Associate Cloud Engineer Study Guide

Contents

Assertion Test

1. Instance templates are used to create a group of individual VMs. The instance templates include:
   A: Machine type, book disk image or container image, zone, and labels
   … are all configuration parameters or attributes of a VM and therefore would be included in an instance group configuration that creates those VMs.

2. The command-line command to create a Cloud Storage bucket is:
   B: gsutil mb
   gsutil the command line for accessing and manipulating Cloud Storage from the command-line. mb is the specific command for creating, or making, a bucket.

3. Your company has an object management policy that requires that objects stored in Cloud Storage be migrated from standard storage to nearline storage 90 days after the object is created. The most efficient way to do this is to:
   D: Create a life cycle management configuration policy specifying an age of 90 days and SetStorageClass as nearline.
   The life cycle configuration policy allows you to specify criteria for migrating data to other storage systems without having to concern yourself with running jobs to actually execute the necessary steps. The other options are inefficient or do not exist.

4. An education client maintains a site where users can upload videos, and your client needs to assure redundancy for the files; therefore, you have created two buckets for Cloud Storage. Which command do you use to synchronize the contents of the two buckets?
   A: gsutil rsync
   gsutil is the command-line tool for working with Cloud Storage. rsync is the specific command in gsutil for synchronizing buckets.

5. VPC resources are which of the following?
   C: Global
   Google operates a global network, and VPCs are resources that can span that global network.

6. A remote component in your network has failed, which results in a transient network error. When you submit a gsutil command, it fails because of a transient error. By default, the command will:
   B: Retry using a truncated binary exponential backoff strategy
   gcloud by default will retry a failed network operation and will wait a long time before each retry. The time to wait is calculated using a truncated binary exponential backoff strategy.

7. All of the following are components of firewall rules except which one?
   C: Time to live (TTL)
   Firewall rules do not have TTL parameters. Direction of traffic, action on match, and protocol are all components of firewall rules.

8. Adding VMs to an instance group can be triggered in an autoscaling policy by all of the following, except which one?
   C: IAM policy violation
   IAM policy violations do not trigger changes in the size of clusters. All other options can be used to trigger a change in cluster size.

9. Your company's finance department is developing a new account management application that requires transactions and the ability to perform relational database operations using fully compliant SQL. Data store options in Google Cloud include:
   A: Spanner and Cloud SQL
   Only Spanner and Cloud SQL databases support transactions and have a SQL interface. Firestore has transactions but does not support fully compliant SQL; it has a SQL-like query language. Cloud Storage does not support transactions or SQL.

10. The marketing department in your company wants to deploy a web application but does not want to have a manage servers or clusters. A good option for them is:
    C: Cloud Run
    … is a serverless service for running containers and allows developers to deploy full applications without having to manage servers or clusters. Compute Engine and Kubernetes Engine require management of servers. Cloud Functions is suitable for short-running Node.js or Python functions but not full applications.

11. Your company is building an enterprise data warehouse and wants SQL query capabilities over petabytes of data, but does not want to manage servers or clusters. A good option for this is:
    B: BigQuery
    … is designed for petabytes-scale analytics and provides a SQL interface.

12. You have been hired as a consultant to a startup in the Internet of Things (IoT) space. The startup will stream large volumes of data into Google Cloud. The data needs to be filtered, transformed, and analyzed before being stored into Google Cloud Firestore. A good option for the stream processing component is:
    B: Cloud Dataflow
    … allows for stream and batch processing of data and is well suited for this kind of ETL work. Dataproc is a managed Hoop and Spark service that is used for big data analytics. Cloud Endpoints is an API service, and Cloud Interconnection is a network service.

13. Preemptible VMs may be shut down at any time but will always be shut down after running:
    C: 24 hours
    If a preemptible machine has not been shut down within 24 hours, Google will stop the instance.

14. You have been tasked with designing an organizational hierarchy for managing departments and their cloud resources. What organizing components are available in Google Cloud?
    A: Organization, folders, projects
    Organizations, folders, and projects are the components used to manage an organizational hierarchy. Buckets, directories, and subdirectories are used to organize storage.

15. During an incident that has caused an application to fail, you suspect some resource may not have appropriate roles granted. The command to list roles granted to a resource is:
    B: gcloud iam list-grantable-roles
    gcloud is the command-line tool for working with IAM, the list-grantable-roles is the correct command (following).

16. The availability of CPU platforms can vary between zones. To get a list of all CPU types available in a particular zone, you should use:
    A: gcloud compute zones describe
    gcloud is the command-line tool for manipulating compute resources, and zones describe is the correct command (following).

17. To create a custom role, a user must posses which role?
    C: iam.roles.create
    … is correct; the other roles do not exist.

18. You have been asked to create a network with 1.000 IP addresses. In the interest of minimizing unused IP addresses, which CIDR suffix would you use to create a network with at least 1.000 addresses but no more than necessary?
    B: /22
    The /22 suffix produces 1.022 usable IP addresses.

19. A team of data scientists have asked for your help setting up an Apache Spark cluster. You suggest they use a managed Google Cloud service instead of managing a cluster themselves on Compute Engine. The service they would use is:
    A: Dataproc
    … is the managed Spark service. Cloud Dataflow is for stream and batch processing of data, BigQuery is for analytics, and Cloud Hadoop is not a Google Cloud service (at all).

20. You have created a web application that allows users to upload files to Cloud Storage. When files are uploaded, you want to check the file size and update the user's total storage used in their account. A serverless option for performing this action on load is:
    D: Cloud Functions
    … respond to events in Cloud Storage, making them a good choice for taking an action after a file is loaded.

21. Your company has just started using Google Cloud, and executives want to have a dedicated connection from your data center to the Google Cloud to allow for large data transfers. Which networking service would you recommend?
    B: Google Cloud Dedicated Interconnect
    … is the only option for a dedicated connection between a customer's data center and a Google data center.

22. You want to have Google Cloud manage cryptographic keys, so you're decided to use Cloud Ke Management Services. Before you start creating cryptographic key, you must:
    A: Enable Google Cloud Key Management Service (KMS) API and set up billing.
    Enabling the Google Cloud KMS API and setting up billing are steps common to using Google Cloud services.

23. In Kubernetes Engine, a node pool is:
    B: A subnet of node instances within a cluster that all have the same configuration.

24. The Google Cloud service for storing and managing Docker containers is:
    C: Container Registry
    The Google Cloud service for storing and managing Docker containers is Artifact Registry. Cloud Build is for creating images. Cloud Source Repositories are private Git repositories hosted on Google Cloud Docker Repository is not a Google Cloud service.

25. Code for Cloud Functions can be written in several languages, including:
    B: Node.js, Python, and Go
    … are three of the languages supported by Cloud Functions.

Chapter 1: Overview of Google Cloud

Google Cloud offers a variety of services for compute, storage, networking, and specialized services. Compute services include virtual machines (VMs) and Kubernetes clusters, while storage services support object and file storage along with caching. Networking services provide virtual private clouds (VPCs) and other services including VPNs, Interconnects, Shared VPC, VPC networking peering, and Direct or Carrier peering. Specialized services include machine learning, Speech-to-Text, and recommendation services. Cloud computing has several advantages over on-premises computing including: renting rather than owning infrastructure, a pay-as-you-go model, elastic resource allocation, and specialized services.

Understand different ways of delivering cloud computing resources. Computing resources can be allocated as individual VMs or clusters of VMs that you manage. You can also use managed Kubernetes clusters that relieve you of some of the operational overload of managing a Kubernetes cluster. Serverless computing options relieve users of any server management. Instead, developers run their code in a containerized environment managed by the cloud provider or in a computer platform designed for short-running code. Developers and DevOps professionals have the most control over resources when they manage their own servers and clusters. Managed services and serverless options are global choices when you do not need control over the computing environment and will get more value from not having to manage compute resources.

Understand the different forms of cloud storage and when to use them. There are four main categories of storage: object, file, block, and in-memory caches. Object storage is designed for highly reliable and durable storage of objects, such as images or data sets. Object storage has more limited functionality than filesystem-based storage systems. File system-based storage provides hierarchical directory storage for files and supports common operating system and filesystem functions. Filesystem services provide network-accessible filesystems that can be accessed by multiple servers. Block storage is used with persistent storage devices, such as SSDs and HDDs. Caches are in-memory data stores used to minimize the latency of retrieving data. They do not provide persistent storage and should never be considered a “system of truth”.

Understand the difference between running an IT environment on-premises or in the cloud. Running an IT environment in the cloud has several advantages, including short-term rental of resources, a pay-as-you-go model, elastic resource allocation, and the ability to use specialized services. The unit cost of cloud resources, such as the cost per minute of a mid-tier server, may be higher in the cloud than on-premises. It is important to understand the cost model of your cloud provider so that you can make decisions about the most efficient distribution of workload between cloud and on-premises resources.

1. Which of the following is an option for choosing a computing resource in Google Cloud?
   B: virtual machine (VM)
   The correct answer is B. A basic unit for purchasing computing resources in Google Cloud is the VM. Option A is incorrect; a cache is a low-latency storage system. Option C is incorrect; a block is a unit of storage on persistent disks. Option D is incorrect; a subnet is a networking abstraction.

2. If you use a cluster that is managed by a cloud provider, which of these will be managed for you by the cloud provider?
   D: All of the above (Monitoring, Networking, and some security management tasks)
   When using managed clusters, the cloud provider will monitor the health of servers, also known as nodes, in the cluster; set up networking between nodes in the cluster, and configure firewall and other security controls.

3. You need serverless computing for file processing, and running the back end of a website; which two products can you choose from Google Cloud?
   B: Cloud Run and Cloud Functions
   Cloud Run is a serverless platform for running containers, and Cloud Functions is a service for executing short-running functions in response to events. Kubernetes Engine is a managed cluster service, and both Kubernetes Engine and Compute Engine require you to configure servers. Neither Compute Engine nor Kubernetes is a serverless option.

4. You have been asked to design a storage system for a web application that allows users to upload large data files to the analyzed by a data analytics workflow. The files should be stored in a high-availability storage system. Filesystem functionality is not required. Which storage system in Google Cloud should be used?
   B: Object storage
   Object storage, like Cloud Storage, provides redundantly stored objects without limits on the amount of data you can store, which makes option B correct. Since filesystem functionality is not required, option D is not a good option. Block storage could be used, but you would have to manage your own replication to ensure high availability and it would cost more than object storage. Caches are transient, in-memory storage and are not high-availability, persistent storage systems.

5. All block storage systems use what block size?
   D: Block size can vary
   Block sizes in a block storage system can vary. Block size is established when a filesystem is created. In Linux, 4 KB block sizes are commonly used.

6. You have been asked to set up network security in a VPC. Your company wants to have multiple subnetworks and limit traffic between the subnetworks. Which network security control would you use to control the flow of traffic between subnets?
   C: Firewall
   Firewalls in Google Cloud are software-defined network controls that limit the flow of traffic into and out of a network or subnetwork. Routers are used to move traffic to appropriate destinations on the network. Identity access management is used for authenticating and authorizing users; it is not relevant to network controls between subnetworks. IP address tables are not a security control.

7. When you create a machine learning service to learn how to classify object using tabular data, what type of servers should you choose to manage compute resources?
   C: No servers; you should use specialized services, which are serverless
   Option C is correct because specialized services in Google Cloud, like AutoML are serverless. Google manages the compute resources used by the services. There is no need for a user to allocate or manage servers.

8. When does investing in servers for extended periods of time, such as committing to use servers for three to five years, work well?
   B: When a company can accurately predict server need for an extended period of time
   Investing in servers works well when an organization can accurately predict the number of servers and other equipment it will need for an extended period and can utilize that equipment consistently. Startups are not established businesses with histories that can guide expected needs in three to five years. It does not matter if a budget is fixed or variable; investing in servers should be based on demand for server capacity.

9. Your company is based in North America and will be running a virtual server for batch processing invoices. What factor determines the unit per minute cost?
   B: The characteristics of the server
   The characteristics of the server, such as the number of virtual servers, the amount of memory, and the region where you run the VM, influence the cost, so option B is correct. Time of day is not a factor, nor is the type of application you run on a VM.
10. You plan to use AutoML to analyze sales data and predict product demand in the near future. You plan to analyze between 1.000 and 2.500 products per hour. How many VMs should you allocate to meet peak demand?
    D: None; AutoML is a serverless service
    AutoML is one of Google Cloud's specialized services. Users of the service do not need to configure any VMs to use the service.

11. You have to run a number of services to support an application. Which of the following is a good deployment model?
    B: Use containers in a managed cluster
    Containers give you the most flexibility for using resources of a cluster efficiently, and orchestration platforms reduce the operations overhead, which makes option B correct. Running in a single VM is not recommended because if the server fails, all services will be down. Using two VMs with one read-only is not useful. Read-only servers are sometimes used with databases, but there was no mention of databases in the question. Using a small VM and upgrading when it is no longer able to keep up with the workload delivers poor-quality service to users and should be avoided.

12. You have created a VM. Which of the following system administration operations are you allowed to perform on it?
    D: All of the above (Configure the filesystem, patch operating system software, and change file and directory permissions)
    All the operations are available to a system administrator after creating a VM.

13. Cloud Filestore is based on what filesystem technology?
    A: Network File System (NFS)
    Cloud Filestore is based on the NFS, which is a distributed file management system. The other options are filesystems supported by Linux.

14. When creating resources in Google Cloud, those resources are always part of what?
    A: Virtual private cloud
    When you create resources, they are created within a VPC. Resources are added to the VPC and are not accessible outside the VPC unless you explicitly configure them to be. A subdomain is related to web domains and not related to the organization of Google Cloud resources. Clusters, such as Kubernetes clusters, may be in your network, but not all resources are necessarily in a cluster.

15. You need to store data for an application and are using a cache. How will the cache affect data retrieval?
    D: Using a cache will reduce latency, since retrieving from a cache is faster than retrieving from SSDs or HDDs.
    Caches use memory, and that makes them the fastest storage type for reading data. Caches are data stores on the back end of distributed systems, not the clients. A cache would have no effect on client-side JavaScript execution. Caches can lose data in the cache if power is lost, and the data would have to be reloaded. Caches can get out of sync with the system of truth because the system of truth could be updated, but the cache may not be updated. Caches have faster read times than SSDs and HDDs.

16. Why can cloud providers offer elastic resource allocation?
    B: Extensive resources and the ability to quickly shift resources between costumers enables public cloud providers to offer elastic allocation more efficiently than can be done in smaller data centers.
    Cloud providers have large capacity and can quickly allocate those resources to different customers. With a mix of customers and workloads, they can optimize the allocation of resources. Option A is incorrect; cloud providers do not take resources from one customer to give them to another, with the exception of preemptible instances. Option C is incorrect; cloud providers usually offer discounts for increased use.

17. What is not a characteristic of specialized services in Google Cloud?
    C: They require monitoring by the user.
    Specialized services are monitored by Google so that users do not have to monitor them. Specialized services provide a specific compute functionality but do not require the user to configure any resources. They also provide APIs.

18. You client's transactions must access a drive attached to a VM that allows for random access to part of files. What kind of storage does the attached drive provide?
    B: Block storage
    Attaches drivers are block storage devices. Cloud Storage is an object storage service and does not attach directly to a VM. NoSQL is a type of database, not a storage system. There is no such thing as SQL storage; SQL is a query language used in relational databases.

19. You are deploying a new relational database to support a web application. Which type of storage system would you use to store data files of the database?
    C: Block storage
    Databases require persistent storage on block devices. Object storage does not provide data block or filesystem storage. Data storage is not a type of storage system. Caches are often used with databases to improve read performance, but they are volatile and are not suitable for persistently storing data files.

20. A user prefers services that require minimal setup; why would you recommend Cloud Storage, Cloud Run, and Cloud Functions?
    B: They are serverless
    All three services are serverless, so the user does not need to configure VMs. Cloud Storage is charged based on time and size of data stored. Cloud Run and Cloud Functions are not restricted to just the Go language.

Chapter 2: Google Cloud Computing Services

Google Cloud provides a full range of services to support information processing including compute resources, storage resources, databases, networking services, identity management and security services, development tools, management and operations services, as well as specialized services to support AI.

Understand the difference between Compute Engine, Kubernetes Engine, App Engine, Cloud Run, and Cloud Functions. Compute Engine is Google's VM service. Users can choose CPUs, memory, persistent disks, and operating systems. They can further customize a VM by adding graphics processing utils for compute-intensive operations. VMs are managed individually or in groups of similar servers.

Kubernetes Engine manages groups of virtual servers and applications that run in containers. Containers are lighter weight than VMs. Kubernetes is called an *orchestration service* because it distributes containers across clusters, monitors cluster health, and scales as proscribed by configurations.

App Engine is Google's PaaS. Developers can run their code in a language-specific sandbox when using the standard environment or in a container when using the flexible environment. App Engine is a serverless service, so customers do not need to specify VM configurations or manage servers.

Cloud Run is a service for running stateless containers. This is a serverless option that provides some of the advantages of Kubernetes without requiring you to deploy your own clusters. Note that Cloud Run does not currently support applications that maintain state in the container.

Cloud Functions is a serverless service that is designed to execute short-running code that responds to events, such as file uploads or messages being published to a message queue. Functions may be written in Node.js or Python.

Understand what is meant by serverless. Serverless means customers using a service do not need to configure, monitor, or maintain the computing resources underlying the service. It does not mean there are no servers involved – there are always physical servers that run applications, functions, and other software. Serverless only refers to not needing to manage those underlying resources.

Understand the difference between object and file storage. Object stores are used to store and access file-based resources. These objects are referenced by a unique identifier, such as a URL. Object stores do not provide block or filesystem services, so they are not suitable for database storage. Cloud Storage is Google Cloud's object storage service.

File storage supports block-based access to files. Files are organized into directories and subdirectories. Google Filestore is based on NFS.

Know the different kind of databases. Databases are broadly divided into relational and NoSQL databases.

Relational databases support transactions, strong consistency, and the SQL query languages. Relational databases have been traditionally difficult to horizontally scale. Cloud Spanner is a global relational database that provides the advantages of relational databases with the scalability previously found only in NoSQL databases.

