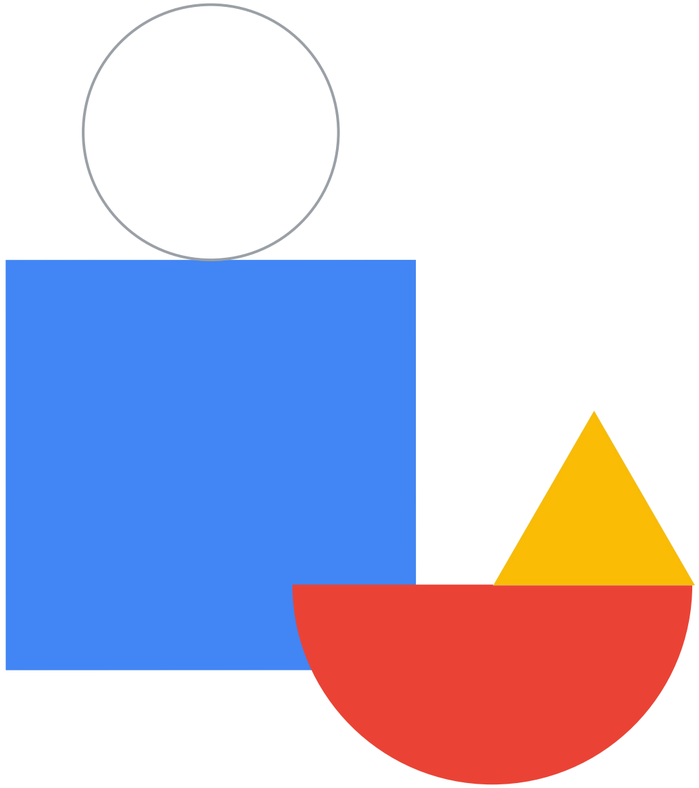
Google Cloud



Preparing for your Professional Cloud Architect Journey

Module 5: Managing Implementation
Module 6: Ensuring solution and operations reliability

Week 6 topics

Managing Implementation

Infrastructure automation journey

Customer starting with IaC and CI/CD

Customer with a high degree of maturity

IaC is not an option, but a necessity...

Increasing demand

Requires rapid scaling of IT infrastructure (in both directions!)

Operational bottlenecks

Large Ops teams need to overcome organizational and technical bottlenecks, such as managing infrastructure **consistently** in scale

Disconnected feedback

Communication gap between software and IT teams

Manual errors

Increased scale leads to greater human errors

IaC: Declarative Infrastructure

Declarative (statement)

"I should have five servers"

VS.

Imperative (command)

"Give me five servers"

GCP services emulators

for local development, testing and validation

- Spanner:
  - https://cloud.google.com/spanner/docs/emulator
  - locally-running, emulated instance of Cloud Spanner to enable local development and testing.
  - https://github.com/GoogleCloudPlatform/cloud-spanner-emulator
- Pub/Sub:
  - https://cloud.google.com/pubsub/docs/emulator
- Bigtable:
  - https://cloud.google.com/bigtable/docs/emulator
- Firestore:
  - https://cloud.google.com/firestore/docs/emulator
- Cloud Run:
  - https://cloud.google.com/run/docs/testing/local

Cloud Shell (Part of Cloud Console)

- A temporary Debian based, Compute Engine virtual machine instance in a web browser
- Built-in code editor
- 5 GB of persistent disk storage
- Pre-installed Google Cloud SDK and other tools
- Web preview functionality
- Built-in authorization for access to Google Cloud Console projects and resources

