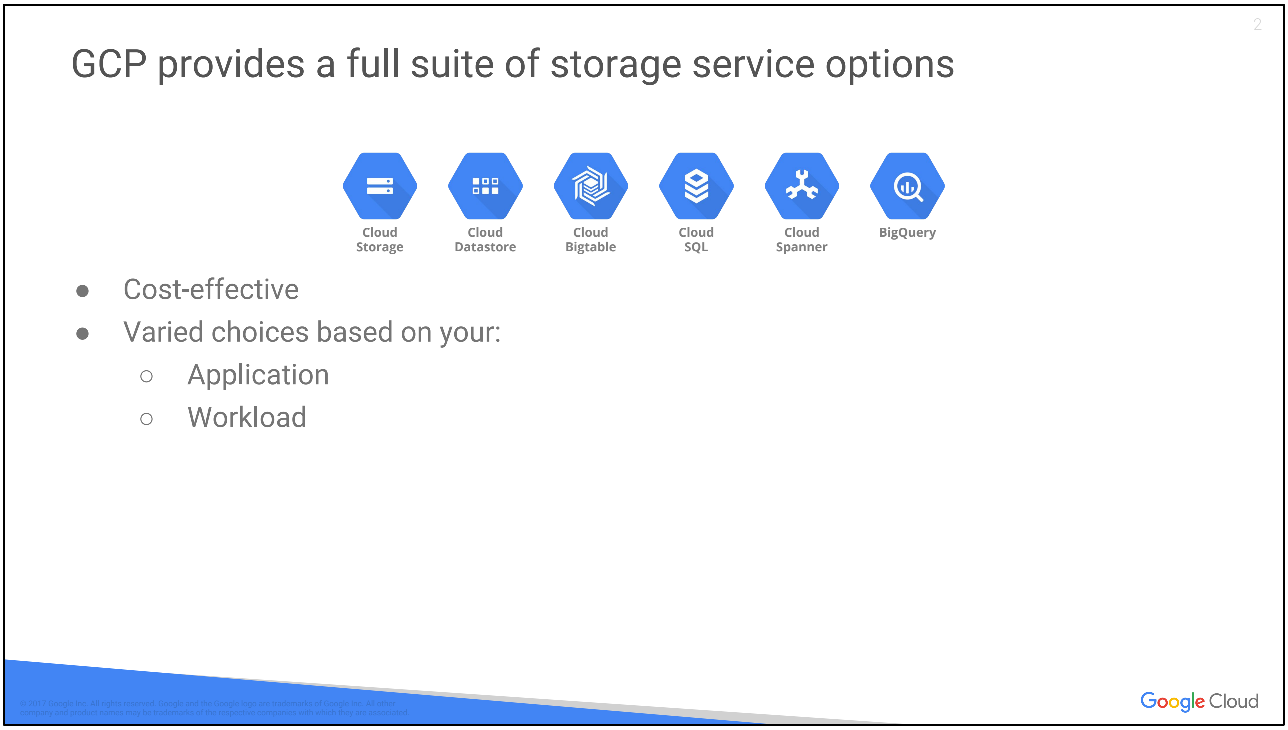
Overview of Data Storage Options

Developing Applications with Google Cloud Platform

CLOUD STORAGE, CLOUD DATASTORE, CLOUD BIGTABLE, CLOUD SQL, CLOUD SPANNER, BIGQUERY, CLOUD SQL PROXY

Google Cloud

Version 2.0
Last modified: 2017-09-12



GCP provides a full suite of storage service options

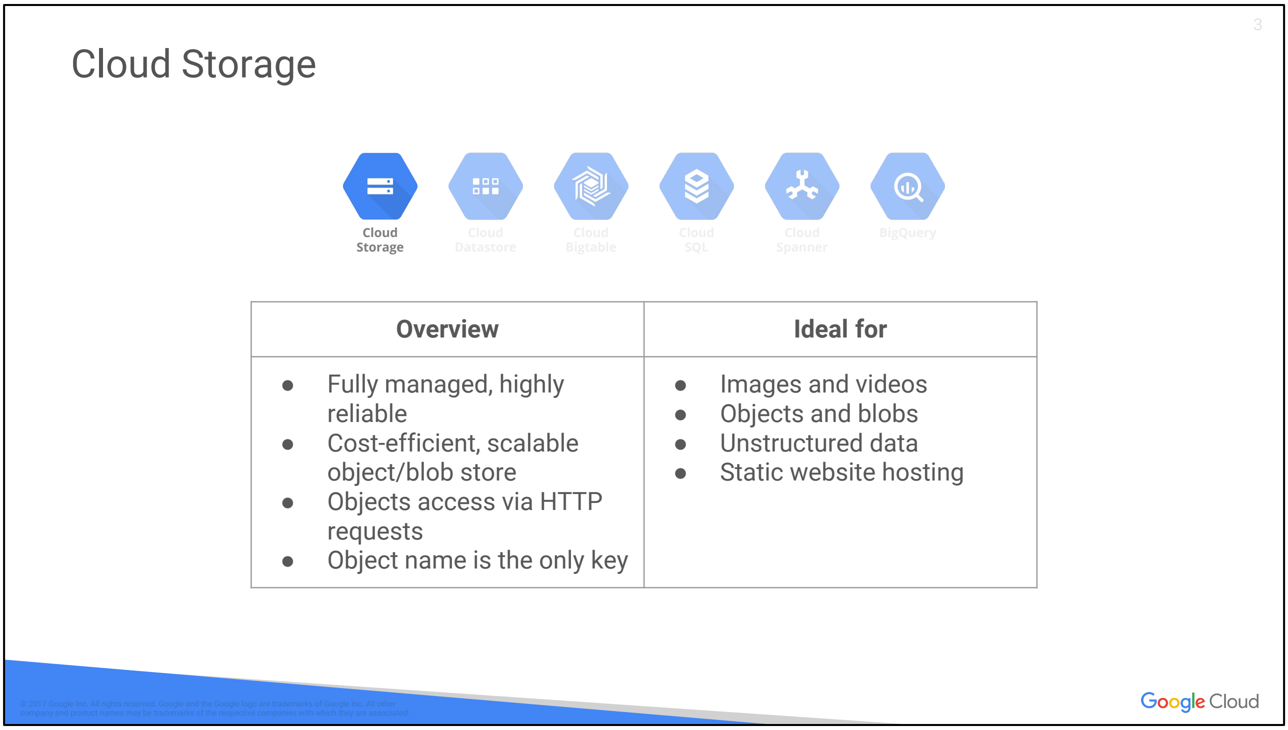
Cloud Storage · Cloud Datastore · Cloud Bigtable · Cloud SQL · Cloud Spanner · BigQuery

