

Preparing for Your Professional Cloud Architect Journey

Module 4: Analyzing and Optimizing Technical
and Business Processes

Week 5 topics

Gen AI - part 2

Data services (Filestore,
Firestore, Memorystore,
Spanner, BigQuery, Bigtable)



Optimizing technical
and business
processes and
procedures

1
2
3
4

Cymbal Retail case
study analysis

Gen AI - Part 2

Google's Unified AI technology stack



Applications

[Gemini Advanced](#) | [Google Workspace apps](#) | [NotebookLM](#)



Agents & Extensions

[Agentspace](#) | [Agent Garden](#) | [3P Connectors](#)



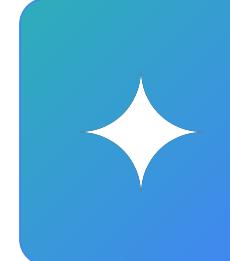
Platform

[Vertex AI](#)



Data Platform

[Multimodal Analytics](#) | [AI Insights](#) | [Data Science](#)



Gen AI Models

[Gemini](#) | [Imagen](#) | [Veo](#) | [Partner](#) | [Open](#)



Infrastructure

[Performance-optimized hardware](#) | [Open software](#) | [Flexible consumption](#)

NotebookLM & NotebookLM Plus

NotebookLM is the ultimate tool for helping you understand the information that matters most to you.

Unlock critical insights faster - **grounded only in the sources you provide.**

NotebookLM

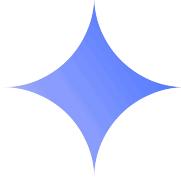
Free for individuals to get started

- ✓ Built with Gemini 1.5
- ✓ Upload PDFs, websites, Google Docs and Slides, YouTube URLs, and more
- ✓ Create one-click summaries, FAQs, timelines, and briefing docs
- ✓ Generate Audio Overviews and listen on-the-go
- ✓ Ask questions for deeper insights and get answers with citations

NotebookLM Plus

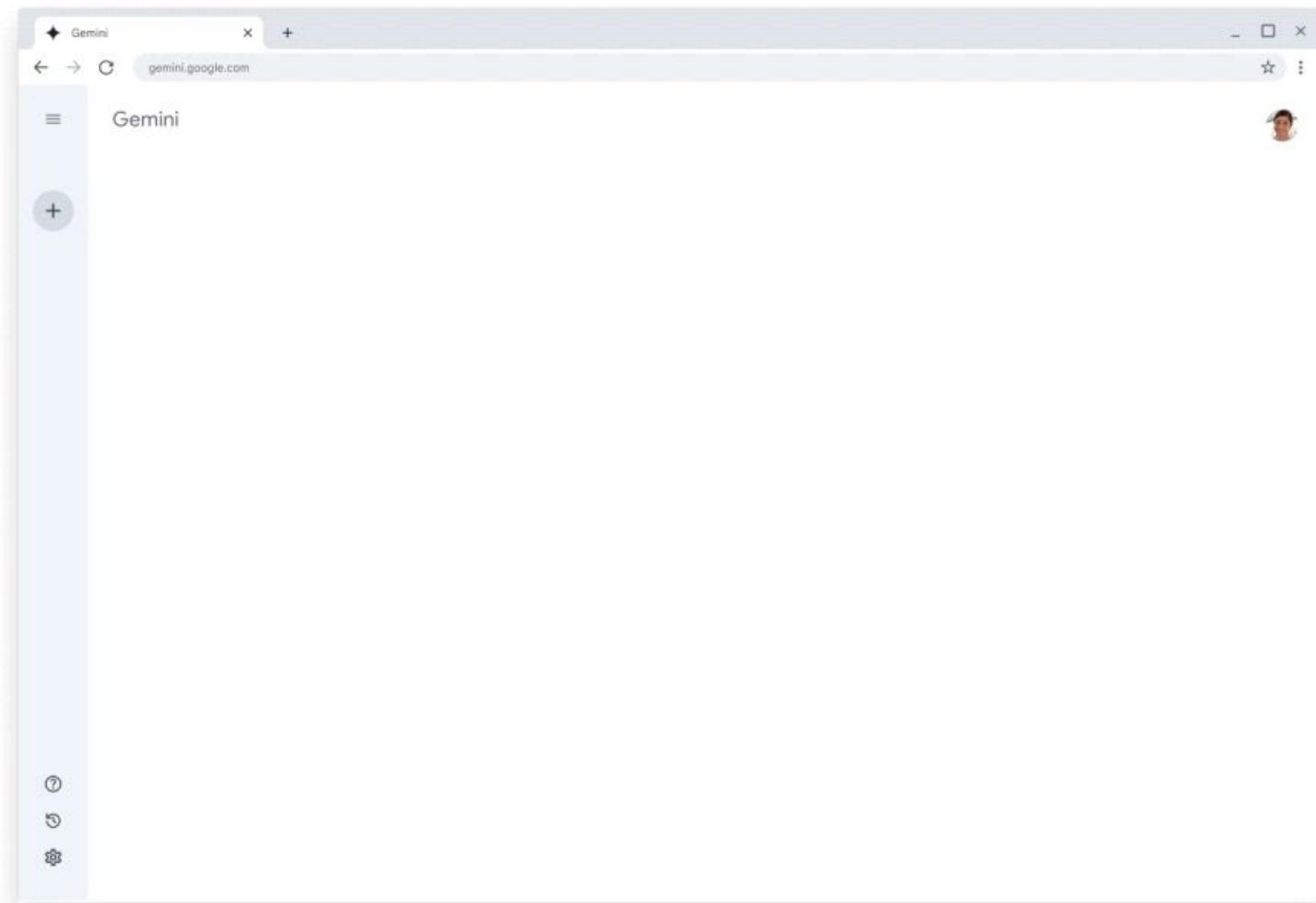
Everything in NotebookLM, plus:

- ✓ Get 5x more Audio Overviews, notebooks, and sources per notebook
- ✓ Customize the style and tone of your notebooks
- ✓ Create shared notebooks for your team and get usage analytics
- ✓ [Learn about more benefits and pricing](#)

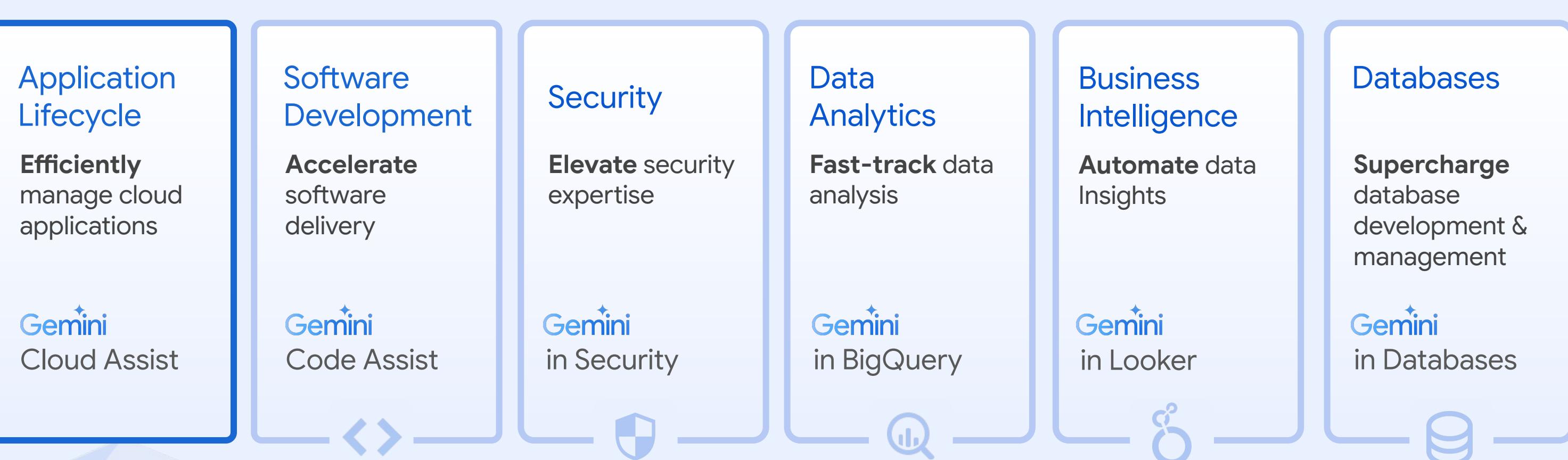
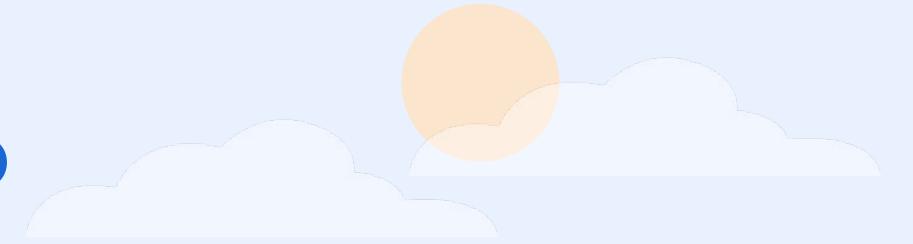


Gemini app

Chat with Gemini to kickstart
brainstorming and planning, or tackle
complex projects like research and
coding.



★ Gemini for Google Cloud Portfolio



Google Workspace with ✨ Gemini

A cloud-native suite of premium tools for work, now with Google AI built into every plan.



**Research, learn,
and tackle
complex work
with Gen AI**



**Find, share,
and manage
files easily**



**Collaborate
and co-edit
in real time**



**Connect
instantly, from
any device**



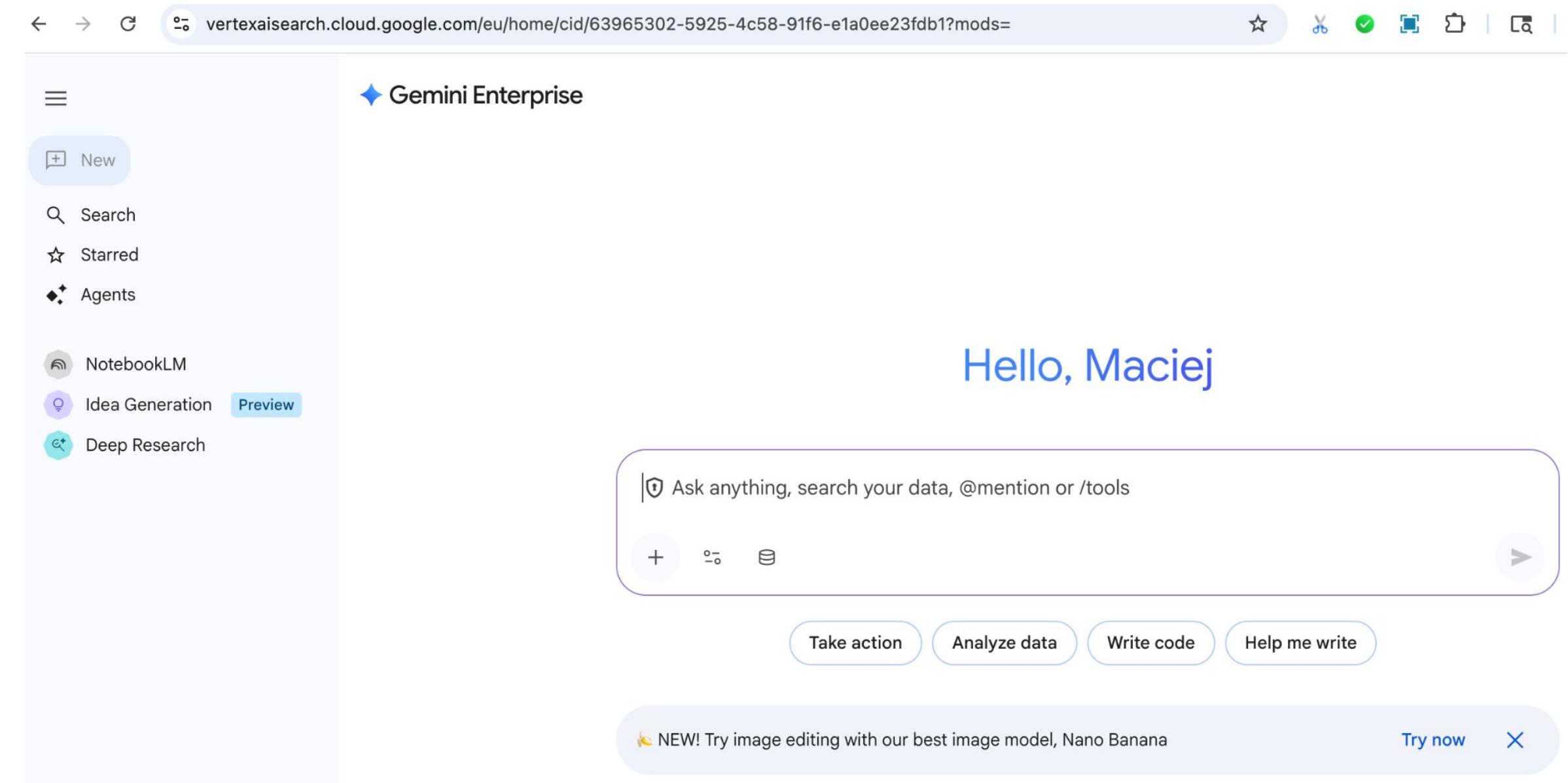
**Stay on top
of things**

AppSheet and Workspace Flows

Build no-code apps and automated AI-powered workflows

Gemini Enterprise

Unlock enterprise expertise for employees with agents that bring together Gemini's advanced reasoning, Google-quality search, and enterprise data.



Gemini Enterprise example use-case

Great. Here are your chosen concepts in a little more detail.

Barritas Energéticas con Sabor a Tortitas para Llevar

Información clave
A la gente le encanta el sabor de las tortitas, pero la vida moderna impide disfrutar de un desayuno completo.

Beneficio
Disfruta del sabor de las tortitas en cualquier lugar con nuestras barritas. Perfectas para desayuno, merienda o reponer fuerzas.

Razón para Creer
Nuestro exclusivo sabor a tortitas y su práctico formato para llevar nos diferencian de las tortitas comunes.

Spanish (Castilian) ▾

Hindi ▾

Translate ▾

Refine

Product Innovation Agent

Would you like to generate some product designs for these concepts?

Write your prompt here...

i Pro

अपना पैनकेक किट बुद बनाएँ

अंतर्दृष्टि
लोग सुविधाजनक नाशे के लिये चाहते हैं, लेकिन अपने स्वाद और आहार के अनुसार कुट का त्वात किये बिना।

लाभ
अपने पॉलीटेक स्लाइड और टॉफिंग के साथ अपना पैनकेक बनाने हमारे लिए अपने काहार के लिए भी उपयुक्त है।

विशुलिकरण का कारण
अपनी स्वाद के लिए प्रिमियम स्लाइड और टॉफिंग में से चुनें।

अंतर्दृष्टि
लोग सुविधाजनक नाशे के लिये चाहते हैं, लेकिन अपने स्वाद और आहार के अनुसार कुट का त्वात किये बिना।

लाभ
हमारे खील किट आपको अपने पॉलीटेक स्लाइड और टॉफिंग के साथ अपने पैनकेक नाशे को नियंत्रित करने की सुविधा देते हैं, यह सब आपकी आहार संस्कृति अवश्यकताओं को पूरा करते हुए।

विशुलिकरण का कारण
हम विशिष्ट व्यवहार के स्वाद, टॉफिंग और आहार विकल्प प्रदान करते हैं ताकि आप अपनी पसंद के अनुसार सही पैनकेक बना सकें।

Pancake-Inspired Granola:

Insight
People crave the taste of pancakes, but often seek healthier breakfast alternatives.

Benefit
Enjoy pancake flavor in a guilt-free way with our Pancake-Inspired Granola. It satisfies your cravings while supporting a healthy lifestyle.

RTB
Our granola delivers authentic pancake taste with wholesome ingredients like oats and maple syrup for a delicious and nutritious start to your day.

la vida moderna impide disfrutar de un desayuno completo.

Beneficio
Disfruta del sabor de las tortitas en cualquier lugar con nuestras barritas. Perfectas para desayuno, merienda o reponer fuerzas.

Razón para Creer
Nuestro exclusivo sabor a tortitas y su práctico formato para llevar nos diferencian de las tortitas comunes.

अपना पैनकेक किट बुद बनाएँ

अंतर्दृष्टि
लोग सुविधाजनक नाशे के लिये चाहते हैं, लेकिन अपने स्वाद और आहार के अनुसार कुट का त्वात किये बिना।

लाभ
हमारे खील किट आपको अपने पॉलीटेक स्लाइड और टॉफिंग के साथ अपने पैनकेक नाशे को नियंत्रित करने की सुविधा देते हैं, यह सब आपकी आहार संस्कृति अवश्यकताओं को पूरा करते हुए।

विशुलिकरण का कारण
हम विशिष्ट व्यवहार के स्वाद, टॉफिंग और आहार विकल्प प्रदान करते हैं ताकि आप अपनी पसंद के अनुसार सही पैनकेक बना सकें।

Pancake-Inspired Granola:

Insight
People crave the taste of pancakes, but often seek healthier breakfast alternatives.

Benefit
Enjoy pancake flavor in a guilt-free way with our Pancake-Inspired Granola. It satisfies your cravings while supporting a healthy lifestyle.

