



Lesson 12

Cloud Services (GCP)

Go to the moon!

Learn About



- Key components
- IaaS, PaaS, and SaaS
- Architecture examples on cloud



Why cloud? (I)

Criteria	Cloud	On-prem
Economy	<ul style="list-style-type: none">• On-going costs (OPEX) are predictable.• It can be easily changed depending on the company's situation.	<ul style="list-style-type: none">• Up-front investments (CAPEX), are mostly redundant.• A IT specialist to manage servers is scarce.
Security	<ul style="list-style-type: none">• Offer systems to discover where personally identifiable information (PII) is stored and who accessed when (Google Cloud DLP).• Encryption systems are ready to use• Identity and Access Management (IAM) ensure authenticated and authorized users can access the system and data	<ul style="list-style-type: none">• It is harder to set up security systems
Resilience	<ul style="list-style-type: none">• Much simpler with a few clicks	<ul style="list-style-type: none">• A backup site is necessary in case of natural disasters• Drills are needed to conduct frequently

Why cloud? (2)

Criteria	Cloud	On-prem
Scalability	<ul style="list-style-type: none">• Can deal with spikes in volume with autoscaling features	<ul style="list-style-type: none">• More difficult, and may require a specialist to tune our infrastructure up and down
Focus	<ul style="list-style-type: none">• Focus on the business	<ul style="list-style-type: none">• Focus on the business, and the infrastructure
Agility	<ul style="list-style-type: none">• Facilitate innovation via quick resource provision and deletion	<ul style="list-style-type: none">• Hard to fulfill this



Which case to use the cloud?

Figure 1.1. Steady growth in resource consumption

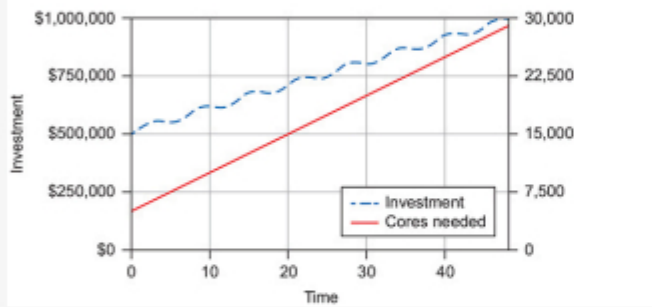
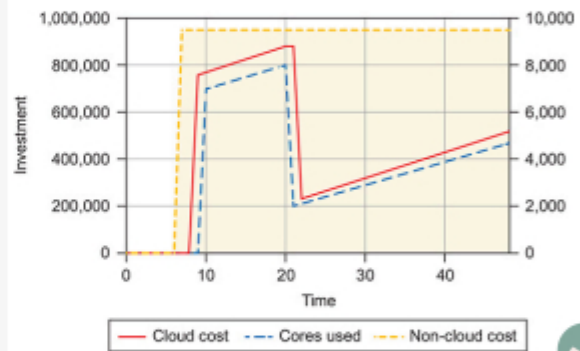
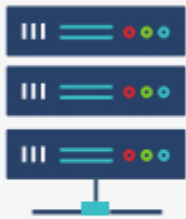


Figure 1.2. Unexpected pattern of resource consumption



Key components



Hardware

- **Purpose:** Provide computing power
- **Structure:** Servers, GPUs, power supply, memory and others



Storage

- **Purpose:** Storage and manage data
- **Structure:** Data is stored across many disks in storage arrays



