

GCP Certification Series: 2.3 Planning and configuring data storage options



Prashanta Paudel

Oct 18, 2018 · 15 min read

To begin the topic I think having the background of the storage systems will be required in order to understand things properly. Let us talk about storage in general.

A storage device is any computing hardware that is used to store, retrieve or move files and objects. It can hold data both permanently or temporarily and can be internal or external to the computer, server or similar computing device.

Hard disk drives, SDD, pen drives, zip drive, floppy drives etc are all example of storage devices.

Storage devices are core component of any computing systems as they are required to store the operating system and user data. Storage devices are required to store all the applications used in personal computer or server.

Storage devices are generally found in various capacity to fit for the various purposes. As we know 8 bits form a byte, all the storage devices sizes are usually multiple of 8.

A standard computer can have various storage devices including RAM, cache, Hard disk as well as Optical disks and USB drives.

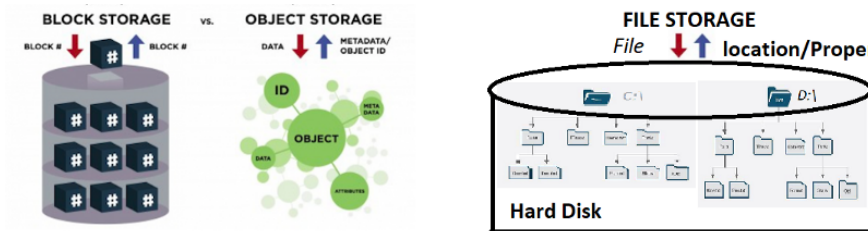
There are two types of storage devices

1. Primary storage devices(RAM): These are in small sizes compared to secondary storage and hold the data temporarily and are internal to the computer. These devices are very fast in data access such as RAM including cache memory.
2. secondary storage devices(HDD): These devices usually have large storage capacity and they store data permanently. These are

usually Hard disks, SSD, Optical disk or USB drive.

Storage Systems

Irrespective of where the storage system is used all the storage devices can be divided into



Type of storage systems

1. File storage———— files and folders
2. Block storage————Databases, OS
3. Object storage————-buckets, multimedia

1. File storage

File storage systems are based on the traditional paper-based file system where files are extracted from folders which are placed in drawers and racks and searched in alphabetical order.

files are hierarchically stored in folders and many subfolders into one root folder. To access the file you have to know the whole path to the last folder and then only you can access the file.

File storage format

The file system can differ from vendor to vendor. For example

- **Windows use FAT,FAT32,NTFS**
- **Mac uses FAT, exFAT, HFS and latest APF**
- **Linux uses ext2, ext3,ext4, XFS, JFS and btrfs.**

FAT- File allocation table

NTFS- New Technology File System

APF- Apple File System

ext2- Extended file system -2

Some file systems are designed for specific applications and work only when used together with the application.

File system and metadata

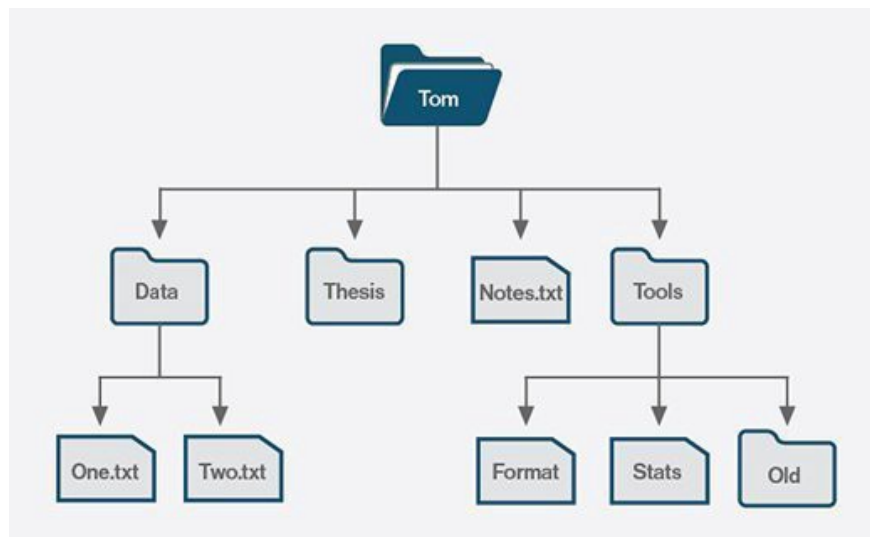
The file system uses metadata to store and pull file information from the storage. Some of the information in metadata could be like

- Creation date
- Modified date
- Owner of the file
- Permissions
- sharing options etc

File storage system

In file storage system you write a file, give it a name, place it inside the folder and nest it under the folder to make it more relevant. In this way, files are organized into a hierarchy, with directories and sub-directories. Each file also has a limited set of metadata associated with it, such as the file name, the date it was created, and the date it was last modified etc.

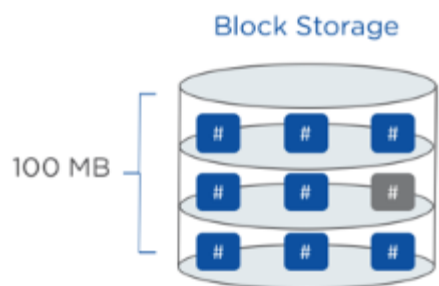
Data is accessed as file IDs (server name + directory path + filename) over a shared network and the storage server manages the data on disk. Network attached storage(NAS) depends on this file system for access and retrieval.



tree file system

This works very well up to a point, but as capacity grows the file model becomes burdensome for two reasons. First, performance suffers beyond a certain capacity. The NAS system itself has the limited processing power, making the processor a bottleneck. The performance also suffers from the massive database—the file lookup tables— that accompany capacity growth.

Block Storage System



In block storage system raw volumes of storage are created and each block can be controlled as an individual hard disk. Data is organized through block IDs (e.g., sector number) and can be organized as a structure (called a file system) or an application-specific structure. The blocks created are controlled by the server-based operating system and each block can be individually formatted with the required file system and are generally accessed by Fibre Channel, iSCSI or Fibre Channel over Ethernet protocols.

