

Total estimated cost

R\$ 616.806,82 / mo

 [Download .csv](#)

 [Duplicate estimate](#)

Next steps

Request a custom quote

Connect with our sales team and learn about any eligible discounts.

[Contact sales](#)

Start your proof of concept

Sign up for Google Cloud, new customers get \$300 in free credits.

[Get started](#)



Cost Estimate Summary

As of 18 de mai. de 2026 • 14:29

Prices in BRL

R\$ 15,84

Secret Manager

R\$ 15,84

Service type Secret Manager

Access operations 10000

R\$ 0,00

Rotation notifications	30	R\$ 7,66
------------------------	----	----------

Active secret versions (across all locations)	30	R\$ 8,18
---	----	----------

AI & ML**R\$ 213.713,78****Gemini Models (Generative AI) R\$ 102.746,78**

Service type Gemini Models

Gemini 3.1 Pro Average input tokens for text	5000	R\$ 8.634,18
--	------	--------------

Gemini 3.1 Pro Average input tokens for image	2000	R\$ 3.453,67
---	------	--------------

Gemini 3.1 Pro Average output tokens for text (response and reasoning)	5000	R\$ 38.853,83
--	------	---------------

Gemini 3.1 Pro Average output tokens for image	1000	R\$ 51.805,10
--	------	---------------

Gemini 3.1 Pro Context window	=>200k tokens	N/A
-------------------------------	---------------	-----

Gemini 3.1 Pro Batch requests	true	N/A
-------------------------------	------	-----

Gemini 3.1 Pro Priority requests	false	N/A
----------------------------------	-------	-----

Gemini 3.1 Pro Requests per day	5000	N/A
---------------------------------	------	-----

Gemini 3.1 Flash Image Batch requests	false	N/A
---------------------------------------	-------	-----

Gemini 3.1 Flash Image Priority requests	false	N/A
--	-------	-----

Gemini 3.1 Flash-Lite Batch requests	false	N/A
--------------------------------------	-------	-----

Gemini 3.1 Flash-Lite Priority requests	false	N/A
---	-------	-----

Gemini 3.0 Pro Context window	<200k tokens	N/A
-------------------------------	--------------	-----

Gemini 3.0 Pro Batch requests	false	N/A
-------------------------------	-------	-----

Gemini 3.0 Pro Priority requests	false	N/A
----------------------------------	-------	-----

Gemini 3.0 Flash Batch requests	false	N/A
---------------------------------	-------	-----

Gemini 3.0 Flash Priority requests	false	N/A
------------------------------------	-------	-----

Gemini 2.5 Pro Context window	<200k tokens	N/A
-------------------------------	--------------	-----

Gemini 2.5 Pro Batch requests	false	N/A
-------------------------------	-------	-----

Gemini 2.5 Pro Priority requests	false	N/A
----------------------------------	-------	-----

Gemini 2.5 Flash Batch requests	false	N/A
---------------------------------	-------	-----

Gemini 2.5 Flash Priority requests	false	N/A
------------------------------------	-------	-----

Gemini 2.5 Flash Lite Batch requests	false	N/A
Gemini 2.5 Flash Lite Priority requests	false	N/A
Gemini Code Assist (Gemini Code Assist)		R\$ 91.971,79
Service type	Gemini Code Assist	
Seats per month	360	N/A
Edition	Enterprise	N/A
Subscription commitment	Annually	R\$ 91.971,79
Language Models (Generative AI)		R\$ 8.634,18
Service type	Language Models	
Gemini 1.0 Pro input Characters	10000	R\$ 2.158,55
Gemini 1.0 Pro output Characters	10000	R\$ 6.475,64
Gemini 1.0 Pro Requests per day	10000	N/A
PaLM 2 Text Unicorn batch discounts	false	N/A
Palm 2 Text Bison Add batch discounts	false	N/A
Imagen Models (Generative AI)		R\$ 10.361,02
Service type	Imagen Models	
Imagen 4 Ultra for Image Generation Requests per day	1000	R\$ 10.361,02
Imagen for Image Generation Images returned per request	1	N/A
Image Editing Images returned per request	1	N/A
Image Upscaling Images returned per request	1	N/A
COMPUTE		R\$ 287.341,60
Apigee		R\$ 53.593,44
Service type	Apigee	
Number of standard API calls	30 Million	R\$ 3.406,36
Number of extensible API calls	30 Million	R\$ 17.031,81
Environment type	Intermediate	R\$ 33.155,26

