



- Expert Verified, Online, **Free**.

MENU

Google Discussions



## Exam Associate Cloud Engineer All Questions

View all questions & answers for the Associate Cloud Engineer exam

[Go to Exam](#)

### EXAM ASSOCIATE CLOUD ENGINEER TOPIC 1 QUESTION 146 DISCUSSION

Actual exam question from Google's Associate Cloud Engineer

Question #: 146

Topic #: 1

[\[All Associate Cloud Engineer Questions\]](#)

You have an application that uses Cloud Spanner as a database backend to keep current state information about users. Cloud Bigtable logs all events triggered by users. You export Cloud Spanner data to Cloud Storage during daily backups. One of your analysts asks you to join data from Cloud Spanner and Cloud Bigtable for specific users. You want to complete this ad hoc request as efficiently as possible. What should you do?

- A. Create a dataflow job that copies data from Cloud Bigtable and Cloud Storage for specific users.
- B. Create a dataflow job that copies data from Cloud Bigtable and Cloud Spanner for specific users.
- C. Create a Cloud Dataproc cluster that runs a Spark job to extract data from Cloud Bigtable and Cloud Storage for specific users.
- D. Create two separate BigQuery external tables on Cloud Storage and Cloud Bigtable. Use the BigQuery console to join these tables through user fields, and apply appropriate filters.

[Show Suggested Answer](#)

by [SSPC](#) at Aug. 13, 2020, 2:02 p.m.

### Comments

Type your comment...

**AmitKM** Highly Voted 4 years, 2 months ago

I think it should be D. <https://cloud.google.com/bigquery/external-data-sources>  
upvoted 38 times

**SSPC** 4 years, 2 months ago

The question says: "Join data from Cloud Spanner and Cloud Bigtable for specific users" You can see the Google documentation in the link <https://cloud.google.com/spanner/docs/export>  
upvoted 3 times

**Eshkrkrkr** 3 years, 11 months ago

Oh my god, SSPC read you your links!

The process uses Dataflow and exports data to a folder in a Cloud Storage bucket. The resulting folder contains a set of Avro files and JSON manifest files. And what next? I will tell - next you read below: Compute Engine: Before running your export job, you must set up initial quotas for Recommended starting values are:

CPUs: 200

In-use IP addresses: 200

Standard persistent disk: 50 TB

Still think its A?

upvoted 4 times

**ESP\_SAP** Highly Voted 4 years, 2 months ago

Correct Answer is (D):

Introduction to external data sources

This page provides an overview of querying data stored outside of BigQuery.

<https://cloud.google.com/bigquery/external-data-sources>

upvoted 28 times

**ESP\_SAP** 4 years, 2 months ago

BigQuery offers support for querying data directly from:

Bigtable

Cloud Storage

Google Drive

Cloud SQL (beta)

upvoted 5 times

**djgodzilla** 3 years, 4 months ago

but here we're not talking about joining Cloud Storage and Cloud Bigtable external tables.

the join happens between a distributed relational database (Spanner) and key-value NoSQL Database (BigTable) .

how's converting Spanner to cloud storage an implicit and trivial step.

upvoted 1 times

**djgodzilla** 3 years, 4 months ago

"The Cloud Spanner to Cloud Storage Text template is a batch pipeline that reads in data from a Cloud Spanner table, optionally transforms the data via a JavaScript User Defined Function (UDF) that you provide, and writes it to Cloud Storage as CSV text files."

<https://cloud.google.com/dataflow/docs/guides/templates/provided-batch#cloudspannertogcstext>

"The Dataflow connector for Cloud Spanner lets you read data from and write data to Cloud Spanner in a Dataflow pipeline"

<https://cloud.google.com/spanner/docs/dataflow-connector>

upvoted 3 times

**ryzior** 2 years, 7 months ago

update:

BigQuery supports the following external data sources:

Bigtable

Cloud Spanner

Cloud SQL

Cloud Storage

Drive

upvoted 6 times

**[Removed]** 4 years, 1 month ago

A parameter named D is the answer

As per your comment D is the answer.

I also agree.

