

Storage in the Cloud

GCP Fundamentals: Core Infrastructure

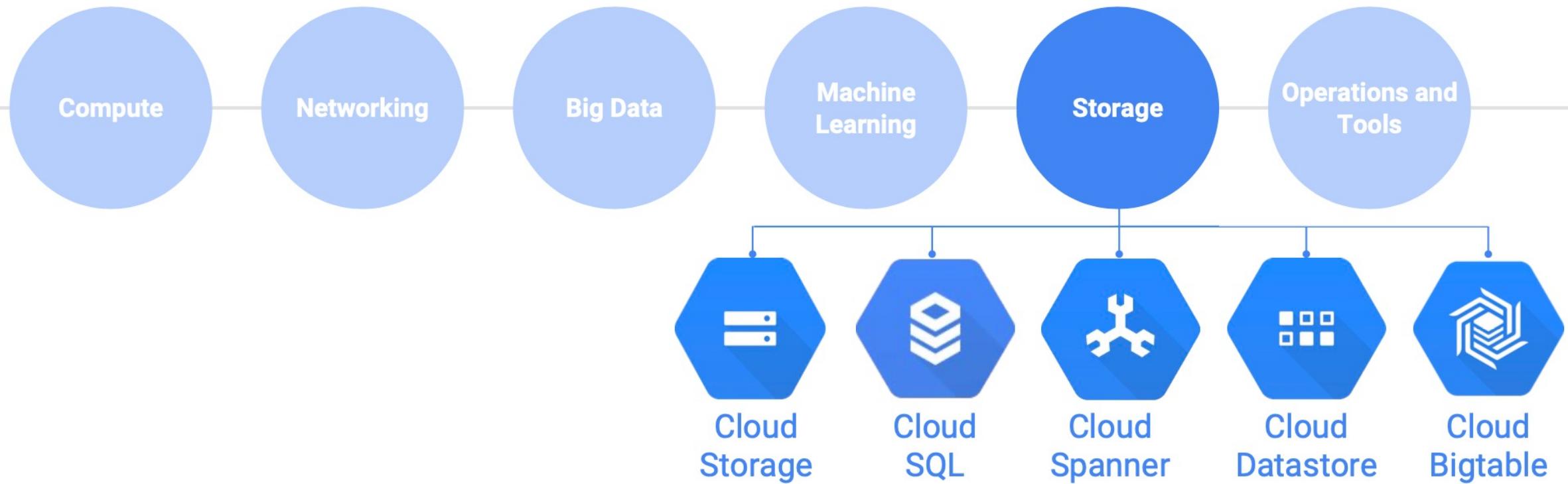


Getting Started with Cloud Storage and Cloud SQL



Google Cloud

Google Cloud Platform



Agenda

Cloud Storage

Cloud Bigtable

Cloud SQL and Cloud Spanner

Cloud Datastore

Comparing storage options

Labs

Cloud Storage is binary large-object storage

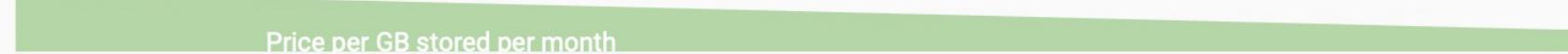
- High performance, internet-scale
- Simple administration
 - Does not require capacity management
- Data encryption at rest
- Data encryption in transit by default from Google to endpoint
- Online and offline import services are available



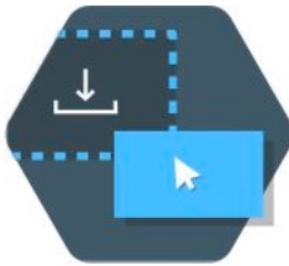
Your Cloud Storage files are organized into buckets

Bucket attributes	Bucket contents
Globally unique name	Files (in a flat namespace)
Storage class	
Location (region or multi-region)	
IAM policies or Access Control Lists	Access Control Lists
Object versioning setting	
Object lifecycle management rules	