Block-Based Disk

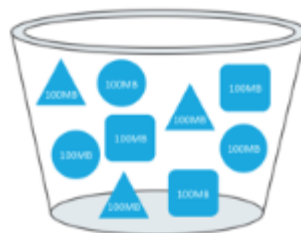


Block level storage is generally deployed in Storage area network (SAN). A client machine connects to a specific volume on the storage service and formats it as if it were a local block device. Block level storage can be used for storing files and can work well for applications like database, Virtual Machine etc. Using traditional file systems on block storage places explicit or practical operational limits on scaling beyond the petabyte range.

Object Storage System

Object storage is a relatively new storage type, designed for unstructured data such as media, documents, logs, backups, application binaries, and VM images. In object storage system the data are stored along with its metadata and unique identifier. This file system doesn't have the hierarchy like the file system or relation to consecutive chunks of hard disk.

Object Storage



Conceptually they are like a persistent key/value store; objects are usually submitted via a REST API call, and an identifier returned. Most object stores allow attaching metadata to objects and aggregating them into containers (or buckets).

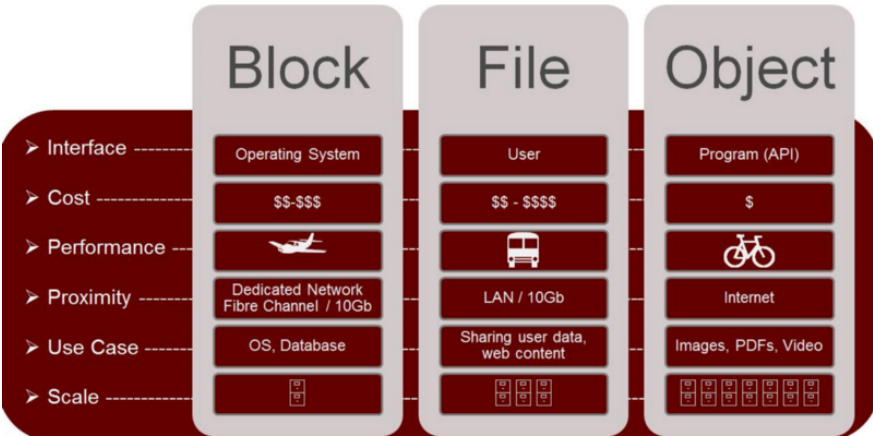
Objects in Object storage are “bundled data” (aka a file) with corresponding metadata. This object gets a unique ID (identifier), that is calculated out of the file content and the metadata.

Applications identify the object via this ID. The many objects inside an object storage system are stored all over the given storage disks. In its pure form, object storage can “only” save one version of a file (object). If a user makes a change another version of the same file is stored as a new object. Because of this reason an object storage is a perfect solution for a backup or archive solution. Or, for example, storage that holds vast amounts of video or movies that are only watched but not changed like for example online movie streaming sites or videos on YouTube.

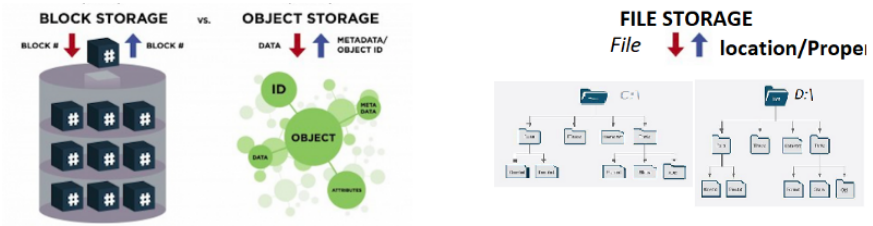
The main difference between the other concepts is that the objects are managed via the application itself that supports Object storage. That means that no real file system is needed here. This layer is obsolete. An application that uses Object storage sends a storage inquiry to the solution where to store the object. The object is then given an address inside the huge storage space and saved there by the application itself.

Because of the much simple management of data—with no real file system in place—Object storage solutions can be scaled up much easier than File storage or Block storage-based systems. You just add some disks in the solution and no big management is needed anymore to have more storage space. That’s the main benefit especially in times of exponential data growth.

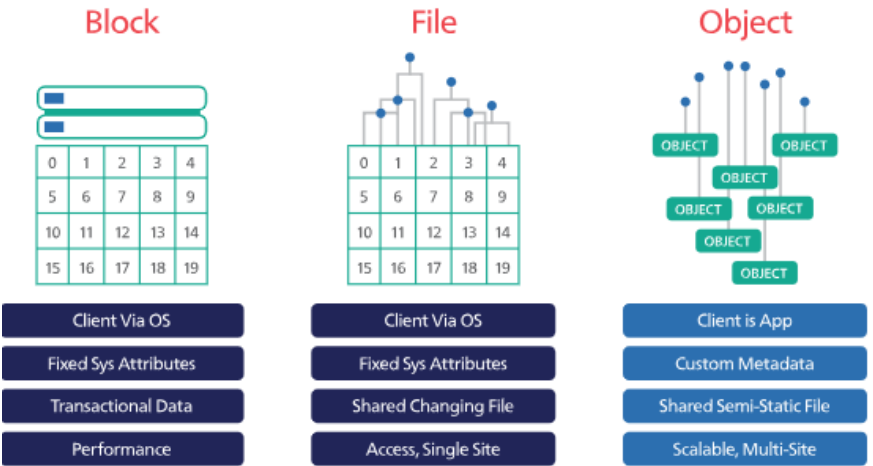
So Object storage is a perfect solution for huge amounts of data and therefore highly used by big cloud service providers like Amazon, Google and others.



Reference: "https://thecustomizewindows.com/wp-content/uploads/2017/09/xObject-Storage-vs-Block-Storage.png.pagespeed.ic.hkN1FTO2ZW.png"



Network Storage Options



reference: https://www.caringo.com/wp-content/uploads/2017/01/File-Object-Storage.svg

NOW LET’S GET BACK TO GOOGLE !!

