



Google Cloud Digital Leader Certification For GCP Beginner





A Cloud Digital Leader can articulate the capabilities of **Google Cloud core products** and **services** and how they **benefit organizations**. The Cloud Digital Leader can also describe **common business use cases** and how **cloud solutions support an enterprise**. The Cloud Digital Leader exam is **job-role agnostic** and **does not require hands-on experience with Google Cloud**

SOURCE: <https://cloud.google.com/certification/guides/cloud-digital-leader>

About Google Cloud Certifications

Foundational certification

Cloud Digital Leader

No hands-on experience
with Google Cloud is
required

Associate certification

Cloud Engineer

Recommended for 6+
months building on
Google Cloud

Professional certification

- Cloud Architect
- Cloud Database Engineer
- Cloud Developer
- Data Engineer
- Cloud DevOps Engineer
- Cloud Security Engineer
- Cloud Network Engineer
- Google Workspace
Administrator
- Machine Learning Engineer

Recommended for 3+ years
industry experience, including 1+
years using Google Cloud

Google Cloud Digital Leader curriculum

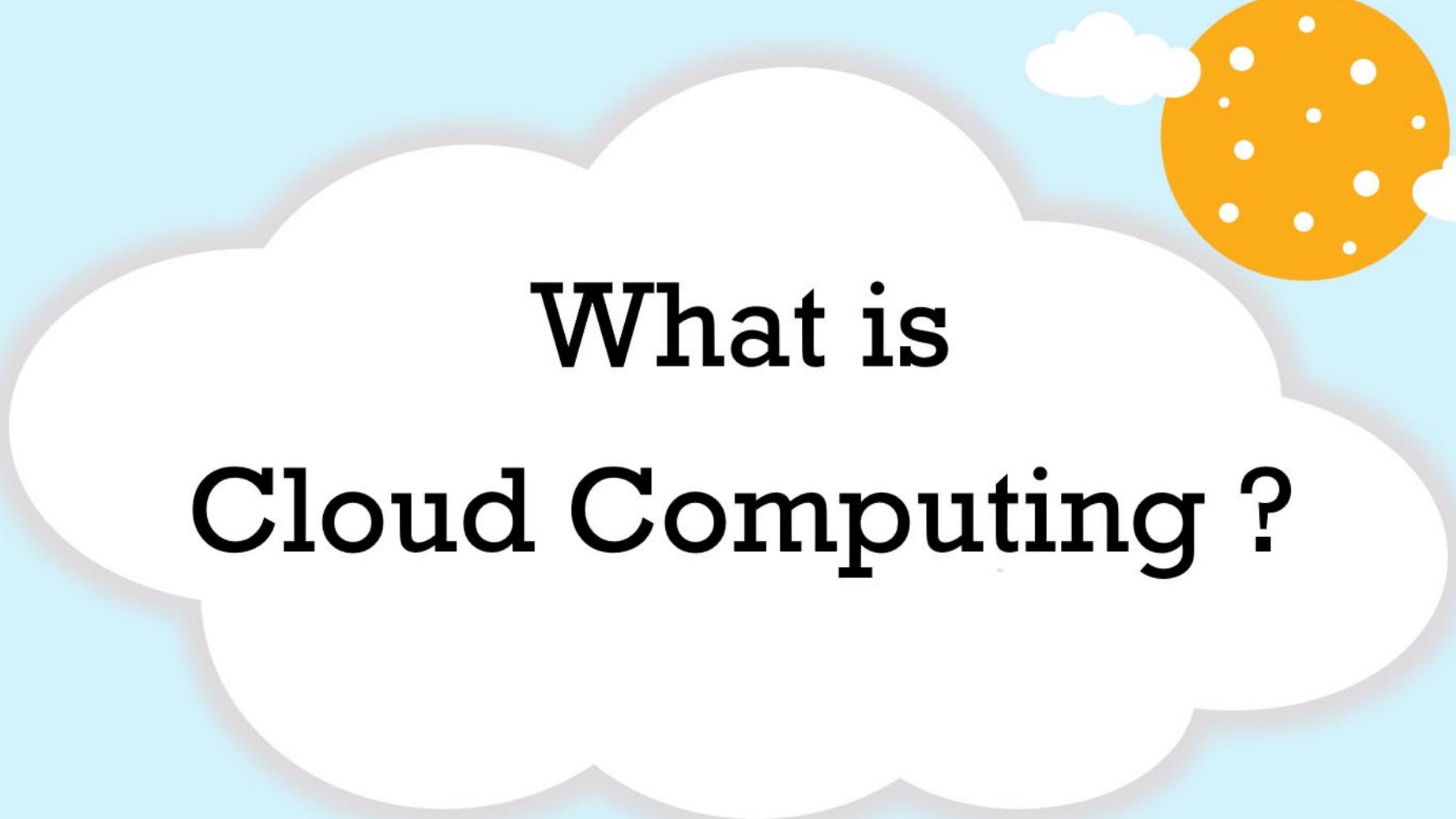
Introduction to digital transformation with Google Cloud
(approximately **10%** of the exam)

Infrastructure and application modernization
(approximately **30%** of the exam)

Innovating with data and Google Cloud
(approximately **30%** of the exam)

Understanding Google Cloud security and operations
(approximately **30%** of the exam)

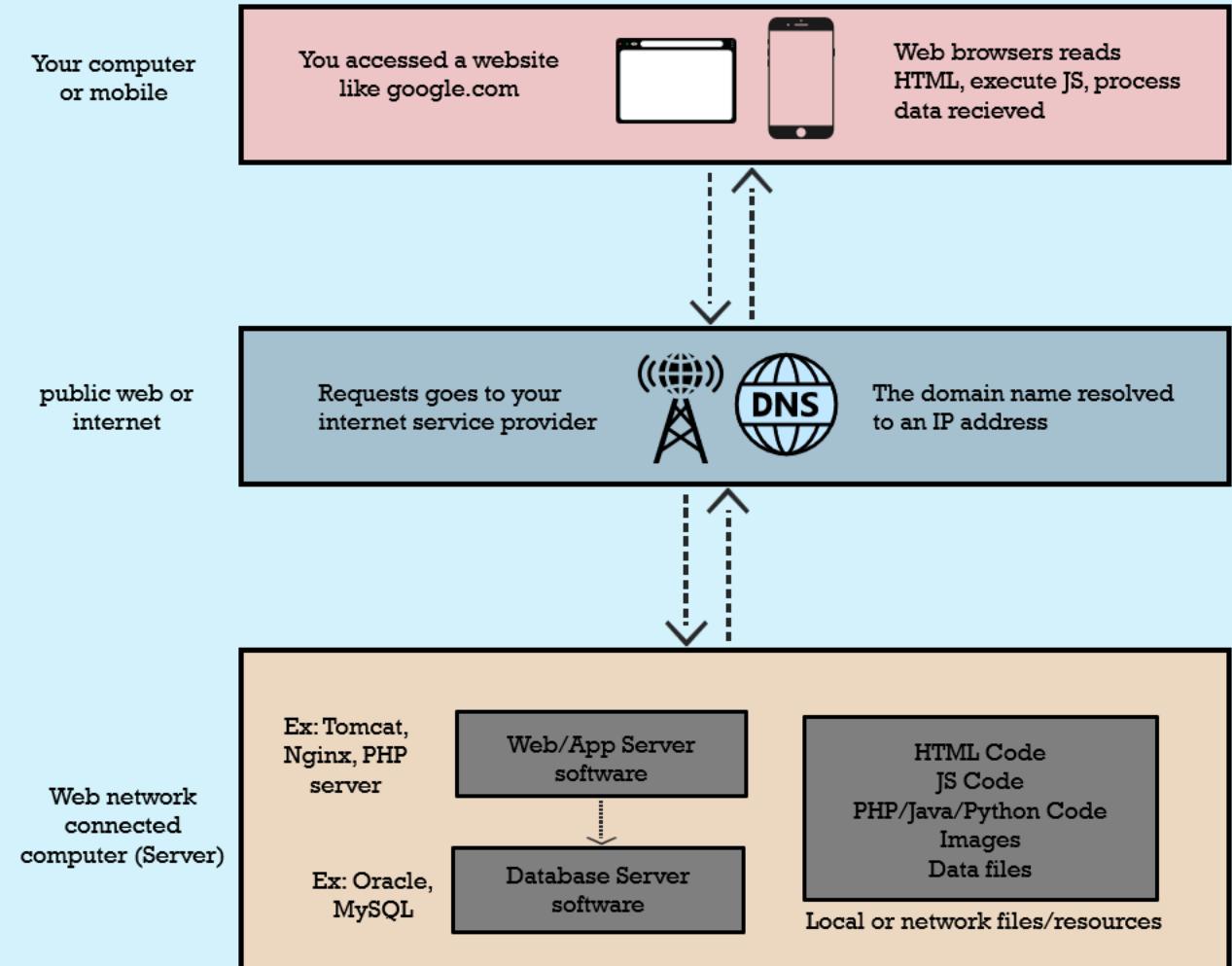
SOURCE: <https://cloud.google.com/certification/guides/cloud-digital-leader>



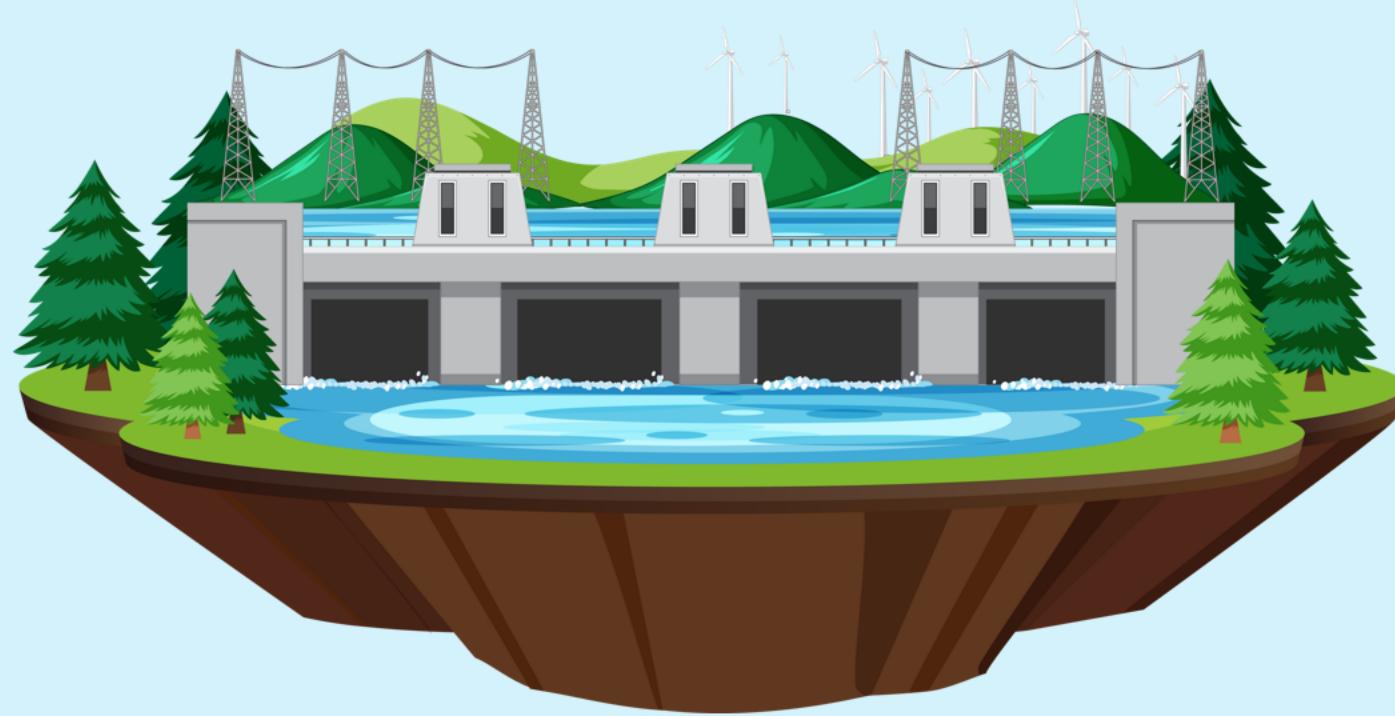
**What is
Cloud Computing ?**

How web or mobile apps works ?

We need machines which are connected to web network to run our web or mobile applications. These machines we call them as **SERVERS**



Just like Web Apps need servers/resources, we all need **POWER** to run our houses, offices, factories etc. But do we generate power by ourself using Thermal/Hydro/Wind approach. Off course not, because it is not economically feasible. What do you do instead? You connect to a power grid, and consume power based on your needs. Finally you will pay to power company based on your usage on monthly basis.



The power company specializes in generating power at scale, and you plug in and use a little slice of that power however you need without worrying about how the power is generated. **CLOUD COMPUTING** works very similarly. Cloud providers like GCP, AWS, Azure works just like a power company, but here the power are the IT resources & services.

What is Cloud Computing ?

Cloud Computing provides on-demand computing resources that are self-service, automated, and elastic to cater the demand.

The Cloud is a metaphor for the network of data centers that store and compute information available through the Internet.

Google, Amazon, and Microsoft are most popular public cloud computing platform providers. These providers own hardware on their facilities, from which they create computing resources that are made available to all organizations, which use their portion of the resources and get billed only for what they use.

Other funny definitions of Cloud



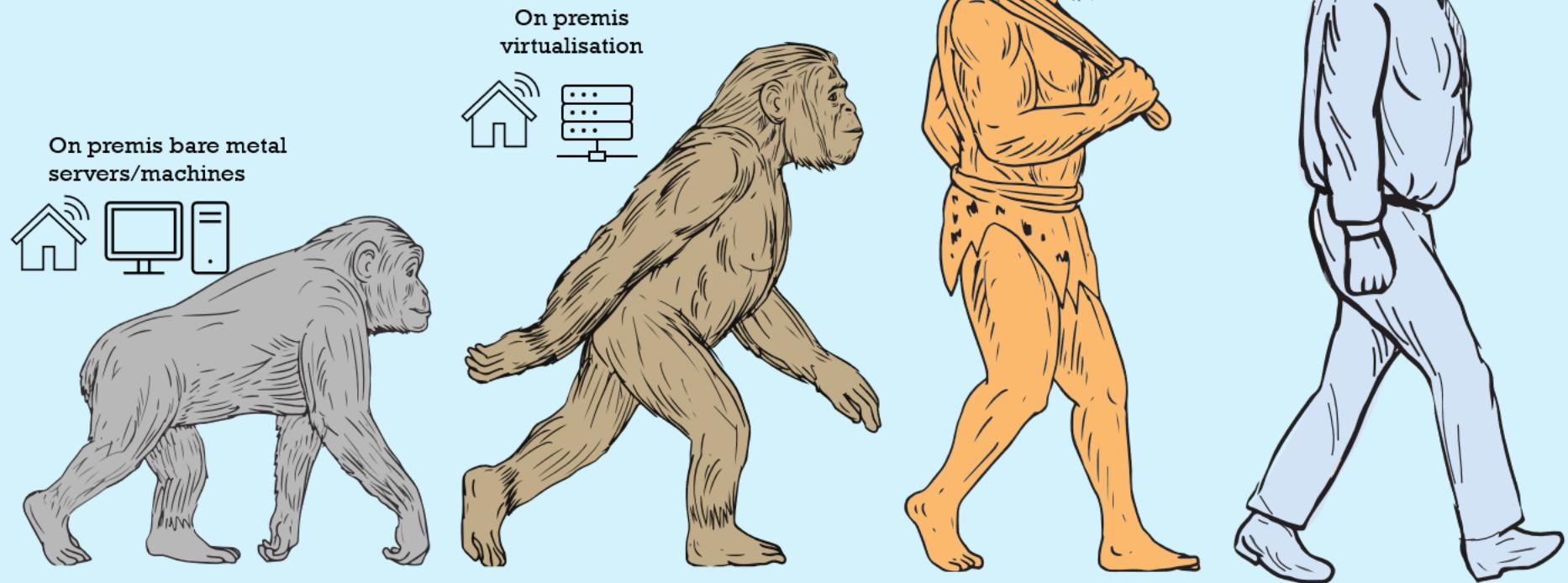
**There's no such thing as the cloud;
it's just someone else's computers.**



**what are the clouds made of?
“Linux servers mostly”**

Evolution of Cloud Computing

The goal of cloud computing is to provide computing resources anywhere, anytime, any size and give a business the power of choice as to the most suitable technology platform for any given workload, business initiative, or scenario that needs to be supported.



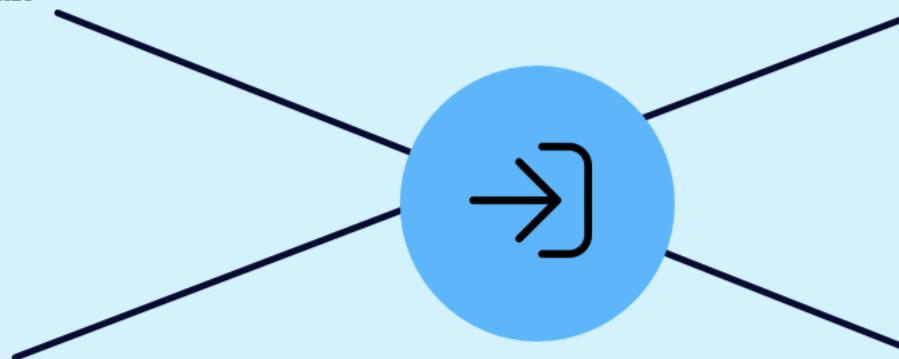
Different ways to interact with Google Cloud

Google Cloud Console

Google Cloud Console is the name Google uses for its web-based UI. You can get to the GCP Console via the link:
<https://console.cloud.google.com/>

Google Cloud SDK & Cloud Shell

If you just really love the command line, then you'll like Cloud Shell and the Cloud SDK. The Cloud SDK is a downloadable set of command-line utilities whereas Cloud Shell is Google's web-based CLI



Google Cloud APIs

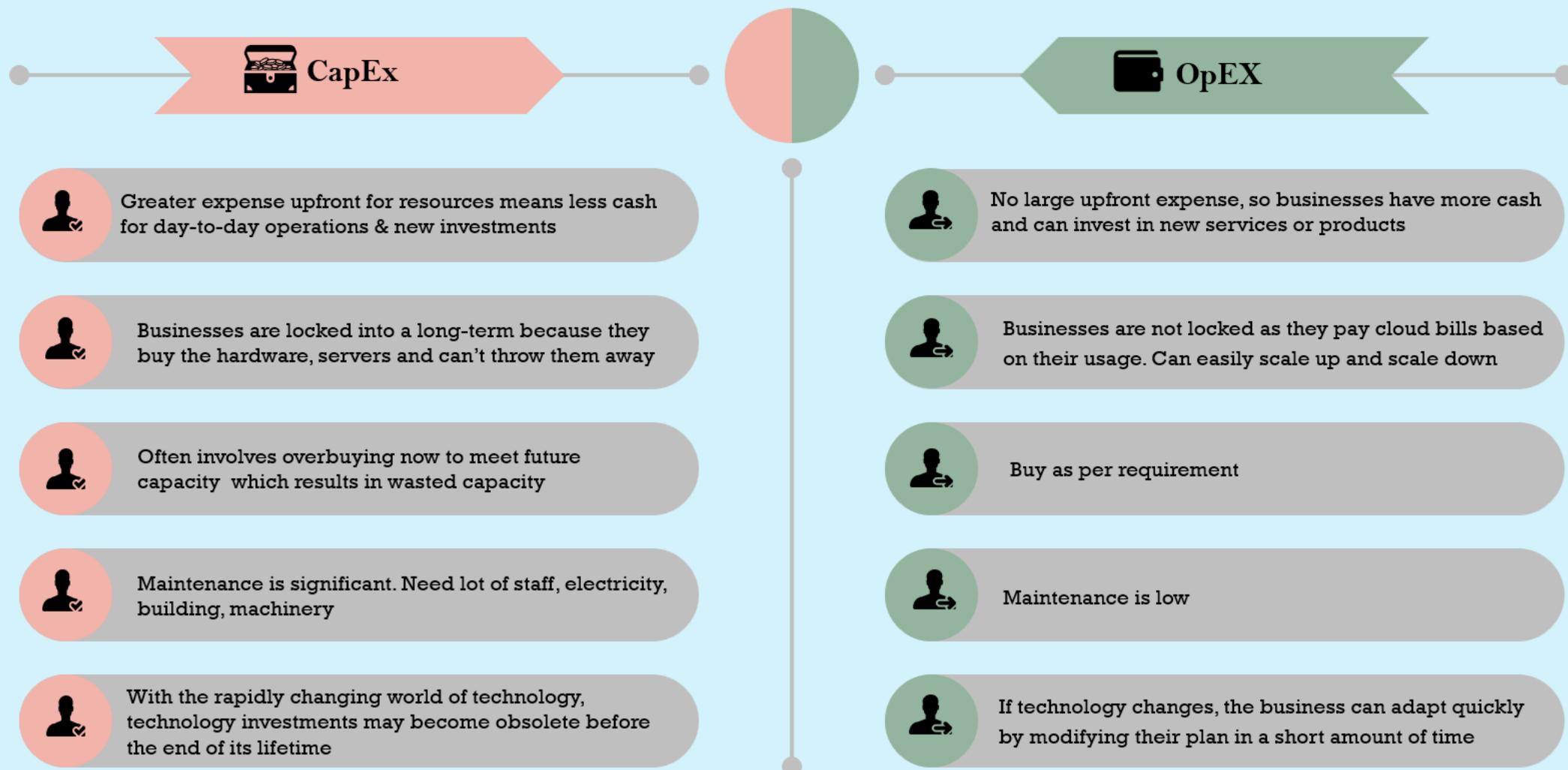
Each GCP service is also a web API service. If you know how to format the correct JSON input, and you send it the right URL, authenticated the right way, then you could use GCP with APIs as well.

Google Cloud mobile client

Google Cloud has a client you can run on your mobile device. It is a quick and easy way you can investigate what's happening in Google Cloud. But it is not the preferred way of interacting with GCP in a general sense. If you want to overview what's happening in the cloud, or perhaps do some quick monitoring, then it's a good tool.



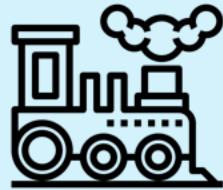
Cloud computing is not only a technical evolution but also a financial evolution; the expenditure model shifts from capital expenditure (CapEx) of hardware (buying upfront before you can use resources) to operating expense (OpEx) and paying as you use resources.



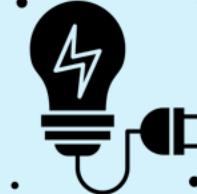
Is Cloud another paradigm shift ?



We as mankind, saw many inventions over the centuries which are acted as paradigm shift. Below are such examples. **Cloud Computing is one such paradigm shift.**



Steam Engines



Electricity



Vehicle



Computer



By the way, what is a paradigm shift ?

It is an important change that happens when the usual way of thinking about or doing something is replaced by a new and different way



These paradigm shift inventions throughout history triggered thousands of innovations ,what are called “**Kondratiev waves**” or “**innovation waves**.”

Like Electricity paved path Radio, TVs, similarly Computer paved paths to Laptops, Tab, Mobiles etc.

Is Cloud another paradigm shift ?



All the paradigm shifts have few things in common.

- ✓ They transform both the supply & the demand at the same time. On the supply side, the new technology can dramatically improve productivity. At the same time, the new technology can increase consumption, which generates more demand
- ✓ Increases the customer expectations
- ✓ Enable new ways of working/living.

For example with electricity, the supply and demand always high because people uses electricity in various forms like for TVs, AC, Heaters, Washing Machine etc. It has risen the customer expectations and create new ways of living/working.



With Cloud Technology, we're right in the middle of another paradigm shift. Because it is transforming how businesses create value, how people work, and ultimately how people live.



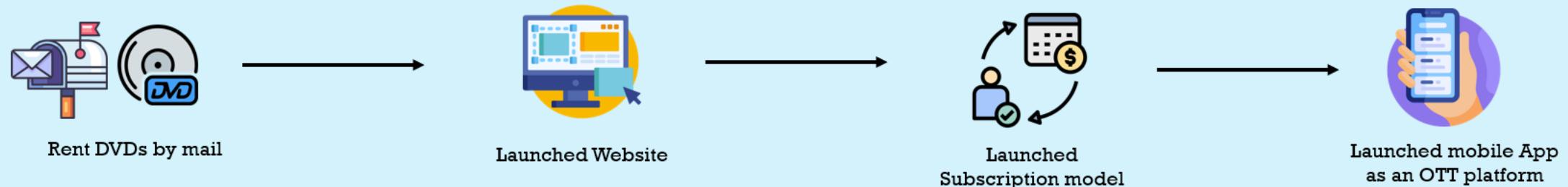
As with any other invention wave, Cloud technology also generating thousands of new innovations such as chat bots, Artificial Intelligence, Machine learning.

Why should Organizations consider Cloud



Organizations or Businesses have to respect the paradigm shifts and align themselves to the new inventions & innovations. Otherwise they will become irrelevant in the market and can't meet the end user expectations.

NETFLIX Story of Netflix which embraces new innovations regularly



Companies failed to embrace new innovations



Why should Organizations consider Cloud



There's no greater danger for any organization than to keep the technology it's always used to refine and perfect what it's always done instead of letting it go and moving on to the next technology platform. With Cloud Technology, it became mandatory for Organization to leverage it.



Abandoning old technology for a new one is commonly referred to as the "**Burning Platform Effect**". It requires organizations to take a leap of faith and to continually adapt as new technologies create new paradigm shifts. Successful companies consistently focus on "**“why”**" they exist, not "**“how”**" they operate.



When an organization takes advantage of new technologies such as cloud to redesign and redefine relationships with their customers, employees, and partners the result is a companywide **digital transformation**.



Companies such as Google Cloud have invested heavily in their own IT infrastructure, creating vast digital spaces to store and process data. Now they're helping other organizations around the world using Cloud by offering them the use of their digital platform to run their applications at scale.

Why should Organizes consider Cloud



The adoption of Cloud technology is not just about survival, it's about thriving in a new technological era. Many organizations are already realizing the following business benefits,

- ✓ global presence and connectivity
- ✓ pay-as-you-go pricing with zero installation, setup, or maintenance fees
- ✓ secure and fast application development and deployment (Agility)
- ✓ shift in the costs from capital expenditure to operational expenditure, or from CapEx to OpEx.
- ✓ modernizing infrastructure regularly
- ✓ compliance levels increases
- ✓ high availability, disaster recovery
- ✓ solving business problems at a much faster speed, and gaining insights from their data.
- ✓ scalability (a.k.a. elasticity) up or down
- ✓ higher fault tolerance with multi-regional capabilities
- ✓ IoT devices able to send & receive data
- ✓ automating processes and enabling real-time collaboration between people globally.

Cloud is revolutionary because it enables every professional to fundamentally rethink and re-imagine how they do business, from collecting data to gaining insights from it, to working with their peers globally, to serving their customers.

What makes Cloud sooo good....



Do you know what makes Cloud sooo good & cool. It is the water they hold & rain they gave.

I am just kidding 😊



Cloud provide unlimited computing/compute power, data storage, data process for any one including individuals. Gone are the days only bigger organization can only afford the larger infrastructure. With Cloud, all the organizations regardless of their size, location has access to unlimited resources globally.



By the way, what is computing & compute power ?

Computing at its most basic is any task that requires a calculation. In the context of Cloud, computing is the ability to process information, to store, retrieve, compare and analyze it, and automate tasks most often done by a computer program.

Compute power refers to the speed at which a computer is able to process data.



In the beginning, giant computers owned by institutions filled entire rooms and took a long time to process small amounts of data. The first change brought computers from institutions to individuals and exponential growth trajectory that doubled computing power every other year inline with **Moore's Law**. This was only the beginning.

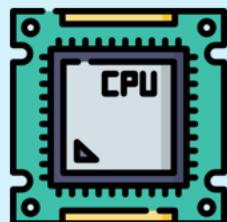
Later on **Moore's Law** has been disrupted twice by radical new designs in chips.

What makes Cloud sooo good....



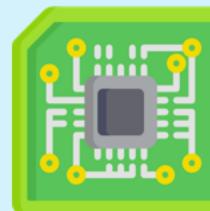
By the way, what is **Moore's Law** ?

In 1965 Gordon Moore observed the number of transistors in a dense integrated circuit will double every 2 years, thereby increasing processing/computing power. Moore's Law survived for over 50 years but it is dead due to radical new designs in chips.



CPU (Central Processing Unit)

TPUs are 50x times more powerful than CPUs



TPU (Tensorflow Processing Unit)

Quantum processor are hundred million times (100mx) more powerful than TPUs



Quantum Processing

EXAMPLE

For examples, a machine learning model that would require a days of training with traditional CPUs only requires 30 mins of training with TPUs. Similarly a machine learning model that would require thousands of years of training with traditional CPUs would now require only a few seconds with quantum computing.



So that Moore's Law has been disrupted twice by radical new designs in chips and now Moore's law considered dead.

What makes Cloud sooo good....



New evolution chips like TPUs & Quantum processors have a limitation though. Right now, the cooling requirements for these new processors can only be met in large industrial environments.



TPUs need to be cooled with pressurized water within the chip, while quantum computing requires absolute zero temperatures to operate. Cloud data centers are the only environment where we can create these conditions at scale.



The “**Internet of Things**” or “**IoT**” refers to everyday objects or devices that are connected to the internet and are able to send and receive data using Cloud. Today we can build highly accurate statistical models to predict complex behaviors and use that information to anticipate intent.



