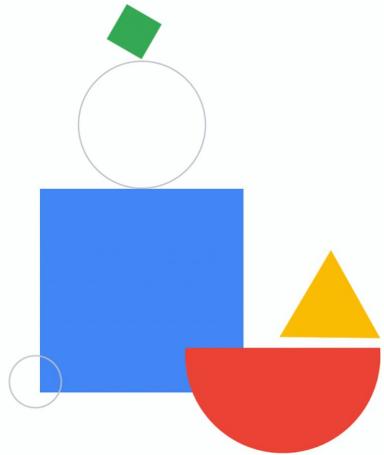


Developing Applications with Google Cloud: Foundations

Module 3: Overview of data storage options



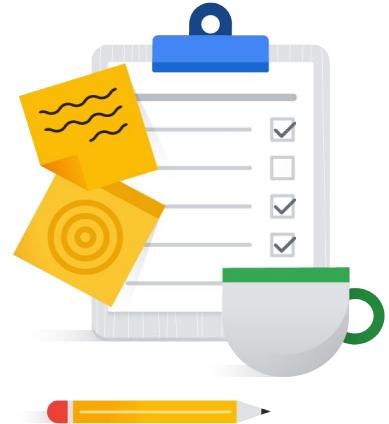
Welcome to Developing Applications with Google Cloud: Foundations, module 3: Overview of data storage options.

Applications can have various types of data. For example, if your application is a social image sharing site, you have to store image files, high volumes of user messages, and transactional data. You generally need to cache frequently accessed data. And you want to collect, query, and analyze all the data to gather business intelligence about your users and product usage patterns.

Agenda

01 Cloud Storage and Databases

02 BigQuery, Memorystore, and Product Comparisons



Google Cloud offers managed services that you can use for each type of data. In this module, you learn about services such as Cloud Storage, Firestore, Bigtable, Cloud SQL, Spanner, and BigQuery. You learn the ideal use cases for each data storage option, as well as use cases for which the option might not be suitable. This knowledge helps you choose the data storage option that meets specific use cases in your applications.

Agenda

01

Cloud Storage and Databases

02

BigQuery, Memorystore, and Product Comparisons



First, we look at the different managed storage services that might be useful in your applications.

Google Cloud provides a full suite of storage service options



Cloud Storage



Firestore



Bigtable



Cloud SQL



AlloyDB



Spanner



BigQuery

- Cost-effective
- Varied choices based on your:
 - Application
 - Workload

You have a full range of cost-effective storage services to choose from when developing with Google Cloud. No one size fits all, and your choice of storage and database solutions will depend on your application and workload.

[Cloud databases: <https://cloud.google.com/products/databases/>
Cloud storage options: <https://cloud.google.com/storage-options/>]

Cloud Storage



Cloud Storage



Firebase



Bigtable



Cloud SQL



AlloyDB



Spanner



BigQuery

Overview

- Fully managed, highly reliable
- Cost-efficient, scalable object store
- Accessed by HTTP request
- Retrieved by object name
- Max object size of 5 TB

Ideal for

- Images and videos
- Objects and blobs
- Unstructured data
- Static website hosting

Cloud Storage is a unified object storage that lets you serve, analyze, and archive data anywhere in the world. Objects are accessed by using HTTP requests, including ranged GETs to retrieve portions of the data. The only key is the object name. There is object metadata but the object itself is treated as unstructured bytes. The scale of the system allows for serving large static content and accepting user uploaded content including videos, photos, and files. Objects can be up to 5 TB each.

Cloud Storage is built for availability, durability, scalability, and consistency. It's an ideal solution for hosting static websites and storing images, videos, objects and blobs, and any unstructured data.

[Cloud Storage documentation: <https://cloud.google.com/storage/docs/>]

Firestore



Cloud
Storage



Firestore



Bigtable



Cloud
SQL



AlloyDB



Spanner



BigQuery

Overview

- Fully managed, serverless, NoSQL
- Scalable
- Mobile and web client libraries
- Real-time updates and offline features

Ideal for

- Mobile and web clients
- Document-oriented data
- Large collections of small documents
- Hierarchical data
- Transactions

Firestore is a fast, fully managed, serverless, NoSQL document database built for automatic scaling, high performance, and ease of application development.

Firestore provides features such as:

- A strongly consistent storage layer,
- A hierarchical collection and document-based data model,
- Real-time updates and offline features, and
- Mobile and Web client libraries

Firestore is built to scale and takes advantage of Google Cloud's powerful infrastructure, with automatic scaling, in response to your application's load.

Firestore is an excellent choice for mobile and web apps, especially those that require flexible data storage or have external user accounts. The data model for Firestore supports flexible, hierarchical data structures. Store your data in documents, organized into collections. Documents can contain complex nested objects in addition to subcollections.

