

Summary

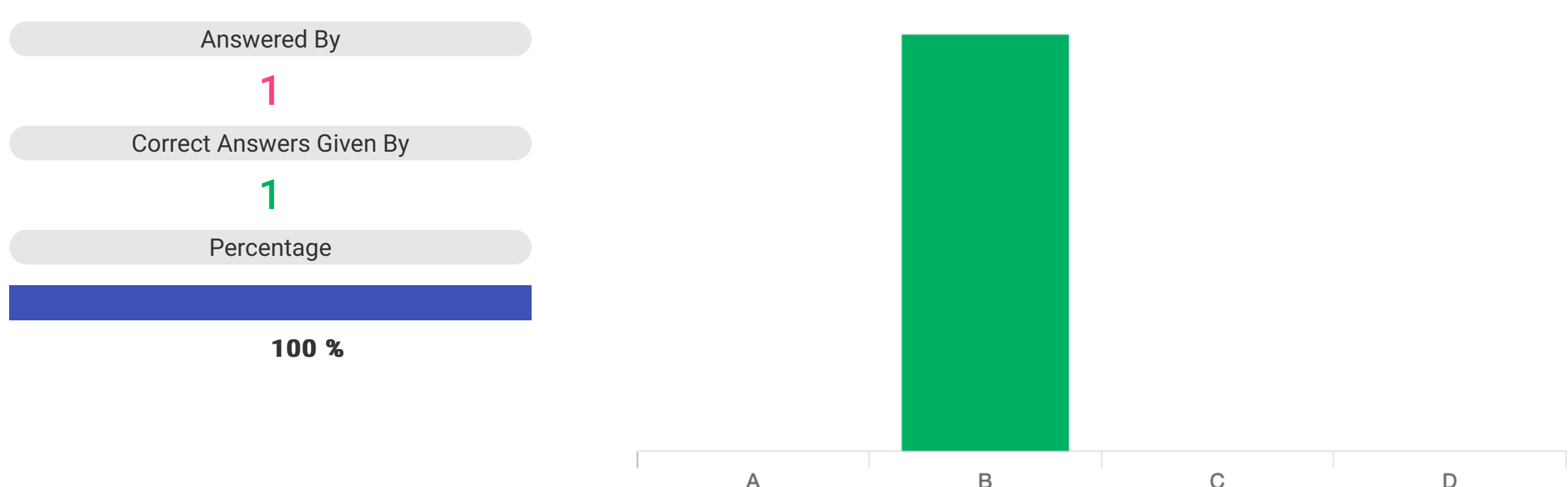
Exit

Question 1:

Your company has successfully migrated to the cloud and wants to analyze their data stream to optimize operations. They do not have any existing code for this analysis, so they are exploring all their options. These options include a mix of batch and stream processing, as they are running some hourly jobs and live- processing some data as it comes in. Which technology should they use for this?

- A. Google Cloud Dataproc
- B. Google Cloud Dataflow
- C. Google Container Engine with Bigtable
- D. Google Compute Engine with Google BigQuery