Number of environments	4	N/A
Number of regions	1	N/A
Number of proxy deployments	5	N/A
Analytics add-on	false	N/A
API Security add-on	false	N/A

GKE (Kubernetes Engine)		R\$ 44.623,60
--------------------------------	--	----------------------

Service type	GKE	
Number of Zonal clusters	1	R\$ 414,44
Boot disk type	SSD persistent disk	N/A
Boot disk size (GiB)	250 GiB	R\$ 10.857,78
Node-time	21900 Hours	N/A
Machine type	t2d-standard-4, vCPUs: 4, RAM: 16 GB	R\$ 33.351,38
Number of Nodes	30	N/A
Operating System / Software	Free: Ubuntu	N/A
Kubernetes Edition	GKE	N/A
Provisioning model	Regular	N/A
Enable Confidential GKE Nodes	false	N/A
Add GPUs	false	N/A
Region	Sao Paulo (southamerica-east1)	N/A
Committed use discount options	None	N/A

Cloud Run		R\$ 62.066,65
------------------	--	----------------------

Service type	Cloud Run	
CPU amount per instance	8 vCPU	R\$ 42.969,22
Memory amount per instance	32 GiB	R\$ 19.097,43
Use case	Model Fine-tuning: Fine-tuning open-source AI models	N/A
Region	southamerica-east1 (São Paulo) - Tier 2	N/A
Resource type	Job	N/A
Adjust your usage estimates	true	N/A

Number of executions per month	730 (1 per hour)	N/A
Execution time per task	1 minute	N/A
Adjust your billing model	false	N/A
Adjust your committed use discounts	false	N/A
Committed use discounts	None	N/A
Advanced options	true	N/A
Number of resources (e.g., services, jobs, or worker pools)	1,000	N/A
Adjust your instance size	false	N/A
Add GPUs	false	N/A
GPU Model	None	N/A
Number of GPUs per instance	0	N/A
Minimum number of instances	0	N/A
Adjust your network internet egress	false	N/A
Network internet data transfer out per month	0 Bytes (B)	N/A

COLAB (Compute Engine)	R\$ 2.052,91
-------------------------------	---------------------

Service type	Instances	
Instance-time	1460 Hours	N/A
Machine type	e2-standard-4, vCPUs: 4, RAM: 16 GB	R\$ 1.763,37
Boot disk type	SSD persistent disk	N/A
Boot disk size (GiB)	100 GiB	R\$ 289,54
Number of Instances	2	N/A
Operating System / Software	Free: Debian, CentOS, CoreOS, Ubuntu or BYOL (Bring Your Own License)	N/A
Provisioning Model	Regular	N/A
Threads per core	2 threads per core	N/A
Enable Confidential VM service	false	N/A
Add GPUs	false	N/A
Region	Sao Paulo (southamerica-east1)	N/A
Committed use discount options	None	N/A

SIGA (Compute Engine)	R\$ 2.876,93
Service type	Instances
Instance-time	730 Hours N/A
Machine type	n4-standard-8, vCPUs: 8, RAM: 32 GB R\$ 2.387,55
Instances using ephemeral public IP	1 R\$ 10,36
Instances using ephemeral public IP	1 N/A
Instances using static public IP	1 R\$ 10,36
Boot disk type	Hyperdisk Balanced N/A
Boot disk size (GiB)	650 GiB R\$ 468,66
Number of Instances	1 N/A
Operating System / Software	N/A
Free: Debian, CentOS, CoreOS, Ubuntu or BYOL (Bring Your Own License)	
Provisioning Model	Regular N/A
Threads per core	2 threads per core N/A
Enable Confidential VM service	false N/A
Add GPUs	false N/A
Region	Sao Paulo (southamerica-east1) N/A
Committed use discount options	None N/A
Processos (Cloud Run)	R\$ 36.902,06
Service type	Cloud Run
Network internet data transfer out per month	32 Gibibytes (GiB) R\$ 21,80
CPU amount per instance	1 vCPU R\$ 29.662,61
Memory amount per instance	512 MiB R\$ 1.544,93
Number of requests per month	10 million R\$ 5.672,73
Use case	Public API / Website: External application data access N/A
Region	southamerica-east1 (São Paulo) - Tier 2 N/A
Resource type	Service N/A
Adjust your usage estimates	false N/A
Traffic shape	

