Professional Cloud Developer

v2309

Quiz questions*

Patterns for scalable and resilient apps

* These are for practice only and are not actual exam questions

Question: You are building an application on Google Kubernetes Engine (GKE) that will handle streaming data. You anticipate varying traffic loads and a need for consistent uptime. What deployment configuration should you adopt to ensure the application remains robust during hardware failures and high-traffic events?

- A. Manually scale the number of pods during high-traffic events and rely on a fixed number of nodes.
- B. Use a Vertical Pod Autoscaler to adjust pod resources and maintain a static node pool.
- C. Rely solely on a Horizontal Pod Autoscaler without adjusting the node pool.
- D. Implement cluster autoscaler for node pool adjustments and utilize a Horizontal Pod Autoscaler for workload scaling.

Question: You are tasked with setting up a gaming platform on Google Kubernetes Engine (GKE) that will host real-time multiplayer games. The games are expected to have sporadic spikes in user traffic, especially during global tournaments. The platform must ensure: • Seamless gameplay experience during high-traffic periods • Continuous availability even if certain hardware components fail. How should you set up the deployment configurations for this platform?

- A. Use a single-zonal node pool and enable vertical pod autoscaling.
- B. Distribute the workload using multiple zonal node pools without autoscaling.
- C. Utilize a regional node pool with manual scaling based on predicted traffic.
- D. Distribute your workload evenly using a multi-zonal node pool.

Question: When setting up your GKE environment, which type of cluster topology allows the control plane and nodes to be replicated across multiple zones within a single region?

- A. Single-zone cluster
- B. Multi-region cluster
- C. Regional cluster
- D. Global cluster

Question: Which Kubernetes feature allows you to provide an accurate signal to Kubernetes for when an application has completed its initialization and is ready to serve end users?

- A. Liveness probe
- B. Readiness probe
- C. Health check

- D. Deployment check

Question: In a GKE cluster, if you want to ensure that replicas of a pod are not scheduled on the same node to avoid a single point of failure, which feature should you use?

- A. Pod affinity
- B. Pod co-location
- C. Pod anti-affinity
- D. Pod distribution

Question: You are architecting a financial analytics platform on Google Kubernetes Engine (GKE). The platform interacts with a data warehouse for fetching financial data. To ensure smooth operations, you want to validate that the platform's containers maintain a stable connection to the data warehouse. If any container encounters a connection issue, a specific script should be triggered within the container to safely close all active sessions. What should be your approach for the Deployment?

- A. Implement a startupProbe to validate the initial connection and a PostStart handler to execute the session closure script.
- B. Use a readinessProbe to ensure continuous data warehouse connectivity and a PostStop handler for the session closure script.
- C. Design the Deployment with a livenessProbe that checks the container's connection to the data warehouse. If a connection issue is detected, a Prestop lifecycle handler should initiate the session closure script.
- D. Set up a health check for monitoring the data warehouse connection and an Init container to handle session closures.

Question: You are tasked with ensuring that a specific logging agent runs on every node in your GKE cluster. Which Kubernetes object would be most suitable for this requirement?

- A. Deployment
- B. StatefulSet
- C. DaemonSet
- D. Job

Question: Which of the following statements about regional clusters in Google Kubernetes Engine (GKE) is true?

- A. Regional clusters have a single control plane in a single zone.
- B. Nodes in a regional cluster can only run in a single zone.
- C. By default, GKE replicates each node pool across two zones of the control plane's region.
- D. Regional clusters offer higher availability than zonal clusters.

Question: After deploying a critical update to your e-commerce platform on Google Kubernetes Engine, you notice a spike in error rates. The application logs are being sent to Cloud Logging, and for metrics collection, you've also implemented Prometheus. To swiftly identify the root cause of the errors and set up real-time alerts without incurring high costs, what steps should you take?

- A. Directly access the GKE nodes and extract the Prometheus metrics, then manually correlate them with Cloud Logging data.
- B. Use Cloud Trace to capture the application traces and correlate them with Cloud Logging.
- C. Use Cloud Monitoring to collect the Prometheus metrics. Configure Cloud Monitoring to create log-based metrics from the logs, and correlate them with the Prometheus data.
- D. Migrate the application to a new GKE cluster and monitor the error rates without using Prometheus.

