

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 29 DISCUSSIO..

Actual exam question from Google's Professional Machine Learning Engineer

Question #: 29

Topic #: 1

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

You have trained a model on a dataset that required computationally expensive preprocessing operations. You need to execute the same preprocessing at prediction time. You deployed the model on AI Platform for high-throughput online prediction. Which architecture should you use?

- A. Validate the accuracy of the model that you trained on preprocessed data. Create a new model that uses the raw data and is available in real time. Deploy the new model onto AI Platform for online prediction.
- B. Send incoming prediction requests to a Pub/Sub topic. Transform the incoming data using a Dataflow job. Submit a prediction request to AI Platform using the transformed data. Write the predictions to an outbound Pub/Sub queue.
- C. Stream incoming prediction request data into Cloud Spanner. Create a view to abstract your preprocessing logic. Query the view every second for new records. Submit a prediction request to AI Platform using the transformed data. Write the predictions to an outbound Pub/Sub queue.
- D. Send incoming prediction requests to a Pub/Sub topic. Set up a Cloud Function that is triggered when messages are published to the Pub/Sub topic. Implement your preprocessing logic in the Cloud Function. Submit a prediction request to AI Platform using the transformed data. Write the predictions to an outbound Pub/Sub queue.

Show Suggested Answer

by [inder0007](#) at June 9, 2021, 9 p.m.

## Comments

Type your comment...

Submit

  **SparkExpedition** Highly Voted  3 years, 3 months ago

Supporting B ..[https://cloud.google.com/architecture/data-preprocessing-for-ml-with-tf-transform-pt1#where\\_to\\_do\\_preprocessing](https://cloud.google.com/architecture/data-preprocessing-for-ml-with-tf-transform-pt1#where_to_do_preprocessing)

   upvoted 29 times

  **inder0007** Highly Voted  3 years, 4 months ago

I think it should b B

   upvoted 13 times

  **q4exam** 3 years, 1 month ago

I also agree with B, this is how I would advise clients to do it as well

   upvoted 3 times

  **bludw** Most Recent  3 months, 3 weeks ago

Selected Answer: D

D. The issue with B is that DataFlow does not work well with high throughput


   upvoted 1 times

  **PhilipKoku** 4 months, 2 weeks ago

Selected Answer: B

B) Pub/Sub + Dataflow

   upvoted 1 times

  **Liting** 1 year, 3 months ago

Selected Answer: B

Went with B, using dataflow for large amount data transformation is the best option

   upvoted 3 times


  **SamuelTsch** 1 year, 3 months ago

Selected Answer: B

I went to B.

A is completely wrong. C: 1st cloud spanner is not designed for high throughput, also it is not for preprocessing. D: cloud function could not be get enough resource to do the high computational transformation.

   upvoted 2 times

  **ashu381** 1 year, 4 months ago

Selected Answer: B

Because the concern here is high throughput and not specifically the latency so better to go with option B

   upvoted 1 times

  **Voyager2** 1 year, 4 months ago

Selected Answer: D

B. Send incoming prediction requests to a Pub/Sub topic. Transform the incoming data using a Dataflow job. Submit a prediction request to AI Platform using the transformed data. Write the predictions to an outbound Pub/Sub queue <https://dataintegration.info/building-streaming-data-pipelines-on-google-cloud>



   upvoted 1 times

  **M25** 1 year, 5 months ago

Selected Answer: B

Went with B

   upvoted 1 times

  **e707** 1 year, 5 months ago

Selected Answer: D

I think it's D as B is not a good choice because it requires you to run a Dataflow job for each prediction request. This is inefficient and can lead to latency issues.

   upvoted 3 times

  **lucaluca1982** 1 year, 5 months ago

Yes i agree Dataflow can introduce latency

   upvoted 3 times

   upvoted 2 times

  **lucaluca1982** 1 year, 6 months ago

**Selected Answer: D**

I go for D. Option B has Dataflow that it is more suitable for batch


   upvoted 1 times

  **SergioRubiano** 1 year, 7 months ago

**Selected Answer: B**

It's B

   upvoted 1 times

  **MithunDesai** 1 year, 10 months ago

**Selected Answer: B**

yes ans B

   upvoted 1 times

  **hiromi** 1 year, 10 months ago

**Selected Answer: B**

B

Pubsub + DataFlow + Vertex AI (AI Platform)

   upvoted 1 times

  **suresh\_vn** 2 years, 2 months ago

**Selected Answer: B**

Should be B. Dataflow is BEST option for preprocessing training , testing data both

   upvoted 2 times

  **sachinxshrivastav** 2 years, 2 months ago

**Selected Answer: B**

Answer should be B

   upvoted 1 times

  **Mohamed\_Mossad** 2 years, 4 months ago

**Selected Answer: B**

- using options eliminatios , A totally wrong , D also not valid as cloud functions is not suitable for heavy data workflows

- answer between B,D will vote for B as dataflow is the best solution while dealing with heavy data workflows

   upvoted 2 times

[Load full discussion...](#)

**Start Learning for free**



**Social Media**

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

[YouTube](#) , [LinkedIn](#)

[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.