← 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 74 DISCUSSION

Actual exam question from Google's Professional Data Engineer

Question #: 74

Topic #: 1

[All Professional Data Engineer Questions]

Your financial services company is moving to cloud technology and wants to store 50 TB of financial time-series data in the cloud. This data is updated frequently and new data will be streaming in all the time. Your company also wants to move their existing Apache Hadoop jobs to the cloud to get insights into this data.

Which product should they use to store the data?

A. Cloud Bigtable

B. Google BigQuery

C. Google Cloud Storage

D. Google Cloud Datastore

Show Suggested Answer

by 👤 YorelNation at *Sept. 2, 2022, 9:42 a.m.*

Comments

Type your comment...

Submit

⊟ 👤 **zellck** [Highly Voted 👍] 2 years, 5 months ago

A is the answer.

https://cloud.google.com/dataproc/docs/concepts/connectors/cloud-bigtable
Bigtable is Google's NoSQL Big Data database service. It's the same database that powers many core Google services, including Search, Analytics, Maps, and Gmail. Bigtable is designed to handle massive workloads at consistent low latency and high throughput, so it's a great choice for both operational and analytical applications, including IoT, user analytics, and financial data analysis.

Bigtable is an excellent option for any Apache Spark or Hadoop uses that require Apache HBase. Bigtable supports the Apache HBase 1.0+ APIs and offers a Bigtable HBase client in Maven, so it is easy to use Bigtable with Dataproc.

👍 ↩ 🏳 upvoted 12 times

---

opt_sub 9 months ago

Gmail is migrated to Spanner now!

👍 ↩ 🏳 upvoted 1 times

Atnafu 2 years, 4 months ago

Hbase concept here us beautiful

👍 ↩ 🏳 upvoted 2 times

---

philli1011 Most Recent ⊘ 1 year, 2 months ago

Every time you hear financial, time series, fast reads and write data, Any of that combinations, think Big Table first.
So A.

👍 ↩ 🏳 upvoted 4 times

Mathew106 1 year, 9 months ago

At first I thought that GCS was the answer but the question does mention that the data is updated frequently. Thereby, it has to be BigTable since we talk about a large amount of data, a streaming application and many individual updates. Storing the data in BigQuery and having to make individual updates doesn't make sense, and neither does running Apache jobs.

If the requirement for updates was not there I would not see any issue with GCS. GCS could serve as a replacement to HDFS and run Hadoop jobs from Dataproc.

👍 ↩ 🏳 upvoted 2 times

KC_go_reply 1 year, 10 months ago

This scenario screams for BigTable.

It's not B) BigQuery or C) Cloud Storage because both aren't supposed to contain data that is updated frequently. Then, we have to decide between A) BigTable and D) Datastore.

It is A) BigTable because
- it is the most suited for real-time / high-frequency updates
- it is similar to HBase, which is commonly used in Hadoop ecosystem stacks to store streaming / time-series data.

👍 ↩ 🏳 upvoted 1 times

AmmarFasih 1 year, 11 months ago

Many here also selected Cloud Storage. But the way I see it BigTable is specifically for low latency, high throughput, mission critical streaming data (financial data is one of them). Also the mentioning of Hadoop that points to HBase functionality if BigTable clarifies the choice more.

👍 ↩ 🏳 upvoted 1 times

Hisayuki 2 years ago

BigTable - a No-SQL database but does not support SQL Querying
Apache HBase - Based on Google's BigTable on top of HDFS and you can migrate Hadoop Apps to Cloud BigTable with the HBase API

👍 ↩ 🏳 upvoted 2 times

---

izekc 2 years ago

A. time series data

👍 ↩ 🏳 upvoted 1 times

---

midgoo 2 years, 2 months ago

Please note that there is Connector for Bigtable for Hadoop

Please note that there is Connector for Bigtable for Hadoop
https://cloud.google.com/dataproc/docs/concepts/connectors/cloud-bigtable

👍 ↩ ⚑ upvoted 1 times

samdhimal 2 years, 3 months ago

Why not Biquery?

Google BigQuery would be the best option for storing and analyzing large amounts of financial time-series data that is frequently updated and streamed in real-time. It is a fully managed, cloud-native data warehouse that allows you to analyze large datasets using SQL-like queries, and it can handle streaming data as well as batch data. Additionally, it can easily integrate with Apache Hadoop to allow your company to run their existing Hadoop jobs in the cloud and gain insights into the data.

