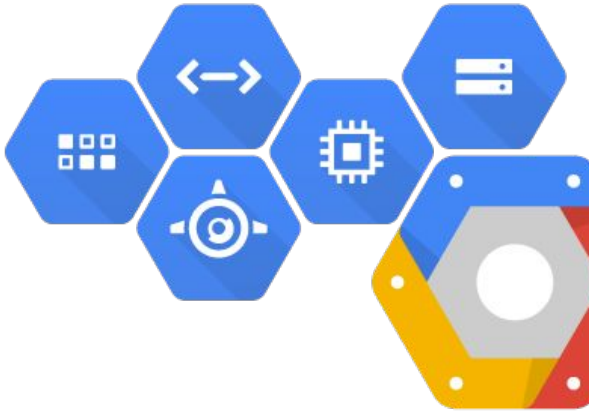
Google

Day 1: Cloud Foundations



Google Cloud Infrastructure
2-Day Training

Welcome to Google Cloud Platform Training!

This 2-day training is designed to provide an overview to help clients adopt Google Cloud Platform.

Day 1

GCP Foundational Architecture

Day 1 is focused on setting up, and configuring GCP including key concepts, terminology, and exercises

- Cloud Identity
- Identity and Access Management
- Network Configuration
- Monitoring And Logging
- Automated Operations
- Billing

Day 2

Using GCP

Day 2 is focused on moving (and developing) client application workloads into GCP including key architectures and technologies

- GKE
- Customer Use Cases
- Path to GCP Certification
- Deconstructing an Architecture Case Study

Expectations & Rules of the Road

- Assume You Have Some Basic Knowledge about Cloud And GCP Infrastructure Products and Offerings
- Please Limit Distractions (e.g., Email) And Leaving The Room
- Ask Questions and Engage!

Google Cloud

Day 1 Agenda

Topic
GCP Introduction
Why Choose Google Cloud Platform
High-Level Conceptual Architecture
Introduction of Client Scenario
Architecting with Google Cloud Platform
Cloud Identity
Identity and Access Management
Networking
Monitoring And Logging
Automated Operations
Billing
Wrap-Up

Google Cloud

GCP Introduction

Why Choose Google Cloud Platform

Why Choose Google Cloud Platform?

Google Cloud Platform enables developers to **build**, **test** and **deploy** applications on Google's *highly-scalable, secure,* and *reliable* infrastructure.

Choose from **computing**, **storage**, **big data/machine learning**, and **application** services for your *web*, *mobile*, *analytics*, and *backend* solutions.

Google

Google's Difference

1

Scaled Infrastructure

- Largest worldwide ISP (1PB/sec)
- 100,000s of miles of fiber
- 4 owned ocean cables

2

Economical

- Customer-friendly pricing
- Managed services provide PAYG with zero ops
- Scale drives lower cost profile

3

Open

- Tensor Flow, Kubernetes, Big Table
- Co-founded CNCF
- 20m lines of open code, across 900 projects

4

Secure

- Found Heartbleed and Rowhammer
- PCI, HIPAA (at no addt'l cost), FedRamp
- Crypto by default

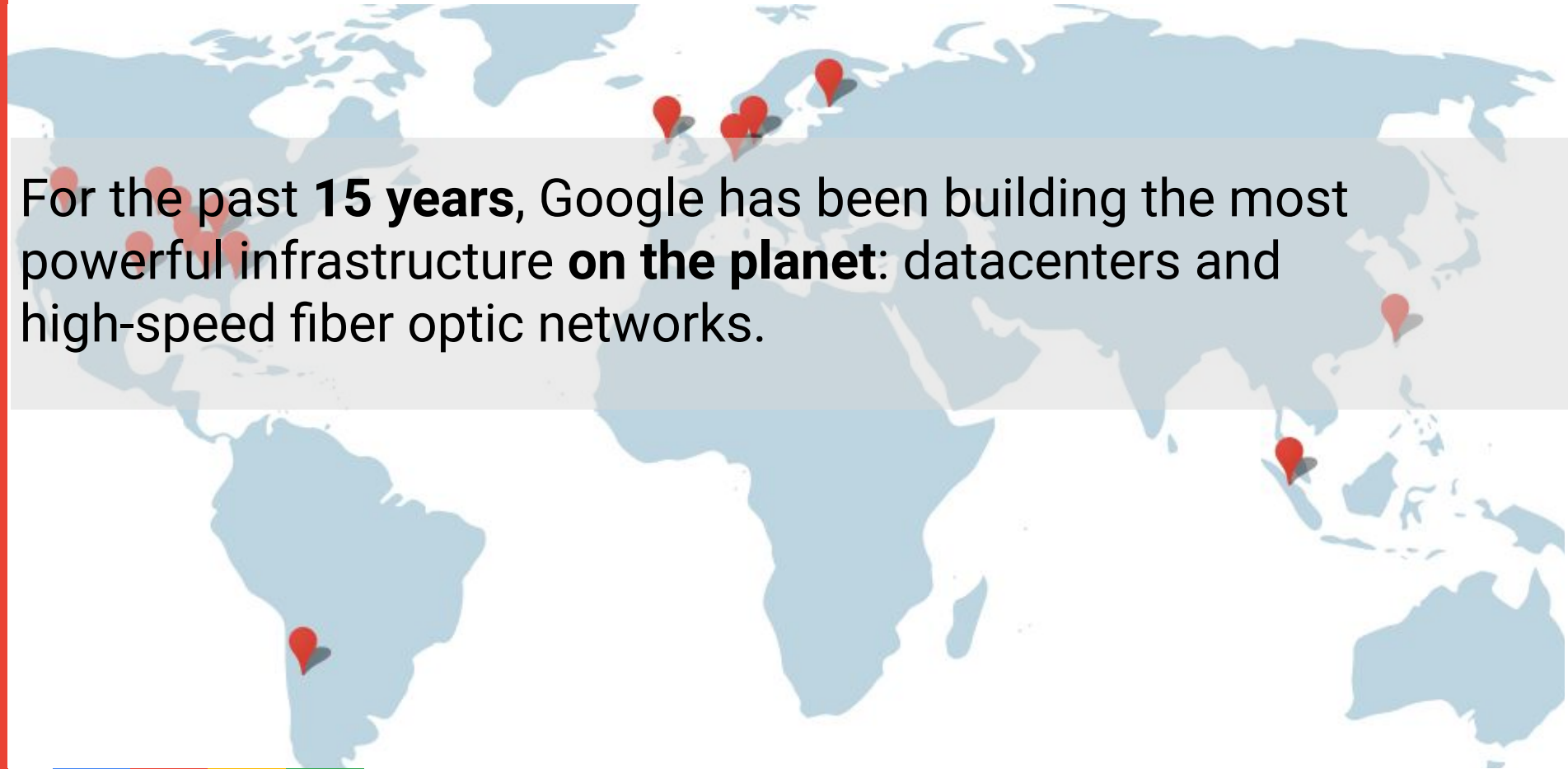
5

Innovative

- Vision, Speech, Natural Language, Prediction, Translation APIs
- 1500 Google projects using ML in prod

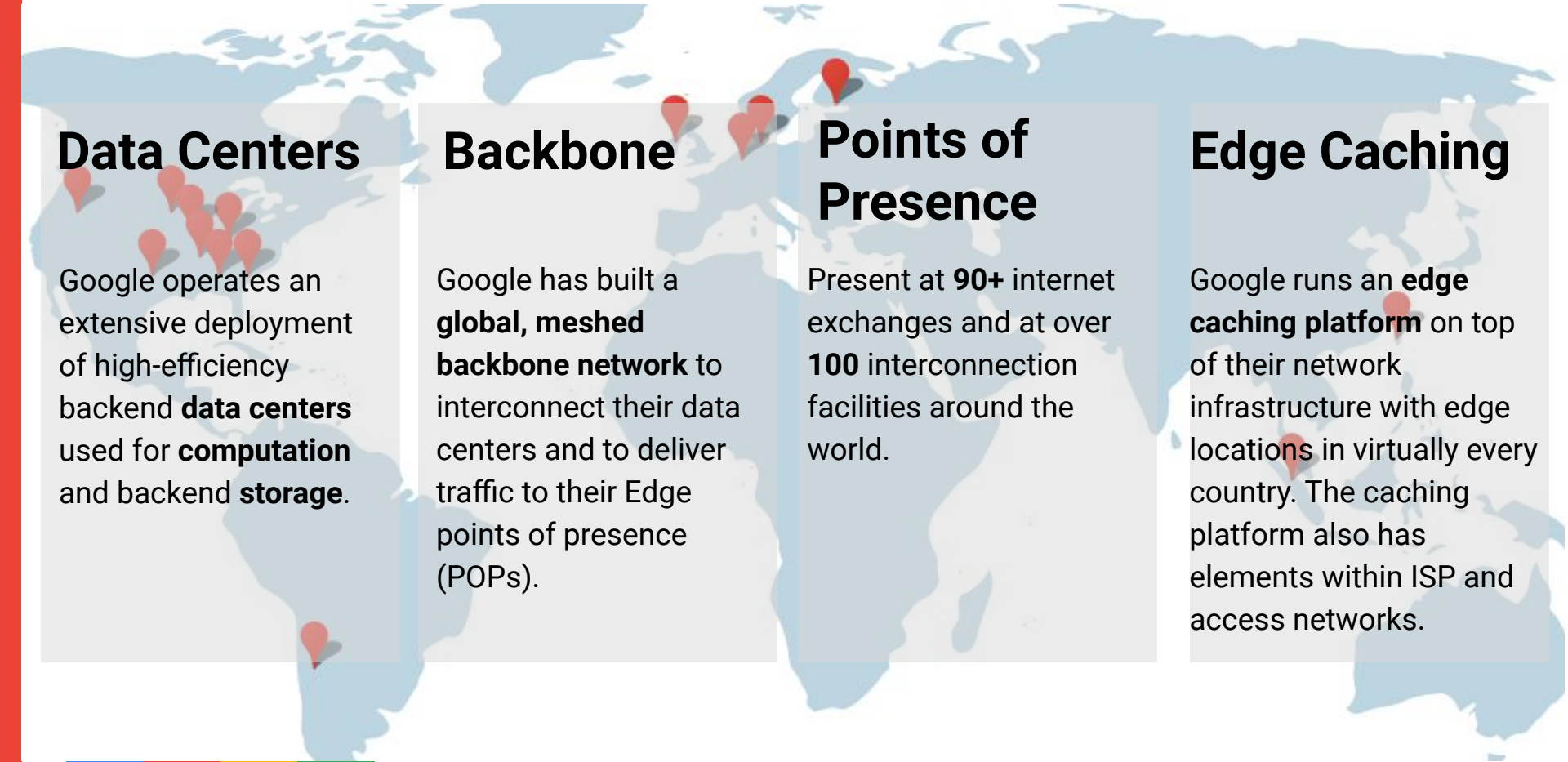
Google

Google's Infrastructure *(1 of 2)*



For the past **15 years**, Google has been building the most powerful infrastructure **on the planet**: datacenters and high-speed fiber optic networks.

Google's Infrastructure *(2 of 2)*



Data Centers

Google operates an extensive deployment of high-efficiency backend **data centers** used for **computation** and backend **storage**.

Backbone

Google has built a **global, meshed backbone network** to interconnect their data centers and to deliver traffic to their Edge points of presence (POPs).

Points of Presence

Present at **90+** internet exchanges and at over **100** interconnection facilities around the world.

Edge Caching

Google runs an **edge caching platform** on top of their network infrastructure with edge locations in virtually every country. The caching platform also has elements within ISP and access networks.

Commitment to Environmental Responsibility

Developing our infrastructure while **respecting our ecosystem**

- Pioneering **data center efficiency**
- Largest private investor in **renewables** (wind, solar)
- First data centers to receive **ISO 14001 certification**
- **100% carbon neutral** since 2007

Innovative, Customer-Friendly Pricing

Granular billing

Sustained-use **discounts**

Automatically reward users who run virtual machines for over **25%** of any calendar month

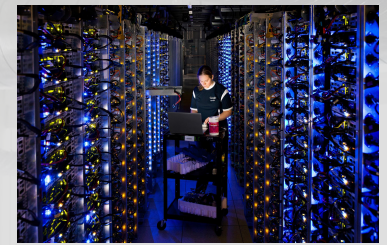
Compute Engine **custom machine types**

Pay only for the resources you need for your application

Commitment to Open APIs and Open Source

Customers should use us because they love us,
not because they are unable to migrate

Security at Scale is part of Google's DNA *(1 of 2)*

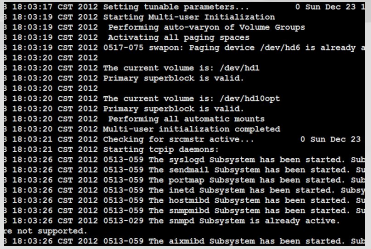


Defense in depth by default at scale

Hardshell perimeter model is insufficient

Trust through transparency

Trust not just with technology, but through transparency



Abstraction and automation

Automate best practices and prevent common mistakes at scale

Google has over **600 security experts**, EC3 Cloud Security Council

Security at Scale is part of Google's DNA *(2 of 2)*

ISO 27001

ISAE 3402 Type II

PCI DSS v3.1

ISO 27017

AICPA SOC

FedRAMP ATO
For G Suite and Google App Engine

ISO 27018

AICPA SOC

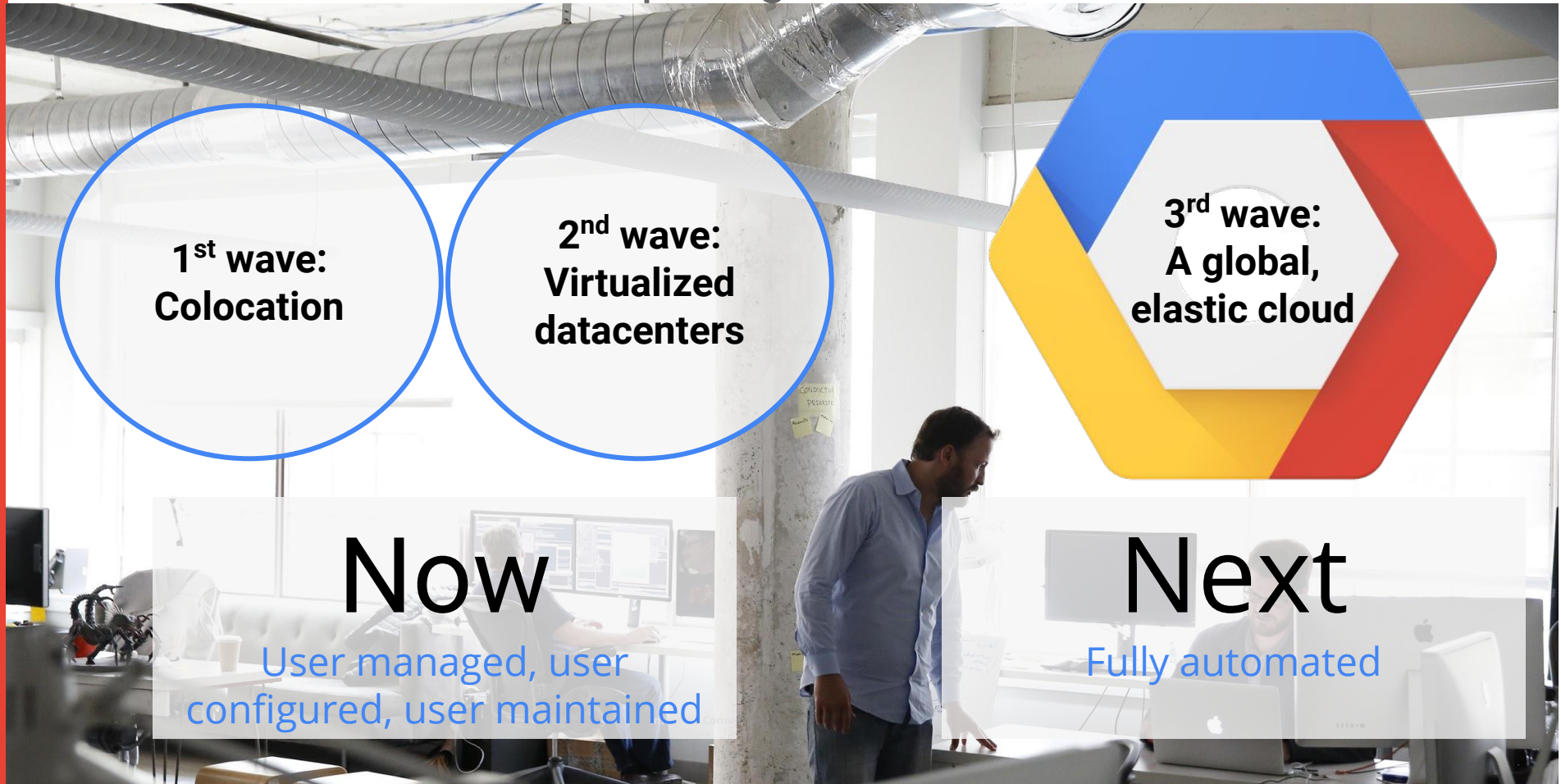
HIPAA

SSAE 15 Type II

Privacy Shield Framework

Google's **infrastructure is trusted and certified** to be in compliance with leading standards setting agencies

The Future of Cloud Computing



Hosting Support Across the Technology Stack

IaaS

Raw compute, storage and network

More granular control

Pay for what you allocate

More management overhead

PaaS

Preset run-times

Java, Go, PHP, Python...

Focus is application logic

Pay for what you use

Less management overhead

Google

Ready to Use Machine Learning Models

Cloud
Translate API

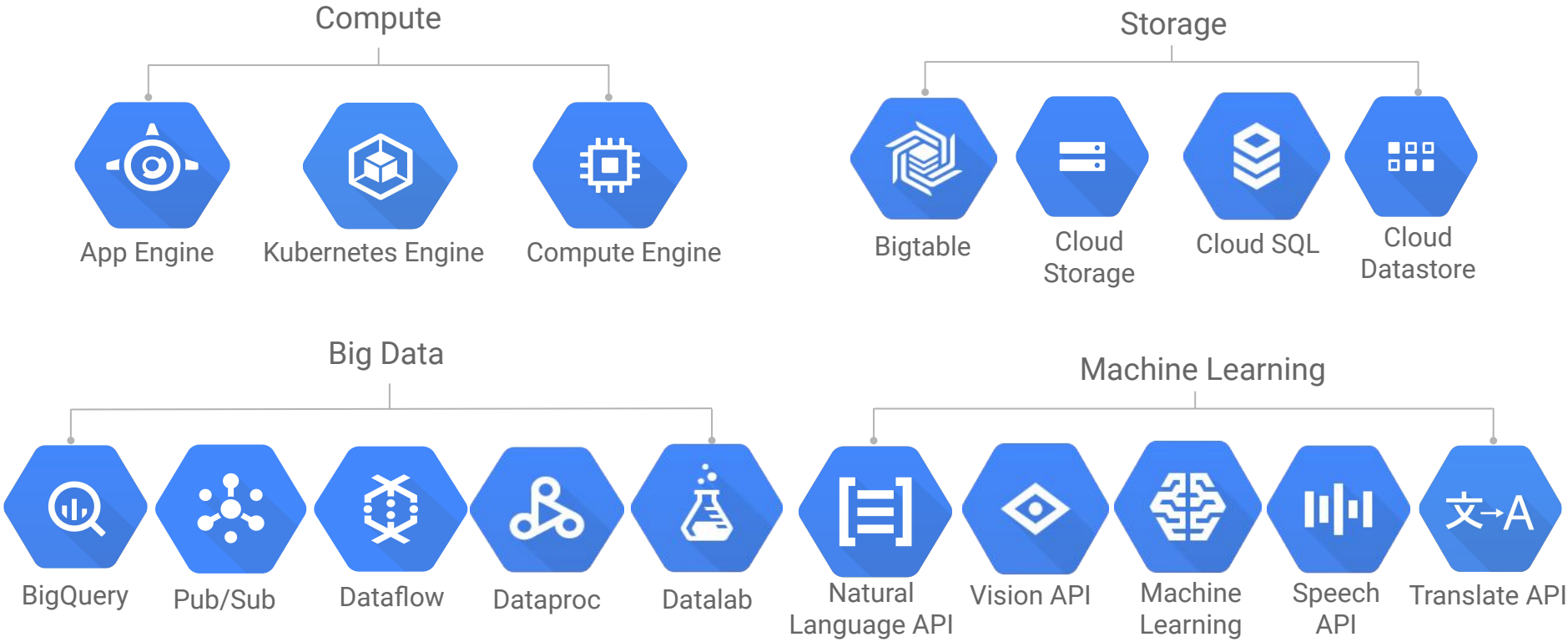
Cloud
Vision API

Cloud
Speech API

Stay tuned…

Google

Google Cloud Platform

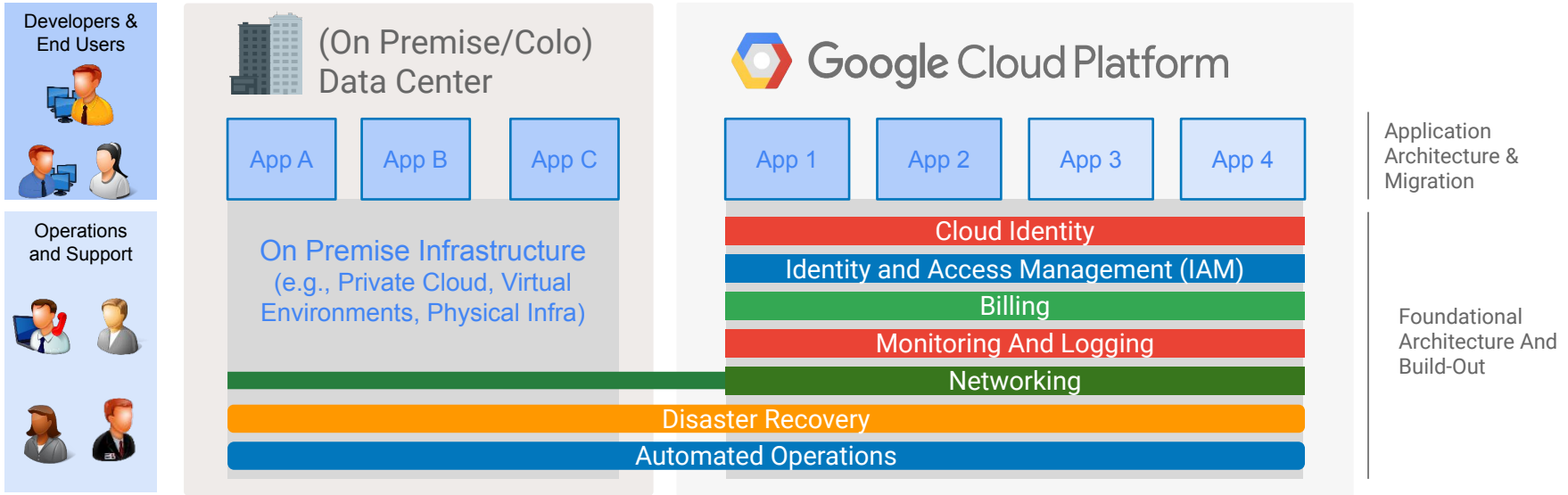


Google

High-Level Conceptual Architecture

High-Level Conceptual Architecture

Following diagram depicts a hybrid architecture adopted by many customers. GCP can be an extension to the current on prem/colo infrastructure, and leveraged for hosting workloads that can be migrated to cloud.



Use Case Introduction

Use Case Introduction

The following scenario will be referenced to at the end of each section. You will be expected to work in groups to design solutions that meet the outlined requirements.

Customer Scenario

- iRobocop is US based electronics retailer with over 200 stores
- Currently they host all their infrastructure out of 2 data centers located in the US (Oregon and California) and plan to expand online operations and improve website performance
- iRobocop wants to migrate 10 applications (including its online store) to cloud and leverage the benefits of moving to cloud such as autoscaling, flexibility, and redundancy. Management would like to keep existing investment already made in certain products such as Active Directory and Splunk, but is open to recommendations for any alternatives
- CTO would like your team to design a hybrid solution that connects their on-prem data center and future GCP infrastructure that meets performance, security and latency objectives
- Additionally, client has asked to leverage managed services wherever feasible and long term vision is to reduce on-premise footprint
- Link to customer scenario details

Architecting with Google Cloud Platform

02

Cloud Identity

Cloud Identity

This section will focus on the following key topics.

Objectives

- Key decisions for configuring Cloud Identity in GCP
- Key decisions for provisioning and managing user accounts.

Key Learnings

- Managing and provisioning user accounts
- Integrating GCP with Active Directory
- Selecting the appropriate authentication method

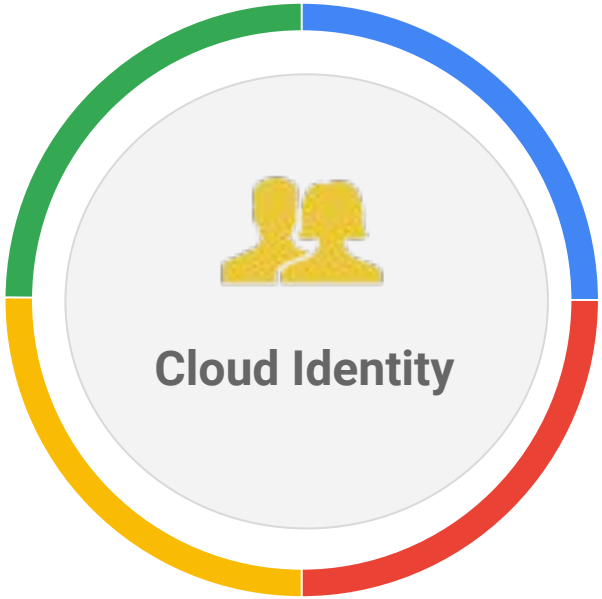
Google Cloud

How is identity managed on-premise?

Discussions related to how identity and user accounts are managed on-premise will helping in understanding how to configure identity in GCP.

How are users provisioned and managed?

Which directory service is used to manage corporate identities?



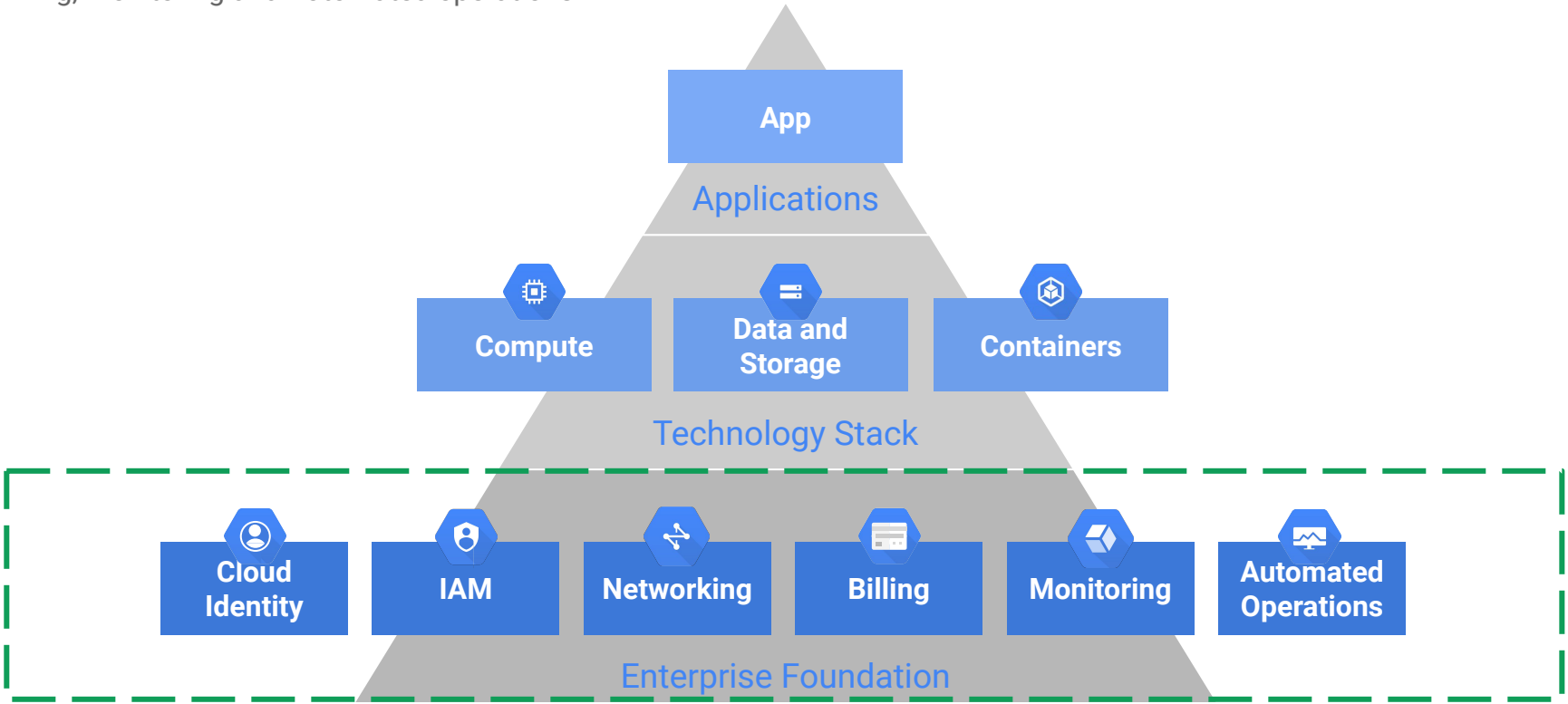
What authentication methods are used?