Question: Which of the following best describes the measure of a system's ability to handle varying amounts of work by adjusting its resources?

- A. Resilience
- B. Redundancy
- C. Scalability
- D. Availability

Question: What is the primary benefit of treating your infrastructure as code (IaC)?

- A. It allows for manual configuration of resources.
- B. It ensures that resources are mutable after deployment.
- C. It encourages storing provisioning and configuration logic in source control.
- D. It promotes the use of third-party tools exclusively.

Question: What is the main advantage of creating immutable infrastructure?

- A. It allows for frequent modifications to resources after deployment.
- B. It promotes configuration drift and snowflake servers.
- C. It leads to more predictable deployments and rollbacks.
- D. It encourages manual provisioning of resources.

Question: Which of the following is NOT a driver from the business side for improving scalability and resilience?

- A. Optimize costs and resource consumption.
- B. Minimize app uptime.
- C. Ensure user demand can be met during high usage periods.

- D. Improve quality and availability of service.

Question: What is a key theme evident in the patterns for building scalable and resilient apps?

- A. Tight coupling
- B. Manual infrastructure provisioning
- C. Data-driven design
- D. Mutable infrastructure

Question: Which Google Cloud service provides an availability Service Level Agreement (SLA) and manages data replication, backups, and the underlying infrastructure for MySQL databases?

- A. Cloud Spanner
- B. Cloud Bigtable
- C. Cloud SQL
- D. Compute Engine

Question: What is a single point of failure (SPOF) in the context of application design?

- A. A component that, if it fails, will not affect the availability of the entire stack.
- B. A component that is replicated across multiple zones.
- C. A component that, if it fails, affects the availability of the entire stack.
- D. A component that is always available.

Question: Which deployment architecture is recommended for workloads that are business-critical and where high availability is essential?

- A. Single-zone

- B. Multi-zone
- C. Multi-region
- D. Hybrid

Question: What are the considerations when designing a distributed architecture with redundant resources?

- A. Only availability requirements
- B. Only operational complexity
- C. Only latency and cost
- D. Operational complexity, latency, and cost

Question: In a single-zone deployment, what happens if a VM in a MIG that hosts a web server or app server instance crashes?

- A. The VM is manually restarted.
- B. The MIG ensures that a new VM is created automatically.
- C. The application becomes unavailable.
- D. The VM is moved to another zone.

Question: In a dual-zone deployment, what happens if the primary Cloud SQL database crashes?

- A. The application becomes unavailable.
- B. The database is manually restarted.
- C. Cloud SQL automatically fails over to the standby database instance.
- D. The database is moved to another zone.

Question: To protect your application against region outages, what should you do with Google Cloud resources?

- A. Distribute them within a single zone.
- B. Distribute them across two or more zones.
- C. Distribute them within a single region.
- D. Distribute them across two or more regions.

Answers to Quiz questions
Patterns for scalable and resilient apps

Question: You are building an application on Google Kubernetes Engine (GKE) that will handle streaming data. You anticipate varying traffic loads and a need for consistent uptime. What deployment configuration should you adopt to ensure the application remains robust during hardware failures and high-traffic events?

- A. Manually scale the number of pods during high-traffic events and rely on a fixed number of nodes.
- B. Use a Vertical Pod Autoscaler to adjust pod resources and maintain a static node pool.
- C. Rely solely on a Horizontal Pod Autoscaler without adjusting the node pool.
- D. Implement cluster autoscaler for node pool adjustments and utilize a Horizontal Pod Autoscaler for workload scaling.

Correct Answer: D. Implement cluster autoscaler for node pool adjustments and utilize a Horizontal Pod Autoscaler for workload scaling.

Explanation: To ensure resilience and high availability, it's crucial to have both node-level and pod-level autoscaling. The cluster autoscaler adjusts the number of nodes in the node pool based on workload demands, while the Horizontal Pod Autoscaler scales the number of pod replicas based on specific metrics, such as CPU utilization.

