← 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 102 DISCUSSION**

Actual exam question from Google's Professional Data Engineer

Question #: 102

Topic #: 1

[All Professional Data Engineer Questions]

You need to create a near real-time inventory dashboard that reads the main inventory tables in your BigQuery data warehouse. Historical inventory data is stored as inventory balances by item and location. You have several thousand updates to inventory every hour. You want to maximize performance of the dashboard and ensure that the data is accurate. What should you do?

A. Leverage BigQuery UPDATE statements to update the inventory balances as they are changing.

B. Partition the inventory balance table by item to reduce the amount of data scanned with each inventory update.

C. Use the BigQuery streaming the stream changes into a daily inventory movement table. Calculate balances in a view that joins it to the historical inventory balance table. Update the inventory balance table nightly.

D. Use the BigQuery bulk loader to batch load inventory changes into a daily inventory movement table. Calculate balances in a view that joins it to the historical inventory balance table. Update the inventory balance table nightly.

Show Suggested Answer

by 👤 **Rajokkiyam** at *March 22, 2020, 6:37 a.m.*

Comments

Type your comment...

Submit

☐ 👤 **MaxNRG** `Highly Voted 👍` 3 years, 4 months ago

A - New correct answer
C - Old correct answer (for 2019)

👍 ↩ 🚩 upvoted 33 times

☐ 👤 **MaxNRG** 1 year, 4 months ago

C is better

The best approach is to use BigQuery streaming to stream the inventory changes into a daily inventory movement table. Then calculate balances in a view that joins the inventory movement table to the historical inventory balance table. Finally, update the inventory balance table nightly (option C).

👍 ↩ 🚩 upvoted 4 times

☐ 👤 **MaxNRG** 1 year, 4 months ago

The key reasons this is better than the other options:

Using BigQuery UPDATE statements (option A) would be very inefficient for thousands of updates per hour. It is better to batch updates.
Partitioning the inventory balance table (option B) helps query performance, but does not solve the need to incrementally update balances.
Using the bulk loader (option D) would require batch loading the updates, which adds latency. Streaming inserts updates with lower latency.
So option C provides a scalable architecture that streams updates with low latency while batch updating the balances only once per day for efficiency. This balances performance and accuracy needs.

👍 ↩ 🚩 upvoted 3 times

☐ 👤 **MaxNRG** 1 year, 4 months ago

Here's why the other options are less suitable:

A. Leverage BigQuery UPDATE statements: While technically possible, this approach is inefficient for frequent updates as it requires individual record scans and updates, affecting performance and potentially causing data race conditions.

B. Partition the inventory balance table: Partitioning helps with query performance for large datasets, but it doesn't address the need for near real-time updates.

D. Use the BigQuery bulk loader: Bulk loading daily changes is helpful for historical data ingestion, but it won't provide near real-time updates necessary for the dashboard.

👍 ↩ 🚩 upvoted 1 times

☐ 👤 **MaxNRG** 1 year, 4 months ago

Option C offers the following advantages:

Streams inventory changes near real-time: BigQuery streaming ingests data immediately, keeping the inventory movement table constantly updated.
Daily balance calculation: Joining the movement table with the historical balance table provides an accurate view of current inventory levels without affecting the actual balance table.
Nightly update for historical data: Updating the main inventory balance table nightly ensures long-term data consistency while maintaining near real-time insights through the view.
This approach balances near real-time updates with efficiency and data accuracy, making it the optimal solution for the given scenario.

👍 ↩ 🚩 upvoted 1 times

☐ 👤 **Yiouk** 1 year, 9 months ago

There are still limitations on DML statements (2023) e.g. only 2 concurrent UPDATES and up to 20 queued hence not appropriate for this scenario:
https://cloud.google.com/bigquery/quotas#data-manipulation-language-statements

👍 ↩ 🚩 upvoted 2 times

☐ 👤 **NeoNitin** 1 year, 9 months ago

option A:what limitation here 1500/perday okay in question we will get max 24 jobs hourly updated okay,
now speed 5 operation /10 sec , 1 operation 2sec , and we are getting new update in 1 hour so we have time 3600 sec and we need to update around 1000 update according to speed take 2000sec still we have 1600 sec rest to getting new update so .
thats why I thing DML is best option for this work

👍 ↩ 🚩 upvoted 2 times

☐ 👤 **Nandababy** 1 year, 4 months ago

In question it mentioned several thousands of updates every hour, several thousands could be 20-30 thousands as well. Where it is mentioned for only 1000 updates?

👍 ↩ 🏳 upvoted 1 times

☐ 👤 **haroldbenites** Highly Voted 👍 4 years, 8 months ago

C is correct.
It says "update Every hour"
And need " accuracy"

👍 ↩ 🏳 upvoted 25 times

☐ 👤 **NeoNitin** 1 year, 9 months ago

option A:what limitation here 1500/perday okay in question we will get max 24 jobs hourly updated okay,
now speed 5 operation /10 sec , 1 operation 2sec , and we are getting new update in 1 hour so we have time 3600 sec
and we need to update around 1000 update according to speed take 2000sec still we have 1600 sec rest to getting new
update so .
thats why I thing DML is best option for this work

👍 ↩ 🏳 upvoted 2 times

☐ 👤 **SamuelTsch** Most Recent ⊙ 6 months, 2 weeks ago

Selected Answer: C

BigQuery is not optimized for updating statement. So, go to C.