Do users in the organization have GCP accounts?

Google

What are the foundational elements for GCP migration?

Before migrating applications to GCP, enterprise foundations must be set up, including Cloud Identity, IAM, Networking, Billing, Monitoring and Automated operations.



Google

What do we need to know?

When evaluating a client's current identity and access management setup, there are key aspects you need to know in order to design identity and access management in GCP.

- Cloud Identity Overview
- Conflict Accounts
- Managing accounts in GCP
- Provisioning user accounts
- Integrating Active Directory with GCDS
- Integrating Active Directory using Federation
- User authentication

Google

What is Cloud Identity?

Cloud Identity is a scalable, secured, and reliable IDaaS solution for GCP user identities. It is the same identity service that powers G Suite and can also be used for 3rd party applications.

Google Cloud Account
Accounts specifically created for Google Cloud Platform whose roles and permissions are managed via the admin console and IAM for specific projects.

G Suite Account
Company-managed corporate identity using corporate credentials. G Suite accounts are recommended for accessing GCP as they grant additional control, audit, and security measures.

Consumer Account
Users with creation managed by individuals and authentication managed by Google. Consumer users are often in the @gmail.com domain; however, Google permits the use of any email address that can be verified.

On-prem Account
Company-managed corporate identity using corporate credentials utilized for an on-premises environment. GCP provides tools to sync between on-prem identities and GCP identities.

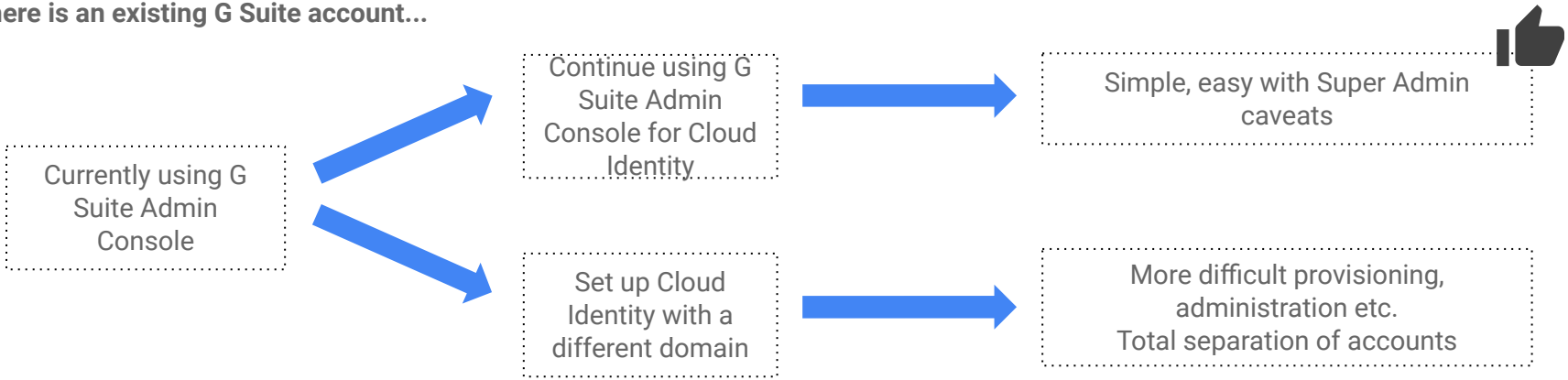
Google

When should a new G Suite account be created?

Some organizations may already have existing G Suite accounts for their users that can be used for GCP, while others will require new G Suite accounts to be created.

If there is not an existing G Suite account...

If there is an existing G Suite account...



Google

How do you resolve conflict accounts?

Users who already have access to GCP (via corporate email id) leads to ‘conflict accounts’; it is imperative to identify and resolve for such accounts.

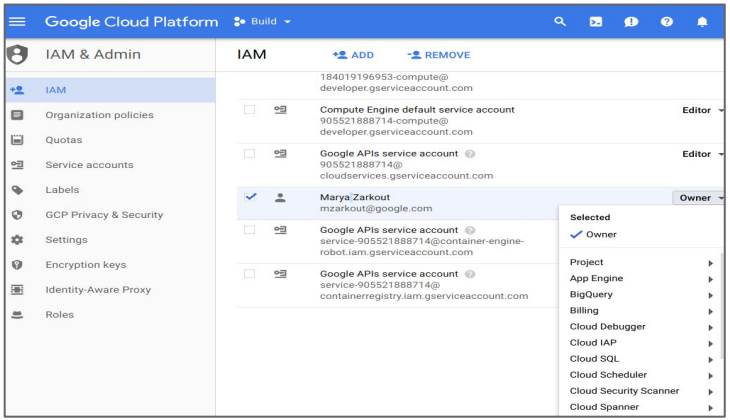
What are they?	A consumer user already using an email address that you would like to provision as an organization-managed user
How they occur?	Before your organization became a Google Cloud customer, a user signed in to use Google services as a consumer user
What to do?	Proactively identify conflict accounts using the transfer tool for unmanaged users. Consider whether a conflict account is a semi-official account using Google business services
Our options?	For accounts that should become organization-managed, request the user to join the domain. **Do not invite users using business services to join your domain!**

Google

How to manage accounts in GCP?

GCP shares its identity and authentication infrastructure with G Suite. The G Suite Admin Console is used to manage user accounts for GCP.

Cloud Console

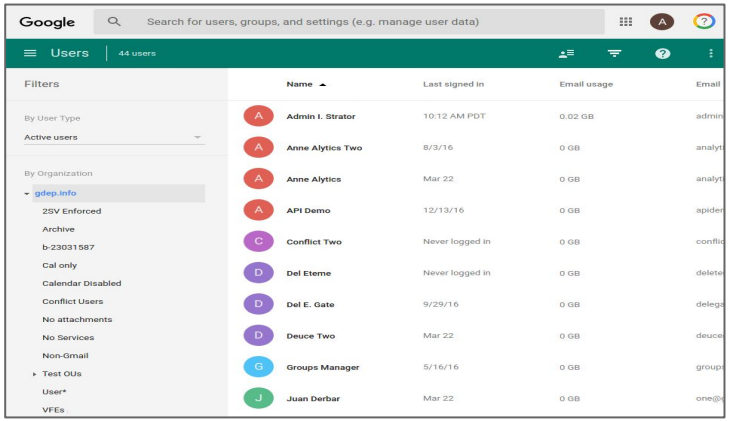


Provision Cloud Platform resources

Access and Identity management roles for Cloud Platform resources

Configure networking and on-premise integration

Admin Console



Create and manage user accounts

Create and manage groups

Enforce authentication (2FA) options

Google

What methods are available for provisioning user accounts?

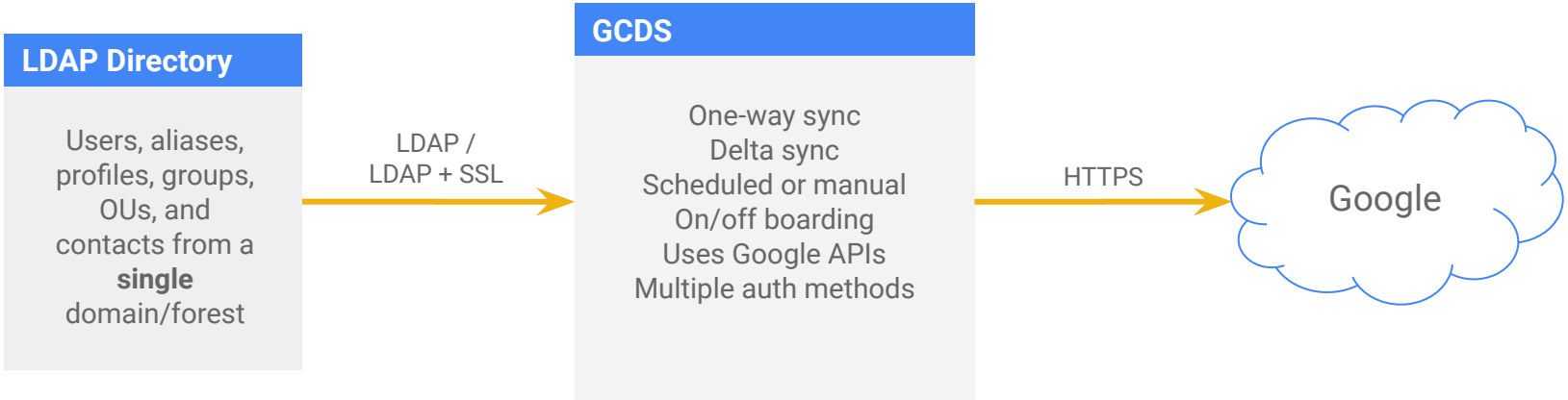
While an organization has several options for provisioning G Suite user accounts, it is recommended to use Google Cloud Directory Sync.

	Person/Role	Effort	
Manual provisioning	G Suite Superadmin	High	Lowest barrier to entry Highest ongoing effort required Not scalable
CSV upload	G Suite Superadmin	Medium	Useful for initial creation More flexible than manual Not scalable
Google Cloud Directory Sync (GCDS)	LDAP Admin + G Suite Superadmin	Medium	Integrates with LDAP Requires no programming Scalable
Admin SDK Directory API	LDAP Admin / Dev team	High	Flexible Requires in-depth programming Scalable
Third-party tools	LDAP Admin / Dev team	Medium	May incur additional cost Scalable

Google

How to integrate with Active Directory using GCDS?

Google Cloud Directory Sync (GCDS) can also be used to synchronize user account information, including user IDs and passwords, in Active Directory or LDAP server with G Suite Admin Console.



Key Considerations

Automatically provisioning users using GCDS is the preferred method to provision users

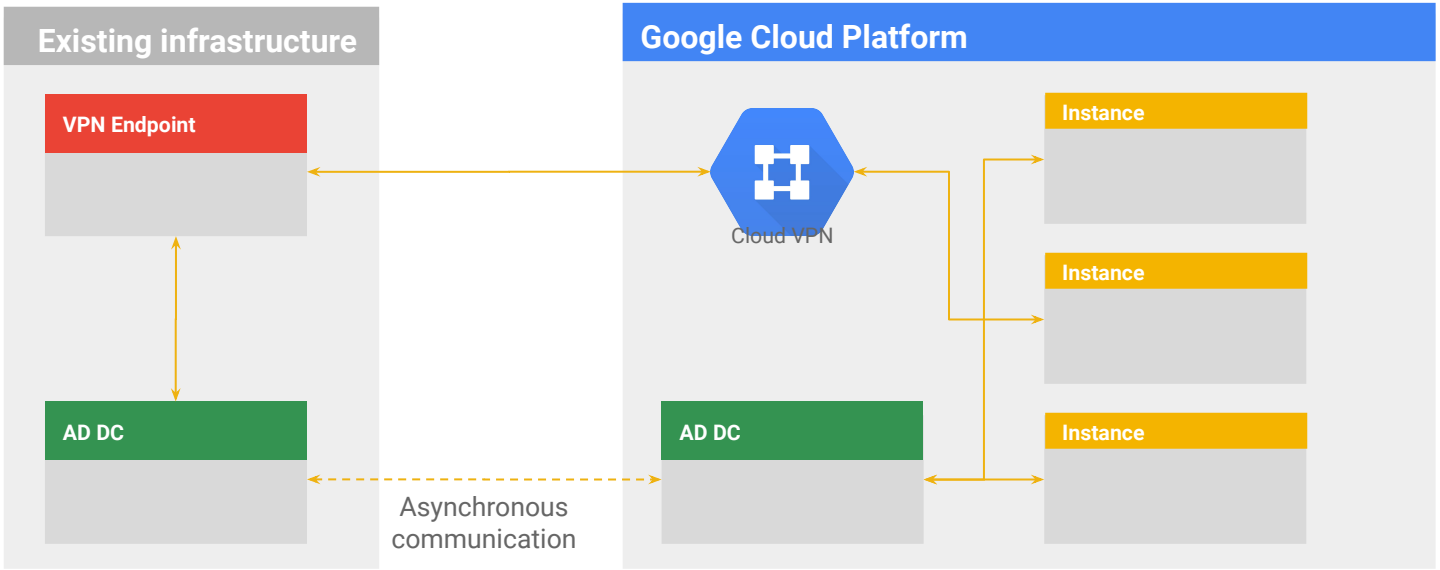
GCDS ensures your Google domain data matches your Active Directory or LDAP server

Only one-way synchronization is feasible i.e. data on LDAP server is never updated/altered

Google

How to integrate with Active Directory using federation?

For customers who would like to host their AD on GCP - Google does not provide a hosted AD service but GCE does support running AD controllers on instances.



Google

What options are available for authenticating user accounts?

For authentication of users, organizations have 3 options with GCP: Google Authentication, 2-Step Verification, and Single Sign-On (SSO).

Google Authentication

Google stored credentials, Google manages all authentication and stores the passwords

Google Authentication with 2-Step Verification

The same as Google Authentication but also using Google's 2 factor authentication tool

Single Sign-on (SSO)

SAML 2.0–based authentication *(should include MFA)*

Google

What are best practices for Cloud Identity?

Key considerations and recommendations related to managing Cloud Identity in GCP.

1. Provision users to Google's directory
2. Resolve conflicting accounts before user provisioning
3. Automate provisioning and deprovisioning of Users and Groups
4. Centralize authentication including multi-factor either through Google or a third party Identity Provider

Google

What are some key decisions to make?

When setting up Cloud Identity, there are several key decisions an organization must make.

1. Which domain(s) will be the primary domain(s)?
2. How will users be provisioned?
3. How will Active Directory be integrated?
4. Which authentication method will be used?

Google

How can I test my understanding of Cloud Identity?

Work within your team to reinforce concepts by applying them to a real life use case. It is essential to consider the customer's business and technical requirements when designing the solution.

Objective

- Help iRobocop synchronize corporate users and groups from AD to Google domain
- Leverage existing tools and services to manage lifecycle (create, update, delete) of users in Google domain
- Help iRobocop security team to manage life cycle of users through one source (AD versus in GCP console)

Problem Statement

- iRobocop GCP onboarding team is handling requests for creating users and assigning IAM roles; they have enabled Cloud Identity so that users can utilize the organization email address
- Process of creating new users and assigning Cloud IAM roles being followed is:
  - new user submits an access request via an internal ticketing system
  - upon approval of the ticket
    - Security team will provision user's ID in AD
    - GCP onboarding team will create user in GCP and assign appropriate GCP IAM role
- iRobocop would like to streamline process of handling access requests to GCP i.e. managing users in 1 location
- **Please design a process and solution in assigning users to GCP**

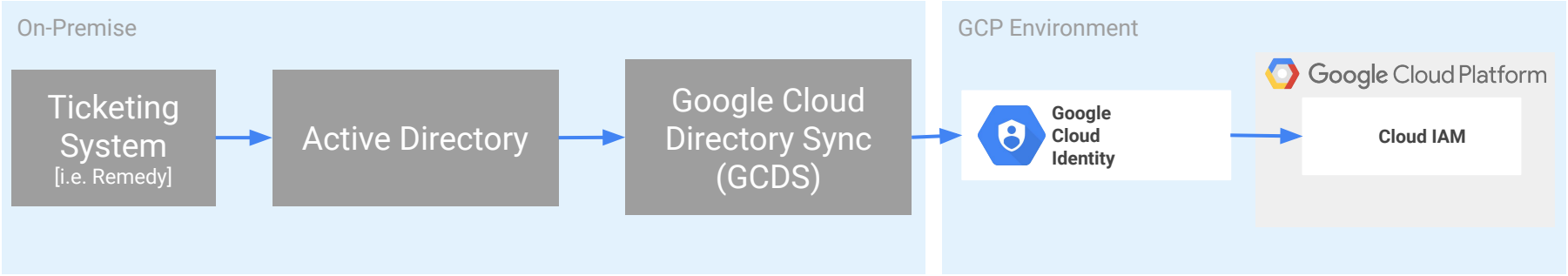
Google

How can I test my understanding of Cloud Identity?

Work within your team to reinforce concepts by applying them to a real life use case. It is essential to consider the customer's business and technical requirements when designing the solution.

Solution

1. User submits an access request via an internal ticketing system
2. Security team upon approval of the ticket will provision and manage user's ID to Google IAM group attributes in AD
3. Configure GCDS job to run at regular intervals to sync to Google domain - upon successful sync job, user is provisioned in GCP
4. Configure GCDS to notify org admin or security team of the sync event
5. On successful sync to Google domain, org admin or security team will assign IAM role to users (incase user does not belong to any IAM groups)



Google

Identity and Access Management

Identity and Access management

This section will focus on the following key topics.

Objectives

- Key considerations for configuring identity and access management in GCP related to assigning permissions
- Key considerations for setting up a project structure

Key Learnings

- Understanding how roles are used to grant permissions in GCP
- Designing project structures in organizations
- Leveraging groups and assigning user roles

Google Cloud

How is identity and access managed for users?

Discussions related to IAM setup in the on-premise infrastructure will help in understanding access for users.

What are various levels of permissions granted?

What are some current user groups used?

Do business units develop applications independently?

How are keys managed and rotated?

Google

What do we need to know?

There are key aspects you need to know in order to design a client's access management strategy in GCP.

- Cloud Resource Manager
- Designing a project structure
- IAM Roles and Custom Roles
- Using groups to manage permissions
- Service Accounts
- Secrets and Key Management
- Resources Quotas

Google Cloud

What is Cloud Resource Manager?

Also known as organization node, Cloud Resource Manager (CRM) helps set organization wide access control policies, and collate various projects under one organization.

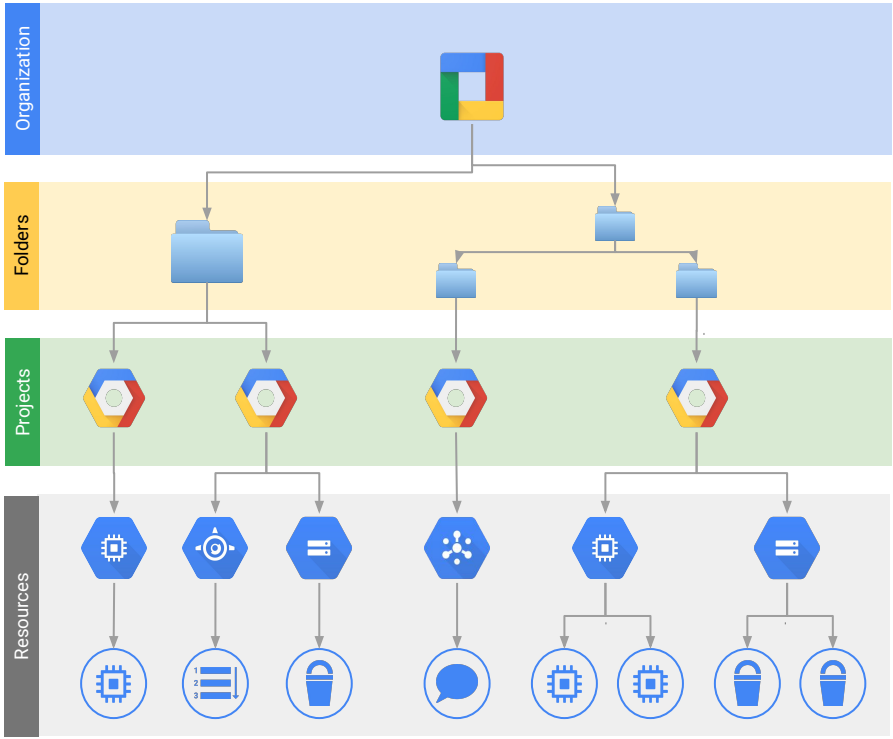
Organization is root node in GCP resource hierarchy and the hierarchical super-node of projects

Folder(s) can contain projects, can be up to four levels deep, or a combination of both

Projects form basis for creating, enabling, and using GCP services including managing APIs, enabling billing, managing IAM, and managing permissions for GCP resources

Resources are the fundamental components that make up all GCP services

Top-down inheritance: additive only



The effective policy for a resource is the union of the policy set at that resource and the policy inherited from its parent.

Google

How to design a project structure?

There are many design principles that must be evaluated when designing the appropriate project structure, including complexity, IAM, and cross-project networking.

	Few projects	Many projects
Project complexity	Low	High
IAM complexity	High	Low
Cross-project network traffic	Less frequent - N/A	More frequent - VPN/XPN
Least privilege	More difficult	Less difficult

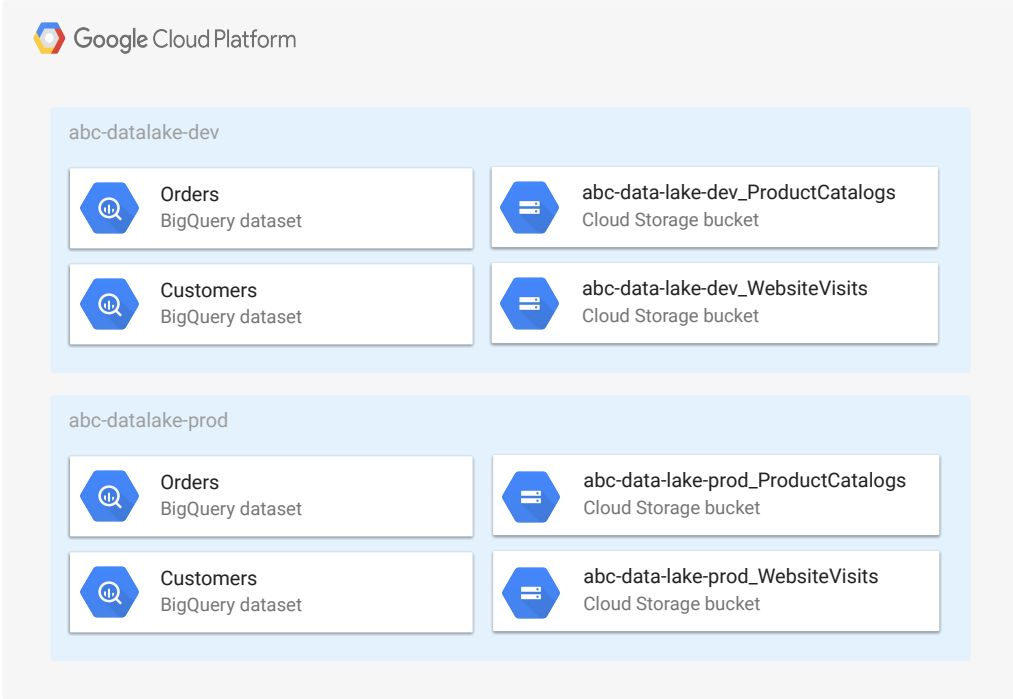
Google

When to use a project per environment?

Using a project per environment is one of the potential approaches. Below we outline the design considerations when using this solution.

Considerations for solution

- Possibilities for different billing accounts, some boundaries
- Separate quotas and limits
- Access management can be different for DEV and PROD
- Misconfiguration issues only impact a single environment
- Separate project-level BQ custom quotas
- Business units and data sensitivity classification mixed



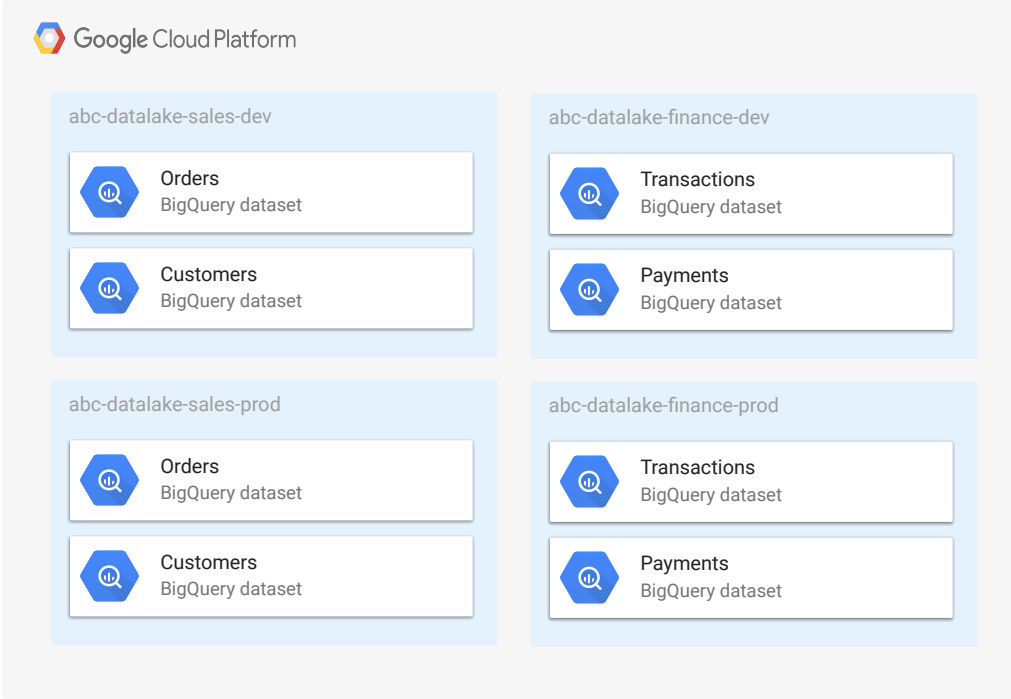
Google

When to use a project per application per environment?

Using a project per application per environment is another common approach. Below we outline the design considerations when using this solution.

Considerations for solution

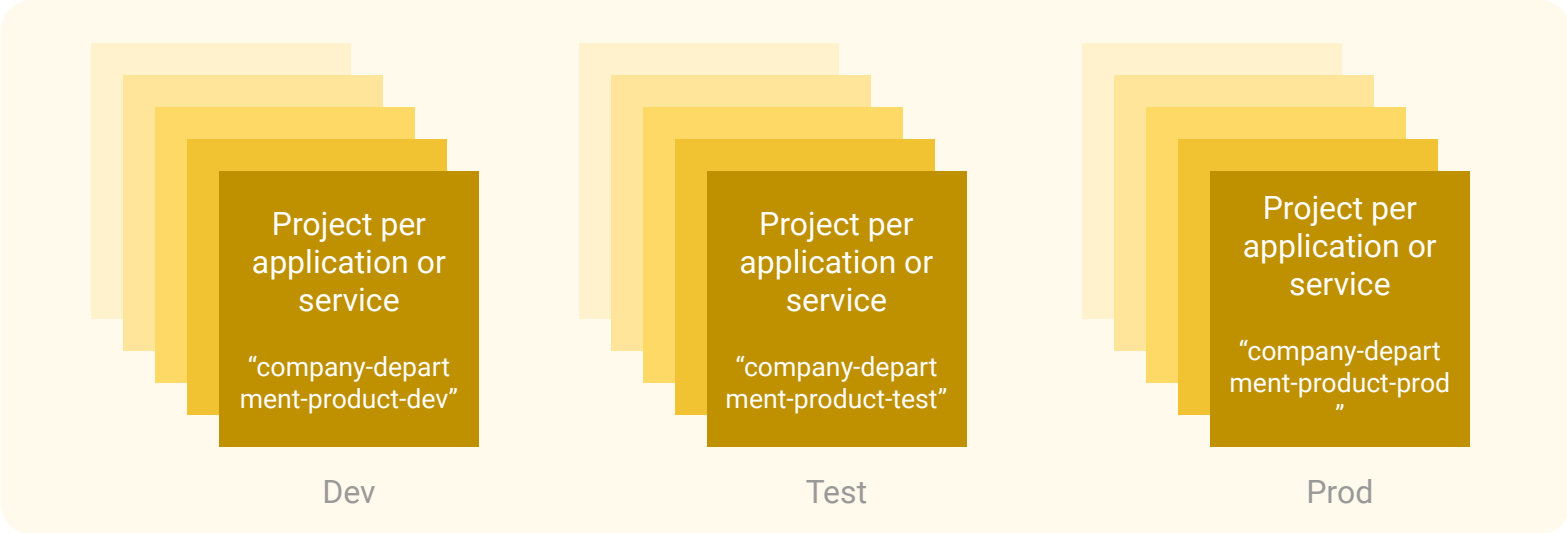
- Possibilities for different billing accounts, some boundaries
- Separate quotas and limits
- Access management can be different for DEV and PROD
- **Misconfiguration issues only impact a single environment and BU**
- Separate project-level BQ custom quotas
- **Potentially more management coordination**



Google

What is the recommended project organization?

