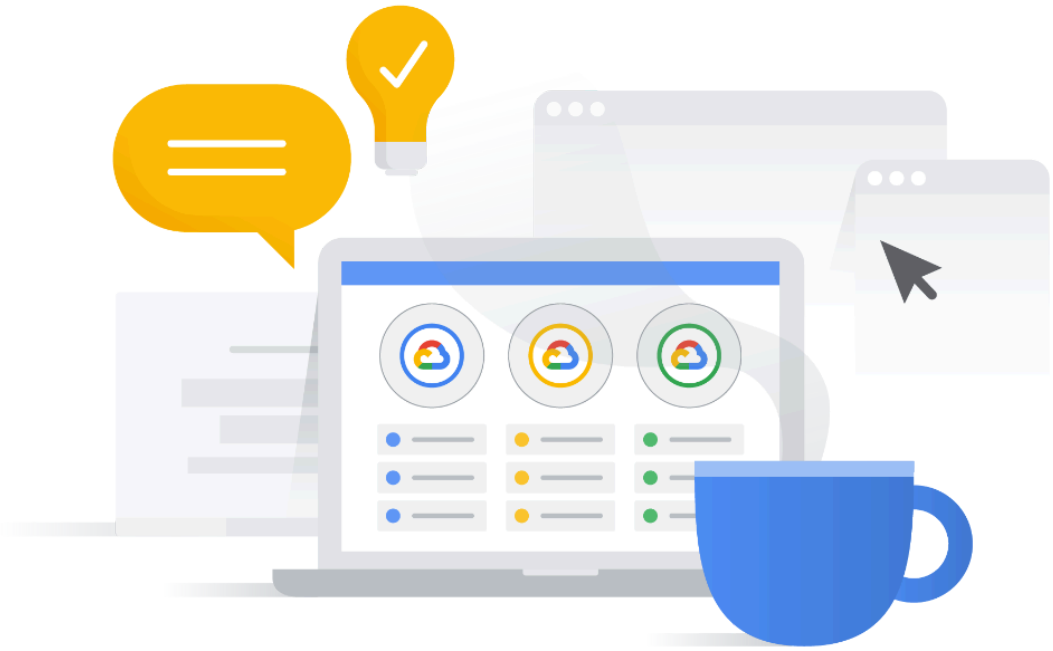
Google Cloud

# Cloud Digital Leader

## Study Guide



v2.0

# Contents

# Introduction

The Google Cloud Digital Leader training and exam are intended for tech-adjacent individuals who want to demonstrate an overall knowledge of cloud technology concepts and Google Cloud.

The exam validates a candidate's ability to complete the following course objectives:

- Identify Google Cloud products and solutions that support digital transformation.
- Explain how cloud technology and data can be used to innovate within organizations.
- Identify how organizations can innovate using Google Cloud's artificial intelligence and machine learning solutions.
- Explain infrastructure and application modernization with Google Cloud.
- Describe the fundamentals of cloud security Google's trusted infrastructure.
- Explain how to optimize cloud costs and achieve operational excellence with Google Cloud.

# Learn more about the exam

A Cloud Digital Leader can distinguish and evaluate the various capabilities of Google Cloud core products and services and how they can be used to achieve desired business goals. A Cloud Digital Leader is well-versed in fundamental cloud concepts and can demonstrate a broad application of cloud computing knowledge in various applications.

The Cloud Digital Leader exam is job-role independent. The exam assesses the knowledge and skills of any individuals who want (or are required to) understand the purpose and application of Google Cloud products.

The Cloud Digital Leader exam assesses knowledge in six areas:

- Digital Transformation with Google Cloud
- Exploring Data Transformation with Google Cloud
- Innovating with Google Cloud Artificial Intelligence
- Modernize Infrastructure and Applications with Google Cloud
- Trust and Security with Google Cloud
- Scaling with Google Cloud Operations

Sign up for the Cloud Digital Leader Learning Path through Google Cloud Skills Boost, Coursera, or Pluralsight.
Prepare for the exam with sample questions.

Learn more about how and where to take the exam on the Cloud Digital Leader website.

