

---

# **Cloud Computing #3**

## **Google Cloud Compute Engine**

---

# Creating a Virtual Machine

<https://google.qwiklabs.com/focuses/3563?parent=catalog>

40 minutes

1 Credit



Rate Lab

GSP001



Google Cloud Self-Paced Labs

# Compute Engine: Qwik Start - Windows

40 minutes

1 Credit



<https://google.qwiklabs.com/focuses/560?parent=catalog>

**GSP093**



Google Cloud Self-Paced Labs

---

# Introducing Google Cloud Platform

## GCP Fundamentals: Core Infrastructure

### Agenda

- **Cloud Computing Overview**
- Virtual Private Cloud (VPC) Network
- Compute Engine
- Quiz and lab

# What is cloud computing?



On-demand  
self-service

No human  
intervention  
needed to get  
resources



Broad network  
access

Access  
from  
anywhere



Resource  
pooling

Provider  
shares  
resources  
to  
customers



Rapid  
elasticity

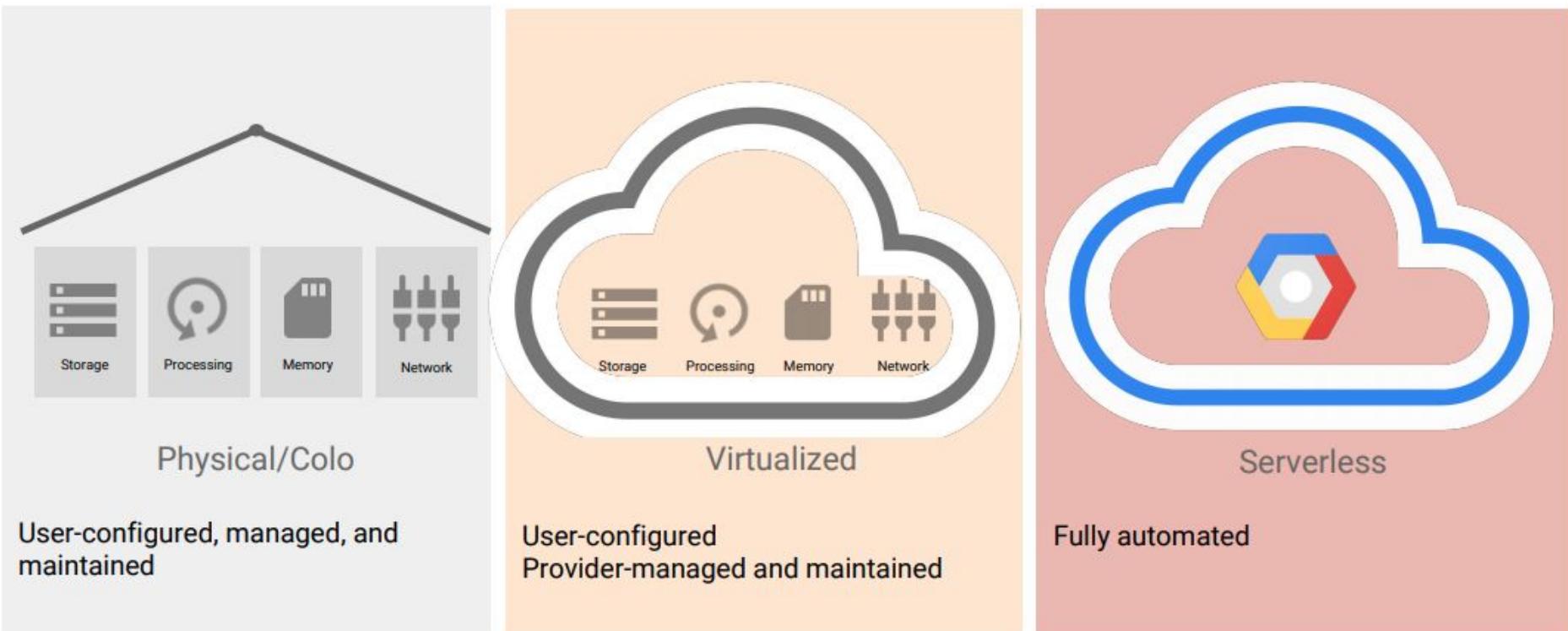
Get more  
resources  
quickly as  
needed



