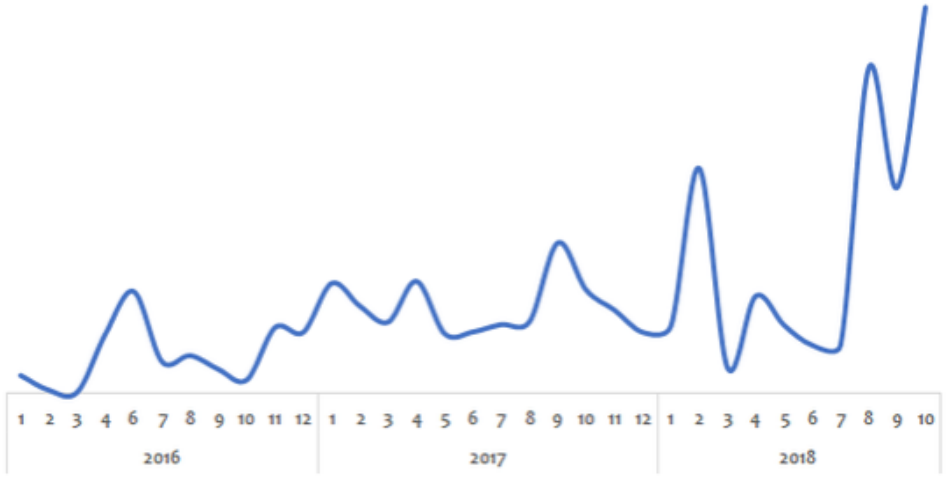
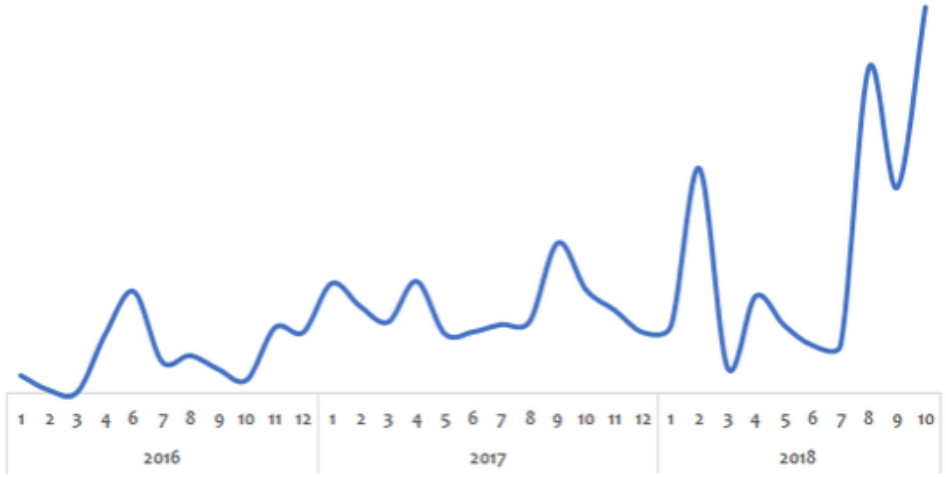
GCP - Getting started

Before the Cloud - Example 1 - Online Shopping App



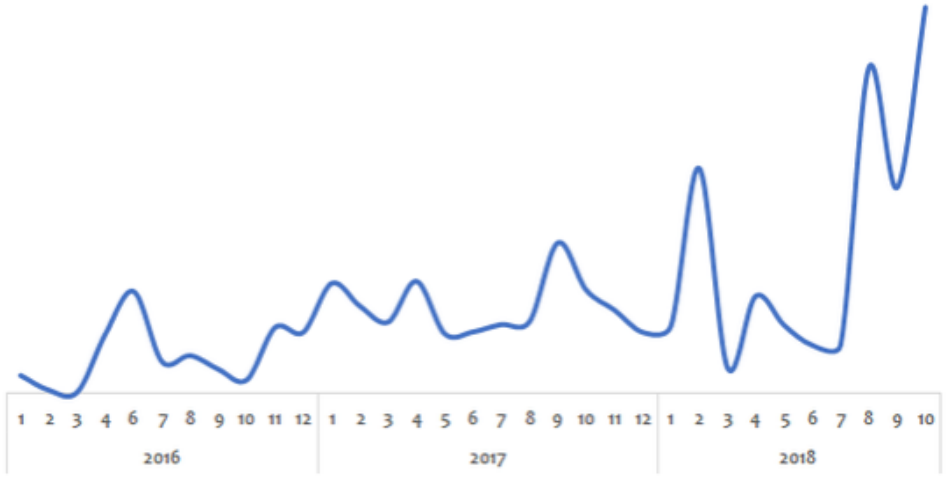
- Challenge:
  - Peak usage during holidays and weekends
  - Less load during rest of the time
- Solution (before the Cloud):
  - **PEAK LOAD provisioning** : **Procure** (Buy) infrastructure **for peak load**
    - What would the infrastructure be doing during periods of low loads?

