

[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 116 DISCUSSION

Actual exam question from Google's Professional Data Engineer

Question #: 116

Topic #: 1

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

Your company is in the process of migrating its on-premises data warehousing solutions to BigQuery. The existing data warehouse uses trigger-based change data capture (CDC) to apply updates from multiple transactional database sources on a daily basis. With BigQuery, your company hopes to improve its handling of CDC so that changes to the source systems are available to query in BigQuery in near-real time using log-based CDC streams, while also optimizing for the performance of applying changes to the data warehouse. Which two steps should they take to ensure that changes are available in the BigQuery reporting table with minimal latency while reducing compute overhead? (Choose two.)

- A. Perform a DML INSERT, UPDATE, or DELETE to replicate each individual CDC record in real time directly on the reporting table.
- B. Insert each new CDC record and corresponding operation type to a staging table in real time.
- C. Periodically DELETE outdated records from the reporting table.
- D. Periodically use a DML MERGE to perform several DML INSERT, UPDATE, and DELETE operations at the same time on the reporting table.
- E. Insert each new CDC record and corresponding operation type in real time to the reporting table, and use a materialized view to expose only the newest version of each unique record.

[Show Suggested Answer](#)

by [ducc](#) at Sept. 3, 2022, 6:25 a.m.

## Comments

Type your comment...

Submit

 **YorelNation** Highly Voted  2 years, 8 months ago



**Selected Answer: BD**

To aim for minimal latency while reducing compute overhead:

B. Insert each new CDC record and corresponding operation type to a staging table in real time.

D. Periodically use a DML MERGE to perform several DML INSERT, UPDATE, and DELETE operations at the same time on the reporting table. (all statements comes from the staging table)

   upvoted 15 times

 **musumusu** Highly Voted  2 years, 2 months ago

B&D

Tricks here: Always choose google recommended approach, Use data first in Staging table then merge with original tables.

   upvoted 6 times

 **Nirca** Most Recent  1 year, 7 months ago

**Selected Answer: CE**

I'm going for E & C - this is the only solution with low TCO.

E - is the best way to work with CDC when real nearline data is needed BQ snapshots can be online! . & C - is good practice to delete old records.

   upvoted 2 times

 **nadavw** 8 months, 2 weeks ago

A isn't correct, as the requirement is "reducing compute overhead"

B isn't correct, as there is no mention of a "staging table" in the scenario

D isn't correct as it's done periodically, and the requirement is "near real-time"


   upvoted 1 times

 **dconesoko** 2 years, 4 months ago

**Selected Answer: BD**

with both the delta table and the main table changes could be queried in near realtime, by using a view that unions both tables and queries the latest record for the given key, eventually the delta table should be merged into the main table and truncated. Google recently introduced datastream that would take away all these headaches.

   upvoted 2 times

 **odacir** 2 years, 4 months ago

**Selected Answer: BD**

The solution is B and D. I perform a similar task in my work, and this is the best way to do it at scale with BigQuery.

   upvoted 2 times

 **zellck** 2 years, 5 months ago

**Selected Answer: BD**

BD is the answer.

[https://cloud.google.com/architecture/database-replication-to-bigquery-using-change-data-capture#overview\\_of\\_cdc\\_data\\_replication](https://cloud.google.com/architecture/database-replication-to-bigquery-using-change-data-capture#overview_of_cdc_data_replication)

Delta tables contain all change events for a particular table since the initial load. Having all change events available can be valuable for identifying trends, the state of the entities that a table represents at a particular moment, or change frequency.

The best way to merge data frequently and consistently is to use a MERGE statement, which lets you combine multiple INSERT, UPDATE, and DELETE statements into a single atomic operation.

   upvoted 4 times

 **NicolasN** 2 years, 5 months ago

I really can't find a correct combination of answers. I'm between the following alternatives, but with no one fitting:

1 [B] and [D]: That's a proposed solution, but as a cost-optimized approach (along with an extra step to "Periodically DELETE outdated records from the STAGING table" - more details on my subsequent reply). Also, I can't imagine how an answer with the word "Periodically" may be compatible with the "minimal latency" requirement.

2 [E] and [C]: It could be a valid approach, but near-real time requirement would demand also for a materialized view refresh. And it seems to contradict the "reducing compute overhead" req.

3 [A] standalone: Provides immediate results but is far from compute-optimized.

upvoted 2 times

**NicolasN** 2 years, 5 months ago

The previous guidelines were here:

[https://cloud.google.com/architecture/database-replication-to-bigquery-using-change-data-capture#immediate\\_consistency\\_approach](https://cloud.google.com/architecture/database-replication-to-bigquery-using-change-data-capture#immediate_consistency_approach)

There were two approaches:

- 1 Immediate consistency approach
- 2 Cost-optimized approach

For approach 1, which is the objective of this question, it proposes:

- a. Insert CDC data into a delta table in BigQuery => that's answer [B]
- b. Create a BigQuery view that joins the main and delta tables and finds the most recent row => there's no answer that fits

For approach 2 it proposes:

- a. Insert CDC data into a delta table in BigQuery => that's answer [B]
- b. Merge delta table changes into the main table and periodically purge merged rows from the delta table - Run Merge statement on a regular interval => that's answer [D]

upvoted 5 times

**NicolasN** 2 years, 5 months ago

Nowadays (Nov. 2022) I don't expect to confront this question in a real exam with this set of answers since the more recent documentation proposes the use of Datastream.

<https://cloud.google.com/blog/products/data-analytics/real-time-cdc-replication-bigquery>

upvoted 4 times

**beanz00** 2 years, 6 months ago

B and E. Typically in a Data Warehouse you don't delete data. Data Warehouse should store full history to see how the data changed over time. All the solutions with 'DELETE' should not be used as this goes against being able to access the history of the data.

upvoted 2 times

**TNT87** 2 years, 6 months ago

<https://www.striim.com/blog/oracle-to-google-bigquery/>

upvoted 1 times

**TNT87** 2 years, 7 months ago

[https://cloud.google.com/architecture/database-replication-to-bigquery-using-change-data-capture#data\\_latency](https://cloud.google.com/architecture/database-replication-to-bigquery-using-change-data-capture#data_latency)

upvoted 1 times

**TNT87** 2 years, 7 months ago

<https://docs.streamsets.com/platform-datacollector/latest/datacollector/UserGuide/Destinations/GBigQuery.html>

upvoted 1 times

**Wasss123** 2 years, 7 months ago

**Selected Answer: BD**

B and D

<https://cloud.google.com/architecture/database-replication-to-bigquery-using-change-data-capture>

upvoted 4 times

**John\_Pongthorn** 2 years, 7 months ago

**Selected Answer: BC**

B D you have to do the both to get it done.

To merge process, you have to perform between the report table and stage a table

upvoted 2 times

**Remi2021** 2 years, 7 months ago

Answers are tricky, official documentation suggests Dataflow or Datafusion path as well as inclusion of DataStreams

<https://cloud.google.com/blog/products/data-analytics/real-time-cdc-replication-bigquery>

upvoted 1 times

**changsu** 2 years, 8 months ago

**Selected Answer: CE**

It costs more to update/delete.

upvoted 2 times

**ducc** 2 years, 8 months ago

**Selected Answer: BE**

BE is correct

Big Query only need to capture change, no need DELETE, UPDATE

upvoted 1 times



## Platform

> [Home](#)

> [Examtopics PRO](#)

> [All Exams](#)

> [Training Courses](#)

