CENTRO INTEGRADO DE TELEMATICA DO EXERCITO

# Estudo Técnico Preliminar 60/2026

## 1. Informações Básicas

Número do processo: 64222.004562/2026-48

## 2. Descrição da necessidade

### 2.1. Contexto no qual se insere a contratação

2.1.1. O Centro Integrado de Telemática do Exército (CITEx) tem como missões previstas no seu regulamento aprovado pela Portaria 077-Cmt Ex, de 24 JAN 2019, entre outras, a de:

> "*I – **prover, gerenciar e manter a infraestrutura física e lógica de Tecnologia da Informação e Comunicações** (TIC) do Sistema de Informação do Exército (SINFOEx) no que diz respeito ao funcionamento oportuno e seguro do Sistema Estratégico de Comando e Controle do Exército (SEC²Ex), da hospedagem dos sistemas corporativos e de outros sistemas computacionais sob sua responsabilidade;*

*IV – **prover os meios necessários e adequados de infraestrutura de TIC para a conexão das organizações militares** (OM) à rede corporativa do Exército, visando, principalmente, ao acesso seguro ao SINFOEx;*" (grifo nosso)

**2.1.2.** O CITEx e suas Organizações Militares Diretamente Subordinadas (OMDS) – 7 Centros de Telemática de Área (CTA), 5 Centros de Telemática (CT) constituem o Sistema de Telemática do Exército (SisTEx). O SisTEx contempla, também, o conjunto de pessoas, processos, ferramentas, subsistemas e demais elementos voltados para a gestão da infraestrutura física e lógica de Tecnologia da Informação e Comunicações (TIC) do Exército Brasileiro (EB).

**2.1.3**. Na condição de órgão central do SisTEx, compete ao CITEx planejar, coordenar, integrar, supervisionar e direcionar tecnicamente a infraestrutura de TIC do Exército Brasileiro, incluindo as atividades relacionadas à padronização tecnológica, racionalização da infraestrutura computacional, governança da TIC e condução das aquisições estratégicas destinadas ao atendimento corporativo do Exército Brasileiro.

**2.1.4.** A operação conduzida pelo SisTEx assenta-se em 3 vertentes: a Rede de Dados Corporativa do Exército (EBNet), a Rede Integrada de Telecomunicação do Exército (RITEx) e a Rede Rádio Fixa (RRF). A EBNet e a RITEx são redes cruciais para as atividades administrativas e operacionais do EB. A RRF complementa as duas redes anteriores, ampliando a capilaridade do sistema e sua resiliência.

**2.1.5**. A EBNet é composta pelo BackBone de Integração (BBI), pelos Backbones Regionais (BBR), e pelas Redes Metropolitanas implantadas nas várias guarnições do EB. Atualmente são mais de 1200

enlaces de dados providos por rádio transceptores, fibra óptica, enlaces metroethernet e túneis criptografados sobre a Internet. Esses enlaces são contratados, próprios ou de parceiros e conectam mais de 670 Organizações Militares (OM) do EB, incluindo-se os Pelotões Especiais de Fronteira (PEF).

**2.1.6**. Os PEF são, via de regra, destacamentos militares situados em regiões afastadas e, muitas vezes, isoladas, cuja missão precípua é a proteção de áreas de fronteiras. A conexão de unidades com essa característica se mostra desafiadora, ao mesmo tempo que é vital para que sejam cumpridas adequadamente as missões que lhes incumbem. Ao SisTEx, compete prover a conectividade de voz e dados e o acesso aos sistemas corporativos do EB, mesmo para esses casos.

**2.1.7**. Além desses, diversas outras instituições que atuam em conjunto com o Exército, entre elas, o Comando de Operações Aeroespaciais (COMAE) e o Ministério da Defesa (MD), estão interligadas à EBNet. Durante operações, como as de Garantia da Lei e da Ordem (GLO) ou Garantia de Votação e Apuração (GVA), outras agências podem operar integradas, cabendo ao SisTEx prover acesso aos meios de comunicação de dados e voz.

**2.1.8.** A EBNet inclui, ainda, a infraestrutura da Rede Vitória-Régia, composta por cabos de fibra óptica subfluviais, lançados na calha do Rio Amazonas e do Rio Negro. Além de integrar as OM, essa rede fornece conectividade e serviços para as comunidades, escolas e hospitais da região Amazônica, por meio de parceria com o Ministério das Comunicações, o Governo do Estado do Amazonas, a Telebrás e com as prefeituras locais.

**2.1.9.** Compõem, ainda, a infraestrutura de TIC do SisTEx 2 (dois) Data Centers – um operado pelo 7º CTA e outro pelo 3° CTA – e mais 10 ambientes de hospedagem ou salas de servidores operadas pelos 10 CTA/CT. Essa estrutura forma o ambiente de hospedagem do sistema, que abriga páginas Web de OM e diversos sistemas corporativos e regionais. Toda essa estrutura atua em apoio a sistemas de informação empregados nas operações e na vida administrativa, como o Sistema Eletrônico de Protocolo de Documentos (SPED) e o EBMail (serviço de correio eletrônico corporativo).

**2.1.10.** Embora o SisTEx seja composto por diversas Organizações Militares distribuídas em todo o território nacional, as aquisições corporativas de infraestrutura de TIC vêm sendo conduzidas, em regra, pelo CITEx, 7º CTA e 3º CTA, Organizações Militares que concentram capacidades administrativas e operacionais voltadas à sustentação da infraestrutura corporativa do sistema.

**2.1.11.** Não obstante a autonomia administrativa dessas Organizações Militares, as aquisições estratégicas de TIC realizadas em benefício do SisTEx permanecem subordinadas à coordenação, orientação técnica e aprovação do CITEx, na qualidade de órgão central do sistema e autoridade responsável pela governança da infraestrutura corporativa de TIC do Exército Brasileiro.

**2.1.12.** Dessa forma, as contratações conduzidas pelo CITEx, 7º CTA ou 3º CTA podem atender demandas corporativas do SisTEx e do Exército Brasileiro como um todo, desde que realizadas conforme planejamento, coordenação e direcionamento técnico do CITEx, observando-se a arquitetura institucional, a padronização tecnológica e as diretrizes técnicas estabelecidas para o ambiente corporativo do Exército Brasileiro.

**2.1.13.** Destaca-se que os Data Centers do Exército Brasileiro (Data Center Ricardo Franco e Data Center Marechal Rondon) são Data Centers com certificação TIER III, registrada pelo Uptime Institute. Essa certificação exige uma infraestrutura de refrigeração, conectividade e alimentação elétrica que garanta disponibilidade mínima de 99,982% aos sistemas hospedados.

**2.1.14.** Nos Data Centers e nas salas de servidores do SisTEx ficam instalados, também, equipamentos e soluções de processamento, armazenamento, backup e segurança que permitem a

hospedagem dos diversos sistemas corporativos ou regionais. Tais sistemas são utilizados pelas OM no cumprimento das suas missões nos níveis tático, operacional ou estratégico.

**2.1.15.** Em razão da natureza corporativa da infraestrutura operada pelo SisTEx, os equipamentos de processamento, armazenamento, backup, segurança e conectividade adquiridos no âmbito dessas contratações não se destinam exclusivamente à Organização Militar contratante, mas ao atendimento integrado das necessidades operacionais, administrativas e estratégicas do Exército Brasileiro, incluindo sistemas corporativos, serviços institucionais e estruturas compartilhadas hospedadas nos ambientes centrais e distribuídos do SisTEx.

**2.1.16.** Outra rede operada pelo SisTEx e de grande importância para o EB é a RITEx. A RITEx permite a interligação por voz entre as diversas OM do EB. Ela atende praticamente todo o efetivo do Exército, através dos cerca de 30 mil terminais telefônicos dispersos em todo o Brasil, proporcionando relevante economia de recursos. Toda a comunicação telefônica da RITEx é realizada pela infraestrutura de TIC do EB. A indisponibilidade dessa rede gera ônus imediato, uma vez que os usuários são, nesse caso, obrigados a utilizar o serviço de telefonia privada para contactar outros usuários do sistema.

**2.1.17.** No Data Center Ricardo Franco do 7º CTA está hospedado o sítio Alfa da solução central de telefonia VoIP da RITEx. Há um sítio redundante (Bravo) hospedado no Data Center Marechal Rondon do 3º CTA, que funciona em paralelo ao sítio Alfa.

**2.1.18.** Considerando que o SisTEx constitui sistema corporativo essencial ao funcionamento das capacidades administrativas, operacionais e estratégicas do Exército Brasileiro, as aquisições de infraestrutura de TIC realizadas sob coordenação do CITEx possuem caráter institucional e corporativo, destinando-se ao atendimento integrado das Organizações Militares do EB, observadas as competências técnicas, operacionais e administrativas estabelecidas no âmbito do sistema.

### 2.2. Descrição da Necessidade

**2.2.1.** Aquisição de Solução de Hardware e Softwares para processamento de Inteligência Artificial com GPU para o Exército Brasileiro.

**2.2.2.** O Estudo Técnico Preliminar tem por objetivo identificar e analisar os cenários para o atendimento da demanda constante do Documento de Formalização da Demanda nº 128/2026, da DPI /CITEx, que trata de solução Hardware e Softwares para processamento de Inteligência Artificial com GPU composta por equipamentos, licenças, serviço de instalação, operação assistida, e demais serviços correlatos, para atender demandas Exército Brasileiro em relação ao processamento de Inteligência Artificial (IA).

**2.2.3.** A presente aquisição tem por base o Plano de Contratações Anual elaborado no ano de 2025, com execução prevista no ano de 2026, materializado pelo Documentos de Formalização da Demanda de número DFD 128/2026.

## 3. Área requisitante

| Área Requisitante | Responsável |
|---|---|
| Divisão de Projetos e Inovação (DPI) | Anderson Lima Muniz Barretto |

## 4. Necessidades de Negócio

**4.1.** A presente contratação decorre da necessidade de o Exército Brasileiro ampliar sua capacidade institucional de processamento computacional acelerado para atendimento às demandas crescentes de inteligência artificial, em ambiente corporativo, seguro, escalável e sob governança.

**4.2.** O objetivo não é apenas adquirir servidores com GPU de forma isolada, mas estruturar uma solução de GPU como serviço, apta a disponibilizar recursos computacionais de alto desempenho de maneira centralizada, compartilhada, controlada e sob demanda para diferentes usuários, equipes e projetos institucionais.

**4.3.** A necessidade de negócio está diretamente relacionada ao aumento do consumo de recursos de Inteligência Artificial no âmbito do Exército, abrangendo aplicações administrativas, analíticas, de apoio à decisão, de processamento documental, de inferência de modelos, de treinamento e de ambientes interativos de desenvolvimento.

**4.4.** A infraestrutura atualmente disponível no Data Center do Exército já contempla servidores voltados à inteligência artificial, porém a capacidade instalada ainda é insuficiente para atender, de forma estruturada, institucionalizada e escalável, ao crescimento das demandas identificadas.

**4.5.** Nesse contexto, a evolução pretendida não se resume à expansão pontual de hardware, mas à constituição de uma plataforma corporativa capaz de transformar a capacidade de processamento em GPU em serviço efetivamente consumível pela Administração.

**4.6.** Sob a ótica do negócio, a contratação busca atender à necessidade de prover uma base tecnológica que permita ao Exército iniciar e expandir o uso de soluções de inteligência artificial com maior autonomia, segurança e soberania sobre dados, modelos e processamento.

**4.7.** A disponibilização dessa capacidade em ambiente próprio contribui para reduzir dependência de soluções fragmentadas, favorecer a padronização do uso institucional de inteligência artificial, ampliar a governança sobre o consumo de recursos computacionais e criar condições para que diferentes iniciativas possam ser desenvolvidas e operadas sobre uma infraestrutura comum, com critérios de controle de acesso, auditoria, observabilidade, cotas, priorização e compartilhamento seguro.

**4.8.** A necessidade também se justifica pela diretriz institucional de modernização da infraestrutura de tecnologia da informação e pela evolução do contexto de transformação digital do Exército.

**4.9.** O tema foi apreciado em instância formal de governança, no âmbito da 4ª Reunião do Comitê de Governança Digital, ocasião em que foi destacada a necessidade de ampliação da capacidade computacional voltada à inteligência artificial, bem como sua aderência às diretrizes estratégicas de evolução tecnológica e suporte a novas capacidades digitais.

**4.10.** Assim, a demanda não decorre de iniciativa isolada, mas de encaminhamento institucional legitimado e associado a prioridades administrativas já reconhecidas.

**4.11.** Além disso, a necessidade de negócio exige que a solução seja concebida com lógica de implantação progressiva.

**4.12.** A solução deverá iniciar por um cenário piloto, mas já com características funcionais, operacionais e arquiteturais compatíveis com crescimento futuro, inclusive para dezenas ou centenas de GPUs.

**4.13.** Essa diretriz revela que a necessidade não é apenas suprir uma carência imediata de processamento, mas construir uma base tecnológica escalável, que produza valor prático no curto prazo e, simultaneamente, permita amadurecimento do modelo operacional e expansão ordenada conforme a evolução da demanda institucional e da disponibilidade de recursos.

**4.14.** Em síntese, a necessidade de negócio consiste em dotar o Exército de uma plataforma corporativa de GPU como serviço capaz de sustentar, de forma segura, padronizada, escalável e governada, o uso institucional de inteligência artificial, atendendo à demanda atual reprimida, preparando o ambiente para o crescimento futuro e fortalecendo a capacidade da Administração de desenvolver, operar e expandir soluções de inteligência artificial em consonância com seus objetivos estratégicos e operacionais.

**4.15.** Em complemento às necessidades de negócio acima descritas, registra-se a existência de previsão de recurso orçamentário para o corrente exercício, conforme DIEx nº 18835-SEGER/6 SCh /EME, de 9 de julho de 2026, do Estado-Maior do Exército, que tratou da solicitação de Previsão de Recurso Orçamentário para aquisição de Solução de Infraestrutura de Inteligência Artificial destinada ao CITEx.

**4.16.** O referido expediente informa que o Estado-Maior do Exército analisou o pleito sob a ótica orçamentária atual e ratificou a Previsão de Recurso Orçamentário (PRO) de R$ 2.000.000,00 (dois milhões de reais), anteriormente concedida por meio do DIEx nº 6373-SEGER/6 SCh/EME, de 10 de março de 2026, destinada ao robustecimento do Data Center Corporativo do Exército, no escopo da Ação Orçamentária 156M.

**4.17.** O mesmo DIEx registra que o pleito complementar de R$ 20.000.000,00 foi reconhecido como iniciativa relevante para a ampliação da capacidade de processamento dedicado à Inteligência Artificial no âmbito do Exército Brasileiro, embora não tenha sido possível a emissão de nova Previsão de Recurso Orçamentário no exercício financeiro de 2026, em razão das restrições impostas à programação orçamentária e financeira daquele exercício.

**4.18.** Posteriormente, por meio do DIEx nº 20668-SEGER/6 SCh/EME, de 27 de julho de 2026, o Estado-Maior do Exército informou a existência de Previsão de Recursos Orçamentários para o exercício de 2027, destinada à ampliação do ambiente de hospedagem do Data Center Corporativo do Exército, no valor total de R$ 40.000.000,00 (quarenta milhões de reais), dos quais R$ 30.000.000,00 estão previstos no âmbito da Ação Orçamentária 156M e R$ 10.000.000,00 no âmbito da Ação Orçamentária 147F, Plano Orçamentário 0001.

**4.19.** Dessa forma, a contratação conta com previsão orçamentária de R$ 2.000.000,00 para o exercício de 2026 e com previsão adicional de recursos para o exercício de 2027 em montante suficiente para suportar a continuidade e a conclusão das etapas de maior vulto financeiro da solução, cuja estimativa global é de aproximadamente R$ 20.000.000,00.

**4.20.** Sob o ponto de vista normativo, a existência de crédito suficiente para fazer frente às parcelas contratuais vincendas no exercício da contratação atende à lógica do art. 150 da Lei nº 14.133/2021, segundo o qual nenhuma contratação será feita sem a caracterização adequada do objeto e sem a indicação dos créditos orçamentários para pagamento das parcelas contratuais vincendas no exercício em que for realizada a contratação.

**4.21.** Também deve ser observado o art. 105 da Lei nº 14.133/2021, que prevê que a duração dos contratos será a prevista em edital e que deverão ser observadas, no momento da contratação e a cada exercício financeiro, a disponibilidade de créditos orçamentários, bem como a previsão no plano plurianual quando a contratação ultrapassar um exercício financeiro.

**4.22.** Assim, a existência de Previsão de Recurso Orçamentário no valor de R$ 2.000.000,00 para o exercício de 2026, complementada pela previsão de recursos para o exercício de 2027, permite a continuidade da instrução processual e a realização do procedimento licitatório, devendo o edital, o contrato e o planejamento da execução prever cronograma físico-financeiro compatível com os créditos disponíveis em cada exercício.

**4.23.** A contratação poderá, portanto, ser estruturada com início da execução ainda no exercício de 2026 e continuidade da execução e dos pagamentos no exercício de 2027, observando-se, em cada exercício financeiro, a existência dos créditos orçamentários necessários para suportar as respectivas parcelas vincendas.

**4.24.** A realização da despesa deverá observar o art. 60 da Lei nº 4.320/1964, segundo o qual é vedada a realização de despesa sem prévio empenho. Assim, a emissão dos empenhos deverá ocorrer previamente à assunção das respectivas despesas, na medida dos créditos disponíveis e conforme o cronograma de execução contratual.

**4.25.** A modalidade de licitação a ser adotada poderá ser o pregão, desde que os bens e serviços a serem contratados sejam caracterizados como comuns, com padrões de desempenho e qualidade objetivamente definidos no edital por meio de especificações usuais de mercado, conforme o regime da Lei nº 14.133/2021. A própria Lei nº 14.133/2021 prevê a adoção do pregão sempre que o objeto possuir padrões de desempenho e qualidade passíveis de definição objetiva pelo edital.

**4.26.** Nesse contexto, a existência de previsão orçamentária para os exercícios de 2026 e 2027 permite o prosseguimento do processo de contratação por meio de pregão, desde que a contratação seja planejada de forma compatível com a execução prevista em cada exercício e com a efetiva disponibilidade dos respectivos créditos orçamentários.

**4.27.** A continuidade do processo de contratação mostra-se, portanto, compatível com a necessidade institucional de iniciar a implantação progressiva da solução de GPU como serviço. Tal estratégia permite à Administração preservar o planejamento, avançar na seleção da solução mais vantajosa, iniciar a execução no exercício corrente e estruturar a execução subsequente conforme o planejamento orçamentário existente.

**4.28.** Dessa forma, as Previsões de Recursos Orçamentários informadas pelo Estado-Maior do Exército constituem elementos suficientes para fundamentar o prosseguimento do processo de contratação, sem prejuízo de que a efetiva contratação, emissão de empenhos, execução e pagamentos observem, em cada exercício financeiro, os créditos orçamentários efetivamente disponíveis e as demais exigências legais aplicáveis.

**4.29.** A presente demanda, portanto, permanece tecnicamente justificada e possui planejamento orçamentário compatível com sua execução, devendo a Administração ajustar o cronograma de contratação, execução, empenho e pagamento às disponibilidades de recursos previstas para os exercícios de 2026 e 2027.

**4.30.** A disponibilização de Previsões de Recursos Orçamentários relacionadas à solução também reforça a opção administrativa pela realização de procedimento licitatório destinado à contratação

efetiva do objeto, e não à formação de ata de registro de preços. No caso concreto, a Administração não busca apenas registrar preços para eventual contratação futura, mas selecionar fornecedor para viabilizar a implantação da solução de infraestrutura de Inteligência Artificial, com início da execução no exercício de 2026 e continuidade no exercício de 2027.

**4.31.** O Sistema de Registro de Preços, regulamentado pelo Decreto nº 11.462/2023, que disciplina os arts. 82 a 86 da Lei nº 14.133/2021, constitui instrumento adequado para situações em que a Administração pretende registrar preços para contratações futuras, com possibilidade de contratações parceladas, demandas recorrentes, atendimento a múltiplos órgãos ou incerteza quanto ao quantitativo a ser efetivamente contratado. No presente caso, entretanto, a existência de previsão orçamentária, a definição do escopo da solução e a intenção concreta de execução indicam maior aderência à contratação ordinária, por pregão, sem formação de ata de registro de preços.

**4.32.** Essa opção tende a conferir maior previsibilidade à contratação, uma vez que sinaliza ao mercado a intenção concreta da Administração de contratar e iniciar a execução do objeto. Diferentemente da ata de registro de preços, que gera expectativa de contratação futura e pode ser executada em montante inferior ao registrado, a contratação ordinária permite vincular de forma mais clara o objeto, o cronograma de execução, o cronograma físico-financeiro, as obrigações contratuais e as condições de pagamento.

**4.33.** Tal previsibilidade é especialmente relevante no mercado de infraestrutura de Inteligência Artificial, caracterizado por alta volatilidade de preços, elevada demanda internacional por GPUs, prazos de fornecimento sujeitos a disponibilidade global, variações cambiais, dependência de componentes importados e rápida evolução tecnológica. Nesse cenário, a realização de licitação voltada à contratação efetiva, em vez da formação de ata de registro de preços, tende a aumentar a segurança comercial do fornecedor quanto à existência de demanda concreta, podendo favorecer a apresentação de melhores condições comerciais, prazos mais realistas e maior compromisso com a alocação de equipamentos, licenças, serviços especializados e suporte.

**4.34.** Além disso, a não adoção do Sistema de Registro de Preços reduz o risco de que quantitativos registrados não venham a ser efetivamente contratados, circunstância que pode diminuir a confiança dos fornecedores nas estimativas da Administração e afetar negativamente a precificação. O Tribunal de Contas da União registra, entre os riscos associados ao uso inadequado de atas de registro de preços, a possibilidade de atas com quantitativos muito superiores aos efetivamente contratados, levando fornecedores a perderem confiabilidade nas estimativas e a deixarem de considerar ganhos de escala em suas propostas, com consequente elevação de preços em futuros registros.

**4.35.** No presente caso, a solução pretendida possui escopo técnico integrado, alta complexidade, necessidade de implantação coordenada, serviços especializados, suporte, validação de arquitetura e execução compatível com cronograma próprio. Essas características reforçam a conveniência de contratação específica, com obrigações claramente vinculadas ao projeto, em vez de registro de preços para contratações futuras e incertas.

**4.36.** Assim, a existência de PRO no valor de R$ 2.000.000,00 para o exercício de 2026, associada à Previsão de Recursos Orçamentários para 2027 no valor total de R$ 40.000.000,00 destinada à ampliação do ambiente de hospedagem do Data Center Corporativo do Exército, constitui elemento concreto de suporte ao planejamento da contratação e reforça a viabilidade orçamentária da execução do projeto, justificando a condução do processo por pregão sem utilização de ata de registro de preços.

**4.37.** Registra-se, contudo, que as previsões orçamentárias deverão ser confirmadas pela Administração antes da assunção das correspondentes obrigações, uma vez que a Previsão de Recurso Orçamentário não se confunde com a efetiva disponibilização do crédito orçamentário.

**4.38.** Dessa forma, antes da publicação do edital e da emissão dos respectivos empenhos, deverá ser verificada a situação atualizada das previsões orçamentárias e a correspondente disponibilidade dos créditos necessários à execução das parcelas previstas para cada exercício.

**4.39.** Essa reavaliação é necessária para assegurar a compatibilidade entre o planejamento da contratação, o cronograma físico-financeiro e os créditos efetivamente disponibilizados. Eventual alteração na disponibilidade de recursos deverá ser considerada no ajuste final da estratégia de contratação, da execução e dos valores a serem empenhados em cada exercício.

**4.40.** Caso haja alteração das disponibilidades orçamentárias antes da fase externa ou durante a execução contratual, a Administração deverá avaliar seus efeitos sobre o cronograma de execução e de empenho, preservados o objeto, a motivação técnica, a vantajosidade e a regularidade do procedimento.

**4.41.** Em síntese, a contratação possui indicação orçamentária compatível com o valor global estimado do projeto, apoiada na PRO de R$ 2.000.000,00 para o exercício de 2026 e na Previsão de Recursos Orçamentários para 2027, que alcança R$ 40.000.000,00 para ampliação do ambiente de hospedagem do Data Center Corporativo do Exército. Esses instrumentos permitem o prosseguimento da instrução e do procedimento licitatório, permanecendo a assunção das obrigações, a emissão dos empenhos e a realização dos pagamentos condicionadas à efetiva disponibilização dos créditos e à observância do cronograma físico-financeiro da contratação.

## 5. Necessidades Tecnológicas

**5.1.** A solução a ser contratada deverá atender às necessidades tecnológicas da Administração de forma integrada, contemplando, em conjunto, componentes de hardware, software, conectividade, armazenamento e serviços especializados, em quantidade e capacidade compatíveis com a implantação inicial da plataforma e com sua evolução futura. A necessidade tecnológica não se limita à aquisição isolada de equipamentos, mas à composição de um ambiente corporativo capaz de disponibilizar processamento em GPU como serviço, com segurança, escalabilidade, governança, observabilidade e condições efetivas de operação institucional.

**5.2.** A solução deverá ser capaz de sustentar, desde a etapa inicial, a prestação efetiva do serviço de GPU sob demanda, assegurando mecanismos de provisionamento, compartilhamento seguro da infraestrutura, definição de cotas, priorização de cargas de trabalho, execução de workloads interativos e em lote, observabilidade, controle de acesso e administração continuada do ambiente. Em razão disso, as necessidades tecnológicas da contratação deverão ser compreendidas como requisitos de plataforma, e não apenas como especificações de desempenho pontual de servidores ou aceleradores.

**5.3.** No plano de hardware, a contratação deverá contemplar infraestrutura de processamento acelerado compatível com workloads de inteligência artificial, inferência, treinamento, ambientes interativos e demais usos previstos para a plataforma. Essa infraestrutura deverá ser acompanhada por recursos computacionais auxiliares destinados ao funcionamento da camada de software, bem como por rede e armazenamento compatíveis com os requisitos de desempenho, disponibilidade, expansão e integração institucional.

**5.4.** No plano de software, a solução deverá contemplar plataforma corporativa apta a prover virtualização, execução de contêineres, orquestração de workloads, gerenciamento de cluster, habilitação e operação de aceleradores gráficos, interface de serviços e autosserviço, além de mecanismos de observabilidade e segurança operacional. Os componentes de software deverão ser compatíveis entre si e suficientes para transformar a infraestrutura acelerada em serviço institucional consumível, observando-se a necessidade de integração entre a camada de processamento, as interfaces de uso, a camada de operações e os elementos de administração da solução.

**5.5.** As necessidades tecnológicas também abrangem a disponibilização de diferentes formas de acesso e consumo da plataforma, de modo a atender perfis distintos de usuários e cenários diversos de operação. Nesse sentido, a solução deverá comportar interfaces compatíveis com acesso por portal web, APIs, automação, ambientes interativos, execução de workloads em contêineres e demais formas de consumo necessárias à utilização da capacidade de processamento em GPU como serviço. Tal diretriz decorre da necessidade de que a solução seja efetivamente utilizável por perfis técnicos e operacionais distintos, e não apenas administrável pela equipe de infraestrutura.

**5.6.** A contratação deverá, ainda, contemplar necessidades tecnológicas relacionadas ao armazenamento e à conectividade da solução. A infraestrutura de armazenamento deverá ser compatível com datasets, imagens, artefatos, modelos, checkpoints, logs, notebooks, bases documentais e demais dados persistentes da plataforma, devendo atender a requisitos de desempenho, capacidade, persistência, disponibilidade e escalabilidade. Da mesma forma, a solução deverá prever redes de frontend, backend e acesso a dados, de forma aderente ao modelo arquitetural da plataforma e às necessidades de comunicação entre usuários, serviços, servidores, aceleradores e armazenamento.

**5.7.** As necessidades tecnológicas da contratação deverão observar as diretrizes institucionais de padronização, interoperabilidade e aderência ao ambiente existente. Em especial, os componentes de software deverão estar em conformidade com a padronização tecnológica adotada pelo órgão e compatíveis com as referências de especificação e desempenho aplicáveis aos processos de contratação pública de bens e serviços de TIC. Da mesma forma, a solução deverá guardar aderência ao planejamento institucional de TIC do CITEx, de modo a contribuir para a evolução coordenada da infraestrutura, da governança e da capacidade computacional do ambiente corporativo.

**5.8.** Também deverão ser observadas necessidades tecnológicas relacionadas à mitigação de dependência excessiva de fornecedor, à preservação da interoperabilidade da camada de software com os demais componentes da arquitetura e à possibilidade de expansão futura sem redesign estrutural, ressalvadas as indicações excepcionais de marca e produto formalmente justificadas no item 7.4.2.

**5.9.** A contratação deverá contemplar, ainda, necessidades tecnológicas de implantação, estabilização, suporte e evolução da solução. Por essa razão, não bastará a aquisição de bens e licenças de software, sendo igualmente necessários serviços especializados de implantação e configuração, capacitação e transferência de conhecimento, operação assistida, suporte e sustentação técnica e atualização tecnológica das equipes da Administração. Tais elementos são indispensáveis para assegurar que a solução entre em produção com condições adequadas de uso e que permaneça operacionalmente sustentável ao longo do tempo.

**5.10.** Em síntese, as necessidades tecnológicas da presente contratação correspondem à disponibilização de uma solução completa e coerente de GPU como serviço, composta por

infraestrutura de processamento, camada lógica de operação, conectividade, armazenamento e serviços especializados, todos compatíveis com a estratégia de implantação progressiva adotada pela Administração.

## 6. Demais requisitos necessários e suficientes à escolha da solução de TIC

### 6.1. Requisitos Técnicos

**6.1.1.** Aquisição de equipamentos, softwares e serviços especializados com o propósito específico de disponibilizar uma solução de GPU como serviço para o Exército Brasileiro, contemplando capacidade computacional acelerada, orquestração, autosserviço, observabilidade, armazenamento, conectividade e sustentação operacional, de forma a atender aos objetivos de padronização, escalabilidade, governança, segurança e continuidade dos serviços corporativos baseados em inteligência artificial.

### 6.2. Requisito de Capacitação

**6.2.1.** A CONTRATADA deverá providenciar treinamento hands-on para toda a solução ofertada, contemplando as melhores práticas de implantação, operação, administração, observabilidade, governança e uso da plataforma de GPU como serviço, abrangendo os equipamentos, softwares e serviços adquiridos..

**6.2.2.** A capacitação deverá ter caráter teórico-prático dirigido, principalmente, ao contexto de atuação da equipe técnica do SisTEx, abrangendo administração da plataforma, operação do ambiente acelerado, gerenciamento de workloads, observabilidade, segurança, uso da camada de autosserviço e sustentação da solução, com carga horária, quantitativos de participantes e modalidade de execução definidos nas especificações técnicas.

**6.2.3.** Os serviços de treinamento deverão ser realizados de segunda a quinta-feira, entre 9h e 15:30h, nas dependências do CITEx.

**6.2.4.** É da total responsabilidade da CONTRATADA a disponibilização de treinamentos específicos para utilização da solução completa, de acordo com as possibilidades apresentadas pelos fabricantes dos componentes envolvidos, de forma que seja possível realizar operação eficiente, administração segura e sustentação adequada da plataforma, mantendo os níveis de disponibilidade, governança e desempenho necessários.

**6.2.5.** As demais condições de treinamento, assim como os locais estão disponíveis nas especificações técnicas, Anexo I a este Estudo.

**6.2.6.** Em atendimento ao previsto no Inciso X do § 1º do Art. 18 da Lei nº 14.133, de 2021, que trata sobre a capacitação de servidores ou de empregados para fiscalização e gestão contratual, os militares envolvidos na fiscalização e gestão contratual deverão obrigatoriamente realizar os cursos de Fiscalização e Gestão de Contratos de TIC, Gestão e Fiscalização de Contratos Administrativos e Gestão de Risco no Setor Público, todos disponíveis no endereço eletrônico: https://www.escolavirtual.gov.br, ou no Portal de Educação do Exército, no endereço eletrônico: https://portaldeeducacao.eb.mil.br/academico.

### 6.3. Requisito Legais

**6.3.1.** A contratação está alinhada com as políticas de gestão e com o planejamento de Tecnologia da Informação e Comunicações adotados pelo Exército Brasileiro, especialmente no âmbito do CITEx, em

consonância com as diretrizes de modernização da infraestrutura, transformação digital, ampliação do uso corporativo de inteligência artificial e fortalecimento da capacidade computacional institucional.

**6.3.2.** A padronização a que se destina esta aquisição considerando o SisTEx como um todo, está alinhada com o previsto no Art. 40, Inciso V letra a) da Lei Nº 14,133, de 01 de abril de 2021, reproduzido a seguir:

> **"Art. 40. O planejamento de compras deverá considerar a expectativa de consumo anual e observar o seguinte:**
>
> **V - atendimento aos princípios:**
>
> **a) da padronização, considerada a compatibilidade de especificações estéticas, técnicas ou de desempenho;"**

**6.3.3.** A presente contratação trata-se de uma aquisição de TIC, considerando o inciso VII do art. 2º da Instrução Normativa SGD/ME nº 94, de 23 de dezembro de 2022, enquadrando-se nos itens 1.5 e 1.6 do Anexo II da referida instrução. O presente Estudo Técnico Preliminar foi confeccionado atendendo às vedações impostas pelo Art. 5º e seus Incisos da Instrução Normativa SGD/ME nº 94, de 23 de dezembro de 2022. A contratação não incide em nenhum dos casos vedados.

**6.3.4.** O presente processo de contratação foi elaborado em atendimento a Instrução Normativa SGD /ME nº 94, de 23 de dezembro de 2022. Em se tratando do Anexo I da referida portaria, a contratação não trata de Solução de autenticação para serviços públicos digitais, Serviços de desenvolvimento, sustentação e manutenção de *softwar*e, Infraestrutura de centro de dados, serviços em nuvem, sala-cofre ou sala segura, Contratação de empresas públicas de tecnologia da informação e comunicação e Serviços de desenvolvimento, sustentação e manutenção de portais na *internet*.

**6.3.4.1.** Em relação ao item Aquisições de ativos de tecnologia da Informação e Comunicação, previsto no item 8 do Anexo I da Instrução Normativa SGD/ME nº 94, de 23 de dezembro de 2022, os itens previstos nessa contratação não constam nos itens que incidem a aplicação da Portaria INMETRO / MDIC número 304 de 06/11/2023.

**6.3.4.2.** Em relação ao item Licenciamento de *software* e serviços agregados, previsto no item 8 do Anexo I da Instrução Normativa SGD/ME nº 94, de 23 de dezembro de 2022, o licenciamento a ser adquirido será específico para os componentes de software que compõem a solução de GPU como serviço, incluindo, conforme aplicável, orquestração, aceleração por GPU, interface de serviços, observabilidade, segurança operacional e demais elementos necessários ao funcionamento da plataforma, e foram considerados os pontos do Anexo I adequados a esse cenário.

**6.3.5.** Esta contratação não incide nas hipóteses vedadas pelos artigos 3º e 4º da Instrução Normativa SGD/ME nº 94, de 23 de dezembro de 2022.

**6.3.6.** A presente aquisição não apresenta outras contratações correlatas e/ou interdependentes, por se tratar da aquisição de uma solução completa que possibilitará a contratação de equipamentos *softwares* e serviços de forma a atender aos objetivos a serem alcançados.

**6.3.7.** A presente contratação se enquadra como uma Aquisição de TIC (Tecnologia da Informação e Comunicações), não sendo necessária a realização de serviços profissionais referentes à Engenharia. Dessa forma, não se faz necessária a exigência de Anotação de responsabilidade Técnica (ART).

**6.3.8.** A presente contratação não contempla itens previstos no Catálogos de Soluções de TIC, previsto no § 2º do Art. 43 da Lei nº 14.133, de 2021.

**6.3.9.** Este Estudo Técnico Preliminar foi confeccionado considerando o que é previsto no Inciso II do Art. 9 e Art. 11º e seus Incisos da Instrução Normativa SGD/ME nº 94, de 23 de dezembro de 2022, assim como o previsto no Art. 18º da Lei nº 14.133, de 2021, no que couber.

**6.3.10.** Este artefato foi elaborado de forma digital, utilizando sistema específico disponibilizado pela Secretaria de Governo Digital do Ministério da Ministério da Gestão e da Inovação em Serviços Públicos, em atendimento ao previsto no § 7º do Art. 9º da Instrução Normativa SGD/ME nº 94, de 23 de dezembro de 2022.

**6.3.11.** As especificações técnicas geradas neste Estudo Técnico Preliminar foram confeccionadas atendendo ao que prevê o Art. 16º da Instrução Normativa SGD/ME nº 94, de 23 de dezembro de 2022.

**6.3.12.** O presente Estudo Técnico considerou, no que era adequado, os Requisitos e Obrigações Quanto a Segurança da Informação e Privacidade, contidos no Item 7 do Anexo I da Instrução Normativa SGD/ME nº 94, de 23 de dezembro de 2022.

**6.3.13.** A presente contratação atende ao previsto no Inciso I do Art. 3º da Instrução Normativa SGD /ME nº 94, de 23 de dezembro de 2022, e trata de uma única Solução de TIC, mesmo que seja composta por vários itens, de forma a possibilitar o atendimento aos objetivos propostos na contratação.

**6.3.14.** Este Estudo Técnico Preliminar foi confeccionado consultando o modelo disponível no endereço https://www.gov.br/governodigital/pt-br/contratacoes/templates-e-listas-de-verificacao, que contém templates e listas de verificação elaborados pelo órgão central do SISP, em atendimento ao previsto no § 2º do Inciso III do Art. 8º da Instrução Normativa SGD/ME nº 94, de 23 de dezembro de 2022, regido pela Lei nº 14.133, de 2021. Adicionalmente, este documento foi elaborado por meio do Sistema ETP Digital, conforme o previsto na Instrução Normativa Nº 58, de 8 de agosto de 2022 da Secretaria de Governo Digital do Ministério da Ministério da Gestão e da Inovação em Serviços Públicos, considerando o previsto no seu Art. 4º.

**6.3.15.** Considerando a Estratégia Federal de Governo Digital, instituída pelo Decreto nº 12.198, de 2024, conforme inciso II do art. 6 da IN SGD/ME nº 94, de 2022, a presente contratação está alinhada aos Incisos IV e VI da Portaria SGD/MGI Nº 6.618, DE 25 DE SETEMBRO DE 2024.

**6.4. Requisito de Manutenção**

**6.4.1. Garantias:**

**6.4.1.1.** Garantia e suporte por período compatível com a vida útil esperada da solução para os equipamentos deverá ser de 60 (sessenta) meses e os Softwares licenciados e serviços atrelados ao funcionamento da plataforma, por um período de 12 (doze) meses. A garantia será conforme o prescrito na Especificação Técnicas dos Itens - Anexo I ao Estudo Técnico Preliminar. A garantia deverá cobrir os defeitos decorrentes de projeto, desenvolvimento e instalação, sem qualquer ônus adicional ao CONTRATANTE.

**6.4.1.2.** Sobre as garantias, as informações encontram-se detalhadas na Especificação Técnicas dos Itens - Anexo I deste Estudo Técnico Preliminar .

**6.4.2.** Considerando os equipamentos que compõem a solução a ser contratada, por se tratar de equipamento e *softwares* a serem aplicados em ambiente de Rede Corporativa, as nomenclaturas de garantia e suporte se confundem e podem trazer interpretação equivocada quanto a real necessidade da administração. Parte dessa confusão é feita pela nomenclatura utilizada pelos fabricantes, a prática de mercado e os termos utilizados nos processos licitatórios.

**6.4.3.** Primeiramente, há que se definir que, pela criticidade e complexidade tecnológica das atividades a que se destinam esses equipamentos, é imperativa a existência de procedimentos para garantir a sua alta disponibilidade, e para tanto serão adotadas medidas como o uso de fontes redundantes. Essas medidas são necessárias, mas não são suficientes, precisando ser complementadas com a continuidade do funcionamento de todo o *hardware* da solução (placas, discos, fontes, conectores, cabos etc.) que podem vir a sofrer falhas imprevistas durante o seu ciclo de vida.

**6.4.4.** Além da questão do *hardware*, os componentes de *software* da solução também têm um processo evolutivo, sujeitos a atualizações de segurança, performance e correção de erros, que porventura podem advir do seu uso normal. Esse é um cenário comum e previsível nesse ambiente.

**6.4.5.** Ressalta-se que a forma de aquisição do item garantia está alinhado com o previsto no Art. 40, Inciso I da Lei Nº 14.133, de 01 de abril de 2021, reproduzido a seguir:

> **"Art. 40. O planejamento de compras deverá considerar a expectativa de consumo anual e observar o seguinte:**
>
> **I - condições de aquisição e pagamento semelhantes às do setor privado;"**

**6.4.6.** A prática de mercado para esses tipos de equipamento é a aquisição de um serviço de garantia que vai proporcionar a cobertura de todo o *hardware* em caso de falhas, bem como o *software*, em caso de falhas e necessidade de correção. Esse serviço é fundamental para garantir a disponibilidade da solução durante o seu ciclo de vida.

**6.4.6.1.** Assim é importante a aquisição deste serviço agregado, cuja prática é reconhecida tanto pela legislação em vigor, quanto pela jurisprudência do TCU, conforme mencionado no PARECER n. 00461/2024/CGAQ/SCGP/CGU/AGU, de 27 de setembro de 2024, manifestação jurídica referente ao processo SRP 90111/2024, da UASG 160091:

> "18. Portanto, aplica-se na íntegra a IN SGD/ME n° 94/2022, uma vez que pretende-se para o item 04 "o licenciamento de software consiste em qualquer forma de aquisição de direitos de uso de software, quer seja por tempo indeterminado (licença perpétua), quer seja por meio de cessão temporária de direito de uso (locação ou subscrição) " (item 1.1 do Anexo I) e serviços agregados para os itens 1, 3 e 4, definidos como "aqueles relacionados ao licenciamento de software, tais como os serviços de atualização de versão, manutenção e suporte técnico" (item 1.2 do mesmo Anexo I).
> 19. Esta norma, foi atualizada em grande parte com base no Acórdão n° 2569/2018 - Plenário (Relator Aroldo Cedraz, Processo 030.236/2016-9), onde realizada auditoria com objetivo avaliar o efeito das práticas comerciais adotadas por grandes fabricantes de software do setor de TI nas contratações realizadas pelo Poder Público.
> 20. Importante para a presente manifestação, o que foi considerado "serviço agregado" no citado Acórdão, cujo conceito foi transferido para a IN 94/2022, nos termos do item 1.2, "d" do Anexo I da IN 94/2022. Ressaltado no referido

> Acórdão que "cabe salientar, por oportuno, que os serviços agregados são normalmente comercializados junto com as licenças na primeira aquisição, quando têm a conotação de "garantia", remetendo-se ao Código de Defesa do Consumidor, sendo a renovação opcional após o fim da vigência do primeiro período contratado . Há variações, mas, normalmente, os três tipos de serviços mencionados podem ser adquiridos como pacote único."

**6.4.7.** A prestação desse serviço deve ser executada por meio de portal do fabricante, onde pode ser consultado o período de cobertura contratado e realizar a abertura de chamados para falhas de funcionamento, substituições de peças defeituosas, consultoria técnica em configuração e solução de problemas e correção de *softwares*.

**6.4.8.** Outro ponto importante para justificar esse tipo de aquisição é a celeridade inerente dos serviços prestados e a origem dos equipamentos. A solução é composta por equipamentos importados de alta complexidade tecnológica e que podem ficar sujeitos a disponibilidade de peças e variações de preços. Um processo licitatório para aquisição de peças de reposição desses equipamentos, além do tempo inerente ao processo, ainda pode se acrescentar de períodos de cerca de 45 (quarenta e cinco) dias úteis, em condições normais, para a importação das peças, gerando indisponibilidades da solução a ser adquirida, o que seria incompatível com as necessidades do SisTEx.

**6.4.9.** Aquisições nesse formato já são uma realidade na administração, onde pode ser tomado como exemplo os itens do Catálogo de Soluções de TIC (https://www.gov.br/governodigital/pt-br/contratacoes/catalogo-de-solucoes-de-tic) para produtos das marcas *VMware*, *Red Hat* e *Microsoft*. No caso da *Red Hat*, por exemplo, o acordo disponibilizado no *site* acima aponta para itens cujo *part number* são de itens com suporte por 36 meses, sendo o pagamento realizado à vista. No caso da *Microsoft*, os itens do catálogo apresentam uma granularidade maior, com itens de *part numbers* distintos com suporte de 12, 24 e 36 meses. Um ponto em comum é que esses itens ofertados no catálogo tratam de pagamento à vista, não existindo a opção de pagamento continuado no formato mensal, reforçando o item 6.4.5, onde o presente estudo considera a prática de mercado.

**6.4.9.1.** No CITEx as aquisições nesse formato também já são uma realidade, tendo como exemplo os processos SRP 8/2022 e SRP 14/2022, SRP 03/2023, SRP 6/2023, SRP 7/2023, SRP 90111/2024, SRP 90113/2024, da UASG 160091 e SRP 11/2022, SRP 3/2023 da UASG 160486.

**6.4.9.2.** Este entendimento também é corroborado pela jurisprudência do TCU, conforme mencionado no PARECER n. 00461/2024/CGAQ/SCGP/CGU/AGU, de 27 de setembro de 2024, manifestação jurídica referente ao processo SRP 90111/2024, da UASG 160091:

> "24. Justificado no ETP que na forma do art. 40, I, da Lei n° 14.133/2021, as compras sempre que possível deverão atender "condições de aquisição e pagamento semelhantes às do setor privado ", motivo pelo qual detalhada a forma de atuação dos grandes fabricantes de software, que exigem pagamento integral do valor da licença (perpétua ou mesmo temporária/locação), mesmo que atualizações e suporte vinculados à cada licença sejam estendidos ao longo do tempo.
> 25. Sobre o tema, o Acórdão n° 2569/2018 - PLENÁRIO reconheceu que de fato é **prática usual de mercado**, em especial para os grandes fabricantes de software, **o pagamento integral do valor da aquisição ou do próprio serviço agregado quando da aquisição**. Não obstante, trouxe algumas ressalvas e considerações sobre o tema, todas incorporadas no Anexo I da IN 94/2022 e que, para as licenças pretendidas, estão todas observadas pela Administração

no caso concreto.

26. Registra-se que a presente licitação será realizada por Sistema de Registro de Preços, de forma que as licenças serão adquiridas em lotes mínimos de 1.000 unidades, até mesmo para autorizar a sua instalação e ativação. Neste ponto, registra-se trecho do voto do Ministro Relator Aroldo Cedraz:

**No caso de licenças de software, entendo que o momento da entrega definitiva é o da ativação da licença**. Consequentemente, o pagamento não pode ser realizado de forma antecipada, sob pena de infringir frontalmente os dispositivos da legislação de licitações e contratos, bem como normas de caráter financeiro (e.g arts. 62 e 63 da Lei 4320/1964)."

**6.4.10.** Outro ponto importante a se considerar é a criticidade da solução de GPU como serviço para o ambiente corporativo do SisTEx, especialmente diante da necessidade de sustentação contínua de workloads institucionais de inteligência artificial, inferência, treinamento e serviços associados. Para a presente aquisição, pretende-se tirar proveito ao máximo os equipamentos envolvidos, pela importância das atividades que serão desempenhadas.

**6.4.11. Aquisição da garantia pelo período de 60 (sessenta) meses para o hardware e 12 (doze) meses para o sfotware**, por ser o prazo de garantia prestada pela maior parte dos fabricantes, e com as necessidades técnicas na própria utilização dos equipamentos, justifica-se a aquisição de garantia elos perídosos definidos.

**6.4.12.** Sobre a forma de pagamento, como já é prática em aquisições de TIC já demonstrada acima no próprio Catálogo de Soluções de TIC (item 6.4.9 deste ETP), e também por estar alinhada ao previsto no Art, 11 Inciso I da Lei 14.133, também **se justifica o pagamento de todo o período de garantia (aquisição única/pagamento à vista), como item específico**, observando-se as especificações técnicas que serão detalhadas neste Estudo Técnico Preliminar.

**6.4.13. Economicidade na aquisição da garantia por 60 (sessenta) meses para o hardware**: Conforme apresentado na Diretrizes para Contratação de Ativos de TIC, a aquisição da garantia por todo o período de vida útil do hardware é mais vantajosa para administração.

**6.4.14.** Dessa forma, conforme já foi descrito, a utilização do **pagamento à vista** pelo serviço de garantia junto com o equipamento é uma **prática do mercado**. Para a presente aquisição, deverá prever no ato convocatório essa forma de pagamento da garantia. O presente Estudo Técnico Preliminar aborda de forma fundamentada a real necessidade e economicidade da medida. Deverá constar no Termo de Referência medidas específicas e suficientes que resguardem a Administração dos riscos inerentes à operação.

**6.4.15.** Os demais requisitos estão definidos na Especificação Técnicas dos Itens, conforme Anexo I a este Estudo Técnico Preliminar.

**6.4.16.** A definição de prazos distintos para a garantia e suporte do hardware e para o licenciamento, suporte e serviços associados à camada de software decorre da diferença técnica entre o ciclo de vida dos equipamentos físicos e o ciclo de evolução das plataformas lógicas que compõem a solução de GPU como serviço.

**6.4.17.** No caso do hardware, a contratação de garantia e suporte pelo período de 60 (sessenta) meses é tecnicamente justificável porque os equipamentos a serem adquiridos constituem a base física da solução, possuem elevado custo unitário, são importados, apresentam alta complexidade tecnológica e deverão permanecer em operação durante todo o ciclo de vida esperado da

infraestrutura. A aquisição de servidores com GPU, redes, componentes internos, fontes, discos, interfaces, módulos e demais elementos físicos demanda proteção contratual compatível com a expectativa de utilização desses ativos pela Administração.

**6.4.18.** A adoção de garantia inferior ao ciclo de vida esperado do hardware poderia expor a Administração a riscos relevantes de indisponibilidade, aumento de custo de manutenção, dificuldade de obtenção de peças de reposição, necessidade de contratações emergenciais ou abertura de novos processos licitatórios para recomposição de componentes críticos. Esse risco é ampliado pelo fato de a solução envolver equipamentos de alta densidade computacional, com componentes especializados e sujeitos a prazos de importação, disponibilidade internacional e variações de preço.

**6.4.19.** Assim, a garantia de 60 (sessenta) meses para o hardware busca assegurar a continuidade operacional da solução durante o período estimado de aproveitamento dos ativos, preservando a disponibilidade dos serviços institucionais de inteligência artificial, reduzindo riscos de paralisação e conferindo maior previsibilidade à sustentação da infraestrutura física.

**6.4.20.** Também se considera que o hardware, uma vez adquirido, tende a possuir ciclo de vida mais estável e previsível do que a camada de software. Embora possam ocorrer atualizações de firmware, drivers e componentes de gerenciamento, a estrutura física principal da solução deverá permanecer em operação por período plurianual, sendo desejável que a cobertura de garantia e suporte acompanhe esse horizonte de utilização.

**6.4.21.** No caso da camada de software, por sua vez, o prazo de 12 (doze) meses para licenciamento, suporte e serviços associados mostra-se tecnicamente mais adequado ao ritmo de evolução das tecnologias de inteligência artificial, orquestração de workloads, aceleração por GPU, observabilidade, segurança, autosserviço, governança, virtualização, contêineres e integração com modelos e frameworks de IA.

**6.4.22.** Diferentemente do hardware, cuja substituição demanda investimento físico relevante e planejamento de longo prazo, as soluções de software aplicadas a ambientes de IA evoluem em ciclos mais curtos, com surgimento frequente de novas funcionalidades, alterações de modelos comerciais, mudanças em arquiteturas de referência, consolidação ou substituição de ferramentas, evolução de frameworks e alteração das melhores práticas de mercado.

**6.4.23.** A contratação de software por 12 (doze) meses permite à Administração assegurar suporte, atualizações, correções, sustentação e estabilidade operacional em período suficiente para implantação, estabilização, maturação do uso e consolidação das demandas institucionais, sem impor compromisso excessivamente longo com tecnologias que podem sofrer mudanças relevantes em curto espaço de tempo.

**6.4.24.** Esse prazo também permite que, ao final do período inicial de 12 (doze) meses, a Administração possa reavaliar a aderência da camada de software às necessidades efetivamente observadas, considerando métricas reais de utilização, perfis de usuários, tipos de workload, consumo de GPU, maturidade operacional, evolução do mercado, surgimento de alternativas tecnológicas e eventual necessidade de ajuste da arquitetura lógica da solução.

**6.4.25.** Sob o ponto de vista técnico, a diferenciação entre 60 (sessenta) meses para hardware e 12 (doze) meses para software preserva a longevidade dos ativos físicos, ao mesmo tempo em que mantém flexibilidade para evolução da camada lógica da plataforma. Essa estratégia reduz o risco de

aprisionamento tecnológico em softwares, modelos de licenciamento ou ferramentas que, em razão da rápida evolução do mercado de IA, possam deixar de ser as mais adequadas ao interesse da Administração antes do término do ciclo de vida do hardware.

**6.4.26.** Além das razões técnicas, a definição desses prazos também considera a limitação da disponibilidade financeira para a contratação. A solução pretendida envolve simultaneamente equipamentos de alto valor agregado, licenças de software, serviços especializados de implantação, suporte, operação assistida, capacitação, integração e sustentação. A contratação de todos os componentes por 60 (sessenta) meses poderia elevar significativamente o dispêndio inicial e comprometer a viabilidade da implantação da solução mínima funcional.

**6.4.27.** Nesse contexto, a priorização de garantia e suporte de 60 (sessenta) meses para o hardware busca proteger o investimento físico mais duradouro e de substituição mais complexa, enquanto a contratação de software e serviços por 12 (doze) meses busca compatibilizar a necessidade de suporte técnico adequado com a disponibilidade orçamentária existente e com a conveniência de reavaliação tecnológica futura.

**6.4.28.** A adoção de prazo de 12 (doze) meses para software e serviços não compromete a continuidade da solução, pois contempla período suficiente para implantação, entrada em operação, estabilização, capacitação das equipes, coleta de métricas reais de uso e consolidação da governança da plataforma. Ao mesmo tempo, evita a imobilização antecipada de recursos em licenças e serviços de longo prazo em um mercado ainda em rápida evolução.

**6.4.29.** A definição proposta, portanto, busca equilibrar três objetivos: assegurar a proteção do investimento físico pelo ciclo de vida esperado dos equipamentos; garantir suporte adequado à camada lógica no período crítico de implantação e maturação da plataforma; e preservar racionalidade econômica diante da limitação de disponibilidade financeira e da necessidade de contratar solução completa, integrada e operacionalmente viável.

**6.4.30.** Dessa forma, a contratação de garantia e suporte por 60 (sessenta) meses para os equipamentos de hardware e de 12 (doze) meses para softwares, licenças e serviços associados mostra-se tecnicamente adequada, economicamente justificável e compatível com a estratégia de implantação progressiva da solução de GPU como serviço, permitindo à Administração maximizar a proteção dos ativos físicos, manter flexibilidade tecnológica e adequar o escopo da contratação à disponibilidade financeira existente.

**6.4.31. Garantia de execução contratual**

**6.4.31.1.** Considerando as características da presente contratação, será exigida garantia de execução contratual, nos termos dos arts. 96 e 98 da Lei nº 14.133, de 2021, destinada a assegurar o fiel cumprimento das obrigações assumidas pela CONTRATADA e a resguardar a Administração quanto aos prejuízos decorrentes de eventual inadimplemento.

**6.4.31.2.** A garantia será fixada em **10% (dez por cento) do valor inicial do contrato**, percentual considerado adequado diante da complexidade técnica e dos riscos envolvidos na execução da solução, nos termos do art. 98 da Lei nº 14.133, de 2021.

**6.4.31.3.** A adoção do percentual de 10% justifica-se pela natureza integrada e estratégica da contratação. O objeto não consiste na simples aquisição de equipamentos, mas na implementação de solução completa de GPU como serviço, envolvendo servidores de elevado valor agregado, softwares,

licenças, armazenamento, conectividade e serviços especializados de planejamento, implantação, integração, capacitação, operação assistida e suporte, todos interdependentes para a efetiva disponibilização da capacidade pretendida pela Administração.

**6.4.31.4.** A interdependência entre os componentes amplia os impactos decorrentes de eventual inexecução total ou parcial. Ainda que determinados equipamentos ou licenças tenham sido entregues, falhas na implantação, integração ou configuração da solução poderão impedir sua entrada em operação e resultar na existência de ativos de elevado valor sem a correspondente capacidade operacional para a qual foram adquiridos.

**6.4.31.5.** Deve-se considerar, ainda, que a contratação será realizada mediante item único e por seu valor global, concentrando em uma única CONTRATADA a responsabilidade pela integração e pelo funcionamento da solução. Eventual inadimplemento relevante poderá ocasionar atrasos na implantação, necessidade de contratação de terceiros para retomada ou conclusão do projeto, custos adicionais de integração e riscos decorrentes da rápida evolução tecnológica e da disponibilidade internacional dos componentes envolvidos.

**6.4.31.6.** A implantação da infraestrutura corporativa de Inteligência Artificial possui elevada relevância institucional para o Exército Brasileiro, constituindo capacidade tecnológica destinada a suportar iniciativas administrativas, analíticas e operacionais e a ampliar a autonomia institucional sobre dados, modelos e processamento. Dessa forma, a não implementação ou a interrupção do projeto poderá gerar prejuízos que ultrapassam o valor dos bens ou serviços individualmente inadimplidos, retardando a disponibilização de capacidade considerada estratégica para a Administração.

**6.4.31.7.** Nesse contexto, entende-se que a garantia correspondente a 10% do valor inicial do contrato é proporcional à complexidade técnica, à concentração das responsabilidades em uma única contratada, à interdependência dos componentes e aos riscos e potenciais prejuízos associados à inexecução da solução, constituindo mecanismo adicional de proteção do interesse público e do investimento realizado pela Administração.

**6.4.31.8.** As modalidades, condições, prazos de apresentação, vigência, atualização, recomposição e hipóteses de execução da garantia serão detalhadas no Termo de Referência, no edital e no instrumento contratual, observadas as disposições dos arts. 96 a 98 da Lei nº 14.133, de 2021.

**6.5. Requisito Temporal**

**6.5.1.** A solução deve ser entregue e instalada em até 12 (doze) meses, contados a partir da assinatura do contrato. Em relação a entrega dos materiais, o mesmo deverá ser feito em até 180 (cento e oitenta) dias após o recebimento da Ordem de Fornecimento dos Bens, somente após a verificação do material entregue é que iniciará a contagem do período de garantia.

**6.5.2.** Será celebrado com a licitante vencedora termo de contrato com vigência de 60 (sessenta) meses, **na forma dos arts. 89 e 105 da Lei nº 14.133, de 2021**, com a finalidade de melhor controlar e fiscalizar a entrega do objeto e a execução dos itens de serviços, excetuando-se os de garantia. Os itens de garantia possuem período de execução de 60 (sessenta) meses, cujo prazo será contado a partir da entrega efetiva do respectivo bem. Sendo assim, o término da garantia poderá ultrapassar o prazo de vigência do contrato, em conformidade com a Orientação Normativa nº 51, de 25 de abril de 2014, alterada pelo Parecer nº 00025/2023/CNLCA/CGU/AGU, de 19 de dezembro de 2023:

> "Enunciado: A garantia legal ou contratual do objeto tem prazo de vigência próprio e desvinculado daquele fixado no contrato, permitindo eventual aplicação de penalidades em caso de descumprimento de alguma de suas condições,

mesmo depois de expirada a vigência contratual.
Referência Legislativa: Arts. 57, 69 e 73, §2º, da Lei nº 8.666, de 1993; arts. 105, 119, 120 e 140, §2º, da Lei nº 14.133, de 2021; e arts. .24, 25 e 50 da Lei nº 8.078, de 11 de setembro de 1990."

**6.5.3.** A presente contratação destina-se à constituição de capacidade corporativa a ser disponibilizada centralizadamente pelo CITEx, podendo seus benefícios alcançar diferentes organizações militares do Exército por meio do consumo compartilhado da plataforma. com recursos a serem empregados de fontes diversas. Dessa forma, a disponibilidade de recursos e os valores a serem destinados para cada organização militar, conferem ao processo um caráter de imprevisibilidade, acarretando entregas de forma parcelada.

**6.5.4.** A evolução da capacidade da solução poderá ocorrer de forma progressiva, em etapas, conforme disponibilidade orçamentária, amadurecimento da demanda e priorização institucional dos casos de uso e das cargas de trabalho, sendo esse mais um fator de imprevisibilidade, em relação a quais organizações serão beneficiadas pela solução.

**6.6. Requisito de Segurança e Privacidade**

**6.6.1.** Os produtos devem estar de acordo com a política de segurança do EB e pela Diretriz de Segurança da Informação e Comunicações de 2014 (DSIC) da CONTRATANTE.

**6.6.2.** Para todos os itens, o licitante vencedor deverá encaminhar as certificações e relatórios requeridos no Termo de Referência (TR), os quais atendem a demandas de segurança da informação da Força Terrestre.

**6.6.3.** A empresa vencedora deverá seguir os requisitos gerais de estruturação de privacidade e segurança elencados a seguir.

**6.6.3.1.** A empresa contratada deverá possuir uma Política de Segurança da Informação (POSIN), ou equivalente, aderente ao disposto na IN GSI/PR nº 1, de 27 de maio de 2020, incluindo políticas ou normas para proteção de dados pessoais vigentes e atualizadas, com processo de revisão periódico formalizado e institucionalizado, de forma a garantir, dentre outros requisitos, o uso de sistemática e procedimentos de segurança da informação para assegurar não apenas a disponibilidade, a integridade, a confidencialidade e a autenticidade, mas também a consistência, a privacidade e a confiabilidade dos dados e informações tratados pela Solução de TIC;

**6.6.3.2.** A empresa contratada deverá realizar avaliação de Impacto na Privacidade, em conjunto com a contratante, avaliação de impacto na privacidade relacionada à Solução de TIC, considerando o descrito pelo relatório de impacto à proteção de dados pessoais, conforme previsto na Lei nº 13.709 /2018, quando da concepção de qualquer novo projeto, produto ou serviço;

**6.6.3.3.** A empresa contratada deverá realizar Análise e Avaliação de Riscos e apresentar à contratante periodicamente uma análise/avaliação de riscos da arquitetura de Solução de TIC, indicando os eventos de risco ao qual o sistema está exposto, baseada em prévia análise de vulnerabilidades dos ativos que compõem a Solução de TIC, resguardando os segredos de negócio, direitos autorais e direitos de propriedade intelectual aplicáveis, conforme metodologia indicada pela contratante;

**6.6.4.** A empresa vencedora deverá seguir os requisitos Arquitetura, Controles de Privacidade e de Segurança da Informação e Matriz de Responsabilidades elencados a seguir.

**6.6.4.1.** A empresa vencedora deverá apresentar, em tempo determinado pela contratante:

**6.6.4.1.1.** Documentação que descreve a arquitetura física e lógica da Solução de TIC;

**6.6.4.1.2.** Uma descrição dos controles de privacidade e segurança da informação implementados em cada componente descrito na arquitetura física e lógica; e

**6.6.4.1.3.** Matriz de responsabilidades descrevendo a atribuição das responsabilidades pela segurança da informação na organização, pela privacidade (encarregado), identificação dos gestores de serviços com dados pessoais, operador(es) de tratamento de dados, relacionada ao objeto da contratação e com relação aos itens aqui descritos.

**6.6.4.2.** A empresa vencedora deverá possuir e implementar um Plano de Continuidade Operacional e um Plano de Contingência relacionados ao objeto contratado, que garantam o nível requerido de continuidade para a segurança da informação durante uma situação adversa;

**6.6.4.3.** A empresa vencedora deverá possuir e implementar um processo de Gestão de Incidentes que registre os incidentes de privacidade e segurança da informação ocorridos e que contemple: a definição de incidente; o escopo da resposta; quando e por quem as autoridades devem ser contatadas; papéis, responsabilidades e autoridades; avaliação de impacto do incidente; medidas para reduzir a probabilidade e mitigar o impacto do incidente; descrição da natureza dos dados pessoais afetados; as informações sobre os titulares de dados pessoais envolvidos; procedimentos para determinar se um aviso para indivíduos afetados e outras entidades designadas (por exemplo, órgãos reguladores) é necessário; além de implementar e manter controles e procedimentos específicos para detecção, tratamento e resposta a incidentes de segurança da informação e de privacidade, de forma a reduzir o nível de risco ao qual o objeto do contrato e/ou a contratante estão expostos, considerando os critérios de aceitabilidade de riscos definidos pela contratante;

**6.6.4.4.** A empresa vencedora deverá possuir implementar os controles necessários para coleta e preservação de evidências de incidentes de segurança da informação e privacidade;

**6.6.5.** A empresa vencedora terá como ações e responsabilidades os itens elencados a seguir:

**6.6.5.1.** Reportar Incidentes. Reportar de imediato à contratante incidentes que envolvam vazamento de dados, indisponibilidade ou comprometimento da informação relacionados à Solução de TIC. É necessário que a comunicação seja feita de forma tempestiva para que as providências sejam tomadas em tempo hábil de modo a solucionar o incidente ou amenizar seus efeitos

**6.6.5.2.** Termo de Compromisso e Ciência. Implementar e manter controles e procedimentos específicos para assegurar completo e absoluto sigilo quanto a todos os dados e informações de que o preposto ou os demais empregados da contratada venham a tomar conhecimento em razão da execução do contrato, de forma a assegurar que seus empregados e outros profissionais sob sua direção e/ou controle respeitem o uso dos dados somente para as finalidades previstas em contrato e as restrições de uso dos ativos utilizado para desenvolvimento e/ou operação da Solução de TIC, cumprindo e fazendo cumprir o disposto nos Termo de Compromisso e Termo(s) de Ciência firmados respectivamente, pelo representante legal e pelo(s) empregado(s) da contratada

**6.6.5.3.** Revogação de Privilégios. Comunicar à contratante, de imediato, a ocorrência de transferência, remanejamento ou demissão de funcionário, para que seja providenciada a revogação de todos os privilégios de acesso aos sistemas, informações e recursos da contratante, porventura colocados à disposição para realização dos serviços contratados.

**6.6.6.** Demais requisitos serão pormenorizados nas Especificação Técnicas dos Itens, Anexo I a este estudo.

**6.7. Requisitos Sociais, Ambientais e Culturais**

**6.7.1.** Que os bens sejam constituídos, no todo ou em parte, por material reciclado, atóxico, biodegradável, conforme ABNT NBR – 15448-1 e 15448-2.

**6.7.2.** Que sejam observados os requisitos ambientais para a obtenção de certificação do Instituto Nacional de Metrologia, Normalização e Qualidade Industrial – INMETRO como produtos sustentáveis ou de menor impacto ambiental em relação aos seus similares.

**6.7.3.** Que os bens devam ser, preferencialmente, acondicionados em embalagem individual adequada, com o menor volume possível, que utilize materiais recicláveis, de forma a garantir a máxima proteção durante o transporte e o armazenamento.

**6.7.4.** Que os bens não contenham substâncias perigosas em concentração acima da recomendada na diretiva *RoHS* (*Restriction of Hazardous Substances*), tais como mercúrio (Hg), chumbo (Pb), cromo hexavalente (Cr(VI)), cádmio (Cd), bifenilpolibromados (PBBs), éteres difenil-polibromados (PBDEs).

**6.7.5.** Para os itens cuja atividade de fabricação ou industrialização é enquadrada no Anexo I da Instrução Normativa IBAMA n° 13, de 23/08/2021, será admitida a oferta de produto cujo fabricante esteja regularmente registrado no Cadastro Técnico Federal de Atividades Potencialmente Poluidoras e /ou Utilizadoras de Recursos Ambientais, instituído pelo artigo 17, inciso II, da Lei n° 6.938, de 1981, devendo ser atendido pelo fabricante, ou representante nacional do fabricante, ou ainda que possua certificações internacionais equivalentes aos exigidos na presente legislação.

**6.7.6.** Os critérios e práticas de sustentabilidade deverão observar o Guia Nacional de Contratações Sustentáveis da AGU em sua edição vigente na data de elaboração do Termo de Referência, com adaptação às características dos equipamentos, softwares e serviços que compõem a solução.

**6.7.7.** O fabricante dos equipamentos deve, **preferencialmente**, fazer parte do Responsible Business (www.responsiblebusiness. org). A aliança é a maior coalisão Alliance global dedicada a realização de negócios responsáveis em sua cadeia de suprimentos, garantindo as melhores práticas sustentáveis e de condições de trabalho. A escolha de uma empresa participante dessa organização garante ao Exército a aquisição de equipamentos fabricados seguindo as melhores práticas globais de sustentabilidade.