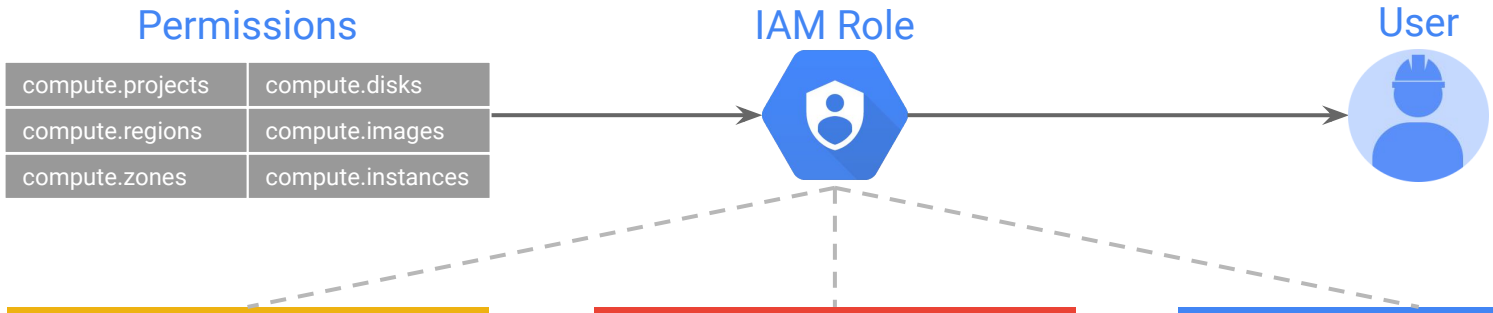
The recommended project organization is to have a project per application or service for each workstream using a consistent naming scheme for the projects.



Google Cloud

What are IAM roles?

Specific IAM role(s) are granted on a per-project basis to users or groups allowing organizations to effectively control permissions, enable onboarding and perform audits.



Control

Granular roles enable very precise permissions management

Enforce separation of duties across Google Cloud Platform

Enable

IAM Predefined roles enable rapid and simple onboarding

Use Custom Roles for further granularity

Utilize Group membership to grant access to resources

Audit

Full log of permissions changes

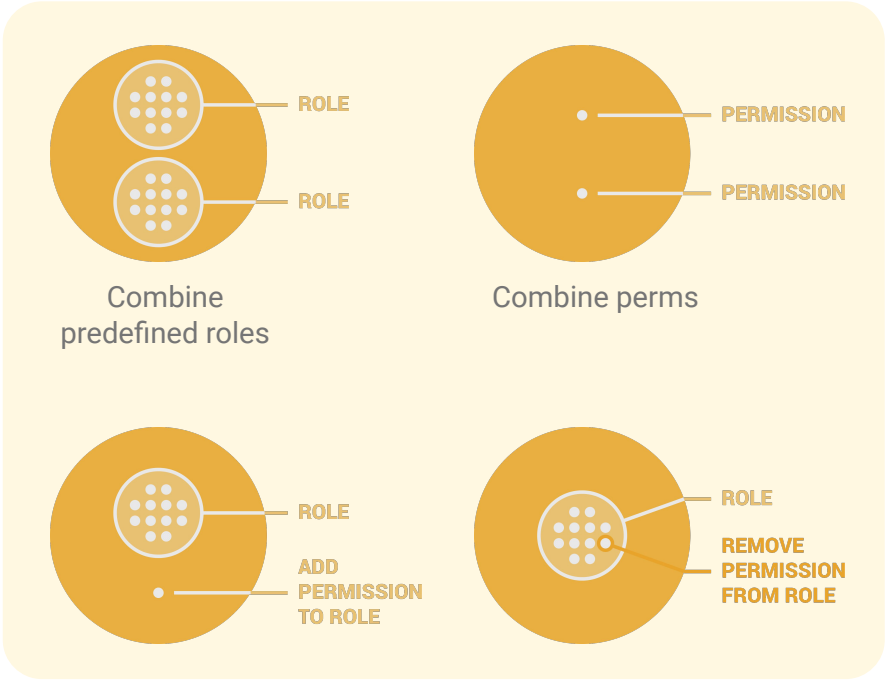
Cloud IAM lets you focus on business policies around your resources, and helps facilitate compliance

Google

What roles are available ?

GCP offers several predefined IAM roles that enables managing access to GCP resources at a granular level i.e.who (identity) has what access (role) for which resource.

- **Primitive roles** are legacy GCP roles that grant broader set of permissions (Owner, Editor, Viewer)
- **Predefined roles** provide permissions (specific actions allowed) bundled together for various job functions.
- **Custom roles** provide granular control over the exact permissions provided to a role



Google

How to leverage groups to manage IAM roles?

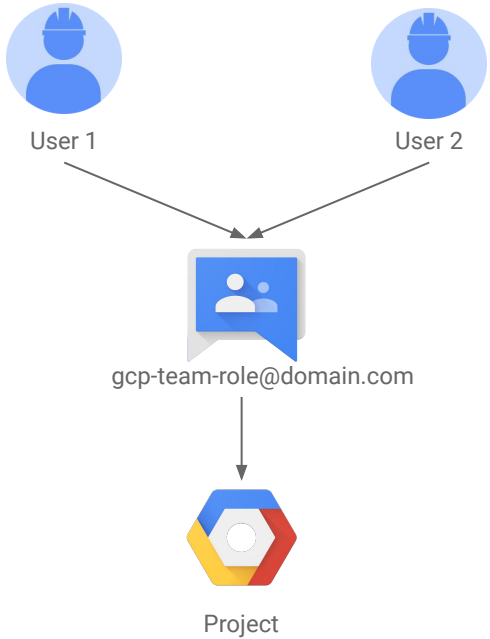
It is recommended best practice to use user groups to manage IAM roles, eliminating the need to manage roles for individual user accounts.

Users can be added to groups to ease the management of roles for individual users:

- Groups can be nested as well
- A Group can be used for each project
- A Group can be used for each team of users

Google Recommends

- Creating Groups for each team
- Using Logical Group naming scheme (can allow for easy automation), for example gcp-{team}-{role} or gcp-{project}-{role}
- Temporary users can be added to the Group and removed when needed



Google

What are the different types of identity in GCP?

In addition to user accounts, there is a second type of identity, called service accounts that are assigned to a machine/application (e.g. GCP resources).

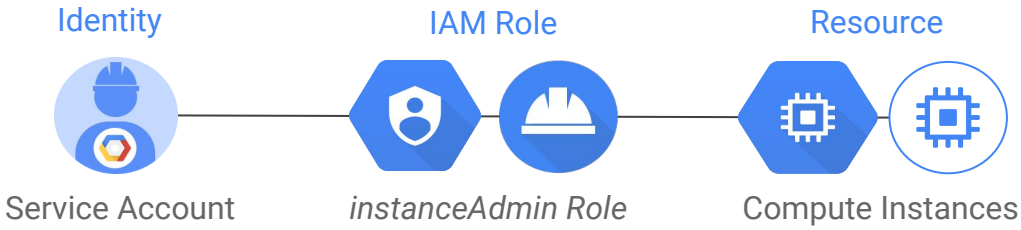
	 User Accounts	 Service Accounts	 Groups	 Domain
Identity	Human	Machine	Both	G Suite Accounts
Authentication	Google password or SSO	Keys	Via User	Via User
Authorization	IAM Roles	IAM Roles	IAM Roles	IAM Roles
Management	Admin Console	Cloud Console	Admin Console	Admin Console

Google

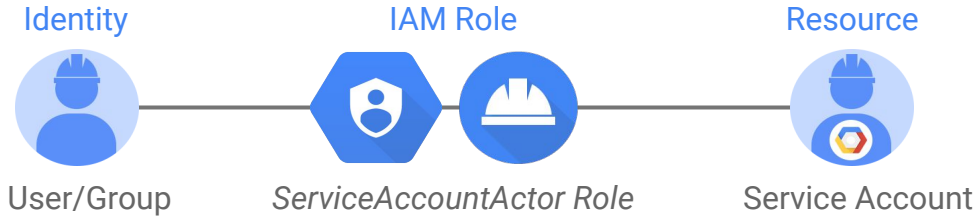
What are service accounts?

Service accounts authenticate applications running on GCE instances to other GCP services. They allow applications to authenticate seamlessly to APIs without the need of hidden keys or user credentials.

Service accounts are identities…



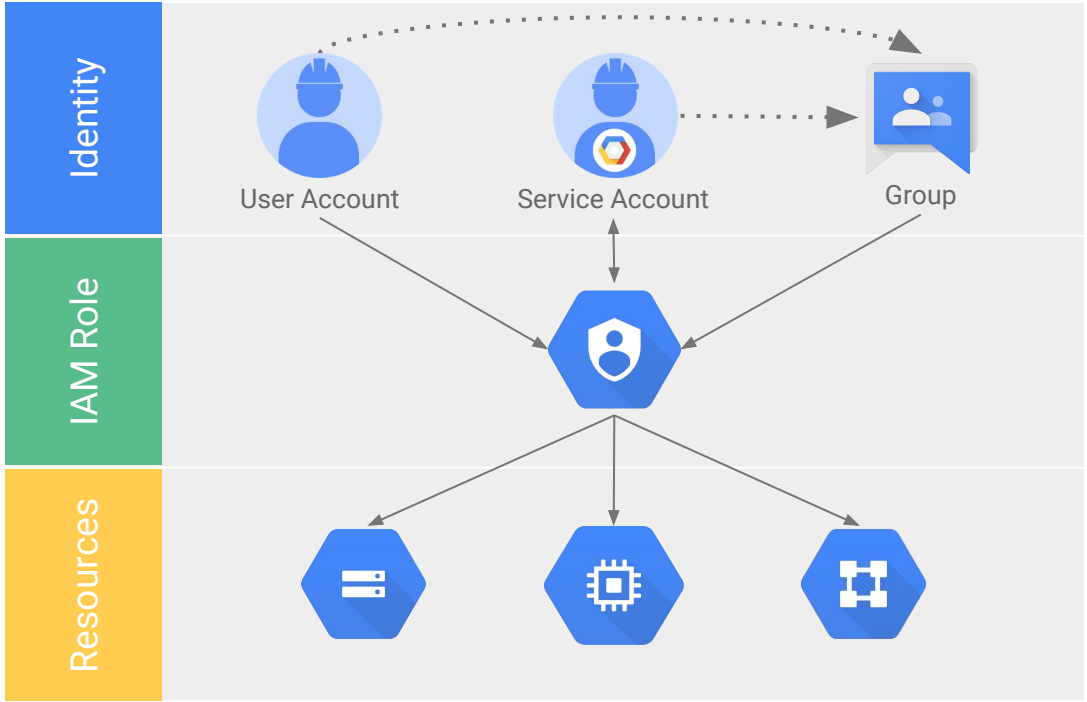
Service accounts are also resources...



Google

How do different types of identities interact?

Users, service accounts, and groups can be assigned IAM roles, and through IAM roles they can be granted access to resources.



Google

What IAM services are audited in GCP?

GCP keeps a record of user, admin, permissions and service accounts activity.

Users

Reports section of admin console

Reports API to generate custom reports

Login audit log in admin console

Admins

Admin SDK audit logs

Configure admin alerts for suspicious logins, admin privilege grants, etc.

Permissions

IAM API used to audit permission changes

Stackdriver Logging shows events

Service accounts

Cloud logs contain access logs

Google

How are keys managed for service accounts?

There are two methods for managing keys, GCP managed keys and customer managed keys; it is recommended to use GCP managed keys wherever possible.

GCP-managed keys

Available when running inside GCP

Keys cannot be downloaded

Keys automatically rotated

Customer-managed keys

Created, downloaded, and managed by users

User must store, distribute, revoke, rotate, and recover keys

Rotate/audit keys via API

Google

What are Secrets and how can they be managed in GCP?

Applications often require access to small pieces of sensitive data at build or run time, referred to as Secrets. There are three methods to manage secrets:

Encrypt w/ Cloud KMS	• Store encrypted secrets at the application layer by using Google Cloud Storage • Requires access to code and corresponding key • Helps protect against insider threat
Google Cloud Storage	• Store secrets in GCS, encrypted at rest • Limits access to smaller set of developers • Ability to audit access
Third-party solution	• Dedicated secret management tool • May have flexibility to rotate keys on your behalf • May require additional investment

Google

What are the benefits of Google Cloud KMS?

For secret management using Cloud KMS, the ideal configuration should minimize unnecessary access and enforce separation of duties.

- **Cloud-hosted key management service** lets you manage encryption for your cloud services the same way you do on-premises.

- You can **generate, use, rotate, and destroy AES256 encryption keys**. Cloud KMS is **integrated with IAM and Cloud Audit Logging** so that you can manage permissions on individual keys and monitor how these are used.

Features

- AES256 key
- Encrypt and decrypt via API
- Automated and at-will key rotation
- Delay for key destruction
- High global availability

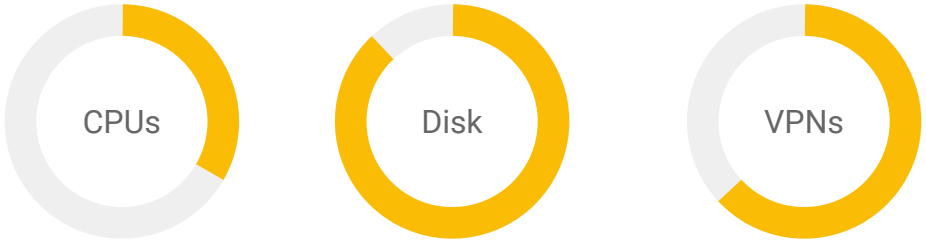
Key versions	**Price ***
Active key versions	$0.06 per month
Key operations	**Price ***
Key use operations (encrypt/decrypt)	$0.03 per 10,000 operations
Key admin operations	Free

* *Prices as of September 2017*

Google

What are Resource quotas?

Compute Engine enforces quotas on resource usage for a variety of reasons. They provide protection from cost overrun and poorly behaved GCP customers.



Quotas provide protection from cost overrun, and can be indicators of bad code

Quotas assist in isolation from other poorly behaved GCP customers

As your use of Google Cloud Platform expands over time, your quotas may increase accordingly

Most quotas are applied per project

Google

How to increase resource quotas?

Small quota increases can be requested via console and usually approved right away; it is recommended to work with your Google account team before significant milestones/go-live dates.

Systematic requests

- Collaborative capacity planning strongly recommended
- Make quota increase requests well ahead of anticipated need

Emergency request

DO NOT DO THIS UNLESS CRITICAL

- Production that exceeds quota limits can be handled on an emergency basis
- Coordinated with either SCE or account manager

Google

What are best practices for IAM?

Key considerations and recommendations related to Identity and Access Management

1. Use projects to control access to and isolate resources as necessary
2. Leverage groups to assign roles to several users
3. Authorize server-to-server interactions with service accounts
4. Google recommends one project per application per environment in order to organize resources

Google

What are some key decisions to make?

When setting up Cloud Identity, there are several key decisions an organization must make.

1. What folder structure will be used to implement IAM and organizational policies?
2. Who will be the organization, folders, and billing admins?
3. How should projects be used to accomplish resource segmentation and billing detail?
4. What quota will be required for resources?
5. How will least-privileged access be implemented?
6. How will service account keys be managed?
7. How will access be audited?
8. How will application secrets be managed and encrypted?

Google

How to access additional resources?

Learn more about IAM through public documentation links, tutorials, and videos.

Documentation

- Basic cloud IAM concepts
- Understanding roles
- Service Accounts
- Best Practices
  - Using Resource Hierarchy for Access Control
  - Understanding Service Accounts
  - Using IAM Securely
  - IAM roles for Billing-related Job Functions
- FAQs

Tutorials

- Managing roles and permissions
- Using service accounts

Videos

- Best practices for IAM on Compute Engine
- Using Google's Identity as a Service (IDaaS) for GCP and G Suite

Google

How can I test my understanding of IAM?

Work within your team to reinforce concepts by applying them to a real life use case. It is essential to consider the customer's business and technical requirements when designing the solution.

Objective

- Help client design an organization node and consider appropriate roles for different customer employees.

Problem Statement

- **Part 1:** Based on workshops conducted by Deloitte team to understand organization structure and collect inventories, help iRobocop come up with RBAC to avoid human error.
  - iRobocop has 3 non-production environments - Dev, QA, and Test
  - Codes is intended to go through rigorous testing in these environments before being deployed to Prod env.
  - Each application has:
    - a different development team
    - maintains their own non-production environments
    - has dedicated production support team
- **Part 2:** Help Security team implement least privilege access by suggesting roles for following employees: System Solutions Architect, Full Stack Ecommerce Developer, Finance Associate, Systems Administrator, Shared Services Engineer.

Google

How can I test my understanding of IAM?

Work within your team to reinforce concepts by applying them to a real life use case. It is essential to consider the customer's business and technical requirements when designing the solution.

Solution: Part 1

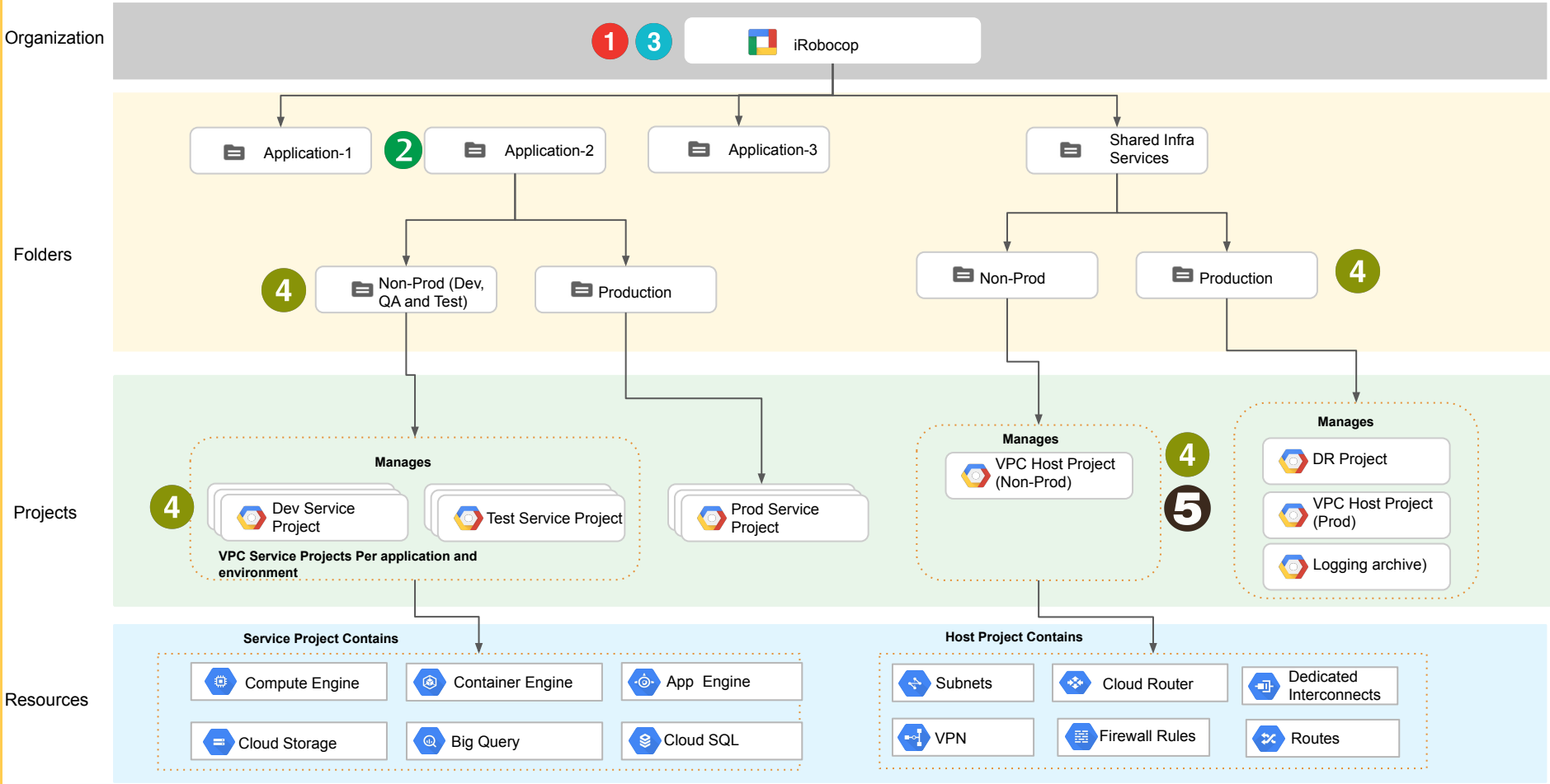
1. Sample Organization Node that segregates application environments and roles (Note: refer next slide)
2. iRobocop have created folder that represent group of applications and share the same development team
3. Parent folder is further divided into production & non-production folders; assign separate Cloud IAM roles to both folders
4. Production folder is restricted for developers. Both Cloud IAM and Organization policies are inherited through hierarchy, hence GCP projects created at each node of the hierarchy, will inherit Cloud IAM roles and policies define at higher level
5. By default, GCP projects created under Production folder will not grant access to developers

Solution: Part 2

1. System Solutions Architect - resourcemanager.organizationAdmin
2. Full Stack Ecommerce Developer - Folder Admin
3. Finance Associate - billing.viewer
4. Systems Administrator - compute.instancAdmin.v1, compute.networkAdmin, compute.securityAdmin
5. Shared Services Engineer - compute.xpnAdmin, compute.networkAdmin, compute.securityAdmin

It is recommended to assign these roles to groups versus individuals.

Google



Google

Networking

Networking

The section will focus on the following key topics

Objectives

- Key considerations for designing networks in GCP
- Key considerations for integrating with an organization's on-premise infrastructure

Key Learnings

- Understanding of GCP network structure in relation to projects, regions and zones
- Controlling ingress and egress traffic
- Enabling a secure connection to a GCP network
- Selecting a connectivity method for integrating with on-premise infrastructure

Google Cloud

What is the current networking footprint?

Understanding the mapping of various components in the on-premise network(s) will help create a blueprint of the current network architecture.

How many networks and subnets do you have?

Are there any specific network speed and latency requirements?

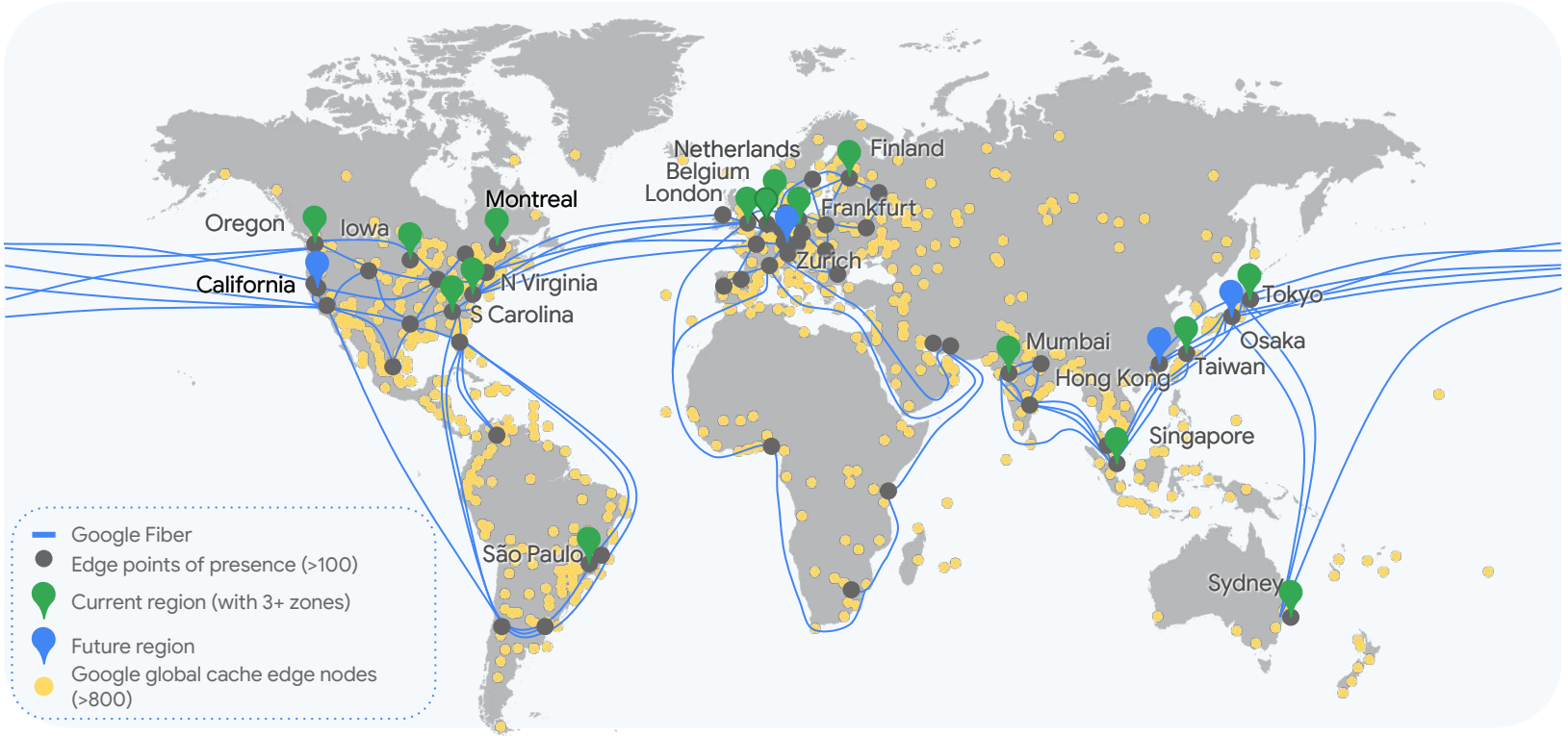
Is a tiered network architecture used? How are applications isolated?

What security features do you have on your network - firewalls, load balancers, bastion host etc.?

Google

What does Google's global network infrastructure look like?

Hundreds of thousands of miles of fiber optic cable connect all of Google's datacenter regions and 100+ points of presence.



Google Cloud

Where are GCP's datacenters located?

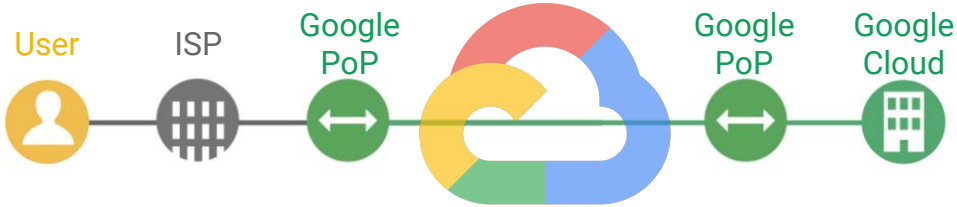
Google have Data centers located across the globe, spanning 18 current Regions, 55 Zones and over 100 points of presence, allowing organizations to build their GCP network in the location of their choice.

Google

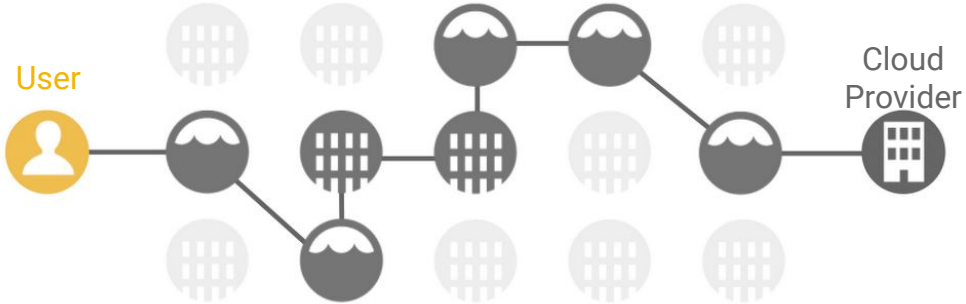
How is connecting to GCP different than other providers?

Unlike Google, most cloud providers use what's called "hot potato" routing i.e. when you're sending out traffic from their infrastructure, they offload it to the public internet as soon as possible.

Google Cloud



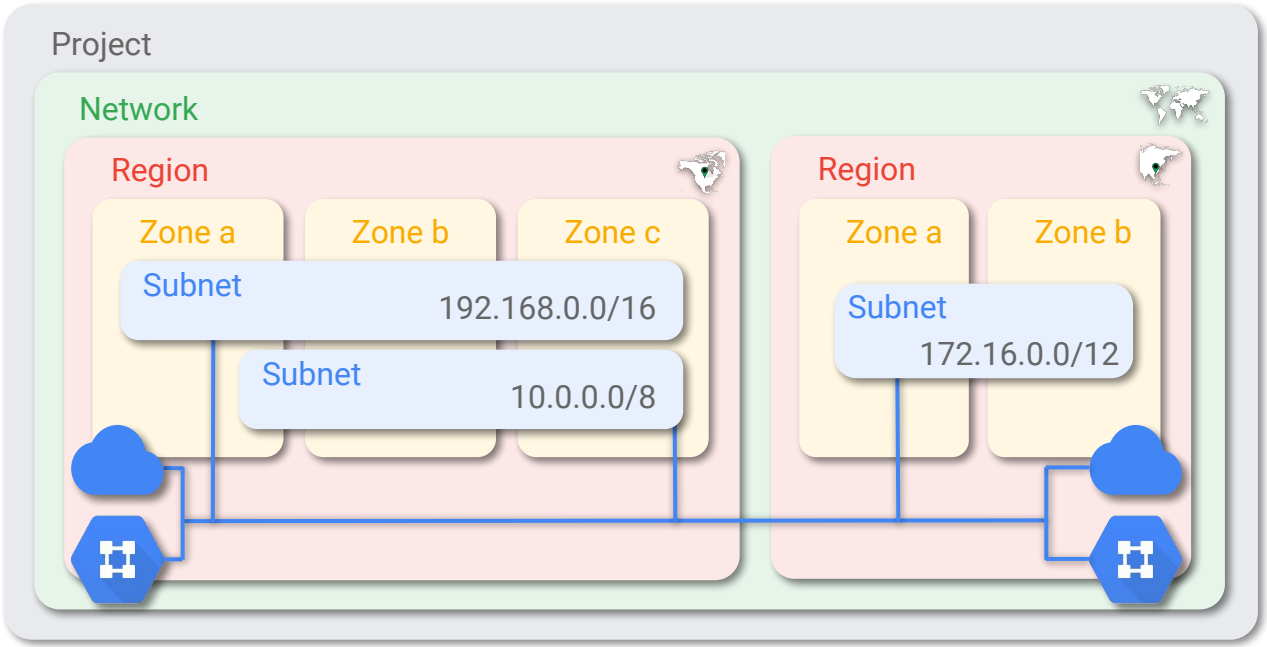
Cloud providers that don't own their network...