Choosing among Cloud Storage classes

	Multi-regional	Regional	Nearline	Coldline
Intended for data that is...	Most frequently accessed	Accessed frequently within a region	Accessed less than once a month	Accessed less than once a year
Availability SLA	99.95%	99.90%	99.00%	99.00%
Access APIs	Consistent APIs			
Access time	Millisecond access			
<u>Storage price</u>	 Price per GB stored per month			
<u>Retrieval price</u>	 Total price per GB transferred			
Use cases	Content storage and delivery	In-region analytics, transcoding	Long-tail content, backups	Archiving, disaster recovery

There are several ways to bring data into Cloud Storage



Online transfer

Self-managed copies
using command-line
tools or drag-and-drop



Storage Transfer Service

Scheduled, managed batch transfers

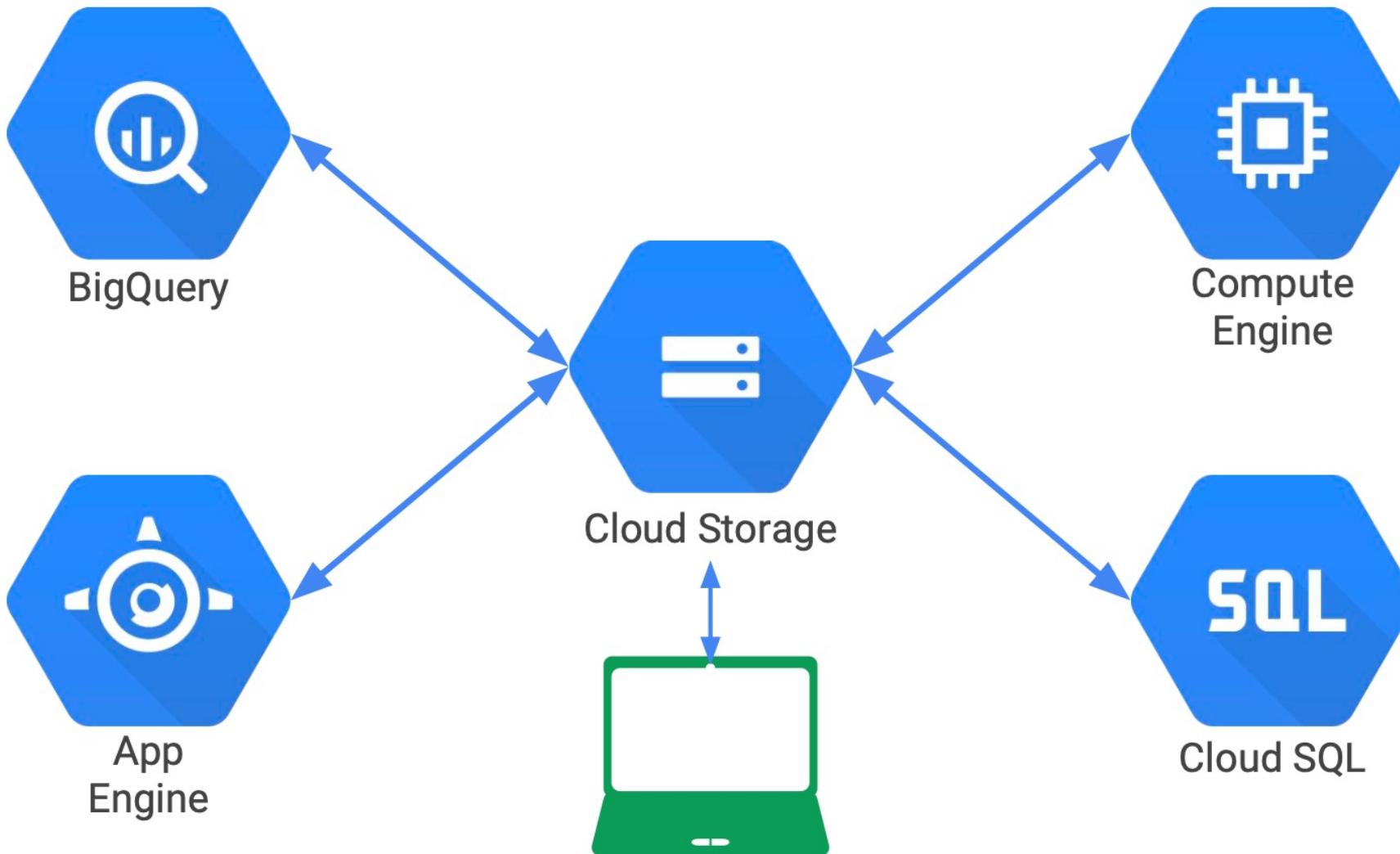


Transfer Appliance Beta

Rackable appliances to securely
ship your data

Cloud Storage works with other GCP services

Import
and
export
tables



Startup
scripts,
images,
and
general
object
storage

Object
storage,
logs, and
Datastore
backups

Import
and
export
tables

