DUMPSARENA

Google Cloud Certified - Associate Cloud Engineer

Google Associate-Cloud-Engineer

Version Demo

Total Demo Questions: 41

Total Premium Questions: 487

Buy Premium PDF

https://dumpsarena.co

sales@dumpsarena.co

sales@dumpsarena.co
dumpsarena.co

Topic Break Down

Topic	No. of Questions
Topic 1, Setting up a cloud solution environment	44
Topic 2, Planning and configuring a cloud solution	139
Topic 3, Deploying and implementing a cloud solution	80
Topic 4, Ensuring successful operation of a cloud solution	96
Topic 5, Configuring access and security	127
Topic 6, Mix Questions	1
Total	487

QUESTION NO: 1

You have created a new project in Google Cloud through the gcloud command line interface (CLI) and linked a billing account. You need to create a new Compute

Engine instance using the CLI. You need to perform the prerequisite steps. What should you do?

A. Create a Cloud Monitoring Workspace.

B. Create a VPC network in the project.

C. Enable the compute googleapis.com API.

D. Grant yourself the IAM role of Compute Admin.

ANSWER: C

Explanation:

Enable the compute googleapis.com API. is the required prerequisite because Compute Engine resources can be created only after the Compute Engine API is enabled in the target project. Creating a project and associating an active billing account make the project eligible to use billable Google Cloud services, but APIs are enabled on a per-project basis. The Compute Engine API provides the service endpoint that the `gcloud compute` command group uses to create and manage virtual machine instances, disks, networks, and related resources.

You can enable it from the CLI with `gcloud services enable compute.googleapis.com --project=PROJECT_ID`. The identity running that command must have permission to enable services, such as through the Service Usage Admin role or an equivalent custom role. After the API is enabled, the CLI can submit the instance-creation request, assuming the user also has the necessary Compute Engine permissions. Google documents enabling APIs through Service Usage at Enable and disable services, and lists the Compute Engine API service name and setup guidance at Enabling and using the Compute Engine API.

QUESTION NO: 2

You are planning to migrate the following on-premises data management solutions to Google Cloud:

â€¢ One MySQL cluster for your main database

â€¢ Apache Kafka for your event streaming platform

â€¢ One Cloud SOL for PostgreSOL database for your analytical and reporting needs

You want to implement Google-recommended solutions for the migration. You need to ensure that the new solutions provide global scalability and require minimal operational and infrastructure management. What should you do?

A. Migrate from MySQL to Cloud SQL, from Kafka to Memorystore, and from Cloud SQL for PostgreSQL to Cloud SQL

B. Migrate from MySQL to Cloud Spanner, from Kafka to Memorystore, and from Cloud SOL for PostgreSQL to Cloud SQL

C. Migrate from MySQL to Cloud SOL, from Kafka to Pub/Sub, and from Cloud SOL for PostgreSQL to BigQuery.

D. Migrate from MySQL to Cloud Spanner, from Kafka to Pub/Sub. and from Cloud SQL for PostgreSQL to BigQuery

ANSWER: D

Explanation:

Migrate from MySQL to Cloud Spanner, from Kafka to Pub/Sub. and from Cloud SQL for PostgreSQL to BigQuery is the best fit because each selected managed service aligns with the stated workload and minimizes infrastructure administration. Cloud Spanner is a fully managed relational database designed for horizontal scalability and strong consistency across regions, making it appropriate when the main transactional database needs global scale and availability. Pub/Sub is a fully managed, globally available messaging service that provides durable asynchronous event ingestion and delivery without requiring management of Kafka brokers, partitions, or cluster capacity. BigQuery is a serverless data warehouse built for

large-scale analytics and reporting; it separates storage and compute and automatically manages the underlying infrastructure, allowing analytical workloads to scale without operating database servers. Together, these services use Google-managed control planes and elastic capacity rather than self-managed clusters, supporting the requirement for minimal operational effort while addressing transactional, streaming, and analytical use cases with purpose-built services. See Cloud Spanner overview and BigQuery introduction.

QUESTION NO: 3

Your company has a large BigQuery table containing several years of event data. Analysts commonly query a limited date range and filter by customer ID. You want to reduce the amount of data scanned by these queries. Which two actions should you take? (Choose two.)

A. Partition the table by the event date or timestamp.

B. Cluster the table by customer ID.

C. Create a separate BigQuery project for each analyst.

D. Export the table to Cloud Storage before analysts run queries.

E. Grant the analysts the BigQuery Admin role.

ANSWER: A B

Explanation:

Partition the table by the event date or timestamp. Partitioning enables BigQuery to prune partitions when a query filters on the partitioning column, so queries that request a limited date range scan only the relevant partitions.

Cluster the table by customer ID. Clustering organizes data based on the clustering columns and can reduce scanned data for queries that filter on customer ID, especially when used with partitioning. See Introduction to partitioned tables and Introduction to clustered tables.

QUESTION NO: 4

You are building an application that stores relational data from users. Users across the globe will use this application. Your CTO is concerned about the scaling requirements because the size of the user base is unknown. You need to implement a database solution that can scale with your user growth with minimum configuration changes. Which storage solution should you use?

A. Cloud SQL

B. Cloud Spanner

C. Cloud Firestore

D. Cloud Datastore

ANSWER: B

Explanation:

Cloud Spanner is the appropriate choice because it is a fully managed, relational database designed to scale horizontally while maintaining strong consistency and SQL support. It is specifically suited to globally distributed applications that need to store relational data and serve users in multiple geographic locations. Cloud Spanner provides automatic sharding, synchronous replication, and automatic scaling capabilities, so the team does not need to redesign the data architecture or manually manage database partitions as the user base grows. Its multi-region configurations can also provide high availability and resilience across geographic regions, which aligns with the requirement for a worldwide application.

Cloud Spanner supports transactional consistency across the database at global scale, allowing the application to retain familiar relational modeling and SQL querying while accommodating uncertain and potentially substantial growth. Capacity

can be adjusted by changing compute capacity, and managed autoscaling is available for eligible configurations, minimizing operational configuration as demand changes. Google describes Spanner as a fully managed, horizontally scalable relational database with strong consistency and global availability. See the Cloud Spanner overview and Spanner managed autoscaling documentation.

QUESTION NO: 5

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

A. 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 created.

B. 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 each VPC.3. Use iptables on this instance to forward traffic from gce-network to the GKE nodes.4. Configure the Compute Engine instance to use the address of proxy in gce-network as endpoint.

C. 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: Internal3. Peer the two VPCs together.4. Configure the Compute Engine instance to use the address of the load balancer that has been created.

D. 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.

ANSWER: C

Explanation:

Creating an internal LoadBalancer Service and peering the two VPC networks is the lowest-effort way to provide private TCP connectivity from the Compute Engine VM to the replicated GKE workload. The Service causes GKE to provision an internal passthrough Network Load Balancer with a private virtual IP address and to distribute incoming TCP connections across the Pods selected by the Service. This gives the VM one stable internal endpoint while allowing the application and cluster autoscaler to manage Pod and node changes without requiring manual proxy configuration or per-node endpoints.

VPC Network Peering is appropriate because the networks have non-overlapping address ranges. Peering exchanges subnet routes between the VPCs, enabling the VM in `gce-network` to reach the internal load balancer virtual IP in the GKE VPC over Google's private network. The VM can then use that internal IP address as its application endpoint. Appropriate firewall rules must permit the TCP port from the VM's source range to the load balancer and its backends. GKE documents the `cloud.google.com/load-balancer-type: Internal` annotation for creating an internal LoadBalancer Service, and Google Cloud documents that internal load-balancer access can be provided from a peered VPC. See GKE internal load balancing and VPC Network Peering.

