

## [2025-November-NewBraindump2go DEA-C01 Exam Prep Free[Q105-Q155]

2025/November Latest Braindump2go DEA-C01 Exam Dumps with PDF and VCE Free Updated Today! Following are some new Braindump2go DEA-C01 Real Exam Questions!

QUESTION 105A company has a data warehouse that contains a table that is named Sales. The company stores the table in Amazon Redshift. The table includes a column that is named city\_name. The company wants to query the table to find all rows that have a city\_name that starts with "San" or "El". Which SQL query will meet this requirement? A. Select \* from Sales where city\_name ~ '\$(San|El)\*'; B. Select \* from Sales where city\_name ~ '(San|El)\*'; C. Select \* from Sales where city\_name ~ '\$(San&El)\*'; D. Select \* from Sales where city\_name ~ '(San&El)\*'; Answer: B

Explanation: This query uses a regular expression pattern with the ~ operator. The caret at the beginning of the pattern indicates that the match must start at the beginning of the string. (San|El) matches either ?San? or ?El?, and \* means zero or more of the preceding element. So this query will return all rows where city\_name starts with either ?San? or ?El?. QUESTION 106A company needs to send customer call data from its on-premises PostgreSQL database to AWS to generate near real-time insights. The solution must capture and load updates from operational data stores that run in the PostgreSQL database. The data changes continuously. A data engineer configures an AWS Database Migration Service (AWS DMS) ongoing replication task. The task reads changes in near real time from the PostgreSQL source database transaction logs for each table. The task then sends the data to an Amazon Redshift cluster for processing. The data engineer discovers latency issues during the change data capture (CDC) of the task. The data engineer thinks that the PostgreSQL source database is causing the high latency. Which solution will confirm that the PostgreSQL database is the source of the high latency? A. Use Amazon CloudWatch to monitor the DMS task. Examine the CDCIncomingChanges metric to identify delays in the CDC from the source database. B. Verify that logical replication of the source database is configured in the postgresql.conf configuration file. C. Enable Amazon CloudWatch Logs for the DMS endpoint of the source database. Check for error messages. D. Use Amazon CloudWatch to monitor the DMS task. Examine the CDCLatencySource metric to identify delays in the CDC from the source database. Answer: D Explanation:

[https://docs.aws.amazon.com/dms/latest/userguide/CHAP\\_Troubleshooting\\_Latency.html](https://docs.aws.amazon.com/dms/latest/userguide/CHAP_Troubleshooting_Latency.html) A high CDCLatencySource metric indicates that the process of capturing changes from the source is delayed. QUESTION 107A lab uses IoT sensors to monitor humidity, temperature, and pressure for a project. The sensors send 100 KB of data every 10 seconds. A downstream process will read the data from an Amazon S3 bucket every 30 seconds. Which solution will deliver the data to the S3 bucket with the LEAST latency? A. Use Amazon Kinesis Data Streams and Amazon Kinesis Data Firehose to deliver the data to the S3 bucket. Use the default buffer interval for Kinesis Data Firehose. B. Use Amazon Kinesis Data Streams to deliver the data to the S3 bucket. Configure the stream to use 5 provisioned shards. C. Use Amazon Kinesis Data Streams and call the Kinesis Client Library to deliver the data to the S3 bucket. Use a 5 second buffer interval from an application. D. Use Amazon Managed Service for Apache Flink (previously known as Amazon Kinesis Data Analytics) and Amazon Kinesis Data Firehose to deliver the data to the S3 bucket. Use a 5 second buffer interval for Kinesis Data Firehose. Answer: D QUESTION 108A company wants to use machine learning (ML) to perform analytics on data that is in an Amazon S3 data lake. The company has two data transformation requirements that will give consumers within the company the ability to create reports. The company must perform daily transformations on 300 GB of data that is in a variety of formats that must arrive in Amazon S3 at a scheduled time. The company must perform one-time transformations of terabytes of archived data that is in the S3 data lake. The company uses Amazon Managed Workflows for Apache Airflow (Amazon MWAA) Directed Acyclic Graphs (DAGs) to orchestrate processing. Which combination of tasks should the company schedule in the Amazon MWAA DAGs to meet these requirements MOST cost-effectively? (Choose two.) A. For daily incoming data, use AWS Glue crawlers to scan and identify the schema. B. For daily incoming data, use Amazon Athena to scan and identify the schema. C. For daily incoming data, use Amazon Redshift to perform transformations. D. For daily and archived data, use Amazon EMR to perform data transformations. E. For archived data, use Amazon SageMaker to perform data transformations. Answer: AD QUESTION 109A retail company uses AWS Glue for extract, transform, and load (ETL) operations on a dataset that contains information about customer orders. The company wants to implement specific validation rules to ensure data accuracy and consistency. Which solution will meet these requirements? A. Use AWS Glue job bookmarks to track the data for accuracy and consistency. B. Create custom AWS Glue Data Quality rulesets to define specific data quality checks. C. Use the built-in AWS Glue Data Quality transforms for standard data quality validations. D. Use AWS Glue Data Catalog to maintain a centralized data schema and metadata repository. Answer: B Explanation: Custom AWS Glue Data Quality rulesets allow you to define precise data quality checks tailored to your specific needs, ensuring that the data meets the required standards of accuracy and consistency. This approach provides flexibility to implement a wide range of validation rules based on your business requirements. QUESTION 110An insurance company stores transaction data that the company compressed with gzip. The company