**6.7.8.** A presente contratação deve levar em consideração às orientações contidas no PARECER n. 00001/2021/CNS/CGU/AGU, em relação a escolha de soluções que atendam os Critérios e Práticas de Sustentabilidade nas Contratações Públicas, bem como as orientações e recomendações contidas na conclusão do referido parecer, em especial quanto a eventuais justificativas quando da impossibilidade de adoção de tais critérios e práticas de sustentabilidade e consulta ao Guia Nacional de Contratações Sustentáveis da Advocacia-Geral da União.

**6.7.9.** A presente contratação foi preparada a partir de interações com fabricantes e fornecedores relevantes para soluções de GPU como serviço, incluindo componentes de hardware, software, conectividade, armazenamento e serviços especializados.

**6.7.10.** Alguns fabricantes internacionais não têm uma operação muito extensa no Brasil, e com isso, atendem apenas as legislações obrigatórias, como o registro na Anatel, mas não atendem legislações cujo atendimento é facultativo, como por exemplo a Portaria 304/2023 do INMETRO. Nesse sentido, foi

feita uma análise sobre os principais pontos relacionados à sustentabilidade e foi identificado o Responsible Business (www.responsiblebusiness.org), que já conta com os principais fabricantes globais de equipamentos de TIC.

**6.7.11.** A participação no Responsible Business garante que o fabricante atende a itens como o uso consciente de materiais em seus produtos e embalagens, a subcontratação de cadeia produtiva que também tenha responsabilidade social e atenda aos itens referentes a condições de trabalho, além da existência de política de logística reversa de seus produtos.

**6.7.12.** Com o objetivo de não restringir a competitividade, alternativamente à participação no Responsible Business, o fabricante deverá comprovar o atendimento aos itens a seguir.

**6.7.12.1.** Deve possuir conformidade com o padrão Energy Star, comprovado através de página impressa extraída do site https://www.energystar.gov/productfinder, com equipamento em nome do fabricante do computador ou possuir conformidade com o Anexo D da Portaria 304/2023 do INMETRO.

**6.7.12.2.** O fabricante do equipamento deve ser associado a ABINEE/ GREEN Eletron para gestão para Logística Reversa de Equipamentos Eletroeletrônicos, comprovado através do site https://www.greeneletron.org.br como associado ou através de declaração da ABINEE/ GREEN Eletron ou fabricante deverá possuir programa para logística reversa de produtos eletroeletrônicos, comprovado por meio de seu site oficial e com atuação no Brasil.

**6.7.12.3.** O fabricante do equipamento deve possuir políticas claras e disponíveis em seu site sobre o não uso de mão de obra escrava moderna e tráfico de pessoas em sua cadeia de suprimentos.

**6.7.12.4.** O fabricante do equipamento deve possuir política de acessibilidade a toda sua documentação de produtos, serviços e sites.

**6.7.12.5.** O fabricante do equipamento deve possuir política de compliance ambiental contemplando no mínimo o atendimento ao framework ISO 14001:2015 SEM, o comprometimento a redução do impacto negativo ao ambiente com sua operação e o compliance às normas internacionais ambientais.

**6.8. Sobre o Parcelamento da Solução**

**6.8.1.** Sobre o parcelamento da solução, em atendimento ao Inciso I do § 2º do Art. 12 da Instrução Normativa SGD/ME nº 94, de 23 de dezembro de 2022, a contratação será dividida tantos itens quanto se comprovarem tecnicamente viável e economicamente vantajoso, inclusive considerando as alíquotas de impostos incidentes em bens (equipamentos e *softwares*) e serviços.

**6.8.2.** Também sobre o parcelamento, em se tratando do inciso VIII do § 1º do Art. 18 da Lei nº 14.133, de 2021, a contratação será parcelada em tantos itens quanto foram adequados, do ponto de vista técnico e econômico, de forma que a composição dos itens atendam ao objetivo da contratação, elencados nesse Estudo Técnico Preliminar.

**6.8.3.** Ainda sobre o parcelamento, esta contratação observa o disposto nos §§ 2º e 3º do Art. 40 e § 1º do Art. 47 da Lei nº 14.133, de 2021, principalmente o que se refere ao custo para a Administração de vários contratos frente às vantagens da redução de custos, com divisão do objeto em itens.

**6.8.4.** A presente aquisição trata de uma solução ou sistema único que deve funcionar por meio do conjunto de partes (equipamentos, *software* e serviços) de forma a atender aos objetivos propostos pela administração.

**6.8.5.** Dessa forma, esta Equipe de Planejamento da Contratação considera inviável o parcelamento da Solução.

**6.8.6.** Outro ponto importante relaciona-se com a **padronização** com a presente aquisição, sendo esperado ganho de escala, maior capacidade de monitoramento e operação, padronização técnica das equipes além de atendimento de objetivo específico da aquisição, qual seja, a disponibilização de solução de maneira centralizada para todo Exército.

**6.8.7.** Para alcançar o nível de maturidade desejado, é necessário aumentar o nível de governança por meio de padronização, o que só é possível com o desenho e operação da infraestrutura de forma homogênea e a aquisição centralizada no CITEx. Essa padronização proporcionará um incremento na governança de TIC facilitando a administração da plataforma, o compartilhamento seguro dos recursos computacionais, a sustentação do ambiente e a evolução coordenada da infraestrutura institucional de IA.

**6.8.8**. A coordenação, padronização e a realização de aquisições centralizadas são uma realidade em diversos órgãos de governo na esfera federal, assim como no CITEx, aumentando a maturidade operacional e otimizando a aplicação dos recursos.

**6.8.9.** Embora os componentes e entregáveis sejam discriminados para fins de especificação técnica, formação de preços, medição, pagamento, controle e fiscalização, não haverá parcelamento do objeto para fins de adjudicação. Todos os componentes deverão ser fornecidos por uma única contratada, mediante item único, em razão de sua interdependência técnica, da necessidade de compatibilidade entre hardware, software e serviços e da conveniência de manutenção de responsabilidade única pela integração e pelo funcionamento da solução.

**6.8.10.** Ainda sobre a formação de grupo único, os itens tratam de solução única de forma que tenham compatibilidade entre si. Com o objetivo de garantir que a instalação da solução contratada seja feita de maneira coordenada e assegurando o atendimento às necessidades do SisTEx, se faz necessário a criação de grupo único sem disponibilização de cotas reservadas.

**6.8.11.** Demonstrada a inviabilidade técnica e a ausência de vantagem econômica do parcelamento, a presente aquisição será adjudicada como item único, com julgamento pelo valor global, sem prejuízo da discriminação dos valores dos componentes e entregáveis na planilha de formação de preços. Os entregáveis adquiridos trabalharão em conjunto entre si e com outros sistemas já existentes no SisTEx. O não atendimento a esta limitação possibilita a ocorrência de problemas de compatibilidade entre versões diferentes de soluções de um mesmo fabricante, podendo ocasionar um aumento considerável na complexidade de implantação, obrigando a coordenação entre empresas distintas para garantir a compatibilidade das soluções, também não seria possível a passagem de conhecimento da solução em funcionamento como um todo e implicaria em um tempo maior para assimilação da solução pela equipe do SisTEx, aumentando o esforço e os riscos para operação.

**6.8.12.** O objeto da presente aquisição não será dividida em cotas em função dos riscos apresentados, ressaltados a seguir: poderá ocorrer problemas de compatibilidade entre versões diferentes de soluções de um mesmo fabricante, ocasionando um aumento considerável na complexidade de implantação, obrigando a coordenação entre empresas distintas para garantir a compatibilidade das soluções; e também não seria possível a passagem de conhecimento da solução funcionando como um todo gerando um tempo maior para assimilação da solução pela equipe do SisTEx e aumentando o esforço e os riscos para operação.

**6.8.13.** Outra questão importante é alcançar a padronização a que se refere o Art. 40, Inciso V letra a) da Lei Nº 14.133, de 01 de abril de 2021, a separação em cotas poderia ocasionar a contratação de mais de uma marca e/ou modelo para um determinado equipamento. A separação é prejudicial ao processo, gerando um grande prejuízo à Administração. Portanto, a fixação de cotas reservadas para ME/EPP não será utilizada e excepciona-se na hipótese do art. 10, inciso II do Decreto nº 8.538, de 2015, a saber:

> *"II - o tratamento diferenciado e simplificado para as microempresas e as empresas de pequeno porte não for vantajoso para a administração pública ou representar prejuízo ao conjunto ou ao complexo do objeto a ser contratado, justificadamente;"*

**6.8.14.** Conforme o Inciso VII do Art. 2º da Instrução Normativa SGD/ME nº 94, de 23 de dezembro de 2022, a presente contratação trata de conjunto de bens e/ou serviços que apoiam processos de negócio mediante a conjugação de recursos de Tecnologia da Informação e Comunicações (TIC), de acordo com as premissas definidas no Anexo II da referida Instrução Normativa. Dessa forma, trata-se de uma Solução de TIC.

**6.8.15.** A presente contratação trata de bens e serviços comuns, pois, embora o objeto seja integrado e tecnicamente complexo, seus padrões de desempenho e qualidade podem ser definidos objetivamente por requisitos funcionais e tecnológicos, quantitativos mínimos, matrizes de compatibilidade, documentação oficial dos fabricantes, níveis de serviço, entregáveis previamente descritos e procedimentos de inspeção e testes de aceitação. O julgamento não dependerá de atribuição de pontuação técnica ou de avaliação subjetiva de metodologias concorrentes, mas da verificação objetiva do atendimento integral às especificações e ao menor preço global, conforme o inciso XIII do art. 6º da Lei nº 14.133/2021.

**6.8.16.** Em razão desse enquadramento, a contratação será realizada por pregão eletrônico, conforme o parágrafo único do art. 25 da IN SGD/ME nº 94/2022.

**6.8.17.** A contratação será realizada por meio de pregão, sem utilização do Sistema de Registro de Preços, mediante adjudicação de item único correspondente à solução integrada. A execução orçamentária poderá ocorrer por empenhos parciais ou sucessivos, compatíveis com as parcelas previstas no cronograma físico-financeiro e com os créditos disponíveis em cada exercício, sem que essa forma de empenho altere a natureza da contratação ordinária ou implique formação de ata de registro de preços.

### 6.9. Sobre o Julgamento das propostas

**6.9.1.** A presente aquisição não contará na fase de julgamento das propostas, o previsto no § 3º do Art. 17 da lei 14.133 de 01 de abril de 2021:

> **"§ 3º Desde que previsto no edital, na fase a que se refere o inciso IV do caput deste artigo, o órgão ou entidade licitante poderá, em relação ao licitante provisoriamente vencedor, realizar análise e avaliação da conformidade da proposta, mediante homologação de amostras, exame de conformidade e prova de conceito, entre outros testes de interesse da Administração, de modo a comprovar sua aderência às especificações definidas no termo de referência ou no projeto básico."**

**6.9.2.** Não será exigida apresentação de amostra ou realização de prova de conceito na fase de julgamento das propostas. A conformidade da oferta será verificada mediante documentação técnica oficial, demonstrativo de atendimento e diligências. Após a entrega e a implantação, a solução será submetida aos procedimentos de inspeção, testes de integração, desempenho, segurança, funcionalidade e aceitação definidos nas Especificações Técnicas e no Termo de Referência, como condição para o recebimento e o pagamento dos respectivos entregáveis.

**6.10. Sobre Atestados de Capacidade Técnica**

**6.10.1.** A exigência de apresentação de atestados de capacidade técnica justifica-se como medida indispensável para assegurar que a empresa contratada possua experiência prévia compatível com a dimensão, a complexidade e a criticidade do objeto ora licitado, em estrita observância ao disposto no artigo 67 da Lei nº 14.133, de 2021. Considerando a natureza da contratação e os riscos envolvidos na implantação de solução de GPU como serviço, cada requisito foi delineado com vistas a garantir a seleção de fornecedor com comprovada competência técnica, mitigando riscos de inexecução, atrasos ou falhas que possam comprometer a continuidade dos serviços estratégicos.

**6.10.2.** Disposições gerais sobre a comprovação da capacidade técnica

**6.10.2.1.** Os Atestados de Capacidade Técnica (ACT) deverão ser emitidos por pessoas jurídicas de direito público ou privado e comprovar que a licitante executou satisfatoriamente fornecimentos ou serviços compatíveis com as parcelas de maior relevância técnica ou de valor significativo da presente contratação.

**6.10.2.2.** Os atestados deverão identificar, sempre que aplicável, a pessoa jurídica emitente, a empresa que executou o objeto, a descrição dos bens fornecidos ou dos serviços prestados, os respectivos quantitativos, o período de execução, o local da prestação e a informação de que as obrigações foram cumpridas satisfatoriamente.

**6.10.2.3.** Não será exigido que os atestados se refiram à mesma marca, modelo, fabricante, arquitetura específica ou conjunto exato de produtos previsto nesta contratação. A aferição da capacidade técnica deverá considerar a similaridade de natureza, complexidade e características relevantes do fornecimento ou serviço anteriormente executado.

**6.10.2.4.** Será admitida a apresentação de um ou mais atestados para comprovação de cada requisito, inclusive emitidos por contratantes distintos, desde que o conjunto documental permita verificar objetivamente o atendimento à experiência mínima exigida.

**6.10.2.5.** Para o requisito que envolva quantitativo de equipamentos, será permitido o somatório dos quantitativos comprovados em diferentes atestados. Para o requisito que envolva período de prestação continuada, serão admitidos períodos sucessivos ou não, comprovados por um ou mais atestados, sem dupla contagem de intervalos concomitantes para fins de apuração do prazo mínimo.

**6.10.2.6.** Quando a descrição constante do atestado não for suficiente para comprovar integralmente a natureza, o quantitativo, o prazo ou a complexidade do objeto executado, a Administração poderá realizar diligência e solicitar documentos complementares, tais como contratos, ordens de serviço, notas fiscais, termos de recebimento, relatórios técnicos, declarações do contratante, projetos, desenhos, planos, registros de chamados ou outros documentos relacionados à execução.

**6.10.2.7.** Os documentos complementares não substituirão o atestado, mas poderão ser utilizados para esclarecer ou detalhar informações relativas à experiência nele declarada, observados o julgamento objetivo, a isonomia e a vedação à apresentação posterior de experiência inexistente na data definida para a habilitação.

**6.10.2.8.** Os requisitos estabelecidos nesta seção destinam-se à comprovação da capacidade técnico-operacional da pessoa jurídica licitante e deverão ser avaliados de forma proporcional às parcelas de maior relevância técnica ou de valor significativo do objeto.

**6.10.3. Atestado de fornecimento de servidores**

**6.10.3.1.** A licitante deverá apresentar Atestado de Capacidade Técnica que comprove o fornecimento satisfatório de 03 (três) servidores ou conjunto de servidores que, destinados ao processamento corporativo, científico, analítico ou de inteligência artificial.

**6.10.3.2.** A comprovação poderá ser realizada mediante fornecimento de diversas quantidades de servidores que totalizem esse quantitativo ou pelo somatório de mais de um atestado, desde que os documentos apresentados permitam identificar objetivamente a quantidade total de servidores fornecidos.

**6.10.3.3.** Não se pretende, com essa exigência, restringir a comprovação a determinado fabricante, modelo, geração, arquitetura, capacidade de memória, fator de forma ou tecnologia de interconexão.

**6.10.3.4.** Serão aceitos fornecimentos de servidores destinados a inteligência artificial, computação de alto desempenho, processamento científico, renderização, análise de dados, virtualização ou outras cargas de processamento.

**6.10.3.5.** A exigência desse atestado justifica-se porque o servidor constitui o componente de maior criticidade técnica e valor unitário da solução. Trata-se de equipamento de alta densidade computacional, energética e térmica, cuja composição envolve CPUs, memória, armazenamento, fontes de alimentação, refrigeração, interfaces de rede de alto desempenho, firmware, BIOS, drivers, gerenciamento remoto, garantia e suporte especializado.

**6.10.3.6.** O quantitativo mínimo de 03 (três) servidores corresponde a 50% do míniom de 07 (sete) servidores previstos. Esse quantitativo representa experiência mínima suficiente para demonstrar capacidade com fornecimento de equipamentos, sem exigir experiência anterior idêntica à totalidade do objeto e sem restringir a participação apenas a empresas que já tenham fornecido exatamente a mesma configuração.

**6.10.3.7.** A exigência não terá por finalidade aferir experiência com uma tecnologia específica, mas comprovar que a licitante já executou fornecimento envolvendo servidores, componentes de alto valor e requisitos diferenciados de configuração, logística, garantia e suporte.

**6.10.3.8.** O atestado deverá indicar a quantidade de servidores fornecidas. Caso essas informações não constem expressamente do documento, poderão ser apresentadas notas fiscais, contratos, propostas comerciais aceitas, termos de recebimento, relação de números de série ou outros documentos emitidos ou reconhecidos pelo contratante original que permitam verificar o atendimento ao quantitativo exigido.

**6.10.4. Atestado de planejamento técnico da implementação de solução de TIC**

**6.10.4.1.** A licitante deverá apresentar Atestado de Capacidade Técnica indicando que executou serviço compatível com o planejamento da implementação de solução de Tecnologia da Informação e Comunicações (TIC) para organização pública ou privada.

**6.10.4.2.** Consideram-se compatíveis os serviços relacionados ao gerenciamento de projetos, análise técnica prévia de ambiente, levantamento de requisitos, definição de arquitetura, planejamento de integração, elaboração de estratégia de implantação, identificação de riscos, definição de responsabilidades, elaboração de planos de testes e produção de documentos técnicos destinados a orientar a implementação de solução de TIC.

**6.10.4.3.** O atestado deverá evidenciar que a licitante participou da fase de concepção, estruturação ou planejamento técnico da solução, não sendo suficiente a comprovação de mera revenda de produtos, prestação de suporte sem atividade de planejamento ou execução de instalação baseada exclusivamente em procedimentos previamente definidos por terceiros.

**6.10.4.4.** A exigência tem por objetivo comprovar a capacidade da licitante para planejar a implantação de soluções de TIC ao longo das principais fases do ciclo de implementação, abrangendo a compreensão do ambiente existente, a definição de requisitos, o desenho da arquitetura, a identificação de integrações, o planejamento da instalação, a definição de testes, a transição para produção e a entrega do ambiente funcional.

**6.10.4.5.** A etapa de planejamento é especialmente crítica no contexto da solução de GPU como serviço, pois envolve a integração coordenada de servidores com GPU, servidores de gerenciamento, plataformas de virtualização e contêineres, orquestração, softwares de aceleração, autosserviço, observabilidade, segurança operacional, redes de *frontend* e *backend*, armazenamento, autenticação e serviços institucionais.

**6.10.4.6.** O planejamento inadequado de qualquer dessas camadas pode resultar em incompatibilidades entre versões, falhas de desempenho, insuficiência de recursos, inadequação da infraestrutura elétrica ou de refrigeração, erros de segmentação de rede, ausência de observabilidade, fragilidades de segurança, perda de suporte oficial e necessidade de retrabalho durante a implantação.

**6.10.4.7.** A comprovação de experiência anterior nessa etapa demonstra que a empresa possui maturidade técnica para conduzir projetos complexos com metodologia adequada, organizar dependências, estabelecer pontos de controle, identificar riscos, coordenar equipes e fabricantes e transformar necessidades institucionais em documentos executáveis.

**6.10.4.8.** Considerando o escopo da presente contratação, serão produzidos diversos documentos técnicos durante a implantação, entre eles a Proposta de Escopo de Trabalho, o Planejamento de Alto Nível da Solução – HLD, a Proposta Técnica de Implementação, o Desenho de Baixo Nível da Solução – LLD, a Documentação de Instalação, o Plano de Testes, o Manual de Operações e a documentação final do ambiente.

**6.10.4.9.** A comprovação da experiência de planejamento não ficará limitada à utilização dessas denominações. Também serão aceitos artefatos técnicos equivalentes que evidenciem a atuação da empresa no planejamento, desenho, gestão ou preparação da implementação de solução de TIC, tais como:

a) Documento de Arquitetura Funcional;

b) Documento de Arquitetura de Integração;

c) Documento de Arquitetura de Solução;

d) Escopo de Trabalho ou *Statement of Work* – SOW;

e) Plano de Gerenciamento do Projeto;

f) Termo de Abertura do Projeto ou *Project Charter*;

g) Plano de Implementação;

h) Plano de Instalação e Configuração;

i) Plano de Migração ou de Transição;

j) Plano de Testes e Validação;

k) Plano de Testes de Aceitação ou *User Acceptance Test Plan* – UAT;

l) Plano de Garantia da Qualidade;

m) Plano de Gerenciamento de Riscos;

n) Plano de Comunicação;

o) Plano de Mudanças;

p) Plano de Alta Disponibilidade e Recuperação de Desastres – HA/DR;

q) Plano de Endereçamento, Roteamento, Segmentação ou Regras de Firewall;

r) Plano de Capacitação e Transferência de Conhecimento;

s) *Definition of Done* – DoD;

t) *Product Backlog*, *Sprint Backlog* ou *Roadmap* técnico;

u) documentação de pipelines de integração ou entrega contínua – CI/CD;

v) relatórios de levantamento, diagnóstico ou avaliação de ambiente;

w) relatórios de testes, validação ou aceite técnico;

x) relatórios de acompanhamento, progresso ou desempenho do projeto;

y) registro de riscos, problemas, decisões ou lições aprendidas;

z) relatório de encerramento do projeto; e

aa) procedimentos operacionais padronizados – SOP.

**6.10.4.10.** A relação de documentos apresentada no subitem anterior possui caráter exemplificativo. Poderão ser aceitos outros artefatos, metodologias ou denominações adotadas pelo mercado, desde que permitam comprovar que a empresa atuou efetivamente no planejamento técnico ou na gestão da implementação de solução de TIC.

**6.10.4.11.** Não será necessário que o atestado mencione todos os documentos relacionados, nem que a solução anteriormente planejada possua exatamente os mesmos componentes desta contratação. O objetivo é comprovar experiência no processo de planejamento técnico e na organização estruturada de uma implementação, e não exigir identidade absoluta entre os projetos.

**6.10.4.12.** Serão aceitos atestados relativos ao planejamento de soluções de Data Center, processamento, armazenamento, backup, redes, segurança cibernética, virtualização, computação em nuvem, plataformas de contêineres, observabilidade, colaboração, comunicações ou outras soluções corporativas de TIC que envolvam integração entre componentes e elaboração de documentação técnica.

**6.10.4.13.** Não será exigido que a experiência anterior se refira especificamente a solução de GPU como serviço ou inteligência artificial. Essa flexibilização preserva a competitividade, uma vez que as competências essenciais que se pretende avaliar são planejamento, desenho técnico, análise de ambiente, gestão de dependências, integração, documentação e controle da implementação.

**6.10.4.14.** A comprovação poderá ser realizada por um único atestado, independentemente da quantidade de equipamentos, usuários, licenças ou locais envolvidos. Também será admitida a apresentação de mais de um atestado quando a documentação conjunta permitir demonstrar as atividades de planejamento exigidas.

**6.10.4.15.** O objetivo desse requisito não é aferir o porte meramente quantitativo da solução anteriormente planejada, mas a natureza e a complexidade técnica do serviço realizado. Essa abordagem evita restringir a competição por critérios de escala que não guardem relação direta com a capacidade de produzir os documentos e conduzir as atividades de planejamento previstas nesta contratação.

**6.10.4.16.** Caso o atestado descreva o objeto de forma genérica, a licitante poderá apresentar, para fins de esclarecimento, contratos, termos de referência, escopos de trabalho, termos de aceite, relatórios, projetos ou outros documentos que demonstrem a efetiva elaboração dos artefatos ou execução das atividades de planejamento.

**6.10.4.17.** A exigência do atestado possui natureza estritamente técnica e busca assegurar que a futura contratada tenha experiência prática na fase de concepção e desenho da solução, etapa indispensável para o sucesso de projeto que reúne múltiplas camadas de hardware, software, rede, segurança, observabilidade e serviços especializados.

**6.10.4.18.** Dessa forma, a solicitação do atestado de planejamento da implementação de solução de TIC constitui medida preventiva destinada a assegurar que a execução contratual seja conduzida por empresa que já tenha vivenciado atividades de concepção, estruturação, documentação e gestão de projetos tecnológicos, reduzindo riscos de incompatibilidade, atrasos, retrabalho, aumento de custos e insucesso da implantação.

**6.10.5. Atestado de prestação continuada de suporte de solução de TIC**

**6.10.5.1.** A licitante deverá apresentar Atestado de Capacidade Técnica que comprove a prestação continuada de serviços de suporte técnico à solução de TIC pelo período mínimo de 12 (doze) meses.

**6.10.5.2.** Consideram-se compatíveis os serviços de suporte prestados de forma continuada a ambientes corporativos de hardware, software, redes, armazenamento, segurança, virtualização, computação em nuvem, plataformas de contêineres, observabilidade, processamento, backup, colaboração ou outras soluções de TIC.

**6.10.5.3.** O serviço anteriormente executado deverá envolver atendimento técnico organizado e continuado, podendo compreender, entre outras atividades, registro e tratamento de incidentes, diagnóstico de falhas, solução de problemas, orientação técnica, manutenção corretiva ou preventiva, atualização controlada, acompanhamento de chamados, escalonamento a fabricantes, monitoramento, emissão de relatórios ou cumprimento de níveis de serviço.

**6.10.5.4.** Não será necessário que o serviço anterior contemple simultaneamente todas as atividades relacionadas no subitem precedente. O atestado deverá, contudo, demonstrar a existência de uma estrutura continuada de suporte e não apenas a execução de atendimento pontual ou isolado.

**6.10.5.5.** A mera comprovação de garantia de produto, desacompanhada de evidência de prestação continuada de atendimento técnico, não será suficiente. Poderá ser aceita garantia acompanhada de suporte quando o atestado ou os documentos complementares demonstrarem a existência de canais de atendimento, gestão de chamados, diagnóstico, escalonamento, manutenção ou outras atividades executadas de forma continuada.

**6.10.5.6.** A comprovação do período mínimo de 12 (doze) meses poderá ser realizada por um único atestado ou pelo conjunto de diferentes atestados, relativos a períodos sucessivos ou não, desde que a documentação permita demonstrar experiência acumulada equivalente ao prazo exigido.

**6.10.5.7.** Para fins de contagem, os períodos concomitantes de suporte prestados deverão ser considerados, desde que prestados para diferentes clientes.

**6.10.5.8.** Não será exigido que o período de 12 (doze) meses tenha sido integralmente executado em um único contrato, para um único cliente, em um único local ou com determinada marca ou fabricante.

**6.10.5.9.** A exigência desse atestado justifica-se porque a presente contratação prevê serviços continuados de suporte da contratada e suporte avançado do fabricante, necessários à sustentação da plataforma após sua entrada em produção.

**6.10.5.10.** A solução de GPU como serviço será composta por múltiplas camadas interdependentes, de modo que incidentes poderão envolver simultaneamente servidores, GPUs, drivers, firmware, sistemas operacionais, redes, orquestração, autosserviço, observabilidade e segurança. A resolução dessas ocorrências exigirá triagem, correlação de eventos, coordenação entre equipes, escalonamento a fabricantes e acompanhamento até a solução definitiva.

**6.10.5.11.** A experiência por período mínimo de 12 (doze) meses permite demonstrar que a licitante já manteve estrutura de suporte durante ciclo operacional relevante, abrangendo variações de demanda, mudanças de versão, atualizações, incidentes recorrentes, substituição ou mobilização de profissionais, elaboração de relatórios e manutenção dos canais de atendimento.

**6.10.5.12.** O prazo estabelecido é proporcional ao objeto, pois representa apenas parte do período de suporte previsto para a futura contratação e é suficiente para aferir experiência efetiva na continuidade do serviço, sem exigir da licitante histórico correspondente à totalidade da vigência contratual.

**6.10.5.13.** A exigência busca reduzir os riscos de descontinuidade do atendimento, demora na resolução de incidentes, ausência de processos de escalonamento, perda de histórico técnico, aplicação inadequada de atualizações, indisponibilidade prolongada e transferência indevida à Administração da responsabilidade pela coordenação dos fabricantes.

**6.10.5.14.** O atestado deverá indicar o objeto do suporte e o período de prestação, com as respectivas datas de início e término ou a indicação de que o serviço permanece em execução. Quando se tratar de contrato ainda vigente, somente será considerado o período efetivamente executado até a data definida para apresentação da documentação de habilitação.

**6.10.5.15.** Caso o período ou a natureza continuada do serviço não conste claramente do atestado, poderão ser apresentados contratos, ordens de serviço, termos de recebimento, relatórios mensais, registros de chamados, comprovantes de faturamento ou outros documentos emitidos ou reconhecidos pelo contratante que permitam confirmar essas informações.

**6.10.5.16.** Dessa forma, o atestado de suporte continuado constitui medida necessária para assegurar que a futura contratada possua experiência não apenas na entrega inicial de produtos ou na execução de serviços pontuais, mas também na manutenção de estrutura técnica capaz de sustentar solução corporativa de TIC durante período prolongado.

**6.10.6.** Disposições finais sobre os Atestados de Capacidade Técnica

**6.10.6.1.** As exigências de Atestados de Capacidade Técnica apresentadas nesta seção resultam de análise técnica realizada à luz da complexidade do objeto, das parcelas de maior relevância da contratação e dos riscos associados à implantação e à sustentação de uma solução corporativa.

**6.10.6.2.** Cada tipo de atestado foi definido com base em característica distinta e essencial da futura execução: o fornecimento de servidores, o planejamento técnico da implementação e a prestação continuada de suporte de solução de TIC.

**6.10.6.3.** O atestado de fornecimento busca demonstrar a capacidade da licitante para entregar equipamentos de elevada complexidade e valor, incluindo a coordenação da cadeia de fornecimento, a observância das configurações contratadas e a disponibilização das garantias correspondentes.

**6.10.6.4.** O atestado de planejamento busca demonstrar que a licitante possui maturidade para analisar ambientes, estruturar arquiteturas, identificar dependências e riscos, elaborar documentos técnicos e organizar metodologicamente a implementação de solução de TIC.

**6.10.6.5.** O atestado de suporte continuado busca demonstrar que a licitante já manteve, durante período relevante, estrutura organizada de atendimento, diagnóstico, escalonamento e acompanhamento de solução tecnológica em produção.

**6.10.6.6.** Os três requisitos refletem, portanto, etapas complementares do ciclo de vida do objeto: fornecimento da infraestrutura, concepção e preparação da implantação e sustentação operacional após a entrada em produção.

**6.10.6.7.** As exigências não possuem caráter de direcionamento a fabricante, marca, modelo ou tecnologia específica. Serão admitidos atestados relativos a soluções tecnicamente equivalentes, emitidos por diferentes contratantes e relacionados a produtos de quaisquer fabricantes, desde que evidenciem experiência compatível com a natureza do requisito correspondente.

**6.10.6.8.** A possibilidade de apresentação de múltiplos atestados, do somatório de quantitativos, da comprovação por períodos sucessivos ou não e da aceitação de diferentes soluções de TIC demonstra que o objetivo da Administração não é limitar a participação, mas aferir a capacidade técnica de forma objetiva, proporcional e vinculada aos riscos da contratação.

**6.10.6.9.** A Administração não exigirá que uma única contratação anterior reproduza integralmente o objeto ora licitado. A capacidade técnica poderá ser demonstrada por experiências distintas, desde que, em conjunto, comprovem o fornecimento, o planejamento e o suporte exigidos.

**6.10.6.10.** Os atestados deverão ser analisados quanto ao seu conteúdo material, considerando a experiência efetivamente comprovada, e não apenas a denominação utilizada pelo emitente. Diferenças de terminologia, metodologia, tecnologia ou estrutura contratual não deverão resultar na rejeição do documento quando estiver demonstrada a compatibilidade técnica da atividade executada.

**6.10.6.11.** Quando necessário, a Administração poderá promover diligências para verificar autenticidade, esclarecer informações e confirmar os quantitativos, períodos e atividades declarados, inclusive mediante contato com os emitentes e análise dos documentos relacionados à execução.

**6.10.6.12.** Em conclusão, as exigências estabelecidas encontram-se vinculadas às parcelas críticas da contratação e destinam-se a reduzir os riscos de inexecução, falhas de planejamento, atraso no fornecimento, implantação inadequada e descontinuidade do suporte.

**6.10.6.13.** Trata-se de medida preventiva e proporcional, destinada a selecionar empresa que demonstre experiência prática nos principais elementos da execução, preservando simultaneamente a competitividade, a isonomia, o julgamento objetivo, a eficiência e a obtenção do resultado mais vantajoso para a Administração.

**6.11. Demonstrativo de Comprovação das Especificações**

**6.11.1.** A exigência do Demonstrativo de Comprovação das Especificações revela-se medida vantajosa para a Administração, uma vez que facilita a análise de propostas e assegura que a comprovação do atendimento às especificações técnicas ocorra de forma objetiva, transparente e padronizada. Ao obrigar que a comprovação seja feita majoritariamente por meio de documentos públicos emitidos pelos fabricantes, como manuais e datasheets, garante-se maior confiabilidade e rastreabilidade das informações apresentadas, evitando-se interpretações subjetivas e reduzindo a possibilidade de equívocos ou de apresentação de declarações genéricas sem respaldo técnico.

**6.11.2.** A utilização prioritária de documentos públicos para comprovação das especificações também promove maior igualdade entre os licitantes, uma vez que esses documentos são de acesso livre e não dependem de tratativas comerciais específicas entre fornecedores e fabricantes. O limite de 10% para a utilização de declarações do fabricante reforça esse princípio, restringindo seu uso apenas a situações excepcionais, quando determinada informação técnica não está disponível em documentos de acesso público. Tal medida confere flexibilidade suficiente para que as empresas consigam comprovar plenamente o atendimento às exigências, sem que isso resulte em barreiras indevidas à competitividade do certame.

**6.11.3.** Adicionalmente, a exigência de tabela padronizada contendo a referência exata ao documento comprobatório, bem como o trecho e a página que evidenciam o atendimento, assegura que o processo de análise seja conduzido com maior eficiência e objetividade, reduzindo o tempo de verificação e prevenindo disputas decorrentes de interpretações divergentes. Essa padronização contribui diretamente para a transparência do processo licitatório, já que possibilita tanto aos órgãos de

controle quanto aos licitantes compreenderem de forma clara os critérios de avaliação utilizados pela Administração.

**6.11.4.** No âmbito legal, a exigência do demonstrativo encontra respaldo na Lei nº 14.133/2021. O art. 5º, inciso IV, prevê como princípio da licitação a isonomia, a seleção da proposta mais vantajosa e o julgamento objetivo, todos plenamente atendidos com a adoção do demonstrativo, que elimina subjetividade e proporciona critérios técnicos claros de análise. O art. 7º, inciso I, que trata da padronização dos instrumentos de contratação, também sustenta a medida, pois o demonstrativo atua como ferramenta uniforme e organizada para avaliação das propostas.

**6.11.5.** O art. 42 estabelece que os licitantes devem comprovar o atendimento às especificações do edital por meio de documentação adequada, sendo que o demonstrativo funciona como instrumento que centraliza e organiza essa comprovação, ampliando a clareza e reduzindo a possibilidade de questionamentos. Já o art. 63, inciso I, que dispõe sobre a habilitação técnica por meio de documentos que evidenciem a capacidade do licitante, também encontra correspondência no uso do demonstrativo, que reforça o caráter objetivo e técnico da habilitação.

**6.11.6.** Cabe ainda mencionar o art. 12, §1º, que veda a adoção de exigências que restrinjam a competitividade, fundamento que respalda a previsão de possibilidade do uso excepcional de declarações do fabricante. Essa flexibilização garante que nenhum licitante seja excluído por motivos formais quando determinada informação não esteja disponível em documentos públicos, preservando assim o equilíbrio entre a necessidade de rigor técnico e a ampla participação.

**6.11.7.** Por fim, o art. 11, inciso II, da Lei nº 14.133/2021, impõe à Administração o dever de planejar suas contratações de modo a garantir a seleção da proposta mais vantajosa e assegurar a execução eficiente do objeto, exatamente o que se pretende com a exigência do demonstrativo. Trata-se, portanto, de medida legítima, vantajosa e juridicamente amparada, que fortalece a transparência, a objetividade e a segurança técnica do processo licitatório, ao mesmo tempo em que preserva a competitividade entre os fornecedores.

**6.12. Requisitos para análise da proposta e habilitação técnica**

**6.12.1.** A Equipe de Planejamento da Contratação, ao elaborar o presente Estudo Técnico Preliminar para a aquisição de Solução de Hardware e Software para Processamento de Inteligência Artificial, identificou a necessidade de estabelecer mecanismos complementares de verificação destinados a assegurar a idoneidade da oferta, a qualificação da equipe responsável pela execução, a autenticidade dos equipamentos, licenças, garantias e serviços e a aderência da solução ofertada à arquitetura e aos requisitos definidos pela Administração.

**6.12.1.1.** A contratação possui elevada complexidade tecnológica, pois não se limita ao fornecimento isolado de servidores com GPU. O objeto compreende a entrega de solução corporativa de GPU como serviço, integrada por servidores de processamento acelerado, servidores destinados à hospedagem da camada de software, virtualização, contêineres, orquestração, aceleração por GPU, autosserviço, observabilidade, segurança operacional, integração com a infraestrutura de rede existente, serviços especializados de planejamento e implantação, operação assistida, suporte continuado e capacitação.

**6.12.1.2.** Em razão dessa composição, a análise da proposta não poderá se limitar à comparação de descrições comerciais genéricas. Será necessário verificar a configuração exata dos equipamentos, a abrangência dos licenciamentos, a compatibilidade entre os componentes, a existência de suporte oficial, a participação dos fabricantes, a qualificação dos profissionais e a correspondência entre os preços apresentados e o conjunto de obrigações assumidas.

**6.12.1.3.** Os documentos previstos nesta seção possuem finalidades distintas. Aqueles relacionados à capacidade e à disponibilidade da equipe técnica serão examinados no âmbito da qualificação técnica, conforme as condições estabelecidas no edital. Os documentos relativos à configuração dos produtos, à comprovação das especificações, à composição de preços, ao licenciamento e à participação dos fabricantes serão analisados como elementos da proposta técnica e comercial.

**6.12.1.4.** A seguir, apresentam-se as justificativas técnicas para a exigência de cada documento ou declaração, os quais se mostram necessários à segurança, à qualidade, à rastreabilidade e à efetividade da futura contratação.

**6.12.2.** Declaração de fornecimento de garantia, licenciamento e suporte oficiais.

**6.12.2.1**. A licitante deverá apresentar declaração de que os equipamentos, licenças, garantias e serviços de suporte que compõem sua proposta serão adquiridos diretamente dos respectivos fabricantes ou por intermédio de distribuidores, revendedores ou canais oficialmente autorizados, quando essa for a forma de comercialização adotada pelo fabricante.

**6.12.2.2.** A exigência possui caráter comprobatório e não se destina a estabelecer vínculo comercial prévio como condição genérica de participação. Sua finalidade é assegurar que os produtos e serviços ofertados sejam legítimos, reconhecidos pelos fabricantes e elegíveis para registro, licenciamento, atualização, garantia e suporte oficial durante os períodos previstos na contratação.

**6.12.2.3.** No contexto da presente solução, a legitimidade da cadeia de fornecimento é especialmente relevante porque parte significativa do valor contratado estará associada a equipamentos de alta complexidade e elevado custo, incluindo servidores com múltiplas GPUs, componentes de rede de alto desempenho, licenças empresariais de virtualização, contêineres, orquestração, aceleração por GPU, autosserviço, observabilidade e segurança operacional.

**6.12.2.4**. A solução deverá contar com garantia de hardware pelo período estabelecido nas especificações e com licenciamento e suporte oficial dos softwares durante toda a vigência aplicável. A ausência de registro válido junto ao fabricante poderá impedir o acesso a atualizações, correções de segurança, novas versões, substituição de peças, abertura de chamados e escalonamento de incidentes às equipes de engenharia.

**6.12.2.5.** No caso dos servidores com GPU, o suporte do fabricante é indispensável para o tratamento de incidentes relacionados a firmware, BIOS, componentes de processamento, fontes, refrigeração, memória, interfaces de rede, armazenamento, drivers, bibliotecas de aceleração e comunicação entre GPUs. A aquisição por cadeia não reconhecida poderá resultar na perda de garantia ou na impossibilidade de atendimento em território nacional.

**6.12.2.6.** Em relação aos softwares, deverão ser fornecidas licenças empresariais, com suporte oficial e cobertura para todas as funcionalidades necessárias ao funcionamento da plataforma. Não serão admitidas licenças irregulares, versões comunitárias utilizadas em substituição a produtos empresariais exigidos, licenças destinadas a demonstração, desenvolvimento ou uso não produtivo, salvo quando expressamente previstas para determinada finalidade.

**6.12.2.7.** A declaração também deverá contemplar o compromisso de que os equipamentos, licenças, garantias e contratos de suporte serão associados à conta corporativa indicada pela Administração nos portais dos fabricantes, sempre que essa funcionalidade estiver disponível. Essa associação permitirá ao CITEx verificar períodos de cobertura, números de série, licenças ativadas, direitos de atualização, contratos de suporte e demais informações necessárias à gestão do ciclo de vida da solução.

**6.12.2.8.** A comprovação definitiva deverá ocorrer antes do recebimento dos respectivos componentes, mediante apresentação de contratos, certificados, registros em portais, números de série, identificadores de licenças, part numbers, SKUs, documentos emitidos pelos fabricantes ou outras evidências que permitam confirmar a efetiva disponibilização das coberturas contratadas.

**6.12.2.9.** A exigência reduz os riscos de fornecimento de equipamentos provenientes de mercado paralelo, licenças reutilizadas ou incompatíveis com a finalidade contratada, garantias assumidas exclusivamente pela revenda sem respaldo do fabricante, serviços não registrados e produtos sem direito a atualização ou escalonamento técnico.

**6.12.2.10.** Quando determinada informação puder ser confirmada diretamente em documentação pública ou portal oficial do fabricante, essa forma de comprovação deverá ser privilegiada. A declaração específica do fabricante ou do canal deverá ser exigida apenas quando necessária para confirmar condição que não possa ser verificada por outros meios ou para materializar compromisso futuro diretamente relacionado à proposta.

**6.12.3.** Declaração de disponibilidade de profissionais certificados e qualificados

**6.12.3.1**. A licitante deverá declarar que dispõe ou disporá, até o momento definido no edital para o início da execução contratual, dos profissionais certificados e qualificados necessários à execução dos serviços previstos nas especificações técnicas.

**6.12.3.2.** A declaração de disponibilidade tem por finalidade assegurar que a empresa vencedora possua condições concretas de mobilizar equipe compatível com a complexidade do objeto, sem exigir, de forma antecipada e desproporcional, a manutenção de vínculo empregatício com todos os profissionais durante a fase de apresentação das propostas.

**6.12.3.3**. A solução envolve múltiplos domínios técnicos, incluindo servidores de alta densidade com GPU, infraestrutura de Data Center, virtualização, contêineres, orquestração, aceleração computacional, redes de frontend e backend, RDMA, RoCE, integração com redes Cisco, armazenamento, observabilidade, segurança operacional, gerenciamento de projetos, planejamento de implantação e suporte continuado.

**6.12.3.4.** Em razão dessa diversidade, será necessária a disponibilização de profissionais com qualificações compatíveis com as respectivas atividades, incluindo, quando exigido nas especificações:

a) profissional com certificação técnica do fabricante do hardware dos servidores, em nível compatível com as atividades de maior complexidade;

b) profissional apto a participar da elaboração ou validação do desenho lógico da solução;

c) especialistas nas plataformas de virtualização, contêineres, orquestração, aceleração por GPU, autosserviço e observabilidade ofertadas;

d) profissional qualificado para a integração com a infraestrutura Cisco de Data Center;

e) profissional com experiência em gerenciamento de projetos de TIC; e

f) especialistas aptos a prestar suporte, realizar escalonamentos e interagir com as estruturas técnicas dos fabricantes.

**6.12.3.5.** A participação de profissionais certificados reduz os riscos de configuração incorreta, utilização de versões incompatíveis, falhas de parametrização, perda de desempenho,

comprometimento da segurança, implantação de procedimentos não suportados e realização de alterações que possam afetar a garantia ou o suporte oficial da solução.

**6.12.3.6.** A exigência de certificação não terá por finalidade restringir a execução apenas a empregados diretos do fabricante. Poderão ser utilizados profissionais da contratada, de empresas subcontratadas nos limites autorizados ou disponibilizados diretamente pelos fabricantes, desde que sejam observadas as responsabilidades previstas no contrato e comprovadas as qualificações exigidas.

**6.12.4.** Demonstrativo de Comprovação das Especificações

**6.12.4.1.** A licitante deverá apresentar Demonstrativo de Comprovação das Especificações, contendo a correspondência entre cada requisito das especificações técnicas e a documentação utilizada para demonstrar seu atendimento.

**6.12.4.2.** Não será suficiente a mera reprodução do texto das especificações, o preenchimento genérico com expressões como “atende”, “conforme solicitado” ou “incluso”, nem a apresentação exclusiva de material publicitário sem elementos técnicos verificáveis.

**6.12.4.3**. Quando a comprovação depender de declaração específica do fabricante, essa declaração deverá identificar a solução, os componentes abrangidos e o requisito comprovado, não sendo suficiente declaração genérica de parceria, capacidade comercial ou conhecimento do procedimento licitatório.

**6.12.4.4.** A exigência do demonstrativo é necessária para evitar propostas constituídas por produtos individualmente disponíveis, mas incapazes de operar de forma integrada ou de atender ao conjunto das funcionalidades requeridas. Também reduz o risco de que a Administração somente identifique incompatibilidades após a assinatura do contrato ou a entrega dos equipamentos.

**6.12.4.5.** O documento auxiliará a equipe técnica na análise objetiva das propostas, na realização de diligências, na comparação entre soluções e na formação das evidências que posteriormente subsidiarão os testes de aceitação e o recebimento contratual.

**6.12.4.6.** A exigência não possui caráter de habilitação técnica da empresa, mas de comprovação de conformidade da solução ofertada, devendo ser examinada na fase de julgamento da proposta técnica.

**6.12.5.** Planilha detalhada de formação e composição de preços

**6.12.5.1.** A licitante deverá apresentar planilha detalhada de formação e composição de preços, compatível com o valor total de sua proposta e com todos os componentes necessários à entrega da solução.

**6.12.5.2.** Embora o objeto possa ser adjudicado como solução única e integrada, a discriminação auxiliar dos valores é necessária para permitir à Administração verificar a coerência econômica da proposta, a abrangência do fornecimento, a incidência tributária, a exequibilidade dos serviços e a correspondência entre os valores e os eventos de pagamento previstos.

**6.12.5.3.** Deverão ser indicados quantitativos, valores unitários, valores totais, período de licenciamento ou suporte, unidade de fornecimento, incidência tributária e demais informações necessárias à compreensão da composição econômica da solução.

**6.12.5.4.** A planilha deverá ser coerente com a lista de materiais e serviços apresentada na proposta técnica, inclusive quanto aos part numbers, SKUs, modelos, quantidades, vigências e níveis de suporte. Divergências relevantes poderão ser objeto de diligência, sem possibilidade de alteração da substância da proposta ou de inclusão posterior de componente cujo custo não tenha sido considerado.

**6.12.5.5.** A discriminação dos valores não descaracteriza a natureza integrada e turn-key do objeto, tampouco transfere à Administração a responsabilidade por adquirir ou integrar separadamente os componentes. Sua finalidade é exclusivamente permitir análise econômica, fiscalização, medição, pagamento, controle patrimonial e rastreabilidade.

**6.12.5.6.** A planilha permitirá identificar eventual ausência de licenças, garantias, componentes, serviços ou horas técnicas necessários ao pleno funcionamento da plataforma, reduzindo o risco de propostas artificialmente reduzidas que posteriormente dependam de acréscimos, aditivos ou aquisições complementares.

**6.12.5.7.** A exigência também subsidiará a análise dos percentuais atribuídos aos diferentes eventos de entrega, implantação e sustentação, de forma que os pagamentos guardem proporcionalidade com os componentes efetivamente fornecidos, ativados, implantados ou prestados.

**6.12.5.8.** Nos serviços continuados, a planilha deverá permitir identificar os valores mensais ou periódicos e os respectivos períodos de cobertura. Nos itens de hardware, software, garantia ou licenciamento pagos em evento único, deverá evidenciar os valores correspondentes e a condição necessária ao respectivo recebimento.

**6.12.5.9.** A planilha constitui instrumento de análise da proposta comercial e de exequibilidade, não devendo ser confundida com documento de habilitação técnica.

**6.12.6.** Proposta comercial com part numbers e componentes verificáveis

**6.12.6.1.** A proposta comercial deverá identificar de forma precisa os produtos, equipamentos, licenças, garantias, subscrições, serviços de suporte e demais componentes que integrarão a solução ofertada.

**6.12.6.2.** Sempre que existente, deverá ser informado o part number, SKU, código comercial, código de configuração ou outra identificação oficial utilizada pelo fabricante. Essa exigência tem por objetivo garantir autenticidade, rastreabilidade e correspondência entre o objeto ofertado, a documentação técnica, a composição de preços e o produto que será efetivamente entregue.

**6.12.6.3.** A utilização de identificadores oficiais impede que a proposta se limite a descrições genéricas ou à reprodução das especificações do edital sem informar o produto concreto utilizado para seu atendimento.

**6.12.6.4.** A proposta deverá apresentar uma lista consolidada de materiais, licenças e serviços, equivalente a uma Bill of Materials – BoM, contendo os códigos, descrições, quantidades e vigências correspondentes.

**6.12.6.5.** Quando determinado produto ou serviço não possuir part number público, poderá ser apresentado código de configuração, descrição oficial, proposta do fabricante, documento de licenciamento, carta de confirmação, BoM validada ou outro elemento que permita identificar inequivocamente o componente ofertado.

**6.12.6.6.** Não se recomenda exigir que todos os códigos estejam obrigatoriamente disponíveis em página pública do fabricante, pois determinadas configurações empresariais, serviços personalizados,

pacotes de suporte, códigos OEM e soluções montadas sob encomenda podem ser comercializados por ferramentas internas. Nesses casos, a autenticidade deverá ser confirmada por documentação oficial ou declaração específica do fabricante.

**6.12.6.7.** A lista deverá abranger não somente os equipamentos principais, mas também componentes auxiliares necessários ao funcionamento da solução, tais como cabos, transceptores, trilhos, acessórios, adaptadores, licenças embarcadas, direitos de uso, garantias, subscrições, serviços de suporte e demais elementos cuja ausência possa impedir a implantação ou a operação.

**6.12.6.8.** A análise conjunta da BoM, do Demonstrativo de Comprovação das Especificações e da planilha de preços permitirá verificar se todos os produtos utilizados para demonstrar conformidade foram efetivamente incluídos e precificados.

**6.12.6.9.** Também será possível verificar se os produtos se encontram vigentes, não estão em condição de fim de venda ou fim de vida incompatível com as exigências, possuem suporte oficial e integram configuração homologada ou validada.

**6.12.6.10.** A exigência reduz o risco de substituição posterior por componente inferior, fornecimento de configuração diversa daquela analisada, ausência de licenciamento necessário, utilização de produto sem suporte ou entrega de solução cuja composição não corresponda à documentação apresentada na licitação.

**6.12.6.11.** A proposta com part numbers constitui documento de identificação e conformidade da oferta, devendo ser examinada durante o julgamento da proposta e novamente confrontada com os equipamentos, licenças e serviços apresentados no recebimento contratual.

**6.12.7.** Síntese e proporcionalidade das exigências

**6.12.7.1.** As exigências descritas nesta seção buscam assegurar quatro dimensões essenciais da contratação: autenticidade dos produtos e serviços, qualificação da equipe executora, conformidade técnica da solução e exequibilidade econômica da proposta.

**6.12.7.2.** A declaração relativa à garantia, ao licenciamento e ao suporte oficial busca assegurar que a Administração receba produtos legítimos e registrados, com acesso aos direitos de atualização, manutenção, substituição de peças e escalonamento técnico contratados.

**6.12.7.3.** A declaração de disponibilidade de profissionais certificados busca demonstrar que a empresa possui condições de mobilizar equipe compatível com a diversidade e a complexidade tecnológica da solução, sem impor vínculo trabalhista antecipado ou restringir indevidamente a participação.

**6.12.7.4.** O Demonstrativo de Comprovação das Especificações permite verificar objetivamente a aderência dos produtos e serviços ofertados, enquanto a proposta com part numbers identifica a configuração concreta que será fornecida e assegura sua rastreabilidade durante a execução.

**6.12.7.5**. A planilha de formação e composição de preços permite analisar a abrangência e a exequibilidade da proposta, verificar a inclusão de todos os componentes e subsidiar a fiscalização, a medição, os pagamentos e os controles administrativo, contábil e patrimonial.

**6.12.7.6.** As exigências não possuem caráter de direcionamento a determinada marca, fabricante, revenda ou integrador. A documentação poderá ser emitida pelos fabricantes correspondentes à solução livremente escolhida por cada licitante, desde que atendidos os requisitos funcionais, técnicos, operacionais e de suporte definidos pela Administração.

**6.12.7.7.** Também não se pretende exigir autorizações comerciais genéricas ou vínculos prévios sem relação com o objeto. As declarações dos fabricantes deverão ser utilizadas apenas quando necessárias para confirmar autenticidade, suporte, compatibilidade, participação direta ou compromisso específico que não possa ser demonstrado por documentação pública ou outro meio objetivo.

**6.12.7.8.** Em conjunto, os documentos permitem que a Administração identifique propostas incompletas, configurações incompatíveis, licenças insuficientes, garantias sem respaldo, serviços não precificados e equipes sem qualificação adequada antes da assinatura do contrato.

**6.12.7.9.** As medidas são proporcionais à elevada complexidade, ao valor, à criticidade e ao caráter integrado da solução de GPU como serviço, contribuindo para a redução dos riscos de inexecução, atraso, incompatibilidade, indisponibilidade, perda de suporte, retrabalho e necessidade de contratações adicionais.

**6.12.7.10.** Em conclusão, os requisitos complementares aqui previstos fortalecem o julgamento objetivo, a transparência, a rastreabilidade e a segurança técnica e econômica do procedimento, assegurando melhores condições para que a Administração selecione proposta apta a produzir o resultado pretendido durante todo o ciclo de implantação e sustentação da plataforma.

### 6.13. Infrações e sanções administrativas e procedimentos para retenção ou glosa no pagamento

**6.13.1.** Considerando que a contratação compreende o fornecimento de solução integrada de hardware, software, licenciamento, suporte e serviços especializados, cuja execução será realizada por meio de ordens de serviço, entregáveis e marcos sucessivos e interdependentes, mostra-se necessário estabelecer, no Termo de Referência, mecanismos objetivos de medição do desempenho da contratada, de adequação dos pagamentos e de responsabilização pelos descumprimentos contratuais.

**6.13.2.** Os modelos de Termo de Referência para contratações de TIC elaborados pela Advocacia-Geral da União e aprovados pela Secretaria de Governo Digital preveem a utilização de indicadores mensuráveis para aferição do cumprimento dos prazos, bem como a aplicação de retenções ou glosas proporcionais aos resultados não alcançados, sem prejuízo das sanções administrativas cabíveis. O modelo aplicável à contratação de serviços de TIC adota, entre seus referenciais, o Índice de Atendimento no Prazo — IAP, calculado a partir da proporção de serviços concluídos tempestivamente, com faixas de ajuste no pagamento conforme o desempenho apurado.

**6.13.3.** De forma complementar, o modelo de Termo de Referência para aquisição de bens de TIC prevê o Indicador de Atraso de Entrega — IAE, destinado a mensurar objetivamente o atraso verificado em cada ordem de fornecimento, tomando como referência a diferença entre o tempo efetivamente utilizado e o prazo estabelecido para sua execução. O mesmo modelo associa o resultado do indicador a faixas progressivas de glosa e, para atrasos de maior gravidade, à possibilidade de aplicação de multa, demonstrando a necessidade de que os efeitos financeiros sejam proporcionais à extensão do descumprimento.

**6.13.4.** A retenção, a glosa ou o redimensionamento do pagamento não possuem, por si sós, natureza de sanção administrativa, mas representam adequação do valor devido ao resultado efetivamente entregue pela contratada. A Instrução Normativa SGD/ME nº 94/2022 estabelece que sejam definidos valores e procedimentos para retenção ou glosa quando a contratada não atingir os valores mínimos aceitáveis, não produzir os resultados esperados, deixar de executar as atividades contratadas ou utilizar recursos em qualidade ou quantidade inferior à exigida, sem prejuízo da aplicação das sanções cabíveis.

**6.13.5.** Em razão das características da solução, deverão ser previstos no Termo de Referência, entre outras ocorrências aplicáveis, os seguintes descumprimentos específicos:

**6.13.5.1. Não atender ao indicador de nível de serviço IAP — Índice de Atendimento no Prazo.** O IAP deverá mensurar, em cada período de referência, a proporção de serviços, solicitações ou entregáveis concluídos dentro dos respectivos prazos, em relação ao quantitativo total cujo encerramento estivesse previsto para o período. O não atingimento da meta estabelecida representa entrega de resultado inferior ao contratado e poderá ensejar ajuste proporcional no pagamento, conforme faixas previamente definidas no Termo de Referência. A permanência do indicador abaixo do nível mínimo, sua reincidência ou a produção de prejuízo relevante ao cronograma, à implantação ou à operação da solução poderá caracterizar inexecução parcial do contrato e justificar a instauração do correspondente processo administrativo sancionador.

**6.13.5.2. Não atender ao indicador de nível de serviço IAE — Indicador de Atraso de Entrega de Ordem de Serviço.** O IAE deverá mensurar o atraso específico de cada Ordem de Serviço — OS ou entregável, mediante comparação entre o prazo estabelecido e o tempo efetivamente utilizado até a entrega aceita pela fiscalização. A previsão desse indicador justifica-se porque o atraso em uma única atividade poderá afetar entregáveis subsequentes, impedir a conclusão de etapas dependentes, postergar testes, treinamentos ou implantação e comprometer o cronograma físico-financeiro da contratação. O Termo de Referência deverá estabelecer faixas progressivas de ajuste no pagamento conforme a extensão do atraso, bem como o limite a partir do qual o descumprimento poderá ser tratado como inexecução parcial ou retardamento injustificado da execução.

**6.13.5.3. Atraso na execução dos serviços.** Sem prejuízo da aferição pelos indicadores IAP e IAE, deverá ser prevista ocorrência específica para o atraso injustificado no início, na execução ou na conclusão dos serviços, atividades, etapas, marcos ou entregáveis previstos no contrato, na Ordem de Serviço, no cronograma aprovado ou nos documentos de planejamento da implementação. Essa ocorrência abrangerá especialmente os prazos que não sejam diretamente mensurados pelos indicadores ou os atrasos que ultrapassem os limites de tolerância neles estabelecidos. O retardamento injustificado da execução ou da entrega do objeto constitui infração administrativa, e o atraso injustificado na execução contratual sujeita a contratada à multa de mora definida no edital ou no contrato.

**6.13.6.** Para o atraso na execução dos serviços, o Termo de Referência deverá estabelecer multa moratória calculada por dia de atraso injustificado, preferencialmente incidente sobre o valor da parcela, do serviço ou do entregável inadimplido, até limite temporal e percentual expressamente definidos. A utilização do valor total do contrato como base de cálculo deverá ficar restrita às hipóteses em que o atraso comprometa a solução como um todo, de modo a preservar a proporcionalidade entre a infração e a consequência aplicada. O modelo da AGU orienta que a multa moratória decorrente do retardamento seja calculada sobre a parcela inadimplida e que seja definido o número máximo de dias de sua incidência.

**6.13.7.** Ultrapassado o limite máximo de atraso estabelecido, ou demonstrado que a mora comprometeu a finalidade da contratação, a Administração poderá avaliar a caracterização de inexecução parcial ou total, a conversão da multa moratória em compensatória, a extinção do contrato e a aplicação das demais sanções legalmente cabíveis, observadas a gravidade da ocorrência e as circunstâncias do caso concreto. A conversão deverá ocorrer em substituição à multa moratória correspondente ao mesmo fato, evitando-se a duplicidade de sanções de idêntica natureza e fundamento. O modelo da AGU prevê expressamente a multa compensatória em substituição à moratória para o retardamento da execução.

**6.13.8.** Os indicadores IAP e IAE e a ocorrência geral de atraso na execução possuem funções complementares. O IAP avaliará o desempenho agregado da contratada durante determinado período; o IAE avaliará o atraso individual de cada Ordem de Serviço ou entregável; e a ocorrência de atraso na execução será aplicada aos marcos, atividades ou obrigações não abrangidos diretamente pelos indicadores, ou quando a gravidade ou a duração da mora ultrapassar o efeito meramente financeiro previsto nas faixas de medição.

**6.13.9.** Para evitar duplicidade indevida, o Termo de Referência deverá estabelecer que um mesmo fato não poderá fundamentar duas multas administrativas com a mesma finalidade. A glosa ou o redimensionamento do pagamento decorrente do IAP ou do IAE poderá coexistir com sanção administrativa quando estiverem presentes os pressupostos para responsabilização, pois a glosa corresponde à adequação do pagamento ao resultado entregue, enquanto a sanção possui finalidade punitiva. A eventual aplicação concomitante deverá ser motivada, proporcional e precedida do procedimento administrativo aplicável.

**6.13.10.** Não deverão ser computados como descumprimento do IAP, do IAE ou como atraso imputável à contratada os períodos decorrentes de ação ou omissão da Administração, indisponibilidade de ambiente ou recurso que caiba à Contratante disponibilizar, suspensão formalmente determinada, alteração contratual que afete o cronograma, caso fortuito, força maior ou dependência externa devidamente demonstrada, desde que a contratada comunique tempestivamente a ocorrência, apresente as evidências correspondentes e adote as medidas razoáveis para redução de seus efeitos.

**6.13.11.** O Termo de Referência deverá apresentar, de maneira clara e objetiva, para cada ocorrência: a descrição da conduta; o indicador ou documento utilizado para sua comprovação; a periodicidade de aferição; a fórmula de cálculo; as metas e faixas de tolerância; a base de incidência da glosa ou multa; os percentuais mínimo e máximo; o limite de dias de atraso; as hipóteses de reincidência; as situações excludentes de responsabilidade; e as condições em que o descumprimento poderá evoluir de ajuste no pagamento para infração administrativa.

**6.13.12.** As sanções administrativas deverão observar os arts. 155 a 163 da Lei nº 14.133/2021, incluindo advertência, multa, impedimento de licitar e contratar e declaração de inidoneidade, conforme o enquadramento e a gravidade da conduta. Na definição e aplicação das sanções deverão ser consideradas a natureza e a gravidade da infração, as peculiaridades do caso concreto, as circunstâncias agravantes e atenuantes, os danos causados à Administração e a existência ou o aperfeiçoamento de programa de integridade. A aplicação de qualquer penalidade deverá ser precedida de processo administrativo que assegure o contraditório e a ampla defesa.

**6.13.13.** Dessa forma, mostra-se tecnicamente necessária e juridicamente adequada a inclusão, no Termo de Referência, das ocorrências relativas ao não atendimento do IAP, ao não atendimento do IAE e ao atraso na execução dos serviços. A medida permitirá relacionar objetivamente o desempenho da contratada aos resultados efetivamente entregues, proteger o cronograma de implantação da solução, proporcionar tratamento proporcional aos diferentes níveis de descumprimento e conferir

maior segurança à fiscalização, à medição, ao pagamento e à eventual responsabilização administrativa.

### 6.14. Planilha de Formação de Preços e Demonstração da Exequibilidade

**6.14.1.** Considerando que a presente contratação será estruturada como item único, com julgamento pelo valor global, mas composta por diferentes entregáveis de hardware, software, licenciamento, garantia, suporte e serviços especializados, deverá ser exigida a apresentação de planilha de formação de preços que permita identificar os valores individuais e compreender a composição econômica da solução ofertada.

**6.14.2.** A planilha de formação de preços terá como finalidade principal fornecer informações suficientes para que a Administração avalie a exequibilidade da proposta, a coerência entre os preços dos diferentes entregáveis, a compatibilidade dos valores com as quantidades, os prazos, as condições de licenciamento, os níveis de suporte e as obrigações técnicas previstas no Termo de Referência.

**6.14.3.** A exigência da planilha mostra-se especialmente relevante em razão da heterogeneidade dos componentes que formam a solução e da necessidade de evitar que preços insuficientes em determinados entregáveis sejam compensados artificialmente por valores excessivos atribuídos a outros componentes, comprometendo a execução, o cronograma físico-financeiro, a manutenção das garantias, a continuidade dos licenciamentos ou a adequada prestação dos serviços.

**6.14.4.** Nas licitações por preço global, cada produto ou serviço que compõe a solução deverá ser discriminado separadamente na proposta, de modo a permitir a identificação de seu preço individual na composição do valor global, conforme previsto no § 4º do art. 12 da Instrução Normativa SGD/ME nº 94 /2022. A Lei nº 14.133/2021 também prevê a desclassificação das propostas que apresentem preços inexequíveis ou que não tenham sua exequibilidade demonstrada quando exigido pela Administração.

**6.14.5.** Não será fornecido pela Administração modelo padronizado ou formulário obrigatório para a planilha de formação de preços. Cada licitante poderá apresentá-la no formato que considerar mais adequado à sua estrutura comercial, contábil e operacional, desde que o documento seja claro, organizado, inteligível e contenha as informações necessárias para demonstrar a exequibilidade da proposta e a compatibilidade entre os valores ofertados e as obrigações assumidas.

**6.14.6.** A liberdade quanto ao formato da planilha não afasta a obrigatoriedade de que sejam individualizados, no mínimo, os preços unitários e totais de todos os entregáveis, as respectivas quantidades, unidades de medida, períodos de licenciamento, prazos de garantia ou suporte e demais elementos necessários à conciliação entre a planilha e o valor global da proposta.

## 7. Estimativa da demanda - quantidade de bens e serviços

### 7.1. Considerações iniciais

**7.1.1.** No Exército Brasileiro (EB), a gestão da infraestrutura de Tecnologia da Informação e Comunicações (TIC) encontra-se sob a responsabilidade do SisTEx, que consiste no conjunto formado pelo CITEx – órgão central do sistema – e pelas suas OMDS, os CTA e os CT, situados nas guarnições dos Comandos de Regiões Militares (RM). Basicamente, os CTA/CT prestam apoio de TIC às Organizações Militares (OM) situadas na área de atuação das RM. Adicionalmente, os CTA apoiam os Comandos Militares de Área de sua região.

**7.1.2.** Vale salientar que o EB possui, atualmente, cerca de 680 (seiscentas e oitenta) OM distribuídas geograficamente por todo o território brasileiro e cada uma possui sistemas de informação que dão suporte à sua vida administrativa e que devem ser hospedados nos CTA/CT devido à disponibilidade de ambiente adequado e seguro, em sintonia com a importância desses sistemas. Além disso, os serviços de TI voltados para o suporte às operações militares permitem a execução rápida e eficaz do ciclo do comando e controle nas operações militares, tais como o Sistema de Rádio Digital Troncalizado (SRDT), o Sistema Pacificador, o Sistema Comando e Controle (C2) em Combate e outros serviços voltados para a logística militar. Com relação à quantidade de usuários, o EB é composto por cerca de 220.000 (duzentos e vinte mil) profissionais na ativa, os quais acessam diuturnamente esses sistemas.

**7.1.3.** Na condição de órgão central do SisTEx, o CITEx é responsável pela gestão orçamentária e pelo direcionamento técnico do sistema. Com o aumento da complexidade das redes e sistemas de TI, está sendo necessária uma coordenação técnica cada vez maior, com o objetivo de padronizar a infraestrutura e as soluções de TIC que atendem ao Exército. Essa padronização tem como principal objetivo a otimização da aplicação de recursos financeiros e de pessoal. Soluções que já se mostraram exitosas em algum dos CTA/CT são grandes candidatas a se tornarem padrão para os demais.

**7.1.4.** Nesse contexto, a evolução da infraestrutura de TIC não pode ser tratada apenas como expansão pontual de capacidade computacional, mas como parte de um processo contínuo de modernização e fortalecimento da base tecnológica do Exército. À medida que cresce a complexidade dos sistemas, das integrações e das necessidades de processamento, torna-se cada vez mais necessária a condução centralizada de projetos que permitam racionalizar investimentos, ampliar a governança, melhorar a operação e evitar fragmentação tecnológica entre as diversas estruturas do sistema.

