

Compute Services Fundamentals



GCE
Compute Engine



GAE
App Engine



GKE
Kubernetes Engine

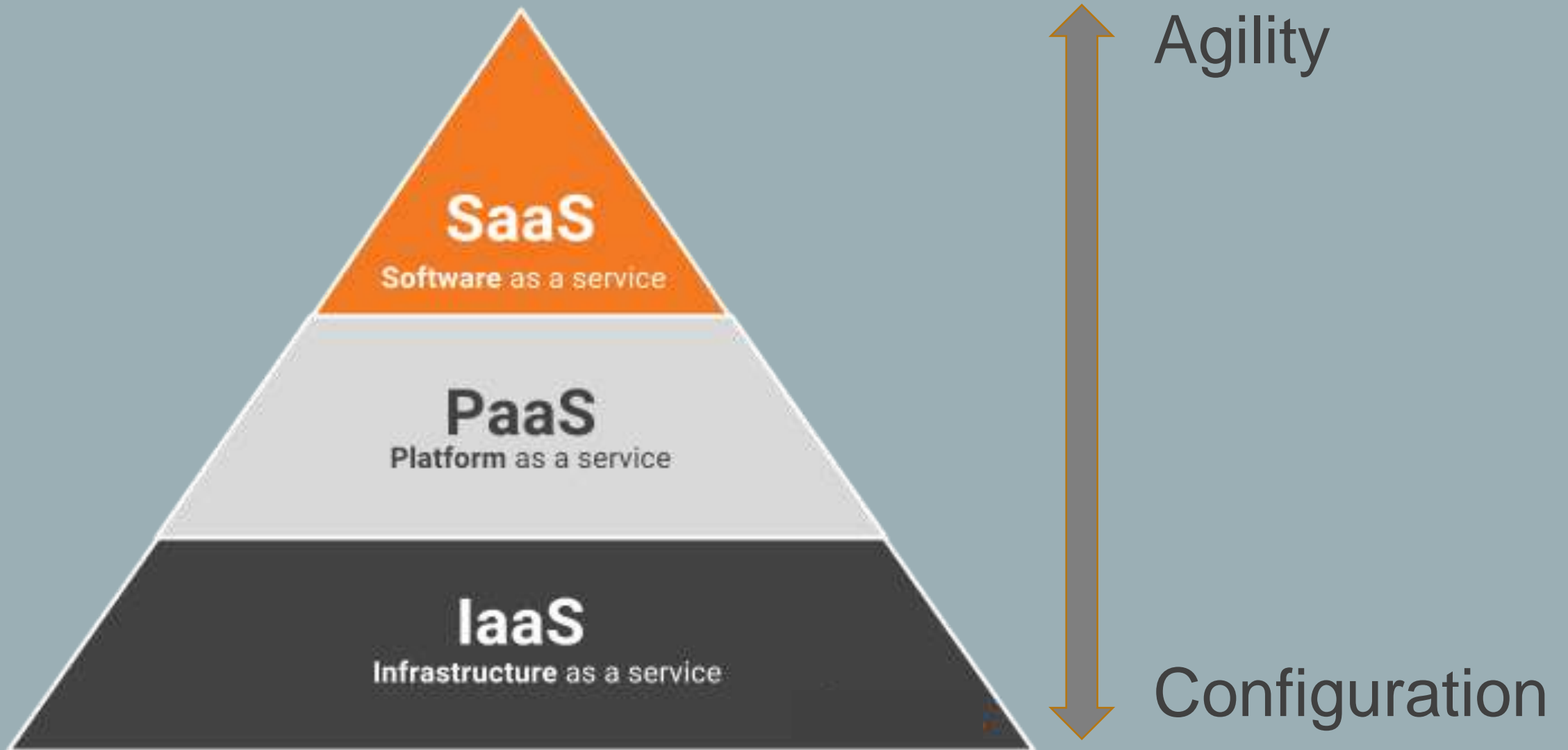


Cloud Function

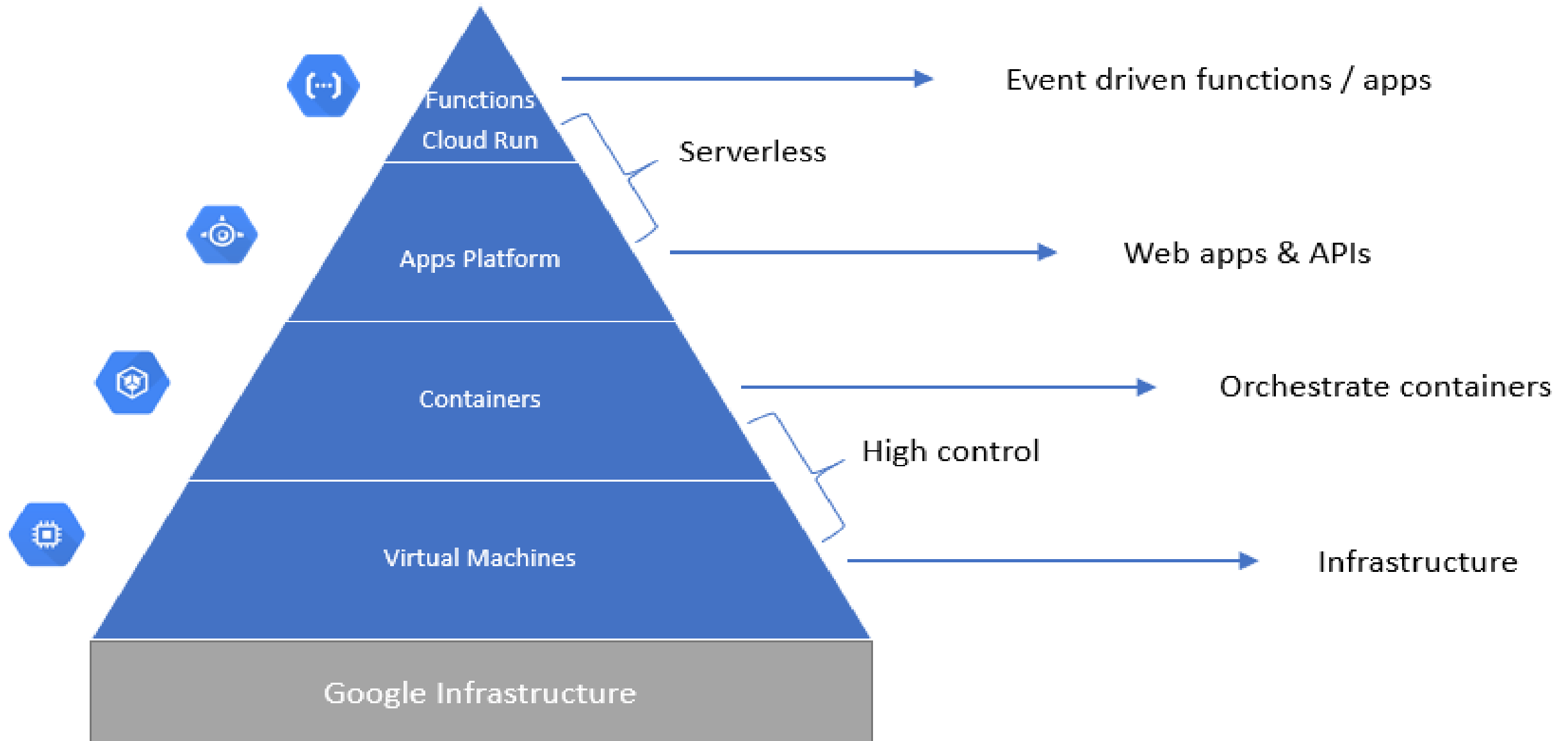


Cloud Run

Services Type Pyramid



Compute Services



GCE Introduction

Virtual machine running in cloud

IaaS offering

Full control for infrastructure customization

Supports custom image

Data at rest is encrypted

Supports network customization

Supports OS customization

GCE Introduction

Supports SSH connection

Sub milli second billing after 1 minute of minimum billing

CDN Support

Network performance defined on number of CPU

Live migration of VM

Discounts like committed use discount

Support load balancing and auto scaling

Auto restart support

GCE Components – Machine Type

General Purpose

- Ideal for typical balanced instances with respect to RAM and CPU
- Best price-performance ratio for a variety of workloads

Memory Optimized

- Ideal for memory-intensive workloads
- Offer more memory per core than other machine types,
- Choose memory up to 12 TB of memory.