needs to query the transaction data for occasional audits. Which solution will meet this requirement in the MOST cost-effective way? A. Store the data in Amazon Glacier Flexible Retrieval. Use Amazon S3 Glacier Select to query the data. B. Store the data in Amazon S3. Use Amazon S3 Select to query the data. C. Store the data in Amazon S3. Use Amazon Athena to query the data. D. Store the data in Amazon Glacier Instant Retrieval. Use Amazon Athena to query the data. Answer: A

QUESTION 111 A data engineer finished testing an Amazon Redshift stored procedure that processes and inserts data into a table that is not mission critical. The engineer wants to automatically run the stored procedure on a daily basis. Which solution will meet this requirement in the MOST cost-effective way? A. Create an AWS Lambda function to schedule a cron job to run the stored procedure. B. Schedule and run the stored procedure by using the Amazon Redshift Data API in an Amazon EC2 Spot Instance. C. Use query editor v2 to run the stored procedure on a schedule. D. Schedule an AWS Glue Python shell job to run the stored procedure. Answer: C

Explanation: <https://docs.aws.amazon.com/redshift/latest/mgmt/query-editor-v2-schedule-query.html>

QUESTION 112 A marketing company collects clickstream data. The company sends the clickstream data to Amazon Kinesis Data Firehose and stores the clickstream data in Amazon S3. The company wants to build a series of dashboards that hundreds of users from multiple departments will use. The company will use Amazon QuickSight to develop the dashboards. The company wants a solution that can scale and provide daily updates about clickstream activity. Which combination of steps will meet these requirements MOST cost-effectively? (Choose two.) A. Use Amazon Redshift to store and query the clickstream data. B. Use Amazon Athena to query the clickstream data. C. Use Amazon S3 analytics to query the clickstream data. D. Access the query data through a QuickSight direct SQL query. E. Access the query data through QuickSight SPICE (Super-fast, Parallel, In-memory Calculation Engine). Configure a daily refresh for the dataset. Answer: BE

Explanation: Athena would be cheaper than Redshift. S3 analytics is irrelevant. The functionality in SPICE should be more cost effective than direct SQL by reducing the frequency and volume of queries.

QUESTION 113 A data engineer is building a data orchestration workflow. The data engineer plans to use a hybrid model that includes some on-premises resources and some resources that are in the cloud. The data engineer wants to prioritize portability and open source resources. Which service should the data engineer use in both the on-premises environment and the cloud-based environment? A. AWS Data Exchange B. Amazon Simple Workflow Service (Amazon SWF) C. Amazon Managed Workflows for Apache Airflow (Amazon MWAA) D. AWS Glue Answer: C

Explanation: Amazon MWAA is a managed service for Apache Airflow, which is an open-source workflow automation tool. Apache Airflow can be used both on-premises and in the cloud, making it ideal for hybrid environments. Using Amazon MWAA allows the data engineer to leverage the managed service in the cloud while maintaining the ability to use the same open-source Airflow setup on-premises, ensuring portability and consistency across environments.

