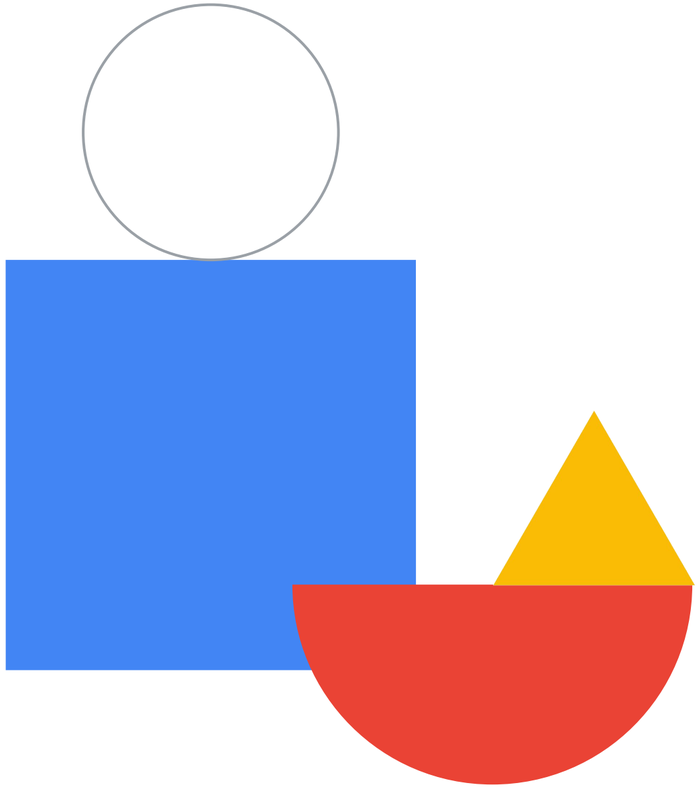
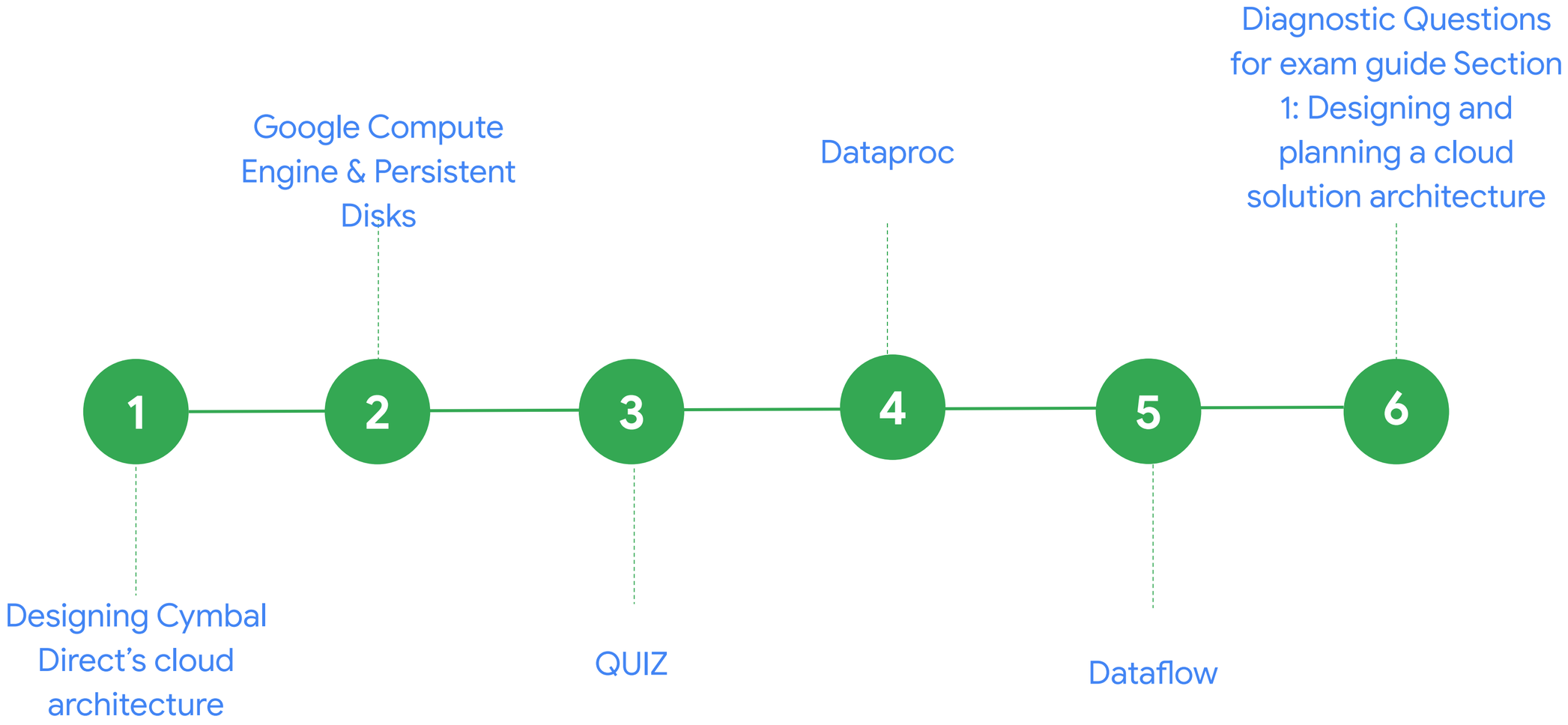
Google Cloud



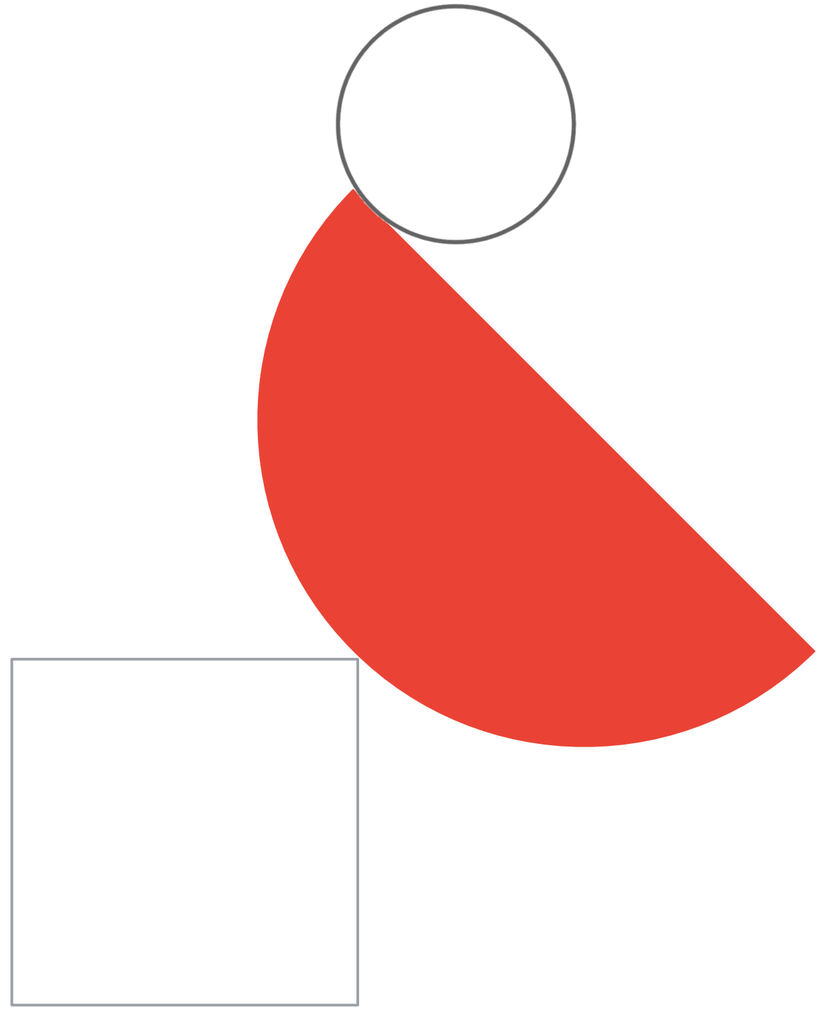
Preparing for Your Professional Cloud Architect Journey

Module 1: Designing and Planning a Cloud Solution Architecture

Week 2 agenda



Designing Cymbal Direct's cloud architecture



Cymbal Direct’s existing environment

Delivery by Drone

- Their website frontend, pilot, and truck management systems run on **Kubernetes.**
- Positional data for drone and truck location is kept in a **MongoDB** database clusters
- Drones stream video to virtual machines via stateful connection

Purchase & Product APIs

- APIs are simply built into **monolithic apps**, and were not designed for partner integration.
- APIs are running on **Ubuntu linux VMs**

Social Media Highlighting

- Single SuSE linux VM
- MySQL DB
- Redis
- Python

Cymbal Direct’s business requirements

- Scale to handle additional demand when expanding into test markets
- Streamline development
- Spend developer time on core business functionality as much as possible
- Let partners order directly via API
- Deploy the social media highlighting service and ensure appropriate content

Cymbal Direct’s technical requirements

- Managed services
- Container-based workloads
- Highly scalable environment
- Standardization where possible
- Existing virtualization infrastructure refactored over time
- Secure partner integration
- Streaming IoT data

Let's Brainstorm

Putting it together: Existing environment

Existing environment

Website frontend, pilot, and truck management systems run on Kubernetes

Technical requirements

Business requirements

Proposed product/ solution

* One row of a much larger spreadsheet

Putting it together: Technical requirements

Existing environment

Website frontend, pilot, and truck management systems run on Kubernetes

Technical requirements (does it...?)

- Move to managed services wherever possible
- Ensure that developers can deploy container based workloads to testing and production environments in a highly scalable environment.
- Standardize on containers where possible

Business requirements

Proposed product/ solution

* One row of a much larger spreadsheet

Putting it together

Existing environment

Website frontend, pilot, and truck management systems run on Kubernetes

Technical requirements (does it...?)

- Move to managed services wherever possible
- Ensure that developers can deploy container based workloads to testing and production environments in a highly scalable environment.
- Standardize on containers where possible

Business requirements (does it...?)

- Easily scale to handle additional demand when needed?
- Streamline development?

Proposed product/ solution

?

* One row of a much larger spreadsheet

Potential solutions

Existing environment

Website frontend, pilot, and truck management systems run on Kubernetes

Technical requirements (does it...?)

- Move to managed services wherever possible
- Ensure that developers can deploy container based workloads to testing and production environments in a highly scalable environment.
- Standardize on containers where possible

Business requirements (does it...?)

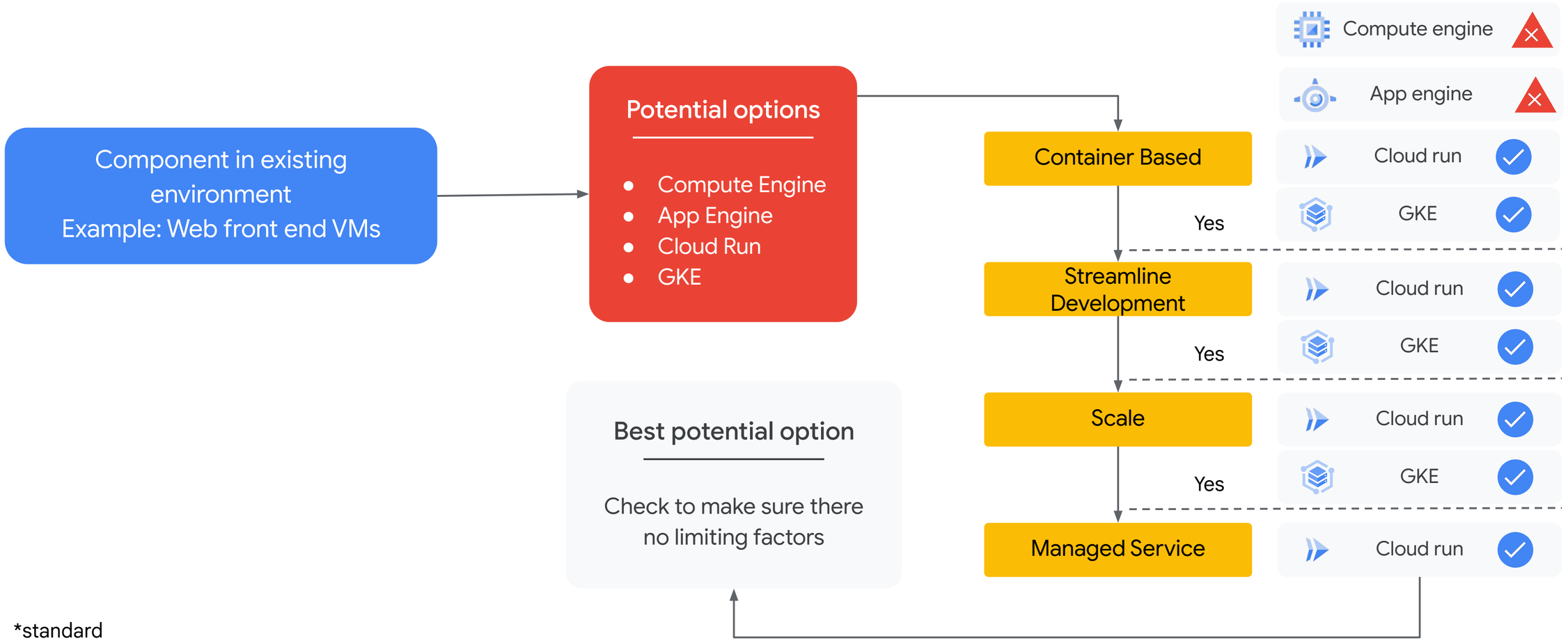
- Easily scale to handle additional demand when needed?
- Streamline development?

Proposed product/ solution

- Global HTTP(s) Load Balancer
- GKE
- Separate projects
- Migration type: lift and shift
- Replace GKE with Cloud Run for website (future)