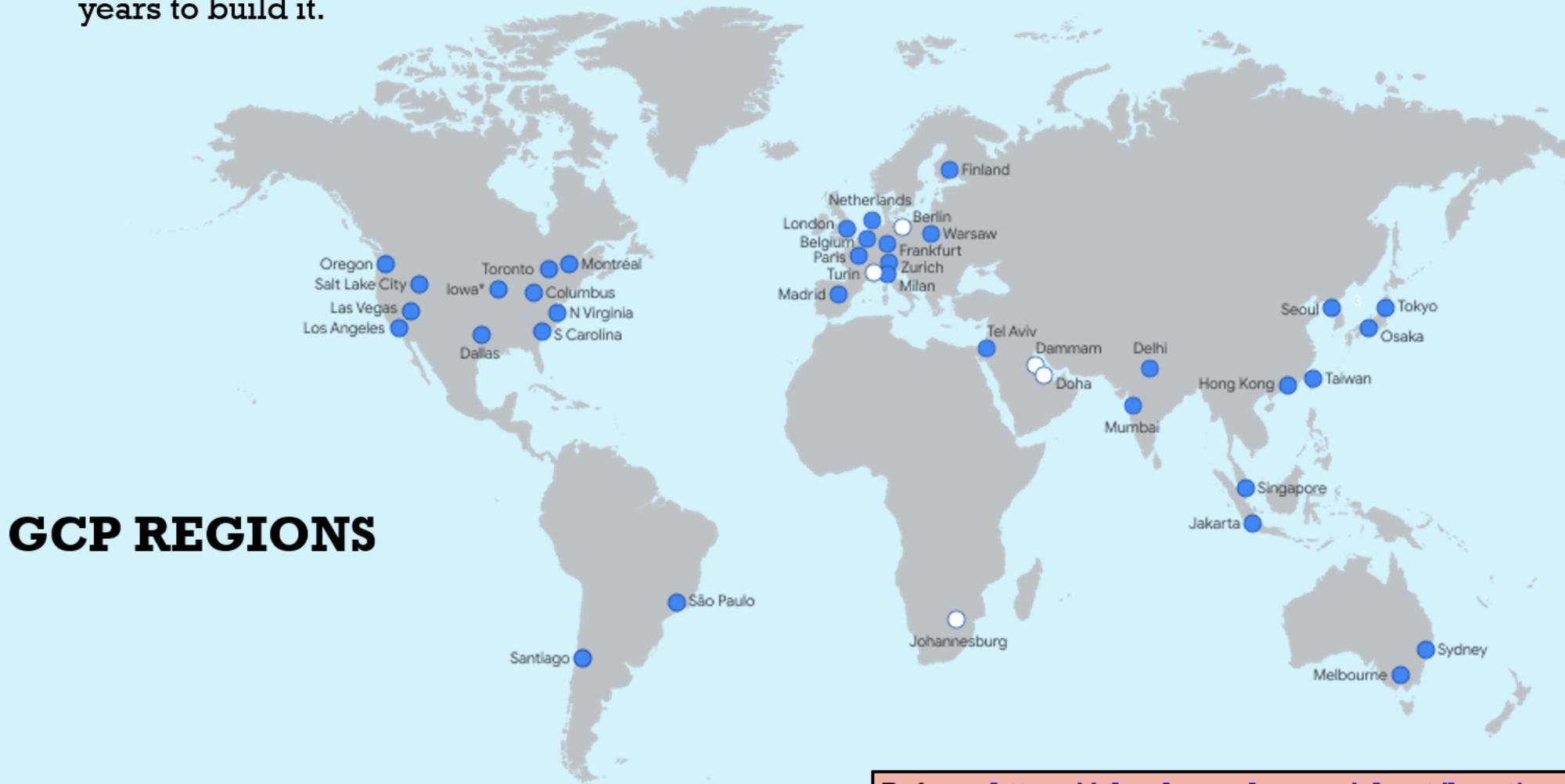
For the foreseeable future, Cloud will be the only option to tap into this vast amount of computing power & Data storing/processing.



Think about electricity, in the first years after its discovery, most users had generators located where they lived or worked. As the industry became more mature, power plants were created that we access through a grid. The same principle applies today to computing. It will mainly be generated in plants, which are the cloud data centers and access to a grid, which is the Internet.

What makes **Google** Cloud sooo good....

Google Cloud is **BIG**. Its infrastructure services are available in locations across North America, South America, Europe, Middle East, Asia, and Australia. Google's network carries as much as 40% of the world's internet traffic every day. In fact, Google's network is the largest of its kind on Earth, and Google has invested billions of dollars over the years to build it.

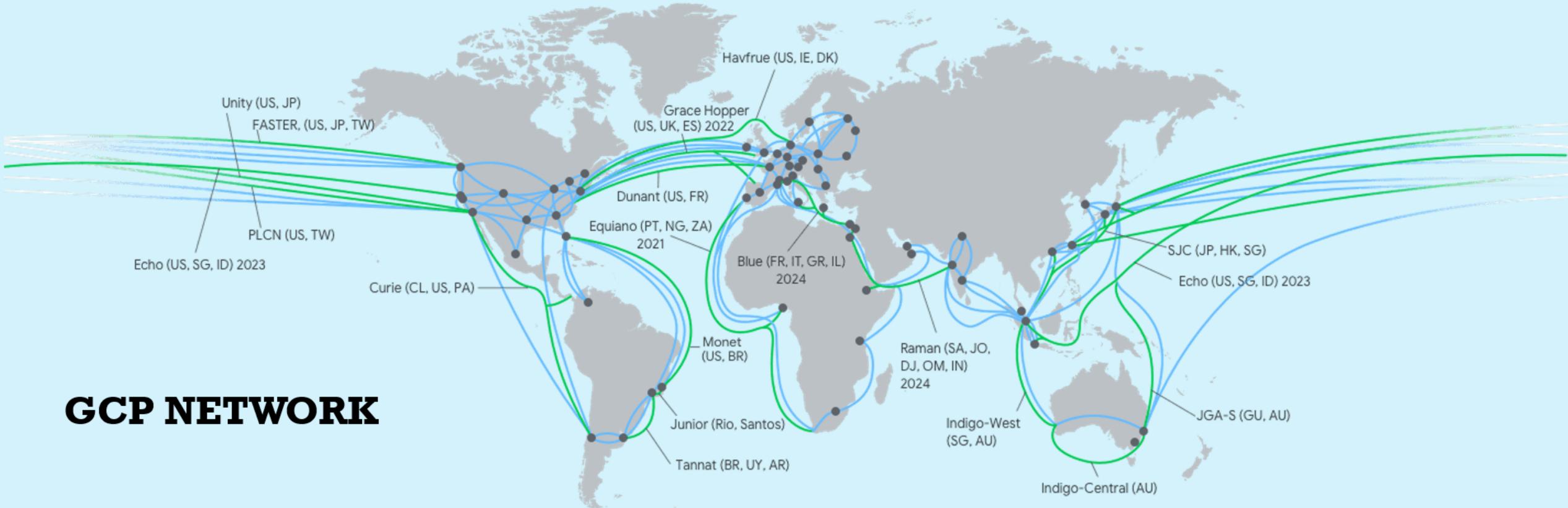


Refer to <https://cloud.google.com/about/locations> for latest data

What makes **Google** Cloud sooo good....



Google Cloud has a vast real-world infrastructure consisting of massive data centers in over **106 zones** across **35 regions** found in over 200 countries, globally connected by a vast network of undersea cables.



What is a Region & Zone in GCP ?

Region is a data center in a particular geographic area of the world. For example **europe-west2** is in London, England & **us-west1** is in Oregon, USA.

Zone is a unit of isolation inside a region. Most of the regions has three isolation zones. For example, europe-west2 region has three zones **europe-west2-a, b, and c** would be the sub-units of the London region.



Typical Google Cloud Region & Zones



Data center
inside a Zone



Cooling &
supporting
infra

What is a Region & Zone in GCP ?



Regions are collections of zones. Zones have high-bandwidth, low-latency network connections to other zones in the same region. In order to deploy fault-tolerant applications that have high availability, Google recommends deploying applications across multiple zones and multiple regions. This helps protect against unexpected failures of components, up to and including a single zone or region.



A zone is a deployment area within a region. The fully-qualified name for a zone is made up of `<region>-<zone>`. For example, the fully qualified name for zone `a` in region `us-central1` is `us-central1-a`.



Speaking physically, every product in Google Cloud is realized in a zone, in multiple zones of a region, in multiple regions, or globally. When you're planning an architecture, one of the things that should be on your to-do list is to determine where exactly to run each of the services you're planning on using.



Choose regions that makes sense for your scenario. For example, if you only have customers in the US, or if you have specific needs that require your data to live in the US, it makes sense to store your resources in zones in the `us-central1` region or zones in the `us-east1` region.



Regions & Zones helps in achieving **High Availability (HA)**, **Disaster Recovery (DR)**, **Backup**

What is High Availability ?



High availability is a quality of computing infrastructure that allows it to continue functioning, even when some of its components fail. This is important for mission-critical systems that cannot tolerate interruption in service, and any downtime can cause damage or result in financial loss.

Highly available systems guarantee a certain percentage of uptime—for example, a system that has 99.9% uptime will be down only 0.1% of the time - 0.365 days or 8.76 hours per year. The number of “nines” is commonly used to indicate the degree of high availability. For example, “five nines” indicates a system that is up 99.999% of the time.

The three elements that are essential to a highly available system are **Redundancy** (redundant component that can take over in case of failure), **Monitoring** (detecting when a component fails or stops responding), **Failover** (switch automatically from the currently active component to a redundant component, if monitoring shows a failure of the active component)

Availability of a system can be calculated using,

$$\text{Availability} = (U / (U + D)) \times 100\%$$

U = Uptime, D= Downtime

What is High Availability ?

Below table shows how a availability percentage converts to yearly downtime,

PERCENTAGE	YEARLY DOWNTIME*
99.9 (Three 9's)	8 Hrs 45m 57s
99.99 (Four 9's)	52m 34s
99.999 (Five 9's)	5m 15s
99.9999 (Six 9's)	31s
99.99999 (Seven 9's)	3s
99.999999 (Eight 9's)	0.3s

* - Approximate time

Steps to calculate yearly downtime based on availability percentage,

- 1 Lets try to calculate yearly downtime for Five 9's which is 99.999. Subtract the given availability from 100. Like shown below 0.001% is the downtime %
 $100 - 99.999 = 0.001\%$
- 2 Take down time % and divide it with 100 as we want to convert the number from % to normal number
 $0.001/100 = 0.00001$
- 3 Take the value from above step and multiply with $365 * 24 * 60 * 60$ (Days*Hours*Minutes*Seconds)
 $0.00001 * 365 * 24 * 60 * 60 = 315s$
- 4 The above value is in seconds. Divide it with 60, to get time in mins
 $315/60 = 5m 15s$

What is DR & Backup ?

Disaster recovery (DR) is the process by which an organization anticipates and addresses technology-related disasters. IT systems in any company can go down unexpectedly due to unforeseen circumstances, such as power outages, natural events, or security issues. Disaster recovery includes a company's procedures and policies to recover quickly from such events.

Outages come in many forms, including the following examples:

- ✓ An earthquake or fire
- ✓ Technology failures
- ✓ System incompatibilities
- ✓ Wars
- ✓ Intentional unauthorized access by third parties

Backup allows you to store business critical data, information, files etc. in a different physical location. This serve critical use cases such as recovery from data corruption, data loss, ransomware recovery, or database cloning etc.

Backups are critical for Safeguarding Data and Apps, Rapid Data Recovery, Reliable Disaster Recovery

For example, Critical applications like Banks, they store their business data, transactional data, customer data at multiple regions to handle business disasters and achieve business continuity all the time.



Cloud Computing Delivery models

Public Cloud

Public cloud, in a nutshell, is a shared entity computing model. Hardware and resources such as compute, storage, and networking are owned by the cloud provider and shared with other tenants on the platform, known as multi-tenant or multi-tenancy.

Private Cloud

Private cloud, in a nutshell, is a dedicated entity computing model. Hardware and resources such as compute, storage, and networking are dedicated to your organization use only; this is single-tenant.

Hybrid Cloud

Hybrid cloud, in a nutshell, is a combination of a shared entity (multi-tenant) computing model and a dedicated entity (single-tenant) computing model. Some computing resources you choose to have running in your private cloud environment and some resources you choose to have running in a public cloud environment based on your needs.



Public Cloud

A public cloud is a pool of virtual resources—developed from hardware owned and managed by a third-party company—that is automatically provisioned and allocated among multiple clients through a self-service interface. It's a straightforward way to scale out workloads that experience unexpected demand fluctuations.

Metered pricing and consumption-based billing and pay-as-you-go monthly usage costs; you only pay for the resources you use, which can allow cost control and cost management. Offer almost unlimited resources at any time. Provides scalability & agility.

Thousands of providers all over the world provides public cloud. But Google Cloud, Amazon Web Services (AWS), Microsoft Azure are among the largest—and most popular—ones today

Least control over security, protection, and compliance; you do not have complete control over security and compliance with the public cloud model.



multi-tenant

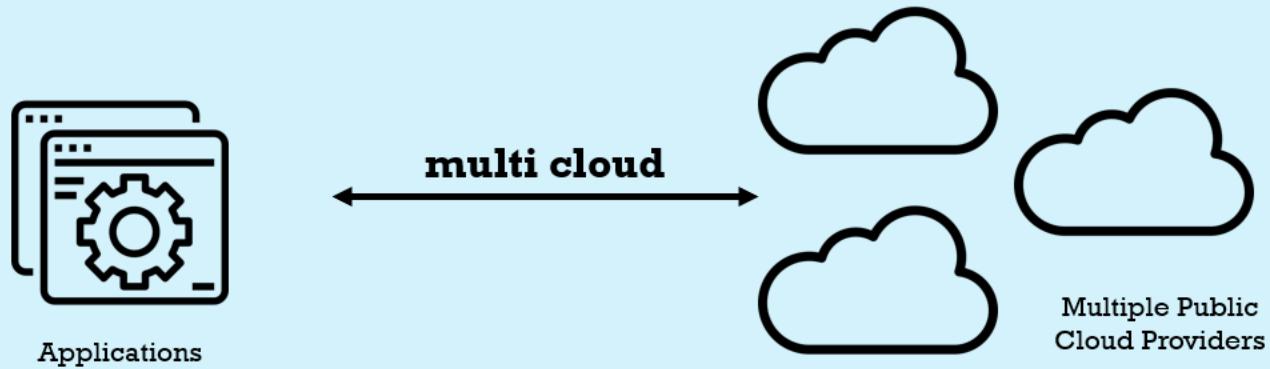


Public cloud is like an apartment block, where you are a tenant that shares the building with other tenants; you pay rent to a landlord for your apartment. In cloud computing, service provider play the role of landlord.



Public Cloud

If you use various public cloud providers like GCP, AWS, Azure, then it is called as **multi cloud**.



Benefits of multi cloud includes the following,

- ✓ Avoids vendor lock-in
- ✓ Organizations will have more options, products, flexibility
- ✓ Possible cost savings
- ✓ Increases redundancy
- ✓ Protect enterprises from outages



Private Cloud

In Private cloud, computing resources/hardware can be in your facility that you own or a third-party hosting provider, colocation data center facilities provider. Alternatively, this could be their hardware that they dedicate to you, which is traditional dedicated server hosting.

■ Private clouds are cloud environments solely dedicated to the end user, usually within the user's firewall. Although private clouds traditionally ran on-premise, organizations are now building private clouds on rented, vendor-owned data centers located off-premise. All clouds become private clouds when the underlying IT infrastructure is dedicated to a single customer with completely isolated access.

■ Providers like VMware, Redhat provides private cloud. In most the scenarios, Organizations set up their own private cloud.

Private cloud provides complete control over hardware, physical resources, security, and compliances.



single-tenant



Private cloud is like a house as opposed to an apartment, you are the single tenant, and you do not share the building with any other tenants. You either own the building or you rent the property and pay a landlord.

Private clouds are very common for domains of medical, research, defence, scientific and manufacturing.



Hybrid Cloud

The hybrid cloud model provides a choice of creating some computing resources created in the service providers' public cloud computing platforms; some resources are created in your on-premises private cloud platform; both these resources are connected via the internet or a private managed network.



The most important hybrid cloud benefits are,

- ✓ Flexibility on expenditure, security, protections and compliance.
- ✓ Improved performance and reduced latency
- ✓ Improved ROI

A hybrid cloud setup may include multiple cloud providers making it both hybrid and multicloud. Using one or more public clouds in addition to your on-premises servers can facilitate and simplify management of your application and help you get more from the cloud.



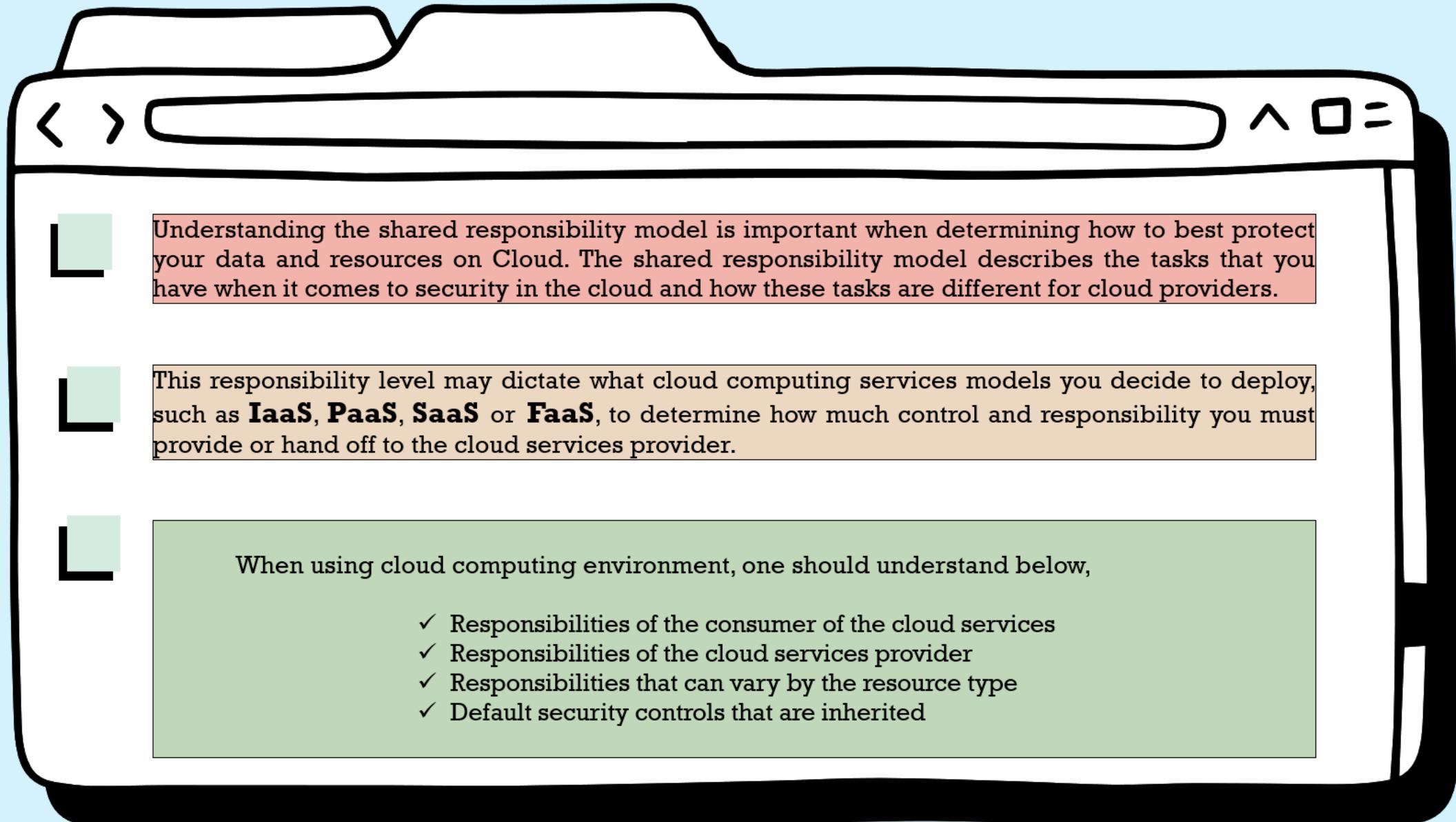
Combination of single-tenant & multi-tenant



A hybrid cloud approach is one of the most common infrastructure setups today because you can continue to use your on-premises servers while also taking advantage of public cloud options like Google Cloud.



Shared Responsibility model



- ✓ Cloud computing is all about abstraction. This abstraction model approach means removing layer(s) that you no longer need to care about; the layer still exists, but it is being handled by somebody else and frees up resources to concentrate on other layers that are of more value.
- ✓ The service models or categories of cloud computing define what layer of access and control the cloud services platform provider is responsible for and what the consumer of the cloud resources is responsible for. Below are the various service models available,



01 Infrastructure as a service (IaaS)

02 Platform as a service (PaaS)

03 Software as a service (SaaS)

04 Function as a service (FaaS) or serverless



IaaS

IaaS provides compute, storage, and network services on demand with pay-as-you-go pricing. You can use IaaS if you plan on migrating an existing on-premises workload to the cloud using lift-and-shift, or if you want to run your application on particular VMs, using specific databases or network configurations.

 In a nutshell, IaaS is a model where you can host your virtual machines and infrastructure services on hardware provided for you and shared with other tenants. The cloud provider is responsible for providing all layers up to and including the hardware.

 You create the virtual machine (install an OS and software), storage, and computing resources as you would in a traditional on-premises computing model.

In IaaS, the bulk of the security responsibilities are consumers, and CSP (Cloud Service Provider) responsibilities are focused on the underlying infrastructure and physical security.


You have direct access and complete control of the virtual machine, the operating system, and any roles/services such as the web server, application server, or SQL server that may be required to be installed/running on the virtual machines, as well as complete control over decisions on networking, security, and protection.

The following are examples of GCP IaaS resources:

Compute Engine, Cloud Storage, Cloud VPN etc.



PaaS

PaaS provides the runtime environment that you can develop and run your applications in. You can use PaaS if you're building an application (such as a website), and want to focus on development not on the underlying infrastructure.



 PaaS provides a ready-to-use environment and platform for faster deployment of web applications, code, business logic, data etc. Using pre-deployed resources, development frameworks, languages, and runtimes provided as a service, there is a quicker time to value and consumption of the service.

 The cloud provider is responsible for providing all layers up to and including the compute. You must still select underlying compute resources sizing to host your app, code, and so on. You would have no direct access or control of the virtual machine or any applications, services, or roles that may be installed.

In PaaS, CSP (Cloud Service Provider) are responsible for more controls than in IaaS, including network controls. Consumer share responsibility with CSP for application-level controls and IAM management. You remain responsible for your data security and client protection.

Since the service provider is responsible for the compute and storage resources layer, it gives you the least control and least flexibility. Still, it requires the least amount of management, administrative, and operations overhead

The following are examples of GCP PaaS resources:

**App Engine, Cloud SQL,
BigQuery,
Google Kubernetes Engine (GKE)**



SaaS

SaaS provides online applications that you can subscribe to or pay for in some way. You can use SaaS applications when your enterprise doesn't have the internal expertise or business requirement to build the application themselves, but does require the ability to process workloads.

- The cloud provider installs the application/solution and is responsible for its updates, scalability, availability, and security.
- The cloud provider is responsible for providing all the layers up to and including the applications; you are not responsible for any layers above other than the configuration and consumption of the app

In SaaS, CSP (Cloud Service Provider) own the bulk of the security responsibilities. Consumer remain responsible for their access controls and the data that they choose to store in the application.

SaaS provides the greatest time to value as there is no development time or resources required to create an application; it can be directly configured as needed and used instantly.

The following are examples of GCP SaaS resources:

Google Workspace,
AI/ML Services,
Third-party SaaS applications
that are available in GCP
Marketplace



FaaS

FaaS provides the platform for developers to run small, single-purpose code (called functions) that run in response to particular events. You would use FaaS when you want particular things to occur based on a particular event. For example, you might create a function that runs whenever data is uploaded to Cloud Storage so that it can be classified.

The term serverless itself is a bit of a misnomer, as in reality, there are servers involved, much like wireless does have wires involved at a certain point in the solution; it's more the fact that the servers do exist, but you don't need to know or care that they exist for you to have your desired outcome met.

Serverless computing enables developers to build applications faster by eliminating the need for them to manage infrastructure. With serverless applications, the cloud service provider automatically provisions, scales, and manages the infrastructure required to run the code

FaaS has a similar shared responsibility list as SaaS. Some of the caveats of FaaS are event-based workloads are the best use case, the execution environment cannot be customized, the cloud provider supports only specific languages and runtimes, long-running tasks are not well suited.

Inside FaaS, you only have control over your application, code, and business logic layer; all other layers are provided as a service that you have no access or control over. Essentially, you take the layers supplied to you and use that to execute (run/launch) your code/workflow.

Inside GCP, **Cloud Functions** is a FaaS application.



Shared responsibilities across various service models

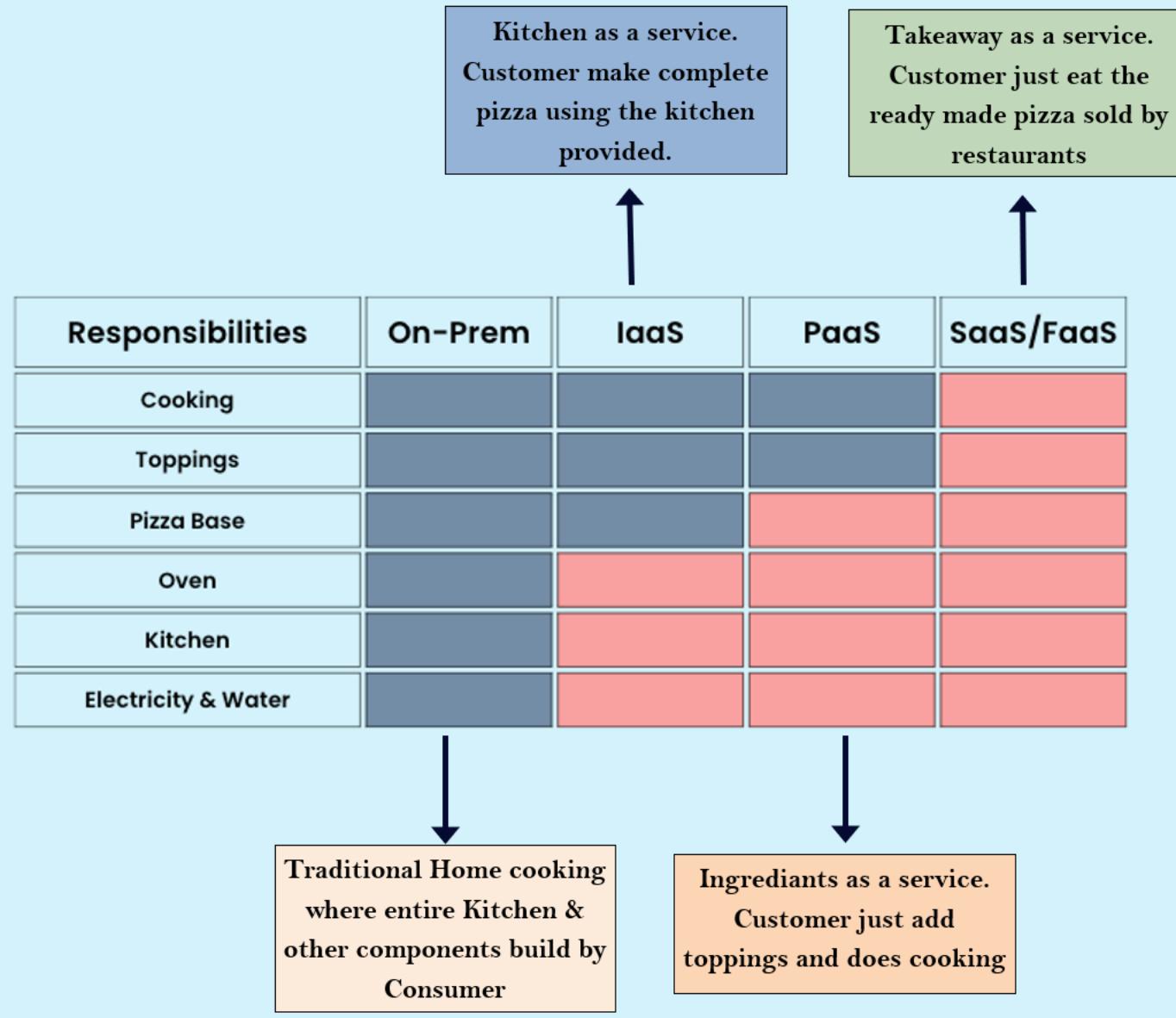
Responsibilities	On-Prem	IaaS	PaaS	SaaS/FaaS
Content				
Access policy				
Usage				
Deployment				
Web application security				
Identity				
Operations				
Access & Authentication				
Network Security				
Guest OS, Data & Content				
Audit logging				
Network				
Storage & encryption				
Hardened kernel & IPC				
Boot				
Hardware				



Customer
Responsibility

Cloud Provider
Responsibility

Service models with Pizza making scenario



Google Cloud Digital Leader curriculum

Introduction to digital transformation with Google Cloud
(approximately **10%** of the exam)

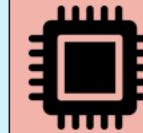
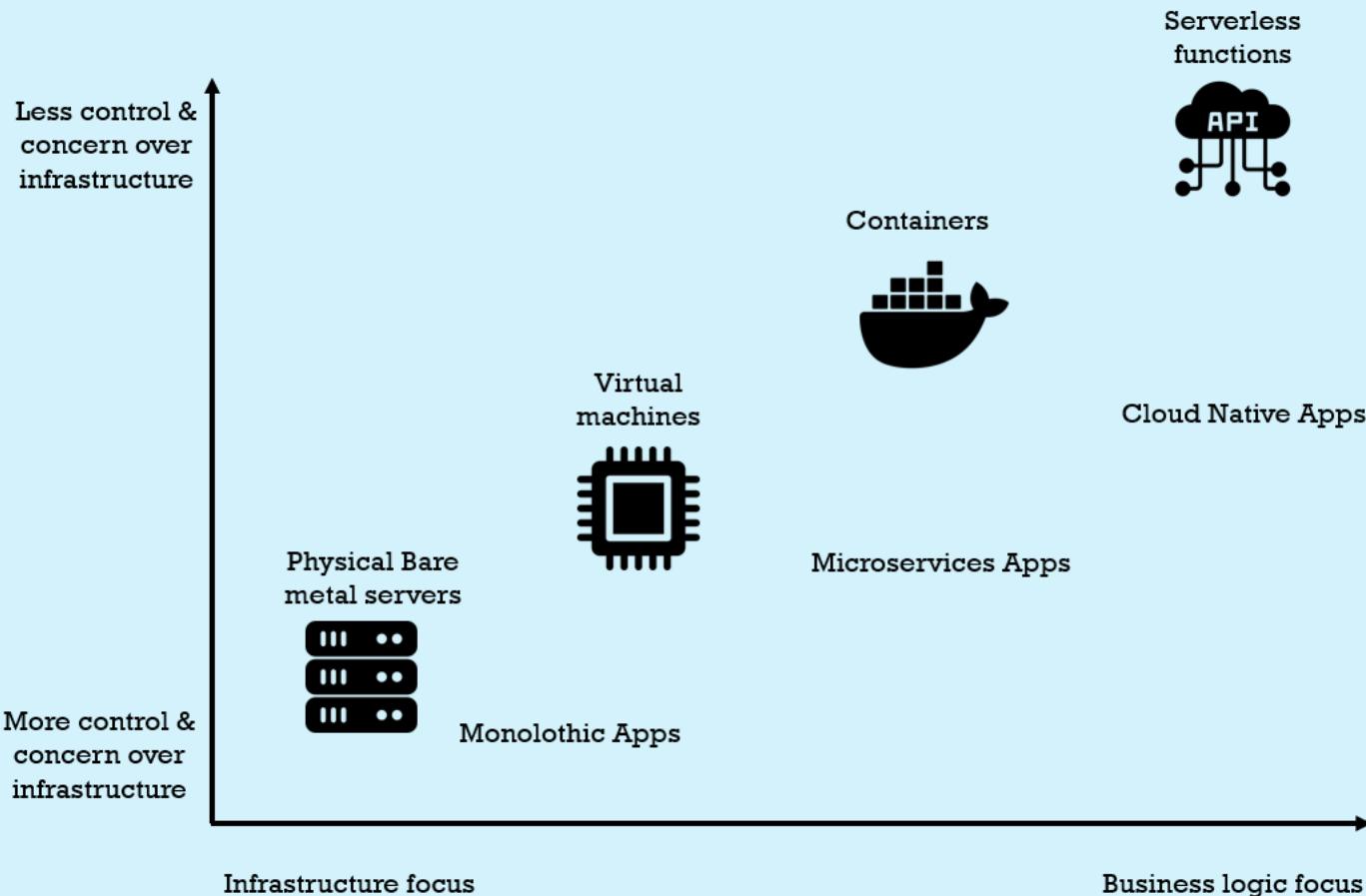
Infrastructure and application modernization
(approximately **30%** of the exam)

Innovating with data and Google Cloud
(approximately **30%** of the exam)

Understanding Google Cloud security and operations
(approximately **30%** of the exam)

SOURCE: <https://cloud.google.com/certification/guides/cloud-digital-leader>

Evolution of Cloud Computing architectures



Virtula machines abstracts the hardware. With a single jumbo bare metal server, we can built multiple VMs/servers using virtualization.



