

[2023 Professional-Data-Engineer Actual Exam Dumps, Professional-Data-Engineer Practice Test [Q156-Q179]



[2023] Professional-Data-Engineer Actual Exam Dumps, Professional-Data-Engineer Practice Test
BraindumpsIT Professional-Data-Engineer dumps & Google Cloud Certified sure practice dumps

Google Professional-Data-Engineer certification is a highly respected credential that validates the knowledge and skills of professionals working in the field of data engineering. Google Certified Professional Data Engineer Exam certification is designed to test the ability of candidates to design, build, operate, and manage data processing systems that are scalable, secure, and reliable. Google Certified Professional Data Engineer Exam certification exam is conducted by Google and is intended for individuals who have a good understanding of the Google Cloud Platform and data engineering best practices.

Q156. Which of the following is not true about Dataflow pipelines?

- * Pipelines are a set of operations
- * Pipelines represent a data processing job
- * Pipelines represent a directed graph of steps
- * Pipelines can share data between instances

Explanation

The data and transforms in a pipeline are unique to, and owned by, that pipeline. While your program can create multiple pipelines, pipelines cannot share data or transforms Reference: <https://cloud.google.com/dataflow/model/pipelines>

Q157. You want to use a database of information about tissue samples to classify future tissue samples as either normal or mutated. You are evaluating an unsupervised anomaly detection method for classifying the tissue samples. Which two characteristic support this method? (Choose two.)

- * There are very few occurrences of mutations relative to normal samples.

- * There are roughly equal occurrences of both normal and mutated samples in the database.
- * You expect future mutations to have different features from the mutated samples in the database.
- * You expect future mutations to have similar features to the mutated samples in the database.
- * You already have labels for which samples are mutated and which are normal in the database.

Unsupervised anomaly detection techniques detect anomalies in an unlabeled test data set under the assumption that the majority of the instances in the data set are normal by looking for instances that seem to fit least to the remainder of the data set.

https://en.wikipedia.org/wiki/Anomaly_detection

Q158. How can you get a neural network to learn about relationships between categories in a categorical feature?

- * Create a multi-hot column
- * Create a one-hot column
- * Create a hash bucket
- * Create an embedding column

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.

Reference: <https://cloudacademy.com/google/introduction-to-google-cloud-machine-learning-engine-course/a-wide-and-deep-model.html>

Q159. You are working on a sensitive project involving private user data

a. You have set up a project on Google Cloud Platform to house your work internally. An external consultant is going to assist with coding a complex transformation in a Google Cloud Dataflow pipeline for your project. How should you maintain users' privacy?

- * Grant the consultant the Viewer role on the project.
- * Grant the consultant the Cloud Dataflow Developer role on the project.
- * Create a service account and allow the consultant to log on with it.
- * Create an anonymized sample of the data for the consultant to work with in a different project.

Q160. 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?

- * Use a fixed-time window with a duration of 60 minutes.
- * Use a sliding time window with a duration of 60 minutes.
- * Use a session window with a gap time duration of 60 minutes.
- * Use a global window with a time based trigger with a delay of 60 minutes.

It will send a message per user after that user is inactive for 60 minutes. Session window works well for capturing a session per user basis.

Q161. You are a head of BI at a large enterprise company with multiple business units that each have different priorities and budgets. You use on-demand pricing for BigQuery with a quota of 2K concurrent on-demand slots per project. Users at your organization sometimes don't get slots to execute their query and you need to correct this. You'd like to avoid introducing new projects to your account.

What should you do?

- * Convert your batch BQ queries into interactive BQ queries.
- * Create an additional project to overcome the 2K on-demand per-project quota.
- * Switch to flat-rate pricing and establish a hierarchical priority model for your projects.
- * Increase the amount of concurrent slots per project at the Quotas page at the Cloud Console.

Reference <https://cloud.google.com/blog/products/gcp/busting-12-myths-about-bigquery>

Q162. You need to choose a database for a new project that has the following requirements:

- * Fully managed
- * Able to automatically scale up
- * Transactionally consistent
- * Able to scale up to 6 TB
- * Able to be queried using SQL

Which database do you choose?

- * Cloud SQL
- * Cloud Bigtable
- * Cloud Spanner
- * Cloud Datastore

<https://cloud.google.com/products/databases>

Q163. Business owners at your company have given you a database of bank transactions. Each row contains the user ID, transaction type, transaction location, and transaction amount. They ask you to investigate what type of machine learning can be applied to the data. Which three machine learning applications can you use?

(Choose three.)

