

🔗 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 52 DISCUSSION

Actual exam question from Google's Professional Data Engineer

Question #: 52

Topic #: 1

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

You are implementing security best practices on your data pipeline. Currently, you are manually executing jobs as the Project Owner. You want to automate these jobs by taking nightly batch files containing non-public information from Google Cloud Storage, processing them with a Spark Scala job on a Google Cloud Dataproc cluster, and depositing the results into Google BigQuery. How should you securely run this workload?

- A. Restrict the Google Cloud Storage bucket so only you can see the files
- B. Grant the Project Owner role to a service account, and run the job with it
- C. Use a service account with the ability to read the batch files and to write to BigQuery
- D. Use a user account with the Project Viewer role on the Cloud Dataproc cluster to read the batch files and write to BigQuery


[Show Suggested Answer](#)

by [rickywck](#) at March 17, 2020, 5:01 a.m.

Comments

Type your comment...

[Submit](#)

  **digvijay** Highly Voted  4 years, 1 month ago

A is wrong, if only I can see the bucket no automation is possible, besides, also needs launch the dataproc job
B is too much, does not follow the security best practices
C has one point missing...you need to submit dataproc jobs.
In D viewer role will not be able to submit dataproc jobs, the rest is ok

Thus....the only one that would work is B! BUT this service account has too many permissions. Should have dataproc editor, write big query and read from bucket

   upvoted 34 times

  **retep007** 2 years, 7 months ago

C doesn't need permission to submit dataproc jobs, it's workload SA. Job can be submitted by any other identity

   upvoted 5 times

  **dambilwa** 3 years, 10 months ago




Hence - Contextually, Option [C] looks to be the right fit

   upvoted 15 times

  **rickywck** Highly Voted  4 years, 1 month ago

Should be C

   upvoted 31 times

  **Mathew106** Most Recent  9 months, 2 weeks ago

Selected Answer: B

We need permissions for submitting dataproc jobs and writing to BigQuery. Project Owner will fix all of that even though it's not a good solution. The rest won't work at all.

   upvoted 1 times

  **Adswerve** 1 year ago

Selected Answer: C

C
Project Owner is too much, violates the principle of least privilege

   upvoted 4 times


  **PolyMoe** 1 year, 3 months ago

Selected Answer: C

C. Use a service account with the ability to read the batch files and to write to BigQuery

It is best practice to use service accounts with the least privilege necessary to perform a specific task when automating jobs. In this case, the job needs to read the batch files from Cloud Storage and write the results to BigQuery. Therefore, you should create a service account with the ability to read from the Cloud Storage bucket and write to BigQuery, and use that service account to run the job.

   upvoted 3 times

  **Mkumar43** 1 year, 4 months ago

Selected Answer: B



B works for the given requirement

   upvoted 1 times

  **Krish6488** 1 year, 4 months ago

Least privilege principle. Option C. job can be submitted or triggered using a Cron or a composer which uses another SA with different set of privileges

   upvoted 2 times

  **DGames** 1 year, 4 months ago

Selected Answer: B


B because we need to run job .. option C mentioned permission about read and write nothing mention to run the job . In case project owner to service account it's similar just running job and doing rest of tasks read and writing as well.

   upvoted 2 times

  **ThomasChoy** 2 years ago

Selected Answer: C

The answer is C because Service Account is the best way to access the BigQuery API if your application can run jobs associated with service credentials rather than an end-user's credentials, such as a batch processing pipeline.
<https://cloud.google.com/bigquery/docs/authentication>

   upvoted 2 times

  **Bhawantha** 2 years, 3 months ago

Selected Answer: C

Data owners cant create jobs or queries. -> B out
We need service Account -> D out
Access only granting me does not solve the problem -> A out
The answer is C. (Minimum rights to perform the job)

👍 ↩ 🚩 upvoted 4 times

🗄️ 👤 **medeis_jar** 2 years, 4 months ago

Selected Answer: C

"taking nightly batch files containing non-public information from Google Cloud Storage, processing them with a Spark Scala job on a Google Cloud Dataproc cluster, and depositing the results into Google BigQuery"

👍 ↩ 🚩 upvoted 1 times

🗄️ 👤 **prasanna77** 2 years, 4 months ago

C should be okay,since he is already a project owner, I guess compute service account created will have access to run the jobs

👍 ↩ 🚩 upvoted 1 times

🗄️ 👤 **MaxNRG** 2 years, 4 months ago

Selected Answer: C

C,
Project Owner role to a service account - is too much

👍 ↩ 🚩 upvoted 1 times

🗄️ 👤 **JG123** 2 years, 5 months ago

Why there are so many wrong answers? Examtopics.com are you enjoying paid subscription by giving random answers from people?

Ans: C

👍 ↩ 🚩 upvoted 6 times

🗄️ 👤 **anji007** 2 years, 6 months ago

Ans: C

See this: https://cloud.google.com/dataproc/docs/concepts/configuring-clusters/service-accounts#dataproc_service_accounts_2

👍 ↩ 🚩 upvoted 3 times

🗄️ 👤 **Blobby** 2 years, 7 months ago

C as service account invoked to read the data into GCS and write to BQ once transformed via Data Proc. Assumes DataProc can inherit SA authorisation to perform transform and propagate.

B seems to violate key IAM principle enforcing least privilege;
<https://cloud.google.com/iam/docs/recommender-overview>

👍 ↩ 🚩 upvoted 4 times

🗄️ 👤 **sumanshu** 2 years, 10 months ago

Vote for 'C'

👍 ↩ 🚩 upvoted 4 times

🗄️ 👤 **sumanshu** 2 years, 10 months ago

Vote for B, (though it's too much access) - But C has one accessing missing (i.e Dataproc job execution) Thus B is correct

👍 ↩ 🚩 upvoted 3 times

[Load full discussion...](#)



Platform

> [Home](#)

> [Examtopics PRO](#)

> [All Exams](#)

> [Training Courses](#)