Containers abstracts the OS which makes them light weight compared to VMs. Containers are being used extensively inside the microservices.



Serverless abstracts the runtime which means we provide only business logic and don't worry about servers. Cloud provides take care of running your code on their compute later.

In order to relavent in the market, Organizations now need to be...



- 1 Innovative
- 2 Agile
- 3 Fast
- 4 Customer-focused



The culture of an organization has a direct impact on employees willingness to innovate.

1

Innovation relies on people being able to try things and failing without judgement.

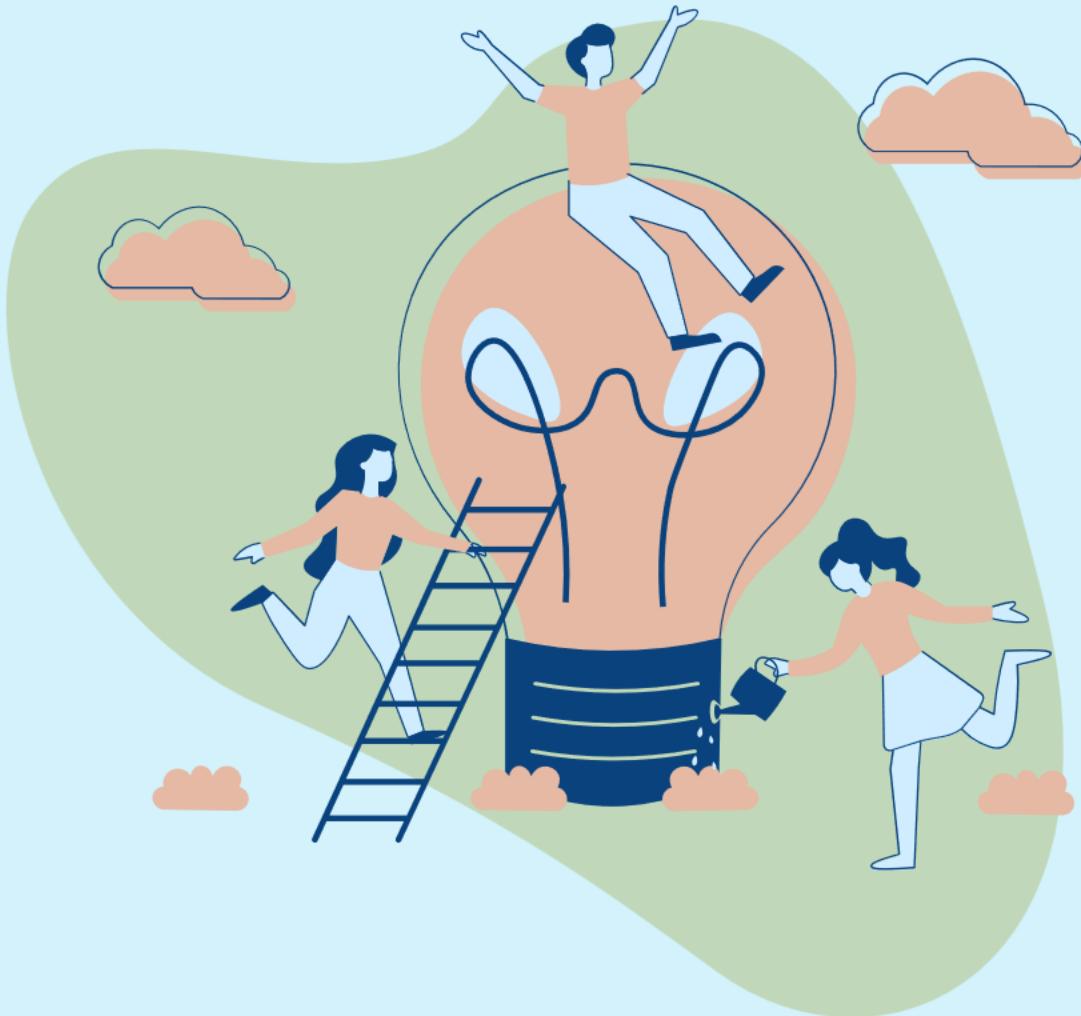
2

The elements that impact innovation are **how teams are structured, how content is managed, and how communication flows across an organization**

3

Innovation will leads to **modernization** of the Organizations.

The Google Cloud Solution Pillars for modernization are:



1

Infrastructure modernization

Google Cloud and its partners offer flexible infrastructure modernization approaches from rehosting customer's existing IT to replatforming.

2

Business applications platform portfolio

With Google Cloud's business application platforms portfolio, businesses can securely unlock their data with APIs, automating processes and creating applications across clouds and on-premises without coding.

3

Application modernization

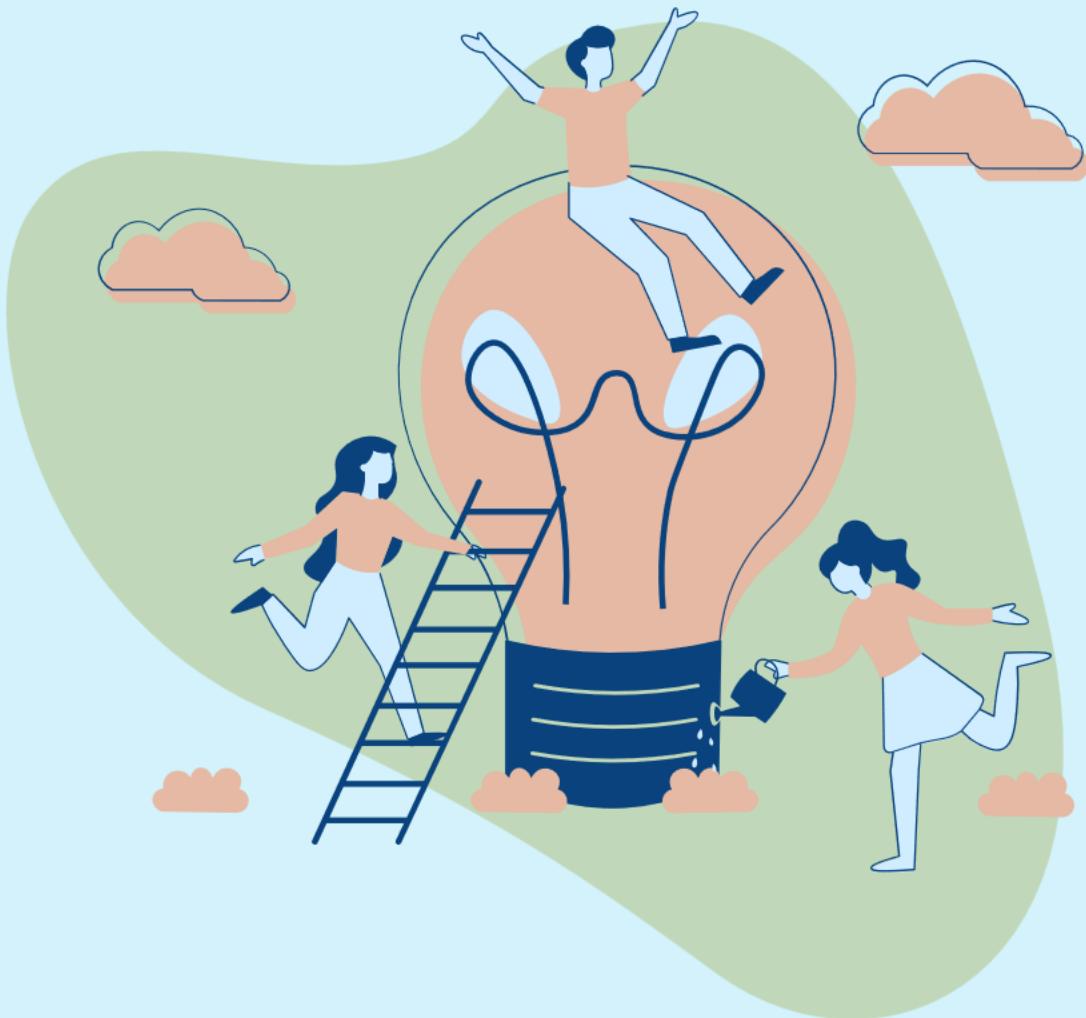
The term 'application' is widely used to refer to programs and software that enable people to perform various digital tasks. Today's customers expect instant access to services wherever they are. The tools within this pillar help businesses develop and run applications anywhere.

4

Database and storage solutions

Google Cloud's database and storage solutions include tools that help businesses migrate and manage enterprise data with security, reliability, high availability, and fully managed data services.

The Google Cloud Solution Pillars for modernization are:



5

Smart analytics

The smart analytics portfolio helps businesses generate instant insights from data at any scale with a serverless, fully managed analytics platform.

6

Artificial intelligence

Google Cloud's artificial intelligence tools are built to enhance innovation and improve productivity, by integrating seamlessly into a company's existing workflow and products.

7

Security

Google Cloud's comprehensive security solutions cover all aspects of protecting your business in this digital era. Businesses are able to detect, investigate, and protect themselves against online threats before attacks result in damage or loss.

To help organizations optimize their cloud adoption, Google developed the Google Cloud Adoption Framework. This best practice guide provides a framework to assess where an organization is in its journey and what they need to do next.

<https://cloud.google.com/adoption-framework>

Categorizing GCP Compute Services

VIRTUAL MACHINES



Compute
Engine



VMWare
Engine



Bare
Metal

CONTAINERS



Kubernetes
Engine



Cloud
Run



Anthos

SERVERLESS



App
Engine



Cloud
Run



Cloud
Functions

Modernizing IT Infrastructure



Owning and operating infrastructure limits an organization's staff in several ways: they have to undertake laborious tasks related to infrastructure, they are using legacy systems that are old, and they cannot scale with any ease. Whereas with Cloud, Organizations are seeing significant financial benefits, as their approach to IT moves from buying fixed capacity to paying only for what they use.



The first option in moving away from an on-premises infrastructure is **colocation**. A business sets up a large data center and then other organizations rent part of that data center. This means organizations no longer have to pay the costs associated with hosting the infrastructure, but they still need to pay to maintain it.



Hardware is often heavily under utilized, even in the colocation model, so engineers packaged applications and their operating systems into a virtual machine.



By the way, what is a virtual machine ? Let's try to understand....

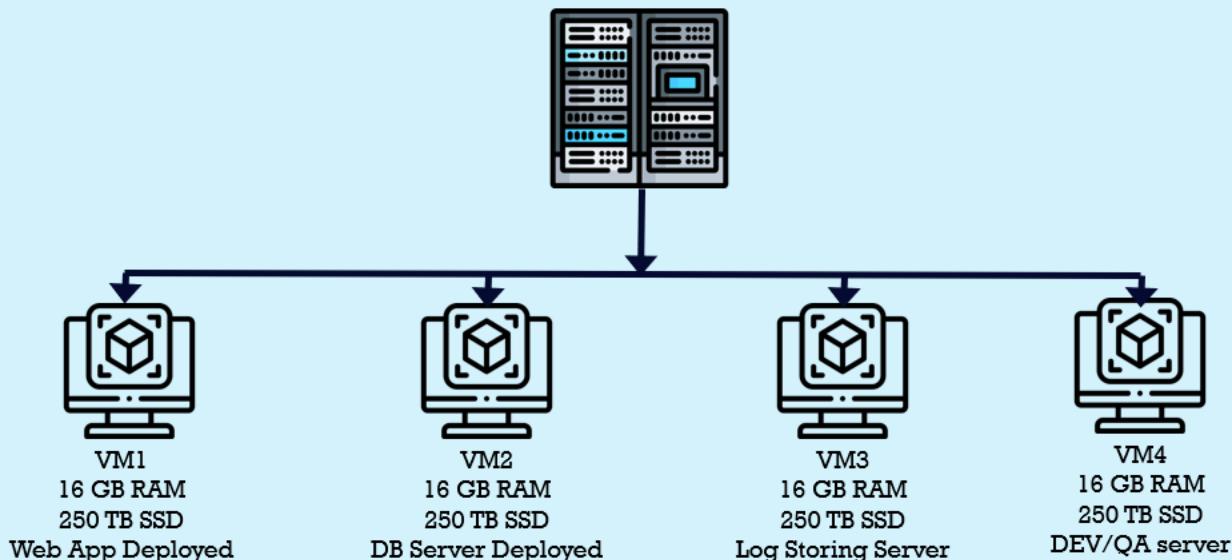
What is a Virtual machine ?



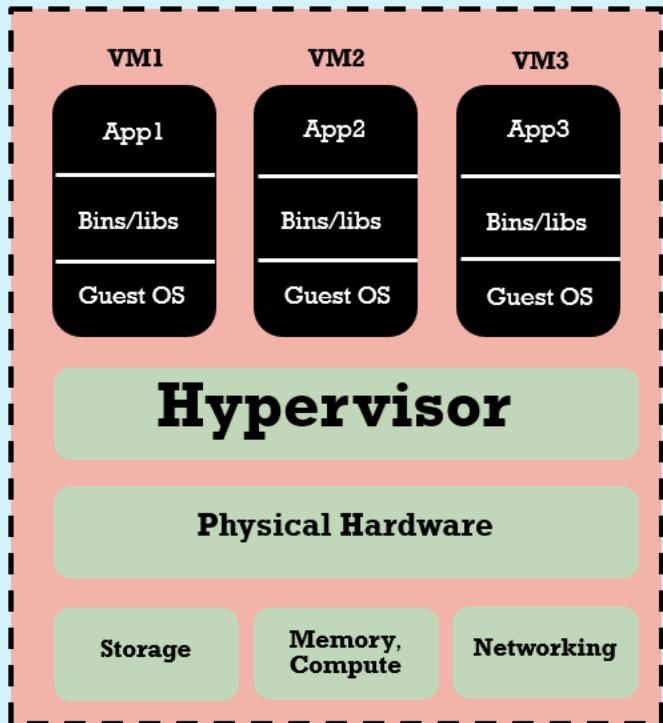
Think like you have a physical bare metal server with a capacity of 64GB RAM & 1000 TB hard disk. Back in the stone age, Developers have no option other than using it as a single machine where they have to install/deploy all the applications into it. There is no separation of Apps & often result in free space/memory getting wasted.



With **virtualization**, a single physical hardware resources can be splitted and **Virtual machines** (VMs) can be created. VMs share and optimize the same pool of computer processing, storage, and networking resources. They also enable businesses to have multiple applications running at the same time on a server.



What is a Virtual machine ?



The **virtualization** software layer that enables creation of VMs is called a **hypervisor**. A hypervisor sits on top of physical hardware, and multiple VMs are built on top of it. It's like having multiple computers that only use one piece of hardware.

Modernizing IT Infrastructure



Many companies are now outsourcing their infrastructure entirely. They are growing to deliver their products and services to customers regionally and globally, and need to scale quickly and securely. Setting up and maintaining data centers and network connections that are optimal for their needs is expensive.



Outsourcing your IT needs at the infrastructure level is called **infrastructure as a service**. If your organization chooses to, it can move some or all of its infrastructure away from physical data centers to virtualized data centers in the cloud.

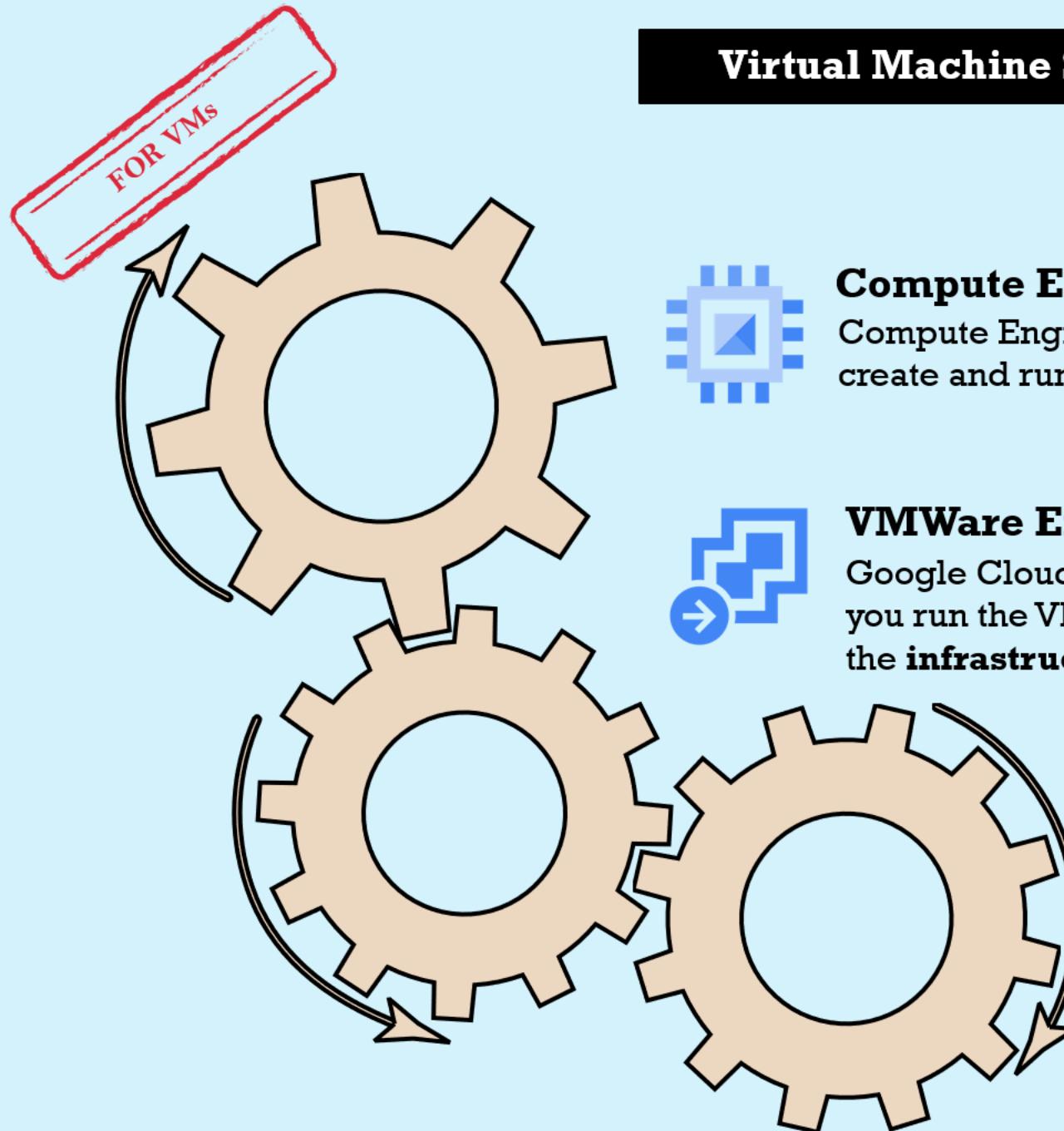


The maintenance work is outsourced to the public cloud provider so it's easier to shift a larger proportion of company expertise to build processes and applications that move the business forward.



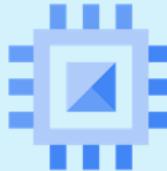
If you want a more managed service, cloud service providers offer something called: **platform as a service**. In this case, you don't have to manage the infrastructure and for some services you only pay for what you use.

Virtual Machine Services in GCP



Compute Engine

Compute Engine is a computing and hosting service that lets you create and run virtual machines on Google infrastructure.



VMWare Engine

Google Cloud VMware Engine is a fully managed service that lets you run the VMware platform in Google Cloud. Google manages the **infrastructure, networking and management** services.

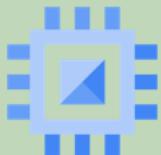


Bare Metal

Bare Metal enables you to migrate specialized workloads to the cloud, while maintaining your existing investments and architecture. This allows you access and integration with Google Cloud services with minimal latency.



Google Compute Engine (GCE)



GCE is someone else's computer. If you are used to ordering servers, or you regularly spin up VMs using technologies such as vSphere, KVM, or Hyper-V, and you want to do virtually the same thing in GCP, then this is the product you're looking for.

- 01 Compute Engine VMs boot quickly
- 02 Come with persistent disk storage
- 03 Deliver consistent performance
- 04 Highly customizable. Choose a predefined configuration or configure your own
- 05 Familiar to organizations currently using on-premises VMs
- 06 Pay per second, easy to scale vertically (bigger machine) and/or horizontally (more machine clones)

FOR VMs



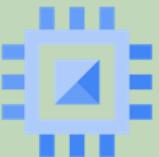
Insightful facts

Using GCE, you can pick the **chipset**, the **vCPU** count, the **memory**, the **size** of the boot disk, what **OS** comes loaded on it, whether it has externally accessible IPs, drive type sizes and counts, and a whole slew of other things. Then, you press the button, and in no time, your server will be up, running, and accessible.



FOR VMs

Google Compute Engine (GCE)



The following are the drawbacks of Compute Engine,

- 01 All that flexibility means more to manage, control, and secure
- 02 Frequently requires specialized skillsets, just like on-premises VMs
- 03 It has a high relative TCO
- 04 You take a big slice of the shared responsibility pie
- 05 Not suitable for microservices kind of applications, better for monolithic apps



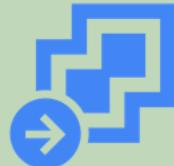
Use GCE when you're doing the following:

- ✓ Lifting and shifting VMs to the cloud with as few changes as possible.
- ✓ Running custom or off-the-shelf workloads with specific OS requirements.
- ✓ You really need the flexibility provided by a VM



VMWare Engine

FOR VMs



Easily lift and shift your VMware-based applications to Google Cloud without changes to your apps, tools, or processes. Includes all the hardware and VMware licenses to run in a dedicated VMware SDDC (software-defined data center) in Google Cloud. Benefits of VMWare engine includes,

- 01 fully dedicated and physically isolated from other customers
- 02 Operational continuity. You can continue to use the VMWare stack but with GCP bare metals
- 03 Infrastructure agility. New bare metal servers can be added on demand with in minutes.
- 04 Scale data center capacity in the cloud and stop managing hardware refreshes
- 05 Reduce risk and cost by migrating to the cloud while still using familiar VMware tools and skills
- 06 Enables high-speed, low-latency access to other Google Cloud services such as BigQuery, AI Platform, Cloud Storage, and more



Insightful facts

- ✓ VMWare engine is a managed VMware service with bare metal infrastructure that runs the VMware software stack on Google Cloud.
- ✓ Your on-premisis VMWare environment can be migrated with just a few clicks. GCP can provide an entire VMware SDDC in about 30 minutes across 14 global regions
- ✓ Suitable for private cloud infrastructure



Bare Metal

FOR VMs

 There are many workloads that are easy to lift and shift to the cloud, but there are also **specialized workloads** (such as **Oracle**) that are difficult to migrate to a cloud environment due to complicated licensing, hardware, and support requirements. **Bare Metal Solution** provides a path to modernize these applications.

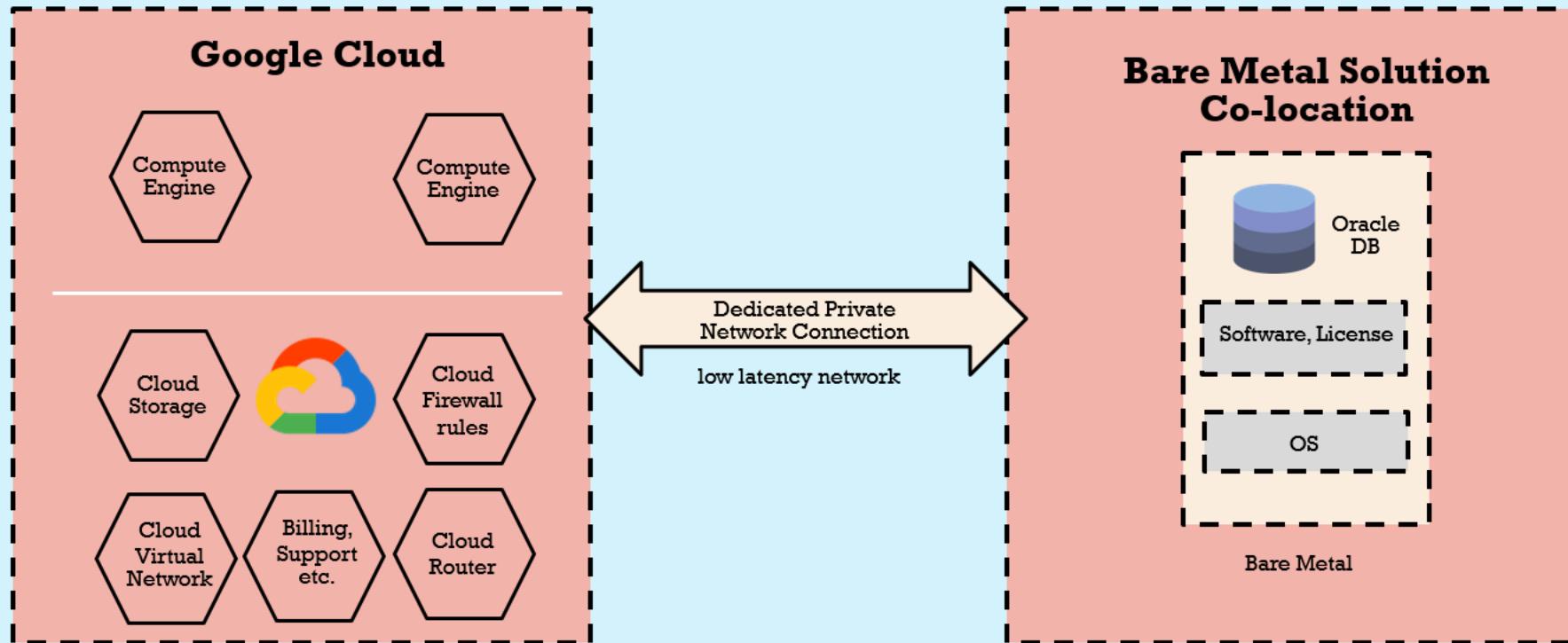
- 01 You can bring your own license of the specialized software such as Oracle
- 02 You are responsible for your software, applications, and data.
- 03 Offers GCP connectivity with 2ms latency
- 04 Primarily used for Oracle databases
- 05 Enables an easier and a faster migration path while maintaining your existing investments and architecture.

Insightful facts

- ✓ Bare Metal Solution provides purpose-built bare metal machines in regional extensions that are connected to Google Cloud by a managed, high-performance connection with a low-latency network fabric. Google Cloud provides and manages the core infrastructure, the network, the physical and network security, and hardware monitoring capabilities in an environment from which you can easily access all Google Cloud services.
- ✓ You can first lift and shift these special workloads to Bare Metal Solution so you can exit your data center and stop managing hardware; then you will be in a great position to modernize your application with Google Cloud.