RTB
Our granola delivers authentic pancake taste with wholesome ingredients like oats and maple syrup for a delicious and nutritious start to your day.

Gen AI Risks

Top 5 AI security & technical risks

1

Prompt Injection

Get the model to execute malicious instructions “injected” inside a prompt.

2

Data Exposure

A model may reveal data that is private to an individual or an organization.

3

Model Theft

A model may be stolen by an attacker.

4

Data Poisoning

Poison a dataset to alter the behavior of a model trained or tuned using it.

5

Model Integrity

Compromise

Tampering the model to change its behavior towards a harmful outcome.

Other AI risks (operational, ethical etc)

Bias	If AI is trained on data that reflects existing societal biases (like against certain groups of people), the AI can learn and even amplify those biases.
Inconsistency & Inaccuracy	AI isn't perfect; it can make mistakes (sometimes called hallucinations). AI also may give different answers to the same question at different times. This can make it hard to rely on its output.
Opacity	Sometimes it's hard to understand why AI makes a particular decision. If we can't see its reasoning, it can be tough to check if it's fair or accurate.
Cybersecurity	AI systems, especially those using models trained on huge amounts of data, can be targets for cyberattacks. This could expose private information or even let bad actors interfere with the AI's results.
Data Privacy	The data used to train AI could include personal data. Using this data without assessing transparency and control considerations is a risk.
Intellectual Property	Some data is protected by copyright or other intellectual property protections. There are risks that this data cannot be legally or ethically used without a proper license.

Potential solutions to AI's Risks

Bias	If AI is trained on data that reflects existing societal biases (like against certain groups of people), the AI can learn and even amplify those biases.	Use diverse data for model training
Inconsistency & Inaccuracy	AI isn't perfect; it can make mistakes (sometimes called hallucinations). AI also may give different answers to the same question at different times. This can make it hard to rely on its output.	Rigorously test and monitor model's performance
Opacity	Sometimes it's hard to understand why AI makes a particular decision. If we can't see its reasoning, it can be tough to check if it's fair or accurate.	Improve model transparency & explainability
Cybersecurity	AI systems, especially those using models trained on huge amounts of data, can be targets for cyberattacks. This could expose private information or even let bad actors interfere with the AI's results.	Think proactively about external risks
Data Privacy	The data used to train AI could include personal data. Using this data without assessing transparency and control considerations is a risk.	Respect legal obligations and best practices in developing the model
Intellectual Property	Some data is protected by copyright or other intellectual property protections. There are risks that this data cannot be legally or ethically used without a proper license.	Ensure appropriate safeguards and transparency when processing personal data

SAIF: Google's approach to securing AI



Expand strong security foundations to the AI ecosystem



Extend detection and response to bring AI into an organization's threat universe



Automate defenses to keep pace with existing and new threats



Harmonize platform level controls to ensure consistent security across the organization



Adapt controls to adjust mitigations and create faster feedback loops for AI deployment



Contextualize AI system risks in surrounding business processes

Model Armor

≡ Google Cloud team-2-prod-service Search (/) for resources, docs, products, and more

Security / Model Armor / Create template

Security Command Ce...

Risk Overview

Threats

Vulnerabilities

Compliance

Assets

Findings

Sources

Access Insights

Posture Management

Detections and Controls

Google SecOps

reCAPTCHA

Model Armor

Web Security Scanner

Marketplace

Release Notes

← Create template

Detections

Malicious URL detection

Identifies web addresses (URLs) that are designed to harm users or systems. These URLs might lead to phishing sites, malware downloads, or other cyberattacks.

Prompt injection and jailbreak detection

Prompt injection is when a malicious actor tries to insert malicious content into a prompt. A jailbreak attempt is when a malicious actor tries to break out of the model's safety controls. Both can make the AI ignore its usual instructions, reveal sensitive information, bypass an AI model's safeguards, or make it perform actions it wasn't designed to do.

Confidence level

Medium and above

For stricter enforcement, set confidence level to Low and above. This will detect most content that is likely to be a prompt injection and/or jailbreak attempt.

Sensitive data protection

Detects sensitive data and helps prevent its accidental exposure from attacks like prompt injection.

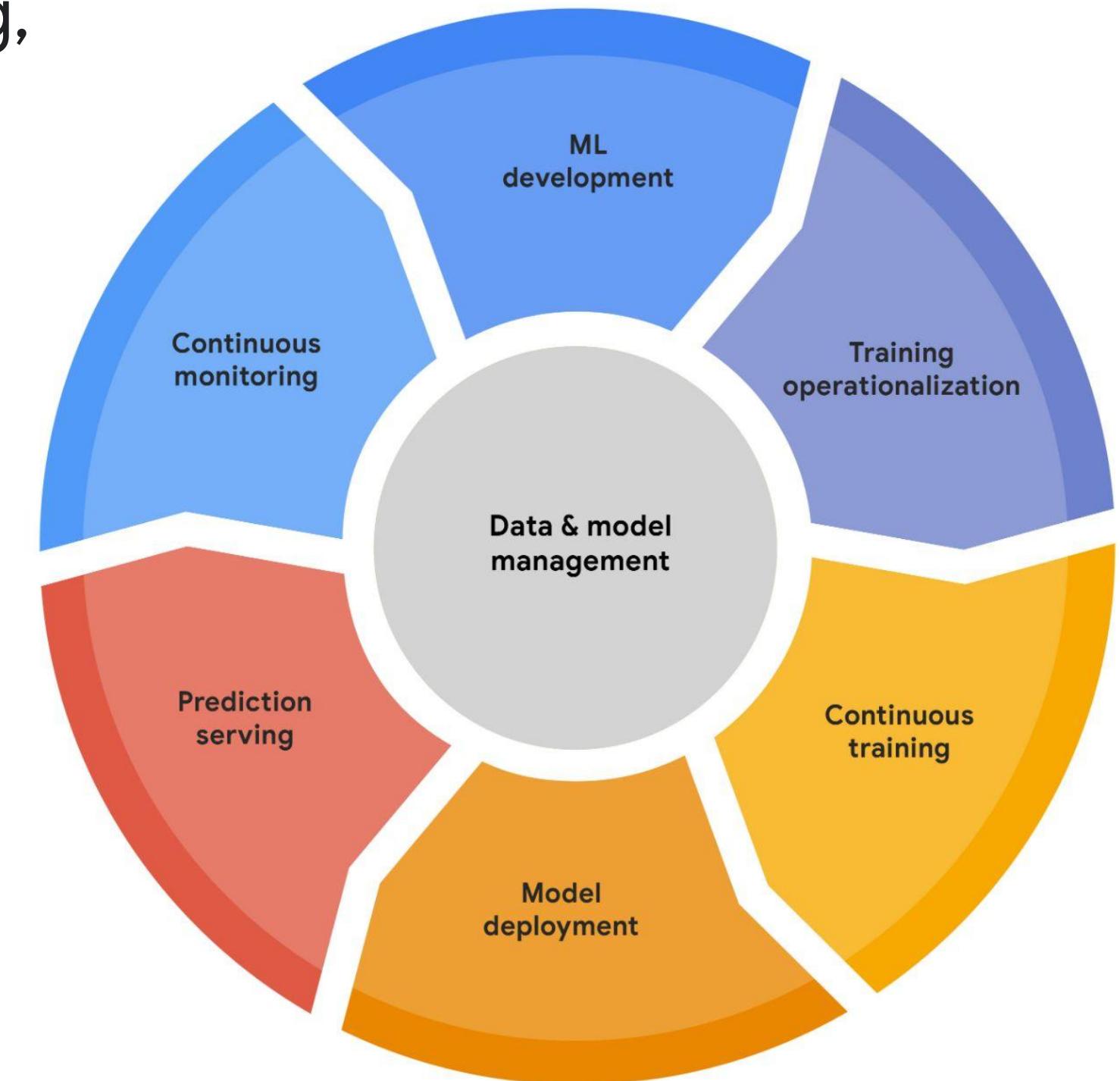
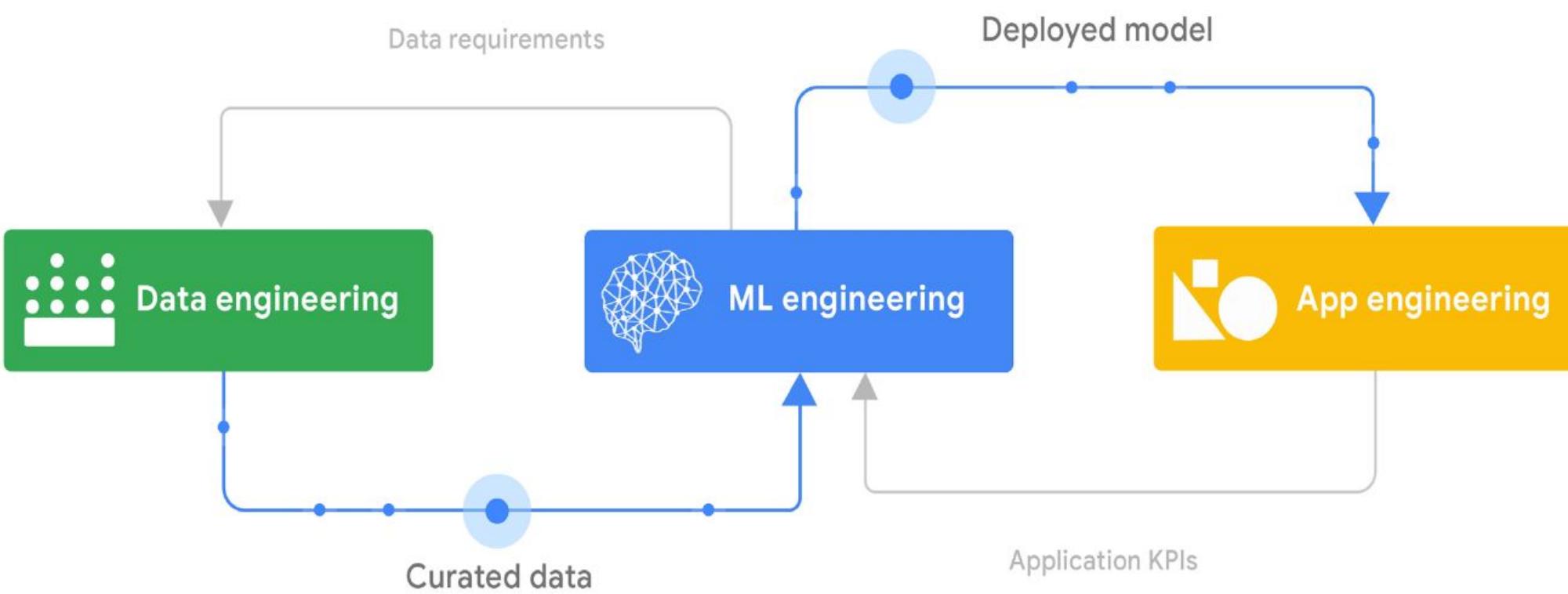
Responsible AI

Confidence level represents how likely it is that the findings match a content filter type. For stricter enforcement, set confidence level to "Low and above" to detect most content that falls into a content filter type.

Customize confidence levels for each content filter below or set confidence level for all content filters.

MLOps

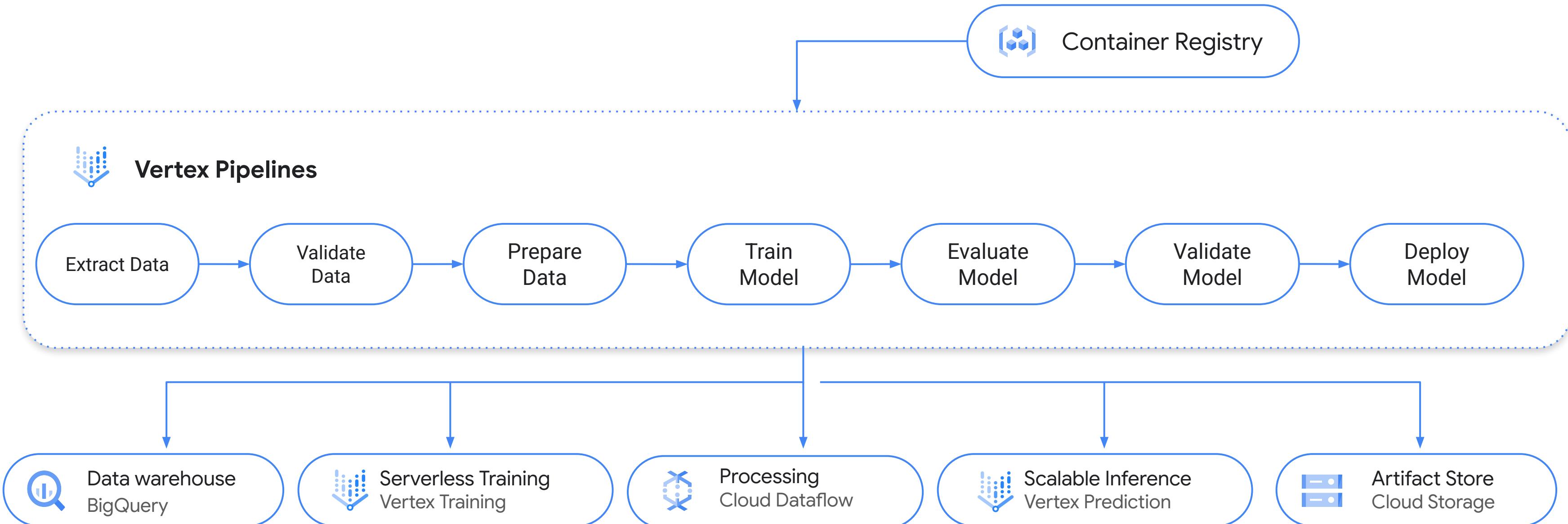
MLOps is an ML engineering culture and practice that aims at unifying a set of **standardized processes and capabilities** for building, deploying, and operationalizing ML systems **rapidly and reliably**.



Full ML Lifecycle Mapping

MLOps Stage	Key Google Cloud Services
1. Gather Data	Pub/Sub, Cloud Storage, Cloud SQL, Cloud Spanner
2. Prepare Data	BigQuery, BigQuery universal catalog
3. Train Model	Vertex AI platform
4. Deploy & Predict	Vertex AI
5. Manage Model	Vertex AI Pipelines, Vertex AI Feature Store, Vertex AI Model Garden

Vertex AI Pipelines: scalable and cost effective



**Optimizing technical and
business processes and
procedures**

Operational excellence

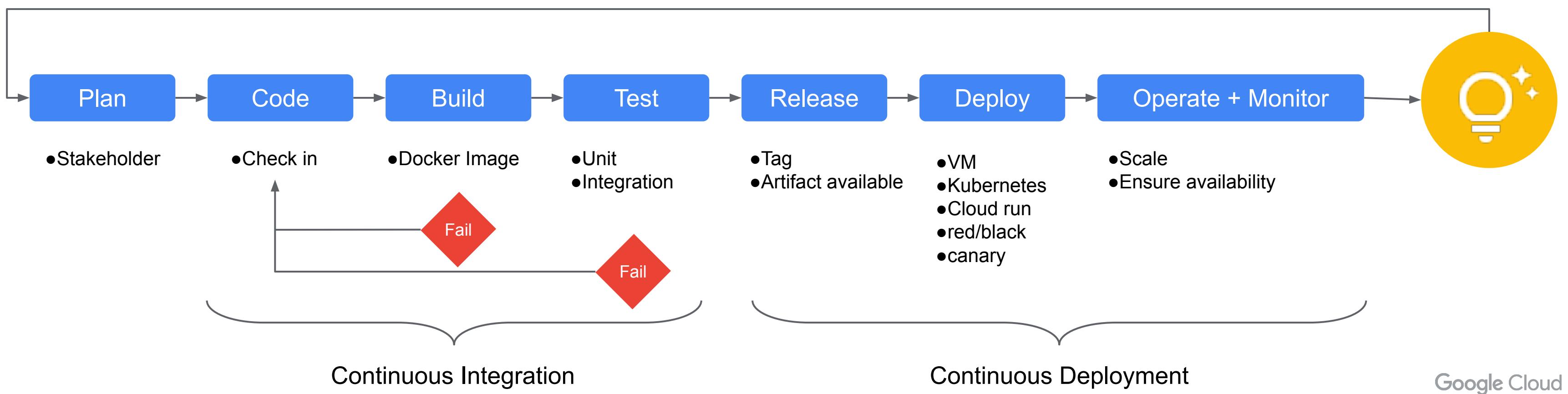
- [CloudOps](#)
- [Managing Incidents](#)
- [Optimizing cloud resources](#)
- [Innovate!!!](#)
- [Secure and responsible AI](#)