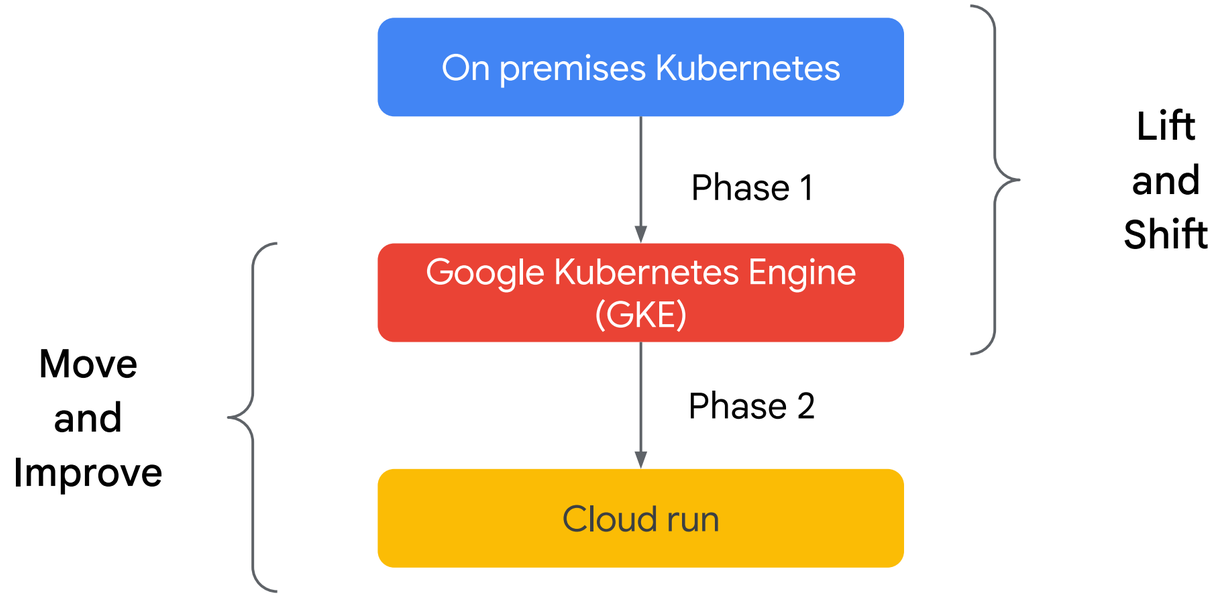
* One row of a much larger spreadsheet

Decision flow diagram

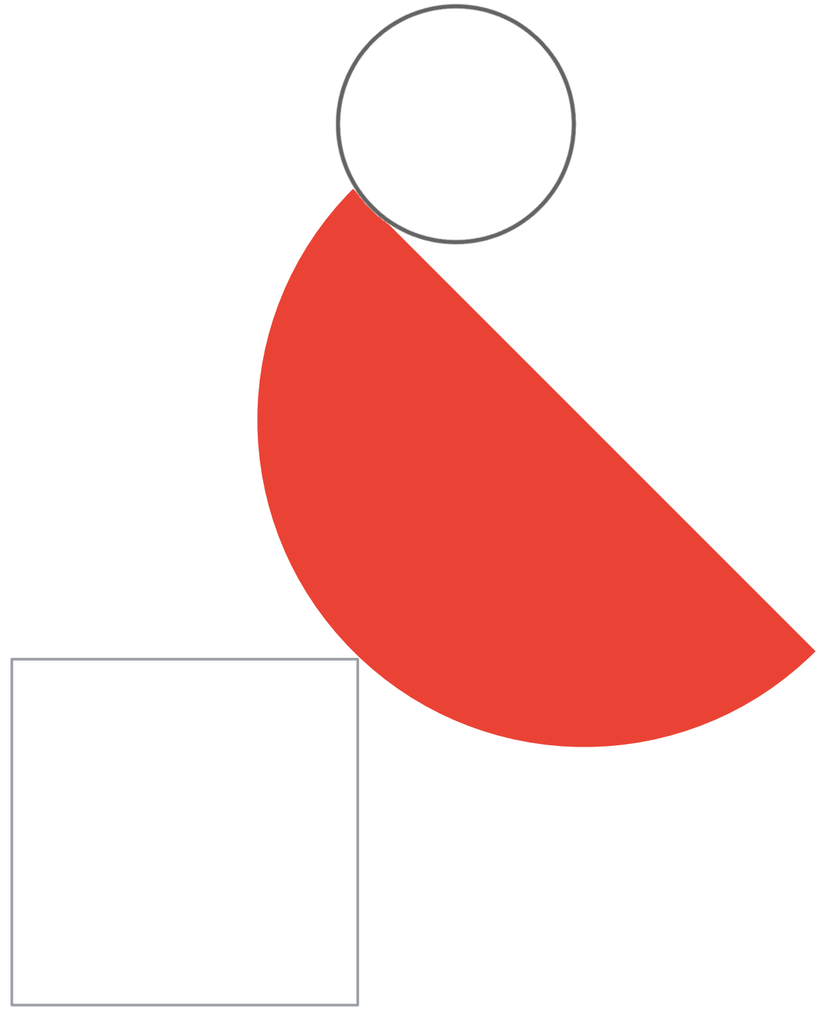


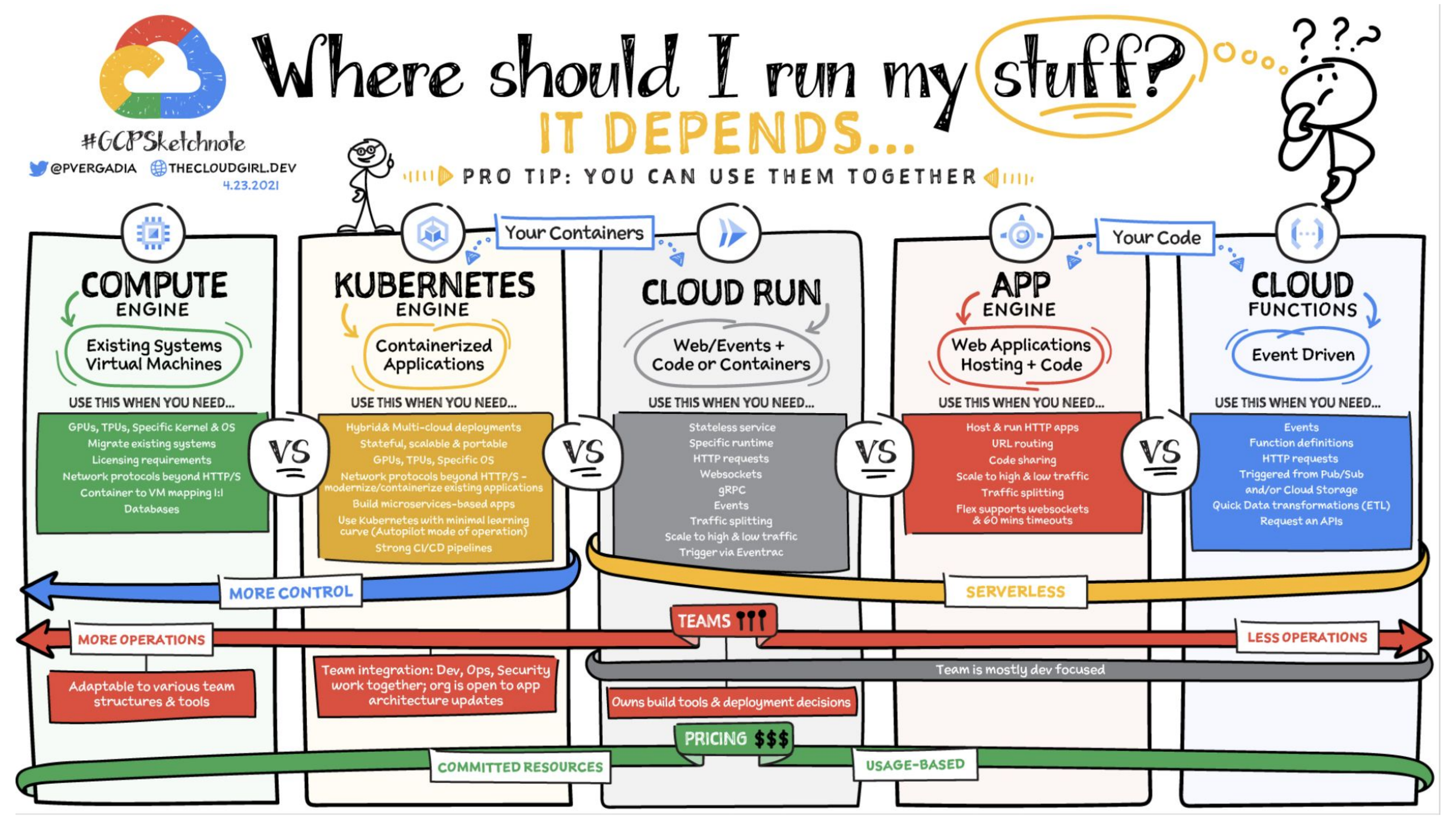
*standard

Planning for migration and the future



Google Compute Engine (GCE)




#GCPSketchnote
@PVERGADIA
THECLOUDGIRL.DEV
4.23.2021
Where should I run my stuff?
IT DEPENDS...
PRO TIP: YOU CAN USE THEM TOGETHER
Your Containers
Your Code
COMPUTE ENGINE
Existing Systems Virtual Machines
USE THIS WHEN YOU NEED...
GPUs, TPUs, Specific Kernel & OS
Migrate existing systems
Licensing requirements
Network protocols beyond HTTP/S
Container to VM mapping 1:1
Databases
VS
KUBERNETES ENGINE
Containerized Applications
USE THIS WHEN YOU NEED...
Hybrid& Multi-cloud deployments
Stateful, scalable & portable
GPUs, TPUs, Specific OS
Network protocols beyond HTTP/S - modernize/containerize existing applications
Build microservices-based apps
Use Kubernetes with minimal learning curve (Autopilot mode of operation)
Strong CI/CD pipelines
VS
CLOUD RUN
Web/Events + Code or Containers
USE THIS WHEN YOU NEED...
Stateless service
Specific runtime
HTTP requests
Websockets
gRPC
Events
Traffic splitting
Scale to high & low traffic
Trigger via Eventrac
VS
APP ENGINE
Web Applications Hosting + Code
USE THIS WHEN YOU NEED...
Host & run HTTP apps
URL routing
Code sharing
Scale to high & low traffic
Traffic splitting
Flex supports websockets & 60 mins timeouts
VS
CLOUD FUNCTIONS
Event Driven
USE THIS WHEN YOU NEED...
Events
Function definitions
HTTP requests
Triggered from Pub/Sub and/or Cloud Storage
Quick Data transformations (ETL)
Request an APIs
MORE CONTROL
SERVERLESS
MORE OPERATIONS
LESS OPERATIONS
TEAMS
Adaptable to various team structures & tools
Team integration: Dev, Ops, Security work together; org is open to app architecture updates
Owns build tools & deployment decisions
Team is mostly dev focused
PRICING $$$
COMMITTED RESOURCES
USAGE-BASED

Google Compute Engine

Infrastructure as a Service (IaaS)

- vCPUs (cores) and Memory (RAM)
- Persistent disks
- Networking
- Linux or Windows

Exam Tips: *GCE is a basic IaaS service, but there are lots of details you're expected to know:*

- *Differences between PD images / snapshots / VM images.*
- *How to troubleshoot VM not booting up properly*
- *Custom image vs public image + startup scripts*
- *VM price differ between regions*
- *PDs are network-attach devices and - as such - consume VM bandwidth.*
- *VM network performance scales with # of vCPUs.*
- *etc...*

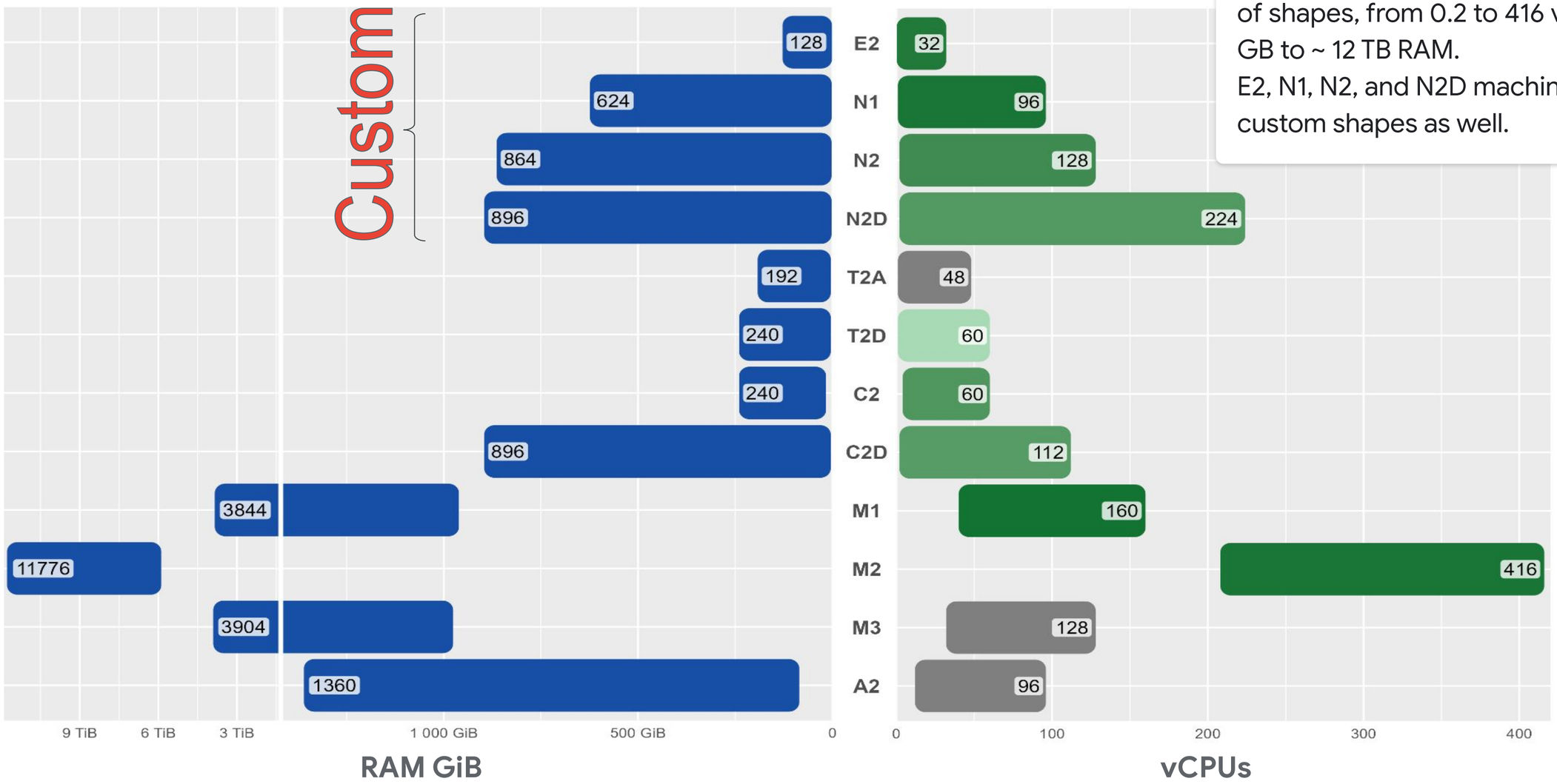
Compute Engine - how to differentiate between families?

Best TCO	Balanced	Scale-out Optimized	Workload-Optimized		
• Web Serving • Steady-state LOB apps • Dev & Test environments • Small prod environments	• Enterprise apps • Medium databases • Web & App Serving	• Scale-out Workloads • Web Serving • Containerized microservices	• EDA • HPC • Scientific Modeling • AAA Gaming	• SAP HANA • Largest in memory DBs • Real-time data analytics • In-memory cache	• ML • HPC • Massive parallelized computation
Cost savings a priority	Leading perf and perf/$	Best Perf/$ for scale out workloads	Highest performance CPUs	Most memory on Compute Engine	Highest performance GPUs
Cost-Optimized (E2)	**General Purpose** (N2 and N2D)	**ScaleOut optimized Tau (T2D, T2A)**	**Compute-Optimized** (C2, C2D)	**Memory-Optimized** (M1, M2, M3)	**Accelerator-Optimized** (A2)

Compute Engine: Max shapes by machine type

Exam Tip: *Custom machines can be used only for some VM families & up to 224vCPU/896 GB RAM*

Instances (VMs) are available in a wide range of shapes, from 0.2 to 416 vCPUs and from 1 GB to ~ 12 TB RAM.
E2, N1, N2, and N2D machine types provide custom shapes as well.

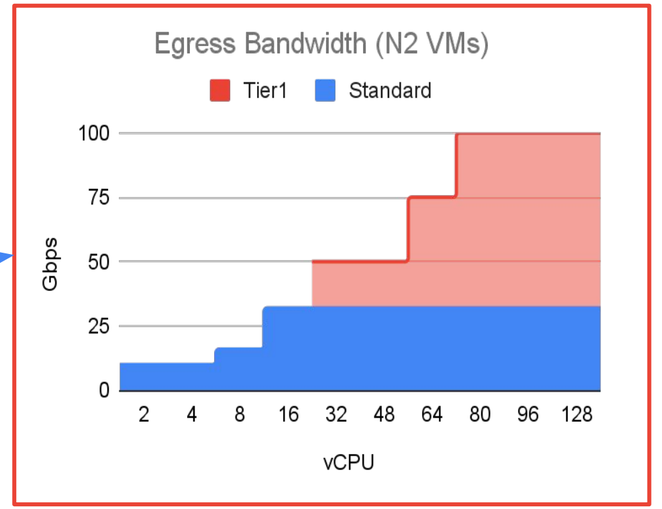


Compute Engine

Network perspective

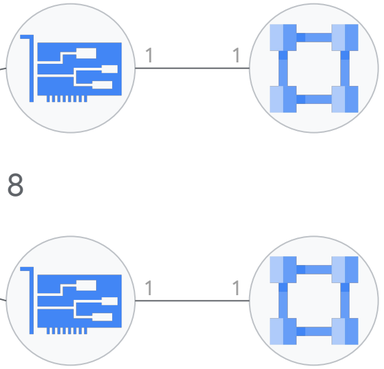
Exam Tips:

- *Network bandwidth limited & dependent on vCPU count (up to ~32Gbps for N2s + Tier1 extends further)*
- *You can expect the best network performance for traffic within the same zone, using internal IP addresses.*
- *Remember about multi-NIC VMs (up to 8)*
- *Storage is a network resource! => Network bandwidth shared between network AND disk activity*



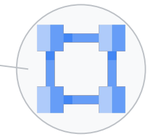
ingress

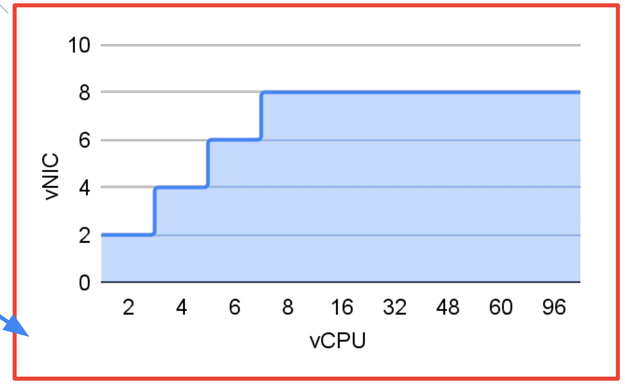
ingress

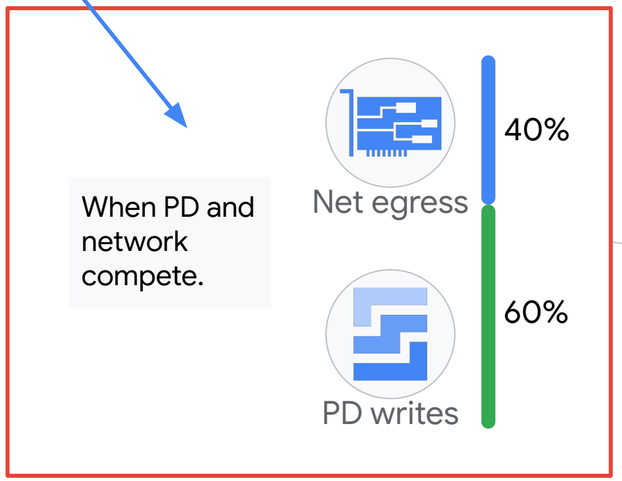


egress

egress



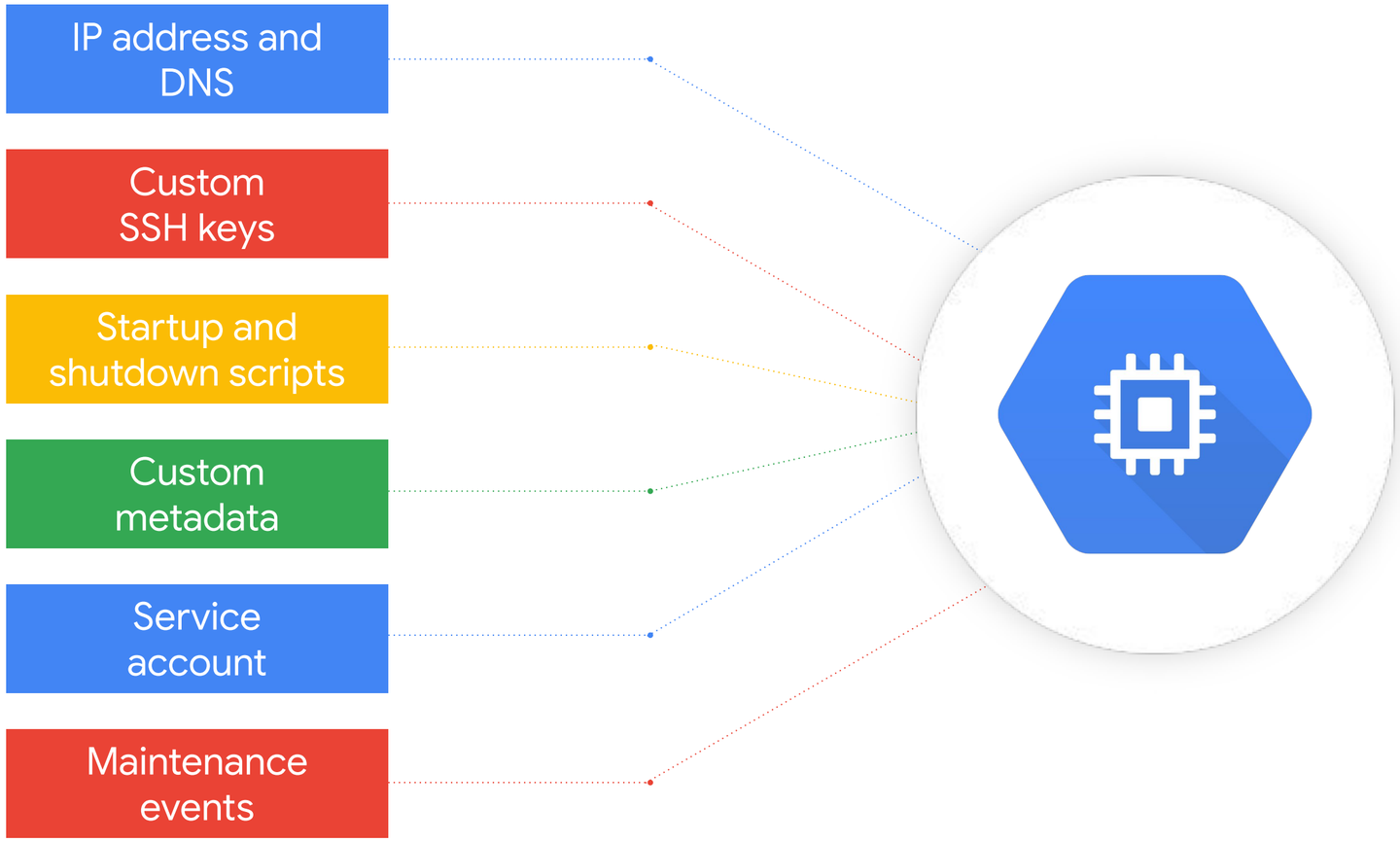




Compute Engine: Metadata Server

- The metadata server stores information about the instance or project.

