Google Cloud Digital Leader certification

Notes obtained from tutorialsdojo cheatsheets

        - Source repositories
        - Container registry
        - Build
    - Google official training

Any Google cloud resources that you allocate and provision must belong to a project. A project is composed of settings, permissions, and other metadata that describe your applications. Resources within a single project can work together straightforward by communicating through an internal network, subject to region-and-zone rules. A project can't access another project's resources unless configured to do so using Shared VPC or VPC Network Peering.

Each Google Cloud project has the following information:

- Project Name – which you will provide
- Project ID – which you can provide or Google Cloud can provide for us
- Project Number – which Google Cloud provides

The cloud infrastructure of GCP is built around:

- 20+ regions
- 70+ zones
- 140+ network edge locations

In GCP, a project is linked to a Cloud Billing Account that enables customers to connect a Payments Profile that includes a payment instrument to which costs are charged and pay for resource usage.

Compute

Google compute engine (GCE)

Linux/Windows based VMs, IaaS.

Machine types:

- General purpose
- Compute optimized
- Memory optimized
- Accelerator optimized

Instance template: save a VM instance's config for later VMs Instance groups: set of VM instances that you can manage as a single entity.

Pricing:

- Reservation
- Disk pricing (persistent disk)
- Preemptible (low-cost, short-term instances to run batch jobs and fault tolerant workloads)

Google App Engine

Fully managed serverless platform to develop and host webapps, supports apps written in many popular languages. PaaS mode.

- Standard env: based on preconfigured containers
- Flexible env: you can manage underlying infra

Google Kubernetes Engine (GKE)

Secured and managed Kubernetes services with auto-scaling and multi-cluster support.

Container orchestration managed by Google.

Cloud Run

Managed serverless compute platform to run stateless HTTP containers that are invokable via web requests or Pub/Sub events. You can deploy container images stored in Container Registry or Artifact Registry.

Similar to AWS Fargate/EKS.

Functions

A pay-as-you-go serverless function as a service (FaaS) to run your code with zero server management. Supports Node.js, Python, Go and Java.

Storage

Local SSD

Local, ephemeral solid-state block storage physically attached to the server that hosts VM instances.

Superior performance, very high IOPS and very low latency.

Local SSDs are designed for temporary storage use cases which makes them suitable for workloads like:

- Media Rendering
- Data Analytics
- Caches
- Processing Space

Persistent disks

Durable block storage devices to host VM instances, like GCE and GKE.

Data on each persistent disk is distributed across several physical disks and is designed for high durability. It stores data redundantly to ensure data integrity.

Filestore

NFS file servers for GCE and GKE. Most commonly used for media rendering, data analytics, and managing shared content.

Storage (GCS)

Object storage service using buckets.

- Lifecycle management
- Versioning
- Retention policies
- Encryption
- Access permissions

Storage classes:

- Standard
- Nearline
  - store objects for >30 days, data accessed max 1/mo
- Coldline
  - store infrequently accessed data within 90 days
- Archive
  - coldest storage, archive and disaster recovery data

`gsutil` tool allows managing GCS via command line.

Database

SQL

A fully managed relational database service. Cloud SQL is available for:

- MySQL
- PostgreSQL
- SQL Server

Scales instantly with a single API call as the data grows.

You can perform an analytics job by using BigQuery to directly query your CloudSQL instance.

Spanner

Relational db that scales horizontally with strong consistency

Synchronous replication across region. Automatic sharding of the data.

Equivalent to AWS Aurora.

BigQuery

Fully managed serverless data warehouse to feed petabyte-scale datasets and run SQL-like queries.

It provides integration with the Apache big data ecosystem allowing Hadoop/Spark and Beam workloads to read or write data directly from BigQuery using Storage API.

Equivalent to AWS Redshift.

BigTable

Fully managed NoSQL db for large analytical and operational workloads. Equivalent to DynamoDB

Networking

DNS

DNS serving, provides DNS lookup, translation from domain names into IP addresses, management of DNS records.

Equivalent to AWS Route 53.

Load Balancing

Put resources behind a single IP address.

- External Load Balancer
  - External HTTP(s)
  - External Network TCP/UDP
  - SSL Proxy Load Balancer (TCP with SSL offload traffic)
  - TCP Proxy
- Internal load balancer
  - Internal HTTP(s)
  - Internal TCP/UDP

CDN

Content Delivery Network accelerates web content by using Google's global edge network. Reduces latency, cost and load for backend services.

Google VPC

A virtual private cloud (VPC) allows you to specify an IP address range for the VPC, add and expand subnets, and configure firewall rules.

Resources inside the same VPC network can communicate with each other by using an internal IPv4 address but is still subject to applicable network firewall rules.

Each VPC network consists of one or more useful IP range partitions called subnets, each subnet associated with a region.