Example of process optimization: current CI/CD pipeline

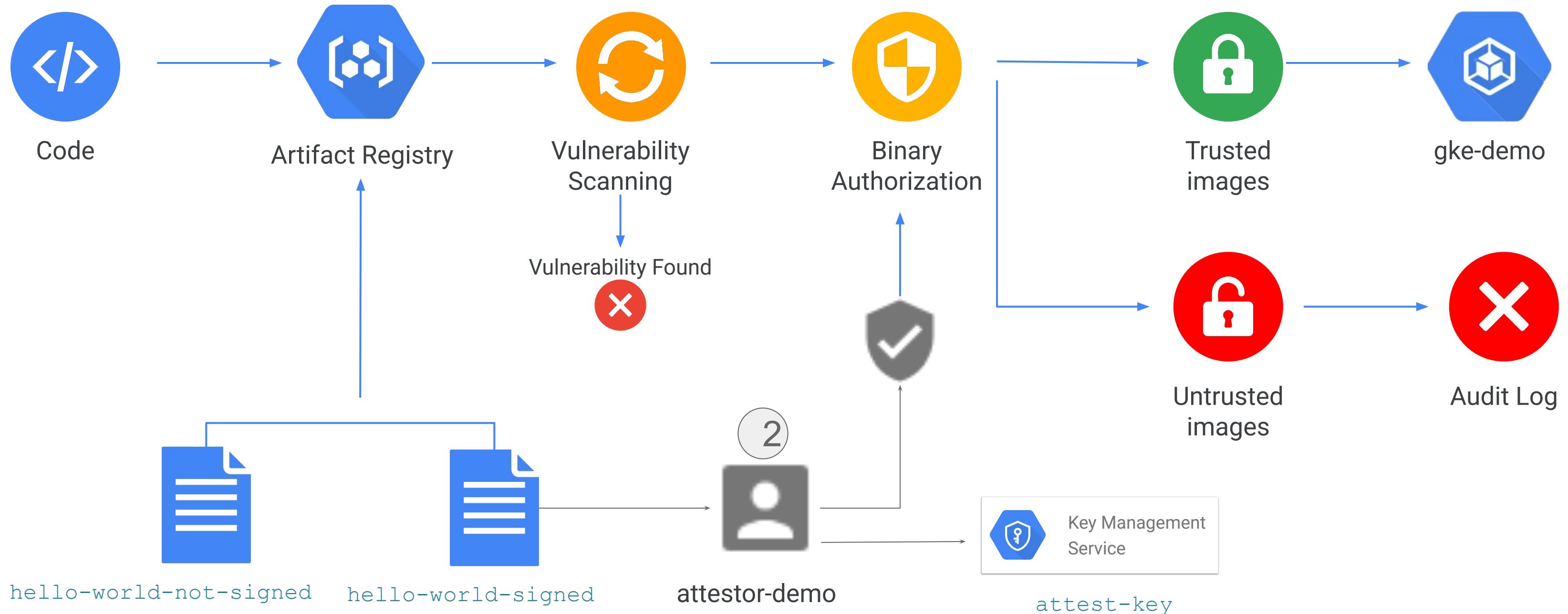
The current build process is:

- Package monolithic application with its dependencies
- Check it in and notify the QA team they need to test it
- Stress test the application to ensure it performs well
- Build a VM image for deployment



Example of process optimization:

Improved CI/CD pipeline



Diagnostic Question Discussion

Your company places a high value on being responsive and meeting customer needs quickly. Their primary business objectives are release speed and agility. You want to reduce the chance of security errors being accidentally introduced.

Which two actions can you take? (Choose two.)

- A. Ensure every code check-in is peer reviewed by a security SME
- B. Use source code security analyzers as part of the CI/CD pipeline
- C. Ensure you have stubs to unit test all interfaces between components
- D. Enable code signing and a trusted binary repository integrated with your CI/CD pipeline
- E. Run a vulnerability security scanner as part of your continuous-integration /continuous-delivery (CI/CD) pipeline

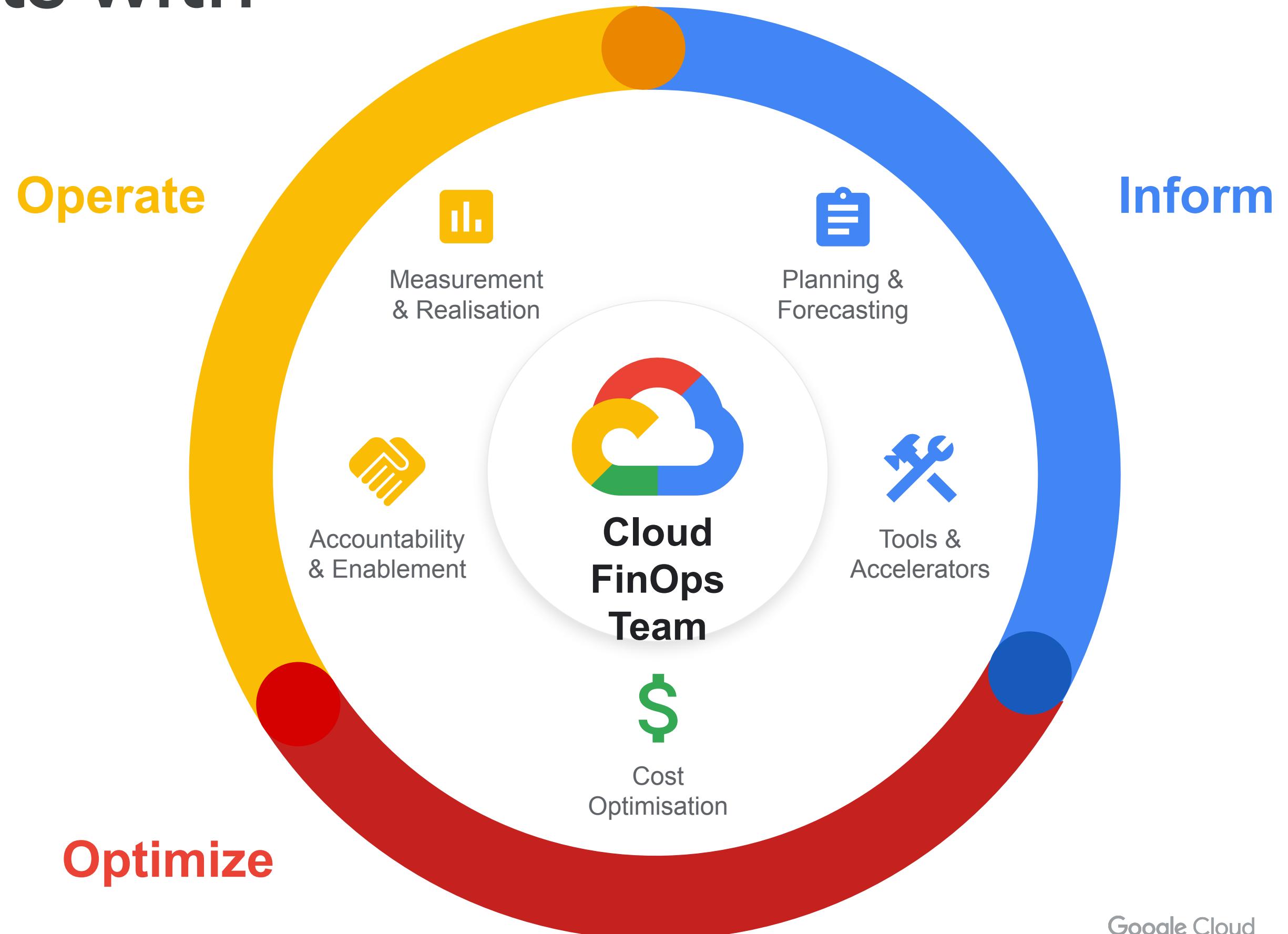
Diagnostic Question Discussion

Your company places a high value on being responsive and meeting customer needs quickly. Their primary business objectives are release speed and agility. You want to reduce the chance of security errors being accidentally introduced.

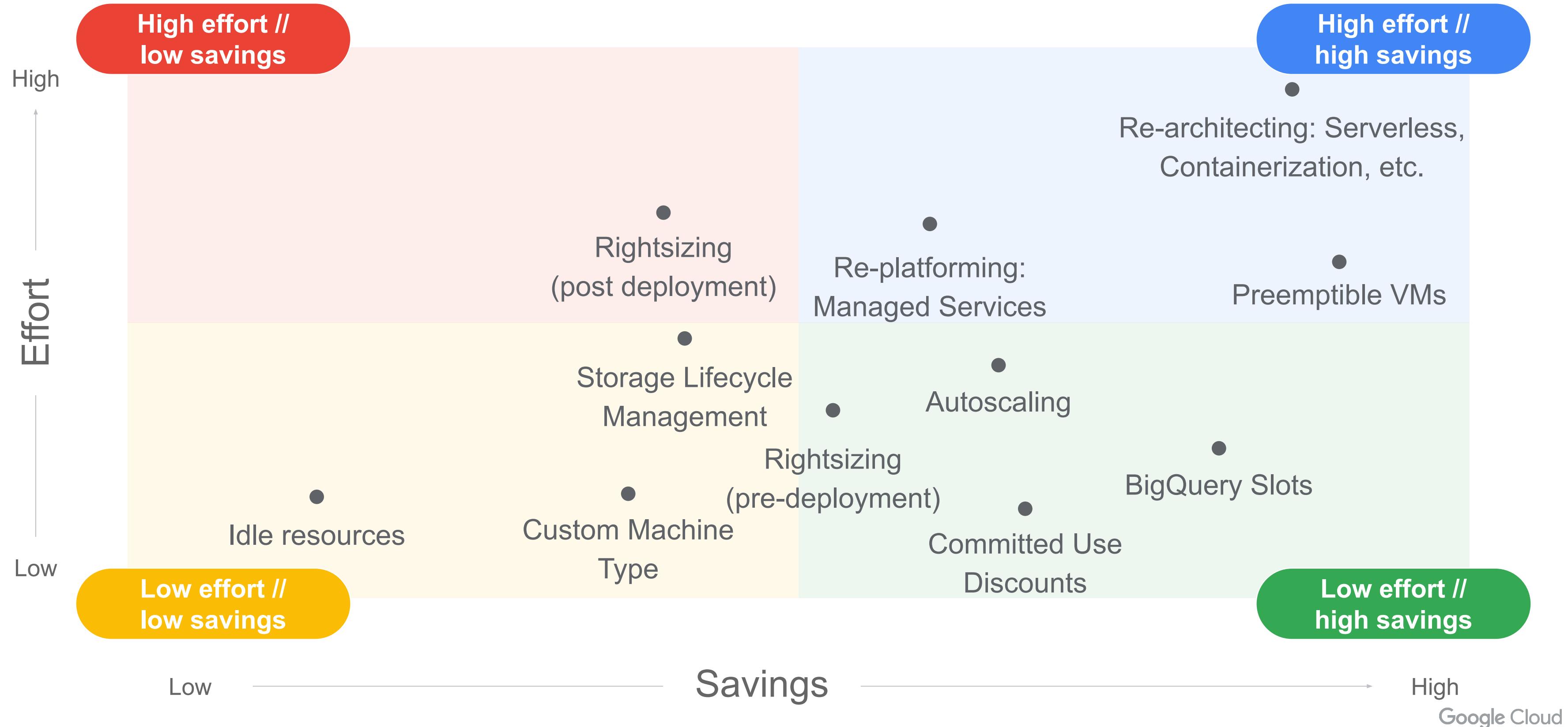
Which two actions can you take? (Choose two.)

- A. Ensure every code check-in is peer reviewed by a security SME
- B. Use source code security analyzers as part of the CI/CD pipeline**
- C. Ensure you have stubs to unit test all interfaces between components
- D. Enable code signing and a trusted binary repository integrated with your CI/CD pipeline
- E. Run a vulnerability security scanner as part of your continuous-integration /continuous-delivery (CI/CD) pipeline**

Optimizing costs with Cloud FinOps Framework



Cost Optimization Matrix



Diagnostic Question Discussion

You are analyzing and defining business processes to support your startup's trial usage of GCP, and you don't yet know what consumer demand for your product will be. Your manager requires you to minimize GCP service costs and adhere to Google best practices.

What should you do?

- A. Utilize free tier and sustained use discounts. Provision a staff position for service cost management.
- B. Utilize free tier and sustained use discounts. Provide training to the team about service cost management.
- C. Utilize free tier and committed use discounts. Provision a staff position for service cost management.
- D. Utilize free tier and committed use discounts. Provide training to the team about service cost management.

Diagnostic Question Discussion

You are analyzing and defining business processes to support your startup's trial usage of GCP, and you don't yet know what consumer demand for your product will be. Your manager requires you to minimize GCP service costs and adhere to Google best practices.

What should you do?

- A. Utilize free tier and sustained use discounts. Provision a staff position for service cost management.
- B. Utilize free tier and sustained use discounts. Provide training to the team about service cost management.**
- C. Utilize free tier and committed use discounts. Provision a staff position for service cost management.
- D. Utilize free tier and committed use discounts. Provide training to the team about service cost management.

Data services (Filestore, Firestore, Memorystore, Spanner, BigQuery, Bigtable)

Filestore

Filestore

Managed NFS, NOT a database



	Filestore Basic	Filestore High Scale	Filestore Enterprise
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%

Diagnostic Question Discussion

You need to deploy a stateful workload on Google Cloud. The workload can scale horizontally, but each instance needs to read and write to the same POSIX filesystem. At high load, the stateful workload needs to support up to 100 MB/s of writes.

What should you do?

- A. Use a persistent disk for each instance.
- B. Use a regional persistent disk for each instance.
- C. Create a Cloud Filestore instance and mount it in each instance.
- D. Create a Cloud Storage bucket and mount it in each instance using gcsfuse.

Diagnostic Question Discussion

You need to deploy a stateful workload on Google Cloud. The workload can scale horizontally, but each instance needs to read and write to the same POSIX filesystem. At high load, the stateful workload needs to support up to 100 MB/s of writes.

What should you do?

- A. Use a persistent disk for each instance.
- B. Use a regional persistent disk for each instance.
- C. Create a Cloud Filestore instance and mount it in each instance.**
- D. Create a Cloud Storage bucket and mount it in each instance using gcsfuse.

Firestore

Firestore: When to use?

Firestore 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	
<u>Offline data persistence</u>		Yes	
Real-time updates		Yes	

Diagnostic Question Discussion

GameCompany wants to move their User Profiles database to Google Cloud Platform.

Which Google Database should they use?

- A. Cloud Spanner
- B. Google BigQuery
- C. Google Cloud Firestore
- D. Google Cloud SQL

Diagnostic Question Discussion

GameCompany wants to move their User Profiles database to Google Cloud Platform.

Which Google Database should they use?

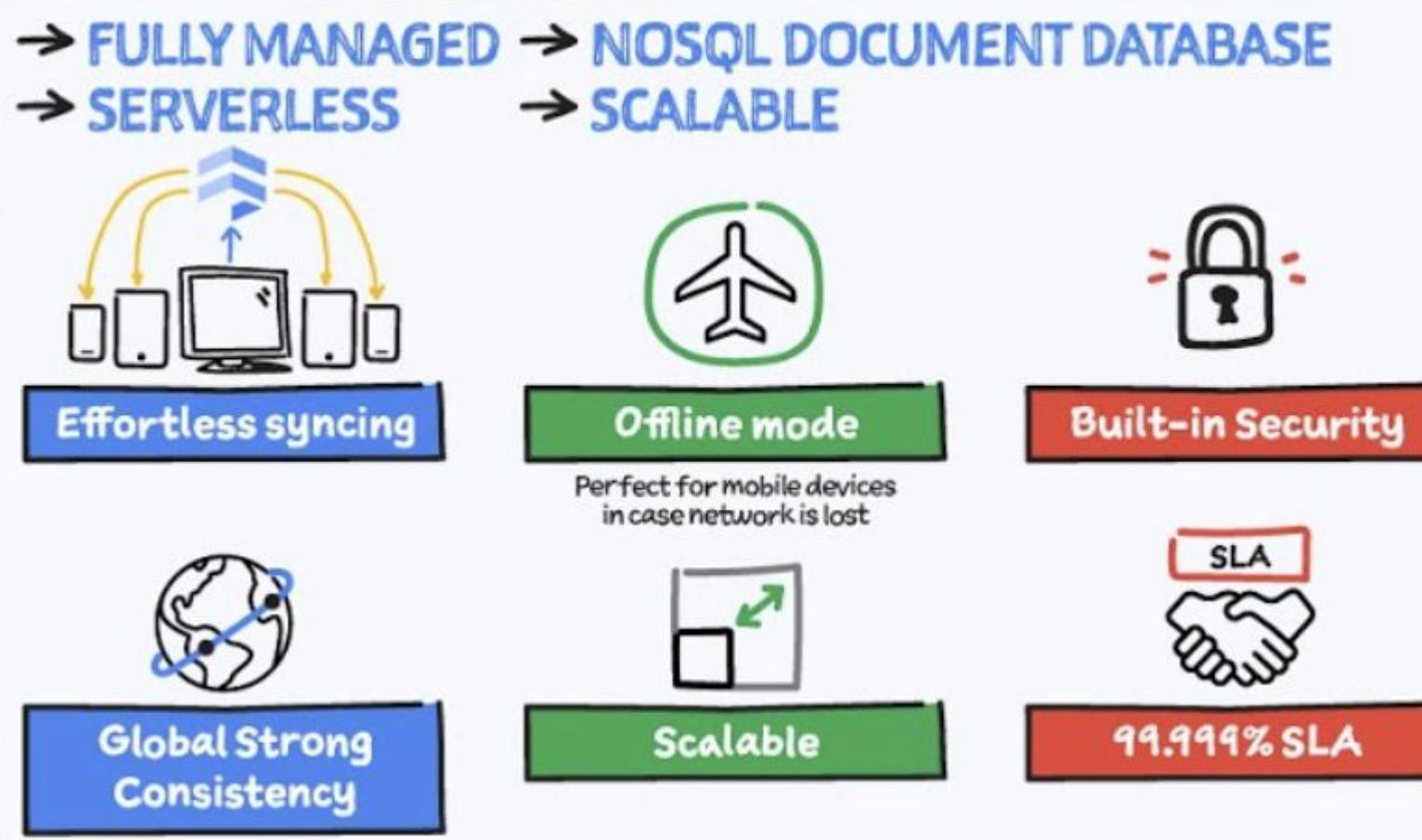
- A. Cloud Spanner
- B. Google BigQuery
- C. Google Cloud Firestore**
- D. Google Cloud SQL