CPU Optimized

- Highest performance per core on Compute Engine
- Optimized for compute-intensive workloads

GCE Components – Machine Type

Custom Machine

- Ideal for workload that requires more processing power or memory than what is offered by the Google-provided types
- Costlier than predefined machine type

GPU

- Ideal for intensive operations that require dedicated processing units or graphics processing
- Only for general-purpose N1 machine types

GCE Components – CPU Platform

- GCE instance uses one of the CPU platforms available on Compute Engine
- CPU Platform determines the CPU processor used in instance
- Compute Engine offers both Intel and AMD processors for your VMs.
- Each zone has default CPU Processor
- Choice of CPU processor also depends on type of machine selected
- Each processor type is not available for all machine type

GPU

- Accelerate specific workloads
- Not supported on shared-core or memory optimized machine type
- Receive sustained use discounts
- Can be added for preemptible machine
- Maximum CPU and memory is zone dependent
- Currently supported for N1 type only
- Number of GPU are limited by project quota

Local SSD Disk

Local SSD

- Physically attached to server
- Local SSD have high IOPS and lower latency than persistent SSD disk
- Ephemeral in nature.
- Data is lost if instance is stopped or deleted
- Data is encrypted at rest
- Good for cached data or redundant data that can be rebuilt easily
- Does not support live migration

Persistent Disk

Persistent Disk

- Durable network storage
- Data is stored in multiple physical disk across zone/region
- Synchronous replication
- Located independently from VM
- Data is encrypted
- Resize disk easily and migrate instance
- Scale in performance
- Snapshot support
- Max 64TB per instance
- Cannot attach to an instance in another project
- Support for live migration
- Can be used as read only disk between multiple VM

Persistent Disk Type

Standard Persistent Disk

- Backed by HDD
- Ideal for efficient and reliable block storage
- Only available within a single zone

SSD persistent disks

- Backed by SSD
- Ideal for fast and reliable block storage

Balanced persistent disk

- Backed by SSD
- Highest performing persistent disk
- Highest cost

Persistent Disk Type

Standard Persistent Disk

- Backed by HDD
- Ideal for efficient and reliable block storage
- Only available within a single zone


SSD persistent disks


- Backed by SSD
- Ideal for fast and reliable block storage
- Max 64TB per instance
- Only available within a single zone


Balanced persistent disk


- Backed by SSD
- Highest performing persistent disk
- Highest cost


Image


 Compute Engine


 VM instances


 Instance groups


 Instance templates


 Sole-tenant nodes


 Machine images


 Disks


 Snapshots


 Images


 TPU


 Migrate for Compute Engine


 Committed use discounts

 Metadata

 Health checks

 Zones

 Marketplace



Images

CREATE IMAGE

REFRESH

CREATE INSTANCE

DEPRECATE

DELETE












SHOW INFO PANEL

LEARN

ImagesImage import historyImage export history

Filter Images

Columns

<input type="checkbox"/>	Name	Location	Size	Disk size	Created by	Family	Creation time
<input type="checkbox"/>	 c0-common-gce-gpu-image-20200128	asia, eu, us	307.24 GB	50 GB	c0-common-gce-gpu-image-20200128	common-gce-gpu-image	Jan 28, 2020, 12:21:55 PM
<input type="checkbox"/>	 c1-deeplearning-common-cu100-v20200728	asia, eu, us	362.24 GB	50 GB	c1-deeplearning-common-cu100-v20200728	common-dl-gpu	Jul 29, 2020, 6:43:04 AM
<input type="checkbox"/>	 c1-deeplearning-common-cu100-v20200825-debian-9	asia, eu, us	405.37 GB	50 GB	Debian	common-dl-gpu-debian-9	Aug 26, 2020, 10:52:39 PM
<input type="checkbox"/>	 c1-deeplearning-common-cu100-v20200826-debian-10	asia, eu, us	173 GB	50 GB	Debian	common-dl-gpu-debian-10	Aug 27, 2020, 12:59:47 AM
<input type="checkbox"/>	 c2-deeplearning-pytorch-1-2-cu100-20191005	asia, eu, us	295.38 GB	30 GB	c2-deeplearning-pytorch-1-2-cu100-20191005	pytorch-1-2-gpu	Oct 7, 2019, 2:29:30 AM
<input type="checkbox"/>	 c2-deeplearning-pytorch-1-3-cu100-20191219	asia, eu, us	338.79 GB	50 GB	c2-deeplearning-pytorch-1-3-cu100-20191219	pytorch-1-3-gpu	Dec 20, 2019, 5:31:28 AM
<input type="checkbox"/>	 c2-deeplearning-pytorch-1-4-cu101-v20200701	asia, eu, us	436.3 GB	50 GB	c2-deeplearning-pytorch-1-4-cu101-v20200701	pytorch-1-4-gpu	Jul 1, 2020, 10:27:59 PM
<input type="checkbox"/>	 c2-deeplearning-pytorch-1-4-cu101-v20200825-debian-9	asia, eu, us	595.48 GB	50 GB	Debian	pytorch-1-4-gpu-debian-9	Aug 26, 2020, 11:44:46 PM
<input type="checkbox"/>	 c2-deeplearning-pytorch-1-4-cu101-v20200826-debian-10	asia, eu, us	269.05 GB	50 GB	Debian	pytorch-1-4-gpu-debian-10	Aug 27, 2020, 1:27:23 AM
<input type="checkbox"/>	 c3-deeplearning-tf-ent-1-15-cu100-v20200728	asia, eu, us	609 GB	50 GB	c3-deeplearning-tf-ent-1-15-cu100-v20200728	tf-ent-1-15-gpu	Jul 29, 2020, 5:54:57 AM
<input type="checkbox"/>	 c3-deeplearning-tf-ent-1-15-cu100-v20200825-debian-9	asia, eu, us	537.52 GB	50 GB	Debian	tf-ent-1-15-gpu-debian-9	Aug 27, 2020, 12:32:35 AM

