Latest Version: 6.0

Question: 1

You have 30 customers, each of whom have a dedicated Kinesis Stream for streaming events. What action can you take so that the Kinesis charges are separated out on the Amazon invoice at the end of the month?
Response:

A. Move each customer into a separate AWS account and use consolidated billing
B. Enable CloudWatch to monitor the streams
C. Call Amazon to do that for you
D. Tag the streams with the name of the customer

Answer: D

Question: 2

A web application emits multiple types of events to Amazon Kinesis Streams for operational reporting. Critical events must be captured immediately before processing can continue, but informational events do not need to delay processing.
What is the most appropriate solution to record these different types of events?
Response:

A. Log all events using the Kinesis Producer Library.
B. Log critical events using the Kinesis Producer Library, and log informational events using the PutRecords API method.
C. Log critical events using the PutRecords API method, and log informational events using the Kinesis Producer Library.
D. Log all events using the PutRecords API method.

Answer: C

Question: 3

An organization uses Amazon Elastic MapReduce(EMR) to process a series of extract-transform-load (ETL) steps that run in sequence. The output of each step must be fully processed in subsequent steps but will not be retained. Which of the following techniques will meet this requirement most efficiently?
Response:

A. Use the s3n URI to store the data to be processed as objects in Amazon S3.

B. Use the EMR File System (EMRFS) to store the outputs from each step as objects in Amazon Simple Storage Service (S3).
C. Load the data to be processed into HDFS, and then write the final output to Amazon S3.
D. Define the ETL steps as separate AWS Data Pipeline activities.

Answer: C

Question: 4

A company is using Amazon Machine Learning as part of a medical software application. The application will predict the most likely blood type for a patient based on a variety of other clinical tests that are available when blood type knowledge is unavailable. What is the appropriate model choice and target attribute combination for this problem?
Response:

A. K-Nearest Neighbors model with a multi-class target attribute.
B. Regression model with a numeric target attribute.
C. Binary Classification with a categorical target attribute.
D. Multi-class classification model with a categorical target attribute.

Answer: D

Question: 5

How are Snowball logs stored?
Response:

A. in a JSON file
B. in a SQLite table
C. in a plaintext file
D. in an XML file

Answer: C

Question: 6

An administrator decides to use the Amazon Machine Learning service to classify social media posts that mention your company into two categories: posts that require a response and posts that do not.
The training dataset of 10,000 posts contains the details of each post, including the timestamp, author, and full text of the post. You are missing the target labels that are required for training.
Which two options will create valid target label data?
Response:

A. Ask the social media handling team to review each post and provide the label.
B. Use the sentiment analysis NLP library to determine whether a post requires a response.
C. Use the Amazon Mechanical Turk web service to publish Human Intelligence Tasks that ask Turk workers to label the posts.
D. Using the a priori probability distribution of the two classes, use Monte-Carlo simulation to generate the labels.

Answer: AC

Question: 7

What are three qualities of a Machine Learning program?
(Select three.)
Response:

A. Machine Learning programs improve accuracy with unsupervised learning.
B. Machine Learning programs are created with data rather than rules.
C. Machine Learning programs improve accuracy with more data.
D. Machine Learning programs are able to alter themselves.

Answer: BCD

Question: 8

An organization needs a data store to handle the following data types and access patterns:
- Key-value access pattern
- Complex SQL queries and transactions
- Consistent reads
- Fixed schema

Which data store should the organization choose?
Response:

A. Amazon S3
B. Amazon Kinesis
C. Amazon DynamoDB
D. Amazon RDS

Answer: D

Question: 9

A customer needs to load a 550-GB data file into an Amazon Redshift cluster from Amazon S3, using the COPY command.
The input file has both known and unknown issues that will probably cause the load process to fail. The customer needs the most efficient way to detect load errors without performing any cleanup if the load process fails.
Which technique should the customer use?
Response:

A. Split the input file into 50-GB blocks and load them separately.
B. Use COPY with NOLOAD parameter.
C. Write a script to delete the data from the tables in case of errors.
D. Compress the input file before running COPY.

Answer: B

Question: 10

Data already in S3 can be encrypted with?
Select two.
Response:

A. CSE-KMS
B. CSE-C
C. SSE-S3
D. SSE KMS

Answer: CD