DAS-C01 Q&As

AWS Certified Data Analytics - Specialty (DAS-C01)

Pass Amazon DAS-C01 Exam with 100% Guarantee

Free Download Real Questions & Answers **PDF** and **VCE** file from:

https://www.lead4pass.com/das-c01.html

100% Passing Guarantee
100% Money Back Assurance

Following Questions and Answers are all new published by Amazon Official Exam Center

QUESTION 1

A data analyst is designing a solution to interactively query datasets with SQL using a JDBC connection. Users will join data stored in Amazon S3 in Apache ORC format with data stored in Amazon OpenSearch Service (Amazon Elasticsearch Service) and Amazon Aurora MySQL.

Which solution will provide the MOST up-to-date results?

A. Use AWS Glue jobs to ETL data from Amazon ES and Aurora MySQL to Amazon S3. Query the data with Amazon Athena.

B. Use Amazon DMS to stream data from Amazon ES and Aurora MySQL to Amazon Redshift. Query the data with Amazon Redshift.

C. Query all the datasets in place with Apache Spark SQL running on an AWS Glue developer endpoint.

D. Query all the datasets in place with Apache Presto running on Amazon EMR.

Correct Answer: C

QUESTION 2

A company stores its sales and marketing data that includes personally identifiable information (PII) in Amazon S3. The company allows its analysts to launch their own Amazon EMR cluster and run analytics reports with the data. To meet compliance requirements, the company must ensure the data is not publicly accessible throughout this process. A data engineer has secured Amazon S3 but must ensure the individual EMR clusters created by the analysts are not exposed to the public internet.

Which solution should the data engineer to meet this compliance requirement with LEAST amount of effort?

A. Create an EMR security configuration and ensure the security configuration is associated with the EMR clusters when they are created.

B. Check the security group of the EMR clusters regularly to ensure it does not allow inbound traffic from IPv4 0.0.0.0/0 or IPv6 ::/0.

C. Enable the block public access setting for Amazon EMR at the account level before any EMR cluster is created.

D. Use AWS WAF to block public internet access to the EMR clusters across the board.

Correct Answer: B

Reference: https://docs.aws.amazon.com/emr/latest/ManagementGuide/emr-security-groups.html

QUESTION 3

A company is building a service to monitor fleets of vehicles. The company collects IoT data from a device in each vehicle and loads the data into Amazon Redshift in near-real time. Fleet owners upload .csv files containing vehicle reference data into Amazon S3 at different times throughout the day. A nightly process loads the vehicle reference data from Amazon S3 into Amazon Redshift. The company joins the IoT data from the device and the vehicle reference data to power reporting and dashboards. Fleet owners are frustrated by waiting a day for the dashboards to update.

Which solution would provide the SHORTEST delay between uploading reference data to Amazon S3 and the change showing up in the owners\\' dashboards?

A. Use S3 event notifications to trigger an AWS Lambda function to copy the vehicle reference data into Amazon Redshift immediately when the reference data is uploaded to Amazon S3.

B. Create and schedule an AWS Glue Spark job to run every 5 minutes. The job inserts reference data into Amazon Redshift.

C. Send reference data to Amazon Kinesis Data Streams. Configure the Kinesis data stream to directly load the reference data into Amazon Redshift in real time.

D. Send the reference data to an Amazon Kinesis Data Firehose delivery stream. Configure Kinesis with a buffer interval of 60 seconds and to directly load the data into Amazon Redshift.

Correct Answer: A

QUESTION 4

A mobile gaming company wants to capture data from its gaming app and make the data available for analysis immediately. The data record size will be approximately 20 KB. The company is concerned about achieving optimal throughput from each device. Additionally, the company wants to develop a data stream processing application with dedicated throughput for each consumer.

Which solution would achieve this goal?

A. Have the app call the PutRecords API to send data to Amazon Kinesis Data Streams. Use the enhanced fan-out feature while consuming the data.

B. Have the app call the PutRecordBatch API to send data to Amazon Kinesis Data Firehose. Submit a support case to enable dedicated throughput on the account.

C. Have the app use Amazon Kinesis Producer Library (KPL) to send data to Kinesis Data Firehose. Use the enhanced fan-out feature while consuming the data.

D. Have the app call the PutRecords API to send data to Amazon Kinesis Data Streams. Host the stream-processing application on Amazon EC2 with Auto Scaling.

Correct Answer: D

QUESTION 5

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.

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

Correct Answer: A

Matching Records with AWS Lake Formation FindMatches Reference:
https://docs.aws.amazon.com/glue/latest/dg/machine-learning.html

QUESTION 6

A company is hosting an enterprise reporting solution with Amazon Redshift. The application provides reporting capabilities to three main groups: an executive group to access financial reports, a data analyst group to run long-running ad-hoc queries, and a data engineering group to run stored procedures and ETL processes. The executive team requires queries to run with optimal performance. The data engineering team expects queries to take minutes.

