



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 developing an application on Google Cloud that will automatically generate subject labels for users' blog posts. You are under competitive pressure to add this feature quickly, and you have no additional developer resources. No one on your team has experience with machine learning. What should you do?

- A. Call the Cloud Natural Language API from your application. Process the generated Entity Analysis as labels.
- B. Call the Cloud Natural Language API from your application. Process the generated Sentiment Analysis as labels.
- C. Build and train a text classification model using TensorFlow. Deploy the model using Cloud Machine Learning Engine. Call the model from your application and process the results as labels.
- D. Build and train a text classification model using TensorFlow. Deploy the model using a Kubernetes Engine cluster. Call the model from your application and process the results as labels.

Correct Answer: B

QUESTION 2

You work for a shipping company that has distribution centers where packages move on delivery lines to route them properly. The company wants to add cameras to the delivery lines to detect and track any visual damage to the packages in transit. You need to create a way to automate the detection of damaged packages and flag them for human review in real time while the packages are in transit. Which solution should you choose?

- A. Use BigQuery machine learning to be able to train the model at scale, so you can analyze the packages in batches.
- B. Train an AutoML model on your corpus of images, and build an API around that model to integrate with the package tracking applications.
- C. Use the Cloud Vision API to detect for damage, and raise an alert through Cloud Functions. Integrate the package tracking applications with this function.
- D. Use TensorFlow to create a model that is trained on your corpus of images. Create a Python notebook in Cloud Datalab that uses this model so you can analyze for damaged packages.

Correct Answer: A

QUESTION 3

You stream order data by using a Dataflow pipeline, and write the aggregated result to Memorystore. You provisioned a Memorystore for Redis instance with Basic Tier, 4 GB capacity, which is used by 40 clients for read-only access. You are expecting the number of read-only clients to increase significantly to a few hundred and you need to be able to support the demand. You want to ensure that read and write access availability is not impacted, and any changes you make can be deployed quickly. What should you do?

- A. Create multiple new Memorystore for Redis instances with Basic Tier (4 GB capacity) Modify the Dataflow pipeline and new clients to use all instances



B. Create a new Memorystore for Redis instance with Standard Tier Set capacity to 4 GB and read replica to No read replicas (high availability only). Delete the old instance.

C. Create a new Memorystore for Memcached instance Set a minimum of three nodes, and memory per node to 4 GB. Modify the Dataflow pipeline and all clients to use the Memcached instance Delete the old instance.

D. Create a new Memorystore for Redis instance with Standard Tier Set capacity to 5 GB and create multiple read replicas Delete the old instance.

Correct Answer: D

The Basic Tier of Memorystore for Redis provides a standalone Redis instance that is not replicated and does not support read replicas. This means that it cannot scale horizontally to handle more read requests, and it does not provide high availability or automatic failover. If the number of read-only clients increases significantly, the Basic Tier instance may not be able to handle the demand and may impact the read and write access availability. Therefore, option A is not a good solution, as it would require creating multiple Basic Tier instances and modifying the Dataflow pipeline and the clients to distribute the load among them. This would increase the complexity and the management overhead of the solution. The Standard Tier of Memorystore for Redis provides a highly available Redis instance that supports replication and read replicas. Replication ensures that the data is backed up in another zone and can fail over automatically in case of a primary node failure. Read replicas allow scaling the read throughput by adding up to five replicas to an instance and using them for read-only queries. The Standard Tier also supports in-transit encryption and maintenance windows. Therefore, option D is the best solution, as it would create a new Standard Tier instance with a higher capacity (5 GB) and multiple read replicas to handle the increased demand. The old instance can be deleted after migrating the data to the new instance. Option B is not a good solution, as it would create a new Standard Tier instance with the same capacity (4 GB) and no read replicas. This would not improve the read throughput or the availability of the solution. Option C is not a good solution, as it would create a new Memorystore for Memcached instance, which is a different service that uses a different protocol and data model than Redis. This would require changing the code of the Dataflow pipeline and the clients to use the Memcached protocol and data structures, which would take more time and effort than migrating to a new Redis instance. References: Redis tier capabilities | Memorystore for Redis | Google Cloud, Pricing | Memorystore for Redis | Google Cloud, What is Memorystore? | Google Cloud Blog, Working with GCP Memorystore - Simple Talk - Redgate Software

QUESTION 4

You are choosing a NoSQL database to handle telemetry data submitted from millions of Internet-of-Things (IoT) devices. The volume of data is growing at 100 TB per year, and each data entry has about 100 attributes. The data processing pipeline does not require atomicity, consistency, isolation, and durability (ACID). However, high availability and low latency are required.

You need to analyze the data by querying against individual fields. Which three databases meet your requirements? (Choose three.)

A. Redis

B. HBase

C. MySQL

D. MongoDB

E. Cassandra

F. HDFS with Hive

Correct Answer: BDF

**QUESTION 5**

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. Introduce data compression for each file to increase the rate file of file transfer.
- B. Contact your internet service provider (ISP) to increase your maximum bandwidth to at least 100 Mbps.
- C. Redesign the data ingestion process to use gsutil tool to send the CSV files to a storage bucket in parallel.
- D. 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.
- E. 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.

Correct Answer: CE

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

[PROFESSIONAL-DATA-ENGINEER Practice Test](#)

[PROFESSIONAL-DATA-ENGINEER Braindumps](#)