Measured  
service

Pay only  
for what  
you  
consume

# How did we get here? Where are we going?



# GCP computing architectures meet you where you are



Compute  
Engine

IaaS



Kubernetes  
Engine

Hybrid



App  
Engine

PaaS



Cloud  
Functions

Serverless  
logic



Managed  
services

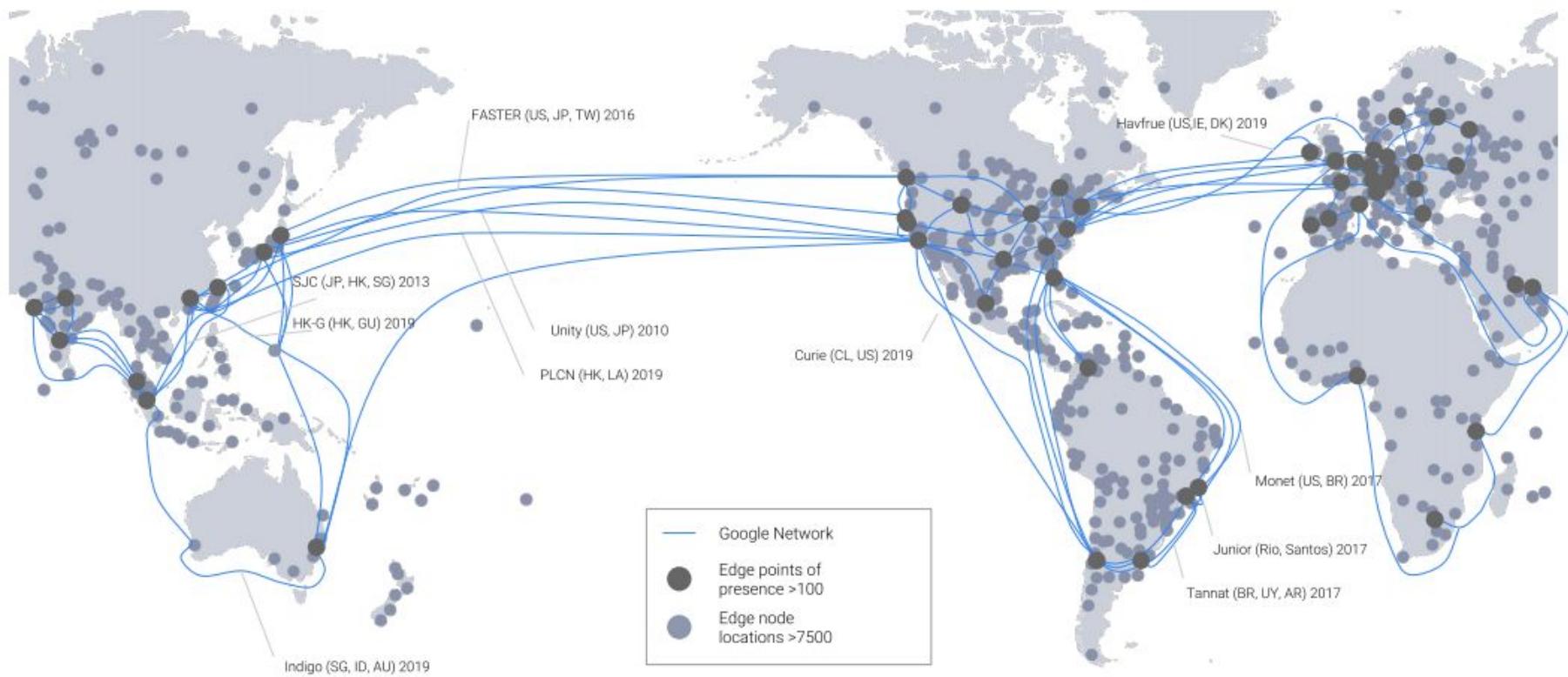
Automated  
elastic  
resources



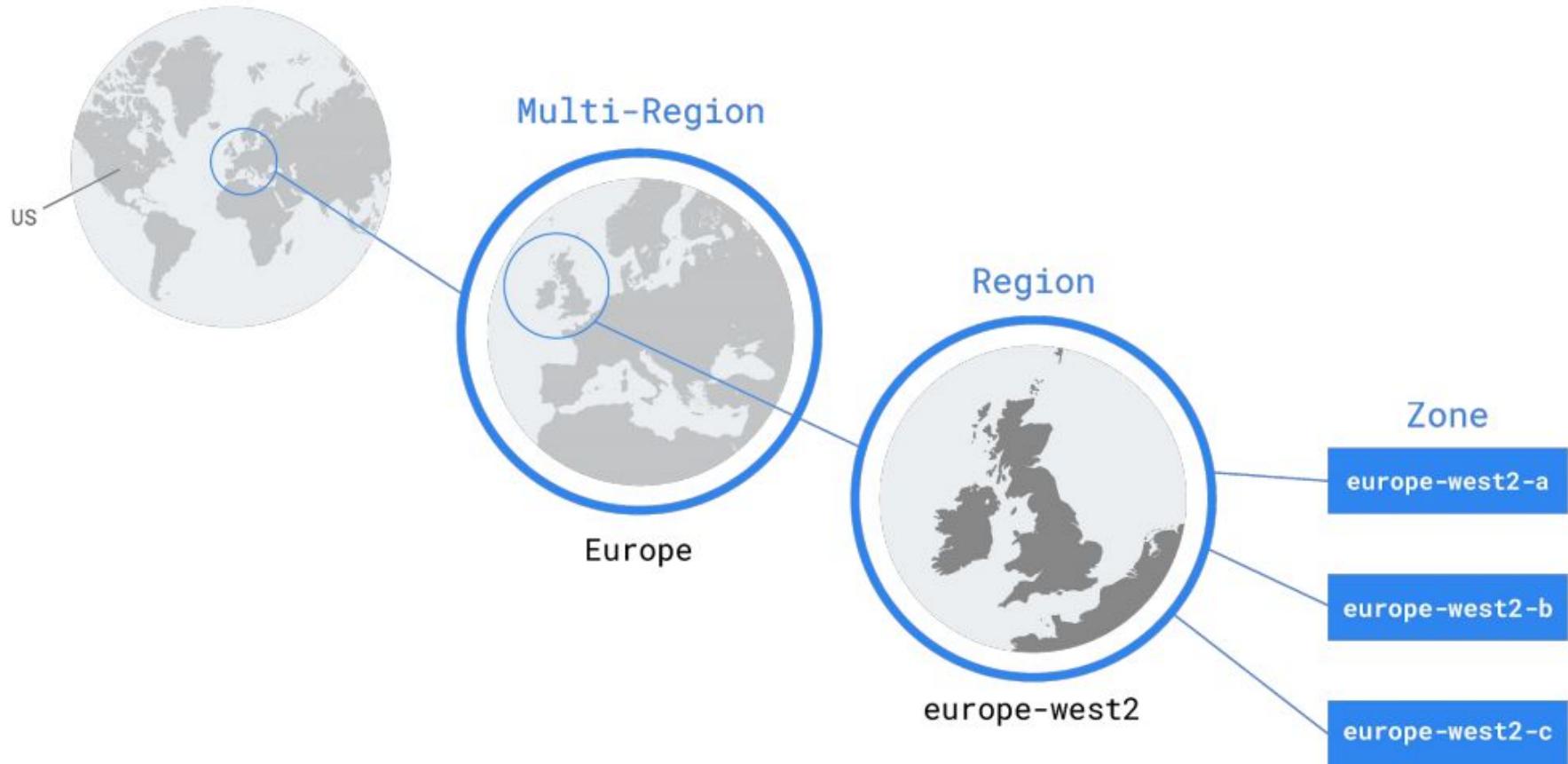
*Toward managed infrastructure*

*Toward dynamic infrastructure*

# Google network: 100,000s of km of fiber cable, 8 subsea cables



# Google Cloud Platform is organized into regions and zones





---

# **Virtual Machines in the Cloud**

## **Google Compute Engine : GCE**

### **Agenda**

- Cloud Computing Overview
- **Using Google Cloud**
- Virtual Private Cloud (VPC) Network
- Compute Engine
- Quiz and lab

# There are four ways to interact with Google Cloud

## Cloud Console

Web user interface



## Cloud Shell and Cloud SDK

Command-line  
interface



## REST-based API

For custom  
applications



## Cloud Mobile App

For iOS and Android



# Cloud Console, Cloud SDK and Cloud Shell

Cloud Console



console.cloud.google.com

	Name ^	Zone	Internal IP	External IP	Connect	
<input type="checkbox"/>	nginxstack-1	us-central1-f	10.128.0.3 (nic0)	35.238.84.245	SSH ▾	⋮
<input type="checkbox"/>	nginxstack-2	us-central1-f	10.128.0.4 (nic0)	35.225.177.18	SSH ▾	⋮
<input type="checkbox"/>	nginxstack-3	us-central1-f	10.128.0.2 (nic0)	35.239.250.238	SSH ▾	⋮

Cloud Shell

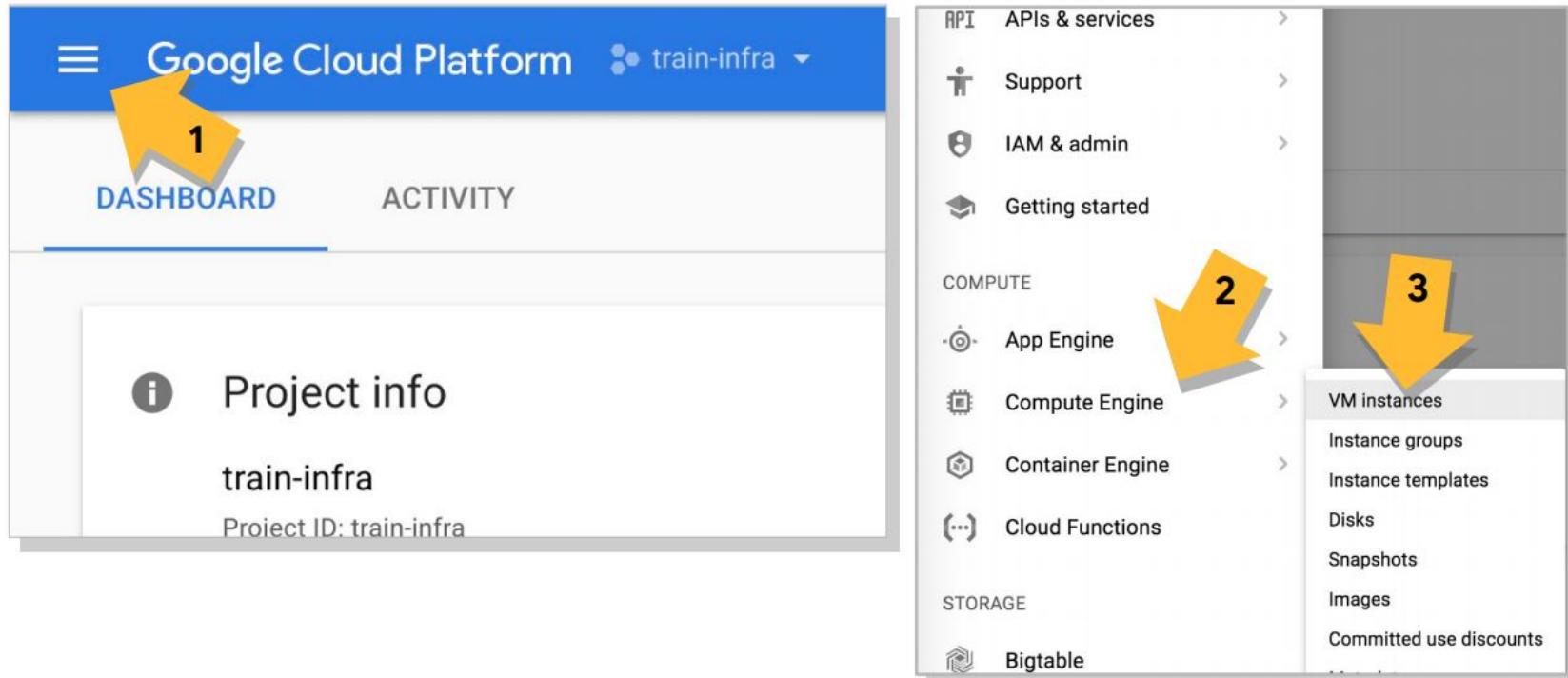
Google Cloud SDK

```
$ gcloud compute instances list
```