Which Amazon Redshift feature meets the requirements for this task?

A. Concurrency scaling

B. Short query acceleration (SQA)

C. Workload management (WLM)

D. Materialized views

Correct Answer: D

Materialized views:

Reference: https://aws.amazon.com/redshift/faqs/

QUESTION 7

A web retail company wants to implement a near-real-time clickstream analytics solution. The company wants to analyze the data with an open-source package. The analytics application will process the raw data only once, but other applications will need immediate access to the raw data for up to 1 year.

Which solution meets these requirements with the LEAST amount of operational effort?

A. Use Amazon Kinesis Data Streams to collect the data. Use Amazon EMR with Apache Flink to consume and process the data from the Kinesis data stream. Set the retention period of the Kinesis data stream to 8.760 hours.

B. Use Amazon Kinesis Data Streams to collect the data. Use Amazon Kinesis Data Analytics with Apache Flink to process the data in real time. Set the retention period of the Kinesis data stream to 8,760 hours.

C. Use Amazon Managed Streaming for Apache Kafka (Amazon MSK) to collect the data. Use Amazon EMR with Apache Flink to consume and process the data from the Amazon MSK stream. Set the log retention hours to 8,760.

D. Use Amazon Kinesis Data Streams to collect the data. Use Amazon EMR with Apache Flink to consume and process the data from the Kinesis data stream. Create an Amazon Kinesis Data Firehose delivery stream to store the data in Amazon S3. Set an S3 Lifecycle policy to delete the data after 365 days.

Correct Answer: B

QUESTION 8

A company has an application that uses the Amazon Kinesis Client Library (KCL) to read records from a Kinesis data stream.

After a successful marketing campaign, the application experienced a significant increase in usage. As a result, a data analyst had to split some shards in the data stream. When the shards were split, the application started throwing an ExpiredIteratorExceptions error sporadically.

What should the data analyst do to resolve this?

A. Increase the number of threads that process the stream records.

B. Increase the provisioned read capacity units assigned to the stream\\'s Amazon DynamoDB table.

C. Increase the provisioned write capacity units assigned to the stream\\'s Amazon DynamoDB table.

D. Decrease the provisioned write capacity units assigned to the stream\\'s Amazon DynamoDB table.

Correct Answer: C

QUESTION 9

A social media company is using business intelligence tools to analyze its data for forecasting. The company is using Apache Kafka to ingest the low-velocity data in near-real time. The company wants to build dynamic dashboards with machine learning (ML) insights to forecast key business trends. The dashboards must provide hourly updates from data in Amazon S3. Various teams at the company want to view the dashboards by using Amazon QuickSight with ML insights. The solution also must correct the scalability problems that the company experiences when it uses its current architecture to ingest data.

Which solution will MOST cost-effectively meet these requirements?

A. Replace Kafka with Amazon Managed Streaming for Apache Kafka. Ingest the data by using AWS Lambda, and store the data in Amazon S3. Use QuickSight Standard edition to refresh the data in SPICE from Amazon S3 hourly and create a dynamic dashboard with forecasting and ML insights.

B. Replace Kafka with an Amazon Kinesis data stream. Use an Amazon Kinesis Data Firehose delivery stream to consume the data and store the data in Amazon S3. Use QuickSight Enterprise edition to refresh the data in SPICE from Amazon S3 hourly and create a dynamic dashboard with forecasting and ML insights.

C. Configure the Kafka-Kinesis-Connector to publish the data to an Amazon Kinesis Data Firehose delivery stream that is configured to store the data in Amazon S3. Use QuickSight Enterprise edition to refresh the data in SPICE from Amazon S3 hourly and create a dynamic dashboard with forecasting and ML insights.

D. Configure the Kafka-Kinesis-Connector to publish the data to an Amazon Kinesis Data Firehose delivery stream that is configured to store the data in Amazon S3. Configure an AWS Glue crawler to crawl the data. Use an Amazon Athena data source with QuickSight Standard edition to refresh the data in SPICE hourly and create a dynamic dashboard with forecasting and ML insights.

Correct Answer: B

Reference: https://noise.getoto.net/tag/amazon-kinesis-data-firehose/

QUESTION 10

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. Run the query from MySQL Workbench instead of Athena directly.

B. Use Athena to extract the data and store it in Apache Parquet format on a daily basis. Query the extracted data.

C. Run a daily AWS Glue ETL job to convert the data files to Apache Parquet and to partition the converted files. Create a periodic AWS Glue crawler to automatically crawl the partitioned data on a daily basis.