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

Google Cloud

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

https://cloud.google.com/storage-options/



Cloud Storage

Cloud Storage · Cloud Datastore · Cloud Bigtable · Cloud SQL · Cloud Spanner · BigQuery

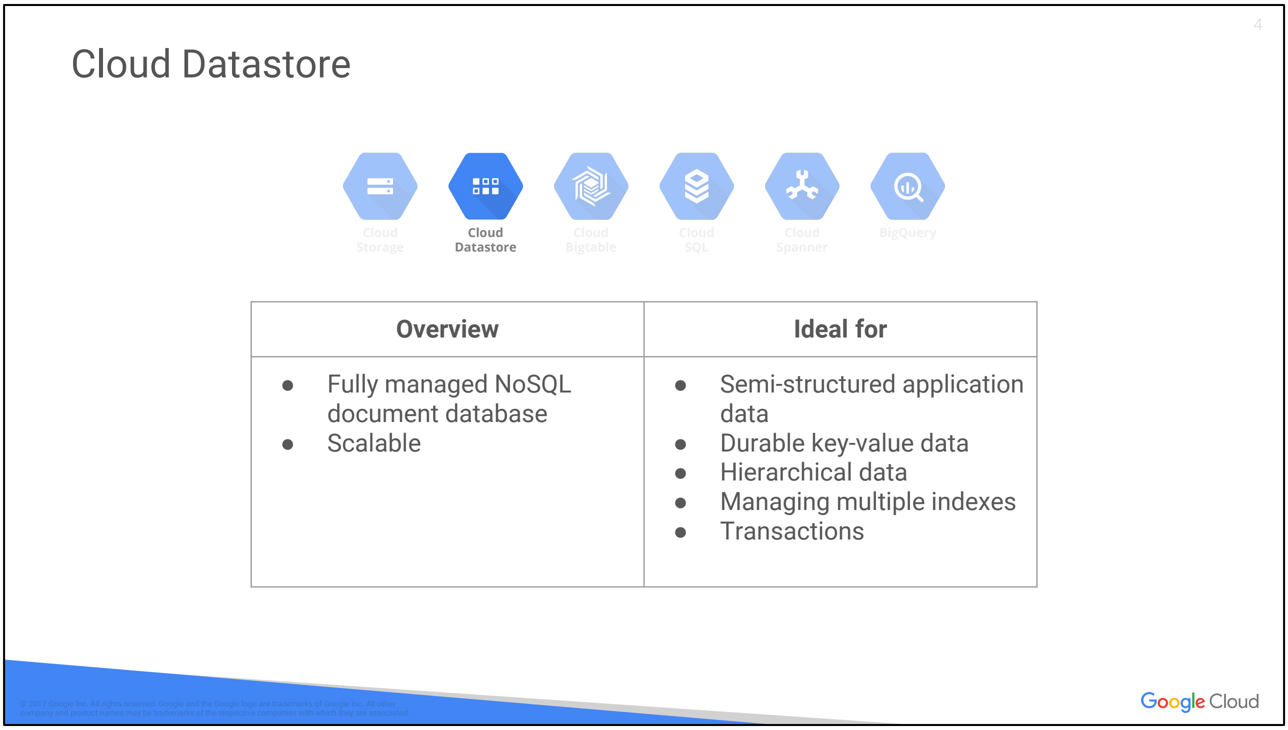
Overview	Ideal for
● Fully managed, highly reliable ● Cost-efficient, scalable object/blob store ● Objects access via HTTP requests ● Object name is the only key	● Images and videos ● Objects and blobs ● Unstructured data ● Static website hosting

Google Cloud

Cloud Storage is a unified object storage for developers and enterprises that allows you to serve, analyze, and archive data. Objects are accessed via 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 just bytes with no structure. The scale of the system allows for serving static content or accepting user uploaded content like photos and videos.

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

For more information, see the Cloud Storage Documentation:
https://cloud.google.com/storage/docs/



Cloud Datastore

Cloud Storage
Cloud Datastore
Cloud Bigtable
Cloud SQL
Cloud Spanner
BigQuery

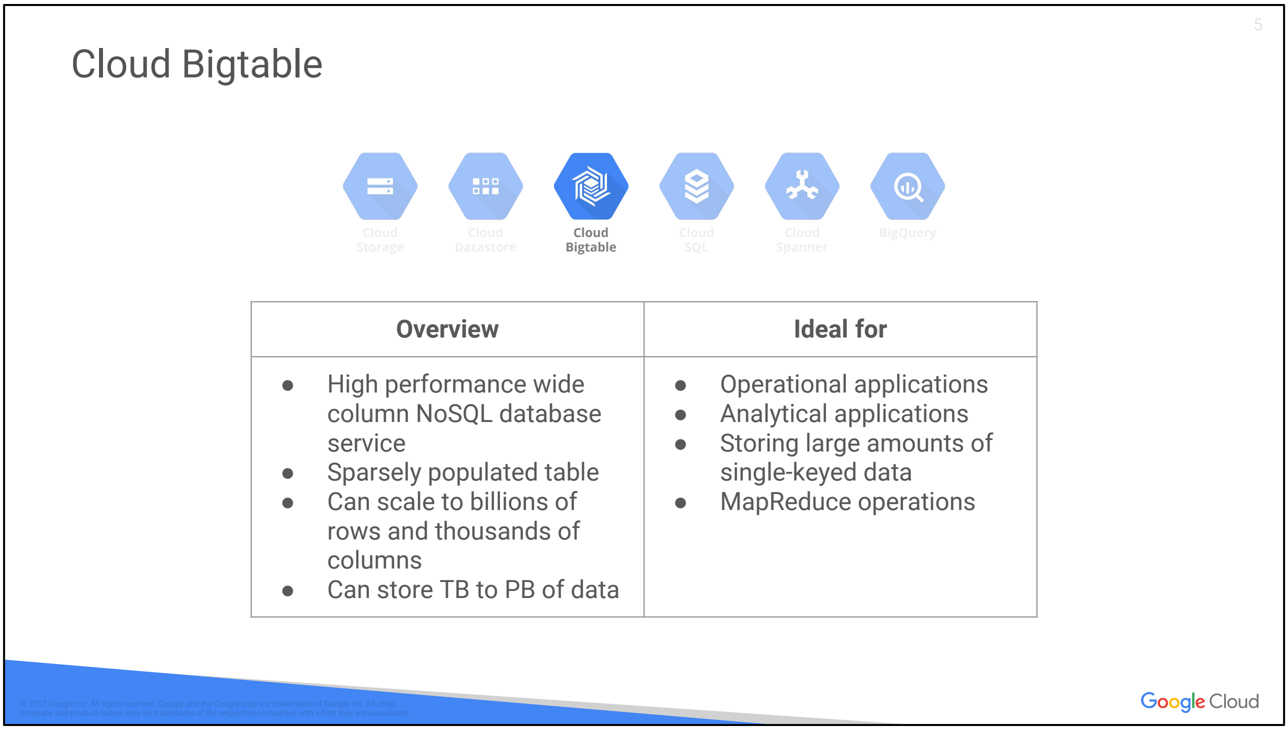
Overview	Ideal for
● Fully managed NoSQL document database ● Scalable	● Semi-structured application data ● Durable key-value data ● Hierarchical data ● Managing multiple indexes ● Transactions