Typical Bare Metal Setup

FOR VMs



The overall idea is clearly to make it easier for enterprises with specialized workloads that can't easily be migrated to the cloud to still benefit from the cloud-based services that need access to the data from these systems. Machine learning is an obvious scenario/example

Welcome to the world of Containers

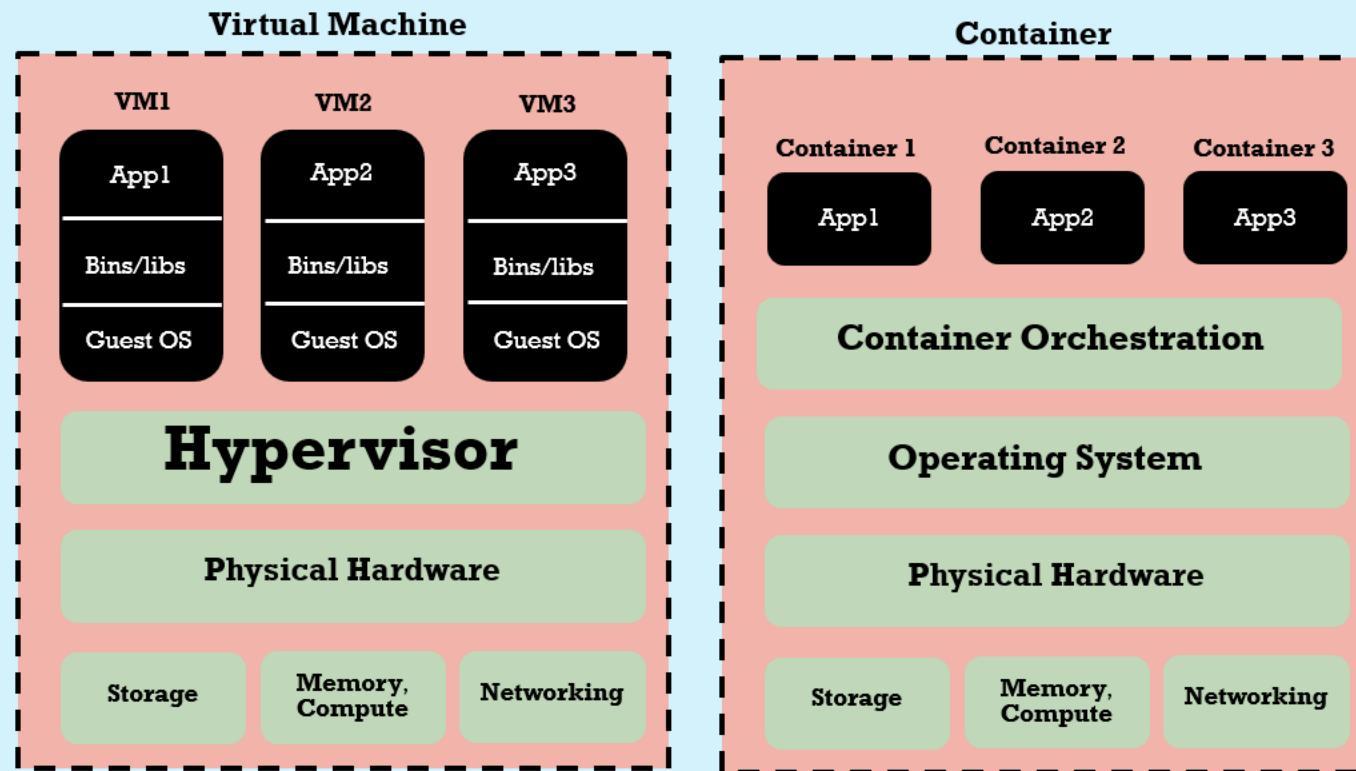


Containers and **virtual machines** are very similar resource **virtualization** technologies. Virtualization is the process in which a system singular resource like RAM, CPU, Disk, or Networking can be ‘virtualized’ and represented as multiple resources.



The key differentiator between containers and virtual machines is that virtual machines virtualize an entire machine down to the hardware layers and containers only virtualize software layers above the operating system level.

- ✓ VMs are heavy in nature as they virtualize hardware and create entirely new machine with separate OS
- ✓ VMs are time consuming to build, regenerate and restart.
- ✓ VMs take lot of storage
- ✓ Idle for traditional monolithic web apps



- ✓ Containers only hold exactly what's needed for the particular application that they support.
- ✓ All containers share same guest OS which makes them light weight
- ✓ Start faster, use less memory, and allow developers to create predictable envs
- ✓ Idle for modern microservices based web apps

Welcome to the world of Containers



A **container** is a loosely isolated environment that allows us to build and run software packages. These software packages include the code and all dependencies to run applications quickly and reliably on any computing environment. We call these packages as **container images**.



Software containerization is an OS virtualization method that is used to deploy and run containers without creating multiple virtual machine (VM). Containers can run on physical hardware, in the cloud, VMs, and across multiple OSs. **Docker** is one such containerization product which helps to create containers.



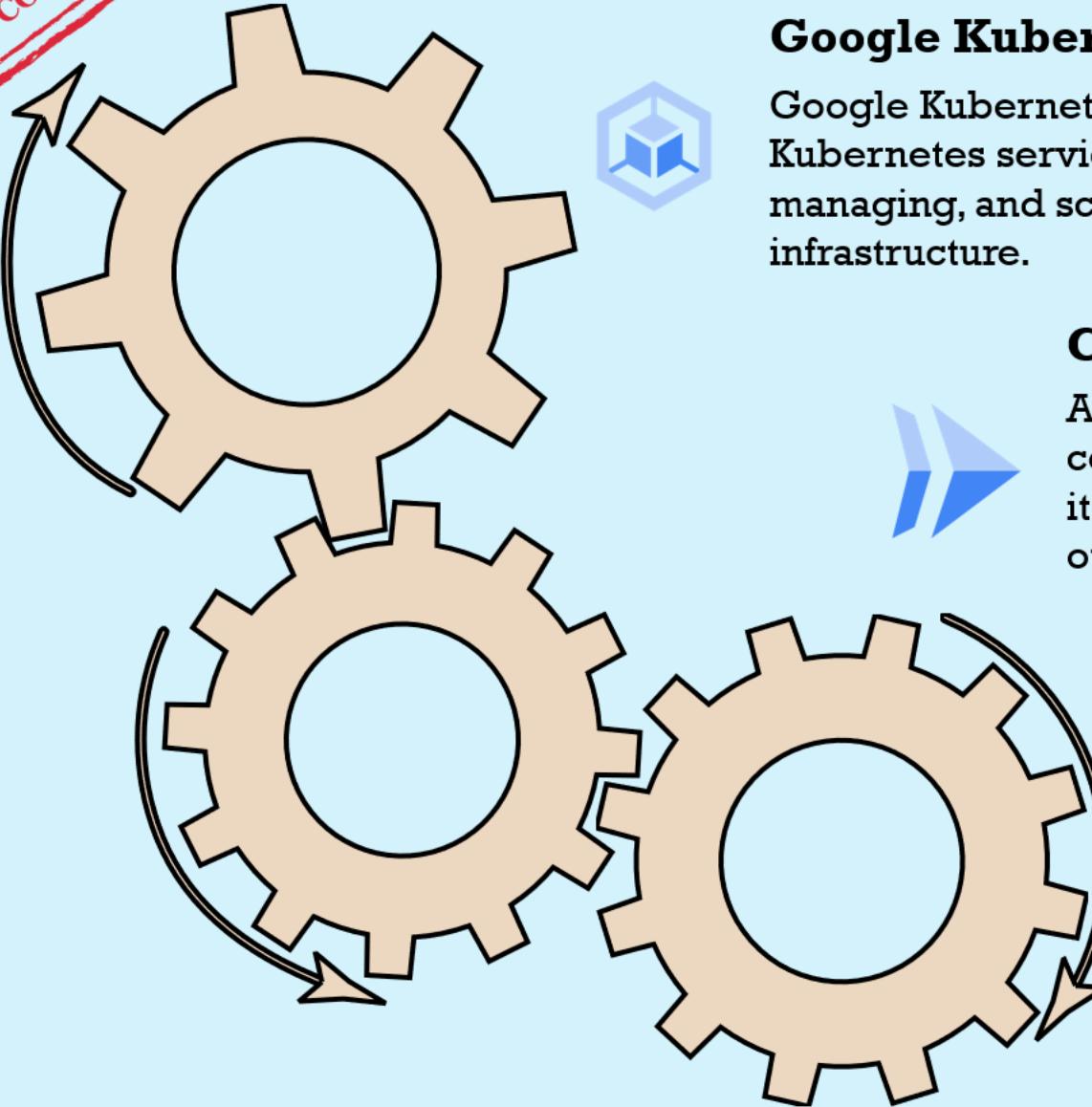
While Docker helps to create containers, we need a manager who can manage 100s of containers in microservices env. **Kubernetes** is one such open-source container orchestration platform originally developed by Google that helps managing all the containers inside a system.



Containers share the OS kernel, start much faster, and use a fraction of the memory compared to booting an entire OS. They help improve portability, shareability, deployment speed, reusability, and make it easier to isolate and repair issues that may arise.

Container Computing Services in GCP

FOR CONTAINERS



Google Kubernetes Engine

Google Kubernetes Engine (GKE) is a secured and fully managed Kubernetes service. It provides an easy to use environment for deploying, managing, and scaling your containerized applications using Google infrastructure.

Cloud Run

A fully managed serverless platform that runs individual containers. You give code or a container to Cloud Run, and it hosts and auto scales as needed to respond to web and other traffic.



Anthos

Anthos is a managed platform for all your application deployments, both traditional as well as cloud native. It enables you to build and manage global fleets and establish operational consistency across them.



Google Kubernetes Engine (GKE)



Google Kubernetes Engine or GKE provides a fully managed environment for deploying, managing, and scaling your containerized applications using Google infrastructure. The GKE environment consists of multiple machines grouped together to form a cluster.

FOR CONTAINERS

- 01 Easy way to get a Kubernetes(K8s) cluster in under 5 minutes.
- 02 Fully managed by Google with a typically lower TCO than GCE. So not as flexible as GCE
- 03 Based on a widely used open source technology
- 04 Very popular with companies modernizing toward containers.
- 05 You just get a cluster of VMs with K8s setup and still need people to configure, deploy the containers etc. If you don't have in-house K8s skills, then go for GKE Autopilot
- 06 If your application isn't containerized, then you can't run it in GKE. Like any k8s cluster, you require specialized k8s skills to manage it.



Insightful facts

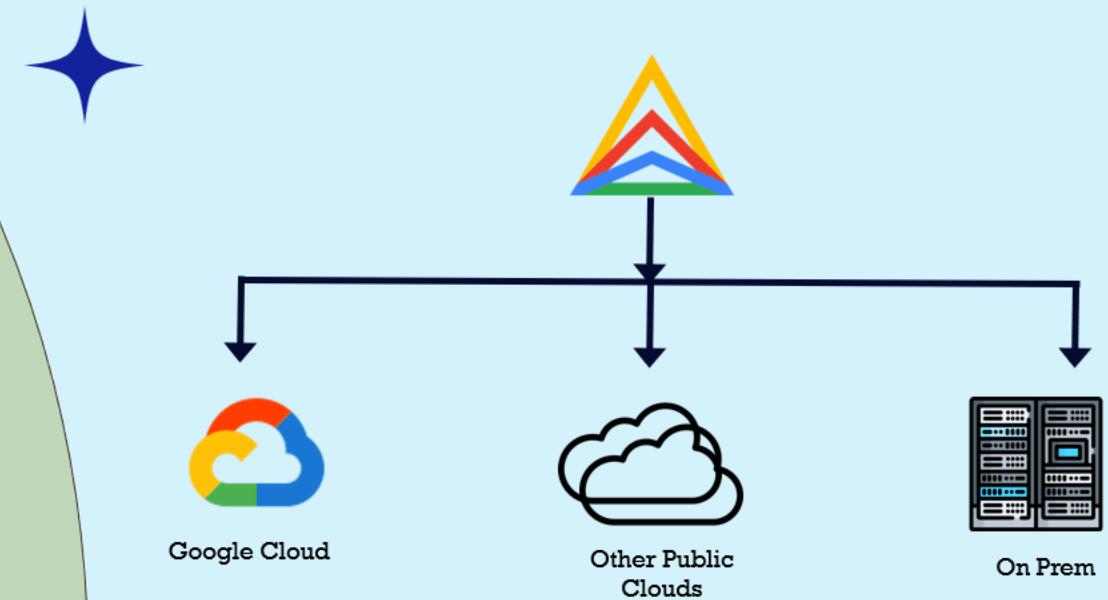
- ✓ GKE allows you to securely speed up app development, streamline operations, and manage infrastructure.
- ✓ GKE creates cluster of VMs to deploy the containers into them. Google takes over the responsibility of managing the OS, the machines, and their security and patching.
- ✓ GKE has two operational modes: **Standard** and **Autopilot**. With Standard you get a managed cluster with VMs that you can see where as with Autopilot, Google uses serverless to run your containers

Anthos

FOR CONTAINERS

Most enterprises have applications, containers in different locations - in their own data centers, in multiple public clouds. With **Anthos**, you get a consistent way to manage your infrastructure, with similar infrastructure management, container management, service management, and policy enforcement across your landscape.

- 01 Build, deploy, and optimize apps on GKE and VMs anywhere - simply, flexibly, and securely
- 02 Consistent development and operations experience for hybrid and multicloud environments
- 03 Accelerate your VM-based app migration journey to containers
- 04 Fully managed service mesh with built-in visibility
- 05 Modernizing your security for hybrid and multi-cloud deployments
- 06 CI/CD Anywhere: Google Cloud, on-premises, or other clouds



What is serverless computing?



Serverless **doesn't** mean there will not be any servers to run the code or process the Data.



? The term **serverless** itself is a bit of a confusion, as in reality, there are servers involved, much like wireless does have wires involved at a certain point in the solution; it's more the fact that the servers do exist, but you don't need to know or care that they exist for you to have your desired outcome met.



Serverless computing enables developers to build applications faster by eliminating the need for them to manage infrastructure. With **serverless** applications, the cloud service provider automatically provisions, scales, and manages the infrastructure required to run the code.

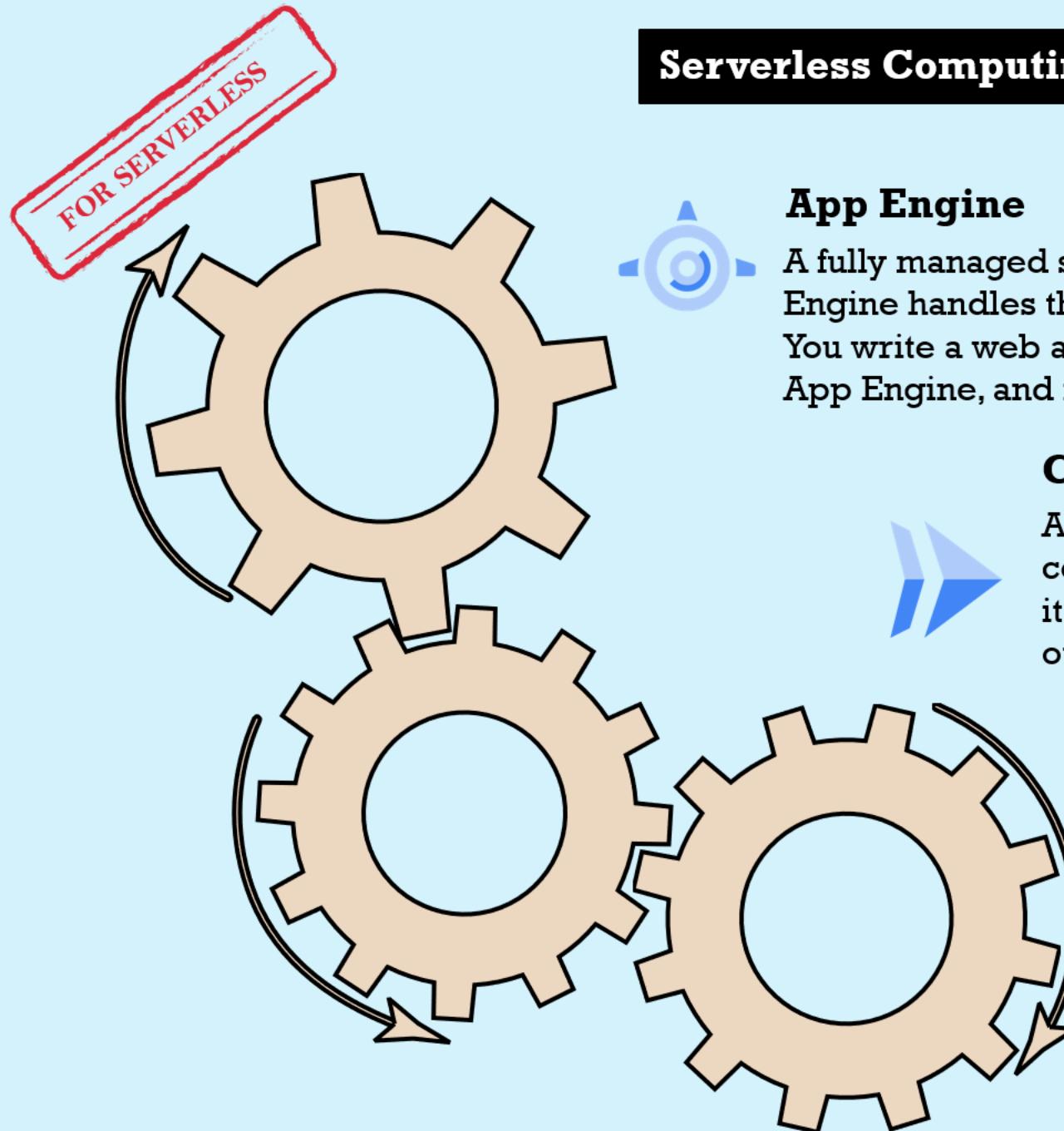


In Serverless there are still servers that execute the code passed down from the runtime layer – it's just that this layer is now further abstracted from you than it was in the IaaS and PaaS models. FaaS model represents Serverless in Cloud.



The benefit of **serverless** is that there are even fewer components to create and manage and allow development teams to focus on writing their core code without considering what's running their code; the provider takes care of automatically provisioning these resources to run the code. This means faster, more productive development teams, less operational overheads for DevOps teams, more significant innovation, and quicker time to value and return on investment in development resources.

Serverless Computing Services in GCP



App Engine

A fully managed serverless platform for complete web applications. App Engine handles the networking, application scaling, and database scaling. You write a web application in one of the supported languages, deploy to App Engine, and it handles scaling, updating versions, and so on.

Cloud Run

A fully managed serverless platform that runs individual containers. You give code or a container to Cloud Run, and it hosts and auto scales as needed to respond to web and other traffic.

Cloud Functions

Cloud Functions is a serverless execution environment for building and connecting cloud services. It offers scalable, pay-as-you-go functions as a service to run your code with zero server management.

App Engine

FOR SERVERLESS



App Engine is a fully managed serverless compute option in Google Cloud that you can use to build and deploy low-latency, highly scalable applications. App Engine makes it easy to host and run your applications.



- 01 Lets app developers build scalable web and mobile back ends in famous programming language on a fully managed serverless platform.
- 02 It scales them from zero to planet scale without you having to manage infrastructure.
- 03 Developer friendly - A fully managed environment lets you focus on code while App Engine manages infrastructure.
- 04 Proprietary to Google, with an App Engine-specific config file
- 05 Runs in a sandbox, so your code has to be particular versions of particular languages



Insightful facts

- ✓ With App Engine, you are restricted to particular versions of particular languages: Java, Node.js, Python, C#, PHP, Ruby, or Go. You upload your source code and a configuration file to App Engine using a command-line tool, and Google creates a sandbox around your application and runs it in that sandbox.
- ✓ Use App Engine if you want to deploy and host a web based application (HTTP/s) in a serverless platform. Examples: web applications, mobile app backends

Cloud Run

FOR SERVERLESS

FOR CONTAINERS



Cloud Run allows you to build applications in your favorite programming language, with your favorite dependencies and tools, and deploy them in seconds. It automatically scaling up and down from zero almost instantaneously-depending on traffic.

- 01 It scales them from zero to planet scale without you having to manage infrastructure.
- 02 Supports the CloudEvents which is not possible with App Engine
- 03 Developer friendly - Build container, put it in a registry somewhere Cloud Run can access, create your service, and you're off to the races.
- 04 Serverless platform for containerized applications
- 05 Easy to deploy and use, with a small learning curve



Insightful facts

- ✓ Use Cloud Run if you just need to deploy a containerized application in a programming language of your choice with HTTP/s and websocket support. Examples: websites, APIs, data processing apps, webhooks.
- ✓ No need to worry about scaling, managing the infrastructure. No Kubernetes, no clusters etc.



Cloud Functions

FOR SERVERLESS



Event-driven serverless functions. You write individual function code and Cloud Functions calls your function when events happen (for example, HTTP, Pub/Sub, and Cloud Storage changes, among others).



- 01 Triggered directly or by cloud events
- 02 Cost-effective (to a point) and scales to zero.
- 03 Runs code with no servers or containers to manage
- 04 Good for writing and running small code snippets that respond to events
- 05 Use cases including ML solutions such as media translation or image recognition for files uploaded in GCS, Integration with third-party services and APIs etc.



Insightful facts

- ✓ If your code is a function and just performs an action based on an event/trigger, then deploying it with Cloud Functions makes sense.
- ✓ Pricing is based on number of events, compute time, memory and ingress/egress requests and costs nothing if the function is idle.

Discounts & Saving Options in GCP



Below are the options that Organization can explore to save some Dollars when using Computing powers in GCP,

- 1 Sustained Use Discounts
- 2 Committed Use Discounts
- 3 Discounts sharing
- 4 Spot & Preemptible Disks

Sustained Use Discounts

%.OFF



Compute Engine offers sustained use discounts on resources that are used for more than 25% of a billing month and are not receiving any other discounts. Whenever you use an applicable resource for more than a fourth of a billing month, you automatically receive a discount for every incremental hour that you continue to use that resource.

The discount increases incrementally with usage and you can get up to a 30% net discount off of the resource cost for instances that run the entire month.



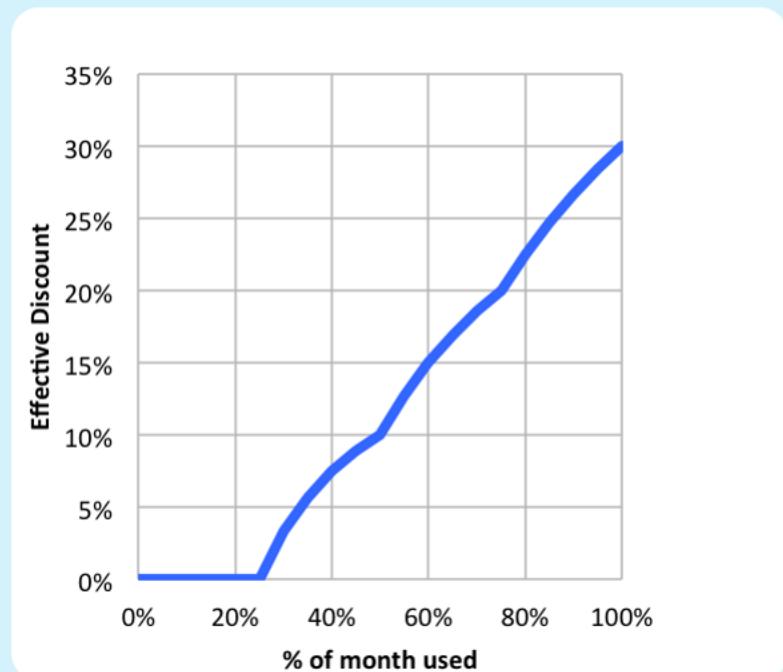
Sustained use discounts are calculated and applied automatically to usage within a billing account, so there is no action required on your part to enable these discounts.

Limitations



- Sustained use discounts automatically apply to VMs created by both Google Kubernetes Engine and Compute Engine.
- Sustained use discounts do not apply to VMs created using the App Engine environments and Dataflow.
- Sustained use discounts do not apply to E2, A2, Tau T2D, or Tau T2A machine types.

The graph shows how your effective discount increases with use, up to 30%



Committed Use Discounts

%.OFF



Google Cloud offers committed use discounts (CUDs) in return for **purchasing or commitment**. You commit either to a minimum amount of resource usage or to a minimum spend amount for a specified term of **one or three years**.

Ideal for **predictable** and **steady** usages



Hardware commitments: You can purchase hardware commitments for resources like vCPUs, memory, GPUs, local SSDs, and sole-tenant nodes. You get a discount of up to 70%

Software license commitments: You can purchase license commitments for applicable premium operating system (OS) licenses. You get a discount of up to 79%



Limitations

- Resource-based and flexible CUDs are available only for resources that are deployed using Compute Engine SKUs, which include VMs that are used by Compute Engine, Google Kubernetes Engine, Dataproc, or Cloud Composer 1.
- You can purchase Compute Engine flexible commitments only at a Cloud Billing account level.
- CANNOT cancel commitments unless we work with Cloud Billing Support team & get approvals



Discounts Sharing



You can purchase a committed use contract for a single project, or purchase multiple contracts which you can share across many projects by enabling **shared discounts**.



The ability to share Committed Use Discounts across projects on the same Cloud Billing account can be easily tracked using attribution in Google's cost management interfaces. To do this, Discount Sharing must first be **enabled**. Once enabled, Google offers two different attribution methods to track which projects are responsible for your usage and benefiting from CUDs. Without Discount Sharing, CUDs will live in projects and only apply to usage in the project the CUD was purchased in.

Proportional attribution (default): Applies discounts directly in proportion to the amount of total eligible usage consumed by each project.

Prioritized attribution: Applies discounts based on a distribution that you specify for specific projects.

Spot & Preemptible Disks

• % OFF



Spot & Preemptible VM instances are available at much lower price of 60-91% discount compared to the price of standard VMs. However, Compute Engine might stop (preempt) these instances if it needs to reclaim the compute capacity for allocation to other VMs.

If your apps are fault-tolerant and can withstand possible instance preemptions, then preemptible instances can reduce your Compute Engine costs significantly. For example, **batch processing** jobs can run on preemptible instances.



IMPORTANT

VM Instances get 30 second warning to save anything before the VM get reclaimed

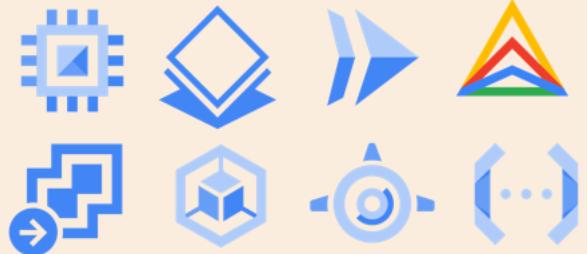
Spot VMs are the **latest version** of preemptible VMs. New and existing preemptible VMs continue to be supported, and preemptible VMs use the same pricing model as Spot VMs. However, Spot VMs provide new features that preemptible VMs do not support. For example, preemptible VMs can only run for up to 24 hours at a time, but Spot VMs do not have a maximum runtime unless you limit the runtime.



Limitations

- Not always available as they are finite Compute Engine resources
- Not covered by any Service Level Agreement and are excluded from the Compute Engine SLA.
- NO Automatic Restarts
- Free Tier credits not applicable

1



So you have understand all the GCP Compute options available and have a plan to Modernizing IT Infrastructure with Google Cloud. Does Modernizing IT Infrastructure alone is enough ? Ofcourse not...



2



To get complete benefits from Cloud, you should also modernize your apps by making them cloud native, adopting latest trends like microservices, containerization, CI/CD, DevOps etc.

Few Tips around Modernizing Applications



Challenges with legacy Apps

- ✓ Updating existing applications that have been built on-premises with a monolithic architecture can be difficult. When an application is updated, the entire application needs to be deployed and tested, even if the change is only small.
- ✓ New enhancements often have to be designed, built, tested, integrated, and deployed. But new needs often compete with existing projects for time and resources.



So what should I do & how do I modernize my Apps

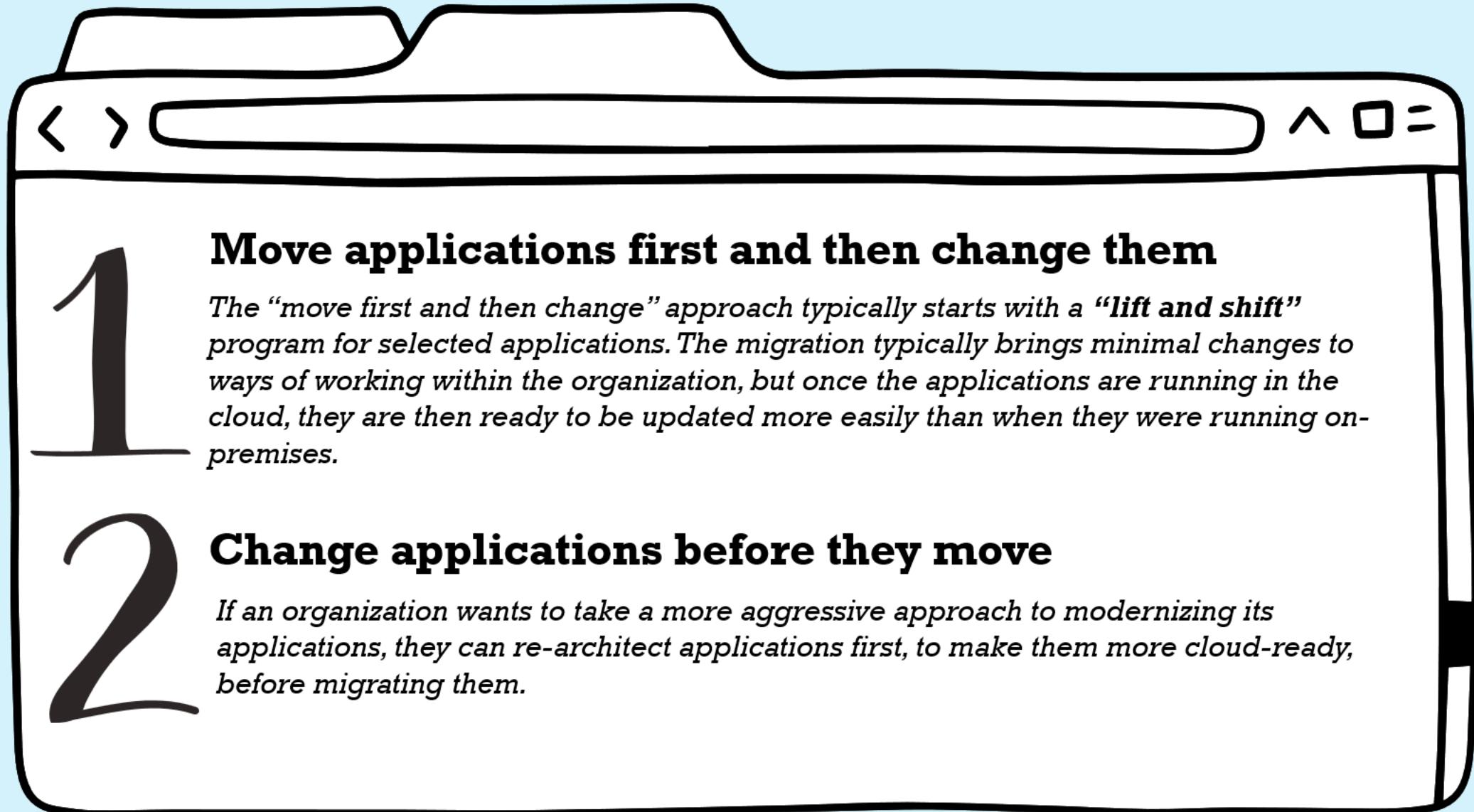
- ✓ Consider developing **cloud-native** applications & avoids the hassle of trying to create something that is constrained by legacy systems and out-dated processes. Building a new application in the cloud frees teams up from worrying about environments so that they can focus on creating features instead.
- ✓ Consider **microservice** architecture which reduces these problems by separating a large application into small, loosely coupled services. This means it's easy to determine where code needs to be changed and the service can be updated, deployed, and scaled independently.
- ✓ Consider **Containerization** & divide an application design into individual compartments. Parts of the code can be updated without affecting the whole application. This builds resilience, because one error doesn't impact the whole application.
- ✓ Consider **Kubernetes** which is an open-source container-orchestration system for automating computer application deployment, scaling, and management.
- ✓ Consider **Serverless** computing for application development. You write the code for the functions you want, and the cloud provider updates and adapts the containers or VMs as needed.
- ✓ Consider adopting an automated continuous integration or '**CI/CD**' that can help you increase your application release velocity and reliability. You can test and roll out changes incrementally. This approach enables you to lower the risk of regressions, debug issues quickly, and roll back to the last stable build if necessary-- all without interrupting service for your users.

