

Free Professional-Data-Engineer Exam Questions - New Professional-Data-Engineer Test Syllabus

- ✓ Professional-Data-Engineer practice test experience prepares you for the real exam and builds your knowledge of exam objectives.
- ✓ All Professional-Data-Engineer questions types found on the vendor exam are included in your practice test.
- ✓ Detailed explanations for both correct and distractor answers reinforce the material.
- ✓ Practice Mode covers all objectives ensuring topics are covered.
- ✓ Certification Mode (timed) prepares you "exam taking" conditions.
- ✓ Instant, drill-down score reports tell you exactly the areas to focus on.
- ✓ Additional related Web references included.

Professional-Data-Engineer
Practice Test
Reasons to Choose **DumpsBuzz**
Professional-Data-Engineer Practice Test.

DOWNLOAD NOW

BTW, DOWNLOAD part of PDFTorrent Professional-Data-Engineer dumps from Cloud Storage: <https://drive.google.com/open?id=1Oj5QqpSSwH22F6p77wFoDWfuZrkqI8VY>

The PDFTorrent Professional-Data-Engineer PDF file contains the real, valid, and updated Google Professional-Data-Engineer exam practice questions. These are the real Professional-Data-Engineer exam questions that surely will appear in the upcoming exam and by preparing with them you can easily pass the final exam. The Professional-Data-Engineer PDF Questions file is easy to use and install. You can use the Professional-Data-Engineer PDF practice questions on your laptop, desktop, tabs, or even on your smartphone and start Professional-Data-Engineer exam preparation right now.

Achieving the Google Professional-Data-Engineer Certification is a significant accomplishment and can help professionals advance their careers in the field of data engineering. Google Certified Professional Data Engineer Exam certification demonstrates to employers and clients that the candidate has the skills and knowledge necessary to design and manage complex data processing systems using the Google Cloud Platform. Google Certified Professional Data Engineer Exam certification also provides access to a community of certified professionals and other resources that can help candidates stay up-to-date with the latest trends and best practices in data engineering.

>> Free Professional-Data-Engineer Exam Questions <<

New Professional-Data-Engineer Test Syllabus & Reliable Professional-Data-Engineer Braindumps

Our Professional-Data-Engineer training braindumps are famous for its wonderful advantages. The content is carefully designed for the Professional-Data-Engineer exam, rich question bank and answer to enable you to master all the test knowledge in a short period of time. Our Professional-Data-Engineer Exam Questions have helped a large number of candidates pass the Professional-Data-Engineer exam yet. Hope you can join us, and we work together to create a miracle.

Google Professional-Data-Engineer exam is a certification exam offered by Google Cloud Platform for data professionals who want to demonstrate their expertise in designing, building, and managing data processing systems on Google Cloud Platform. It is a highly valued certification in the industry and is especially relevant for those looking to work with Big Data. Professional-Data-Engineer exam tests a candidate's knowledge of various data engineering tools and technologies, and passing the exam demonstrates that the candidate has the skills and knowledge to design and implement data solutions on Google Cloud Platform.

Google Professional-Data-Engineer Exam is designed to test the skills and knowledge of professionals who work with Google Cloud Platform. It is designed for those who have experience in designing, building, and managing data processing systems, and for those who work with machine learning models and data analysis tools. Professional-Data-Engineer exam is designed to test the ability of professionals to use Google Cloud Platform tools to develop and implement solutions that meet the needs of businesses.

Google Certified Professional Data Engineer Exam Sample Questions (Q315-

Q320):

NEW QUESTION # 315

Your company has a hybrid cloud initiative. You have a complex data pipeline that moves data between cloud provider services and leverages services from each of the cloud providers. Which cloud-native service should you use to orchestrate the entire pipeline?

- A. Cloud Dataflow
- B. Cloud Dataprep
- C. **Cloud Composer**
- D. Cloud Dataproc

Answer: C

Explanation:

Cloud Composer uses airflow which is open source and can help to orchestrate jobs.

NEW QUESTION # 316

How can you get a neural network to learn about relationships between categories in a categorical feature?