QUESTION NO: 6

You have created an application that is packaged into a Docker image. You want to deploy the Docker image as a workload on Google Kubernetes Engine. What should you do?

A. Upload the image to Cloud Storage and create a Kubernetes Service referencing the image.

B. Upload the image to Cloud Storage and create a Kubernetes Deployment referencing the image.

C. Upload the image to Container Registry and create a Kubernetes Service referencing the image.

D. Upload the image to Artifact Registry and create a Kubernetes Deployment referencing the image.

ANSWER: D

Explanation:

Upload the image to Artifact Registry and create a Kubernetes Deployment referencing the image. is correct because a GKE workload must obtain its container image from a container image registry that the cluster can access, and Artifact Registry is Google Cloud's current managed service for storing and distributing Docker and OCI images. After pushing the image to an Artifact Registry repository, the Deployment manifest specifies that image in its container definition. Kubernetes then uses the Deployment controller to create and maintain the required Pods, including replacing failed Pods and supporting declarative updates to the application.

A Deployment is the appropriate Kubernetes workload resource for a typical stateless containerized application because it defines the desired replica count and Pod template. The image reference should use the Artifact Registry hostname and repository path, such as `REGION-docker.pkg.dev/PROJECT/REPOSITORY/IMAGE:TAG`. The GKE node or workload identity must also have permission to read the repository, normally through the Artifact Registry Reader role. Google recommends Artifact Registry for container images; Container Registry has been shut down for writes and is no longer the service to use for new image deployments. See Deploying workloads to GKE and Store Docker container images in Artifact Registry.

QUESTION NO: 7

Your company ' s management has chosen Google Cloud as its strategic cloud provider. You need to provision Google Cloud identities for all employees in the IT and Data departments but not to other departments. This access should automatically update when employees change departments, join, or leave the company. All employees are managed centrally in an LDAP directory, with one organizational unit (OU) for each department. What should you do?

A. Create a CSV file with the users from the relevant OUs. Upload the CSV file to Cloud Identity. Instruct your Human Resources department to repeat this process for any employee changes.

B. Develop a containerized application that reads user information from the relevant OUs in the LDAP directory. Use the Admin SDK to synchronize user accounts in Identity Platform with the LDAP directory. Deploy the application as a scheduled Cloud Run job.

C. Develop a containerized application that reads user information from the relevant OUs in the LDAP directory. Use the Directory API to create accounts in Cloud Identity for new users. Deploy the application as a scheduled Cloud Run job.

D. Configure Google Cloud Directory Sync to synchronize the relevant OUs between the LDAP directory and Cloud Identity. Set up a scheduled task that triggers a provisioning run.

ANSWER: D

Explanation:

Configure Google Cloud Directory Sync to synchronize the relevant OUs between the LDAP directory and Cloud Identity. Set up a scheduled task that triggers a provisioning run. Google Cloud Directory Sync (GCDS) is designed specifically to provision and manage Google Workspace or Cloud Identity users and groups from an existing LDAP directory. It can be configured to query only selected LDAP organizational units, which limits provisioning to the IT and Data populations rather than creating identities for every employee in the directory.

GCDS supports recurring synchronization through an external scheduler, such as Windows Task Scheduler or cron. Each run compares the configured LDAP source with Cloud Identity and applies the required changes. Therefore, new employees in the selected department OUs can receive accounts automatically, users removed from those OUs or from LDAP can be deprovisioned according to the synchronization configuration, and employees who move between departments are brought into or removed from the in-scope population based on their current LDAP placement. This provides centralized lifecycle management using the company's existing authoritative identity directory, with no need for custom account-provisioning code. Google documents GCDS as the tool for synchronizing user accounts and groups from LDAP directories: Google Cloud Directory Sync overview. The supported scheduling approach is described in Automate synchronization with GCDS.

QUESTION NO: 8

You are building an application that processes data files uploaded from thousands of suppliers. Your primary goals for the application are data security and the expiration of aged data. You need to design the application to:

• Restrict access so that suppliers can access only their own data.

• Give suppliers write access to data only for 30 minutes.

• Delete data that is over 45 days old.

You have a very short development cycle, and you need to make sure that the application requires minimal maintenance. Which two strategies should you use? (Choose two.)

A. Build a lifecycle policy to delete Cloud Storage objects after 45 days.

B. Use signed URLs to allow suppliers limited time access to store their objects.

C. Set up an SFTP server for your application, and create a separate user for each supplier.

D. Build a Cloud function that triggers a timer of 45 days to delete objects that have expired.

E. Develop a script that loops through all Cloud Storage buckets and deletes any buckets that are older than 45 days.

ANSWER: A B

Explanation:

Build a lifecycle policy to delete Cloud Storage objects after 45 days because Cloud Storage Object Lifecycle Management automatically performs configured actions as objects meet specified age conditions. An age-based Delete rule removes objects once they are older than the configured number of days, without requiring application code, scheduled jobs, or operational maintenance. This directly implements the 45-day retention requirement using a managed Cloud Storage capability. Google documents lifecycle configuration at https://cloud.google.com/storage/docs/lifecycle.

Use signed URLs to allow suppliers limited time access to store their objects because a signed URL grants time-limited access to a specific Cloud Storage resource without giving the supplier broad bucket-level IAM permissions. The application can create a separate signed upload URL for each supplier and object, set the URL expiration to 30 minutes, and use object naming or prefixes that associate the upload with that supplier. This minimizes development and maintenance while limiting upload authority to the intended object and time window. Cloud Storage signed URLs support specifying an expiration time and are documented at https://cloud.google.com/storage/docs/access-control/signed-urls.

QUESTION NO: 9

You are deploying an application to GKE. The application needs credentials for an external payment service. You must avoid storing long-lived credentials in container images or Kubernetes manifest files, and the application must retrieve only the secret it needs. Which two actions should you take? (Choose two.)

A. Store the payment-service credential in Secret Manager.

B. Create a dedicated Google service account, grant it Secret Manager Secret Accessor on the required secret, and use Workload Identity Federation for GKE.

C. Add the credential as an environment variable directly in the Deployment manifest.

D. Include the credential in the container image during the Docker build process.

E. Create a service account key for the default Compute Engine service account and mount the key in every Pod.

ANSWER: A B

Explanation:

Store the payment-service credential in Secret Manager. Secret Manager is designed to centrally store, manage, and audit sensitive values such as API keys and passwords. This keeps the secret out of container images and source-controlled Kubernetes manifests.

Create a dedicated Google service account for the workload and grant it the Secret Manager Secret Accessor role only on the required secret. Configure GKE Workload Identity Federation for GKE so that the Kubernetes service account used by the application can authenticate as that Google service account. This provides short-lived credentials and least-privilege secret access without distributing service account keys. See Secret Manager overview and Workload Identity Federation for GKE.

QUESTION NO: 10

Your company implemented BigQuery as an enterprise data warehouse. Users from multiple business units run queries on this data warehouse. However, you notice that query costs for BigQuery are very high, and you need to control costs. Which two methods should you use? (Choose two.)

A. Split the users from business units to multiple projects.

B. Apply a user- or project-level custom query quota for BigQuery data warehouse.

C. Create separate copies of your BigQuery data warehouse for each business unit.

D. Split your BigQuery data warehouse into multiple data warehouses for each business unit.

E. Change your BigQuery pricing model from on-demand to capacity-based pricing. Apply the appropriate number of slots to each project.

ANSWER: B E

Explanation:

Applying a user- or project-level custom query quota for the BigQuery data warehouse directly limits the amount of query data that a specified user or project can consume within a day. This creates a guardrail against unexpectedly expensive on-demand queries while allowing administrators to allocate appropriate query capacity to individual business units. Custom query quotas are especially useful where different teams share data but need independently enforceable usage limits.

Using BigQuery capacity-based pricing with an appropriate number of slots assigned to each project changes spending from query-by-query, bytes-scanned charges to predictable capacity commitments. BigQuery reservations let an organization allocate dedicated or shared slot capacity to workloads and projects, helping prevent one business unit's workload from consuming capacity intended for another. Administrators can size commitments according to observed demand and use autoscaling where appropriate, making the cost model more controllable for sustained enterprise warehouse usage. Google has replaced the former flat-rate terminology with capacity-based pricing, but the cost-control principle remains the same. See BigQuery custom query quotas and BigQuery reservations and capacity management.

QUESTION NO: 11

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?

A. Create a Google Kubernetes Engine cluster, and use horizontal pod autoscaling to scale the application.

B. Create an instance template, and use the template in a managed instance group with autoscaling configured.

C. Create an instance template, and use the template in a managed instance group that scales up and down based on the time of day.

D. Use a set of third-party tools to build automation around scaling the application up and down, based on Stackdriver CPU usage monitoring.

ANSWER: B

Explanation:

Create an instance template, and use the template in a managed instance group with autoscaling configured. This is the native Compute Engine approach for running an application directly on virtual machines while automatically adjusting capacity in response to CPU utilization. An instance template defines the repeatable VM configuration, including the machine type, boot disk image, startup script, service account, network settings, and application installation. A managed instance group uses that template to create identical VM instances, maintain the desired fleet state, and replace unhealthy instances automatically.

Compute Engine autoscaling can use average CPU utilization as its target metric. When demand raises the group's observed CPU utilization above the configured target, the managed instance group adds instances; when utilization falls, it removes excess instances while observing stabilization behavior. This avoids custom automation and provides a managed, operationally efficient scaling mechanism. For quicker scale-out, the group can be configured with an appropriate initialization period and, where needed, proactive instance initialization so instances are prepared ahead of anticipated demand. See Compute Engine autoscaling documentation and managed instance group documentation.

QUESTION NO: 12

You want to configure a solution for archiving data in a Cloud Storage bucket. The solution must be cost-effective. Data with multiple versions should be archived after 30 days. Previous versions are accessed once a month for reporting. This archive data is also occasionally updated at month-end. What should you do?

A. Add a bucket lifecycle rule that archives data with newer versions after 30 days to Coldline Storage.

B. Add a bucket lifecycle rule that archives data with newer versions after 30 days to Nearline Storage.

C. Add a bucket lifecycle rule that archives data from regional storage after 30 days to Coldline Storage.

D. Add a bucket lifecycle rule that archives data from regional storage after 30 days to Nearline Storage.

ANSWER: B

Explanation:

Add a bucket lifecycle rule that archives data with newer versions after 30 days to Nearline Storage. This approach uses Cloud Storage Object Lifecycle Management to automatically change the storage class of older object versions after the required 30-day period. A lifecycle condition targeting objects that have newer versions ensures that the rule applies to noncurrent versions rather than the active version of an object, which preserves the current data for normal use.

Nearline Storage is the appropriate archival class for this access pattern because it is designed for data accessed roughly once per month or less. The prior versions are retrieved monthly for reporting and can also be modified at month-end, so they remain available without the longer minimum storage duration associated with colder archival classes. Nearline has a 30-day minimum storage duration, aligning directly with the required archival timing and supporting a cost-effective balance between storage charges and the expected monthly access pattern. Lifecycle rules perform this transition automatically as eligible object versions age, reducing administrative overhead while retaining versioned data for reporting and updates.

See Cloud Storage Object Lifecycle Management and Cloud Storage storage classes.

QUESTION NO: 13

Your company uses BigQuery to store and analyze data. Upon submitting your query in BigQuery, the query fails with a quotaExceeded error. You need to diagnose the issue causing the error. What should you do?

Choose 2 answers

A. Search errors in Cloud Audit Logs to analyze the issue.

B. Configure Cloud Trace to analyze the issue.

C. View errors in Cloud Monitoring to analyze the issue.

D. Use the information schema views to analyze the underlying issue.

E. Use BigQuery BI Engine to analyze the issue.

ANSWER: A C

Explanation:

“Search errors in Cloud Audit Logs to analyze the issue.” is correct because BigQuery audit logs record job and API activity, including failed requests. Reviewing the relevant BigQuery job or API request in Cloud Logging can identify the principal, project, operation, timestamp, and error details associated with the quota failure. This provides the operational context needed to determine which workload or request pattern is consuming the limited resource.

“View errors in Cloud Monitoring to analyze the issue.” is also correct because Google Cloud exposes BigQuery quota usage and limit metrics through Cloud Monitoring. These metrics allow administrators to compare current quota consumption with the configured limit, identify whether the problem is a sustained quota exhaustion or a traffic spike, and determine the affected quota dimension. Monitoring can also support alerting before usage reaches a limit, helping teams correlate the failure with changes in query volume or workload behavior. Together, audit logs provide request-level evidence and Cloud Monitoring provides quota-level usage visibility, which are complementary for diagnosing a `quotaExceeded` error. See Google’s BigQuery quota troubleshooting guidance and its documentation for monitoring BigQuery.

QUESTION NO: 14

You are building a product on top of Google Kubernetes Engine (GKE). You have a single GKE cluster. For each of your customers, a Pod is running in that cluster, and your customers can run arbitrary code inside their Pod. You want to maximize the isolation between your customers’ Pods. What should you do?

A. Use Binary Authorization and whitelist only the container images used by your customers’ Pods.

B. Use the Container Analysis API to detect vulnerabilities in the containers used by your customers’ Pods.

C. Create a GKE node pool with a sandbox type configured to gvisor. Add the parameter runtimeClassName: gvisor to the specification of your customers’ Pods.

D. Use the cos_containerd image for your GKE nodes. Add a nodeSelector with the value cloud.google.com/gke-os-distribution: cos_containerd to the specification of your customers’ Pods.

ANSWER: C

Explanation:

Create a GKE node pool with a sandbox type configured to gvisor. Add the parameter `runtimeClassName: gvisor` to the specification of your customers’ Pods. GKE Sandbox uses gVisor, a container runtime sandbox designed for workloads that execute untrusted or potentially hostile code. Rather than allowing a container workload to interact as directly with the host kernel as a standard container runtime does, gVisor adds an isolation layer that intercepts and handles many system calls in a user-space kernel. This reduces the host attack surface and strengthens the boundary between tenant workloads and the underlying node.

For this configuration to take effect, the Pods must be scheduled onto a node pool that supports GKE Sandbox, and each applicable Pod must request the gVisor RuntimeClass through `runtimeClassName: gvisor`. This is particularly appropriate for a multi-tenant product where customers are allowed to run arbitrary code, because the primary requirement is stronger runtime isolation rather than image admission control or vulnerability reporting. GKE documents GKE Sandbox as an option for running untrusted workloads with an extra security boundary. See GKE Sandbox documentation and the gVisor documentation.