There are so many data storage options in Google. One thing to note is that the option you choose also depends on the data you are going to operate. For example, if your data are user input in text and needs to save for a short period of time, creating the bucket could be your solution whereas if you are thinking of database then Cloud SQL could be the solution. Each of the Google solutions has its own significant benefits over the other.

Data could be structured or unstructured, anything from text files, mp3, pictures or database etc.

Google provides

1. **Object storage for different needs and price points**
2. **Block storage for VM's**
3. **File storage for applications that need shared filesystems**
4. **Managed MySQL**
5. **Globally scalable NoSQL**
6. **Archival storage for long-term storage**

Different Storage products from Google

Unstructured Data

1. **Google Cloud Storage**
2. **Cloud Filestore**
3. **Persistent Disk**
4. **Cloud Memorystore**

Structured Data

SQL

1. **Cloud SQL**
2. **Cloud Spanner**

NoSQL

1. **Cloud Bigtable**
2. **Cloud Firestore**
3. **Cloud Datastore**

Let us briefly look at what Google has to say about the storage products so you will have a brief idea about what we are going to learn.

PRODUCT	DESCRIPTION	GOOD FOR	COMMON WORKLOADS
 Persistent Disk	Fully-managed, price-performant block storage that is suitable for virtual machines and containers.	<ul style="list-style-type: none"> Block storage for Google Compute Engine and Google Kubernetes Engine Snapshots for data backup 	<ul style="list-style-type: none"> Disks for virtual machines Sharing read-only data across multiple virtual machines Rapid, durable backups of running virtual machines
 Google Cloud Storage	A scalable, fully-managed, highly reliable, and cost-efficient object / blob store.	<ul style="list-style-type: none"> Images, pictures, and videos Objects and blobs Unstructured data 	<ul style="list-style-type: none"> Storing and streaming multimedia Storage for custom data analytics pipelines Archive, backup, and disaster recovery
 Google Cloud Bigtable	A scalable, fully-managed NoSQL wide-column database that is suitable for both real-time access and analytics workloads.	<ul style="list-style-type: none"> Low-latency read/write access High-throughput analytics Native time series support 	<ul style="list-style-type: none"> IoT, finance, adtech Personalization, recommendations Monitoring Geospatial datasets Graphs
 Google Cloud Datastore	A scalable, fully-managed NoSQL document database for your web and mobile applications.	<ul style="list-style-type: none"> Semi-structured application data Hierarchical data Durable key-value data 	<ul style="list-style-type: none"> User profiles Product catalogs Game state
 Google Cloud SQL	A fully-managed MySQL and PostgreSQL database service that is built on the strength and reliability of Google's infrastructure.	<ul style="list-style-type: none"> Web frameworks Structured data OLTP workloads 	<ul style="list-style-type: none"> Websites, blogs, and content management systems (CMS) Business Intelligence (BI) applications ERP, CRM, and eCommerce applications Geospatial applications
 Google Cloud Spanner	Mission-critical, relational database service with transactional consistency, global scale and high availability.	<ul style="list-style-type: none"> Mission-critical applications High transactions Scale + Consistency requirements 	<ul style="list-style-type: none"> Adtech Financial services Global supply chain Retail
 Google BigQuery	A scalable, fully-managed Enterprise Data Warehouse (EDW) with SQL and fast response times.	<ul style="list-style-type: none"> OLAP workloads up to petabyte-scale Big Data exploration and processing Reporting via Business Intelligence (BI) tools 	<ul style="list-style-type: none"> Analytical reporting on large data Data Science and advanced analyses Big Data processing using SQL

References: <https://cloud.google.com/storage-options/>

We will discuss all of these options and use cases one by one

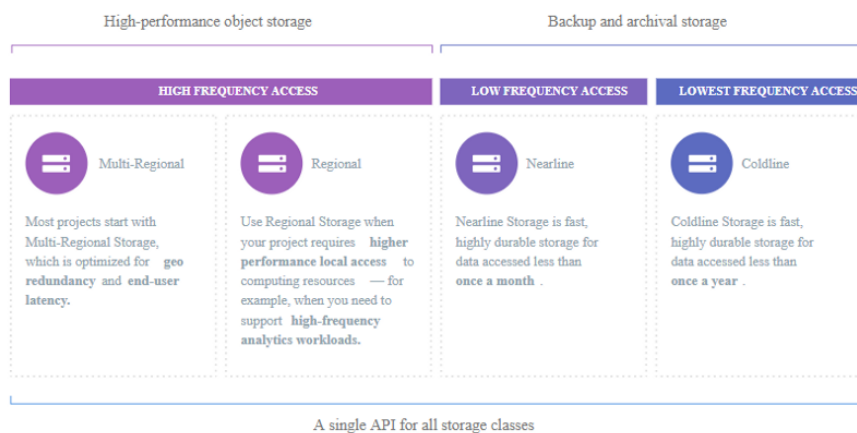
1. Google Cloud Storage



Google Cloud Storage

Google cloud storage is the most widely used storage options in GCP. It is Object-based storage and highly accessible in different ways. It Integrates into your app with API. The price depends on the storage class.

Data stored is highly secured and durable with zero carbon emission. The storage is divided into 4 classes and two storage types.



reference: <https://cloud.google.com/images/storage/storage-classes-desktop.svg>

Storage Class Comparison

Storage Class	Multi-Regional	Regional	Nearline	Coldline
Price	\$0.026 - \$0.036 per GB/month	\$0.02 - \$0.035 per GB/month	\$0.01 - \$0.02 per GB/month	\$0.007 - \$0.014 per GB/month
Good for	For highest availability of frequently accessed data	For data accessed frequently within a region	For data accessed less than once a month	Data accessed less than once a year
Redundancy	Geo-redundant	Regional, redundant across availability zones	Regional	Regional
Availability	99.95% SLA	99.9% SLA	99% SLA	99% SLA
Durability	99.99999999%			
Responsiveness	Millisecond access consistent API			
Use cases	Content storage and delivery, business continuity	Store data and run data analytics within a region	Store infrequently accessed content	Archive storage, backup and recovery
Applications	Video, Multimedia, Business continuity	Transcoding, Data analytics, Compute intensive data processing	Backup long-tail content, Rarely accessed docs	Archive source file backup, Disaster recovery