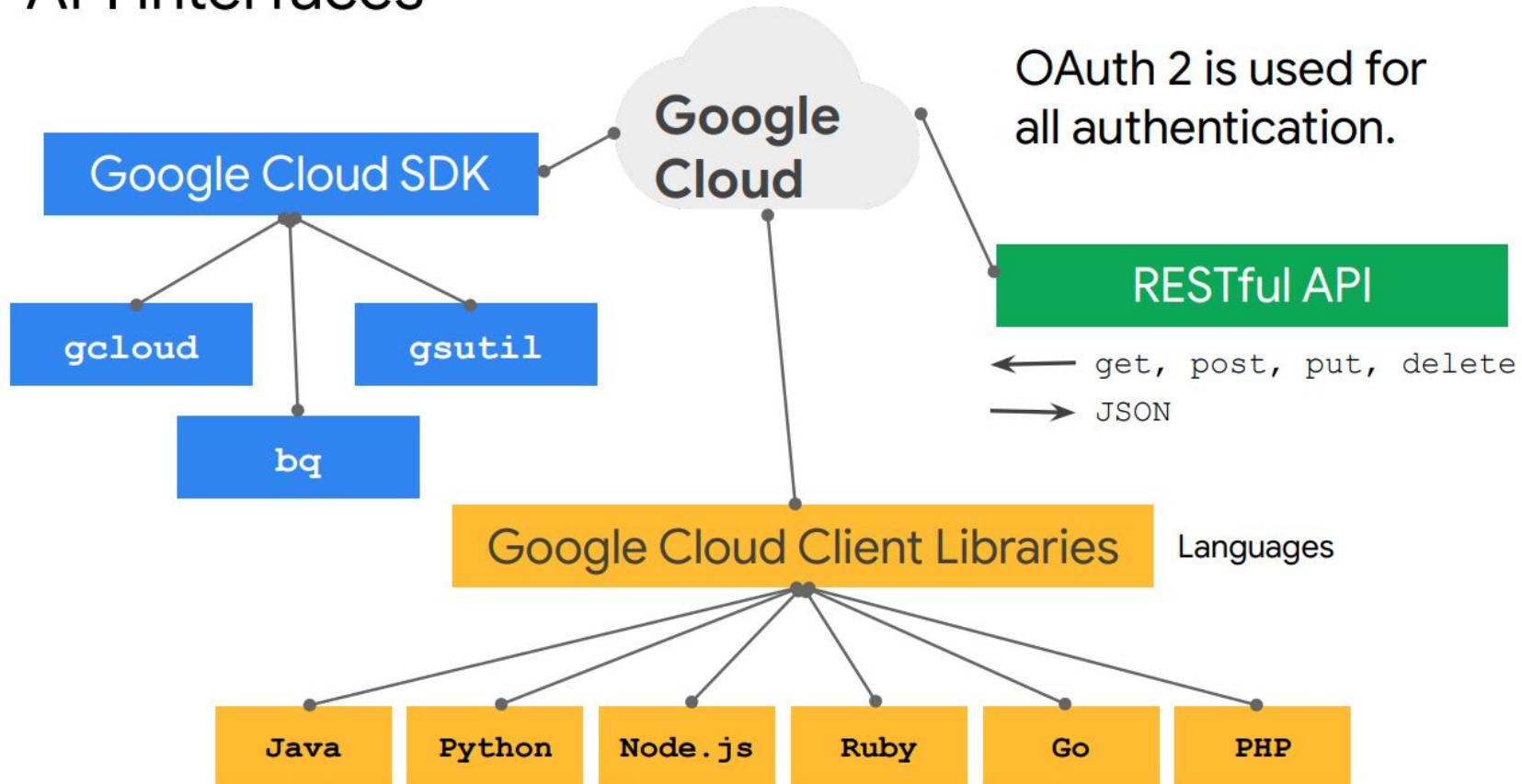
NAME	ZONE	INTERNAL_IP	EXTERNAL_IP
nginxstack-1	us-central1-f	10.128.0.3	35.238.84.245
nginxstack-2	us-central1-f	10.128.0.4	35.225.177.18
nginxstack-3	us-central1-f	10.128.0.2	35.239.250.238

# How to interpret lab instructions

“On the Navigation menu, click **Compute Engine > VM instances**”

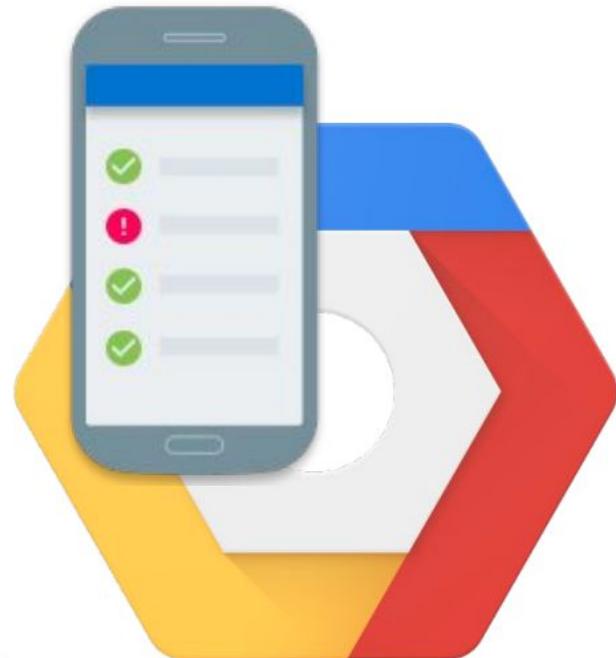


# API interfaces



# Cloud Mobile App

- Manage virtual machines and database instances.
- Manage apps in App Engine.
- Manage your billing.
- Visualize your projects with a customizable dashboard.



---

# **Virtual Machines in the Cloud**

## **Google Compute Engine : GCE**

### **Agenda**

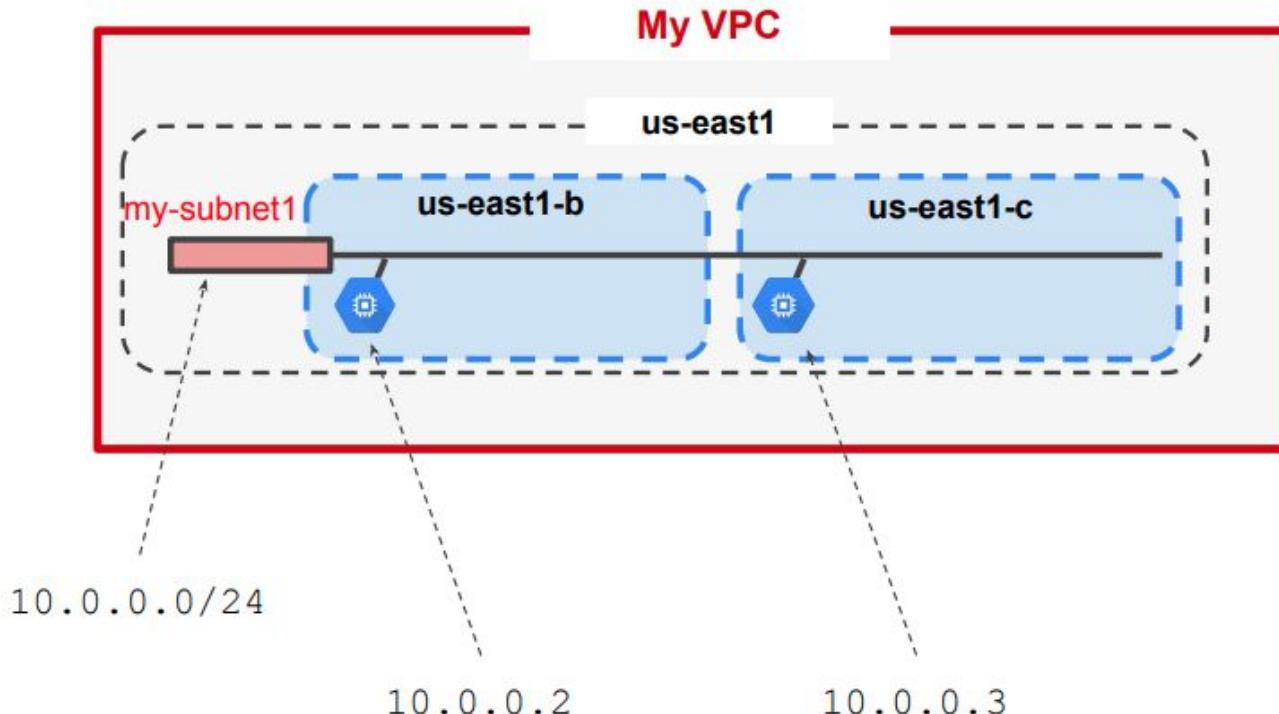
- Cloud Computing Overview
- Using Google Cloud
- **Virtual Private Cloud (VPC) Network**
- Compute Engine
- Quiz and lab

## Virtual Private Cloud Networking

- Each VPC network is contained in a GCP project.
- You can provision Cloud Platform resources, connect them to each other, and isolate them from one another.