Before the Cloud - Example 2 - Startup



- Challenge:
  - Startup suddenly becomes popular
  - How to handle the **sudden increase** in load?
- Solution (before the Cloud):
  - **Procure** (Buy) infrastructure assuming they would be successful
    - What if they are not successful?

Before the Cloud - Challenges



- High cost of procuring infrastructure
- Needs ahead of time planning (**Can you guess the future?**)
- Low infrastructure utilization (**PEAK LOAD** provisioning)
- Dedicated infrastructure maintenance team (**Can a startup afford it?**)

Silver Lining in the Cloud

In28 Minutes

- How about **provisioning (renting) resources** when you want them and releasing them back when you do not need them?
  - **On-demand resource provisioning**
  - Also called **Elasticity**

Cloud - Advantages

In28 Minutes

- Trade **"capital expense"** for **"variable expense"**
- Benefit from massive **economies of scale**
- Stop **guessing** capacity
- Stop spending money running and maintaining data centers
- **"Go global"** in minutes

Google Cloud Platform (GCP)

In28 Minutes

- **One of the Top 3** cloud service providers
- Provides a number of services (200+)
- Reliable, secure and highly-performant:
  - Infrastructure that powers 8 services with over 1 Billion Users: Gmail, Google Search, YouTube etc
- One thing I love : **"cleanest cloud"**
  - Net carbon-neutral cloud (electricity used matched 100% with renewable energy)
- The entire course is all about GCP. You will learn it as we go further.

Google Cloud

Best path to learn GCP!

In28 Minutes

- Cloud applications make use of multiple GCP services
- There is **no single path** to learn these services independently
- HOWEVER, we've worked out a simple path!