QUESTION NO: 15

You need to extract text from audio files by using the Speech-to-Text API. The audio files are pushed to a Cloud Storage bucket. You need to implement a fully managed, serverless compute solution that requires authentication and aligns with Google-recommended practices. You want to automate the call to the API by submitting each file to the API as the audio file arrives in the bucket. What should you do?

A. Run a Kubernetes job to scan the bucket regularly for incoming files, and call the Speech-to-Text API for each unprocessed file.

B. Create an App Engine standard environment triggered by Cloud Storage bucket events to submit the file URI to the Google Speech-to-Text API.

C. Run a Python script by using a Linux cron job in Compute Engine to scan the bucket regularly for incoming files, and call the Speech-to-Text API for each unprocessed file.

D. Create a Cloud Function triggered by Cloud Storage bucket events to submit the file URI to the Google Speech-to-Text API.

ANSWER: D

Explanation:

Create a Cloud Function triggered by Cloud Storage bucket events to submit the file URI to the Google Speech-to-Text API. A Cloud Storage event trigger invokes the serverless function whenever an object is finalized in the bucket, so each newly uploaded audio file can be processed immediately without polling or managing infrastructure. The event payload supplies details such as the bucket and object name; the function can construct the `gs://` URI required for asynchronous Speech-to-Text recognition requests. Cloud Functions is fully managed and scales automatically in response to incoming events. Authentication should use the function's attached service account, granting it the least-privilege IAM permissions needed to read the relevant Cloud Storage objects and invoke Speech-to-Text. This avoids embedding service-account keys or implementing custom credential handling. Google documents Cloud Storage triggers for Cloud Run functions at Cloud Storage event triggers. Speech-to-Text documentation describes asynchronous recognition using audio stored in Cloud Storage, including the required Cloud Storage URI, at Asynchronous speech recognition.

QUESTION NO: 16

You are migrating a production-critical on-premises application that requires 96 vCPUs to perform its task. You want to make sure the application runs in a similar environment on GCP. What should you do?

A. When creating the VM, use machine type n1-standard-96.

B. When creating the VM, use Intel Skylake as the CPU platform.

C. Create the VM using Compute Engine default settings. Use gcloud to modify the running instance to have 96 vCPUs.

D. Start the VM using Compute Engine default settings, and adjust as you go based on Rightsizing Recommendations.

ANSWER: A

Explanation:

When creating the VM, use machine type n1-standard-96. is correct because a Compute Engine machine type defines the virtual hardware resources assigned to a VM, including its number of vCPUs and memory. The `n1-standard-96` predefined machine type provides exactly 96 vCPUs and 360 GB of memory, allowing the migrated workload to run with the required processor capacity from its initial deployment. Selecting an appropriately sized machine type before creating the production VM is important because the application's stated requirement is explicit and capacity must be available when the workload starts.

Compute Engine predefined machine types are designed to provide predictable, documented combinations of compute and memory resources. The N1 machine series includes the 96-vCPU standard configuration, making it directly suitable for this requirement. A CPU platform selection can control the underlying processor generation where supported, but it does not itself allocate 96 vCPUs; the selected machine type performs that function. For production migration planning, use a machine type that meets the application's known resource requirement and validate any additional needs, such as memory, disk performance, networking, and regional quota, before cutover. See the Compute Engine N1 machine type documentation and the VM instance creation documentation.

QUESTION NO: 17

You are developing a new application and are looking for a Jenkins installation to build and deploy your source code. You want to automate the installation as quickly and easily as possible. What should you do?

A. Deploy Jenkins through the Google Cloud Marketplace.

B. Create a new Compute Engine instance. Run the Jenkins executable.

C. Create a new Kubernetes Engine cluster. Create a deployment for the Jenkins image.

D. Create an instance template with the Jenkins executable. Create a managed instance group with this template.

ANSWER: A

Explanation:

Deploy Jenkins through the Google Cloud Marketplace. Google Cloud Marketplace provides packaged deployment solutions that are designed to simplify provisioning software on Google Cloud. Selecting a Jenkins offering from the marketplace gives you a guided deployment flow that creates the required infrastructure and configures the application with substantially less manual setup than building a virtual-machine, Kubernetes, or managed-instance-group deployment yourself. This is the most appropriate choice when the stated priority is installing Jenkins as quickly and easily as possible.

Marketplace solutions commonly use deployment templates and expose configuration choices during deployment, allowing a team to select a suitable environment and then launch the solution from a standardized package. Jenkins can then be configured with its required credentials, source repositories, and pipeline definitions after the initial deployment. Google Cloud also documents Jenkins deployment architectures for teams that need more control over Kubernetes configuration, but those approaches entail creating and managing the underlying cluster and Jenkins resources. For a new application with no stated custom infrastructure requirements, the marketplace deployment is the fastest automated starting point. See Deploying VM products from Google Cloud Marketplace and Jenkins on Google Kubernetes Engine.

QUESTION NO: 18

Your application accepts requests that require slow processing by a Cloud Run service. The API must return quickly, and failed processing attempts must be retried automatically. You want to manage the retry behavior without implementing a custom retry system. Which two actions should you take? (Choose two.)

A. Create a Cloud Tasks queue and add a task for each processing request.

B. Configure the queue to deliver HTTP tasks to the Cloud Run service and configure a retry policy.

C. Increase the request timeout of the API until it can wait for all processing to finish.

D. Create a Cloud Storage lifecycle policy to retry failed requests.

E. Run a cron job on a Compute Engine VM that polls the API for unfinished requests.

ANSWER: A B

Explanation:

Create a Cloud Tasks queue and add a task for each processing request. Cloud Tasks provides asynchronous task delivery, allowing the API to enqueue work and return without waiting for the slow processing operation to complete.

Configure the queue to deliver HTTP tasks to the Cloud Run service and configure an appropriate retry policy. Cloud Tasks retries failed task deliveries according to the queue retry settings, while Cloud Run processes each delivered request. See Cloud Tasks overview and Configure queues.

QUESTION NO: 19

Your security team must retain a record of all administrative changes made in a production project for seven years. The records must be stored separately from the project and protected from accidental deletion. Which two actions should you take? (Choose two.)

A. Create a Cloud Logging sink that exports Admin Activity audit logs to a Cloud Storage bucket in a separate project.

B. Configure a seven-year retention policy on the destination Cloud Storage bucket.

C. Disable Data Access audit logs to reduce the volume of audit records.

D. Export the audit logs manually from Logs Explorer at the end of each month.

E. Grant the production application's service account the Storage Admin role on the destination bucket.

ANSWER: A B

Explanation:

Create a Cloud Logging sink that exports the project's Admin Activity audit logs to a dedicated Cloud Storage bucket in a separate project. Admin Activity audit logs record administrative actions, including resource configuration and IAM policy changes. A sink continuously exports matching log entries to the selected destination.

Configure a retention policy on the destination bucket for seven years. A Cloud Storage retention policy prevents objects from being deleted or replaced until the configured retention period has elapsed. Storing the archive in a separate project also provides an administrative boundary between the production workload and the audit-log destination. See Cloud Audit Logs, Export logs with sinks, and Cloud Storage retention policies.

QUESTION NO: 20

A contractor requires access to a single Google Cloud project for a two-week engagement. The contractor needs to view Compute Engine instances but must not retain access after the engagement ends. You want to automate the access expiration by using IAM. What should you do?

