

Metodologia de Estimativa de Custos Cloud (GCP)

Data Warehouse, BI e Plataformas de IA

1. Objetivo

Este documento define o **método padrão** para estimar custos de infraestrutura em Google Cloud Platform (GCP) para qualquer projeto que envolva:

- Integração de múltiplas fontes de dados
- Data Lake
- Data Warehouse
- BI, dashboards e agentes de IA

O objetivo é garantir que todas as propostas usem **critérios técnicos, auditáveis e replicáveis**.

2. Arquitetura base considerada

Todas as estimativas partem de uma arquitetura Lakehouse padrão:

Camada	Serviço	Função
Bronze	Cloud Storage	Dados brutos extraídos de APIs, bancos e arquivos
Silver	BigQuery	Dados tratados, normalizados e limpos
Gold	BigQuery	Métricas, BI, dashboards e IA
ETL	Cloud Run	Extração, transformação e carga diária

Essa arquitetura é assumida como **baseline** para qualquer cliente.

3. Quais custos entram na conta

Sempre consideramos três blocos:

Bloco	O que mede
BigQuery	Quanto de dados são armazenados e lidos
Cloud Storage	Quanto de dados brutos são guardados
Cloud Run	Quanto de processamento é usado para ETL

Outros custos (Power BI, Looker, licenças) são tratados separadamente.

4. Como estimar o volume de dados de um cliente

A estimativa parte de três perguntas:

4.1 Quantas fontes de dados existem?

Exemplos:

- Plataformas de mídia
- Analytics
- CRMs
- ERPs
- Social
- Sistemas internos

Cada fonte gera dados diariamente.

4.2 Desde quando os dados serão carregados?

Perguntas-chave:

- A ingestão começa do zero?
- Vai trazer histórico?
- A partir de que ano?

Exemplo:

- “Dados a partir de 2025”
- “Últimos 3 anos”
- “Somente dados novos”

Isso impacta diretamente o tamanho do Data Lake e do Data Warehouse.

4.3 Com que frequência os dados são atualizados?

Padrão mais comum:

- Atualização **diária**

Isso significa:

- 30 cargas por mês por fonte
 - Crescimento contínuo do histórico
-

4.4 Tabela de referência para estimar volume por tipo de fonte

Para permitir uma estimativa inicial rápida, usamos uma **régua padrão de volume** por tipo de fonte, baseada em projetos reais de marketing, CRM e analytics.

Esses valores representam o **volume bruto médio gerado por uma fonte por mês**.

Tipo de fonte	Volume bruto típico
Web Analytics (GA4, Firebase, etc)	20 a 50 GB / mês
Mídia paga (Google Ads, Meta Ads, TikTok, LinkedIn)	5 a 15 GB / mês por plataforma
Social orgânico (Instagram, Facebook, YouTube, etc)	2 a 8 GB / mês
CRM (RD, Hubspot, Pipedrive, Moskit, etc)	2 a 6 GB / mês
ERP / Vendas	5 a 20 GB / mês
Dados financeiros	1 a 5 GB / mês
Logs e eventos técnicos	10 a 50 GB / mês

Esses números já consideram:

- Histórico de campanhas
- Métricas, dimensões e eventos
- Dados incrementais diários

Eles são usados apenas para **dimensionamento inicial**.

4.5 Como estimar o volume total de um cliente

O cálculo é feito assim:

1. Identificar as fontes do cliente
2. Classificar cada fonte pelo tipo acima
3. Multiplicar pelo número de meses de histórico

Exemplo genérico

Cliente tem:

- GA4
- Google Ads
- Meta Ads
- TikTok Ads
- Um CRM

Histórico desde janeiro de 2025 até hoje: 12 meses

Estimativa:

Fonte	Tipo	Volume mensal
GA4	Web Analytics	40 GB
Google Ads	Mídia paga	10 GB
Meta Ads	Mídia paga	12 GB
TikTok Ads	Mídia paga	6 GB
CRM	CRM	4 GB

Total mensal \approx 72 GB

Total histórico (12 meses) \approx 864 GB \approx 0,86 TB

Esse valor alimenta diretamente o cálculo do **Data Lake (Cloud Storage)**.

4.6 Como converter volume bruto em BigQuery

Após processados (Silver + Gold), os dados normalmente ficam mais compactos.

