

Google Professional-Data-Engineer인기덤프공부 & Professional-Data-Engineer최신버전인기덤프자료



그리고 Itcertkr Professional-Data-Engineer 시험 문제집의 전체 버전을 클라우드 저장소에서 다운로드할 수 있습니다:
https://drive.google.com/open?id=1Bc_AXpaRtJymSyQP9IuS_QrHw0sJU0yf

Itcertkr의 Google인증 Professional-Data-Engineer덤프를 구매하여 공부한지 일주일만에 바로 시험을 보았는데 고득점으로 시험을 패스했습니다. 이는 Itcertkr의 Google인증 Professional-Data-Engineer덤프를 구매한 분이 전해온 최소익입니다. 다른 자료 필요없이 단지 저희 Google인증 Professional-Data-Engineer덤프로 이렇게 어려운 시험을 일주일만에 패스하고 자격증을 취득할 수 있습니다. 덤프가격도 다른 사이트보다 만만하여 부담없이 덤프마련이 가능합니다. 구매전 무료샘플을 다운받아 보시면 믿음을 느낄 것입니다.

Google Professional-Data-Engineer 인증 시험은 Google Cloud Platform에서 데이터 처리 시스템을 설계하고 구축하는 전문가를 증명하고자 하는 개인을 위해 설계되었습니다. Google 인증된 전문 데이터 엔지니어로서, 이 분야에서 전문가로 인정받으며 Google Cloud 기술을 활용한 확장 가능하고 효율적인 데이터 처리 시스템을 만드는 기술을 갖추게 됩니다. 이 인증은 데이터 엔지니어, 데이터 아키텍트, 솔루션 아키텍트 및 클라우드에서 데이터 처리 시스템을 구축하고 관리하는 기술을 검증하고자 하는 모든 사람들에게 이상적입니다.

Google Professional-Data-Engineer 자격증 시험은 데이터 엔지니어링, 데이터 분석, 기계 학습 및 대용량 데이터 처리와 같은 다양한 분야의 지식과 기술을 검증하는 온라인 시험입니다. 이 시험은 객관식 문항으로 구성되어 있으며, Google Cloud Platform, 그 서비스 및 기능에 대한 철저한 이해력이 필요합니다. 후보자는 비즈니스 요구 사항을 충족하는 데이터 처리 시스템을 설계하고 구현하는 능력을 증명해야 합니다.

>> Google Professional-Data-Engineer인기덤프공부 <<

시험대비 Professional-Data-Engineer인기덤프공부 덤프 최신문제

일반적으로 Professional-Data-Engineer인증시험은 IT업계전문가들이 끊임없는 노력과 지금까지의 경험으로 연구하여 만들어낸 제일 정확한 시험문제와 답들입니다. 마침 우리 Itcertkr의 문제와 답들은 모두 이러한 과정을 거쳐서 만들어진 아주 완벽한 시험대비문제집들입니다. 우리의 문제집으로 여러분은 충분히 안전이 시험을 패스하실 수 있습니다. 우리 Itcertkr의 문제집들은 모두 100%보장 도를 자랑하며 만약 우리 Itcertkr의 제품을 구매하였다면 Google Professional-Data-Engineer관련 시험패스와 자격증취득은 근심하지 않으셔도 됩니다. 여러분은 IT업계에서 또 한층

업그레이드 될것입니다.

최신 Google Cloud Certified Professional-Data-Engineer 무료 샘플문제 (Q154-Q159):

질문 # 154

You want to migrate an on-premises Hadoop system to Cloud Dataproc. Hive is the primary tool in use, and the data format is Optimized Row Columnar (ORC). All ORC files have been successfully copied to a Cloud Storage bucket. You need to replicate some data to the cluster's local Hadoop Distributed File System (HDFS) to maximize performance. What are two ways to start using Hive in Cloud Dataproc?

(Choose two.)

- A. Run the gsutil utility to transfer all ORC files from the Cloud Storage bucket to HDFS. Mount the Hive tables locally.
- B. Leverage Cloud Storage connector for Hadoop to mount the ORC files as external Hive tables. Replicate external Hive tables to the native ones.
- C. Run the gsutil utility to transfer all ORC files from the Cloud Storage bucket to the master node of the Dataproc cluster. Then run the Hadoop utility to copy them to HDFS. Mount the Hive tables from HDFS.
- D. Load the ORC files into BigQuery. Leverage BigQuery connector for Hadoop to mount the BigQuery tables as external Hive tables. Replicate external Hive tables to the native ones.
- E. Run the gsutil utility to transfer all ORC files from the Cloud Storage bucket to any node of the Dataproc cluster. Mount the Hive tables locally.

정답: C,E

설명:

HDFS lies on datanode, data on masternode needs to be copied on datanode.

질문 # 155

You have a data analyst team member who needs to analyze data by using BigQuery. The data analyst wants to create a data pipeline that would load 200 CSV files with an average size of 15MB from a Cloud Storage bucket into BigQuery daily. The data needs to be ingested and transformed before being accessed in BigQuery for analysis. You need to recommend a fully managed, no-code solution for the data analyst. What should you do?

- A. Create a Cloud Run function and schedule it to run daily using Cloud Scheduler to load the data into BigQuery.
- B. Build a custom Apache Beam pipeline and run it on Dataflow to load the file from Cloud Storage to BigQuery and schedule it to run daily using Cloud Composer.
- C. Create a pipeline by using BigQuery pipelines and schedule it to load the data into BigQuery daily.
- D. Use the BigQuery Data Transfer Service to load files from Cloud Storage to BigQuery, create a BigQuery job which transforms the data using BigQuery SQL and schedule it to run daily.