  - Metadata request/response never leaves the physical host.
  - Metadata information is encrypted on the way to the virtual machine host.
  - Metadata server can generate a signed token for apps to verify the instance identity.



Compute Engine: Spot (Preemptible) VMs

Made for batch, fault-tolerant, and high throughput computing

Super-low-cost, short-term instances

- Up to 91% less than standard instances
- No maximum duration, may be preempted with 30-seconds notice (preemptible: max 24 hrs)
- Simple to use with graceful termination

Exam Tips:

- *Those use-cases usually pop up at the exam with regards to Spot VMs / Preemptibles.*
- *Can also be used in GKE clusters!*

Ideal for a variety of stateless, fault-tolerant workloads

- Genomics, pharmaceuticals
- Physics, math, computational chemistry
- Data processing (for example, with Hadoop or Cloud DataProc)
- Image handling, rendering, and media transcoding
- Monte Carlo simulations
- Financial services

Compute Engine: automate start & stop activities

Executed from metadata, either directly or from file:

- Startup:
  - gcloud compute instances create VM_NAME \
    --image-project=debian-cloud \
    --image-family=debian-10 \
    --metadata=startup-script='#! /bin/bash
    apt update
    apt -y install apache2
    cat <<EOF > /var/www/html/index.html
    <html><body> Linux startup script added directly. </body></html>
    EOF'

- Shutdown:
  - gcloud compute instances create example-instance --metadata-from-file=shutdown-script=FILE_PATH

To see output of startup/shutdown script:

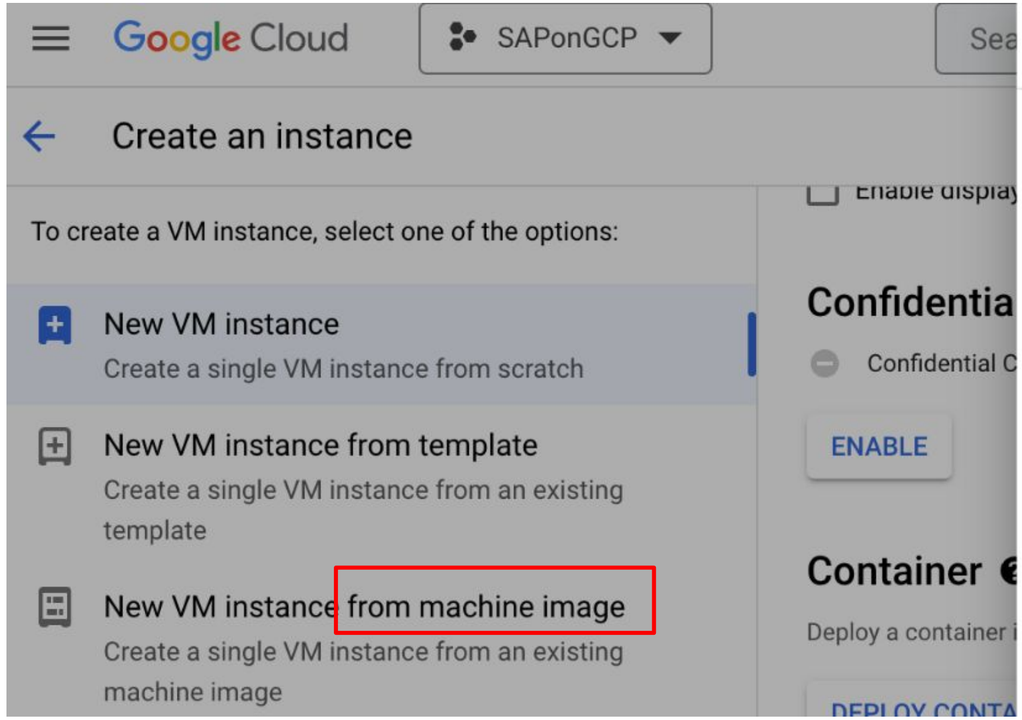
- gcloud compute instances create example-instance --metadata shutdown-script="#! /bin/bash
  > # Shuts down Apache server
  > /etc/init.d/apache2 stop"

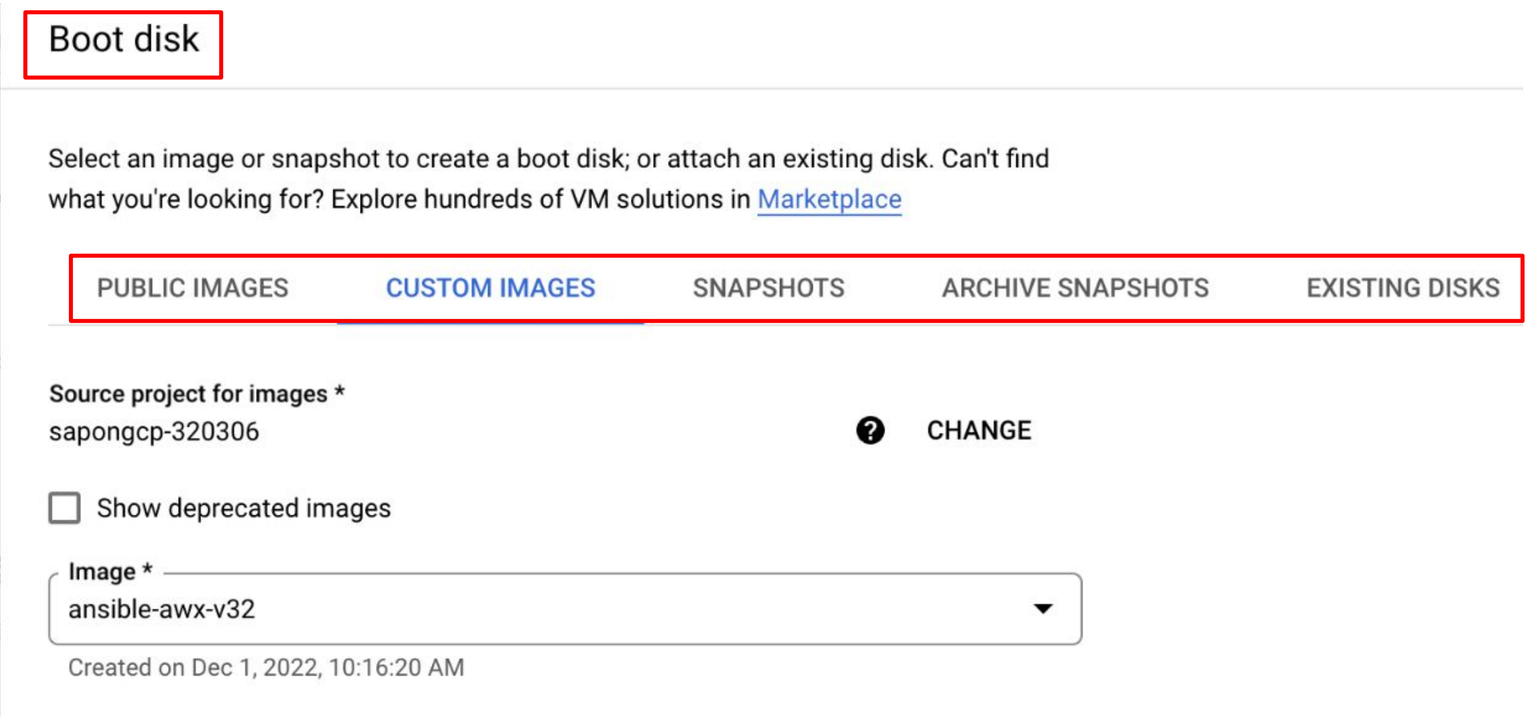
Exam Tips:

- *Startup / shutdown scripts are best-effort only!*
- *Startup / shutdown scripts are always run by root (Linux) / System (Windows)*
- *Shutdown scripts are especially useful for:*
  - *MIGs (to copy back processed data or logs before a VM goes down).*
  - *Spot / Preemptible VMs, which are much more vulnerable to be stopped.*
- *Startup / shutdown scripts can be set on VM or project (!!!) level -> will trigger for every VM. VM-level always take precedence (if exists, project-level script is not executed)*
- *Shutdown scripts have timeouts:*
  - *90s for standard instances*
  - *30s for Spot / Preemptible instances*

Compute Engine creation

public OS image vs custom OS image vs snapshot vs machine image





Exam Tips:

- *Custom images should be centralized and controlled from lifecycle perspective (know what are image families and image states*
- *Public / Custom OS image IS NOT the same as "machine image"*
- *You can create a VM based on all of those options (public / custom OS image, snapshot, existing disk, machine image)*
- *You can 'automate' post-processing with startup script, regardless of how boot disk was created.*

Shielded VMs

Exam Tips: *Using Shielded VMs is a best practice in GCP!*

Secure Boot	vTPM	Integrity Monitoring	Result/implications
ON	ON	ON	Most secure. Allows for use of vTPM for data encryption using vTPM protected key, Secure Boot to prevent malicious rootkits and bootkits, and Integrity Monitoring to alert to any changes in boot process. Secure Boot may not be compatible with customers drivers or other software.
OFF	ON	ON	Default when creating a GCP VM. Allows for use of vTPM for data encryption using vTPM protected key and Integrity Monitoring to alert to any changes in boot process. If customer has unsigned drivers or low level software this is the most secure option as Secure Boot would not be compatible.
OFF	OFF	OFF	Least secure. No benefits of Shielded VM. This is not recommended.

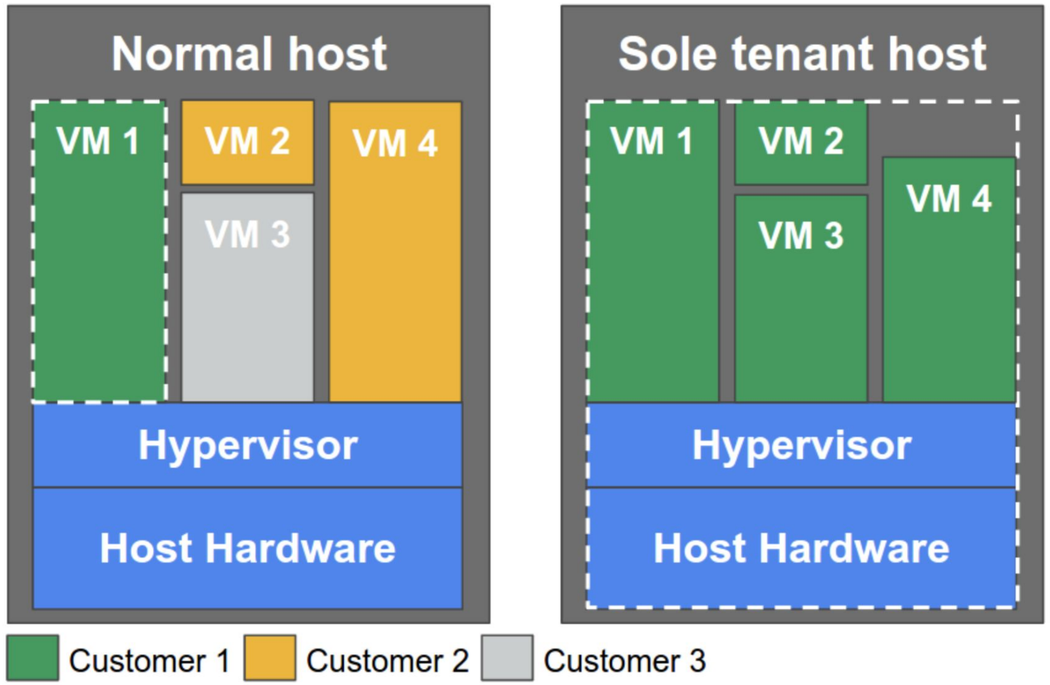
```
gcloud compute instances update instance name \
```

Feature	Flag to Turn On	Flag to Turn Off
Secure boot	--shielded-secure-boot	--no-shielded-secure-boot
vTPM (measure boot)	--shielded-vtpm	--no-shielded-vtpm
Integrity monitoring	--shielded-integrity-monitoring	--no-shielded-integrity-monitoring

Sole-Tenant Nodes

Regular VMs on regular machines, dedicated specifically to your workloads.

- Dedicated hardware
- Mix-and-match VMs to consume host resources
- Full access to host resources for 10% premium*

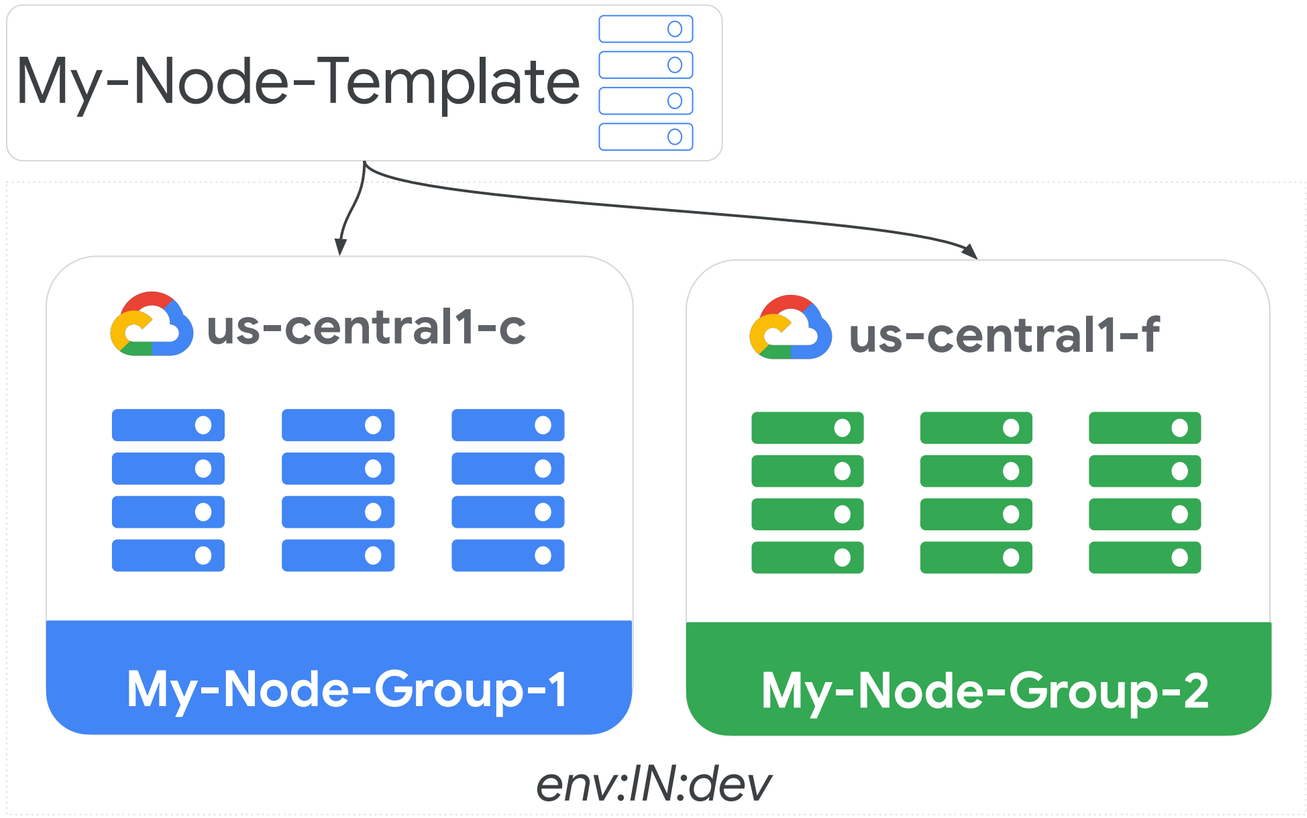


*10% Premium based on on-demand price

Quick Start for Sole-Tenant Nodes (1/2)

Each Sole-Tenant Node has a 1:1 mapping to a physical Host and represents a reserved host.

