



EXAMPLE REPORT

GCP Cost Analysis July 2024



Table of Contents

Table of Contents	2
Exec Summary	3
Recommendations	3
Compute Engine	3
Cloud SQL	4
Observations	4
Idle Compute Engine resources	4
Trending up Cloud Storage resources	5
Introduction	6
Compute Engine	7
Committed Use Discounts	7
FlexCUD Analysis	10
Cloud SQL	12
Spend-based committed use discounts	12
Cloud Storage	14
Storage Cost Trends	14
Using Autoclass	15
Lifecycle Rules	18
Storage Location	19
Actions and next steps	21
Tools and resources	23

Exec Summary

The findings in the document below highlight some key actions to save you money on your GCP bill. We've split these into key recommendations and observations. We've also created a new report for you called [Cost by Service, SKU & Project name](#) in order to help you with cost reporting in the long run.

Recommendations

Compute Engine

If you continue to have predictable & steady workloads, we recommend evaluating CUDs further. You can purchase another 1-year commitment in exchange for a substantial discount on your Compute Engine usage. Eg.

- Create an **additional** commitment of type **General Purpose N2** in region **eu-west1** for **217 vCPU** and **1245 GB** memory to save up to **\$19,609.20** a year with a 1-year commitment.

On the other hand, if you have less stable workloads that are not suitable for standard CUDs like above, consider using FlexCUDs. FlexCUDs are a spend-based (or flexible) commitment for Compute Engine of a minimum amount of hourly spend. You can create a FlexCUD on top of your current, existing resource-based CUDs and both will apply. Eg.

- Create a **new** commitment of type **Flexible CUDs** for up to **\$30.33/hr** to get a 28% discount over an 1-year commitment and save up to **\$73,374.36** a year across all regions.

FlexCUD pricing has a 9% premium over standard CUDs and can coexist with standard CUDs.

DoIT recommends having a mixture of resource-based and flexCUDs.



Cloud SQL

If you have a predictable steady-state workload, please re-evaluate CUDs, as you can purchase a one or a three-year commitment in exchange for a substantial discount on your Cloud SQL usage. Eg.

- Save **\$17,928** per year, by purchasing an **additional** hourly on-demand **1-year** commitment of **\$8.30** per hour in **europa-west1**.
- Save **\$2,656.80** per year, by purchasing an **additional** hourly on-demand **1-year** commitment of **\$1.23** per hour in **us-east4**.

Observations

Idle Compute Engine resources

Project: **example-uat**

- Save **\$790.39** per month by deleting the idle **project-euwest2-a---reservation** reservation in **europa-west2-a**.

Project: **example-microservices-qa**

- Save **\$258.39** per month by snapshotting and deleting the idle **dds-nfs-disk** persistent disk in **europa-west2-b**.

Project: **example-microservices-prd**

- Save **\$235.16** per month by snapshotting and deleting the idle **dd1-dsf-disk** persistent disk in **us-east4-b**.

Project: **example-johndoe**

- Save **\$198.27** per month by deleting the idle **asdfq123-disk2-poc** image in **all regions**.

Project: **example-testing**

- Save **\$105.32** per month by shutting down the idle **studio-ftp-02** VM in **europa-west1-b**.

Project: **example-api**

- Save **\$827.91** per month by snapshotting and deleting the following idle persistent disks;
 - ◆ **processor-machine** in europa-west4.



- ◆ **render-node-6** in europe-west4-a.
- ◆ **render-node-debug** in europe-west4-a.
- ◆ **render-node-worker** in europe-west4-a
- ◆ **render-node-assets-rendering** in europe-west4-a
- ◆ **render-node-2016-6-3712** in europe-west4-a
- ◆ **render-node-2016-9** in europe-west1-b
- ◆ **render-node-2016-6** in us-central1-a
- ◆ **render-node-2016-8** in us-central1-a
- ◆ **render-node-2016-7** in us-central1-a

Trending up Cloud Storage resources

The following buckets are upwards trending in storage bytes:

- folder-capture-draft
- photo-data-storage
- cloud-dev
- asd-asset-transfer
- storage-bucket-element
- lake-photography

Consider implementing lifecycle rules or looking into the [Autoclass](#) feature which helps reduce storage costs and achieve price predictability in a simple and automated way. **Autoclass is now supported for existing buckets too.**

There's also a higher spike in *Download Worldwide Destinations* SKU in the following buckets:

- Project: **example-microservices-prd**
 - ◆ alex-prd
 - ◆ john-prd
 - ◆ doe-prd
- Project: **example-test**
 - ◆ capture-testing

Introduction

We have examined the billing data across your GCP projects. The analysis is based on the billing data collected during the period of July 2024. At this point, we are focusing on Compute Engine, Cloud SQL and Cloud Storage costs as these represent the major part of your monthly spend and are an area where we can work together to apply some cost saving strategies.



[Service Breakdown](#)



Compute Engine

Committed Use Discounts

Effort ●●● Savings ●●●

Google Cloud offers committed use discounts (CUDs) in return for purchasing committed use contracts (also known as commitments). When you purchase a commitment, you commit either to a minimum amount of resource usage or to a minimum spend amount for a specified term of one or three years.