Image

Public Image

- Public images are provided and maintained by Google, open source communities, and third-party vendors
- Most public images at no additional cost, but there are some premium images that do add additional cost

Custom Image

- Custom image is a boot disk image that you own and control access to
- It can be created in on premise and then used in cloud to create instance
- No cost to your instances for using custom image
- Storage charge for keeping custom image in your project
- Can be created in on-prem environment
- Can be imported from on-prem or other cloud provider like AWS

Image

Boot disk

Select an image or snapshot to create a boot disk; or attach an existing disk. Can't find what you're looking for? Explore hundreds of VM solutions in [Marketplace](#).

Public images

Custom images

Snapshots

Existing disks

Operating system

Container Optimized OS

Version

Container-Optimized OS 77-12371.1073.0 stable

Kernel: ChromiumOS-4.19.112 Kubernetes: 1.15.3 Docker: 19.03.1 Family: cos-77-
Its, supports Shielded VM features ?

Boot disk type ?

Size (GB) ?

Balanced persistent disk

Max disk size: 65,536 GB

SSD persistent disk

Max disk size: 65,536 GB

Standard persistent disk

Max disk size: 65,536 GB

10

Image & Snapshot

Boot disk

Select an image or snapshot to create a boot disk; or attach an existing disk. Can't find what you're looking for? Explore hundreds of VM solutions in [Marketplace](#).

Public images

Custom images

Snapshots

Existing disks

Show images from

Select a project ▼

Boot disk type ?