# Google Cloud VPC networks are global; subnets are regional



---



# **Virtual Machines in the Cloud**

## **Google Compute Engine : GCE**

### **Agenda**

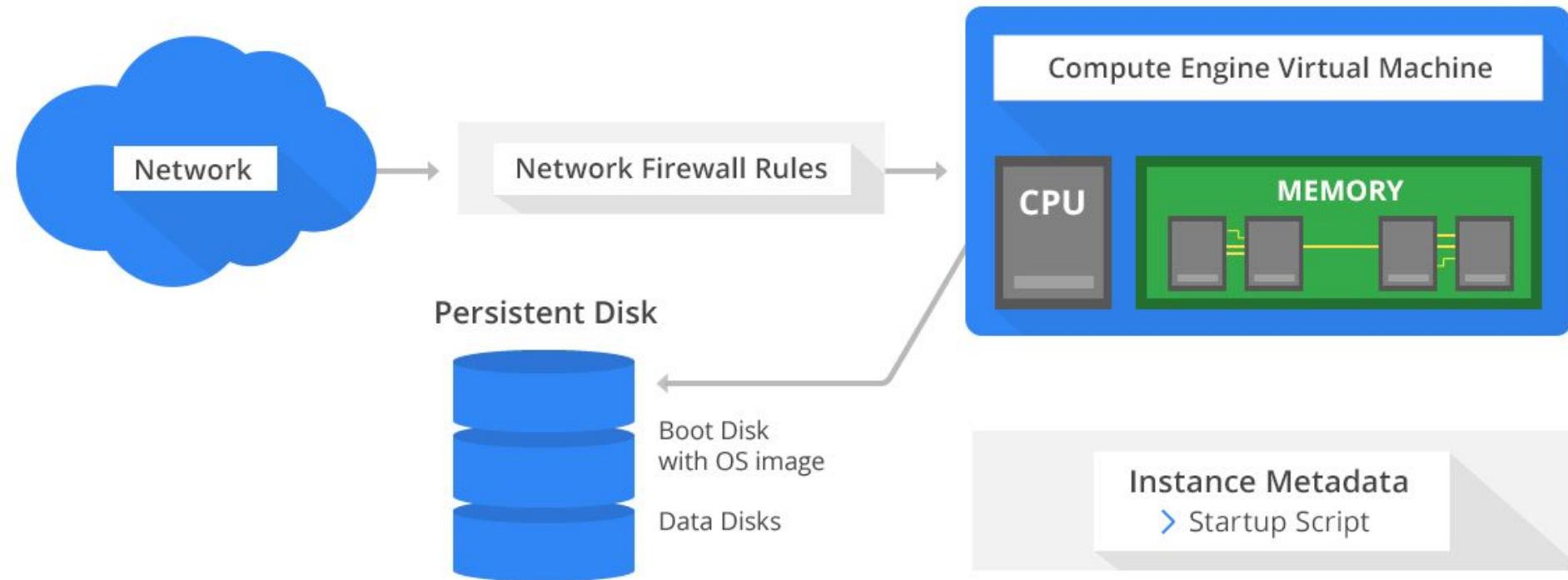
- Virtual Private Cloud (VPC) Network
- Compute Engine
- Quiz and lab

# Google Cloud compute and processing options

					
	Compute Engine	Kubernetes Engine	App Engine Standard	App Engine Flexible	Cloud Functions
Language support	Any	Any	Python Node.js Go Java PHP	Python Node.js Go Java PHP Ruby .NET Custom Runtimes	Python Node.js Go
Usage model	IaaS	IaaS PaaS	PaaS	PaaS	Microservices Architecture
Scaling	Server Autoscaling	Cluster	Autoscaling managed servers		Serverless
Primary use case	General Workloads	Container Workloads	Scalable web applications Mobile backend applications		Lightweight Event Actions

# Google Compute Engine

---



# Compute Engine

Infrastructure as a Service (IaaS)

Predefined or custom machine types:

- vCPUs (cores) and Memory (RAM)
- Persistent disks: HDD, SSD, and Local SSD
- Networking
- Linux or Windows



Compute Engine

# Compute Engine features

- Instance metadata
- Startup scripts

## Machine rightsizing

- Recommendation engine for optimum machine size
- Stackdriver statistics
- New recommendation 24 hrs after VM create or resize

## Global load balancing:

- Multiple regions for availability



## Availability policies:

- Live migrate
- Auto restart

- Per-second billing
- Sustained use discounts
- Committed use discounts

## Preemptible:

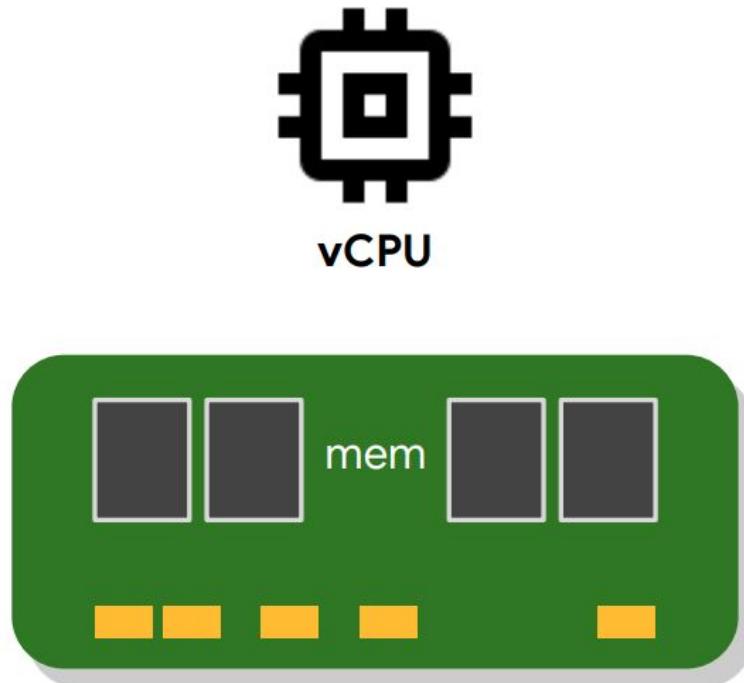
- Up to 80% discount
- No SLA

# Compute

Several machine types

- Network throughput scales 2 Gbps per vCPU (small exceptions)
- Theoretical max of 32 Gbps with 16 vCPU or 100 Gbps with T4 or V100 GPUs

A vCPU is equal to 1 hardware hyper-thread



# Storage

## Disks

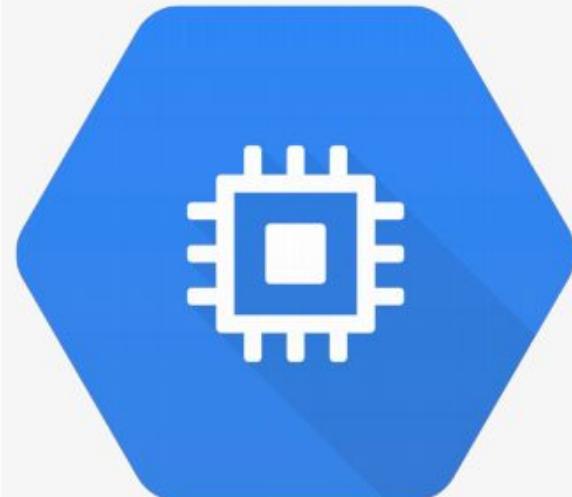
- Standard, SSD, or Local SSD
- Standard and SSD PDs scale in performance for each GB of space allocated

Resize disks or migrate instances with no downtime



Compute Engine offers managed virtual machines

- High CPU, high memory, standard and shared-core machine types
- Persistent disks
  - Standard, SSD, local SSD
  - Snapshots
- Resize disks with no downtime
- Instance metadata and startup scripts





## Break 1

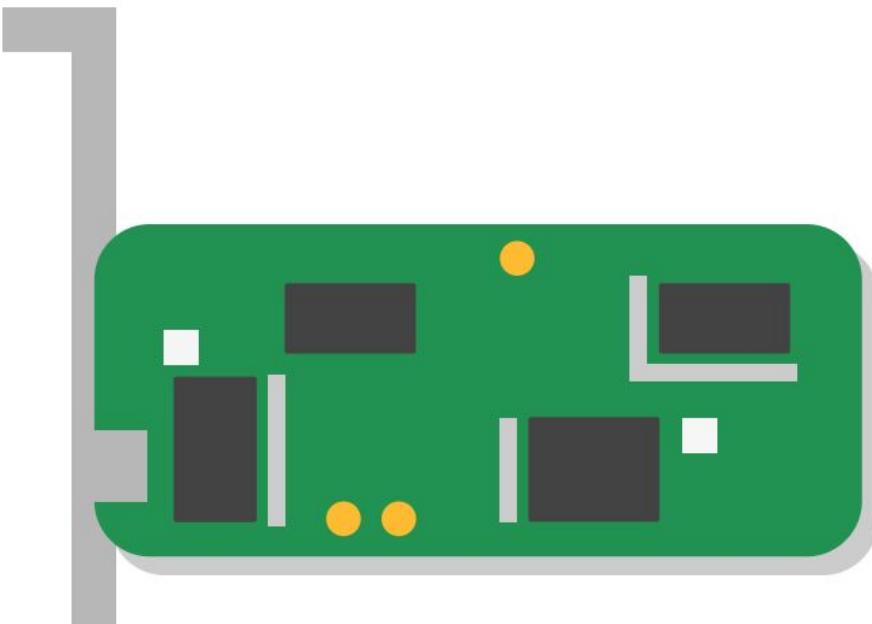
5 mins.

