

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 4 DISCUSSION

Actual exam question from Google's Professional Machine Learning Engineer

Question #: 4

Topic #: 1

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

You want to rebuild your ML pipeline for structured data on Google Cloud. You are using PySpark to conduct data transformations at scale, but your pipelines are taking over 12 hours to run. To speed up development and pipeline run time, you want to use a serverless tool and SQL syntax. You have already moved your raw data into Cloud Storage. How should you build the pipeline on Google Cloud while meeting the speed and processing requirements?

- A. Use Data Fusion's GUI to build the transformation pipelines, and then write the data into BigQuery.
- B. Convert your PySpark into SparkSQL queries to transform the data, and then run your pipeline on Dataproc to write the data into BigQuery.
- C. Ingest your data into Cloud SQL, convert your PySpark commands into SQL queries to transform the data, and then use federated queries from BigQuery for machine learning.
- D. Ingest your data into BigQuery using BigQuery Load, convert your PySpark commands into BigQuery SQL queries to transform the data, and then write the transformations to a new table.

Show Suggested Answer

by [ralf\\_cc](#) at July 8, 2021, 12:35 p.m.

### Comments

Type your comment...

Submit

  **nunzio144** Highly Voted  1 month ago

It should be D .... Data Fusion is not SQL syntax ....

   upvoted 21 times

  **q4exam** 3 years, 1 month ago

Agree, BQ is the only serverless that support SQL

   upvoted 4 times

  **A4M** 2 years, 9 months ago

Needs to be D as the most suitable answer given the req's in question Datafusion is more of a no code Data transformation tool

   upvoted 1 times

  **Celia20210714** Highly Voted  3 years, 3 months ago

ANS: A

<https://cloud.google.com/data-fusion#section-1>

- Data Fusion is a serverless approach leveraging the scalability and reliability of Google services like Dataproc means Data Fusion offers the best of data integration capabilities with a lower total cost of ownership.

- BigQuery is serverless and supports SQL.

- Dataproc is not serverless, you have to manage clusters.

- Cloud SQL is not serverless, you have to manage instances.

   upvoted 12 times

  **TornikePirveli** 2 months, 1 week ago

By your logic it should be D, because BQ is fully serverless and supports SQL

   upvoted 1 times

  **q4exam** 3 years, 1 month ago

Data Fusion is not serverless, it create dataproc to execute the job .... I think the answer is C

   upvoted 1 times

  **mousseUwU** 3 years ago

Data Fusion is serverless: <https://cloud.google.com/data-fusion#all-features>

   upvoted 3 times

  **tavva\_prudhvi** 1 year, 7 months ago

I think you're only viewing the sentence "A serverless approach leveraging the scalability and reliability of Google services like Dataproc means Data Fusion offers the best of data integration capabilities with a lower total cost of ownership", The sentence implies that Data Fusion leverages a serverless approach, but it does not explicitly state that Data Fusion itself is serverless. It states that Data Fusion offers the best of data integration capabilities by using a serverless approach that leverages the scalability and reliability of Google services like Dataproc. So, while Data Fusion may not be fully serverless, it is designed to take advantage of serverless capabilities through its integration with Google services.

   upvoted 2 times

  **mousseUwU** 3 years ago

Agree, A is correct

   upvoted 2 times

  **LeumaS\_NoswaY** Most Recent  1 month ago

B. You need Cloud Dataproc to transform the data from PySpark to Spark SQL

   upvoted 1 times

  **TornikePirveli** 2 months, 1 week ago

Serverless, SQL syntax -> BigQuery, simple as that

   upvoted 2 times

  **jsalvasoler** 2 months, 2 weeks ago

I am very curious. Why are the solutions (when I click Reveal Solution) generally WRONG?

   upvoted 2 times

  **tadeupan** 3 months, 1 week ago

option D because needs a serverless solution and sql syntax and BigQuery offer this. Dataproc is not serverless, so B is incorrect, D is correct option.

   upvoted 2 times

  **Yerke** 2 months, 2 weeks ago

 **TORKO** 3 months, 2 weeks ago

**Selected Answer: D**

There's an updated version of this question in the official Google Cloud certified PMLE study guide. Option D is marked as correct

   upvoted 2 times

  **TornikePirveli** 2 months, 1 week ago

Can you link the updated version? On Amazon it's still 1st version and marked B

   upvoted 1 times

  **PhilipKoku** 4 months, 2 weeks ago

**Selected Answer: D**

The best approach is option D: Ingest data into BigQuery and use SQL queries for transformations. This leverages BigQuery's serverless capabilities, efficient processing, and seamless integration with other Google Cloud services.

   upvoted 2 times

  **fragkris** 10 months, 3 weeks ago

**Selected Answer: D**

D - BigQuery is the only serverless and SQL-syntax option.



   upvoted 1 times

  **Sum\_Sum** 11 months, 1 week ago

**Selected Answer: D**

D - as BQ is server less and supports SQL  
none of the other options match both criteria

   upvoted 2 times

  **12112** 1 year, 3 months ago

**Selected Answer: D**

I'll go with D.

   upvoted 1 times

  **M25** 1 year, 5 months ago

**Selected Answer: D**

Went with D

   upvoted 3 times

  **asava** 1 year, 7 months ago

**Selected Answer: B**

BQ is the serverless solution

   upvoted 3 times

  **TornikePirveli** 2 months, 1 week ago

But using dataproc is not serverless, so answer should be D

   upvoted 1 times

  **mellowed** 1 year, 9 months ago

Correct option is D

   upvoted 1 times

  **ssaporylo** 1 year, 9 months ago

Vote D

   upvoted 1 times

  **ares81** 1 year, 9 months ago

**Selected Answer: A**

It should be A.

   upvoted 1 times

  **EFIGO** 1 year, 11 months ago

**Selected Answer: D**

Data Fusion is not in SQL syntax, so no A;  
Dataproc is not serverless, so no B;  
Passing through Cloud SQL is useless, just go with BigQuery, so no C;  
D is correct

   upvoted 4 times

[Load full discussion...](#)

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.