# Outline of learning path content

*For learners who have taken Google Cloud's instructor-led training the terminology differs slightly. Each course listed here refers to one module and each module refers to a lesson.*

## Course 1

### Digital Transformation with Google Cloud

**Module 1**: Why Cloud Technology is Transforming Business

**Module 2**: Fundamental Cloud Concepts

**Module 3**: Cloud Computing Models and Shared Responsibility

## Course 2

### Exploring Data Transformation with Google Cloud

**Module 1**: The Value of Data

**Module 2**: Google Cloud Data Management Solutions

**Module 3**: Making Data Useful and Accessible

## Course 3

### Innovating with Google Cloud Artificial Intelligence

**Module 1**: AI and ML Fundamentals

**Module 2**: Google Cloud's AI and ML Solutions

## Course 4

### Modernize Infrastructure and Applications with Google Cloud

**Module 1**: Important Cloud Migration Terms

**Module 2**: Modernizing Infrastructure in the Cloud

**Module 3**: Modernizing Applications in the Cloud

## Course 5

### Trust and Security with Google Cloud

**Module 1:** Trust and Security in the Cloud

**Module 2:** Google's Trusted Infrastructure

**Module 3:** Google Cloud's Trust Principles and Compliance

## Course 6

### Scaling with Google Cloud Operations

**Module 1:** Financial Governance and Managing Cloud Costs

**Module 2:** Operational Excellence and Reliability at Scale

**Module 3:** Sustainability with Google Cloud

# Course 1: Digital Transformation with Google Cloud

## Module 1: Why Cloud Technology is Revolutionizing Business

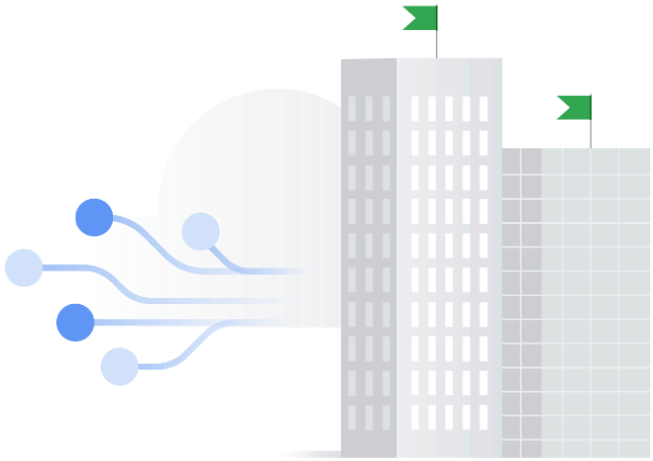
This module covers innovations throughout history that led to paradigm shifts, and looks at how cloud technology is enabling a new wave of digital transformation. It defines some important terms that you hear throughout the course, and it describes the benefits of adopting cloud technology to digitally transform a business. It also provides real-world examples and discusses the challenges that lead to a digital transformation.

## Module 2: Fundamental cloud concepts

This module explains some fundamental cloud concepts and explains how migrating to the cloud affects your organization's flexibility, agility, reliability, and total cost of ownership. It explores the different types of infrastructure and various use cases for them. It also describes the Google Cloud network, how it's structured, and how performance is measured and impacted.

## Module 3: Cloud computing models and shared responsibility

This module explains the different cloud computing models: IaaS (infrastructure as a service), PaaS (platform as a service), and SaaS (software as a service). It identifies the benefits and tradeoffs of implementing each computing model and how to choose the most appropriate one to meet business needs. It also explores the shared responsibility between an organization and its cloud provider in hardware, software, and security.



**Key terms:**

Cloud, cloud technology, compute power, computing, data, innovation, transformation cloud, total cost of ownership (TCO), CapEx vs OpEx, network, regions, zones, IaaS, PaaS, SaaS, shared responsibility

**Additional reading**

What is cloud computing?

What is digital transformation?

Cloud computing 101: Frequently asked questions

What is Infrastructure as a Service (IaaS)?