Google Cloud

Datastore is GCP’s fully managed NoSQL document database. It is designed to automatically scale to very large data sets so your applications can maintain high performance as they receive more traffic. Datastore adds many features on top of Bigtable, like managing multiple indexes over each entity, cross row transactions, and regional replication. This often makes it a better choice for high-value application data like user profiles or shopping carts/orders.

Datastore is an ideal solution when your application data is semi-structured or hierarchical and when you need to store durable key-value data.

For more information, see the Datastore Documentation:
https://cloud.google.com/datastore/docs/



Cloud Bigtable

Cloud Storage
Cloud Datastore
Cloud Bigtable
Cloud SQL
Cloud Spanner
BigQuery

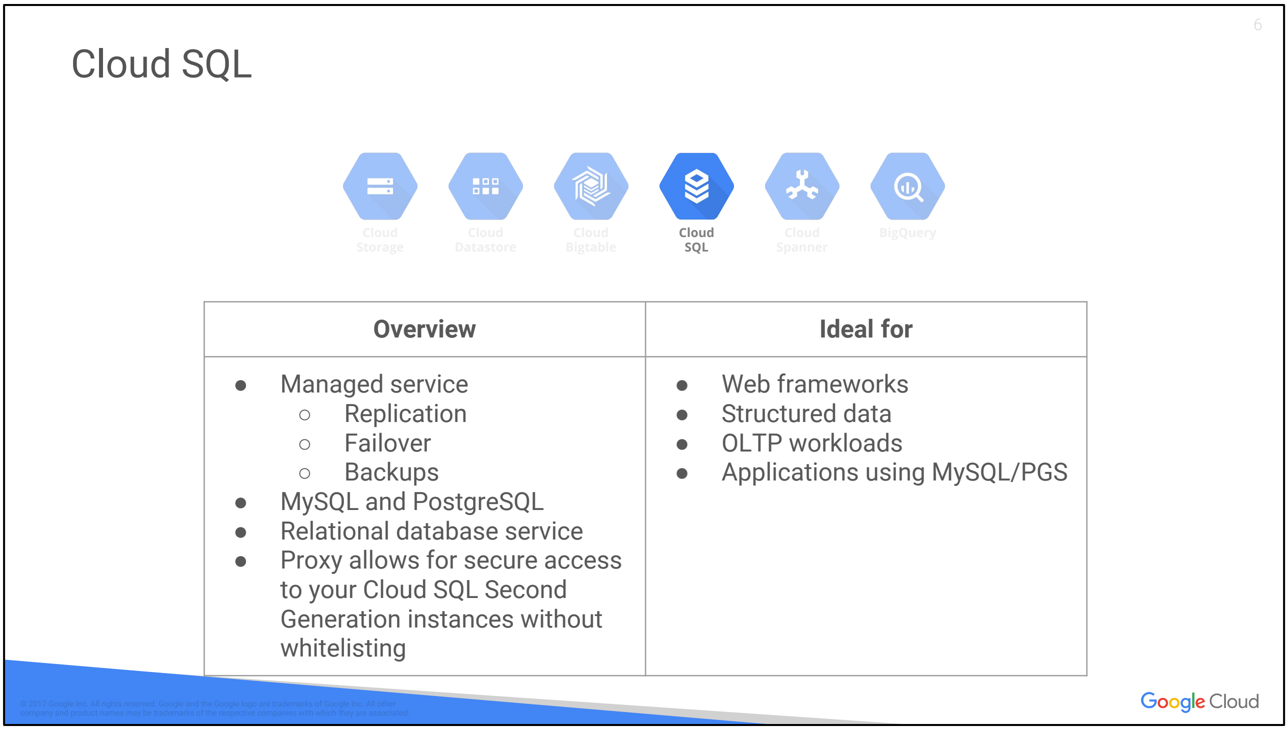
Overview	Ideal for
● High performance wide column NoSQL database service ● Sparsely populated table ● Can scale to billions of rows and thousands of columns ● Can store TB to PB of data	● Operational applications ● Analytical applications ● Storing large amounts of single-keyed data ● MapReduce operations

Google Cloud

Bigtable is a high performance NoSQL database service. It is a sparsely populated table that can scale to billions of rows and thousands of columns. Bigtable can store terabytes to petabytes of data. Similar to Apache HBASE, Bigtable is built for fast key-value lookup and scanning over a defined key range. It is similar to a spreadsheet that gives you access to any set of columns from contiguous rows by searching only the value in the first column (the key). Updates to individual rows are atomic. Due to the fast lookup and write speed of Bigtable, it's great for user behavior.

Bigtable supports operational and analytical applications and is ideal for storing large amounts of single-keyed data and MapReduce operations.

For more information, see the Bigtable Documentation:
https://cloud.google.com/bigtable/docs/



Cloud SQL

Cloud Storage
Cloud Datastore
Cloud Bigtable
Cloud SQL
Cloud Spanner
BigQuery

Overview	Ideal for
● Managed service ○ Replication ○ Failover ○ Backups ● MySQL and PostgreSQL ● Relational database service ● Proxy allows for secure access to your Cloud SQL Second Generation instances without whitelisting	● Web frameworks ● Structured data ● OLTP workloads ● Applications using MySQL/PGS

