Free Questions for BDS-C00 by go4braindumps

Shared by Pope on 09-08-2024

For More Free Questions and Preparation Resources

Check the Links on Last Page

Question 1

Question Type: MultipleChoice

An organization is developing a mobile social application and needs to collect logs from all devices on which it is installed. The organization is evaluating the Amazon Kinesis Data Streams to push logs and Amazon EMR to process data. They want to store data on HDFS using the default replication factor to replicate data among the cluster, but they are concerned about the durability of the data. Currently, they are producing 300 GB of raw data daily, with additional spikes during special events. They will need to scale out the Amazon EMR cluster to match the increase in streamed data.

Which solution prevents data loss and matches compute demand?

Options:

A- Use multiple Amazon EBS volumes on Amazon EMR to store processed data and scale out the Amazon EMR cluster as needed.

B- Use the EMR File System and Amazon S3 to store processed data and scale out the Amazon EMR cluster as needed.

C- Use Amazon DynamoDB to store processed data and scale out the Amazon EMR cluster as needed.

D- use Amazon Kinesis Data Firehose and, instead of using Amazon EMR, stream logs directly into Amazon Elasticsearch Service.

Answer:

B

Question 2

Question Type: MultipleChoice

An organization currently runs a large Hadoop environment in their data center and is in the process of creating an alternative Hadoop environment on AWS, using Amazon EMR.

They generate around 20 TB of data on a monthly basis. Also on a monthly basis, files need to be grouped and copied to Amazon S3 to be used for the Amazon EMR environment. They have multiple S3 buckets across AWS accounts to which data needs to be copied. There is a 10G AWS Direct Connect setup between their data center and AWS, and the network team has agreed to allocate

Options:

A- Use an offline copy method, such as an AWS Snowball device, to copy and transfer data to Amazon S3.

B- Configure a multipart upload for Amazon S3 on AWS Java SDK to transfer data over AWS Direct Connect.

C- Use Amazon S3 transfer acceleration capability to transfer data to Amazon S3 over AWS Direct Connect.

D- Setup S3DistCop tool on the on-premises Hadoop environment to transfer data to Amazon S3 over AWS Direct Connect.

Answer:

D

Question 3

Question Type: MultipleChoice

An organization has added a clickstream to their website to analyze traffic. The website is sending each page request with the PutRecord API call to an Amazon Kinesis stream by using the page name as the partition key. During peak spikes in website traffic, a support engineer notices many ProvisionedThroughputExcededException events in the application logs.

What should be done to resolve the issue in the MOST cost-effective way?

Options:

A- Create multiple Amazon Kinesis streams for page requests to increase the concurrency of the clickstream.

B- Increase the number of shards on the Kinesis stream to allow for more throughput to meet the peak spikes in traffic.

C- Modify the application to use on the Kinesis Producer Library to aggregate requests before sending them to the Kinesis stream.

D- Attach more consumers to the Kinesis stream to process records in parallel, improving the performance on the stream.

Answer:

B

Explanation:

https://aws.amazon.com/kinesis/data-streams/faqs/

Question 4

Question Type: MultipleChoice

An organization would like to run analytics on their Elastic Load Balancing logs stored in Amazon S3 and join this data with other tables in Amazon S3. The users are currently using a BI tool connecting with JDBC and would like to keep using this BI tool.

Which solution would result in the LEAST operational overhead?

Options:

A- Trigger a Lambda function when a new log file is added to the bucket to transform and load it into Amazon Redshift. Run the VACUUM command on the Amazon Redshift cluster every night.

B- Launch a long-running Amazon EMR cluster that continuously downloads and transforms new files from Amazon S3 into its HDFS storage. Use Presto to expose the data through JDBC.

C- Trigger a Lambda function when a new log file is added to the bucket to transform and move it to another bucket with an optimized data structure. Use Amazon Athena to query the optimized bucket.

D- Launch a transient Amazon EMR cluster every night that transforms new log files and loads them into Amazon Redshift.

Answer:

C

Question 5

Question Type: MultipleChoice

An Operations team continuously monitors the number of visitors to a website to identify any potential system problems. The number of website visitors varies throughout the day. The site is more popular in the middle of the day and less popular at night.

Which type of dashboard display would be the MOST useful to allow staff to quickly and correctly identify system problems?

Options:

A- A vertical stacked bar chart showing today's website visitors and the historical average number of website visitors.

B- An overlay line chart showing today's website visitors at one-minute intervals and also the historical average number of website visitors.

C- A single KPI metric showing the statistical variance between the current number of website visitors and the historical number of website visitors for the current time of day.

D- A scatter plot showing today's website visitors on the X-axis and the historical average number of website visitors on the Y-axis.

Answer:

C

Question 6

Question Type: MultipleChoice

An organization is using Amazon Kinesis Data Streams to collect data generated from thousands of temperature devices and is using AWS Lambda to process the data. Devices generate 10 to 12 million records every day, but Lambda is processing only around 450 thousand records. Amazon CloudWatch indicates that throttling on Lambda is not occurring.

What should be done to ensure that all data is processed? (Choose two.)

