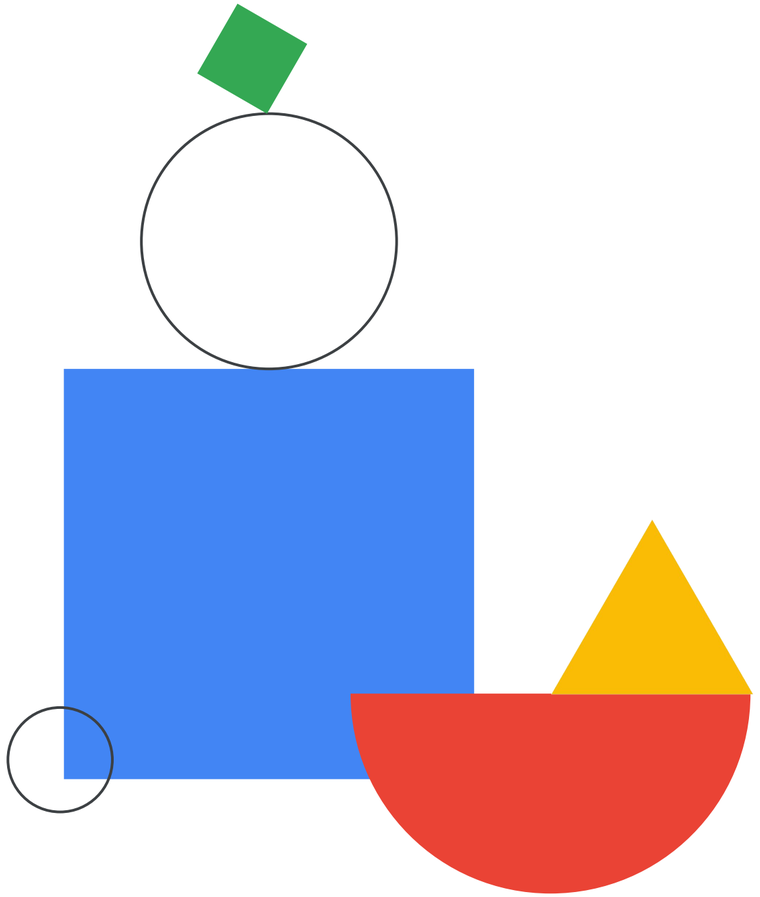
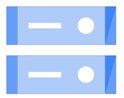
Google Cloud

Storage and Database Services



Storage and database services

Object



Cloud Storage

Good for: Binary or object data

Such as: Images, media serving, backups

File

Filestore

Good for: Network Attached Storage (NAS)

Such as: Latency sensitive workloads

Relational

Cloud SQL

Good for: Web frameworks

Such as: CMS, eCommerce

Spanner

Good for: RDBMS + scale, HA, HTAP

Such as: User metadata, Ad/Fin/ MarTech

AlloyDB

Good for: Hybrid transactional and analytical processing

Such as: Machine learning, Generative AI

Non-relational

Firestore

Good for: Hierarchical, mobile, web

Such as: User profiles, game state

Bigtable

Good for: Heavy read + write, events,

Such as: AdTech, financial, IoT

Warehouse

BigQuery

Good for: Enterprise data warehouse

Such as: Analytics, dashboards

Redis

Memorystore

Good for: Automating complex Redis and Memcached tasks

Such as: Enabling high availability, failover, patching

Scope

Infrastructure Track

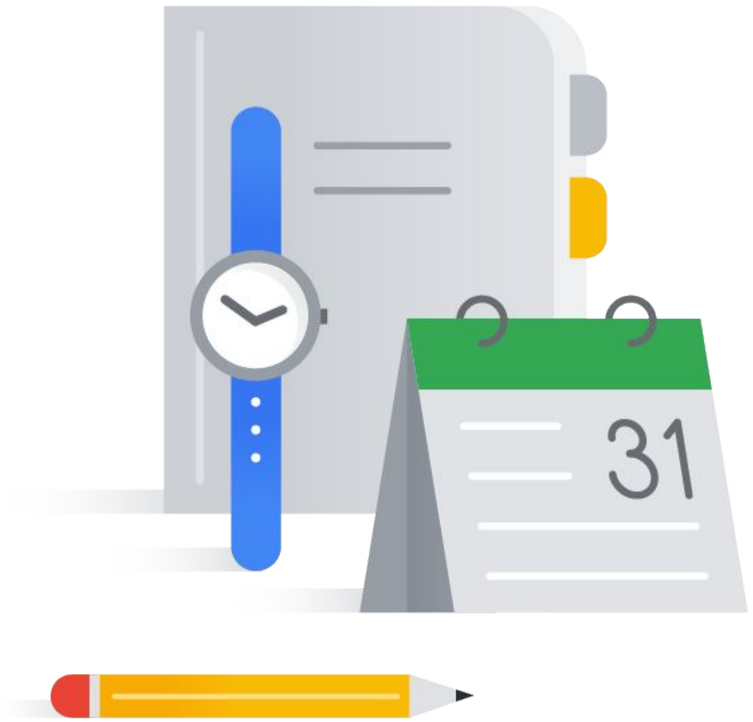
- Service differentiators
- When to consider using each service
- Set up and connect to a service

Data Engineering Track

- How to use a database system
- Design, organization, structure, schema, and use for an application
- Details about how a service stores and retrieves structured data

Agenda

01	Cloud Storage and Filestore Lab: Cloud Storage
02	Cloud SQL Lab: Implementing Cloud SQL
03	Spanner
04	AlloyDB
05	Firestore
06	Bigtable
07	Memorystore



Cloud Storage and Filestore

Cloud Storage is an object storage service

Use cases:

- Website content
- Storing data for archiving and disaster recovery
- Distributing large data objects to users via direct download

Key features:

- Scalable to exabytes
- Time to first byte in milliseconds
- Very high availability across all storage classes
- Single API across storage classes

Overview of storage classes

	Standard	Nearline	Coldline	Archive
Use case	“Hot” data and/or stored for only brief periods of time like data-intensive computations	Infrequently accessed data like data backup, long-tail multimedia content, and data archiving	Infrequently accessed data that you read or modify at most once a quarter	Data archiving, online backup, and disaster recovery
Minimum storage duration	None	30 days	90 days	365 days
Retrieval cost	None	$0.01 per GB	$0.02 per GB	$0.05 per GB
Availability SLA	99.95% (multi/dual) 99.90% (region)	99.90% (multi/dual) 99.00% (region)		99.90% (multi/dual) 99.00% (region)
Durability	99.999999999%			

Cloud Storage overview

Buckets

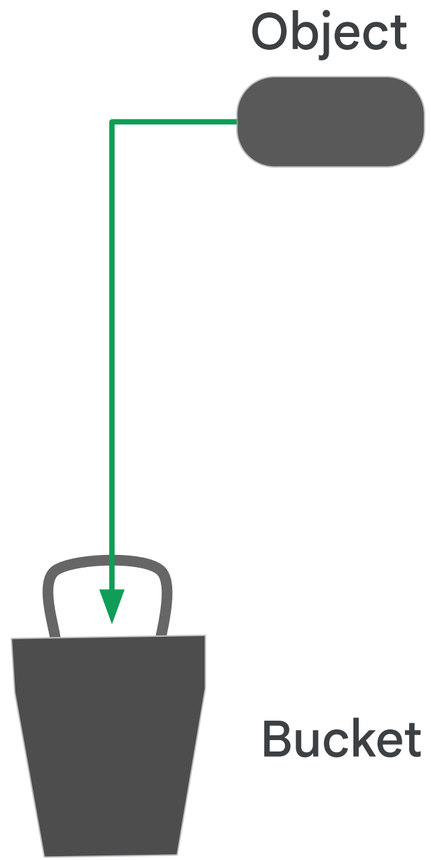
- Naming requirements
- Cannot be nested

Objects

- Inherit storage class of bucket when created
- No minimum size; unlimited storage

Access

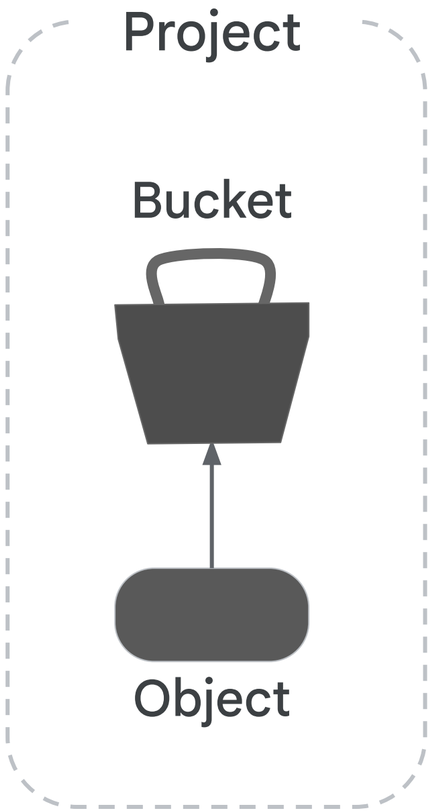
- `gcloud storage` command
- (RESTful) JSON API or XML API



Changing default storage classes

- Default class is applied to new objects
- Regional bucket can never be changed to Multi-Region/Dual-Region
- Multi-Regional bucket can never be changed to Regional
- Objects can be moved from bucket to bucket
- Object Lifecycle Management can manage the classes of objects

Access control



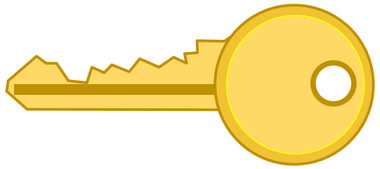
IAM

Identity and Access Management

ACLs

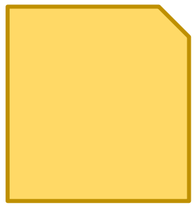
Access Control Lists

Signed URL



Signed and Timed Cryptographic Key

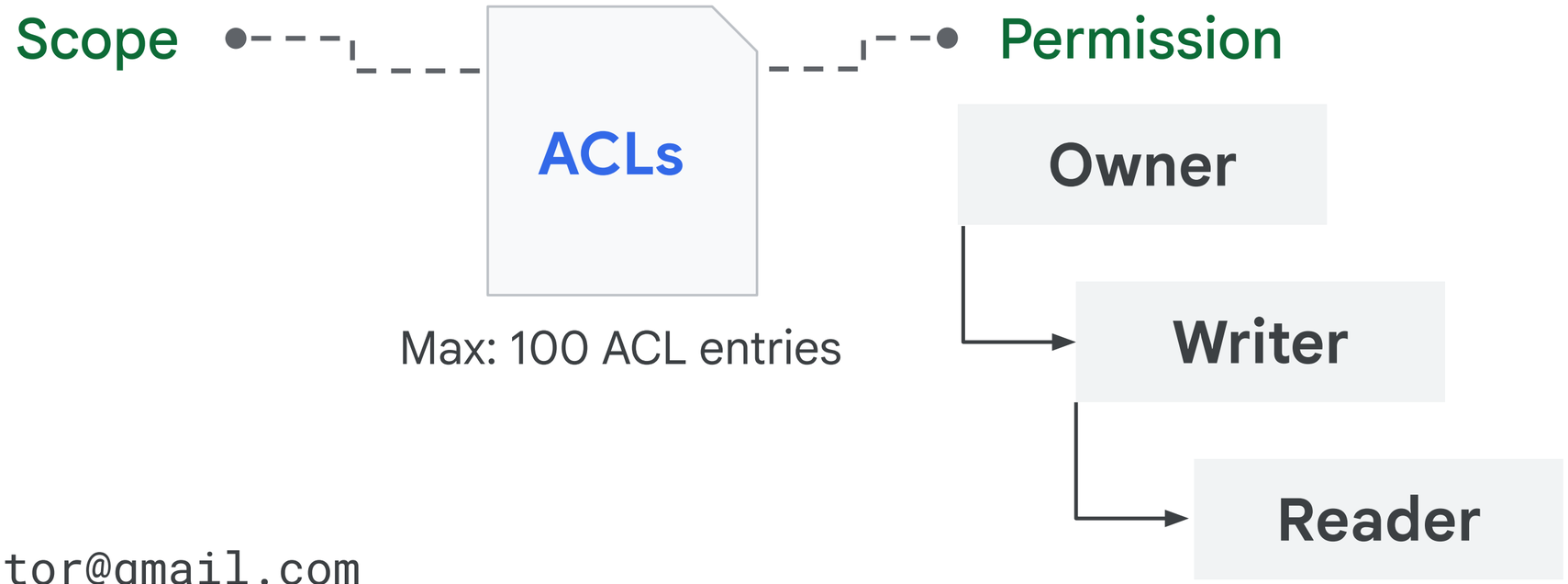
Signed Policy Document



Controls File Upload Policy

Can be used together

Access control lists (ACLs)