Setting up GCP Account

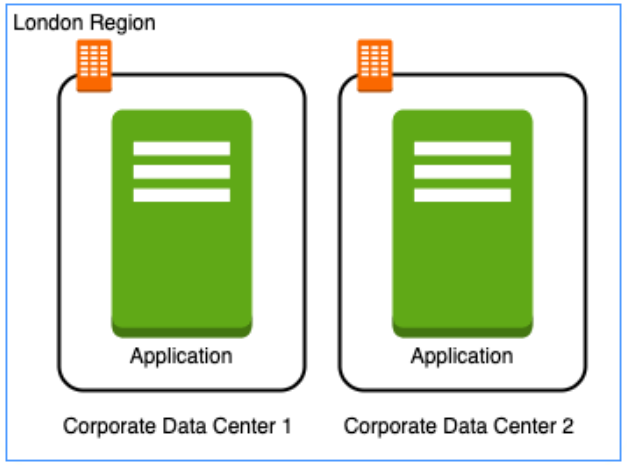
- Create GCP Account

Regions and Zones

Regions and Zones

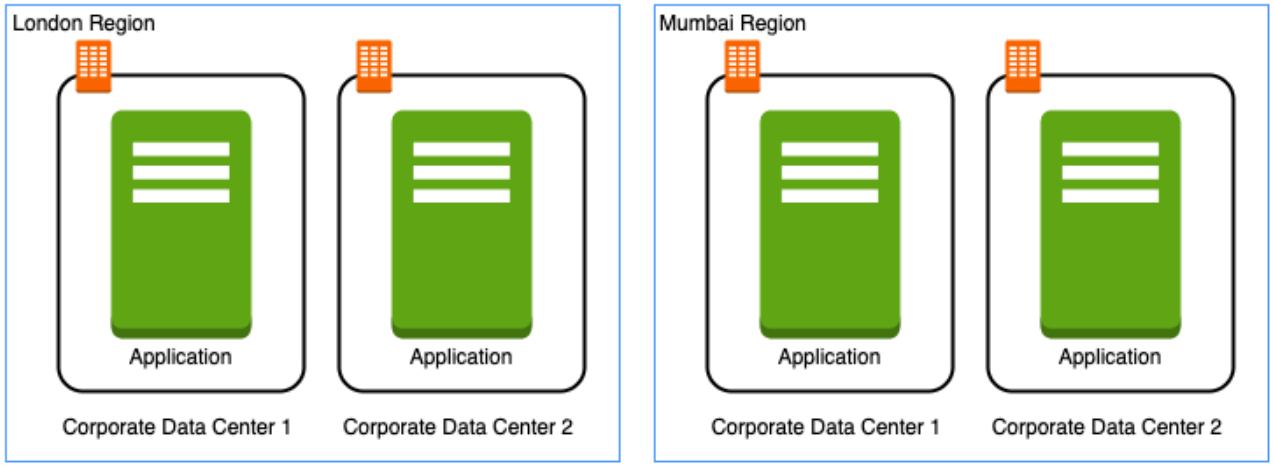
- Imagine that your application is deployed in a data center in London
- What would be the challenges?
  - Challenge 1 : Slow access for users from other parts of the world (**high latency**)
  - Challenge 2 : What if the data center crashes?
    - Your application goes down (**low availability**)

Multiple data centers



- Let's **add in one more data center** in London
- What would be the challenges?
  - Challenge 1 : Slow access for users from other parts of the world
  - Challenge 2 (**SOLVED**) : What if one data center crashes?
    - Your application is **still available** from the other data center
  - Challenge 3 : What if **entire region** of London is unavailable?
    - Your application goes down

Multiple regions



- Let's add a new region : Mumbai
- What would be the challenges?
  - Challenge 1 (**PARTLY SOLVED**) : Slow access for users from other parts of the world
    - You can solve this by adding deployments for your applications in other regions
  - Challenge 2 (SOLVED) : What if one data center crashes?
    - Your application is still live from the other data centers
  - Challenge 3 (**SOLVED**) : What if entire region of London is unavailable?
    - Your application is served from Mumbai

Regions

- Imagine setting up data centers in different regions around the world
  - **Would that be easy?**
- (Solution) Google provides **20+ regions** around the world
  - Expanding every year
- **Region** : Specific geographical location to host your resources
- **Advantages**:
  - High Availability
  - Low Latency
  - Global Footprint
  - Adhere to government **regulations**

Zones

- How to achieve high availability in the same region (or geographic location)?
  - Enter **Zones**
- Each Region has three or more **zones**
- (Advantage) **Increased availability and fault tolerance** within same region
- (Remember) Each Zone has **one or more discrete clusters**
  - **Cluster** : distinct physical infrastructure that is housed in a data center
- (Remember) Zones in a region are connected through **low-latency** links

Regions and Zones examples

In28 Minutes

New Regions and Zones are constantly added

Region Code	Region	Zones	Zones List
us-west1	The Dalles, Oregon, North America	3	us-west1-a us-west1-b us-west1-c
europe-north1	Hamina, Finland, Europe	3	europe-north1-a, europe-north1-b europe-north1-c
asia-south1	Mumbai, India APAC	3	asia-south1-a, asia-south1-b asia-south1-c

Compute

Compute Engine Fundamentals

Google Compute Engine (GCE)

In 28 Minutes

- In corporate data centers, applications are deployed to physical servers
- Where do you deploy applications in the cloud?
  - Rent virtual servers
  - **Virtual Machines** - Virtual servers in GCP
  - **Google Compute Engine (GCE)** - Provision & Manage Virtual Machines

