1. Question

You have a project for your App Engine application that serves a development environment. The required testing has succeeded and you want to create a new project to serve as your production environment. What should you do?

- Deploy your application again using gcloud and specify the project parameter with the new project name to create the new project.
- Use gcloud to create the new project and to copy the deployed application to the new project.
- **Use gcloud to create the new project, and then deploy your application to the new project.**
- Create a Deployment Manager configuration file that copies the current App Engine deployment into a new project.

Unattempted

Use gcloud to create the new project and to copy the deployed application to the new project. is not right.
You can use gcloud to create a new project but you can not copy a deployed application from one project to another. This feature is not offered by Google App Engine.

Create a Deployment Manager configuration file that copies the current App Engine deployment into a new project. is not right.
The deployment manager configuration file contains configuration about the resources that need to be created in Google cloud, however, it does not offer the feature to copy app engine deployment into a new project.

Deploy your application again using gcloud and specify the project parameter with the new project name to create the new project. is not right.
You can deploy using gcloud app deploy and target it to a different project using –project flag. However, you can only deploy to an existing project as the gcloud app deploy command is unable to create a new project if it doesn't already exist.

Use gcloud to create the new project, and then deploy your application to the new project. is the right answer.
You can deploy to a different project by using –project flag.
By default, the service is deployed the current project configured via:
$ gcloud config set core/project PROJECT

To override this value for a single deployment, use the –project flag:
$ gcloud app deploy ~/my_app/app.yaml –project=PROJECT
Ref: https://cloud.google.com/sdk/gcloud/reference/app/deploy

.

2. Question

.

You have a single binary application that you want to run on Google Cloud Platform. You decided to automatically scale the application based on underlying infrastructure CPU usage. Your organizational policies require you to use virtual machines directly. You need to ensure that the application scaling is operationally efficient and completed as quickly as possible. What should you do?

.

- Create a Google Kubernetes Engine cluster, and use horizontal pod autoscaling to scale the application.
- **Create an instance template, and use the template in a managed instance group with autoscaling configured.**
- Create an instance template, and use the template in a managed instance group that scales up and down based on the time of day.
- Use a set of third-party tools to build automation around scaling the application up and down, based on Stackdriver CPU usage monitoring

Unattempted

Our requirements are
1. Use Virtual Machines directly (i.e. not container-based)
2. Scale Automatically
3. Scaling is efficient & is quick

Create a Google Kubernetes Engine cluster, and use horizontal pod autoscaling to scale the application. is not right.
We want to use virtual machines directly. And although GKE uses virtual machines under the hood for its GKE cluster, the autoscaling is totally different – it uses scaling at VMs (cluster auto-scaling) as well as at pod level (horizontal and vertical pod autoscaling).

Create an instance template, and use the template in a managed instance group that scales up and down based on the time of day. is not right.
Scaling based on time of the day may be insufficient especially when there is a sudden surge of requests (causing high CPU utilization) or if the requests go down suddenly (resulting in low CPU usage). Our requirements state we need to scale automatically i.e.

we need autoscaling solution that scales up and down based on CPU usage which is indicative of the volume of requests processed but scaling based on time of the day is not indicative of the load (CPU) on the system and is therefore not right.

Use a set of third-party tools to build automation around scaling the application up and down, based on Stackdriver CPU usage monitoring. is not right.
While this can be done, it is not the most efficient solution when Google's own services offer this functionality and can do it more efficiently as they are all natively integrated.

Create an instance template, and use the template in a managed instance group with autoscaling configured. is the right answer.
Managed instance groups offer autoscaling capabilities that let you automatically add or delete instances from a managed instance group based on increases or decreases in load (CPU Utilization in this case). Autoscaling helps your apps gracefully handle increases in traffic and reduce costs when the need for resources is lower. You define the autoscaling policy and the autoscaler performs automatic scaling based on the measured load (CPU Utilization in this case). Autoscaling works by adding more instances to your instance group when there is more load (upscaling), and deleting instances when the need for instances is lowered (downscaling).
Ref: https://cloud.google.com/compute/docs/autoscaler

.

3. Question

.

You have a virtual machine that is currently configured with 2 vCPUs and 4 GB of memory. It is running out of memory. You want to upgrade the virtual machine to have 8GB of memory. What should you do?

.

- **Stop the VM, increase the memory to 8 GB and start the VM**
- Rely on live migration to move the workload to a machine with more memory.
- Stop the VM, change the machine type to n1-standard-2 and start the VM
- Use gcloud to add metadata to the VM. Set the key to required-memory-size and the value to 8 GB

Unattempted

Rely on live migration to move the workload to a machine with more memory. is not right.

Live migration migrates your running instances to another host in the same zone so that Google can perform maintenance such as a software or hardware update. It can not be used for changing machine type.
Ref: https://cloud.google.com/compute/docs/instances/live-migration

Use gcloud to add metadata to the VM. Set the key to required-memory-size and the value to 8 GB. is not right.
There is no such setting as required-memory-size.

Stop the VM, change the machine type to n1-standard-2 and start the VM. is not right.
n1-standard-2 instance offers less than 8 GB (7.5 GB to be precise) so this falls short of the required memory.
Ref: https://cloud.google.com/compute/docs/machine-types

Stop the VM, increase the memory to 8 GB and start the VM. is the right answer.
In Google compute engine, if predefined machine types don't meet your needs, you can create an instance with custom virtualized hardware settings. Specifically, you can create an instance with a custom number of vCPUs and custom memory, effectively using a custom machine type. Custom machine types are ideal for the following scenarios:
1. Workloads that aren't a good fit for the predefined machine types that are available to you.
2. Workloads that require more processing power or more memory but don't need all of the upgrades that are provided by the next machine type level.
In our scenario, we only need a memory upgrade. Moving to a bigger instance would also bump up the CPU which we don't need so we have to use a custom machine type. It is not possible to change memory while the instance is running so you need to first stop the instance, change the memory and then start it again. See below a screenshot that shows how CPU/Memory can be customized for an instance that has been stopped.
Ref: https://cloud.google.com/compute/docs/instances/creating-instance-with-custom-machine-type

- **4. Question**

- You have a web application deployed as a managed instance group based on an instance template. You modified the startup script used in the instance template and would like the existing instances to pick up changes from the new startup scripts. Your web application is currently serving live web traffic. You want to propagate the startup script changes to all instances in the managed instances group while minimizing effort,

minimizing cost and ensuring that the available capacity does not decrease. What would you do?

.

- Delete instances in the managed instance group (MIG) one at a time and rely on auto-healing to provision an additional instance.
- **Perform a rolling-action replace with max-unavailable set to 0 and max-surge set to 1**
- Create a new managed instance group (MIG) based on a new template. Add the group to the backend service for the load balancer. When all instances in the new managed instance group are healthy, delete the old managed instance group
- Perform a rolling-action start-update with max-unavailable set to 1 and max-surge set to 0

Unattempted

Perform a rolling-action start-update with max-unavailable set to 1 and max-surge set to 0. is not right.
You can carry out a rolling action start update to fully replace the template by executing a command like
gcloud compute instance-groups managed rolling-action start-update instance-group-1 –zone=us-central1-a –version template=instance-template-1 –canary-version template=instance-template-2,target-size=100%
which updates the instance-group-1 to use instance-template-2 instead of instance-template-1 and have instances created out of instance-template-2 serve 100% of traffic. However, the values specified for maxSurge and maxUnavailable mean that we will lose capacity which is against our requirements.
maxSurge specifies the maximum number of instances that can be created over the desired number of instances. If maxSurge is set to 0, the rolling update can not create additional instances and is forced to update existing instances. This results in a reduction in capacity and therefore does not satisfy our requirement to ensure that the available capacity does not decrease during the deployment.
maxUnavailable – specifies the maximum number of instances that can be unavailable during the update process. When maxUnavailable is set to 1, the rolling update updates 1 instance at a time. i.e. it takes 1 instance out of service, updates it, and puts it back into service. This results in a reduction in capacity while the instance is out of service. Example – if we have 10 instances in service, this combination of setting results in 1 instance at a time taken out of service for replacement while the remaining 9 continue to serve live traffic. That's a reduction of 10% in available capacity.
Ref: https://kubernetes.io/docs/concepts/workloads/controllers/deployment/#max-unavailable
Ref: https://kubernetes.io/docs/concepts/workloads/controllers/deployment/#max-surge

Create a new managed instance group (MIG) based on a new template. Add the group to the backend service for the load balancer. When all instances in the new managed instance group are healthy, delete the old managed instance group. is not right.
While the end result is the same, we have a period of time where the traffic is served by instances from both the old managed instances group (MIG) which doubles our cost and increases effort and complexity.

Delete instances in the managed instance group (MIG) one at a time and rely on auto-healing to provision an additional instance. is not right.
While this would result in the same eventual outcome, there are two issues with this approach. First, deleting an instance one at a time would result in a reduction in capacity which is against our requirements. Secondly, deleting instances manually one at a time is error-prone and time-consuming. One of our requirements is to "minimize the effort" but deleting instances manually and relying on auto-healing health checks to provision them back is time-consuming and could take a lot of time depending on the number of instances in the MIG and the startup scripts executed during bootstrap.

Perform a rolling-action replace with max-unavailable set to 0 and max-surge set to 1. is the right answer.
This option achieves the outcome in the most optimal manner. The replace action is used to replace instances in a managed instance group. When maxUnavailable is set to 0, the rolling update can not take existing instances out of service. And when maxSurge is set to 1, we let the rolling update spin a single additional instance. The rolling update then puts the additional instance into service and takes one of the existing instances out of service for replacement. There is no reduction in capacity at any point in time.

Ref: https://kubernetes.io/docs/concepts/workloads/controllers/deployment/#max-unavailable
Ref: https://kubernetes.io/docs/concepts/workloads/controllers/deployment/#max-surge
Ref: https://cloud.google.com/sdk/gcloud/reference/alpha/compute/instance-groups/managed/rolling-action/replace

-

5. Question

-

You have a web application deployed as a managed instance group. You have a new version of the application to gradually deploy. Your web application is currently serving live web traffic. You want to ensure that the available capacity does not decrease during the deployment. What would you do?

.

- Create a new instance template with the new application version. Update the existing managed instance group with the new instance template. Delete the instances in the managed instance group to allow the managed instance group to recreate the instance using the new instance template.
- Perform a rolling-action start-update with maxSurge set to 0 and maxUnavailable set to 1
- Create a new managed instance group with an updated instance template. Add the group to the backend service for the load balancer. When all instances in the new managed instance group are healthy, delete the old managed instance group.
- **Perform a rolling-action start-update with maxSurge set to 1 and maxUnavailable set to 0**

Unattempted

Our requirements are
1. Deploy a new version gradually
2. Ensure available capacity does not decrease during deployment

Create a new managed instance group with an updated instance template. Add the group to the backend service for the load balancer. When all instances in the new managed instance group are healthy, delete the old managed instance group. is not right.
First of all instance templates can not be updated. So the phrase updated instance template rules out this option.
Ref: https://cloud.google.com/compute/docs/instance-templates/

Create a new instance template with the new application version. Update the existing managed instance group with the new instance template. Delete the instances in the managed instance group to allow the managed instance group to recreate the instance using the new instance template. is not right.
If we follow these steps, we end up with a full fleet of instances belonging to the new managed instances group (i.e. based on the new template) behind the load balancer, but our requirement to gradually deploy the new version is not met. In addition, deleting the existing instances of the managed instance group would almost certainly result in an outage to our application which is not desirable when we are serving live web traffic.

Perform a rolling-action start-update with maxSurge set to 0 and maxUnavailable set to 1 is not right.
maxSurge specifies the maximum number of instances that can be created over the desired number of instances. If maxSurge is set to 0, the rolling update can not create additional instances and is forced to update existing instances. This results in a reduction in capacity and therefore does not satisfy our requirement to ensure that the available capacity does not decrease during the deployment.

maxUnavailable – specifies the maximum number of instances that can be unavailable during the update process. When maxUnavailable is set to 1, the rolling update updates 1 instance at a time. i.e. it takes 1 instance out of service, updates it, and puts it back into service. This results in a reduction in capacity while the instance is out of service. Example – if we have 10 instances in service, this combination of setting results in 1 instance at a time taken out of service for an upgrade while the remaining 9 continue to serve live traffic. That's a reduction of 10% in available capacity and does not satisfy our requirement to ensure that the available capacity does not decrease during the deployment.

Perform a rolling-action start-update with maxSurge set to 1 and maxUnavailable set to 0 is the right answer.
This is the only option that satisfies our two requirements – deploying gradually and ensuring the available capacity does not decrease. When maxUnavailable is set to 0, the rolling update can not take existing instances out of service. And when maxSurge is set to 1, we let the rolling update spin a single additional instance. The rolling update then puts the additional instance into service and takes one of the existing instances out of service for the upgrade. There is no reduction in capacity at any point in time. And the rolling upgrade upgrades 1 instance at a time so we gradually deploy the new version. Example – if we have 10 instances in service, this combination of setting results in 1 additional instance put into service (resulting in 11 instances serving traffic), then a older instance taken out of service (resulting in 10 instances serving traffic) and puts the upgraded instance back into service (resulting in 11 instances serving traffic). The rolling upgrade continues updating the remaining 9 instances one at a time. Finally, when all 10 instances have been upgraded, the additional instance that is spun up is deleted. We still have 10 instances serving live traffic but now on the new version of code.
Ref: https://kubernetes.io/docs/concepts/workloads/controllers/deployment/#max-unavailable
Ref: https://kubernetes.io/docs/concepts/workloads/controllers/deployment/#max-surge

- **6. Question**

- You have a web application deployed as a managed instance group. You noticed some of the compute instances are running low on memory. You suspect this is due to JVM memory leak and you want to restart the compute instances to reclaim the leaked memory. Your web application is currently serving live web traffic. You want to ensure that the available capacity does not go below 80% at any time during the restarts and you want to do this at the earliest. What would you do?

-

- Stop instances in the managed instance group (MIG) one at a time and rely on autohealing to bring them back up.
- Perform a rolling-action replace with max-unavailable set to 20%.
- **Perform a rolling-action restart with max-unavailable set to 20%.**
- Perform a rolling-action reboot with max-surge set to 20%.

Unattempted

Perform a rolling-action reboot with max-surge set to 20%. is not right.
reboot is not a supported action for rolling updates. The supported actions are replace, restart, start-update and stop-proactive-update.
Ref: https://cloud.google.com/sdk/gcloud/reference/beta/compute/instance-groups/managed/rolling-action

Perform a rolling-action replace with max-unavailable set to 20%. is not right.
Performing a rolling-action replace – Replaces instances in a managed instance group. While this resolves the JVM memory leak issue, recreating the instances is a little drastic when the same result can be achieved with the simple restart action. One of our requirements is to "do this at the earliest " but recreating instances might take a lot of time depending on the number of instances and startup scripts; certainly more time than restart action.
Ref: https://cloud.google.com/sdk/gcloud/reference/beta/compute/instance-groups/managed/rolling-action

Stop instances in the managed instance group (MIG) one at a time and rely on autohealing to bring them back up. is not right.
While this would result in the same eventual outcome, it is manual, error-prone and time-consuming. One of our requirements is to "do this at the earliest" but stopping instances manually is time-consuming and could take a lot of time depending on the number of instances in the MIG. Also, relying on autohealing health checks to detect the failure and spin up the instance adds to the delay.

Perform a rolling-action restart with max-unavailable set to 20%. is the right answer.
This option achieves the outcome in the most optimal manner. The restart action restarts instances in a managed instance group. By performing a rolling restart with max-unavailable set to 20%, the rolling update restarts instances while ensuring there is at least 80% available capacity. The rolling update carries on restarting all the remaining instances until all instances in the MIG have been restarted.
Ref: https://cloud.google.com/sdk/gcloud/reference/alpha/compute/instance-groups/managed/rolling-action/restart

.

7. Question

.

You have a website hosted on App Engine standard environment. You want 1% of your users to see a new test version of the website. You want to minimize complexity. What should you do?

.

- Create a new App Engine application in the same project. Deploy the new version in that application. Configure your network load balancer to send 1% of the traffic to that new application.
- Create a new App Engine application in the same project. Deploy the new version in that application. Use the App Engine library to proxy 1% of the requests to the new version.
- Deploy the new version in the same application and use the --migrate option.
- **Deploy the new version in the same application and use the --splits option to give a weight of 99 to the current version and a weight of 1 to the new version.**

Unattempted

Deploy the new version in the same application and use the –migrate option. is not right.
migrate is not a valid flag for the gcloud app deploy command.
Ref: https://cloud.google.com/sdk/gcloud/reference/app/deploy
Also, gcloud app versions migrate, which is a valid command to migrate traffic from one version to another for a set of services, is not suitable either as we only want to send 1% traffic.
https://cloud.google.com/sdk/gcloud/reference/app/versions/migrate

Create a new App Engine application in the same project. Deploy the new version in that application. Use the App Engine library to proxy 1% of the requests to the new version. is not right.
While this can be done, we are increasing complexity and do not meet our requirement "minimize complexity". There is an out of the box option in the app engine to split traffic in a seamless way.

Create a new App Engine application in the same project. Deploy the new version in that application. Configure your network load balancer to send 1% of the traffic to that new application. is not right.
Instances that participate as backend VMs for network load balancers must be running the appropriate Linux guest environment, Windows guest environment, or other processes that provide equivalent functionality. Network load balancer is not suitable for App Engine

standard environment which is container-based and provide us specific runtimes without any promise on the underlying guest environments.

Deploy the new version in the same application and use the –splits option to give a weight of 99 to the current version and a weight of 1 to the new version. is the right answer.
You can use traffic splitting to specify a percentage distribution of traffic across two or more of the versions within a service. Splitting traffic allows you to conduct A/B testing between your versions and provides control over the pace when rolling out features.
For this scenario, we can split the traffic as shown below, sending 1% to v2 and 99% to v1
by executing the command gcloud app services set-traffic service1 –splits v2=1,v1=99
Ref: https://cloud.google.com/sdk/gcloud/reference/app/services/set-traffic

.

8. Question

.

You have a workload running on Compute Engine that is critical to your business. You want to ensure that the data on the boot disk of this workload is backed up regularly. You need to be able to restore a backup as quickly as possible in case of disaster. You also want older backups to be cleaned automatically to save on cost. You want to follow Google-recommended practices. What should you do?

.

- Create a Cloud Function to create an instance template.
- **Create a snapshot schedule for the disk using the desired interval.**
- Create a cron job to create a new disk from the disk using gcloud.
- Create a Cloud Task to create an image and export it to Cloud Storage.

Unattempted

Create a Cloud Function to create an instance template. is not right.
This does not fulfil our requirement of backing up data on boot disk 'regularly'.

Create a cron job to create a new disk from the disk using gcloud. is not right.
Like above, this does not fulfil our requirement of backing up data on boot disk 'regularly'.

Create a Cloud Task to create an image and export it to Cloud Storage. is not right.
Like above, this does not fulfil our requirement of backing up data on boot disk 'regularly'.

Create a snapshot schedule for the disk using the desired interval. is the right answer.
Create snapshots to periodically back up data from your zonal persistent disks or regional persistent disks. To reduce the risk of unexpected data loss, consider the best practice of setting up a snapshot schedule to ensure your data is backed up on a regular schedule.
Ref: https://cloud.google.com/compute/docs/disks/create-snapshots
You can also delete snapshots on a schedule by defining a snapshot retention policy. A snapshot retention policy defines how long you want to keep your snapshots. If you choose to set up a snapshot retention policy, you must do so as part of your snapshot schedule.
Ref: https://cloud.google.com/compute/docs/disks/scheduled-snapshots#retention_policy

.

9. Question

.

You have an application deployed in a GKE Cluster as a Kubernetes workload with Daemon Sets. Your application has become very popular and is now struggling to cope up with increased traffic. You want to add more pods to your workload and want to ensure your cluster scales up and scales down automatically based on volume. What should you do?

.

- Perform a rolling update to modify machine type from n1-standard-2 to n1-standard-4.
- **Enable autoscaling on Kubernetes Engine.**
- Enable Horizontal Pod Autoscaling for the Kubernetes deployment.
- Create another identical Kubernetes workload and split traffic between the two workloads.

Unattempted

Enable Horizontal Pod Autoscaling for the Kubernetes deployment. is not right.
Horizontal Pod Autoscaling can not be enabled for Daemon Sets, this is because there is only one instance of a pod per node in the cluster. In a replica deployment, when Horizontal Pod Autoscaling scales up, it can add pods to the same node or another node within the cluster. Since there can only be one pod per node in the Daemon Set workload, Horizontal Pod Autoscaling is not supported with Daemon Sets.
Ref: https://cloud.google.com/kubernetes-engine/docs/concepts/daemonset

Create another identical Kubernetes cluster and split traffic between the two workloads. is not right.
Creating another identical Kubernetes cluster is going to double your costs; at the same time, there is no guarantee that this is enough to handle all the traffic. Finally, it doesn't satisfy our requirement of "cluster scales up and scales down automatically"

Perform a rolling update to modify machine type from n1-standard-2 to n1-standard-4. is not right.
While increasing the machine type from n1-standard-2 to n1-standard-4 gives the existing nodes more resources and processing power, we don't know if that would be enough to handle the increased volume of traffic. Also, it doesn't satisfy our requirement of "cluster scales up and scales down automatically"
Ref: https://cloud.google.com/compute/docs/machine-types

Enable autoscaling on Kubernetes Engine. is the right answer.
GKE's cluster autoscaler automatically resizes the number of nodes in a given node pool, based on the demands of your workloads. As you add nodes to a node pool, DaemonSets automatically add Pods to the new nodes as needed. DaemonSets attempt to adhere to a one-Pod-per-node model.
Ref: https://cloud.google.com/kubernetes-engine/docs/concepts/cluster-autoscaler

.

10. Question

.

You have an application on a general-purpose Compute Engine instance that is experiencing excessive disk read throttling on its Zonal SSD Persistent Disk. The application primarily reads large files from disk. The disk size is currently 350 GB. You want to provide the maximum amount of throughput while minimizing costs. What should you do?

.

- Increase the size of the disk to 1 TB.
- Increase the allocated CPU to the instance.
- **Migrate to use a Local SSD on the instance.**
- Migrate to use a Regional SSD on the instance.

Unattempted

Migrate to use a Regional SSD on the instance. is not right.
Migrating to a regional SSD would actually make it worse. At the time of writing, the Read IOPS for a Zonal standard persistent disks is 7,500 and the Read IOPS reduces to 3000 for a Regional standard persistent disks which reduces the throughput.
Ref: https://cloud.google.com/compute/docs/disks/performance

Increase the size of the disk to 1 TB. is not right.
The performance of SSD persistent disks scales with the size of the disk.
Ref: https://cloud.google.com/compute/docs/disks/performance#cpu_count_size

However, there is no guarantee that increasing the disk to 1 TB will increase the throughput in this scenario as disk performance also depends on the number of vCPUs on VM instance.
Ref: https://cloud.google.com/compute/docs/disks/performance#ssd_persistent_disk_performance_by_disk_size
Ref: https://cloud.google.com/compute/docs/disks/performance#machine-type-disk-limits
For example, consider a 1,000 GB SSD persistent disk attached to an instance with an N2 machine type and 4 vCPUs. The read limit based solely on the size of the disk is 30,000 IOPS. However, because the instance has 4 vCPUs, the read limit is restricted to 15,000 IOPS.

Increase the allocated CPU to the instance. is not right.
In Compute Engine, machine types are grouped and curated for different workloads. Each machine type is subject to specific persistent disk limits per vCPU. Increasing the vCPU count increases the Read IOPS
https://cloud.google.com/compute/docs/disks/performance#machine-type-disk-limits
However, there is no guarantee that increasing CPU will definitely increase the throughput in this scenario as disk performance could be limited by disk size.
Ref: https://cloud.google.com/compute/docs/disks/performance#ssd_persistent_disk_performance_by_disk_size
Ref: https://cloud.google.com/compute/docs/disks/performance#machine-type-disk-limits
For example, consider a 1,000 GB SSD persistent disk attached to an instance with an N2 machine type and 48 vCPUs.
The read limit based solely on the vCPU count is 60,000 IOPS. However, because the instance has 1000 GB SSD, the read limit is restricted to 30,000 IOPS.

Migrate to use a Local SSD on the instance. is the right answer.
Local SSDs are physically attached to the server that hosts your VM instance. Local SSDs have higher throughput and lower latency than standard persistent disks or SSD persistent disks. The performance gains from local SSDs require certain trade-offs in availability, durability, and flexibility. Because of these trade-offs, Local SSD storage isn't automatically replicated and all data on the local SSD might be lost if the instance terminates for any reason.
Ref: https://cloud.google.com/compute/docs/disks#localssds
Ref: https://cloud.google.com/compute/docs/disks/performance#type_comparison

.

11. Question

.

You have an application running in App Engine standard environment. You want to add a custom C# library to enhance the functionality of this application. However, C# isn't supported by App Engine standard. You want to maintain the serverless aspect of your application. What should you do? Choose 2 answers.

.

- **Containerize your new application and deploy it to a Cloud Run on GKE environment.**
- **Containerize your new application and deploy it to a Cloud Run environment.**
- Containerize your new application and deploy it to a App Engine flexible environment.
- Containerize your new application and deploy it to a Google Kubernetes Engine environment.
- Split your application into different functions. Deploy your application as separate cloud functions in Google Cloud Functions (GCP) environment.

Unattempted

App engine standard currently supports Python, Java, Node.js, PHP, Ruby and Go.
Ref: https://cloud.google.com/appengine/docs/standard/
The question already states C# isn't supported by App Engine. Our requirement is to ensure we maintain the serverless aspect of our application.

Split your application into different functions. Deploy your application as separate cloud functions in Google Cloud Functions (GCP) environment is not right.
Cloud Functions is a serverless platform where you can run the code in the cloud without having to provision servers. You split your application functionality into multiple functions, and each of these is defined as a cloud function. Cloud Functions don't support C#. Supported runtimes are Python, Node.js and Go.
Ref: https://cloud.google.com/functions

Containerize your new application and deploy it to a App Engine flexible environment is not right.
While App Engine flexible lets us customize runtimes or provide our own runtime by supplying a custom Docker image or Dockerfile from the open-source community, it uses compute engine virtual machines so it is not serverless.
Ref: https://cloud.google.com/appengine/docs/flexible/

Containerize your new application and deploy it to a Google Kubernetes Engine environment. is not right.
GKE i.e. Google Kubernetes Clusters uses compute engine virtual machines so it is not

serverless.
Ref: https://cloud.google.com/kubernetes-engine

Containerize your new application and deploy it to a Cloud Run environment. is the right answer.
Cloud Run is a fully managed compute platform that automatically scales your stateless containers. Cloud Run is serverless: it abstracts away all infrastructure management, so you can focus on what matters most—building great applications. Run your containers in fully managed Cloud Run or on Anthos, which supports both Google Cloud and on?premises environments. Cloud Run is built upon an open standard, Knative, enabling the portability of your applications.
Ref: https://cloud.google.com/run

Containerize your new application and deploy it to a Cloud Run on GKE environment. is the right answer.
Cloud Run implements the Knative serving API, an open-source project to run serverless workloads on top of Kubernetes. That means you can deploy Cloud Run services anywhere Kubernetes runs. And if you need more control over your services (like access to GPU or more memory), you can also deploy these serverless containers in your own GKE cluster instead of using the fully managed environment. When using the fully managed environment, Cloud Run on GKE is serverless.

Ref: https://cloud.google.com/blog/products/serverless/cloud-run-bringing-serverless-to-containers

- **12. Question**

- You have an application running in Google Kubernetes Engine (GKE) with cluster autoscaling enabled. This application exposes a TCP endpoint. There are several replicas of the application. You have a Compute Engine instance in the same region but in another Virtual Private Cloud (VPC) called pt-network that has no overlapping CIDR range with the other VPC. The instance needs to connect to the application on GKE. You want to minimize effort. What should you do?

- 1. In GKE, create a service of type LoadBalancer that uses the application's pods as backend. 2. Set the service's externalTrafficPolicy to Cluster. 3. Configure the Compute Engine instance to use the address of the load balancer that has been assigned.

- 1. In GKE, create a service of type NodePort that uses the application's pods as backend. 2. Create a Compute Engine instance called proxy with 2 network interfaces one in VPC. 3. Use iptables on the instance to forward traffic from pt-network to the GKE nodes. 4. Configure the Compute Engine instance to use the address of proxy in pt-network as endpoint.
- **1. In GKE, create a Service of type LoadBalancer that uses the application's Pods as backend. 2. Add an annotation to this service cloud.google.com/load-balancer-type: Internal 3. Peer the two VPCs together 4. Configure the Compute Engine instance to use the address of the load balancer that has been created.**
- 1. In GKE, create a Service of type LoadBalancer that uses the application's Pods as backend. 2. Add a Cloud Armor Security Policy to the load balancer that whitelists the internal IPs of the MIG's instances. 3. Configure the Compute Engine instance to use the address of the load balancer that has been created.

Unattempted

While it may be possible to set up the networking to let the compute engine instance in pt-network communicate with pods in the GKE cluster in multiple ways, we need to look for an option that minimizes effort. Generally speaking, this means using Google Cloud Platform services directly and configuring them to achieve the intended outcome; over setting up a service ourselves, installing/managing/upgrading it ourselves which is manual, error-prone, time-consuming and add to operational overhead.

1. In GKE, create a service of type LoadBalancer that uses the application's pods as backend.
2. Set the service's externalTrafficPolicy to Cluster.
3. Configure the Compute Engine instance to use the address of the load balancer that has been assigned. is not right.

In GKE, services are used to expose pods to the outside world. There are multiple types of services. The three common types are – NodePort, ClusterIP, and LoadBalancer (there are two more service types – ExternalName and Headless which are not relevant in this context). We do not want to create a Cluster IP as this is not accessible outside the cluster. And we do not want to create NodePort as this results in exposing a port on each node in the cluster; and as we have multiple replicas, this will result in them trying to open the same port on the nodes which fail. The compute engine instance in pt-network needs a single point of communication to reach GKE. This is achieved by creating a service of type LoadBalancer. This gives the service a public IP that is externally accessible.

Ref: https://cloud.google.com/kubernetes-engine/docs/how-to/exposing-apps

externalTrafficPolicy denotes how the service should route external traffic – including public access. Rather than trying to explain, I'll point you to a very good blog that does a great job of answering how this works. https://www.asykim.com/blog/deep-dive-into-kubernetes-external-traffic-policies

Since we have cluster autoscaling enabled, we can have more than 1 node and possibly multiple replicas running on each node. So externalTrafficPolicy set to Cluster plays well

with our requirement.
Finally, we configure the compute engine to use the (externally accessible) address of the load balancer.
So this certainly looks like an option, but is it the best option that minimizes effort? One of the disadvantages of this option is that it exposes the pods publicly by using a service of type LoadBalancer. We want our compute engine to talk to the pods, but do we really want to expose our pods to the whole world? Maybe not!! Let's look at the other options to find out if there is something more relevant and secure.

1. In GKE, create a service of type NodePort that uses the application's pods as backend.
2. Create a Compute Engine instance called proxy with 2 network interfaces one in VPC.
3. Use iptables on the instance to forward traffic from pt-network to the GKE nodes.
4. Configure the Compute Engine instance to use the address of proxy in pt-network as endpoint. is not right.
For reasons explained in the above option, we don't want to create a service of type NodePort. This opens up a port on each node for each replica (pod). If we choose to do this, the compute engine doesn't have a single point to contact. Instead, it would need to contact the GKE cluster nodes individually – and that is bound to have issues because we have autoscaling enabled and the nodes may scale up and scale down as per the scaling requirements. New nodes may have different IP addresses to the previous nodes, so unless the Compute engine is continuously supplied with the IP addresses of the nodes, it can't reach them. Moreover, we have multiple replicas and it is possible we might have multiple replicas of the pod on the same node in which case they all can't open the same node port – once a node port is opened by one replica (pod), it can't be used by other replicas on the same node. So this option can be ruled out without going into the rest of the answer.

1. In GKE, create a Service of type LoadBalancer that uses the application's Pods as backend.
2. Add a Cloud Armor Security Policy to the load balancer that whitelists the internal IPs of the MIG's instances.
3. Configure the Compute Engine instance to use the address of the load balancer that has been created. is not right.
Creating a service of type LoadBalancer and getting a Compute Engine instance to use the address of the load balancer is fine, but Cloud Armor is not required. You could certainly use Cloud Armor to set up a whitelist policy to only let traffic through from the compute engine instance, but hang on – this option says "MIG instances". We don't have a managed instance group. The question mentions a single instance but not MIG. If we were to assume the single instance is part of a MIG, i.e. a MIG with a single instance, this option works too. It is more secure than the first option discussed in the explanation but at the same time more expensive. Let's look at the other option to see if it provides a secure yet cost-effective way of achieving the same.

1. In GKE, create a Service of type LoadBalancer that uses the application's Pods as backend.
2. Add an annotation to this service cloud.google.com/load-balancer-type: Internal
3. Peer the two VPCs together
4. Configure the Compute Engine instance to use the address of the load balancer that has been created. is the right answer.

Creating a service of type LoadBalancer and getting a Compute Engine instance to use the address of the load balancer is fine. We covered this previously in the first option in the explanations section.

Adding the annotation cloud.google.com/load-balancer-type: Internal makes your cluster's services accessible to applications outside of your cluster that use the same VPC network and are located in the same Google Cloud region. So this improves security by not allowing public access, however, the compute engine is located in a different VPC so it can't access.

Ref: https://cloud.google.com/kubernetes-engine/docs/how-to/internal-load-balancing

But peering the VPCs together enables the compute engine to access the load balancer IP. And peering is possible because they do not use overlapping IP ranges. Peering essentially links up the two VPCs and resources inside the VPCs can communicate with each other as if they were all in a single VPC. More info about VPC peering: https://cloud.google.com/vpc/docs/vpc-peering

So this option is the right answer. It provides a secure and cost-effective way of achieving our requirements. There are several valid answers but this option is more correct than the others.

.

13. Question

.

You have an application that looks for its licensing server on the IP 10.0.3.21. You need to deploy the licensing server on the Compute Engine. You do not want to change the configuration of the application and want the application to be able to reach the licensing server. What should you do?

.

- Use the IP 10.0.3.21 as a custom ephemeral IP address and assign it to the licensing server.
- Start the licensing server with an automatic ephemeral IP address, and then promote it to a static internal IP address.
- Reserve the IP 10.0.3.21 as a static public IP address using gcloud and assign it to the licensing server.
- **Reserve the IP 10.0.3.21 as a static internal IP address using gcloud and assign it to the licensing server.**

Unattempted

Reserve the IP 10.0.3.21 as a static public IP address using gcloud and assign it to the licensing server. is not right.
The private network range is defined by IETF (Ref: https://tools.ietf.org/html/rfc1918) and includes 10.0.0.0/8. So all IP Addresses from 10.0.0.0 to 10.255.255.255 belong to this internal IP range. As the IP of interest 10.0.3.21 falls within this range, it can not be reserved as a public IP Address.

Use the IP 10.0.3.21 as a custom ephemeral IP address and assign it to the licensing server. is not right.
Ephemeral IP address is the public IP Address assigned to compute instance. An ephemeral external IP address is an IP address that doesn't persist beyond the life of the resource. When you create an instance or forwarding rule without specifying an IP address, the resource is automatically assigned an ephemeral external IP address.
Ref: https://cloud.google.com/compute/docs/ip-addresses#ephemeraladdress
The private network range is defined by IETF (Ref: https://tools.ietf.org/html/rfc1918) and includes 10.0.0.0/8. So all IP Addresses from 10.0.0.0 to 10.255.255.255 belong to this internal IP range. As the IP of interest 10.0.3.21 falls within this range, it can not be used as a public IP Address (ephemeral IP is public).

Start the licensing server with an automatic ephemeral IP address, and then promote it to a static internal IP address. is not right.
When a compute instance is started with public IP, it gets an ephemeral IP address. An ephemeral external IP address is an IP address that doesn't persist beyond the life of the resource.
Ref: https://cloud.google.com/compute/docs/ip-addresses#ephemeraladdress
You can promote this ephemeral address into a Static IP address but this will be an external IP address and not an internal one.
Ref: https://cloud.google.com/compute/docs/ip-addresses/reserve-static-external-ip-address#promote_ephemeral_ip

Reserve the IP 10.0.3.21 as a static internal IP address using gcloud and assign it to the licensing server. is right.
This is the only option that lets us reserve IP 10.0.3.21 as a static internal IP address because it falls within the standard IP Address range as defined by IETF (Ref: https://tools.ietf.org/html/rfc1918). This includes the range 10.0.0.0/8 so all IP Addresses from 10.0.0.0 to 10.255.255.255 belong to this internal IP range. Since we can now reserve this IP Address as a static internal IP address, it can be assigned to the licensing server in the VPC so that the application is able to reach the licensing server.

.

14. Question

.

You have an application that receives SSL-encrypted TCP traffic on port 443. Clients for this application are located all over the world. You want to minimize latency for the clients. Which load balancing option should you use?

.

- HTTPS Load Balancer
- **Network Load Balancer**
- SSL Proxy Load Balancer
- Internal TCP/UDP Load Balancer. Add a firewall rule allowing ingress traffic from 0.0.0.0/0 on the target instances.

Unattempted

SSL Proxy Load Balancer. is not right.
Google says "SSL Proxy Load Balancing is intended for non-HTTP(S) traffic. For HTTP(S) traffic, we recommend that you use HTTP(S) Load Balancing."
So this option can be ruled out.
Ref: https://cloud.google.com/load-balancing/docs/ssl

Internal TCP/UDP Load Balancer. Add a firewall rule allowing ingress traffic from 0.0.0.0/0 on the target instances. is not right.
Internal TCP Load Balancing is a regional load balancer that enables you to run and scale your services behind an internal load balancing IP address that is accessible only to your internal virtual machine (VM) instances. Since we need to serve public traffic, this load balancer is not suitable for us.
Ref: https://cloud.google.com/load-balancing/docs/internal

HTTPS Load Balancer. is not right.
The HTTPS load balancer terminates TLS in locations that are distributed globally, so as to minimize latency between clients and the load balancer. If you require geographic control over where TLS is terminated (which is our scenario with clients located all over the world), you should use Google Cloud Network Load Balancing instead, and terminate TLS on backends that are located in regions appropriate to your needs.
Ref: https://cloud.google.com/load-balancing/docs/https#control-tls-termination

Network Load Balancer. is the right answer.
Google Cloud external TCP/UDP Network Load Balancing (after this referred to as Network Load Balancing) is a regional, non-proxied load balancer. The network load balancer supports any and all ports. You can use Network Load Balancing to load balance

TCP and UDP traffic. Because the load balancer is a pass-through load balancer, your backends terminate the load-balanced TCP connection or UDP packets themselves. For example, you might run an HTTPS web server on your backends (which is our scenario) and use a Network Load Balancing to route requests to it, terminating TLS on your backends themselves.
Ref: https://cloud.google.com/load-balancing/docs/network
Also, the latency is minimized when using network load balancer. Because load balancing takes place in-region and traffic is merely forwarded, there is no significant latency impact compared with the no-load-balancer option.
Ref: https://cloud.google.com/load-balancing/docs/tutorials/optimize-app-latency#network_load_balancing

.

15. Question

.

You have an application that uses Cloud Spanner as a backend database. The application has a very predictable traffic pattern. You want to automatically scale up or down the number of Spanner nodes depending on traffic. What should you do?

.

- Create a cron job that runs on a scheduled basis to review stackdriver monitoring metrics, and then resize the Spanner instance accordingly.
- Create a Stackdriver alerting policy to send an alert to oncall SRE emails when Cloud Spanner CPU exceeds the threshold. SREs would scale resources up or down accordingly.
- Create a Stackdriver alerting policy to send an alert to Google Cloud Support email when Cloud Spanner CPU exceeds your threshold. Google support would scale resources up or down accordingly.
- **Create a Stackdriver alerting policy to send an alert to webhook when Cloud Spanner CPU is over or under your threshold. Create a Cloud Function that listens to HTTP and resizes Spanner resources accordingly.**

Unattempted

Create a cron job that runs on a scheduled basis to review stackdriver monitoring metrics, and then resize the Spanner instance accordingly. is not right.
While this works and does it automatically , it does not follow Google's recommended practices.
Ref: https://cloud.google.com/spanner/docs/instances "Note: You can scale the number of nodes in your instance based on the Cloud Monitoring metrics on CPU or storage utilization in conjunction with Cloud Functions."

Create a Stackdriver alerting policy to send an alert to oncall SRE emails when Cloud Spanner CPU exceeds the threshold. SREs would scale resources up or down accordingly. is not right.
This does not follow Google's recommended practices.
Ref: https://cloud.google.com/spanner/docs/instances "Note: You can scale the number of nodes in your instance based on the Cloud Monitoring metrics on CPU or storage utilization in conjunction with Cloud Functions."

Create a Stackdriver alerting policy to send an alert to Google Cloud Support email when Cloud Spanner CPU exceeds your threshold. Google support would scale resources up or down accordingly. is not right.
This does not follow Google's recommended practices.
Ref: https://cloud.google.com/spanner/docs/instances "Note: You can scale the number of nodes in your instance based on the Cloud Monitoring metrics on CPU or storage utilization in conjunction with Cloud Functions."

Create a Stackdriver alerting policy to send an alert to webhook when Cloud Spanner CPU is over or under your threshold. Create a Cloud Function that listens to HTTP and resizes Spanner resources accordingly. is the right answer.
For scaling the number of nodes in Cloud spanner instance, Google recommends implementing this base on the Cloud Monitoring metrics on CPU or storage utilization in conjunction with Cloud Functions.
Ref: https://cloud.google.com/spanner/docs/instances

.

