Question 1

A manufacturing company has many IoT devices in different facilities across the world The company is using Amazon Kinesis Data Streams to collect the data from the devices

The company's operations team has started to observe many WnteThroughputExceeded exceptions The operations team determines that the reason is the number of records that are being written to certain shards The data contains device ID capture date measurement type, measurement value and facility ID The facility ID is used as the partition key

Which action will resolve this issue?

Options:

A- Change the partition key from facility ID to a randomly generated key

B- Increase the number of shards

C- Archive the data on the producers' side

D- Change the partition key from facility ID to capture date

Answer:

B

Question 2

An analytics software as a service (SaaS) provider wants to offer its customers business intelligence
The provider wants to give customers two user role options

* Read-only users for individuals who only need to view dashboards

* Power users for individuals who are allowed to create and share new dashboards with other users

Which QuickSight feature allows the provider to meet these requirements'?

Options:

A- Embedded dashboards

B- Table calculations

C- Isolated namespaces

D- SPICE

Answer:

A

Question 3

A bank is using Amazon Managed Streaming for Apache Kafka (Amazon MSK) to populate real-time data into a data lake The data lake is built on Amazon S3, and data must be accessible from the data lake within 24 hours Different microservices produce messages to different topics in the cluster The cluster is created with 8 TB of Amazon Elastic Block Store (Amazon EBS) storage and a retention period of 7 days

The customer transaction volume has tripled recently and disk monitoring has provided an alert that the cluster is almost out of storage capacity

What should a data analytics specialist do to prevent the cluster from running out of disk space1?

Options:

A- Use the Amazon MSK console to triple the broker storage and restart the cluster

B- Create an Amazon CloudWatch alarm that monitors the KafkaDataLogsDiskUsed metric Automatically flush the oldest messages when the value of this metric exceeds 85%

C- Create a custom Amazon MSK configuration Set the log retention hours parameter to 48 Update the cluster with the new configuration file

D- Triple the number of consumers to ensure that data is consumed as soon as it is added to a topic.

Answer:

B

Question 4

Question Type: **MultipleChoice**

A hospital is building a research data lake to ingest data from electronic health records (EHR) systems from multiple hospitals and clinics. The EHR systems are independent of each other and do not have a common patient identifier. The data engineering team is not experienced in machine learning (ML) and has been asked to generate a unique patient identifier for the ingested records.

Which solution will accomplish this task?

Options:

A- An AWS Glue ETL job with the FindMatches transform

B- Amazon Kendra

C- Amazon SageMaker Ground Truth

D- An AWS Glue ETL job with the ResolveChoice transform

Answer:

A

Explanation:

Matching Records with AWS Lake Formation FindMatches

Question 5

An online retail company uses Amazon Redshift to store historical sales transactions. The company is required to encrypt data at rest in the clusters to comply with the Payment Card Industry Data Security Standard (PCI DSS). A corporate governance policy mandates management of encryption keys using an on-premises hardware security module (HSM).

Which solution meets these requirements?

Options:

A- Create and manage encryption keys using AWS CloudHSM Classic. Launch an Amazon Redshift cluster in a VPC with the option to use CloudHSM Classic for key management.

B- Create a VPC and establish a VPN connection between the VPC and the on-premises network. Create an HSM connection and client certificate for the on-premises HSM. Launch a cluster in the VPC with the option to use the on-premises HSM to store keys.

C- Create an HSM connection and client certificate for the on-premises HSM. Enable HSM encryption on the existing unencrypted cluster by modifying the cluster. Connect to the VPC where the Amazon Redshift cluster resides from the on-premises network using a VPN.

D- Create a replica of the on-premises HSM in AWS CloudHSM. Launch a cluster in a VPC with the option to use CloudHSM to store keys.

Answer:

B

Question 6

Question Type: **MultipleChoice**

A company with a video streaming website wants to analyze user behavior to make recommendations to users in real time Clickstream data is being sent to Amazon Kinesis Data Streams and reference data is stored in Amazon S3 The company wants a solution that can use standard SQL quenes The solution must also provide a way to look up pre-calculated reference data while making recommendations

Which solution meets these requirements?

Options:

A- Use an AWS Glue Python shell job to process incoming data from Kinesis Data Streams Use the Boto3 library to write data to Amazon Redshift

B- Use AWS Glue streaming and Scale to process incoming data from Kinesis Data Streams Use the AWS Glue connector to write data to Amazon Redshift

C- Use Amazon Kinesis Data Analytics to create an in-application table based upon the reference data Process incoming data from Kinesis Data Streams Use a data stream to write results to Amazon Redshift

D- Use Amazon Kinesis Data Analytics to create an in-application table based upon the reference data Process incoming data from Kinesis Data Streams Use an Amazon Kinesis Data Firehose delivery stream to write results to Amazon Redshift

Answer:

D

Question 7

A company uses Amazon kinesis Data Streams to ingest and process customer behavior information from application users each day. A data analytics specialist notices that its data stream is throttling. The specialist has turned on enhanced monitoring for the Kinesis data stream and has verified that the data stream did not exceed the data limits. The specialist discovers that there are hot shards

Which solution will resolve this issue?