Step 1: Reserve Sole-Tenant Node(s)




```
// 1. CREATE NODE TEMPLATE
$ gcloud compute sole-tenancy \
node-templates create my-node-template \
--node-type n1-node-96-624 \
--region us-central1
```

```
// 2. CREATE NODE GROUP OF 3 NODES
$ gcloud compute sole-tenancy \
node-groups create my-node-group-1 \
--node-template my-node-template \
--target-size 3 \
--zone us-central1-c
```

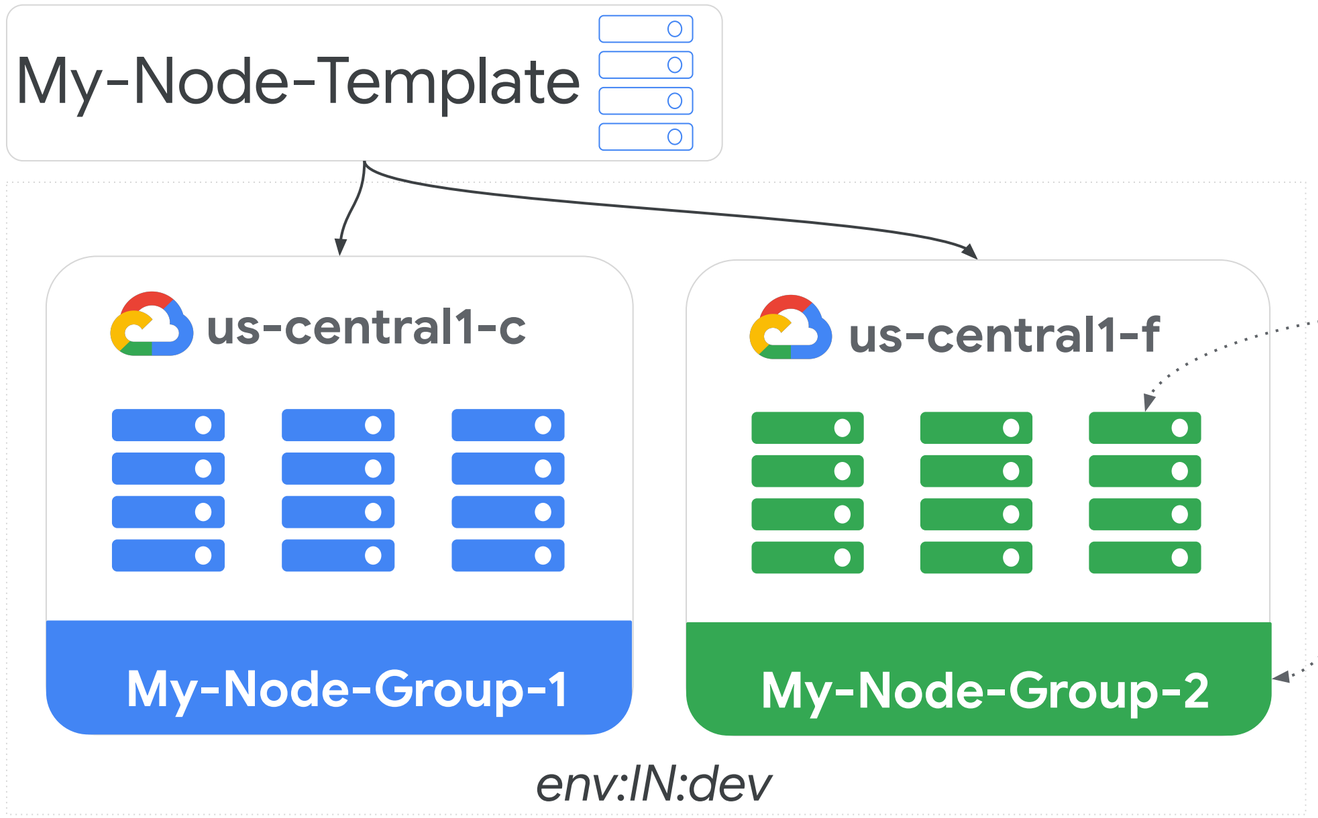
```
// 2b. [FOR ILLUSTRATION] CREATE ANOTHER
$ gcloud compute sole-tenancy \
node-groups create my-node-group-2 \
--node-template my-node-template \
--target-size 3 \
--zone us-central1-f
```

Google Cloud

Quick Start for Sole-Tenant Nodes (2/2)

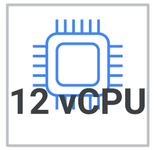
Exam Tips: *More info on provisioning VMs on sole-tenant nodes can be found here.*

Step 1: Reserve Sole-Tenant Node(s)



2. Schedule Instance(s)

VM



3 ways to schedule:

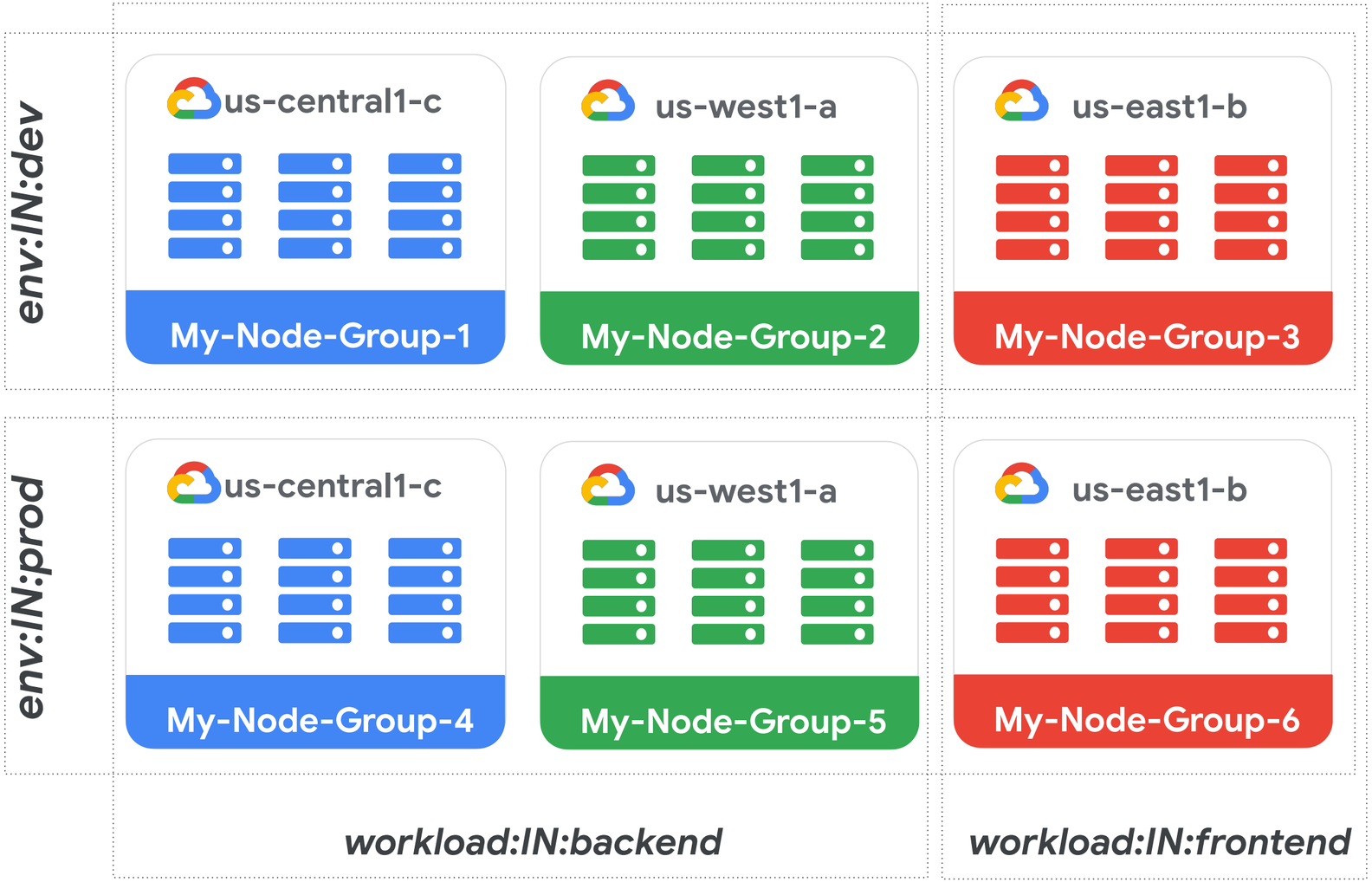
```
// SCHEDULE ONTO A SPECIFIC NODE
$ gcloud compute instances create \
INSTANCE_NAME --node=NODE_NAME
```

```
// SCHEDULE ON ANY NODE IN NODE GROUP
$ gcloud compute instances create \
INSTANCE_NAME \
--node-group=NODE_GROUP_NAME
```

```
// SCHEDULE ONTO ANY HOST WITH
// MATCHING LABELS
$ gcloud compute instances create
INSTANCE_NAME --zone ZONE
--node-affinity-file=FILE.json
```

Google Cloud

Sole-Tenant Nodes: Using Node Affinity Labels

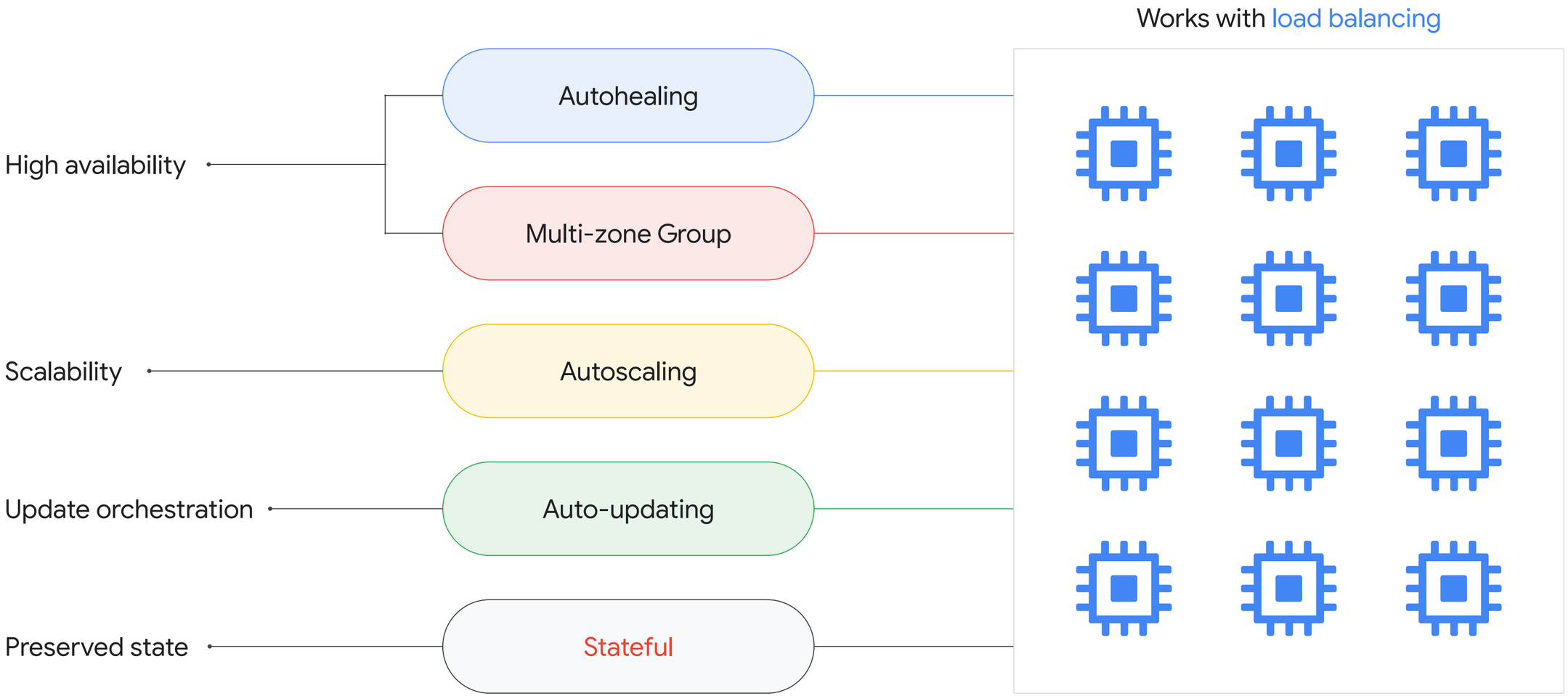



```
// SCHEDULE ONTO ANY HOST WITH
// MATCHING LABELS
$ gcloud compute instances create
INSTANCE_NAME --zone ZONE
--node-affinity-file=FILE.json
```

```
// SCHEDULE ONTO ANY HOST WITH
// IN MY-NODE-GROUP-1
$ gcloud compute instances create
INSTANCE_NAME --zone ZONE
--node-group my-node-group-1
```

Managed Instance Groups: Run VMs at Scale

Up to thousands of VMs



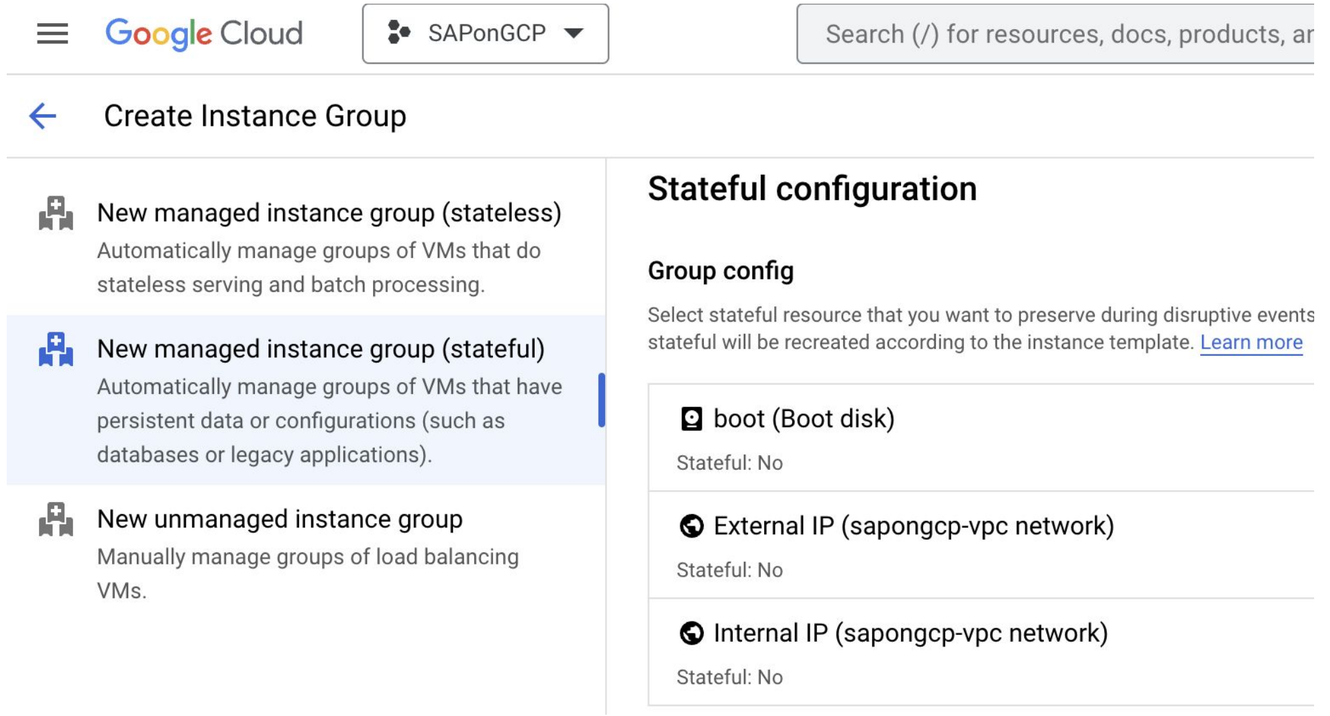
Exam Tips: *pros & cons of "ready" custom OS image vs public image + startup scripts*

Stateful vs stateless

And why stateless is usually preferred...

Exam Tips:

- *Here a look at this document.*
- *Prefer stateless. Use stateful only when necessary, eg:*
  - *Databases*
  - *Data processing apps (Kafka etc)*
  - *Legacy monoliths*



Stateful	Stateless (PREFERRED!)
Each server retains information about it's client sessions, such as the current state of an application or the content of a user shopping cart.	Server does not retain any information about the client sessions. Each request made by the client is treated as an independent transaction, and the server does not maintain any memory of previous requests.
Not perfect if we have multiple backends that can serve the requests...	Greater scalability and flexibility.
Can scale up easily. Can't scale down easily since each server keeps its' state.	Ability to scale up & down easily

Choosing instance groups for Compute Engine

Type of Instance Group		Properties of Instances	Feature
Unmanaged		Heterogeneous	
Managed		Homogeneous	Instance Templates Autoscaling
	Zonal	Same zone	Latency consistency
	Regional	Different zones	Reliability

Exam Tip:

- *Unmanaged are used to group EXISTING, different VMs under one "umbrella" and balance traffic to healthy ones only. For example, used in lift&shift migrations.*
- *You can't update existing instance template (need to create a new one)*
- *Know the difference between scale-out and scale-up!*

```
gcloud compute instance-groups managed create [*INSTANCE_GROUP_NAME*] \
    --size=[*SIZE*] \
    --template=[*INSTANCE_TEMPLATE_NAME*] \
    --zone=[*ZONE*]
```

MIG - Autoscaling

CPU Utilization

Treats the target CPU utilization level as a fraction of the average use of all vCPUs over time in the instance group

External HTTPS Capacity

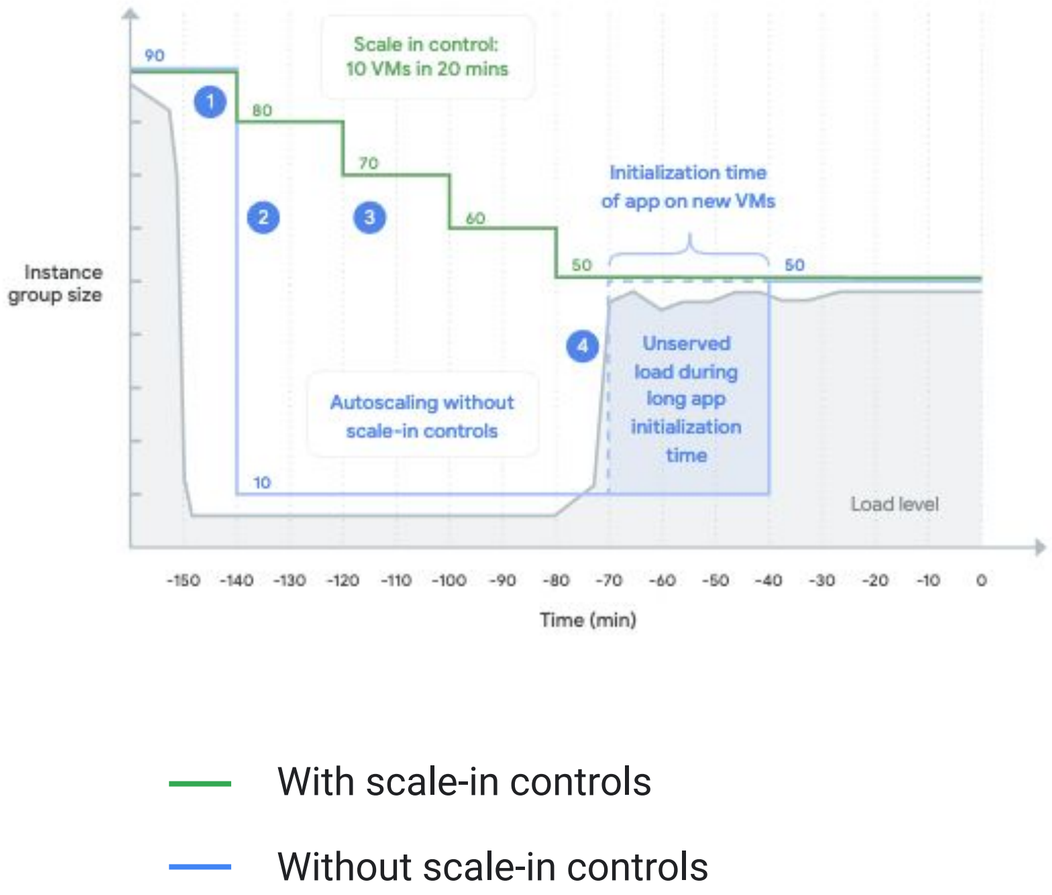
Autoscaling works with maximum backend utilization and maximum requests per second/instance