What are some foundational GCP Networking components?

A VPC network, sometimes just called a “network,” is a virtual version of a physical network, like a data center network. It provides connectivity for your GCP resources.

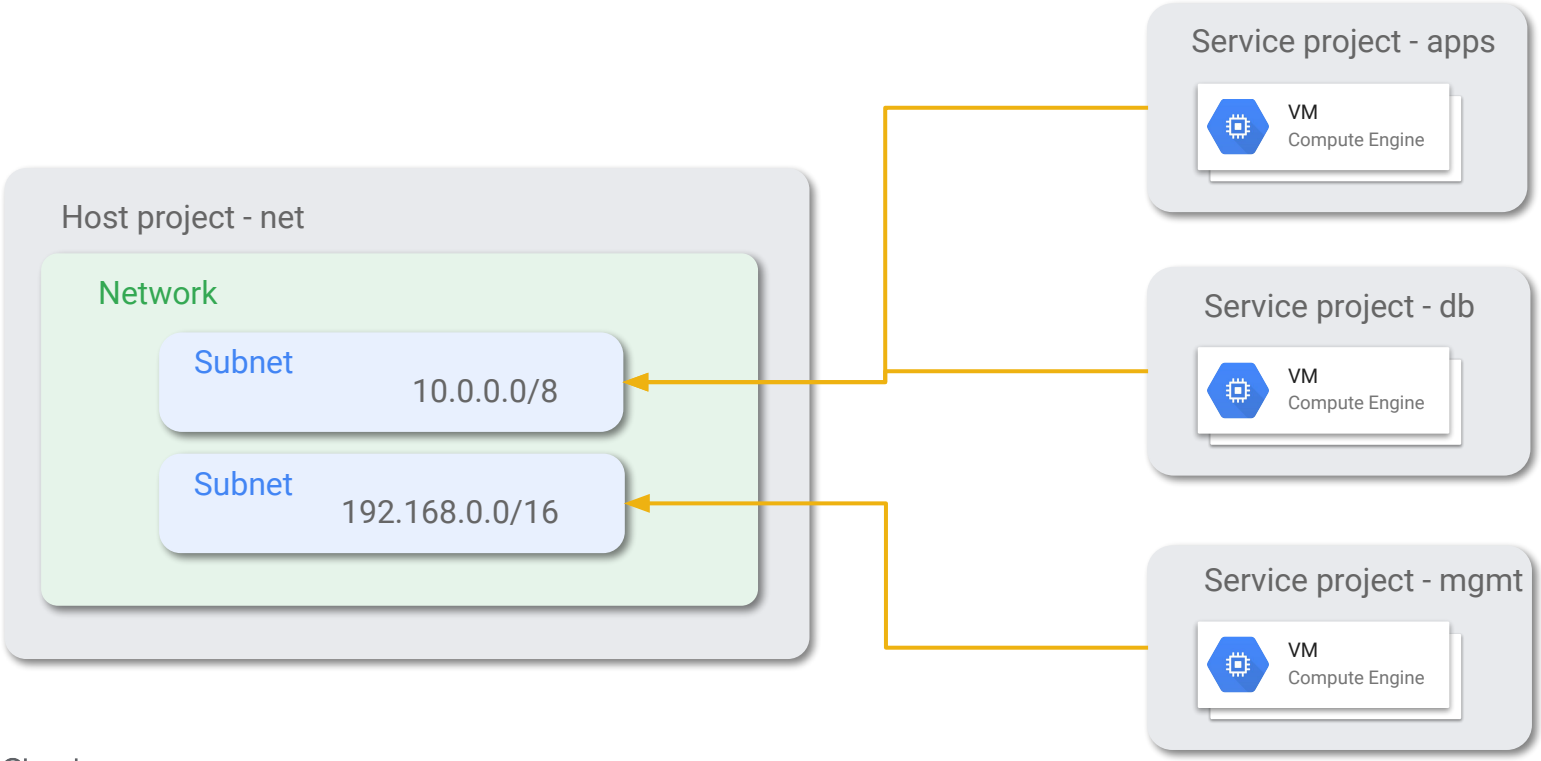
- Networks live within projects
- Networks may consist of multiple regions
- Regions may consist of multiple zones
- Subnets may span zones
- All elements of a network may talk to one another *(depending on firewall rules)*



Google Cloud

How to share a network across related projects?

Networking across projects is achieved through the use of Shared VPC that consist of a host project containing the network, and service projects using this shared network.

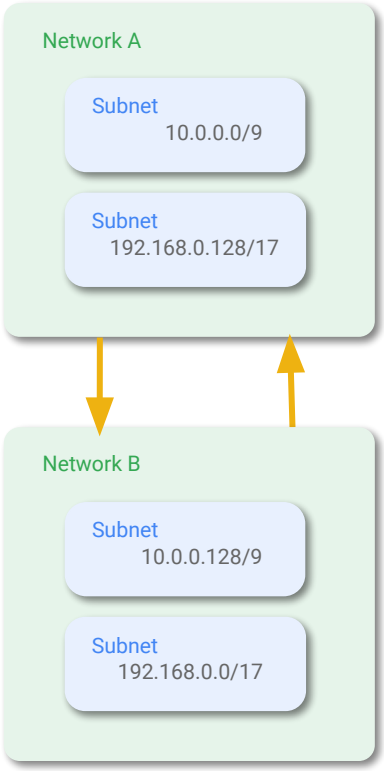


Google Cloud

How to connect two GCP networks?

Network Peering allows private RFC 1918 connectivity across two VPC networks regardless of whether or not they belong to the same project or the same organization.

- **Allows private RFC1918 connectivity across two virtual networks belonging to the same or different projects**
  - No public IP or VPN needed
  - Lower latency
  - Each project retains its own network resources (unlike VPC shared network)
  - No egress charge between public IPs
  - Internal services are not exposed to the public internet
- **Use for hosting services that need to be accessible by other virtual networks in different projects**



Google Cloud

What are some key concepts for IP addressing?

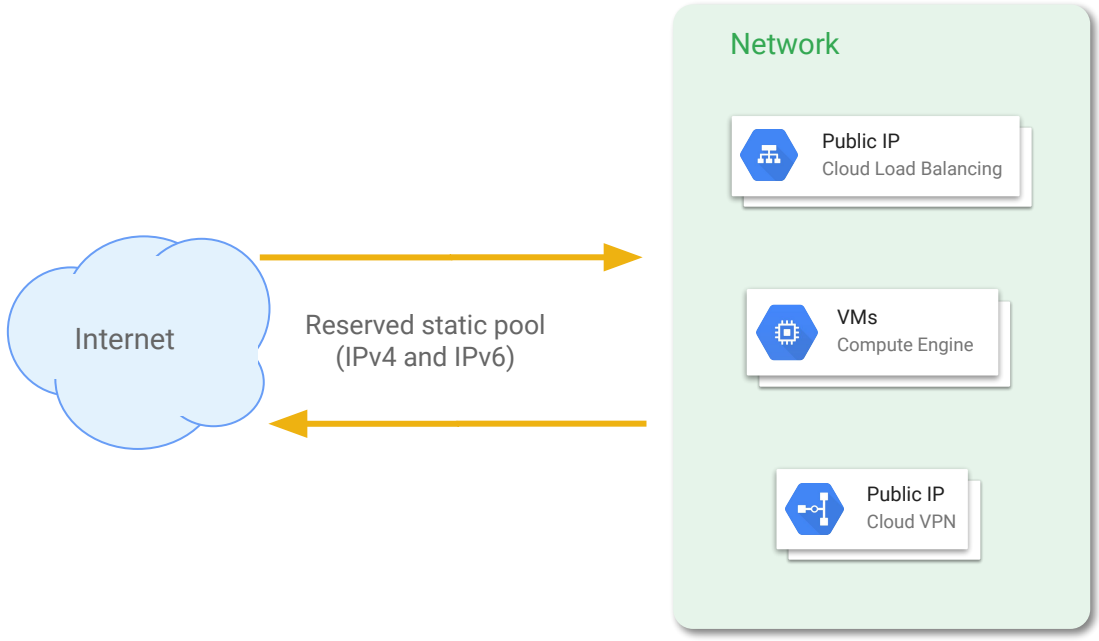
Some GCP resources can be assigned an IP address (static and/or ephemeral) s that enables it to communicate externally (internet) or internal GCP resources.

- All *VM instance resources* receive an internal RFC 1918 IP address (reserved or ephemeral)
  - Internal addresses are allocated from the CIDR range specified for its subnetwork
  - Google reserves 4 addresses from every CIDR range; the smallest allowable range is /29
- VM instance resources can be given an external IP address (reserved or ephemeral)
  - Reserved IP addresses are regional and will be used when the instance is started
  - Ephemeral IP addresses are assigned every time an instance is started
- *External load balancers* receive only public IP addresses (regional or global with anycast)
- *Internal load balancers* receive only internal IP addresses

What are the requirements for reserving external IPs?

The key considerations mentioned below should be kept in mind when requesting external IPs for GCP resources.

- GCP allows customers to reserve static Public IPv4 and IPv6 addresses
- Reservations are made on per-IP basis. Pools cannot be reserved, but IPs can be requested via API
- IPs that are reserved but not assigned are billed at an hourly rate (Currently, $0.010/hour)



Google Cloud

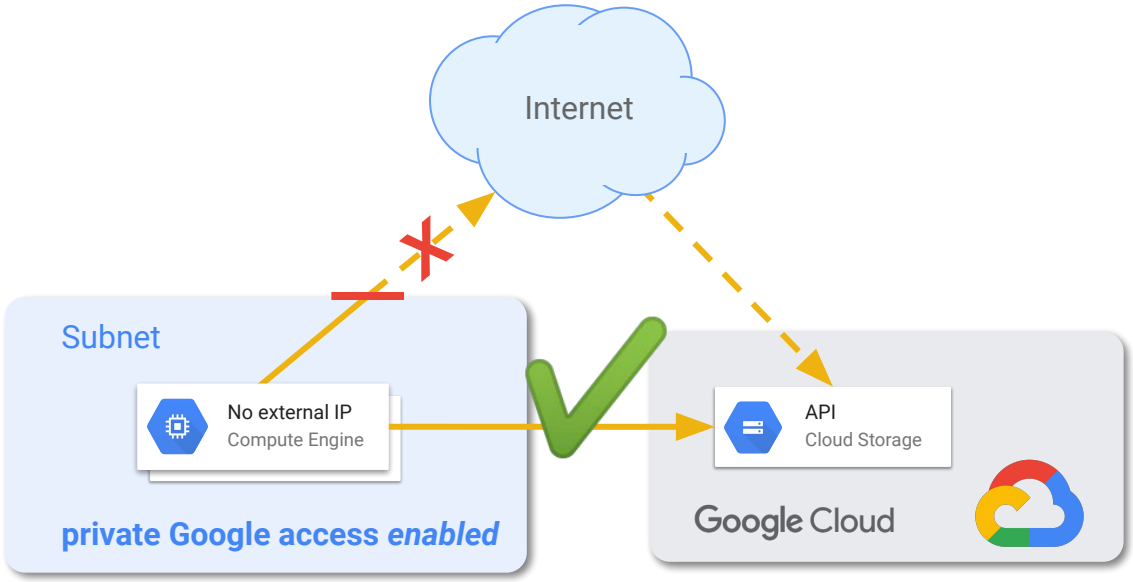
Why enable Private Google access on VMs?

Private Google access is an option available for each subnetwork. When it is enabled, instances in the subnetwork can communicate with public Google API endpoints even if the instances don't have external IP addresses.

Compute instances require public IP addresses to communicate directly with resources outside of their network

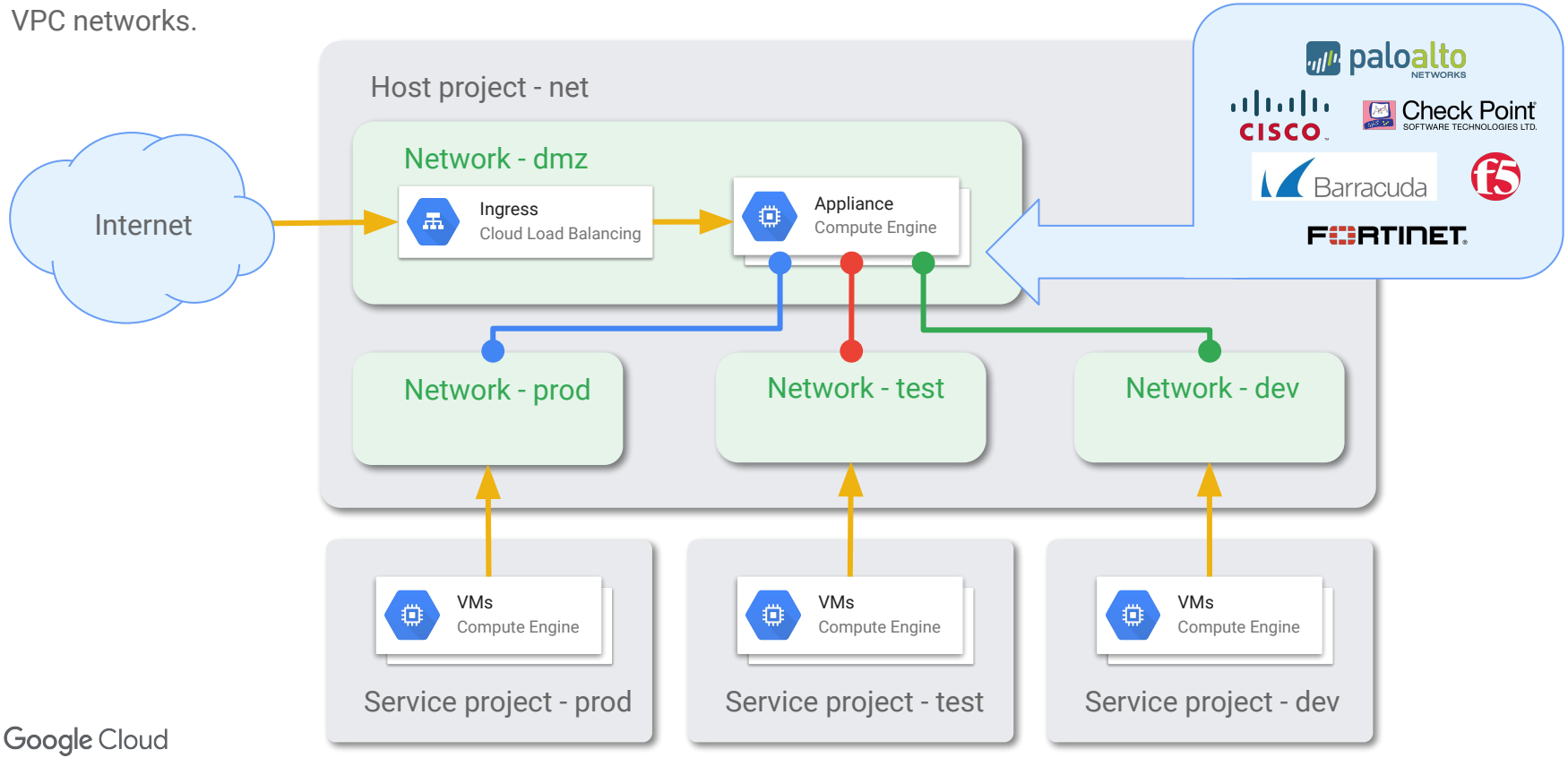
Problem:
Instances without public IP addresses can't access Google Cloud's public API endpoints

Solution:
Enable **private Google access** in the subnetwork the instance is attached to



Can a VM be configured with multiple VPCs?

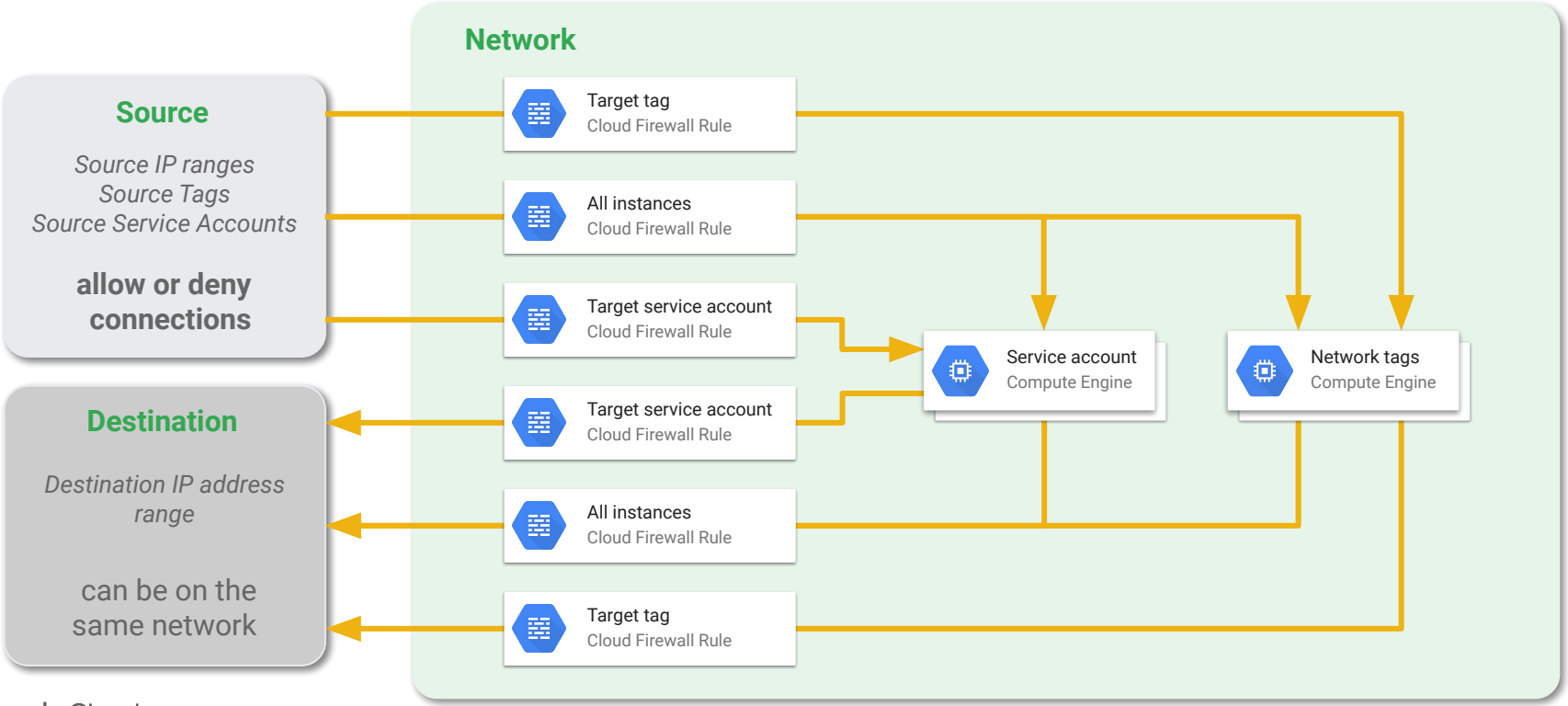
Multi-NIC appliances enable additional network interfaces to be attached to VMs, giving that instance access to different VPC networks.



Google Cloud

How to control network traffic?

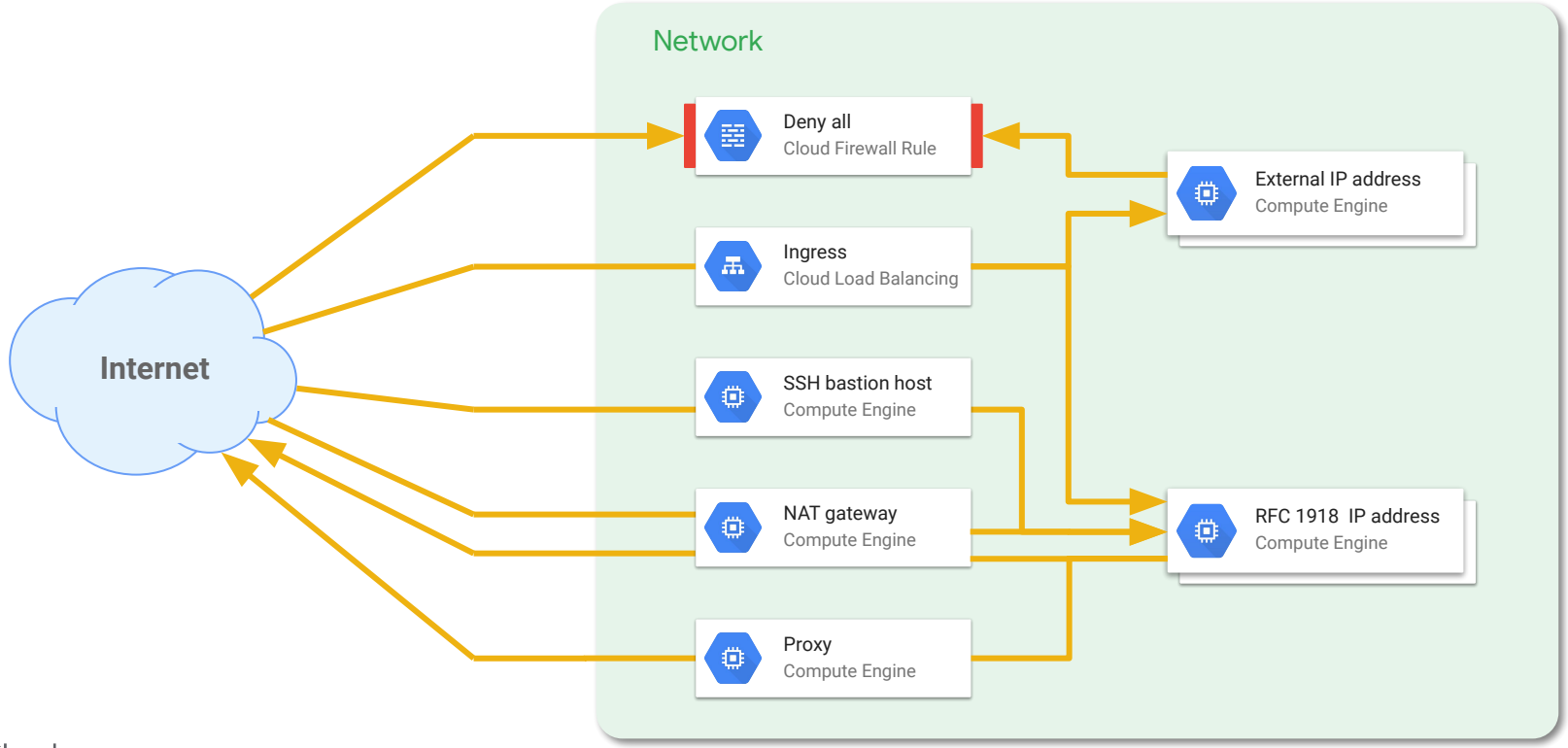
Firewall rules can be used to control ingress and egress traffic for GCP networks.



Google Cloud

How to connect to instances without external IPs?

Several methods can be used to control external access including a Bastion Host, a NAT instance or a Proxy instance.



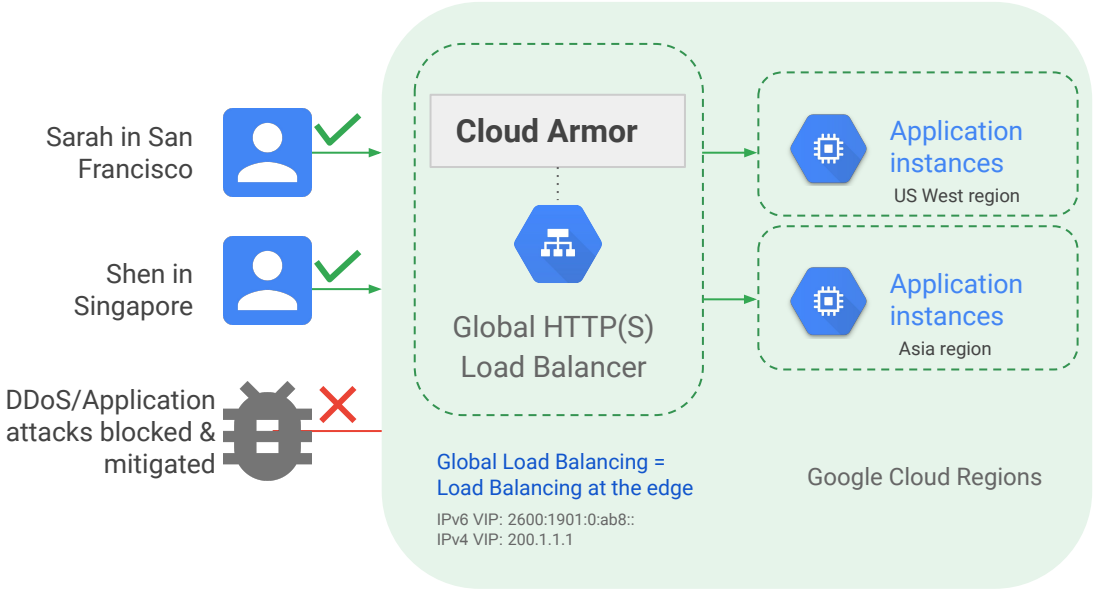
Google Cloud

What is GCP's solution for DDOS attacks?

Beta

Cloud Armor works with Global HTTP(S) Load Balancer to provide built-in defenses against Infrastructure DDoS Attacks.

- Mitigate infrastructure DDoS attacks
- Allow or block traffic with predefined and custom rules
- Defend against application level attacks
- Integrate with a rich ecosystem of security partners



Google Cloud

What are some other steps for preventing DDoS attacks?

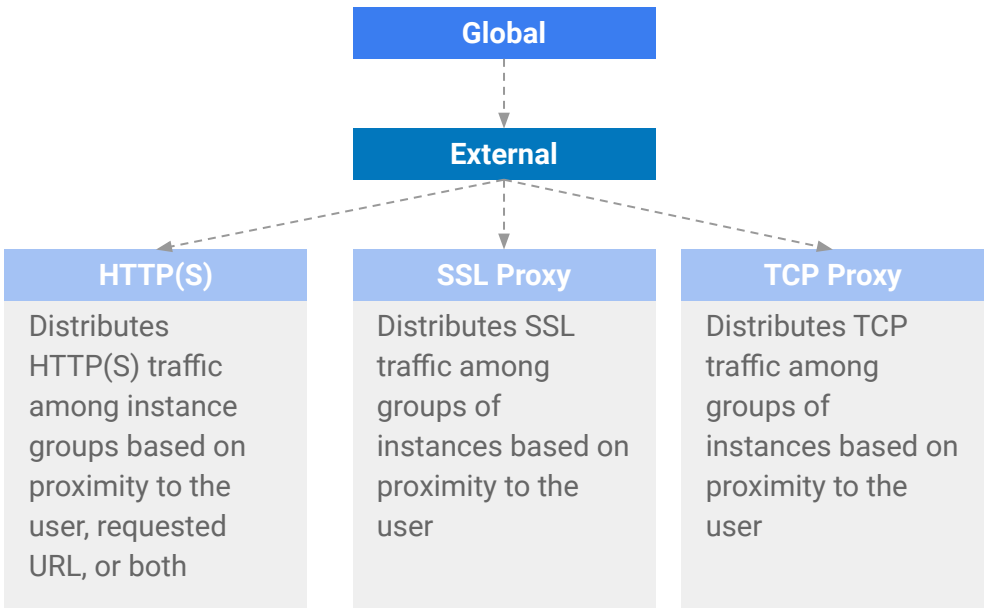
While Google has several measures in place to prevent DDoS attacks, mitigating these attacks is a shared responsibility between Google and the customer.