Network

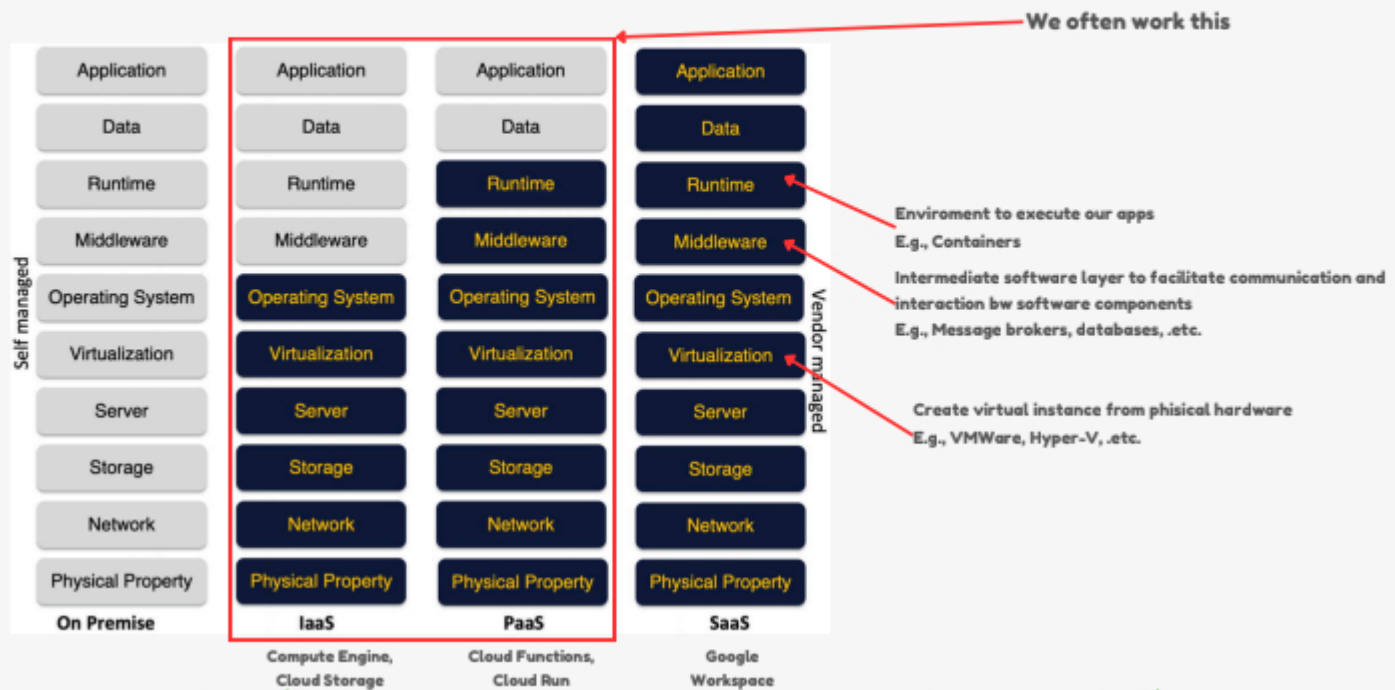
- **Purpose:** Connect cloud services over the Internet
- **Structure:** physical wires, switches, routers and others



Virtualization

- **Purpose:** Abstract machine's resources
- **Structure:** Hypervisors sit on top of physical hardware

IaaS, PaaS, and SaaS

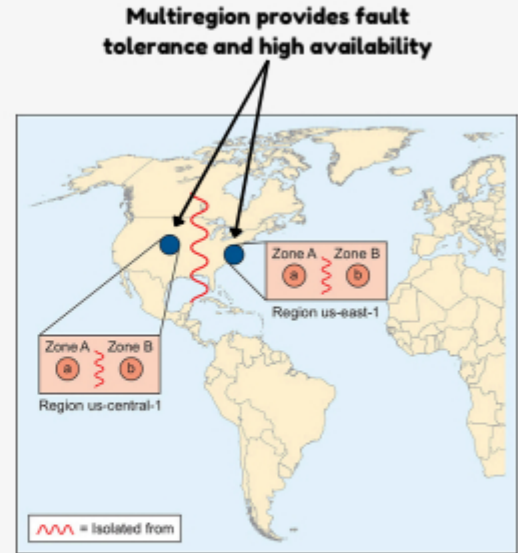


Data Center (I)

on the way



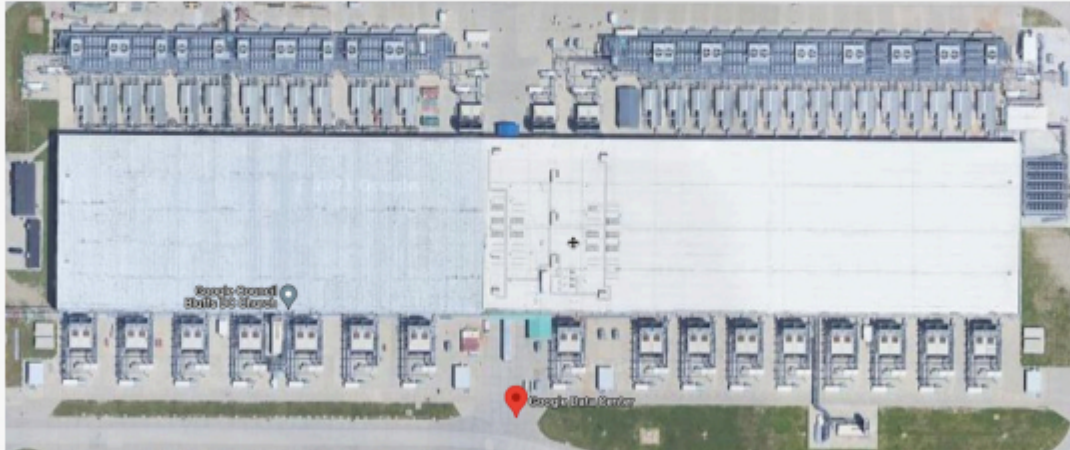
GCP has 15 different **regions** across the world, from the US, Brazil, Western Europe, India, and East Asia to Australia



