

GOOGLE CLOUD PLATFORM

# COMPUTE ENGINE

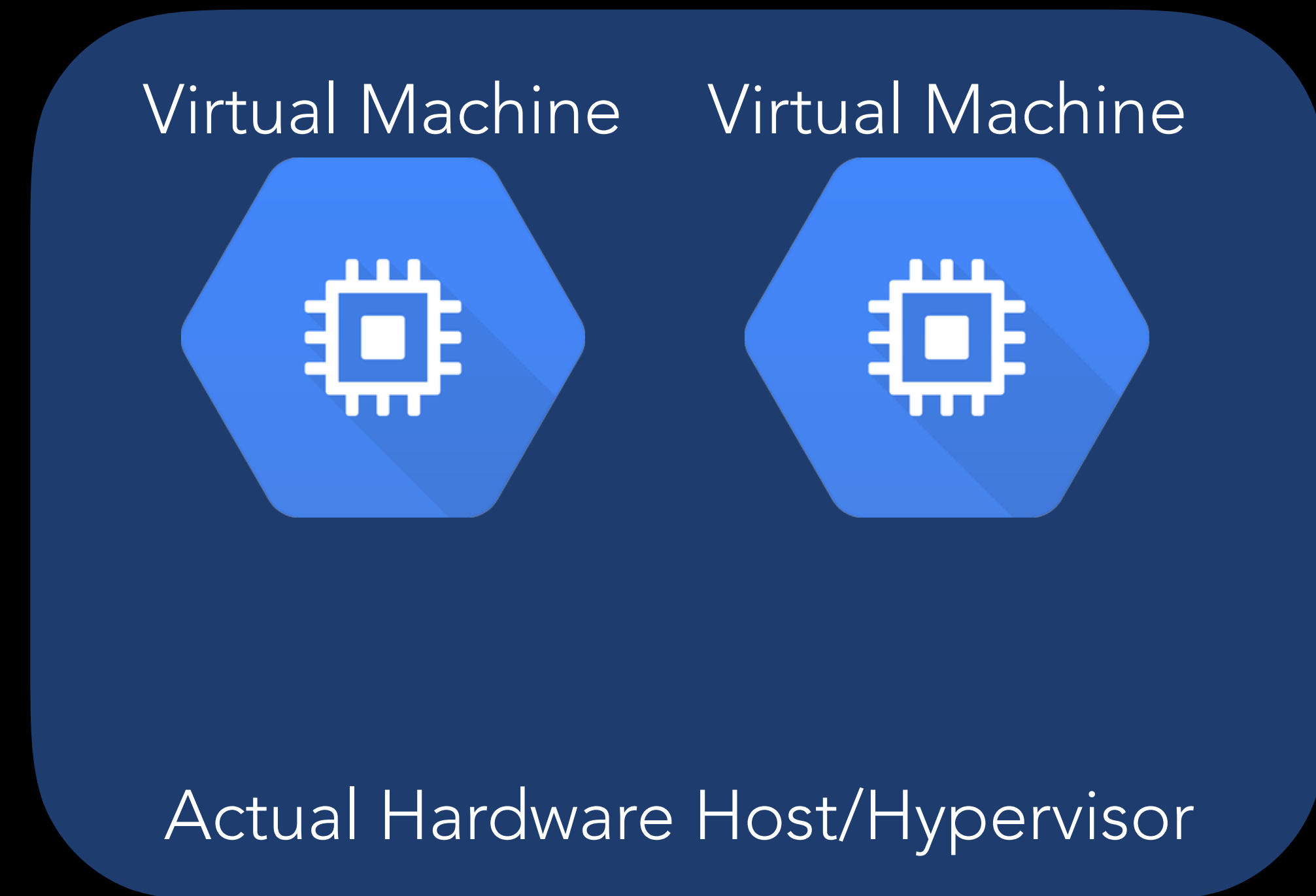






# COMPUTE ENGINE = VIRTUAL MACHINES (VM)

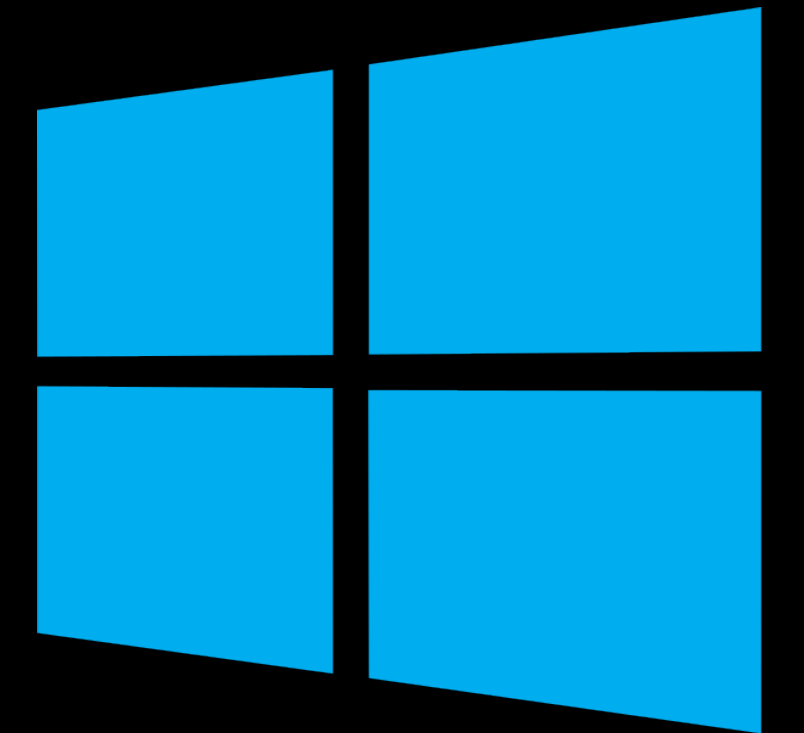
- Compute Engine is Google's Virtual Machine offering (similar to Amazon's EC2)
- Typically billed by the second based on how much CPU & RAM is used
  - "vCPU" is the unit of measure of CPU - equivalent to a single hardware hyper-thread on the underlying hardware (you have your choice)
  - RAM is measured in Gigabytes
  - Committed Use & Sustained Use Discounts
  - You'll also pay for attached block storage (disks) - but those aren't tied to the machine definition as closely and exist separately from the machine