[Firestore documentation: <https://cloud.google.com/firestore/docs/>]

Bigtable



Overview

- High performance wide column NoSQL database service
- Sparsely populated table
- Can scale to billions of rows and thousands of columns
- Can store TB to PB of data

Ideal for

- Operational applications
- Analytical applications
- Storing large amounts of single-keyed data
- MapReduce operations

Bigtable is a high-performance NoSQL database service. It's a sparsely populated table that can scale to billions of rows and thousands of columns. Bigtable can store terabytes to petabytes of data. Bigtable is built for fast key-value lookup and supports consistent sub-10ms latency. Due to the fast lookup and write speed of Bigtable, it's great for storing user behavior.

Bigtable is ideal for operational and analytical applications and is ideal for storing large amounts of single-keyed data and performing MapReduce operations. Bigtable offers seamless scaling: changes to the deployment configuration are immediate, so there's no downtime during reconfiguration.

Bigtable supports the open source industry standard HBase API.

[Bigtable documentation: <https://cloud.google.com/bigtable/docs/>]

Cloud SQL



Cloud Storage



Firestore



Bigtable



Cloud SQL



AlloyDB



Spanner



BigQuery

Overview

- Relational database service
- Managed service (replication, failover, backups)
- MySQL, PostgreSQL, and SQL Server
- Easy and secure access to Cloud SQL instances by using Cloud SQL Auth Proxy

Ideal for

- Web frameworks
- Structured data
- Online transaction processing (OLTP) workloads
- Applications already using MySQL, PostgreSQL, or SQL Server

Cloud SQL is Google Cloud's managed relational database service. With Cloud SQL, Google manages replication, failover, and backups of your databases so you can focus on your MySQL, PostgreSQL, or SQL Server-compatible applications. Cloud SQL lets you easily configure replication and backups to protect your data. You can replicate a primary instance to one or more read replicas. A read replica is a copy of the primary that reflects changes to the primary instance in almost-real time. You can enable automatic failover to make your database highly available.

The Cloud SQL Auth Proxy works by having a local client, called the proxy, running in the local environment. An application communicates with the proxy by using the standard database protocol for your database. The proxy uses a secure tunnel to communicate with its companion process running on the server. Cloud SQL Auth Proxy provides secure access to your Cloud SQL instances without you having to configure allowed IP addresses or SSL certificates.

Cloud SQL is ideal for web frameworks, applications requiring structured data, and online transaction processing (OLTP) workloads. It is ideal for applications using MySQL, PostgreSQL, or SQL Server, with minimal refactoring required for migration to Google Cloud.

[<https://cloud.google.com/sql/docs/>]

[<https://cloud.google.com/sql/docs/mysql/replication/>]

[<https://cloud.google.com/sql/docs/mysql/backup-recovery/backups>]

[<https://cloud.google.com/sql/docs/mysql/sql-proxy>]

AlloyDB



Cloud Storage



Firestore



Bigtable



Cloud SQL



AlloyDB



Spanner



BigQuery

Overview

- High performance, distributed relational database service
- Compute and storage are separate
- Full PostgreSQL compatibility
- High availability with automatic scaling
- Secure access using the AlloyDB Auth Proxy

Ideal for

- Applications that require high performance
- Hybrid transactional/analytical processing (HTAP) workloads
- Fast-response analytic workloads using Columnar Engine
- Applications already using PostgreSQL

AlloyDB is a fully managed, high performance PostgreSQL database service from Google Cloud. It combines the best of Google with PostgreSQL.

Historically, a PostgreSQL database runs on a single VM with an attached disk. Scaling this type of architecture is difficult. With AlloyDB, Google separates the compute and storage for the database. Google databases like Bigtable and Spanner have traditionally used this same approach, allowing them to scale extremely well. AlloyDB gains the same benefits, allowing the database to scale while maintaining high performance for both reads and writes.

AlloyDB is 4 times faster than standard PostgreSQL for transactional workloads. The Columnar Engine of AlloyDB results in 100 times faster performance than standard PostgreSQL for analytical queries, with zero impact on performance when running business intelligence, reporting, and hybrid transactional and analytical processing workloads (HTAP).

AlloyDB has full PostgreSQL compatibility and high availability with automatic scaling. Google manages replication, failover, and backups of your databases so you can focus on your applications.

The AlloyDB Auth Proxy, which is similar to the Cloud SQL Auth Proxy, provides secure access to the database without having to configure allowed IP addresses or SSL certificates.

