



- 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 182 DISCUSSI...

Actual exam question from Google's Professional Machine Learning Engineer

Question #: 182

Topic #: 1

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

You are deploying a new version of a model to a production Vertex AI endpoint that is serving traffic. You plan to direct all user traffic to the new model. You need to deploy the model with minimal disruption to your application. What should you do?

- A. 1. Create a new endpoint
- 2. Create a new model. Set it as the default version. Upload the model to Vertex AI Model Registry
- 3. Deploy the new model to the new endpoint
- 4. Update Cloud DNS to point to the new endpoint
- B. 1. Create a new endpoint
- 2. Create a new model. Set the parentModel parameter to the model ID of the currently deployed model and set it as the default version. Upload the model to Vertex AI Model Registry
- 3. Deploy the new model to the new endpoint, and set the new model to 100% of the traffic.
- C. 1. Create a new model. Set the parentModel parameter to the model ID of the currently deployed model. Upload the model to Vertex AI Model Registry.
- 2. Deploy the new model to the existing endpoint, and set the new model to 100% of the traffic
- D. 1. Create a new model. Set it as the default version. Upload the model to Vertex AI Model Registry
- 2. Deploy the new model to the existing endpoint

Show Suggested Answer

Comments

Type your comment...

Submit

  **fitri001** 6 months, 2 weeks ago

Selected Answer: C

Minimal Downtime: By deploying the new model to the existing endpoint, you avoid any service interruptions caused by creating and switching to a completely new endpoint.

Versioning: Setting the parentModel parameter allows you to track the lineage of your models and easily revert to the previous version if needed.

Traffic Control: Vertex AI lets you control traffic allocation between different versions of a model deployed on the same endpoint. Setting the new model to 100% traffic directs all user requests to the new version.

   upvoted 2 times

  **fitri001** 6 months, 2 weeks ago

A. Creating a New Endpoint: This approach introduces downtime as you need to switch DNS records to point to the new endpoint.

B. Creating a New Endpoint with Default Version: While using a parentModel helps with versioning, creating a new endpoint still leads to service disruption.

D. Deploying to Existing Endpoint Without Traffic Control: This might cause unexpected behavior if the new model isn't ready for production traffic.

   upvoted 1 times

  **guilhermebutzke** 8 months, 3 weeks ago

Selected Answer: C

My Answer: C

In the context of deploying machine learning models, setting the `parentModel` parameter to the model ID of the currently deployed model means that the new model being deployed is created as a child model or an iteration of the existing model. This allows the new model to inherit certain properties or characteristics from the existing model, such as the architecture, hyperparameters, or feature transformations.

Create a new Endpoint is Unnecessary.

   upvoted 2 times

  **sonicclasps** 9 months ago

Selected Answer: D

Optionally set this model as the default version. The default version is preselected whenever the model is used for prediction (although you can still select other versions).

<https://cloud.google.com/vertex-ai/docs/model-registry/versioning>

   upvoted 1 times

  **BlehMaks** 9 months, 3 weeks ago

Selected Answer: C

a,c -creating new endpoint is an unnecessary disruption to the application

d - doesn't work, two models are on the same endpoint and traffic is still going through the old model

   upvoted 1 times

  **pikachu007** 9 months, 4 weeks ago

Selected Answer: C

Leverages existing endpoint: Using the same endpoint maintains the same endpoint URL, avoiding DNS updates and potential service interruptions.

Gradual traffic transition: Vertex AI allows you to gradually shift traffic between model versions, ensuring a smooth transition without impacting users.

Clear versioning: Setting parentModel establishes a relationship between the new model and the existing one, aiding in organization and tracking model lineage.

   upvoted 3 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.

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.