# VM SOFTWARE: OPERATING SYSTEMS

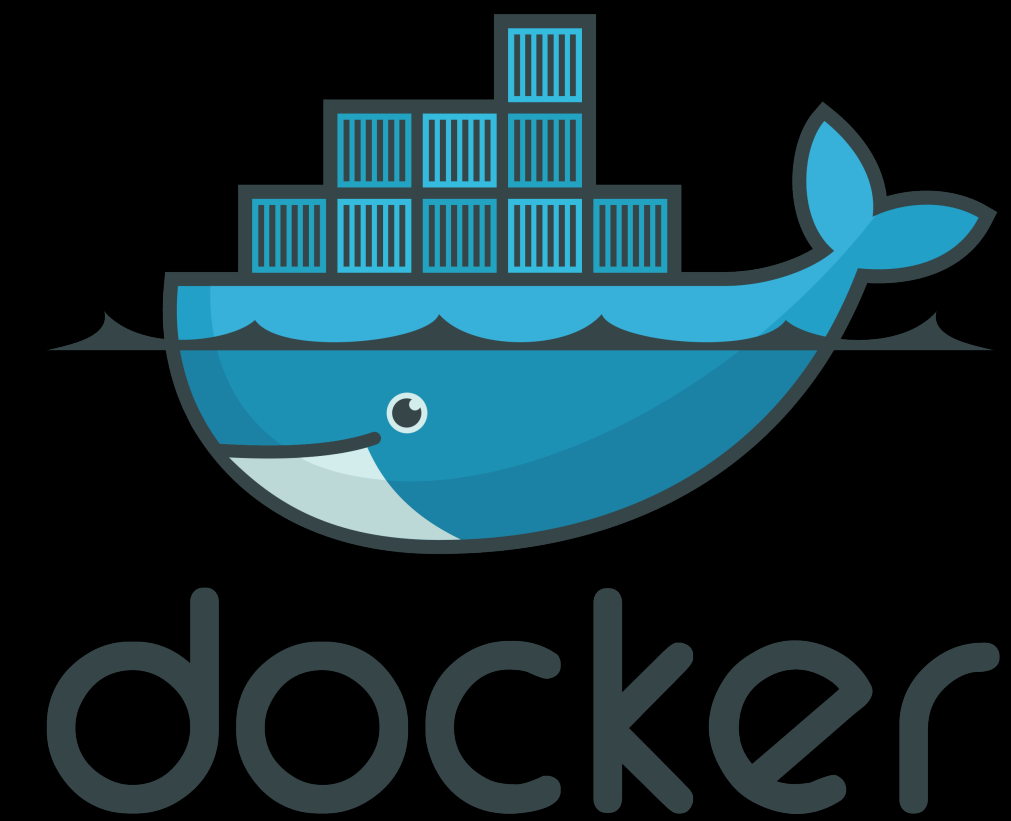
- Public Images offer hundreds of options of compatible Linux & Windows offerings/versions - both free & premium/licensed
- Third Party vendors offer free & paid images bundling OS & common applications (ERP, CRM, etc packages)
- Private Images allow you to build your own OS/ Software stack for deployment to VMs including your custom configuration





# VM SOFTWARE: CONTAINER

- Compute Engine allows you to specify a single container image (stored in a remote repository) and a configuration for that image in the consistent Docker format to run on a VM
- The VM runs a Google-maintained image of Container OS (COS) that is very minimal, ports defined in the Docker configuration are exposed
- The VM boots under COS and then starts the Docker Image
- Disks maybe attached using normal Docker Volumes

