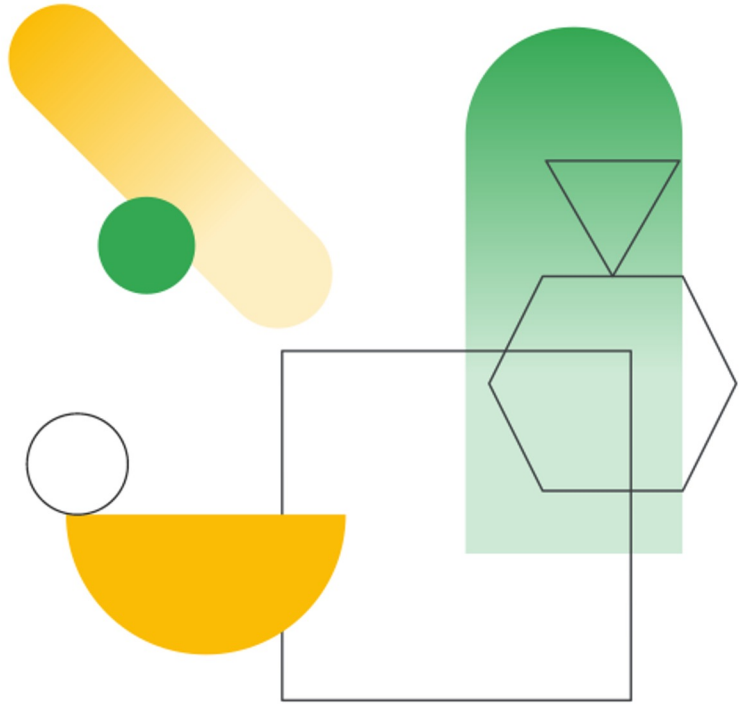
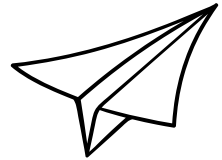
Google Cloud

Google CLoud Platform Crash Course



Course timeline



Day 1

- Intro to GCP, Resources Hierarchy, APIs & Billing.
- Google Cloud SDK.
- Identity and Access management (IAM)

Day 2

- Networking - Basics
- Compute Services - GCE

Day 3

- Networking - Advanced
- Compute Services - GKE

Day 4

- Storage and Databases
- Serverless services and other selected services
- Final Task

Day 5

- Big Data Services
- Q&As

Course evaluation

30%

Get you hands dirty and please don’t exhaust Sarah & Eslam 🥺

60%

Do you think you can get them all?

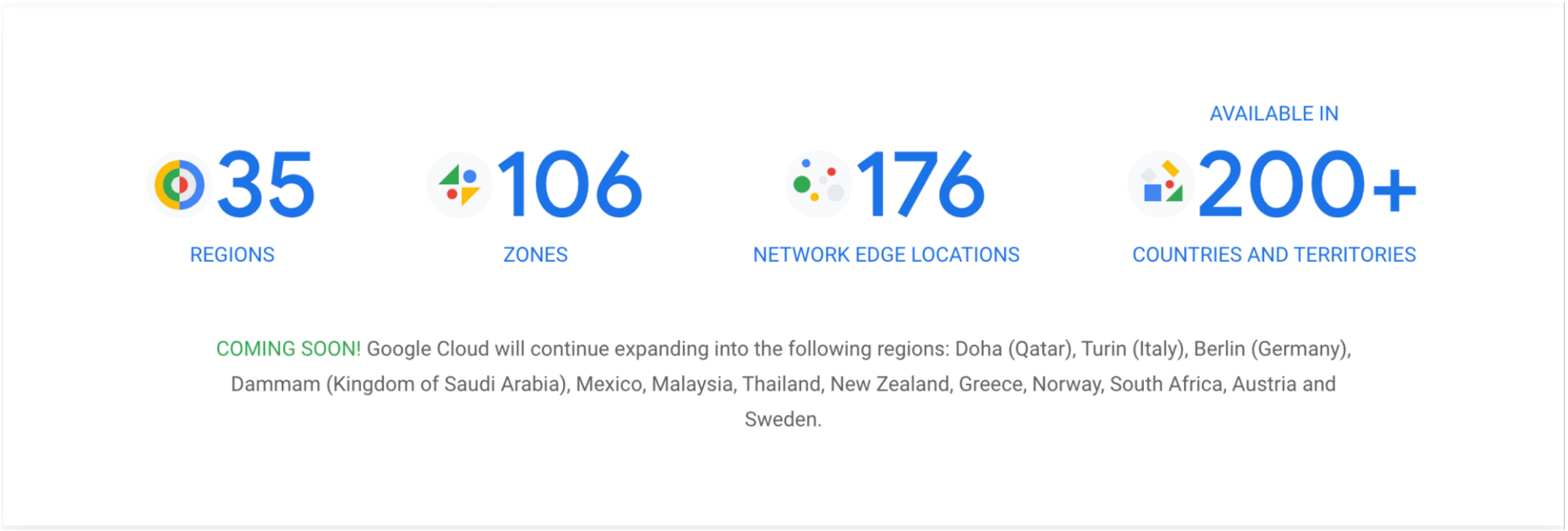
Well, think again! because no one ever did 😈

10%

That's a gift ❤️

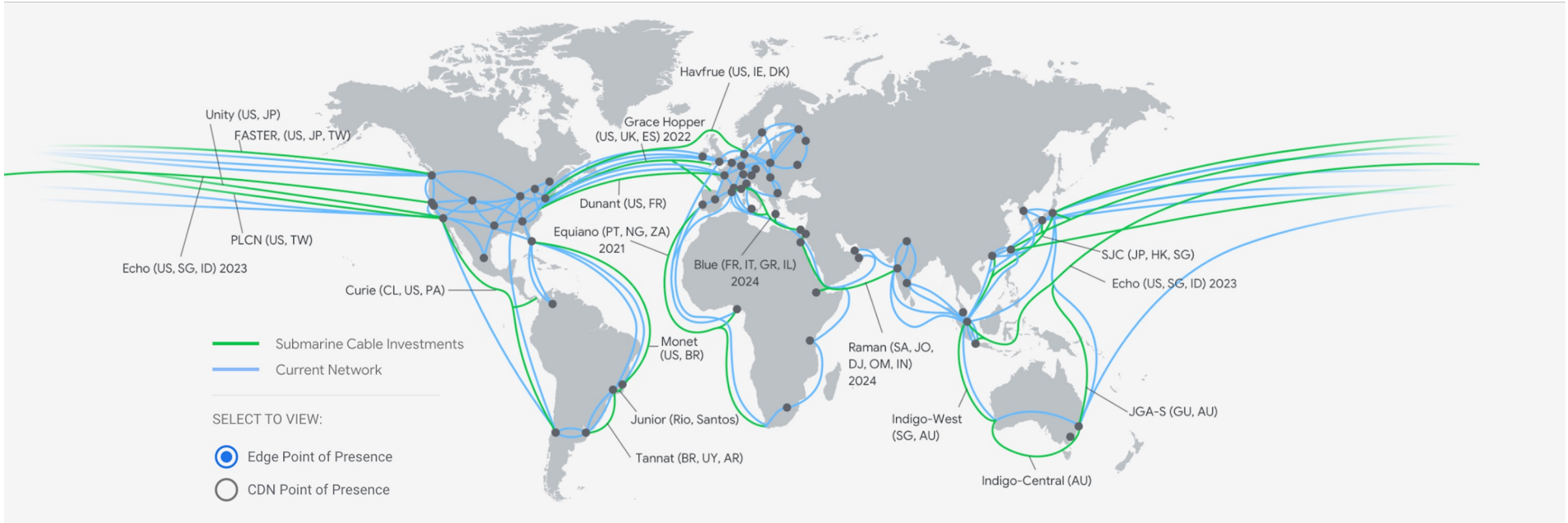
Intro to
Google Cloud Platform (GCP)

GCP locations and Network



GCP locations and Network – cont.

GCP locations and Network – cont.



GCP top services

Compute Services	Serverless (FaaS): Cloud Functions PaaS: App Engine	Containers: GKE & Cloud Run VMs: Compute Engine
Storage Services	Object Storage: Google Storage Buckets. Instance Storage: Persistent Disks. SQL: Cloud SQL & Cloud Spanner	No-SQL: Cloud Datastore/Firestore & Cloud BigTable Analytics: Cloud BigQuery
Big Data Services	Cloud Dataproc Cloud Dataflow Cloud Composer	Cloud Transfer Cloud BigQuery	Cloud Pub/Sub Cloud Data Fusion

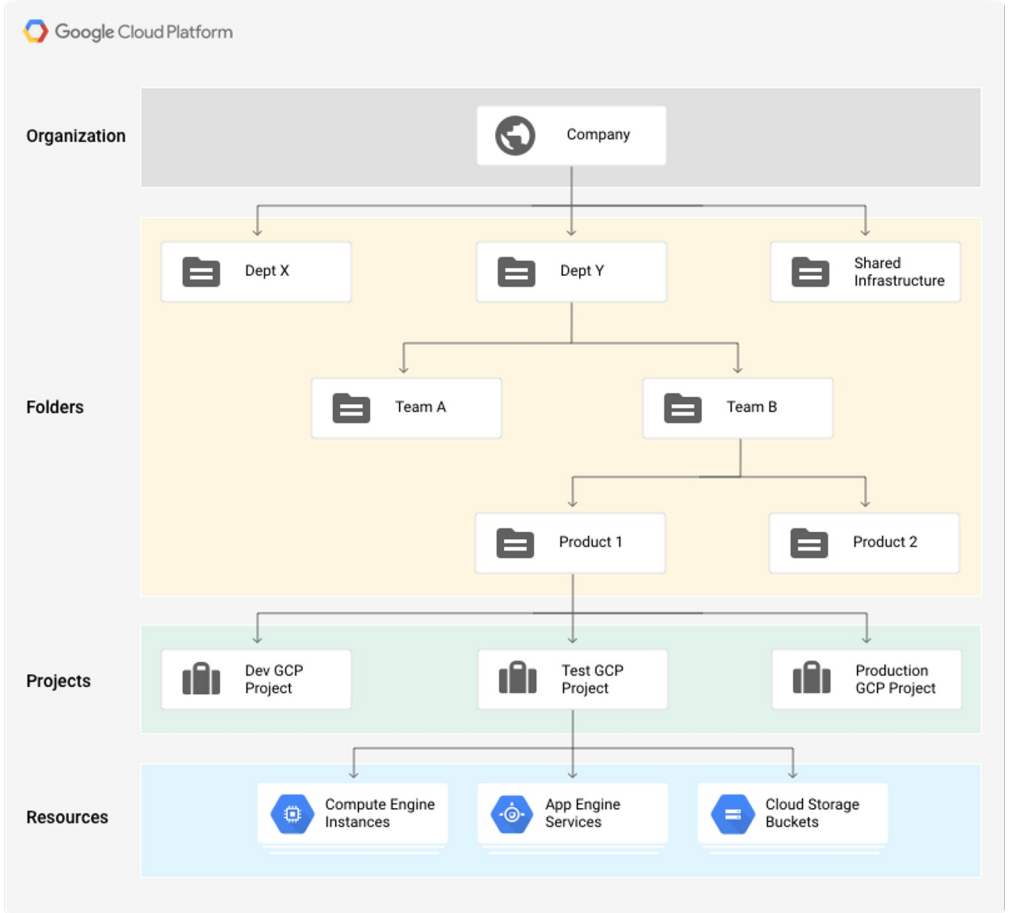
[Google Cloud Platform Services Summary]

First look at GCP console

https://console.cloud.google.com

GCP Resources Hierarchy, Enabling APIs & Setup Billing

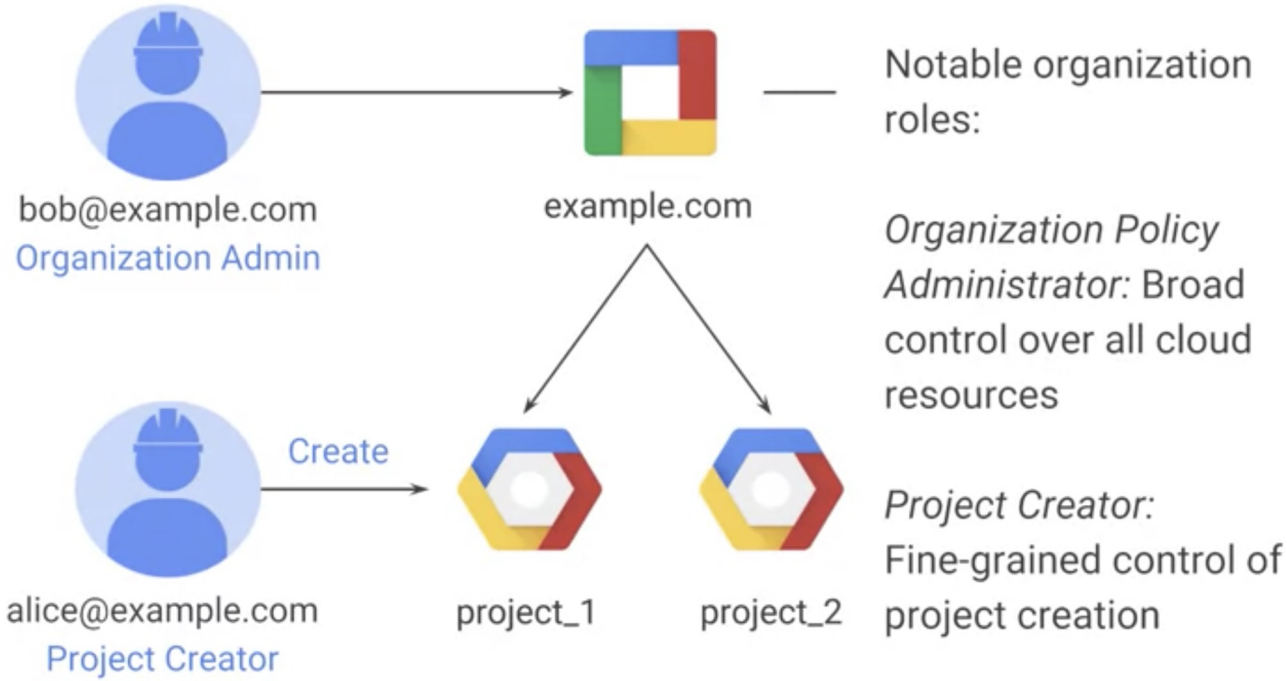
Resources Hierarchy



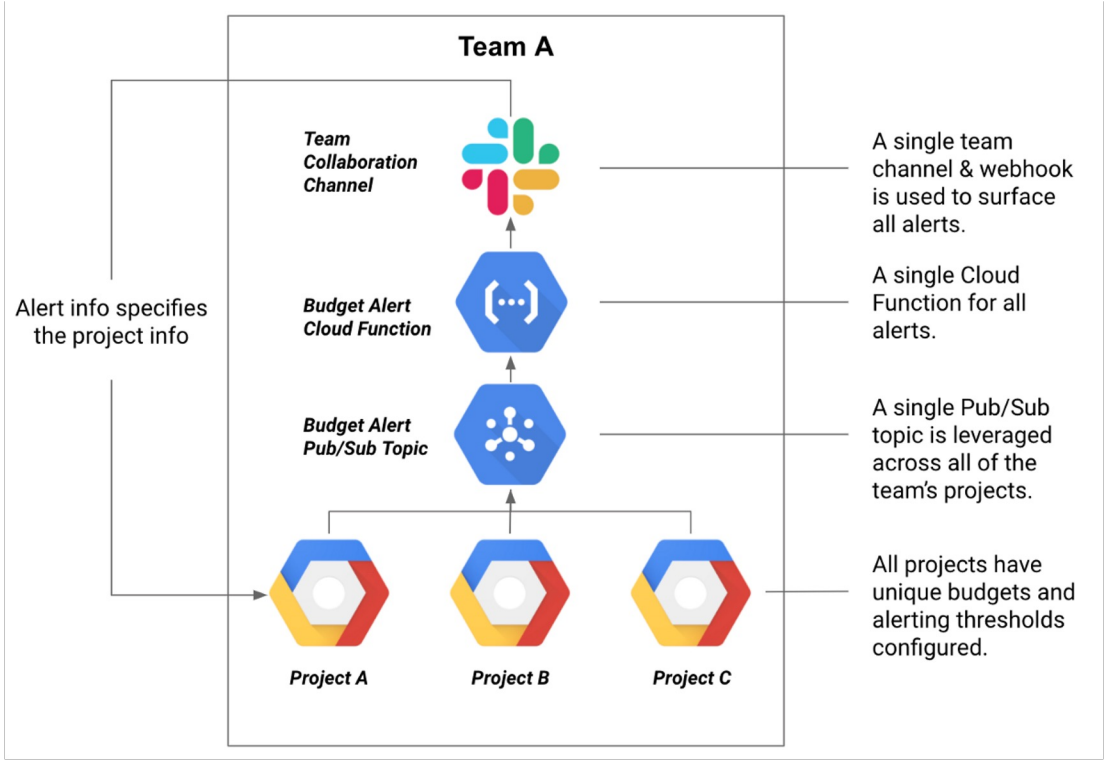
Resources Hierarchy – cont.