Google Cloud products list page

Google's Guide to Innovation

Google Cloud Adoption Framework Impact Study

Managing Change in the Cloud

Cloud Customer Case Studies

Google Cloud whitepapers

Cloud locations

Google Cloud infrastructure (interactive web tool)

# Course 2: Exploring Data Transformation with Google Cloud

## Module 1: The Value of Data

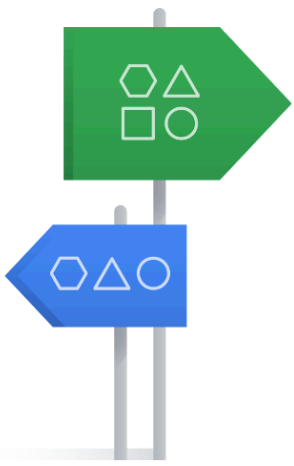
This module explores the important role that data plays in an organization's digital transformation. It examines how data generates business insights, drives decision making, and creates new value. It explains how the cloud unlocks business value from all types of data, including structured data and previously untapped unstructured data, and also examines how data governance is essential to a successful data journey.

## Module 2: Google Cloud data management solutions

This module explores Google Cloud data management products and solutions, and demonstrates how they apply to different business use cases. It reviews the differences between Google Cloud data management options, including data types and common business use cases, and explains storage classes in Cloud Storage. The module also describes how organizations can migrate or modernize a database in the cloud.

## Module 3: Making data useful and accessible

This module examines how smart analytics, business intelligence tools, and streaming analytics can add value in different business use cases. It outlines the benefits of using BigQuery and explores how Looker democratizes access to data by empowering individuals to self-serve business intelligence and create insights. This module also explains how streaming analytics in real time makes data more useful and generates business value, and it reviews key Google Cloud products that modernize data pipelines.



**Key terms:**

Data, data management, data value chain, data governance, structured data, unstructured data, semi-structured data, databases, data warehouses, data lakes, database migration, business intelligence, streaming analytics

**Additional reading**

What is a data lake?

What is a data warehouse?

What is Cloud Storage?

What is object storage?

What is big data?

What is streaming analytics?

What is business intelligence?

What is ETL?

Build a modern, unified data analytics platform with Google Cloud

Principles and best practices for data governance in the cloud

# Course 3: Innovating with Google Cloud Artificial Intelligence

## Module 1: AI and ML Fundamentals

Artificial intelligence and machine learning can provide many benefits to a business, but it's important to understand the fundamentals before starting any AI or ML initiative. This module explores many of those fundamental concepts, including the difference between artificial intelligence and machine learning, and how machine learning differs from data analytics and business intelligence. It also outlines the different types of problems that AI solutions are suited to solve and examines the importance of data quality, responsible AI practices, and explainable AI.

## Module 2: Google Cloud's AI and ML Solutions

This module explores four options to build ML models with Google Cloud: BigQuery ML, pre-trained APIs, AutoML, and custom training. It explains how BigQuery ML can be used to create and execute ML models in BigQuery by using standard SQL queries, and outlines the Natural Language API, Vision API, Translation API, Speech-to-Text API, and Text-to-Speech API. The module also investigates ways organizations can use their own data to train custom ML models with AutoML, and explains how creating custom models with Google Cloud's Vertex AI can help businesses stand out from the competition.

**Key terms:**

Artificial intelligence (AI), machine learning (ML), data quality, responsible AI, explainable AI, ML models

**Additional reading**

What is artificial intelligence?

What is machine learning?

AI vs ML

Google Cloud's AI Adoption Framework

Google Cloud Responsible AI

# Course 4: Modernize Infrastructure and Applications with Google Cloud

## Module 1: Important Cloud Migration Terms

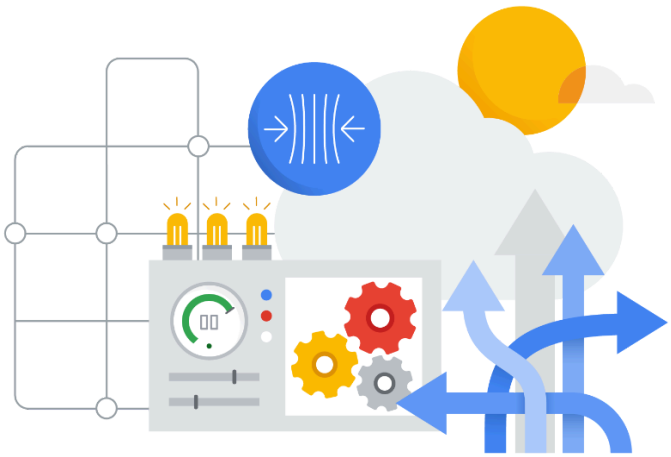
This module lays the groundwork for understanding cloud migration by introducing essential terminology. These terms serve as a foundation for further exploration into the modernization of infrastructure and applications in the cloud.

