

🔗 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 [damaldon](#) at Sept. 2, 2022, 9:40 a.m.

### Comments

Type your comment...

[Submit](#)

👤 [jkhong](#) Highly Voted 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>),

there are already Airflow Operators for Dataprep which we should be using instead - <https://airflow.apache.org/docs/apache-airflow-providers-google/stable/operators/cloud/dataprep.html>

👍 ↩ 🚩 upvoted 12 times

🗄️ 👤 **midgoo** Highly Voted 1 year, 7 months ago

**Selected Answer: D**

Since the load job execution time is unexpected, schedule the Dataprep based on a fixed time window may not work. When the Dataprep job run the first time, we can find the Dataflow job for that in the console. We can use that to create the Template --> With the help of the Composer to determine if the load job is completed, we can then trigger the Dataflow job

👍 ↩ 🚩 upvoted 9 times

🗄️ 👤 **TVH\_Data\_Engineer** Most Recent 10 months, 2 weeks ago

**Selected Answer: A**

Dataprep by Trifacta allows you to schedule the execution of recipes. You can set up a cron schedule directly within Dataprep to automatically run your recipe at specified intervals, such as daily. WHY NOT D ? : This option involves significant additional complexity. Exporting the Dataprep job as a Dataflow template and 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 straightforward

👍 ↩ 🚩 upvoted 4 times

🗄️ 👤 **musumusu** 1 year, 8 months ago

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

👍 ↩ 🚩 upvoted 3 times

🗨️ 👤 **jroig\_** 1 year, 9 months ago

**Selected Answer: C**

Why not C?

👍 ↩ 🚩 upvoted 1 times

🗨️ 👤 **zelck** 1 year, 11 months ago

**Selected Answer: D**

D is the answer.

👍 ↩ 🚩 upvoted 1 times

🗨️ 👤 **anicloudgirl** 1 year, 11 months ago

**Selected Answer: A**

It's A. You can set it directly in Dataprep a job and it will use Dataflow under the hood.

👍 ↩ 🚩 upvoted 4 times

🗨️ 👤 **anicloudgirl** 1 year, 11 months ago

It's A. You can set it directly in Dataprep a job and it will use Dataflow under the hood. No need to export nor incorporate into a Composer job.

Dataprep by trifacta - <https://docs.trifacta.com/display/DP/cron+Schedule+Syntax+Reference>

Dataprep job uses dataflow - <https://cloud.google.com/dataprep>

👍 ↩ 🚩 upvoted 2 times

🗨️ 👤 **jkhong** 1 year, 10 months ago

The question mentions after a load job with variable time, i dont think setting a dataprep cron job can address the issue of variable load times

👍 ↩ 🚩 upvoted 4 times

🗨️ 👤 **cloudmon** 1 year, 12 months ago

**Selected Answer: D**

It's D

👍 ↩ 🚩 upvoted 2 times

🗨️ 👤 **John\_Pongthorn** 2 years, 1 month ago

**Selected Answer: D**

Dataprep and Dataflow are same family

👍 ↩ 🚩 upvoted 2 times

🗨️ 👤 **AWSandeep** 2 years, 2 months ago

**Selected Answer: D**

D. Export the Dataprep job as a Dataflow template, and incorporate it into a Composer job.

Reveal Solution

👍 ↩ 🚩 upvoted 4 times

[Load full discussion...](#)



## Platform

> Home

> Examtopics PRO

> All Exams

> Training Courses