- A. Create a hash bucket
- B. Create a one-hot column
- C. **Create an embedding column**
- D. Create a multi-hot column

Answer: C

Explanation:

There are two problems with one-hot encoding. First, it has high dimensionality, meaning that instead of having just one value, like a continuous feature, it has many values, or dimensions. This makes computation more time-consuming, especially if a feature has a very large number of categories. The second problem is that it doesn't encode any relationships between the categories. They are completely independent from each other, so the network has no way of knowing which ones are similar to each other.

Both of these problems can be solved by representing a categorical feature with an embedding column. The idea is that each category has a smaller vector with, let's say, 5 values in it. But unlike a one-hot vector, the values are not usually 0. The values are weights, similar to the weights that are used for basic features in a neural network. The difference is that each category has a set of weights (5 of them in this case).

You can think of each value in the embedding vector as a feature of the category. So, if two categories are very similar to each other, then their embedding vectors should be very similar too.

NEW QUESTION # 317

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. **Create an S3-compatible storage endpoint in your network, and use Google Cloud Storage Transfer Service to transfer on-premises data to the designated storage bucket.**
- C. Introduce data compression for each file to increase the rate file of file transfer.
- D. Contact your internet service provider (ISP) to increase your maximum bandwidth to at least 100 Mbps.
- 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,B

NEW QUESTION # 318

You are administering a BigQuery on-demand environment. Your business intelligence tool is submitting hundreds of queries each day that aggregate a large (50 TB) sales history fact table at the day and month levels. These queries have a slow response time and are exceeding cost expectations. You need to decrease response time, lower query costs, and minimize maintenance. What should you do?

- A. Enable BI Engine and add your sales table as a preferred table.
- B. Create a scheduled query to build sales day and sales month aggregate tables on an hourly basis.
- C. Build authorized views on top of the sales table to aggregate data at the day and month level.
- D. **Build materialized views on top of the sales table to aggregate data at the day and month level.**

Answer: D

Explanation:

To improve response times and reduce costs for frequent queries aggregating a large sales history fact table, materialized views are a highly effective solution. Here's why option A is the best choice:

* Materialized Views:

* Materialized views store the results of a query physically and update them periodically, offering faster query responses for frequently accessed data.

* They are designed to improve performance for repetitive and expensive aggregation queries by precomputing the results.

* Efficiency and Cost Reduction:

* By building materialized views at the day and month level, you significantly reduce the computation required for each query, leading to faster response times and lower query costs.

* Materialized views also reduce the need for on-demand query execution, which can be costly when dealing with large datasets.

* Minimized Maintenance:

* Materialized views in BigQuery are managed automatically, with updates handled by the system, reducing the maintenance burden on your team.

Steps to Implement:

* Identify Aggregation Queries:

* Analyze the existing queries to identify common aggregation patterns at the day and month levels.

* Create Materialized Views:

* Create materialized views in BigQuery for the identified aggregation patterns. For example `CREATE MATERIALIZED VIEW project.dataset.sales_daily_summary AS SELECT DATE(transaction_time) AS day, SUM(amount) AS total_sales FROM project.dataset.sales GROUP BY day; CREATE MATERIALIZED VIEW project.dataset.sales_monthly_summary AS SELECT EXTRACT(YEAR FROM transaction_time) AS year, EXTRACT(MONTH FROM transaction_time) AS month, SUM(amount) AS total_sales FROM project.dataset.sales GROUP BY year, month;`

* Query Using Materialized Views:

* Update existing queries to use the materialized views instead of directly querying the base table.

Reference Links:

* [BigQuery Materialized Views](#)

* [Optimizing Query Performance](#)

NEW QUESTION # 319

Your neural network model is taking days to train. You want to increase the training speed. What can you do?

- A. Subsample your training dataset.
- B. Subsample your test dataset.
- C. Increase the number of input features to your model.
- D. **Increase the number of layers in your neural network.**

Answer: D

Explanation:

Explanation/Reference: <https://towardsdatascience.com/how-to-increase-the-accuracy-of-a-neural-network-9f5d1c6f407d>

NEW QUESTION # 320

.....

New Professional-Data-Engineer Test Syllabus: <https://www.pdftorrent.com/Professional-Data-Engineer-exam-prep-dumps.html>

BONUS!!! Download part of PDFTorrent Professional-Data-Engineer dumps for free: <https://drive.google.com/open?id=1Oj5QqpSSwH22F6p77wFoDWfuZrkqJ8VY>