A. Grant the contractor the Project Owner role and create a Cloud Scheduler job to remove it after two weeks.

B. Grant the contractor the Compute Viewer role with an IAM condition that has an expiration time.

C. Create a service account key for the contractor and delete the key after two weeks.

D. Add the contractor to the project as a Billing Account Viewer.

ANSWER: B

Explanation:

Grant the contractor the Compute Viewer role with an IAM condition that has an expiration time. IAM Conditions allow an administrator to grant a role only when a specified condition evaluates to true. A time-based condition can limit the role binding so it is effective only until the end of the engagement.

This approach gives the contractor the required read-only access while avoiding a manual task to remove the role at a later date. The role should be granted at the project scope only if project-level instance viewing is required, and the condition should use the appropriate expiration timestamp. See the IAM Conditions overview.

QUESTION NO: 21

Your organization has a dedicated person who creates and manages all service accounts for Google Cloud projects. You need to assign this person the minimum role for projects. What should you do?

A. Add the user to roles/iam.roleAdmin role.

B. Add the user to roles/iam.securityAdmin role.

C. Add the user to roles/iam.serviceAccountUser role.

D. Add the user to roles/iam.serviceAccountAdmin role.

ANSWER: D

Explanation:

Add the user to roles/iam.serviceAccountAdmin role. This predefined IAM role is designed for administrators who need to create, update, delete, and otherwise manage service accounts within a project. It includes the permissions required to administer service account resources, such as creating service accounts and managing their metadata, while limiting the scope to service-account administration rather than granting broad control over unrelated IAM policy or project resources.

Applying the role at the project level allows the designated administrator to manage service accounts for that project, which matches the stated operational responsibility. This follows Google Cloud's least-privilege principle: grant a predefined role containing the permissions necessary for the job, at the narrowest resource level that supports the required work. If the person also needs to manage service-account keys, Google Cloud documents that Service Account Admin includes permissions for service-account and key administration. The role does not by itself grant the ability to impersonate a service account to access resources; that is a separate capability and should be granted only when required.

For the role's included permissions and supported resource scope, see the Google Cloud IAM roles reference. For guidance on applying least privilege, see Using IAM securely.

QUESTION NO: 22

You are building an archival solution for your data warehouse and have selected Cloud Storage to archive your data. Your users need to be able to access this archived data once a quarter for some regulatory requirements. You want to select a cost-efficient option. Which storage option should you use?

A. Coldline Storage

B. Nearline Storage

C. Regional Storage

D. Multi-Regional Storage

ANSWER: A

Explanation:

Coldline Storage is the appropriate cost-efficient Cloud Storage class for archived data that users expect to retrieve about once per quarter. Cloud Storage storage classes are intended for different access-frequency patterns while maintaining the same high durability, availability, low-latency access characteristics, and consistent APIs. Coldline is designed for cold data that is accessed no more than roughly once per quarter, such as backups, disaster-recovery data, and retained business records. Its lower storage price makes it a suitable fit for a data-warehouse archive with quarterly regulatory retrieval requirements.

When selecting a Cloud Storage class, access frequency is important because Coldline has retrieval and early-deletion charges, including a 90-day minimum storage duration. In this scenario, the expected quarterly access pattern aligns directly with the service's intended Coldline use case, so its lower ongoing storage cost provides an efficient archival solution without sacrificing direct online access when regulatory users need the data. Google documents Coldline as appropriate for data read or modified at most once per quarter. See the Cloud Storage storage classes documentation and the Cloud Storage pricing documentation for class characteristics and applicable retrieval and duration charges.

QUESTION NO: 23

Your company stores data from multiple sources that have different data storage requirements. These data include:

1. Customer data that is structured and read with complex queries

2. Historical log data that is large in volume and accessed infrequently

3. Real-time sensor data with high-velocity writes, which needs to be available for analysis but can tolerate some data loss

You need to design the most cost-effective storage solution that fulfills all data storage requirements. What should you do?

A. Use Spanner for all data.

B. Use Cloud SQL for customer data, Cloud Storage (Coldline) for historical logs, and BigQuery for sensor data.

C. Use Cloud SQL for customer data, Cloud Storage (Archive) for historical logs, and Bigtable for sensor data.

D. Use Firestore for customer data, Cloud Storage (Nearline) for historical logs, and Bigtable for sensor data.

ANSWER: C

Explanation:

Use Cloud SQL for customer data, Cloud Storage (Archive) for historical logs, and Bigtable for sensor data. is the most cost-effective design because it selects a managed service aligned with each workload's access pattern and performance needs. Cloud SQL is appropriate for structured customer records that require relational modeling and complex SQL queries, while retaining the operational simplicity of a managed relational database.

Cloud Storage Archive is designed for data accessed less than once per year and has the lowest storage price among Cloud Storage classes. It is therefore well suited to retaining large historical log collections that are rarely retrieved. Retrieval and early-deletion charges are acceptable trade-offs when the primary requirement is economical long-term retention rather than frequent access.

Bigtable is designed for very high-throughput, low-latency workloads and is a strong fit for continuously arriving sensor or time-series data. Its scalable wide-column architecture can ingest high-velocity writes and retain data in a form that remains accessible for downstream analysis. This combination avoids using a premium relational service for bulk time-series ingestion and applies lower-cost archival storage to infrequently accessed logs. See the Cloud Storage class documentation and the Bigtable overview.

QUESTION NO: 24

You have been asked to create robust Virtual Private Network (VPN) connectivity between a new Virtual Private Cloud (VPC) and a remote site. Key requirements include dynamic routing, a shared address space of 10.19.0.1/22, and no overprovisioning of tunnels during a failover event. You want to follow Google-recommended practices to set up a high availability Cloud VPN. What should you do?

A. Use a custom mode VPC network, configure static routes, and use active/passive routing

B. Use an automatic mode VPC network, configure static routes, and use active/active routing

C. Use a custom mode VPC network, use Cloud Router Border Gateway Protocol (BGP) routes, and use active/passive routing

D. Use an automatic mode VPC network, use Cloud Router Border Gateway Protocol (BGP) routes and configure policy-based routing

ANSWER: C

Explanation:

Use a custom mode VPC network, use Cloud Router Border Gateway Protocol (BGP) routes, and use active/passive routing. A custom-mode VPC is appropriate when the organization must explicitly control subnet address ranges, such as allocating the specified 10.19.0.0/22 address block (the usable network CIDR is aligned to the /22 boundary). This avoids relying on the predefined subnet ranges created by an auto-mode network and supports coordinated IP planning with the remote environment.

High-availability Cloud VPN uses two interfaces and two VPN tunnels, and Cloud Router exchanges routes dynamically through BGP. Dynamic routing lets the VPN automatically withdraw or advertise routes as connectivity changes, avoiding manual static-route administration during a tunnel failure. For the stated traffic-engineering requirement, active/passive routing directs traffic through one preferred tunnel while retaining the second tunnel as a failover path. BGP route priorities can establish that preference, and, when the preferred path becomes unavailable, Cloud Router and the peer can converge

on the backup path. Google documents HA VPN's Cloud Router and BGP requirements, as well as supported active/passive topology behavior, in the Cloud VPN overview and Cloud VPN topologies documentation.

QUESTION NO: 25

You just installed the Google Cloud CLI on your new corporate laptop. You need to list the existing instances of your company on Google Cloud. What must you do before you run the gcloud compute instances list command?

