

Amazon Web Services

Exam Questions DAS-C01

AWS Certified Data Analytics - Specialty



NEW QUESTION 1

A market data company aggregates external data sources to create a detailed view of product consumption in different countries. The company wants to sell this data to external parties through a subscription. To achieve this goal, the company needs to make its data securely available to external parties who are also AWS users.

What should the company do to meet these requirements with the LEAST operational overhead?

- A. Store the data in Amazon S3. Share the data by using presigned URLs for security.
- B. Store the data in Amazon S3. Share the data by using S3 bucket ACLs.
- C. Upload the data to AWS Data Exchange for storage.
- D. Share the data by using presigned URLs for security.
- E. Upload the data to AWS Data Exchange for storage.
- F. Share the data by using the AWS Data Exchange sharing wizard.

Answer: A

NEW QUESTION 2

A company hosts an Apache Flink application on premises. The application processes data from several Apache Kafka clusters. The data originates from a variety of sources, such as web applications, mobile apps, and operational databases. The company has migrated some of these sources to AWS and now wants to migrate the Flink application. The company must ensure that data that resides in databases within the VPC does not traverse the internet. The application must be able to process all the data that comes from the company's AWS solution, on-premises resources, and the public internet.

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

- A. Implement Flink on Amazon EC2 within the company's VPC. Create Amazon Managed Streaming for Apache Kafka (Amazon MSK) clusters in the VPC to collect data that comes from applications and databases within the VPC. Use Amazon Kinesis Data Streams to collect data that comes from the public internet. Configure Flink to have sources from Kinesis Data Streams, Amazon MSK, and any on-premises Kafka clusters by using AWS Client VPN or AWS Direct Connect.
- B. Implement Flink on Amazon EC2 within the company's VPC. Use Amazon Kinesis Data Streams to collect data that comes from applications and databases within the VPC and the public internet. Configure Flink to have sources from Kinesis Data Streams and any on-premises Kafka clusters by using AWS Client VPN or AWS Direct Connect.
- C. Create an Amazon Kinesis Data Analytics application by uploading the compiled Flink jar file. Use Amazon Kinesis Data Streams to collect data that comes from applications and databases within the VPC and the public internet. Configure the Kinesis Data Analytics application to have sources from Kinesis Data Streams and any on-premises Kafka clusters by using AWS Client VPN or AWS Direct Connect.
- D. Create an Amazon Kinesis Data Analytics application by uploading the compiled Flink jar file. Create Amazon Managed Streaming for Apache Kafka (Amazon MSK) clusters in the company's VPC to collect data that comes from applications and databases within the VPC. Use Amazon Kinesis Data Streams to collect data that comes from the public internet. Configure the Kinesis Data Analytics application to have sources from Kinesis Data Stream.
- E. Amazon MSK and any on-premises Kafka clusters by using AWS Client VPN or AWS Direct Connect.

Answer: D

NEW QUESTION 3

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

- 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: BCD

NEW QUESTION 4

A company's marketing team has asked for help in identifying a high performing long-term storage service for their data based on the following requirements:

- > The data size is approximately 32 TB uncompressed.
- > There is a low volume of single-row inserts each day.
- > There is a high volume of aggregation queries each day.
- > Multiple complex joins are performed.
- > The queries typically involve a small subset of the columns in a table.

Which storage service will provide the MOST performant solution?

- A. Amazon Aurora MySQL
- B. Amazon Redshift
- C. Amazon Neptune
- D. Amazon Elasticsearch

Answer: B

NEW QUESTION 5

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?

- 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

NEW QUESTION 6

A data analytics specialist is building an automated ETL ingestion pipeline using AWS Glue to ingest compressed files that have been uploaded to an Amazon S3 bucket. The ingestion pipeline should support incremental data processing. Which AWS Glue feature should the data analytics specialist use to meet this requirement?

- A. Workflows
- B. Triggers
- C. Job bookmarks
- D. Classifiers

Answer: C

NEW QUESTION 7

A company wants to use an automatic machine learning (ML) Random Cut Forest (RCF) algorithm to visualize complex real-world scenarios, such as detecting seasonality and trends, excluding outliers, and imputing missing values. The team working on this project is non-technical and is looking for an out-of-the-box solution that will require the LEAST amount of management overhead. Which solution will meet these requirements?

- A. Use an AWS Glue ML transform to create a forecast and then use Amazon QuickSight to visualize the data.
- B. Use Amazon QuickSight to visualize the data and then use ML-powered forecasting to forecast the key business metrics.
- C. Use a pre-build ML AMI from the AWS Marketplace to create forecasts and then use Amazon QuickSight to visualize the data.
- D. Use calculated fields to create a new forecast and then use Amazon QuickSight to visualize the data.