Type : SINGLE SELECTION

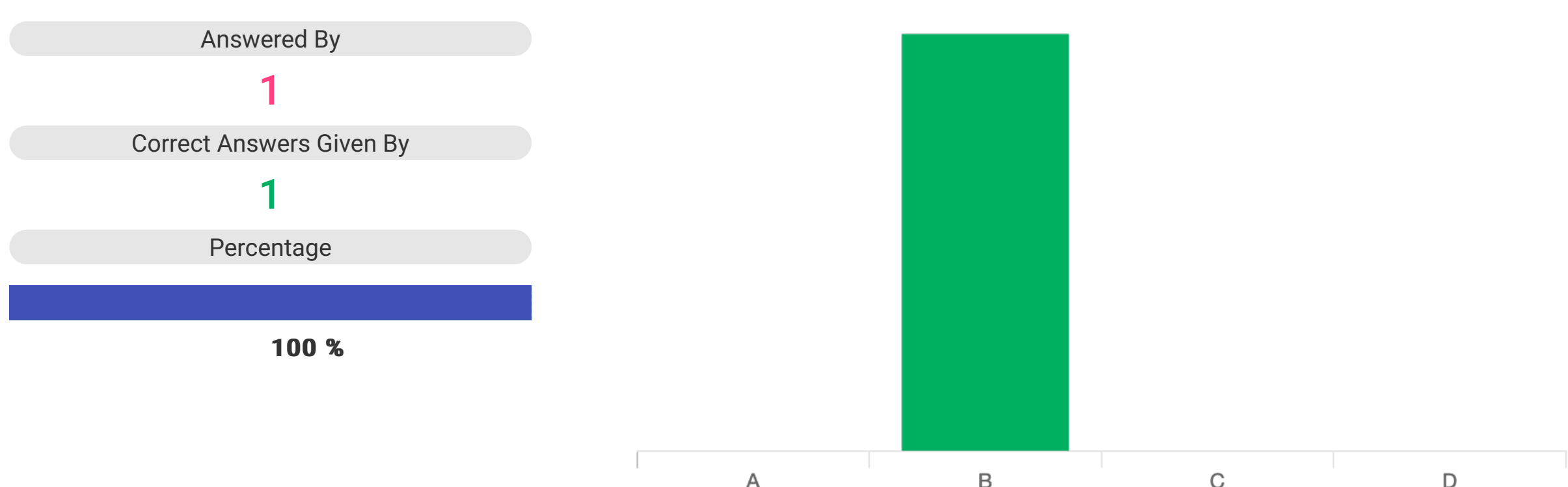


Question 2:

Your company is forecasting a sharp increase in the number and size of Apache Spark and Hadoop jobs being run on your local datacenter. You want to utilize the cloud to help you scale this upcoming demand with the least amount of operations work and code change.

- A. Google Cloud Dataflow
- B. Google Cloud Dataproc
- C. Google Compute Engine
- D. Google Kubernetes Engine

Type : SINGLE SELECTION



Question 3:

You need to migrate Hadoop jobs for your company's Data Science team without modifying the underlying infrastructure. You want to minimize costs and infrastructure management effort. What should you do?

- A. Create a Dataproc cluster using standard worker instances.
- B. Create a Dataproc cluster using preemptible worker instances.
- C. Manually deploy a Hadoop cluster on Compute Engine using standard instances.
- D. Manually deploy a Hadoop cluster on Compute Engine using preemptible instances.

Type : SINGLE SELECTION

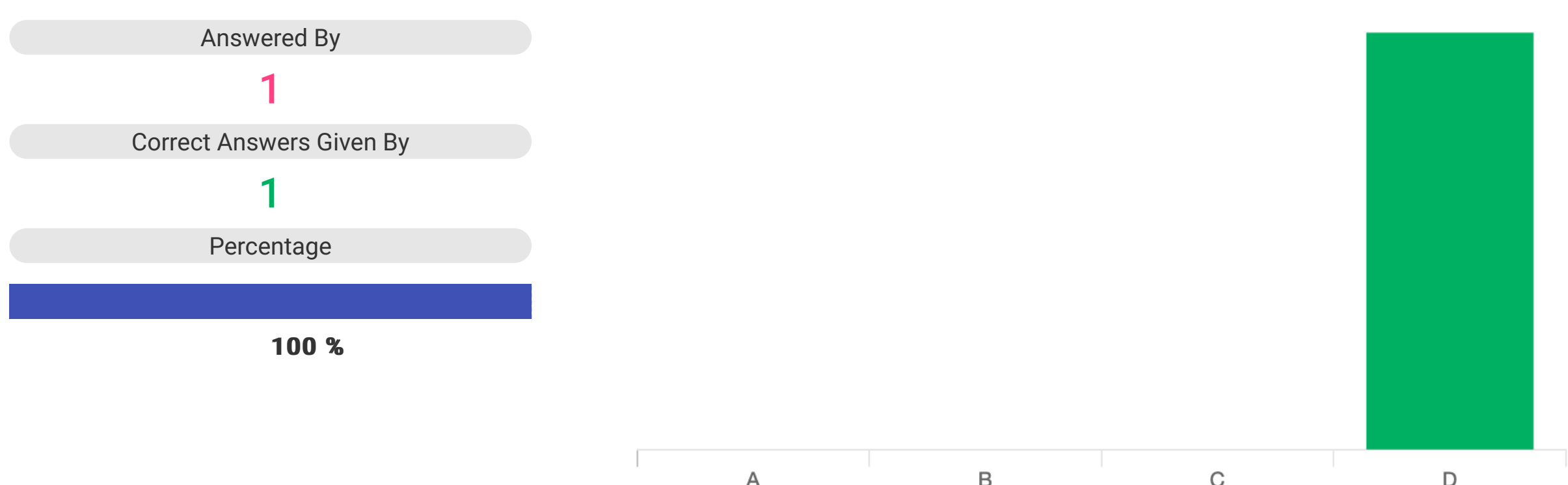


Question 4:

Your company is designing its data lake on Google Cloud and wants to develop different ingestion pipelines to collect unstructured data from different sources. After the data is stored in Google Cloud, it will be processed in several data pipelines to build a recommendation engine for end users on the website. The structure of the data retrieved from the source systems can change at any time. The data must be stored exactly as it was retrieved for reprocessing purposes in case the data structure is incompatible with the current processing pipelines. You need to design an architecture to support the use case after you retrieve the data. What should you do?

- A. Send the data through the processing pipeline, and then store the processed data in a BigQuery table for reprocessing.
- B. Store the data in a BigQuery table. Design the processing pipelines to retrieve the data from the table.
- C. Send the data through the processing pipeline, and then store the processed data in a Cloud Storage bucket for reprocessing.
- D. Store the data in a Cloud Storage bucket. Design the processing pipelines to retrieve the data from the bucket.

Type : SINGLE SELECTION

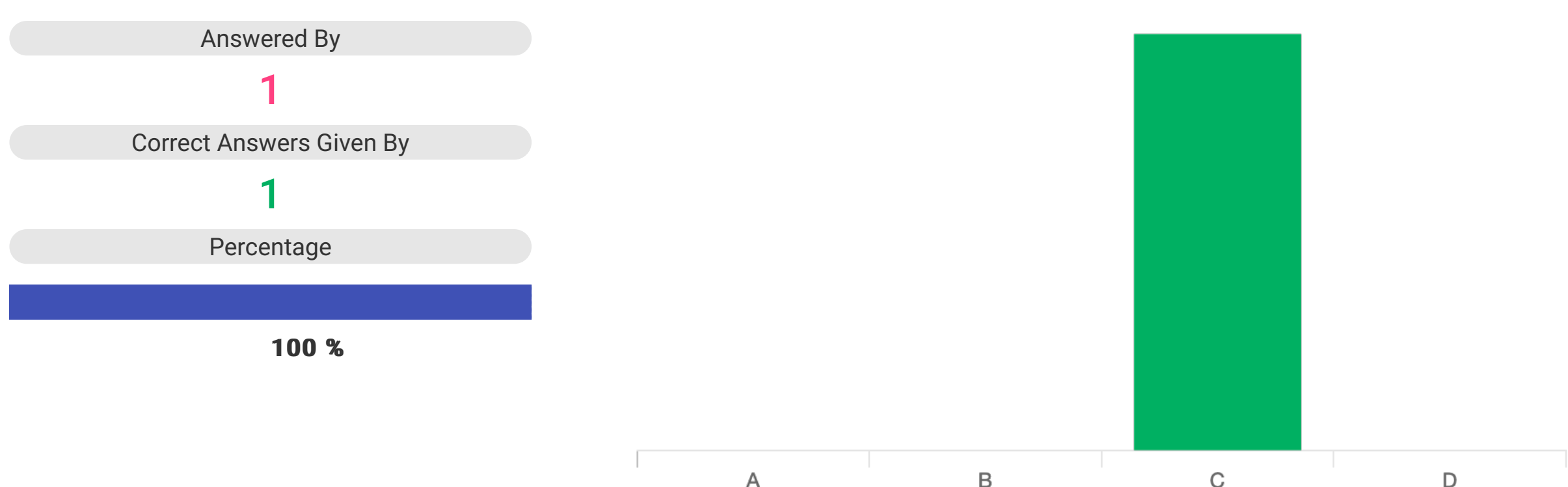


Question 5:

You've been asked to create a solution where managers can analyze sales trends by combining data from a relational system in an on-premises SQL Server database with website activity from Google Analytics and social media posts that mention your company. Which solution below seems most feasible and cost effective?

