GCP

Google Cloud

Professional Cloud Developer

© ANKIT MISTRY – GOOGLE CLOUD

Google Certified Professional Cloud Developer

© ANKIT MISTRY – GOOGLE CLOUD

Professional Cloud Developer

- Pay attention for 5 minutes, before we dive in.
- Challenging certification, and course is long so have patience.
- Good to have basic IT skill, but I will start from scratch in GCP
- Learn by Doing
- 100+ Hands-on Lab

© ANKIT MISTRY – GOOGLE CLOUD

GCP certifications

https://cloud.google.com/certification/cloud-developer

© ANKIT MISTRY – GOOGLE CLOUD

Cloud Cost for this course

- $0 – for GCP account
- GCP Free trial
- $300 for next 3 months https://cloud.google.com/free
- Length: Two hours
- Registration fee: $200 (plus tax where applicable)
- Languages: English, Japanese
- Exam format: Multiple choice and multiple select,

© ANKIT MISTRY – GOOGLE CLOUD

Udemy tips

BY ANKIT MISTRY

© ANKIT MISTRY – GOOGLE CLOUD

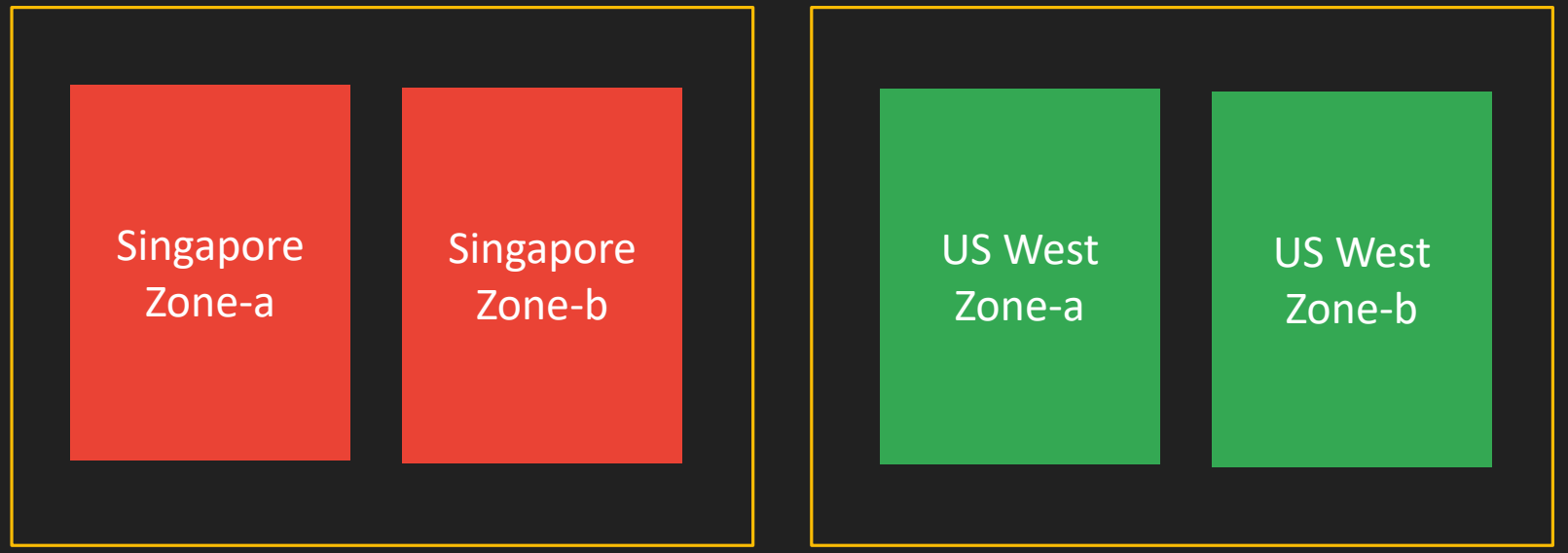
Getting Started with GCP

- Google Cloud Overview
- Setup Free Trial GCP Account
- GCP Regions & Zones

© ANKIT MISTRY – GOOGLE CLOUD

Zones & Regions

- Low latency
- Follow Government rules
- High availability
- Disaster recovery



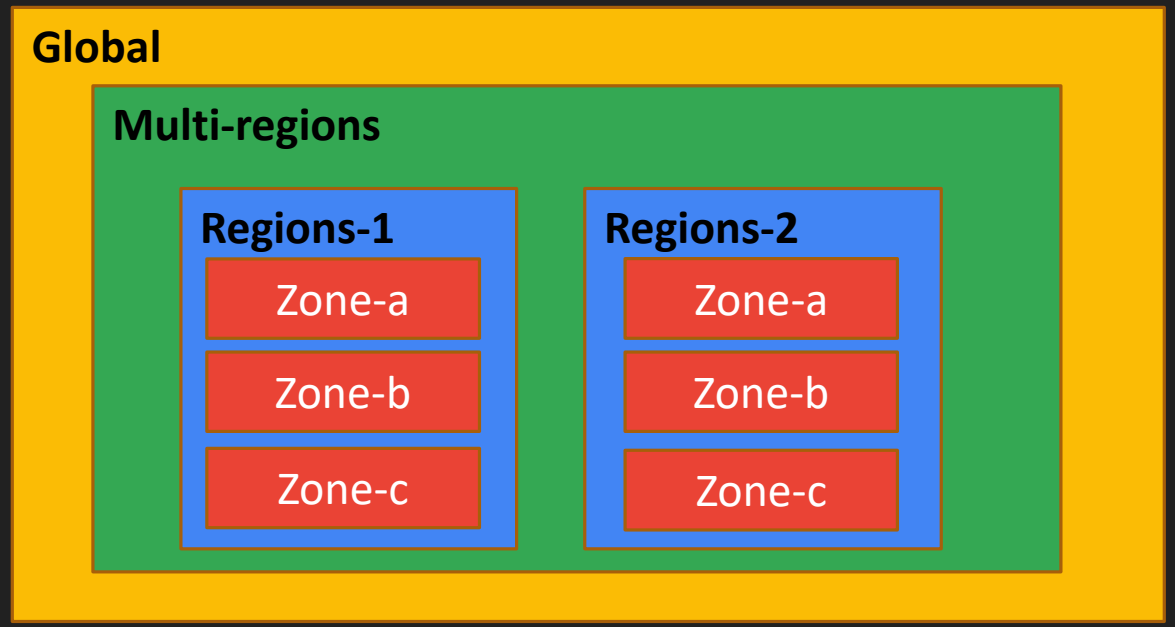
© ANKIT MISTRY – GOOGLE CLOUD

GCP (Zones & Region)

Fascinating Number: Google Is Now 40% Of The Internet (forbes.com)

- Zones – Independent data Center
- Region – Geographical area
- Multi-region : Collection of Geographical
- Global - Anywhere

Global Locations - Regions & Zones | Google Cloud



© ANKIT MISTRY – GOOGLE CLOUD

© ANKIT MISTRY – GOOGLE CLOUD

Different way to access GCP

- Web Console
- CLI – Command line interface
- Language SDK – Python, Java, Go, Node …
- Mobile

© ANKIT MISTRY – GOOGLE CLOUD

Console & Cloud Shell

- Web Console walkthrough
- Cloud Shell
  - Free of cost linux Machine
- Cloud SDK – Python code demo

© ANKIT MISTRY – GOOGLE CLOUD

Google Cloud Shell

- Linux based virtual machine
- Easiest way to get connect with cloud using Command Line
- 5 GB of Persistence storage – at $HOME
- Cloud Shell is ephemeral
- Free of cost (But for limited duration)
- Attach with Google Cloud account – not with Project
- By default fully authenticated & secure
- Pre installed – lots of utilities
  - Cloud SDK already installed
  - gcloud, Python, Docker, Node, Java, MySQL & many more
  - Other utilities like bq, kubectl, gsutil,cbt
- Built-in Editor
- Web Preview
- File management
- Boost mode
- Install additional tool
- Repair Cloud Shell

© ANKIT MISTRY – GOOGLE CLOUD

Mobile APP

BY ANKIT MISTRY

© ANKIT MISTRY – GOOGLE CLOUD

Create Project

BY ANKIT MISTRY

© ANKIT MISTRY – GOOGLE CLOUD

© ANKIT MISTRY – GOOGLE CLOUD

Google Cloud SDK

- Libraries and tools for interacting with Google Cloud products and services.
- Cloud SDK by default installed on Cloud Shell
- You can use Cloud SDK locally or from Cloud Shell (No Installation)
- SDK contains
  - SDK Client Libraries for popular programming languages
  - Google Cloud Command Line Interface (gcloud CLI)
  - Other utilities – bq, gsutil, kubectl
- Free of cost
- Installation of Cloud SDK - Local

© ANKIT MISTRY – GOOGLE CLOUD

gcloud – command line tool

- Basics of gcloud command
  - Command groups
  - root commands
  - Global Flags
- gcloud configuration
- Initialize & authenticate
  - gcloud init, gcloud auth
- Manage Google Cloud SDK
- alpha & beta channel command
- other command line tools – bq, gsutil, kubectl

gcloud GROUP SUBGROUP ACTION FLAGS

gcloud compute instances list
gcloud compute regions list

© ANKIT MISTRY – GOOGLE CLOUD

© ANKIT MISTRY – GOOGLE CLOUD

Google Cloud APIs

- Cloud APIs allow you to automate your workflows by using your favorite language.
- Use these Cloud APIs with REST calls or client libraries in popular programming languages.
- By Default not all API are enabled.
- You can enable/disable API
- There are APIs by default enabled
  - Visit : https://cloud.google.com/service-usage/docs/enabled-service#default
- For Quick check – use API Explorer
- How to access API
  - Raw Rest based call or grpc
  - Programming language SDK

© ANKIT MISTRY – GOOGLE CLOUD

Google Cloud Python SDK

- Google Cloud Client Libraries help us to make programmatic call to Google Cloud
- Programming language Support
  - Python, JAVA, go, node
- Demo :
  - create bucket with python code
  - make sure Cloud SDK is installed
  - Assign proper role to service account
  - set env variable GOOGLE_APPLICATION_CREDENTIALS = service account keys
  - install google cloud storage python lib
    - pip3 install --upgrade google-cloud-storage

© ANKIT MISTRY – GOOGLE CLOUD

© ANKIT MISTRY – GOOGLE CLOUD

Google Cloud billing

- Billing accounts & Dashboard overview
- Budgets & billing alerts
- Exports Billing Data
- Pricing calculator - estimate pricing
- Resource Quota

© ANKIT MISTRY – GOOGLE CLOUD

Billing Account

- Account which pay for resources
- Multiple billing account can be created
- Each Project must be attached with one Billing account
- Without billing account one can not provision resource inside Projects
- Billing account has one or more than one project associated
- Each billing account has single invoice (1 project or Multiple Project)
- With GCP account creation
  - Default is “My Billing account” & attached Default Project – My First Project
- Let’s see in action

© ANKIT MISTRY – GOOGLE CLOUD

GCP Pricing Calculator

BY ANKIT MISTRY

© ANKIT MISTRY – GOOGLE CLOUD

Resource Quota

- Quota defined resource Limits
- Quota defined at
  - Regional
  - Global
- You can increase Quota Limit
  - Raise Ticket, but beyond that there are some hard limit
- Let’s see in action

© ANKIT MISTRY – GOOGLE CLOUD

© ANKIT MISTRY – GOOGLE CLOUD

Resource Management

BY ANKIT MISTRY

© ANKIT MISTRY – GOOGLE CLOUD

Section Overview

- Resource Hierarchy
  - Resources, Project, Folder & Organization
- Create Project/Folder – with Org/No Org Node

© ANKIT MISTRY – GOOGLE CLOUD

Resource Hierarchy

Resources

Project

Folder

Organization

© ANKIT MISTRY – GOOGLE CLOUD

Resources

- Anything you provision/create inside GCP is resource
  - VM, Bucket, Firewall rule, Notebook instance
- Resource can be
  - Zonal
  - Regional – Multi zonal
  - Dual Region
  - Multi Region
  - Global

© ANKIT MISTRY – GOOGLE CLOUD

Projects

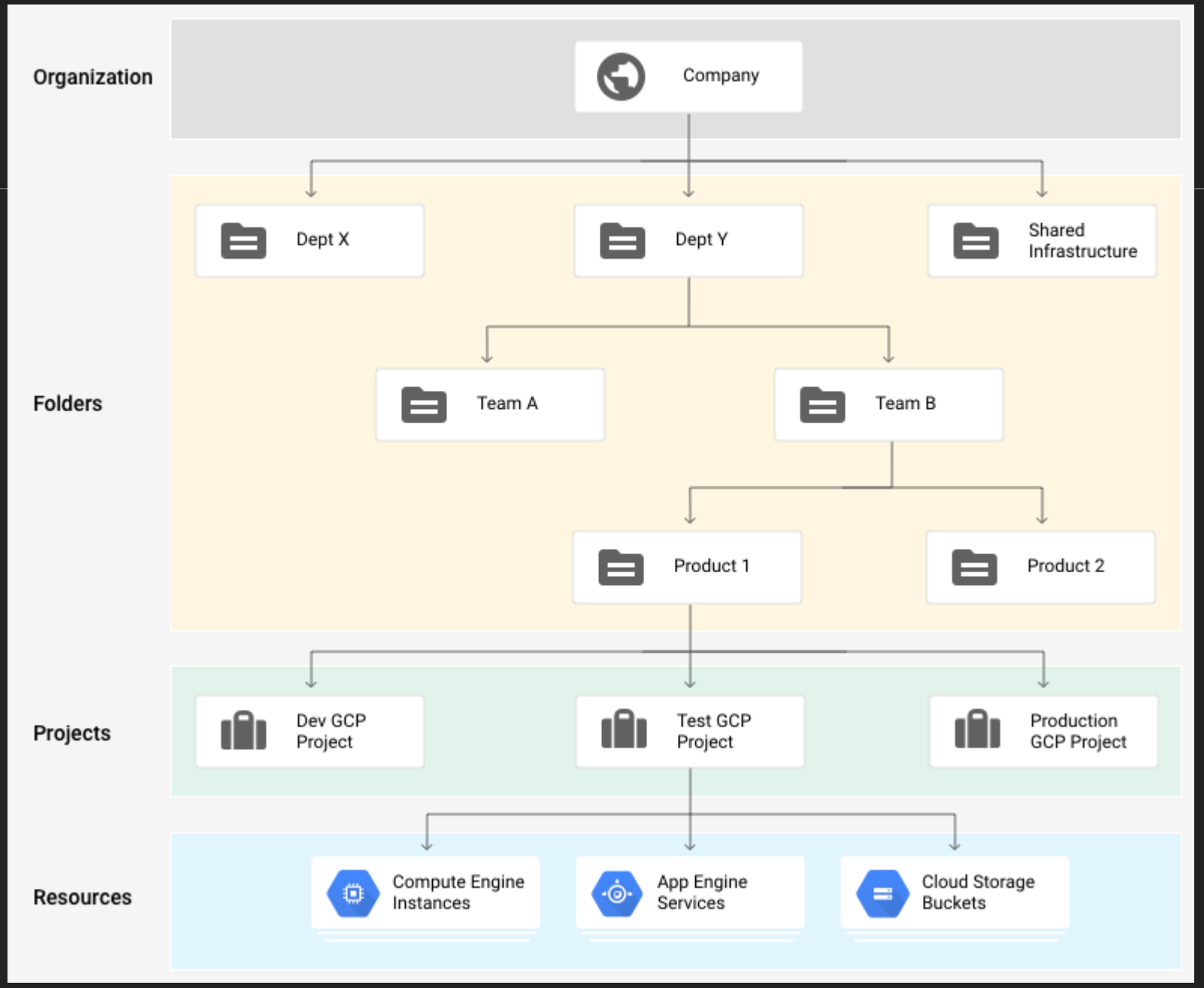
- Container for Resources
- All the resource created must be under some project
- Without Project, No resource can be created
- Project creation is Free
- With GCP account creation – “My First Project” is default Project created
- Project attributes
  - Project name – can be changed
  - Project ID – Can not be changed
  - Project number

© ANKIT MISTRY – GOOGLE CLOUD

Folders & Organization

- Folders Container for Project
- Group Multiple project in same Folder
- Organization is root node

© ANKIT MISTRY – GOOGLE CLOUD



© ANKIT MISTRY – GOOGLE CLOUD

[Hands-on] Create Projects, Folder

BY ANKIT MISTRY

© ANKIT MISTRY – GOOGLE CLOUD

© ANKIT MISTRY – GOOGLE CLOUD

Google Cloud IAM

BY ANKIT MISTRY

© ANKIT MISTRY – GOOGLE CLOUD

Section Overview

- Define What IAM is
- Identity – members
- Service account
- Roles, permission
- Assign Role to members
- Service account & Compute engine
- Service account keys

