◉ 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 32 DISCUSSION**

Actual exam question from Google's Professional Data Engineer

Question #: 32

Topic #: 1

[All Professional Data Engineer Questions]

Your company is running their first dynamic campaign, serving different offers by analyzing real-time data during the holiday season. The data scientists are collecting terabytes of data that rapidly grows every hour during their 30-day campaign. They are using Google Cloud Dataflow to preprocess the data and collect the feature (signals) data that is needed for the machine learning model in Google Cloud Bigtable. The team is observing suboptimal performance with reads and writes of their initial load of 10 TB of data. They want to improve this performance while minimizing cost. What should they do?

A. Redefine the schema by evenly distributing reads and writes across the row space of the table.

B. The performance issue should be resolved over time as the site of the BigDate cluster is increased.

C. Redesign the schema to use a single row key to identify values that need to be updated frequently in the cluster.

D. Redesign the schema to use row keys based on numeric IDs that increase sequentially per user viewing the offers.

Show Suggested Answer

by [deleted] at *March 20, 2020, 2:15 p.m.*

Comments

Type your comment...

Submit

☐ 👤 **IsaB** 🔵 **Highly Voted** 👍 4 years, 7 months ago

I hate it when I read the question, than I think oh easy and I KNOW the answer, then I look at the choices and the answer I thought of is just not there at all... and I realize I absolutely have no idea :'D

👍 ↩ 🏳 upvoted 53 times

☐ 👤 **[Removed]** 🔵 **Highly Voted** 👍 5 years, 1 month ago

Correct A

👍 ↩ 🏳 upvoted 23 times

☐ 👤 **[Removed]** 5 years, 1 month ago

https://cloud.google.com/bigtable/docs/performance#troubleshooting
If you find that you're reading and writing only a small number of rows, you might need to redesign your schema so that reads and writes are more evenly distributed.

👍 ↩ 🏳 upvoted 20 times

☐ 👤 **meh_33** 🔵 **Most Recent** ⊘ 8 months, 3 weeks ago

Believe me all questions were from Exam topic all were there yesterday in exam. But yes dont go with starting questions mainly focus questions after 200 and latest questions are at last page.

👍 ↩ 🏳 upvoted 3 times

☐ 👤 **09878d5** 9 months, 2 weeks ago

Selected Answer: A

B is a Lie
C and D are actually not recommended
A is correct as it will help in even distribution of load and avoid hotspots

👍 ↩ 🏳 upvoted 1 times

☐ 👤 **JOKKUNO** 1 year, 5 months ago

Improving performance in Google Cloud Bigtable involves optimizing the schema design to distribute the load efficiently across the clusters. Given the scenario, the best option would be:

A. Redefine the schema by evenly distributing reads and writes across the row space of the table.

Explanation:

Distributing reads and writes evenly across the row space helps prevent hotspots and ensures that the load is spread evenly, avoiding performance bottlenecks.
Google Cloud Bigtable's performance is influenced by how well the data is distributed across the tablet servers, and evenly distributing the load can lead to better performance.
This approach aligns with best practices for designing scalable and performant Bigtable schemas.

👍 ↩ 🏳 upvoted 3 times

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

Selected Answer: A

The comment from hilel_eth totally makes sense to me. I would go with A

👍 ↩ 🏳 upvoted 1 times

☐ 👤 **hkris909** 1 year, 8 months ago

Guys, how relevant are these questions, as of Aug 14, 2023 Could we clear the PDE exam with these set of questions?

👍 ↩ 🏳 upvoted 7 times

☐ 👤 **roty** 1 year, 4 months ago

HEY DID U CLEAR THE EXAM

👍 ↩ 🏳 upvoted 2 times

☐ 👤 **FP77** 1 year, 9 months ago

Selected Answer: A

A is the only one that makes sense and is correct

👍 ↩ 🏳 upvoted 1 times

☐ 👤 **Mathew106** 1 year, 9 months ago

I understand why it could be A. But why not B also? Is it because of the typo saying BigDate instead of BigTable?

👍 ↩ 🏳 upvoted 1 times

☐ 👤 **Adswerve** 2 years ago

Selected Answer: A

A to avoid hot-spotting https://cloud.google.com/bigtable/docs/schema-design

👍 ↩ 🏳 upvoted 2 times

☐ 👤 **Brillianttyagi** 2 years, 4 months ago

A -
Make sure you're reading and writing many different rows in your table. Bigtable performs best when reads and writes are evenly distributed throughout your table, which helps Bigtable distribute the workload across all of the nodes in your cluster. If reads and writes cannot be spread across all of your Bigtable nodes, performance will suffer.

https://cloud.google.com/bigtable/docs/performance#troubleshooting

👍 ↩ 🚩 **upvoted 6 times**

⊟ 👤 **hilel_eth** 2 years, 4 months ago

A good way to improve read and write performance in a database system like Google Cloud Bigtable is to redefine the schema of the table so that reads and writes are evenly distributed across the row space of the table. This can help reduce bottlenecks in processing capacity and improve efficiency in table management. In addition, by evenly distributing read and write operations, it can prevent the accumulation of operations in one part of the table, which can improve the overall performance of the system.

👍 ↩ 🚩 **upvoted 3 times**

⊟ 👤 **Pime13** 2 years, 9 months ago

https://cloud.google.com/bigtable/docs/keyvis-overview#what-is-keyvis

To accomplish these goals, Key Visualizer can help you complete the following tasks:

Check whether your reads or writes are creating hotspots on specific rows

👍 ↩ 🚩 **upvoted 4 times**

⊟ 👤 **Arkon88** 3 years, 2 months ago

A is correct
https://cloud.google.com/bigtable/docs/performance#troubleshooting

If you find that you're reading and writing only a small number of rows, you might need to redesign your schema so that reads and writes are more evenly distributed.

👍 ↩ 🚩 **upvoted 3 times**

⊟ 👤 **samdhimal** 3 years, 3 months ago

correct answer -> Redefine the schema by evenly distributing reads and writes across the row space of the table.

Make sure you're reading and writing many different rows in your table. Bigtable performs best when reads and writes are evenly distributed throughout your table, which helps Bigtable distribute the workload across all of the nodes in your cluster. If reads and writes cannot be spread across all of your Bigtable nodes, performance will suffer.
If you find that you're reading and writing only a small number of rows, you might need to redesign your schema so that reads and writes are more evenly distributed.

Reference: https://cloud.google.com/bigtable/docs/performance#troubleshooting

👍 ↩ 🚩 **upvoted 2 times**

⊟ 👤 **MaxNRG** 3 years, 5 months ago

A as the schema needs to be redesigned to distribute the reads and writes evenly across each table.
Refer GCP documentation - Bigtable Performance:
https://cloud.google.com/bigtable/docs/performance
The table's schema is not designed correctly. To get good performance from Cloud Bigtable, it's essential to design a schema that makes it possible to distribute reads and writes evenly across each table. See Designing Your Schema for more information.
https://cloud.google.com/bigtable/docs/schema-design
Option B is wrong as increasing the size of cluster would increase the cost.
Option C is wrong as single row key for frequently updated identifiers reduces performance
Option D is wrong as sequential IDs would degrade the performance.
A safer approach is to use a reversed version of the user's numeric ID, which spreads traffic more evenly across all of the nodes for your Cloud Bigtable table.

👍 ↩ 🚩 **upvoted 11 times**

⊟ 👤 **anji007** 3 years, 6 months ago

Ans: A

👍 ↩ 🚩 **upvoted 1 times**

Load full discussion...

Platform

> Home

> Examtopics PRO

> All Exams

> Training Courses