Multiregion provides fault tolerance and high availability

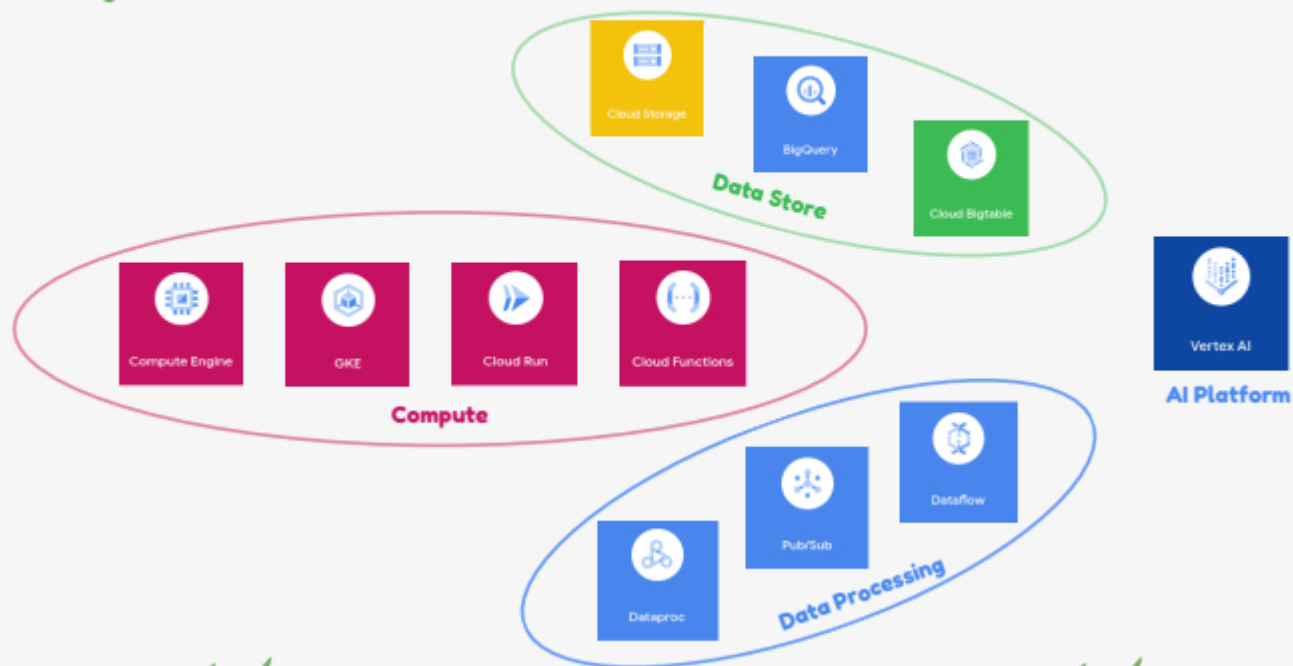
- **Region** ~ a city (us-east1), while **zone** ~ one or more nearby buildings (us-east1-a), refer [here](#) for more details.
- **Zones** can be down by power outages, while **regions** by disasters

Data Center (2)



Google Cloud us-central1 data center with 4 zones (the largest region)

Popular Services



Machine type families

Workload type						
General-purpose workloads				Compute-optimized	Memory-optimized	Accelerator-optimized
E2	N2, N2D, N1	C3	Tau T2D, Tau T2A	H3, C2, C2D	M3, M2, M1	A2, G2
Day-to-day computing at a lower cost	Balanced price/performance across a wide range of machine types	Consistently high performance for a variety of workloads	Best per-core performance/cost for scale-out workloads	Ultra high performance for compute-intensive workloads	Highest memory to compute ratios for memory-intensive workloads	Optimized for accelerated high performance computing workloads



Machine types

Each family has multiple types:

tasks require
more cpu

tasks require
more ram

C3 standard

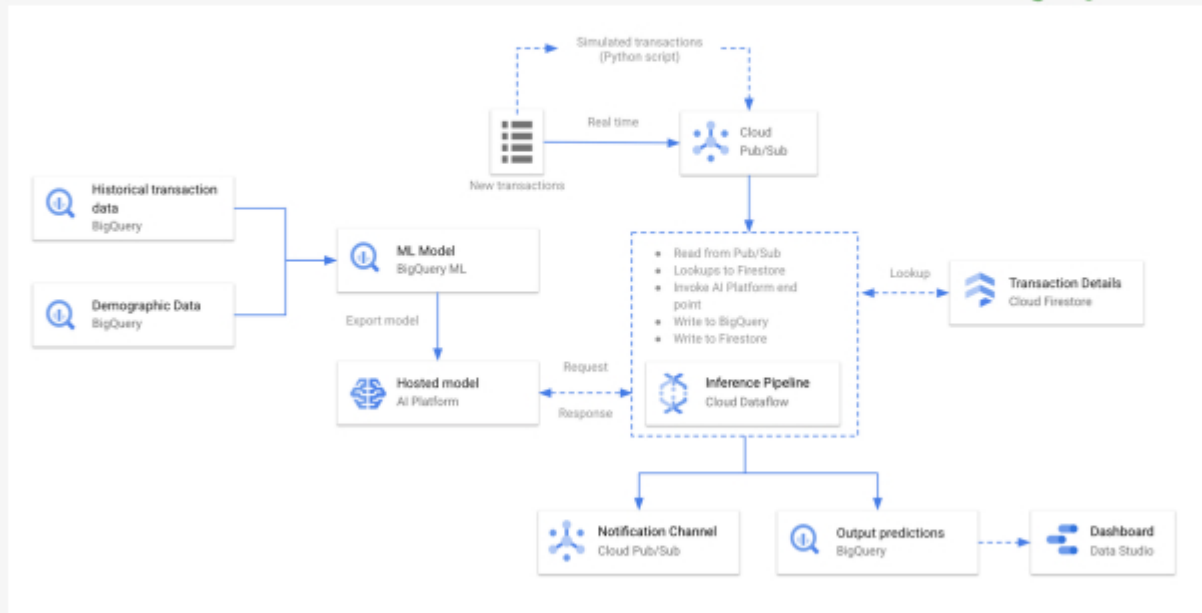
C3 highcpu

C3 highmem

C3 with Local SSD

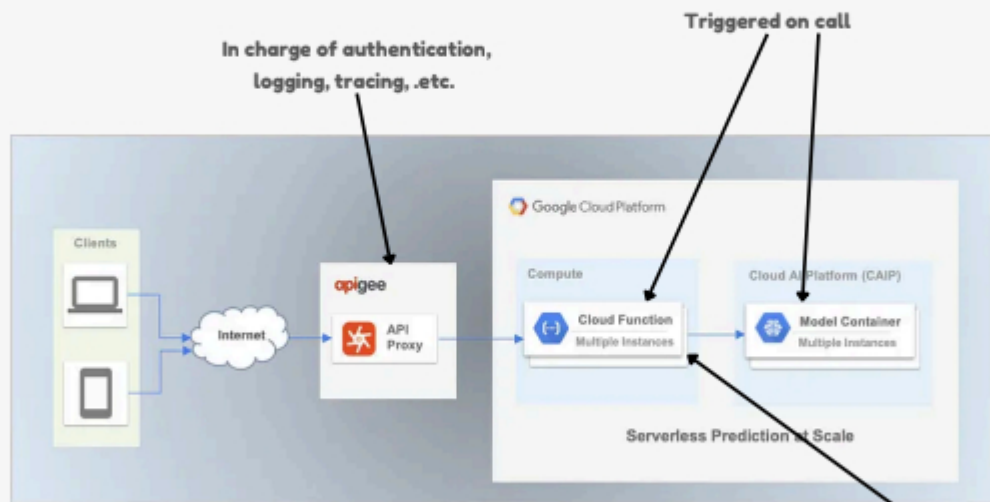
Machine types	vCPUs*	Memory (GB)	Default egress bandwidth (Gbps)‡	Tier_1 egress bandwidth (Gbps)#
c3-standard-4	4	16	up to 23	N/A
c3-standard-8	8	32	up to 23	N/A
c3-standard-22	22	88	up to 23	N/A
c3-standard-44	44	176	up to 32	up to 50
c3-standard-88	88	352	up to 62	up to 100
c3-standard-176	176	704	up to 100	up to 200

ML architectures on cloud (I)



Credit card fraud detection high level architecture


ML architectures on cloud (2)



Adverse drug reaction risks prediction high level architecture

Add more features to model serving such as preprocessing, routing by model versions, logging, .etc.

References

- Cloud Computing Basics: A Non-Technical Introduction 1st ed. Edition
 - Google Cloud Platform in Action
 - <https://cloud.google.com/learn/what-is-iaas>
 - <https://googlecloudcheatsheet.withgoogle.com/>
 - <https://github.com/priyankavergadia/google-cloud-4-words>
 - The Ultimate Guide to Building a Google Cloud Foundation
 - Google Cloud Helpful Links
 - <https://www.redhat.com/en/topics/cloud-computing/what-is-cloud-infrastructure>
 - <https://medium.com/google-cloud/13-most-common-google-cloud-reference-architectures-23630b46326d>
 - https://maelfabien.github.io/bigdata/gcps_4/#from-on-premise-to-gcp
 - <https://cloud.google.com/architecture>
- 

Thank You!



Created by
Quan Dang