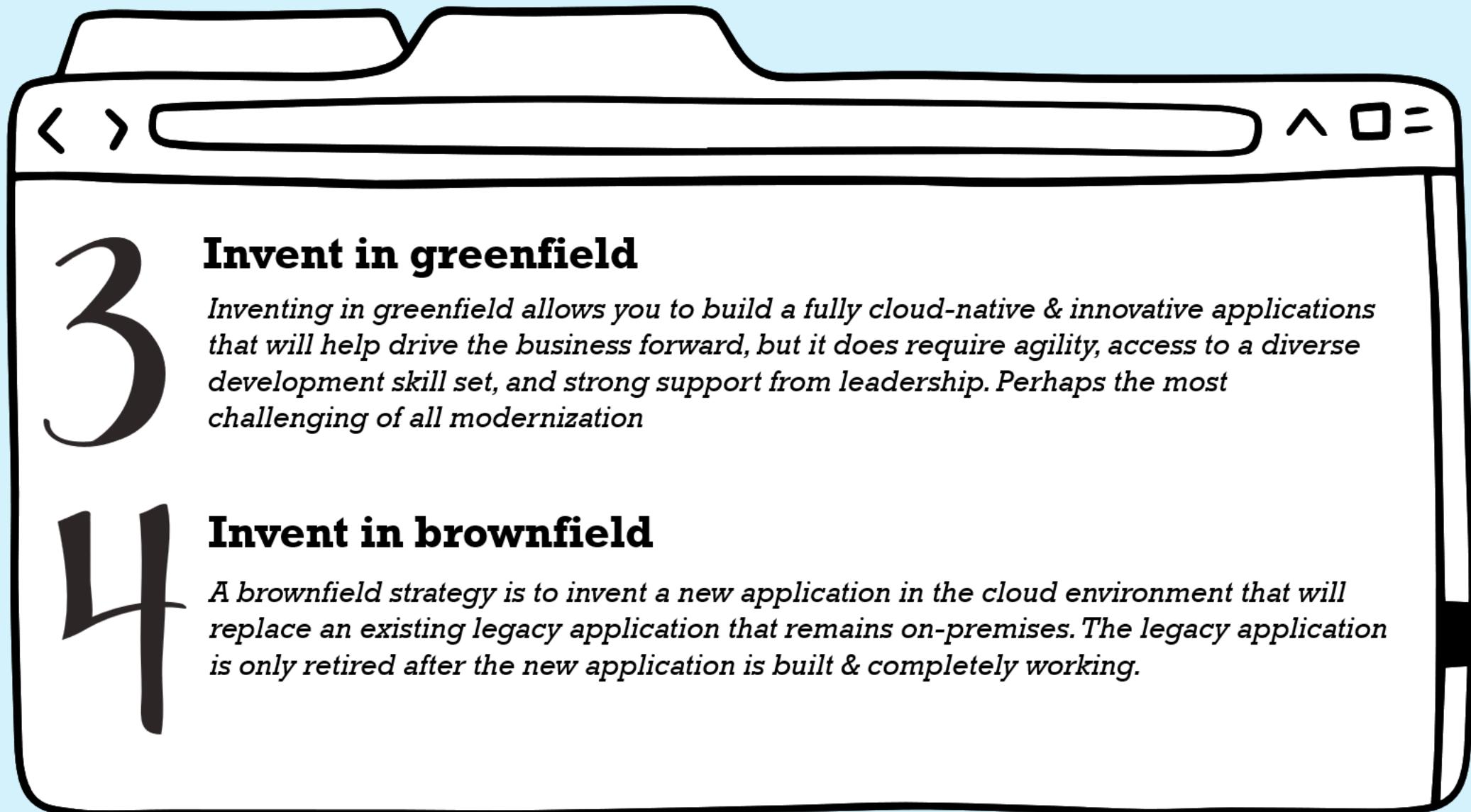
Modernizing Applications with Google Cloud

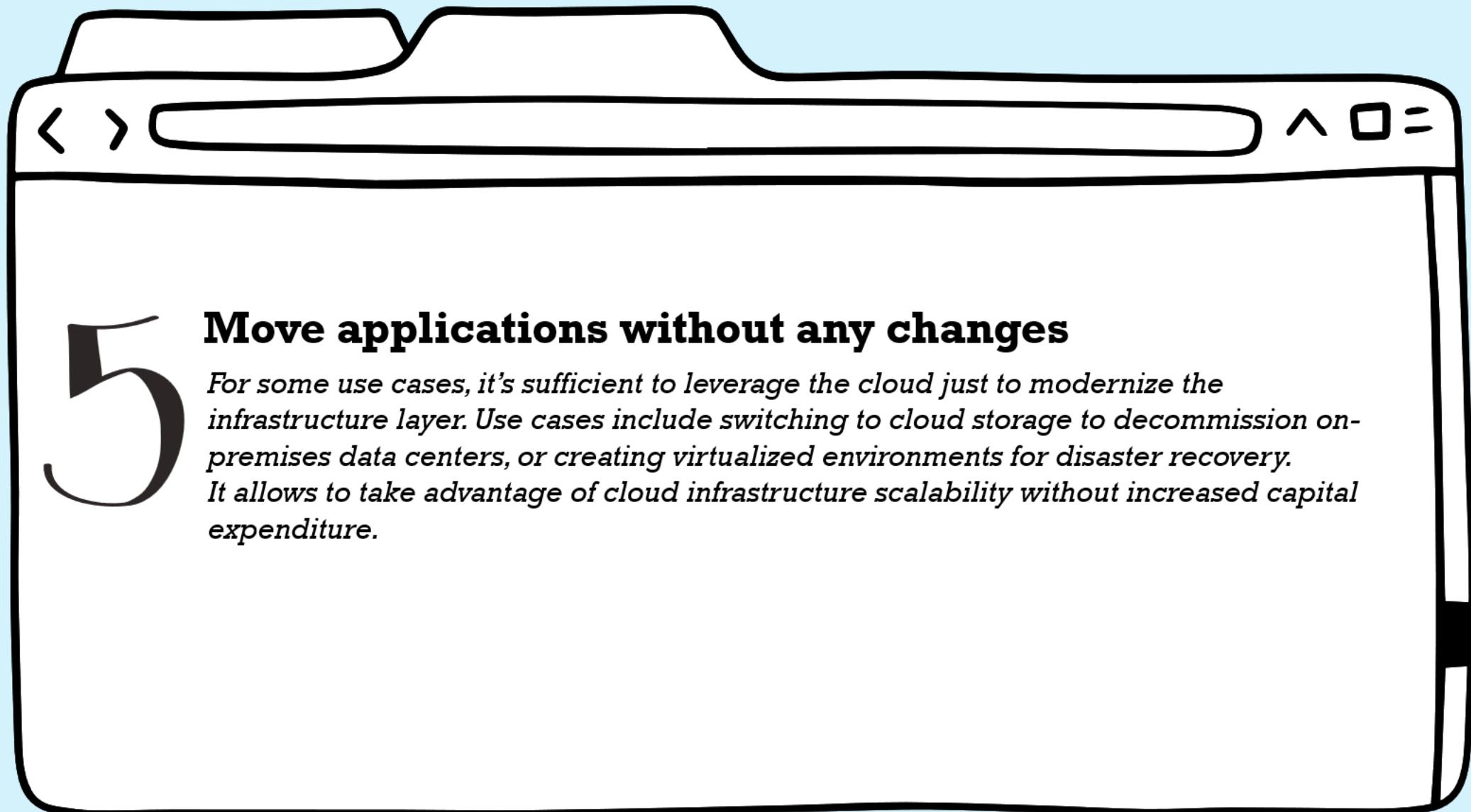
Google Cloud has identified & recommend five common patterns that businesses can adopt when they want to modernize their applications.

- 1 Move applications first and then change them
- 2 Change applications before they move
- 3 Invent in greenfield
- 4 Invent in brownfield
- 5 Move applications without any changes











What is a Legacy App ?

A legacy system is outdated computing software and/or hardware that is still in use, because it is mission critical. But often it is not equipped to deliver new services or upgrades at the speed and scale that users expect. It also often can't connect to newer systems.

01

NOT CLOUD FRIENDLY

Legacy Apps are not cloud friendly and it is impossible to migrate them to cloud & adopt any modern technologies due to its complexity, time consuming etc.

02

CANNOT MEET USER EXPECTATIONS

Legacy Apps doesn't provide the speed, scalability that are required to deliver seamless, digital experiences that consumers now expect in cloud era.

03

CAN'T SERVE REAL TIME DATA

Legacy systems are not designed to serve real-time data as is expected in today's digital world which is a huge limitation that Organizations face

04

LEGACY DATA IS IMPORTANT

Data that is locked in legacy systems limits the growth of the company, its partners, and its users.

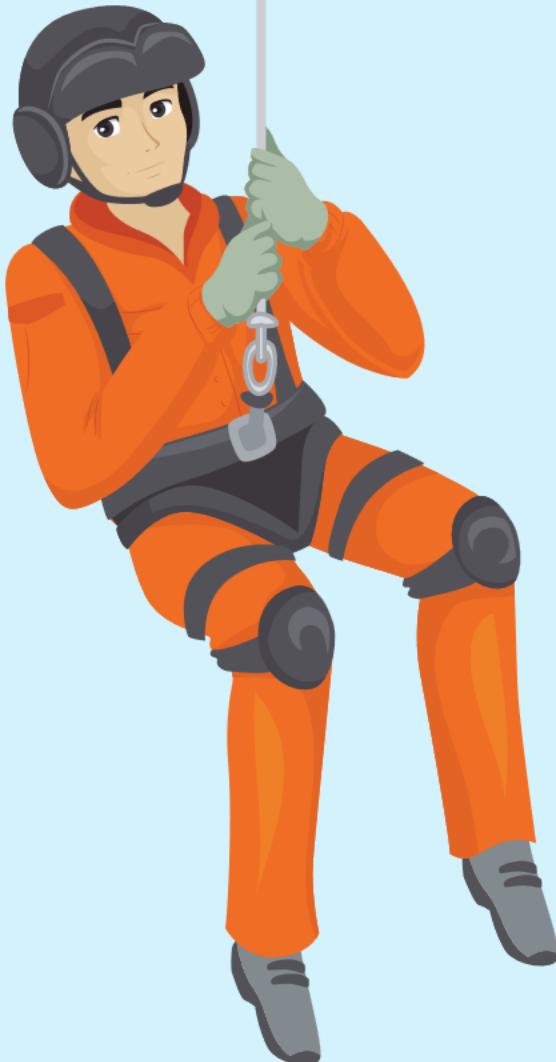
05

HOLDS INNOVATION

With all the limitations, legacy systems tend to hold organizations back from using digital technologies to innovate or improve IT efficiency.

Legacy Apps & APIs

APIs to the rescue

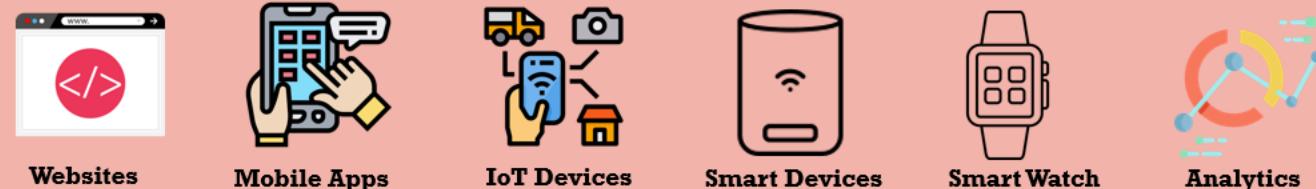


An **API (Application programming interface)** is a piece of software that connects different applications and enables information to flow between systems, so businesses can unlock value and create new services. Think of it as “middleware” that delivers a request from one server and returns the response to another.

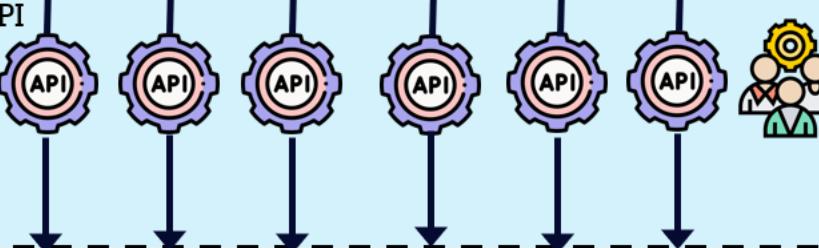
- 1 APIs can connect new Cloud based apps to legacy apps. Using the APIs, the data in the legacy system can be exposed to new aged apps like AI/ML.
- 2 APIs expose data in a way that protects the integrity of the legacy systems and enables secure and governed access to the underlying data.
- 3 Inside an Organization, there can be websites, mobile apps, IoT devices, third party apps, legacy apps, internal apps etc. All these apps can be connected with the help of APIs regardless of which technology, language are used to develop the apps.
- 4 With the 100s of APIs, a digital ecosystem will form inside a Organization. This includes vendors, third party suppliers, customers, and applications.
- 5 A robust, well connected, and multi faceted digital ecosystem enables businesses to create and monetize new digital experiences.

Legacy Apps & APIs

Modern Apps that are capable of handling new demands at faster pace



APIs are built and managed by the API teams within the enterprise. App developers leverage those APIs to integrate with backend services and other legacy service endpoints.



The required time and effort to manage 100s of APIs securely and at scale becomes costly. It requires a bigger team with good skills

Legacy Apps that holds lots of insightful data



Apigee

FOR APIs



Apigee is a fully featured API management platform that enables application developers and API providers to create connected digital experiences for end users. Apigee bridges the gap between legacy systems.

- 01 Apigee supports the entire API lifecycle and is capable of connecting legacy systems and cloud applications.
- 02 Integrates with Google Cloud services. Exposes services to broader use as well as data
- 03 API Analytics, tracks and monetizes developer usage. Includes a developer portal for testing and security
- 04 Hosts APIs on-prem, on Google Cloud, or on Apigee hybrid
- 05 Enables you to take advantage of the scaling, monitoring, and security features built into the Google Cloud Platform (GCP)



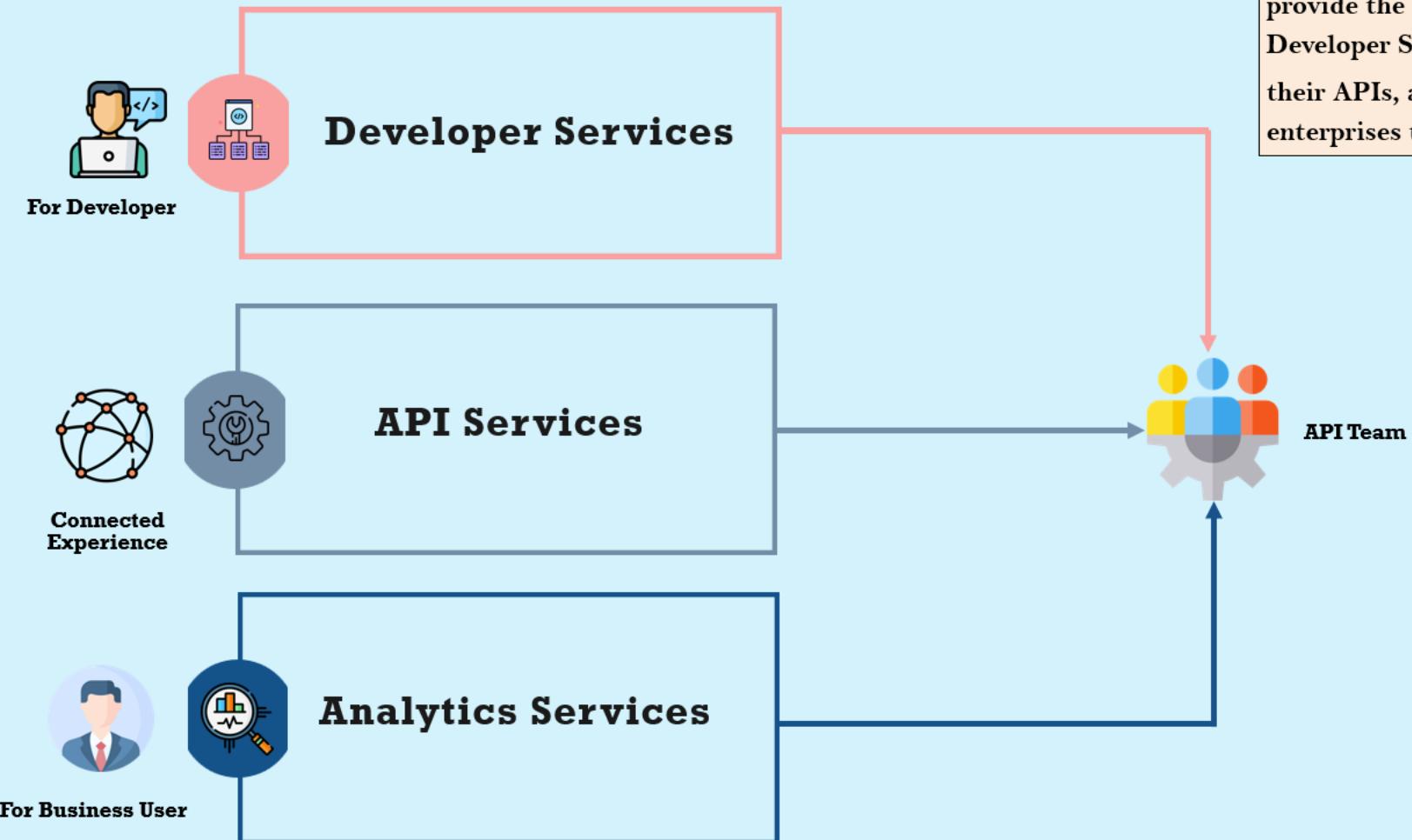
Insightful facts

Apigee API management helps **modernize** your applications and **monetize** your business channels. It helps **control** an application's access to the backend data and services. It also offers application developers with **tools** they need to access the API and helps API providers with **tools** they need to manage and provision the APIs.



Apigee Services

FOR APIs



The Apigee platform includes API services that provide the runtime API gateway functionality, Developer Services that allow developers to utilize their APIs, and Analytics Services that enable enterprises to report on APIs.

Google Cloud Digital Leader curriculum

Introduction to digital transformation with Google Cloud
(approximately **10%** of the exam)

Infrastructure and application modernization
(approximately **30%** of the exam)

Innovating with data and Google Cloud
(approximately **30%** of the exam)

Understanding Google Cloud security and operations
(approximately **30%** of the exam)

SOURCE: <https://cloud.google.com/certification/guides/cloud-digital-leader>

The Value of DATA

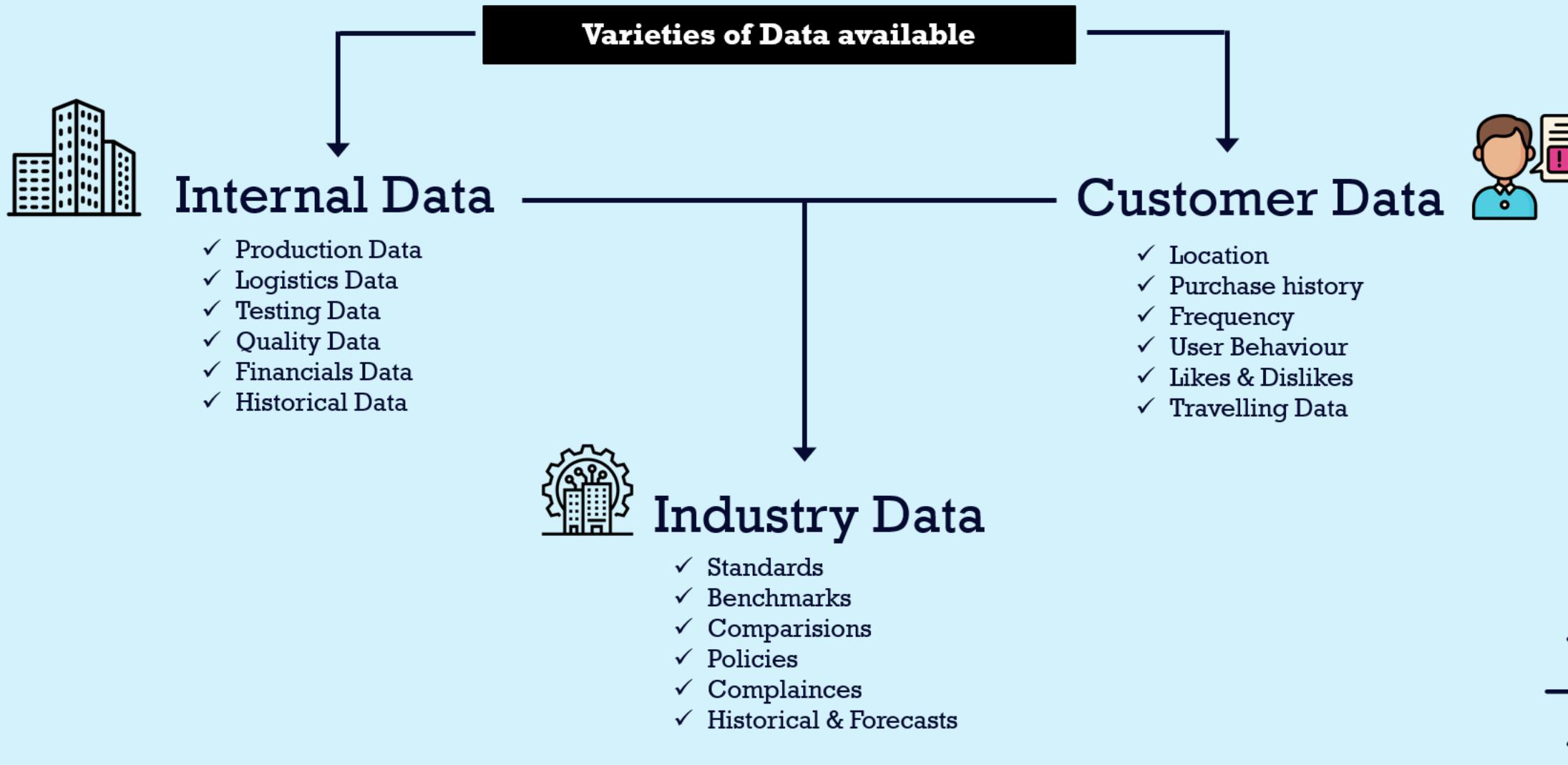
Data is any information that is useful to an organization. Examples include: documents, emails, audio files, video files, images, and data in DBs. Capturing and leveraging this data to unlock business value is central to digital transformation.

- 1 Businesses now have access to data like never before. This includes **internal information** (data from inside their organization) and **external information** (**customer** and **industry data**).
- 2 Large enterprises with traditional IT infrastructures face several **limitations** when it comes to leveraging the value of data.
- 3 Businesses that adopt Cloud, can now consume, store and process **terabytes of data in real-time**, and run queries—that is, requests to retrieve and use data, **instantly**.
- 4 With Cloud, Data can now be **distributed** across a global network. Multiple data centers create resilience against **data loss or service disruption**, without any extra overhead for businesses.
- 5 When coming to Data, GCP offers **Economies of scale**, **Rapid elasticity**, **Automation** & **global Data access**





Organizations can take advantage of the digital data from a variety of sources to take better decisions.



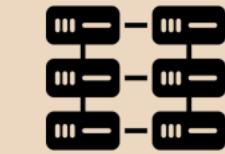
Limitations of handling Data with Traditional Infra

Scaling resources on demand

- both up and down is not possible.



Processing great amounts
of data, in varying formats is not
possible.



Storing, retrieving, and analyzing
existing data is not real time &
always limited by the resources
available.



Using AI/ML & making predictions
based on historical & incoming data
is not possible

Types of Data

STRUCTURE DATA

Structured data is highly organized and maintained in a consistent, predefined format. It can be easily stored & managed in databases. Examples include,

- ✓ Customers Data
- ✓ Transactional Data
- ✓ Credit Card Data
- ✓ Employees Data etc.

Stored in SQL databases, excel, data warehouses. Quantitative & searchable data.



SEMI STRUCTURED DATA

Semi-structured data has some consistent characteristic or organizational properties, but is not as rigid as structured data. Examples include,

- ✓ Json Data
- ✓ XML Data
- ✓ Emails

Moderately simple to analyze. Stored in NoSQL databases, data lakes.



UNSTRUCTURED DATA

Unstructured data has no consistency, organization and usually stored in BLOB* format. It stored in its own native format without a specific data model.

- ✓ Documents, Comments
- ✓ Images
- ✓ Audio files
- ✓ Videos

Stored in datalakes, applications. Qualitative & difficult to analyze or search



*BLOB - Binary large object

Handling Data with Cloud Infra

1

Organizations rely on structured, semi-structured, and unstructured data to gain insight and make intelligent decisions. However, unstructured data has historically been very difficult to analyze. But no more with Cloud.

2

With the right cloud tools, businesses can extract value from unstructured data by using APIs to create structure. APIs are a set of functions that integrate different platforms, with different types of data, so that new insights can be uncovered.

3

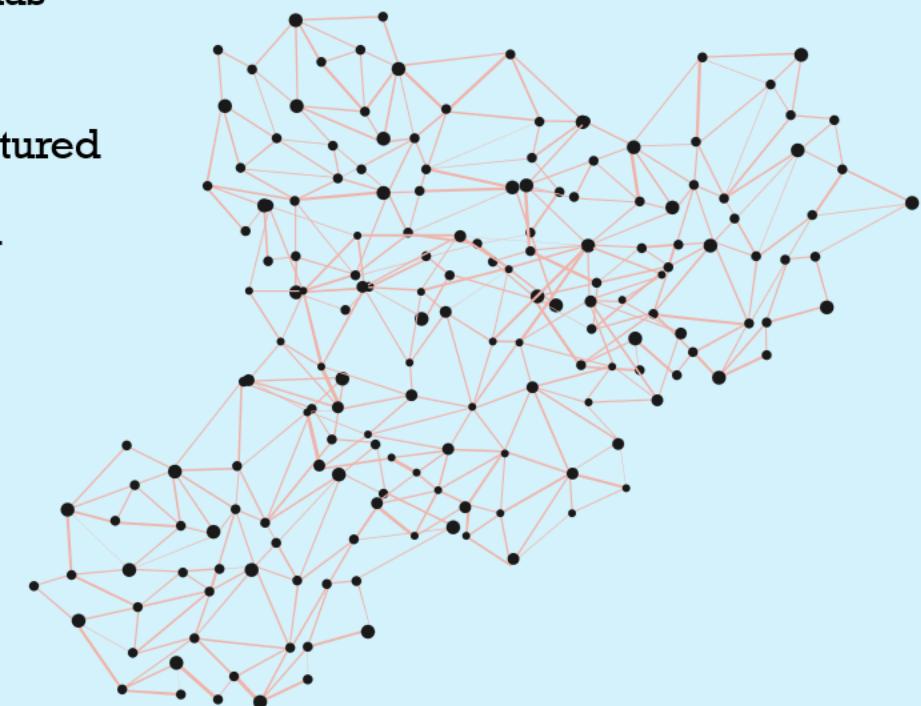
Modern processing power makes it possible to analyze and categorize unstructured data.

4

Cloud offers unlimited storage and can be scaled in and out at real time based on the need.

5

GCP offers various products to store the data based on the type. This includes relational DBs, NoSQL DBs, Datawarehouses, Data lakes, analytical tools.



Categorizing GCP Data Services

SQL Databases



Cloud SQL



Cloud
Spanner

NoSQL Databases



Cloud
BigTable



FireStore

Data Warehouse



Big Query

Data Lake



Cloud
Storage

BI & Analytics



Looker

What is a Database ?

- 1 Database is an organized collection of data, generally stored in tables, columns and accessed electronically from a computer system
- 2 Databases offer scalability & Data integrity. Allow businesses to rollback transactions to see data history.
- 3 Data integrity, or transactional integrity, refers to the accuracy and consistency of data stored in a database. Data integrity is achieved by implementing a set of rules when a database is first designed and through ongoing error checking and validation routines as data is collected.
- 4 Databases are generally used to store the transactional data of an application like Order Details, Customer Details, Invoice Details etc.
- 5 Maintaining Databases in a traditional infrastructure is a pain process where a DBA need to monitor & take care of the endless maintenance required to run a database including setting up servers, applying patches and updates, configuring replication and managing backups

PERSON ID	NAME	EMAIL
456	John	j@gmail.com
457	Peter	p@gmail.com
458	Anna	a@gmail.com

Cloud SQL

Managed SQL DB

 Cloud SQL is a fully managed relational database for **MySQL**, **PostgreSQL** and **SQL Server**. It reduces maintenance cost and automates database provisioning, storage capacity management, backups and out of the box high availability and Disaster recovery/failover.

- 01 Easily integrates with existing applications and Google Cloud services like GKE, GCE, App Engine, BigQuery etc.
- 02 Offers security, high availability and durability.
Automatically scale storage up to 30 TB.
- 03 **Database Migration Service(DMS)** simplifies migration of MySQL and PostgreSQL databases from on-premises, Google Compute Engine, and other clouds to Cloud SQL.
- 04 Automatic replication with automatic failover, backup, and point-in-time recovery
- 05 The data in CloudSQL is automatically encrypted at rest and in transit. External connections can be enforced to be SSL-only.

Insightful facts

- ✓ Supports 99.95% SLA. Google Cloud continuously monitors the Cloud SQL instance with a heartbeat signal and when a master fails an automatic failover is triggered to another zone in your selected region in case of an outage. You can also create replicas across regions to protect from regional failure.
- ✓ Cloud SQL Insights, a free tool, helps detect, diagnose, and identify problems in a query for Cloud SQL databases. It provides self-service, intuitive monitoring, and diagnostic information that goes beyond detection to help you to identify the root cause of performance problems.



