Free Questions for BDS-C00

Shared by Walsh on 04-10-2024

For More Free Questions and Preparation Resources

Question 1

An administrator needs to design a strategy for the schema in a Redshift cluster. The administrator needs to determine the optimal distribution style for the tables on the Redshift schema.

In which two circumstances would choosing EVEN distribution be most appropriate? (Select two)

Options:

A- When the tables are highly denormalized and do NOT participate in frequent joins

B- When data must be grouped based on a specific key on a defined slice

C- When data transfer between nodes must be eliminated

D- When a new table has been loaded and it is unclear how it will be joined to dimension tables

Answer:

A, D

Explanation:

Question 2

Question Type: MultipleChoice

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

* Bicycle organization points

* Bicycle destination points

* Mileage between the points

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

* Number of slots available and taken at each station 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 station must be located to provide the most riders access to bicycles. How should this task be performed?

Options:

A- Move the data from Amazon S3 into Amazon EBS-backed volumes and EC2 Hardoop 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 platform a SQL query that outputs the most popular bicycle stations.

C- Persist the data on Amazon S3 and use a transits 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 insistences to run a spark job that perform a stochastic gradient descent optimization over EMBFS.

Answer:

D

Question 3

Managers in a company need access to the human resources database that runs on Amazon Redshift, to run reports about their employees. Managers must only see information about their direct reports.

Which technique should be used to address this requirement with Amazon Redshift?

Options:

A- Define an IAM group for each employee as an IAM user in that group and use that to limit the access.

B- Use Amazon Redshift snapshot to create one cluster per manager. Allow the managers to access only their designated clusters.

C- Define a key for each manager in AWS KMS and encrypt the data for their employees with their private keys.

D- Define a view that uses the employee's manager name to filter the records based on current user names.

Answer:

D

Explanation:

https://www.intermix.io/blog/iam-to-generate-temporary-amazon-redshift-passwords/

https://docs.aws.amazon.com/redshift/latest/mgmt/generating-iam-credentials-steps.html

Question 4

Question Type: **MultipleChoice**

Which data store should the organization choose?

Options:

A- Amazon Relational Database Service (RDS)

B- Amazon Redshift

C- Amazon DynamoDB

D- Amazon Elasticsearch

Answer:

C

Question 5

Question Type: **MultipleChoice**

An organization needs a data store to handle the following data types and access patterns

* Faceting

* Search

* Flexible schema (JSON) and fixed schema

* Noise word elimination

Which data store should the organization choose?

Options:

A- Amazon Relational Database Service (RDS)

B- Amazon Redshift

C- Amazon DynamoDB

D- Amazon Elasticsearch Service

Answer:

D

Question 6

Question Type: **MultipleChoice**

A web-hosting company is building a web analytics tools to capture clickstream data from all of the websites hosted within its platform and to provide near-real-time business intelligence. This entire system is built on AWS services. The web-hosting company is interested in using Amazon kinesis to collect this data and perform sliding window analytics. What is the most reliable and fault-tolerant technique to get each website to send data to Amazon Kinesis with every click?

Options:

A- After receiving a request each web server sends it to Amazon kinesis using the Amazon kinesis PutRecord APL Use the SessionID as a parturition key and set up a loop to retry until a success response is received

B- After receiving a request each web server sends it to Amazon kinesis using the Amazon Kinesis Producer Library addRecord method

C- Each web server bluffers the request until the count reaches 500 and sends them to Amazon kinesis using the Amazon kinesis PutRecord API call

D- After receiving a request each web server sends it to Amazon Kinesis using the Amazon kinesis PutRecord API. Use the exponential back off algorithm for retries until a successful response is received

Answer:

A

Explanation:

Question 7

Question Type: MultipleChoice

An Amazon EMR cluster using EMRFS has access to Megabytes of data on Amazon S3, originating from multiple unique data sources. The customer needs to query common fields across some of the data sets to be able to perform interactive joins and then display results quickly.

Which technology is most appropriate to enable this capability?

Options:

A- Presto

B- MicroStrategy

C- Pig

D- R Studio

Answer:

A

Explanation:

https://dzone.com/articles/getting-introduced-with-presto