Resource: [Autoscaling a cluster | Google Kubernetes Engine (GKE) - Google Cloud](#)

Question: You are tasked with setting up a gaming platform on Google Kubernetes Engine (GKE) that will host real-time multiplayer games. The games are expected to have sporadic spikes in user traffic, especially during global tournaments. The platform must ensure: • Seamless gameplay experience during high-traffic periods • Continuous availability even if certain hardware components fail. How should you set up the deployment configurations for this platform?

- A. Use a single-zonal node pool and enable vertical pod autoscaling.
- B. Distribute the workload using multiple zonal node pools without autoscaling.
- C. Utilize a regional node pool with manual scaling based on predicted traffic.
- D. Distribute your workload evenly using a multi-zonal node pool.

Correct Answer: D. Distribute your workload evenly using a multi-zonal node pool.

Explanation: Distributing the workload using a multi-zonal node pool ensures that the application is spread across multiple zones. This not only provides high availability by ensuring that the application remains accessible even if one zone experiences issues but also helps in handling traffic spikes by distributing the load.
Resource: [About node pools | Google Kubernetes Engine (GKE) | Google Cloud](#)

Question: When setting up your GKE environment, which type of cluster topology allows the control plane and nodes to be replicated across multiple zones within a single region?

- A. Single-zone cluster
- B. Multi-region cluster

- C. Regional cluster
- D. Global cluster

Correct Answer: C. Regional cluster

Explanation: A regional cluster in GKE replicates the control plane and nodes across multiple zones within a single region, offering higher availability and resilience against zone-specific issues.
Resource: [About cluster configuration choices | Google Kubernetes Engine (GKE)](#)

Question: Which Kubernetes feature allows you to provide an accurate signal to Kubernetes for when an application has completed its initialization and is ready to serve end users?

- A. Liveness probe
- B. Readiness probe
- C. Health check
- D. Deployment check

Correct Answer: B. Readiness probe

Explanation: A readiness probe is designed to let Kubernetes know when your application is truly ready to accept traffic. It ensures that Kubernetes doesn't send traffic to an application that is still initializing and not yet ready to serve requests.
Resource: [Configure Liveness, Readiness and Startup Probes | Kubernetes](#)

Question: In a GKE cluster, if you want to ensure that replicas of a pod are not scheduled on the same node to avoid a single point of failure, which feature should you use?

- A. Pod affinity

- B. Pod co-location
- C. Pod anti-affinity
- D. Pod distribution

Correct Answer: C. Pod anti-affinity

Explanation: Pod anti-affinity allows you to specify rules that prevent certain pods from being co-located on the same node. This ensures that if a node fails, not all replicas of a pod are lost, providing higher availability.
Resource: [Assigning Pods to Nodes | Kubernetes](#)

Question: You are architecting a financial analytics platform on Google Kubernetes Engine (GKE). The platform interacts with a data warehouse for fetching financial data. To ensure smooth operations, you want to validate that the platform's containers maintain a stable connection to the data warehouse. If any container encounters a connection issue, a specific script should be triggered within the container to safely close all active sessions. What should be your approach for the Deployment?

- A. Implement a startupProbe to validate the initial connection and a PostStart handler to execute the session closure script.
- B. Use a readinessProbe to ensure continuous data warehouse connectivity and a PostStop handler for the session closure script.
- C. Design the Deployment with a livenessProbe that checks the container's connection to the data warehouse. If a connection issue is detected, a Prestop lifecycle handler should initiate the session closure script.
- D. Set up a health check for monitoring the data warehouse connection and an Init container to handle session closures.

Correct Answer: C. Design the Deployment with a livenessProbe that checks the container's connection to the data warehouse. If a connection issue is detected, a Prestop lifecycle handler should initiate the session closure script.

Explanation: The livenessProbe ensures that the container maintains a stable connection to the data warehouse. If the probe detects a failure, it indicates that the container isn't functioning correctly. Before Kubernetes terminates such a container, the Prestop lifecycle handler is invoked to safely close all active sessions.
Resource: [Attach Handlers to Container Lifecycle Events | Kubernetes](#), [Configure Liveness, Readiness and Startup Probes](#)

Question: You are tasked with ensuring that a specific logging agent runs on every node in your GKE cluster. Which Kubernetes object would be most suitable for this requirement?