Options:
A- Use a random partition key to ingest the records.

B- Increase the number of shards Split the size of the log records.

C- Limit the number of records that are sent each second by the producer to match the capacity of the stream.

D- Decrease the size of the records that are sent from the producer to match the capacity of the stream.

Answer:
A

Question 8

A company recently created a test AWS account to use for a development environment The company also created a production AWS account in another AWS Region As part of its security testing the company wants to send log data from Amazon CloudWatch Logs in its production account to an Amazon Kinesis data stream in its test account

Which solution will allow the company to accomplish this goal?

Options:

A- Create a subscription filter in the production accounts CloudWatch Logs to target the Kinesis data stream in the test account as its destination In the test account create an 1AM role that grants access to the Kinesis data stream and the CloudWatch Logs resources in the production account

B- In the test account create an 1AM role that grants access to the Kinesis data stream and the CloudWatch Logs resources in the production account Create a destination data stream in Kinesis Data Streams in the test account with an 1AM role and a trust policy that allow CloudWatch Logs in the production account to write to the test account

C- In the test account, create an 1AM role that grants access to the Kinesis data stream and the CloudWatch Logs resources in the production account Create a destination data stream in Kinesis Data Streams in the test account with an 1AM role and a trust policy that allow CloudWatch Logs in the production account to write to the test account

D- Create a destination data stream in Kinesis Data Streams in the test account with an 1AM role and a trust policy that allow CloudWatch Logs in the production account to write to the test account Create a subscription filter in the production accounts CloudWatch Logs to target the Kinesis data stream in the test account as its destination

Answer:

D

Question 9

A bank wants to migrate a Teradata data warehouse to the AWS Cloud The bank needs a solution for reading large amounts of data and requires the highest possible performance. The solution also must maintain the separation of storage and compute

Which solution meets these requirements?

Options:

A- Use Amazon Athena to query the data in Amazon S3

B- Use Amazon Redshift with dense compute nodes to query the data in Amazon Redshift managed storage

C- Use Amazon Redshift with RA3 nodes to query the data in Amazon Redshift managed storage

D- Use PrestoDB on Amazon EMR to query the data in Amazon S3

Answer:

C

Question 10

A company is reading data from various customer databases that run on Amazon RDS. The databases contain many inconsistent fields For example, a customer record field that is place_id in one database is location_id in another database. The company wants to link customer records across different databases, even when many customer record fields do not match exactly

Which solution will meet these requirements with the LEAST operational overhead?

Options:

A- Create an Amazon EMR cluster to process and analyze data in the databases Connect to the Apache Zeppelin notebook, and use the FindMatches transform to find duplicate records in the data.

B- Create an AWS Glue crawler to crawl the databases. Use the FindMatches transform to find duplicate records in the data Evaluate and tune the transform by evaluating performance and results of finding matches

C- Create an AWS Glue crawler to crawl the data in the databases Use Amazon SageMaker to construct Apache Spark ML pipelines to find duplicate records in the data

D- Create an Amazon EMR cluster to process and analyze data in the databases. Connect to the Apache Zeppelin notebook, and use Apache Spark ML to find duplicate records in the data. Evaluate and tune the model by evaluating performance and results of finding duplicates

Answer:

B

Question 11

Question Type: **MultipleChoice**

A company is building an analytical solution that includes Amazon S3 as data lake storage and Amazon Redshift for data warehousing. The company wants to use Amazon Redshift Spectrum to query the data that is stored in Amazon S3.

Which steps should the company take to improve performance when the company uses Amazon Redshift Spectrum to query the S3 data files? (Select THREE)

Use gzip compression with individual file sizes of 1-5 GB

Options:

A- Use a columnar storage file format

B- Partition the data based on the most common query predicates

C- Split the data into KB-sized files.

D- Keep all files about the same size.

E- Use file formats that are not splittable

Answer:

B, C, D

Question 12

Question Type: **MultipleChoice**

An advertising company has a data lake that is built on Amazon S3. The company uses AWS Glue Data Catalog to maintain the metadat

a. The data lake is several years old and its overall size has increased exponentially as additional data sources and metadata are stored in the data lake. The data lake administrator wants to implement a mechanism to simplify permissions management between Amazon S3 and the Data Catalog to keep them in sync

Which solution will simplify permissions management with minimal development effort?

Options:

A- Set AWS Identity and Access Management (1AM) permissions tor AWS Glue

B- Use AWS Lake Formation permissions

C- Manage AWS Glue and S3 permissions by using bucket policies

D- Use Amazon Cognito user pools.

Answer:

B

To Get Premium Files for DAS-C01 Visit

https://www.p2pexams.com/products/das-c01

For More Free Questions Visit

https://www.p2pexams.com/amazon/pdf/das-c01

20% DISCOUNT