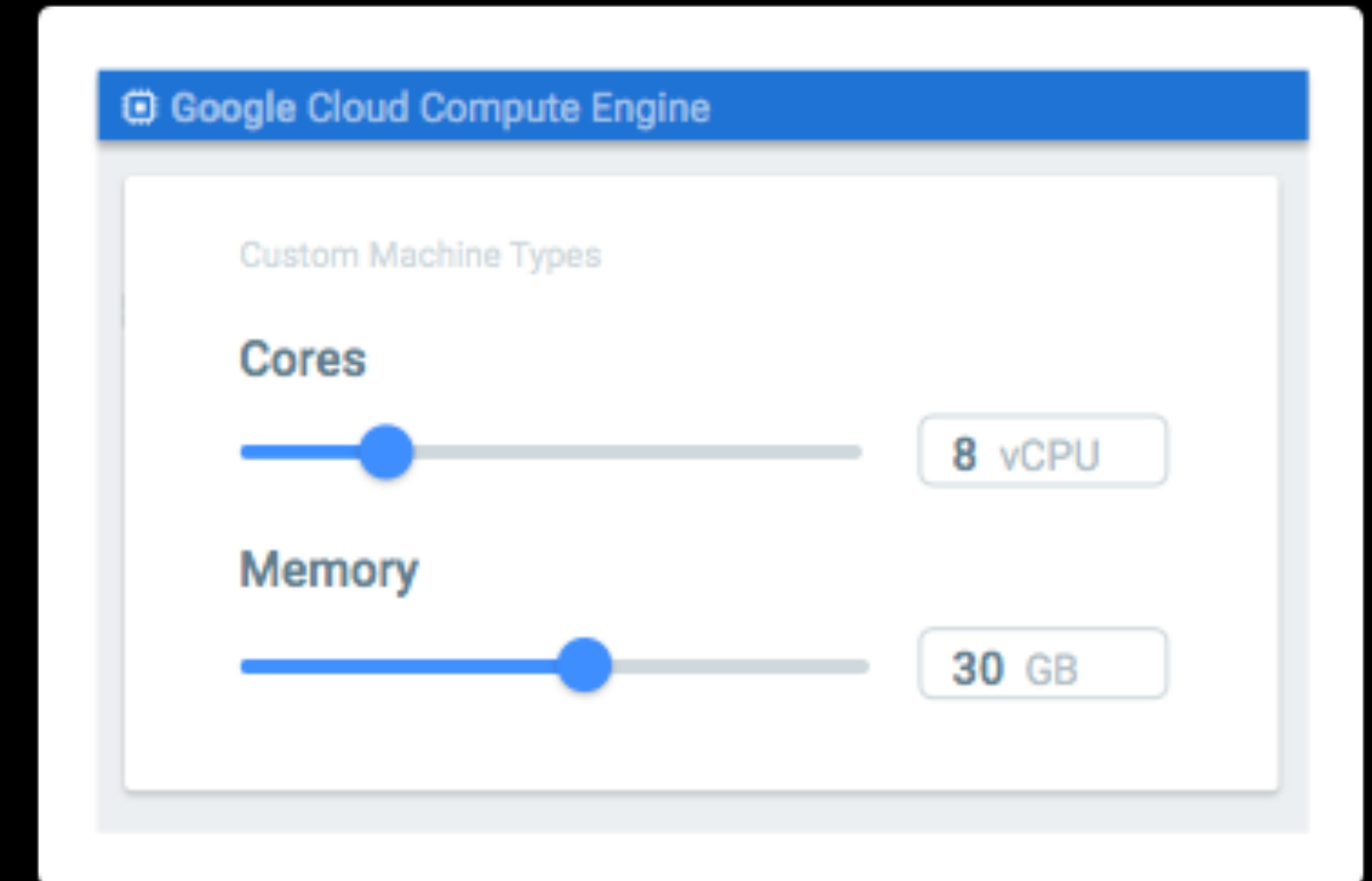




# MACHINE TYPES & SIZES

- Variety of machine types

- Standard
- High-Mem
- High-CPU
- Ultra-Mem
- Shared



- Google also offers “custom machine types”

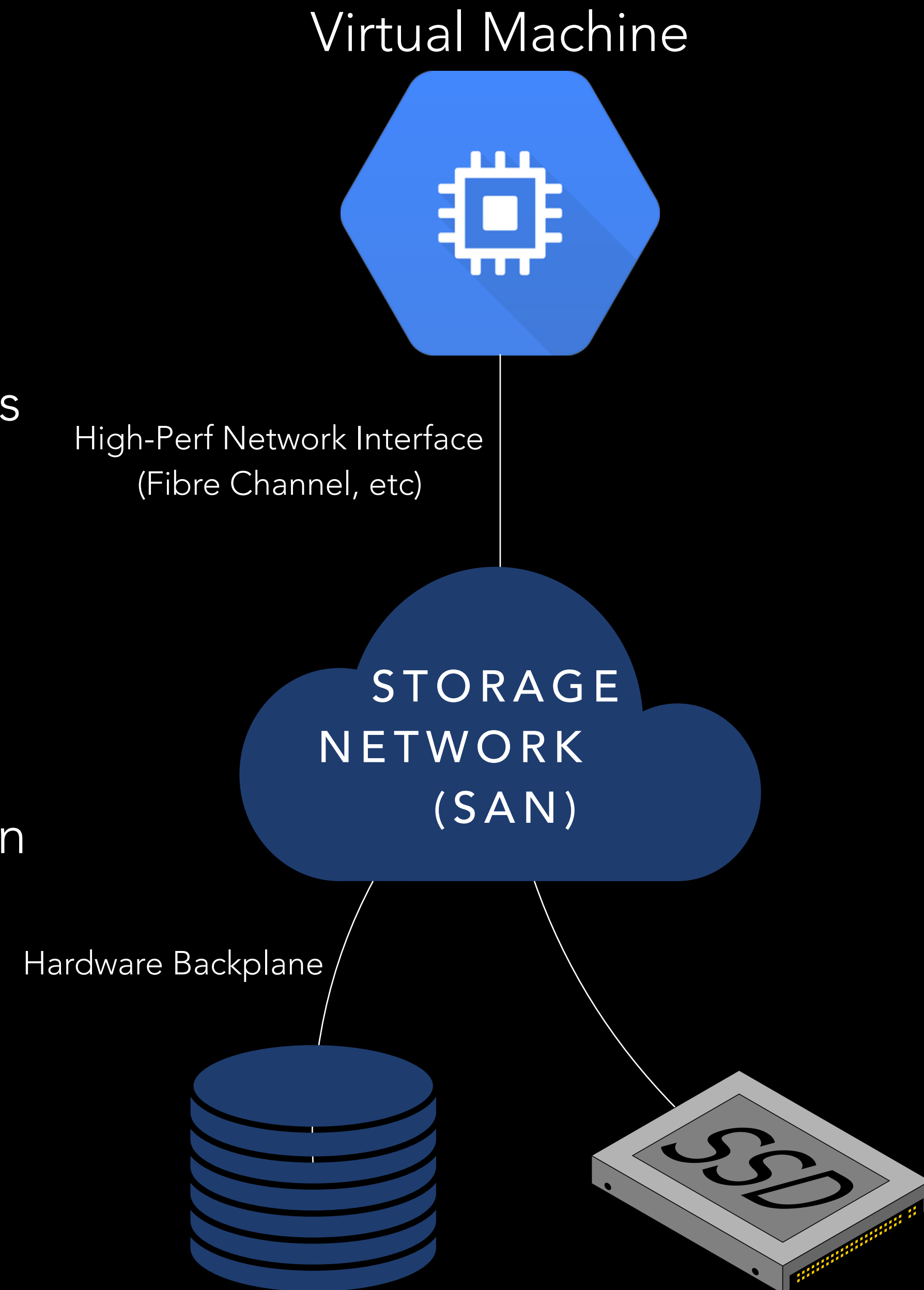
- Within certain ratios of vCPU:RAM you pick how much vCPU you need and how much RAM you need
- Slightly more expensive than pre-defined machines, but could be cheaper if your needs fall “in the middle” on either vCPU or RAM
- Consult the Pricing Calculator on Google’s website for latest information and to model your options

