Google Cloud

Choosing Storage and Data Solutions

Learning objectives

- Choose the appropriate Google Cloud data storage service based on use case, durability, availability, scalability, and cost.
- Store binary data with Cloud Storage.
- Store relational data using Cloud SQL and Spanner.
- Store NoSQL data using Firestore and Cloud Bigtable.
- Cache data for fast access using Memorystore.
- Aggregate data for queries and reports using BigQuery as a data warehouse.

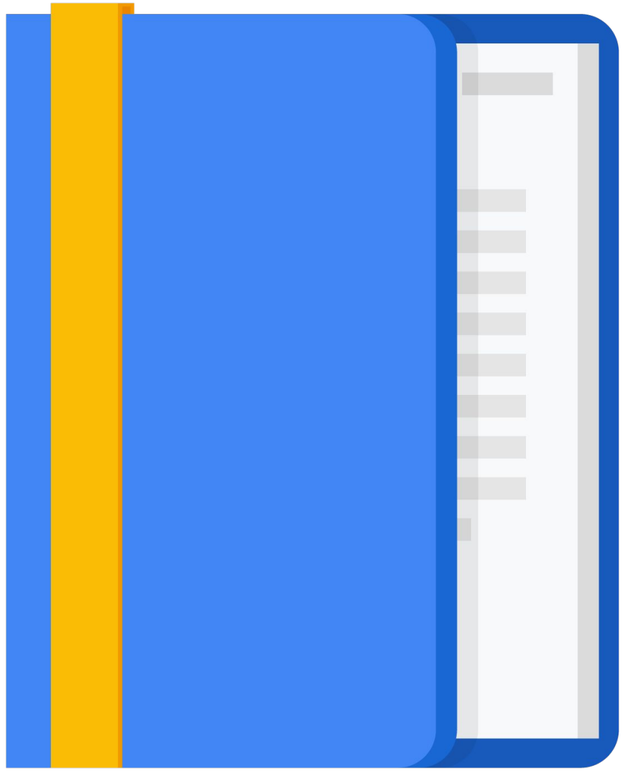
Agenda

Key Storage Characteristics

Design Activity #6

Choosing Google Cloud Storage and Data Solutions

Design Activity #7



Google Cloud–managed storage and database portfolio

Relational		NoSQL		Object	Warehouse	In memory
 Cloud SQL	 Cloud Spanner	 Firestore	 Cloud Bigtable	 Cloud Storage	 BigQuery	 Memorystore

Different data storage services have different availability SLAs

Storage Choice	Availability SLA %
Cloud Storage (multi-region bucket)	>=99.95
Cloud Storage (regional bucket)	99.9
Cloud Storage (coldline)	99.0
Spanner (multi-region)	99.999
Spanner (single region)	99.99
Firestore (multi-region)	99.999
Firestore (single region)	99.99

Durability represents the odds of losing data

Preventing data loss is a shared responsibility.

Storage Choice	Google Cloud Provides	What you should do
Cloud Storage	11 9's durability Versioning (optional)	Turn versioning on
Disks	Snapshots	Schedule snapshot jobs
Cloud SQL	Automated machine backups Point-in-time recovery Failover server (optional)	Run SQL database backups
Spanner	Automatic replication	Run export jobs
Firestore	Automatic replication	Run export jobs

The amount of data and number of reads and writes is important when selecting a data storage service

Some services scale horizontally by adding nodes.

- Bigtable
- Spanner

Some services scale vertically by making machines larger.

- Cloud SQL
- Memorystore

Some services scale automatically with no limits.

- Cloud Storage
- BigQuery
- Firestore

Do you need strong consistency?

Strongly consistent databases update all copies of data within a transaction.

Ensures everyone gets the latest copy of the data on reads.

- Storage
- Cloud SQL
- Spanner
- Firestore

Eventually consistent databases update one copy of the data and the rest asynchronously.

Can handle a large volume of writes.

