Google Cloud Platform

Google Cloud Platform Storage Options

Google Cloud Platform Fundamentals

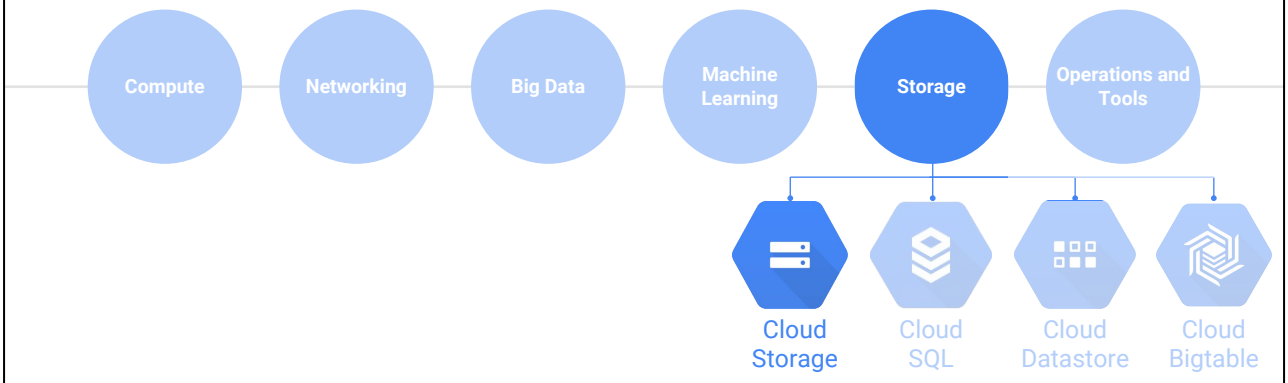
V2.1

Timing: Approximately 45 minutes

Agenda

1. Google Cloud Storage
2. Google Cloud Bigtable
3. Google Cloud SQL
4. Comparing Storage Options
5. Quiz & Lab

Google Cloud Platform



Notes:
Google Cloud Platform has many storage options that satisfy nearly every customer use case. Cloud Datastore was discussed previously. In this module, we turn our attention to the remaining core storage options: Google Cloud Storage, Google Cloud SQL, and Google Cloud Bigtable. First, is Cloud Storage.

Google Cloud Storage (1 of 2)

- High performance, internet-scale, immutable BLOB (binary large object) storage
- *Not* a file system (but can be accessed as one via 3rd party tools such as Cloud Storage Fuse)

Notes:
Google Cloud Storage offers developers and IT organizations durable and highly available object storage.
With no minimum fee and a pay for what you use model, Google Cloud Storage capacity pricing is the most cost effective in the market.

Google Cloud Storage (2 of 2)

- Simple administration and does not require capacity management
- Data encryption in-flight and at rest
- All storage classes accessed through the same APIs

Notes:
Store your data on Google's infrastructure with very high level of durability and availability. Google Cloud Storage stores and replicates your data allowing a high level of persistence. Google Cloud Storage is almost infinitely scalable. Whether you're supporting a small application or building a large, multi-petabyte system, Cloud Storage can handle it.

Cloud Storage Classes

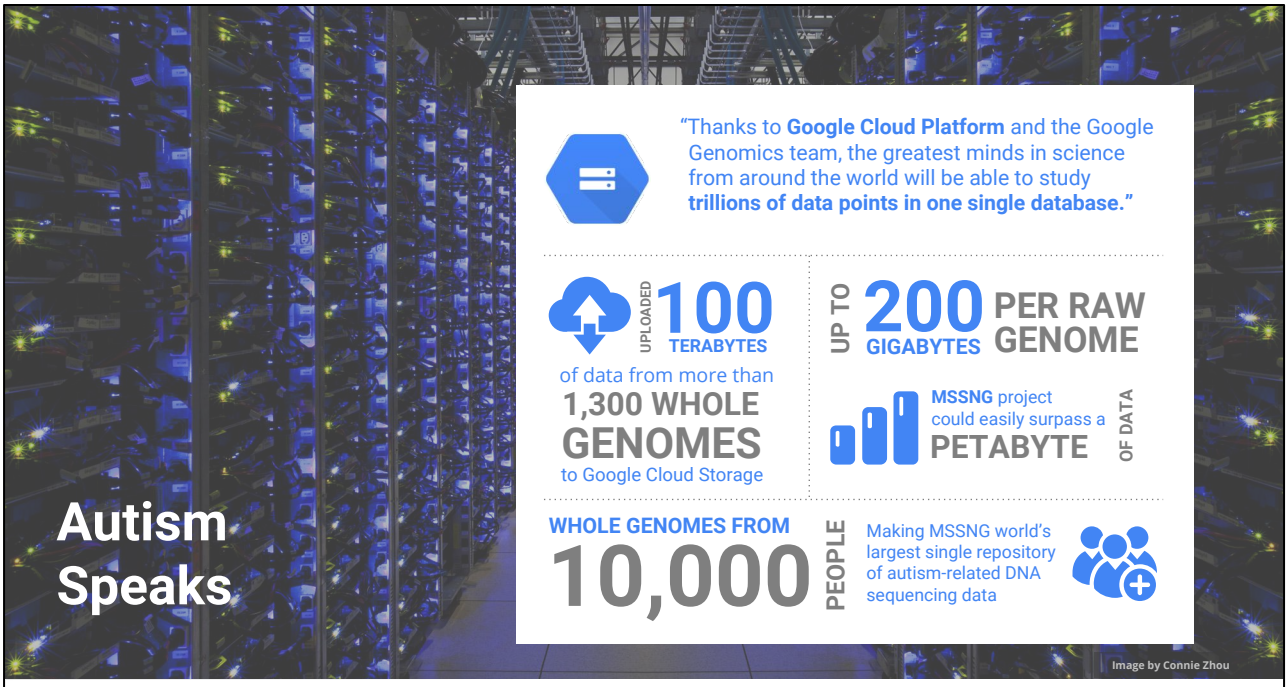
Standard Storage provides the highest durability, availability and **performance** with **low latency** and is ideal for use with website content distribution and video streaming