👍 ↩ 🏳 upvoted 1 times

☐ 👤 **edre** 9 months, 2 weeks ago

Selected Answer: C

The answer is C because the requirement is near real-time

👍 ↩ 🏳 upvoted 1 times

☐ 👤 **MaxNRG** 1 year, 4 months ago

Selected Answer: C

The best approach is to use BigQuery streaming to stream the inventory changes into a daily inventory movement table.
Then calculate balances in a view that joins the inventory movement table to the historical inventory balance table. Finally,
update the inventory balance table nightly (option C).

👍 ↩ 🏳 upvoted 1 times

☐ 👤 **MaxNRG** 1 year, 4 months ago

The key reasons this is better than the other options:

Using BigQuery UPDATE statements (option A) would be very inefficient for thousands of updates per hour. It is better to
batch updates.

Partitioning the inventory balance table (option B) helps query performance, but does not solve the need to incrementally
update balances.

Using the bulk loader (option D) would require batch loading the updates, which adds latency. Streaming inserts updates
with lower latency.

So option C provides a scalable architecture that streams updates with low latency while batch updating the balances only
once per day for efficiency. This balances performance and accuracy needs.

👍 ↩ 🏳 upvoted 1 times

☐ 👤 **rocky48** 1 year, 5 months ago

Selected Answer: C

Option C.

Using the BigQuery streaming to stream changes into a daily inventory movement table and calculating balances in a view
that joins it to the historical inventory balance table can help you achieve the desired performance and accuracy. You can
then update the inventory balance table nightly. This approach can help you avoid the overhead of scanning large amounts
of data with each inventory update, which can be time-consuming and resource-intensive.
Leveraging BigQuery UPDATE statements to update the inventory balances as they are changing (option A) can be
resource-intensive and may not be the most efficient way to achieve the desired performance.

👍 ↩ 🏳 upvoted 3 times

☐ 👤 **AnonymousPanda** 1 year, 5 months ago

Selected Answer: C

As per other answers C

👍 ↩ 🏳 upvoted 1 times

☐ 👤 **Nirca** 1 year, 6 months ago

Selected Answer: A

Simple and will work

👍 ↩ 🚩 upvoted 1 times

⊟ 👤 **odacir** 1 year, 6 months ago

Selected Answer: C

Answer is C. Why because "Update" limits is 1500/per day, and the question say: You have several thousand updates to inventory every hour. So is impossible to use updates all the time.

👍 ↩ 🚩 upvoted 2 times

⊟ 👤 **Nirca** 1 year, 7 months ago

Selected Answer: A

A. Leverage BigQuery UPDATE statements to update the inventory balances as they are changing - is so simple and RIGHT!

👍 ↩ 🚩 upvoted 1 times

⊟ 👤 **brookpetit** 1 year, 7 months ago

Selected Answer: C

C is more universal and sustainable

👍 ↩ 🚩 upvoted 2 times

⊟ 👤 **ZZHZZH** 1 year, 9 months ago

Selected Answer: C

UPDATE is too expensive. Joining main and delta tables is the right wat to capture data change.

👍 ↩ 🚩 upvoted 3 times

⊟ 👤 **euro202** 1 year, 10 months ago

Selected Answer: C

I think the answer is C. The question is about maximizing performance and accuracy, it's ok if we need expensive JOINs. BigQuery has a daily quota of 1500 UPDATEs, and the question talks about several thousand updates every hour.

👍 ↩ 🚩 upvoted 2 times

⊟ 👤 **jackdbd** 1 year, 7 months ago

DML statements do not count toward the number of table modifications per day.
https://cloud.google.com/bigquery/quotas#data-manipulation-language-statements

So I would go with A.

👍 ↩ 🚩 upvoted 1 times

⊟ 👤 **jackdbd** 1 year, 7 months ago

Sorry, wrong link. Here is the correct one: https://cloud.google.com/bigquery/quotas#standard_tables

👍 ↩ 🚩 upvoted 1 times

⊟ 👤 **vaga1** 1 year, 10 months ago

Selected Answer: A

C create a view that joins to a table seems dumb to me

👍 ↩ 🚩 upvoted 1 times

⊟ 👤 **forepick** 1 year, 11 months ago

Selected Answer: C

Too frequent updates are way too expensive in an OLAP solution. This is much more likely to stream changes to the table(s) and aggregate these changes in the view.

https://stackoverflow.com/questions/74657435/bigquery-frequent-updates-to-a-record

👍 ↩ 🚩 upvoted 3 times

⊟ 👤 **streeeber** 2 years ago

Selected Answer: C

Has to be C.
DML has hard limit of 1500 operations per table per day: https://cloud.google.com/bigquery/quotas#standard_tables

👍 ↩ 🚩 upvoted 1 times

⊟ 👤 **lucaluca1982** 2 years, 1 month ago

Selected Answer: C

Update action is not efficient

👍 ↩ 🚩 upvoted 2 times

⊟ 👤 **NeoNitin** 1 year, 9 months ago

option A:what limitation here 1500/perday okay in question we will get max 24 jobs hourly updated okay.

now speed 5 operation /10 sec , 1 operation 2sec , and we are getting new update in 1 hour so we have time 3600 sec and we need to update around 1000 update according to speed take 2000sec still we have 1600 sec rest to getting new update so .
thats why I thing DML is best option for this work

👍 ↩ 🚩 **upvoted 1 times**

Load full discussion...