Cloud Spanner

Managed SQL DB

Cloud Spanner is another fully managed database service, and it's designed for **global scale**. Data is automatically and instantly copied across regions. This replication means that if one region goes offline, the organization's data can still be retrieved from another region.

- 01 Provides unlimited scale, global consistency, and up to 99.999% availability
- 02 Processes up to 1 billion requests per second.
- 03 Ideal for gaming, global financial, ecommerce apps etc.
- 04 Support for two SQL dialects: Google Standard SQL and PostgreSQL.
- 05 Offers fully managed, mission-critical, relational database service that offers transactional consistency at global scale, automatic, synchronous replication for high availability

Insightful facts

Cloud Spanner is the only enterprise-grade, globally-distributed, and strongly-consistent database service built for the cloud, specifically to combine the benefits of relational database structure with non-relational horizontal scale. It is a unique database that combines transactions, SQL queries, and relational structure with the scalability that you typically associate with non-relational or NoSQL databases.

What is a NoSQL Database ?

1

NoSQL databases (aka "not only SQL") are non-tabular databases and store data differently than relational tables.

2

They come in a variety of types based on their data model. The main types are document, key-value, wide-column, and graph.

3

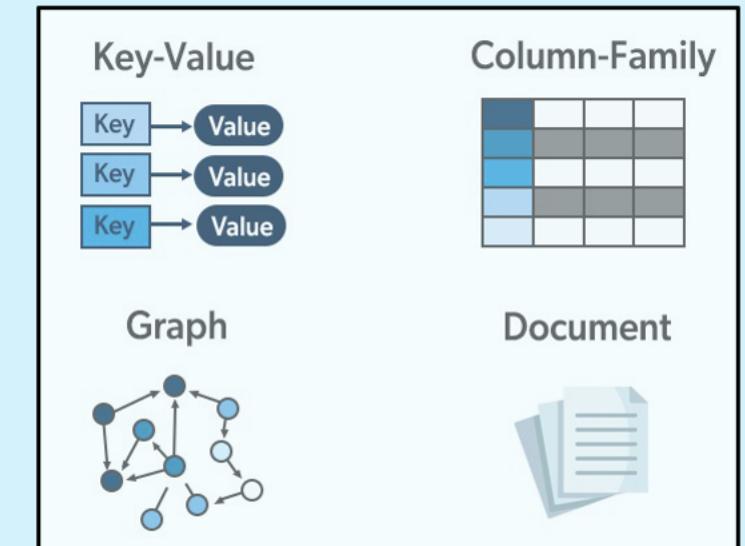
They provide flexible schemas and scale easily with large amounts of data and high user loads.

4

Supports storage of structured and semi-structured data

5

A common misconception is that NoSQL databases or non-relational databases don't store relationship data well. NoSQL databases can store relationship data — they just store it differently than relational databases do.



Cloud BigTable

Managed NoSQL DB



Cloud Bigtable is a fully managed wide-column NoSQL database that scales to petabyte-scale. It's optimized for low latency, large numbers of reads and writes, and maintaining performance at scale. It offers really low latency of the order of single-digit milliseconds.



- 01 Handles large amounts of data in a key-value store and supports high read and write at low latency
- 02 Linearly scalable. Throughput can be adjusted by adding/removing nodes. Each node provides up to 10,000 operations per second.
- 03 It's the same database that powers many core Google services, including Search, Analytics, Maps, and Gmail.
- 04 Can scale to billions of rows and thousands of columns, enabling you to store terabytes or even petabytes of data.
- 05 Provides real time analytics for huge workloads which are common in Gaming, IoT, AdTech, FinTech and ML based personalizations



Insightful facts

Cloud BigTable is an ideal data source for time series and MapReduce-style operations. Bigtable supports the open-source HBase API standard to easily integrate with the Apache ecosystem including HBase, Beam, Hadoop and Spark. It also integrates with Google Cloud ecosystem including Memorystore, BigQuery, Dataproc, Dataflow and more.

FireStore

Managed Serverless
NoSQL DB



Firestore is a serverless, fully managed NoSQL document database that scales from zero to global scale without configuration or downtime.



- 01 Accelerate development of mobile, web, and IoT apps with direct connectivity to the database
- 02 Supports effortless real time data synchronization with changes in your database as they happen
- 03 Robust support for offline mode, so your users can keep interacting with your app even when the internet isn't available or is unreliable.
- 04 Integration with Firebase and Google Cloud services like Cloud Functions and BigQuery, serverless data warehouse.
- 05 Use Firestore backend-as-a-service, to build a feature or an application quickly

Insightful facts

Possible uses:

- ✓ Live asset and activity tracking
- ✓ Real-time analytics
- ✓ Media and product catalogs
- ✓ Social user profiles and gaming leaderboards



Categorizing GCP Data Services

SQL Databases



Cloud SQL



Cloud
Spanner

NoSQL Databases



Cloud
BigTable



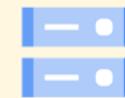
FireStore

Data Warehouse



Big Query

Data Lake



Cloud
Storage

BI & Analytics



Looker

Different ways to store Data

Database	Data warehouse	Data lake
✓ A database stores the current data required to power an application.	✓ A data warehouse stores current and historical data from one or more systems in a predefined and fixed schema, which allows business analysts and data scientists to easily analyze the data.	✓ A data lake stores current and historical data from one or more systems in its raw form, which allows data scientists to easily analyze the data. It tends to serve many purposes.
✓ Data available in real time		
✓ Used for operational and transactional		✓ May be available in real time based on the processes or setup
✓ Ex: Relational databases (Oracle, MySQL, Microsoft SQL Server & PostgreSQL)	✓ May not be up-to-date based on frequency of ETL(Extract, Transform, Load) processes	✓ Used for Analytics
Document databases (MongoDB & CouchDB)	✓ Used for Analytics	✓ Ex: Google Cloud Storage, AWS S3 etc.
	✓ Ex: Google BigQuery, Amazon Redshift, Snowflake	

BigQuery

Managed Serverless
Data warehouse



BigQuery is the Google Cloud enterprise data warehouse designed to help organizations to run large scale analytics with ease and quickly unlock actionable insights. As a fully-managed data warehouse, BigQuery takes care of the infrastructure so you can focus on analyzing your data up to petabyte-scale.

- 01 It is a serverless, multi-regional, multicloud, SQL column-store data warehouse
- 02 Scales to handle terabytes of data in seconds and petabytes in minutes
- 03 Built-in integration for machine learning and backbone for Business Intelligence (BI) Engine
- 04 Supports real-time analytics with streams from Pub/Sub, Dataflow, and Datastream
- 05 You can ingest data into BigQuery either through batch uploading or by streaming data directly to unlock real-time insights. BigQuery supports SQL, which you're likely already familiar with.



Insightful facts

- ✓ BigQuery is **serverless**. It means that resources, such as compute power, are automatically provisioned behind the scenes as needed to run your queries. So businesses do not pay for compute power unless they are actually running a query.
- ✓ Pub/Sub and DataFlow can work together to bring unstructured data into the cloud and transform it into semi-structured data. This transformed data can then be sent directly from Dataflow to BigQuery, where it is made immediately available for analysis.



Cloud Storage

Cloud Storage is a global, secure, and scalable object store for immutable data such as images, text, videos, and other file formats. You can add data to it or retrieve data from it as often as your application needs. It also offers automatic object versioning, so you can restore older versions of objects—which can be especially helpful as protection against accidental deletion.

- • 01 Often used to hold 'back-up' data, which helps businesses build resilience against unexpected harm affecting their data.
- • 02 Also used to hold data that is historic and not relevant to day-to-day business operations.
- • 03 Offers multi-regional storage. It's ideal for serving content to users worldwide with low latency. Supports to store any amount of data
- • 04 Offers four storage classes that can be choosed based on budget, availability, and access frequency.
- • 05 Lets you store your data in three types of locations: Regional, Multi-region & Dual-regions.

Data Lake for Unstructured Data

Insightful facts

- ✓ By default 100% of data in Cloud Storage is automatically encrypted at rest and in transit with no configuration required by customers. You can grant permission to specific members and teams or make the objects fully public for use cases such as websites.
- ✓ If you want more direct control over encryption you have two additional key management options available to you that go beyond the built-in encryption that Google manages for you:
 - *Customer-managed encryption keys (CMEK)*
 - *Customer-supplied encryption keys (CSEK)*

Choosing a location for Cloud Storage

Regional

All of your data is stored redundantly in a single region. Regional buckets usually offer the lowest monthly storage price and are suitable for a wide range of use cases, including high-performance analytics where it is important to co-locate your compute and storage in the same region.

Multi-region

All of your data is stored redundantly across a continent but it's not visible which specific regions your data is in. Availability is higher than regional because your data can be served from more than one region. Multi-regions cost a little more than single regions, but are great choices for content serving to the Internet.

Dual-regions

All of your data is stored in two specific regions. Dual-regions provide the best of regions and multi-regions — providing you with high availability and protection against regional failures while also giving you the high-performance characteristics of regional storage. Business-critical workloads are often best implemented on top of dual-regions.



No matter the location that you select, all four storage classes are available to you so that you can optimize your costs over time, storing your most active “hot” data in Standard and moving it down to colder classes as it becomes older and less frequently accessed.

Cloud Storage Classes

Standard

For high-performance, frequent access, and highest availability

Nearline

Ideal for data you plan to read or modify on average once per month or less

Coldline

Best for objects or data you plan to access at most once every 90 days

Archive

Least expensive choice for objects that you plan to access less than once a year



It will costs a bit more to use standard storage because it is designed for short-lived and/or frequently accessed data. Nearline, coldline, and archive storage offer a lower monthly storage cost for longer-lived and less frequently accessed data.

Security of Data in Cloud Storage



By default 100% of data in Cloud Storage is automatically encrypted at rest and in transit with no configuration required by customers. You can grant permission to specific members and teams or make the objects fully public for use cases such as websites.

If you want more direct control over encryption you have two additional key management options available to you that go beyond the built-in encryption that Google manages for you:

Customer-managed encryption keys (CMEK)

You can use CMEK via Google Cloud Key Management Service (KMS). You can define access controls to encryption keys, establish rotation policies, and gather additional logging into encryption/decryption activities.

In both the default and customer-managed case, Google remains the root of trust for encryption/decryption activities.

Customer-supplied encryption keys (CSEK)

You can use CSEK in which Google is no longer in the root of trust. Using CSEK comes with some additional risk of data loss, as Google cannot help you decrypt data if you lose your encryption keys.

Great Data Power comes with Great Data Responsibility



Handling volumes and diversity of data comes with its own ethical considerations and requires alternative ways of thinking about security. Not all information that can be captured, should be captured. Businesses are accountable for making responsible decisions about which data they collect, store, and analyze.

Ethics

Just because you can gather all the data, doesn't mean you should. Businesses must be responsible. Consider the source of the data, how it is being collected and where it's stored.

Secure Privacy

If it's personal or sensitive data (PII) about an individual, it needs to be securely collected, encrypted when stored in the cloud, and protected from external threats.

Follow Regulations

All the regulations of industry, country, region must be followed when handling the data.

Ex: HIPPA, PCI DSS, CCPA

AI & ML

Ethical and fair considerations are particularly important and applicable when you work with Artificial Intelligence (AI) and Machine Learning.



Human bias can influence the way datasets are collected, combined, and used. It's always important to include strategies to remove unconscious biases as you start to leverage data to build new business value.

Categorizing GCP Data Services

SQL Databases



Cloud SQL



Cloud
Spanner

NoSQL Databases



Cloud
BigTable



FireStore

Data Warehouse



Big Query

Data Lake



Cloud
Storage

BI & Analytics



Looker

Looker



Looker is a Google Cloud business intelligence solution. It's a data platform that sits on top of any analytics database and makes it simple to describe your data and define business metrics.



- 01 A business intelligence software and big data analytics platform. Includes integrated built-in and custom data visualizations and dashboards
- 02 Helps explore, analyze, and share realtime business analytics. Access, analyze, and act on the up-to-date, trusted version of your data
- 03 Integrates with BigQuery, BigTable, CloudStorage, other cloud platforms, Google Drive etc.
- 04 Build data experiences that empower users and reduce reliance on your data teams by up to 99%
- 05 Improve productivity, decision-making, and innovation by delivering more insights to more users with embedded analytics

Business Intelligence & Analytics

Insightful facts

Sometimes data analysis tools is not accessible by anyone outside the data engineering or data analysis teams & some tools let everyone in the business perform their own data analysis. But they can only perform their analysis with portions of the available data. This means that only a few people, or possibly no one, has a full view of the company's business data. Where as with **Looker**, once you have a reliable source of truth for your business data, anyone on your team can analyse and explore it, ask and answer their own questions, create visualisations, and explore row level details.



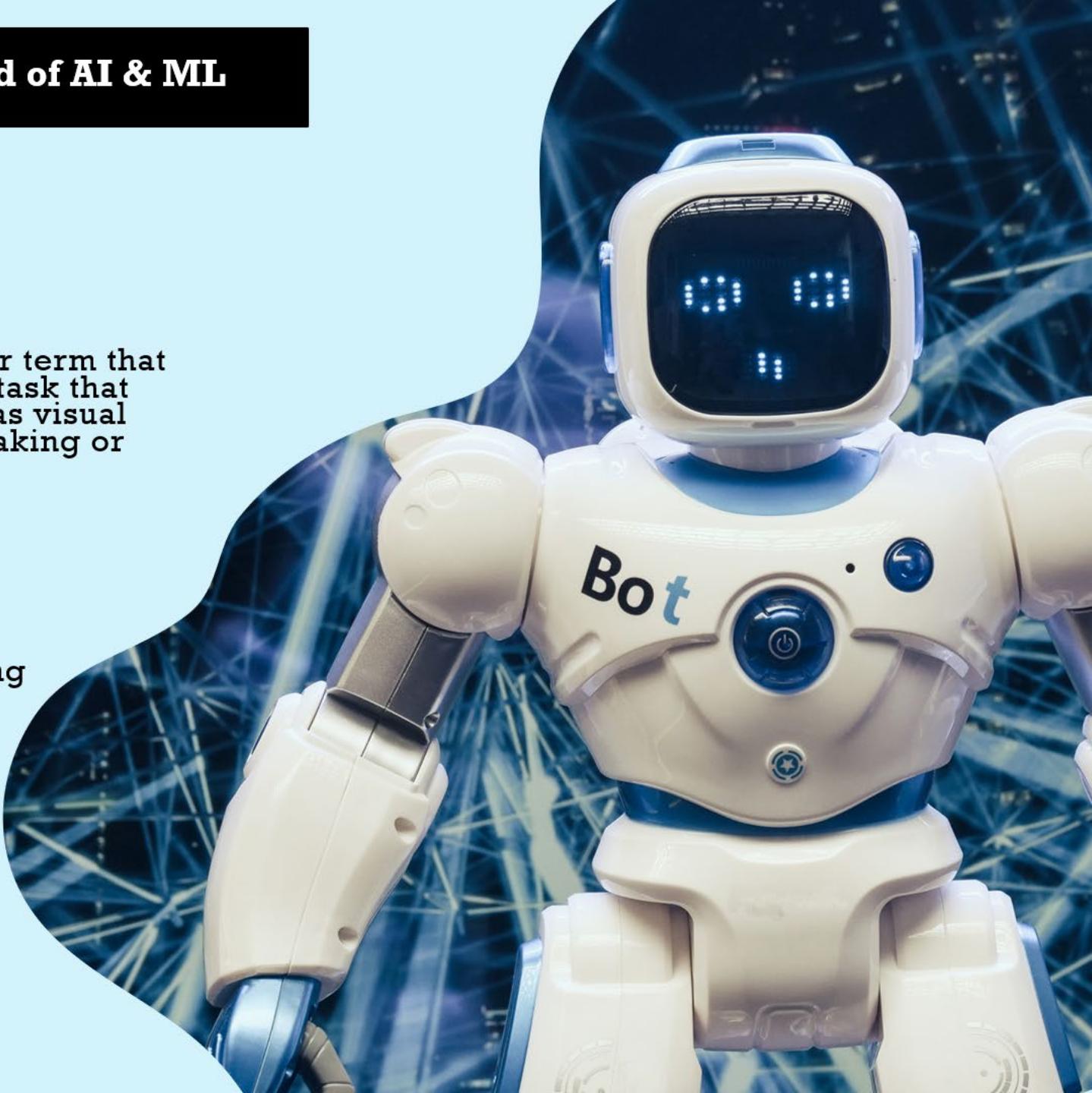
Welcome to the world of AI & ML

★ What is Artificial Intelligence ?

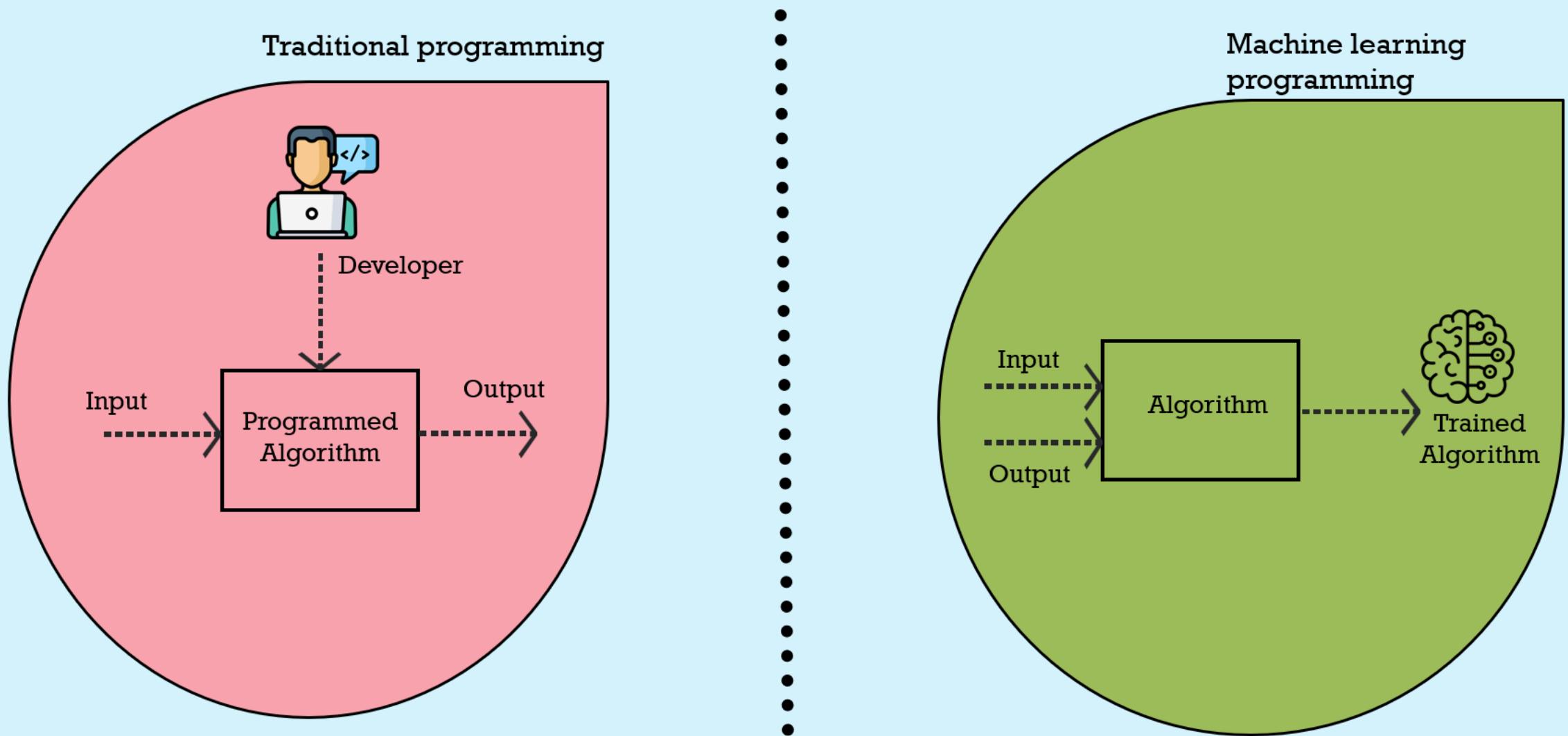
Artificial intelligence or AI is a broad field or term that describes any kind of machine capable of a task that normally requires human intelligence, such as visual perception, speech recognition, decision making or translation between languages.

★ What is Machine Learning ?

Machine learning (ML) is a branch of AI that enables computers to “self-learn” from training data and improve over time, without being explicitly programmed



Traditional vs ML programming

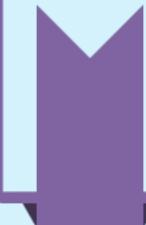


Benefits of AI & ML



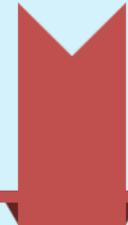
Replacing rule-based apps

With AI & ML, we can teach machines on how to react or analyze with any kind of scenarios which is not the case with traditional programmes



Automating processes

AI & ML is being used in many industries to automate the process of manufacturing, processing & analysis etc.



Understands unstructured data

When trained properly, AI & ML models can understand & process any kind of Unstructured data which is impossible with traditional programmes



Personalized Customer experience

With AI & ML, we can analyze the user behaviour, history, metrics and provide a personalized behaviour inside the applications



Predictions Analysis

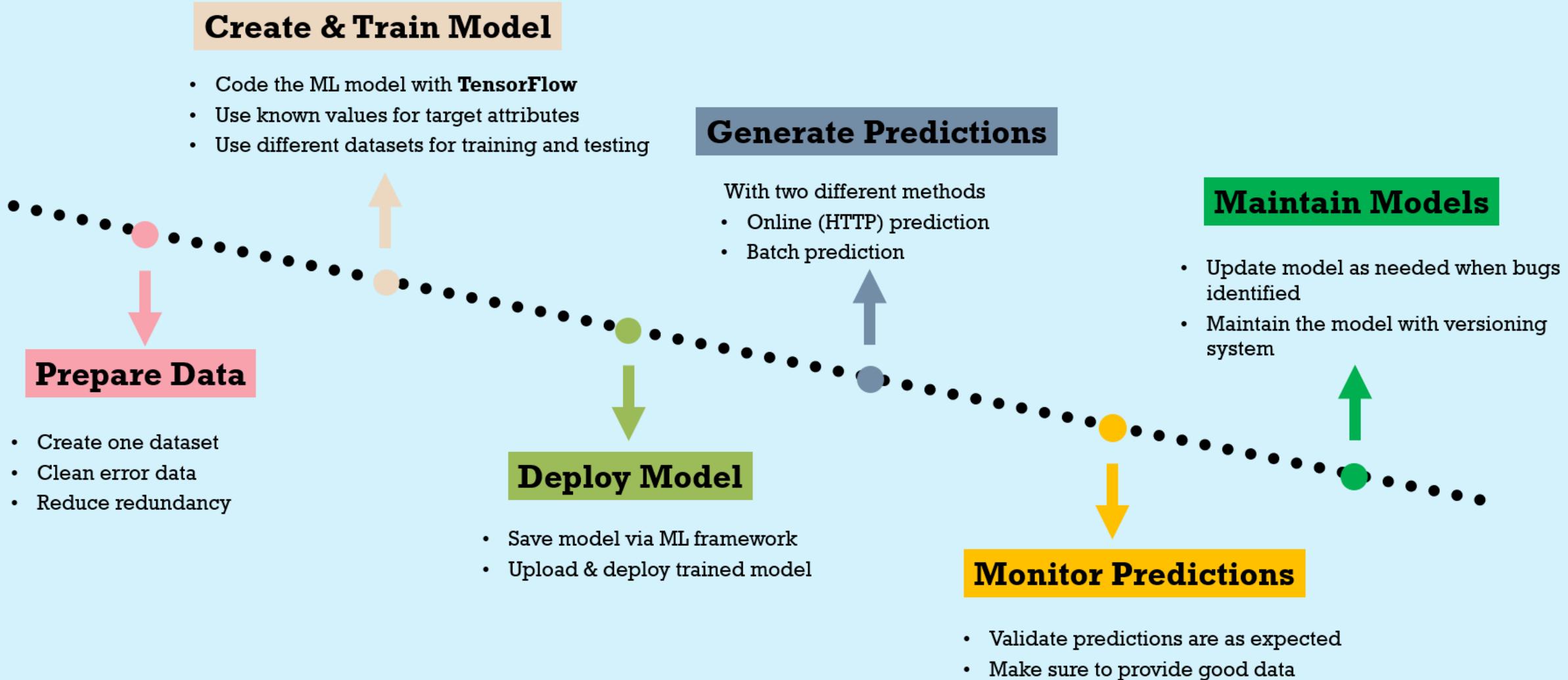
AI & ML can be used for Demand forecasting, fraud detections, credit scoring, Medical imaging, Trade forecasting etc.



More Human time

AI & ML can assist humans in analyzing data, daily activities, Medical diagnosis which improves the productivity of humans.

HOW DOES ML WORKS ?

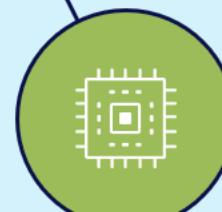


Challenges building ML models

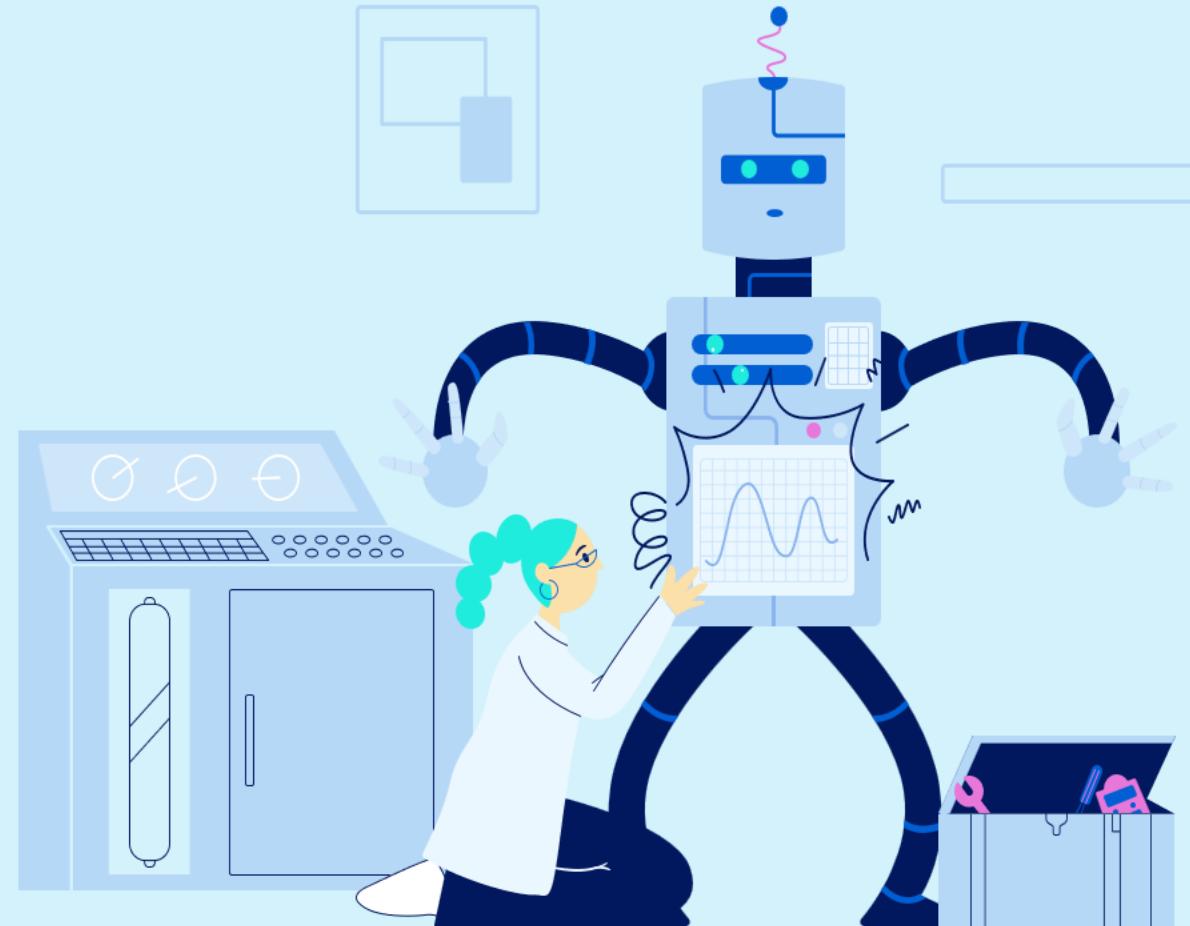
The accuracy of any ML predictions depends on large volumes of data that are free of bugs. So cleaning data is an essential step but time-consuming phase of building an ML model, and involves removing irrelevant, duplicate data as well as outliers and anomalies.