## Module 2: Modernizing Infrastructure in the Cloud

This module explores the options for and advantages of running compute workloads in the cloud, and explaining containers and the business value of serverless computing. It investigates why modernization and migration to the cloud are important steps in an organization's transformation journey, and how each application might have a different path. This module also examines the options for and advantages of running compute workloads in the cloud, and also discusses the advantages of using containers and serverless computing in application modernization.

## Module 3: Modernizing Applications in the Cloud

This module focuses on application modernization and the business value of application programming interfaces (APIs). It also explores the business reasons for choosing hybrid or multi-cloud strategies and how GKE Enterprise enables these strategies.



**Key terms:**

Infrastructure modernization, virtual machines, containers, serverless computing, applications, rehosting, APIs, hybrid cloud, multi-cloud

**Additional reading**

Google Cloud overview

Google Cloud geography and regions

Best practices for enterprise organizations

Google Cloud setup checklist

What are containers?

What is a hybrid cloud?

What is a virtual machine?

Migrate workloads to the public cloud: an essential guide & checklist

Where should I run my stuff? Choosing a Google Cloud compute option

CIO's Guide to Application Migration

Google Cloud Networking overview blog

# Course 5: Trust and Security with Google Cloud

## Module 1: Trust and Security in the Cloud

This module explores the importance of control, compliance, confidentiality, integrity, and availability in a cloud security model. It explains the difference between cloud security and traditional on-premises security, and also examines today's top cybersecurity threats and how they impact businesses.

## Module 2: Google's Trusted Infrastructure

This module examines the benefits of data centers designed and built by Google and explores the role of encryption in securing an organization's data. It explains the difference between authentication, authorization, and auditing, and outlines how Google products can help protect against network attacks, including distributed denial-of-service (DDoS) by using Google Cloud Armor.

## Module 3: Google Cloud's Trust Principles and Compliance

This module explores how Google Cloud's trust principles, transparency reports, and independent third-party audits support customer trust. It examines the importance of data sovereignty and data residency and the options Google Cloud offers organizations to control where their data is stored. This module also discusses how the Google Cloud compliance resource center and Compliance Reports Manager supports industry and regional compliance needs.

**Key terms:**

Encryption, zero trust model, least privileged model, phishing attack, physical damage, malware, virus, ransomware, unsecured third party systems, data at rest, data in transit, two-step verification, SecOps, compliance

**Additional reading**

What is cloud security?

What is encryption?

Google Cloud security foundations guide

Google security

Google Infrastructure Security Design Overview

Voyage to the modern workplace

Encryption at rest in Google Cloud

Encryption in transit in Google Cloud

# Course 6: Scaling with Google Cloud Operations

## Module 1: Financial Governance and Managing Cloud Costs

This module explores how cloud financial governance best practices provides predictability and control over cloud resources and defines important cloud cost-management terms and concepts. It also examines the benefits of using the resource hierarchy to control access and considers how resource quota policies, budget threshold rules, and Cloud Billing Reports can help control cloud consumption.

## Module 2: Operational excellence and reliability at scale

This module discusses the benefits of modernizing operations by using Google Cloud and defines the main cloud operations, cloud reliability, DevOps, and SRT terms. It explains the importance of designing resilient, fault-tolerant, and scalable infrastructure and processes and also explores how Google Cloud Customer Care supports organizations as they move to the cloud.

## Module 3: Sustainability with Google Cloud

This module explores Google Cloud's commitment to sustainability and how it supports organizations' sustainability goals.