16. Question

.

You have an application that uses Cloud Spanner as a database backend to keep current state information about users. Cloud Bigtable logs all events triggered by users. You export Cloud Spanner data to Cloud Storage during daily backups. One of your analysts asks you to join data from Cloud Spanner and Cloud Bigtable for specific users. You want to complete this ad hoc request as efficiently as possible. What should you do?

.

- Create a dataflow job that copies data from Cloud Bigtable and Cloud Storage for specific users.
- Create a Cloud Dataproc cluster that runs a Spark job to extract data from Cloud Bigtable and Cloud Storage for specific users.

- **Create two separate BigQuery external tables on Cloud Storage and Cloud Bigtable. Use the BigQuery console to join these tables through user fields, and apply appropriate filters.**
- Create a dataflow job that copies data from Cloud Bigtable and Cloud Spanner for specific users.

Unattempted

Our requirements are to join user sessions with user events efficiently. We need to look for an option that is primarily a Google service and provides this feature out of the box or with minimal configuration.
Create two separate BigQuery external tables on Cloud Storage and Cloud Bigtable. Use the BigQuery console to join these tables through user fields, and apply appropriate filters is the right answer.
Big query lets you create tables that reference external data sources such as Bigtable and Cloud Storage. You can then join up these two tables through user fields and apply appropriate filters. You can achieve the end result with minimal configuration using this option.
Ref: https://cloud.google.com/bigquery/external-data-sources

Create a dataflow job that copies data from Cloud Bigtable and Cloud Spanner for specific users. is not right.
Cloud dataflow does not support Cloud Spanner. Cloud Dataflow SQL supports reading from Pub/Sub topics, Cloud Storage file sets, and BigQuery tables.

Create a Cloud Dataproc cluster that runs a Spark job to extract data from Cloud Bigtable and Cloud Storage for specific users. is not right.
While it is certainly possible to do this using a Spark job, it is complicated as we would have to come up with the code/logic to extract the data and certainly not straightforward.

Create a dataflow job that copies data from Cloud Bigtable and Cloud Storage for specific users. is not right.
This is possible but it is not as efficient as using Big Query.
Ref: https://cloud.google.com/dataflow/docs/guides/sql/dataflow-sql-intro

Here is some more documentation around this option, some of the issues are
1. Dataflow SQL expects CSV files in Cloud Storage filesets. CSV files must not contain a header row with column names; the first row in each CSV file is interpreted as a data record. – but our question doesn't say how the exported data is stored in cloud storage.
2. You can only run jobs in regions that have a Dataflow regional endpoint. Our question doesn't say which region. Ref: https://cloud.google.com/dataflow/docs/concepts/regional-

endpoints.
3. Creating a Dataflow job can take several minutes – unlike Big Query external tables which can be created very easily.
Too many unknowns. Otherwise, this option is a good option.
Here is some more information if you'd like to get a better understanding of how to use Cloud Dataflow to achieve this result.
Cloud Dataflow SQL lets you use SQL queries to develop and run Dataflow jobs from the BigQuery web UI. You can join streams (such as Pub/Sub) and snapshotted datasets (such as BigQuery tables and Cloud Storage filesets); query your streams or static datasets with SQL by associating schemas with objects, such as tables, Cloud Storage filesets and Pub/Sub topics; and write your results into a BigQuery table for analysis and dashboarding.
Cloud Dataflow SQL supports multiple data sources including Cloud Storage and Big Query tables which are of interest for this scenario.
https://cloud.google.com/dataflow/docs/guides/sql/data-sources-destinations

.

17. Question

.

You have an instance group that you want to load balance. You want the load balancer to terminate the client SSL session. The instance group is used to serve a public web application over HTTPS. You want to follow Google recommended practices. What should you do?

.

- Configure an external TCP proxy load balancer.
- Configure an external SSL proxy load balancer.
- Configure an internal TCP load balancer.
- **Configure an HTTP(S) load balancer.**

Unattempted

Configure an internal TCP load balancer. is not right.
Internal TCP Load Balancing is a regional load balancer that enables you to run and scale your services behind an internal load balancing IP address that is accessible only to your internal virtual machine (VM) instances. Since we need to serve public traffic, this load balancer is not suitable for us.
Ref: https://cloud.google.com/load-balancing/docs/internal

Configure an external SSL proxy load balancer. is not right.
Google says "SSL Proxy Load Balancing is intended for non-HTTP(S) traffic. For HTTP(S) traffic, we recommend that you use HTTP(S) Load Balancing."

So this option can be ruled out.
Ref: https://cloud.google.com/load-balancing/docs/ssl

Configure an external TCP proxy load balancer. is not right.
Google says "TCP Proxy Load Balancing is intended for non-HTTP traffic. For HTTP traffic, use HTTP Load Balancing instead. For proxied SSL traffic, use SSL Proxy Load Balancing." So this option can be ruled out.
Ref: https://cloud.google.com/load-balancing/docs/tcp

Configure an HTTP(S) load balancer. is the right answer.
This is the only option that fits all requirements. It can serve public traffic, can terminate SSL at the load balancer and follows google recommended practices.
? "The backends of a backend service can be either instance groups or network endpoint groups (NEGs), but not a combination of both."
? "An external HTTP(S) load balancer distributes traffic from the internet"
? "The client SSL session terminates at the load balancer."
? "For HTTP traffic, use HTTP Load Balancing instead."
Ref: https://cloud.google.com/load-balancing/docs/https

.

18. Question

.

You have an object in a Cloud Storage bucket that you want to share with an external company. The object contains sensitive data. You want access to the content to be removed after four hours. The external company does not have a Google account to which you can grant specific user-based access privileges. You want to use the most secure method that requires the fewest steps. What should you do?

.

- Configure the storage bucket as a static website and furnish the object's URL to the company. Delete the object from the storage bucket after four hours.
- Create a new Cloud Storage bucket specifically for the external company to access. Copy the object to that bucket. Delete the bucket after four hours have passed.
- **Create a signed URL with a four-hour expiration and share the URL with the company.**
- Set object access to 'public' and use object lifecycle management to remove the object after four hours.

Unattempted

Set object access to 'public' and use object lifecycle management to remove the object after four hours. is not right.
While the external company can access the public objects from the bucket, it doesn't stop bad actors from accessing the data as well. Since the data is "sensitive" and we want to follow a "secure method", we shouldn't do this.

Configure the storage bucket as a static website and furnish the object's URL to the company. Delete the object from the storage bucket after four hours. is not right.
The static website is public by default. While the external company can access the objects from the static website, it doesn't stop bad actors from accessing the data as well. Since the data is "sensitive" and we want to follow a "secure method", we shouldn't do this.

Create a new Cloud Storage bucket specifically for the external company to access. Copy the object to that bucket. Delete the bucket after four hours have passed. is not right.
Even if we were to create a separate bucket for the external company to access, since the company does not have a google account, the only way to have them access this separate bucket is by enabling public access which we can't because of the nature of data (sensitive) and is against standard security practices.

Create a signed URL with a four-hour expiration and share the URL with the company. is the right answer.
This is the only option that fits all requirements. When we generate a signed URL, we can specify an expiry and only users with the signed URL can view/download the objects, and they don't need a google account.
Ref: https://cloud.google.com/storage/docs/access-control/signed-urls
This page provides an overview of signed URLs, which you use to give time-limited resource access to anyone in possession of the URL, regardless of whether they have a Google account.

.

19. Question

.

You have annual audits every year and you need to provide external auditors access to the last 10 years of audit logs. You want to minimize the cost and operational overhead while following Google recommended practices. What should you do? (Select Three)

.

- **Grant external auditors Storage Object Viewer role on the logs storage bucket.**

- Set a custom retention of 10 years in Stackdriver logging and provide external auditors view access to Stackdriver Logs.
- **Export audit logs to Cloud Storage via an audit log export sink.**
- Export audit logs to BigQuery via an audit log export sink.
- Export audit logs to Cloud Filestore via a Pub/Sub export sink.
- **Configure a lifecycle management policy on the logs bucket to delete objects older than 10 years**

Unattempted

Export audit logs to Cloud Filestore via a Pub/Sub export sink. is not right.
Storing logs in Cloud Filestore is expensive. In Cloud Filestore, Standard Tier pricing costs $0.2 per GB per month and Premium Tier pricing costs $0.3 per GB per month. In comparison, Google Cloud Storage offers several storage classes that are significantly cheaper.
Ref: https://cloud.google.com/bigquery/pricing
Ref: https://cloud.google.com/storage/pricing

Set a custom retention of 10 years in Stackdriver logging and provide external auditors view access to Stackdriver Logs. is not right.
While it is possible to configure a custom retention period of 10 years in Stackdriver logging, storing logs in Stackdriver is expensive compared to Cloud Storage. Stackdriver charges $0.01 per GB per month, whereas something like Cloud Storage Coldline Storage costs $0.007 per GB per month (30% cheaper) and Cloud Storage Archive Storage costs 0.004 per GB per month (60% cheaper than Stackdriver)
Ref: https://cloud.google.com/logging/docs/storage#pricing
Ref: https://cloud.google.com/storage/pricing

Export audit logs to BigQuery via an audit log export sink. is not right.
Storing logs in BigQuery is expensive. In BigQuery, Active storage costs $0.02 per GB per month and Long-term storage costs $0.01 per GB per month. In comparison, Google Cloud Storage offers several storage classes that are significantly cheaper.
Ref: https://cloud.google.com/bigquery/pricing
Ref: https://cloud.google.com/storage/pricing

Export audit logs to Cloud Storage via an audit log export sink. is the right answer.
Among all the storage solutions offered by Google Cloud Platform, Cloud storage offers the best pricing for long term storage of logs. Google Cloud Storage offers several storage classes such as Nearline Storage ($0.01 per GB per Month) Coldline Storage ($0.007 per GB per Month) and Archive Storage ($0.004 per GB per month) which are significantly cheaper than the storage options covered by the above options above.
Ref: https://cloud.google.com/storage/pricing

Grant external auditors Storage Object Viewer role on the logs storage bucket. is the right answer.
You can provide external auditors access to the logs in the bucket by granting the Storage Object Viewer role which allows them to read any object stored in any bucket.
Ref: https://cloud.google.com/storage/docs/access-control/iam

Configure a lifecycle management policy on the logs bucket to delete objects older than 10 years. is the right answer.
You need to archive log files for 10 years but you don't need log files older than 10 years. And since you also want to minimize costs, it is a good idea to set up a lifecycle management policy on the bucket to delete objects that are older than 10 years. Livecycle management configuration is a set of rules which apply to current and future objects in the bucket. When an object meets the criteria of one of the rules, Cloud Storage automatically performs a specified action (delete in this case) on the object.
Ref: https://cloud.google.com/storage/docs/lifecycle

.

20. Question

.

You have asked your supplier to send you a purchase order and you want to enable them to upload the file to a cloud storage bucket within the next 4 hours. Your supplier does not have a Google account. You want to follow Google recommended practices. What should you do?

.

- Create a JSON key for the Default Compute Engine Service Account. Execute the command gsutil signurl -m PUT -d 4h gs:///**.
- Create a service account with just the permissions to upload files to the bucket. Create a JSON key for the service account. Execute the command gsutil signurl -httpMethod PUT -d 4h gs:///**.
- **Create a service account with just the permissions to upload files to the bucket. Create a JSON key for the service account. Execute the command gsutil signurl -m PUT -d 4h gs:///po.pdf.**
- Create a service account with just the permissions to upload files to the bucket. Create a JSON key for the service account. Execute the command gsutil signurl -d 4h gs:///.

Unattempted

Create a service account with just the permissions to upload files to the bucket. Create a JSON key for the service account. Execute the command gsutil signurl -d 4h gs:///. is not right.

This command creates signed URLs for retrieving existing objects. This command does not specify a HTTP method and in the absence of one, the default HTTP method is GET.
Ref: https://cloud.google.com/storage/docs/gsutil/commands/signurl

Create a service account with just the permissions to upload files to the bucket. Create a JSON key for the service account. Execute the command gsutil signurl -httpMethod PUT -d 4h gs:///**. is not right.
gsutil signurl does not accept -httpMethod parameter.
$ gsutil signurl -d 4h -httpMethod PUT keys.json gs://gcp-ace-lab-255520/*
CommandException: Incorrect option(s) specified. Usage:
The HTTP method can be provided through -m flag.
Ref: https://cloud.google.com/storage/docs/gsutil/commands/signurl

Create a JSON key for the Default Compute Engine Service Account. Execute the command gsutil signurl -m PUT -d 4h gs:///**. is not right.
Using the default compute engine service account violates the principle of least privilege. The recommended approach is to create a service account with just the right permissions needed and create JSON keys for this service account to use with gsutil signurl command.

Create a service account with just the permissions to upload files to the bucket. Create a JSON key for the service account. Execute the command gsutil signurl -m PUT -d 4h gs:///po.pdf. is the right answer.
This command correctly creates a signed url that is valid for 4 hours and allows PUT (through the -m flag) operations on the file po.pdf in the bucket. The supplier can then use the signed URL to upload a file to this bucket within 4 hours.
Ref: https://cloud.google.com/storage/docs/gsutil/commands/signurl

.

21. Question

.

You have been asked to create a new Kubernetes Cluster on Google Kubernetes Engine that can autoscale the number of worker nodes as well as pods. What should you do? (Select 2)

.

- **Create a GKE cluster and enable autoscaling on Kubernetes Engine.**
- Create Compute Engine instances for the workers and the master and install Kubernetes. Rely on Kubernetes to create additional Compute Engine instances when needed.

- Create a GKE cluster and enable autoscaling on the instance group of the cluster.
- Configure a Compute Engine instance as a worker and add it to an unmanaged instance group. Add a load balancer to the instance group and rely on the load balancer to create additional Compute Engine instances when needed.
- **Enable Horizontal Pod Autoscaling for the Kubernetes deployment.**

Unattempted

Create a GKE cluster and enable autoscaling on the instance group of the cluster. is not right.
GKE's cluster auto-scaler automatically resizes the number of nodes in a given node pool, based on the demands of your workloads. However, we should not enable Compute Engine autoscaling for managed instance groups for the cluster nodes. GKE's cluster auto-scaler is separate from Compute Engine autoscaling.
Ref: https://cloud.google.com/kubernetes-engine/docs/concepts/cluster-autoscaler

Configure a Compute Engine instance as a worker and add it to an unmanaged instance group. Add a load balancer to the instance group and rely on the load balancer to create additional Compute Engine instances when needed. is not right.
When using GKE to manage your Kubernetes clusters, you can not add manually created compute instances to the worker node pool. A node pool is a group of nodes within a cluster that all have the same configuration. Node pools use a NodeConfig specification.
Ref: https://cloud.google.com/kubernetes-engine/docs/concepts/node-pools
Moreover, Unmanaged instance groups do not autoscale. An unmanaged instance group is simply a collection of virtual machines (VMs) that reside in a single zone, VPC network, and subnet. An unmanaged instance group is useful for grouping together VMs that require individual configuration settings or tuning.
Ref: https://cloud.google.com/compute/docs/instance-groups/creating-groups-of-unmanaged-instances

Create Compute Engine instances for the workers and the master and install Kubernetes. Rely on Kubernetes to create additional Compute Engine instances when needed. is not right.
When using Google Kubernetes Engine, you can not install master node separately. The cluster master runs the Kubernetes control plane processes, including the Kubernetes API server, scheduler, and core resource controllers. The master's lifecycle is managed by GKE when you create or delete a cluster.
Ref: https://cloud.google.com/kubernetes-engine/docs/concepts/cluster-architecture
Also, you can not add manually created compute instances to the worker node pool. A node pool is a group of nodes within a cluster that all have the same configuration. Node pools use a NodeConfig specification.
Ref: https://cloud.google.com/kubernetes-engine/docs/concepts/node-pools

Create a GKE cluster and enable autoscaling on Kubernetes Engine. is the right answer. GKE's cluster autoscaler automatically resizes the number of nodes in a given node pool, based on the demands of your workloads. You don't need to manually add or remove nodes or over-provision your node pools. Instead, you specify a minimum and maximum size for the node pool, and the rest is automatic. When demand is high, cluster autoscaler adds nodes to the node pool. When demand is low, cluster autoscaler scales back down to a minimum size that you designate. This can increase the availability of your workloads when you need it while controlling costs.
Ref: https://cloud.google.com/kubernetes-engine/docs/concepts/cluster-autoscaler

Enable Horizontal Pod Autoscaling for the kubernetes deployment. is the right answer. Horizontal Pod Autoscaler scales up and scales down your Kubernetes workload by automatically increasing or decreasing the number of Pods in response to the workload's CPU or memory consumption, or in response to custom metrics reported from within Kubernetes or external metrics from sources outside of your cluster. Horizontal Pod Autoscaling cannot be used for workloads that cannot be scaled, such as DaemonSets.
Ref: https://cloud.google.com/kubernetes-engine/docs/concepts/horizontalpodautoscaler

.

22. Question

.

You have been asked to migrate a docker application from datacenter to cloud. Your solution architect has suggested uploading docker images to GCR in one project and running an application in a GKE cluster in a separate project. You want to store images in the project img-278322 and run the application in the project prod-278986. You want to tag the image as acme_track_n_trace:v1. You want to follow Google-recommended practices. What should you do?

.

- Run gcloud builds submit --tag gcr.io/img-278322/acme_track_n_trace
- **Run gcloud builds submit --tag gcr.io/img-278322/acme_track_n_trace:v1**
- Run gcloud builds submit --tag gcr.io/prod-278986/acme_track_n_trace
- Run gcloud builds submit --tag gcr.io/prod-278986/acme_track_n_trace:v1

Unattempted

Run gcloud builds submit –tag gcr.io/img-278322/acme_track_n_trace. is not right.
This command tags the image as acme_track_n_trace:latest but we want to tag the image as acme_track_n_trace:v1.
Ref: https://cloud.google.com/sdk/gcloud/reference/builds/submit

Run gcloud builds submit –tag gcr.io/prod-278986/acme_track_n_trace. is not right.
This command tags the image as acme_track_n_trace:latest but we want to tag the image as acme_track_n_trace:v1. This command also upload the image to the wrong project.
Ref: https://cloud.google.com/sdk/gcloud/reference/builds/submit

Run gcloud builds submit –tag gcr.io/prod-278986/acme_track_n_trace:v1. is not right.
This command also upload the image to the wrong project.
Ref: https://cloud.google.com/sdk/gcloud/reference/builds/submit

Run gcloud builds submit –tag gcr.io/img-278322/acme_track_n_trace:v1. is the right answer.
This command correctly tags the image as acme_track_n_trace:v1 and uploads the image to the img-278322 project.
Ref: https://cloud.google.com/sdk/gcloud/reference/builds/submit

.

23. Question

.

You have designed a solution on Google Cloud Platform (GCP) that uses multiple GCP products. Your company has asked you to estimate the costs of the solution. You need to provide estimates for the monthly total cost. What should you do?

.

- **For each GCP product in the solution, review the pricing details on the products pricing page. Use the pricing calculator to total the monthly costs for each GCP product.**
- For each GCP product in the solution, review the pricing details on the products pricing page. Create a Google Sheet that summarizes the expected monthly costs for each product.
- Provision the solution on GCP. Leave the solution provisioned for 1 week. Navigate to the Billing Report page in the Google Cloud Platform Console. Multiply the 1 week cost to determine the monthly costs.
- Provision the solution on GCP. Leave the solution provisioned for 1 week. Use Stackdriver to determine the provisioned and used resource amounts. Multiply the 1 week cost to determine the monthly costs.

Unattempted

Provision the solution on GCP. Leave the solution provisioned for 1 week. Use Stackdriver to determine the provisioned and used resource amounts. Multiply the 1 week cost to determine the monthly costs. is not right.

By provisioning the solution on GCP, you are going to incur costs. Our requirement is to just estimate the costs and this can be done by using Google Cloud Pricing Calculator.
Ref: https://cloud.google.com/products/calculator

Provision the solution on GCP. Leave the solution provisioned for 1 week. Use Stackdriver to determine the provisioned and used resource amounts. Multiply the 1 week cost to determine the monthly costs. is not right.
By provisioning the solution on GCP, you are going to incur costs. Our requirement is to just estimate the costs and this can be done by using Google Cloud Pricing Calculator.
Ref: https://cloud.google.com/products/calculator

For each GCP product in the solution, review the pricing details on the products pricing page. Create a Google Sheet that summarizes the expected monthly costs for each product. is not right.
This would certainly work but is a manual task. Why use this when you can use Google Cloud Pricing Calculator to achieve the save?
Ref: https://cloud.google.com/products/calculator

For each GCP product in the solution, review the pricing details on the products pricing page. Use the pricing calculator to total the monthly costs for each GCP product. is the right answer.
You can use the Google Cloud Pricing Calculator to total the estimated monthly costs for each GCP product. You don't incur any charges for doing so.
Ref: https://cloud.google.com/products/calculator

.