Agenda

Cloud Storage

Cloud Bigtable

Cloud SQL and Cloud Spanner

Cloud Datastore

Comparing storage options

Integrations with other services

Labs

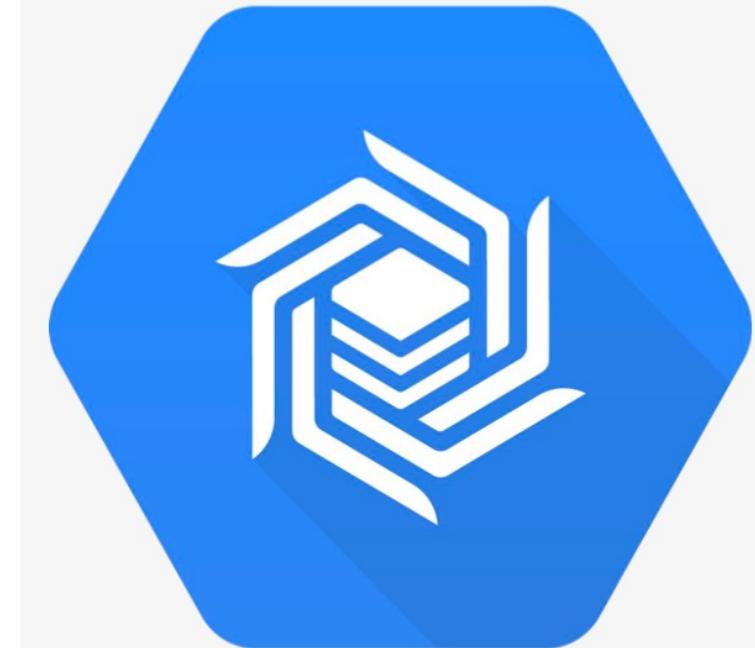
Cloud Bigtable is managed NoSQL

- Fully managed NoSQL, wide-column database service for terabyte applications
- Integrated
 - Accessed using HBase API
 - Native compatibility with big data, Hadoop ecosystems

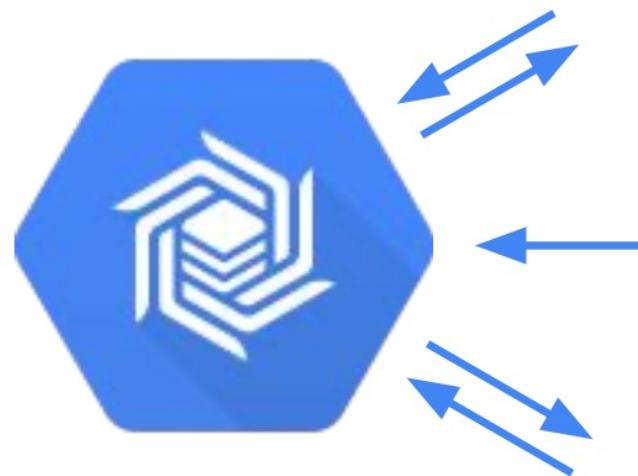


Why choose Cloud Bigtable?

- Replicated storage
- Data encryption in-flight and at rest
- Role-based ACLs
- Drives major applications such as Google Analytics and Gmail



Bigtable Access Patterns



Application API

Data can be read from and written to Cloud Bigtable through a data service layer like Managed VMs, the HBase REST Server, or a Java Server using the HBase client. Typically this will be to serve data to applications, dashboards, and data services.



Streaming

Data can be streamed in (written event by event) through a variety of popular stream processing frameworks like Cloud Dataflow Streaming, Spark Streaming, and Storm.