Examples:

- `collaborator@gmail.com`
- `allUsers`
- `allAuthenticatedUsers`

Signed URLs

- “Valet key” access to buckets and objects via ticket:
  - Ticket is a cryptographically signed URL
  - Time-limited
  - Operations specified in ticket: HTTP GET, PUT, DELETE (not POST)
  - Any user with URL can invoke permitted operations
- Example using private account key and `gcloud storage`:

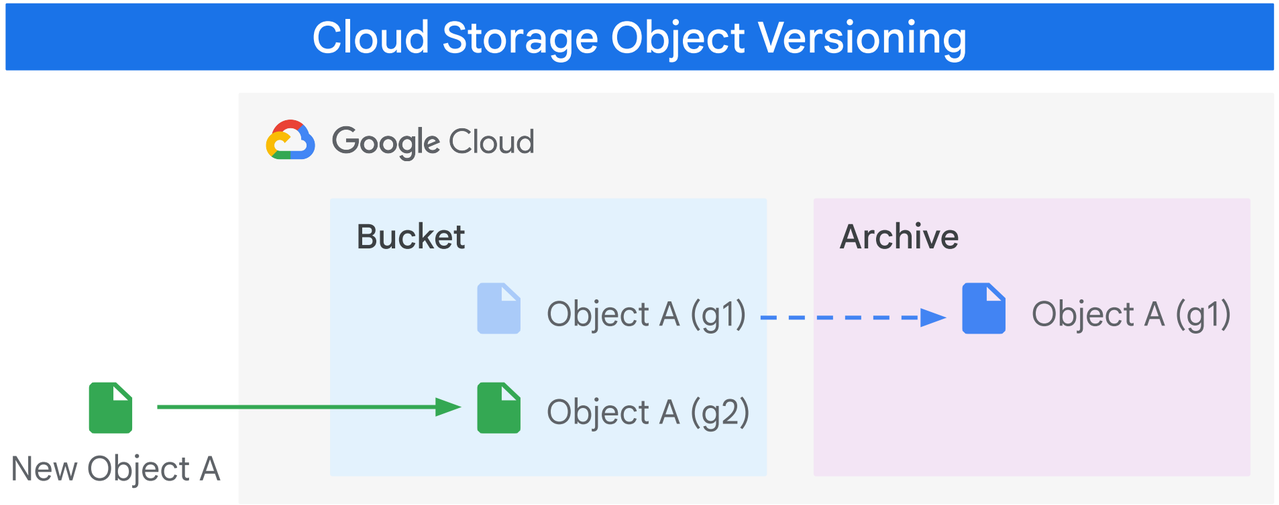
```
gcloud storage signurl -d 10m path/to/privatekey.p12
gs://bucket/object
```

Cloud Storage features

- Customer-supplied encryption key (CSEK)
  - Use your own key instead of Google-managed keys
- Object Lifecycle Management
  - Automatically delete or archive objects
- Object Versioning
  - Maintain multiple versions of objects
- Directory synchronization
  - Synchronizes a VM directory with a bucket
- Object change notifications using Pub/Sub
- Autoclass

Object Versioning supports the retrieval of objects that are deleted or overwritten

- Objects are immutable.
- Object Versioning:
  - Maintain a history of modifications of objects.
  - List archived versions of an object, restore an object to an older state, or delete a version.



Soft Delete overview

- Provides default bucket-level protection from:
  - Accidental deletion
  - Malicious deletion
- Retains overwritten or changed data.
- Is enabled by default with a 7 day retention duration.

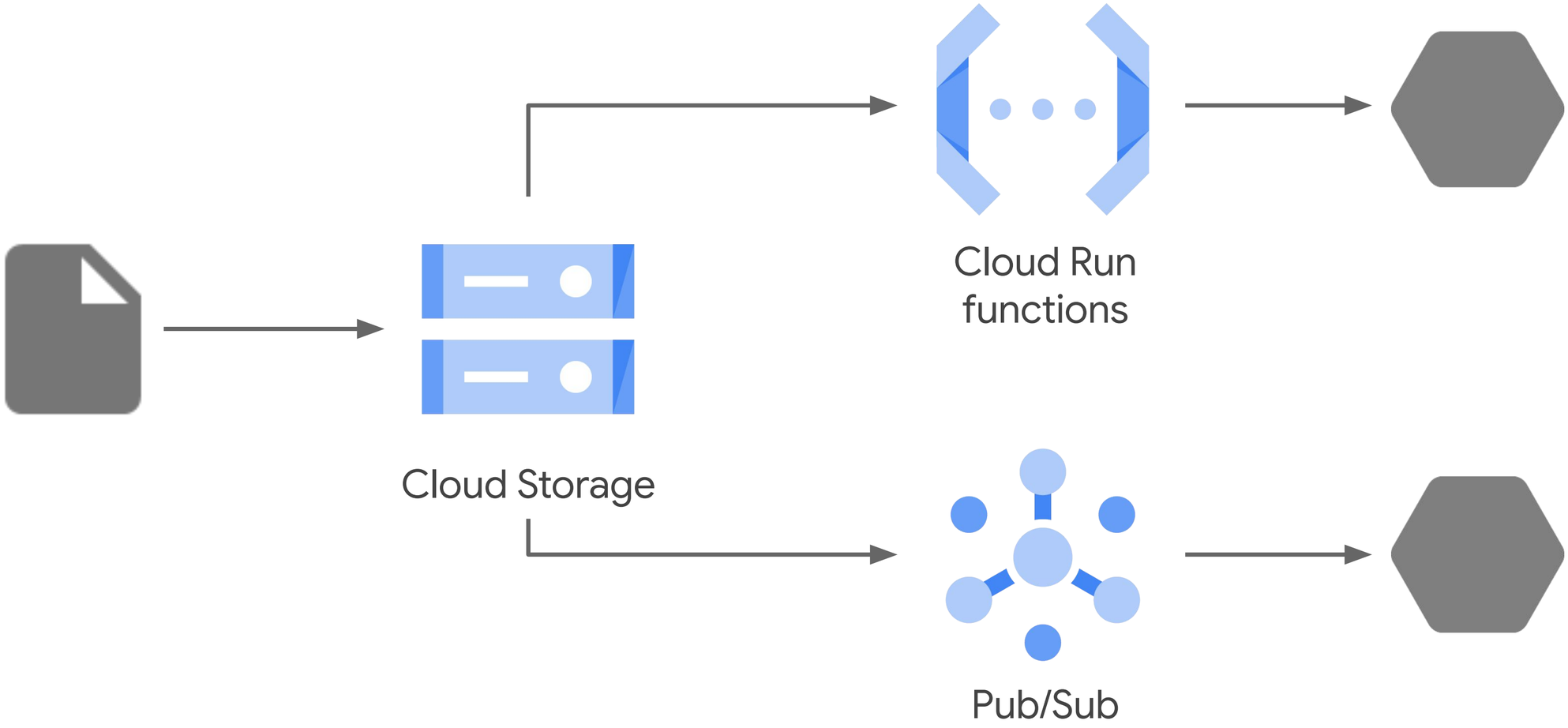
Object Lifecycle Management policies specify actions to be performed on objects that meet certain rules

- Assign a lifecycle management configuration to a bucket.
- Example use cases:
  - Downgrade storage class on objects older than a year.
  - Delete objects created before a specific date.
  - Keep only the 3 most recent versions of an object.
- Object inspection occurs in asynchronous batches.
- Changes can take 24 hours to apply.

Object Retention Lock

- Lets you define data retention requirements on a per-object basis.
- Retention configuration governs how long the object must be retained.
- Helps with regulatory and compliance requirements.

Pub/Sub notifications for Cloud Storage



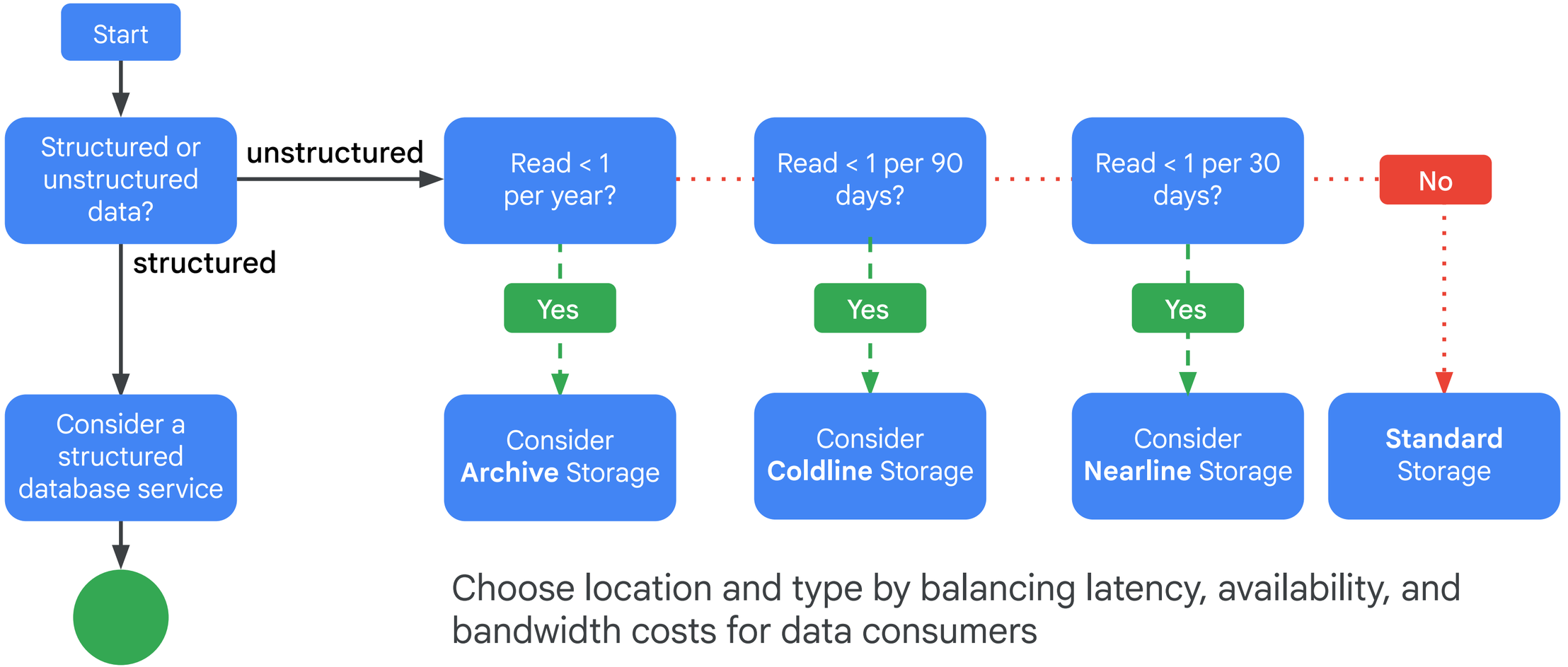
Data import services

- **Transfer Appliance:** Rack, capture and then ship your data to Google Cloud.
- **Storage Transfer Service:** Import online data (another bucket, an S3 bucket, or web source).
- **Offline Media Import:** Third-party provider uploads the data from physical media.

