DUMPSARENA

AWS Certified Data Analytics - Specialty

Amazon AWS DAS-C01

Version Demo

Total Demo Questions: 10

Total Premium Questions: 164

Buy Premium PDF

https://dumpsarena.co

sales@dumpsarena.co

sales@dumpsarena.co
dumpsarena.co

QUESTION NO: 1

A data analyst runs a large number of data manipulation language (DML) queries by using Amazon Athena with the JDBC driver. Recently, a query failed after it ran for 30 minutes. The query returned the following message:

java.sql.SQLException: Query timeout

The data analyst does not immediately need the query results. However, the data analyst needs a long-term solution for this problem.

Which solution will meet these requirements?

A. Split the query into smaller queries to search smaller subsets of data

B. In the settings for Athena, adjust the DML query timeout limit

C. In the Service Quotas console, request an increase for the DML query timeout

D. Save the tables as compressed .csv files

ANSWER: C

Explanation:

Reference: https://docs.aws.amazon.com/athena/latest/ug/service-limits.html

QUESTION NO: 2

A marketing company is using Amazon EMR clusters for its workloads. The company manually installs third-party libraries on the clusters by logging in to the master nodes. A data analyst needs to create an automated solution to replace the manual process.

Which options can fulfill these requirements? (Choose two.)

A. Place the required installation scripts in Amazon S3 and execute them using custom bootstrap actions.

B. Place the required installation scripts in Amazon S3 and execute them through Apache Spark in Amazon EMR.

C. Install the required third-party libraries in the existing EMR master node. Create an AMI out of that master node and use that custom AMI to re-create the EMR cluster.

D. Use an Amazon DynamoDB table to store the list of required applications. Trigger an AWS Lambda function with DynamoDB Streams to install the software.

E. Launch an Amazon EC2 instance with Amazon Linux and install the required third-party libraries on the instance. Create an AMI and use that AMI to create the EMR cluster.

ANSWER: A C

QUESTION NO: 3

A retail company's data analytics team recently created multiple product sales analysis dashboards for the average selling price per product using Amazon QuickSight. The dashboards were created from .csv files uploaded to Amazon S3. The team is now planning to share the dashboards with the respective external product owners by creating individual users in Amazon QuickSight. For compliance and governance reasons, restricting access is a key requirement. The product owners should view only their respective product analysis in the dashboard reports.

Which approach should the data analytics team take to allow product owners to view only their products in the dashboard?

A. Separate the data by product and use S3 bucket policies for authorization.

B. Separate the data by product and use IAM policies for authorization.

C. Create a manifest file with row-level security.

D. Create dataset rules with row-level security.

ANSWER: B

QUESTION NO: 4

An airline has been collecting metrics on flight activities for analytics. A recently completed proof of concept demonstrates how the company provides insights to data analysts to improve on-time departures. The proof of concept used objects in Amazon S3, which contained the metrics in .csv format, and used Amazon Athena for querying the data. As the amount of data increases, the data analyst wants to optimize the storage solution to improve query performance.

Which options should the data analyst use to improve performance as the data lake grows? (Choose three.)

A. Add a randomized string to the beginning of the keys in S3 to get more throughput across partitions.

B. Use an S3 bucket in the same account as Athena.

C. Compress the objects to reduce the data transfer I/O.

D. Use an S3 bucket in the same Region as Athena.

E. Preprocess the .csv data to JSON to reduce I/O by fetching only the document keys needed by the query.

F. Preprocess the .csv data to Apache Parquet to reduce I/O by fetching only the data blocks needed for predicates.

ANSWER: A C E

QUESTION NO: 5

A manufacturing company wants to create an operational analytics dashboard to visualize metrics from equipment in near-real time. The company uses Amazon Kinesis Data Streams to stream the data to other applications. The dashboard must automatically refresh every 5 seconds. A data analytics specialist must design a solution that requires the least possible implementation effort.

Which solution meets these requirements?

A. Use Amazon Kinesis Data Firehose to store the data in Amazon S3. Use Amazon QuickSight to build the dashboard.

B. Use Apache Spark Streaming on Amazon EMR to read the data in near-real time. Develop a custom application for the dashboard by using D3.js.

C. Use Amazon Kinesis Data Firehose to push the data into an Amazon OpenSearch Service (Amazon Elasticsearch Service) cluster. Visualize the data by using an OpenSearch Dashboards (Kibana).