Durable Reduced Availability Storage offers the **same durability** as Standard Storage but with a lower availability SLA at a **reduced cost**

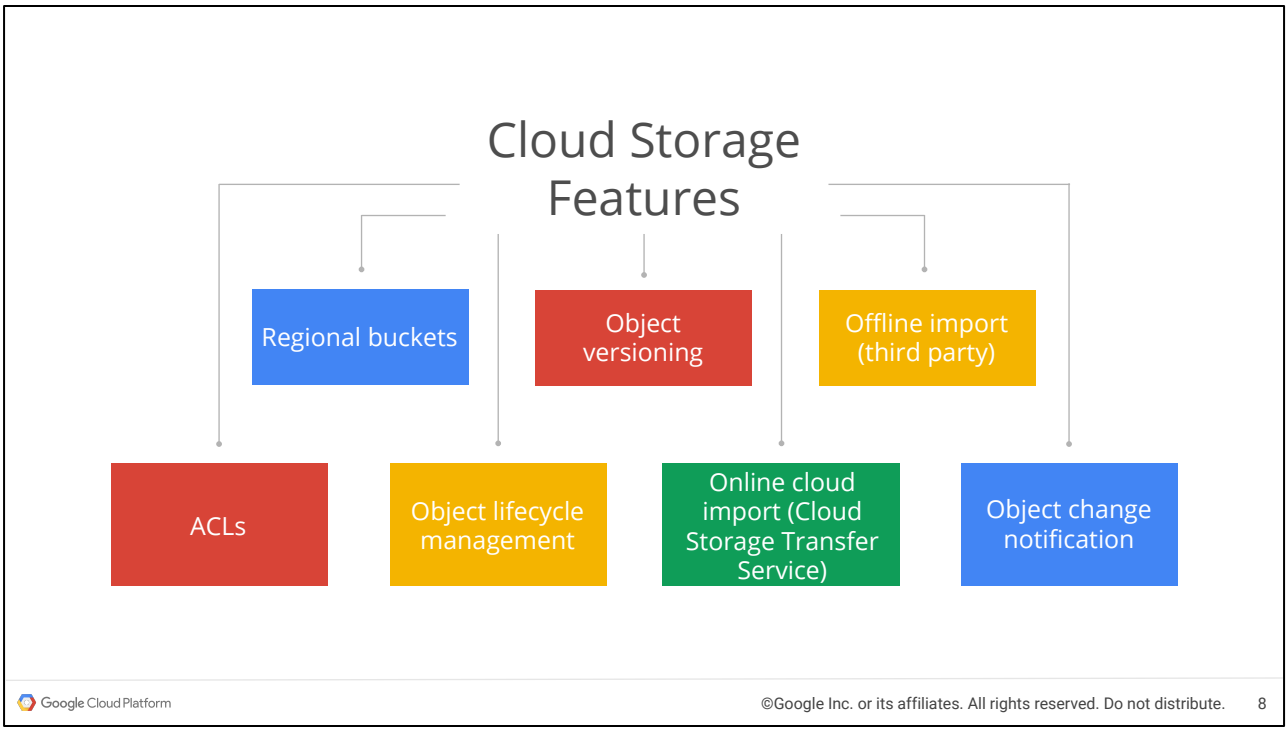
Nearline Storage offers **low-cost,** highly durable storage service for data **archiving**, online **backup**, and **disaster recovery**, without having to wait hours or days to retrieve or access your data

Each option comes with detailed pricing

Notes:
Google Cloud Storage offers three types of storage: Standard Storage, Durable Reduced Availability (DRA) Storage, and Nearline Storage with different availability, latency, and throughput characteristics. All storage classes offer the same high level of durability.



Notes:
Read more about Autism Speaks here:
https://cloud.google.com/customers/autism-speaks/.



Notes:
Regional and multi-regional buckets

A Cloud Storage bucket location can be a multi-region location or a regional location. Typically, a good location for your bucket balances latency, availability, and bandwidth costs for applications and users of the bucket data.

A regional location is a specific geographic location within a multi-region location. You can create a bucket configured for all storage classes in the following regional locations:

- asia-east1 — Eastern Asia-Pacific
- europe-west1 — Western Europe
- us-central1 — Central United States
- us-east1 — Eastern United States

A multi-region location is a location that spans multiple regional locations. You can configure a bucket for all storage classes in the following multi-region locations:

- asia — Asia Pacific
- eu — European Union
- us — United States

Object versioning
Google Cloud Storage allows you to enable Object Versioning at the bucket level. Once enabled, a history of modifications (overwrite / delete) of objects is kept for all objects in the bucket. You can list archived versions of an object, restore an object to an older state, or permanently delete a version, as needed. All objects have generation properties associated with them to help you identify the objects and perform safe read-modify-write updates on them as well as perform conditional operations on them.

ACLs
Google Cloud Storage uses access control lists (ACLs) to manage bucket and object access. ACLs are the mechanism you use to share objects with other users and allow other users to access your buckets and objects.
An ACL consists of one or more entries, where each entry grants permissions to a scope. Permissions define the actions that can be performed against a bucket or object (for example, read or write); the scope defines who the permission applies to (for example, a specific user or group of users). Scopes are sometimes referred to as grantees. The maximum number of ACL entries you can create for a bucket or object is 100. When the ACL scope is a group or domain, it counts as one ACL entry regardless of how many users are in the group or domain.