**Key terms:**

Resource hierarchy, DevOps, SRE, latency, traffic, saturation, errors, SLO, SLA, SLI, monitoring, logging

**Additional reading**

A guide to financial governance in the cloud

Resource hierarchy

Increasing business value with better IT operations: A guide to site reliability engineering (SRE)

Exploring container security: the shared responsibility model in GKE

# Glossary

## Course 1

**Bandwidth**: A measure of how much data a network can transfer in a given time.

**Capital expenditures (CapEx)**: Upfront business expenses put toward fixed assets. Organizations buy these items once, and they benefit their business for years.

**Cloud technology/computing** – The technology and processes needed to store, manage, and access data that is transferred over the Cloud (as opposed to data that remains on your computer's hard drive).

**Compute power**: The speed at which a computer can process data.

**Computing**: A machine's ability to process, store, retrieve, compare and analyze information, and to automate tasks often done by computer programs (otherwise known as software or applications).

**Data** – Any information that is useful to an organization. Can be numbers on a spreadsheet, text in an email, audio or video recordings, images, or even ideas in employees' heads. Includes internal and external information.

**Digital transformation** – When an organization uses new technologies to redesign and redefine relationships with their customers, employees, and partners. Digital transformation uses modern digital technologies—including all types of public, private, and hybrid cloud platforms—to create or modify business processes, culture, and customer experiences to meet changing business and market dynamics.

**Infrastructure as a service (IaaS)**: A computing model that offers the on-demand availability of almost infinitely scalable infrastructure resources, such as compute, networking, storage, and databases as services over the internet.

**Network latency**: The time it takes for data to travel from one point to another. Often measured in milliseconds, latency, sometimes called lag, describes delays in communication over a network.

**On-premises IT infrastructure** – This refers to hardware and software applications that are hosted on-site, located, and operated within an organization's data center to serve their unique needs.

**Open source**: Software with source code that is publicly accessible and free for anyone to use, modify, and share.

**Open standard**: Software that follows particular specifications that are openly accessible and usable by anyone.

**Operating expenses (OpEx)**: Recurring costs for a more immediate benefit. This represents the day-to-day expenses to run a business.

**Platform as a service (PaaS)**: A computing model that offers a cloud-based platform for developing, running, and managing applications.

**Private cloud**: When an organization has virtualized servers in its own data centers, or those of a private cloud provider, to create its own private dedicated environment.

**Public cloud**: Where on-demand computing services and infrastructure are managed by a third-party provider, such as Google Cloud, and shared with multiple organizations or "tenants" through the public internet.

**Regions**: Independent geographic areas where Google Cloud resources are deployed, composed of zones.

**Shared responsibility model**: A model in which the responsibility to secure data is shared between a business and the cloud provider. The cloud service provider is the data processor, whereas the organization is the data controller.

**Software as a service (SaaS)**: A computing model that offers an entire application, managed by a cloud provider, through a web browser

**The cloud**: A metaphor for the network of data centers that store and compute information available through the internet. It includes the complex web of software, computers, networks, and security systems involved.

**Total cost of ownership (TCO)**: A comprehensive assessment of all layers within the infrastructure and other associated costs across the business over time. Includes acquiring hardware and software, management and support, communications, and user expenses, and the cost of service downtime, training, and other productivity losses.

**Zone**: A geographic area where Google Cloud resources are deployed.

## Course 2

**Business intelligence**: The process of collecting, analyzing, and interpreting data to make better business decisions

**Database**: An organized collection of data generally stored in tables and accessed electronically from a computer system. Built and optimized to enable the efficient ingestion of large amounts of data from many different sources.

**Data lake**: Data lake is a repository designed to store, process, and secure large amounts of structured, semistructured, and unstructured data. It can store data in its native format and process any variety of it, ignoring size limits and serves many purposes, such as exploratory data analysis.

**Data point**: A piece of information (for example, a customer purchase or return).

**Data**: Raw or unprocessed information that can be used to derive insights.

**Dataset**: Aggregated data points of one category (for example, customer transactions).

**Data value chain**: The sequence of activities involved in transforming data into value for an organization

**Data warehouse**: The central hub for all business data, it assembles data from multiple sources, including databases. When combined with connector tools, it can transform unstructured data into semi-structured data that can be used for analysis. Data warehouses are built to rapidly analyze and report massive and multi-dimensional datasets on an ongoing basis, in real-time.