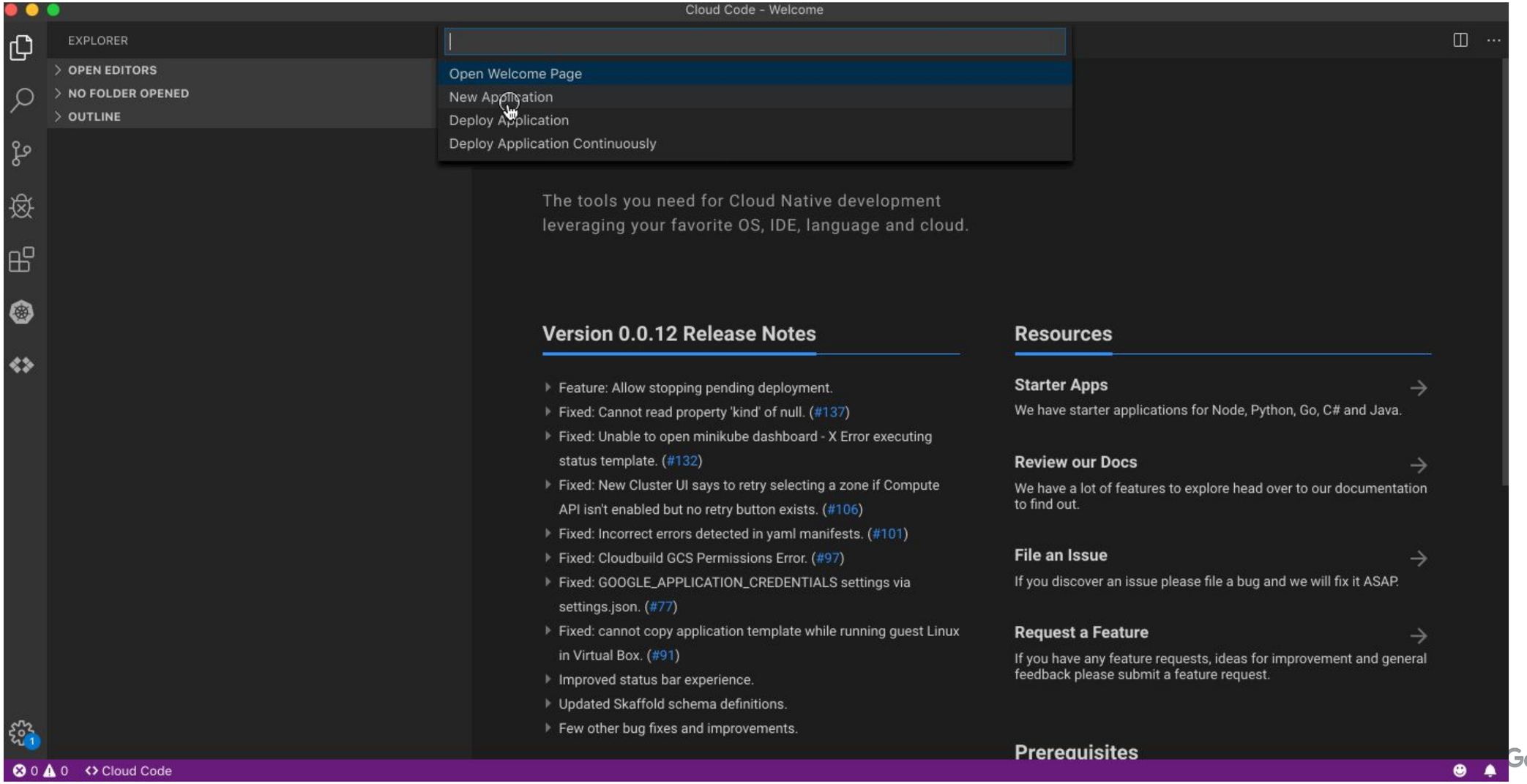
Cloud Shell Editor

Cloud Shell also provides a VS Code-like web IDE.

The IDE runs on the Cloud Shell VM, and serves a web interface on port 970. The Cloud Shell UI iframes this web interface and provides additional UI and features around it.

We maintain a custom Cloud Shell extension that runs on the IDE and manages communication with the rest of the Cloud Shell UI. This extension allows for interactions such as managing the app lifecycle (loading, reconnect) in the Cloud Shell UI, or allowing Neos to open specific files in the Editor or spotlight UI elements.

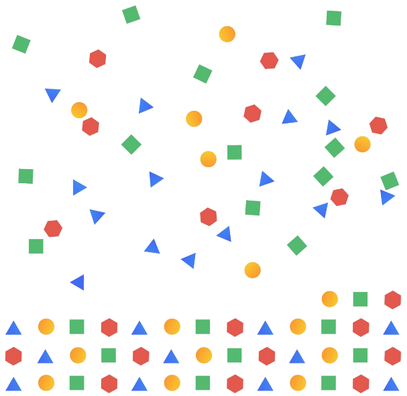
Cloud Code


Cloud Code - Welcome
EXPLORER
OPEN EDITORS
NO FOLDER OPENED
OUTLINE
Open Welcome Page
New Application
Deploy Application
Deploy Application Continuously
The tools you need for Cloud Native development
leveraging your favorite OS, IDE, language and cloud.
Version 0.0.12 Release Notes
Feature: Allow stopping pending deployment.
Fixed: Cannot read property 'kind' of null. (#137)
Fixed: Unable to open minikube dashboard - X Error executing status template. (#132)
Fixed: New Cluster UI says to retry selecting a zone if Compute API isn't enabled but no retry button exists. (#106)
Fixed: Incorrect errors detected in yaml manifests. (#101)
Fixed: Cloudbuild GCS Permissions Error. (#97)
Fixed: GOOGLE_APPLICATION_CREDENTIALS settings via settings.json. (#77)
Fixed: cannot copy application template while running guest Linux in Virtual Box. (#91)
Improved status bar experience.
Updated Skaffold schema definitions.
Few other bug fixes and improvements.
Resources
Starter Apps
We have starter applications for Node, Python, Go, C# and Java.
Review our Docs
We have a lot of features to explore head over to our documentation to find out.
File an Issue
If you discover an issue please file a bug and we will fix it ASAP.
Request a Feature
If you have any feature requests, ideas for improvement and general feedback please submit a feature request.
Prerequisites
0 0 Cloud Code

Ensuring solution and operations excellence

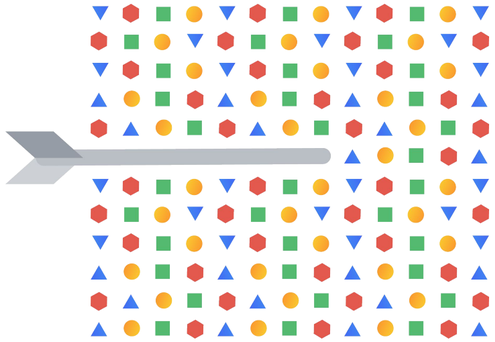
Solution reliability optimization

Make sure to go through
Well-Architected Framework: Reliability pillar



Chaos Engineering

- Creates a culture of reliability
- Crashes systems intentionally to build resiliency
- **Service Mesh can help you here!**



Penetration testing

- Mimics the behavior of hackers to attack your own environment

Diagnostic Question Discussion

You are developing your microservices application on Google Kubernetes Engine. During testing, you want to validate the behavior of your application in case a specific microservice should suddenly crash.

What should you do?

A. Add a taint to one of the nodes of the Kubernetes cluster. For the specific microservice, configure a pod anti-affinity label that has the name of the tainted node as a value.

B. Use Istio's fault injection on the particular microservice whose faulty behavior you want to simulate.

C. Destroy one of the nodes of the Kubernetes cluster to observe the behavior.

D. Configure Istio's traffic management features to steer the traffic away from a crashing microservice.

Diagnostic Question Discussion

You are developing your microservices application on Google Kubernetes Engine. During testing, you want to validate the behavior of your application in case a specific microservice should suddenly crash.