24. Question

.

You have files in a Cloud Storage bucket that you need to share with your suppliers. You want to restrict the time that the files are available to your suppliers to 1 hour. You want to follow Google recommended practices. What should you do?

.

- Create a service account with just the permissions to access files in the bucket. Create a JSON key for the service account. Execute the command gsutil signurl -p 60m gs:///.
- **Create a service account with just the permissions to access files in the bucket. Create a JSON key for the service account. Execute the command gsutil signurl -d 1h gs:///**.**

- Create a JSON key for the Default Compute Engine Service Account. Execute the command gsutil signurl -t 60m gs:///*.* .
- Create a service account with just the permissions to access files in the bucket. Create a JSON key for the service account. Execute the command gsutil signurl -m 1h gs:///*.

Unattempted

Create a JSON key for the Default Compute Engine Service Account. Execute the command gsutil signurl -t 60m gs:///*.* is not right.
gsutil signurl does not support -t flag. Executing the command with -t flag fails as shown.
$ gsutil signurl -t 60m keys.json gs://gcp-ace-lab-255520/*.*
CommandException: Incorrect option(s) specified. Usage:
Ref: https://cloud.google.com/storage/docs/gsutil/commands/signurl
Also, using the default compute engine service account violates the principle of least privilege. The recommended approach is to create a service account with just the right permissions needed and create JSON keys for this service account to use with gsutil signurl command.

Create a service account with just the permissions to access files in the bucket. Create a JSON key for the service account. Execute the command gsutil signurl -p 60m gs:///. is not right.
With gsutil signurl, -p is used to specify the key store password instead of prompting for the password. It can not be used to pass a time value. Executing the command with -p flag fails as shown.
$ gsutil signurl -p 60m keys.json gs://gcp-ace-lab-255520/*.*
TypeError: Last argument must be a byte string or a callable.
Ref: https://cloud.google.com/storage/docs/gsutil/commands/signurl

Create a service account with just the permissions to access files in the bucket. Create a JSON key for the service account. Execute the command gsutil signurl -m 1h gs:///*. is not right.
With gsutil signurl, -m is used to specify the operation e.g. PUT/GET etc. Executing the command with -m flag fails as shown.
$ gsutil signurl -m 1h keys.json gs://gcp-ace-lab-255520/*.*
CommandException: HTTP method must be one of[GET|HEAD|PUT|DELETE|RESUMABLE]
Ref: https://cloud.google.com/storage/docs/gsutil/commands/signurl

Create a service account with just the permissions to access files in the bucket. Create a JSON key for the service account. Execute the command gsutil signurl -d 1h gs:///**. is the right answer.
This command correctly specifies the duration that the signed url should be valid for by using the -d flag. The default is 1 hour so omitting the -d flag would have also resulted in

the same outcome. Times may be specified with no suffix (default hours), or with s = seconds, m = minutes, h = hours, d = days. The max duration allowed is 7d.
Ref: https://cloud.google.com/storage/docs/gsutil/commands/signurl

.

25. Question

.

You have one GCP account running in your default region and zone and another account running in a non-default region and zone. You want to start a new Compute Engine instance in these two Google Cloud Platform accounts using the command line interface. What should you do?

.

- Activate two configurations using gcloud configurations activate [NAME]. Run gcloud config list to start the Compute Engine instances.
- Activate two configurations using gcloud configurations activate [NAME]. Run gcloud configurations list to start the Compute Engine instances.
- **Create two configurations using gcloud config configurations create [NAME]. Run gcloud config configurations activate [NAME] to switch between accounts when running the commands to start the Compute Engine instances.**
- Create two configurations using gcloud config configurations create [NAME]. Run gcloud configurations list to start the Compute Engine instances.

Unattempted

Create two configurations using gcloud config configurations create [NAME]. Run gcloud configurations list to start the Compute Engine instances. is not right.
gcloud configurations list is an invalid command. To list the existing named configurations, you need to execute gcloud config configurations list but this does not start the compute engine instances.
Ref: https://cloud.google.com/sdk/gcloud/reference/config/configurations/list

Activate two configurations using gcloud configurations activate [NAME]. Run gcloud configurations list to start the Compute Engine instances. is not right.
gcloud configurations list is an invalid command. To list the existing named configurations, you need to execute gcloud config configurations list but this does not start the compute engine instances.
Ref: https://cloud.google.com/sdk/gcloud/reference/config/configurations/list

Activate two configurations using gcloud configurations activate [NAME]. Run gcloud config list to start the Compute Engine instances. is not right.

gcloud configurations activate [NAME] activates an existing named configuration. It can't be used to activate two configurations at the same time. Moreover, gcloud config list lists Cloud SDK properties for the currently active configuration. It does not start the Compute Engine instances.
Ref: https://cloud.google.com/sdk/gcloud/reference/config/configurations/activate
Ref: https://cloud.google.com/sdk/gcloud/reference/config/list

Create two configurations using gcloud config configurations create [NAME]. Run gcloud config configurations activate [NAME] to switch between accounts when running the commands to start the Compute Engine instances. is the right answer.

Each gcloud configuration has a 1 to 1 relationship with the region (if a region is defined). Since we have two different regions, we would need to create two separate configurations using gcloud config configurations create
Ref: https://cloud.google.com/sdk/gcloud/reference/config/configurations/create

Secondly, you can activate each configuration independently by running gcloud config configurations activate [NAME]
Ref: https://cloud.google.com/sdk/gcloud/reference/config/configurations/activate

Finally, while each configuration is active, you can run the gcloud compute instances start [NAME] command to start the instance in the configuration's region.
https://cloud.google.com/sdk/gcloud/reference/compute/instances/start

.

26. Question

.

You have one project called ptech-sa where you manage all your service accounts. You want to be able to use a service account from this project to take snapshots of VMs running in another project called ptech-vm. What should you do?

.

- When creating the VMs, set the service account's API scope for Compute Engine to read/write.
- **Grant the service account the IAM Role of Compute Storage Admin in the project called ptech-vm.**
- Download the private key from the service account, and add it to each VMs custom metadata.

- Download the private key from the service account, and add the private key to each VM's SSH keys.

Unattempted

Download the private key from the service account, and add it to each VMs custom metadata. is not right.
Adding service accounts private key (JSON file) to VMs custom metadata has no effect. Metadata entries are key-value pairs and do not influence any other behavior.
Ref: https://cloud.google.com/compute/docs/storing-retrieving-metadata

Download the private key from the service account, and add the private key to each VM's SSH keys. is not right.
Adding service accounts private key to the VMs SSH keys does not influence any other behavior. SSH keys are used for SSHing to the instance.
https://cloud.google.com/compute/docs/instances/adding-removing-ssh-keys

When creating the VMs, set the service account's API scope for Compute Engine to read/write. is not right.
The scopes can be modified only when using compute engine default service account.
Ref: https://cloud.google.com/compute/docs/access/service-accounts#default_service_account
See the screenshot below.
The scopes can not be modified when using a non-default service account. See the screenshot below.
Since we want to use service accounts from another project, it is safe to say they are not the default compute service accounts of this project and hence it is not possible to customize the scopes.

Grant the service account the IAM Role of Compute Storage Admin in the project called ptech-vm. is the right answer.
Compute Storage Admin role provides permissions to create, modify, and delete disks, images, and snapshots. If the service account in ptech-sa is granted the IAM Role of Compute Storage Admin in the project called ptech-vm, it can take snapshots and carry out other activities as defined by the role.
Ref: https://cloud.google.com/compute/docs/access/iam#compute.storageAdmin

.

27. Question

.

You have production and test workloads that you want to deploy on Compute Engine. Production VMs need to be in a different subnet than the test VMs. All the VMs must be

able to reach each other over internal IP without creating additional routes. You need to set up VPC and the 2 subnets. Which configuration meets these requirements?

- Create 2 custom VPCs, each with a single subnet. Create each subnet in a different region and with a different CIDR range.
- **Create a single custom VPC with 2 subnets. Create each subnet in a different region and with a different CIDR range.**
- Create 2 custom VPCs, each with a single subnet. Create each subnet in the same region and with the same CIDR range.
- Create a single custom VPC with 2 subnets. Create each subnet in the same region and with the same CIDR range.

Unattempted

Create 2 custom VPCs, each with a single subnet. Create each subnet in a different region and with a different CIDR range. is not right.
We need to get our requirements working with 1 VPC, not 2 !!

Create 2 custom VPCs, each with a single subnet. Create each subnet in the same region and with the same CIDR range. is not right.
We need to get our requirements working with 1 VPC, not 2 !!

Create a single custom VPC with 2 subnets. Create each subnet in the same region and with the same CIDR range. is not right.
We can not create two subnets in one VPC with the same CIDR range. "Primary and secondary ranges for subnets cannot overlap with any allocated range, any primary or secondary range of another subnet in the same network, or any IP ranges of subnets in peered networks." Ref: https://cloud.google.com/vpc/docs/using-vpc#subnet-rules

Create a single custom VPC with 2 subnets. Create each subnet in a different region and with a different CIDR range. is the right answer.
When we create subnets in the same VPC with different CIDR ranges, they can communicate automatically within VPC. "Resources within a VPC network can communicate with one another by using internal (private) IPv4 addresses, subject to applicable network firewall rules"
Ref: https://cloud.google.com/vpc/docs/vpc

28. Question

You have sensitive data stored in three Cloud Storage buckets and have enabled data access logging. You want to verify activities for a particular user for these buckets, using the fewest possible steps. You need to verify the addition of metadata labels and which files have been viewed from those buckets. What should you do?

.

- View the bucket in the Storage section of the GCP Console.
- Using the GCP Console, filter the Activity log to view the information.
- **Using the GCP Console, filter the Stackdriver log to view the information.**
- Create a trace in Stackdriver to view the information.

Unattempted

Our requirements are – sensitive data, verify access, fewest possible steps.

Using the GCP Console, filter the Activity log to view the information. is not right.
Since data access logging is enabled, you can see relevant log entries in both activity Logs as well as stack driver logs. However, verifying what has been viewed/updated is not straightforward in activity logs. Activity logs display a list of all actions and you can restrict this down to a user and further filter by specifying Data access as the Activity types and GCS Bucket as the Resource type. But that is the extent of the filter functionality in Activity logs. It is not possible to restrict the activity logs to just the three buckets that we are interested in. Secondly, it is not possible to restrict the activity logs to just the gets and updates. So we'd have to go through the full list to identify activities of interest before verifying them which is a manual process and can be time taking depending on the number of activities in the list.
Ref: https://cloud.google.com/storage/docs/audit-logs

View the bucket in the Storage section of the GCP Console. is not right.
The bucket page in the GCP console does not show the logs.

Create a trace in Stackdriver to view the information. is not right.
Stackdriver trace is not supported on google cloud. Stackdriver Trace runs on Linux in the following environments: Compute Engine, Google Kubernetes Engine (GKE), App Engine flexible environment, App Engine standard environment.
Ref: https://cloud.google.com/trace/docs/overview

Using the GCP Console, filter the Stackdriver log to view the information. is the right answer.
Data access logs is already enabled, so we already record all API calls that read the configuration or metadata of resources, as well as user-driven API calls that create, modify, or read user-provided resource data. Data Access audit logs do not record the

data-access operations on resources that are publicly shared (available to All Users or All Authenticated Users) or that can be accessed without logging into Google Cloud.

Since we are dealing with sensitive data, it is safe to assume that these buckets are not publicly shared and therefore enabling Data access logging logs all data-access operations on resources. These logs are sent to Stackdriver where they can be viewed by applying a suitable filter.

Unlike activity logs, retrieving the required information to verify is easier and quicker through Stackdriver as you can apply filters such as

```
resource.type=''gcs_bucket''
(resource.labels.bucket_name=''gcp-ace-lab-255520" OR
resource.labels.bucket_name=''gcp-ace-lab-255521" OR
resource.labels.bucket_name=''gcp-ace-lab-255522")
(protoPayload.methodName=''storage.objects.get'' OR
protoPayload.methodName=''storage.objects.update'')
protoPayload.authenticationInfo.principalEmail=''test.gcp.labs.user@gmail.com''
```

and query just the gets and updates, for specific buckets for a specific user. This involves fewer steps and is more efficient.
Data access logging is not enabled by default and needs to be enabled explicitly. The screenshot below shows a screenshot for enabling the data access logging for Google Cloud Storage.

.

29. Question

.

You have successfully created a development environment in a project for an application. This application uses Compute Engine and Cloud SQL. Now, you need to create a production environment for this application. The security team has forbidden the existence of network routes between these 2 environments, and asks you to follow Google-recommended practices. What should you do?

.

- **Create a new project, enable the Compute Engine and Cloud SQL APIs in that project, and replicate the setup you have created in the development environment.**

- Create a new production subnet in the existing VPC and a new production Cloud SQL instance in your existing project, and deploy your application using those resources.
- Create a new project, modify your existing VPC to be a Shared VPC, share that VPC with your new project, and replicate the setup you have in the development environment in that new project, in the Shared VPC.
- Ask the security team to grant you the Project Editor role in an existing production project used by another division of your company. Once they grant you that role, replicate the setup you have in the development environment in that project.

Unattempted

Create a new project, modify your existing VPC to be a Shared VPC, share that VPC with your new project, and replicate the setup you have in the development environment in that new project, in the Shared VPC. is not right.
A Shared VPC allows an organization to connect resources from multiple projects to a common Virtual Private Cloud (VPC) network so that they can communicate with each other securely and efficiently using internal IPs from that network. This goes totally against the recommendations of the security team.
Ref: https://cloud.google.com/vpc/docs/shared-vpc

Create a new production subnet in the existing VPC and a new production Cloud SQL instance in your existing project, and deploy your application using those resources. is not right.
You can't achieve complete isolation between development and production environments. When configuration access in Cloud SQL, while you can grant any application access to a Cloud SQL instance by authorizing the public IP addresses that the application uses to connect, you can not specify a private network (for example, 10.x.x.x) as an authorized network. The compute engine instances use their private IP addresses to reach out to Cloud SQL and because of the above limitation, we can't prevent the development compute engine reach out to production MySQL and vice versa. Since the security team has forbidden the existence of network routes between these 2 environments, having the production and development environments in a single project is not an option.
https://cloud.google.com/sql/docs/mysql/connect-external-app#appaccessIP

Ask the security team to grant you the Project Editor role in an existing production project used by another division of your company. Once they grant you that role, replicate the setup you have in the development environment in that project. is not right.
While this would technically isolate the development environment from the production environment, your production application is running in a project that is also hosting production applications of another division of your company. This goes against Google's recommended practices. You can use folders to isolate requirements for different departments and teams in the parent organization. And you have separate projects under

the folders so as per Google recommendations we should be deploying the production application to a separate project that is just for one company division/department.
Ref: https://cloud.google.com/docs/enterprise/best-practices-for-enterprise-organizations#define-hierarchy

Create a new project, enable the Compute Engine and Cloud SQL APIs in that project, and replicate the setup you have created in the development environment. is the right answer.
This aligns with Google's recommended practices. By creating a new project, we achieve complete isolation between development and production environments; as well as isolate this production application from production applications of other departments.
Ref: https://cloud.google.com/docs/enterprise/best-practices-for-enterprise-organizations#define-hierarchy

.

30. Question

.

You have three gcloud configurations – one for each of development, test and production projects. You want to list all the configurations and switch to a new configuration. With the fewest steps possible, what's the fastest way to switch to the correct configuration?

.

- To list configurations - gcloud config list To activate a configuration - gcloud config activate.
- To list configurations - gcloud configurations list To activate a configuration - gcloud configurations activate
- To list configurations - gcloud configurations list To activate a configuration - gcloud config activate.
- **To list configurations - gcloud config configurations list To activate a configuration - gcloud config configurations activate.**

Unattempted

To list configurations – gcloud configurations list
To activate a configuration – gcloud configurations activate. is not right.
gcloud configurations list does not list configurations. To list existing configurations, you need to execute gcloud config configurations list.
Ref: https://cloud.google.com/sdk/gcloud/reference/config/configurations/list
gcloud configurations activate does not activate a named configuration. To activate a configuration, you need to execute gcloud config configurations activate
Ref: https://cloud.google.com/sdk/gcloud/reference/config/configurations/activate

To list configurations – gcloud config list
To activate a configuration – gcloud config activate. is not right.
gcloud config list does not list configurations. It lists the properties of the existing configuration. To list existing configurations, you need to execute gcloud config configurations list.
Ref: https://cloud.google.com/sdk/gcloud/reference/config/configurations/list
gcloud config activate does not activate a named configuration. To activate a configuration, you need to execute gcloud config configurations activate
Ref: https://cloud.google.com/sdk/gcloud/reference/config/configurations/activate

To list configurations – gcloud configurations list
To activate a configuration – gcloud config activate. is not right.
gcloud configurations list does not list configurations. To list existing configurations, you need to execute gcloud config configurations list.
Ref: https://cloud.google.com/sdk/gcloud/reference/config/configurations/list
gcloud config activate does not activate a named configuration. To activate a configuration, you need to execute gcloud config configurations activate
Ref: https://cloud.google.com/sdk/gcloud/reference/config/configurations/activate

To list configurations – gcloud config configurations list
To activate a configuration – gcloud config configurations activate. is the right answer.
The two commands together achieve the intended outcome. gcloud config configurations list – lists existing named configurations and gcloud config configurations activate – activates an existing named configuration
Ref: https://cloud.google.com/sdk/gcloud/reference/config/configurations/list
Ref: https://cloud.google.com/sdk/gcloud/reference/config/configurations/activate
See an example below
$ gcloud config configurations list
NAME IS_ACTIVE ACCOUNT PROJECT DEFAULT_ZONE DEFAULT_REGION
dev-configuration False gcp-ace-lab-dev
prod-configuration False gcp-ace-lab-prod
test-configuration True gcp-ace-lab-test

$ gcloud config configurations activate prod-configuration
Activated [prod-configuration].

$ gcloud config configurations list
NAME IS_ACTIVE ACCOUNT PROJECT DEFAULT_ZONE DEFAULT_REGION
dev-configuration False gcp-ace-lab-dev
prod-configuration True gcp-ace-lab-prod
test-configurationFalsegcp-ace-lab-test

31. Question

You have two compute instances in the same VPC but in different regions. You can SSH from one instance to another instance using their external IP address but not their internal IP address. What could be the reason for SSH failing on the internal IP address?

- The internal IP address is disabled.
- **The combination of compute instance network tags and VPC firewall rules allow SSH from 0.0.0.0 but denies SSH from the VPC subnets IP range.**
- The compute instances are not using the right cross-region SSH IAM permissions
- The compute instances have a static IP for their internal IP.

Unattempted

The compute instances have a static IP for their internal IP. is not right.
Static internal IPs shouldn't be a reason for failed SSH connections. With all networking set up correctly, SSH works fine on Static internal IPs.
Ref: https://cloud.google.com/compute/docs/ip-addresses#networkaddresses

The internal IP address is disabled. is not right.
Every compute instance has one or more internal IP addresses so this option is not correct.

The compute instances are not using the right cross-region SSH IAM permissions. is not right.
There is no such thing as cross region SSH IAM permissions.

The combination of compute instance network tags and VPC firewall rules allow SSH from 0.0.0.0 but denies SSH from the VPC subnets IP range. is the right answer.
The combination of compute instance network tags and VPC firewall rules can certainly result in SSH traffic being allowed on the external IP range but disabled from subnets IP range. The firewall rule can be configured to allow SSH traffic from 0.0.0.0/0 but deny traffic from the VPC range e.g. 10.0.0.0/8. In this case, all SSH traffic from within the VPC is denied but external SSH traffic (i.e. on external IP) is allowed.
Ref: https://cloud.google.com/vpc/docs/using-firewalls

32. Question

You have two compute instances in the same VPC but in different regions. You can SSH from one instance to another instance using their internal IP address but not their external IP address. What could be the reason for SSH failing on external IP address?

- **The combination of compute instance network tags and VPC firewall rules only allow SSH from the subnets IP range.**
- The external IP address is disabled.
- The compute instances have a static IP for their external IP.
- The compute instances are not using the right cross-region SSH IAM permissions

Unattempted

The compute instances have a static IP for their external IP. is not right.
Not having a static IP is not a reason for failed SSH connections. When the firewall rules are set up correctly, SSH works fine on compute instances having ephemeral IP Address.

The external IP address is disabled. is not right.
Our question states SSH doesn't work on external IP addresses so it is safe to assume they already have an external IP. Therefore, this option is not correct.

The compute instances are not using the right cross-region SSH IAM permissions. is not right.
There is no such thing as cross region SSH IAM permissions.

The combination of compute instance network tags and VPC firewall rules only allow SSH from the subnets IP range. is the right answer.
The combination of compute instance network tags and VPC firewall rules can certainly result in SSH traffic being allowed from only subnets IP range. The firewall rule can be configured to allow SSH traffic from just the VPC range e.g. 10.0.0.0/8. In this scenario, all SSH traffic from within the VPC is accepted but external SSH traffic is blocked.
Ref: https://cloud.google.com/vpc/docs/using-firewalls

33. Question

You have two Kubernetes resource configuration files.
deployments.yaml – creates a deployment
service.yaml – sets up a LoadBalancer service to expose the pods.

You don't have a GKE cluster in the development project and you need to provision one. Which of the commands fail with an error in Cloud Shell when you are attempting to create a GKE cluster and deploy the YAML configuration files to create a deployment and service. (Select Two)

.

- **gcloud container clusters create cluster-1 --zone=us-central1-a gcloud container clusters get-credentials cluster-1 --zone=us-central1-a kubectl apply -f [deployment.yaml,service.yaml]**
- **gcloud container clusters create cluster-1 --zone=us-central1-a gcloud container clusters get-credentials cluster-1 --zone=us-central1-a kubectl apply -f deployment.yaml&&service.yaml**
- gcloud config set compute/zone us-central1-a gcloud container clusters create cluster-1 gcloud container clusters get-credentials cluster-1 --zone=us-central1-a kubectl apply -f deployment.yaml kubectl apply -f service.yaml
- gcloud container clusters create cluster-1 --zone=us-central1-a gcloud container clusters get-credentials cluster-1 --zone=us-central1-a kubectl apply -f deployment.yaml kubectl apply -f service.yaml
- gcloud container clusters create cluster-1 --zone=us-central1-a gcloud container clusters get-credentials cluster-1 --zone=us-central1-a kubectl apply -f deployment.yaml,service.yaml

Unattempted

gcloud container clusters create cluster-1 –zone=us-central1-a
gcloud container clusters get-credentials cluster-1 –zone=us-central1-a
kubectl apply -f deployment.yaml
kubectl apply -f service.yaml. is not right (i.e. commands executes successfully)
You create a cluster by running gcloud container clusters create command. You then fetch credentials for a running cluster by running gcloud container clusters get-credentials command. Finally, you apply the kubernetes resource configuration by running kubectl apply -f
Ref: https://cloud.google.com/sdk/gcloud/reference/container/clusters/create
Ref: https://cloud.google.com/sdk/gcloud/reference/container/clusters/get-credentials
Ref: https://kubernetes.io/docs/reference/generated/kubectl/kubectl-commands#apply

gcloud container clusters create cluster-1 –zone=us-central1-a
gcloud container clusters get-credentials cluster-1 –zone=us-central1-a
kubectl apply -f deployment.yaml,service.yaml. is not right (i.e. commands executes successfully)
Like above, but the only difference is that both configurations are applied in the same statement. With kubectl apply, you can apply the configuration from a single file or

multiple files or even a complete directory.
Ref: https://kubernetes.io/docs/reference/generated/kubectl/kubectl-commands#apply

gcloud config set compute/zone us-central1-a
gcloud container clusters create cluster-1
gcloud container clusters get-credentials cluster-1 –zone=us-central1-a
kubectl apply -f deployment.YAML
kubectl apply -f service.yaml. is not right (i.e. commands executes successfully)
Like above, but the only difference is in how the compute zone is set. In this scenario, you set the zone us-central1-a as the default zone so when you don't pass a zone property to the gcloud container clusters create command, it takes the default zone which is us-central1-a.
Ref: https://kubernetes.io/docs/reference/generated/kubectl/kubectl-commands#apply

gcloud container clusters create cluster-1 –zone=us-central1-a
gcloud container clusters get-credentials cluster-1 –zone=us-central1-a
kubectl apply -f [deployment.yaml,service.yaml]. is the right answer (i.e. commands fail)
kubectl apply can apply the configuration from a single file or multiple files or even a complete directory. When applying configuration from multiple files, the file names need to be separated by a comma. In this scenario, the filenames are passed as a list and Kubernetes treats the list as literal so looks for
files "[deployment.yaml" and "service.yaml]" which it doesn't find.
Ref: https://kubernetes.io/docs/reference/generated/kubectl/kubectl-commands#apply

gcloud container clusters create cluster-1 –zone=us-central1-a
gcloud container clusters get-credentials cluster-1 –zone=us-central1-a
kubectl apply -f deployment.yaml&&service.yaml. is the right answer (i.e. commands fail)
kubectl apply can apply the configuration from a single file or multiple files or even a complete directory. When applying configuration from multiple files, the file names need to be separated by a comma. In this scenario, the filenames are separated by && and kubernetes treats the && as literal so it looks for the
file "deployment.yaml&&service.yaml" which it doesn't find.
Ref: https://kubernetes.io/docs/reference/generated/kubectl/kubectl-commands#apply

- **34. Question**

- You have two Kubernetes resource configuration files.
deployments.yaml – creates a deployment
service.YAML – sets up a LoadBalancer service to expose the pods.

You don't have a GKE cluster in the development project and you need to provision one. Which of the commands below would you run in Cloud Shell to create a GKE cluster and deploy the YAML configuration files to create a deployment and service?

.

- gcloud container clusters create cluster-1 --zone=us-central1-a gcloud container clusters get-credentials cluster-1 --zone=us-central1-a kubectl create -f deployment.yaml kubectl create -f service.yaml
- **gcloud container clusters create cluster-1 --zone=us-central1-a gcloud container clusters get-credentials cluster-1 --zone=us-central1-a kubectl apply -f deployment.yaml kubectl apply -f service.yaml**
- gcloud container clusters create cluster-1 --zone=us-central1-a gcloud container clusters get-credentials cluster-1 --zone=us-central1-a gcloud gke apply -f deployment.yaml gcloud gke apply -f service.yaml
- kubectl container clusters create cluster-1 --zone=us-central1-a kubectl container clusters get-credentials cluster-1 --zone=us-central1-a kubectl apply -f deployment.yaml kubectl apply -f service.yaml

Unattempted

kubectl container clusters create cluster-1 –zone=us-central1-a
kubectl container clusters get-credentials cluster-1 –zone=us-central1-a
kubectl apply -f deployment.yaml
kubectl apply -f service.yaml.
is not right.
kubectl doesn't support kubectl container clusters create command. kubectl can not be used to create GKE clusters. To create a GKE cluster, you need to execute gcloud container clusters create command.
Ref: https://cloud.google.com/sdk/gcloud/reference/container/clusters/create

gcloud container clusters create cluster-1 –zone=us-central1-a
gcloud container clusters get-credentials cluster-1 –zone=us-central1-a
kubectl create -f deployment.yaml
kubectl create -f service.yaml.

is not right.
kubectl doesn't support kubectl create command. The YAML file contains the cluster resource configuration. You don't create the configuration, instead, you apply the configuration to the cluster. The configuration can be applied by running kubectl apply command
Ref: https://kubernetes.io/docs/reference/generated/kubectl/kubectl-commands#apply

gcloud container clusters create cluster-1 –zone=us-central1-a
gcloud container clusters get-credentials cluster-1 –zone=us-central1-a
gcloud gke apply -f deployment.yaml
gcloud gke apply -f service.yaml.

is not right.
gcloud doesn't support gcloud gke apply command. The YAML file contains the cluster resource configuration. You don't create the configuration, instead, you apply the configuration to the cluster. The configuration can be applied by running kubectl apply command
Ref: https://kubernetes.io/docs/reference/generated/kubectl/kubectl-commands#apply

gcloud container clusters create cluster-1 –zone=us-central1-a
gcloud container clusters get-credentials cluster-1 –zone=us-central1-a
kubectl apply -f deployment.yaml
kubectl apply -f service.yaml.

is the right answer.
You create a cluster by running gcloud container clusters create command. You then fetch credentials for a running cluster by running gcloud container clusters get-credentials command. Finally, you apply the Kubernetes resource configuration by running kubectl apply -f
Ref: https://cloud.google.com/sdk/gcloud/reference/container/clusters/create
Ref: https://cloud.google.com/sdk/gcloud/reference/container/clusters/get-credentials
Ref: https://kubernetes.io/docs/reference/generated/kubectl/kubectl-commands#apply

- **35. Question**

- You have two workloads on GKE (Google Kubernetes Engine) – create-order and dispatch-order. create-order handles the creation of customer orders, and dispatch-order handles dispatching orders to your shipping partner. Both create-order and dispatch-order workloads have cluster autoscaling enabled. The create-order deployment needs to access (i.e. invoke web service of) dispatch-order deployment. dispatch-order deployment cannot be exposed publicly. How should you define the services?

- Create a Service of type NodePort for dispatch-order and an Ingress Resource for that Service. Have create-order use the Ingress IP address.

- Create a Service of type LoadBalancer for dispatch-order and an Ingress Resource for that Service. Have create-order use the Ingress IP address.
- Create a Service of type LoadBalancer for dispatch-order. Have create-order use the Service IP address.
- **Create a Service of type ClusterIP for dispatch-order. Have create-order use the Service IP address.**

Unattempted

Create a Service of type LoadBalancer for dispatch-order. Have create-order use the Service IP address. is not right.
When you create a Service of type LoadBalancer, the Google Cloud controller configures a network load balancer that is publicly available. Since we don't want our service to be publicly available, we shouldn't create a Service of type LoadBalancer
Ref: https://cloud.google.com/kubernetes-engine/docs/how-to/exposing-apps

Create a Service of type LoadBalancer for dispatch-order and an Ingress Resource for that Service. Have create-order use the Ingress IP address. is not right.
When you create a Service of type LoadBalancer, the Google Cloud controller configures a network load balancer that is publicly available. Since we don't want our service to be publicly available, we shouldn't create a Service of type LoadBalancer
Ref: https://cloud.google.com/kubernetes-engine/docs/how-to/exposing-apps

Create a Service of type NodePort for dispatch-order and an Ingress Resource for that Service. Have create-order use the Ingress IP address. is not right.
Exposes the Service on each Node's IP at a static port (the NodePort). If the compute instance has public connectivity, the dispatch-order can be accessed publicly which is undesirable. Secondly, dispatch-order has auto-scaling enabled so we shouldn't create a service of NodePort. If autoscaler spins up another pod on the node, it fails to initialize as the port on the node is already taken by an existing pod on the same node.
Ref: https://cloud.google.com/kubernetes-engine/docs/how-to/exposing-apps

Create a Service of type ClusterIP for dispatch-order. Have create-order use the Service IP address. is the right answer.
ClusterIP exposes the Service on a cluster-internal IP that is only reachable within the cluster. This satisfies our requirement that dispatch-order shouldn't be publicly accessible. create-order which is also located in the same GKE cluster can now access the ClusterIP of the service to reach dispatch-order.
Ref: https://kubernetes.io/docs/concepts/services-networking/service/

.

36. Question

.

You host a production application in Google Compute Engine in the us-central1-a zone. Your application needs to be available 24*7 all through the year. The application suffered an outage recently due to a Compute Engine outage in the zone hosting your application. Your application is also susceptible to slowness during peak usage. You have been asked for a recommendation on how to modify the infrastructure to implement a cost-effective and scalable solution that can withstand zone failures. What would you recommend?

.

- Use Managed instance groups with instances in a single zone. Enable Autoscaling on the Managed instance group.
- Use Managed instance groups with preemptible instances across multiple zones. Enable Autoscaling on the Managed instance group.
- **Use Managed instance groups across multiple zones. Enable Autoscaling on the Managed instance group.**
- Use Unmanaged instance groups across multiple zones. Enable Autoscaling on the Unmanaged instance group.

Unattempted

Use Managed instance groups with preemptible instances across multiple zones. Enable Autoscaling on the Managed instance group. is not right.
A preemptible VM runs at a much lower price than normal instances and is cost-effective. However, Compute Engine might terminate (preempt) these instances if it requires access to those resources for other tasks. Preemptible instances are not suitable for production applications that need to be available 24*7.
Ref: https://cloud.google.com/compute/docs/instances/preemptible

Use Unmanaged instance groups across multiple zones. Enable Autoscaling on the Unmanaged instance group. is not right.
Unmanaged instance groups do not autoscale. An unmanaged instance group is simply a collection of virtual machines (VMs) that reside in a single zone, VPC network, and subnet. An unmanaged instance group is useful for grouping together VMs that require individual configuration settings or tuning.
Ref: https://cloud.google.com/compute/docs/instance-groups/creating-groups-of-unmanaged-instances

Use Managed instance groups with instances in a single zone. Enable Autoscaling on the Managed instance group. is not right.
While enabling auto-scaling is a good idea, autoscaling would spin up instances in the same zone. Should there be a zone failure, all instances of the managed instance group would be unreachable and cause the application to be unreachable. Google recommends

you distribute your resources across multiple zones to tolerate outages.
Ref: https://cloud.google.com/compute/docs/regions-zones

Use Managed instance groups across multiple zones. Enable Autoscaling on the Managed instance group. is the right answer.
Distribute your resources across multiple zones and regions to tolerate outages. Google designs zones to be independent of each other: a zone usually has power, cooling, networking, and control planes that are isolated from other zones, and most single failure events will affect only a single zone. Thus, if a zone becomes unavailable, you can transfer traffic to another zone in the same region to keep your services running.
Ref: https://cloud.google.com/compute/docs/regions-zones
In addition, a managed instance group (MIG) contains offers auto-scaling capabilities that let you automatically add or delete instances from a managed instance group based on increases or decreases in load. Autoscaling helps your apps gracefully handle increases in traffic and reduce costs when the need for resources is lower. Autoscaling works by adding more instances to your instance group when there is more load (upscaling), and deleting instances when the need for instances is lowered (downscaling).
Ref: https://cloud.google.com/compute/docs/autoscaler/

.

37. Question

.

You host a static website in Cloud Storage. Recently you began to include links to PDF files on this site. Currently, when users click on links to these PDF files, their browser prompts them to save the file to their machine locally. However, you want the clicked PDF files to be displayed within the browser window directly without prompting the user to save the files locally. What should you do?

.

- **Set Content-Type metadata to application/pdf on the PDF file objects**
- Enable Cloud CDN on the website frontend.
- Add a label to the storage bucket with a key of Content-Type and a value of application/pdf.
- Enable Share publicly on the PDF file objects

Unattempted

Set Content-Type metadata to application/pdf on the PDF file objects is the right answer. Content-Type allows browsers to render the object properly. If the browser prompts users to save files to their machine, it is likely the browser does not see the Content-Type as application/pdf. Setting this would ensure the browser displays PDF files within the browser instead of popping up a download dialog.

Ref: https://cloud.google.com/storage/docs/gsutil/addlhelp/WorkingWithObjectMetadata#content-type_1

Enable Cloud CDN on the website frontend. is not right.
CDN helps with caching content at the edge but doesn't help the browser in displaying pdf files.

Enable Share publicly on the PDF file objects. is not right.
The fact that the browser lets users download the file suggests the browser is able to reach out and download the file. Sharing publicly wouldn't make any difference.

Add a label to the storage bucket with a key of Content-Type and a value of application/pdf. is not right.
Bucket labels are key: value metadata pairs that allow you to group your buckets along with other Google Cloud resources such as virtual machine instances and persistent disks. They don't determine the file's content type.

.

38. Question

.

You installed Stackdriver Logging agent on all compute instances. You now need to forward logs from all Compute Engine instances to a BigQuery dataset called pt-logs. You want to minimize cost. What should you do?

.

- 1. Give the BigQuery Data Editor role on the pt-logs dataset to the service accounts used by your instances. 2. Update your instances' metadata to add the following value: logs-destination: bq://pt-logs.
- 1. In Stackdriver Logging, create a logs export with a Cloud Pub/Sub topic called logs as a sink. 2. Create a Cloud Function that is triggered by messages in the logs topic. 3. Configure that Cloud Function to drop logs that are not from Compute Engine and to insert Compute Engine logs in the pt-logs dataset.
- **1. In Stackdriver Logging, create a filter to view only Compute Engine logs. 2. Click Create Export. 3. Choose BigQuery as Sink Service, and the pt-logs dataset as Sink Destination.**
- 1. Create a Cloud Function that has the BigQuery User role on the pt-logs dataset. 2. Configure this Cloud Function to create a BigQuery Job that executes this query: INSERT INTO dataset.pt-logs (timestamp, log) SELECT timestamp, log FROM

compute.logs WHERE timestamp > DATE_SUB(CURRENT_DATE(), INTERVAL 1 DAY) 3. Use Cloud Scheduler to trigger this Cloud Function once a day.

Unattempted

1. Give the BigQuery Data Editor role on the pt-logs dataset to the service accounts used by your instances.
2. Update your instances' metadata to add the following value: logs-destination: bq://pt-logs. is not right.
Among other things, roles/bigquery.dataEditor lets you Create, update, get, and delete the dataset's tables. However, setting a metadata tag logs-destination to bq://pt-logs has no effect on how the logs are generated or forwarded. The stack driver agent is already installed so the logs are forwarded to stack driver logging and not to the BigQuery dataset. Metadata entries are key-value pairs and do not influence this behavior.
Ref: https://cloud.google.com/compute/docs/storing-retrieving-metadata

1. In Stackdriver Logging, create a logs export with a Cloud Pub/Sub topic called logs as a sink.
2. Create a Cloud Function that is triggered by messages in the logs topic.
3. Configure that Cloud Function to drop logs that are not from Compute Engine and to insert Compute Engine logs in the pt-logs dataset. is not right.
While the end result meets our requirement, this option involves more steps, it is inefficient and expensive. Triggering a cloud function for each log message and then dropping messages that are not relevant (i.e. not compute engine logs) is inefficient. We are paying for cloud function execution for all log entries when we are only interested in compute engine logs. Secondly, triggering a cloud function and then have that insert into the BigQuery dataset is also inefficient and expensive when the same can be achieved directly by configuring BigQuery as the sink destination – we don't pay for cloud function executions. Using this option, we are unnecessarily paying for Cloud Pub/Sub and Cloud Functions.
Ref: https://cloud.google.com/logging/docs/export/configure_export_v2
Ref: https://cloud.google.com/logging/docs/view/advanced-queries

1. Create a Cloud Function that has the BigQuery User role on the pt-logs dataset.
2. Configure this Cloud Function to create a BigQuery Job that executes this query:
INSERT INTO dataset.pt-logs (timestamp, log)
SELECT timestamp, log FROM compute.logs
WHERE timestamp > DATE_SUB(CURRENT_DATE(), INTERVAL 1 DAY)
3. Use Cloud Scheduler to trigger this Cloud Function once a day. is not right.
The role roles/bigquery.user provides permissions to run jobs, including queries, within the project. A cloud function with this role can execute queries in BigQuery, however, the logs are not available in BigQuery in compute.logs so you can not query compute engine logs by running SELECT timestamp, log FROM compute.logs.
Ref: https://cloud.google.com/bigquery/docs/access-control

1. In Stack driver Logging, create a filter to view only Compute Engine logs.
2. Click Create Export.
3. Choose BigQuery as Sink Service, and the pt-logs dataset as Sink Destination. is the right answer.
In stack driver logging, it is possible to create a filter to just query the compute engine logs which is what we are interested in.
Ref: https://cloud.google.com/logging/docs/view/advanced-queries
You can then export these logs into a sink that has the BigQuery dataset configured as the destination.
https://cloud.google.com/logging/docs/export/configure_export_v2
This way, just the logs that we need are exported to BigQuery. This option is the most efficient of all options and uses features provided by GCP out of the box.

.

39. Question

.

You need a dynamic way of provisioning VMs on the Compute Engine. The exact specifications will be in a dedicated configuration file. You want to follow Google recommended practices. Which method should you use?

.

- Unmanaged Instance Group
- **Deployment Manager**
- Managed Instance Group
- Cloud Composer

Unattempted

Unmanaged Instance Group. is not right.
Unmanaged instance groups let you load balance across a fleet of VMs that you manage yourself. But it doesn't help with dynamically provisioning VMs.
Ref: https://cloud.google.com/compute/docs/instance-groups#unmanaged_instance_groups

Cloud Composer. is not right.
Cloud Composer is a fully managed workflow orchestration service that empowers you to author, schedule, and monitor pipelines that span across clouds and on-premises data centers. Cloud Composer is deeply integrated within the Google Cloud Platform, giving users the ability to orchestrate their full pipeline. Cloud Composer has robust, built-in integration with many products, including BigQuery, Cloud Dataflow, Cloud Dataproc, Cloud Datastore, Cloud Storage, Cloud Pub/Sub, and AI Platform.
Ref: https://cloud.google.com/composer

Managed Instance Group. is not right.
Managed instance groups (MIGs) let you operate apps on multiple identical VMs. You can make your workloads scalable and highly available by taking advantage of automated MIG services, including autoscaling, autohealing, regional (multiple zones) deployment, and automatic updating. While MIG dynamically provisions virtual machines based on scaling policy, it doesn't satisfy our requirement of "dedicated configuration file"
Ref: https://cloud.google.com/compute/docs/instance-groups#managed_instance_groups

Deployment Manager. is the right answer.
Google Cloud Deployment Manager allows you to specify all the resources needed for your application in a declarative format using YAML. You can also use Python or Jinja2 templates to parameterize the configuration and allow reuse of common deployment paradigms such as a load-balanced, auto-scaled instance group. You can deploy many resources at one time, in parallel. Using the deployment manager, you can apply a Python/Jinja2 template to create a MIG/auto-scaling policy that dynamically provisions VM. And our other requirement of "dedicated configuration file" is also met. Using the deployment manager for provisioning results in a repeatable deployment process. By creating configuration files that define the resources, the process of creating those resources can be repeated over and over with consistent results. Google recommends we script our infrastructure and deploy using Deployment Manager
Ref: https://cloud.google.com/deployment-manager

.