Compute Engine - Features

- Create and manage lifecycle of Virtual Machine (VM) instances
- **Load balancing** and **auto scaling** for multiple VM instances
- **Attach storage** (& network storage) to your VM instances
- Manage **network connectivity and configuration** for your VM instances
- **Our Goal**:
  - Setup VM instances as HTTP (Web) Server
  - Distribute load with Load Balancers

Compute Engine Hands-on

- Let's create a few VM instances and play with them
- Let's check out the lifecycle of VM instances
- Let's use SSH to connect to VM instances

Compute Engine Machine Family

In 28 Minutes

- What type of hardware do you want to run your workloads on?
- Different Machine Families for Different Workloads:
  - **General Purpose (E2, N2, N2D, N1)** : Best price-performance ratio
    - Web and application servers, Small-medium databases, Dev environments
  - **Memory Optimized (M2, M1)**: Ultra high memory workloads
    - Large in-memory databases and In-memory analytics
  - **Compute Optimized (C2)**: Compute intensive workloads
    - Gaming applications

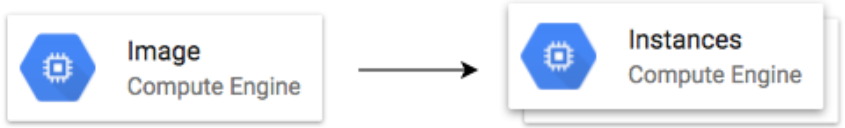
Compute Engine Machine Types

In28 Minutes

Machine name	vCPUs[1]	Memory (GB)	Max number of persistent disks (PDs)[2]	Max total PD size (TB)	Local SSD	Maximum egress bandwidth (Gbps)[3]
e2-standard-2	2	8	128	257	No	4
e2-standard-4	4	16	128	257	No	8
e2-standard-8	8	32	128	257	No	16
e2-standard-16	16	64	128	257	No	16
e2-standard-32	32	128	128	257	No	16

- How much CPU, memory or disk do you want?
  - Variety of machine types are available for each machine family
  - Let's take an example : **e2-standard-2**:
    - **e2** - Machine Type Family
    - **standard** - Type of workload
    - **2** - Number of CPUs
- Memory, disk and networking capabilities increase along with vCPUs

Image



- What operating system and what software do you want on the instance?
- Type of Images:
  - **Public Images**: Provided & maintained by Google or Open source communities or third party vendors
  - **Custom Images**: Created by you for your projects

Compute Engine Hands-on : Setting up a HTTP server

In 28 Minutes

```
#! /bin/bash
sudo su
apt update
apt -y install apache2
sudo service apache2 start
sudo update-rc.d apache2 enable
echo "Hello World" > /var/www/html/index.html
echo "Hello world from $(hostname) $(hostname -I)" > /var/www/html/index.html
```

- Commands:
  - `sudo su` - execute commands as a root user
  - `apt update` - Update package index - pull the latest changes from the APT repositories
  - `apt -y install apache2` - Install apache 2 web server
  - `sudo service apache2 start` - Start apache 2 web server
  - `echo "Hello World" > /var/www/html/index.html` - Write to index.html
  - `$(hostname)` - Get host name
  - `$(hostname -I)` - Get host internal IP address

Internal and External IP Addresses

In 28 Minutes

- **External** (Public) IP addresses are **Internet addressable**.
- **Internal** (Private) IP addresses are **internal** to a corporate network
- You CANNOT have two resources with same public (External) IP address.
  - HOWEVER, two different corporate networks CAN have resources with same Internal (private) IP address
- All **VM instances** are assigned at least one Internal IP address
- Creation of External IP addresses can be enabled for VM instances
  - (Remember) When you stop an VM instance, External IP address is lost
- **DEMO**: VM instances - Internal and External IPs

Static IP Addresses

In28 Minutes

- Scenario : How do you get a constant External IP address for a VM instance?
  - Quick and dirty way is to assign an Static IP Address to the VM!
- **DEMO**: Using Static IP Address with an VM instance

Static IP Addresses - Remember

In 28 Minutes

- Static IP **can be switched** to another VM instance in same project
- Static IP **remains attached** even if you stop the instance. You have to manually detach it.
- Remember : You are **billed for** an Static IP when **you are NOT using it**!
  - Make sure that you explicitly release an Static IP when you are not using it.