What should you do?

A. Add a taint to one of the nodes of the Kubernetes cluster. For the specific microservice, configure a pod anti-affinity label that has the name of the tainted node as a value.

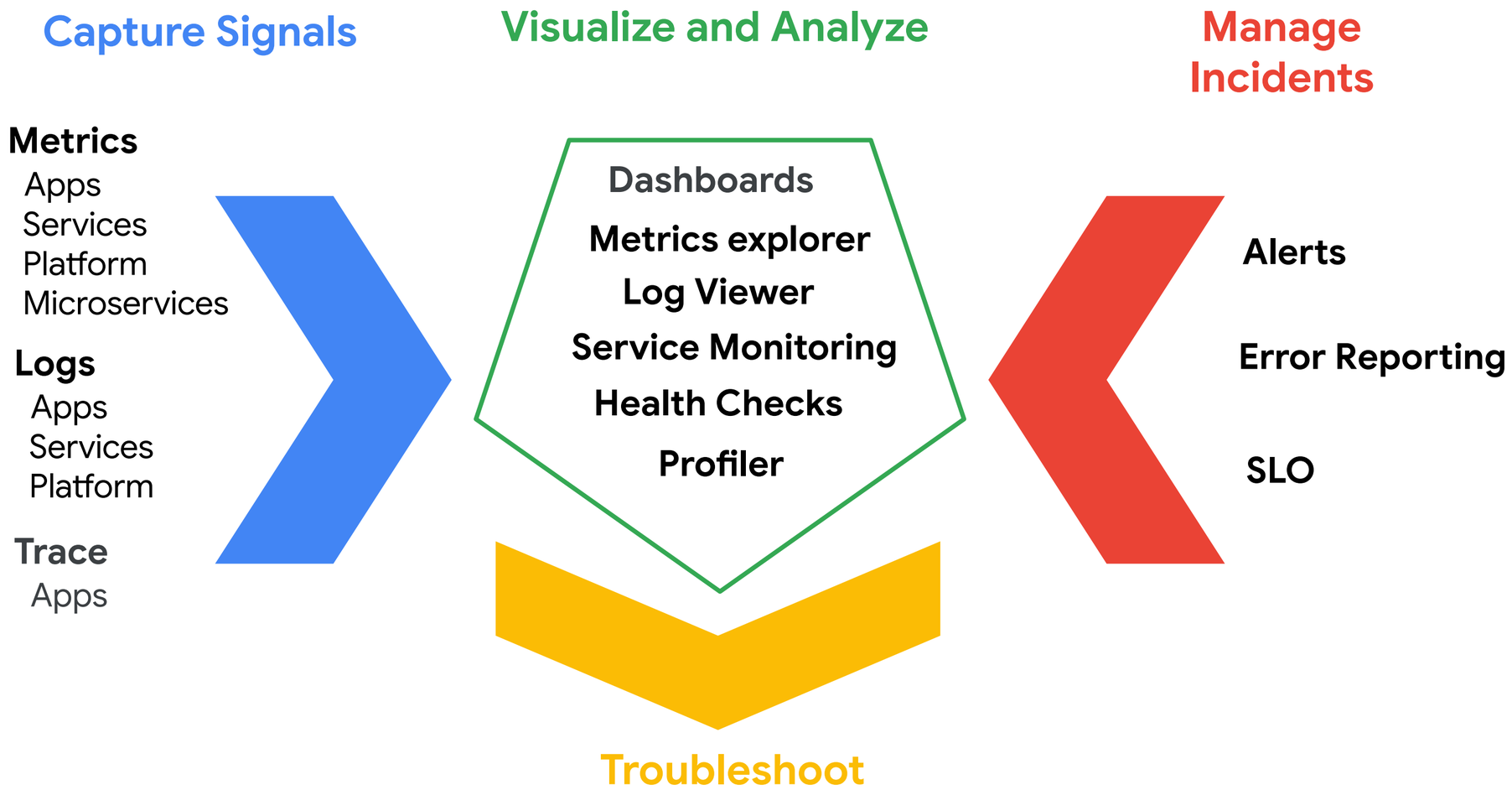
B. Use Istio's fault injection on the particular microservice whose faulty behavior you want to simulate.