AlloyDB is ideal for applications requiring high performance and PostgreSQL compatibility. The predictable and high performance is perfect for applications that want to perform a mix of transactional and analytical processing.

[<https://cloud.google.com/alloydb/docs/overview>]

[<https://cloud.google.com/alloydb/docs/cross-region-replication/about-cross-region-replication>]

[<https://cloud.google.com/alloydb/docs/backup/overview>]

[<https://cloud.google.com/alloydb/docs/auth-proxy/overview>]

Spanner



Overview

- Mission-critical relational database service
- Transactional consistency
- Global scale and multi-region replication
- High availability
- 99.999% SLA

Ideal for

- Mission-critical OLTP applications
- High transactions
- Scale and consistency requirements

Spanner is Google Cloud's fully managed relational database service that offers both strong consistency and horizontal scalability. It's designed for mission-critical online transactional processing, or OLTP, applications. Spanner provides automatic, synchronous replication for high availability. Spanner is built for multi-region replication and offers one of the highest SLAs in the industry: 99.999%.

Spanner is ideal for applications with relational, structured, and semi-structured data that require high availability, strong consistency, and transactional reads and writes.

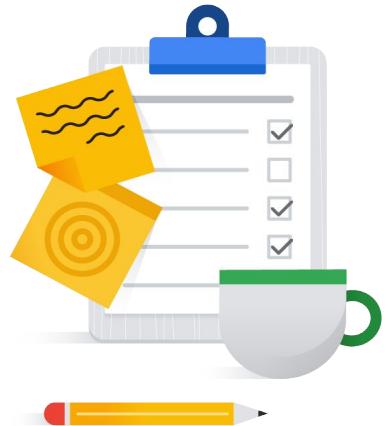
[<https://cloud.google.com/spanner/docs/>]

[<https://cloud.google.com/spanner/docs/schema-design>]

Agenda

01 Cloud Storage and Databases

02 [BigQuery, Memorystore, and Product Comparisons](#)



Next, we look at BigQuery and Memorystore, and then compare the storage services.

BigQuery



Cloud Storage



Firestore



Bigtable



Cloud SQL



AlloyDB



Spanner



BigQuery

Overview

- Fully managed, serverless enterprise data warehouse for analytics
- Built-in machine learning, geospatial analysis, and business intelligence
- Petabyte scale
- Fast response times

Ideal for

- Online Analytical Processing (OLAP) workloads
- Big data exploration and processing
- Reporting with Business Intelligence (BI) tools

BigQuery is a fully managed, serverless enterprise data warehouse for analytics. BigQuery has built-in features like machine learning, geospatial analysis, and business intelligence. BigQuery can scan terabytes in seconds, and petabytes in minutes.

It's a great solution for Online Analytical Processing, or OLAP, workloads, for big data exploration and processing, and for reporting with Business Intelligence tools.

Cache your application data with Memorystore

- Automate complex tasks for Redis and Memcached caching engines
- Fully protocol compatible with each engine
- Ideal for high-performance, scalable web applications, gaming, and stream processing
- Fully managed service
- Google-grade security



Memorystore

Applications that you run on Google Cloud can achieve high levels of performance by using either Redis or Memcached without the burden of managing complex deployments. Memorystore supports both of these highly scalable, available, and secure open source caching engines, and is fully protocol compatible with each engine.

Memorystore is ideal for scalable web applications, gaming, and stream processing, where a distributed in-memory data store allows for fast, real-time access or processing of data.

As a fully managed service, provisioning, replication, failover, and patching are all automated. You can also monitor instances and set up alerts with Cloud Monitoring.

Memorystore can be protected from the internet by using VPC networks and internal IP addresses. Memorystore also integrates with Identity and Access Management (IAM).

Product	Simple Description	Ideal for	Not Ideal for
 Cloud Storage	Binary/object store	Large or rarely accessed unstructured data	Structured data, building fast apps
 Firestore	Real-time NoSQL database to store and sync data	Mobile, web, multi-user, IoT & real-time applications	Analytic data, heavy writes
 Bigtable	High-volume, low-latency database	“Flat,” heavy read/write, or analytical data	Highly structured or transactional data
 Cloud SQL	Well-understood VM-based RDBMS	Web frameworks, existing applications	Scaling, analytics, heavy writes
 AlloyDB	Highly scalable PostgreSQL database	High performance, availability, and scale	Non-PostgreSQL apps
 Spanner	Global relational database service	Low-latency transactional systems	Analytics data
 BigQuery	Autoscaling, managed analytic data warehouse	Interactive analysis of static datasets	Transactional workloads
 Memorystore	Fully managed, scalable caching for applications	High performance data access	Persistent data

Here are the Google Cloud storage options at a glance. To choose the right storage option for your application, it's important to understand what a product is and isn't ideal for by design.

