

 Google Discussions



Exam Professional Data Engineer All Questions

View all questions & answers for the Professional Data Engineer exam

[Go to Exam](#)



EXAM PROFESSIONAL DATA ENGINEER TOPIC 1 QUESTION 299 DISCUSSION

Actual exam question from Google's Professional Data Engineer

Question #: 299

Topic #: 1

[\[All Professional Data Engineer Questions\]](#)

You have an upstream process that writes data to Cloud Storage. This data is then read by an Apache Spark job that runs on Dataproc. These jobs are run in the us-central1 region, but the data could be stored anywhere in the United States. You need to have a recovery process in place in case of a catastrophic single region failure. You need an approach with a maximum of 15 minutes of data loss (RPO=15 mins). You want to ensure that there is minimal latency when reading the data. What should you do?

- A. 1. Create two regional Cloud Storage buckets, one in the us-central1 region and one in the us-south1 region.
2. Have the upstream process write data to the us-central1 bucket. Use the Storage Transfer Service to copy data hourly from the us-central1 bucket to the us-south1 bucket.
3. Run the Dataproc cluster in a zone in the us-central1 region, reading from the bucket in that region.
4. In case of regional failure, redeploy your Dataproc clusters to the us-south1 region and read from the bucket in that region instead.
- B. 1. Create a Cloud Storage bucket in the US multi-region.
2. Run the Dataproc cluster in a zone in the us-central1 region, reading data from the US multi-region bucket.
3. In case of a regional failure, redeploy the Dataproc cluster to the us-central2 region and continue reading from the same bucket.
- C. 1. Create a dual-region Cloud Storage bucket in the us-central1 and us-south1 regions.
2. Enable turbo replication.
3. Run the Dataproc cluster in a zone in the us-central1 region, reading from the bucket in the us-south1 region.
4. In case of a regional failure, redeploy your Dataproc cluster to the us-south1 region and continue reading from the same bucket.

D. 1. Create a dual-region Cloud Storage bucket in the us-central1 and us-south1 regions

1. Create a dual-region Cloud Storage bucket in the us-central1 and us-south1 regions.
2. Enable turbo replication.
3. Run the Dataproc cluster in a zone in the us-central1 region, reading from the bucket in the same region.
4. In case of a regional failure, redeploy the Dataproc clusters to the us-south1 region and read from the same bucket.

Show Suggested Answer

by  scaenrui at Jan. 4, 2024, 2:34 p.m.

Comments

Type your comment...

Submit

  **raaad** Highly Voted 10 months ago

Selected Answer: D

- Rapid Replication: Turbo replication ensures near-real-time data synchronization between regions, achieving an RPO of 15 minutes or less.
- Minimal Latency: Dataproc clusters can read from the bucket in the same region, minimizing data transfer latency and optimizing performance.
- Disaster Recovery: In case of regional failure, Dataproc clusters can seamlessly redeploy to the other region and continue reading from the same bucket, ensuring business continuity.

   upvoted 6 times

  **JyoGCP** Most Recent 8 months, 2 weeks ago

Selected Answer: D

Option D

   upvoted 1 times

  **Matt_108** 9 months, 3 weeks ago

Selected Answer: D

Option D, answers all needs from the request

   upvoted 2 times

  **scaenrui** 10 months ago

Selected Answer: D

- D.
1. Create a dual-region Cloud Storage bucket in the us-central1 and us-south1 regions.
 2. Enable turbo replication.
 3. Run the Dataproc cluster in a zone in the us-central1 region, reading from the bucket in the same region.
 4. In case of a regional failure, redeploy the Dataproc clusters to the us-south1 region and read from the same bucket.

   upvoted 3 times



Platform

> Home

> Examtopics PRO

> All Exams

> Training Courses