QUESTION 114 A gaming company uses a NoSQL database to store customer information. The company is planning to migrate to AWS. The company needs a fully managed AWS solution that will handle high online transaction processing (OLTP) workload, provide single-digit millisecond performance, and provide high availability around the world. Which solution will meet these requirements with the LEAST operational overhead? A. Amazon Keyspaces (for Apache Cassandra) B. Amazon DocumentDB (with MongoDB compatibility) C. Amazon DynamoDB D. Amazon Timestream Answer: C

QUESTION 115 A data engineer creates an AWS Lambda function that an Amazon EventBridge event will invoke. When the data engineer tries to invoke the Lambda function by using an EventBridge event, an AccessDeniedException message appears. How should the data engineer resolve the exception? A. Ensure that the trust policy of the Lambda function execution role allows EventBridge to assume the execution role. B. Ensure that both the IAM role that EventBridge uses and the Lambda function's resource-based policy have the necessary permissions. C. Ensure that the subnet where the Lambda function is deployed is configured to be a private subnet. D. Ensure that EventBridge schemas are valid and that the event mapping configuration is correct. Answer: B

Explanation: The lambda resource based policy must allow the events principle to invoke the lambda function. Amazon SQS, Amazon SNS, Lambda, CloudWatch Logs, and EventBridge bus targets do not use roles, and permissions to EventBridge must be granted via a resource policy. <https://docs.aws.amazon.com/eventbridge/latest/userguide/eb-run-lambda-schedule.html#eb-schedule-create-rule>

QUESTION 116 A company uses a data lake that is based on an Amazon S3 bucket. To comply with regulations, the company must apply two layers of server-side encryption to files that are uploaded to the S3 bucket. The company wants to use an AWS Lambda function to apply the necessary encryption. Which solution will meet these requirements? A. Use both server-side encryption with AWS KMS keys (SSE-KMS) and the Amazon S3 Encryption Client. B. Use dual-layer server-side encryption with AWS KMS keys (DSSE-KMS). C. Use server-side encryption with customer-provided keys (SSE-C) before files are uploaded. D. Use server-side encryption with AWS KMS keys (SSE-KMS). Answer: B

Explanation: <https://docs.aws.amazon.com/AmazonS3/latest/userguide/UsingDSSEncryption.html>

QUESTION 117 A data engineer notices that Amazon Athena queries are held in a queue before the queries run. How can the data engineer prevent the queries from queueing? A.

Increase the query result limit.B. Configure provisioned capacity for an existing workgroup.C. Use federated queries.D. Allow users who run the Athena queries to an existing workgroup.Answer: BExplanation:

<https://aws.amazon.com/blogs/aws/introducing-athena-provisioned-capacity/>QUESTION 118A data engineer needs to debug an AWS Glue job that reads from Amazon S3 and writes to Amazon Redshift. The data engineer enabled the bookmark feature for the AWS Glue job.The data engineer has set the maximum concurrency for the AWS Glue job to 1.The AWS Glue job is successfully writing the output to Amazon Redshift. However, the Amazon S3 files that were loaded during previous runs of the AWS Glue job are being reprocessed by subsequent runs.What is the likely reason the AWS Glue job is reprocessing the files?A. The AWS Glue job does not have the s3:GetObjectAcl permission that is required for bookmarks to work correctly.B. The maximum concurrency for the AWS Glue job is set to 1.C. The data engineer incorrectly specified an older version of AWS Glue for the Glue job.D. The AWS Glue job does not have a required commit statement.Answer: DExplanation:

<https://docs.aws.amazon.com/glue/latest/dg/glue-troubleshooting-errors.html#error-job-bookmarks-reprocess-data>QUESTION 119 An ecommerce company wants to use AWS to migrate data pipelines from an on-premises environment into the AWS Cloud. The company currently uses a third-party tool in the on-premises environment to orchestrate data ingestion processes.The company wants a migration solution that does not require the company to manage servers. The solution must be able to orchestrate Python and Bash scripts. The solution must not require the company to refactor any code.Which solution will meet these requirements with the LEAST operational overhead?A. AWS LambdaB. Amazon Managed Workflows for Apache Airflow (Amazon MWAA)C. AWS Step FunctionsD. AWS GlueAnswer: BExplanation:All of the components contained in the outer box (in the image below) appear as a single Amazon MWAA environment in your account. The Apache Airflow Scheduler and Workers are AWS Fargate (Fargate) containers that connect to the private subnets in the Amazon VPC for your environment. Each environment has its own Apache Airflow metadata database managed by AWS that is accessible to the Scheduler and Workers Fargate containers via a privately-secured VPC endpoint.<https://docs.aws.amazon.com/mwaa/latest/userguide/what-is-mwaa.html>QUESTION 120A retail company stores data from a product lifecycle management (PLM) application in an on-premises MySQL database. The PLM application frequently updates the database when transactions occur.The company wants to gather insights from the PLM application in near real time. The company wants to integrate the insights with other business datasets and to analyze the combined dataset by using an Amazon Redshift data warehouse.The company has already established an AWS Direct Connect connection between the on-premises infrastructure and AWS.Which solution will meet these requirements with the LEAST development effort?A. Run a scheduled AWS Glue extract, transform, and load (ETL) job to get the MySQL database updates by using a Java Database Connectivity (JDBC) connection. Set Amazon Redshift as the destination for the ETL job.B. Run a full load plus CDC task in AWS Database Migration Service (AWS DMS) to continuously replicate the MySQL database changes. Set Amazon Redshift as the destination for the task.C. Use the Amazon AppFlow SDK to build a custom connector for the MySQL database to continuously replicate the database changes. Set Amazon Redshift as the destination for the connector.D. Run scheduled AWS DataSync tasks to synchronize data from the MySQL database. Set Amazon Redshift as the destination for the tasks.Answer: B

Explanation:<https://aws.amazon.com/ko/blogs/apn/change-data-capture-from-on-premises-sql-server-to-amazon-redshift-target/>QUESTION 121A marketing company uses Amazon S3 to store clickstream data. The company queries the data at the end of each day by using a SQL JOIN clause on S3 objects that are stored in separate buckets.The company creates key performance indicators (KPIs) based on the objects. The company needs a serverless solution that will give users the ability to query data by partitioning the data. The solution must maintain the atomicity, consistency, isolation, and durability (ACID) properties of the data.Which solution will meet these requirements MOST cost-effectively?A. Amazon S3 SelectB. Amazon Redshift SpectrumC. Amazon AthenaD. Amazon EMRAnswer: CQUESTION 122A company wants to migrate data from an Amazon RDS for PostgreSQL DB instance in the eu-east-1 Region of an AWS account named Account\_A. The company will migrate the data to an Amazon Redshift cluster in the eu-west-1 Region of an AWS account named Account\_B.Which solution will give AWS Database Migration Service (AWS DMS) the ability to replicate data between two data stores?A. Set up an AWS DMS replication instance in Account\_B in eu-west-1.B. Set up an AWS DMS replication instance in Account\_B in eu-east-1.C. Set up an AWS DMS replication instance in a new AWS account in eu-west-1.D. Set up an AWS DMS replication instance in Account\_A in eu-east-1.Answer: B