When a user requests access to a bucket or object, the Google Cloud Storage system reads the bucket or object ACL and determines whether to allow or reject the access request. If the ACL grants the user permission for the requested operation, the request is allowed. If the ACL does not grant the user permission for the requested operation, the request fails and a 403 Forbidden error (Access Denied) is returned.

Offline and online import
Offline Media Import / Export is a third party solution that allows you to load data into Google Cloud Storage by sending your physical media, such as hard disk drives (HDDs), tapes, and USB flash drives, to a third party service provider who uploads data on your behalf. Offline Media Import / Export is helpful if you're limited to a slow, unreliable, or expensive Internet connection.

Offline import is available through third party providers:
https://cloud.google.com/storage/docs/offline-media-import-export
Cloud Storage Transfer Service enables you to import large amounts of online data into Google Cloud Storage quickly and cost-effectively. To use Cloud Storage Transfer Service, you set up a transfer from a data source to data sink. Data sources can be an Amazon Simple Storage Service (Amazon S3) bucket, an HTTP/HTTPS location, or another Google Cloud Storage bucket.

Data sinks are always a Google Cloud Storage bucket.

Example uses of Cloud Storage Transfer Service include:

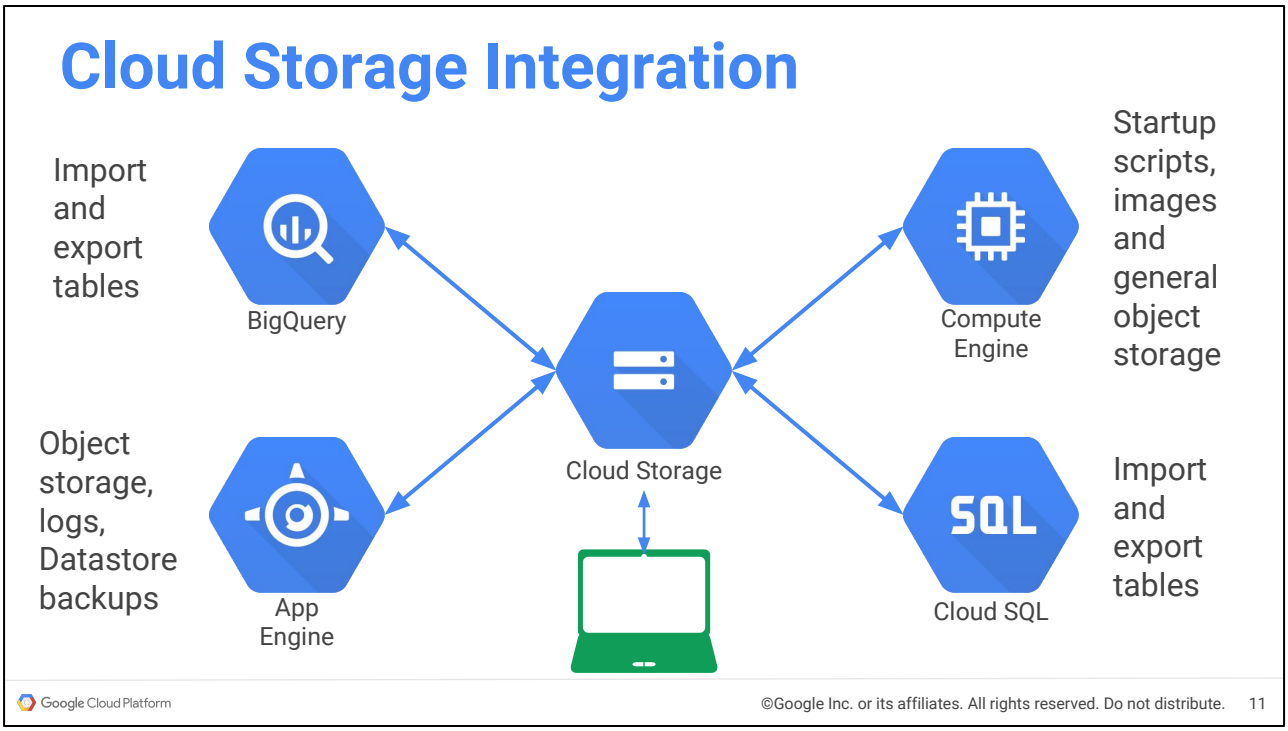
- Backing up data to a Google Cloud Storage bucket from other storage providers.
- Moving data from a Standard Storage bucket to a Nearline Storage bucket to lower your storage costs.

Object lifecycle management
You can define lifecycle management policies to let Google Cloud Storage perform automatic actions on your objects based on certain conditions. Here are some example use cases:

- Delete objects older than 365 days.
- Delete objects created before January 1, 2013.
- Keep only the 3 most recent versions of each object in a bucket with versioning enabled.

Object change notification
A client application can send a request to watch for changes to the objects in a particular bucket. Completing a watch request creates a new notification channel. A notification channel is the means by which a notification message is sent to an application watching a bucket. Only one type of notification channel, a web hook, is currently supported.

After a notification channel is initiated, Google Cloud Storage notifies the application any time an object is added, updated, or removed from the bucket. For example, when you add a new picture to a bucket, an application could be notified to create a thumbnail.



Notes:
Google Cloud Storage is tightly integrated with many of the Google Cloud Platform products and services. Cloud Storage is often the ingestion point for data being moved into the cloud, and is frequently the long-term storage location for data.