© ANKIT MISTRY – GOOGLE CLOUD

IAM

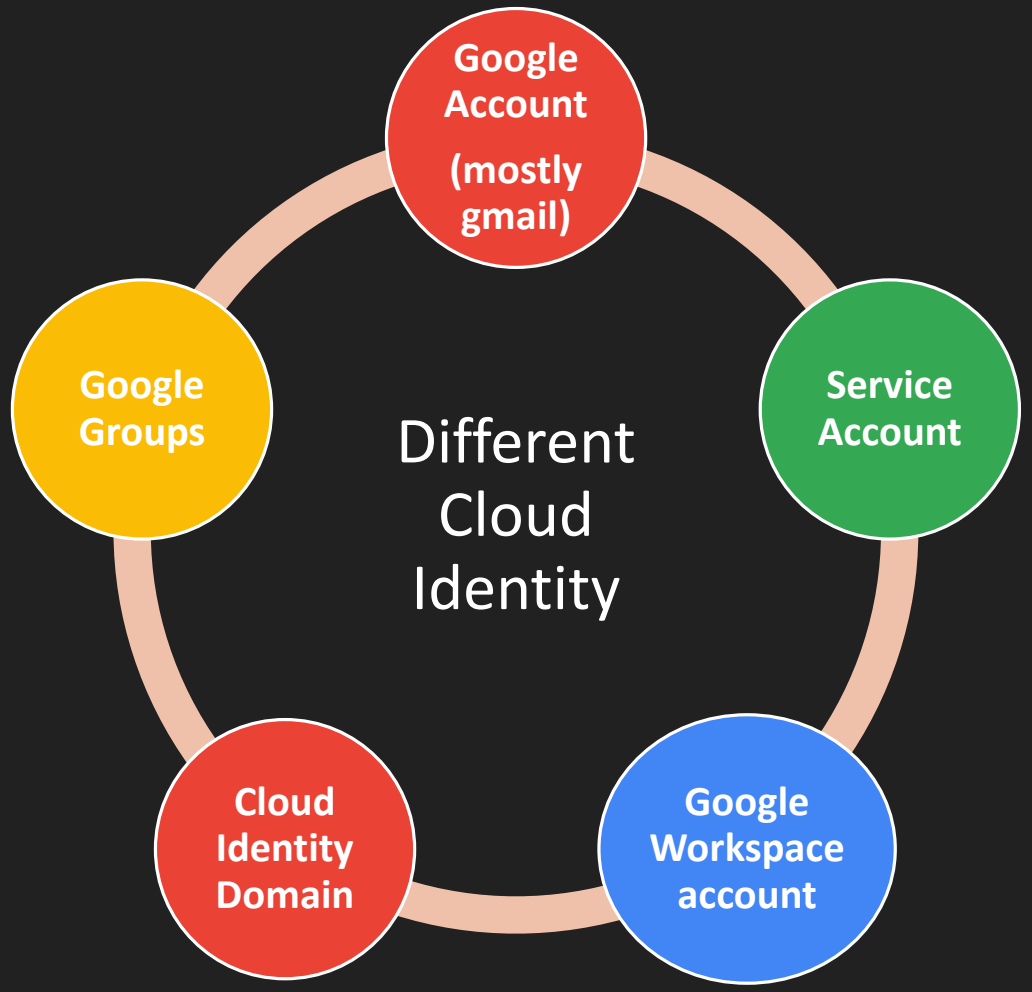
- Identity & access management
- Security framework in GCP
- IAM Defines
  - Who can do What on Which resources.
  - Who - Identity - Member - Email
  - What – Roles (Collection of Permissions)
  - Which (Resources, Compute, Appengine, BigQuery)

X can Create VM in Compute Engine

Y can Delete, Create Bucket in Cloud Storage

© ANKIT MISTRY – GOOGLE CLOUD

Identity - members



- allAuthenticatedUsers
- allUsers

© ANKIT MISTRY – GOOGLE CLOUD

Service Account

- For non human – like for Apps, services, VM
- Service Account is identity for Compute engine
- No Password.
- Service account have keys for authentication
- Max 10 keys per Service Account
- Max 100 Service Account per project
- Service account is not member of Google workspace, Cloud identity
- Let’s see in action

© ANKIT MISTRY – GOOGLE CLOUD

Roles & Permission

BY ANKIT MISTRY

© ANKIT MISTRY – GOOGLE CLOUD

Permissions

- Most Granular Operation performed on Different GCP Resources
  - Cloud Storage
    - View Bucket
    - Create Bucket
- Permissions are represented as
  - **service.resource.verb**
  - pubsub.subscriptions.consume
  - pubsub.topics.publish
- Permission can not be assigned directly to user

© ANKIT MISTRY – GOOGLE CLOUD

Role

- Role = Collection of permission
- Role can be assigned to identity
- Role can be classified into 3 category

Primitive Role	Predefined Role	Custom Role

© ANKIT MISTRY – GOOGLE CLOUD

Primitive Role

- Project level role
- Broad level access
- No Resource specific role
- Viewer, Editor, Owner role
- Viewer = Read only Resources
- Editor = Viewer + Write
- Owner = editor + Access control + billing
- Let’s Explore

© ANKIT MISTRY – GOOGLE CLOUD

Pre-defined Role

- Product specific role
- Granular level access for specific resources
- Google manages all role
- Google defined all role
  - more than 800+ across all GCP Products
- Assign multiple such role to identity
  - Bigtable Reader
  - Dataproc Editor
- Let’s Explore

© ANKIT MISTRY – GOOGLE CLOUD

Custom Role

- If roles requirement doesn’t satisfy by predefined role
- Custom roles can be created by
  - Combining different permission from different product
  - Combine permission of roles
  - Remove some permission from existing predefined role
  - Add some permission from existing predefined role
- Demo
  - Create Custom role which can
  1. create vm
  2. create storage bucket

© ANKIT MISTRY – GOOGLE CLOUD

Role assignment

- Policy
  - Policy binds user with role
- Role can be assigned at Project, Folder or at Org level
- Policy are inherited
- let’s assign some role to
  - Google account
  - Service account
  - Workspace/identity account
  - Groups

© ANKIT MISTRY – GOOGLE CLOUD

Service Account + Compute Engine

BY ANKIT MISTRY

© ANKIT MISTRY – GOOGLE CLOUD

© ANKIT MISTRY – GOOGLE CLOUD

Types of Compute model

- Infrastructure as a Service (IaaS)
- Platform as a Service (PaaS)
- Software as a Service (SaaS)
- Container as a Service (CaaS)

© ANKIT MISTRY – GOOGLE CLOUD

Types of Compute model

On-Premises	IAAS	PAAS	SAAS
Application	Application	Application	Application
Data	Data	Data	Data
Runtime	Runtime	Runtime	Runtime
Middleware	Middleware	Middleware	Middleware
O/S	O/S	O/S	O/S
Virtualization	Virtualization	Virtualization	Virtualization
Servers	Servers	Servers	Servers
Storage	Storage	Storage	Storage
Networking	Networking	Networking	Networking

You Manage

Cloud Provider

© ANKIT MISTRY – GOOGLE CLOUD

IAAS, PAAS, SAAS

- IAAS
  - Only Infrastructure provided by GCP, AWS
  - Provisioning Virtual machine
  - Full flexibility & complete control over machine
  - you need to manage everything yourself.
  - Server management, Upgrade OS, Deploy Application
- PAAS
  - No Server management – GCP will take care
  - Auto scaling, Auto Healing feature
  - You just Focus on Application
- SAAS
  - Google drive
  - GCP will take care everything
  - You are responsible for content you upload on drive & Some feature configuration to access feature
  - You are consumer for Drive App
- CAAS
  - Container as a Service
  - Cloud Run, GKE, App engine Flexible
- FAAS
  - Function as a Service
  - Deploy function - PAAS

© ANKIT MISTRY – GOOGLE CLOUD

Shared Responsibility Model

- Google Responsibility to secure cloud, app, data is one aspect
- As a cloud user, also responsible to secure individual resources
- It is shared responsibility between user & GCP
- GCP provide feature like Encryption at rest & transit, KMS, IAM to secure Data

© ANKIT MISTRY – GOOGLE CLOUD

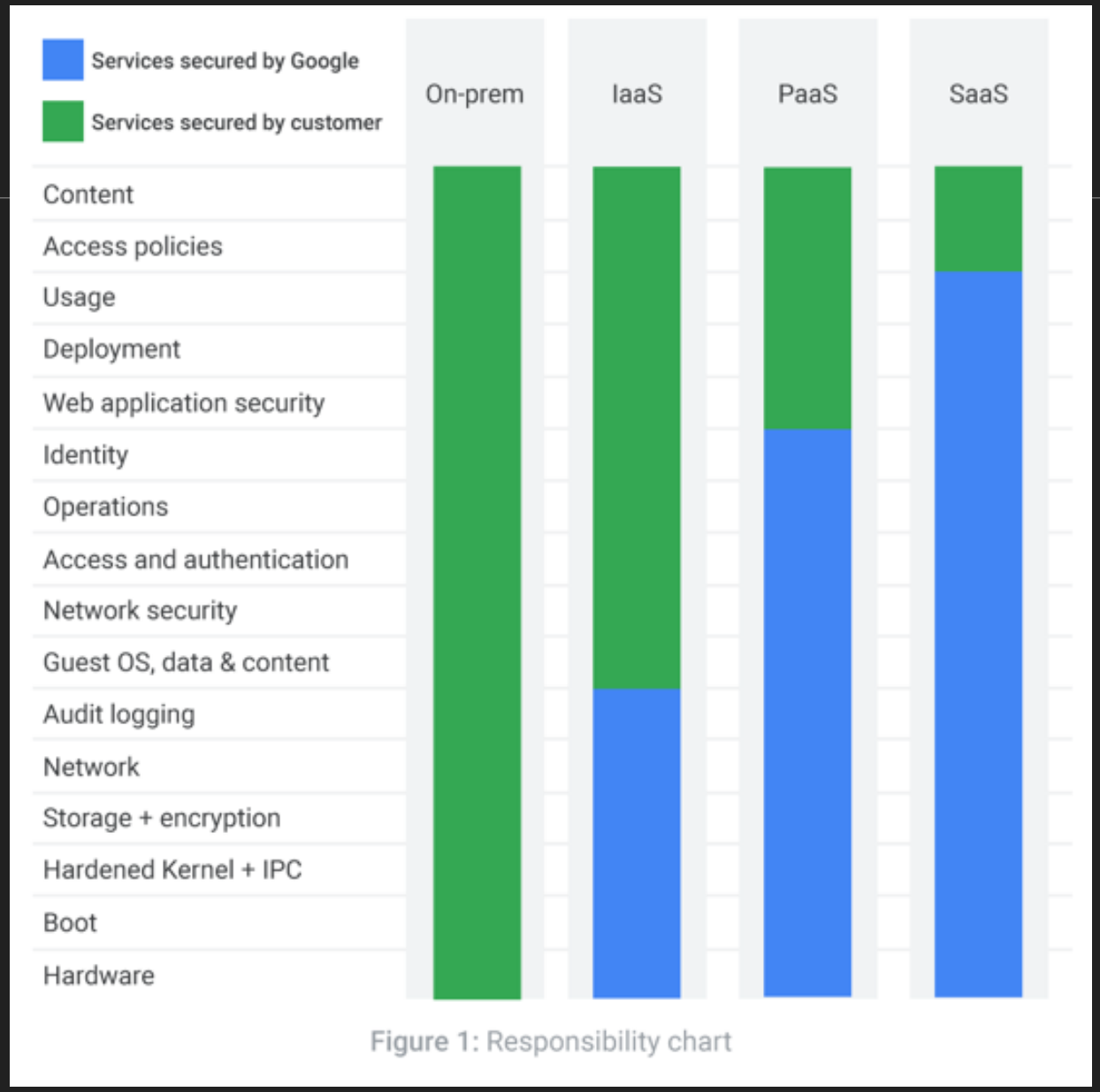


Figure 1: Responsibility chart

https://cloud.google.com/security/incident-response

© ANKIT MISTRY – GOOGLE CLOUD

Cost vs responsibility

Towards Lesser Responsibility Cost will increase

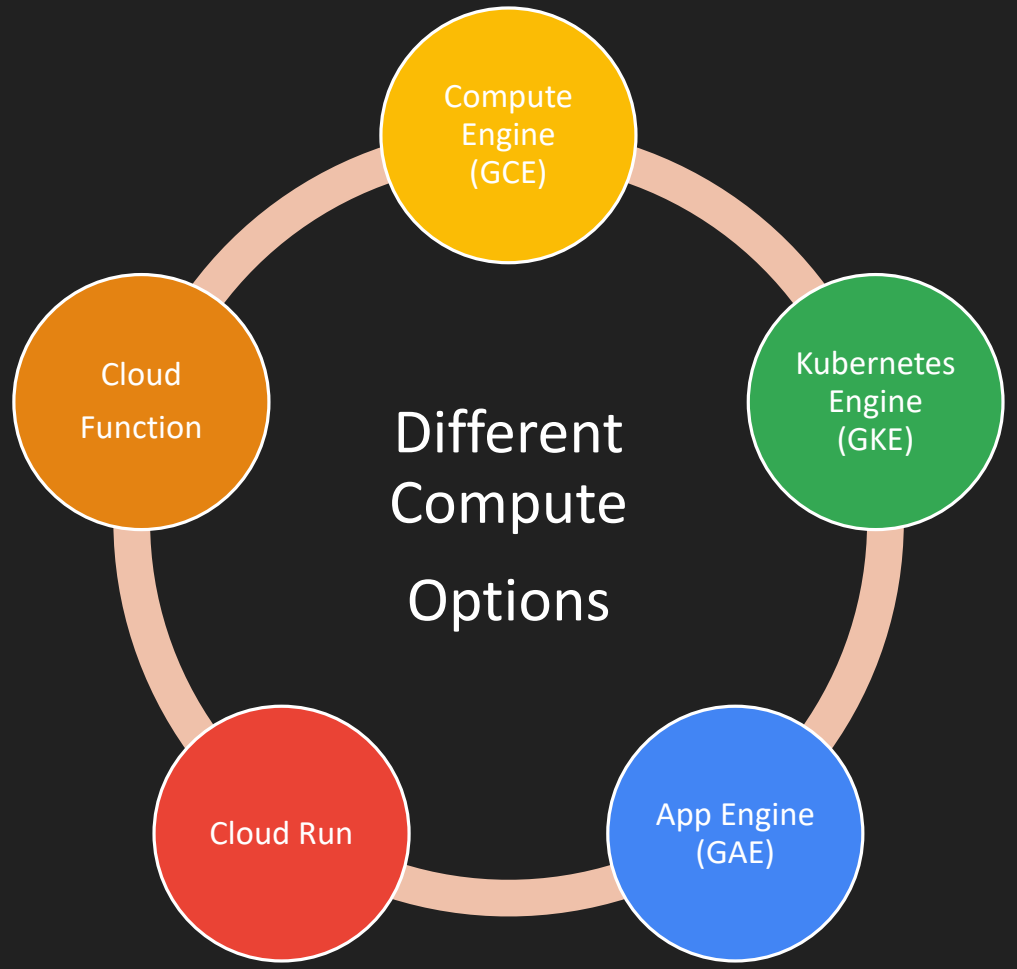
On-Premises	IAAS	PAAS	SAAS
Application	Application	Application	Application
Data	Data	Data	Data
Runtime	Runtime	Runtime	Runtime
Middleware	Middleware	Middleware	Middleware
O/S	O/S	O/S	O/S
Virtualization	Virtualization	Virtualization	Virtualization
Servers	Servers	Servers	Servers
Storage	Storage	Storage	Storage
Networking	Networking	Networking	Networking

You Manage

Cloud Provider

© ANKIT MISTRY – GOOGLE CLOUD

Compute options in GCP



Less Flexibility, control

- Cloud Function (FAAS)
- Cloud Run (CAAS)
- App Engine (GAE) (PAAS)
- Kubernetes Engine (GKE) (CAAS)
- Compute Engine (GCE) (IAAS)

More Flexibility

© ANKIT MISTRY – GOOGLE CLOUD

© ANKIT MISTRY – GOOGLE CLOUD

Google Compute Engine

BY ANKIT MISTRY

© ANKIT MISTRY – GOOGLE CLOUD

Google Compute Engine

- IAAS – Infrastructure as a service
- Where to deploy application in Cloud
  - Compute engine
- Compute engine requires
  - CPU
  - Memory
  - Storage
  - Networking
- Let’s create first virtual machine
  - Regions, Zones, Machine family, Machine Types, OS Image, Disk

© ANKIT MISTRY – GOOGLE CLOUD

Google Compute Engine

- Managing compute engine – lifecycle of compute engine
- Let's SSH into GCE
  - Cloud console
  - CLI
- Machine Family & Types
- Create VM from CLI
- Setup nginx webserver on VM
- Access Webapp with External IP address + Private IP
- Startup script
- Instance template
  - Simplify VM creation
  - Define all VM parameter once & create VM multiple time
- Snapshots, Custom image

© ANKIT MISTRY – GOOGLE CLOUD

Google Compute Engine

- Attach extra disk
- Machine Image
  - A machine image is a Compute Engine resource that stores all the configuration, metadata, permissions, and data from multiple disks of a virtual machine (VM) instance.
- Detach Disk
- Availability Policy
  - On-host maintenance, Automatic Restart, Live migration

© ANKIT MISTRY – GOOGLE CLOUD

Optimize GCE Cost

- Flat-rate
- Committed use discounts[CUD]
- Sustained use discounts[SUD]
- Preemptible VM

© ANKIT MISTRY – GOOGLE CLOUD

Flat-rate, committed use discounts[CUD], sustained use discounts[SUD]