정답: D

설명:

The requirements are for a daily scheduled load, ingest, and transformation, and specifically a fully managed, no-code solution.

* Ingest (Load): The BigQuery Data Transfer Service (DTS) is the fully managed, serverless, and no-code solution for batch loading files (including CSV from Cloud Storage) into BigQuery on a schedule. This is the "ingest" part.

* Transform: After loading the raw data into a staging table using DTS, the transformation can be done using BigQuery SQL. This transformation query can then be automated using a Scheduled Query in BigQuery, which is also a fully managed and no-code feature that runs on a schedule.

* Fully Managed & No-Code: Both DTS for Cloud Storage and Scheduled Queries are native BigQuery features that are fully managed and configured through the console without requiring code, directly meeting the constraints.

* Correcting other options:

* A (Cloud Run + Script): Cloud Run requires writing a custom Python script, which violates the no-code requirement.

* C (Dataflow + Apache Beam + Cloud Composer): This is a powerful, highly scalable ETL solution, but it requires writing custom code (Apache Beam) and requires setting up and managing a workflow orchestrator (Cloud Composer/Airflow), which violates both the fully managed (Dataflow is serverless, but the code/pipeline itself is custom and needs maintenance) and no-code requirements.

* D (BigQuery pipelines): "BigQuery pipelines" is not a distinct, official product name in the Google Cloud documentation that fulfills a no-code scheduled ETL. The closest product is the combination of DTS and Scheduled Queries, as described in option B.

Reference: Google Cloud Documentation on BigQuery Data Transfer Service and Scheduled Queries:

"The BigQuery Data Transfer Service automates data movement into BigQuery on a scheduled, managed basis... The BigQuery Data Transfer Service supports loading data from Cloud Storage in one of the following formats: Comma-separated values (CSV)..." (Source: What is BigQuery Data Transfer Service? and Introduction to Cloud Storage transfers)

"A scheduled query is a query that BigQuery automatically runs at regular intervals. When you configure a scheduled query, you specify the GoogleSQL SELECT statement to run, the destination table for the query results, and the frequency of the query." (Source: Scheduling queries) This combination delivers a fully managed, no-code ELT (Extract-Load-Transform) pipeline.

질문 # 156

You are working on a sensitive project involving private user data

a. You have set up a project on Google Cloud Platform to house your work internally. An external consultant is going to assist with coding a complex transformation in a Google Cloud Dataflow pipeline for your project. How should you maintain users' privacy?

- A. Grant the consultant the Viewer role on the project.
- **B. Create a service account and allow the consultant to log on with it.**
- C. Grant the consultant the Cloud Dataflow Developer role on the project.
- D. Create an anonymized sample of the data for the consultant to work with in a different project.

정답: B

질문 # 157

Your chemical company needs to manually check documentation for customer order. You use a pull subscription in Pub/Sub so that sales agents get details from the order. You must ensure that you do not process orders twice with different sales agents and that you do not add more complexity to this workflow. What should you do?

- **A. Use Pub/Sub exactly-once delivery in your pull subscription.**
- B. Create a new Pub/Sub push subscription to monitor the orders processed in the agent's system.
- C. Use a Deduplicate PTransform in Dataflow before sending the messages to the sales agents.
- D. Create a transactional database that monitors the pending messages.

정답: A

설명:

Pub/Sub exactly-once delivery is a feature that guarantees that subscriptions do not receive duplicate deliveries of messages based on a Pub/Sub-defined unique message ID. This feature is only supported by the pull subscription type, which is what you are using in this scenario. By enabling exactly-once delivery, you can ensure that each order is processed only once by a sales agent, and that no order is lost or duplicated. This also simplifies your workflow, as you do not need to create a separate database or subscription to monitor the pending or processed messages. Reference:

Exactly-once delivery | Cloud Pub/Sub Documentation

Cloud Pub/Sub Exactly-once Delivery feature is now Generally Available (GA)

질문 # 158

You have a data pipeline with a Cloud Dataflow job that aggregates and writes time series metrics to Cloud Bigtable. This data feeds a dashboard used by thousands of users across the organization. You need to support additional concurrent users and reduce the amount of time required to write the data

a. Which two actions should you take? (Choose two.)

- **A. Increase the number of nodes in the Cloud Bigtable cluster**
- B. Modify your Cloud Dataflow pipeline to use the Flatten transform before writing to Cloud Bigtable
- **C. Increase the maximum number of Cloud Dataflow workers by setting maxNumWorkers in PipelineOptions**
- D. Configure your Cloud Dataflow pipeline to use local execution
- E. Modify your Cloud Dataflow pipeline to use the CoGroupByKey transform before writing to Cloud Bigtable

정답: A,C

설명:

References:

Professional-Data-Engineer최신버전 인기 덤프자료 : https://www.itcertkr.com/Professional-Data-Engineer_exam.html

- 그리고 Itcertkr Professional-Data-Engineer 시험 문제집의 전체 버전을 클라우드 저장소에서 다운로드할 수 있습니다:
<https://drive.google.com/open?id=1Bc AXpaRtJymSyQP9IuS QrHw0sJU0yf>