Integration between Cloud Storage and other services includes using Cloud Storage to:

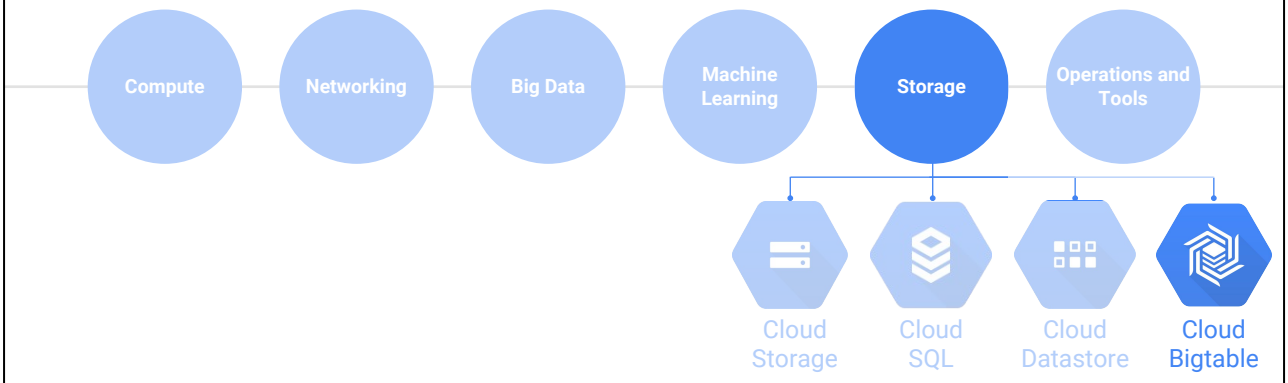
- Store tables imported into BigQuery and to store tables exported from BigQuery
- Store App Engine logs, Cloud Datastore backups, and objects used by App Engine applications (such as images)
- Store instance startup scripts, Compute Engine images, and objects used by Compute Engine applications
- Store tables imported into Cloud SQL and to store tables exported from Cloud SQL

For more information on Cloud Storage integration with other Google Cloud Platform services, see:
https://cloud.google.com/storage/docs/google-integration.

Agenda

1. Google Cloud Storage
2. Google Cloud Bigtable
3. Google Cloud SQL
4. Comparing Storage Options
5. Quiz & Lab

Google Cloud Platform



Notes:
The next storage option on Google Cloud Platform is Google Cloud Bigtable.

Google Cloud Bigtable (1 of 2)

- Fully managed, NoSQL, wide-column database service for large-workload applications - Terabytes to petabytes
- Integrated
  - Accessed using HBase API
  - Native compatibility with big data, Hadoop ecosystems

Notes:
Cloud Bigtable is offered as a fully managed service, meaning you spend your time developing valuable applications instead of configuring and tuning your database for performance and scalability. In addition, Google's own Bigtable operations team monitors the service to ensure issues are addressed quickly.

Cloud Bigtable is ideal for applications that need very high throughput and scalability for non-structured key/value data, where each value is typically no larger than 10 MB. Cloud Bigtable also excels as a storage engine for batch MapReduce operations, stream processing/analytics, and machine-learning applications.

You can use Cloud Bigtable to store and query all of the following types of data:

- Marketing data such as purchase histories and customer preferences.
- Financial data such as transaction histories, stock prices, and currency exchange rates.
- Internet of Things data such as usage reports from energy meters and home appliances.
- Time-series data such as CPU and memory usage over time for multiple servers.

Cloud Bigtable is offered through the same open source API as HBase, the native Hadoop database. This enables portability of applications between HBase and Bigtable.

Google Cloud Bigtable (2 of 2)

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

Notes:
Cloud Bigtable is Google's NoSQL Big Data database service. It's the same database that powers many core Google services, including Search, Analytics, Maps, and Gmail. Bigtable is designed to handle massive workloads at consistent low latency and high throughput, so it's a great choice for both operational and analytical applications, including IoT, user analytics, and financial data analysis.

Customers frequently choose Bigtable if the data is:
Big
- Large quantities (>1TB) of semi-structured or structured data

Fast
- Data is high throughput or rapidly changing

NoSQL
- Transactions, strong relational semantics not required

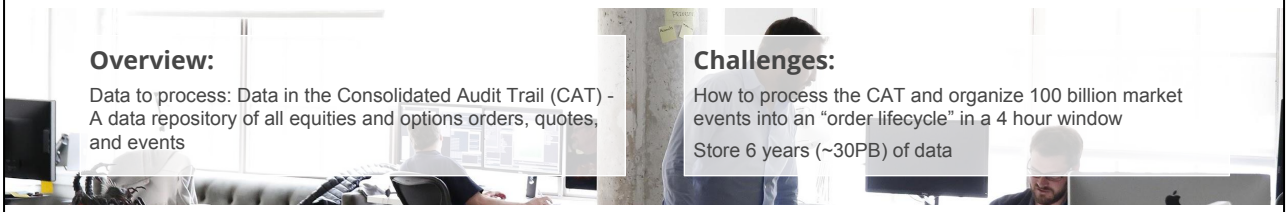
And especially if it is:
Time series
- Data is time-series or has natural semantic ordering

Big data
- You run asynchronous batch or real-time processing on the data

Machine learning

- You run machine learning algorithms on the data

SUNGARD



Overview:

Data to process: Data in the Consolidated Audit Trail (CAT) - A data repository of all equities and options orders, quotes, and events

Challenges:

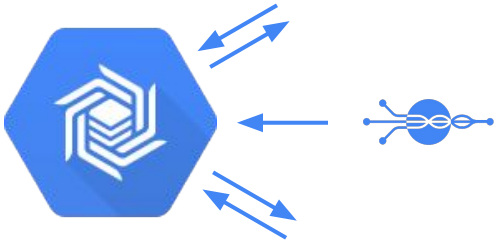
How to process the CAT and organize 100 billion market events into an "order lifecycle" in a 4 hour window

Store 6 years (~30PB) of data

Cloud Bigtable to process and run queries and tolerate volume increases

Notes:
Read more about Sungard's implementation of Bigtable here: https://cloud.google.com/bigtable/pdf/FISCATCaseStudy.pdf?authuser=2.