BY ANKIT MISTRY

© ANKIT MISTRY – GOOGLE CLOUD

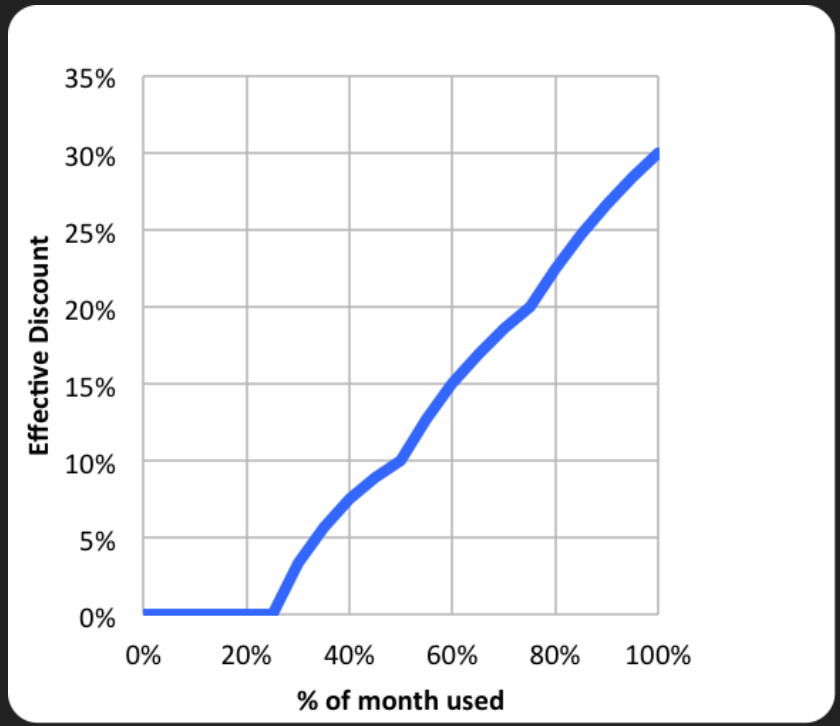
Flat Rate

- Pay for what you use
- No Special Discount
- In Compute Engine :
  - E2 and A2 category of Machine

© ANKIT MISTRY – GOOGLE CLOUD

Sustained use discounts[CUD]

- Sustained use discounts are automatic discounts for running specific Compute Engine resources a significant portion of the billing month
- Applies to N1, N2 machine types
  - Not applicable to other machine type
- If you use at least 25% of month
- Only on GKE & VM Instances
- Let's see in action



© ANKIT MISTRY – GOOGLE CLOUD

Committed use discounts[CUD]

- Let’s say your workload is predictable
- you can commit for 1 year or 3 year
- Get up to 70% of discount.
- Only on GKE & VM Instances
- Can not cancel commitments
- Let’s see in action

© ANKIT MISTRY – GOOGLE CLOUD

Preemptible VM

- Just like Other virtual machine
- Short lived cheaper virtual machine
- Provision Pre-emptible VM When
  - Workload is fault tolerant
  - Not require 100% high availability
  - Cost is critical
- up to 80% discount
- max life is 24 hours
- Not always available
- Google give you 30 sec warning before auto shutdown
  - Regular VM has higher priority than Preemptible VM
- Let’s see how to configure it

© ANKIT MISTRY – GOOGLE CLOUD

Instance Group

- How to create Multiple virtual machine
  - not just 1 or 2, but let's say 100 or even 1000
- Instance group – Group of virtual machine
- Types of Instance group
  - Managed instance group
    - Identical VM
    - Created using instance template
    - Auto scaling, healing kind of feature
  - Unmanaged instance group
    - All VM are not identical

© ANKIT MISTRY – GOOGLE CLOUD

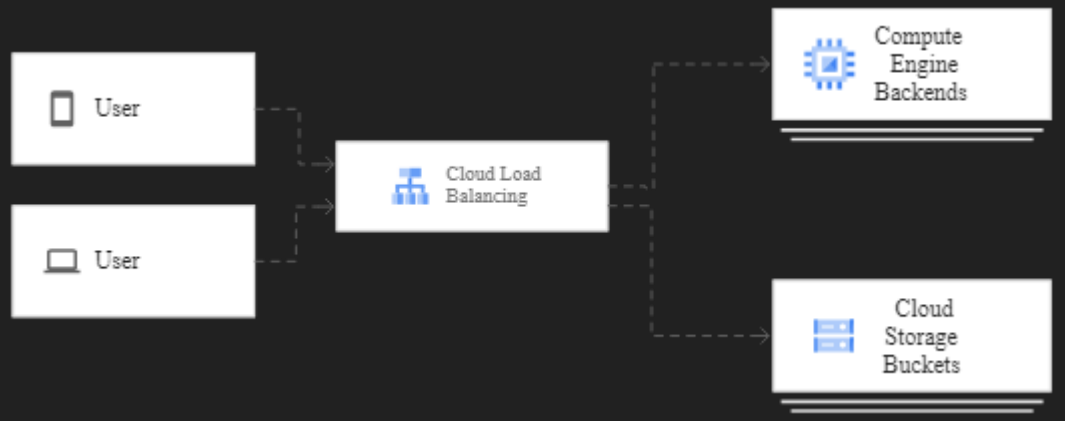
Instance Group Demo

- Managed Instance Group Demo
  - create instance template – with custom startup script (v1.0 script)
  - Provision MIG from instance template
  - Deploy new version v2.0 with another template
  - Rollback to v1.0
- Unmanaged instance Group demo
  - Create 2 VM manually
  - One with apache2 install
  - Other with nginx installed
  - Provision unmanaged instance group and attach above 2 VM

© ANKIT MISTRY – GOOGLE CLOUD

Cloud Load balancing

How to distribute traffic between instance group VM



- External vs Internal load balancing
- regional vs Global Load balancer
- TCP, UDP, SSL, HTTP/HTTPS load balancer

© ANKIT MISTRY – GOOGLE CLOUD

© ANKIT MISTRY – GOOGLE CLOUD

Google APP Engine

BY ANKIT MISTRY

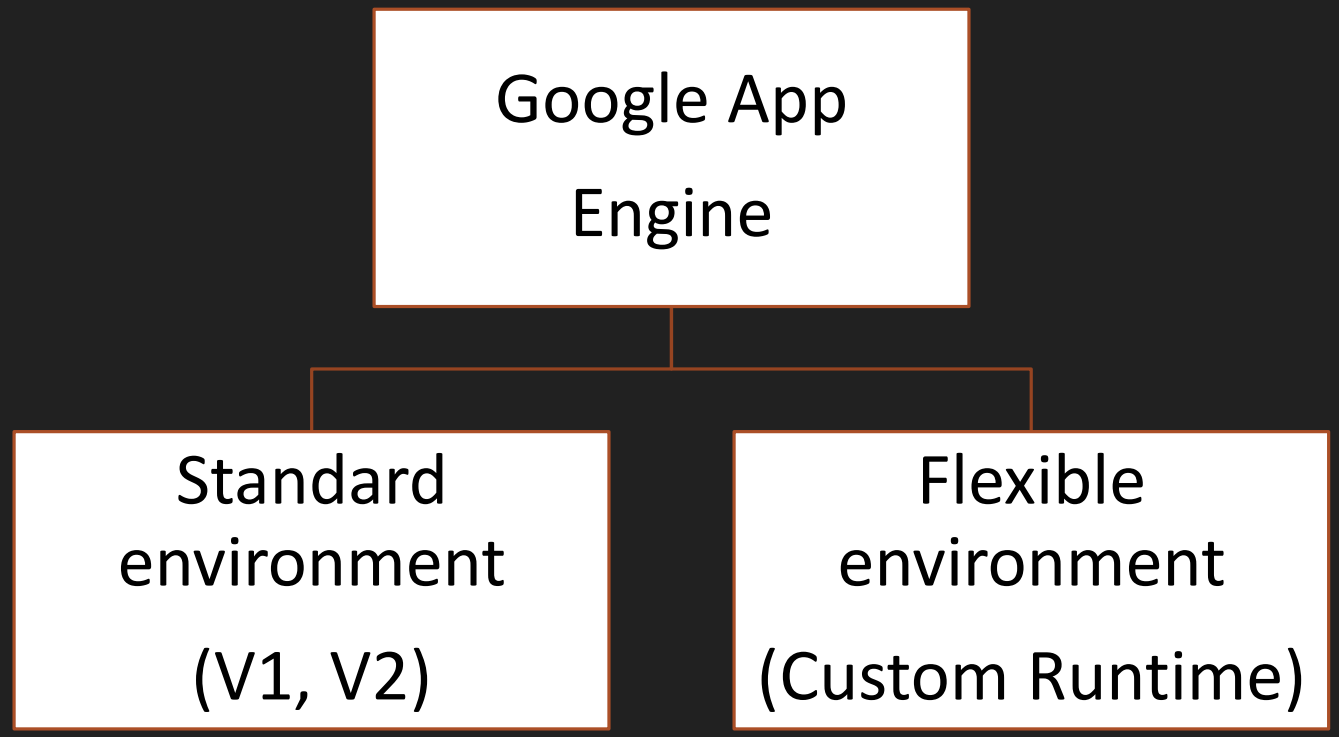
© ANKIT MISTRY – GOOGLE CLOUD

Google App Engine

- Do you want to deploy web app with some fixed runtime, without thinking about server, scaling
- PAAS solution
- Fully managed, no server management (serverless)
- due to serverless, lesser responsibility
- Developer can focus on Development, rather than infrastructure
- It’s oldest Google Cloud Product
- Simplest way to deploy Http based web application
- Auto scaling, Automatic Load balancing
- Support for Java, Python, .NET, Node, Ruby
- One can use custom runtime with container in flexible environment

© ANKIT MISTRY – GOOGLE CLOUD

App Engine Environment

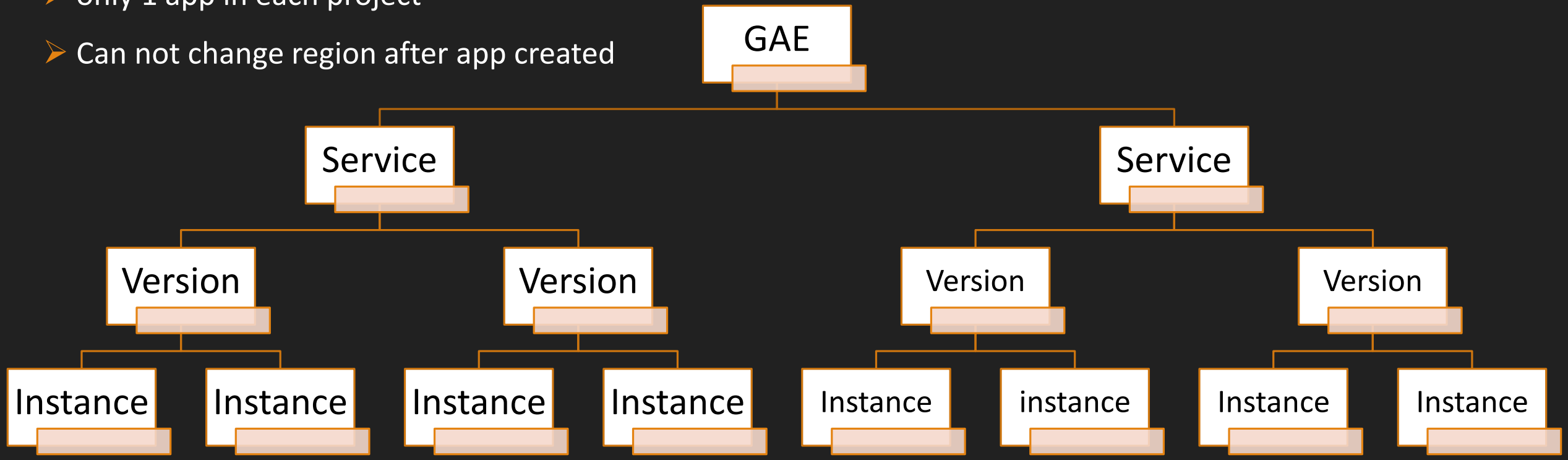


https://cloud.google.com/appengine/docs/the-appengine-environments

© ANKIT MISTRY – GOOGLE CLOUD

App Engine Hierarchy

- Regional Service
- only 1 app in each project
- Can not change region after app created



© ANKIT MISTRY – GOOGLE CLOUD

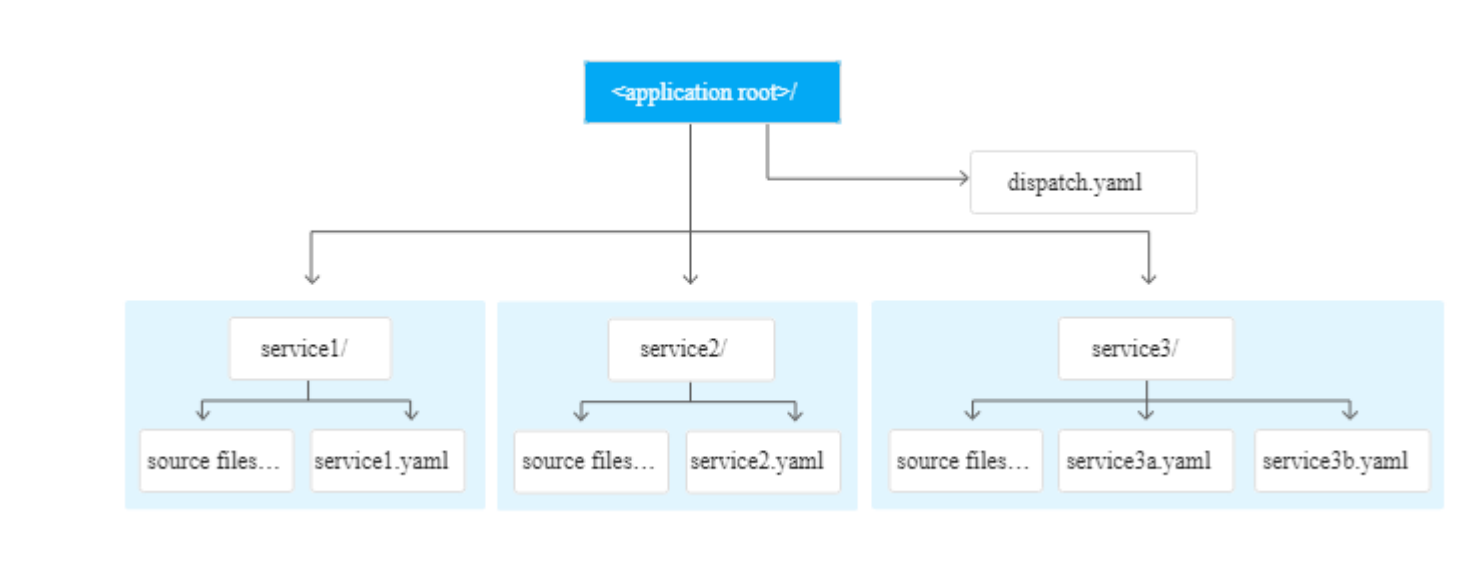
App Engine - Scaling

- Basic
  - Cold start – instance will be created when application receive request.
  - Recommended for adhoc workload
- Automatic
  - App engine takes care about new instances require or destroyed based on workload based on request rate, response latencies, and other application metrics
  - Configure - Max Instances and Min Instances
- Manual
  - No automation, manually specify number of instances required
- All these configuration can be specified in app.yaml

https://cloud.google.com/appengine/docs/standard/python/how-instances-are-managed

© ANKIT MISTRY – GOOGLE CLOUD

App Engine – Code Structure



https://cloud.google.com/appengine/docs/standard/python3/configuration-files

© ANKIT MISTRY – GOOGLE CLOUD

app.yaml

```
runtime: python37 # or another supported version

service : my-service

env_variables:
  BUCKET_NAME: "example-gcs-bucket"

service_account :  ---

automatic_scaling:
  target_cpu_utilization: 0.65
  min_instances: 5
  max_instances: 100
  min_pending_latency: 30ms
  max_pending_latency: automatic
  max_concurrent_requests: 50
```

https://cloud.google.com/appengine/docs/standard/python3/config/appref

© ANKIT MISTRY – GOOGLE CLOUD

Google App Engine Demo

- Deploying application to App Engine
  - Standard – Scale down to 0 instances
- Auto scaling demo
- App versioning – canary deployment (Traffic splitting)
- Deploy another services

© ANKIT MISTRY – GOOGLE CLOUD

© ANKIT MISTRY – GOOGLE CLOUD

Google Cloud Run

- Serverless – fully managed
- Containerized App
- Best of App Engine Standard + Container
- App versioning – canary deployment (Traffic splitting)
- Cloud Run Demo

© ANKIT MISTRY – GOOGLE CLOUD

© ANKIT MISTRY – GOOGLE CLOUD

Google Cloud Function

- Single purpose micro services
- Event based trigger
  - Http
  - Pub sub
  - object upload in Cloud storage
- Deploy code as function

© ANKIT MISTRY – GOOGLE CLOUD

Google Cloud Function

Pub Sub

Cloud Storage

© ANKIT MISTRY – GOOGLE CLOUD

Google Cloud Function (Hands-on)

