Data Loss Prevention API

- automatically detect a sensitive data and mask. Los Angeles[LOCATION], HIROKO[PERSON_NAME],

Private Google Access

- VM that only have internal IP addresses (no external IP addresses) can use Private Google Access.
- They can reach the external IP addresses of Google APIs and services.
- VM + internal IP
- No effect public IP
- Private Google Access -> Internal IP -> can access an external Google services
- subnet-a, Private Google Access ON, 10.240.0.2 ---> Can access
- subnet-a, Private Google Access ON, 10.240.0.3 + public IP ---> CAN access
- subnet-b, Private Google Access OFF, 192.168.1.2 ---> CANNOT access
- subnet-b, Private Google Access OFF, 192.168.1.3 + public IP ---> CAN access
- https://cloud.google.com/vpc/docs/private-google-access#example

Cloud VPN

- Virtual Private Network(VPN), Virtual Private Cloud(VPC)
- on-prem + VPC = Cloud VPN
- VPC: two subnets(us-east1, us-west1) => communicate internal IP with routing, Cloud gateway, on-prem gateway, two tunnels
- Cloud VPN = static or dynamic routes=> set up Cloud Router with BGP(border gateway protocol)
- Cloud VPN gateway with dynamic routing ---> HA VPN Border Gateway Protocol (BGP).
- The highest level of availability, use HA VPN whenever possible.
- Cloud DNS offers DNS forwarding zones and DNS server policies to allow lookups of DNS names between your on-premises and Google Cloud environment. [1]

1. DNS forwarding zones

Billing

Billing Accout Administrator

- Manage billing accounts (but not create them).
- Who can create? billing.accounts.create role. -> Billing Account Creator

Billing Account User

- Link projects to billing accounts.

  > This role has very restricted permissions, so you can grant it broadly. When granted in combination with Project Creator, the two roles allow a user to create new projects linked to the billing account on which the Billing Account User role is granted. Or, when granted in combination with the Project Billing Manager role, the two roles allow a user to link and unlink projects on the billing account on which the Billing Account User role is granted.

- With Project Creator ==> create new projects linked to the billing account
- With Project Billing Manager ==> link/unlink projects on the billing account

Project Billing Manager

- resourcemanager.projects.createBillingAssignment
- resourcemanager.projects.deleteBillingAssignment
- Link/unlink the project to/from a billing account.
- When granted in conjunction with the Billing Account User role, provides access to assign a project's billing account or disable its billing.

  > When granted in combination with the Billing Account User role, the Project Billing Manager role allows a user to attach the project to the billing account, but does not grant any rights over resources. Project Owners can use this role to allow someone else to manage the billing for the project without granting them resource access.

- A billing account is assosicated with org.
- Billing alert - Avoid surprises on your bill, who -> Billing Account Administrator, Billing Account Costs Manager
- Moving a project from one organization resource to another won't impact billing, and charges will continue against the old billing account. []
- Billing data -> export BQ -> visualize = Data Studio
- Q: You need to create a new billing account and then link it with an existing Google Cloud Platform project. What should you do?
- A: Billing Account Admin can NOT create a billing account,link/unlink is Project Billing Manager

links:

1. Overview of Cloud Billing roles in IAM
2. Create a budget
3. Change the billing account for a project

microservices

- microservices, automation

Machine Family

- M1 machine type: Memory-optimized - "M", Compute-optimized - "C", cost-optimized: "E", Balanced price/performance: "N"
- Local SSD - When you stop a VM, all data on the local SSD is discarded.
- Unlike Persistent Disks, Local SSDs are physically attached to the server on VM.

Cloud SQL

- Backup(Data Protection) 1)automate backup is everyday setup only. 2) point-in-time(on-demand)
- Transactional and a single physical location = Cloud SQL.
- Region: us-central1, Single zone - in case of outage, no failover no recommended.
- Region: us-central1, dMultiple zones Automatic failover to another zone - recommended.