Project ID	Globally unique	Chosen by you	Immutable
Project name	Need not be unique	Chosen by you	Mutable
Project number	Globally unique	Assigned by GCP	Immutable

Resources Hierarchy – cont.



Setup billing budgets and billing Alerts



[Create, edit, or delete budgets and budget alerts | Cloud Billing | Google Cloud]

[Overview of Cloud Billing access control]

Google APIs

Google Cloud APIs
Compute Engine API
BigQuery API
Cloud Storage Service
Cloud Datastore API
Cloud Deployment Manager API
Cloud DNS API
More

Google Maps APIs
Google Maps Android API
Google Maps SDK for iOS
Google Maps JavaScript API
Google Places API for Android
Google Places API for iOS
Google Maps Roads API
More

Google Apps APIs
Drive API
Calendar API
Gmail API
Sheets API
Google Apps Marketplace SDK
Admin SDK
More

Mobile APIs
Google Cloud Messaging
Google Play Game Services
Google Play Developer API
Google Places API for Android

Social APIs
Google+ API
Blogger API
Google+ Pages API
Google+ Domains API

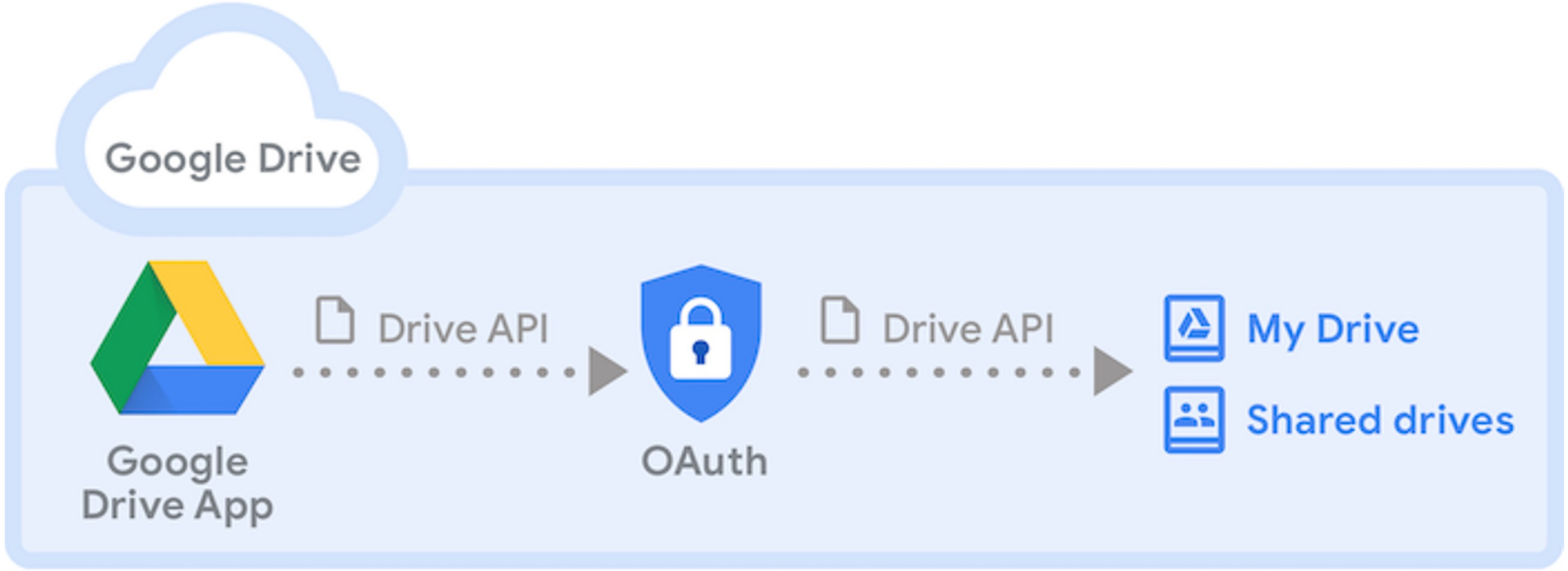
YouTube APIs
YouTube Data API
YouTube Analytics API
YouTube Reporting API

Advertising APIs
AdSense Management API
DCM/DFA Reporting And Trafficking API
Ad Exchange Seller API
Ad Exchange Buyer API
DoubleClick Search API
DoubleClick Bid Manager API

Other popular APIs
Analytics API
Translate API
Custom Search API
URL Shortener API
PageSpeed Insights API
Fusion Tables API
Web Fonts Developer API

https://cloud.google.com/apis
https://developers.google.com/apis-explorer

Google APIs - Cont.



Cloud Shell & Cloud SDK

Cloud SDK is a set of tools that you can use to manage resources and applications hosted on Google Cloud. These tools include the gcloud, gsutil, and bq command-line tools

Quickstart: Getting started with Cloud SDK | Cloud SDK Documentation | Google Cloud

Installing Cloud SDK

https://shell.cloud.google.com

The gcloud tool

The gcloud command-line tool is the primary CLI tool to create and manage Google Cloud resources. You can use this tool to perform many common platform tasks either from the command line or in scripts and other automations.

The `gcloud` tool and Cloud SDK

The `gcloud` tool is a part of the Cloud SDK. Before you can use the `gcloud` tool, you must download and install the Cloud SDK on your system and initialize Cloud SDK .

By default, the Cloud SDK installs the `gcloud` tool commands that are at the General Availability level. Additional functionality is available in Cloud SDK components named `alpha` and `beta`. These components allow you to use the `gcloud` tool to work with Cloud Bigtable, Dataflow and other parts of the Google Cloud at earlier release levels than General Availability.

The `gcloud` tool releases have the same version number as the Cloud SDK. The current Cloud SDK version is 367.0.0. You can download and install previous versions of the Cloud SDK from the download archive.

The gcloud tool - cont.

The `gcloud` tool is a tree; non-leaf nodes are command groups and leaf nodes are commands. (Also, tab completion works for commands and resources!)

Most gcloud commands follow the following format:

```
gcloud + release level (optional) + component + entity + operation + positional args + flags
```

For example: `gcloud + compute + instances + create + example-instance-1 + --zone=us-central1-a`

[The gcloud tool cheat sheet | Cloud SDK Documentation]

https://cloud.google.com/sdk/docs/cheatsheet#understanding_commands

The gcloud tool - cont.

SYNOPSIS

```
gcloud GROUP | COMMAND [--account=ACCOUNT] [--billing-project=BILLING_PROJECT]
  [--configuration=CONFIGURATION] [--flags-file=YAML_FILE] [--flatten=[KEY,...]]
  [--format=FORMAT] [--help] [--project=PROJECT_ID] [--quiet, -q] [--verbosity=VERBOSITY;
  default="warning"] [--version, -v] [-h] [--access-token-file=ACCESS_TOKEN_FILE]
  [--impersonate-service-account=SERVICE_ACCOUNT_EMAILS] [--log-http]
  [--trace-token=TRACE_TOKEN] [--no-user-output-enabled]
```

gcloud | Cloud SDK Documentation

The gcloud tool - cont.

The `gcloud` tool commands have the following release levels:

Release level	Label	Description
General Availability	None	Commands are considered fully stable and available for production use. For advance notice of changes to commands that break current functionality, see the release notes.
Beta	`beta`	Commands are functionally complete, but could still have some outstanding issues. Breaking changes to these commands can be made without notice.
Alpha	`alpha`	Commands are in early release and may change without notice.

The `alpha` and `beta` components are not installed by default when you install the Cloud SDK. You must install these components separately using the `gcloud components install` command. If you try to run an alpha or beta command and the corresponding component is not installed, the `gcloud` tool prompts you to install it.

The gcloud tool - cont.

Managing Cloud SDK properties

Send feedback

Properties are settings that govern the behavior of the `gcloud` command-line tool and other Cloud SDK tools.

You can use properties to define a per-product or per-service setting such as the account used by the `gcloud` tool and other Cloud SDK tools for authorization, the default region to use when working with Compute Engine resources, or the option to turn off automatic Cloud SDK component update checks. Properties can also be used to define `gcloud` tool preferences like verbosity level and prompt configuration for `gcloud` tool commands.

The gcloud tool - cont.

Configurations

A configuration is a named set of Cloud SDK properties. The `gcloud` tool uses a configuration named `default` as the initial active configuration; `default` is suitable for most use cases. However, you can also create additional configurations and switch between them as required.

Listing properties

To list the properties in the active configuration, run `gcloud config list`:

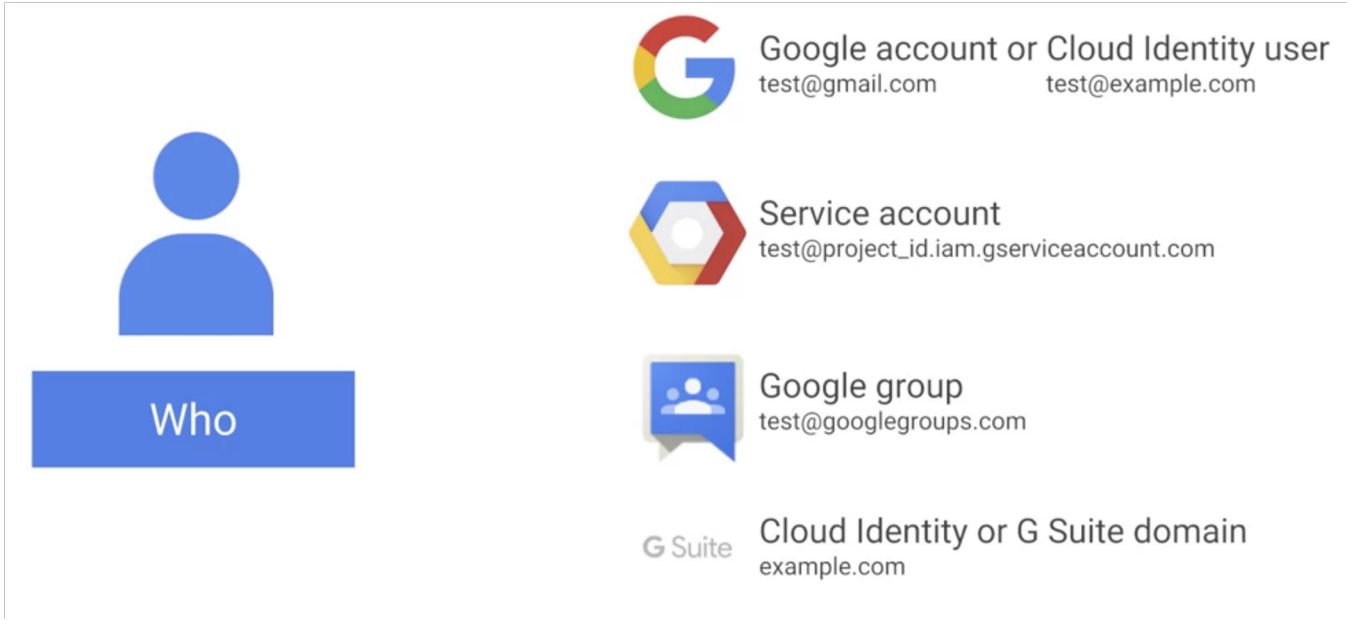
Lab 1.1

1. Explore Google Cloud Console.
2. Setup a billing method on your google account.
3. Create a GCP project.
4. Assign your billing account to your project.
5. Setup project budget.
6. Setup billing alerts.
7. Using cloud shell, list all projects and set default project.
8. Install and configure gcloud SDK on your pc.
9. List all projects using gcloud SDK and set default project

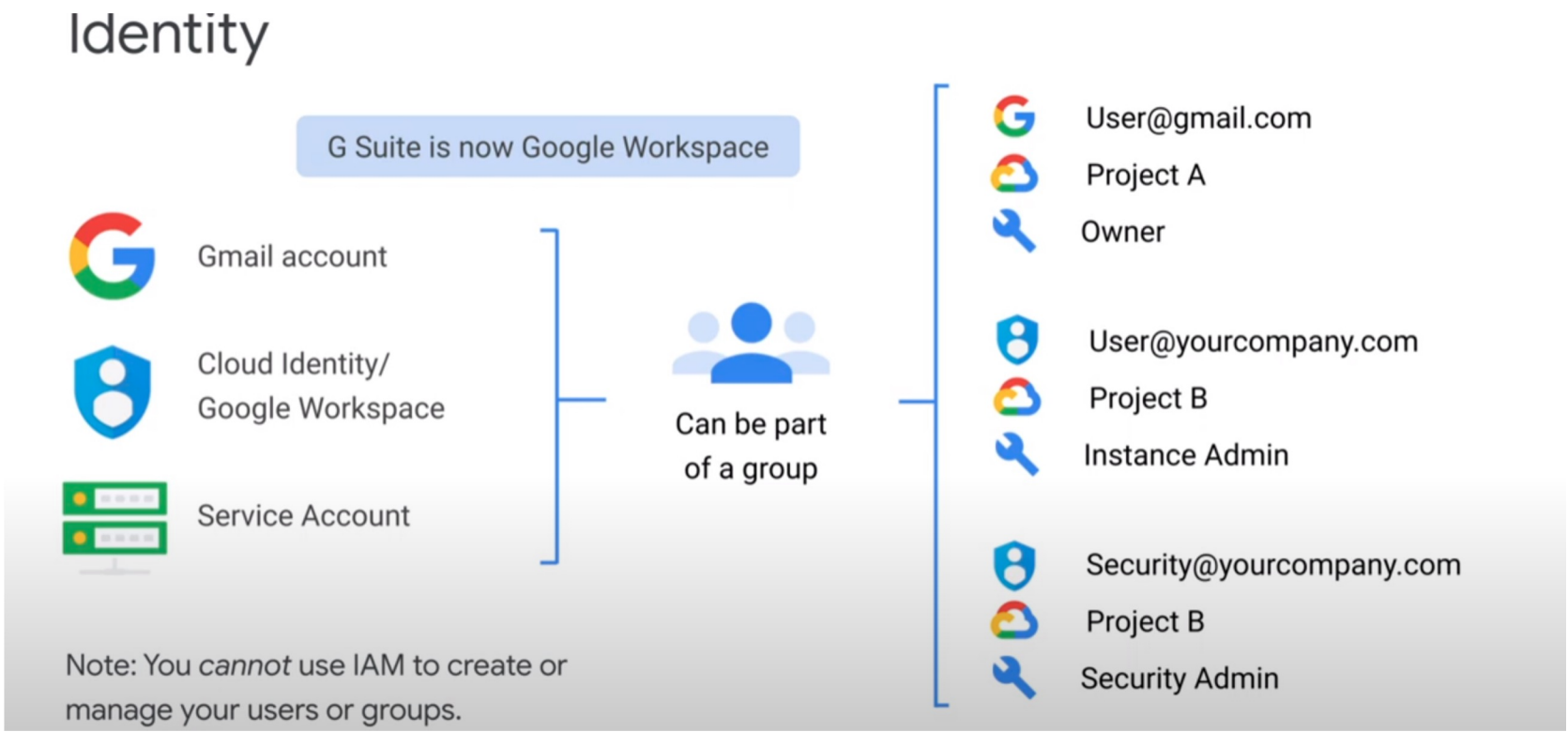
Identity and Access Management (IAM)

What is IAM?