C. Destroy one of the nodes of the Kubernetes cluster to observe the behavior.

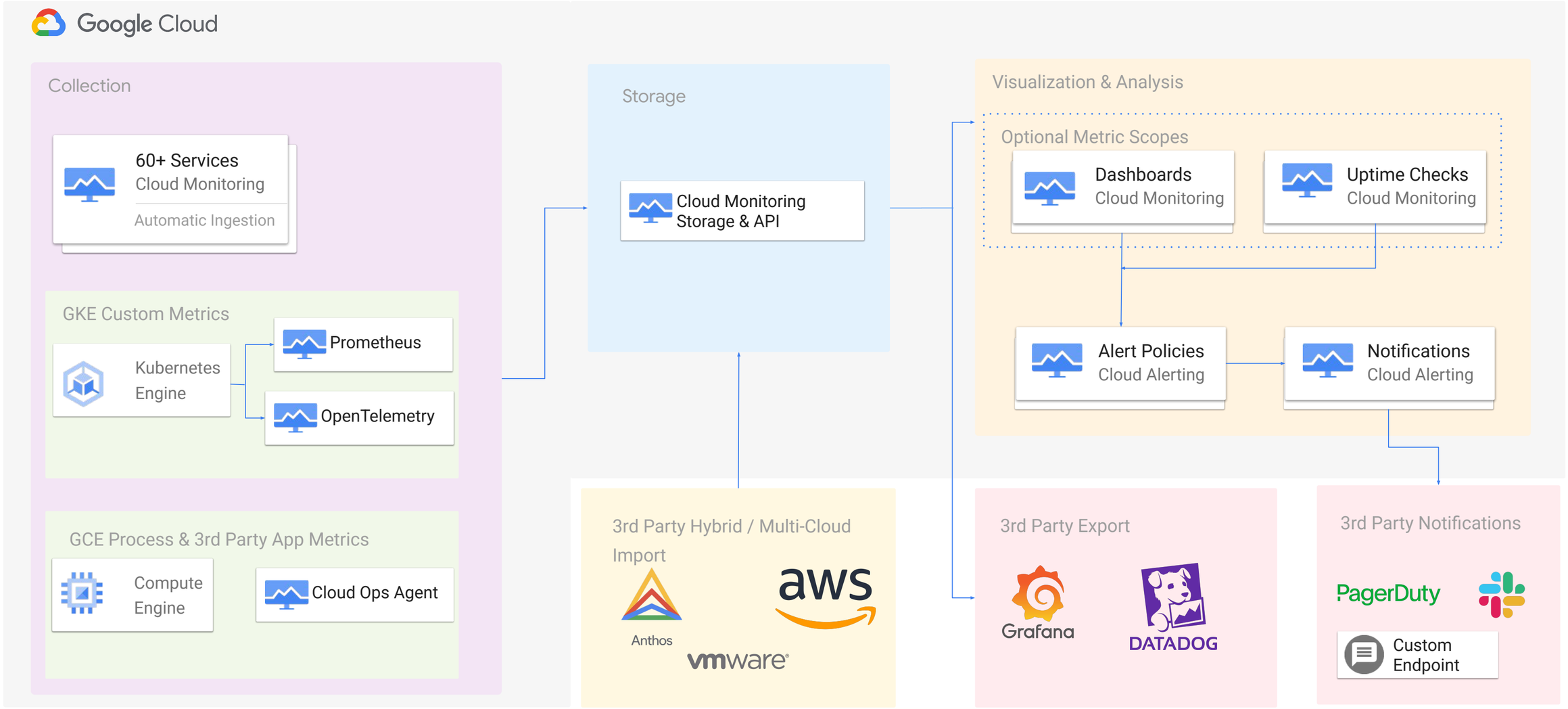
D. Configure Istio's traffic management features to steer the traffic away from a crashing microservice.

https://istiobyexample.dev/fault-injection/

Observability in GCP: The big picture



Cloud monitoring architecture and integration



Business Continuity = Planning for failure

A well-designed system can answer the question: "What happens when a zone or region has a 1, 5, 10, or 30 minute outage?" This should be considered at many layers, including:

- What will my customers experience during an outage?
- How will I detect that an outage is happening?
- What happens to my application during an outage?
- What happens to my data during an outage?
- What happens to my other applications due to an outage (due to cross-dependencies)?
- What do I need to do in order to recover after an outage is resolved? Who does it?
- Who do I need to notify about an outage, within what time period?

Resource	Examples	Availability design goal	Implied downtime
Zonal	Compute Engine, Persistent Disk	99.9%	8.75 hours / year
Regional	Regional Cloud Storage, Replicated Persistent Disk, Regional Google Kubernetes Engine	99.99%	52 minutes / year

High Availability for...

- Compute Engine ⇒ ?
- GKE ⇒ ?
- Cloud Run ⇒ ?
- Cloud SQL ⇒ ?
- Cloud Spanner ⇒ ?
- Cloud Storage => ?

High Availability for...

- Compute Engine ⇒ regional MIGs, Load Balancers
- GKE ⇒ ?
- Cloud Run ⇒ ?
- Cloud SQL ⇒ ?
- Cloud Spanner ⇒ ?
- Cloud Storage => ?

High Availability for...

- Compute Engine ⇒ regional MIGs, Load Balancers
- GKE ⇒ Regional clusters, Load Balancers
- Cloud Run ⇒ ?
- Cloud SQL ⇒ ?
- Cloud Spanner ⇒ ?
- Cloud Storage => ?

High Availability for...

- Compute Engine ⇒ regional MIGs, Load Balancers
- GKE ⇒ Regional clusters, Load Balancers
- Cloud Run ⇒ automatic scaling, regional resource
- Cloud SQL ⇒ ?
- Cloud Spanner ⇒ ?
- Cloud Storage => ?

High Availability for...

- Compute Engine ⇒ regional MIGs, Load Balancers
- GKE ⇒ Regional clusters, Load Balancers
- Cloud Run ⇒ automatic scaling, regional resource
- Cloud SQL ⇒ HA “checkbox” for multi-zone
- Cloud Spanner ⇒ ?
- Cloud Storage => ?

High Availability for...

- Compute Engine ⇒ regional MIGs, Load Balancers
- GKE ⇒ Regional clusters, Load Balancers
- Cloud Run ⇒ automatic scaling, regional resource
- Cloud SQL ⇒ HA “checkbox” for multi-zone
- Cloud Spanner ⇒ Multi-instance deployment with automatic failover
- Cloud Storage => ?

High Availability for...