Break

# Networking

## Robust networking features

- Default, custom networks
- Inbound/outbound firewall rules
  - IP based
  - Instance/group tags
- Regional HTTPS load balancing
- Network load balancing
  - Does not require pre-warming
- Global and multi-regional subnetworks



# VM access

## Linux: SSH

- SSH from Cloud Console or CloudShell via Cloud SDK
- SSH from computer or third-party client and generate key pair
- Requires firewall rule to allow tcp:22

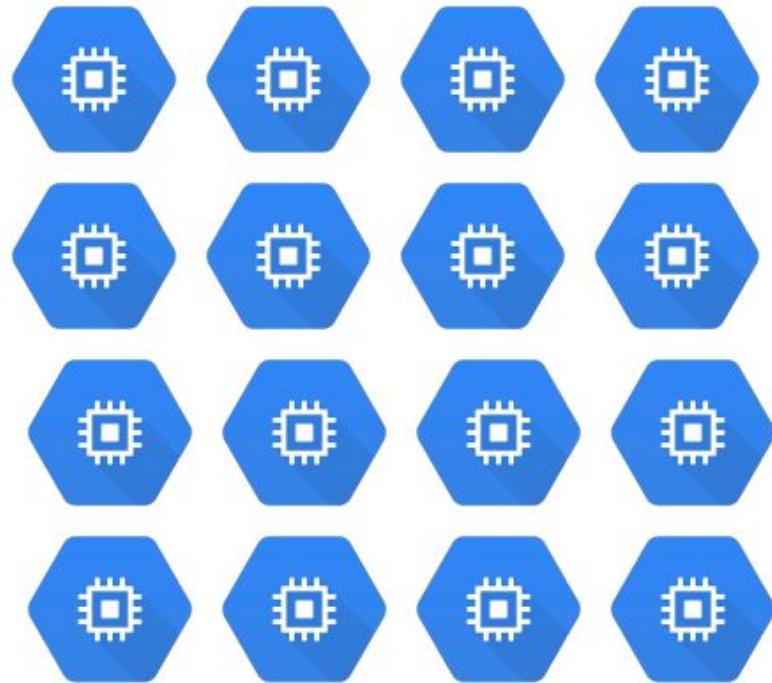
## Windows: RDP

- RDP clients
- Powershell terminal
- Requires setting the Windows password
- Requires firewall rule to allow tcp:3389

## Scale up or scale out with Compute Engine

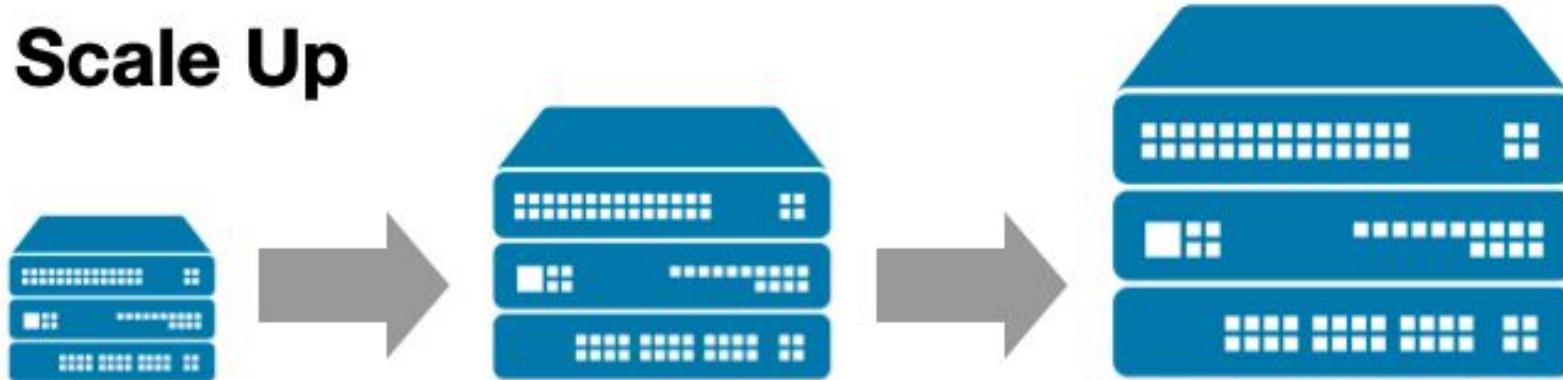


*Use big VMs for memory- and compute-intensive applications*



*Use Autoscaling for resilient, scalable applications*

## Scale Up



## Scale Out



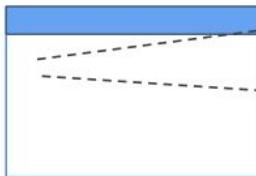
---

## Compute Option ( vCPU and Memory )

- Create a VM
- Machine Type
- Region and Zone
- Pricing
- Sustained Discount
- Preemptible

# Creating a VM

1



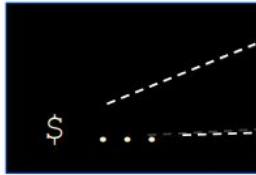
VM instances

CREATE INSTANCE

IMPORT VM

console.google.com

2



\$ gcloud compute instances create [instance-name]

command line  
*including Cloudshell*

3

REST API

Many VM options

- Project
- Region
- Zone
- Subnetwork
- Machine type
- Disk options
- Image
- IP options

# Machine types

Predefined machine types: Ratio of GB of memory per vCPU

- Standard
- High-memory
- High-CPU
- Memory-optimized
- Compute-optimized
- Shared-core

Custom machine types:

- You specify the amount of memory and number of vCPUs.

# Standard machine types

Machine name	vCPUs	Memory (GB)	Max # PD	Max total PD size
n1-standard-1	1	3.75	128	257 TB
n1-standard-2	2	7.50		
n1-standard-4	4	15		
n1-standard-8	8	30		
n1-standard-16	16	60		
n1-standard-32	32	120		
n1-standard-64	64	240		
n1-standard-96	96	360		

3.75 GB of memory

1 vCPU

# Shared-core machine types

Machine name	vCPUs	Memory (GB)	Max # PD	Max total PD size
f1-micro	0.2	0.60	16	3 TB
g1-small	0.5	1.70		

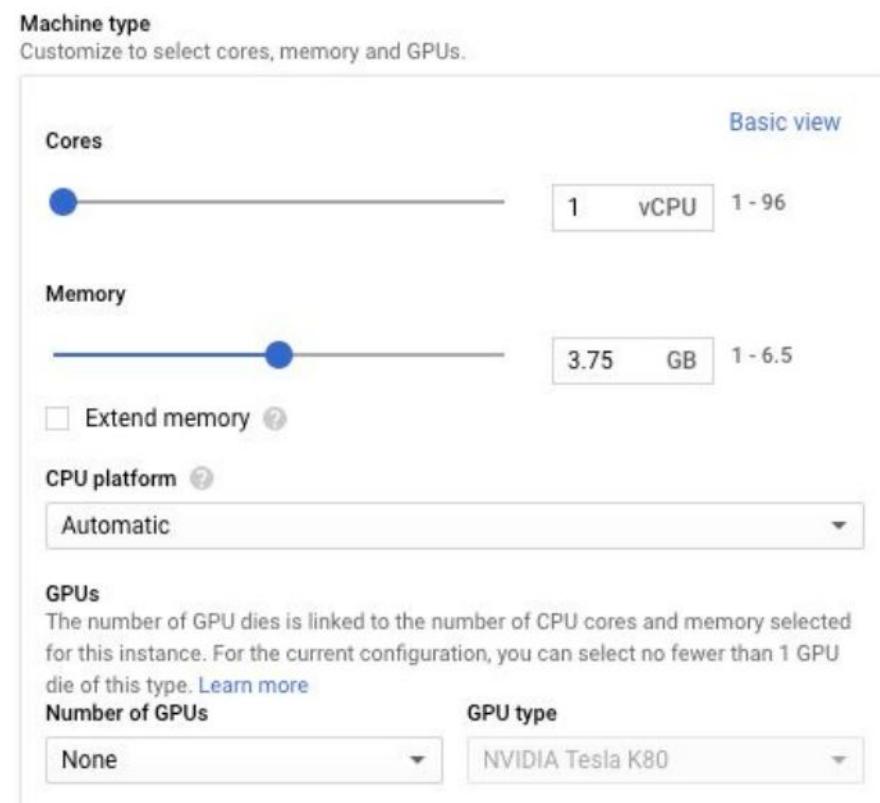
# Creating custom machine types

When to select custom:

- Requirements fit between the predefined types
- Need more memory or more CPU

Customize the amount of memory and vCPU for your machine:

- Either 1 vCPU or even number of vCPU
- 0.9 GB per vCPU, up to 6.5 GB per vCPU (default)
- Total memory must be multiple of 256 MB