IAM (cont.)

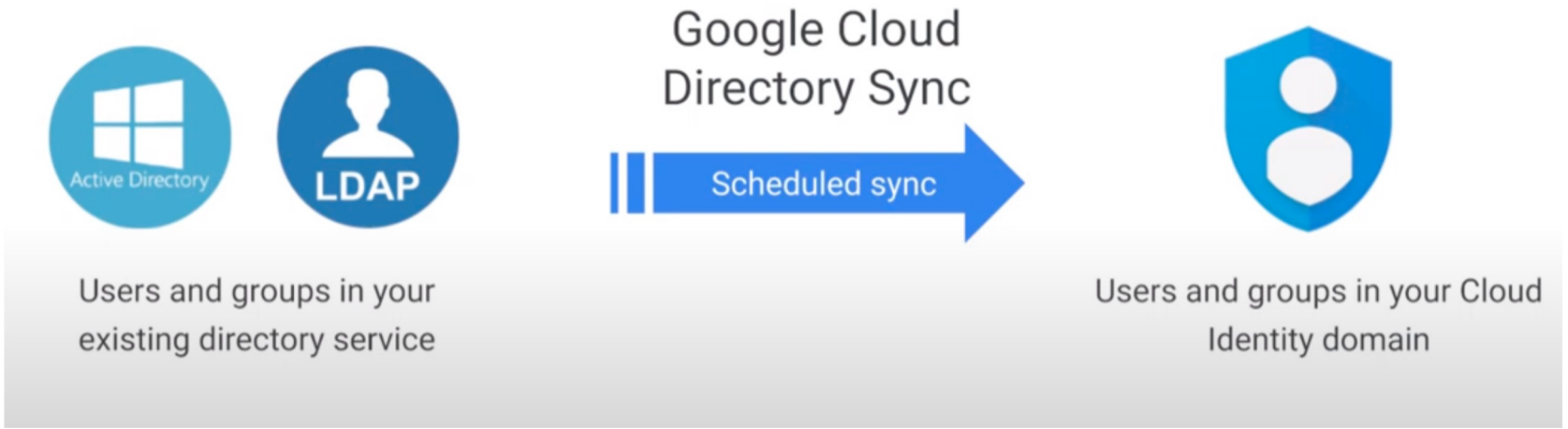

Who
Google account or Cloud Identity user
test@gmail.com
test@example.com
Service account
test@project_id.iam.gserviceaccount.com
Google group
test@googlegroups.com
G Suite
Cloud Identity or G Suite domain
example.com

IAM (cont.)

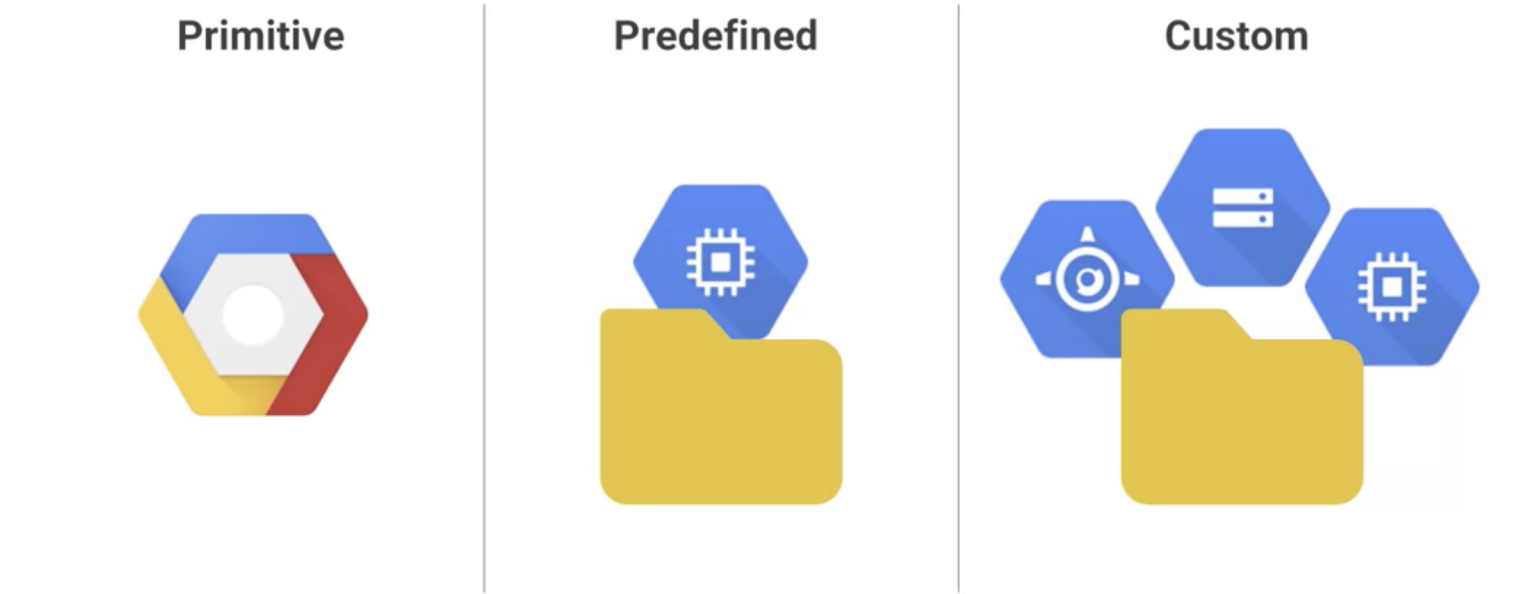

Identity
G Suite is now Google Workspace
Gmail account
Cloud Identity/
Google Workspace
Service Account
Can be part
of a group
User@gmail.com
Project A
Owner
User@yourcompany.com
Project B
Instance Admin
Security@yourcompany.com
Project B
Security Admin
Note: You cannot use IAM to create or
manage your users or groups.

IAM (cont.)

What if you already have a different corporate directory?



IAM (cont.) – Types of Roles



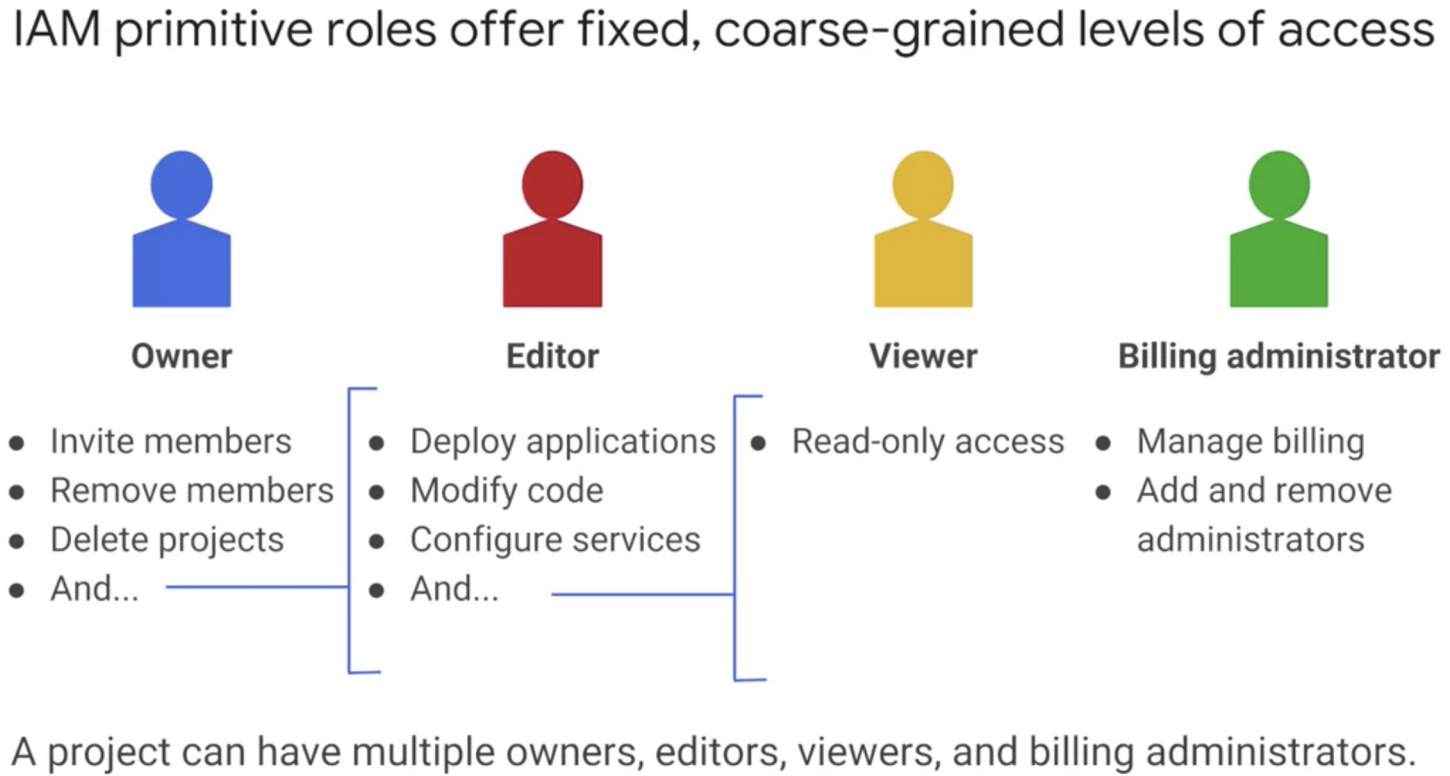
Types of roles (cont.) - Primitive Roles

IAM primitive roles apply across all GCP services in a project

can do what

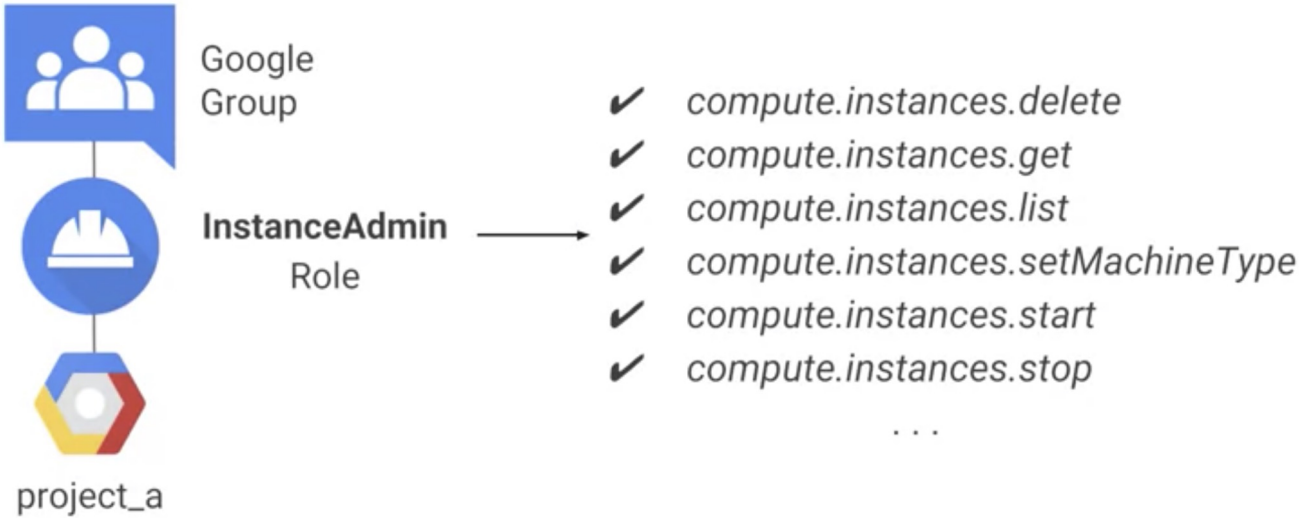
on all resources

Primitive Roles (cont.)



