

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 240 DISCUSSI...

Actual exam question from Google's Professional Machine Learning Engineer

Question #: 240

Topic #: 1

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

You work at a mobile gaming startup that creates online multiplayer games. Recently, your company observed an increase in players cheating in the games, leading to a loss of revenue and a poor user experience. You built a binary classification model to determine whether a player cheated after a completed game session, and then send a message to other downstream systems to ban the player that cheated. Your model has performed well during testing, and you now need to deploy the model to production. You want your serving solution to provide immediate classifications after a completed game session to avoid further loss of revenue. What should you do?

- A. Import the model into Vertex AI Model Registry. Use the Vertex Batch Prediction service to run batch inference jobs.
- B. Save the model files in a Cloud Storage bucket. Create a Cloud Function to read the model files and make online inference requests on the Cloud Function.
- C. Save the model files in a VM. Load the model files each time there is a prediction request, and run an inference job on the VM.
- D. Import the model into Vertex AI Model Registry. Create a Vertex AI endpoint that hosts the model, and make online inference requests.




Show Suggested Answer

by [pikachu007](#) at Jan. 13, 2024, 8:30 a.m.

Comments

Type your comment...

Submit

  **guilhermebutzke** Highly Voted  8 months, 3 weeks ago

My answer: D

A: Not correct: Batch Prediction is designed for offline processing of large datasets, not for immediate real-time predictions needed in this scenario.

B: Not correct: While Cloud Functions offer real-time processing, loading the model files each time might introduce latency, especially for larger models

C: Not correct: Using a VM is less scalable and more complex to manage compared to other options.

D: CORRECT: Vertex AI Model Registry ensures proper model management, versioning, and access control while Vertex AI endpoint provides a highly scalable and managed solution for real-time online inference, ensuring immediate predictions after game sessions.

   upvoted 6 times

  **fitri001** Most Recent  6 months, 3 weeks ago

Selected Answer: D

Low Latency: Vertex AI Endpoints are specifically designed for low-latency online inference. They offer automatic scaling and efficient resource allocation, ensuring quick responses to game session completion signals.

Real-time Decisions: This deployment method allows your game backend to send data from finished game sessions to the Vertex AI endpoint in near real-time. The endpoint can then make classifications (cheater or not cheater) promptly.

Managed Service: Vertex AI handles the infrastructure management and scaling of your model, freeing you from managing servers or virtual machines (VMs).

   upvoted 4 times


  **fitri001** 6 months, 3 weeks ago

A. Vertex Batch Prediction: Batch prediction is designed for offline processing of large datasets, not real-time inference on individual game sessions.

B. Cloud Function with Model Files: While Cloud Functions can be triggered by events, reading the model files each time and running inference can introduce latency. This might not be ideal for immediate classifications.

C. Model Files in a VM: Loading the model on a VM for each inference request incurs significant overhead and latency. This approach is not suitable for real-time processing.

   upvoted 1 times

  **pikachu007** 9 months, 3 weeks ago

Selected Answer: D

Option A: Batch prediction is too slow for your needs.

Option B: Cloud Functions are ideal for short-lived tasks, not for continuously serving models. Loading the model on every request would be inefficient.

Option C: VMs offer less scalability and management overhead compared to Vertex AI.

   upvoted 3 times

  **sonicclasps** 9 months, 1 week ago

although the game is multiplayer, and you could submit requests for all the players in the game that just ended, as a batch. So I think A is also an option

   upvoted 1 times



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.

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.