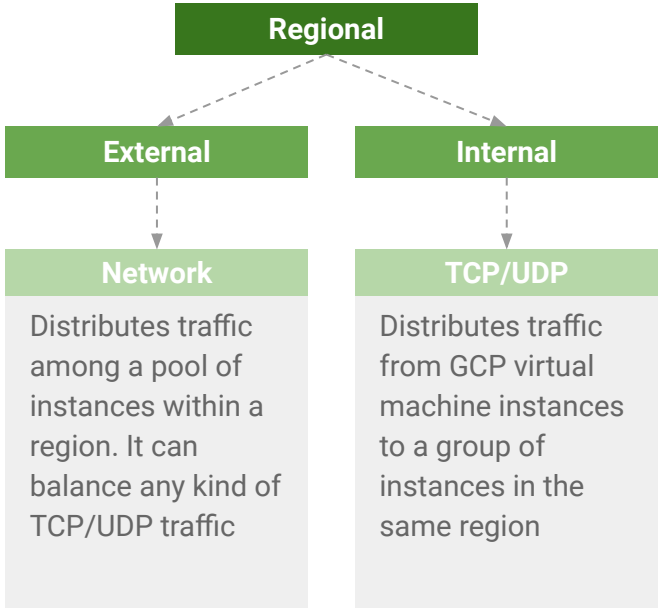
Attack surface	Reduce the attack surface on GCP by reducing externally facing resources
Internal traffic	Isolate internal traffic from the outside world
Load balancing	Use proxy-based load balancing to distribute load across resources
Scaling	Ensure that your apps scale well to handle the increased load
CDN Offloading	Offload static content to a CDN (such as Cloud CDN) to minimize impact
Third party	Deploy third-party DDoS protection if necessary
Rate limits and Quotas	Be aware of the role API rate limits and quotas play in protection against DDoS

Google

How is traffic distributed in GCP?

Software-defined load balancers can be used to distribute network traffic among instances. There are five load balancers available depending on the type of traffic that needs to be distributed.





Google

How to connect on-premise infrastructure with GCP?

GCP offers five connectivity options to enable a shorter connection between on premise infrastructure and GCP.

Public Internet/VPN

- Simple
- Resilient
- Use Google's existing edge network

Direct Peering

- Utilizes existing BGP route selection and internet routing
- Greater control of peering facilities
- Requirements for peering

Carrier Peering

- Offers benefits of peering when requirements cannot be met
- Service provider partners can provide SLA

Dedicated Interconnect

- Lower GCP egress cost
- Dedicated bandwidth and SLA
- Private space RFC-1918 addressing
- Control of circuit location

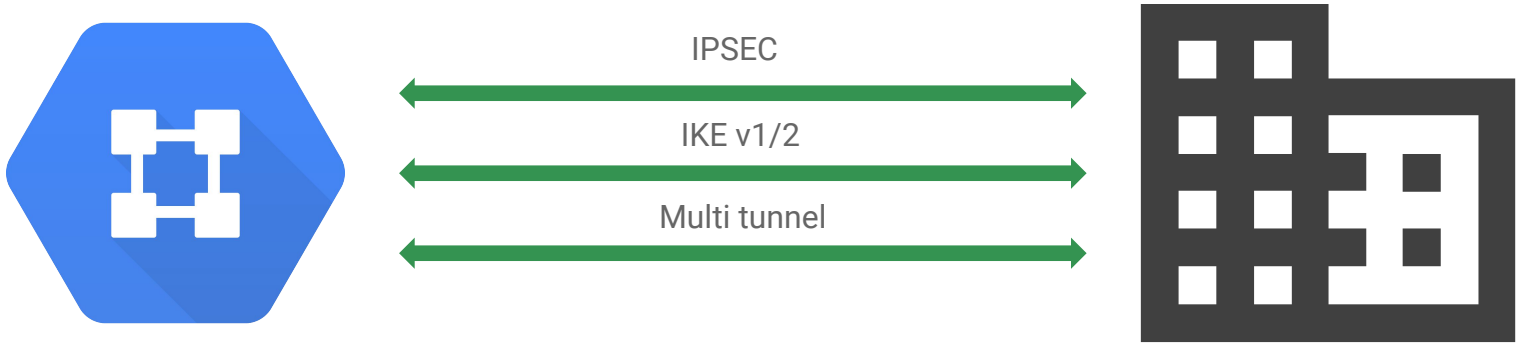
Partner Interconnect

- Work with a service provider when connectivity to Dedicated Interconnect cannot be met
- Private connection
- Provides Layers 2 and 3 connectivity

Google Cloud

How to enable a connection between private networks?

VPN tunnels can be used to connect private address spaces together, including on premise networks, other cloud platforms and other networks in GCP.



Google recommends reserving RFC 1918 IP space to avoid overlap.

Google

Cloud router

Dynamic routing

BGP-driven SDN tool to implement dynamic routes between GCP and other platforms

Dynamic BGP routes for easy traffic between on-prem networks and GCP

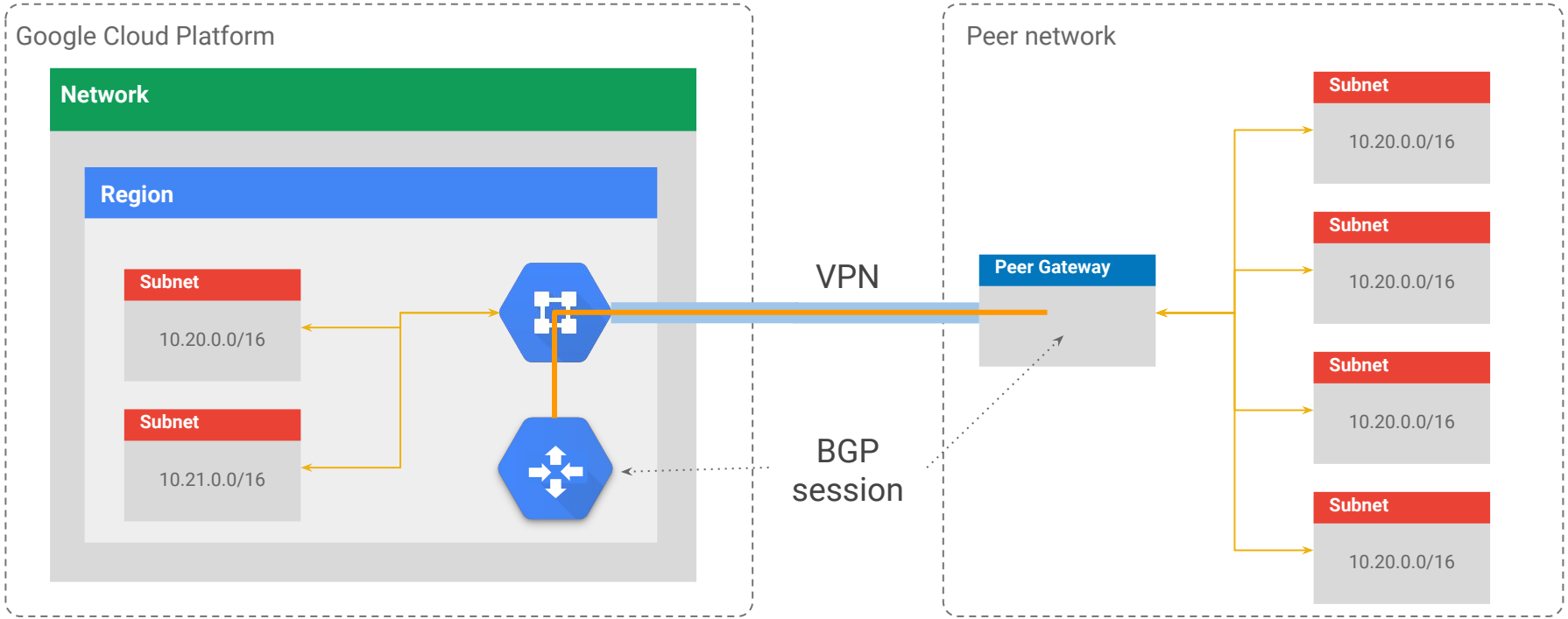
Delivered via Google's software-defined network virtualization

Eliminates need to restart VPN tunnels on topology changes

Google Cloud

How is dynamic routing enabled?

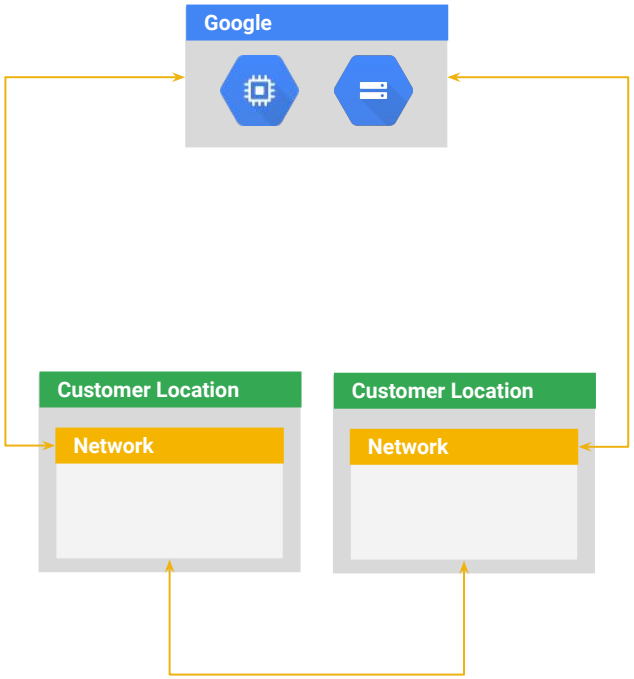
Adding Cloud Router to Cloud VPN deployments enables GCP and on-premise networks to automatically discover each other via BGP.



Google

What is Direct Peering?

Direct peering allows an organization to establish a direct connection between their infrastructure and Google's network using one of Google's global edge locations.



Requirements

Registered ASN from one to five regional internet registries

Announce public IPs /24 or larger to Google AS15169 using eBGP

Connect to Google in one of our private peering facilities or over an Internet exchange where AS15169 is present.

Have a completed ASN record at peeringdb.com

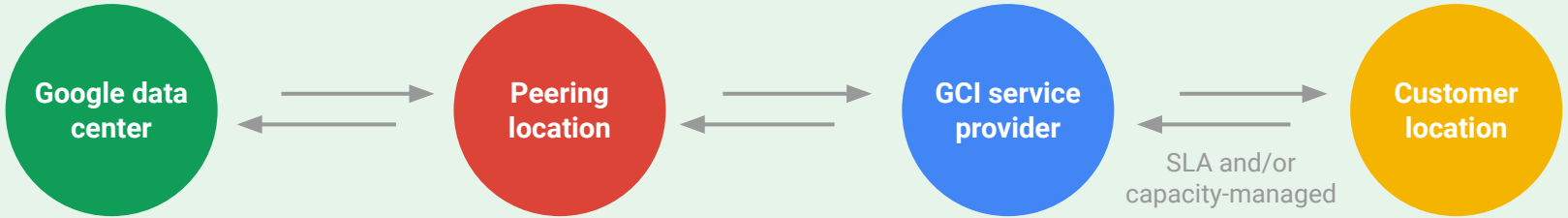
Provide a 24 x 7 Network Operations Center contact

Google

What is Carrier Peering?

Carrier Peering allows clients to take advantage of direct connection between the networks of their service provider and Google, providing higher availability and lower latency for business traffic as it travels from their systems to Google.

Carrier peering has fewer requirements than direct peering



Google Cloud

What is public versus “private” direct peering?

“Private” direct peering is dedicated network capacity only. It does not provide additional security.

Public	“Private”
• Uses Internet Exchange Point (IX or IXP). • Connect to multiple peers using one or more physical connections • IXP may require payment	• Uses public addresses — no additional security • Direct connection between two networks • Usually over dedicated fiber • Dedicated capacity • No fees except for cross-connects

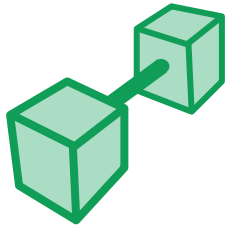
Google recommends multiple peering locations to establish redundancy in case the peering link goes down.

Google

How to establish a direct connection to Google?

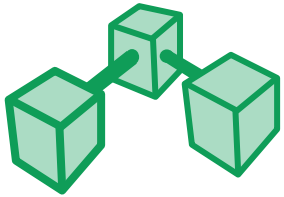
Cloud Interconnect provides low latency, highly available connections that enables clients to reliably transfer data between on-premises and VPC networks. There are two options available extending the on-prem environment.

- Provides access to private (e.g. RFC1918) network addresses
- Enables easy hybrid cloud deployment
- Does not require the use of and management of hardware VPN devices
- 99.99% and 99.9% SLA With Redundant Designs



Dedicated Interconnect

Connect N X 10G transport circuits for private connectivity to Google Cloud

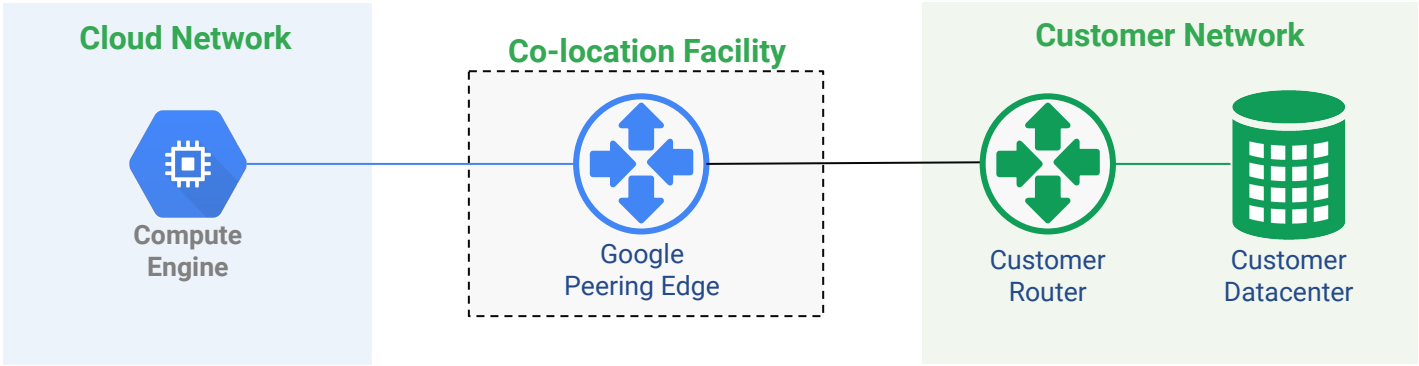


Partner Interconnect

Provide dedicated bandwidth (50Mbps - 10Gbps) to customers on a service providers network

What is Dedicated Interconnect?

Dedicated Interconnect provides direct physical connections between on-prem network and Google's network. This enables clients to transfer large amounts of data between network, instead of purchasing additional bandwidth over public Internet.

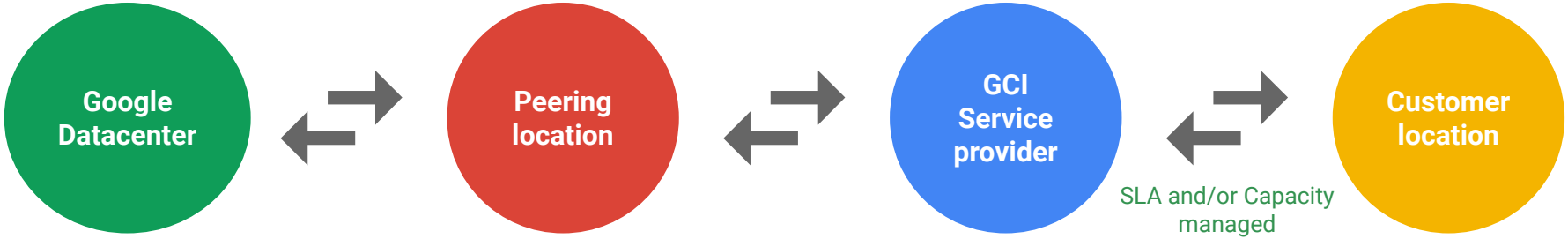


*Many organizations do not want public facing IPs as a security measure. Private Interconnect, aka **Dedicated Interconnect**, offers organizations a direct connection between their private RFC1918 space and Google's network. Private interconnect is also available through a Partner Carrier.*

Google

What is Partner Interconnect?

If the organization is unable to meet requirements for Direct Peering, organizations can chose Carrier Interconnect and connect to Google through a partner service provider.

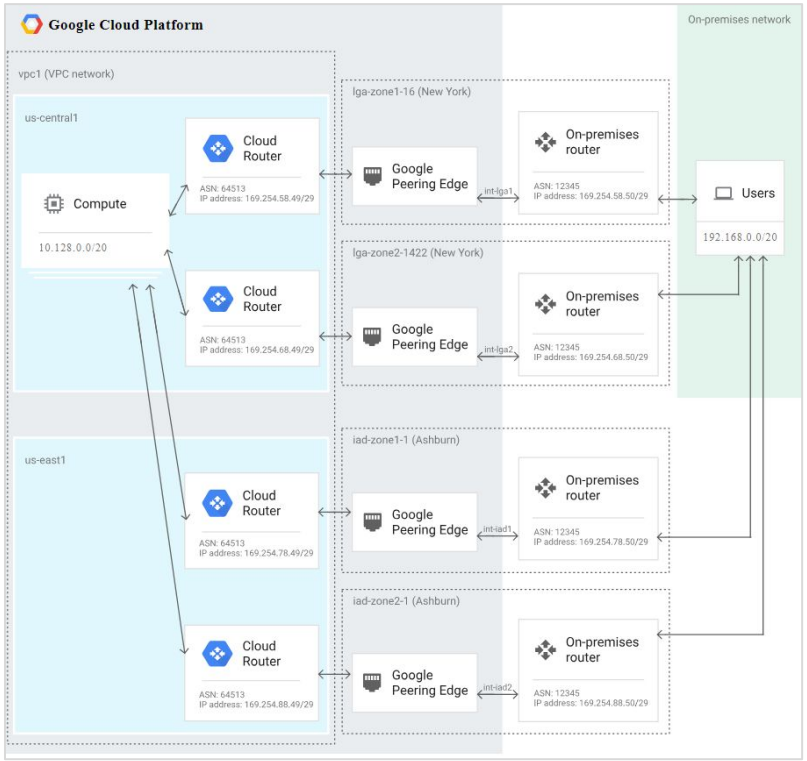


- *Google allows service providers to establish multiple redundant peering ports at selected POP locations.*
- *The connection to Google is facilitated and managed by a GCI Service Provider.*

How to design high availability Interconnect *(99.99% SLA)*?

Google recommends this configuration for production-level applications, such as mission-critical operations that have a low tolerance for downtime.

- Four interconnects, two interconnects in one metro (city) and two interconnects in another metro:
  - Interconnects that are in the same metro must be in different metropolitan availability zones
- Four Cloud Routers, two in each region:
  - Each Cloud Router must be attached to a different interconnect (four different VLAN attachments)
- The dynamic routing mode for the VPC network must be global



Google Cloud

How do the connectivity options compare?

Organizations must evaluate their requirements and determine which options best serves their needs.

GCP Service	Service Provider	Registered ASN	Public IP required?	Private IP (RFC1918)	Services	Google Charges Customers	SLA
Direct Peering	No	Required	Yes, /24 or less specific	requires Cloud VPN	All, but access to Compute Engine private networks requires Cloud VPN	None	None
Private Interconnect	No	No	not supported	Yes	Compute Engine	$2500 MRC	99.9%
Carrier Interconnect (Public)	Yes	Optional, ISP can provide	Optional, ISP can provide	Yes, but with NAT	All, but access to Compute Engine private networks requires Cloud VPN	None	None
Partner Interconnect (Private)	Yes	No	No	Yes	Compute Engine	MRC depending on data rate	Partner provided

Google

Internal DNS

TBD

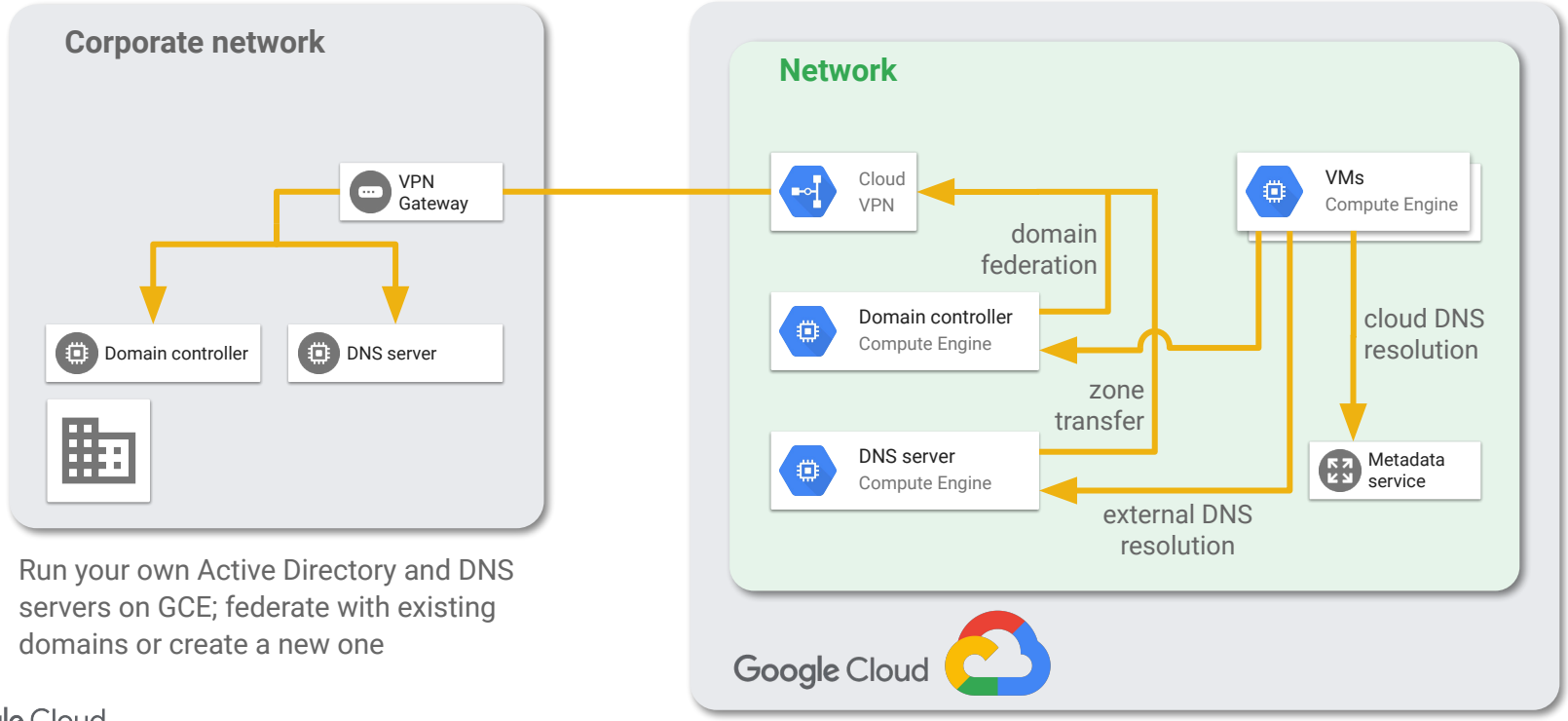
- All VM instance resources have access to internal metadata
  - Metadata includes resolving instance names to internal IP addresses

- Resolution for resources outside of the project requires an internal DNS
  - Google does not currently provide an internal DNS solution
    - An internal DNS system can be deployed in GCE
    - An on-premise internal DNS system can be accessed over VPN

Google Cloud

Directory and DNS integration

Google does not provide a hosted AD service but GCE does support running AD controllers inside VMs. So, a client can either create a GCE-only forest or federate an existing one into the platform over VPN.



Run your own Active Directory and DNS servers on GCE; federate with existing domains or create a new one

What are best practices related to networking?

Key considerations and recommendations related to networking in GCP.

1. Create multiple networks within projects to isolate groups of VM instances
2. Use subnets to control the address space in which VM instances are created
3. Achieve further isolation between subnetworks by using firewall rules and routes
4. Use iptables and routes to limit or filter egress traffic if necessary
5. Use Cloud Router for dynamic routing over Cloud VPN where possible
6. Use Cloud VPN or Private Interconnect to connect private address spaces together if necessary

Google

What are some key decisions to make?

When designing the networking target state, there are several key decisions an organization must make.

1. Public internet, direct peering, carrier interconnect or private interconnect?
2. What IP address space will be reserved for use in GCP?
3. What redundancy will be in place? Multi-zone or multi-region?
4. What internal DNS configuration will be used?
5. Will BGP via Cloud Router be implemented?

Google

How to access additional resources?

Learn more about networking through public documentation links, tutorials, labs and videos.

Documentation

- VPN Overview
- Shared VPC (XPN)
- Advanced VPC Concepts
- Load Balancing and Scaling
- Cloud Delivery Network
- Cloud Interconnect
- Cloud DNS

Tutorials

- Networking 101
- Networking 102
- Setup Network and HTTP Load Balancers
- Customize Network Topology with Subnetworks

Videos

- A cloud networking blueprint for securing your workloads
- Cloud networking solutions that support hybrid cloud workloads
- Secure, private environments in the cloud & on-prem with Virtual Clouds

Google

How can I test my understanding of Networking?

Work within your team to reinforce concepts by applying them to a real life use case. It is essential to consider the customer's business and technical requirements when designing the solution.

Objective

- Provide a simple, scalable and secure VPC network design to meet iRobocop requirements.

Problem Statement

- While evaluating different public cloud providers and subsequent POC's with limited number of applications, some key issues were observed:
  1. Application developers built projects with subnet CIDR ranges that overlap with other applications in different project due to lack of network visibility, causing failure of network traffic.
  2. Application developers implemented firewall rules that don't meet iRobocop objective of least privilege access.
  3. In some cases, the environment was open to the internet due to misconfiguration thereby exposing to security vulnerabilities.
  4. In order to provide communication across projects hosting different applications, network team had to configure and maintain multiple VPN tunnels causing increase in complexity of design and support overhead.

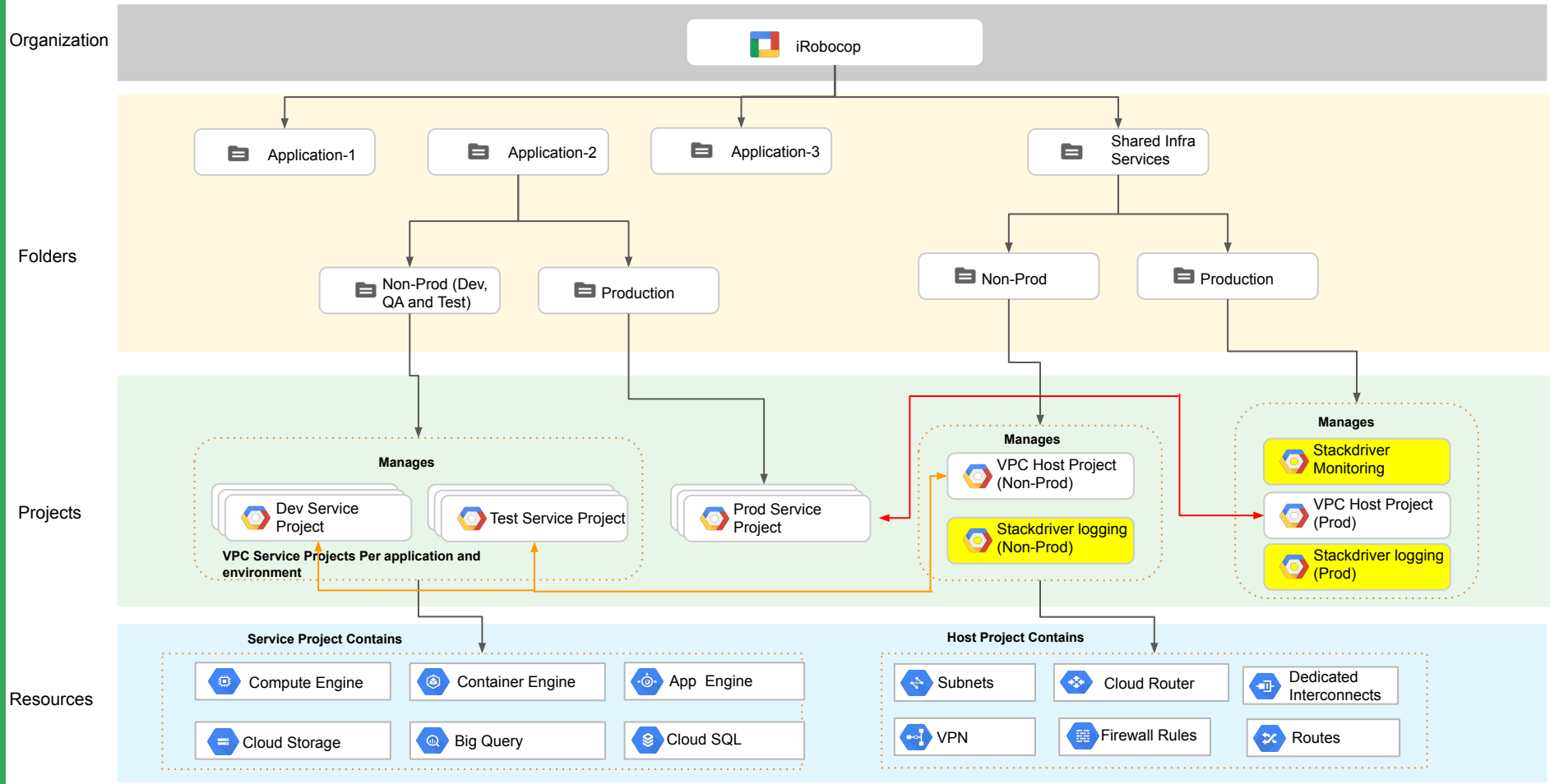
Google

How can I test my understanding of Networking?