NoSQL databases are designed to be horizontally scalable. Other features, such as strong consistency and support for standard SQL, are often sacrificed to achieve scalability and low-latency query responses. NoSQL databases may be key-value stores like Cloud Memorystore, document databases like Cloud Firestore, or wide-column databases such as Cloud Bigtable.

Understand virtual private clouds. A VPC is a logical isolation of an organization's cloud resources within a public cloud. In Google Cloud, VPCs are global; they are not restricted to a single zone or region. All traffic between Google Cloud services can be transmitted over the Google network without the need to send traffic over the public Internet.

Understand load balancing. Load balancing is the process of distributing a workload across a group of servers. Load balancers can route workload based on network-level or application-level rules. Google Cloud load balancers can distribute workloads globally.

Understand developer and management tools. Developer tools support common workflows in software engineering, including using version control for software, building containers to run applications and services, and making containers available to other developers and orchestration systems, such as Kubernetes Engine.

Management tools, such as Cloud Monitoring and Cloud Logging, are designed to provide systems administration information to developers and operators who are responsible for ensuring applications are available and operating as expected.

Know the types of specialized services offered by Google Cloud. Google Cloud includes a growing list of specialized services for data analytics as well as AI and machine learning.

Know the differences between on-premises and public cloud computing. On-premises computing is computing, storage, networking, and related services that occur on infrastructure managed by a company or organization for its own use. Hardware may be located literally on the premises in a computer building or in a third-party colocation facility. Colocation facilities provide power, cooling, and physical security, but the customers of the colocation facility are responsible for all the setup and management of the infrastructure.

Public cloud computing uses infrastructure and services provided by the cloud provider such as Google, AWS, or Microsoft. The cloud provider maintains all physical hardware and facilities. It provides a mix of services, such as VMs that are configured and maintained by customers and serverless offerings that enable customers to focus on application development, while the cloud provider takes on more responsibility for maintaining the underlying compute infrastructure.

1. You are planning to deploy an SaaS application for customers in North America, Europe, and Asia. To maintain scalability, you will need to distribute workload across servers in multiple regions. Which Google Cloud service would you use to implement the workload distributions?

   C: Cloud Load Balancing

   … distributes workloads within and across regions, provides health checks, and implements autoscaling. Cloud DNS provides domain name service, such as translating a URL like www.example.com to an IP address. Cloud CDN distributes content across regions to reduce latency when delivering content to users across the globe.

2. You have decided to deploy a set of microservices using containers. The microservices will maintain state in the container. You could install and manage Docker on Compute Engine instances, but you would rather have Google Cloud provide some container management services. Which are two Google Cloud services that allow you to run containers in a managed service?

   C: Kubernetes Engine and Cloud Run environment

   Cloud Run allows you to run containers in a serverless service. Kubernetes Engine is an orchestration platform for running containers. Both provide container management services and support stateful applications. Cloud Run allows for running containers in a managed service but does not currently support managing state within the container. The App Engine standard environment runs applications in language-specific sandboxes and is not a general container management system. Cloud Functions is a serverless service for running code in response to events.

3. Why would an API developer want to use the Apigee API platform?

   D: A and B (To get the benefits of routing and rate-limiting and authentication services)

   The Apigee API platform provides policy-based rate-limiting and routing services to help accommodate spikes in traffic. It also provides OAuth 2.0 and SAML authentication. It does not provide version control; Cloud Source Repositories is the service used for version control.

4. You are developing an API to the public Internet and are concerned that your service will be subject to DDoS attacks. Which Google Cloud service should you consider to protect your API?

   A: Cloud Armor

   … builds on Google Cloud's load balancing services to provide the ability to allow or restrict access based on IP address, deploy rules to counter cross-site scripting attacks, and provide countermeasures to SQL injection attacks. Cloud CDN is a content distribution service, not a security service. Identity and access management is a security service, but it is for authorization, not denial-of-service mitigation. VPSs are used to restrict network access to an organization's resources, but it does not have features to mitigate denial-of-service attacks. Also, Cloud CDN acts as a first line of defense in the case of DDoS attacks

5. You have an application that uses a Pub/Sub message queue to maintain a list of tasks that are to be processed by another application. The application that consumes messages from the Pub/Sub queue removes the message only after completing the task. It takes approximately 10 seconds to complete a task. It is not a problem if two or more VMs perform the same task. What is a cost-effective configuration for processing this workload?
   A: Use preemptible VMs
   This is a good use case for preemptible VMs because they could reduce the cost of running the second application without the risk of losing work. Since tasks are delete from the queue only after they are completed, if a preemptible VM is shut down before completing the task another VM can perform the task. Also, there is no harm in running a task more than once, so if two VMs do the same task, it will not adversely affect the output of the application. DataProc is a managed Hadoop and Spark cluster, and Spanner is a globally scalable relational database; neither are appropriate products for this task.

6. Your department is deploying an application that has a database back end. You are concerned the read load of the database server and want to have data available in memory to reduce the time to respond to queries and to reduce the load on the database server. Which Google Cloud service would you use to keep data in memory?
   B: Cloud Memorystore
   … is the Google Cloud managed service for caching data in memory using either Redis or memcached. Cloud SQL is a relational database service and might be a good option for the back-end database. Cloud Spanner is a global relational database and is a good option when you need a globally scalable, relational database. Cloud Firestore is a document database suitable for product catalogs, user profiles, and other semi-structured data.

7. The Cloud SDK can be used to configure and manage resources in which of the following services?
   D: All of the above (Compute Engine, Cloud Storage, and network firewalls)
   All three of the services listed, Compute Engine, Cloud Storage, and network firewalls, can be managed and configured using Cloud SDK.

8. What server configuration is required to use Cloud Functions?
   D: None
   Cloud Functions is a serverless product, so no configuration is required.

9. You have been assigned the task of consolidating log data generated by each instance of an application. Which management and observability tools would you use?
   D: Cloud Logging
   The Cloud Logging service is used to consolidate and manage logs generated by applications and servers.

10. Which specialized services are most likely to be used to build a data warehousing platform that requires complex extraction, transformation, and loading operations on batch data as well as processing streaming data?

    B: Data analytics

    The data analytics set of specialized services includes products that help with extraction, transformation, and loading (ETL) and work with both batch and streaming data. The Apigee API platform is used for managing APIs and does not meet the needs described. AI and machine learning might be useful for analyzing data in the data warehouse, but the services in that set are not always helpful for ETL operations. Cloud SDK is used to control services but by itself is not a directly able to perform the operations needed.

11. Your company has deployed 100.000 Internet of Things (IoT) sensors to collect data on the state of equipment in several factories. Each sensor will collect and send data to a data store every 5 seconds. Sensors will run continuously. Daily reports will produce data on the maximum, minimum, and average values for each metric collected on each sensor. There is no need to support transactions in this application. Which database product would you recommend?

    B: Cloud Bigtable

    … is designed to accept billions of rows of data. Spanner is a relational database and supports transactions, but they are not needed. Cloud SQL MySQL and Cloud SQL PostgreSQL would be difficult to scale to this level of read and write performance.

12. You are the lead developer on a medical application that uses patients' smartphones to capture biometric data. The app is required to collect data and store it on the smartphone when data cannot be reliably transmitted to the back-end application. You want to minimize the amount of development you have to do to keep data synchronized between smartphones and back-end data stores. Which data store option should you recommend?

    A: Cloud Firestore

    … is a database service that can synchronize data between mobile devices and centralized storage. Spanner is a global relational database for large-scale applications that require transaction support in highly scaled databases. Cloud CDN is a distributed storage system for reducing latency when delivering static content to web application users. Cloud SQL could be used but would require more custom development to synchronize data between mobile devices and the centralized data store.

13. A software engineer comes to you for a recommendation. They have implemented a machine learning algorithm to identify cancerous cells ins medical images. The algorithm is computationally intensive, makes many floating-point calculations, requires immediate access to large amounts of data, and cannot be easily distributed over multiple servers. What kind of Compute Engine configuration would you recommend?

    B: High memory, high CPU, and high GPU

    A computationally intensive application obviously requires high CPUs, but the fact that there are many floating-point calculations indicates that a GPU should be used. You might consider running this in a cluster, but the work is not easily distributed over multiple servers, so you will need to have a single server capable of handling the load. Immediate access to large amounts of data indicates that a high-memory machine should be recommended.

14. You are tasked with mapping the authentication and authorization policies of your on-premises applications to Google Cloud's authentication and authorization mechanisms. The Google Cloud documentation states that an identity must be authenticated in order to grant permissions to that identity. What does the term identity refer to?
    B: User
    Identities are abstractions of users. They can also represent characteristics of processes that run on behalf of a human user or a VM in the Google Cloud; these are known as service accounts. Identities are not related to VM IDs. Roles are collections of privileges that can be granted to identities.

15. A client is developing an application that will need to analyze large volumes of text information. The client is not expert in text mining or working with language. What Google Cloud service would you recommend they use?
    C: Natural language
    … services provide functionality for analyzing text. Vertex AI is a unique platform for building machine learning models, but since the client is not an expert in machine learning, a specialized service such as Natural Language is a better option. Recommendation AI is used to make product recommendations to customers. Text-to-Speech is a service for converting natural language text to human-sounding speech.

16. Data scientist in your company want to use a machine learning library available only in Apache Spark. They want to minimize the amount of administration and DevOps work. How would you recommend they proceed?
    B: Use Cloud Dataproc

17. Database designers at your company are debating the best way to move a database to Google Cloud. The database supports an application with a global user base. Users expect support for transactions and the ability to query data using commonly used query tools. The database designers decide that any database service they choose will need to support ANSI SQL 2011 and global transactions. Which database service would you recommend?
    B: Cloud Spanner
    … supports ANSI SQL 2011 and global transactions. Cloud SQL supports standard SQL but does not have global transactions. Firestore and Bigtable are NoSQL databases.

18. Which specialized service supports both batch and stream processing workflows?
    A: Cloud Dataproc
    … is designed to execute workflows in both batch and streaming modes, which makes option A correct. BigQuery is a data warehouse service. Firestore is a document database. AutoML is a machine learning service.

19. You have a Python application you would like to run in a scalable environment with the least amount of management overhead. Which Google Cloud product would you select?
    C: App Engine standard environment
    … provides a serverless Python sandbox that scales automatically. App Engine flexible environment runs containers and requires more configuration. Cloud Engine and Kubernetes Engine both require significant management and monitoring.

20. A product manager at you company reports that customers are complaining about the reliability of one of your applications. The application is crashing periodically, but developers have not found a common pattern that triggers the crashes. They are concerned that they do not have good insight into the behavior of the application and want to perform a detailed review of all crash data. Which observability tool would you use to view consolidated crash information?

    D: Error Reporting

    … consolidates crash information. Cloud Monitoring collects metrics on application and server performance. Logging is a log management service. Cloud Dataproc is not an observability tool but is a managed Hadoop and Spark service.

Chapter 3: Projects, Service Accounts, and Billing

The central abstraction for managing Google Cloud resources is the resource hierarchy. It consists of three levels: organization, folder, and project. The Organization Policy Service and IAM together control access to an organization's resources. Billing accounts store information about how to pay charges for resources used. A billing accounts is associated with one or more projects. Google Cloud uses APIs to make services programmatically accessible and most are not enabled by default.

Understand the Google Cloud resource hierarchy. All resources are organized within your resource hierarchy. You can define the resource hierarchy using one organization and multiple folders and projects. Folders are useful for grouping departments, and other groups manage their projects separately. Projects contain resources such as VMs and cloud storage buckets. Projects must have billing accounts associated with them to use services that are not free.

Understand organization policies. Organization policies restrict resources in the resource hierarchy. Policies include constraints, which are rules that define what can or cannot be done with a resource. For example, a constraint can be set to block access to the serial port on all VMs in a project. Also, understand the policy evaluation process and how to override inherited policies.

Understand service accounts and how they are used. Service accounts are identities that are not associated with a specific user but can be assigned to a resource, like a VM. Resources that are assigned a service account can perform operations that the service account has permission to perform.

Understand Google Cloud Billing. Billing must be enabled to use services and resources beyond free services. Billing associates a billing method, such as a credit card or invoicing information, with a project. All costs associated with resources in a project are billed to the project's billing account. A billing account can be associated with more than one project. You manage your billing through the Billing API.

1. You are designing cloud applications for a healthcare provider. The records management application will manage medical information for patients. Access to this data is limited to a small number of employees. The billing department application will have insurance and payment information. Another group of employees will have access to billing information. In addition, the billing system will have two components: a private insurance billing system and a government payer billing system. Government regulations require that software used to bill the government must be isolated from other software systems. Which of the following resource hierarchies would meet these requirements and provide the most flexibility to adapt to changing requirements?
   A: One organization, with folders for records management and billing. The billing folder would have private insurer and government payer folders within it. Common constrains would be specified in organization-level policies. Other policies would be defined at the appropriate folder.
   Option A, the correct answer, separates the two main applications into their own folders and further separating private insurance from government payer by using folders for each. This satisfies the regulatory need to keep the government payer software isolated from other software.

2. When you create a hierarchy, you can have more than one of which structures?
   C: Folder and project
   Resource hierarchies have a single organization at the root, which makes option C correct. Below that, there are folders that can contain other folders or projects. Folders can contain multiple folders and multiple projects.

3. You are designing an application that uses a series of services to transform data from its original form into a format suitable for use in a data warehouse. Your transformation application will write to the message queue as it processes each input file. You don't want to give users permission to write to the message queue. You could allow the application to write to the message queue by using which of the following?
   B: Service accounts
   … are designed to give applications or VMs permission to perform tasks. Billing accounts are for associating charges with a payment method. Folders are part of resource hierarchy and have nothing to do with enabling an application to perform a task. Messaging accounts are a fictitious option.
4. Your company has several policies that need to be enforces for all projects. You decide to apply policies to the resource hierarchy. Not long after you apply the policies, an engineer finds that an application that had worked prior to implementing policies is no longer working. The engineer would like you to create an exception for the application. How can you override a policy inherited from another entity in the resource hierarchy?
   A: Inherited policies can be overridden by defining a policy at a folder or project level
   Service accounts and billing accounts are not part of the resource hierarchy and are not involved in overriding policies.

5. Constraints are used in resource hierarchy policies. Which of the following are types of constraints allowed?
   E: All of the above (Allow a specific set of values, deny a specific set of values, deny a value and all its child values, and allow all allowed values)

6. A team with four members wants you to set up a project that needs only general permissions for all resources. You are granting each person a basic role for different levels of access, depending on their responsibilities in the project. Which of the following are not included as basic roles in Google Cloud?
   B: Publisher
   Owner, Editor, and Viewer are the three basic roles in Google Cloud.

7. You are deploying a new custom application and want to delegate some administration tasks to DevOps engineers. They do not need all the privileges of a full application administrator, but they do need a subset of those privileges. What kind of role should you use to grant those privileges?
   D: Custom
   Basic roles only include Owner, Editor, and Viewer permissions. Predefined roles are designed for Google Cloud products and services, like App Engine and BigQuery. For a custom application, you can create sets of privileges that give the user with that role as much permissions as needed but not more.

8. An app for a finance company needs access to a database and a Cloud Storage bucket. There is no predefined role that grants all the needed permissions without granting some permissions that are not needed. You decide to create a custom role. When defining custom roles, you should follow which of the following principles?
   D: Least privilege
   Users should have only the privileges that are needed to carry out their duties. This is the principle of least privilege. Rotation of duties is another security principle related to having different people perform a task at different times. Defense in depth is the practice of using multiple security controls to protect the same asset.

9. How many organizations can you create in a resource hierarchy?
   A: 1

10. You are contracted by the finance department of your company for advice on how to automate payments for Google Cloud services. What kind of account would you recommend setting up?
    B: Billing account
    In option B, the correct answer, the billing account is used to specify payment information and should be used to set up automatic payments. Service accounts are used to grant privileges to a VM and are not related to billing and payments. Resource accounts and edit accounts to not exist.

11. You are experimenting with Google Cloud for your company. You do not have permissions to incur costs. How can you experiment with Google Cloud without incurring charges?
    C: You can use only free services in Google Cloud
    Google Cloud offers a free service level for many products, which makes option C the correct answer. You can use these services without having to set up a billing account. Google charges for serverless products, such as Cloud Functions and App Engine, when customers exceed the amount allowed under the free tier.

12. The CFO of your company is concerned that they will learn of unusually high cloud computing bills only after charges have been incurred. What mechanism in Google Cloud could be used to address the CFO's concern?

    C: Budgeting and Alerting

    … allows you to specify a budget. When specified percentages of that budget is spent, alerts can be generated. Cloud Monitoring is an observability service for application and infrastructure performance, not billing. Cloud Logging is an observability service for collecting information about events in services and infrastructure. Policy Constraints are a mechanism for restricting how resources can be used.

13. A large enterprise is planning to use Google Cloud across several subdivisions. Each subdivision is managed independently and has its own budget. Most subdivisions plan to spend tens of thousands of dollars per month. How would you recommend they set up their billing account(s)?

    D: Use multiple invoiced billing accounts

    Large enterprises should use invoicing when incurring large charges, which makes option D the right answer. A self-service account is appropriate only for amounts that are within the credit limits of credit cards. Since the subdivisions are independently managed and have their own budgets, each should have its own billing account.

14. An application administrator is responsible for managing all resources in a project. They want to delegate responsibility for several service accounts to another administrator. If additional service accounts are created, the other administrator should manage those as well. What is the best way to delegate privileges needed to manage the service accounts?

    A: Grant iam.serviceAccountUser toe the administrator at the project level

    When a user is granted iam.serviceAccountUser at the project level, that user can manage all service accounts in the project, so option A is correct. If a new service account is created, they will automatically have privileges to manage that service account. You could grant iam.serviceAccountUser to the administrator at the service account level, but that would require setting the role for all service accounts. If a new service accounts is created, the application administrator would have to grant iam.serviceAccountUser to the other administrator on the new service account.