D. Use AWS Glue streaming ETL to store the data in Amazon S3. Use Amazon QuickSight to build the dashboard.

ANSWER: B

Explanation:

Reference: https://aws.amazon.com/blogs/big-data/analyze-a-time-series-in-real-time-with-aws-lambda-amazon-kinesis-and-amazon-dynamodb-streams/

QUESTION NO: 6

A company is providing analytics services to its sales and marketing departments. The departments can access the data only through their business intelligence (BI) tools, which run queries on Amazon Redshift using an Amazon Redshift internal user to connect. Each department is assigned a user in the Amazon Redshift database with the permissions needed for that department. The marketing data analysts must be granted direct access to the advertising table, which is stored in Apache Parquet format in the marketing S3 bucket of the company data lake. The company data lake is managed by AWS Lake Formation. Finally, access must be limited to the three promotion columns in the table. Which combination of steps will meet these requirements? (Choose three.)

A. Grant permissions in Amazon Redshift to allow the marketing Amazon Redshift user to access the three promotion columns of the advertising external table.

B. Create an Amazon Redshift Spectrum IAM role with permissions for Lake Formation. Attach it to the Amazon Redshift cluster.

C. Create an Amazon Redshift Spectrum IAM role with permissions for the marketing S3 bucket. Attach it to the Amazon Redshift cluster.

D. Create an external schema in Amazon Redshift by using the Amazon Redshift Spectrum IAM role. Grant usage to the marketing Amazon Redshift user.

E. Grant permissions in Lake Formation to allow the Amazon Redshift Spectrum role to access the three promotion columns of the advertising table.

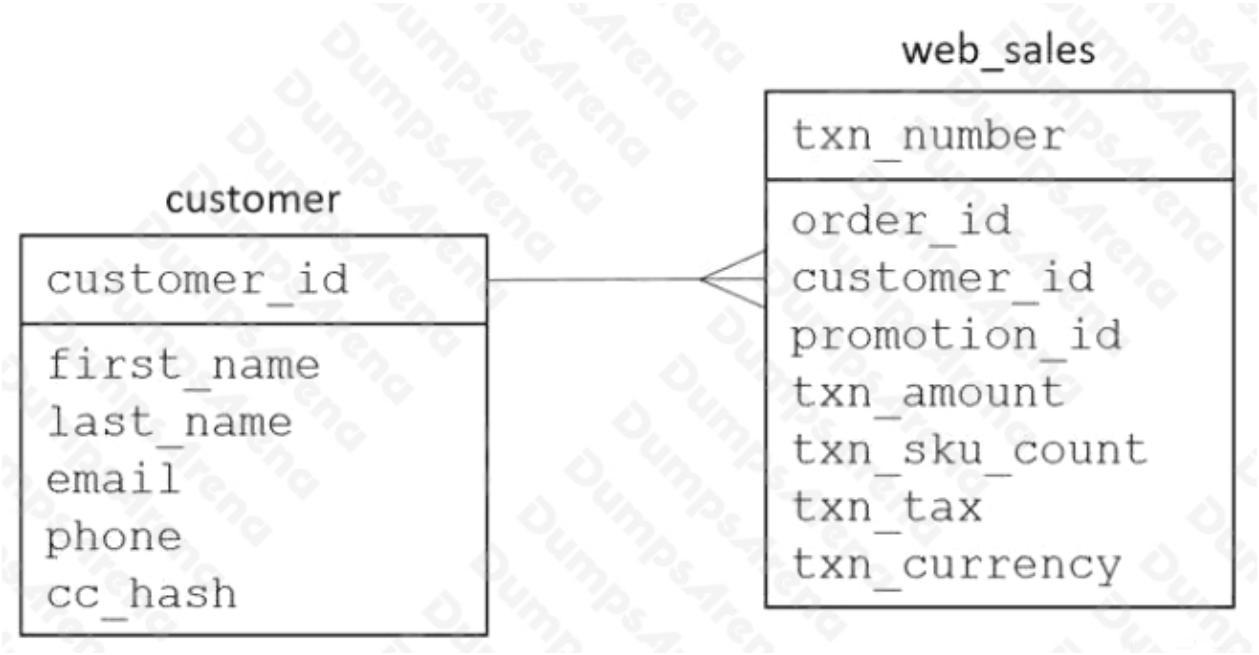
F. Grant permissions in Lake Formation to allow the marketing IAM group to access the three promotion columns of the advertising table.

ANSWER: B D E

QUESTION NO: 7

A retail company is using an Amazon S3 bucket to host an ecommerce data lake. The company is using AWS Lake Formation to manage the data lake.