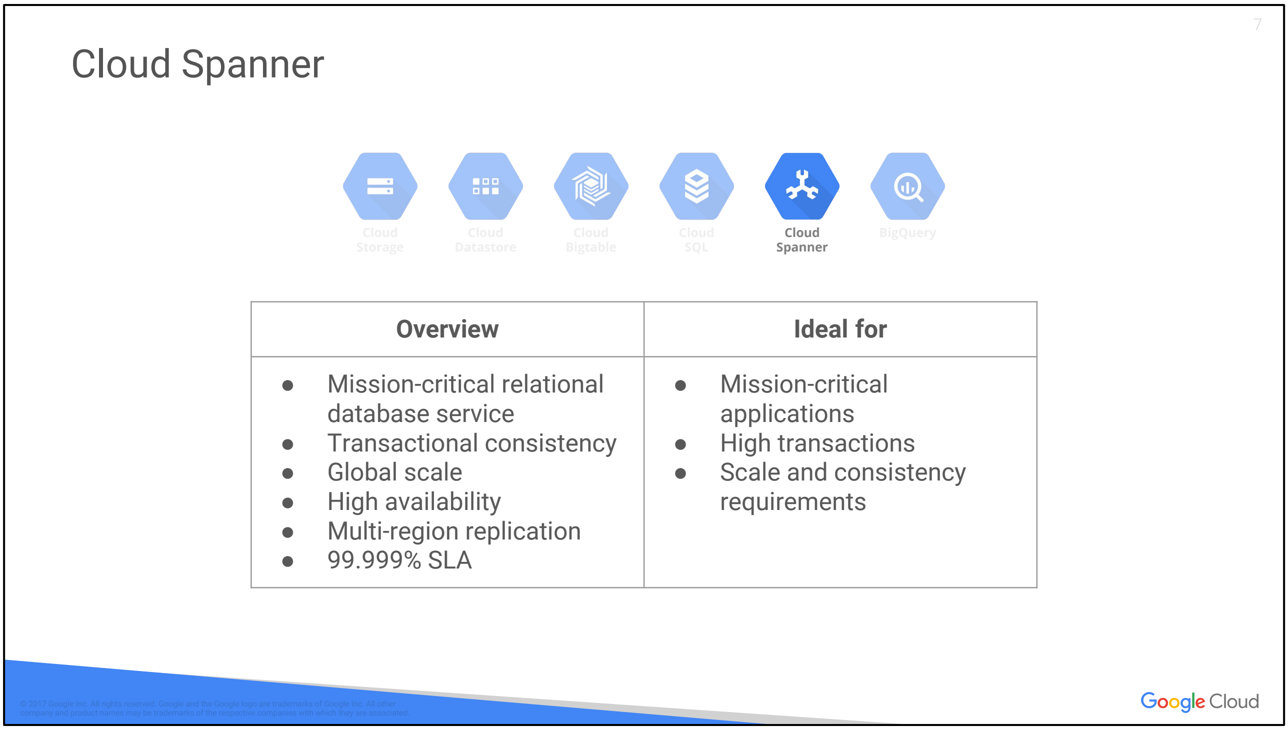
Google Cloud

Cloud SQL is GCP's relational database service. It is a managed service that lets Google manage replication, failover, and backups of your databases so you can focus on your MySQL- or PostgreSQL-compatible applications. Cloud SQL lets you easily configure replication and backups to protect your data. You can replicate a master instance to one or more read replicas. A read replica is a copy of the master that reflects changes to the master instance in almost-real time. You can enable automatic failover to make your database highly available. Backups allow you to restore your Cloud SQL instance to recover lost data or recover from a problem with your instance. You can enable automated backups for any instance that contains data that you need to protect from loss or damage.

The Cloud SQL Proxy works by having a local client, called the proxy, running in the local environment. Your application communicates with the proxy with the standard database protocol used by your database. The proxy uses a secure tunnel to communicate with its companion process running on the server. Cloud SQL Proxy provides secure access to your Cloud SQL Second Generation instances without your having to whitelist IP addresses or configure SSL. The proxy uses the Cloud SQL API to authenticate with the Google Cloud Platform. You must enable the API before using the proxy, and you must provide the proxy with a valid user account.

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, with minimal refactoring required for migration to GCP.

For more information, see:
Google Cloud SQL Documentation: https://cloud.google.com/sql/docs/
Replication Options: https://cloud.google.com/sql/docs/mysql/replication/
Google Cloud SQL Backups:
https://cloud.google.com/sql/docs/mysql/backup-recovery/backups
Cloud SQL Proxy: https://cloud.google.com/sql/docs/mysql/sql-proxy



Cloud Spanner

Cloud Storage · Cloud Datastore · Cloud Bigtable · Cloud SQL · **Cloud Spanner** · BigQuery

Overview	Ideal for
● Mission-critical relational database service ● Transactional consistency ● Global scale ● High availability ● Multi-region replication ● 99.999% SLA	● Mission-critical applications ● High transactions ● Scale and consistency requirements

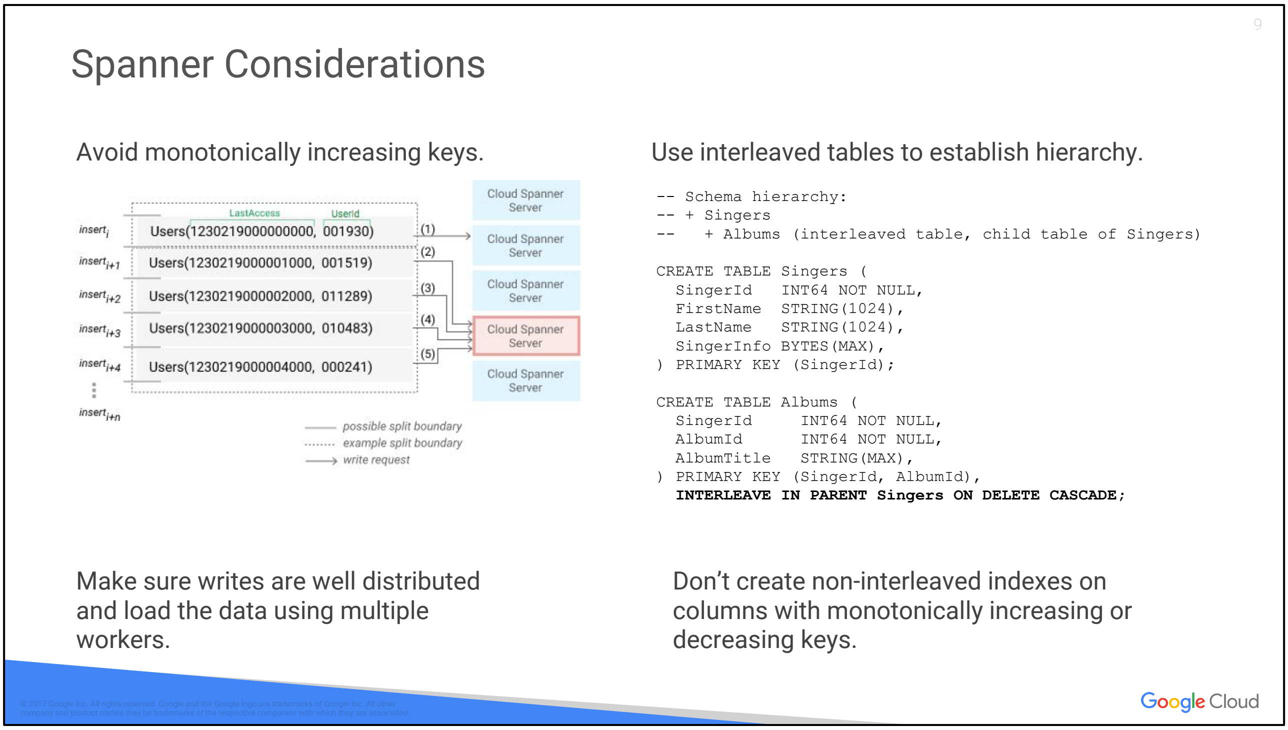
Google Cloud

Google Cloud Spanner is GCP's fully managed relational database service offering both strong consistency and horizontal scalability. It is designed for mission-critical OLTP applications. Cloud Spanner provides automatic, synchronous replication for high ability. 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.