Http, Cloud Storage, Pubsub

BY ANKIT MISTRY

© ANKIT MISTRY – GOOGLE CLOUD

© ANKIT MISTRY – GOOGLE CLOUD

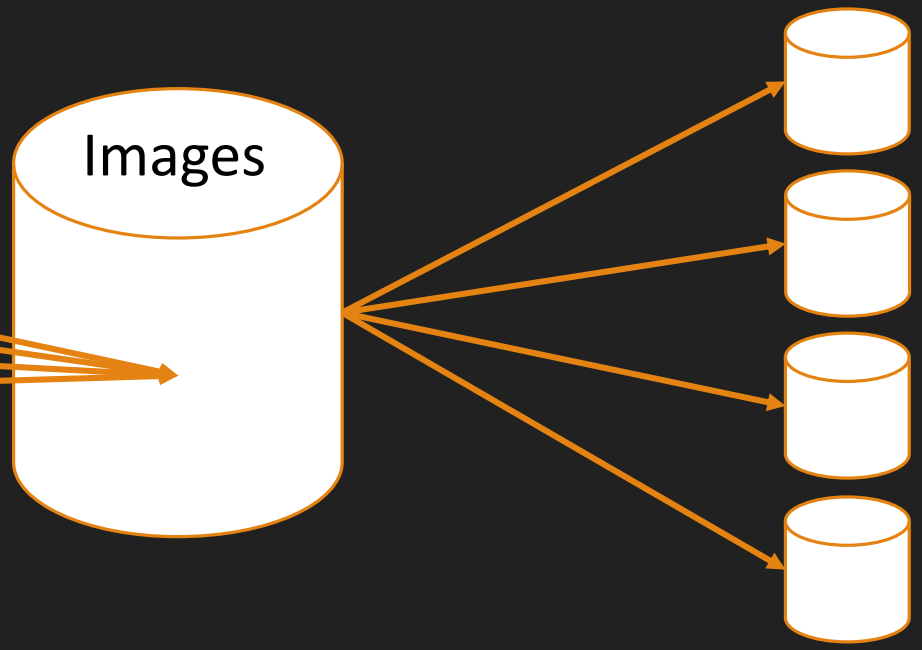
Container & Registry

BY ANKIT MISTRY

© ANKIT MISTRY – GOOGLE CLOUD

Container

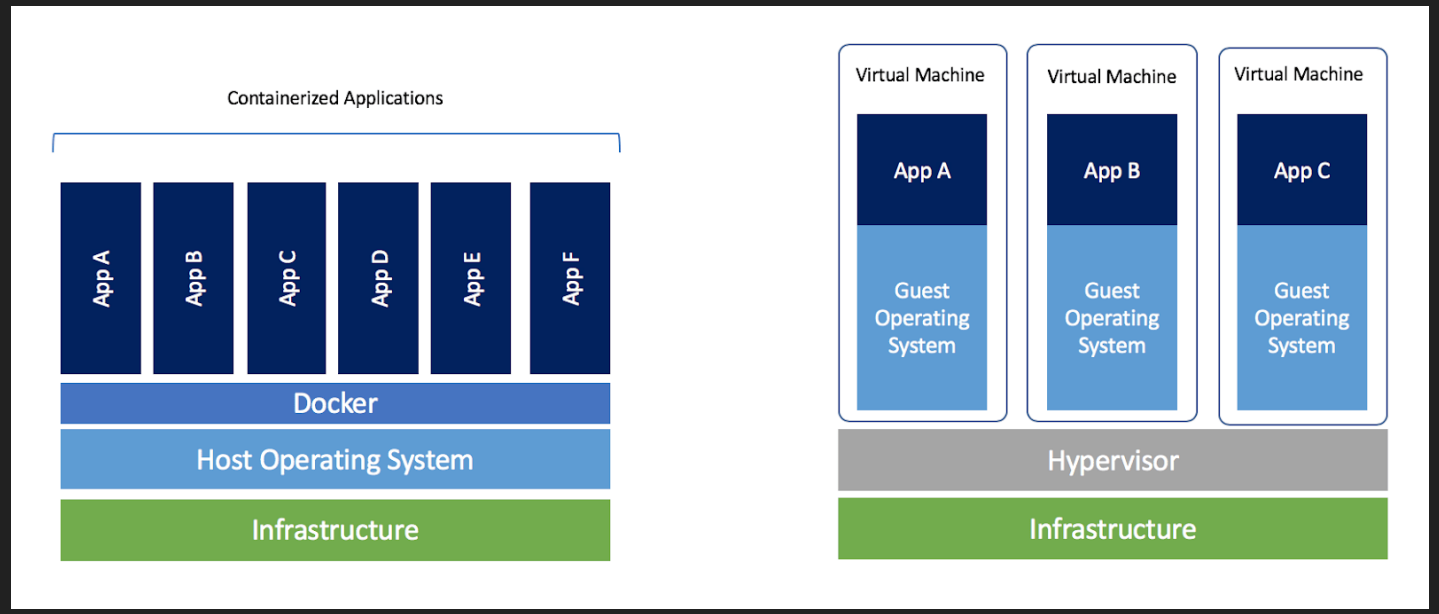
- Software shipping technology
- Let’s say building JAVA APP
  - java runtime (JRE)
  - All library dependency
  - Network configuration
  - Runtime DLL
- Combine all this thing into one single bucket & ship
- Compare with oops
  - Images like class (Blueprint)
  - Container like objects



© ANKIT MISTRY – GOOGLE CLOUD

VM vs Containers

- Container are lightweight
- Easily portable to any public cloud, VM, bare metal
- For Micro service deployment, lightweight containers are preferred
- Fast CI/CD cycle
- All major public cloud providers has services to deploy container
- In GCP
  - VM
  - Cloud Run
  - GKE



https://www.docker.com/blog/containers-replacing-virtual-machines/

© ANKIT MISTRY – GOOGLE CLOUD

Docker

- Container are abstract concept.
- Docker is specific implementation of Container concept.
- Create Docker Images, from Images can create multiple containers
- Here you packaged app in images
- Container use image to start application
- Containers run on any operating system – prefer Linux based
- It works exactly same independent of OS, machine, Environment
- Lightweight compared to VM
- Easier to maintain & deploy
- Docker works with any language, runtime, OS

© ANKIT MISTRY – GOOGLE CLOUD

Docker workflow

Docker Installation

Dockerfile + Source code

Docker Image

Docker Container

Container Registry

Artifact Registry

AWS ECR

© ANKIT MISTRY – GOOGLE CLOUD

Container Registry

- Online storage space for Docker images
- Docker Hub inside Google Cloud
- You can store Docker images, pull images & push images, tag images
- GCP recently introduce next level registry –
  - Artifact registry
  - It can store not just Docker image but many more thing like NPM, maven
- Naming convention :
  - HostName/ProjectID/imagename:Tag - gcr.io/[ProjectID]/nginx:1.0
- Binary authorization can be used to detect vulnerabilities & enforce deployment policies.
- No IAM Role defined at granular level
- No Region specific Repository
- Pricing – store in GCS

© ANKIT MISTRY – GOOGLE CLOUD

Create First Docker images

- Node Application
- Simple Hello world will be returned.
- Filename
  - server.js
  - Dockerfile
- Run app with -> node server.js
- docker build -t myapp:v1.0 .
- Push Images
  - Docker hub
  - Container registry
  - artifact registry

© ANKIT MISTRY – GOOGLE CLOUD

Artifact Registry

- Artifact Registry comes with fine-grained access control via Cloud IAM
- Multiple Repository per project
- Regional & Multi-region repositories
- It can store not just Docker image but many more thing like NPM, maven, Python
- asia-southeast1-docker.pkg.dev/[ProjectID]/[repo]/nginx:v1.0
- Create Repo (Not Required for Container Registry)
- Configure :
  - gcloud auth configure-docker asia-southeast1-docker.pkg.dev
- Let’s see
  - How to configure via gcloud
  - Push image to AR

© ANKIT MISTRY – GOOGLE CLOUD

Cloud Build

- With Cloud Build
  - Rent machine in Google cloud to build image & push
  - No Docker required at local side
  - With gcloud command (build + push in single command)
  - gcloud builds submit . --tag=gcr.io/$DEVSHELL_PROJECT_ID/myapp:v1.0

© ANKIT MISTRY – GOOGLE CLOUD

Deploy Docker Image to GCE

BY ANKIT MISTRY

© ANKIT MISTRY – GOOGLE CLOUD

Google Kubernetes Engine

BY ANKIT MISTRY

© ANKIT MISTRY – GOOGLE CLOUD

Kubernetes

- Let's say
  - you want to create 100's of container to scale your app
  - need some automate approach which fully manage all container lifecycle
- Docker developed Docker swarm
- In parallel Google internally developed Borg for Container management
- Google made Borg open sources becomes Kubernetes.
- Kubernetes is the solution for managing container
- Open-Source Container orchestration system in 2014.
- k8S – 8 letter between first and last character
- No vendor lock in – can be deployed anywhere AWS, Azure, Google cloud, on-premises
- Written in Go language.

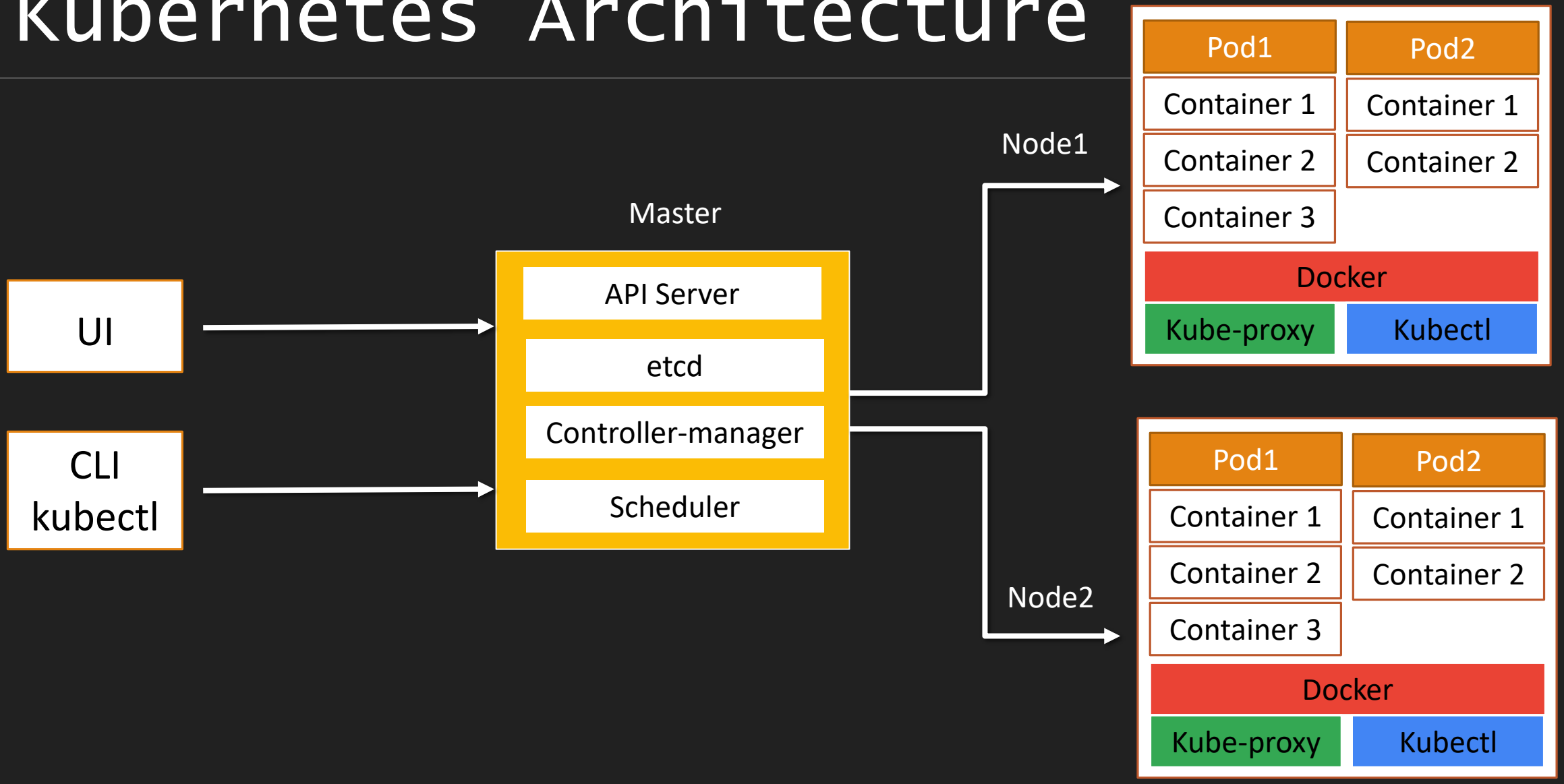
© ANKIT MISTRY – GOOGLE CLOUD

Google Kubernetes Engine

- GKE - Google Kubernetes Engine
- Why Google Developed GKE
  - Master – slave architecture
  - Installation is tedious task
  - Google build managed Kubernetes service
  - One can provision cluster in matter of minutes
- Fully managed services in Google Cloud
  - Whole operation, setup, installation work being take care by Google
- Launched in 2015
- Google is first to bring Containerization in Cloud

© ANKIT MISTRY – GOOGLE CLOUD

Kubernetes Architecture



© ANKIT MISTRY – GOOGLE CLOUD

[Hands-on] GKE

- Create two GKE Cluster
  - From Cloud Console
  - From gcloud cli
- Deploy Workload (Docker image From CR/AR)
- Expose as a service
- Scale services
  - Horizontal Pod scaling
  - Node Scaling
- Adding New Node Pool
- Set New Docker Image
- Configure AutoScale
- ConfigMap & Secret

© ANKIT MISTRY – GOOGLE CLOUD

Declarative Configuration

- Configuration is written in YAML Format
- Each YAML has 4 main components.
- kubectl apply –f file.yaml

deploy.yaml

```yaml
apiVersion: apps/v1
kind: Deployment
metadata:
  name: desc-deployment
spec:
  selector:
    matchLabels:
      app: desc-dep-1
  template:
    metadata:
      labels:
        app: desc-dep-1
    spec:
      containers:
      - name: nginx-con
        image: nginx
        resources:
          requests:
            memory: "32Mi"
            cpu: "100m"
          limits:
            memory: "128Mi"
            cpu: "500m"
        ports:
        - containerPort: 80
```

© ANKIT MISTRY – GOOGLE CLOUD

© ANKIT MISTRY – GOOGLE CLOUD

Google Cloud Storage

© ANKIT MISTRY – GOOGLE CLOUD

Google Cloud Storage

- Object storage solution in GCP
- Unstructured Data storage
  - Image
  - Video
  - Binary File, etc…
- Cloud storage can be used for long term archival storage
- Can be access object over http, Rest API
- No capacity planning required
  - Scale to Exabyte
- Unlimited data can be stored
- By Default Data is encrypted at rest
- In transit also by default encryption.

© ANKIT MISTRY – GOOGLE CLOUD

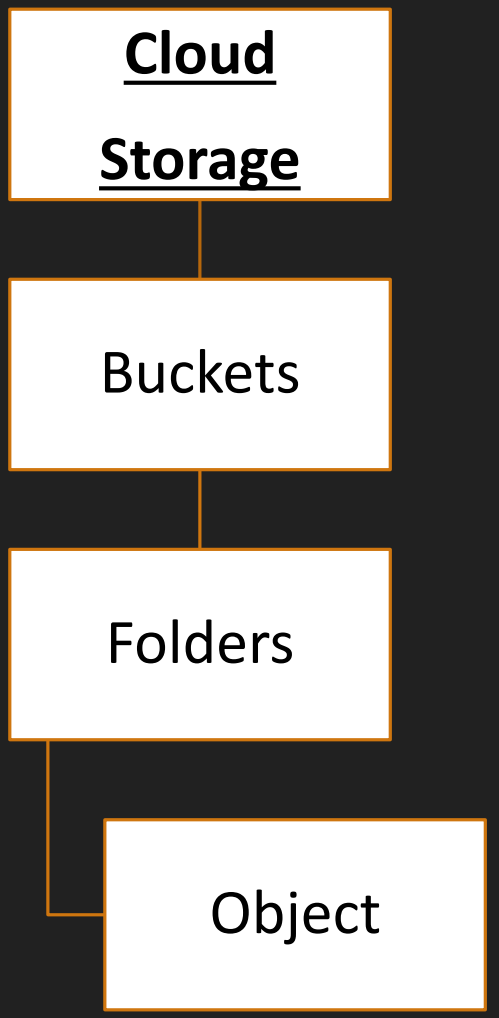
Google Cloud Storage

- No minimum Object Size
- Max object size is 5 TB
- High Durability – 99.999999999% annual
- Object can be Globally access
- Single API to access across multiple storage class
- Data is geo - redundant
  - Due to Multiregional
  - Dual-Region storage

© ANKIT MISTRY – GOOGLE CLOUD

Object Organization

- Global unique name for bucket
- Example access URL :
  - https://storage.cloud.google.com/[bucket]/[objectname]
- Bucket name appear in URL
- So, be careful while naming bucket
- Does not store anything like file system
  - Folder are virtual
- Bucket level lock with data retention policy
- Object are immutable
- Object can be versioned



© ANKIT MISTRY – GOOGLE CLOUD

Storage Location

Region	Dual-region	Multi-region
• Lowest latency within a single region • Replicated data across multiple zone in single region	• High availability and low latency across 2 regions (Paired region) • Auto-failover	• Highest availability across continent area – US, EU, Asia • Auto-failover

© ANKIT MISTRY – GOOGLE CLOUD

Storage Class

- How frequently access data
- How much amount of data

Standard

- Good for Hot data
- High frequency access
- Storage Costliest
- Access cost is very low
- Low latency
- SLA :
  - 99.95% Multi/Dual
  - 99.9% Regional

Near line

- Low Frequency access Once in a 30 days
- Storage is Cheaper than standard
- Access cost will increase
- Back up
- SLA : 99.9% Multi/Dual
  - 99.0% for Regional

Cold line

- Very low frequency to access
- Once in 90 days
- Storage is Cheaper than Near line
- SLA :
  - 99.9% Multi/Dual
  - 99.0% for Regional

Archive

- Offline data
- Backup
- Data access is once in year
- Storage Cheapest
- Access cost very high
- No SLA

© ANKIT MISTRY – GOOGLE CLOUD

[Hands-on]
Google Cloud Storage

© ANKIT MISTRY – GOOGLE CLOUD

Object Lifecycle management

- Based on condition what action needs to perform on object.
- Condition
  - Object age
  - Object file type
  - after some specific date
- Action
  - Transition to different storage class for high performance
  - Like – Standard to Nearline
  - Coldline to Delete

© ANKIT MISTRY – GOOGLE CLOUD

Secure Data with Encryption

- Encryption
  - Google managed Encryption keys
    - No Configuration
    - Fully managed
  - Customer managed Encryption keys
    - Create keyring in Cloud KMS
    - key will be managed by customer. Like Key rotation
  - Customer supplied Encryption keys
    - We will generate Key with : openssl rand =base64 32
    - gsutil – encrypt with CSEK

© ANKIT MISTRY – GOOGLE CLOUD

Object Versioning

- Help to prevent accidental deletion of object
- Enable/Disable versioning at bucket level
- Get access to older version with (object key + version number)
- If you don’t need earlier version, delete it & reduce storage cost
- If you don’t specify version number, always retrieve latest version
- Let’s see in action

© ANKIT MISTRY – GOOGLE CLOUD

Controlling access

- Who can do what on GCS at what level
- Permissions
- Apply at Bucket level
  - Uniform level access
    - No Object level permission
    - Apply uniform at all object inside bucket
  - Fine grained permission
    - Access Control List – ACL For Each object Separately
- Apply Project level
  - IAM
  - Different Predefined Role
    - Storage Admin
    - Storage Object Admin
    - Storage Object Creator
    - Storage Object Viewer
  - Create Custom Role
- Assign Bucket level Role
  - Select bucket & assign role
  - To user
  - To other GCP services or product

© ANKIT MISTRY – GOOGLE CLOUD

Bucket Retention Policy

- Minimum duration for which bucket will be protected from
  - Deletion
  - modification
- Let’s see How to Configure it.