40. Question

.

You need to assign a Cloud Identity and Access Management (Cloud IAM) role to an external auditor. The auditor needs to have permissions to review your Google Cloud Platform (GCP) Audit Logs and also to review your Data Access logs. What should you do?

.

- Assign the auditor the IAM role roles/logging.privateLogViewer. Perform the export of logs to Cloud Storage.
- **Assign the auditor the IAM role roles/logging.privateLogViewer. Direct the auditor to also review the logs for changes to Cloud IAM policy.**
- Assign the auditor's IAM user to a custom role that has logging.privateLogEntries.list permission. Perform the export of logs to Cloud Storage.
- Assign the auditor's IAM user to a custom role that has logging.privateLogEntries.list permission. Direct the auditor to also review the logs for changes to Cloud IAM policy.

Unattempted

Google Cloud provides Cloud Audit Logs, which is an integral part of Cloud Logging. It consists of two log streams for each project: Admin Activity and Data Access, which are generated by Google Cloud services to help you answer the question of "who did what, where, and when?" within your Google Cloud projects.
Ref: https://cloud.google.com/iam/docs/job-functions/auditing#scenario_external_auditors

To view Admin Activity audit logs, you must have one of the following Cloud IAM roles in the project that contains your audit logs:
- Project Owner, Project Editor, or Project Viewer.
- The Logging Logs Viewer role.
- A custom Cloud IAM role with the logging.logEntries.list Cloud IAM permission.
https://cloud.google.com/iam/docs/audit-logging#audit_log_permissions

To view Data Access audit logs, you must have one of the following roles in the project that contains your audit logs:
- Project Owner.
- Logging's Private Logs Viewer role.
- A custom Cloud IAM role with the logging.privateLogEntries.list Cloud IAM permission.
https://cloud.google.com/iam/docs/audit-logging#audit_log_permissions
——
Assign the auditor's IAM user to a custom role that has logging.privateLogEntries.list permission. Perform the export of logs to Cloud Storage. is not right.
logging.privateLogEntries.list provides permissions to view Data Access audit logs but this does not grant permissions to view Admin activity logs.
Ref: https://cloud.google.com/logging/docs/access-control#console_permissions

Assign the auditor's IAM user to a custom role that has logging.privateLogEntries.list permission. Direct the auditor to also review the logs for changes to Cloud IAM policy. is not right.
logging.privateLogEntries.list provides permissions to view Data Access audit logs but this does not grant permissions to view Admin activity logs.
Ref: https://cloud.google.com/logging/docs/access-control#console_permissions

Assign the auditor the IAM role roles/logging.privateLogViewer. Perform the export of logs to Cloud Storage. is not right.
roles/logging.privateLogViewer is the right role and lets the auditor review admin activity and data access logs but exporting logs to Cloud Storage indicates that we want the auditor to review logs from Cloud Storage and not the logs within Cloud Logging console. In this scenario, unless the auditor is assigned a role that lets them access the relevant cloud storage buckets, they wouldn't be able to view log information in the buckets.

Assign the auditor the IAM role roles/logging.privateLogViewer. Direct the auditor to also review the logs for changes to Cloud IAM policy. is the right answer.
roles/logging.privateLogViewer (Private Logs Viewer) includes everything from roles/logging.viewer, plus the ability to read Access Transparency logs and Data Access audit logs. This lets the auditor review the admin activity and data access logs in Cloud Logging console.
Ref: https://cloud.google.com/logging/docs/access-control

.

41. Question

.

You need to configure IAM access audit logging in BigQuery for external auditors. You want to follow Google-recommended practices. What should you do?

.

- Add the auditors group to two new custom IAM roles.
- Add the auditor user accounts to the 'logging.viewer' and 'bigQuery.dataViewer' predefined IAM roles.
- Add the auditor user accounts to two new custom IAM roles.
- **Add the auditors group to the 'logging.viewer' and 'bigQuery.dataViewer' predefined IAM roles.**

Unattempted

Add the auditor user accounts to the 'logging.viewer' and 'bigQuery.dataViewer' predefined IAM roles. is not right.
Since auditing happens several times a year, we don't want to repeat the process of granting multiple roles to multiple users every time. Instead, we want to define a group with the required grants (a one time task) and assign this group to the auditor users during the time of the audit.

Add the auditor user accounts to two new custom IAM roles. is not right.
Google already provides roles that fit the external auditing requirements so we don't need to create custom roles. Nothing stops us from creating custom IAM roles to achieve the same purpose, but this doesn't follow "Google-recommended practices"

Add the auditors group to two new custom IAM roles. is not right.
Google already provides roles that fit the external auditing requirements so we don't need to create custom roles. Nothing stops us from creating custom IAM roles to achieve the same purpose, but this doesn't follow "Google-recommended practices"

Add the auditors group to the 'logging.viewer' and 'bigQuery.dataViewer' predefined IAM roles. is the right answer.
For external auditors, Google recommends we grant logging.viewer and bigquery.dataViewer roles. Since auditing happens several times a year to review the organization's audit logs, it is recommended we create a group with these grants and assign the group to auditor user accounts during the time of the audit.

.

42. Question

.

You need to create a custom IAM role for use with a GCP service. All permissions in the role must be suitable for production use. You also want to clearly share with your organization the status of the custom role. This will be the first version of the custom role. What should you do?

.

- **Use permissions in your role that use the SUPPORTED support level for role permissions. Set the role stage to ALPHA while testing the role permissions.**
- Use permissions in your role that use the SUPPORTED support level for role permissions. Set the role stage to BETA while testing the role permissions.
- Use permissions in your role that use the TESTING support level for role permissions. Set the role stage to ALPHA while testing the role permissions.
- Use permissions in your role that use the TESTING support level for role permissions. Set the role stage to BETA while testing the role permissions.

Unattempted

When setting support levels for permissions in custom roles, you can set to one of SUPPORTED, TESTING or NOT_SUPPORTED.
SUPPORTED -The permission is fully supported in custom roles.
TESTING – The permission is being tested to check its compatibility with custom roles. You can include the permission in custom roles, but you might see unexpected behavior. Not recommended for production use.
Ref: https://cloud.google.com/iam/docs/custom-roles-permissions-support
Since we want the role to be suitable for production use, we need "SUPPORTED" and not "TESTING".

In terms of role stage, the stage transitions from ALPHA –> BETA –> GA
Ref: https://cloud.google.com/iam/docs/understanding-custom-roles#testing_and_deploying
Since this is the first version of custom role, we start with "ALPHA".

The only option that satisfies "ALPHA" stage with "SUPPORTED" support level is Use permissions in your role that use the SUPPORTED support level for role permissions. Set the role stage to ALPHA while testing the role permissions.

.

43. Question

.

You need to create a custom VPC with a single subnet. The subnet's range must be as large as possible. Which range should you use?

.

- 10.0.0.0/8
- 192.168.0.0/16
- 172.16.0.0/12
- 0.0.0.0/0

Unattempted

The private network range is defined by IETF (Ref: https://tools.ietf.org/html/rfc1918) and adhered to by all cloud providers. The supported internal IP Address ranges are

1. 24-bit block 10.0.0.0/8 (16777216 IP Addresses)
2. 20-bit block 172.16.0.0/12 (1048576 IP Addresses)
3. 16-bit block 192.168.0.0/16 (65536 IP Addresses)

10.0.0.0/8 gives you the largest range - 16777216 IP Addresses.

.

44. Question

.

You need to create a new billing account and then link it with an existing Google Cloud Platform project. What should you do?

.

- Verify that you are Project Billing Manager for the GCP project. Update the existing project to link it to the existing billing account.
- Verify that you are Billing Administrator for the billing account. Update the existing project to link it to the existing billing account.

- **Verify that you are Project Billing Manager for the GCP project. Create a new billing account and link the new billing account to the existing project.**
- Verify that you are Billing Administrator for the billing account. Create a new project and link the new project to the existing billing account.

Unattempted

Our requirement is to link an existing google cloud project with a new billing account.

Verify that you are Billing Administrator for the billing account. Create a new project and link the new project to the existing billing account. is not right.
We do not need to create a new project.

Verify that you are Project Billing Manager for the GCP project. Update the existing project to link it to the existing billing account. is not right.
We want to link the project with a new billing account so is not right.

Verify that you are Billing Administrator for the billing account. Update the existing project to link it to the existing billing account. is not right.
We want to link the project with a new billing account so is not right.

Verify that you are Project Billing Manager for the GCP project. Create a new billing account and link the new billing account to the existing project. is the right answer.
The purpose of Project Billing Manager is to Link/unlink the project to/from a billing account. It is granted at the organization or project level. Project Billing Manager role allows a user to attach the project to the billing account, but does not grant any rights over resources. Project Owners can use this role to allow someone else to manage the billing for the project without granting them resource access.
Ref: https://cloud.google.com/billing/docs/how-to/billing-access

.

45. Question

.

You need to create an autoscaling managed instance group for an HTTPS web application. You want to make sure that unhealthy VMs are recreated. What should you do?

.

- In the Instance Template, add the label health-check.
- In the Instance Template, add a startup script that sends a heartbeat to the metadata server.

- Select Multi-Zone instead of Single-Zone when creating the Managed Instance Group.
- **Create a health check on port 443 and use that when creating the Managed Instance Group.**

Unattempted

Our requirement is to ensure unhealthy VMs are recreated.

Select Multi-Zone instead of Single-Zone when creating the Managed Instance Group. is not right.
You can create two types of MIGs: A zonal MIG, which deploys instances to a single zone and a regional MIG, which deploys instances to multiple zones across the same region. However, this doesn't help with recreating unhealthy VMs.
Ref: https://cloud.google.com/compute/docs/instance-groups

In the Instance Template, add the label health-check. is not right.
Metadata entries are key-value pairs and do not influence any other behavior.
Ref: https://cloud.google.com/compute/docs/storing-retrieving-metadata

In the Instance Template, add a startup script that sends a heartbeat to the metadata server. is not right.
The startup script is executed only at the time of startup. Whereas we need something like a liveness check that monitors the status of the server periodically to identify if the VM is unhealthy. So this is not going to work.
Ref: https://cloud.google.com/compute/docs/startupscript

Create a health check on port 443 and use that when creating the Managed Instance Group. is the right answer.
To improve the availability of your application and to verify that your application is responding, you can configure an auto-healing policy for your managed instance group (MIG). An auto-healing policy relies on an application-based health check to verify that an application is responding as expected. If the auto healer determines that an application isn't responding, the managed instance group automatically recreates that instance. Since our application is a HTTPS web application, we need to set up our health check on port 443 which is the standard port for HTTPS.
Ref: https://cloud.google.com/compute/docs/instance-groups/autohealing-instances-in-migs

.

46. Question

.

You need to deploy an application, which is packaged in a container image, in a new project. The application exposes an HTTP endpoint and receives very few requests per day. You want to minimize costs. What should you do?

.

- **Deploy the container on Cloud Run.**
- Deploy the container on Cloud Run on GKE.
- Deploy the container on App Engine Flexible.
- Deploy the container on Google Kubernetes Engine, with cluster autoscaling and horizontal pod autoscaling enabled.

Unattempted

Deploy the container on Cloud Run on GKE. is not right.
Cloud Run on GKE can scale the number of pods to zero. The number of nodes per cluster cannot scale to zero and these nodes are billed in the absence of requests.
Ref: https://cloud.google.com/serverless-options

Deploy the container on Google Kubernetes Engine, with cluster autoscaling and horizontal pod autoscaling enabled. is not right.
Like above, while you can set up the pod autoscaler to scale back the pods to zero, the number of nodes per cluster cannot scale to zero and these nodes are billed in the absence of requests. If you specify the minimum node pool size of zero nodes, an idle node pool can scale down completely. However, at least one node must always be available in the cluster to run system Pods.
Ref: https://cloud.google.com/kubernetes-engine/docs/concepts/cluster-autoscaler

Deploy the container on App Engine Flexible. is not right.
App Engine flexible environment instances are Compute Engine virtual machines. This means you can't truly scale down to zero and compute instances are billed in the absence of requests.
Ref: https://cloud.google.com/appengine/docs/flexible

Deploy the container on Cloud Run. is the right answer.
Cloud Run is a fully managed compute platform that automatically scales your stateless containers. Cloud Run is serverless. Cloud Run abstracts away all infrastructure management. It automatically scales up and down from zero depending on traffic almost instantaneously. Cloud Run only charges you for the exact resources you use.
Ref: https://cloud.google.com/run

.

47. Question

You need to enable traffic between multiple groups of Compute Engine instances that are currently running two different GCP projects. Each group of Compute Engine instances is running in its own VPC. What should you do?

- Verify that both projects are in a GCP Organization. Create a new VPC and add all instances.
- **Verify that both projects are in a GCP Organization. Share the VPC from one project and request that the Compute Engine instances in the other project use this shared VPC.**
- Verify that you are the Project Administrator of both projects. Create two new VPCs and add all instances.
- Verify that you are the Project Administrator of both projects. Create a new VPC and add all instances.

Unattempted

Verify that both projects are in a GCP Organization. Share the VPC from one project and request that the Compute Engine instances in the other project use this shared VPC. is the right answer.
All other options make no sense. Shared VPC allows an organization to connect resources from multiple projects to a common Virtual Private Cloud (VPC) network so that they can communicate with each other securely and efficiently using internal IPs from that network. When you use Shared VPC, you designate a project as a host project and attach one or more other service projects to it.
Ref: https://cloud.google.com/vpc/docs/shared-vpc
All other options make no sense.

48. Question

You need to grant access for three users so that they can view and edit table data on a Cloud Spanner instance. What should you do?

- Run gcloud iam roles describe roles/spanner.databaseUser. Add the users to the role.
- Run gcloud iam roles describe roles/spanner.viewer -- project my-gcp-ace-project. Add the users to a new group. Add the group to the role.

- **Run gcloud iam roles describe roles/spanner.databaseUser. Add the users to a new group. Add the group to the role.**
- Run gcloud iam roles describe roles/spanner.viewer -- project my-gcp-ace-project. Add the users to the role.

Unattempted

Our requirements
1. View and Edit table data
2. 3 users (i.e. multiple users)

3 users should give us the idea that we do not want to assign roles/permissions at the user level and instead want to do it based on groups so that we can create one group with all the required permissions and all such users who need this access can be assigned to the group.
Ref: https://cloud.google.com/iam/docs/reference/rest/v1/Policy#Binding
Ref: https://cloud.google.com/iam/docs/granting-changing-revoking-access#granting-gcloud-manual

Run gcloud iam roles describe roles/spanner.databaseUser. Add the users to the role. is not right.
We are looking for an option that assigns users to a group (in order to maximise reuse and minimize maintenance overhead). This option assigns users to the role so is not the right answer.

Run gcloud iam roles describe roles/spanner.viewer — project my-gcp-ace-project. Add the users to the role. is not right.
We are looking for an option that assigns users to a group (in order to maximise reuse and minimize maintenance overhead). This option assigns users to the role so is not the right answer.

Run gcloud iam roles describe roles/spanner.viewer — project my-gcp-ace-project. Add the users to a new group. Add the group to the role. is not right.
Adding users to a group and granting the role to the group is the right way forward. But the role used in this option is spanner.viewer which allows viewing all Cloud Spanner instances (but cannot modify instances), and allows viewing all Cloud Spanner databases (but cannot modify databases and cannot read from databases). Since we required edit access as well, this option is not right.
Ref: https://cloud.google.com/spanner/docs/iam

Run gcloud iam roles describe roles/spanner.databaseUser. Add the users to a new group. Add the group to the role. is the right answer.

Adding users to a group and granting the role to the group is the right way forward. In addition, we assign the role spanner.databaseUser which allows Read from and write to the Cloud Spanner database; execute SQL queries on the database, including DML and Partitioned DML; and View and update schema for the database. This is the only option that grants the right role to a group and assigns users to the group.

.

49. Question

.

You need to host an application on a Compute Engine instance in a project shared with other teams. You want to prevent the other teams from accidentally causing downtime on that application. Which feature should you use?

.

- Use a Shielded VM.
- Use a Preemptible VM.
- Use a sole-tenant node.
- **Enable deletion protection on the instance.**

Unattempted

Use a Shielded VM. is not right.
Shielded VMs are virtual machines (VMs) on Google Cloud hardened by a set of security controls that help defend against rootkits and boot kits. Using Shielded VMs helps protect enterprise workloads from threats like remote attacks, privilege escalation, and malicious insiders. But shielded VMs don't offer protection for accidental termination of the instance.
Ref: https://cloud.google.com/shielded-vm

Use a Preemptible VM. is not right.
A preemptible VM is an instance that you can create and run at a much lower price than normal instances. However, Compute Engine might terminate (preempt) these instances if it requires access to those resources for other tasks. Preemptible instances are excess Compute Engine capacity, so their availability varies with usage. Preemptible VMs don't offer protection for accidental termination of the instance.
Ref: https://cloud.google.com/compute/docs/instances/preemptible

Use a sole-tenant node. is not right.
Sole-tenancy lets you have exclusive access to a sole-tenant node, which is a physical Compute Engine server that is dedicated to hosting only your project's VMs. Use sole-tenant nodes to keep your VMs physically separated from VMs in other projects, or to group your VMs together on the same host hardware. Sole-tenant nodes don't offer

protection for accidental termination of the instance.
Ref: https://cloud.google.com/compute/docs/nodes

Enable deletion protection on the instance. is the right answer.
As part of your workload, there might be certain VM instances that are critical to running your application or services, such as an instance running a SQL server, a server used as a license manager, and so on. These VM instances might need to stay running indefinitely so you need a way to protect these VMs from being deleted. By setting the deletionProtection flag, a VM instance can be protected from accidental deletion. If a user attempts to delete a VM instance for which you have set the deletionProtection flag, the request fails. Only a user that has been granted a role with compute.instances.create permission can reset the flag to allow the resource to be deleted.
Ref: https://cloud.google.com/compute/docs/instances/preventing-accidental-vm-deletion

.

50. Question

.

You need to manage multiple Google Cloud Platform (GCP) projects in the fewest steps possible. You want to configure the Google Cloud SDK command line interface (CLI) so that you can easily manage multiple GCP projects. What should you?

.

- **1. Create a configuration for each project you need to manage. 2. Activate the appropriate configuration when you work with each of your assigned GCP projects.**
- 1. Create a configuration for each project you need to manage. 2. Use gcloud init to update the configuration values when you need to work with a non-default project
- 1. Use the default configuration for one project you need to manage. 2. Activate the appropriate configuration when you work with each of your assigned GCP projects.
- 1. Use the default configuration for one project you need to manage. 2. Use gcloud init to update the configuration values when you need to work with a non-default project.

Unattempted

1. Create a configuration for each project you need to manage.
2. Activate the appropriate configuration when you work with each of your assigned GCP projects. is the right answer.

gcloud configurations enable you to manage multiple projects in gcloud cli using the fewest possible steps,
Ref: https://cloud.google.com/sdk/gcloud/reference/config

For example, we have two projects
$ gcloud projects list
PROJECT_ID NAME PROJECT_NUMBER
project-1-278333 project-1-278333 85524215451
project-2-278333 project-2-278333 25349885274

We create configuration for each project. For project-2-278333,
$ gcloud config configurations create project-1-config
$ gcloud config set project project-1-278333

And for project-2-278333,
$ gcloud config configurations create project-2-config
$ gcloud config set project project-2-278333

We now have two configurations, one for each project.
$ gcloud config configurations list
NAME IS_ACTIVE ACCOUNT PROJECT COMPUTE_DEFAULT_ZONE
COMPUTE_DEFAULT_REGION
cloudshell-4453 False
project-1-config False project-1-278333
project-2-config True project-2-278333

To activate configuration for project-1,
$ gcloud config configurations activate project-1-config
Activated [project-1-config].
$ gcloud config get-value project
Your active configuration is: [project-1-config]
project-1-278333

To activate configuration for project-2,
$ gcloud config configurations activate project-2-config
Activated [project-2-config].
$ gcloud config get-value project
Your active configuration is: [project-2-config]
project-2-278333

.

51. Question

.

You need to monitor resources that are distributed over different projects in Google Cloud Platform. You want to consolidate reporting under the same Stackdriver Monitoring dashboard. What should you do?

.