**Metadata** – Information about objects (for example, about images or audio).

**Object storage** – A data storage architecture for large stores of unstructured data, designating each piece of data as an object (for example, audio or multimedia files).

**Semi-structured data**: Data that falls somewhere between structured and unstructured data. It's organized into a hierarchy, but without full differentiation or any particular ordering. Examples include emails, HTML, JSON, and XML files.

**Streaming analytics**: The process of analyzing data in real time as it is being generated.

**Structured data** – Highly organized, quantitative data (for example, names or credit card numbers). Easily stored and managed in databases.

**Unstructured data** – Data that has no organization and tends to be qualitative (e.g. word processing documents or images). Can be stored as objects, which consist of the data in its native format along with metadata such as unique identifiers.

## Course 3

**Artificial intelligence (AI)**: A broad field or term that describes any kind of machine capable of a task that normally requires human intelligence, such as visual perception, speech recognition, decision-making, or translation between languages.

**Data quality**: The degree to which data is complete, unique, timely, valid, accurate, and consistent.

**Explainable AI**: Techniques that make AI models more transparent and understandable to humans.

**Machine learning (ML)**: A branch within the field of AI. Computers that can "learn" from data and make predictions or decisions without being explicitly programmed to do so by using algorithms or models to analyze data. These algorithms use historical data as input to predict new output values.

**ML models**: Mathematical models that are used to make predictions or decisions based on data.

**Responsible AI**: An approach to AI development and deployment that considers the ethical, social, and environmental implications of AI.

## Course 4

**Application (or app)**: A computer program or software that is designed to perform a specific digital task, typically used or run by an end-user. In this digital age, customers expect applications to be intuitive, well-functioning, and efficient.

**Application programming interface (API)** – A piece of software that interfaces with or connects different applications and enables information to flow between systems. Unlike a user interface, which connects a computer to a person, an API connects computers or pieces of software to each other. One purpose of APIs is to hide the internal details of how a system works, exposing only those parts a developer wants to allow a user or program to interface with. In this way APIs can help organizations to adapt to modern business needs by allowing access to older legacy systems.

**Container**: Follows the same principle as a VM, providing an isolated environment to run software services and optimize resources from one piece of hardware. Containers are more efficient than VMs because they do not recreate a full representation of the hardware, but only recreate or virtualize the operating system.

**Hybrid cloud**: An IT infrastructure that combines on-premises infrastructure with cloud infrastructure, allowing organizations to leverage the benefits of both environments.

**Kubernetes**: An open source cluster management system that provides automated container orchestration.

**Multi-cloud**: An IT infrastructure that uses multiple public cloud providers, such as Google Cloud, to achieve greater flexibility, scalability, and cost savings.

**Rehosting**: Moving an application or system from one environment to another, such as from on-premises to the cloud, without making any changes to the application or system itself.

**Serverless computing**: A cloud computing execution model in which the cloud provider allocates machine resources on demand and takes care of the servers on behalf of their customers. Businesses provide code for the function that they want to run and the cloud provider handles all infrastructure management. Resources such as compute power are automatically provisioned behind the scenes as needed.

**Virtual machines (VM)** – A VM is a virtualized instance of a server that re-creates the functionality of a dedicated physical server. It uses a partitioned space inside a physical server which makes it easy to optimize and reallocate resources and allow multiple systems to run on the same hardware.

## Course 5

**Availability**: The duration for which the cloud service provider guarantees that client's data and services are up and running or accessible.

**Compliance**: The act of adhering to laws, regulations, or standards.

**Defense-in-depth**: The cloud service provider manages the security of its infrastructure and its data centers, and customers gain the benefits of their infrastructure's multiple built-in security layers.

**Encryption**: The process of encoding data stored in the cloud to safeguard it from unauthorized access.

**Least privilege model**: A security principle that grants users only the minimum permissions necessary to perform their tasks.

**Malware**: Software designed to harm a computer system, such as viruses, ransomware, and spyware.

**Phishing**: An attempt to obtain personal information or credentials by sending fraudulent emails or messages.

**Privacy**: The data an organization or an individual has access to, and who they can share that data with.

**SecOps**: A collaborative approach to security that combines IT security and operations teams.

**Security**: The policies, procedures, and controls put in place to keep data and infrastructure safe.

**Two-step verification**: A security measure that requires users to enter a second verification code, such as a code sent to their phone, in addition to their password to log in.