- IP addresses can be external and internal
- Firewall rules can be ingress or egress rules

Hybrid connectivity

> Cloud Interconnect

Provides low latency, highly available connections that enable you to reliably transfer data between on-prem and VPCs

> Direct Peering

Connects your on-premises network to Google services, including Google Cloud products that can be exposed via one or more public IP addresses.

> Carrier peering

Enables you to access Google applications, such as Google Workspace, by using a service provider to obtain enterprise-grade network services that connect your infrastructure to Google.

> Cloud VPN

Securely extends your peer network to Google’s network through an IPsec VPN tunnel.

Router

Used to define custom dynamic routes and scales with your network traffic. It utilizes Border Gateway Protocol (BGP) to exchange routes between your Virtual Private Cloud (VPC) network and your on-premises network

Similar to AWS Direct Connect

Security and Identity services

IAM

Lets you authorize who can take specific actions on resources to give you full control and visibility on your Google Cloud services centrally.

Identity

It's an API for provisioning and managing identity resources. It's a unified identity, access, app, and endpoint management (IAM/EMM) platform that helps IT and security teams maximize end-user efficiency, protect company data, and transition to a digital workspace.

Give users one-click access to apps with Single Sign-On (SSO).

Armor

Helps protect apps and websites against DDoS attacks. Detects and mitigates attacks against Load Balancing workloads.

Comes with predefined rules for protection against OWASP Top 10 risks. Easily monitor the metrics associated with your policies in the Cloud Monitoring dashboard.

KMS

Key Management Service is a cloud-hosted key management service that allows you to manage encryption keys.

Secret Manager

Secure way to store API keys, passwords, certificates and other sensitive data in Google Cloud.

Data and analytics

Looker

Business intelligence, embedded analytics and data application platform. Can be used for real-time dashboards.

Pub/Sub

Fully-managed, real-time messaging service for event driven systems that allows you to send and receive messages between independent apps.

Equivalent to AWS SNS.

Dataprep

Intelligent data service for exploring, cleaning and preparing (un-)structured data for analysis, reporting and ML.

Cloud Dataprep enables users to collaborate on similar flow objects in real-time or to create copies for other team members to use for independent tasks.

Explore your data through interactive visual distributions to assist in your discovery, cleansing, and transformation process.

Dataflow

Fully-managed serverless stream and batch data processing service, including Apache Beam SDK.

Dataproc

Fully-managed Spark, Hadoop, Presto and OSS clusters. USed for ETL jobs.

Management tools

Logging

Fully managed service for real-time log management. Helps you to securely store, search, analyze, and alert on all of your log data and events.

Write any custom log, from any source, into Cloud Logging using the public write APIs.

Monitoring

Collects metrics, events and metadata of apps. They are displayed with rich query language that helps identify issues and uncover significant patterns.

Deployment Manager

Infrastructure deployment service that automates the creation and management of Google Cloud resources. Equivalent to AWS CloudFormation.

Developer tools

Source repositories

Fully managed git repository where you can securely manage your code.

Container registry

Container image repository to manage Docker images, perform vulnerability analysis and define fine-grained access control.

Build

Build, test and deploy on GCP's serverless CICD platform. Fully serverless platform that helps you build your custom development workflows for building, testing, and deploying.

Google official training

Focus areas for culture transformation

- Talent
- Environment
- Structure
- Strategy
- Empowerment
- Innovation

Principles to scale innovation mindset

- Access
  - Fast and easy access
- Engagement
  - Valued up-to-date content
- Customization
  - Product adapts to individual needs and preferences
- Communication
  - Users want to communicate with services

To ensure transformational outcomes: focus on the user, think 10x, launch and iterate

10x mindset: fundamentally rethink business problems and solutions by a factor of 10.

- Data lake is a repository of raw data and tend to hold backup data
- Databases efficiently ingest large amounts of real-time data
- Data warehouses rapidly analyze multi-dimensional datasets

High-quality, bug-free data has: coverage, cleanliness, completeness (The availability of sufficient data about the world to replace human knowledge).

> Migration

When migrating to cloud, IT maintenance work is outsorced to the cloud provider.

- Move, then change

  - Conservative approach
- Change, then move
  - More aggressive, make apps cloud-ready while they're on premises, then move
- Invent in greenfield
  - Build an entirely new infra and apps in the cloud
  - Build apps at the same time as infra
  - Requires agility, access to a diverse development skillset and strong support from leadership
- Invent in brownfield
  - Invent new app in cloud that will replace an existing legacy app that remains on premises, and is retired only after the new app is built
  - Adds redundancy, minimizes risk, increased costs
- Move without changes

> Costs

Best practices for managing cloud costs:

- Visibility
  - Visualize current spending, trends, forecasting
- Accountability
  - You can allocate costs to individual departments and teams
  - Accountablity culture
- Control
  - Control who has the ability to spend and view costs
- Intelligence
  - Google cloud makes intelligent recommendations

Google's data security design consists of:

- Sharding data
- Encrypting each piece of data
- Encryption the data encryption key