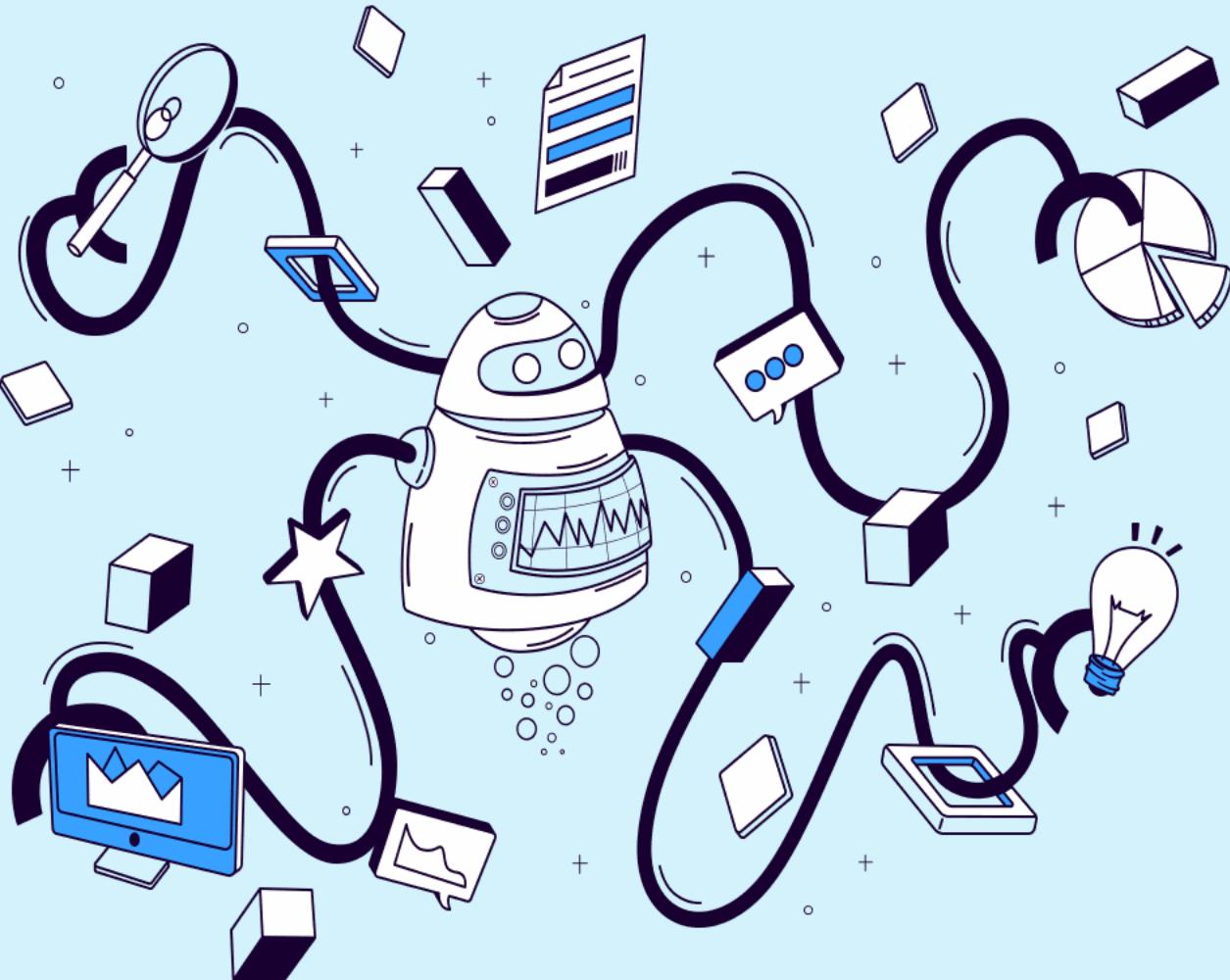
To avoid bugs in ML models, Engineers put lot of efforts in **data cleanliness/consistency & data completeness/coverage**. Often a single Organization alone will not have enough data to train an AI model. So building the data & training AI model will take years with traditional approaches.



Need to use hundreds of data points correctly labeled to train an ML model with a percentage earmarked for testing; retrain the model with additional data as needed.



Intro to GCP AI Ecosystem



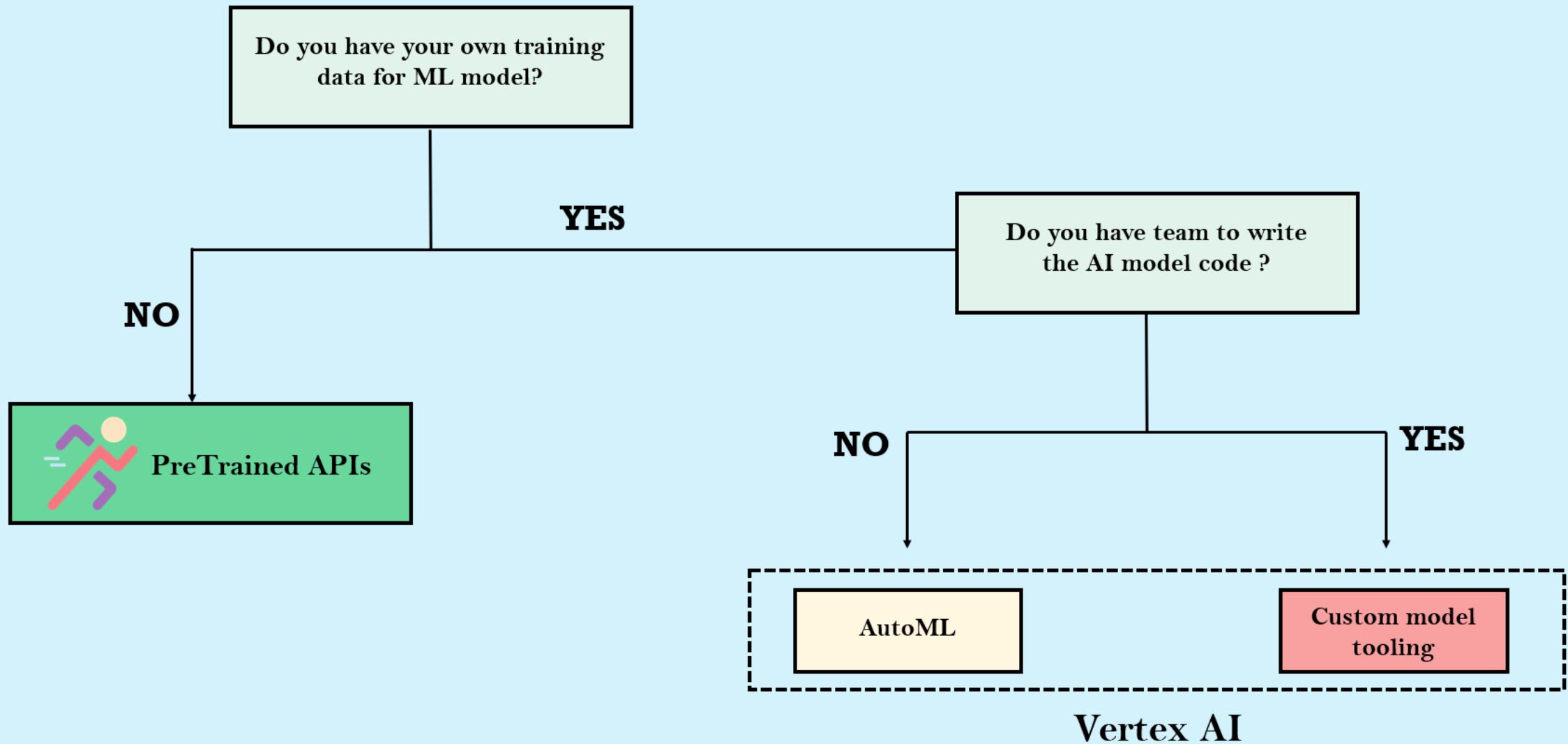
Google Cloud democratizes AI by providing a range of ML and AI solutions that enable businesses to leverage the power of ML and AI, without the traditional costs and efforts.

Google Cloud AI Platform is a unified, simply managed platform that makes ML easy to adopt. It provides modern ML services, with the ability to generate tailored models and use pre-trained models.

TensorFlow takes advantage of Tensor Processing Units (TPU), hardware devices designed to accelerate ML workloads with TensorFlow by 15-30x. Because you pay only for what you use, there's no up-front capital investment required.

The **AI Hub** is a hosted repository of plug-and-play AI components, including end-to-end AI pipelines and out-of-the-box algorithms.

Intro to GCP AI Ecosystem



Vertex AI

Vertex AI brings together the Google Cloud services for building ML under one, unified UI and API. In Vertex AI, you can now easily train and compare models using AutoML or custom code training. Whichever option you choose for training, you can save models, deploy models, and request predictions with Vertex AI.

- 01 Build, deploy, and scale machine learning (ML) models faster, with fully managed ML tools for any use case.
- 02 Implement **MLOps** practices to efficiently scale, manage, monitor, and govern your ML workloads
- 03 Reduce training time and costs with optimized infrastructure
- 04 All your models are stored in one central model repository. Deploy model to URL for predictions
- 05 Data scientists can move faster with purpose-built tools for training, tuning, and deploying ML models.

Unified AI + ML

Insightful facts

You can use Vertex AI to manage the following stages in the ML workflow:

- ✓ Create a dataset and upload data.
- ✓ Train an ML model on your data:
- ✓ Upload and store your model in Vertex AI.
- ✓ Deploy your trained model to an endpoint for serving predictions.
- ✓ Send prediction requests to your endpoint.
- ✓ Specify a prediction traffic split in your endpoint.
- ✓ Manage your models and endpoints.

AutoML



AutoML enables developers with limited machine learning expertise to train high-quality models specific to their business needs. Build your own custom machine learning model in minutes.

01 AutoML Tables

02 AutoML Image

03 AutoML Video

04 AutoML Text

05 AutoML Translation

For Machine Learning



Insightful facts

Data types supported include:

- ✓ Text data
- ✓ Tabular data
- ✓ Image data
- ✓ Video data

AutoML Products

AutoML Tabular

Handles wide range of tabular data primitives

Ingest structured data to predict a value in the target column

AutoML Image

Use REST and RPC APIs

Detect objects, where they are, and how many

Classify images using custom labels

AutoML Video

Annotate video using custom labels

Streaming video analysis

Shot change detection

Object detection and tracking

AutoML Text

Integrated REST API

Reveal the meaning of text

Custom sentiment analysis

Large dataset support

AutoML Translation

Dynamically detect and translate between languages

Integrated REST and gRPC APIs

Supports 50 language pairs

Translate with custom models



Pre-trained models return results almost instantaneously

Speech-to-Text

Accurately convert speech into text with an API powered by the best of Google's AI research and technology.

Transcribe your content with accurate captions

Text-to-Speech

Convert text into natural-sounding speech using an API powered by the best of Google's AI technologies

Engage users with voice user interface in your devices and applications

Translation AI

Make your content and apps multilingual with fast, dynamic machine translation

Reach global markets through internationalization of your products

Natural Language AI

Get insightful text analysis with machine learning that extracts, analyzes, and stores text

Apply natural language understanding (NLU) to apps with Natural Language API

Video AI

Precise video analysis that recognize over 20,000 objects, places, and actions in video

Gain near real time insights with streaming video annotation and object-based event triggers

Media Translation

Add real-time audio translation directly to your content and applications

Media Translation API supports 12 languages.

Recommendations AI

Deliver highly personalized product recommendations at scale

Same AI that powers recommendations on Google Ads, Google Search, and YouTube.

Contact Center AI/ Dialogflow

Lifelike conversational AI with state-of-the-art virtual agents

One comprehensive development platform for chatbots and voicebots

Vision AI

Offers powerful pre-trained machine learning models. Assign labels to images and quickly classify them into millions of predefined categories. Detect objects, read printed and handwritten text, and build valuable metadata into your image catalog.

Document AI

Analyze documents like Resume, application letters etc.

Cloud talent solution which uses AI with job search and talent acquisition capabilities.

Google Cloud Digital Leader curriculum

Introduction to digital transformation with Google Cloud
(approximately **10%** of the exam)

Infrastructure and application modernization
(approximately **30%** of the exam)

Innovating with data and Google Cloud
(approximately **30%** of the exam)

Understanding Google Cloud security and operations
(approximately **30%** of the exam)

SOURCE: <https://cloud.google.com/certification/guides/cloud-digital-leader>

Financial Governance in the Cloud



Cloud technology can provide organizations with the means to make more dynamic decisions and accelerate innovation, but managing cloud costs requires **vigilance** and **real-time monitoring** in parallel.

Budgeting is no longer a **one-time operational process** completed annually. Because of the variable nature of cloud resources and their costs, spending must be **monitored and controlled on an ongoing basis**.

In a **small** organization, there might be **one person or team** responsible for managing all aspects of the cloud infrastructure and associated finances, from budgeting to procurement, tracking, optimization, and so on.

In **large** organizations, **Technology teams** are often using cloud resources, but they don't necessarily factor cost into their decision-making. **Finance teams** control cloud costs, but may struggle to understand or to keep up with cloud spend on a daily, weekly, or monthly basis.

Financial Governance in the Cloud



To solve this problem consider the solution through three lenses: **People, Process, Technology**



People

A **Cloud Center of Excellence** can be formed to manage cloud costs effectively. This group can be a partnership across **finance, technology, and business functions**. This partnership would consist of several experts who ensure that best practices are in place across the organization and there is visibility into the ongoing cloud spend.



Process

As part of the **Process** & cloud cost management, we need to consider the following **questions**:

- 1) What cloud resources are being used and by whom?
- 2) What are the associated resource costs?
- 3) How do these costs measure against the broader business strategy?



Technology

Google Cloud brings its own **native** tools to help organizations monitor and manage their costs. In fact, these tools enable organizations to gain greater visibility, drive a culture of accountability for cloud spending, control costs to reduce the risk of overspending, and provide intelligent recommendations to optimize costs and usage.

Financial Governance with the help of GCP tools

Visibility

Before organizations can optimize their cloud costs, they first need to understand what they're currently spending, whether there are any trends, and what their forecasted costs are. This means they need visibility into their cloud costs. GCP provides following for the same,

- Built-in reporting tools
- Custom dashboards
- Pricing calculator

Control

Organizations should also have precise access **permissions** in place to ensure that only authorized individuals in an organization have the power to create cloud resources. Creating **budgets** and **alerts** to notify key stakeholders when spending is getting off track is an important practice to keep costs under control.

The goals of the cost management tools are to provide:



Accountability

Because cloud spending can be done by any team, it's important to establish a culture of accountability for costs across the organization. This can be done by defining clear ownership for projects and sharing cost views with the departments and teams that are using cloud resources.

Intelligence

Organizations can make smart spending decisions with intelligent recommendations delivered by Google Cloud. These are tailored to each organization and help **optimize usage, save time on management, and minimize costs**. The recommendations can easily be applied for immediate cost savings and greater efficiency.

Best Practices & Advantages



Best practices and advantages for using available Google Cloud tools for ongoing cost management are as follows,



- 1 Identify the individual or team that will manage costs. Learn the difference between invoices and cost tools
- 2 Determine costs for projects before they're underway with the online **Pricing Calculator**.
- 3 Use cost management tools available in the Billing account for accountability & to set up budget alerts when costs reach a certain level.
- 4 Follow the advice of the Recommender to identify underused or idle resources
- 5 Apply labels to resources to identify them by which team should be billed for their use

TCO complexity with Cloud



Total cost of ownership (TCO) is a method used to calculate the various costs to host, run, integrate, secure and manage workloads over their lifetime.



Whether an organization is moving to the cloud for the first time or moving from a single cloud provider to multiple providers, how they calculate the total cost of ownership of the IT infrastructure will vary.

TCO calculation will be Simple, Infrequent, predictable



On-premises



Single Cloud



Multi Cloud



Multi Cloud +
On-Premises

TCO calculation will be Complex, Agile, unpredictable



When organizations use public cloud services, much of their capital expenditure shifts toward a pay-as-you-go OpEx model. But if they keep some of their business running on-premises and some running on public cloud. The **total cost of ownership (TCO)** for them would be more complex.

FEW JARGONS



What is Privacy?

Privacy, in the context of cloud technology, refers to the data an organization or an individual has access to and who they can share that data with.



What is Security?

Security in the cloud refers to the policies, procedures and controls put in place to keep data safe.



What is Compliance?

Compliance is about meeting standards set by a third party. This third party might be a regulatory authority, or it might be an international standards organization. Compliance is especially important in highly regulated industries such as Healthcare or Finance, where there is lot of sensitive data.



What is Availability?

It refers to how much time the cloud service provider guarantees data and services will be running or accessible. The availability of a service is typically documented as a percentage of time per year.

Security risks that Organizations face



Cyber attacks are bigger than ever. Many groups might use sophisticated methods to gain access to an organization's data. These attacks have become possible because almost every organization is digitally connected with their customers and employees globally.



Traditional on-premises systems generally rely on a perimeter-based security approach. The boundary around all of the data is protected by security features. Once someone is inside that security perimeter, they are deemed trustworthy, and therefore have access to everything.



Phishing attackers do research to gather information about you or anyone in your organization, then craft highly targeted emails to trick these people into thinking that the messages are genuine. These people are scammed into downloading malicious attachments, giving up their password, or sharing sensitive data.



Data can be lost, damaged, or destroyed by **viruses** or **malware**. Malware is derived from the terms malicious software. Hackers develop malicious software to infect and gain access to the victim computer without the user's consent. Alternatively, a set of files can be rendered unavailable to its intended users via **ransomware** until the ransom amount is paid.

Security risks that Organizations face



With **Physical Damage** organizations can still be responsible for data losses even when there is damage to the physical hard disk, there are power losses, or natural disasters such as floods, fires, and earthquakes.



Third-party systems are often used to address common business needs, without adequate security measures and regular checks. The **Unsecured third-party systems** systems can pose a threat to data security.



Lack of expert knowledge is a risk that Organizations face regularly. At the rate that technology is changing, investing in the right expertise to assess, develop, implement, and maintain data security plans is essential for businesses to stay ahead of potential data security threats.

When an organization manages its data in its own data centers, that organization is then responsible for all aspects of its security.

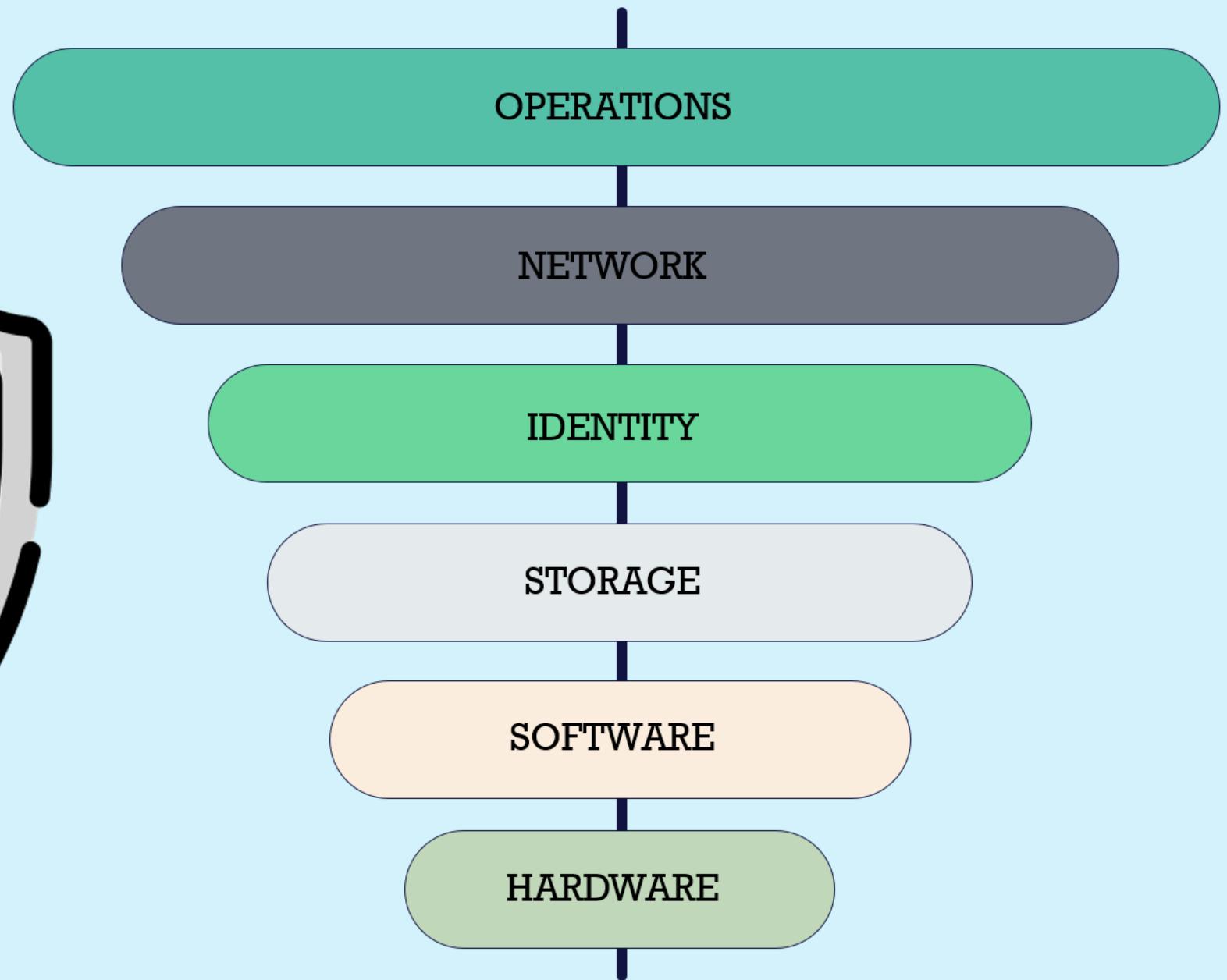
Google commitment to keep data secure & private

- 1 Google does not sell customer data to third parties & does not use for advertising. In Google Cloud, you own your Data, not Google.
- 2 All customer data is encrypted by default. The data at rest & in transit always encrypted.
- 3 Google Cloud guards against insider staff accessing to your data by following secure approaches.
- 4 Google Cloud never give any government entity “backdoor” access to your data
- 5 Google Cloud privacy practices are audited by external agencies by considering international standards.

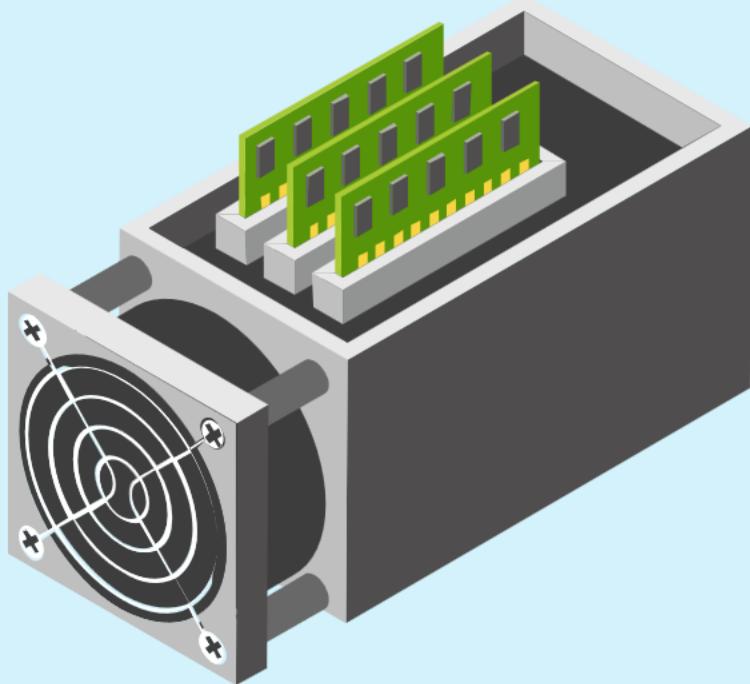


The advantage of using cloud technology is that the responsibility to secure data is shared between a business and the cloud provider. The cloud provider typically becomes the **data processor** & the organization is the **data controller**.

Google Cloud's multilayer approach to security



HARDWARE



1

Google designs its own servers, its storage, and its networking gear. It manufactures almost all of its own hardware, and third parties never see the overall process. The hardware is housed in these high-security data centers that are located around the world.

2

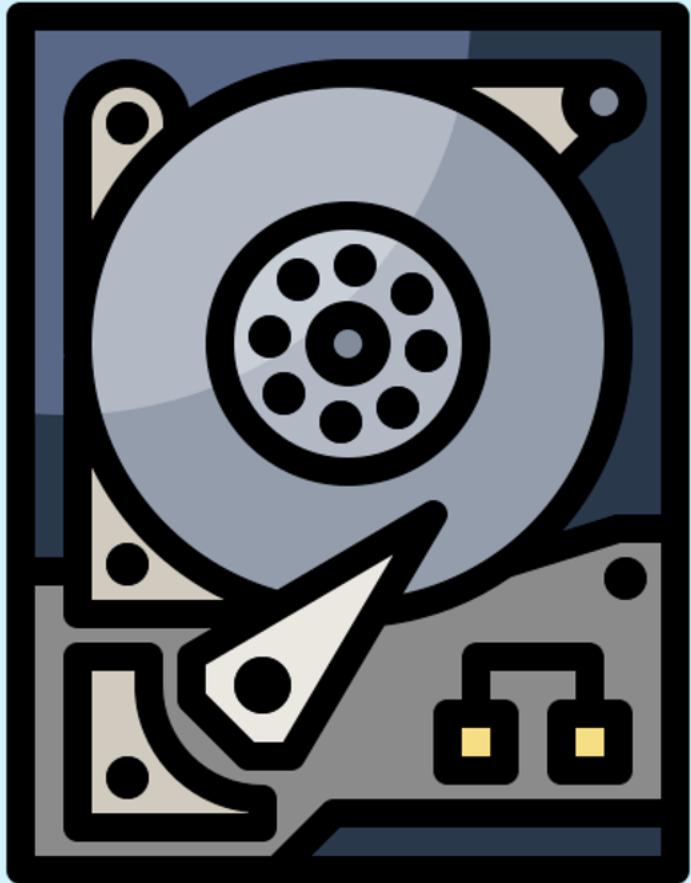
New server builds have a chip, called **Titan**, embedded. Titan checks a machine for integrity every time it boots up.

SOFTWARE



The Titan microcontroller continues to verify the operating systems and the rest of the deployed software stack. The server is not allowed onto the network and it holds zero data until its health is confirmed.

STORAGE



1

Storage is closely connected to the idea of data encryption at rest. Encryption at rest protects data when it is stored on physical media. ALL data at rest is also encrypted by default to help guard against unauthorized access.

2

The defense-in-depth process for storing data in Google Cloud is:

1. Data is broken into many pieces in memory.
2. These pieces, or “chunks”, are encrypted with their own data encryption key or ‘DEK’.
3. These DEKs are then encrypted a second time with key encryption key or ‘KEK’.
4. Encrypted chunks and wrapped KEKs are distributed across Google’s infrastructure.

3

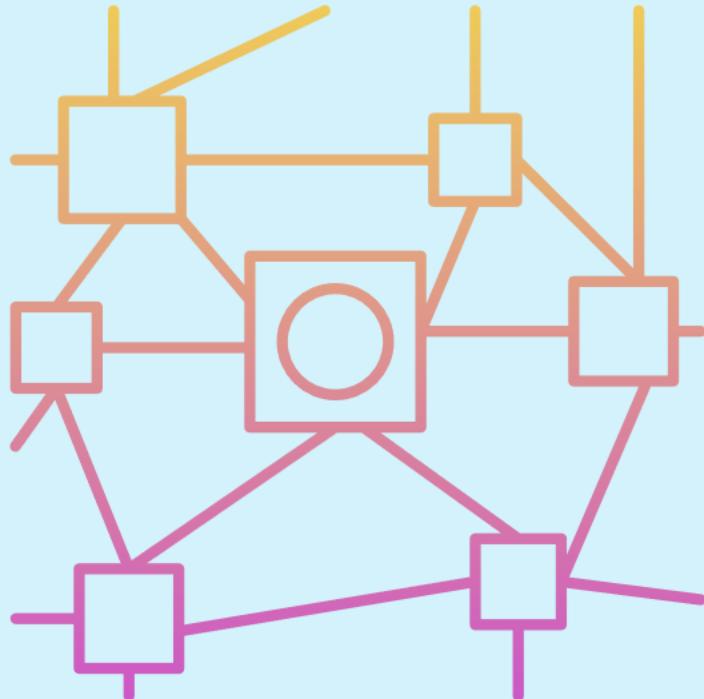
In the unlikely event that someone compromises an encryption key, they could only access one tiny piece of data, which, without all of the other pieces, would be unreadable.

IDENTITY



Instead of relying on the traditional perimeter approach to security, Google Cloud operates a zero-trust model. This means that every user and every machine that tries to access data or services must strongly authenticate identity at each stage for each file.

NETWORK



Anyone accessing the cloud does so via a network. Encryption in transit protects data as it moves across a network. Multiple layers of defense are in place to help protect customers against network attacks, like DDoS attacks.

OPERATIONS



At Google, a global operations team of more than 900 security experts monitor the system 24 hours a day, 365 days a year. Their role is to detect attacks and other issues and to respond to them.

Identity Access Management (IAM)

An Identity Access Management policy, or IAM policy, is made of three parts:

Who can do what on which resource



IT teams need to have a complete understanding of who can access what data/resource. Wherever possible, they need to establish **granular access policies**. They need to define who can do what, and on what cloud resource.

The “**who**” part of an IAM policy can be a Google account, a Google group, a service account, or a Google Workspace or Cloud Identity domain.

The “**can do what**” part is defined by an IAM role. This conveys what action an account/user can perform on the resource. Actions are like view, update, delete etc.

The **on which resource** tells on which cloud resource that an account/user can perform an action. Resources are like data, Compute Engines, GKE, reports etc.

Identity Access Management (IAM)

Who

Principal

A principal can be a Google Account (for end users), a service account (for applications and compute workloads), a Google group, or a Google Workspace account or Cloud Identity domain that can access a resource. Each principal has its own identifier, which is typically an email address.

can do what

Role

A role is a collection of permissions. Permissions determine what operations are allowed on a resource. When you grant a role to a principal, you grant all the permissions that the role contains.

There are three kinds of roles in Cloud IAM: **Primitive**, **Predefined**, **Custom**

on which resource

Policy

The allow policy is a collection of role bindings that bind one or more principals to individual roles.

When you want to define who (principal) has what type of access (role) on a resource, you create an allow policy and attach it to the resource.

IAM ROLES IN GOOGLE CLOUD

Primitive

Roles historically available in the Google Cloud console. These roles are **Owner**, **Editor**, and **Viewer**.

If you're a viewer on a given resource, you can examine it, but not change its state. If you're an editor, you can do everything a viewer can do plus change its state. And if you're an owner, you can do everything an editor can do plus manage roles and permissions on the resource.

Predefined

Roles that give finer-grained access control than the basic roles. For example, the predefined role **Pub/Sub Publisher** (`roles/pubsub.publisher`) provides access to only publish messages to a Pub/Sub topic.

Predefined roles give fine-grained access to specific Google Cloud resources and help prevent unwanted access to other resources.

Custom

Roles that you create to tailor permissions to the needs of your organization when predefined roles don't meet your needs.

Custom roles help you enforce the principle of least privilege, because they help to ensure that the principals in your organization have only the permissions that they need.

Google Cloud recommends using a “least-privilege” model, in which each person in your organization is given the minimal amount of privilege needed to do their job.

Assigning IAM roles to groups

Assigning roles to each individual may be a painful process. Let's see how we can overcome this ?



