Google Cloud Platform Storage Options

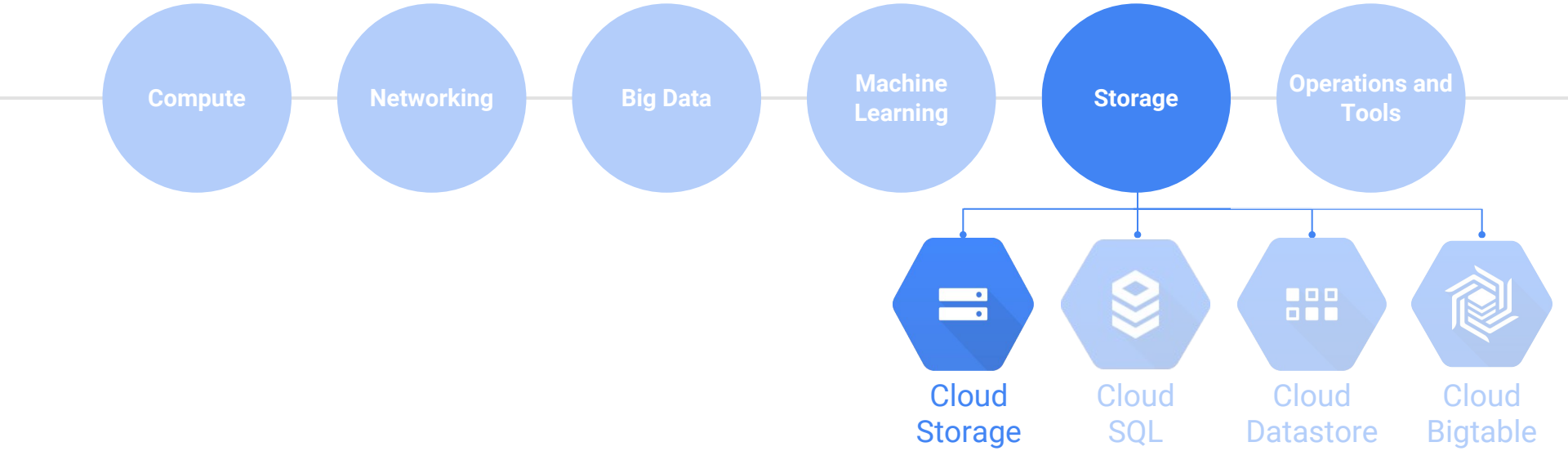
Google Cloud Platform Fundamentals

V2.0

Agenda

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

Google Cloud Platform



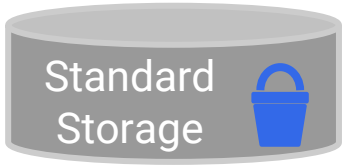
Google Cloud Storage (1 of 2)

- 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)

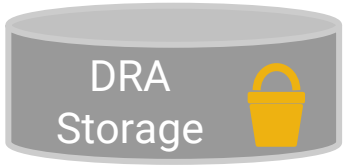
Google Cloud Storage (2 of 2)

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

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

Autism Speaks

"Thanks to **Google Cloud Platform** and the Google Genomics team, the greatest minds in science from around the world will be able to study **trillions of data points in one single database.**"

UPLOADED **100** TERABYTES of data from more than **1,300 WHOLE GENOMES** to Google Cloud Storage