- Bigtable
- Memorystore replicas

Calculate the total cost per GB when choosing a storage service

- Bigtable and Spanner would be too expensive for storing smaller amounts of data
- Firestore is less expensive per GB, but you also pay for reads and writes
- Cloud Storage is relatively cheap, but you can't run a database in storage
- BigQuery storage is relatively cheap, but doesn't provide fast access to records and you have to pay for running queries

You need to choose the right storage solutions for each of your microservices based their requirements.

Activity 6: Defining storage characteristics

Refer to your Design and Process Workbook.

- Determine the storage characteristics for each of your case-study services.

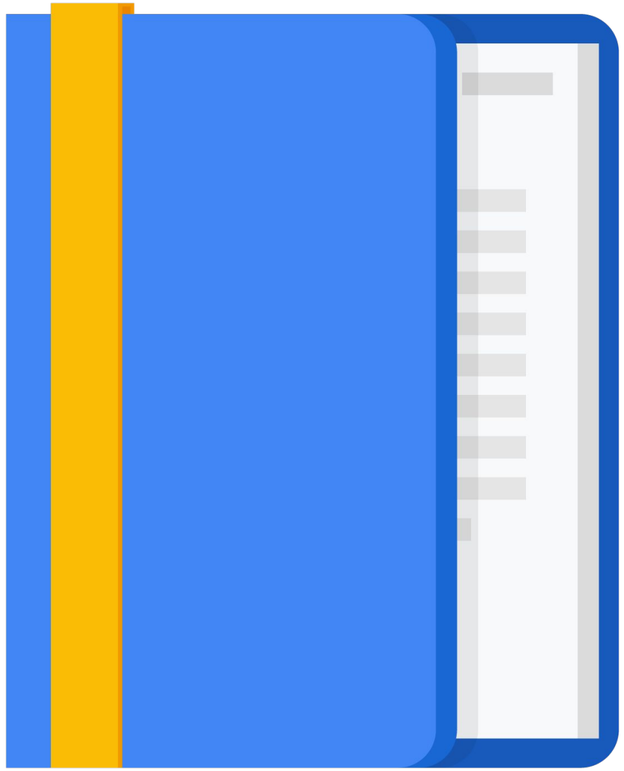
Agenda

Key Storage Characteristics

Design Activity #6

Choosing Google Cloud Storage and Data Solutions

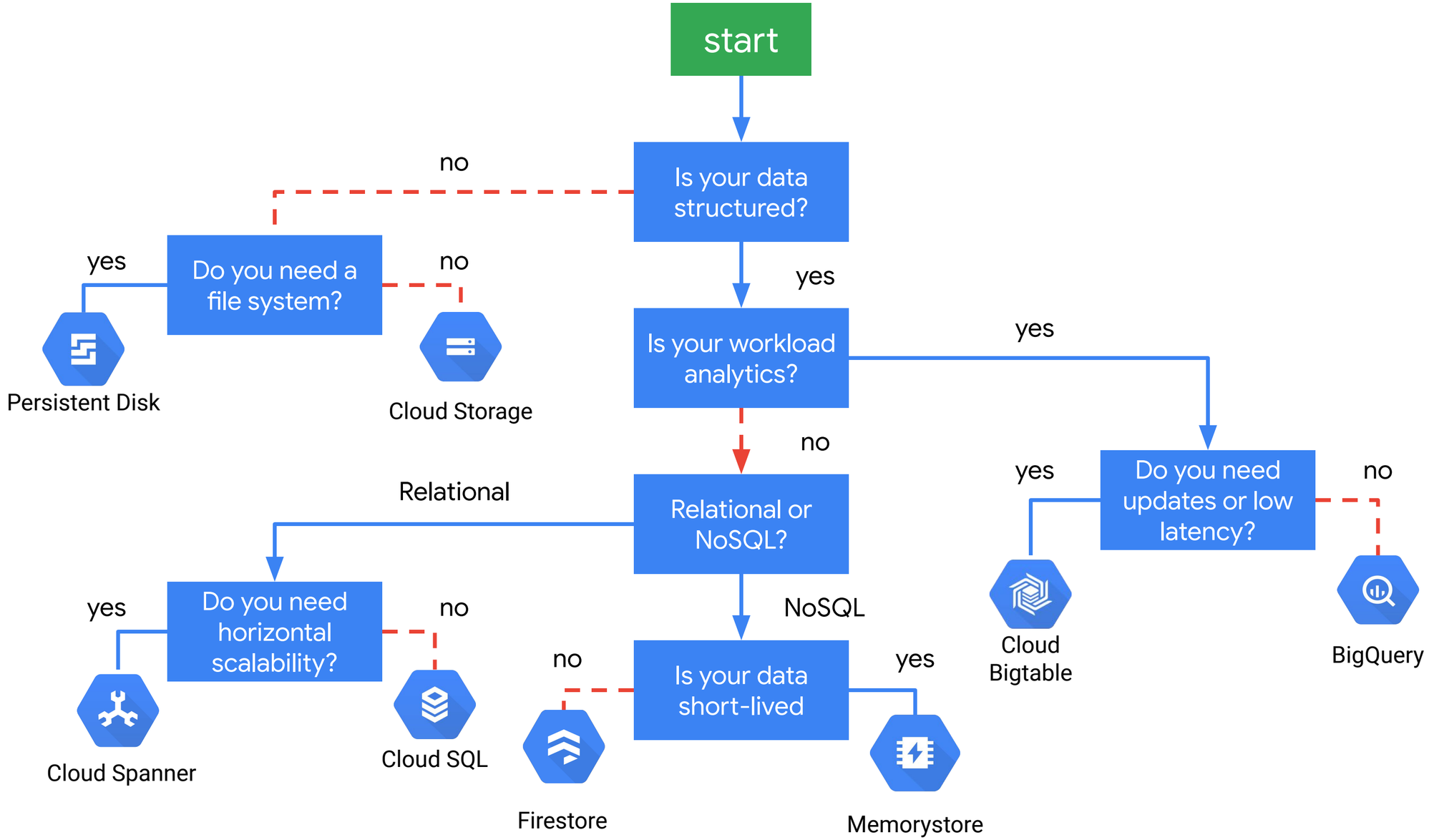
Design Activity #7



Google Cloud storage and database portfolio

Relational		NoSQL		Object	Warehouse	In memory
 Cloud SQL	 Cloud Spanner	 Firestore	 Cloud Bigtable	 Cloud Storage	 BigQuery	 Memorystore
Good for: Web frameworks	**Good for:** RDBMS+scale, HA, HTAP	**Good for:** Hierarchical, mobile, web	**Good for:** Heavy read + write, events	**Good for:** Binary object data	**Good for:** Enterprise data warehouse	**Good for:** Caching for Web/Mobile apps
Such as: CMS, eCommerce	**Such as:** User metadata, Ad/Fin/MarTech	**Such as:** User profiles, Game State	**Such as:** AdTech, financial, IoT	**Such as:** Images, media serving, backups	**Such as:** Analytics, dashboards	**Such as:** Game state, user sessions
Scales to 30 TB **MySQL** **PostgreSQL** **SQL Server**	**Scales infinitely** **Regional or multi-regional**	**Completely managed** **Document database**	**Scales infinitely** **Wide-column** **NoSQL**	**Completely managed** **Infinitely scalable**	**Completely Managed** **SQL analysis**	**Managed Redis DB**
Fixed schema	Fixed schema	Schemaless	Schemaless	Schemaless	Fixed schema	Schemaless