References: <https://cloud.google.com/storage/features/>

Key Features



Single API across storage classes



Scalable to exabytes of data



Designed for 99.99999999% durability



Very high availability across all storage classes



Time to first byte in milliseconds



Strongly consistent listing

references: <https://cloud.google.com/products/storage/>

Use cases

1. Hosting music and videos
2. Hosting images and website content
3. Mobile apps development

4. Video transcoding
5. Genomics
6. General data analytics and compute
7. Archival data
8. Multimedia backup storage
9. Disaster recovery

Google says:{

Choosing a region and zone

Certain Compute Engine resources live in regions or zones. A region is a specific geographical location where you can run your resources. Each region has one or more zones; most regions have three or more zones. For example, the us-central1 region denotes a region in the Central United States that has zones us-central1-a, us-central1-b, us-central1-c, and us-central1-f.

Resources that live in a zone, such as instances or persistent disks, are referred to as zonal resources. Other resources, like static external IP addresses, are regional. Regional resources can be used by any resources in that region, regardless of zone, while zonal resources can only be used by other resources in the same zone.

For example, disks and instances are both zonal resources. To attach a disk to an instance, both resources must be in the same zone. Similarly, if you want to assign a static IP address to an instance, the instance must be in the same region as the static IP.

You choose which region or zone hosts your resources, which controls where your data is stored and used. Choosing a region and zone is important for several reasons:


Handling failures Distribute your resources across multiple zones and regions to tolerate outages. Google designs zones to be independent of each other: a zone usually has power, cooling, networking, and control planes that are isolated from other zones, and most single failure events will affect only a single zone. Thus, if a zone becomes unavailable, you can transfer traffic to another zone in the same region to keep your

services running. Similarly, if a region experiences any disturbances, you should have backup services running in a different region. For more information about distributing your resources and designing a robust system, see *Designing Robust Systems*. Decreased network latency decrease network latency, you might want to choose a region or zone that is close to your point of service. For example, if you mostly have customers on the East Coast of the US, then you might want to choose a primary region and zone that is close to that area and a backup region and zone that is also close by.}

Migrate your Data

Online, offline, and partner solutions to move data fast


[VIEW DATA TRANSFER PRODUCTS](#) [VIEW DOCUMENTATION](#)



Move from cloud to cloud

Cloud Storage Transfer Service transfers data from an online data source to Google Cloud Storage. Your data source can be an Amazon Simple Storage Service (Amazon S3) bucket, an HTTP/HTTPS location, or a Google Cloud Storage bucket.

[LEARN MORE](#)



Move from on-prem to cloud

Transfer Appliance is a rackable, high capacity storage server that you set up in your datacenter. Fill it with data and ship it to an ingest location where the data is uploaded to Google Cloud Storage. Choose from 100TB or 480TBs of raw capacity per appliance to move your data quickly.

[LEARN MORE](#)

References: <https://cloud.google.com/storage/migrate/>

Object lifecycle

Object lifecycles refer to what to do with the data once it is old or doesn't require frequently. There are several options to select like storing in nearline or cold line or delete it after custom time.

2. Cloud Filestore—Beta



cloud Filestore

The name itself filestore refers to a managed file storage service for the application that requires a filesystem interface and a shared file system for data. It gives native experience for using Network attached storage(NAS) with Google Compute Engine.

BENEFITS

Fast

Cloud Filestore offers low latency for file operations. For workloads that are latency sensitive, like content management systems, databases, random I/O, or other metadata intensive applications, Filestore provides high IOPS with minimal variability in performance.

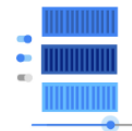


Consistent

With Cloud Filestore, you pay a predictable price for predictable performance. Users independently pick the IOPS and the storage capacity you need with Filestore, which enables you to tune your filesystem for a particular workload. The performance you experience for a particular workload will be consistent over time.

Simple






Cloud Filestore is a fully managed, NoOps service that is integrated with the rest of the Google Cloud portfolio. You can easily mount Filestore file shares on Compute Engine VMs. Filestore is also tightly integrated with Google Kubernetes Engine so your containers can reference the same shared data.



References: <https://cloud.google.com/filestore/>

Create and manage Cloud Filestore instances by using the GCP console or the command-line tool, and interact with the NFS file share on the instance by using standard operating system commands.

USE CASES

 <h4>Rendering</h4> <p>You can easily mount Cloud Filestore file shares on Compute Engine instances, enabling visual effects artists to collaborate on the same file share. As rendering workflows typically run across fleets ("render farms") of compute machines, all of which mount a shared filesystem, Filestore and Compute Engine can scale to meet your job's rendering needs.</p>	 <h4>Application migrations</h4> <p>Many on-premises applications require a filesystem interface to data. As these applications continue to migrate to the cloud, Filestore can support a broad range of enterprise applications that need a shared filesystem.</p>	 <h4>Web content management</h4> <p>Web developers creating websites and blogs that serve file content to their audience will find it easy to integrate Filestore with web software like WordPress.</p>
 <h4>Media processing</h4> <p>Graphic design, video and image editing, and other media workflows use files as an input and files as the output. Filestore helps creators access shared storage to manipulate and produce large files.</p>	 <h4>Home directories</h4> <p>Users across your organization probably need to access and share common data sets. You can host file content in Cloud Filestore and enable shared access to that data.</p>	

References: <https://cloud.google.com/filestore/>

3. Cloud SQL



Cloud SQL

Cloud SQL is a fully managed database service that makes it easy to set up, maintain, manage and administer your relational database. Cloud SQL supports

1. PostgreSQL and
2. MySQL

It offers high performance, scalable and convenience for users to create and implement a database for their applications. A database is

structured data it will be hosted in Google cloud but can be accessed from anywhere.



The infographic is divided into six horizontal sections, each with an icon and a title. 1. 'Focus on Your Application' features a database cylinder icon. 2. 'Simple & Fully Managed' features a gear with a checkmark icon. 3. 'Performance & Scalability' features a bar chart icon. 4. 'Reliability & Security' features an icon of three servers with padlocks. 5. 'Discounts Without Lock-in' features a stopwatch icon. Each section contains a short paragraph describing the benefit.