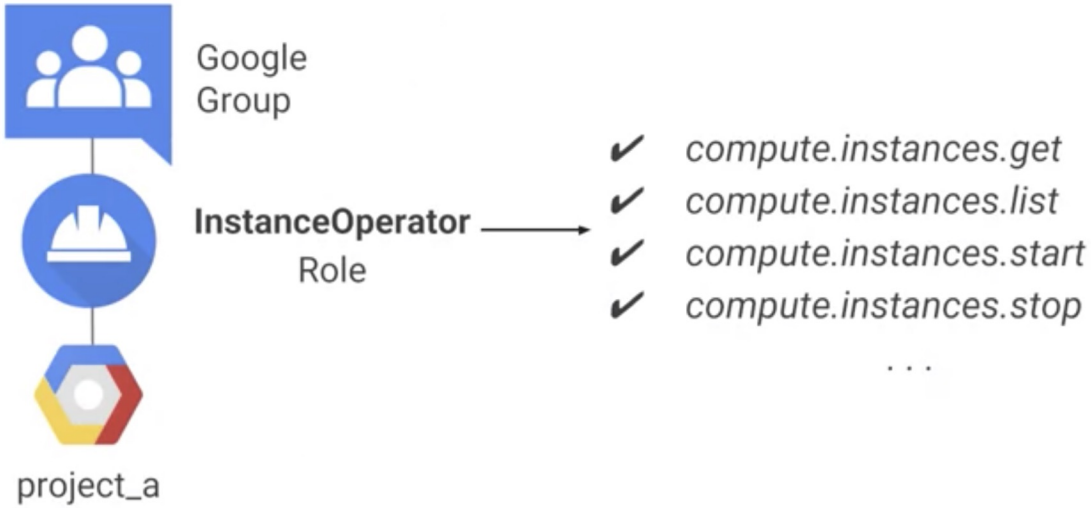
Types of roles (cont.), Pre-defined Roles

IAM predefined roles offer more fine-grained permissions on particular services



Types of roles (cont.), Custom Roles

IAM custom roles let you define a precise set of permissions



Custom Roles (cont.), notes

- Can’t be used at folders level.
- Can be used only at project level or at organization level.
- Custom roles are not maintained by google; You completely manage these roles.
- To create custom roles, you need to have the permission “iam.roles.create”

Custom Roles (cont.), Creating custom roles

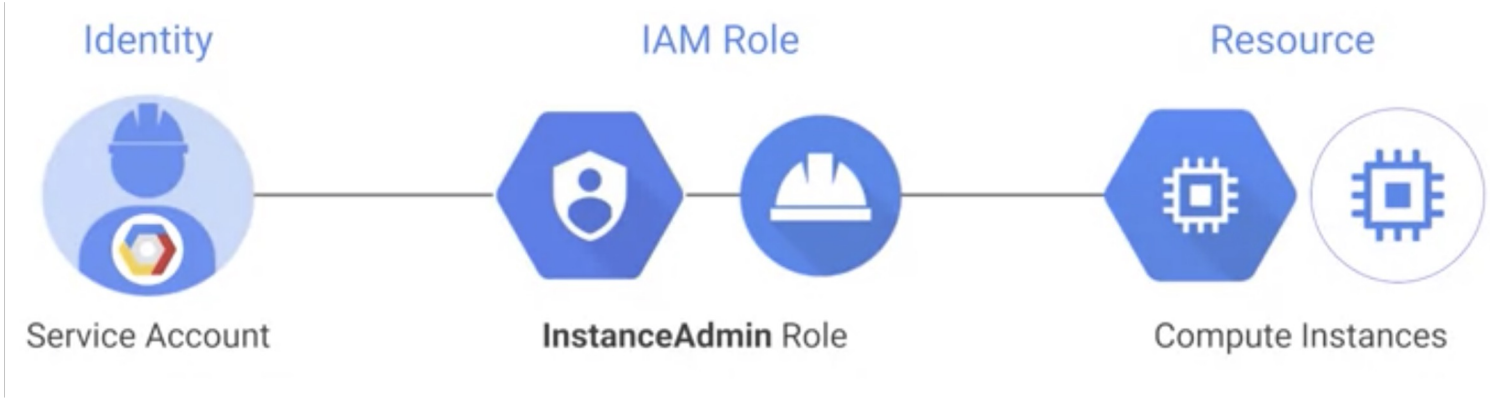
- To create a custom role at the organization level, execute the following command:

```
gcloud iam roles create role-id --organization=organization-id \
  --file=yaml-file-path
```

- To create a custom role at the project level, execute the following command:

```
gcloud iam roles create role-id --project=project-id \
  --file=yaml-file-path
```

IAM (cont.) - Service Accounts (SAs)



Service Accounts (Cont.)

- A service account is an identity that an instance or an application can use to run API requests on your behalf.
- Two types of service accounts are available to Compute Engine instances:
  - User-managed service accounts
  - Google-managed service accounts

IAM (cont.) – Policies

- IAM policies
- Docs 1 & Docs 2

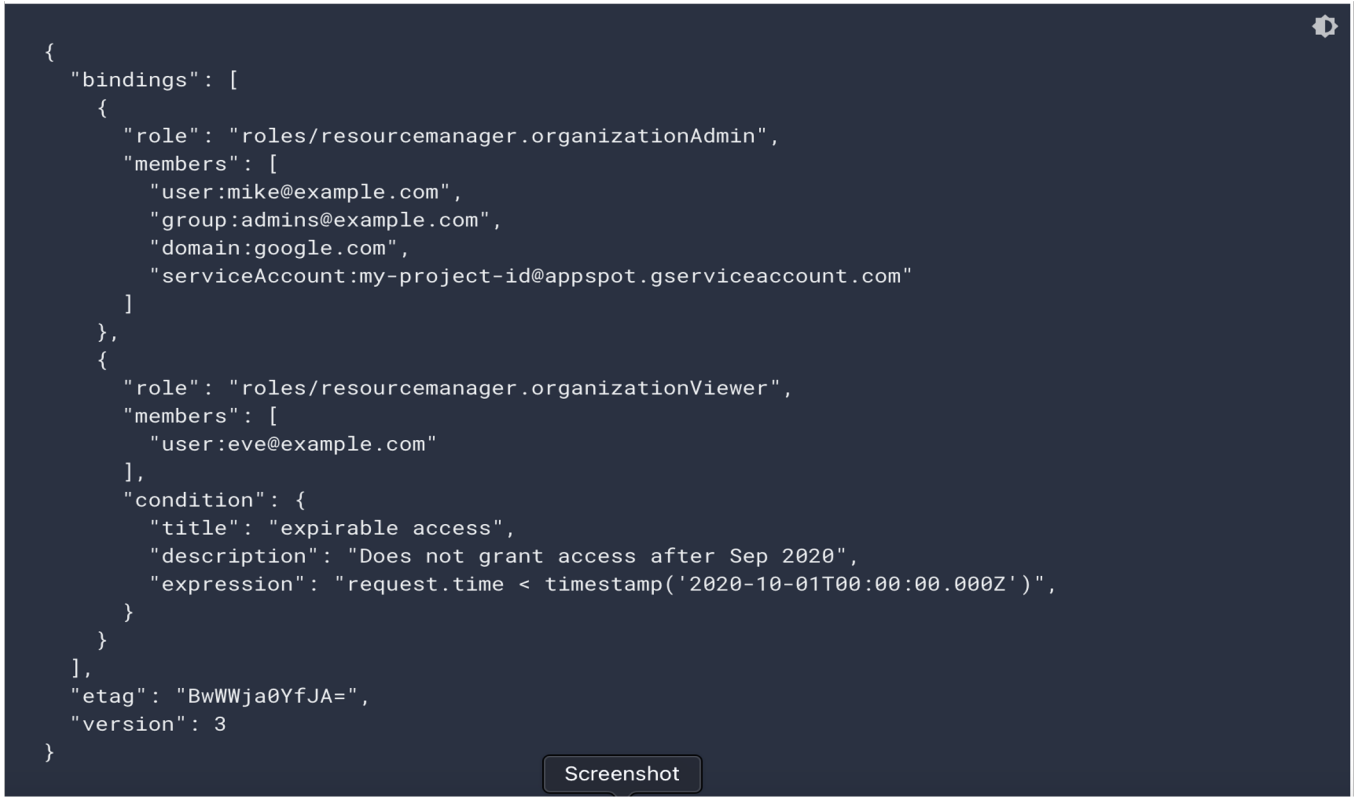
Policy

Send feedback

An Identity and Access Management (IAM) policy, which specifies access controls for Google Cloud resources.

A `Policy` is a collection of `bindings`. A `binding` binds one or more `members` to a single `role`. Members can be user accounts, service accounts, Google groups, and domains (such as G Suite). A `role` is a named list of permissions; each `role` can be an IAM predefined role or a user-created custom role.

IAM Policies (cont.)



IAM (cont.) – Best Practices

- Always consider applying the concept of Least Privilege.
- Never use primitive roles unless there is no other way.
- Rotate SAs keys and store them in a safe place like Vault.
- Enable Audit Logs.
- Apply IAM policies as much as needed at the Organization level.

- Resources:
  - Securely using IAM
  - Policy Intelligence tools
  - Video 1 & Video 2

Lab 1.2

1. From Cloud console, do the following:
    I. Create custom role named "my-custom-role-1" with the following permissions only:
        - Iam.roles.get
        - Iam.roles.list
2. From Cloud console, Explore primitive and pre-defined roles and their permissions.
3. From Cloud console, Create a service account with id "my-first-serviceaccount".
4. From Cloud console, Assign the custom role "my-custom-role-1" to the service account "my-first-serviceaccount"
5. Using gcloud,
    I. List all roles on your project.
    II. Describe the predefined role "roles/compute.viewer" and view its details & permissions
    III. Describe the custom role "my-custom-role-1" and view its details & permissions.
    IV. List all authenticated accounts.
    V. Activate the service account "my-first-serviceaccount".
    VI. List all authenticated accounts again.
    VII. Using this service account, try to list all roles on your project.
    VIII. Try to delete custom role "my-custom-role-1"

Networking in GCP

Topics to cover:

- Global vs. Regional vs. Zonal resources.
- Networking tiers
- Virtual Private Clouds (VPCs).
- IP addressing & Subnets CIDRs (optional).
- Firewall rules.

Global vs. Regional vs. Zonal resources

Global	Regional	Zonal
Global resources are accessible by any resource in any zone within the same project	Regional resources are accessible by any resources within the same region	Resources that are unique to a zone and are only usable by other resources in the same zone
• Addresses • Images • Snapshots • Instance templates • VPC network • Firewalls • Routes	• Addresses • Subnets • Regional MIGs • Regional disks	• Instances • Persistent disks • Machine types • Zonal MIGs

Global, regional, and zonal resources | Compute Engine Documentation | Google Cloud

Networking tiers

Optimize cloud network for performance or price.
GCP is the first major public cloud to offer a tiered cloud network.

Premium
Give users exceptional high performing network experience by using Google's global network.

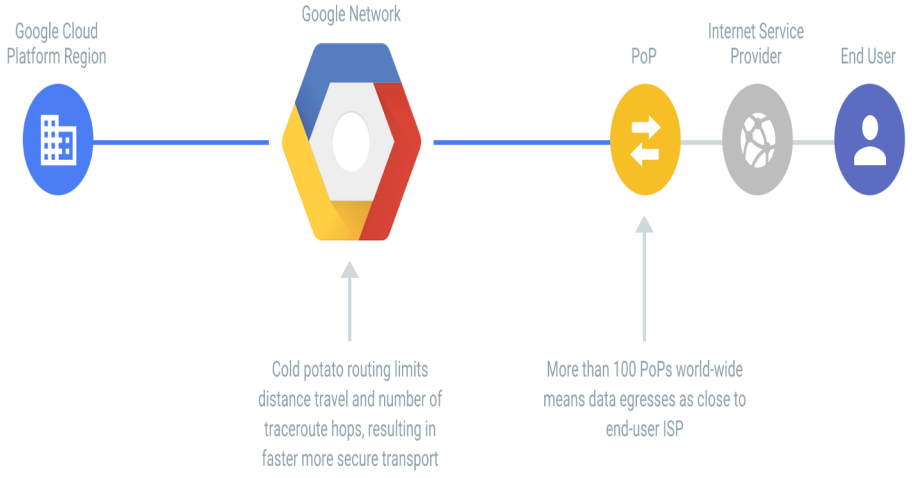
VS

Standard
Get control over network costs, while still delivering performance comparable with other cloud providers.

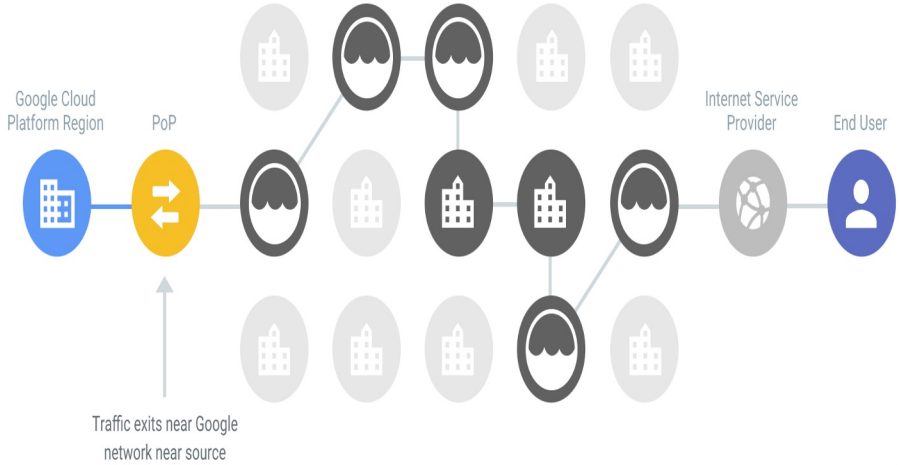
Category	Premium	Standard
Network	High performance routing	Lower performance network
Network Services	Network services such as Cloud Load Balancing are global (single VIP for backends in multiple regions)	Network services such as Cloud Load Balancing are regional (one VIP per region)
Service Level	High performance and reliability	Performance and availability comparable to other public cloud providers (lower than premium)
Use Case	Performance, reliability, global footprint and user experience are your main considerations	Cost is your main consideration, and you're willing to trade-off some network performance

Networking tiers - Cont.

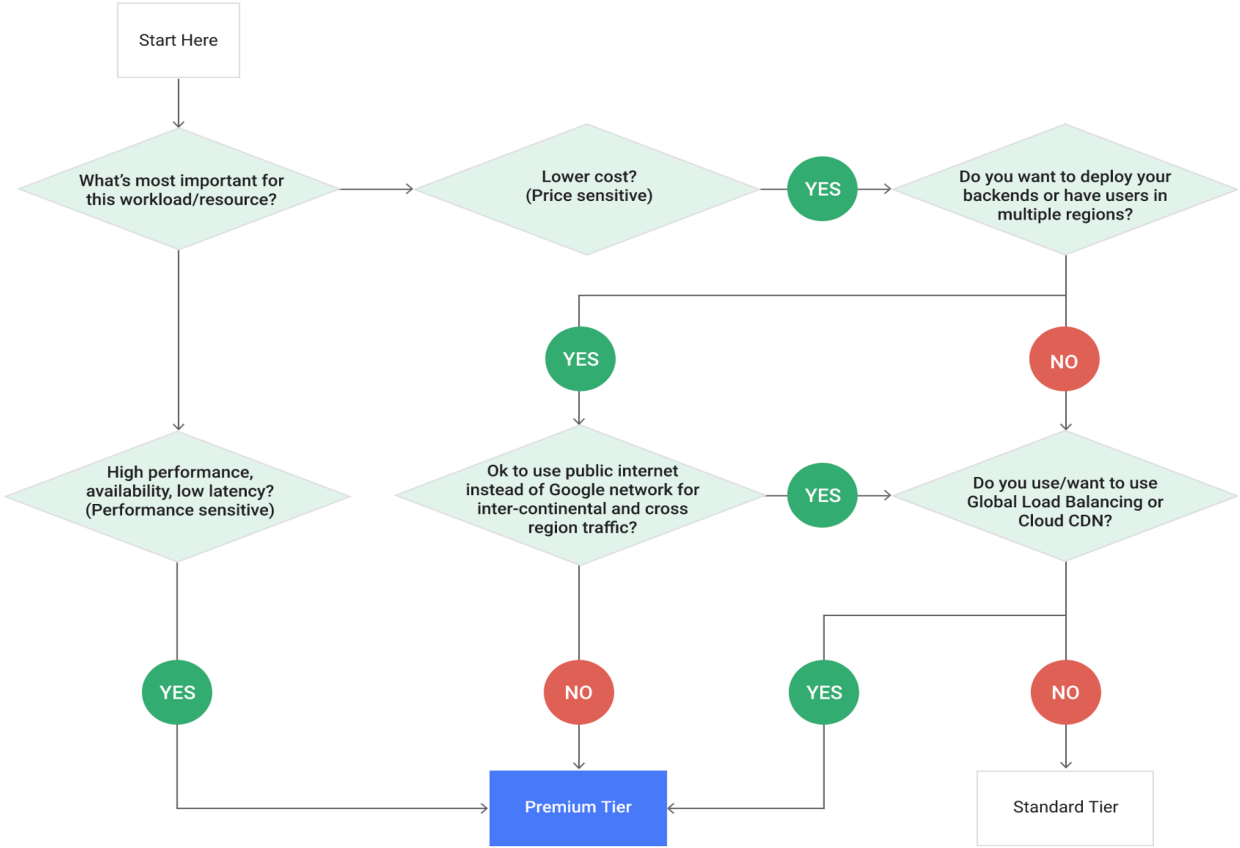
Premium Tier delivers GCP traffic over Google's well-provisioned, low latency, highly reliable global network. This network consists of an extensive global private fiber network with over 100 points of presence (POPs) across the globe. By this measure, Google's network is the largest of any public cloud provider.



