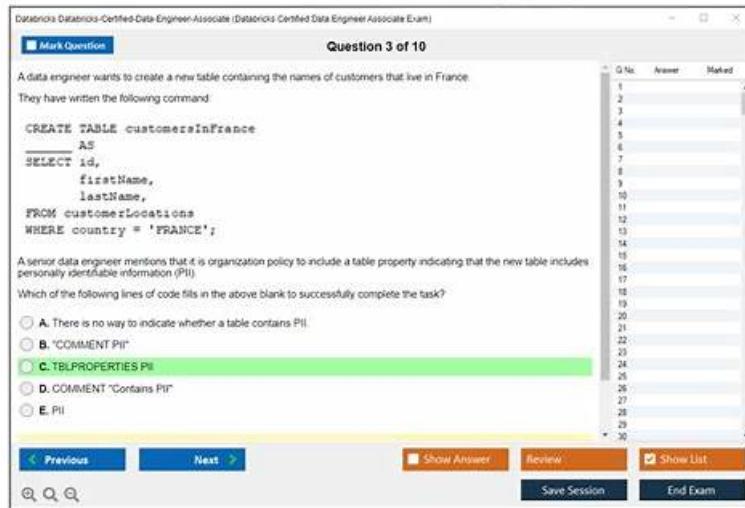


Professional-Data-Engineer Latest Test Camp & Reliable Professional-Data-Engineer Exam Prep



P.S. Free 2026 Google Professional-Data-Engineer dumps are available on Google Drive shared by ActualTorrent:
<https://drive.google.com/open?id=1dJK-TuXGQOXARBeSIRfQzNN82JKlj4Q4>

Our company abides by the industry norm all the time. By virtue of the help from professional experts, who are conversant with the regular exam questions of our latest Professional-Data-Engineer exam torrent we are dependable just like our Professional-Data-Engineer test prep. They can satisfy your knowledge-thirsty minds. And our Professional-Data-Engineer quiz torrent is quality guaranteed. By devoting ourselves to providing high-quality practice materials to our customers all these years we can guarantee all content is of the essential part to practice and remember. To sum up, our latest Professional-Data-Engineer Exam Torrent are perfect paragon in this industry full of elucidating content for exam candidates of various degree to use. Our results of latest Professional-Data-Engineer exam torrent are startlingly amazing, which is more than 98 percent of exam candidates achieved their goal successfully.

To prepare for the Google Professional-Data-Engineer Certification Exam, candidates must have a solid understanding of data engineering concepts and techniques. They must also have experience working with the Google Cloud Platform and be familiar with the tools and services offered by Google. There are many resources available to help candidates prepare for the exam, including online courses, study guides, practice exams, and hands-on training.

[**>> Professional-Data-Engineer Latest Test Camp <<**](#)

Reliable Professional-Data-Engineer Exam Prep - Reliable Professional-Data-Engineer Exam Cost

Our Professional-Data-Engineer practice guide well received by the general public for immediately after you have made a purchase for our Professional-Data-Engineer exam prep, you can download our Professional-Data-Engineer study materials to make preparations for the exams. It is universally acknowledged that time is a key factor in terms of the success of exams. The more time you spend in the preparation for Professional-Data-Engineer Learning Engine, the higher possibility you will pass the exam.

Google Professional-Data-Engineer Certification Exam is a highly recognized certification program that validates the knowledge and skills of professionals in the field of data engineering. Google Certified Professional Data Engineer Exam certification is designed to demonstrate the ability of data engineers to design, build, and maintain data processing systems, as well as to troubleshoot and optimize them for performance and cost-effectiveness. Google Certified Professional Data Engineer Exam certification exam covers a range of topics, including data processing systems, data storage and management, data analysis and machine learning, and security and compliance.

Google Certified Professional Data Engineer Exam Sample Questions (Q12-Q17):

NEW QUESTION # 12

Your infrastructure team has set up an interconnect link between Google Cloud and the on-premises network.

You are designing a high-throughput streaming pipeline to ingest data in streaming from an Apache Kafka cluster hosted on-premises. You want to store the data in BigQuery, with as minimal latency as possible.

What should you do?

- A. Use Dataflow, write a pipeline that reads the data from Kafka, and writes the data to BigQuery.
- B. Use a proxy host in the VPC in Google Cloud connecting to Kafka. Write a Dataflow pipeline, read data from the proxy host, and write the data to BigQuery.
- C. Setup a Kafka Connect bridge between Kafka and Pub/Sub. Use a Google-provided Dataflow template to read the data from Pub/Sub, and write the data to BigQuery.
- D. **Setup a Kafka Connect bridge between Kafka and Pub/Sub. Write a Dataflow pipeline, read the data from Pub/Sub, and write the data to BigQuery.**

Answer: D

Explanation:

Here's a detailed breakdown of why this solution is optimal and why others fall short:

Why Option C is the Best Solution:

* **Kafka Connect Bridge:** This bridge acts as a reliable and scalable conduit between your on-premises Kafka cluster and Google Cloud's Pub/Sub messaging service. It handles the complexities of securely transferring data over the interconnect link.

* **Pub/Sub as a Buffer:** Pub/Sub serves as a highly scalable buffer, decoupling the Kafka producer from the Dataflow consumer. This is crucial for handling fluctuations in message volume and ensuring smooth data flow even during spikes.