Cloud Storage provides strong global consistency

- Read-after-write
- Read-after-metadata-update
- Read-after-delete
- Bucket listing
- Object listing

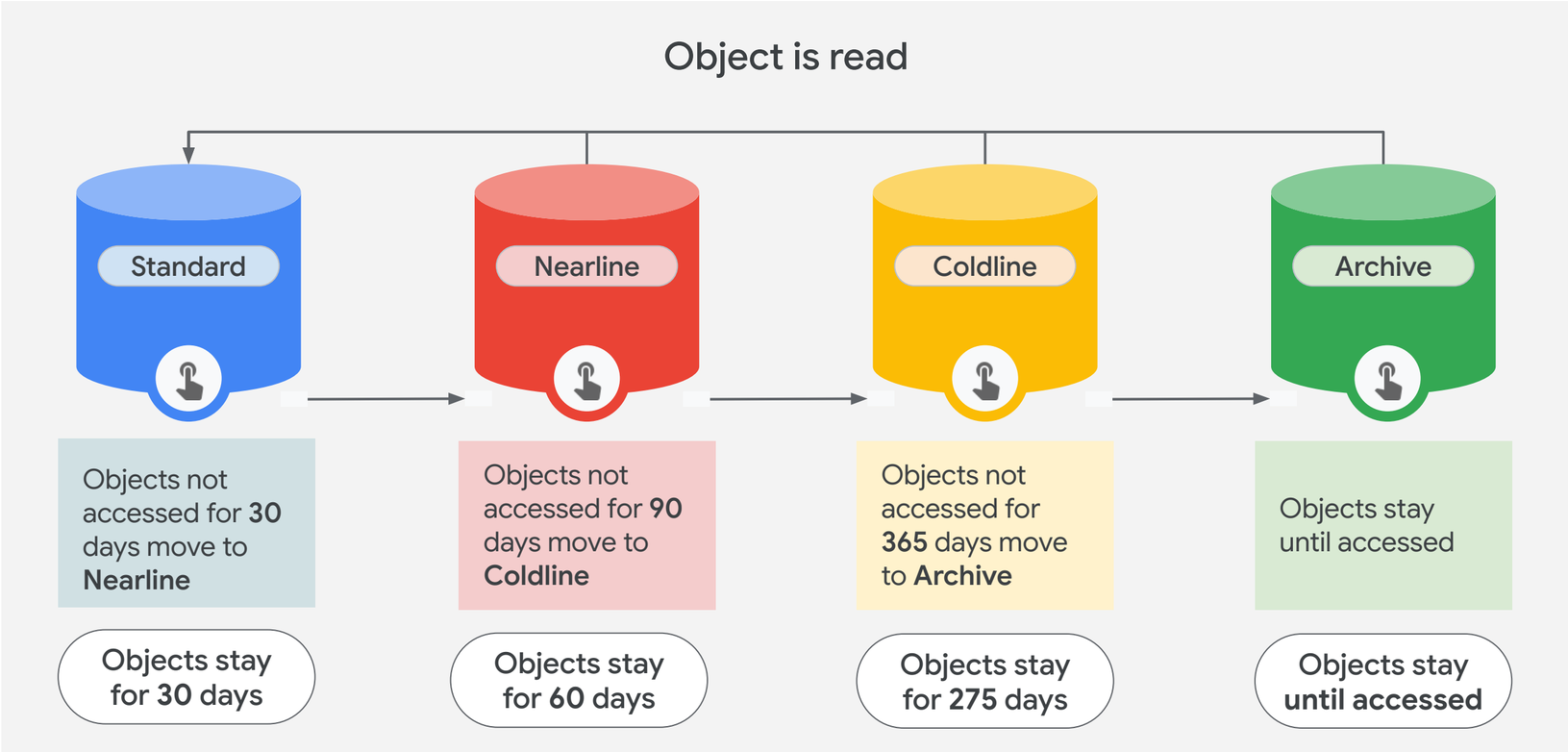
Choosing a storage class



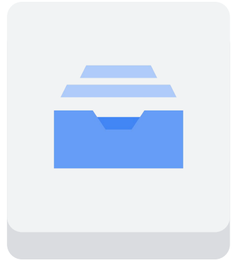
Choose location and type by balancing latency, availability, and bandwidth costs for data consumers

Autoclass storage in Google Cloud

Autoclass transitions objects in your bucket to appropriate storage classes based on the access pattern of each object.



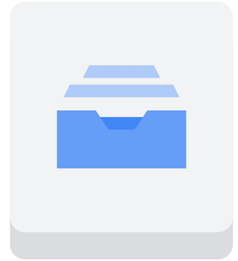
Filestore is a managed file storage service for applications



Filestore

- Fully managed network attached storage (NAS) for Compute Engine and GKE instances.
- Predictable performance.
- Full NFSv3 support.
- Scales to 100s of TBs for high-performance workloads.

Filestore has many use cases

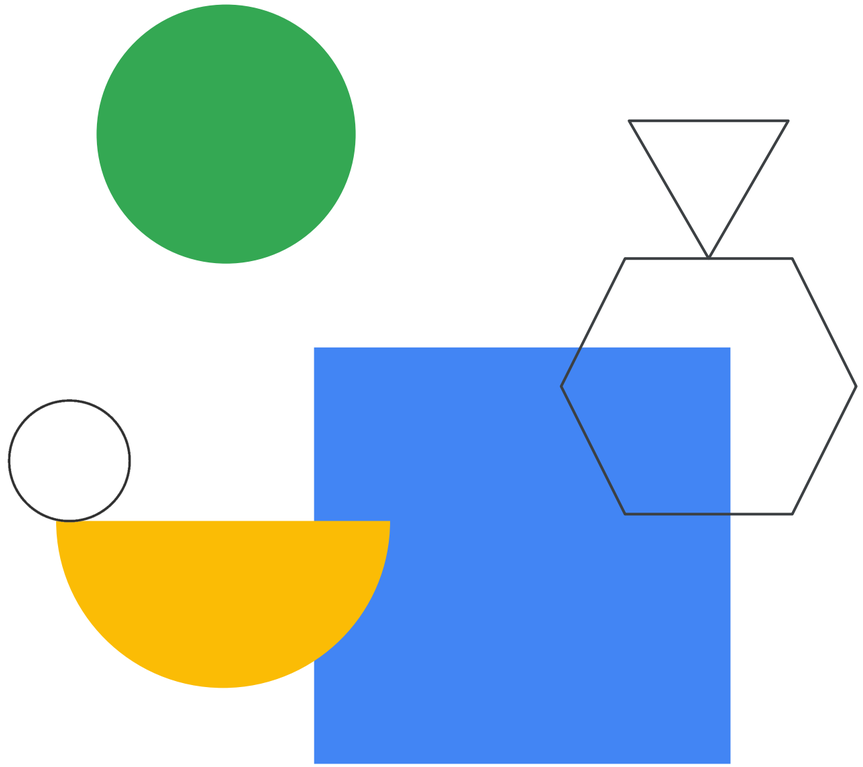


Filestore

- Application migration
- Media rendering
- Electronic Design Automation (EDA)
- Data analytics
- Genomics processing
- Web content management

Lab Intro

Cloud Storage



Lab objectives

01 Create and use buckets

02 Set access control lists to restrict access

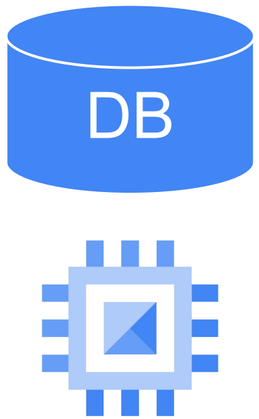
03 Use your own encryption keys

04 Implement version controls

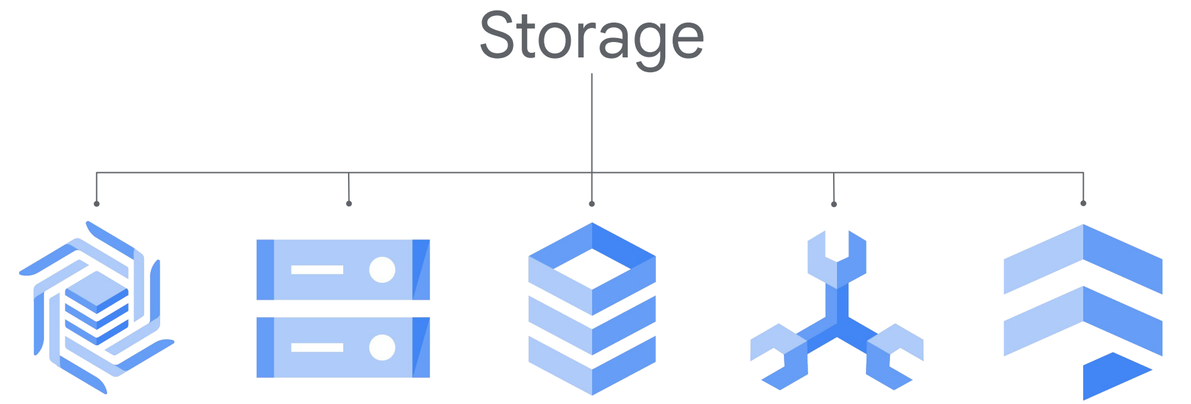
05 Use directory synchronization

Cloud SQL

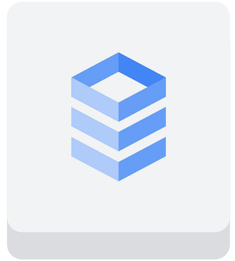
Build your own database solution or use a managed service



Compute Engine



Cloud SQL is a fully managed database service (MySQL, PostgreSQL, or Microsoft SQL Server)



Cloud SQL

- Patches and updates automatically applied
- You administer MySQL users
- Cloud SQL supports many clients
  - `gcloud sql`
  - App Engine, Google Workspace scripts
  - Applications and tools
    - SQL Workbench, Toad
    - External applications using standard MySQL drivers

Cloud SQL instance

Performance:

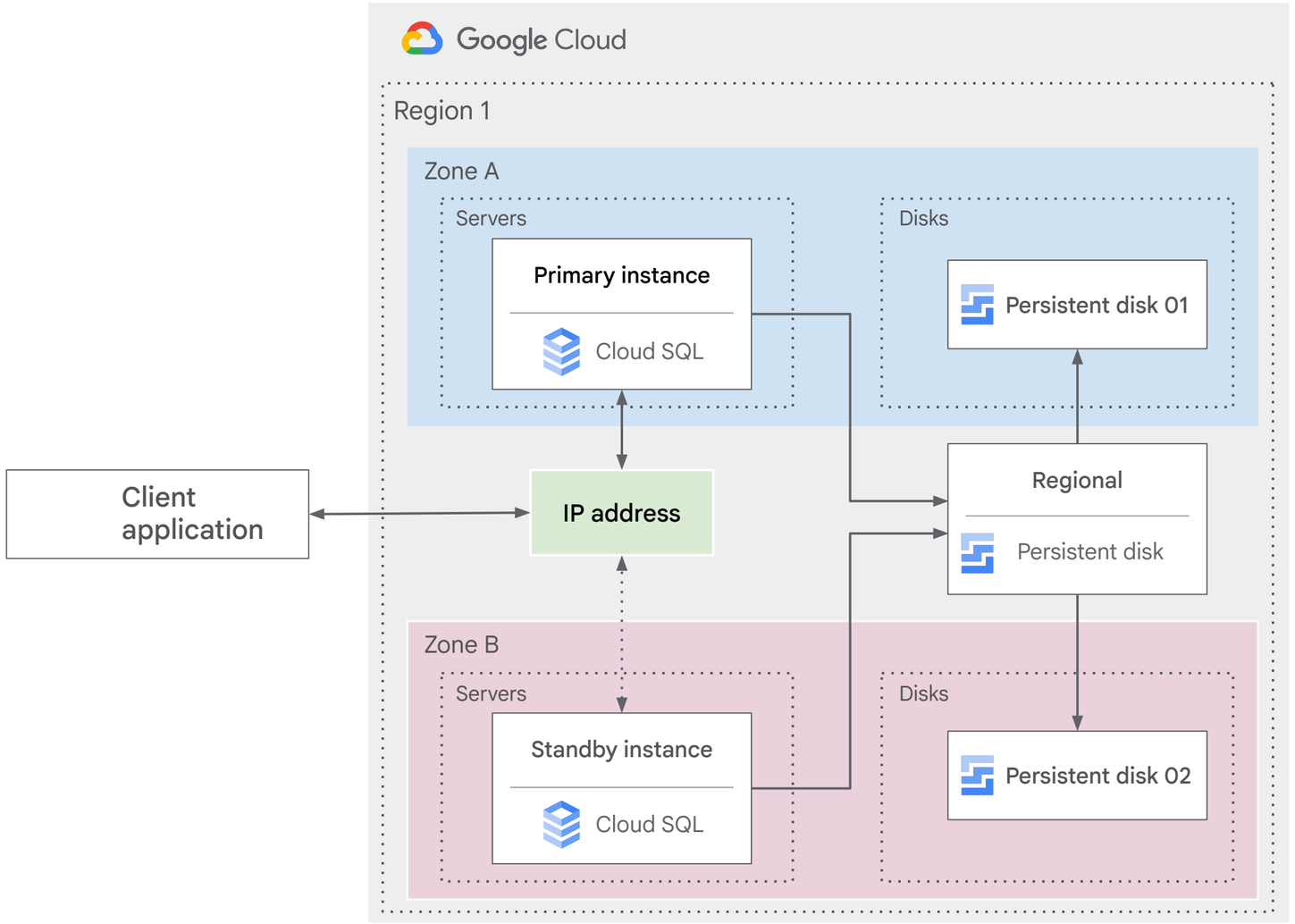
- 64 TB of storage
- 60,000 IOPS
- 624 GB of RAM
- Scale out with read replicas

