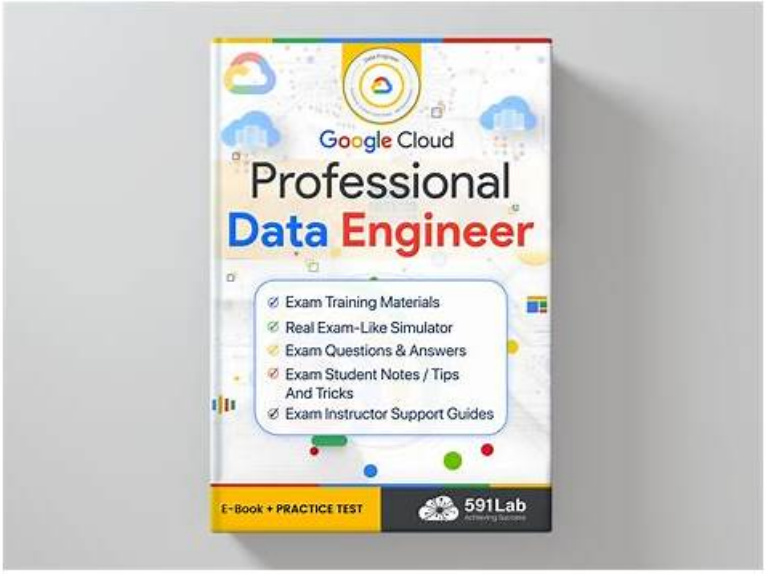
Exam Google Professional-Data-Engineer Testking - Exam Professional-Data-Engineer Review



P.S. Free & New Professional-Data-Engineer dumps are available on Google Drive shared by DumpsTests: https://drive.google.com/open?id=143PWc4NlCgptIz-boHokTxsTpYZE_prK

We pay emphasis on variety of situations and adopt corresponding methods to deal with. More successful cases of passing the Professional-Data-Engineer exam can be found and can prove our powerful strength. As a matter of fact, since the establishment, we have won wonderful feedback and ceaseless business, continuously working on developing our Professional-Data-Engineer Test Prep. We have been specializing Professional-Data-Engineer exam dumps many years and have a great deal of long-term old clients, and we would like to be a reliable cooperator on your learning path and in your further development.

Google Professional-Data-Engineer Exam is a rigorous test of an individual's skills and knowledge in data engineering on Google Cloud technologies. As demand for skilled data engineering professionals continues to grow, the certification can open up many lucrative job opportunities for those looking to make their mark in the industry. Google Certified Professional Data Engineer Exam certification process requires practical experience, extensive preparation, and dedication to acquiring skills that are highly valued in today's rapidly evolving technology ecosystem.

>> Exam Google Professional-Data-Engineer Testking <<

Exam Professional-Data-Engineer Review - Professional-Data-Engineer Dump Torrent

You can directly refer our Professional-Data-Engineer study materials to prepare the exam. Once the newest test syllabus is issued by the official, our experts will quickly make a detailed summary about all knowledge points of the real Professional-Data-Engineer exam in the shortest time. All in all, our Professional-Data-Engineer Exam Quiz will help you grasp all knowledge points. Not only our professional expert have simplified the content of the subject for you to understand fully, but also our Professional-Data-Engineer practice guide will help you pass the exam smoothly.

The Google Professional-Data-Engineer exam consists of multiple-choice and scenario-based questions that test the candidates' understanding of GCP data engineering services and best practices for data engineering. Candidates have two hours and thirty minutes to complete the exam. Professional-Data-Engineer Exam is available in English, Japanese, Spanish, and Portuguese.

Google Certified Professional Data Engineer Exam Sample Questions (Q275-Q280):

NEW QUESTION # 275
Which of these is not a supported method of putting data into a partitioned table?

- A. Use ORDER BY to put a table's rows into chronological order and then change the table's type to "Partitioned".
- B. If you have existing data in a separate file for each day, then create a partitioned table and upload each file into the appropriate partition.
- C. Run a query to get the records for a specific day from an existing table and for the destination table, specify a partitioned table ending with the day in the format "$YYYYMMDD".
- D. Create a partitioned table and stream new records to it every day.