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, a Java Server using the HBase client. Typically this will be to serve data to applications, dashboards and data services.

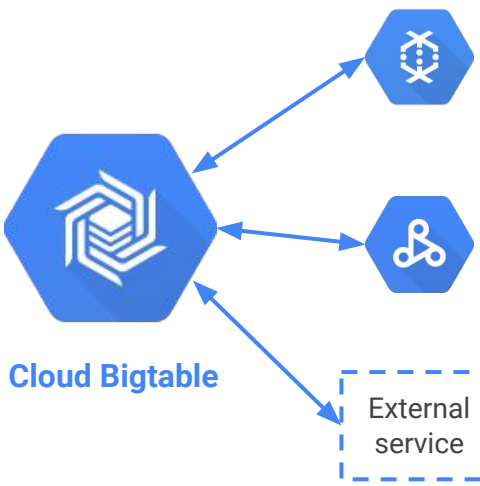
Streaming

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, Spark. Often, summarized or newly calculated data is written back to Cloud Bigtable or to a downstream database.

Cloud Bigtable Integration



Google Cloud Dataflow

Use Cloud Dataflow connector for Bigtable for batch and streaming operations in pipelines

Google Cloud Dataproc

Use Bigtable HBase client to integrate Hadoop jobs with Cloud Dataproc

On-premises, cloud-based Hadoop

Use Bigtable HBase client to integrate with Hadoop clusters

Notes:

Dataflow Connector for Cloud Bigtable

The Cloud Dataflow connector for Cloud Bigtable makes it possible to use Cloud Bigtable in a Cloud Dataflow pipeline. You can use the connector for both batch and streaming operations. The connector is written in Java and is built on the HBase client for Cloud Bigtable. You can find the connector on GitHub in the repository GoogleCloudPlatform/cloud-bigtable-client.

For more information on the Cloud Dataflow connector, see the full API documentation.

Cloud Bigtable and the HBase API

One way to communicate with Cloud Bigtable is through the Cloud Bigtable HBase client, which is a customized version of the standard Apache HBase client. The Cloud Bigtable HBase client allows you to write Java applications that communicate with Cloud Bigtable through the open-source HBase API. Source code for the client is available in the GitHub repository GoogleCloudPlatform/cloud-bigtable-client.

When you use the Cloud Bigtable HBase client, you don't need to run the HBase server. You are simply using a modified version of the HBase client interface. Google runs and maintains the servers for the Cloud Bigtable

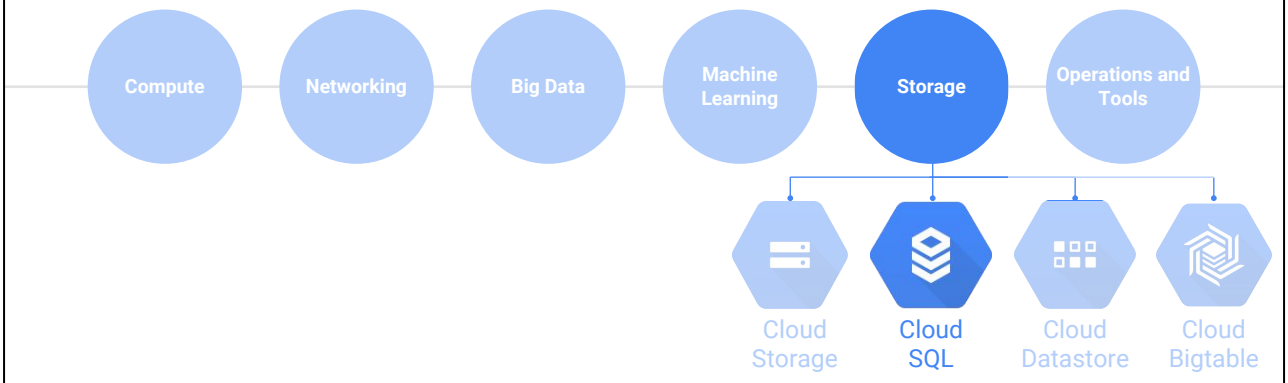
service. Because HBase and Cloud Bigtable are not identical, there are minor differences between the standard HBase API and the HBase API for Cloud Bigtable. See Differences between the HBase and Cloud Bigtable APIs for details.

For more information on the Cloud Bigtable HBase client, see: https://cloud.google.com/bigtable/docs/bigtable-and-hbase.

Agenda

1. Google Cloud Storage
2. Google Cloud Bigtable
3. Google Cloud SQL
4. Comparing Storage Options
5. Quiz & Lab

Google Cloud Platform



Notes:
The next storage option on Google Cloud Platform is Google Cloud SQL.