For Compute Engine, you receive deeply discounted prices for your VM instances in return for your 1-year or 3-year commitments. Depending on your resource usage requirements, you can purchase commitments and receive CUDs for Compute Engine resources in either of the following ways:

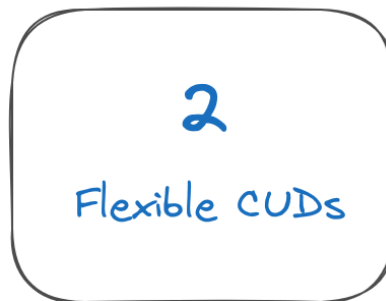
- Resource-based committed use discounts (or resource-based CUDs) are **ideal for predictable and steady state resource usage**. You receive these CUDs when you purchase a resource-based commitment and commit to use a minimum level of Compute Engine resources in a particular region. For more information, see [Resource-based CUDs](#).
- Compute Engine flexible committed use discounts (or flexible CUDs) are spend-based CUDs that are **ideal for scenarios where you have more predictable Google Cloud spend needs**. You receive *flexible CUDs* when you purchase a *spend-based (or flexible) commitment* for Compute Engine and commit to a minimum amount of hourly spend. For more information, see [Flexible CUDs](#).

	Compute Engine Standard CUDs	Compute Engine Flex CUDs
Scope	Purchased in a project by default Billing account CUDs can be enabled	Purchased on the billing account level by default
Purchase unit	Resource based (for example: N1 vcpu, GB memory, local SSD, GPU) Purchased in terms of the underlying resources	Spend based (for example: \$100) Purchased in terms of \$/hour of equivalent on-demand spend
Discount off on-demand rate	1 year discount 37% 3 year discount 55%	1 year discount 28% 3 year discount 46%
Machine family eligibility	Applies to a specific machine family	Applies to most general purpose and compute optimized machine families
Regional eligibility	Applies to a specific region	Applies to ALL regions

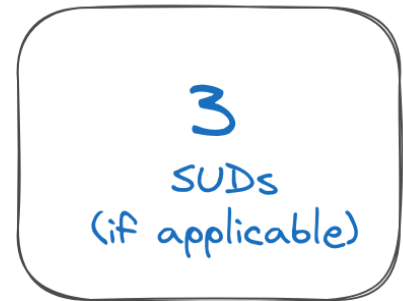
Order of application of discounts in billing*



Apply Standard CUDs to eligible usage (machine family and region specific)



On demand usage that wasn't discount by standard CUDs can be eligible for flex CUDs

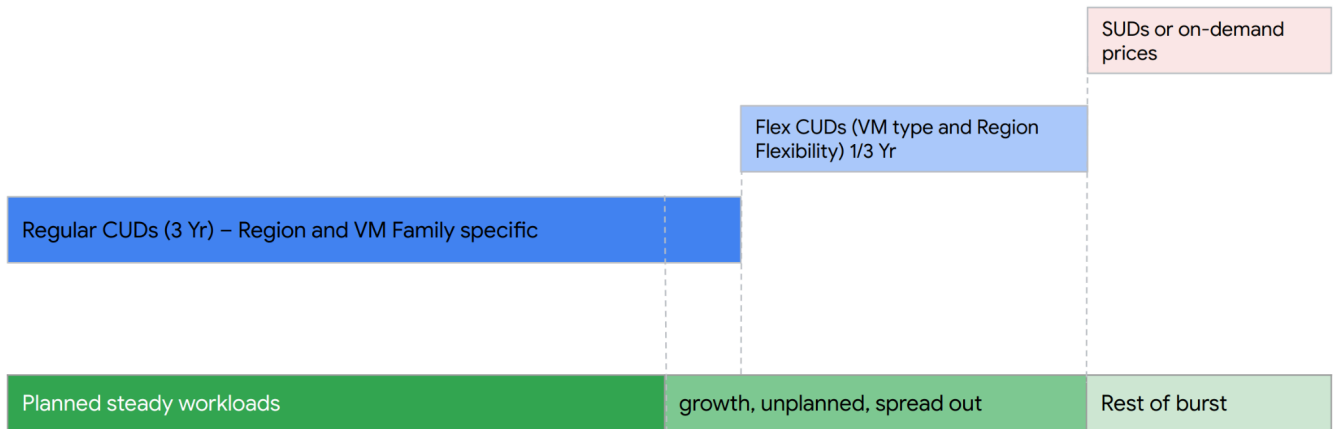


The remaining usage not covered by any CUD can be eligible for SUDs

*At any given point, a resource is eligible for only one kind of discount.



You can have a combination of Standard resource based CUDs to cover your most stable resource usage and Flexible spend based CUDs to cover your more flexible resource usage. Every hour, Standard CUDs will apply first to any eligible usage followed by flexible CUDs to optimize the use of your committed use discounts. Finally, any usage overage or usage not eligible to be covered by either type of CUDs, will be charged based on your on-demand rates.



Applying a mix of Standard and Flexible CUDs to maximize savings



FlexCUD Analysis

In July 2024, your total on-demand Compute Engine usage was \$43,154.00 - that's \$56.19 per hour. Compared to our last cost optimization exercise in December 2023, this is an increase of **44.76%**.

For the price of **\$30.33** per hour, you can benefit from savings on the above number (of **\$6,114.53** per month) on a 1-year commitment, which totals **\$73,374.36** in savings in a year!