Choose 2 answers

A. Run gcloud auth login and complete the browser-based sign-in flow.

B. Create a Google Cloud service account, and download the service account key. Place the key file in a folder on your machine where gcloud CLI can find it.

C. Download your Cloud Identity user account key. Place the key file in a folder on your machine where gcloud CLI can find it.

D. Run gcloud config set compute/zone $my_zone to set the default zone for gcloud CLI.

E. Run gcloud config set project $my_project to set the default project for gcloud CLI.

ANSWER: A E

Explanation:

Run `gcloud auth login` and complete the browser-based sign-in to create local user credentials for the Google Cloud CLI. The CLI then uses those credentials to call Google Cloud APIs as the signed-in user. That user must have the appropriate IAM permission to view Compute Engine instances in the target project, such as permission provided by a suitable Compute Engine viewer role. Google documents the interactive user-credential flow in its Authorize the gcloud CLI guide.

Run `gcloud config set project $my_project` to set the active configuration's default project to the project containing the instances. The `gcloud compute instances list` command uses that configured project when no explicit `--project` flag is supplied. Compute Engine instances are zonal resources, but listing them does not require a default zone because the command can return instances across the project's zones. Configuring the project therefore supplies the required resource scope while avoiding the need to repeat a project flag on each command. The behavior and available flags are documented in the gcloud compute instances list reference.

QUESTION NO: 26

You host your website on Compute Engine. The number of global users visiting your website is rapidly expanding. You need to minimize latency and support user growth in multiple geographical regions. You also want to follow Google-recommended practices and minimize operational costs. Which two actions should you take?

Choose 2 answers

A. Use a Network Load Balancer

B. Deploy your VMs in multiple Google Cloud regions closest to your users' geographical locations.

C. Use an external Application Load Balancer in Global mode.

D. Use an external Application Load Balancer in Regional mode.

E. Deploy all of your VMs in a single Google Cloud region with the largest available CIDR range.

ANSWER: B C

Explanation:

Deploying VMs in multiple Google Cloud regions closest to your users' geographical locations places application capacity nearer to users. This reduces network distance and enables the service to grow independently in the regions where demand is increasing. Compute Engine managed instance groups can provide resilient, autoscaled VM backends in each selected region, helping match capacity to traffic without manually operating individual instances.

Using an external Application Load Balancer in Global mode provides the global entry point needed for this multi-region architecture. A global external Application Load Balancer uses a single anycast IP address, so users connect through Google's global network and are directed to an appropriate healthy backend based on the load-balancing configuration. It supports routing traffic across backends in multiple regions and can steer requests away from unavailable backends. The load balancer is a managed Google Cloud service, which avoids the operational overhead of deploying and maintaining self-managed global traffic-routing infrastructure while allowing the application tier to scale regionally.

Google documents global external Application Load Balancers and their global anycast virtual IP addressing at https://cloud.google.com/load-balancing/docs/https. Guidance for deploying and autoscaling VM workloads with managed instance groups is available at https://cloud.google.com/compute/docs/instance-groups.

QUESTION NO: 27

You built an application on Google Cloud Platform that uses Cloud Spanner. Your support team needs to monitor the environment but should not have access to table data. You need a streamlined solution to grant the correct permissions to your support team, and you want to follow Google-recommended practices. What should you do?

A. Add the support team group to the roles/monitoring.viewer role

B. Add the support team group to the roles/spanner.databaseUser role.

C. Add the support team group to the roles/spanner.databaseReader role.

D. Add the support team group to the roles/stackdriver.accounts.viewer role.

E. Add the support team group to the roles/spanner.viewer role.

ANSWER: E

Explanation:

Add the support team group to the `roles/spanner.viewer` role. This predefined Cloud Spanner IAM role is designed for personnel who need visibility into the Cloud Spanner environment without access to database contents. It grants permissions to view Cloud Spanner resources and their configuration, such as instances, databases, backups, and operations, enabling a support team to inspect the service environment and assist with operational monitoring. Crucially, the role does not include permissions to read rows from tables or otherwise access application data.

Using a Google-managed predefined role is the streamlined, recommended approach because it applies least privilege without requiring the organization to create, test, and maintain a custom role. Assigning the role to a Google group also makes access management simpler: team membership can be managed centrally, and the appropriate Cloud Spanner visibility is inherited by members of that group. Google documents `roles/spanner.viewer` as the Viewer role for Cloud Spanner and lists its permissions in the Cloud Spanner IAM roles reference. See Cloud Spanner predefined IAM roles and Understand IAM roles.

QUESTION NO: 28

Your organization wants to prevent users from creating resources in unapproved regions. The restriction must apply consistently to all existing and future projects in the Production folder. Which two actions should you take? (Choose two.)

A. Define an Organization Policy constraint that restricts allowed resource locations to the approved regions.

B. Apply the Organization Policy to the Production folder.

C. Add labels that identify approved regions to each project in the Production folder.

D. Create a separate billing account for each approved region.

E. Grant the Project Editor role only to users who deploy resources in approved regions.

ANSWER: A B

Explanation:

Define an Organization Policy constraint that restricts allowed resource locations to the approved regions. Organization Policy provides centralized control over supported resource configurations, including restrictions on permitted resource locations.

Apply the policy to the Production folder. Organization Policy constraints applied at a folder are inherited by the projects and resources beneath that folder. This provides a consistent control for both existing and future projects in the folder without configuring each project separately. See Organization Policy overview and Restricting resource locations.

QUESTION NO: 29

You host a static public website in a Cloud Storage bucket. You need to serve the site through a custom domain over HTTPS and provide a globally distributed entry point. Which two actions should you take? (Choose two.)

A. Configure a global external Application Load Balancer with the Cloud Storage bucket as a backend bucket.

B. Configure a Google-managed SSL certificate for the custom domain on the load balancer.

C. Assign an external IP address directly to the Cloud Storage bucket.

D. Deploy a Compute Engine instance in every region to serve the bucket contents.

E. Create a private Cloud DNS zone for the public website domain.

ANSWER: A B

Explanation:

Configure a global external Application Load Balancer with the Cloud Storage bucket as a backend bucket. The load balancer provides a globally distributed external IP address and can route HTTP(S) requests to content in the backend bucket.

Configure a Google-managed SSL certificate for the custom domain on the load balancer. Google-managed certificates automate certificate provisioning and renewal for eligible domain names. After DNS for the domain points to the load balancer, users can access the static site through HTTPS. See External Application Load Balancer with a backend bucket and Google-managed SSL certificates.

QUESTION NO: 30

You run an application on Google Kubernetes Engine. Pods must access a Cloud Storage bucket without using service account key files. Each workload must receive only the permissions required for its own bucket access. Which two actions should you take? (Choose two.)

A. Enable Workload Identity Federation for GKE on the cluster or node pool.

B. Create a dedicated Kubernetes service account and bind it to a Google service account that has the required Cloud Storage role on the bucket.

C. Store a Google service account key in the container image.

D. Grant the default Compute Engine service account the Owner role on the project.

E. Add the Cloud Storage access key to the Pod specification as an environment variable.

ANSWER: A B

Explanation:

Enable Workload Identity Federation for GKE on the cluster or node pool. Workload Identity Federation for GKE lets Kubernetes workloads obtain Google Cloud credentials using a Kubernetes service account rather than a downloaded service account key.

Create a dedicated Kubernetes service account for the workload and bind it to a dedicated Google service account with the required Cloud Storage IAM role on the specific bucket. This provides workload-specific identity and supports least-privilege access. See Use Workload Identity Federation for GKE.