Google Cloud SQL (1 of 2)

- Google-managed MySQL relational database in the cloud
- Pay-per-use model
- REST API for management
- Affordability and performance

Notes:
Google Cloud SQL is an easy-to-use service that delivers fully managed MySQL databases. It lets you hand off to Google the mundane, but necessary and often time consuming tasks — like applying patches and updates, managing backups and configuring replications — so you can put your focus on building great applications. And because we use standard MySQL, it's easy to connect from just about any application, anywhere. It has all the capabilities and functionality of MySQL, with a few additional features and a few unsupported features as listed below. Google Cloud SQL is easy to use-- it doesn't require any software installation or maintenance.

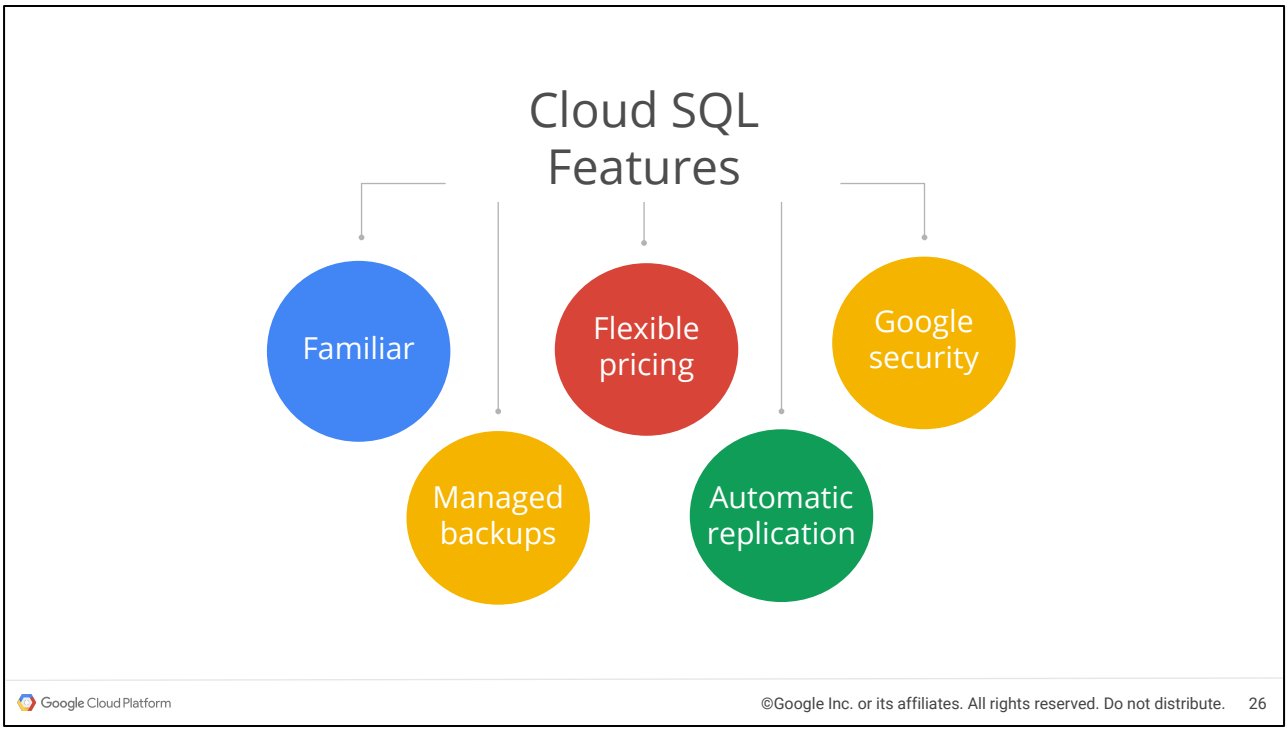
You can connect to a Google Cloud SQL instance from:

- MySQL Client
- Third-party tools like SQL Workbench or Toad for MySQL
- External applications using standard MySQL database drivers
- Google App Engine applications
- Google Apps Script scripts

Google Cloud SQL (2 of 2)

- Google security
- Vertical scaling (read and write)
- Horizontal scaling (read)
- Seamless integration with App Engine and Compute Engine

Notes:
Cloud SQL customer's data is encrypted when on Google's internal networks and when stored in database tables, temporary files, and backups. Every Cloud SQL instance includes a network firewall, allowing you to control network access to your database instance by granting access. Cloud SQL instances are accessible from just about any application, anywhere. Easily connect from App Engine, Compute Engine, and your workstation. Easily scale up to 16 processor cores and more than 100GB of RAM. Quickly scale out with read replicas.



Notes:
Familiar
Build and deploy for the cloud faster because Cloud SQL offers standard MySQL databases. Use standard connection drivers and built-in migration tools to get started quickly. Cloud SQL supports most MySQL statements and functions:

- Stored procedures
- Triggers
- Views

Cloud SQL does not support:

- User-defined functions
- MySQL-esque replication
- Statements and functions related to files and plugins

Flexible pricing
For First Generation instances, Google offers two billing plans: Packages and Per Use. Which plan you should select depends on how your database will be used, but generally it is more economical to use the Package plan if your instance is used for more than 450 hours each month. If your instance is used sporadically, you can reduce costs by using the Per-use plan and setting your activation policy to ON DEMAND. Learn more.

Cloud SQL second generation is in beta as of this writing. Pricing information during beta is as follows:

- Instance pricing
- Storage pricing
- Network pricing

Google security
Cloud SQL customer's data is encrypted when on Google's internal networks and when stored in database tables, temporary files, and backups. Every Cloud SQL instance includes a network firewall, allowing you to control network access to your database instance by granting access.

Managed backups
Google Cloud SQL takes care of securely storing your backed-up data and makes it easy for you to restore from a backup and perform a point-in-time recovery to specific state of an instance. Google Cloud SQL retains up to 7 backups for each instance, which is included in the cost of your instance.