Eligible usage for commitments - July 1 – August 1, 2024

Active commitment	Savings	Utilization	Coverage
None	None	N/A	N/A



Cost	Selected time range actual	Hourly average
--- Compute Engine Flexible commitment at on-demand rates		
Total cost at on-demand rates	\$43,154.00	\$56.19
--- Compute Engine Flexible commitment at on-demand rates	\$0.00	\$0.00
Compute Engine Flexible commitment cost	\$0.00	\$0.00
■ Compute Engine Flexible commitment covered costs	\$0.00	\$0.00
Eligible costs not covered by commitments	\$43,154.00	\$56.19

FlexCUD are automatically applying to all of your vCPU and RAM usage for all General Purpose and Compute-optimized VMs across all regions. This gives you the flexibility to use any machine type within these two VMs families across all regions, without having to modify your commitments.

Flex CUDs are ideal for customers with less predictable resource needs while having predictable cloud spend. It's ideal for situations where customers need to have flexibility to use GCE but don't want to be restricted to one region or one machine type. It's ideal for customers who want to decouple their cloud investment financial decision from the workload specific VM choice.

Customers will commit to a consistent amount of spend, measured in \$/hour of equivalent on-demand spend, for a one or three year term. In exchange they receive a discounted rate on the applicable usage covered by their commitment.

Eg. Company ABC commits to spending \$100 every hour for the next 3 years. They are able discount the N2,N1, E2 and C2 usage across us-central1 and us-east1.

Purchase a committed use discount

[LEARN](#)

Committing to an ongoing, minimum spend gives you steep product discounts. [Learn more](#)

Additional information

Product *

Compute Flexible Committed Use Discounts

You can purchase Committed Use Discounts by using the gcloud CLI, the [Google Cloud Console](#), or the API.

Select a billing account *

doit.itg.co.uk



Compute Flexible Committed Use Discounts

Google Compute services
[View product details](#)

Commitment details

Commitment name *

CostOpts_teamITG

What period?

1 year

Up to 28% discount

[Learn more](#)

3 years

Up to 46% discount

[Learn more](#)

Hourly on-demand commitment *

\$ 30.33

This commitment is based on the on-demand price. This commitment will apply to eligible vCPU and memory usage for all General-purpose and Compute-optimized machine types. [Learn more](#)

Commitment summary

Hourly commitment cost	\$21.8376
28% off on demand price	
30 day commitment cost	\$15,729.072
Estimated 30 day savings	\$6,114.528

Compute Flexible Committed Use Discounts

Current commitment

Use your historical usage to find the commitment level that's right for you. [View cost report](#)

Committed use discounts only apply to the [specific resources](#) in scope for the commitment you are purchasing.

You agree to pay the applicable [monthly fees](#) for the term selected above. The term will not automatically renew. These monthly fees are non-cancelable and will apply regardless of actual usage. [Learn more](#)

Please review the Google Cloud Platform's [Service Specific Terms](#) regarding Committed Units before proceeding.

PURCHASE

CANCEL

Cloud SQL

Spend-based committed use discounts

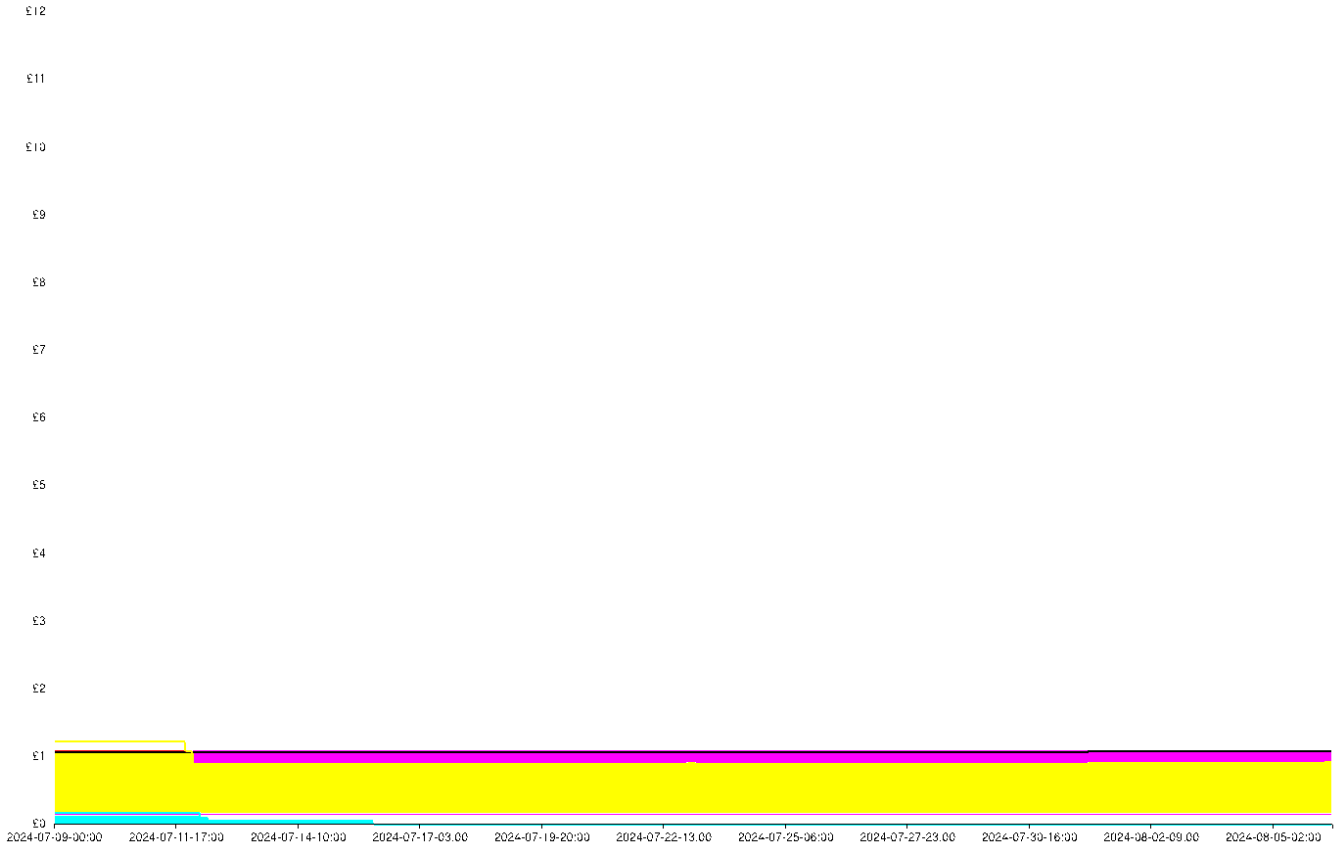
Committed use discounts (CUDs) for Cloud SQL provide deeply discounted prices in exchange for committing to continuously use database instances in a particular region for a one- or three-year term.