Answer: A

NEW QUESTION 8

An online retail company with millions of users around the globe wants to improve its ecommerce analytics capabilities. Currently, clickstream data is uploaded directly to Amazon S3 as compressed files. Several times each day, an application running on Amazon EC2 processes the data and makes search options and reports available for visualization by editors and marketers. The company wants to make website clicks and aggregated data available to editors and marketers in minutes to enable them to connect with users more effectively. Which options will help meet these requirements in the MOST efficient way? (Choose two.)

- A. Use Amazon Kinesis Data Firehose to upload compressed and batched clickstream records to Amazon Elasticsearch Service.
- B. Upload clickstream records to Amazon S3 as compressed file
- C. Then use AWS Lambda to send data to Amazon Elasticsearch Service from Amazon S3.
- D. Use Amazon Elasticsearch Service deployed on Amazon EC2 to aggregate, filter, and process the data.Refresh content performance dashboards in near-real time.
- E. Use Kibana to aggregate, filter, and visualize the data stored in Amazon Elasticsearch Service
- F. Refresh content performance dashboards in near-real time.
- G. Upload clickstream records from Amazon S3 to Amazon Kinesis Data Streams and use a Kinesis Data Streams consumer to send records to Amazon Elasticsearch Service.

Answer: AD

NEW QUESTION 9

An Amazon Redshift database contains sensitive user data. Logging is necessary to meet compliance requirements. The logs must contain database authentication attempts, connections, and disconnections. The logs must also contain each query run against the database and record which database user ran each query. Which steps will create the required logs?

- A. Enable Amazon Redshift Enhanced VPC Routing
- B. Enable VPC Flow Logs to monitor traffic.
- C. Allow access to the Amazon Redshift database using AWS IAM roles
- D. Log access using AWS CloudTrail.
- E. Enable audit logging for Amazon Redshift using the AWS Management Console or the AWS CLI.
- F. Enable and download audit reports from AWS Artifact.

Answer: C

NEW QUESTION 10

A company stores Apache Parquet-formatted files in Amazon S3. The company uses an AWS Glue Data Catalog to store the table metadata and Amazon Athena to query and analyze the data. The tables have a large number of partitions. The queries are only run on small subsets of data in the table. A data analyst adds new time partitions into the table as new data arrives. The data analyst has been asked to reduce the query runtime. Which solution will provide the MOST reduction in the query runtime?

- A. Convert the Parquet files to the csv file format. Then attempt to query the data again
- B. Convert the Parquet files to the Apache ORC file format
- C. Then attempt to query the data again
- D. Use partition projection to speed up the processing of the partitioned table
- E. Add more partitions to be used over the table
- F. Then filter over two partitions and put all columns in the WHERE clause

Answer: C

NEW QUESTION 11

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: CD

Explanation:

<https://aws.amazon.com/blogs/big-data/under-the-hood-scaling-your-kinesis-data-streams/>

NEW QUESTION 12

An IoT company wants to release a new device that will collect data to track sleep overnight on an intelligent mattress. Sensors will send data that will be uploaded to an Amazon S3 bucket. About 2 MB of data is generated each night for each bed. Data must be processed and summarized for each user, and the results need to be available as soon as possible. Part of the process consists of time windowing and other functions. Based on tests with a Python script, every run will require about 1 GB of memory and will complete within a couple of minutes.

Which solution will run the script in the MOST cost-effective way?

- A. AWS Lambda with a Python script
- B. AWS Glue with a Scala job
- C. Amazon EMR with an Apache Spark script
- D. AWS Glue with a PySpark job

Answer: A

NEW QUESTION 13

A company has developed an Apache Hive script to batch process data stored in Amazon S3. The script needs to run once every day and store the output in Amazon S3. The company tested the script, and it completes within 30 minutes on a small local three-node cluster.

Which solution is the MOST cost-effective for scheduling and executing the script?

- A. Create an AWS Lambda function to spin up an Amazon EMR cluster with a Hive execution step.
- B. Set `KeepJobFlowAliveWhenNoSteps` to false and disable the termination protection flag.
- C. Use Amazon CloudWatch Events to schedule the Lambda function to run daily.
- D. Use the AWS Management Console to spin up an Amazon EMR cluster with Python Hive, and Apache Oozie.
- E. Hive, and Apache Oozie.
- F. Set the termination protection flag to true and use Spot Instances for the core nodes of the cluster.
- G. Configure an Oozie workflow in the cluster to invoke the Hive script daily.
- H. Create an AWS Glue job with the Hive script to perform the batch operation.
- I. Configure the job to run once a day using a time-based schedule.
- J. Use AWS Lambda layers and load the Hive runtime to AWS Lambda and copy the Hive script. Schedule the Lambda function to run daily by creating a workflow using AWS Step Functions.