Cloud Spanner and Cloud SQL differ in that Spanner uses real primary keys and has the notion of interleaved child records instead of foreign keys.

For more information, see the Cloud Spanner Documentation:
https://cloud.google.com/spanner/docs/



Spanner Considerations

Avoid monotonically increasing keys.

Make sure writes are well distributed and load the data using multiple workers.

Use interleaved tables to establish hierarchy.

```
-- Schema hierarchy:
-- + Singers
--   + Albums (interleaved table, child table of Singers)

CREATE TABLE Singers (
  SingerId   INT64 NOT NULL,
  FirstName  STRING(1024),
  LastName   STRING(1024),
  SingerInfo BYTES(MAX),
) PRIMARY KEY (SingerId);

CREATE TABLE Albums (
  SingerId     INT64 NOT NULL,
  AlbumId      INT64 NOT NULL,
  AlbumTitle   STRING(MAX),
) PRIMARY KEY (SingerId, AlbumId),
  INTERLEAVE IN PARENT Singers ON DELETE CASCADE;
```

Don’t create non-interleaved indexes on columns with monotonically increasing or decreasing keys.

Google Cloud

Avoid creating hotspots in your database in Cloud Spanner when choosing a primary key. Hotpots can be created by choosing a column whose value monotonically increases as the first key part because it results in all inserts occurring at the end of your key space. Cloud Spanner divides data among servers by key ranges, so all your inserts will be directed at a single server. This is pictured in the first diagram.

Possible fixes to this situation include:

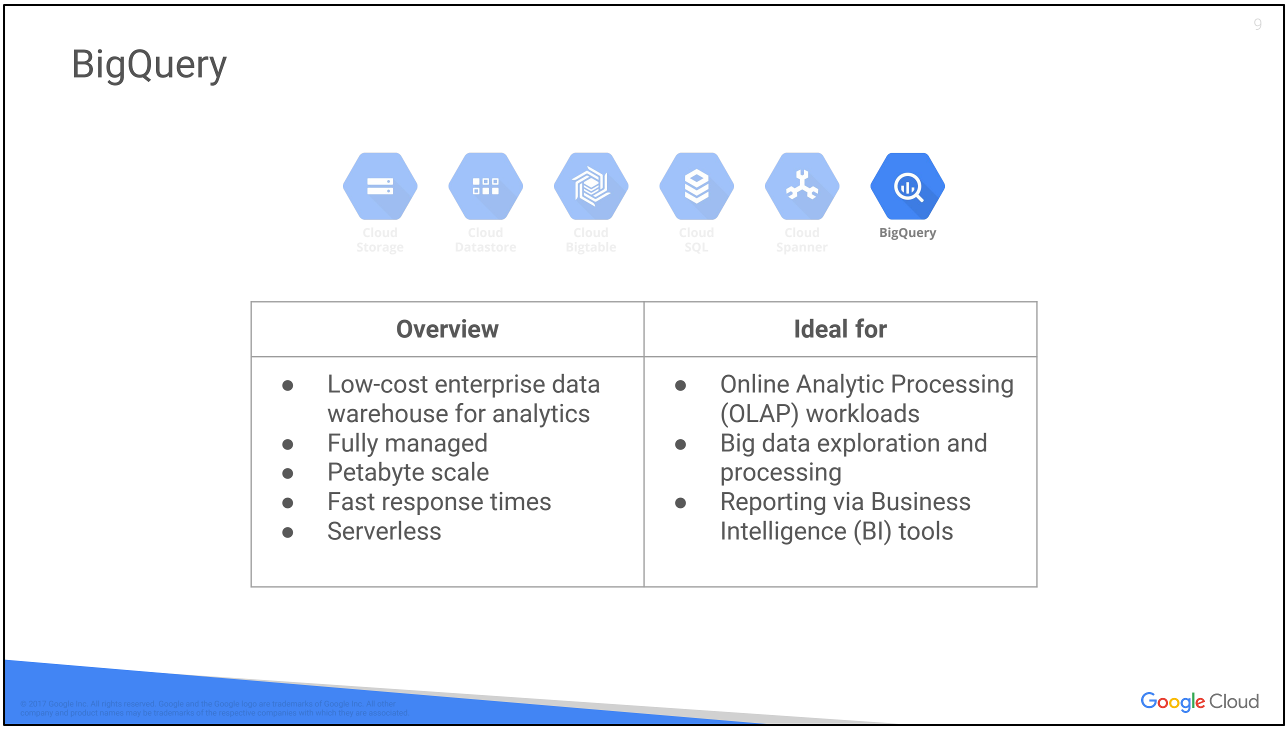
- swapping the order of the keys
- hashing the key and spreading the writes among N shard

For more information, see: https://cloud.google.com/spanner/docs/schema-design

An interleaved table is a table that you declare to be a child of another table because you want the rows of the child table to be physically stored together with the associated parent row. The prefix of the primary key of a child table must be the primary key of the parent table. The example shows how to define a table *Albums* as a child of *Singers*, where Singers is at the root of the database hierarchy.

When loading data in Cloud Spanner, make sure that writes are well distributed and load the data using multiple workers.

Avoid creating non-interleaved indexes on columns with monotonically increasing or decreasing keys, and create indexes after you bulk load your data.