Cloud Monitoring Metrics

Per Instance or Per Group
Standard or custom metrics
Not for log-based metrics

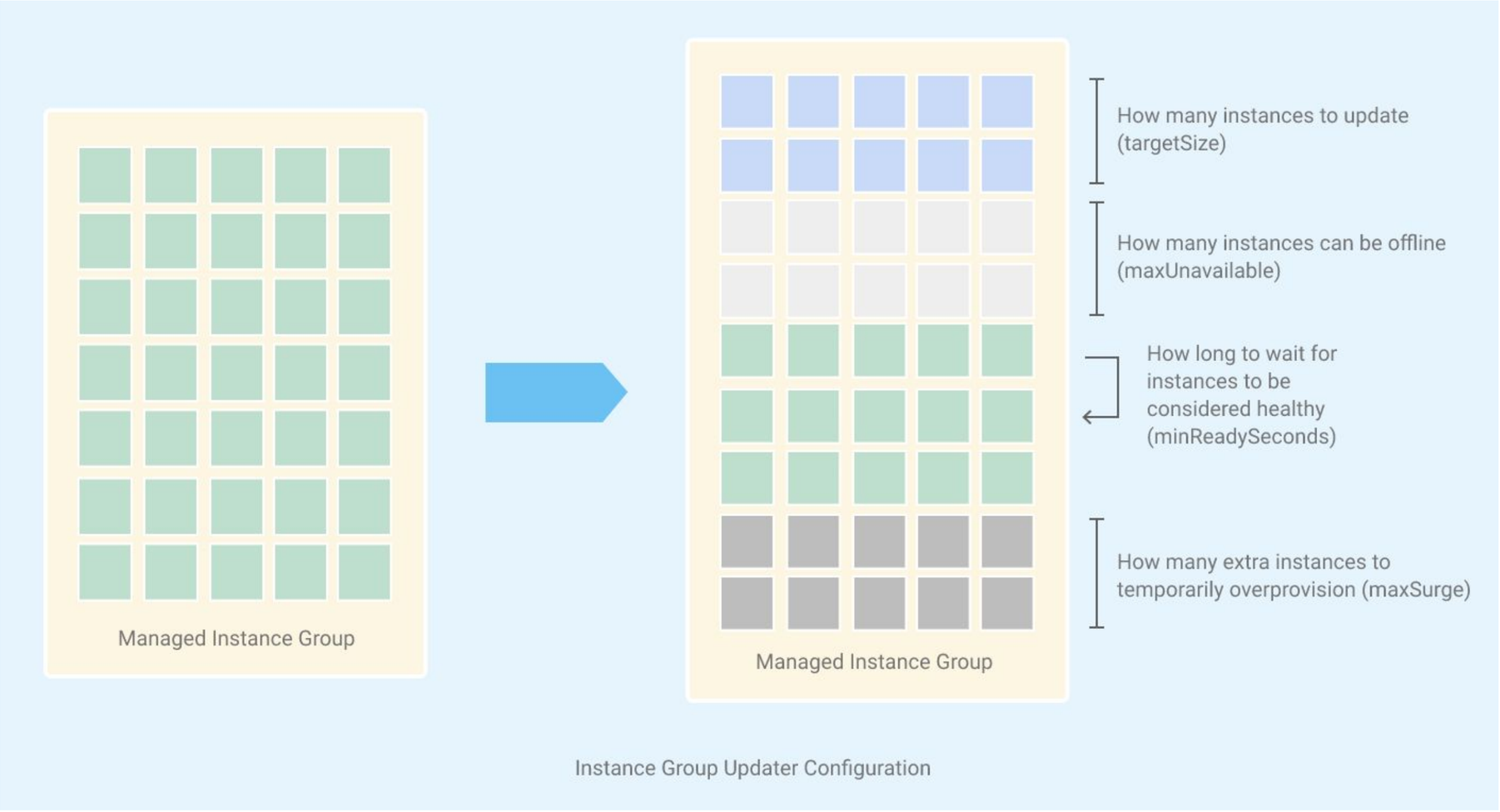
Schedules

Additional autoscaler
Up to 128 schedules
Min instances
Duration
Start time & Recurrence



Updating MIGs

= implementing new image versions



Exam Tip: *Know WELL how to rollout new versions to MIGs, incl. canary & rollback strategies*

VM Pricing and cost optimization

Sustained Use Discounts (SUD)

Up to 30% savings on Compute Engine and Cloud SQL

Committed Use Discounts (CUD)

Up to 70% savings without upfront fees or instance-type lock-in

Spot / Preemptible VM instances

Up to 91% savings on workloads that can be interrupted, like data mining and data processing

Per second billing

Up to 38% savings by paying per second, not per hour

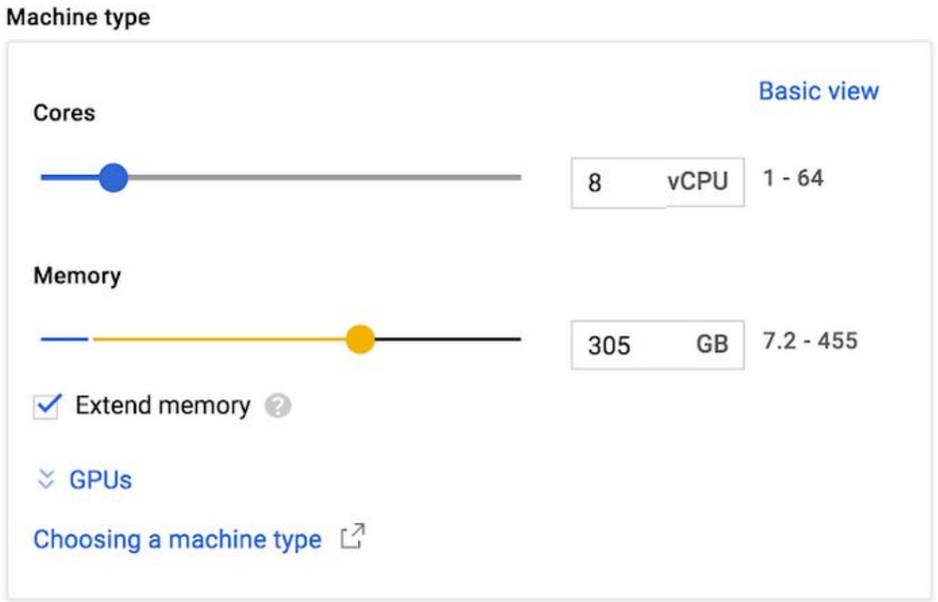
Network Service Tiers

Pick performance and get 70% more bandwidth than other clouds, or pick cost savings and save up to 9% compared to other clouds

Rightsizing (incl. choosing optimal GCE families) and Custom Machine Types

Exam Tips:

- Common pattern for optimization costs for unused PDs: you can create a snapshot, and delete the disk to reduce the maintenance cost of that disk by 35% to 92%.
- For premium OS, you're billed for license per vCPU per second.
- Bring Your Own License is an option for some OSes
- Use Extended memory to save on OS license costs.



Migrate for Compute Engine

Lift&Shift your VMWare, AWS, Azure workloads to GCE

- Purpose-built, enterprise-grade
- Migrate from on-prem or other clouds
- Proven at scale, having migrated customers w/ thousands of workloads
- Success across healthcare, energy, government, manufacturing, and more

Agentless

Nothing to install on source machines

Minimize complexity, reduce IT labor requirements by 5+ hours per server, keep migrations on track.

Streaming

Migrate storage while apps run in GCP

Eliminate long upfront data transfers and unpredictable maintenance windows, enabling fast time-to-cloud and reduced downtime.

Frictionless

Automate migration and in-cloud conversion

Reduce touch points for IT, provide uninterrupted experience for line of business owners and end users.

Persistent Disks best practices, tips&tricks

Exam Tips:

- Use "**--no-boot-disk-auto-delete**" parameter if you don't want boot / OS disk to be deleted if a VM gets deleted.
- CMEK and CSEK can be used to encrypt PDs. Have a look at how to use CSEK for a PD.
- Avoid using ext3 filesystems in Linux (poor performance under heavy write loads). Prefer ext4.
- You can share a PD across multiple VMs at the same time in **read-only mode**.
    - If read-write required, prefer a managed solution such as Filestore or utilize GCS.
- You can share a SSD PD across two N2 VMs at the same time in read-write mode. In Preview as of Q1 '23 -> should NOT be covered on the exam.
- To recover from a corrupted disk / Os not booting properly, follow this procedure.
    - On high-level, just attach the corrupted disk as non-boot disk to another VM and troubleshoot.
- For special use-cases (app needs a RAM disk with exceptionally low latency and high throughput and cannot just use the VM memory), you can create a tmpfs filesystem by allocating some VM memory as a RAM disk.
- If needed, you can attach 1-24 **local SSDs** (ephemeral = data is lost if VM stops; each 375 GB and physically attached to the server that hosts your VM instance). **Local SSDs are NOT the same as PD SSDs**!!! More information here.

Compute Engine

#GCPSketchnotes

@PVERGADIA THECLOUDGIRL.DEV

What is Compute Engine?

CUSTOMIZABLE VIRTUAL MACHINES IN GOOGLE CLOUD

THERE ARE 2 MACHINE TYPES

PREDEFINED
n1-standard-1 =
(1 vCPU, 3.75 GB memory)

CUSTOM
vCPU + Memory

CORES — 1 vCPU — 1-96

MEMORY — 3.75 GB — 1-6.5

THERE ARE 3 MACHINE TYPE FAMILIES

GENERAL PURPOSE Machine Type
- General Servers
- Websites
- Databases

COMPUTE OPTIMIZED Machine Type
- High Performance Computing
- Gaming
- Electronic Design Automation
- Single Threaded Applications

MEMORY OPTIMIZED Machine Type
- Large In-Memory Databases (SAP HANNA)
- In-Memory Analytics

Compute Engine Use case (example)

- FIG. 1 — Websites
- FIG. 2 — Databases
- FIG. 3 — Legacy Monolithic Apps
- FIG. 4 — Containers
- FIG. 5 — Windows Apps

Compute Engine PRICING

SUSTAINED USE SAVINGS
Automatic discounts for running VMs a significant portion of the month

COMMITTED USE DISCOUNT
Up to 57% savings with no up-front cost

PREEMPTIBLE VMs
Up to 80% savings and run batch jobs & fault-tolerant workloads

RIGHT SIZE RECOMMENDATIONS
Suggests resizing for efficiency and cost

How does it WORK??

CREATE

region
+ zone
+ machine type (cpu & memory)
= Instance

BACKUPS

- Automatic Scheduled snapshots
- Live migration — Keep apps running during maintenance

HIGH AVAILABILITY

- Automatic failover to another zone or region (ZONE A, AUTOMATIC FAILOVERS, ZONE B)
- MANAGED INSTANCE GROUPS (MIGs)
  - Autohealing
  - Load balancing
  - Automatic updates
  - Support for stateful workloads
  - Container support
- AUTOSCALER - 3 types of policies:
  1. CPU utilization = more than 60% → create new instance
  2. HTTP(S) load balancers service capacity — Requests per second or utilization
  3. Cloud monitoring metrics

Persistent Disk

#GCPSketchnotes

@PVERGADIA THECLOUDGIRL.DEV

3.29.2020

What is Persistent Disk?

HIGH PERFORMANCE BLOCK STORAGE

- ✓ Transparently Resize
- ✓ Quick backup

Persistent Disk

SSD/HDD storage

Compute Engine

Kubernetes Engine

- ✓ Supports Simultaneous readers

Which Block storage is right for your app?

LOWER COST ← → PERFORMANCE

Standard PD	Balanced PD	Performance (PD-SSD)	Extreme PD	Local SSD
HDD	SSD	SSD	SSD	Local SSD
Standard throughput	Best price per GB	Best price per IOPS	Best performance	Lowest Latency
Current 'cost sensitive'	Best fit for most workloads	Performance sensitive workloads	Uncompromising performance	For caches and hot layers

(How to pick) based on availability needs.

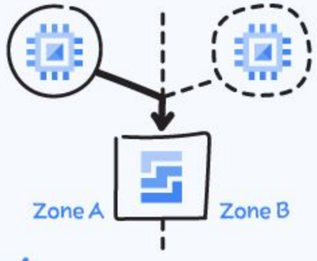
Local SSD

- ✓ Ephemeral
- ✓ Stateless workloads, or replication managed at app or database layer

Persistent Disk

- ✓ Durable, support snapshots
- ✓ Most workloads

Regional Persistent Disk

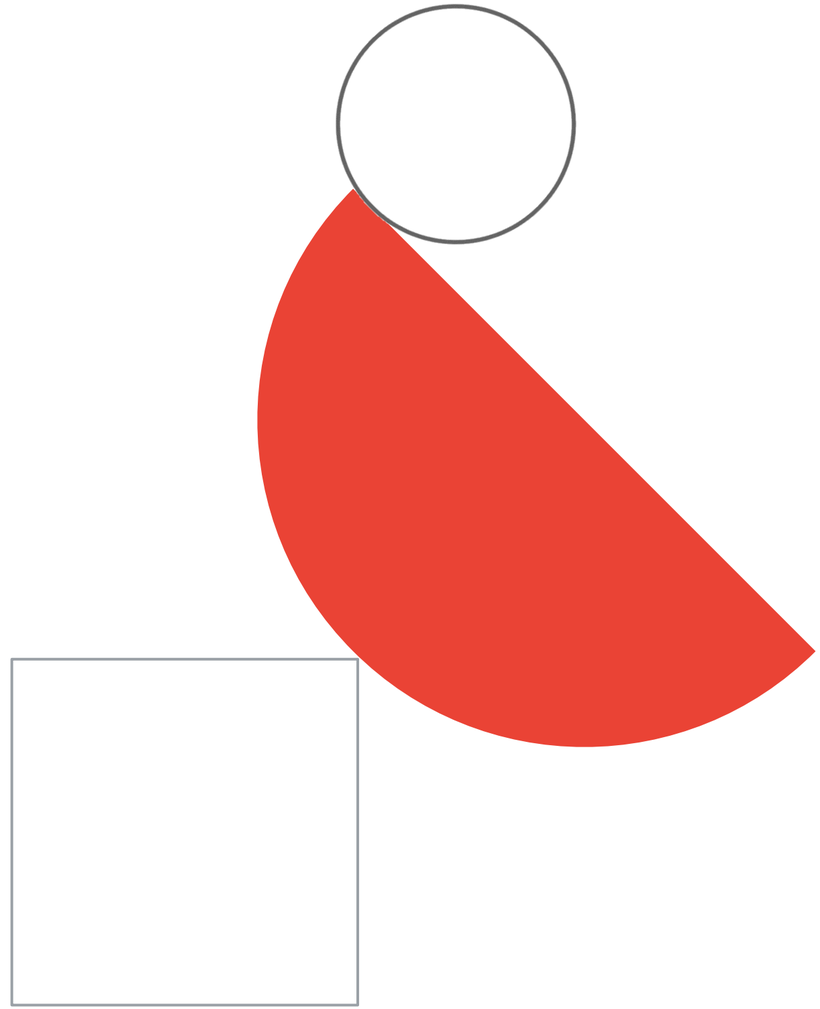


- ✓ Durable & Highly Available
- ✓ Mission-critical workloads with RPO/RTO near 0

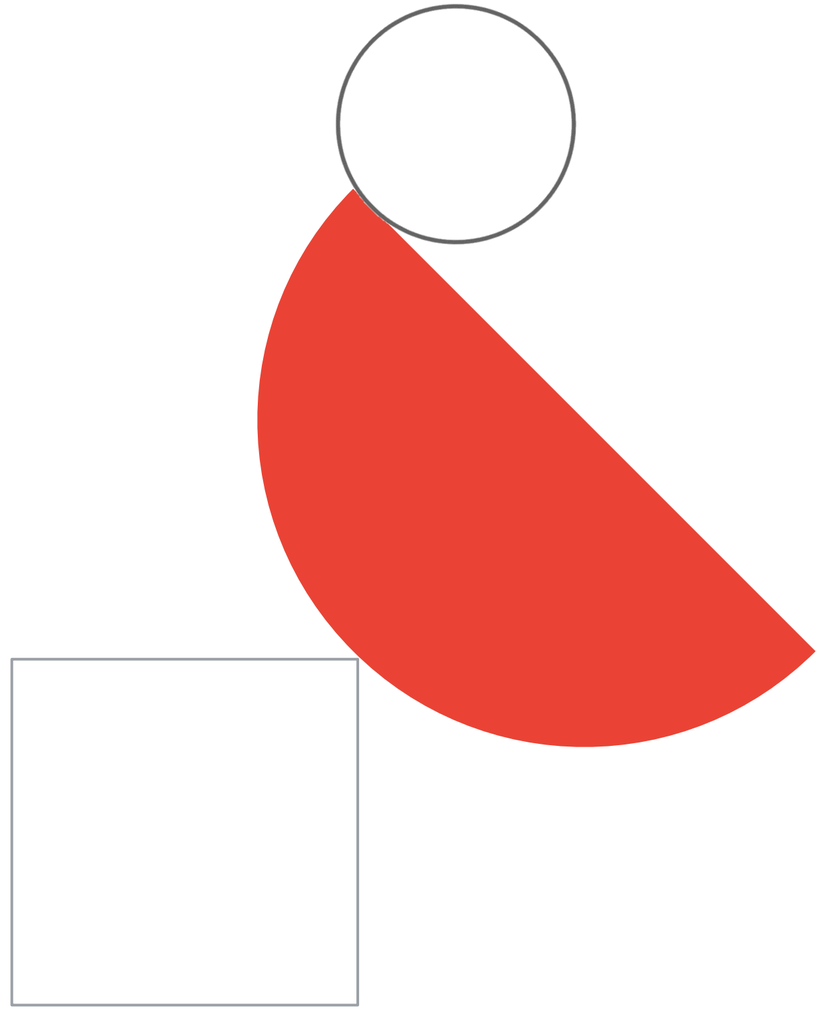
Persistent Disk «Use case example»

Standard PD	Balanced PD	Performance (PD-SSD)	Extreme PD	Local SSD
Cost sensitive workloads	Most enterprise apps	Most databases	SAP HANA	Scale out analytics
Scale out analytics (Hadoop, Kafka)	LOB apps	Persistent cache	Oracle	Media rendering
	Boot disks	Scale-out analytics	Largest in-memory DBs	Other use cases where ephemeral scratch space is required
	Webserving			

QUIZ time!