QUESTION NO: 31

You want to deploy a new containerized application into Google Cloud by using a Kubernetes manifest. You want to have full control over the Kubernetes deployment, and at the same time, you want to minimize configuring infrastructure. What should you do?

A. Deploy the application on GKE Autopilot.

B. Deploy the application on GKE Standard.

C. Deploy the application on Cloud Functions.

D. Deploy the application on Cloud Run.

ANSWER: A

Explanation:

Deploy the application on GKE Autopilot. Autopilot is the appropriate GKE operating mode when the team wants to deploy and manage an application through standard Kubernetes resources, such as Deployment, Service, ConfigMap, and Ingress manifests, without taking on routine node infrastructure administration. You retain control over the workload's Kubernetes configuration: container images, replica counts, resource requests, scheduling-related settings permitted by Autopilot, networking objects, and rollout behavior can all be defined declaratively in the manifest and managed through Kubernetes tooling.

At the same time, Google manages the underlying worker-node infrastructure, including provisioning capacity for Pods, node lifecycle operations, and much of the cluster infrastructure configuration. This substantially reduces the operational work required before and during application deployment. Autopilot also applies platform guardrails and resource-management practices that help provide a secure, production-oriented Kubernetes environment while preserving the familiar Kubernetes deployment model. Google describes Autopilot as a mode in which Google manages the infrastructure, enabling users to focus on workloads rather than nodes. See the GKE Autopilot overview and the GKE cluster mode selection guidance.

QUESTION NO: 32

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

A. Create an HTTP load balancer with a backend configuration that references an existing instance group. Set the health check to healthy (HTTP).

B. Create an HTTP load balancer with a backend configuration that references an existing instance group. Define a balancing mode and set the maximum RPS to 10.

C. Create a regional managed instance group. Configure its autohealing policy with an HTTP health check that has a 10-second check interval and an unhealthy threshold of 3.

D. Create a managed instance group. Verify that the autoscaling setting is on.

ANSWER: C

Explanation:

Create a regional managed instance group and configure its autohealing policy with an HTTP health check that has a 10-second check interval and an unhealthy threshold of 3. A regional managed instance group is the Compute Engine group type that manages instances across multiple zones in a region. Its autohealing policy monitors each VM by using the designated health check and recreates a VM when that check reports it as unhealthy.

The requested timing maps directly to health-check settings: a 10-second interval causes probes to occur every 10 seconds, and an unhealthy threshold of 3 requires three consecutive failed probe attempts before the VM is considered unhealthy. The managed instance group then performs the repair action by recreating the affected VM. This configuration provides instance repair independently of traffic distribution and keeps the group at its configured target size. Google recommends using application-based health checks, such as HTTP checks, for autohealing when application responsiveness—not merely VM process status—is the condition that matters.

See Autohealing instances in managed instance groups and Regional managed instance groups.

QUESTION NO: 33

Your company implemented BigQuery as an enterprise data warehouse. Users from multiple business units run queries on this data warehouse. However, you notice that query costs for BigQuery are very high, and you need to control costs. Which two methods should you use? (Choose two.)

A. Split the users from business units to multiple projects.

B. Apply a user- or project-level custom query quota for BigQuery data warehouse.

C. Create separate copies of your BigQuery data warehouse for each business unit.

D. Split your BigQuery data warehouse into multiple data warehouses for each business unit.

E. Change your BigQuery query model from on-demand to flat rate. Apply the appropriate number of slots to each Project.

ANSWER: B E

Explanation:

Applying a user- or project-level custom query quota for BigQuery data warehouse is an effective governance control for on-demand query costs. BigQuery lets administrators set custom daily query-usage quotas, including quotas scoped to a project or a specific user. These limits restrict the number of bytes that can be processed by queries during a day, helping prevent an individual user, workload, or business unit from consuming an unexpected amount of billed query capacity. Quotas can therefore enforce a defined spending boundary while allowing teams to continue using the shared warehouse.

Changing your BigQuery query model from on-demand to flat rate and applying the appropriate number of slots to each Project addresses cost predictability. Under capacity-based pricing, organizations pay for dedicated BigQuery compute capacity rather than paying per byte processed by each query. Allocating that capacity to reservations and projects enables the organization to govern how much compute each business unit can use, while making costs more stable for sustained or high-volume analytical workloads. Google's current terminology is capacity-based pricing rather than flat-rate pricing, but the intended cost-management approach is the same. See BigQuery custom quotas and BigQuery pricing.

QUESTION NO: 34

You need to create and manage service accounts for your workloads running on Google Cloud. You want to follow Google-recommended practices. What should you do?

Choose 2 answers

A. Create as few service accounts as possible.

B. Delete any unused service accounts immediately.

C. Create single-purpose service accounts.

D. Manage service accounts as resources.

E. Use random names for the service accounts.

ANSWER: C D

Explanation:

Creating single-purpose service accounts follows the principle of least privilege and provides clear identity boundaries between workloads. Each workload, component, or application function can receive only the IAM roles it needs, rather than sharing a broadly privileged identity with unrelated workloads. This also improves auditing because Cloud Audit Logs can more clearly associate an action with the specific workload identity that performed it. Google recommends using single-purpose service accounts whenever practical, especially when workloads have different permission requirements.

Managing service accounts as resources means treating them as lifecycle-managed cloud assets: define ownership, grant access through IAM policy, review permissions regularly, and use consistent provisioning and governance practices. A service account is both an identity and a resource that has its own IAM policy, so its administration should be controlled and monitored just like other production resources. This approach supports secure delegation, reliable operations, and clear accountability across teams and environments. Google's guidance on service-account lifecycle and identity management describes these practices in the service account best practices and the service account management documentation.

QUESTION NO: 35

Your company has separate Google Cloud projects for application teams. A central networking team must manage the VPC networks, subnetworks, and firewall rules. Application teams must be able to deploy Compute Engine instances into the centrally managed networks. Which two actions should you take? (Choose two.)

A. Create a Shared VPC host project for the central networking team.

B. Attach the application projects as service projects to the Shared VPC host project.

C. Create a separate VPC network in each application project and configure VPC Network Peering between all projects.

D. Grant the Compute Network Admin role to each application team on the central networking project.

E. Configure Cloud NAT in each application project to share the central VPC network.

ANSWER: A B

Explanation:

Create a Shared VPC host project for the central networking team. The host project contains the shared VPC network resources, including subnetworks, routes, and firewall rules, and centralizes their administration.

Attach the application projects as service projects to the Shared VPC host project. Workloads in service projects can then use the shared subnetworks while the networking team retains control of the network configuration. See Shared VPC overview.

QUESTION NO: 36

You are operating a Google Kubernetes Engine (GKE) cluster for your company where different teams can run non-production workloads. Your Machine Learning (ML) team needs access to Nvidia Tesla P100 GPUs to train their models. You want to minimize effort and cost. What should you do?

A. Ask your ML team to add the "accelerator: gpu" annotation to their pod specification.

B. Recreate all the nodes of the GKE cluster to enable GPUs on all of them.

C. Create your own Kubernetes cluster on top of Compute Engine with nodes that have GPUs. Dedicate this cluster to your ML team.

D. Add a new, GPU-enabled node pool to the GKE cluster. Ask your ML team to add the `cloud.google.com/gke-accelerator: nvidia-tesla-p100` node selector to their pod specification.