**7.1.5.** Nos últimos anos, a evolução das soluções de inteligência artificial passou a representar uma nova e relevante demanda para a infraestrutura institucional de TIC. Diferentemente de cargas tradicionais de virtualização, banco de dados ou aplicações corporativas convencionais, os workloads de inteligência artificial, especialmente aqueles relacionados a inferência acelerada, treinamento, ajuste fino de modelos, processamento documental avançado, análise semântica e ambientes interativos de desenvolvimento, demandam recursos computacionais especializados, com elevado poder de processamento paralelo, grande largura de banda e mecanismos específicos de orquestração, compartilhamento e governança.

**7.1.6.** Nesse cenário, a simples aquisição de servidores com GPU, de forma isolada e sem coordenação arquitetural, não se mostra suficiente para atender às necessidades da Administração. O que se busca é a constituição de uma plataforma corporativa de GPU como serviço, capaz de transformar a capacidade computacional acelerada em serviço institucional efetivamente consumível, com mecanismos de autosserviço, controle de acesso, observabilidade, auditoria, cotas, priorização, compartilhamento seguro e possibilidade de expansão progressiva.

**7.1.7.** A motivação para essa aquisição também está relacionada à necessidade de preservar maior soberania institucional sobre dados, modelos, fluxos de processamento e bases documentais utilizadas em soluções de inteligência artificial. Para o Exército, a construção de capacidade própria de processamento acelerado em ambiente sob controle institucional contribui para maior segurança, melhor aderência às políticas de acesso e proteção da informação, maior governança sobre a infraestrutura e menor dependência de arranjos dispersos ou de soluções externas inadequadas à natureza das informações tratadas.

**7.1.8.** Outro ponto relevante é que o uso de inteligência artificial no âmbito do Exército tende a se expandir de forma progressiva e heterogênea. Em um primeiro momento, poderão coexistir demandas por ambientes interativos de desenvolvimento, inferência de modelos, provas de conceito, automação documental, recuperação semântica de conhecimento e experimentação técnica. Em etapa posterior, poderão surgir demandas mais intensivas em treinamento distribuído, uso concorrente por múltiplas áreas, modelos institucionais especializados e maior escala de usuários. Tal variabilidade reforça a necessidade de que a solução seja concebida com arquitetura escalável, implantação progressiva e governança centralizada.

**7.1.9.** A contratação pretendida também deve ser compreendida no contexto mais amplo de amadurecimento do ambiente computacional do SisTEx. Assim como ocorreu em outras frentes de modernização, a adoção de capacidade institucional de GPU como serviço exige visão sistêmica, padronização técnica, coordenação central e integração com a infraestrutura já existente. Essa abordagem permite reduzir assimetrias, evitar ilhas tecnológicas, simplificar capacitação das equipes, melhorar as condições de operação e sustentar futuras expansões com maior racionalidade técnica e econômica.

**7.1.10.** Por todos esses fatores, o CITEx, como órgão central do SisTEx, entende que a aquisição de solução de GPU como serviço deve ser conduzida de forma centralizada, planejada e progressiva, permitindo ao Exército iniciar a utilização institucional de inteligência artificial em ambiente próprio, com escala piloto, mas já com características técnicas, funcionais e operacionais compatíveis com a expansão futura da capacidade computacional.

**7.1.11.** A presente contratação, portanto, não se destina apenas a suprir uma carência pontual de hardware, mas a estruturar uma base tecnológica corporativa apta a sustentar, no presente e no futuro, a evolução do uso de inteligência artificial no Exército Brasileiro.

### 7.2. Situação atual da capacidade institucional para inteligência artificial

**7.2.1.** O ambiente de TIC do SisTEx já dispõe de infraestrutura relevante de datacenter, virtualização, armazenamento, backup e conectividade, voltada ao suporte dos sistemas corporativos do Exército. Todavia, no que se refere especificamente à oferta institucional de capacidade computacional acelerada para inteligência artificial, ainda não há uma plataforma corporativa estruturada de GPU como serviço, com características de compartilhamento seguro, autosserviço, observabilidade, governança e expansão modular.

**7.2.2.** A situação atual é marcada pela existência de capacidade limitada para atendimento de demandas de IA, sem que isso represente, ainda, uma solução institucional plenamente estruturada para suportar de forma padronizada o crescimento das necessidades de inferência, treinamento, ambientes interativos, processamento documental e demais workloads acelerados. Em outras palavras, já existe percepção concreta da demanda e já existem iniciativas relacionadas ao tema, mas ainda não se dispõe de uma base corporativa suficientemente robusta para ofertar esse serviço em escala institucional.

**7.2.3.** Atualmente, o ambiente conta com **12 (doze) GPUs L40S**, disponibilizadas para atendimento de projeto ainda em fase de implementação e não concluído. Embora essa capacidade inicial represente avanço importante no amadurecimento institucional do tema, ela ainda não permite extrair, com segurança e profundidade suficientes, todos os indicadores necessários à especificação precisa de uma demanda corporativa de inteligência artificial.

**7.2.4.** Ainda não se encontram plenamente disponíveis, com base em operação madura e contínua, métricas como quantidade efetiva de usuários concorrentes, consumo por perfil de uso, tokens por segundo requeridos, tamanho e porte dos modelos a serem executados, volume e perfil das bases de dados, frequência de processamento, simultaneidade entre workloads, necessidade de treinamento distribuído, comportamento de inferência em escala e impacto real sobre armazenamento, rede e observabilidade.

**7.2.5.** Esse cenário de incerteza é reforçado pelo fato de que outras demandas de inteligência artificial já identificadas pelo Exército ainda se encontram em fases mais iniciais de desenvolvimento, prototipação ou estruturação funcional. Assim, embora já exista clareza institucional quanto à direção estratégica de ampliação do uso de IA, ainda não se dispõe, em todos os casos, de soluções prontas e em produção que permitam estimativas exatas e definitivas de consumo computacional.

**7.2.6.** Não se mostra adequado, contudo, aguardar que todas essas soluções de IA estejam finalizadas e em operação para somente então iniciar o processo de aquisição da infraestrutura necessária ao seu suporte. Caso essa opção fosse adotada, o resultado prático seria o adiamento, em pelo menos cerca de um ano, da disponibilização de capacidade computacional própria apta a suportar a entrada em produção dessas iniciativas, comprometendo o ritmo de evolução tecnológica pretendido para o Exército.

**7.2.7.** Em razão disso, a estratégia institucional mais adequada consiste em iniciar a implantação da capacidade de GPU como serviço de forma progressiva, a partir de ambiente inicial funcional, que permita simultaneamente gerar valor prático imediato e produzir os dados necessários ao refinamento das fases subsequentes. Tal abordagem é compatível com a necessidade de reduzir riscos de planejamento, evitar superdimensionamentos prematuros e, ao mesmo tempo, impedir que a ausência de infraestrutura adequada se torne fator de atraso para a operacionalização das soluções de inteligência artificial em desenvolvimento.

**7.2.8.** Também deve ser considerado que o consumo de processamento de GPU em nuvem pública não se mostra, em regra, adequado ao nível de soberania exigido para os dados e aplicações do Exército, sobretudo em áreas relacionadas às atribuições constitucionais da Força, ao tratamento de informações sensíveis, ao apoio à decisão, à gestão documental institucional e a outros contextos nos quais a preservação do controle sobre dados, modelos e fluxos de processamento é requisito essencial.

**7.2.9.** Em determinadas iniciativas, a utilização de soluções em nuvem pública poderá, inclusive, inviabilizar ou restringir tecnicamente o emprego da inteligência artificial pretendida, seja pela incidência de guardrails próprios das plataformas públicas, seja pelo risco de exposição indevida de informações sensíveis, seja ainda pela possibilidade de utilização inadequada de dados institucionais em contextos incompatíveis com as necessidades da Administração, inclusive em cenários de treinamento, ajuste ou retroalimentação de modelos públicos de IA.

**7.2.10.** Dessa forma, a situação atual evidencia dois aspectos simultâneos: de um lado, já existe demanda concreta e capacidade inicial limitada, insuficiente para atendimento corporativo; de outro, ainda não estão plenamente consolidados os indicadores que permitiriam especificação exata e definitiva da demanda futura.

**7.2.11.** É justamente dessa combinação entre necessidade já identificada e incerteza ainda existente que decorre a opção por implantação progressiva, em ambiente próprio e sob controle institucional, como forma de permitir aprendizado, medição, governança e expansão ordenada da capacidade computacional acelerada do Exército.

**7.3. Evolução pretendida para o ambiente institucional de processamento acelerado**

**7.3.1.** A evolução pretendida consiste na implantação de uma plataforma corporativa de GPU como serviço, apta a disponibilizar recursos de processamento acelerado de forma centralizada, compartilhada, escalável, segura e sob demanda para os diversos usuários, projetos e áreas do Exército.

**7.3.2.** Essa evolução deverá ocorrer de maneira progressiva, começando por uma etapa inicial com quantitativo reduzido de GPUs, porém já em ambiente plenamente funcional, e prosseguindo por fases de crescimento modular, conforme a maturidade do uso institucional, a consolidação das demandas efetivas e a disponibilidade de recursos. A estratégia visa compatibilizar prudência administrativa, restrição orçamentária e construção de uma base tecnológica apta a crescer de forma ordenada, sem necessidade de redesign estrutural da solução.

**7.3.3.** O objetivo técnico e institucional é que o Exército passe a dispor de uma plataforma na qual a capacidade de GPU possa ser consumida como serviço, com múltiplas formas de acesso, mecanismos de controle e compartilhamento, e suporte a diferentes perfis de workload. Com isso, pretende-se ampliar a governança sobre o uso institucional de IA, reduzir fragmentação de iniciativas, criar base comum para evolução tecnológica e permitir que futuras expansões ocorram com maior precisão, à luz do comportamento real da demanda.

**7.3.4.** A implantação progressiva também permitirá que a Administração valide, em ambiente real, parâmetros que ainda não se encontram plenamente consolidados, como perfil predominante de workload, simultaneidade de uso, pressão sobre armazenamento e rede, comportamento de inferência, necessidade de segmentação por perfis de uso e quantitativos efetivamente necessários para as etapas subsequentes. A solução inicial, portanto, deverá produzir valor operacional e, simultaneamente, funcionar como base de aferição técnica para o crescimento futuro da plataforma.

**7.3.5.** Em síntese, a aquisição ora estudada representa etapa estruturante da evolução da infraestrutura de TIC do Exército, voltada à constituição de capacidade institucional de inteligência artificial em ambiente próprio. Seu propósito é viabilizar, em base segura e controlada, o início e a expansão do uso corporativo de processamento acelerado, fortalecendo a soberania tecnológica, a modernização da infraestrutura e a capacidade do SisTEx de sustentar novos serviços digitais e novas formas de apoio computacional à missão do Exército Brasileiro.

**7.4. Considerações Sobre as Demandas**

**7.4.1**. As demandas de alto nível apresentadas no item 4 deste estudo técnico estão de acordo com as Especificação Técnicas dos Itens, Anexo I ao presente Estudo Técnico Preliminar, sendo atendidas na plenitude as necessidades da contratação.

**7.4.2.** Por se tratar de contratação cujos demais componentes possuem características usuais de mercado e podem ser fornecidos por diferentes fabricantes, não será utilizada a excepcionalidade de indicação de marca ou modelo, ressalvadas exclusivamente as unidades de processamento gráfico — GPUs, que deverão ser da marca NVIDIA, modelo B300, e a Solução de Software Específica para Aceleração por GPU, correspondente ao Entregável 6, que deverá contemplar o produto NVIDIA AI Enterprise. As indicações encontram fundamento no art. 41, inciso I, alíneas “b” e “c”, da Lei nº 14.133 /2021 e decorrem das necessidades de compatibilidade com o ecossistema tecnológico já adotado, integração entre hardware e software, suporte oficial, capacidade de memória e processamento,

arquitetura multi-GPU e redução dos riscos de implantação, conforme matriz comparativa e evidências de mercado juntadas aos autos, que deverão demonstrar também a existência de pluralidade de fornecedores aptos a comercializar e integrar os produtos indicados.

**7.5. Desenho da Solução**

**7.5.1.** A presente contratação tem por objetivo estruturar, no âmbito do Exército Brasileiro, uma solução de **GPU como serviço**, entendida como modelo no qual a capacidade de processamento acelerado deixa de ser tratada como recurso isolado de hardware e passa a ser disponibilizada como serviço corporativo, acessível de forma padronizada, controlada, compartilhada, escalável e sob demanda. Nesse modelo, o usuário não depende de aquisições descentralizadas ou de arranjos locais de infraestrutura para utilizar GPU, passando a consumir recursos computacionais por meio de plataforma institucional com mecanismos de orquestração, autosserviço, controle de acesso, observabilidade e governança.

**7.5.2.** A adoção desse modelo se justifica não apenas pela necessidade de prover capacidade computacional acelerada, mas por sua aderência às práticas atualmente consolidadas no mercado para atendimento de demandas de inteligência artificial em ambiente corporativo. Em vez de tratar cada iniciativa como projeto isolado de infraestrutura, o modelo de GPU como serviço permite organizar o consumo da capacidade acelerada em base comum, compartilhável entre diferentes áreas, projetos e perfis de usuário, com maior racionalidade operacional, melhor uso dos recursos, maior segurança e melhores condições de expansão futura.

**7.5.3.** O desenho da solução parte de uma visão de longo prazo segundo a qual o ambiente institucional deverá estar apto a evoluir para capacidade da ordem de **até 500 GPUs**, sem perda de coerência arquitetural, sem ruptura do modelo operacional e sem necessidade de redefinição estrutural da plataforma a cada expansão. Essa perspectiva de longo prazo é indispensável para compatibilizar o crescimento esperado das demandas de inteligência artificial com a necessidade de preservar continuidade técnica, padronização, governança e viabilidade de sustentação do ambiente.

**7.5.4.** No entanto, a estratégia de implantação adotada não busca materializar de imediato toda a capacidade final vislumbrada. A primeira fase do projeto deverá ser concebida com referência de expansão para **até 80 GPUs**, correspondente a um ambiente composto por **10 servidores de processamento com GPU**, cada qual com **8 GPUs**, sem prejuízo de que a implantação inicial se inicie de forma muito mais contida. Tal diretriz decorre da necessidade de compatibilizar o planejamento estratégico de crescimento com a realidade orçamentária, com o amadurecimento progressivo do uso institucional e com a necessidade de preservar capacidade de ajuste entre etapas.

**7.5.5.** Dentro dessa primeira fase, a contratação deverá iniciar por uma **primeira parte com apenas 8 GPUs**, mas já em ambiente plenamente funcional como plataforma de GPU como serviço. A escolha por começar com 8 GPUs não decorre de redução do escopo conceitual da solução, mas de estratégia deliberada de implantação progressiva. Busca-se, com isso, viabilizar entrada em operação em menor escala, produzir valor prático imediato, reduzir riscos de adoção, permitir aprendizado institucional, coletar indicadores reais de uso e, simultaneamente, manter a arquitetura, a camada de software, a governança e os mecanismos de operação já compatíveis com a expansão para 80 GPUs e, posteriormente, para capacidade ainda superior.

**7.5.6.** A vantagem desse arranjo é permitir que a Administração não fique dependente da maturidade completa de todas as futuras soluções de inteligência artificial para iniciar a construção da infraestrutura que as suportará. Ao começar com ambiente piloto, porém funcional, o Exército poderá validar padrões de uso, perfil de workloads, consumo por usuário, pressão sobre armazenamento e

rede, comportamento de inferência e necessidades de observabilidade, sem atrasar em cerca de um ano a entrada em produção das iniciativas de IA que já se encontram em desenvolvimento. Ao mesmo tempo, a solução já nascerá com base tecnológica capaz de sustentar crescimento controlado até a referência de 80 GPUs sem necessidade de reformulação estrutural.

**7.5.7.** A legitimidade institucional da contratação encontra respaldo na **4ª Reunião do Comitê de Governança Digital (CGD)**, na qual foi registrada a necessidade de ampliação da capacidade computacional voltada à inteligência artificial para atendimento à crescente demanda institucional por soluções baseadas em IA, incluindo aplicações administrativas, analíticas e operacionais, bem como para fortalecimento da soberania sobre os dados e da capacidade de processamento em ambiente próprio. Na mesma oportunidade, foi destacado o alinhamento do projeto com a modernização da infraestrutura de TI do Exército e com as diretrizes estratégicas de evolução tecnológica e suporte a novas capacidades digitais. A ata registra, ainda, que o projeto foi submetido ao Comitê, tendo sido deliberada a continuidade de sua preparação, o que reforça a legitimidade, aderência estratégica e prioridade administrativa da iniciativa.

**7.5.8.** Do ponto de vista funcional, a solução deverá ser capaz de prover, em alto nível, capacidade de GPU sob demanda; compartilhamento seguro da infraestrutura entre múltiplos usuários e projetos; definição de cotas e políticas de utilização; particionamento de aceleradores quando aplicável; suporte a cargas com uso simultâneo de múltiplas GPUs, inclusive de forma distribuída; fila e priorização de workloads; suporte a diferentes modelos de execução de IA; catálogo de imagens padronizadas; integração com diretório corporativo; registro de logs, métricas operacionais e auditoria; medição de consumo; portal web e APIs para autosserviço; expansão horizontal da infraestrutura; e, adicionalmente, capacidades voltadas à utilização institucional de modelos de linguagem, interfaces simplificadas de consumo, bases de conhecimento e funcionalidades documentais, conforme a evolução da plataforma.

**7.5.9. Detalhamento da Solução**

**7.5.9.1. Visão geral do detalhamento da solução**

**7.5.9.1.1.** Após a evolução do entendimento técnico e funcional da demanda, decorrente da análise interna conduzida pelo CITEx, das interações com fabricantes e fornecedores do mercado, bem como das articulações realizadas com outras áreas do Exército interessadas no tema, chegou-se a uma formulação estruturada da solução em camadas. Tal abordagem busca representar, de maneira mais clara e aderente à realidade de implantação e operação, os diferentes níveis que compõem uma plataforma corporativa de GPU como serviço.

**7.5.9.1.2.** A organização em camadas permite compreender a solução não apenas sob a ótica da infraestrutura física, mas também como um conjunto integrado de mecanismos de acesso, componentes lógicos, recursos computacionais, capacidades complementares externas e estruturas permanentes de operação. Essa forma de representação favorece a visualização do papel de cada grupo de componentes, melhora a compreensão das relações entre os elementos da arquitetura e contribui para a definição mais consistente dos itens de aquisição e dos requisitos de implantação, sustentação e expansão.

**7.5.9.1.3.** A Figura 1 apresenta, em alto nível, a visão arquitetural da solução de GPU como serviço, evidenciando sua organização em camadas e a forma pela qual essas camadas se articulam para transformar a capacidade de processamento acelerado em um serviço institucional consumível, escalável e governado. Em vez de representar apenas uma infraestrutura composta por servidores com GPU, a figura demonstra que a solução deve ser compreendida como uma plataforma integrada,

na qual diferentes componentes físicos, lógicos e operacionais atuam de forma coordenada para atender às demandas de processamento em GPU do Exército.

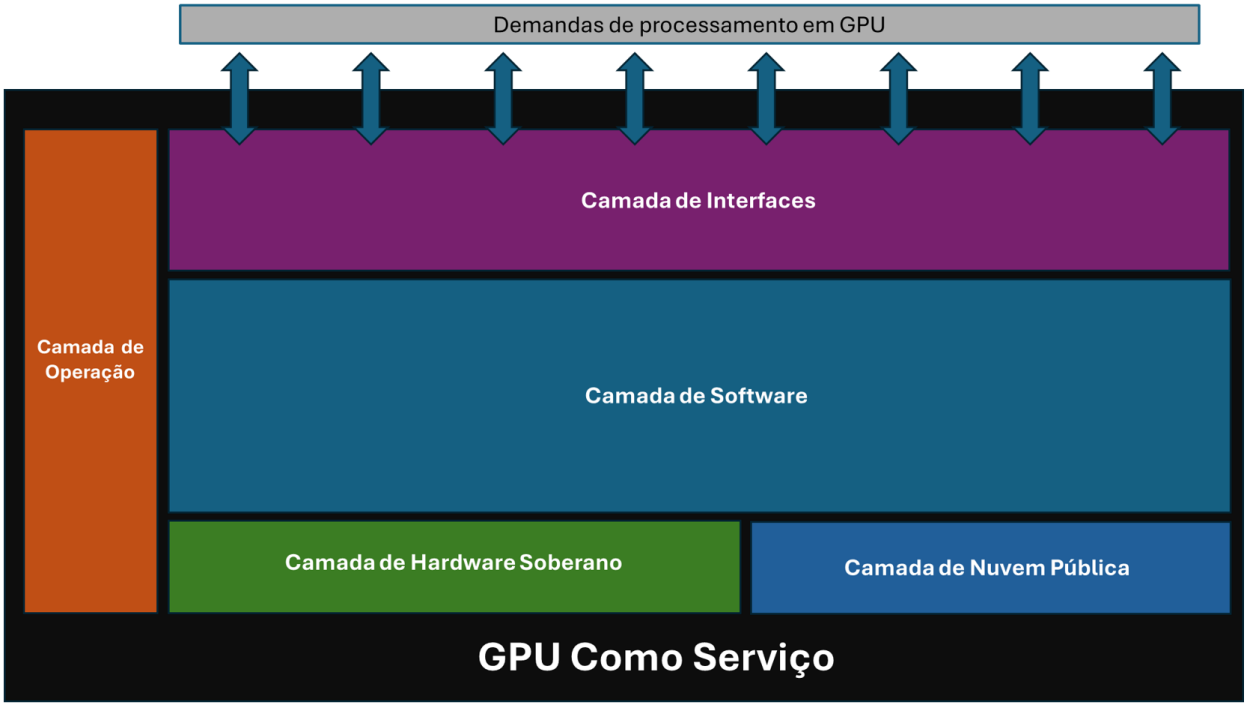



**Figura 1:** Visão arquitetural da solução de GPU como Serviço

**7.5.9.1.4.** Na parte superior da figura, as “Demandas de processamento em GPU” representam o conjunto de necessidades institucionais que motivam a solução, incluindo workloads de treinamento, inferência, desenvolvimento, experimentação e demais usos compatíveis com o ambiente. As setas indicam que essas demandas são absorvidas pela plataforma por meio da Camada de Interfaces, que constitui o ponto de entrada e de consumo do serviço. É por essa camada que usuários, aplicações e equipes técnicas acessam a capacidade de processamento em GPU, utilizando diferentes formas de interação, conforme o perfil de uso e o grau de abstração desejado.

**7.5.9.1.5.** Logo abaixo, a Camada de Software representa o núcleo lógico da solução. É nessa camada que se concentram os componentes responsáveis por viabilizar a operação da plataforma, organizar o uso dos recursos acelerados, sustentar a orquestração dos workloads e assegurar que a infraestrutura computacional seja efetivamente entregue como um serviço. Em termos práticos, essa camada faz a ligação entre as interfaces de consumo e os recursos físicos ou complementares que compõem a base da arquitetura, sendo essencial para transformar hardware acelerado em uma plataforma de GPU como serviço propriamente dita.

**7.5.9.1.6.** Na base da figura, a solução se apoia em dois blocos principais de capacidade computacional. O primeiro deles é a Camada de Hardware Soberano, que representa a infraestrutura física instalada em ambiente próprio, sob controle institucional, e que constitui a base principal da capacidade de processamento, rede e armazenamento da solução. O segundo é a Camada de Nuvem Pública, representada como capacidade complementar, passível de uso quando necessário para expansão, elasticidade ou apoio a cenários específicos. A disposição desses dois blocos na figura evidencia que a arquitetura pode combinar recursos próprios e externos, preservando, contudo, a centralidade da base soberana como elemento estruturante da solução.

**7.5.9.1.7.** Ao lado de toda a arquitetura, a Camada de Operação aparece como elemento transversal, demonstrando que segurança, observabilidade, pessoas e serviços especializados não devem ser entendidos como componentes isolados, mas como capacidades permanentes de sustentação da

plataforma. Essa camada é responsável por garantir que a solução opere de forma segura, monitorada, administrável e aderente às necessidades institucionais ao longo do tempo. Sua posição na figura reforça que a operação não está restrita a um único ponto da arquitetura, mas permeia e sustenta todas as demais camadas.

**7.5.9.1.8.** Assim, a solução de GPU Como Serviço tem por objetivo não apenas disponibilizar GPUs, mas estruturar um ambiente institucional no qual a capacidade de processamento acelerado possa ser consumida como serviço, de maneira padronizada, controlada, escalável e compatível com o contexto de transformação digital do Exército.

**7.5.9.1.9.** A estrutura analítica da solução, tal como aqui apresentada, não substitui o detalhamento técnico posterior dos componentes e dos itens de aquisição, mas estabelece a referência arquitetural sobre a qual a contratação deverá ser compreendida. Trata-se, portanto, de um desenho orientador, destinado a explicitar como os diversos elementos da futura solução se articulam para entregar, de forma efetiva, a capacidade institucional de processamento em GPU sob demanda, com governança, escalabilidade, segurança e aderência ao ambiente do SisTEx.

**7.5.9.1.10.** A Figura 2 apresenta a visão consolidada, em alto nível, da solução de GPU como serviço, organizada em camadas funcionais e suas subdivisões, que, em conjunto, estruturam a entrega da capacidade de processamento acelerado como um serviço institucional. Essa representação tem por objetivo facilitar a compreensão global da arquitetura, evidenciando os principais blocos que compõem as camadas da solução e a forma como eles se relacionam para atender às demandas de processamento em GPU do Exército.

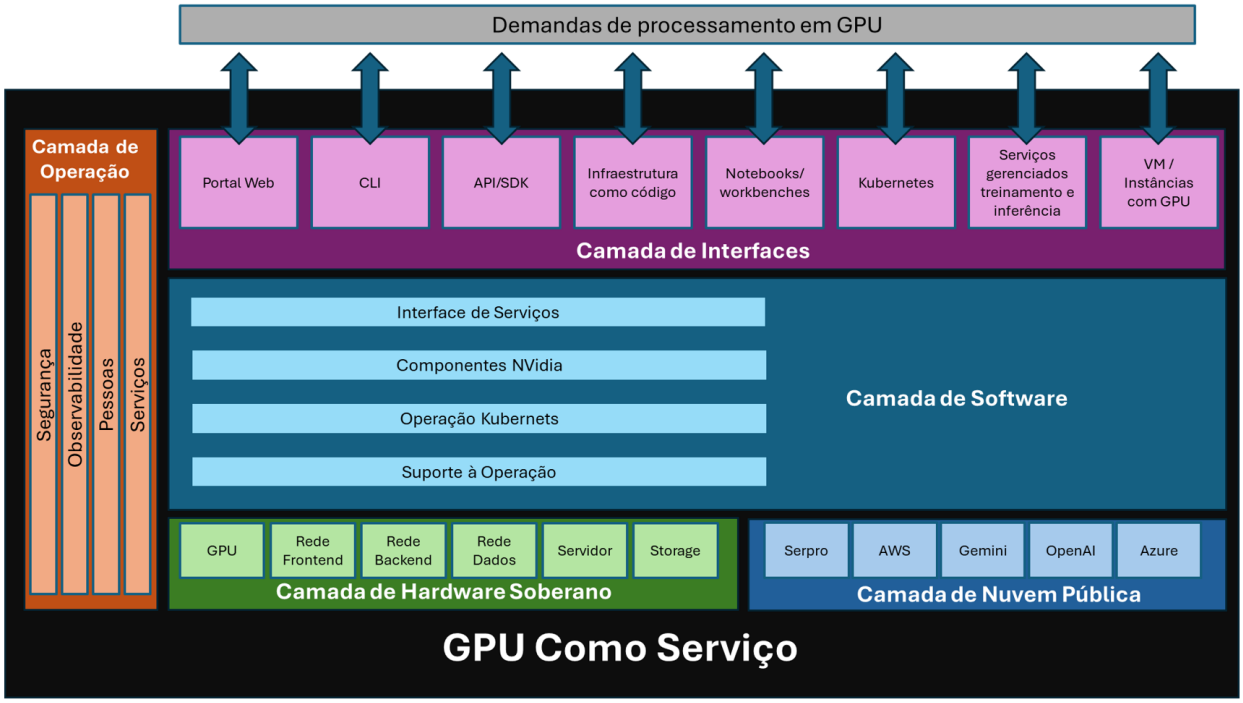



**Figura 2:** Detalhamento das camadas da solução de GPU como serviço.

**7.5.9.1.11.** Conforme se observa, a solução está estruturada em camadas já apresentadas anteriormente e seus componentes. Tais camadas serão, nas subseções seguintes, subdivididas em seus principais componentes internos, de modo a permitir melhor entendimento de sua composição, de suas responsabilidades e de sua contribuição para o funcionamento da solução como um todo.

**7.5.9.2.** Camada de Interfaces

**7.5.9.2.1.** A Camada de Interfaces corresponde ao conjunto de meios pelos quais o usuário técnico, o desenvolvedor, a equipe de dados, o administrador ou mesmo uma aplicação passa a consumir a capacidade de processamento em GPU disponibilizada pela solução. Trata-se, portanto, da camada mais próxima da experiência de uso, sendo responsável por materializar o acesso ao serviço de GPU de forma padronizada, controlada e compatível com os diferentes perfis de consumo previstos para a plataforma.

**7.5.9.2.2.** O funcionamento dessa camada deve permitir que o usuário compreenda, de forma objetiva, que é por meio dela que a capacidade computacional acelerada será efetivamente acessada. A solução não deve ser percebida apenas como um conjunto de servidores GPU em bastidor, mas como um serviço institucional acessível por diferentes formas de interação, a depender do perfil técnico do usuário, do tipo de workload e do grau de automação requerido.

**7.5.9.2.3.** Nesse contexto, o portal web cumpre o papel de interface gráfica principal da solução, permitindo solicitação, acompanhamento e administração do consumo de GPU de maneira simples e intuitiva. É por meio dele que o usuário pode interagir com o serviço sem depender, necessariamente, de comandos técnicos ou integrações programáticas.

**7.5.9.2.4.** A CLI representa a interface textual e automatizável da plataforma, voltada a usuários com maior perfil técnico e a cenários que demandem agilidade, repetibilidade e automação operacional. Seu papel é permitir o consumo da GPU como serviço por meio de comandos padronizados e integráveis a rotinas de trabalho.

**7.5.9.2.5.** As APIs e SDKs constituem o ponto de integração entre a solução de GPU como serviço e outras aplicações, sistemas corporativos, pipelines e automações. Por meio delas, a plataforma expõe capacidades programáticas para provisionamento, consumo, acompanhamento e orquestração do uso da infraestrutura acelerada.

**7.5.9.2.6.** A interface de infraestrutura como código permite que ambientes e recursos da solução sejam descritos de forma declarativa, reprodutível e padronizada. Essa interface é particularmente importante para cenários de automação, governança e repetição controlada de ambientes.

**7.5.9.2.7.** Os notebooks e workbenches representam a forma interativa de acesso ao processamento em GPU, especialmente voltada a atividades de experimentação, desenvolvimento e testes. Nessa modalidade, a solução entrega ao usuário um ambiente controlado de trabalho, no qual a GPU é consumida como serviço em sessões interativas.

**7.5.9.2.8.** A interface baseada em Kubernetes viabiliza o consumo da capacidade acelerada por meio de contêineres, jobs, pods e serviços, permitindo que workloads sejam executados de forma escalável, orquestrada e aderente a arquiteturas modernas de nuvem e plataforma.

**7.5.9.2.9.** Os serviços gerenciados de treinamento e inferência representam uma forma mais abstrata de acesso à solução, permitindo ao usuário consumir capacidades de execução de modelos, treinamento e inferência sem necessidade de administrar diretamente todos os aspectos da infraestrutura subjacente.

**7.5.9.2.10.** Por fim, a interface de VMs ou instâncias com GPU corresponde à forma de acesso mais próxima do modelo tradicional de infraestrutura, oferecendo maior controle direto sobre o ambiente computacional, quando esse perfil de uso se mostrar necessário.

**7.5.9.2.11.** Em conjunto, esses elementos mostram que a Camada de Interfaces é responsável por transformar a infraestrutura acelerada em um serviço efetivamente acessível. É essa camada que

assegura que diferentes perfis de usuário possam consumir GPU sob demanda, cada qual pela forma de interação mais compatível com seu contexto técnico e operacional.

**7.5.9.3.** Camada de Software

**7.5.9.3.1.** A Camada de Software corresponde ao conjunto de componentes lógicos que transforma a infraestrutura acelerada em uma plataforma operacional de GPU como serviço. É nessa camada que se concentram os mecanismos responsáveis por organizar o acesso aos recursos, realizar a orquestração dos workloads, habilitar o uso das GPUs, sustentar os serviços de plataforma e garantir que a solução opere de forma coerente, escalável e governada.

**7.5.9.3.2.** Enquanto a Camada de Interfaces representa os meios de consumo do serviço, a Camada de Software representa o núcleo lógico que faz esse consumo funcionar. Em outras palavras, é essa camada que conecta a intenção de uso do usuário à alocação real dos recursos computacionais acelerados.

**7.5.9.3.3.** A Interface de Serviços cumpre a função de organizar o acesso às capacidades disponibilizadas pela plataforma, estruturando as funções e recursos que poderão ser consumidos por usuários e sistemas. Ela atua como camada organizadora da lógica de exposição dos serviços da plataforma.

**7.5.9.3.4.** Os Componentes Nvidia correspondem ao conjunto de elementos necessários para habilitar o uso efetivo das GPUs, incluindo drivers, runtimes, bibliotecas e demais funções específicas de aceleração por hardware. Trata-se da camada que permite que os aceleradores sejam corretamente reconhecidos, operados e consumidos pelos workloads da solução.

**7.5.9.3.5.** A Operação Kubernetes representa a camada de orquestração dos ambientes, workloads e serviços sobre a infraestrutura acelerada. Sua função é permitir a execução organizada, isolada, escalável e controlada das cargas de trabalho, servindo como elemento central de automação e distribuição do uso dos recursos computacionais.

**7.5.9.3.6.** O Suporte à Operação compreende os componentes auxiliares necessários para gestão, automação, controle e sustentação da plataforma. Incluem-se aqui as funções que não constituem diretamente o núcleo da aceleração, mas que são indispensáveis para o funcionamento estável e administrável do ambiente como um serviço institucional.

**7.5.9.3.7.** Em conjunto, esses elementos fazem com que a solução deixe de ser apenas um parque de hardware acelerado e passe a se comportar como uma plataforma de GPU como serviço propriamente dita. É essa camada que viabiliza o provisionamento, a execução, a administração e a expansão lógica do ambiente.

**7.5.9.4.** Camada de Hardware Soberano

**7.5.9.4.1.** A Camada de Hardware Soberano corresponde ao conjunto de recursos físicos instalados em ambiente próprio, sob controle institucional, destinado a prover a base material de processamento, conectividade e armazenamento necessária à operação da solução de GPU como serviço. Essa camada representa o núcleo físico da arquitetura e é essencial para assegurar soberania operacional, previsibilidade de desempenho e aderência ao ambiente corporativo do Exército.

**7.5.9.4.2.** É nessa camada que se materializa a capacidade computacional acelerada propriamente dita. Sem ela, as camadas superiores não teriam como entregar os serviços previstos. Por essa razão, seu desenho deve ser compreendido como parte estruturante da solução, e não apenas como aquisição de equipamentos isolados.

**7.5.9.4.3.** O Servidor com GPU é o elemento responsável por executar os workloads que demandam processamento acelerado. Trata-se da unidade principal de capacidade computacional da solução, onde se concentram os aceleradores gráficos utilizados em tarefas de treinamento, inferência, desenvolvimento e demais cargas de trabalho compatíveis com a plataforma.

**7.5.9.4.4.** O Servidor para Orquestração sustenta os serviços de controle, gerenciamento e coordenação da plataforma. Sua função é assegurar que a lógica de operação do ambiente permaneça segregada dos nós de processamento principal, contribuindo para estabilidade, governança e continuidade operacional.

**7.5.9.4.5.** A Rede Frontend é responsável por conectar usuários, portais, APIs e demais serviços de acesso à solução. Ela representa a face de entrada da plataforma e deve ser capaz de suportar o tráfego associado às interações dos usuários e sistemas com o serviço.

**7.5.9.4.6.** A Rede Backend sustenta a comunicação interna entre os Servidores com GPU para comunicações específica entre as GPU e demais componentes que precisem usar acesso direto às GPU.

**7.5.9.4.7.** A Rede de Acesso a Dados viabiliza o tráfego entre processamento, armazenamento e fontes de dados. Trata-se de uma camada crítica para workloads que dependem de leitura intensiva de datasets, gravação de checkpoints, acesso a artefatos e circulação eficiente de informações entre computação e storage.

**7.5.9.4.8.** Os Storages representam a base persistente de armazenamento da solução, devendo suportar datasets, imagens, artefatos, logs, modelos e demais dados persistentes relevantes à operação. Seu papel é garantir que a solução disponha de capacidade adequada de persistência, desempenho e retenção para os diferentes perfis de uso previstos.

**7.5.9.4.9.** Em síntese, a Camada de Hardware Soberano representa a base física institucional sobre a qual a solução será construída, permitindo que a plataforma opere com capacidade própria, integração ao ambiente do SisTEx e perspectiva de crescimento modular e controlado.

**7.5.9.5.** Camada de Nuvem Pública

**7.5.9.5.1.** A Camada de Nuvem Pública corresponde ao conjunto de serviços externos que podem complementar ou expandir, quando necessário, a capacidade computacional da solução. Essa camada não substitui a base soberana da arquitetura, mas funciona como capacidade adicional ou complementar em cenários específicos, conforme critérios técnicos, operacionais, orçamentários e institucionais.

**7.5.9.5.2.** A presença dessa camada decorre do entendimento de que, embora a solução tenha como base uma infraestrutura própria, poderá haver situações em que o uso de serviços externos de processamento de IA se mostre conveniente ou necessário. Isso pode ocorrer, por exemplo, em demandas pontuais de expansão, experimentação, homologação, comparação tecnológica ou atendimento complementar a cargas não absorvidas de imediato pelo ambiente soberano.

**7.5.9.5.3.** Os provedores de serviços de processamento de IA representam, nessa lógica, fontes externas de capacidade adicional, capazes de ofertar GPU sob demanda e serviços correlatos. Seu papel, dentro do desenho da solução, é complementar a capacidade da arquitetura, e não redefinir sua natureza principal.

**7.5.9.5.4.** A utilização dessa camada deve ser compreendida como uma extensão possível da solução, devendo sempre observar os requisitos de segurança, governança, integração e compatibilidade institucional aplicáveis. Dessa forma, a nuvem pública passa a ser entendida como capacidade complementar de elasticidade, sem descaracterizar a diretriz de construção de uma base própria e soberana de GPU como serviço.

**7.5.9.6.** Camada de Operações

**7.5.9.6.1.** A Camada de Operações corresponde ao conjunto de capacidades permanentes necessárias para manter a solução segura, monitorada, sustentada e efetivamente utilizável ao longo do tempo. Trata-se de camada transversal à arquitetura, cujo papel é assegurar que a solução não apenas exista do ponto de vista técnico, mas também opere de forma contínua, controlada e aderente às necessidades institucionais.

**7.5.9.6.2.** Essa camada é essencial porque uma solução de GPU como serviço não se esgota na disponibilização de hardware e software. Para que a plataforma produza valor institucional de forma contínua, é necessário que exista uma estrutura de operação capaz de proteger, monitorar, administrar, sustentar e evoluir o ambiente.

**7.5.9.6.3.** A dimensão de Segurança compreende os mecanismos e práticas destinados a proteger acessos, dados, workloads, componentes e integrações da solução. Seu papel é assegurar que a plataforma opere em conformidade com os padrões institucionais de proteção, segregação e controle de acesso.

**7.5.9.6.4.** A dimensão de Observabilidade corresponde ao acompanhamento da saúde, do desempenho, do uso, dos eventos e das falhas da plataforma. É por meio dela que a Administração poderá monitorar o comportamento do ambiente, diagnosticar problemas, acompanhar tendências de consumo e apoiar decisões de expansão ou correção.

**7.5.9.6.5.** A dimensão de Pessoas representa o conjunto de perfis humanos responsáveis pela administração, suporte, sustentação e governança da solução. Inclui, portanto, as equipes técnicas e operacionais necessárias para manter o ambiente funcional, seguro e aderente ao modelo de serviço.

**7.5.9.6.6.** A dimensão de Serviços compreende as atividades de implantação, suporte, manutenção, evolução e operação assistida necessárias à sustentação prática da solução. Trata-se dos serviços especializados que garantem que a plataforma seja implantada, ajustada, estabilizada e mantida de forma compatível com a complexidade do ambiente.

**7.5.9.6.7.** A Camada de Operações, portanto, deve ser entendida como elemento indispensável para a viabilidade de longo prazo da solução. É ela que garante que a arquitetura em camadas não seja apenas tecnicamente concebida, mas efetivamente sustentada como serviço institucional ao longo de todo o seu ciclo de vida.

**7.5.9.7.** Síntese do desenho da solução

**7.5.9.7.1.** A estrutura em camadas ora apresentada consolida o entendimento de que a solução de GPU como serviço deve ser concebida como uma plataforma integrada, composta por diferentes níveis

funcionais e tecnológicos que se complementam mutuamente. A Camada de Interfaces representa o ponto de consumo do serviço. A Camada de Software organiza logicamente a operação do ambiente. A Camada de Hardware Soberano materializa a base física sob controle institucional. A Camada de Nuvem Pública oferece capacidade complementar de expansão. E a Camada de Operações assegura a sustentação contínua do ambiente.

**7.5.9.7.2.** Em conjunto, essas camadas demonstram que a solução não deve ser compreendida como simples aquisição de servidores com GPU, mas como composição coerente de infraestrutura, software, interfaces de consumo, capacidade complementar e mecanismos permanentes de operação. O desenho da solução, portanto, estabelece a referência arquitetural sobre a qual deverão se apoiar a contratação, a implantação inicial, a expansão futura da plataforma e a consolidação do modelo institucional de GPU como serviço.

## 7.6. Estratégia de Contratação e Forma de Implantação da Solução

### 7.6.1. Finalidade da estratégia de contratação

**7.6.1.1.** A presente contratação deverá ser compreendida como a primeira fase da disponibilização institucional da capacidade de GPU como serviço para o Exército, concebida para estabelecer a base tecnológica, lógica e operacional necessária à futura expansão da solução. Seu propósito não se restringe à aquisição pontual de equipamentos, mas à estruturação de uma plataforma corporativa padronizada, escalável e passível de crescimento sob demanda.

**7.6.1.2.** Nesse contexto, a estratégia adotada busca assegurar que a Administração disponha, desde o início, de uma solução efetivamente operacional, ainda que implantada de forma progressiva. Pretende-se, assim, viabilizar a entrada em funcionamento de um ambiente piloto real, capaz de produzir resultados práticos imediatos, sem comprometer a coerência arquitetural necessária à expansão futura.

**7.6.1.3.** A forma de contratação aqui proposta também tem por objetivo compatibilizar três fatores centrais: a necessidade de iniciar rapidamente a disponibilização da capacidade de processamento acelerado; a prudência administrativa e orçamentária inerente à implantação de uma solução dessa natureza; e a necessidade de preservar uma arquitetura de referência tecnicamente robusta, apta a crescer de forma ordenada ao longo do tempo.

### 7.6.2. Arquitetura de referência e capacidade-alvo da primeira fase

**7.6.2.1.** Para fins de arquitetura de referência e dimensionamento global da solução, deverá ser considerada, nesta primeira fase do projeto, a perspectiva de atendimento a um ambiente composto por 10 (dez) servidores de processamento com GPU, cada qual com 8 (oito) GPUs, perfazendo um total de 80 (oitenta) GPUs ao término da expansão prevista para essa etapa do projeto.

**7.6.2.2.** A escolha da arquitetura deverá tomar como referência servidores com 8 (oito) GPUs B300, devendo tal opção ser compreendida como estratégia de centralização da capacidade computacional. Essa diretriz visa maximizar a densidade de processamento por nó, racionalizar o uso de espaço físico no datacenter, otimizar o consumo energético por unidade de capacidade entregue e simplificar a operação, a integração e a sustentação do ambiente.

**7.6.2.3.** A adoção de aceleradores de maior capacidade também se justifica pela necessidade de suportar workloads de maior porte, cenários de inferência mais exigentes, cargas multi-GPU e futuras expansões da solução sem necessidade de redesenho estrutural da arquitetura. Busca-se, com isso, assegurar que a base tecnológica implantada já nasça compatível com o crescimento esperado da demanda institucional por processamento acelerado.

### 7.6.3. Escolha pela NVIDIA em detrimento de AMD e Intel

**7.6.3.1.** A escolha pela arquitetura NVIDIA, em detrimento das alternativas atualmente ofertadas por AMD e Intel, decorre da necessidade de privilegiar a solução com maior maturidade de ecossistema para uso corporativo de inteligência artificial, maior disponibilidade de softwares compatíveis, maior aderência a workloads acelerados já consolidados no mercado e maior previsibilidade de integração com a camada de software prevista para a plataforma.

**7.6.3.2.** Tal diretriz não decorre apenas de comparação de hardware isolado, mas da constatação de que, para fins de constituição de uma solução de GPU como serviço em ambiente corporativo, a disponibilidade de bibliotecas, runtimes, frameworks, drivers, ferramentas de observabilidade, operadores, microserviços de inferência e demais componentes associados à arquitetura NVIDIA apresenta, no estágio atual do mercado, maior grau de consolidação e suporte.

**7.6.3.3.** Adicionalmente, a adoção da arquitetura NVIDIA preserva maior compatibilidade com os principais softwares corporativos considerados para a solução, reduzindo riscos de integração, simplificando a sustentação do ambiente e ampliando a previsibilidade técnica da implantação inicial e das futuras expansões.

**7.6.3.4.** Também deverá ser considerado que grande parte das referências de mercado para ambientes corporativos de GPU como serviço, plataformas de inferência, frameworks de aceleração e ecossistemas de operação multi-GPU se estruturam prioritariamente sobre o ecossistema CUDA e sobre componentes específicos da NVIDIA, o que reforça a racionalidade da escolha para o contexto da presente contratação.

**7.6.3.5.** Dessa forma, a escolha pela NVIDIA se justifica pela combinação entre maturidade tecnológica, aderência ao estado atual do mercado, robustez do ecossistema, previsibilidade de integração e menor risco para a implantação de uma solução corporativa de GPU como serviço no âmbito do Exército.

### 7.6.4. Escolha pela B300

**7.6.4.1.** A escolha pela GPU B300 decorre da necessidade de privilegiar aceleradores de maior capacidade computacional, maior capacidade de memória e maior aderência aos cenários futuros de crescimento da demanda institucional.

**7.6.4.2.** Tal escolha se justifica pela necessidade de suportar workloads de maior porte, inferência mais exigente, treinamento distribuído, ajuste fino de modelos e demais cenários nos quais a disponibilidade de capacidade computacional e memória por acelerador exerçam impacto direto sobre a eficiência da solução.

**7.6.4.3.** A adoção da B300 também se mostra vantajosa em relação a modelos de gerações anteriores ou de menor porte, como forma de reduzir a necessidade de fragmentação excessiva da capacidade entre múltiplos nós, aumentar a densidade computacional por servidor e preservar maior longevidade técnica da infraestrutura.

**7.6.4.4.** Busca-se, com isso, assegurar que a capacidade inicialmente implantada já se estabeleça sobre base tecnológica compatível com as necessidades de médio prazo da solução, evitando obsolescência precoce e reduzindo a probabilidade de redesign arquitetural em fases futuras.

### 7.6.5. Escolha pelo servidor com 8 GPUs

**7.6.5.1.** A escolha por servidores com 8 (oito) GPUs decorre da estratégia de centralização da capacidade computacional acelerada em nós de maior densidade, o que permite melhor aproveitamento do espaço físico, da infraestrutura elétrica, da refrigeração e da interconectividade de alto desempenho entre aceleradores.

**7.6.5.2.** Essa opção também se justifica pelo fato de que servidores com 8 GPUs representam perfil amplamente disponível entre os principais fabricantes de hardware voltado a inteligência artificial, o que favorece padronização, disponibilidade de mercado, comparabilidade entre ofertas e futuras expansões compatíveis com a arquitetura adotada.

**7.6.5.3.** Em termos funcionais, a concentração de 8 GPUs em um único servidor favorece workloads que demandam uso simultâneo de múltiplas GPUs em um mesmo nó, reduzindo dependência de comunicação externa entre servidores e permitindo melhor aproveitamento das interconexões internas de alta velocidade entre os aceleradores.

**7.6.5.4.** Essa diretriz também contribui para simplificação da operação, maior previsibilidade de desempenho e maior eficiência no suporte a cargas de trabalho mais exigentes, tanto em inferência quanto em treinamento e processamento distribuído.

### 7.6.6. Elementos que deverão compor a solução

**7.6.6.1.** Deverão compor a solução, em seu desenho global de referência, os seguintes elementos: 10 (dez) servidores com GPU; servidor ou conjunto de servidores destinados à hospedagem das soluções de software necessárias ao funcionamento completo da plataforma; storage; equipamentos de rede de frontend, backend e acesso a dados; e itens de serviço especializados, abrangendo implantação, operação assistida, capacitação, treinamento, suporte e sustentação da solução.

**7.6.6.2.** No que se refere às camadas lógicas da solução, deverão ser previstos, no mínimo, os seguintes grupos de software: Software de Virtualização e Container; Softwares NVIDIA; Software para Interface de Serviços; e Software de Observabilidade. Tais elementos deverão ser entendidos como componentes estruturantes da plataforma, necessários para transformar a infraestrutura acelerada em um serviço institucional de GPU como serviço, com governança, autosserviço, controle, segurança e capacidade de expansão.

### 7.6.7. Diretriz de aquisição de software

**7.6.7.1.** Todos os softwares a serem adquiridos no âmbito da presente contratação deverão, obrigatoriamente, ser softwares de prateleira, em versão corporativa, com suporte oficial do fabricante, direito a atualização, correções, documentação técnica e mecanismos formais de sustentação do produto.

**7.6.7.2.** Não deverão ser aceitos, para fins de atendimento à solução, softwares livres sem suporte oficial do fabricante, tampouco soluções customizadas ou desenvolvidas sob medida como elemento central da arquitetura. Essa diretriz visa reduzir riscos de implantação e operação, elevar a previsibilidade da sustentação, assegurar continuidade tecnológica, simplificar a governança contratual e garantir maior robustez ao ambiente corporativo de GPU como serviço.

### 7.6.8. Princípio de crescimento sob demanda

**7.6.8.1.** Embora o desenho da solução deva considerar, desde sua origem, a capacidade de crescimento até o limite de 10 (dez) servidores com GPU, a estratégia de contratação e implantação observará o princípio de crescimento sob demanda. Isso significa que a solução deverá ser concebida, contratada e implantada com visão de plataforma completa, porém com expansão física progressiva da capacidade computacional.

**7.6.8.2.** Assim, inicialmente, será adquirida apenas a parcela da solução necessária à entrada em operação do ambiente piloto, sem prejuízo de que os componentes estruturantes já sejam dimensionados para sustentar a futura expansão. Tal abordagem permite reduzir risco de superdimensionamento inicial, distribuir melhor o investimento ao longo do tempo e amadurecer o modelo operacional antes da ampliação integral da capacidade física.

**7.6.8.3.** Essa estratégia também se justifica pela própria natureza evolutiva da demanda de processamento em GPU. O comportamento real dos workloads, a quantidade de usuários concorrentes, o perfil das aplicações desenvolvidas, o porte dos modelos empregados e a distribuição entre inferência, treinamento e ambientes interativos somente poderão ser validados com maior precisão após o início da operação efetiva da plataforma.

### 7.6.9. Justificativas para execução em etapas

**7.6.9.1.** A execução da primeira fase em etapas se justifica, em primeiro lugar, pela necessidade de permitir aprendizado institucional progressivo. A implantação inicial de uma solução dessa natureza tende a gerar informações relevantes sobre padrões reais de consumo, gargalos operacionais, necessidades de ajuste fino da arquitetura e comportamento dos diferentes perfis de uso. A adoção de etapas sucessivas permite que essas informações sejam incorporadas ao processo de expansão da solução.

**7.6.9.2.** A execução escalonada também se justifica pela necessidade de adequação progressiva às demandas dos softwares e serviços que serão desenvolvidos e implantados sobre a plataforma. À medida que as equipes usuárias passarem a utilizar a infraestrutura, poderão surgir demandas específicas relativas a maior memória de GPU, maior largura de banda de rede, diferentes perfis de storage, maior densidade de workloads distribuídos ou novos padrões de uso que recomendem ajustes na forma de expansão da capacidade.

**7.6.9.3.** Outro fundamento relevante para a implantação em etapas é a própria evolução tecnológica do mercado de hardware e software acelerado por GPU. Em um cenário de rápida evolução tecnológica, a contratação escalonada permite que a Administração preserve a coerência da arquitetura geral, sem abrir mão da possibilidade de realizar ajustes técnicos entre etapas, quando isso se mostrar necessário para melhor aderência ao estado da arte, às novas funcionalidades disponibilizadas pelos fabricantes ou às necessidades operacionais efetivamente verificadas.

**7.6.9.4.** A execução em etapas também favorece melhor gestão de risco técnico e financeiro. Em vez de concentrar toda a expansão física da solução em um único momento, a Administração poderá validar a aderência da arquitetura, a qualidade da implantação, a efetividade da operação assistida, o comportamento da plataforma em produção e a suficiência dos componentes lógicos antes de realizar aquisições adicionais de capacidade.

**7.6.9.5.** Dessa forma, a implantação escalonada não deve ser entendida como fragmentação indevida da solução, mas como estratégia de execução compatível com o modelo de crescimento sob

demanda, com a maturação progressiva das necessidades institucionais e com a preservação da flexibilidade técnica necessária ao ajuste fino da solução entre seus ciclos de expansão.

**7.6.10. Componentes estruturantes e componentes expansíveis**

**7.6.10.1.** A lógica de contratação deverá, portanto, separar claramente os componentes que precisam nascer com visão de plataforma completa daqueles que poderão crescer por blocos de capacidade.

**7.6.10.2.** Em termos práticos, os elementos de software, orquestração, interfaces, segurança, observabilidade, governança e hospedagem da camada lógica deverão ser contratados já com capacidade compatível com a solução ampliada ou, quando aplicável, com licenciamento aderente à estratégia de crescimento definida para a plataforma. Isso decorre do fato de que tais componentes estruturam o ambiente como serviço e não apenas sua capacidade física momentânea.

**7.6.10.3.** Por outro lado, os componentes mais diretamente associados à expansão física da capacidade computacional — especialmente servidores adicionais com GPU, expansão de storage e complementos da infraestrutura de rede backend e de acesso a dados — poderão ser adquiridos progressivamente, à medida que a demanda institucional se materialize e justifique o incremento da capacidade instalada.

**7.6.11. Etapa inicial de implantação**

**7.6.11.1.** Na etapa inicial, deverão ser adquiridos 1 (um) servidor com GPU, o servidor destinado à hospedagem das soluções de software da plataforma, os softwares licenciados para toda a solução ou pela quantidade de GPU contratada, conforme o modelo de licenciamento aplicável, a rede de frontend, o storage compatível com a capacidade inicialmente instalada, bem como os itens de serviço necessários à implantação, parametrização, treinamento, operação assistida e sustentação inicial.

**7.6.11.2.** Essa etapa deverá ser suficiente para permitir o funcionamento real da plataforma de GPU como serviço, ainda em escala piloto, porém já com os componentes lógicos e operacionais da solução em funcionamento. O objetivo é que a Administração disponha, desde o início, de uma solução efetivamente utilizável, capaz de atender às primeiras demandas institucionais e de produzir subsídios concretos para o planejamento da expansão subsequente.

**7.6.11.3.** Ainda que a implantação física inicial contemple apenas um servidor com GPU, os componentes de software e os elementos de arquitetura que estruturam a plataforma deverão ser especificados e dimensionados de modo compatível com a futura expansão para os 10 (dez) servidores previstos na arquitetura de referência.

**7.6.12. Etapas subsequentes de crescimento**

**7.6.12.1.** No segundo momento da expansão, conforme a evolução das demandas e a necessidade de ampliação da capacidade computacional, deverão ser adquiridos novos servidores com GPU, a rede de backend completa, as licenças de software adicionais eventualmente necessárias, quando cabíveis, e os serviços correspondentes de expansão, integração, parametrização e sustentação.

**7.6.12.2.** Essa etapa deverá consolidar a infraestrutura de interconexão de alto desempenho entre os nós de processamento, preparando a solução para suportar cargas distribuídas, uso coordenado de múltiplas GPUs e cenários mais intensivos de treinamento e inferência.

**7.6.12.3.** No terceiro momento, deverá ser prevista a continuidade da expansão por meio da aquisição de novos servidores com GPU, de eventuais complementos de storage, de ajustes adicionais de

infraestrutura e das licenças e serviços necessários à manutenção da coerência da solução com a arquitetura de referência.

**7.6.12.4.** Esse crescimento deverá ocorrer de forma modular e controlada, preservando a compatibilidade com os padrões tecnológicos adotados, com a camada lógica já implantada e com a estratégia de centralização e escalabilidade definida para o projeto.

**7.6.13. Possibilidade de ajustes entre etapas**

**7.6.13.1.** Entre uma etapa e outra, a Administração poderá promover os ajustes técnicos necessários à melhor adequação da solução às demandas efetivamente observadas, desde que preservados os princípios de padronização, interoperabilidade, compatibilidade arquitetural e aderência aos objetivos da contratação.

**7.6.13.2.** Tais ajustes poderão decorrer, por exemplo, da identificação de novos perfis de workload, da necessidade de maior capacidade de armazenamento, da revisão de parâmetros de observabilidade, de ajustes na estratégia de alocação de recursos, da evolução dos softwares desenvolvidos sobre a plataforma ou da incorporação de funcionalidades tecnológicas mais adequadas ao contexto institucional no momento de cada expansão.

**7.6.13.3.** Essa possibilidade de ajuste entre etapas é parte da racionalidade da estratégia de contratação proposta e não descaracteriza a solução como um todo. Ao contrário, constitui mecanismo de adaptação controlada, voltado a assegurar que a expansão da plataforma ocorra de forma mais aderente à realidade operacional e à evolução tecnológica, sem comprometer a unidade arquitetural da solução.

**7.6.14. Síntese da estratégia de contratação**

**7.6.14.1.** A forma de contratação aqui proposta busca conciliar, de um lado, a prudência administrativa e orçamentária inerente à implantação inicial de uma nova capacidade institucional e, de outro, a necessidade de estruturar desde já uma solução tecnicamente robusta, passível de expansão ordenada e compatível com a visão de médio prazo da disponibilização de GPU como serviço para o Exército.

**7.6.14.2.** O objetivo é assegurar que o investimento inicial produza resultado prático imediato, sem comprometer a coerência arquitetural, a padronização, a governança e a sustentabilidade da futura expansão da solução. Ao mesmo tempo, a execução em etapas permitirá que a Administração incorpore aprendizado progressivo, ajuste a expansão às demandas efetivamente verificadas e preserve flexibilidade técnica diante da evolução dos softwares desenvolvidos e da própria evolução tecnológica do setor.

## 7.7. Diretrizes de Padronização, Interoperabilidade e Mitigação de Dependência Excessiva de Fornecedor

**7.7.1. Mitigação de dependência excessiva de fornecedor**

**7.7.1.1.** A solução deverá ser concebida e contratada de modo a evitar, na maior medida possível, a dependência excessiva de fabricantes e fornecedores, de forma a preservar a flexibilidade técnica e administrativa ao longo do ciclo de vida da plataforma.

**7.7.1.2.** Para esse fim, os componentes de software da solução deverão, sempre que tecnicamente possível, operar de forma desacoplada da marca do servidor e baseados em padrões, interfaces e

mecanismos de integração que permitam sua utilização sobre infraestrutura compatível de diferentes fabricantes.

**7.7.1.3.** A arquitetura da solução não deverá impor, como requisito implícito ou decorrência indevida de implementação, que futuras expansões da capacidade computacional somente possam ocorrer com servidores de um único fabricante, quando houver no mercado alternativas tecnicamente compatíveis com a mesma estrutura de GPUs e com os requisitos funcionais da solução.

**7.7.1.4.** Assim, ressalvada a necessidade de compatibilidade técnica, homologação dos componentes, suporte oficial e aderência aos requisitos de desempenho, a solução deverá admitir, em futuras aquisições, a incorporação de servidores de diferentes fabricantes, desde que tais equipamentos contemplem a mesma arquitetura de aceleradores adotada nesta contratação, capacidade equivalente de integração à plataforma e compatibilidade com a camada de software, com a rede e com os demais elementos da solução.

**7.7.1.5.** A mesma diretriz deverá ser observada em relação aos componentes de armazenamento, de forma que a solução não seja estruturada de modo a restringir indevidamente futuras expansões a um único fabricante de storage, desde que os equipamentos adicionais sejam compatíveis com os padrões técnicos, protocolos, requisitos de desempenho, mecanismos de integração e modelos operacionais definidos para o ambiente.

### 7.7.2. Agnosticidade da camada de software em relação ao hardware

**7.7.2.1.** Os softwares que compõem a solução deverão ser especificados com foco em funcionalidades, interoperabilidade, suporte oficial e compatibilidade com a arquitetura adotada, e não de forma a vincular indevidamente a plataforma a modelos específicos de servidores ou de storage, salvo quando tal vinculação for estritamente necessária por razões técnicas devidamente justificadas.

**7.7.2.2.** Nesse contexto, a camada de software deverá ser, tanto quanto possível, agnóstica em relação ao hardware subjacente, operando de forma compatível com diferentes equipamentos que atendam aos requisitos de processamento, memória, conectividade, aceleração e suporte exigidos pela solução.

**7.7.2.3.** Tal diretriz visa assegurar que a evolução futura da plataforma possa ocorrer com maior liberdade de composição, ampliando a competitividade nas expansões subsequentes, reduzindo o risco de aprisionamento tecnológico e preservando a capacidade da Administração de selecionar, em cada etapa, os equipamentos mais adequados ao interesse público, desde que compatíveis com a arquitetura de referência.

**7.7.2.4.** A observância dessa diretriz não afasta a necessidade de padronização. Ao contrário, busca-se compatibilizar dois objetivos complementares: de um lado, manter coerência arquitetural e padronização operacional da solução; de outro, evitar que essa padronização seja indevidamente convertida em dependência absoluta de um único fabricante de hardware.

### 7.7.3. Padronização de rede do Data Center do Exército

**7.7.3.1.** No que se refere à infraestrutura de rede, deverá ser observada a padronização de rede de data center em implementação nos Data Centers do Exército, na qual os equipamentos de rede são do fabricante Cisco.

**7.7.3.2.** Assim, os componentes de conectividade previstos para a solução, especialmente aqueles relacionados às redes de frontend, backend e acesso a dados, deverão ser concebidos de forma

aderente a essa padronização, preservando compatibilidade técnica, operacional e administrativa com a infraestrutura de rede já adotada pela Administração.

**7.7.3.3.** A manutenção dessa padronização se justifica, em primeiro lugar, pela necessidade de preservar homogeneidade tecnológica no ambiente de rede do Data Center, reduzindo a complexidade de integração, administração e suporte da solução.

**7.7.3.4.** A adoção de uma única linha tecnológica de rede no ambiente do Data Center favorece a uniformidade de configuração, a previsibilidade de operação, a simplificação do troubleshooting, a redução da curva de aprendizado das equipes técnicas e o melhor aproveitamento do conhecimento institucional já acumulado pela Administração.

**7.7.3.5.** Essa diretriz também contribui para maior eficiência na gestão de sobressalentes, na aplicação de políticas de segurança, na padronização de monitoramento, na execução de mudanças controladas e na manutenção da coerência entre os diversos segmentos da infraestrutura de rede já implantada ou em implantação no Data Center do Exército.

**7.7.3.6.** Do ponto de vista técnico, a manutenção da padronização de rede reduz o risco de incompatibilidades operacionais, diminui a necessidade de tratamentos específicos por fabricante e favorece a integração da nova solução aos processos já estabelecidos de administração, observabilidade, suporte e governança da infraestrutura de Data Center.

**7.7.3.7.** Em ambiente de missão crítica e de alta densidade computacional, como o pretendido para a solução de GPU como serviço, essa uniformidade é especialmente relevante, pois a rede exerce papel central no desempenho dos workloads distribuídos, no acesso a dados, na comunicação entre nós e na sustentação dos serviços de plataforma. A preservação da padronização adotada no Data Center do Exército contribui, portanto, para maior estabilidade operacional, menor risco de integração e melhor governança da solução ao longo de sua expansão.

**7.7.3.8.** A manutenção da padronização de rede no Data Center do Exército, com adoção de equipamentos do fabricante já utilizado na infraestrutura corporativa em implantação, justifica-se não apenas por razões técnicas e operacionais, mas também por aderência ao regime jurídico de contratações públicas estabelecido pela Lei nº 14.133/2021.

**7.7.3.9.** Com efeito, o art. 40, inciso I, determina que o planejamento das compras observe requisitos como qualidade, rendimento, compatibilidade, durabilidade e segurança. De igual modo, o art. 40, inciso V, alínea “a”, estabelece que o planejamento deverá considerar as condições de guarda e armazenamento e atender ao princípio da padronização.

**7.7.3.10.** Nesse contexto, a aquisição e a implementação dos equipamentos de rede do Data Center do Exército foram concebidas em conformidade com essa diretriz legal de padronização do ambiente. Tal orientação visa preservar homogeneidade tecnológica, compatibilidade entre componentes e previsibilidade de operação.

**7.7.3.11.** Assim, a observância da padronização já adotada não constitui mera preferência administrativa, mas medida tecnicamente justificável e juridicamente compatível com a Lei nº 14.133 /2021, na medida em que contribui para melhor desempenho, maior segurança operacional e redução de riscos de integração e sustentação do ambiente.

**7.8. Tratativas com fabricantes**

**7.8.1.** Para subsidiar a elaboração deste Estudo Técnico Preliminar e reduzir os riscos inerentes à contratação de solução de elevada complexidade tecnológica, foram realizadas tratativas técnicas com os fabricantes Dell, Lenovo, HPE, Cisco e Supermicro, conforme materializado no Anexo II – Tratativas com os Fabricantes ao presente Estudo Técnico Preliminar.

**7.8.2.** Tais tratativas tiveram por finalidade ampliar a compreensão da Administração acerca do estado atual das soluções disponíveis no mercado, das arquiteturas ofertadas para ambientes de processamento acelerado, dos componentes normalmente empregados na composição de soluções de GPU como serviço e dos limites técnicos, operacionais e comerciais existentes para o atendimento da demanda institucional do Exército.

**7.8.3.** A realização dessas interações se mostra especialmente importante no contexto desta contratação, uma vez que a solução pretendida não se restringe à aquisição de servidores com GPU, mas envolve a composição coordenada de múltiplos elementos de hardware, software, conectividade, armazenamento, observabilidade, segurança operacional e serviços especializados de implantação e sustentação.

**7.8.4.** Em razão dessa complexidade, a interlocução prévia com fabricantes relevantes do mercado constitui medida de mitigação de risco técnico, pois permite à Administração identificar com maior clareza quais soluções possuem maturidade de mercado, quais arquiteturas se mostram efetivamente aderentes ao modelo de GPU como serviço e quais ofertas apresentam condições de entregar, de maneira integrada, os resultados pretendidos pelo Exército.

**7.8.5.** Essa providência também contribui para reduzir riscos de especificação inadequada, subdimensionamento, incompatibilidades entre componentes, dependência excessiva de integrações improvisadas, baixa previsibilidade de suporte e dificuldades futuras de expansão da solução. Em síntese, trata-se de medida de planejamento voltada a conferir maior robustez, segurança e coerência técnica à futura contratação.

**7.8.6.** No âmbito dessas tratativas, foi elaborado o documento de requisitos da solução de GPU como serviço, constante do Anexo III a este Estudo Técnico Preliminar, com o objetivo de consolidar, em linguagem técnica e funcional, as necessidades da Administração e permitir que os fabricantes pudessem analisar a demanda institucional de forma estruturada.