Choice:

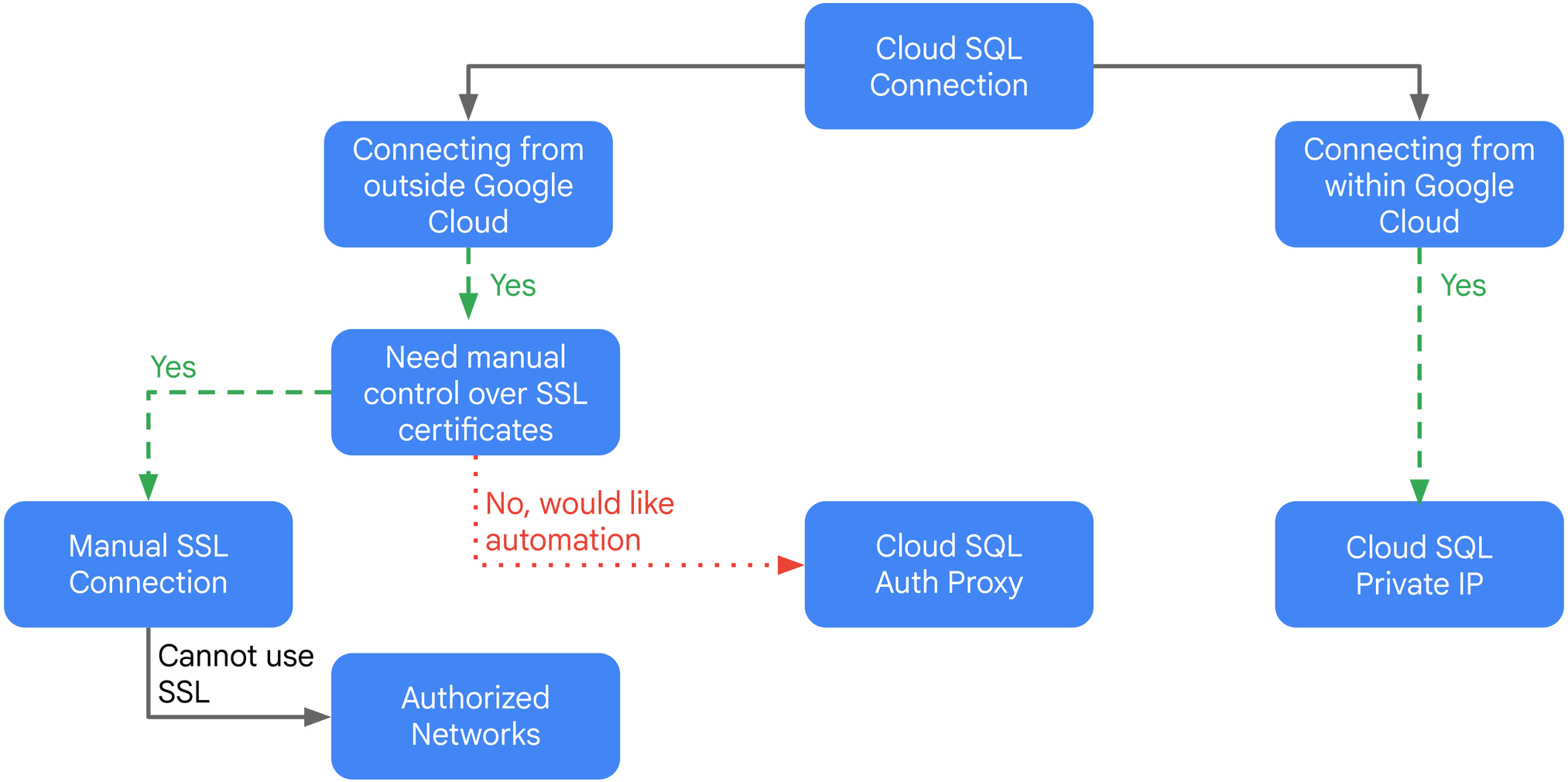
- MySQL 5.6, 5.7, or 8.0 (default)
- PostgreSQL 9.6, 10, 11, 12, 13, 14 or 15 (default)
- Microsoft SQL Server 2017 or 2019 (Standard default)

Cloud SQL services

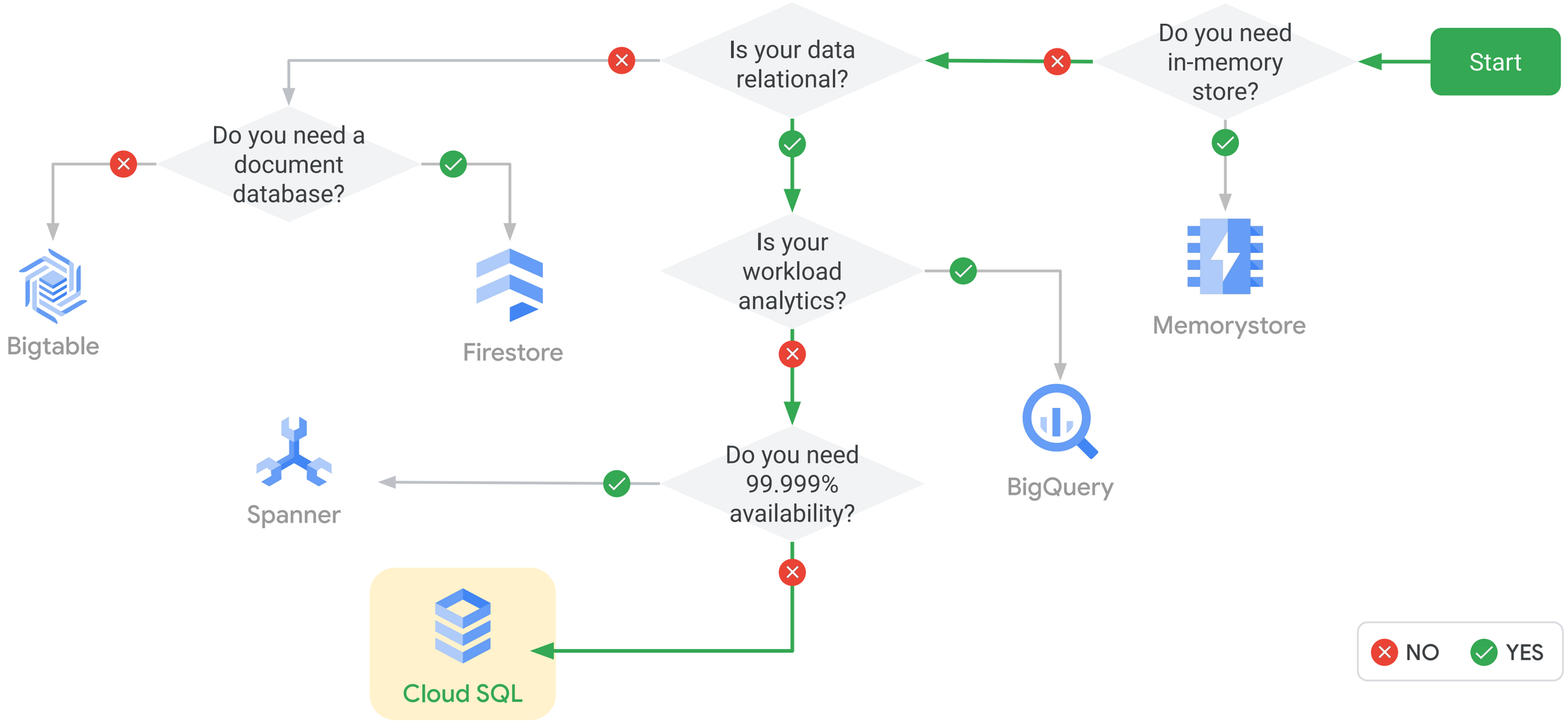
- HA configuration
- Backup service
- Import/export
- Scaling
  - Up: Machine capacity
  - Out: Read replicas