- Regardless of machine type - you can start & stop and change machine sizes\*



# STORAGE - PERSISTENT

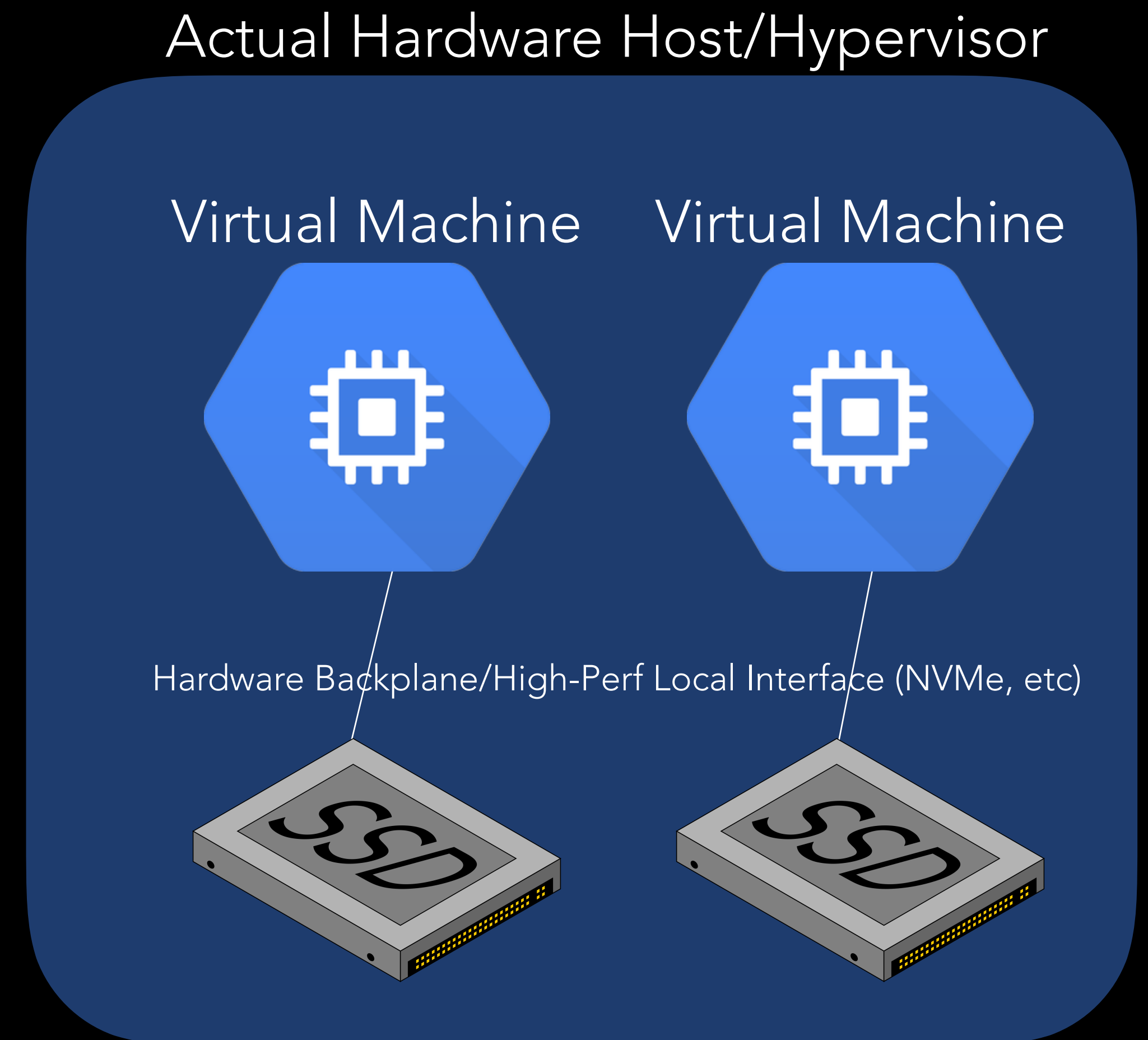
- Persistent storage is based on a SAN device that lives relatively close to the VM - highly performant
- Backed by Hard Drives (HDD, Spinning Magnetic Media) or SSDs (Solid State Disks)
  - Vary in pricing & performance
  - Both are measured by "IOPS" and "Throughput" - how many operations/sec a volume can handle and how much data they can read or write either in a "burst" or as a sustained operation
  - SSD almost always outperforms HDD, but is currently much more expensive
- Regardless of type, sold by the GB
- Supports Snapshots - for backup and replication