Answer: A

Explanation:
You cannot change an existing table into a partitioned table. You must create a partitioned table from scratch. Then you can either stream data into it every day and the data will automatically be put in the right partition, or you can load data into a specific partition by using "$YYYYMMDD" at the end of the table name.

NEW QUESTION # 276
You are developing an Apache Beam pipeline to extract data from a Cloud SQL instance by using JdbcIO. You have two projects running in Google Cloud. The pipeline will be deployed and executed on Dataflow in Project A.
The Cloud SQL instance is running jn Project B and does not have a public IP address. After deploying the pipeline, you noticed that the pipeline failed to extract data from the Cloud SQL instance due to connection failure. You verified that VPC Service Controls and shared VPC are not in use in these projects. You want to resolve this error while ensuring that the data does not go through the public internet. What should you do?

- A. Add the external IP addresses of the Dataflow worker as authorized networks in the Cloud SOL instance.
- B. Set up VPC Network Peering between Project A and Project B. Add a firewall rule to allow the peered subnet range to access all instances on the network.
- C. Set up VPC Network Peering between Project A and Project B. Create a Compute Engine instance without external IP address in Project B on the peered subnet to serve as a proxy server to the Cloud SQL database.
- D. Turn off the external IP addresses on the Dataflow worker. Enable Cloud NAT in Project A.

Answer: C

Explanation:
Option A is incorrect because VPC Network Peering alone does not enable connectivity to Cloud SQL instances with private IP addresses. You also need to configure private services access and allocate an IP address range for the service producer network1.
Option B is incorrect because Cloud NAT does not support Cloud SQL instances with private IP addresses. Cloud NAT only provides outbound connectivity for resources that do not have public IP addresses, such as VMs, GKE clusters, and serverless instances2.
Option C is correct because it allows you to use a Compute Engine instance as a proxy server to connect to the Cloud SQL database over the peered network. The proxy server does not need an external IP address because it can communicate with the Dataflow workers and the Cloud SQL instance using internal IP addresses. You need to install the Cloud SQL Auth proxy on the proxy server and configure it to use a service account that has the Cloud SQL Client role.
Option D is incorrect because it requires you to assign public IP addresses to the Dataflow workers, which exposes the data to the public internet. This violates the requirement of ensuring that the data does not go through the public internet. Moreover, adding authorized networks does not work for Cloud SQL instances with private IP addresses.

NEW QUESTION # 277
You want to migrate an Apache Spark 3 batch job from on-premises to Google Cloud. You need to minimally change the job so that the job reads from Cloud Storage and writes the result to BigQuery. Your job is optimized for Spark, where each executor has 8 vCPU and 16 GB memory, and you want to be able to choose similar settings. You want to minimize installation and management effort to run your job. What should you do?

- A. Execute the job from a new Compute Engine VM.
- B. Execute the job in a new Dataproc cluster.
- C. Execute as a Dataproc Serverless job.
- D. Execute the job as part of a deployment in a new Google Kubernetes Engine cluster.