Cloud Dataproc



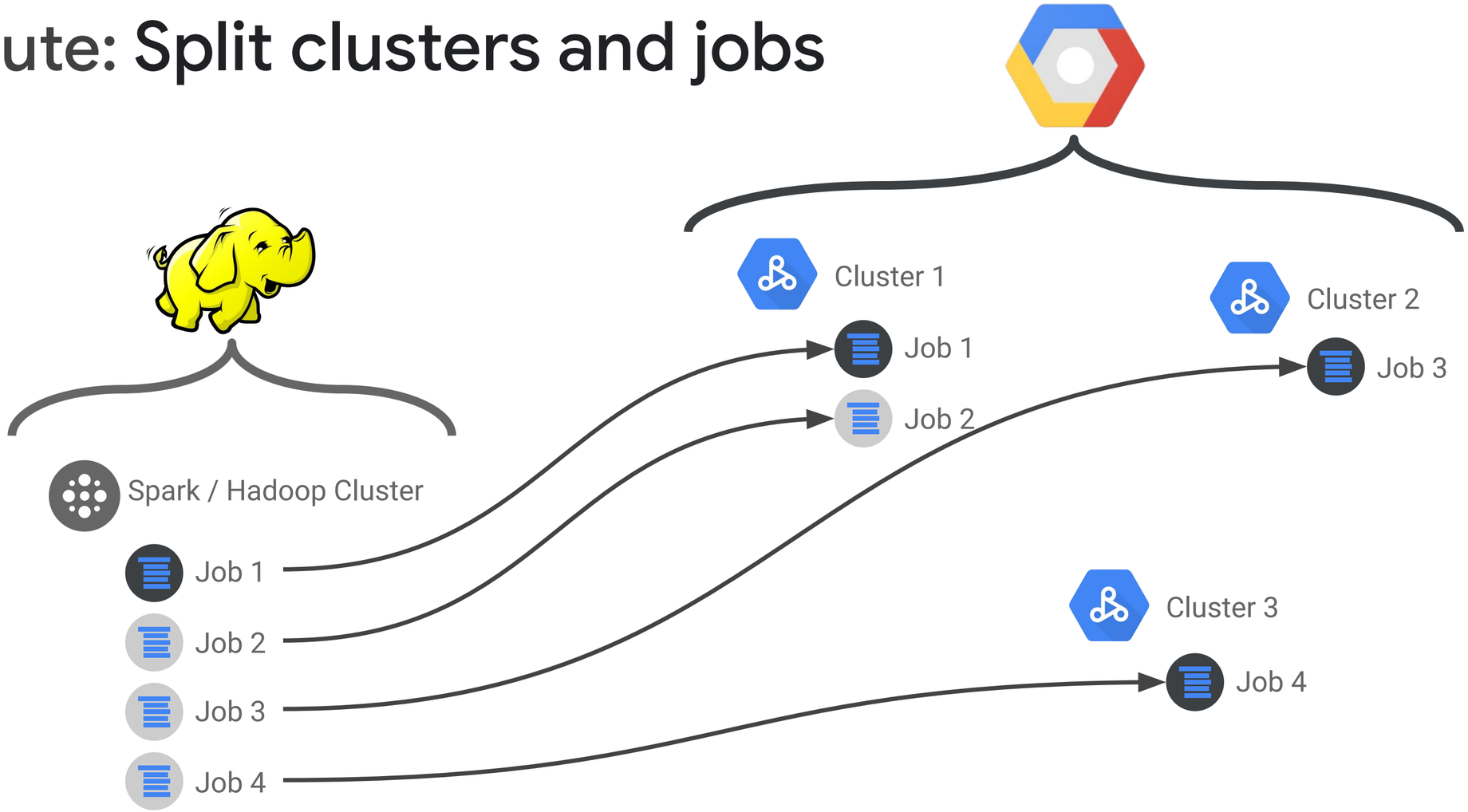
The benefits of Hadoop/Spark on Cloud

On premises	On compute engine	Cloud Dataproc
Custom code	Custom code	Custom code
Monitoring/Health	Monitoring/Health	Monitoring/Health
Dev integration	Dev integration	Dev integration
Scaling	Scaling	Scaling
Job submission	Job submission	Job submission
GCP connectivity	GCP connectivity	GCP connectivity
Deployment	Deployment	Deployment
Creation	Creation	Creation

Self-managed Google managed

Exam Tip: *if exam question mentions Apache Hadoop / Spark / Pig / Hive, plus it's clear that the customer already invested in building the pipelines in on-premises and does not want to lose it, you should probably go with Dataproc.*

Flexible compute: Split clusters and jobs



Exam Tips:

- *When thinking about **Dataproc**, you should really think about per-job, ephemeral, auto-scaling clusters with auto-shutdown after the task is completed.*
- *Using Spot/Preemptible VMs for **secondary** Dataproc workers is a common pattern.*
- *Switching from HDFS to GCS is also a best practice in most cases.*

Dataproc #GCPSketchnote

@PVERGADIA THECLOUDGIRL.DEV

12.07.2020

Managing Apache Hadoop cluster is hard!

ERIN

We need a tool that make it easy to gain insights from our data.

Yes, it is slow, difficult & expensive!

SAM

Dataproc!

How is a typical Hadoop/Spark cluster installation different from Dataproc?

Setup
Configure
Optimize
Utilize
Monitor
Shutdown

Typical Cluster Activities

90 seconds from start to finish

Dataproc

Simple deployment
Rapid development
Cost-effective use

Cluster Dataproc Activities

Familiar open source tools

Customizable machines
Easily autoscale up or down

Ephemeral clusters on-demand
Save cost

Tightly Integrated with
Google Cloud Platform

Rapid cluster creation
Spin up a cluster in 90 seconds

How do I move Spark/Hadoop jobs to Dataproc?

Copy Data → Update File Path → Go

GCS

from hdfs:// to gs://

Run your Spark, Hadoop, Hive, or Pig script

Can I right-size clusters per use case?

Storage
Cloud Storage

Hive Metastore
Cloud SQL

Dataproc	Dataproc	Dataproc	Dataproc
375GB local SSD	100GB PD	500GB PD-SSD	150GB PD
MapReduce ETL	Machine learning	Hive analytics	Business reporting

How do I use ephemeral & long-lived clusters to save cost?

Clusters per job

Cluster Dataproc
Cluster Dataproc
Cluster Dataproc
Cluster Dataproc

Client
Client
Client

Edge Nodes
Compute Engine

Storage
Cloud Storage

Semi-long-lived clusters - group and select by label

Production (1.2)
Prod 1 Dataproc
Prod 2 Dataproc

Development (Preview)
Dev cluster
Dataproc

Clients

How does it work?

Disaggregates storage & compute

External Applications
Application Logs

Data Sources
Storage Cloud Storage
Storage BigQuery
Storage Cloud Bigtable

Production Dataproc

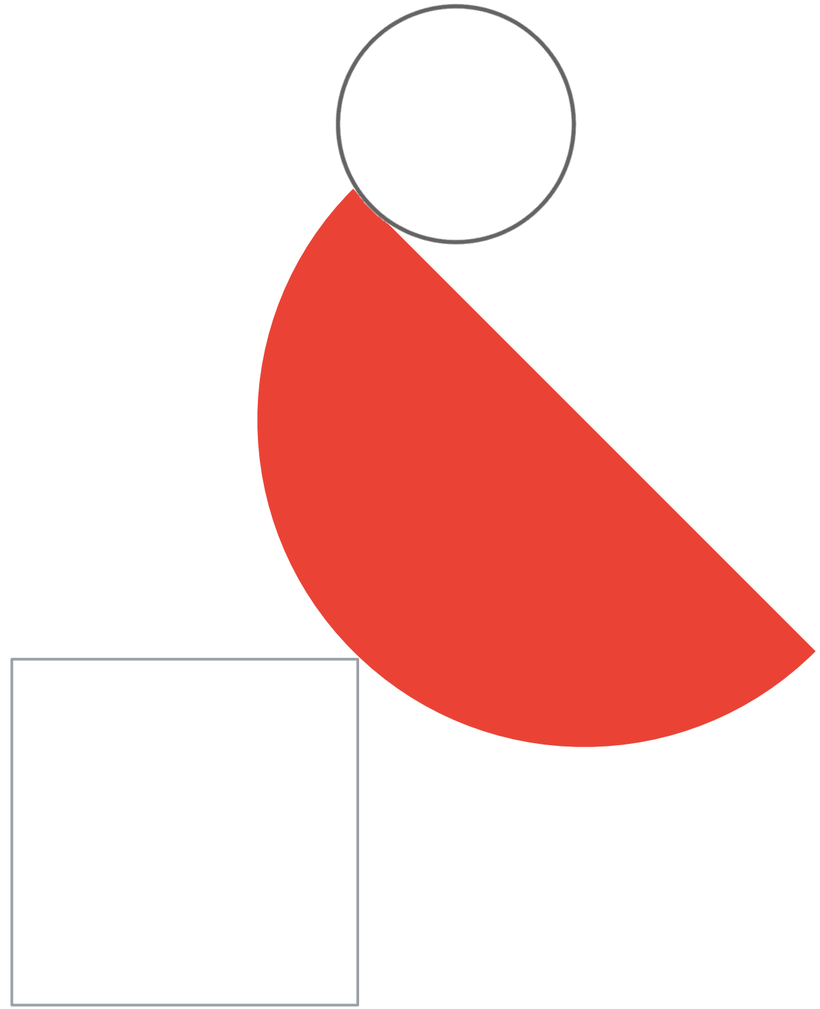
External Applications
Test Dataproc
Development Dataproc

Cluster monitoring
Monitor Stackdriver
Logs Logging

Data Sources
Storage Cloud Storage
Storage BigQuery
Storage Cloud Bigtable

Data science
Analysis Notebook

Cloud Dataflow



Cloud Dataflow

The fully-managed, serverless, auto-optimized data processor that simplifies development and management of stream and batch pipelines

Exam Tips:

- *if exam question mentions Apache Beam -> most probably answer is Dataflow.*
- *When you're starting from scratch with ETL, Dataflow is preferred over Dataproc!*
- *KEY thing about Dataflow: it's able to serve BOTH batch and streaming within a SINGLE pipeline.*

Stream Analytics

- Works with Cloud Pub/Sub to deliver stream analytics
- Real-time data processing with "exactly-once" semantic

Unified with Batch

- No more Lambda architecture
- Apache Beam provides unified batch & streaming
- Reuse skills, tools and code

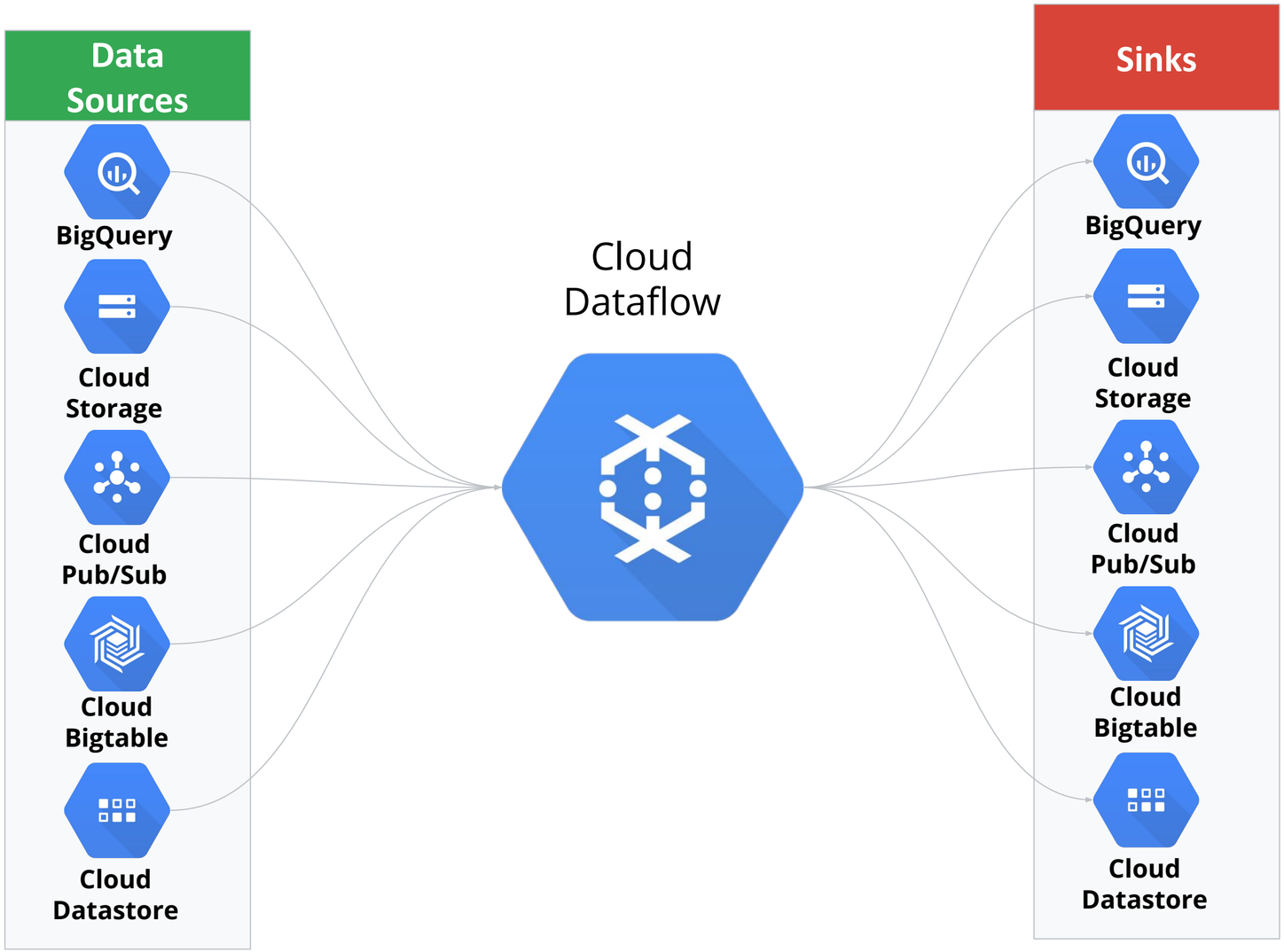
Open Source ensures Portability

- Pipelines written in Beam API are portable
- Runners include Dataflow, Flink, Samza and Spark

Auto-optimizations

- No more cluster management
- Submit a job and Dataflow auto-optimizes resources
- Makes pipelines faster and cost-effective