# STORAGE - LOCAL

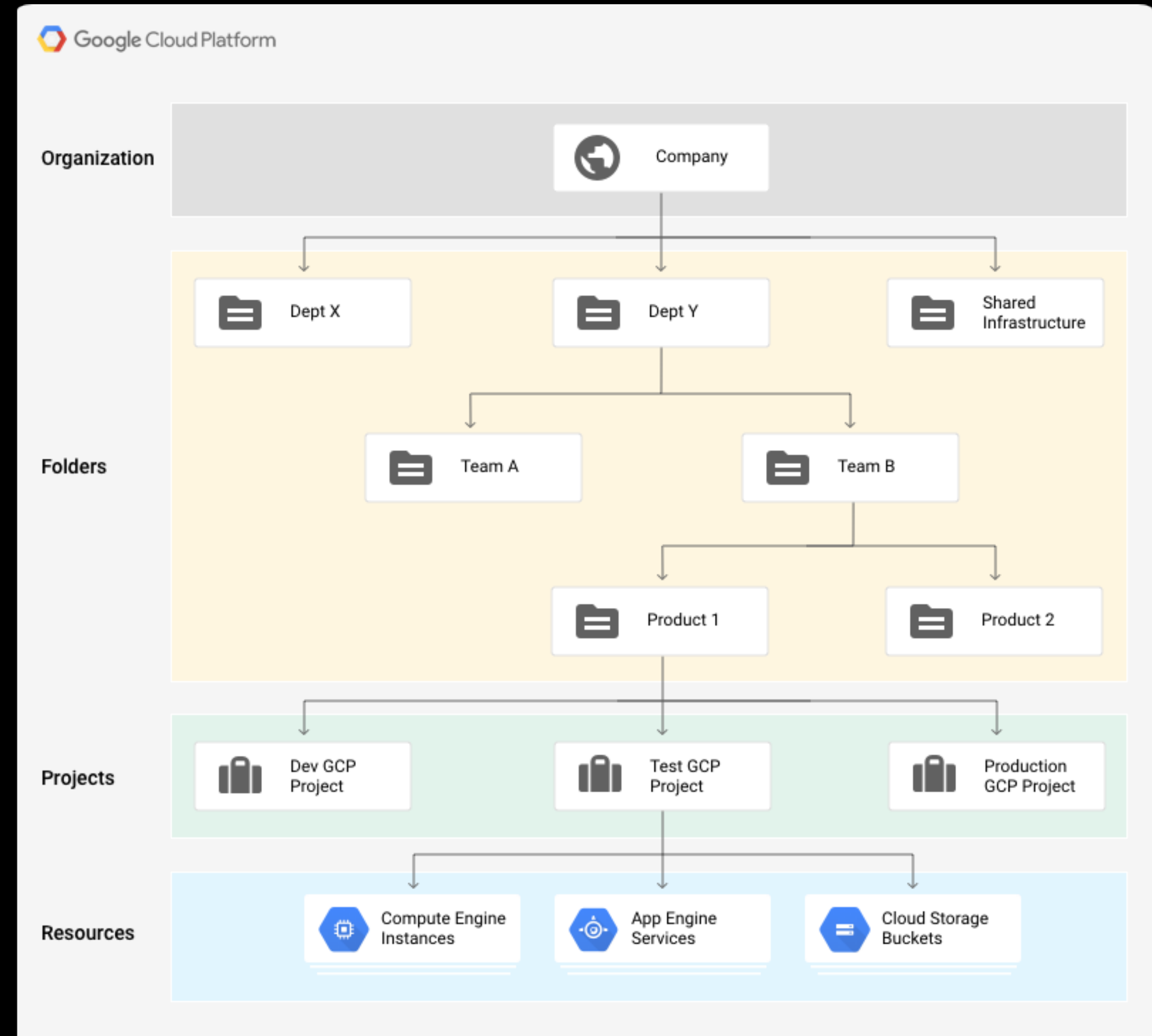
- An SSD that is physically attached to the system running the VM - no network!
- It loses its contents when the VM is started or stopped for any reason, including maintenance
- Sold in 375GB increments with a limit on attachments by machine type
- Ultra high performance, good for temporary workloads, caches, etc
- Can't be moved between VMs





# A VM IS A RESOURCE

- All Google Cloud Platform resources (not just VMs) exist inside a “project” - a project groups together resources so you can share access and billing easily and so resources can utilize discovery to easily connect to each other
- A VM is a resource inside a project