**7.8.7.** O compartilhamento prévio desse documento com os fabricantes teve por objetivo não restringir a competitividade, mas justamente permitir que a futura contratação seja preparada com maior equilíbrio e transparência, em condições de igualdade entre os potenciais ofertantes, com base em necessidades claramente explicitadas pela Administração e em uma visão comum sobre o escopo da solução pretendida.

**7.8.8.** Nesse contexto, é importante destacar que o objetivo institucional de uma solução de GPU como serviço não pode ser compreendido como mera aquisição de partes isoladas. A Administração não busca, neste caso, a simples reunião de servidores, softwares e serviços tratados de forma desconexa, mas sim a constituição de uma solução única e integrada, capaz de operar como plataforma corporativa de processamento acelerado, com funcionamento harmônico entre seus componentes e aderência aos resultados esperados.

**7.8.9.** Por essa razão, a Administração entende como indispensável que a solução completa, considerada em sua composição de hardware, software e serviços, seja homologada, validada ou

formalmente suportada pelos fabricantes dos componentes principais da arquitetura, especialmente pelo fabricante do hardware base de processamento. Tal diretriz não se destina a restringir a competitividade do certame, mas a assegurar que o conjunto da solução tenha viabilidade técnica comprovada, suporte coerente, previsibilidade de integração e menor risco de insucesso na implantação e operação.

**7.8.10.** A exigência de solução homologada ou validada pelos fabricantes também se justifica pela necessidade de reduzir o risco de que a Administração contrate um arranjo formado por componentes individualmente disponíveis no mercado, porém sem validação conjunta, sem garantia efetiva de interoperabilidade plena, sem clareza de responsabilidades entre fornecedores e sem previsibilidade adequada de evolução e sustentação ao longo do ciclo de vida da solução.

**7.8.11.** Em contratações dessa natureza, o risco não reside apenas na aquisição de um componente inadequado, mas principalmente na contratação de um conjunto tecnicamente frágil, cuja operação dependa de integrações casuísticas, coordenação excessiva entre múltiplos fornecedores ou esforço contínuo da Administração para estabilizar uma solução que já deveria nascer validada. Assim, a diretriz de homologação e validação deve ser compreendida como medida de proteção ao interesse público e de atendimento à real necessidade da Administração, com o menor risco técnico e operacional possível.

**7.8.12. Tratativas com a Dell Technologies**

**7.8.12.1.** No âmbito das tratativas realizadas para subsidiar a elaboração deste Estudo Técnico Preliminar, a Administração encaminhou à Dell Technologies, em 11 de março de 2026, a primeira versão do documento de requisitos da solução de GPU como serviço, com o objetivo de permitir a análise técnica da demanda e o eventual amadurecimento conjunto das especificações aplicáveis à futura contratação.

**7.8.12.2.** Em resposta, a Dell Technologies informou que, antes do início das interações técnicas, seria necessária a indicação de representante do CITEx para assinatura de documento denominado Letter of Assurance – LOA, ou Carta de Garantia, relacionado às normas internacionais de exportação aplicáveis à empresa.

**7.8.12.3.** Segundo a manifestação da Dell Technologies, tal documento seria exigido em razão da natureza da instituição demandante e das regras de controle de exportações aplicáveis a empresas americanas, em especial no contexto de fornecimentos a usuários finais militares fora dos Estados Unidos da América.

**7.8.12.4.** A Dell Technologies informou, ainda, que a ausência de ratificação quanto à assinatura da LOA impediria a Dell ou eventual revenda Dell de prosseguir com as próximas etapas, inclusive em fases que pudessem resultar em entrega, faturamento ou execução contratual.

**7.8.12.5.** A Administração respondeu que, no âmbito do CITEx, não havia previsão nem autorização para assinatura desse tipo de documento, solicitando manifestação da Dell Technologies quanto ao prosseguimento ou não das tratativas técnicas.

**7.8.12.6.** Em nova manifestação, a Dell Technologies reiterou que, por estar sujeita à legislação de controle de exportações dos Estados Unidos, a ausência de LOA assinada por representante competente da Administração inviabilizaria a continuidade do fornecimento, mesmo em caso de eventual adjudicação, esclarecendo tratar-se, segundo a empresa, de restrição regulatória e não de posição comercial.

**7.8.12.7.** Diante desse cenário, a Administração informou que, caso não houvesse devolutiva técnica até o prazo estabelecido, a ausência de manifestação seria considerada negativa formal da Dell Technologies em participar da análise de requisitos e da definição das especificações da contratação.

**7.8.12.8.** Considerando que, no material constante do Anexo II, não houve apresentação de contribuição técnica da Dell Technologies sobre a arquitetura pretendida, composição da solução, aderência ao modelo de GPU como serviço, requisitos de software, desenho de infraestrutura, serviços especializados ou validação de solução completa, conclui-se que a Dell Technologies optou por não participar materialmente das tratativas técnicas destinadas ao amadurecimento das especificações.

**7.8.12.9.** Essa conclusão não constitui juízo de incapacidade técnica da Dell Technologies, tampouco impede, por si só, eventual participação da empresa ou de suas revendas no futuro procedimento licitatório, desde que atendidas as condições editalícias e os requisitos técnicos aplicáveis.

**7.8.12.10.** Contudo, para fins de planejamento da contratação, registra-se que a Dell Technologies não contribuiu tecnicamente para o processo de definição da solução de GPU como serviço pretendida pela Administração, uma vez que condicionou o prosseguimento das tratativas à assinatura prévia de documento não previsto no âmbito do CITEx.

**7.8.12.11.** Assim, diferentemente de fabricantes que prosseguiram nas interações técnicas e apresentaram subsídios, documentos ou propostas de arquitetura, a Dell Technologies não forneceu elementos técnicos que pudessem ser utilizados pela Administração para avaliar, no âmbito deste Estudo Técnico Preliminar, eventual solução completa, integrada, homologada ou validada para atendimento ao modelo de GPU como serviço pretendido pelo Exército Brasileiro.

**7.8.12.12.** Por esse motivo, as tratativas com a Dell Technologies foram consideradas encerradas sem contribuição técnica material para o refinamento das especificações, permanecendo registradas no Anexo II apenas como evidência do esforço de consulta ao mercado e da negativa da empresa em prosseguir na fase de análise de requisitos sem a prévia assinatura da Letter of Assurance – LOA.

**7.8.13. Tratativas com a Lenovo**

**7.8.13.1**. No âmbito das tratativas realizadas para subsidiar a elaboração deste Estudo Técnico Preliminar, a Administração encaminhou à Lenovo, em 11 de março de 2026, a primeira versão do documento de requisitos da solução de GPU como serviço, com o objetivo de permitir a análise técnica da demanda e eventual contribuição para o amadurecimento das especificações da futura contratação.

**7.8.13.2.** Em razão de dificuldades operacionais relacionadas ao recebimento da mensagem inicial, o documento foi novamente encaminhado à Lenovo em 16 de março de 2026, oportunidade em que a empresa confirmou o recebimento e informou que já estaria analisando tecnicamente o descritivo encaminhado pela Administração.

**7.8.13.3.** Na sequência, houve tentativa de agendamento de reunião técnica para tratar da demanda, tendo sido indicadas datas para realização de reunião remota, com o objetivo de possibilitar o aprofundamento das discussões sobre a arquitetura pretendida, a composição da solução, os requisitos de hardware, software, serviços, suporte e validação necessários ao modelo de GPU como serviço.

**7.8.13.4.** Contudo, apesar da manifestação inicial de análise do descritivo e da tentativa de agendamento, não houve avanço material nas tratativas técnicas com a Lenovo em prazo compatível com o andamento do planejamento da contratação.

**7.8.13.5**. Em 4 de maio de 2026, considerando que as tratativas com outros fabricantes já se encontravam em estágio mais avançado, a Administração informou à Lenovo que, caso não houvesse retorno até o final daquela semana, não seria mais possível realizar tratativas técnicas sobre o desenho da solução.

**7.8.13.6.** Em resposta, a Lenovo informou que a Administração deveria seguir a tramitação do processo com as informações já disponíveis dos demais concorrentes, registrando que, naquele momento, encontrava dificuldades para finalizar sua análise e que não seria adequado atrasar o projeto da Administração.

**7.8.13.7.** A manifestação da Lenovo foi interpretada, para fins de planejamento da contratação, como ausência de interesse em prosseguir nas tratativas técnicas destinadas ao desenho da solução e ao refinamento das especificações, ao menos no prazo necessário ao desenvolvimento deste Estudo Técnico Preliminar.

**7.8.13.8.** Registra-se que a Lenovo não apresentou, no âmbito dessas tratativas, contribuição técnica material que permitisse à Administração avaliar, com base em documentos próprios da fabricante, uma arquitetura completa, integrada, homologada ou validada para atendimento ao modelo de GPU como serviço pretendido pelo Exército Brasileiro.

**7.8.13.9.** A ausência de prosseguimento nas tratativas não constitui juízo de incapacidade técnica da Lenovo, tampouco impede eventual participação futura da empresa ou de suas revendas no procedimento licitatório, desde que atendidos os requisitos técnicos, jurídicos e editalícios aplicáveis.

**7.8.13.10.** Entretanto, para fins de elaboração deste Estudo Técnico Preliminar, deve-se registrar que a Lenovo não participou de forma efetiva do processo de amadurecimento do desenho da solução, não tendo fornecido, em tempo útil, elementos técnicos suficientes para subsidiar a definição da arquitetura, da composição da plataforma, dos requisitos de software, dos serviços especializados ou das condições de validação da solução completa.

**7.8.13.11.** Dessa forma, diferentemente dos fabricantes que prosseguiram nas interações técnicas e apresentaram documentos, propostas de arquitetura ou subsídios técnicos para análise da Administração, as tratativas com a Lenovo foram encerradas sem contribuição material para o refinamento das especificações.

**7.8.13.12.** Assim, as tratativas com a Lenovo permanecem registradas no Anexo II como evidência do esforço de consulta ao mercado, da disponibilização isonômica do documento de requisitos e da decisão da própria fabricante de não prosseguir, em tempo hábil, na análise técnica necessária ao desenho da solução de GPU como serviço objeto deste Estudo Técnico Preliminar.

**7.8.14. Tratativas com a HPE**

**7.8.14.1.** No âmbito das tratativas realizadas para subsidiar a elaboração deste Estudo Técnico Preliminar, foram mantidas interações técnicas com a Hewlett Packard Enterprise – HPE, com participação efetiva da empresa Drive A, na condição de revenda e integradora associada à proposta apresentada.

**7.8.14.2.** As tratativas com a HPE tiveram início em março de 2026 e se estenderam por período considerável, envolvendo reuniões, encaminhamento de versões do documento de requisitos, solicitações de detalhamento técnico, pedido de estimativa comercial e posterior envio de minuta de especificações técnicas para comentários.

**7.8.14.3.** Durante esse período, a Administração buscou obter da HPE contribuições capazes de subsidiar o desenho de uma solução completa de GPU como serviço, abrangendo não apenas o servidor de processamento com GPU, mas também a camada de software, a camada de serviços, a integração com rede, armazenamento, suporte, operação assistida, autosserviço, observabilidade e demais componentes necessários à implantação da plataforma.

**7.8.14.4.** A documentação encaminhada pela HPE/Drive A incluiu proposta técnica denominada HPE AI Factory, na qual foi apresentada visão de arquitetura para ambiente de IA, contemplando, em alto nível, infraestrutura, camada operacional, control plane, workloads de IA, segurança, operações de ciclo de vida e observabilidade. A proposta também apresentou uma arquitetura lógica com blocos de software, personas de uso, fluxo de criação de tenants, atribuição de recursos e painéis de administração.

**7.8.14.5.** Também foi encaminhada estimativa comercial pela Drive A, datada de 22 de maio de 2026, contendo valores de referência para itens de hardware, software, armazenamento, redes, serviços, capacitação, operação assistida, suporte e treinamento especializado, com valor total estimado de R$ 47.481.100,00.

**7.8.14.6.** Registra-se, contudo, que as tratativas com a HPE foram marcadas por sucessivas postergações e perda de prazos. Em 14 de maio de 2026, foi concedido prazo final até 22 de maio de 2026 para encaminhamento de material técnico relacionado especificamente à contratação em preparação. Em 23 de maio de 2026, após o decurso do prazo, a Drive A encaminhou o material, registrando pedido de desculpas pelo atraso e informando o esforço realizado pela equipe formada por profissionais da HPE e da própria Drive A para concluir a documentação.

**7.8.14.7.** Em razão do atraso e da necessidade de prosseguimento célere da contratação, a Administração chegou a registrar o encerramento das tratativas técnicas mantidas com a HPE, consignando que, até aquele momento, não havia sido encaminhado material que demonstrasse adequação concreta da solução apresentada às demandas específicas do CITEx, tendo as interações se limitado, em essência, à apresentação do portfólio e das soluções da própria HPE, sem o devido alinhamento ao cenário e aos requisitos indicados pela Administração.

**7.8.14.8.** Não obstante esse registro, a Administração manteve postura colaborativa e prosseguiu na avaliação do material encaminhado, agradecendo a disponibilidade da equipe HPE/Drive A e permitindo que eventuais contribuições adicionais fossem analisadas, desde que houvesse tempo hábil e aderência ao estágio de maturidade já alcançado no processo de planejamento.

**7.8.14.9.** Posteriormente, em 28 de maio de 2026, foi encaminhada à HPE/Drive A a primeira versão das especificações para aquisição, com solicitação expressa para que fossem apresentadas sugestões de revisão, especialmente quanto à parte de serviços, à participação efetiva do fabricante, à parte de hardware, à camada de software e aos detalhes da solução apresentada. A Administração destacou, naquela oportunidade, que a parte de serviços já se encontrava praticamente especificada, mas que seria necessária revisão e sugestão, principalmente quanto aos pontos em que haveria participação efetiva do fabricante.

**7.8.14.10.** Apesar dessa solicitação, a contribuição posteriormente encaminhada pela HPE concentrou-se, essencialmente, em comentários relacionados à parte de hardware e gerenciamento dos servidores HPE ProLiant XD, tendo sido expressamente registrado pela HPE que havia diferenças significativas entre o gerenciamento dos servidores HPE ProLiant tradicionais e o HPE ProLiant XD, antiga linha Cray. Quanto aos demais itens, a HPE informou que aguardaria o texto para análise técnica.

**7.8.14.11.** Em 25 de junho de 2026, a Administração registrou que a expectativa era que a HPE apresentasse sugestões relativas aos itens ainda não detalhados no documento, especialmente itens de software e de infraestrutura complementar, bem como contribuições referentes à parte de switches, com base no que a HPE efetivamente entregaria. Na ocasião, foi solicitado novo envio de sugestões, em razão de as contribuições encaminhadas até então não terem contemplado integralmente tais aspectos.

**7.8.14.12.** Assim, para fins de registro no presente Estudo Técnico Preliminar, verifica-se que a contribuição efetiva da HPE para o refinamento das especificações técnicas foi limitada. Embora tenha havido apresentação de proposta técnica e comercial abrangente, a revisão específica da minuta de especificações encaminhada pela Administração não trouxe contribuições substanciais quanto à camada de serviços da contratação, tampouco observações relevantes sobre os serviços especializados de implantação, operação assistida, suporte, capacitação, participação do fabricante, documentação, monitoramento ou encerramento do projeto.

**7.8.14.13.** Da mesma forma, não foram apresentadas, na fase de revisão das especificações técnicas, contribuições específicas e suficientemente detalhadas sobre a camada de software a ser exigida na contratação, incluindo virtualização, contêineres, orquestração, aceleração por GPU, autosserviço, interface de consumo, observabilidade e segurança operacional. A documentação técnica anteriormente encaminhada tratou tais temas em nível de proposta de arquitetura e portfólio, mas não resultou, em tempo útil, em sugestões objetivas para a especificação técnica da Administração.

**7.8.14.14.** Em consequência, as contribuições efetivamente aproveitáveis da HPE concentraram-se, predominantemente, em pontos de hardware e gerenciamento do servidor HPE ProLiant XD, sem que tenha havido o mesmo nível de detalhamento ou contribuição em relação à solução completa de GPU como serviço, considerada em sua dimensão integrada de hardware, software, serviços, suporte, operação e governança.

**7.8.14.15.** Cumpre registrar, de forma positiva, a participação efetiva da Drive A nas tratativas, especialmente no acompanhamento das comunicações, no envio da proposta técnica e da estimativa comercial e na interlocução com a Administração. A documentação demonstra que a Drive A atuou como participante ativa do processo de apresentação da solução HPE, inclusive registrando o esforço conjunto de profissionais da HPE e da Drive A na preparação do material encaminhado ao CITEx.

**7.8.14.16.** Não obstante a participação da Drive A e a disponibilidade manifestada pelas equipes envolvidas, a Administração deve registrar que a contribuição técnica recebida não foi suficiente para caracterizar participação plenamente efetiva da HPE no amadurecimento de todos os componentes da solução, especialmente no que se refere aos itens de software e serviços, que são essenciais para transformar servidores com GPU em uma plataforma corporativa de GPU como serviço.

**7.8.14.17**. Esse registro não constitui juízo de incapacidade técnica da HPE. Registra-se, contudo, que, posteriormente às interações descritas nesta subseção, a HPE, por intermédio da Drive A, manifestou formalmente a decisão de não prosseguir na participação relacionada à presente contratação, conforme documentação juntada ao Anexo II. Tal manifestação será considerada, para fins deste planejamento, como encerramento das tratativas com a HPE e a Drive A. A desistência manifestada não configura, por si só, impedimento jurídico à eventual participação futura da HPE, de suas revendas ou integradores no procedimento licitatório, caso venha a ocorrer reconsideração da decisão e sejam integralmente atendidos os requisitos técnicos, jurídicos e editalícios aplicáveis.

**7.8.14.18.** Na manifestação formal posteriormente encaminhada pela Drive A, após análise e discussões técnicas conduzidas com especialistas da HPE no Brasil e no exterior, foi informado que o

atendimento às especificações da contratação, especialmente aos pré-requisitos estabelecidos, demandaria quantidade considerada elevada de customizações dos produtos e soluções HPE, bem como integrações com produtos de terceiros. Segundo a manifestação recebida, essas condições tornariam inviável a continuidade da participação sob os aspectos técnico e de competitividade, razão pela qual foi comunicada a desistência de prosseguir no projeto.

**7.8.14.19.** A manifestação da HPE/Drive A deve ser compreendida como decisão daquelas empresas acerca da viabilidade de participação com a arquitetura por elas analisada diante do conjunto de especificações submetido à avaliação, não constituindo, isoladamente, demonstração de impossibilidade técnica absoluta de atendimento ao objeto por produtos HPE nem conclusão da Administração quanto à incapacidade técnica do fabricante. Para fins deste Estudo Técnico Preliminar, contudo, deixa de ser considerada a expectativa de novas contribuições da HPE ou da Drive A para o refinamento das especificações.

**7.8.14.20.** Dessa forma, para fins de instrução do presente Estudo Técnico Preliminar, as tratativas com a HPE devem ser compreendidas como interações relevantes para conhecimento de mercado e obtenção de subsídios pontuais, especialmente quanto a hardware e gerenciamento de servidores, mas com contribuição limitada para a definição integral da arquitetura da solução de GPU como serviço, notadamente nas camadas de software e serviços especializados, tendo sido posteriormente encerradas por iniciativa da HPE/Drive A, que comunicou formalmente a desistência de prosseguir na participação relacionada à presente contratação.

**7.8.14.21.** Em síntese, a HPE/Drive A apresentou proposta técnica e estimativa comercial, participou das interações e forneceu subsídios relacionados à sua abordagem de AI Factory, mas não apresentou, em tempo útil e de forma suficientemente detalhada, contribuições efetivas sobre a parte de serviços das especificações técnicas nem sobre a camada de software da solução, razão pela qual tais pontos foram amadurecidos pela Administração com base em outras referências técnicas, nas necessidades institucionais do CITEx e nos demais subsídios disponíveis no processo. Ao término das tratativas, após nova análise das especificações, a HPE/Drive A comunicou não possuir interesse em prosseguir na participação, em razão da quantidade de customizações e integrações com produtos de terceiros que entendeu necessárias ao atendimento do objeto e dos impactos decorrentes sobre a viabilidade técnica e a competitividade de sua solução.

**7.8.14.22.** Posteriormente à conclusão da versão deste Estudo Técnico Preliminar, a Administração manteve abertas as tratativas com a HPE e a Drive A e, em 14 de julho de 2026, encaminhou a versão mais recente das especificações técnicas para nova análise, facultando às empresas a apresentação de ajustes que entendessem pertinentes, especialmente aqueles destinados a preservar a abertura da especificação, ampliar a competitividade e eliminar requisitos que eventualmente pudessem inviabilizar a participação de solução baseada na arquitetura HPE.

**7.8.14.23.** Em 18 de julho de 2026, foi encaminhada nova atualização das especificações técnicas, acompanhada de solicitação expressa para que a HPE e a Drive A verificassem a viabilidade de atendimento às condições propostas e informassem objetivamente a existência, ou não, de entregáveis, requisitos técnicos, prazos, condições de habilitação ou obrigações de participação do fabricante que fossem considerados impeditivos à apresentação de proposta no futuro procedimento licitatório.

**7.8.14.24.** Até 22 de julho de 2026, não havia sido recebido retorno formal da HPE ou da Drive A quanto à existência de requisitos impeditivos, tampouco confirmação expressa de que a solução por elas representada atenderia integralmente à versão atualizada das especificações técnicas. Posteriormente, em 28 de julho de 2026, a Drive A encaminhou manifestação formal, após análise

conjunta com especialistas da HPE no Brasil e no exterior, comunicando a decisão de desistir de prosseguir na participação relacionada à contratação, pelos motivos detalhados nos subitens seguintes.

**7.8.14.25.** A manifestação posteriormente recebida esclareceu a posição da HPE/Drive A quanto à continuidade das tratativas. Segundo informado, a decisão foi tomada após análise técnica conduzida com especialistas da HPE no Brasil e no exterior, tendo sido identificada, para atendimento às especificações analisadas, necessidade de quantidade considerada elevada de customizações de produtos e soluções HPE e de integrações com produtos de terceiros, circunstâncias que, na avaliação das empresas, inviabilizariam sua participação sob os aspectos técnico e de competitividade.

**7.8.14.26.** A Administração registra os motivos apresentados pela HPE/Drive A como elemento adicional do levantamento de mercado e da avaliação dos riscos da contratação. A manifestação não será interpretada automaticamente como necessidade de alteração das especificações, uma vez que a definição dos requisitos deve permanecer vinculada às necessidades institucionais e ao resultado pretendido. Contudo, por indicar que um dos fabricantes relevantes consultados considerou o conjunto de requisitos incompatível com uma participação técnica e comercialmente competitiva baseada em sua arquitetura, a informação deverá ser considerada na avaliação final da competitividade e na validação das especificações antes da abertura da fase externa.

**7.8.14.27.** Dessa forma, permanece registrado que a Administração facultou à HPE e à Drive A sucessivas oportunidades de participação no refinamento das especificações e de manifestação sobre eventuais barreiras técnicas à sua participação, inclusive mediante o encaminhamento das versões atualizadas dos documentos para análise. Após essas oportunidades, a HPE/Drive A comunicou formalmente sua decisão de não prosseguir na participação relacionada à presente contratação. Consideram-se, portanto, encerradas, por iniciativa das próprias empresas, as tratativas técnicas destinadas ao refinamento da solução HPE no âmbito deste planejamento.

**7.8.14.28.** Registra-se que a desistência foi manifestada pela própria HPE/Drive A e não decorreu de ato da Administração que impedisse sua continuidade nas tratativas. Ao contrário, conforme documentado nesta subseção e no Anexo II, foram encaminhados documentos de requisitos, sucessivas versões das especificações e solicitações expressas para que fossem identificados requisitos considerados impeditivos, facultando-se a apresentação de propostas de ajuste tecnicamente justificadas.

**7.8.14.29.** A manifestação recebida também não permite concluir que produtos ou arquiteturas HPE sejam, em abstrato, incapazes de atender soluções de processamento acelerado ou de Inteligência Artificial. O que se registra é que, diante da configuração específica e do conjunto de requisitos desta contratação, a HPE/Drive A avaliou que seriam necessárias customizações e integrações com produtos de terceiros em quantidade suficiente para tornar sua participação técnica e comercialmente inviável.

**7.8.14.30.** A proposta técnica e comercial anteriormente apresentada pela Drive A permanecerá nos autos e poderá ser utilizada como evidência histórica das alternativas analisadas e dos preços apresentados durante o levantamento de mercado. Entretanto, após a manifestação de desistência, essa proposta não deverá ser interpretada como demonstração de disponibilidade atual da solução HPE para participação no certame, nem como confirmação de que a arquitetura nela descrita atende integralmente às especificações consolidadas da contratação.

### 7.8.15. Tratativas com a Supermicro

**7.8.15.1.** As tratativas com a Supermicro tiveram início em momento posterior às interações realizadas com os demais fabricantes consultados neste processo de planejamento. Tal circunstância decorreu do fato de a Supermicro não ter participado, até então, de processos licitatórios de servidores conduzidos pelo CITEx, não havendo histórico prévio de interlocução institucional equivalente ao existente com outros fabricantes tradicionalmente atuantes nas contratações do Centro.

**7.8.15.2.** Não obstante esse início posterior, considerando a relevância internacional da Supermicro no fornecimento de hardware para ambientes de alto desempenho, processamento acelerado e infraestrutura de inteligência artificial, a Administração entendeu adequado oportunizar sua participação nas tratativas técnicas, de modo a ampliar o conhecimento de mercado e permitir que eventual solução baseada em hardware Supermicro pudesse ser avaliada no contexto da contratação pretendida.

**7.8.15.3.** As tratativas envolveram a participação de duas revendas, Scherm e Planejeti, bem como representante da Supermicro em cópia nas comunicações. A documentação constante do Anexo II demonstra que as reuniões e os encaminhamentos técnicos foram conduzidos predominantemente pela Scherm, com participação da Planejeti na interlocução comercial e no agendamento das interações.

**7.8.15.4.** A primeira reunião registrada no conjunto documental ocorreu em 21 de maio de 2026, com participação de representantes da Supermicro, Scherm, Planejeti e CITEx. Na sequência, a Administração solicitou a transcrição da reunião e prosseguiu no envio de versões atualizadas do documento de requisitos e das especificações técnicas para análise.

**7.8.15.5.** Em 28 de maio de 2026, foi encaminhada à Supermicro, Scherm e Planejeti a primeira versão das especificações para aquisição, tendo a Administração destacado que a parte de serviços já se encontrava praticamente especificada, mas que se esperava revisão e sugestões, especialmente quanto aos pontos em que haveria participação efetiva do fabricante. Na mesma oportunidade, foi solicitado que a parte de hardware e software contivesse os detalhes da solução apresentada, bem como que fosse encaminhada precificação mais assertiva e explicação sobre a forma como os softwares interagiriam para entregar os objetivos da contratação.

**7.8.15.6.** Em resposta, foram encaminhadas dúvidas técnicas pela Scherm, com foco na implementação da plataforma em ambiente air-gapped ou com conectividade controlada, espelhamento de repositórios RPM/APT/PyPI, validação de proveniência de modelos, observabilidade com Datadog em modo on-premise, retenção de logs de auditoria, certificados TLS emitidos pela CA interna do Exército, sincronização de tempo por NTP interno e resolução DNS pelos servidores internos do EB.

**7.8.15.7.** Essas dúvidas indicaram que a proposta apresentada pela Scherm dependia de diversas definições complementares de arquitetura, integração, segurança, observabilidade e operação em ambiente restrito, especialmente no que se refere à implantação de mirrors internos, à validação de modelos, à retenção de logs, à observabilidade e à operação da plataforma sem dependência de serviços externos.

**7.8.15.8.** Posteriormente, a Scherm encaminhou proposta técnica e comercial de GPU como serviço, apresentando-se como parceira Supermicro, NVIDIA e Red Hat, com solução estruturada em camadas de operações, interfaces, software e hardware soberano. A proposta indicava, entre outros

componentes, servidores GPU Supermicro, Red Hat OpenShift Platform Plus, OpenShift AI, NVIDIA AI Enterprise, Run Workload Orchestration, NVIDIA Base Command Manager, Datadog, NIM, NeMo, storage interno e posterior possibilidade de storage externo.

**7.8.15.9.** O diagrama de arquitetura posteriormente encaminhado pela Scherm apresentou a solução como operação 100% on-premises, sem comunicação com serviços externos em produção, com integração a LDAP/AD, CA interna, NTP e DNS do Exército, além de interfaces por Run Portal, OpenShift Console, JupyterHub e APIs REST. O mesmo diagrama apontou Datadog Private on-premise, Kubecost Enterprise, mirrors RPM/APT/PyPI/NGC e auditoria imutável como componentes da arquitetura proposta.

**7.8.15.10.** Apesar da abrangência aparente da proposta, permaneceram dúvidas relevantes quanto à natureza da solução apresentada, especialmente no que se refere à existência de uma solução completa do tipo turn-key, homologada e suportada pela Supermicro como fabricante do hardware, abrangendo a totalidade da stack de software, serviços, integrações, observabilidade, auditoria, chargeback, operação airgap e interfaces de consumo.

**7.8.15.11.** Em 25 de junho de 2026, após o recebimento das sugestões, a Administração registrou que todos os softwares deveriam possuir licenciamento pelo período requerido e solicitou esclarecimentos adicionais sobre a forma como os softwares interagiriam para entregar os objetivos da contratação.

**7.8.15.12.** Em 26 de junho de 2026, após análise preliminar, a Administração questionou a arquitetura de comunicação indicada no material, especialmente quanto à conectividade entre o servidor de orquestração, os demandantes e o servidor com GPU, uma vez que o documento sugeria uso da rede de backend para parte dessas interações, o que demandava esclarecimento quanto à aderência ao desenho esperado de segregação e funcionamento da solução.

**7.8.15.13.** Na mesma data, a Administração solicitou esclarecimentos adicionais considerados essenciais, incluindo: apresentação de matriz de suporte fabricante a fabricante, indicando exatamente qual contrato cobriria cada componente, inclusive Loki, Triton, Kubecost, mirrors e auditoria; comprovação documental do fabricante para a alegação de Datadog fully on-premise e zero SaaS; documentação pública da Supermicro de validação de uso de toda a stack de software ofertada; detalhamento da camada de interfaces, cobrindo CLI, infraestrutura como código, VMs ou instâncias GPU, serviços gerenciados e experiência unificada de autosserviço; e detalhamento do que seria produto pronto e do que dependeria de integração ou customização, inclusive chargeback, auditoria imutável, proveniência de modelos e operação airgap.

**7.8.15.14.** Tais questionamentos foram centrais para a avaliação da proposta, pois buscavam justamente distinguir uma solução efetivamente homologada, validada e suportada pelo fabricante de hardware de uma composição integradora formada por componentes de múltiplos fornecedores, cuja interoperabilidade e sustentação dependeriam predominantemente das revendas responsáveis pela montagem da solução.

**7.8.15.15.** A Administração reconhece que a Supermicro possui documentação pública relacionada a AI Factory e SuperClusters baseados em arquiteturas de referência NVIDIA, com oferta de soluções em escala de rack e cluster, voltadas a desempenho, compatibilidade e implantação de ambientes de IA.

**7.8.15.16.** Todavia, a existência de documentação pública geral da Supermicro sobre AI Factory não se confundiu, nas tratativas realizadas no Brasil, com a apresentação de uma solução completa, turn-key, homologada e suportada pela Supermicro para a totalidade da stack proposta pela Scherm. A proposta

efetivamente apresentada nas tratativas locais não demonstrou, de forma suficiente, que a combinação de hardware Supermicro, softwares de terceiros, observabilidade, chargeback, auditoria, mirrors, operação airgap, interfaces e serviços especializados estivesse formalmente validada e suportada pelo fabricante do hardware como solução integrada.

**7.8.15.17.** Observa-se, ainda, que a arquitetura apresentada pela Scherm indicou a própria Scherm como ponto focal de operação e suporte em português, bem como responsável pela estruturação da camada de operações, implantação, capacitação, operação assistida e suporte, o que reforça a percepção de que a responsabilidade prática pela integração da solução permaneceria majoritariamente a cargo das revendas e não do fabricante Supermicro como provedor de solução completa.

**7.8.15.18.** Esse ponto é especialmente relevante porque a presente contratação busca justamente o cenário oposto: uma solução do tipo turn-key, integrada, homologada ou formalmente validada pelo fabricante dos componentes principais, com participação efetiva do fabricante na implantação, nos entregáveis críticos, no suporte e na sustentação da plataforma, reduzindo riscos de integração, indefinição de responsabilidades e dependência de customizações não suportadas.

**7.8.15.19.** Na fase de revisão das especificações técnicas, as contribuições encaminhadas pela Scherm trouxeram sugestões relacionadas a requisitos de hardware, rede backend, plataforma de orquestração, aceleração por GPU, autosserviço, observabilidade e licenciamento. Todavia, não foram identificadas contribuições equivalentes sobre a especificação da parte de serviços, tampouco sobre a forma de participação do fabricante Supermicro na execução das atividades de planejamento, desenho, implementação, validação, operação assistida, suporte, documentação e encerramento do projeto.

**7.8.15.20.** Essa ausência de manifestação específica sobre a participação do fabricante nos serviços é incompatível com a diretriz adotada pela Administração, segundo a qual a efetiva participação do fabricante constitui mecanismo essencial de mitigação de risco, especialmente em solução de alta complexidade tecnológica e com múltiplas camadas de hardware, software, rede, segurança, observabilidade e operação.

**7.8.15.21.** Após o envio dos questionamentos de 26 de junho de 2026, não houve resposta suficiente para afastar integralmente as dúvidas apresentadas pela Administração. Em 3 de julho de 2026, a Administração registrou que, pelas tratativas realizadas, vários pontos cruciais da demanda não eram atendidos pela proposta enviada e tampouco haviam sido apontados como impeditivos na revisão realizada.

**7.8.15.22.** Em 6 de julho de 2026, diante de questionamento posterior da Planejeti sobre quais seriam os pontos de maior dúvida, a Administração consignou que os e-mails enviados desde 26 de junho haviam ficado sem resposta e informou que, para fins do Estudo Técnico Preliminar e da fase interna da licitação, já possuía as informações necessárias para prosseguir.

**7.8.15.23.** Dessa forma, para fins de instrução do presente Estudo Técnico Preliminar, conclui-se que as tratativas com a Supermicro, conduzidas no Brasil por intermédio das revendas Scherm e Planejeti, permitiram conhecer proposta baseada em hardware Supermicro e em componentes de software de terceiros, mas não demonstraram, de forma satisfatória, a existência de solução completa, turn-key, homologada e suportada pela Supermicro para atendimento integral ao modelo de GPU como serviço pretendido pelo Exército Brasileiro.

**7.8.15.24.** A solução apresentada permaneceu com integrações e detalhamentos relevantes em aberto, especialmente quanto à matriz de suporte entre fabricantes, à validação da stack completa pela Supermicro, à comprovação da observabilidade integralmente on-premise, à distinção entre produtos prontos e customizações, à camada unificada de interfaces, à operação airgap, à auditoria, ao chargeback e à responsabilidade do fabricante nas atividades de implantação e sustentação.

**7.8.15.25.** Assim, embora se reconheça a relevância do hardware Supermicro no mercado internacional de processamento acelerado e a existência de documentação pública do fabricante relacionada a AI Factory, a proposta apresentada no âmbito das tratativas nacionais não se caracterizou como solução completa, homologada e suportada pelo fabricante para todos os componentes da arquitetura, transferindo risco substancial de integração, interoperabilidade, operação e sustentação para as revendas.

**7.8.15.26.** Esse cenário não se mostra aderente ao objetivo da presente contratação, que busca reduzir o risco técnico por meio de solução integrada, validada, suportada e entregue em regime turn-key, com participação efetiva do fabricante nos pontos críticos da implantação e da sustentação.

**7.8.15.27.** O registro dessas limitações não constitui juízo de incapacidade técnica da Supermicro, da Scherm ou da Planejeti, tampouco impede eventual participação futura do fabricante ou de suas revendas no procedimento licitatório, desde que atendidos integralmente os requisitos técnicos, jurídicos e editalícios aplicáveis.

**7.8.15.28.** Contudo, para fins de definição da arquitetura de referência e de fundamentação do planejamento da contratação, as tratativas com a Supermicro foram consideradas insuficientes para sustentar o desenho da solução pretendida pela Administração, em razão da ausência de comprovação adequada de solução completa, turn-key, homologada e suportada pelo fabricante, bem como da permanência de dúvidas relevantes não plenamente esclarecidas no curso das interações.

**7.8.16. Tratativas com a Cisco**

**7.8.16.1.** No âmbito das tratativas realizadas para subsidiar a elaboração deste Estudo Técnico Preliminar, a Cisco foi o fabricante que atuou de forma mais intensa, estruturada e colaborativa no apoio à definição técnica da solução de GPU como serviço pretendida pela Administração.

**7.8.16.2.** As interações com a Cisco envolveram a participação direta de equipe multidisciplinar do fabricante, contemplando profissionais das áreas de infraestrutura de nuvem e IA, engenharia de soluções, segurança cibernética, observabilidade e portfólio, conforme evidenciado na documentação encaminhada sob o título “Arquitetura de Fábrica de IA Segura – CITEx – Cisco Secure AI Factory with NVIDIA”.

**7.8.16.3.** A disponibilidade de engenheiros e especialistas do fabricante ao longo das tratativas permitiu à Administração aprofundar a análise de requisitos de hardware, rede, armazenamento, software, observabilidade, segurança, suporte, serviços especializados, operação assistida, licenciamento e evolução futura da solução.

**7.8.16.4.** Diferentemente de tratativas mais restritas a apresentações genéricas de portfólio ou a contribuições pontuais sobre hardware, a Cisco apresentou material específico para o contexto do CITEx, incluindo arquitetura de referência de Fábrica de IA Segura, exemplos de dimensionamento, blocos de software, componentes de infraestrutura, desenho lógico, opções de armazenamento, serviços especializados de implantação e contribuições diretas para a minuta de especificações técnicas.

**7.8.16.5.** A arquitetura apresentada pela Cisco baseou-se no conceito de Cisco Secure AI Factory with NVIDIA, descrito como desenho modular de referência que combina infraestrutura de alto desempenho com segurança e observabilidade full-stack, incluindo Cisco Compute com NVIDIA Accelerated Computing, Cisco AI Networking, plataforma Kubernetes, NVIDIA AI Enterprise, Splunk Observability, gestão e operações e segurança de workloads.

**7.8.16.6.** A documentação apresentada também indicou, de forma estruturada, os principais blocos da solução, incluindo Cisco AI PODs, opções de servidores UCS baseados em NVIDIA HGX, MGX ou RTX PRO, gerenciamento por Cisco Intersight, Cisco Data Center Switching, Nexus Dashboard, Isovalent, opções de storage parceiro, Red Hat OpenShift, NVIDIA AI Enterprise, NeMo, NIM Blueprints, orquestração de workloads de IA e GPU, Splunk Enterprise/ITSI, Splunk Enterprise Security, Cisco AI Defense e demais componentes de segurança e observabilidade.

**7.8.16.7.** A Cisco também apresentou exemplo de bill of materials inicial para uma configuração de 8 GPUs, contemplando servidor Cisco UCS C880A com NVIDIA HGX/B300, appliances Cisco UCS C225 para cluster Qumulo, servidores de gerenciamento para Intersight, control cluster, Kubernetes, Run, Rafay e Splunk Heavy Forwarder, além de licenças de NVIDIA AI Enterprise, Run, Rafay AI/ML, Red Hat, Isovalent, Splunk, Cisco Intersight e licenças Nexus.

**7.8.16.8.** Um aspecto relevante das tratativas com a Cisco foi a capacidade demonstrada de articular e integrar outros fabricantes e fornecedores especializados em componentes complementares da solução, em especial Qumulo e Rafay, reforçando a viabilidade de construção de solução integrada e turn-key com múltiplos fabricantes, sob coordenação técnica do fabricante principal.

**7.8.16.9.** No caso da Qumulo, a Cisco coordenou reunião específica e encaminhou posteriormente o material produzido pelo time da Qumulo, incluindo apresentação institucional e visão de dimensionamento realizada para o CITEx. O material de dimensionamento apresentou cluster Qumulo com 10 nós, 4.387,43 TB úteis, escalabilidade de capacidade e indicadores de desempenho, demonstrando esforço concreto de integração da camada de armazenamento ao desenho da solução de IA.

**7.8.16.10.** A apresentação Qumulo and Cisco for CITEx demonstrou a atuação conjunta entre Cisco e Qumulo para plataforma de dados de alto desempenho, com aplicação em AI/ML, HPC, simulação, analytics, log aggregation, backup/recovery e outros fluxos de dados críticos, reforçando a abordagem de ecossistema integrado para armazenamento em ambientes de inteligência artificial.

**7.8.16.11.** No caso da Rafay, as tratativas também foram conduzidas em articulação com a Cisco, com participação de representantes da Rafay e encaminhamento de material sobre GPU Platform as a Service, destacando capacidades de gerenciamento multicloud de GPUs, governança, multi-tenancy e autosserviço para desenvolvedores.

**7.8.16.12.** A participação de Qumulo e Rafay nas tratativas, capitaneada pela Cisco, é relevante porque evidencia a capacidade do fabricante de coordenar componentes de diferentes fornecedores dentro de uma arquitetura integrada, reduzindo o risco de que a Administração tenha que conduzir isoladamente a compatibilização entre hardware, rede, armazenamento, software, autosserviço, observabilidade e suporte.

**7.8.16.13.** Essa abordagem é coerente com a diretriz da contratação de buscar solução do tipo turn-key, na qual a contratada deverá entregar plataforma funcional e integrada, com responsabilidades claras sobre desenho, implantação, validação, suporte, operação assistida, documentação e sustentação.

**7.8.16.14.** A documentação de suporte da Cisco também reforça a possibilidade de tratamento da solução em nível integrado, ao prever Solution Level Support, com expertise centralizada e gerenciamento de incidentes envolvendo produtos Cisco e produtos de provedores terceiros aprovados como Solution Alliance Partners.

**7.8.16.15.** No tocante aos serviços especializados, a Cisco apresentou descritivo detalhado para desenvolvimento de requisitos, design da solução, plano de implementação, implantação, configuração, documentação, operação e suporte, com atividades e entregáveis específicos, tais como questionário de pré-implantação, workshop presencial de design, análise de lacunas, Documento de Requisitos da Solução, Documento de Design da Solução e Plano de Implementação.

**7.8.16.16.** O descritivo de serviços apresentado pela Cisco contemplou a revisão da documentação existente do CITEx, coleta de requisitos de negócio, técnicos e operacionais, requisitos de escalabilidade, lacunas de recursos, requisitos do cluster de gerenciamento, instalação do Nexus Dashboard e Nexus Dashboard Fabric Controller, desenho da fabric de frontend, requisitos de computação, instalação de servidores com GPUs, sistemas operacionais suportados, containers, drivers, firmware, CUDA e NCCL, além de integração com plataforma de dados Qumulo.

**7.8.16.17.** No desenvolvimento do design da solução, a documentação Cisco previu cluster de gerenciamento OpenShift com alta disponibilidade e escalabilidade, implantação do Cisco Nexus Dashboard, desenho da fabric de frontend, estratégia de rede de armazenamento com VLANs, MTU e políticas de QoS, arquitetura de computação com nós C880A e configurações de GPU e integração com Cisco Intersight.

**7.8.16.18.** Ainda no campo dos serviços, a documentação apresentada incluiu atividades de configuração de servidor UCS C880A com GPU, instalação de sistema operacional suportado, atualização de firmware e drivers, instalação de CUDA e NCCL, implantação de Red Hat OpenShift Container Platform, implantação da pilha NVIDIA AI Enterprise para workloads de IA, aceleração por GPU e Run, além de gerenciamento centralizado por Cisco Intersight em modelo SaaS ou Appliance.

**7.8.16.19.** A Cisco, portanto, não se limitou a apresentar os componentes de hardware da solução. Houve contribuição efetiva sobre a forma de prestar serviços especializados, sobre os entregáveis necessários ao ciclo de implantação, sobre a necessidade de workshops técnicos, sobre critérios de design, sobre integração com ferramentas de gerenciamento, sobre os componentes de software e sobre a operação da solução como plataforma de IA.

**7.8.16.20.** No que se refere aos itens de hardware, as contribuições da Cisco abrangeram servidores de processamento com GPU, servidores de gerenciamento, rede de frontend, rede backend de alta velocidade, armazenamento parceiro, gerenciamento por Intersight, switches Nexus, integração com Qumulo e desenho de arquitetura para ambientes com NVIDIA HGX/B300, incluindo discussões de conectividade, desempenho, escalabilidade e requisitos de rede para workloads de treinamento e inferência.

**7.8.16.21.** No que se refere aos itens de software, a Cisco apresentou uma composição mais abrangente e alinhada ao objetivo de GPU como serviço, contemplando, entre outros, Red Hat OpenShift, NVIDIA AI Enterprise, Run, Rafay, Splunk, Isovalent, Cisco Intersight, Nexus Dashboard, componentes de segurança, observabilidade e gerenciamento, além de referências à integração com AI Defense e demais elementos de proteção da camada de IA.

**7.8.16.22.** Também foram apresentadas contribuições sobre licenciamento, contemplando licenças de software associadas à configuração inicial, incluindo NVIDIA AI Enterprise, Run, Rafay AI/ML, Red Hat,

Isovalent, Splunk, Cisco Intersight e licenças Nexus, o que permitiu à Administração visualizar melhor a estrutura lógica necessária para que a solução não se restringisse à aquisição de servidores com GPU.

**7.8.16.23.** A participação da Cisco também se materializou na revisão da minuta de especificações encaminhada pela Administração. Em 28 de maio de 2026, o CITEx encaminhou a primeira versão das especificações para aquisição e solicitou sugestões, especialmente quanto à parte de serviços, participação efetiva do fabricante, hardware e software. Em 23 de junho de 2026, a Cisco encaminhou primeira revisão com sugestões nas partes em aberto e informou que marcaria reunião para colher impressões e realizar ajustes na semana seguinte.

**7.8.16.24.** Posteriormente, em 2 de julho de 2026, por intermédio da Teltec, foi encaminhado documento atualizado com sugestões, evidenciando continuidade das interações e refinamento das contribuições à especificação técnica. Em 3 de julho de 2026, a Administração registrou o recebimento das considerações, informou que todos os pontos haviam sido considerados e agradeceu o apoio no processo de especificações da solução.

**7.8.16.25.** Registra-se, ainda, a participação do fornecedor Teltec nas tratativas, atuando em conjunto com a Cisco na interlocução comercial e no encaminhamento de contribuições atualizadas à Administração. A documentação demonstra a participação de representantes da Teltec nas comunicações, inclusive com envio formal de documento atualizado contendo sugestões à especificação técnica.

**7.8.16.26.** A participação da Teltec foi relevante para dar suporte à materialização das contribuições técnicas em formato aproveitável para o processo de planejamento, funcionando como elo de interlocução entre fabricante, integrador/fornecedor e Administração, sem afastar a participação direta dos engenheiros e especialistas da Cisco.

**7.8.16.27.** Para fins de comparação com as demais tratativas realizadas, a contribuição da Cisco foi a mais completa e aderente ao objetivo da contratação, pois abordou de forma integrada componentes de hardware, software, licenças, serviços especializados, suporte, armazenamento parceiro, autosserviço, observabilidade, segurança, gerenciamento, rede e operação.

**7.8.16.28.** A amplitude das contribuições encaminhadas pela Cisco permitiu à Administração amadurecer a compreensão de que a solução de GPU como serviço deve ser tratada como plataforma integrada e não como aquisição isolada de servidores com GPU. Essa contribuição foi especialmente relevante para reforçar a necessidade de arquitetura validada, participação efetiva do fabricante, serviços especializados, licenciamento empresarial, suporte oficial e responsabilidade integrada pela entrega.

**7.8.16.29.** A atuação da Cisco também evidenciou a importância de que a solução possua capacidade de integração entre diferentes fabricantes e tecnologias, especialmente quando envolver storage, observabilidade, segurança, orquestração de workloads, autosserviço e infraestrutura acelerada. A coordenação de interações com Qumulo e Rafay demonstrou, na prática, uma abordagem mais compatível com a entrega de solução turn-key.

**7.8.16.30.** A documentação apresentada pela Cisco, contudo, não deve ser interpretada como imposição de marca ou como definição exclusiva de arquitetura, mas como subsídio técnico relevante para a elaboração do Estudo Técnico Preliminar e das especificações, observada a necessidade de preservar a competitividade, a isonomia e a possibilidade de atendimento por soluções equivalentes que comprovem o mesmo nível de integração, validação, suporte e aderência aos requisitos da Administração.

**7.8.16.31.** Assim, para fins de instrução do presente Estudo Técnico Preliminar, registra-se que as tratativas com a Cisco foram as que apresentaram maior maturidade técnica, maior disponibilidade de especialistas do fabricante, maior abrangência de contribuições e maior aderência ao conceito de solução completa de GPU como serviço, especialmente em razão da abordagem integrada de hardware, software, rede, armazenamento, segurança, observabilidade, suporte e serviços especializados.

**7.8.16.32.** Em síntese, a Cisco, com participação do fornecedor Teltec e com articulação de fabricantes e parceiros como Qumulo e Rafay, contribuiu de forma substancial para o amadurecimento do desenho da solução, oferecendo subsídios técnicos relevantes para a definição de uma arquitetura integrada, validada, suportada e compatível com o objetivo de contratação turn-key pretendido pela Administração.

**7.8.16.33.** Como desdobramento das tratativas técnicas e comerciais mantidas com a Cisco e a Teltec, foi recebida, em 17 de julho de 2026, a proposta comercial inicial TTEC3067/2026, versão 1.0, para fornecimento de solução de GPU como serviço baseada em arquitetura Cisco, contemplando componentes de hardware, garantias, licenças de software e os serviços especializados necessários à implantação e à sustentação inicial da plataforma.

**7.8.16.34.** A proposta inicial, versão 1.0, apresentou valor global de R$ 26.700.000,00 (vinte e seis milhões e setecentos mil reais) e prazo de validade de 60 dias, constituindo, naquele momento, o menor valor global entre os orçamentos formalmente recebidos pela Administração. Para fins comparativos, a estimativa anteriormente encaminhada pela Drive A para solução baseada em HPE apresentou valor global de R$ 47.481.100,00. Os valores desta proposta inicial foram posteriormente substituídos, para fins de referência do planejamento, pelos constantes da versão 2.0, conforme registrado nos subitens seguintes.

**7.8.16.35.** Em razão de sua aderência à composição da contratação, o orçamento encaminhado pela Teltec foi inicialmente utilizado como parâmetro técnico-orçamentário para estruturação do planejamento. Após o recebimento da Proposta TTEC3067/2026, versão 2.0, esta passa a constituir a referência comercial mais atual das tratativas com a Teltec, sem prejuízo da realização e da consolidação da pesquisa de preços na forma da regulamentação aplicável e da possibilidade de alteração do valor estimado definitivo da contratação.

**7.8.16.36.** A composição individualizada dos valores atribuídos aos 33 entregáveis também será utilizada como subsídio para a elaboração do cronograma físico-financeiro da contratação, especialmente para a definição da participação percentual de cada entregável no valor global, dos marcos de execução e aceite, das parcelas de pagamento e da previsão de desembolso ao longo da vigência contratual. A utilização dessa composição permitirá relacionar os pagamentos à efetiva execução e ao recebimento de cada parcela da solução, conferindo maior previsibilidade orçamentária e financeira à Administração.

**7.8.16.37.** Após a análise inicial da proposta, a Administração identificou a necessidade de aperfeiçoamento de determinados elementos, especialmente quanto às quantidades de alguns entregáveis, à consistência dos valores, à distribuição percentual entre os componentes e à adequação dos períodos de licenciamento e prestação dos serviços. Em razão disso, foi solicitada à Teltec a apresentação de versão revisada da proposta, bem como manifestação acerca da viabilidade do cronograma físico-financeiro e da existência de eventuais requisitos impeditivos à participação de solução Cisco. Essas solicitações resultaram nas manifestações posteriormente recebidas e na apresentação da Proposta TTEC3067/2026, versão 2.0.

**7.8.16.38.** Na mesma oportunidade, a Administração encaminhou a versão atualizada das especificações e solicitou à Cisco e à Teltec que verificassem a viabilidade de atendimento à distribuição física e financeira proposta, bem como que apresentassem manifestação expressa quanto à inexistência ou à identificação de requisitos e entregáveis que pudessem impedir a participação de solução Cisco no procedimento licitatório.

**7.8.16.39.** Posteriormente, em 30 de julho de 2026, a Teltec, com a Cisco em cópia na comunicação, apresentou resposta aos questionamentos formulados pela Administração. Na manifestação, informou ser possível atender à distribuição percentual proposta e ao cronograma físico-financeiro constante da versão atualizada das Especificações Técnicas e declarou que, até aquele momento, não havia identificado entregáveis ou requisitos impeditivos à sua participação com solução baseada em arquitetura Cisco.

**7.8.16.40.** Em comunicação subsequente, a Teltec propôs ajuste pontual na distribuição percentual dos Entregáveis 31, 32 e 33, sem alteração da participação conjunta desses componentes no valor global da solução. O percentual do Entregável 31 — Serviço de Suporte Avançado Mensal do Fabricante foi ajustado de 3,60% para 3,50%; o do Entregável 32 — Serviço de Suporte Mensal do Fornecedor, de 5,40% para 5,00%; e o do Entregável 33 — Serviço Especializado de Treinamento Avançado e Atualização Tecnológica, de 1,00% para 1,50%, permanecendo os três entregáveis, conjuntamente, com participação de 10,00% no valor global.

**7.8.16.41.** A redistribuição proposta foi considerada adequada pela Administração, uma vez que preserva o valor global, não altera os quantitativos ou o escopo dos entregáveis, mantém a participação econômica conjunta originalmente destinada aos três componentes e resulta de avaliação comercial apresentada pelo próprio fornecedor acerca da adequada distribuição dos custos entre suporte do fabricante, suporte do fornecedor e atualização tecnológica.

**7.8.16.42.** Em 30 de julho de 2026, foi recebida a Proposta Comercial TTEC3067/2026, versão 2.0, estruturada segundo os 33 entregáveis da contratação e contemplando as quantidades, períodos e distribuição econômica revisados. A proposta apresenta valor global de **R$ 19.977.500,00** (dezenove milhões, novecentos e setenta e sete mil e quinhentos reais) e prazo de validade de 60 dias. Para os Entregáveis 31, 32 e 33 foram apresentados, respectivamente, valores totais de R$ 699.212,50, R$ 998.875,00 e R$ 299.662,50.

**7.8.16.43.** A manifestação quanto à inexistência de impeditivos identificados, a confirmação da viabilidade do cronograma e da distribuição percentual e a apresentação da proposta revisada constituem evidências adicionais de exequibilidade técnica e comercial da arquitetura avaliada nas tratativas com a Cisco/Teltec. Tais elementos não representam habilitação prévia, homologação da solução, escolha antecipada de fornecedor ou garantia de atendimento no futuro certame, permanecendo obrigatória a comprovação integral, pela licitante, de todos os requisitos previstos no instrumento convocatório.

**7.8.17.** Síntese consolidada das tratativas e efeitos sobre o desenho da solução

**7.8.17.1.** Durante as tratativas realizadas, os fabricantes Dell e Lenovo, embora tenham indicado dispor de ofertas relacionadas ao conceito de Fábrica de IA, potencialmente aderentes ao modelo de GPU como serviço, optaram por não participar do processo de desenho da solução, deixando de contribuir tecnicamente para o amadurecimento da especificação a partir do documento de requisitos encaminhado pela Administração.

**7.8.17.2.** Esse fato não impede, por si só, a futura participação desses fabricantes no certame, mas evidencia que, no processo de preparação técnica da contratação, não houve colaboração material desses fornecedores para o refinamento do desenho da solução pretendida pelo Exército, diferentemente do que se verificou com outros fabricantes.

**7.8.17.3.** Os fabricantes Cisco e HPE deram prosseguimento às tratativas em relação às etapas iniciais de consulta, analisaram materiais encaminhados pela Administração e apresentaram documentos técnicos relacionados ao tema, os quais constam do Anexo II a este Estudo Técnico Preliminar. As tratativas com a Cisco permaneceram ativas durante o refinamento da solução, enquanto a HPE, por intermédio da Drive A, comunicou posteriormente sua decisão de não prosseguir na participação relacionada à contratação, em razão das customizações e integrações de terceiros que considerou necessárias ao atendimento das especificações e de seus impactos sobre a viabilidade técnica e a competitividade da solução HPE. Os documentos anteriormente recebidos de ambos os fabricantes permanecem como insumos relevantes para conhecimento e análise do mercado.

**7.8.17.4.** A desistência da HPE/Drive A constitui resultado relevante das tratativas de mercado. Além de registrar a posição comercial das empresas, a manifestação evidencia que, para a arquitetura analisada pela HPE, o atendimento integral ao conjunto de requisitos demandaria nível de customização e de integração com produtos de terceiros considerado excessivo pelo próprio fabricante e por sua integradora. Essa constatação reforça a importância atribuída pela Administração à contratação de solução completa, integrada, validada e suportada, de modo a evitar que parcela substancial do risco de integração seja transferida ao CITEx.

**7.8.17.5.** O fabricante Supermicro, por sua vez, embora seja reconhecido como relevante fornecedor de hardware para ambientes de alto desempenho e processamento acelerado, não apresentou solução completa homologada para atendimento ao modelo de GPU como serviço pretendido pela Administração. Sua abordagem está concentrada, predominantemente, na oferta de hardware, apoiando-se em rede de parceiros e fornecedores terceiros para composição de soluções aderentes ao escopo funcional e operacional da contratação.

**7.8.17.6.** Essa característica representa fator de risco relevante para a Administração, na medida em que a ausência de solução completa homologada e validada tende a transferir para o integrador ou, em última instância, para a própria contratante, parcela maior do risco de composição, interoperabilidade, sustentação e responsabilização técnica pela solução como um todo.

**7.8.17.7.** Em outras palavras, quando não há solução integralmente homologada, existe maior probabilidade de ocorrência de zonas cinzentas de responsabilidade entre fornecedores, dificuldades de suporte conjunto, dependência excessiva de integração específica e menor previsibilidade quanto ao comportamento da solução em operação real, especialmente em cenários de expansão futura.

**7.8.17.8.** Por esse motivo, a abordagem baseada apenas em oferta de hardware, ainda que tecnicamente relevante sob o ponto de vista do servidor ou do acelerador isoladamente considerado, não se mostra suficiente para fundamentar o desenho de uma solução de GPU como serviço para o Exército. Em razão disso, a Supermicro não será considerada como base para o desenho da solução no presente Estudo Técnico Preliminar.

**7.8.17.9.** Não obstante, o material técnico encaminhado pelo fabricante Supermicro, por representar contribuição informativa relevante sobre seu posicionamento de mercado e sua oferta de hardware, encontra-se igualmente juntado no Anexo II a este Estudo Técnico Preliminar.

**7.8.17.10.** Assim, as tratativas realizadas com os fabricantes cumpriram função relevante no processo de planejamento da contratação, permitindo à Administração avaliar não apenas a existência de hardware compatível, mas, principalmente, a maturidade das soluções ofertadas, a existência de validação entre componentes, o nível de homologação das arquiteturas e o grau de risco associado à adoção de cada abordagem.

**7.8.17.11.** Em conclusão, a presente seção evidencia que a Administração adotou providências concretas para conhecer o mercado, dialogar com fabricantes relevantes e reduzir o risco técnico da futura contratação. Também demonstra que a diretriz de considerar apenas soluções completas, integradas e homologadas não decorre de preferência indevida por fabricante, mas da necessidade objetiva de estruturar contratação compatível com a criticidade, a complexidade e os objetivos institucionais da solução de GPU como serviço pretendida pelo Exército Brasileiro.

**7.8.17.12.** Cumpre ainda esclarecer que a estratégia desenhada neste Estudo Técnico Preliminar não representa vedação definitiva à participação de qualquer fabricante de hardware nas futuras contratações relacionadas à expansão da capacidade de processamento acelerado do Exército.

**7.8.17.13.** Ao contrário, a própria arquitetura e a estratégia de contratação aqui adotadas foram concebidas de forma a permitir que, nas expansões previstas para a solução, sejam realizadas contratações com especificações abertas, padronização funcional e efetiva competitividade entre os fabricantes de hardware, desde que observados os requisitos técnicos, de compatibilidade e de interoperabilidade definidos pela Administração.

**7.8.17.14.** No estágio atual do projeto, entretanto, a diretriz de privilegiar solução completa, integrada e homologada está diretamente relacionada à mitigação de risco. Trata-se de opção de planejamento voltada a reduzir a probabilidade de insucesso na implantação inicial de uma solução de elevada complexidade tecnológica, situada em fronteira técnica de mercado e ainda dependente de forte coordenação entre múltiplos componentes de hardware, software e serviços especializados.

**7.8.17.15.** Nessa fase inicial, admite-se, portanto, que a busca pela maior mitigação de risco possa implicar redução pontual do grau de competitividade prática no processo de desenho e amadurecimento da solução, não por intenção de direcionamento, mas pela necessidade de assegurar maior probabilidade de êxito na implantação da primeira base operacional de GPU como serviço para o Exército.

**7.8.17.16.** A lógica adotada é a de que, uma vez implantada a solução inicial, absorvido o conhecimento necessário à sua operação, estabilizados os processos de sustentação e consolidadas as premissas de uso institucional, o nível de risco técnico e operacional das expansões subsequentes será significativamente menor.

**7.8.17.17.** Nesse cenário futuro, a Administração poderá atribuir prioridade ainda maior à competitividade, tornando viável a participação mais ampla dos diversos fabricantes de hardware, inclusive daqueles que não participaram deste processo de especificação ou que, neste momento, não disponham de solução completa homologada para servir de base ao desenho inicial da plataforma.

**7.8.17.18.** Assim, a diretriz ora adotada deve ser compreendida como medida de prudência técnica aplicável ao momento inicial da contratação, e não como restrição permanente de mercado. Seu objetivo é permitir que a Administração estabeleça, com menor risco, a primeira base da solução de GPU como serviço, preservando, para as etapas futuras de expansão, ambiente mais favorável à ampla competitividade entre fabricantes e à realização de contratações tecnicamente abertas e economicamente vantajosas.

**7.8.18.** Sobre o não direcionamento da contratação

**7.8.18.1.** A presente contratação será estruturada como licitação aberta e competitiva, sem restrição prévia de marca ou modelo para os demais componentes de hardware e software, ressalvadas exclusivamente a exigência de GPUs da marca NVIDIA, modelo B300, e a exigência do produto NVIDIA AI Enterprise para atendimento ao Entregável 6 Solução de Software Específica para Aceleração por GPU. As referidas indicações encontram fundamento no art. 41, inciso I, da Lei nº 14.133/2021 e nas justificativas de padronização, compatibilidade, integração, suporte oficial e mitigação de riscos apresentadas neste Estudo Técnico Preliminar.

**7.8.18.2.** A diretriz adotada pela Administração é permitir a participação de quaisquer fornecedores, revendas, integradores ou fabricantes capazes de ofertar solução completa, validada, suportada e aderente aos requisitos da contratação, desde que comprovem, de forma objetiva, o atendimento integral às especificações técnicas e às condições de execução estabelecidas no instrumento convocatório.

**7.8.18.3.** Registra-se, desde logo, que a presente contratação envolve tema de elevada complexidade técnica, caráter inovador e baixa maturidade relativa no mercado nacional, especialmente por tratar da implantação de uma plataforma corporativa de GPU como serviço, composta por servidores acelerados, rede de alta velocidade, orquestração, autosserviço, observabilidade, segurança operacional, integração com diretórios institucionais, suporte especializado, serviços de implantação, operação assistida e participação efetiva de fabricantes.

**7.8.18.4**. Em contratações dessa natureza, é possível que alguns requisitos técnicos sejam atendidos com maior facilidade por determinados ecossistemas tecnológicos ou por soluções que já possuam maior maturidade de integração entre hardware, software, serviços e suporte. Essa circunstância, por si só, não caracteriza direcionamento, desde que os requisitos estejam tecnicamente justificados, vinculados à real necessidade da Administração, descritos de forma objetiva e necessários à redução dos riscos de implantação e operação.

**7.8.18.5.** No caso concreto, a Administração adotou medidas prévias de mitigação do risco de direcionamento, incluindo a consulta a múltiplos fabricantes relevantes do mercado, o encaminhamento de documento de requisitos para análise, a realização de reuniões técnicas, a solicitação de contribuições sobre hardware, software e serviços, o registro das tratativas no Anexo II e a consolidação das justificativas técnicas que fundamentam cada grupo de requisitos da solução.

**7.8.18.6.** As tratativas demonstraram que os fabricantes responderam de formas distintas à oportunidade de contribuir para o amadurecimento da especificação. Alguns declinaram ou não prosseguiram materialmente nas interações; outros apresentaram contribuições pontuais ou incompletas; a HPE/Drive A, após participar das tratativas e apresentar proposta, comunicou posteriormente sua desistência, por considerar que o atendimento às especificações exigiria elevada quantidade de customizações e integrações de terceiros, comprometendo sua viabilidade técnica e competitividade; e um dos fabricantes consultados apresentou contribuição mais abrangente, com maior disponibilidade de especialistas, maior detalhamento de arquitetura e maior articulação com outros fornecedores de componentes complementares.

**7.8.18.7.** A maior contribuição técnica de determinado fabricante no processo de planejamento não decorreu de preferência prévia da Administração, mas da efetiva participação do mercado nas tratativas. A Administração oportunizou a participação de diferentes fabricantes e considerou os subsídios recebidos conforme sua pertinência técnica, sua aderência à demanda institucional e sua utilidade para redução de riscos da contratação.

**7.8.18.8.** Ainda que, na fase externa, algum interessado venha a alegar que determinados requisitos sejam mais aderentes a determinado fabricante ou ecossistema tecnológico, a análise da especificação deve considerar o conjunto do processo de planejamento, a motivação técnica dos requisitos, a complexidade da solução, a necessidade de validação entre componentes, a exigência de suporte oficial e a busca pela implantação de solução turn-key efetivamente funcional.

**7.8.18.9.** A Administração não busca reproduzir especificação proprietária, restringir a disputa ou favorecer fornecedor específico, mas contratar solução capaz de atender, com menor risco técnico possível, à necessidade institucional identificada. A eventual redução do universo de soluções aptas a atender requisitos essenciais não se confunde com restrição indevida, quando tais requisitos são proporcionais, motivados e necessários ao resultado pretendido.

**7.8.18.10.** A vedação contida no art. 2º, inciso I, do Decreto nº 7.174, de 12 de maio de 2010, dirige-se às especificações que direcionem ou favoreçam a contratação de fornecedor específico. O mesmo dispositivo também veda especificações que não representem a real demanda de desempenho do órgão ou entidade ou que não explicitem métodos objetivos de mensuração do desempenho dos bens e serviços de informática e automação. A presente contratação foi estruturada justamente para observar esses parâmetros, com especificações vinculadas à real demanda da Administração e critérios técnicos passíveis de comprovação objetiva.

**7.8.18.11.** Assim, eventual aderência mais direta de determinado fabricante a algum requisito técnico deverá ser analisada à luz da necessidade pública que originou o requisito, e não de forma isolada. O parâmetro de avaliação deve ser se a exigência é necessária à entrega de uma plataforma integrada, validada, segura, suportada, observável, expansível e operacionalmente sustentável, e não simplesmente se todos os fabricantes consultados possuem solução idêntica ou maturidade equivalente.

**7.8.18.12.** A Administração também considerou que, mesmo quando algum componente de determinada marca venha a se mostrar tecnicamente necessário por razões de padronização, compatibilidade, integração ou mitigação de risco, tal circunstância não conduz automaticamente à contratação de fornecedor único. No mercado de TIC, equipamentos, licenças, serviços e suporte de determinados fabricantes podem ser ofertados por múltiplas revendas, integradores e parceiros autorizados, preservando disputa entre fornecedores e possibilidade de ampla participação na licitação.

**7.8.18.13.** Essa consideração é especialmente relevante à luz do Decreto nº 7.174/2010, pois a vedação legal se dirige ao favorecimento de fornecedor específico. Quando há pluralidade de potenciais fornecedores aptos a comercializar, integrar, licenciar, suportar ou prestar serviços relacionados a determinado ecossistema tecnológico, a eventual adoção de requisitos compatíveis com esse ecossistema deve ser examinada em conjunto com a justificativa técnica, a padronização existente, a compatibilidade requerida e a real possibilidade de competição no fornecimento.

