BDS-C00 Q&As

AWS Certified Big Data - Speciality (BDS-C00)

Pass Amazon BDS-C00 Exam with 100% Guarantee

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

https://www.leads4pass.com/aws-certified-big-data-specialty.html

100% Passing Guarantee
100% Money Back Assurance

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

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

QUESTION 1

A city has been collecting data on its public bicycle share program for the past three years. The 5PB dataset currently resides on Amazon S3. The data contains the following datapoints:

1.

Bicycle origination points

2.

Bicycle destination points

3.

Mileage between the points

4.

Number of bicycle slots available at the station (which is variable based on the station location)

5.

Number of slots available and taken at a given time

The program has received additional funds to increase the number of bicycle stations available. All data is regularly archived to Amazon Glacier.

The new bicycle stations must be located to provide the most riders access to bicycles.

How should this task be performed?

A. Move the data from Amazon S3 into Amazon EBS-backed volumes and use an EC-2 based Hadoop cluster with spot instances to run a Spark job that performs a stochastic gradient descent optimization.

B. Use the Amazon Redshift COPY command to move the data from Amazon S3 into Redshift and perform a SQL query that outputs the most popular bicycle stations.

C. Persist the data on Amazon S3 and use a transient EMR cluster with spot instances to run a Spark streaming job that will move the data into Amazon Kinesis.

D. Keep the data on Amazon S3 and use an Amazon EMR-based Hadoop cluster with spot instances to run a Spark job that performs a stochastic gradient descent optimization over EMRFS.

Correct Answer: B

QUESTION 2

Company A operates in Country X. Company A maintains a large dataset of historical purchase orders that contains personal data of their customers in the form of full names and telephone numbers. The dataset consists of 5 text files, 1TB each. Currently the dataset resides on-premises due to legal requirements of storing personal data in-country. The

research and development department needs to run a clustering algorithm on the dataset and wants to use Elastic Map Reduce service in the closest AWS region. Due to geographic distance, the minimum latency between the on-premises system and the closet AWS region is 200 ms.

Which option allows Company A to do clustering in the AWS Cloud and meet the legal requirement of maintaining personal data in-country?

A. Anonymize the personal data portions of the dataset and transfer the data files into Amazon S3 in the AWS region. Have the EMR cluster read the dataset using EMRFS.

B. Establish a Direct Connect link between the on-premises system and the AWS region to reduce latency. Have the EMR cluster read the data directly from the on-premises storage system over Direct Connect.

C. Encrypt the data files according to encryption standards of Country X and store them on AWS region in Amazon S3. Have the EMR cluster read the dataset using EMRFS.

D. Use AWS Import/Export Snowball device to securely transfer the data to the AWS region and copy the files onto an EBS volume. Have the EMR cluster read the dataset using EMRFS.

Correct Answer: B

QUESTION 3

Does Amazon RDS allow direct host access via Telnet, Secure Shell (SSH), or Windows Remote Desktop Connection?

A. Yes

B. No

C. Depends on if it is in VPC or not

Correct Answer: B

QUESTION 4

A clinical trial will rely on medical sensors to remotely assess patient health. Each physician who participates in the trial requires visual reports each morning. The reports are built from aggregations of all the sensor data taken each minute.

What is the most cost-effective solution for creating this visualization each day?

A. Use Kinesis Aggregators Library to generate reports for reviewing the patient sensor data and generate a QuickSight visualization on the new data each morning for the physician to review.

B. Use a transient EMR cluster that shuts down after use to aggregate the sensor data each night and generate a QuickSight visualization on the new data each morning for the physician to review.

C. Use Spark streaming on EMR to aggregate the patient sensor data in every 15 minutes and generate a QuickSight visualization on the new data each morning for the physician to review.

D. Use an EMR cluster to aggregate the patient sensor data each night and provide Zeppelin notebooks that look at the new data residing on the cluster each morning for the physician to review.

Correct Answer: D

QUESTION 5

An Amazon Redshift Database is encrypted using KMS. A data engineer needs to use the AWS CLI to create a KMS encrypted snapshot of the database in another AWS region.

Which three steps should the data engineer take to accomplish this task? (Choose three.)

A. Create a new KMS key in the destination region.

B. Copy the existing KMS key to the destination region.

C. Use CreateSnapshotCopyGrant to allow Amazon Redshift to use the KMS key from the source region.

D. In the source region, enable cross-region replication and specify the name of the copy grant created.

E. In the destination region, enable cross-region replication and specify the name of the copy grant created.

F. Use CreateSnapshotCopyGrant to allow Amazon Redshift to use the KMS key created in the destination region.

Correct Answer: ADF