Key points to keep in mind:

- Cloud SQL committed use discounts only apply to vCPUs and memory.
- Cloud SQL committed use discounts do not apply to storage, backups, IP Addresses, network egress, or licensing.
- Cloud SQL committed use discounts apply to all instances in a given region.
- Cloud SQL committed use discounts are measured in dollars per hourly on-demand commitment.

You can purchase a Cloud SQL CUD on the [Google Cloud Console billing page](#). Select the Commitments tab, then select the PURCHASE action at the top. After purchasing a CUD, the commitment is effective starting within the next hour. The discounts are automatically applied to eligible usage in the region you specified.

Cloud SQL CUDs are measured in dollars per-hourly on-demand commitment. You can find the consumption per hour by using the DoIT Console Report as seen below.



[Cloud SQL CUD Eligible Consumption](#)

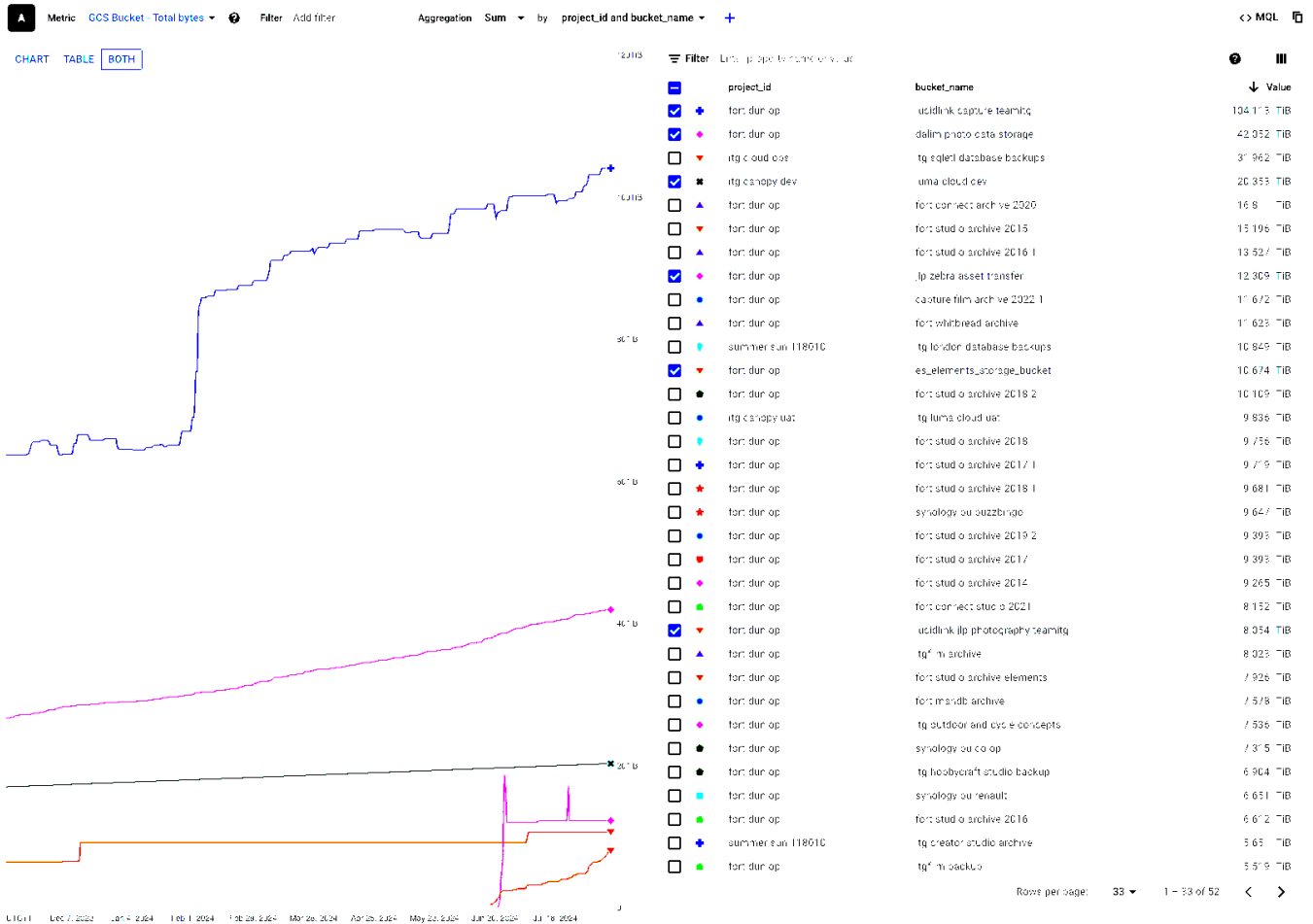
As seen in the report, and to cover 90% of usage, we recommend committing to the following regions:

- Save **\$17,928** per year, by purchasing an **additional** hourly on-demand **1-year** commitment of **\$8.30** per hour in **europa-west1**.
- Save **\$2,656.80** per year, by purchasing an **additional** hourly on-demand **1-year** commitment of **\$1.23** per hour in **europa-west4**.

Cloud Storage

Storage Cost Trends

Comparing bucket storage between periods of our last exercise in December 2023 and this one, your bucket storage costs increased by **80%**.



Total bytes in monitoring-project - [metrics explorer view](#)

Utilizing the metric scope in **test-monitoring** project, we see the following buckets are trending up in storage bytes:

- capture-test
- asset-transfer
- photo-data-storage
- storage-bucket
- cloud-dev
- lake-photography

Using Autoclass

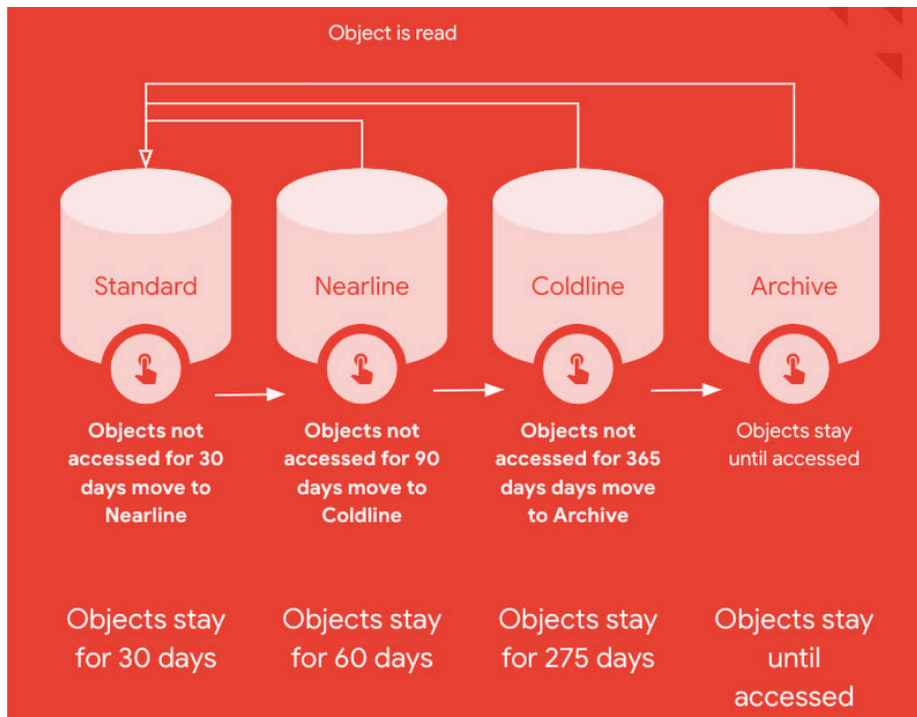
Effort ● ● ● Savings ● ● ●

[Autoclass](#) is an easy to use bucket-level setting that *simplifies and automates lifecycle management of all your Cloud Storage data based on last access time.*

This is very useful for workloads with unpredictable and unknown access patterns. It automatically moves data that is not accessed to colder storage classes to reduce at-rest storage cost. When cold data is accessed, it is automatically promoted to Standard storage class to optimize the operations costs for future accesses.

Autoclass automatically delivers cost savings and removes inefficiencies by moving data to the storage class with the most favorable pricing for customer workloads.

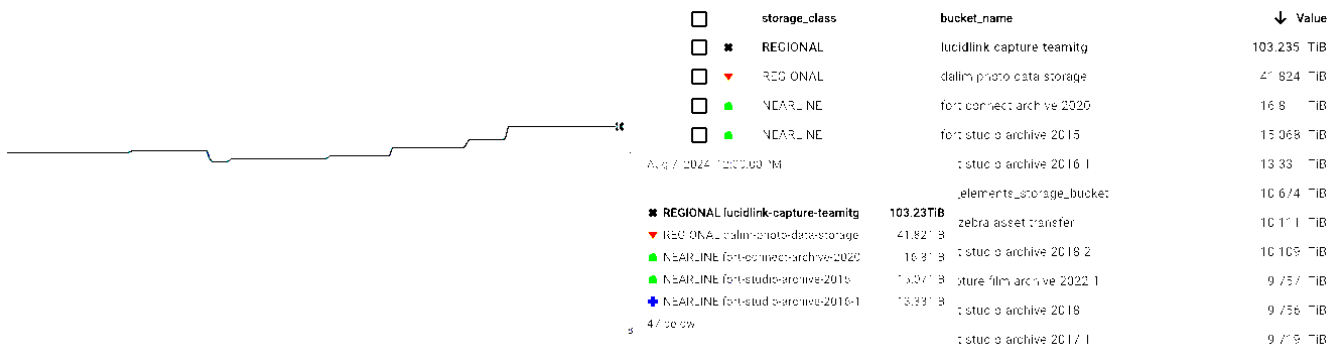
It *helps organizations achieve price predictability* by removing surcharges associated with colder storage classes - there are no early deletion fees, retrieval fees or operations charges for class transitions when Autoclass is enabled on a bucket.



