

# Latest Brindumps Google Professional-Data-Engineer Book | Exam Professional-Data-Engineer Simulator Fee



P.S. Free & New Professional-Data-Engineer dumps are available on Google Drive shared by TestkingPass:  
<https://drive.google.com/open?id=1SSv28BCiwsWtWdL68OVXdW6gnyHNve7E>

If you're still learning from the traditional old ways and silently waiting for the test to come, you should be awake and ready to take the exam in a different way. Study our Professional-Data-Engineer training materials to write "test data" is the most suitable for your choice, after recent years show that the effect of our Professional-Data-Engineer Guide Torrent has become a secret weapon of the examinee through qualification examination, a lot of the users of our Professional-Data-Engineer guide torrent can get unexpected results in the examination. Now, I will briefly introduce some details about our Professional-Data-Engineer guide torrent for your reference.

Google Professional-Data-Engineer certification is a valuable credential that can help professionals stand out in the competitive field of data engineering. It demonstrates that the certified individual has the expertise and skills required to design, build, and manage data processing systems effectively. As data becomes increasingly important in organizations of all sizes and industries, the demand for certified data engineers is expected to grow, making this certification a worthwhile investment for individuals looking to advance their career in the field.

Google Professional-Data-Engineer certification exam comprises of multiple-choice and multiple-select questions that require a thorough understanding of Google Cloud Platform services such as BigQuery, Google Cloud Storage, and Google Cloud Dataflow. Professional-Data-Engineer Exam also tests an individual's knowledge of data processing patterns and best practices, understanding of machine learning models and algorithms, and proficiency in designing and deploying solutions that meet business requirements.

## Difficulty in Attempting Google Professional Data Engineer Exam Certification

If the user has successfully passed the **professional-data-engineer practice exam** and has been through **professional-data-engineer exam dumps** then the certification exam will not be too much difficult as the user has shown aptitude for understanding complicated processes.

>> Latest Brindumps Google Professional-Data-Engineer Book <<

## Professional-Data-Engineer Test Preparation & Professional-Data-Engineer Exam Questions & Professional-Data-Engineer Test Prep

Many users report to us that they are very fond of writing their own notes while they are learning. This will enhance their memory and make it easier to review. Our Professional-Data-Engineer exam questions have created a PDF version of the Professional-Data-Engineer practice material to meet the needs of this group of users. You can print the PDF version of the Professional-Data-Engineer learning guide so that you can carry it with you. As long as you have time, you can take it out to read and write your own experience.

## Google Certified Professional Data Engineer Exam Sample Questions (Q381-Q386):

### NEW QUESTION # 381

You are designing a data mesh on Google Cloud with multiple distinct data engineering teams building data products. The typical data curation design pattern consists of landing files in Cloud Storage, transforming raw data in Cloud Storage and BigQuery datasets, and storing the final curated data product in BigQuery datasets. You need to configure Dataplex to ensure that each team can access only the assets needed to build their data products. You also need to ensure that teams can easily share the curated data product. What should you do?

- A. 1 Create a single Dataplex virtual lake and create a single zone to contain landing, raw, and curated data.  
2 Provide each data engineering team access to the virtual lake.
- B. 1 Create a Dataplex virtual lake for each data product, and create multiple zones for landing, raw, and curated data.  
2. Provide the data engineering teams with full access to the virtual lake assigned to their data product.
- C. 1 Create a single Dataplex virtual lake and create a single zone to contain landing, raw, and curated data. 2 Build separate assets for each data product within the zone.  
3. Assign permissions to the data engineering teams at the zone level.
- D. 1 Create a Dataplex virtual lake for each data product, and create a single zone to contain landing, raw, and curated data.  
2. Provide the data engineering teams with full access to the virtual lake assigned to their data product.