Keep the following in mind when planning how you will use your instance and your restore needs:

- To perform a restore of your instance from a backup, you must have automated backups enabled.
- To perform a point-in-time recovery of your instance, you must have automated backups and binary logging enabled for the instance.
- Restoring from a backup updates the instance to its state at the time when the backup was taken.

You can enable or disable automated backups and binary logging for an instance at any time.

Automatic replication
Google Cloud SQL supports the following read replica scenarios:

- Cloud SQL instances replicating from a Cloud SQL master instance. Replicas are other instances in the same project and location as the master instance. This feature is in Beta.
- Cloud SQL instances replicating from an external master instance. The master instance is external to Google Cloud SQL. For example, it can be outside the Google network or in a Google Compute Engine instance. This feature is in Beta.
- External MySQL instances replicating from a Cloud SQL master instance. External replicas are in hosting environments, outside of Cloud SQL.

Cloud SQL Integration

Cloud SQL can be used with App Engine using standard drivers like Connector/J for Java or MySQLdb for Python.

App Engine applications are authorized to access Cloud SQL, and the instance can be configured to follow one 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 - to stay close to the Compute Engine infrastructure.

Cloud SQL can be used with external applications and clients by authorizing IP addresses or networks using CIDR notation.

Standard tools like MySQL Workbench can be used to administer databases. External read replicas can be configured.

Cloud SQL Second Generation

- Same features as first generation with higher performance and storage capacity at lower cost
  - Up to 7X throughput and 20X storage capacity of first generation instances
  - Less expensive than first generation for most use cases

Notes:
Google Cloud SQL is a fully-managed database service that makes it easy to set-up, maintain, manage and administer your relational MySQL databases in the cloud. Our Cloud SQL Second Generation offers improved performance, scalability, and convenience. Hosted on Google Cloud Platform, Cloud SQL provides a database infrastructure for applications running anywhere.

Differences between Second Generation and First Generation instances
If you have been using First Generation instances, the following list of differences between First Generation instances and Second Generation instances can help you plan for the differences between the two instance types.

Pricing differences

- Cloud SQL Second Generation does not provide a per-use pricing package; instance pricing is determined by the machine type. With the introduction of per-minute billing and sustained use discounts, Cloud SQL Second Generation can be more cost effective for many workloads.

On Demand activation policy

- Second Generation instances do not support the On Demand activation policy. Only Always On and Off are supported.

IPv6 connectivity

- IPv6 connectivity is not supported for Second Generation instances. IPv4 connectivity is fully supported.

Storage engine support

- For Second Generation instances, InnoDB is the only supported storage engine.

GTID Replication

- Second Generation instances use GTID replication. This means that certain SQL statements and operations are not allowed. For more information, see Differences between Cloud SQL and standard MySQL functionality.

MySQL user and database administration

- For Second Generation instances, you cannot create and administer MySQL user accounts and databases from the Google Cloud Platform Console, the CLoud SDK, or the Cloud SQL API. You use the MySQL Client administration tool to manage users and databases.

MySQL versions

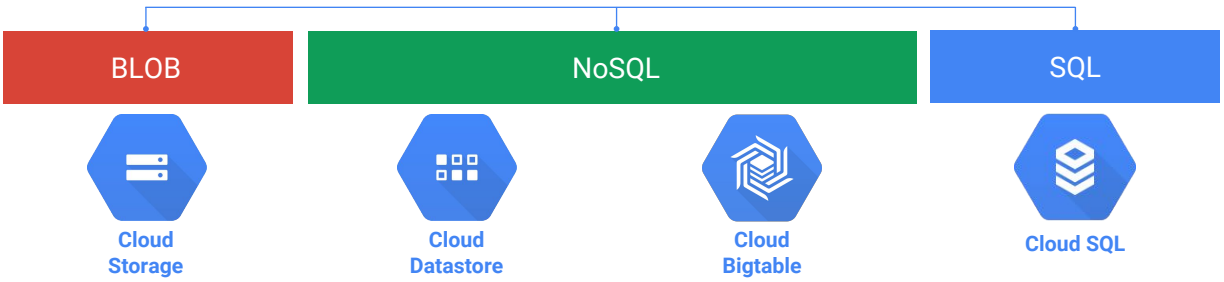
- Second Generation instances support MySQL 5.7 (the default) and MySQL 5.6.

For more information on Cloud SQL second generation, see: https://cloud.google.com/sql/docs/1st-2nd-gen-differences.

Agenda

1. Google Cloud Storage
2. Google Cloud Bigtable
3. Google Cloud SQL
4. Comparing Storage Options
5. Quiz & Lab

Comparing Storage Options (1 of 2)



Good for: Structured and unstructured binary or object data	**Good for:** Getting started, App Engine applications	**Good for:** "Flat" data, Heavy read/write, events, analytical data	**Good for:** Web frameworks, existing applications
Use cases: Images, large media files, backups	**Use cases:** User profiles, product catalog	**Use cases:** AdTech, Financial and IoT data	**Use cases:** User credentials, customer orders

Notes:
Google Cloud Platform delivers various storage service offerings which remove much of the burden of building and managing storage and infrastructure. Like our other cloud services, cloud storage will free you to focus on doing what you do best and differentiating at the application or service layer.

Our storage offerings range from SQL, NoSQL, Blob and Block storage depending on what you are trying to do, and it's easy to mix and match.