BigQuery

Cloud Storage · Cloud Datastore · Cloud Bigtable · Cloud SQL · Cloud Spanner · BigQuery

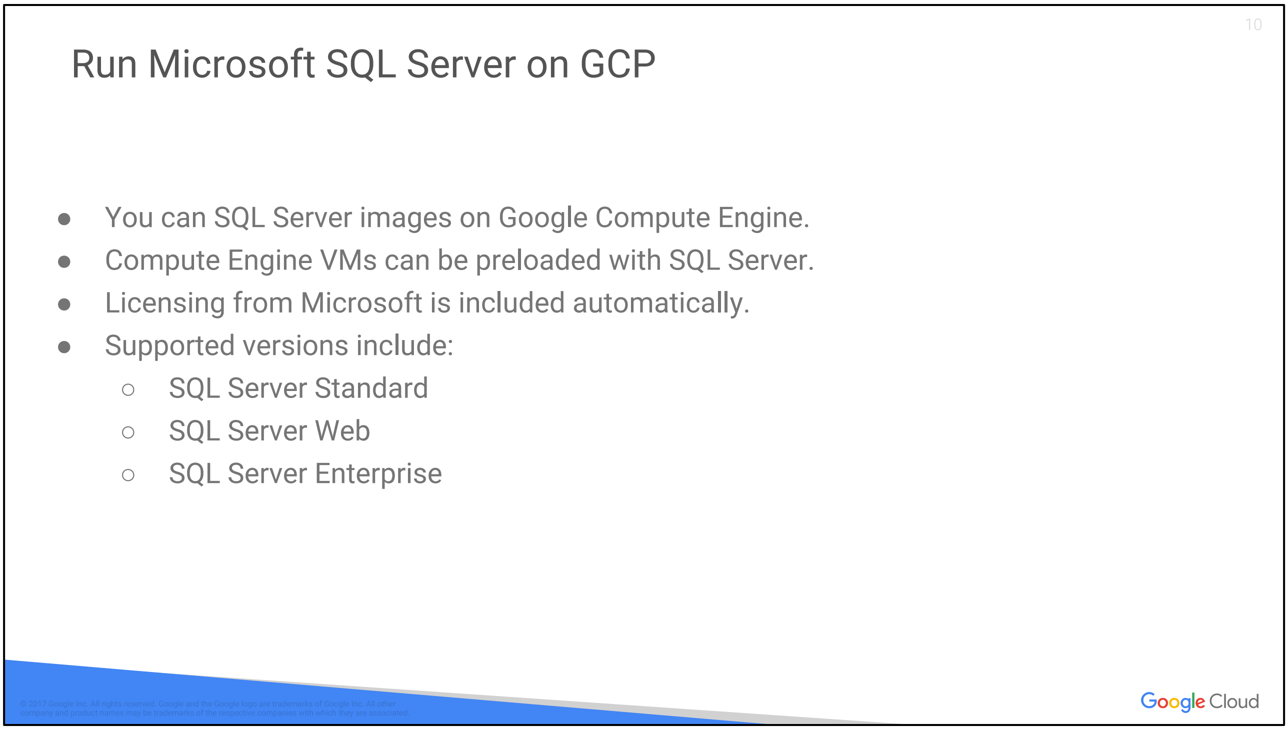
Overview	Ideal for
● Low-cost enterprise data warehouse for analytics ● Fully managed ● Petabyte scale ● Fast response times ● Serverless	● Online Analytic Processing (OLAP) workloads ● Big data exploration and processing ● Reporting via Business Intelligence (BI) tools

Google Cloud

BigQuery is a low-cost enterprise data warehouse for analytics. It is a fully managed service, meaning you don’t need to worry about administration of your data warehouse. BigQuery can scan TB in seconds and PB in minutes.

It is a great solution for Online Analytic Processing (OLAP) workloads, for big data exploration and processing, and to report via Business Intelligence (BI) tools.

For more information, see the BigQuery Documentation:
https://cloud.google.com/bigquery/docs/



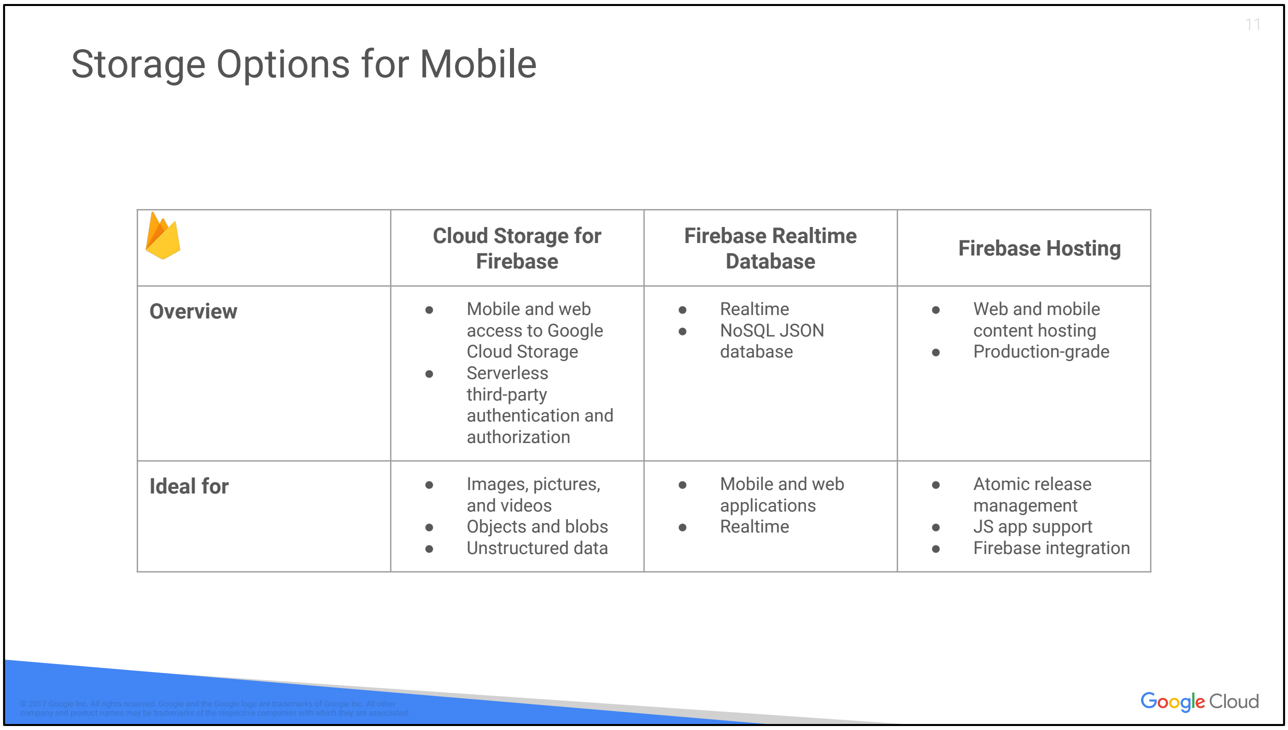
Run Microsoft SQL Server on GCP

- You can SQL Server images on Google Compute Engine.
- Compute Engine VMs can be preloaded with SQL Server.
- Licensing from Microsoft is included automatically.
- Supported versions include:
  - SQL Server Standard
  - SQL Server Web
  - SQL Server Enterprise

Google Cloud

You can run your Microsoft SQL Server deployment on GCP. Compute Engine VMs come preloaded with SQL Server, and licensing from Microsoft is included automatically. Supported versions include SQL Server Standard, SQL Server Web, and SQL Server Enterprise. Microsoft SQL Server on GCP is not a managed service like Cloud SQL or Spanner.