- A. Deployment
- B. StatefulSet
- C. DaemonSet
- D. Job

Correct Answer: C. DaemonSet

Explanation: DaemonSets are designed to ensure that a copy of a specific Pod runs on all (or some) nodes in a cluster. They are ideal for tasks like logging agents or monitoring services that need to be present on every node.
Resource: [DaemonSet in Google Kubernetes Engine](#)

Question: Which of the following statements about regional clusters in Google Kubernetes Engine (GKE) is true?

- A. Regional clusters have a single control plane in a single zone.
- B. Nodes in a regional cluster can only run in a single zone.
- C. By default, GKE replicates each node pool across two zones of the control plane's region.
- D. Regional clusters offer higher availability than zonal clusters.

Correct Answer: D. Regional clusters offer higher availability than zonal clusters.

Explanation: Regional clusters in GKE have multiple replicas of the control plane, running in multiple zones within a given region. This design provides higher availability compared to zonal clusters, which have a single control plane in one zone.
Resource: [About cluster configuration choices | Google Kubernetes Engine (GKE) | Google Cloud](#)

Question: After deploying a critical update to your e-commerce platform on Google Kubernetes Engine, you notice a spike in error rates. The application logs are being sent to Cloud Logging, and for metrics collection, you've also implemented Prometheus. To swiftly identify the root cause of the errors and set up real-time alerts without incurring high costs, what steps should you take?

- A. Directly access the GKE nodes and extract the Prometheus metrics, then manually correlate them with Cloud Logging data.
- B. Use Cloud Trace to capture the application traces and correlate them with Cloud Logging.
- C. Use Cloud Monitoring to collect the Prometheus metrics. Configure Cloud Monitoring to create log-based metrics from the logs, and correlate them with the Prometheus data.
- D. Migrate the application to a new GKE cluster and monitor the error rates without using Prometheus.

Correct Answer: C. Use Cloud Monitoring to collect the Prometheus metrics. Configure Cloud Monitoring to create log-based metrics from the logs, and correlate them with the Prometheus data.

Explanation: By using Prometheus metrics with Cloud Monitoring, you can leverage the powerful visualization and alerting capabilities of Cloud Monitoring. Creating log-based metrics from Cloud Logging allows you to correlate log data

with Prometheus metrics, providing a comprehensive view of your application's performance and health.
Resource:Google Cloud Managed Service for Prometheus

Question: Which of the following best describes the measure of a system's ability to handle varying amounts of work by adjusting its resources?

- A. Resilience
- B. Redundancy
- C. Scalability
- D. Availability

Correct Answer: C. Scalability

Explanation: Scalability is the measure of a system's ability to handle varying amounts of work by adding or removing resources from the system.
Resource: Scalable and Resilient Apps on Google Cloud

Question: What is the primary benefit of treating your infrastructure as code (IaC)?

- A. It allows for manual configuration of resources.
- B. It ensures that resources are mutable after deployment.
- C. It encourages storing provisioning and configuration logic in source control.
- D. It promotes the use of third-party tools exclusively.

Correct Answer: C. It encourages storing provisioning and configuration logic in source control.

Explanation: Infrastructure as code (IaC) encourages treating infrastructure provisioning and configuration in the same way as application code, storing it in source control for discoverability, versioning, and auditing.
Resource: [Scalable and Resilient Apps on Google Cloud](#)

Question: What is the main advantage of creating immutable infrastructure?

- A. It allows for frequent modifications to resources after deployment.
- B. It promotes configuration drift and snowflake servers.
- C. It leads to more predictable deployments and rollbacks.
- D. It encourages manual provisioning of resources.

Correct Answer: C. It leads to more predictable deployments and rollbacks.

Explanation: Immutable infrastructure mandates that resources never be modified after they're deployed. This approach leads to more predictable deployments and rollbacks and mitigates common issues in mutable infrastructures.
Resource: [Scalable and Resilient Apps on Google Cloud](#)

Question: Which of the following is NOT a driver from the business side for improving scalability and resilience?

- A. Optimize costs and resource consumption.
- B. Minimize app uptime.
- C. Ensure user demand can be met during high usage periods.
- D. Improve quality and availability of service.