<https://cloud.google.com/products/firestore?hl=en#common-uses>

Firestore

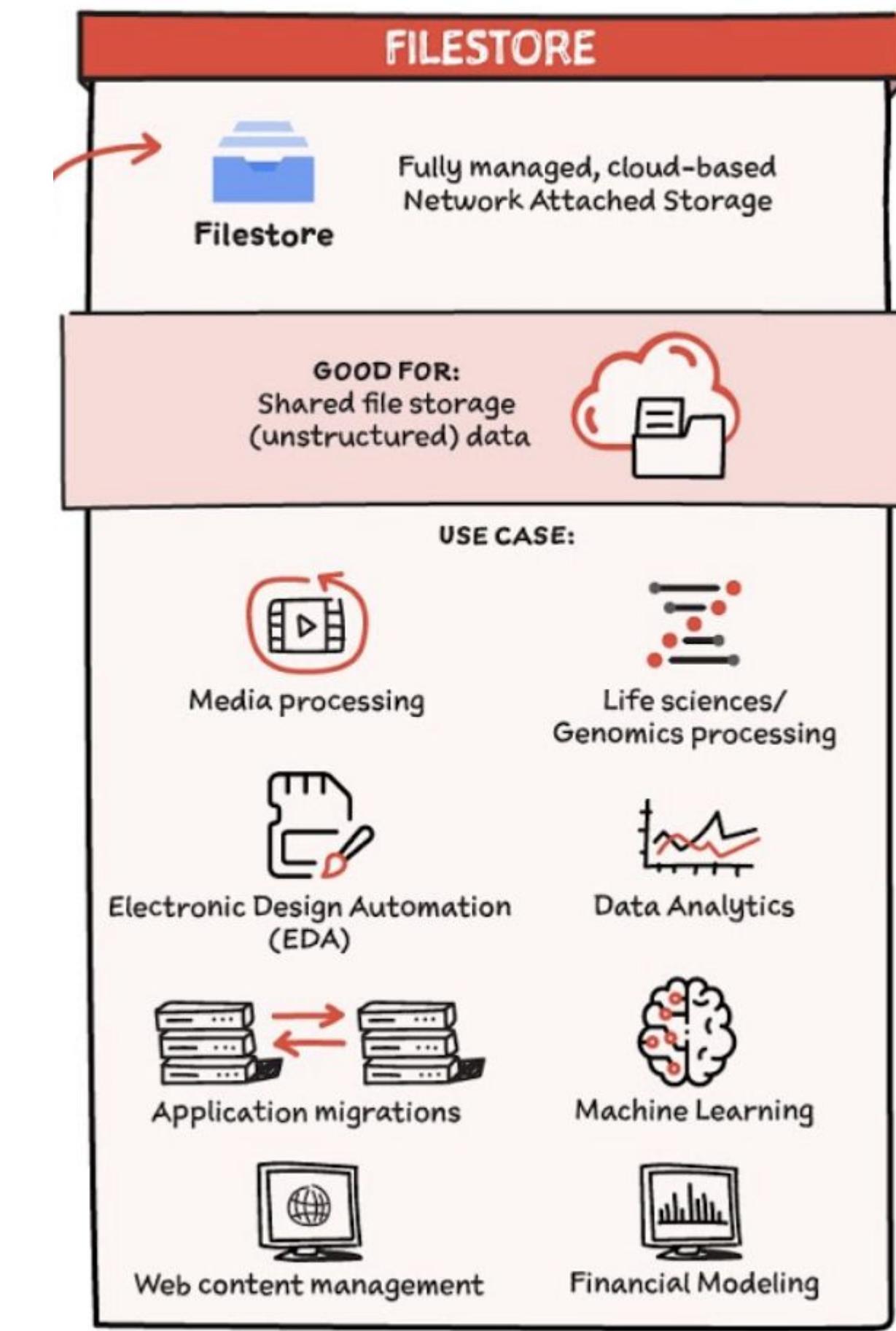
VS

Filestore



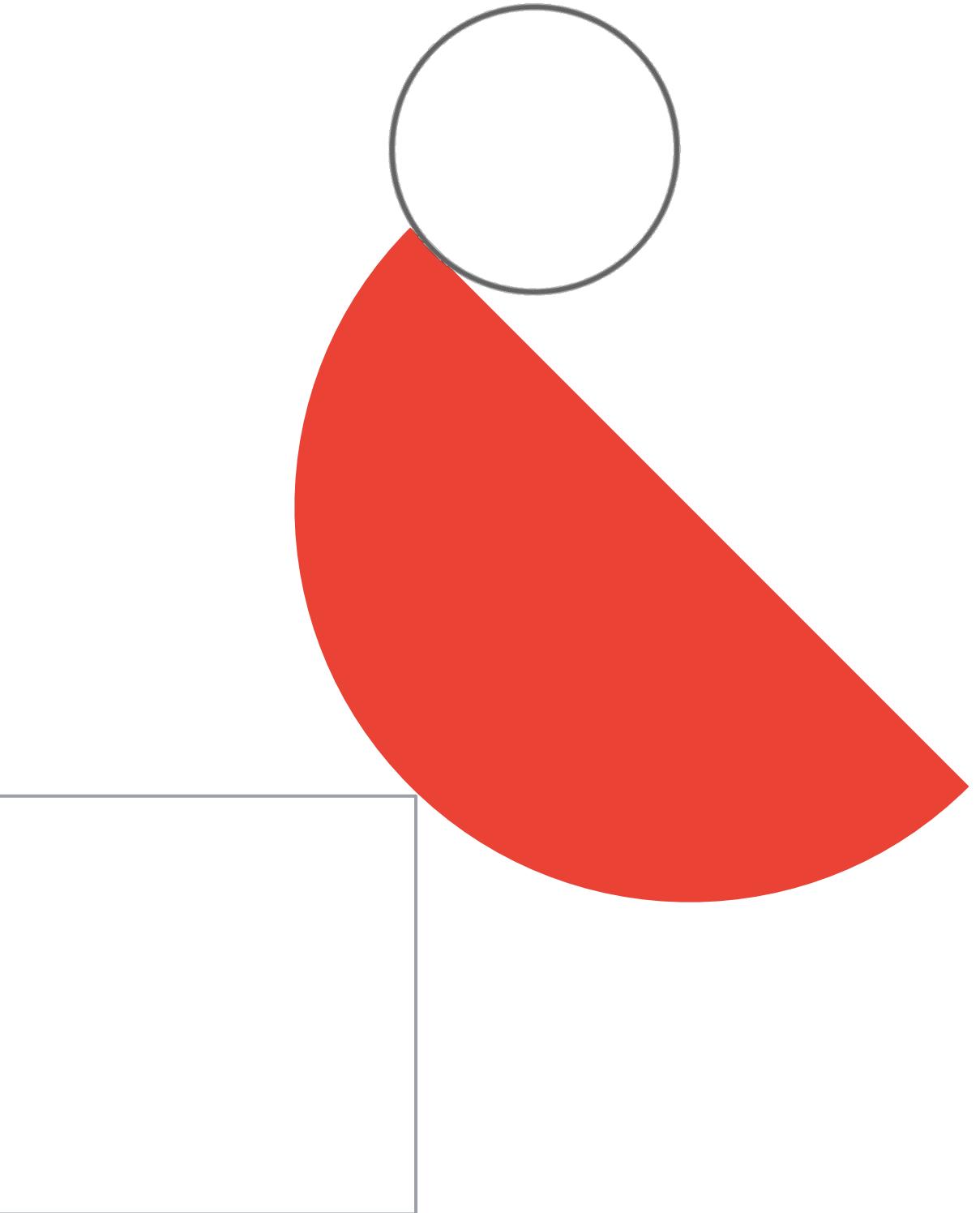
... vs Firebase

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 ***



Firebase is Google's complete app development platform

Complete = it provides different products to:

- Build apps
- Test apps
- Implement authentication ([Firebase Authentication](#) can be a part of PCA exam on very high-level!)

- Run apps
- Run analytics
- Personalize apps
- And more...



iOS



Android



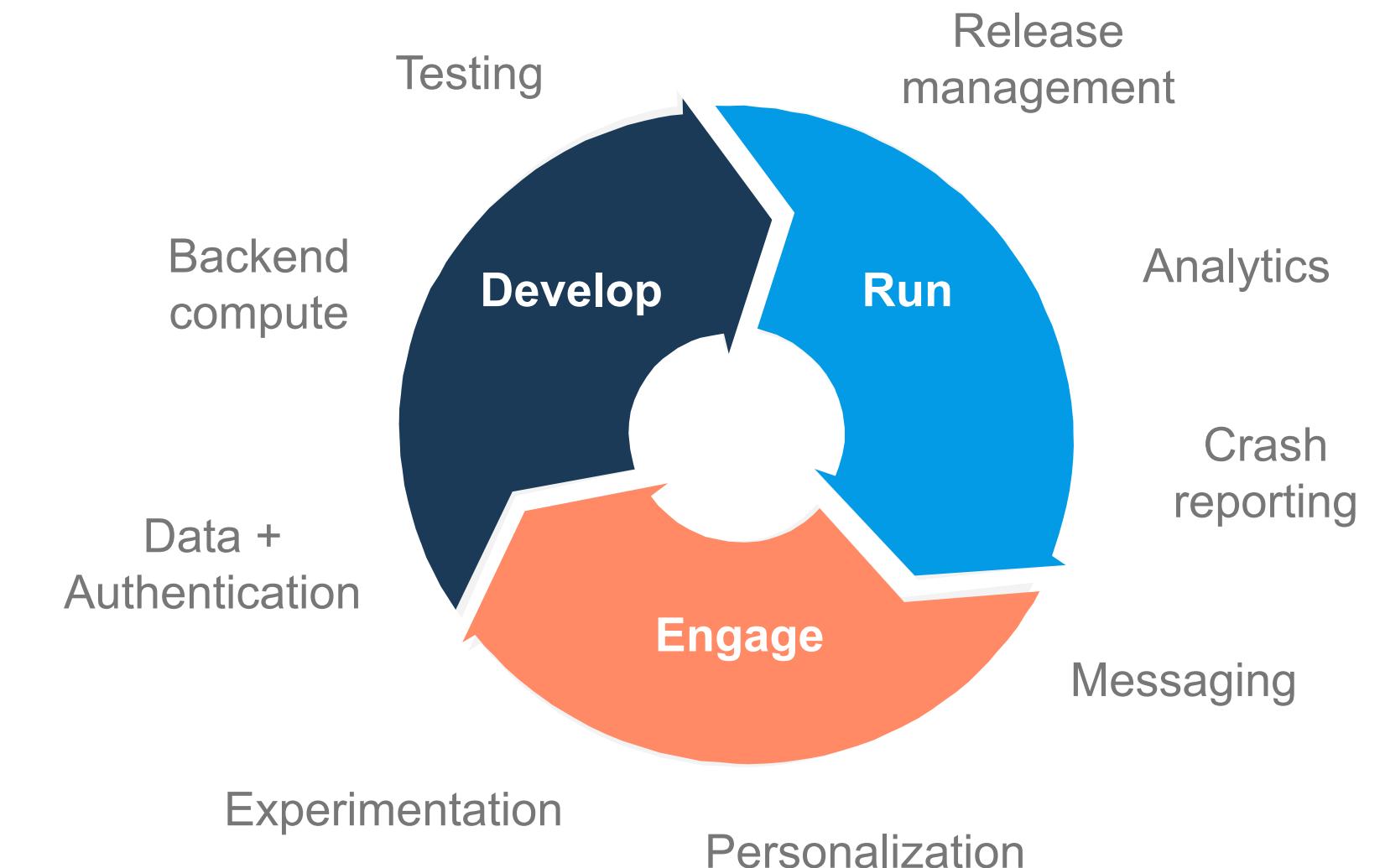
Web



C++



Unity



Exam Tip: Firestore is usually a part of Firebase-based app (for storing and syncing data)

Google Cloud

Diagnostic Question Discussion

Your development team has created a mobile game app. You want to test the new mobile app on Android and iOS devices with a variety of configurations.

You need to ensure that testing is efficient and cost-effective.

What should you do?

- A. Create Android and iOS VMs on Google Cloud, install the mobile app on the VMs, and test the mobile app.
- B. Upload your mobile app to the Firebase Test Lab, and test the mobile app on Android and iOS devices.
- C. Create Android and iOS containers on Google Kubernetes Engine (GKE), install the mobile app on the containers, and test the mobile app.
- D. Upload your mobile app with different configurations to Firebase Hosting and test each configuration.

Diagnostic Question Discussion

Your development team has created a mobile game app. You want to test the new mobile app on Android and iOS devices with a variety of configurations.

You need to ensure that testing is efficient and cost-effective.

What should you do?

- A. Create Android and iOS VMs on Google Cloud, install the mobile app on the VMs, and test the mobile app.
- B. Upload your mobile app to the Firebase Test Lab, and test the mobile app on Android and iOS devices.**
- C. Create Android and iOS containers on Google Kubernetes Engine (GKE), install the mobile app on the containers, and test the mobile app.
- D. Upload your mobile app with different configurations to Firebase Hosting and test each configuration.

<https://firebase.google.com/docs/test-lab>

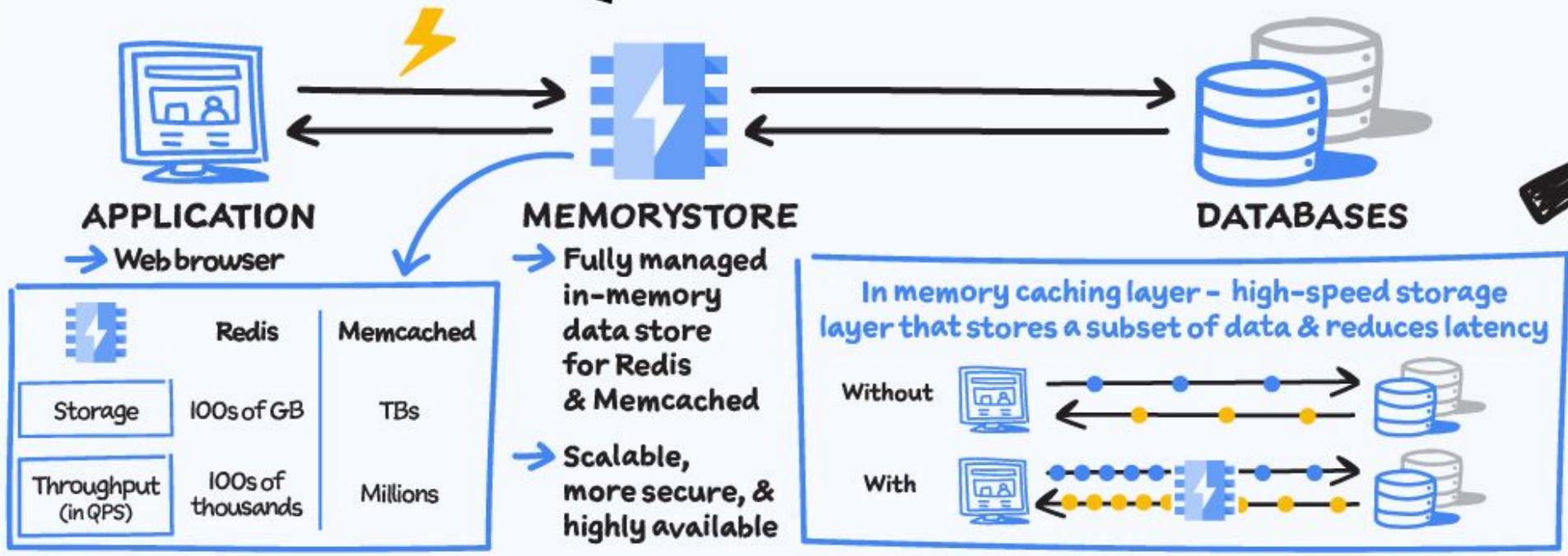
Memorystore



Memorystore #GCPSketchnote

@PVERGADIA THECLOUDGIRL.DEV 6.29.2021

What is Memorystore?



CLOUD MEMORYSTORE USE CASES



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



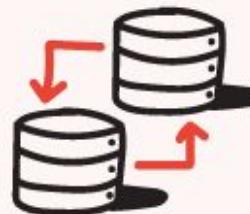
FEATURES & CAPABILITIES



SECURE BY DEFAULT



SEAMLESS SCALE & HA



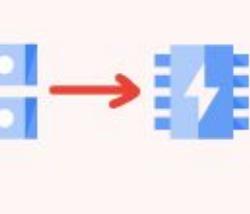
DEEP INSIGHTS



BACKUP DATA



NO CODE CHANGES



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

Instance Auth, Data encrypted in-transit

Standard high availability instances are replicated across zones

Monitor instances using cloud operations

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

OSS compliance allows using Memorystore without any code changes

Spanner

What workloads Cloud Spanner fits 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

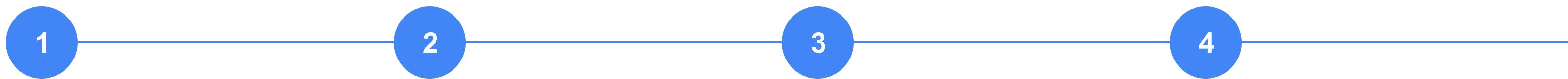
Write once and automatically replicate your data to multiple regions.