**7.8.18.14.** Sem prejuízo disso, a Administração reconhece que a existência de múltiplas revendas ou integradores não autoriza, por si só, especificações artificiais ou desnecessárias. Por essa razão, os requisitos da solução foram vinculados a necessidades técnicas concretas, tais como integração entre hardware e software, suporte oficial, licenciamento empresarial, observabilidade, operação assistida, segurança, autosserviço, capacidade de expansão, compatibilidade com ambiente de Data Center e redução de riscos de implantação.

**7.8.18.15.** A especificação também busca evitar situações comumente associadas a questionamentos por direcionamento, cerceamento da competitividade ou quebra de isonomia, tais como indicação de marca sem justificativa, reprodução de características meramente comerciais de determinado

fabricante, exigência de documentos sem pertinência técnica, requisitos de desempenho sem método objetivo de aferição, certificações desproporcionais, exigências de qualificação incompatíveis com o objeto, restrições injustificadas à participação de revendas ou integradores e ausência de critérios claros para comprovação de atendimento.

**7.8.18.16.** Para esse fim, os requisitos deverão ser avaliados preferencialmente por sua finalidade técnica e pelo resultado a ser entregue, admitindo-se soluções equivalentes sempre que comprovem objetivamente o mesmo nível de desempenho, integração, validação, suporte, segurança, governança, observabilidade e sustentabilidade operacional exigido pela Administração, ressalvadas as exigências expressas relativas às GPUs NVIDIA B300 e ao NVIDIA AI Enterprise.

**7.8.18.17.** Quando forem exigidas arquiteturas de referência, declarações de fabricante, matrizes de compatibilidade, documentação pública, suporte oficial, licenciamento empresarial ou participação efetiva do fabricante nos serviços críticos, tais exigências deverão ser compreendidas como instrumentos de mitigação de risco técnico e não como barreiras artificiais de competição. A finalidade é evitar que a Administração contrate um conjunto de componentes que, embora individualmente disponíveis no mercado, não possuam validação conjunta, suporte coordenado ou responsabilidade clara pela integração.

**7.8.18.18.** A exigência de solução turn-key também deve ser compreendida nesse contexto. A Administração pretende contratar plataforma funcional, integrada e suportada, e não apenas adquirir servidores, softwares e serviços de forma desconexa. A experiência demonstrada nas tratativas reforçou que, em solução de GPU como serviço, a ausência de validação entre os componentes pode transferir indevidamente para a Administração o risco de integração, operação e sustentação.

**7.8.18.19.** A fase externa da licitação poderá trazer pedidos de esclarecimento ou impugnações relacionados ao grau de abertura da especificação. Tais manifestações deverão ser examinadas tecnicamente, considerando-se a documentação de planejamento, as tratativas de mercado, a motivação dos requisitos e a possibilidade de atendimento por soluções equivalentes ou por diferentes fornecedores aptos a entregar a solução requerida.

**7.8.18.20.** Caso algum questionamento demonstre, de forma objetiva, que determinado requisito pode ser redigido de maneira mais funcional, sem prejuízo do atendimento da necessidade, a Administração poderá avaliar eventual ajuste. Por outro lado, quando o requisito estiver diretamente relacionado à mitigação de risco, à compatibilidade, à padronização, à segurança, ao suporte oficial ou à viabilidade da solução turn-key, sua manutenção estará tecnicamente fundamentada no próprio planejamento da contratação.

**7.8.18.21.** A Administração deverá, portanto, preservar o equilíbrio entre competitividade e segurança técnica. A ampliação da disputa é objetivo relevante, mas não pode levar à supressão de requisitos essenciais à entrega do resultado pretendido, especialmente quando a contratação envolve infraestrutura crítica, processamento de IA, dados sensíveis, ambientes de Data Center, múltiplos fabricantes, integração complexa e necessidade de sustentação de longo prazo.

**7.8.18.22.** A eventual existência de requisitos de alta exigência técnica deve ser compreendida como consequência da natureza do objeto e do resultado pretendido, e não como preferência subjetiva por determinado fabricante. A vantajosidade da contratação deve considerar custo total de propriedade, probabilidade de sucesso da implantação, suporte, maturidade da solução, risco operacional, segurança, expansão futura e capacidade de entrega de serviço efetivo à Administração.

**7.8.18.23.** Dessa forma, o planejamento ora realizado documenta que a Administração adotou cautelas compatíveis com a legislação aplicável, com os princípios da isonomia, competitividade, julgamento objetivo, motivação, eficiência, planejamento, economicidade e seleção da proposta apta a gerar o resultado de contratação mais vantajoso para a Administração, nos termos dos arts. 5º e 11 da Lei nº 14.133/2021.

**7.8.18.24.** Em síntese, a presente contratação não tem por finalidade favorecer fabricante, marca, modelo, revenda ou integrador específico. O desenho da solução foi construído a partir da necessidade pública, das tratativas de mercado, da análise técnica dos riscos e da busca por solução completa, integrada, validada e suportada, capaz de implantar a primeira base institucional de GPU como serviço do Exército Brasileiro com segurança, governança, previsibilidade de sustentação e possibilidade de expansão futura.

**7.8.18.25.** Como medida adicional de mitigação do risco de direcionamento e de identificação de eventual restrição indevida à competitividade, a Administração facultou aos fabricantes Cisco e HPE, com a participação de seus respectivos fornecedores e integradores Teltec e Drive A, a oportunidade de analisar a versão mais recente das especificações técnicas e apontar requisitos, entregáveis, prazos, condições de habilitação, exigências de certificação ou formas de participação do fabricante que pudessem constituir impedimento à apresentação de proposta.

**7.8.18.26.** Essa oportunidade foi concedida após sucessivas interações técnicas com os fabricantes, o encaminhamento de diferentes versões dos documentos de requisitos e das especificações e a consideração das contribuições anteriormente recebidas. Buscou-se, assim, permitir que os participantes com maior envolvimento no processo de amadurecimento da solução realizassem uma última avaliação da exequibilidade técnica das condições estabelecidas. Posteriormente, a HPE/Drive A apresentou manifestação conclusiva de desistência, conforme registrado na subseção 7.8.14.

**7.8.18.27.** Até 28 de julho de 2026, ainda não havia sido recebida manifestação conclusiva dos participantes consultados quanto a todos os aspectos da versão final das especificações. Em 28 de julho de 2026, a HPE/Drive A comunicou formalmente a decisão de não prosseguir na participação relacionada à contratação, pelos motivos registrados na subseção 7.8.14. Posteriormente, em 30 de julho de 2026, a Teltec, com a Cisco em cópia, informou ser possível atender à distribuição percentual e ao cronograma físico-financeiro propostos e declarou que, até aquele momento, não havia identificado entregáveis ou requisitos impeditivos à participação com solução Cisco.

**7.8.18.28.** A manifestação da HPE/Drive A não invalida, por si só, as conclusões técnicas deste Estudo Técnico Preliminar, permanecendo, contudo, como elemento de atenção para a avaliação da competitividade. Em sentido complementar, a resposta posteriormente encaminhada pela Teltec, com a Cisco em cópia, constitui evidência de que a arquitetura por elas analisada considera exequíveis o conjunto de especificações, a distribuição físico-financeira e os prazos propostos. Essa manifestação reduz o risco de inexequibilidade técnica do desenho da contratação.

**7.8.18.29.** Caso sejam apresentados posteriormente, por fabricantes, fornecedores, integradores ou demais agentes do mercado, apontamentos tecnicamente fundamentados acerca das especificações, a Administração poderá analisá-los e promover os ajustes cabíveis, desde que as alterações preservem a necessidade pública, o desempenho mínimo requerido, a segurança, a compatibilidade, o suporte oficial, a responsabilidade integrada pela solução, a isonomia entre os participantes e a competitividade do certame.

**7.8.18.30.** Dessa forma, a oportunidade concedida aos fabricantes reforça que as especificações não foram tratadas como documento imutável ou construído para favorecer solução previamente determinada, mas como instrumento passível de aperfeiçoamento até a fase externa, mediante motivação técnica e observância dos princípios da isonomia, competitividade, planejamento, eficiência e seleção da proposta mais vantajosa. O texto atual do ETP permanece, portanto, suficiente para fundamentar o prosseguimento da contratação.

**7.8.18.31.** A desistência da HPE/Drive A permanece elemento de atenção para a fase final do planejamento, especialmente quanto ao risco de concentração do universo de soluções aptas a atender ao objeto. Em sentido complementar, a manifestação da Teltec, com a Cisco em cópia, confirmou a viabilidade da distribuição físico-financeira e informou a inexistência, até aquele momento, de requisitos impeditivos à participação com solução Cisco. Tal manifestação constitui evidência relevante da exequibilidade do objeto para uma das arquiteturas analisadas.

**7.8.18.32.** A circunstância de um fabricante consultado ter desistido de prosseguir não determina, por si só, a retirada ou flexibilização dos requisitos por ele considerados de difícil atendimento. Cada requisito deverá ser mantido, alterado ou suprimido conforme sua necessidade objetiva para o resultado pretendido, sua proporcionalidade e as evidências técnicas constantes dos autos, preservando-se simultaneamente a segurança técnica da contratação e a maior competitividade possível.

### 7.9. Arquitetura de referência

#### 7.9.1. Considerações gerais sobre a composição da solução

**7.9.1.1.** A estimativa da quantidade de bens e serviços necessários à presente contratação deve ser compreendida a partir da arquitetura de referência da solução de GPU como serviço e da estratégia institucional de implantação progressiva adotada para o projeto.

**7.9.1.2.** Não se trata, portanto, de estimativa baseada apenas em quantitativo isolado de servidores ou aceleradores, mas da composição coerente de todos os elementos físicos, lógicos e operacionais indispensáveis para transformar capacidade computacional acelerada em serviço institucional consumível, escalável, seguro e governado.

**7.9.1.3.** Nesse contexto, a contratação deverá ser entendida como aquisição de uma solução única e integrada, composta por componentes de hardware, software, armazenamento, conectividade e serviços especializados, organizados de forma a permitir a implantação inicial da plataforma e sua posterior expansão modular.

**7.9.1.4.** A quantidade de bens e serviços, assim, deve refletir não apenas a capacidade computacional pretendida, mas também as necessidades de orquestração, observabilidade, autosserviço, proteção de dados, suporte operacional e evolução da solução ao longo do tempo.

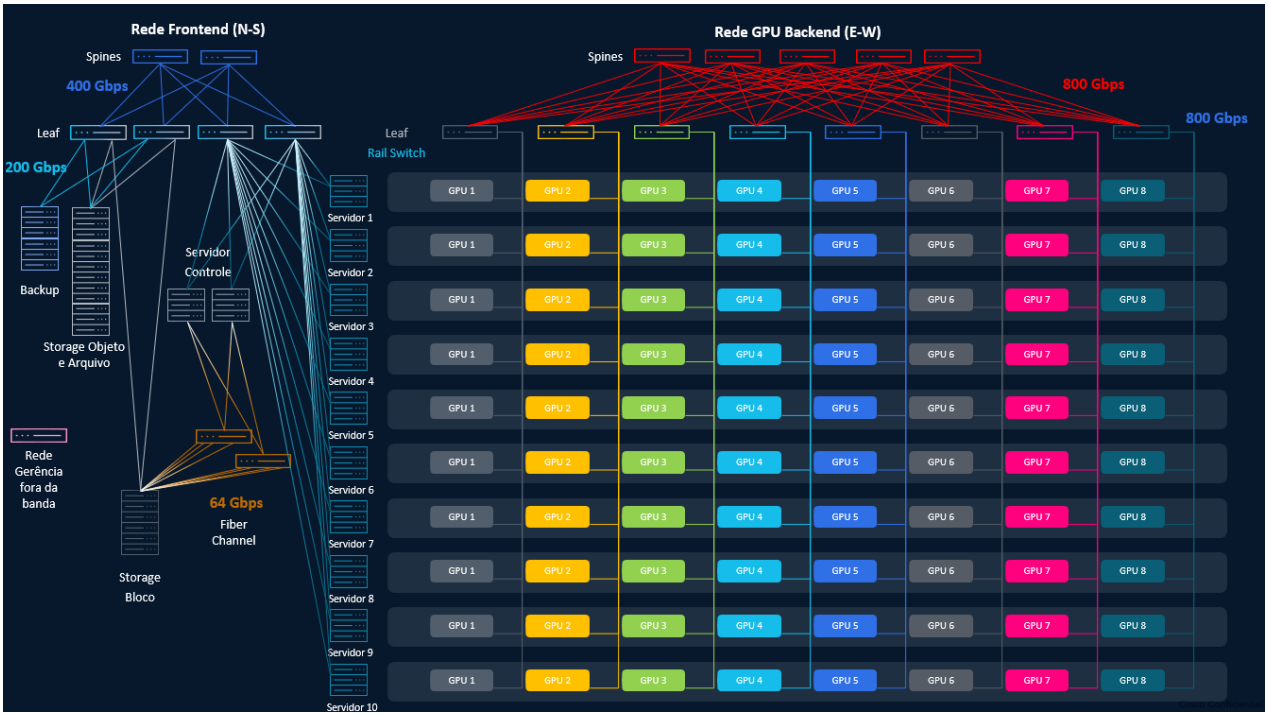



**Figura 3:** Diagrama de equipamentos da Fase 1.

**7.9.1.5.** A **Figura 3** representa, em alto nível, os componentes de hardware da Fase 1, devendo ser utilizada como referência para compreensão da arquitetura física que fundamenta a presente estimativa quantitativa.

### 7.9.2. Capacidade computacional de referência

**7.9.2.1.** A arquitetura física de referência adotada para a Fase 1 da solução considera um ambiente-alvo composto por **10 (dez) servidores de processamento com GPU**, cada qual com **8 (oito) GPUs**, totalizando **80 (oitenta) GPUs** ao término dessa fase.

**7.9.2.2.** Essa capacidade representa a referência de expansão do ambiente, embora a implantação inicial da solução ocorra com quantitativo reduzido, em consonância com a estratégia de crescimento sob demanda já definida neste Estudo Técnico Preliminar.

**7.9.2.3.** A estrutura de 10 (dez) servidores com 08 (oito) GPUs por nó deverá ser interpretada como a base da capacidade computacional acelerada da solução, destinada a suportar workloads de inteligência artificial com diferentes perfis de consumo, incluindo inferência, treinamento, ajuste fino, desenvolvimento interativo, ambientes de experimentação, serviços de inferência em produção e cargas distribuídas com uso simultâneo de múltiplas GPUs.

**7.9.2.4.** O quantitativo de servidores e aceleradores, portanto, não se destina apenas a prover poder computacional bruto, mas a constituir a base física de uma plataforma corporativa de GPU como serviço.

**7.9.2.5.** A solução não se resume aos servidores com GPU. A arquitetura de referência evidencia que o ambiente deverá ser composto, de forma articulada, por servidores de processamento acelerado, servidores de gerenciamento, múltiplas redes especializadas, storages com perfis distintos, appliance de backup, softwares estruturantes e serviços especializados de implantação, capacitação, operação assistida e sustentação.

### 7.9.3. Componentes de hardware de processamento e gerenciamento

**7.9.3.1.** O primeiro componente estruturante da solução corresponde ao **servidor de processamento com GPU**, que deverá constituir o núcleo da capacidade computacional acelerada da plataforma.

**7.9.3.2.** Esse item deverá ser dimensionado para suportar cargas de inferência, treinamento, ajuste fino, desenvolvimento interativo, execução de modelos de linguagem de maior porte e workloads que demandem uso simultâneo de múltiplas GPUs.

**7.9.3.3.** Deverá, ainda, ser compatível com a arquitetura lógica da solução, com os mecanismos de orquestração, com a infraestrutura de rede de alto desempenho e com a estratégia de expansão modular da plataforma.

**7.9.3.4.** Em razão da estratégia de centralização da capacidade computacional, os servidores de processamento deverão ser de alta densidade, compatíveis com a arquitetura de 08 (oit0) GPUs por nó e com os requisitos de conectividade, energia, refrigeração e integração previstos para o ambiente.

**7.9.3.5.** Essa diretriz busca maximizar a densidade de processamento por servidor, simplificar a topologia de interconexão, favorecer workloads intra-nó e preservar coerência arquitetural para as fases subsequentes de expansão.

**7.9.3.6.** O segundo componente corresponde ao **servidor para hospedagem da camada de software da solução ou servidor de controle**, que deverá disponibilizar capacidade computacional **sem GPU**, dedicada à execução dos softwares responsáveis pela operação da plataforma.

**7.9.3.7.** Nesses servidores deverão ser hospedados os componentes de orquestração, automação, portal, APIs, gestão da solução, observabilidade, segurança operacional, serviços auxiliares e demais elementos lógicos necessários ao funcionamento do ambiente de GPU como serviço.

**7.9.3.8.** A separação entre servidores de processamento com GPU e servidores de controle se justifica por razões de governança, estabilidade e eficiência operacional.

**7.9.3.9.** Busca-se, com isso, evitar que workloads acelerados disputem recursos com os serviços de plataforma e, ao mesmo tempo, assegurar que a camada lógica da solução possua capacidade computacional própria, estável e administrável para sustentar o funcionamento institucional do ambiente.

### 7.9.4. Redes especializadas da solução

**7.9.4.1.** Do ponto de vista da conectividade, a arquitetura foi organizada em **quatro redes específicas**, cada qual com finalidade técnica própria: **rede de backend ou leste-oeste (East-West)**, **rede frontend ou norte-sul (North-South)**, **rede de gerência** e **rede SAN Fibre Channel**.

**7.9.4.2.** Essa segregação se justifica pela necessidade de separar fluxos com características distintas de latência, largura de banda, criticidade, segurança e padrão de consumo, evitando contenção indevida entre tráfegos de naturezas diferentes e assegurando melhor previsibilidade de desempenho da solução.

#### 7.9.4.3. Rede de backend ou leste-oeste

**7.9.4.3.1.** A rede de backend ou leste-oeste deverá ser a malha de comunicação de mais alto desempenho da arquitetura, destinada ao tráfego intensivo entre os servidores com GPU.

**7.9.4.3.2.** Ela deverá suportar comunicação entre nós em cenários de treinamento distribuído, troca de dados entre aceleradores, sincronização de parâmetros, inferência paralela e demais workloads que demandem comunicação coordenada entre múltiplos servidores e múltiplas GPUs.

**7.9.4.3.3.** Por essa razão, trata-se da rede mais crítica da arquitetura sob o ponto de vista de desempenho computacional.

**7.9.4.3.4.** Essa rede deverá ser compatível com a tecnologia **RDMA**, sigla para **Remote Direct Memory Access**, ou acesso remoto direto à memória.

**7.9.4.3.5.** Em termos práticos, o RDMA permite que um equipamento acesse diretamente áreas de memória de outro equipamento pela rede, com mínima intervenção da CPU e com baixa sobrecarga de software.

**7.9.4.3.6.** Em ambientes de alto desempenho, essa característica é particularmente importante porque reduz latência, melhora a eficiência da comunicação entre nós e favorece workloads distribuídos que dependem de troca intensiva e frequente de dados entre servidores e aceleradores.

**7.9.4.3.7.** No contexto da solução de GPU como serviço, o uso de RDMA sobre Ethernet na rede de backend é requisito relevante para viabilizar comunicação eficiente entre os nós de processamento, especialmente em cenários de treinamento distribuído, execução multi-GPU entre servidores distintos, sincronização de parâmetros e demais fluxos de dados que exijam comportamento próximo ao de uma malha de alto desempenho.

**7.9.4.3.8.** Para essa rede, a solução deverá adotar topologia **spine-leaf** com uso do conceito de **optimized rail leaf**, de modo a estruturar a comunicação entre GPUs e servidores de forma compatível com ambientes de alto desempenho baseados em **RDMA sobre Ethernet**.

**7.9.4.3.9.** Nessa malha, a conectividade entre os servidores e os switches leaf deverá ser de **800 Gbps** , e os equipamentos spine também deverão operar em **800 Gbps**, em arranjo **não blocante**, de forma a evitar oversubscription e preservar o desempenho esperado para workloads multi-GPU e distribuídos.

**7.9.4.3.10.** O conceito de **optimized rail leaf** corresponde à organização da rede em “trilhos” lógicos, nos quais interfaces equivalentes dos diversos servidores são conectadas a switches leaf específicos, preservando alinhamento e previsibilidade do tráfego entre os nós.

**7.9.4.3.11.** Em termos práticos, isso significa que cada conjunto de interfaces associado a determinada posição lógica de GPU ou de adaptador de rede é distribuído de forma padronizada ao longo da malha, reduzindo assimetria de caminhos, favorecendo balanceamento do tráfego e melhorando a eficiência das comunicações coletivas entre servidores.

**7.9.4.3.12.** A exigência de rede não blocante na malha de backend decorre da necessidade de assegurar que a capacidade agregada de comunicação entre os nós de processamento não seja reduzida por gargalos na camada de interconexão.

**7.9.4.4. Rede frontend ou norte-sul**

**7.9.4.4.1.** A rede frontend ou norte-sul deverá ser a malha responsável pela comunicação entre os servidores da solução, os usuários, as APIs, os portais, os serviços de acesso, os storages de arquivos e objetos, os appliances de backup e os demais elementos de infraestrutura que não demandam o comportamento extremo de baixa latência da rede leste-oeste.

**7.9.4.4.2.** Essa rede deverá suportar o tráfego de acesso aos serviços da plataforma, o acesso aos dados, a comunicação com os softwares de operação e o fluxo entre computação e armazenamento.

**7.9.4.4.3.** A topologia da rede frontend também deverá ser estruturada em **spine-leaf**, com velocidade de conectividade de até **200 Gbps** entre os equipamentos terminais e a malha, e enlaces de **400 Gbps** entre switches spine e leaf.

**7.9.4.4.4.** Diferentemente da rede de backend, essa rede não precisa ser do tipo não blocante, uma vez que seu perfil de uso admite maior elasticidade e não está diretamente associado à troca intensiva de dados entre GPUs distribuídas.

**7.9.4.4.5.** Ainda assim, deverá ser dimensionada com capacidade suficiente para suportar os fluxos previstos de dados, acesso aos serviços e comunicação com os sistemas de armazenamento.

**7.9.4.4.6.** A rede frontend também deverá **suportar RDMA**, uma vez que será por meio dela que os servidores com GPU acessarão, de forma eficiente, os storages de arquivos e objetos utilizados pelas soluções de inteligência artificial.

**7.9.4.4.7.** Em outras palavras, embora a malha leste-oeste seja a mais crítica para a comunicação entre nós de processamento, a malha norte-sul também deverá oferecer suporte a acesso direto de memória sobre a rede para reduzir a sobrecarga de CPU e melhorar o desempenho do tráfego de dados entre os servidores com GPU e os storages destinados a datasets, modelos, checkpoints, artefatos e demais conteúdos consumidos pelos workloads de IA.

**7.9.4.5. Rede de gerência**

**7.9.4.5.1.** A rede de gerência deverá ser logicamente segregada das demais e destinada exclusivamente às funções de administração, configuração, monitoramento, suporte técnico, atualização de firmware, acesso aos controladores de hardware, gestão fora da banda e demais atividades de operação da infraestrutura.

**7.9.4.5.2.** Sua existência é indispensável para assegurar governança, segurança administrativa e isolamento entre tráfego operacional e tráfego de produção da solução.

**7.9.4.6. Rede SAN Fibre Channel**

**7.9.4.6.1.** A rede SAN Fibre Channel deverá ser empregada para atendimento às demandas de armazenamento em bloco, especialmente voltadas a máquinas virtuais, volumes persistentes, demandas específicas de contêineres e demais componentes da solução que requeiram acesso a storage em bloco com maior previsibilidade de desempenho.

**7.9.4.6.2.** Para esse fim, a conectividade adequada deverá ser em **64 Gbps Fibre Channel**, de modo a garantir maior desempenho para workloads de infraestrutura e armazenamento crítico.

**7.9.5. Armazenamento da solução**

**7.9.5.1.** A solução deverá contemplar **duas abordagens distintas de armazenamento**, compatíveis com a natureza heterogênea das cargas de trabalho previstas.

#### 7.9.5.2. Storage para arquivos e objetos

**7.9.5.2.1.** A primeira delas corresponde ao **storage para arquivos e objetos**, destinado ao consumo pelas soluções de inteligência artificial, incluindo datasets, artefatos, checkpoints, modelos, imagens, documentos, logs e demais dados não estruturados ou semiestruturados.

**7.9.5.2.2.** Tal solução deverá ser compatível com acesso em arquivo e objeto, bem como possuir compatibilidade com **RDMA**, de forma a elevar o desempenho do acesso aos dados pelos servidores de processamento quando tecnicamente aplicável.

**7.9.5.2.3.** A exigência de compatibilidade com arquivo, objeto e RDMA para esse storage decorre do fato de que workloads de IA, especialmente em ambientes Linux, Kubernetes e processamento acelerado, dependem frequentemente de acesso compartilhado a grandes volumes de dados não estruturados, exigindo alta largura de banda, baixa latência relativa, concorrência entre múltiplos consumidores e integração eficiente com a camada computacional.

#### 7.9.5.3. Storage para bloco

**7.9.5.3.1.** A segunda abordagem corresponde ao **storage para bloco**, destinado ao armazenamento de máquinas virtuais, serviços de gerenciamento da solução, volumes persistentes de contêineres e demais componentes que demandem armazenamento em bloco.

**7.9.5.3.2.** Para essa finalidade, a arquitetura deverá prever conectividade por meio de **rede SAN Fibre Channel** e, quando aplicável, também por **iSCSI**, de modo a atender tanto aos servidores de gerenciamento quanto aos servidores com GPU, conforme o perfil de consumo de cada componente da plataforma.

**7.9.5.4.** Além da capacidade bruta, a solução de armazenamento deverá ser dimensionada em coerência com o modelo de crescimento sob demanda da plataforma, considerando o volume de dados de IA, o crescimento progressivo dos datasets, a retenção de artefatos, a necessidade de persistência dos serviços de plataforma e a futura expansão da capacidade computacional instalada.

**7.9.5.5.** A estimativa quantitativa de storage, portanto, deverá estar vinculada à quantidade de GPUs efetivamente implantadas em cada etapa, observadas as necessidades técnicas da arquitetura e os requisitos de desempenho da solução.

### 7.9.6. Proteção de dados e backup

**7.9.6.1.** A arquitetura deverá prever, ainda, **appliance de backup** com características técnicas compatíveis com a retenção, proteção e recuperação de dados em **arquivo, objeto e bloco**, de forma aderente aos perfis de armazenamento utilizados pela solução.

**7.9.6.2.** Tal componente é essencial para garantir proteção dos dados operacionais, dos artefatos de IA, dos ambientes de gerenciamento, dos volumes persistentes e dos demais elementos críticos ao funcionamento da plataforma.

**7.9.6.3.** A solução de backup deverá ser compatível com o porte da arquitetura, com os requisitos de retenção institucional, com os volumes de dados esperados e com a diversidade de formatos armazenados.

**7.9.6.4.** Em ambiente dessa natureza, o backup não deve ser compreendido apenas como cópia de segurança de arquivos isolados, mas como mecanismo estruturado de proteção da plataforma como

um todo, inclusive para fins de continuidade operacional, recuperação de falhas e preservação de informações institucionais.

**7.9.7. Componentes lógicos estruturantes da solução**

**7.9.7.1.** Além dos componentes físicos representados na arquitetura de referência, a solução deverá contemplar componentes lógicos indispensáveis para transformar a infraestrutura acelerada em uma plataforma corporativa de GPU como serviço.

**7.9.7.2.** Nesse contexto, a estimativa de bens e serviços não se limita a servidores, redes e storages, devendo abranger também os softwares necessários à operação, ao consumo, à governança e à sustentação do ambiente.

**7.9.7.3. Software de virtualização, contêineres e orquestração**

**7.9.7.3.1.** O primeiro desses componentes corresponde à **solução de software de virtualização, contêineres e orquestração**, que deverá prover a base operacional para hospedagem de máquinas virtuais, execução de workloads em contêineres, gerenciamento de cluster, automação de implantação, escalabilidade e operação padronizada da plataforma.

**7.9.7.3.2.** Trata-se da camada que viabiliza a administração centralizada dos recursos computacionais e a distribuição controlada dos workloads sobre a infraestrutura disponível.

**7.9.7.3.3.** Esse componente é essencial porque a solução de GPU como serviço precisa suportar diferentes formas de consumo da capacidade computacional, incluindo ambientes virtualizados, workloads cloud-native, serviços de inferência, jobs de treinamento, ambientes interativos e mecanismos de automação.

**7.9.7.3.4.** Assim, a plataforma de virtualização e contêineres deverá oferecer recursos corporativos de disponibilidade, orquestração, segurança, governança e expansão horizontal, constituindo a base de execução lógica do ambiente.

**7.9.7.4. Software específico para aceleração por GPU**

**7.9.7.4.1.** O segundo componente corresponde ao software específico para aceleração por GPU, que deverá ser atendido pelo produto NVIDIA AI Enterprise, em razão da indicação excepcional formalmente justificada no item 7.4.2.

**7.9.7.4.2.** Essa camada deverá prover os elementos necessários para habilitar, operar e sustentar o uso corporativo das GPUs, incluindo drivers, runtimes, bibliotecas, componentes de aceleração, ferramentas de integração com a plataforma de orquestração e demais recursos necessários à execução estável de workloads de inteligência artificial.

**7.9.7.4.3.** A presença desse componente se justifica porque a simples existência do hardware acelerado não é suficiente para disponibilizar uma solução de GPU como serviço.

**7.9.7.4.4.** É indispensável que a infraestrutura conte com software especializado capaz de expor corretamente os recursos de GPU, suportar frameworks e ambientes de IA, integrar-se à camada de orquestração e oferecer base tecnológica compatível com desenvolvimento, inferência, treinamento, particionamento de aceleradores e observabilidade específica de recursos acelerados.

**7.9.7.5. Software para interface de serviços e autosserviço**

**7.9.7.5.1.** O terceiro componente corresponde ao **software para interface de serviços e autosserviço**, responsável por organizar a forma como os usuários e sistemas consumirão a capacidade disponibilizada pela plataforma.

**7.9.7.5.2.** Esse elemento deverá oferecer mecanismos de autosserviço, portal web, catálogo de serviços, integração por APIs, controle de acesso, gestão por perfis, acompanhamento de consumo, abstração da complexidade da infraestrutura e facilidades para solicitação e uso de ambientes com GPU.

**7.9.7.5.3.** Tal componente é necessário porque o objetivo institucional da contratação não é apenas manter uma infraestrutura acelerada em funcionamento, mas transformá-la em serviço efetivamente consumível pelas diversas áreas do Exército.

**7.9.7.5.4.** Em outras palavras, a plataforma precisa oferecer uma camada de interação e governança que permita aos usuários autorizados acessar a capacidade computacional de maneira organizada, padronizada e controlada, sem dependência excessiva de intervenção manual da equipe de infraestrutura para cada demanda operacional.

**7.9.7.6. Software de observabilidade e segurança operacional**

**7.9.7.6.1.** O quarto componente corresponde ao **software de observabilidade e segurança operacional**, que deverá prover visibilidade operacional sobre toda a solução, abrangendo infraestrutura, workloads, consumo de recursos, eventos, falhas, desempenho, trilhas de auditoria e comportamento dos aceleradores.

**7.9.7.6.2.** Essa camada é indispensável para permitir monitoramento contínuo da plataforma, diagnóstico de problemas, análise de capacidade, geração de indicadores gerenciais e suporte à tomada de decisão quanto à evolução do ambiente.

**7.9.7.6.3.** Em uma solução de GPU como serviço, a observabilidade não deve ser tratada como elemento acessório, mas como requisito estruturante.

**7.9.7.6.4.** A complexidade do ambiente, a diversidade de componentes, a necessidade de governança, a criticidade da operação e o caráter progressivo da expansão tornam indispensável a existência de ferramenta capaz de consolidar métricas, logs, alertas e telemetria da solução como um todo, incluindo os recursos específicos de processamento acelerado.

**7.9.8. Serviços especializados associados à solução**

**7.9.8.1.** A estimativa da contratação deverá considerar, ainda, os **serviços especializados** indispensáveis à implantação e à sustentação da solução.

**7.9.8.2.** Tais serviços não constituem elemento acessório, mas parte integrante da estratégia de disponibilização de GPU como serviço, uma vez que o sucesso da contratação depende não apenas do fornecimento dos bens e das licenças, mas também da correta implantação, integração, estabilização e transferência de conhecimento para a Administração.

**7.9.8.3.** Nesse contexto, deverão ser considerados, no mínimo, os seguintes grupos de serviços: **serviços especializados de implantação e configuração**, **serviços de capacitação e transferência de conhecimento**, **serviços de operação assistida**, **serviços de suporte e sustentação técnica** e **treinamento especializado de atualização tecnológica**.

**7.9.8.4.** Esses itens compõem o conjunto de serviços previsto para a contratação e deverão ser dimensionados em coerência com a quantidade de GPUs efetivamente implantadas em cada etapa, com a complexidade da arquitetura e com a necessidade de amadurecimento progressivo da operação institucional da plataforma.

**7.9.8.5.** Os serviços especializados de implantação e configuração deverão assegurar a instalação coordenada da solução, a integração entre hardware, softwares, redes e storages, a parametrização dos componentes, a validação funcional do ambiente e sua entrada em operação inicial.

**7.9.8.6.** Os serviços de capacitação e transferência de conhecimento deverão assegurar que as equipes da Administração recebam treinamento compatível com a operação da solução, com os processos de administração da plataforma, com o uso da camada de autosserviço, com os mecanismos de observabilidade e com as melhores práticas de sustentação do ambiente.

**7.9.8.7.** Os serviços de operação assistida deverão acompanhar a entrada em funcionamento da plataforma, reduzir riscos de adoção, apoiar a estabilização do ambiente, analisar consumo de GPU, crescimento da demanda, gargalos de processamento, comportamento das filas de workloads e oportunidades de ajuste fino da solução, gerando insumos para decisões sobre expansão e reconfiguração.

**7.9.8.8.** Os serviços de suporte e sustentação técnica deverão abranger atendimento corretivo, manutenção preventiva, atualização controlada, apoio à administração do ambiente, correção de falhas, esclarecimento de dúvidas técnicas e suporte à continuidade operacional, assegurando que a solução permaneça em condições adequadas de funcionamento ao longo do período contratual.

**7.9.8.9.** O treinamento especializado de atualização tecnológica deverá ser previsto como mecanismo de atualização contínua das equipes da Administração, abrangendo evolução de arquiteturas de GPU, novas funcionalidades dos softwares adquiridos, melhores práticas de operação de ambientes acelerados, tendências tecnológicas em inteligência artificial, otimização de workloads e evolução dos mecanismos de inferência e treinamento.

**7.9.8.10.** Seu objetivo é reduzir risco de obsolescência de conhecimento e fortalecer a governança da solução ao longo do tempo.

**7.9.9. Consolidação da estratégia quantitativa da contratação**

**7.9.9.1.** Em síntese, a definição quantitativa dos itens da contratação deverá observar a arquitetura de referência da solução e a estratégia de crescimento sob demanda, considerando não apenas a quantidade de servidores e GPUs, mas também a necessidade de switches, interfaces de rede, conectividade SAN, capacidade de storage em arquivo, objeto e bloco, servidores de gerenciamento, appliance de backup, softwares estruturantes e serviços especializados de implantação, integração, configuração, capacitação, operação assistida e sustentação.

**7.9.9.2.** A estimativa de quantidade de bens e serviços deverá, portanto, refletir a lógica segundo a qual a solução de GPU como serviço constitui plataforma integrada, composta por elementos complementares e articulados, e não mera reunião de componentes isolados.

**7.9.9.3.** O atendimento à demanda institucional depende da composição coerente desses elementos, observadas as necessidades técnicas da arquitetura, os requisitos de desempenho, a compatibilidade entre componentes e o modelo de expansão progressiva da plataforma.

**7.10. Quantidade de bens e serviços**

**7.10.1.** A definição dos quantitativos de bens e serviços da presente contratação foi realizada considerando a necessidade de implantação de uma solução efetivamente completa de GPU como serviço, composta por hardware, softwares, integração, implantação, suporte, operação assistida, capacitação e demais serviços especializados indispensáveis à entrada em operação da plataforma.

**7.10.1.1.** A composição dos itens foi orientada pelo princípio de contratação de solução integrada do tipo turn-key, segundo o qual a contratada deverá entregar a plataforma em condições de funcionamento conjunto, assumindo a responsabilidade pela integração entre todos os componentes de hardware, software, serviços, suporte, documentação, capacitação e operação assistida necessários à efetiva disponibilização de GPU como serviço.

**7.10.1.2.** A adoção do modelo turn-key é fator essencial para o sucesso da contratação, pois a solução pretendida não corresponde à simples aquisição de servidores com GPU, mas à implantação de uma plataforma corporativa integrada, na qual a camada de hardware acelerado, a camada de software, a rede, os mecanismos de autosserviço, a observabilidade, a segurança operacional e os serviços especializados precisam operar de forma coerente, suportada e validada.

**7.10.1.3.** Nesse contexto, a validação da solução pelo fabricante do hardware dos servidores e a participação efetiva do fabricante no desenho, implantação, suporte e evolução da plataforma constituem elementos essenciais para mitigação de riscos técnicos. A complexidade do ambiente de inteligência artificial, especialmente em soluções de alta densidade com GPUs, exige que a arquitetura ofertada não dependa exclusivamente de montagem, compatibilização ou customização realizada pela licitante ou por terceiros sem respaldo formal do fabricante.

**7.10.1.4.** A exigência de participação efetiva do fabricante busca assegurar que a solução esteja aderente a arquiteturas de referência, boas práticas de implantação, matrizes de compatibilidade, recomendações de operação e condições de suporte oficialmente reconhecidas. Tal diretriz reduz riscos de incompatibilidade, indefinição de responsabilidades, falhas de integração, dificuldades de sustentação e perda de garantia ou suporte em caso de incidentes críticos.

**7.10.1.5.** Dessa forma, os quantitativos ora definidos devem ser compreendidos em conjunto com a premissa de que os componentes fornecidos deverão compor uma solução única, integrada, homologada ou formalmente validada pelo fabricante, com suporte oficial e com evidências documentais que demonstrem a compatibilidade entre hardware, softwares, rede, serviços e demais elementos necessários à operação da plataforma.

**7.10.2.** A necessidade orçamentária estimada para a contratação foi indicada ao escalão superior na ordem de R$ 20.000.000,00, sem que tenha sido emitida, até o momento, correspondente Previsão de Recurso Orçamentário. A PRO atualmente ratificada para o exercício de 2026 é de R$ 2.000.000,00, valor que orienta o cronograma de execução e desembolso inicial, permanecendo as parcelas de maior vulto condicionadas à efetiva disponibilidade dos créditos necessários nos exercícios subsequentes.

**7.10.3.** Em razão dessa limitação orçamentária, as tratativas realizadas com fabricantes e integradores buscaram identificar a configuração mínima capaz de entregar uma solução completa, funcional e operacionalmente utilizável, ainda que restrita em capacidade inicial, evitando-se a aquisição isolada

de equipamentos sem a correspondente camada lógica, de governança, de autosserviço, de observabilidade e de sustentação técnica.

**7.10.4.** A estratégia adotada não consiste em reduzir o escopo funcional da solução, mas em limitar a escala inicial da capacidade computacional, preservando os componentes essenciais para que a plataforma seja implantada como GPU como serviço desde a primeira fase, com possibilidade de crescimento futuro de acordo com o uso efetivo, a maturidade operacional da Administração e a disponibilidade de novos recursos orçamentários.

**7.10.5.** Assim, optou-se por estruturar a contratação com 01 (um) servidor de processamento com GPU, 01 (um) conjunto de servidores para hospedagem da camada de software da solução, os softwares corporativos necessários à operação da plataforma, serviço de integração com a rede de frontend existente, serviços especializados de implantação e operação assistida, suporte continuado e capacitação técnica avançada.

**7.10.5.1.** A estruturação da contratação nesses termos busca preservar a competitividade entre fabricantes de soluções completas de IA, sem impor que todos os potenciais fornecedores forneçam, nesta etapa, todos os componentes físicos já existentes ou já padronizados no Data Center do Exército em processos próprios. Ao mesmo tempo, mantém-se a exigência de que a solução ofertada seja compatível com o ambiente existente e entregue sob responsabilidade integrada da contratada.

**7.10.5.2.** Desse modo, evita-se tanto a fragmentação excessiva da contratação quanto a imposição de barreiras desnecessárias à participação de fabricantes distintos, preservando-se a lógica de solução turn-key, a responsabilidade única pela integração e a necessidade de validação formal da compatibilidade entre os componentes ofertados e a infraestrutura institucional.

**7.10.5.3.** A aquisição será composta de uma primeira parte com 8 (oito) GPUs dentro da primeira fase de 80 (oitenta) GPU. A primeira parte que será a o objetivo dessa contratação está detalhada na Figura 4.

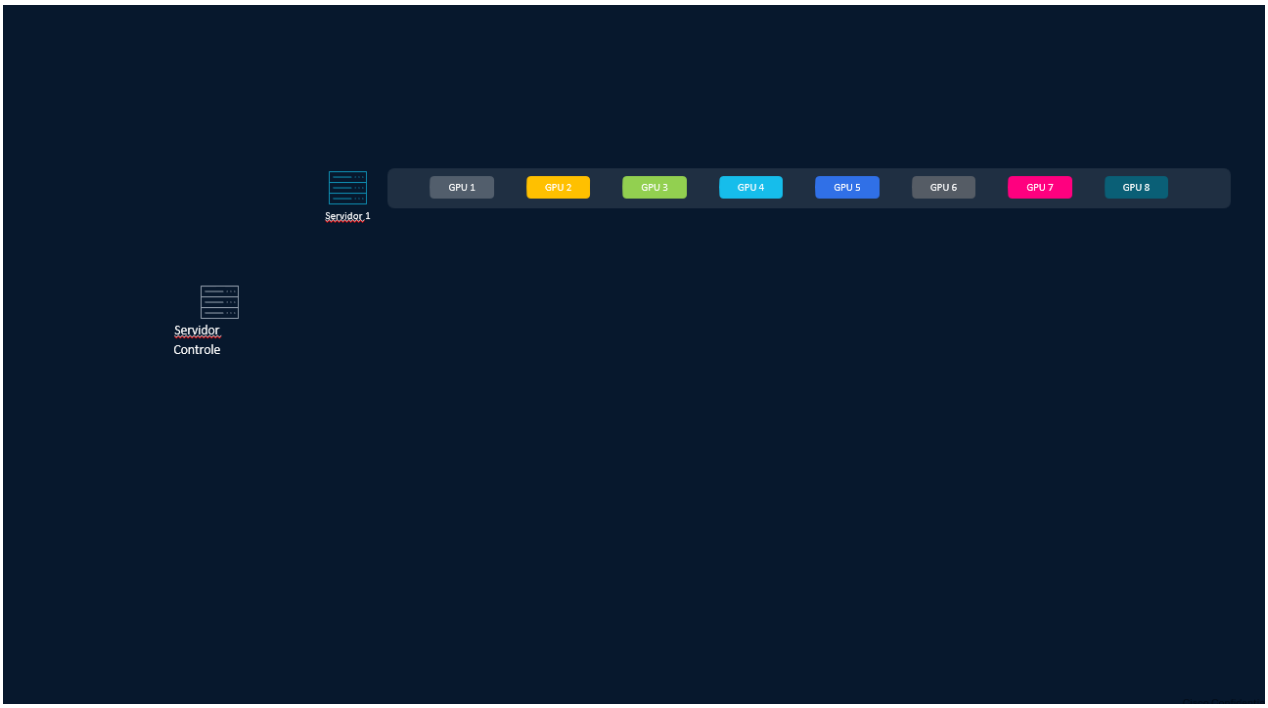



**Figura 5:** Diagrama da Primeira Parte.

**7.10.6.** A quantidade inicial de 01 (um) servidor de processamento com GPU justifica-se por representar a menor unidade prática de implantação de capacidade acelerada compatível com a

arquitetura pretendida, permitindo disponibilizar um ambiente real de inferência, treinamento, ajuste fino, desenvolvimento interativo e execução de workloads de inteligência artificial, sem que seja necessário aguardar a formação de uma demanda plenamente consolidada para dezenas de GPUs.

**7.10.7.** A escolha por um servidor de alta densidade com 8 (oito) GPUs permite iniciar a operação com capacidade suficiente para validar o modelo de uso institucional, coletar métricas reais de consumo, avaliar perfis de usuários, identificar gargalos, medir concorrência, testar políticas de cotas, estabelecer padrões operacionais e subsidiar futuras expansões com base em evidências concretas.

**7.10.8.** A adoção de 8 (oito) GPUs em um único servidor também preserva a coerência arquitetural da solução, pois permite suportar workloads que demandem múltiplas GPUs no mesmo nó, reduzindo complexidade inicial de rede backend multinó, ao mesmo tempo em que mantém aderência à estratégia de crescimento para ambientes com múltiplos servidores GPU em fases futuras.

**7.10.9.** No que se refere à garantia e ao suporte dos componentes de hardware, definiu-se prazo de 60 (sessenta) meses para os itens 1 a 4, abrangendo o servidor de processamento com GPU, seu componente de garantia e licenciamento, o servidor para hospedagem da camada de software da solução e seu respectivo componente de garantia e licenciamento.

**7.10.10.** A definição de 60 (sessenta) meses para o hardware justifica-se pela elevada criticidade, complexidade técnica, custo unitário e expectativa de vida útil dos servidores envolvidos, especialmente do servidor de processamento com GPU, cuja substituição prematura ou descontinuidade de suporte poderia comprometer a continuidade da plataforma e elevar o custo total de propriedade.

**7.10.11.** Por outro lado, definiu-se prazo de 12 (doze) meses para os softwares e serviços, em razão da necessidade de equilibrar sustentação técnica, atualização tecnológica, restrição orçamentária e flexibilidade para evolução futura da solução, considerando que o mercado de plataformas de inteligência artificial, orquestração, autosserviço e observabilidade apresenta ritmo acelerado de mudança.

**7.10.12.** Essa distinção entre garantia de hardware por 60 (sessenta) meses e softwares/serviços por 12 (doze) meses é compatível com a estratégia de implantação progressiva adotada pela Administração, pois preserva o investimento estrutural em equipamentos de alto valor e, simultaneamente, permite reavaliar, renovar ou expandir os componentes lógicos e os serviços de sustentação de acordo com o amadurecimento do uso efetivo da plataforma.

**7.10.13.** A opção por não adquirir, nesta etapa inicial, uma solução externa dedicada de armazenamento de grande porte decorre da incerteza quanto ao perfil real de demanda de dados, datasets, checkpoints, modelos, artefatos, bases documentais, imagens e workloads que serão efetivamente processados pela plataforma na fase inicial.

**7.10.14.** Em substituição à aquisição imediata de storage específico, optou-se por prever discos internos de alta performance no servidor de processamento com GPU, de modo a disponibilizar capacidade local suficiente para a etapa inicial, reduzir custo de implantação, manter aderência à previsão orçamentária e permitir que a Administração colete métricas reais antes de definir eventual expansão de armazenamento externo.

**7.10.15.** Essa decisão é tecnicamente adequada para a fase inicial, pois o ambiente parte de 01 (um) servidor de processamento com GPU, no qual parte relevante das cargas poderá utilizar armazenamento local de alta performance para datasets temporários, cache, staging, checkpoints, artefatos de desenvolvimento e demais dados de uso imediato.

**7.10.16.** A solução de armazenamento interno, entretanto, não deve ser compreendida como substituto definitivo de uma arquitetura corporativa de armazenamento para IA. Trata-se de medida de racionalização da primeira etapa, compatível com a restrição orçamentária e com a ausência de métricas maduras de demanda, sem prejuízo de futura aquisição de storage externo quando o crescimento da plataforma, a multiplicidade de servidores GPU ou o padrão real de uso assim exigirem.

**7.10.17.** Desse modo, os quantitativos definidos para esta contratação devem ser compreendidos como a menor composição tecnicamente coerente para entrega de uma solução completa de GPU como serviço, e não como o dimensionamento final da capacidade institucional de processamento de inteligência artificial do Exército Brasileiro.

**7.10.18.** O Item 1 – Servidor de Processamento com GPU, componente de hardware, será contratado na quantidade de 01 (uma) unidade. Esse item constitui o núcleo da capacidade computacional acelerada da solução, sendo responsável pelo processamento principal dos workloads de inteligência artificial, incluindo inferência, treinamento, ajuste fino, desenvolvimento interativo e execução de modelos de maior complexidade.

**7.10.19.** A especificidade técnica do Item 1 decorre da necessidade de disponibilizar, desde a primeira etapa, hardware compatível com arquitetura de referência para ambientes corporativos de IA, contendo GPUs de alto desempenho, grande volume de memória de vídeo, processadores, memória RAM, interfaces de rede e armazenamento local suficientes para sustentar cargas reais de trabalho de inteligência artificial.

**7.10.20.** A quantidade de 01 (uma) unidade do Item 1 é tecnicamente suficiente para a etapa inicial porque permite implantar a plataforma em ambiente produtivo controlado, iniciar a prestação do serviço, validar o modelo operacional, medir a utilização efetiva e preparar futuras expansões sem incorrer, neste momento, no custo de aquisição de múltiplos servidores GPU.

**7.10.20.1.** A exigência de que o servidor de processamento com GPU esteja aderente à arquitetura de referência de IA e seja homologado pelo fabricante do servidor decorre da necessidade de assegurar que GPUs, CPUs, memória, interfaces de rede, fontes, refrigeração, firmware, BIOS, gerenciamento, drivers e camada de aceleração operem de forma integrada e suportada.

**7.10.20.2.** A validação pelo fabricante é especialmente relevante nesse item em razão da elevada densidade energética, térmica e computacional do equipamento, bem como da necessidade de integração com softwares de virtualização, contêineres, orquestração, aceleração por GPU, observabilidade e autosserviço. Em caso de falha ou degradação de desempenho, a existência de validação formal reduz risco de indefinição de responsabilidades entre integrador, fabricante de hardware, fabricante de software e Administração.

**7.10.21.** O Item 2 – Servidor de Processamento com GPU, componente de garantia e licenciamento, será contratado na quantidade de 01 (uma) unidade, vinculado ao hardware do Item 1. Sua finalidade é assegurar a cobertura de garantia, suporte, direito de uso, licenciamento embarcado ou demais componentes necessários à plena operação do servidor GPU pelo período definido para o hardware.

**7.10.22.** A contratação segregada do componente de garantia e licenciamento do servidor de GPU permite evidenciar o custo de sustentação do equipamento de maior criticidade da solução além de possibilitar incidência de impostos adequada quando da emissão de notas fiscais por parte do fornecedor, assegurando que a Administração disponha de suporte compatível com a complexidade, densidade energética, criticidade operacional e custo de reposição do ambiente acelerado.

**7.10.23.** O Item 3 – Servidor para Hospedagem da Camada de Software da Solução, componente de hardware, será contratado na quantidade de 01 (uma) unidade. Esse servidor ou conjunto de servidores terá a função de hospedar os componentes lógicos necessários ao funcionamento da plataforma, tais como orquestração, gerenciamento, automação, observabilidade, segurança, interfaces de serviços, catálogos, portais, APIs e demais elementos de controle. Adicionalmente, esse item tem a dependência do fabricante da solução ofertada, devendo seguir as especificações funcionais e serem adequados à solução apresentada pelo fabricante.

**7.10.23.1.** O servidor para hospedagem da camada de software deverá ser dimensionado, desde a entrega inicial, para atender não apenas ao servidor GPU adquirido nesta primeira etapa, mas à fase 1 completa da arquitetura, correspondente a até 10 (dez) servidores de processamento com GPU, totalizando até 80 (oitenta) GPUs.

**7.10.23.2.** Essa diretriz justifica-se porque a camada de software constitui componente estruturante da solução e não deve precisar ser redesenhada ou substituída a cada expansão incremental do ambiente. O dimensionamento inicial do servidor de orquestração para a capacidade de até 80 GPUs permite preservar a continuidade arquitetural, reduzir retrabalho, evitar novas interrupções de implantação e assegurar que o crescimento futuro ocorra por acréscimo de capacidade acelerada, e não por reconstrução da camada de controle.

**7.10.23.3.** A exigência de capacidade imediata para a fase 1 completa também é coerente com a estratégia de restrição orçamentária adotada, pois concentra o investimento inicial nos componentes estruturantes da plataforma e permite que futuras expansões de GPU utilizem a mesma base lógica, administrativa e operacional já implantada.

**7.10.24.** A separação física ou lógica entre o servidor de processamento com GPU e o servidor de hospedagem da camada de software é necessária para preservar estabilidade, governança, segurança e desempenho da plataforma, evitando que os serviços de controle, administração e autosserviço concorram indevidamente com os workloads de IA executados no servidor acelerado.

**7.10.25.** A quantidade de 01 (uma) unidade do Item 3 justifica-se por ser suficiente para a fase inicial da contratação, na qual haverá apenas 01 (um) servidor GPU, sem prejuízo de que o dimensionamento interno do equipamento, sua configuração de disponibilidade e sua capacidade de expansão sejam compatíveis com o crescimento futuro da solução. O fornecedor deverá indicar na proposta técnica o detalhamento do servidor ou mesmo do conjunto de servidores para completo atendimento a este item da contratação

**7.10.26.** O Item 4 – Servidor para Hospedagem da Camada de Software da Solução, componente de garantia e licenciamento, será contratado na quantidade de 01 (uma) unidade, vinculado ao Item 3. Esse item assegura a cobertura de garantia, suporte e demais licenças ou direitos necessários à operação contínua da camada de software instalada no servidor de orquestração.

**7.10.27.** A contratação do Item 4 é necessária porque a indisponibilidade da camada de software comprometeria o consumo da capacidade de GPU como serviço, ainda que o servidor acelerado permanecesse fisicamente operacional. Portanto, o suporte ao servidor de orquestração deve acompanhar a criticidade da plataforma como um todo. A separação em item de hardware e software e serviços também tem a função de de possibilitar incidência de impostos adequada quando da emissão de notas fiscais por parte do fornecedor

**7.10.28.** O Item 5 – Solução de Software de Virtualização, Contêineres e Orquestração, será contratado na quantidade de 01 (uma) solução. Esse item constitui a plataforma-base para execução

coordenada de workloads, gerenciamento de cluster, virtualização, contêineres, alta disponibilidade, isolamento, escalabilidade e integração com os demais componentes da arquitetura.

**7.10.28.1.** Para atendimento do Item 5, poderá ser proposta uma solução composta por mais de um produto de software, desde que a combinação apresentada seja integralmente compatível, suportada e validada pelo fabricante da solução ou pelo fabricante do hardware dos servidores, mediante documentação pública, arquitetura de referência, matriz oficial de compatibilidade ou declaração específica do fabricante.

**7.10.28.2.** A possibilidade de composição por mais de um software decorre da natureza das plataformas modernas de IA, nas quais capacidades de virtualização, contêineres, orquestração, cluster management e integração com workloads acelerados podem ser atendidas por produtos complementares. Entretanto, tal flexibilidade não afasta a exigência de validação formal da combinação ofertada, uma vez que a Administração não pretende assumir risco de integração entre produtos não testados em conjunto ou dependentes de customização exclusiva da licitante.

**7.10.29.** A quantidade de 01 (uma) solução de virtualização, contêineres e orquestração justifica-se porque a contratação busca implantar uma plataforma única, corporativa e padronizada, ainda que dimensionada inicialmente para a quantidade de GPUs efetivamente adquirida nesta etapa. O licenciamento deverá cobrir todos os recursos necessários ao atendimento das funcionalidades exigidas.

**7.10.30.** O Item 6 – Solução de Software Específica para Aceleração por GPU, trata especificamente do licencimaneto da NVIDIA para possibilitar a solução de GPU como serviço pretendida, e será contratado na quantidade de 01 (uma) solução. Esse item é indispensável para habilitar o uso adequado das GPUs pela camada de software, contemplando drivers, runtimes, bibliotecas, operadores, integrações, ferramentas de gerenciamento, componentes de inferência, treinamento e demais recursos de aceleração necessários ao ambiente de IA.

**7.10.30.1.** Para atendimento do Item 6, poderá ser proposta combinação de softwares e componentes específicos de aceleração da NVIDIA, desde que todos estejam cobertos por licenciamento empresarial, suporte oficial e validação formal com o hardware ofertado e com a plataforma de orquestração. Essa validação deverá demonstrar que os componentes de aceleração, drivers, runtimes, operadores, bibliotecas, telemetria, particionamento, inferência e treinamento funcionam de forma integrada e suportada.

**7.10.30.2.** A exigência de validação do conjunto é necessária porque a camada de aceleração por GPU constitui elo crítico entre o hardware físico e os workloads de inteligência artificial. Incompatibilidades nessa camada podem comprometer desempenho, estabilidade, observabilidade, segurança, isolamento entre usuários, atualização de drivers e suporte a frameworks, razão pela qual não se admite solução baseada em composição informal de componentes sem respaldo do fabricante.

**7.10.30.3.** A validação do conjunto deve ser feita por documentação específica do fabricante de hardware ou da própria NVIDIA.

**7.10.31.** A contratação desse entregável justifica-se porque o hardware acelerado, isoladamente, não é suficiente para disponibilizar GPU como serviço. É necessário que exista camada lógica suportada oficialmente pelo fabricante, capaz de expor a capacidade das GPUs aos workloads de forma controlada, segura, observável e compatível com os frameworks e aplicações de inteligência artificial. Para essa finalidade, deverá ser fornecido o NVIDIA AI Enterprise, em conformidade com a indicação excepcional de produto e com as justificativas técnicas e legais apresentadas no item 7.4.2.

**7.10.32.** O Item 7 – Solução de Software para Interface de Serviços e Autosserviço, será contratado na quantidade de 01 (uma) solução. Esse item tem por finalidade transformar a capacidade de processamento em serviço efetivamente consumível por usuários, equipes e projetos autorizados, por meio de portal, APIs, catálogo de serviços, políticas de acesso, cotas e mecanismos de solicitação e acompanhamento de recursos.

**7.10.32.1.** Para atendimento do Item 7, poderá ser proposta solução composta por mais de um software, desde que o conjunto ofertado seja apresentado como camada funcionalmente integrada, validada pelo fabricante e suportada de forma empresarial. A combinação deverá demonstrar capacidade de prover autosserviço, portal, APIs, catálogo, controle de cotas, segregação multiusuário, governança de consumo e integração com a camada de orquestração e aceleração por GPU.

**7.10.32.2.** A validação desse conjunto é necessária porque a camada de interface de serviços é responsável por materializar o conceito de GPU como serviço. Caso essa camada seja composta por produtos desconectados ou por integrações customizadas sem suporte formal, haverá risco de que a solução entregue seja apenas administrável por especialistas, mas não efetivamente consumível pela Administração de forma padronizada, governada e escalável.

**7.10.33.** A contratação de 01 (uma) solução de interface de serviços e autosserviço justifica-se pela necessidade de evitar que o uso da infraestrutura dependa exclusivamente de intervenção manual da equipe de infraestrutura. A solução deve permitir governança do consumo, padronização dos ambientes e acesso controlado à capacidade computacional, conforme os perfis de uso definidos pela Administração.

**7.10.34.** O Item 8 – Solução de Software de Observabilidade e Segurança Operacional, será contratado na quantidade de 01 (uma) solução. Esse item deverá prover visibilidade sobre infraestrutura, workloads, consumo de recursos, eventos, falhas, desempenho, trilhas de auditoria, comportamento dos aceleradores, segurança operacional e indicadores gerenciais.

**7.10.34.1.** Para atendimento do Item 8, poderá ser proposta solução composta por mais de um software, desde que a combinação apresentada seja oficialmente suportada, licenciada em regime empresarial e validada para operação conjunta com a plataforma de GPU como serviço. A comprovação poderá ocorrer por documentação pública do fabricante, arquitetura de referência, matriz de compatibilidade ou declaração específica que indique a responsabilidade pelo suporte da combinação ofertada.

**7.10.34.2.** A exigência de validação é especialmente relevante para observabilidade e segurança operacional, pois esse item deverá coletar, correlacionar e apresentar dados de servidores, GPUs, workloads, orquestração, APIs, portais, logs, eventos, alertas, consumo, auditoria e indicadores de capacidade. Uma solução formada por múltiplos produtos sem validação formal poderia gerar lacunas de monitoramento, perda de rastreabilidade, dificuldade de troubleshooting e indefinição de responsabilidade em caso de incidentes.

**7.10.35.** A contratação desse item é indispensável porque a solução de GPU como serviço precisa ser monitorada continuamente, tanto para fins de disponibilidade e desempenho quanto para fins de governança, auditoria, segurança, planejamento de capacidade, detecção de gargalos e suporte à tomada de decisão sobre futuras expansões.

**7.10.36.** O Item 9 – Serviço de Integração com Equipamentos de Rede de Frontend, será contratado na quantidade de 01 (um) serviço. Esse item visa assegurar que a solução seja adequadamente integrada à infraestrutura de rede do Data Center, permitindo o acesso dos usuários, sistemas, APIs, portais e serviços de administração ao ambiente de GPU como serviço.

**7.10.37.** A opção por contratar serviço de integração com a rede de frontend, e não necessariamente novos equipamentos de rede nesta etapa, decorre da estratégia de compatibilização com a previsão orçamentária e de aproveitamento da infraestrutura existente, desde que preservados os requisitos de desempenho, segurança, segmentação, disponibilidade e aderência à arquitetura institucional.

**7.10.37.1.** A contratação do Item 9, como serviço de integração com equipamentos de rede de frontend, representa a alternativa mais viável para compatibilizar a padronização já existente do Data Center do Exército com a ampliação da competitividade da presente contratação.

**7.10.37.2.** Os equipamentos de rede Cisco já foram padronizados no Data Center em processo específico, de modo que a presente contratação não deve impor, de forma desnecessária, que um fornecedor vinculado a outro fabricante de solução de IA, como HPE, Dell, Lenovo, Supermicro ou outro, forneça switches Cisco como parte de sua proposta. Tal imposição poderia restringir indevidamente a participação de fabricantes capazes de entregar a solução de GPU como serviço, mas que não atuam ordinariamente como fornecedores principais da rede Cisco já padronizada.

**7.10.37.3.** Por outro lado, a não inclusão dos switches de frontend como fornecimento direto nesta contratação não afasta a necessidade de integração plena com a rede existente. A contratada deverá demonstrar que a solução ofertada é compatível com os equipamentos e softwares de rede Cisco em uso pela Contratante, garantindo que a plataforma de GPU como serviço opere de forma integrada, segura, suportada e aderente ao ambiente institucional.

**7.10.37.4.** Dessa forma, o Item 9 preserva o princípio turn-key da aquisição, pois mantém sob responsabilidade da contratada a integração da solução com a rede de frontend existente, sem obrigar que todos os potenciais fornecedores passem a fornecer equipamentos de rede de fabricante distinto daquele de sua solução principal. A contratada continuará responsável pela entrega funcional da plataforma, pela compatibilidade da solução com a rede institucional e pelo suporte à integração necessária à operação.

**7.10.37.5.** Essa modelagem reduz risco de restrição indevida à competição, evita duplicidade de aquisição de switches já padronizados, preserva a arquitetura de rede do Data Center do Exército e, simultaneamente, exige comprovação formal de compatibilidade com a solução ofertada, por meio de documentação pública, matriz de compatibilidade, arquitetura validada ou declaração específica do fabricante da solução.

**7.10.38.** O Item 10 – Reunião Inicial, será contratado na quantidade de 01 (uma) unidade de serviço. Esse item marca o início formal da execução contratual, viabilizando alinhamento entre contratante e contratada quanto a escopo, cronograma, responsáveis, metodologia de execução, riscos, entregáveis, pontos de controle e condições de aceite.

**7.10.38.1.** A participação da contratada e, quando aplicável, de representantes ou especialistas do fabricante na reunião inicial é necessária para assegurar que as premissas de solução validada, responsabilidades de integração, matriz de suporte, procedimentos de escalonamento e pontos críticos de implantação sejam esclarecidos desde o início da execução contratual.

**7.10.39.** O Item 11 – Gerenciamento de Projetos, será contratado na quantidade de 01 (um) serviço. Sua necessidade decorre da complexidade da solução, da multiplicidade de componentes de hardware, software e serviços, da participação de fabricantes, da criticidade do ambiente de Data Center e da necessidade de controle coordenado de prazos, riscos, responsabilidades e entregáveis.

**7.10.39.1.** O gerenciamento de projetos deverá contemplar a coordenação entre contratada, fabricantes envolvidos e Administração, de modo a preservar o caráter turn-key da contratação. Essa coordenação é indispensável para evitar lacunas entre responsabilidades de hardware, software, rede, implantação, suporte e operação assistida, assegurando que todos os componentes sejam entregues como uma solução única e funcional.

**7.10.40.** O Item 12 – Proposta de Escopo de Trabalho, será contratado na quantidade de 01 (um) entregável. Esse item deverá consolidar, em documento formal, o escopo detalhado das atividades, responsabilidades, premissas, restrições, cronograma, produtos esperados e critérios de aceitação, reduzindo riscos de interpretação divergente durante a execução.

**7.10.40.1.** A Proposta de Escopo de Trabalho deverá indicar claramente quais atividades serão realizadas diretamente pela contratada, quais contarão com participação do fabricante, quais entregáveis deverão possuir anuência ou validação do fabricante e quais evidências serão apresentadas para comprovar a aderência da solução ao desenho validado.

**7.10.41.** O Item 13 – Levantamento da Situação Atual, será contratado na quantidade de 01 (um) serviço. Esse levantamento é necessário para identificar as condições reais do ambiente de instalação, incluindo infraestrutura física, energia, refrigeração, rede, segurança, integração com serviços institucionais, dependências técnicas e restrições operacionais.

**7.10.41.1.** A participação de especialistas do fabricante no levantamento da situação atual, quando, justifica-se pela necessidade de avaliar se as condições físicas e lógicas do ambiente são compatíveis com a solução validada, especialmente quanto a energia, refrigeração, rede, conectividade, integração com diretórios, segurança, observabilidade e requisitos de operação de servidores de alta densidade.

**7.10.42.** O Item 14 – Tomada de Conhecimento da Documentação Existente, será contratado na quantidade de 01 (um) serviço. Esse item justifica-se pela necessidade de que a contratada compreenda a documentação técnica, padrões internos, normas de operação, arquitetura de Data Center, procedimentos de segurança e demais referências institucionais aplicáveis à implantação da solução.

**7.10.42.1**. A tomada de conhecimento da documentação existente é relevante para que a contratada e os fabricantes envolvidos compreendam as restrições e padrões institucionais do Exército antes da definição do desenho final. Isso reduz o risco de elaboração de arquitetura incompatível com o ambiente real e favorece a implantação de solução validada sem depender de adaptações improvisadas durante a execução.

**7.10.43.** O Item 15 – Apontamento de Oportunidades de Melhoria, será contratado na quantidade de 01 (um) entregável. Esse item permitirá que, a partir do levantamento do ambiente e da análise documental, sejam identificadas oportunidades de ajuste, mitigação de riscos, melhoria de segurança, otimização de desempenho e aperfeiçoamento operacional antes da implantação definitiva.

**7.10.43.1.** Quando as oportunidades de melhoria estiverem relacionadas a arquitetura, compatibilidade, integração, desempenho, suporte ou segurança da solução ofertada, deverá haver participação ou manifestação do fabricante, de modo a assegurar que as recomendações permaneçam aderentes à arquitetura validada e às melhores práticas oficiais da solução.

**7.10.44.** O Item 16 – Definição de Requisitos da Solução, será contratado na quantidade de 01 (um) serviço. Esse item é necessário para transformar as necessidades institucionais, técnicas e operacionais em requisitos executáveis, alinhando a expectativa da Administração à arquitetura de referência, aos produtos ofertados, às limitações da primeira etapa e às possibilidades de expansão futura.

**7.10.44.1.** A participação do fabricante no desenvolvimento ou validação dos requisitos da solução justifica-se porque os requisitos de GPU como serviço envolvem interação entre hardware acelerado, drivers, orquestração, autosserviço, observabilidade, rede, segurança e suporte. A validação pelo fabricante reduz o risco de definição de requisitos tecnicamente incompatíveis com a solução ofertada ou dependentes de customizações sem suporte oficial.

**7.10.45.** O Item 17 – Treinamento Inicial da Solução, será contratado como atividade de capacitação introdutória vinculada à implantação. Sua finalidade é preparar a equipe técnica da Administração para acompanhar o projeto, compreender os principais componentes, validar as decisões de desenho e participar de forma efetiva das etapas de implantação e aceite.