- **Configure a single Stackdriver account, and link all projects to the same account.**
- Use Shared VPC to connect all projects, and link Stackdriver to one of the projects.
- Configure a single Stackdriver account for one of the projects. In Stackdriver, create a Group and add the other project names as criteria for that Group.
- For each project, create a Stackdriver account. In each project, create a service account for that project and grant it the role of Stackdriver Account Editor in all other projects.

Unattempted

Use Shared VPC to connect all projects, and link Stackdriver to one of the projects. is not right.
Linking Stackdriver to one project brings metrics from that project alone. A Shared VPC allows an organization to connect resources from multiple projects to a common Virtual Private Cloud (VPC) network so that they can communicate with each other securely and efficiently using internal IPs from that network. But it does not help in linking all projects to a single Stackdriver workspace/account.
Ref: https://cloud.google.com/vpc/docs/shared-vpc

For each project, create a Stackdriver account. In each project, create a service account for that project and grant it the role of Stackdriver Account Editor in all other projects. is not right.
Stackdriver monitoring does not use roles to gather monitoring information from the project. Instead, the Stackdriver Monitoring agent, which is a collectd-based daemon, gathers system and application metrics from virtual machine instances and sends them to Monitoring. In this case, as each project is linked to a separate Stackdriver account, it is not possible to have a consolidated view of all monitoring.
Ref: https://cloud.google.com/monitoring/agent

Configure a single Stackdriver account for one of the projects. In Stackdriver, create a Group and add the other project names as criteria for that Group. is not right.
As the other projects are not linked to the stack driver, they can't be monitored.

Moreover, you can not add projects to Stackdriver groups. Groups provide a mechanism for alerting on the behavior of a set of resources, rather than on individual resources. For example, you can create an alerting policy that is triggered if some number of resources in the group violates a particular condition (for example, CPU load), rather than having each resource inform you of violations individually.
Ref: https://cloud.google.com/monitoring/groups

Configure a single Stackdriver account, and link all projects to the same account. is the right answer.
You can monitor resources of different projects in a single Stackdriver account by creating a Stackdriver workspace. A Stackdriver workspace is a tool for monitoring resources contained in one or more Google Cloud projects or AWS accounts. Each Workspace can have between 1 and 100 monitored projects, including Google Cloud projects and AWS accounts. A Workspace accesses metric data from its monitored projects, but the metric data and log entries remain in the individual projects.
Ref: https://cloud.google.com/monitoring/workspaces

.

52. Question

.

You need to produce a list of the enabled Google Cloud Platform APIs for a GCP project using the gcloud command line in the Cloud Shell. The project name is my-project. What should you do?

.

- **Run gcloud projects list to get the project ID, and then run gcloud services list --project .**
- Run gcloud init to set the current project to my-project, and then run gcloud services list --available.
- Run gcloud info to view the account value, and then run gcloud services list --account .
- Run gcloud projects describe to verify the project value, and then run gcloud services list --available.

Unattempted

Run gcloud init to set the current project to my-project, and then run gcloud services list –available. is not right.
–available return the services available to the project to enable and not the services that are enabled. This list will include any services that the project has already enabled plus the services that the project can enable.
Ref: https://cloud.google.com/sdk/gcloud/reference/services/list

Also, to set the current project, you need to use gcloud config set project Ref: https://cloud.google.com/sdk/gcloud/reference/config/set
gcloud init is used for initializing or reinitializing gcloud configurations.
https://cloud.google.com/sdk/gcloud/reference/init

Run gcloud info to view the account value, and then run gcloud services list –account . is not right.
We aren't passing any project id to the command so it would fail with the error shown below. (n.b. it is possible this command succeeds if you have an active gcloud configuration that has set the project so rather than accepting value from –project parameter, the command would obtain the project info from the gcloud configuration. The command shown below is run when no configuration is active).
gcloud services list –account
Errors with the following error.
ERROR: (gcloud.services.list) The project property is set to the empty string, which is invalid.
To set your project, run:
$ gcloud config set project PROJECT_ID
or to unset it, run:
$ gcloud config unset project

Run gcloud projects describe to verify the project value, and then run gcloud services list –available. is not right.
–available return the services available to the project to enable and not the services that are enabled. This list will include any services that the project has already enabled plus the services that the project can enable.
Ref: https://cloud.google.com/sdk/gcloud/reference/services/list

Run gcloud projects list to get the project ID, and then run gcloud services list –project . is the right answer.
For the gcloud services list command, –enabled is the default.
So running
gcloud services list –project is the same as running
gcloud services list –project –enabled
which would get all the enabled services for the project.

.

53. Question

.

You need to provide a cost estimate for a Kubernetes cluster using the GCP pricing calculator for Kubernetes. Your workload requires high IOPS, and you will also be using disk snapshots. You start by entering the number of nodes, average hours, and average days. What should you do next?

.

- **Fill in local SSD. Fill in persistent disk storage and snapshot storage.**
- Fill in local SSD. Add estimated cost for cluster management.
- Select Add GPUs. Fill in persistent disk storage and snapshot storage.
- Select Add GPUs. Add estimated cost for cluster management.

Unattempted

Fill in local SSD. Add estimated cost for cluster management. is not right.
You don't need to add an estimated cost for cluster management as the calculator automatically applies this. At the time of writing this, GKE clusters accrue a management fee of $0.10 per cluster per hour, irrespective of cluster size or topology. One zonal cluster per billing account is free. GKE cluster management fees do not apply to Anthos GKE clusters.
Ref: https://cloud.google.com/kubernetes-engine/pricing

Select Add GPUs. Add estimated cost for cluster management. is not right.
You don't need to add an estimated cost for cluster management as the calculator automatically applies this. At the time of writing this, GKE clusters accrue a management fee of $0.10 per cluster per hour, irrespective of cluster size or topology. One zonal cluster per billing account is free. GKE cluster management fees do not apply to Anthos GKE clusters.
Ref: https://cloud.google.com/kubernetes-engine/pricing

Select Add GPUs. Fill in persistent disk storage and snapshot storage. is not right.
GPUs don't help us with our requirement of high IOPS. Compute Engine provides graphics processing units (GPUs) that you can add to your virtual machine instances to accelerate specific workloads on your instances such as machine learning and data processing. But this doesn't help increase IOPS.
Ref: https://cloud.google.com/compute/docs/gpus

Fill in local SSD. Fill in persistent disk storage and snapshot storage. is the right answer.
The pricing calculator for Kubernetes Engine offers us the ability to add GPUs as well as specify Local SSD requirements for estimation. GPUs don't help us with our requirement of high IOPS but Local SSD does.
Ref: https://cloud.google.com/products/calculator

GKE offers always-encrypted local solid-state drive (SSD) block storage. Local SSDs are physically attached to the server that hosts the virtual machine instance for very high input/output operations per second (IOPS) and very low latency compared to persistent disks.
Ref: https://cloud.google.com/kubernetes-engine
Once you fill in the local SSD requirement, you can fill in persistent disk storage and snapshot storage.

.

54. Question

.

You need to reduce GCP service costs for a division of your company using the fewest possible steps. You need to turn off all configured services in an existing GCP project. What should you do?

.

- **1. Verify that you are assigned the Project Owners IAM role for this project. 2. Locate the project in the GCP console, click Shut down and then enter the project ID.**
- 1. Verify that you are assigned the Project Owners IAM role for this project. 2. Switch to the project in the GCP console, locate the resources and delete them.
- 1. Verify that you are assigned the Organization Administrator IAM role for this project. 2. Locate the project in the GCP console, enter the project ID and then click Shut down.
- 1. Verify that you are assigned the Organization Administrator IAM role for this project. 2. Switch to the project in the GCP console, locate the resources and delete them.

Unattempted

1. Verify that you are assigned the Organization Administrator IAM role for this project. 2. Locate the project in the GCP console, enter the project ID and then click Shut down. is not right.
Organization Admin role provides permissions to get and list projects but not shutdown projects.
Ref: https://cloud.google.com/iam/docs/understanding-roles#resource-manager-roles

1. Verify that you are assigned the Organization Administrator IAM role for this project. 2. Switch to the project in the GCP console, locate the resources and delete them. is not right.
Organization Admin role provides permissions to get and list projects but not delete

projects.
Ref: https://cloud.google.com/iam/docs/understanding-roles#resource-manager-roles

1. Verify that you are assigned the Project Owner IAM role for this project.
2. Switch to the project in the GCP console, locate the resources and delete them. is not right.
The primitive Project Owner role provides permissionst to delete project
https://cloud.google.com/iam/docs/understanding-roles#primitive_roles
But locating all the resources and deleting them is a manual task, time consuming and error prone. Our goal is to accomplish the same but with fewest possible steps

1. Verify that you are assigned the Project Owner IAM role for this project.
2. Locate the project in the GCP console, click Shut down and then enter the project ID. is the right answer.
The primitive Project Owner role provides permissionst to delete project
https://cloud.google.com/iam/docs/understanding-roles#primitive_roles
You can shut down projects using the Cloud Console. When you shut down a project, this immediately happens: All billing and traffic serving stops, You lose access to the project, The owners of the project will be notified and can stop the deletion within 30 days, The project will be scheduled to be deleted after 30 days. However, some resources may be deleted much earlier.
Ref: https://cloud.google.com/resource-manager/docs/creating-managing-projects#shutting_down_projects

.

55. Question

.

You need to run an important query in BigQuery but expect it to return a lot of records. You want to find out how much it will cost to run the query. You are using on-demand pricing. What should you do?

.

- Run a select count (*) to get an idea of how many records your query will look through. Then convert that number of rows to dollars using the Pricing Calculator.
- Arrange to switch to Flat-Rate pricing for this query, then move back to on-demand.
- Use the command line to run a dry run query to estimate the number of bytes returned. Then convert that bytes estimate to dollars using the Pricing Calculator.
- **Use the command line to run a dry run query to estimate the number of bytes read. Then convert that bytes estimate to dollars using the Pricing Calculator.**

Unattempted

Arrange to switch to Flat-Rate pricing for this query, then move back to on-demand. is not right.
The cost of acquiring a big query slot (associated with flat-rate pricing) is significantly higher than our requirement here to run a single important query or just to know how much it would cost to run that query. BigQuery offers flat-rate pricing for customers who prefer a stable monthly cost for queries rather than paying the on-demand price per TB of data processed. You enroll in flat-rate pricing by purchasing slot commitments, measured in BigQuery slots. Slot commitments start at 500 slots and the price starts from $10000. Your queries consume this slot capacity, and you are not billed for bytes processed.
Ref: https://cloud.google.com/bigquery/pricing#flat_rate_pricing

Use the command line to run a dry run query to estimate the number of bytes returned. Then convert that bytes estimate to dollars using the Pricing Calculator. is not right.
Under on-demand pricing, BigQuery doesn't charge for the query execution based on the output of the query (i.e. bytes returned) but on the number of bytes processed (also referred to as bytes read or bytes scanned) in order to arrive at the output of the query. You are charged for the number of bytes processed whether the data is stored in BigQuery or in an external data source such as Cloud Storage, Google Drive, or Cloud Bigtable. On-demand pricing is based solely on usage. You are charged for the bytes scanned even if your query itself doesn't return any data.
Ref: https://cloud.google.com/bigquery/pricing

Run a select count (*) to get an idea of how many records your query will look through. Then convert that number of rows to dollars using the Pricing Calculator. is not right.
This is not as practical as identifying the number of records your query will look through (i.e. scan/process) is not straightforward. Plus BigQuery supports external data sources such as Cloud Storage, Google Drive, or Cloud Bigtable; and the developer cost associated with identifying this information from various data sources is significant, not practical and sometimes not possible.

Use the command line to run a dry run query to estimate the number of bytes read. Then convert that bytes estimate to dollars using the Pricing Calculator. is the right answer.
BigQuery pricing is based on the number of bytes processed/read. Under on-demand pricing, BigQuery charges for queries by using one metric: the number of bytes processed (also referred to as bytes read). You are charged for the number of bytes processed whether the data is stored in BigQuery or in an external data source such as Cloud Storage, Google Drive, or Cloud Bigtable. On-demand pricing is based solely on usage.
Ref: https://cloud.google.com/bigquery/pricing

56. Question

You need to select and configure compute resources for a set of batch processing jobs. These jobs take around 2 hours to complete and are run nightly. You want to minimize service costs. What should you do?

- Select Compute Engine. Use VM instance types that support micro bursting.
- Select GKE. Use a single node cluster with a small instance type
- **Select Compute Engine. Use preemptible VM instances of the appropriate standard machine types.**
- Select GKE. Use a three-node cluster with micro instance types.

Unattempted

Requirements – achieve end goal while minimizing service costs.

Select GKE. Use a single node cluster with a small instance type is not right.
We do not know if a small instance is capable of handling all the batch volume. Plus this is not the most cost-effective of the options.

Select GKE. Use a three-node cluster with micro instance types is not right.
We do not know if three micro instances are capable of handling all the batch volume. Plus this is not the most cost-effective of the options.

Select Compute Engine. Use VM instance types that support micro bursting is not right.
We can use an instance that supports micro bursting but we have a job that runs for 2 hours. Bursting is suitable for short periods.

Select Compute Engine. Use preemptible VM instances of the appropriate standard machine types is the right answer.
We minimize the cost by selecting a preemptible instance of the appropriate type. If the preemptible instance is terminated, the next nightly run picks up the unprocessed volume.

57. Question

You need to set up a policy so that videos stored in a specific Cloud Storage Regional bucket are moved to Coldline after 90 days and then deleted after one year from their creation. How should you set up the policy?

.

- **Use Cloud Storage Object Lifecycle Management using Age conditions with SetStorageClass and Delete actions. Set the SetStorageClass action to 90 days and the Delete action to 365 days.**
- Use gsutil rewrite and set the Delete action to 365 days.
- Use gsutil rewrite and set the Delete action to 275 days (365-90).
- Use Cloud Storage Object Lifecycle Management using Age conditions with SetStorageClass and Delete actions. Set the SetStorageClass action to 90 days and the Delete action to 275 days (365 - 90)

Unattempted

Use gsutil rewrite and set the Delete action to 275 days (365-90). is not right.
gsutil rewrite is used to change the storage class of objects within a bucket through overwriting the object. It does not support Delete action.
Ref: https://cloud.google.com/storage/docs/changing-storage-classes

Use gsutil rewrite and set the Delete action to 365 days. is not right.
gsutil rewrite is used to change the storage class of objects within a bucket through overwriting the object. It does not support Delete action.
Ref: https://cloud.google.com/storage/docs/changing-storage-classes

Use Cloud Storage Object Lifecycle Management using Age conditions with SetStorageClass and Delete actions. Set the SetStorageClass action to 90 days and the Delete action to 275 days (365 – 90). is not right.
Object Lifecycle Management does not rewrite an object when changing its storage class. This means that when an object is transitioned to Nearline Storage, Coldline Storage, or Archive Storage using the SetStorageClass feature, any subsequent early deletion and associated charges are based on the original creation time of the object, regardless of when the storage class changed.
If however, the change of storage class is done manually using a rewrite, the creation time of the objects is the new creation time since they are rewritten. In such a case, you would need to apply a lifecycle delete action of 275 days.
Ref: https://cloud.google.com/storage/docs/lifecycle

Use Cloud Storage Object Lifecycle Management using Age conditions with SetStorageClass and Delete actions. Set the SetStorageClass action to 90 days and the Delete action to 365 days. is the right answer.

Object Lifecycle Management does not rewrite an object when changing its storage class. This means that when an object is transitioned to Nearline Storage, Coldline Storage, or Archive Storage using the SetStorageClass feature, any subsequent early deletion and associated charges are based on the original creation time of the object, regardless of when the storage class changed.
Ref: https://cloud.google.com/storage/docs/lifecycle

.

58. Question

.

You need to set up permissions for a set of Compute Engine instances to enable them to write data into a particular Cloud Storage bucket. You want to follow Google-recommended practices. What should you do?

.

- **Create a service account with an access scope. Use the access scope 'https://www.googleapis.com/auth/cloud-platform'.**
- Create a service account with an access scope. Use the access scope 'https://www.googleapis.com/auth/devstorage.write_only'.
- Create a service account and add it to the IAM role 'storage.objectAdmin' for that bucket.
- Create a service account and add it to the IAM role 'storage.objectCreator' for that bucket.

Unattempted

Our requirements are
1. Google recommended practices
2. Multiple compute engine instances to write data to a bucket.

Create a service account with an access scope. Use the access scope 'https://www.googleapis.com/auth/devstorage.write_only'. is not right.
There is no scope called "write_only".
Ref: https://cloud.google.com/storage/docs/authentication

Create a service account and add it to the IAM role 'storage.objectCreator' for that bucket. is not right.
You can't add a service account to a role. The relationship is the other way round. You grant roles to the service account. See below a screenshot of the role.
Ref: https://cloud.google.com/iam/docs/granting-roles-to-service-accounts

Create a service account and add it to the IAM role 'storage.objectAdmin' for that bucket. is not right.
You can't add a service account to a role. The relationship is the other way round. You grant roles to the service account.
Ref: https://cloud.google.com/iam/docs/granting-roles-to-service-accounts

Create a service account with an access scope. Use the access scope 'https://www.googleapis.com/auth/cloud-platform'. is the right answer.
cloud-platform role lets you view and manage data across all Google Cloud services. For Cloud Storage, this is the same as devstorage.full-control which allows full control over data, including the ability to modify IAM policies.
Ref: https://cloud.google.com/storage/docs/authentication

.

59. Question

.

You need to trigger a budget alert for Compute Engine charges on one of the three Google Cloud Platform projects that you manage. All three projects are linked to a single billing account. What should you do?

.

- **Verify that you have Billing Account Administrator role. Select the associated billing account and create a budget and alert for the appropriate project.**
- Verify that you have Billing Account Administrator role. Select the associated billing account and create a budget and a custom alert.
- Verify that you have Project Billing Manager role. Select the associated billing account and create a budget for the appropriate project.
- Verify that you have Project Billing Manager role. Select the associated billing account and create a budget and a custom alert.

Unattempted

Verify that you have Project Billing Manager role. Select the associated billing account and create a budget for the appropriate project. is not right.
Project Billing Manager role allows a user to attach the project to the billing account, but does not grant any rights over resources. This role does not provide user permissions to view spending, create budgets and alerts.
Ref: https://cloud.google.com/billing/docs/how-to/billing-access#overview-of-cloud-billing-roles-in-cloud-iam

Verify that you have Project Billing Manager role. Select the associated billing account and create a budget and a custom alert. is not right.
Project Billing Manager role allows a user to attach the project to the billing account, but does not grant any rights over resources. This role does not provide user permissions to view spending, create budgets and alerts.
Ref: https://cloud.google.com/billing/docs/how-to/billing-access#overview-of-cloud-billing-roles-in-cloud-iam

Verify that you have Billing Account Administrator role. Select the associated billing account and create a budget and a custom alert. is not right.
Billing Account Administrator role enables a user to view spend and set budget alerts. But the budget here isn't scoped to the single project that we are interested in. Since the single billing account is linked to all three projects, this results in budget alerts being triggered for Compute Engine usage on all three projects – which is against our requirements.
Ref: https://cloud.google.com/billing/docs/how-to/billing-access#overview-of-cloud-billing-roles-in-cloud-iam
Ref: https://cloud.google.com/billing/docs/how-to/budgets#budget-scope

Verify that you have Billing Account Administrator role. Select the associated billing account and create a budget and alert for the appropriate project. is the right answer.
Billing Account Administrator role enables a user to view spend and set budget alerts. In addition, the budget here is scoped to a single project. Therefore, when the compute engine spend exceeds the budget threshold in the project, we send an alert, and this only works for the scoped project, and not all projects linked to the billing account.
Ref: https://cloud.google.com/billing/docs/how-to/billing-access#overview-of-cloud-billing-roles-in-cloud-iam
Ref: https://cloud.google.com/billing/docs/how-to/budgets#budget-scope

.

60. Question

.

You need to update a deployment in Deployment Manager without any resource downtime in the deployment. Which command should you use?

.

- **gcloud deployment-manager deployments update --config**
- gcloud deployment-manager deployments create --config
- gcloud deployment-manager resources create --config
- gcloud deployment-manager resources update --config

Unattempted

gcloud deployment-manager resources create –config . is not right.
gcloud deployment-manager resources command does not support the action create. The supported actions are describe and list. So this option is not right.
Ref: https://cloud.google.com/sdk/gcloud/reference/deployment-manager/resources

gcloud deployment-manager resources update –config . is not right.
gcloud deployment-manager resources command does not support the action update. The supported actions are describe and list. So this option is not right.
Ref: https://cloud.google.com/sdk/gcloud/reference/deployment-manager/resources

gcloud deployment-manager deployments create –config . is not right.
gcloud deployment-manager deployments create – creates a deployment but we want to update a deployment. So this option is not right.
Ref: https://cloud.google.com/sdk/gcloud/reference/deployment-manager/deployments/create

gcloud deployment-manager deployments update –config . is the right answer.
gcloud deployment-manager deployments update – updates a deployment based on a provided config file and fits our requirement.
https://cloud.google.com/sdk/gcloud/reference/deployment-manager/deployments/update

.