# Choosing region and zone

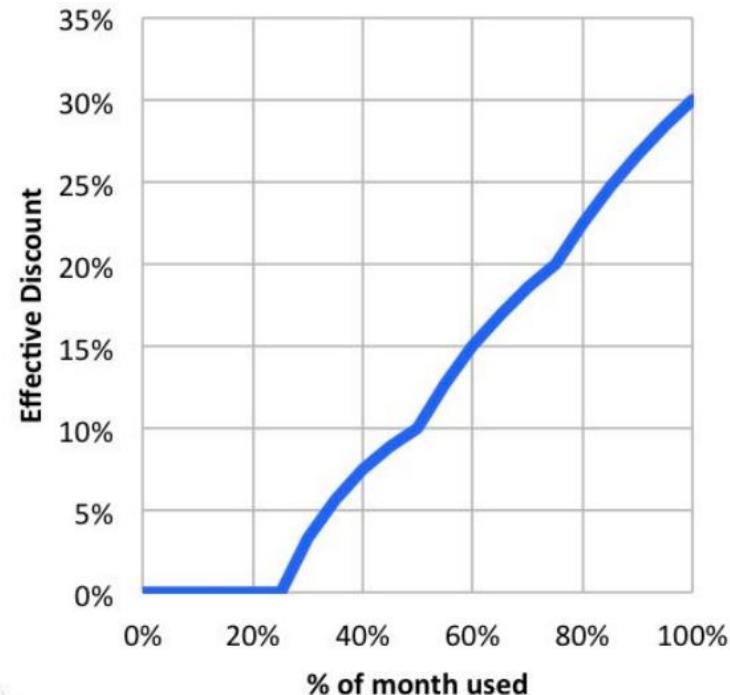


# Pricing

- Per-second billing, with minimum of 1 minute
  - vCPUs, GPUs, and GB of memory
- Resource-based pricing:
  - Each vCPU and each GB of memory is billed separately
- Discounts:
  - Sustained use
  - Committed use
  - Preemptible VM instances
- Recommendation Engine
  - Notifies you of underutilized instances
- Free usage limits

## Sustained use discounts

Usage Level (% of month)	% at which incremental is charged
0% - 25%	100% of base rate
25% - 50%	80% of base rate
50% - 75%	60% of base rate
75% - 100%	40% of base rate



Up to 30% net discount for instances that run the entire month

# Sustained use discounts (example)



Discount for 4 vCPUs

10%

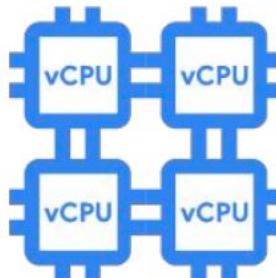
20%

30%

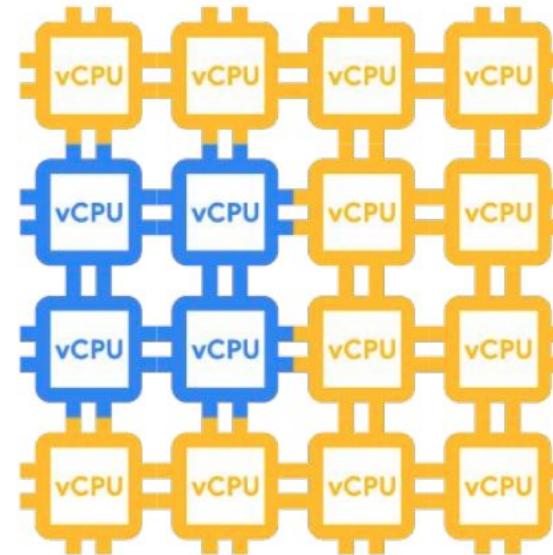
Discount for additional 12 vCPUs

10%

n1-standard-4



n1-standard-4+12



Week

# Preemptible

- Lower price for interruptible service (up to 80%)
- VM might be terminated at any time
  - No charge if terminated in the first minute
  - 24 hours max
  - 30-second terminate warning, but not guaranteed
    - *Time for a shutdown script*
- No live migrate; no auto restart
- You can request that CPU quota for a region be split between regular and preemption
  - Default: preemptible VMs count against region CPU quota



# Image

- What's in an image?
- Images

# What's in an image?

- Boot loader
- Operating system
- File system structure
- Software
- Customizations



# Images

- Public base images
  - Google, third-party vendors, and community; Premium images (p)
  - Linux
    - CentOS, CoreOS, Debian, RHEL(p), SUSE(p), Ubuntu, openSUSE, and FreeBSD
  - Windows
    - Windows Server 2019(p), 2016(p), 2012-r2(p)
    - SQL Server pre-installed on Windows(p)
- Custom images
  - Create new image from VM: pre-configured and installed SW
  - Import from on-prem, workstation, or another cloud
  - Management features: image sharing, image family, deprecation

---

## Disk Options

- Boot disk
- Persistent disks
- Local SSD disks are physically attached to a VM
- RAM disk
- Summary of disk options

# Boot disk

- VM comes with a single root persistent disk.
- Image is loaded onto root disk during first boot:
  - Bootable: you can attach to a VM and boot from it.
  - Durable: can survive VM terminate.
- Some OS images are customized for Compute Engine.
- Can survive VM deletion if “Delete boot disk when instance is deleted” is disabled.

# Persistent disks

Network storage appearing as a block device

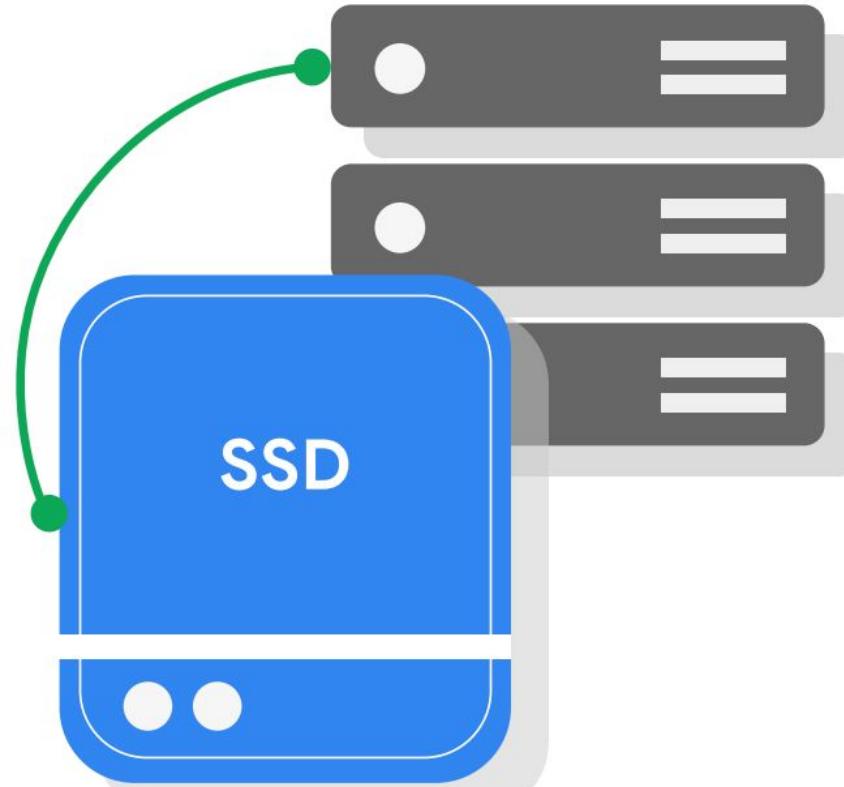
- Attached to a VM through the network interface
- Durable storage: can survive VM terminate
- Bootable: you can attach to a VM and boot from it
- Snapshots: incremental backups
- Performance: Scales with size

## Features

- HDD (magnetic) or SSD (faster, solid-state) options
- Disk resizing: even running and attached!
- Can be attached in read-only mode to multiple VMs
- Zonal or Regional
- Encryption keys:
  - Google-managed
  - Customer-managed
  - Customer-supplied

# Local SSD disks are physically attached to a VM

- More IOPS, lower latency, and higher throughput than persistent disk
- 375-GB disk up to eight, total of 3 TB
- Data survives a reset, but not a VM stop or terminate
- VM-specific: cannot be reattached to a different VM



# RAM disk

- tmpfs
- Faster than local disk, slower than memory
  - Use when your application expects a file system structure and cannot directly store its data in memory
  - Fast scratch disk, or fast cache
- Very volatile; erase on stop or restart
- May need a larger machine type if RAM was sized for the application
- Consider using a persistent disk to back up RAM disk data