**7.10.45.1.** O treinamento inicial deverá, quando aplicável, contar com conteúdo, instrutores, material ou validação do fabricante, de forma a assegurar que os conhecimentos transmitidos estejam aderentes às boas práticas oficiais da solução e não apenas a procedimentos específicos do integrador.

**7.10.46.** O Item 18 – Planejamento de Alto Nível da Solução (High Level Design - HLD), será contratado na quantidade de 01 (um) entregável. Esse item deverá estabelecer a visão arquitetural da solução, a distribuição dos componentes, as integrações principais, os fluxos de comunicação, os domínios de segurança, a estratégia de implantação e a aderência da solução à infraestrutura institucional.

**7.10.46.1.** O Planejamento de Alto Nível deverá contar com participação ou anuência do fabricante, especialmente quanto à arquitetura de referência, matriz de compatibilidade, dimensionamento da camada de controle, integração com rede de frontend, expansão futura para até 80 GPUs, requisitos de energia e refrigeração, suporte aos softwares ofertados e aderência às melhores práticas de implantação.

**7.10.46.2.** Essa exigência reduz risco de que o HLD represente apenas um desenho elaborado pelo integrador, sem respaldo do fabricante responsável pela solução. Em uma contratação turn-key de alta complexidade, o HLD deve materializar uma arquitetura viável, suportada, validada e compatível com o ciclo de vida da solução.

**7.10.47.** O Item 19 – Proposta Técnica de Implementação, será contratado na quantidade de 01 (um) entregável. Esse documento deverá detalhar a forma de execução da implantação, incluindo sequência de atividades, pré-requisitos, janelas técnicas, responsabilidades, riscos, critérios de validação e dependências entre hardware, software, rede e serviços.

**7.10.47.1.** A Proposta Técnica de Implementação deverá demonstrar a participação do fabricante na definição dos procedimentos críticos, especialmente nas atividades que envolvam instalação, parametrização, atualização, integração e validação de componentes estruturantes da plataforma.

Essa exigência é necessária para reduzir risco de falhas de implementação e assegurar que a solução seja implantada conforme práticas oficialmente suportadas.

**7.10.48.** O Item 20 – Desenho de Baixo Nível da Solução (Low Level Design - LLD), será contratado na quantidade de 01 (um) entregável. Esse item é necessário para converter o planejamento de alto nível em parâmetros técnicos detalhados, tais como endereçamento, segmentação, integrações, políticas, configurações, componentes instalados, fluxos de autenticação, observabilidade, automação e segurança.

**7.10.48.1.** O Desenho de Baixo Nível deverá ser submetido à anuência ou validação do fabricante, sempre que envolver configurações estruturantes da solução, matriz de compatibilidade, melhores práticas de segurança, integração entre softwares, parâmetros de desempenho, rede, GPU, orquestração, observabilidade, autosserviço e suporte.

**7.10.48.2.** Essa exigência justifica-se porque o LLD contém as definições que efetivamente orientarão a implantação. Em solução de GPU como serviço, decisões inadequadas de baixo nível podem comprometer desempenho, estabilidade, segurança, escalabilidade e suporte. A validação do fabricante reduz esse risco e reforça a responsabilidade pela entrega turn-key.

**7.10.49.** O Item 21 – Preparação da Documentação de Instalação, será contratado na quantidade de 01 (um) entregável. Sua finalidade é documentar, previamente à execução, os procedimentos de instalação, configuração, integração, testes, rollback, aceite e transição operacional, reduzindo riscos durante a implantação.

**7.10.49.1.** A documentação de instalação deverá refletir procedimentos suportados pelo fabricante, evitando práticas não documentadas, scripts exclusivos não homologados, customizações sem respaldo ou dependências técnicas que possam comprometer o suporte futuro da solução.

**7.10.50.** O Item 22 – Atualização da Documentação Técnica Existente, será contratado na quantidade de 01 (um) serviço. Esse item justifica-se porque a nova plataforma deverá ser incorporada à documentação institucional do ambiente de Data Center, garantindo rastreabilidade, padronização, continuidade operacional e suporte futuro pela equipe da Administração.

**7.10.50.1.** A atualização da documentação técnica existente deverá registrar, quando aplicável, a referência aos documentos de fabricante, matrizes de compatibilidade, versões de software, modelos de hardware, contratos de suporte, procedimentos de escalonamento e evidências de validação da solução, de forma a preservar rastreabilidade técnica durante todo o ciclo de vida da plataforma.

**7.10.51.** O Item 23 – Treinamento Antes da Implementação, será contratado na quantidade de 01 (uma) atividade. Esse treinamento é necessário para assegurar que a equipe técnica compreenda, antes da implantação, os procedimentos, riscos, componentes, etapas de execução e critérios de validação, permitindo participação ativa da Administração durante a implementação.

**7.10.51.1.** Esse treinamento deverá enfatizar o desenho validado, a arquitetura turn-key, as responsabilidades da contratada, os pontos de integração com fabricantes, os procedimentos de verificação e os cuidados necessários durante a implantação, de modo a preparar a equipe da Administração para acompanhar a execução e identificar desvios relevantes.

**7.10.52.** O Item 24 – Implementação, será contratado na quantidade de 01 (um) serviço. Esse item compreende a instalação, configuração, integração e ativação da solução, envolvendo os servidores, softwares, rede, observabilidade, segurança, autosserviço, aceleração por GPU e demais componentes necessários à disponibilização da plataforma.

**7.10.52.1.** A implementação deverá ocorrer com participação efetiva da contratada e, quando aplicável, de especialistas do fabricante, especialmente nas atividades críticas de instalação, configuração, integração, validação, testes, atualização e ativação da solução. Tal exigência decorre da necessidade de assegurar que a plataforma seja implantada conforme arquitetura validada e permaneça coberta por suporte oficial.

**7.10.52.2.** A participação do fabricante nessa etapa reduz risco de falhas de configuração, incompatibilidades entre versões, erros de parametrização, baixa performance, problemas de observabilidade, limitações de suporte e necessidade de retrabalho após a entrada em produção.

**7.10.53.** O Item 25 – Treinamento Após a Implementação, será contratado na quantidade de 01 (uma) atividade. Sua finalidade é capacitar a equipe com base no ambiente efetivamente implantado, permitindo aprendizagem prática sobre operação, administração, monitoramento, troubleshooting, gestão de workloads, uso do autosserviço e sustentação da plataforma.

**7.10.53.1.** O treinamento após a implementação deverá utilizar a solução efetivamente entregue, contemplando os fluxos operacionais reais, a matriz de suporte, os mecanismos de escalonamento ao fabricante, os limites da solução, os procedimentos de atualização, as rotinas de observabilidade e os cuidados para preservar a aderência ao desenho validado.

**7.10.54.** O Item 26 – Atualização da Documentação, será contratado na quantidade de 01 (um) serviço. Esse item deverá registrar a configuração final implantada, eventuais ajustes realizados durante a execução, desvios autorizados, resultados de testes, informações operacionais e demais elementos necessários à manutenção do conhecimento institucional.

**7.10.54.1.** A atualização final da documentação deverá evidenciar a configuração efetivamente aceita pela Administração, bem como as validações realizadas pela contratada e, quando aplicável, pelo fabricante. Essa medida assegura que a documentação final corresponda à solução realmente implantada e não apenas ao desenho inicialmente proposto.

**7.10.55.** O Item 27 – Manual de Operações, será contratado como entregável técnico indispensável à sustentação da solução. Esse manual deverá consolidar procedimentos de operação, administração, abertura de chamados, resposta a eventos, monitoramento, gestão de usuários, gestão de cotas, atualização, contingência e demais rotinas necessárias à operação contínua.

**7.10.55.1.** O Manual de Operações deverá refletir as melhores práticas do fabricante e da solução validada, distinguindo procedimentos ordinários da Administração, procedimentos a serem executados pela contratada, procedimentos sujeitos à anuência do fabricante e rotinas que possam impactar garantia, suporte, desempenho ou segurança da plataforma.

**7.10.56.** O Item 28 – Operação Assistida, será contratado na quantidade de 01 (um) serviço. Esse item é necessário para acompanhar a entrada em produção da plataforma, apoiar a estabilização operacional, realizar ajustes finos, corrigir inconsistências, orientar a equipe da Administração e reduzir riscos no período inicial de uso efetivo.

**7.10.56.1.** A operação assistida deverá contar, quando aplicável, com apoio ou escalonamento ao fabricante, especialmente nos casos em que sejam identificados problemas de desempenho, incompatibilidade, comportamento anômalo de software, instabilidade de drivers, falhas de integração, necessidade de ajuste fino ou dúvidas relacionadas às melhores práticas da arquitetura validada.

**7.10.56.2.** Essa participação é necessária para que a estabilização do ambiente não dependa exclusivamente de tentativas locais da contratada, preservando o suporte oficial, reduzindo tempo de

resolução de incidentes e assegurando que os ajustes realizados permaneçam compatíveis com a solução homologada.

**7.10.57.** A operação assistida deverá gerar insumos para decisões futuras sobre expansão, pois permitirá avaliar consumo real de GPU, perfil dos usuários, comportamento das filas de workloads, necessidade de armazenamento, gargalos de rede, uso de memória, padrões de inferência, demanda de treinamento e demais métricas relevantes.

**7.10.58.** O Item 29 – Monitoramento da Solução, será contratado na quantidade de 01 (um) serviço. Esse item deverá apoiar a verificação contínua da disponibilidade, desempenho, segurança, consumo de recursos, eventos e indicadores operacionais, permitindo acompanhamento técnico da plataforma durante o período de estabilização e sustentação.

**7.10.58.1.** O monitoramento da solução deverá ser configurado de forma aderente às recomendações do fabricante e às necessidades da Administração, contemplando métricas de infraestrutura, GPUs, workloads, filas, consumo, disponibilidade, eventos, logs, alertas, tendências de uso e indicadores relevantes para planejamento de capacidade e suporte técnico.

**7.10.58.2.** A validação da estratégia de monitoramento pelo fabricante, quando aplicável, é relevante porque a observabilidade da plataforma será um dos principais meios de identificar problemas, acionar suporte, comprovar desempenho, planejar expansão e preservar disponibilidade da solução.

**7.10.59.** O Item 30 – Encerramento do Projeto, será contratado na quantidade de 01 (um) entregável ou serviço. Esse item visa formalizar a conclusão da implantação, consolidar documentação, registrar pendências, validar entregas, confirmar critérios de aceite, transferir conhecimento e encerrar a fase de projeto de forma organizada.

**7.10.59.1.** O encerramento do projeto deverá consolidar evidências de que a solução foi entregue em conformidade com o escopo turn-key, com os documentos de validação, aceites, evidências de suporte, registros de licenciamento, documentação final, matriz de escalonamento e comprovações de participação ou anuência do fabricante nos entregáveis aplicáveis.

**7.10.60.** O Entregável 31 — Serviço de Suporte Avançado Mensal do Fabricante será contratado pelo período de 12 (doze) meses. Esse entregável deverá assegurar acesso a suporte especializado do fabricante, atualização, orientação técnica, análise de problemas, apoio à resolução de incidentes e sustentação dos componentes críticos da solução.

**7.10.60.1.** A contratação do suporte avançado mensal do fabricante é necessária porque a solução de GPU como serviço será composta por componentes de alta complexidade, integrando servidores de GPU, camada de software, orquestração, aceleração, autosserviço, observabilidade, segurança e interfaces com a infraestrutura institucional. Em ambiente dessa natureza, a resolução de incidentes críticos pode depender de conhecimento especializado do fabricante, acesso a bases técnicas internas, análise de bugs, recomendações de versão, avaliação de compatibilidade e suporte a melhores práticas de operação.

**7.10.60.2.** O suporte avançado do fabricante complementa a garantia técnica do hardware e o suporte ordinário dos softwares, agregando capacidade de análise especializada, priorização de incidentes, orientação preventiva, revisão de ambiente, avaliação de vulnerabilidades e apoio técnico qualificado para manter a solução aderente ao desenho validado ao longo do contrato.

**7.10.60.3.** A contratação desse serviço reduz risco de indisponibilidade prolongada, degradação de desempenho, aplicação inadequada de atualizações, perda de aderência à arquitetura validada e

dificuldades de troubleshooting em incidentes que envolvam múltiplas camadas da solução. Também permite que a equipe do SisTEx amadureça gradualmente sua capacidade de operação, apoiada por especialistas com domínio da solução fornecida.

**7.10.61.** A contratação de 12 (doze) meses de suporte avançado do fabricante é compatível com a estratégia de implantação progressiva, pois garante sustentação suficiente para a fase de adoção e amadurecimento da plataforma, sem comprometer excessivamente o orçamento inicial com um ciclo mais longo de serviços e softwares em um mercado de rápida evolução tecnológica.

**7.10.61.1.** O prazo de 12 (doze) meses também permite avaliar, com base em dados reais, a efetividade da solução, a evolução da demanda, a necessidade de expansão de GPUs, o consumo de armazenamento, a maturidade operacional, o perfil de incidentes e a adequação da stack de software, subsidiando decisões futuras sobre renovação, ampliação ou eventual readequação da sustentação contratual.

**7.10.62.** O Item 32 – Serviço de Suporte Mensal do Fornecedor, será contratado pelo período de 12 (doze) meses. Esse item deverá complementar o suporte do fabricante, assegurando atendimento em português, apoio operacional, acompanhamento da solução, interlocução técnica, triagem de incidentes, orientação à Administração e coordenação com os fabricantes envolvidos.

**7.10.62.1.** O suporte mensal do fornecedor é necessário porque, embora o suporte do fabricante seja essencial para questões especializadas, a Administração também necessita de interlocutor técnico responsável pela visão integrada da solução entregue, capaz de atuar na triagem inicial, correlacionar eventos entre componentes, apoiar a equipe local, acionar fabricantes, acompanhar chamados, orientar procedimentos e manter histórico operacional.

**7.10.62.2.** Esse serviço preserva o princípio turn-key durante a fase de sustentação, uma vez que a contratada permanece responsável por apoiar a Administração na operação da solução integrada, evitando que o CITEx tenha que coordenar isoladamente múltiplos fabricantes, softwares, garantias, chamados e responsabilidades técnicas.

**7.10.62.3.** O suporte mensal do fornecedor também contribui para manter a documentação atualizada, registrar lições aprendidas, apoiar a análise de tendências, orientar boas práticas, acompanhar incidentes recorrentes e garantir que os ajustes operacionais realizados não comprometam a aderência da solução ao desenho validado pelo fabricante.

**7.10.63.** A contratação do suporte mensal do fornecedor justifica-se porque a solução será entregue em regime integrado, envolvendo múltiplos componentes de hardware, software, rede, segurança, observabilidade e serviços. A existência de suporte continuado reduz risco de descontinuidade operacional e evita que a Administração assuma, prematuramente, toda a carga de sustentação de uma plataforma ainda em amadurecimento.

**7.10.63.1.** A coexistência entre suporte avançado do fabricante e suporte mensal do fornecedor não representa sobreposição indevida, mas divisão complementar de responsabilidades. O fabricante atua com profundidade técnica, acesso a engenharia, melhores práticas oficiais, análise de bugs, compatibilidade e suporte especializado; o fornecedor atua como ponto focal operacional, integrador da solução, responsável pela coordenação, triagem, acompanhamento, suporte em português e interface cotidiana com a Administração.

**7.10.63.2.** Essa combinação é particularmente importante na fase inicial do projeto, em que a Administração ainda estará desenvolvendo maturidade operacional sobre GPU como serviço. A contratação conjunta desses serviços reduz risco de interrupções, acelera a resolução de problemas,

melhora a transferência de conhecimento e cria condições para que a equipe do SisTEx assuma progressivamente maior protagonismo na operação.

**7.10.64.** O Item 33 – Serviço Especializado de Treinamento Avançado e Atualização Tecnológica, será contratado na quantidade de 8 (oito) participações, distribuídas ao longo de 24 (vinte e quatro) meses, conforme previsto na especificação técnica. Esse item visa assegurar atualização tecnológica continuada da equipe da Administração em temas de inteligência artificial, GPU como serviço, infraestrutura acelerada, redes, segurança, observabilidade, automação e operação de plataformas corporativas de IA.

**7.10.64.1.** O treinamento avançado deverá estar alinhado às tecnologias efetivamente entregues ou às tecnologias diretamente relacionadas à evolução da solução, preferencialmente em eventos, cursos ou capacitações promovidos por fabricantes, ecossistemas tecnológicos ou entidades reconhecidas, de modo a manter a equipe da Administração atualizada quanto às melhores práticas, arquiteturas de referência, funcionalidades emergentes, riscos tecnológicos e caminhos de expansão.

**7.10.65.** A contratação do treinamento avançado justifica-se pela natureza estratégica e emergente da tecnologia envolvida. A implantação da solução não se limita à entrega de equipamentos e licenças, exigindo apropriação institucional de conhecimento, acompanhamento da evolução do ecossistema tecnológico, formação de multiplicadores e capacidade de tomada de decisão qualificada sobre futuras expansões.

**7.10.66.** A quantidade de bens e serviços definida para esta etapa, portanto, não objetiva esgotar a capacidade futura de processamento de IA do Exército, mas criar a base mínima, completa e operacional para que a Administração possa iniciar o uso institucional de GPU como serviço com segurança, governança, suporte, observabilidade e capacidade de evolução.

**7.10.67.** A composição dos itens preserva a lógica de uma contratação integrada e turn-key, na qual a contratada deverá entregar os componentes de hardware, software e serviços em condições de funcionamento conjunto, com participação dos fabricantes, suporte oficial, documentação, transferência de conhecimento e responsabilidade pela integração dos elementos fornecidos.

**7.10.68.** A opção por adquirir menor quantidade de hardware acelerado, sem eliminar os componentes essenciais de software e serviços, reflete a busca por equilíbrio entre necessidade tecnológica, prudência administrativa, restrição orçamentária e estratégia de crescimento baseada em dados reais de utilização.

**7.10.69.** Dessa forma, a Administração evita dois riscos opostos: de um lado, adquirir apenas equipamentos de GPU sem plataforma de consumo, governança e sustentação; de outro, contratar uma solução superdimensionada, sem métricas consolidadas de uso e sem plena maturidade da demanda institucional.

**7.10.70.** Em síntese, a quantidade de bens e serviços proposta representa a configuração mínima viável para implantação inicial de uma solução completa de GPU como serviço, compatível com a previsão orçamentária disponível, aderente à estratégia de expansão futura e suficiente para permitir que o Exército Brasileiro inicie, de forma segura, governada e mensurável, a disponibilização institucional de capacidade computacional acelerada para inteligência artificial.

**7.10.70.1.** A configuração mínima viável aqui descrita deve ser compreendida em conjunto com as premissas de solução turn-key, validação pelo fabricante, participação efetiva dos fabricantes nos entregáveis críticos, licenciamento empresarial, suporte oficial e comprovação documental da compatibilidade dos componentes ofertados. Esses elementos são indispensáveis para que a

Administração não assuma, inadvertidamente, o risco de integrar por conta própria uma solução de alta complexidade tecnológica.

**7.10.70.2.** Portanto, a economicidade buscada nesta etapa não se confunde com simplificação indevida da solução. A racionalização adotada limita a capacidade inicial de processamento e posterga aquisições cuja demanda ainda não está plenamente caracterizada, como storage externo dedicado de maior porte, mas preserva os elementos essenciais para que a plataforma seja efetivamente entregue como serviço institucional, validado, suportado, expansível e operacionalmente sustentável.

**7.11. Atuação do fabricante**

**7.11.1.** Contextualização e finalidade.

**7.11.1.1.** A presente contratação não se limita ao fornecimento isolado de servidores, aceleradores gráficos, licenças ou serviços de instalação. Seu objetivo consiste na implantação de uma solução corporativa de GPU como serviço, integrada por componentes de hardware, software, virtualização, contêineres, orquestração, aceleração computacional, redes, autosserviço, observabilidade, segurança operacional, suporte, capacitação e serviços especializados.

**7.11.1.2.** A funcionalidade da plataforma depende da operação conjunta e coerente desses elementos, da utilização de versões oficialmente suportadas, da observância das arquiteturas de referência aplicáveis e da existência de responsabilidades claramente definidas para integração, implantação, atualização e sustentação do ambiente.

**7.11.1.3.** Nesse contexto, a participação do fabricante do hardware dos servidores, doravante referido como fabricante principal da solução, constitui mecanismo de mitigação dos riscos de incompatibilidade, indefinição de responsabilidades, utilização de configurações não homologadas, perda de garantia ou suporte, falhas de implementação, baixo desempenho e dependência de customizações desenvolvidas exclusivamente pela contratada ou por terceiros.

**7.11.1.4.** A atuação do fabricante busca assegurar que a solução seja projetada, implantada, validada e sustentada conforme documentação oficial, matrizes de compatibilidade, recomendações técnicas, arquiteturas de referência, procedimentos de segurança e práticas de suporte reconhecidas pelo próprio fabricante.

**7.11.1.5.** A intensidade dessa atuação deverá variar conforme a natureza de cada serviço. Em determinados itens será necessária a participação direta de especialistas do fabricante; em outros, será suficiente a anuência ou validação formal de documentos; em atividades de capacitação, poderá ser exigida a utilização de profissionais certificados; e, nos itens de menor risco técnico, a relação ocorrerá apenas de forma indireta, mediante utilização de documentação, recomendações e bases oficiais.

**7.11.1.6.** A participação do fabricante não afastará nem reduzirá a responsabilidade integral da futura contratada pela execução do objeto. A contratada deverá permanecer como responsável única perante a Administração pela entrega da solução completa, pela coordenação dos fabricantes envolvidos, pelo cumprimento dos prazos, pela qualidade dos serviços e pela correção de falhas de integração ou funcionamento.

**7.11.1.7.** Não deverá ser admitida a transferência para a Administração da responsabilidade de integrar os componentes fornecidos, tampouco a utilização de divergências entre fabricantes, distribuidores, integradores ou desenvolvedores como justificativa para o descumprimento de obrigação contratual.

**7.11.1.8.** Para os itens em que a atuação direta, anuência ou validação do fabricante seja considerada indispensável, as especificações técnicas deverão prever a apresentação, juntamente com a proposta, de documento que identifique a solução ofertada.

**7.11.2.** Item 9 – Serviço de Integração com Equipamentos de Rede de Frontend

**7.11.2.1.** A atuação dos fabricantes nesse item está relacionada à necessidade de assegurar a compatibilidade entre os servidores, os componentes da solução de inteligência artificial e a infraestrutura de rede de Data Center Cisco já existente na Administração.

**7.11.2.2.** Embora os switches de frontend não integrem o fornecimento principal, eles constituem componente indispensável à operação da plataforma. A futura contratada deverá assumir a responsabilidade pela integração da solução com os equipamentos e softwares de rede existentes, preservando o caráter turn-key da contratação.

**7.11.2.3.** O fabricante principal da solução deverá possuir documentação ou matriz de compatibilidade aplicável à integração com o ambiente Cisco. Na ausência de documentação pública suficientemente específica, poderá ser exigida declaração formal do fabricante confirmando que a solução ofertada é compatível e suportada quando integrada aos equipamentos indicados pela Administração.

**7.11.2.4.** Para a análise da proposta, considera-se adequado prever a apresentação da matriz ou documentação pública de compatibilidade, com indicação precisa das páginas ou endereços eletrônicos utilizados. Alternativamente, poderá ser aceita declaração oficial de compatibilidade emitida pelo fabricante.

**7.11.2.5.** Deverá ser apresentada, declaração de que a futura contratada disponibilizará profissional com certificação Cisco Certified Internetwork Expert – Data Center durante as atividades de planejamento, implantação e suporte.

**7.11.3.** Item 10 – Reunião Inicial

**7.11.3.1.** A reunião inicial não exige, como regra, participação presencial ou remota direta do fabricante, uma vez que seu objetivo principal consiste na apresentação das equipes, na estruturação da governança do projeto, no alinhamento dos prazos, na definição dos canais de comunicação e na formalização das responsabilidades.

**7.11.3.2.** A relação com o fabricante ocorrerá indiretamente, por meio das certificações, declarações, matrizes de suporte, compromissos de participação e demais documentos apresentados pela contratada.

**7.11.3.3.** Durante a reunião inicial deverão ser confirmados os itens que contarão com participação direta do fabricante, os entregáveis sujeitos à sua validação, os procedimentos de escalonamento e os profissionais certificados disponibilizados para o projeto.

**7.11.3.4.** Para a análise da proposta, não se mostra necessária declaração autônoma do fabricante exclusivamente para esse item. Será suficiente a apresentação da matriz consolidada de atuação do fabricante e da relação preliminar dos perfis profissionais e certificações que serão utilizados na execução.

**7.11.4.** Item 11 – Gerenciamento de Projetos

**7.11.4.1.** O gerenciamento do projeto será executado pela contratada, não havendo necessidade de que o fabricante assuma a gestão administrativa ou contratual da execução.

**7.11.4.2.** A relação com o fabricante decorre da necessidade de coordenar sua participação nas atividades técnicas, controlar os entregáveis sujeitos à sua validação, acompanhar a disponibilização das licenças, garantias e serviços de suporte e consolidar as evidências mantidas nos portais oficiais.

**7.11.4.3.** O gerenciamento deverá controlar a agenda de participação dos especialistas do fabricante, os prazos de obtenção de anuências, a matriz de responsabilidades, os procedimentos de escalonamento e os registros relativos a números de série, contratos de suporte, chamados e licenciamento.

**7.11.5.** Item 12 – Proposta de Escopo de Trabalho

**7.11.5.1.** A Proposta de Escopo de Trabalho será elaborada pela contratada e deverá consolidar, de maneira clara, o papel dos fabricantes no projeto.

**7.11.5.2.** O documento deverá identificar as atividades executadas exclusivamente pela contratada, aquelas que contarão com especialistas do fabricante, os entregáveis que dependerão de anuência ou validação e as obrigações atendidas por meio dos serviços oficiais de suporte.

**7.11.5.3.** Essa definição prévia é necessária para evitar lacunas de escopo, interpretações divergentes e transferência indevida de responsabilidades durante a execução.

**7.11.6.** Item 13 – Levantamento da Situação Atual

**7.11.6.1.** O levantamento da situação atual será conduzido pela contratada, não havendo necessidade de participação direta do fabricante em todas as atividades.

**7.11.6.2.** Entretanto, considerando a elevada densidade elétrica, térmica e computacional dos servidores de GPU, a contratada deverá consultar ou envolver o fabricante sempre que forem identificadas condições físicas ou lógicas capazes de comprometer a aderência da solução à arquitetura homologada.

**7.11.6.3.** Entre essas condições incluem-se energia, refrigeração, espaço físico, racks, cabeamento, redes, firmware, integração com diretórios, serviços de nomes, sincronização de tempo, certificados digitais, observabilidade e segurança operacional.

**7.11.7.** Item 14 – Tomada de Conhecimento da Documentação Existente

**7.11.7.1.** A tomada de conhecimento da documentação existente será realizada pela contratada, sem necessidade de participação direta do fabricante.

**7.11.7.2.** A relação com os fabricantes ocorrerá quando a documentação institucional contiver configurações, versões, integrações ou restrições que precisem ser confrontadas com matrizes de compatibilidade, recomendações e práticas oficialmente suportadas.

**7.11.7.3.** Caso sejam identificadas divergências entre a documentação existente e os requisitos da arquitetura homologada, a contratada deverá registrá-las e, quando necessário, consultar o fabricante antes de propor alterações que possam afetar garantia, desempenho, segurança ou suporte.

**7.11.8.** Item 15 – Apontamento de Oportunidades de Melhoria

**7.11.8.1.** A atuação do fabricante nesse item será predominantemente indireta. As oportunidades de melhoria deverão ser fundamentadas em normas, documentação oficial, recomendações de

segurança, avisos de produto, arquiteturas de referência e melhores práticas aplicáveis à solução ofertada.

**7.11.8.2.** Não será necessária anuência formal do fabricante para cada recomendação, pois o documento possui natureza consultiva. Entretanto, não poderá ser apresentada como prática do fabricante orientação que não esteja apoiada em documentação verificável ou que dependa de customização sem suporte oficial.

**7.11.8.3.** Quando a recomendação afetar arquitetura, compatibilidade, desempenho, garantia ou condições de suporte, deverá ser prevista consulta ou manifestação do fabricante antes de sua implementação.

**7.11.9.** Item 16 – Definição dos Requisitos da Solução

**7.11.9.1.** Esse item deverá contar com participação direta do fabricante, que apoiará a coleta de informações do ambiente, a identificação dos objetivos de negócio e a transformação das necessidades institucionais em requisitos tecnicamente compatíveis com a solução.

**7.11.9.2.** A participação deverá abranger, no mínimo, os aspectos relacionados ao cluster de gerenciamento, redes de frontend e backend, servidores com GPU, sistemas operacionais, contêineres, drivers, bibliotecas de aceleração, armazenamento, observabilidade, autosserviço e protocolos de acesso.

**7.11.9.3.** Essa atuação é necessária para reduzir o risco de definição de requisitos incompatíveis com a arquitetura ofertada, com os limites de suporte ou com a estratégia de expansão futura.

**7.11.10.** Item 17 – Treinamento Inicial da Solução

**7.11.10.1.** O treinamento inicial poderá ser ministrado por profissionais da contratada, desde que possuam certificações vigentes e compatíveis com as tecnologias abrangidas.

**7.11.10.2.** A exigência de certificação busca assegurar que a apresentação conceitual da arquitetura, dos componentes e das funcionalidades seja realizada por profissionais qualificados e alinhada às práticas oficiais, sem exigir que o instrutor seja necessariamente empregado do fabricante.

**7.11.10.3.** Não será obrigatória assinatura ou anuência do fabricante no material didático, desde que sejam utilizadas referências oficiais e conteúdos tecnicamente compatíveis com a solução.

**7.11.11.** Item 18 – Planejamento de Alto Nível da Solução

**7.11.11.1.** O Planejamento de Alto Nível, ou High Level Design – HLD, deverá contar com participação e anuência formal do fabricante.

**7.11.11.2.** Essa participação é necessária porque o HLD estabelecerá a arquitetura geral, a distribuição dos componentes, as integrações, os domínios de segurança, os requisitos de escalabilidade e as principais decisões que orientarão as etapas subsequentes.

**7.11.11.3.** O fabricante deverá avaliar a aderência do desenho à arquitetura de referência, às matrizes de compatibilidade, aos limites de suporte, aos requisitos de energia e refrigeração, às redes e à estratégia de expansão futura.

**7.11.11.4.** A validação deverá ser formalizada por assinatura no documento, declaração de aprovação que identifique a versão analisada ou outro mecanismo equivalente que permita associar inequivocamente a anuência ao HLD produzido.

**7.11.12.** Item 19 – Proposta Técnica de Implementação

**7.11.12.1.** A Proposta Técnica de Implementação será elaborada pela contratada, mas os procedimentos críticos e o Plano de Testes deverão contar com participação do fabricante.

**7.11.12.2.** Essa participação tem por finalidade assegurar que a sequência de implantação, as versões de firmware e software, os testes de integração, os critérios de desempenho, as dependências e os procedimentos de reversão estejam alinhados às práticas oficialmente suportadas.

**7.11.12.3.** O fabricante deverá apoiar a definição dos testes relacionados aos servidores, GPUs, interfaces de rede, drivers, bibliotecas de aceleração, comunicação entre GPUs, RDMA, RoCE e demais componentes estruturantes.

**7.11.13.** Item 20 – Desenho de Baixo Nível da Solução

**7.11.13.1.** O Desenho de Baixo Nível, ou Low Level Design – LLD, deverá ser elaborado pela contratada e contar com respaldo formal do fabricante.

**7.11.13.2.** O LLD conterá as configurações, versões, parâmetros, políticas, endereçamentos, fluxos de autenticação, segmentações, integrações e demais definições que serão efetivamente aplicadas na implantação.

**7.11.13.3.** O fabricante deverá revisar os elementos estruturantes relacionados a compatibilidade, segurança, desempenho, GPU, rede, orquestração, observabilidade, autosserviço e suporte.

**7.11.13.4.** A validação deverá identificar claramente a versão do documento analisada e poderá ocorrer por assinatura, declaração de aprovação ou mecanismo equivalente.

**7.11.14.** Item 21 – Preparação da Documentação de Instalação

**7.11.14.1.** A documentação de instalação deverá ser preparada pela contratada e contar com respaldo formal do fabricante.

**7.11.14.2.** A participação do fabricante é necessária para assegurar que os procedimentos de instalação, atualização, configuração, teste, reversão e recuperação sejam oficialmente reconhecidos e não dependam de práticas improvisadas ou customizações sem suporte.

**7.11.14.3.** Os procedimentos críticos deverão ser revisados e formalmente validados pelo fabricante antes da implantação.

**7.11.15.** Item 22 – Atualização da Documentação Técnica Existente

**7.11.15.1.** A atualização da documentação técnica existente será executada pela contratada, não sendo obrigatória participação direta ou assinatura do fabricante.

**7.11.15.2.** As informações relativas a versões, matrizes de compatibilidade, configurações suportadas, contratos de suporte e ciclo de vida dos componentes deverão, contudo, ser obtidas de fontes oficiais.

**7.11.15.3.** Sempre que a documentação incorporar decisão anteriormente validada pelo fabricante, deverá ser preservada a referência à correspondente evidência, mantendo-se a rastreabilidade entre o desenho aprovado, a implantação e o estado final registrado.

**7.11.16.** Item 23 – Treinamento Antes da Implementação

**7.11.16.1.** O treinamento anterior à implementação poderá ser ministrado por profissionais da contratada, desde que possuam certificações vigentes do fabricante nas tecnologias abrangidas.

**7.11.16.2.** O conteúdo deverá estar alinhado ao HLD, ao LLD, à Proposta Técnica de Implementação e à documentação de instalação previamente validada.

**7.11.16.3.** Não será obrigatória a presença de empregado do fabricante.

**7.11.17.** Item 24 – Implementação

**7.11.17.1.** A implementação deverá contar, nos pontos críticos definidos no plano de implantação, com a participação de especialistas do fabricante do hardware dos servidores. A comprovação da disponibilidade desses profissionais será exigida antes do início da implementação, permanecendo a contratada integralmente responsável pela execução.

**7.11.17.2.** A exigência decorre da criticidade das atividades de instalação, parametrização, integração, validação e ativação dos servidores com GPU e dos componentes de software, rede, armazenamento e aceleração computacional.

**7.11.17.3.** Os especialistas do fabricante deverão atuar nos pontos críticos definidos no plano de implementação, incluindo a validação de firmware, drivers, CUDA, NCCL, comunicação entre GPUs, RDMA ou RoCE, interfaces de rede, desempenho e integração dos componentes.

**7.11.17.4.** Essa participação deverá assegurar que a implantação seja realizada conforme a arquitetura validada e que a solução permaneça coberta pelo suporte oficial.

**7.11.18.** Item 25 – Treinamento Após a Implementação

**7.11.18.1.** O treinamento posterior à implementação poderá ser ministrado por profissionais da contratada, desde que certificados pelo fabricante e com domínio do ambiente efetivamente implantado.

**7.11.18.2.** O conteúdo deverá abranger a arquitetura final, as configurações aplicadas, os procedimentos operacionais, o monitoramento, o tratamento de incidentes e os fluxos de suporte e escalonamento.

**7.11.18.3.** A participação direta do fabricante como instrutor não será obrigatória.

**7.11.19.** Item 26 – Atualização da Documentação

**7.11.19.1.** A documentação final e o as-built serão produzidos pela contratada, não havendo obrigação de autoria ou assinatura direta pelo fabricante.

**7.11.19.2.** A documentação deverá registrar as validações do fabricante aplicáveis, identificar as alterações autorizadas durante a implantação e demonstrar que o estado final permaneceu aderente ao desenho homologado.

**7.11.19.3.** Alterações estruturantes que não tenham sido previamente submetidas ao processo de validação não poderão ser apresentadas como integrantes da solução aceita.

**7.11.20.** Item 27 – Manual de Operações

**7.11.20.1.** O Manual de Operações, ou Operations Run Book, deverá ser elaborado pela contratada em conjunto com o fabricante.

**7.11.20.2.** A atuação conjunta é necessária para que os procedimentos operacionais, as rotinas de monitoramento, a resposta a eventos, a gestão de recursos e os mecanismos de suporte estejam alinhados ao funcionamento real da solução e às práticas oficialmente reconhecidas.

**7.11.20.3.** O fabricante deverá apoiar a definição dos procedimentos que possam afetar garantia, desempenho, segurança, disponibilidade ou suporte, distinguindo as atividades ordinárias da Administração, as atividades da contratada e aquelas que exigem escalonamento.

**7.11.21.** Item 28 – Operação Assistida

**7.11.21.1.** Durante a operação assistida, a contratada deverá disponibilizar equipe com conhecimento da solução e mobilizar especialistas do fabricante sempre que necessário à estabilização, ao ajuste de desempenho ou à investigação de falhas.

**7.11.21.2.** Nas duas primeiras semanas da operação assistida deverá ser disponibilizado recurso técnico remoto do fabricante para prestar aconselhamento técnico, consultivo e orientativo à equipe do CITEx.

**7.11.21.3.** Essa participação busca reduzir o período de estabilização, permitir o tratamento rápido de incompatibilidades e assegurar que os ajustes realizados permaneçam compatíveis com a arquitetura homologada.

**7.11.22.** Item 29 – Monitoramento da Solução

**7.11.22.1.** A definição, implementação e parametrização do monitoramento serão de responsabilidade da contratada, não havendo exigência de participação direta e permanente do fabricante.

**7.11.22.2.** Os indicadores, limiares, alertas, eventos e mecanismos de coleta deverão observar a documentação oficial e as capacidades suportadas pelos componentes fornecidos.

**7.11.22.3.** Quando determinado parâmetro depender de informação não pública ou de recomendação específica, a contratada deverá acionar o fabricante pelos canais de suporte contratados.

**7.11.23.** Item 30 – Encerramento do Projeto

**7.11.23.1.** No encerramento do projeto, a contratada deverá atuar conjuntamente com o fabricante na realização do workshop de transferência de conhecimento, com duração mínima de oito horas.

**7.11.23.2.** A participação do fabricante tem por objetivo consolidar o entendimento das tecnologias implantadas, apresentar recomendações finais, esclarecer os processos de suporte e transferir os conhecimentos necessários à operação contínua da plataforma.

**7.11.23.3.** O fabricante deverá disponibilizar especialistas com conhecimento dos produtos e da arquitetura efetivamente implementada, evitando que o evento se limite a apresentação comercial ou genérica.

**7.11.24.** Item 31 – Serviço de Suporte Avançado Mensal do Fabricante

**7.11.24.1.** Esse item apresenta o vínculo mais direto com o fabricante e deverá ser prestado pelo próprio fabricante, por estrutura de atendimento oficialmente constituída e controlada por ele.

**7.11.24.2.** O serviço deverá complementar a garantia técnica e o suporte ordinário, abrangendo a solução completa, inclusive componentes próprios e componentes de terceiros homologados ou integrados à arquitetura suportada.

**7.11.24.3.** A atuação deverá contemplar gerenciamento operacional, priorização e acompanhamento de chamados, coordenação de especialistas, gestão de inventário, acompanhamento do ciclo de vida, análise de falhas de engenharia, avisos de produto, vulnerabilidades, portais, apoio a RMA, emissão de relatórios e sessões de orientação.

**7.11.24.4.** Deverá ser disponibilizado engenheiro de suporte especializado, vinculado ao fabricante, com certificação oficial vigente e acesso às ferramentas internas, bases técnicas, equipes de engenharia e demais recursos necessários ao tratamento de casos complexos.

**7.11.25.** Item 32 – Serviço de Suporte Mensal da Contratada

**7.11.25.1.** O suporte mensal será prestado pela contratada, mas deverá possuir integração operacional com o suporte do fabricante.

**7.11.25.2.** Caberá à contratada realizar o primeiro atendimento, diagnosticar as ocorrências, abrir chamados junto ao fabricante, acompanhar as escalações, fornecer as informações técnicas necessárias e permanecer responsável pelo incidente até sua efetiva solução.

**7.11.25.3.** A existência de chamado aberto no fabricante não poderá ser utilizada para transferir responsabilidade ou encerrar prematuramente o atendimento da contratada.

**7.11.25.4.** A contratada deverá manter disponível profissional especialista com certificação oficial vigente, emitida ou reconhecida pelo fabricante, compatível com a solução ofertada e com as atividades de suporte de maior complexidade. Será aceita certificação sucessora ou de nível superior reconhecida pelo respectivo programa de certificação.

**7.11.26.** Item 33 – Serviço Especializado de Treinamento Avançado e Atualização Tecnológica

**7.11.26.1.** A relação com os fabricantes nesse item decorre da participação dos profissionais da Administração em eventos técnicos oficiais ou reconhecidos pelo ecossistema tecnológico da solução.

**7.11.26.2.** A previsão inclui participação no NVIDIA GTC e em evento técnico internacional de grande porte promovido pelo fabricante do hardware ou por ecossistema tecnológico de relevância equivalente.

**7.11.26.3.** Essas atividades têm por finalidade proporcionar acesso às atualizações de arquitetura, novas funcionalidades, laboratórios, práticas de operação, segurança, observabilidade, redes de alto desempenho, treinamento, inferência e demais temas necessários à evolução da plataforma.

**7.12.** A Tabela 1 apresenta, de forma resumida, os itens a serem adquiridos nessa contratação. Após a definição da melhor solução de TIC a ser contratada, os itens da tabela serão mais detalhados nas Especificações Técnicas do Itens, Anexo I a este estudo.

Tabela 1: Resumo dos itens.

| Item | Descrição | Qtd |
|---|---|---|
| 1 | SERVIDOR DE PROCESSAMENTO COM GPU – Componente de Hardware | 1 |
| 2 | SERVIDOR DE PROCESSAMENTO COM GPU – Componente de Garantia e Licenciamento | 1 |
| 3 | SERVIDOR PARA HOSPEDAGEM DA CAMADA DE SOFTWARE DA SOLUÇÃO – Componente de Hardware | 1 |
| 4 | SERVIDOR PARA HOSPEDAGEM DA CAMADA DE SOFTWARE DA SOLUÇÃO – Componente de Garantia e Licenciamento | 1 |
| 5 | SOLUÇÃO DE SOFTWARE DE VIRTUALIZAÇÃO, CONTÊINERES E ORQUESTRAÇÃO | 1 |
| 6 | SOLUÇÃO DE SOFTWARE ESPECÍFICA PARA ACELERAÇÃO POR GPU | 1 |
| 7 | SOLUÇÃO DE SOFTWARE PARA INTERFACE DE SERVIÇOS E AUTOSSERVIÇO | 1 |
| 8 | SOLUÇÃO DE SOFTWARE DE OBSERVABILIDADE E SEGURANÇA OPERACIONAL | 1 |
| 9 | SERVIÇO DE INTEGRAÇÃO COM EQUIPAMENTOS DE REDE DE FRONTEND | 1 |
| 10 | REUNIÃO INICIAL | 1 |
| 11 | GERENCIAMENTO DE PROJETOS | 1 |
| 12 | PROPOSTA DE ESCOPO DE TRABALHO | 1 |
| 13 | LEVANTAMENTO DA SITUAÇÃO ATUAL | 1 |
| 14 | TOMADA DE CONHECIMENTO DA DOCUMENTAÇÃO EXISTENTE | 1 |
| 15 | APONTAMENTO DE OPORTUNIDADES DE MELHORIA | 1 |
| 16 | DEFINIÇÃO DE REQUISITOS DA SOLUÇÃO | 1 |
| 17 | TREINAMENTO INICIAL DA SOLUÇÃO | 1 |
| 18 | PLANEJAMENTO DE ALTO NÍVEL DA SOLUÇÃO | 1 |
| 19 | PROPOSTA TÉCNICA DE IMPLEMENTAÇÃO | 1 |
| 20 | DESENHO DE BAIXO NÍVEL DA SOLUÇÃO | 1 |

| | | |
|---|---|---|
| 21 | PREPARAÇÃO DA DOCUMENTAÇÃO DE INSTALAÇÃO | 1 |
| 22 | ATUALIZAÇÃO DA DOCUMENTAÇÃO TÉCNICA EXISTENTE | 1 |
| 23 | TREINAMENTO ANTES DA IMPLEMENTAÇÃO | 1 |
| 24 | IMPLEMENTAÇÃO | 1 |
| 25 | TREINAMENTO APÓS A IMPLEMENTAÇÃO | 1 |
| 26 | ATUALIZAÇÃO DA DOCUMENTAÇÃO | 1 |
| 27 | MANUAL DE OPERAÇÕES | 1 |
| 28 | OPERAÇÃO ASSISTIDA | 1 |
| 29 | MONITORAMENTO DA SOLUÇÃO | 1 |
| 30 | ENCERRAMENTO DO PROJETO | 1 |
| 31 | SERVIÇO DE SUPORTE AVANÇADO MENSAL DO FABRICANTE | 36 |
| 32 | SERVIÇO DE SUPORTE MENSAL DO FORNECEDOR | 36 |
| 33 | SERVIÇO ESPECIALIZADO DE TREINAMENTO AVANÇADO E ATUALIZAÇÃO TECNOLÓGICA | 24 |

## 8. Levantamento de soluções

**8.1.** Em atendimento ao Inciso II do Art. 11º Instrução Normativa SGD/ME nº 94, de 23 de dezembro de 2022, que trata de "análise comparativa das soluções de TIC", para a solução pesquisada levou-se em consideração os seguintes aspectos, conforme detalhado a seguir.

**8.1.1.** Sobre necessidades similares em outros órgãos ou entidades da Administração Pública e as soluções adotadas, em atendimento ao disposto no art. 10 do Decreto nº 11.462, de 31 MAR 23, considerando a especificidade técnica da contratação, não foram encontradas outras contratações equivaletes por meio de consultas no site https://www2.comprasnet.gov.br/siasgnet-irp /consultarIRPComprasNet.do?method=iniciar.

**8.1.2.** Diante do exposto, considerando a grande especificidade técnica da solução e as informações disponíveis para consulta de outros processos de contratação, atendendo ao disposto no art. 10 do Decreto nº 11.462, de 31 MAR 23, não foi identificado pelo CITEx nenhuma IRP em andamento ou outro órgão da administração com características técnicas que seja compatível com a demanda do SisTEx, dessa forma a equipe de planejamento considera que não é conveniente a participação em outro processo licitatório.

**8.1.3.** Sobre as alternativas do mercado, foram feitas consultas aos principais fabricnates de Hardware, conforme já descrito neste estudo.

**8.1.4.** Sobre a existência de *softwares* disponíveis conforme descrito na Portaria STI/MP nº 46, de 28 de setembro de 2016, e suas atualizações, a solução não é adequada a ser atendida com um único *software*, por se tratar de uma solução completa composta por *hardware, software* e serviços.

**8.1.5.** Sobre as políticas, os modelos e os padrões de governo, a exemplo dos Padrões de Interoperabilidade de Governo Eletrônico - ePing, Modelo de Acessibilidade em Governo Eletrônico - eMag, Padrões *Web* em Governo Eletrônico - ePwg, padrões de *Design System* de governo, Infraestrutura de Chaves Públicas Brasileira - ICP-Brasil e Modelo de Requisitos para Sistemas Informatizados de Gestão Arquivística de Documentos - e-ARQ Brasil, as mesmas foram consideradas quando aplicáveis.

**8.1.5.1.** Em relação aos Padrões de Interoperabilidade de Governo Eletrônico – ePing, a presente contratação levou em consideração as Especificações Técnicas dos Componentes do ePing, contidos na Parte II da arquitetura disponível em https://eping.governoeletronico.gov.br/, se atendo principalmente aos itens de Rede / Transporte, Comunicação de dados, Criptografia, Serviços de Redes, Resposta a Incidentes de Segurança da Informação.

**8.1.5.2.** Ainda sobre o ePing, a presente contratação atende às indicações de Padrões de Interoperabilidade de Governo Eletrônico – ePing, nas modalidades de Adotado e Recomendado sempre que adequado ao escopo da contratação.

**8.1.6.** Sobre as necessidades de adequação do ambiente do órgão ou entidade para viabilizar a execução contratual, não serão necessárias adequações dos ambientes.

**8.1.7.** Sobre os diferentes modelos de prestação do serviço ou a contratação como seriço, será apresentada uma análise como alternativa da contratação de serviços de nuvem pública para fins de análise.

**8.1.8.** Sobre os diferentes tipos de soluções em termos de especificação, composição ou características dos bens e serviços integrantes, a aquisição trata de uma solução completa, composta por várias partes e tecnologias, todas disponíveis no mercado e por mais de um fabricante.

**8.1.9.** Sobre a ampliação ou substituição da solução implantada, o ambiente atual de GPU é heterogêneo, com falta de capacidade e não padronizado, adicionalmente os equipamentos estão alocados para implementação de projetos específicos e não estão disponíveis para todo o Exército, sendo adequada a aquisição de nova solução, em detrimento da ampliação.

**8.1.10** Sobre as diferentes métricas de prestação do serviço e de pagamento, a solução não permite muitas variações, sendo as soluções possíveis analisadas nesse estudo.

**8.2. Identificação das Soluções**

**8.2.1.** Solução 1: Aquisição de Hardware Software e Serviços para disponibilização de uma soluão de GPU como Serviço para o Exército.

**8.2.2.** Solução 2: Contratação de solução de processamento de IA em nuvem pública.

**8.3. Análise de solução alternativa baseada em nuvem pública**

**8.3.1.** Para fins de avaliação de alternativa de contratação, foi realizada análise preliminar de contratação de capacidade equivalente em nuvem pública, considerando como referência a capacidade computacional inicial pretendida para a solução local, composta por 01 (um) servidor com 08 (oito) GPUs NVIDIA B300.

**8.3.2.** A análise tem por objetivo verificar, em termos comparativos, o custo de contratação de capacidade computacional equivalente em provedores de nuvem pública, quando houver preço público disponível, ou, alternativamente, estimar o volume de processamento em tokens que a solução poderia disponibilizar e comparar esse volume com serviços de IA generativa precificados por token.

**8.3.3.** Para padronização da análise de custo total da alternativa em nuvem pública, adotou-se o período de 12 (doze) meses, em regime de uso contínuo 24x7, totalizando 8.760 horas no período, correspondentes a 730 horas mensais multiplicadas por 12 meses.

**8.3.4.** Para conversão dos valores originalmente expressos em dólar norte-americano, adotou-se a cotação referencial de US$ 1,00 = R$ 5,1664, correspondente à taxa de câmbio dólar comercial para compra — média, referente a 06/07/2026, conforme série do Ipeadata com fonte no Banco Central do Brasil, disponível no endereço: https://www.ipeadata.gov.br/ExibeSerie.aspx?module=M&serid=38590.

**8.3.5. Alternativa 1 — Oracle Cloud Infrastructure – OCI**

**8.3.5.1.** Para a alternativa Oracle Cloud Infrastructure – OCI, foi identificada oferta diretamente comparável à capacidade inicial pretendida para a solução local, por meio do shape BM.GPU.B300.8. A documentação pública da Oracle informa que esse shape possui 8 GPUs NVIDIA B300 Tensor Core, com 263 GB de memória por GPU, além de 2 CPUs Intel Emerald Rapids 8592+ de 64 núcleos. Assim, a equivalência adotada é direta: a solução local de referência, composta por 1 servidor com 8 GPUs NVIDIA B300, foi comparada a 1 instância OCI BM.GPU.B300.8, também composta por 8 GPUs NVIDIA B300. A documentação técnica da instância está disponível no endereço: https://docs.oracle.com/en-us/iaas/Content/Compute/References/computeshapes.htm.

**8.3.5.2.** A lista pública de preços da OCI Brasil apresenta o shape BM.GPU.B300.8 como instância com 8x NVIDIA B300 263 GB, arquitetura Blackwell e rede 8x800 Gb/seg. O preço público identificado na lista brasileira foi de R$ 78,5355 por GPU/hora. Considerando 8 GPUs, o custo horário da instância completa é de R$ 628,284 por hora. A lista pública de preços da OCI Brasil está disponível no endereço: https://www.oracle.com/br/cloud/price-list/.

**8.3.5.3.** A precificação utilizada pode ser conferida na página pública Lista de Preços OCI Brasil – Oracle Cloud Infrastructure, disponível no endereço: https://www.oracle.com/br/cloud/price-list/, na qual deve ser pesquisado o termo BM.GPU.B300.8. A caracterização técnica da instância pode ser conferida na página pública Compute Shapes – Oracle Cloud Infrastructure Documentation, disponível no endereço: https://docs.oracle.com/en-us/iaas/Content/Compute/References/computeshapes.htm, também mediante pesquisa pelo termo BM.GPU.B300.8.

**8.3.5.4.** Como referência adicional, a lista global de preços da Oracle também apresenta o shape BM.GPU.B300.8 com preço de US$ 15,00 por GPU/hora. Nessa referência internacional, a instância completa com 8 GPUs teria custo de US$ 120,00 por hora, resultando em US$ 1.051.200,00 em 12 meses de uso contínuo. Convertido pela cotação referencial adotada de R$ 5,1664 por dólar, esse valor corresponderia a aproximadamente R$ 5.430.919,68. Para fins desta análise, contudo, foi adotado como referência principal o valor da lista pública da OCI Brasil, por estar expresso diretamente

em reais. A lista global de preços da Oracle está disponível no endereço: https://www.oracle.com/cloud/price-list/.

**8.3.5.5.** Valor convertido em reais para a solução por 12 meses. Com base na lista pública da OCI Brasil, a alternativa OCI teria custo de R$ 628,284 por hora para a instância completa com 8 GPUs. Considerando uso contínuo por 8.760 horas, o custo aproximado da alternativa OCI para 12 meses é de R$ 5.503.767,84, considerando exclusivamente a capacidade bruta de processamento equivalente a 8 GPUs NVIDIA B300.

**8.3.5.6.** Esse valor não inclui custos adicionais de armazenamento persistente, backup, transferência de dados, tráfego de saída, conectividade dedicada, suporte avançado, licenças de softwares de IA, observabilidade, segurança, operação assistida, serviços profissionais, retenção de logs, integração com diretórios institucionais, adequações de soberania de dados ou eventuais encargos tributários, cambiais e contratuais.

**8.3.6. Alternativa 2 — Amazon Web Services – AWS EC2 P6-B300**

**8.3.6.1.** Para a alternativa Amazon Web Services – AWS, foi identificada a família de instâncias Amazon EC2 P6-B300, acelerada por GPUs NVIDIA Blackwell Ultra B300. A documentação pública da AWS informa que as instâncias P6-B300 fornecem 8 GPUs NVIDIA Blackwell Ultra, 2,1 TB de memória HBM3e, 6,4 Tbps de rede EFA, 300 Gbps de throughput ENA dedicado e 4 TB de memória do sistema, sendo voltadas a treinamento e inferência de IA em larga escala. Assim, a equivalência adotada também é direta: a solução local de referência, composta por 1 servidor com 8 GPUs NVIDIA B300, foi comparada a 1 instância AWS p6-b300.48xlarge, composta por 8 GPUs NVIDIA Blackwell Ultra B300. A documentação da família de instâncias P6 está disponível no endereço: https://aws.amazon.com/ec2/instance-types/p6/.

**8.3.6.2.** A tabela de detalhes do produto da AWS indica que a instância p6-b300.48xlarge possui 8 GPUs Blackwell Ultra, 2.144 GB de memória de GPU, 192 vCPUs, 4.096 GiB de memória de sistema, 8 discos locais de 3,84 TB e 6,4 Tbps de largura de banda de rede. A página técnica da AWS sobre as instâncias P6 pode ser consultada no endereço: https://aws.amazon.com/ec2/instance-types/p6/.

**8.3.6.3.** A precificação utilizada foi obtida na página pública Amazon EC2 Capacity Blocks for ML Pricing, disponível no endereço: https://aws.amazon.com/ec2/capacityblocks/pricing/, na qual a AWS informa que, a partir de 1º de julho de 2026, a taxa horária por acelerador para P6-B300 será de US$ 14,04 nas regiões disponíveis, exceto AWS GovCloud. As características técnicas da instância foram verificadas na página pública Amazon EC2 P6e UltraServers and P6 instances, disponível no endereço: https://aws.amazon.com/ec2/instance-types/p6/, que descreve as instâncias P6-B300 e sua composição de GPUs, memória, rede e armazenamento local.

**8.3.6.4.** Para a estimativa de custo, considerou-se o valor de US$ 14,04 por acelerador/hora. Como a instância possui 8 aceleradores B300, o custo horário estimado da instância completa é de US$ 112,32 por hora. Para o período de 12 meses em uso contínuo, correspondente a 8.760 horas, o custo total estimado da capacidade computacional é de US$ 983.923,20.

**8.3.6.5.** Valor convertido em reais para a solução por 12 meses. Aplicando-se a cotação referencial de US$ 1,00 = R$ 5,1664, o custo de US$ 983.923,20 corresponde a aproximadamente R$ 5.083.340,82. Assim, o custo aproximado da alternativa AWS P6-B300 para 12 meses é de R$ 5,08 milhões, considerando exclusivamente a capacidade bruta de processamento equivalente a 8 GPUs B300.

**8.3.6.6.** Esse valor não inclui custos de sistema operacional eventualmente aplicáveis, armazenamento compartilhado, storage de alta performance, tráfego de rede, transferência de dados, conectividade

dedicada, suporte, segurança, licenças de software, operação assistida, observabilidade, backup, retenção de logs, serviços profissionais ou integração com ambientes institucionais.

### 8.3.7. Alternativa 3 — Google Cloud/Gemini API com estimativa por tokens

**8.3.7.1.** Para a alternativa Google, foi analisada uma abordagem baseada em consumo de IA generativa por tokens, utilizando preços públicos da Gemini API/Google AI. Essa alternativa não representa contratação de infraestrutura bruta de GPU equivalente, mas foi utilizada como referência econômica complementar, tendo em vista que a precificação pública direta por GPU B300/GB300 não se mostrou disponível, de forma simples e diretamente comparável, para contratação equivalente em Google Cloud.

**8.3.7.2.** A documentação pública do Google Cloud informa que a família A4X Max utiliza NVIDIA GB300 Grace Blackwell Ultra Superchips, enquanto a família A4X utiliza NVIDIA GB200 Grace Blackwell Superchips. A documentação também informa que a criação de instâncias A4X Max exige reserva de capacidade e que cada máquina a4x-maxgpu-4g-metal possui 4 GPUs, 1.116 GB de memória HBM3e de GPU, 144 vCPUs, 960 GB de memória de instância, 12.000 GiB de SSD local, 6 NICs físicos e 3.600 Gbps de largura de banda máxima de rede. A documentação dos tipos de máquina com GPU do Google Cloud está disponível no endereço: https://cloud.google.com/compute/docs/gpus.

**8.3.7.3.** Como a oferta Google analisada não foi precificada por GPU de forma equivalente à OCI e à AWS, adotou-se uma estimativa por volume de tokens. Para isso, partiu-se da referência pública de que um sistema NVIDIA DGX H100 com 8 GPUs alcançou 24.525 tokens por segundo, e aplicou-se o fator de desempenho 3,9 vezes superior indicado em referência pública de desempenho para infraestrutura Azure ND GB200 v6 em relação à geração anterior. Com base nessa aproximação, a capacidade equivalente estimada para um conjunto de 8 GPUs Blackwell foi calculada em 95.647,5 tokens por segundo. Considerando uso contínuo por 12 meses, ou 8.760 horas, o volume total estimado é de aproximadamente 3,016 trilhões de tokens. A referência de desempenho da infraestrutura Azure ND GB200 v6 está disponível no endereço: https://techcommunity.microsoft.com/blog/azurehighperformancecomputingblog/azure%E2%80%99s-nd-gb200-v6-delivers-record-performance-for-inference-workloads/4399253.

**8.3.7.4.** A precificação utilizada para a alternativa por tokens foi obtida na página pública Gemini Developer API Pricing, disponível no endereço: https://ai.google.dev/gemini-api/docs/pricing, que apresenta os valores por 1 milhão de tokens para modelos Gemini. Para o modelo Gemini 2.5 Flash, a tabela pública informa preço de US$ 0,30 por 1 milhão de tokens de entrada e US$ 2,50 por 1 milhão de tokens de saída, incluindo tokens de raciocínio no preço de saída. A caracterização técnica das famílias A4X Max e A4X pode ser conferida na página pública GPU machine types – Google Cloud Compute Engine Documentation, disponível no endereço: https://cloud.google.com/compute/docs/gpus.

**8.3.7.5.** Considerando apenas a capacidade de geração de tokens de saída, o volume estimado de 3,016 trilhões de tokens corresponde a 3.016.339,56 milhões de tokens. Aplicando o preço de US$ 2,50 por 1 milhão de tokens de saída do Gemini 2.5 Flash, o custo estimado para consumo equivalente é de aproximadamente US$ 7.540.848,90 em 12 meses.

**8.3.7.6.** Valor convertido em reais para a solução por 12 meses. Aplicando-se a cotação referencial de US$ 1,00 = R$ 5,1664, o custo de US$ 7.540.848,90 corresponde a aproximadamente R$ 38.959.041,76. Nessa hipótese, o custo aproximado da alternativa Google/Gemini para 12 meses é de R$ 38,96 milhões, considerando consumo equivalente integralmente precificado como tokens de saída.

**8.3.7.7.** Em uma segunda hipótese simplificada, considerando composição média de 50% de tokens de entrada e 50% de tokens de saída, o custo médio ponderado seria de US$ 1,40 por 1 milhão de tokens, correspondente à média entre US$ 0,30 e US$ 2,50. Nessa hipótese, o custo estimado para o mesmo volume total de tokens seria de aproximadamente US$ 4.222.875,38 em 12 meses. Convertido pela cotação referencial adotada de R$ 5,1664 por dólar, esse valor corresponde a aproximadamente R$ 21.817.063,38. Assim, mesmo na hipótese simplificada de composição média entre entrada e saída, o custo aproximado da alternativa Google/Gemini para 12 meses é de R$ 21,82 milhões.

**8.3.7.8.** A análise por tokens deve ser interpretada com cautela. O consumo de API de IA generativa não é funcionalmente equivalente à contratação de infraestrutura própria de GPU como serviço, pois não entrega, por si só, ambiente soberano de execução, treinamento, ajuste fino, controle integral de modelos, isolamento completo de dados, governança local, observabilidade institucional, operação air-gapped, integração com diretórios internos, nem capacidade de execução de múltiplos frameworks e workloads próprios.

**8.3.8. Racional de valor de referência da alternativa equivalente em nuvem pública**

**8.3.8.1.** Para fins de consolidação da análise econômica da alternativa de contratação em nuvem pública, adotou-se como valor de referência a média aritmética simples dos custos estimados das três alternativas analisadas: Oracle Cloud Infrastructure – OCI, Amazon Web Services – AWS e Google Cloud/Gemini API.

**8.3.8.2.** A utilização da média das três alternativas tem por objetivo estabelecer um valor referencial aproximado para a contratação de capacidade equivalente em nuvem pública pelo período de 12 meses, em regime de uso contínuo 24x7, sem vincular a análise a um único provedor e sem assumir que a Administração necessariamente contrataria a alternativa mais barata ou mais cara disponível.

**8.3.8.3.** Para a alternativa OCI BM.GPU.B300.8, foi considerado o custo aproximado de R$ 5.503.767,84 em 12 meses, com base na lista pública de preços da OCI Brasil para instância equivalente a 8 GPUs NVIDIA B300.

**8.3.8.4.** Para a alternativa AWS EC2 P6-B300, foi considerado o custo aproximado de R$ 5.083.340,82 em 12 meses, obtido a partir do preço público de US$ 14,04 por acelerador/hora, aplicado a 8 aceleradores B300 e convertido para reais pela cotação referencial de US$ 1,00 = R$ 5,1664.

**8.3.8.5.** Para a alternativa Google Cloud/Gemini API, a análise por tokens resultou em uma faixa estimada entre R$ 21.817.063,38 e R$ 38.959.041,76 em 12 meses, conforme a hipótese de composição entre tokens de entrada e saída. Para fins de obtenção de um valor único a ser utilizado na média das três alternativas, adotou-se o ponto médio dessa faixa, correspondente a R$ 30.388.052,57.

**8.3.8.6.** Assim, considerando os valores de R$ 5.503.767,84 para OCI, R$ 5.083.340,82 para AWS e R$ 30.388.052,57 para Google/Gemini, a média aritmética simples das três alternativas resulta em R$ 13.658.387,08.

**8.3.8.7.** Dessa forma, para fins de referência econômica da alternativa equivalente em nuvem pública, o valor aproximado para contratação por 12 meses pode ser estimado em R$ 13,66 milhões, considerando a média das três opções avaliadas.

**8.3.8.8.** Como análise de sensibilidade, caso se adotasse para a alternativa Google/Gemini a hipótese inferior da faixa, correspondente à composição média de 50% de tokens de entrada e 50% de tokens de saída, o valor médio das três alternativas seria de aproximadamente R$ 10.801.390,68. Caso se

adotasse a hipótese superior, considerando todo o volume como tokens de saída, o valor médio das três alternativas seria de aproximadamente R$ 16.515.383,47.

**8.3.8.9.** Assim, a média central de R$ 13,66 milhões deve ser compreendida como valor referencial aproximado para a alternativa equivalente em nuvem pública por 12 meses, enquanto a faixa de sensibilidade entre R$ 10,80 milhões e R$ 16,52 milhões demonstra a variação possível conforme o modelo de consumo adotado, especialmente na alternativa baseada em tokens.

**8.3.8.10.** Esse valor de referência não contempla todos os custos acessórios necessários à disponibilização de uma solução completa de GPU como serviço em nuvem pública, tais como armazenamento persistente, backup, tráfego de saída, conectividade dedicada, suporte avançado, licenças corporativas, observabilidade, segurança, operação assistida, serviços profissionais, retenção de logs, integração com diretórios institucionais, controles de soberania, auditoria, customização, governança de consumo e eventuais encargos tributários, cambiais e contratuais.

**8.3.8.11.** Dessa forma, o valor médio de R$ 13,66 milhões representa apenas uma referência econômica aproximada da capacidade equivalente em nuvem pública para 12 meses, não devendo ser interpretado como custo total final de contratação de uma solução completa. Ainda assim, a estimativa permite demonstrar que a alternativa em nuvem pública, em cenário de uso contínuo, apresenta custo recorrente relevante e deve ser analisada com cautela quando comparada à implantação de capacidade própria de GPU como serviço sob governança institucional.

**8.3.8.12.** Registra-se, ainda, que os custos médios estimados podem variar de forma considerável caso sejam considerados outros componentes de hardware, software e serviços normalmente associados à disponibilização de capacidade equivalente em nuvem pública. A depender do modelo de contratação e da arquitetura adotada, a alternativa em nuvem poderia demandar recursos adicionais de armazenamento de alto desempenho, armazenamento persistente, backup, snapshots, transferência de dados, tráfego de saída, conectividade dedicada, suporte premium, ferramentas de segurança, observabilidade, gestão de identidades, retenção de logs, auditoria, governança de consumo, serviços profissionais, operação assistida e integração com o ambiente institucional.

**8.3.8.13.** Também deve ser considerado que as ofertas de nuvem pública podem apresentar diferentes modelos de cobrança, variando conforme região, disponibilidade de capacidade, reserva prévia, compromisso de uso, contratação sob demanda, descontos por volume, encargos de suporte, custos de transferência de dados e serviços complementares associados. Assim, embora a média calculada seja adequada como referência preliminar, ela não elimina a possibilidade de variações relevantes no custo final da alternativa em nuvem pública.

**8.3.8.14.** Uma alternativa adicional seria a contratação da capacidade em nuvem pública em regime estritamente sob demanda, com pagamento conforme a utilização efetiva de usuários, projetos, horas de GPU, volume de tokens, armazenamento, tráfego e serviços associados. Em tese, esse modelo poderia reduzir custos em cenários de uso eventual, baixa utilização ou demandas pontuais, ao evitar a manutenção contínua de capacidade computacional dedicada.

**8.3.8.15.** Contudo, no presente momento, não é possível quantificar de forma suficientemente confiável a alternativa de consumo sob demanda, pois as demandas institucionais relacionadas ao uso de inteligência artificial ainda se encontram em fase de crescimento, descoberta e maturação. O perfil de consumo possui componente variável relevante, podendo se alterar significativamente conforme o surgimento de novos projetos, modelos, bases documentais, usuários, políticas de uso, necessidades de treinamento, inferência, ajuste fino, experimentação e execução de aplicações institucionais.