Standard Tier delivers GCP traffic over a transit ISP's network with the latency and reliability typical of transit ISPs, and with a network quality comparable to that of other public clouds, at a lower price than our Premium Tier. Also provide only regional network services in Standard tier, such as the new regional Cloud Load Balancing service.



Networking tiers - Cont.

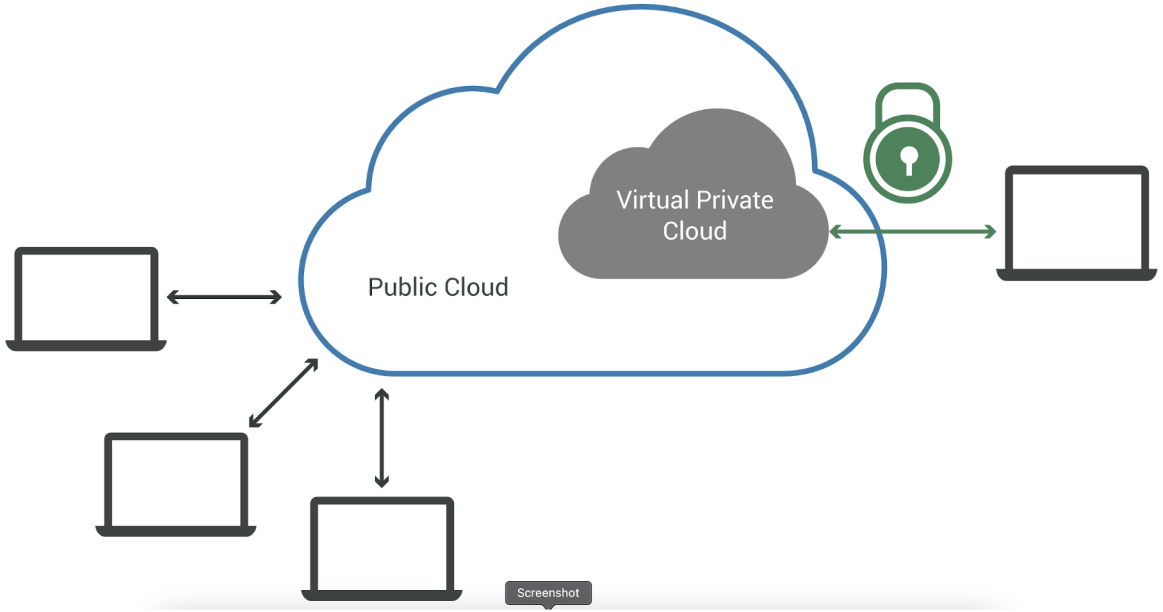


Networking tiers - Cont.

Features	Premium Tier	Standard Tier
Plain VM instance	Yes - Regional	Yes - Regional
HTTP(S) Load Balancing (LB)	Yes - Global only	Yes - Regional
TCP/SSL Proxy LB (non-HTTP traffic)	Yes - Global only	Yes - Regional
Network / Internal LB	Yes - Regional VIP (+ Client can be anywhere)	Yes - Regional VIP (+ Client needs to be in same region)
Google Cloud Storage, Google Kubernetes Engine	Yes	Yes - Regional but only via LB
Cloud CDN	Yes	No
Cloud VPN/Cloud Router	Yes	No

Virtual Private Clouds (VPCs)

A VPC isolates computing resources from the other computing resources available in the public cloud.



Virtual Private Clouds (VPCs)

```
gcloud compute networks create NETWORK \
    --subnet-mode=auto \
    --bgp-routing-mode=DYNAMIC_ROUTING_MODE \
    --mtu=MTU
```

```
gcloud compute networks subnets create SUBNET \
    --network=NETWORK \
    --range=PRIMARY_RANGE \
    --region=REGION
```

[Using VPC networks | Google Cloud]

Firewall rules

Ingress (inbound) rule					
Priority	**Action**	**Enforcement**	**Target (defines the destination)**	**Source**	**Protocols and ports**
Integer from `0` to `65535`, inclusive; default `1000`	`allow` or `deny`	`enabled` (default) or `disabled`	The *target* parameter specifies the destination; it can be one of the following: • All instances in the VPC network • Instances by network tag • Instances by service account	One of the following: • Range of IPv4 addresses; default is any (`0.0.0.0/0`) • Instances by network tag • Instances by service account • Range of IPv4 address *and* instances by network tag • Range of IPv4 address *and* instances by service account	Specify a protocol or a protocol and a destination port. If not set, the rule applies to all protocols and destination ports.

Egress (outbound) rule					
Priority	**Action**	**Enforcement**	**Target (defines the source)**	**Destination**	**Protocols and ports**
Integer from `0` to `65535`, inclusive; default `1000`	`allow` or `deny`	`enabled` (default) or `disabled`	The *target* parameter specifies the source; it can be one of the following: • All instances in the VPC network • Instances by network tag • Instances by service account	Any network or a specific range of IPv4 addresses; default is any (`0.0.0.0/0`)	Specify a protocol or a protocol and a destination port. If not set, the rule applies to all protocols and destination ports.

Firewall rules

EXAMPLES

To create a firewall rule allowing incoming TCP traffic on port 8080, run:

```
$ gcloud compute firewall-rules create FooService --allow=tcp:8080
    --description="Allow incoming traffic on TCP port 8080" --direction=INGRESS
```

To create a firewall rule that allows TCP traffic through port 80 and determines a list of specific IP address blocks that are allowed to make inbound connections, run:

```
$ gcloud compute firewall-rules create "tcp-rule" --allow=tcp:80
    --source-ranges="10.0.0.0/22,10.0.0.0/14" --description="Narrowing TCP traffic"
```

To list existing firewall rules, run:

```
$ gcloud compute firewall-rules list
```

Notes on Firewall rules priorities

- Priority is a numerical value which determines whether the rule is applied. Only the highest priority (lowest priority number) rule whose other components match traffic is applied; conflicting rules with lower priorities are ignored.

- The highest priority rule applicable to a target for a given type of traffic takes precedence. Target specificity does not matter. For example, a higher priority ingress rule for certain destination ports and protocols intended for all targets overrides a similarly defined rule with lower priority for the same destination ports and protocols intended for specific targets.

- A rule with a deny action overrides another with an allow action *only if the two rules have the same priority.* Using relative priorities, it is possible to build allow rules that override deny rules, and deny rules that override allow rules.

Notes on Firewall rules priorities - cont.

Consider the following example where two firewall rules exist:

- An ingress rule from sources `0.0.0.0/0` (anywhere) applicable to all targets, all protocols, and all destination ports, having a `deny` action and a priority of `1000`.
- An ingress rule from sources `0.0.0.0/0` (anywhere) applicable to specific targets with the tag `webserver`, for traffic on TCP 80, with an `allow` action.

The priority of the second rule determines whether TCP traffic to port 80 is allowed for the `webserver` targets:

- If the priority of the second rule is set to a number *greater than* `1000`, it has a *lower* priority, so the first rule denying all traffic applies.
- If the priority of the second rule is set to `1000`, the two rules have identical priorities, so the first rule denying all traffic applies.
- If the priority of the second rule is set to a number *less than* `1000`, it has a *higher* priority, thus allowing traffic on TCP 80 for the `webserver` targets. Absent other rules, the first rule would still deny other types of traffic to the `webserver` targets, and it would also deny all traffic, including TCP 80, to instances *without* the `webserver` tag.

The previous example demonstrates how you can use priorities to create selective `allow` rules and global `deny` rules to implement a security best practice of least privilege.

Google Compute Engine (VMs)

Topics to cover:

- What is Google Compute Engine
- Create and manage VMs from cloud console and using gcloud tool.
- Accessing a VM.
- VMs types.
- Attaching additional disks and GPUs.
- Creating snapshots and custom images.
- Instance groups (IGs)

What is Compute Engine?

Customizable virtual machines in Google Cloud

Compute Engine - cont.

- Create and manage VMs from cloud console and using gcloud tool.
- Accessing a VM.
- VMs types.
- Attaching additional disks and GPUs.
- Creating snapshots and custom images.

gcloud compute instances create | Cloud SDK Documentation

Instance groups (IGs):

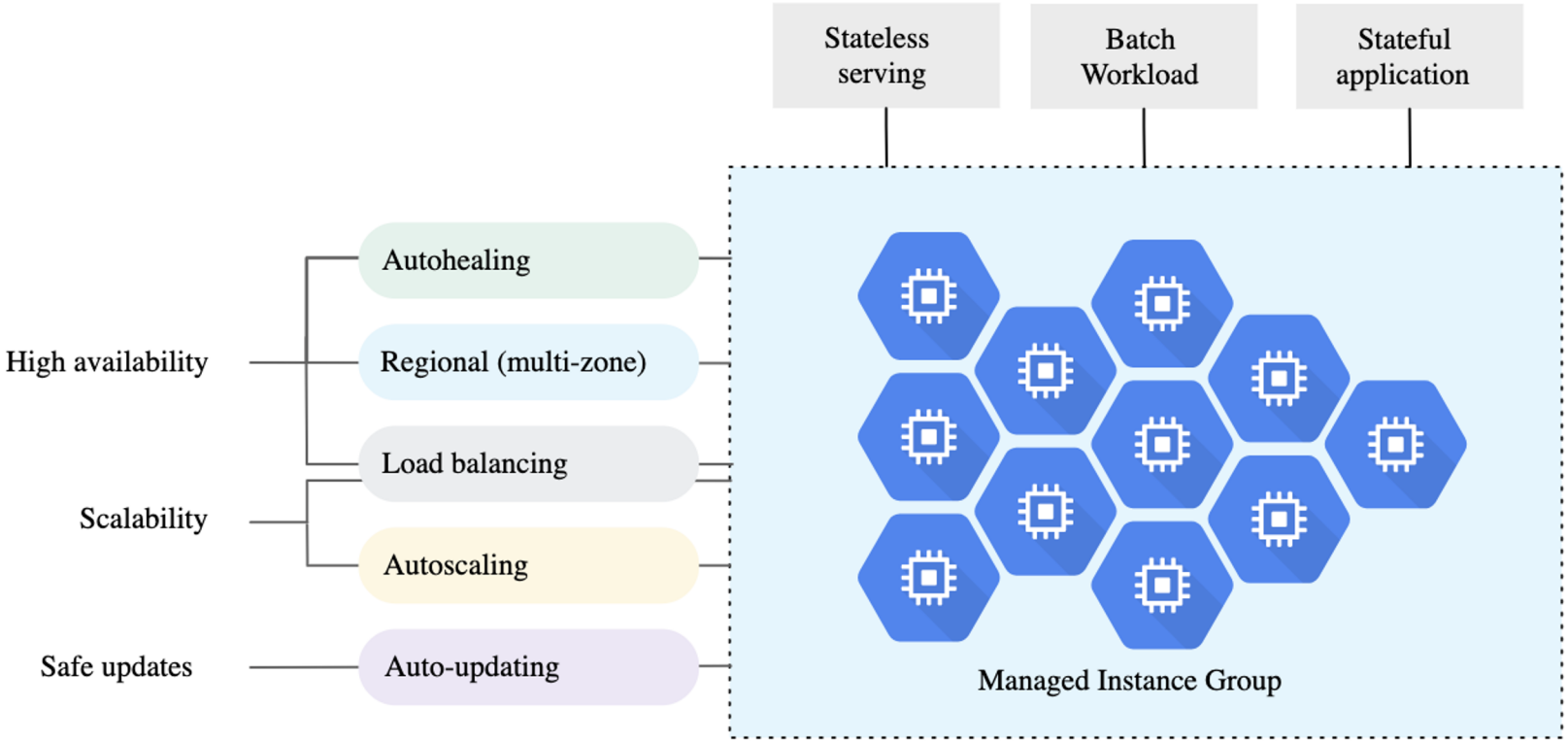
Types of Instance Groups:

- Managed instance groups (MIGs)
- Unmanaged instance groups

`

Creating MIGs - Demo

Managed instance groups (MIGs):



Lab 2.1

1. From Cloud console, create a VPC named “auto-vpc” with auto-mode enabled,
   How many subnets created?
2. From Cloud console, create a VPC named “custom-vpc” with auto-mode disabled and create two subnets.
3. Using gcloud tool list all available VPCs and list subnets of each VPC.
4. Block internet access from you VPC using firewall rules.
5. Create a firewall rule to allow incoming SSH requests from internet to all instances in your vpc.
6. Modify the previous firewall rule to allow only ssh requests coming through Google’s IAP servers.

Lab 2.2

1. Create a VM with public ip then:
   - In two different ways, SSH into this VM.
   - Enforce SSH into this VM to be IAP protected.
2. Create a VM **without** public ip then:
   - SSH into this vm.
   - update system packages (is it possible?)
3. Create a VM **with** public ip then:
   - SSH into this vm
   - Update system packages.
   - Setup Nginx Web Server and test your setup.
   - Create a custom image from this VM named “custom-img-nginx”.
4. Create MIG (min 3 and max 5) of a template using the custom image “custom-img-nginx”.

Networking in GCP (Advanced topics)

Topics to cover:

- Cloud Routing.
- Cloud NAT
- Load Balancing
- Identity Aware Proxy (IAP)
- Cloud DNS
- Virtual Private Network (VPN)
- VPC Service Controls (VPC-SC)
- VPC Peering, Shared VPC and Cloud Interconnect

Routing in GCP

Every VPC network uses a scalable, distributed virtual routing mechanism. There is no physical device that's assigned to the network. Some routes can be applied selectively, but the routing table for a VPC network is defined at the VPC network level.