Connecting to a Cloud SQL instance

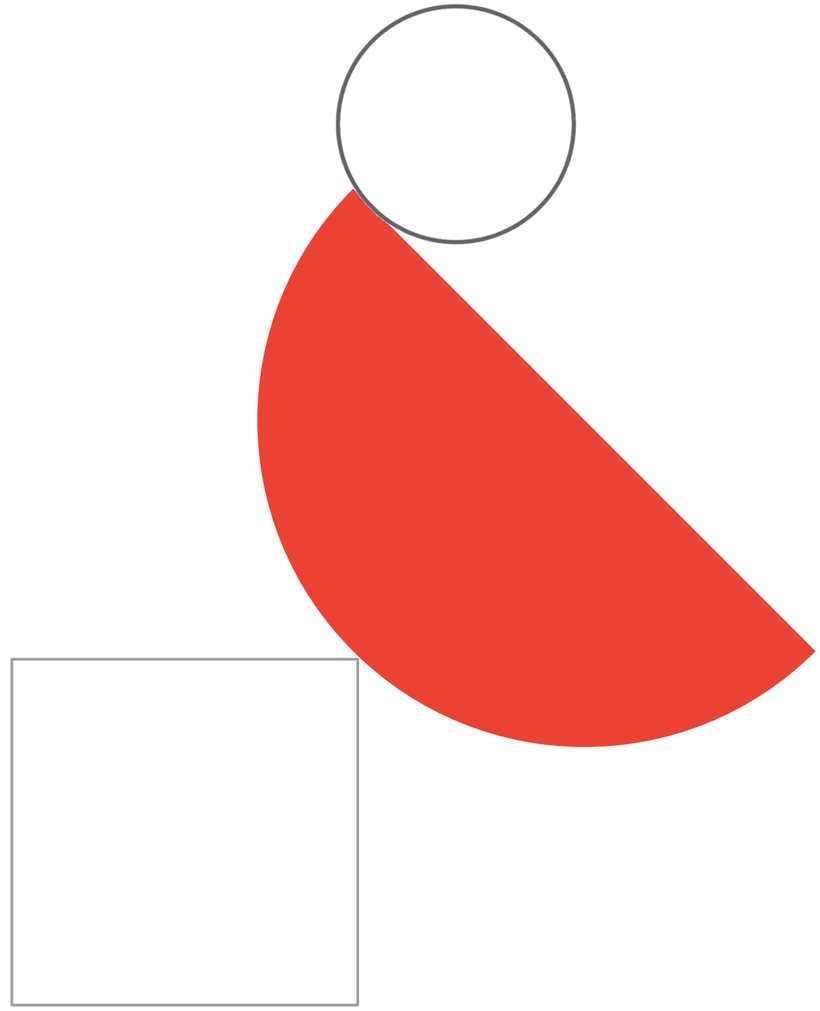


Choosing Cloud SQL



Lab Intro

Implementing Cloud SQL



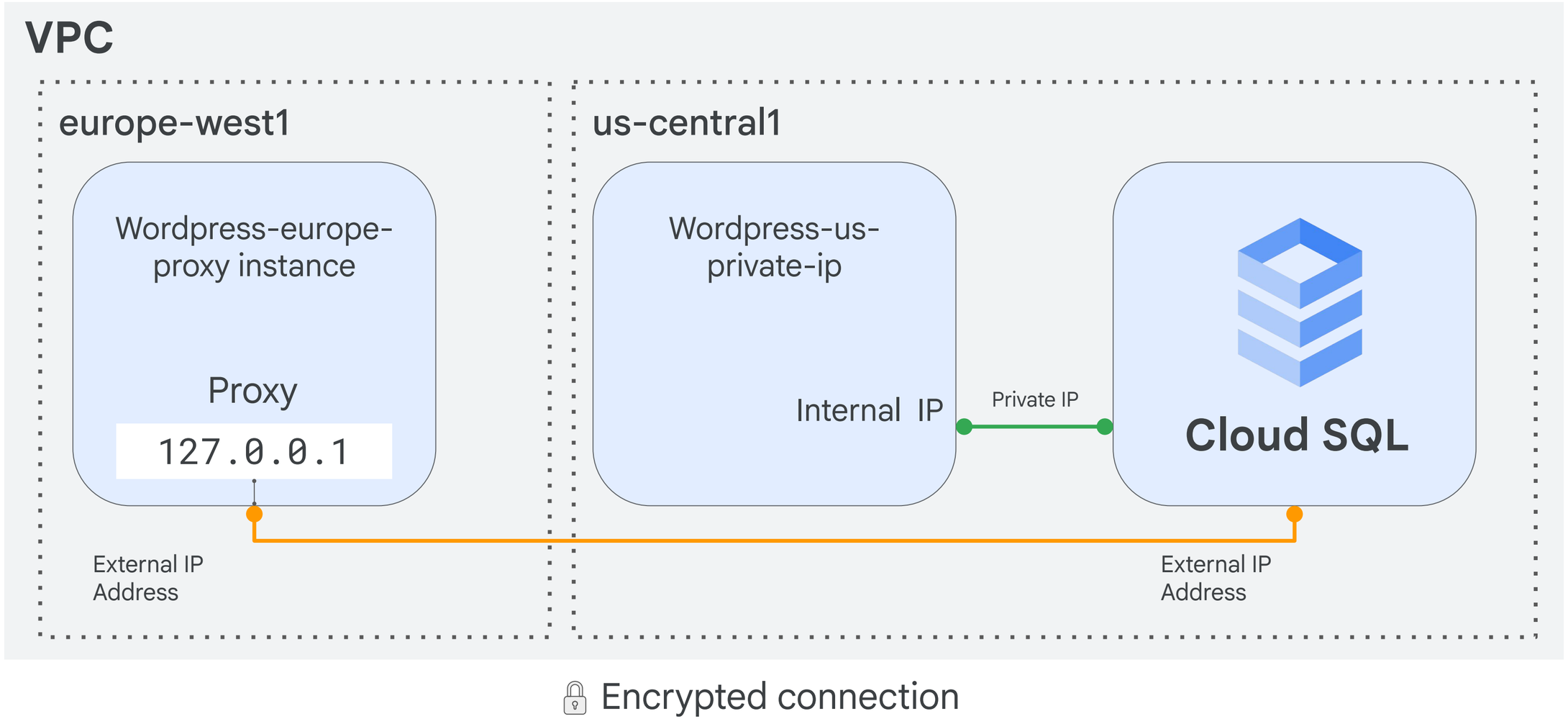
Lab objectives

01 Create a Cloud SQL database

02 Configure a virtual machine to run a proxy

03 Create a connection between an application and Cloud SQL

04 Connect an application to Cloud SQL using Private IP address



Spanner

Spanner combines the benefits of relational database structure with non-relational horizontal scale

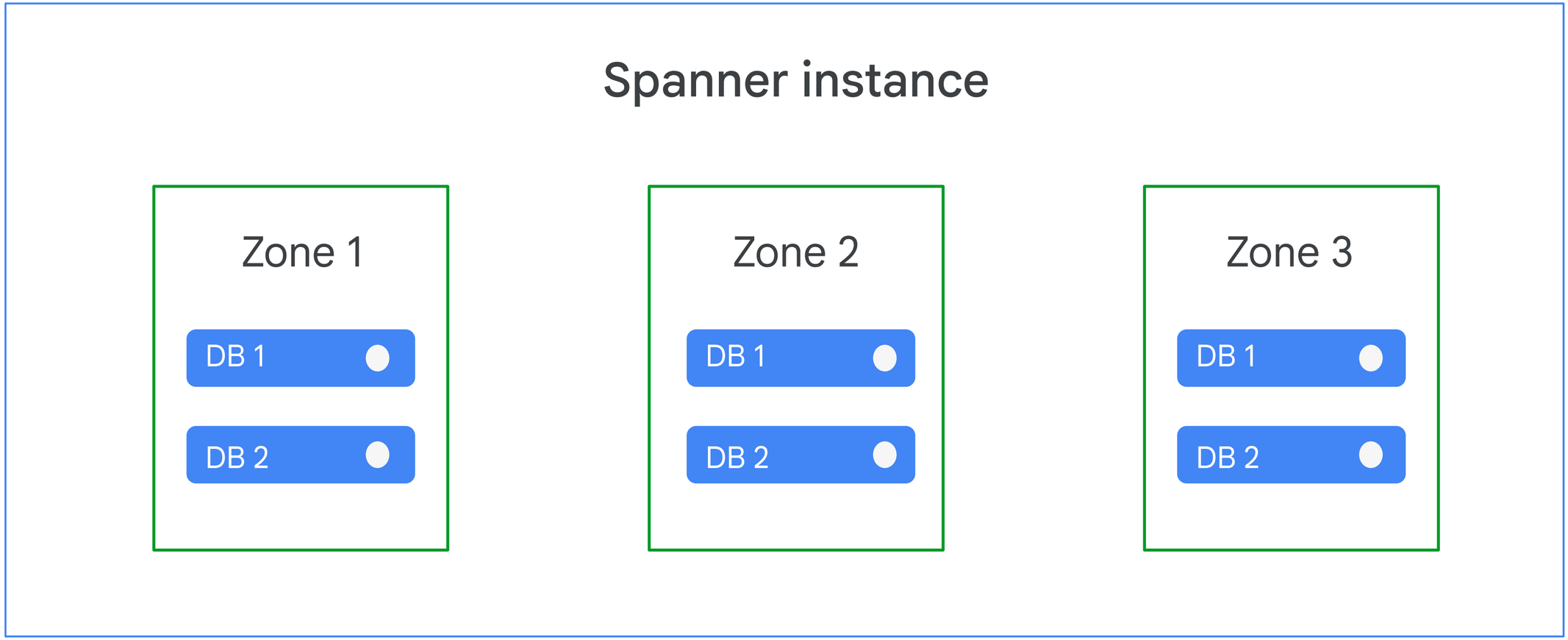
Spanner

- Scale to petabytes
- Strong consistency
- High availability
- Used for financial and inventory applications
- Monthly uptime
  - Multi-regional: 99.999%
  - Regional: 99.99%

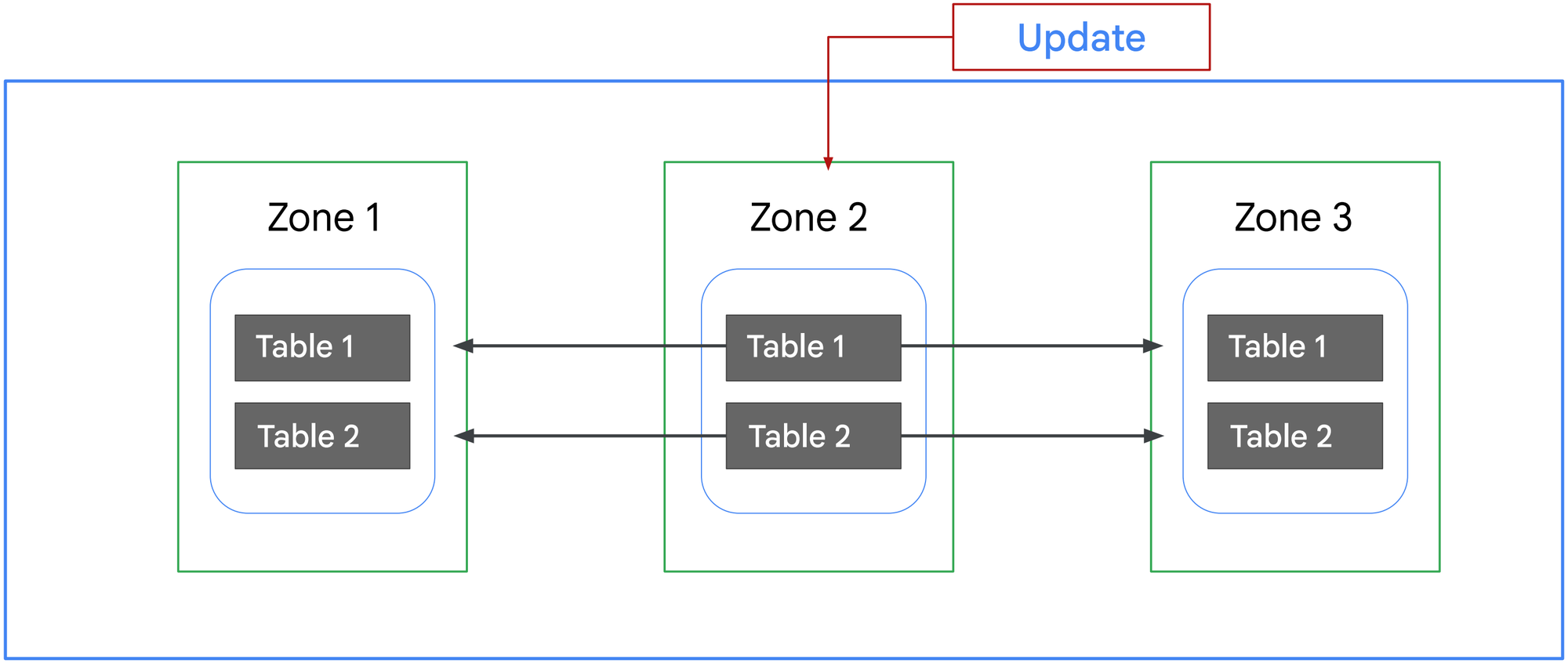
Characteristics

	Spanner	Relational DB	Non-Relational DB
Schema	Yes	Yes	No
SQL	Yes	Yes	No
Consistency	Strong	Strong	Eventual
Availability	High	Failover	High
Scalability	Horizontal	Vertical	Horizontal
Replication	Automatic	Configurable	Configurable