Batch Processing

Data can be read from and written to Cloud Bigtable through batch processes like Hadoop MapReduce, Dataflow, or Spark. Often, summarized or newly calculated data is written back to Cloud Bigtable or to a downstream database.

Agenda

Cloud Storage

Cloud Bigtable

Cloud SQL and Cloud Spanner

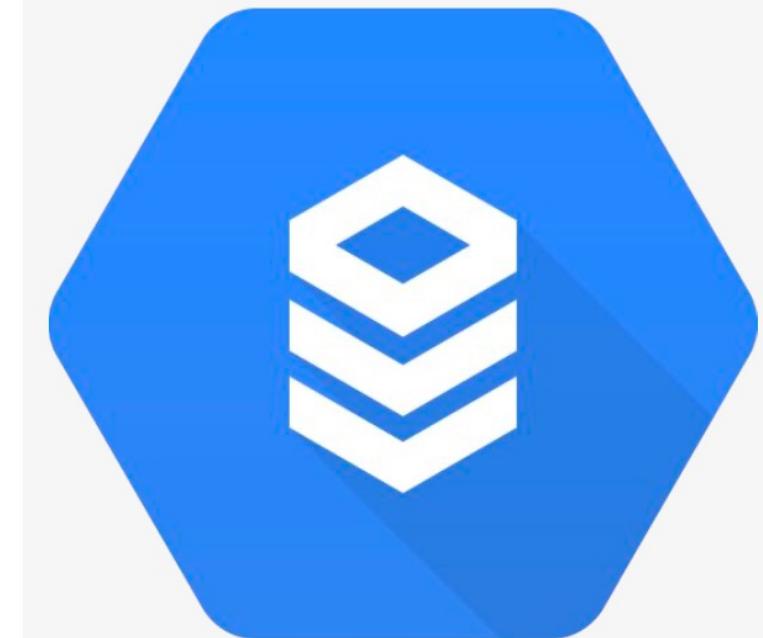
Cloud Datastore

Comparing storage options

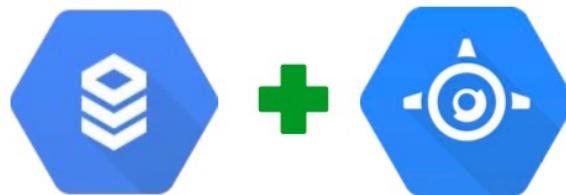
Labs

Cloud SQL is a managed RDBMS

- Offers MySQL and PostgreSQL databases as a service
- Automatic replication
- Managed backups
- Vertical scaling (read and write)
- Horizontal scaling (read)
- Google security

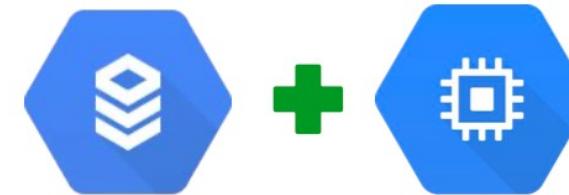


Cloud SQL can be used with other GCP services



Cloud SQL can be used with App Engine using standard drivers.

You can configure a Cloud SQL instance to follow an App Engine application.



Compute Engine instances can be authorized to access Cloud SQL instances using an external IP address.

Cloud SQL instances can be configured with a preferred zone.



Cloud SQL can be used with external applications and clients.

Standard tools can be used to administer databases.

External read replicas can be configured.

Cloud Spanner is a horizontally scalable RDBMS

Cloud Spanner supports:

- Automatic replication
- Strong global consistency
- Managed instances with high availability
- SQL (ANSI 2011 with extensions)