15. You work for a retailer with a large number of stores. Every night the stores upload daily sales data. You have been tasked with creating a service that verifies the uploads every night. You decide to use a service account. Your manager questions the security of your proposed solution, particularly about authenticating the service account. You explain the authentication mechanism used by service accounts. What authentication mechanism is used?

    C: Encryption keys

    When a service account is created, Google generates encryption keys for authentication, making option C correct. Usernames and passwords are not an option for service accounts. Two-factor authentication is an authentication practice that requires two forms of authentication, such as a username password pair and a code from an authentication device. Biometrics cannot be used by services and is not an option.

16. What objects in Google Cloud are sometimes treated as resources and sometimes as identities?

    B: Service accounts

    … are resources that are managed by administrators, but they also function as identities that can be assigned roles, which makes option B the correct answer. Billing accounts are not related to identities. Projects are not identities; they cannot take on roles. Roles are resources but not identities. They can take on privileges, but those privileges are used only when they are attached to an identity.

17. You plan to develop a web application using products from the Google Cloud that already include established roles for managing permissions such as read-only access or the ability to delete old versions. Which of the following roles offers these capabilities?

    B: Predefined roles

    … are defined for a particular product, such as Cloud Run or Compute Engine, so option B is the right answer. They bundle privileges often needed together when managing or using a service. Basic roles are building blocks for other roles. Custom roles are created by users to meets their particular needs.

18. You are reviewing a new Google Cloud account created for use by the finance department. An auditor has questions about who can create projects by default. You explain who has privileges to create projects by default. Who is included?

    B: All users

    By default all users in an organization can create projects, which makes option B correct. The role resourcemanager.projects.create allows to create projects. The billing account is not associated with creating projects.

19. How many projects can be created in an account?

    D: Each account has a limit determined by Google

    The maximum number of organizations is determined on a per-account basis by Google, so option D is the correct answer. If you need additional organizations, you can contact Google and ask for an increase in your limit.

20. You are planning how to grant privileges to users of your company's Google Cloud account. You need to document what each user will be able to do. Auditors are most concerned about a role called Organization Administrator. You explain that users with that role can perform a number of tasks, which include all of the following except which one?

    B: Determining what permissions a user should be assigned

    Users with the Organization Administrator role are not necessarily responsible for determining what permissions should be assigned to users. That is determined based on the person's role in the organization and the security policies established within the organization, which makes option B correct.

Chapter 4: Introduction to Computing in Google Cloud

Google Cloud offers several computing options. The options vary in the level of control that you, as a user of Google Cloud, have over the computing platform. Generally, with more control comes more responsibility and management overhead. Your objective when choosing a computing platform is to choose one that meets your requirements while minimizing DevOps overhead and cost. Compute Engine is the Google Cloud service that lets you provision VMs. You can choose from predefined configurations, or you can create a custom configuration with the best combination of virtual CPUs and memory for your needs. If you can tolerate some disruption in VM functioning, you can save a significant amount of memory by using preemptible VMs.

Modern software applications are built on multiple services that may have different computing requirements and change on different life cycle. Kubernetes Engine runs clusters of servers that can be used to run a variety of services while efficiently allocating work to servers as needed. Kubernetes Engine also provides monitoring, scaling, and remediation when something goes wrong with a VM in the cluster. As enterprise adopt Kubernetes and run multiple clusters, they can turn to Anthos for managing Kubernetes clusters in Google Cloud, other clouds, and on-premises. Cloud Run is a managed service for running stateless containers. If you do not need the full functionality and feature-richness of Kubernetes Engine, Cloud Run is a good option for deploying stateless contains. Loosely coupled applications may be strung together to implement complex workflows. Often, we want each component to be independent of others. In such cases, we often need to execute "glue" code that moves workload from one stage to another. Cloud Functions is the serverless computing option designed to meet this need.

Understand how images are used to create instances of VMs and how VMs are organized in projects. Instances run images, which contain operating system, libraries, and other code. When you create an instance, you specify a project to contain the instance.

Know that Google Cloud has multiple geographic regions and regions have one or more zones. VMs run in zones. A region is a geographical location, such as asia-east1, Europe-west2, and us-east4. The zones within a region are linked by low-latency, high-bandwidth network connections.

Understand what preemptible VMs are and when they are appropriate to use. Also understand when *not* to use them. Google Cloud offers an option called a preemptible VM for workloads that can be disrupted without creating problems.

Understand the difference between the App Engine standard and flexible environment. The standard environment runs a language-specific platform, and the App Engine flexible environment allows you to run custom containers. App Engine is well suited for HTTP(S)-based applications.

Know that Kubernetes is a container orchestration platform. It also runs containers in a cluster.

Understand Kubernetes. It provides load balancing, automatic scaling, logging, and node health checks and repair. Also know that Anthos is used to manage multiple Kubernetes clusters across Google Cloud, other clouds, and on-premises.

Understand Cloud Run. Cloud Run is a managed service for running stateless containers and is a good option when you do not need the full functionality of Kubernetes Engine.

Understand Cloud Functions. This service is used to run programs in response to events, such as file upload or a message being added to a queue.

1. You are deploying a Python web application to Google Cloud. The application uses only custom code and basic Python libraries. You expect to have sporadic use of the application for the foreseeable future and want to minimize both the cost of running the application and the DevOps overhead of managing the application. Which computing service is the best option for running the application?

   B: App Engine standard environment

   The App Engine standard environment can run Python applications, which can autoscale down to no instances when there is no load and thereby minimize costs. Compute Engine and the App Engine flexible environment both require more configuration management than the App Engine standard environment. Kubernetes Engine is used when a cluster of servers is needed to support large or multiple applications using the same computing resources.

2. Your manager is concerned about the rate at which the department is spending on cloud services. You suggest that your team use preemptible VMs for all the following except which one?

   A: Database servers

   … require high availability to respond to queries from users or applications. Preemptible machines are certain to shut down in at most 24 hours unless they are spot VMs. A batch processing job with no fixed time requirements could use preemptible machines as long as the VM is restarted. High-performance computing clusters can use preemptible machines because work on a preemptible machine can be automatically rescheduled for another node on the cluster when a server is preempted.

3. What parameters need to be specified when creating a VM in Compute Engine?

   A: Project and zone

   VMs are created in projects, which are part of the resource hierarchy. They are also located in geographic regions and data centers, so a zone is specified as well. Usernames and admin roles are not specified during creation. The billing account is tied to a project and so does not have to be specified when the VM is created.

4. Your company has licensed a third-party software package that runs on Linux. You will run multiple instances of the software in Docker containers. Which of the following Google Cloud services could you use to deploy this software package?
   C: Compute Engine, Kubernetes Engine, and the App Engine flexible environment only
   Compute Engine can run Docker containers if you install Docker on the VM. Kubernetes and the App Engine flexible environment support Docker containers. The App Engine standard environment provides language-specific runtime environments and does not allow customers to specify custom Docker images for use.

5. You can specify packages to install into a Docker container by including commands in which file?
   B: Dockerfile

6. Which of the following could be managed using Anthos?
   D: Kubernetes clusters in Google Cloud, AWS, and on-premises
   Anthos is a managed service for administrating Kubernetes clusters in Google Cloud, other clouds, and on-premises.

7. Your manager is making a presentation to executives in your company advocating that you start using Kubernetes Engine. You suggest that the manager highlight all the features Kubernetes provides to reduce the workload on DevOps engineers. You describe several features, including all of the following except which one?
   B: Security scanning for vulnerabilities
   Kubernetes provides load balancing, scaling, and automatic upgrading of software. It does not provide vulnerability scanning. Google Cloud's Web Security Scanner service and the Container Analysis service can detect vulnerabilities, but they are separate from Kubernetes Engine.

8. Your company is about to release an online service that builds on a new user interface experience driven by a set of services that will run on your servers. A separate set of services manage authentication and authorization. A third set of services keeps track of account information. All three set of services must be highly reliable and scale to meet demand. Which of the Google Cloud services is the best option for deploying this?
   D: Kubernetes Engine
   The scenario described is a good fit for Kubernetes. Each of the groups of services can be structured in pods and deployed using Kubernetes deployment. Kubernetes Engine manages node health, load balancing, and scaling. App Engine standard environment has language-specific sandboxed and is not a good fit for this use case. Cloud Functions is designed for short-running even processing and is not the kind of continuous processing needed in this scenario. Compute Engine could meet the requirements of this use case, but it would require more effort on the part of application administrators and DevOps professionals to configure load balancers, monitor health, and manage software deployments.

9. A mobile application uploads images for analysis, including identifying objects in the image and extracting text that may be embedded in the image. A third party has created the mobile application, and you have developed the image analysis service. You both agree to use Cloud Storage to store images. You want to keep the two services completely decoupled, but you need a way to invoke the image analysis as soon as possible as an image is uploaded. How should this be done?
   B: Write a function in Python that is invoked by Cloud Functions when a new image file is written to the Cloud Storage bucket that receives new images. The function should submit the URL of the uploaded file to the image analysis service. The image analysis service will then load the image from Cloud Storage, perform analysis, and generate results, which can be saved to Cloud Storage.
   This is an ideal use case for Cloud Functions. The cloud function is triggered by a file upload event. The cloud function calls the image processing service. With this setup, the two services are independent. No additional servers are required.

10. Your team is developing a new pipeline to analyze a stream of data from sensors on manufacturing devices. The old pipeline occasionally corrupted data because parallel threads overwrote data written by other threads. You decide to use Cloud Functions as part of the pipeline. As a developer of a Cloud Function, what do you have to do to prevent multiple invocations of the function from interfering with each other?
    D: Nothing. Google Cloud ensures that functions invocations do not interfere with each other
    Each invocation of a cloud function runs in a secure, isolated runtime environment. There is no need to check whether other invocations are running. With the Cloud Functions service, there is no way for a developer to control code execution at the process or thread level.

11. A client of yours processes personal and health information for hospitals. All health information needs to be protected according to government regulations. Your client wants to move their application to Google Cloud but wants to use the encryption library that they have used in the past. You suggest that all VMs running the application have the encryption library installed. Which kind of image would you use for that?
    A: Custom image
    You would create a custom image after you installed the custom code, in this case the encryption library. A public image does not contain custom code, but it could be used as the base that you add custom code to. Both CentOS and Ubuntu are Linux distributions. You could use either as the base image that you add custom code to, but on their own, they do not have custom code.

12. What is the lowest level of the resource hierarchy?
    B: Project

13. Your company is seeing a marked increase in the rate of customer growth in Europe. Latency is becoming an issue because your application is running in us-central1. You suggest deploying your services to a region in Europe. You have several choices. You should consider all the following factors except which one?
    D: Reliability
    All Google regions have the same level of service level agreement, so reliability is the same. Costs may differ between regions. Regulations may require that data stay within a geographic area, such as the European Union. Latency is a consideration when you want a region that is close to end users or data you will need is already stored in a particular region.

14. What role gives users full control over Compute Engine instances?
    B: Compute Admin role
    The Compute Engine Admin role gives users complete control over instances. Compute Engine Security Admin gives users the privilege to create, modify, and delete SSL certificates and firewall rules.

15. Which of the following are limitations of a preemptible VM?
    D: All of the above (Will be terminated within 24 hours, may not always be available and availability may vary across zones and regions, and cannot migrate to a regular VM)
    Preemptible VMs will be terminated after 24 hours with the exception of spot VMs. Google does not guarantee that preemptible VMs will be available. Once an instance is started as preemptible machine, it cannot migrate to a regular VM. You could, however, save a snapshot and use that to create a new regular instance.

16. Which of the following would eliminate Cloud Run as an option for deploying an application in Google Cloud?
    B: The application stores data about a session in memory for use across multiple request during a session.
    The application maintains state and therefore cannot run in Cloud Run. Cloud Run is a managed service for running application in containers, including Docker-based containers. Containers can run applications written in a variety of languages.

17. When using the App Engine standard environment, which of the following languages' runtime is not supported?
    C: C
    The C programming language is not supported in the App Engine standard environment. If you need to run a C application, it can be compiled and run in a container running in the App Engine flexible environment.

18. You want to be sure all services in a Kubernetes Engine cluster use the same authentication and monitoring services. What service would you use?
    B: Anthos Service Math
    … is a managed service that enables consistent security and monitoring services in Kubernetes clusters. Cloud Functions is used for event processing. App Engine standard and App Engine flexible environment are services for running containerized applications.

19. You are developing a set of virtual machines in Compute Engine. You want to ensure that malware does not compromise the operating system, so you want to validate boot integrity. What feature of Compute Engine would you enable?
    B: vTPM
    … verifies the boot integrity of Compute Engine instances and is used to prevent rootkits and other malicious software from compromising the operating system. Customer-supplied encryption keys are used to encrypt data at rest. Sole tenancy limits which instances can run on a server but does not validate boot integrity. Identity and access management is used to assign roles and permissions to control access to resources in Google Cloud.

20. A client has brought in to help reduce their DevOps overhead. Engineers are spending too much time patching servers and optimizing server utilization. They want to move to serverless platforms as much as possible. Your client has heard of Cloud Functions and wants to use them. You recommend all the following types of applications except which one?
    A: Long-running data warehouse data load procedures
    Cloud Functions is best suited for event-driven processing, such as a file being uploaded to Cloud Storage or an event being written to a Pub/Sub queue. Long-running jobs, such as loading data into a data warehouse, are better suited to Compute Engine or App Engine.

Chapter 5: Computing with Compute Engine Virtual Machines

Google Cloud Console is web-based graphical user interface for managing Google Cloud resources. Cloud SDK is a command-line package that allows engineers to manage cloud resources from the command-line of their local device. Cloud Shell is a web-based terminal interface to VMs. Cloud SDK is installed in Cloud Shell. When creating a VM, you must specify a number of parameters, including a name for the VM, a region and zone where the VM will run, a machine type that specifies the number of vCPUs and the amount of memory, and a boot disk that includes an operating system. gcloud is the top-level command of the hierarchical command structure in Cloud SDK. Common tasks when managing VMs are starting and stopping instances, using SSH to access a terminal on the VM, monitoring, and tracking the cost of the VM.

Understand how to use Cloud Console and Cloud SDK to create, start, and stop VMs. Parameters that you will need to provide when creating a VM include name, machine type, region, zone, and boot disk. Understand the need to create a VM in a project.

Know how to configure a spot VM using Cloud Console and the gcloud commands. Know when to use a spot VM and when not to. Know that spot VMs cost up to 80 percent less than standard VMs.

Know the purpose of advanced options, including Shielded VMs and advanced boot disk configurations. Know that advanced options provide additional security. Understand the kinds of protections provided.

Know how to use gcloud compute instances commands to list, start, and stop VMs. Know the structure of gcloud commands. gcloud commands start with gcloud followed by a service, such as compute, followed by a resource type, such as instances, followed by a command or verb, like create, list, or describe.

Understand how to monitor a VM. Know where to find CPU utilization, network monitoring, and disk monitoring in the VM instances pages of the console. Know the difference between listing and describing instances with a gcloud command.

Know the factors that determine the cost of a VM. Know that Google charges by the second with a 1-minute minimum. Understand that the costs of a machine type may be different in different locations. Know that cost is based on the number of vCPUs and memory.

1. You have just opened the Google Cloud console at http://console.google.com. You have authenticated with the user you want to use. What is one of the first things you should do before performing tasks on VMs?
   C: Verify that the selected project is the on you want to work with
   You should verify the project selected because all operations you perform will apply to resources in the selected project, making option C the correct answer. You do not need to open Cloud Shell unless you want to work with the command line, and if you did, you should verify that the project is correctly selected first. Logging into a VM using SSH is one of the tasks that requires you to be working with the correct project to see the VMs associated with that project, so logging in via SSH should not happen before verifying the project. The list of VMs in the VM instance window is a list of VMs in the current project. You should verify which project you are using to ensure you are viewing the set of VMs you think you are using.

2. What is a one-time task you will need to complete before using the console?
   A: Set up billing
   You will need to set up billing if it is not already enabled when you start using the console, so option A is the right answer. You may create a project, but you will be able to do this only if billing is enabled. You do not need to create a storage bucket to work with the console. Specifying a default zone is not a one-time task; you may change zones throughout the lift of your project.

3. A colleague has asked for your assistance setting up a test environment in Google Cloud. They have never worked in Google Cloud. You suggest starting with a single VM. Which of the following is the minimal set of information you will need?
   B: A name for the VM, a machine type, a region, and a zone

4. An architect has suggested a particular machine type for your workload. You are in the console creating a VM and you don't see the machine type in the list of available machine types. What could be the reason for this?
   B: That machine type is not available in the zone you specified
   Different zones may have different machine types available, so you will need to specify a region first and then a zone to determine the set of machine types available. If the machine type does not appear in the list, it is not available in that zone.

5. Your manager asks for your help with understanding cloud computing costs. Your team runs dozens of VMs for three different applications. Two of the applications are for use by the marketing department and one is used by the finance department. Your manager wants a way to bill each department for the cost of the VMs used for their applications. What would you suggest to help solve this problem?
   C: Labels and descriptions
   … help you track your own attributes of resources. As the number of servers grows, it can become difficult to trach which VMs are used for which applications and services. Labels and a general description will help administrators track numbers of VMs and their related costs.

6. If you wanted to set the preemptible property using Cloud Console, in which section of the Create An Instance page would you find the option?
   A: Availability Policy
   The Availability Policy section within the Management tab is where you set the preemptible option. Identity and API Access is used to control the VM's access to Google Cloud APIs and which service account is used with the VM. Sole Tenancy is used if you need to run your VMs on physical servers that only run your VMs. Networking is used to set network tags and change the network interface.

7. You need to set up a server with high level of security. You want to be prepared in case of attacks on your server by someone trying to inject a rootkit (a kind of malware that can alter the operating system). Which option should you select when creating a VM?
   B: Shielded VM
   … is an advanced set of security controls that includes Integrity Monitoring, a check to ensure boot images have not been tampered with, which makes option B the right answer. Firewalls are used to control ingress and egress of network traffic to a server or subnet. SSH keys are used for authenticating users across servers within a project. Boot disk integrity control service is a fictional feature.