Focus on Your Application

Let Google manage your database, so you can focus on your applications. Cloud SQL is perfect for Wordpress sites, e-commerce applications, CRM tools, geospatial applications, and any other application that is compatible with MySQL or PostgreSQL.

Simple & Fully Managed

Cloud SQL is easy to use. It doesn't require any software installation. It automates all your backups, replication, patches, and updates - while ensuring greater than 99.95% availability, anywhere in the world.

Performance & Scalability

Cloud SQL delivers high performance and scalability with up to 10TB of storage capacity, 40,000 IOPS, and 416GB of RAM per instance.

Reliability & Security

Easily configure replication and backups to protect your data. Go further by enabling automatic failover to make your database highly available (HA). Your data is automatically encrypted and Cloud SQL is SSAE 16, ISO 27001, PCI DSS v3.0, and HIPAA compliant.

Discounts Without Lock-in

Cloud SQL offers per-second billing, automatic sustained use discounts, and instance sizes to fit any budget. Database instances are easy to stop and start. There is no up-front commitment, and with sustained use discounts, you'll automatically get discounted prices for databases that run continuously.

References: <https://cloud.google.com/sql/>

CLOUD SQL FEATURES

Cloud SQL is a fully-managed MySQL and PostgreSQL database service.

Scalability

Easily scale up to 64 processor cores and more than 400GB of RAM. Quickly scale out with read replicas.

High Performance

Designed to scale from small development workloads up to performance-intensive workloads.

Integrated

Cloud SQL instances are accessible from just about any application, anywhere. Easily connect from [App Engine](#), [Compute Engine](#), and your workstation.

Fully Managed

Replicated, managed and backed-up, so you can make better use of your time.

Security

Cloud SQL data is encrypted when on Google's internal networks and when stored in database tables, temporary files, and backups. Cloud SQL supports private connectivity with [Virtual Private Cloud \(VPC\)](#) and every Cloud SQL instance includes a network firewall, allowing you to control public network access to your database instance. Learn more about Google Cloud Platform's [comprehensive security architecture](#).

Standard APIs

Build and deploy for the cloud faster because Cloud SQL offers standard MySQL and PostgreSQL databases. Use standard connection drivers and built-in migration tools to get started quickly.

Availability Protection

Live migration makes maintenance of our underlying infrastructure [transparent](#). For isolation from failures, [High Availability](#) provides continuous health-checking and automatically fails over if an instance is not healthy.

Partnerships & Integrations

Take advantage of our growing [partner ecosystem](#) and tools to make working with Cloud SQL even easier. Our partners can help you streamline the process of loading your data, create rich visualizations for meaningful insights, and monitor and manage your databases.

References: <https://cloud.google.com/sql/>

Cloud SQL Pricing depends on Regions and Zones of the GCP

Use cases

Cloud SQL can be used for managing

PRODUCT	DESCRIPTION	GOOD FOR	COMMON WORKLOADS
 Google Cloud SQL	A fully-managed MySQL and PostgreSQL database service that is built on the strength and reliability of Google's infrastructure.	<ul style="list-style-type: none"> Web frameworks Structured data OLTP workloads 	<ul style="list-style-type: none"> Websites, blogs, and content management systems (CMS) Business Intelligence (BI) applications ERP, CRM, and eCommerce applications Geospatial applications

4. Cloud Bigtable



Cloud Bigtable


Cloud Bigtable is a fully managed NoSQL database service that can scale up to petabyte for analytical and operational workloads. It has

very low latency as small as 10ms. It has replication option in case of zonal failures.

The most widely used area for Bigtable is Machine Learning and AI.

It is easily integrable with current Big data tools like Hadoop and HBase.

CLOUD BIGTABLE BENEFITS

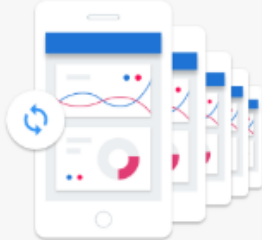




Fast and performant

Use Cloud Bigtable as the storage engine for large-scale, low-latency applications as well as throughput-intensive data processing and analytics.

Seamless scaling and replication

Provision and scale to hundreds of petabytes, and smoothly handle millions of operations per second. Changes to the deployment configuration are immediate, so there's no downtime during reconfiguration. Replication adds high availability for live serving apps, and workload isolation for serving vs. analytics.



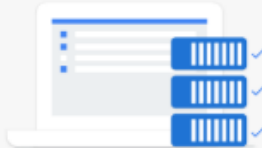


Simple and integrated

Cloud Bigtable integrates easily with popular big data tools like [Hadoop](#), [Cloud Dataflow](#), and [Cloud Dataproc](#). Plus, Cloud Bigtable supports the open source industry standard [HBase API](#), which makes it easy for your development teams to get started.

Fully managed

Because we manage the database and handle the configuring and tuning, you can focus on developing applications.



Reference: <https://cloud.google.com/bigtable/>

Use Cases

Cloud Bigtables are used in Ad tech, Fintech, and IoT

5.Cloud Spanner



Cloud Spanner

Cloud Spanner is a fully managed, mission-critical relational database system. It has special properties like

1. High Availability
2. Transactional consistency
3. global scale

With traditional relational semantics like schemas, ACID transaction, SQL with automatic and synchronous replication cloud spanner is the only product of its kind in the market.

Cloud spanner is first horizontally scalable. strongly consistent, relational database service.

Cloud Spanner is the only enterprise-grade, globally-distributed, and strongly consistent database service built for the cloud specifically to combine the benefits of a relational database structure with non-relational horizontal scale. This combination delivers high-performance transactions and strong consistency across rows, regions, and continents with an industry-leading 99.999% availability SLA, no planned downtime, and enterprise-grade security. Cloud Spanner revolutionizes database administration and management and makes application development more efficient.