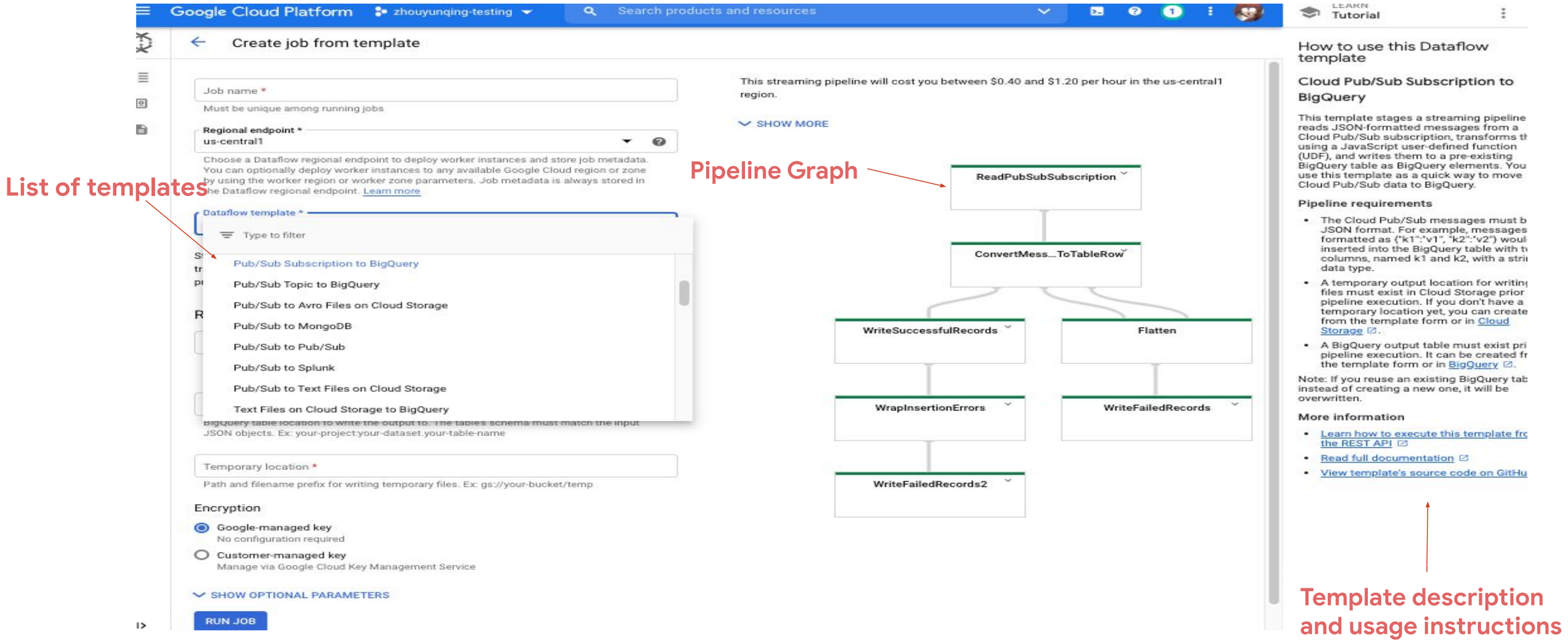
Data Sources and Sinks for Dataflow



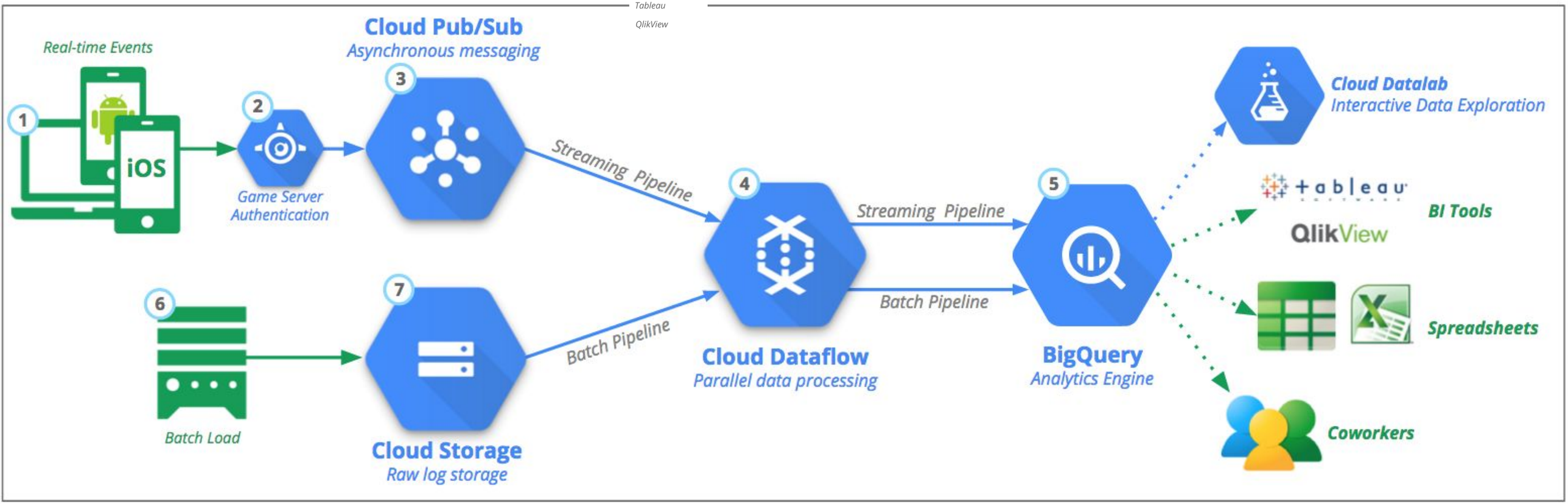
Exam Tips:

- *Dataflow does NOT store data! There is always a Source and a Sink*

Dataflow: Google Provides Templates for different use-cases



Example architecture for data analytics



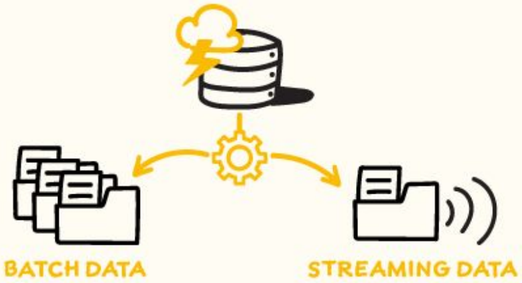
Dataflow

#GCPSketchnote

@PVERGADIA THECLOUDGIRL.DEV

What is Dataflow?

SERVERLESS, DATA PROCESSING SERVICE FOR BOTH STREAMING AND BATCH DATA



CREATE DATA PROCESSING PIPELINE

Apache Beam (SDK)

DEPLOY & EXECUTE

Dataflow

AUTOMATE
Operations & Management

Data Pipeline → Dataflow Job

How does it WORK?

DATA PROCESSING PIPELINE

STEP 1
Read Data Source

Files, BigQuery, Bigtable, Pub/Sub, Custom Source

STEP 2
Transform Data

PCollection 1 → PCollection 2 → PCollection 3

STEP 3
Write Data Sink

Files, BigQuery, Bigtable, Pub/Sub, Custom Source

Dataflow

WORKER VIRTUAL MACHINES EXECUTE THE DATA PROCESSING

Worker Virtual Machines

Autoscale up or down dynamically as needed

Dataflow Streaming Engine

Separate Compute | Separate Storage

BATCH PIPELINES SCALE SEAMLESSLY, WITHOUT ANY TUNING REQUIRED.

Dataflow → FEATURE Service-based Dataflow Shuffle → Data Grouping & Joining → Scale to 100s of TBs

HOW TO USE Dataflow

1 DATAFLOW SQL VIA BIGQUERY

- Use SQL from BigQuery web UI
- Read from pub/sub, cloud storage or BigQuery
- Write to BigQuery

2 DATAFLOW TEMPLATES

- Share pipeline with teams
- Easy & repeatable Pre built templates

3 AI PLATFORM NOTEBOOKS

- Use latest data science and machine learning frameworks

SECURITY for Dataflow Pipelines

EXACTLY-ONCE PROCESSING

0 Avoid risk of data loss

1 Turn off public IPs

2 VPC controls

3 Customer managed encryption key (CMEK)

Dataflow PRICING

DATAFLOW
Per second billing

DATAFLOW JOB
Batch or Streaming
(FlexRS) flexible resource scheduling = low cost

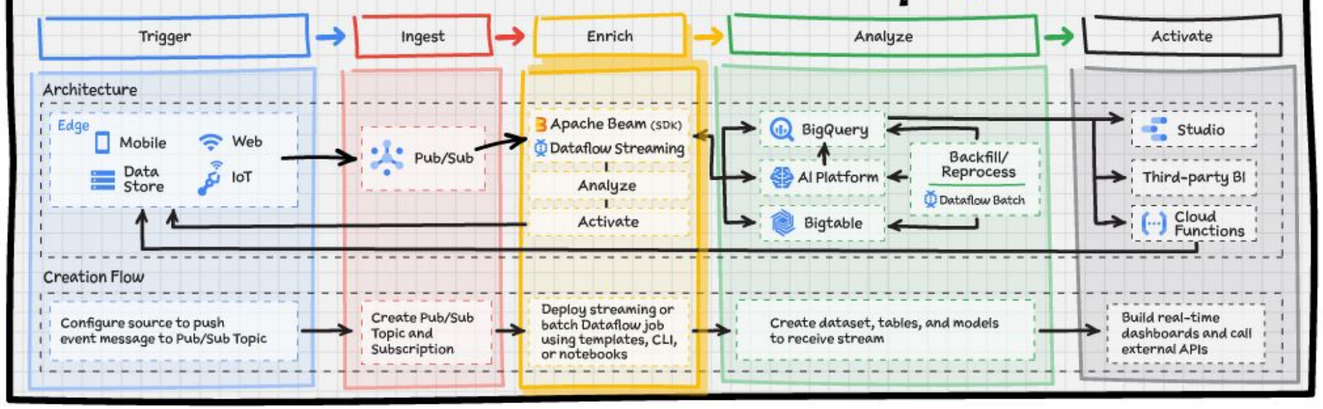
AUTOSCALING
Worker(s)
Better price to performance

CUSTOMIZABLE
vCPU + Memory + Storage
Each billed on per second basis

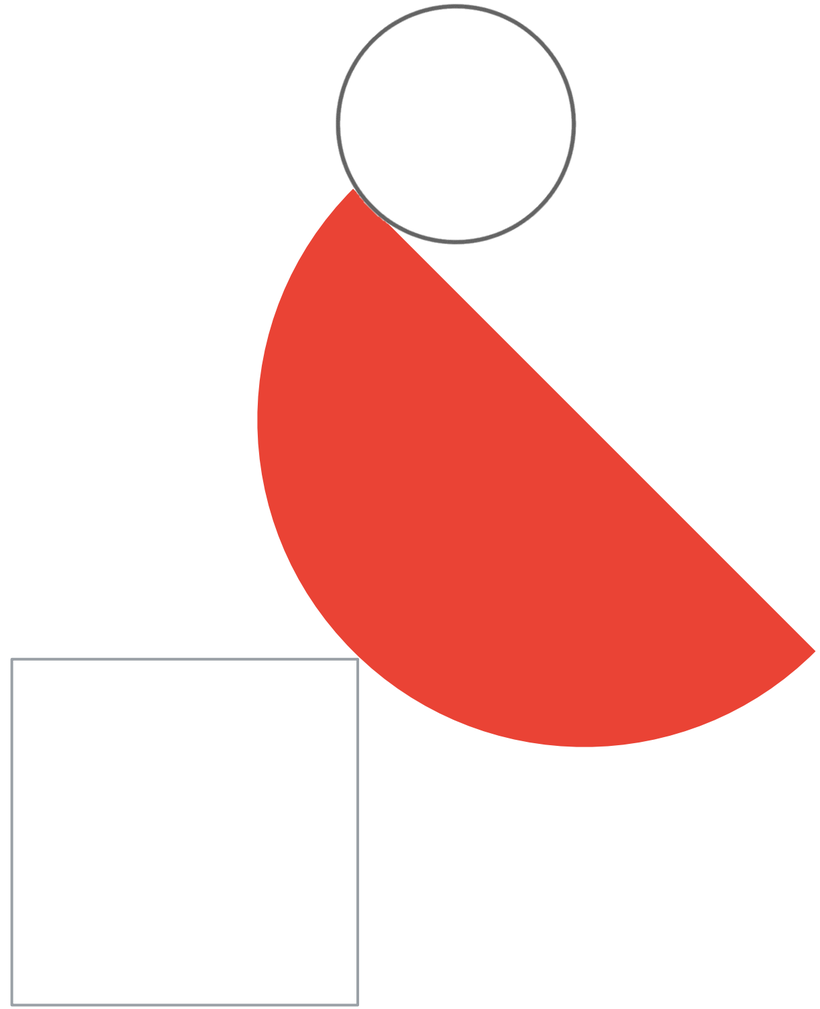
Dataflow Use case example

Streaming analytics

Real-time AI



[optional] Links to useful materials



Optional materials 1

[READING]

- get a feeling of the differences between PD snapshots, images and machine images (important from exam perspective: it's good to know what is global/regional, what can be used to create VMs, how to share those resources between projects / regions etc).
- What is GCP metadata server?
- Sole-tenant nodes
- How stateful workloads are different from stateless workloads
- How to achieve HA with Regional Persistent Disks and what a "--force-attach" is.
- Image management best practices | Compute Engine Documentation | Google Cloud
- Best practices for persistent disk snapshots | Compute Engine Documentation | Google Cloud
- Encrypt disks with customer-supplied encryption keys | Compute Engine Documentation | Google Cloud

Optional materials 2

[VIDEOS]

- Networking 102 (Cloud Routing and VPC Peering): Cloud OnAir: CE Chat: Google Cloud Networking 102 - Cloud Routing and VPC Peering
- What is Persistent Disk?: What is Persistent Disk? #GCPSketchnote
- Introduction to Virtual Machines (Next '19): Introduction to Virtual Machines (Cloud Next '19)
- Best Practices for GCE: Best Practices for GCE Enterprise Deployments (Cloud Next '19)
- VM Manager overview: What is VM Manager?
- GCE Managed Instance Groups: Using managed instance groups
- All you need to know about Migrate for Compute Engine: Migrate for Compute Engine
- Effective autoscaling with Managed Instance Groups: Effective autoscaling with managed instance groups
- Shared VPC: Level Up From Zero Episode 4: Shared VPC
- BeyondCorp overview: BeyondCorp Enterprise in a minute
- App Engine introduction: Get to know Google App Engine
- Pub/Sub overview: Cloud Pub/Sub Overview - ep. 1
- Cloud security basics: Top 3 access risks in Cloud Security
- What is DNS?: What is DNS? | How a DNS Server (Domain Name System) works | DNS Explained

Optional materials 3

[PODCASTS]

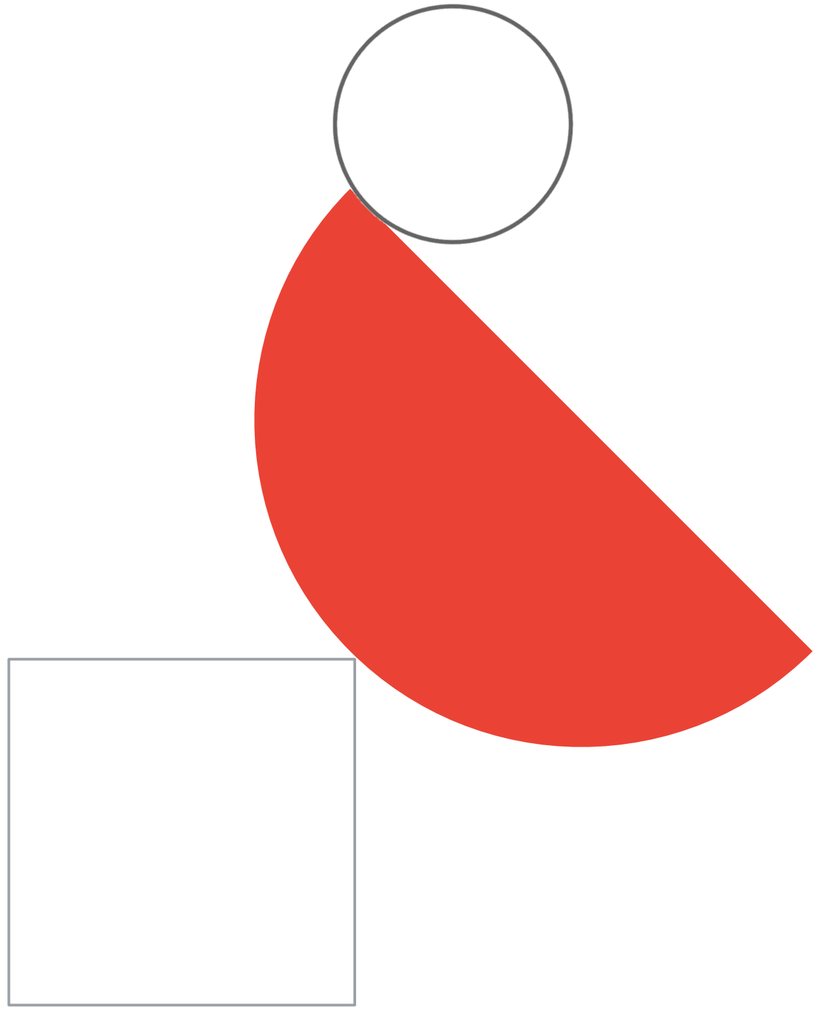
- Firestore intro, plus differences between SQL and NoSQL databases
- BeyondCorp

[DEEP DIVES]

- What is envelope encryption?
- Stateful Managed Instance Groups
- Key Management Service deep dive
- BeyondProd security model (evolution of BeyondCorp model)

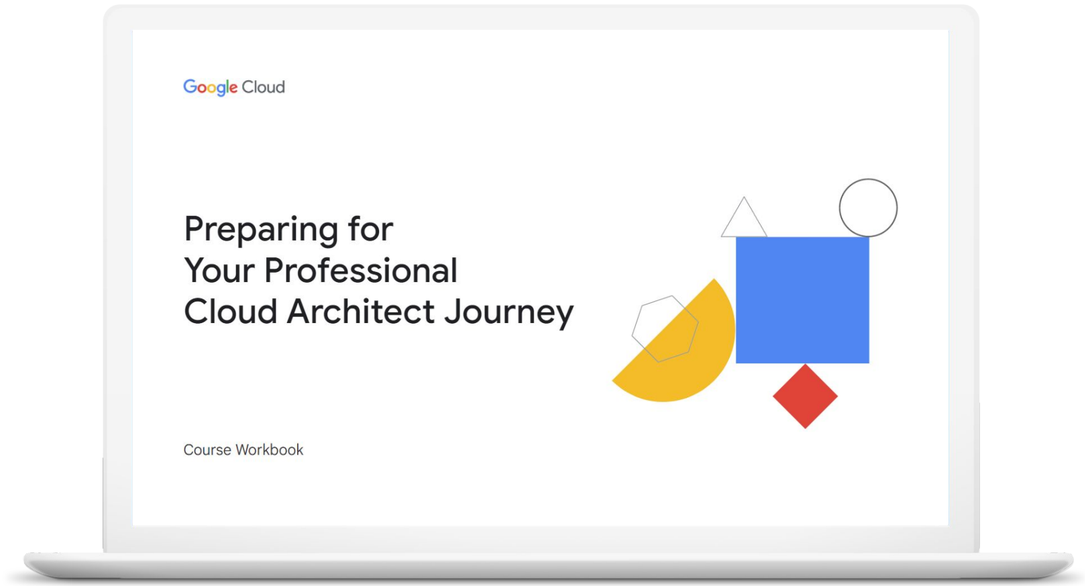
Diagnostic Questions

for Exam Guide Section 1: Designing and planning a cloud solution architecture



PCA Exam Guide Section 1:

Designing and planning a cloud solution architecture



1.1 Designing a solution infrastructure that meets business requirements

1.2 Designing a solution infrastructure that meets technical requirements

1.3 Designing network, storage, and compute resources

1.4 Creating a migration plan

1.5 Envisioning future solution improvements

1.1 Designing a solution infrastructure that meets business requirements

Considerations include:

- Business use cases and product strategy
- Cost optimization
- Supporting the application design
- Integration with external systems
- Movement of data
- Design decision trade-offs
- Build, buy, modify, or deprecate
- Success measurements (e.g., key performance indicators [KPI], return on investment [ROI], metrics)
- Compliance and observability

1.1 Diagnostic Question 01 Discussion

Cymbal Direct drones continuously send data during deliveries. You need to process and analyze the incoming telemetry data. **After processing, the data should be retained, but it will only be accessed once every month or two.** Your CIO has issued a directive to incorporate managed services wherever possible. You want a **cost-effective solution to process the incoming streams of data.**

What should you do?