Work within your team to reinforce concepts by applying them to a real life use case. It is essential to consider the customer's business and technical requirements when designing the solution.

Solution:

1. Create a new folder called "shared services" to provide centralized networking for all iRobocop applications running in GCP
2. Create subfolders for production and non-production under shared services folder to further segregate based on environment and support personnel
3. Leverage Shared VPC architecture for each environment to segregate between production and non-production traffic. Implementation of Shared VPC architecture provides network team with a centralized administrative and access control of configuring IP subnets and their allocation to application teams.
4. With Shared VPC approach, security team has centralized control of configuring, maintaining and auditing firewall rules from a single host project
5. Shared VPC networks also allow different projects to access shared network resources such as a VPN or Cloud Interconnect connection on the host project
6. Configuration of appropriate firewall rules in the host project can allow applications hosted in different service projects to communicate if necessary without need of a VPN connection
7. This is a very simple network architecture that can get exponentially complicated as we introduce additional networking components e.g. load balancers (Google, 3rd party), on-prem connectivity, firewall rules (egress, ingress), restricting traffic (NAT Gateway)



Google

Monitoring and Logging

Monitoring and Logging

This section will focus on the following key topics.

Objectives

- Key decisions for the set-up of logging and monitoring for GCP resources based on the organization's requirements.
- Key decisions for the configuration of logging and monitoring for GCP resources
- Considerations around the organization's requirements

Key Learnings

- Understanding GCP best practices for monitoring and logging
- Identifying commonly tracked GCP metrics
- Monitoring and logging via Stackdriver
- Implementing archiving policy via GCP

Google Cloud

How are resource activities monitored and logged?

Resource monitoring conversations will identify performance tracking metrics and logging processes.

How are resources, performance and applications monitored?

What 3rd party tools are being used to monitor resources?

What are some benchmarks and metrics being tracked?

Are there any regulatory retention policy for logs?

Google Cloud

What do we need to know?

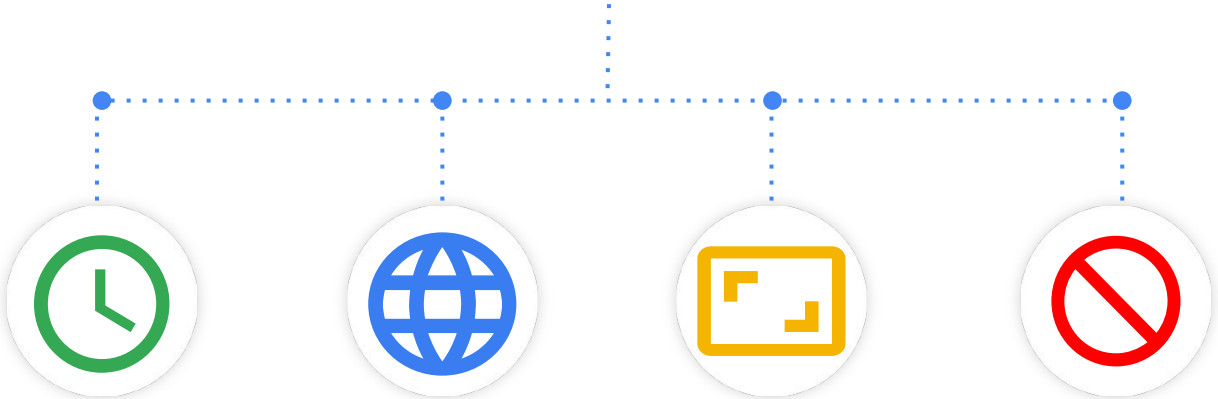
When evaluating a client's monitoring and logging requirement, there are key aspects you need to know in order to set up monitoring and logging in GCP.

- Stackdriver monitoring
- Monitoring Resources
- Logging in GCP
- Third party tools for monitoring and logging
- Stackdriver Trace
- Stackdriver Debugger
- Auditing requirements
- Compliance requirements

Google Cloud

What are the four golden signals? *(1 of 2)*

The four parameters below are common metrics considered whilst evaluating performance of your GCP solution.



Latency

The time it takes to **service a request** - response time, including queue/wait time (ms).

Traffic

A measure of **how much demand** is being placed on your system, measured in a high-level, system-specific metric.

Saturation

How **overloaded** something is - similar to utilization, but more directly measured by queue depth, concurrency, etc.

Errors

The **error rate**, in errors / second for applications, micro-services, infrastructure resources and cloud services.

Google Cloud
Google

What are the four golden signals? *(2 of 2)*

The parameters below have the following business and technical impacts.

Latency

- Key contributor to **user experience**
- Indicator of **emerging issues**
- Indicator of growing **capacity demands**

Traffic

- Direct indicator of **site activity**
- Historical trends used for **forecasting**
- Capacity decisions based on the **demand lifecycle**
- Monitor **cloud infrastructure spend**

Saturation

- Indicator of reaching **maximum capacity**
- Analysis helps inform **autoscaling decisions**
- Predict **performance degradation** when demand exceeds capacity

Errors

- Application errors show **something is wrong**
- Infrastructure errors can indicate **misconfiguration** or **capacity issues**
- Services errors can indicate **permissions errors** and **SLO violations**

Google Cloud

What are some application server sample metrics?

In this case example, consider how each metric impacts business and technical performance.

Latency

- # requests waiting for a thread
- Thread execution duration (s)
- Average garbage collection duration over time (s)

Traffic

- # of currently active requests
- # requests processed per second
- # of currently active sessions

Saturation

- % memory utilization
- % thread pool utilization
- % cache utilization
- Garbage collection frequency

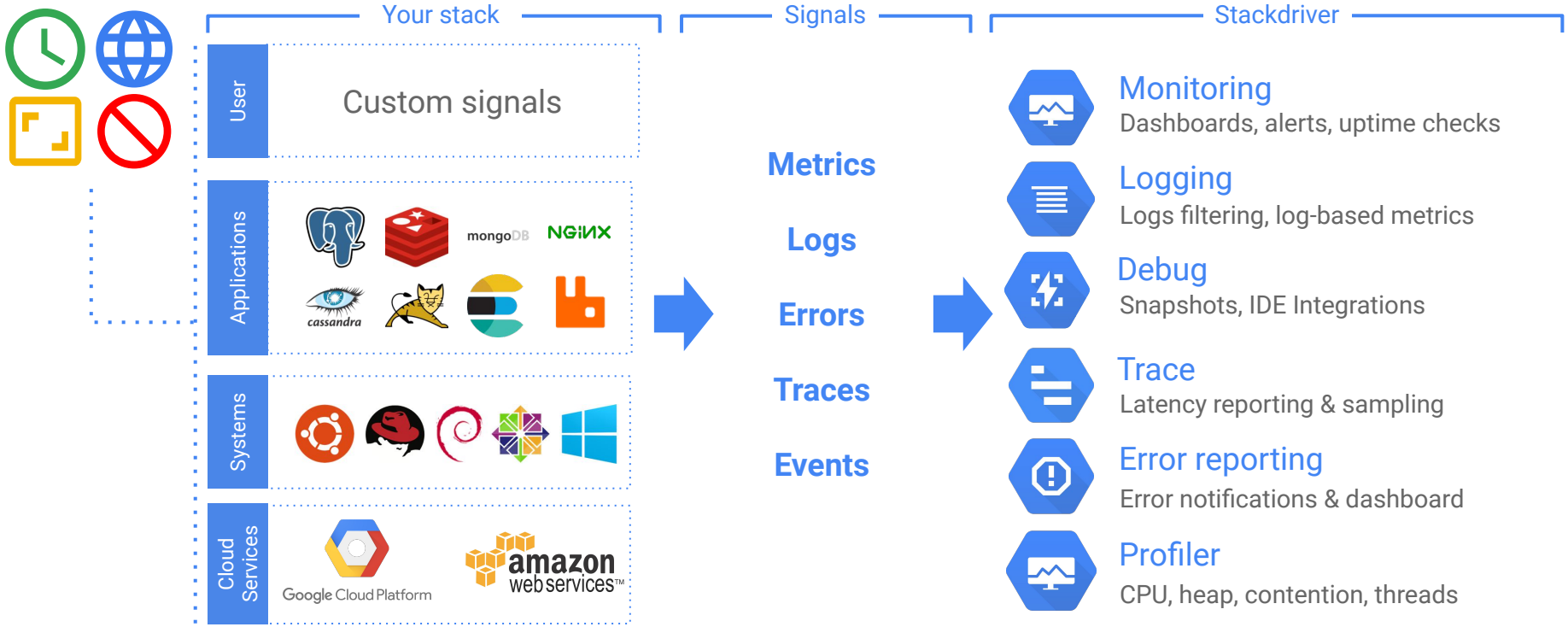
Errors

- # of 5xx class errors
- # of 4xx class responses
- # of stack traces
- # of OutOfMemory exceptions
- # of unclean peer connection resets

Google Cloud

What does Google offer?

Google provides a menu of monitoring and logging tools, as follows.



Google Cloud

What is Stackdriver?

Google Stackdriver provides powerful resource and performance monitoring, logging, and diagnostics for your AWS and GCP resources.

Monitoring

Endpoint checks to internet-facing services

Uptime checks for URLs, groups, or resources

Plugins for many major stacks (Apache, MySQL, CouchDB etc.)

Logging

Filter, search, and view

Define metrics, dashboards, and alerts

Export to BigQuery, Google Cloud Storage, and Pub/Sub

Trace

Distributed trace system collecting latency data

Captures trace from all compute resources

Find performance bottlenecks

Fast, automatic issue detection

Debugger

Debug in Production without impacting the application or users

Multiple sources option

Sharing a debugging session as easy as sending a URL

Audit and Compliance

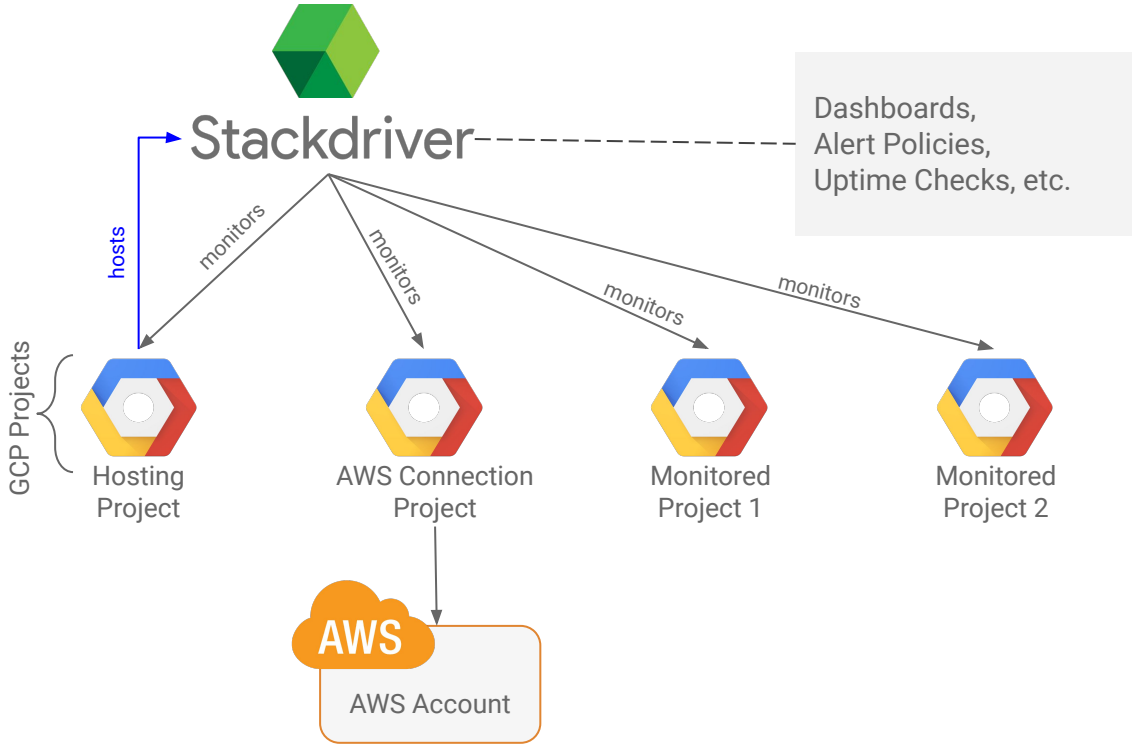
Audit logs that capture all the admin and access events

Google

How is monitoring enabled?

A Stackdriver account, stored in a separate GCP project, holds the monitoring configuration for resources being monitored. GCP Projects and AWS accounts must be associated with a Stackdriver account.

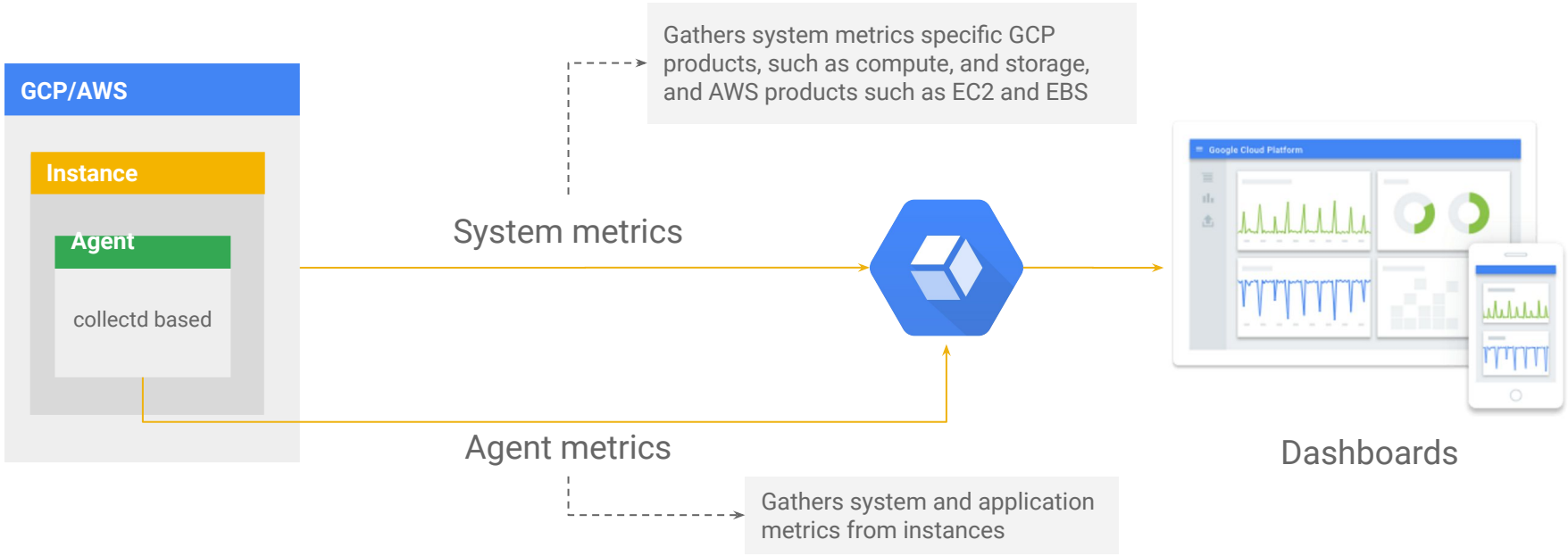
Stackdriver is hosted in a GCP Project, the hosting project, which holds the monitoring configuration for the Stackdriver account.



Google Cloud

How are resources monitored?

Identified metrics can be be tracked by Stackdriver, which offers wide-ranging monitoring of GCP resources to generate awareness and visibility across the platform.



Google recommends one StackDriver “account” per environment: dev, test, demo, prod, etc.

Google Cloud

How is application performance monitored?

Application performance may be monitored using the following GCP Stackdriver products.

Debugger

Debugger provides **snapshots**, **breakpoints** and **dynamic log statements**.

It can remove **multiple deployment steps** from root cause analysis iteration cycles.

Profiler

Profiler operates at extreme scale and isn't limited to analyzing single transactions.

It can detect **slow performance** in your application and helps reduce **costs**.

Trace

Trace provides **distributed tracing capabilities** to analyze your applications.

It can help find **performance bottlenecks** in your application

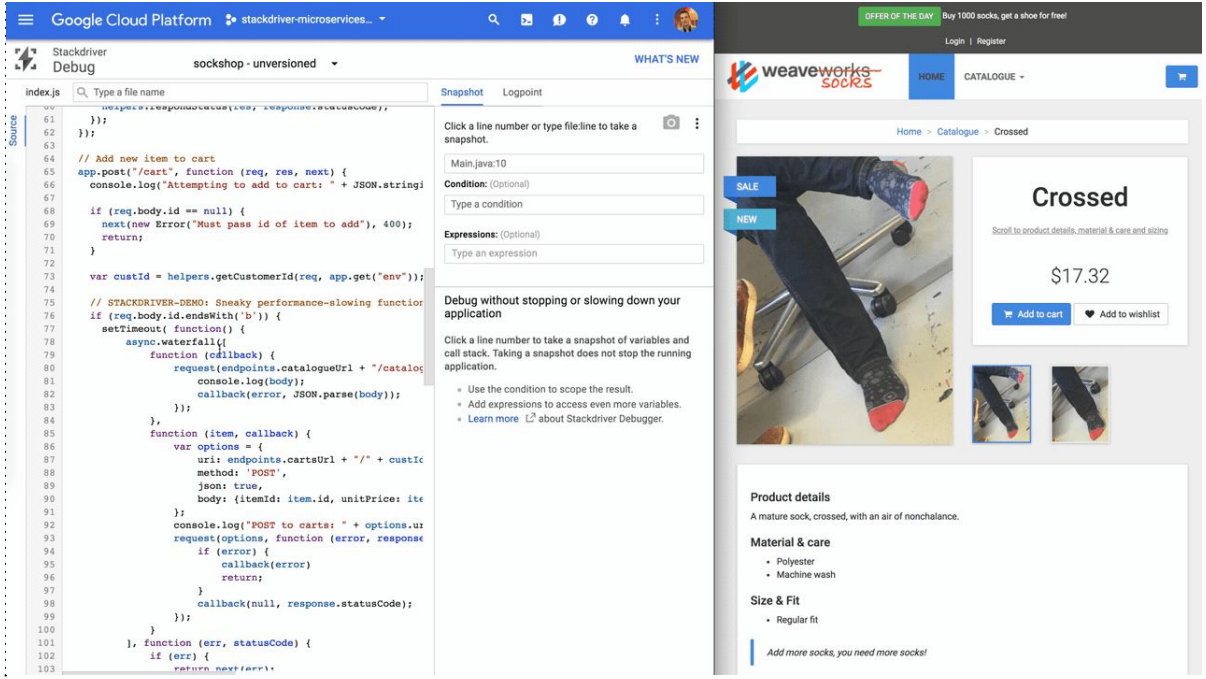
Google Cloud

How does the Stackdriver Debugger work?

Stackdriver Debugger debugs applications whilst running in production without stopping or slowing them down.

Real-Time Production Debugging

- Real-time app debugging
- Debug snapshots, breakpoints, conditional debugging
- Integrations with popular IDEs
- Multiple version control sources (GitHub, Google Cloud Source, Bitbucket, Gitlab)



Google Cloud

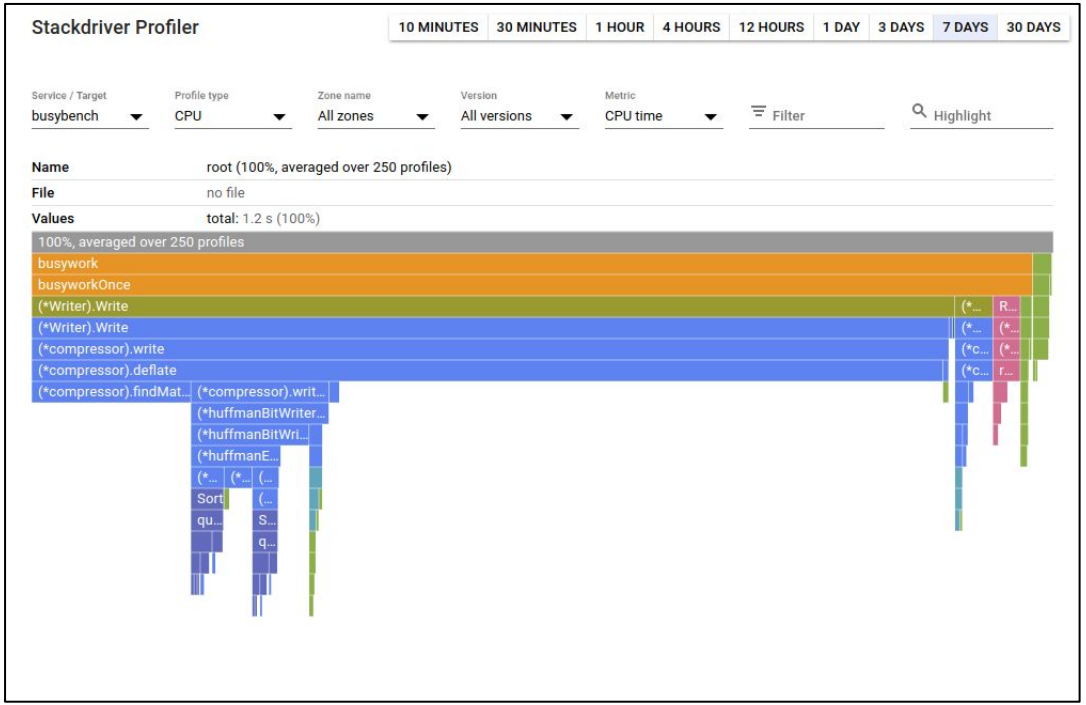
How does the Stackdriver Profiler work?

Beta

Stackdriver Profiler monitors CPU to identify latency and inefficiency, improve application bottlenecks, and reduce resource consumption.

Continuous CPU and heap profiling

- Low-impact **production profiling**
- Broad **platform support** (VMs, GAE, GCE, GKE)
- Support for **Java**, **Go**, **Python**, **NodeJS**
- Understand your applications' **call patterns**
- Improve **performance** and reduce **costs**



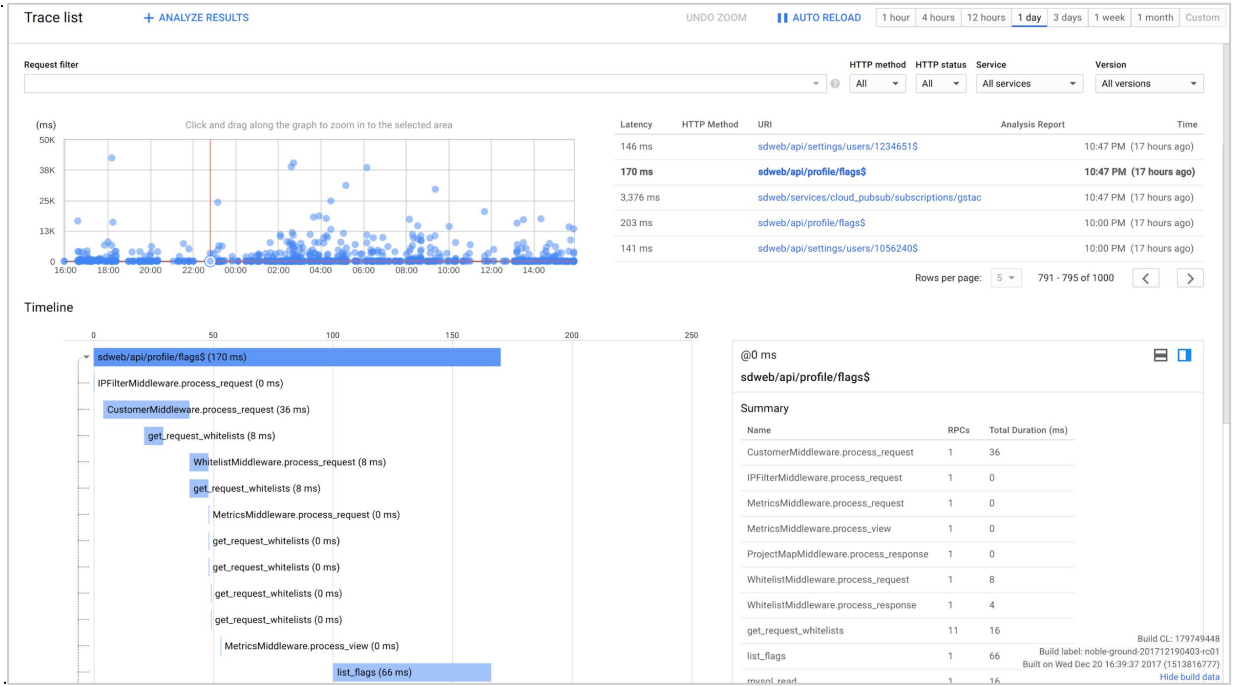
Google Cloud

How does the Stackdriver Trace work?

Stackdriver Trace inspects detailed latency information to find bottlenecks and quickly identify root causes.

Distributed tracing system to collect latency data from applications

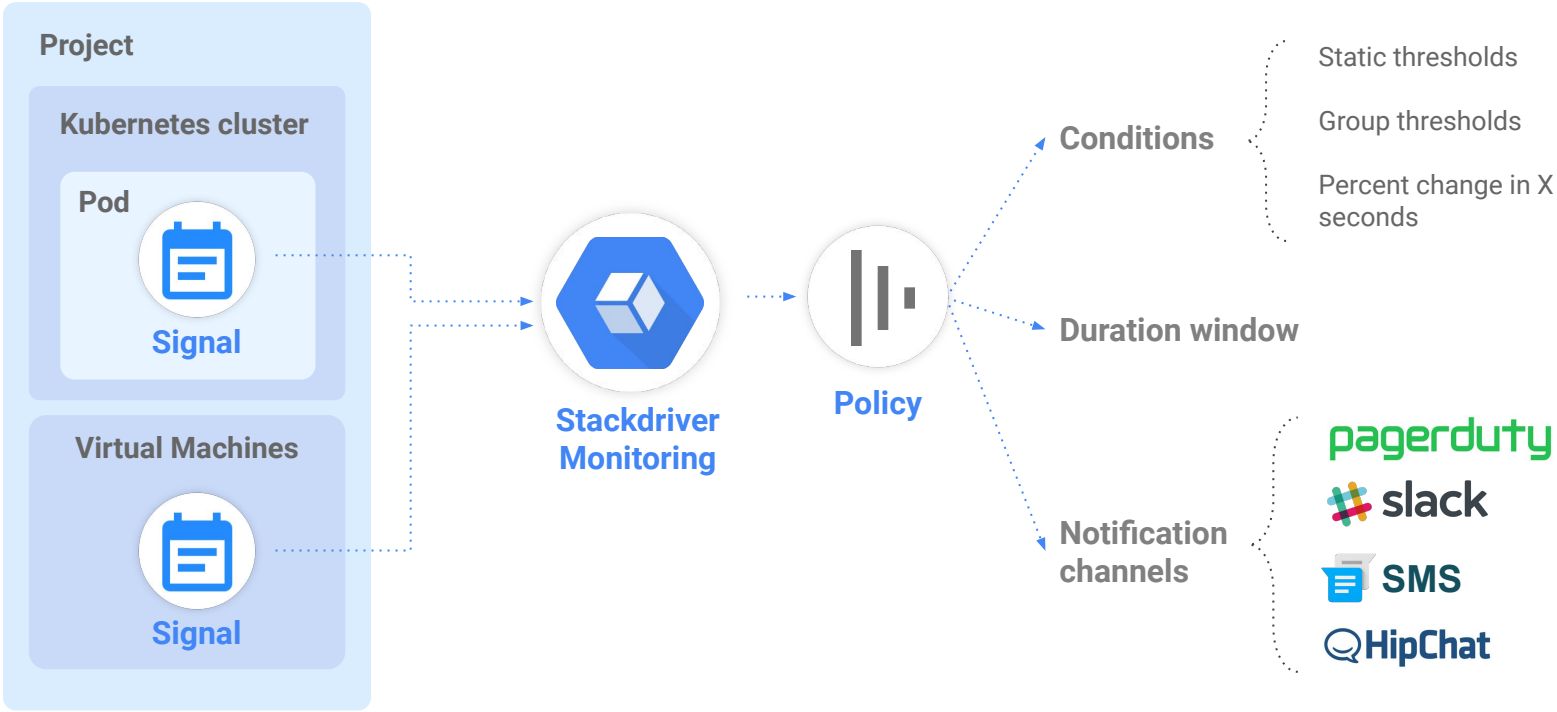
- Integrated with **App Engine**
- Support for **Compute Engine** and **Kubernetes Engine**
- Near real-time **performance insights**
- In-depth **latency reports**
- Find **performance bottlenecks** in your apps
- Automatic **issue detection** and **alerting**



Google Cloud

How does the Stackdriver alerting process work?