**Answer: B**

Explanation:

This option is the best way to configure Dataplex for a data mesh architecture, as it allows each data engineering team to have full ownership and control over their data products, while also enabling easy discovery and sharing of the curated data across the organization<sup>1</sup>2. By creating a Dataplex virtual lake for each data product, you can isolate the data assets and resources for each domain, and avoid conflicts and dependencies between different teams<sup>3</sup>. By creating multiple zones for landing, raw, and curated data, you can enforce different security and governance policies for each stage of the data curation process, and ensure that only authorized users can access the data assets<sup>4</sup>5. By providing the data engineering teams with full access to the virtual lake assigned to their data product, you can empower them to manage and monitor their data products, and leverage the Dataplex features such as tagging, quality, and lineage.

Option A is not suitable, as it creates a single point of failure and a bottleneck for the data mesh, and does not allow for fine-grained access control and governance for different data products<sup>2</sup>. Option B is also not suitable, as it does not isolate the data assets and resources for each data product, and assigns permissions at the zone level, which may not reflect the different roles and responsibilities of the data engineering teams<sup>3</sup>4. Option C is better than option A and B, but it does not create multiple zones for landing, raw, and curated data, which may compromise the security and quality of the data products<sup>5</sup>. Reference:

1: Building a data mesh on Google Cloud using BigQuery and Dataplex | Google Cloud Blog

2: Data Mesh - 7 Effective Practices to Get Started - Confluent

3: Best practices | Dataplex | Google Cloud

4: Secure your lake | Dataplex | Google Cloud

5: Zones | Dataplex | Google Cloud

[6]: Managing a Data Mesh with Dataplex - ROI Training

### NEW QUESTION # 382

MJTelco Case Study

Company Overview

MJTelco is a startup that plans to build networks in rapidly growing, underserved markets around the world. The company has patents for innovative optical communications hardware. Based on these patents, they can create many reliable, high-speed backbone links with inexpensive hardware.

Company Background

Founded by experienced telecom executives, MJTelco uses technologies originally developed to overcome communications challenges in space. Fundamental to their operation, they need to create a distributed data infrastructure that drives real-time analysis and incorporates machine learning to continuously optimize their topologies. Because their hardware is inexpensive, they plan to overdeploy the network allowing them to account for the impact of dynamic regional politics on location availability and cost.

Their management and operations teams are situated all around the globe creating many-to-many relationship between data consumers and provides in their system. After careful consideration, they decided public cloud is the perfect environment to support their needs.

Solution Concept

MJTelco is running a successful proof-of-concept (PoC) project in its labs. They have two primary needs:

Scale and harden their PoC to support significantly more data flows generated when they ramp to more than 50,000 installations.

Refine their machine-learning cycles to verify and improve the dynamic models they use to control topology definition.

MJTelco will also use three separate operating environments - development/test, staging, and production - to meet the needs of running experiments, deploying new features, and serving production customers.

#### Business Requirements

Scale up their production environment with minimal cost, instantiating resources when and where needed in an unpredictable, distributed telecom user community.

Ensure security of their proprietary data to protect their leading-edge machine learning and analysis.

Provide reliable and timely access to data for analysis from distributed research workers

Maintain isolated environments that support rapid iteration of their machine-learning models without affecting their customers.

#### Technical Requirements

Ensure secure and efficient transport and storage of telemetry data

Rapidly scale instances to support between 10,000 and 100,000 data providers with multiple flows each.

Allow analysis and presentation against data tables tracking up to 2 years of data storing approximately 100m records/day

Support rapid iteration of monitoring infrastructure focused on awareness of data pipeline problems both in telemetry flows and in production learning cycles.

#### CEO Statement

Our business model relies on our patents, analytics and dynamic machine learning. Our inexpensive hardware is organized to be highly reliable, which gives us cost advantages. We need to quickly stabilize our large distributed data pipelines to meet our reliability and capacity commitments.

#### CTO Statement

Our public cloud services must operate as advertised. We need resources that scale and keep our data secure. We also need environments in which our data scientists can carefully study and quickly adapt our models. Because we rely on automation to process our data, we also need our development and test environments to work as we iterate.

#### CFO Statement

The project is too large for us to maintain the hardware and software required for the data and analysis.