© ANKIT MISTRY – GOOGLE CLOUD

Signed URL

- Temporary access
- you can give access to user who doesn't have Google Account.
- URL expired after time period defined.
- Max period for which URL is valid is 7 days.
- gsutil signurl -d 10m -u gs://<bucket>/<object>

© ANKIT MISTRY – GOOGLE CLOUD

GCS - Pricing

- Storage Pricing
- Data access Pricing
- Go to Cloud Console & create Bucket, observe pricing

© ANKIT MISTRY – GOOGLE CLOUD

© ANKIT MISTRY – GOOGLE CLOUD

Encryption

BY ANKIT MISTRY

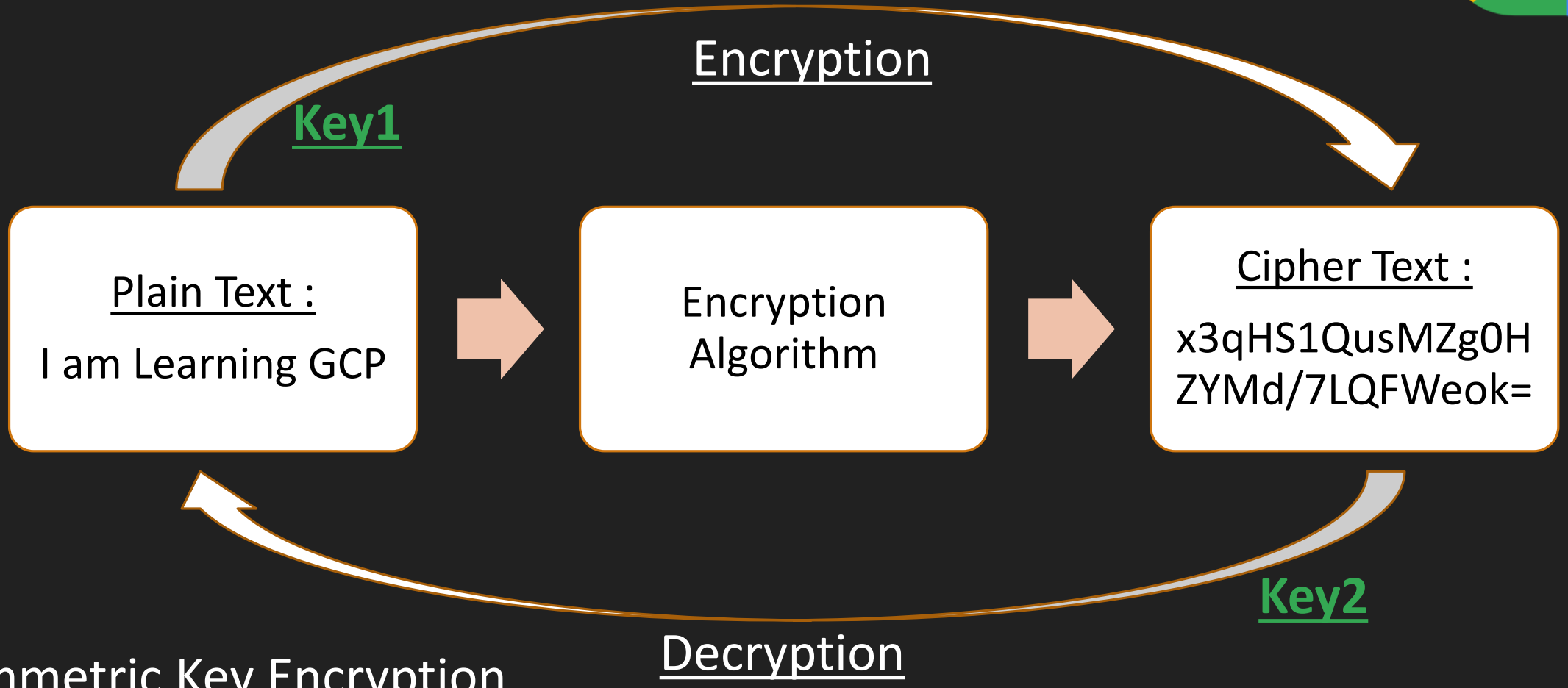
© ANKIT MISTRY – GOOGLE CLOUD

Why Encryption

- In GCP, Data stored at
  - GCS
  - Persistent Disk, SSD
  - File Server
  - Database File
- If Let's say hacker get access to your hard Disk?

© ANKIT MISTRY – GOOGLE CLOUD

Encryption



Symmetric Key Encryption

Asymmetric Key Encryption

© ANKIT MISTRY – GOOGLE CLOUD

When Encryption

- Data at Rest
  - Data Situated at GCS, Database
- Data in Motion
  - Data transfer from one network to another
  - Within GCP or Outside of GCP
- Data in Use
  - Data situated in RAM.
  - Memory Store, In memory Data Processing

© ANKIT MISTRY – GOOGLE CLOUD

Cloud KMS

BY ANKIT MISTRY

© ANKIT MISTRY – GOOGLE CLOUD

What are the things need to encrypt

- What are the things need to encrypt
  - Data
  - Keys
    - Envelope Encryption
- Client Side
  - Encryption that occurs before data is sent to Cloud Storage - GCP.
- Server Side
  - Encryption that occurs after Google Cloud receives your data

© ANKIT MISTRY – GOOGLE CLOUD

Cloud KMS

- Manage encryption keys on Google Cloud.
- 3 ways of managing keys
  - Google-managed encryption keys
  - Customer-managed encryption keys
  - Customer-supplied encryption keys

© ANKIT MISTRY – GOOGLE CLOUD

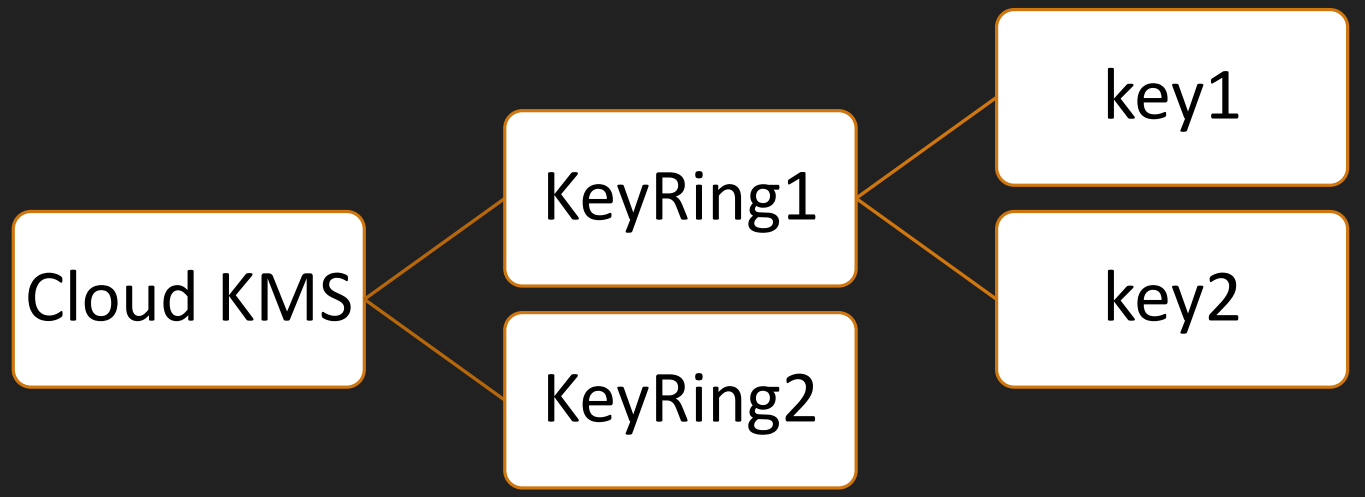
Google-managed encryption keys

- By Default Encryption
- Server side encryption – before data written to disk
- No Additional Configuration required
- Encrypt data using AES-256
- Google manage rotation policy

© ANKIT MISTRY – GOOGLE CLOUD

Customer-managed encryption keys

- Keys generated by Google Cloud KMS
- Customer has control over
  - Rotation policy
  - HSM/Software based keys



© ANKIT MISTRY – GOOGLE CLOUD

Customer-supplied encryption keys

- Complete control over encryption keys
- If keys lost, data can not be recovered
- To generate keys,
  - openssl rand -base64 32
- Can not create bucket from Cloud console
- gsutil -o 'GSUtil:encryption_key='**keys**

© ANKIT MISTRY – GOOGLE CLOUD

© ANKIT MISTRY – GOOGLE CLOUD

Application Secrets

- While building Application we need to store
  - Database password, Some API Keys
- It’s not good idea to store in code or some config file
- Solution : Secret manager
- Dynamically grab secret inside code from secret manager
- Let’s see in action

© ANKIT MISTRY – GOOGLE CLOUD

Get App Secret inside Cloud Function

BY ANKIT MISTRY

© ANKIT MISTRY – GOOGLE CLOUD

© ANKIT MISTRY – GOOGLE CLOUD

Few Database Concept

- Relational data
  - OLTP
  - OLAP
- RTO vs RPO
- Vertical vs Horizontal Scaling
- Availability & Durability

© ANKIT MISTRY – GOOGLE CLOUD

OLTP & OLAP

© ANKIT MISTRY – GOOGLE CLOUD

OLTP

- OLTP – Online Transaction Processing
- Simple Query
- Large number of small transaction
- Traditional RDBMS
- Database modification
- Popular Database - MySQL, PostgreSQL, Oracle, MSSQL
- ERP, CRM, Banking application
- GCP - Cloud SQL, Cloud spanner

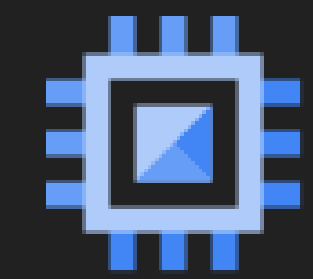
© ANKIT MISTRY – GOOGLE CLOUD

OLAP

- OLAP – Online Analytical Processing
- Data warehousing
- Data is collected from multiple sources
- Complex Query
- Data analysis
- Google Cloud Big Query – Petabyte Data warehouse
- Reporting Application, Web click analysis, BI Dashboard app

© ANKIT MISTRY – GOOGLE CLOUD

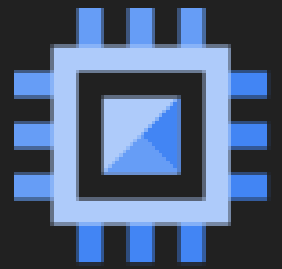
Vertical – Horizontal Scaling



8 VCPU - 16 GB

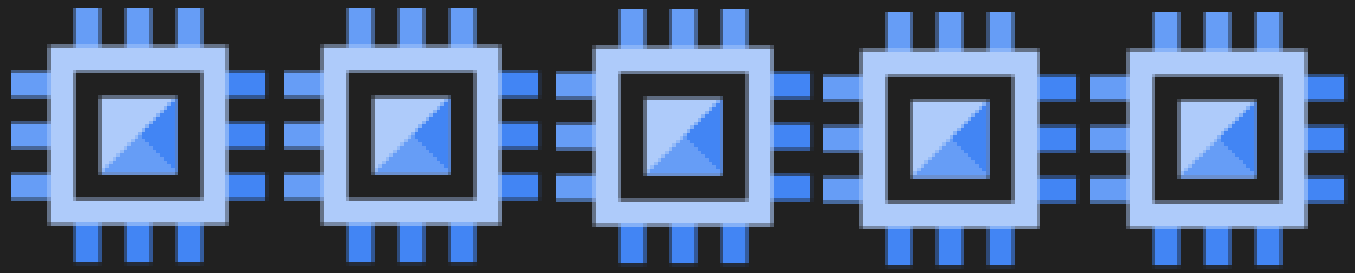
Vertical Scaling

12 VCPU - 24 GB



8 VCPU - 16 GB

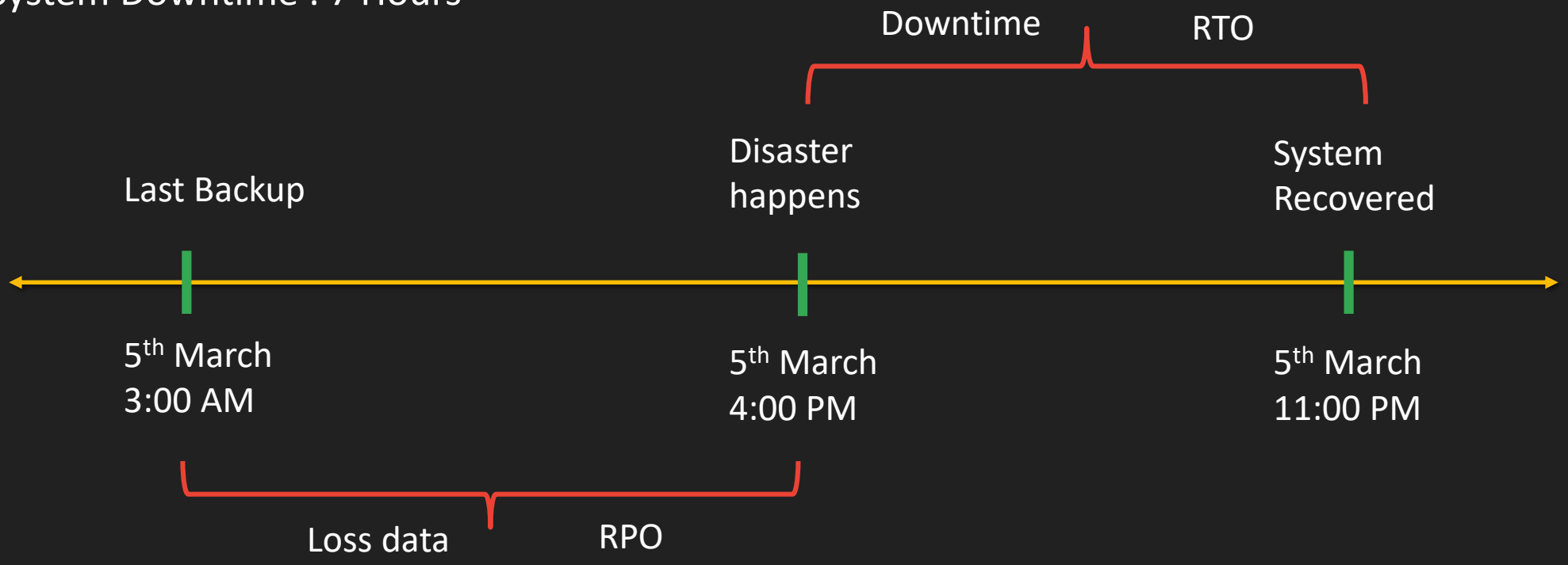
Horizontal Scaling



© ANKIT MISTRY – GOOGLE CLOUD

RTO & RPO

- Data loss : 13 hours
- System Downtime : 7 Hours



© ANKIT MISTRY – GOOGLE CLOUD

RTO & RPO

- RTO – Recovery Time objective
  - Maximum time for which system can be down
- RPO - Recovery Point objective
  - Maximum time for which organization can tolerate Dataloss

© ANKIT MISTRY – GOOGLE CLOUD

RTO & RPO

- RTO – Recovery Time objective
  - Maximum time for which system can be down
- RPO - Recovery Point objective
  - Maximum time for which organization can tolerate Dataloss

© ANKIT MISTRY – GOOGLE CLOUD

Durability

- If you loose data means
  - business is down
  - No business afford to loose data
- How healthy & resilient your data is
- Object Storage provider measure durability in terms of number of 9's
- Example : 99.999999999999% - 11 9's
- That means that even with one billion objects, you would likely go a hundred years without losing a single one!
- https://cloud.google.com/blog/products/storage-data-transfer/understanding-cloud-storage-11-9s-durability-target

© ANKIT MISTRY – GOOGLE CLOUD

Availability

- If region goes down where your data stored
  - Replicate data across many region
- How much amount of time data is up/available to access.
- Data replicated across multiple regions, means higher Availability
- SLA – service level agreement
- SLA – 99.99% : four 9's
- https://uptime.is
- https://cloud.google.com/terms/sla

© ANKIT MISTRY – GOOGLE CLOUD

GCP Database products

Structure data	• Cloud SQL • Cloud Spanner • BigQuery
Semi-structure data	• BigTable • Datastore • Memorystore

© ANKIT MISTRY – GOOGLE CLOUD

© ANKIT MISTRY – GOOGLE CLOUD

Relational Database in GCP

BY ANKIT MISTRY

© ANKIT MISTRY – GOOGLE CLOUD

Google Cloud SQL

© ANKIT MISTRY – GOOGLE CLOUD

Google Cloud SQL

- Fully managed Relational database services for MySQL, PostgreSQL & SQL Server
- Lift & shift above database
- Regional Database with 99.95% SLA
- Storage up to 30 TB
- Scale up to 96 core & 416 GB Memory
- No Horizontal Scaling
- Data is encrypted with Google managed key or CMEK
- Cloud SQL can be accessed from anywhere like – App Engine, Compute Engine…
- Used for storing Transactional database
- Ecommerce, CRM kind application backend.