But can BigQuery read backed up data? , as we have backup data on Cloud storage, did not get any evidence in the link you shared.

upvoted 2 times

**kuracpalac** **Most Recent** 8 months, 1 week ago

**Selected Answer: B**

The Q says that an analyst wants to analyze data about a user from 2 different sources, which Dataflow will give you, plus as Google states, it allows you more time analyzing stuff and less time fiddling with setting things up, which option D is talking about, which is wrong per the asked Q.

upvoted 1 times

**thewalker** 11 months ago

**Selected Answer: D**

D is apt and possible.

upvoted 2 times

**KC\_go\_reply** 1 year, 4 months ago

**Selected Answer: D**

BigQuery is powerful. If you have data in one of the popular sources like Cloud Storage or Bigtable, it is much more efficient - both for cost and computation - to create an external table on those data sources, than to copy their data around.

Besides that, also keep in mind that table clones and snapshots are much more efficient than full table copy etc.

upvoted 3 times

**Praxii** 1 year, 5 months ago

**Selected Answer: B**

I go for option B. As in option D, the data is backed up data and not the most recent data.

upvoted 1 times

**Bobbybash** 1 year, 8 months ago

**Selected Answer: B**

B. Create a dataflow job that copies data from Cloud Bigtable and Cloud Spanner for specific users.

To join data from Cloud Spanner and Cloud Bigtable for specific users, creating a dataflow job that copies data from both sources is the most efficient option. This approach allows you to process the data in parallel, and you can take advantage of Cloud Dataflow's autoscaling feature to handle large volumes of data. You can use Cloud Dataflow to read data from Cloud Bigtable and Cloud Spanner, join the data based on the user fields, and write the output to a new location or send it to the analyst. Option A (copying data from Cloud Storage) does not provide data from Cloud Spanner, and option C (running a Spark job on a Dataproc cluster) involves higher overhead costs. Option D (using BigQuery external tables) is not efficient for ad hoc requests, as data is exported from Spanner to Cloud Storage during backups, so there may be a delay in data availability.

upvoted 2 times

**anolive** 1 year, 11 months ago

**Selected Answer: D**

I think is D, but not 100% sure, because D does not have any information about the specific user like others options.

upvoted 1 times

**Charumathi** 2 years ago

**Selected Answer: D**

D is the correct answer,

An external data source is a data source that you can query directly from BigQuery, even though the data is not stored in BigQuery storage.

BigQuery supports the following external data sources:

Amazon S3  
Azure Storage  
Cloud Bigtable  
Cloud Spanner  
Cloud SQL  
Cloud Storage  
Drive

upvoted 2 times

**DualCore573** 2 years ago

**Selected Answer: D**

D makes sense as the BigQuery external tables are made for such use cases, and "efficient" keyword makes sense to use

D makes sense as the BigQuery external tables are made for such use cases, and efficient keyword makes sense to use this way as resources used are less.

upvoted 1 times

**soaresleo** 2 years, 2 months ago

First of all, using Dataflow can perhaps be effective, but NOT efficient, specially because of costs.

Second:

"To query Cloud Bigtable data using a permanent external table, you: Create a table definition file (for the API or bq command-line tool); Create a table in BigQuery linked to the external data source; Query the data using the permanent table."

Source: <https://cloud.google.com/bigquery/docs/external-data-bigtable#:~:text=To%20query%20Cloud%20Bigtable%20data,data%20using%20the%20permanent%20table>

Third:

"To query a Cloud Storage external data source, provide the Cloud Storage URI path to your data and create a table that references the data source."

Source:

<https://cloud.google.com/bigquery/docs/external-data-cloud-storage>

Correct answer: D.

upvoted 1 times

**jeffangel28** 2 years, 2 months ago

**Selected Answer: D**

"efficiently as possible" -> use the least amount of resources and achieve the same result... so I think it's D

upvoted 1 times

**tomis2** 2 years, 3 months ago

**Selected Answer: D**

Most "cloud" solution is D

upvoted 1 times

**sabbella** 2 years, 7 months ago

**Selected Answer: D**

option d

upvoted 1 times

**sabbella** 2 years, 7 months ago

**Selected Answer: D**

Option is D

upvoted 1 times

**rljjhk** 2 years, 8 months ago

**Selected Answer: B**

I think it is B.

The data in Cloud storage is not up to date as backup window is daily. SO, there are chances is missing one day worth of data.

As it is mentioned as "efficiently" instead of quickly, I would choose "B".

upvoted 3 times

**obeythefist** 2 years, 7 months ago

How does this create a "join" between the two tables?

upvoted 1 times

**BigQuery** 2 years, 7 months ago

why do you think one cannot join 2 subsets of data in dataflow Its meant for processing sets of data.

upvoted 1 times

**sasithra** 2 years, 8 months ago

B is correct

upvoted 1 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.

---