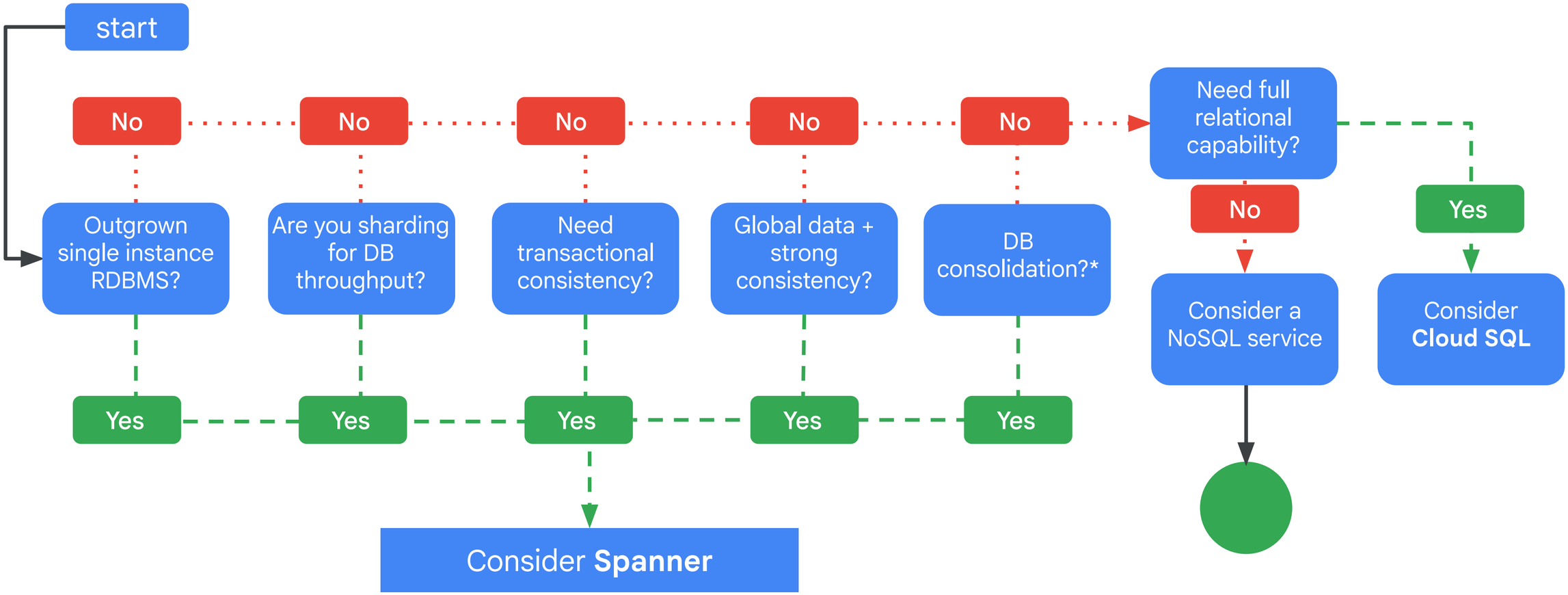
Spanner architecture



Data replication is synchronized across zones using Google's global fiber network

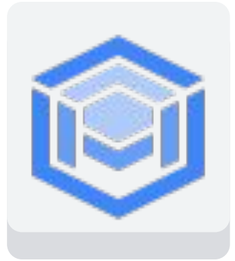


Choosing Spanner



AlloyDB

AlloyDB is a fully managed database service



AlloyDB

- Fully managed database service
- Fast transactional processing
- High availability
- Real-time business insights

Firestore

Firestore is a NoSQL document database

Firestore

- Simplifies storing, syncing, and querying data
- Mobile, web, and IoT apps at global scale
- Live synchronization and offline support
- Security features
- ACID transactions
- Multi-region replication
- Powerful query engine

Firestore is the next generation of Datastore

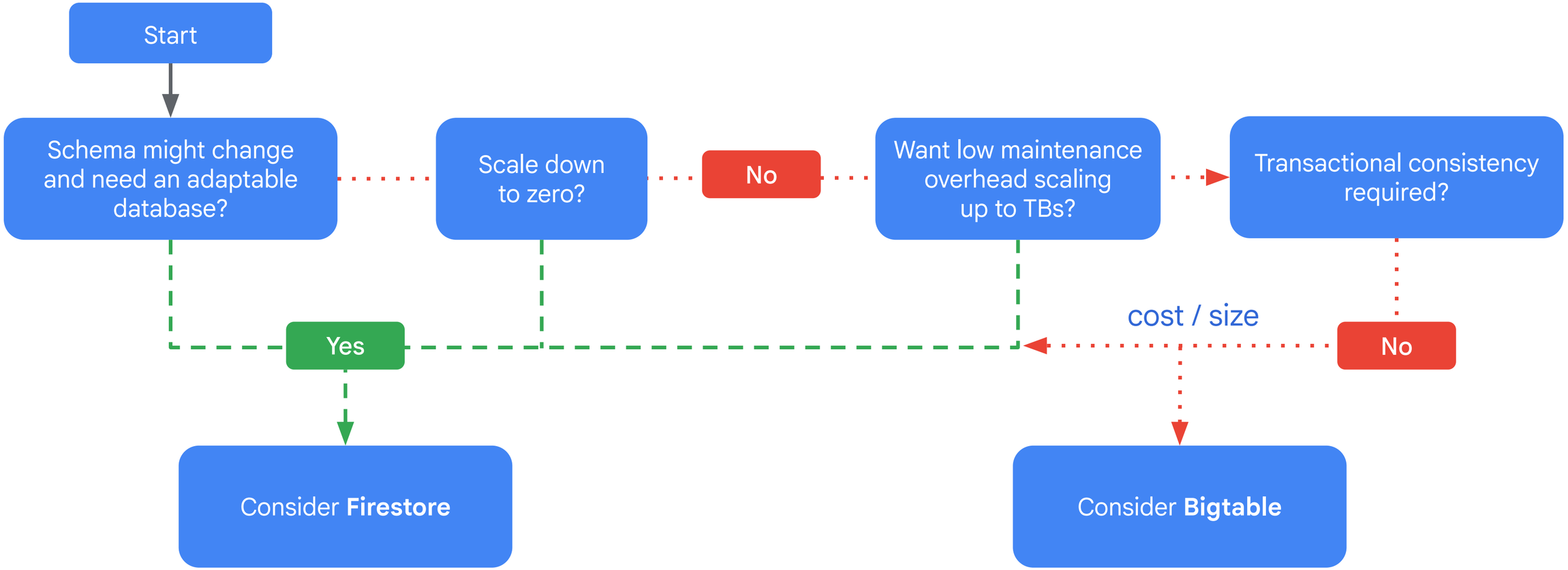
Datastore mode (new server projects):

- Compatible with Datastore applications
- Strong consistency
- No entity group limits

Native mode (new mobile and web apps):

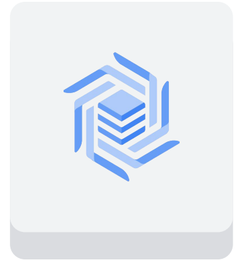
- Strongly consistent storage layer
- Collection and document data model
- Real-time updates
- Mobile and Web client libraries

Choosing Firestore



Bigtable

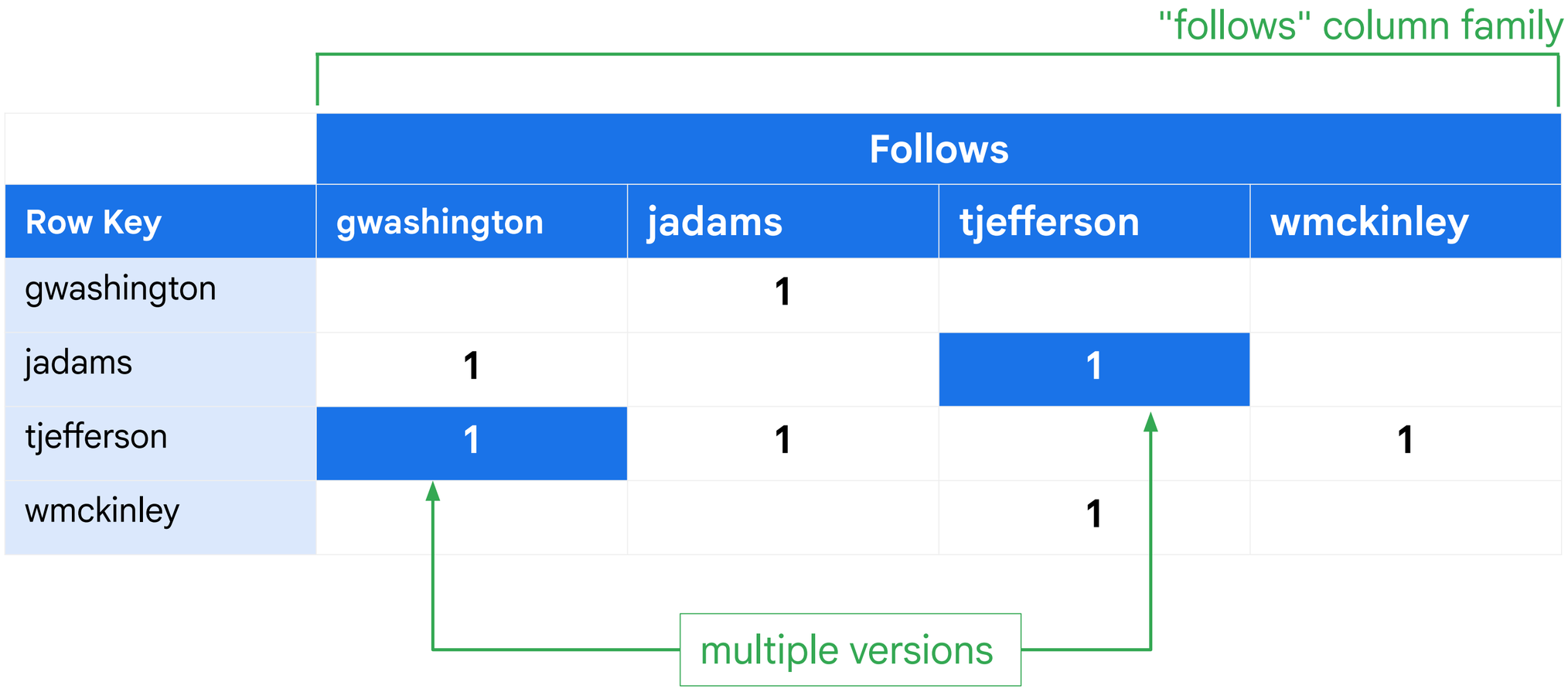
Bigtable is a NoSQL big data database service



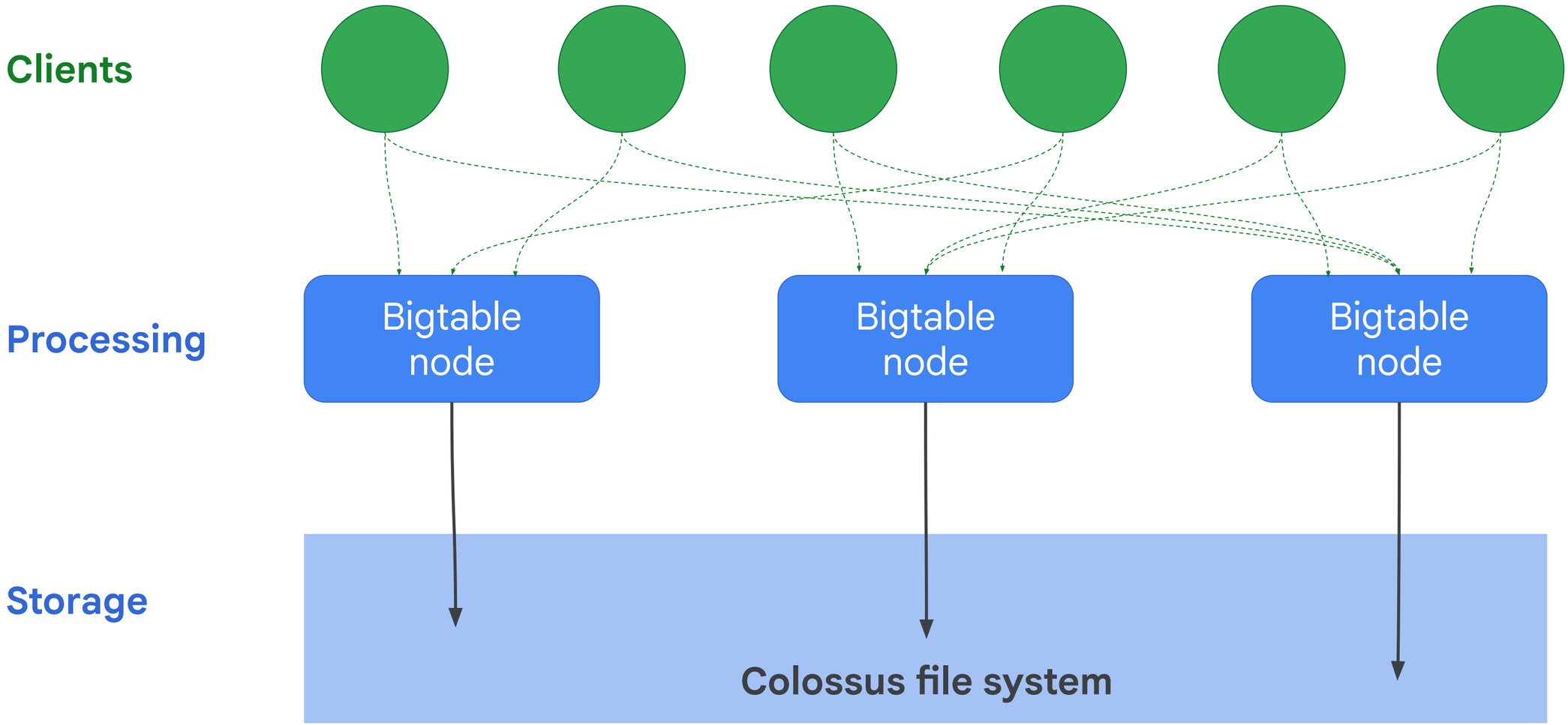
Bigtable

- Petabyte-scale
- Consistent sub-10ms latency
- Seamless scalability for throughput
- Learns and adjusts to access patterns
- Ideal for Ad Tech, Fintech, and IoT
- Storage engine for ML applications
- Easy integration with open source big data tools

