



- Expert Verified, Online, Free.

MENU



Google Discussions



## Exam Professional Machine Learning Engineer All Questions

View all questions & answers for the Professional Machine Learning Engineer exam

Go to Exam

### EXAM PROFESSIONAL MACHINE LEARNING ENGINEER TOPIC 1 QUESTION 262 DISCUSSI...

Actual exam question from Google's Professional Machine Learning Engineer

Question #: 262

Topic #: 1

[\[All Professional Machine Learning Engineer Questions\]](#)

You recently deployed a model to a Vertex AI endpoint and set up online serving in Vertex AI Feature Store. You have configured a daily batch ingestion job to update your featurestore. During the batch ingestion jobs, you discover that CPU utilization is high in your featurestore's online serving nodes and that feature retrieval latency is high. You need to improve online serving performance during the daily batch ingestion. What should you do?

- A. Schedule an increase in the number of online serving nodes in your featurestore prior to the batch ingestion jobs
- B. Enable autoscaling of the online serving nodes in your featurestore
- C. Enable autoscaling for the prediction nodes of your DeployedModel in the Vertex AI endpoint
- D. Increase the worker\_count in the ImportFeatureValues request of your batch ingestion job

Show Suggested Answer

by [pikachu007](#) at Jan. 13, 2024, 3:47 p.m.

## Comments

Type your comment...

Submit

🗨️ 👤 **tardigradum** 2 months, 1 week ago

Selected Answer: A

Agree with bobjr

👍 ↩️ 🚩 upvoted 1 times

🗨️ 👤 **Prakzz** 3 months, 3 weeks ago

Selected Answer: A

<https://cloud.google.com/vertex-ai/docs/featurestore/managing-featurestores>

Specifically mentioned here that --> If CPU utilization is consistently high, consider increasing the number of online serving nodes for your featurestore.

👍 ↩️ 🚩 upvoted 1 times

🗨️ 👤 **bobjr** 4 months, 2 weeks ago

Selected Answer: A

Gemini + Perplexity ai + ChatGPT votes A

Because : B. Enable Autoscaling: While autoscaling can be useful, it might not react quickly enough to sudden spikes in traffic during batch ingestion. Scheduling the increase ensures that the resources are available when needed.

👍 ↩️ 🚩 upvoted 2 times

🗨️ 👤 **cruise93** 6 months ago

Selected Answer: D

This question is valid for the Legacy feature store.

[https://cloud.google.com/vertex-ai/docs/featurestore/ingesting-batch#import\\_job\\_performance](https://cloud.google.com/vertex-ai/docs/featurestore/ingesting-batch#import_job_performance)

👍 ↩️ 🚩 upvoted 1 times

🗨️ 👤 **pinimichele01** 5 months, 4 weeks ago

"CPU utilization is high in your featurestore's online serving nodes"

👍 ↩️ 🚩 upvoted 1 times

🗨️ 👤 **daidai75** 9 months ago

Selected Answer: B

[https://cloud.google.com/vertex-ai/docs/featurestore/managing-featurestores?](https://cloud.google.com/vertex-ai/docs/featurestore/managing-featurestores?&_gl=1*sswg5e*_ga*NDE2OTc3OTAzLjE3MDU4OTQ5OTE.*_ga_WH2QY8WWF5*MTcwNTkzNDM0NS40LjAuMTcwNTkzNDM0NS4wLjAuMA..&_ga=2.242492743.-416977903.1705894991#online_serving_nodes)

[\\_gl=1\\*sswg5e\\*\\_ga\\*NDE2OTc3OTAzLjE3MDU4OTQ5OTE.\\*\\_ga\\_WH2QY8WWF5\\*MTcwNTkzNDM0NS40LjAuMTcwNTkzNDM0NS4wLjAuMA..&\\_ga=2.242492743.-416977903.1705894991#online\\_serving\\_nodes](https://cloud.google.com/vertex-ai/docs/featurestore/managing-featurestores?&_gl=1*sswg5e*_ga*NDE2OTc3OTAzLjE3MDU4OTQ5OTE.*_ga_WH2QY8WWF5*MTcwNTkzNDM0NS40LjAuMTcwNTkzNDM0NS4wLjAuMA..&_ga=2.242492743.-416977903.1705894991#online_serving_nodes)

👍 ↩️ 🚩 upvoted 2 times

🗨️ 👤 **b1a8fae** 9 months ago

Selected Answer: B

Vertex AI Feature Store provides two options for online serving: Bigtable and optimized online serving. Both options support autoscaling, which means that the number of online serving nodes can automatically adjust to the traffic demand. By enabling autoscaling, you can improve the online serving performance and reduce the feature retrieval latency during the daily batch ingestion. Autoscaling also helps you optimize the cost and resource utilization of your featurestore.

👍 ↩️ 🚩 upvoted 2 times

🗨️ 👤 **pikachu007** 9 months, 1 week ago

Selected Answer: B

Option A: Manually scheduling node increases requires prior knowledge of batch ingestion times and might not be as responsive to unexpected workload spikes.

Option C: Autoscaling prediction nodes in the Vertex AI endpoint might help with model prediction latency but doesn't directly address feature retrieval latency from the featurestore.

Option D: Increasing worker\_count in the batch ingestion job could speed up ingestion but might further strain online serving nodes, potentially worsening latency.

👍 ↩️ 🚩 upvoted 1 times

🗨️ 👤 **ieeva** 9 months, 1 week ago

Hey Pikachu,

did you pass the exam or are you preparing? I am as well preparing and I have noticed that in many question you chose the same answer I would chose, but which is not the indicated answer of my Udemy Course Exam Preparation.

Thanks and Best

👍 ↩️ 🚩 upvoted 1 times

🗨️ 👤 **pikachu007** 9 months, 1 week ago

Yes passed

👍 ↩️ 🚩 upvoted 5 times

🗨️ 👤 **Sunny\_M** 9 months ago

Hi @pikachu007 May I ask when did you pass the exam? was it after they updated the new questions? I need to take the

exam ASAP, just want to make sure the new questions are valid?

   upvoted 1 times

**Start Learning for free**



### Social Media

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

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

[Pinterest](#)



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.