8. All of the following parameters can be set when adding an additional disk through Google Cloud Console, except:
   C: Block size
   … is not an option in under Additional Disks. Encryption key management, disk type, and the option of specifying a source image are all available options.

9. You lead a team of cloud engineers who maintain cloud resources for several departments in your company. You have noticed a problem with configuration drift. Some machine configurations are no longer in the same state as they were when created. You can't find notes or documentation on how the changes were made or why. What practice would you implement to solve this problem?
   B: Write scripts using gcloud commands to change configuration and store those scripts in a version control system
   Using version-controlled scripts is the best approach of the four options. Scripts can be documented with reasons for the changes, and they can be run repeatedly on different machines to implement the same change. This reduces the chance of error when manually entering a command. You could become a bottleneck to making changes, changes cannot be made when you are unavailable, and your memory may not be a reliable way to track all configuration changes.

10. When you're using the Cloud SDK command-line interface, which of the following is part of commands for administrating resources in Compute Engine?
    A: gcloud compute instances

11. A newly hired cloud engineer is trying to understand what VMs are running in a particular project. How could the engineer get summary information on each VM running in a project?
    B: execute the command gcloud compute instances list

12. When creating a VM using the command-line, how should you specify label for the VM?
    B: Use the –labels option with labels in the format of KEYS=VALUE

13. In the boot disk advanced configuration, which operations can you specify when creating a new VM?
    C: Add a new disk and attach an existing disk

14. You have acquired a 10 GB data set from a third-party research firm. A group of data scientists would like to access this data from their statistics programs written in R. R works well with Linux and Windows filesystems, and the data scientists are familiar with file operations in R. The data scientists would each like to have their own dedicated VM with the data available in the VM's filesystem. What is a way to make this data readily available on a VM and minimize the steps the data scientists will have to take?
    B: Create VMs using a source image created from a disk with the data on it
    10 GB of data is small enough to store on a single disk. By creating an image of a disk within the data stored on it, you can specify that source image when creating a VM.

15. The Networking tab of the Create VM form is where you would perform which of the following operations?
    B: Add a network interface to the VM

16. You want to create a VM using the gcloud command. What parameter would you include to specify the type of boot disk?
    A: boot-disk-type

17. Which of the following commands will create a VM with four CPUs that is named web-server-1?
    A: gcloud compute instances create –machine-type=n1-standard-4 webserver-1

18. Which of the following commands will stop a VM named web-server-1?
    C: gcloud compute instances stop web-server-1

19. You have created an Ubuntu VM and want to log into the VM to install some software packages. Which network service would you use to access the VM?
    B: SSH
    … is a service for connecting to a remote server and logging into a terminal window. Once logged in, you would have access to a command line, so option B is the right answer.

20. Your management team is considering three different cloud providers. You have been asked to summarize billing and cost information to help the management team compare cost structures between clouds. Which of the following would you mention about the cost of VMs in Google Cloud?
    A: VMs are billed in 1 -second increments, cost varies with the number of CPUs and amount of memory in a machine type, you can create custom machine types, preemptible VMs cost up to 80 percent less than standard VMs, and Google offers discounts for sustained usage.

Chapter 6: Managing Virtual Machines

In this chapter, you learned how to manage a single VM instance and instance groups. Single VM instances can be created, configured, stopped, started, and deleted using Cloud Console or gcloud commands from Cloud Shell or your local machine if you have SDK installed. Snapshots are copies of disks and are useful as backups and for copying data to other instances. Images are complete backups of a boot disk, so are used to create VMs. Snapshots made from boot disk can also be used to create a VM. The main command used to manage VMs is the gcloud compute instances command. gcloud uses a hierarchical structure to order the command begins with gcloud, followed by an entity type such as instances or snapshots. An operation is then specified, such as create, delete, list, or describe. GPUs can be attached to instances that have GPU libraries installed in the operating system. GPUs are used for compute-intensive tasks, such as building machine learning models. Instance groups are groups of instances that are managed together. Managed instance groups have instances that are the same. These groups support load balancing and autoscaling.

Understand how to navigate Cloud Console. Cloud Console is the graphical interface for working with Google Cloud. You can create, configure, delete, and list VM instances from the Compute Engine area of the console.

Understand how to install Cloud SDK. Cloud SDK allows you to configure default environment variables, such as a preferred zone, and issue commands from the command line. If you use Cloud Shell, Cloud SDK is already installed.

Know how to create a VM in the console and at the command-line. You can specify machine type, choose an image, and configure disks with the console. You can use commands at the command-line to list and describe, and you can find the same information in the console. Understand when to use customized images and how to deprecate them. Images are copies of contents of a disk, and they are used to create VMs. Deprecated marks an image as no longer supported.

Understand why GPUs are used and how to attach them to a VM. GPUs are used for compute-intensive operations; a common use case for using GPUs is machine learning. It is best to use an image that has GPU libraries installed. Understand how to determine which locations have GPUs available, because there are some restrictions. The CPU must be compatible with GPU selected, and GPUs cannot be attached to shared memory machines. Know how GPU costs are charged.

Understand images and snapshots. Snapshots save the contents of disks for backup and data-sharing purposes. Images save the operating system and related configurations so that you can create identical copies of the instance.

Understand instance groups and instance groups templates. Instance groups are sets of instances managed as a single entity. Instance group templates specify the configuration of an instance group and the instance in it. Managed instance groups support autoscaling and load balancing.

1. Which page in Google Cloud Console would you use to create a single instance of a VM?
   A: Compute Engine

2. You view a list of Linux VM instances in the console. All have public AP addresses assigned. You notice that the SSH option is disabled for one of the instances. Why might that be the case?
   B: The instance is stopped
   Instances can be stopped, and when they are, then you cannot connect to them via SSH, which makes option B the correct answer. Starting the instance will enable SSH access.

3. You have noticed unusually slow response time when issuing commands to a Linux server, and you decide to reboot the machine. Which command would you use in the console to reboot?
   B: Reset
   The Reset command can be used to restart a VM. The properties of the VM will not change, but data in memory will be lost. There is no Reboot, Restart, Shutdown, or Startup option in the console.

4. In the console, you can filter the list of VM instances by which of the following?
   C: Labels, status, or deletion prevention
   … are all available for filtering. You can also filter by internal IP, external IP, zone, network, deletion protection, and member of a managed or unmanaged instance group.

5. You will build several machine learning models on an instance and attaching GPU to the instance. When you run your machine learning models they take an unusually long time too run. It appears that GPU is not being used. What could be the cause of this?
   A: GPU libraries are not installed
   To function properly, the operating system must have GPU libraries installed. The operating system does not have to be Ubuntu-based, and there is no need to have at least eight CPUs in an instance before you can attach and use a GPU. Available disk space does not determine whether or not a GPU is used.

6. When you add a GPU to an instance, you must ensure that:
   A: The GPU and CPU are compatible
   If you add a GPU to a VM, you must have compatible CPUs and GPUs. The instance does not need to be preemptible and it can not have nonboot disks attached.

7. You are using snapshots to save copies of a 100 GB disk. You make a snapshot and then add 10 GB of data. You create a second snapshot. How much storage is used in total for the two snapshots (assume no compression)?
   B: 110 GB, with 100 GB for the first and 10 GB for the second
   When you first create a snapshot, Google Cloud will make a full copy of the data on the persistent disk. The next time you create a snapshot from that disk, Google Cloud will only copy the data that has changed since the last snapshot.

8. You have decided to delegate the task of making backup snapshots to a member of your team. What role would you need to grant to your team member to create snapshots?
   D: Compute Storage Admin

9. The source of an image may be:
   C: Disks, snapshots, or another image
   Images can be created from disks, snapshots, cloud storage files, a virtual disk, or another image.

10. You have built images using Ubuntu 18.04 and now want users to start using Ubuntu 20.04. You do not want to just delete images based on Ubuntu 18.04, but you want users to know they should start using Ubuntu 20.04. What feature of images would you use to accomplish this?
    B: Deprecated
    … marks the image as no longer supported and allows you to specify a replacement image to use going forward. Deprecated images are available for use but may not be patched for security flaws or have other updates.

11. You want to generate a list of VMs in your inventory and have the results in JSON format. What command would you use?
    C: gcloud compute instances list --format=json

12. You would like to understand details of how Google Cloud starts a virtual machine instance. Which optional parameter would you use when starting an instance to display those details?
    B: --async

13. Which command line will delete an instance named ch06-instance-3?
    C: gcloud compute instances delete ch06-instance-3

14. You are about to delete an instance named ch06-instance-1 but want to keep its boot disk. You do not want to keep other attached disks. What gcloud command would you use?
    A: gcloud compute instances delete ch06-instance-1 --keep-disks=boot

15. You want to view a list of fields you can use to sort a list of instances. What command would you use to see the field names?
    B: gcloud compute instances describe

16. 16. You are deploying an application that will need to scale and be highly available. Which of these Compute Engine components will help achieve scalability and high availability?
    B: Instance groups
    … are sets of VMs that can be configured to scale and are used with load balancers, which contribute to improving availability. Preemptible instances are not highly available because they can be shut down at any time by Google Cloud. Cloud Storage is not a Compute Engine component.

17. Before creating an instance group, what do you need to create?
    B: Instance template
    An instance template is used to specify how the instance group should be created.

18. How would you delete an instance group using the command line?
    B: gcloud compute instance-templates delete

19. What can be the basis for scaling up an instance group?
    C: Network latency, load balancing capacity, and CPU utilization
    You can configure an autoscaling policy to trigger adding or removing instances based on CPU utilization, monitoring metric, load balancing capacity, or queue-based workloads. Disk, network latency, and memory can trigger scaling if monitoring metrics on those resources are configured.

20. An architect is moving a legacy application to Google Cloud and wants to minimize the changes to the existing architecture while administrating the cluster as a single entity. The legacy application runs on a load-balanced cluster that runs nodes with two different configurations. The two configurations are required because of design decisions made several years ago. The load on the application is consistent, so there is rarely a need to scale up or down. What Google Cloud Compute Engine resource would you recommend using?
    B: Unmanaged instance groups
    … are available for limited use cases such as this. Unmanaged instance groups are not recommended in general. Managed instance groups are the recommended way to use instance groups, but the two different configurations prevent their use.

Chapter 7: Introduction to Kubernetes Engine

Kubernetes Engine is a container orchestration system for deploying applications to run in clusters. Kubernetes is architected with a single cluster manager and worker nodes. Kubernetes uses the concept of pods, or instances running a container. It is possible to run multiple containers in a pod, but this is usually only done when the two containers are tightly coupled. ReplicaSets are controllers for ensuring that the correct number of pods are running. Deployments are sets of identical pods. StatefulSets are a type of deployment used for stateful applications. Kubernetes clusters can be deployed through Cloud Console or by using gcloud commands. You deploy applications by bundling the application in a container and using the console or the kubectl command to create a deployment that runs the applications on the cluster. Cloud Operations Suite includes Cloud Monitoring and Cloud Logging, which is used to monitor instances in clusters.

Understand that Kubernetes is a container orchestration system. Kubernetes Engine is a Google Cloud product that provides Kubernetes to Google Cloud customers. Kubernetes manages containers that run in a set in a set of VM instances.

Understand that Kubernetes uses a control plane to manage nodes and workloads. Kubernetes uses the control plane to coordinate execution and monitor the health of pods. If there is a problem with a pod, the control plane can correct the problem and reschedule the disrupted job.

Be able to describe pods. Pods are single instances of a running process, services provide a level of indirection between pods and clients calling services in the pods, a ReplicaSet is a kind of controller that ensures that the right number of pods are running, and a deployment is a set of identical pods.

Kubernetes can be deployed using Cloud Console or using gcloud commands. gcloud commands manipulate the Kubernetes Engine service, whereas kubectl commands are used to manage the internal state of clusters from the command-line. The base command for working with Kubernetes Engine is cloud container. Note that gcloud and kubectl have different command syntaxes. kubectl commands specify a verb and then a resource, as in kubectl scale deployment …, whereas gcloud specifies a resource before the verb, as in gcloud container clusters create. Deployments are created using Cloud Console or at the command-line using a YAML specification.

Be able to define Kubernetes objects. Deployments are sets of identical pods. StatefulSets are a type of deployment used for stateful applications. Kubernetes is monitored using Cloud Operations. Cloud Operations can be configured to generate alerts and notify you on a variety of channels. To monitor the state of a cluster, you can create a policy that monitors a metric, like CPU utilization, and have notifications sent to email or other channels.

1. A new engineer is asking for clarification about when is it best to use Kubernetes and when to use instance groups. You point out that Kubernetes uses instance groups. What purpose do instance groups play in a Kubernetes cluster?
   C: They create sets of VMs that can be managed as a unit
   Kubernetes creates instance groups as part of the process of creating a cluster.

2. What components are required in a Kubernetes cluster?
   A: A control plane and nodes to execute workloads
   A Kubernetes cluster has a single control plane and one or more nodes to execute workloads. There is no monitoring node in Kubernetes, but it does generate metrics that can be sent to Cloud Monitoring.

3. What is a pod in Kubernetes?
   C: A single instance of a running application in a cluster
   Pods are single instances of a running application in a cluster. Pods run containers but are not simply sets of containers. Application code runs in containers that are deployed in pods. Pods are not controllers, so they cannot manage communication with clients and Kubernetes services.

4. You have developed an application that calls a service running in a Kubernetes cluster. The service runs in pods that can be terminated if they are unhealthy and replaced with other pods that might have a different IP address. How should you code your application to ensure it functions properly in this situation?
   B: Communicate with Kubernetes Services so that applications do not have to be coupled to specific pods
   Services are Kubernetes components providing API endpoints that allow applications to discover pods running in a particular application.

5. You have noticed that an application's performance has degraded significantly. You have recently made some configuration changes to resources in your Kubernetes cluster and suspect that those changes have altered the number of pods running in the cluster. Where would you look for details on the number of pods that should be running?
   A: Deployment config
   A deployment config specifies how many nodes to create in a ReplicaSet. Cloud Operations Suite is a monitoring and logging service that monitors but does not control Kubernetes clusters. Container Runtime is a component of Kubernetes that is responsible for running containers. Jobs is an abstraction of workloads and is not tied to the number of pods running in a cluster.

6. You are deploying a high-availability application in Kubernetes Engine. You want to maintain availability even if there is a major network outage in a data center. What feature of Kubernetes Engine would you employ?
   B: Regional clusters
   … are available in Kubernetes Engine and are used to provide resiliency to an application. If load is not distributed across zones or regions, it does not help to add resiliency across data centers.

7. You want to write a script to deploy a Kubernetes cluster with GPUs. You have deployed clusters before, but you are not sure about all the required parameters. You need to deploy this script as quickly as possible. What is one way to develop this script quickly?
   A: Use the GPU template in the Kubernetes Engine cloud console to generate the gcloud command to create the cluster
   Starting with an existing template, filling in parameters, and generating the gcloud command is the most reliable way. Using an existing script and add parameters for attaching GPUs may work, but multiple parameters that are needed for your configuration may not be in the script you start with. There may be some trial and error with this option.

8. What gcloud command will create a cluster named ch07-cluster-1 with four nodes?
   A: gcloud container clusters create ch07-cluster-1 --num-nodes=4

9. When using Create Deployment from Cloud Console, which of the following cannot be specified for a deployment?
   C: Time to Live (TTL)
   … is not an attribute of deployments. Application name, container image, and initial command can all be specified.

10. Deployment configuration files created in Cloud Console use what type of file format?
    B: YAML

11. What command is used to run a Docker image on a cluster?
    C: kubectl run
    gcloud is not used to manipulate Kubernetes processes.

12. What command would you use to have 10 replicas of a deployment named ch07-app-deploy?
    C: kubectl scale deployment ch07-app-deploy --replicas-10

13. Cloud Operations Suite is used for what operations on Kubernetes clusters?
    D: Notifications, monitoring, and logging
    Cloud Operations Suite is a comprehensive monitoring, logging, alerting, and notification service that can be used to monitor Kubernetes clusters.

14. You want to use Cloud Logging and Cloud Monitoring with your GKE clusters. What must you do to enable this when creating a cluster?
    D: Nothing; metrics and logs are sent to Cloud Logging and Cloud Monitoring by default
    GKE sends metrics and logs to Cloud Monitoring and Cloud Logging by default, so you do not need to do anything other than accept the default configuration for monitoring and logging.

15. What popular open source monitoring tool is available in Google Cloud as a managed service?
    A: Prometheus

16. You want to create a Kubernetes Engine cluster and want to minimize the amount of configuring and infrastructure management. What kind of cluster would you create?
    B: Autopilot mode clusters
    … require the least configuration and infrastructure management.

17. You want the greatest degree of control over your Kubernetes cluster. What kind of cluster would you create?
    A: Standard mode clusters
    … require you to make configurations and infrastructure choices. Autopilot mode clusters are preconfigured optimized infrastructure and do not give you as much control over configuration and infrastructure as standard mode clusters do.

18. You want to create a Kubernetes cluster, but you do not want GKE to automatically upgrade the cluster. How would you configure the cluster?
    B: With static channel
    B is the correct answer because with a static channel configuration, GKE will not automatically upgrade the cluster.

19. You are attempting to execute commands to initiate a deployment on a Kubernetes cluster. The commands are not having any effect. You suspect that a Kubernetes component is not functioning correctly. What component could be the problem?
    A: The Kubernetes API
    All interactions with the cluster are done through the master using the Kubernetes API. If an action is to be taken on a node, the command is issued by the control plane.

20. You have deployed an application to a Kubernetes cluster. You have noticed that several pods are starved for resources for a period of time and the pods are shut down. When resources are available, new instantiations of those pods are created. Clients are still able to connect to pods even though the new pods have different API addresses from the pods that were terminated. Which Kubernetes component makes this possible?
    A: Services
    … provide a level of indirection to accessing pods. Pods are ephemeral. Clients connect to services, which can discover pods. ReplicaSets and StatefulSets provide managed pods.