Spanner

- Spanner, CPU utilization, Cloud Monitoring, scaling
- Spanner is used for global scaling.
- primary key -> automatically created
- you should be careful when choosing a primary key in the schema design to not accidentally create hotspots in your database. One cause of hotspots is having a column whose value **monotonically** increases as the first key part,
- AutoScaling - Alerts for high CPU utilization.(45%, 65%)

1. Choose a primary key to prevent hotspots
2. https://cloud.google.com/architecture/autoscaling-cloud-spanner

VM

- SSH connection/VM: **enable-osLogin=true** with roles/compute.osLogin or roles/compute.osAdminLogin.
- https://medium.com/infrastructure-adventures/centralized-ssh-login-to-google-compute-engine-instances-d00f8654f379
- RDP: Windows login, reset password, download RDP client
- Network tags: it makes FW enable in a vm.
- batch job, preemptible vm
- (maintenance)Maintenance occurs - On host maintenance=Migrate VM instance, Automatic restart=ON
- (maintenance) Live Migration - The reason A is not correct is because live migration simply moves an existing VM between hosts, no attributes or properties are changed otherwise. Hence, you cannot live migrate from 1 VM type to another.
- Stop VM when increasing the memeory 4GB -> 8GB

- disk - persistant disc Local SSD
- Q:how to login using Cloud Identity Proxy for VM Access a paticular instance
- A:"without allowing other instances" , the other instances are created with default compute engine service account. So you must create a new independant service account
- Q: multiple VMs without public IP, you want to access all VMs without having to configure specific access on the existing and new instances-->
- A: You can connect to Linux instances that don't have an external IP address by tunneling SSH traffic through IAP.[1]
- "Risks of manual key management

  > If you create and manage public SSH keys yourself through the Cloud Console, the gcloud command-line tool, or the API, you must keep track of the used keys and delete the public SSH keys for users who no longer have access. For example, if a team member leaves your project, remove their public SSH keys from metadata, so they can't continue to access your instances. https://cloud.google.com/compute/docs/instances/access-overview

- (SA)Create a new service account rather than using the Compute Engine default service account.[2]
- (SA)each VM + SA with with the minimum permissions necessary.[2]

https://cloud.google.com/compute/docs/access/create-enable-service-accounts-for-instances#best_practices

1. Tunneling SSH connections
2. Authenticate workloads using service accounts

MIGs

- gradually deploy - maxUnavailable(How many instanes can be offline), maxSurge(How many extra instances to temporarily create)

```
gcloud compute instance-groups managed rolling-action start-update
INSTANCE_GROUP_NAME
--version=template=[INSTANCE_TEMPLATE_NAME]
--max-surge=[MAX_SURGE]
--max-unavailable=[MAX_UNAVAILABLE]
```

- Autoscaling policies: CPU utilization, Monitoring metrics, Queue-based workload, Load balancing capacity
- Autohealing - Autohealing recreates VM instances
- Initial delay - 300 seconds (5 min)
- Pro Tip: Use separate health checks for load balancing and for autohealing. Health checks for load balancing detect unresponsive instances and direct traffic away from them. Health checks for autohealing detect and recreate failed instances. [1]

- If your MIG cannot create or recreate instances - 1. The boot disk already exists, 2. The instance template is not valid[2]
- Rolling Update in MIG - gradually update a template in MIG
- Canary update - We roll out a software update to a small part of the users first,

1. https://cloud.google.com/compute/docs/tutorials/high-availability-autohealing#create_the_health_check
2. https://cloud.google.com/compute/docs/troubleshooting/troubleshooting-migs

LB

- HTTP LB --> global, HTTP or Port(80/8080), HTTPs on port 443, URL maps
- SSL Proxy LB --> global, encrypted, non-HTTP(s) traffic, TCP with SSL offload.
- TCP proxy LB --> glocal, unencrypted, non-HTTP traffic, TCP.
- Network LB --> regional, non-proxied LB, Traffic: UDP, TCP/UDP ports.
- Internal TCP/UDP --> regional, private LB, VM in same region, TCP/UDP traffic

> Q61:You have an instance group that you want to load balance. You want the load balancer to terminate the client SSL session. The instance group is used to serve a public web application over HTTPS.

- A: Configure an HTTP(S) load balancer -> global, HTTP or Port(80, 8080), HTTPs port 443, URL map**
- B: Configure an internal TCP load balancer. -> regional, private LB, TCP/UDP
- C: Configure an external SSL proxy load balancer.-> global, encrypted, non-HTTP(s)
- D: Configure an external TCP proxy load balancer. ->global, unencrypted, non-HTTP traffi

> Q148: You have an application that receives SSL-encrypted TCP traffic on port 443. Clients for this application are located all over the world. You want to minimize latency for the clients. Which load balancing option should you use?

- A: HTTPS Load Balancer -> global, HTTP or port - 80, 8080, HTTPS on 443, url map
- B: Network Load Balancer ->regional, non-proxied LB, UDP, TCP/SSL ports
- C: SSL Proxy Load Balancer --> global, encrypted, non-HTTP(s), TCP.**
- D: Internal TCP/UDP Load Balancer. -> regional, private LB, VM in same region, TCP/UDP traffic

> Q188: Your company developed a mobile game that is deployed on Google Cloud. Gamers are connecting to the game with their personal phones over the Internet. The game sends UDP packets to update the servers about the gamers actions while they are playing in multiplayer mode. Your game backend can scale over multiple virtual machines (VMs), and you want to expose the VMs over a single IP address. What should you do?