For more information, see: SQL Server images on GCE



Storage Options for Mobile

	Cloud Storage for Firebase	Firebase Realtime Database	Firebase Hosting
Overview	• Mobile and web access to Google Cloud Storage • Serverless third-party authentication and authorization	• Realtime • NoSQL JSON database	• Web and mobile content hosting • Production-grade
Ideal for	• Images, pictures, and videos • Objects and blobs • Unstructured data	• Mobile and web applications • Realtime	• Atomic release management • JS app support • Firebase integration

Google Cloud

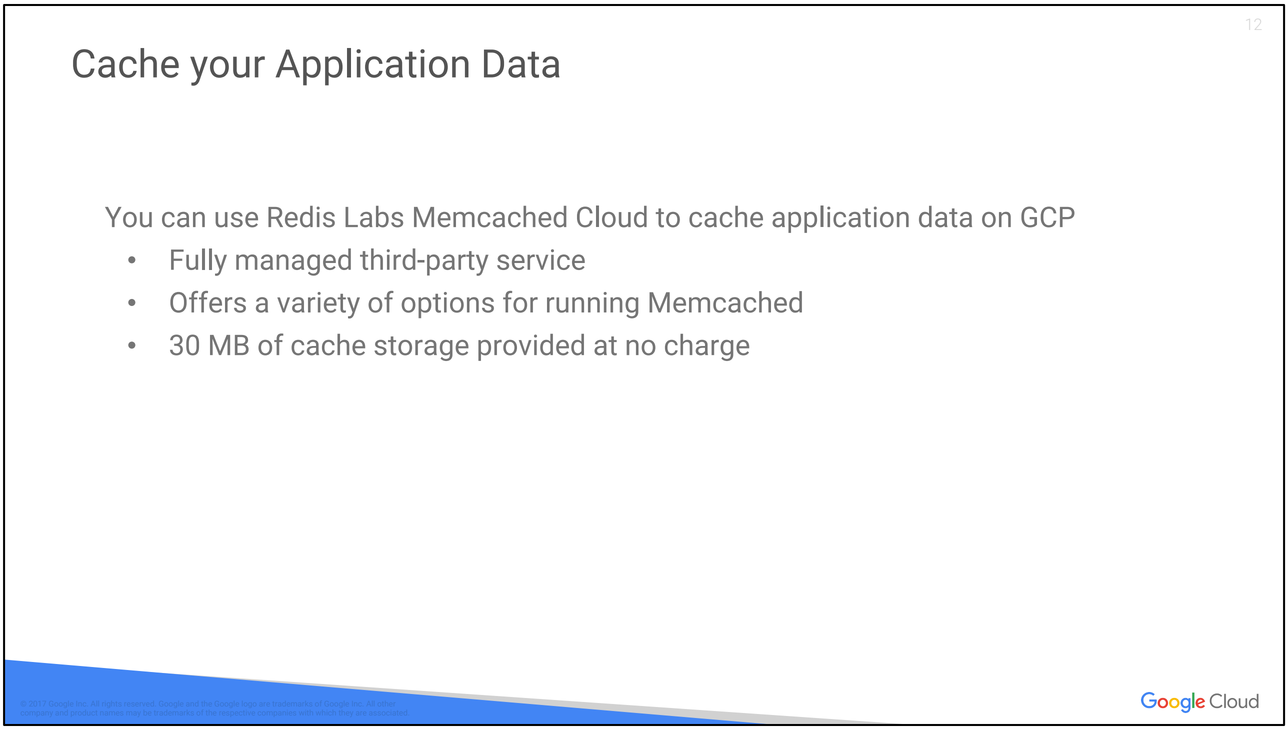
Firebase is a mobile and web application development platform with a few storage options for web and mobile development with GCP.

Cloud Storage for Firebase stores user-generated data and files to Google Cloud Storage. The Firebase SDKs for Cloud Storage provide simple and intuitive authentication for developers. Ideal use cases include saving user-generated images, pictures, videos, objects, and blogs from your mobile or web applications.

Firebase Realtime Database allows you to store and sync data with Firebase's NoSQL cloud database. Data is synced across all clients in real time so that it remains available when your app goes offline. It is an ideal storage solution for mobile and web applications that require responsiveness when offline.

Firebase Hosting is a fast and secure way to host static resources for your web app. It is ideal for URL rewriting, atomic release management, and firebase integration of your web or mobile app.

Cloud Storage for Firebase Documentation: https://firebase.google.com/docs/storage/
Firebase Realtime Database Documentation:
https://firebase.google.com/docs/database/
Firebase Hosting Documentation: https://firebase.google.com/docs/hosting/