UP TO **200** GIGABYTES **PER RAW GENOME**

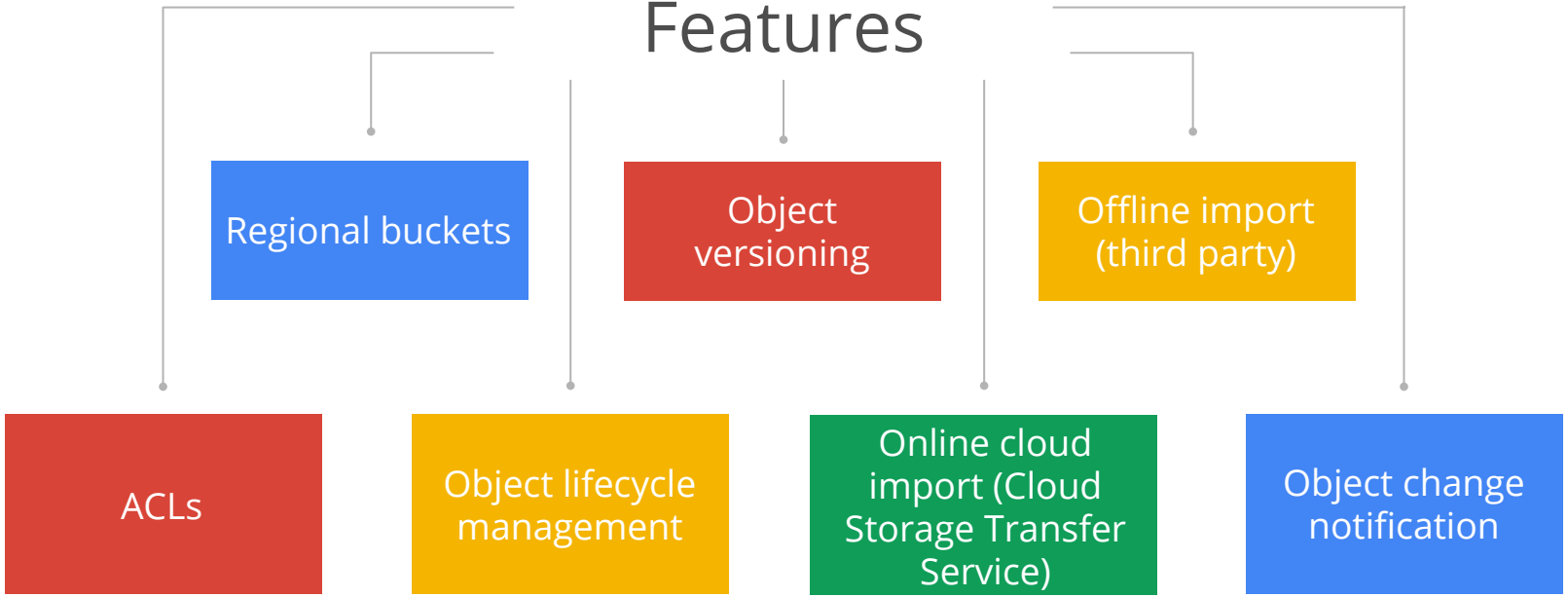
MSSNG project could easily surpass a **PETABYTE** OF DATA