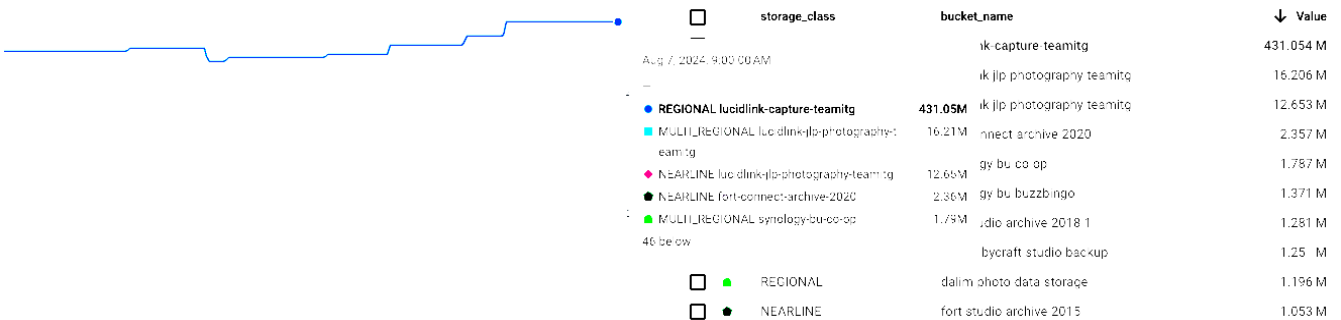


Autoclass can be enabled when you create a new Cloud Storage bucket by using CLI, API, or Cloud Console. Once enabled on a bucket, Autoclass will manage the lifecycle of all objects that are uploaded to the bucket until disabled.

If you choose to disable Autoclass, all the objects in the bucket will remain in their current storage class. This disablement doesn't require any data movement or any changes to storage classes of objects and hence, incurs no charge.



total bytes in example-project - [metrics explorer view](#)



object count in testing-project - [metrics explorer view](#)

- You can now enable Autoclass for [existing](#) buckets, which means you have an option to use this new automated feature or use [Object Lifecycle Management](#) instead.
- Changing the storage class of objects will fall under Class A operations and will increase the cloud spent during the transition. Consider deleting the objects if they are no longer required instead of moving the objects to a different storage class.



Here's also a list of current **project | bucket** with Autoclass enabled:

→ example-project| test-creative

```
terminalStorageClass=NEARLINE;  
terminalStorageClassUpdateTime=2024-05-16T14:36:52.178000+00:00;
```

→ example-project| lake-photography

```
terminalStorageClass=NEARLINE;  
terminalStorageClassUpdateTime=2024-05-17T10:22:48.869000+00:00;
```

→ example-project | asset-transfer

```
terminalStorageClass=NEARLINE;  
terminalStorageClassUpdateTime=2024-06-19T08:31:58.049000+00:00;
```

→ example-project | photography-testing

```
terminalStorageClass=ARCHIVE;  
terminalStorageClassUpdateTime=2024-05-10T16:03:35.589000+00:00;
```

→ example-project | alex-test1

```
terminalStorageClass=NEARLINE;  
terminalStorageClassUpdateTime=2024-05-16T14:46:46.931000+00:00;
```

→ example-project-group | file-storage-poc

```
terminalStorageClass=NEARLINE;  
terminalStorageClassUpdateTime=2024-04-24T11:11:29.444000+00:00;
```

We believe re-evaluating Autoclass would drastically decrease your Cloud Storage bills.



Lifecycle Rules

Outside of Autoclass, Google Cloud also provides object lifecycle rules that automatically move objects to different storage classes based on a set of attributes, such as their creation date or live state.

Every time you shift to a lower storage tier, you decrease the storage costs by **~50%**, as seen below.

	North America	South America	Europe	Asia	Indonesia	Australia
Location	Standard Storage (per GB per Month)		Nearline Storage (per GB per Month)	Coldline Storage (per GB per Month)		Archive Storage (per GB per Month)
Warsaw (europe-central2)	\$0.023		\$0.013	\$0.006		\$0.0025
Finland (europe-north1)	\$0.020		\$0.010	\$0.004		\$0.0012
Belgium (europe-west1)	\$0.020		\$0.010	\$0.004		\$0.0012

If you haven't done so already and in order to get the best cost efficiency out of this, you could automate the process by using a Cloud Function that will migrate the idle bucket to a less expensive storage class, and trigger the function by using a payload intended to mock a notification received from a Cloud alerting policy.

[Object Lifecycle Management | Cloud Storage](#)

[Manage object lifecycles | Cloud Storage](#)

Note: The lifecycle rules do not take into account whether the objects have been accessed or not.