Also, we cannot afford to staff an operations team to monitor so many data feeds, so we will rely on automation and infrastructure. Google Cloud's machine learning will allow our quantitative researchers to work on our high-value problems instead of problems with our data pipelines.

You need to compose visualization for operations teams with the following requirements:

Telemetry must include data from all 50,000 installations for the most recent 6 weeks (sampling once every minute)

The report must not be more than 3 hours delayed from live data.

The actionable report should only show suboptimal links.

Most suboptimal links should be sorted to the top.

Suboptimal links can be grouped and filtered by regional geography.

User response time to load the report must be <5 seconds.

You create a data source to store the last 6 weeks of data, and create visualizations that allow viewers to see multiple date ranges, distinct geographic regions, and unique installation types. You always show the latest data without any changes to your visualizations. You want to avoid creating and updating new visualizations each month. What should you do?

- A. Look through the current data and compose a series of charts and tables, one for each possible combination of criteria.
- B. Load the data into relational database tables, write a Google App Engine application that queries all rows, summarizes the data across each criteria, and then renders results using the Google Charts and visualization API.
- C. Look through the current data and compose a small set of generalized charts and tables bound to criteria filters that allow value selection.
- D. Export the data to a spreadsheet, compose a series of charts and tables, one for each possible combination of criteria, and spread them across multiple tabs.

**Answer: C**

#### NEW QUESTION # 383

If you're running a performance test that depends upon Cloud Bigtable, all the choices except one below are recommended steps.

Which is NOT a recommended step to follow?

- A. Do not use a production instance.
- B. Use at least 300 GB of data.
- C. Run your test for at least 10 minutes.
- D. Before you test, run a heavy pre-test for several minutes.

**Answer: A**

Explanation:

If you're running a performance test that depends upon Cloud Bigtable, be sure to follow these steps as you plan and execute your test:

Use a production instance. A development instance will not give you an accurate sense of how a production instance performs under load.

Use at least 300 GB of data. Cloud Bigtable performs best with 1 TB or more of data.

However, 300 GB of data is enough to provide reasonable results in a performance test on a 3-node cluster. On larger clusters, use 100 GB of data per node.

Before you test, run a heavy pre-test for several minutes. This step gives Cloud Bigtable a chance to balance data across your nodes based on the access patterns it observes.

Run your test for at least 10 minutes. This step lets Cloud Bigtable further optimize your data, and it helps ensure that you will test reads from disk as well as cached reads from memory.

Reference: <https://cloud.google.com/bigtable/docs/performance>

#### NEW QUESTION # 384

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 the master node of the Dataproc cluster. Then run the Hadoop utility to copy them to HDFS. Mount the Hive tables from HDFS.
- 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 HDFS. Mount the Hive tables locally.
- D. 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.
- E. 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.

**Answer: A,D**

#### NEW QUESTION # 385

You are designing a basket abandonment system for an ecommerce company. The system will send a message to a user based on these rules:

No interaction by the user on the site for 1 hour

☐ Has added more than \$30 worth of products to the basket  
☐ Has not completed a transaction

You use Google Cloud Dataflow to process the data and decide if a message should be sent. How should you design the pipeline?

- A. Use a fixed-time window with a duration of 60 minutes.
- B. Use a sliding time window with a duration of 60 minutes.
- C. Use a session window with a gap time duration of 60 minutes.
- D. Use a global window with a time based trigger with a delay of 60 minutes.

**Answer: D**

#### NEW QUESTION # 386

The more you practice with our Professional-Data-Engineer simulating exam, the more compelling you may feel. Even if you are lack of time, these Professional-Data-Engineer practice materials can speed up your pace of review. Our Professional-Data-Engineer guide questions are motivating materials especially suitable for those exam candidates who are eager to pass the exam with efficiency. And we can claim that with our Professional-Data-Engineer study braindumps for 20 to 30 hours, you will be bound to pass the exam.

[illegible]

DOWNLOAD the newest TestkingPass Professional-Data-Engineer PDF dumps from Cloud Storage for free:  
<https://drive.google.com/open?id=1SSv28BCiwsWdL68OVXdW6gnvHNve7E>