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 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



Relational

Horizontal Scale

99.999% SLA, fully managed, and scalable

Non-Relational



How does Spanner provide global consistency? >>>

SPANNER GLOBAL CONSISTENCY

TrueTime

Synchronizes clocks in all machines across datacenters



Google's Global Network

Fast & redundant



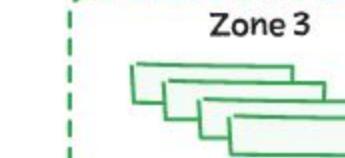
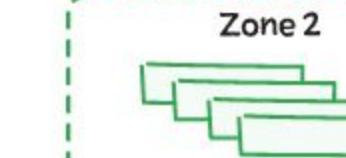
How does Cloud Spanner work?

This Spanner instance contains 4-nodes

REGIONAL INSTANCE

Compute nodes

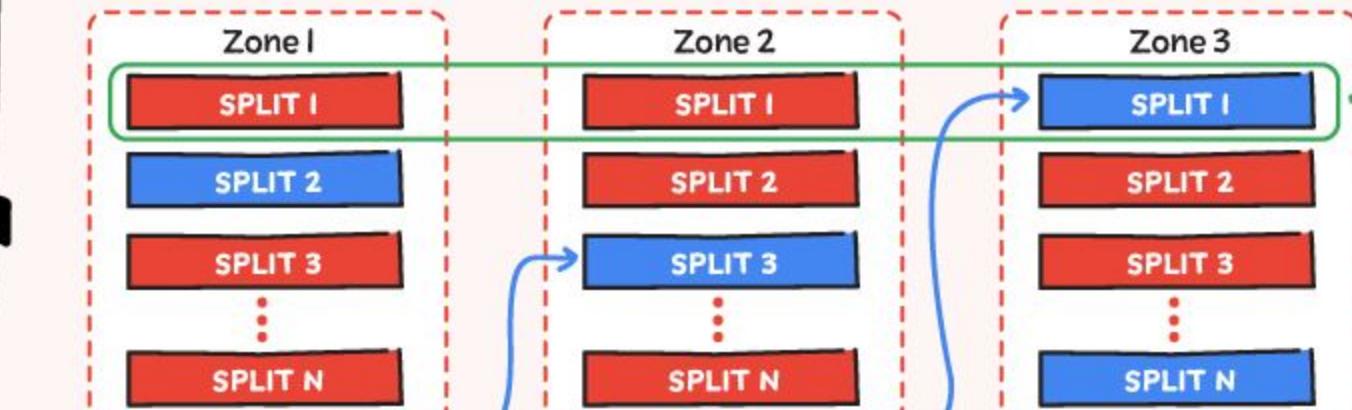
Storage



This Spanner instance is hosting 2 databases across 3 zones

How does Spanner provide high availability* & scalability*

Zero downtime for planned maintenance or schema changes



5 Different splits can have leaders in different zones

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

Leader manages writes for that split

All replicas serve reads

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	[$-\infty, 3]$	[4, 224]	...	[2457, ∞)

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

3 Paxos group for split 1

Diagnostic Question Discussion

Your organization is developing an application that will manage payments and online bank accounts located around the world. The most critical requirement for your database is that each transaction is handled consistently. Your organization anticipates almost unlimited growth in the amount of data stored.

- A. Cloud SQL
- B. Cloud Storage
- C. Cloud Firestore
- D. Cloud Spanner

Which Google Cloud product should your organization choose?

Diagnostic Question Discussion

Your organization is developing an application that will manage payments and online bank accounts located around the world. The most critical requirement for your database is that each transaction is handled consistently. Your organization anticipates almost unlimited growth in the amount of data stored.

- A. Cloud SQL
- B. Cloud Storage
- C. Cloud Firestore
- D. **Cloud Spanner**

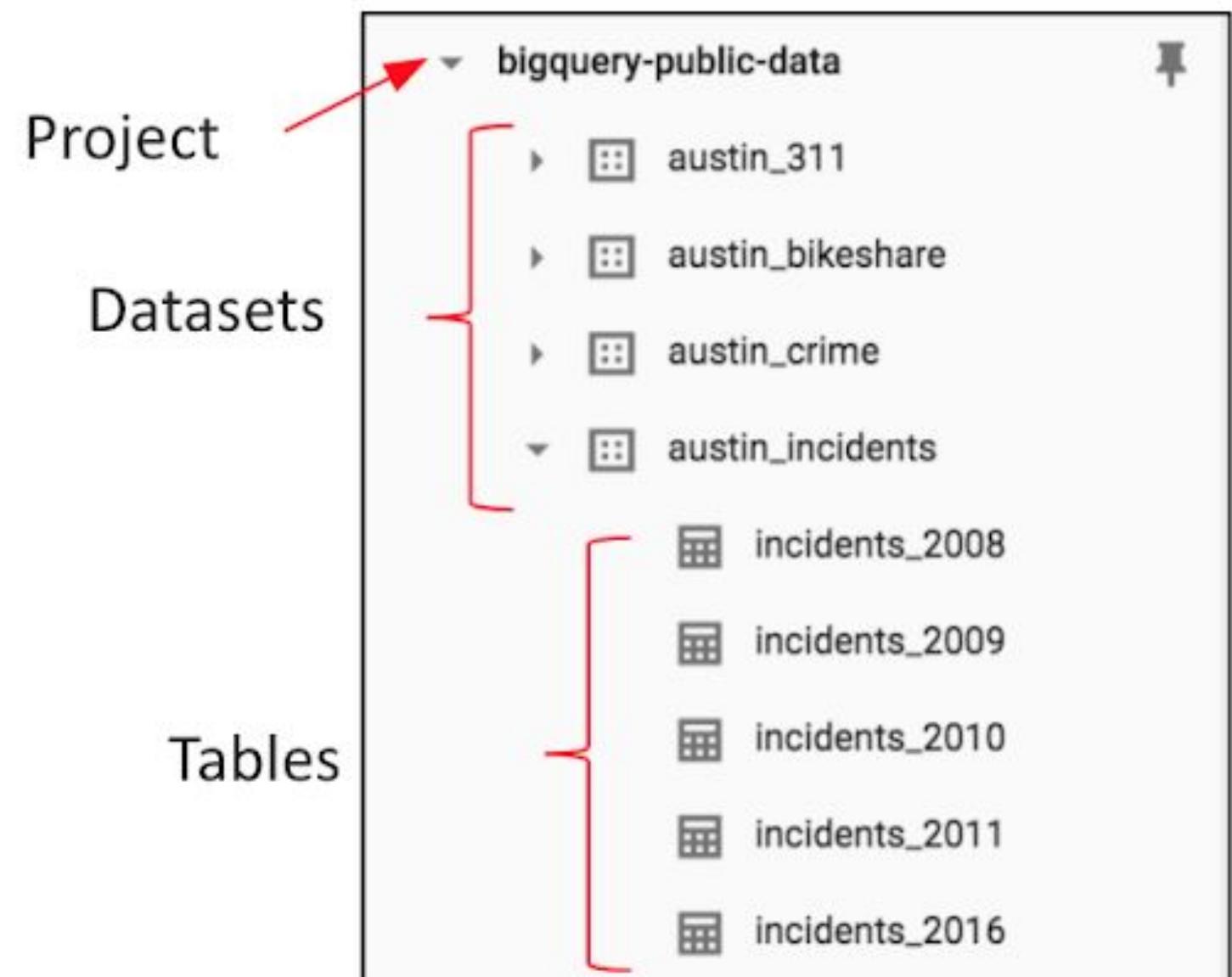
Which Google Cloud product should your organization choose?

Bigquery

BigQuery hierarchy

Project -> Dataset -> Tables (-> Partitions)

- For each query, BigQuery executes a full-column scan.
- BigQuery performance and query costs are based on the amount of data scanned.
- You can set the **geographic location of a Dataset** at creation time only.
- All tables that are referenced in a query must be stored in datasets in the same location.
- When you copy a table (bq cp), the datasets that contain the source table and destination table must reside in the same location.
 - You can copy a dataset (NOT with bq cp, but with BigQuery Data Transfer Service) within a region or from one region to another
- Dataset names are case-sensitive



BigQuery: Controlling access to datasets

Common BigQuery predefined roles

Exam Tips: It's a common practice to have a Dataset in one project and perform queries from another one (split billing!).

Admin	Full Access to all datasets
Data Editor	Access to edit all contents of the datasets
Data Owner	Full access to datasets and all of their contents
Data Viewer	Access to view datasets and all of their contents
Job User	Access to run jobs
Metadata Viewer	Access to view table and dataset metadata
User	Access to run queries and create datasets
Read Sessions User	Access to create and use read sessions

Capability	<u>dataViewer</u>	<u>dataEditor</u>	<u>dataOwner</u>	<u>user</u>	<u>jobUser</u>	<u>admin</u>
List/get projects	✓	✓	✓	✓	✓	✓
List tables	✓	✓	✓	✓	✗	✓
Get table data/metadata	✓	✓	✓	✗	✗	✓
Create tables	✗	✓	✓	✗	✗	✓
Modify/delete tables	✗	✓	✓	✗	✗	✓
List/get datasets	✓	✓	✓	✓	✗	✓
Create new datasets	✗	✓	✓	✓	✗	✓
Modify/delete datasets	✗	✗	✓		Self-created datasets	✗
Create jobs/queries	✗	✗	✗	✓	✓	✓
Cancel jobs	✗	✗	✗		Self-created jobs	Self-created jobs
Get/list saved queries	✗	✗	✗	✓	✗	✓
Create/update/delete saved queries	✗	✗	✗	✗	✗	✓
Get transfers	✗	✗	✗	✓	✗	✓
Create/update/delete transfers	✗	✗	✗	✗	✗	✓

BigQuery: Controlling access to datasets

You can grant access at the following BigQuery resource levels:

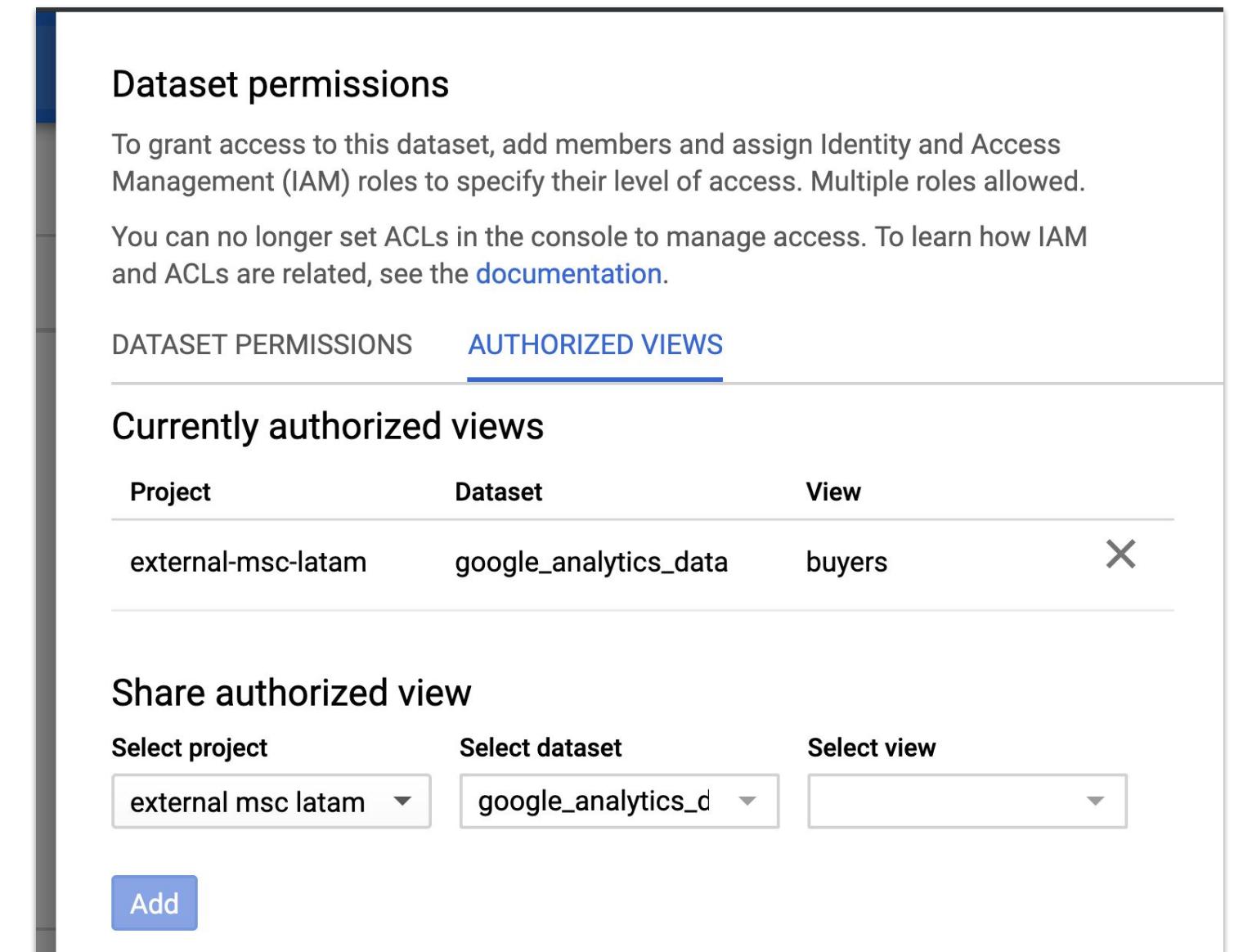
- organization or Google Cloud project level
- dataset level
- table or view level
 - a. [Authorized Views](#)
- You can also restrict access to data on more granular level by using the following methods:
 - a. [column-level access control](#)
 - b. [dynamic data masking](#) (aka “some **columns** may be hidden, depending on privileges”)
 - i. Works together with column-level security.
 - ii. no need to modify existing queries by excluding the columns that the user cannot access
 - c. [row-level security](#) (aka “some rows may be hidden, depending on privileges”)
 - i. One table can have multiple row-level access policies. Row-level access policies can coexist on a table with column-level security as well as dataset-level, table-level, and project-level access controls.

BigQuery: Controlling access to datasets

Authorized Views

1. **View:** View is a virtual table defined by a SQL query. When you create a view, you query it in the same way you query a table
2. **Query:** When a user queries the view, the query results contain data only from the tables and fields specified in the query that defines the view.
3. **Authorized Views:** An authorized view allows you to share query results with particular users and groups without giving them access to the underlying tables.

Exam Tip: Authorized Views were especially useful when there were no table/column-level permissions. However, they're still often-used way to selectively share access to datasets (and they pop up on the exam!).
MAKE SURE TO UNDERSTAND [HOW TO CREATE AND SHARE SUCH A VIEW.](#)



The screenshot shows the 'Dataset permissions' page for a dataset named 'google_analytics_data'. The 'AUTHORIZED VIEWS' tab is selected. It displays a table with one row:

Project	Dataset	View
external-msc-latam	google_analytics_data	buyers

Below the table is a 'Share authorized view' section with three dropdown menus: 'Select project' (set to 'external msc latam'), 'Select dataset' (set to 'google_analytics_d'), and 'Select view' (empty). A blue 'Add' button is located at the bottom left of this section.

BigQuery - Data Transfer Service

Mostly useful for **regular** data transfers to BigQuery

- BigQuery Data Transfer Service automates data movement **from** various sources **into** BigQuery on a scheduled, managed basis.
- You can initiate data backfills to recover from any outages or gaps.

The screenshot shows the 'Create transfer' interface in the BigQuery Data Transfer Service. At the top, there is a back arrow labeled 'Create transfer'. Below it, a section titled 'Source type' with the sub-instruction 'Choose a data source from the list below'. A dropdown menu is open, showing a list of available sources. The menu includes a 'Filter Type to filter' option and a list of sources: Amazon S3, Campaign Manager (formerly DCM), Dataset Copy, Google Ad Manager (formerly DFP), Google Ads - Preview, Google Ads (formerly AdWords), Google Cloud Storage, and Google Merchant Center. At the bottom of the dropdown, there is a link 'Can't find what you're looking for? [Explore Data Sources](#)'.