- A: Configure an SSL Proxy load balancer in front of the application servers.-> global, encrypted, non-HTTP(s)
- B: Configure an Internal UDP load balancer in front of the application servers.-> regional, private LB, VM in same region. TCP/UDP
- C: Configure an External HTTP(s) load balancer in front of the application servers.-> global, HTTP 80/8080, HTTPS 443, URL map

- D: Configure an External Network load balancer in front of the application servers. -> UDP, TCP/UDP ports,**

FW

- Request -> egress, Response -> ingress.
- egress is leaving
- priority: 0 to 65535
- Highest priority: 65535
- Default priority: 1000
- you always need to enable the ingress traffic as this is **never** enabled by default.
- Implied rules - allow all egress, and **deney** all ingress

Cloud Storage

- Storage Admin - Grants full control of buckets and objects.
- Storage Object Admin - Grants full control over objects.
- Storage Object Creator - create objects in a bucket.
- Bucket public -> a signed URL
- failover
- Sensitive data -> Enable Data Access audit log[2]
- Google Cloud services write audit logs ->"Who did what, where, and when?" -> Admin Activity audit log/Data Access audit log
- lifecycle - Nearline, Coldline, and Archive
- enable them to write data into a particular Cloud Storage bucket --> Storage Object Creator
- Cloud Audit Logs with Cloud Storage
- gsutil rsync <source_location> <destination_location>.

1. IAM roles for Cloud Storage
2. Cloud Audit Logs with Cloud Storage

Audit log

Role	Audit Log Name
Logging.viewer	Admin Activity
Logging.viewer	Policy Denied
Logging.viewer	System Event
Logging.privateLogViewer	Data Access

- Has all permissions of Logging.privateLogViewer
- Data access audit log - disable as a default because it can be quite large
- except for BigQuery Data Access audit logs

1. Audit Logs: Querying Logs, Pricing and Retention

IAM

- 4 Members: Google Account, SA, Google Group, Google Workspace /Cloud Identity

Member	Email
Google Account	userid@gmail.com
SA	12345678@cloudservices.gserviceaccount.com
Google Group	groupname@googlegroups.com
Google Workspace Domain/Cloud Identity	test@example.com

- Cloud Identity - verify thrid party authentication
- G Suite = Google Workspace
- use predefined roles and create groups to control access to multiple users
- A Google group is a named collection of Google accounts
- A service account is an account that belongs to your application instead of to an individual end user
- A Google group is a named collection of Google accounts.
- A Workspace domain represents a virtual group of all the Google accounts
- Google Cloud customers who are not Workspace customers can get these same capabilities through Cloud Identity. Cloud Identity lets you manage users and groups using the Google Admin Console, but you do not pay for or receive Workspace's collaboration products such as Gmail, Docs, Drive, and Calendar.

SA

- To use a service account outside of Google Cloud, such as on other platforms or on-premises, you must first establish the identity of the service account. Public/private key pairs provide a secure way of accomplishing this goal.
- https://cloud.google.com/iam/docs/creating-managing-service-account-keys#iam-service-account-keys-create-rest

IAP(Identity-Aware Proxy)

- Identity Aware Proxy API(IAP)- https://cloud.google.com/iap/docs/external-identities
- Proxy(=dairi)
- external identities with Identity-Aware Proxy (IAP) instead of Google accounts.

- external identity: email/password, OAuth (Google, Facebook, Twitter, GitHub, Microsoft, etc.
- This is useful if your application is already using an external authentication system,
- IAP controls access to App Engine apps and VMs
- Q: Your company runs its Linux workloads on Compute Engine instances. Your company will be working with a new operations partner that does not use Google Accounts. You need to grant access to the instances to your operations partner so they can maintain the installed tooling.
- A:Enable Cloud IAP for the Compute Engine instances, and add the operations partner as a Cloud IAP Tunnel User.

1. Youtube - #GCP #IAP
2. https://cloud.google.com/architecture/identity/migrating-consumer-accounts

VPC

- auto or custom mode
- auto mode - one subnet from each region is automatically created within it.
- custom mode: you have to create a subnet
- VPC and the 2 subnets -> custom
- they need to communicate via private addresses which cannot be achieved with 2 VPCs without Network Peering)
- Two VPCs -> Network Peering
- Shared VPC - > a host project and service projects are using same VPC.
- you always need to enable the ingress traffic as this is never enabled by default.