Each VM instance has a controller that is kept informed of all applicable routes from the network's routing table. Each packet leaving a VM is delivered to the appropriate next hop of an applicable route based on a routing order. When you add or delete a route, the set of changes is propagated to the VM controllers by using an eventually consistent design.

Routes overview | VPC | Google Cloud

Routes types

- System-generated routes:
  - Default routes
  - Subnet Routes
- Custom Routes:
  - Static routes
  - Dynamic routes
- Peering routes:
  - Peering subnet routes
  - Peering custom routes.

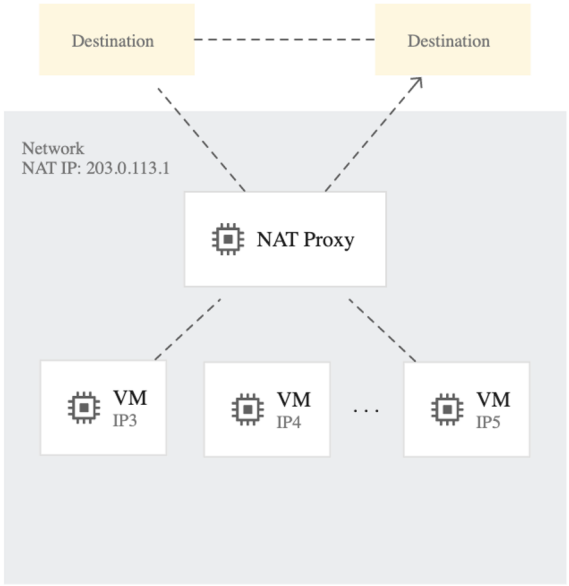
https://cloud.google.com/vpc/docs/routes#types_of_routes

Cloud Router

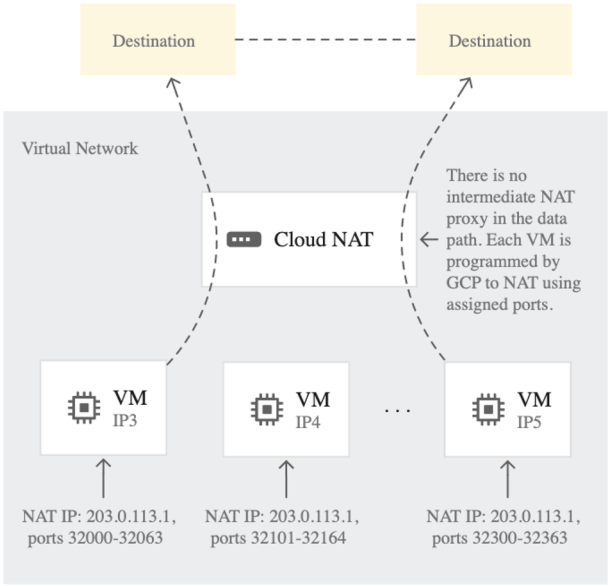
- Cloud Router is used to dynamically exchange routes between two VPCs or between VPC and on-premises networks.
- Cloud Router uses “Border Gateway Protocol” to exchange routing information between the networks.
- A Cloud Router also serves as the control plane for Cloud NAT.
- A Cloud Router is used with services like:
  - HA VPN
  - Classic VPN if your on-premises VPN gateway supports BGP.
  - Dedicated Interconnect & Partner Interconnect
- Direct Peering and Carrier Peering do not use Cloud Routers.

Cloud Router overview

Cloud NAT



1. Typical NAT Proxies



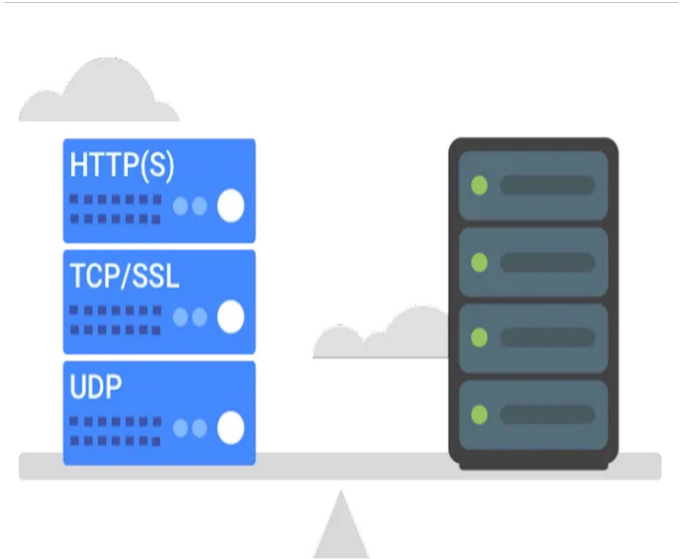
2. Google Cloud NAT

[Cloud NAT overview]

[Protect Your Network with Cloud NAT]

Cloud Load Balancing

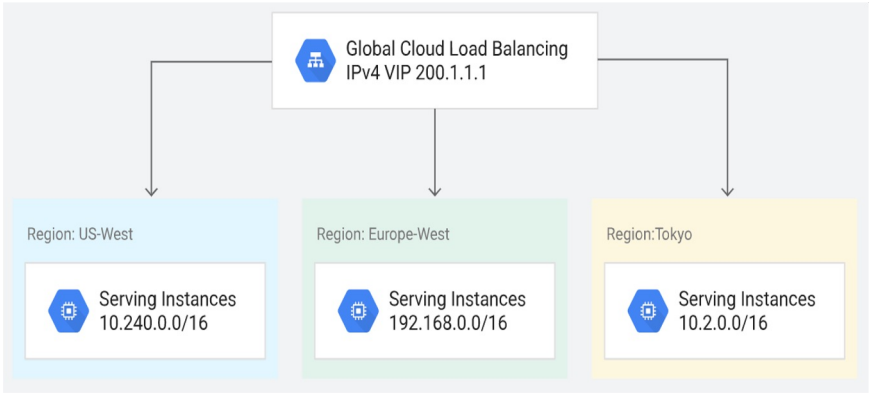
High performance, scalable load balancing on Google Cloud Platform.



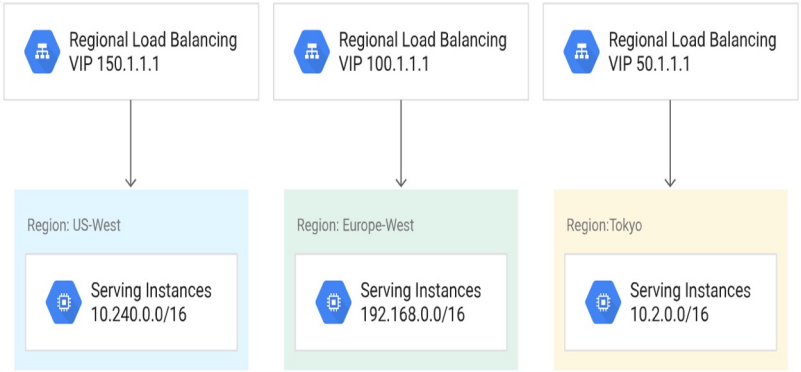
[Cloud Load Balancing documentation]

Cloud Load Balancing - Cont.

Premium Networking Tier



Standard Networking Tier



Cloud Load Balancing - Cont.

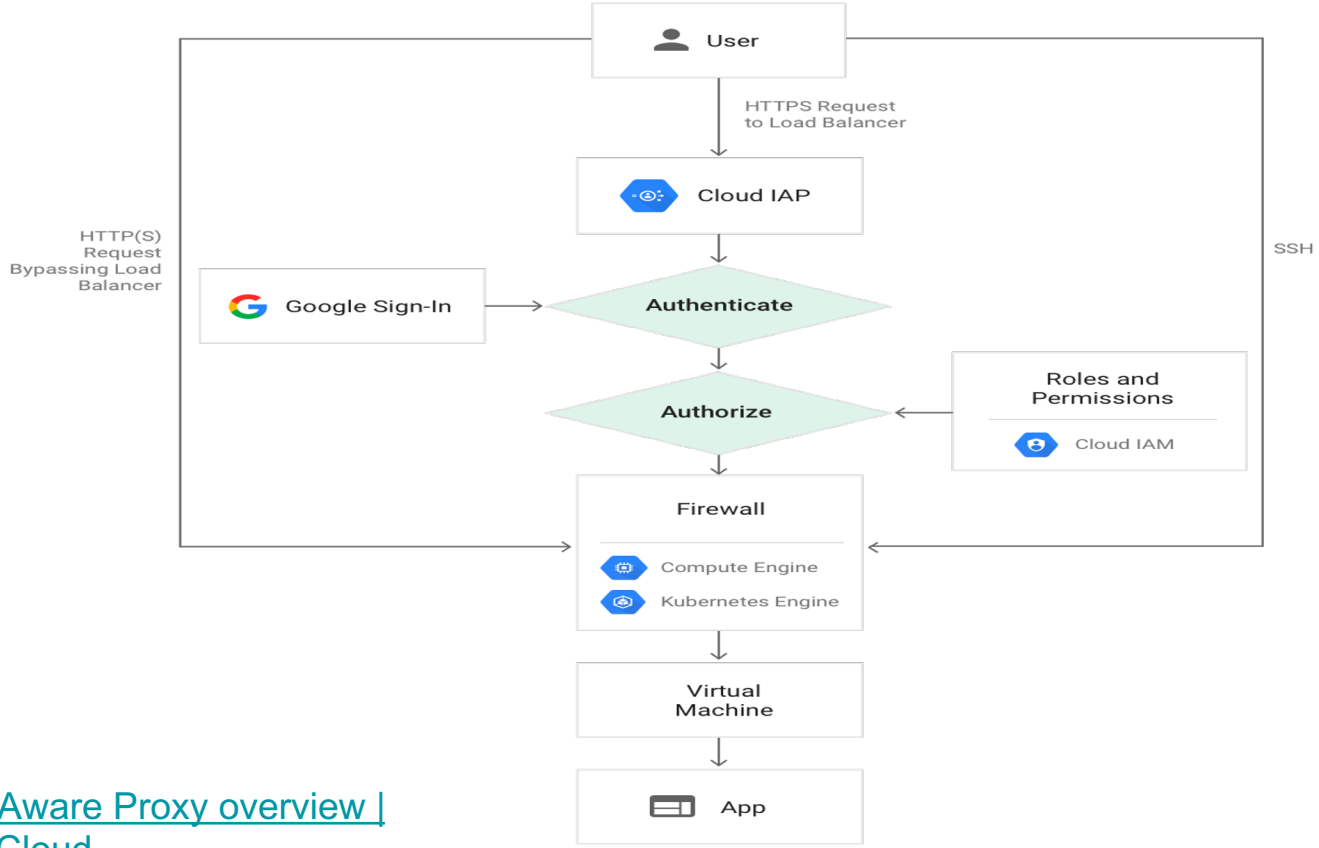
Load balancer types:

- External HTTP(S) Load Balancing (global and regional modes)
- Internal HTTP(S) Load Balancing
- External TCP/UDP Network Load Balancing
- Internal TCP/UDP Load Balancing
- SSL Proxy Load Balancing
- TCP Proxy Load Balancing

Choosing a load balancer

Decision tree for choosing a load balancer

Identity Aware Proxy (IAP)



[Identity-Aware Proxy overview | Google Cloud]

Lab 3.1

1. In previous lab you created the VPC “auto-vpc”, How many routes created for this VPC? Can you delete any of these routes?
2. In previous lab you created a VPC named “custom-vpc” How many routes created for this VPC?
3. How would you block internet access from you vpc using routes?
4. Add a NAT gateway on any of the subnets in your VPC.

Lab 3.2

1. In previous lab you created an MIG of a template using the custom image “custom-img-nginx”, Create a Global (or Regional) HTTP Load balancer to access your MIGs Nginx setup.
2. Try to configure IAP at the load balancer level to protect your ingress access. Is it possible to have IAP enabled for HTTP resources?

Google Kubernetes Engine (GKE)

Topics to cover:

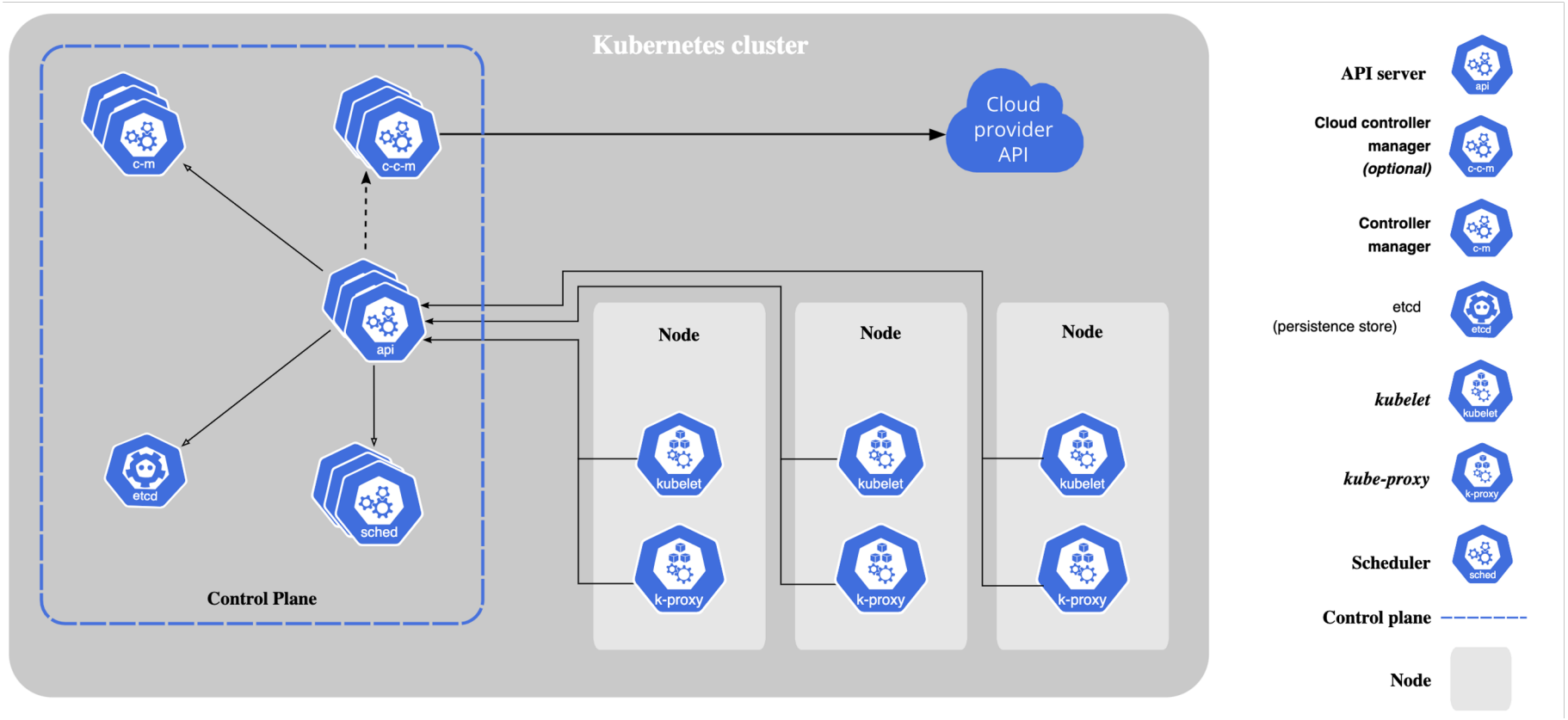
- What is GKE?
- Create and Manage GKE clusters.
- Installing kubectl and connect to GKE cluster
- Viewing GKE cluster state.
- Monitoring deployments on GKE.