Source *

Filter Type to filter

Amazon S3

Campaign Manager (formerly DCM)

Dataset Copy

Google Ad Manager (formerly DFP)

Google Ads - Preview

Google Ads (formerly AdWords)

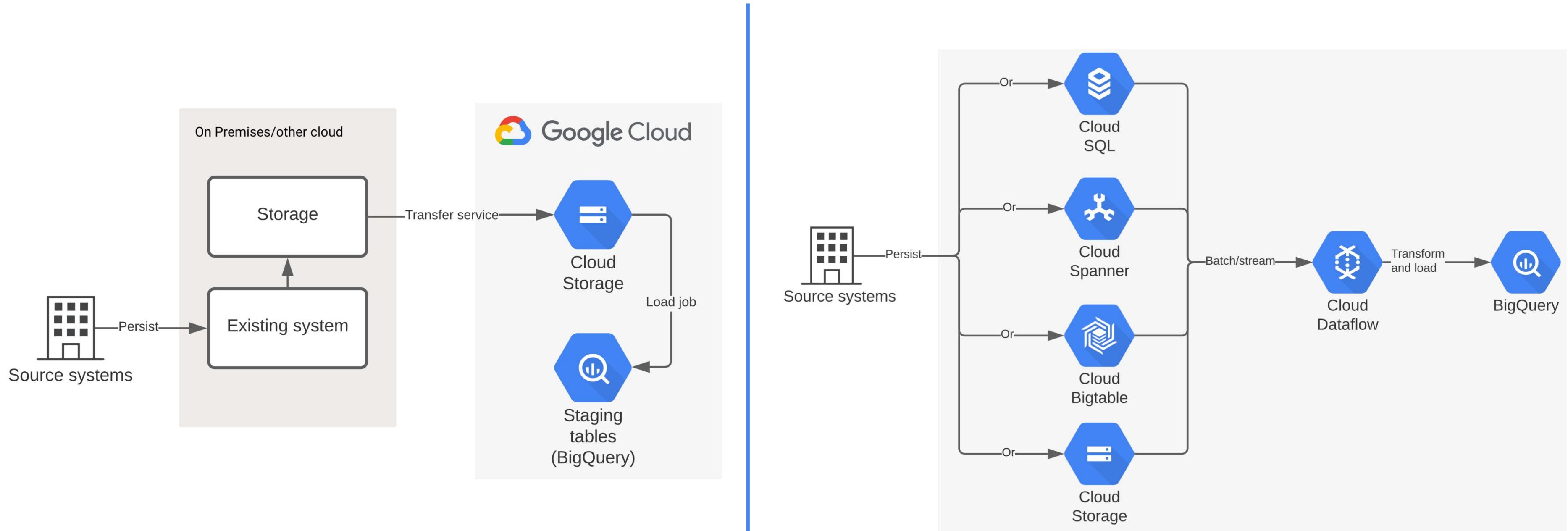
Google Cloud Storage

Google Merchant Center

Can't find what you're looking for? [Explore Data Sources](#)

BigQuery - Batch vs Streaming inserts

Most common architectures



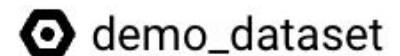
Exam Tip: There is additional cost for streaming (both inserts and reads) in BigQuery.

BigQuery: Sharing Datasets with others

AllAuthenticatedUsers

The special setting [allAuthenticatedUsers](#) makes a dataset public. Authenticated users must use BigQuery within their own project and have access to run BigQuery jobs so that they can query the Public Dataset. The billing for the query goes to their project, even though the query is using public or shared data. In summary, [the cost of a query is always assigned to the active project from where the query is executed.](#)

Resource



demo_dataset

Add principals

Principals are users, groups, domains, or service accounts. [Learn more about principals in IAM](#)

New principals

allAuthenticatedUsers



Assign roles

Roles are composed of sets of permissions and determine what the principal can do with this resource. [Learn more](#)

Role *

BigQuery Data Viewer



Access to view datasets and all of their contents

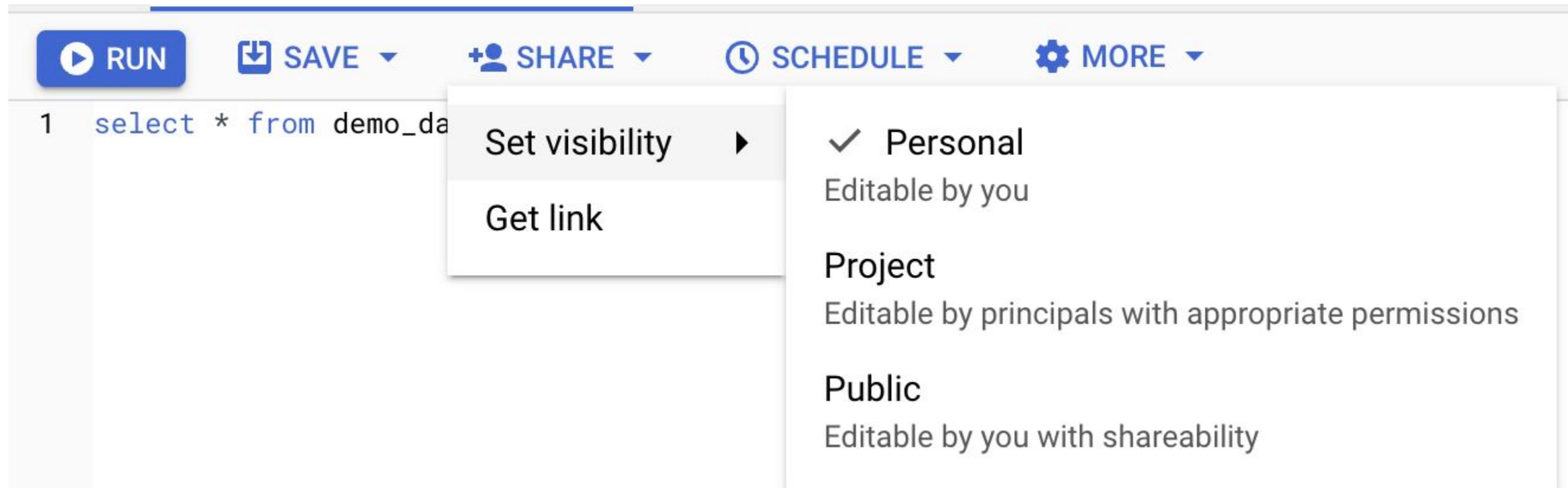
+ ADD ANOTHER ROLE

SAVE

CANCEL

BigQuery: Sharing **Queries** with others

Mostly for collaboration



- Query needs to be saved first, before it's shared;
- Can share incomplete / invalid queries -> collaboration;
- Project-level saved queries are visible to principals with the required [permissions](#);
- Public saved queries are visible to anyone with a link to the query;

BigQuery: Scheduling queries

Mostly useful for regular execution

- Scheduled queries use features of BigQuery Data Transfer Service.
- If the destination table for your results doesn't exist when you set up the scheduled query, BigQuery attempts to create the table for you.
- You can set up a scheduled query to authenticate as a service account.

Details and schedule

Name for scheduled query *
scheduled_query_1

Schedule options

Repeats *

Daily

At *

13:00

UTC

Start now Start at set time

Start date and run time

1/15/23, 9:08 AM

CET 

End never Schedule end time

End date

CET 

Destination for query results

Set a destination table for query results

Dataset *

sapongcp-320306.demo_dataset

SAVE

CANCEL

BigQuery: Query results **caching**

Limit Access by Data Lifecycle Stages

- Query results are cached to improve performance and reduce costs for repeated queries
- Cache is per user
- Still subject to quota policies
- Cache results have a size limit of 128 MB compressed
- No charge for queries that use cached results
- Results are cached for approximately 24 hours
- Lifetime extended when a query returns a cached result
- Use of cached results can be turned off (useful for benchmarking)

BigQuery: table/partition (automatic) data expiration

Can be set for dataset / table / partition

Best practice for data lifecycle management.

Expiration in BigQuery automatically implements retention policy.

- [Dataset expiration](#)
 - = “default table expiration time” for a dataset
- [Table expiration](#)
 - If Dataset expiration is set, each table inherits this setting by default
- [Partition expiration](#):
 - The setting applies to all partitions in the table, but is calculated independently for each partition based on the partition time.
 - At any point after a table is created, you can update the table's partition expiration

Dataset info

Dataset ID	simoahava-com.analytics_206575074
Created	Aug 27, 2019, 2:44:32 PM UTC+3
Default table expiration	60 days
Last modified	Nov 15, 2022, 11:05:11 AM UTC+2
Data location	EU

BigQuery: Table Partitioning

Partitioning versus sharding:

- Table sharding is the practice of storing data in multiple tables, using a naming prefix such as [PREFIX]_YYYYMMDD. **Partitioning is recommended over table sharding, because partitioned tables perform better.**

You can partition BigQuery tables by:

- Time-unit column: Tables are partitioned based on a TIMESTAMP, DATE, or DATETIME column in the table.
- Ingestion time: Tables are partitioned based on the timestamp when BigQuery ingests the data.
- Integer range: Tables are partitioned based on an integer column.

c2	c3	eventDate
		2018-01-01
		2018-01-02
		2018-01-03
		2018-01-04
		2018-01-05

```
SELECT * FROM ...
WHERE eventDate BETWEEN
"2018-01-03" AND
"2018-01-04"
```

BigQuery: Table Clustering

c1	userId	c3	
			2018-01-01
			2018-01-02
			2018-01-03
			2018-01-04
			2018-01-05

```
SELECT c1, c3 FROM ... WHERE userId BETWEEN 52 and 63  
AND eventDate BETWEEN "2018-01-03" AND "2018-01-04"
```

BigQuery - table partitioning vs clustering

Decision making

- Clustering gives you more granularity than partitioning alone allows
- Use clustering if your queries commonly use filters or aggregation against multiple particular columns.

Use case	Recommendation
You're using on-demand pricing and require strict cost guarantees before running queries.	Partitioned tables
Your segment size is less than 1 GB after partitioning the table.	Clustered tables
You require a large number of partitions beyond the BigQuery limits	Clustered tables
Frequent mutations in your data modify a large number of partitions.	Clustered tables
You frequently run queries to filter data on certain fixed columns.	Partitions plus clustering

Diagnostic Question Discussion

To be compliant with European GDPR regulation, Cymbal Bank is required to delete data generated from its European customers after a period of 36 months when it contains personal data. In the new architecture, this data will be stored in both Cloud Storage and BigQuery.

What should you do?

- A. Create a BigQuery table for the European data, and set the table retention period to 36 months. For Cloud Storage, use gsutil to enable lifecycle management using a DELETE action with an Age condition of 36 months.
- B. Create a BigQuery table for the European data, and set the table retention period to 36 months. For Cloud Storage, use gsutil to create a SetStorageClass to NONE action when with an Age condition of 36 months.
- C. Create a BigQuery time-partitioned table for the European data, and set the partition expiration period to 36 months. For Cloud Storage, use gsutil to enable lifecycle management using a DELETE action with an Age condition of 36 months.
- D. Create a BigQuery time-partitioned table for the European data, and set the partition expiration period to 36 months. For Cloud Storage, use gsutil to create a SetStorageClass to NONE action with an Age condition of 36 months.

Diagnostic Question Discussion

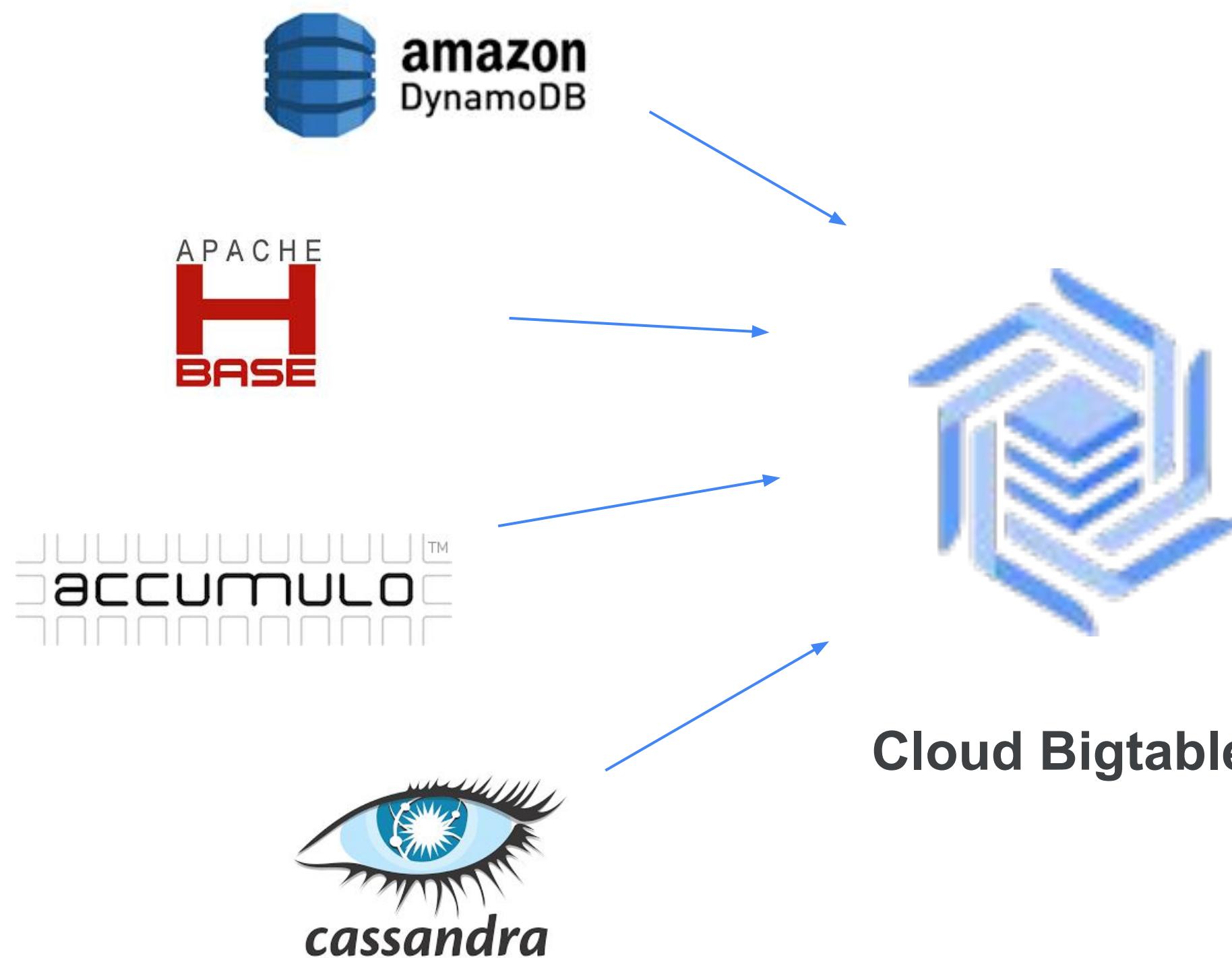
To be compliant with European GDPR regulation, Cymbal Bank is required to delete data generated from its European customers after a period of 36 months when it contains personal data. In the new architecture, this data will be stored in both Cloud Storage and BigQuery.

What should you do?

- A. Create a BigQuery table for the European data, and set the table retention period to 36 months. For Cloud Storage, use gsutil to enable lifecycle management using a DELETE action with an Age condition of 36 months.
- B. Create a BigQuery table for the European data, and set the table retention period to 36 months. For Cloud Storage, use gsutil to create a SetStorageClass to NONE action when with an Age condition of 36 months.
- C. **Create a BigQuery time-partitioned table for the European data, and set the partition expiration period to 36 months. For Cloud Storage, use gsutil to enable lifecycle management using a DELETE action with an Age condition of 36 months.**
- D. Create a BigQuery time-partitioned table for the European data, and set the partition expiration period to 36 months. For Cloud Storage, use gsutil to create a SetStorageClass to NONE action with an Age condition of 36 months.

Bigtable

Bigtable is a common migration target for **key-value**, **wide-column** and **time-series databases**



- Petabyte-scale
- **fully managed NoSQL database service** for use cases where **low latency** random data access, **scalability** and **reliability** are critical.
- **scales seamlessly**
- **integrates with the Apache[®] ecosystem** and supports the **HBase™ API**.

What is Bigtable good for?

Use Case Examples

- **Time-series** data, such as CPU and memory usage over time for multiple servers.
- **Marketing data**, such as purchase histories and customer preferences.
- **Financial data**, such as transaction histories, stock prices, and currency exchange rates.
- **Internet of Things** data, such as usage reports from energy meters and home appliances.
- **Graph data**, such as information about how users are connected to one another.

Applications that need...