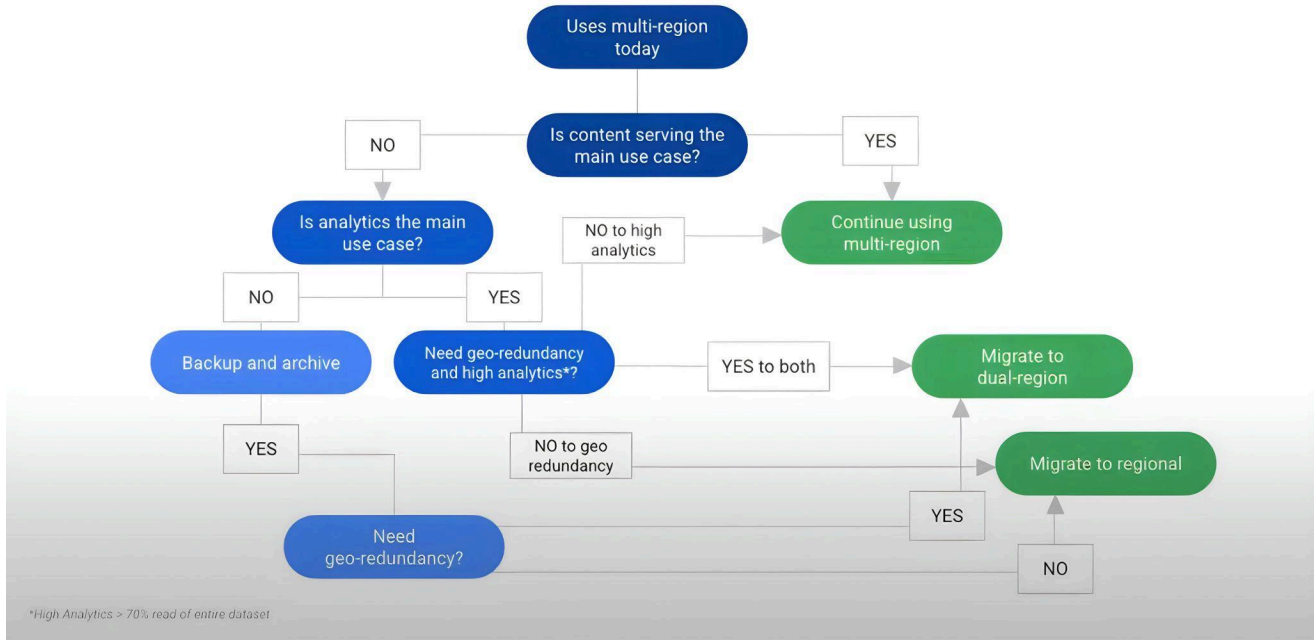
Storage Location

When you choose the location for a Cloud Storage bucket, consider the differences in availability, price, and performance, as shown in the following table.

	Regional	Dual-region	Multi-region
Availability ¹	<ul style="list-style-type: none"> Data redundancy across availability zones (synchronous) RTO=0: automated failover and failback on zonal failure (no need to change storage paths) 	<ul style="list-style-type: none"> Higher availability than regional Data redundancy across regions (asynchronous) Turbo replication option for replication within 15 minutes RTO=0: automated failover and failback on regional failure (no need to change storage paths) 	<ul style="list-style-type: none"> Higher availability than regional Data redundancy across regions (asynchronous) RTO=0: automated failover and failback on regional failure (no need to change storage paths)
Performance	<ul style="list-style-type: none"> 200 Gbps (per region, per project) Scalable to many Tbps by requesting higher bandwidth quota 	<ul style="list-style-type: none"> 200 Gbps (per region, per project) Scalable to many Tbps by requesting higher bandwidth quota 	<ul style="list-style-type: none"> 50 Gbps (per region, per project) Limited performance scaling, variable performance for reads
Pricing	<ul style="list-style-type: none"> Lowest storage price No replication charges No egress charges when reading data inside the same region 	<ul style="list-style-type: none"> Highest storage price Replication charges apply on write No egress charges when reading data within either region 	<ul style="list-style-type: none"> Higher storage price than regional, but lower than dual-region Replication charges apply on write Egress charges always apply when reading data

1. The service level objective (SLO) for each location option depends on the storage class of the bucket. See the [Cloud Storage Service Level Agreement \(SLA\)](#).

Location recommendation flowchart



One thing to keep in mind when considering this option is that [storage in multi-regional locations](#) allow for better performance and higher availability, but comes at a premium and could increase network egress charges, depending on your application’s design. During the application design phase, this is an important factor to consider.

Another option when you’re thinking about performance is buckets in regional locations, a good choice if your region is relatively close to your end users. You can select a specific region that your data will reside in, and get guaranteed redundancy within that region. This location type is typically a safe bet when you have a team working in a particular area and accessing a dataset with relatively high frequency. This is the most commonly used storage location type that we see, as it handles most workloads’ needs quite well. It’s fast to access, redundant within the region, and affordable overall as an object store.