61. Question

.

You plan to deploy an application on an autoscaled managed instances group. The application uses a tomcat server and runs on port 8080. You want to access the application on https://www.example.com. You want to follow Google recommended practices. What services would you use?

.

- Google Domains, Cloud DNS private zone, HTTP(S) Load Balancer
- Google Domains, Cloud DNS private zone, SSL Proxy Load Balancer
- Google DNS, Google CDN, SSL Proxy Load Balancer
- **Google Domains, Cloud DNS, HTTP(S) Load Balancer**

Unattempted

To serve traffic on https://www.example.com, we have to first own the domain example.com. We can use Google Domains service to register a domain.
Ref: https://domains.google/

Once we own example.com domain, we need to create a zone http://www.example.com. We can use Cloud DNS, which is a scalable, reliable, and managed authoritative Domain Name System (DNS) to create a DNS zone.
Ref: https://cloud.google.com/dns

Once the http://www.example.com zone is set up, we need to create a DNS (A) record to point to the public IP of the Load Balancer. This is also carried out in Cloud DNS.

Finally, we need a load balancer to front the autoscaled managed instances group. Google recommends we use HTTP(S) Load Balancer for this requirement as "SSL Proxy Load Balancing is intended for non-HTTP(S) traffic. For HTTP(S) traffic, we recommend that you use HTTP(S) Load Balancing."
Ref: https://cloud.google.com/load-balancing/docs/ssl

So the correct answer is Google Domains, Cloud DNS, HTTP(S) Load Balancer

- **62. Question**

- You ran the following commands to create two compute instances.
gcloud compute instances create instance1
gcloud compute instances create instance2
Both compute instances were created in europe-west2-a zone but you want to create them in other zones. Your active gcloud configuration is as shown below.
$ gcloud config list
[component_manager]
disable_update_check = True
[compute]
gce_metadata_read_timeout_sec = 5
zone = europe-west2-a
[core]
account = gcp-ace-lab-user@gmail.com
disable_usage_reporting = False
project = gcp-ace-lab-266520
[metrics]
environment = devshell

You want to modify the gcloud configuration such that you are prompted for a zone when you execute the create instance commands above. What should you do?

.

- **gcloud config unset compute/zone**
- gcloud config set zone ""
- gcloud config set compute/zone ""
- gcloud config unset zone

Unattempted

gcloud config unset zone. is not right.
gcloud config does not have a core/zone property. The syntax for this command is gcloud config unset SECTION/PROPERTY. If SECTION is missing from the command, SECTION is defaulted to core. We are effectively trying to run the command gcloud config unset core/zone but the core section doesn't have a property called zone, so this command fails.
$ gcloud config unset zone
ERROR: (gcloud.config.unset) Section [core] has no property [zone].
Ref: https://cloud.google.com/sdk/gcloud/reference/config/unset

gcloud config set zone "". is not right.
gcloud config does not have a core/zone property. The syntax for this command is gcloud config set SECTION/PROPERTY VALUE. If SECTION is missing, SECTION is defaulted to core. We are effectively trying to run gcloud config set core/zone "" but the core section doesn't have a property called zone, so this command fails.
$ gcloud config set zone ""
ERROR: (gcloud.config.unset) Section [core] has no property [zone].
Ref: https://cloud.google.com/sdk/gcloud/reference/config/set

gcloud config set compute/zone "". is not right.
This command uses the correct syntax but it doesn't unset the compute/zone property correctly. Instead it sets it to "" in gcloud configuration. When the gcloud compute instances create command runs, it picks the zone value from this configuration property which is "" and attempts to create an instance in "" zone and fails because zone "" doesn't exist. gcloud doesn't treat "" zone as an unset value. The zone must be explicitly unset if it is to be removed from the configuration.
$ gcloud config set compute/zone ""
$ gcloud compute instances create instance1
Zone: Expected type (,) for field id, found projects/compute-challenge-lab-266520/zones/ (type)
Ref: https://cloud.google.com/sdk/gcloud/reference/config/set

gcloud config unset compute/zone. is the right answer.
This command uses the correct syntax and correctly unsets the zone in gcloud configuration. The next time gcloud compute instances create command runs, it knows there is no default zone defined in gcloud configuration and therefore prompts for a zone before the instance can be created.
Ref: https://cloud.google.com/sdk/gcloud/reference/config/unset

.

63. Question

.

You recently deployed a new application in Google App Engine to serve production traffic. After analyzing logs for various user flows, you uncovered several issues in your application code and have developed a fix to address the issues. Parts of your proposed fix could not be validated in the pre-production environment by your testing team as some of the scenarios can only be validated by an end-user with access to specific data in your production environment. In the company's weekly Change Approval Board meeting, concerns were raised that the fix could possibly take down the application. It was unanimously agreed that while the fix is risky, it is a necessary change to the application. You have been asked to suggest a solution that minimizes the impact of the change going wrong. You also want to minimize costs. What should you do?

.

- **Deploy a new version of the application, and use traffic splitting to send a small percentage of traffic to it.**
- Set up a second Google App Engine service, and then update a subset of clients to hit the new service.
- Deploy the new application version temporarily, capture logs and then roll it back to the previous version.
- Create a second Google App Engine project with the new application code, and onboard users gradually to the new application.

Unattempted

Deploy the new application version temporarily, capture logs and then roll it back to the previous version. is not right.
Deploying a new application version and promoting it would result in your new version serving all production traffic. If the code fix doesn't work as expected, it would result in the application becoming unreachable to all users. This is a risky approach and should be avoided.

Create a second Google App Engine project with the new application code, and onboard users gradually to the new application. is not right.

You want to minimize costs. This approach effectively doubles your costs as you have to pay for two identical environments until all users are moved over to the new application. There is an additional overhead of manually onboarding users to the new application which could be expensive as well as time-consuming.

Set up a second Google App Engine service, and then update a subset of clients to hit the new service. is not right.
It is not straightforward to update a set of clients to hit the new service. When users access an App Engine service, they use an endpoint like https://SERVICE_ID-dot-PROJECT_ID.REGION_ID.r.appspot.com. Introducing a new service introduces a new URL and getting your users to use the new URL is possible but involves effort and coordination. If you want to mask these differences to the end-user, then you have to make changes in the DNS and use a weighted algorithm to split the traffic between the two services based on the weights assigned.
Ref: https://cloud.google.com/appengine/docs/standard/python/splitting-traffic
Ref: https://cloud.google.com/appengine/docs/standard/python/an-overview-of-app-engine
This approach also has the drawback of doubling your costs until all users are moved over to the new service.

Deploy a new version of the application, and use traffic splitting to send a small percentage of traffic to it. is the right answer.
This option minimizes the risk to the application while also minimizing the complexity and cost. When you deploy a new version to App Engine, you can choose not to promote it to serve live traffic. Instead, you could set up traffic splitting to split traffic between the two versions – this can all be done within Google App Engine. Once you send a small portion of traffic to the new version, you can analyze logs to identify if the fix has worked as expected. If the fix hasn't worked, you can update your traffic splitting configuration to send all traffic back to the old version. If you are happy your fix has worked, you can send more traffic to the new version or move all user traffic to the new version and delete the old version.

Ref: https://cloud.google.com/appengine/docs/standard/python/splitting-traffic
Ref: https://cloud.google.com/appengine/docs/standard/python/an-overview-of-app-engine

.

64. Question

.

You recently deployed a new version of an application to App Engine and then discovered a bug in the release. You need to immediately revert to the prior version of the application. What should you do?

.

- On the App Engine page of the GCP Console, select the application that needs to be reverted and click Revert.
- **On the App Engine Versions page of the GCP Console, route 100% of the traffic to the previous version.**
- Deploy the original version as a separate application. Then go to App Engine settings and split traffic between applications so that the original version serves 100% of the requests.
- Run gcloud app restore.

Unattempted

Run gcloud app restore. is not right.
restore action is not supported by gcloud app command.
Ref: https://cloud.google.com/sdk/gcloud/reference/app/deploy

On the App Engine page of the GCP Console, select the application that needs to be reverted and click Revert. is not right.
Revert option is not present on the App Engine page of the GCP Console.

Deploy the original version as a separate application. Then go to App Engine settings and split traffic between applications so that the original version serves 100% of the requests. is not right.
Each application in the app engine is different and it is not possible to split traffic between applications in App Engine. You can use traffic splitting to specify a percentage distribution of traffic across two or more of the versions within a service but not across applications.
Ref: https://cloud.google.com/appengine/docs/standard/python/splitting-traffic

On the App Engine Versions page of the GCP Console, route 100% of the traffic to the previous version. is the right answer
You can roll back to a previous version in the app engine GCP console. Go back to the list of versions and check the box next to the version that you want to receive all traffic and click the MAKE DEFAULT button located above the list. Traffic immediately switches over to the selected version.
Ref: https://cloud.google.com/community/tutorials/how-to-roll-your-app-engine-managed-vms-app-back-to-a-previous-version-part-1

.

65. Question

.

You significantly changed a complex deployment manager template and want to confirm that the dependencies of all defined resources are properly met before committing it to the project. You want the most rapid feedback on your changes. What would you do?

.

- Use granular logging statements within the Deployment Manager template authored in Python.
- Monitor activity of the Deployment Manager executing on Stackdriver logging page of the GCP console.
- Execute the Deployment manager template against a separate project with the same configuration, and monitor for failures
- **Execute the Deployment Manager template using the --preview option in the same project, and observe the status of interdependent resources**

Unattempted

Requirements – confirm dependencies, rapid feedback.

Use granular logging statements within the Deployment Manager template authored in Python. is not right.
Deployment Manager doesn't provide the ability to set granular logging statements. Moreover, if that was possible the logging statements wouldn't be written to a log file until the template is applied and it is already too late as the template is applied and we haven't had a chance to confirm that the dependencies of all defined resources are properly met

Monitor activity of the Deployment Manager executing on Stackdriver logging page of the GCP console. is not right.
This doesn't give us a chance to confirm that the dependencies of all defined resources are properly met before executing it.

Execute the Deployment manager template against a separate project with the same configuration, and monitor for failures. is not right.
While we can identify whether dependencies are met by monitoring the failures, it is not rapid. We need rapid feedback on changes and we want that before changes are committed (i.e. applied) to the project

Execute the Deployment Manager template using the –preview option in the same project, and observe the status of interdependent resources. is the right answer.
After we have written a configuration file, we can preview the configuration before you create a deployment. Previewing a configuration lets you see the resources that Deployment Manager would create but does not actually instantiate any actual resources. In gcloud command-line, you use the create sub-command with the –preview flag to preview configuration changes.
Ref: https://cloud.google.com/deployment-manager

.

66. Question

.

You want to add a new auditor to a Google Cloud Platform project. The auditor should be allowed to read, but not modify, all project items. How should you configure the auditor's permissions?

.

- Create a custom role with view-only project permissions. Add the user's account to the custom role.
- Create a custom role with view-only service permissions. Add the user's account to the custom role.
- **Select the built-in IAM project Viewer role. Add the user's account to this role.**
- Select the built-in IAM service Viewer role. Add the user's account to this role.

Unattempted

Select the built-in IAM project Viewer role. Add the user's account to this role. Is the right answer

The primitive role roles/viewer provides read access to all resources in the project. The permissions in this role are limited to Get and list access for all resources. As we have an out of the box role that exactly fits our requirement, we should use this.
Ref: https://cloud.google.com/resource-manager/docs/access-control-proj
It is advisable to use the existing GCP provided roles over creating custom roles with similar permissions as this becomes a maintenance overhead. If GCP modifies how permissions are handled or adds/removes permissions, the default GCP provided roles are automatically updated by Google whereas if they were custom roles, the responsibility is with us and this adds to the operational overhead and needs to be avoided.

.

67. Question

.

You want to configure 10 Compute Engine instances for availability when maintenance occurs. Your requirements state that these instances should attempt to automatically restart if they crash. Also, the instances should be highly available including during system maintenance. What should you do?

.

- Create an instance group for the instance. Verify that the Advanced creation options setting for do not retry machine creation is set to off.
- Create an instance group for the instances. Set the Autohealing health check to healthy (HTTP).
- Create an instance template for the instances. Set Automatic Restart to off. Set On-host maintenance to Terminate VM instances. Add the instance template to an instance group.
- **Create an instance template for the instances. Set the Automatic Restart to on. Set the On-host maintenance to Migrate VM instance. Add the instance template to an instance group.**

Unattempted

Requirements
1. 10 instances – indicates we need to look for MIG (Managed Instances Group) where we can configure healing/scaling settings. All options talk about creating an instance group so this point isn't of much use, unfortunately.
2. Highly available during system maintenance – indicates we need to look for Live Migration.
3. Automatically restart on crash – indicates we need to look for options that enable automatic restarts.

Create an instance template for the instances.
Set Automatic Restart to off.
Set On-host maintenance to Terminate VM instances.
Add the instance template to an instance group. is not right.
If Automatic Restart is off, then the compute engine instances are not automatically restarted. This results in loss of capacity and if GCP decides to start system maintenance on all instances at the same time, all instances are down and this does not meet our requirement "Highly available during system maintenance" so this option is not right.

Create an instance group for the instances.
Set the Autohealing health check to healthy (HTTP). is not right.

While auto-healing helps with the recreation of VM instances when needed, it doesn't Live-migrate the instances so our requirement of "highly available including during system maintenance" is not met. More info about Autohealing – Auto-healing allows the recreation of VM instances when needed. You can use a health check to recreate a VM instance if the health check finds it unresponsive. If you don't select a health check, Compute Engine will recreate VM instances only when they're not running.
Ref: https://cloud.google.com/compute/docs/instance-groups/?hl=en_GB#managed_instance_groups_and_autohealing

Create an instance group for the instance.
Verify that the Advanced creation options setting for do not retry machine creation is set to off. is not right.
Like above – this option doesn't Live-migrate the instances so our requirement of "highly available including during system maintenance" is not met.

Create an instance template for the instances.
Set the Automatic Restart to on.
Set the On-host maintenance to Migrate VM instance.
Add the instance template to an instance group. is the right option.
Enabling automatic restart ensures that compute engine instances are automatically restarted when they crash. And Enabling "Migrate VM Instance" enables live migrates i.e. compute instances are migrated during system maintenance and remain running during the migration.
Automatic Restart – If your instance is set to terminate when there is a maintenance event, or if your instance crashes because of an underlying hardware issue, you can set up Compute Engine to automatically restart the instance by setting
the automaticRestart field to true. This setting does not apply if the instance is taken offline through a user action, such as calling sudo shutdown, or during a zone outage.
Ref: https://cloud.google.com/compute/docs/instances/setting-instance-scheduling-options#autorestart
Enabling the Migrate VM Instance option migrates your instance away from an infrastructure maintenance event, and your instance remains running during the migration. Your instance might experience a short period of decreased performance, although generally, most instances should not notice any difference. This is ideal for instances that require constant uptime and can tolerate a short period of decreased performance.
Ref: https://cloud.google.com/compute/docs/instances/setting-instance-scheduling-options#live_migrate

.

68. Question

.

You want to configure a cost-effective solution for archiving objects in a Cloud Storage bucket. Noncurrent versions should be archived after 30 days. Non-current versions are accessed once a month for reporting. This archived objects are also occasionally updated at month-end. What should you do?

.

- Add a bucket lifecycle rule that archives noncurrent versions after 30 days to Coldline Storage.
- **Add a bucket lifecycle rule that archives noncurrent versions after 30 days to Nearline Storage.**
- Add a bucket lifecycle rule that archives objects from regional storage after 30 days to Coldline Storage.
- Add a bucket lifecycle rule that archives objects from regional storage after 30 days to Nearline Storage.

Unattempted

We don't know what the current storage class is. In the absence of this information and considering the 4 options provided, it is safe to assume that objects are currently in Regional or Multi-Regional buckets. We want to archive noncurrent versions after 30 days and you need to read and modify on average once per month

Add a bucket lifecycle rule that archives noncurrent versions after 30 days to Coldline Storage. is not right.
Coldline Storage is ideal for data you plan to read or modify at most once a quarter. Our requirement states we need to read or modify every month so Coldline Storage is not an ideal storage class for our requirement.
Ref: https://cloud.google.com/storage/docs/storage-classes#coldline

Add a bucket lifecycle rule that archives objects from regional storage after 30 days to Coldline Storage. is not right.
Coldline Storage is ideal for data you plan to read or modify at most once a quarter. Our requirement states we need to read or modify every month so Coldline Storage is not an ideal storage class for our requirement. Moreover, we don't want to archive live versions, we want to archive just the noncurrent versions.
Ref: https://cloud.google.com/storage/docs/storage-classes#coldline

Add a bucket lifecycle rule that archives objects from regional storage after 30 days to Nearline Storage. is not right.
While Nearline Storage is ideal for data you plan to read or modify on average once per month or less, we don't want to archive live versions, we want to archive just the

noncurrent versions.
Ref: https://cloud.google.com/storage/docs/storage-classes#nearline

Add a bucket lifecycle rule that archives noncurrent versions after 30 days to Nearline Storage. is the right answer.
Nearline Storage is ideal for data you plan to read or modify on average once per month or less. And this option archives just the noncurrent versions which is what we want to do.
https://cloud.google.com/storage/docs/storage-classes#nearline

.

69. Question

.

You want to configure an SSH connection to a single Compute Engine instance for users in the dev1 group. This instance is the only resource in this particular Google Cloud Platform project that the dev1 users should be able to connect to. What should you do?

.

- **Set metadata to enable-oslogin=true for the instance. Grant the dev1 group the compute.osLogin role. Direct them to use the Cloud Shell to ssh to that instance.**
- Set metadata to enable-oslogin=true for the instance. Set the service account to no service account for that instance. Direct them to use the Cloud Shell to ssh to that instance.
- Enable block project wide keys for the instance. Generate an SSH key for each user in the dev1 group. Distribute the keys to dev1 users and direct them to use their third-party tools to connect.
- Enable block project wide keys for the instance. Generate an SSH key and associate the key with that instance. Distribute the key to dev1 users and direct them to use their third-party tools to connect.

Unattempted

You have multiple ways to connect to instances. More information can be found here: https://cloud.google.com/compute/docs/instances/access-overview

Enable block project wide keys for the instance. Generate an SSH key for each user in the dev1 group. Distribute the keys to dev1 users and direct them to use their third-party tools to connect. is not right.
Generating SSH keys for users is fine but unless the SSH keys are added to the instance, users would not be able to SSH to the server. If you need your instance to ignore project-wide public SSH keys and use only the instance-level keys, you can block project-wide

public SSH keys from the instance. This allows only users whose public SSH key is stored in instance-level metadata to access the instance.
Ref: https://cloud.google.com/compute/docs/instances/adding-removing-ssh-keys#edit-ssh-metadata

Enable block project wide keys for the instance. Generate an SSH key and associate the key with that instance. Distribute the key to dev1 users and direct them to use their third-party tools to connect. is not right.
While this is possible, sharing SSH keys is a strict NO from a security point of view as this breaks auditing. Should one of the developers create a disaster (either accidental or malicious), your security admin would be unable to identify which of the users in dev1 group caused the issue.

Set metadata to enable-oslogin=true for the instance. Set the service account to no service account for that instance. Direct them to use the Cloud Shell to ssh to that instance. is not right.
After you enable OS Login on one or more instances in your project, those VMs accept connections only from user accounts that have the necessary IAM roles in your project or organization. In this case, since we have not granted either of these roles – roles/compute.osLogin or roles/compute.osAdminLogin role, users of dev1 group can't SSH to the server.
Ref: https://cloud.google.com/compute/docs/instances/managing-instance-access#configure_users

Set metadata to enable-oslogin=true for the instance. Grant the dev1 group the compute.osLogin role. Direct them to use the Cloud Shell to ssh to that instance. is the right answer.
After you enable OS Login on one or more instances in your project, those VMs accept connections only from user accounts that have the necessary IAM roles in your project or organization. In this case, we are granting the group compute.osLogin which lets them log in as non-administrator account. And since we are directing them to use Cloud Shell to ssh, we don't need to add their SSH keys to the instance metadata.
Ref: https://cloud.google.com/compute/docs/instances/managing-instance-access#configure_users
Ref: https://cloud.google.com/compute/docs/instances/managing-instance-access#add_oslogin_keys

- **70. Question**
-

You want to configure auto-healing for network load balancer for a group of Compute Engine instances that run in multiple zones using the fewest possible steps. You need to configure the recreation of VMs if they are unresponsive after 3 attempts of 10 seconds each. What should you do?

.

- Create a HTTP load balancer with a backend configuration that references an existing instance group. Define a balancing mode and set the maximum RPS to 10
- Create a HTTP load balancer with a backend configuration that references an existing instance group. Set the health check to healthy (HTTP)
- Create a managed instance group. Verify that the auto-scaling setting is on.
- **Create a managed instance group. Set the Autohealing health check to healthy (HTTP)**