Answer: C

NEW QUESTION 14

A company analyzes historical data and needs to query data that is stored in Amazon S3. New data is generated daily as .csv files that are stored in Amazon S3. The company's analysts are using Amazon Athena to perform SQL queries against a recent subset of the overall data. The amount of data that is ingested into Amazon S3 has increased substantially over time, and the query latency also has increased.

Which solutions could the company implement to improve query performance? (Choose two.)

- A. Use MySQL Workbench on an Amazon EC2 instance, and connect to Athena by using a JDBC or ODBC connector.
- B. Run the query from MySQL Workbench instead of Athena directly.
- C. Use Athena to extract the data and store it in Apache Parquet format on a daily basis.
- D. Query the extracted data.
- E. Run a daily AWS Glue ETL job to convert the data files to Apache Parquet and to partition the converted file.
- F. Create a periodic AWS Glue crawler to automatically crawl the partitioned data on a daily basis.
- G. Run a daily AWS Glue ETL job to compress the data files by using the .gzip format.
- H. Query the compressed data.
- I. Run a daily AWS Glue ETL job to compress the data files by using the .lzo format.
- J. Query the compressed data.

Answer: BC

NEW QUESTION 15

A company operates toll services for highways across the country and collects data that is used to understand usage patterns. Analysts have requested the ability to run traffic reports in near-real time. The company is interested in building an ingestion pipeline that loads all the data into an Amazon Redshift cluster and alerts operations personnel when toll traffic for a particular toll station does not meet a specified threshold. Station data and the corresponding threshold values are stored in Amazon S3.

Which approach is the MOST efficient way to meet these requirements?

- A. Use Amazon Kinesis Data Firehose to collect data and deliver it to Amazon Redshift and Amazon Kinesis Data Analytics simultaneously.

- B. Create a reference data source in Kinesis Data Analytics to temporarily store the threshold values from Amazon S3 and compare the count of vehicles for a particular toll station against its corresponding threshold value
- C. Use AWS Lambda to publish an Amazon Simple Notification Service (Amazon SNS) notification if the threshold is not met.
- D. Use Amazon Kinesis Data Streams to collect all the data from toll station
- E. Create a stream in Kinesis Data Streams to temporarily store the threshold values from Amazon S3. Send both streams to Amazon Kinesis Data Analytics to compare the count of vehicles for a particular toll station against its corresponding threshold value
- F. Use AWS Lambda to publish an Amazon Simple Notification Service (Amazon SNS) notification if the threshold is not met
- G. Connect Amazon Kinesis Data Firehose to Kinesis Data Streams to deliver the data to Amazon Redshift.
- H. Use Amazon Kinesis Data Firehose to collect data and deliver it to Amazon Redshift
- I. Then, automatically trigger an AWS Lambda function that queries the data in Amazon Redshift, compares the count of vehicles for a particular toll station against its corresponding threshold values read from Amazon S3, and publishes an Amazon Simple Notification Service (Amazon SNS) notification if the threshold is not met.
- J. Use Amazon Kinesis Data Firehose to collect data and deliver it to Amazon Redshift and Amazon Kinesis Data Analytics simultaneously
- K. Use Kinesis Data Analytics to compare the count of vehicles against the threshold value for the station stored in a table as an in-application stream based on information stored in Amazon S3. Configure an AWS Lambda function as an output for the application that will publish an Amazon Simple Queue Service (Amazon SQS) notification to alert operations personnel if the threshold is not met.

Answer: D

NEW QUESTION 16

A large retailer has successfully migrated to an Amazon S3 data lake architecture. The company's marketing team is using Amazon Redshift and Amazon QuickSight to analyze data, and derive and visualize insights. To ensure the marketing team has the most up-to-date actionable information, a data analyst implements nightly refreshes of Amazon Redshift using terabytes of updates from the previous day.

After the first nightly refresh, users report that half of the most popular dashboards that had been running correctly before the refresh are now running much slower. Amazon CloudWatch does not show any alerts.

What is the MOST likely cause for the performance degradation?

- A. The dashboards are suffering from inefficient SQL queries.
- B. The cluster is undersized for the queries being run by the dashboards.
- C. The nightly data refreshes are causing a lingering transaction that cannot be automatically closed by Amazon Redshift due to ongoing user workloads.
- D. The nightly data refreshes left the dashboard tables in need of a vacuum operation that could not be automatically performed by Amazon Redshift due to ongoing user workloads.

