



PROFESSIONAL-DATA-ENGINEER^{Q&As}

Professional Data Engineer on Google Cloud Platform

**Pass Google PROFESSIONAL-DATA-ENGINEER
Exam with 100% Guarantee**

Free Download Real Questions & Answers **PDF** and **VCE** file from:

<https://www.pass4itsure.com/professional-data-engineer.html>

100% Passing Guarantee
100% Money Back Assurance

Following Questions and Answers are all new published by Google
Official Exam Center



VCE & PDF

Pass4itSure.com

<https://www.pass4itsure.com/professional-data-engineer.html>

2024 Latest pass4itsure PROFESSIONAL-DATA-ENGINEER PDF and VCE
dumps Download

- ⚙️ **Instant Download** After Purchase
- ⚙️ **100% Money Back** Guarantee
- ⚙️ **365 Days** Free Update
- ⚙️ **800,000+** Satisfied Customers



**QUESTION 1**

You are selecting services to write and transform JSON messages from Cloud Pub/Sub to BigQuery for a data pipeline on Google Cloud. You want to minimize service costs. You also want to monitor and accommodate input data volume that will vary in size with minimal manual intervention. What should you do?

- A. Use Cloud Dataproc to run your transformations. Monitor CPU utilization for the cluster. Resize the number of worker nodes in your cluster via the command line.
- B. Use Cloud Dataproc to run your transformations. Use the diagnose command to generate an operational output archive. Locate the bottleneck and adjust cluster resources.
- C. Use Cloud Dataflow to run your transformations. Monitor the job system lag with Stackdriver. Use the default autoscaling setting for worker instances.
- D. Use Cloud Dataflow to run your transformations. Monitor the total execution time for a sampling of jobs. Configure the job to use non-default Compute Engine machine types when needed.

Correct Answer: B

QUESTION 2

Your company needs to upload their historic data to Cloud Storage. The security rules don't allow access from external IPs to their on-premises resources. After an initial upload, they will add new data from existing on-premises applications every day. What should they do?

- A. Execute gsutil rsync from the on-premises servers.
- B. Use Cloud Dataflow and write the data to Cloud Storage.
- C. Write a job template in Cloud Dataproc to perform the data transfer.
- D. Install an FTP server on a Compute Engine VM to receive the files and move them to Cloud Storage.

Correct Answer: B

QUESTION 3

Your company's on-premises Apache Hadoop servers are approaching end-of-life, and IT has decided to migrate the cluster to Google Cloud Dataproc. A like-for-like migration of the cluster would require 50 TB of Google Persistent Disk per node. The CIO is concerned about the cost of using that much block storage. You want to minimize the storage cost of the migration. What should you do?

- A. Put the data into Google Cloud Storage.
- B. Use preemptible virtual machines (VMs) for the Cloud Dataproc cluster.
- C. Tune the Cloud Dataproc cluster so that there is just enough disk for all data.



D. Migrate some of the cold data into Google Cloud Storage, and keep only the hot data in Persistent Disk.

Correct Answer: B

QUESTION 4

You need to move 2 PB of historical data from an on-premises storage appliance to Cloud Storage within six months, and your outbound network capacity is constrained to 20 Mb/sec. How should you migrate this data to Cloud Storage?

A. Use Transfer Appliance to copy the data to Cloud Storage

B. Use gsutil cp to compress the content being uploaded to Cloud Storage

C. Create a private URL for the historical data, and then use Storage Transfer Service to copy the data to Cloud Storage

D. Use trickle or ionice along with gsutil cp to limit the amount of bandwidth gsutil utilizes to less than 20 Mb/sec so it does not interfere with the production traffic

Correct Answer: A

QUESTION 5

You are configuring networking for a Dataflow job. The data pipeline uses custom container images with the libraries that are required for the transformation logic preinstalled. The data pipeline reads the data from Cloud Storage and writes the data to BigQuery. You need to ensure cost-effective and secure communication between the pipeline and Google APIs and services. What should you do?

A. Leave external IP addresses assigned to worker VMs while enforcing firewall rules.

B. Disable external IP addresses and establish a Private Service Connect endpoint IP address.

C. Disable external IP addresses from worker VMs and enable Private Google Access.

D. Enable Cloud NAT to provide outbound internet connectivity while enforcing firewall rules.

Correct Answer: C

Private Google Access allows VMs without external IP addresses to communicate with Google APIs and services over internal routes. This reduces the cost and increases the security of the data pipeline. Custom container images can be stored in Container Registry, which supports Private Google Access. Dataflow supports Private Google Access for both batch and streaming jobs. References: Private Google Access overview Using Private Google Access and Cloud NAT Using custom containers with Dataflow

[Latest PROFESSIONAL-DATA-ENGINEER Dumps](#)

[PROFESSIONAL-DATA-ENGINEER PDF Dumps](#)

[PROFESSIONAL-DATA-ENGINEER Study Guide](#)