**8.3.8.16.** Para estimar adequadamente uma alternativa sob demanda, seriam necessárias informações mais maduras sobre quantidade de usuários, perfis de workload, simultaneidade, horas efetivas de GPU, volume de tokens de entrada e saída, sazonalidade de uso, tipos de modelos empregados, volume de armazenamento associado, tráfego de rede, retenção de dados, requisitos de suporte e padrões de expansão. Tais informações ainda não estão disponíveis com precisão suficiente para fundamentar estimativa robusta nesta fase do planejamento.

**8.3.8.17.** Por essa razão, para os fins deste Estudo Técnico Preliminar, adotou-se o cenário de uso contínuo por 12 meses como metodologia objetiva, conservadora e comparável entre as alternativas avaliadas. Embora esse cenário não represente necessariamente o padrão exato de consumo futuro, ele permite estabelecer uma ordem de grandeza econômica para a contratação de capacidade equivalente em nuvem pública e compará-la, de forma consistente, com a estratégia de implantação de capacidade própria de GPU como serviço.

**8.3.8.18.** Dessa forma, a estimativa realizada é considerada adequada para a análise a ser desenvolvida neste Estudo Técnico Preliminar, pois utiliza referências públicas de preço, critérios explícitos de equivalência, período uniforme de comparação, cotação cambial definida e premissas transparentes quanto ao escopo incluído e excluído da avaliação. A metodologia adotada não pretende esgotar todas as possibilidades de contratação em nuvem pública, mas fornecer parâmetro técnico-econômico suficiente para subsidiar a tomada de decisão na fase de planejamento.

**8.3.8.19.** Em síntese, o valor médio de R$ 13,66 milhões deve ser interpretado como referência aproximada da alternativa equivalente em nuvem pública para 12 meses, sujeita a variações conforme o modelo de contratação, os serviços adicionais incluídos e o perfil real de consumo. Ainda assim, a estimativa é suficiente para demonstrar, no âmbito deste Estudo Técnico Preliminar, que a alternativa em nuvem pública, especialmente em cenário de uso contínuo, apresenta custo recorrente relevante e deve ser analisada com cautela frente à implantação de capacidade própria, integrada e sob governança institucional.

**8.3.8.20.** Para ampliar a comparabilidade entre a alternativa em nuvem pública e a aquisição de infraestrutura própria, deve-se considerar que o servidor de processamento com GPU a ser adquirido permanecerá disponível para utilização pela Administração durante o período total de 60 (sessenta) meses, correspondente ao prazo de garantia e suporte previsto para o componente de hardware. Embora os licenciamentos de software e determinados serviços tenham vigência inicial de 12 (doze) meses, a capacidade computacional instalada continuará integrando o patrimônio e permanecendo à disposição do Exército durante os 48 (quarenta e oito) meses subsequentes.

**8.3.8.21.** Diferentemente da aquisição do hardware, na qual o desembolso inicial assegura a disponibilidade do equipamento durante sua vida útil e o período de garantia contratado, a alternativa em nuvem pública exige pagamentos recorrentes enquanto a capacidade computacional permanecer disponível. Por essa razão, para estimar o valor equivalente da disponibilidade da capacidade em nuvem durante os mesmos 60 (sessenta) meses, projetou-se o valor anual de referência obtido para os 48 (quarenta e oito) meses adicionais.

**8.3.8.22.** A projeção foi realizada de forma linear, mantendo-se as mesmas condições de preço, taxa de câmbio, regime de utilização contínua 24x7 e composição das alternativas consideradas na estimativa de 12 (doze) meses. Não foram aplicados reajustes, inflação, variação cambial, descontos decorrentes de compromissos de longo prazo ou alterações futuras nos preços públicos dos provedores, de modo que o resultado deve ser compreendido como referência econômica baseada nas premissas utilizadas neste Estudo Técnico Preliminar.

**8.3.8.23.** Considerando o valor médio de referência de **R$ 13.658.387,08** para 12 (doze) meses, o custo correspondente aos 48 (quarenta e oito) meses adicionais é estimado em **R$ 54.633.548,32**, equivalente a quatro vezes o valor anual. Somado ao custo do período inicial de 12 (doze) meses, o valor estimado para manutenção de capacidade equivalente em nuvem pública durante 60 (sessenta) meses totaliza **R$ 68.291.935,40**.

**8.3.8.24.** Na projeção individual das alternativas, a manutenção da capacidade equivalente durante 60 (sessenta) meses resultaria em aproximadamente **R$ 27.518.839,20** para a Oracle Cloud Infrastructure — OCI; **R$ 25.416.704,10** para a Amazon Web Services — AWS; e **R$ 151.940.262,85** para a alternativa Google Cloud/Gemini API, considerando, nesta última, o ponto médio entre as hipóteses de composição dos tokens de entrada e saída. A média dessas alternativas permanece compatível, ressalvadas diferenças decorrentes de arredondamento, com o valor de referência de aproximadamente **R$ 68,29 milhões**.

**8.3.8.25.** Aplicando-se a mesma projeção à análise de sensibilidade, o valor equivalente para 60 (sessenta) meses poderia variar entre aproximadamente **R$ 54.006.953,40**, na hipótese inferior da alternativa baseada em tokens, e **R$ 82.576.917,35**, na hipótese superior. Essa faixa reforça a possibilidade de variação relevante dos custos de nuvem conforme o perfil de consumo, a proporção entre tokens de entrada e saída e as condições comerciais aplicáveis durante o período.

**8.3.8.26.** Dessa forma, para fins de comparação com a disponibilidade do hardware próprio durante os 60 (sessenta) meses de garantia e suporte, adota-se como novo valor de referência da alternativa equivalente em nuvem pública o montante de **R$ 68.291.935,40**, correspondente a aproximadamente **R$ 68,29 milhões**. Esse valor permanece limitado à capacidade computacional considerada na análise e não contempla integralmente os custos acessórios necessários à disponibilização de uma solução completa de GPU como serviço em nuvem, tais como armazenamento persistente, backup, conectividade dedicada, transferência de dados, suporte avançado, licenças corporativas, segurança, observabilidade, integração, operação assistida e serviços profissionais.

## 9. Análise comparativa de soluções

**9.1.** Nesta subseção, será apresentada a análise comparativa em relação às soluções 1 e 2.

**9.2.** A análise de soluções, trata de uma solução completa, que abarca a combinação de hardware, softwares e serviços para atendimento aos objetivos da contratação.

**9.3.** Para a Solução 1 foram recebidas, durante o levantamento de mercado, proposta da Drive A associada à arquitetura HPE, no valor de R$ 47.481.100,00, e proposta inicial da Teltec associada à arquitetura Cisco, no valor de R$ 26.700.000,00. Posteriormente, após o refinamento das especificações, das quantidades, da distribuição físico-financeira e dos períodos dos entregáveis, a Teltec apresentou a Proposta TTEC3067/2026, versão 2.0, de 30 de julho de 2026, no valor global de R$ 19.977.500,00, que passa a constituir a referência comercial mais atual das tratativas relacionadas à arquitetura Cisco. A proposta da Drive A e a versão 1.0 da proposta da Teltec permanecem nos autos como registros históricos da evolução do levantamento de mercado.

**9.4.** A Proposta TTEC3067/2026, versão 2.0, será utilizada como referência preliminar para a composição e a distribuição dos valores dos entregáveis, por refletir o escopo, os quantitativos, os

períodos e a distribuição físico-financeira atualmente adotados no planejamento. Sua utilização não representa escolha antecipada de fornecedor, definição do valor estimado definitivo ou afastamento da necessidade de realização da pesquisa de preços nos termos da regulamentação aplicável.

**9.5.** Para a solução 2 será utilizado o valor já apresentado na seção anterior de R$ 68.290.000,00.

**9.6.** Os valores disponíveis nesta etapa possuem caráter preliminar. A estimativa definitiva deverá ser formada antes da fase externa mediante pesquisa de preços que utilize os parâmetros da IN SEGES /ME nº 65/2021, preferencialmente de forma combinada, com cesta ampla e diversificada, análise crítica dos escopos, tratamento estatístico, memória de cálculo e justificativa objetiva para eventual inclusão ou exclusão de valores. Caso permaneça a impossibilidade de obtenção de três preços comparáveis, essa circunstância deverá ser formalmente justificada e aprovada pela autoridade competente. A Proposta TTEC3067/2026, versão 2.0, no valor global de R$ 19.977.500,00, constitui uma das fontes disponíveis para essa análise, não substituindo a formação da estimativa definitiva mediante pesquisa de preços.

### 9.7. Análise da prestação de Serviços nas duas soluções e cronograma de desembolso financeiro

**9.7.1.** A análise da prestação de serviços considera as diferenças existentes entre as duas soluções avaliadas. Na Solução 1, a contratação compreende o fornecimento de equipamentos, licenças de software, garantias e serviços especializados necessários à implantação, integração, capacitação, operação assistida e sustentação inicial de uma plataforma própria de GPU como serviço. Na Solução 2, a capacidade computacional é disponibilizada integralmente como serviço em nuvem pública, sem transferência da propriedade dos equipamentos para a Administração e com desembolso recorrente durante todo o período em que a capacidade permanecer disponível.

#### 9.7.2. Solução 1 — Aquisição de hardware, software e serviços para disponibilização de solução própria de GPU como serviço

**9.7.2.1.** Para análise dos serviços da Solução 1, foram considerados os valores constantes da Proposta Comercial TTEC3067/2026, versão 2.0, encaminhada pela Teltec Solutions, abrangendo os Entregáveis 9 a 33 das Especificações Técnicas.

**9.7.2.2.** O Entregável 9 — Serviço de Integração com Equipamentos de Rede de Frontend, destinado à integração da solução com os equipamentos e softwares de rede de frontend existentes na Administração, possui valor estimado de R$ 39.955,00.

**9.7.2.3.** O Entregável 10 — Reunião Inicial, destinado ao alinhamento formal das equipes, do escopo, do cronograma, das responsabilidades, dos riscos e das condições de execução, possui valor estimado de R$ 19.977,50.

**9.7.2.4.** O Entregável 11 — Gerenciamento de Projetos, destinado à coordenação técnica e administrativa, ao acompanhamento dos prazos, riscos, entregas e participantes e ao apoio à fiscalização contratual, possui valor estimado de R$ 159.820,00.

**9.7.2.5.** O Entregável 12 — Proposta de Escopo de Trabalho, correspondente à elaboração do documento formal de escopo, limites, entregas, responsabilidades, cronograma e critérios de aceitação do projeto, possui valor estimado de R$ 139.842,50.

**9.7.2.6.** O Entregável 13 — Levantamento da Situação Atual, destinado à identificação e à documentação das condições físicas, lógicas, operacionais e de segurança do ambiente de implantação, possui valor estimado de R$ 99.887,50.

**9.7.2.7.** O Entregável 14 — Tomada de Conhecimento da Documentação Existente, destinado à análise formal da arquitetura, dos processos, das normas, dos registros e dos documentos técnicos institucionais, possui valor estimado de R$ 119.865,00.

**9.7.2.8.** O Entregável 15 — Apontamento de Oportunidades de Melhoria, destinado à identificação de recomendações relacionadas à segurança, ao desempenho, à continuidade, à governança e à eficiência operacional, possui valor estimado de R$ 119.865,00.

**9.7.2.9.** O Entregável 16 — Definição de Requisitos da Solução, destinado à consolidação dos requisitos de negócio, técnicos, operacionais, de escalabilidade, computação, armazenamento, rede e integração, possui valor estimado de R$ 79.910,00.

**9.7.2.10.** O Entregável 17 — Treinamento Inicial da Solução, destinado à apresentação dos conceitos fundamentais, da arquitetura geral e do posicionamento dos componentes no ambiente institucional, possui valor estimado de R$ 79.910,00.

**9.7.2.11.** O Entregável 18 — Planejamento de Alto Nível da Solução, correspondente à elaboração do High Level Design — HLD, contendo a arquitetura de referência, o planejamento da implantação, a segurança, o roteamento, a integração e as diretrizes gerais da solução, possui valor estimado de R$ 679.235,00.

**9.7.2.12.** O Entregável 19 — Proposta Técnica de Implementação, destinado ao detalhamento das etapas, dos procedimentos, dos cenários de teste e da abordagem técnica para implantação, possui valor estimado de R$ 459.482,50.

**9.7.2.13.** O Entregável 20 — Desenho de Baixo Nível da Solução, correspondente à elaboração do Low Level Design — LLD, contendo configurações, diagramas, conexões, parâmetros, planos de contingência e procedimentos detalhados, possui valor estimado de R$ 599.325,00.

**9.7.2.14.** O Entregável 21 — Preparação da Documentação de Instalação, destinado à elaboração dos procedimentos de instalação, migração, configuração, testes e reversão aplicáveis a todos os componentes, possui valor estimado de R$ 279.685,00.

**9.7.2.15.** O Entregável 22 — Atualização da Documentação Técnica Existente, destinado à atualização dos documentos institucionais afetados pela implantação, incluindo inventários, desenhos, LLDs e demais registros técnicos, possui valor estimado de R$ 39.955,00.

**9.7.2.16.** O Entregável 23 — Treinamento Antes da Implementação, destinado à preparação técnico-operacional da equipe da Administração para acompanhamento das atividades de instalação e configuração, possui valor estimado de R$ 39.955,00.

**9.7.2.17.** O Entregável 24 — Implementação, destinado à instalação, parametrização, configuração, integração, testes e validação de todos os equipamentos e softwares da solução, possui valor estimado de R$ 1.997.750,00.

**9.7.2.18.** O Entregável 25 — Treinamento Após a Implementação, destinado à transferência de conhecimento sobre a arquitetura efetivamente implantada, sua operação, administração, manutenção e suporte, possui valor estimado de R$ 79.910,00.

**9.7.2.19.** O Entregável 26 — Atualização da Documentação, destinado à elaboração dos documentos finais e do as-built da solução efetivamente implantada, possui valor estimado de R$ 59.932,50.

**9.7.2.20.** O Entregável 27 — Manual de Operações, correspondente à elaboração do Operations Run Book com procedimentos de operação, monitoramento, gerenciamento e manutenção da solução, possui valor estimado de R$ 199.775,00.

**9.7.2.21.** O Entregável 28 — Operação Assistida, destinado à estabilização inicial do ambiente, à realização de ajustes técnicos e à transferência prática de conhecimento operacional, possui valor estimado de R$ 599.325,00.

**9.7.2.22.** O Entregável 29 — Monitoramento da Solução, destinado à definição, configuração, integração e validação das métricas, alarmes, eventos e indicadores operacionais, possui valor estimado de R$ 59.932,50.

**9.7.2.23.** O Entregável 30 — Encerramento do Projeto, destinado à consolidação das entregas, da documentação, das lições aprendidas e da transferência final de conhecimento, possui valor estimado de R$ 239.730,00.

**9.7.2.24.** O Entregável 31 — Serviço de Suporte Avançado Mensal do Fabricante, destinado à prestação de suporte especializado, gestão de incidentes, análise de causa raiz, avaliação preventiva e apoio à sustentação dos componentes críticos, possui quantidade de 12 mensalidades, com valor médio unitário aproximado de R$ 58.267,71, totalizando R$ 699.212,50, correspondente a 3,50% do valor global da proposta.

**9.7.2.25.** O Entregável 32 — Serviço de Suporte Mensal do Fornecedor, destinado à manutenção corretiva e preventiva, ao atendimento de incidentes, ao suporte presencial e remoto, à otimização e ao acompanhamento operacional da solução, possui quantidade de 12 mensalidades, com valor médio unitário aproximado de R$ 83.239,58, totalizando R$ 998.875,00, correspondente a 5,00% do valor global da proposta.

**9.7.2.26.** O Entregável 33 — Serviço Especializado de Treinamento Avançado e Atualização Tecnológica, destinado à participação de militares em eventos técnicos de referência relacionados à Inteligência Artificial, infraestrutura, GPU como serviço e tecnologias integrantes da solução, possui quantidade de 8 participações, com valor médio unitário aproximado de R$ 37.457,81, totalizando R$ 299.662,50, correspondente a 1,50% do valor global da proposta.

**9.7.2.27.** A soma dos valores individualizados dos Entregáveis 9 a 33 corresponde a R$ 8.190.775,00, representando a parcela da proposta diretamente relacionada aos serviços de integração, planejamento, projeto, implementação, documentação, capacitação, operação assistida, monitoramento e suporte.

**9.7.2.28.** Aos serviços devem ser adicionados os demais componentes da Solução 1: Entregável 1 — Servidor de Processamento com GPU, componente de hardware, no valor de R$ 3.396.175,00; Entregável 2 — Garantia e Licenciamento do Servidor de Processamento com GPU por 60 meses, no valor de R$ 3.396.175,00; Entregável 3 — Servidor para Hospedagem da Camada de Software, componente de hardware, no valor de R$ 998.875,00; Entregável 4 — Garantia e Licenciamento desse servidor por 60 meses, no valor de R$ 998.875,00; Entregável 5 — Software de Virtualização, Contêineres e Orquestração por 12 meses, no valor de R$ 1.398.425,00; Entregável 6 — Software Específico para Aceleração por GPU por 12 meses, no valor de R$ 399.550,00; Entregável 7 — Software para Interface de Serviços e Autosserviço por 12 meses, no valor de R$ 599.325,00; e

Entregável 8 — Software de Observabilidade e Segurança Operacional por 12 meses, no valor de R$ 599.325,00. Esses componentes totalizam R$ 11.786.725,00.

**9.7.2.29.** A soma aritmética dos valores individualizados dos 33 entregáveis da Proposta Comercial TTEC3067/2026, versão 2.0, resulta em R$ 19.977.500,00, coincidindo com o valor global declarado na proposta e eliminando a divergência aritmética identificada na versão anteriormente apresentada. Para os Entregáveis 31, 32 e 33, os valores unitários indicados possuem arredondamento em centavos, devendo ser considerados, para fins de composição do valor global, os respectivos valores totais constantes da proposta.

**9.7.2.30.** Para fins de projeção do desembolso no horizonte de 60 meses adotado para a análise comparativa, considera-se que a maior parte dos pagamentos ocorrerá no primeiro ano, em razão da entrega dos equipamentos, da disponibilização das licenças e da execução dos serviços de planejamento, projeto, instalação, configuração, treinamento, documentação e implementação. Os serviços mensais de suporte e as parcelas residuais de capacitação e sustentação se estenderão pelo segundo ano, de acordo com os respectivos marcos de início, medição e aceite. Nos períodos subsequentes, os equipamentos permanecerão disponíveis e cobertos pelas condições de garantia e suporte do hardware contratadas por 60 meses, observadas as condições estabelecidas nas Especificações Técnicas.

**9.7.2.31.** Com base no cronograma físico-financeiro preliminar, projeta-se para o Ano 1 desembolso de R$ 18.129.581,25, correspondente a 90,75% do valor global da proposta, abrangendo os equipamentos, garantias, licenciamentos, softwares, serviços de integração e a maior parte dos entregáveis de planejamento, implementação e capacitação previstos para o período.

**9.7.2.32.** Para o Ano 2, projeta-se desembolso de R$ 1.847.918,75, correspondente aos 9,25% restantes do valor global, relacionado principalmente aos pagamentos mensais dos serviços de suporte, às capacitações ainda previstas e aos últimos marcos de execução e aceite.

**9.7.2.33.** Para os Anos 3, 4 e 5, não são projetados novos desembolsos no âmbito da composição comercial apresentada pela Teltec, uma vez que a proposta contempla somente 12 meses para os softwares e serviços mensais, embora a garantia e o suporte ordinário dos componentes de hardware permaneçam vigentes por 60 meses. Eventuais renovações de softwares ou contratações adicionais de serviços após os períodos inicialmente previstos dependerão de nova avaliação técnica, pesquisa de preços, disponibilidade orçamentária e instrumento contratual próprio.

**9.7.2.34.** Dessa forma, a projeção da Solução 1 corresponde a R$ 18.129.581,25 no Ano 1, R$ 1.847.918,75 no Ano 2 e R$ 0,00 nos Anos 3, 4 e 5, totalizando R$ 19.977.500,00 ao longo do horizonte de cinco anos considerado na análise. Essa distribuição representa uma projeção anual agregada, devendo os pagamentos efetivos observar os marcos, percentuais, eventos de medição e condições de recebimento definidos no cronograma físico-financeiro da contratação.

### 9.7.3. Solução 2 — Contratação de capacidade equivalente em nuvem pública

**9.7.3.1.** O racional utilizado para a obtenção do valor da Solução 2 encontra-se detalhado na seção 8.3 deste Estudo Técnico Preliminar, na qual foram analisadas referências públicas de preços da Oracle Cloud Infrastructure — OCI, da Amazon Web Services — AWS e do Google Cloud/Gemini API, considerando capacidade equivalente à disponibilização contínua de 8 GPUs NVIDIA B300.

**9.7.3.2.** A Solução 2 possui natureza exclusivamente de prestação de serviços. Nessa alternativa, a Administração não adquire servidores, GPUs ou demais ativos físicos, não ocorre incorporação de equipamentos ao patrimônio e não permanece capacidade computacional disponível após a

interrupção dos pagamentos. A utilização da infraestrutura depende da continuidade da contratação e do pagamento recorrente ao provedor de nuvem.

**9.7.3.3.** Para comparação com a Solução 1, foi considerado o período total de 60 meses, correspondente ao prazo durante o qual os equipamentos próprios permaneceriam disponíveis e cobertos pela garantia contratada. O valor de referência da capacidade equivalente em nuvem pública para esse período foi estimado em **R$ 68.291.935,40**, conforme metodologia e premissas descritas na seção 8.3.

**9.7.3.4.** Como os valores utilizados foram calculados em regime de uso contínuo e não foram considerados reajustes, inflação, variação cambial ou alteração futura dos preços dos provedores, a projeção do desembolso foi realizada de forma linear, adotando-se o valor anual de **R$ 13.658.387,08**.

**9.7.3.5.** Assim, para o **Ano 1**, projeta-se desembolso de **R$ 13.658.387,08**; para o **Ano 2**, **R$ 13.658.387,08**; para o **Ano 3**, **R$ 13.658.387,08**; para o **Ano 4**, **R$ 13.658.387,08**; e para o **Ano 5**, **R$ 13.658.387,08**, totalizando **R$ 68.291.935,40** durante os 60 meses.

**9.7.3.6.** Os valores da Solução 2 representam somente a referência econômica da capacidade computacional analisada e não contemplam integralmente os custos acessórios necessários à disponibilização de uma plataforma completa de GPU como serviço em nuvem pública, tais como armazenamento persistente, backup, transferência de dados, tráfego de saída, conectividade dedicada, suporte avançado, licenças corporativas, segurança, observabilidade, retenção de logs, integração com diretórios institucionais, governança de consumo, operação assistida e serviços profissionais.

**9.7.4. Comparação da prestação dos serviços e do desembolso financeiro**

**9.7.4.1.** A Solução 1 apresenta desembolso financeiro concentrado nos dois primeiros anos, especialmente no primeiro, em razão da aquisição e da implantação da infraestrutura. Em contrapartida, os equipamentos permanecem disponíveis para a Administração durante os 60 meses de vigência e passam a constituir capacidade computacional própria, sob controle e governança institucionais.

**9.7.4.2.** A Solução 2 distribui o desembolso de maneira uniforme durante os cinco anos, mas exige pagamento recorrente para manutenção da disponibilidade da capacidade computacional, sem incorporação de ativos ao patrimônio e sem disponibilidade residual da infraestrutura após o encerramento da contratação.

**9.7.4.3.** Considerando os valores de referência atualmente adotados, a Solução 1 implica desembolso global de **R$ 19.977.500,00**, enquanto a Solução 2 implica desembolso estimado de R$ 68.291.935,40 durante o período de 60 meses, resultando em diferença de **R$ 48.314.435,40** em favor da implantação da capacidade própria, sem considerar os custos acessórios da nuvem pública não incluídos na estimativa.

**9.7.4.4.** A análise demonstra, portanto, que, embora a Solução 1 exija maior concentração inicial de recursos, ela proporciona disponibilidade prolongada dos ativos, controle institucional da infraestrutura e menor desembolso total no horizonte de cinco anos. A Solução 2 apresenta menor necessidade de aquisição inicial de bens, porém mantém custo recorrente durante todo o período e não resulta na formação de capacidade computacional própria ao término da contratação.

## 9.8. Solução 1: Aquisição de hardware, software e serviços para disponibilização de uma solução de GPU como serviço para o Exército Brasileiro

### 9.8.1. Descrição

**9.8.1.1.** A Solução 1 consiste na aquisição de uma plataforma própria, integrada e do tipo *turn-key* para disponibilização de capacidade de processamento computacional acelerado com GPU como serviço, contemplando servidor de processamento equipado com 8 (oito) GPUs NVIDIA B300, infraestrutura destinada à hospedagem da camada de gerenciamento, softwares corporativos, garantias, licenciamentos e serviços especializados de planejamento, implantação, integração, capacitação, operação assistida, monitoramento e suporte.

**9.8.1.2.** A solução será instalada em ambiente *on-premises* do Exército Brasileiro e deverá ser integrada aos serviços institucionais de rede, identidade, segurança, armazenamento, monitoramento e gerenciamento, permitindo a disponibilização controlada, segura, governada e mensurável de recursos computacionais para cargas de treinamento, inferência, ajuste fino, desenvolvimento interativo, processamento documental e demais aplicações de Inteligência Artificial.

**9.8.1.3.** Os equipamentos permanecerão disponíveis para utilização pela Administração durante o período de 60 (sessenta) meses, correspondente ao prazo de garantia e suporte contratado para os componentes de hardware. A contratação inicial dos softwares, licenças e serviços observará os períodos e quantitativos definidos nas Especificações Técnicas e na proposta utilizada como referência, preservando-se a possibilidade de reavaliação e renovação da camada lógica conforme a evolução tecnológica, as métricas de utilização e a disponibilidade orçamentária.

### 9.8.2. Custo Total de Propriedade

**9.8.2.1.** Para fins da análise econômica desta solução, foi adotado, como referência preliminar, o valor global de **R$ 19.977.500,00**, correspondente à Proposta Comercial TTEC3067/2026, versão 2.0, de 30 de julho de 2026, encaminhada pela Teltec Solutions para os 33 entregáveis que compõem a solução, sem prejuízo da estimativa definitiva decorrente da pesquisa de preços.

**9.8.2.2.** O valor individual de cada entregável, sua finalidade, quantidade e distribuição ao longo da execução contratual encontram-se detalhados na seção 9.7 deste Estudo Técnico Preliminar. Nesta seção, considera-se apenas o valor global da solução, para fins de comparação consolidada com a alternativa de contratação em nuvem pública.

**9.8.2.3.** O valor de referência considera a aquisição e a implantação da capacidade própria e a disponibilidade dos equipamentos pelo prazo de 60 (sessenta) meses, não devendo ser confundido com eventual custo futuro de renovação de licenças de software ou contratação de serviços adicionais que não estejam abrangidos pelos quantitativos e períodos constantes da proposta comercial.

### 9.8.3. Vantagens

**9.8.3.1.** Formação de capacidade computacional própria e permanente durante a vida útil dos equipamentos, sem dependência de pagamentos recorrentes por hora de GPU para manutenção da disponibilidade da infraestrutura.

**9.8.3.2.** Menor custo global no horizonte de 60 (sessenta) meses, considerando os valores de referência utilizados na comparação com a alternativa em nuvem pública.

**9.8.3.3.** Incorporação dos equipamentos ao patrimônio da Administração, com manutenção da capacidade computacional instalada após a conclusão da etapa de implantação e dos serviços inicialmente contratados.

**9.8.3.4.** Maior soberania e controle institucional sobre dados, modelos, bases documentais, resultados de processamento, registros operacionais, credenciais e fluxos de informação.

**9.8.3.5.** Possibilidade de operação em ambiente *on-premises*, com conectividade externa restrita ou controlada e suporte a cenários segregados ou *air-gapped*, conforme os requisitos de segurança aplicáveis.

**9.8.3.6.** Maior previsibilidade de desempenho, disponibilidade e latência, uma vez que a capacidade computacional permanece dedicada à Administração e não depende da disponibilidade regional de recursos em provedores de nuvem.

**9.8.3.7.** Possibilidade de execução de diferentes tipos de cargas de trabalho, bibliotecas, modelos, *frameworks* e aplicações próprias, sem limitação às funcionalidades disponibilizadas por APIs ou serviços gerenciados específicos.

**9.8.3.8.** Maior possibilidade de customização e integração com os serviços institucionais de identidade, rede, segurança, armazenamento, observabilidade, auditoria, automação e gerenciamento.

**9.8.3.9.** Desenvolvimento e retenção de conhecimento técnico pela equipe do Exército Brasileiro, mediante participação no planejamento, implantação, treinamentos, operação assistida, monitoramento e sustentação da plataforma.

**9.8.3.10.** Possibilidade de expansão modular da infraestrutura com base em métricas reais de utilização, demanda de usuários, perfis de cargas de trabalho e evolução das necessidades institucionais.

**9.8.3.11.** Redução dos riscos associados a custos de transferência de dados, tráfego de saída, armazenamento persistente e consumo não controlado, normalmente cobrados de forma adicional em ambientes de nuvem pública.

**9.8.4. Desvantagens**

**9.8.4.1.** Necessidade de investimento financeiro concentrado, principalmente durante o primeiro ano da contratação, em razão da aquisição dos equipamentos, das licenças e da execução dos serviços de implantação.

**9.8.4.2.** Maior complexidade técnica e administrativa para planejamento, aquisição, integração, implantação, operação e sustentação de uma solução composta por múltiplas camadas de hardware, software, rede, segurança e serviços especializados.

**9.8.4.3.** Necessidade de disponibilização de infraestrutura física adequada, incluindo espaço em Data Center, energia elétrica, refrigeração, conectividade, segurança física e capacidade de monitoramento.

**9.8.4.4.** Responsabilidade institucional pela administração cotidiana da plataforma após o período de implantação, ainda que haja suporte da contratada e dos fabricantes durante os períodos contratados.

**9.8.4.5.** Possibilidade de obsolescência tecnológica dos componentes ao longo do ciclo de vida, especialmente diante da rápida evolução das arquiteturas de GPU e dos softwares de Inteligência Artificial.

**9.8.4.6.** Menor elasticidade imediata quando comparada à nuvem pública, pois a expansão da capacidade dependerá de novo planejamento, disponibilidade orçamentária, aquisição, entrega e instalação de equipamentos adicionais.

**9.8.4.7.** Necessidade de reavaliar e, quando necessário, renovar licenças de software e serviços especializados após o término dos respectivos períodos inicialmente contratados.

### 9.9. Solução 2: Contratação de solução de processamento de Inteligência Artificial em nuvem pública

#### 9.9.1. Descrição

**9.9.1.1.** A Solução 2 consiste na contratação de capacidade computacional de processamento acelerado em nuvem pública, disponibilizada integralmente como serviço por provedor externo, considerando capacidade equivalente à configuração inicial da solução própria, composta por 8 (oito) GPUs NVIDIA B300.

**9.9.1.2.** Nessa alternativa, a Administração não adquire servidores, GPUs ou outros ativos físicos, sendo o acesso à capacidade condicionado à manutenção da contratação e aos pagamentos recorrentes pelo consumo ou pela reserva dos recursos computacionais.

**9.9.1.3.** A estimativa foi construída a partir de referências públicas de preços da Oracle Cloud Infrastructure — OCI, Amazon Web Services — AWS e Google Cloud/Gemini API, utilizando período uniforme de comparação, regime de uso contínuo e as demais premissas detalhadas nas seções 8.3 e 9.7 deste Estudo Técnico Preliminar.

#### 9.9.2. Custo Total de Propriedade

**9.9.2.1.** Para fins da análise econômica da Solução 2, foi adotado o valor total de referência de **R$ 68.291.935,40 (sessenta e oito milhões, duzentos e noventa e um mil, novecentos e trinta e cinco reais e quarenta centavos)** para disponibilização de capacidade equivalente durante 60 (sessenta) meses.

**9.9.2.2.** O racional de cálculo, as referências de preços, as premissas de equivalência, os valores anuais e a projeção do desembolso durante os cinco anos encontram-se detalhados nas seções 8.3 e 9.7 deste Estudo Técnico Preliminar. Nesta seção, considera-se apenas o valor total consolidado para fins de comparação entre as soluções, conforme a Tabela 2.

**9.9.2.3.** O valor estimado não contempla integralmente os custos acessórios que poderão incidir sobre uma contratação efetiva de nuvem pública, tais como armazenamento persistente, armazenamento de alto desempenho, backup, *snapshots*, tráfego de saída, transferência de dados, conectividade dedicada, suporte avançado, licenças corporativas, observabilidade, segurança, retenção de logs, auditoria, integração, governança de consumo, operação assistida e serviços profissionais.

**Tabela 2: Comparativo consolidado dos valores das soluções**

| Solução | Descrição resumida | Período considerado | Valor total de referência |
|---|---|---|---|
| **Solução 1** | Aquisição de hardware, software, garantias, licenciamentos e serviços para implantação de capacidade própria de GPU como serviço | 60 meses | **R$ 19.977.500,00** |
| **Solução** | Contratação de capacidade equivalente de processamento de IA em nuvem | | **R$** |

| | | | |
|---|---|---|---|
| **2** | pública | 60 meses | **68.291.935,40** |

**9.9.2.4.** A diferença entre os valores totais das soluções é de **R$ 48.314.435,40**, sendo a Solução 2 aproximadamente R$ 48,31 milhões mais onerosa no período de 60 meses, mesmo sem a inclusão integral dos custos acessórios necessários à disponibilização de uma plataforma completa de GPU como serviço em nuvem pública.

**9.9.3. Vantagens**

**9.9.3.1.** Menor necessidade de investimento inicial em aquisição de equipamentos, uma vez que o pagamento é realizado de forma recorrente conforme o modelo contratado.

**9.9.3.2.** Possibilidade de disponibilização da capacidade sem a necessidade de aquisição, entrega, instalação e incorporação patrimonial de servidores e demais componentes físicos.

**9.9.3.3.** Maior elasticidade para ampliação ou redução da capacidade computacional, desde que os recursos estejam disponíveis na região e no provedor contratados.

**9.9.3.4.** Possibilidade de atendimento a demandas temporárias, sazonais ou de curta duração sem necessidade de aquisição permanente de capacidade equivalente ao pico de utilização.

**9.9.3.5.** Transferência ao provedor de parte das responsabilidades relacionadas à infraestrutura física, energia, refrigeração, manutenção dos equipamentos e atualização dos componentes do Data Center.

**9.9.3.6.** Acesso mais rápido a novas gerações de aceleradores, quando disponibilizadas comercialmente pelo provedor, sem necessidade de substituição dos ativos físicos da Administração.

**9.9.3.7.** Possibilidade de utilização complementar para ampliação temporária da capacidade própria, execução de cargas não sensíveis ou atendimento a situações excepcionais de indisponibilidade ou demanda elevada.

**9.9.4. Desvantagens**

**9.9.4.1.** Custo total de referência significativamente superior ao da aquisição de capacidade própria no período de 60 meses.

**9.9.4.2.** Ausência de incorporação de ativos ao patrimônio da Administração, uma vez que a capacidade computacional deixa de estar disponível após a interrupção dos pagamentos ou o encerramento da contratação.

**9.9.4.3.** Dependência contínua do provedor, da disponibilidade regional dos recursos, das condições comerciais, dos modelos de cobrança e das políticas de prestação do serviço.

**9.9.4.4.** Possibilidade de variação dos custos em razão de reajustes, alterações de preços, flutuação cambial, mudanças de região, necessidade de reserva de capacidade ou utilização superior à inicialmente estimada.

**9.9.4.5.** Existência de custos adicionais de armazenamento, backup, transferência de dados, tráfego de saída, conectividade, suporte, segurança e serviços complementares, não integralmente contemplados na estimativa utilizada.

**9.9.4.6.** Menor controle institucional sobre a localização física da infraestrutura, os processos internos de operação, a administração dos equipamentos e determinados mecanismos de segurança e auditoria.

**9.9.4.7.** Maior complexidade para atendimento a requisitos de soberania, classificação da informação, proteção de dados, segregação, custódia de modelos, retenção de logs e controle integral dos fluxos de processamento.

**9.9.4.8.** Dependência de conectividade externa com capacidade, disponibilidade, segurança e latência compatíveis com as cargas de treinamento, inferência e transferência de grandes volumes de dados.

**9.9.4.9.** Limitação ou inviabilidade de operação integralmente *air-gapped*, especialmente para cargas, modelos, documentos ou bases de dados que não possam ser processados fora do ambiente institucional.

**9.9.4.10.** Risco de aprisionamento tecnológico e comercial em relação às APIs, formatos, serviços gerenciados, mecanismos de identidade, ferramentas de observabilidade e modelos de cobrança específicos do provedor.

**9.9.4.11.** Menor previsibilidade do custo efetivo quando o perfil de consumo institucional ainda não está consolidado, especialmente quanto à simultaneidade de usuários, horas de GPU, volumes de tokens, armazenamento, tráfego e utilização de serviços complementares.

**9.9.4.12.** A comparação econômica demonstra que a alternativa em nuvem pública pode ser adequada como recurso complementar, temporário ou elástico, mas se mostra menos vantajosa como substituta integral da capacidade própria em cenário de disponibilidade contínua durante 60 meses.

## 10. Registro de soluções consideradas inviáveis

**10.1.** A contratação exclusiva de capacidade em nuvem pública foi considerada inviável como solução principal e substitutiva integral da infraestrutura própria, na configuração analisada, por não atender de forma suficiente e simultânea aos requisitos de soberania e controle institucional sobre dados e modelos, operação em ambientes segregados ou sem conectividade externa, integração integral com os serviços internos, execução de múltiplos frameworks e workloads próprios e previsibilidade de custos para utilização contínua. Essa conclusão não impede a utilização futura da nuvem pública como capacidade complementar, temporária ou elástica, para cargas compatíveis com os requisitos de segurança e governança institucional.

## 11. Análise comparativa de custos (TCO)

**11.1. Os custos das soluções analisadas são apresentados em valores atuais, sem a aplicação de reajustes, inflação, atualização monetária, variação cambial, descontos financeiros ou cálculo de valor presente líquido. A metodologia adotada permite comparar as alternativas a partir das referências de preços disponíveis e das premissas uniformes definidas neste Estudo Técnico Preliminar.**

**11.2.** Para a análise comparativa, foi considerado o período total de 60 (sessenta) meses, correspondente ao prazo durante o qual os equipamentos adquiridos na Solução 1 permanecerão disponíveis para utilização pela Administração e cobertos pela garantia e pelo suporte do hardware.

**11.3.** O custo da Solução 1 foi obtido, como referência preliminar, a partir da Proposta Comercial TTEC3067/2026, versão 2.0, encaminhada pela Teltec Solutions em 30 de julho de 2026, no valor global de **R$ 19.977.500,00**. Esse valor abrange os 33 entregáveis da contratação, incluindo equipamentos, garantias, licenciamentos, softwares e serviços especializados. A composição individual dos entregáveis, suas descrições, quantidades, valores e projeções de desembolso encontra-se detalhada na seção 9.7 deste Estudo Técnico Preliminar.

**11.4.** O custo da Solução 2 foi obtido mediante o racional apresentado na seção 8.3, com base nas referências públicas de preços da Oracle Cloud Infrastructure — OCI, Amazon Web Services — AWS e Google Cloud/Gemini API. Para o período de 60 (sessenta) meses, o valor total estimado da capacidade computacional equivalente em nuvem pública é de **R$ 68.291.935,40 (sessenta e oito milhões, duzentos e noventa e um mil, novecentos e trinta e cinco reais e quarenta centavos)**.

**11.5.** A projeção anual dos custos das duas soluções é apresentada na Tabela 3.

**Tabela 3 — Custos Totais de Propriedade das soluções**

| Solução | Ano 1 | Ano 2 | Ano 3 | Ano 4 | Ano 5 | Total |
|---|---|---|---|---|---|---|
| **Solução 1** | R$ 18.129.581,25 | R$ 1.847.918,75 | R$ 0,00 | R$ 0,00 | R$ 0,00 | **R$ 19.977.500,00** |
| **Solução 2** | R$ 13.658.387,08 | R$ 13.658.387,08 | R$ 13.658.387,08 | R$ 13.658.387,08 | R$ 13.658.387,08 | **R$ 68.291.935,40** |

**11.6.** Na Solução 1, o desembolso financeiro encontra-se concentrado principalmente no primeiro ano, em razão da aquisição dos equipamentos, da disponibilização das licenças e da execução da maior parte dos serviços de planejamento, integração, implementação, capacitação, documentação e operação assistida. Para o Ano 1, estima-se o desembolso de R$ 18.129.581,25, correspondente a 90,5% do valor global da solução.

**11.7.** Para o Ano 2 da Solução 1, estima-se o desembolso de **R$** 1.847.918,75, correspondente aos 9,25% restantes do valor global, relacionado à conclusão dos entregáveis, aos últimos marcos de recebimento e aos serviços mensais e atividades que se estendam além do primeiro ano.

**11.8.** Para os Anos 3, 4 e 5, não foram projetados novos desembolsos na Solução 1, uma vez que o valor global da proposta utilizada como referência contempla a totalidade dos equipamentos e das garantias de hardware por 60 (sessenta) meses, mas prevê apenas os quantitativos e períodos iniciais de softwares e serviços indicados na proposta comercial.

**11.9.** A ausência de valores projetados para os Anos 3, 4 e 5 não significa que não possam surgir despesas adicionais durante esse período. A contratação inicial dos serviços pelo período de 12 (doze) meses decorre da necessidade de compatibilizar o escopo da solução com a disponibilidade orçamentária existente. Após o término da cobertura inicialmente contratada, poderão ser realizadas novas contratações para renovação de licenças, ampliação do período de suporte, sustentação, capacitação, atualização tecnológica ou prestação de outros serviços necessários ao funcionamento da solução.

**11.10.** As eventuais contratações posteriores dependerão da análise das necessidades efetivamente demonstradas pelas métricas de utilização, da maturidade operacional do ambiente, da evolução tecnológica, da avaliação de conveniência e vantajosidade e da disponibilidade de recursos orçamentários. Como esses elementos ainda não podem ser estimados com precisão suficiente, os respectivos valores não foram adicionados ao custo da Solução 1 nesta fase do planejamento.

**11.11.** Na Solução 2, por se tratar exclusivamente de prestação de serviço em nuvem pública, o desembolso é recorrente durante todo o período em que a capacidade computacional permanecer disponível. Não ocorre aquisição de equipamentos nem incorporação de ativos ao patrimônio, e a interrupção dos pagamentos implica a indisponibilidade da capacidade contratada.

**11.12.** Para a Solução 2, adotou-se o valor anual de referência de **R$ 13.658.387,08**, calculado para 12 (doze) meses de disponibilidade contínua da capacidade equivalente. A projeção linear desse valor durante cinco anos resulta em:

**11.12.1.** Ano 1: **R$ 13.658.387,08**;

**11.12.2.** Ano 2: **R$ 13.658.387,08**;

**11.12.3.** Ano 3: **R$ 13.658.387,08**;

**11.12.4.** Ano 4: **R$ 13.658.387,08**; e

**11.12.5.** Ano 5: **R$ 13.658.387,08**.

**11.13.** A multiplicação do valor anual de **R$ 13.658.387,08** pelos cinco anos considerados resulta no custo total estimado de **R$ 68.291.935,40** para a Solução 2. Essa projeção foi realizada sem aplicação de reajustes, inflação, variação cambial ou alteração futura dos preços dos provedores.

**11.14.** O valor estimado da Solução 2 não contempla integralmente os custos acessórios que poderiam ser necessários à disponibilização de uma plataforma completa de GPU como serviço em nuvem pública, incluindo armazenamento persistente e de alto desempenho, backup, *snapshots*, transferência de dados, tráfego de saída, conectividade dedicada, suporte avançado, licenças corporativas, observabilidade, segurança, retenção de logs, auditoria, integração com diretórios institucionais, governança de consumo, operação assistida e serviços profissionais.

**11.15.** Comparando-se os valores totais, a Solução 2 apresenta custo estimado **R$ 48.314.435,40** superior ao da Solução 1 durante o período considerado. Esse acréscimo corresponde a aproximadamente **241,84%** do valor da Solução 1, de modo que o custo estimado da alternativa em nuvem pública equivale a aproximadamente **3,42 vezes** o valor global atualmente utilizado como referência para a aquisição da infraestrutura própria.

**11.16.** Sob outra perspectiva, a adoção da Solução 1 representa redução estimada de aproximadamente **70,75%** em relação ao valor da Solução 2, consideradas as premissas e os escopos utilizados nesta análise.

**11.17.** Deve-se reconhecer que os escopos temporais não são integralmente simétricos. Na Solução 1, o hardware permanecerá disponível e coberto por garantia durante 60 meses, enquanto determinados softwares e serviços foram inicialmente precificados por 12 meses. Na Solução 2, a estimativa pressupõe prestação contínua da capacidade computacional durante os 60 meses. Assim, eventual ampliação dos períodos de licenciamento e de prestação dos serviços da Solução 1 poderá aumentar seu custo total de propriedade.

**11.18.** Não obstante essa ressalva, a diferença de valores identificada é significativa e a estimativa da Solução 2 ainda não incorpora todos os custos acessórios de uma plataforma completa em nuvem pública. Dessa forma, com base nos dados atualmente disponíveis, a Solução 1 permanece economicamente mais vantajosa no horizonte analisado, mesmo considerando a possibilidade de futuras contratações complementares de software e serviços.

**11.19.** Além da vantagem econômica, a Solução 1 resulta na formação de capacidade computacional própria, com incorporação dos equipamentos ao patrimônio, maior controle sobre dados e modelos, possibilidade de operação em ambiente segregado, previsibilidade de desempenho e menor dependência de pagamentos recorrentes para manutenção da disponibilidade da infraestrutura.

**11.20.** Conclui-se, portanto, que a **Solução 1 — Aquisição de hardware, software e serviços para disponibilização de uma solução de GPU como serviço para o Exército Brasileiro**, com custo total inicial estimado de **R$ 19.977.500,00**, apresenta a melhor relação entre custo, disponibilidade, controle institucional, soberania, possibilidade de expansão e formação de capacidade própria, quando comparada à **Solução 2 — Contratação de processamento de Inteligência Artificial em nuvem pública**, cujo custo de referência para 60 meses foi estimado em **R$ 68.291.935,40**.

## 12. Descrição da solução de TIC a ser contratada

**12.1.** Foi escolhida a solução 1, por atendimento aos requisitos técnicos e por ser a solução que gera maior economicidade para Administração.

**12.2.** As tratativas de mercado permitiram identificar arquiteturas potencialmente capazes de subsidiar o atendimento da necessidade. Entre os fabricantes que participaram das interações, a solução associada à Cisco apresentou maior aderência ao conjunto de requisitos e maior amadurecimento durante o planejamento. A HPE, por intermédio da Drive A, apresentou proposta e contribuições ao longo das tratativas, mas posteriormente comunicou sua decisão de não prosseguir na participação relacionada à contratação, pelos motivos já registrados. Posteriormente, a Teltec, com a Cisco em cópia, confirmou a possibilidade de atendimento à distribuição físico-financeira proposta e informou não ter identificado, até aquele momento, requisitos impeditivos à participação com solução Cisco. Essa circunstância não representa prévia habilitação, homologação ou garantia de atendimento da arquitetura Cisco, que deverá, assim como qualquer solução ofertada no certame, comprovar integralmente sua conformidade com as especificações técnicas, as condições de execução e os critérios de aceitação definidos no instrumento convocatório.

## 13. Estimativa de custo total da contratação

**Valor (R$):** 19.977.500,00

**13.1.** Para fins desta etapa do planejamento, adota-se como estimativa preliminar da contratação o valor de **R$ 19.977.500,00** (dezenove milhões, novecentos e setenta e sete mil e quinhentos reais), correspondente ao valor global da Proposta Comercial TTEC3067/2026, versão 2.0, de 30 de julho de 2026, sem prejuízo de sua revisão após a conclusão da pesquisa de preços.

**13.2.** As especificações técnicas atualizadas foram submetidas a fabricantes e fornecedores durante o desenvolvimento deste Estudo Técnico Preliminar. A Proposta TTEC3067/2026, versão 2.0, foi recebida como resposta às tratativas realizadas e constitui uma das referências disponíveis para a

pesquisa de preços. A estimativa definitiva deverá ser estabelecida a partir da pesquisa de preços da contratação, razão pela qual o valor final poderá ser superior ou inferior à referência atualmente adotada.

**13.3.** Considerando a Proposta Comercial TTEC3067/2026, versão 2.0, a estimativa preliminar atualmente utilizada para a contratação é de R$ 19.977.500,00, situando-se abaixo do limite de R$ 20.000.000,00 previsto para submissão da contratação à análise de alçada da Secretaria de Governo Digital. A aplicabilidade da IN SGD/MGI nº 6, de 29 de março de 2023, deverá, contudo, ser novamente verificada após a conclusão da pesquisa de preços e definição do valor global estimado do objeto. Caso a estimativa definitiva resulte em valor igual ou superior a R$ 20.000.000,00, deverão ser adotados os procedimentos de submissão e aprovação previstos na regulamentação antes do prosseguimento da contratação.

**13.4.** Registra-se que o valor de R$ 19.977.500,00 decorre de proposta comercial revisada apresentada pela Teltec após as solicitações da Administração destinadas à compatibilização da proposta com os quantitativos, períodos, estrutura dos entregáveis e cronograma físico-financeiro da contratação. Não foi estabelecido pela Administração valor inferior a R$ 20.000.000,00 como condição, meta ou parâmetro para a revisão comercial. A definição da incidência ou não da análise de alçada deverá decorrer exclusivamente do valor global estimado obtido ao final da pesquisa de preços, independentemente de seu posicionamento em relação ao limite regulamentar.

## 14. Justificativa técnica da escolha da solução

**14.1.** A aquisição da solução 1 mostra-se viável e atende adequadamente às demandas de negócio formuladas, os benefícios pretendidos são apropriados, os riscos envolvidos são administráveis e a área requisitante priorizará o fornecimento de todos os elementos aqui relacionados, necessários à consecução dos benefícios pretendidos, pelo que recomenda-se a aquisição proposta.

## 15. Justificativa econômica da escolha da solução

**15.1.** A aquisição da solução 1 mostra-se viável, os custos previstos são compatíveis e caracterizam a economicidade devido às demandas, a área requisitante priorizará o fornecimento de todos os elementos aqui relacionados necessários à consecução dos benefícios pretendidos, pelo que recomendamos a aquisição proposta.

## 16. Benefícios a serem alcançados com a contratação

**16.1.** Atendimento aos objetivos elencados no item 4 deste Estudo Técnico Preliminar, mediante a implantação de capacidade própria, integrada, segura, governada e expansível de processamento computacional acelerado por GPU para aplicações de Inteligência Artificial, de forma a disponibilizar recursos de treinamento, inferência, ajuste fino, desenvolvimento interativo e processamento de grandes volumes de dados; ampliar a soberania e o controle institucional sobre dados, modelos e cargas de trabalho; assegurar níveis adequados de desempenho, disponibilidade, segurança da informação, observabilidade, auditoria e controle de acesso; possibilitar o compartilhamento eficiente dos recursos entre usuários e projetos; promover a integração com a infraestrutura e os serviços

corporativos existentes no Data Center do Exército; desenvolver a capacidade técnica das equipes responsáveis pela operação e sustentação da plataforma; e permitir a expansão modular da solução conforme a evolução das necessidades institucionais e das tecnologias de Inteligência Artificial.

## 17. Providências a serem Adotadas

**17.1.** Deverão ser adotadas as providências necessárias para o prosseguimento do processo de aquisição.

**17.2.** Os militares envolvidos na fiscalização e gestão contratual deverão obrigatoriamente realizar os cursos de Fiscalização e Gestão de Contratos de TIC, Gestão e Fiscalização de Contratos Administrativos e Gestão de Risco no Setor Público, todos disponíveis no endereço eletrônico: https://www.escolavirtual.gov.br, ou no Portal de Educação do Exército, no endereço eletrônico: https://portaldeeducacao.eb.mil.br/academico.

### 17.3. Sobre o modo de disputa do pregão e orçamento

**17.3.1.** A presente seção tem por objetivo discutir de forma técnica a melhor abordagem do modo de disputa, conforme Art. 56 da Lei Nº 14.133, de 01 de abril de 2021 e sobre o orçamento, conforme Art. 24 da Lei Nº 14.133, de 01 de abril de 2021.

**17.3.2.** Embora a presente contratação estabeleça indicação de marca e modelo exclusivamente para as GPUs NVIDIA B300 e para o produto NVIDIA AI Enterprise, os demais componentes de hardware e software possuem especificações abertas, admitindo-se a participação de diferentes fabricantes de servidores capazes de atender integralmente aos requisitos técnicos. Nesse contexto, cabe a este Estudo Técnico Preliminar manifestar-se sobre o modo de disputa e o tratamento do orçamento, especialmente diante da inexistência de consenso na Administração quanto à utilização dessas metodologias em contratações com solução integrada e restrição pontual de marca e modelo.

**17.3.3**. Os fabricantes consultados durante os estudos técnicos atuam no Brasil com estrutura de distribuidores e fornecedores, ou seja, não fazem venda direta, porém, os preços são controlados pelos fabricantes, o que torna a competitividade em relação aos valores dos equipamentos diretamente ligada ao método de formação de preço de cada fabricante. Esse método depende de vários fatores que tem pouco ou nenhum controle por parte da Administração, alguns desses fatores identificados são: proximidade do término do ano fiscal, interesse em aumentar a presença no Brasil, estratégias de expansão regional, interesse em ter o Exército como cliente, etc.

**17.3.4**. O CITEx (UASG 160091) e suas unidades subordinadas, no caso o 3º CTA (UASG 160486) e 7º CTA (UASG 160749), vem realizando aquisições centralizadas, sob a coordenação técnica do CITEx. A seguir, na Tabela 4, encontra-se um resumo dessas contratações, com ou sem restrição de marca, o modo de disputa, tipo de orçamento relação de valores entre hardware e software com garantia e o percentual de desconto em relação ao valor de referência.

Tabela 4: Resumos dos processos centralizados do CITEx.

| Nr | Processo | UASG | Pregão | Marca e Modelo ou Aberto | Modo de Disputa | Orçamento sigiloso | Relação HW/SW | % Desconto |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 19 | 18 - 2026 Antivírus - CITEx | 160091 | PE 90065 /2026 | Marca e Modelo | Aberto | Não | N/A | 0% |
| 18 | 17 - 2025 - Secure E-mail Gateway - 7º CTA | 160749 | PE 90012 /2025 | Aberto | Aberto / Fechado | Não | N/A | 21% |
| 17 | 16 - 2025 - Switch Data Center - CITEx | 160091 | PE 90152 /2025 | Aberto | Aberto | Não | 44% / 56% | 37% |
| 16 | 15 - 2025 - Switch SAN - 7º CTA | 160749 | PE 90010 /2025 | Marca e Modelo | Aberto / Fechado | Não | 68% / 32% | 7% |
| 15 | 14 - 2025 - Computadores - CITEx | 160091 | PE 90009 /2025 | Aberto | Aberto / Fechado | Não | N/A | 14% |
| 14 | 13 - 2024 - Computadores Deserto - CITEx | 160091 | PE 90072 /2024 | Aberto | Aberto / Fechado | Sim | N/A | N/A |
| 13 | 12 - 2025 - Storage Relançamento - 7º CTA | 160749 | PE 90006 /2025 | Aberto | Aberto / Fechado | Não | N/A | 31% |
| 12 | 11 - 2024 - Storage - 7º CTA | 160749 | PE 90005 /2024 | Aberto | Aberto / Fechado | Não | N/A | 1% |
| 11 | 10 - 2024 - Zimbra - CITEx | 160091 | PE 07 /2023 | Marca e Modelo | Aberto | Não | N/A | 5% |
| 10 | 09 - 2024 - BBI 100G - CITEx | 160091 | PE 90113 /2024 | Marca e Modelo | Aberto | Não | 26% / 74% | 1% |
| 9 | 08 - 2023 - Segurança / UTM (Roteadores e Firewalls) - CITEx | 160091 | PE 03 /2023 | Aberto | Aberto / Fechado | Sim | 63% / 37% | 54% |
| | | | | | | | | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 8 | Data Center 3º CTA | 160486 | PE 07 /2023 | Aberto | Aberto | Não | N/A | 0% |
| 7 | 07 - 2023 - WAF - CITEx | 160091 | PE 04 /2023 | Marca e Modelo | Aberto / Fechado | Sim | N/A | 7% |
| 6 | 06 - 2023 - Computadores - CITEx | 160091 | PE 05 /2023 | Aberto | Aberto / Fechado | Sim | N/A | 24% |
| 5 | 05 - 2023 - Hospedagem Complemento - 3º CTA | 160486 | PE 08 /2023 | Marca e Modelo | Aberto | Não | N/A | 15% |
| 4 | 04 - 2023 - Hospedagem relançamento de itens - 3º CTA | 160486 | PE 3 /2023 | Marca e Modelo | Aberto | Não | N/A | 38% |
| 3 | 03 - 2023 - Hospedagem - 3º CTA | 160486 | PE 11 /2022 | Marca e Modelo | Aberto | Não | N/A | 43% |
| 2 | 02 - 2023 - BBI - CITEx | 160091 | PE 14 /2022 | Marca e Modelo | Aberto | Não | 36% / 64% | 5% |
| 1 | 01 - 2022 - Backup - CITEx | 160091 | PE 08 /2022 | Marca e Modelo | Aberto | Não | 10% / 90% | 4% |

**17.3.5.** As contratações integralmente realizadas com restrição de marca e modelo correspondem aos itens 1, 2, 7, 10, 11, 16 e 19 da Tabela 4. Além dessas, os itens 3, 4 e 5 correspondem a pregões que combinaram itens com especificações abertas e itens com restrição de marca e modelo. Essas contratações foram realizadas com e sem orçamento sigiloso e mediante a adoção tanto do modo de disputa aberto quanto do modo aberto e fechado. Nas contratações integralmente restritas, os descontos verificados variaram entre 0% e 7%, demonstrando pouca diferença entre o valor máximo estimado e o valor final do pregão. Nos pregões que combinaram itens abertos e restritos, os descontos globais foram mais elevados, mas os dados disponíveis não permitem atribuir a redução especificamente aos itens com restrição de marca e modelo. Nesse contexto, a experiência das contratações indica que a realização de consulta prévia com fornecedores, integradores e fabricantes, buscando a apresentação de propostas comerciais que já contemplem descontos significativos, pode exercer maior influência sobre o preço final do que o tipo de orçamento ou o modo de disputa adotado.

**17.3.6.** As contratações integralmente realizadas sem restrição de marca e modelo correspondem aos itens 6, 8, 9, 12, 13, 14, 15, 17 e 18 da Tabela 4. Também devem ser considerados os itens com

especificações abertas constantes dos pregões mistos identificados nos itens 3, 4 e 5. Nesses processos, foi possível perceber uma diferença na dinâmica da etapa competitiva. Os certames realizados pelo modo de disputa aberto e fechado apresentaram, em regra, menor quantidade de lances durante a etapa aberta, observando-se uma tendência de os fornecedores aguardarem a etapa fechada para apresentar sua melhor oferta. Nos processos realizados exclusivamente pelo modo aberto, houve maior quantidade de lances sucessivos e disputa mais prolongada entre os participantes. Além disso, no modo aberto e fechado, o prazo reduzido para apresentação do lance final fechado pode elevar o risco de erros de digitação ou de formulação da proposta pelos fornecedores. A IN SEGES/ME nº 73/2022 caracteriza o modo aberto pela apresentação de lances públicos e sucessivos e o modo aberto e fechado pela combinação desses lances com uma oferta final fechada.

**17.3.7.** Quanto ao orçamento sigiloso, adotado nos itens 6, 7, 9 e 14 da Tabela 4, deve-se considerar que, nas aquisições compostas por grupos ou por soluções integradas, existe a possibilidade de a proposta vencedora pelo valor global conter preços individuais superiores aos valores máximos admitidos pela Administração. Como os licitantes não conhecem previamente os valores máximos de cada componente, essa situação pode resultar na desclassificação da proposta globalmente mais vantajosa ou exigir a realização de negociações e ajustes posteriores, com risco de insucesso da contratação. Diante desse cenário, os dados analisados não demonstram que o uso do orçamento sigiloso tenha produzido vantagem objetiva para a Administração. Ao contrário, para objetos compostos por diferentes componentes, os riscos decorrentes da ausência de conhecimento dos preços máximos individualizados podem superar os potenciais benefícios do sigilo.

**17.3.8.** A adoção do modo de disputa aberto e fechado, utilizada nos itens 6, 7, 9, 12, 13, 14, 15, 16 e 18 da Tabela 4, também apresenta riscos que devem ser considerados. Nesse modo, apenas os participantes enquadrados nas condições estabelecidas pelo sistema podem apresentar um único lance final e fechado. Em mercados caracterizados pela atuação de fabricantes por intermédio de canais de revenda, existe a possibilidade de um mesmo fabricante mobilizar diversos parceiros comerciais para participar da fase aberta, fazendo com que os fornecedores classificados para a fase fechada representem, na prática, uma mesma solução ou fabricante. Esse cenário pode ser agravado pela existência de mecanismos comerciais de registro de oportunidade, descontos diferenciados ou proteção de negócios, ainda que implementados de forma indireta. Tais circunstâncias podem reduzir a competição efetiva e prejudicar a observância dos princípios da igualdade, da competitividade, do julgamento objetivo, da transparência e da busca da proposta mais vantajosa, previstos no art. 5º da Lei nº 14.133, de 1º de abril de 2021.

**17.3.9.** As contratações centralizadas e de elevado valor comumente atraem fornecedores que não possuem experiência suficiente, estrutura operacional ou capacidade técnica para atender integralmente aos requisitos de habilitação e execução contratual. Essa situação é especialmente relevante na presente contratação, que envolve fornecimento de componentes tecnológicos e prestação de serviços de elevada especificidade. Diante desse cenário, mostra-se adequado preservar a participação do maior número possível de empresas durante toda a fase de lances, evitando que fornecedores com efetiva capacidade técnica sejam excluídos da etapa competitiva em razão da dinâmica do modo de disputa. A ampliação da competição durante a fase de lances também aumenta a possibilidade de obtenção de preços mais vantajosos e reduz o risco de concentração das propostas em fornecedores vinculados a um mesmo fabricante.

**17.3.10.** Nas contratações em que foi identificada a relação entre hardware e software, correspondentes aos itens 1, 2, 9, 10, 16 e 17 da Tabela 4, realizou-se a segregação entre a parcela de hardware e a parcela correspondente aos componentes de software e garantia, consideradas conjuntamente. Essa separação tem por finalidade permitir o adequado enquadramento tributário dos

componentes da solução e a correspondente estruturação do cronograma físico-financeiro. Nas contratações que continham mais de um item combinado, foi utilizada como referência a proporção relativa ao item combinado de maior valor. As relações históricas apresentaram elevada dispersão, variando entre 10% de hardware e 90% de software e garantia e 68% de hardware e 32% de software e garantia, não sendo possível identificar uma proporção predominante ou suficientemente estável. Diante dessa variação e considerando a natureza integrada da solução a ser contratada, será adotada, para fins de planejamento, a proporcionalidade de 50% do valor para hardware e 50% para software e garantia.

**17.3.11.** Em relação aos descontos obtidos, verifica-se que cada licitação apresenta características mercadológicas, comerciais e temporais próprias, não sendo possível estabelecer uma relação causal direta entre o modo de disputa, o tipo de orçamento e o percentual final de redução. A amostra indica que as licitações realizadas com especificações abertas tendem a apresentar diferenças mais elevadas entre o valor máximo estimado e o preço final do pregão, quando comparadas às contratações integralmente restritas a marca e modelo. Entretanto, a formação do preço depende de diversos fatores alheios ao controle da Administração, como o interesse comercial do fabricante na contratação, a quantidade de concorrentes, a disponibilidade de estoque, as políticas de desconto concedidas aos canais de venda, a variação cambial, o posicionamento estratégico do fabricante e o período do ano em que o certame é realizado. A proximidade do encerramento de períodos mensais, trimestrais ou anuais e a necessidade de cumprimento de metas comerciais pelo fabricante ou por seus parceiros podem resultar na concessão de descontos significativamente superiores aos normalmente praticados. Assim, os percentuais históricos devem ser utilizados como elementos de referência, mas não como garantia de que resultados semelhantes serão obtidos na presente contratação.

**17.3.12.** Ressalta-se que o conjunto de processos apresentado na Tabela 4 compreende contratações associadas às características próprias do Sistema de Telemática do Exército — SisTEx, realizadas, em sua maioria, de forma centralizada, em âmbito nacional e destinadas ao atendimento de organizações militares distribuídas por todo o território nacional. Essas aquisições envolvem soluções de TIC de elevada complexidade, integração entre hardware, software, licenciamento, garantia e serviços especializados, além de requisitos de padronização, interoperabilidade, suporte e sustentação compatíveis com as necessidades institucionais do Exército Brasileiro. Por essa razão, o conjunto analisado constitui o espaço amostral mais adequado para subsidiar a presente análise e a tomada de decisão, uma vez que reflete, de maneira mais próxima, as particularidades técnicas, operacionais, administrativas e comerciais das contratações da Força. Embora processos conduzidos por outros órgãos ou entidades possam fornecer informações complementares relevantes, tais contratações não apresentam, necessariamente, o mesmo grau de similaridade ou aderência às especificidades do SisTEx e às condições de execução de compras nacionais centralizadas destinadas a todo o Exército.

**17.3.13.** Considerando os resultados das contratações analisadas, o espaço amostral disponível, as características específicas de cada objeto e a inexistência de dados concretos que demonstrem efetiva vantajosidade para a Administração na adoção do orçamento sigiloso, esta Equipe de Planejamento recomenda que a presente contratação seja realizada **sem orçamento sigiloso** e mediante a utilização do **modo de disputa aberto**. Essa configuração permite que os participantes apresentem lances públicos e sucessivos durante toda a etapa competitiva, amplia a possibilidade de competição efetiva e mitiga o risco de exclusão prematura de empresas tecnicamente capacitadas. A adoção dessas condições é considerada, sob o ponto de vista técnico e do planejamento da contratação, medida adequada para ampliar a transparência, favorecer a competitividade e reduzir os riscos capazes de comprometer o sucesso do procedimento licitatório.

**17.3.14.** Recomenda-se, ainda, que o cronograma físico-financeiro da contratação não estabeleça proporção para a parcela correspondente ao hardware e ao software e à garantia, porém, considerados conjuntamente deverão alcançar o percentual definido para a combinação dos itens. Essa divisão deverá ser adotada para fins de planejamento, medição, pagamento e adequado tratamento tributário dos componentes da solução, sem prejuízo de sua caracterização como item único, integrado e funcionalmente indivisível. A falta de proporcionalidade decorre da elevada variação identificada nas contratações anteriores e da inexistência de uma relação histórica predominante que permita adotar, com segurança, percentual definido.

**17.4. Sobre a estruturação do objeto em item único, critérios de pagamento, inclusão em carga e empenhos parcelados**

**17.4.1.** A presente contratação deverá ser estruturada como **item único**, denominado **Solução de Hardware e Software para Processamento de Inteligência Artificial**, composto pelo conjunto integrado de equipamentos, softwares, licenças, garantias, suporte, serviços especializados, operação assistida, capacitação, integração e demais componentes necessários à efetiva disponibilização da solução de GPU como serviço para o Exército Brasileiro.

**17.4.2.** A opção por item único decorre da própria natureza da solução pretendida. A Administração não busca a aquisição isolada de servidores com GPU, licenças independentes ou serviços autônomos, mas a entrega de uma solução integrada, funcional, validada, suportada e apta a operar como plataforma corporativa de processamento de Inteligência Artificial sob governança institucional.

**17.4.3.** A divisão do objeto em itens autônomos poderia comprometer a finalidade da contratação, pois aumentaria o risco de incompatibilidade entre componentes, indefinição de responsabilidades, conflitos de suporte entre fabricantes e fornecedores, dificuldades de integração, atrasos na implantação, lacunas de garantia e transferência indevida, para a Administração, do papel de integradora da solução.

**17.4.4.** No caso concreto, os componentes de hardware, software, licenciamento, suporte e serviços especializados possuem forte interdependência técnica e operacional. O servidor com GPU somente atenderá à necessidade institucional se estiver integrado à camada de orquestração, aos softwares de aceleração, aos mecanismos de autosserviço, à observabilidade, à segurança, ao suporte, à governança, à capacitação e à operação assistida. Da mesma forma, os softwares e serviços somente produzirão resultado prático se forem implantados sobre infraestrutura compatível, validada e suportada.