- Very high throughput
- Scalability
- Non-Structured key/value data where each value is no larger than 10MB

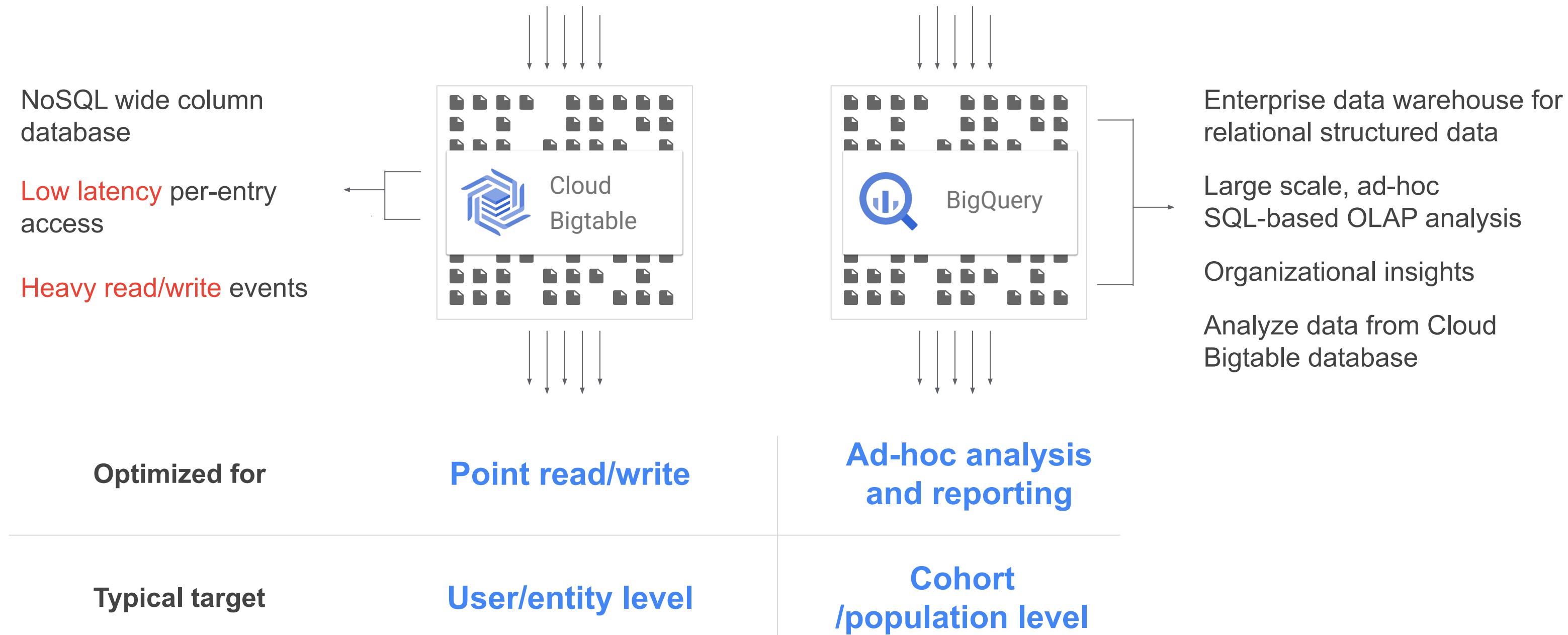
Storage Engine

- Batch MapReduce
- Stream Processing/Analytics
- ML applications

Exam Tip: types of apps where you'd consider using Bigtable:
recommendation engines, personalizing user experience, Internet of Things, real-time analytics, fraud detection, migrating from HBase or Cassandra, Fintech, gaming, high-throughput data streaming for creating / improving ML models.

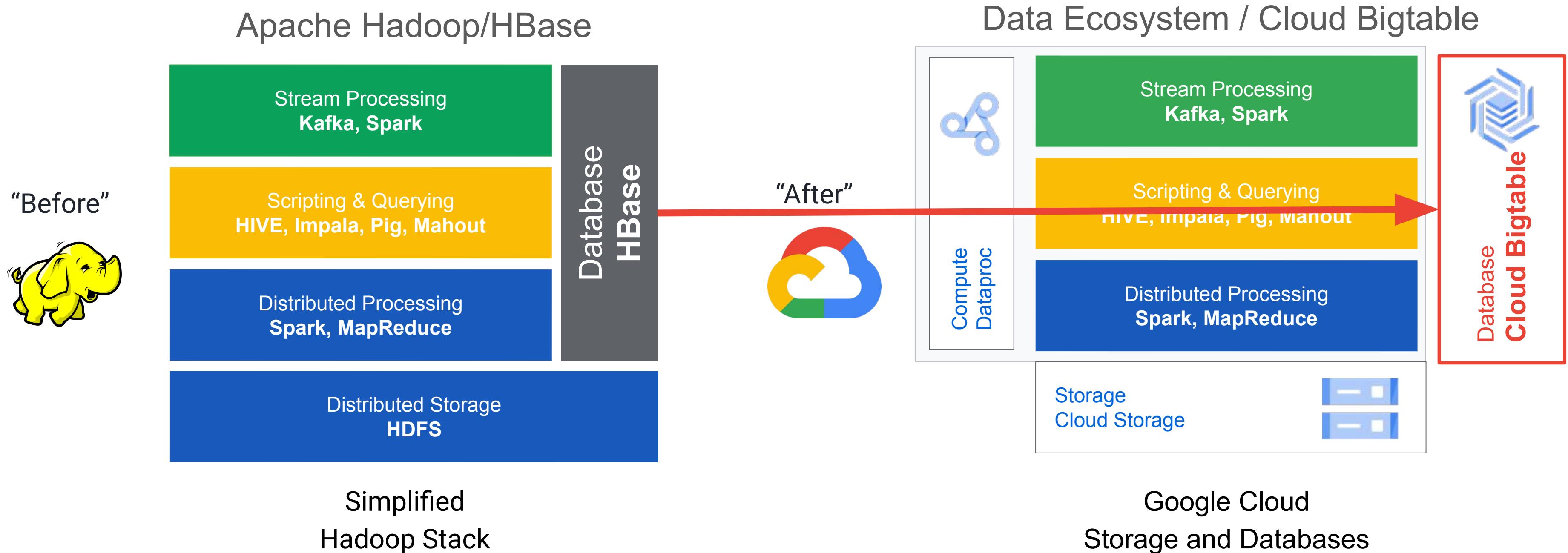
Bigtable for analytics... ?

Bigtable vs BigQuery



Exam Tip: BigTable might be optimal for “real-time analytics”, when you need to make decisions on events as they’re happening.

Bigtable: Hadoop migration and modernization



Exam Tip: Main goal: decoupling of storage & compute. As a consequence, you can treat Dataproc clusters as job-specific / ephemeral

Google Cloud

What is Bigtable not good for?



Not good for...

- Not a relational database
- No SQL Queries or Joins
- No Multi-Row Transactions



Considerations

- You need full SQL support for OLTP
 - consider Spanner or CloudSQL
- Interactive querying for OLAP
 - consider BigQuery
- Need to store immutable blobs larger than 10MB (e.g. movies, images)
 - consider Cloud Storage

Diagnostic Question Discussion

You want to optimize the performance of an accurate, real-time, weather-charting application. The data comes from 50,000 sensors sending 10 readings a second, in the format of a timestamp and sensor reading.

- A. Google BigQuery
- B. Google Cloud SQL
- C. Google Cloud Bigtable
- D. Google Cloud Storage

Where should you store the data?

Diagnostic Question Discussion

You want to optimize the performance of an accurate, real-time, weather-charting application. The data comes from 50,000 sensors sending 10 readings a second, in the format of a timestamp and sensor reading.

Where should you store the data?

- A. Google BigQuery
- B. Google Cloud SQL
- C. **Google Cloud Bigtable**
- D. Google Cloud Storage

Diagnostic Question Discussion

Your company has an application running on Google Cloud that is collecting data from thousands of physical devices that are globally distributed. Data is published to Pub/Sub and streamed in real time into an SSD Cloud Bigtable cluster via a Dataflow pipeline. The operations team informs you that your Cloud Bigtable cluster has a hotspot, and queries are taking longer than expected. You need to resolve the problem and prevent it from happening in the future.

What should you do?

- A. Advise your clients to use HBase APIs instead of NodeJS APIs.
- B. Delete records older than 30 days.
- C. Review your RowKey strategy and ensure that keys are evenly spread across the alphabet.
- D. Double the number of nodes you currently have.

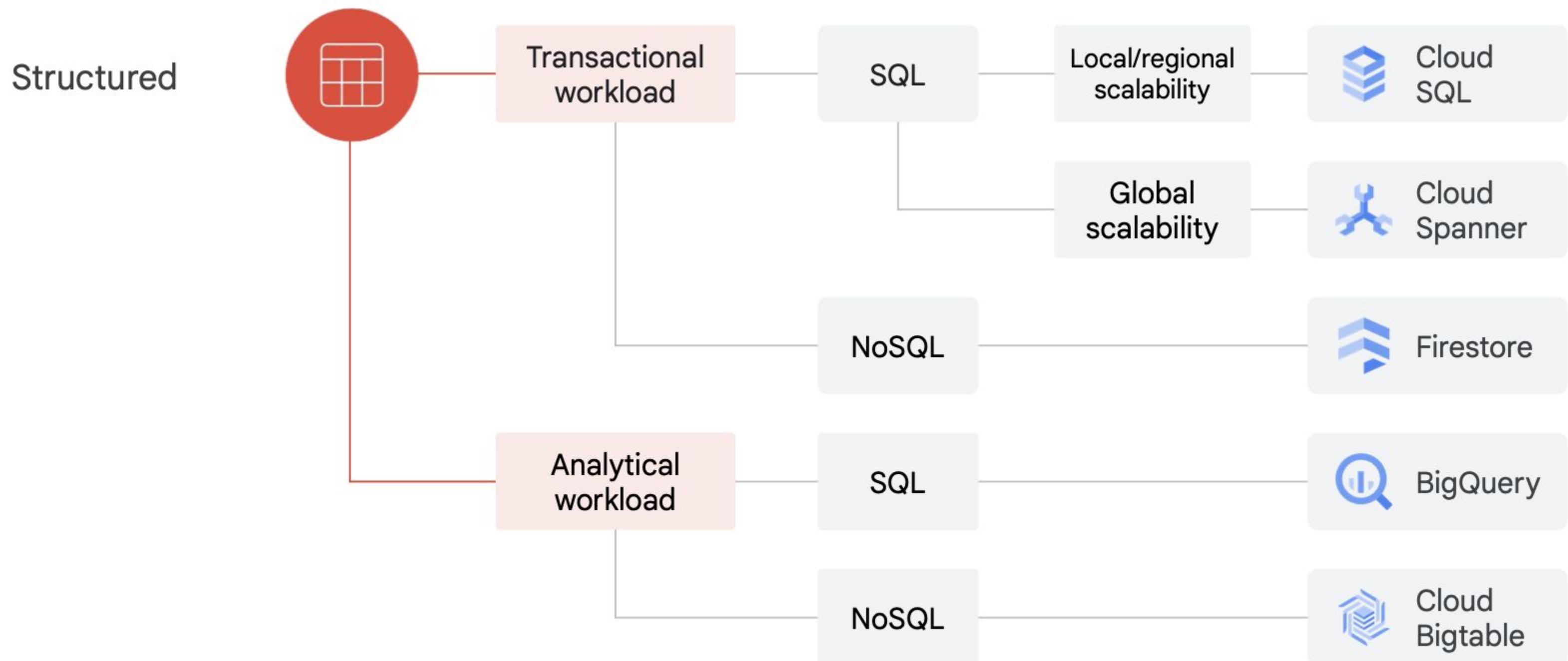
Diagnostic Question Discussion

Your company has an application running on Google Cloud that is collecting data from thousands of physical devices that are globally distributed. Data is published to Pub/Sub and streamed in real time into an SSD Cloud Bigtable cluster via a Dataflow pipeline. The operations team informs you that your Cloud Bigtable cluster has a hotspot, and queries are taking longer than expected. You need to resolve the problem and prevent it from happening in the future.

What should you do?

- A. Advise your clients to use HBase APIs instead of NodeJS APIs.
- B. Delete records older than 30 days.
- C. **Review your RowKey strategy and ensure that keys are evenly spread across the alphabet.**
- D. Double the number of nodes you currently have.

GCP: storage service decision tree



Cymbal Retail

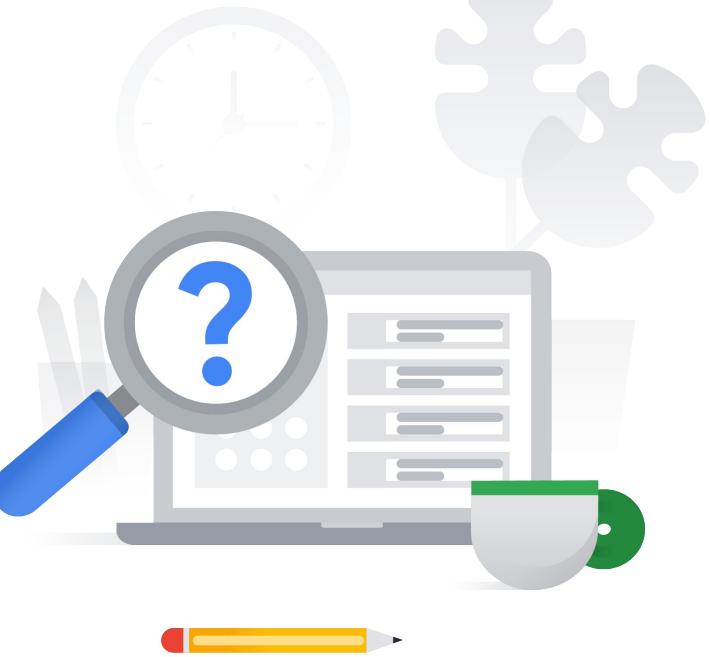
case study



Proposed Technical Solution

- **Catalog and Content Enrichment:**
 - Vertex AI with Gemini Models to extract structured attributes, generate compelling product descriptions and enrich data.
 - [Vertex AI Search for Commerce](#) - service specifically designed for retail and e-commerce (power search, recommendations, catalog improvement) with [HTML-enabled](#).
 - [Cloud Vision API](#) and [Document AI](#) (with [Human-in-the-Loop](#)... deprecated?) for extracting information from visual and document-based supplier data. [Imagen on Vertex AI](#) for image creation
 - Storage part: Cloud Storage for data hosting, and BigQuery for storing and normalizing the final attributes.
- **Conversational Commerce with Product Discovery:**
 - [Conversational Commerce agent \(with IVR enabled\)](#) and [Security Settings](#) for data connected to [Vertex AI Search for Commerce](#) service to use product catalog as its primary source of information. Have this agent deployed to website and mobile app by following [developer's guide](#) and / or using [Dialogflow CX](#).
 - Potentially, implement [CCAI](#) to move away from legacy IVR and manual agent processes, reduce call center costs, and implement modern AI-powered conversational agents for product discovery and sales conversion
- **Technical Stack Modernization:**
 - Migrate on-prem Kubernetes to GKE (Autopilot) or Cloud Run where possible.
 - Migrate relational databases (MySQL, MS SQL) to managed services like Cloud SQL. BigQuery might be suitable for consolidating product catalog and customer data, overcoming data silos and supporting the Gen AI needs. Memorystore can manage Redis instances.
 - Utilize tools like Cloud Data Fusion or Cloud Composer for modern, efficient orchestration and transformation of data, replacing legacy ETL processes. Modernize 3rd party integrations using an API management platform (Apigee).
 - Modernize monitoring with Cloud Logging, Cloud Monitoring, Uptime Checks, Alerting Policies, [Managed Service for Prometheus](#) etc
 - Security and compliance: KMS, encrypted hybrid connectivity, [SDP](#) for sensitive data, [VPC-SC perimeters around sensitive APIs](#)

[Cymbal Retail case study] Diagnostic Question #1

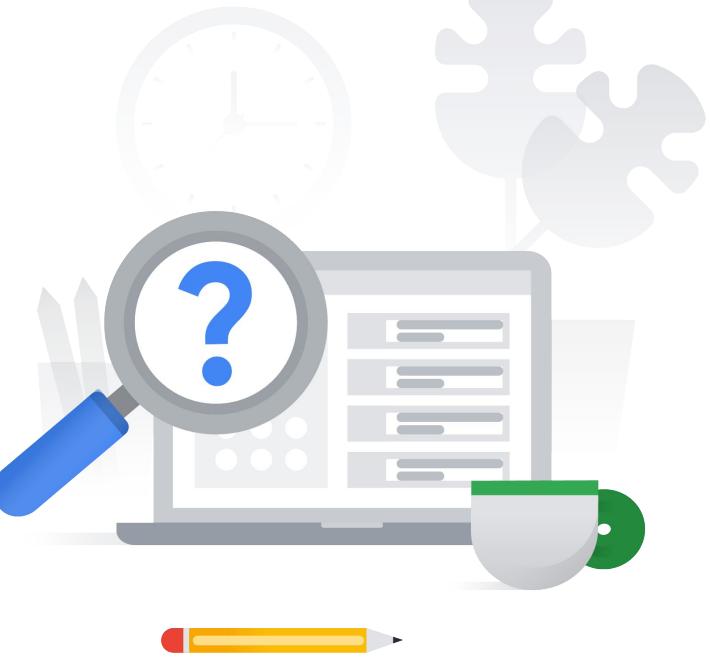


Cymbal Retail's current environment has data silos across various databases (MySQL, Microsoft SQL Server, Redis, and MongoDB), which limits a unified view of the customer. They want to create a centralized data platform to get a holistic view of their customers and enable advanced analytics.

- A. Cloud SQL
- B. Bigtable
- C. Firestore
- D. BigQuery

Which Google Cloud service would be the most appropriate foundation for this centralized data platform?

[Cymbal Retail case study] Diagnostic Question #1

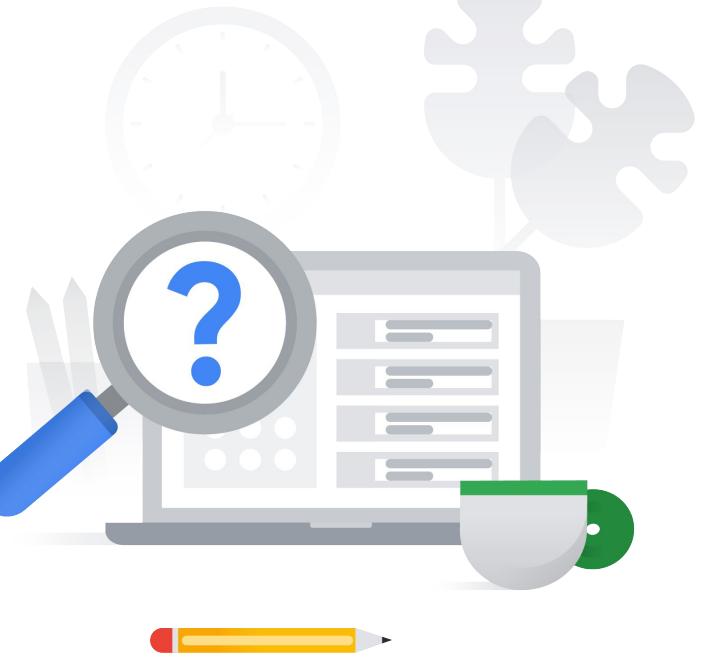


Cymbal Retail's current environment has data silos across various databases (MySQL, Microsoft SQL Server, Redis, and MongoDB), which limits a unified view of the customer. They want to create a centralized data platform to get a holistic view of their customers and enable advanced analytics.