Correct Answer: B. Minimize app uptime.

Explanation: Business drivers aim to optimize costs, minimize downtime, meet user demand, improve service quality, and ensure user experience and trust are maintained.

Question: What is a key theme evident in the patterns for building scalable and resilient apps?

- A. Tight coupling
- B. Manual infrastructure provisioning
- C. Data-driven design
- D. Mutable infrastructure

Correct Answer: C. Data-driven design

Explanation: Data-driven design is one of the key themes in the patterns for building scalable and resilient apps. It emphasizes the importance of collecting metrics to understand app behavior and making decisions based on data.
Resource: [Scalable and Resilient Apps on Google Cloud](#)

Question: Which Google Cloud service provides an availability Service Level Agreement (SLA) and manages data replication, backups, and the underlying infrastructure for MySQL databases?

- A. Cloud Spanner
- B. Cloud Bigtable
- C. Cloud SQL
- D. Compute Engine

Correct Answer: C. Cloud SQL

Explanation: Cloud SQL provides a MySQL database as a managed service, offering an availability SLA and handling data replication, backups, and the underlying infrastructure.
Resource: [Scalable and Resilient Apps on Google Cloud](#)

Question: What is a single point of failure (SPOF) in the context of application design?

- A. A component that, if it fails, will not affect the availability of the entire stack.
- B. A component that is replicated across multiple zones.
- C. A component that, if it fails, affects the availability of the entire stack.
- D. A component that is always available.

Correct Answer: C. A component that, if it fails, affects the availability of the entire stack.

Explanation: If any component of an application stack depends on a single infrastructure resource and that resource fails, it can affect the availability of the entire stack. Such a component is termed as a single point of failure (SPOF).
Resource: [Google Cloud Infrastructure Reliability Guide](#)

Question: Which deployment architecture is recommended for workloads that are business-critical and where high availability is essential?

- A. Single-zone
- B. Multi-zone
- C. Multi-region
- D. Hybrid

Correct Answer: C. Multi-region

Explanation: For workloads that are business-critical and where high availability is essential, such as retail and social media applications, a multi-region deployment is recommended.
Resource: [Google Cloud Infrastructure Reliability Guide](#)

Question: What are the considerations when designing a distributed architecture with redundant resources?

- A. Only availability requirements
- B. Only operational complexity
- C. Only latency and cost
- D. Operational complexity, latency, and cost

Correct Answer: D. Operational complexity, latency, and cost

Explanation: When designing a distributed architecture with redundant resources, besides the availability requirements, one must also consider the effects on operational complexity, latency, and cost.
Resource: [Google Cloud Infrastructure Reliability Guide](#)

Question: In a single-zone deployment, what happens if a VM in a MIG that hosts a web server or app server instance crashes?

- A. The VM is manually restarted.
- B. The MIG ensures that a new VM is created automatically.
- C. The application becomes unavailable.
- D. The VM is moved to another zone.

Correct Answer: B. The MIG ensures that a new VM is created automatically.

Explanation: If a VM that hosts a web server or app server instance crashes in a single-zone deployment, the Managed Instance Group (MIG) ensures that a new VM is created automatically.
Resource: [Google Cloud Infrastructure Reliability Guide](#)

Question: In a dual-zone deployment, what happens if the primary Cloud SQL database crashes?

- A. The application becomes unavailable.
- B. The database is manually restarted.
- C. Cloud SQL automatically fails over to the standby database instance.
- D. The database is moved to another zone.

Correct Answer: C. Cloud SQL automatically fails over to the standby database instance.

Explanation: In a dual-zone deployment, if the primary Cloud SQL database crashes, Cloud SQL automatically fails over to the standby database instance.
Resource: [Google Cloud Infrastructure Reliability Guide](#)

Question: To protect your application against region outages, what should you do with Google Cloud resources?

- A. Distribute them within a single zone.
- B. Distribute them across two or more zones.
- C. Distribute them within a single region.
- D. Distribute them across two or more regions.

Correct Answer: D. Distribute them across two or more regions.

Explanation: To protect your application against region outages, you can distribute the Google Cloud resources across two or more regions.

Resource: [Google Cloud Infrastructure Reliability Guide](#)