© ANKIT MISTRY – GOOGLE CLOUD

Google Cloud SQL

- No maintenance & auto update
- Back-up Database
  - On-demand Backup
  - Schedule backup
- Database migration service (DMS)
  - migrate data from different SQL system to Cloud SQL
- Point-in Time Recovery
- Scale with Read replicas – To transfer workload to other instance
- Export data
  - gcloud utility or Cloud Console
  - In SQL/CSV format

© ANKIT MISTRY – GOOGLE CLOUD

[Hands-on]
Create Google Cloud SQL

© ANKIT MISTRY – GOOGLE CLOUD

[Hands-on] Connect Cloud SQL & IP Whitelisting

© ANKIT MISTRY – GOOGLE CLOUD

[Hands-on] Data migration to Cloud SQL Demo

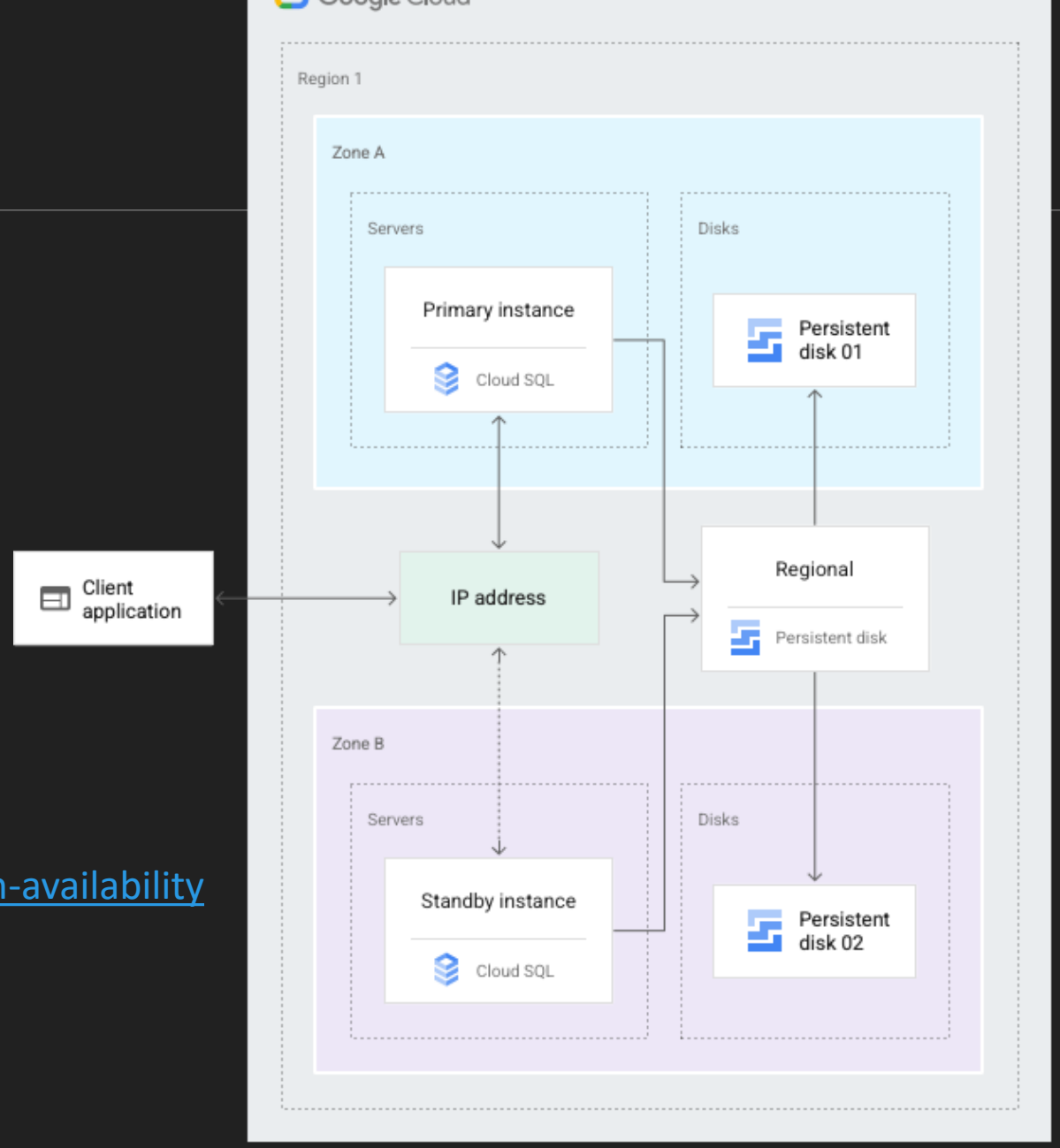
© ANKIT MISTRY – GOOGLE CLOUD

[Hands-on]
Cloud SQL failover demo

© ANKIT MISTRY – GOOGLE CLOUD

Google Cloud SQL Failover

https://cloud.google.com/sql/docs/mysql/high-availability



© ANKIT MISTRY – GOOGLE CLOUD

Cloud SQL Explore

© ANKIT MISTRY – GOOGLE CLOUD

Cloud SQL Export

© ANKIT MISTRY – GOOGLE CLOUD

Google Cloud Spanner

© ANKIT MISTRY – GOOGLE CLOUD

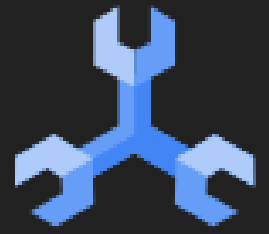
Google Cloud Spanner

- Distributed & scalable solution for RDBMS in GCP
- Fully managed, Mission critical application
- Horizontal Scalability
- use when Data volume > 2 TB
- Costlier than Cloud SQL
- Cloud SQL has just Read replicas,
  - where as in cloud spanner horizontal read/write across region
- Highly scalable, Petabyte scale
- Data is strongly typed.
  - Must define schema database
  - Datatype for each column of each table must be defined.
- 99.999% availability
- Cloud native solution – specific to GCP
  - Lift & Shift not possible, Not recommended
- Spanner = Cloud SQL + Horizontal Scalable
- Scale to petabyte
- Regional/ Multi-region level instance can be created
- Data export
  - can not export with gcloud
  - Cloud Console or Cloud Dataflow Job

© ANKIT MISTRY – GOOGLE CLOUD

Spanner vs RDBMS-SQL

	Spanner	Cloud SQL
Availability	High	During failover little downtime
Scalable	Horizontal	vertical
Price	Costly	Cheaper than spanner
SQL/Schema Support	yes	yes
Replication	High	Only Read Replica



© ANKIT MISTRY – GOOGLE CLOUD

[Hands-on] Cloud Spanner

© ANKIT MISTRY – GOOGLE CLOUD

Cloud Spanner Demo

- Create Spanner Instance
- Create database edu_db
- Create 2 Table
  - Author
    - AuthorID
    - AuthorName
  - Book
    - BookId
    - Bookname
    - AuthorId

© ANKIT MISTRY – GOOGLE CLOUD

After Job Done make sure to delete Spanner Instance

© ANKIT MISTRY – GOOGLE CLOUD

Which database to use when

- Cloud SQL
  - Lift & Shift SQL based system
  - CRM, Ecommerce App
  - Max Data size is 30 TB
- Cloud Spanner
  - Horizontal scalability
  - Low latency
  - High scalability in terms of storage + compute
  - if Data Storage requirement is beyond TB

© ANKIT MISTRY – GOOGLE CLOUD

© ANKIT MISTRY – GOOGLE CLOUD

Cloud VPC & Subnets

BY ANKIT MISTRY

© ANKIT MISTRY – GOOGLE CLOUD

Networking Basics

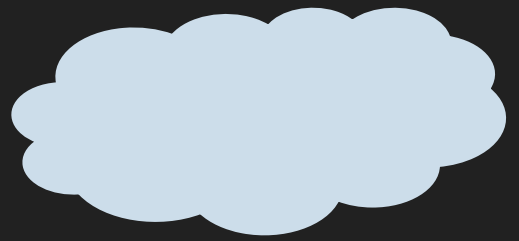
- What is Network
- IP Address & CIDR ranges
- RFC 1918 standard

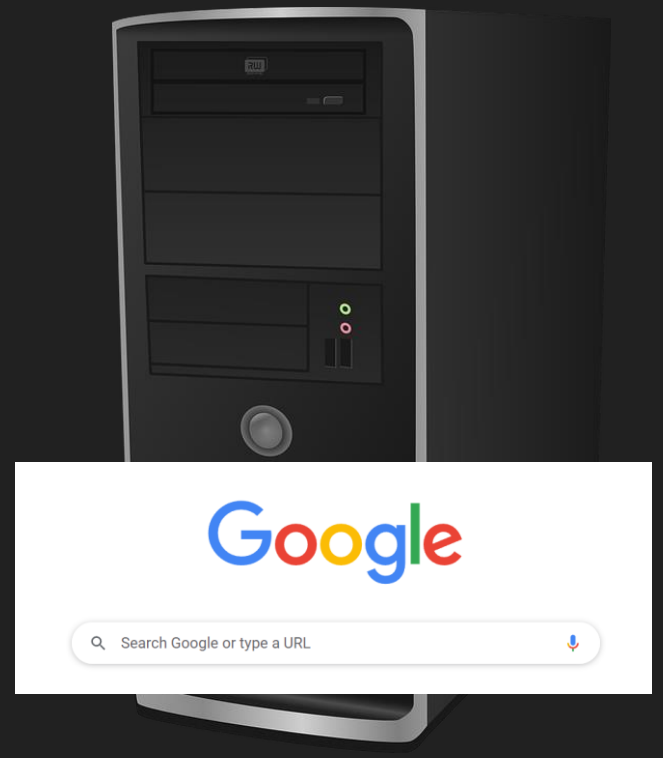
© ANKIT MISTRY – GOOGLE CLOUD

Network

© ANKIT MISTRY – GOOGLE CLOUD

Home Network

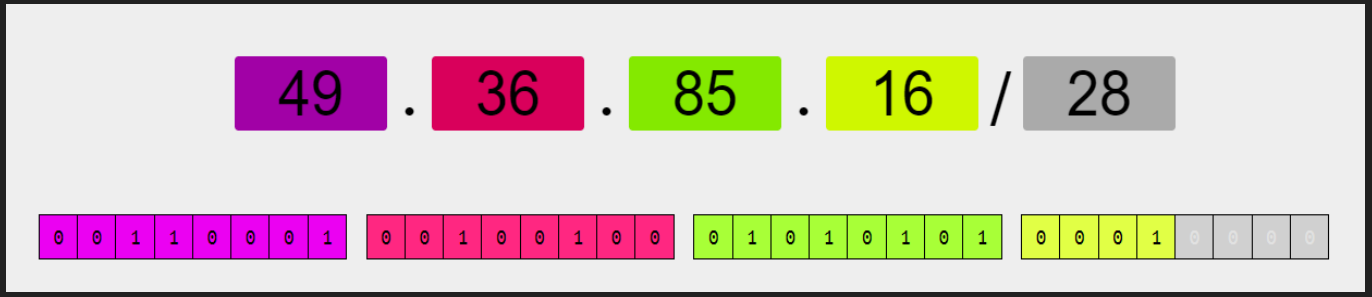




© ANKIT MISTRY – GOOGLE CLOUD

IP address

49.36.84.16



- 32 Bit representation
- IPV4 - 4 number
- 4 Billion address can be represented
- Advanced – IPV6
- many more IP can be represented – 2^128
- Your machine IP : https://api.ipify.org/

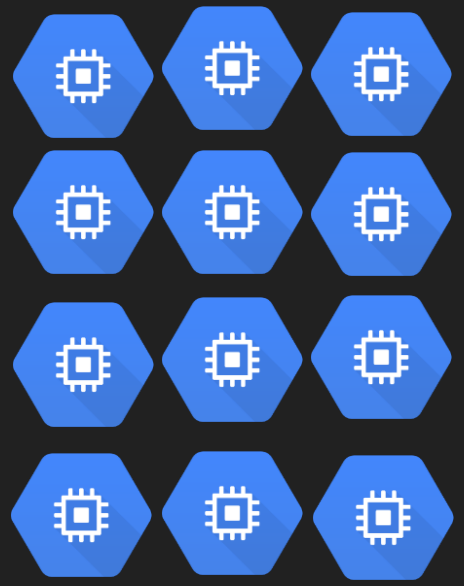
Ref : https://cidr.xyz/

© ANKIT MISTRY – GOOGLE CLOUD

CIDR notation

Classless Inter-Domain Routing

123.52.36.47



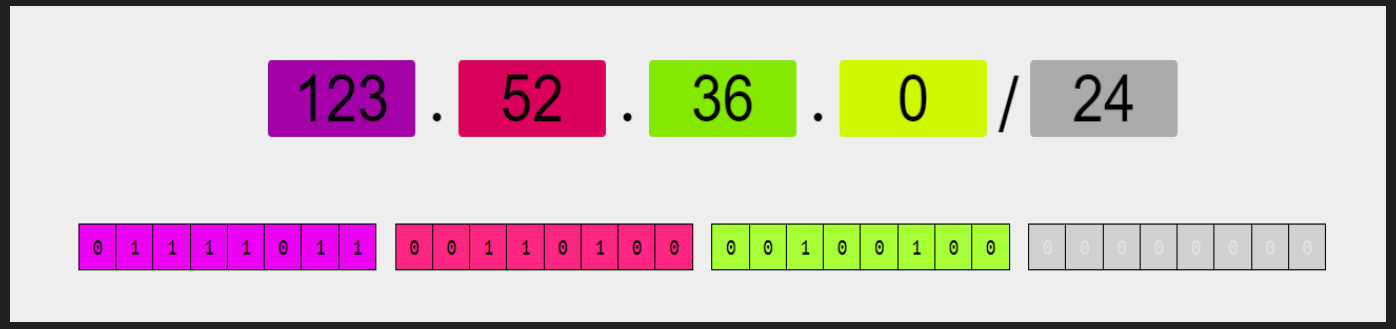
123.52.36.0
123.52.36.1
123.52.36.2
123.52.36.3
123.52.36.4
123.52.36.5
123.52.36.6
123.52.36.7
123.52.36.8
123.52.36.9
123.52.36.10
123.52.36.11

123.52.36.0 / 24

123.52.36.0/24

CIDR notation

123.52.36.0/24



123.52.36.0
123.52.36.1
123.52.36.2
123.52.36.3
123.52.36.4
| |
| |
| |
| |
| |
123.52.36.254
123.52.36.255

CIDR Notation

123.52.36.0/28	28 bits are fixed	4 bits are variable	Total IP address – 2^4 = 16
123.52.36.0/31	31 bits are fixed	1 bit is variable	Total IP address – 2^1 = 2
0.0.0.0/32	32 bits are fixed	0 bits are variable	Total IP address – 2^0 = 1
0.0.0.0/0	0 bits are fixed	32 bits are variable	Total IP address – 2^{32} = 4,294,967,296

RFC 1918

Standard for Private IP addressing

Class	Internal Address Range	CIDR Prefix
A	10.0.0.0 – 10.255.255.255	10.0.0.0/8
B	172.16.0.0 – 172.31.255.255	172.16.0.0/12
C	192.168.0.0 – 192.168.255.255	192.168.0.0/16

