



- Expert Verified, Online, Free.

MENU

Google Discussions



Exam Associate Cloud Engineer All Questions

View all questions & answers for the Associate Cloud Engineer exam

Go to Exam

EXAM ASSOCIATE CLOUD ENGINEER TOPIC 1 QUESTION 249 DISCUSSION

Actual exam question from Google's Associate Cloud Engineer

Question #: 249

Topic #: 1

[\[All Associate Cloud Engineer Questions\]](#)

Your manager asks you to deploy a workload to a Kubernetes cluster. You are not sure of the workload's resource requirements or how the requirements might vary depending on usage patterns, external dependencies, or other factors. You need a solution that makes cost-effective recommendations regarding CPU and memory requirements, and allows the workload to function consistently in any situation. You want to follow Google-recommended practices. What should you do?

- A. Configure the Horizontal Pod Autoscaler for availability, and configure the cluster autoscaler for suggestions.
- B. Configure the Horizontal Pod Autoscaler for availability, and configure the Vertical Pod Autoscaler recommendations for suggestions.
- C. Configure the Vertical Pod Autoscaler recommendations for availability, and configure the Cluster autoscaler for suggestions.
- D. Configure the Vertical Pod Autoscaler recommendations for availability, and configure the Horizontal Pod Autoscaler for suggestions.

Show Suggested Answer

by [KelvinToo](#) at Dec. 31, 2023, 6:52 p.m.

Comments

Type your comment...

Submit

Urbanvzla Highly Voted 9 months, 2 weeks ago

Selected Answer: B

Horizontal Pod Autoscaler (HPA): It automatically scales the number of pods in a deployment, replica set, or stateful set based on observed CPU utilization (or, with custom metrics support, on some other application-provided metrics). This helps maintain availability by ensuring that your application has the necessary number of pods to handle the workload.

Vertical Pod Autoscaler (VPA): It automatically adjusts the CPU and memory reservations for your pods to help "right size" your applications. This is particularly useful when you're unsure of the resource requirements. VPA makes recommendations for the appropriate CPU and memory settings based on usage patterns, which can be very effective for cost optimization.

This combination ensures that your workload is both horizontally scalable (to handle changes in demand) and vertically optimized (to use resources efficiently), following Google-recommended practices for Kubernetes workloads.

upvoted 7 times

PiperMe Most Recent 7 months, 3 weeks ago

Selected Answer: B

I believe B is the best choice: HPA ensures availability by scaling the number of pods based on metrics (like CPU utilization). The VPA analyzes resource utilization and provides recommendations for CPU and memory requests and limits. This is key for right-sizing your pods for optimal cost efficiency.

upvoted 2 times

Lakshvenkat 9 months, 2 weeks ago

Selected Answer: D

D is the correct answer

upvoted 1 times

kuracpalac 7 months, 3 weeks ago

VPA is not recommended as per Google requirements, so that answer must be wrong.

upvoted 1 times

Cynthia2023 9 months, 3 weeks ago

Selected Answer: B

Horizontal Pod Autoscaler (HPA): HPA automatically adjusts the number of pods in a deployment based on observed CPU utilization (or other select metrics). This is crucial for maintaining the availability of your workload, especially if the workload experiences varying levels of traffic or load. HPA ensures that there are enough pods to handle the load, scaling out (adding more pods) when demand is high and scaling in (removing pods) when demand is low.

Vertical Pod Autoscaler (VPA) Recommendations: VPA automatically adjusts the CPU and memory reservations for pods in a deployment. It can operate in a mode where it only provides recommendations (without automatically applying them), which is useful for understanding the resource needs of your workload. VPA recommendations can guide you in setting appropriate CPU and memory limits based on the observed usage of your workload.

upvoted 3 times

Cynthia2023 9 months, 3 weeks ago

A. Cluster Autoscaler for Suggestions: While the Cluster Autoscaler is useful for scaling the number of nodes in the cluster, it doesn't provide recommendations on the CPU and memory requirements for individual pods.

C. VPA for Availability, Cluster Autoscaler for Suggestions: VPA can automatically adjust pod sizes, but using it for ensuring availability might lead to frequent and potentially disruptive pod restarts. The Cluster Autoscaler is again more about node-level scaling rather than providing pod resource recommendations.

D. VPA for Availability, HPA for Suggestions: This configuration isn't ideal as VPA's primary function isn't about maintaining high availability but rather about optimizing resource allocation. HPA, on the other hand, is specifically designed for scaling the number of pods based on load, which is directly related to availability.

upvoted 1 times

kaby1987 9 months, 3 weeks ago

Selected Answer: B

Ans is B

B. Configure the Horizontal Pod Autoscaler for availability, and configure the Vertical Pod Autoscaler recommendations for suggestions.

This approach allows you to manage the number of pods based on the workload (HPA) and get optimal CPU and memory settings for each pod (VPA), which is in line with Google-recommended practices for managing Kubernetes workloads with uncertain resource requirements. This combination ensures that your workload can function consistently in varying situations by automatically adjusting both the quantity of pods and the resources each pod is allocated.

upvoted 2 times

KelvinToo 9 months, 3 weeks ago

Selected Answer: D

ChatGPT says option D,
By configuring VPA for resource recommendations based on actual usage patterns and HPA for scaling pod instances based on demand, you can ensure that your workload is both cost-effective and capable of adapting to varying resource requirements, all while following Google-recommended practices for Kubernetes workloads.

upvoted 1 times

PiperMe 7 months, 3 weeks ago

Chat strikes again. Option B provides a more tailored and Google-recommended approach given the uncertainty about the workload's resource needs. It prioritizes establishing an efficient baseline with VPA before relying on HPA for scaling.

upvoted 1 times

Start Learning for free

Social Media

[Facebook](#) , [Twitter](#)

[YouTube](#) , [Reddit](#)

[Pinterest](#)

EXAMTOPICS

We are the biggest and most updated IT certification exam material website.

Using our own resources, we strive to strengthen the IT professionals community for free.



© 2024 ExamTopics

ExamTopics doesn't offer Real Microsoft Exam Questions. ExamTopics doesn't offer Real Amazon Exam Questions. ExamTopics Materials do not contain actual questions and answers from Cisco's Certification Exams.

CFA Institute does not endorse, promote or warrant the accuracy or quality of ExamTopics. CFA® and Chartered Financial Analyst® are registered trademarks owned by CFA Institute.