In today's always-on, globally-distributed world, IT and developer efficiency, measured in the app downtime and time to market, is one of an organization's most precious resources. The challenge of efficiently managing app database backends while at the same time giving developers the tools they need to build efficiently was previously a challenge.

Cloud Spanner: The best of the relational and non-relational worlds

	CLOUD SPANNER	TRADITIONAL RELATIONAL	TRADITIONAL NON-RELATIONAL
Schema	✓ Yes	✓ Yes	✗ No
SQL	✓ Yes	✓ Yes	✗ No
Consistency	✓ Strong	✓ Strong	✗ Eventual
Availability	✓ High	✗ Failover	✓ High
Scalability	✓ Horizontal	✗ Vertical	✓ Horizontal
Replication	✓ Automatic	⚙️ Configurable	⚙️ Configurable

References: <https://cloud.google.com/spanner/>

Use Cases

Customers across industries can use Cloud Spanner to deliver value to their customers:

USE CASE	BEFORE CLOUD SPANNER	WITH CLOUD SPANNER
Financial trading	Inconsistencies lead to potential monetary loss during reconciliation. Global synchronous replication of trades is not feasible.	Cost savings and a consistent, unified, global view.
Insurance	Inconsistencies lead to incomplete views of customers.	Up-to-date customer views provide more accurate, real-time data.
Global call centers	Eventual and out-of-touch.	Real-time and up-to-date.
Supply-chain management and manufacturing	Global supply chain presents an inconsistent global view and/or data must be shipped in batches.	Global, real-time, consistent view enables real-time decision making.
Telecom and billing	Processing capacity limited to finite scale-up compute resources.	Scale-out allows improved processing speed.
Logistics and Transportation	Regional reach with many systems glued together.	Global reach with lower latency and a consistent view.
Gaming	Each server or cluster is its own universe.	Consistent, global view delivers a unified experience.
E-Commerce (High Availability)	Limited availability SLA or no SLA guarantees. In practice, potential missed sales.	Guaranteed max of 5 minutes of downtime (including planned downtime) on paper and in practice.

References: <https://cloud.google.com/spanner/>

CLOUD SPANNER FEATURES

The first horizontally scalable, globally consistent, relational database service

Global Scale

Horizontally scalable across rows, regions, and continents, from 1 to hundreds or thousands of nodes.

Fully Managed

Ease of deployment at every scale and every stage. Synchronous replication and maintenance are automatic and built-in.

Relational Semantics

Everything you would expect from a relational database—schemas, ACID transactions, and SQL queries (ANSI 2011).

Multi-Language Support

Client libraries in C#, Go, Java, Node.js, PHP, Python, and Ruby. JDBC driver for connectivity with popular third-party tools.

Transactional Consistency

Purpose-built for external, strong, global transactional consistency.

Enterprise Grade Security

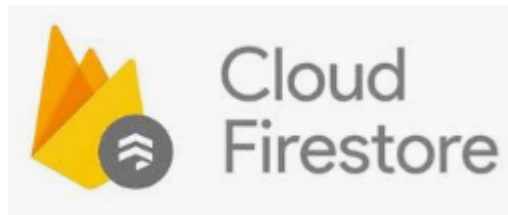
Data-layer encryption, IAM integration for access and controls, and audit logging.

Highly Available

Whenever, wherever, your data is highly available.

References: <https://cloud.google.com/spanner/>

6. Cloud Firestore — Beta



Cloud Firestore



Cloud Firestore is a fast, fully managed, serverless, cloud-native NoSQL document database that simplifies storing, syncing and querying data for mobile, web, and IoT apps at global scale. It's client libraries provide live synchronization and offline support, as well as its security features and integrations with firebase and GCP, accelerates building truly serverless apps

Build serverless apps

Cloud Firestore in Datastore mode is fully compatible with your Cloud Datastore applications. It delivers massive improvements for Datastore users—up to 99.999% availability SLA, no more eventual consistency, no more entity group limits on writes per second, and no more cross-entity group transaction limits.

You can talk directly to Cloud Firestore from your mobile or web clients for a truly serverless solution. No need to set up an intermediary server to manage access to your data. This is great for prototyping, iterating, and getting a production system up and running quickly. Cloud Firestore is a cloud-native database, which provides an automatically scaling solution built from the ground up to take advantage of Google Cloud Platform's powerful infrastructure.

Sync data across devices, on or offline

With Cloud Firestore, your applications can be updated in near real time when data on the backend changes. This is not only great for

building collaborative multi-user mobile applications but also means you can keep your data in sync with individual users who might want to use your app from multiple devices.

Cloud Firestore has full offline support, so you can access and make changes to your data, and those changes will be synced to the cloud when the client comes back online. Built-in offline support leverages local cache to serve and store data, so your app remains responsive regardless of network latency or internet connectivity.

Simple and effortless

Cloud Firestore's robust client libraries make it easy for you to update and receive new data while worrying less about establishing network connections or unforeseen race conditions. It can scale effortlessly as your app grows. Cloud Firestore allows you to run sophisticated queries against your data. This gives you more flexibility in the way you structure your data and can often mean that you have to do less filtering on the client, which keeps your network calls and data usage more efficient.

Enterprise-grade, scalable NoSQL

Cloud Firestore is a fast and fully managed NoSQL cloud database. It is built to scale and takes advantage of GCP's powerful infrastructure, with automatic horizontal scaling in and out, in response to your application's load. Security access controls for data are built in and enable you to handle data validation via a configuration language.

Cloud Firestore features

Security

Cloud Firestore has built-in security access controls for data and enables simple data validation via a configuration language.

Datastore mode

Cloud Firestore supports the Datastore API. You won't need to make any changes to your existing Datastore apps, and you can expect the same performance characteristics and pricing with the added benefit of strong consistency.

Native mode

Built explicitly for serverless mobile and web applications, Native Mode offers real-time data synchronization, offline support, and Security Rules to enable direct-from-client access.

Automatic upgrade

Your Cloud Datastore databases will be upgraded seamlessly and automatically shortly after the GA release of Cloud Firestore. No code changes required, and no downtime for your app.

ACID transactions

Cloud Firestore has support for transactions, so if any of the operations in the transaction fails (and cannot be retried) the whole transaction will fail.