© ANKIT MISTRY – GOOGLE CLOUD

VPC & Subnets

BY ANKIT MISTRY

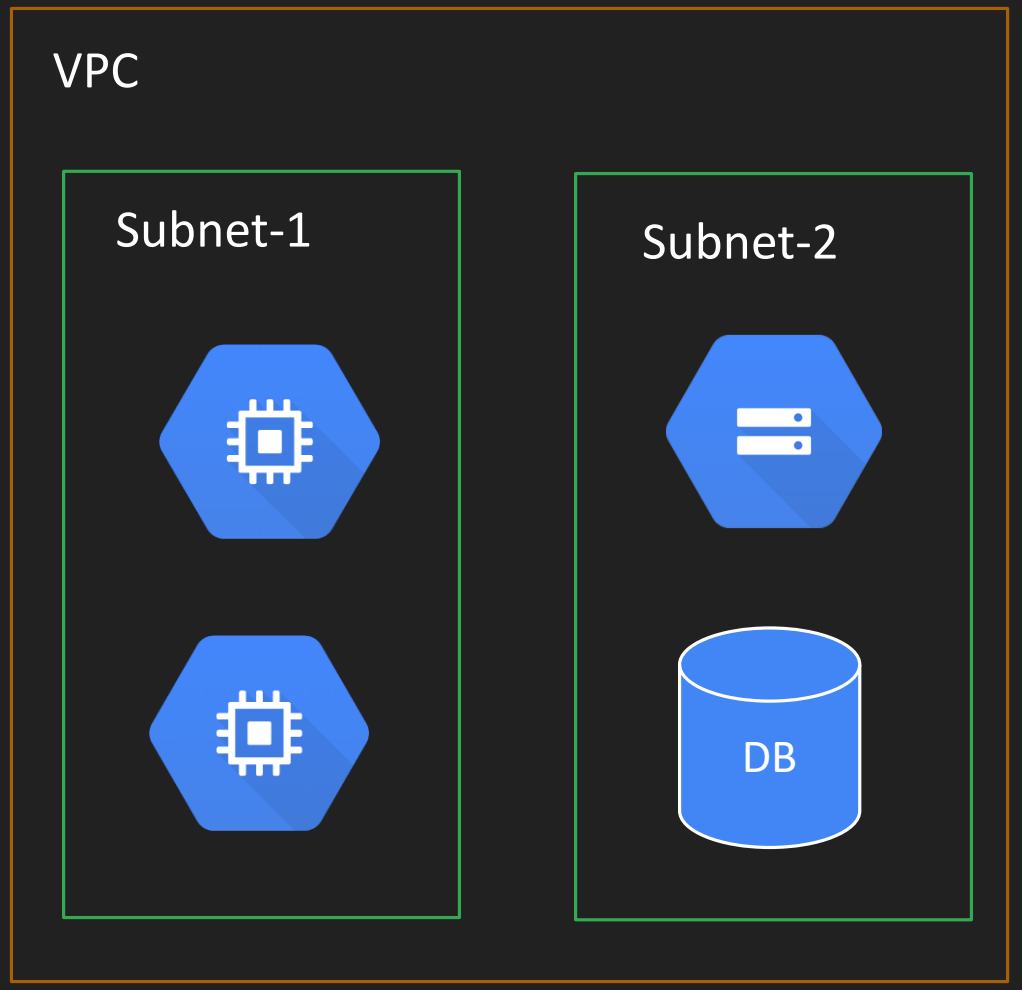
© ANKIT MISTRY – GOOGLE CLOUD

VPC – Subnetworks

- No Network -> No Cloud
- Virtual version of a physical network
- Networks are part of projects
- It's Global resources
  - Does not belong to any Region
- Placeholder to keep your resources
- Max 5 VPC per project
- No IP Assigned to VPC
- Network contain subnets
- Subnets are used for segregate resources
- Subnets has IP ranges
  - Expressed as CIDR notation
- VPC must have minimum one subnet
- Subnet belongs to one single region in GCP

© ANKIT MISTRY – GOOGLE CLOUD

VPC - Subnetworks



© ANKIT MISTRY – GOOGLE CLOUD

Default VPC

BY ANKIT MISTRY

© ANKIT MISTRY – GOOGLE CLOUD

Skip Default Network Creation Org Policy

BY ANKIT MISTRY

© ANKIT MISTRY – GOOGLE CLOUD

Create VM with Default VPC

BY ANKIT MISTRY

© ANKIT MISTRY – GOOGLE CLOUD

Avoid Default VPC

BY ANKIT MISTRY

© ANKIT MISTRY – GOOGLE CLOUD

Avoid Default VPC

- Lots of unnecessary subnets
- Same name – confusion
- Broad ranges in IP address
- Can not delete subnet
- Default Firewall rules are broad
- Can not go beyond /16

© ANKIT MISTRY – GOOGLE CLOUD

Reserved IP Address in Subnet

Reserved IP addresses in IPv4 subnet ranges

There are four reserved IP addresses in each subnet's primary IPv4 range. There are no reserved IP addresses in the secondary IPv4 ranges.

Reserved IP address	Description	Example
Network	First address in the primary IP range for the subnet	`10.1.2.0` in `10.1.2.0/24`
Default gateway	Second address in the primary IP range for the subnet	`10.1.2.1` in `10.1.2.0/24`
Second-to-last address	Second-to-last address in the primary IP range for the subnet that is reserved by Google Cloud for potential future use	`10.1.2.254` in `10.1.2.0/24`
Broadcast	Last address in the primary IP range for the subnet	`10.1.2.255` in `10.1.2.0/24`

https://cloud.google.com/vpc/docs/subnets#ipv4-ranges

© ANKIT MISTRY – GOOGLE CLOUD

Types of VPC

Default	Auto	Custom
• Created when compute engine API enabled • Every project has default VPC • There is one subnet per regions	• With Auto mode, Default VPC can be created • Fixed subnetwork ranges per region • Can expand from /20 to /16 • Default firewall can be added easily.	• No Subnet automatically created • Subnet creation manual • Custom IP range allocation • No necessary to create subnet in each region

© ANKIT MISTRY – GOOGLE CLOUD

Create Default Network - Auto Mode

BY ANKIT MISTRY

© ANKIT MISTRY – GOOGLE CLOUD

Create Custom VPC

BY ANKIT MISTRY

© ANKIT MISTRY – GOOGLE CLOUD

Create VM with Custom VPC

BY ANKIT MISTRY

© ANKIT MISTRY – GOOGLE CLOUD

© ANKIT MISTRY – GOOGLE CLOUD

Cloud IAP

BY ANKIT MISTRY

© ANKIT MISTRY – GOOGLE CLOUD

IAP

- Identity aware proxy
- IAP provides a single point of control for managing user access to web applications and cloud resources.
- Manage Http & SSH based resources
- Demo1
  - SSH with just Private IP address
  - Protect Compute Engine SSH Resources, Assign secured tunnel user role
- Demo2
  - Secure Google App engine http resources
  - Assign web app user role
- Demo3
  - Firewall rule - allow SSH to VM(Private IP only) just from browser

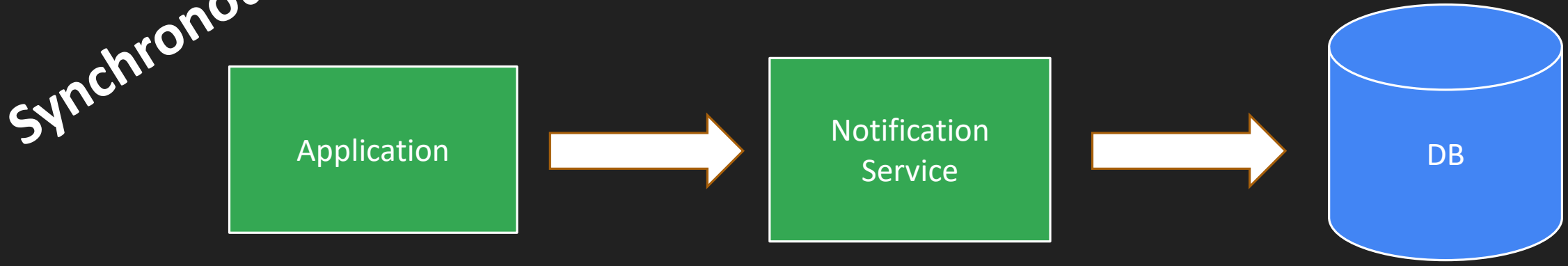
© ANKIT MISTRY – GOOGLE CLOUD

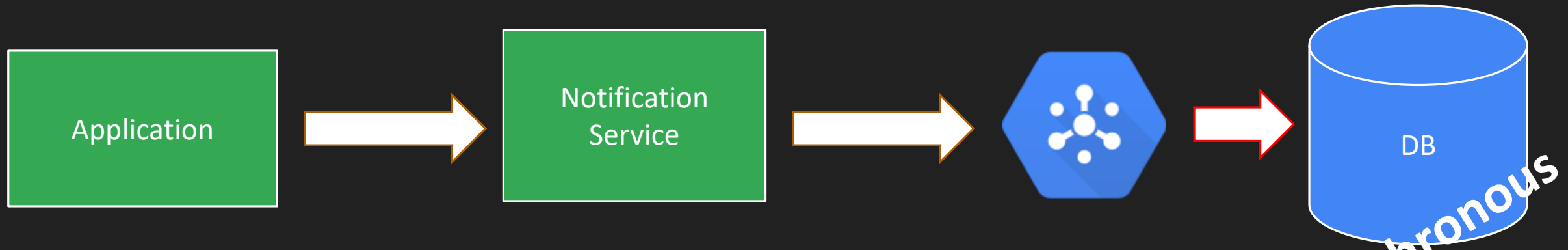
Google Cloud PubSub

© ANKIT MISTRY – GOOGLE CLOUD

PubSub

Synchronous





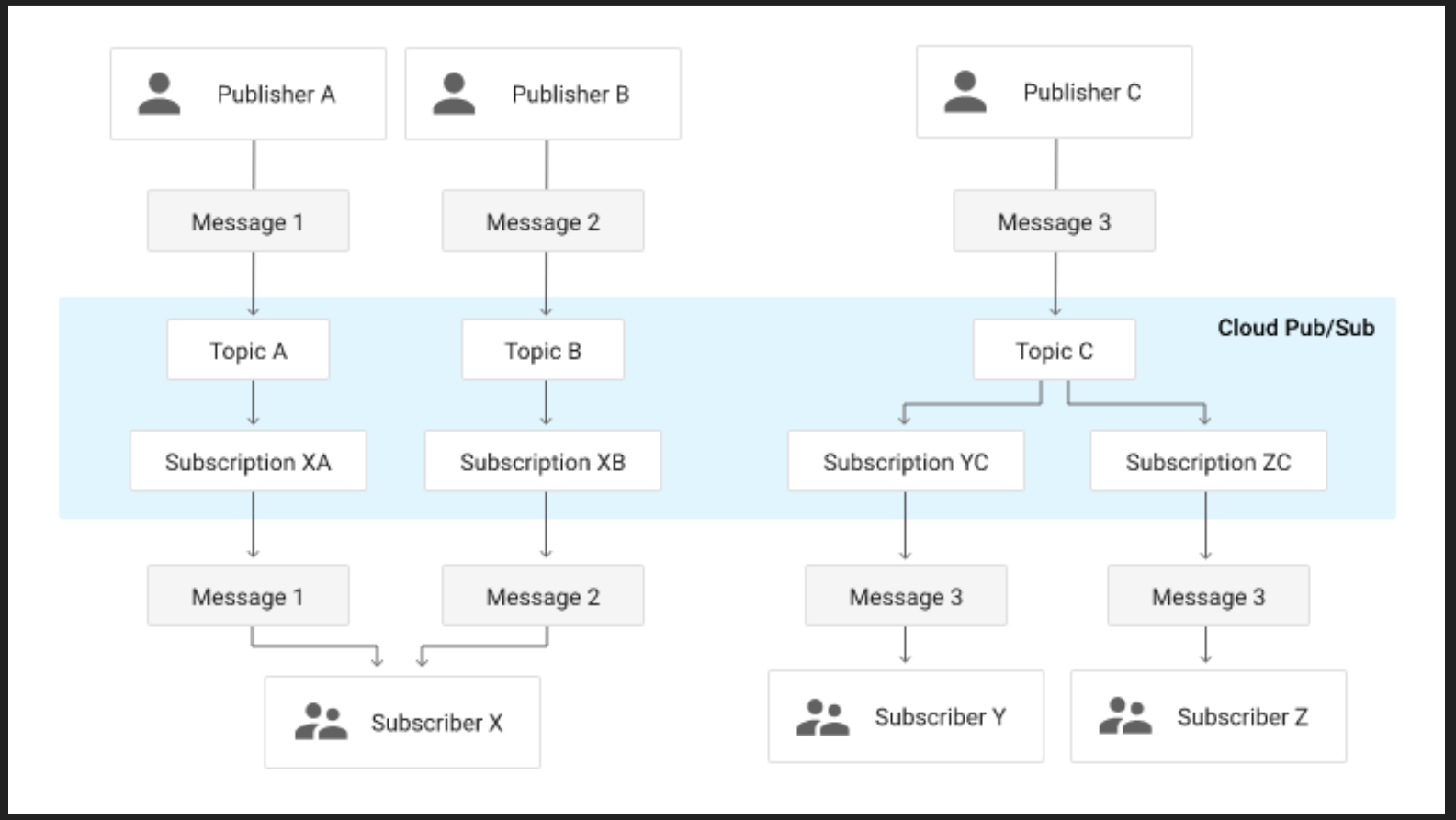
Asynchronous

© ANKIT MISTRY – GOOGLE CLOUD

How PubSub works

- Fully-managed asynchronous messaging service
- Scale to billions of message per day
- Publisher – App send message to Topic
- Push & Pull way to access messages
  - Pull – Subscriber pull message
  - Push – Message will be sent to subscriber via webhook
- One topic – Multiple Subscriber
- One subscriber – Multiple Topic

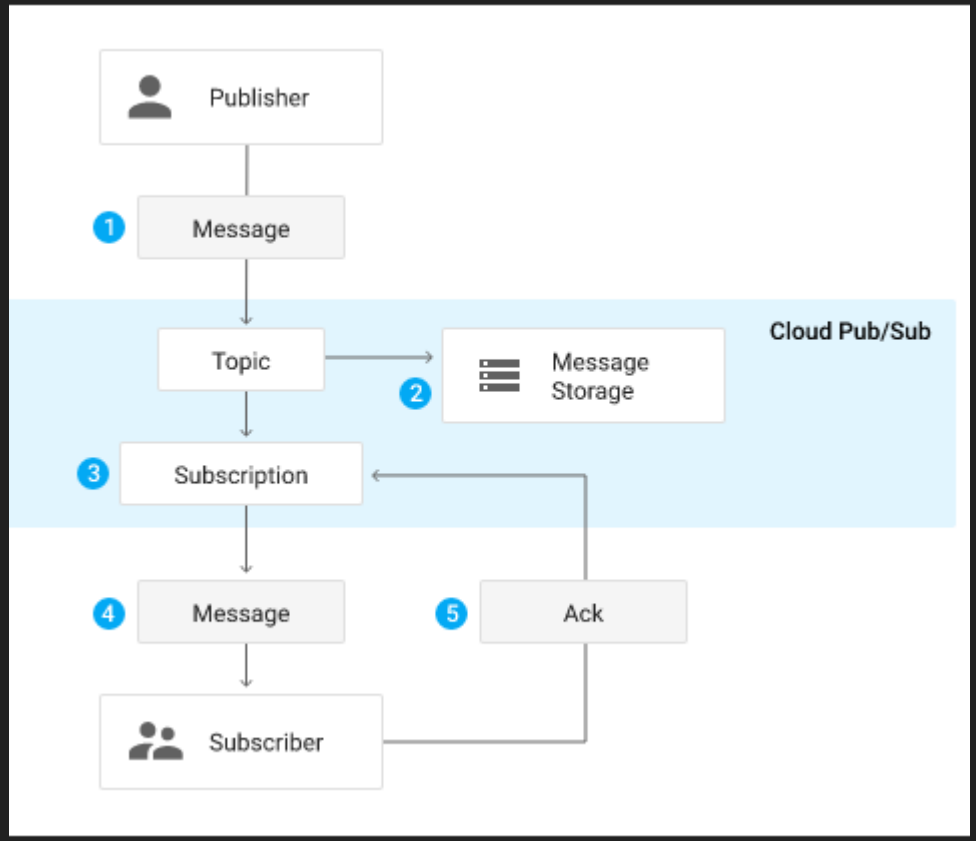
https://cloud.google.com/pubsub/docs/overview



© ANKIT MISTRY – GOOGLE CLOUD

Cloud PubSub

- Fully-managed Pubsub system inside Google Cloud
- Serverless
- Auto-scaling and auto-provisioning with support from zero to hundreds of GB/second
- Topic – Storage reference
- Publisher send message to topic at pubsub.googleapis.com
- Push – Pull way to access message
- Once subscriber receive message ack is sent.
- Cloud Pubsub act as staging environment for many GCP services



© ANKIT MISTRY – GOOGLE CLOUD

Advantage PubSub

- Durability of data will increase
- Highly Scalable, Scalable
- Decoupling between both system (Publisher & Subscriber)
  - Application don’t synchronously communicate with Notification service
  - Application (Publisher) is not dependent on Notification service (Subscriber)

© ANKIT MISTRY – GOOGLE CLOUD

[Hands-on] Cloud PubSub

© ANKIT MISTRY – GOOGLE CLOUD

© ANKIT MISTRY – GOOGLE CLOUD

SRE - Site Reliability engineering

BY ANKIT MISTRY

© ANKIT MISTRY – GOOGLE CLOUD

SRE

- History of Software Development Cycle
- DevOps & SRE
- Role of SRE
- Eliminating Toil
- Blameless Postmortem
- SLI, SLO & SLA
- Error Budgets

© ANKIT MISTRY – GOOGLE CLOUD

History

- History of Software Development Cycle
- DevOps & SRE
- Role of SRE
- Eliminating Toil
- Blameless Postmortem
- SLI, SLO & SLA
- Error Budgets

© ANKIT MISTRY – GOOGLE CLOUD

History

DEVELOPERS

- Developer write code
- Update software
- Adding new feature
- Don’t bother about stability
- They want to push code faster to Prod

OPERATORS

- Operator know how to deploy & monitor application
- Operators don’t know how to write code
- They know how to assemble code
- Solve Production issue
- How to scale Application
- They love less updates

© ANKIT MISTRY – GOOGLE CLOUD

Devops

- DevOps is a set of practices, guidelines and culture
  - which designed to reduce the gap between software development and software operations.
- If Both team work together, productivity will increase
- DevOps established five goals.
  - Reduce organizational silos
  - Accept failure as normal
  - Implement gradual changes
  - Leverage tooling and automation
  - Measure everything

© ANKIT MISTRY – GOOGLE CLOUD

SRE

- There is problem with devops
  - Goal of Devops is broad.
  - Devops does not define how to implement it.
- This is How SRE comes.
- Devops is Philosophy where SRE is implementation of Devops ‘s Philosophy.
- class SRE implements DevOps
- SRE Practices
  - SRE Role
  - blameless postmortems
  - error budget
  - reduce toil
  - track service level metrics , SLIs, SLOs, and SLAs.