This slide includes a simple description of the products and use cases that are ideal for each product. Use cases that are not ideal for each product are also listed.

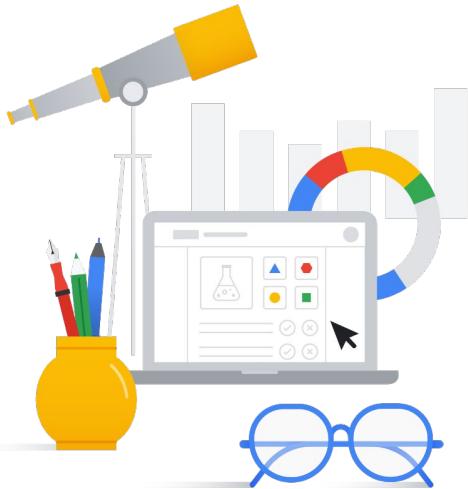
Product	Read/Write latency	Typical size	Storage type
 Cloud Storage	Medium (100s of ms)	Any	Object
 Firestore	Medium (10s of ms)	<200 TB	Document
 Bigtable	Low (ms)	2 TB - 10 PB	Key-Value
 Cloud SQL	Low (ms)	< 64 TB	Relational
 AlloyDB	Low (ms)	< 64 TB	Relational
 Spanner	Low (ms)	Any	Relational
 BigQuery	High (s)	Any	Columnar
 Memorystore	Very Low (sub-ms)	1-TB quota per region	Object

Other considerations for choosing a storage option for your application include read/write latency, typical size of your data, and storage type.

When you're designing the database needs of your applications, remember that you are not limited to a single database. Choose the database that is most suitable for each use case. Size limits are per database, so you can exceed the size limits shown here by splitting your data into multiple databases.

Refer to the table or Google Cloud documentation to identify the best storage option for your application.

[Cloud Storage Options: <https://cloud.google.com/products/storage>]



Hands-on lab:

Storing Application Data in Firestore and Cloud Storage

In this lab, "Storing Application Data in Firestore and Cloud Storage," you create a bookshelf application written in Python. You update the application to use Cloud Client Libraries to store book data in a Firestore database. You also use Cloud Client Libraries to upload book cover images to a Cloud Storage bucket.

In this module, you learned ...



- ✓ Cloud Storage: Managed service for unstructured data, ideal for static websites and large objects.
- ✓ Firestore: Fully managed, scalable, and serverless document database with great developer experience, live synchronization, offline support, and transactions.

Google Cloud has a rich set of services to enable you to store, query, and manage different types of application data.

Cloud Storage is a managed service for storing unstructured data. It's ideal for hosting static websites and storing large objects like videos or photos.

Firestore is a fully managed, scalable, and serverless document database that scales to meet any demand. It offers a great developer experience for mobile and web applications, with support for live synchronization, offline support, and transactions.

In this module, you learned ...



- ✓ **Bigtable:** High-performance NoSQL database for fast access to very large amounts of data with high read and write throughput, ideal for large analytical and operational workloads.
- ✓ **Cloud SQL:** Fully managed service for MySQL, PostgreSQL, and SQL Server, useful for migrating apps to Google Cloud without refactoring data access.
- ✓ **AlloyDB:** Fully managed, PostgreSQL service, useful for apps requiring high performance and for hybrid transactional/analytical processing (HTAP) workloads.

Bigtable is a high-performance NoSQL database service, ideal for fast access to very large amounts of data with high read and write throughput. It's ideal for large analytical and operational workloads.

Cloud SQL is a fully managed database service that helps you set up, maintain, manage, and administer your MySQL, PostgreSQL, and SQL Server databases. It's very useful for migrating existing applications to Google Cloud without having to refactor all of your data access.

AlloyDB is a fully managed, high performance PostgreSQL database service that is useful for applications that require a mix of transactional and analytical processing.

In this module, you learned ...



Spanner: Fully managed relational database, strong consistency, horizontal scalability. Ideal for low-latency transactional systems, especially global scale.

BigQuery: Serverless enterprise data warehouse that scales with your data, great solution for OLAP workloads.

Memorystore: Fully managed, scalable, secure, and highly available in-memory service for caching application data.

Spanner is Google Cloud's fully managed relational database service offering both strong consistency and horizontal scalability. It's ideal for low-latency transactional systems, especially those systems with global scale.

BigQuery is Google Cloud's serverless enterprise data warehouse that scales with your data. It's a great solution for OLAP workloads.

Memorystore is a fully managed, scalable, secure, and highly available in-memory service for caching application data.

With all of these options, you can find managed services to handle all of your application's data needs.