- Compute Engine ⇒ regional MIGs, Load Balancers
- GKE ⇒ Regional clusters, Load Balancers
- Cloud Run ⇒ automatic scaling, regional resource
- Cloud SQL ⇒ HA “checkbox” for multi-zone
- Cloud Spanner ⇒ Multi-instance deployment with automatic failover
- Cloud Storage => regional / dual/ multi-region bucket, optional replication

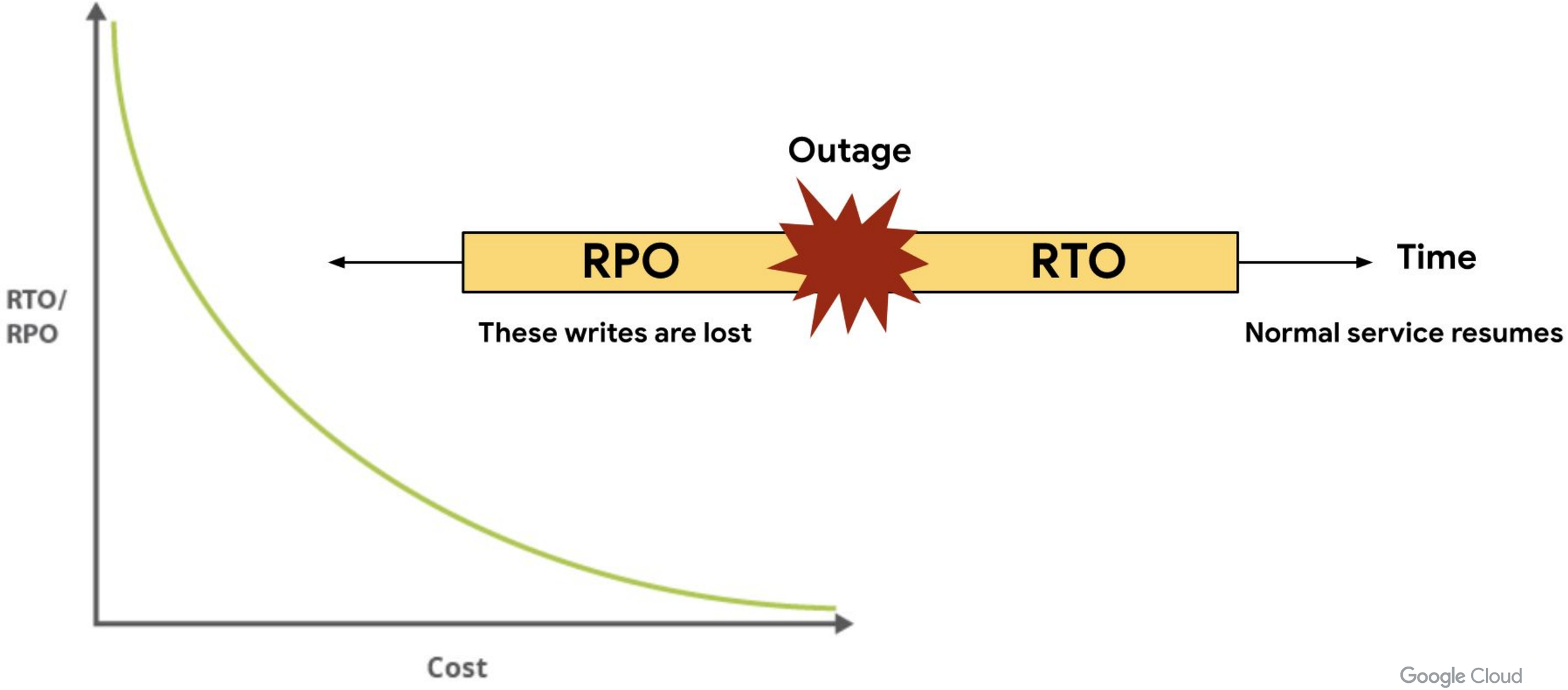
TIP

Product Reference for Business Continuity

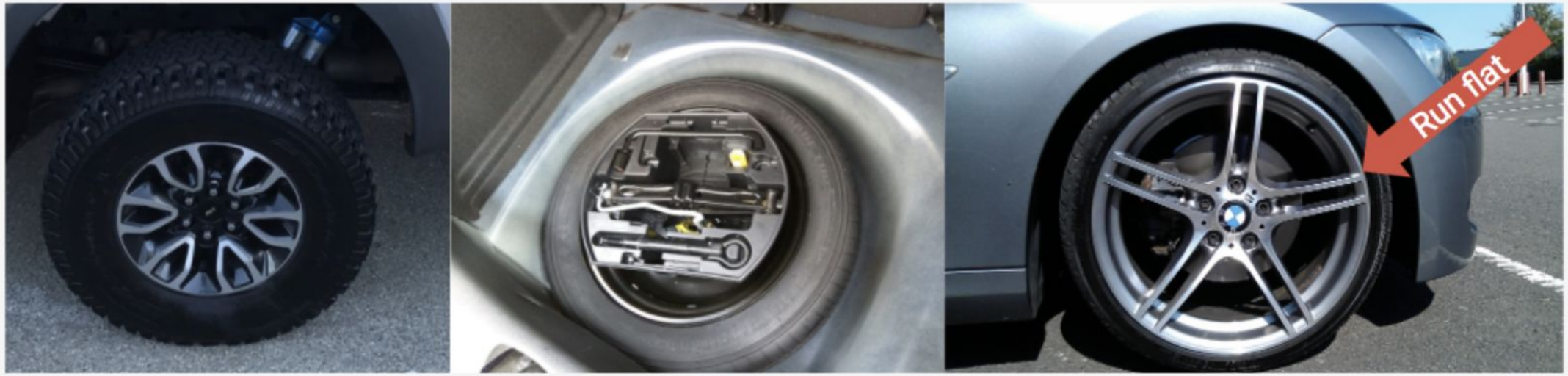
TIP

The more managed a service is, the easier it is to ensure at least 99,99% SLA

Disaster Recovery architecture is driven by RPO and RTO



Disaster Recovery architecture is driven by RPO and RTO



Application criticality	% of Apps	Example apps	Zone outage	Region outage
Tier 1 (most important)	5%	Typically global or external customer-facing applications such as real-time payments and eCommerce storefronts.	RTO Zero RPO Zero	RTO Zero RPO Zero
Tier 2	35%	Typically regional applications or important internal applications such as CRM or ERP.	RTO 15mins RPO 15mins	RTO 1hr RPO 1hr
Tier 3 (least important)	60%	Typically team or departmental applications, such as back office, leave booking, internal travel, accounting, and HR.	RTO 1hr RPO 1hr	RTO 12hrs RPO 12hrs

EXAMPLE: Disaster Recovery for Cloud SQL instance...

