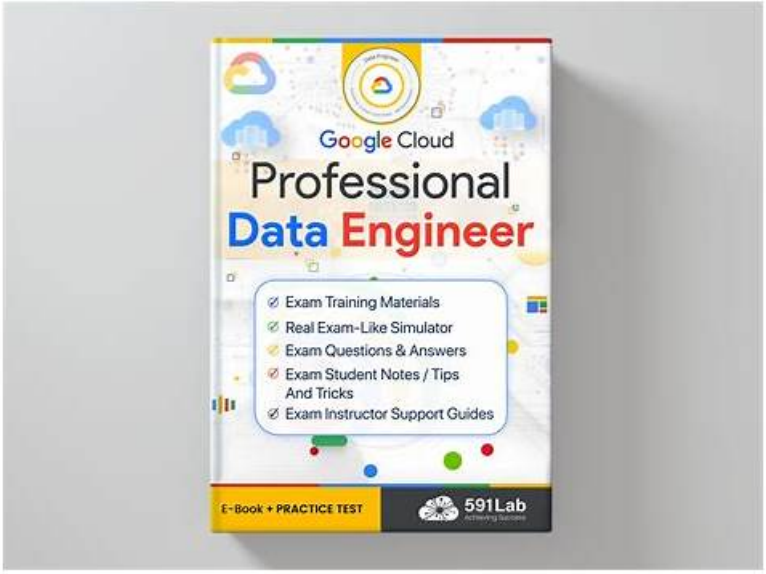
Google Professional-Data-Engineer Valid Exam Online, Latest Professional-Data-Engineer Test Notes



DOWNLOAD the newest BraindumpQuiz Professional-Data-Engineer PDF dumps from Cloud Storage for free: https://drive.google.com/open?id=1R6rWCRiJzgSlUzZNKJRexO5LHUzVGrnh

In accordance with the actual exam, we provide the latest Professional-Data-Engineer exam dumps for your practices. With the latest Professional-Data-Engineer test questions, you can have a good experience in practicing the test. Moreover, you have no need to worry about the price, we provide free updating for one year and half price for further partnerships, which is really a big sale in this field. After your payment, we will send the updated Professional-Data-Engineer Exam to you immediately and if you have any question about updating, please leave us a message on our Professional-Data-Engineer exam questions.

Google Professional-Data-Engineer certification is a highly recognized credential in the field of data engineering. It is designed to demonstrate a candidate's expertise in designing, building, and maintaining data processing systems on Google Cloud Platform. Google Certified Professional Data Engineer Exam certification is intended for experienced data professionals who are seeking to enhance their skills and knowledge in the areas of data engineering, cloud computing, and data analytics.

Google Professional-Data-Engineer Certification has become a must-have for data engineering professionals who work with Google Cloud. Google Certified Professional Data Engineer Exam certification validates their knowledge and skills in designing and building data processing systems, as well as their ability to analyze and use machine learning models. Google Certified Professional Data Engineer Exam certification also helps professionals to stand out in a competitive job market and advance their careers.

>> Google Professional-Data-Engineer Valid Exam Online <<

Hot Professional-Data-Engineer Valid Exam Online | Reliable Professional-Data-Engineer: Google Certified Professional Data Engineer Exam 100% Pass

Success does not come only from the future, but it continues to accumulate from the moment you decide to do it. At the moment you choose Professional-Data-Engineer practice quiz, you have already taken the first step to success. The next thing you have to do is stick with it. Professional-Data-Engineer Training Materials will definitely live up to your expectations. Not only our Professional-Data-Engineer study materials contain the latest exam questions and answers, but also the pass rate is high as 98% to 100%.

Achieving the Google Certified Professional Data Engineer certification demonstrates to employers and colleagues that an individual has the knowledge and skills necessary to design and build data processing systems on Google Cloud Platform. Google Certified Professional Data Engineer Exam certification can open up new career opportunities and help individuals advance in their current roles.

