Google Discussions

EXAM PROFESSIONAL DATA ENGINEER TOPIC 1 QUESTION 284 DISCUSSION

Actual exam question from Google's Professional Data Engineer
Question #: 284
Topic #: 1
[All Professional Data Engineer Questions]

You have a network of 1000 sensors. The sensors generate time series data: one metric per sensor per second, along with a timestamp. You already have 1 TB of data, and expect the data to grow by 1 GB every day. You need to access this data in two ways. The first access pattern requires retrieving the metric from one specific sensor stored at a specific timestamp, with a median single-digit millisecond latency. The second access pattern requires running complex analytic queries on the data, including joins, once a day. How should you store this data?

A. Store your data in BigQuery. Concatenate the sensor ID and timestamp, and use it as the primary key.

B. Store your data in Bigtable. Concatenate the sensor ID and timestamp and use it as the row key. Perform an export to BigQuery every day.

C. Store your data in Bigtable. Concatenate the sensor ID and metric, and use it as the row key. Perform an export to BigQuery every day.

D. Store your data in BigQuery. Use the metric as a primary key.

Show Suggested Answer

by scaenruy at *Jan. 4, 2024, 7:37 a.m.*

Comments

Type your comment...

Submit

raaad Highly Voted 1 year, 3 months ago

Selected Answer: B

- Bigtable excels at incredibly fast lookups by row key, often reaching single-digit millisecond latencies.
- Constructing the row key with sensor ID and timestamp enables efficient retrieval of specific sensor readings at exact timestamps.
- Bigtable's wide-column design effectively stores time series data, allowing for flexible addition of new metrics without schema changes.
- Bigtable scales horizontally to accommodate massive datasets (petabytes or more), easily handling the expected data growth.

upvoted 16 times

fitri001 Most Recent 10 months, 3 weeks ago

Selected Answer: B

agree with raaad

upvoted 2 times

hanoverquay 1 year, 1 month ago

Selected Answer: B

voted b

upvoted 1 times

JyoGCP 1 year, 2 months ago

Selected Answer: B

Option B

upvoted 1 times

Matt_108 1 year, 3 months ago

Selected Answer: B

Option B - agree with raaad

upvoted 1 times

scaenruy 1 year, 4 months ago

Selected Answer: B

B. Store your data in Bigtable. Concatenate the sensor ID and timestamp and use it as the row key. Perform an export to BigQuery every day.

upvoted 3 times

Smakyel79 1 year, 3 months ago

Based on your requirements, Option B seems most suitable. Bigtable's design caters to the low-latency access of time-series data (your first requirement), and the daily export to BigQuery enables complex analytics (your second requirement). The use of sensor ID and timestamp as the row key in Bigtable would facilitate efficient access to specific sensor data at specific times.

upvoted 4 times