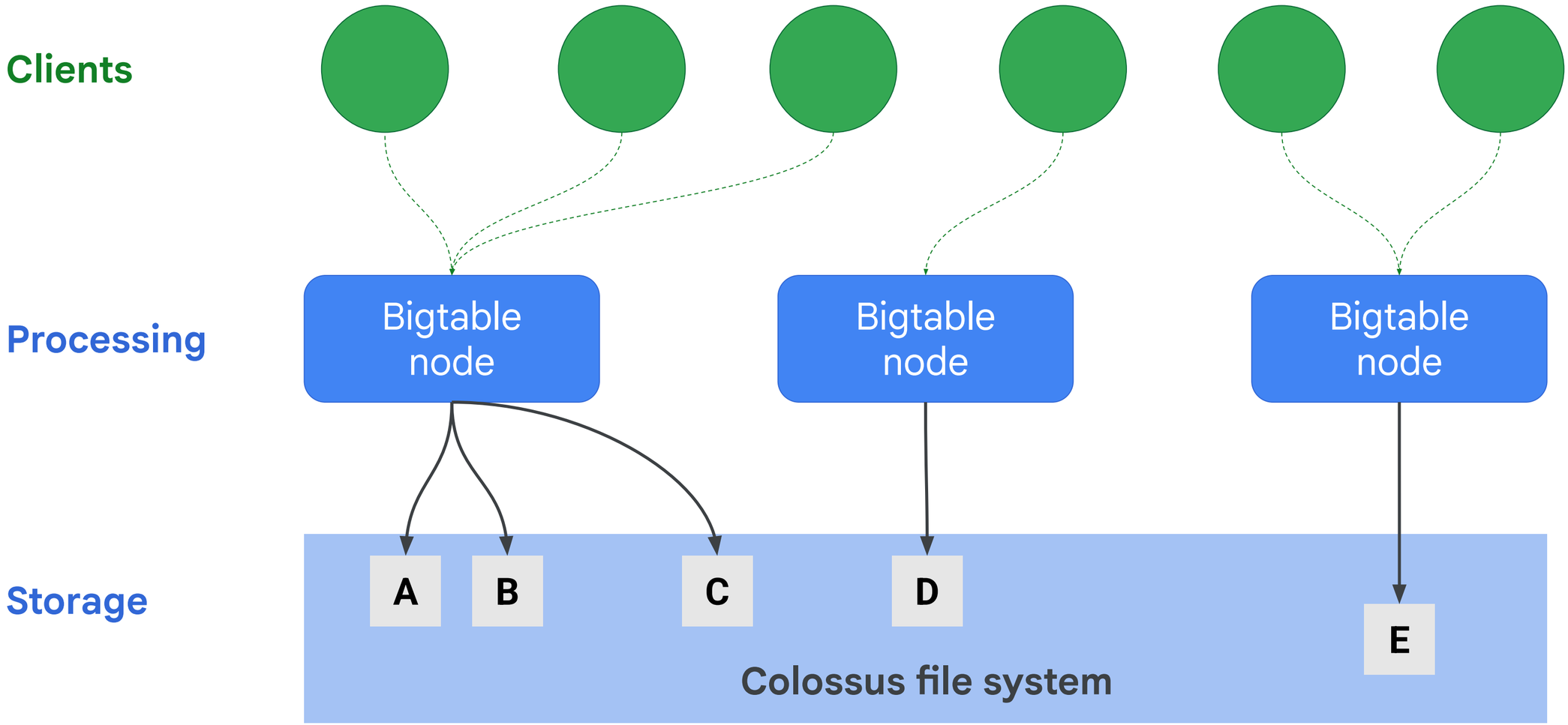
Bigtable storage model



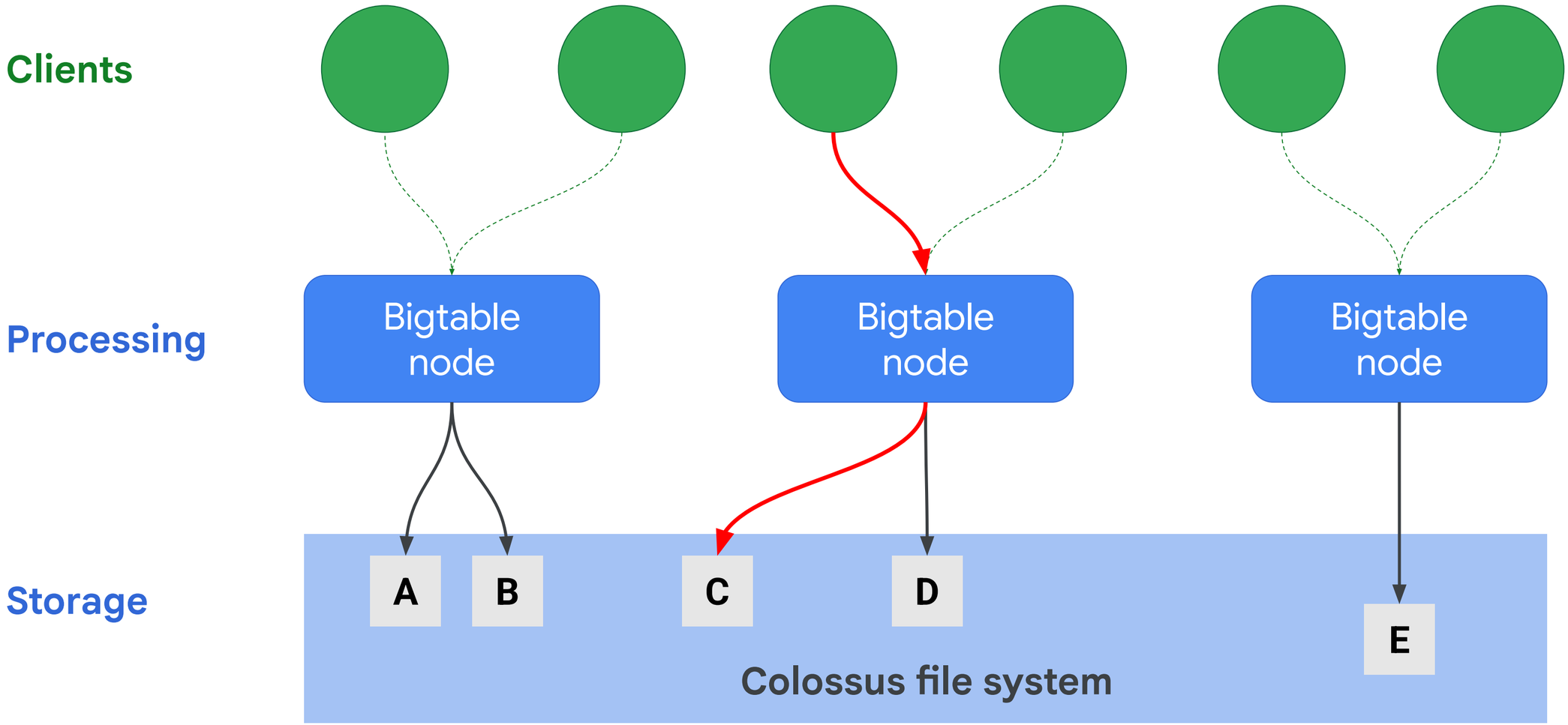
Processing is separated from storage



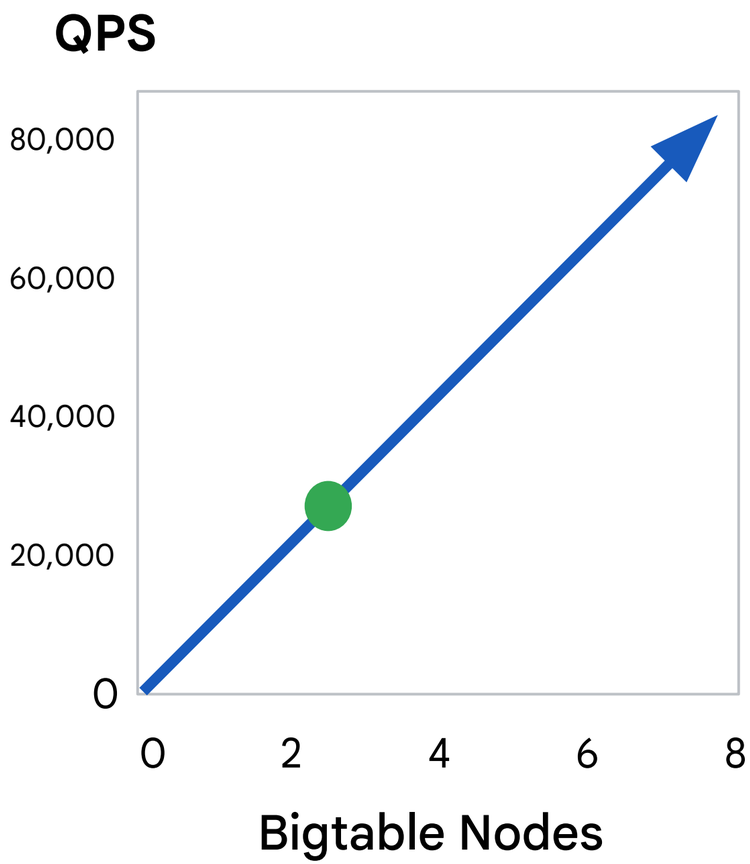
Learns access patterns

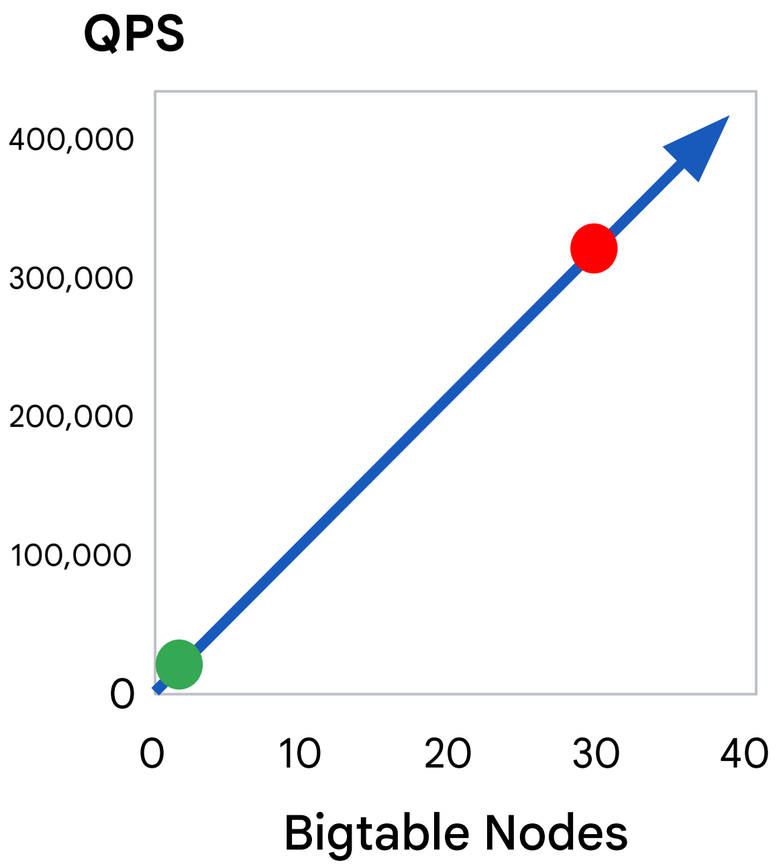


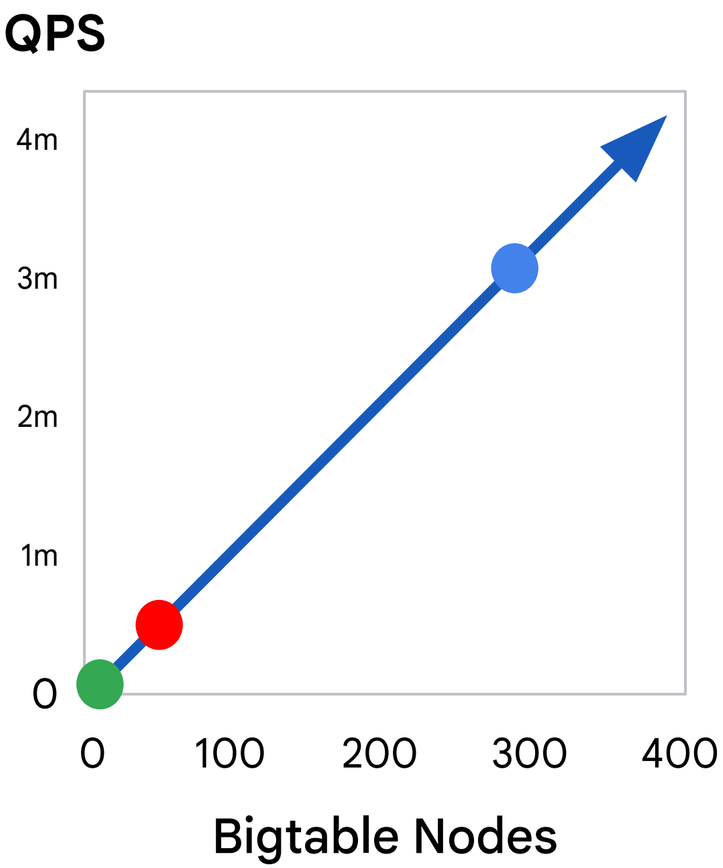
Rebalances without moving data



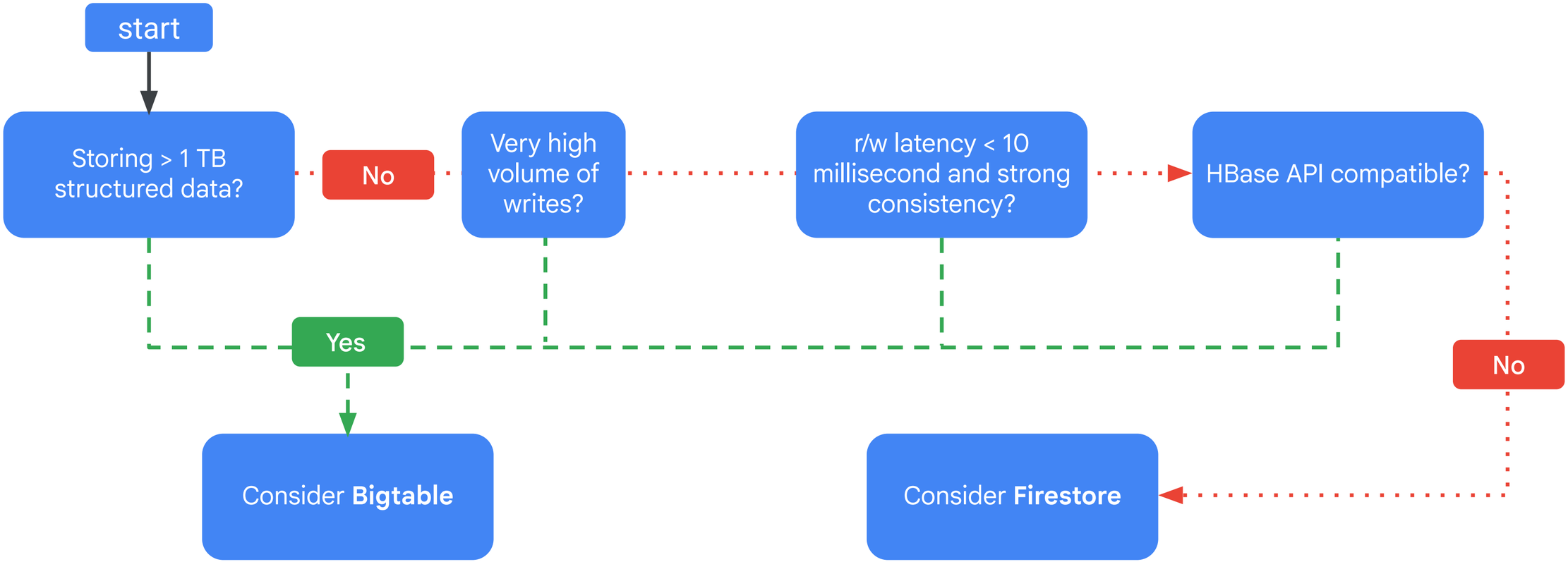
Throughput scales linearly







Choosing Bigtable



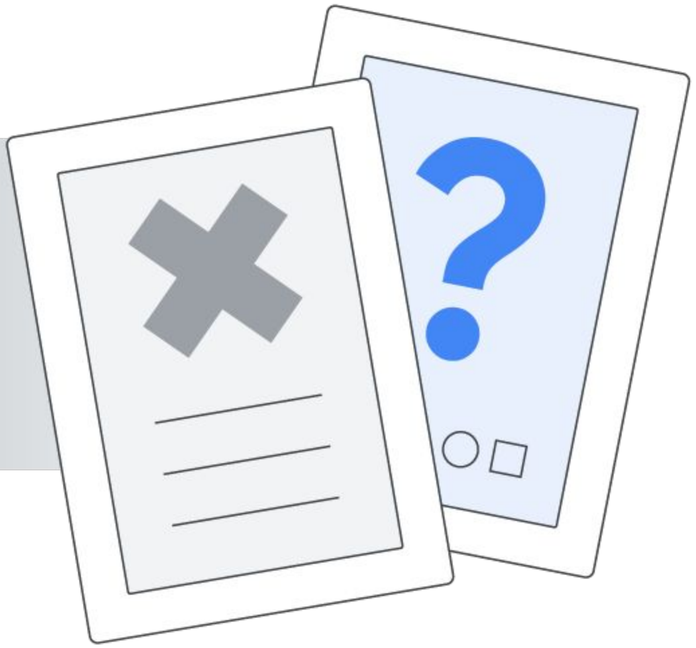
- Bigtable scales UP well
- Firestore scales DOWN well

Memorystore

Memorystore is a fully managed Redis service

Memorystore

- In-memory data store service
- Focus on building great apps
- High availability, failover, patching, and monitoring
- Sub-millisecond latency
- Instances up to 300 GB
- Network throughput of 12 Gbps
- Easy Lift-and-Shift



Quiz

Question #1

Question

What data storage service might you select if you just needed to migrate a standard relational database running on a single machine in a data center to the cloud?

A. Cloud SQL

B. BigQuery

C. Persistent Disk

D. Cloud Storage

Question #1

Answer

