04 GCP Pricing and Billing -KirkYagami

GCP primarily uses a pay-as-you-go billing approach. This means you pay for what you use. Once you consume a service, GCP keeps track of your usage, and you're billed accordingly.

Example Consumes

- **Compute Engine:** Billed for VM instances by the second (with a 1-minute minimum), based on the type of instance and the region.
- **Storage:** Costs vary by storage type (like Persistent Disk, Cloud Storage, etc.) and the amount of data stored.
- **Data Transfer:** You're typically charged for data going out of GCP to the internet or to other cloud providers.

Google Compute

[Pricing | Compute Engine: Virtual Machines (VMs) | Google Cloud](#)
Google Cloud Platform (GCP) offers robust cloud computing services through Compute Engine, which provides scalable and flexible virtual machines (VMs). Understanding the pricing and billing of GCP Compute Engine is essential for effective cloud cost management and budgeting.

1. Overview of Compute Engine Pricing

Compute Engine pricing is based on several factors, including VM instance type, machine configuration, usage duration, and additional features. Key pricing components are:

- **VM Instance Type**: The cost varies depending on the type and size of the VM instance you choose. Instances are categorized into general-purpose, compute-optimized, memory-optimized, and accelerator-optimized.
- **Machine Configuration**: Includes the number of vCPUs, amount of memory, and type of storage. Prices differ based on these specifications.
- **Usage Duration**: Compute Engine charges based on per-second billing, with a minimum of 1 minute. This allows for precise billing aligned with actual usage.
- **Additional Features**: Such as IP addresses, persistent disks, and network usage also influence costs.

2. VM Pricing Components

2.1. Instance Types and Pricing

- **General-Purpose**: Ideal for a wide range of applications. Examples include N1 and N2 machine types. Prices are based on the instance's vCPUs and memory.
- **Compute-Optimized**: Designed for compute-intensive workloads. Examples include C2 machine types. Higher cost reflects the increased compute power.
- **Memory-Optimized**: Suited for applications requiring high memory. Examples include M2 machine types. Pricing is higher due to the large memory allocation.
- **Accelerator-Optimized**: Includes GPU and TPU instances. Pricing is influenced by the type and number of accelerators used.

2.2. Storage Pricing

- **Persistent Disks**: Charged based on the size and type (standard or SSD). Prices are per GB per month.
- **Local SSDs**: Charged based on the number of SSDs and the duration of use.
- **Snapshot Storage**: Snapshots of persistent disks incur additional costs based on storage used.

2.3. Network Pricing

- **Egress Traffic**: Charged for data transferred out of GCP to the internet or other regions. Costs vary based on the volume and destination of the data.
- **Ingress Traffic**: Generally free, but some exceptions apply depending on the source.

3. Discounts and Pricing Plans

3.1. Sustained Use Discounts

Compute Engine provides automatic discounts for instances that run for a significant portion of the billing month. The discount increases with the duration of usage.

3.2. Committed Use Discounts

Offers substantial savings (up to 70%) for committing to a one-year or three-year term for VM instances. This is ideal for predictable workloads.

3.3. Preemptible VMs

Cost-effective option for batch processing or fault-tolerant workloads. Preemptible VMs are significantly cheaper than regular instances but can be terminated by GCP at any

time.

4. Real-World Scenarios

4.1. Scenario 1: Web Application Hosting

A company hosts a web application requiring a general-purpose VM with moderate traffic. They choose an N2 instance with 2 vCPUs and 8 GB of memory. The application requires 100 GB of persistent disk storage.

- **VM Cost**: Based on hourly rates for the N2 instance.
- **Persistent Disk Cost**: Based on the size and type of disk storage.
- **Network Cost**: Depending on egress traffic to users.

4.2. Scenario 2: Data Processing Pipeline

A data processing pipeline runs compute-intensive tasks and uses GPU acceleration. The pipeline operates continuously, and the team opts for a C2 instance with 4 vCPUs and an NVIDIA Tesla T4 GPU.

- **VM Cost**: Includes the cost of the C2 instance and the GPU.
- **Persistent Disk Cost**: For storing intermediate data and results.
- **Network Cost**: For transferring large datasets between storage and processing units.

4.3. Scenario 3: Development and Testing

A development team needs temporary VM instances for testing purposes. They choose preemptible VMs due to their lower cost.

- **VM Cost**: Based on the lower rate for preemptible VMs.
- **Storage Cost**: For temporary storage needs during testing.
- **Network Cost**: If significant data transfer is involved.

5. Cost Management and Optimization

- **Monitoring and Alerts**: Use Google Cloud's monitoring tools to track and set alerts for VM usage and costs.
- **Rightsizing**: Regularly review and adjust VM sizes based on actual usage to avoid overprovisioning.

- **Billing Reports**: Analyze detailed billing reports to identify cost drivers and optimize spending.

Conclusion

Understanding the intricacies of GCP Compute Engine pricing is crucial for managing cloud costs effectively. By leveraging different pricing models and optimizing resource usage, organizations can achieve significant savings while meeting their computing needs. Regular monitoring and adjustments based on real-world scenarios help ensure cost efficiency and financial control in cloud operations.

Google Cloud Storage Pricing

Please read the [documentation](#)

Cloud Storage pricing is calculated according to the following elements:

- **Data storage—** the volume of data retained in your buckets. Storage rates differ according to the storage class of your data and the place of your buckets.
- **Network usage—** the volume of data read from or moved from one bucket to another.
    - Free, if Data transfer within Google Cloud. [Cloud Storage](#)
- **Operation usage—** the activities you undertake in Cloud Storage, including listing the objects in the buckets.
- **Retrieval and early deletion fees—** relate to data retained in the Coldline, Archive, and Nearline storage classes.
- **Inter-region replication—** applies to data replicated over locations.

As a component of the Google Cloud Free Tier, Cloud Storage offers resources at no cost —up to a certain limit. Such usage limits are applicable during and after the free trial time. Monthly Limits for Free Usage are:

- **Network egress—** up to 1 GB for each Google Cloud Platform egress destination from North America (not including Australia or China)
- **Standard storage—** up to five GB-months
- Up to 5,000 Class A operations (active data operations like INSERT and UPDATE)
- Up to 50,000 Class B operations (passive data operations like GET)

In addition, Google Persistent Disks, which offer reliable, high-performance block storage that can be attached to Google Cloud VMs. Persistent Disks cost $0.040 per

GB/month for standard magnetic disks, $0.170 per GB/month for SSD, and offer additional options such as additional IOPS and multi-region redundancy.

A *retrieval fee* applies when you read, copy, move, or rewrite object data or metadata that is stored using Nearline storage, Coldline storage, or Archive storage. This cost is in addition to any network charges associated with reading the data.

The following table shows the retrieval rates for each storage class:

Standard storage	Nearline storage	Coldline storage	Archive storage
$0 per GB	$0.01 per GB	$0.02 per GB	$0.05 per GB

Cloud Storage Always Free usage limits

As part of the Google Cloud Free Tier, Cloud Storage provides resources that are free to use up to specific limits. These usage limits are available both during and after the free trial period. If you are no longer in the free trial period, usage beyond these Always Free limits is charged according to the pricing tables above.

Resource	Monthly Free Usage Limits1
Standard storage	5 GB-months
Class A Operations	5,000
Class B Operations	50,000
Data transfer	100 GB from North America to each Google Cloud Data transfer destination (Australia and China excluded)

Cloud Storage Always Free quotas apply to usage in `US-WEST1`, `US-CENTRAL1`, and `US-EAST1` regions. Usage is aggregated across these 3 regions. Always Free is subject to change.