- A. Export the data to CSV files and put them in Cloud Storage. Use a Spark cluster in DataProc to analyze the data using SparkSQL.
- B. Write an App Engine program in Python to gather the data from each source and store it in Firestore and expose the data via a REST API. Then, create web-based reports using an Angular front end.
- C. Use Dataflow to transfer the data into BigQuery tables. Create reports using Google Data Studio.
- D. Write n application that copies the data to Pub/Sub and then use an AppEngine Application to process the messages and store them in Cloud Storage. Use a DataProc cluster to run the analysis and display the results using a Web site.

Type : SINGLE SELECTION

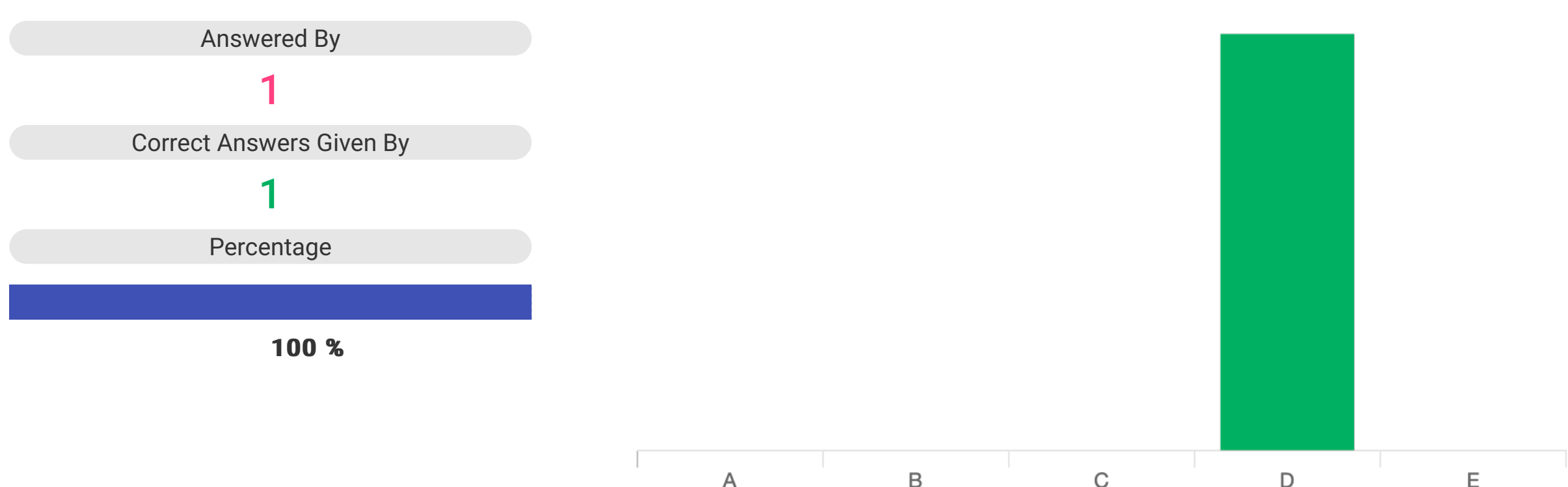


Question 6:

When creating a Dataproc cluster, an HDFS cluster is automatically created using the disks that are attached to the workers. This HDFS cluster can be used to store the data that is being analyzed. Why might you prefer to use Cloud Storage instead?

- A. Using Cloud Storage is cheaper.
- B. By separating compute and storage, the Dataproc cluster can be deleted when jobs are finished.
- C. With Cloud Storage, you pay for the amount of storage used. The HDFS cluster uses persistent disks, so you pay for what you allocate.
- D. All of the above
- E. None of the above. HDFS provides lower latency since the disks are attached to the workers. Therefore, for most workloads it would be preferred.

