EXAM PROFESSIONAL DATA ENGINEER TOPIC 1 QUESTION 182 DISCUSSION

Actual exam question from Google's Professional Data Engineer
Question #: 182
Topic #: 1
[All Professional Data Engineer Questions]

You are migrating your data warehouse to Google Cloud and decommissioning your on-premises data center. Because this is a priority for your company, you know that bandwidth will be made available for the initial data load to the cloud. The files being transferred are not large in number, but each file is 90 GB.
Additionally, you want your transactional systems to continually update the warehouse on Google Cloud in real time. What tools should you use to migrate the data and ensure that it continues to write to your warehouse?

A. Storage Transfer Service for the migration; Pub/Sub and Cloud Data Fusion for the real-time updates

B. BigQuery Data Transfer Service for the migration; Pub/Sub and Dataproc for the real-time updates

C. gsutil for the migration; Pub/Sub and Dataflow for the real-time updates

D. gsutil for both the migration and the real-time updates

Show Suggested Answer

by AWSandeep at *Sept. 2, 2022, 9:55 p.m.*

Comments

Type your comment...

Submit

zellck Highly Voted 1 year, 11 months ago

Selected Answer: C

C is the answer.

https://cloud.google.com/architecture/migration-to-google-cloud-transferring-your-large-datasets#gsutil_for_smaller_transfers_of_on-premises_data
The gsutil tool is the standard tool for small- to medium-sized transfers (less than 1 TB) over a typical enterprise-scale network, from a private data center to Google Cloud.

upvoted 11 times

AzureDP900 1 year, 10 months ago

Agreed

thx for sharing link

upvoted 1 times

musumusu 1 year, 8 months ago

what is wrong with A, there is no cost constraint

upvoted 1 times

AWSandeep Highly Voted 2 years, 2 months ago

Selected Answer: C

C. gsutil for the migration; Pub/Sub and Dataflow for the real-time updates

Use Gsutil when there is enough bandwidth to meet your project deadline for less than 1 TB of data. Storage Transfer Service is for much larger volumes for migration. Moreover, Cloud Data Fusion and Dataproc are not ideal for real-time updates. BigQuery Data Transfer Service does not support all on-prem sources.

upvoted 8 times

shangning007 Most Recent 4 months, 2 weeks ago

Selected Answer: A

According to the latest documentation, "Generally, you should use gcloud storage commands instead of gsutil commands. The gsutil tool is a legacy Cloud Storage CLI and minimally maintained."
We should remove the presence of gsutil in the questions.

upvoted 2 times

TVH_Data_Engineer 10 months, 2 weeks ago

Selected Answer: C

Considering the requirement for handling large files and the need for real-time data integration, Option C (gsutil for the migration; Pub/Sub and Dataflow for the real-time updates) seems to be the most appropriate. gsutil will effectively handle the large file transfers, while Pub/Sub and Dataflow provide a robust solution for real-time data capture and processing, ensuring continuous updates to your warehouse on Google Cloud.

upvoted 1 times

MaxNRG 10 months, 2 weeks ago

Selected Answer: C

Option C is the best choice given the large file sizes for the initial migration and the need for real-time updates after migration.

Specifically:

gsutil can transfer large files in parallel over multiple TCP connections to maximize bandwidth. This works well for the 90GB files during initial migration.
Pub/Sub allows real-time messaging of updates that can then be streamed into Cloud Dataflow. Dataflow provides scalable stream processing to handle transforming and writing those updates into BigQuery or other sinks.

upvoted 1 times

MaxNRG 10 months, 2 weeks ago

Option A is incorrect because Storage Transfer Service is better for scheduled batch transfers, not ad hoc large migrations.

Option B is incorrect because BigQuery Data Transfer Service is more focused on scheduled replication jobs, not ad hoc migrations.

Option D would not work well for real-time updates after migration is complete.

So option C leverages the right Google cloud services for the one-time migration and ongoing real-time processing.

upvoted 2 times

xiangbobopopo 1 year ago

Selected Answer: C

agree with C

upvoted 1 times

TNT87 2 years, 1 month ago

https://cloud.google.com/architecture/migration-to-google-cloud-transferring-your-large-datasets#gsutil_for_smaller_transfers_of_on-premises_data
Answer C

upvoted 4 times

YorelNation 2 years, 2 months ago

Selected Answer: C

C seems legit

upvoted 3 times