BigQuery

- dry-run select * from coffee.coffee_dataset -->304538 bytes of data
- External table (source: Cloud Storage)
- BigQuery Job User((roles/bigquery.jobUser)) - can run query[1]

1. https://cloud.google.com/bigquery/docs/access-control#bigquery.jobUser

App Engine

- You specify the scaling type in your app's app.yaml.
- app.yaml: runtime, URL, scaling
- gradually deploy - a rolling-action start-update with maxSurge
- --split
- version in a same project
- Important: Each Cloud project can contain only a single App Engine application, and once created you cannot change the location of your App Engine application.[1]

1. https://cloud.google.com/appengine/docs/flexible/managing-projects-apps-billing#create

Deployment Manager

- gcloud deployment-manager deployments update my-deployment --config=new_config.yaml
- gcloud deployment-manager deployments update example-deployment
  --config configuration-file.yaml
  --preview
- **--preview**:Preview the requested create without actually instantiating the underlying resources. (default=False)
- A type provider exposes all of the resources of a third-party API to Deployment Manager as base types that you can use in your configurations. These types must be directly served by a **RESTful API** that supports Create, Read, Update, and Delete (CRUD). Youtube - Cloud Deployment Manager
- gcloud deployment-manager deployments create quickstart-deployment --config vm.yaml

```
resources:
- name: vm-created-by-deployment-manager
  type: compute.v1.instance
  properties:
    zone: us-central1-a
    machineType: zones/us-central1-a/machineTypes/n1-standard-1
    disks:
    - deviceName: boot
      type: PERSISTENT
      boot: true
      autoDelete: true
      initializeParams:
        sourceImage: projects/debian-cloud/global/images/family/debian-11
    networkInterfaces:
    - network: global/networks/default
```

Stackdriver

- different projects, monitor a single report
- Monitoring, Error Reporting, Tracing, Cloud Debugger
- Cloud Debugging - Inspect an application without stopping it or slowing it down significantly.
- Error Reporting - Aggregate and display errors for running cloud services

Monitoring

- Installing Logging agent
- Installing Monitoring agent

BigTable

- time series database

Cloud Function

```
gcloud functions deploy Hello
   --trigger-bucket=STORAGE_BUCKET  #Cloud Storage trigger
   --trigger-http                   #HTTP trigger
   --trigger-topic=TOPIC_NAME [1]   #PubSub trigger
```

- Q: You have created a **code snippet** that should be triggered **whenever a new file** is uploaded to a Cloud Storage bucket. You want to deploy this code snippet. What should you do?
- A:Use Cloud Functions and configure the bucket as a trigger resource.

Links:

1. https://cloud.google.com/functions/docs/calling/storage#event_types

Cross projects

- **Service accounts are both identities and resources**. Because service accounts are identities, you can let a service account access resources in your project by granting it a role, just like you would for any other principal.
- Q: You manage an App Engine Service that aggregates and visualizes data from BigQuery. The application is deployed with the default App Engine Service account. The data that needs to be visualized resides in a different project managed by another team. You do not have access to this project, but you want your application to be able to read data from the BigQuery dataset. What should you do?
- A: Ask the other team to grant your default App Engine Service account the role of BigQuery Data Viewer.
- Q:An application generates daily reports in a Compute Engine virtual machine (VM). The VM is in the project corp-iot-insights. Your team operates only in the project corp-aggregate-reports and needs a copy of the daily exports in the bucket corp-aggregate-reports-storage. You want to configure access so that the daily reports from the VM are available in the bucket corp-aggregate-reports-storage and use as few steps as possGrant the VM Service Account the role Storage Object Creator on corp-aggregate-reports-storage. ible while following Google-recommended practices. What should you do?
- A:Grant the VM Service Account the role Storage Object Creator on corp-aggregate-reports-storage.

gcloud commands

- (Subnet) gcloud compute networks subnets expand-ip-range SUBNET --region=us-central1 --prefix-length=16
- (GKE) gcloud container node-pools create node-pool-1 --cluster=example-cluster --preemptible

- (VM) gcloud compute instances create [INSTANCE_NAME] --deletion-protection
- (Config) gcloud config configurations create my-config
- (Config) gcloud config set compute/region us-west4
- (Config) gcloud config set compute/zone us-west4-b
- gcloud compute regions list
- gcloud compute zones list
- gcloud config unset compute/zone
- gcloud config unset compute/region
- (Config) gcloud config list -> view project id
- (IAM) gcloud projects get-iam-policy react-app-demo
- (IAM) gcloud iam roles copy --source="roles/spanner.databaseAdmin" --destination=CustomSpannerDbAdmin --dest-project=PROJECT_ID
- (IAM) gcloud iam roles describe roles/spanner.databaseUser
- (IAM) gcloud iam roles list
- (IAM) gcloud iam service-accounts list
- (Cloud Function) gcloud functions deploy Hello --http-trigger
- (Cloud Function) gcloud functions deploy Hello --trigger-topic=mytopic
- “sudo apt-get install”? sudo apt-get install command is used to download the latest version of your desired application from an online software repository pointed to by your sources.list configuration file and and install that application on your Linux machine.