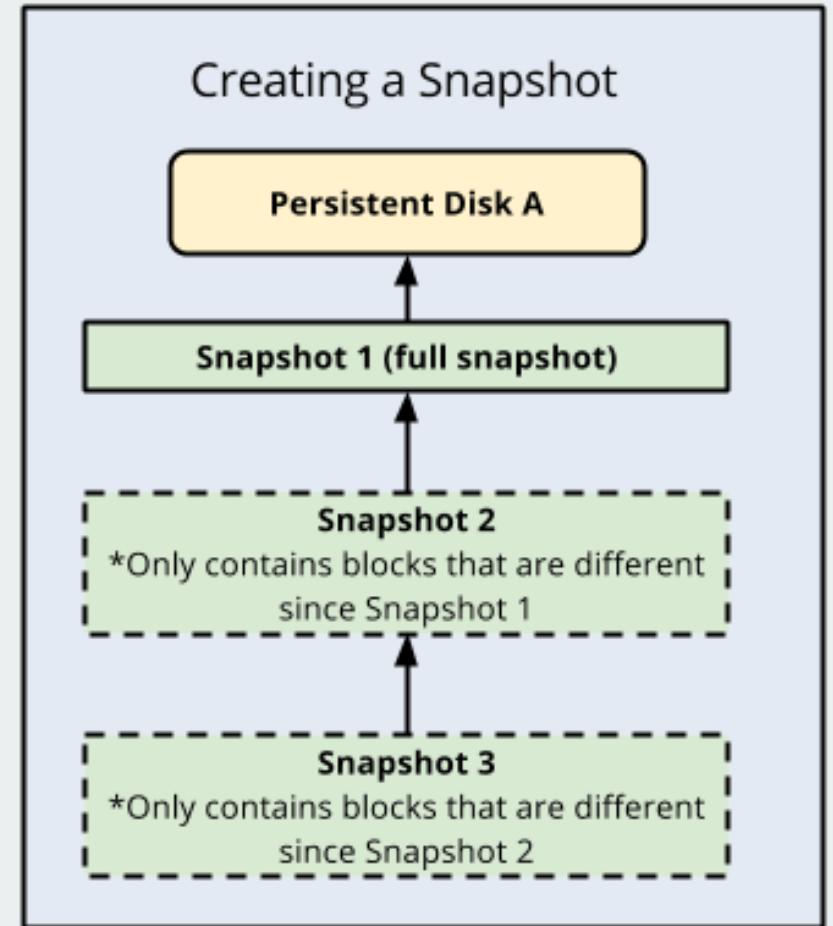
Standard persistent disk ▼

Size (GB) ?

10

Snapshot

- Incremental back of disk data
- Create snapshot schedule to avoid data loss
- Stored in cloud storage
- Can't change the storage location of an existing snapshot.
- Global resource
- Can be shared across project



Networking

Firewall

Add tags and firewall rules to allow specific network traffic from the Internet.

- ☐ Allow HTTP traffic
- ☐ Allow HTTPS traffic

Management Security Disks Networking Sole Tenancy

Network tags (Optional)

Hostname


Set a custom hostname for this instance or leave it default. Choice is permanent.

Network interfaces


Network interface is permanent.




 Add network interface

 To create another network interface you need to have a new network first.


Networking

Network interfaces 
Network interface is permanent


Network interface

Network 

default


Subnetwork 

default (10.128.0.0/20)


Primary internal IP 


Ephemeral (Automatic)


 Show alias IP ranges


External IP 

Ephemeral


Network Service Tier 

☒ Premium (Current project-level tier, [change](#)) 

☐ Standard (us-central1) 

IP forwarding 

Off

Public DNS PTR Record 

☐ Enable

PTR domain name

Done

Cancel

Start Up Script

- Piece of code that will always be executed when your instance boots up
- It can perform actions like
 - Environment configuration
 - Downloading dependencies
 - Initializing services
 - downloading updates
- Bash and Python script supported
- Can be used for both running and new VM
- Can be linked to cloud storage

Meta Data

- Meta data is key value pair
- Stored in metadata server
- Metadata server can be reached just inside from VM
- VM can read this metadata by querying metadata server
- Both default metadata and custom metadata is supported
- Generally used with startup and shutdown script

Preemptible VM

- Discounted VM in Google cloud
- Uses surplus inventory available in zone in cloud
- Maximum 80% discount
- Can use it for maximum 24 hours
- Allocated based on bidding process
- Can be preempted with short notice of 30 seconds
- Discount are given only for compute resources
- Storage resource attached to VM are not preemptible resources
- Data will not be lost if persistent disk is used to store data
- Preemptible instances can't live migrate to a regular VM instance