* **Custom Dataflow Pipeline:** Writing a custom Dataflow pipeline gives you the flexibility to implement any necessary transformations or enrichments to the data before it's written to BigQuery. This is often required in real-world streaming scenarios.

* **Minimal Latency:** By using Pub/Sub as a buffer and Dataflow for efficient processing, you minimize the latency between the data being produced in Kafka and being available for querying in BigQuery.

Why Other Options Are Not Ideal:

* **Option A:** Using a proxy host introduces an additional point of failure and can create a bottleneck, especially with high-throughput streaming.

* **Option B:** While Google-provided Dataflow templates can be helpful, they might lack the customization needed for specific transformations or handling complex data structures.

* **Option D:** Dataflow doesn't natively connect to on-premises Kafka clusters. Directly reading from Kafka would require complex networking configurations and could lead to performance issues.

Additional Considerations:

* **Schema Management:** Ensure that the schema of the data being produced in Kafka is compatible with the schema expected in BigQuery. Consider using tools like Schema Registry for schema evolution management.

* **Monitoring:** Set up robust monitoring and alerting to detect any issues in the pipeline, such as message backlogs or processing errors.

By following Option C, you leverage the strengths of Kafka Connect, Pub/Sub, and Dataflow to create a high-throughput, low-latency streaming pipeline that seamlessly integrates your on-premises Kafka data with BigQuery.

NEW QUESTION # 13

You have an Oracle database deployed in a VM as part of a Virtual Private Cloud (VPC) network. You want to replicate and continuously synchronize 50 tables to BigQuery. You want to minimize the need to manage infrastructure. What should you do?

- A. Deploy Apache Kafka in the same VPC network, use Kafka Connect Oracle change data capture (CDC), and the Kafka Connect Google BigQuery Sink Connector.
- B. Deploy Apache Kafka in the same VPC network, use Kafka Connect Oracle Change Data Capture (CDC), and Dataflow to stream the Kafka topic to BigQuery.
- C. Create a Pub/Sub subscription to write to BigQuery directly Deploy the Debezium Oracle connector to capture changes in the Oracle database, and sink to the Pub/Sub topic.
- D. **Create a Datastream service from Oracle to BigQuery, use a private connectivity configuration to the same VPC network, and a connection profile to BigQuery.**

Answer: D

Explanation:

Datastream is a serverless, scalable, and reliable service that enables you to stream data changes from Oracle and MySQL databases to Google Cloud services such as BigQuery, Cloud SQL, Google Cloud Storage, and Cloud Pub/Sub. Datastream

captures and streams database changes using change data capture (CDC) technology. Datastream supports private connectivity to the source and destination systems using VPC networks. Datastream also provides a connection profile to BigQuery, which simplifies the configuration and management of the data replication. Reference:

Datastream overview

Creating a Datastream stream

Using Datastream with BigQuery

NEW QUESTION # 14

The YARN ResourceManager and the HDFS NameNode interfaces are available on a Cloud Dataproc cluster ____.

- A. master node
- B. worker node
- C. application node
- D. conditional node

Answer: A

Explanation:

The YARN ResourceManager and the HDFS NameNode interfaces are available on a Cloud Dataproc cluster master node. The cluster master-host-name is the name of your Cloud Dataproc cluster followed by an -m suffix-for example, if your cluster is named "my- cluster", the master-host-name would be "my-cluster-m".

Reference: [https://cloud.google.com/dataproc/docs/concepts/cluster-web- interfaces#interfaces](https://cloud.google.com/dataproc/docs/concepts/cluster-web-interfaces#interfaces)

NEW QUESTION # 15

You are migrating your data warehouse to BigQuery. You have migrated all of your data into tables in a dataset. Multiple users from your organization will be using the dat

a. They should only see certain tables based on their team membership. How should you set user permissions?

- A. Create authorized views for each team in the same dataset in which the data resides, and assign the users/groups data viewer access to the authorized views
- B. Create SQL views for each team in the same dataset in which the data resides, and assign the users/groups data viewer access to the SQL views
- C. **Assign the users/groups data viewer access at the table level for each table**
- D. Create authorized views for each team in datasets created for each team. Assign the authorized views data viewer access to the dataset in which the data resides. Assign the users/groups data viewer access to the datasets in which the authorized views reside

Answer: C

NEW QUESTION # 16

You are developing an application on Google Cloud that will automatically generate subject labels for users' blog posts. You are under competitive pressure to add this feature quickly, and you have no additional developer resources. No one on your team has experience with machine learning. What should you do?

- A. Build and train a text classification model using TensorFlow. Deploy the model using Cloud Machine Learning Engine. Call the model from your application and process the results as labels.
- B. Call the Cloud Natural Language API from your application. Process the generated Sentiment Analysis as labels.
- C. Build and train a text classification model using TensorFlow. Deploy the model using a Kubernetes Engine cluster. Call the model from your application and process the results as labels.
- D. **Call the Cloud Natural Language API from your application. Process the generated Entity Analysis as labels.**

Answer: D

Explanation:

As time is less, use cloud NLP and entity is used to label general subjects, sentiment label for sentiment analysis.

NEW QUESTION # 17

Reliable Professional-Data-Engineer Exam Prep: <https://www.actualtorrent.com/Professional-Data-Engineer-questions-answers.html>

BTW, DOWNLOAD part of ActualTorrent Professional-Data-Engineer dumps from Cloud Storage:
<https://drive.google.com/open?id=1djk-TuXG00XARBeSIRfOzNN82JKj4Q4>