Google Cloud



Google Cloud

Transferring data into Google Cloud can be challenging

	1 Mbps	10 Mbps	100 Mbps	1 Gbps	10 Gbps	100 Gbps
1 GB	3 hrs	18 mins	2 mins	11 secs	1 sec	0.1 sec
10 GB	30 hrs	3 hrs	18 mins	2 mins	11 secs	1 sec
100 GB	12 days	30 hrs	3 hrs	18 mins	2 mins	11 secs
1 TB	124 days	12 days	30 hrs	3 hrs	18 mins	2 mins
10 TB	3 years	124 days	12 days	30 hrs	3 hrs	18 mins
100 TB	34 years	3 years	124 days	12 days	30 hrs	3 hrs
1 PB	340 years	34 years	3 years	124 days	12 days	30 hrs
10 PB	3,404 years	340 years	34 years	3 years	124 days	12 days
100 PB	34,048 years	3,404 years	340 years	34 years	3 years	124 days

For smaller or scheduled data uploads, use the Cloud Storage Transfer Service

Import online data to Cloud Storage

- Amazon S3
- HTTP/HTTPS Location
- Transfer data between Cloud Storage buckets

Scheduled jobs

- One time or recurring, import at a scheduled time of day
- Options for delete objects not in source or after transfer
- Filter on file name, creation date

Use the Storage Transfer Service for on-premises data for large-scale uploads from your data center

- Install on-premises agent on your servers
- Agent runs in a Docker container
- Set up a connection to Google Cloud
- Requires a minimum of 300 Mbps bandwidth

- Scales to billions of files and 100s of TBs
- Secure
- Automatic retires
- Logged
- Easy to monitor via the Cloud Console

Use Transfer Appliance for large amounts of data

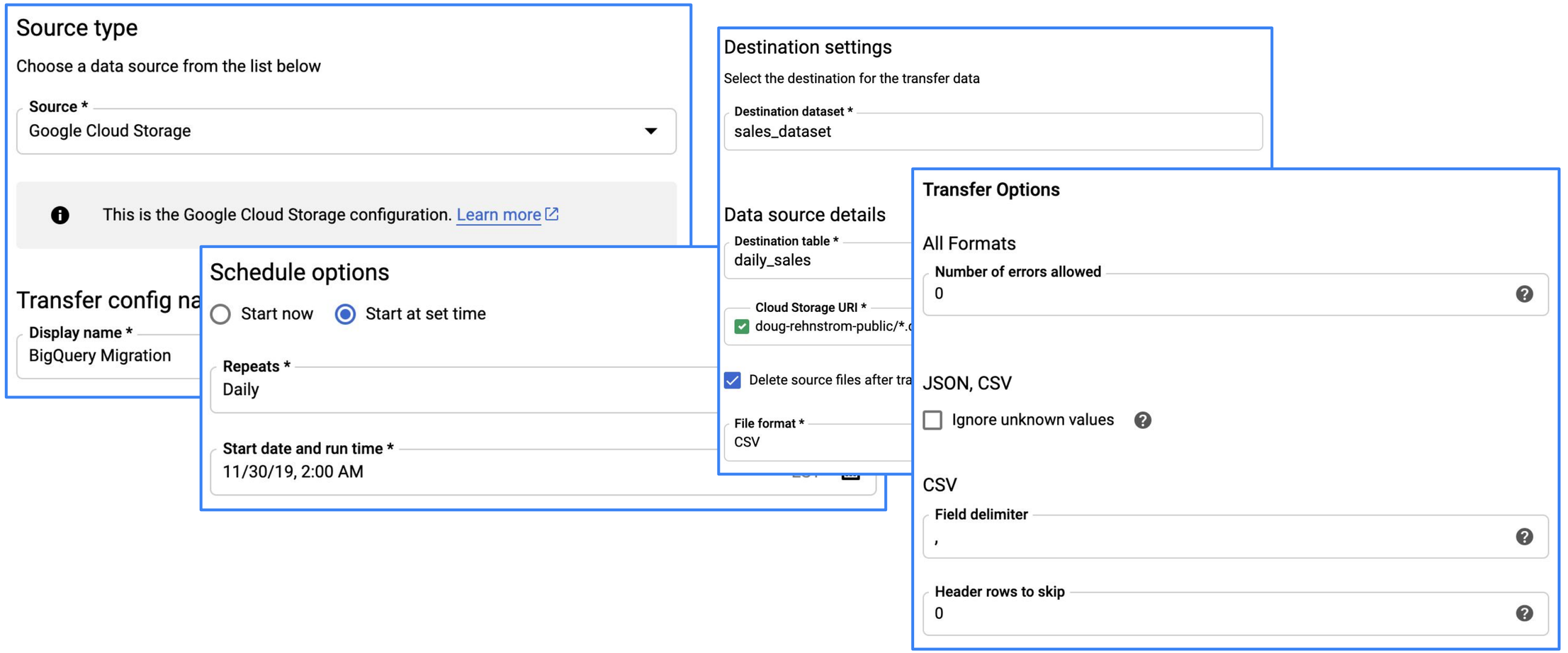
Rackable device up to 1PB shipped to Google.