Multi-region replication

With automatic multi-region replication and strong consistency, your data is safe and available, even when disasters strike.

Powerful query engine

Cloud Firestore allows you to run sophisticated queries against your NoSQL data. This gives you more flexibility in the way you structure your data.

Built for cloud-native applications

Typical workloads include mobile and web applications, collaborative multi-user applications, IoT asset tracking and real-time analytics, retail product catalogs, communications, social user profiles and activity, and gaming leaderboards.

References: <https://cloud.google.com/firestore/>

7. Cloud Datastore



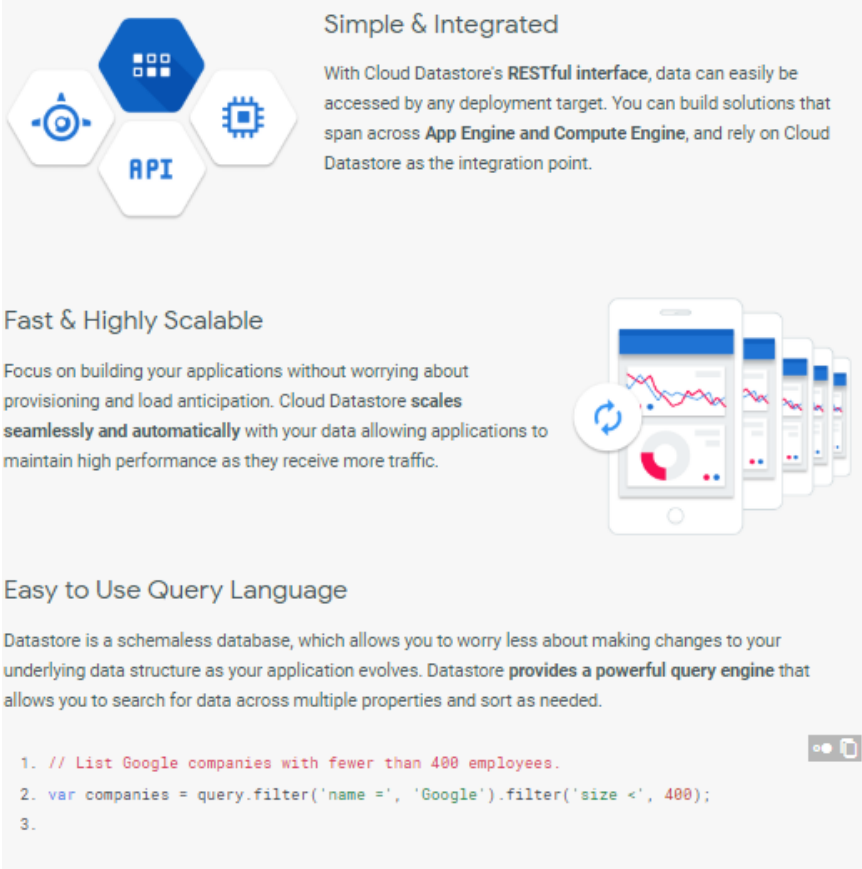
Cloud Datastore

Cloud Datastore is a highly scalable NoSQL database for your web and mobile applications.

Google Cloud Datastore gives you an elastic, highly available document-oriented database as a service. It is fully managed so you don't need to deploy, update, configure, or manage your database solution. Cloud Datastore comes with a rich admin dashboard, a powerful query engine as well as multiple methods to access the database which makes it ideal for mobile and web workloads.

Highly Scalable NoSQL Database

Cloud Datastore is a highly-scalable NoSQL database for your applications. Cloud Datastore **automatically handles sharding and replication**, providing you with a highly available and durable database that scales automatically to handle your applications' load. Cloud Datastore provides a myriad of capabilities such as **ACID transactions, SQL-like queries, indexes and much more.**



Simple & Integrated

With Cloud Datastore's **RESTful interface**, data can easily be accessed by any deployment target. You can build solutions that span across **App Engine and Compute Engine**, and rely on Cloud Datastore as the integration point.

Fast & Highly Scalable

Focus on building your applications without worrying about provisioning and load anticipation. Cloud Datastore **scales seamlessly and automatically** with your data allowing applications to maintain high performance as they receive more traffic.

Easy to Use Query Language

Datastore is a schemaless database, which allows you to worry less about making changes to your underlying data structure as your application evolves. Datastore **provides a powerful query engine** that allows you to search for data across multiple properties and sort as needed.

```
1. // List Google companies with fewer than 400 employees.
2. var companies = query.filter('name =', 'Google').filter('size <', 400);
3.
```

References: <https://cloud.google.com/datastore/>

CLOUD DATASTORE FEATURES

Cloud Datastore is a highly-scalable NoSQL database for your web and mobile applications

<p>Rich Admin Dashboard</p> <p>View entity statistics, query your database, view indexes, and backup/restore your data.</p>	<p>Diverse Data Types</p> <p>Datastore supports a variety of data types, including integers, floating-point numbers, strings, dates, and binary data among others.</p>
<p>Multiple Access Methods</p> <p>Access your data via our JSON API, open-source clients, or community maintained ORMs (Objectify, NDB).</p>	<p>ACID Transactions</p> <p>Ensure the integrity of your data by executing multiple datastore operations in a single transaction with ACID characteristics, so all the grouped operations succeed or all fail.</p>
<p>Fully Managed</p> <p>Cloud Datastore is fully managed, which means Google automatically handles sharding and replication in order to provide you with a highly available and consistent database.</p>	

references: <https://cloud.google.com/datastore/>

8. Persistent Disk




Persistent Disk

Persistent Disk is a high-performance block storage service suitable for Virtual Machines and container storage. It offers unmatched price/performance. You only pay for capacity and you are never charged for provisioned IOPS. Additionally, Persistent Disk offers multi-reader mounts and on-demand volume resizing to simplify operations.


Fast and Flexible Block Storage

Google Persistent Disk is durable and high performance block storage for the Google Cloud Platform. Persistent Disk provides SSD and HDD storage which can be attached to instances running in either Google Compute Engine or Google Kubernetes Engine. Storage volumes can be transparently resized, quickly backed up, and offer the ability to support simultaneous readers.