**17.4.5.** Assim, a estruturação em item único não tem por objetivo restringir a competitividade, mas assegurar que a disputa ocorra em torno da entrega do resultado pretendido pela Administração: uma solução completa de hardware e software para processamento de Inteligência Artificial, entregue em regime integrado, com responsabilidade clara da contratada pela compatibilidade, funcionamento, implantação, suporte e sustentação inicial da plataforma.

**17.4.6.** Essa diretriz encontra respaldo na lógica de planejamento da Lei nº 14.133/2021, segundo a qual o parcelamento deve ser adotado quando for tecnicamente viável e economicamente vantajoso, devendo a Administração considerar a responsabilidade técnica, o custo de múltiplos contratos, a ampliação da competição e a prevenção de concentração de mercado. No presente caso, embora a competição seja preservada no nível da solução integrada, o parcelamento autônomo dos componentes não se mostra tecnicamente adequado, pois poderia comprometer a responsabilidade única pela entrega da solução de GPU como serviço.

**17.4.7.** A existência de subcomponentes técnicos, descritos nas especificações, não altera a natureza unitária da contratação. Os itens, subitens, camadas, serviços, licenças e componentes descritos no Estudo Técnico Preliminar, no Termo de Referência e na Especificação Técnica deverão ser compreendidos como partes integrantes de uma única solução, utilizadas para fins de caracterização técnica, planejamento, medição, fiscalização, recebimento, composição de preço, pagamento e controle patrimonial.

**17.4.8.** Para fins de julgamento, adjudicação e contratação, deverá prevalecer o valor global do item único, sem prejuízo da apresentação, pela licitante vencedora, de planilha de composição de preços que discrimine os valores dos componentes internos da solução, incluindo hardware, garantias, softwares, licenças, suporte, serviços especializados, operação assistida, capacitação e demais elementos previstos no objeto.

**17.4.9.** A planilha de composição de preços não terá por finalidade transformar a contratação em múltiplos itens independentes, mas permitir maior transparência na formação do preço, possibilitar a aferição de exequibilidade, subsidiar a fiscalização contratual, orientar os eventos de pagamento, permitir a correta instrução dos recebimentos e apoiar os registros patrimoniais e contábeis da solução.

**17.4.10.** Para fins de pagamento contratual, o Termo de Referência deverá prever que o valor total da solução será distribuído em percentuais correspondentes aos principais grupos de entrega, sem prejuízo da manutenção do objeto como item único. Esses percentuais deverão refletir a participação relativa de cada grupo no valor total contratado e deverão constar da proposta vencedora, observada a estrutura mínima definida pela Administração.

**17.4.11.** A distribuição percentual deverá contemplar, no mínimo, os seguintes grupos de composição: hardware da solução; garantias e suporte de hardware; softwares, licenças e subscrições; serviços especializados de planejamento, desenho, implantação, integração, configuração e documentação; treinamentos e capacitações; operação assistida; suporte mensal do fabricante; suporte mensal do fornecedor; e treinamento avançado ou atualização tecnológica, quando aplicável.

**17.4.12.** Os pagamentos deverão estar vinculados a eventos objetivos de entrega, ativação, implantação, aceite e prestação de serviços, sendo vedado o pagamento por objeto não entregue, não ativado ou não recebido. A liquidação da despesa deverá observar a efetiva comprovação da entrega ou da prestação correspondente, conforme as regras da Lei nº 4.320/1964 e os critérios de recebimento provisório e definitivo definidos no instrumento convocatório e no contrato.

**17.4.13.** No caso dos componentes físicos, o pagamento deverá estar condicionado à entrega, conferência, instalação, quando aplicável, verificação de conformidade e recebimento pela Administração. No caso de softwares, licenças e subscrições, o pagamento deverá estar condicionado à comprovação da ativação, disponibilização, vinculação à conta institucional, validade do período contratado e acesso aos respectivos portais, mídias, chaves, contratos de suporte ou instrumentos equivalentes.

**17.4.14.** No caso dos serviços especializados, o pagamento deverá ser vinculado à conclusão e ao aceite dos entregáveis correspondentes, tais como reuniões iniciais, questionários, documentos de requisitos, HLD, LLD, plano de implementação, documentação de instalação, implantação, operação assistida, manuais, treinamentos, relatórios, transferência de conhecimento, monitoramento e encerramento do projeto.

**17.4.15.** No caso dos suportes mensais, caso previstos como prestação continuada ao longo de 12 (doze) meses, o pagamento deverá observar a periodicidade e as condições de prestação definidas no

Termo de Referência, podendo ser realizado de forma mensal ou conforme cronograma físico-financeiro aprovado, desde que haja comprovação da disponibilidade do serviço, dos canais de atendimento, dos contratos de suporte, dos níveis de serviço e das demais obrigações assumidas.

**17.4.16.** A estruturação do pagamento por percentuais e eventos não descaracteriza a natureza unitária do objeto. Trata-se de mecanismo de medição e fiscalização da execução contratual, destinado a assegurar que a Administração pague de forma proporcional aos componentes efetivamente entregues, ativados, implantados ou prestados, sem fragmentar a adjudicação nem transferir para a Administração a responsabilidade de integrar componentes fornecidos por contratados distintos.

**17.4.17.** A caracterização do objeto como item único para fins de licitação, adjudicação e responsabilidade contratual não afasta a necessidade de classificação contábil e patrimonial individualizada dos componentes. Os bens físicos deverão ser incorporados ao patrimônio pelos respectivos valores de aquisição, acrescidos dos custos diretamente atribuíveis à sua colocação em condições de funcionamento, quando cabível. Licenças, subscrições, garantias, suportes, treinamentos e demais serviços deverão receber o tratamento contábil próprio de sua natureza, conforme orientação dos setores de patrimônio, contabilidade e controle interno.

**17.4.18.** A planilha de composição de preços deverá permitir a segregação dos valores necessários à correta classificação patrimonial, contábil e orçamentária, sem descaracterizar a responsabilidade única da contratada pela entrega da solução integrada.

**17.4.19.** A Administração poderá, para fins de controle interno, rastreamento, inventário, manutenção, suporte e gestão de ciclo de vida, registrar informações auxiliares relativas a número de série, modelo, part number, licenças, contratos de suporte, períodos de garantia, softwares ativados, serviços executados e componentes vinculados à solução. Tais registros auxiliares não afastam a diretriz de que a inclusão em carga deverá considerar o valor total da solução contratada, observadas as orientações dos setores competentes de patrimônio, contabilidade e controle interno.

**17.4.20.** Considerando a previsão de execução parcial ainda no exercício de 2026 e continuidade de execução em 2027, a contratação poderá ser acompanhada de empenhos parciais, parcelados ou sucessivos, compatíveis com a disponibilidade orçamentária de cada exercício, sem que isso altere a natureza de item único da solução contratada.

**17.4.21.** A possibilidade de empenhos parcelados decorre da distinção entre a unidade do objeto contratado e a execução orçamentária da despesa. A contratação poderá ter objeto único e contrato único, ao mesmo tempo em que os empenhos sejam emitidos de forma compatível com as parcelas executadas, os créditos disponíveis e o cronograma físico-financeiro aprovado.

**17.4.22.** Nos termos do art. 60 da Lei nº 4.320/1964, é vedada a realização de despesa sem prévio empenho, sendo permitido, no § 3º do mesmo artigo, o empenho global de despesas contratuais e outras sujeitas a parcelamento. Dessa forma, a despesa contratual poderá ser empenhada de modo compatível com a execução parcelada, desde que previamente ao nascimento da obrigação correspondente e dentro dos limites do crédito orçamentário disponível.

**17.4.23.** Adicionalmente, o art. 27 do Decreto nº 93.872/1986 estabelece que as despesas relativas a contratos, convênios, acordos ou ajustes de vigência plurianual serão empenhadas, em cada exercício financeiro, pela parte nele a ser executada. Tal previsão permite que, em contratos cuja execução ultrapasse o exercício financeiro, o empenho seja realizado conforme a parcela de execução correspondente a cada exercício, preservando a regularidade orçamentária.

**17.4.24.** A Lei nº 14.133/2021 também é compatível com essa lógica, ao prever, em seu art. 105, que a duração dos contratos será a prevista em edital, devendo ser observada, no momento da contratação e a cada exercício financeiro, a disponibilidade de créditos orçamentários. O art. 150 da mesma lei dispõe que nenhuma contratação será feita sem a caracterização adequada do objeto e sem a indicação dos créditos orçamentários para pagamento das parcelas contratuais vincendas no exercício em que for realizada a contratação.

**17.4.25.** Assim, para o exercício de 2026, o empenho deverá limitar-se às parcelas compatíveis com o crédito orçamentário disponível e com a execução prevista para o exercício. As parcelas cuja execução e pagamento estejam previstas para 2027 deverão observar a disponibilidade de créditos orçamentários próprios naquele exercício, a serem indicados e empenhados oportunamente, nos termos da legislação orçamentária e financeira aplicável.

**17.4.26.** Essa estrutura não constitui fracionamento indevido da contratação, pois não há divisão artificial do objeto nem tentativa de afastar modalidade licitatória, critério de julgamento ou regime jurídico aplicável. O objeto permanece único e integrado, enquanto a execução orçamentária se organiza por parcelas, eventos e exercícios financeiros, conforme o cronograma físico-financeiro e a disponibilidade de créditos.

**17.4.27.** O edital e o contrato deverão explicitar que a contratação será realizada por item único, com julgamento pelo valor global da solução, mas com exigência de planilha de composição de preços, critérios de medição por eventos, percentuais de pagamento, regras de recebimento e possibilidade de empenhos compatíveis com a disponibilidade orçamentária de cada exercício.

**17.4.28.** O cronograma físico-financeiro deverá ser elaborado de forma a compatibilizar a entrega dos componentes, a ativação das licenças, a execução dos serviços, a operação assistida, os suportes mensais, os treinamentos e os respectivos pagamentos com os créditos orçamentários disponíveis, evitando assunção de obrigação sem prévio empenho e assegurando a adequada liquidação da despesa.

**17.4.29.** A contratada deverá estar ciente de que a emissão de empenhos poderá ocorrer de forma parcelada, conforme a disponibilidade orçamentária e o cronograma de execução, sem que isso implique alteração do objeto, redução da responsabilidade global pela solução ou afastamento das obrigações de integração, compatibilidade, suporte e funcionamento previstas no contrato.

**17.4.30.** Dessa forma, a estruturação da contratação como item único, associada à definição de percentuais de pagamento, inclusão em carga pelo valor total da solução e possibilidade de empenhos parcelados por exercício e por evento de execução, mostra-se compatível com a natureza integrada do objeto, com a legislação orçamentária e financeira aplicável e com a necessidade de preservar a responsabilidade da contratada pela entrega de uma solução completa, funcional, validada e suportada.

### 17.5. Cronograma Físico-Financeiro

**17.5.1.** O cronograma físico-financeiro da contratação foi elaborado com a finalidade de relacionar a sequência prevista para a execução dos entregáveis às respectivas parcelas de pagamento, proporcionando à Administração uma visão consolidada da evolução física do objeto e da distribuição estimada dos desembolsos ao longo da vigência contratual.

**17.5.2.** Durante a realização dos estudos de mercado, os fornecedores Drive A e Teltec apresentaram propostas comerciais destinadas a subsidiar a estimativa de preços, a análise da composição da solução e a avaliação da viabilidade de execução do objeto. A proposta da Drive A permanece como

referência histórica em razão da posterior desistência da HPE/Drive A. A Teltec, após proposta inicial no valor de R$ 26.700.000,00, apresentou em 30 de julho de 2026 a Proposta TTEC3067/2026, versão 2.0, no valor global de R$ 19.977.500,00, que passa a representar a referência comercial mais atual dessa tratativa.

**17.5.3.** A proposta revisada da Teltec foi precedida de maior interação técnica durante a análise e o aperfeiçoamento das especificações. Além da atualização comercial, a empresa confirmou a viabilidade da distribuição físico-financeira proposta e informou não ter identificado, até aquele momento, requisitos impeditivos à participação com solução Cisco.

**17.5.4.** Por essas razões, a versão 2.0 da proposta da Teltec será utilizada como principal referência preliminar para a distribuição percentual dos valores entre os entregáveis. Essa utilização possui natureza exclusivamente metodológica e não representa escolha antecipada de fornecedor, direcionamento da contratação ou substituição da pesquisa de preços.

**17.5.5.** A proposta foi empregada como fonte de dados para estimar a participação econômica relativa de cada componente, permitindo que o cronograma físico-financeiro fosse construído a partir de valores concretos apresentados pelo mercado e de uma solução que se aproximou do escopo técnico pretendido pela Administração.

**17.5.6.** Além da análise das propostas comerciais, foi realizada consulta formal quanto à viabilidade dos prazos e da sequência de execução previstos. A consulta abrangeu tanto o fabricante Cisco, em conjunto com o fornecedor Teltec, quanto o fabricante HPE, em conjunto com o fornecedor Drive A.

**17.5.7.** A consulta aos dois conjuntos formados por fabricante e fornecedor teve por finalidade verificar se o prazo de fornecimento dos equipamentos, a elaboração dos documentos técnicos, a participação dos fabricantes, a implantação da solução, a operação assistida e o início dos serviços continuados poderiam ser executados conforme a sequência proposta pela Administração.

**17.5.8.** As consultas realizadas forneceram subsídios para avaliação do cronograma inicialmente proposto. A HPE/Drive A comunicou posteriormente sua desistência, razão pela qual sua proposta não constitui confirmação atual da exequibilidade do cronograma para aquela arquitetura. Por outro lado, a Teltec, com a Cisco em cópia, confirmou expressamente ser possível atender à distribuição percentual e ao cronograma físico-financeiro propostos e informou não ter identificado requisitos impeditivos à participação com solução Cisco. Essa manifestação constitui evidência de exequibilidade do cronograma para a arquitetura analisada, sem prejuízo de que os prazos definitivos sejam vinculados às condições do edital, à proposta vencedora e ao plano de execução aprovado após a contratação.

**17.5.9.** Para os itens 1 a 4, que representam os servidores e seus respectivos componentes de garantia e licenciamento, foi aplicada a relação de 50% para o componente de hardware e 50% para o componente de garantia, licenciamento e suporte pelo período estabelecido.

**17.5.10.** Essa distribuição decorre da metodologia já examinada neste Estudo, a partir da análise de contratações similares, e busca reconhecer que o resultado pretendido pela Administração não depende apenas da entrega física dos servidores, mas também da existência de garantia, licenciamento, atualização e suporte oficial durante o ciclo de vida contratado.

**17.5.11.** Assim, o valor agregado de cada conjunto formado pelo equipamento e por sua correspondente cobertura foi distribuído em partes iguais. No caso do servidor de processamento com GPU, os itens 1 e 2 receberam, o percentual de 34% do valor da contratação. Para o servidor destinado à hospedagem da camada de software, os itens 3 e 4 receberam o percentual de 10% do valor global. Embora a proposta considerada como referência, apresentada pela Teltec, tenha adotado

distribuição de 50% entre o componente de hardware e o componente correspondente de software, garantia, licenciamento e suporte, os pagamentos referentes aos componentes de cada conjunto serão realizados de forma conjunta, de modo a preservar a composição econômica apresentada pelo fornecedor e a adequada distribuição tributária aplicável às diferentes parcelas que integram a solução.

**17.5.12.** Após a aplicação dessa metodologia e da participação econômica relativa dos demais entregáveis, foram realizados arredondamentos nos percentuais. Os arredondamentos tiveram por finalidade facilitar a medição, o cálculo das parcelas, o acompanhamento financeiro e a fiscalização contratual, sem alterar o valor global da contratação.

**17.5.13.** A distribuição final preserva o total de 100% do valor contratado e mantém proporcionalidade entre a relevância econômica dos componentes, a complexidade técnica das atividades, o esforço necessário à sua execução e os resultados efetivamente entregues à Administração.

**17.5.14.** Com base nessas premissas, foi estabelecido o cronograma físico-financeiro constante da Tabela 5.

Tabela 5 — Cronograma físico-financeiro por entregável

| ENTREGÁVEL | Descrição | Qtd | Percentual | Previsão de pagamento em dias após assinatura do contrato |
|---|---|---|---|---|
| 1 | SERVIDOR DE PROCESSAMENTO COM GPU – Componente de Hardware | 1 | 34,00% | 180 |
| 2 | SERVIDOR DE PROCESSAMENTO COM GPU – Componente de Garantia e Licenciamento para 60 meses | 1 | | 180 |
| 3 | SERVIDOR PARA HOSPEDAGEM DA CAMADA DE SOFTWARE DA SOLUÇÃO – Componente de Hardware | 1 | 10,00% | 180 |
| 4 | SERVIDOR PARA HOSPEDAGEM DA CAMADA DE SOFTWARE DA SOLUÇÃO – Componente de Garantia e Licenciamento para 60 meses | 1 | | 180 |
| 5 | SOLUÇÃO DE SOFTWARE DE VIRTUALIZAÇÃO, CONTÊINERES E ORQUESTRAÇÃO para 12 meses | 1 | 7,00% | 180 |
| 6 | SOLUÇÃO DE SOFTWARE ESPECÍFICA PARA ACELERAÇÃO POR GPU para 12 meses | 1 | 2,00% | 180 |
| | | | | |

| | | | | |
|---|---|---|---|---|
| 7 | SOLUÇÃO DE SOFTWARE PARA INTERFACE DE SERVIÇOS E AUTOSSERVIÇO para 12 meses | 1 | 3,00% | 180 |
| 8 | SOLUÇÃO DE SOFTWARE DE OBSERVABILIDADE E SEGURANÇA OPERACIONAL para 12 meses | 1 | 3,00% | 180 |
| 9 | SERVIÇO DE INTEGRAÇÃO COM EQUIPAMENTOS DE REDE DE FRONTEND | 1 | 0,20% | 360 |
| 10 | REUNIÃO INICIAL | 1 | 0,10% | 5 |
| 11 | GERENCIAMENTO DE PROJETOS | 1 | 0,80% | 360 |
| 12 | PROPOSTA DE ESCOPO DE TRABALHO | 1 | 0,70% | 30 |
| 13 | LEVANTAMENTO DA SITUAÇÃO ATUAL | 1 | 0,50% | 45 |
| 14 | TOMADA DE CONHECIMENTO DA DOCUMENTAÇÃO EXISTENTE | 1 | 0,60% | 45 |
| 15 | APONTAMENTO DE OPORTUNIDADES DE MELHORIA | 1 | 0,60% | 60 |
| 16 | DEFINIÇÃO DE REQUISITOS DA SOLUÇÃO | 1 | 0,40% | 70 |
| 17 | TREINAMENTO INICIAL DA SOLUÇÃO | 1 | 0,40% | 70 |
| 18 | PLANEJAMENTO DE ALTO NÍVEL DA SOLUÇÃO | 1 | 3,40% | 100 |
| 19 | PROPOSTA TÉCNICA DE IMPLEMENTAÇÃO | 1 | 2,30% | 120 |
| 20 | DESENHO DE BAIXO NÍVEL DA SOLUÇÃO | 1 | 3,00% | 150 |
| 21 | PREPARAÇÃO DA DOCUMENTAÇÃO DE INSTALAÇÃO | 1 | 1,40% | 170 |
| | | | | |

| | | | | |
|---|---|---|---|---|
| 22 | ATUALIZAÇÃO DA DOCUMENTAÇÃO TÉCNICA EXISTENTE | 1 | 0,20% | 180 |
| 23 | TREINAMENTO ANTES DA IMPLEMENTAÇÃO | 1 | 0,20% | 200 |
| 24 | IMPLEMENTAÇÃO | 1 | 10,00% | 245 |
| 25 | TREINAMENTO APÓS A IMPLEMENTAÇÃO | 1 | 0,40% | 260 |
| 26 | ATUALIZAÇÃO DA DOCUMENTAÇÃO | 1 | 0,30% | 280 |
| 27 | MANUAL DE OPERAÇÕES | 1 | 1,00% | 280 |
| 28 | OPERAÇÃO ASSISTIDA | 1 | 3,00% | 340 |
| 29 | MONITORAMENTO DA SOLUÇÃO | 1 | 0,30% | 350 |
| 30 | ENCERRAMENTO DO PROJETO | 1 | 1,20% | 360 |
| 31 | SERVIÇO DE SUPORTE AVANÇADO MENSAL DO FABRICANTE | 12 | 3,50% | mensalidade a partir de 360 dias após início das atividades durante 12 meses, cada mensalidade correspondente a 1/12 de 3,50% do valor total |
| 32 | SERVIÇO DE SUPORTE MENSAL DO FORNECEDOR | 12 | 5,00% | mensalidade a partir de 360 dias após início das atividades durante 12 meses, cada mensalidade correspondente a 1/12 de 5,00% do valor total |
| 33 | SERVIÇO ESPECIALIZADO DE TREINAMENTO AVANÇADO E | 8 | 1,50% | de acordo com a prestação do serviço e calendário de eventos podendo ocorrer no período de 24 meses de |

| | | | | |
|---|---|---|---|---|
| | ATUALIZAÇÃO TECNOLÓGICA POR ALUNO | | | execução contratual, cada unidade correspondente a 0,1875% do valor total |
| Total | | | 100% | - |

**17.5.15.** A coluna relativa à previsão de pagamento em dias após a assinatura do contrato também representa a referência cronológica para a entrega dos produtos ou para a conclusão das atividades correspondentes. Essa indicação permite relacionar a evolução física da solução ao desembolso financeiro previsto.

**17.5.16.** Os prazos indicados deverão ser contados a partir da assinatura do contrato ou da emissão da ordem de início, conforme vier a ser definido no Termo de Referência e no instrumento contratual. Para fins da projeção mensal apresentada nesta seção, considerou-se que o início das atividades coincide com a assinatura do contrato.

**17.5.17.** A indicação de determinado prazo ou mês no cronograma não implica pagamento automático. Cada parcela permanecerá condicionada à efetiva execução do fornecimento ou serviço, à apresentação da documentação exigida, à realização dos testes aplicáveis, ao recebimento ou aceite pela fiscalização e ao cumprimento das demais condições previstas no contrato.

**17.5.18.** Caso um entregável seja concluído antes do prazo máximo previsto e esteja integralmente apto ao recebimento, o pagamento poderá ocorrer antecipadamente, desde que haja disponibilidade orçamentária e sejam observados os procedimentos de medição, liquidação e aceite.

**17.5.19.** Da mesma forma, eventual atraso na entrega, na validação ou no aceite de um produto deslocará o correspondente pagamento, sem que a simples previsão constante do cronograma gere direito ao recebimento antes do cumprimento integral da obrigação.

**17.5.20. Projeção mensal de desembolso**

**17.5.20.1.** Para demonstrar a distribuição estimada dos pagamentos ao longo do contrato, foi elaborada uma projeção mensal considerando, exclusivamente para fins de planejamento, a assinatura do contrato em **1º de novembro de 2026** e a contagem dos prazos em dias corridos.

**17.5.20.2.** Para os itens 31 e 32, considerou-se que os serviços continuados de suporte terão início após a conclusão da fase de implantação, aproximadamente 360 dias depois do início das atividades, e que o primeiro pagamento mensal ocorrerá após o primeiro mês efetivamente prestado. Dessa forma, projetaram-se 12 mensalidades entre novembro de 2027 e outubro de 2028.

**17.5.20.3.** Cada mensalidade do Entregável 31 corresponderá a 1/12 do percentual total de 3,50% atribuído ao entregável, enquanto cada mensalidade do Entregável 32 corresponderá a 1/12 do percentual total de 5,00%, resultando em desembolso mensal conjunto equivalente a aproximadamente **0,708333%** do valor global durante os 12 meses de suporte.

**17.5.20.4.** Para o item 33, que contempla oito participações em treinamentos avançados e eventos de atualização tecnológica, a data de execução dependerá dos calendários oficiais dos eventos e da efetiva prestação do serviço. Para a projeção orçamentária, as oito unidades do Entregável 33 foram

distribuídas trimestralmente entre janeiro de 2027 e outubro de 2028, correspondendo cada unidade a **0,1875%** do valor global.

**17.5.20.5.** Essa distribuição do item 33 possui caráter meramente estimativo e poderá ser alterada conforme as datas dos eventos, a conveniência administrativa e o planejamento de capacitação, sem alteração do percentual total de 1,50% reservado ao entregável.

**17.5.20.6.** Adotando-se como referência o valor global de **R$ 19.977.500,00**, a projeção mensal apresenta-se conforme a Tabela 6.

**Tabela 6 — Projeção mensal do desembolso contratual**

| Mês de referência | Entregáveis ou parcelas considerados | Percentual no mês | Valor estimado | Percentual acumulado |
|---|---|---|---|---|
| nov/26 | Item 10 | 0,1000% | R$ 19.977,50 | 0,1000% |
| dez/26 | Itens 12, 13, 14 e 15 | 2,4000% | R$ 479.460,00 | 2,5000% |
| jan/27 | Itens 16, 17 e uma unidade do item 33 | 0,9875% | R$ 197.277,81 | 3,4875% |
| fev/27 | Item 18 | 3,4000% | R$ 679.235,00 | 6,8875% |
| mar/27 | Itens 19 e 20 | 5,3000% | R$ 1.058.807,50 | 12,1875% |
| abr/27 | Itens 1 a 8, 21, 22 e uma unidade do item 33 | 60,7875% | R$ 12.143.822,81 | 72,9750% |
| mai/27 | Item 23 | 0,2000% | R$ 39.955,00 | 73,1750% |
| jun/27 | Sem desembolso programado | 0,0000% | R$ 0,00 | 73,1750% |
| jul/27 | Itens 24, 25 e uma unidade do item 33 | 10,5875% | R$ 2.115.117,81 | 83,7625% |
| ago/27 | Itens 26 e 27 | 1,3000% | R$ 259.707,50 | 85,0625% |
| set/27 | Sem desembolso programado | 0,0000% | R$ 0,00 | 85,0625% |
| out/27 | Itens 9, 11, 28, 29, 30 e uma unidade do item 33 | 5,6875% | R$ 1.136.220,31 | 90,7500% |
| nov/27 | 1ª mensalidade dos itens 31 e 32 | 0,708333% | R$ 141.507,29 | 91,458333% |
| dez/27 | 2ª mensalidade dos itens 31 e 32 | 0,708333% | R$ 141.507,29 | 92,166667% |
| jan/28 | 3ª mensalidade dos itens 31 e 32 e uma unidade do item 33 | 0,895833% | R$ 178.965,10 | 93,062500% |
| fev/28 | 4ª mensalidade dos itens 31 e 32 | 0,708333% | R$ 141.507,29 | 93,770833% |
| mar/28 | 5ª mensalidade dos itens 31 e 32 | 0,708333% | R$ 141.507,29 | 94,479167% |
| abr/28 | 6ª mensalidade dos itens 31 e 32 e uma unidade do item 33 | 0,895833% | R$ 178.965,10 | 95,375000% |
| mai/28 | 7ª mensalidade dos itens 31 e 32 | 0,708333% | R$ 141.507,29 | 96,083333% |
| | | | R$ | |

| | | | | |
|---|---|---|---|---|
| jun/28 | 8ª mensalidade dos itens 31 e 32 | 0,708333% | 141.507,29 | 96,791667% |
| jul/28 | 9ª mensalidade dos itens 31 e 32 e uma unidade do item 33 | 0,895833% | R$ 178.965,10 | 97,687500% |
| ago/28 | 10ª mensalidade dos itens 31 e 32 | 0,708333% | R$ 141.507,29 | 98,395833% |
| set/28 | 11ª mensalidade dos itens 31 e 32 | 0,708333% | R$ 141.507,29 | 99,104167% |
| out/28 | 12ª mensalidade dos itens 31 e 32 e uma unidade do item 33 | 0,895833% | R$ 178.965,14 | 100,000000% |
| **Total** | | **100,0000%** | **R$ 19.977.500,00** | **100,0000%** |

**17.5.20.7.** A projeção demonstra concentração relevante de desembolso em abril de 2027, mês em que se estima o pagamento dos itens 1 a 8, correspondentes aos equipamentos e às principais soluções de software, além dos entregáveis 21, 22 e de uma unidade do item 33.

**17.5.20.8.** O desembolso estimado para abril de 2027 corresponde a 60,7875% do valor global. Essa concentração decorre do prazo de até 180 dias previsto para a entrega dos componentes estruturantes da plataforma e deverá ser considerada no planejamento orçamentário do exercício de 2027.

**17.5.20.9.** Até o término da implantação e do encerramento do projeto, previsto para outubro de 2027, o desembolso acumulado estimado será de 90,75%. O saldo de 9,25% será destinado às mensalidades remanescentes dos Entregáveis 31 e 32 e às unidades remanescentes do Entregável 33.

**17.5.21. Compatibilidade com a previsão orçamentária de 2026 e o planejamento de recursos para 2027**

**17.5.21.1.** Considerando a assinatura do contrato no início de novembro de 2026 e o cumprimento dos prazos previstos, os pagamentos estimados para o exercício de 2026 corresponderão aos itens 10, 12, 13, 14 e 15.

**17.5.21.2.** Esses pagamentos representam, conjuntamente, 2,50% do valor total da contratação, correspondendo, com base no valor global de R$ 19.977.500,00, ao montante estimado de **R$ 499.437,50**.

**17.5.21.3.** O valor estimado para desembolso em 2026 encontra-se integralmente abrangido pela PRO de R$ 2.000.000,00, permanecendo margem estimada de **R$ 1.500.562,50** em relação ao crédito previsto para o exercício.

**17.5.21.4.** A compatibilidade entre a PRO e o desembolso inicialmente projetado demonstra a viabilidade preliminar de início das atividades em 2026. A assinatura do contrato e a emissão das ordens correspondentes permanecerão condicionadas à confirmação da disponibilidade orçamentária, à indicação dos créditos e ao prévio empenho das parcelas cuja execução gere obrigação no exercício, bem como à compatibilidade do compromisso plurianual com os instrumentos de planejamento aplicáveis.

**17.5.21.5.** Os pagamentos de maior vulto ocorrerão a partir do exercício de 2027 e dependerão da existência dos créditos orçamentários correspondentes às parcelas que vencerem em cada exercício financeiro.

**17.5.21.6.** Em complemento à Previsão de Recurso Orçamentário disponível para o exercício de 2026, encontra-se em curso a elaboração do Plano de Descentralização de Recursos no âmbito do Programa Estratégico do Exército Inteligência Artificial e Defesa Cibernética — PIADefCiber, destinado a orientar a descentralização dos recursos necessários à presente contratação no exercício financeiro de 2027.

**17.5.21.7.** O Plano de Descentralização de Recursos em elaboração contempla a integralidade da necessidade orçamentária estimada para a presente contratação no exercício de 2027, constituindo elemento relevante para o planejamento da continuidade da execução contratual e para o atendimento dos desembolsos de maior vulto previstos no cronograma físico-financeiro.

**17.5.21.8.** A inclusão da contratação no planejamento de descentralização de recursos do PIADefCiber demonstra seu alinhamento com as prioridades do Programa Estratégico e reforça a perspectiva de continuidade orçamentária do projeto, especialmente porque a maior parcela dos equipamentos, softwares e serviços especializados deverá ser entregue, executada e paga durante o exercício de 2027.

**17.5.21.9.** Essa previsão também reduz o risco de descontinuidade entre o início das atividades em 2026 e as etapas subsequentes de fornecimento, planejamento, implementação e sustentação da solução, permitindo que a Administração organize antecipadamente a distribuição dos créditos de acordo com os marcos físicos e financeiros da contratação.

**17.5.21.10.** O planejamento de recursos do PIADefCiber deverá considerar, em especial, a concentração estimada de desembolso em abril de 2027, quando se prevê o pagamento dos equipamentos e das principais soluções de software, bem como os pagamentos posteriores relacionados à implementação, à operação assistida, ao encerramento do projeto, ao suporte e à atualização tecnológica.

**17.5.21.11.** Registra-se, contudo, que o Plano de Descentralização de Recursos se encontra em fase de elaboração. A previsão nele contida não se confunde com a efetiva disponibilização do crédito orçamentário, que somente estará caracterizada após a aprovação do planejamento e a formalização da correspondente descentralização de recursos pelos instrumentos orçamentários aplicáveis.

**17.5.21.12.** Após a efetiva descentralização, a emissão dos empenhos deverá ser compatibilizada com o cronograma físico-financeiro, com os créditos disponíveis e com as parcelas cuja execução esteja prevista para cada exercício, observando-se que a previsão de recursos não autoriza pagamento antecipado nem afasta a necessidade de recebimento, medição, liquidação e aceite dos respectivos entregáveis.

**17.5.21.13.** Considerando que o cronograma estimativo contempla serviços continuados e atividades de capacitação com execução até o exercício de 2028, a Administração deverá avaliar, na etapa de consolidação orçamentária e contratual, o tratamento aplicável às parcelas que ultrapassarem o exercício de 2027, observando a legislação orçamentária, a disponibilidade de créditos em cada exercício e as regras relativas à emissão de empenhos e à inscrição de despesas em restos a pagar.

**17.5.21.14.** Assim, a estrutura orçamentária projetada para a contratação está apoiada em dois elementos complementares: a PRO de R$ 2.000.000,00, que oferece cobertura suficiente para os desembolsos inicialmente estimados para 2026, e o Plano de Descentralização de Recursos do PIADefCiber, em elaboração, que contempla a integralidade dos recursos planejados para a continuidade da contratação no exercício de 2027.

**17.5.21.15.** A existência desses dois instrumentos reforça a viabilidade orçamentária do planejamento, sem prejuízo da necessidade de confirmação da disponibilidade efetiva dos créditos antes da assunção das correspondentes obrigações e da permanente compatibilização entre execução física, empenho, liquidação e pagamento.

**17.5.21.16.** A projeção mensal deverá ser atualizada antes da fase externa da contratação e novamente após a assinatura do contrato, considerando a data efetiva de início, a proposta da empresa vencedora, o plano de execução aprovado, os calendários dos eventos do item 33 e a disponibilidade orçamentária dos exercícios envolvidos.

**17.5.21.17.** Eventuais alterações nas datas de entrega ou na sequência de execução poderão modificar a distribuição mensal dos pagamentos, mas não poderão alterar os percentuais totais dos entregáveis sem a correspondente formalização e observância das regras contratuais aplicáveis.

**17.5.21.18.** O cronograma deverá, portanto, ser utilizado como instrumento de planejamento, acompanhamento e controle, mantendo-se permanentemente vinculado à execução física, à medição, ao aceite dos entregáveis e à disponibilidade orçamentária da Administração.

**17.5.21.19.** Adicionalmente, por meio do DIEx nº 20668-SEGER/6 SCh/EME, de 27 de julho de 2026, o Estado-Maior do Exército informou a existência de Previsão de Recursos Orçamentários — PRO para o exercício de 2027, destinada à ampliação do ambiente de hospedagem do Data Center Corporativo do Exército, no valor total de R$ 40.000.000,00 (quarenta milhões de reais), na Natureza da Despesa 44.90.00. Desse montante, R$ 30.000.000,00 estão previstos no âmbito da Ação Orçamentária 156M — Modernização e Transformação Estratégica da Força Terrestre, e R$ 10.000.000,00 no âmbito da Ação Orçamentária 147F, Plano Orçamentário 0001 — Implantação do Sistema de Defesa Cibernética na esfera da Força Terrestre.

**17.5.21.20.** A referida PRO constitui elemento adicional de suporte ao planejamento orçamentário da presente contratação para o exercício de 2027, considerando sua vinculação à ampliação da capacidade do ambiente de hospedagem do Data Center Corporativo do Exército e a concentração, naquele exercício, das etapas de maior vulto financeiro previstas no cronograma físico-financeiro. A existência dessa previsão reforça a perspectiva de disponibilidade de recursos para a continuidade da execução contratual, sem que isso represente, nesta etapa, a efetiva disponibilização dos créditos ou autorização para assunção de obrigações além dos limites orçamentários formalmente descentralizados.

**17.5.21.21.** O próprio DIEx nº 20668-SEGER/6 SCh/EME condiciona as medidas subsequentes relacionadas à execução das despesas à evolução do processo, destacando a necessidade de manter o Estado-Maior do Exército informado quando o procedimento licitatório estiver concluído, os valores exatos dos objetos estiverem definidos e as Unidades Gestoras Executoras estiverem em condições de realizar os respectivos empenhos. Dessa forma, a utilização dos recursos previstos deverá permanecer compatibilizada com a conclusão da contratação, a definição do valor efetivamente contratado, a correspondente descentralização dos créditos e a emissão dos empenhos pelas Unidades Gestoras competentes.

**17.5.21.22.** Assim, a PRO para 2027, em conjunto com os demais instrumentos de planejamento orçamentário mencionados neste Estudo Técnico Preliminar, reforça a viabilidade orçamentária da contratação e oferece referência concreta para o planejamento dos desembolsos previstos para aquele

exercício, permanecendo a execução das despesas condicionada à efetiva disponibilização dos créditos, à observância do cronograma físico-financeiro e ao cumprimento das etapas de empenho, liquidação e pagamento aplicáveis.

**17.6. Excepcionalidade de indicação de marca e modelo**

**17.6.1.** Objetivo e delimitação da excepcionalidade da indicação de marca e modelo para as GPUs NVIDIA B300 e para o NVIDIA AI Enterprise.

**17.6.1.1.** A presente seção tem por objetivo consolidar as justificativas técnicas, operacionais e jurídicas para a indicação excepcional de marca e modelo aplicável a componentes específicos da solução de GPU como serviço, sem prejuízo das referências e justificativas complementares apresentadas nas demais seções deste Estudo Técnico Preliminar.

**17.6.1.2.** A excepcionalidade de indicação de marca e modelo ficará restrita aos seguintes componentes:

**17.6.1.2.1.** às unidades de processamento gráfico — GPUs integrantes do servidor de processamento acelerado, que deverão ser da marca **NVIDIA**, modelo **B300**; e

**17.6.1.2.2.** à Solução de Software Específica para Aceleração por GPU, correspondente ao Entregável 6, que deverá ser atendida pelo produto **NVIDIA AI Enterprise**, em versão corporativa, oficialmente licenciada e suportada pelo fabricante.

**17.6.1.3.** A restrição não alcançará a marca ou o modelo dos servidores que hospedarão as GPUs, nem os demais componentes de hardware, software, conectividade, armazenamento ou serviços da solução. Poderão ser ofertados servidores e produtos de diferentes fabricantes, desde que atendam integralmente às Especificações Técnicas, sejam oficialmente compatíveis com as GPUs NVIDIA B300 e com o NVIDIA AI Enterprise e constituam solução integrada, validada e suportada.

**17.6.1.4.** A indicação não representa escolha antecipada de fornecedor, revenda, distribuidor, integrador ou fabricante de servidor. A disputa permanecerá aberta entre todas as empresas capazes de fornecer, integrar, implantar e sustentar os produtos indicados, juntamente com os demais componentes necessários à entrega da solução completa.

**17.6.1.5.** A excepcionalidade deverá ser compreendida de forma estrita e limitada aos dois componentes expressamente identificados nesta seção. Para todos os demais componentes, deverão ser preservadas especificações funcionais e de desempenho abertas, admitindo-se produtos de quaisquer fabricantes que atendam objetivamente aos requisitos da contratação.

**17.6.2. Fundamentos legais e jurisprudenciais**

**17.6.2.1.** O art. 41, inciso I, da Lei nº 14.133, de 1º de abril de 2021, admite que, nas licitações que envolvam o fornecimento de bens, a Administração indique excepcionalmente uma ou mais marcas ou modelos, desde que a decisão seja formalmente justificada e esteja enquadrada em uma das hipóteses previstas na legislação.

**17.6.2.2.** No caso concreto, a indicação encontra fundamento no art. 41, inciso I, alínea **“b”**, da Lei nº 14.133/2021, em razão da necessidade de manter compatibilidade tecnológica e operacional com plataformas, padrões e conhecimentos já adotados ou em implantação pela Administração, especialmente o ecossistema NVIDIA e CUDA associado às GPUs NVIDIA L40S atualmente existentes no ambiente institucional.

**17.6.2.3.** A indicação também encontra fundamento no art. 41, inciso I, alínea **“c”**, da Lei nº 14.133 /2021, uma vez que, consideradas conjuntamente as necessidades de compatibilidade, capacidade de processamento e memória, arquitetura multi-GPU, suporte oficial, integração entre hardware e software, licenciamento empresarial e redução dos riscos de implantação, a combinação formada pelas GPUs NVIDIA B300 e pelo NVIDIA AI Enterprise foi identificada, durante o planejamento, como a única capaz de atender simultaneamente ao conjunto específico de necessidades da contratação, sendo os produtos comercializados por mais de um fornecedor, revenda ou integrador.

**17.6.2.4.** A decisão também guarda relação com o princípio da padronização previsto no art. 40, inciso V, alínea “a”, da Lei nº 14.133/2021, segundo o qual o planejamento das compras deverá considerar a compatibilidade de especificações técnicas e de desempenho, sem prejuízo da observância da isonomia, da competitividade, da motivação, do julgamento objetivo e da seleção da proposta mais vantajosa.

**17.6.2.5.** Em matéria de contratações de TIC, deverá ser observado o art. 2º do Decreto nº 7.174, de 12 de maio de 2010, que veda especificações que direcionem ou favoreçam a contratação de fornecedor específico, que não representem a real demanda de desempenho da Administração ou que não estabeleçam métodos objetivos de mensuração. A presente indicação não se destina a favorecer fornecedor determinado, mas a definir dois componentes técnicos considerados indispensáveis, permanecendo aberta a competição entre diferentes empresas capazes de fornecê-los e integrá-los à solução.

**17.6.2.6.** A definição também é compatível com a Instrução Normativa SGD/ME nº 94, de 23 de dezembro de 2022, que atribui à Equipe de Planejamento da Contratação a responsabilidade pela especificação dos requisitos tecnológicos, incluindo arquitetura, hardware, software, padrões de interoperabilidade, implantação, garantia, manutenção, segurança e demais elementos necessários à solução de TIC. A excepcionalidade decorre, portanto, da análise da arquitetura integral e dos riscos da contratação, e não de preferência comercial dissociada da necessidade institucional.

**17.6.2.7.** A Súmula nº 270 do Tribunal de Contas da União estabelece que, nas licitações referentes a compras, inclusive de softwares, é possível indicar marca quando a medida for estritamente necessária ao atendimento de exigências de padronização e estiver acompanhada de justificativa prévia. De forma complementar, o Acórdão nº 113/2016 — Plenário registra que a indicação deverá estar amparada em razões técnicas, motivadas e documentadas, que demonstrem a necessidade da marca específica para satisfação do interesse público.

### 17.6.3. Contexto tecnológico preexistente

**17.6.3.1.** O ambiente institucional já possui 12 (doze) GPUs NVIDIA L40S destinadas a projeto de Inteligência Artificial em fase de implementação. Embora essa capacidade ainda não constitua plataforma corporativa completa de GPU como serviço, sua existência demonstra que o Exército já iniciou a adoção do ecossistema tecnológico NVIDIA para processamento acelerado.

**17.6.3.2.** A infraestrutura existente e os projetos em andamento envolvem conhecimentos, procedimentos operacionais, bibliotecas, drivers, ferramentas, imagens de software, práticas de monitoramento e cargas de trabalho estruturadas ou planejadas sobre o ecossistema NVIDIA e CUDA.

**17.6.3.3.** A adoção de arquitetura distinta na implantação da primeira plataforma corporativa de GPU como serviço exigiria a criação e sustentação simultânea de ecossistemas diferentes, com procedimentos próprios de instalação, configuração, atualização, monitoramento, treinamento, diagnóstico, suporte e desenvolvimento de aplicações.

**17.6.3.4.** Essa fragmentação poderia reduzir o aproveitamento dos conhecimentos já adquiridos, elevar o esforço de capacitação, ampliar a quantidade de combinações tecnológicas a serem validadas e transferir para a Administração riscos adicionais de compatibilidade e sustentação.

**17.6.3.5.** A compatibilidade considerada nesta seção não se limita à possibilidade de funcionamento físico do acelerador no servidor. Abrange a compatibilidade tecnológica e operacional entre aceleradores, drivers, runtimes, bibliotecas, frameworks, contêineres, operadores, ferramentas de orquestração, observabilidade, suporte e procedimentos utilizados durante todo o ciclo de vida da solução.

**17.6.4. Justificativa da escolha da marca NVIDIA**

**17.6.4.1.** Durante a análise das alternativas de aceleradores ofertadas por NVIDIA, AMD e Intel, verificou-se que a implantação de uma solução corporativa de GPU como serviço não poderia ser avaliada apenas mediante comparação isolada de capacidade computacional do hardware.

**17.6.4.2.** A solução exige um ecossistema completo, formado por drivers, runtimes, bibliotecas, frameworks, ferramentas de gerenciamento, operadores para ambientes de contêineres, mecanismos de particionamento, recursos de observabilidade, microserviços de inferência, documentação, atualizações e suporte empresarial.

**17.6.4.3.** No estágio atual do planejamento e com base nas informações obtidas durante as tratativas de mercado, o ecossistema NVIDIA apresentou maior maturidade e disponibilidade de componentes compatíveis com os softwares corporativos e com os modelos de operação considerados para a plataforma, proporcionando maior previsibilidade para implantação, operação, suporte e expansão.

**17.6.4.4.** Grande parte dos frameworks, bibliotecas e soluções empregados nas cargas de Inteligência Artificial consideradas pela Administração encontra-se estruturada ou otimizada para utilização do ecossistema CUDA. A manutenção dessa arquitetura reduz a necessidade de conversão, adaptação, revalidação ou substituição de componentes desenvolvidos ou planejados sobre esse padrão.

**17.6.4.5.** A escolha não decorre da afirmação genérica de que somente a NVIDIA produz aceleradores aptos à execução de Inteligência Artificial. Decorre da conclusão de que, no contexto específico da presente contratação, a arquitetura NVIDIA é a única identificada como capaz de atender simultaneamente aos requisitos de compatibilidade preexistente, integração com a camada lógica, suporte empresarial, padronização operacional, implantação multi-GPU e mitigação dos riscos técnicos considerados pela Administração.

**17.6.4.6.** Dessa forma, a indicação da marca NVIDIA constitui medida de compatibilidade e redução de risco para a implantação inicial da plataforma, e não preferência subjetiva baseada exclusivamente em notoriedade, participação de mercado ou conveniência comercial.

**17.6.5. Justificativa da escolha do modelo NVIDIA B300**

**17.6.5.1.** A escolha específica do modelo NVIDIA B300 decorre da necessidade de disponibilizar capacidade adequada para workloads de maior porte, inferência intensiva, treinamento distribuído, ajuste fino de modelos, processamento documental, ambientes interativos e demais cargas que possam exigir grande capacidade computacional e de memória por acelerador.

**17.6.5.2.** A seleção considera a estratégia de implantação de servidor com 8 (oito) GPUs, na qual a comunicação de alta velocidade entre os aceleradores instalados no mesmo nó é necessária para suportar cargas multi-GPU, reduzir a fragmentação da capacidade entre diferentes servidores e proporcionar maior previsibilidade de desempenho.

**17.6.5.3.** A maior densidade computacional por servidor permite melhor aproveitamento do espaço físico, da alimentação elétrica, da refrigeração e da conectividade disponíveis no Data Center, além de reduzir a quantidade de nós necessária para disponibilizar determinada capacidade de processamento.

**17.6.5.4.** A adoção da B300 também busca preservar maior longevidade tecnológica para o investimento, evitando que a configuração inicial seja implantada sobre geração de aceleradores com menor capacidade de atender às demandas de médio prazo ou que exija redesign arquitetural prematuro nas etapas futuras de expansão.

**17.6.5.5.** A escolha do modelo não se baseia exclusivamente em uma métrica isolada de desempenho. Foram considerados, conjuntamente, capacidade computacional, memória disponível, arquitetura multi-GPU, compatibilidade com o software corporativo, suporte oficial, integração com servidores homologados, possibilidade de expansão e redução do risco de obsolescência.

**17.6.5.6.** Por essas razões, não será admitida, para o componente GPU, a substituição por outro modelo apresentado apenas como similar, equivalente ou de desempenho superior em determinado benchmark, uma vez que a decisão considera o conjunto de características técnicas, operacionais, arquiteturais e de compatibilidade que fundamentam a indicação excepcional.

**17.6.6. Justificativa da escolha do NVIDIA AI Enterprise**

**17.6.6.1.** O hardware acelerado, isoladamente, não é suficiente para disponibilizar GPU como serviço. É necessária uma camada de software capaz de habilitar, administrar, particionar, monitorar e disponibilizar os recursos das GPUs às cargas de trabalho, de forma integrada à plataforma de virtualização, contêineres e orquestração.

**17.6.6.2.** O NVIDIA AI Enterprise constitui plataforma corporativa de software voltada ao desenvolvimento, à implantação e ao gerenciamento de aplicações de Inteligência Artificial em ambientes de Data Center, nuvem e borda. Sua composição abrange componentes da camada de aplicação e da camada de infraestrutura, incluindo drivers, operadores, ferramentas de gerenciamento de workloads, mecanismos de particionamento e componentes construídos sobre CUDA e CUDA-X, com suporte empresarial, correções de segurança e atualizações de manutenção.

**17.6.6.3.** A exigência busca assegurar que os componentes necessários à operação das GPUs sejam disponibilizados em versões oficialmente validadas e suportadas pelo fabricante, com matriz de compatibilidade, documentação técnica, atualização, correções de segurança, acesso ao suporte empresarial e mecanismos formais de escalonamento de incidentes.

**17.6.6.4.** A documentação oficial do NVIDIA AI Enterprise contempla suporte à arquitetura NVIDIA B300 em configuração de oito GPUs, inclusive com recursos relacionados a virtualização, particionamento e operação de workloads acelerados, reforçando a integração tecnológica entre o hardware indicado e a camada de software exigida.

**17.6.6.5.** Softwares comunitários, componentes isolados ou produtos de terceiros poderão implementar funcionalidades específicas semelhantes às encontradas no NVIDIA AI Enterprise. Contudo, não foram identificados como substitutos capazes de fornecer, sob responsabilidade e suporte oficial do

fabricante das GPUs, a mesma combinação de drivers, bibliotecas, operadores, ferramentas de gerenciamento, matrizes de compatibilidade, atualizações e suporte empresarial exigida para o Entregável 6.

**17.6.6.6.** A indicação do NVIDIA AI Enterprise não alcança as demais camadas de software. Permanecerá aberta a escolha dos produtos destinados à virtualização, contêineres, orquestração, autosserviço, observabilidade e segurança operacional, desde que sejam compatíveis, integrados, licenciados e oficialmente suportados na arquitetura ofertada.

**17.6.6.7.** O licenciamento deverá abranger a totalidade das GPUs e dos ambientes previstos no escopo, durante o período definido nas Especificações Técnicas, com direito de utilização dos componentes necessários, atualizações, correções e suporte oficial correspondente.

**17.6.7. Atendimento à hipótese do art. 41, inciso I, alínea “b”**

**17.6.7.1.** A hipótese prevista na alínea “b” encontra-se atendida porque a Administração já adotou e se encontra em processo de implementação de capacidade baseada em GPUs NVIDIA L40S, acompanhada do correspondente ecossistema tecnológico.

**17.6.7.2.** A manutenção da arquitetura NVIDIA permite preservar compatibilidade com os padrões já adotados ou em implantação, incluindo conhecimentos técnicos, procedimentos operacionais, bibliotecas, frameworks, imagens de contêineres, processos de atualização, mecanismos de observabilidade e formas de suporte.

**17.6.7.3.** A compatibilidade reduz a necessidade de manter ambientes paralelos, desenvolver procedimentos específicos para diferentes fabricantes de aceleradores, duplicar capacitações e realizar novas validações para cada combinação de driver, biblioteca, framework e ferramenta de gerenciamento.

**17.6.7.4.** A exigência também contribui para a padronização da primeira plataforma corporativa de GPU como serviço, permitindo que a Administração concentre os esforços iniciais de capacitação, governança, segurança, observabilidade e suporte em uma arquitetura coerente e controlada.

**17.6.7.5.** A medida é proporcional porque não estende a padronização ao fabricante dos servidores nem aos demais componentes da solução, restringindo-se ao acelerador e à camada de software diretamente vinculada ao seu funcionamento corporativo.

**17.6.8. Atendimento à hipótese do art. 41, inciso I, alínea “c”**

**17.6.8.1.** A hipótese prevista na alínea “c” encontra-se atendida porque a avaliação realizada não identificou outra combinação de marca, modelo e software capaz de satisfazer, simultaneamente, o conjunto específico de requisitos estabelecido pela Administração, especialmente:

**17.6.8.1.1.** compatibilidade com o ecossistema NVIDIA já adotado ou em implantação;

**17.6.8.1.2.** capacidade para execução de workloads de maior porte e cargas multi-GPU;

**17.6.8.1.3.** elevada capacidade computacional e de memória por acelerador;

**17.6.8.1.4.** implantação em servidor de alta densidade com 8 (oito) GPUs;

**17.6.8.1.5.** disponibilização de camada de software oficialmente suportada pelo fabricante das GPUs;

**17.6.8.1.6.** integração com ambientes de virtualização, contêineres e orquestração;

**17.6.8.1.7.** mecanismos de particionamento, compartilhamento, gerenciamento e observabilidade dos recursos acelerados;

**17.6.8.1.8.** acesso a atualizações, correções de segurança, matrizes de compatibilidade e suporte empresarial; e

**17.6.8.1.9.** aproveitamento dos conhecimentos, padrões, ferramentas e cargas de trabalho relacionados ao ecossistema tecnológico já adotado.

**17.6.8.2.** A conclusão de exclusividade técnica refere-se ao atendimento simultâneo desse conjunto de requisitos e não à capacidade genérica de outros aceleradores executarem aplicações de Inteligência Artificial.

**17.6.8.3.** Os produtos indicados não são comercializados por fornecedor único. As tratativas realizadas demonstraram a existência de diferentes fabricantes de servidores, revendas, distribuidores e integradores capazes de estruturar ofertas baseadas no ecossistema NVIDIA, inclusive mediante arquiteturas associadas a Cisco, HPE e Supermicro, sem prejuízo da participação de outros fabricantes e canais que atendam às condições do edital.

**17.6.8.4.** A existência de diferentes canais de fornecimento preserva a disputa entre empresas, ainda que os dois componentes críticos permaneçam vinculados à marca e ao modelo indicados.

**17.6.8.5.** Antes da abertura da fase externa, deverá ser juntada aos autos comprovação atualizada de que as GPUs NVIDIA B300 e o NVIDIA AI Enterprise podem ser fornecidos por mais de uma empresa independente, mediante consultas, propostas, declarações de canais, documentação pública ou outros elementos idôneos.

**17.6.9. Interdependência entre as GPUs e o software de aceleração**

**17.6.9.1.** As indicações da GPU NVIDIA B300 e do NVIDIA AI Enterprise não constituem duas escolhas independentes. Ambos integram uma única decisão arquitetural destinada a assegurar compatibilidade entre o hardware acelerador e a camada corporativa de software responsável por sua habilitação, administração e disponibilização aos workloads.

**17.6.9.2.** A utilização de GPU de outro fabricante exigiria correspondente substituição ou adaptação de drivers, runtimes, bibliotecas, operadores, ferramentas de gerenciamento e mecanismos de suporte, descaracterizando a arquitetura de referência e ampliando os riscos de integração.

**17.6.9.3.** De forma semelhante, a utilização das GPUs B300 sem a camada corporativa oficialmente suportada pelo fabricante transferiria para a contratada ou para a Administração a responsabilidade pela seleção, integração e sustentação de componentes isolados, aumentando o risco de incompatibilidades, lacunas de suporte e indefinição de responsabilidades.

**17.6.9.4.** A associação entre os dois componentes permite estabelecer matriz clara de compatibilidade, responsabilidade técnica, atualização, correção de vulnerabilidades e escalonamento de incidentes, coerente com a exigência de solução integrada e do tipo *turn-key*.

**17.6.10. Preservação da competitividade e ausência de direcionamento a fornecedor**

**17.6.10.1.** A indicação de marca e modelo não deverá ser confundida com indicação de fornecedor. A licitação permanecerá competitiva entre todas as empresas capazes de fornecer os produtos indicados e entregar a solução integralmente aderente às exigências técnicas, comerciais e contratuais.

**17.6.10.2.** Não será exigido servidor de marca determinada. A licitante poderá ofertar servidor de qualquer fabricante, desde que a configuração com 8 (oito) GPUs NVIDIA B300 esteja oficialmente homologada ou qualificada, disponha de suporte para toda a configuração e atenda aos requisitos de desempenho, conectividade, garantia e expansão previstos nas Especificações Técnicas.

**17.6.10.3.** Também não haverá restrição de marca ou modelo para o servidor destinado à camada de gerenciamento, processadores, memórias, unidades de armazenamento, interfaces, equipamentos de rede e demais componentes físicos.

**17.6.10.4.** Ressalvado o NVIDIA AI Enterprise, os demais softwares poderão ser ofertados por diferentes fabricantes, desde que atendam às funcionalidades, integrações, condições de licenciamento, suporte oficial e requisitos de segurança estabelecidos.

**17.6.10.5.** As tratativas de mercado foram realizadas com Dell, Lenovo, HPE, Cisco e Supermicro, com o objetivo de conhecer diferentes arquiteturas e permitir o aperfeiçoamento das especificações. A consulta a vários fabricantes e a disponibilização prévia do documento de requisitos constituem medidas destinadas a ampliar a transparência e reduzir o risco de especificação construída para fornecedor determinado.

**17.6.10.6.** A participação dos fabricantes nas tratativas não representa habilitação prévia, preferência ou garantia de atendimento. Todos os licitantes deverão comprovar integralmente a conformidade da solução ofertada durante o procedimento licitatório.

**17.6.10.7.** Não será exigida carta de exclusividade emitida pela própria licitante ou declaração de que seu produto seja exclusivo no mercado. A demonstração do atendimento às hipóteses legais decorre dos estudos, comparações, documentos técnicos e evidências reunidos pela Administração.

**17.6.10.8.** Registra-se que, durante o planejamento, a HPE/Drive A comunicou sua desistência de prosseguir na participação relacionada à contratação, em razão do volume de customizações e integrações de produtos de terceiros que entendeu necessário para atendimento ao conjunto das especificações. Essa manifestação não se relaciona, isoladamente, à indicação das GPUs NVIDIA B300 ou do NVIDIA AI Enterprise, uma vez que as arquiteturas de diferentes fabricantes analisadas durante as tratativas também se apoiaram no ecossistema NVIDIA. A desistência será considerada, contudo, na avaliação global da competitividade da solução antes da abertura da fase externa.

**17.6.10.9.** Em sentido complementar à manifestação da HPE/Drive A, a Teltec, com a Cisco em cópia, informou posteriormente não ter identificado, até aquele momento, requisitos impeditivos à participação com solução Cisco e confirmou a viabilidade da distribuição percentual e do cronograma físico-financeiro propostos. Tal manifestação constitui evidência de que a excepcionalidade relativa às GPUs NVIDIA B300 e ao NVIDIA AI Enterprise, juntamente com os demais requisitos da solução, é tecnicamente atendível pela arquitetura analisada, sem que isso represente demonstração de exclusividade, escolha antecipada de fornecedor ou comprovação, por si só, de competição suficiente no futuro certame.

**17.6.11. Mitigação da dependência excessiva de fornecedor**

**17.6.11.1.** Embora a contratação estabeleça indicação excepcional de dois componentes, a arquitetura deverá preservar mecanismos de interoperabilidade, portabilidade de dados, exportação de configurações, uso de interfaces documentadas, integração por APIs e adoção de padrões amplamente utilizados.

**17.6.11.2.** Os dados, modelos, imagens, artefatos, notebooks, configurações e demais elementos produzidos ou utilizados pela Administração não poderão ser mantidos exclusivamente em formatos que impeçam sua exportação ou migração para outras plataformas.

**17.6.11.3.** A restrição deverá ser compreendida como medida aplicável à implantação inicial e à formação da primeira base corporativa da solução, não constituindo decisão definitiva de que todas as futuras expansões deverão necessariamente manter o mesmo fabricante ou modelo.

**17.6.11.4.** As expansões futuras deverão ser precedidas de nova análise técnica e de mercado, considerando a evolução das tecnologias, as métricas reais de utilização, os custos de migração, a compatibilidade com a plataforma implantada, os riscos operacionais e a possibilidade de adoção de especificações mais abertas.

**17.6.11.5.** O próprio ETP registra que a estratégia inicial busca reduzir o risco de implantação, preservando para etapas futuras a possibilidade de contratações mais abertas e competitivas, desde que mantidas a compatibilidade e a coerência arquitetural da plataforma.

**17.6.12. Comprovações a serem previstas no Termo de Referência**

**17.6.12.1.** O Termo de Referência deverá declarar expressamente que as únicas indicações obrigatórias de marca e modelo são as GPUs NVIDIA B300 e o NVIDIA AI Enterprise.

**17.6.12.2.** A licitante deverá identificar precisamente os equipamentos, softwares, licenças, códigos de produto, quantidades e períodos de suporte que integram sua proposta.

**17.6.12.3.** A conformidade deverá ser demonstrada mediante documentação oficial dos fabricantes, incluindo catálogos, fichas técnicas, matrizes de compatibilidade, listas de sistemas qualificados ou certificados, documentos de licenciamento e condições de suporte.

**17.6.12.4.** A licitante deverá comprovar que a configuração do servidor com 8 (oito) GPUs NVIDIA B300 possui suporte oficial para a totalidade dos componentes ofertados e que as versões do NVIDIA AI Enterprise, do sistema operacional, da plataforma de virtualização ou contêineres e dos demais softwares constam das matrizes de compatibilidade aplicáveis.

**17.6.12.5.** Deverá ser comprovado que o licenciamento do NVIDIA AI Enterprise abrange a totalidade das GPUs e dos ambientes previstos na contratação e inclui suporte, atualização e correções durante o período exigido.

**17.6.12.6.** A Administração poderá realizar diligências junto à NVIDIA, aos fabricantes de servidores, distribuidores e canais de comercialização para confirmar a autenticidade da documentação, a cobertura das licenças, a compatibilidade, a disponibilidade dos produtos e a existência de suporte oficial.

**17.6.12.7.** Eventuais referências a NVIDIA ou a outras marcas em requisitos que não correspondam aos dois componentes restritos deverão ser interpretadas apenas como descrição de compatibilidade

ou referência arquitetural, admitindo-se soluções equivalentes quando não houver declaração expressa de obrigatoriedade.

### 17.7.13. Conclusão sobre a excepcionalidade

**17.7.13.1.** A indicação das GPUs NVIDIA B300 e do NVIDIA AI Enterprise encontra-se formalmente delimitada, tecnicamente motivada e vinculada às necessidades reais da Administração, não decorrendo de preferência subjetiva ou de intenção de beneficiar fornecedor, revenda, integrador ou fabricante de servidor específico.

**17.7.13.2.** O atendimento à alínea “b” do inciso I do art. 41 da Lei nº 14.133/2021 decorre da necessidade de preservar compatibilidade com o ecossistema NVIDIA e CUDA já adotado ou em implantação, permitindo o aproveitamento dos conhecimentos, padrões, ferramentas, cargas de trabalho e procedimentos institucionais existentes.

**17.7.13.3.** O atendimento à alínea “c” decorre da conclusão de que, no conjunto específico de requisitos definidos para a contratação, a combinação formada pelas GPUs NVIDIA B300 e pelo NVIDIA AI Enterprise é a única identificada como capaz de atender simultaneamente às exigências de capacidade, compatibilidade, integração, arquitetura multi-GPU, licenciamento empresarial, suporte oficial e mitigação dos riscos de implantação, sendo os produtos passíveis de comercialização por diferentes fornecedores.

**17.7.13.4.** A restrição mostra-se proporcional porque se limita aos componentes diretamente relacionados à arquitetura de processamento acelerado e ao seu suporte corporativo, permanecendo abertas as especificações dos servidores e dos demais componentes de hardware, software e serviços.

**17.7.13.5.** Dessa forma, considera-se juridicamente admissível e tecnicamente necessária a indicação excepcional da marca NVIDIA, modelo B300, e do produto NVIDIA AI Enterprise, nos termos do art. 41, inciso I, alíneas “b” e “c”, da Lei nº 14.133/2021, observadas as comprovações, atualizações e cautelas previstas nesta seção.

## 18. Declaração de Viabilidade

Esta equipe de planejamento declara **viável** esta contratação.

**18.1. Justificativa da Viabilidade**

### 18.1. Justificativa da Viabilidade

**18.1.1.** Considerando a necessidade institucional caracterizada neste Estudo Técnico Preliminar, a possibilidade de implantação progressiva, a existência de arquitetura técnica apta a atender aos requisitos de negócio e tecnológicos, as tratativas realizadas com o mercado, a manifestação da Teltec quanto à viabilidade do cronograma e à inexistência de requisitos impeditivos identificados para participação com solução Cisco, bem como a apresentação de proposta comercial revisada no valor global de R$ 19.977.500,00, a Equipe de Planejamento considera viável o prosseguimento da contratação da Solução 1.

## 19. Responsáveis

Todas as assinaturas eletrônicas seguem o horário oficial de Brasília e fundamentam-se no §3º do Art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Despacho: A solução escolhida, objeto do presente ETP, atende aos requisitos estratégicos, técnicos e de negócios, necessários para atender a demanda do SisTEx.

**MARCELO FIGUEIRA DE VASCONCELOS**

Integrante Técnico

*Assinou eletronicamente em 25/08/2026 às 07:34:18.*

Despacho: A solução escolhida, objeto do presente ETP, atende aos requisitos estratégicos, técnicos e de negócios, necessários para atender a demanda do SisTEx.

**ANDERSON LIMA MUNIZ BARRETTO**

Integrante Requisitante

*Assinou eletronicamente em 25/08/2026 às 10:33:58.*

Despacho: A solução escolhida, objeto do presente ETP, atende aos requisitos estratégicos e de negócios, necessários para atender a demanda do SisTEx.

**LOIANE GOMES PIMENTA**

Integrante Administrativo

*Assinou eletronicamente em 26/08/2026 às 09:50:52.*

Despacho: Aprovo o presente Estudo Técnico Preliminar; atesto sua conformidade às disposições da IN SGD/ME nº 94, de 23 de dezembro de 2022; e aprovo a padronização sugerida neste Estudo.

**RODOLFO ROQUE SALGUERO DE LA VEGA FILHO**

Autoridade Máxima de TIC

*Assinou eletronicamente em 24/08/2026 às 17:00:25.*

Despacho: Aprovo o presente Estudo Técnico Preliminar.

**RAFAEL MOREIRA DE OLIVEIRA**

Autoridade competente


*Assinou eletronicamente em 26/08/2026 às 16:28:18.*

# Lista de Anexos

Atenção: Apenas arquivos nos formatos ".pdf", ".txt", ".jpg", ".jpeg", ".gif" e ".png" enumerados abaixo são anexados diretamente a este documento.

- Anexo I - ETP - Anexo I - Especificação Técnica v 2.0 em 17 Agosto 26.pdf (875.64 KB)
- Anexo II - ETP - Anexo II - Tratativas com Fabricantes Completo 8AGO26 - Parte 1.pdf (9.94 MB)
- Anexo III - ETP - Anexo II - Tratativas com Fabricantes Completo 8AGO26 Parte 2.pdf (9.79 MB)
- Anexo IV - ETP - Anexo II - Tratativas com Fabricantes Completo 8AGO26 Parte 3.pdf (9.96 MB)
- Anexo V - ETP - Anexo II - Tratativas com Fabricantes Completo 8AGO26 Parte 4.pdf (9.83 MB)
- Anexo VI - ETP - Anexo II - Tratativas com Fabricantes Completo 8AGO26 Parte 5.pdf (9.45 MB)
- Anexo VII - ETP - Anexo II - Tratativas com Fabricantes Completo 8AGO26 Parte 6.pdf (9.87 MB)
- Anexo VIII - ETP - Anexo II - Tratativas com Fabricantes Completo 8AGO26 Parte 7.pdf (9.85 MB)
- Anexo IX - ETP - Anexo II - Tratativas com Fabricantes Completo 8AGO26 Parte 8.pdf (9.94 MB)
- Anexo X - ETP - Anexo II - Tratativas com Fabricantes Completo 8AGO26 Parte 9.pdf (859.49 KB)
- Anexo XI - ETP - Anexo III - Requisitos Solução GPU como Serviço em 17 Maio 26.pdf (693.06 KB)
- Anexo XII - ETP - Anexo IV - Orçamentos Recebidos v2.pdf (4.28 MB)