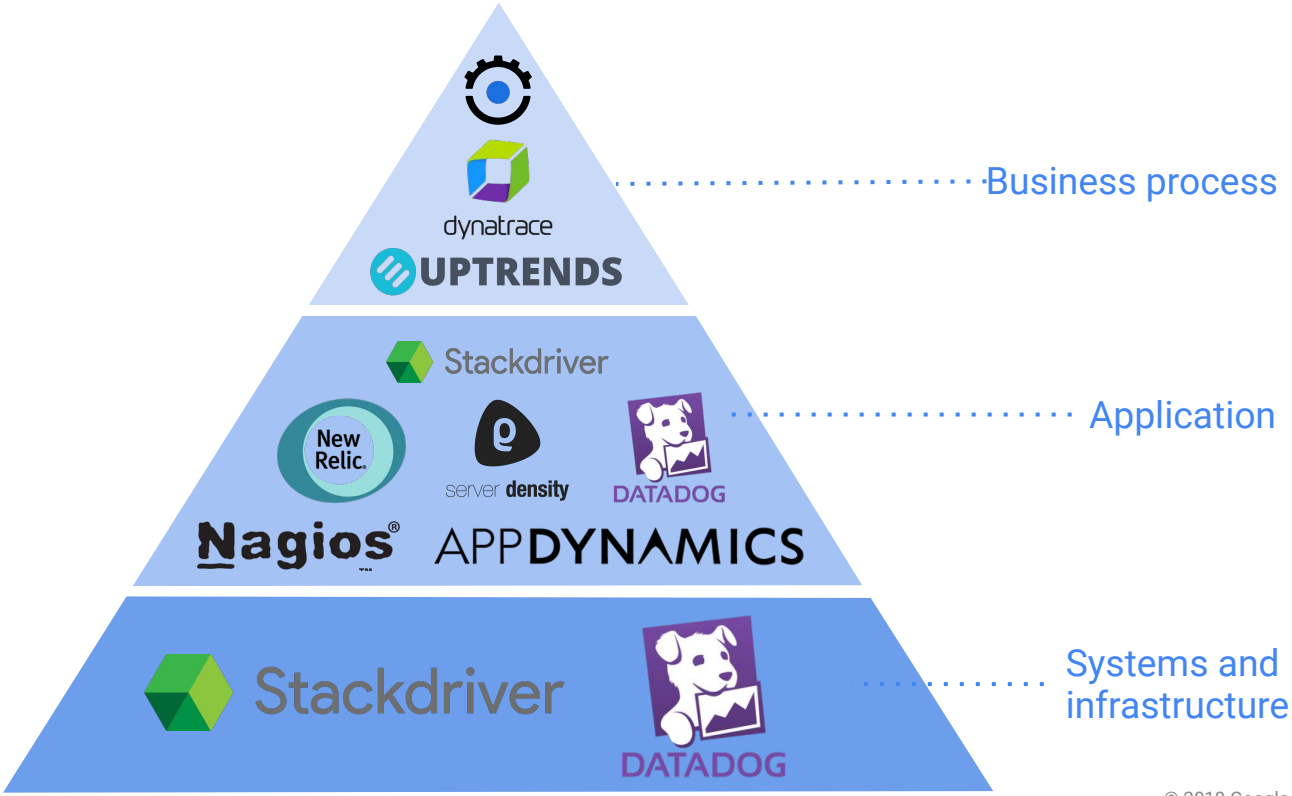
The below diagram illustrates the alerting process using Stackdriver.



Google Cloud

What other tools can be used for monitoring and logging?

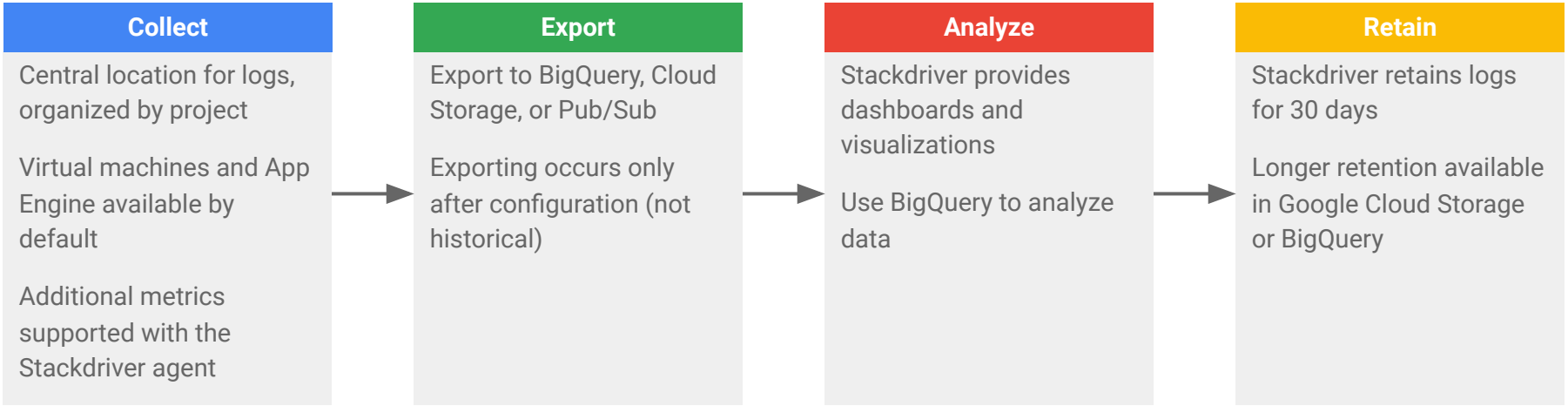
Third-party tools can also be leveraged in conjunction with Stackdriver to meet additional monitoring and logging needs.



Google Cloud

How is logging handled in GCP?

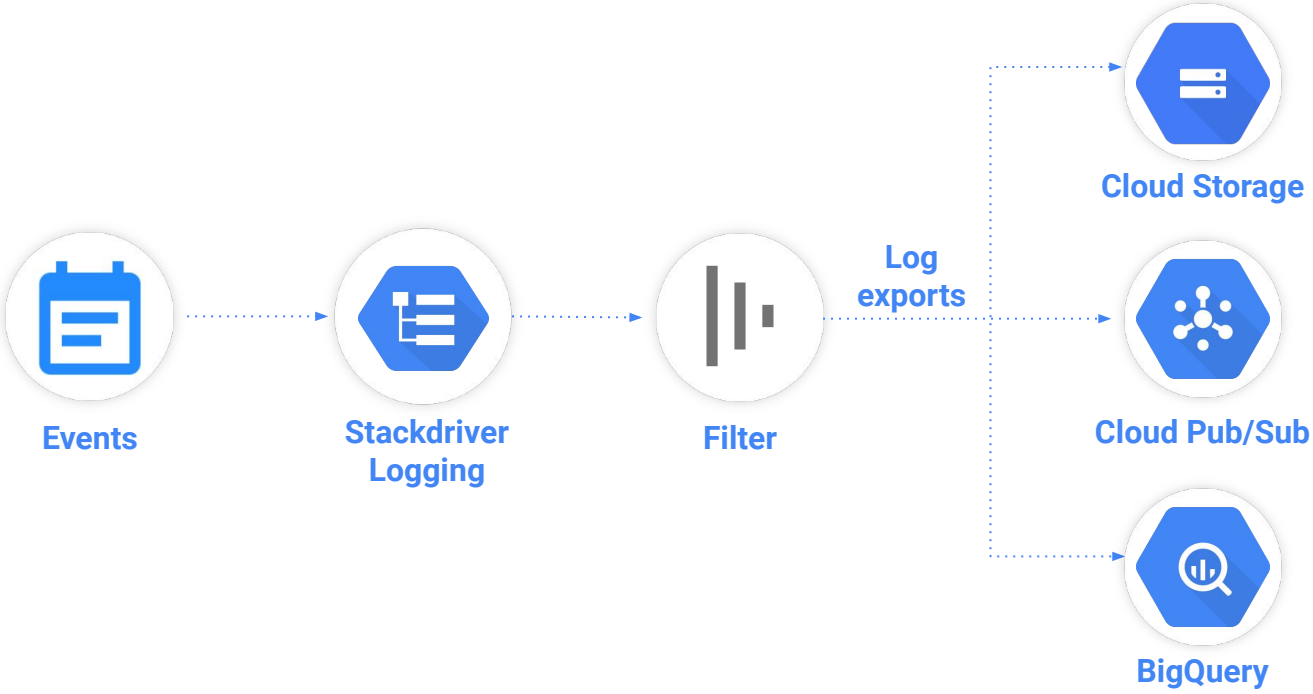
Stackdriver Logging provides functionality to store, search, analyze, and alert on log data to determine who did what, where and when on GCP.



Google Cloud

How are logs exported?

Log exports may be published to the following three outputs, via the process diagrammed below.



Preferred way to **archive logs** for **long-term analysis** and **auditing**.

Provides **cheap storage** and **integrations** with most Google Cloud Platform services.

Preferred way to **integrate** with other logging tools.

It is also popular for **real-time log streaming analysis**, where performance matters.

Preferred way to analyze **very large datasets**.

It is also a good option when your logs are coming from **many sources**, like in **IoT** applications.

Google

What are key considerations related to compliance?

Adhering to the following best practices will help organizations remain compliant while running their applications in GCP.

Separation of duties (SoD)

- Copy logs to a project with different ownership than the source
- Logs can identify SoD violations

Least privilege

- Restrict the usage of owner role for projects and log-buckets
- Use the editor role, which can deploy applications and modify / configure resources

Non-repudiation

- Cloud Storage automatically encrypts all data before it is written to disk
- Additional fortification can be implemented by object-versioning log buckets

Google

How does Stackdriver enable audits?

Stackdriver can be used to access audit logs that capture all the admin and data access events within GCP, allowing organizations to answer the question of “who did what, where and when?” in GCP.

Users

Reports section of Admin Console

Reports API to generate custom reports

Login audit log in Admin Console

Admins

Admin SDK audit logs

Configure Admin Alerts for suspicious logins, admin privilege grants, etc.

Permissions

IAM API used to audit permission changes

Stackdriver logging shows events

Service accounts

Cloud logs contain access logs

Google Cloud

What are common audit logging use cases?

Audit logging may encompass the three common use cases, as noted below.

Admin Activity

- Record API calls modifying **configuration** or **metadata**
- Default retention is **300 days**
- Used for **auditing** and **forensic analysis**
- Available at **no charge**

EXAMPLE RECORD

```
Object: /buckets/XYZ
Action: CREATE OBJECT
Actor : devops-service-account
```

Always enabled

Data access

- Record API calls that create, modify, or read **user-provided data**
- Default retention is **30 days**

EXAMPLE RECORD

```
Object: /buckets/XYZ
Action: READ OBJECT
Actor : employee@my-org.com
```

Needs to be enabled

Access transparency

- Oversight of Google **access to your resources**
- Data protection **control**
- Access **justification**
- Resource **identification**

EXAMPLE RECORD

```
Object: /buckets/XYZ
Action: READ
Reason: Ticket #12345
```

Needs to be enabled

Google

How can I design an audit logging strategy?

In designing an audit logging strategy, consider the following parameters.

Drivers

- IAM policy changes
- Sensitive data access
- Security configuration changes
- Ad-hoc retrospective and forensic analysis
- Regulatory compliance reporting obligations

Best practices

- Auditable events should egress origin network ASAP
- Aim to segregate production and audit infrastructure from each other
- Store and forward to be resilient to network partitions

Tiers of audit logs

- Each distinct Google Cloud service maintains its own audit logs
- Audit logs are maintained for both data and administrative operations
- Audit logs are updated gradually, not necessarily reflected in real-time

Google Cloud

What are best practices for monitoring and logging?

Adhering to recommended best practices for logging processes and performance monitoring enables long-term success of a GCP project.

1. Use Cloud Logging as a centralized location for logs and export to BigQuery for analysis
2. Monitor access of system resources, accounts and domain
3. Monitor administrative actions (i.e. operations performed in the domain admin console)
4. Prevent unwanted changes to logs through the principles of least privilege, non-repudiation and separation of duties
5. Use Stackdriver Monitoring to monitor resources and provide alerts
6. Export logs to BigQuery for analysis and long term storage

Google Cloud

What are some key decisions to make?

When designing the networking target state, there are several key decisions an organization must make.

1. What are the “must-have” infrastructure metrics to be collected for your most important applications?
2. Will you prefer your cloud-based logging to remain separate from other logging, or would you prefer a consolidated log?
3. Is there any trending or periodic reporting you require?
4. What are your retention requirements for metrics, for the OS, and for application log file entries? Will an external platform be leveraged?
5. What monitoring and visualization tools will be used?
6. Who will have access to logs?
7. Are there requirements to build strict compliance rules for the workloads in scope?

Google

How to access additional resources?

Learn more about monitoring and logging through public documentation links, tutorials, and videos.

Documentation

- Stackdriver Monitoring
- Stackdriver Logging
- Stackdriver Error Reporting
- Stackdriver Debugger
- Stackdriver Trace
- Stackdriver Metrics List

Tutorials

- Monitoring Cloud Infrastructure with Stackdriver
- Using Stackdriver's monitoring and logging to get better visibility into your application's health

Videos

- Metrics that matter
- Alerting best practices - the thin line between informing and over-informing
- Stackdriver: monitor, diagnose, and fix

Google Cloud

How can I test my understanding of Monitoring and Logging?

Work within your team to reinforce concepts by applying them to a real life use case. It is essential to consider the customer's business and technical requirements when designing the solution.

Objective

- Help iRobocop team implement Stackdriver Logging and Monitoring best practices.

Problem Statement

- iRobocop hosts many 3rd party applications like Redis, Cassendra and Kafka. They would like to monitor and collect logs from these applications in addition to default monitoring and logs. For compliance reasons they would like to store all logs for future analysis.
- iRobocop need help in collecting logs and setting up alerts for critical applications. They also would like to log only relevant information in stackdriver. Additionally they would like to reduce the overall cost of logging.

Google Cloud
Google

How can I test my understanding of Monitoring and Logging?

Work within your team to reinforce concepts by applying them to a real life use case. It is essential to consider the customer's business and technical requirements when designing the solution.

Solution:

1. Create pubsub for exporting stackdriver logs directly to GCS bucket. Setup bucket policies based on compliance requirements to move log entries from nearline to coldline.
    - Note: Refer to this page.
2. Create appropriate exclusion filters to exclude entries that are not required.
3. Install and configure stackdriver monitoring agent on VMs running Redis, Cassendra and Kafka

Automated Operations

Automated Operations

This section will focus on the following key topics.

Objectives

- Review configuration and resource management
- Discuss continuous integration and continuous delivery as part of the software development lifecycle

Key Learnings

- Leveraging CI/CD release process in software development
- Reviewing common tools and models for automation
- Understanding configuration management in GCP
- Managing resources using Deployment Manager

Google Cloud

What are operational challenges faced by organizations?

Organizations face several IT operational challenges that stem from being able to to scale rapidly and release frequently.

Increasing Demand

Requires rapid scaling of IT infrastructure

Operational Bottlenecks

Large Ops teams need to overcome organizational and technical bottlenecks

Disconnected Feedback Loops

Communication gap between software and IT teams

Manual Errors

Increased scale leads to greater human errors

Google

How are pipelines used in CI/CD?

Pipelines are a critical infrastructure setup to ensure rapid and consistent deployment of code

Build

The codebase is checked out to a specific version and artifacts (e.g., Docker containers) are built.

Deploy

Artifacts are deployed into an environment.

Test

Tests (unit, integration, vulnerability, performance, etc.) are performed to ensure application quality.

Approve

User determines whether a pipeline should proceed to the next stage.

A **pipeline** is a process that takes a version of an application codebase and performs all steps necessary to release it to production. Pipelines can be triggered manually, or they can be triggered automatically when changes are pushed to the codebase. They are generally composed of separate stages, most commonly these four.

Google Cloud

What is an effective release lifecycle?

To address IT operational challenges, a more effective release lifecycle can be implemented to enable developers and operations teams to automate software delivery and infrastructure changes.

Standard release lifecycle

A more effective release lifecycle

Continuous Integration

Continuous Delivery

Configuration Management

What is Continuous Integration (CI)?

CI allows multiple developers quickly integrate their code into the production code.

Purpose

Continuous integration facilitates the automated integration, testing, and building of software and infrastructure code in preparation for deployment to a staging or production environment

Benefits

Quick incremental feedback to the developers with automated testing

Reduce the number of bugs put into production with dependable builds

Reduce manual testing required to get code production ready

Popular Tools

Code repositories

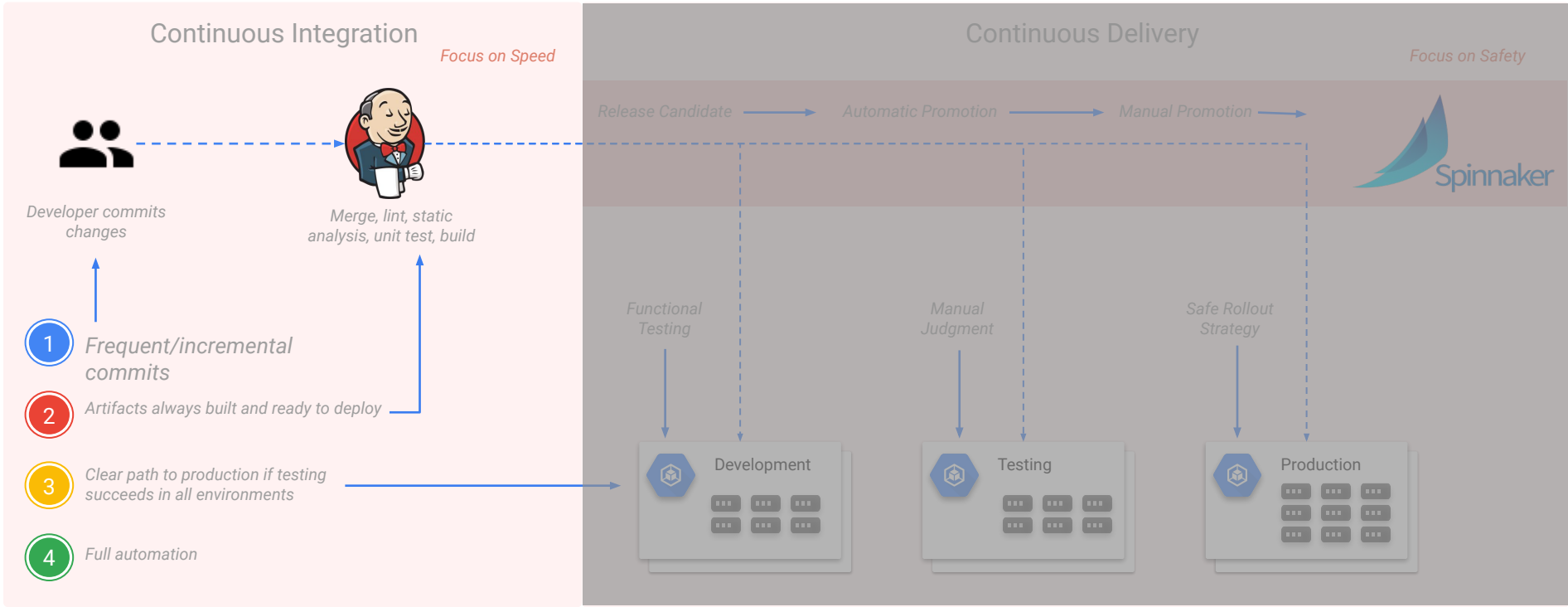
Github, Bitbucket, Cloud Source Repository

CI pipeline tools

Jenkins, Bamboo, Travis CI

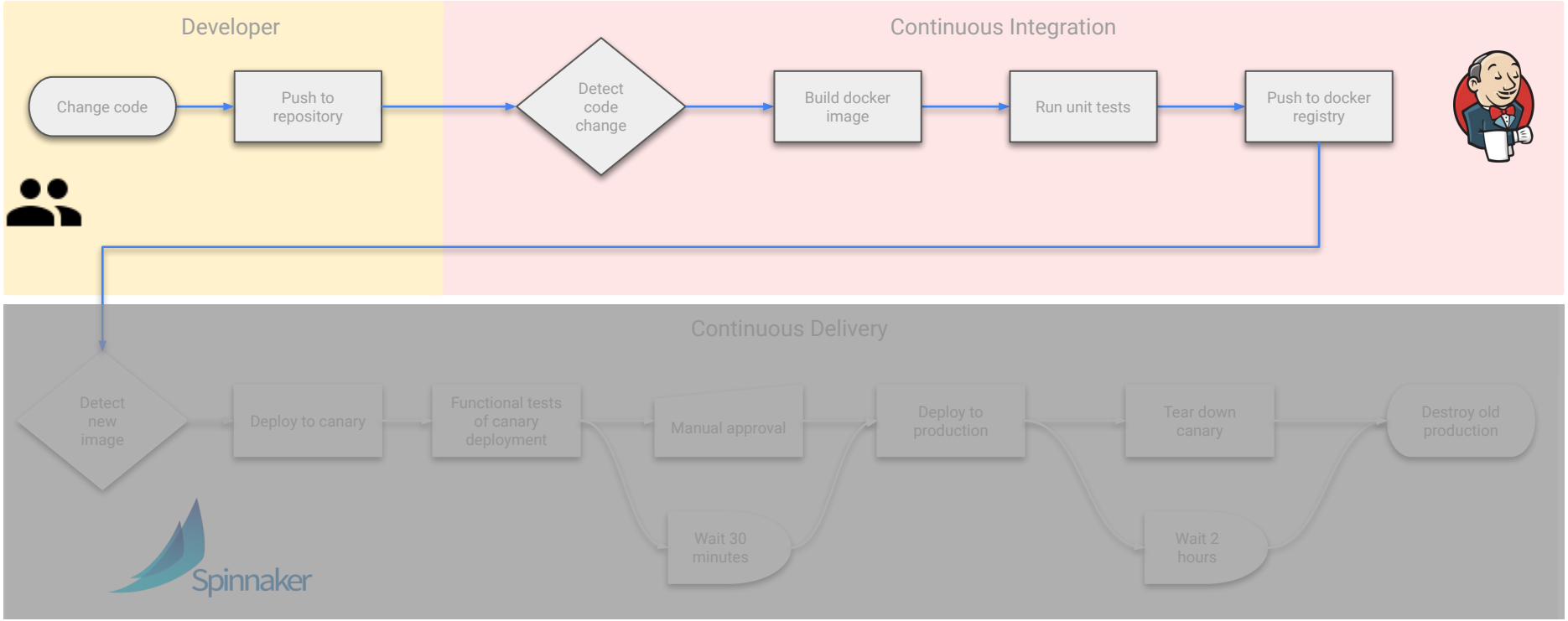
What is the Continuous Integration (CI) process?

CI allows developers to simultaneously push their code to the master repository to have Jenkins integrate it together before it gets tested



What is the CI Pipeline Process Flow?

Once the code has been merged a Docker image will be built to test the code before pushing the code to the package step. Docker image will be deleted once test succeeds or fails



Google

How can Jenkins Software help Continuous Integration?

Jenkins is the most popular open source continuous integration tool although there are many others that do similar jobs.

- Jenkins is an open source automation server written in Java.
- Jenkins helps to automate the non-human part of the software development process, with continuous integration and facilitating technical aspects of continuous delivery.
- It supports version control tools, including AccuRev, CVS, Subversion, Git, Mercurial, Perforce, ClearCase and RTC
- Can execute Apache Ant, Apache Maven and sbt based projects as well as arbitrary shell scripts and Windows batch commands.

What are some Jenkins strengths and weaknesses?

Strengths and weaknesses to consider when evaluating Jenkins for continuous integration

Strengths

- Mature product, extensive community support
- Simple architecture
- Entirely pipeline-as-code (Jenkinsfile), UI is just a view
- Deep level of control in pipelines (groovy-based)
- Pipeline libraries for sharing and reuse
- Slaves created as pods during job run, offering efficient resource usage

Weaknesses

- Lacks integration with Kubernetes objects, often supplemented with tools like Helm
- HA options not natively supported

What is Continuous Delivery?

Continuous delivery packages the code once tests have been completed and releases to desired environment

Purpose

Continuous delivery provides the ability to do **frequent and automated promotion of code** through the deployment pipeline with **quality gates** at each stage

Benefits

Test and deliver software at high velocity with more **frequent releases**

Kick off a release process with **minimal effort**

Roll back quickly if issues appear in production

Tools & Methodologies

Tools

Spinnaker, GoCD, ConcourseCI, Codeship

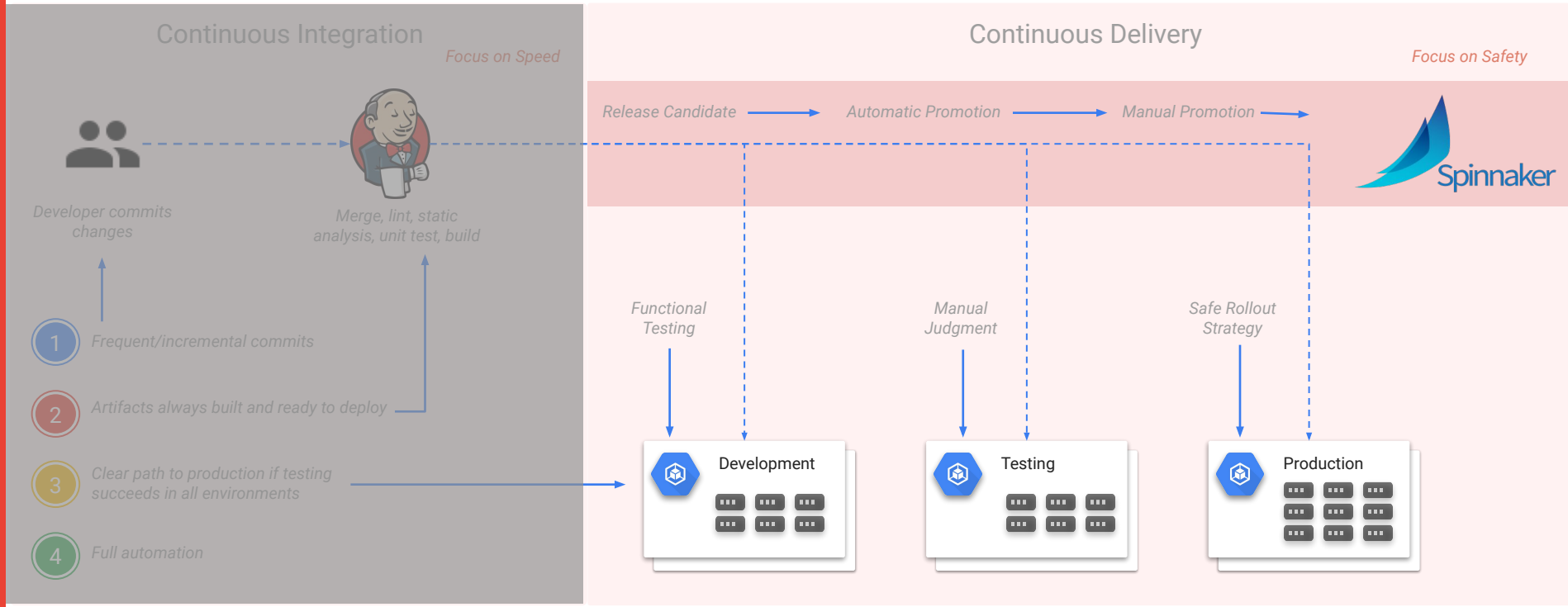
Methodologies

Integration testing, performance testing, user-acceptance testing, canary releases, blue-green deployments

Google

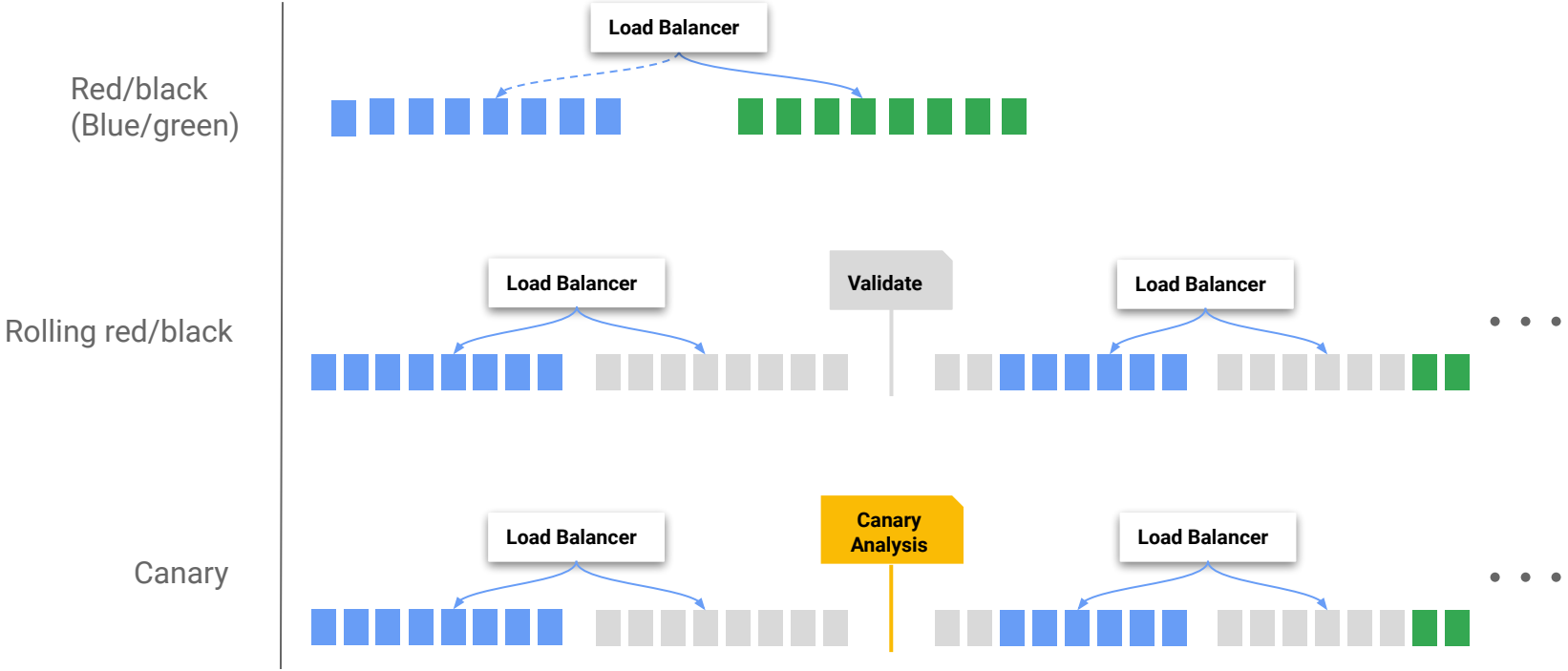
What is the Continuous Delivery (CD) process?

Continuous delivery packages code and sets the release cycle for the environment. From there it can be promoted to production



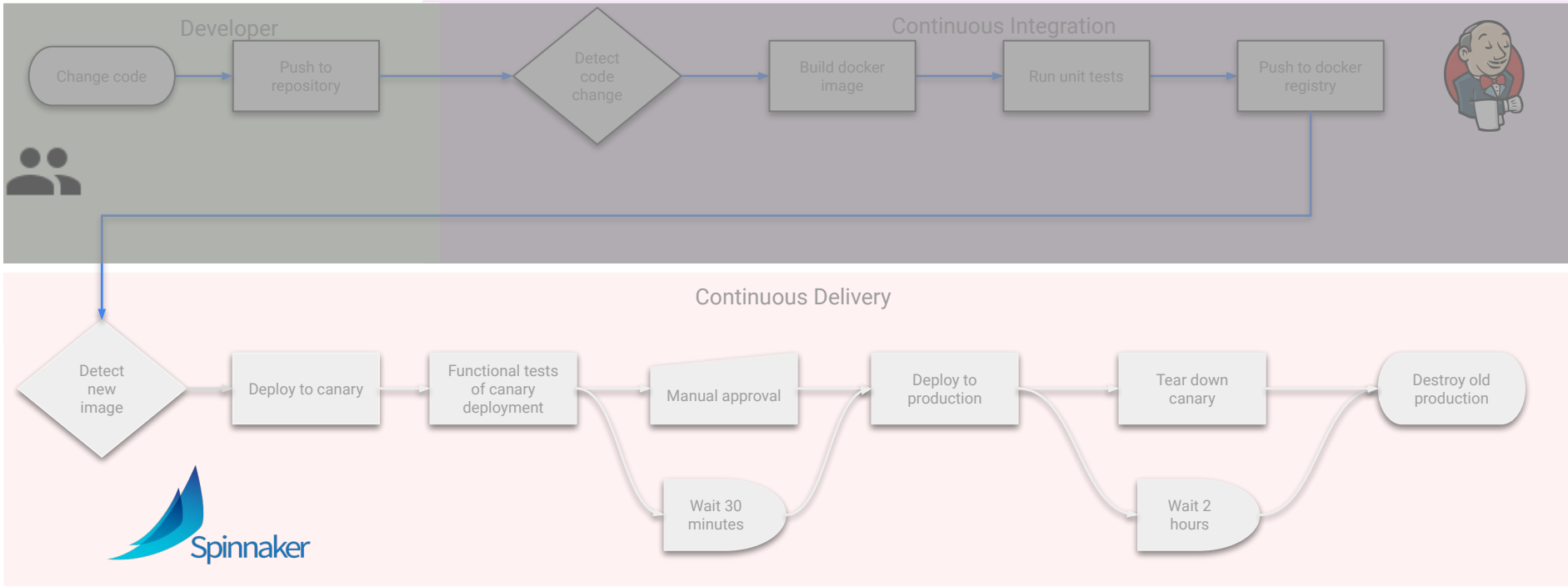
What are some deployment strategies?

There are various deployment methods that allow for testing out new code changes with a easy roll back plan



What is the pipeline flow for deployment management?

There are multiple ways a pipeline can deploy code. One way mentioned below is the canary method, which will release code to a percentage of the environment for a period of time and slowly roll out the changes to the whole environment



Google

How is Spinnaker used for continuous deployment?

Spinnaker was originally developed by Netflix and then was open sourced to the community. Google quickly became one of the biggest contributors to the project.

- Spinnaker is an open source, multi-cloud continuous delivery platform for releasing software changes with high velocity and confidence.
- Deploy across multiple cloud providers including AWS EC2, Kubernetes, Google Compute Engine, Google Kubernetes Engine, Google App Engine, Microsoft Azure, and Openstack, with Oracle Bare Metal and DC/OS coming soon.
- Create deployment pipelines that run integration and system tests, spin up and down server groups, and monitor your rollouts. Trigger pipelines via git events, Jenkins, Travis CI, Docker, CRON, or other Spinnaker pipelines.
- Create and deploy immutable images for faster rollouts, easier rollbacks, and the elimination of hard to debug configuration drift issues. Leverage an immutable infrastructure in the cloud with built-in deployment strategies such as red/black and canary deployments.

What are some Spinnaker strengths and weaknesses?

Strengths and weaknesses to consider when evaluating Spinnaker for continuous deployment

Strengths

- Tight integration with Kubernetes workload resources (Pod, ReplicaSet, Service, Ingress, etc.)
- Agnostic on deployment destination (Kubernetes Engine, GCE, AWS EC2, etc.)
- Active contributions by Netflix, Google, CoreOS, and others
- Detailed roadmap for pipeline templates and pipeline-as-code (currently alpha)
- HA options available

Weaknesses

- Fairly new product that lacks a high level of community support
- Primarily for CD, not CI — often still need Jenkins or another CI tool
- Relatively complex architecture
- Currently, largely UI based

Google Cloud

What is Configuration Management?

Configuration management is essential for after the code is deployed. This is used to keep servers and environments in sync

Purpose

Provision, manage, and **standardize** infrastructure deployments and software configurations **at scale**

Benefits

Increase **efficiency, stability, and control** with better monitoring of resources

Revert back to past deployments quickly when failures arise

Popular Tools

Provisioning

Google Cloud Deployment Manager, Terraform

Configuration

Chef, Puppet, Ansible

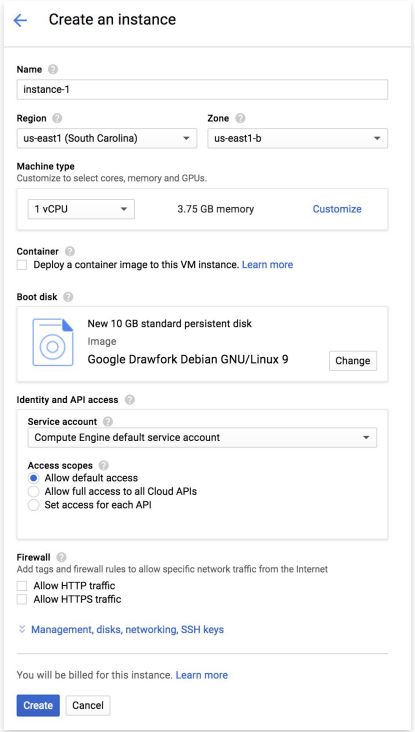
Google

What is a recommended provisioning strategy?

Manual configuration has many disadvantages, for example, human error. End users can easily miss a step or deploy infrastructure out of sync from the rest of the environment. With automated provisioning with tools like Terraform, it allows you to use source control and ensure all configurations are deployed consistently.

Manual

Not recommended



Automated

Recommended

HashiCorp Terraform

Cloud Deployment Manager

```
resource "google_compute_instance" "default" {
  name         = "test"
  machine_type = "n1-standard-1"
  zone         = "us-central1-a"

  tags = ["foo", "bar"]

  boot_disk {
    initialize_params {
      image = "debian-cloud/debian-8"
    }
  }

  // Local SSD disk
  scratch_disk {
  }

  network_interface {
    network = "default"

    access_config {
      // Ephemeral IP
    }
  }

  metadata {
    foo = "bar"
  }

  metadata_startup_script = "echo hi > /test.txt"

  service_account {
    scopes = ["userinfo-email", "compute-ro", "storage-ro"]
  }
}
```

Google

How does Deployment Manager provision resources?

Deployment Manager enables repeatable template driven deployments that utilize a declarative approach to configure resources comprising the application.

Repeatable deployments

Create configuration files to define resources which can be repeated consistent results. Eg. Python or Jinja2 templates

Declarative Language

A declarative approach allows the user to specify what the configuration should be and let the system figure out the steps to take.

Focus on the application

The user can focus on the set of resources which comprise the application or service instead of deploying each resource separately.

Template driven deployments

Templates allow the use of parameterized building blocks to create abstractions or sets of resources that are typically deployed together.

Google

What is a recommended configuration management strategy?

When it comes to configuration management, it is generally good practice to look to automate but there are use cases where it may be more beneficial to manually provision and/or configure

Manual

Not recommended

With pressure to move quickly, software is often configured by hand.

This can quickly lead to systems that are difficult and expensive to maintain and cannot be replicated for capacity expansion, development, or testing.

Automated

Recommended

There are many tools to manage these processes automatically. These tools increase efficiency, stability, and control by improving visibility and tracking. They offer a record of all configurations deployed to a fleet and provide greater agility and faster problem resolution.

Google

How is configuration management handled in GCP?

Configuration management can be accomplished by utilizing 3rd party tools, custom scripts or pre-baked images

Open Source/Third-Party Tools

Install and configure software by utilizing 3rd party config management tools such as Ansible, Puppet, or Chef.

Tools are installed, managed, and **maintained by customer**.

Startup Scripts, Images, Custom Boot Disks

Specify a startup script in your **instance metadata** and **execute it on startup** of the VM instance.

Create a **snapshot as a foundation** for booting other instances.

Google

What are some key decisions to make?

When designing continuous integration and continuous delivery for your target state, there are several key decisions an organization must make.

1. How will resources be provisioned in GCP?
2. What tools will leverage for CI/CD?
3. How will resources be provisioned in GCP?
4. How will GCP configurations be managed?
5. What tools will be leveraged for deployments?

Google

What are best practices for automated operations?

Key considerations and recommendations related to automated operations.

1. Use version control software to manage source code, test scripts, config files, and infrastructure-as-code
2. Check in, build, and integrate code frequently
3. Standardize tools and processes used across the organization
4. Enable simple methods for deploying ready builds and automate when possible
5. Monitor build and test results for failures and leverage metrics for continuous improvement
6. Always test code in an environment as similar to production as possible

Google

How to access additional resources?

Learn more about automate operations through public documentation links, tutorials, and videos.

Documentation

- Continuous Delivery Tool Integrations
- Jenkins on Container Engine
- Compute Engine Management with Puppet, Chef, Salt, and Ansible
- Continuous Deployment to App Engine Using Bitbucket Pipelines

Tutorials

- Continuous Deployment to Container Engine using Jenkins

Videos

- Scalable Deployments and Updates in Compute Engine
- Gaining full control over your organization's cloud resources
- Spinnaker: continuous delivery from first principles to production

Google

How can I test my understanding of Automated Operations?

Work within your team to reinforce concepts by applying them to a real life use case. It is essential to consider the customer's business and technical requirements when designing the solution.

Objective

- To be able to identify when to automate or manually deploy an environment and to identify various components of a pipeline.

Problem Statement

1. Manual vs Automate:
    a. Deploy GCE instance to learn
    b. All servers need to be in CIS compliance
    c. Highly complex environment with the potential to be deployed to multiple environments
    d. Deploy a Cloud Function for learning purposes
    e. Highly complex environment - one-time deployment

2. Identify which part of the pipeline below fits.
    a. What part of the lifecycle would you use Jenkins for typically?
    b. What part of the lifecycle would you use Spinnaker for typically?
    c. What are some examples of Configuration Management tool?
    d. When would you use a tool like Puppet over Deployment Manager

Google

How can I test my understanding of Automated Operations?

Work within your team to reinforce concepts by applying them to a real life use case. It is essential to consider the customer's business and technical requirements when designing the solution.

Solution:

Scenario	Answer	Explanation
Deploy GCE instance to learn		
All servers need to be in CIS compliance		
Complex environment with potential to be deployed to multiple environments		
Deploy a Cloud Function for learning purposes		
Highly complex environment - one-time deployment		

Google

How can I test my understanding of Automated Operations?

Work within your team to reinforce concepts by applying them to a real life use case. It is essential to consider the customer's business and technical requirements when designing the solution.

Solution:

Question	Answer	Explanation
What part of the lifecycle would you use Jenkins for typically?		
What part of the lifecycle would you use Spinnaker for typically?		
What are some examples of Configuration Management tool?		
When would you use a tool like Puppet over Deployment Manager		

Google

Billing

Billing

This section will focus on the following key topics.

Objectives

- Key considerations for designing billing strategy in GCP
- Key considerations for managing billing reporting and analysis in GCP

Key Learnings

- Understanding of GCP billing process and structure
- Evaluating billing reports
- Controlling spending and costs
- Leveraging billing user roles for billing management
- Leveraging billing user roles for billing management
- Understanding labels
- Exporting billing data

Google Cloud

What do we need to know?

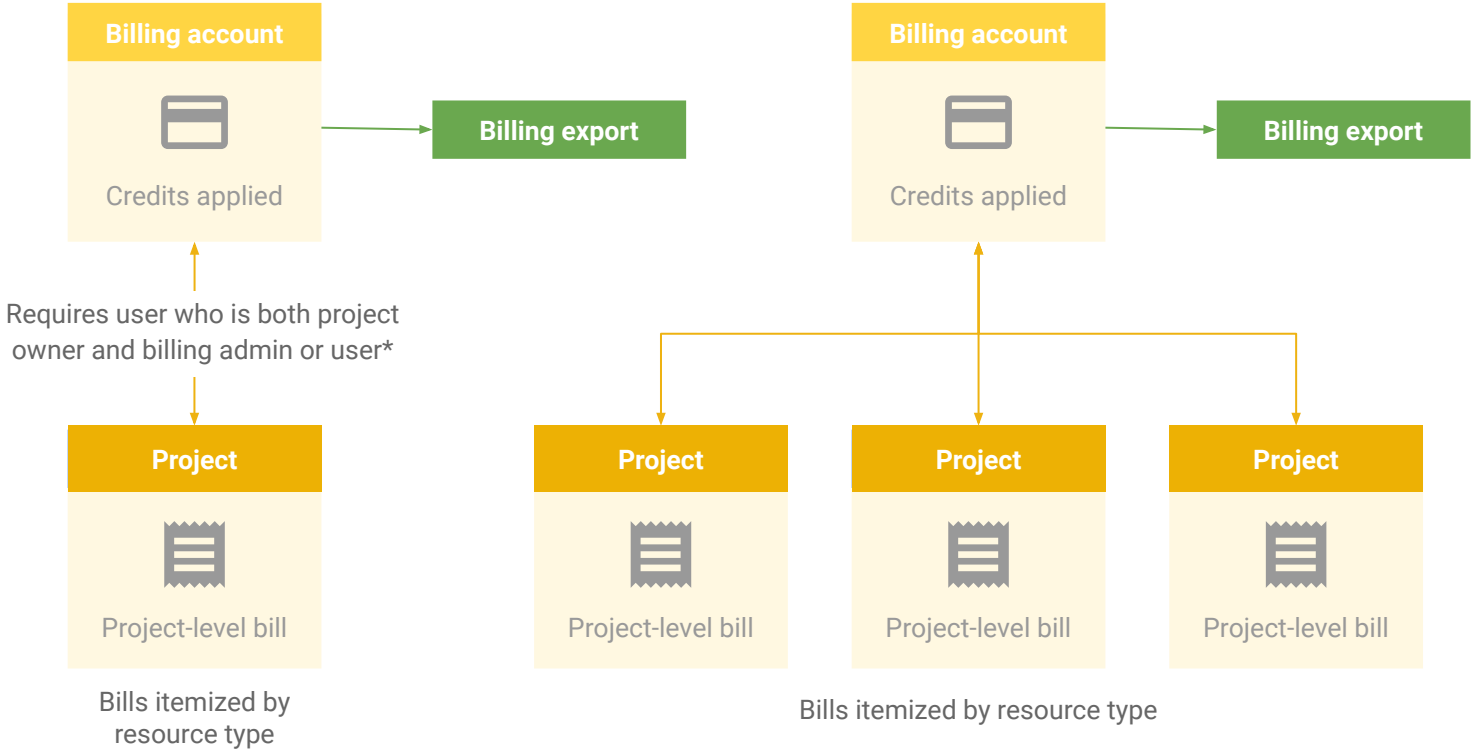
There are key aspects you need to know in order to design a client's billing strategy in GCP.

- How billing is managed in GCP
- Billing reports
- Using labels to customize reports
- Exporting and analyzing billing data
- Budgets and spending limits
- Managing billing

Google Cloud

How is billing managed in GCP?

Billing accounts are attached to organizations / projects and are billed for GCP resource consumption for those projects. It is possible to create multiple billing accounts.



Google

How can a billing report be customized?

Billing labels can be assigned to GCP resources to create customized billing reports (e.g. business unit, cost center) based on organizational needs and accounting strategy.

What resources can be labeled?

- Virtual machine instances
- Forwarding rules (Alpha)
- Images
- Persistent disks
- Persistent disk snapshots
- Static external IP addresses (Alpha)
- VPN tunnels (Alpha)

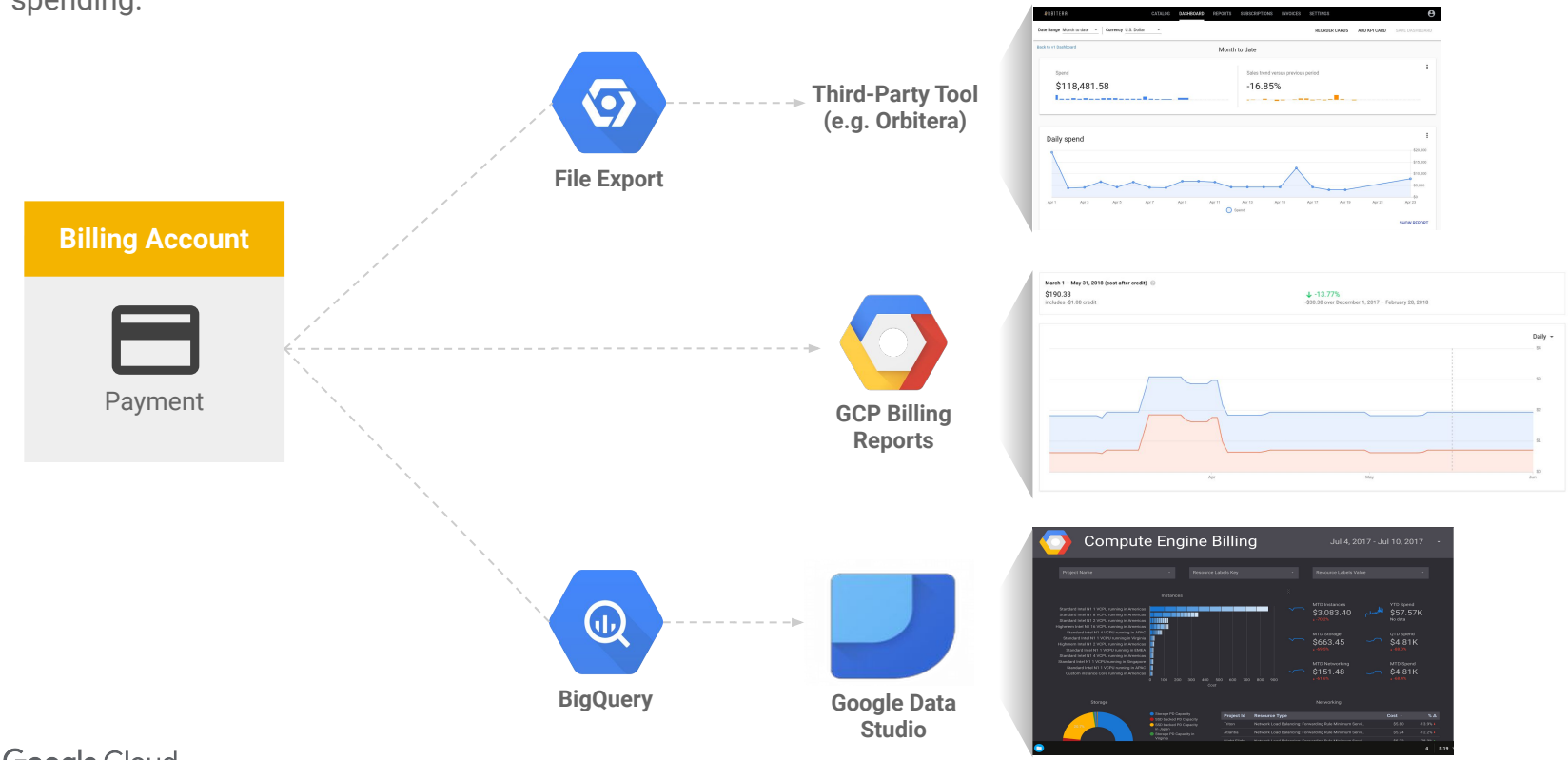
Considerations for applying labels

- Focus on label consistency
  Apply labels programmatically if possible
- Make labels a simple, standard set of values that are useful both technically and business value–wise
  Despite the ability to apply up to 64 labels per resource, try to stick with no more than 5 labels
- Follow.json format of -l <key>:<value>
  - No more than 63 characters each
  - Only contain lowercase letters, numeric characters, underscores, and dashes

Google Cloud

How can billing reports be used to drive insights?

Data from billing reports can be exported to analytics tools in order to analyze spending patterns and predict future spending.



Google Cloud

Exports to BigQuery

Billing dataset in BQ enables users to slice and dice billing data to meet reporting requirements of various business units.

Setup

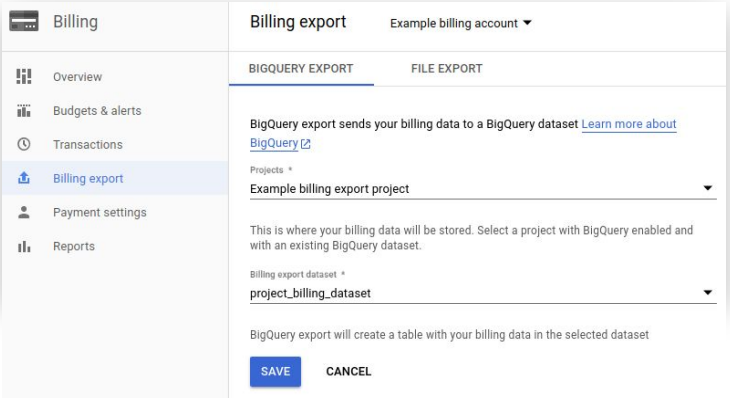
- Simple one-time setup

Project

- Project to host the BQ dataset

Billing export dataset

- Dataset where table will be created
- If none exist, you can create one



Results | Explanation | Job Information

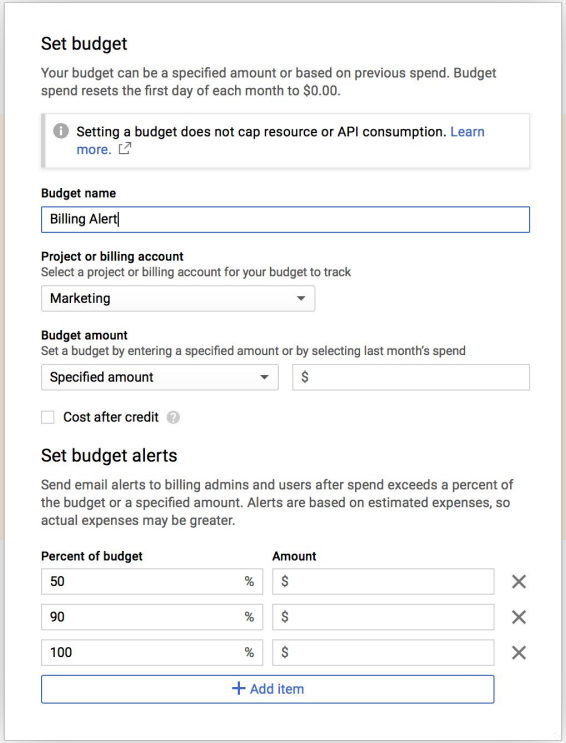
Row	project_labels_key	project_labels_value	cost_total	usage_total
1	pubsub	metric	40.88075199999999	3.2135524418585517E17
2	team	sales	2.905382	1.1376485747702972E16
3	gce-enforcer-fw-opt-out	testing-customer-use-cases	187.70291399999994	2.07529806639022797E18
4	testprojectlabel		44.522840999999985	1.0482929879316314E16
5	null	null	5272.94248	2.7390326190994706E19
6	team	marketing	44.522840999999985	1.0482929879316314E16
7	cost_center	34910481	2.905382	1.1376485747702972E16

Google

What are budgets and spending limits?

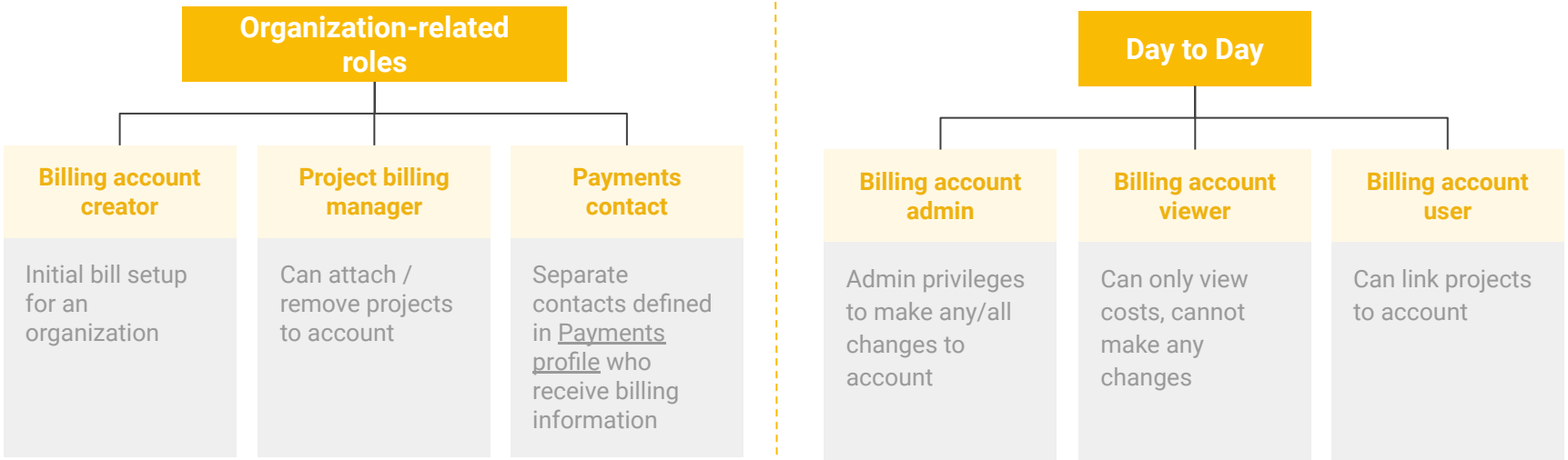
The analysis of data from billing reports can be used to determine daily usage budgets and spending limits to gain more control over resource and application costs.

- A **budget** can be set for a project to send alerts
- Budgets may be set for a billing account or for a project. After the monthly budget is set, custom thresholds for alerts may be set (e.g., 50%, 75%, 100%)
- Reaching a budget or threshold has no resource restriction implications. All GCP resources continue to function normally.



What Billing IAM roles are available?

Billing roles can be managed at the organization level using the Org Billing Account Admin/Creator/User roles.



Google Cloud

What are best practices for billing?

Adhering to recommended best practices for logging processes and performance monitoring enables long-term success of a GCP project.

1. Naming convention for projects and resources makes it easy to see which teams or products are consuming resources
2. Use project labels to categorize resources enabling ability to query GCP usage data
3. Control spending by setting budgets appropriate for your projects
4. Set alerts for monthly spending thresholds
5. Developing a formal automated process to request quota increases to help with stronger enforcement
6. Exporting reports to analytics and visualization tools will help analyze spending patterns and predict future spending

Google Cloud

What are some key decisions to make?

Keep the following key parameters in mind whilst developing a billing strategy in GCP.

1. How will resource utilization be tracked?
2. Which billing accounts are required?
3. How will budgets be used?
4. Will billing need to be integrated with existing systems?

Google

How to access additional resources?

Learn more about billing through public documentation links, tutorials, and videos.

Documentation

- IAM roles for Billing-related Job Functions
- Pricing details on each GCP product
- GCP Billing API
- Get Started with the Google Cloud Billing API
- Best practices for billing

Tutorials

- Open an Account and Manage Billing and Projects

Videos

- Guide to Google Cloud Platform billing
- Saving Money on Compute Engine

Google Cloud

Wrap-Up

Questions?