👍 ↩ ⚑ upvoted 1 times

samdhimal 2 years, 3 months ago

A. Google Bigtable is a fully managed, NoSQL, wide-column database that is designed for large scale, low-latency workloads. It is well suited for use cases such as real-time analytics, IoT, and gaming, but it may not be the best fit for storing and analyzing large amounts of financial time-series data that is frequently updated and streamed in real-time. It lacks built-in support for SQL-like queries, which is a standard way of analyzing data in Data Warehousing and Business Intelligence. It is more focused on handling high-performance low-latency workloads, while BigQuery is focused on providing an easy and cost-effective way to analyze large amounts of data using SQL-like queries. Additionally, Bigtable doesn't provide built-in support for running Apache Hadoop jobs, and it would require additional work to integrate it with Hadoop and set it up for data warehousing and Business Intelligence use cases.

👍 ↩ ⚑ upvoted 2 times

samdhimal 2 years, 3 months ago

C. Google Cloud Storage is an object storage service that allows you to store and retrieve large amounts of unstructured data, such as video, audio, images and other files. It is not a data warehouse and does not provide built-in support for SQL-like queries, which is a standard way of analyzing data in Data Warehousing and Business Intelligence. It would not be suitable for storing and analyzing large amounts of financial time-series data that is frequently updated and streamed in real-time.

D. Google Cloud Datastore is a fully-managed, NoSQL document database that allows you to store, retrieve, and query data. It is not a data warehouse and does not provide built-in support for SQL-like queries, which is a standard way of analyzing data in Data Warehousing and Business Intelligence. It would not be suitable for storing and analyzing large amounts of financial time-series data that is frequently updated and streamed in real-time.

👍 ↩ ⚑ upvoted 2 times

samdhimal 2 years, 3 months ago

Can someone clarify why Bigtable and Not Bigquery? Super Confused.

👍 ↩ ⚑ upvoted 1 times

Oleksandr0501 2 years ago

Yes, it is possible to analyze data in Bigtable. Bigtable is a distributed NoSQL database that is designed to handle large volumes of structured data with high read and write throughput. While Bigtable itself does not provide analysis tools, it is often used in combination with other tools and technologies to perform analysis on the stored data.

👍 ↩ ⚑ upvoted 1 times

Load full discussion…

desertlotus1211 2 years, 3 months ago

https://cloud.google.com/bigtable/docs/schema-design-time-series

👍 ↩ ⚑ upvoted 2 times

Yazar97 2 years, 5 months ago

Time series data = Bigtable... So it's A

👍 ↩ ⚑ upvoted 3 times

Jay_Krish 2 years, 5 months ago

Selected Answer: A

Option A seems right

👍 ↩ ⚑ upvoted 1 times

drunk_goat82 2 years, 5 months ago

Selected Answer: A

Big Table has a HBase compliant API and is transactional unlike GCS.

👍 ↩ ⚑ upvoted 1 times

solar_maker 2 years, 5 months ago

Selected Answer: A

Selected Answer: A

BigTable can take in data from dataproc, spark and hadoop
https://cloud.google.com/dataproc/docs/concepts/connectors/cloud-bigtable#using_with

👍 ↩ 🚩 upvoted 1 times

⊟ 👤 **cloudmon** 2 years, 6 months ago

It must be C because of the existing Hadoop jobs

👍 ↩ 🚩 upvoted 3 times

⊟ 👤 **cloudmon** 2 years, 5 months ago

On 2nd thought, it's Bigtable: https://cloud.google.com/dataproc/docs/concepts/connectors/cloud-bigtable

👍 ↩ 🚩 upvoted 6 times

⊟ 👤 **pluiedust** 2 years, 7 months ago

I think it is C

👍 ↩ 🚩 upvoted 2 times

⊟ 👤 **maia01** 2 years, 7 months ago

Use Datarproc with Cloud Storage in combo with HDFS
https://cloud.google.com/dataproc/docs/concepts/dataproc-hdfs

👍 ↩ 🚩 upvoted 2 times

⊟ 👤 **euro202** 1 year, 10 months ago

Answer is A: Hadoop doesn't mean Dataproc + HDFS. This scenario is about time series that is a use-case for BigTable. Coincidentally BigTable is the best solution for migration of HBase...

👍 ↩ 🚩 upvoted 1 times

Load full discussion...