Options:

A- Increase the BatchSize value on the EventSource, and increase the memory allocated to the Lambda function.

B- Decrease the BatchSize value on the EventSource, and increase the memory allocated to the Lambda function.

C- Create multiple Lambda functions that will consume the same Amazon Kinesis stream.

D- Increase the number of vCores allocated for the Lambda function.

E- Increase the number of shards on the Amazon Kinesis stream.

Answer:

A, E

Explanation:

https://tech.trivago.com/2018/07/13/aws-kinesis-with-lambdas-lessons-learned/

Question 7

Question Type: MultipleChoice

An organization is currently using an Amazon EMR long-running cluster with the latest Amazon EMR release for analytic jobs and is storing data as external tables on Amazon S3.

The company needs to launch multiple transient EMR clusters to access the same tables concurrently, but the metadata about the Amazon S3 external tables are defined and stored on the long-running cluster.

Which solution will expose the Hive metastore with the LEAST operational effort?

Options:

A- Export Hive metastore information to Amazon DynamoDB hive-site classification to point to the Amazon DynamoDB table.

B- Export Hive metastore information to a MySQL table on Amazon RDS and configure the Amazon EMR hive-site classification to point to the Amazon RDS database.

C- Launch an Amazon EC2 instance, install and configure Apache Derby, and export the Hive metastore information to derby.

D- Create and configure an AWS Glue Data Catalog as a Hive metastore for Amazon EMR.

Answer:

B

Question 8

Question Type: MultipleChoice

An organization is setting up a data catalog and metadata management environment for their numerous data stores currently running on AWS. The data catalog will be used to determine the structure and other attributes of data in the data stores. The data stores are composed of Amazon RDS databases, Amazon Redshift, and CSV files residing on Amazon S3. The catalog should be populated on a scheduled basis, and minimal administration is required to manage the catalog.

How can this be accomplished?

Options:

A- Set up Amazon DynamoDB as the data catalog and run a scheduled AWS Lambda function that connects to data sources to populate the database.

B- Use an Amazon database as the data catalog and run a scheduled AWS Lambda function that connects to data sources to populate the database.

C- Use AWS Glue Data Catalog as the data catalog and schedule crawlers that connect to data sources to populate the database.

D- Set up Apache Hive metastore on an Amazon EC2 instance and run a scheduled bash script that connects to data sources to populate the metastore.

Answer:

C

Explanation:

https://docs.aws.amazon.com/glue/latest/dg/populate-data-catalog.html

Question 9

Question Type: MultipleChoice

An organization is soliciting public feedback through a web portal that has been deployed to track the number of requests and other important data. As part of reporting and visualization, AmazonQuickSight connects to an Amazon RDS database to virtualize data. Management wants to understand some important metrics about feedback and how the feedback has changed over the last four weeks in a visual representation.

What would be the MOST effective way to represent multiple iterations of an analysis in Amazon QuickSight that would show how the data has changed over the last four weeks?

Options:

A- Use the analysis option for data captured in each week and view the data by a date range.

B- Use a pivot table as a visual option to display measured values and weekly aggregate data as a row dimension.

C- Use a dashboard option to create an analysis of the data for each week and apply filters to visualize the data change.

D- Use a story option to preserve multiple iterations of an analysis and play the iterations sequentially.

Answer:

D

Explanation:

https://docs.aws.amazon.com/quicksight/latest/user/working-with-stories.html

Question 10

Question Type: MultipleChoice

How should an Administrator BEST architect a large multi-layer Long Short-Term Memory (LSTM) recurrent neural network (RNN) running with MXNET on Amazon EC2? (Choose two.)

Options:

A- Use data parallelism to partition the workload over multiple devices and balance the workload within the GPUs.

B- Use compute-optimized EC2 instances with an attached elastic GPU.

C- Use general purpose GPU computing instances such as G3 and P3.

D- Use processing parallelism to partition the workload over multiple storage devices and balance the workload within the GPUs.

Answer:

A, C

Explanation:

https://aws.amazon.com/blogs/machine-learning/parallelizing-across-multiple-cpu-gpus-to-speed-up-deep-learning-inference-at-the-edge/

Question 11

Question Type: MultipleChoice

An organization needs to design and deploy a large-scale data storage solution that will be highly durable and highly flexible with respect to the type and structure of data being stored. The data to be stored will be sent or generated from a variety of sources and must be persistently available for access and processing by multiple applications.

What is the most cost-effective technique to meet these requirements?

Options:

A- Use Amazon Simple Storage Service (S3) as the actual data storage system, coupled with appropriate tools for ingestion/acquisition of data and for subsequent processing and querying.

B- Deploy a long-running Amazon Elastic MapReduce (EMR) cluster with Amazon Elastic Block Store (EBS) volumes for persistent HDFS storage and appropriate Hadoop ecosystem tools for processing and querying.

C- Use Amazon Redshift with data replication to Amazon Simple Storage Service (S3) for comprehensive durable data storage, processing and querying.

D- Launch an Amazon Relational Database Service (RDS), and use the enterprise grade and capacity of the Amazon Aurora Engine for storage processing and querying.

Answer:

A