# Summary of disk options

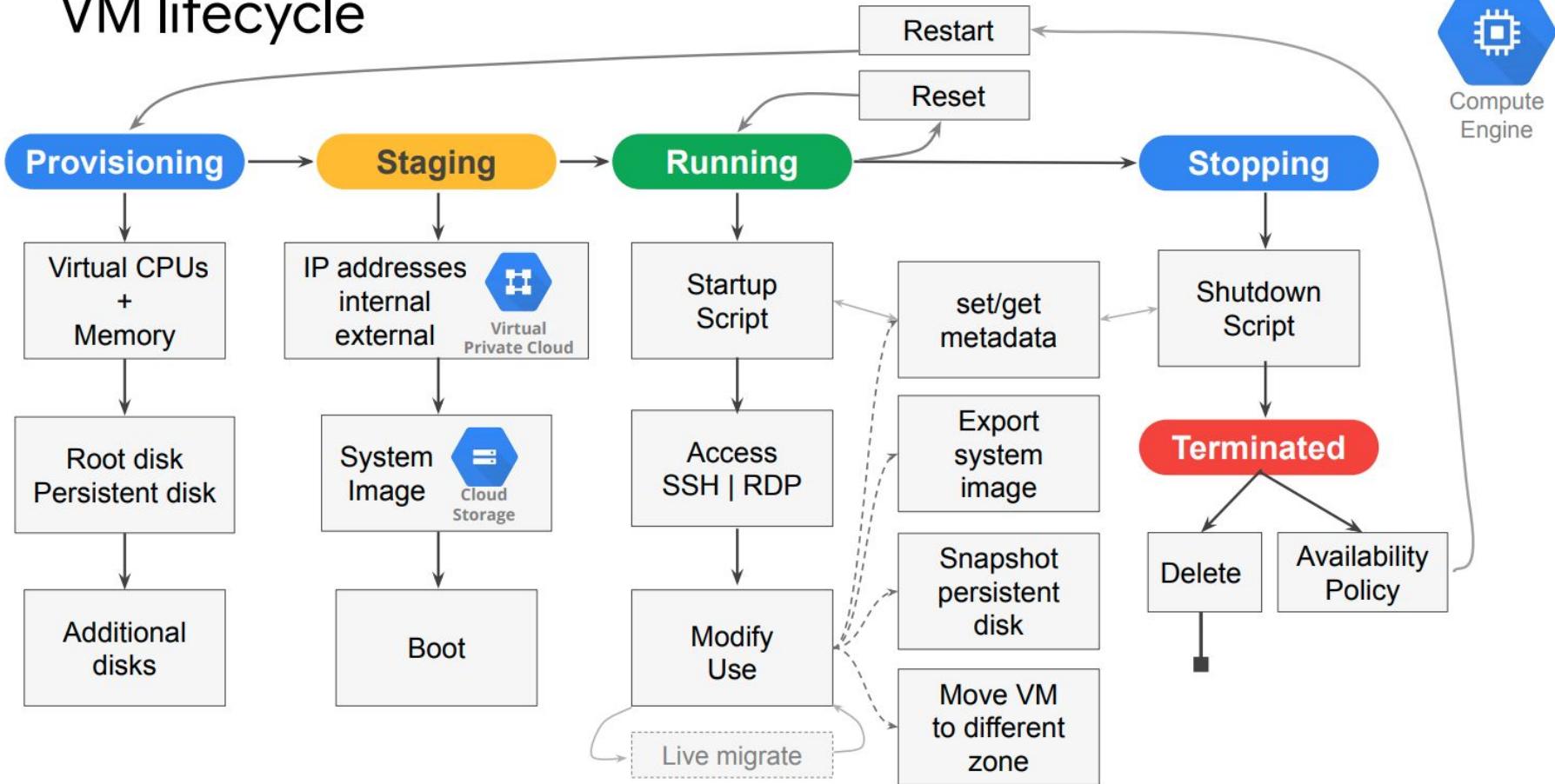
	Persistent disk HDD	Persistent disk SSD	Local SSD disk	RAM disk
Data redundancy	Yes	Yes	No	No
Encryption at rest	Yes	Yes	Yes	N/A
Snapshotting	Yes	Yes	No	No
Bootable	Yes	Yes	No	Not
Use case	General, bulk file storage	Very random IOPS	High IOPS and low latency	low latency and risk of data loss

---

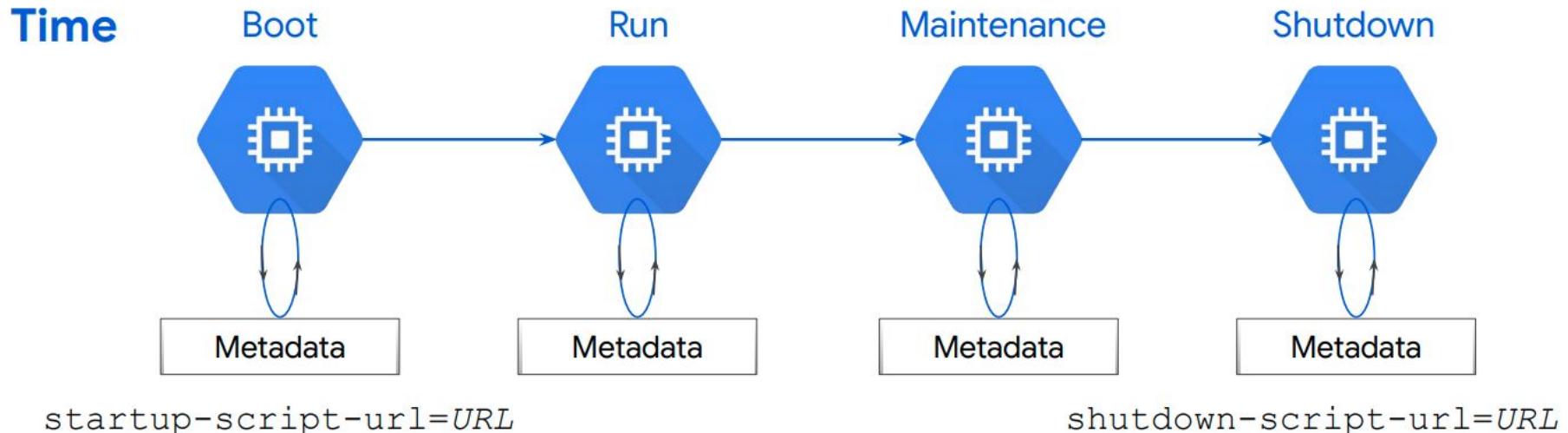
## Common Compute Engine Action

- Metadata and script
- Move an instance to a new zone
- Snapshot: Backup critical data
- Snapshot: Migrate data between zones
- Snapshot: Transfer to SSD to improve performance
- Resize persistent disk