Daily peak and trough: Traffic builds and fades over 12 hours daily, becoming idle for the remaining 12 hours	N/A
Adjust your billing model false	N/A
Billing Charged only when processing requests. CPU is limited outside of requests.	N/A
Adjust your committed use discounts false	N/A
Committed use discounts None	N/A
Advanced options true	N/A
Number of resources (e.g., services, jobs, or worker pools) 250	N/A
Adjust your instance size false	N/A
Add GPUs false	N/A
GPU Model None	N/A
Number of GPUs per instance 0	N/A
Adjust how your service handles traffic false	N/A
Execution time per request 400 milliseconds	N/A
Number of concurrent requests per instance 20	N/A
Minimum number of instances 0	N/A
Adjust your network internet egress true	N/A

Processos de IA (Cloud Run)	R\$ 36.880,26
------------------------------------	----------------------

Service type Cloud Run	
CPU amount per instance 1 vCPU	R\$ 29.662,61
Memory amount per instance 512 MiB	R\$ 1.544,93
Number of requests per month 10 million	R\$ 5.672,73
Use case Public API / Website: External application data access	N/A
Region southamerica-east1 (São Paulo) - Tier 2	N/A
Resource type Service	N/A
Adjust your usage estimates false	N/A
Traffic shape	
Daily peak and trough: Traffic builds and fades over 12 hours daily, becoming idle for the remaining 12 hours	N/A
Adjust your billing model false	N/A

Billing	Charged only when processing requests. CPU is limited outside of requests.	N/A
Adjust your committed use discounts	false	N/A
Committed use discounts	None	N/A
Advanced options	true	N/A
Number of resources (e.g., services, jobs, or worker pools)	250	N/A
Adjust your instance size	false	N/A
Add GPUs	false	N/A
GPU Model	None	N/A
Number of GPUs per instance	0	N/A
Adjust how your service handles traffic	false	N/A
Execution time per request	400 milliseconds	N/A
Number of concurrent requests per instance	20	N/A
Minimum number of instances	0	N/A
Adjust your network internet egress	false	N/A
Network internet data transfer out per month	0 Bytes (B)	N/A

LLM (Cloud Run)	R\$ 11.478,08
------------------------	---------------

Service type	Cloud Run	
CPU amount per instance	4 vCPU	R\$ 7.946,36
Memory amount per instance	16 GiB	R\$ 3.531,72
Use case	Batch AI Inference: Asynchronous, GPU-accelerated LLM inference	N/A
Region	southamerica-east1 (São Paulo) - Tier 2	N/A
Resource type	Job	N/A
Adjust your usage estimates	true	N/A
Number of executions per month	500,000	N/A
Execution time per task	1 minute	N/A
Adjust your billing model	false	N/A
Adjust your committed use discounts	false	N/A
Committed use discounts	Compute Flexible CUD - 3 Year (46% discount)	N/A

Advanced options	false	N/A
Number of resources (e.g., services, jobs, or worker pools)	1	N/A
Adjust your instance size	false	N/A
Add GPUs	false	N/A
GPU Model	None	N/A
Number of GPUs per instance	0	N/A
Adjust how your service handles traffic	false	N/A
Minimum number of instances	0	N/A
Adjust your network internet egress	false	N/A
Network internet data transfer out per month	0 Bytes (B)	N/A

Patch de Dados (Cloud Run) R\$ 9.185,63

Service type	Cloud Run	
Network internet data transfer out per month	N/A	R\$ 0,96
CPU amount per instance	1 vCPU	R\$ 8.701,27
Memory amount per instance	512 MiB	R\$ 483,40
Use case	Server-side tagging: Run a Google Tag Manager server container	N/A
Region	southamerica-east1 (São Paulo) - Tier 2	N/A
Resource type	Service	N/A
Adjust your usage estimates	false	N/A
Inbound events per month	100,000	N/A
Outbound events per month	100,000	N/A
First party serving enabled (via server container)	true	N/A
Average events per session	4.5	N/A
Estimate type	Lower Bound Estimate	N/A
Billing	Charged for the entire lifecycle of instances. Full CPU for the entire lifetime of each instance.	N/A
Adjust your committed use discounts	false	N/A
Committed use discounts	Compute Flexible CUD - 3 Year (46% discount)	N/A
Advanced options	true	N/A