Simplify VM HTTP server setup

In28 Minutes

- How do we **reduce** the **number of steps** in creating an VM instance and setting up a HTTP Server?
- Let's explore a few options:
  - **Startup script**
  - **Instance Template**
  - **Custom Image**

Bootstrapping with Startup script

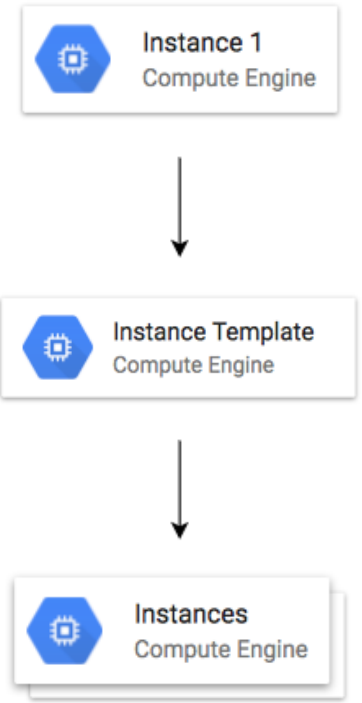
In 28 Minutes

```
#!/bin/bash
apt update
apt -y install apache2
echo "Hello world from $(hostname) $(hostname -I)" > /var/www/htr
```

- **Bootstrapping**: Install OS patches or software when an VM instance is launched.
- In VM, you can configure **Startup script** to bootstrap
- **DEMO** - Using Startup script

Instance templates

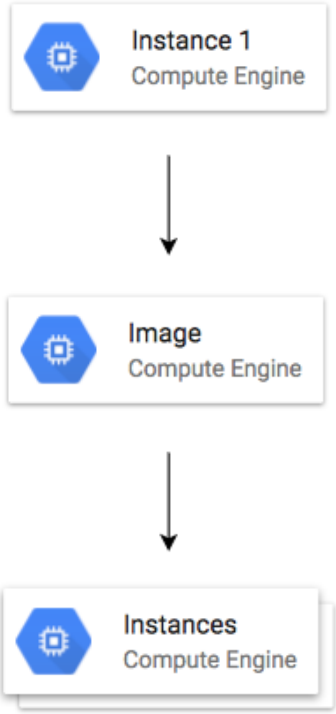
- Why do you need to specify all the VM instance details (Image, instance type etc) **every time** you launch an instance?
  - How about creating a **Instance template**?
  - Define **machine type, image, labels, startup script** and other properties
- Used to create **VM instances** and **managed instance groups**
  - Provides a **convenient way** to create similar instances
- **CANNOT** be updated
  - To make a change, copy an existing template and modify it
- (Optional) Image family can be specified (example - debian-9):
  - Latest non-deprecated version of the family is used
- **DEMO** - Launch VM instances using Instance templates



Reducing Launch Time with Custom Image

In28 Minutes

- Installing OS patches and software at launch of VM instances **increases boot up** time
- How about creating a custom image with OS patches and software **pre-installed**?
  - Can be created from an instance, a persistent disk, a snapshot, another image, or a file in Cloud Storage
  - Can be shared across projects
  - (Recommendation) Deprecate old images (& specify replacement image)
  - (Recommendation) **Hardening an Image** - Customize images to your corporate security standards
- **Prefer** using **Custom Image** to **Startup script**
- **DEMO** : Create a Custom Image and using it in an Instance Template



Compute Engine Scenarios

In28 Minutes

Scenario	Solution
What are the pre-requisites to be able to create a VM instance?	1. Project 2. Billing Account 3. Compute Engines APIs should be enabled
You want dedicated hardware for your compliance, licensing, and management needs	Sole-tenant nodes
I have 1000s of VM and I want to automate OS patch management, OS inventory management and OS configuration management (manage software installed)	Use "VM Manager"
You want to login to your VM instance to install software	You can SSH into it
You do not want to expose a VM to internet	Do NOT assign an external IP Address
You want to allow HTTP traffic to your VM	Configure Firewall Rules