- * Supervised learning to determine which transactions are most likely to be fraudulent.
- * Unsupervised learning to determine which transactions are most likely to be fraudulent.
- * Clustering to divide the transactions into N categories based on feature similarity.
- * Supervised learning to predict the location of a transaction.
- * Reinforcement learning to predict the location of a transaction.
- * Unsupervised learning to predict the location of a transaction.

Q164. 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.

MJTelco needs you to create a schema in Google Bigtable that will allow for the historical analysis of the last 2 years of records. Each record that comes in is sent every 15 minutes, and contains a unique identifier of the device and a data record. The most common query is for all the data for a given device for a given day. Which schema should you use?

* Rowkey: date#device_id

Column data: data_point

* Rowkey: date

Column data: device_id,data_point

* Rowkey: device_id

Column data: date, data_point

* Rowkey: data_point

Column data: device_id,date

* Rowkey: date#data_point

Column data: device_id

Q165. 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 into. 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 parity 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 our 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.

Flowlogic'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?

- * Cloud Pub/Sub, Cloud Dataflow, and Cloud Storage
- * Cloud Pub/Sub, Cloud Dataflow, and Local SSD
- * Cloud Pub/Sub, Cloud SQL, and Cloud Storage
- * Cloud Load Balancing, Cloud Dataflow, and Cloud Storage

Q166. Which of these is NOT a way to customize the software on Dataproc cluster instances?

- * Set initialization actions
- * Modify configuration files using cluster properties
- * Configure the cluster using Cloud Deployment Manager
- * Log into the master node and make changes from there

You can access the master node of the cluster by clicking the SSH button next to it in the Cloud Console.

You can easily use the `--properties` option of the `dataproc` command in the Google Cloud SDK to modify many common configuration files when creating a cluster.

When creating a Cloud Dataproc cluster, you can specify initialization actions in executables and/or scripts that Cloud Dataproc will run on all nodes in your Cloud Dataproc cluster immediately after the cluster is set up.

[<https://cloud.google.com/dataproc/docs/concepts/configuring-clusters/init-actions>]

Q167. An online retailer has built their current application on Google App Engine. A new initiative at the company mandates that they extend their application to allow their customers to transact directly via the application.

They need to manage their shopping transactions and analyze combined data from multiple datasets using a business intelligence (BI) tool. They want to use only a single database for this purpose. Which Google Cloud database should they choose?

- * BigQuery
- * Cloud SQL
- * Cloud BigTable
- * Cloud Datastore

Explanation/Reference: <https://cloud.google.com/solutions/business-intelligence/>

Q168. To give a user read permission for only the first three columns of a table, which access control method would you use?

- * Primitive role
- * Predefined role
- * Authorized view
- * It's not possible to give access to only the first three columns of a table.

An authorized view allows you to share query results with particular users and groups without giving them read access to the

underlying tables. Authorized views can only be created in a dataset that does not contain the tables queried by the view.

When you create an authorized view, you use the view's SQL query to restrict access to only the rows and columns you want the users to see.

Q169. You are planning to use Google's Dataflow SDK to analyze customer data such as displayed below. Your project requirement is to extract only the customer name from the data source and then write to an output PCollection.

Tom,555 X street

Tim,553 Y street

Sam, 111 Z street

Which operation is best suited for the above data processing requirement?

- * ParDo
- * Sink API
- * Source API
- * Data extraction

Explanation

In Google Cloud dataflow SDK, you can use the ParDo to extract only a customer name of each element in your PCollection.

Reference: <https://cloud.google.com/dataflow/model/par-do>

Q170. What Dataflow concept determines when a Window's contents should be output based on certain criteria being met?

- * Sessions
- * OutputCriteria
- * Windows
- * Triggers

Triggers control when the elements for a specific key and window are output. As elements arrive, they are put into one or more windows by a Window transform and its associated WindowFn, and then passed to the associated Trigger to determine if the Windows contents should be output.

Q171. Each analytics team in your organization is running BigQuery jobs in their own projects. You want to enable each team to monitor slot usage within their projects. What should you do?

- * Create a Stackdriver Monitoring dashboard based on the BigQuery metric query/scanned_bytes
- * Create a Stackdriver Monitoring dashboard based on the BigQuery metric slots/allocated_for_project
- * Create a log export for each project, capture the BigQuery job execution logs, create a custom metric based on the totalSlotMs, and create a Stackdriver Monitoring dashboard based on the custom metric
- * Create an aggregated log export at the organization level, capture the BigQuery job execution logs, create a custom metric based on the totalSlotMs, and create a Stackdriver Monitoring dashboard based on the custom metric

Q172. Your company is performing data preprocessing for a learning algorithm in Google Cloud Dataflow.