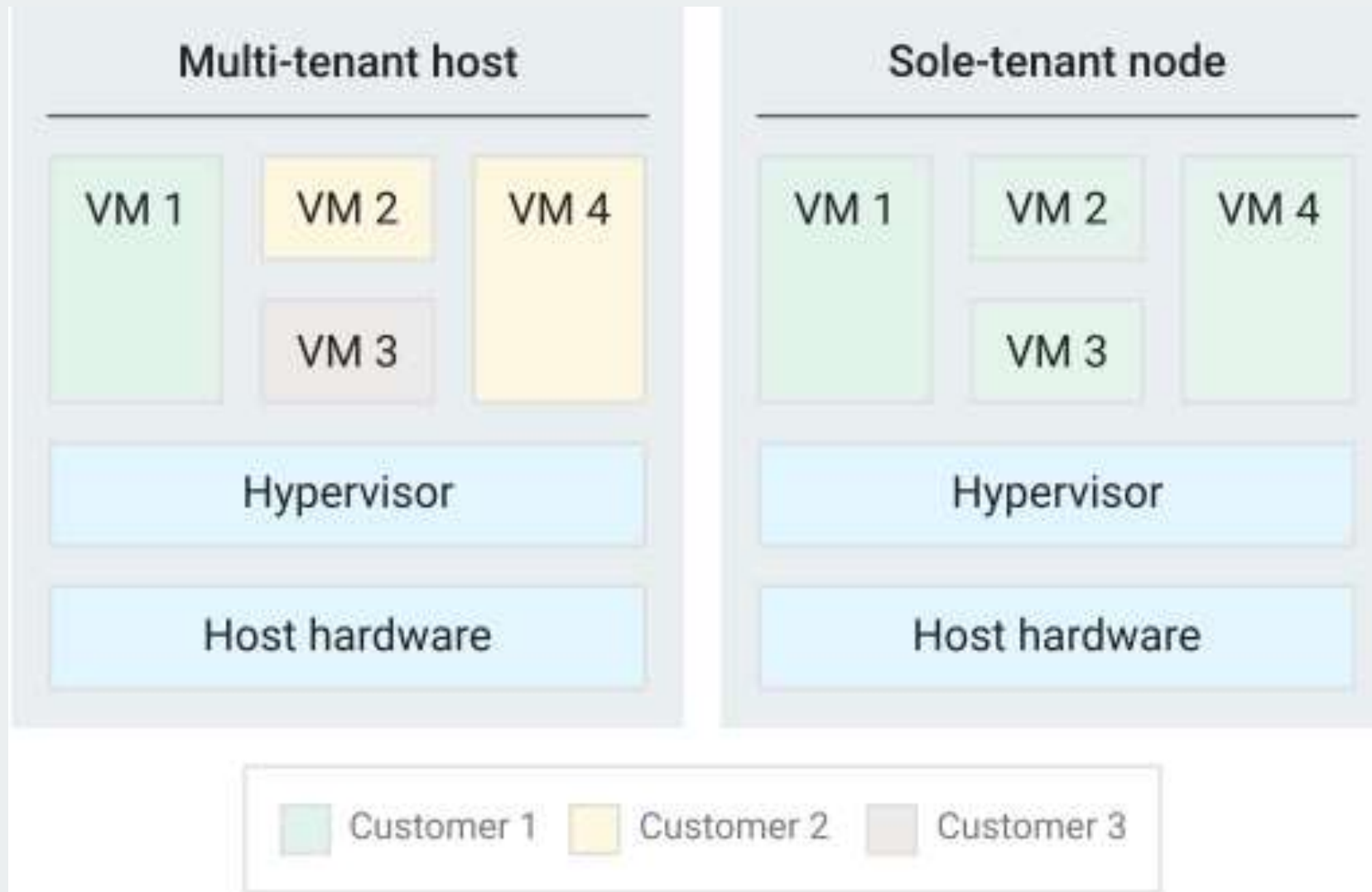
On Host Maintenance & Auto Restart

On Host maintenance determines what should happen to your VM during infrastructure maintenance.

Termination or Live migration

Automatic restart feature can automatically restart VM instances if they are terminated due to non-user-initiated reason

Sole Tenancy



Instance Template

- Instance template is a resource that you can use to create Virtual Machine (VM)
- Helps in creating identical resources
- Does not define location of VM to be created
- Just defines the resources configuration required to create VM like machine type, boot disk
- It's a global resource
- Can not update existing instance template