Explanation:[https://docs.aws.amazon.com/dms/latest/userguide/CHAP\\_Target.Redshift.html#CHAP\\_Target.Redshift.Prerequisites](https://docs.aws.amazon.com/dms/latest/userguide/CHAP_Target.Redshift.html#CHAP_Target.Redshift.Prerequisites)QUESTION 123A company uses Amazon S3 as a data lake. The company sets up a data warehouse by using a multi-node Amazon Redshift cluster. The company organizes the data files in the data lake based on the data source of each data file.The company loads all the data files into one table in the Redshift cluster by using a separate COPY command for each data file location. This approach takes a long time to load all the data files into the table. The company must increase the speed of the data ingestion. The company does not want to increase the cost of the process.Which solution will meet these requirements?A. Use a provisioned Amazon EMR

cluster to copy all the data files into one folder. Use a COPY command to load the data into Amazon Redshift.B. Load all the data files in parallel into Amazon Aurora. Run an AWS Glue job to load the data into Amazon Redshift.C. Use an AWS Glue job to copy all the data files into one folder. Use a COPY command to load the data into Amazon Redshift.D. Create a manifest file that contains the data file locations. Use a COPY command to load the data into Amazon Redshift.Answer: DExplanation:

[https://docs.aws.amazon.com/redshift/latest/dg/c\\_best-practices-single-copy-command.html](https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-single-copy-command.html)

<https://docs.aws.amazon.com/redshift/latest/dg/loading-data-files-using-manifest.html>QUESTION 124A company plans to use Amazon Kinesis Data Firehose to store data in Amazon S3. The source data consists of 2 MB .csv files. The company must convert the .csv files to JSON format. The company must store the files in Apache Parquet format.Which solution will meet these requirements with the LEAST development effort?A. Use Kinesis Data Firehose to convert the .csv files to JSON. Use an AWS Lambda function to store the files in Parquet format.B. Use Kinesis Data Firehose to convert the .csv files to JSON and to store the files in Parquet format.C. Use Kinesis Data Firehose to invoke an AWS Lambda function that transforms the .csv files to JSON and stores the files in Parquet format.D. Use Kinesis Data Firehose to invoke an AWS Lambda function that transforms the .csv files to JSON. Use Kinesis Data Firehose to store the files in Parquet format.Answer: BExplanation:By using the built-in transformation and format conversion features of Kinesis Data Firehose, you achieve the desired result with minimal custom development, thereby meeting the requirements efficiently and cost-effectively.QUESTION 125A company is using an AWS Transfer Family server to migrate data from an on-premises environment to AWS. Company policy mandates the use of TLS 1.2 or above to encrypt the data in transit.Which solution will meet these requirements?A. Generate new SSH keys for the Transfer Family server. Make the old keys and the new keys available for use.B. Update the security group rules for the on-premises network to allow only connections that use TLS 1.2 or above.C. Update the security policy of the Transfer Family server to specify a minimum protocol version of TLS 1.2.D. Install an SSL certificate on the Transfer Family server to encrypt data transfers by using TLS 1.2.Answer: CExplanation:

<https://docs.aws.amazon.com/transfer/latest/userguide/security-policies.html>QUESTION 126

A company wants to migrate an application and an on-premises Apache Kafka server to AWS. The application processes incremental updates that an on-premises Oracle database sends to the Kafka server. The company wants to use the replatform migration strategy instead of the refactor strategy.Which solution will meet these requirements with the LEAST management overhead?A. Amazon Kinesis Data StreamsB. Amazon Managed Streaming for Apache Kafka (Amazon MSK) provisioned clusterC. Amazon Kinesis Data FirehoseD. Amazon Managed Streaming for Apache Kafka (Amazon MSK) ServerlessAnswer: D

QUESTION 127A data engineer is building an automated extract, transform, and load (ETL) ingestion pipeline by using AWS Glue. The pipeline ingests compressed files that are in an Amazon S3 bucket. The ingestion pipeline must support incremental data processing.Which AWS Glue feature should the data engineer use to meet this requirement?A. WorkflowsB. TriggersC. Job bookmarksD. ClassifiersAnswer: CExplanation:

<https://docs.aws.amazon.com/glue/latest/dg/monitor-continuations.html>

QUESTION 128A banking company uses an application to collect large volumes of transactional data. The company uses Amazon Kinesis Data Streams for real-time analytics. The company's application uses the PutRecord action to send data to Kinesis Data Streams.A data engineer has observed network outages during certain times of day. The data engineer wants to configure exactly-once delivery for the entire processing pipeline.Which solution will meet this requirement?A. Design the application so it can remove duplicates during processing by embedding a unique ID in each record at the source.B. Update the checkpoint configuration of the Amazon Managed Service for Apache Flink (previously known as Amazon Kinesis Data Analytics) data collection application to avoid duplicate processing of events.C. Design the data source so events are not ingested into Kinesis Data Streams multiple times.D. Stop using Kinesis Data Streams. Use Amazon EMR instead. Use Apache Flink and Apache Spark Streaming in Amazon EMR.Answer: A

QUESTION 129A company stores logs in an Amazon S3 bucket. When a data engineer attempts to access several log files, the data engineer discovers that some files have been unintentionally deleted.The data engineer needs a solution that will prevent unintentional file deletion in the future.Which solution will meet this requirement with the LEAST operational overhead?A. Manually back up the S3 bucket on a regular basis.B. Enable S3 Versioning for the S3 bucket.C. Configure replication for the S3 bucket.D. Use an Amazon S3 Glacier storage class to archive the data that is in the S3 bucket.Answer: B

QUESTION 130A telecommunications company collects network usage data throughout each day at a rate of several thousand data points each second. The company runs an application to process the usage data in real time. The company aggregates and stores the data in an Amazon Aurora DB instance.Sudden drops in network usage usually indicate a network outage. The company must be able to identify sudden drops in network usage so the company can take immediate remedial actions.Which solution will meet this requirement with the LEAST latency?A. Create an AWS Lambda function to query Aurora for drops in network usage. Use Amazon EventBridge to automatically invoke the Lambda function every minute.B. Modify the processing application to publish the data to an Amazon Kinesis data stream. Create an Amazon Managed Service for Apache Flink (previously



known as Amazon Kinesis Data Analytics) application to detect drops in network usage.C. Replace the Aurora database with an Amazon DynamoDB table. Create an AWS Lambda function to query the DynamoDB table for drops in network usage every minute. Use DynamoDB Accelerator (DAX) between the processing application and DynamoDB table.D. Create an AWS Lambda function within the Database Activity Streams feature of Aurora to detect drops in network usage.

Answer: B

QUESTION 131A data engineer is processing and analyzing multiple terabytes of raw data that is in Amazon S3. The data engineer needs to clean and prepare the data. Then the data engineer needs to load the data into Amazon Redshift for analytics. The data engineer needs a solution that will give data analysts the ability to perform complex queries. The solution must eliminate the need to perform complex extract, transform, and load (ETL) processes or to manage infrastructure. Which solution will meet these requirements with the LEAST operational overhead?

A. Use Amazon EMR to prepare the data. Use AWS Step Functions to load the data into Amazon Redshift. Use Amazon QuickSight to run queries.

B. Use AWS Glue DataBrew to prepare the data. Use AWS Glue to load the data into Amazon Redshift. Use Amazon Redshift to run queries.

C. Use AWS Lambda to prepare the data. Use Amazon Kinesis Data Firehose to load the data into Amazon Redshift. Use Amazon Athena to run queries.

D. Use AWS Glue to prepare the data. Use AWS Database Migration Service (AWS DMS) to load the data into Amazon Redshift. Use Amazon Redshift Spectrum to run queries.

Answer: B

QUESTION 132A company uses an AWS Lambda function to transfer files from a legacy SFTP environment to Amazon S3 buckets. The Lambda function is VPC enabled to ensure that all communications between the Lambda function and other AWS services that are in the same VPC environment will occur over a secure network. The Lambda function is able to connect to the SFTP environment successfully. However, when the Lambda function attempts to upload files to the S3 buckets, the Lambda function returns timeout errors. A data engineer must resolve the timeout issues in a secure way. Which solution will meet these requirements in the MOST cost-effective way?

A. Create a NAT gateway in the public subnet of the VPC. Route network traffic to the NAT gateway.

B. Create a VPC gateway endpoint for Amazon S3. Route network traffic to the VPC gateway endpoint.

C. Create a VPC interface endpoint for Amazon S3. Route network traffic to the VPC interface endpoint.

D. Use a VPC internet gateway to connect to the internet. Route network traffic to the VPC internet gateway.

Answer: B

Explanation: While interface endpoints is a viable solution, it can be more complex and expensive compared to a gateway endpoint. VPC interface endpoints charge per hour and per gigabyte of data transferred.