WHOLE GENOMES FROM 10,000 PEOPLE

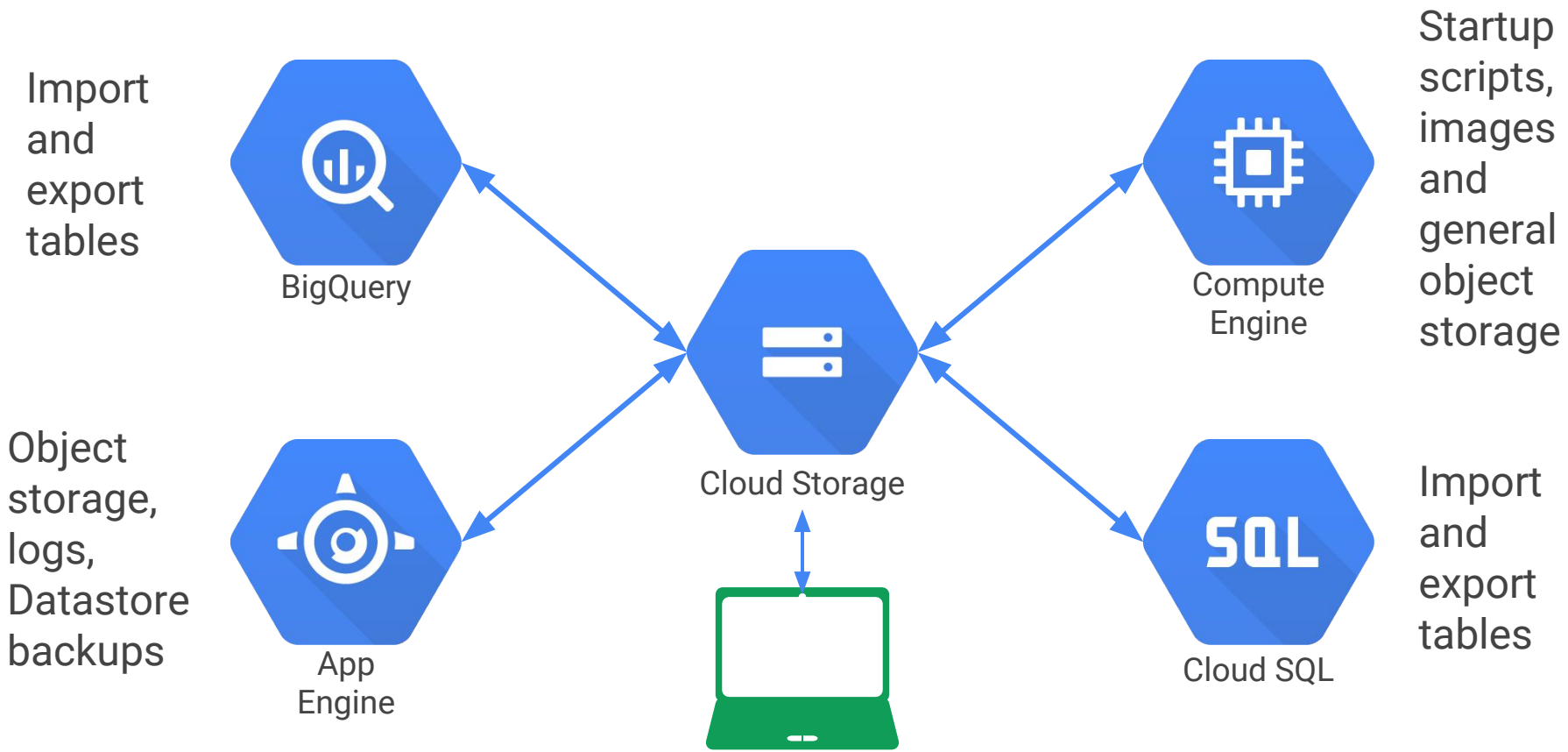
Making MSSNG world's largest single repository of autism-related DNA sequencing data

Image by Connie Zhou

Cloud Storage Features



Cloud Storage Integration



Agenda

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

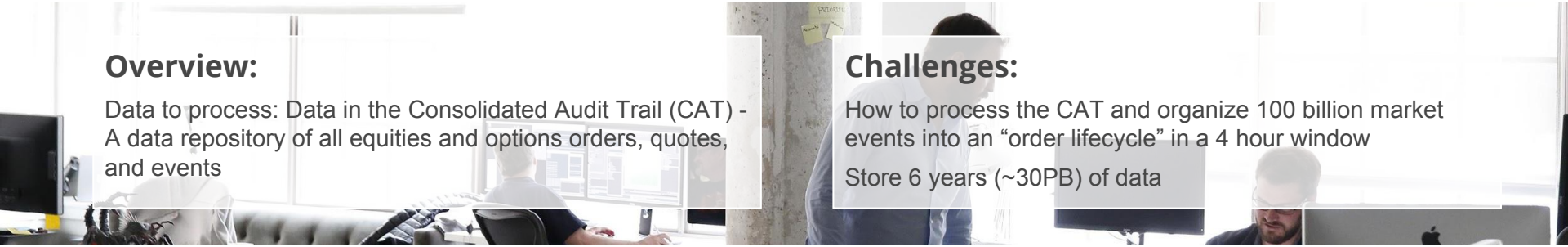
Google Cloud Bigtable (1 of 2)

- 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

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

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

6 BILLION
MARKET EVENTS
WRITTEN **PER HOUR**

1.7 GIGs
PER SECOND

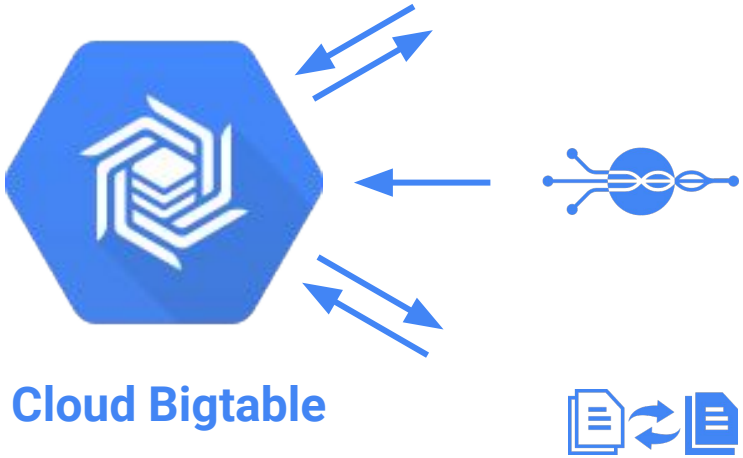
6 TBs
PER HOUR

10 BN
WRITTEN
PER HOUR BURSTS

1.7 GIGABYTES
PER SECOND

10 TERABYTES
PER HOUR

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.

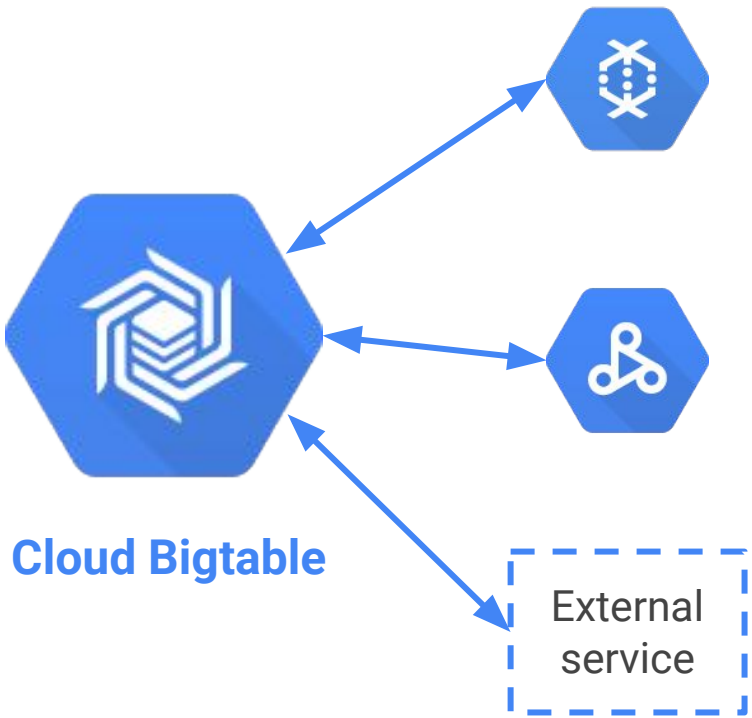
Streaming

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

Agenda

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

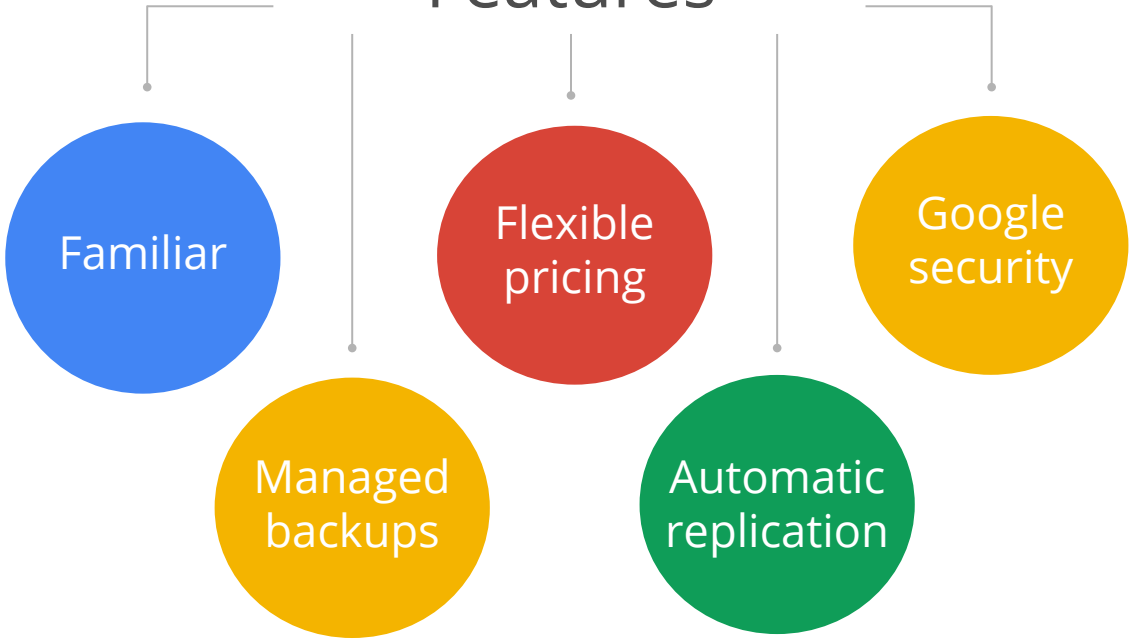
Google Cloud SQL (1 of 2)

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

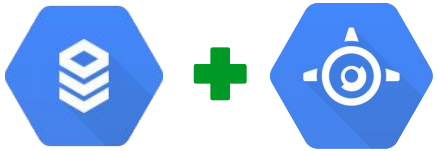
Google Cloud SQL (2 of 2)

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

Cloud SQL Features

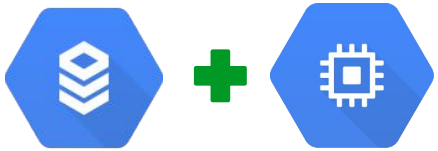


Cloud SQL Integration



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.

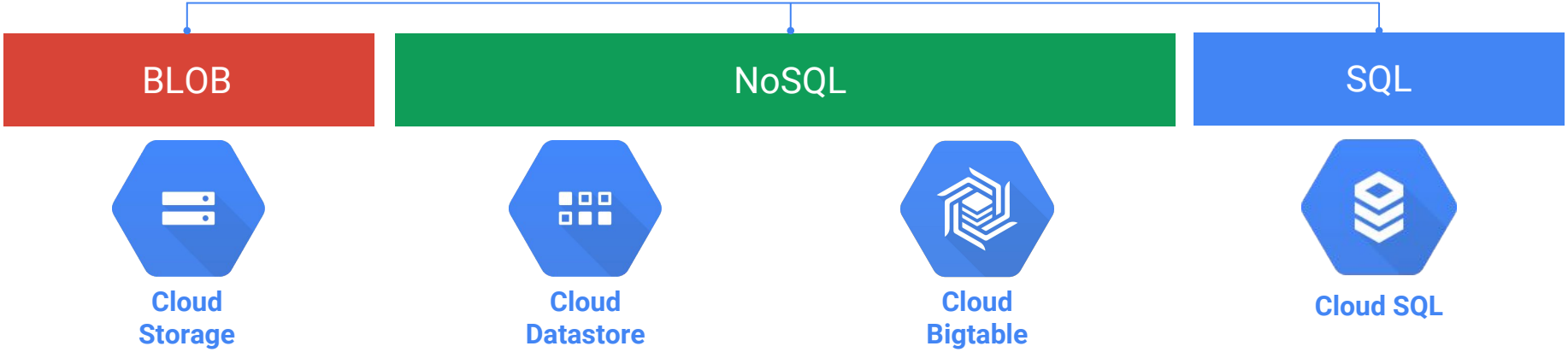
Cloud SQL can be used with external applications and clients by authorizing IP addresses or networks using CIDR notation.

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

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

Comparing Storage Options (2 of 2)

	Cloud Datastore	Cloud Storage	Cloud SQL (1st 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

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?

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

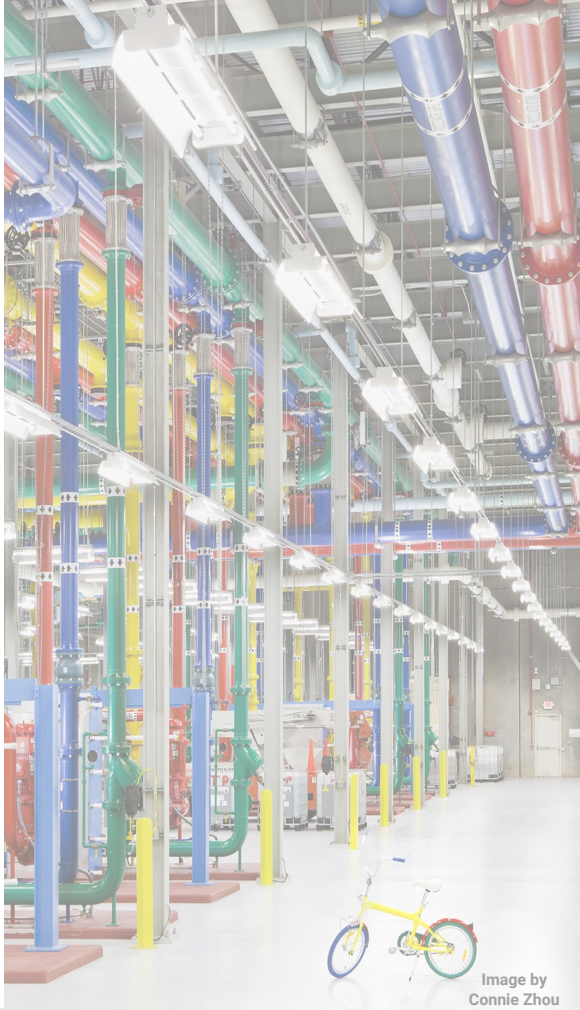
2. You manufacture devices with sensors and need to stream huge amounts of data from these devices to a storage option in the cloud. Which Google Cloud Platform storage option is the best choice for your application?

   Answer: Google Cloud Bigtable

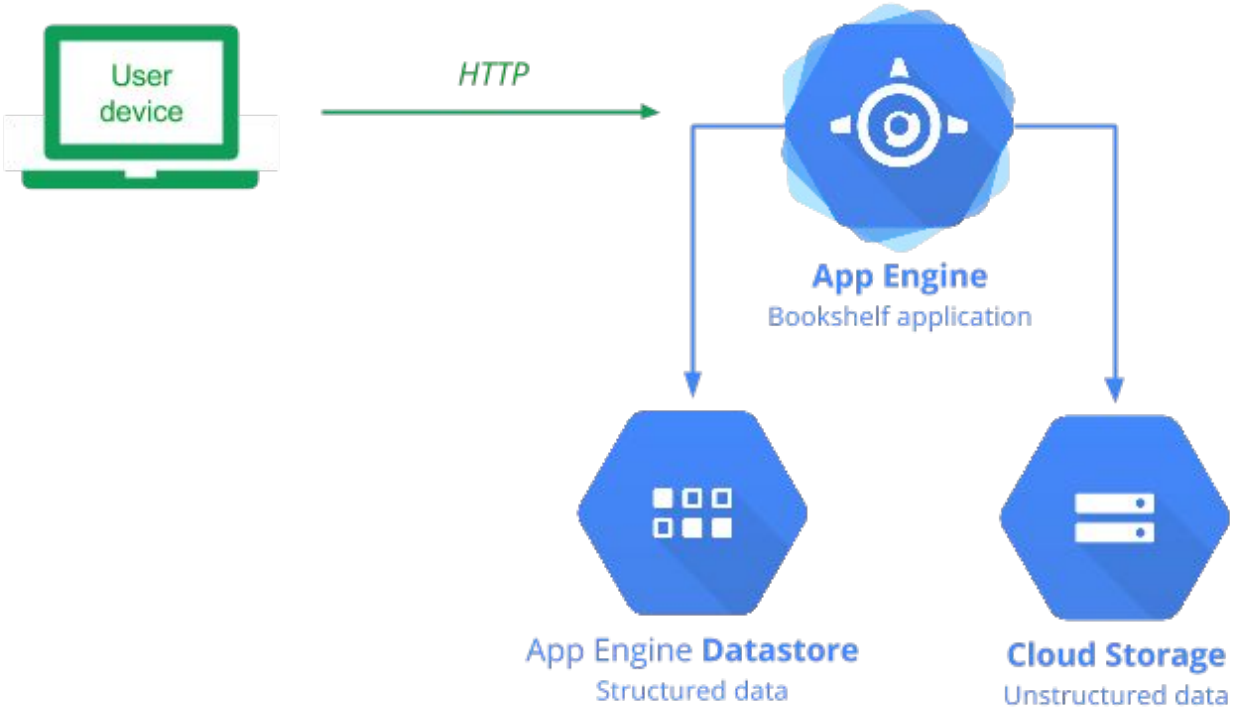
Lab (1 of 2)

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)



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/getting-started
- Overview of Cloud Bigtable
  https://cloud.google.com/bigtable/docs/api-overview


cloud.google.com