Instance Group

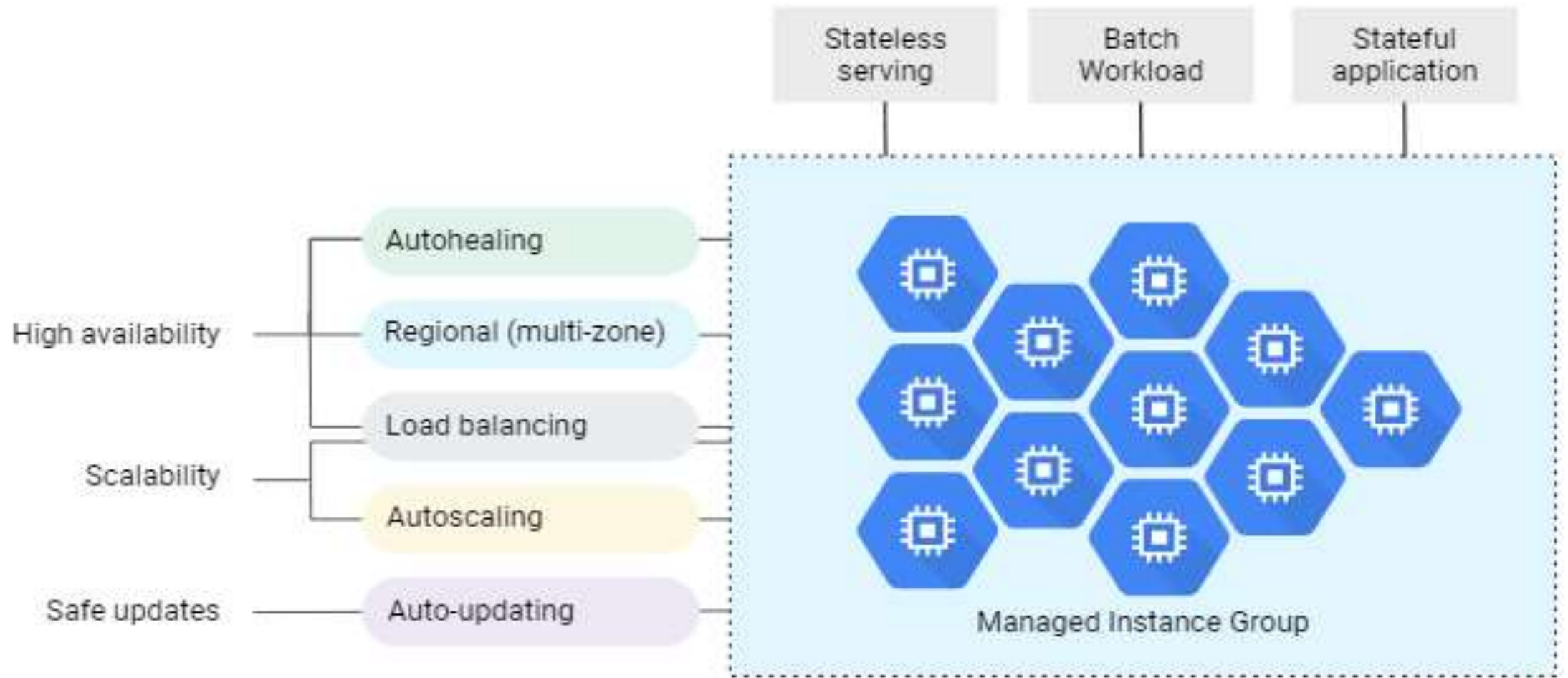
An instance group is a collection of virtual machine (VM) instances that you can manage as a single entity.

Managed Instance Group

- Regional
- Zonal

Unmanaged Instance group

Managed Instance Group



Unmanaged Instance Group

- Unmanaged instance groups can contain heterogeneous instances
- Don't use instance template
- Don't support features like autoscaling, autohealing, rolling update support, multi-zone support
- Not good for deploying highly scalable and available application
- Load balancing is supported