Cache your Application Data

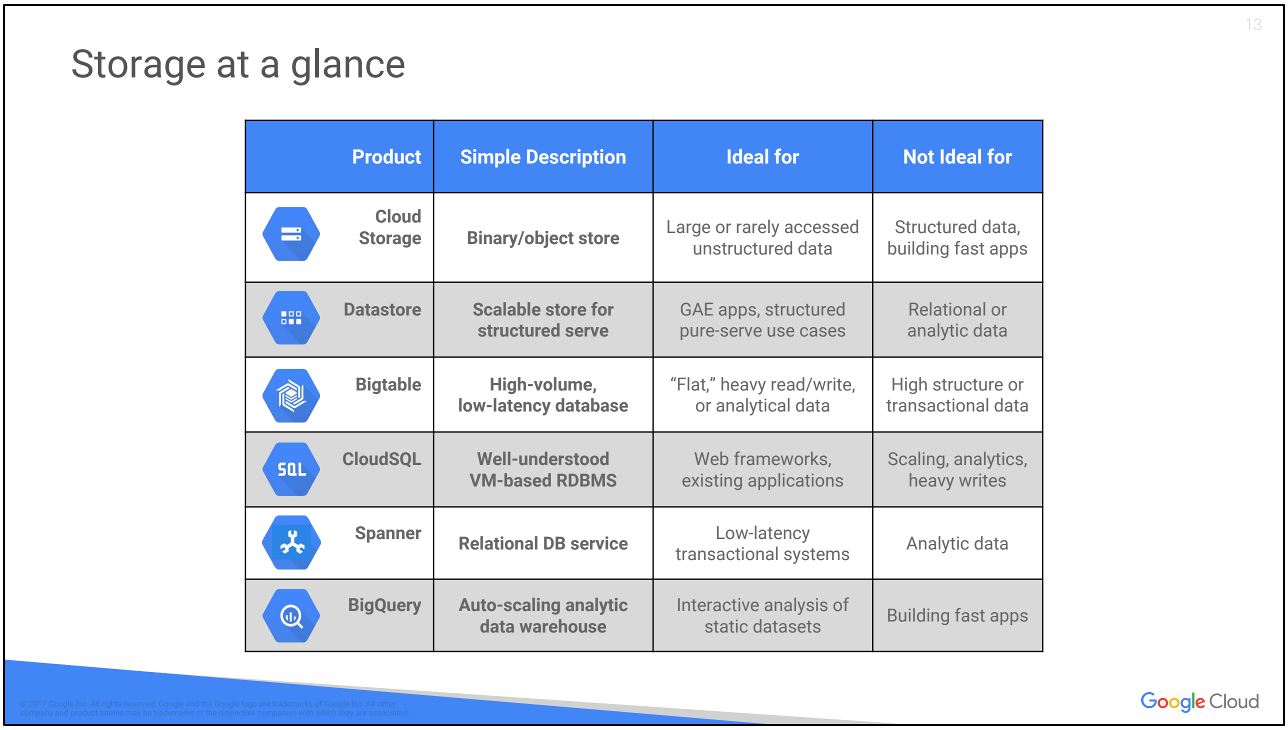
You can use Redis Labs Memcached Cloud to cache application data on GCP

- Fully managed third-party service
- Offers a variety of options for running Memcached
- 30 MB of cache storage provided at no charge

Google Cloud

You can cache application data on GCP by using third-party Redis Labs Memcached Cloud. Memcached Cloud is a fully managed service, providing a variety of options for running Memcached. Redis Labs provides 30 MB of cache storage at no charge.

For more information, see:
https://cloud.google.com/appengine/docs/flexible/python/using-redislabs-memcache
https://redislabs.com/products/memcached-cloud/

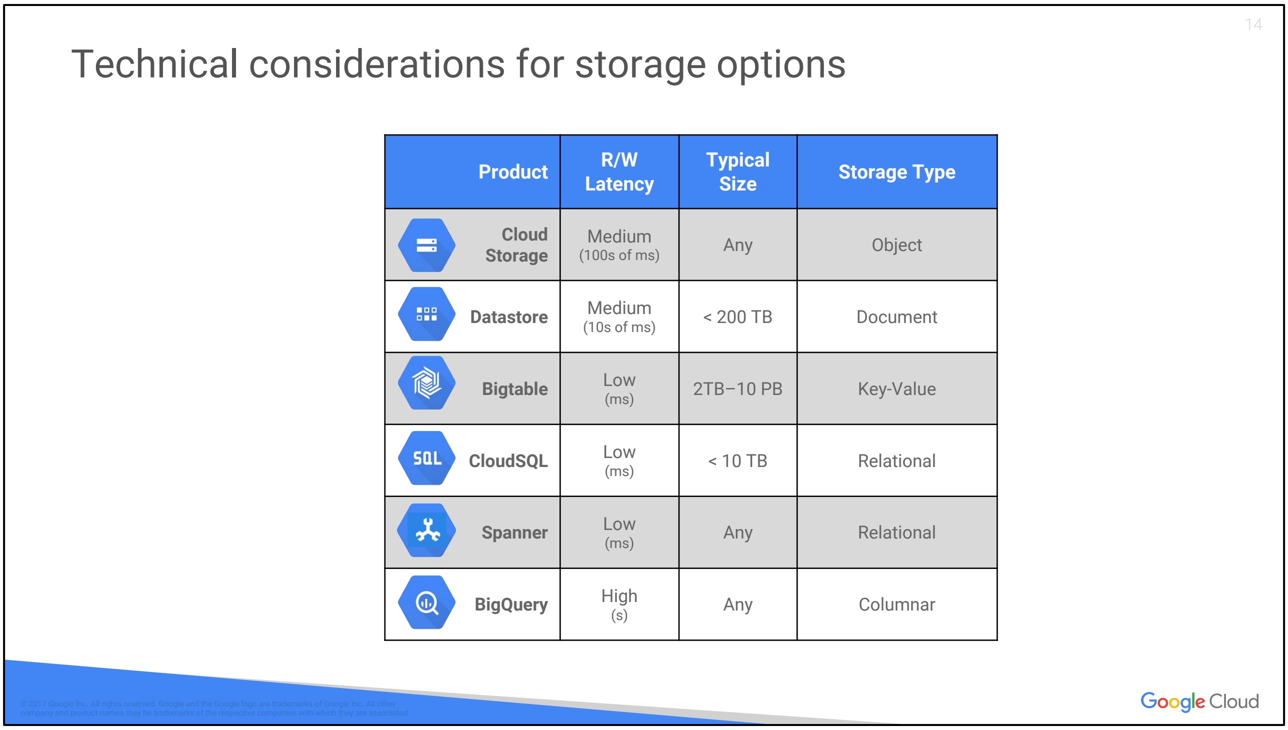


Storage at a glance

Product	Simple Description	Ideal for	Not Ideal for
Cloud Storage	**Binary/object store**	Large or rarely accessed unstructured data	Structured data, building fast apps
Datastore	**Scalable store for structured serve**	GAE apps, structured pure-serve use cases	Relational or analytic data
Bigtable	**High-volume, low-latency database**	“Flat,” heavy read/write, or analytical data	High structure or transactional data
CloudSQL	**Well-understood VM-based RDBMS**	Web frameworks, existing applications	Scaling, analytics, heavy writes
Spanner	**Relational DB service**	Low-latency transactional systems	Analytic data
BigQuery	**Auto-scaling analytic data warehouse**	Interactive analysis of static datasets	Building fast apps

Google Cloud

Here are the GCP storage options at a glance. When you are choosing the right storage option for your application, it’s important to understand what a product is and isn’t ideal for by design.



Technical considerations for storage options

Product	R/W Latency	Typical Size	Storage Type
Cloud Storage	Medium (100s of ms)	Any	Object
Datastore	Medium (10s of ms)	< 200 TB	Document
Bigtable	Low (ms)	2TB–10 PB	Key-Value
CloudSQL	Low (ms)	< 10 TB	Relational
Spanner	Low (ms)	Any	Relational
BigQuery	High (s)	Any	Columnar

Google Cloud

Other considerations for choosing a storage option for your application include the interface, read/write latency, typical size of your data, and storage type. Refer to the table to identify the best storage option for your application.