A. Ingest data with IoT Core, process it with **Dataprep,** and store it in a **Coldline Cloud Storage bucket.**

B. Ingest data with IoT Core, and then publish to Pub/Sub. Use **Dataflow** to process the data, and store it in a **Nearline Cloud Storage bucket.**

C. Ingest data with IoT Core, and then publish to Pub/Sub. Use **BigQuery** to process the data, and store it in a **Standard Cloud Storage bucket.**

D. Ingest data with IoT Core, and then store it in **BigQuery.**

1.1 Diagnostic Question 01 Discussion

Cymbal Direct drones continuously send data during deliveries. You need to process and analyze the incoming telemetry data. **After processing, the data should be retained, but it will only be accessed once every month or two.** Your CIO has issued a directive to incorporate managed services wherever possible. You want a **cost-effective solution to process the incoming streams of data.**

What should you do?

A. Ingest data with IoT Core, process it with **Dataprep,** and store it in a **Coldline Cloud Storage bucket.**

B. Ingest data with IoT Core, and then publish to Pub/Sub. Use **Dataflow** to process the data, and store it in a **Nearline Cloud Storage bucket.**

C. Ingest data with IoT Core, and then publish to Pub/Sub. Use **BigQuery** to process the data, and store it in a **Standard Cloud Storage bucket.**

D. Ingest data with IoT Core, and then store it in **BigQuery.**

1.1 Diagnostic Question 02 Discussion

Customers need to have a g**ood experience** when accessing your web application so they will continue to use your service. You want to **define key performance indicators (KPIs)** to establish a service level objective (SLO).

Which KPI could you use?

A. Eighty-five percent of customers are **satisfied users**
B. Eighty-five percent of **requests succeed when aggregated over 1 minute**
C. **Low latency** for > 85% of requests when aggregated over 1 minute
D. Eighty-five percent of **requests are successful**

1.1 Diagnostic Question 02 Discussion

Customers need to have a g**ood experience** when accessing your web application so they will continue to use your service. You want to **define key performance indicators (KPIs)** to establish a service level objective (SLO).

Which KPI could you use?

A. Eighty-five percent of customers are **satisfied users**

B. Eighty-five percent of **requests succeed when aggregated over 1 minute**

C. **Low latency** for > 85% of requests when aggregated over 1 minute

D. Eighty-five percent of **requests are successful**

1.1 Designing a solution infrastructure that meets business requirements

Resources to start your journey

Google Cloud Architecture Framework: System design

SRE Books

1.2 Designing a solution infrastructure that meets technical requirements

Considerations include:

- High availability and failover design
- Elasticity of cloud resources with respect to quotas and limits
- Scalability to meet growth requirements
- Performance and latency

1.2 Diagnostic Question 03 Discussion

Cymbal Direct developers have written a new application. Based on initial usage estimates, you decide to run the application on **Compute Engine instances with 15 Gb of RAM and 4 CPUs.** These instances **store persistent data locally.** After the application runs for several months, historical data indicates that the **application requires 30 Gb of RAM**. Cymbal Direct management wants you to make adjustments that will **minimize costs.**

What should you do?

A. Stop the instance, and then use the command gcloud compute instances **set-machine-type VM_NAME --machine-type e2-standard-8.** Start the instance again.

B. Stop the instance, and then use the command gcloud compute instances **set-machine-type VM_NAME --machine-type e2-standard-8.** Set the instance's metadata to: **preemptible: true**. Start the instance again.

C. Stop the instance, and then use the command gcloud compute instances **set-machine-type VM_NAME --machine-type 2-custom-4-30720**. Start the instance again.

D. Stop the instance, and then use the command gcloud compute instances **set-machine-type VM_NAME --machine-type 2-custom-4-30720**. Set the instance's metadata to: **preemptible: true.** Start the instance again.

1.2 Diagnostic Question 03 Discussion

Cymbal Direct developers have written a new application. Based on initial usage estimates, you decide to run the application on **Compute Engine instances with 15 Gb of RAM and 4 CPUs.** These instances **store persistent data locally.** After the application runs for several months, historical data indicates that the **application requires 30 Gb of RAM**. Cymbal Direct management wants you to make adjustments that will **minimize costs.**

What should you do?

A. Stop the instance, and then use the command gcloud compute instances **set-machine-type VM_NAME --machine-type e2-standard-8.** Start the instance again.

B. Stop the instance, and then use the command gcloud compute instances **set-machine-type VM_NAME --machine-type e2-standard-8.** Set the instance's metadata to: **preemptible: true**. Start the instance again.

C. Stop the instance, and then use the command gcloud compute instances **set-machine-type VM_NAME --machine-type 2-custom-4-30720**. Start the instance again.

D. Stop the instance, and then use the command gcloud compute instances **set-machine-type VM_NAME --machine-type 2-custom-4-30720**. Set the instance's metadata to: **preemptible: true.** Start the instance again.

1.2 Designing a solution infrastructure that meets technical requirements

Resources to start your journey

Google Cloud Architecture Framework: System design

1.3 Designing network, storage, and compute resources

Considerations include:

- Integration with on-premises/multicloud environments
- Cloud-native networking (VPC, peering, firewalls, container networking)
- Choosing data processing technologies
- Choosing appropriate storage types (e.g., object, file, databases)
- Choosing compute resources (e.g., preemptible, custom machine type, specialized workload)
- Mapping compute needs to platform products

1.3 Diagnostic Question 04 Discussion

You are creating a new project. You plan to set up a Dedicated interconnect between two of your data centers in the near future and want to ensure that your resources are only deployed to the **same regions** where your data centers are located. You need to make sure that you **don't have any overlapping IP addresses** that could cause conflicts when you set up the interconnect. You want to use **RFC 1918 class B address space.**

What should you do?

A. Create a new project, **leave the default network in place**, and then use the default 10.x.x.x network range to create subnets in your desired regions.

B. Create a new project, delete the default VPC network, **set up an auto mode VPC network**, and then use the default 10.x.x.x network range to create subnets in your desired regions.

C. Create a new project, delete the default VPC network, **set up a custom mode VPC network**, and then use IP addresses in the **172.16.x.x address range** to create subnets in your desired regions.

D. Create a new project, delete the default VPC network, **set up the network in custom mode**, and then use IP addresses in the **192.168.x.x address range** to create subnets in your desired zones. **Use VPC Network Peering** to connect the zones in the same region to create regional networks.

1.3 Diagnostic Question 04 Discussion

You are creating a new project. You plan to set up a Dedicated interconnect between two of your data centers in the near future and want to ensure that your resources are only deployed to the **same regions** where your data centers are located. You need to make sure that you **don't have any overlapping IP addresses** that could cause conflicts when you set up the interconnect. You want to use **RFC 1918 class B address space.**

What should you do?

A. Create a new project, **leave the default network in place**, and then use the default 10.x.x.x network range to create subnets in your desired regions.

B. Create a new project, delete the default VPC network, **set up an auto mode VPC network**, and then use the default 10.x.x.x network range to create subnets in your desired regions.

C. Create a new project, delete the default VPC network, **set up a custom mode VPC network**, and then use IP addresses in the **172.16.x.x address range** to create subnets in your desired regions.

D. Create a new project, delete the default VPC network, **set up the network in custom mode**, and then use IP addresses in the **192.168.x.x address range** to create subnets in your desired zones. **Use VPC Network Peering** to connect the zones in the same region to create regional networks.

1.3 Diagnostic Question 05 Discussion

Cymbal Direct is working with Cymbal Retail, a separate, autonomous division of Cymbal with different staff, networking teams, and data center. Cymbal Direct and Cymbal Retail are **not in the same Google Cloud organization**. Cymbal Retail needs access to Cymbal Direct's web application for making bulk orders, but the **application will not be available on the public internet**. You want to ensure that **Cymbal Retail has access to your application with low latency**. You also want to avoid **egress network charges** if possible.

What should you do?

A. Verify that the subnet range Cymbal Retail is using **doesn't overlap** with Cymbal Direct's subnet range, and then **enable VPC Network Peering for the project.**

B. If Cymbal Retail does not have access to a Google Cloud data center, **use Carrier Peering** to connect the two networks.

C. Specify Cymbal Direct's project as the **Shared VPC** host project, and then configure Cymbal Retail's project as a service project.

D. Verify that the subnet Cymbal Retail is using has the **same** IP address range with Cymbal Direct's subnet range, and then **enable VPC Network Peering for the project.**

1.3 Diagnostic Question 05 Discussion

Cymbal Direct is working with Cymbal Retail, a separate, autonomous division of Cymbal with different staff, networking teams, and data center. Cymbal Direct and Cymbal Retail are **not in the same Google Cloud organization**. Cymbal Retail needs access to Cymbal Direct's web application for making bulk orders, but the **application will not be available on the public internet**. You want to ensure that **Cymbal Retail has access to your application with low latency**. You also want to avoid **egress network charges** if possible.

What should you do?

A. Verify that the subnet range Cymbal Retail is using **doesn't overlap** with Cymbal Direct's subnet range, and then **enable VPC Network Peering for the project.**

B. If Cymbal Retail does not have access to a Google Cloud data center, **use Carrier Peering** to connect the two networks.

C. Specify Cymbal Direct's project as the **Shared VPC** host project, and then configure Cymbal Retail's project as a service project.

D. Verify that the subnet Cymbal Retail is using has the **same** IP address range with Cymbal Direct's subnet range, and then **enable VPC Network Peering for the project.**

1.3 Diagnostic Question 06 Discussion

Cymbal Direct's employees will use **Google Workspace**. Your current on-premises network **cannot meet the requirements to connect** to Google's public infrastructure.

What should you do?

A. Order a **Dedicated Interconnect** from a Google Cloud partner, and ensure that proper routes are configured.

B. Connect the network to a Google point of presence, and enable **Direct Peering.**

C. Order a **Partner Interconnect** from a Google Cloud partner, and ensure that proper routes are configured.

D. Connect the on-premises network to Google’s public infrastructure via a partner that supports **Carrier Peering.**

1.3 Diagnostic Question 06 Discussion

Cymbal Direct's employees will use **Google Workspace**. Your current on-premises network **cannot meet the requirements to connect** to Google's public infrastructure.

What should you do?

A. Order a **Dedicated Interconnect** from a Google Cloud partner, and ensure that proper routes are configured.

B. Connect the network to a Google point of presence, and enable **Direct Peering.**

C. Order a **Partner Interconnect** from a Google Cloud partner, and ensure that proper routes are configured.

D. Connect the on-premises network to Google’s public infrastructure via a partner that supports **Carrier Peering.**

1.3 Diagnostic Question 07 Discussion

Cymbal Direct is evaluating database options to store the **analytics data** from its experimental drone deliveries. You're currently using a small cluster of MongoDB NoSQL database servers. You want to move to a **managed NoSQL database service with consistent low latency that can scale throughput seamlessly and can handle the petabytes of data** you expect after expanding to additional markets.

What should you do?

A. Extract the data from MongoDB. Insert the data into **Firestore** using Datastore mode.

B. Create a **Bigtable** instance, extract the data from MongoDB, and insert the data into Bigtable.

C. Extract the data from MongoDB. Insert the data into **Firestore** using Native mode.

D. Extract the data from MongoDB, and insert the data into **BigQuery**.

1.3 Diagnostic Question 07 Discussion

Cymbal Direct is evaluating database options to store the **analytics data** from its experimental drone deliveries. You're currently using a small cluster of MongoDB NoSQL database servers. You want to move to a **managed NoSQL database service with consistent low latency that can scale throughput seamlessly and can handle the petabytes of data** you expect after expanding to additional markets.

What should you do?

A. Extract the data from MongoDB. Insert the data into **Firestore** using Datastore mode.

B. Create a **Bigtable** instance, extract the data from MongoDB, and insert the data into Bigtable.

C. Extract the data from MongoDB. Insert the data into **Firestore** using Native mode.

D. Extract the data from MongoDB, and insert the data into BigQuery.

1.3 Designing network, storage, and compute resources

Resources to start your journey

[Choose and manage compute | Architecture Framework | Google Cloud]

[Design your network infrastructure | Architecture Framework | Google Cloud]

[Select and implement a storage strategy | Architecture Framework | Google Cloud]

[Google Cloud documentation]

1.4 Creating a migration plan

Considerations include:

- Integrating solutions with existing systems
- Migrating systems and data to support the solution
- Software license mapping
- Network planning
- Testing and proofs of concept
- Dependency management planning

1.3 Diagnostic Question 08 Discussion

You are working with a client who is using **Google Kubernetes Engine (GKE)** to migrate applications from a virtual machine–based environment to a microservices-based architecture. Your client has a **complex legacy application that stores a significant amount of data on the file system of its VM.** You do not want to re-write the application to use an external service to store the file system data.

What should you do?

A. In Cloud Shell, create a YAML file defining your **Deployment** called deployment.yaml. Create a Deployment in GKE by running the command kubectl apply -f deployment.yaml

B. In Cloud Shell, create a YAML file defining your **Container** called build.yaml. Create a Container in GKE by running the command gcloud builds submit –config build.yaml .

C. In Cloud Shell, create a YAML file defining your **StatefulSet** called statefulset.yaml. Create a StatefulSet in GKE by running the command kubectl apply -f statefulset.yaml

D. In Cloud Shell, create a YAML file defining your **Pod** called pod.yaml. Create a Pod in GKE by running the command kubectl apply -f pod.yaml

1.3 Diagnostic Question 08 Discussion

You are working with a client who is using **Google Kubernetes Engine (GKE)** to migrate applications from a virtual machine–based environment to a microservices-based architecture. Your client has a **complex legacy application that stores a significant amount of data on the file system of its VM.** You do not want to re-write the application to use an external service to store the file system data.

What should you do?

A. In Cloud Shell, create a YAML file defining your **Deployment** called deployment.yaml. Create a Deployment in GKE by running the command `kubectl apply -f deployment.yaml`

B. In Cloud Shell, create a YAML file defining your **Container** called build.yaml. Create a Container in GKE by running the command `gcloud builds submit –config build.yaml .`

C. In Cloud Shell, create a YAML file defining your **StatefulSet** called statefulset.yaml. Create a StatefulSet in GKE by running the command `kubectl apply -f statefulset.yaml`

D. In Cloud Shell, create a YAML file defining your **Pod** called pod.yaml. Create a Pod in GKE by running the command `kubectl apply -f pod.yaml`

1.4 Diagnostic Question 09 Discussion

You are working in a mixed environment of VMs and Kubernetes. **Some of your resources are on-premises, and some are in Google Cloud.** Using containers as a part of your CI/CD pipeline has sped up releases significantly. You want to start **migrating some of those VMs to containers** so you can get similar benefits. You want to **automate the migration process** where possible.

What should you do?

A. **Manually create a GKE cluster**, and then use **Migrate to Containers (Migrate for Anthos)** to set up the cluster, import VMs, and convert them to containers.

B. Use **Migrate to Containers (Migrate for Anthos)** to automate the creation of **Compute Engine instances** to import VMs and convert them to containers.

C. **Manually create a GKE cluster.** Use **Cloud Build** to import VMs and convert them to containers.

D. Use **Migrate for Compute Engine** to import VMs and convert them to containers.

1.4 Diagnostic Question 09 Discussion

You are working in a mixed environment of VMs and Kubernetes. **Some of your resources are on-premises, and some are in Google Cloud.** Using containers as a part of your CI/CD pipeline has sped up releases significantly. You want to start **migrating some of those VMs to containers** so you can get similar benefits. You want to **automate the migration process** where possible.

What should you do?

A. **Manually create a GKE cluster**, and then use **Migrate to Containers (Migrate for Anthos)** to set up the cluster, import VMs, and convert them to containers.

B. Use **Migrate to Containers (Migrate for Anthos)** to automate the creation of **Compute Engine instances** to import VMs and convert them to containers.

C. **Manually create a GKE cluster.** Use **Cloud Build** to import VMs and convert them to containers.

D. Use **Migrate for Compute Engine** to import VMs and convert them to containers.

1.4 Creating a migration plan

Resources to start your journey

[Migrate to Containers | Google Cloud]

[Migration to Google Cloud: Choosing your migration path]

[Migrating to the cloud: a guide and checklist]

[Cloud Migration Products & Services]

[Application Migration | Google Cloud]

1.5 Envisioning future solution improvements

Considerations include:

- Cloud and technology improvements
- Evolution of business needs
- Evangelism and advocacy

1.5 Diagnostic Question 10 Discussion

Cymbal Direct has created a proof of concept for a social integration service that highlights images of its products from social media. **The proof of concept is a monolithic application** running on a single SuSE Linux virtual machine (VM). **The current version requires increasing the VM's CPU and RAM in order to scale.** You would like to **refactor the VM so that you can scale out instead of scaling up.**

What should you do?

A. **Move the existing codebase and VM provisioning scripts to git**, and attach external persistent volumes to the VMs.

B. Make sure that the application declares any **dependent requirements** in a requirements.txt or equivalent statement so that they can be referenced in a startup script. Specify the startup script in a **managed instance group** template, and use an autoscaling policy.

C. Make sure that the application declares any **dependent requirements** in a requirements.txt or equivalent statement so that they can be referenced in a startup script, and **attach external persistent volumes to the VMs**.

D. **Use containers instead of VMs**, and use a **GKE autoscaling deployment.**

1.5 Diagnostic Question 10 Discussion

Cymbal Direct has created a proof of concept for a social integration service that highlights images of its products from social media. **The proof of concept is a monolithic application** running on a single SuSE Linux virtual machine (VM). **The current version requires increasing the VM's CPU and RAM in order to scale.** You would like to **refactor the VM so that you can scale out instead of scaling up.**

What should you do?

A. **Move the existing codebase and VM provisioning scripts to git**, and attach external persistent volumes to the VMs.

B. Make sure that the application declares any **dependent requirements** in a requirements.txt or equivalent statement so that they can be referenced in a startup script. Specify the startup script in a **managed instance group** template, and use an autoscaling policy.

C. Make sure that the application declares any **dependent requirements** in a requirements.txt or equivalent statement so that they can be referenced in a startup script, and **attach external persistent volumes to the VMs**.

D. **Use containers instead of VMs**, and use a **GKE autoscaling deployment.**

1.5 Envisioning future solution improvements

Resources to start your journey

[Twelve-factor app development on Google Cloud | Cloud Architecture Center]

Make sure to...

Enjoy the journey as much as the destination!