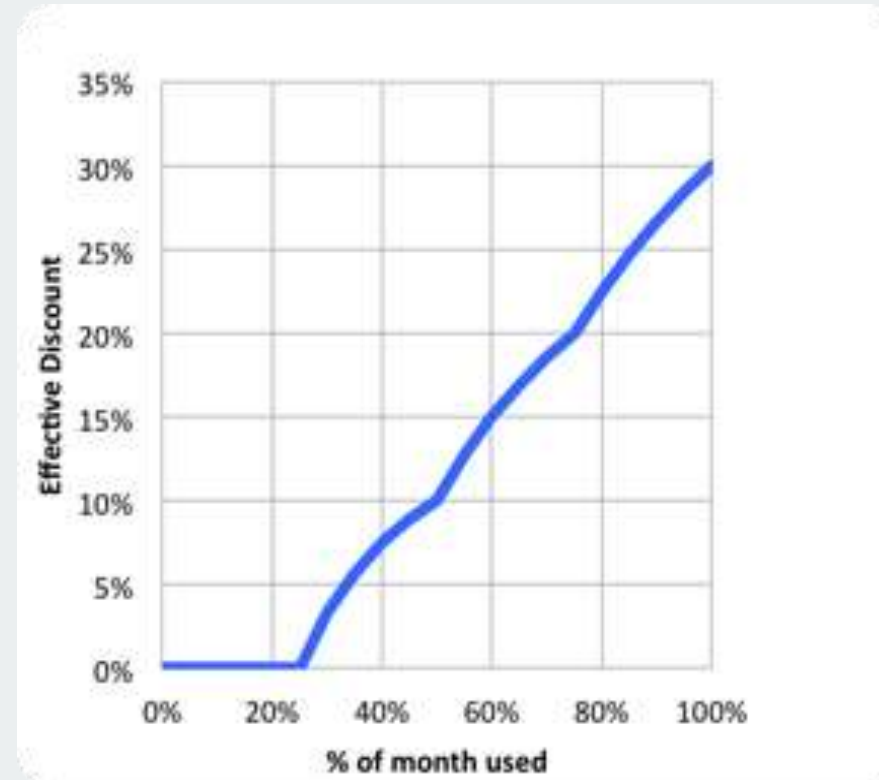
Pricing

- Compute Engine are priced based on individual resource usage
- Custom machine are charged more than predefined machine
- Preemptible machine can be discounted up to 80%
- Multiple VM can be combined to inferred VM to get more discount
- Minimum chargeable interval is 1 minute and charged per minute after this
- User can get discounts like sustained use discount or committed use discount

Sustained Use Discount

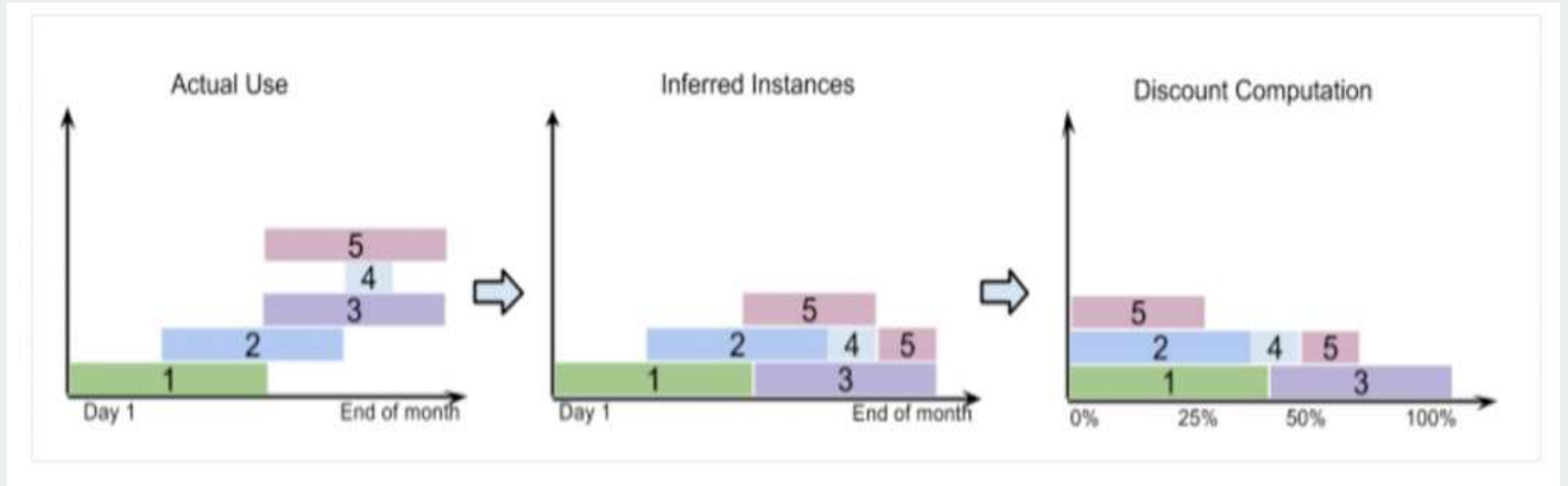
- Discount based on incremental usage
- More usage mean more discount

Resource	Usage level (% of month)	% at which incremental is charged	Incremental rate (USD/hour) example: n1-standard-1 instance
• General-purpose N1 predefined and custom machine types	0%–25%	100% of base rate	\$0.0475
• Memory-optimized machine types	25%–50%	80% of base rate	\$0.0380
• Shared-core machine types	50%–75%	60% of base rate	\$0.0285
• Sole-tenant nodes	75%–100%	40% of base rate	\$0.0190



Inferred Instance

Discount due to combination of usage of similar VM



Committed use Discount

- Discount due to long term commitment to use VM
- Allows user to get VM at deep discount than normal price
- Ideal for predictable workload
- Applicable to both predefined and custom machine
- No upfront cost required