Agenda

Cloud Storage

Cloud Bigtable

Cloud SQL and Cloud Spanner

Cloud Datastore

Comparing storage options

Labs

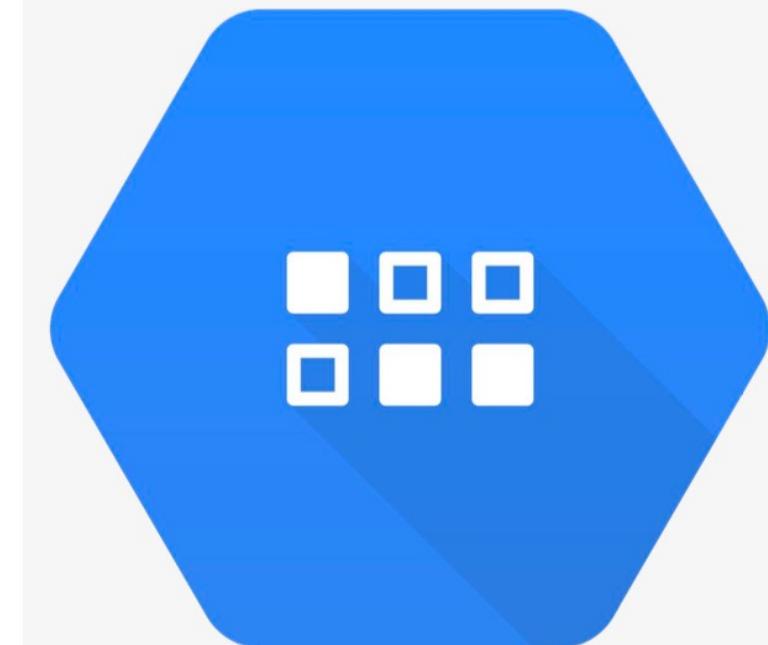
Cloud Datastore is a horizontally scalable NoSQL DB

- NoSQL designed for application backends
- Fully managed
 - Uses a distributed architecture to automatically manage scaling
- Built-in redundancy
- Supports ACID transactions



Google Cloud Datastore: benefits

- Schemaless access
 - No need to think about underlying data structure
- Local development tools
- Includes a free daily quota
- Access from anywhere through a [RESTful interface](#)



Agenda

Cloud Storage

Cloud Bigtable

Cloud SQL and Cloud Spanner

Cloud Datastore

Comparing storage options

Integrations with other services

Labs

Comparing storage options: technical details

	Cloud Datastore	Bigtable	Cloud Storage	Cloud SQL	Cloud Spanner	BigQuery
Type	NoSQL document	NoSQL wide column	Blobstore	Relational SQL for OLTP	Relational SQL for OLTP	Relational SQL for OLAP
Transactions	Yes	Single-row	No	Yes	Yes	No
Complex queries	No	No	No	Yes	Yes	Yes
Capacity	Terabytes+	Petabytes+	Petabytes+	Up to ~10 TB	Petabytes	Petabytes+
Unit size	1 MB/entity	~10 MB/cell ~100 MB/row	5 TB/object	Determined by DB engine	10,240 MiB/row	10 MB/row

Comparing storage options: use cases

	Cloud Datastore	Cloud Bigtable	Cloud Storage	Cloud SQL	Cloud Spanner	BigQuery
Type	NoSQL document	NoSQL wide column	Blobstore	Relational SQL for OLTP	Relational SQL for OLTP	Relational SQL for OLAP
Best for	Semi-structured application data, durable key-value data	“Flat” data, Heavy read/write, events, analytical data	Structured and unstructured binary or object data	Web frameworks, existing applications	Large-scale database applications (> ~2 TB)	Interactive querying, offline analytics
Use cases	Getting started, App Engine applications	AdTech, Financial and IoT data	Images, large media files, backups	User credentials, customer orders	Whenever high I/O, global consistency is needed	Data warehousing

Agenda

Google Cloud Storage

Google Cloud Bigtable

Google Cloud SQL and Google Cloud Spanner

Google Cloud Datastore

Comparing Storage Options

Integrations with Storage Services

Labs

Lab Instructions

In this lab you will create a Google Cloud Storage bucket and place an image in it. You'll also configure an application running in Google Compute Engine to use a database managed by Google Cloud SQL and to reference the image in the Cloud Storage bucket.

- Create a Cloud Storage bucket and place an image into it
- Create a Cloud SQL instance and configure it
- Connect to a Cloud SQL instance from a web server
- Use an image stored in a Cloud Storage bucket in a web page

More resources

Overview of Cloud Storage

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

Getting started with Google Cloud SQL

<https://cloud.google.com/sql/docs/quickstart>

Cloud Bigtable

<https://cloud.google.com/stackdriver/docs/>

Cloud Spanner

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

Cloud Datastore

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