ANSWER: D

Explanation:

Add a new, GPU-enabled node pool to the existing GKE cluster and schedule the ML workloads onto it using the `cloud.google.com/gke-accelerator: nvidia-tesla-p100` node selector. This keeps the current shared cluster in place while isolating GPU capacity to a purpose-built node pool. GKE node pools can be created with GPU accelerators, including NVIDIA Tesla P100 GPUs, so only the nodes intended for model training require the more expensive GPU resources. The ML team should also request the required number of GPUs in each container's resource limits, using `nvidia.com/gpu`, so Kubernetes schedules the workload only where sufficient accelerator capacity is available. The accelerator node selector ensures the Pod targets nodes configured for the requested GPU type. This approach minimizes operational effort because GKE continues to manage the cluster and avoids building or maintaining a separate Kubernetes installation. It also controls cost by allowing non-ML workloads to remain on standard nodes and by retaining one GKE cluster rather than creating another cluster solely for GPU use. See GKE: Run GPUs and GKE pricing.

QUESTION NO: 37

You are deploying a production Cloud SQL for PostgreSQL instance. The application requires automatic failover if a zonal failure occurs and the ability to restore the database to a specific time within the backup retention period after an operator error. Which two actions should you take? (Choose two.)

A. Configure the Cloud SQL instance for high availability.

B. Enable automated backups and point-in-time recovery.

C. Use a read replica as the only recovery solution for accidental data deletion.

D. Store the database on a Local SSD attached to a Compute Engine instance.

E. Disable binary logging to reduce storage costs.

ANSWER: A B

Explanation:

Configure the Cloud SQL instance for high availability. A highly available Cloud SQL instance uses a primary instance and a standby instance in different zones within the selected region. Cloud SQL can fail over to the standby instance when a zonal or instance failure occurs.

Enable automated backups and point-in-time recovery. Point-in-time recovery uses automated backups and transaction logs to restore the database to a chosen time within the configured recovery window. This is appropriate for recovering from logical errors such as accidental data modification or deletion. See Cloud SQL high availability and Cloud SQL point-in-time recovery.

QUESTION NO: 38

You are about to deploy a new Enterprise Resource Planning (ERP) system on Google Cloud. The application holds the full database in-memory for fast data access, and you need to configure the most appropriate resources on Google Cloud for this application. What should you do?

A. Provision preemptible Compute Engine instances.

B. Provision Compute Engine instances with GPUs attached.

C. Provision Compute Engine instances with local SSDs attached.

D. Provision Compute Engine instances with M1 machine type.

ANSWER: D

Explanation:

Provision Compute Engine instances with M1 machine type. An ERP workload that keeps its entire database in memory is constrained primarily by available RAM, rather than by CPU throughput or specialized accelerators. M1 machines are memory-optimized Compute Engine machine types, designed to provide a high memory-to-vCPU ratio and large total memory capacities. This lets the application retain a much larger working dataset in RAM, minimizing database reads from persistent storage and supporting the low-latency access pattern required by an in-memory database.

Machine selection should match the workload's dominant resource requirement. For this application, selecting a memory-optimized VM provides the capacity needed for the database buffer and application processes while allowing the required CPU and network resources to scale with the chosen machine size. Google Compute Engine documents memory-optimized machine families, including M1, as intended for workloads such as in-memory databases and large-memory enterprise applications. Review the available memory-optimized machine configurations and regional availability before deployment, then select a size that accommodates the database's memory requirement plus operating-system and application overhead. See Google Cloud memory-optimized machine families and Compute Engine machine resource documentation.

QUESTION NO: 39

All development (dev) teams in your organization are located in the United States. Each dev team has its own Google Cloud project. You want to restrict access so that each dev team can only create cloud resources in the United States (US). What should you do?

A. Create a folder to contain all the dev projects. Create an organization policy to limit resource locations to US locations.

B. Create an organization to contain all the dev projects. Create an Identity and Access Management (IAM) policy to limit the resources in US regions.

C. Create an Identity and Access Management < IAM) policy to restrict the resources locations in the US. Apply the policy to all dev projects.

D. Create an Identity and Access Management (IAM)policy to restrict the resources locations in all dev projects. Apply the policy to all dev roles.

ANSWER: A

Explanation:

Create a folder to contain all the dev projects. Create an organization policy to limit resource locations to the US. This approach uses the Google Cloud Resource Location Restriction organization policy constraint, `gcp.resourceLocations`, which controls the geographic locations where supported resources may be created. The policy can allow the US location group, such as `in:us-locations`, thereby permitting supported US regions and multi-regions while preventing creation in locations outside the United States. Organization Policy is specifically designed to enforce centrally managed constraints on resource configuration and can be attached to an organization, folder, or project. Applying the restriction at the folder level gives every development project in that folder a consistent, inherited location rule, including projects added to the folder later. This provides a scalable governance boundary without requiring individual project-by-project configuration. Administrators should verify that the intended Google Cloud services support the Resource Location Restriction constraint and account for dependent resources that may have location requirements. Google Cloud documents the constraint's behavior, supported value formats, and inheritance model in its Resource Location Restriction documentation. General guidance for creating and applying policies at organization-resource hierarchy levels is available in the Organization Policy overview.

QUESTION NO: 40

You are setting up a Windows VM on Compute Engine and want to make sure you can log in to the VM via RDP. What should you do?

A. After the VM has been created, use your Google Account credentials to log in into the VM.

B. After the VM has been created, use gcloud compute reset-windows-password to retrieve the login credentials for the VM.

C. When creating the VM, add metadata to the instance using 'windows-password' as the key and a password as the value.

D. After the VM has been created, download the JSON private key for the default Compute Engine service account. Use the credentials in the JSON file to log in to the VM.

ANSWER: B

Explanation:

After the VM has been created, use `gcloud compute reset-windows-password` to retrieve the login credentials for the VM. This is the standard Compute Engine method for establishing or recovering local Windows credentials for Remote Desktop Protocol access. The command creates or resets the password for the specified Windows user account and returns a temporary password that can be used with the VM's external IP address in an RDP client. For example, an administrator can run `gcloud compute reset-windows-password VM_NAME --user USERNAME --zone ZONE`, then use the resulting username and password to sign in through RDP.

This approach uses the Google guest environment installed on supported Windows images to securely manage the local account password. The caller must have the required Compute Engine and IAM permissions, and the VM must be reachable for RDP, typically through a firewall rule allowing TCP port 3389 from an appropriately restricted source range. Google documents the RDP connection and password-reset workflow at Connecting to Windows VMs using RDP and the command syntax at gcloud compute reset-windows-password reference.

QUESTION NO: 41

Your compliance team requires that financial reports stored in a Cloud Storage bucket cannot be deleted or replaced for five years. After the retention period ends, administrators must be able to delete the reports. Which two actions should you take? (Choose two.)

A. Configure a bucket retention policy with a retention period of five years.

B. Lock the retention policy after verifying the configured retention period.

C. Enable object versioning and delete all noncurrent object versions every day.

D. Configure a lifecycle rule that deletes objects after five years.

E. Change the bucket storage class to Archive.

ANSWER: A B

Explanation:

Configure a bucket retention policy with a retention period of five years. A retention policy prevents objects from being deleted or replaced until they reach the configured retention period.

Lock the retention policy after confirming that the policy is correct. Locking makes the retention period irreversible and prevents it from being reduced or removed, which is appropriate when the retention requirement must be enforced. Objects can still be deleted after their individual retention periods have expired. See Bucket Lock retention policies.