Chapter 8: Managing Standard Mode Kubernetes Clusters

In this chapter, you learned how to perform basic management tasks for working with Kubernetes clusters, nodes, pods, and services. The chapter also describe how to list the contents of container image repositors. You learned how to pin services in the Cloud Console menu, view the status of Kubernetes clusters, and view image repository and image details using gloud commands. This chapter also described how to modify and remove nodes and pods. You also saw the benefits of autoscaling in a real-world scenario. Both Cloud Condole and Cloud SDK, including Cloud Shell, can be used to add, remove, and modify nodes, pods, and services. They both can be used to review the contents of an imagine repository. Some of the most useful commands include gcloud container clusters create and gcloud container clusters resize. The kubectl command is used to modify Kubernetes resources such as deployments and pods.

Know how to view the status of a Kubernetes cluster. Use Cloud Console to list clusters and drill down into clusters to see details of the cluster, including node, pod, and container details. Know the gcloud container clusters command and its options.

Understand how to add, modify, and remove nodes. Use Cloud Console to modify nodes and know now to add and remove nodes by changing deployments. Use the gcloud container clusters resize command to add and remove nodes.

Understand how to add, modify, and remove nodes. Use Cloud Console to modify pods and to add and remove pods by changing deployments. Use kubectl get deployments to list deployments, kubectl scale deployment to modify the number of deployments, and kubectl autoscale deployment to enable Autoscaling.

Understand how to add, modify, and remove Services. Use Cloud Console to modify Services and add and remove Services by changing deployments. Use kubectl create deployment to start Services and kubectl expose deployment to make a Service accessible outside the cluster. Delete a service using the kubectl delete service command.

1. You are running several microservices in a Kubernetes cluster. You have noticed some performance degradation. After reviewing some logs, you begin to think the cluster may be improperly configured, and you open Cloud Console to investigate. How do you see the details of a specific cluster?
   B: Click the cluster name

2. You are viewing the details of a cluster in Cloud Console and want to see how many vCPUs are available in the cluster. Where would you look for that information?
   A: Node Pools section of the Nodes Details page

3. You have been assigned to help diagnose performance problems with applications running on several Kubernetes clusters. The first thing you want to do is understand, at a high level, the characteristics of the clusters. Which command should you use?
   B: gcloud container clusters list

4. When you first try to use the kubectl command, you get an error message indicating that the resource cannot be found or you connect to the cluster. What command would you use to to try to eliminate the error?
   B: gcloud container clusters get-credentials

5. An engineer joined your team and is not aware of your team's standards for creating clusters and other Kubernetes objects. In particular, the engineer has not properly labeled several clusters. You want to modify the labels on the cluster from Cloud Console. How would you do it?
   C: Click the Edit menu option

6. You receive a page in the middle of the night informing you that several services running on a Kubernetes cluster have high latency when responding to API requests. You review monitoring data and determine that there are not enough resources in the cluster to keep up with the load. You decide to add six more VMs to the cluster. What parameters will you need to specify when you issue the cluster resize command?
   D: All of the above (Cluster size, cluster name, and node pool name)
   When resizing, the gcloud container clusters resize command requires the name of the cluster and the node pool to modify. The size is required to specify how many nodes should be running.

7. You want to modify the number of pods in a cluster. What is the best way to do that?
   B: Modify deployments
   Pods are used to implement replicas of a deployment. It is a best practice to modify the deployments, which are configured with a specification of the number of replicas that should always run.

8. You want to see a list of deployments. Which option from the Kubernetes Engine navigation menu would you select?
   C: Workloads

9. What actions are available from the Actions menu when viewing deployment details?
   B: Autoscale, Expose, and Rolling Update
   There are four actions available for deployments: Autoscale, Expose, Rolling Update, and Scale

10. What is the command to list deployments from the command-line?
    C: kubectl get deployments

11. What parameters of a deployment can be set in the Create Deployment page in Cloud Console?
    D: All of the above (Container image, cluster name, and application name)
    You can specify container image, cluster name, and application name along with the labels, initial command, and namespace.

12. Where can you view a list of applications when using Cloud Console?
    A: In the Deployment Details page

13. What kubectl command is used to create deployment?
    A: run
    kubectl run is the command used to start a deployment. It takes a name for the deployment, an image, and a port specification.

14. You are supporting machine learning engineers who are testing a series of classifiers. They have five classifiers, called ml-classifier-1, ml-classifier-2, etc. They have found that ml-classifier-3 is not functioning as expected, and they would like it removed from the cluster. What would you do to delete a service called ml-classifier-3?
    A: Run the command kubectl delete service ml-classifier-3

15. What service is responsible for managing container images?
    C: Artifact Registry
    The Artifact Registry is the service for managing images that can be used in other services, including Kubernetes Engine and Compute Engine. Both Kubernetes Engine and Compute Engine use images but do not manage them.

16. What command is used to list container images in the command line?
    A: gcloud container images list
    kubectl is for managing Kubernetes resources, not Google Cloud resources like container images.

17. A data warehouse designer wants to deploy an extract, transform, and load process to Kubernetes. The designer provided you with a list of libraries that should be installed, including drivers for GPUs. You have a number of container images that you think may meet the requirements. How could you get a detailed description of each of those containers?
    B: Run the command gcloud container images describe

18. You have just created a deployment and want applications outside the cluster to have access to pods provided by the deployment. What do you need to do to the deployment?
    B: Issue a kubectl expose deployment command
    The kubectl expose deployment command makes a service accessible.

19. You have deployed an application to a Kubernetes cluster that processes sensor data from a fleet of delivery vehicles. The volume of incoming data depends on the number of vehicles making deliveries. The number of vehicles making deliveries is dependent on the number of customer orders. Customer orders are high during daytime hours, holiday seasons, and when major advertising campaigns are run. You want to make sure you have enough nodes running to handle the load, but you want to keep your costs down. How should you configure your Kubernetes cluster?
    B: Enable Autoscaling
    Autoscaling is the most cost-effective and least burdensome way to respond to changes in demand for a service.

20. When using Kubernetes Engine, which of the following might a cloud engineer need to configure?
    B: Nodes, pods, services, clusters, and container images
    Cloud engineers working with Kubernetes will need to be familiar with working with clusters, nodes, pods, and container images. They will also need to be familiar with deployment.

Chapter 9: Computing with Cloud Run and App Engine

Cloud Run is a serverless, managed service for running containerized applications. Cloud Run supports any application that can be run in a container. Cloud Run services are used when your code is used to respond to web requests or events. Cloud Run jobs are used when the code executes until a workload is complete. When working with services or jobs, you can configure several categories of parameters, including container, connections, and security settings. App Engine Standard is a serverless platform for running applications in language-specific environments. As a cloud engineers, you are expected to know how to deploy and scale App Engine applications. App Engine applications consist of services, versions, and instances. You can have multiple versions running at one time. You can split traffic between versions and have all traffic automatically migrate to a new version. App Engine applications are configured through app.yaml configuration files. You can specify the language environment, scaling parameters, and other parameters to customize your deployment. App Engine is no longer listed in the Cloud Associate Cloud Engineer Exam Guide, but it is included here because cloud engineers should be familiar with this popular Google Cloud service.

Be able to describe Cloud Run as a serverless service for running containers. Cloud Run is a serverless, managed service for deploying, scaling, and managing services. Although there are no servers to configure, you can specify parameters to control the number of instances running at any time, the security used to process the service, as well as connection configuration details.

Know how Cloud Run services are used to run long-lived services like websites and API servers. Cloud Run services run containers continuously. You have the option to pay only for CPU resources used when responding to requests, or you can choose to have a container always available and pay for the time the CPU resources are allocated.

Know how Cloud Run jobs are used to run tasks, such as loading data into a database. Cloud Run jobs are configured similarly to Cloud Run services. You can specify that jobs are multiple containers running simultaneously. This is useful when running parallelizable workloads.

Be able to describe the structure of App Engine Standard applications. These consists of services, versions, and instances. Services usually provide a single function. Versions are different versions of code running in the App Engine environment. Instances are managed instances running the service.

Know how to deploy an App Engine app. This includes configuring the App Engine environment using the app.yaml file. Know that a project can have only one App Engine app at a time. Know how to use the gcloud the gcloud app deploy command.

Understand the various scaling options. Three scaling options are autoscaling, basic scaling, and manual scaling. Only autoscaling and basic scaling are dynamic. Manual scaling creates resident instances. Autoscaling allows for more configuration options than basic scaling.

1. You want to provide your customers with an API to allow them to query a database with proprietary industry data. You want your developers to focus on adding new features and not on administering services. Which of the following Google Cloud services would you choose?
   C: Cloud Run services
   … are managed, serverless services for running containers and are designed to support containers that run continuously, which is what is needed for an API service.

2. You are working for a biomedical research group that has several hundred data files stored in Cloud Storage. They have a statistical analysis program that analyzes a data file and writes the output to another Cloud Storage bucket. They have agreed with you that deploying the program in a container is the best option, but they are unsure which Google Cloud service to use to run the container. What would you recommend?
   D: Cloud Run jobs
   … are managed, serverless services for running containers and are designed to support executables that run until a task is completed. Kubernetes Engine is used to run containers but is best suited to running large numbers of containers in complicated environments, such as environments that need to support multiple namespaces. Cloud Engine could be used but requires mode administration than using Cloud Run. App Engine Flexible could be used to run a container, but Cloud Run is preferred over App Engine.

3. You are working for a climate change research group that has tens of thousands of public weather files stored in Cloud Storage. They are building a model to predict sea levels in the near future. The data in each file can be analyzed independently of other files. They plan to use Cloud Run jobs for this task. What feature of Cloud Run jobs would you recommend they use?
   B: Array jobs
   … in Cloud Run allow multiple containers to run and process the workload in parallel. Since the data in each file is independent of the data in other files, they can be processed in any order and in parallel.

4. An application administrator has asked for your help with configuring a Cloud Run service. The application administrator would like to have all client requests routed to the same container if possible. How would you suggest the administrator accomplish this?
   C: Configure the connection in the Cloud Run service to support session affinity
   By supporting session affinity in the Connection configuration, Cloud Run will route all requests from a client to the same container if possible.

5. You are deploying a service on Cloud Run. The service has access to personal identifiable information (PII) and for compliance reasons, you do not want to expose the service to any traffic outside of internal traffic in your Google Cloud environment. What ingress configuration would you use?
   A: Internal
   An internal ingress (Internal traffic only) setting will restrict network traffic to internal Google Cloud traffic. Internal (traffic) and (traffic from) Cloud Load Balancing would allow traffic entering trough external load balancing. An all ingress (All traffic) setting would allow all traffic.

6. You want to use a service account specifically created for a Cloud Run service. Where would you specify that in the cloud console?
   B: On the Security tab

7. A group of developers need the ability to deploy new versions of a service running in Cloud Run. How would you configure that access?
   A: Using IAM
   You would grant a role with appropriate permissions to a group that includes the developers who need access.

8. Your team deployed a Cloud Run service last month that accesses a Cloud SQL database. The database team has changed their system and now use a Memcached cache running in Cloud Memorystore. You have to change your Cloud Run service to access the Cloud Memorystore cache. What would you use to do that?
   C: VPC Connection
   … enables the use of Serverless VPC Access to connect your Cloud Run service to other resources in your VPC.

9. A service is deployed to Cloud Run services and will communicate with clients using gRPC. What should you configure to enable this protocol to work with the service?
   D: HTTP/2 end-to-end

10. What Google Cloud services can be used to store and access container images accessible from Cloud Run?
    B: Container Registry and Artifact Registry

11. You have designed a microservice that you want to deploy a production. Before it can be deployed, you have to review how you will manage the service life cycle. The architect is particularly concerned about how you will deploy updates to the service with minimal disruption. What aspect of App Engine components would you use to minimize disruptions during updates to the service?
    B: Versions
    … support migration. An app can have multiple versions, and by deploying with the --migrate parameter, you can migrate traffic to the new version. Services are a higher-level abstraction and represent the functionality of a microservice. An app may have multiple services, but they serve different purposes. Instances execute code in a version. Instances may be added and removed as needed, but they will run only one version of a service.

12. You have just released an application running in App Engine Standard. You notice that there are peak demand periods in which you need up to 12 instances, but most of the time 5 instances are sufficient. What is the best way to ensure that you have enough instances to meet demand without spending more than you have to?
    A: Configure your app for autoscaling and specify max instances of 12 and min instances of 5
    Autoscaling enables setting a maximum and minimum number of instances. Basic scaling does not support maximum and minimum instances.

13. What command should you use to deploy an App Engine app from the command-line?
    B: gcloud app deploy

14. You have deployed a Django 1.5 Python application to App Engine. This version of Django requires Python 3. For some reason, App Engine is trying to run the application using Python 2. What file would you check and possibly modify to ensure that Python 3 is used with this application?
    B: app.yaml
    The app.yaml file is used to configure an App Engine application.

15. You are concerned that as users make connections to your application, the performance will degrade. You want to make sure instances are added to your App Engine application when there are more than 20 concurrent requests. What parameter would you specify in app.yaml?
    A: max_concurrent_requests
    … lets you specify the maximum number of concurrent requests before another instance is started.

16. What parameters can be configured with basic scaling?
    C: idle_timeout and max_instances
    Basic scaling only allows for idle time and maximum instances. min_instances is not supported. target_throughput_utilization is an autoscaling parameter, not a basic scaling parameter.

17. The runtime parameter in app.yaml is used to specify what?
    C: The language runtime environment
    The runtime parameter specifies the language environment to execute in. The script to execute is specifies by the script parameter. The URL to access the application is based on the project name and the domain appshot.com.

18. You work for a startup, and costs are a major concern. You are willing to take a slight performance hit if it will safe you money. How should you configure the scaling for your apps running in App Engine?
    A Use dynamic instances by specifying autoscaling or basic scaling
    Using dynamic instances by specifying autoscaling or basic scaling will automatically adjust the number of instances in use based on load.

19. What parameter to gcloud app services set-traffic is used to specify the method to use when splitting traffic?
    B: --split-by

20. What are valid methods for splitting traffic in App Engine?
    D: By IP address, HTTP cookies, and randomly

Chapter 10: Computing with Cloud Functions

Cloud Functions can be created using either Google Cloud Console or the command-line. To use Cloud Functions, it is important to understand the relationship between events, triggers, and functions. Events are actions that happen to the cloud. Different services have different types of events. Triggers are how you indicate you want to execute a function when an event occurs. Cloud Functions refer to the code that is executed when an event occurs that has a trigger defined for it.

Know the relationship between events, triggers, and functions. Events are actions that happen, such as a file is uploaded to Cloud Storage or a message is written to a Cloud Pub/Sub topic. Triggers are declarations that an action should be taken when an event occurs. Cloud Functions associated with triggers define what actions are taken when an event occurs.

Know when to use Cloud Functions. Cloud Functions is a service that supports single-purpose functions that respond to events in the cloud. Cloud Run is also a serverless computing option, but it is used to deploy multifunction applications, including those that users interact with directly.

Know the runtimes and generations supported in Cloud Functions. Cloud Functions support the following runtimes: Node.js, Python, Go, Java, .NET, Ruby, and PHP. There are two generations of Cloud Functions, the original is known as First Generation and the other Second Generation. Second Generation Cloud Functions have fewer constraints and more functionality.

Know the parameters for defining a Cloud Functions on a Cloud Storage event. Parameters for Cloud Storage include the following:

- Cloud function name
- Memory allocated for the function
- Trigger
- Topic
- Source of the function code
- Runtime
- Source code
- Name of the function to execute

Know the gcloud commands for working with Cloud Functions. These include the following:

- gcloud functions only
- gcloud functions delete

1. A product manager is proposing a new application that will require several back-end services and three business logic services. Each service will provide a single function, and it will require several of these services to complete a business task. Service execution time is dependent on the size of input and is expected to take up to 90 minutes in some cases. Which GCP product is a good serverless option for running this related service?
   C: Cloud Run
   … is a serverless service for running containerized applications that run continuously and provide an endpoint. This is unlike Cloud Functions, which is designed to support single-purpose functions that operate independently and in response to isolated events in the Google Cloud and complete within a specified period of time. Compute Engine is not a serverless option. Cloud Storage is not a computing product.

2. You have been asked to deploy a cloud function to reformat image files as soon as they are uploaded to Cloud Storage. You notice after a few hours that 10 percent of the files are not processed correctly. After reviewing the files that failed, you realize they are all substantially larger than average. What could be the cause of the failures?
   C: The timeout it too low to allow enough time to process large files
   A timeout period that is too low would explain why the smaller files are processed in time but the larger are not. If only 10 percent of the files are failing, then it is not a syntax error or the wrong runtime selected. Those errors affect all files, not just the larger ones. Similarly, if there was a permission problem with the Cloud Storage bucket, it would affect all files.

3. When an action occurs in GCP, such as a file written to Cloud Storage or a message being added to a Cloud Pub/Sub topic, what is that action called?
   B: An event

4. All of the following generate events that can be triggered using Cloud Functions, except which one?
   C: SSL
   The three GCP products (Cloud Storage, Cloud Pub/Sub, and Firebase) listed do generate events that can have triggers associated with them.

5. Which runtimes are supported in Cloud Functions?
   D: Node.js, Python, Go, Java, .NET, Ruby, and PHP only

6. An HTTP trigger can be invoked by making a request using which of the following?
   C: DELETE, POST, and GET
   HTTP requests using GET, POST, PUT, DELETE, and OPTIONS can invoke an HTTP trigger in Cloud Functions.

7. What types of events are available to Cloud Functions working with Cloud Storage?
   D: Upload or finalize, delete, archive, and metadata update
   The four events supported in Cloud Storage are:
   google.storage.object.finalize
   google.storage.object.delete
   google.storage.object.archive
   google.storage.object.metadataUpdate

8. You are tasked with designing a function to execute in Cloud Functions. The function will need more than the default amount of memory and should be applied only when a finalize event occurs after a file is uploaded to Cloud Storage. The function should only apply its logic to files with a standard image file type. Which of the following required features cannot be specified in a parameter and must be implemented in the function code?
   C: File type to apply the function to
   There is no option to specify the file type to apply the function to. You can, however, specify the bucket to which the function is applied to. You could only save files or the types you want processed in that bucket, or you could have your function check file type and then execute the rest of the function or not, based on type.

9. How much memory can be allocated to a Cloud Function when using Second-Generation functions?
   D: 128 MB to 16 GB

10. How long can a Second-Generation event type Cloud Function run by default before timing out?
    B: 1 minute
    By default, Cloud Functions can run for up to 1 minute before timing out. You can, however, set the timeout parameter for a Cloud Function for periods of up to 9 minutes before timing out.

11. You want to use the command-line to manage Cloud Functions that will be written in Python. What command should you run to ensure your command-line SDK is up to date?
    A: gcloud components install

12. You want to create a Cloud Function to transform audio files into different formats. The audio files will be uploaded into Cloud Storage. You want to start transformations as soon as the files finish uploading. Which trigger would you specify in the Cloud Function to cause it to execute after the file is uploaded?
    A: google.storage.object.finalize

13. You are defining a Cloud Function to write a record to a database when a file in Cloud Storage is archived. What parameters will you have to set when creating that function?
    C: runtime, trigger-resource, trigger-event only

14. You would like to stop using a Cloud Function and delete it from your project. Which command would you use from the command-line to delete a Cloud Function?
    A: gcloud functions delete

15. You have been asked to deploy a Cloud Function to work with Cloud Pub/Sub. As you review the Python code, you notice a reference to a Python function called base64.b64decode. Why would a decode function be required in a Pub/Sub Cloud Function?
    B: Messages in Pub/Sub topics are encoded to allow binary data to be used in places where text data is expected. Messages need to be decoded to access the data in the message.
    Messages are stored in the text format, base64, so that binary data can be stored in the message in a text format.

16. Which of these commands will deploy a Python Cloud Function called pub_sub_function_test?
    C: gcloud functions deploy pub_sub_function_test --runtime python37 --trigger-topic gcp-ace-exam-test-topic
    Option C is correct because it includes the name of the function, the runtime environment, and the name of the Pub/Sub topic.

17. When specifying a Cloud Storage Cloud Function, you have to specify an event type, such as finalize, delete or archive. When specifying a Cloud Pub/Sub Cloud Function, you do not have to specify an event type. Why is this the case?
    B: Cloud Pub/Sub has triggers on only the event type, when a message is published

18. Your company has a web application that allows job seekers to upload résumé files. Some files are in Microsoft Word, some are PDFs, and others are text files. You would like to store all résumés as PDFs. How could you do this in a way that minimizes the time between upload and conversion and with minimal amount of coding?
    B: Implementing a Cloud Function on Cloud Storage to execute on a finalize event. The function checks the file type, and if it is not PDF, the function calls a PDF converter function and writes the PDF version to the bucket that has the original
    The correct answer is B because it uses a Cloud Storage finalize event to trigger conversion if needed. There is a minimal delay between the time the file is uploaded and when it is converted.

19. What are options for uploading code to a Cloud Function?
    D: All of the above (Inline editor, zip upload, and cloud source repository)

20. What type of trigger allows developers to use HTTP POST, GET, and PUT calls to invoke a Cloud Function?
    A: HTTP

Chapter 11: Planning Storage in the Cloud

When planning Cloud Storage, consider the typed of storage systems and types of data models. The storage systems provide the hardware and basic organizational structure used for storing data. The data models organize data into logical structures that determine how data is stored and queried within a database. The main storage systems available in Google Cloud are Memorystore, a managed cache service, and persistent disks, which are network-accessible disks for VMs in Compute Engine and Kubernetes Engine. Cloud Storage is Google Cloud's object storage system. The primary data models are object, relational, and NoSQL. NoSQL databases in Google Cloud are further subdivided into document and wide-column databases. Cloud Storage uses an object data model. Cloud SQL and Cloud Spanner use relational databases for transactions processing applications. BigQuery uses a relational model for data warehouse and analytic applications. Firestore is a document database. Bigtable is a wide-column table. When choosing data storage systems, consider read and write patterns, consistency requirements, transaction support, cost, and latency.

Know the major storage system types, including caches, persistent disks, and object storage. Caches are used to improve applications performance by reducing the need to read from databases on disk. Caches are limited by the amount of available memory. Persistent disks are network devices that are attached to VMs. Persistent disks may be attached to multiple VMs in read-only mode. Object storage is used for storing files for shared access and long-term storage.

Know the major kind of data models. Relational databases are used for transaction processing systems that require transaction support and strong consistency. Cloud SQL and Cloud Spanner are relational databases used for transaction processing applications. BigQuery uses an analytical model but is designed for data warehouses and analytics. The object model is an alternative to a filesystem model. Objects, stored as files, are treated as atomic units. NoSQL data models include document data models and wide-column models. Firestore is a document database. Bigtable is a wide-column database.

Know the various classes in Cloud Storage. Standard, nearline, coldline, and archive are the four storage classes. Standard is designed for data that is accessed frequently (more than once per month) or only stored in Cloud Storage for a short time. Nearline is designed for infrequent access, less than once per month. Coldline storage is designed for long-term storage, with files being accessed less than once per 90 days. Archive storage is designed for data that is not accessed more frequently than once per year. Nearline, Coldline, and Archive storage incur retrieval charges in addition to charges based on the size of the data.

Know that cloud applications may require more than one kind data store. For example, an application may need a cache to reduce latency when querying data in Cloud SQL, object storage for the long-term storage of data files, and BigQuery for data warehousing reporting and analysis.

Know that you can apply lifecycle configurations on Cloud Storage buckets. Lifecycles are used to delete files and change storage class. Standard class objects can be changed to Nearline, Coldline, or Archive. Nearline storage can change to Coldline and Archive. Coldline can be changed to Archive.

Know the characteristics of different data stores that help you determine which is the best option for your requirements. Read and write patterns, consistency requirements, transaction support, cost, and latency are often factors.

1. You are tasked with defining life cycle configuration on buckets in Cloud Storage. You need to consider all possible options for transitioning from one storage class to another. All of the following transactions are allowed except for which one?
   D: Archive to Standard
   An Archive storage class cannot be changed to Standard or to any of the other storage classes. All other options are allowed.

2. Your manager has asked for your help in reducing Cloud Storage charges. You know that some of the files stored in Cloud Storage are rarely accessed more than every 90 days. What kind of storage would you recommend for those files?
   C: Coldline
   The goal is to reduce cost, so you would want to use the least costly storage option. Coldline is designed for objects that are accessed no more than once every 90 days. Nearline and Standard costs more than Coldline, so those are not good options. Archive should only be used for objects accessed no more than once per year.

3. You are working with a startup developing analytics software for IoT data. You need to ingest large volumes of data consistently and store it for several months. The startup has several applications that will need to query this data. Volumes are expected to grow to petabyte volumes. Which database should you use?
   B: Bigtable
   … is a wide-column database that can ingest large volumes of data consistently, so answer B is correct. It also supports low-millisecond latency, making it a good choice for supporting querying. Cloud Spanner is a global rational database that is not suitable for high-speed ingestion of large volumes of data. Firestore is an object data model and not a good fit for IoT or other time series data. BigQuery is an analytics database and not designed for ingestion of large volumes of data with short write latencies.

4. A software developer on your team is asking for your help improving the query performance of a database application. The developer is using a Cloud SQL MySQL database and is willing to modify some parts of the application but wants to continue to use a relational database. Which options would you recommend?
   A: Memorystore and SSD persistent disks
   Memorystore is a managed cache. The cache can be used to store the results of queries. Follow-on queries that reference the data stored in the cache can read it from the cache, which is much faster than reading from persistent disks. SSDs have significantly lower latency than hard disk drives and should be used for performance-sensitive applications like databases.

5. You are creating a set of persistent disks to store data for exploratory data analysis. The disks will be mounted on a VM in the us-west2-a zone. The data is historical data retrieved from Cloud Storage. The data analytics do not need peak performance and are more concerned about cost than performance. The data will be stored in a local relational database. Which type of storage would you recommend?
   B: HDDs
   … are the better choice for persistent disks for a local database when performance is not the primary concern and you are trying to keep costs down, so option B is correct.

6. Which of the following statements about Cloud Storage is not true?
   B: Lifecycle configurations can be used to change storage class from Archive to Standard

7. When using versioning on a bucket, what is the latest version of the object called?
   A: Live version

8. A product manager has asked for your advice on which database services might be options for a new application. Transactions and support for tabular data are important. Ideally, the database would support common query tools. What databases would you recommend the product manager consider?
   B: Cloud SQL and Cloud Spanner
   Both Cloud SQL and Cloud Spanner are relational databases and are very well suited for transaction-processing applications.

9. The Cloud SQL service provides fully managed relational databases. What two types of databases are available in Cloud SQL?
   C: MySQL and PostgreSQL

10. Which of the following Cloud Spanner configurations would have the highest hourly cost?
    D: Located in nam-eur-asia1
    The multiregional and multi-super-regional location of nam-eur-asia1 is the most expensive.

11. Which of the following are database services that do not require you to specify configuration information for VMs?
    D: BigQuery and Firestore
    … are fully managed services that do not require you to specify configuration information for VMs. Cloud SQL and Bigtable require you to specify some configuration information for VMs.

12. What kind of data model is used in Firestore?
    B: Document

13. You have been tasked with creating a data warehouse for your company. It must support tens of petabytes of data and use SQL for query language. Which managed database service would you choose?
    A: BigQuery
    … is a managed service designed for data warehouses and analytics. It uses standard SQL for querying, which makes option A the right answer. Bigtable can support the volume of data described but does not use SQL as a query language.

14. A team of mobile developers is developing a new application. It will require synchronizing data between mobile devices and a back-end database. Which database service would you recommend?
    B: Firestore
    … is a document database that has mobile supporting features, like data synchronization.

15. A product manager is considering a new set of features for an application that will require additional storage. What feature of storage would you suggest the product manager consider?
    D: They are all relevant considerations (Read and write patterns, cost, and consistency)
    In addition to read and write patterns, cost, and consistency, you should consider transaction support and latency.

16. What is the maximum size of a Memorystore cache when using Redis?
    B: 300 GB
    Memorystore can be configured to use between 1 GB and 300 GB of memory.

17. Once a bucket has its storage class set to Archive, what are other storage classes it can transition to?
    D: None of the above (Standard, Nearline, and Coldline)
    Once a bucket is set to Archive, it cannot be changed to another storage class. Standard can change to Nearline, Coldline, and Archive. Nearline can change to Coldline and Archive. Coldline can change to Archive.

18. Before you start storing data in BigQuery, what must you create?
    A: A dataset
    To use BigQuery to store data, you must have a dataset to store it, which makes option A the right answer. Buckets are used by Cloud Storage, not BigQuery. You do not manage persistent disk when using BigQuery. An entity is a data structure in Firestore, not BigQuery.

19. What features can you configure when running a MySQL database in Cloud SQL?
    D: All of the above (Machine type, maintenance windows, and failover replicas)
    With a MySQL database, you can configure the MySQL version, connectivity, machine type, automatic backups, failover replicas, database flags, maintenance windows, and labels, so option D is correct.

20. A colleague is wondering why some storage charges are so high. They explain that they have moved all their storage to Nearline and Coldline storage and then costs increased. They routinely access most of the objects on any given day. What is one possible reason the storage costs are higher than expected?
    A: Nearline and Coldline incur access charges
    Access charges are used with Nearline and Coldline storage, which makes option A the right answer. There is no transfer charge involved.

Chapter 12: Deploying Storage in Google Cloud

In this chapter, you learned how to perform basic deployments and management tasks for a number of Google Cloud services, including Cloud SQL, Cloud Datastore, BigQuery, Bigtable, Cloud Spanner, Cloud Pub/Sub, Cloud Dataproc, and Cloud Storage. While gcloud is often used, several of the services have their own command-line tools.

Understand how to initialize Cloud SQL and Cloud Spanner. Cloud SQL and Cloud Spanner are two managed relational databases for transaction processing systems. BigQuery is an analytical database designed for data warehouse and analytics. Understand the need to create databases and tables. Know that SQL is used to query these databases.

Understand how to initialize Cloud Firestore and Cloud Bigtable. These are two NoSQL offerings. You can add small amounts of data to Cloud Firestore through the console and query it with a SQL-like language called GQL. Cloud Bigtable is a wide-column database that does not support SQL. Bigtable is managed with the cbt command-line tool.

Know how to export data from BigQuery, estimate the cost of a query, and monitor jobs in BigQuery. BigQuery is designed to work with petabyte-scale data warehouses. SQL is used to query data. Know how to export data using the console. Understand that the bq command-line, not gcloud, is the tool for working with BigQuery from the command-line.

Know how to convert Cloud Storage bucket storage classes. Life cycle policies can change storage classes of buckets when events occur, such as a period of time passes. Know that gsutil rewrite is used to change the storage class of a bucket interactively. Know how to use the console and the command-line to move and rename objects.

Understand that Pub/Sub is a message queue. Applications write data to topics, and applications receive messages through subscriptions to topics. Subscriptions can be push or pull. Unread messages have a retention period after which they are deleted.

Understand that Cloud Dataproc is a managed Spark and Hadoop service. These platforms are used for big data analytics, machine learning, and large-scale batch jobs, such as large volume extraction, transformation, and load operations. Spark is a good option for analyzing transaction data, but data must be loaded into Spark from its source system.

Know the four command-line tools: gcloud, gsutil, bq, and cbt. gcloud is used for most products but not all. gsutil and the newer gcloud storage commands are used to work with Cloud Storage from the command-line. If you want to work with BigQuery from the command-line, you need to use bq. To work with Bigtable, you use the cbt command.

1. Cloud SQL is a fully managed relational database service, but database administrators still have to perform some tasks. Which of the following tasks do Cloud SQL users need to perform?
   C: Creating databases
   … is the responsibility of database administrators or other users of Cloud SQL, so option C is correct. Goggle applies security patches and performs other maintenance. Google Cloud performs regularly scheduled backups. Database administrators need to schedule backups, but Google Cloud makes sure they are performed on schedule. Cloud SQL users can not use SSH to connect to a Cloud SQL server, so they can not tune the operating system. That is not a problem; Google takes care of that.

2. Which of the following commands is used to create a backup of a Cloud SQL database?
   A: gcloud sql backups create
   Cloud SQL is controlled using the gcloud command; the sequence of terms in gcloud commands is gcloud followed by the service, in this case SQL; followed by a resource, in this case backups; and a command or verb, in this case create.

3. Which of the following commands will run an automatic backup at 3:00 a.m on an instance called ace-exam-mysql?
   A: gcloud sql instances patch ace-exam-mysql --backup-start-time 03:00

4. What is the query language used by Firestore in Datastore mode?
   C: GQL
   Datastore mode uses a SQL-like query language called GQL.

5. What is the correct command-line structure to export data from Firestore?
   C: gcloud firestore export gs://[BUCKET_NAME]
   Option C is the correct command. Is has the correct base command, gcloud firestore, followed by the name of a Cloud Storage bucket to hold the export file.

6. When you enter a query into the BigQuery query form, BigQuery analyzes the query and displays an estimate of what metric?
   C: Amount of data scanned
   BigQuery displays an estimate of the amount of data scanned. This is important because BigQuery charges for data scanned in queries.

7. You want to get an estimate of the volume of data scanned by BigQuery from the command line. Which option shows the command structure you should use?
   B: bg --location=LOCATION] query --use_legacy_sql=false --dry_run [SQL_QUERY]
   Option B shows the correct bq command structure, which includes location and the --dry_run option. This option calculates an estimate without actually running the query.

8. You are using Cloud Console and want to check on some job running in BigQuery. You navigate to the BigQuery part of the console. Which menu item would you click to view jobs?
   A: Personal History or Project History

9. You want to estimate the cost of running a BigQuery query. What two services within Google Cloud will you need to use?
   C: BigQuery and Pricing Calculator
   BigQuery provides an estimate of the amount of data scanned, and the Pricing Calculator gives a cost estimate for scanning that volume of data.

10. You have just created a Cloud Spanner instance. You have been tasked with creating a way to store data about a product catalog. What is the next step after creating a Cloud Spanner instance that you would perform to enable you to load data?
    B: Create a database within the instance
    Option B is correct; the next step is to create a database within the instance. Once a database is created, tables can be created, and data can be loaded into tables.

11. You have created a Cloud Spanner instance and database. According to Google best practices, how often should you update VM packages using apt-get?
    D: Never; Cloud Spanner is a managed service

12. Your software team is developing a distributed application and wants to send messages from one application to another. Once the consuming application reads a message, is should be deleted. You want your system to be robust to failure, so messages should be available for at least three days before they are discarded. Which Google Cloud service is the best designed to support this use case?
    C: Cloud Pub/Sub
    This use case is well suited to Pub/Sub, so Option C is correct. It involves sending messages to the topic, and the subscription model is a good fit. Pub/Sub has a retention period to support the three-day retention period.

13. Your manager asks you to set up a bare-bones Pub/Sub system as a sandbox for new developers to learn about messaging systems. What are the two resources within Pub/Sub you will need to create?
    C: Topics and subscriptions
    Pub/Sub works with topics, which receive and hold messages, and subscriptions, which make messages available to consuming applications; therefore, option C is correct.

14. Your company is launching an IoT service and will receive large volumes of streaming data. You have to store this data in Bigtable. You want to explore the Bigtable environment from the command-line. What command would you run to ensure you have command-line tools installed?
    C: gcloud components install cbt

15. You need to create a table called iot-ingest-data in Bigtable. What command would you use?
    A: cbt createtable iot-ingest-data
    You would need to use a cbt command, which is the command-line tool for working with Bigtable.