Regra prática:

Volume BigQuery \approx 30% a 60% do volume bruto

Se o Data Lake tem 0,86 TB:

- BigQuery tende a ficar entre 260 GB e 520 GB
-

4.7 Como estimar leitura do BigQuery (BI, dashboards, IA)

O volume lido depende do uso de dashboards, relatórios e modelos de IA.

Regra prática:

Dados lidos por mês = 5 a 10 vezes o volume armazenado

Exemplo:

- 400 GB armazenados
- Totalizando: 2 a 4 TB lidos por mês

Esse número vai direto na calculadora do BigQuery.

4.8 Como estimar processamento (Cloud Run)

Cada fonte gera:

- 1 job por dia
- 30 execuções por mês

Regra prática:

Número de fontes	Execuções por mês
5 fontes	150 execuções
10 fontes	300 execuções
20 fontes	600 execuções

Quanto maior o volume de dados, maior o tempo de execução, mas o custo cresce quase linearmente.

Esse número é usado como input na calculadora do Cloud Run.

5. Como dimensionar o Data Lake (Cloud Storage)

O Data Lake guarda:

- Dados brutos
- Históricos
- Arquivos intermediários

A conta é:

Volume mensal estimado x Meses de histórico

Exemplo:

- 100 GB por mês
- 10 meses de histórico
- Totalizando: 1.000 GB (1 TB)

Esse número é usado diretamente na calculadora do Cloud Storage.

6. Como dimensionar o Data Warehouse (BigQuery)

O BigQuery tem dois tipos de custo:

6.1 Armazenamento

É o tamanho dos dados tratados (Silver + Gold).

Regra prática:

BigQuery Storage = 30% a 60% do volume bruto do Data Lake

Se o Data Lake tem 1 TB, o DW tende a ficar entre 300 GB e 600 GB.

6.2 Leitura (queries e BI)

Esse é o principal custo do BigQuery.

Regra prática:

Dados lidos por mês = 5 a 10 vezes o volume armazenado

Exemplo:

- 1 TB armazenado
- Totalizando: 5 a 10 TB lidos por mês (dashboards, relatórios, IA)

Esse valor entra na calculadora de BigQuery.

7. Como dimensionar o processamento (Cloud Run)

O Cloud Run executa:

- Pipelines de extração
- Transformações
- Cargas

Parâmetros usados:

- Execução diária
- Uma execução por fonte
- Volume médio de dados

Regra prática:

Para pipelines diários multi-fonte, o custo fica baixo e cresce quase linearmente com o volume.

Isso é simulado na calculadora usando:

- Tráfego periódico
 - Número de execuções por mês
 - Tempo médio de cada job
-

8. Como calcular o custo total

1. Calcular:
 - Cloud Storage
 - BigQuery Storage
 - BigQuery Leitura
 - Cloud Run
2. Somar os três:

Custo Total = Storage + BigQuery + ETL

Esse é o valor mensal da infraestrutura.

Para calcularmos esse valor, podemos usar diretamente a calculadora do GCP, encontrada no link: <https://cloud.google.com/products/calculator>

9. Como distribuir custo por fonte

Para projetos com múltiplas fontes, os custos são rateados proporcionalmente ao volume típico de dados que cada uma gera.

Fontes de analytics e mídia geram mais dados do que CRMs ou social.

O rateio permite:

- Mostrar quanto cada canal “custa”
 - Justificar investimentos
 - Criar visão de ROI por fonte
-

10. Como considerar créditos gratuitos do GCP

Novas contas do GCP recebem créditos iniciais.

Esses créditos:

- São usados automaticamente
- Cobrem os primeiros meses
- Atrasam o início da cobrança real

Eles **não mudam o custo real**, apenas o fluxo de faturamento inicial.

11. Uso eventual de Máquinas Virtuais (Compute Engine)

Em alguns projetos pode ser necessário incluir uma **máquina virtual (VM)** no Google Cloud para suportar integrações específicas, conectores legados ou processos que não funcionam bem em ambientes serverless.

Nesses casos, utiliza-se o **Compute Engine**.

O custo de uma VM depende basicamente de:

- Quantidade de CPU e memória
- Tipo de disco
- Tempo de uso mensal

Na prática:

Uma VM pequena, operando continuamente, costuma custar entre **US\$ 40 e US\$ 120 por mês**.

Quando aplicável, esse valor deve ser **somado ao custo total da infraestrutura**.