Numerous data logs are being generated during this step, and the team wants to analyze them. Due to the dynamic nature of the campaign, the data is growing exponentially every hour.

The data scientists have written the following code to read the data for a new key features in the logs.

BigQueryIO.Read

```
.named('ReadLogData');
```

```
.from('clouddataflow-readonly:samples.log_data');
```

You want to improve the performance of this data read. What should you do?

- * Specify the TableReference object in the code.
- * Use .fromQuery operation to read specific fields from the table.
- * Use of both the Google BigQuery TableSchema and TableFieldSchema classes.
- * Call a transform that returns TableRow objects, where each element in the PCollection represents a single row in the table.

Q173. All Google Cloud Bigtable client requests go through a front-end server _____ they are sent to a Cloud Bigtable node.

- * before
- * after
- * only if
- * once

Explanation

In a Cloud Bigtable architecture all client requests go through a front-end server before they are sent to a Cloud Bigtable node.

The nodes are organized into a Cloud Bigtable cluster, which belongs to a Cloud Bigtable instance, which is a container for the cluster. Each node in the cluster handles a subset of the requests to the cluster.

When additional nodes are added to a cluster, you can increase the number of simultaneous requests that the cluster can handle, as well as the maximum throughput for the entire cluster.

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

Q174. 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.)

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

Q175. You are migrating your data warehouse to BigQuery. You have migrated all of your data into tables in a dataset. Multiple users from your organization will be using the data. They should only see certain tables based on their team membership. How should you set user permissions?

- * Assign the users/groups data viewer access at the table level for each table

- * Create SQL views for each team in the same dataset in which the data resides, and assign the users/groups data viewer access to the SQL views
- * Create authorized views for each team in the same dataset in which the data resides, and assign the users/groups data viewer access to the authorized views
- * Create authorized views for each team in datasets created for each team. Assign the authorized views data viewer access to the dataset in which the data resides. Assign the users/groups data viewer access to the datasets in which the authorized views reside

Q176. What is the recommended action to do in order to switch between SSD and HDD storage for your Google Cloud Bigtable instance?

- * create a third instance and sync the data from the two storage types via batch jobs
- * export the data from the existing instance and import the data into a new instance
- * run parallel instances where one is HDD and the other is SDD
- * the selection is final and you must resume using the same storage type

Explanation

When you create a Cloud Bigtable instance and cluster, your choice of SSD or HDD storage for the cluster is permanent. You cannot use the Google Cloud Platform Console to change the type of storage that is used for the cluster.

If you need to convert an existing HDD cluster to SSD, or vice-versa, you can export the data from the existing instance and import the data into a new instance. Alternatively, you can write a Cloud Dataflow or Hadoop MapReduce job that copies the data from one instance to another.

Reference: <https://cloud.google.com/bigtable/docs/choosing-ssd-hdd->

Q177. 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?

- * 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.
- * Use Cloud Dataproc to run your transformations. Use the diagnose command to generate an operational output archive. Locate the bottleneck and adjust cluster resources.
- * Use Cloud Dataflow to run your transformations. Monitor the job system lag with Stackdriver. Use the

default autoscaling setting for worker instances.

- * 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.

Q178. Case Study 2 – MJTelco

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.

MJTelco's Google Cloud Dataflow pipeline is now ready to start receiving data from the 50,000 installations. You want to allow Cloud Dataflow to scale its compute power up as required. Which Cloud Dataflow pipeline configuration setting should you update?

- * The zone
- * The number of workers
- * The disk size per worker
- * The maximum number of workers

Q179. Which row keys are likely to cause a disproportionate number of reads and/or writes on a particular node in a Bigtable cluster (select 2 answers)?

- * A sequential numeric ID
- * A timestamp followed by a stock symbol
- * A non-sequential numeric ID
- * A stock symbol followed by a timestamp

Explanation

using a timestamp as the first element of a row key can cause a variety of problems.

In brief, when a row key for a time series includes a timestamp, all of your writes will target a single node; fill that node; and then move onto the next node in the cluster, resulting in hotspotting.

Suppose your system assigns a numeric ID to each of your application's users. You might be tempted to use the user's numeric ID as the row key for your table. However, since new users are more likely to be active users, this approach is likely to push most of your traffic to a small number of nodes.

[<https://cloud.google.com/bigtable/docs/schema-design>]

Reference:

https://cloud.google.com/bigtable/docs/schema-design-time-series#ensure_that_your_row_key_avoids_hotspotti

Professional-Data-Engineer Actual Questions and Braindumps:

https://www.braindumpsit.com/Professional-Data-Engineer_real-exam.html