Actions and next steps

Increase usage of Spend-based Committed Use Discounts (CUDs) (P3)

Effort

Savings

Insights: Cloud SQL CUDs give you a 25% discount off on-demand pricing for a one-year commitment and a 52% discount off on-demand pricing for a three-year commitment. Cloud SQL CUDs apply to all Cloud SQL database instance CPU and memory usage in the region in which you have committed, except shared CPU machine types (such as db-f1-micro and db-g1-small). Cloud SQL CUDs do not apply to storage, backups, IP Addresses, network egress, or licensing.

Configure lifecycle policy to cloud storage buckets (P3)

Effort

Savings

Insights: Configure object lifecycle management to delete the objects that are no longer required and consider using Autoclass for new buckets, which will automatically transition the objects to different storage classes based on the access pattern.

Remove Unused Compute Engine Resources (P4)

Effort

Savings

Insights: Backup and delete the unused Compute Engine resources to reduce the cost and Only pay for compute you need.

Leverage Spot VMs (P4)

Effort

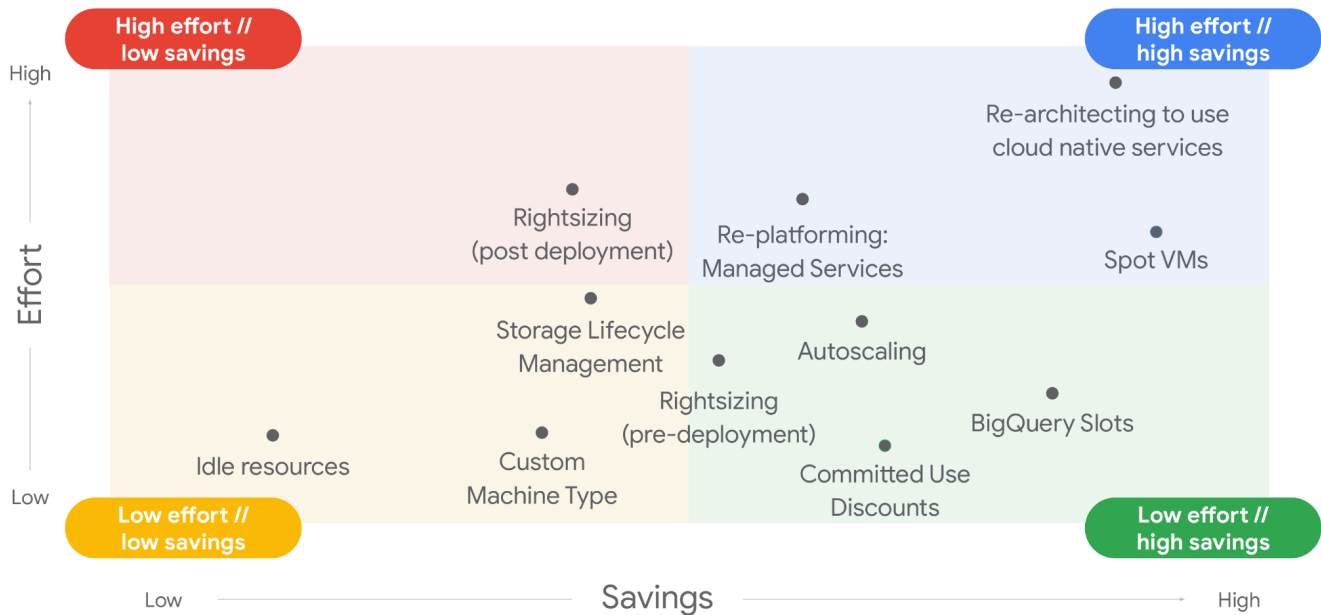
Savings

Up to 80% Savings Capability vs Standard VM price. Based on the use cases, Spot VMs are better fit for batch jobs and fault-tolerant workloads.

Cost Optimization Recommendations

Compute Optimizations	Storage Optimizations	BigQuery Optimizations	Others Optimizations
<p>Resource Optimizations</p> <ul style="list-style-type: none"> Rightsize VM Recs Idle VM Recs Idle Persistent Disk VM Scheduler Newer generation instances Custom machine type <p>Pricing Efficiency</p> <ul style="list-style-type: none"> Compute CUD Recs (save upto 57%) Compute SUD Discounts (Save upto 30%) Spot VMs (save upto 91%) 	<p>Resource Optimizations</p> <ul style="list-style-type: none"> Object Lifecycle Management Object versioning Snapshot retention and cleanup <p>Pricing Efficiency</p> <ul style="list-style-type: none"> Storage Classes Autoclass 	<p>Resource Optimizations</p> <ul style="list-style-type: none"> BigQuery Partitioning & Clustering Data retention and clean up for active storage BigQuery Caching <p>Pricing Efficiency</p> <ul style="list-style-type: none"> Flex Slots BigQuery Slots Recs 	<ul style="list-style-type: none"> Autoscaling Cloud SQL Insights Unused IPs Optimize licensing cost <ul style="list-style-type: none"> BYOL (Save upto 55%) Extended Memory Network Service Tiers Network Intelligence Center BigTable Autoscaler Dataproc optimization Redis MemoryStore optimization Spanner Query Optimizer Dataflow FlexRS (save upto 40%)

Cost Optimization matrix



Tools and resources

- [A tool](#) for converting SKU to unit and cost by id can be useful in exploring costs
- DoIT [Cloud Management Portal - Analytics](#)
- DoIT [Cloud Analytics Documentation](#) - Analyze your cloud consumption and costs
- Sharing committed use discounts across projects [\[1\]](#) [\[2\]](#)