Type : SINGLE SELECTION

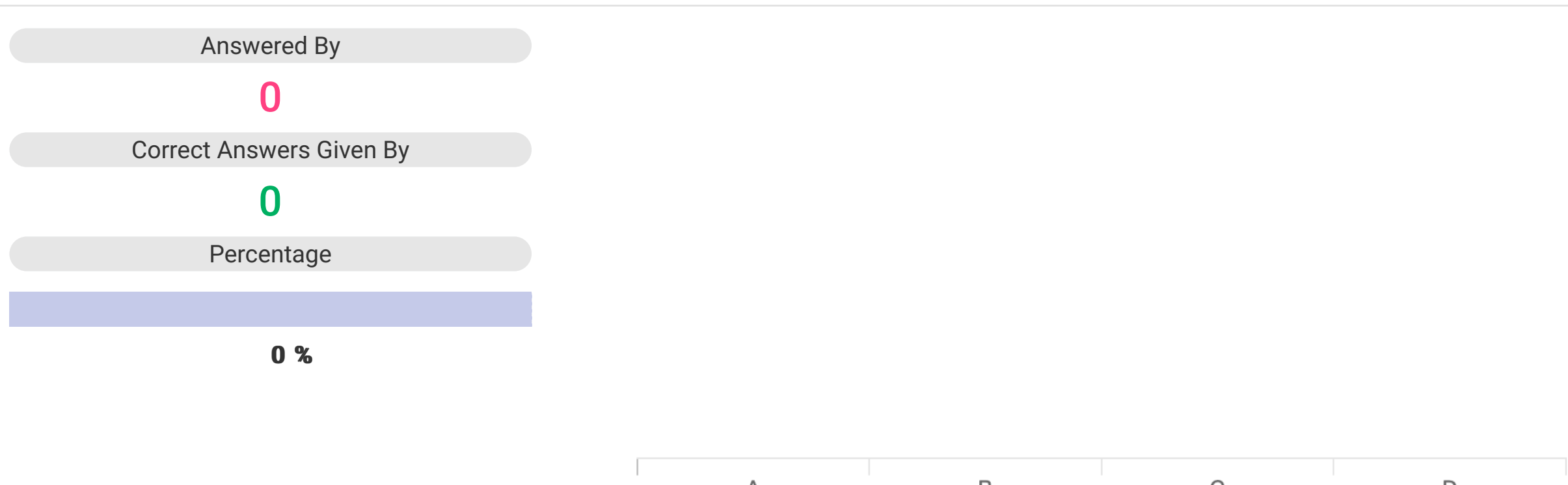


Question 7:

Why might you use some number of preemptible workers when configuring a Dataproc cluster?

- A. Preemptible machines have more throughput per CPU than normal machines.
- B. Preemptible machines utilize GPUs for faster processing of Big Data workloads.
- C. Preemptible machines are sold at a discount, but may be shut off if Google needs the resources.
- D. All of the above

Type : SINGLE SELECTION



Question 8:

You've been asked to help design a system that will run in Google Cloud. The company has stores in multiple countries, and they have an online store. They want to be able to analyze sales information close to real time. They want to collect every sale as they happen, store them, and then create web-based reports that allow managers to see the information. Of the choices below, which would you recommend?

- A. From the stores and the website, send every sale to Pub/Sub. Create a Dataflow pipeline to process the messages and then store them in BigQuery. Create reports using Google Data Studio.
- B. Use App Engine Standard. Create a REST API that receives all the sales from the stores and the website. Save each sale into DataStore. Create a Python website for the managers that provides the information.
- C. Create a CRON job in App Engine Standard that runs every few minutes. Read the new sales from the on-premises database and write the data into BigQuery. Create a web application using Angular that generates the reports using the BigQuery data. Deploy the Angular app from a Cloud Storage bucket.
- D. Build a web application that runs in both the stores and the website. Store that sales information in a MySQL database running in Cloud SQL. Create reports for the managers in this application. Deploy the application using Compute Engine and set up autoscaling and load balancing to handle peak periods.

Type : SINGLE SELECTION

