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 213 DISCUSSION

Actual exam question from Google's Professional Data Engineer
Question #: 213
Topic #: 1
[All Professional Data Engineer Questions]

Your company's customer_order table in BigQuery stores the order history for 10 million customers, with a table size of 10 PB. You need to create a dashboard for the support team to view the order history. The dashboard has two filters, country_name and username. Both are string data types in the BigQuery table. When a filter is applied, the dashboard fetches the order history from the table and displays the query results. However, the dashboard is slow to show the results when applying the filters to the following query:

```
SELECT date, order, status FROM customer_order
WHERE country = '<country_name>' AND username = '<username>'
```

How should you redesign the BigQuery table to support faster access?

A. Cluster the table by country and username fields.

B. Cluster the table by country field, and partition by username field.

C. Partition the table by country and username fields.

D. Partition the table by _PARTITIONTIME.

Show Suggested Answer

by e70ea9e at *Dec. 30, 2023, 9:38 a.m.*

Comments

Type your comment...

Submit

Ryannn23 3 months ago

Selected Answer: A

Partition on String is not available in BQ - excludes B and C
Partition by ingest time not useful as the query is filtering on 2 other columns - excludes D
Correct answer A: cluster on both fields.

upvoted 1 times

niujo 8 months ago

Why not D? if u partition by date and int going to be the best option??

upvoted 1 times

b3e59c2 4 months ago

Because our query filtering is relating to country and user, and nothing to do with time. A partition by time will provide no performance increase in this case.

upvoted 1 times

JyoGCP 8 months, 3 weeks ago

Selected Answer: A

If country is represented by an integer code, then partition by country and cluster by username would be a better solution. As country code is a string, available best solution is "A. Cluster the table by country and username fields."

upvoted 3 times

datapassionate 9 months, 3 weeks ago

Selected Answer: A

Correct answer: A. Cluster the table by country and username fields.

Why not B and C - > Intiger is required for partitioning
https://cloud.google.com/bigquery/docs/partitioned-tables#integer_range

upvoted 4 times

Matt_108 9 months, 3 weeks ago

Selected Answer: A

A: the fields are both strings, which are not supported for partitioning. Moreover, the fields are regularly used in filters, which is where clustering really improves performance

upvoted 3 times

SanjeevRoy91 7 months, 2 weeks ago

Is not mandatory to have partitioning for clustering?

upvoted 1 times

Takshashila 10 months ago

Selected Answer: B

Clustering can also be done after partiton?

upvoted 1 times

chambg 7 months, 4 weeks ago

Yes but the partition is done on username field which has 10 million values. Since a BQ table can only have 4000 it is not suitable

upvoted 2 times

raaad 10 months, 1 week ago

Selected Answer: A

- Clustering organizes the data based on the specified columns (in this case, country_name and username).
- When a query filters on these columns, BigQuery can efficiently scan only the relevant parts of the table

upvoted 4 times

e70ea9e 10 months, 1 week ago

Selected Answer: A

country and username --> cluster

upvoted 3 times