16. Cloud Dataproc is a managed service for which two big data platforms?
    B: Spark and Hadoop

17. Your department has been asked to analyze large batches of data every night. The jobs will run for about three to four hours. You want to shut down resources as soon as the analysis is done, so you decide to write a script to create a Dataproc cluster every night at mighnight. What command would you use to create a cluster called spark-nightly-analysis in the us-west2-a zone?
    B: gcloud dataproc clusters create spark-nightly-analysis --zone us-west2-a

18. You have a number of buckets containing old data that is hardly ever used. You do not want to delete it, but you want to minimize the cost of storing it. You decide to change the storage class to Coldline for each of those buckets. What is the command structure that you would use?
    B: gsutil rewrite -s [STORAGE_CLASS] gs://[PATH_TO_OBJECT]

19. You want to rename an object stored in a bucket. What command structure would you use?
    B: gsutil mv gs://[BUCKET_NAME]/[OLD_OBJECT_NAME]
    gs://[BUCKET_NAME]/[NEW_OBJECT_NAME]

20. An executive is your company emails you asking about creating a recommendation system that will help sell more products. The executive has heard there are some Google Cloud solutions that may be good fits for this problem. What Google Cloud service would you recommend the executive look into?
    A: Cloud Dataproc, especially Spark and its machine learning library

Chapter 13: Loading Data into Storage

Cloud Storage is organized around object in buckets. The gsutil command and Cloud Console can be used to upload data as well as move it between buckets. The database services provide import and export utilities. Each supports a variety of file formats. Cloud Pub/Sub can be used to decouple applications and improve resiliency to spikes in load. Know that Cloud Spanner uses the Dataflow service for importing and exporting. There can be additional charges when using Dataflow and moving data between regions.

Know how to load data into and move data around Cloud Storage. Cloud Storage is widely used for a variety of use cases, including long-term storage and archiving, file transfers, and data sharing. Understand the structure of gsutil commands, which is different from gcloud. gsutil commands start with gsutilfollowed by an operation, such as copy or make bucket. Be sure to know the syntax of the copy (cp), move (mv), and make bucket (mb) commands. You can copy files from Cloud Storage to VMs, and vice versa. Also, know that the gsutil acl ch -u command is used to change permissions on objects. You can use the gsutil acl ch command to change permissions on a Cloud Storage bucket.

Understand how import and export work with Cloud SQL. Importing and exporting data from databases are common operations. You can perform imports and exports from the console and from the command-line.

Know that you can export entities from a Cloud Firestore. Exports and imports are done at the database level when in Native mode and at the level of namespaces when the database is in Datastore mode.

Understand how to export and import data from BigQuery. BigQuery has a range of options for the source of data to import. Data can be compressed when exported to save on space. BigQuery can export data in multiple formats, including CSV, JSON, and Avro. Know that the bq command is used for importing and exporting from the command-line.

Know that Pub/Sub is used to send messages between services. Pub/Sub allows for greater resiliency to fluctuations in load. If one service lags, its work can accumulate in a Pub/Sub queue without forcing the service that generates that data to wait.

1. Which of the following commands is used to create buckets in Cloud Storage?
   C: gsutil mb
   gsutil is the command-line utility for working with Cloud Storage.

2. You need to copy files from your local device to a bucket in Cloud Storage. What command would you use? Assume you have Cloud SDK installed on your local computer.
   B: gsutil cp

3. You are migrating a large number of files from a local storage system to Cloud Storage and want to use the Cloud Console instead of writing a script. Which of the following Cloud Storage operations can you perform in the console?
   C: Upload files and folders

4. A software developer asks for your help exporting data from a Cloud SQL database. The developer tells you which database to export and which bucket to store the export file in but has not mentioned which file format should be used for the export file. What are the options for the export file format?
   D: CSV and SQL

5. A database administrator has asked for an export of a MySQL database in Cloud SQL. The database administrator will load the data into another relational database and would like to do it with the least amount of work. Specifically, the loading method should not require the database administrator to define a schema. What file format would you recommend for this task?
   A: SQL
   SQL-format exports a database as a series of SQL data definition commands. These commands can be executed in another relational database without having to first create a schema.

6. Which command will export a MySQL database called ace-exam-mysql1 to a file called ace-exam-mysql-export.sql in a bucket named ace-exam-bucket1?
   C: gcloud sql export sql ace-exam-mysql gs://ace-exam-bucket1/ace-exam-mysql-export.sql --database=mysql

7. Which of the following file formats is not an option for an export when exporting from BigQuery?
   B: XML

8. Which of the following file formats is not supported when importing data into BigQuery?
   D: YAML

9. You have received a large data set from an Internet of Things (IoT) system. You want to use BigQuery to analyze the data. What command-line command would you use to make data available for analysis in BigQuery?
   A: bg load --autodetect --source_format=[FORMAT] [DATASET].[TABLE] [PATH_TO_SOURCE]

10. You have set up a Cloud Spanner process to export data to Cloud Storage. You notice that each time the process runs you incur charges for another Google Cloud service, which you think is related to the export process. What other Google Cloud service might be incurring your part?
    B: Dataflow
    … is a pipeline service for processing streaming and batch data that implements workflows used by Cloud Spanner.

11. As a developer on a project using Bigtable for an IoT application, you will need to export data from Bigtable to make some data available for analysis with another tool. What would you use to export the data, assuming you want to minimize the amount of effort required on your part?
    A: A Java program designed for importing and exporting data from Bigtable
    Bigtable is exported using a compiled Java program, so option A is correct.

12. You have just exported from a Dataproc cluster. What have you exported?
    C: Configuration data about the cluster
    Exporting from Dataproc exports data about the cluster configuration, which makes option C correct.

13. A team of data scientists has requested access to data stored in Bigtable so that they can train machine learning models. They explain that Bigtable does not have the features required to build machine learning models. Which of the following Google Cloud services are they most likely to use to build machine learning models?
    C: Dataproc
    … supports Apache Spark, which has libraries for machine learning.

14. Which of the following is the correct command to create a Pub/Sub topic?
    A: gcloud pubsub topics create

15. Which of the following commands will create a subscription on the topic ace-exam-topic1?
    C: gcloud pubsub subscriptions create --topic=ace-exap-topic1 ace-exam-sub1

16. What is one of the direct advantages of using a message queue in distributed systems?
    B: It decouples services, so if one lags, it does not cause other services to lag
    Using a message queue between services decouples the services, so if one lags it does not cause other services to lag, which makes option B correct.

17. To ensure you have installed beta gcloud commands, which command should you run?
    B: gcloud components install beta

18. What parameter is used to tell BigQuery to automatically detect the schema of a file on import?
    A: autodetect

19. The compression options Deflate and Snappy are available for what file types when exporting from BigQuery?
    A: Avro
    … supports Deflate and Snappy compressions. CSV supports Gzip and no compression. XML and Thrift are not export file type options.

20. You want to read a message from a Pub/Sub topic and acknowledge reading that message in the same command. Which of the following would you use?
    A: gcloud pubsub subcriptions pull --auto-ack

Chapter 14: Networking in the Cloud: Virtual Private Clouds and Virtual Private Networks

VPCs define networks in the Google Cloud to link your Google Cloud resources. VPNs in Google Cloud are used to link your Google Cloud networks to your internal networks.

Know that VPCs are logical data centers in the cloud and that VPNs are secure connections between your VPC subnets and your internal network. Your cloud resources are in a VPC. VPCs have subnets and routing rules for routing traffic between subnets. You control the flow of traffic using firewall rules.

Know that VPCs create subnets in each region when in auto mode. You can create additional subnets. Each subnet has a range of IP addresses. Firewall rules are applied to subnets, also called networks. Routers can be configured to learn just regional routes or global routes.

Understand how to read and calculate CIDR notation. CIDR notation represents a subnet mask and the size of available IP address in the IP range. The smaller the subnet mask size, which is the number after the slash in CIDR block, the more IP addresses are available. The format of the CIDR address is an IP address followed by a slash, followed by the size of the subnet mask, such as 10.0.0.0/8.

Know that VPCs can be created using gcloud commands. A VPC can be created with gcloud compute networks create. A shared VPC can be created using gcloud beta compute shared-vpc. Shared VPCs can be shared at the network or the folder level. You will need to bind identity and access management (IAM) policies at the organizational or folder level to enable Shared VPC Admin roles. VPC peering can be used for interproject connectivity.

Understand that you can add network interfaced to a VM. You can configure these interfaces to use a particular subnet. You can assign ephemeral or static IP addresses.

Know that firewall rules control the flow of network traffic. Firewall rules consist of direction, priority, action, target, source/destination, protocols and port, and enforcement status. Firewall rules are applied to a subnet.

Know how to create a VPN with Cloud Console. VPNs route traffic between your cloud resources and your internal network. VPNs include gateways, forwarding rules, and tunnels. Both Classic and High Availability (HA) VPNs are available.

1. What kind of resources are VPCs in Google Cloud?
   D: Global
   VPS are global. By default, they have subnets to all regions. Resources in any region can be accessed through the VPC.

2. You have been tasked with defining CIDR ranges to use with a project. The project includes two VPCs with several subnets in each VPS. How many CIDR ranges will you need to define?
   B: One for each subnet
   IP ranges are assigned to subnets. Each subnet is assigned an IP range for its exclusive use. IP ranges are assigned network structures, not zones and regions. VPCs can have multiple subnets but each subnet has its own address range.

3. The legal department needs to isolate its resources on its own VPC. You want to have the network provide routing to any other service available on the global network. The VPC network has not learned global routes. What parameter may have been missed when creating the VPC subnet?
   B: Dynamic routing
   … is the parameter that specifies whether routes are learned regionally or globally.

4. The command used to create a VPC from the command-line is:
   A: gcloud compute networks create

5. You have created several subnets. Most of them are sending logs to Cloud Logging. One subnet is not sending logs. What option may have been misconfigured when creating the subnet that is not forwarding logs?
   A: Flow Logs
   The Flow Logs option of the create vpc command determines whether logs are sent to Cloud Logging, so option A is correct.

6. At what levels of the resource hierarchy can a shared VPC be created?
   C: Organizations and folders
   Shared VPCs can be created at the organization or folder level of the resource hierarchy, so option C is correct.

7. You are using Cloud Console to create a VM that you want to exist in a custom subnet you just created. What section of the Create Instance page would you use to specify the custom subnet?
   A: Networking tab of the Management, Security, Disks, Networking, Sole Tenancy section
   The correct answer is the Networking tab of the Management, Security, Disks, Networking, Sole Tenancy section of the page, which makes option A correct.

8. You want to implement interproject communication between VPCs. Which feature of VPCs would you use to implement this?
   A: VPC network peering

9. You want to limit traffic to a set of instances. You decide to set a specific network tag on each instance. What part of a firewall rule can reference the network tag to determine the set of instances affected by the rule?
   B: Target
   The target can be all instances in a network, instances with network tags, or instances using a specific service account, so option B is correct.

10. What part of a firewall rule determines whether a rule applies to incoming or outgoing traffic?
    D: Direction
    … specifies whether the rule is applied to incoming or outing traffic, which makes option D the right answer.

11. You want to define a CIDR range that applies to all destination addresses. What IP address would you specify?
    A: 0.0.0.0/0

12. You are using gcloud to create a firewall rule. Which command would you use?
    B: gcloud compute firewall-rules create

13. You are using gcloud to create a firewall rule. Which parameter would you use to specify the subnet it should apply to?
    B: --network

14. An application development team is deploying a set of specialized service endpoints and wants to limit traffic so that only traffic going to one of the endpoints is allowed trough by firewall rules. The service endpoints will accept any UDP traffic, and each endpoint will use a port in the range of 20 000-30 000. Which of the following commands would you use?
    A: gcloud compute firewall-rules create fwr1 --allow=udp:20000-30000 --direction=ingress

15. You have a rule to allow inbound traffic to a VM. You want it to apply only if there is not another rule that would deny that traffic. What priority should you give this rule?
    D: 65635
    Option D is correct because it is the largest number allowed in the range of values for priorities. The larger the number, the lower the priority. Having the lowest priority will ensure that other rules that match will apply.

16. You want to create a VPC using Cloud Console. What section of Cloud Console should you use?
    C: Hybrid Connectivity

17. Your company needs to ensure they have at least a 99.99 percent availability SLA for networking between on-premises network and a VPC in Google Cloud. What should you use to ensure you have this level of availability?
    B: HA VPN
    … is a VPN that can provide a 99.99 percent availability SLA and connect on-premises networks to Google Cloud.

18. You want the router on a tunnel you are creating to learn routes from all Google Cloud regions on the network. What feature of Google Cloud routing would you enable?
    A: Global dynamic routing
    … is used to learn all routes on a network.

19. What gcloud command would you use to create tunnels for a VPN?
    B: gcloud compute vpn-tunnels create

20. You are using gcloud to create a VPN. Which command(s) would you use?
    D: gcloud compute forwarding-rile,
    gcloud compute target-vpn-gateways, and
    gcloud compute vpn-tunnels

Chapter 15: Networking in the Cloud: DNS, Load Balancing, Google Private Access, and IP Addressing

The Associate Cloud Engineer exam may test your knowledge of Cloud DNS, load balancing, and managing IP addresses. Cloud DNS is an authoritative name service for mapping ping domain names to IP addresses. You can set up public or private DNS zones. You will also need to be familiar with load balancing and the different types of load balancers. Some load balancers are regional, and some are global. Some are for internal use only, and others support external sources of traffic.

Understand that Cloud DNS is used to map domain names to IP addresses. I you want to support queries from the Internet, use a public DNS zone. Use a private DNS zone only if you want to accept queries from resources in your project.

Know that DNS entries, like example.com, can have multiple records associated with them. The A record specifies the address of a DNS resolver that maps domain names to IP addresses. CNAME records store the canonical name of the domain.

Know how load balancers are distinguished. Load balancers are distinguished based on global versus regional load balancing, external versus internal load balancing, and the protocols supported. Global balancers distribute load across regions, whereas regional load balancers work within a region. Internal load balancers balance traffic only from within Google Cloud, not external sources. Some load balancers are protocol-specific, such as HTTP and SSL load balancers.

Know the types of load balancers and when they should be used. HTTP(S), SSL, Proxy, TCP Proxy, and TCP/UDP. Load balancers distribute load regionally or globally. Internal load balancers distribute load from internal traffic. External load balancers distribute load from external traffic.

- HTTP(S) balances HTTP and HTTPS load
- SSL Proxy terminates SSL/TLS connections
- TCP Proxy terminates TCP sessions
- TCP/UDP balances TCP/UDP traffic on private networks hosting internal VMs

Understand that configuring a load balancer can require configuring both the front end and back end. The network load balancer can be configured by specifying a forwarding rule that routes traffic to the load balancer to VMs in the target pool.

Know Google Private Access options. Private Google Access is used for private access to most Google Cloud Services, while Private Service Access is used with third-party services and some Google Cloud service. Private Service Connect uses a VPC endpoint for forwarding traffic to Google Cloud services. Serverless VPC Access allows Cloud Run, Cloud Functions, and App Engine Standard to reach VMs with private addresses.

Know how to increase the number of IP addresses in a subnet. Use the gcloud compute network subnets expand-ip-range command to increase IP addresses in a subnet. The number of addresses can only increase. The expand-ip-range command cannot be used to decrease the number of addresses.

Know how to reserve an IP address using the console and the gcloud compute address create command. Reserved IP addresses continue to be available to your project even if they are not attached to a resource. Know the difference between Premium and Standard tier network services.

1. What record type is used to specify the IPv4 address of a domain?
   B: A
   The A record is used to map a domain name to an IPv4 address, so option B is correct. The AAAA record is used for IPv6 addresses.

2. The CEO of your startup just read a news report about a company that was attacked by something called cache poisoning. The CEO wants to implement additional security measures to reduce the risk of DNS spoofing and cache poisoning. What would you recommend?
   A: Using DNSSEC
   DNSSEC is a secure protocol designed to prevent spoofing and cache poisoning, so option A is correct.

3. What do the TTL parameters specify in a DNS record?
   A: Time a record can exist in a cache before it should be queried again

4. What command is used to create a DNS zone in the command-line?
   B: gcloud dns managed-zones create

5. What parameter is used to make a DNS zone private?
   B: --visibility=private

6. Which load balancers provide global load balancing?
   C: Global External HTTP(S) Load Balancing, Global External HTTP(S) Load Balancing (classic), SSL Proxy, and TCP Proxy

7. Which regional load balancer balances HTTP(S) regionally on Premium tier networking only?
   D: Internal HTTP(S) Load Balancing
   … distributes traffic regionally on Premium tier networking.

8. You are configuring a load balancer and want to implement private load balancing. Which option would you select?
   A: Only Between My VMs
   In the console there is an option to select between From Internet To My VMs and Only Between My VMs. This is the option to indicate private or public.

9. What two components need to be configured when creating a TCP Proxy load balancer?
   B: Front end and back end

10. A health check is used to check what resources?
    B: VMs
    Health checks monitor the health of VMs used with load balancers, so option B is correct.

11. Where do you specify the ports on a TCP Proxy load balancer that should have their traffic forwarded?
    B: Front end
    You specify ports to forward when configuring the frond end. The back end is where you configure how traffic is routed to VMs.

12. What command is used to create a network load balancer at the command-line?
    A: gcloud compute forwarding-rules create

13. A team is setting up a web service for internal use. The want to use the same IP address for the foreseeable future. What type of IP address would you assign?
    C: Static
    Static addresses are assigned until they are released. Internal and external addresses determine whether traffic is routed into and out of the subnet. External addresses can have traffic reach them from the Internet; internal addresses cannot.

14. You are starting up a VM to experiment with a new Python data science library. You will use SSH to connect to the VM, use the Python interpreter interactively for a while, and then shut down the machine. What type of IP address would you assign to the VM?
    A: Ephemeral
    An ephemeral address is sufficient since resources outside the subnet will not need to reach the VM and you can SSH into the VM from the console.

15. You have created a subnet called sn1 using 192.168.0.0 with 65.534 addresses. You realized that you will not need that many addresses, and you would like to reduce that number to 254. Which of the following commands would you use?
    D: There is no command to reduce the number of IP addresses available
    You cannot reduce the number of addresses using any of the commands. A new subnet with less addresses must be created instead.

16. You have created a subnet called sn1 using 192.168.0.0. You want to have 14 addresses. What prefix length would you use?
    B: 28
    The prefix length specifies the length in bits of the subnet mask. The remaining bits of the IP address are used for device addresses. Since there are 32 bits in an IP address, you subtract the length of the mask to get the number of bits used to represent the address. 16 is equal to 2*, so you need 4 bits to represent 14 addresses. 32 – 4 is 28, so option B is correct.