D. Run a daily AWS Glue ETL job to compress the data files by using the .gzip format. Query the compressed data.

E. Run a daily AWS Glue ETL job to compress the data files by using the .lzo format. Query the compressed data.

Correct Answer: BC

Reference: https://www.upsolver.com/blog/apache-parquet-why-use https://aws.amazon.com/blogs/big-data/work-with-partitioned-data-in-aws-glue/

QUESTION 11

A company needs to collect streaming data from several sources and store the data in the AWS Cloud. The dataset is heavily structured, but analysts need to perform several complex SQL queries and need consistent performance. Some of the data is queried more frequently than the rest. The company wants a solution that meets its performance requirements in a cost-effective manner.

Which solution meets these requirements?

A. Use Amazon Managed Streaming for Apache Kafka to ingest the data to save it to Amazon S3. Use Amazon Athena to perform SQL queries over the ingested data.

B. Use Amazon Managed Streaming for Apache Kafka to ingest the data to save it to Amazon Redshift. Enable Amazon Redshift workload management (WLM) to prioritize workloads.

C. Use Amazon Kinesis Data Firehose to ingest the data to save it to Amazon Redshift. Enable Amazon Redshift workload management (WLM) to prioritize workloads.

D. Use Amazon Kinesis Data Firehose to ingest the data to save it to Amazon S3. Load frequently queried data to Amazon Redshift using the COPY command. Use Amazon Redshift Spectrum for less frequently queried data.

Correct Answer: B

Reference: https://aws.amazon.com/about-aws/whats-new/2019/

QUESTION 12

A company wants to collect and process events data from different departments in near-real time. Before storing the data in Amazon S3, the company needs to clean the data by standardizing the format of the address and timestamp columns. The data varies in size based on the overall load at each particular point in time. A single data record can be 100 KB-10 MB.

How should a data analytics specialist design the solution for data ingestion?

A. Use Amazon Kinesis Data Streams. Configure a stream for the raw data. Use a Kinesis Agent to write data to the stream. Create an Amazon Kinesis Data Analytics application that reads data from the raw stream, cleanses it, and stores the output to Amazon S3.

B. Use Amazon Kinesis Data Firehose. Configure a Firehose delivery stream with a preprocessing AWS Lambda function for data cleansing. Use a Kinesis Agent to write data to the delivery stream. Configure Kinesis Data Firehose to deliver the data to Amazon S3.

C. Use Amazon Managed Streaming for Apache Kafka. Configure a topic for the raw data. Use a Kafka producer to write data to the topic. Create an application on Amazon EC2 that reads data from the topic by using the Apache Kafka consumer API, cleanses the data, and writes to Amazon S3.

D. Use Amazon Simple Queue Service (Amazon SQS). Configure an AWS Lambda function to read events from the SQS queue and upload the events to Amazon S3.

Correct Answer: B

QUESTION 13

A global pharmaceutical company receives test results for new drugs from various testing facilities worldwide. The results are sent in millions of 1 KB-sized JSON objects to an Amazon S3 bucket owned by the company. The data engineering team needs to process those files, convert them into Apache Parquet format, and load them into Amazon Redshift for data analysts to perform dashboard reporting. The engineering team uses AWS Glue to process the objects, AWS Step Functions for process orchestration, and Amazon CloudWatch for job scheduling.

More testing facilities were recently added, and the time to process files is increasing.

What will MOST efficiently decrease the data processing time?

A. Use AWS Lambda to group the small files into larger files. Write the files back to Amazon S3. Process the files using AWS Glue and load them into Amazon Redshift tables.

B. Use the AWS Glue dynamic frame file grouping option while ingesting the raw input files. Process the files and load them into Amazon Redshift tables.

C. Use the Amazon Redshift COPY command to move the files from Amazon S3 into Amazon Redshift tables directly. Process the files in Amazon Redshift.

D. Use Amazon EMR instead of AWS Glue to group the small input files. Process the files in Amazon EMR and load them into Amazon Redshift tables.

Correct Answer: A

Reference: https://docs.aws.amazon.com/prescriptive-guidance/latest/patterns/build-an-etl-service-pipeline-to-load-data-

incrementally-from-amazon-s3-to-amazon-redshift-using-aws-glue.html

To Read the Whole Q&As, please purchase the Complete Version from Our website.

Try our product !

100% Guaranteed Success
100% Money Back Guarantee
365 Days Free Update
Instant Download After Purchase
24x7 Customer Support
Average 99.9% Success Rate
More than 800,000 Satisfied Customers Worldwide
Multi-Platform capabilities - Windows, Mac, Android, iPhone, iPod, iPad, Kindle

We provide exam PDF and VCE of Cisco, Microsoft, IBM, CompTIA, Oracle and other IT Certifications.
You can view Vendor list of All Certification Exams offered:

https://www.lead4pass.com/allproducts

Need Help

Please provide as much detail as possible so we can best assist you.
To update a previously submitted ticket: