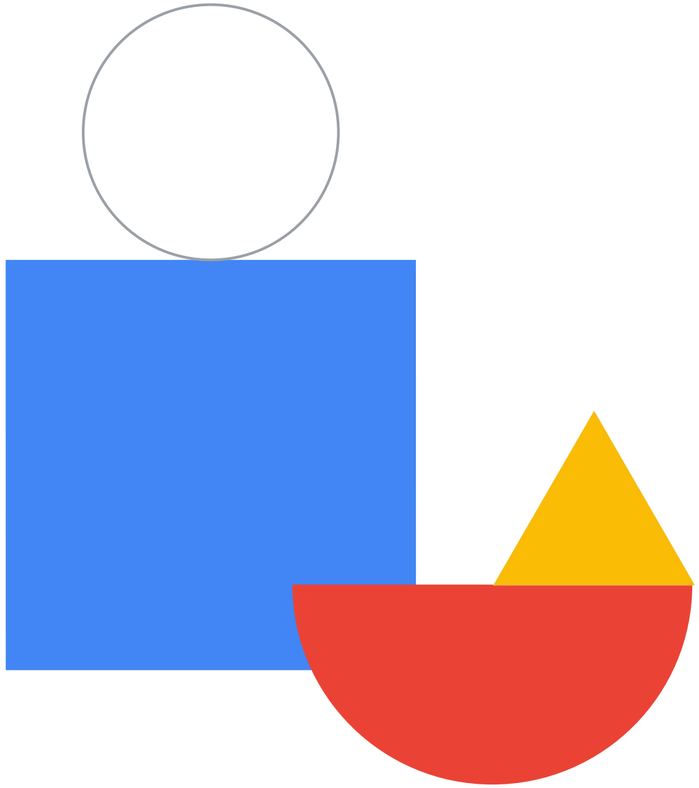
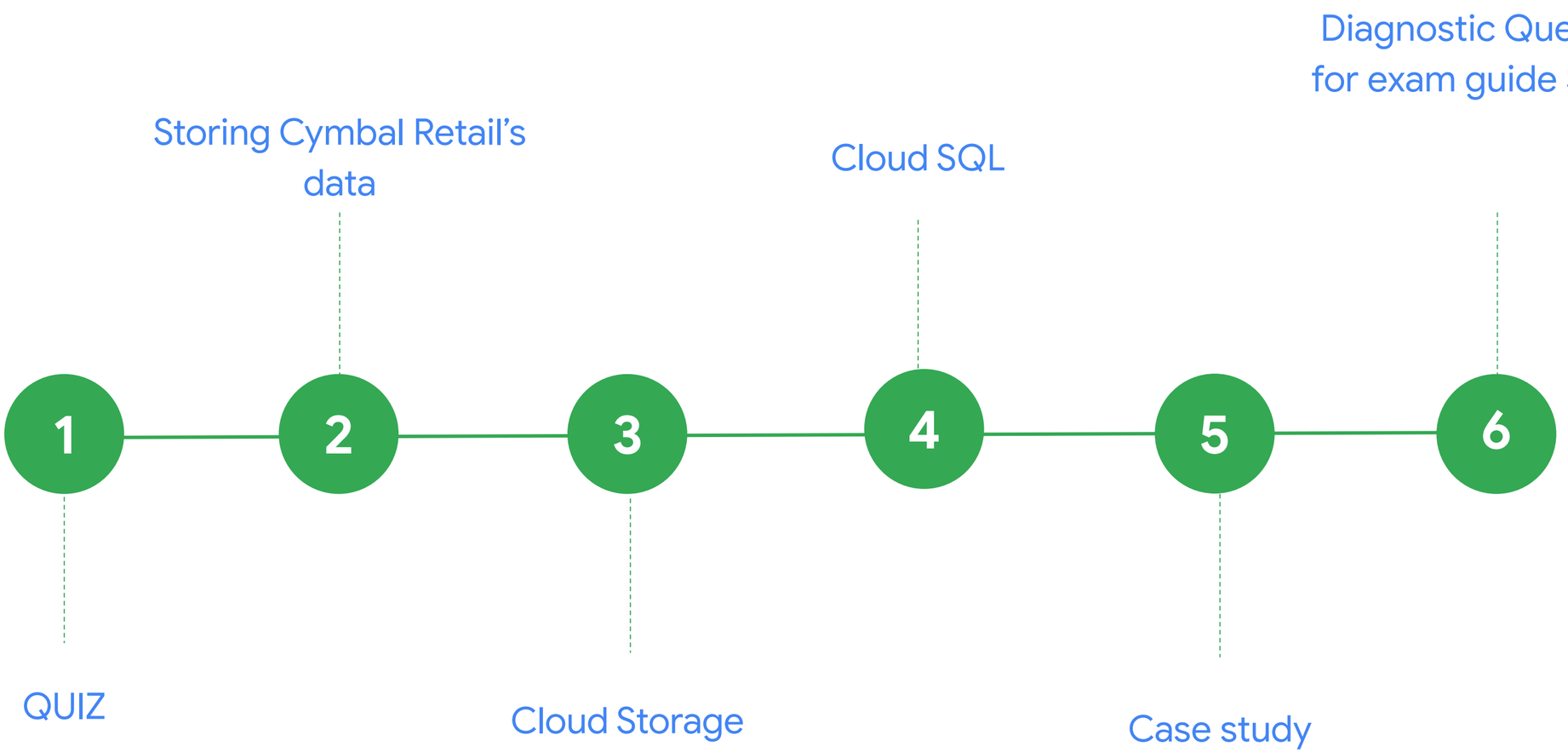
Google Cloud



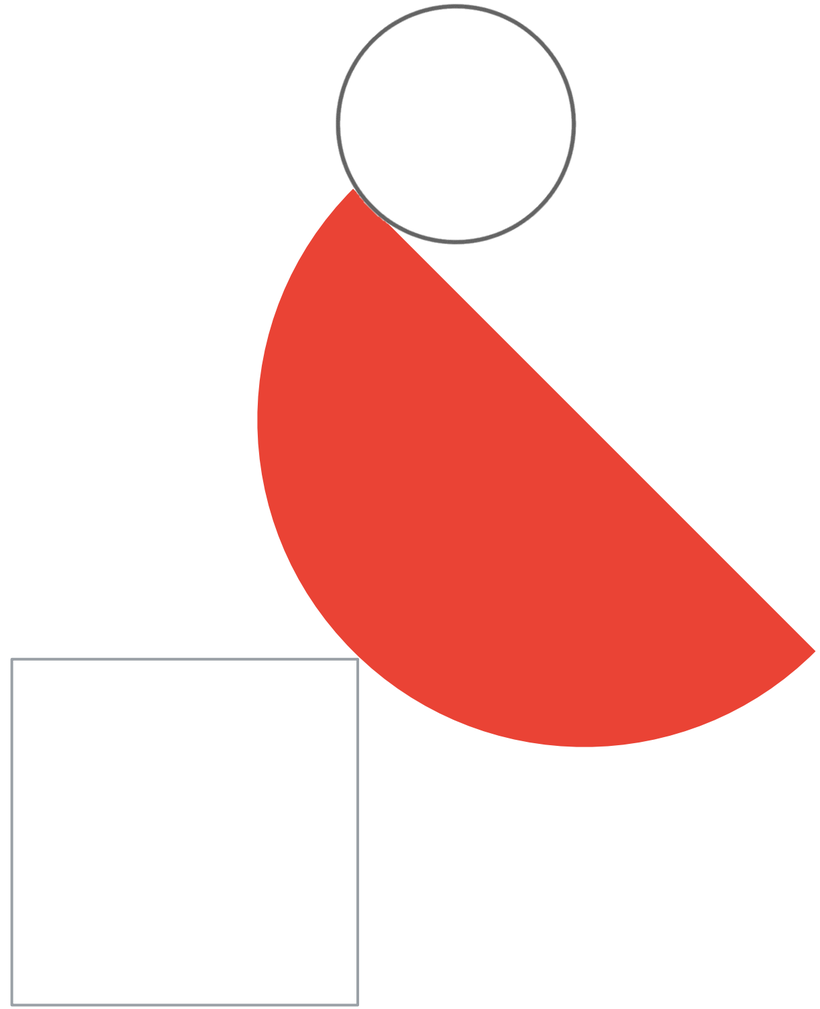
Preparing for Your Professional Data Engineer Journey

Module 3: Storing the Data

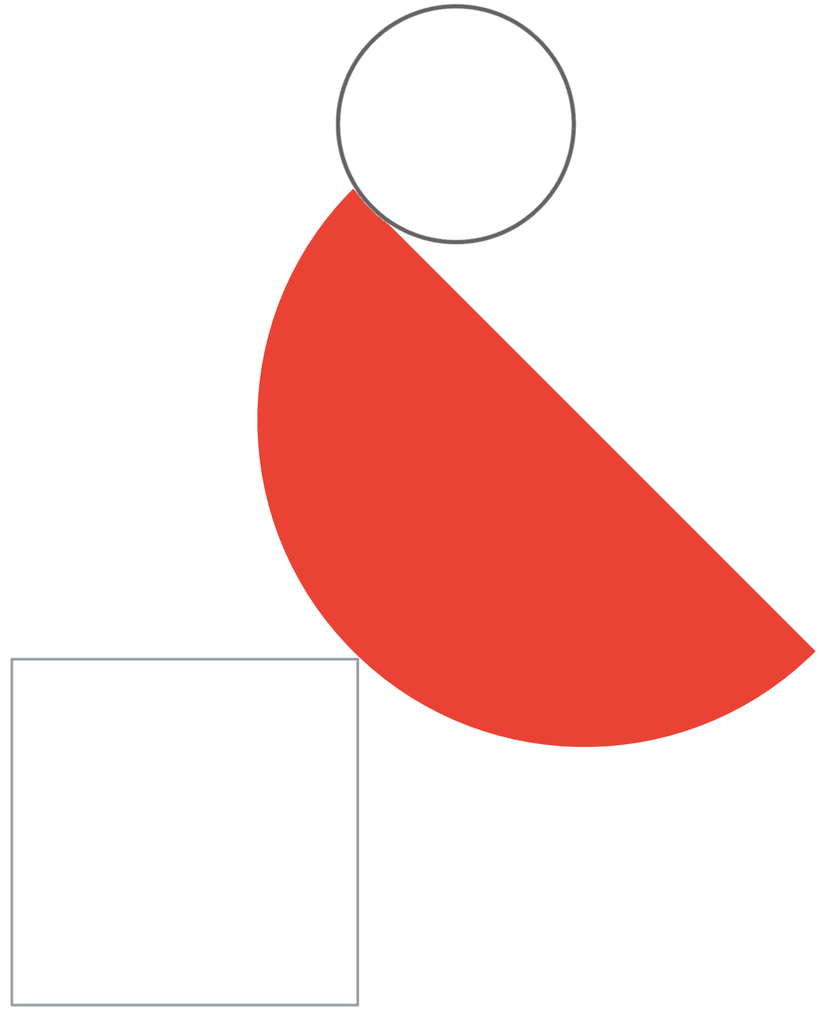
Week 4 agenda

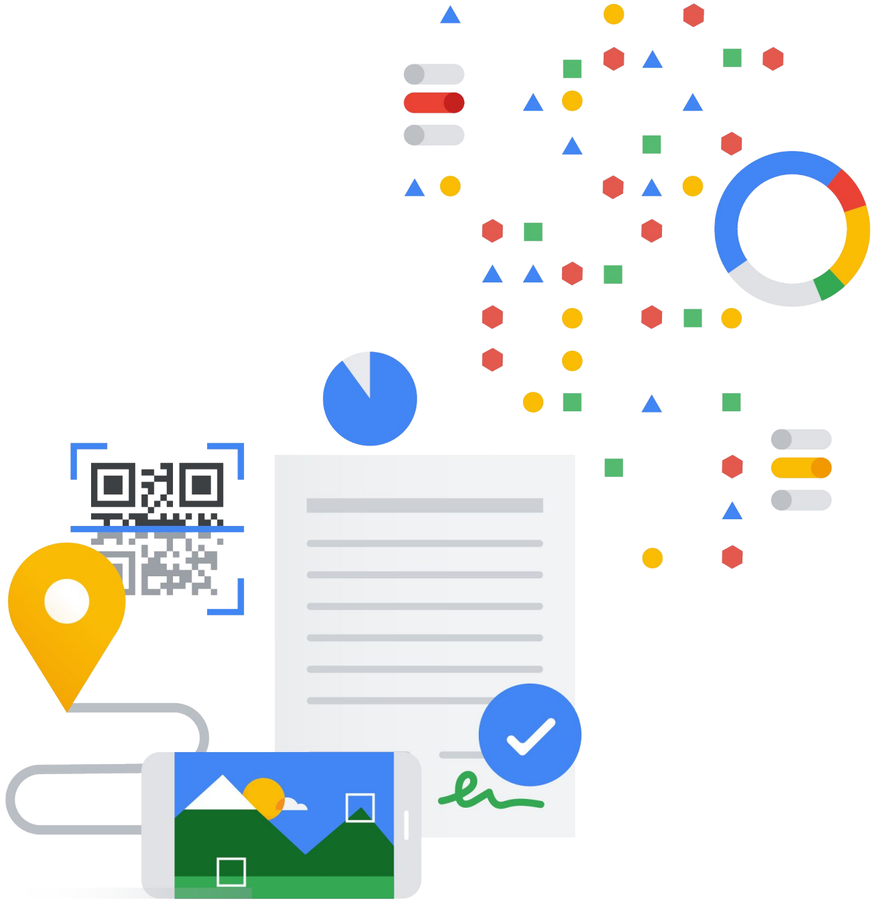


QUIZ time!



Storing Cymbal Retail's data



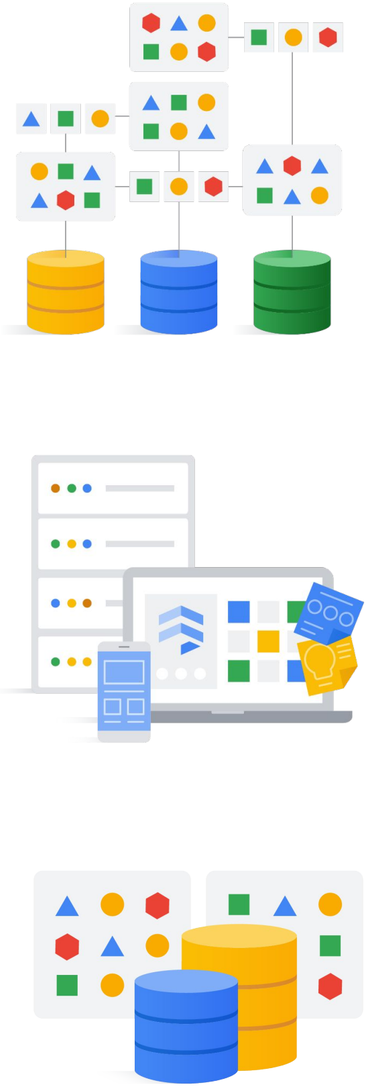


How does Cymbal Retail use data?

Cymbal Retail ingests, stores, processes, and analyzes many types of data—documents, images, text, and video.

Choosing appropriate storage solutions on Google Cloud

You need to ensure the right solution to help ensure efficiency of access and cost.

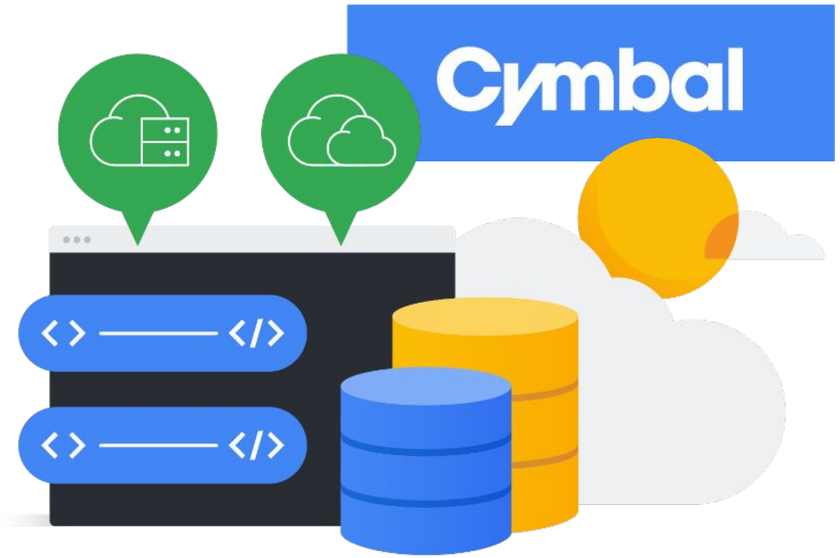


Ensuring the data lifecycle complies with data privacy rules

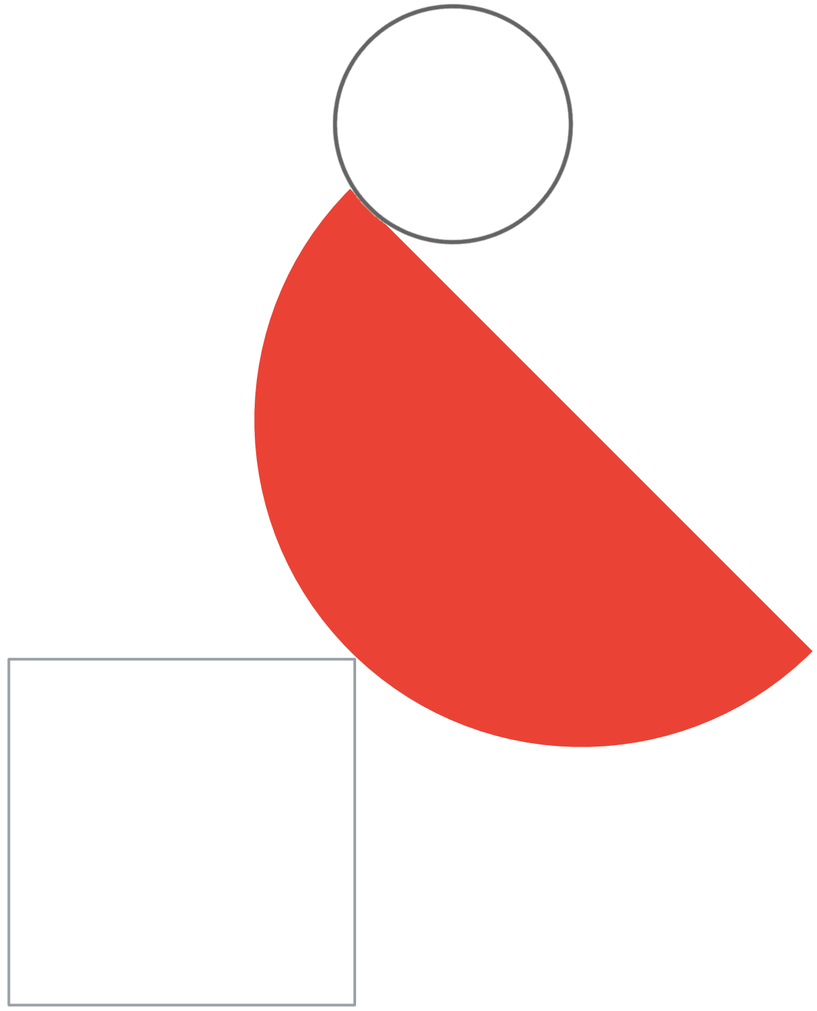
- Cymbal Retail needs to comply with industry and regional data privacy laws.
- Data must follow a well-defined lifecycle.
- You need to set controls to prevent accidental breaches.

Evolving data architecture as business needs change

- What are the goals of the business?
- How do data-driven decisions map to business goals?



Google Cloud Storage (GCS)



Know these GCS features well!

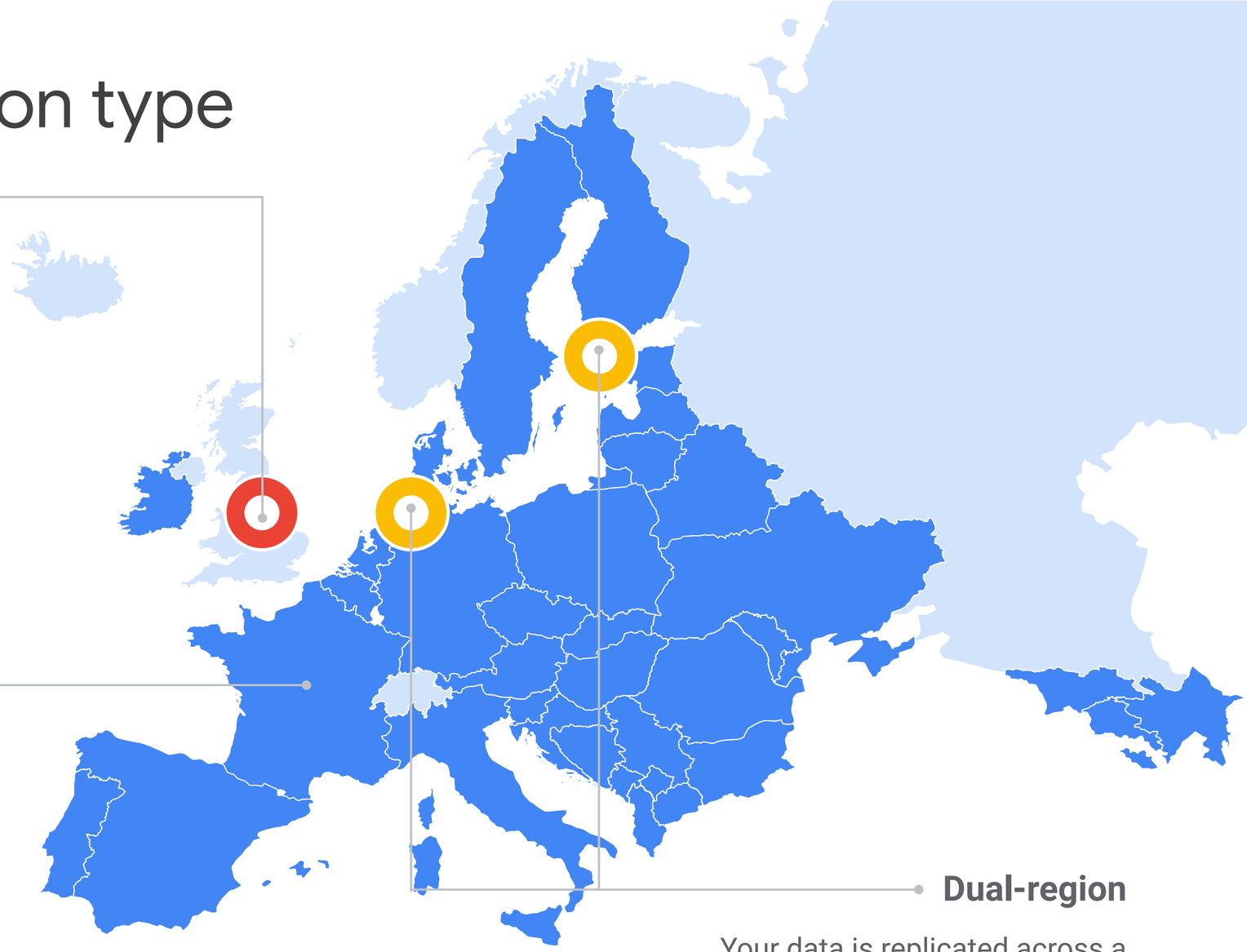
Object versioning and retention policies cannot be on at the same time. If you want to allow objects to be modified, choose object versioning. If you want to prevent deletions or changes to objects, set a retention policy.

- Controlling object lifecycle:
  - Retention policy (best for compliance)
  - Object Hold (prevent individual objects from being deleted)
  - Object Versioning (aka “automatic backups with retention policy; be aware of additional costs)
  - Object Lifecycle Management (aka “object TTL” / downgrade class to optimize costs)
- ACLs (read, write, full control on buckets or an object).
- Objects are immutable.
- Location constraints on buckets.
- Object change notifications (useful for automation, along with Pub/Sub).
- Resumable uploads (can restart from the last successful chunk).
- Strong consistency except for cached objects.
- Storage class set at object level (fine-grained performance/cost control without moving data to different buckets).
- Cloud Storage Triggers to handle events in Cloud Functions.
- Streaming uploads to GCS.
- For data encryption, GCS supports GMEK, CMEK and CSEK (most services do not support CSEK!)

GCS: Choosing a location type

Regional

Your data is stored in a specific region with replication across availability zones in that region. Good for colocating compute and storage for high performance (eg. data analytics).



Multi-Region

Your data is distributed redundantly across US, EU, or Asia. Good for serving content to end users and when you want automatic failover.

Dual-region

Your data is replicated across a specific pair of regions. Good for when you need colocated compute and storage and automatic DR.

Exam Tip:

- *Know the use-cases for each of those choices.*
- *Dual-region buckets are only available for selected regions!*

Google Cloud Storage: where to use each storage class

Standard		Nearline	Coldline	Archive
In multi-region locations for serving content globally.	**In regional locations for data accessed frequently or high throughput needs**	**For data access less than once a month**	**For data accessed roughly less than once a quarter**	**For long term retention**
Streaming videos	Video transcoding	Serving rarely accessed docs	Serve rarely used data	Regulatory archives
Images	Genomics	Backup	Movie archive	Tape replacement
Websites	General data analytics & compute		Disaster recovery	
Documents				

Google Cloud Storage: autoclass

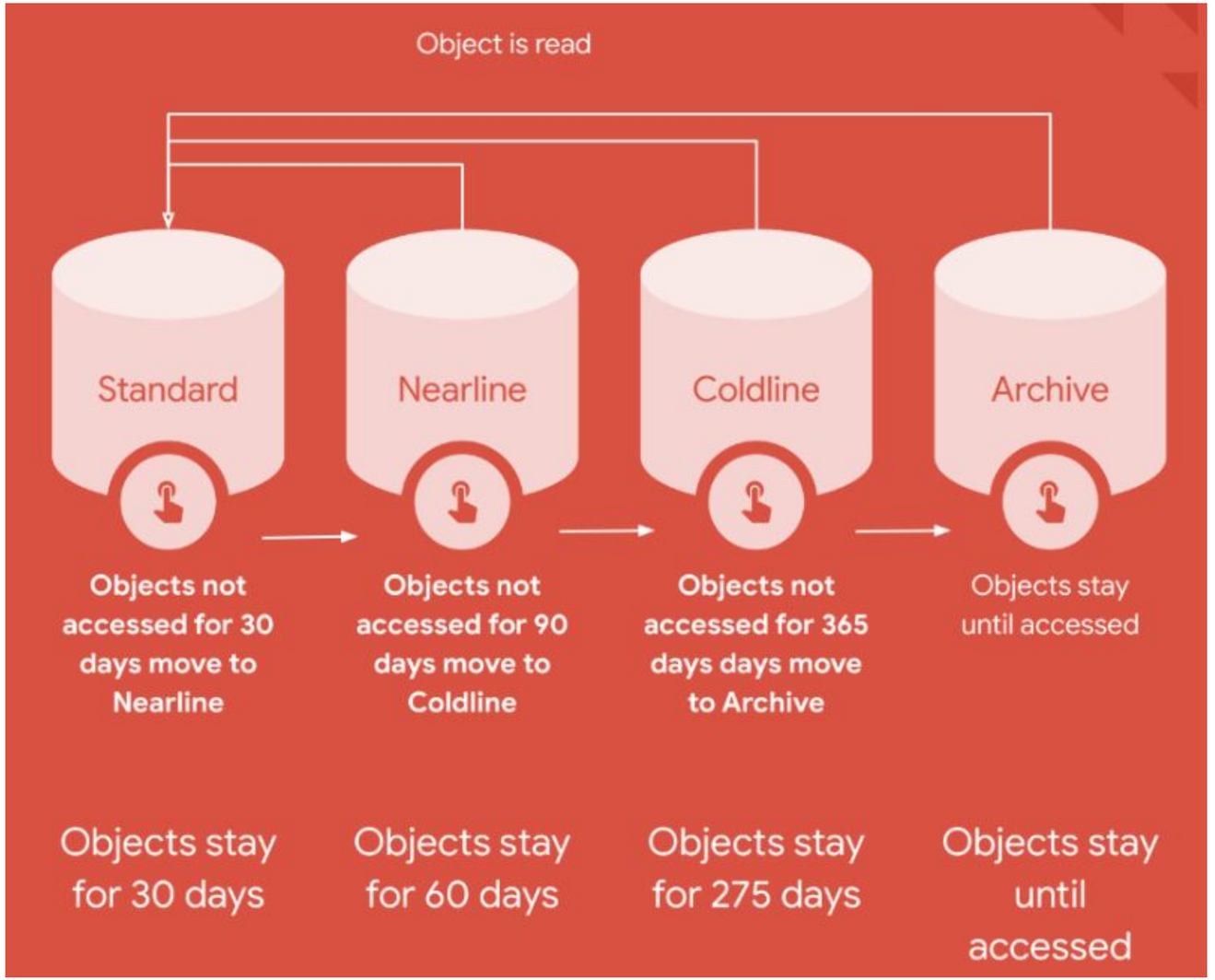
Automatically transition objects to colder storage classes based on usage patterns and transition back to Standard on access.

Not too “intelligent” as of now.

- **Choose a storage class for your data**

  A storage class sets costs for storage, retrieval, and operations, with minimal differences in uptime. Choose if you want objects to be managed automatically or specify a default storage class based on how long you plan to store your data and your workload or use case. Learn more

  - (●) **Autoclass**
    Automatically transitions each object to hotter or colder storage based on object-level activity, to optimize for cost and latency. Recommended if usage frequency may be unpredictable. Can be changed to a default class at any time. Pricing details
  - ( ) **Set a default class**
    Applies to all objects in your bucket unless you manually modify the class per object or set object lifecycle rules. Best when your usage is highly predictable. Can't be changed to Autoclass once the bucket is created.
    - ( ) Standard
      Best for short-term storage and frequently accessed data
    - ( ) Nearline
      Best for backups and data accessed less than once a month
    - ( ) Coldline
      Best for disaster recovery and data accessed less than once a quarter
    - ( ) Archive
      Best for long-term digital preservation of data accessed less than once a year



Exam Tip: *autoclass is a new feature (GA Q4 ‘22) and will not covered on the exam*

Best Practices on Storage Class Selection

Consider retention period and access frequency

Retention Period

Access Frequency	<1 mo	1–3 mo	3–12 mo	>12 mo
>12/yr	Standard	Standard	Standard	Standard
4–12/yr	Standard	Nearline	Nearline	Nearline
1–4/yr	Standard	Nearline	Coldline	Coldline
<1/yr	Standard	Nearline	Coldline	Archive

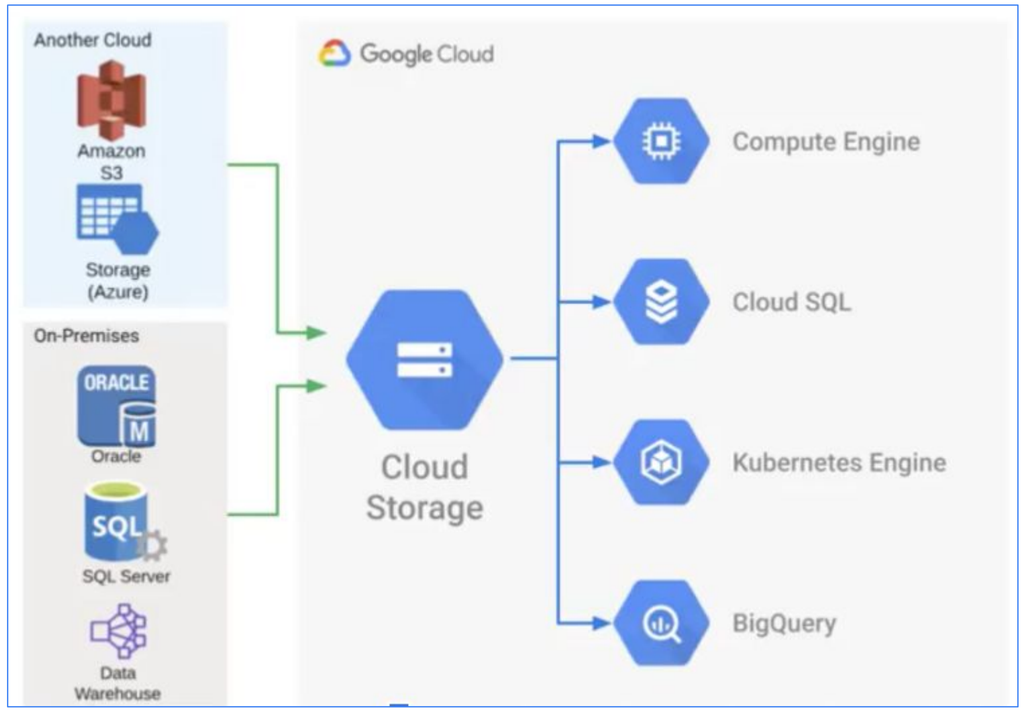
Choose the right tool to move data to GCS...

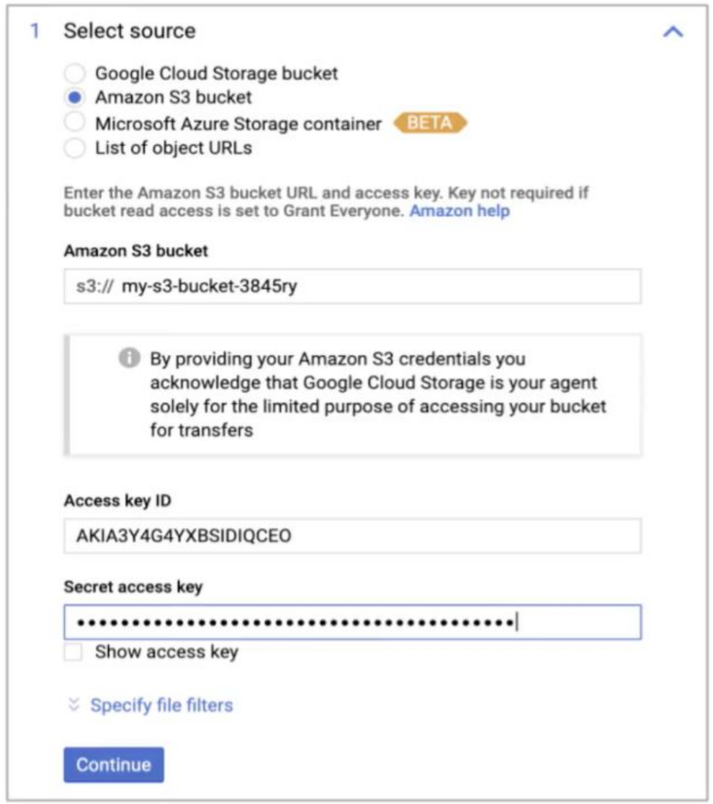
Where you're moving data from	Scenario	Suggested products
Another cloud provider (for example, Amazon Web Services or Microsoft Azure) to Google Cloud	—	Storage Transfer Service
Cloud Storage to Cloud Storage (two different buckets)	—	Storage Transfer Service
Your private data center to Google Cloud	Enough bandwidth to meet your project deadline for less than 1 TB of data	gsutil
Your private data center to Google Cloud	Enough bandwidth to meet your project deadline for more than 1 TB of data	Storage Transfer Service for on-premises data
Your private data center to Google Cloud	Not enough bandwidth to meet your project deadline	Transfer Appliance

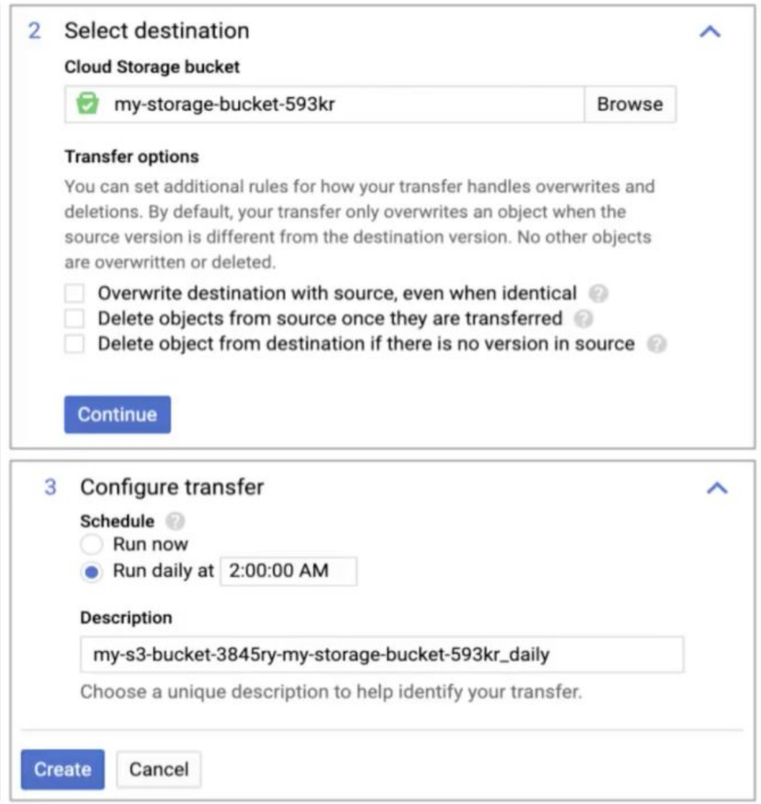
Exam Tips:

- *Depending on size and throughput, use gsutil / Transfer Service (low cost) / Transfer Appliance*
- ***When using Transfer Appliance, you need to execute so-called "rehydration" process*** *which will decrypt and uncompress before it's put to a destination bucket.*

Cloud Storage: Storage Transfer Service







Exam Tip: *Storage Transfer Service can be set up one-off (e.g., move bucket to new location) and recurring (e.g., back up from S3 to GCS with rsync-like semantics)*

Cloud Storage

#GCPSketchnote

@PVERGADIA THECLOUDGIRL.DEV

8.8.2020

What is Cloud Storage?

A GLOBAL, SECURE AND SCALABLE OBJECT STORE

GLOBAL SECURE SCALABLE

.png .jpg .mpg .mov .txt other

How does it WORK?

WORKFLOW

Image → Bucket (PHOTOS) → Project → Organization

4 STORAGE CLASSES

Based on Budget, Availability and Access Frequency

STANDARD	NEARLINE	COLDLINE	ARCHIVE
Frequent access High Availability	Once a month	Once a quarter	Once a year

New Version >30 Days >90 Days

OBJECT LIFECYCLE MANAGEMENT

AUTO VERSIONING → v1 → v2 → v3

Cloud Storage Use case example

STANDARD → Website + Streaming + Document

NEARLINE → Backups + Disaster Recovery

COLDLINE → Backups + Disaster Recovery

ARCHIVE → Long Term Archiving

STREAMING DATA, BATCH DATA → AI PLATFORM, BIGQUERY

SECURITY for Cloud Storage

- ☑ Encryption at rest
- ☑ Bring your own encryption key
  - CMEK - Customer Managed
  - CSEK - Customer Supplied

Cloud Storage PRICING

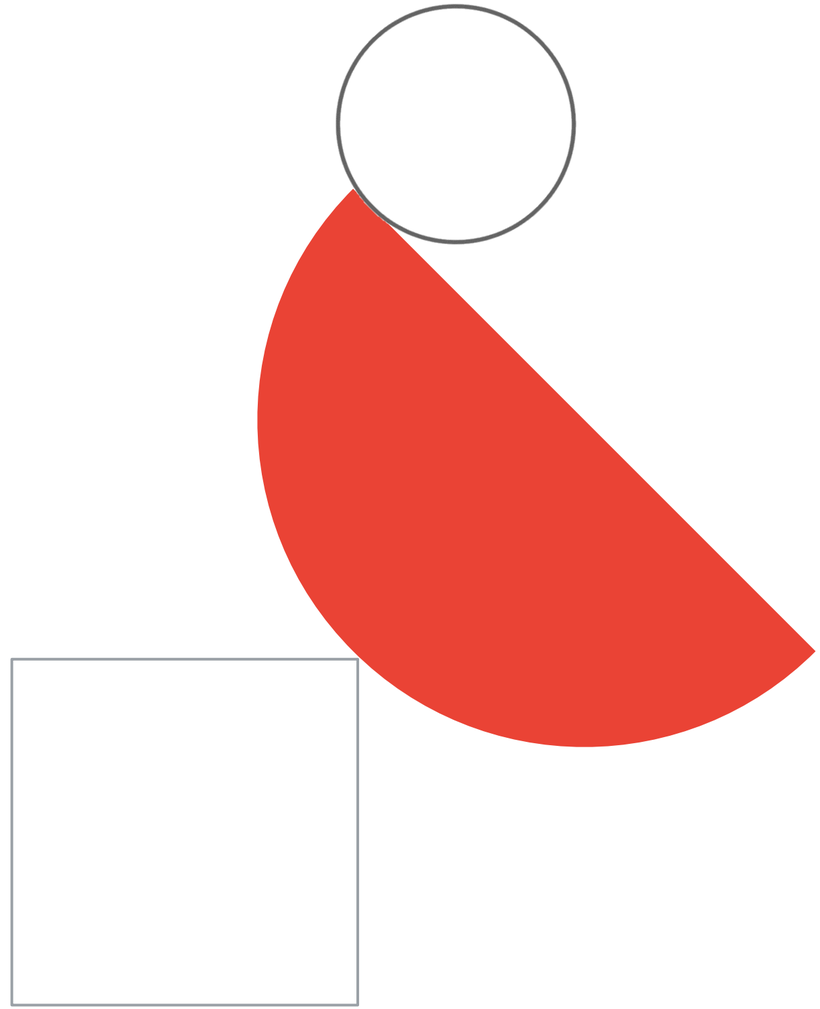
STANDARD	NEARLINE	COLDLINE	ARCHIVE
$$$$	$$$	$$	$
Automatic Redundancy Frequent Access			

How to USE Cloud Storage

ONLINE TRANSFER	TRANSFER SERVICE	TRANSFER APPLIANCE
gsutil, API, UI <gsutil cp image.png gs://my-bucket>	Transfer data from other clouds & on-premise	Hardware for >100tb data transfer

image.png API

Filestore

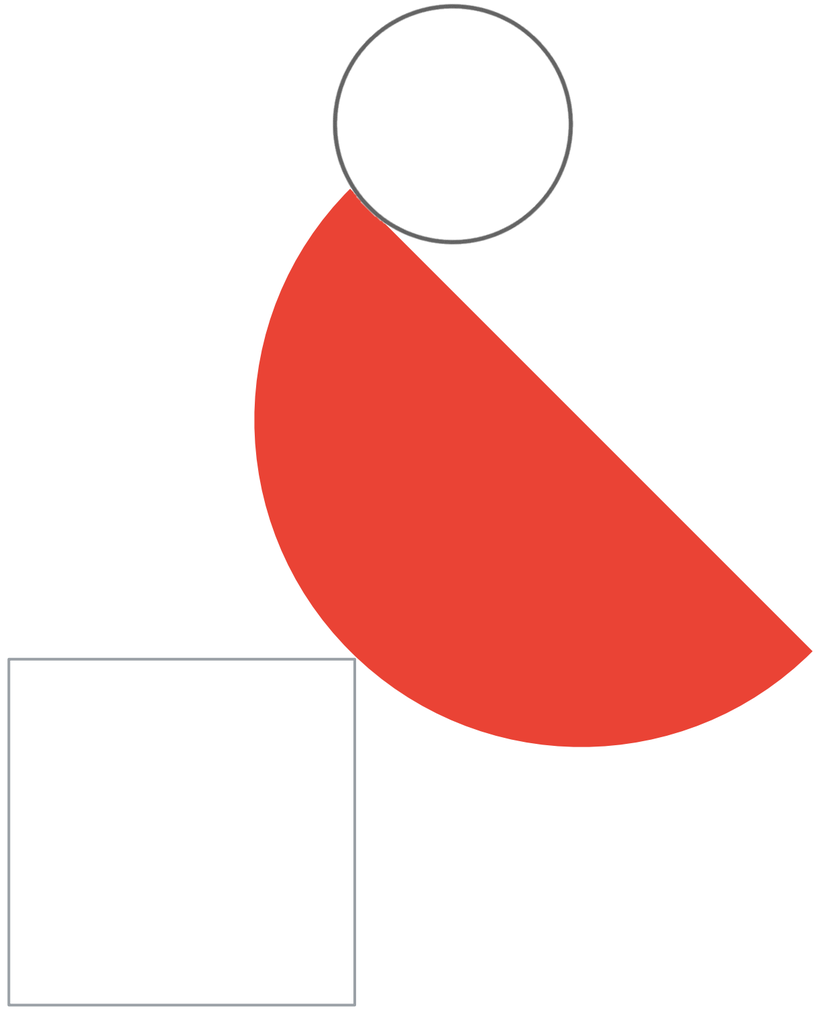


Filestore

Managed NFS, NOT a database

	Filestore Basic (GA)	Filestore High Scale (Public Preview)	Filestore Enterprise (GA)
Workloads	File sharing, Software Dev, and Web Hosting	HPC, Financial Modeling, Pharma, and Analytics	SAP, GKE, and ‘Lift & Shift Apps”
Capacity	1 - 64 TiB	10 - 100 TiB	1 - 10 TiB
Scale	Scale-up	Scale-out	Scale-out
Capacity Management	Grow	**Grow & Shrink**	**Grow & Shrink**
Max Performance (Throughput | IOPS)	1.2GiB/s | 60k	26GiB/s | 920k	1.2GiB/s | 120k
Data Protection	**Backups**	None	**Snapshots**
Availability SLA	99.9%	99.9%	**99.99%**

Firestore



Firestore: When to use?

Datastore is ideal for applications that rely on **highly available structured data** at scale.

Ideal Use Cases:

- Product catalogs that provide real-time inventory and product details for a retailer.
- User profiles that deliver a customized experience based on the user's past activities and preferences.
- Transactions based on ACID properties

Non-Ideal Use Cases:

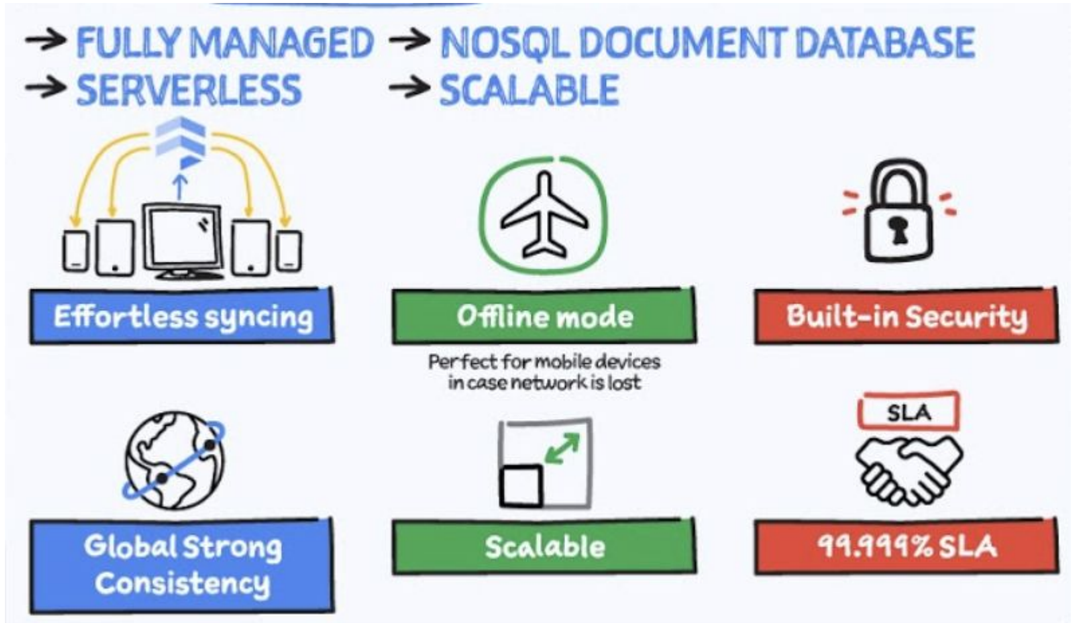
- OLTP relational database with full SQL support. *Consider:* Cloud SQL
- Data isn't highly structured or no need for ACID transactions. *Consider:* Cloud Bigtable
- Interactive querying in an online analytical processing (OLAP) system. *Consider:* BigQuery
- Unstructured data such as images or movies, *Consider:* Cloud Storage

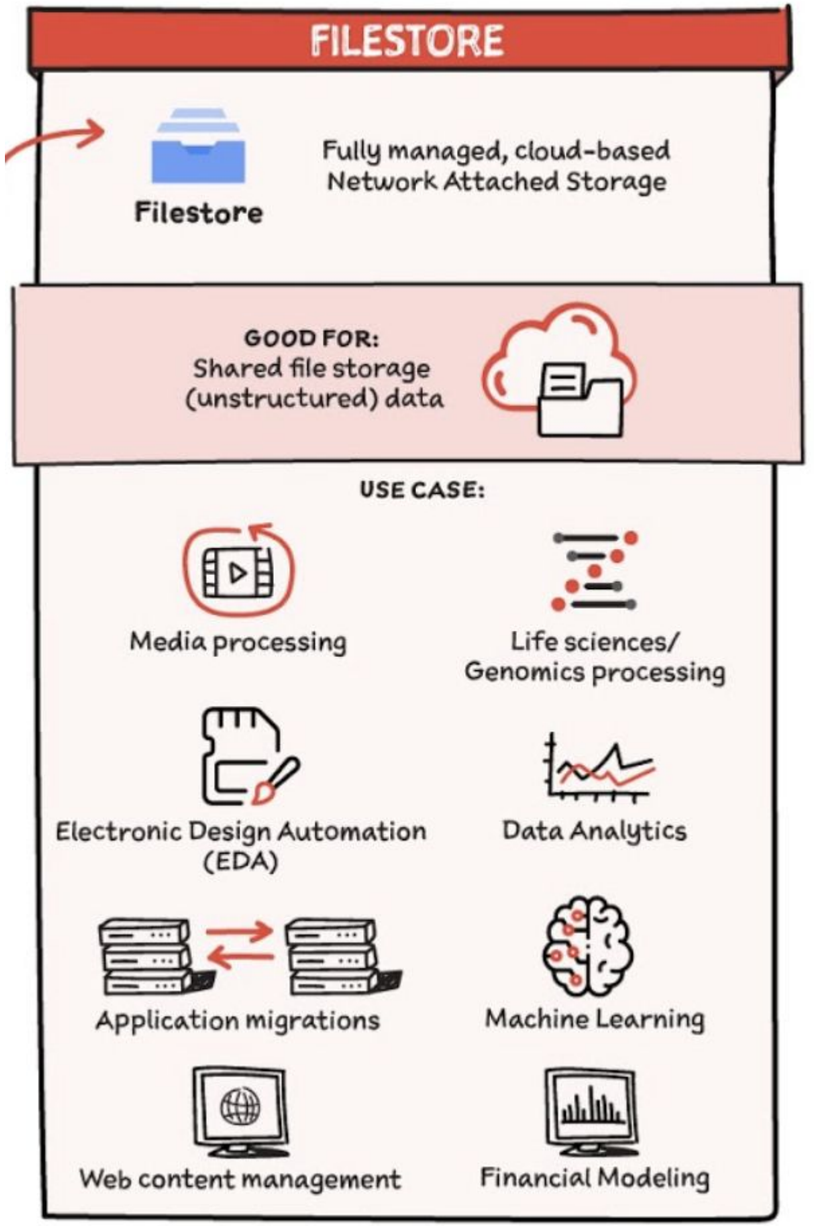
Firestore: Datastore mode vs Firestore (native) mode

	Both	Native Mode (only)	Datastore Mode (only)
Data model	Strong consistency	Documents and collections	Entities, kinds, ancestor queries/results
Performance limits	No read limits	10K writes/sec 500 documents/txn	
API		Firestore (Documents)	Datastore (Entities)
Security	IAM	Firebase Rules	
Offline data persistence		**Yes**	
Real-time updates		Yes	

Firestore or Datastore - comparison

Firestore vs Filestore



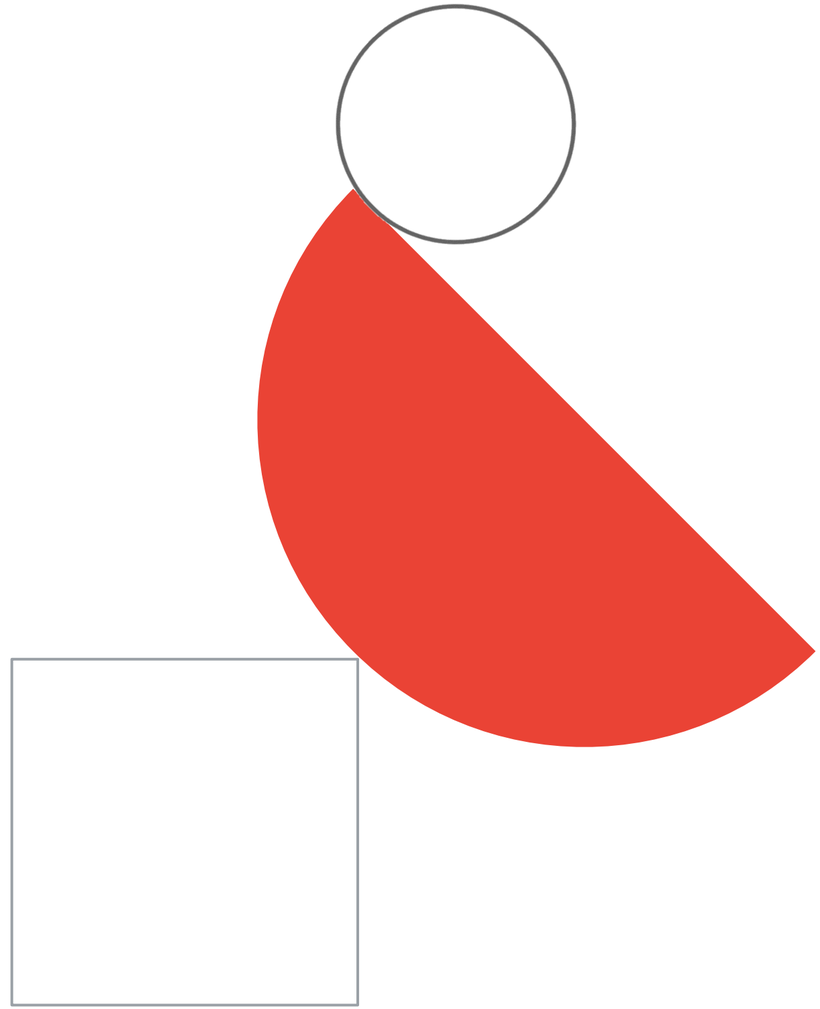


… vs Firebase

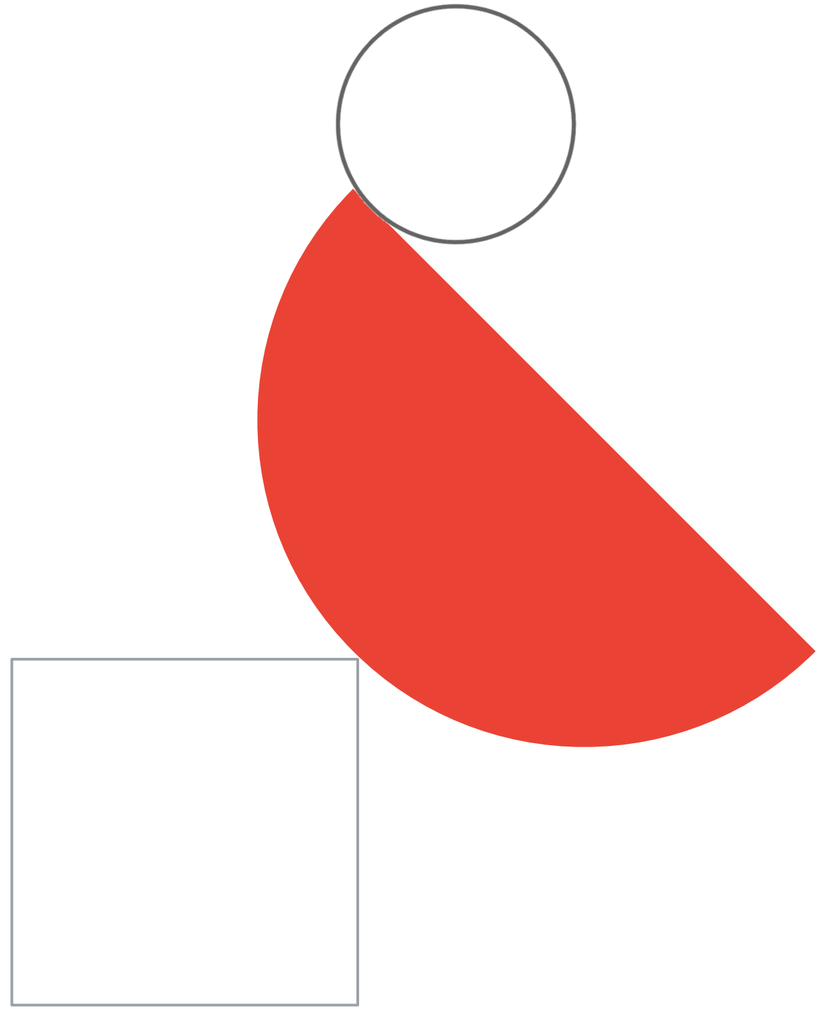
Exam Tip: *Firestore is a NoSQL Database, but Firebase is a development platform with a ton of additional features that uses Firestore. Make sure to differentiate between them!*

Firebase

* Platform, NOT a database *



Memorystore



Memorystore #GCPSketchnote

@PVERGADIA THECLOUDGIRL.DEV

6.29.2021

What is Memorystore?

- Web browser

- Fully managed in-memory data store for Redis & Memcached
- Scalable, more secure, & highly available

	Redis	Memcached
Storage	100s of GB	TBs
Throughput (in QPS)	100s of thousands	Millions

In memory caching layer - high-speed storage layer that stores a subset of data & reduces latency

What is your applications' availability need?

MEMORYSTORE FOR REDIS

BASIC TIER

Single Redis instance, ideal for caching use cases

- Instance health monitoring & automatic recovery from failures
- No SLA

STANDARD TIER

Replicated Redis instance, increased availability

- One secondary replica deployed across zones, protection from zone failures
- Seamless scale up down
- 99.9% availability SLA

FEATURES & CAPABILITIES

SECURE BY DEFAULT

Data is protected from the internet using VPC networks, private IP & IAM integration

Instance Auth, Data encrypted in-transit

SEAMLESS SCALE & HA

Standard high availability instances are replicated across zones

DEEP INSIGHTS

Monitor instances using cloud operations

BACKUP DATA

Easily backup instance data or import data into Memorystore from GCS buckets using RDB files

NO CODE CHANGES

OSS compliance allows using Memorystore without any code changes

CLOUD MEMORYSTORE USE CASES

GAMING

- Session Store
- Distributed Lock

SOCIAL MEDIA

- Content Caching
- Fast Data Ingestion
- Jobs and Queues
- Stream Processing

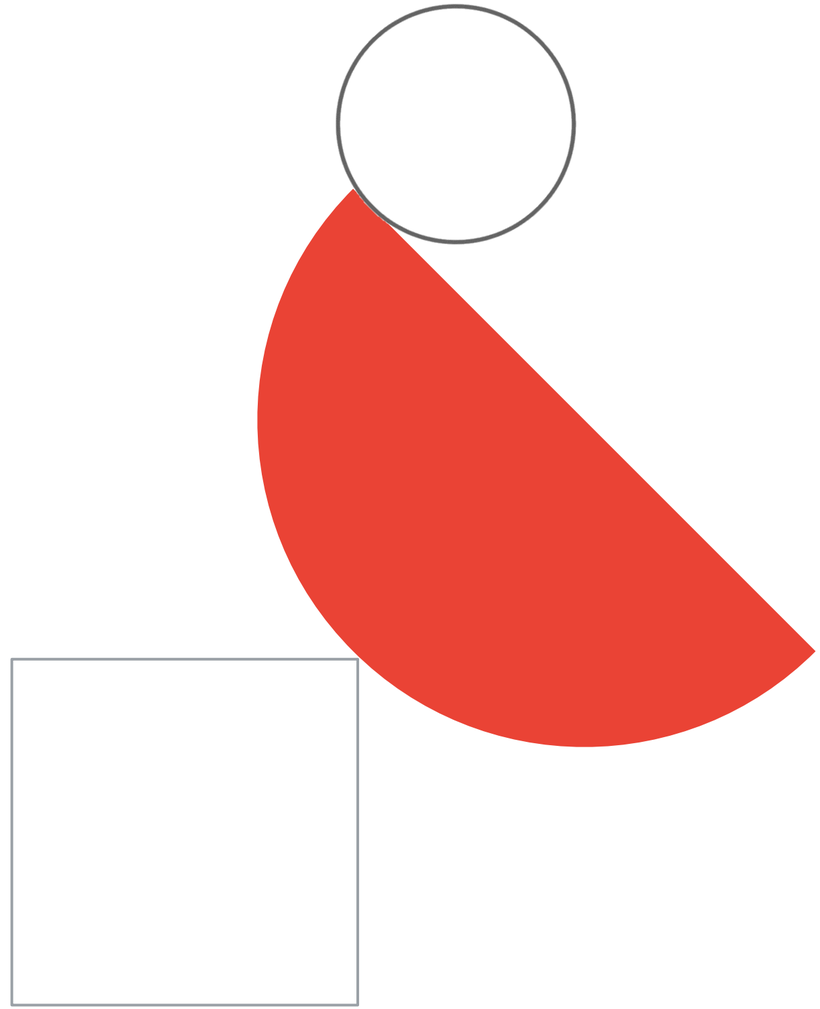
IoT

- Real-time Analytics
- Counters
- Leaderboard
- Page Ranking

ANALYTICS

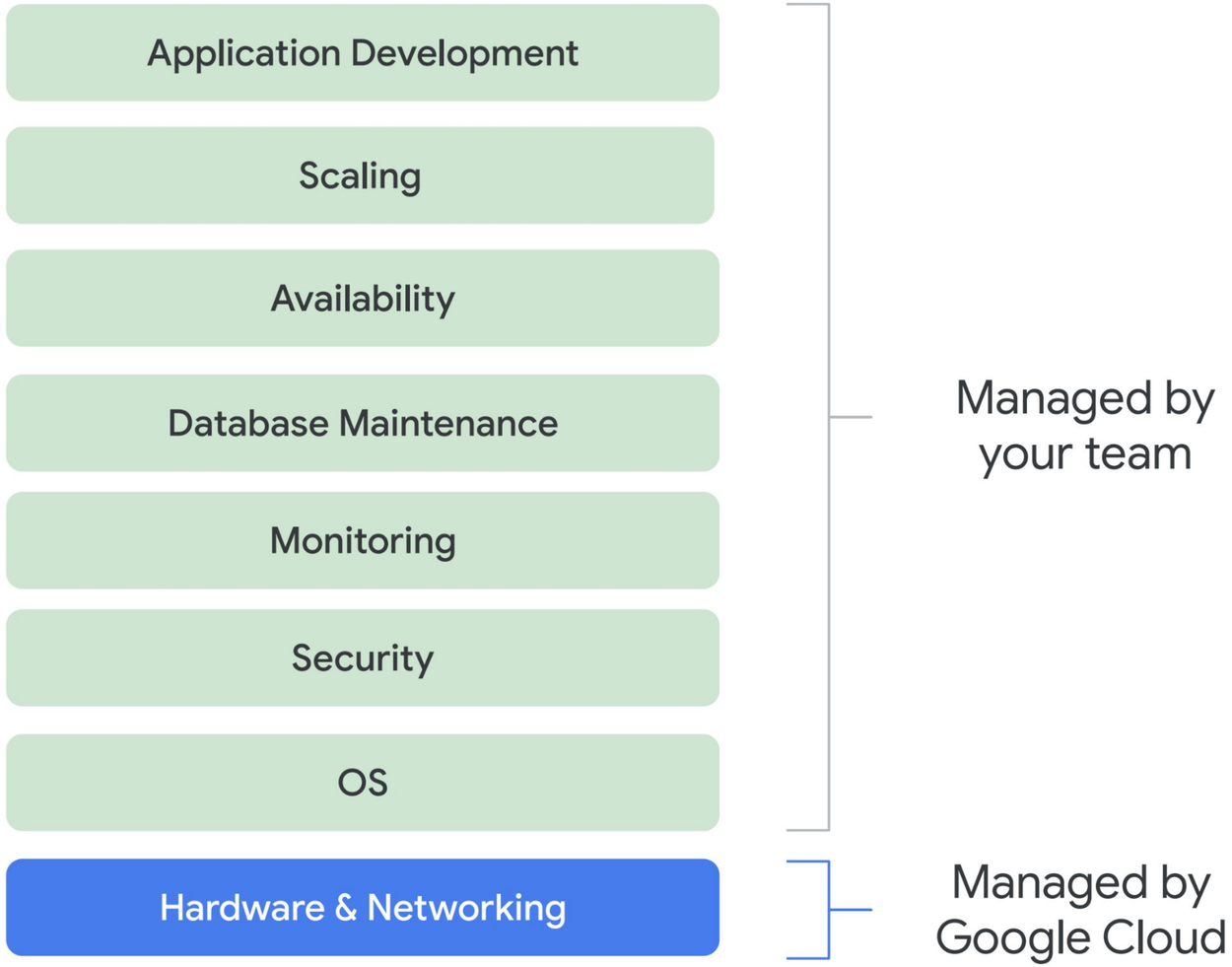
- Recommendation Engine

Cloud SQL



Managed databases: the “why”

Self-managed DBs on GCE VMs



Exam Tip: *custom OS images / startup scripts / products from GCP Marketplace etc...*

I can automate the installation, but you still need to...

- Provision resources
- Install database engine
- Configure it
- Patch & update
- Handle high availability
- Implement & manage backup & restore
- Resize when needed
- Implement monitoring
- ... and so on

Exam Tip: *Hence* ***self-managed databases are NOT a preferred option from the exam perspective****... unless you have a valid reason.*

Managed database services.

Exam Tip: *using managed services is usually a preferred approach from exam perspective (unless you're running into some constraints / have special needs.*

Application Development

Managed by your team

Scaling

Availability

Database Maintenance

Monitoring

Security

OS

Hardware & Networking

99.999% SLA

Online scaling

Easy global replication

Automatic sharding

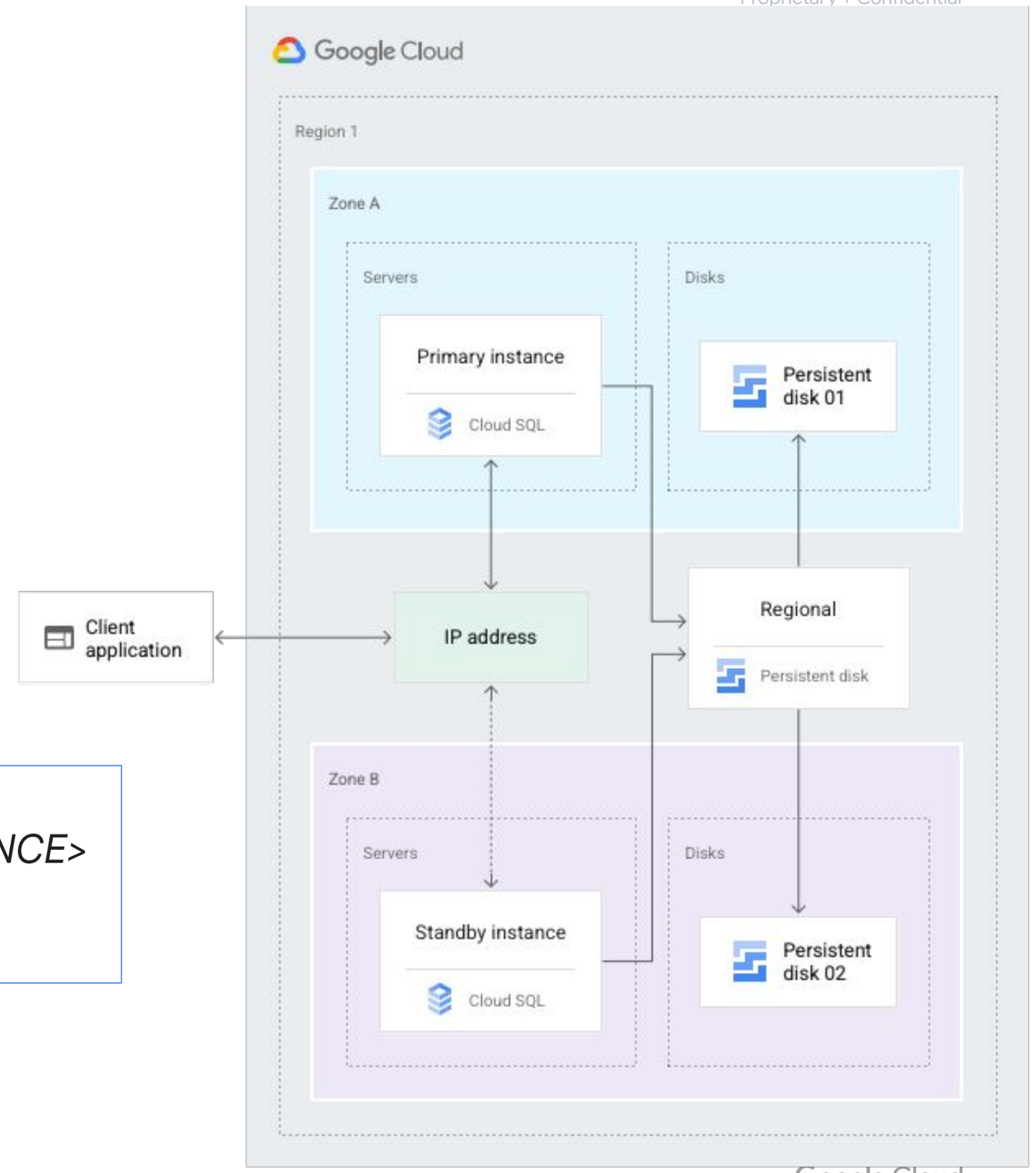
Automatic failure recovery

Built into cloud-native databases

Cloud SQL High availability

- Primary and secondary in different zones within the configured region
- Synchronous replication to each zone's persistent disk
- If heartbeat of the primary instance is unavailable for ~60 sec → automatic failover
- The persistent disk is then attached to the standby instance
- Less than 3 min of unavailability, the same IP address for the client application

Exam Tip: *Know how to:*
- *Initiate failover: gcloud sql instances failover <PRIM_INSTANCE>*
- *Check if instance is / isn't set for HA: gcloud sql instances describe (availabilityType = REGIONAL / ZONAL)*

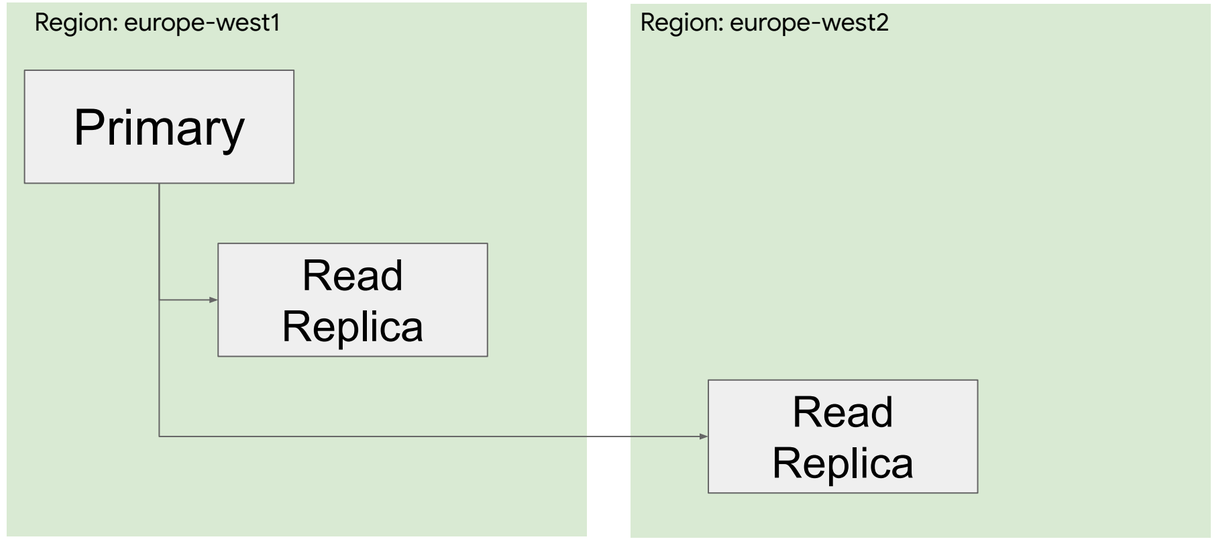


MySQL PostgreSQL Microsoft SQL Server

Cloud SQL: Read Replicas

Use cases: Disaster Recovery / offload analytics workloads / migrate between platforms or regions

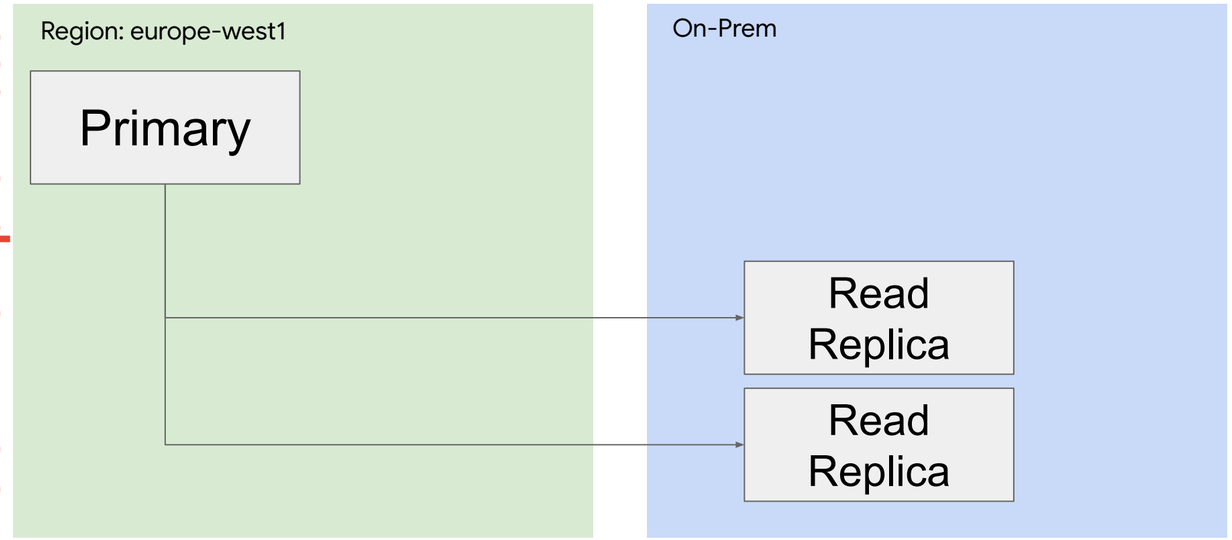
Read Replica
GCP -> GCP



Benefits & Use Cases

- Additional Read capacity (read only)
- Analytics target (adding secondary indexes)
- Read replicas can be different machine types than primary (never less vcpus for postgres)
- Settings of primary are propagated to replicas incl. root pwd &user table changes
- No load balancing between replicas
- Mysql Parallel replication (read replica side)

External Read Replica
GCP -> on-premises



Benefits & Use Cases

- Reduce latency for external connection
- Analytics target
- Migration path to other platforms
- In case of e.g. network outage on-prem the replication lag might be too large and replicas need to be recreated

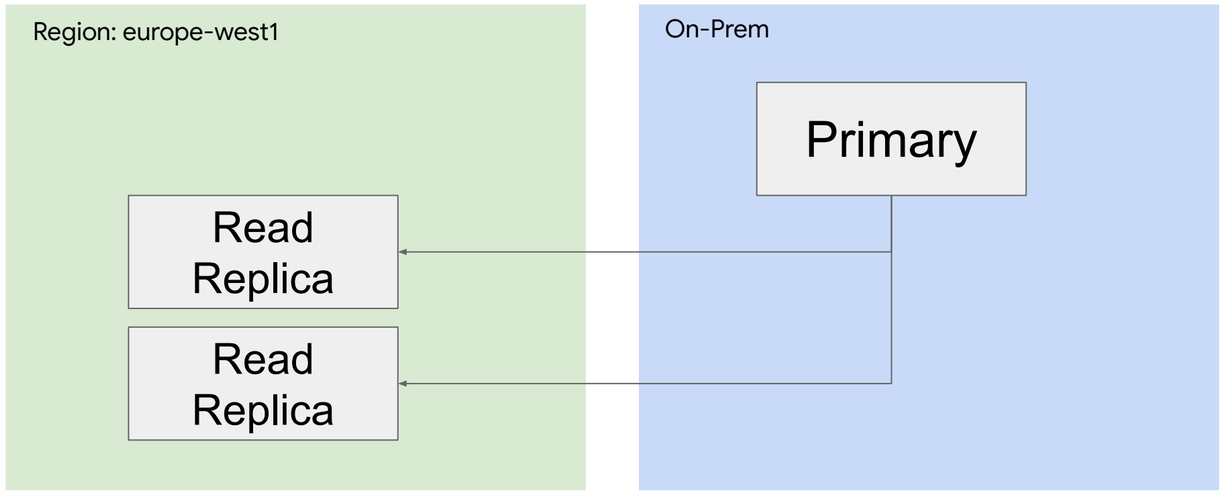
Exam Tip: *Docus on replicating TO external server*

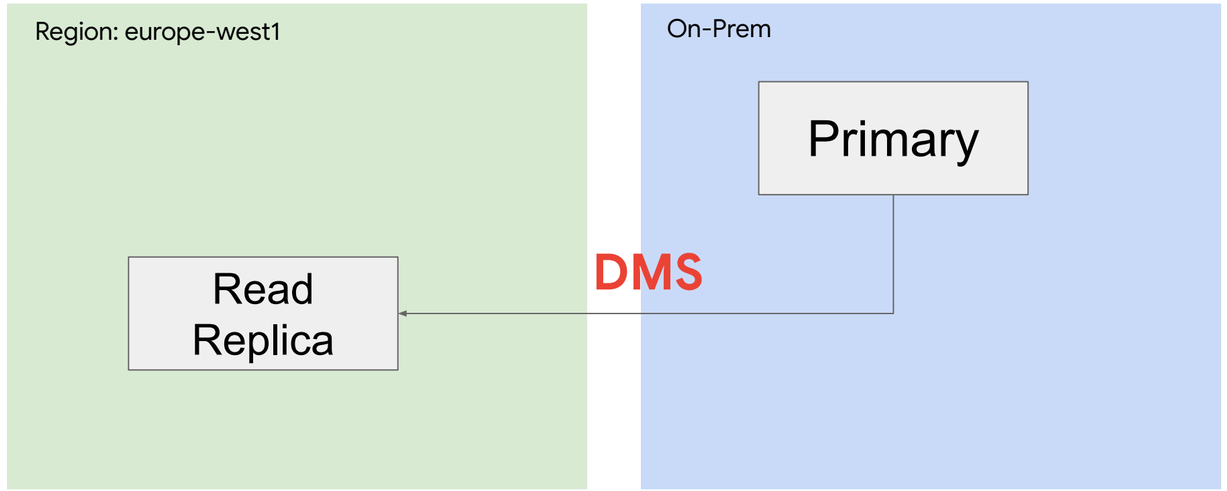
Cloud SQL: Read Replicas

Use cases: Disaster Recovery / offload analytics workloads / migrate between platforms or regions

Replication from external server

On-premises -> GCP





MYSQL Benefits & Use Cases

- Migration path to Cloud SQL with minimum downtime
- Data replication to GCP
- Offloading admin overhead of replicas to GCP
- Analytics target
- Parallel replication (read replica)

POSTGRES - DMS

- Use DMS to replicate from an external DB Server to a Cloud SQL Read replica (One-off Migration or Continuous cdc replication)
- DB Source can be self-managed DB (on-prem or IaaS), Aws Rds, Aurora, Cloud SQL

Exam Tip: *Docus on replicating FROM external server*

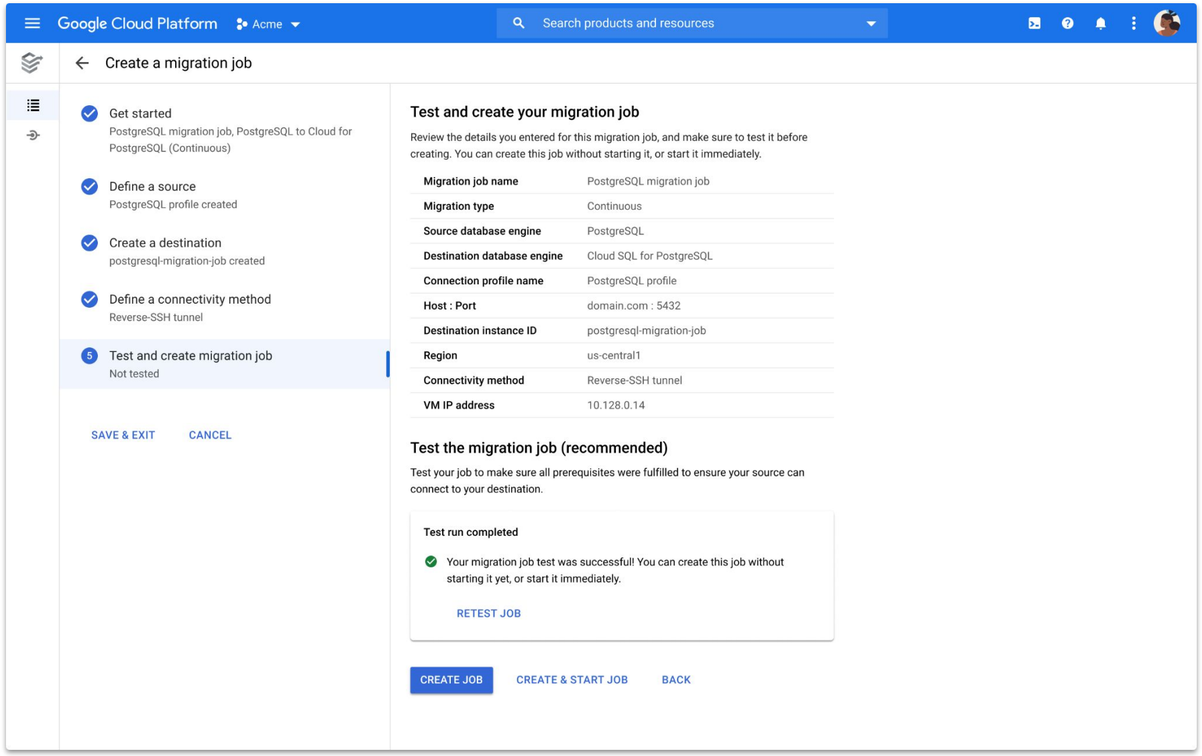
Simplified migration with Data Migration Service (DMS)

Continuous migration path to Cloud SQL with minimal downtime

Simplest way to migrate to Cloud SQL:

- No migration servers to manage
- Baked-in configuration support
- Native replication method

Secure connection options for encrypted data and using private networking



Exam Tip: *Make sure to be familiar with this overview: Database Migration Service.*

How to reduce Cloud SQL replication lag

Steps to enable parallel replication:

- On a read replica, disable replication.
- On the read replica, set the flags for parallel replication (Use the gcloud command to set the flags. The Google Cloud console option is disabled when replication is disabled):
  - Slave_parallel_workers, slave_parallel_type, slave_preserve_commit_order, slave_pending_jobs_size_max
- On the read replica, enable replication, using gCloud command or console
- Optionally, on the primary instance, set the flags to optimize performance for parallel replication.

Exam Tips:

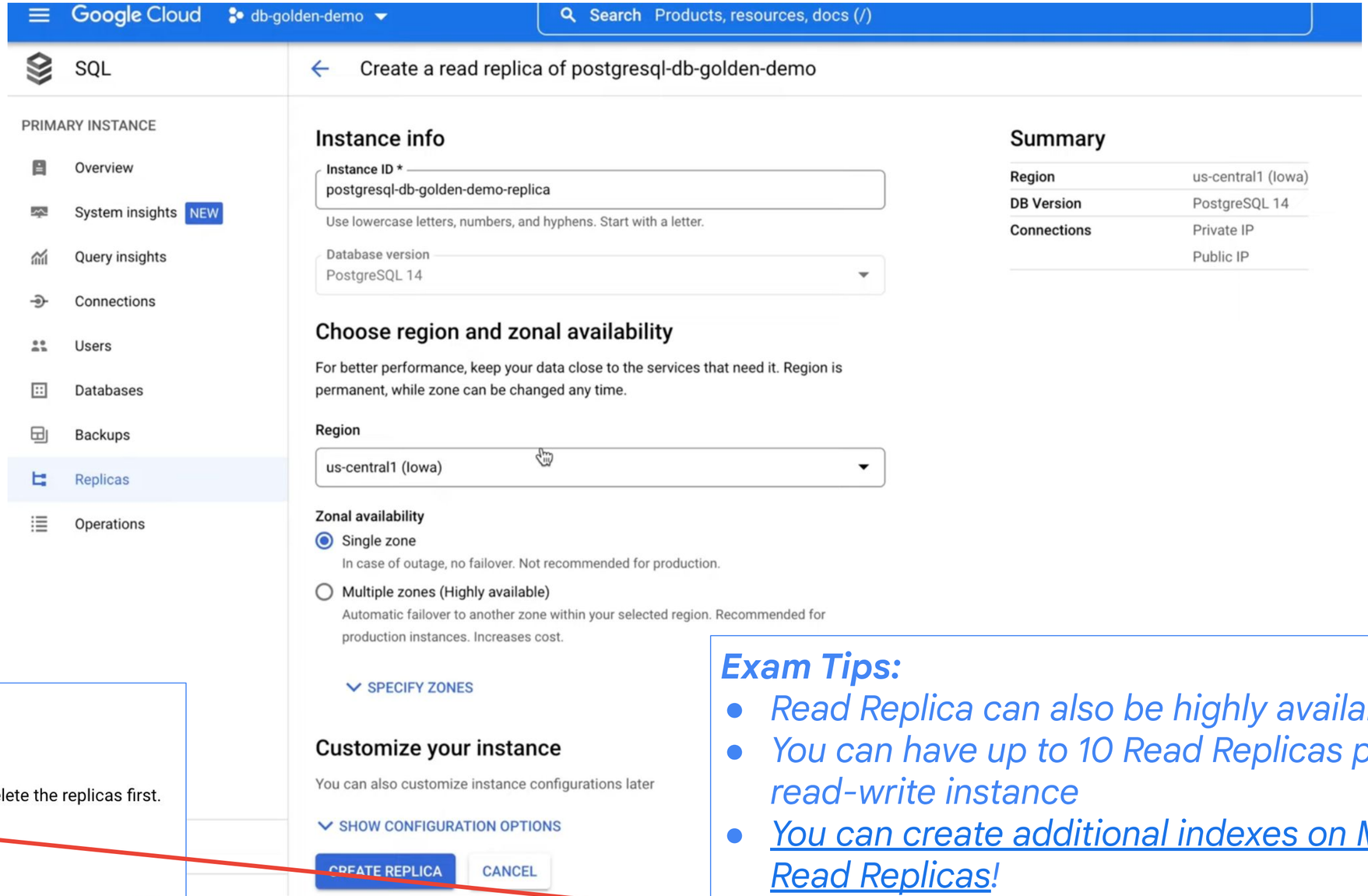
- *Have a look at replication lag topic (MySQL / PostgreSQL / SQL Server)*
- *You can use the **replica_lag** and **network_lag** metrics to monitor replication lag*
- *There are two ways to make a MySQL replica apply changes faster:*
  - *Parallel replication*
  - *High performance flushing*

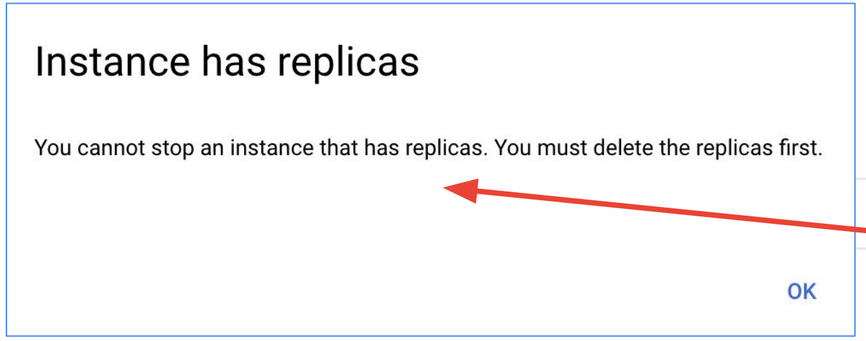
Exam Tips:

- *Know how to enable & disable replication using gcloud!*
  - *gcloud sql instances patch REPLICA_NAME --no-enable-database-replication*
  - *gcloud sql instances patch REPLICA_NAME --enable-database-replication*
- *Know how to set a flag:*
  - *gcloud sql instances patch INSTANCE_NAME --database-flags=FLAG1=VALUE1*

Cloud SQL DR with cross-region Read Replicas

How to create





Exam Tips:

- *Read Replica can also be highly available.*
- *You can have up to 10 Read Replicas per read-write instance*
- *You can create additional indexes on MySQL Read Replicas!*
- *!!!!*

Google Cloud

Cloud SQL: edit instance

Zone	Y	The possible values depend on the region.
Database version	N	Console string / API enum string **PostgreSQL 14** POSTGRES_14 **PostgreSQL 13** (default) POSTGRES_13 **PostgreSQL 12** POSTGRES_12 **PostgreSQL 11** POSTGRES_11 **PostgreSQL 10** POSTGRES_10 **PostgreSQL 9.6** POSTGRES_9_6
Set password policy	Y	Configured or not.
Private IP	After it is enabled, it cannot be disabled.	Enabled or disabled.
Public IP	Y	Enabled or disabled.
Authorized networks	Y	If Public IP is enabled, IP addresses authorized to connect to the instance. You can also specify this value as an IP address range, in CIDR notation.
Private path for Google Cloud services	Y	Enabled or disabled.
Machine type	Y	Select from Shared core, Lightweight, Standard (Most common), or High memory. Select the **Custom** radio button to create a custom machine type. Learn more

Exam Tips:

- *Can be done with command: gcloud sql instances patch INSTANCE_NAME –<setting_name> <value>*
- *Have a look at the parameter list, know the most important ones and focus if those can be changed AFTER instance creation or not.*
- *Storage (=regional PD) size CANNOT be reduced (just like with normal VMs)*

Cloud SQL - Performance & scaling

Machine Type

Choose a preset or customize your own. For better performance, choose a machine type with enough memory to hold your largest table.

Shared core

- (•) 1 vCPU, 0.614 GB
- ( ) 1 vCPU, 1.7 GB

...

- (•) Custom

vCPUs * 96 vCPU (1 - 96)

Memory * 624 GB (86.5 - 624)

Storage

Storage type

Choice is permanent. Storage type affects performance.

- ( ) SSD (Recommended)
  Most popular choice. Lower latency than HDD with higher QPS and data throughput.
- (•) HDD
  Lower performance than SSD with lower storage rates.

Storage capacity

10 - 65,536 GB. Higher capacity improves performance, up to the limits set by the machine type. Capacity can't be decreased later.

- (•) 10 GB

...

Storage

Storage type

Choice is permanent. Storage type affects performance.

- (•) SSD (Recommended)
  Most popular choice. Lower latency than HDD with higher QPS and data throughput.
- ( ) HDD
  Lower performance than SSD with lower storage rates.

Storage capacity

10 - 65,536 GB. Higher capacity improves performance, up to the limits set by the machine type. Capacity can't be decreased later.

- ( ) 10 GB
- ( ) 20 GB
- ( ) 100 GB
- ( ) 200 GB
- (•) Custom

65536 GB (10 - 65,536)

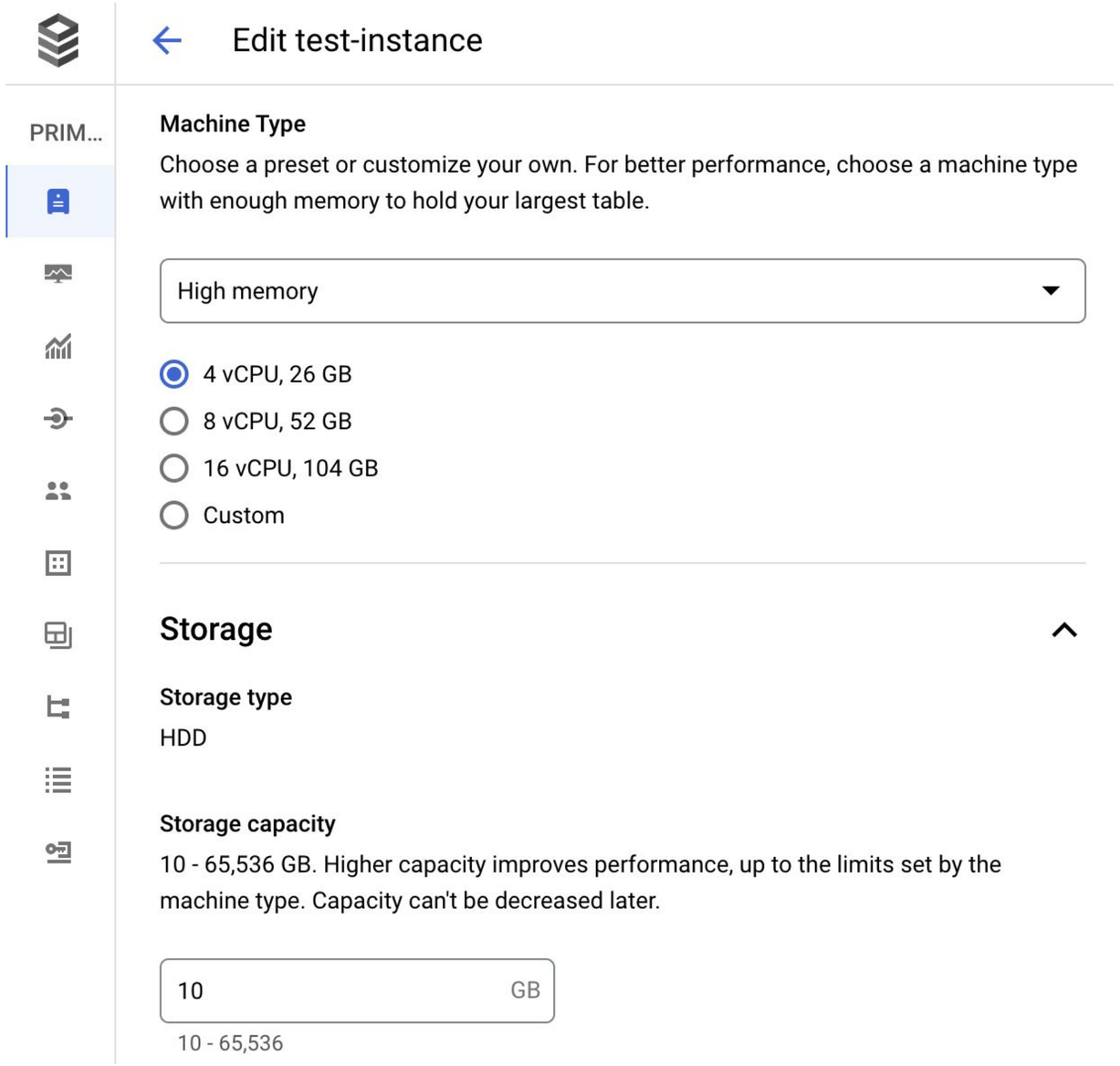
[x] Enable automatic storage increases

If enabled, whenever you are nearing capacity, storage will be incrementally (and permanently) increased. Learn more

Cloud SQL: storage

Can’t change between HDD <-> SSD

- SSD recommended
- Performance scales with size
- HDD / SSD decision is permanent
  - Change requires creating a separate instance and migrating data...
- Balanced disks NOT available for Cloud SQL
- Secondary (HA) instances are of the same size & disk size&type
- Read Replicas can be of larger (not smaller!) size (vCPUs/memory), but the same disk size&type



Cloud SQL: GCP-native Backup and Restore

Types

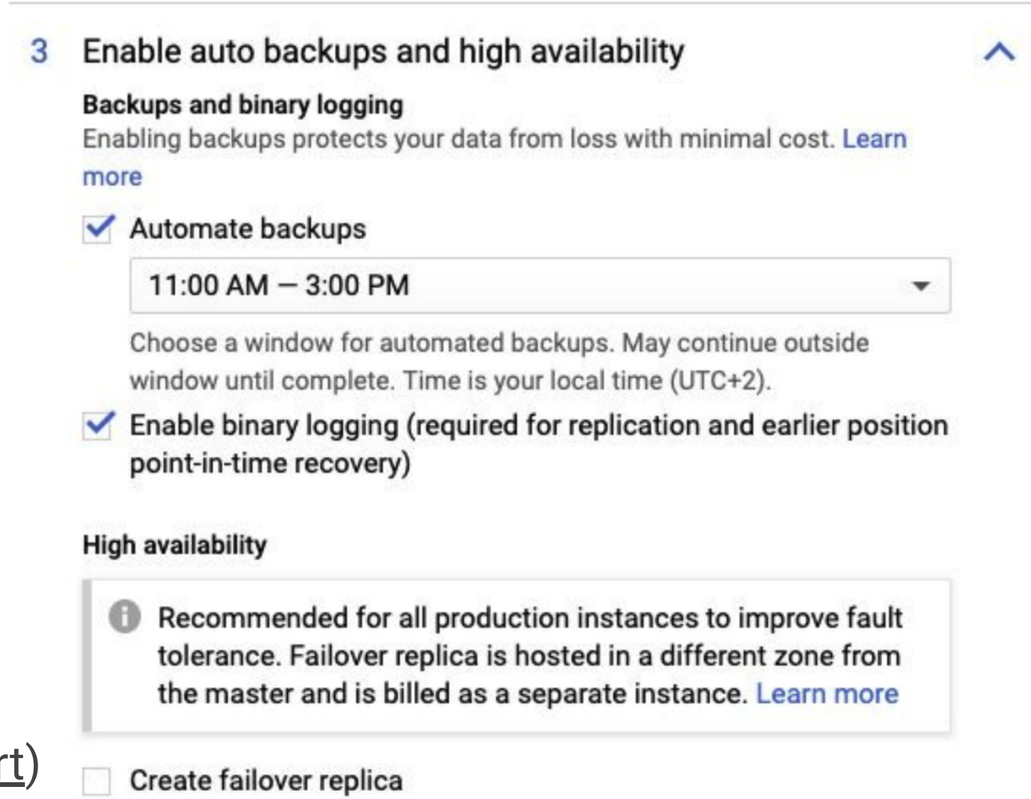
- **On-Demand**
  - Create disk-level snapshot backup at any time
  - Not deleted automatically

- **Automated**
  - 4 hour backup window (e.g. 11am - 3pm)
  - Schedule when instance has least activity
  - If data has not changed since last backup then no backup is taken

Characteristics

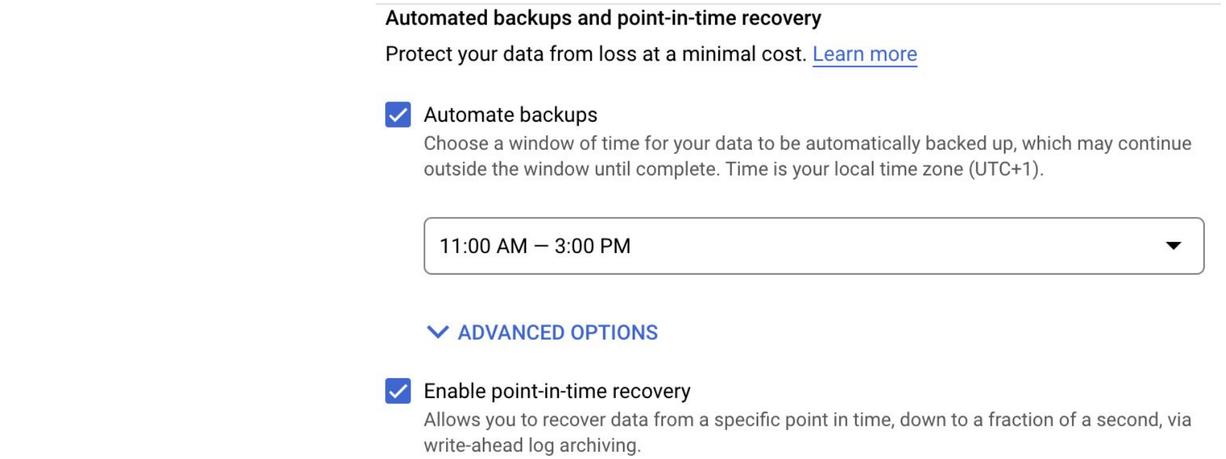
- Up to 365 daily automated backups for each instance. (Only up to 7 days for binlog/WAL files)
- Incremental Backups
- Storage used by backups is charged at a reduced rate. (see pricing)
- Backups can not be exported - only instance data (see doc for export)
- Backups are deleted after instance is deleted (Data export required to retain data; read replica for export without perf. impact)
- Backups are disk-level snapshots stored on GCS

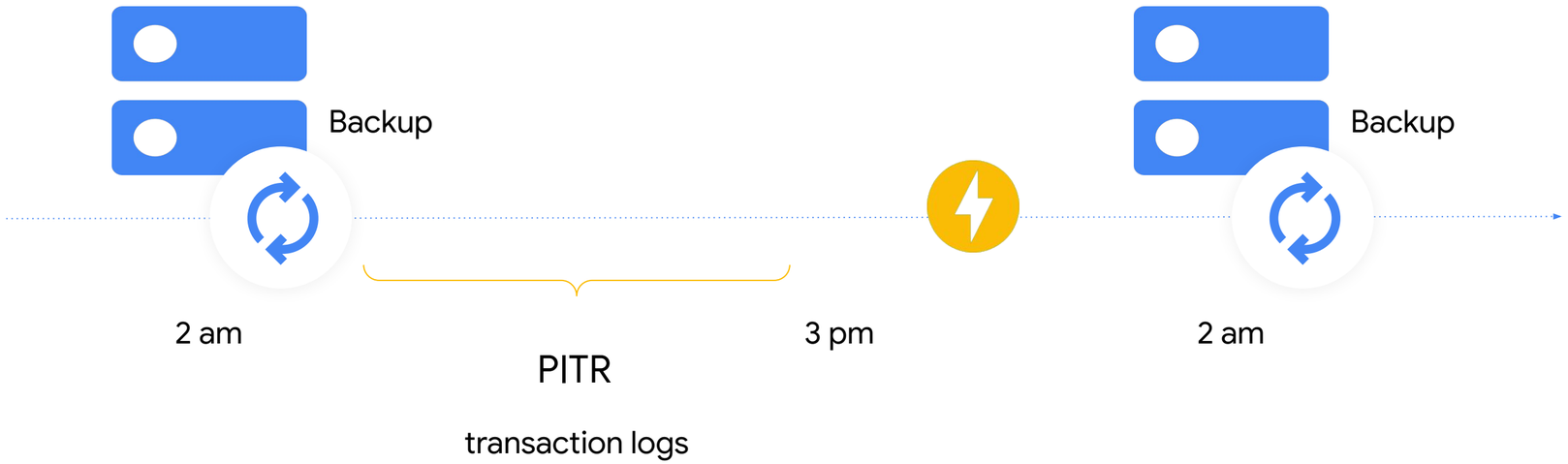
Exam Tip: *Cloud SQL backup window is NOT the same as maintenance window.*



Cloud SQL: Point-in-time recovery

recover an instance to a specific point in time



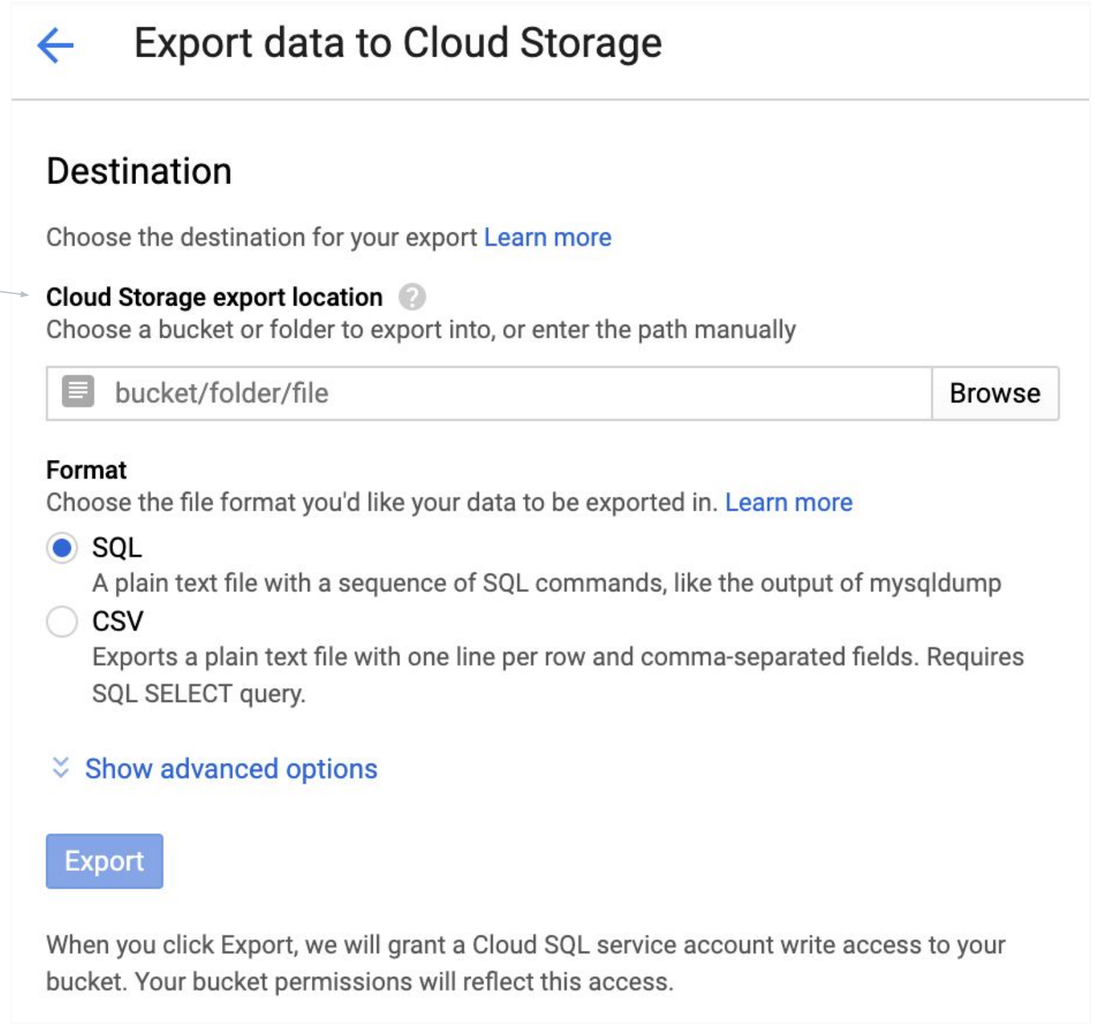


Cloud SQL: Data Export

Export (to GCS)

- ➔ SQL - high fidelity
- ➔ CSV - database-agnostic

Note:
Export to different buckets in different project is also supported if you have granted the service account with write permissions on this bucket.



Exam Tip: *For regular & automatic Cloud SQL exports, use Cloud Functions and Cloud Scheduler.*

Connection Options (external apps)

Exam Tip: *Make sure to know when to use which pattern: Cloud SQL Connection options*

Exam Tip: *Common solution to questions about connectivity from a GKE cluster to Cloud SQL*

Connection option	Secure, encrypted?	More information	Notes
Public IP address with SSL	Yes	• Configuring SSL for Instances • Configuring access for IP connections • Connect mysql client using SSL	SSL certificate management required.
Public IP address without SSL	No	• Configuring access for IP connections	Not recommended for production instances.
Cloud SQL Proxy	Yes	• Connecting from an external application using the Cloud SQL Proxy • Connecting mysql Client Using the Cloud SQL Proxy • About the Cloud SQL Proxy	
Cloud SQL Proxy Docker image	Yes	• Connecting mysql Client Using the Cloud SQL Proxy Docker Image • About the Cloud SQL Proxy	
JDBC Socket Library	Yes	• External connections with Java • JDBC socket factory GitHub page	Java programming language only.
Go Proxy Library	Yes	• External connections with Go • Cloud SQL Proxy GitHub page	Go programming language only.
Cloud Shell	No	• Using the mysql client in the Cloud Shell	Uses the Cloud SQL Proxy to easily connect from the Google Cloud Console. Best for quick administration tasks requiring the `mysql` command-line tool.
Apps Script	Yes	• External connections with Apps Script • Apps Script sample GitHub page	Apps Script can connect to external databases through the JDBC service, a wrapper around the standard Java Database Connectivity technology.

Cloud SQL

IP address assignment

Private IP:

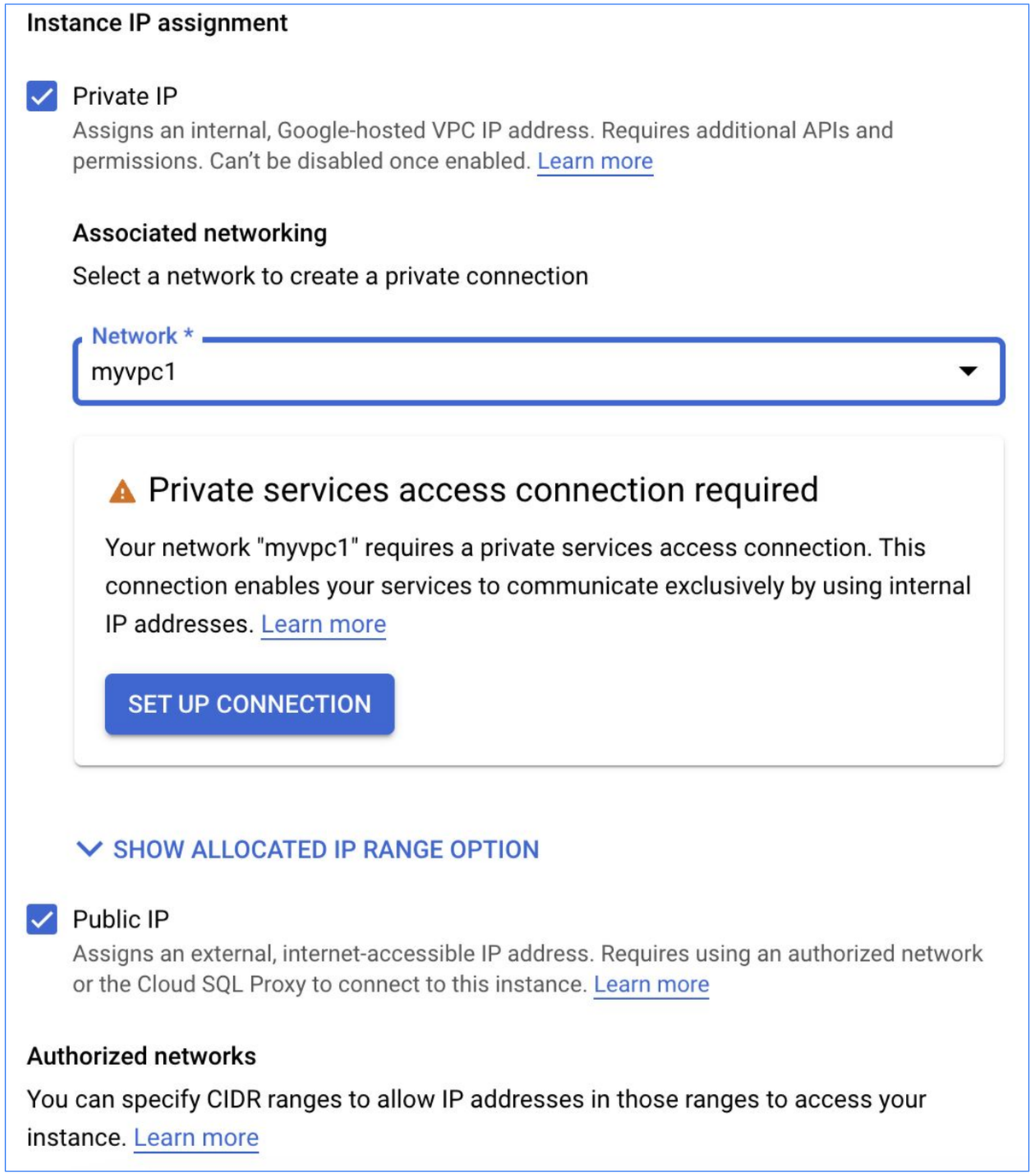
- Preferred when client is coming from resource with internal visibility (not necessarily from the same VPC!
- IPv4 address accessible from VPC
- Connections **may** be configured to use Cloud SQL proxy or self-managed SSL certificates
- Low latency and increased security

Public IP:

- IPv4 address accessible from the public network
- Connections **must** be authorized using either the Cloud SQL Auth proxy or authorized networks

Exam Tip: *Cloud SQL instances can have* ***both*** *a public and a private IP address. If private IP address is configured, Private Service Access (technically: VPC peering) is configured underneath.*

MUST-WATCH VIDEO

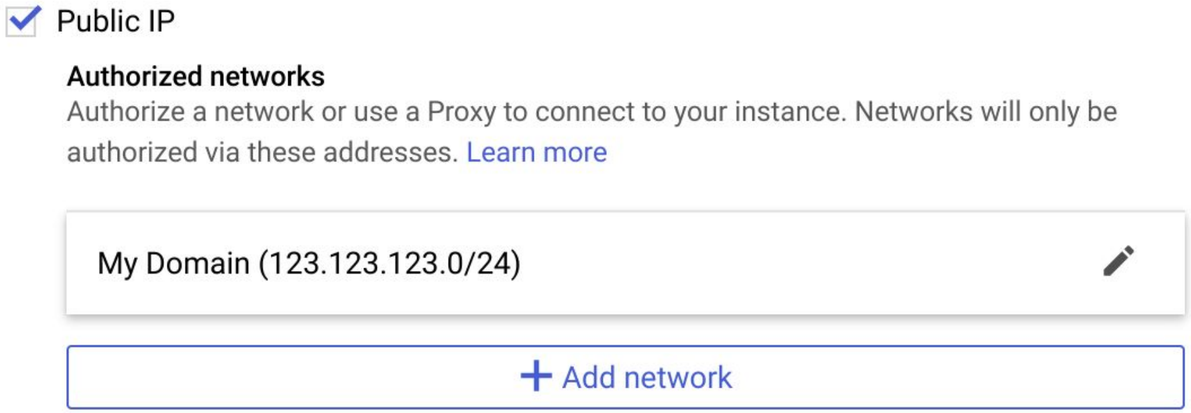


Cloud SQL: Public IP & risk mitigation

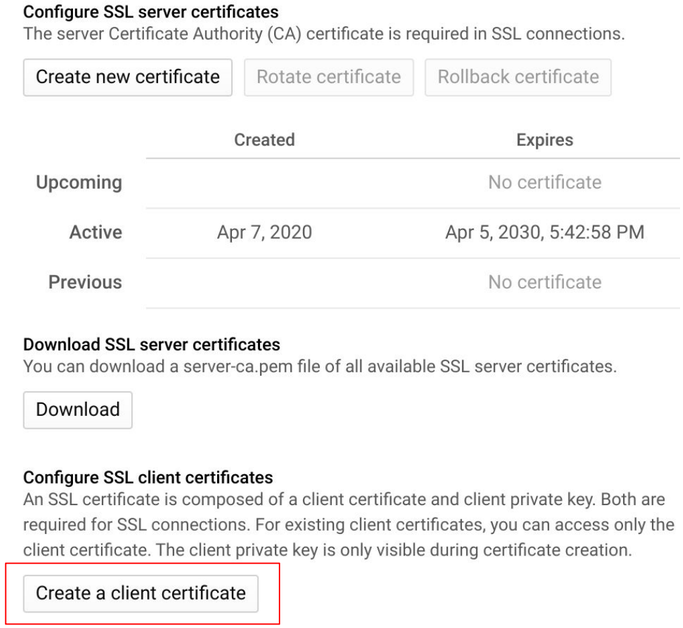
Access by public IP, without SSL : maximum risk for your data !!

Risk mitigation options:

Set an authorized network domain



Use SSL Certificate for transit data




```
psql "sslmode=verify-ca sslrootcert=server-ca.pem \
      sslcert=client-cert.pem sslkey=client-key.pem \
      hostaddr=01.23.45.67 \
      user=postgres dbname=postgres"
```

Cloud SQL: Private IP

Characteristics

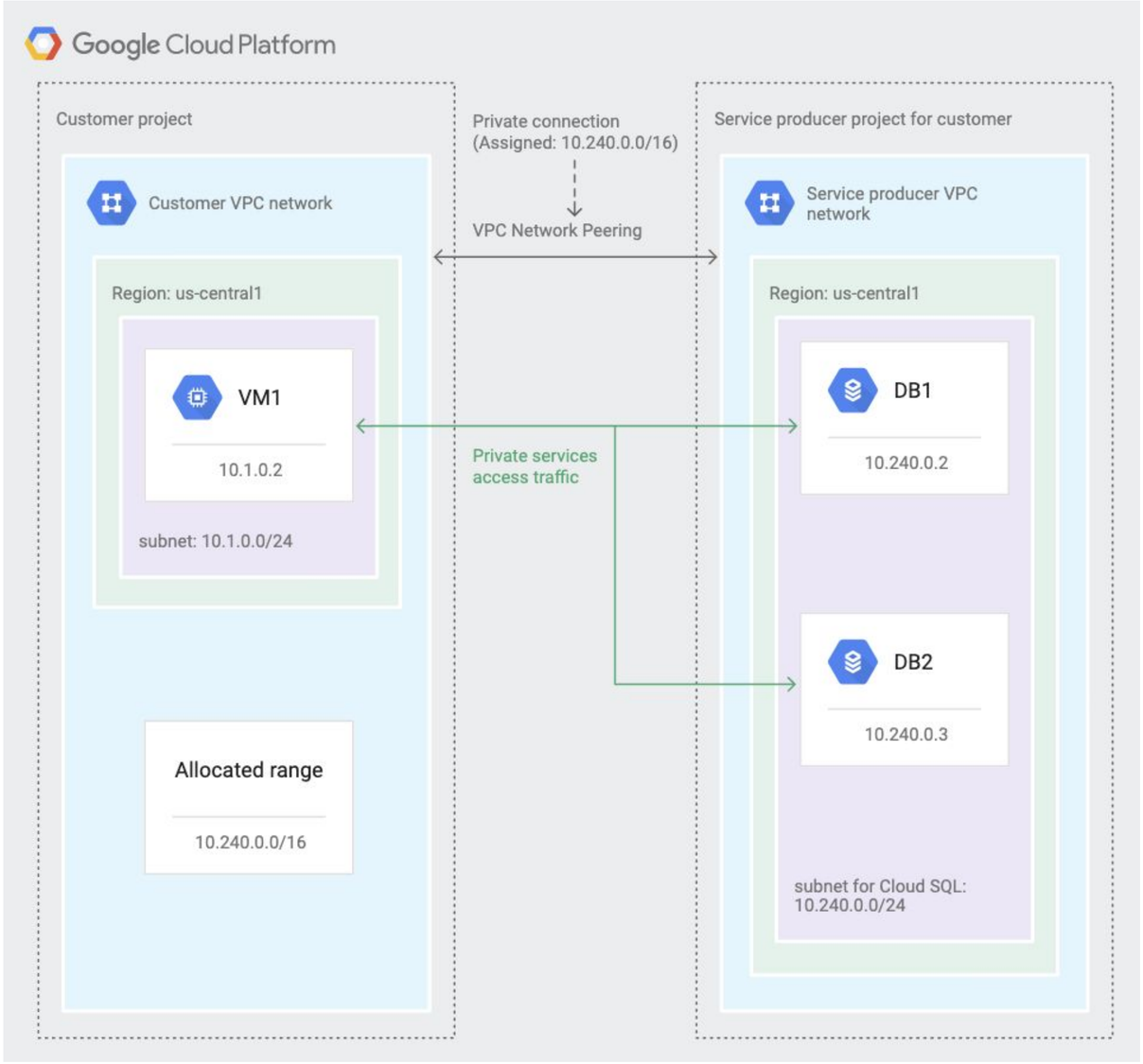
- Allows VM instances in your VPC network to use internal IP addresses to reach the service resources that have internal IP addresses
  → i.e. Connect CloudSQL to GCE or GKE instances
- **GCP uses network peering to create connection**

Benefits

- **Lower network latency**
  Best performance
- **Improved network security**
  Traffic is never exposed to public internet
- **Lower egress cost**
  Regular network pricing still applies to all traffic.

Limitations

- **Max 25 Peering connection per VPC network**
- **Transitive peering is not supported.**
- **Subnet in Peered VPC cannot overlap with CloudSQL**



Cloud SQL

Access authentication

- Basic roles (should NOT be used!):
  - Owner (Full access and control for all Google Cloud resources)
  - Editor (Read-write access to all Google Cloud resources)
  - Viewer (Read-only access to all Google Cloud resources)

Common Cloud SQL IAM Roles:

Role (predefined)	Privileges	For who/which service
Cloud SQL Admin	Full control for all Cloud SQL resources.	DBA Team / DB owner
Cloud SQL Editor	Manage Cloud SQL resources. No ability to see or modify permissions, nor modify users or sslCerts. No ability to import data or restore from a backup, nor clone, delete, or promote instances. No ability to start or stop replicas. No ability to delete databases, replicas, or backups.	DB Operator
Cloud SQL Viewer	View all Cloud SQL resources (read-only)	Audit, Security, DevOps Team
Cloud SQL Client	Connectivity access to Cloud SQL instances from App Engine and the Cloud SQL Proxy. Not required for accessing an instance using IP addresses.	Apps service (AppEngine, CloudSQL Auth Proxy)

Exam Tip: *Understand permissions in predefined Cloud SQL roles: Admin / Editor / Viewer / Client.*

Cloud SQL

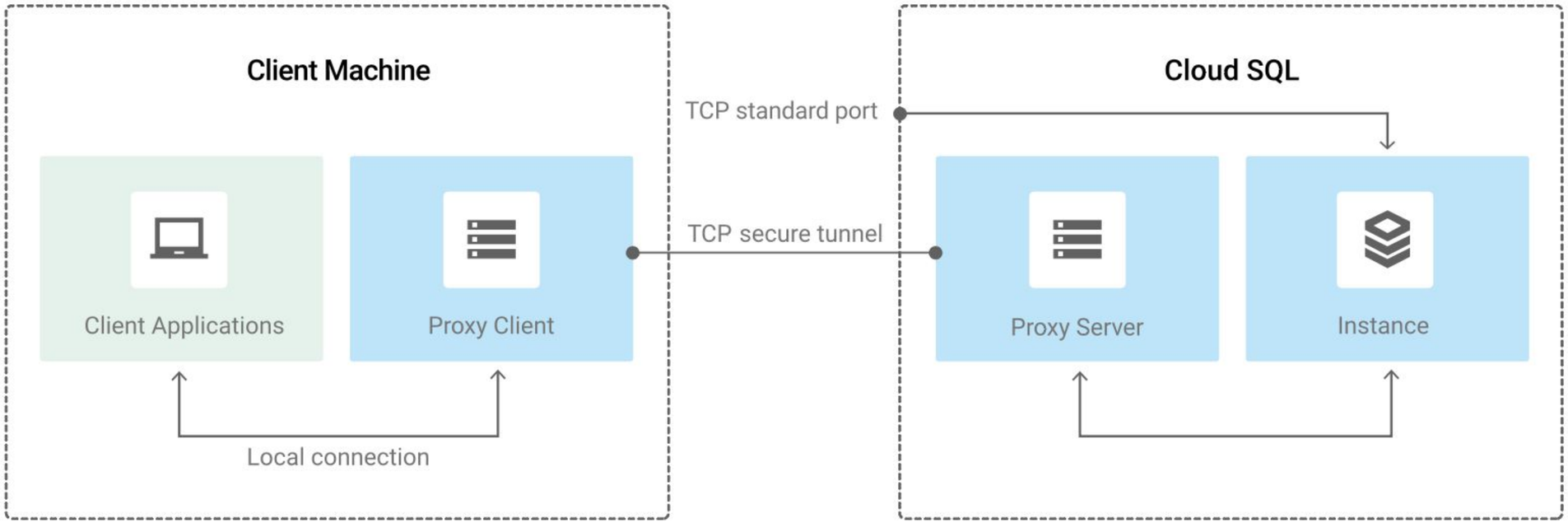
Access authorization -> Cloud SQL Auth proxy details

How the Cloud SQL Auth proxy works?

- a local client running in the local environment + companion process running on the server.
- doesn't provide connection pooling, but can be paired with other connection pooling to increase efficiency.
- also available as a Docker container.

Exam Tips:

- *Cloud SQL Proxy is the recommended option even when connecting to Cloud SQL behind a Private IP (because of strong encryption and authentication using IAM)*



CloudSQL

#GCPSketchnotes

@PVERGADIA THECLOUDGIRL.DEV

What is CloudSQL?

FULLY MANAGED RELATIONAL DB FOR

MySQL PostgreSQL SQL Server

How does it WORK?

CREATE

region
+ zone
+ machine type (cpu & memory)
+ storage (SSD or HDD)
= Instance

- [x] Enable automatic storage increases

BACKUPS

- [x] Automate backups [time windows]
- [x] Enable point-in-time recovery

HIGH AVAILABILITY

Automatic failover to another zone or region

ZONE A
AUTOMATIC FAILOVERS
ZONE B

How do you USE it?

CONNECTIVITY WITH OTHER SERVICES

CloudSQL connects to App Engine, GKE, GCE, Cloud Run and Workstation

CloudSQL

Steps to MIGRATE

EXISTING MYSQL DB TO CLOUDSQL

1 2 3 4

1. PROVIDE details on your data source
2. CREATE a Cloud SQL read replica
3. SYNC read replica with source
4. PROMOTE read replica to primary instance (optional)

SECURITY for CloudSQL?

ENCRYPTED AT REST & IN MOTION

CloudSQL PRICING

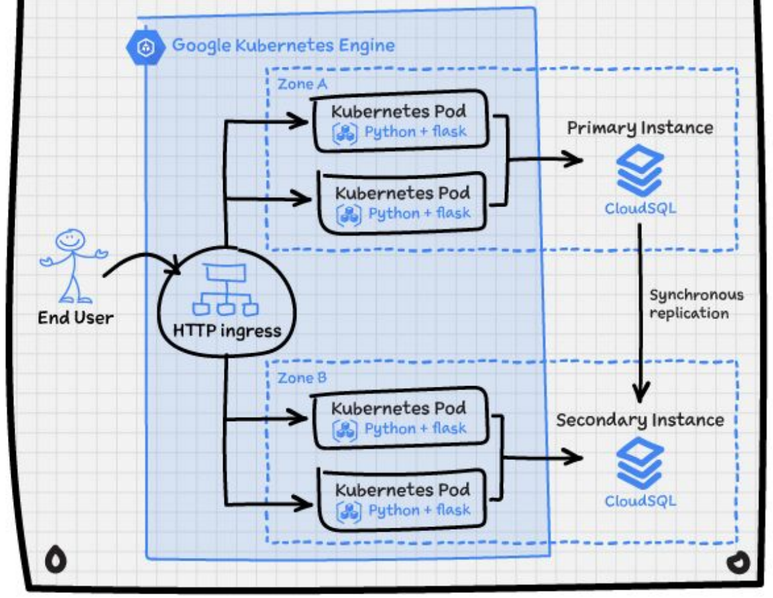
IN GENERAL PRICING IS A FACTOR OF:

Instance + CPU + Memory + Storage + Network Cost

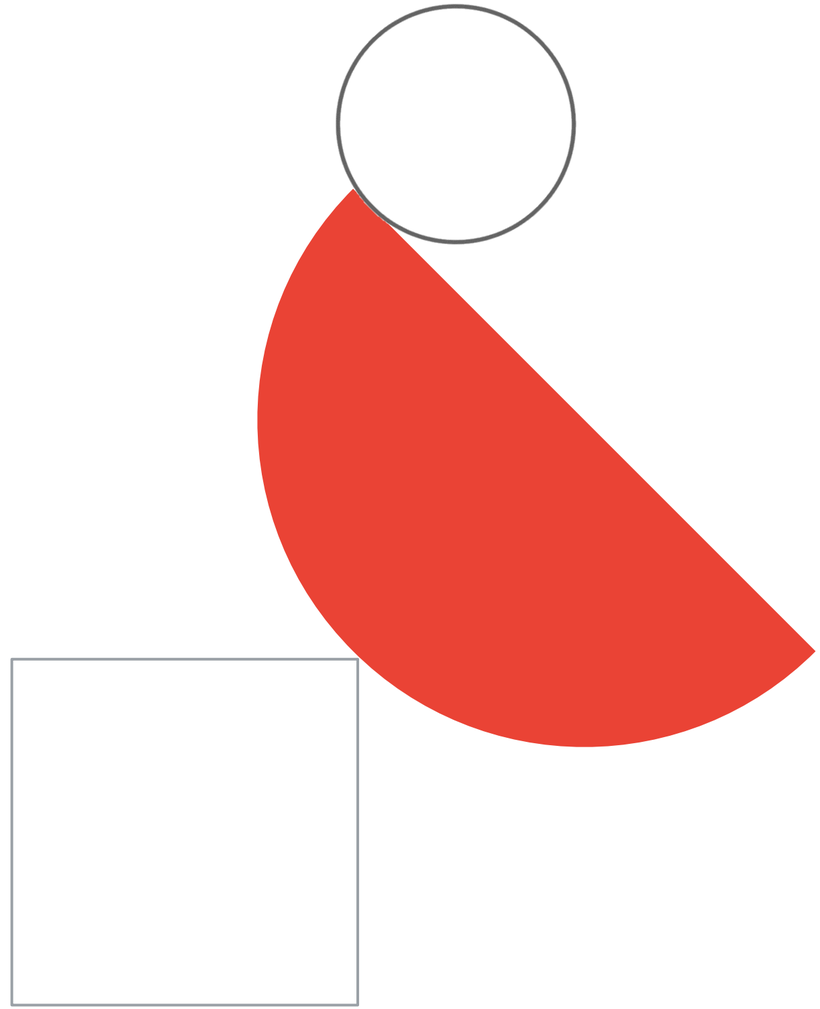
$

CloudSQL Use case (example)

- [x] Online Transaction processing (OLTP)
- [x] Frequent queries
- [x] Fast response times



Spanner



What workloads fit Cloud Spanner best?

01

Sharded RDBMS

Manually sharding is difficult. People do it to achieve scale. Cloud Spanner gives you relational data and scale.

02

Scalable relational data

Scalable relational database. Instead of moving to NoSQL, move from one relational database to a more scalable relational database.

03

Manageability/HA

Highly automated. Online Schema changes and patching. No planned downtime and comes with up to a 99.999% availability SLA.

04

Multi-region

Write once and automatically replicate your data to multiple regions.

Most customers use regional instances, but multi-region is there if you need it.

When Cloud Spanner fits less well

TIP

It’s NOT a straightforward thing to migrate a different RDBMS to Cloud Spanner. Be familiar with challenges on high level.

1. Lift and shift
2. Lots of in-database business logic (triggers, stored procedures)
3. Compatibility needed
4. App is very sensitive to very low latency (micro/nano/low single digit ms)

   Lots of analytics / OLAP type of queries / workloads

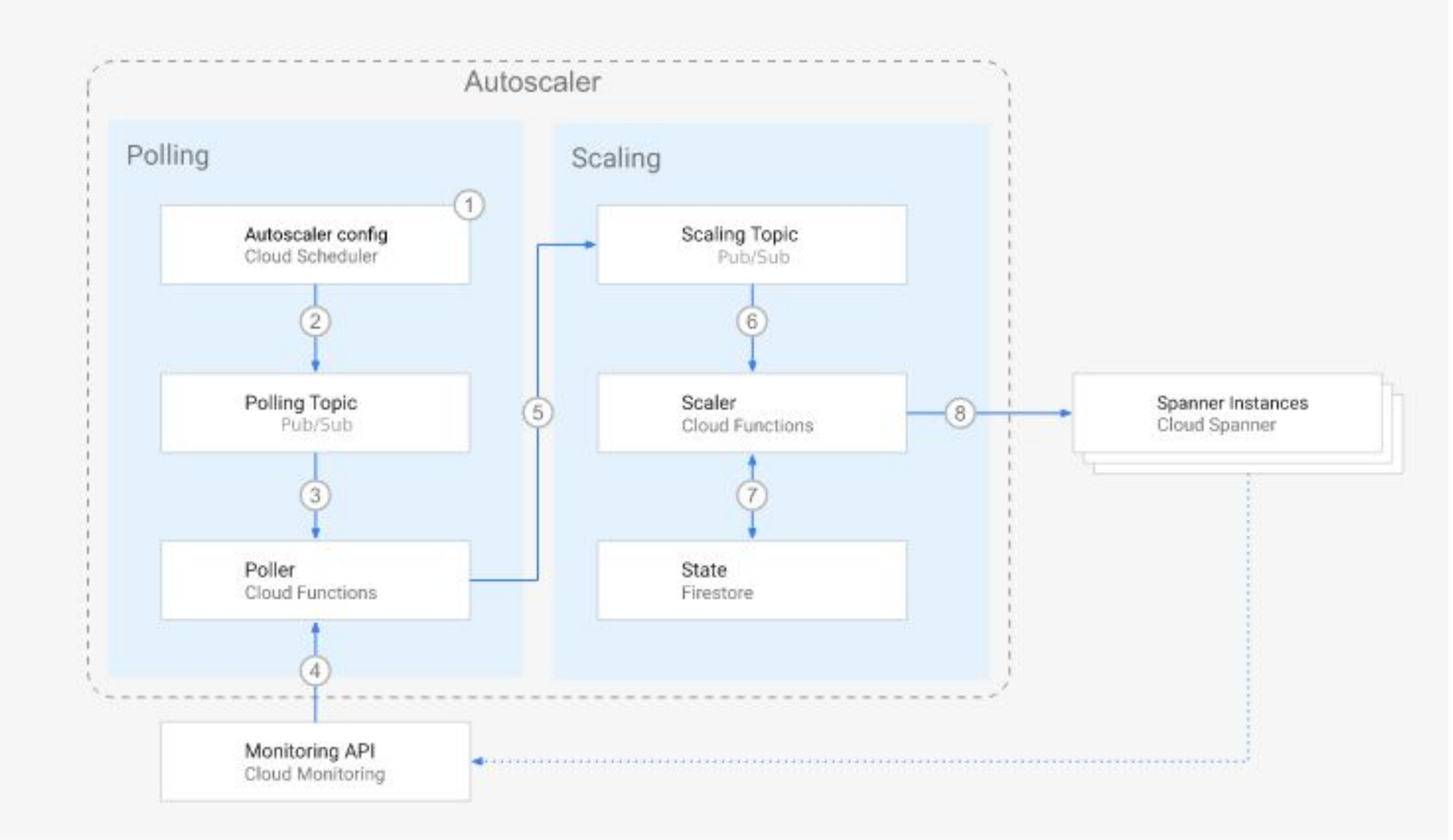
Cloud Spaner - (auto)scaling

Some DIY (still) required...

- The Autoscaler architecture consists of Cloud Scheduler, two Pub/Sub topics, two Cloud Functions, and Firestore. The Cloud Monitoring API is used to obtain CPU utilization and storage metrics for Spanner instances.

Exam Tips:

- *Do-It-Yourself still needed to automatically scale Spanner*
- *Scale Spanner nodes mostly based on CPU utilization metrics.*



Google Cloud

Cloud Spanner

#GCPSketchnote

@PVERGADIA THECLOUDGIRL.DEV

5.7.2021

What is Cloud Spanner?

- ✓ FULLY MANAGED
- ✓ HORIZONTALLY SCALABLE
- ✓ GLOBALLY CONSISTENT
- ✓ RELATIONAL DATABASE
- ✓ MULTI-VERSION DATABASE

Relational Semantics

Schemas, ACID transactions, SQL

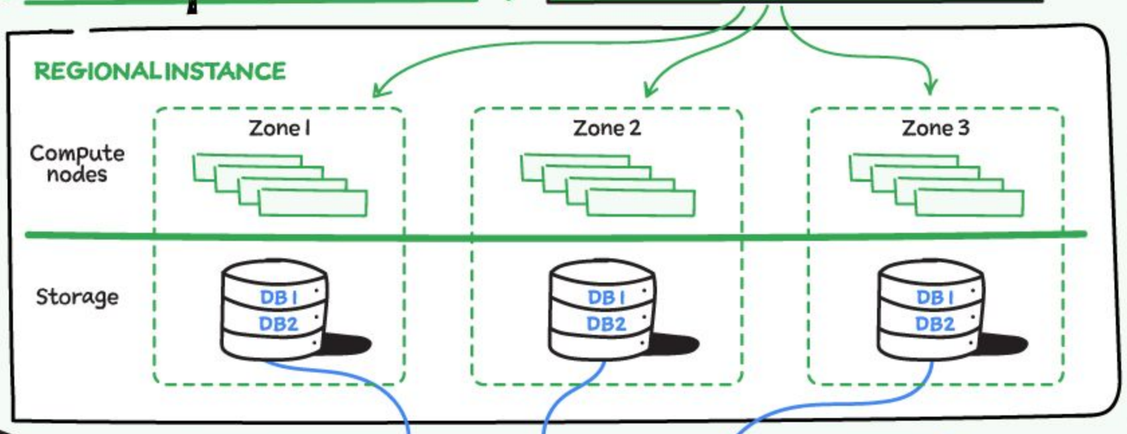
Horizontal Scale

99.999% SLA, fully managed, and scalable

Relational | Non-Relational

How does Cloud Spanner work?

This Spanner instance contains 4-nodes



This Spanner instance is hosting 2 databases across 3 zones

How does Spanner provide «global consistency?»

SPANNER GLOBAL CONSISTENCY

TrueTime

Synchronizes clocks in all machines across datacenters

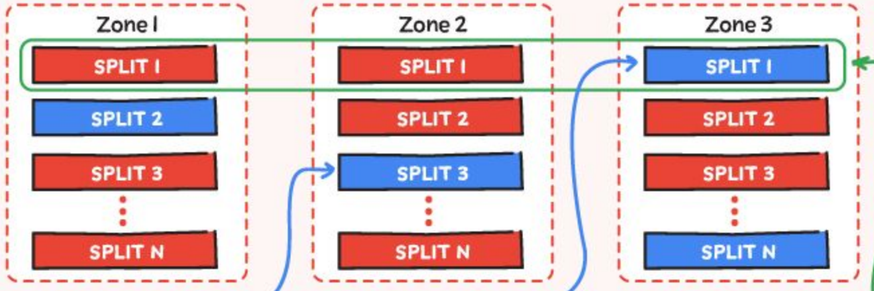
+

Google's Global Network

Fast & redundant

How does Spanner provide high availability* & scalability*?

Zero downtime for planned maintenance or schema changes



(1) **WHAT IS A SPLIT?**

Each table in the database is broken down into several splits using ranges of the primary key

SPLIT ID	0	1	...	n
KEY RANGE	[-∞,3]	[4,224]	...	[2457,∞]

(2) Splits are re-balanced dynamically based on amount of data + load

(3) Paxos group for split 1

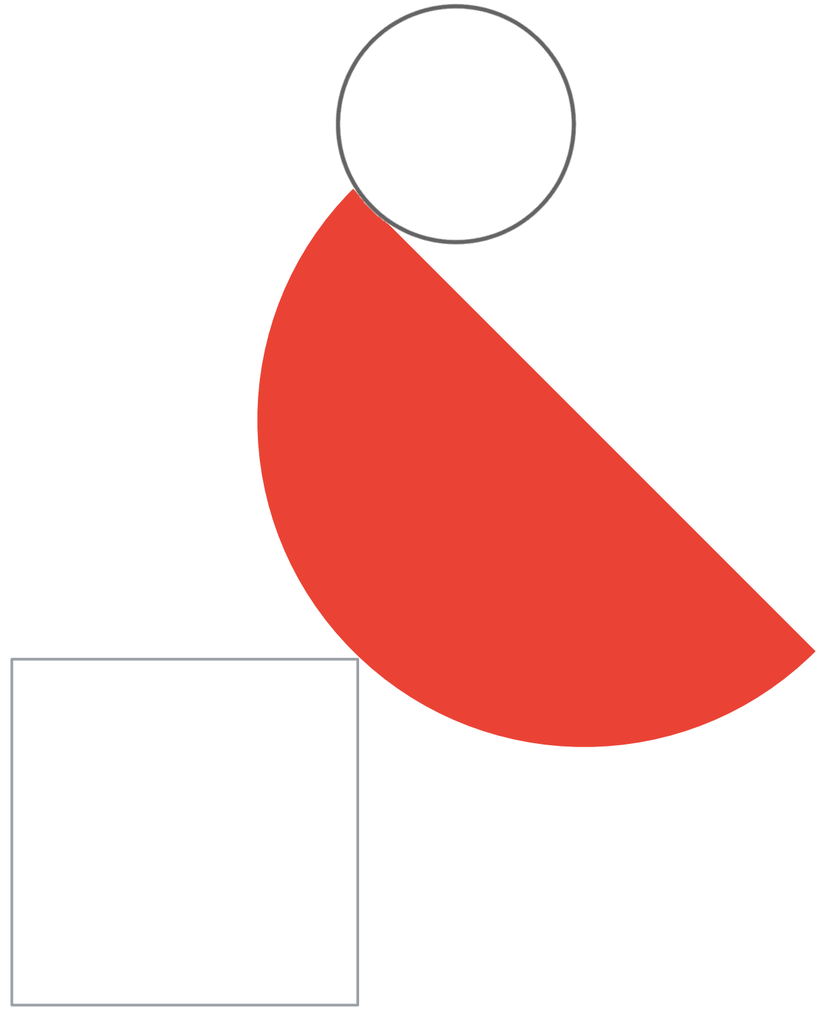
(4) Replicas participate in voting for writes to achieve a majority quorum (2 out of 3) and serve reads

(5) Different splits can have leaders in different zones

Leader manages writes for that split

All replicas serve reads

[optional] Links to useful materials



Optional materials 1

[READING]

- Boto configuration file | Cloud Storage
- Use customer-supplied encryption keys | Cloud Storage
- Object change notification | Cloud Storage
- Read about Database Migration Service.
- What are the options for connecting to Cloud SQL instance:
  a. https://cloud.google.com/sql/docs/mysql/connect-overview
  b. https://cloud.google.com/sql/docs/postgres/external-connection-methods
- How to choose optimal AI/ML path in GCP?

[VIDEOS]

- Google Cloud Storage options: Difference between object store, block store and file store | Google Cloud Storage options
- GCS Offline Transfer Appliance: Introducing Google Cloud's Transfer Appliance
- How to transfer data to GCS: How to transfer data to Google Cloud? #GCPSketchnote
- Authentication controls for Cloud Storage
- What's new with Cloud SQL

Optional materials 2

- [IMPORTANT TO KNOW] Different patterns for connecting to Cloud SQL: Cloud SQL: Concepts of Networking
- Great demo of how to centralize network management and set up Shared VPC in GCP: Level Up From Zero Episode 4: Shared VPC
- Accelerating cloud migrations with managed databases: Accelerating cloud migration with managed databases
- [Highly recommended] Choose your database on Google Cloud: Choose your database on Google Cloud
- Introducing Database Migration Service: Introducing Database Migration Service
- How to achieve high resiliency and availability with GCP: How to achieve high resiliency and availability with Google Cloud infrastructure
- Deploying MongoDB via GCP Marketplace: Deploying MongoDB from Google Cloud Marketplace

Optional materials 3

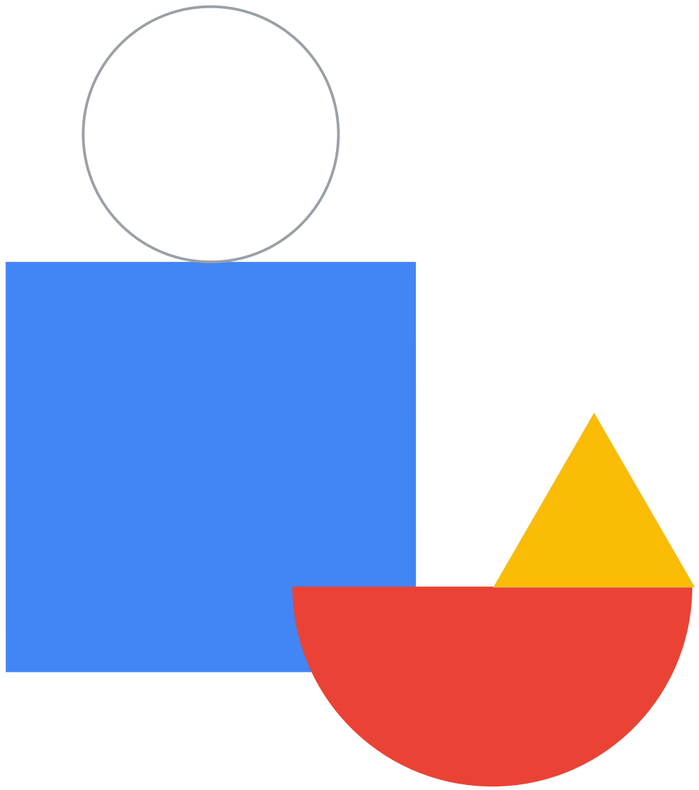
[PODCASTS]

- Cloud SQL
- Database Migration Service
- Beam and Spark

[DEEP DIVES]

- The battle of relational and non-relational databases | SQL vs NoSQL Explained: The battle of relational and non-relational databases | SQL vs NoSQL Explained
- [video] How to accelerate migration to GCP: Tools and services to accelerate your migration to Google Cloud
- 5 ways Google can help you succeed in the multicloud world.

Google Cloud

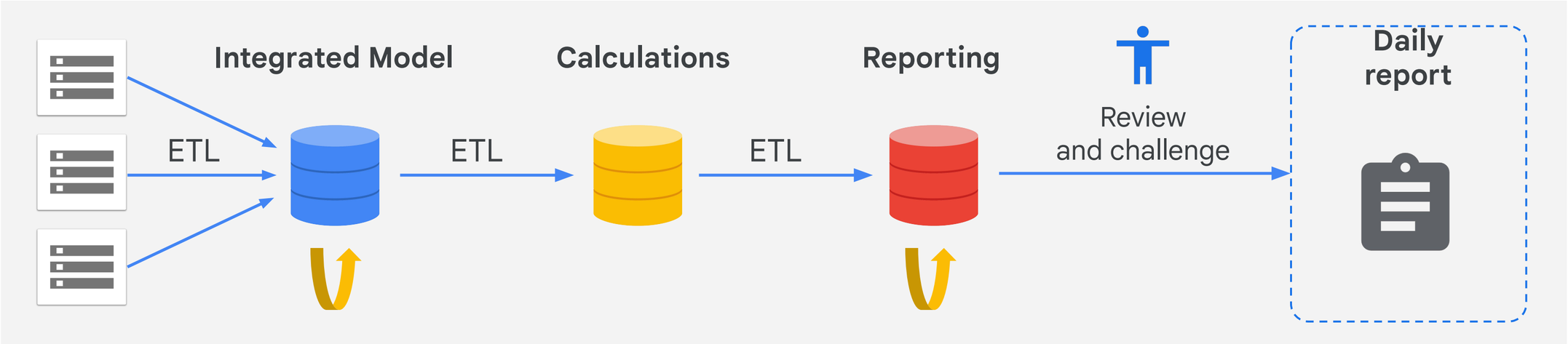


Case Study - I

Data engineer case study 01:

A customer had this interesting business requirement...

A daily reporting pipeline with multiple sources and complex dependencies
Human intervention to data check quality, inputs, and proceed to next stage
Need daily updates on yesterday's data, but takes >24 hours to run.



Data engineer case study 01:

We mapped that to technical requirements like this...

BigQuery and Cloud Composer (aka Apache Airflow)

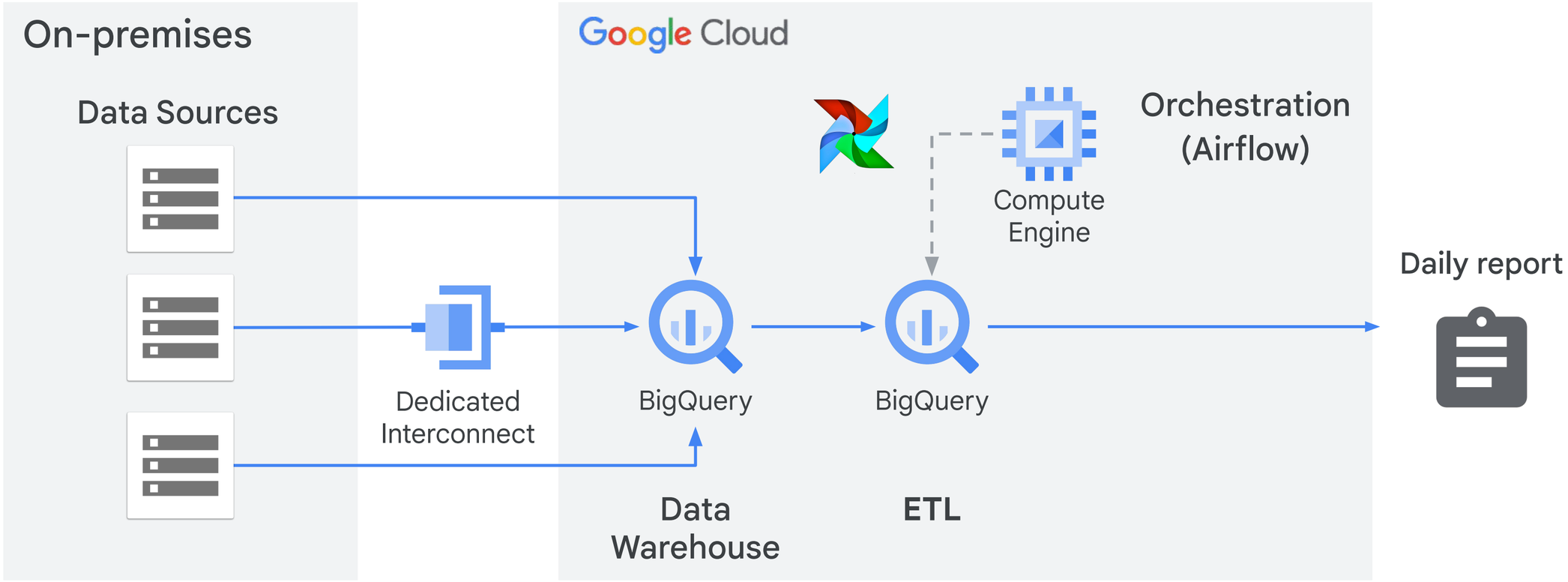
BigQuery: Reduce overall time to run with BigQuery as data warehouse and analytics engine.

Apache Airflow: Control to automate pipeline, handle dependencies as code, start query when preceding queries were done.

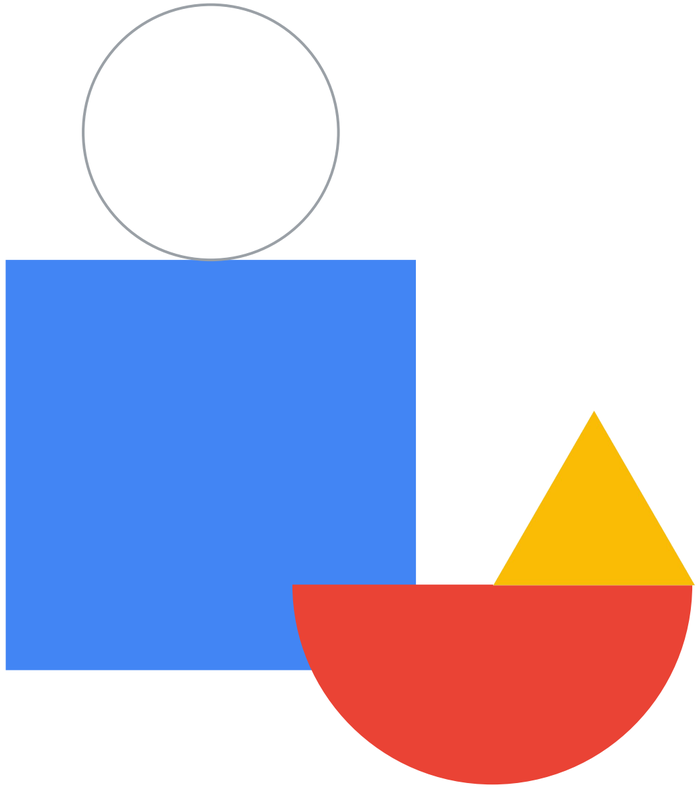
Data engineer case study 01:

And this is how we implemented that technical requirement

Common data warehouse in BigQuery. Apache Airflow to automate query dependencies.



Google Cloud

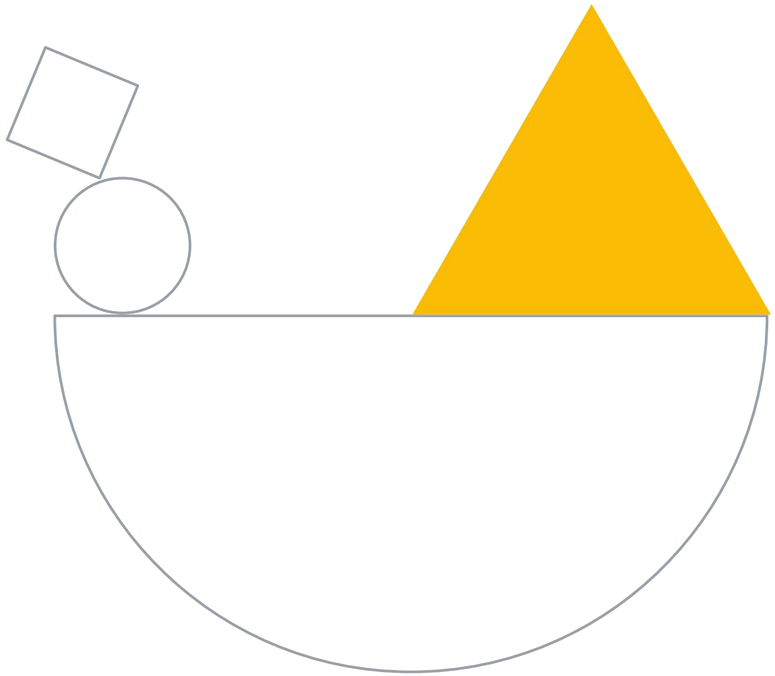


Case Study - II

A customer had this interesting business requirement...

- Capture data reading and updates events to know who, what, when, and where.
- Separation of who manages the data and who can read the data.
- Allocate costs appropriately; costs to read/process vs. costs to store.
- Prevent exfiltration of data to other Google Cloud projects and to external systems.

Diagnostic questions



Please complete the diagnostic questions now

- Forms are provided for you to answer the diagnostic questions.
- The instructor will provide you a link to the forms.
- The diagnostic questions are also available in the workbook.

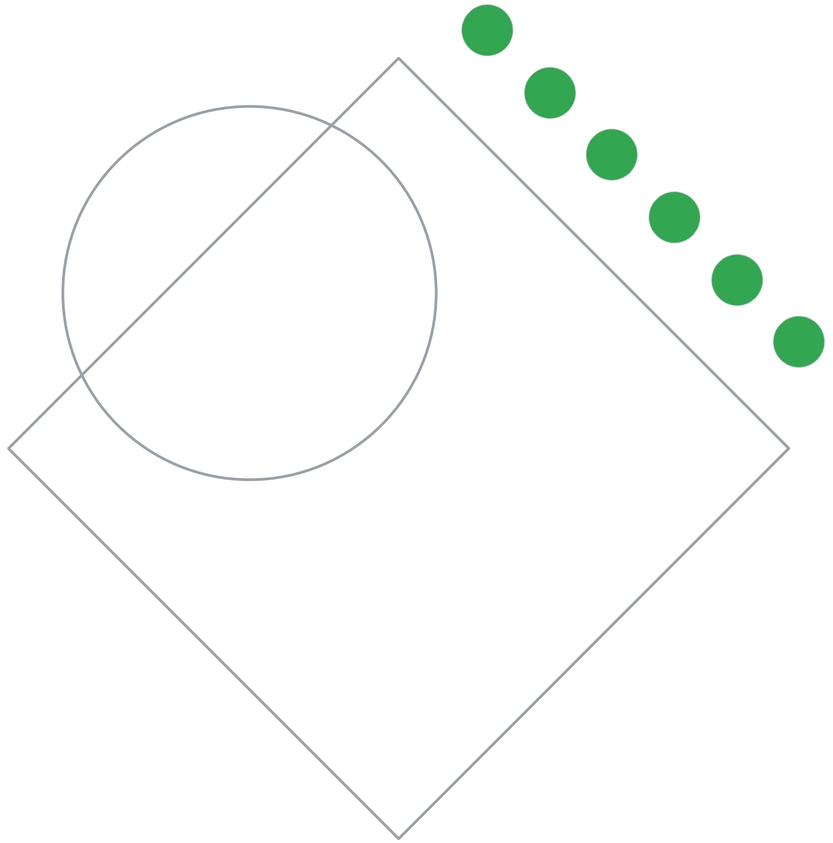
Module agenda

01 Storing Cymbal Retail's data

02 Diagnostic questions

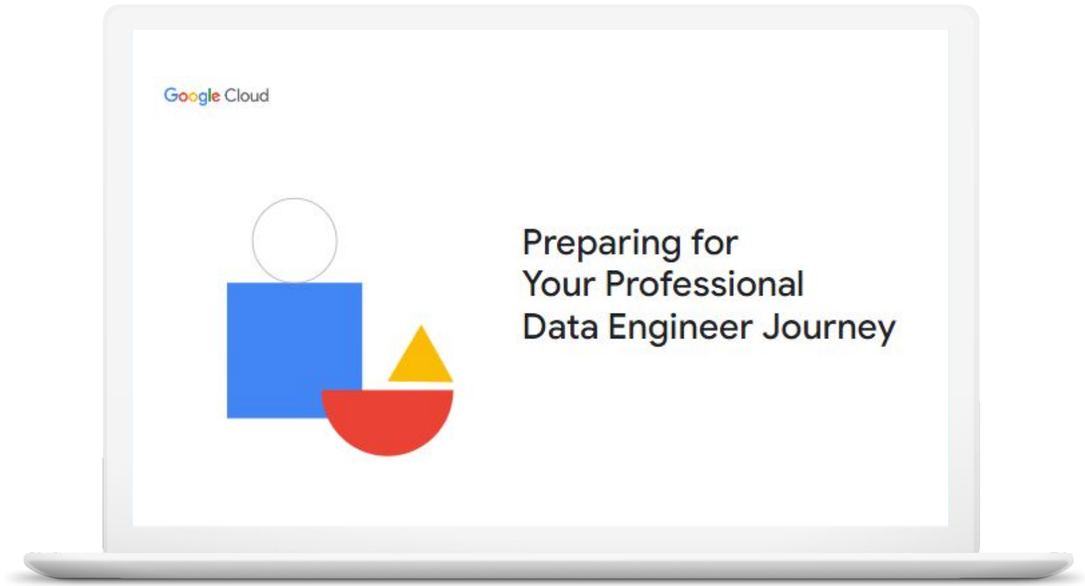
03 **Review and study planning**

Review and study planning



Your study plan:

Storing the data



3.1 Selecting storage systems

3.2 Planning for using a data warehouse

3.3 Using a data lake

3.4 Designing for a data mesh

3.1 Selecting storage systems

Considerations include:

- Analyzing data access patterns
- Choosing managed services (e.g., Bigtable, Cloud Spanner, Cloud SQL, Cloud Storage, Firestore, Memorystore)
- Planning for storage costs and performance
- Lifecycle management of data

3.1 Diagnostic Question 01 Discussion

You need to choose a data storage solution to support a transactional system. Your customers are primarily based in one region. You want to reduce your administration tasks and focus engineering effort on building your business application.

What should you do?

A. Use Cloud Spanner.

B. Use Cloud SQL.

C. Install a database of your choice on a Compute Engine VM.

D. Create a Cloud Storage bucket with a regional bucket.

3.1 Diagnostic Question 01 Discussion

You need to choose a data storage solution to support a **transactional** system. Your customers are primarily based in **one region**. You want to **reduce your administration tasks and focus engineering effort** on building your business application.

What should you do?

A. Use Cloud Spanner.

B. Use Cloud SQL.

C. Install a database of your choice on a Compute Engine VM.

D. Create a Cloud Storage bucket with a regional bucket.

3.1 Diagnostic Question 01 Discussion

You need to choose a data storage solution to support a **transactional** system. Your customers are primarily based in **one region**. You want to **reduce your administration tasks and focus engineering effort** on building your business application.

What should you do?

A. Use Cloud Spanner.

B. Use Cloud SQL.

C. Install a database of your choice on a Compute Engine VM.

D. Create a Cloud Storage bucket with a regional bucket.

3.1 Diagnostic Question 02 Discussion

You need to store data long term and use it to create quarterly reports.

What storage class should you choose?

A. Standard storage class
B. Nearline storage class
C. Coldline storage class
D. Archive storage class

3.1 Diagnostic Question 02 Discussion

You need to store data long term and use it to create **quarterly** reports.

What storage class should you choose?

A. Standard storage class
B. Nearline storage class
C. Coldline storage class
D. Archive storage class

3.1 Diagnostic Question 02 Discussion

You need to store data long term and use it to create **quarterly** reports.

What storage class should you choose?

A. Standard storage class
B. Nearline storage class
C. Coldline storage class
D. Archive storage class

Link

3.1 Selecting storage systems

Courses

Google Cloud Big Data and Machine Learning Fundamentals

- Big Data and Machine Learning on Google Cloud

Modernizing Data Lakes and Data Warehouses on Google Cloud

- Introduction to data engineering
- Building a data lake
- Building a data warehouse

Building Resilient Streaming Analytics Systems on Google Cloud

- High-Throughput BigQuery and Bigtable Streaming Features

Documentation

Cloud SQL for MySQL, PostgreSQL, and SQL Server

What is Cloud SQL?

Storage classes | Google Cloud

3.2 Planning for using a data warehouse

Considerations include:

- Designing the data model
- Deciding the degree of data normalization
- Mapping business requirements
- Defining architecture to support data access patterns

3.2 Diagnostic Question 03 Discussion

You have several large tables in your transaction databases. You need to move all the data to BigQuery for the business analysts to explore and analyze the data.

How should you design the schema in BigQuery?

A. Retain the data on BigQuery with the same schema as the source.
B. Combine all the transactional database tables into a single table using outer joins.
C. Redesign the schema to normalize the data by removing all redundancies.
D. Redesign the schema to denormalize the data with nested and repeated data.

3.2 Diagnostic Question 03 Discussion

You have several **large tables** in your transaction databases. You need to **move all the data to BigQuery** for the business analysts to explore and analyze the data.

How should you design the schema in BigQuery?

A. Retain the data on BigQuery with the same schema as the source.
B. Combine all the transactional database tables into a single table using outer joins.
C. Redesign the schema to normalize the data by removing all redundancies.
D. Redesign the schema to denormalize the data with nested and repeated data.

3.2 Diagnostic Question 03 Discussion

You have several **large tables** in your transaction databases. You need to **move all the data to BigQuery** for the business analysts to explore and analyze the data.

How should you design the schema in BigQuery?

A. Retain the data on BigQuery with the same schema as the source.
B. Combine all the transactional database tables into a single table using outer joins.
C. Redesign the schema to normalize the data by removing all redundancies.
D. Redesign the schema to denormalize the data with nested and repeated data.

Link

3.2 Diagnostic Question 04 Discussion

You are ingesting data that is spread out over a wide range of dates into BigQuery at a fast rate. You need to partition the table to make queries performant.

What should you do?

A. Create an ingestion-time partitioned table with daily partitioning type.

B. Create an ingestion-time partitioned table with yearly partitioning type.

C. Create an integer-range partitioned table.

D. Create a time-unit column-partitioned table with yearly partitioning type.

3.2 Diagnostic Question 04 Discussion

You are ingesting data that is **spread out over a wide range of dates** into BigQuery at a fast rate. You need to partition the table to make queries performant.

What should you do?

A. Create an ingestion-time partitioned table with daily partitioning type.

B. Create an ingestion-time partitioned table with yearly partitioning type.

C. Create an integer-range partitioned table.

D. Create a time-unit column-partitioned table with yearly partitioning type.

3.2 Diagnostic Question 04 Discussion

You are ingesting data that is **spread out over a wide range of dates** into BigQuery at a fast rate. You need to partition the table to make queries performant.

What should you do?

A. Create an ingestion-time partitioned table with daily partitioning type.

B. Create an ingestion-time partitioned table with yearly partitioning type.

C. Create an integer-range partitioned table.

D. Create a time-unit column-partitioned table with yearly partitioning type.

Link 1
Link 2

3.2 Diagnostic Question 05 Discussion

Your analysts repeatedly run the same complex queries that combine and filter through a lot of data on BigQuery. The data changes frequently. You need to reduce the effort for the analysts.

What should you do?

A. Create a dataset with the data that is frequently queried.

B. Create a view of the frequently queried data.

C. Export the frequently queried data into a new table.

D. Export the frequently queried data into Cloud SQL.

3.2 Diagnostic Question 05 Discussion

Your analysts **repeatedly run the same complex queries** that combine and filter through a lot of data on BigQuery. The **data changes frequently.** You need to reduce the effort for the analysts.

A. Create a dataset with the data that is frequently queried.

B. Create a view of the frequently queried data.

C. Export the frequently queried data into a new table.

D. Export the frequently queried data into Cloud SQL.

What should you do?

3.2 Diagnostic Question 05 Discussion

Your analysts **repeatedly run the same complex queries** that combine and filter through a lot of data on BigQuery. The **data changes frequently.** You need to reduce the effort for the analysts.

What should you do?

A. Create a dataset with the data that is frequently queried.

B. Create a view of the frequently queried data.

C. Export the frequently queried data into a new table.

D. Export the frequently queried data into Cloud SQL.

3.2 Planning for using a data warehouse

Courses

[Modernizing Data Lakes and Data Warehouses on Google Cloud]

- Building a data warehouse

[Building Resilient Streaming Analytics Systems on Google Cloud]

- Advanced BigQuery functionality and performance

Skill Badges

[Build and Optimize Data Warehouses with BigQuery]

Documentation

[Introduction to optimizing query performance | BigQuery | Google Cloud]

[Introduction to partitioned tables | BigQuery | Google Cloud]

[Creating partitioned tables | BigQuery | Google Cloud]

[Introduction to views | BigQuery | Google Cloud]

3.3 Using a data lake

Considerations include:

- Managing the lake (configuring data discovery, access, and cost controls)
- Processing data
- Monitoring the data lake

3.3 Diagnostic Question 06 Discussion

You have data that is ingested daily and frequently analyzed in the first month. Thereafter, the data is retained only for audits, which happen occasionally every few years. You need to configure cost-effective storage.

What should you do?

A. Create a bucket on Cloud Storage with object versioning configured.
B. Create a bucket on Cloud Storage with Autoclass configured.
C. Configure a data retention policy on Cloud Storage.
D. Configure a lifecycle policy on Cloud Storage.

3.3 Diagnostic Question 06 Discussion

You have data that is **ingested daily and frequently analyzed** in the first month. Thereafter, the data is **retained only for audits,** which happen occasionally every few years. You need to configure **cost-effective storage**.

What should you do?

A. Create a bucket on Cloud Storage with object versioning configured.

B. Create a bucket on Cloud Storage with Autoclass configured.

C. Configure a data retention policy on Cloud Storage.

D. Configure a lifecycle policy on Cloud Storage.

3.3 Diagnostic Question 06 Discussion

You have data that is **ingested daily and frequently analyzed** in the first month. Thereafter, the data is **retained only for audits,** which happen occasionally every few years. You need to configure **cost-effective storage**.

What should you do?

A. Create a bucket on Cloud Storage with object versioning configured.

B. Create a bucket on Cloud Storage with Autoclass configured.

C. Configure a data retention policy on Cloud Storage.

D. Configure a lifecycle policy on Cloud Storage.

3.3 Diagnostic Question 07 Discussion

You have data stored in a Cloud Storage bucket. You are using both Identity and Access Management (IAM) and Access Control Lists (ACLs) to configure access control. Which statement describes a user's access to objects in the bucket?

A. The user has no access if IAM denies the permission.

B. The user only has access if both IAM and ACLs grant a permission.

C. The user has access if either IAM or ACLs grant a permission.

D. The user has no access if either IAM or ACLs deny a permission.

Which statement describes a user's access to objects in the bucket?

3.3 Diagnostic Question 07 Discussion

You have data stored in a Cloud Storage bucket. You are **using both Identity and Access Management (IAM) and Access Control Lists (ACLs) to configure access control**. Which statement describes a user's access to objects in the bucket?

Which statement describes a user's access to objects in the bucket?

A. The user has no access if IAM denies the permission.

B. The user only has access if both IAM and ACLs grant a permission.

C. The user has access if either IAM or ACLs grant a permission.

D. The user has no access if either IAM or ACLs deny a permission.

3.3 Diagnostic Question 07 Discussion

You have data stored in a Cloud Storage bucket. You are **using both Identity and Access Management (IAM) and Access Control Lists (ACLs) to configure access control**. Which statement describes a user's access to objects in the bucket?

Which statement describes a user's access to objects in the bucket?

A. The user has no access if IAM denies the permission.

B. The user only has access if both IAM and ACLs grant a permission.

C. The user has access if either IAM or ACLs grant a permission.

D. The user has no access if either IAM or ACLs deny a permission.

3.3 Diagnostic Question 08 Discussion

A manager at Cymbal Retail expresses concern about unauthorized access to objects in your Cloud Storage bucket. You need to evaluate all access on all objects in the bucket.

What should you do?

A. Review the Admin Activity audit logs.

B. Enable and then review the Data Access audit logs.

C. Route the Admin Activity logs to a BigQuery sink and analyze the logs with SQL queries.

D. Change the permissions on the bucket to only trusted employees.

3.3 Diagnostic Question 08 Discussion

A manager at Cymbal Retail expresses **concern about unauthorized access to objects in your Cloud Storage bucket.** You need to evaluate all access on all objects in the bucket.

What should you do?

A. Review the Admin Activity audit logs.

B. Enable and then review the Data Access audit logs.

C. Route the Admin Activity logs to a BigQuery sink and analyze the logs with SQL queries.

D. Change the permissions on the bucket to only trusted employees.

Types of entries in Cloud Storage audit logs

Admin Activity logs

- Modify configuration of project, bucket or object
- Creating and deleting buckets

Data Access logs

- Admin_read
  - Listing buckets and bucket information
- Data_read
  - Listing object data and object information
- Data_write
  - Creating and deleting objects

3.3 Diagnostic Question 08 Discussion

A manager at Cymbal Retail expresses **concern about unauthorized access to objects in your Cloud Storage bucket.** You need to evaluate all access on all objects in the bucket.

What should you do?

A. Review the Admin Activity audit logs.

B. Enable and then review the Data Access audit logs.

C. Route the Admin Activity logs to a BigQuery sink and analyze the logs with SQL queries.

D. Change the permissions on the bucket to only trusted employees.

3.3 Using a data lake

Courses

Modernizing Data Lakes and Data Warehouses on Google Cloud
- Building a data lake

Documentation

Cloud Storage

Object Lifecycle Management | Cloud Storage

Overview of access control | Cloud Storage

Cloud Audit Logs with Cloud Storage | Google Cloud

3.4 Designing for a data mesh

Considerations include:

- Building a data mesh based on requirements by using Google Cloud tools (e.g., Dataplex, Data Catalog, BigQuery, Cloud Storage)
- Segmenting data for distributed team usage
- Building a federated governance model for distributed data systems

3.4 Diagnostic Question 09 Discussion

Cymbal Retail has accumulated a large amount of data. Analysts and leadership are finding it difficult to understand the meaning of the data, such as BigQuery columns. Users of the data don't know who owns what. You need to improve the searchability of the data.

What should you do?

A. Create tags for data entries in Cloud Catalog.
B. Rename BigQuery columns with more descriptive names.
C. Export the data to Cloud Storage with descriptive file names.
D. Add a description column corresponding to each data column.

3.4 Diagnostic Question 09 Discussion

Cymbal Retail has accumulated a large amount of data. Analysts and leadership are finding it **difficult to understand the meaning of the data,** such as BigQuery columns. **Users of the data don't know who owns what.** You need to improve the searchability of the data.

What should you do?

A. Create tags for data entries in Cloud Catalog.
B. Rename BigQuery columns with more descriptive names.
C. Export the data to Cloud Storage with descriptive file names.
D. Add a description column corresponding to each data column.

3.4 Diagnostic Question 09 Discussion

Cymbal Retail has accumulated a large amount of data. Analysts and leadership are finding it **difficult to understand the meaning of the data,** such as BigQuery columns. **Users of the data don't know who owns what.** You need to improve the searchability of the data.

What should you do?

A. Create tags for data entries in Cloud Catalog.
B. Rename BigQuery columns with more descriptive names.
C. Export the data to Cloud Storage with descriptive file names.
D. Add a description column corresponding to each data column.

3.4 Diagnostic Question 10 Discussion

You have large amounts of data stored on Cloud Storage and BigQuery. Some of it is processed, but some is yet unprocessed. You have a data mesh created in Dataplex. You need to make it convenient for internal users of the data to discover and use the data.

What should you do?

A. Create a lake for Cloud Storage data and a zone for BigQuery data.
B. Create a lake for BigQuery data and a zone for Cloud Storage data.
C. Create a lake for unprocessed data and assets for processed data.
D. Create a raw zone for the unprocessed data and a curated zone for the processed data.

3.4 Diagnostic Question 10 Discussion

You have large amounts of data stored on Cloud Storage and BigQuery. **Some of it is processed, but some is yet unprocessed. You have a data mesh created in Dataplex.** You need to make it convenient for internal users of the data to discover and use the data.

What should you do?

A. Create a lake for Cloud Storage data and a zone for BigQuery data.

B. Create a lake for BigQuery data and a zone for Cloud Storage data.

C. Create a lake for unprocessed data and assets for processed data.

D. Create a raw zone for the unprocessed data and a curated zone for the processed data.

3.4 Diagnostic Question 10 Discussion

You have large amounts of data stored on Cloud Storage and BigQuery. **Some of it is processed, but some is yet unprocessed. You have a data mesh created in Dataplex.** You need to make it convenient for internal users of the data to discover and use the data.

What should you do?

A. Create a lake for Cloud Storage data and a zone for BigQuery data.

B. Create a lake for BigQuery data and a zone for Cloud Storage data.

C. Create a lake for unprocessed data and assets for processed data.

D. Create a raw zone for the unprocessed data and a curated zone for the processed data.

3.4 Designing for a data mesh

Courses

[Modernizing Data Lakes and Data Warehouses on Google Cloud]

- Introduction to data engineering

[Building Batch Data Pipelines on Google Cloud]

- Introduction to building batch data pipelines

Skill Badges

[Data Catalog Fundamentals]

Documentation

[Tags and tag templates | Data Catalog Documentation | Google Cloud]

[Quickstart: Tag a BigQuery table by using Data Catalog]

[Dataplex overview | Google Cloud]

Knowledge Check 1

Cymbal Retail collects large amounts of data that is useful for improving business operations. The company wants to store and analyze this data in a serverless and cost-effective manner using Google Cloud. The analysts need to use SQL to write the queries.
What tool can you use to meet these requirements?

A. Data Fusion

B. BigQuery

C. Cloud Spanner

D. Memorystore

Knowledge Check 1

Cymbal Retail collects large amounts of data that is useful for improving business operations. The company wants to store and analyze this data in a serverless and cost-effective manner using Google Cloud. The analysts need to use SQL to write the queries.
What tool can you use to meet these requirements?

A. Data Fusion

B. BigQuery

C. Cloud Spanner

D. Memorystore

Knowledge Check 2

Cymbal Retail also collects large amounts of structured, semistructured, and unstructured data. The company wants a centralized repository to store this data in a cost-effective manner using Google Cloud. What tool can you use to meet these requirements?

A. Bigtable

B. Dataflow

C. Cloud Storage

D. Cloud SQL

Knowledge Check 2

Cymbal Retail also collects large amounts of structured, semistructured, and unstructured data. The company wants a centralized repository to store this data in a cost-effective manner using Google Cloud. What tool can you use to meet these requirements?

A. Bigtable

B. Dataflow

C. Cloud Storage

D. Cloud SQL

Link