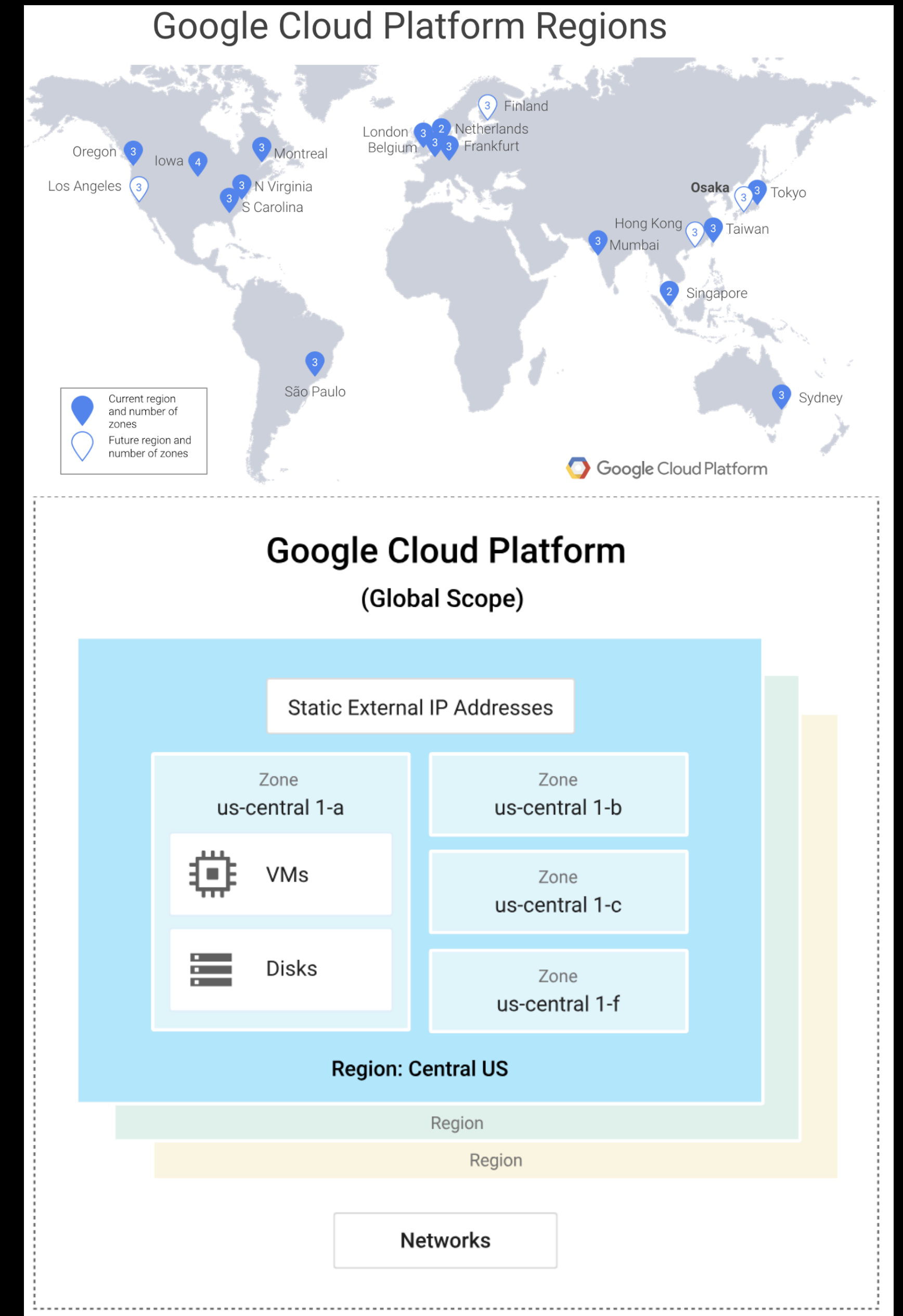






# WHERE A VM LIVES

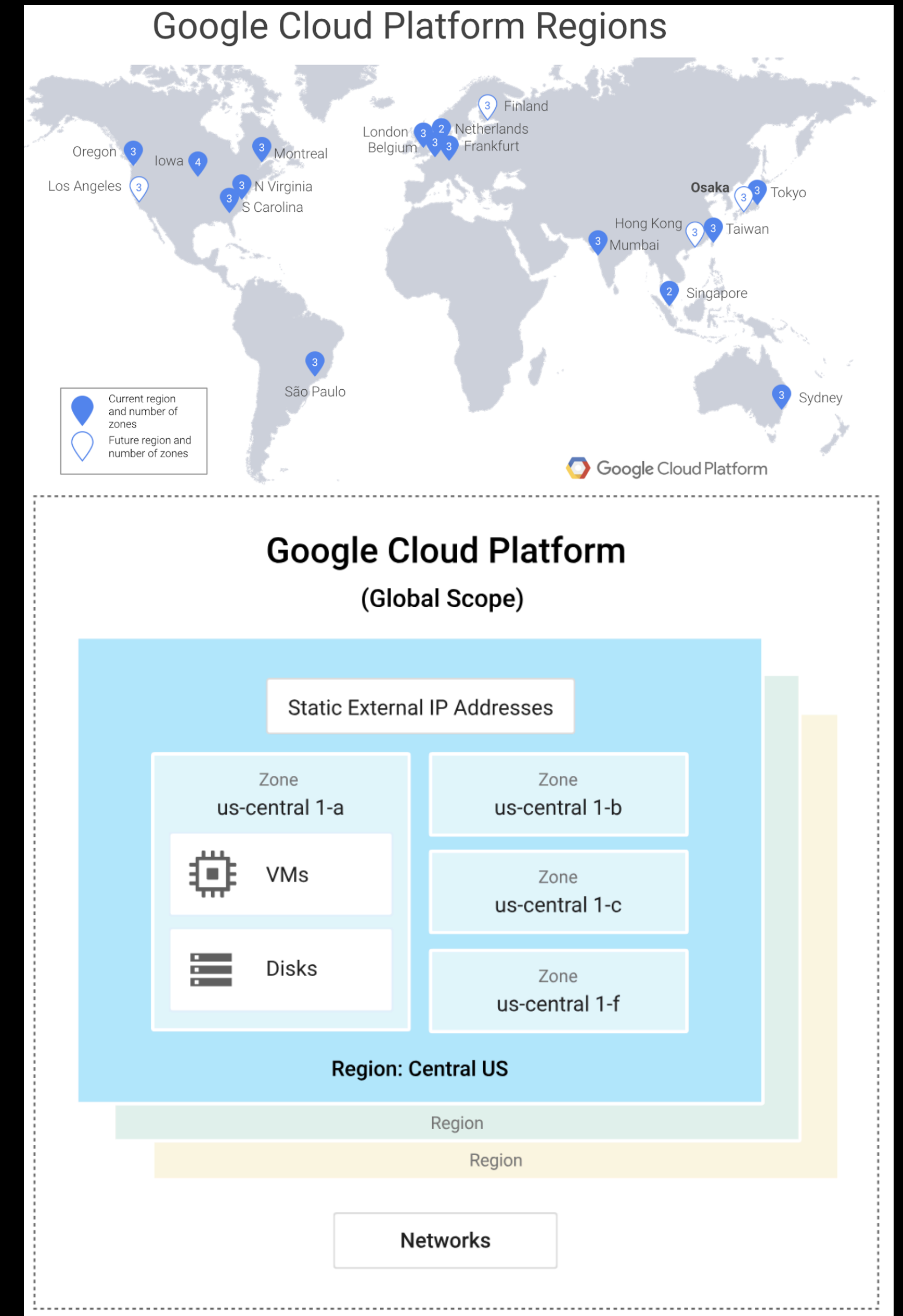
- A VM is launched in a given zone
  - A "zone" lives inside a "region"
  - A "region" is a given data center in a certain part of the world (think "a city")
- A "zone" is given building or room inside that data center that has its own power, security, fire suppression, heating/cooling, etc - ideally (and within reason), an incident such as a fire in a given zone would not affect another zone
- An outage in a "region" would be something major that would take down a few city blocks, etc for example
- New regions & zones are coming up all the time - consult the GCP website for the latest list





# WHERE A VM LIVES

- Each zone has different capabilities that change
  - Type of CPUs offered
  - Type of GPUs offered
  - Network capabilities
- Consult Google's website for an up-to-date listing

