Google Cloud Platform (GCP) – Overview & Key Concepts

✅ Overview

Google Cloud Platform (GCP) is **Google's cloud computing suite** of services that allows users to **build, deploy, and scale applications, websites, and services** on the same infrastructure that Google uses internally for products like Search, YouTube, and Gmail.

It provides **Infrastructure as a Service (IaaS)**, **Platform as a Service (PaaS)**, and **Software as a Service (SaaS)** offerings with a strong emphasis on **data, machine learning, and developer tools**.

🔑 Key Concepts

1. Global Infrastructure

- **Regions**: Specific geographic locations where GCP services are available (e.g., us-central1, asia-southeast1).
- **Zones**: Each region has 2+ zones to support high availability and fault tolerance.
- **Projects**: The main organizational unit for resource management in GCP.
- **Resources**: Services like VMs, databases, and storage created within projects.

---

2. Core Services

Category	Key Services
Compute	Compute Engine, App Engine, Cloud Functions, Cloud Run, Kubernetes Engine (GKE)
Storage	Cloud Storage, Persistent Disks, Filestore, Cloud Storage Nearline/Coldline
Networking	VPC, Cloud Load Balancing, Cloud DNS, Cloud CDN, Cloud Interconnect
Databases	Cloud SQL, Cloud Spanner, Firestore, Bigtable
AI/ML	Vertex AI, AutoML, AI APIs (Vision, Speech, NLP)
Analytics	BigQuery, Dataflow, Dataproc, Pub/Sub

---

3. Identity, Security & Governance

Area	Key Concepts
IAM (Identity and Access Management)	Assign roles and permissions to users and resources.
Service Accounts	Used by applications to access resources securely.
Organization Policies	Set constraints on resources across the org hierarchy.
Cloud Identity	Manage users and groups across GCP and Google Workspace.
Security Tools	VPC Service Controls, Cloud Armor, Data Loss Prevention (DLP), Security Command Center

---

4. Application Development & DevOps

Service	Description
App Engine	PaaS for deploying web apps without managing infrastructure.
Cloud Functions	Serverless functions triggered by events (e.g., HTTP, Pub/Sub).
Cloud Run	Serverless container hosting for HTTP-based apps.
Google Kubernetes Engine (GKE)	Managed Kubernetes service for container orchestration.
Cloud Build	Continuous integration and delivery (CI/CD) service.
Artifact Registry	Store and manage Docker images, Maven/Gradle artifacts, etc.
Operations Suite (formerly Stackdriver)	Monitoring, logging, and diagnostics across apps and infrastructure.

---

5. Data, AI & Analytics

📊 Data Services

Service	Purpose
Cloud SQL	Managed relational database (MySQL, PostgreSQL, SQL Server).
Cloud Spanner	Horizontally scalable, globally distributed relational DB.
Firestore / Firebase Realtime DB	NoSQL databases for mobile/web apps.
Bigtable	Low-latency, high-throughput NoSQL database (used in Google Search).

🤖 AI & ML Services

Service	Description
Vertex AI	Unified platform for building, training, and deploying ML models.
AutoML	Train custom ML models with minimal coding.
AI APIs	Pre-trained models for vision, speech, translation, and more.
TPUs (Tensor Processing Units)	Google's specialized hardware for ML workloads.

📈 Analytics Services

Service	Description
BigQuery	Serverless, highly scalable data warehouse for fast SQL analytics.
Dataflow	Stream/batch data processing using Apache Beam.
Pub/Sub	Asynchronous messaging service for event-driven systems.
Dataproc	Managed Hadoop/Spark cluster for big data workloads.

---

6. Billing & Cost Management

Tool	Description
Billing Account	Centralized payment method for GCP resources.

Tool	Description
Budgets & Alerts	Set usage or cost limits and receive email alerts.
Pricing Calculator	Estimate GCP costs before deployment.

🚀 Why Use GCP?

- Runs on **Google's global infrastructure** (same as Gmail and YouTube)
- Industry-leading **AI/ML and analytics tools** (BigQuery, Vertex AI)
- Fully managed **Kubernetes (GKE)** and **serverless offerings**
- Strong integration with **open-source tools** and **DevOps pipelines**
- Transparent and competitive **pricing**