Google Certified Professional Data Engineer Exam Sample Questions (Q286-Q291):

NEW QUESTION # 286
Your company built a TensorFlow neutral-network model with a large number of neurons and layers. The model fits well for the training data. However, when tested against new data, it performs poorly. What method can you employ to address this?

- A. Serialization
- B. Dropout Methods
- C. Dimensionality Reduction
- D. Threading

Answer: B

Explanation:
Explanation/Reference: https://medium.com/mlreview/a-simple-deep-learning-model-for-stock-price-prediction-using- tensorflow-30505541d877

NEW QUESTION # 287
Your company produces 20,000 files every hour. Each data file is formatted as a comma separated values (CSV) file that is less than 4 KB. All files must be ingested on Google Cloud Platform before they can be processed. Your company site has a 200 ms latency to Google Cloud, and your Internet connection bandwidth is limited as 50 Mbps. You currently deploy a secure FTP (SFTP) server on a virtual machine in Google Compute Engine as the data ingestion point. A local SFTP client runs on a dedicated machine to transmit the CSV files as is. The goal is to make reports with data from the previous day available to the executives by 10:00 a.m. each day. This design is barely able to keep up with the current volume, even though the bandwidth utilization is rather low.
You are told that due to seasonality, your company expects the number of files to double for the next three months. Which two actions should you take? (Choose two.)

- A. Redesign the data ingestion process to use gsutil tool to send the CSV files to a storage bucket in parallel.
- B. Introduce data compression for each file to increase the rate file of file transfer.
- C. Contact your internet service provider (ISP) to increase your maximum bandwidth to at least 100 Mbps.
- D. Create an S3-compatible storage endpoint in your network, and use Google Cloud Storage Transfer Service to transfer on-premices data to the designated storage bucket.
- E. Assemble 1,000 files into a tape archive (TAR) file. Transmit the TAR files instead, and disassemble the CSV files in the cloud upon receiving them.

Answer: A,D

NEW QUESTION # 288
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
* 10 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's management has determined that the current Apache Kafka servers cannot handle the data volume for their real-time inventory tracking system. You need to build a new system on Google Cloud Platform (GCP) that will feed the proprietary tracking software. The system must be able to ingest data from a variety of global sources, process and query in real-time, and store the data reliably. Which combination of GCP products should you choose?

- A. Cloud Pub/Sub, Cloud Dataflow, and Cloud Storage
- B. Cloud Pub/Sub, Cloud SQL, and Cloud Storage
- C. Cloud Load Balancing, Cloud Dataflow, and Cloud Storage
- D. Cloud Dataflow, Cloud SQL, and Cloud Storage
- E. Cloud Pub/Sub, Cloud Dataflow, and Local SSD

Answer: B

NEW QUESTION # 289
You are preparing an organization-wide dataset. You need to preprocess customer data stored in a restricted bucket in Cloud Storage. The data will be used to create consumer analyses. You need to follow data privacy requirements, including protecting certain sensitive data elements, while also retaining all of the data for potential future use cases. What should you do?

- A. Use customer-managed encryption keys (CMEK) to directly encrypt the data in Cloud Storage. Use federated queries from BigQuery. Share the encryption key by following the principle of least privilege.
- B. Use the Cloud Data Loss Prevention API and Dataflow to detect and remove sensitive fields from the data in Cloud Storage. Write the filtered data in BigQuery.
- C. Use Dataflow and Cloud KMS to encrypt sensitive fields and write the encrypted data in BigQuery. Share the encryption key by following the principle of least privilege.
- D. Use Dataflow and the Cloud Data Loss Prevention API to mask sensitive data. Write the processed data in BigQuery.

