

■ MENU

C

G 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 104 DISCUSSION

Actual exam question from Google's Professional Data Engineer

Question #: 104

Topic #: 1

[All Professional Data Engineer Questions]

You used Dataprep to create a recipe on a sample of data in a BigQuery table. You want to reuse this recipe on a daily upload of data with the same schema, after the load job with variable execution time completes. What should you do?

- A. Create a cron schedule in Dataprep.
- B. Create an App Engine cron job to schedule the execution of the Dataprep job.
- C. Export the recipe as a Dataprep template, and create a job in Cloud Scheduler.
- D. Export the Dataprep job as a Dataflow template, and incorporate it into a Composer job.

Show Suggested Answer

by Admaldon at Sept. 2, 2022, 9:40 a.m.

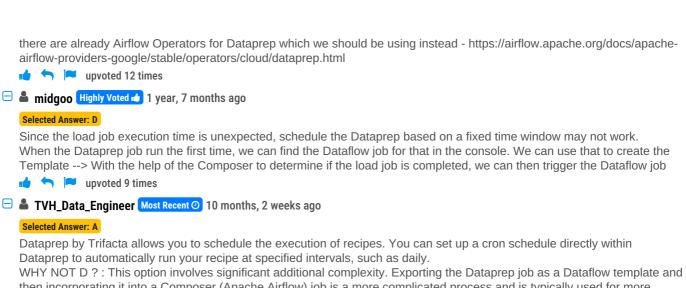
Comments

Type your comment...

Submit

☐ ♣ jkhong Highly Voted 1 1 year, 10 months ago

I'd pick D because it's the only option which allows variable execution (since we need to execute the dataprep job only after the prior load job). Although D suggests the export of Dataflow templates, this discussion suggests that the export option is no longer available (https://stackoverflow.com/questions/72544839/how-to-get-the-dataflow-template-of-a-dataprep-job),



then incorporating it into a Composer (Apache Airflow) job is a more complicated process and is typically used for more complex orchestration needs that go beyond simple scheduling.

upvoted 2 times

MaxNRG 10 months, 3 weeks ago

Selected Answer: D

We have external dependency "after the load job with variable execution time completes" which requires DAG -> Airflow (Cloud Composer)

The reasons:

A scheduler like Cloud Scheduler won't handle the dependency on the BigQuery load completion time

Using Composer allows creating a DAG workflow that can:

Trigger the BigQuery load

Wait for BigQuery load to complete

Trigger the Dataprep Dataflow job

Dataflow template allows easy reuse of the Dataprep transformation logic

Composer coordinates everything based on the dependencies in an automated workflow

upvoted 1 times

□ ♣ rocky48 11 months ago

Selected Answer: D

I'd pick D because it's the only option which allows variable execution

upvoted 1 times

😑 🏜 gaurav0480 1 year, 2 months ago

The key here is "after the load job with variable execution time completes" which means the execution of this job depends on the completion of another job which has a variable execution time. Hence D

upvoted 3 times

god_brainer 1 year, 2 months ago

This approach ensures the dynamic triggering of the Dataprep job based on the completion of the preceding load job, ensuring data is processed accurately and in sequen

upvoted 1 times

Adswerve 1 year, 6 months ago

Selected Answer: A

A is correct. D is too complicated.

A is correct, because you can schedule a job right from Dataprep UI.

https://cloud.google.com/blog/products/gcp/scheduling-and-sampling-arrive-for-google-cloud-dataprep Scheduling and sampling arrive for Google Cloud Dataprep

Throughout our early releases, users' most common request has been Flow scheduling. As of Thursday's release, Flows can be scheduled with minute granularity at any frequency.

upvoted 2 times

🖃 🏜 lucaluca1982 1 year, 7 months ago

Selected Answer: C

I think C it is more straighforward

upvoted 4 times

🖃 🏜 musumusu 1 year, 8 months ago

Answer C: Use Recipe Template feature of dataprep. Don't need to change the service.

