



- 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 278 DISCUSSI...

Actual exam question from Google's Professional Machine Learning Engineer

Question #: 278

Topic #: 1

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

You need to train an XGBoost model on a small dataset. Your training code requires custom dependencies. You want to minimize the startup time of your training job. How should you set up your Vertex AI custom training job?

- A. Store the data in a Cloud Storage bucket, and create a custom container with your training application. In your training application, read the data from Cloud Storage and train the model.
- B. Use the XGBoost prebuilt custom container. Create a Python source distribution that includes the data and installs the dependencies at runtime. In your training application, load the data into a pandas DataFrame and train the model.
- C. Create a custom container that includes the data. In your training application, load the data into a pandas DataFrame and train the model.
- D. Store the data in a Cloud Storage bucket, and use the XGBoost prebuilt custom container to run your training application. Create a Python source distribution that installs the dependencies at runtime. In your training application, read the data from Cloud Storage and train the model.

Show Suggested Answer

by [guilhermebutzke](#) at Feb. 19, 2024, 3:10 p.m.

Comments

Type your comment...

Submit

  **guilhermebutzke** Highly Voted  8 months ago

Selected Answer: A

My Answer: A

Focus on “training code requires custom dependencies” and “minimize the startup time of your training job”, the best choice is A because use custom container and read the data from GCS is the faster way

   upvoted 5 times

  **Foxy2021** Most Recent  1 week, 1 day ago

I select D: While A could work, D is the optimal solution because it balances efficiency, ease of setup, and performance. It minimizes startup time by leveraging Google's prebuilt XGBoost container and offers flexibility by installing custom dependencies at runtime. This approach avoids the overhead of building and maintaining a custom container from scratch, which is unnecessary for a small dataset with only specific custom dependency needs.

   upvoted 1 times

  **wences** 3 weeks, 2 days ago

Selected Answer: A

The fastest way is to have most of the things already installed, so that is why option A fits the best

   upvoted 1 times

  **omribt** 4 months ago

Selected Answer: C

The focus is on startup time, and the dataset is small, so the container should still be of reasonable size. Downloading data from Cloud Storage introduces a delay.

   upvoted 2 times

  **bobjr** 4 months, 2 weeks ago

Selected Answer: C

The dataset is small, xgboost is implemented in python... (correcting my error A answer)

   upvoted 1 times

  **bobjr** 4 months, 2 weeks ago

Selected Answer: A

The dataset is small, xgboost is implemented in python...

   upvoted 1 times

  **omermahgoub** 6 months, 1 week ago


Selected Answer: A

Given the focus on minimizing startup time, and based on the information about XGBoost prebuilt container dependencies available here <https://cloud.google.com/vertex-ai/docs/training/pre-built-containers#xgboost>

A: Separate Data and Custom Container is the best approach for minimizing startup time, especially for small datasets. Separating data in Cloud Storage keeps the container image lean, leading to faster download and startup compared to bundling data within the container.

B. The prebuilt Container could have unnecessary components, potentially increasing the image size and impacting startup time.

   upvoted 4 times

  **CHARLIE2108** 7 months ago

Why not C?

   upvoted 1 times


  **tavva_prudhvi** 6 months, 3 weeks ago

Because, Including the data in the container image is not recommended as it increases the image size and makes it less reusable.

   upvoted 3 times

  **raidenrock** 5 months, 3 weeks ago

But the description mentioned it is a small dataset and requires minimizing latency which makes C the best per requirement, there is no mentioning to make the container reusable whatsoever

   upvoted 1 times

  **Yan_X** 7 months, 2 weeks ago

Selected Answer: B

B

XGBoost prebuilt customer container already includes XGBoost library and all of its dependencies. Python source distribution to avoid overhead of reading the data from Cloud storage the 2nd time. Load data to a Pandas DataFrame is convenient to work with Python. Pandas is for data analysis and manipulation.

   upvoted 2 times

  **tavva_prudhvi** 6 months, 3 weeks ago

However, the question specifically says that the training code requires custom dependencies beyond those included in the prebuilt container. Therefore, using the prebuilt container alone would not be sufficient in this case.

&

regarding the use of a Python source distribution to avoid reading data from Cloud Storage multiple times, it's important to consider the trade-off between startup time and potential performance gains. While including the data in the source distribution might save some time during training, it also increases the size of the container and can lead to longer startup times. For small datasets, the overhead of reading data from Cloud Storage is typically negligible compared to the benefits of a smaller container and faster startup.

   upvoted 2 times

  **tavva_prudhvi** 6 months, 3 weeks ago

Also, creating a Python source distribution that includes the data and installs the dependencies at runtime can increase startup time since dependencies have to be installed every time the job runs

   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.

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.