© ANKIT MISTRY – GOOGLE CLOUD

SRE Role

- Specific job role
- Old operator role -> SRE Role
- A Site Reliability Engineer is basically the result of asking a software engineer to design an operations team
- SRE requires experience in both development as well as operations
- SRE spends half of their time doing ops-related work
  - production issues, attending call, performing manual interventions
- SRE spends other half of their time in development task, Scaling system, automation
- Compared to old operator, both SRE & Developer share responsibility of Prod Server
- SREs build the tools that developers use to compile, test, and deploy their code. (CI/CD Pipeline)
- Developers and SREs work together to fix issue

© ANKIT MISTRY – GOOGLE CLOUD

Blameless postmortems

- One of goal of DevOps is accept failure as normal
- Failure is un-avoided, However good system you design.
- Once you change system, risk is involved
- If your rate of change is zero, risk is also zero. But that means you are stopping growth.
- Need to balance between change & risk.
- You can take it as opportunity to grow business, if Things break, fix it.
- Fix will teach you lot of thing, minimize future issue.
- In SRE, you can accomplish with Blameless postmortems

© ANKIT MISTRY – GOOGLE CLOUD

Blameless postmortems (Cntd…)

- Idea behind Blameless postmortems
  - is to analyze system failure
  - Root cause behind it.
  - Discuss about what has happened exactly
  - What action need to be performed.
- Not to look for someone who can be blamed.
- Assumption is – everyone had good intentions
- Some postmortems question need to be asked.
  - When incident begin & end?
  - How incident get notified
  - Who are all involved
  - Which system are affected
  - What is root cause of failure
  - How to avoid in future

© ANKIT MISTRY – GOOGLE CLOUD

Blameless postmortems (Cntd…)

- Accept that With Human error are involved.
- Blameless postmortems is
  - Honest Communication with other team member so that similar incident can be avoided in future

© ANKIT MISTRY – GOOGLE CLOUD

Toil

- One of goal of DevOps is leverage tooling and automation
- There is lots of task are manual, laborious.
- Task like Password Change, Copy Files, Creating new Folders, Restart Servers
- These type of task are considered as Toil.
- Identifying Toil is important.
- Not all Task are Toil.
- There are task which is laborious but not necessary is toil.
- Toil is related
  - Prod system
  - Manual, repetitive & automatable task

© ANKIT MISTRY – GOOGLE CLOUD

Toil (Cntd…)

- SRE want to reduce Toil by automation.
- Task like
  - Automate CI/CD Pipeline
  - Schedule Jobs
  - Write some Automation scripts
  - Automate testing
  - No manual Provisioning hardware
- If Repetitive task automated, It should be automated
- Due to Automation, more resource can work something more interesting
- SRE should spend significant amount of time in reducing toil.

© ANKIT MISTRY – GOOGLE CLOUD

Error budget

- One of goal of DevOps is implement gradual change
- Why outage occurs
  - Added new feature, change, new hardware, security patches
- More change leads to less stable system
- How to balance between change & stability
- We have to define metric for high system reliability.
- It is business Problem
- how much can the service fail before it begins to have a significant negative impact?
- How quickly do we need to be able to release new features?
- Depending on target, need to define error budget

© ANKIT MISTRY – GOOGLE CLOUD

Error budget (Cntd…)

- Anytime your service is down, time require to recover it will be consumed from error budget
- After you define error budget
  - as long as you are within error budget, you are good to go for more changes
  - Once you run out of error budget, need to hold all future changes for deployment & make system stable first
- Larger error budget
  - means more downtime for service acceptable,
  - frequent changes possible.
- Less error budget,
  - means less downtime for service acceptable,
  - lesser changes allowed.
- Error budget make sure smaller & gradual changes deployed.

© ANKIT MISTRY – GOOGLE CLOUD

SLI, SLO & SLA

BY ANKIT MISTRY

© ANKIT MISTRY – GOOGLE CLOUD

SLO

- Service level objective
- It is internal objective of team
- SLO is something everyone in org want to achieve
- Error Budget is directly related to SLO
- It kind of complement to Error Budget
- Error – 3% means service is down 3% at max
- SLO – 97% means service should be up for 97%
- Error Budget + SLO = 100%
- Define SLO with respect to latency, Availability, Response Time

© ANKIT MISTRY – GOOGLE CLOUD

SLI

- Service level indicator
- Indicator internal to team
- SLI needs to be compared against SLO
- SLI are metrics which track over time (generally 5 minutes interval)
- SLI ranges from 0 to 100%

$$SLI = \frac{Total\ Good\ Event}{Total\ Valid\ Event}\ X\ 100$$

- Let's say SLO – 96%
  - 96% of request should be serve within 300 ms latency.
- If Current SLI is 95% or anything less than 96%, system is under performing.
- SLI help us to find which service are not performing as per SLO
- Good SLI leads customer happy

© ANKIT MISTRY – GOOGLE CLOUD

SLI (Cntd…)

- Good SLI leads customer happy
- If Changes to SLI does not impact customer, SLI definition is not worth
- Different signal to track
  - Latency
  - Traffic
  - Errors
  - Saturation
  - Availability of system
- Selecting right SLO & SLI will lead to success

© ANKIT MISTRY – GOOGLE CLOUD

SLA

- Service level agreement
- It is contract with consequences of failing to meet the SLOs they contain
- SLO & SLA are quite similar
- But your SLAs should not be the same as your SLOs
- SLO is an internal objective,
  - If you can not meet SLO, team can slow down changes
- SLAs violations are shared with your customers
  - If you can not meet SLA, compensate need to be provided to customers
- https://cloud.google.com/terms/sla
- SLI should be higher than SLO & SLA, means current indicator shows services are performing as expected

© ANKIT MISTRY – GOOGLE CLOUD

SLA (Cntd…)

- If SLI goes below SLO, slow own
- If SLI goes below SLA, notify customer & compensate
- Higher SLA Good but more likely you will violate it
- Lesser SLA means You will meet but customer will have less confident in your services
- Google recommendation in case very high SLA
  - Down your service for some time

© ANKIT MISTRY – GOOGLE CLOUD

© ANKIT MISTRY – GOOGLE CLOUD

CI/CD Pipeline

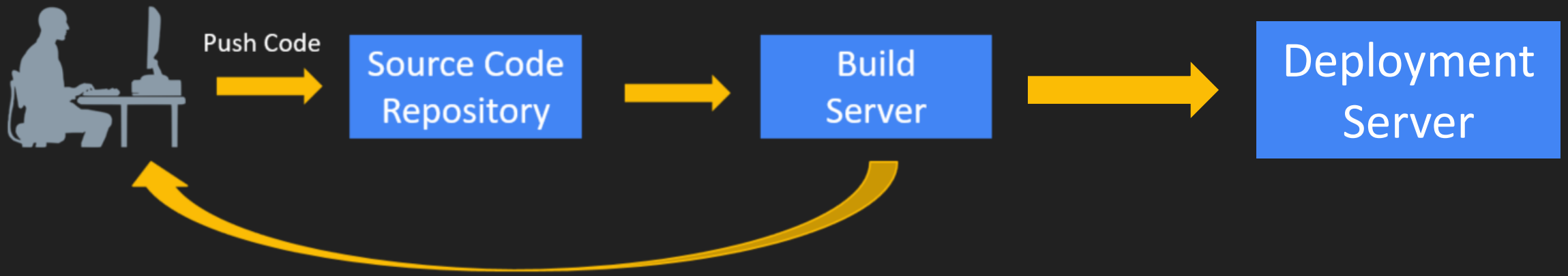
BY ANKIT MISTRY

© ANKIT MISTRY – GOOGLE CLOUD

Continuous integration

© ANKIT MISTRY – GOOGLE CLOUD

Continuous Deployment



© ANKIT MISTRY – GOOGLE CLOUD

Continuous Deployment vs Continuous Delivery

- Continuous Deployment
  - Fully automated, no manual intervention
  - Code is continuously build & deploy
- Continuous Delivery
  - Release to Production
  - May involve manual approval
  - It will make sure delivery are often & fast
  - Before Continuous Delivery, frequency of release usually one in 3-month
  - Now, Possible to release 5 times in day

© ANKIT MISTRY – GOOGLE CLOUD

Different Services for CI/CD

BY ANKIT MISTRY

© ANKIT MISTRY – GOOGLE CLOUD

Source Code management

Source Code Management

Cloud Source Repository

© ANKIT MISTRY – GOOGLE CLOUD

Build

Jenkins

Build

Teamcity

circleci

Cloud Build

© ANKIT MISTRY – GOOGLE CLOUD

Artifact Storage

Artifact Storage

Container Registry

Artifact Registry

© ANKIT MISTRY – GOOGLE CLOUD

Deployment

Deployment

Compute Engine

Kubernetes

App Engine

Cloud Run

Cloud Function

© ANKIT MISTRY – GOOGLE CLOUD

Cloud Build

BY ANKIT MISTRY

© ANKIT MISTRY – GOOGLE CLOUD

Cloud Build

- Serverless CI/CD Platform in GCP
- Fully managed CI/CD workflows
- Build software quickly across all programming languages, including Java, Go, Node.js and more
- Deploy across multiple environments such as VMs, serverless, Kubernetes, or Firebase
- Nice integration with other GCP services
  - SCR – Source code repository
  - Various compute platform
- All sequence of steps can be defined in cloudbuild.yaml file
- Every step has some cloud builder defined.
  - https://github.com/GoogleCloudPlatform/cloud-builders
  - https://github.com/GoogleCloudPlatform/cloud-builders-community
- How to create builder??
- Daily free quota

© ANKIT MISTRY – GOOGLE CLOUD

CI/CD Pipeline Demo

© ANKIT MISTRY – GOOGLE CLOUD

© ANKIT MISTRY – GOOGLE CLOUD

Cloud Operation Tool

BY ANKIT MISTRY

© ANKIT MISTRY – GOOGLE CLOUD

Operation Tool

- Operation like Monitoring, Logging
- Why Logging – Monitoring is required
- What is Logging
- Kinds of Log – Audit Logs
- Log Collection
- Log Routing
- Log Export
- Cloud Monitoring – Metrics, Dashboard, Uptime check, Alerts
- Cloud Debugger, Trace, Profiler, Error Reporting

© ANKIT MISTRY – GOOGLE CLOUD

Why such tool

- Software Development + Maintenance
- Everyone want their software run smoothly
- But No software is bug free
- issues come at dev stage, Test or Prod level
- How to find root cause behind it
- You need to continuously monitor resources
  - Space is sufficient
  - Is application is slow
  - Is CPU usage going beyond 90%
  - Who did What with Prod (even if by mistake)
- So to know all those answer & many more, such tool is required

© ANKIT MISTRY – GOOGLE CLOUD

Cloud Monitoring

- Monitor various cloud Resources
- Different Metrics can be measured
- Monitor one or more GCP Project or AWS Account
- Workspace
  - Multiple metrics can be added
- Default workspace & custom workspace
- Let’s see in action – Monitoring UI

© ANKIT MISTRY – GOOGLE CLOUD

Cloud Logging

- Log management tool
- Fully managed service
- Store Exabyte scale data
- Log can collected from multiple source
- Search & analyze log
- Let’s explore Logging UI
  - Logs Explorer, Dashboard, Log Metrics, Logs Router

© ANKIT MISTRY – GOOGLE CLOUD

Types of Cloud Audit Logging

who did what, when, where

Admin activity

- By Default Enabled
- Administrative action
- 400 days
- Free
- Create VM, Delete VM
- Can not Configure, Can not Disable

System Event

- By Default Enabled
- Generated by Google System
- 400 days
- Free
- VM Migration, Preemtive VM
- Can not Configure, Can not Disable

Data Access

- By Default **Not** Enabled
- Create, modify Resource Data
- 30 days
- **Not** Free
- Create Object in Bucket
- Can be disable

Policy Denied

- By Default Enabled
- Google Service denies access
- 30 days
- **Not** Free
- Security violation
- Can not be disabled. But can be excluded with Filters

[Hands-on] Cloud audit Logging

BY ANKIT MISTRY

© ANKIT MISTRY – GOOGLE CLOUD

Explore Audit Log Structure

BY ANKIT MISTRY

© ANKIT MISTRY – GOOGLE CLOUD

Log Collection

- Log read/write via gcloud SDK
- Automatically
  - Cloud Run, GKE, App Engine
- Logging Agent
  - For Compute Engine on Google cloud / AWS VM
  - Legacy agent/ Ops agent
- Cloud Logging API
  - Python/Java SDK
  - From On-premises

© ANKIT MISTRY – GOOGLE CLOUD

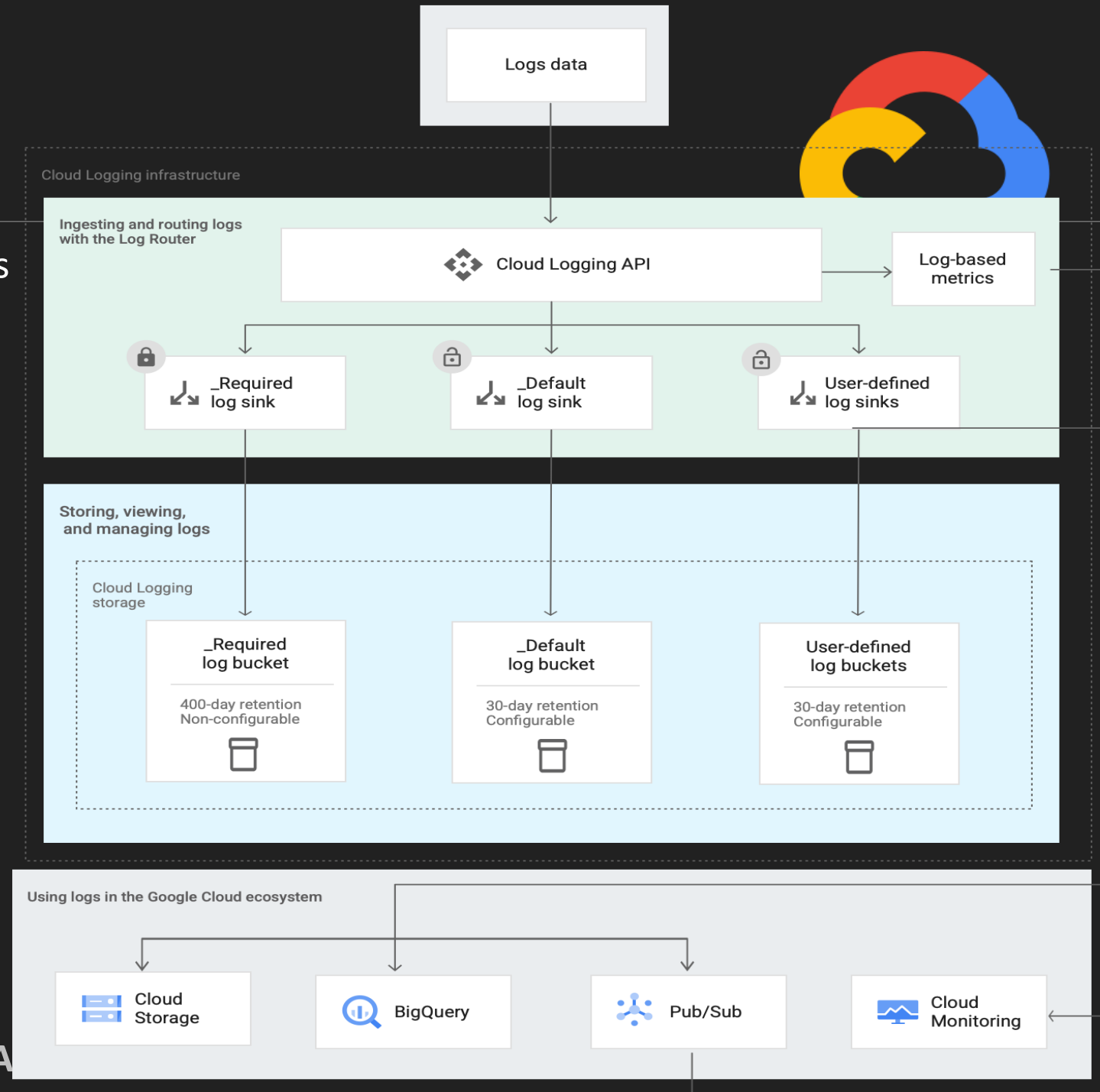
[Hands-on] Log based metrics

BY ANKIT MISTRY

© ANKIT MISTRY – GOOGLE CLOUD

Log Router

- Log arrives at Log Router from various sources
- From Router, diverted to various sink
- Two types of Log Bucket
  - _Required
  - _Default
- Logs can be routed to User defined Bucket
- Sinks
  - BigQuery
  - Cloud Storage
  - PubSub



© A

More Ops/Dev Tool

- Cloud error reporting – detect error
- Cloud Debugger – Find state of running application
- Cloud Trace - latency
- Cloud Profiler – How much resource consumed

© ANKIT MISTRY – GOOGLE CLOUD

THANK YOU

© ANKIT MISTRY – GOOGLE CLOUD