Google Discussions

EXAM PROFESSIONAL DATA ENGINEER TOPIC 1 QUESTION 125 DISCUSSION

Actual exam question from Google's Professional Data Engineer
Question #: 125
Topic #: 1
[All Professional Data Engineer Questions]

You have a petabyte of analytics data and need to design a storage and processing platform for it. You must be able to perform data warehouse-style analytics on the data in Google Cloud and expose the dataset as files for batch analysis tools in other cloud providers. What should you do?

A. Store and process the entire dataset in BigQuery.

B. Store and process the entire dataset in Bigtable.

C. Store the full dataset in BigQuery, and store a compressed copy of the data in a Cloud Storage bucket.

D. Store the warm data as files in Cloud Storage, and store the active data in BigQuery. Keep this ratio as 80% warm and 20% active.

Show Suggested Answer

by [deleted] at *March 22, 2020, 12:07 p.m.*

Comments

Type your comment...

Submit

Rajokkiyam Highly Voted 4 years, 1 month ago

Answer C.

upvoted 34 times

AJKumar Highly Voted 3 years, 10 months ago

A and B can be eliminated right away as they do not talk about providing for other cloud providers. between C and D. The question says nothing about warm or cold data-rather that data should be made available for other providers--C--can fulfill this condition. Answer C.

upvoted 24 times

AzureDP900 1 year, 4 months ago

Agree with C

upvoted 1 times

zbyszek1 Most Recent 7 months, 2 weeks ago

For me A. I can use export from BQ to Cloud Storage. There is no need to store two copies of data.

upvoted 1 times

spicebits 5 months, 4 weeks ago

If you export data from BQ to GCS then you will have two copies and you will be in the same architecture as answer C.

upvoted 4 times

vamgcp 9 months, 1 week ago

Selected Answer: B

It can be C or D , but I will go with C as storing the full dataset in BigQuery and a compressed copy of the data in Cloud Storage is a good way to balance performance and cost.

upvoted 2 times

forepick 11 months, 1 week ago

Selected Answer: C

Best answer is C, although BQ can query gzipped files stored on GCS directly.
Maybe this double storage makes it a bit more highly available.

upvoted 2 times

izekc 12 months ago

Selected Answer: D

D is much more accurate.

upvoted 1 times

jkhong 1 year, 4 months ago

Selected Answer: C

D → does not guarantee 100% queryable or accessible/available

upvoted 1 times

zellck 1 year, 5 months ago

Selected Answer: C

C is the answer.

upvoted 1 times

Smaks 1 year, 9 months ago

You can read streaming data from Pub/Sub, and you can write streaming data to Pub/Sub or BigQuery.
Thus Cloud Storage is not a proper sink for streaming pipeline.
I vote for B, since it is possible to convert unstructured data and store in BQ

upvoted 1 times

Smaks 1 year, 9 months ago

ignore this comment, please

upvoted 10 times

Aslkdup 2 years, 2 months ago

BQ can reach files at google storage as external table. so my answer is D. (If data was smaller than this, I would choose C)

upvoted 1 times

Bhawantha 2 years, 3 months ago

Selected Answer: C

both requirements are full filled.

upvoted 2 times

MaxNRG 2 years, 3 months ago

Selected Answer: D

D: BigQuery + Cloud Storage

upvoted 1 times

jkhong 1 year, 4 months ago

D → does not guarantee 100% queryable or accessible/available

upvoted 2 times

medeis_jar 2 years, 3 months ago

Selected Answer: C

"You must be able to perform data warehouse-style analytics on the data in Google Cloud and expose the dataset as files for batch analysis tools in other cloud providers?"
Analytics -> BQ
Exposing -> GCS

upvoted 7 times

JG123 2 years, 5 months ago

Correct: C

upvoted 2 times

xiaofeng_0226 2 years, 9 months ago

vote for C

upvoted 3 times

sumanshu 2 years, 10 months ago

Vote for 'C'

A - Only Half requirement fulfil, expose as a file not getting fulfiled
B - Not a warehouse
C. Both requirements fulfiled...Bigquery and GCS
D. Both requirement fulfiled...but what if other cloud provider wants to analysis on rest 80% of the data. -

So out of 4 options, C looks okay

upvoted 8 times

gcper 3 years, 2 months ago

C

BigQuery for analytics processing and Cloud Storage for exposing the data as files

upvoted 3 times

Load full discussion...