- Cold:
- Warm:
- Hot:

EXAMPLE: Disaster Recovery for Cloud SQL instance...

- Cold: backups offloaded to another region
- Warm:
- Hot:

EXAMPLE: Disaster Recovery for Cloud SQL instance...

- Cold: backups offloaded to another region
- Warm: read-only replicas in a different region with asynchronous replication
- Hot:

EXAMPLE: Disaster Recovery for Cloud SQL instance...

- Cold: backups offloaded to another region
- Warm: read-only replicas in a different region with asynchronous replication
- Hot: … none available out of the box. Alternatives:
  - Migrate to Cloud Spanner ?
  - MySQL on 2 GCE VMs (NOT Cloud SQL) with DRBD, load balancer in front and automatic failover. Details here.
  - Other Do-It-Yourself options

Diagnostic Question Discussion

You are designing a multi-region disaster recovery solution for a critical web application deployed on Google Cloud. The application requires high availability and minimal downtime in the event of a regional outage. You need to ensure that the application remains accessible to users in a different region if one region fails.

What should you do?

A. Deploy the application to a single region using Compute Engine virtual machine (VM) instances and configure cross-region backups to a different region.

B. Deploy the application to a single region using Cloud Functions and configure a Cloud Storage bucket in a different region for static content.

C. Deploy the application to multiple regions using Google Kubernetes Engine (GKE) with regional clusters and configure a global load balancer to distribute traffic across the regions.

D. Deploy the application to multiple zones within a single region using Compute Engine VM instances and configure a regional load balancer to distribute traffic across the zones.

Diagnostic Question Discussion

You are designing a multi-region disaster recovery solution for a critical web application deployed on Google Cloud. The application requires high availability and minimal downtime in the event of a regional outage. You need to ensure that the application remains accessible to users in a different region if one region fails.

What should you do?

A. Deploy the application to a single region using Compute Engine virtual machine (VM) instances and configure cross-region backups to a different region.

B. Deploy the application to a single region using Cloud Functions and configure a Cloud Storage bucket in a different region for static content.

C. Deploy the application to multiple regions using Google Kubernetes Engine (GKE) with regional clusters and configure a global load balancer to distribute traffic across the regions.

D. Deploy the application to multiple zones within a single region using Compute Engine VM instances and configure a regional load balancer to distribute traffic across the zones.

“Where should I run my stuff?” game

Where should I run my stuff?

- Containers =>

Where should I run my stuff?

- Containers => GKE, Cloud Run, (App Engine)

Where should I run my stuff?

- Containers => GKE, Cloud Run, (App Engine)
- Specific OS, licensing, backup =>

Where should I run my stuff?

- Containers => GKE, Cloud Run, (App Engine)
- Specific OS, licensing, backup => GCE

Where should I run my stuff?

- Containers => GKE, Cloud Run, (App Engine)
- Specific OS, licensing, backup => GCE
- Hybrid, multi-cloud, portability =>

Where should I run my stuff?

- Containers => GKE, Cloud Run, (App Engine)
- Specific OS, licensing, backup => GCE
- Hybrid, multi-cloud, portability => GKE (Anthos), Cloud Run

Where should I run my stuff?

- Containers => GKE, Cloud Run, (App Engine)
- Specific OS, licensing, backup => GCE
- Hybrid, multi-cloud, portability => GKE (Anthos), Cloud Run
- Web applications =>

Where should I run my stuff?

- Containers => GKE, Cloud Run, (App Engine)
- Specific OS, licensing, backup => GCE
- Hybrid, multi-cloud, portability => GKE (Anthos), Cloud Run
- Web applications => Cloud Run (App Engine)

Where should I run my stuff?

- Containers => GKE, Cloud Run, (App Engine)
- Specific OS, licensing, backup => GCE
- Hybrid, multi-cloud, portability => GKE (Anthos), Cloud Run
- Web applications => Cloud Run (App Engine)
- Event-based processing =>

Where should I run my stuff?

- Containers => GKE, Cloud Run, (App Engine)
- Specific OS, licensing, backup => GCE
- Hybrid, multi-cloud, portability => GKE (Anthos), Cloud Run
- Web applications => Cloud Run (App Engine)
- Event-based processing => Cloud Functions

Where should I run my stuff?

- Containers => GKE, Cloud Run, (App Engine)
- Specific OS, licensing, backup => GCE
- Hybrid, multi-cloud, portability => GKE (Anthos), Cloud Run
- Web applications => Cloud Run (App Engine)
- Event-based processing => Cloud Functions
- Want to squeeze every drop from provisioned resources =>

Where should I run my stuff?

- Containers => GKE, Cloud Run, (App Engine)
- Specific OS, licensing, backup => GCE
- Hybrid, multi-cloud, portability => GKE (Anthos), Cloud Run
- Web applications => Cloud Run (App Engine)
- Event-based processing => Cloud Functions
- Want to squeeze every drop from provisioned resources => GCE, GKE

Where should I run my stuff?

- Containers => GKE, Cloud Run, (App Engine)
- Specific OS, licensing, backup => GCE
- Hybrid, multi-cloud, portability => GKE (Anthos), Cloud Run
- Web applications => Cloud Run (App Engine)
- Event-based processing => Cloud Functions
- Want to squeeze every drop from provisioned resources => GCE, GKE
- Prefer managed / serverless / focus on code =>

Where should I run my stuff?

- Containers => GKE, Cloud Run, (App Engine)
- Specific OS, licensing, backup => GCE
- Hybrid, multi-cloud, portability => GKE (Anthos), Cloud Run
- Web applications => Cloud Run (App Engine)
- Event-based processing => Cloud Functions
- Want to squeeze every drop from provisioned resources => GCE, GKE
- Prefer managed / serverless / focus on code => Cloud Functions, Cloud Run, App Engine

KnightMotives Automotive

case study

Proposed Technical Solutions

- Hybrid cloud strategy -> GKE + Anthos / Cloud Service Mesh
- ERP: migrate to GCP (using GCE), or modernize by choosing 3rd party, cloud-based solution
- Network:
  - NCC for connectivity between plants and headquarters.
  - IoT platform for vehicle connectivity + Google AI Edge & small models (Gemma? Nano?) deployed directly to vehicles
- In-vehicle experience:
  - release Android Automotive OS for a consistent experience across the board.
  - Build IoT pipeline (Pub/Sub -> Dataflow -> BigQuery) with a custom AI model (complete lifecycle via Vertex AI) for predictive maintenance, personalized driver settings, and advanced driver-assistance systems
- Delightful experience for dealers and customers:
  - Rebuild the online ordering system as a cloud-native application on Google Kubernetes Engine (GKE) / Cloud Run for scalability and reliability. Use Firestore or Cloud SQL as the backend database.
  - Develop web-based apps for inventory management, sales analytics dashboards (using Looker), and a streamlined service process.
  - Improve customer experience with Vertex AI Conversation chatbots for customer support
- Focus on security and risk management: variety of options depending on the system (compute / storage / AI / …) like SCC, Model Armor, VPC-SC, SDP and more.
- Data monetization and insights:
  - Break down data silos by consolidating all corporate data into BigQuery.
  - Expose APIs via Apigee for monetization

[KnightMotives Automotive] Diagnostic Question #1

KnightMotives wants to create a consistent in-vehicle user experience across all its models, including BEV, hybrid, and ICE vehicles. The platform needs to support AI-powered features and be easily updatable.

Which technology should they adopt as the foundation for their new in-vehicle system?

A. A custom Linux-based OS

B. Android Automotive OS

C. A proprietary real-time operating system (RTOS)

D. A web-based application running on an in-vehicle browser

[KnightMotives Automotive] Diagnostic Question #1

KnightMotives wants to create a consistent in-vehicle user experience across all its models, including BEV, hybrid, and ICE vehicles. The platform needs to support AI-powered features and be easily updatable.

Which technology should they adopt as the foundation for their new in-vehicle system?

A. A custom Linux-based OS

B. Android Automotive OS

C. A proprietary real-time operating system (RTOS)

D. A web-based application running on an in-vehicle browser

[KnightMotives Automotive] Diagnostic Question #2

KnightMotives has experienced past data breaches and must adhere to strict EU data protection regulations. They need a centralized tool to manage security policies and detect potential vulnerabilities across their GCP environment.

Which GCP service should they use?

A. Cloud Armor

B. Cloud Identity and Access Management (IAM)

C. Security Command Center

D. Cloud Key Management Service (KMS)

[KnightMotives Automotive] Diagnostic Question #2

KnightMotives has experienced past data breaches and must adhere to strict EU data protection regulations. They need a centralized tool to manage security policies and detect potential vulnerabilities across their GCP environment.

Which GCP service should they use?

A. Cloud Armor

B. Cloud Identity and Access Management (IAM)

C. Security Command Center

D. Cloud Key Management Service (KMS)

[KnightMotives Automotive] Diagnostic Question #3

KnightMotives plans to develop a chatbot to improve the customer service experience. This chatbot should be able to handle natural language queries related to sales, service, and vehicle features.

Which GCP service should be used to build this conversational AI?

A. Conversational AI

B. Cloud Natural Language API

C. Cloud Translation API

D. AutoML Tables

[KnightMotives Automotive] Diagnostic Question #3

KnightMotives plans to develop a chatbot to improve the customer service experience. This chatbot should be able to handle natural language queries related to sales, service, and vehicle features.

Which GCP service should be used to build this conversational AI?

A. Conversational AI

B. Cloud Natural Language API

C. Cloud Translation API

D. AutoML Tables

[KnightMotives Automotive] Diagnostic Question #4

KnightMotives's dealers have no budget for new equipment , but they need modern tools to be successful. The solution for new dealer tools must be accessible without requiring any local hardware installation or upgrades.

How should KnightMotives deploy these new tools?

A. As a desktop application that dealers must install.

B. As a mobile app for iOS and Android.

C. As a web application hosted on Cloud Run.

D. As a thick client application running on-premises at the dealerships.

[KnightMotives Automotive] Diagnostic Question #4

KnightMotives's dealers have no budget for new equipment , but they need modern tools to be successful. The solution for new dealer tools must be accessible without requiring any local hardware installation or upgrades.

How should KnightMotives deploy these new tools?