QUESTION 133A company reads data from customer databases that run on Amazon RDS. The databases contain many inconsistent fields. For example, a customer record field that is named place\_id in one database is named location\_id in another database. The company needs to link customer records across different databases, even when customer record fields do not match. Which solution will meet these requirements with the LEAST operational overhead?

A. Create a provisioned Amazon EMR cluster to process and analyze data in the databases. Connect to the Apache Zeppelin notebook. 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 the performance and results.

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

D. Create a provisioned Amazon EMR cluster to process and analyze data in the databases. Connect to the Apache Zeppelin notebook. Use an Apache Spark ML model to find duplicate records in the data. Evaluate and tune the model by evaluating the performance and results.

Answer: B

QUESTION 134A finance company receives data from third-party data providers and stores the data as objects in an Amazon S3 bucket. The company ran an AWS Glue crawler on the objects to create a data catalog. The AWS Glue crawler created multiple tables. However, the company expected that the crawler would create only one table. The company needs a solution that will ensure the AWS Glue crawler creates only one table. Which combination of solutions will meet this requirement? (Choose two.)

A. Ensure that the object format, compression type, and schema are the same for each object.

B. Ensure that the object format and schema are the same for each object. Do not enforce consistency for the compression type of each object.

C. Ensure that the schema is the same for each object. Do not enforce consistency for the file format and compression type of each object.

D. Ensure that the structure of the prefix for each S3 object name is consistent.

E. Ensure that all S3 object names follow a similar pattern.

Answer: AD

QUESTION 135An application consumes messages from an Amazon Simple Queue Service (Amazon SQS) queue. The application experiences occasional downtime. As a result of the downtime, messages within the queue expire and are deleted after 1 day. The message deletions cause data loss for the application. Which solutions will minimize data loss for the application? (Choose two.)

A. Increase the message retention period.

B. Increase the visibility timeout.

C. Attach a dead-letter queue (DLQ) to the SQS queue.

D. Use a delay queue to delay message delivery.

E. Reduce message processing time.

Answer: AC

QUESTION 136A company is creating near real-time dashboards to visualize time series data. The company ingests data into Amazon Managed Streaming for Apache Kafka (Amazon MSK). A customized data pipeline consumes the data. The pipeline then writes data to Amazon Keyspaces (for Apache Cassandra), Amazon OpenSearch Service, and Apache Avro objects in Amazon S3. Which solution will make the data available for the data visualizations with the

LEAST latency?A. Create OpenSearch Dashboards by using the data from OpenSearch Service.B. Use Amazon Athena with an Apache Hive metastore to query the Avro objects in Amazon S3. Use Amazon Managed Grafana to connect to Athena and to create the dashboards.C. Use Amazon Athena to query the data from the Avro objects in Amazon S3. Configure Amazon Keyspaces as the data catalog. Connect Amazon QuickSight to Athena to create the dashboards.D. Use AWS Glue to catalog the data. Use S3 Select to query the Avro objects in Amazon S3. Connect Amazon QuickSight to the S3 bucket to create the dashboards.

Answer: A

QUESTION 137A data engineer maintains a materialized view that is based on an Amazon Redshift database. The view has a column named load\_date that stores the date when each row was loaded. The data engineer needs to reclaim database storage space by deleting all the rows from the materialized view. Which command will reclaim the MOST database storage space?A. DELETE FROM materialized\_view\_name where 1=1B. TRUNCATE materialized\_view\_nameC. VACUUM table\_name where load\_date<=current\_date materializedviewD. DELETE FROM materialized\_view\_name where load\_date<=current\_date

Answer: B

QUESTION 138A media company wants to use Amazon OpenSearch Service to analyze real-time data about popular musical artists and songs. The company expects to ingest millions of new data events every day. The new data events will arrive through an Amazon Kinesis data stream. The company must transform the data and then ingest the data into the OpenSearch Service domain. Which method should the company use to ingest the data with the LEAST operational overhead?A. Use Amazon Kinesis Data Firehose and an AWS Lambda function to transform the data and deliver the transformed data to OpenSearch Service.B. Use a Logstash pipeline that has prebuilt filters to transform the data and deliver the transformed data to OpenSearch Service.C. Use an AWS Lambda function to call the Amazon Kinesis Agent to transform the data and deliver the transformed data OpenSearch Service.D. Use the Kinesis Client Library (KCL) to transform the data and deliver the transformed data to OpenSearch Service.