- A. Cloud SQL
- B. Bigtable
- C. Firestore
- D. **BigQuery**

Which Google Cloud service would be the most appropriate foundation for this centralized data platform?

[Cymbal Retail case study] Diagnostic Question #2

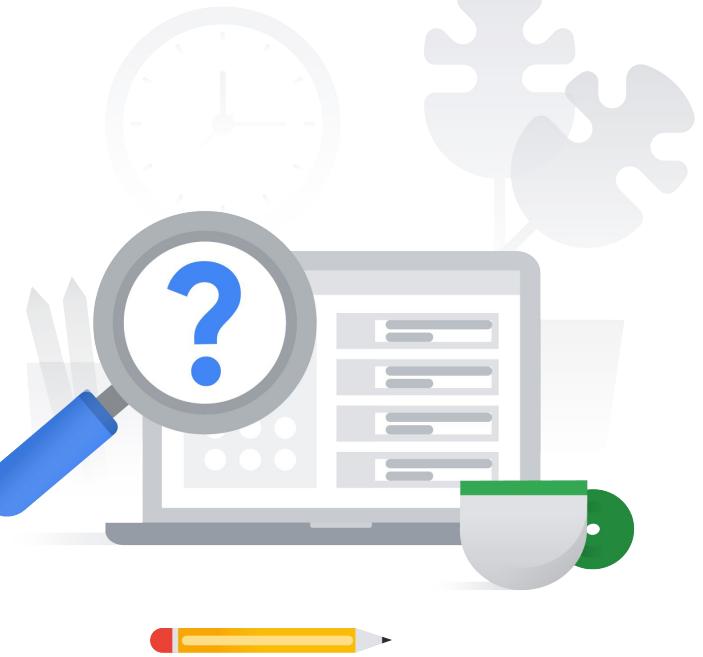


Cymbal Retail's custom-built web application experiences significant traffic fluctuations. They want to ensure that the application can handle traffic spikes without manual intervention while minimizing costs. The application is containerized and runs on Kubernetes.

Which GKE feature should they leverage to achieve this?

- A. Horizontal Pod Autoscaler (HPA)
- B. Vertical Pod Autoscaler (VPA)
- C. Cluster Autoscaler
- D. A combination of HPA and Cluster Autoscaler

[Cymbal Retail case study] Diagnostic Question #2

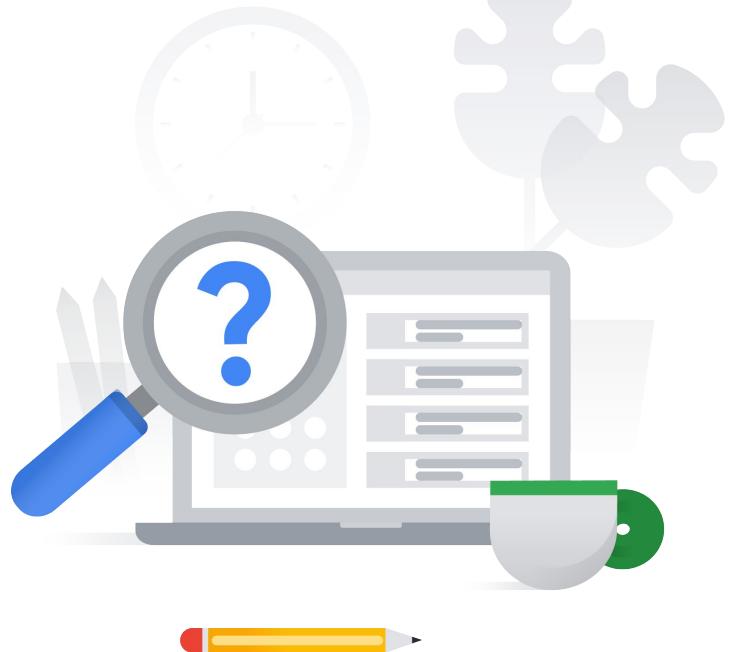


Cymbal Retail's custom-built web application experiences significant traffic fluctuations. They want to ensure that the application can handle traffic spikes without manual intervention while minimizing costs. The application is containerized and runs on Kubernetes.

Which GKE feature should they leverage to achieve this?

- A. Horizontal Pod Autoscaler (HPA)
- B. Vertical Pod Autoscaler (VPA)
- C. Cluster Autoscaler
- D. A combination of HPA and Cluster Autoscaler**

[Cymbal Retail case study] Diagnostic Question #3

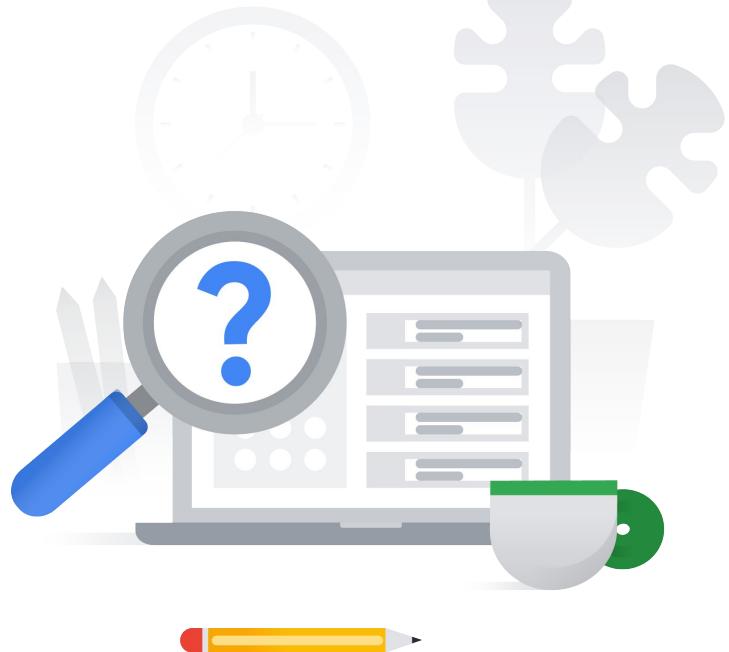


Cymbal Retail wants to implement a solution that can automatically generate high-quality images of their products in different settings and styles for their e-commerce website.

Which Vertex AI service and model type would be most appropriate for this task?

- A. Vertex AI Language with a text generation model.
- B. Vertex AI with a generative adversarial network (GAN) model.
- C. Vertex AI Vision with an image generation model (e.g., Imagen).
- D. Vertex AI Vision with a classification model.

[Cymbal Retail case study] Diagnostic Question #3

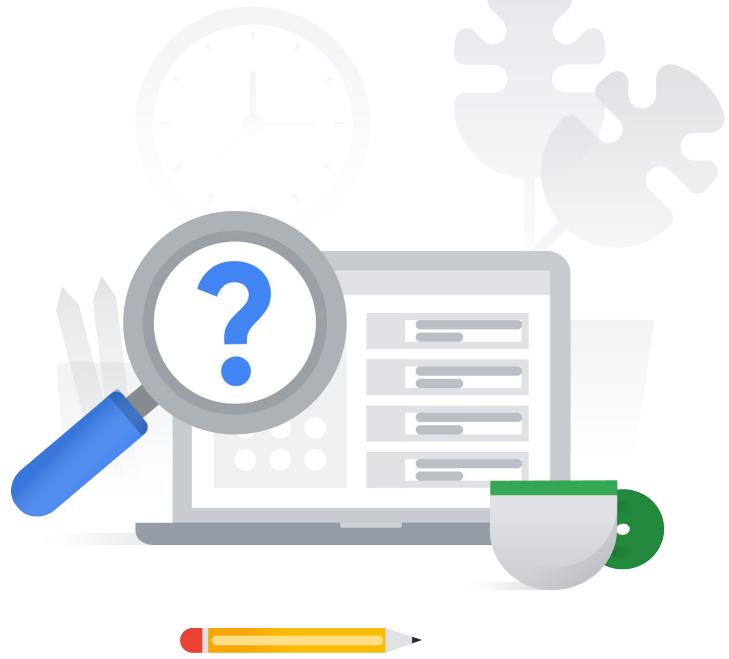


Cymbal Retail wants to implement a solution that can automatically generate high-quality images of their products in different settings and styles for their e-commerce website.

Which Vertex AI service and model type would be most appropriate for this task?

- A. Vertex AI Language with a text generation model.
- B. Vertex AI with a generative adversarial network (GAN) model.
- C. Vertex AI Vision with an image generation model (e.g., Imagen).**
- D. Vertex AI Vision with a classification model.

[Cymbal Retail case study] Diagnostic Question #4

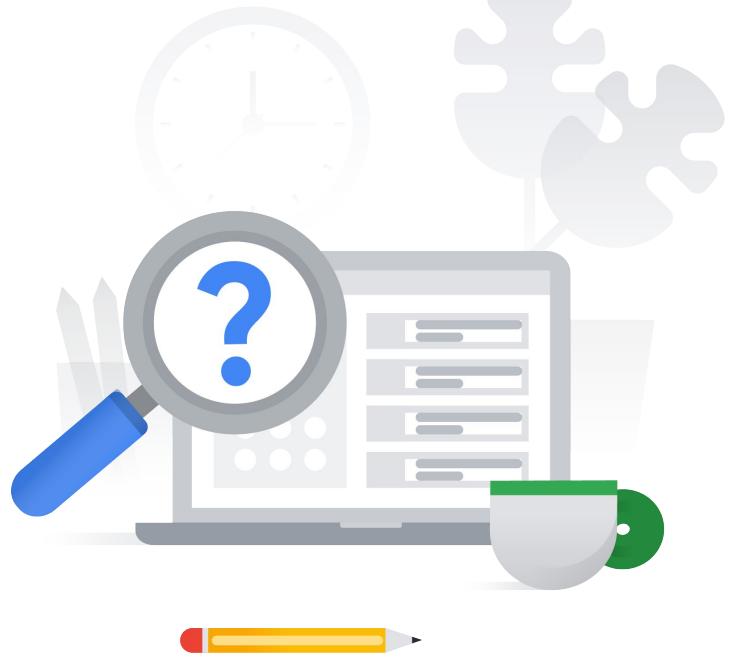


Cymbal Retail wants to use Generative AI to provide their customer service agents with real-time assistance. The system should listen to customer calls, transcribe them in real-time, and suggest relevant answers and solutions from a knowledge base.

Which combination of Google Cloud services would be best for this solution?

- A. Speech-to-Text API, Vertex AI Language, and Cloud Storage.
- B. Contact Center AI (CCAI) Platform, which integrates Speech-to-Text, Dialogflow, and Agent Assist.
- C. Dialogflow CX, Speech-to-Text API, and BigQuery
- D. Cloud Pub/Sub, Cloud Functions, and the Speech-to-Text API.

[Cymbal Retail case study] Diagnostic Question #4

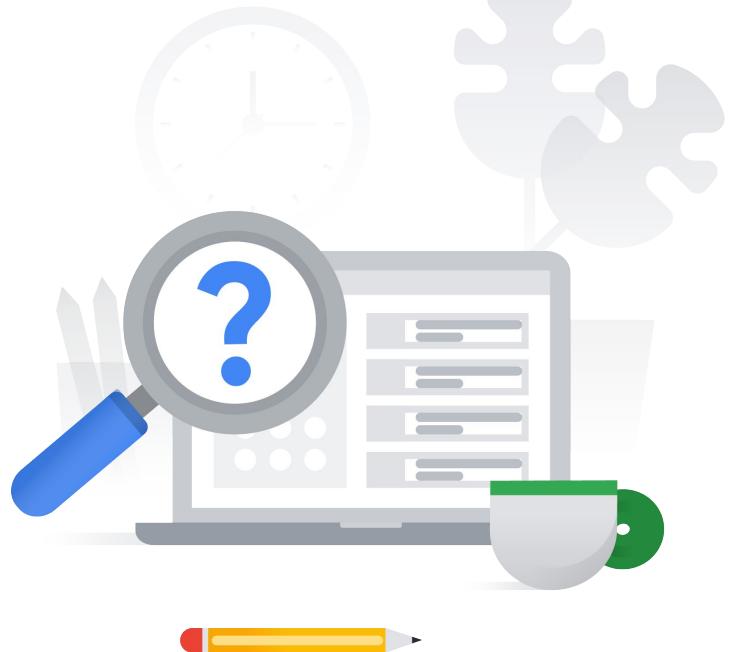


Cymbal Retail wants to use Generative AI to provide their customer service agents with real-time assistance. The system should listen to customer calls, transcribe them in real-time, and suggest relevant answers and solutions from a knowledge base.

Which combination of Google Cloud services would be best for this solution?

- A. Speech-to-Text API, Vertex AI Language, and Cloud Storage.
- B. **Contact Center AI (CCAI) Platform, which integrates Speech-to-Text, Dialogflow, and Agent Assist.**
- C. Dialogflow CX, Speech-to-Text API, and BigQuery
- D. Cloud Pub/Sub, Cloud Functions, and the Speech-to-Text API.

[Cymbal Retail case study] Diagnostic Question #5

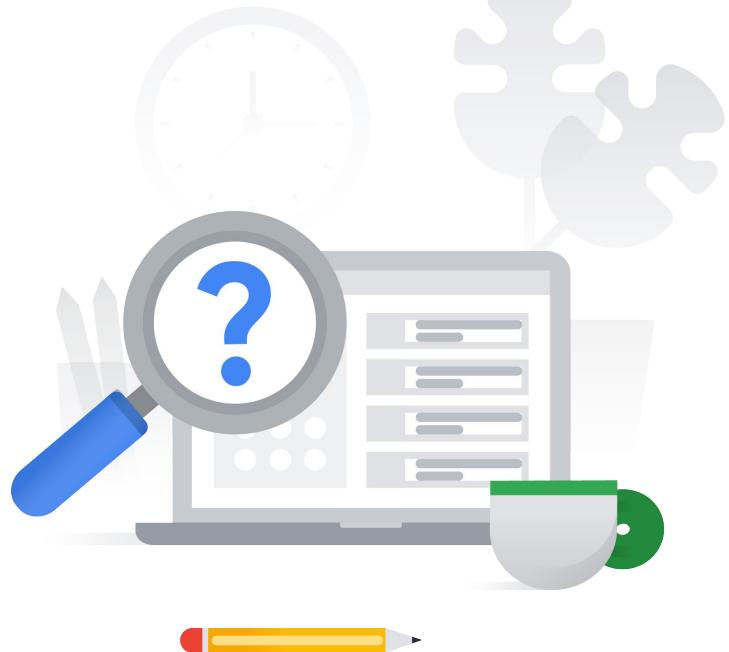


Cymbal Retail is building a product recommendation engine using Generative AI. The model should be able to recommend products based on a user's natural language query (e.g., "show me some stylish and comfortable shoes for a summer wedding").

Which approach would be most effective?

- A. Use Vertex AI Search's semantic search capabilities to match user queries with relevant products.
- B. Fine-tune a large language model (LLM) on Cymbal Retail's product catalog and customer reviews.
- C. Use a traditional collaborative filtering model trained on user purchase data.
- D. Build a custom model using TensorFlow and host it on Vertex AI Training.

[Cymbal Retail case study] Diagnostic Question #5



Cymbal Retail is building a product recommendation engine using Generative AI. The model should be able to recommend products based on a user's natural language query (e.g., "show me some stylish and comfortable shoes for a summer wedding").

Which approach would be most effective?

- A. Use Vertex AI Search's semantic search capabilities to match user queries with relevant products.
- B. Fine-tune a large language model (LLM) on Cymbal Retail's product catalog and customer reviews.
- C. Use a traditional collaborative filtering model trained on user purchase data.
- D. Build a custom model using TensorFlow and host it on Vertex AI Training.

Optional materials 1

[READING]

- Make sure you know the [differences between BigQuery and BigTable](#).
- Be aware [how BigQuery table partitioning works](#).

[VIDEOS]

- Cloud Networking 104 (Load Balancers): [Cloud OnAir: Networking 104 - Everything You Need to Know About Load Balancers on GCP](#)
- [Querying external data with BigQuery](#)
- BigQuery: [What is BigQuery?](#)
- [IMPORTANT] Sharing BigQuery data with others: [Protect data with authorized views](#)
- BigTable: [What is Cloud Bigtable?](#)
- BigTable: [What can you do with Bigtable?](#)
- Cloud Spanner [5 min]: [What is Cloud Spanner | Cloud Spanner Explained | Cloud Native Relational Database](#)
- Cloud Spanner [2x5min]: [How to set up a Cloud Spanner instance](#) & [Cloud Spanner: Database deep dive](#)
- Introduction to Firestore: [Introduction to Firestore | NoSQL Document Database](#)
- What is Cloud Build?: [What is Cloud Build? #GCPSketchnote](#)
- [Three ways to improve CI/CD in your serverless app](#)

Make sure to...

Enjoy the journey as much
as the destination!