What is GKE? (Cont.)

GKE is a managed service for deploying kubernetes clusters with the benefit of:

- Google Cloud load-balancing for Compute Engine instances
- Node pools to designate subsets of nodes within a cluster for additional flexibility
- Automatic scaling of your cluster's node instance count
- Automatic upgrades for your cluster's node software
- Node auto-repair to maintain node health and availability
- Logging and monitoring with Google Cloud's operations suite for visibility into your cluster

What is GKE?


Kubernetes cluster
c-m
c-c-m
Cloud provider API
api
etcd
sched
Control Plane
Node
kubelet
k-proxy
Node
kubelet
k-proxy
Node
kubelet
k-proxy
API server
Cloud controller manager (optional)
Controller manager
etcd (persistence store)
kubelet
kube-proxy
Scheduler
Control plane
Node

What is GKE? (Cont.)

Types of GKE clusters:

- Zonal clusters (Single-Zone Or Multi-Zone):
  - A single-zone cluster has a single control plane running in one zone. This control plane manages workloads on nodes running in the same zone.
  - A multi-zonal cluster has a single replica of the control plane running in a single zone, and has nodes running in multiple zones.
- Regional clusters:
  - A regional cluster has multiple replicas of the control plane, running in multiple zones within a given region. Nodes also run in each zone where a replica of the control plane runs.

What is GKE? (Cont.)

Modes of GKE clusters:

- Standard mode:
  You manage the cluster's underlying infrastructure, giving you node configuration flexibility

- Autopilot mode:
  GKE provisions and manages the cluster's underlying infrastructure, including nodes and node pools, giving you an optimized cluster with a hands-off experience.

GKE Standard vs Autopilot

Introducing GKE Autopilot

Create and Manage GKE clusters:

Create a zonal cluster:

```
gcloud container clusters create example-cluster \
    --zone us-central1-a \
    --node-locations us-central1-a,us-central1-b,us-central1-c
```

Create a regional cluster:

```
gcloud container clusters create my-regional-cluster --region us-west1
```

- how-to: create zonal cluster
- how-to: create regional cluster

Installing kubectl and connect to GKE cluster:

1- Install kubectl using gcloud tool:

```
→  .ssh gcloud components install kubectl
```

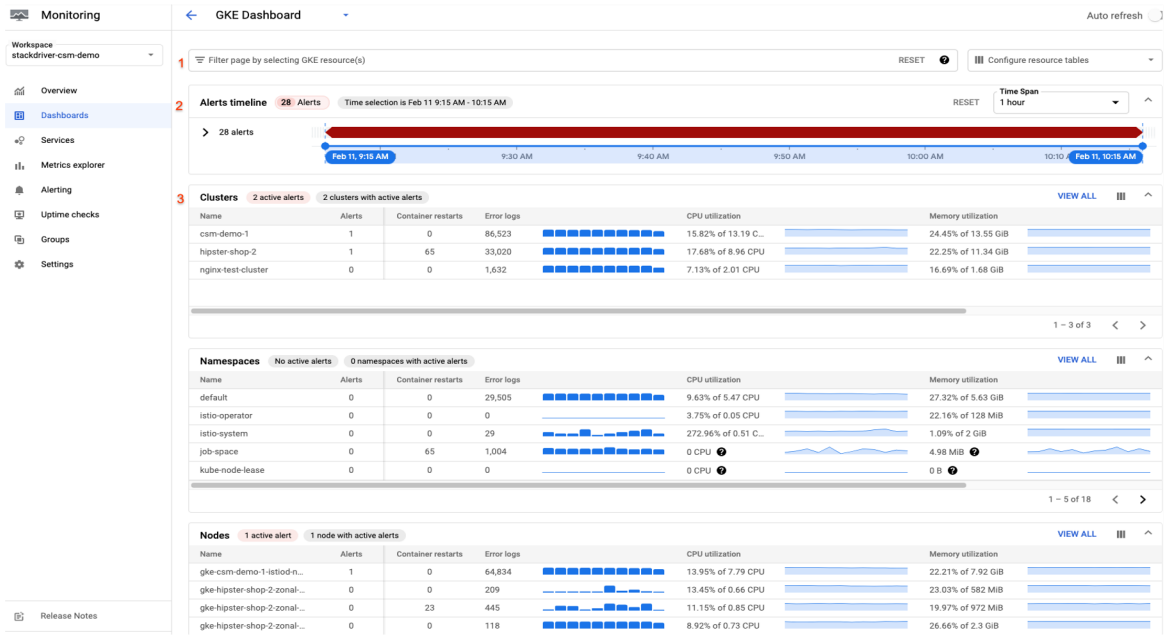
2- Connect to the cluster:

Generating a ”kubeconfig” entry

```
gcloud container clusters get-credentials cluster-name
```

[gcloud components install | Cloud SDK Documentation]

Viewing GKE cluster stats and Monitoring your deployments.



[Observing your GKE clusters | Operations Suite | Google Cloud]

Lab 3.3

1. Create a **private** standard GKE cluster.
2. Deploy Nginx as a deployment using latest Nginx docker image on Docker Hub.
3. Expose your Nginx deployment using Kubernetes LoadBalancer Service.
4. What is the type of GCP Load Balancer that is created for your LB service?
5. Use kubectl to view container logs.
6. Use cloud logging service to view container logs. [hint: search about cloud logging service for gke]
7. (Bonus) setup a HTTP load balancer for your deployment using the kubernetes ingress resource. (hint: link)
8. Create an autopilot GKE cluster with public control plane.
9. Enforce the cluster's control plane to accept only connections from your local machine.

Storage

Google Cloud storage and database portfolio

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

Google Cloud storage and database - Cont.

Some services scale horizontally by adding nodes.

- Bigtable
- Spanner

Some services scale vertically by making machines larger.

- Cloud SQL
- Memorystore

Some services scale automatically with no limits.

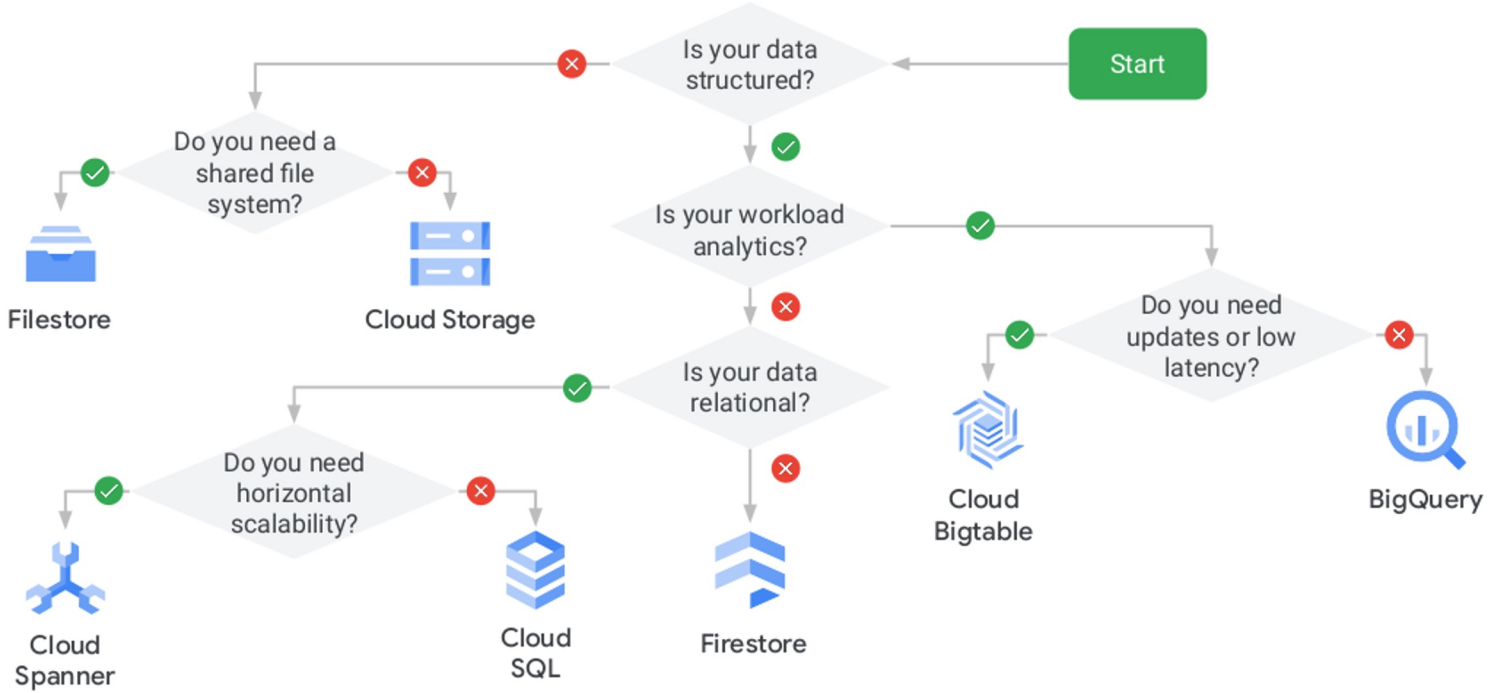
- Cloud Storage
- BigQuery
- Firestore

Google Cloud storage and database - Cont.

Preventing data loss is a shared responsibility.

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

Google Cloud storage and database - Cont.



Transferring data into Google Cloud

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

Transferring data into Google Cloud - Cont.

Google provides two types of services for storage transfer:

1. Online Data transfer: **Storage Transfer Service**, Import online data to Cloud Storage:
   - Amazon S3
   - HTTP/HTTPS Location
   - Transfer data between Cloud Storage buckets
   - On-prem server

1. Offline Data transfer: **Transfer Appliance**: Rackable device up to 1PB shipped to Google.

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

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

Lab 3.2

1. Using gsutil:
   - Create 3 buckets.
   - Enable Versioning for them.
   - Upload a file into bucket-1 then copy it from bucket-1 into bucket-2 & bucket-3.
   - Delete the file from bucket-1
2. Host a static website on a standard public GCS bucket [hint: link].
3. Deploy MySQL private instance and connect to it then create a new database.

More GCP Services

Topics to cover

- Cloud Container Registry (GCR)
- Cloud Artifact Registry
- Cloud Functions
- Cloud Run
- App Engine
- Cloud StackDriver (Cloud logging, Cloud Monitoring, etc)

Lab 3.3

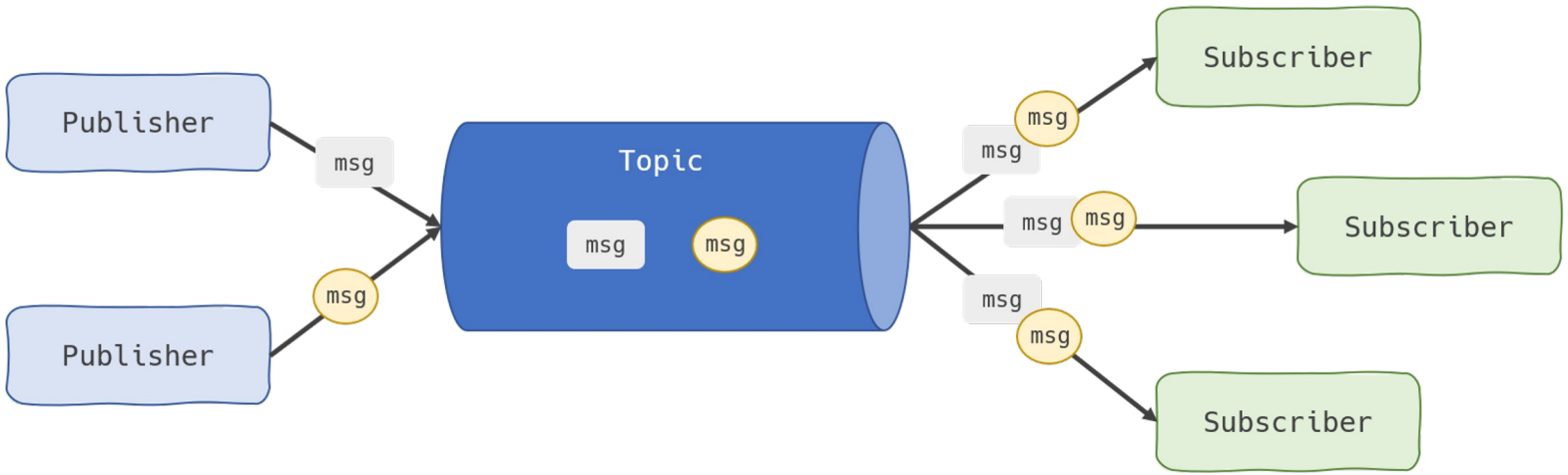
1. Using gcloud & Docker:
   - Configure Docker & gcloud to work with GCR of your project. [hint: link]
   - Push Nginx docker image to GCR (make the image private).
   - Pull this image into a k8s setup or on a VM (hint: attach a SA on ur vm or gke with correct iam role).
2. Using Cloud Functions:
   - Create a Function that runs whenever a file is uploaded to a cloud storage bucket. [hint: link]
3. Using Cloud Run:
   - Run a pre-built docker image (pulled from GCR) [hint: link]
   - Build and Run any sample app [hint: link]
4. Using App Engine:
   - Run the sample hello-world python app [link]

GCP Big Data Services

Topics to cover

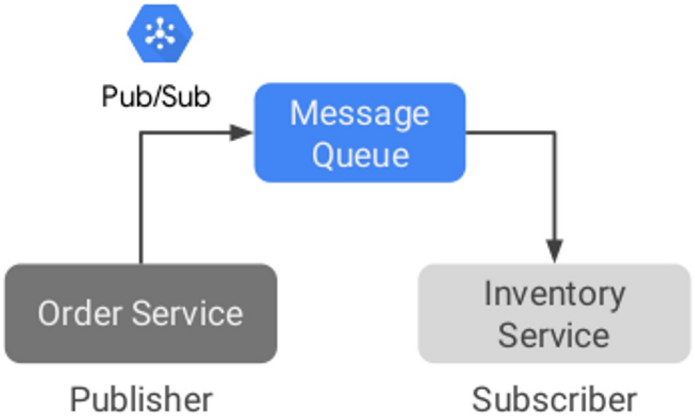
- Cloud Pub/Sub
- Cloud Dataproc
- Cloud Dataflow
- Cloud Composer
- Cloud Data Fusion
- Cloud BigQuery

Cloud Pub/Sub

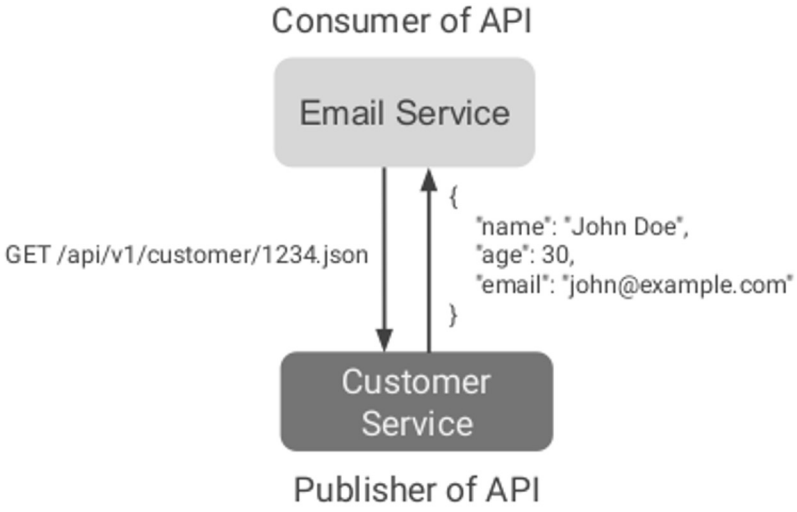


Cloud Pub/Sub - Cont.