Answer: D

Explanation:
The core requirements are to protect sensitive data elements (data privacy) while retainingalldata for potential future use, and then using this preprocessed data for consumer analyses.
* Retaining All Data:This immediately makes option B (remove sensitive fields) unsuitable because it involves data loss.
* Protecting Sensitive Data for Analysis & Future Use:Masking is a de-identification technique that redacts or replaces sensitive data with a substitute, allowing the data structure and usability for analysis to be maintained without exposing the original sensitive values. This aligns with protecting data while still making it usable.
* Cloud Data Loss Prevention (DLP) API:This service is specifically designed to discover, classify, and protect sensitive data. It offers various de-identification techniques, including masking.
* Dataflow:This is a serverless, fast, and cost-effective service for unified stream and batch data processing. It's well-suited for transforming large datasets, such as those read from Cloud Storage, and can integrate with the DLP API for de-identification.
* Writing to BigQuery:BigQuery is an ideal destination for an organization-wide dataset for consumer analyses.
Therefore, using Dataflow to read the data from Cloud Storage, leveraging the Cloud DLP API tomask(a form of de-identification) the sensitive elements, and then writing the processed (masked) data to BigQuery is the most appropriate solution. This approach protects privacy for the consumer analyses dataset while the original, unaltered data can still be retained in the restricted Cloud Storage bucket for future use cases that might require access to the original sensitive information (under strict governance).
Let's analyze why other options are less suitable:
* Option B:"Remove sensitive fields" means data loss, which contradicts the requirement to retain all data for potential future use cases.
* Option C:Encrypting sensitive fields with Cloud KMS and writing them to BigQuery is a valid way to protect data. However, for "consumer analyses," masked data is generally more directly usable than encrypted data. Analysts would typically work with de-identified (e.g., masked) data rather than directly querying encrypted fields and managing decryption keys for analytical purposes. While decryption is possible, masking often provides a better balance of privacy and utility for broad analysis.
The question also implies creating a datasetforanalysis, where masking makes the data ready-to-use for that purpose. The original data remains in Cloud Storage.
* Option D:Using CMEK encrypts the entire object in Cloud Storage at rest. While this protects the data in Cloud Storage, federated queries from BigQuery would access the raw, unmasked data (assuming decryption occurs seamlessly). This doesn't address the preprocessing requirement of protectingcertain sensitive data elementswithin the data itself for theconsumer analysesdataset. The goal is to create a de- identified dataset for analysis, not just secure the raw data at rest.
Reference:
Google Cloud Documentation: Cloud Data Loss Prevention > De-identification overview. "De-identification is the process of removing identifying information from data. Cloud DLP uses de-identification techniques such as masking, tokenization, pseudonymization, date shifting, and more to help you protect sensitive data." Google Cloud Documentation: Cloud Data Loss Prevention > Basic de-identification > Masking. "Masking hides parts of data by replacing characters with a symbol, such as an asterisk (*) or hash (#)." Google Cloud Documentation: Dataflow > Overview. "Dataflow is a fully managed streaming analytics service that minimizes latency, processing time, and cost through autoscaling and batch processing." Google Cloud Solution: Automating the de-identification of PII in large-scale datasets using Cloud DLP and Dataflow. This solution guide explicitly outlines using Dataflow and DLP API for de-identifying (including masking) data from Cloud Storage and loading it into BigQuery. "You can use Cloud DLP to scan data for sensitive elements andthen apply de-identification techniques such as redaction, masking, or tokenization." and "This tutorial uses Dataflow to orchestrate the de-identification process."

NEW QUESTION # 290
Your company has recently grown rapidly and now ingesting data at a significantly higher rate than it was previously. You manage the daily batch MapReduce analytics jobs in Apache Hadoop. However, the recent increase in data has meant the batch jobs are falling behind. You were asked to recommend ways the development team could increase the responsiveness of the analytics without increasing costs. What should you recommend they do?

- A. Rewrite the job in Pig.
- B. Increase the size of the Hadoop cluster.
- C. Rewrite the job in Apache Spark.
- D. Decrease the size of the Hadoop cluster but also rewrite the job in Hive.

Answer: A

NEW QUESTION # 291
.....