A. As a desktop application that dealers must install.

B. As a mobile app for iOS and Android.

C. As a web application hosted on Cloud Run.

D. As a thick client application running on-premises at the dealerships.

Optional materials

[VIDEOS]

- [HIGHLY RECOMMENDED] Example of how to define architecture for a serverless finance system: Designing a serverless finance system on Google Cloud
- Why you shouldn't aim at 100% uptime and what is an error budget: Why you shouldn't aim for 100% uptime
- SLIs, SLOs, SLAs in 8 mins: SLIs, SLOs, SLAs, oh my! (class SRE implements DevOps)
- DevOps vs SRE: What's the Difference Between DevOps and SRE? (class SRE implements DevOps)
- Cloud Operations Suite services: Cloud operations spotlight
- Private Service Connect: What is Private Service Connect?
- How to secure your cloud environment: How to secure your cloud environment
- Securing customer data: Securing customer data
- Network Connectivity Test: Get started with Connectivity Test in Network Intelligence Center
- Apigee: Intro to Apigee API management
- Apigee X: Introduction to Apigee X
- Securing hardware in GCP: Securing your hardware for your software
- Firewall Insights: Get Started with Firewall Insights in Network Intelligence Center
- Best Practices for Cloud Monitoring: Best Practices for Cloud Monitoring

BONUS CONTENT

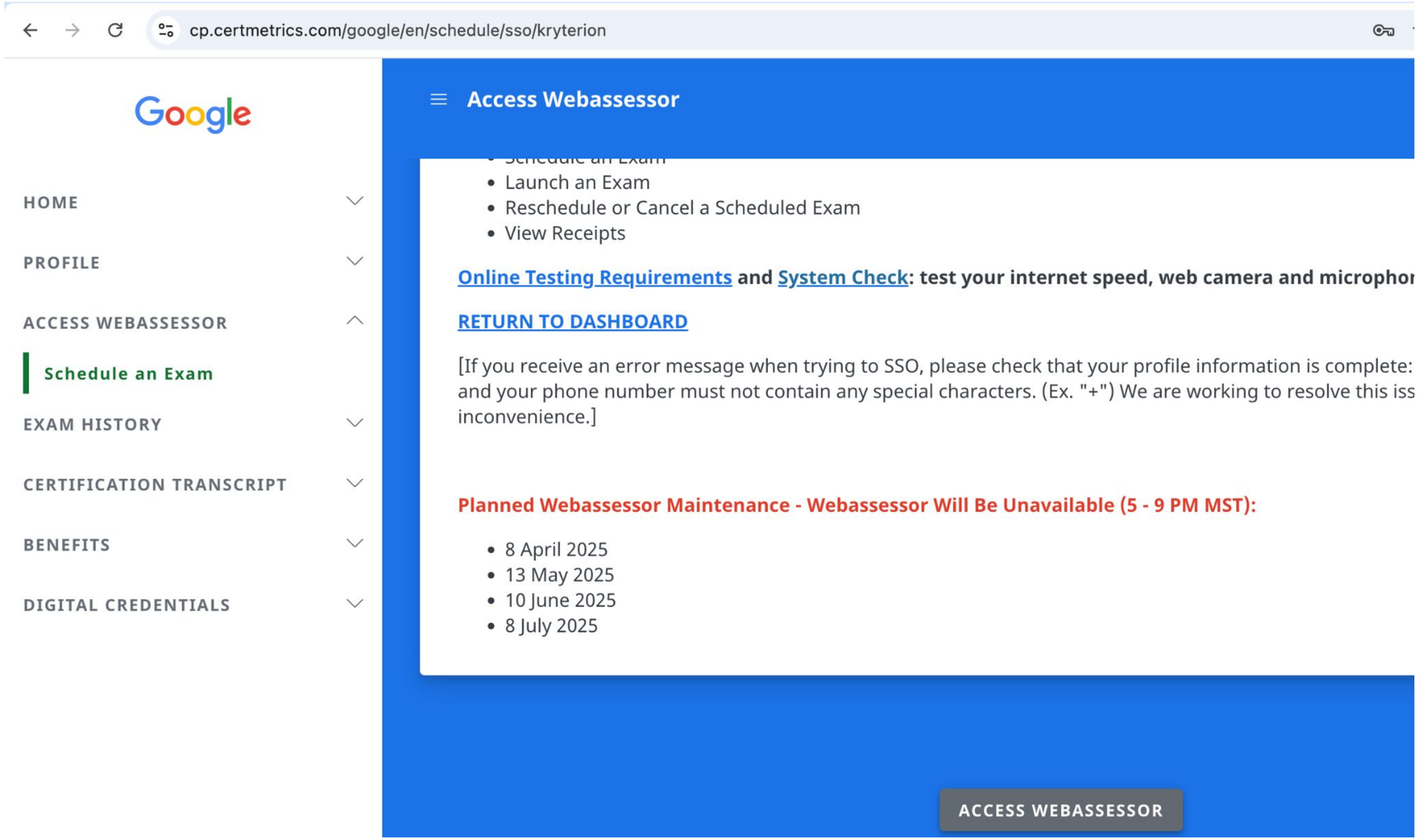
Bonus quiz

Pre-exam quiz 1

Pre-exam quiz 2

~30 exam-like questions which should help you evaluate your exam-readiness.

How to register for the exam?



Start by creating an account on Certmetrics - Google Cloud Certification platform

Taking the test: options comparison

Online

Pros

- Take it from home. No need to plan commute.
- Flexible Hours
- Can be rescheduled if needed

Cons

- Can be overwhelming during inspection
- Need extra 15 mins if you are taking a GCP test for the first time.
- Windows / MacOS / Chromebooks supported. Linux and virtual machines NOT supported!
- Personal computer strongly recommended

In-Person

Pros

- No hassle. Just walk in and start your test
- Can be rescheduled if needed
- Distractions free

Cons

- Slots fill up fast. Sign up as quickly as possible
- Plan a commute.

Click here for online testing requirements.

Q & A

Make sure to…

Enjoy the journey as much as the destination!