Answer: D

Explanation:

<https://github.com/awsdocs/amazon-redshift-developer-guide/issues/21>

NEW QUESTION 17

A regional energy company collects voltage data from sensors attached to buildings. To address any known dangerous conditions, the company wants to be alerted when a sequence of two voltage drops is detected within 10 minutes of a voltage spike at the same building. It is important to ensure that all messages are delivered as quickly as possible. The system must be fully managed and highly available. The company also needs a solution that will automatically scale up as it covers additional cities with this monitoring feature. The alerting system is subscribed to an Amazon SNS topic for remediation.

Which solution meets these requirements?

- A. Create an Amazon Managed Streaming for Kafka cluster to ingest the data, and use an Apache Spark Streaming with Apache Kafka consumer API in an automatically scaled Amazon EMR cluster to process the incoming data
- B. Use the Spark Streaming application to detect the known event sequence and send the SNS message.
- C. Create a REST-based web service using Amazon API Gateway in front of an AWS Lambda function. Create an Amazon RDS for PostgreSQL database with sufficient Provisioned IOPS (PIOPS). In the Lambda function, store incoming events in the RDS database and query the latest data to detect the known event sequence and send the SNS message.
- D. Create an Amazon Kinesis Data Firehose delivery stream to capture the incoming sensor data
- E. Use an AWS Lambda transformation function to detect the known event sequence and send the SNS message.
- F. Create an Amazon Kinesis data stream to capture the incoming sensor data and create another stream for alert message
- G. Set up AWS Application Auto Scaling on both
- H. Create a Kinesis Data Analytics for Java application to detect the known event sequence, and add a message to the message stream
- I. Configure an AWS Lambda function to poll the message stream and publish to the SNS topic.

Answer: D

NEW QUESTION 18

Once a month, a company receives a 100 MB .csv file compressed with gzip. The file contains 50,000 property listing records and is stored in Amazon S3 Glacier. The company needs its data analyst to query a subset of the data for a specific vendor.

What is the most cost-effective solution?

- A. Load the data into Amazon S3 and query it with Amazon S3 Select.
- B. Query the data from Amazon S3 Glacier directly with Amazon Glacier Select.
- C. Load the data to Amazon S3 and query it with Amazon Athena.
- D. Load the data to Amazon S3 and query it with Amazon Redshift Spectrum.

Answer: A

NEW QUESTION 19

A company's data analyst needs to ensure that queries executed in Amazon Athena cannot scan more than a prescribed amount of data for cost control purposes. Queries that exceed the prescribed threshold must be canceled immediately.

What should the data analyst do to achieve this?

- A. Configure Athena to invoke an AWS Lambda function that terminates queries when the prescribed threshold is crossed.
- B. For each workgroup, set the control limit for each query to the prescribed threshold.

- C. Enforce the prescribed threshold on all Amazon S3 bucket policies
- D. For each workgroup, set the workgroup-wide data usage control limit to the prescribed threshold.

Answer: B

Explanation:

<https://docs.aws.amazon.com/athena/latest/ug/manage-queries-control-costs-with-workgroups.html>

NEW QUESTION 20

A company analyzes its data in an Amazon Redshift data warehouse, which currently has a cluster of three dense storage nodes. Due to a recent business acquisition, the company needs to load an additional 4 TB of user data into Amazon Redshift. The engineering team will combine all the user data and apply complex calculations that require I/O intensive resources. The company needs to adjust the cluster's capacity to support the change in analytical and storage requirements.

Which solution meets these requirements?

- A. Resize the cluster using elastic resize with dense compute nodes.
- B. Resize the cluster using classic resize with dense compute nodes.
- C. Resize the cluster using elastic resize with dense storage nodes.
- D. Resize the cluster using classic resize with dense storage nodes.

Answer: C

NEW QUESTION 21

.....

Thank You for Trying Our Product

We offer two products:

1st - We have Practice Tests Software with Actual Exam Questions

2nd - Questions and Answers in PDF Format

DAS-C01 Practice Exam Features:

- * DAS-C01 Questions and Answers Updated Frequently
- * DAS-C01 Practice Questions Verified by Expert Senior Certified Staff
- * DAS-C01 Most Realistic Questions that Guarantee you a Pass on Your First Try
- * DAS-C01 Practice Test Questions in Multiple Choice Formats and Updates for 1 Year

100% Actual & Verified — Instant Download, Please Click
[Order The DAS-C01 Practice Test Here](#)