- Cloud SQL gives you fully managed MySQL so you have relational DB and a more traditional approach to queries.
- Cloud Datastore provides a nearly infinitely scalable, schemaless solution.
- If you want a disk you can mount Persistent Disk as a block store that can be used by Compute Engine.
- For just pure data and blobs, Cloud Storage can deliver what you need.
- Google Cloud Bigtable offers companies a fast, fully managed, infinitely scalable NoSQL database service ideal for web, mobile and IoT applications.

Comparing Storage Options (2 of 2)

	Cloud Datastore	Cloud Storage	Cloud SQL (1st and 2nd Generation)	Bigtable
Storage type	NoSQL, document	Object (BLOB) store	Relational SQL	NoSQL, wide-column
Overall capacity	Terabytes +	Petabytes +	up to 500 GB	Petabytes +
Unit size	1 megabyte / entity	5 TB / object	Standard MySQL limits	Recommended - Individual values: ~10 MB per cell All values per row: ~100 MB
Transactions	Yes	No	Yes	No
Complex queries	No	No	Yes	No

Notes:

- Cloud Bigtable is not a relational database; it does not support SQL queries or joins, nor does it support multi-row transactions. Also, it is not a good solution for small amounts of data (< 1 TB).
- If you need full SQL support for an online transaction processing (OLTP) system, consider Google Cloud SQL.
- If you need interactive querying in an online analytical processing (OLAP) system, consider Google BigQuery.
- If you need to store immutable blobs larger than 10 MB, such as large images or movies, consider Google Cloud Storage.
- If you need to store highly structured objects, or if you require support for ACID transactions and SQL-like queries, consider Cloud Datastore.

Agenda

1. Google Cloud Storage
2. Google Cloud Bigtable
3. Google Cloud SQL
4. Comparing Storage Options
5. Quiz & Lab

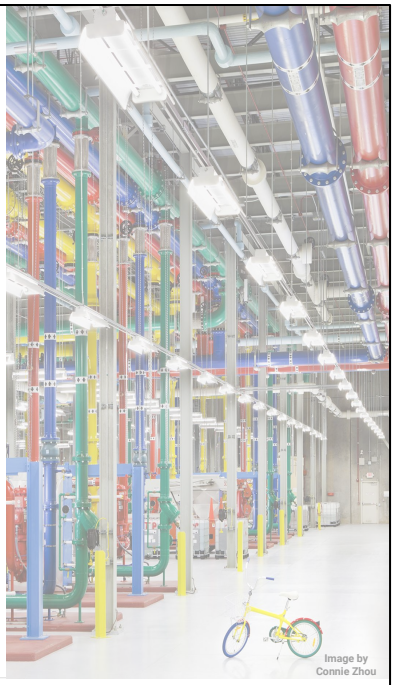
Quiz

1. You are developing an application that transcodes large video files. Which Google Cloud Platform storage option is the best choice for your application?
2. You manufacture devices with sensors and need to stream huge amounts of data from these devices to a storage option in the cloud. Which Google Cloud Platform storage option is the best choice for your application?

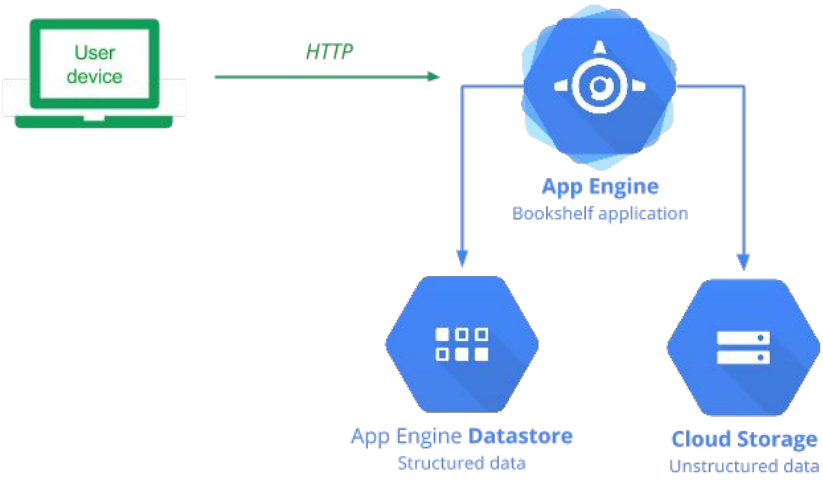
Lab (1 of 2)

Deploy the Bookshelf application to App Engine using Cloud Storage to store application data.

1. Review the application code and create a Cloud Storage bucket
2. Deploy the Bookshelf application to App Engine using Cloud Shell
3. Test the application in your browser and view the objects in Cloud Storage



Lab (2 of 2)



Notes:
Bookshelf is a simple Python web application that lets users create and manage a list of books. The application uses the Flask web microframework to coordinate reading and writing to storage. This version of Bookshelf uses Cloud Datastore to persistently store book data and uses Cloud Storage to store images (book covers).

Resources

- Overview: Cloud Storage
  https://cloud.google.com/storage/
- DevBytes - File storage in the cloud
  https://www.youtube.com/watch?v=vyIap827rHs
- Cloud SQL: Features, case studies, pricing, & documentation
  https://cloud.google.com/sql/
- Getting started with Google Cloud SQL
  https://cloud.google.com/sql/docs/quickstart
- Overview of Cloud Bigtable
  https://cloud.google.com/bigtable/docs/overview

Quiz Answers

1. You are developing an application that transcodes large video files. Which storage option is the best choice for your application?

   Answer: Google Cloud Storage

2. You manufacture devices with sensors and need to stream huge amounts of data from these devices to a storage option in the cloud. Which Google Cloud Platform storage option is the best choice for your application?

   Answer: Google Cloud Bigtable


cloud.google.com