Publishers and subscribers are loosely coupled



Consumers of HTTP APIs should bind loosely with publisher payloads

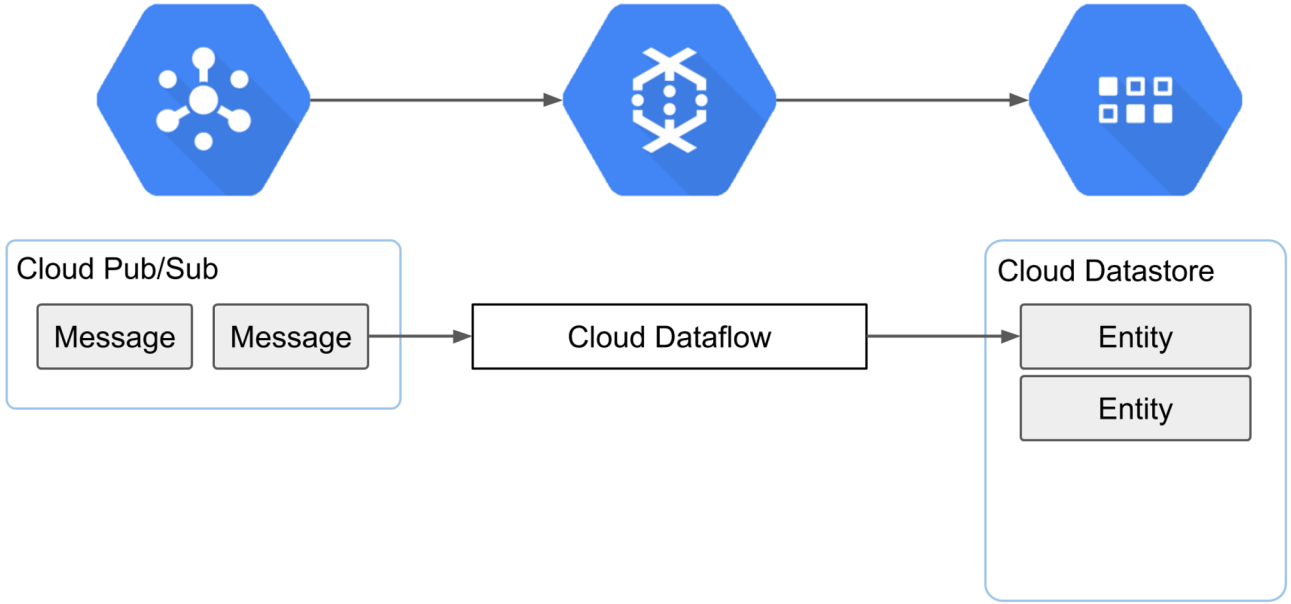


Cloud Dataproc

Managed Spark and Hadoop cluster

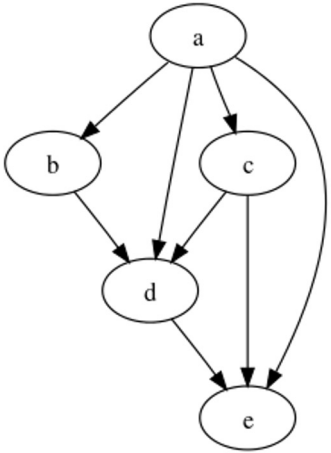
Cloud Dataflow

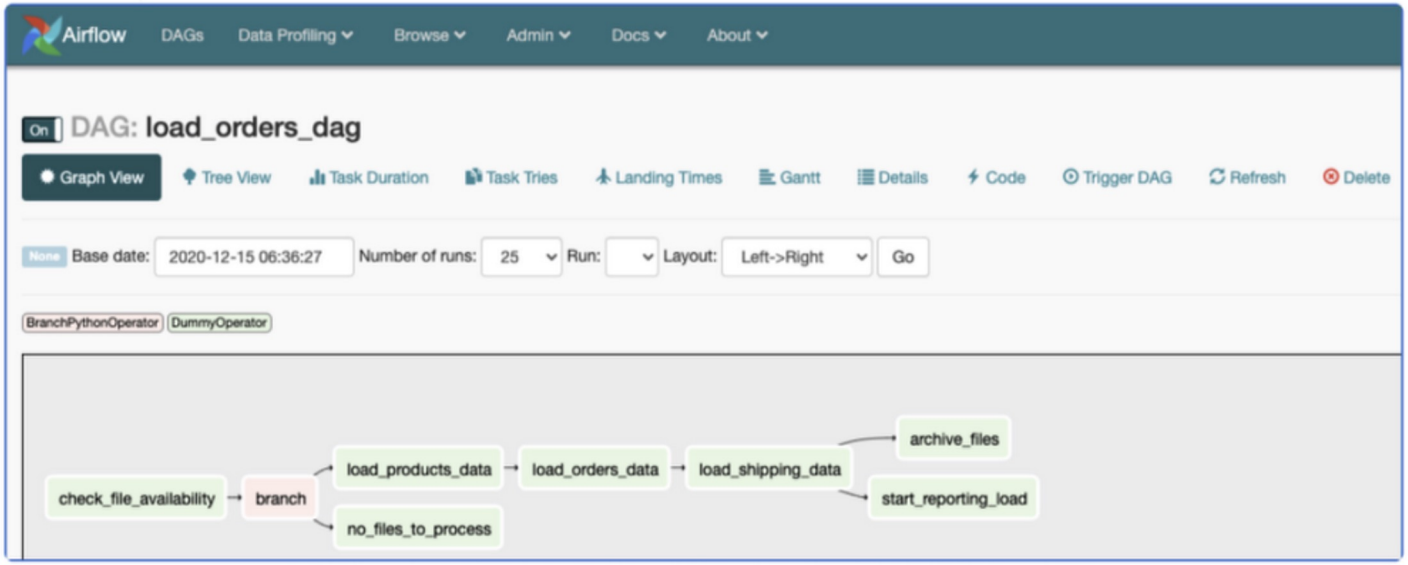
Stream/batch data processing (Based on Apache Beam)



Cloud Composer

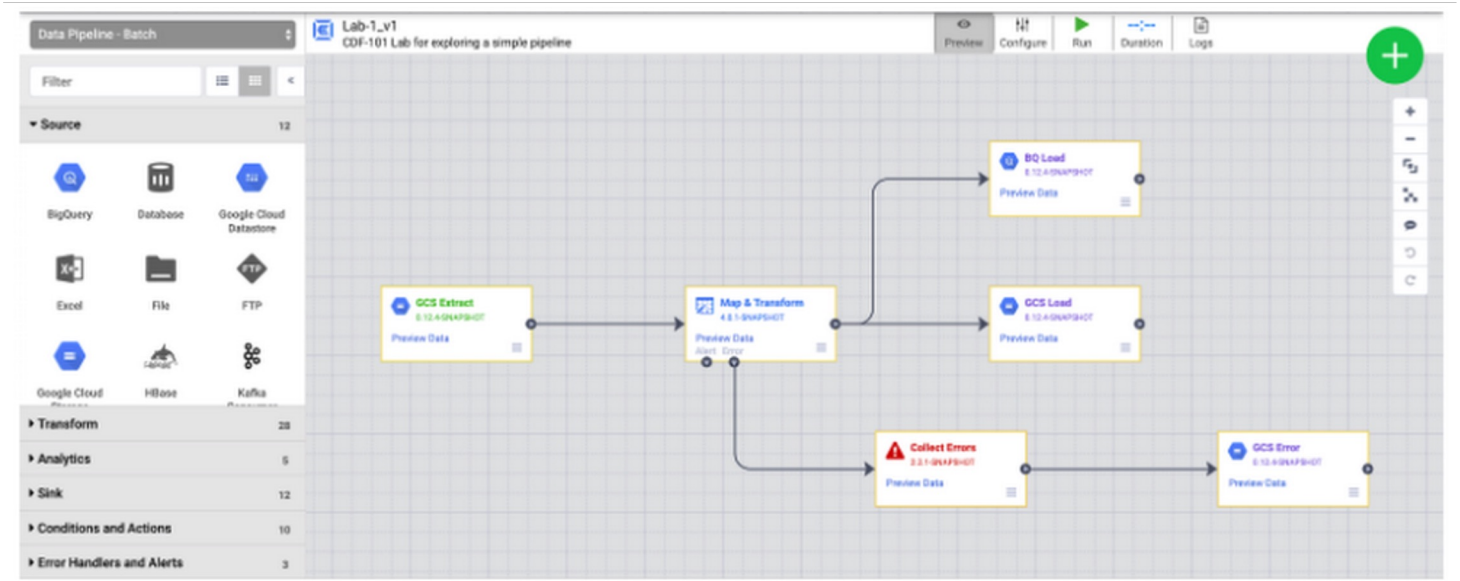
- Managed workflow orchestration service based on Apache AirFlow
- Fully integrated with many other GCP services
- Workflows in Airflow are represented in the form of a Direct Acyclic Graph (DAG)
- A DAG is simply a set of tasks that needs to be performed



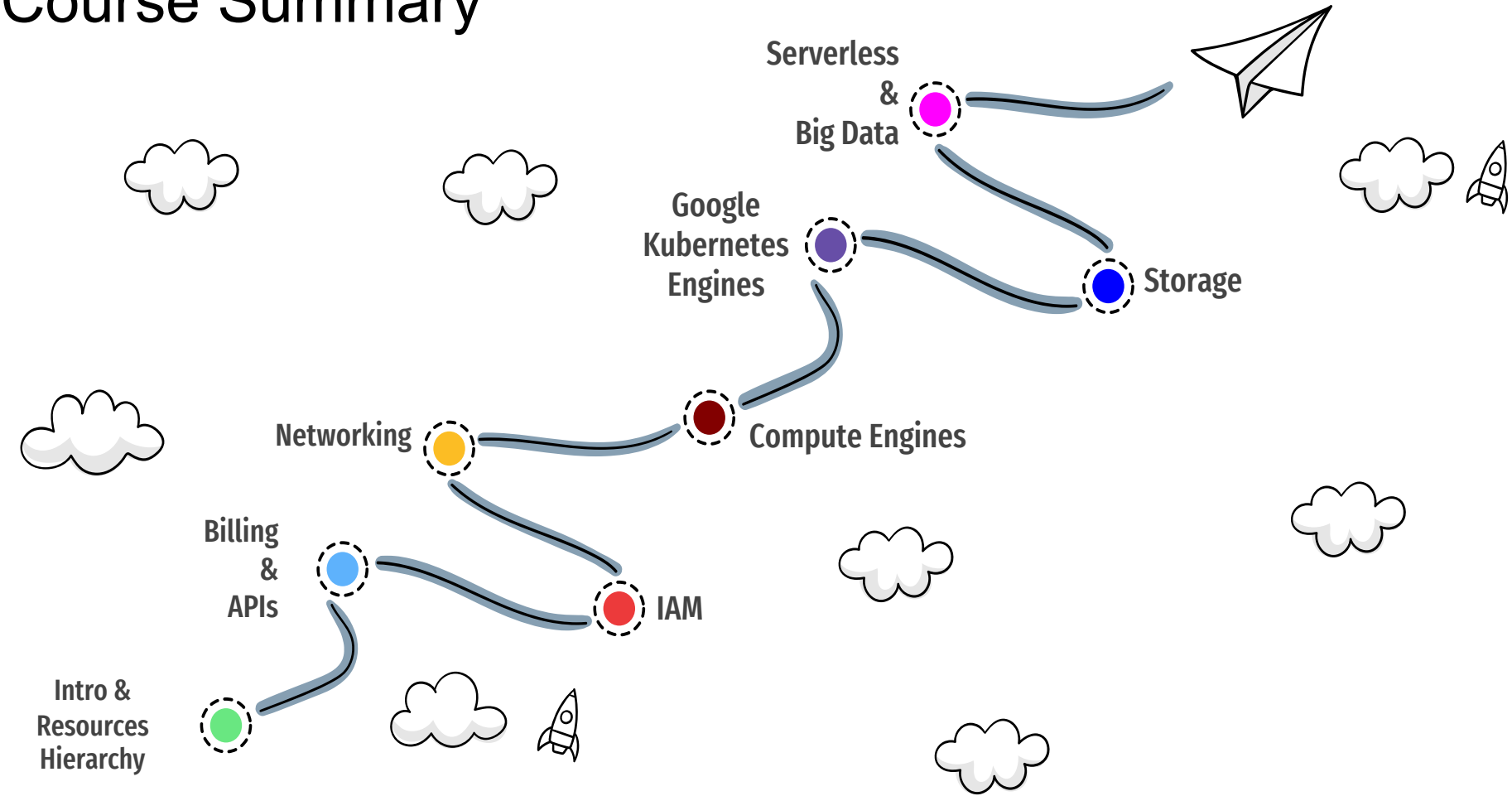


Cloud DataFusion

- GUI based data integration service for building and managing data pipelines (based on CDAP)
- Provides a wide variety of out of the box connectors to sources on GCP, other public clouds and on-premise sources



Course Summary



Resources:

- GCP services:
  - GCP regions and zones
  - GCP description of services
  - GCP list of products
  - AWS vs Azure vs GCP
  - GCP services explained in 4 words
  - Comparison of resources hierarchy between AWS & GCP
- Anthos:
  - https://www.youtube.com/watch?v=Qtwt7QcW4J8
- Study Resources:
  - Google Cloud Platform Fundamentals: Core Infrastructure (Recommended for beginners)
  - https://www.coursera.org/professional-certificates/cloud-engineering-gcp
  - Google Cloud Tech – YouTube channel

Resources (cont.):

- Hands-on Labs and quests:
  - Qwiklabs & Qwiklabs-free
- More study resources:
  - A Cloud Guru – GCP ACE
  - Dan Sullivan - GCP ACE Prep. Course
  - GCP official study guide
- GCP professional certifications
- Google Cloud Platform Podcast
- Best practices for Compute Engine regions selection

Thank You