Industry Leading Price & Performance

Persistent Disk offers industry-leading price performance for both HDD and SSD to satisfy your needs whether you're optimizing for latency sensitive workloads or high throughput workloads. HDD offers low-cost storage with a focus on affordability for large devices for which bulk throughput is of primary importance. SSD offers consistently high performance for both random access workloads and bulk throughput. Both types can be up to 64TB in size.




Share Data Easily

Persistent Disk offers unique multi-reader capability. With multi-reader mounting, many virtual machines can read data from a single Persistent Disk with no performance degradation. Multi-reader mounting makes sharing content a snap.



Better Snapshot Performance

For maximum flexibility and minimal effort, snapshots are geo-replicated and available for restore in all regions by default. Snapshots of a block device can take place in minutes rather than hours.



references: <https://cloud.google.com/persistent-disk/>

PERSISTENT DISK FEATURES

Durable

Persistent Disk is designed for high durability. It stores data redundantly to ensure data integrity.

Independent Volumes

Your storage is located independently from your virtual machine instances, so you can detach or move your disks to keep your data even after you delete your instances.

Volume Size

Each persistent disk can be up to 64 TB in size, so there is no need to manage arrays of disks to create large logical volumes.

Online Resize

Online growth allows volumes to grow on-demand without the need to restart virtual machines or reattach volumes.

References: <https://cloud.google.com/persistent-disk/>

9. Cloud Memorystore



Cloud Memstore

Cloud Memorystore is a fully managed in-memory data store service built on scalable, more secure, and highly available infrastructure managed by Google. Use Cloud Memorystore to build application caches that provide sub-millisecond data access. Cloud Memorystore is compatible with the Redis protocol, allowing easy migration with zero code change.

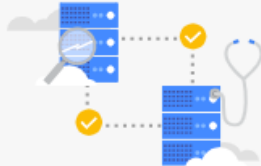


Focus on Building Great Apps

Spend more time writing code and less time managing by taking advantage of the power of open source Redis. Cloud Memorystore automates complex tasks like enabling high availability, failover, patching, and monitoring so that you can spend more time coding.

Scale as Needed

With Cloud Memorystore for Redis, you can easily achieve the sub-millisecond latency and throughput your applications need. Start with the lowest tier and smallest size, and then grow your instance effortlessly with minimal impact to application availability. Cloud Memorystore can support instances up to 300 GB and network throughput of 12 Gbps.

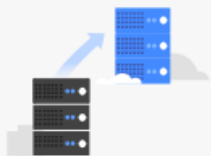


Highly Available

High availability instances are replicated across two zones and provide a 99.9% availability SLA. Instances are monitored constantly, and with automatic failover, applications experience minimal disruption.

Google Grade Security

Cloud Memorystore instances are isolated and protected from the internet using private IPs and are further secured using IAM role-based access control.



Easy Lift and Shift

Cloud Memorystore for Redis is fully compatible with the Redis protocol. You can lift and shift your applications from open source Redis to Cloud Memorystore without any code changes. There is no need to learn new tools since all existing tools and client libraries just work.

References: <https://cloud.google.com/memorystore/>

CLOUD MEMORYSTORE FEATURES

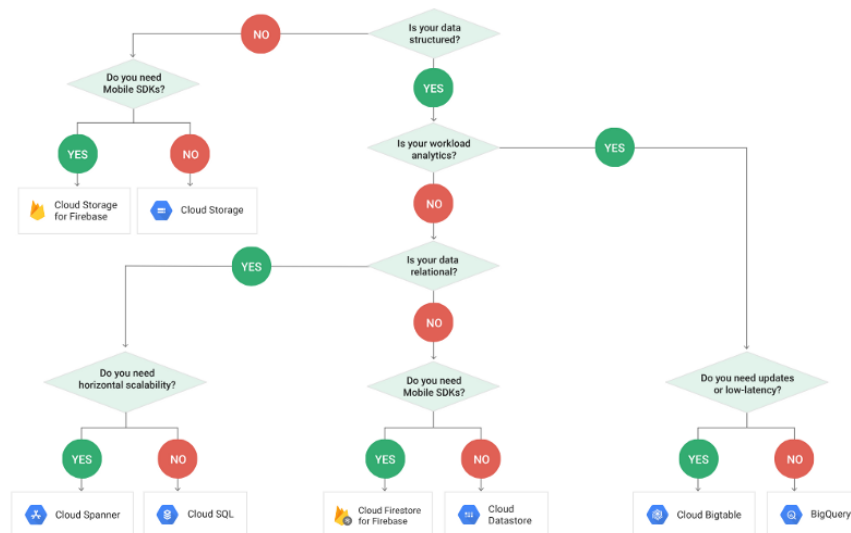
Service Tiers Choose the right tier that fits your cost and availability requirements for Redis instance.	Security Cloud Memorystore for Redis is protected from the internet using VPC networks and private IP and comes with IAM integration—all designed to protect your data. Systems are monitored 24/7/365, ensuring your applications and data are protected.
Fully Managed Provisioning, replication, failover, and patching are all automated which drastically reduces the time you spend doing devops.	Stackdriver Integration Get insights into how Cloud Memorystore for Redis is performing using Stackdriver.
Highly Available Standard Tier Cloud Memorystore for Redis instances provide a 99.9% availability SLA with automatic failover to ensure that your instance is highly available.	Protocol Compatibility Redis protocol compatibility makes it easy to switch your applications with no code changes.

References: <https://cloud.google.com/memorystore/>

FINALLY, CHOOSING A STORAGE OPTION

Different applications or website or apps have different requirements and different platforms in which they work. Performance requirements are one of the important things to consider while choosing the solution. Google has as we have seen from the above a wide range of storage options whether it is structured or unstructured data.

To determine which solution to take Google has built a flow diagram which will lead to one of the storage options.



Reference: <https://cloud.google.com/images/storage-options/flowchart.svg>

In this way, we can choose one solution from a large set of options.

See you in 2.4!