Number of resources (e.g., services, jobs, or worker pools)	25	N/A
Adjust your instance size	false	N/A
Add GPUs	false	N/A
GPU Model	None	N/A
Number of GPUs per instance	0	N/A
Adjust how your service handles traffic	false	N/A
Minimum number of instances	2	N/A

Batch AI (Cloud Run) R\$ 6.445,38

Service type	Cloud Run	
CPU amount per instance	2 vCPU	R\$ 5.800,85
Memory amount per instance	2 GiB	R\$ 644,54
Use case	Batch AI Inference: Asynchronous, GPU-accelerated LLM inference	N/A
Region	southamerica-east1 (São Paulo) - Tier 2	N/A
Resource type	Job	N/A
Adjust your usage estimates	true	N/A
Number of executions per month	730 (1 per hour)	N/A
Execution time per task	1 minute	N/A
Adjust your billing model	false	N/A
Adjust your committed use discounts	false	N/A
Committed use discounts	Compute Flexible CUD - 3 Year (46% discount)	N/A
Advanced options	true	N/A
Number of resources (e.g., services, jobs, or worker pools)	1,000	N/A
Adjust your instance size	true	N/A
Add GPUs	false	N/A
GPU Model	None	N/A
Number of GPUs per instance	0	N/A
Minimum number of instances	10	N/A
Adjust your network internet egress	false	N/A

Network internet data transfer out per month 0 Bytes (B) N/A

Worker pool (Cloud Run) R\$ 21.236,64

Service type Cloud Run

CPU amount per instance 1 vCPU R\$ 20.131,08

Memory amount per instance 512 MiB R\$ 1.105,56

Use case Custom: User-defined configuration N/A

Region southamerica-east1 (São Paulo) - Tier 2 N/A

Resource type Worker Pool N/A

Instance count 100 N/A

Adjust your billing model false N/A

Committed use discounts None N/A

Advanced options false N/A

Number of resources (e.g., services, jobs, or worker pools) 1 N/A

Adjust your instance size false N/A

Add GPUs false N/A

GPU Model None N/A

Number of GPUs per instance 0 N/A

Adjust how your service handles traffic false N/A

Minimum number of instances 0 N/A

Adjust your network internet egress false N/A

Network internet data transfer out per month 0 Bytes (B) N/A

DATA ANALYTICS R\$ 57.826,81

On-Demand (BigQuery) R\$ 3.247,74

Service type On-Demand

Amount of data queried 10 TiB R\$ 574,82

Active logical storage 20 TiB R\$ 2.672,92

Storage write API 100 GiB R\$ 0,00

Location type	Region	N/A
Location	Sao Paulo (southamerica-east1)	N/A
Pub/Sub		R\$ 223,13
Service type	Pub/Sub	
Amount of published data daily	5 GiB	N/A
Message delivery type	BigQuery	R\$ 31,51
Number of subscriptions	2	R\$ 84,32
Topic retention days	7	R\$ 53,65
Average retention window days	7	N/A
Number of subscriptions with retained acknowledged messages	1	R\$ 53,65
Number of snapshots used per month	30	N/A
Dataflow Classic (Dataflow)		R\$ 9.222,49
Service type	Dataflow Classic	
Amount of memory	15	N/A
Number of job run hours	365 Hours	R\$ 1.647,51
Number of vCPUs	4	N/A
Number of job run hours	365 Hours	R\$ 7.426,78
Storage amount	15 GiB	R\$ 148,20
Region	Iowa (us-central1)	N/A
Job Type	Batch	N/A
Enable Dataflow Shuffle	false	N/A
Number of worker nodes	16	N/A
Machine type	n1-standard-4, vCPUs: 4, RAM: 15 GB	N/A
Storage Type	SSD	N/A
Add GPUs	false	N/A
Editions (BigQuery)		R\$ 45.133,45
Service type	Editions	
Slot commitments	200	R\$ 39.786,32