Use Transfer Appliance if uploading your data would take too long.

Secure:

- You control the encryption key.
- Google securely erases the appliance after use.

There’s also a transfer service for BigQuery



Google Cloud

Activity 7: Choosing Google Cloud storage and data services

Refer to your Design and Process Workbook.

- Choose the storage services for each of your case-study services.

Quiz

Describe the difference between strong and eventual consistency.

Quiz

Describe the difference between strong and eventual consistency.

In a distributed database, there are multiple copies of the data. With strong consistency, when one copy changes, all the copies change before it is available for read. With eventual consistency, when data is updated, one copy changes and then the other copies are changed asynchronously.

Eventual consistency allows for faster writes, but it is possible for users to get stale data on a read.

Quiz

You need to store user preferences, product information, and reviews for a website you are building. There won't be a huge amount of data. What would be a simple, cost-effective, managed solution?

A. Firestore

B. Spanner

C. Cloud SQL

D. BigQuery

Quiz

You need to store user preferences, product information, and reviews for a website you are building. There won't be a huge amount of data. What would be a simple, cost-effective, managed solution?

A. Firestore

B. Spanner

C. Cloud SQL

D. BigQuery

Quiz

You are a global financial services company with users all over the world. You need a database service that can provide low latency worldwide with strong consistency. Which service might you choose?

A. Firestore

B. Spanner

C. Cloud SQL

D. BigQuery

Quiz

You are a global financial services company with users all over the world. You need a database service that can provide low latency worldwide with strong consistency. Which service might you choose?

A. Firestore

B. Spanner

C. Cloud SQL

D. BigQuery

Quiz

You want to analyze sales trends. To help achieve this, you want to combine data from your on-premises Oracle database with Google Analytics data and your web server logs. Where might you store the data so it is both easy to query and cost-effective?

A. Firestore

B. Spanner

C. Cloud SQL

D. BigQuery

Quiz

You want to analyze sales trends. To help achieve this, you want to combine data from your on-premises Oracle database with Google Analytics data and your web server logs. Where might you store the data so it is both easy to query and cost-effective?

A. Firestore

B. Spanner

C. Cloud SQL

D. BigQuery

Quiz

Currently, you are using Firestore to store information about products, reviews, and user sessions. You'd like to speed up data access in a simple, cost-effective way. What would you recommend?

A. Move the data to Spanner.

B. Move the data to BigQuery.

C. Move the data to Cloud Bigtable.

D. Cache the data using Memorystore.

Quiz

Currently, you are using Firestore to store information about products, reviews, and user sessions. You'd like to speed up data access in a simple, cost-effective way. What would you recommend?

A. Move the data to Spanner.

B. Move the data to BigQuery.

C. Move the data to Cloud Bigtable.

D. Cache the data using Memorystore.

Review

Choosing Storage and Data Solutions

More resources

Google Cloud Storage Options
https://cloud.google.com/products/storage/

Google Cloud Data Options
https://cloud.google.com/products/databases/


Google Cloud