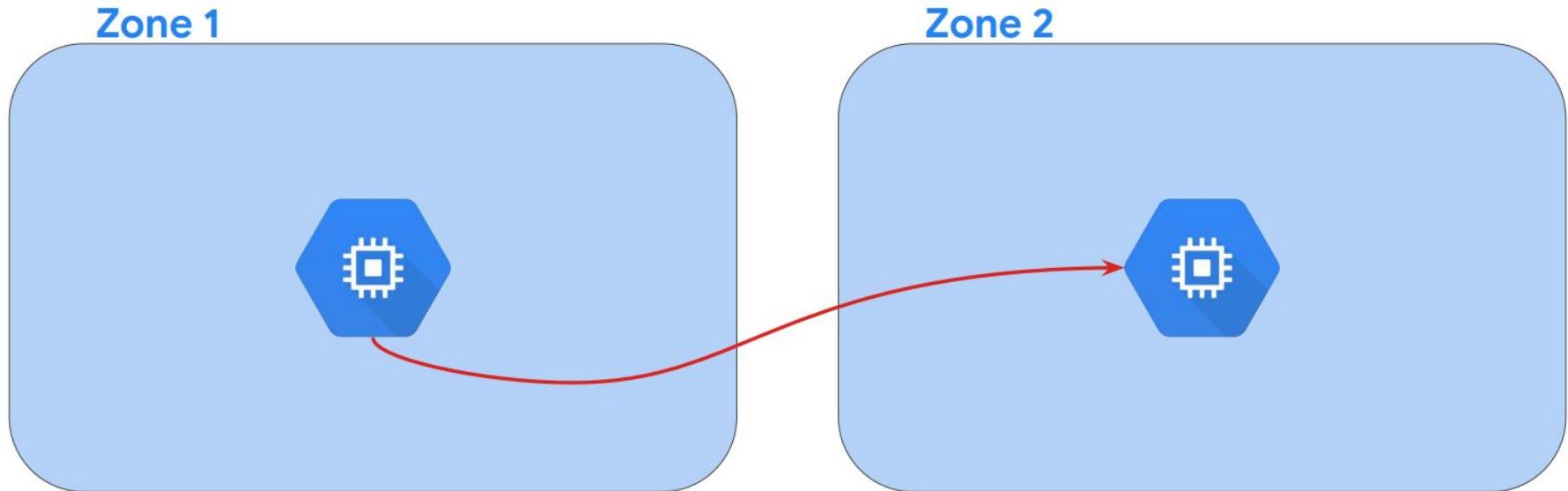
# VM lifecycle



# Metadata and scripts



## Move an instance to a new zone

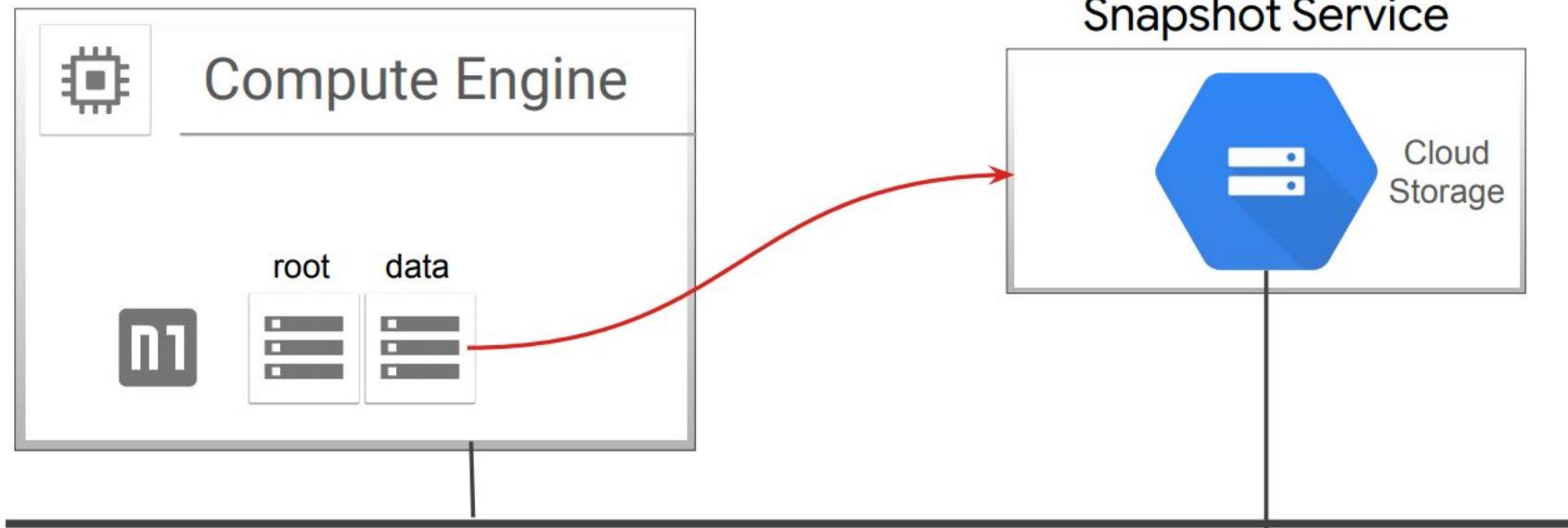


`gcloud compute instances move`

# Move an instance to a new zone

- Automated process (moving within region):
  - `gcloud compute instances move`
  - Update references to VM; not automatic
- Manual process (moving between regions):
  - Snapshot all persistent disks on the source VM.
  - Create new persistent disks in destination zone restored from snapshots.
  - Create new VM in the destination zone and attach new persistent disks.
  - Assign static IP to new VM.
  - Update references to VM.
  - Delete the snapshots, original disks, and original VM.

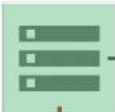
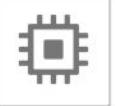
# Snapshot: Back up critical data



# Snapshot: Migrate data between zones

Zone 1

Compute Engine



Zone 2

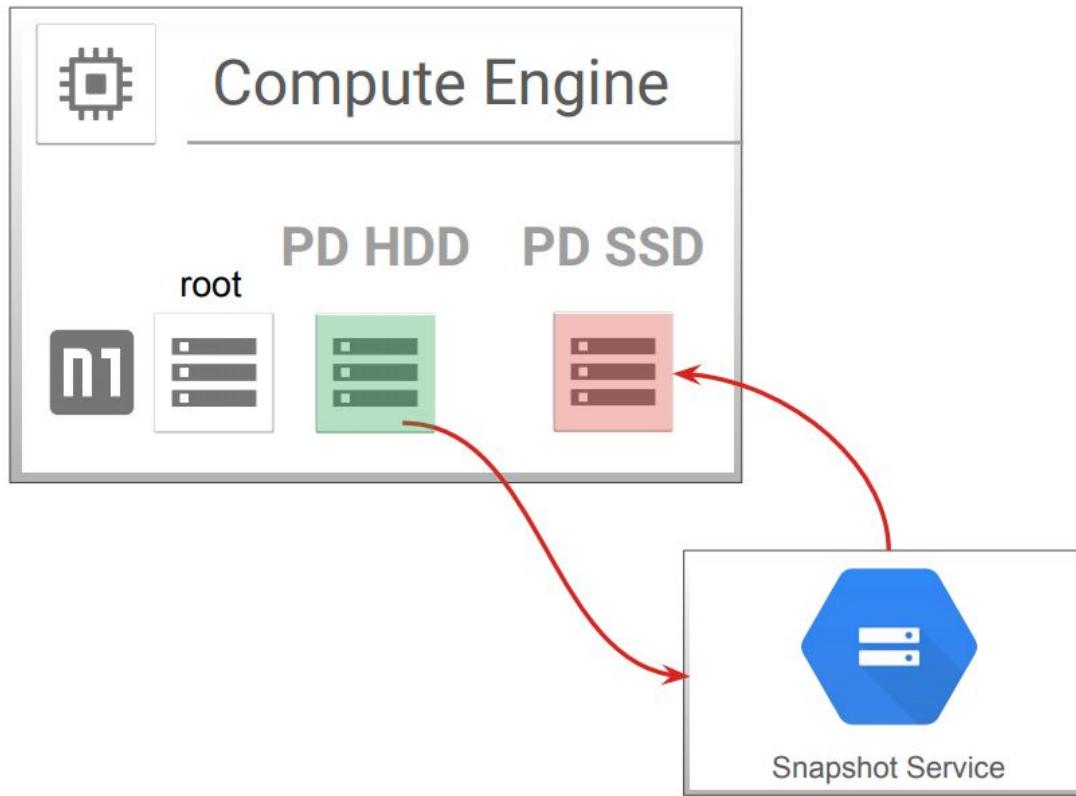
Compute Engine



Snapshot Service



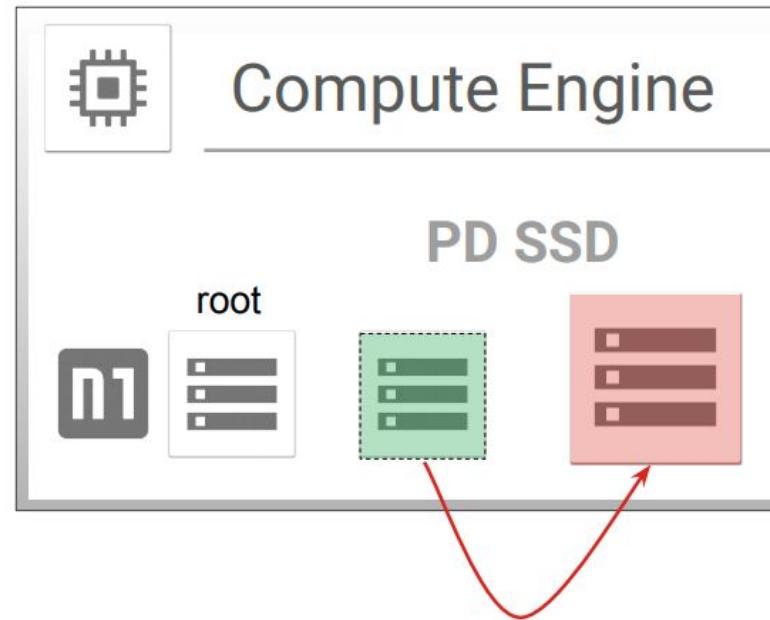
# Snapshot: Transfer to SSD to improve performance



# Persistent disk snapshots

- Snapshot is not available for local SSD.
- Creates an *incremental* backup to Cloud Storage.
  - Not visible in *your* buckets; managed by the snapshot service.
  - Consider cron jobs for periodic incremental backup.
- Snapshots can be restored to a new persistent disk.
  - New disk can be in another region or zone in the same project.
  - Basis of VM migration: "moving" a VM to a new zone.
    - Snapshot doesn't back up VM metadata, tags, etc.

# Resize persistent disk



You can grow disks, but never shrink them!



## Break 2

5 mins.

Break

# Pricing Calculation

<https://cloud.google.com/products/calculator>

The screenshot shows the Google Cloud Pricing Calculator interface. At the top, there's a blue header bar with the title "Google Cloud Pricing Calculator" on the left and a note "Prices are up to date. Last update: 18-January-2021" on the right. Below the header is a navigation bar with icons for various services: COMPUTE ENGINE, APP ENGINE, KUBERNETES ENGINE, CLOUD RUN, VMWARE ENGINE, CLOUD STORAGE, NETWORKING EGRESS, CLOUD LOAD INT BALANCING, and more. An "Estimate" button is located on the right side of this bar. Below the navigation bar is a search bar with the placeholder "Search for a product you are interested in." To the right of the search bar is a magnifying glass icon. The main body of the calculator is titled "Instances". It contains several input fields with dropdown menus and question mark icons for help: "Number of instances \*", "What are these instances for?", "Operating System / Software" (with a note "Free: Debian, CentOS, CoreOS, Ubuntu, or other User Provided OS"), "Machine Class" (set to "Regular"), "Machine Family" (set to "General purpose"), "Series" (set to "E2"), and "Machine type".

---

# Creating a Virtual Machine

<https://google.qwiklabs.com/focuses/3563?parent=catalog>

40 minutes

1 Credit



Rate Lab

GSP001



Google Cloud Self-Paced Labs

# Compute Engine: Qwik Start - Windows

40 minutes

1 Credit



<https://google.qwiklabs.com/focuses/560?parent=catalog>

**GSP093**



Google Cloud Self-Paced Labs