Active logical storage	40 TiB	R\$ 5.347,14
Location type	Region	N/A
Location	Iowa (us-central1)	N/A
Edition	Enterprise	N/A
Maximum slots	Medium (200 slots)	N/A
Baseline slots	100	N/A
Commitment	1 Year	N/A
Average utilization of autoscale slots	50	N/A

DATABASES **R\$ 15.602,26**

AlloyDB for PostgreSQL (AlloyDB for PostgreSQL)		R\$ 15.602,26
Service type	AlloyDB for PostgreSQL	
Storage amount	2 TiB	R\$ 5.545,05
Backup amount	Equal to Storage amount	N/A
Storage amount	2 TiB	R\$ 1.848,63
Primary vCPU	4	R\$ 8.208,58
Region	Sao Paulo (southamerica-east1)	N/A
Machine series	C4A - Google Axion ARM - Best price-performance	N/A
Enable High Availability Configuration	true	N/A
Enable Read Pool (N2)	false	N/A
Enable Read Pool (C4A)	false	N/A
Enable Read Pool (C4)	false	N/A
Enable Read Pool (Z3 High Cache)	false	N/A
Enable Read Pool (Z3 Standard Cache)	false	N/A
Committed use discount options	None	N/A

NETWORKING **R\$ 16.635,15**

IP Address (Networking)	R\$ 2.258,62
Service type IP Address	
Number of IP addresses in use on standard VM instances 50	R\$ 1.015,38
Static IP addresses (assigned but unused) 20	R\$ 1.243,24
Region Sao Paulo (southamerica-east1)	N/A
Cloud Load Balancing (Networking)	R\$ 2.575,42
Service type Cloud Load Balancing	
Amount of inbound data 10 TB	R\$ 634,48
Amount of outbound data 1 GiB	R\$ 0,07
Amount of data processed 1 TB	R\$ 63,45
Number of forwarding rules 20	R\$ 1.089,98
Number of proxy instances 5	R\$ 787,44
Load balancer scope Regional	N/A
Region Sao Paulo (southamerica-east1)	N/A
Region Sao Paulo (southamerica-east1)	N/A
Cloud CDN	R\$ 533,72
Service type Cloud CDN	
South America 1 TiB	R\$ 523,22
All other destinations (including Mexico, Central America, and Middle East) 10 GiB	R\$ 5,11
Within each of Asia Pacific, South America, Middle East, Africa, and Oceania (including Hong Kong) 10 GiB	R\$ 1,14
Monthly cache lookup requests 1 Million	R\$ 4,26
Cloud VPN	R\$ 1.554,15
Service type Cloud VPN	
Number of tunnels 5	R\$ 1.554,15
Region Sao Paulo (southamerica-east1)	N/A
NAT Gateway (Networking)	R\$ 9.713,24

Service type NAT Gateway

Number of assigned VM instances 10

R\$ 1.864,98

Amount of data processed 30 TiB

R\$ 7.848,26

Cloud NAT type Private NAT

N/A

OPERATIONS**R\$ 3.535,69****Cloud Logging (Cloud Operations)****R\$ 3.462,45**

Service type Cloud Logging

Logging storage amount 1 TiB

R\$ 2.764,83

Logging retention duration 1

R\$ 697,62

Cloud Monitoring (Cloud Operations)**R\$ 73,24**

Service type Cloud Monitoring

Volume of metrics data (MiB) 200 MiB

R\$ 73,24

I know the volume (MiB) of metrics true

N/A

SECURITY**R\$ 1.788,34****Cloud Armor****R\$ 1.788,34**

Service type Cloud Armor

Number of requests (globally-scoped) 20 Million

R\$ 85,16

Number of security policies 50

R\$ 1.419,32

Number of rules 50

R\$ 283,86

Service tier Standard

N/A

STORAGE**R\$ 20.347,34****Cloud Storage****R\$ 20.347,34**

Service type Cloud Storage

Data Transfer within Google Cloud	100 GiB	R\$ 0,00
Total amount of storage	100 TiB	R\$ 20.347,34
Location type	Region	N/A
Location	Sao Paulo (southamerica-east1)	N/A
Storage class	Standard Storage	N/A
Source region	North America	N/A
Destination region	North America	N/A

Total estimated cost**R\$ 616.806,82 / mo**

The estimated fees provided by Google Cloud Pricing Calculator are for discussion purposes only and are not binding on either you or Google. Your actual fees may be higher or lower than the estimate.