17. You want all your network traffic to route over the Google network and not traverse the public Internet. What level of network service should you choose?
    C: Premium
    … is the network service level that routes all traffic over the Google network.

18. You have a website hosted on a Compute Engine VM. Users can access the website using the domain name you provided. You do some maintenance work on the VM and stop the server and restart it. Now users cannot access the website. No other changes have occurred on the subnet- What might be the cause of the problem?
    B: You used an ephemeral instead of a static IP address
    Stopping and starting a VM will release ephemeral IP addresses, so option B is correct. Use a static IP address to have the same IP address across reboots.

19. You are deploying a distributed system. Messages will be passed between Compute Engine VMs using a reliable UDP protocol. All VMs are in the same region. You want to use the load balancer that best fits these requirements. Which kind of load balancer would you use?
    A: Internal TCP/UDP
    … is a good option. It is a regional load balancer that supports UDP.

20. You want to use Cloud Console to review the records in a DNS entry. What section of Cloud Console would you navigate to?
    B: Network Services

Chapter 16: Deploying Applications with Cloud Marketplace and Cloud Foundation Toolkit

Cloud Marketplace and Cloud Deployment Manager are designed to make it easy to deploy resources in Google Cloud. Cloud Marketplace is where third-party vendors can offer deployable applications based on proprietary or open source software. When an application is deployed from Cloud Marketplace, resources such as VMs, storage buckets, and persistent disks are created automatically without additional human intervention. Deployment Manager gives cloud engineers the ability to define configuration files that describe the resources they would like to deploy. Cloud engineers can then use gcloud commands to deploy the resources and list their status. Deployment Manager is especially useful in organizations where you want to easily deploy resources without requiring users of those resources to understand the details of how to configure Google Cloud resources. The Cloud Foundation Toolkit provides templates and blueprints that encode best practices for deploying solutions and individual resources to Google Cloud. The Configuration Connector add-on to Kubernetes allows you to manage Google Cloud resources using Kubernetes.

Understand how to browse for solutions using the Cloud Marketplace section of Cloud Console. You can use filters to narrow your search to specific kinds of solutions, such as operating systems and developer tools. There may be multiple options for a single application, such as WordPress. This is because multiple vendors provide configurations. Review the description of each to understand which best fits your needs.

Know how to deploy a solution in Cloud Marketplace. Understand how to configure a Cloud Marketplace deployment in Cloud Console. Understand that when you launch a solution, you may be prompted for application-specific configurations. For example, with WordPress you may be prompted to install phpMyAdmin. You may also have the opportunity to configure common configuration attributes, such as the machine type and boot disk type.

Understand how to use the Deployment Manager section of the console to monitor deployment. It may be a few minutes from the time you launch a configuration to the time it is ready to use. Note that once the application is ready, you may be prompted for additional information, such as a username and password to log in.

Know that Deployment Manager is a Google Cloud service for creating configuration files that define resources to use with an application. These configuration files use YAML syntax. They are made up of resource specifications that use key-value pairs to define properties of the resource.

Know that resources in a configuration file are defined using a name, type, and set of properties. The properties vary by type. The machine type can be defined using just a URL that points to a type of machine available in a region. Disks have multiple properties, including a device name, a type, and whether the disk is a boot disk.

Know that you can use templates with configuration files. If your configuration files are getting long or complicated, you can modularize them using templates. Templates define resources and can be imported into other templates. Templates are text files written in Jinja2 or Python.

Know how to launch a deployment configuration file using gcloud deployment-manager deployment create. You can review the status of a deployment using gcloud deployment-manager deployments-describe.

Know the purpose of Cloud Foundation Toolkit and Config Connector. Cloud Foundation Toolkit is an open source project with blueprints and example configurations that capture Google Cloud-recommended best practices for deploying solutions. Config Connector is a Kubernetes add-on for managing Google Cloud resources from Kubernetes.

1. What are the categories of Cloud Marketplace solutions?
   D: Data sets, operating systems, and developer tools

2. You want to use Terraform for managing infrastructure as code and you would also like to follow Google Cloud-recommended best practices. What would you use to start implementing such a solution?
   B: Cloud Foundation Toolkit

3. Where do you navigate to launch a Cloud Marketplace solution?
   A: Overview page of the solution

4. You want to quickly identify the set of operating systems available in Cloud Marketplace. Which of these steps would help with that?
   B: Use filters in Cloud Marketplace

5. You want to use Cloud Marketplace to deploy a WordPress site. You notice there is more than one WordPress option. Why is that?
   B: Multiple vendors may offer the same application

6. You have used Cloud Marketplace to deploy a WordPress site and would now like to deploy a database. You notice that the configuration page for the database is different from the one used with WordPress. Why is that?
   C: Configuration properties are based on the application you are deploying and will be different depending on what application you are deploying

7. You have been asked by your manager to deploy a WordPress site. You expect heavy traffic, and your manager wants to make sure the VM hosting the WordPress site has enough resources. Which resources can you configure when launching a WordPress site using Cloud Marketplace?
   D: All of the above (Machine type, disk type, and disk size)

8. You would like to define as code the configuration of a set of application resources. What is the Google Cloud service for creating resources using a configuration file made up of resource specifications defined in YAML syntax?
   B: Cloud Deployment Manager

9. What file format is used to define Cloud Deployment Manager configuration files?
   D: YAML

10. A Deployment Manager configuration file starts with what word?
    B: resources

11. Which of the following are used to define a resource in a Cloud Deployment Manager configuration file?
    D: Type, properties, and name

12. What properties may be set when defining a disk on a VM?
    D: A device name, a Boolean indicating a boot disk, and a Boolean indicating autodelete

13. You need to look up what images are available in the zone in which you want to deploy a VM. What command would you use?
    A: gcloud compute images list

14. You want to use a template file with Cloud Deployment Manager. You expect the file to be complicated. What language would you use?
    D: Python

15. What command launches a deployment?
    A: gcloud deployment-manager deployments create

16. A DevOps engineer is noticing a spike in CPU utilization on your servers. You explain that you have just launched a deployment. You would like to show the DevOps engineer the details of a deployment you just launched. What command would you use?
    C: gcloud deployment-manager deployments describe

17. If you expand the More link in the Networking section when deploying a Cloud Marketplace solution, what will you be able to configure?
    A: IP addresses

18. What are the license types referenced in Cloud Marketplace?
    D: Free, flat hourly, usage fee, and bring your own license (BYOL)

19. Which license type will add charges to your Google Cloud bill when using Cloud Marketplace with this type of license?
    B: Flat hourly and usage fees

20. You are deploying a Cloud Marketplace application that includes an LAMP stack. What software will this deploy?
    D: Apache, MySQL, Linux, and PHP

Chapter 17: Configure Access and Security

Access controls in Google Cloud are managed using IAM, basic roles, and scopes. The three basic roles are Owner, Editor, and Viewer. They provide coarse-grained access controls to resources. Scopes are access controls that apply to instances of VMs. They are used to limit operations that can be performed by an instance. The set of operations that an instance can perform is determined by the scopes assigned and the roles assigned to a service account used by the instance. IAM provides predefined roles. These roles are groups by service. The roles are designed to provide the minimal set of permissions needed to carry out a logical task, such as writing to a bucket or deploying an App Engine application. When predefined roles do not meet your needs, you can define custom roles. Service accounts are used to enable VMs to perform operations with a set of permissions. The permissions are granted to service accounts through the roles assigned to the service account. You can use the default service account provided by Google Cloud for an instance or you can assign your own.

Know the three types of roles: basic, predefined, and custom. Basic roles include Owner, Editor, and Viewer. These were developed prior to the release of IAM. Predefined roles are IAM roles. Permissions are assigned to these roles, and then the roles are assigned to users, groups, and service accounts. Custom roles include permissions selected by the user creating the custom role.

Understand that scopes are a type of access control applied to VM instances. The VM can only perform operations allowed by scopes and IAM roles assigned to the service account of the instance. You can use IAM roles to constrain scopes and use scopes to constrain IAM roles.

Know how to view roles assigned to identities. You can use the Roles tab in the IAM & Admin section of the console to list the identities assigned particular roles. You can also use the gcloud projects get-iam-policy command to list roles assigned to users in a project.

Understand that IAM roles support separation of duties and the principle of least privilege. Basic roles did not support least privilege and separation of duties because they are too coarse-grained. Separation of duties ensures that two or more people are required to complete a sensitive task.

Know how to use gcloud iam roles describe to view details of a role, including permissions assigned to a role. You can also view roles users have been granted by drilling down into a role in the Roles pages of the IAM & Admin section of the console. When working with IAM, you will be using the gcloud command when working from the command-line.

Understand the different options for accessing scopes when creating an instance. The options are Default Access, Full Access, and Set Access For Each API. If you aren't sure which to use, you can grant full access, but be sure to limit what the instance can do by assigning roles that constrain allowed operations.

Know that Cloud Logging collections logging events. They can be filtered and displayed in the Logging section of Cloud Console. You can filter by resource, type of log, log level, and period of time to display.

1. What does IAM stand for?
   B: Identity and access management

2. When you navigate to IAM & Admin in Cloud Console, what appears in the main body of the page?
   A: Members and roles assigned

3. Why are basic roles classified in a category in addition to IAM?
   B: They were created before IAM

4. A developer intern is confused about what roles are used for. You describe IAM roles as a collection of what?
   B: Permissions

5. You want to list roles assigned to users in a project called ace-exam-project. What gcloud command would you use?
   C: gcloud projects get-iam-policy ace-exam-project

6. You are working in the form displayed after clicking the Add link on the IAM page of IAM & Admin in Cloud Console. There is a field called New Members. What items would you enter in that parameter?
   B: Individual users or groups

7. You have been assigned the App Engine Deployer role. What operations can you perform?
   D: Read application configuration and settings and write new versions.

8. You want to list permissions in a role using Cloud Console. Where would you go to see that?
   B: IAM & Admin; select Roles. Check the box next to a role to display the permissions in that role

9. You are meeting with an auditor to discuss security practices in the cloud. The auditor asks how you implement several best practices. You describe how IAM predefined roles help to implement which security best practice(s)?
   D: Option A and B (Least privilege, and separation of duties)

10. What launch stages are available when creating custom roles?
    D: Alpha, beta, general availability, and disabled

11. What is the gcloud command used to create a custom role?
    B: gcloud iam roles create

12. A DevOps engineer is confused about the purpose of scopes. Scopes are access controls that are applied to what kind of resources?
    B: VM instances

13. A scope is identified using what kind of identifier?
    C: A URL beginning with www.googleapis.com/auth

14. A VM instance is trying to read from a Cloud Storage bucket. Reading the bucket is allowed by IAM roles granted to the service account of the VM. Reading buckets in denied by the scopes assigned to the VM. What will happen if the VM tries to read from the bucket?
    C: The read will not execute because both scopes and IAM roles are applied to determine what operations can be performed

15. What are the options for setting scopes in a VM?
    B: Allow Default Access, Allow Full Access, and Set Access For Each API

16. What gcloud command would you use to set scopes?
    B: gcloud compute instances set-service-account

17. What gcloud command would you use to assign a service account when creating a VM?
    A: gcloud compute instances create [INSTANCE_NAME] --service-account [SERVICE_ACCOUNT_EMAIL]

18. What Google Cloud service is used to view audit logs?
    C: Cloud Logging

19. What options are available for filtering log messages when viewing audit logs?
    B: Resource, type of log, log level, and period of time only

20. An auditor needs to review audit logs. You assign read-only permission to a custom role you create for auditors. What security best practice are you following?
    B: Least privilege

Chapter 18: Monitoring, Logging, and Cost Estimating

As a cloud engineer, you are responsible for monitoring the health and performance of applications and clous services. Google Cloud provides multiple tools, including monitoring, logging, and tracing services. Cloud Monitoring allows you to define alerts on metrics, such as CPU utilization, so that you can be notified if part of your infrastructure is not performing as expected. Cloud Logging collects, stores, and manages log entries. Logs can be store in Cloud Logging-provided buckets or user-defined buckets. Log messages can be routed to Cloud Storage, BigQuery, or Cloud Pub/Sub. Cloud Trace provides distributed tracing services to identify slow-running parts of code. You can always get the status of Google Cloud services at the Google Cloud Status Dashboard at https://status.cloud.google.com. The Pricing Calculator is designed to help you estimate the cost of services in the Google Cloud. It is available at https://cloud.google.com/products/calculator.

Understand the need for monitoring and the role of metrics. Metrics provide data on the state of applications and infrastructure. You create conditions, like CPU exceeding 80 percent for 5 minutes, to trigger alerts. Alerts are delivered by notification channels. Google Cloud has a substantial number of predefined metrics, but you can create custom metrics as well.

Know how to collect, store, filter, and display log data using Cloud Logging. Logs can come from virtually any source. Logging keeps log data in the Default bucket for 30 days unless a custom retention policy is specified. If you need to keep log data longer than that, you need to export the data to a log sink. Log sinks may be a Cloud Storage bucket, a BigQuery data set, or a Cloud Pub/Sub topic.

Know how to filter logs. Logs can contain a large amount of data. Use filters to search for text or labels, limit log entries by log type and severity, and restrict the time range to a period of interest.

Log entries are hierarchical. Cloud Logging shows a single-line summary for a log entry by default, but you can drill down into the details of a log entry. Use the Expand All and Collapse All options to quickly view or hide the full details of a log entry.

Know how to use the Cloud Trace distributed tracing service. Software developers include Cloud Trace code in their applications to record trace data. Trace data can be viewed as individual traces, or you can create reports that include parameters specifying a subnet of traces you want to include.

Know where Google Cloud publishes the status of services. The Google Cloud Status page includes a list of all services, their current status, and the status over the near past. If there is an incident in a service, you will find additional details on the impact and root cause of the problem.

Know how to use the Pricing Calculator to estimate the cost of resources and services in the Google Cloud. The calculator is available at https://cloud.google.com/products/calculator. There is a separate calculator for each service. Each service has its own set of parameters for estimating costs. The Pricing Calculator allows you to estimate the cost of multiple services and generate a total estimate for all those services.

1. What Cloud Operations service is used to generate alerts when the CPU utilization of a VM exceeds 90 percent?
   B: Cloud Monitoring
   The Monitoring service is used to set a threshold on metrics and generate alerts when a metric exceeds a threshold for a specified period of time.

2. You have just created a virtual machine, and you would like to collect detailed metrics about the VM. What do you need to do to the VM to have this happen?
   B: Install the Ops Agent on the VM
   You would install the Ops Agent on the VM. The agent will collect data and send it to Cloud Monitoring and Cloud Logging.

3. Where can Cloud Monitoring be used to monitor resources?
   D: In Google Cloud, Amazon Web Services, and on-premises data centers

4. You are responsible for the reliability and availability of several services running in Kubernetes Engine. You have determined that you need to monitor several metrics to get information on the state of the services. You would like to see all of these metrics displayed as line charts, one for each metric. All of the line charts should be available on a single-page view. What would you use to create such a page view?
   A: Cloud Monitoring Dashboard

5. You have created a condition of CPU utilization, and you want to receive notifications. Which of the following are options?
   D: Email, PagerDuty, and Webhooks
   PagerDuty is a popular DevOps tool.

6. When you create a policy to notify you of a potential problem with your infrastructure, you can specify optional documentation. Why would you bother putting documentation in that form?
   D: Option B and C (It can help you or a colleague understand the purpose of the policy, and it can contain information that would help someone diagnose and correct the problem)
   The documentation is useful for documenting the purpose of the policy and for providing guidance for solving the problem.

7. What is alert fatigue, and why is it a problem?
   A: Too many alert notifications are sent for events that do not require human intervention, and eventually DevOps engineers begin to pay less attention to notifications

8. How long is a log data stored in the Default bucket of Cloud Logging?
   C: 30 days

9. You need to store log entries for a larger period of time than Cloud Logging retains them in the Default bucket. What is the best option for preserving log data?
   B: Create a user-defined bucket and configure a retention policy

10. Which of the following are options for logging sinks?
    D: Cloud Storage bucket, BigQuery data set, and Cloud Pub/Sub topic

11. Which of the following can be used to filter log entries when viewing logs in Cloud Logging?
    D: Log query, resource type, severity, and time

12. Which of the following is not a standard log level that can be used to filter log viewings?
    B: Halted
    There is no such standard log level status. Statuses include Critical, Error, Warning, Info, and Debug.

13. You are viewing log entries and spot one that looks suspicious. You are familiar with that kind of log entry, and want to find out what, specifically, is in a field called metadataRequest. What would you do?
    A: Expand the metadataRequest field in the JSON structure of the message

14. What Cloud Operations service is best for identifying where bottlenecks exist in your application?
    C: Cloud Trace
    … is a distributed tracing application that provides details on how long different parts of code run, so option C is correct. Monitoring is used to notify DevOps engineers when resources are not functioning as expected. Logging is for collecting, storing, and viewing log data, and although log entries might help diagnose bottlenecks, they are not specifically designed for that purpose.

15. There is a performance problem in a microservice. You have reviewed application outputs but cannot identify the problem. What Cloud Operations service would you use to gain insight into the performance of the services throughout execution?
    C: Google Trace
    … is a distributed tracing application that provides details on how longer different parts of code run.

16. You believe there may be a problem with BigQuery in the us-central zone. Where would you go to check the status of the BigQuery service for the quickest access to details?
    B: Check https://status.cloud.google.com

17. You would like to estimate the cost of Google Cloud resources you will be using. Which services would require you to have information on the virtual machines you will be using?
    B: Compute Engine and Kubernetes Engine

18. You are generating an estimate of the cost of using BigQuery. One of the parameters is Query Pricing. You have to specify a value in TB units. What is the value you are specifying?
    C: The amount of data scanned by the query

19. Why do you need to specify the operating system to be used when estimating the cost of a VM?
    B: Some operating systems incur a cost
    Some operating systems, like Microsoft Windows Server, require a license, so option B is correct. Google sometimes has arrangements with vendors to collect fees for using proprietary software.

20. Which types of log messages are sent to the Required log sink?
    D: Admin activity, system events, and access transparency

---

Good luck!