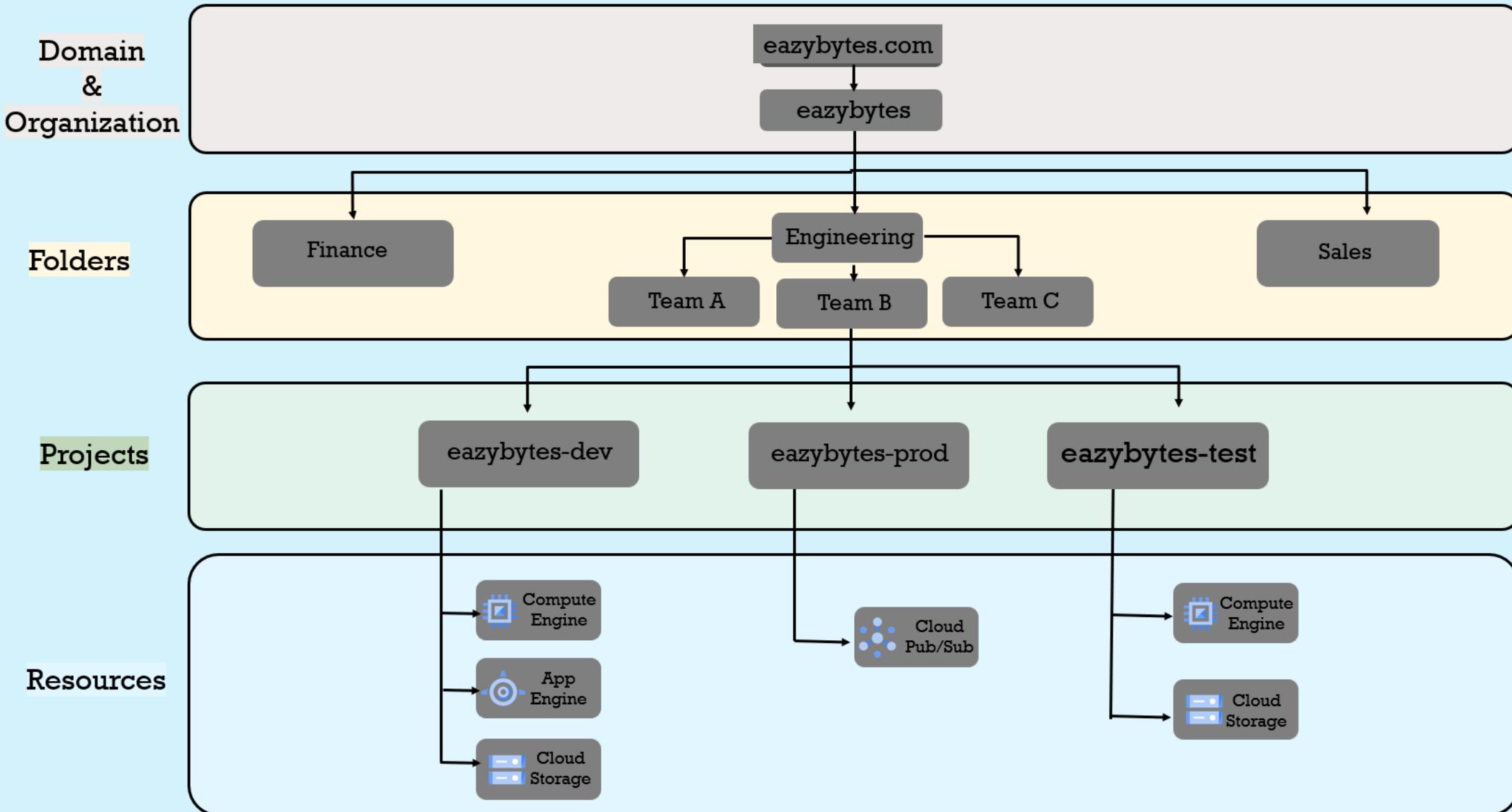
An organization can easily map job functions within the organization to specific groups. Each group can then be given specific roles for specific resources. Users get access only to what they need to do their job, and admins can grant default permissions to entire groups of users.

In the cloud environment, a **project** is the basis for enabling and Using Google Cloud capabilities, like managing APIs, enabling billing, adding and removing collaborators, and enabling other Google services.

Once the resources are created under project, we can form a **Resource hierarchy** which refers to the way your IT team can organize your business' Google Cloud environment and how that service structure maps to your organization's actual structure.

For example, with a resource hierarchy, IT teams can manage access and permissions for groups of related resources by teams or by projects or by both. .

Resource hierarchy in GCP



Resource hierarchy in GCP

Domain & Organization

Everything managed in Google Cloud is under a domain and an organization. The domain is handled through Cloud Identity and helps manage user profiles. The organization is managed through the Cloud Console and lets administrators see and control Google Cloud resources and permissions.

Folders

Folder resources optionally provide an additional grouping mechanism and isolation boundaries between projects. They can be seen as sub-organizations within the organization resource. Folder resources can be used to model different legal entities, departments, and teams within a company. Each team folder could contain additional sub-folders to represent different applications/teams.

Projects

Projects belong to the organization rather than the user that created them. Projects are used for grouping Google Cloud resources like Cloud Storage buckets. It can inherit permissions from any folders above it as well as from the organization at the top, making it easy to set organization-wide rules and policies that cascade down and are enforced throughout the hierarchy.

Resources

Resources inherit the policies of the parent resource. If you set a policy at the organization level, it is inherited by all its child folder and project resources, and if you set a policy at the project level, it is inherited by all its child resources. You can set an IAM policy at the organization level, the folder level, the project level, or (in some cases) the resource level.

Keeping Deployed Apps Operational

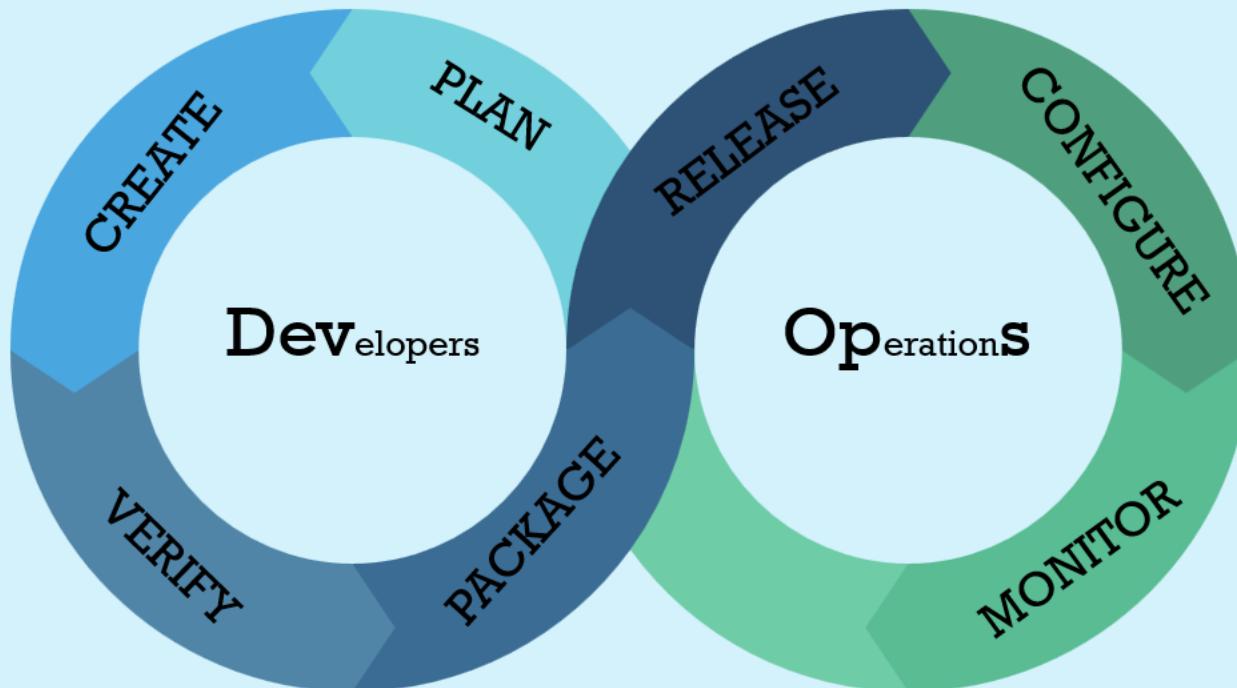


Developers aim is to release new functions frequently, increase core business value with new features, and release fixes fast. In contrast, operators are expected to keep systems stable, and so they often prefer to work more slowly to ensure reliability and consistency.

Service disruptions, unexpected system failure, planned down times are all common for any Organization. What makes difference is, how well the Organization/team are ready to handle these kind of situations

For organizations to thrive in the cloud, they'll need to adapt the best practices in IT operations. These practices including **DevOps culture**, **Site Reliability Engineering** & **Adjust expectations for service availability**

What is DevOps ?



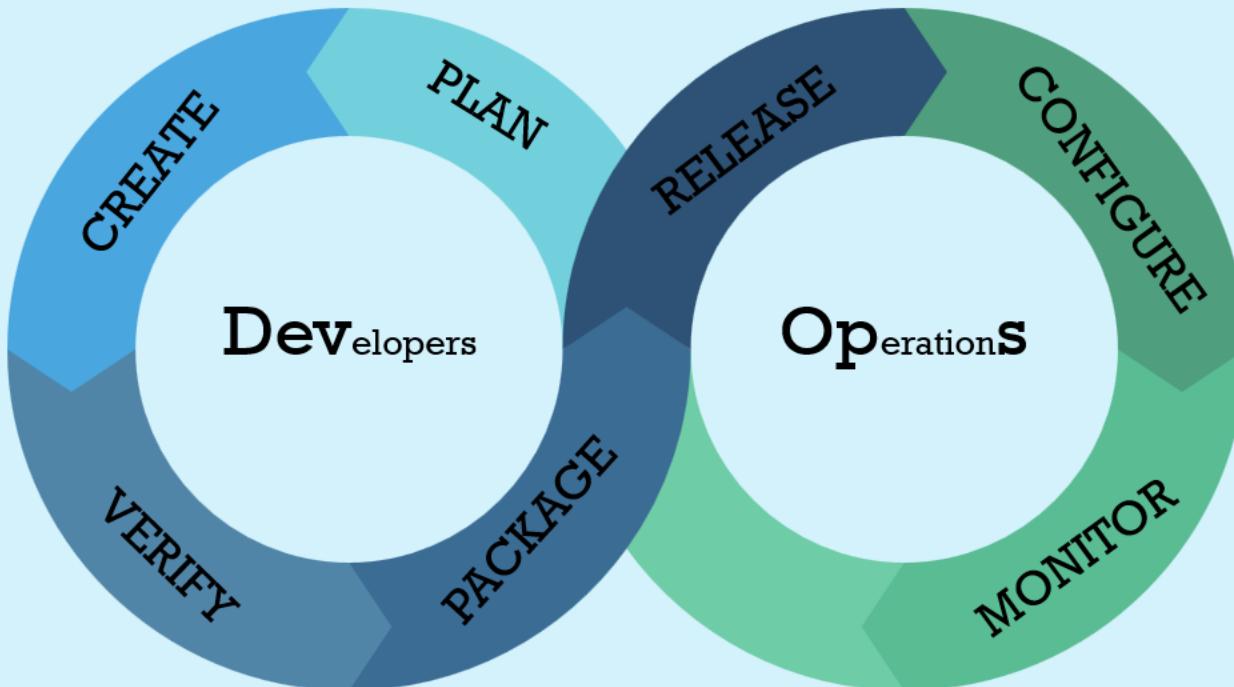
1

DevOps is a philosophy that seeks to create a more collaborative and accountable culture within developer and operations teams. The philosophy highlights how IT teams can operate, but doesn't give explicit guidance on how an organization should implement practices to be successful.

2

In other words, DevOps is the combination of application developers and operations, where the teams work together through the entire service lifecycle in an ongoing fashion.

Objectives of DevOps



- 1 Reduce silos between teams
- 2 Accept failure as normal.
- 3 Implement gradual change
- 4 Leverage tooling and automation
- 5 Measure everything

What is SRE ?



1

Site Reliability Engineering (or SRE) is a discipline that applies aspects of software engineering to operations. The goals of SRE are to create ultra-scalable and highly reliable software systems.

2

SRE is an **implementation** of DevOps, initiated and fostered by Google, that stresses quantitative targets to create and operate reliable applications.

3

DevOps convey **WHAT** to do where as SRE convey **HOW** to do. In programming language terminology, we can say,
class SRE implements DevOps

How SRE helps achieving DevOps Objectives

1

Reduce silos between teams

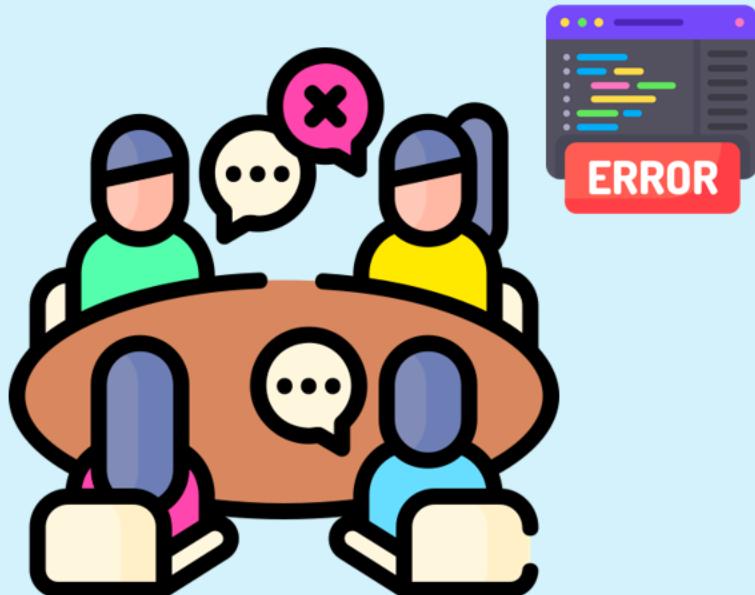


SRE emphasizes **shared ownership** of production between developers and operations. Together, they define service level objectives or SLOs, calculate error budgets, determine reliability, and order work priorities.

How SRE helps achieving DevOps Objectives

2

Accept failure as normal



SREs believe that accepting failure as normal helps to build an **iterative, collaborative culture**. One way this is done is by holding a blameless “lessons learned” discussion after an incident occurs.

3



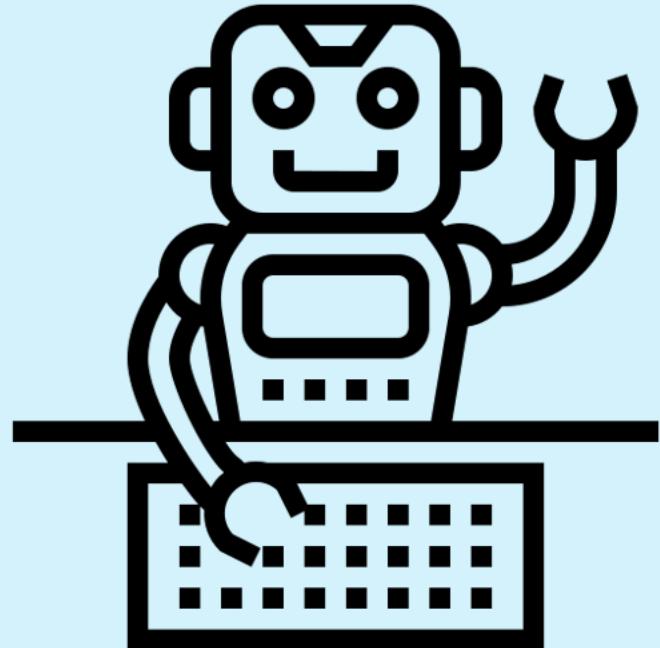
Implement gradual change

When implementing gradual changes, SREs aim to **reduce the cost of failure** by rolling out changes to a small percentage of users before making them generally available. This promotes more prototyping and launching iteratively.

How SRE helps achieving DevOps Objectives

4

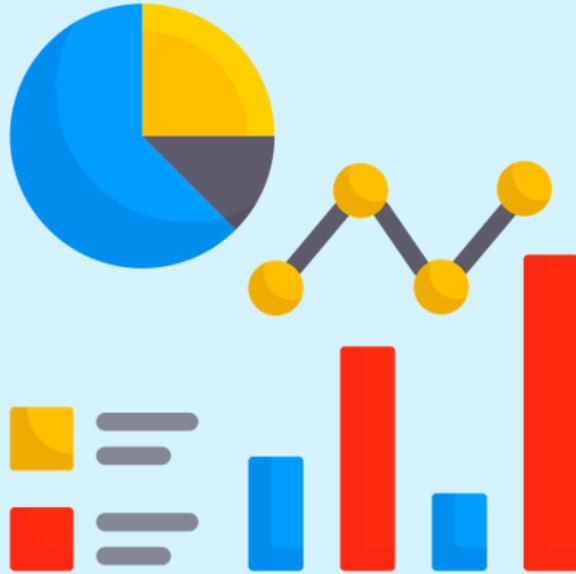
Leverage tooling and automation



In order to leverage tooling and automation, SREs focus on **toil automation**. In software engineering, toil is a type of work that is tied to running a production service. Toil automation, therefore, reduces the amount of manual, repetitive work.

How SRE helps achieving DevOps Objectives

5 Measure everything



'Measure everything' means tracking everything related to toil, reliability, and the health of their systems.

How SRE helps achieving DevOps Objectives



1

To foster SRE practices, organizations need a culture of **goal setting, transparency, and data-driven decision making**. They also need the tools to monitor their cloud environment, and to identify whether they are meeting their service level objectives.

2

First rule of SRE is 100% reliability is **NOT** a good objective. It shifts the mindset from '100% availability' to 99.99% or 99.999% availability. This is because, in order to roll out updates, operators have to take a system offline. Ensuring 100% service availability is also incredibly expensive for any business. This means that at some point the marginal cost of reliability exceeds the marginal value of reliability.

How Cloud providers measure service availability



- 1 Service level agreement (SLA)
- 2 Service level indicators (SLI)
- 3 Service level objectives (SLO)

What is SLA ?

1

Service level agreement (SLA)

SLA committed by
Cloud Provider

Response
time

0 milli
second 100 ms 200 ms 300 ms

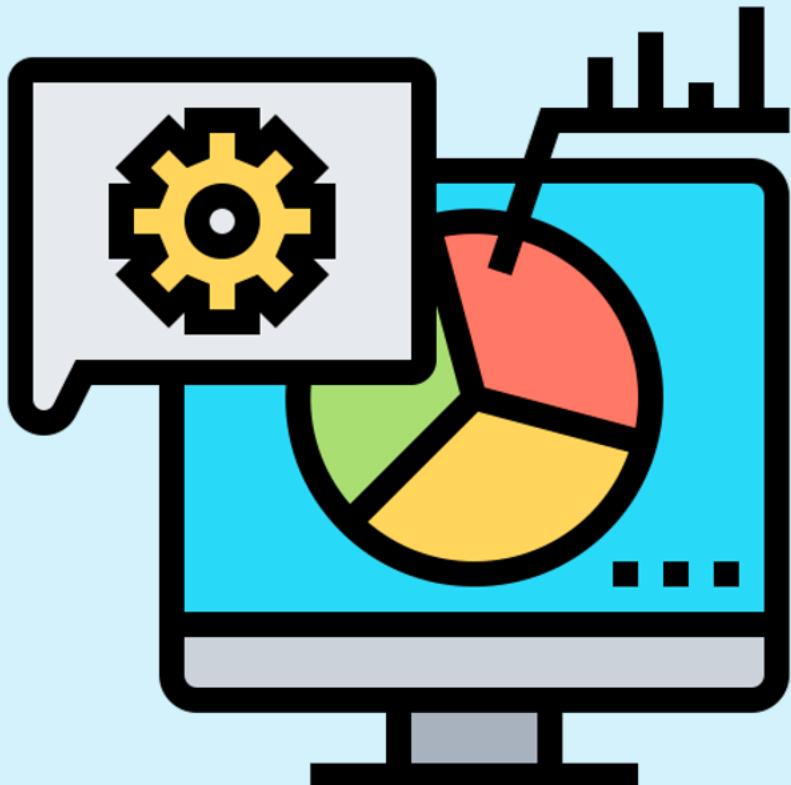
- ✓ A contractual commitment between the cloud service provider and the customer. The SLA provides the baseline level for the quality, availability, and reliability of that service. If the baseline service is not met by the provider, end users and end customers would be affected. The cloud provider would incur a cost usually paid out to the customer.
- ✓ For example, Google can commit all the traffic between the DB, Storage Systems, Compute Engine will be at a max of 300 ms. Anything beyond that, Google will pay back to the clients based on the agreement.

GCP SLA Details: <https://cloud.google.com/terms/sla>

What is SLI ?

2

Service level indicators (SLI)

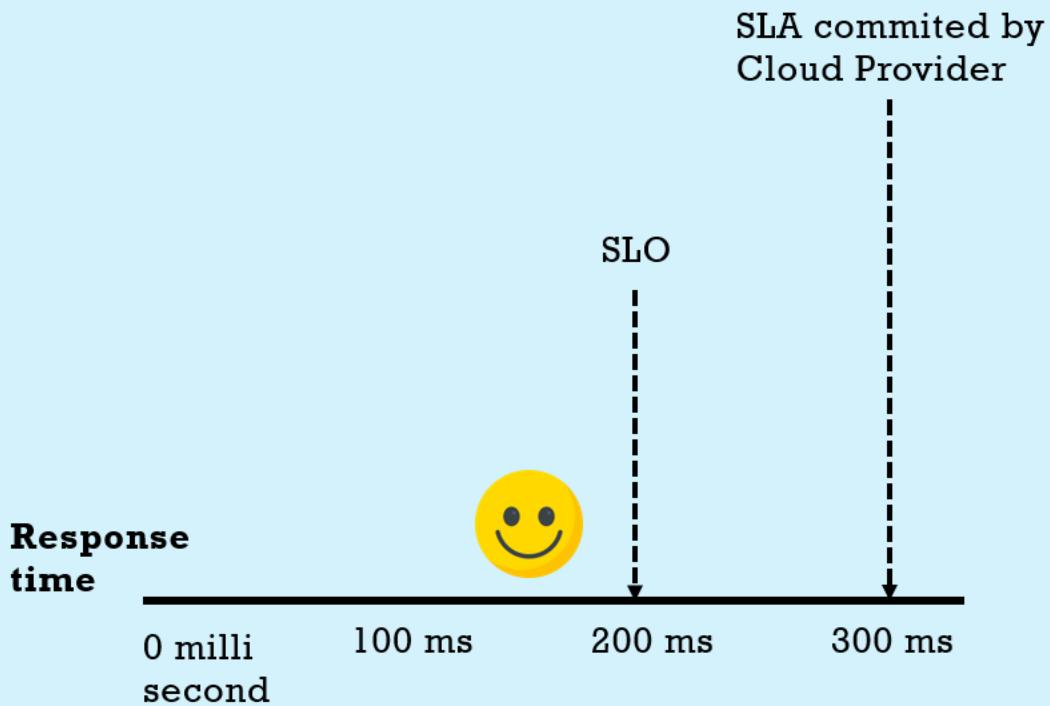


- ✓ An SLI describes the performance of some aspect of your service. It often include reliability, latency (which means delays in the system) and errors.
- ✓ SLIs are metrics over time—specific to a user journey, such as request/response, data processing, or storage—that show how well a service is doing.
- ✓ Example of SLIs include,
 - *An availability SLI is the ratio of the number of successful responses to the number of all responses.*
 - *A latency SLI is the ratio of the number of calls below a latency threshold to the number of all calls.*
 - *Request Latency: How long it takes to return a response to a request*
 - *Failure Rate: A fraction of all rates received (unsuccessful requests/all requests)*

What is SLO ?

3

Service level objectives (SLO)



- ✓ SLO is a key element which is always within the SLA. The goal for the cloud service performance level, shared between the cloud provider and a customer. If the service performance meets or exceeds the SLO, it means that end users, customers, and internal stakeholders are all happy.
- ✓ The SLO defines what qualifies as good service & you specify SLOs based on the SLIs. In other words, an SLO is a target value for an SLI, measured over a period of time. SLOs are always measured by SLIs. **SLI <= SLO**
- ✓ An SLO is built on the following kinds of information:
 - An SLI, which measures the performance of the service.
 - A performance goal, which specifies the desired level of performance.
 - A time period, called the compliance period, for measuring how the SLI compares to the performance goal.
- ✓ For example, you might have requirements like these:
 - Latency can't exceed 200 ms over a rolling 30-day period.
 - The system must have 99% availability measured over a calendar week.

How Cloud providers measure service availability

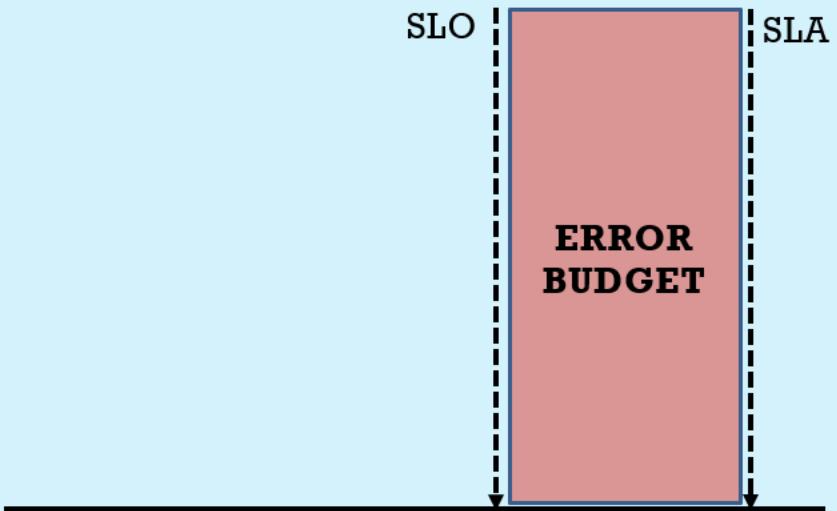


SLA – the agreement you made with your clients or users

SLOs – the objective your team must hit to meet the SLA

SLIs – the real numbers on your service performance

What is Error Budget ?



The **error budget** is typically the space between the SLA and the SLO. This error budget gives developers clarity into how many failed fixes they can attempt without affecting the end user experience.

For example, if your SLA is 99% & SLO is 99.5%, then the 0.5% will come under error budget. If the team meets an availability of 99%, then this means they used all the error budget whereas if they meet an availability of 99.3%, then 0.2% will be the left over budget.

The development team can 'spend' this error budget in any way they like. If the product is currently running flawlessly, with few or no errors, they can launch whatever they want, whenever they want. Conversely, if they have met or exceeded the error budget and are operating at or below the defined SLA, all launches are frozen until they reduce the number of errors to a level that allows the launch to proceed.

Google Cloud's operations suite

The tools in Google Cloud's operations suite are available in two major categories:

Operations-focused tools



Cloud Monitoring



Cloud Logging



Error Reporting



Service/SLO Monitoring

Application performance management tools



Cloud Debugger



Cloud Trace



Cloud Profiler

Google Cloud's operations suite



Cloud Monitoring

Cloud Monitoring is the foundation for Site Reliability Engineering because it provides visibility into the performance, uptime, and overall health of cloud-powered applications.

It collects measurements of the GCP resources used and visualizes them. Questions that Cloud Monitoring answers:

- How well are my resources performing?
- Are my applications meeting their SLOs and SLAs?
- Is there anything wrong that requires immediate action?



Cloud Logging

Google Cloud Logging is a fully managed service that performs at scale and can ingest application and system log data, as well as custom log data from GKE, environments, Virtual Machines, and Google Cloud services. It helps in storing, viewing, and interacting with logs.

All Google Cloud services, from GKE, to BigQuery, to Cloud Spanner, stream metrics and logs into the Google Cloud Monitoring and Cloud Logging components.



Error Reporting

Error Reporting aggregates and displays errors produced in your running cloud services. Using the centralized error management interface, you can find your application's top or new errors so that you can fix the root causes faster.

You can report errors from your application by sending them directly to Cloud Logging with proper formatting or by calling an Error Reporting API endpoint that sends them for you.



Service/SLO Monitoring

SLO monitoring helps you monitor the health of Google Cloud services by providing the tools to set up alerting policies on the performance of service-level objectives (SLOs).



Cloud Debugger (Deprecated)

Cloud Debugger helps monitor application performance. IT teams can inspect the state of a running application in real time, without stopping or slowing it down. This means that end users are not affected while a developer searches the source code. IT teams can use it to understand the behavior of their code in production and analyze its state to find those hard-to-find bugs.

Supports Snapshot feature captures the application state at a specific line location, along with variables.



Cloud Trace

Cloud Trace is another Google Cloud solution for monitoring application performance. It is a distributed tracing system that helps developers debug or fix and optimize their code. Collects latency data from your apps for display and analysis. Helps in identifying bottlenecks



Cloud Profiler

Cloud Profiler is a statistical, low-overhead profiler that **continuously gathers CPU usage and memory-allocation information** from your production applications. It attributes that information to the application's source code, helping you identify the parts of the application consuming the most resources, and otherwise illuminating the performance characteristics of the code.



Exam Preparation

1

Go to <https://www.cloudskillsboost.google>

- Read the official certification overview
- Review the exam guide
- Take the official free course from Google Cloud
- Revise Student slides provided

2

Attempt practice exams, quizzes

3

Make sure to revise the slides provided by me multiple times & be clear with all the products along with their features provided.

Enrolling for exam



1 Go to <https://www.webassessor.com/googlecloud>

- Create an account with personal email & login
- Enroll into the exam by paying \$99 USD (vary based on your Country)
- It is a 90-minute exam with 50 MCQ
- Can select either English or Japanese
- Can select either On-site (Need to go to the nearest exam center) or online (Can write exam from home but please note it is a online proctor exam)

2 If you opted for Online exam, make sure you have a private room in the home to attempt the exam along with a working webcam & headphones.

3 Online proctor will examine the room and may request to remove mobile phone, books etc. Be prepared !!!

Tips for the exam



- 1** Mark difficult or not sure questions for later review. At the end, you can review such questions & answer
- 2** While answering a question, follow below rules
 - Eliminate obvious wrong answers
 - Identify the significant clauses in the question
 - Always consider Google Cloud best practices
 - Always look for the most efficient method & least expensive method
 - Apply common sense while answering the non technical & scenario based questions
- 3** Make sure to monitor time & review all answers once quickly. Don't try to close the exam well ahead. When confident & done, submit the exam.
- 4** Immediate grade where PASS/FAIL will be provided. But official confirmation will be shared after few days from Google Cloud. No marks will be provided.
- 5** If you PASS, a BIG CONGRATULATIONS !!!

If you don't pass an exam, you can take it again after 14 days. If you don't pass the second time, you must wait 60 days before you can take it a third time. If you don't pass the third time, you must wait 365 days before taking it again. Every retake attempt requires payment again.

THANK YOU