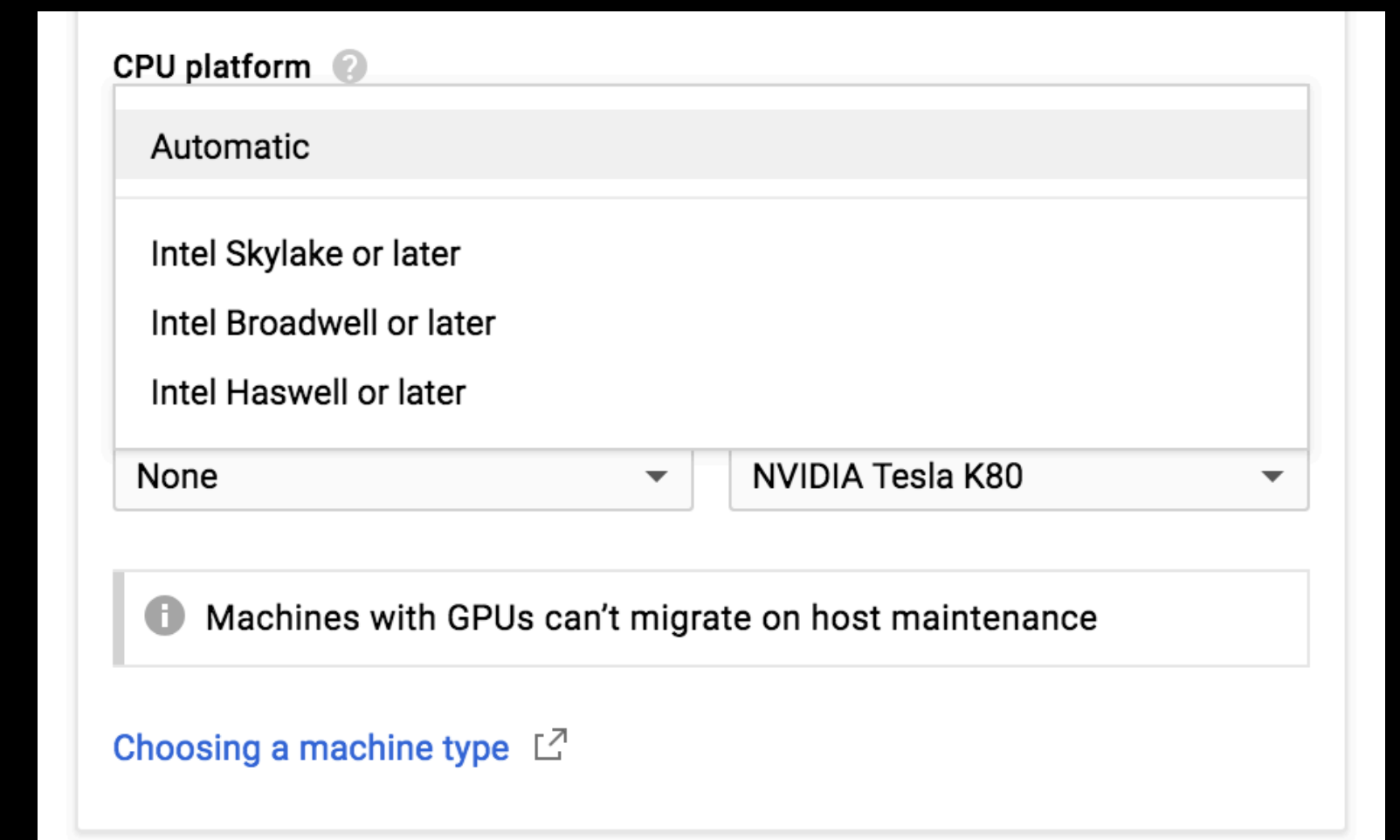






# THE LIFE OF A VM

- Standard vs. Preemptible
  - Standard Instances run until you stop them and may live in a “stopped” state to be restarted
  - Preemptible Instances are technically exactly the same but much cheaper as they are sold based on excess capacity and maybe stopped at any time and will always only live at least 24 hours
- Sole vs. Shared Tenancy





# THE LIFE OF A VM

- All VMs start by being defined in the Console, command line, or in the API by an external process
  - Select name, type, image (operating system/software stack), location (zone), network(s), disks, tags/labels, tenancy

VM Created







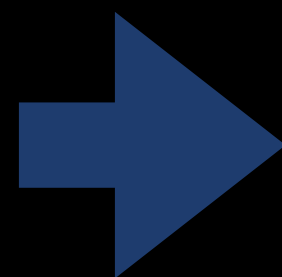
# THE LIFE OF A VM

- VMs all start/boot like a regular system, based on options, they may start:
  - On a specific processor/microarchitecture type and/or with GPUs
  - With a given attachment of disks
  - Connected to given networks (static IPs, custom subnets, behind firewalls, etc)
  - On a machine shared with others or dedicated only for your VMs

VM Created



VM Started/Running





# THE LIFE OF A VM

- Sometimes, things go wrong, or maintenance needs to happen
  - Underlying hardware
  - Hypervisor security fix/maintenance
  - Scheduled facilities maintenance requiring system shutdown
- Compute Engine offers seamless migration for most resources (GPU-equipped VMs cannot)
  - VM is moved without any impact to processes inside it to new hardware

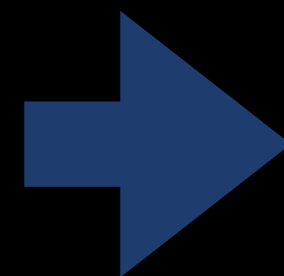
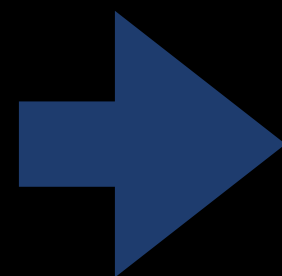
VM Created



VM Started/Running



Migrated?