A data analytics specialist must provide access to a new business analyst team. The team will use Amazon Athena from the AWS Management Console to query data from existing web_sales and customer tables in the ecommerce database. The team needs read-only access and the ability to uniquely identify customers by using first and last names. However, the team must not be able to see any other personally identifiable data. The table structure is as follows:



Which combination of steps should the data analytics specialist take to provide the required permission by using the principle of least privilege? (Choose three.)

A. In AWS Lake Formation, grant the business_analyst group SELECT and ALTER permissions for the web_sales table.

B. In AWS Lake Formation, grant the business_analyst group the SELECT permission for the web_sales table.

C. In AWS Lake Formation, grant the business_analyst group the SELECT permission for the customer table. Under columns, choose filter type “Include columns” with columns fisrt_name, last_name, and customer_id.

D. In AWS Lake Formation, grant the business_analyst group SELECT and ALTER permissions for the customer table. Under columns, choose filter type “Include columns” with columns fisrt_name and last_name.

E. Create users under a business_analyst IAM group. Create a policy that allows the lakeformation:GetDataAccess action, the athena:* action, and the glue:Get* action.

F. Create users under a business_analyst IAM group. Create a policy that allows the lakeformation:GetDataAccess action, the athena:* action, and the glue:Get* action. In addition, allow the s3:GetObject action, the s3:PutObject action, and the s3:GetBucketLocation action for the Athena query results S3 bucket.

ANSWER: B D F

QUESTION NO: 8

A company has developed several AWS Glue jobs to validate and transform its data from Amazon S3 and load it into Amazon RDS for MySQL in batches once every day. The ETL jobs read the S3 data using a DynamicFrame. Currently, the ETL developers are experiencing challenges in processing only the incremental data on every run, as the AWS Glue job processes all the S3 input data on each run.

Which approach would allow the developers to solve the issue with minimal coding effort?

A. Have the ETL jobs read the data from Amazon S3 using a DataFrame.

B. Enable job bookmarks on the AWS Glue jobs.

C. Create custom logic on the ETL jobs to track the processed S3 objects.

D. Have the ETL jobs delete the processed objects or data from Amazon S3 after each run.

ANSWER: D

QUESTION NO: 9

A company launched a service that produces millions of messages every day and uses Amazon Kinesis Data Streams as the streaming service.

The company uses the Kinesis SDK to write data to Kinesis Data Streams. A few months after launch, a data analyst found that write performance is significantly reduced. The data analyst investigated the metrics and determined that Kinesis is throttling the write requests. The data analyst wants to address this issue without significant changes to the architecture. Which actions should the data analyst take to resolve this issue? (Choose two.)

A. Increase the Kinesis Data Streams retention period to reduce throttling.

B. Replace the Kinesis API-based data ingestion mechanism with Kinesis Agent.

C. Increase the number of shards in the stream using the UpdateShardCount API.

D. Choose partition keys in a way that results in a uniform record distribution across shards.

E. Customize the application code to include retry logic to improve performance.

ANSWER: A C

QUESTION NO: 10

An ecommerce company ingests a large set of clickstream data in JSON format and stores the data in Amazon S3. Business analysts from multiple product divisions need to use Amazon Athena to analyze the data. The company's analytics team must design a solution to monitor the daily data usage for Athena by each product division. The solution also must produce a warning when a division exceeds its quota.

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

A. Use a CREATE TABLE AS SELECT (CTAS) statement to create separate tables for each product division. Use AWS Budgets to track Athena usage. Configure a threshold for the budget. Use Amazon Simple Notification Service (Amazon SNS) to send notifications when thresholds are breached.

B. Create an AWS account for each division. Provide cross-account access to an AWS Glue Data Catalog to all the accounts. Set an Amazon CloudWatch alarm to monitor Athena usage. Use Amazon Simple Notification Service (Amazon SNS) to send notifications.

C. Create an Athena workgroup for each division. Configure a data usage control for each workgroup and a time period of 1 day. Configure an action to send notifications to an Amazon Simple Notification Service (Amazon SNS) topic.

D. Create an AWS account for each division. Configure an AWS Glue Data Catalog in each account. Set an Amazon CloudWatch alarm to monitor Athena usage. Use Amazon Simple Notification Service (Amazon SNS) to send notifications.

ANSWER: A

Explanation:

Reference: https://docs.aws.amazon.com/athena/latest/ug/ctas-console.html