GKE

- Cluster -> Nodes(VM) -> Pod(container), Pod(container), Pod(container) [1]
- a cluster consists of at least one control plane and multiple worker machines called nodes.[1]
- Cluster type: Autopilot and Standard
- To deploy your app to the GKE cluster you created, you need two Kubernetes objects.[3]

Kubernetes objects	Description
Deployment	to define your app.
Service	to define how to access your app

HelloApp using Node.js

1. Node.js Hello App (index.js)
2. Containerizing an app with Cloud Build
    - Dockerfile
    - get Project ID
    - Store your container in Artifact Registry
    - Build your container image using Cloud Build
3. Create a GKE cluster
4. Deploying to GKE using Deployment and Service
    - deployment.yaml
    - service.yaml
5. View the app - http://EXTERNAL_IP

```
gcloud config get-value project

gcloud artifacts repositories create hello-repo \
    --project=PROJECT_ID \
    --repository-format=docker \
    --location=REGION \

gcloud builds submit \
  --tag REGION-docker.pkg.dev/PROJECT_ID/hello-repo/helloworld-gke .

gcloud container clusters create-auto helloworld-gke \
  --region REGION

kubectl get nodes

kubectl apply -f deployment.yaml

kubectl get deployments

NAME               READY   UP-TO-DATE   AVAILABLE   AGE
hello-deployment   1/1     1            1           20s

kubectl get pods

kubectl apply -f service.yaml

kubectl get services
NAME         TYPE           CLUSTER-IP      EXTERNAL-IP     PORT(S)
AGE
hello        LoadBalancer   10.22.222.222   35.111.111.11   80:32341/TCP
1m
kubernetes   ClusterIP      10.22.222.1     <none>          443/TCP
20m

//http://35.111.111.11
```

DaemonSets

- A DaemonSet ensures that all (or some) Nodes run a copy of a Pod.
- "Every Node" is the keyword here, which is what DaemonSet is used for
- DaemonSets attempt to adhere to a one-Pod-per-node model
- B is right: https://kubernetes.io/docs/concepts/workloads/controllers/daemonset/ Some typical uses of a DaemonSet are:
- running a cluster storage daemon on every node
- running a logs collection daemon on every node
- running a node monitoring daemon on every node

Reasons for a Pod Status Pending:[7]

- Troubleshooting Reason #1: Not enough CPU

- Troubleshooting Reason #2: Not enough memory
- Troubleshooting Reason #3: Not enough CPU and memory

```
kubectl get pods
$ kubectl get pods
NAME                                          READY   STATUS
RESTARTS   AGE
echoserver-657f6fb8f5-wmgj5        0/1     Pending            0
1d
```

```
kubectl describe pod echoserver-657f6fb8f5-wmgj5
kubectl get nodes
kubectl describe node gke-gar-3-pool-1-9781becc-bdb3
```

GKE Sandbox

- GKE Sandbox on your cluster to isolate untrusted workloads in sandboxes on the node. GKE Sandbox is built using gVisor, an open source project.[8]
- click Security and select the Enable sandbox with gVisor checkbox.

Commands - Cluster

```
gcloud container clusters create-auto hello-cluster
kubectl get nodes
gcloud config set container/cluster hello-cluster
```

Command - Config

```
kubectl config use-context black
kubectl config view
```

- Why use node auto-provisioning Node auto-provisioning automatically manages and auto scales a set of node pools on the user's behalf. Without node auto-provisioning, GKE starts new nodes only from user-created node pools. With node auto-provisioning, new node pools are created and deleted automatically.https://cloud.google.com/kubernetes-engine/docs/how-to/node-auto-provisioning

GKE Links:

1. Standard cluster architecture
2. Quickstart-1
3. Quickstart-2
4. Kubernetes Service vs Deployment
5. Commands for Cluster

6. ClusterIP-NodePort-LB
7. Kubernetes Troubleshooting Walkthrough - Pending Pods
8. GKE Sandbox
9. Service vs Deployment