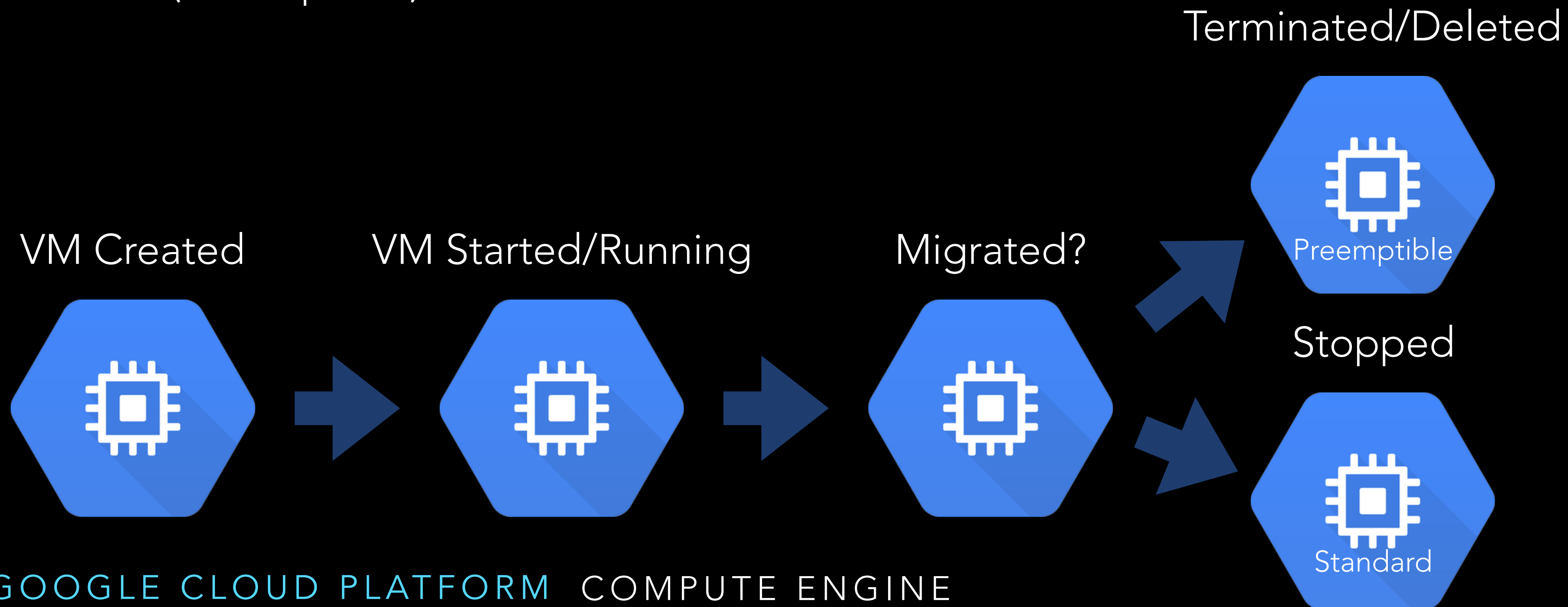






# THE LIFE OF A VM

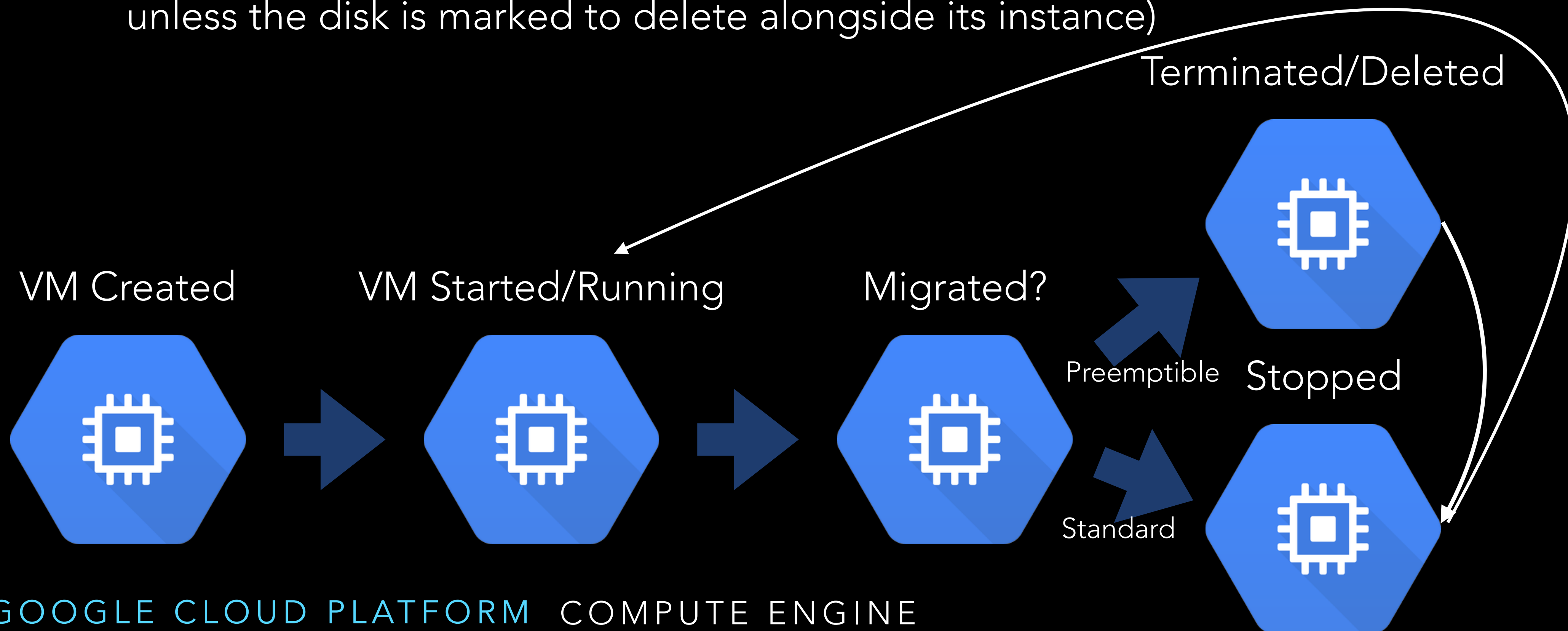
- Based on VM type, now, it either:
  - Continues running until stopped (Standard)
  - Is terminated at 24 hrs or when GCP usage spikes, whichever comes first (Preemptible)





# THE LIFE OF A VM

- You may restart a stopped instance (you are only billed for storage, IP addresses, etc while an instance is stopped, not its vCPU or RAM)
- You may delete a stopped instance (and decide what to do with the other resources unless the disk is marked to delete alongside its instance)







# PRACTICAL EXAMPLE

- Let's bring up a Linux (Debian) VM
- We'll start with the default VM then edit it to
  - Include an additional persistent disk
  - Add more vCPU & RAM





# PRACTICAL EXAMPLE

- There are several ways you could accomplish this
  - Command line (gcloud) command
  - REST API
  - Google Cloud Console (Web)
- Overwhelmingly, most users create resources using the Google Cloud Console we explored earlier
- We'll use the Google Cloud Console but will also examine the REST & Command Line equivalents through the Cloud Console interface