Answer: B

NEW QUESTION # 278
Case Study: 2,
Flowlogistic Case Study
Company Overview
Flowlogistic is a leading logistics and supply chain provider. They help businesses throughout the world manage their resources and transport them to their final destination. The company has grown rapidly, expanding their offerings to include rail, truck, aircraft, and oceanic shipping.
Company Background
The company started as a regional trucking company, and then expanded into other logistics market.
Because they have not updated their infrastructure, managing and tracking orders and shipments has become a bottleneck. To improve operations, Flowlogistic developed proprietary technology for tracking shipments in real time at the parcel level. However, they are unable to deploy it because their technology stack, based on Apache Kafka, cannot support the processing volume. In addition, Flowlogistic wants to further analyze their orders and shipments to determine how best to deploy their resources.
Solution Concept
Flowlogistic wants to implement two concepts using the cloud:
Use their proprietary technology in a real-time inventory-tracking system that indicates the location of their loads Perform analytics on all their orders and shipment logs, which contain both structured and unstructured data, to determine how best to deploy resources, which markets to expand info. They also want to use predictive analytics to learn earlier when a shipment will be delayed.
Existing Technical Environment
Flowlogistic architecture resides in a single data center:
Databases
8 physical servers in 2 clusters
SQL Server - user data, inventory, static data
3 physical servers
Cassandra - metadata, tracking messages
10 Kafka servers - tracking message aggregation and batch insert
Application servers - customer front end, middleware for order/customs 60 virtual machines across 20 physical servers Tomcat - Java services Nginx - static content Batch servers Storage appliances iSCSI for virtual machine (VM) hosts Fibre Channel storage area network (FC SAN) ?SQL server storage Network-attached storage (NAS) image storage, logs, backups Apache Hadoop /Spark servers Core Data Lake Data analysis workloads
20 miscellaneous servers
Jenkins, monitoring, bastion hosts,
Business Requirements
Build a reliable and reproducible environment with scaled panty of production. Aggregate data in a centralized Data Lake for analysis Use historical data to perform predictive analytics on future shipments Accurately track every shipment worldwide using proprietary technology Improve business agility and speed of innovation through rapid provisioning of new resources Analyze and optimize architecture for performance in the cloud Migrate fully to the cloud if all other requirements are met Technical Requirements Handle both streaming and batch data Migrate existing Hadoop workloads Ensure architecture is scalable and elastic to meet the changing demands of the company.
Use managed services whenever possible
Encrypt data flight and at rest
Connect a VPN between the production data center and cloud environment SEO Statement We have grown so quickly that our inability to upgrade our infrastructure is really hampering further growth and efficiency. We are efficient at moving shipments around the world, but we are inefficient at moving data around.
We need to organize our information so we can more easily understand where our customers are and what they are shipping.
CTO Statement
IT has never been a priority for us, so as our data has grown, we have not invested enough in our technology. I have a good staff to manage IT, but they are so busy managing our infrastructure that I cannot get them to do the things that really matter, such as organizing our data, building the analytics, and figuring out how to implement the CFO' s tracking technology.
CFO Statement
Part of our competitive advantage is that we penalize ourselves for late shipments and deliveries. Knowing where out shipments are at all times has a direct correlation to our bottom line and profitability.
Additionally, I don't want to commit capital to building out a server environment.
Flowlogistic wants to use Google BigQuery as their primary analysis system, but they still have Apache Hadoop and Spark workloads that they cannot move to BigQuery. Flowlogistic does not know how to store the data that is common to both workloads. What should they do?

- A. Store the common data in BigQuery as partitioned tables.
- B. Store he common data in the HDFS storage for a Google Cloud Dataproc cluster.
- C. Store the common data encoded as Avro in Google Cloud Storage.
- D. Store the common data in BigQuery and expose authorized views.

Answer: D

NEW QUESTION # 279
If a dataset contains rows with individual people and columns for year of birth, country, and income, how many of the columns are continuous and how many are categorical?

- A. 1 continuous and 2 categorical
- B. 3 continuous
- C. 3 categorical
- D. 2 continuous and 1 categorical

Answer: D

Explanation:
The columns can be grouped into two types-categorical and continuous columns:
A column is called categorical if its value can only be one of the categories in a finite set. For example, the native country of a person (U.S., India, Japan, etc.) or the education level (high school, college, etc.) are categorical columns.
A column is called continuous if its value can be any numerical value in a continuous range. For example, the capital gain of a person (e.g. $14,084) is a continuous column.
Year of birth and income are continuous columns. Country is a categorical column.
You could use bucketization to turn year of birth and/or income into categorical features, but the raw columns are continuous.
Reference: https://www.tensorflow.org/tutorials/wide#reading_the_census_data

NEW QUESTION # 280
.....

- gritacademy.us, pct.edu.pk, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, www.stes.tyc.edu.tw, www.tttttt456.com, project.gabus.lt, www.stes.tyc.edu.tw, kemono.im, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, Disposable vapes

BTW, DOWNLOAD part of DumpsTests Professional-Data-Engineer dumps from Cloud Storage: https://drive.google.com/open?id=143PWc4NlCgptIz-boHokTxsTpYZE_prK