**Zero trust model**: A security approach that assumes no entity or user is trustworthy and requires continuous verification before granting access.

## Course 6

**DevOps**: Developer operations. A philosophy that seeks to create a more collaborative and accountable culture within developer and operations teams. Five objectives of DevOps include reducing silos, accepting failure as normal, implementing gradual change, leveraging tooling and automation, and measuring everything.

**Latency**: The time it takes for a system to respond to a request.

**Log file**: A text file where applications (including the operating system) write events. Log files make it easier for developers, DevOps, and system administrators to get insights and identify the root cause of issues within applications and the infrastructure.

**Logging**: A process that allows IT teams to analyze selected logs and accelerate application troubleshooting.

**Monitoring**: Gathering predefined sets of metrics or logs. Monitoring is the foundation for site reliability engineering because it provides visibility into the performance, uptime, and overall health of cloud powered applications.

**Resource hierarchy**: How an IT team can organize a business's Google Cloud environment and how that service structure maps to the organization's actual structure. It determines what resources users can access.

**Saturation**: The point at which a system is no longer able to handle any more requests.

**SLA (Service Level Agreement)**: A contract between a service provider and a customer that specifies the level of service that will be provided.

**SLI (Service Level Indicator)**: A quantitative measure of a particular aspect of service performance, such as latency or error rate.

**SLO (Service Level Objective)**: A target for a particular SLI, such as a maximum latency of 200 milliseconds.

**SRE**: Site reliability engineering. A discipline that applies aspects of software engineering to operations. The goals of SRE are to create scalable and highly reliable software systems. Best practices central to SRE align with DevOps objectives.

**Traffic**: The amount of data or number of requests that a system is handling.

# List of Google products and solutions

Learn more about Google Cloud products at cloud.google.com/products.

Learn more about Google Cloud solutions at cloud.google.com/solutions.

**Apigee API Management**: A platform for developing and managing APIs.

**App Engine**: A platform for building scalable web applications and mobile backends.

**Bare Metal**: Infrastructure to run specialized workloads on Google Cloud.

**BigQuery**: Google Cloud's leading data warehouse solution.

**Bigtable**: Google's NoSQL big data database service.

**Cloud Functions**: An event-driven compute platform for cloud services and apps.

**Cloud Identity**: A unified platform for IT administrators to manage user devices and apps.

**Cloud Logging**: An audit, platform, and application logs management tool.

**Cloud Monitoring**: A tool monitoring infrastructure and application health with rich metrics.

**Cloud Profiler**: Continuous CPU and heap profiling to improve performance and reduce costs.

**Cloud Run**: A fully managed environment for running containerized apps.

**Spanner**: A fully managed Google Cloud database service designed for global scale.

**Cloud SQL**: Google Cloud's database service (relational database management service).

**Cloud Storage**: Google Cloud's object storage service for structured, semi-structured, and structured data. One of several products used in data lake solutions.

**Cloud Trace**: A tracing system collecting latency data from applications.

**Compute Engine**: Virtual machines running in Google's data center.

**Cost Management**: Tools for monitoring, controlling, and optimizing business costs.

**Dataflow**: A fully managed streaming analytics service that creates a pipeline to process both streaming data and batch data.

**Firebase**: An app development software to build, improve, and grow mobile and web apps.

**Google Cloud console**: A web-based interface for managing and monitoring cloud apps.

**Google Kubernetes Engine**: An open source container orchestration system for automating computer application deployment, scaling, and management.

**Looker**: Google Cloud's business intelligence solution.

**Pub/Sub**: A distributed messaging service that can receive messages from various device streams such as gaming events, IoT devices, and application streams. The name is short for Publisher/Subscriber.

**TensorFlow**: An end-to-end open source platform for machine learning, with a comprehensive, flexible ecosystem of tools, libraries and community resources, originally created by Google.

**Vertex AI**: A unified platform for training, hosting and managing ML models. Features include AutoML and custom training.

**VMware Engine**: An engine for migrating and running VMware workloads natively on Google Cloud.