What data storage service might you select if you just needed to migrate a standard relational database running on a single machine in a data center to the cloud?

A. **Cloud SQL** 

B. BigQuery

C. Persistent Disk

D. Cloud Storage

Question #2

Question

Which Google Cloud data storage service offers ACID transactions and can scale globally?

A. Cloud Storage

B. Cloud CDN

C. Spanner

D. Cloud SQL

Question #2

Answer

Which Google Cloud data storage service offers ACID transactions and can scale globally?

A. Cloud Storage

B. Cloud CDN

C. **Spanner** 

D. Cloud SQL

Question #3

Question

Which data storage service provides data warehouse service for storing data but also offers an interactive SQL interface for querying the data?

A. BigQuery

B. Dataproc

C. Datalab

D. Cloud SQL

Question #3

Answer

Which data storage service provides data warehouse service for storing data but also offers an interactive SQL interface for querying the data?

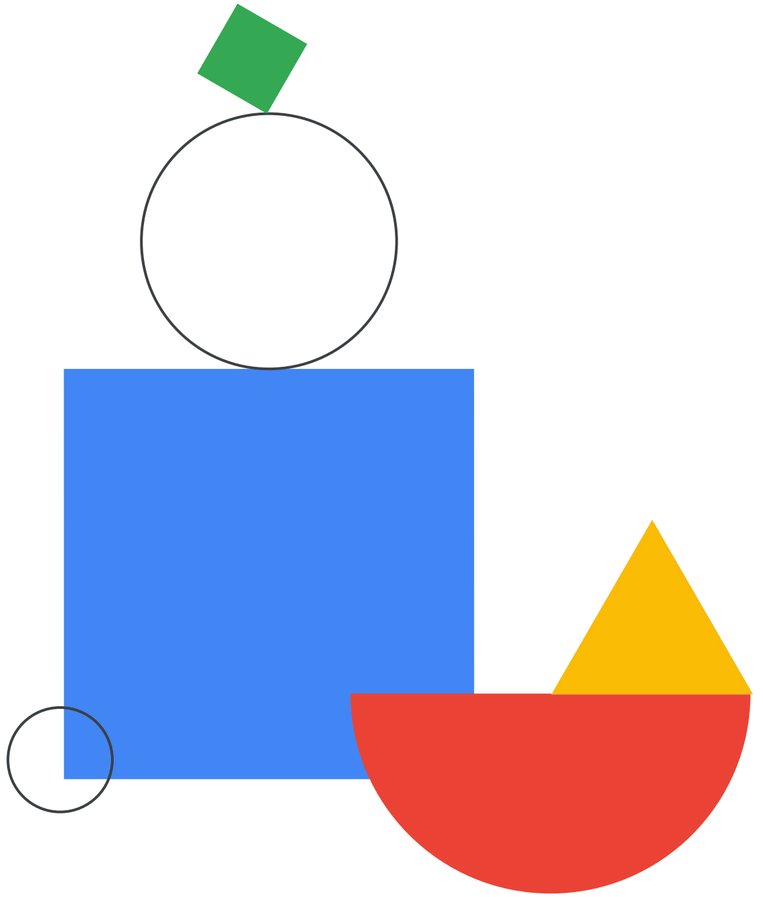
A. **BigQuery** 

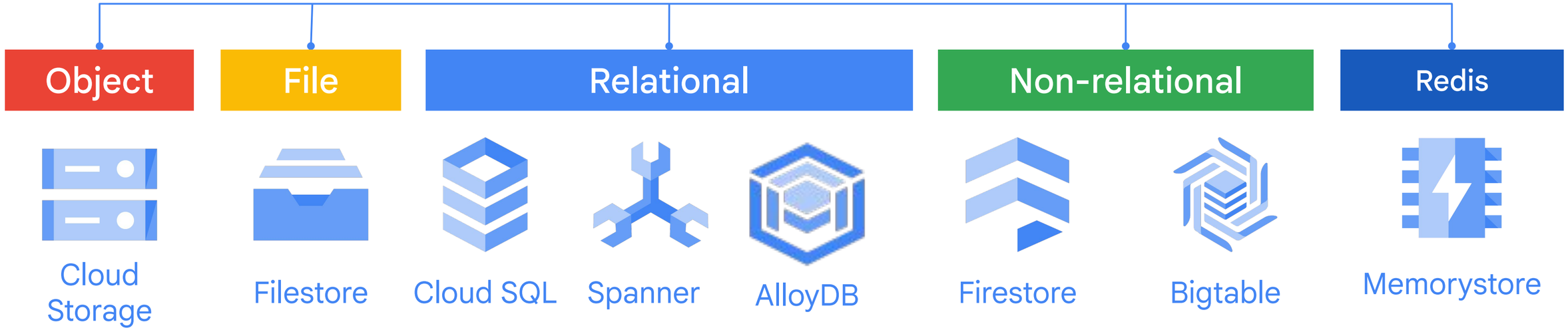
B. Dataproc

C. Datalab

D. Cloud SQL

Review: Storage and Database Services





Decision chart