Answer: A

Explanation: Amazon Kinesis Data Firehose is a fully managed service that reliably loads streaming data into data lakes, data stores and analytics services like OpenSearch Service. It can automatically scale to match the throughput of your data and requires no ongoing administration.

QUESTION 139A company stores customer data tables that include customer addresses in an AWS Lake Formation data lake. To comply with new regulations, the company must ensure that users cannot access data for customers who are in Canada. The company needs a solution that will prevent user access to rows for customers who are in Canada. Which solution will meet this requirement with the LEAST operational effort?A. Set a row-level filter to prevent user access to a row where the country is Canada.B. Create an IAM role that restricts user access to an address where the country is Canada.C. Set a column-level filter to prevent user access to a row where the country is Canada.D. Apply a tag to all rows where Canada is the country. Prevent user access where the tag is equal to "Canada".

Answer: A

QUESTION 140A company has implemented a lake house architecture in Amazon Redshift. The company needs to give users the ability to authenticate into Redshift query editor by using a third-party identity provider (IdP). A data engineer must set up the authentication mechanism. What is the first step the data engineer should take to meet this requirement?A. Register the third-party IdP as an identity provider in the configuration settings of the Redshift cluster.B. Register the third-party IdP as an identity provider from within Amazon Redshift.C. Register the third-party IdP as an identity provider for AWS Secrets Manager. Configure Amazon Redshift to use Secrets Manager to manage user credentials.D. Register the third-party IdP as an identity provider for AWS Certificate Manager (ACM). Configure Amazon Redshift to use ACM to manage user credentials.

Answer: A

Explanation: To enable users to authenticate into the Amazon Redshift query editor using a third-party identity provider (IdP), the data engineer must first register that IdP within the configuration settings of the Redshift cluster itself. Amazon Redshift natively supports integrating with external identity providers to manage user authentication. By registering the third-party IdP directly in the Redshift cluster settings, it establishes the trust relationship needed for Redshift to rely on that IdP for authenticating users when they log into the query editor.

QUESTION 141A company currently uses a provisioned Amazon EMR cluster that includes general purpose Amazon EC2 instances. The EMR cluster uses EMR managed scaling between one to five task nodes for the company's long-running Apache Spark extract, transform, and load (ETL) job. The company runs the ETL job every day. When the company runs the ETL job, the EMR cluster quickly scales up to five nodes. The EMR cluster often reaches maximum CPU usage, but the memory usage remains under 30%. The company wants to modify the EMR cluster configuration to reduce the EMR costs to run the daily ETL job. Which solution will meet these requirements MOST cost-effectively?A. Increase the maximum number of task nodes for EMR managed scaling to 10.B. Change the task node type from general purpose EC2 instances to memory optimized EC2 instances.C. Switch the task node type from general purpose Re instances to compute optimized EC2 instances.D. Reduce the scaling cooldown period for the provisioned EMR cluster.

Answer: C

Explanation: Since the ETL job reaches maximum CPU usage but not memory usage, switching from general-purpose instances to compute-optimized instances (such as C5 or C6g instances) can provide better performance per dollar for CPU-bound workloads.

QUESTION 142A company uploads .csv files to an Amazon S3 bucket. The company's data platform team has set up an AWS Glue crawler to perform data discovery and to create the tables and schemas. An AWS Glue job

writes processed data from the tables to an Amazon Redshift database. The AWS Glue job handles column mapping and creates the Amazon Redshift tables in the Redshift database appropriately. If the company reruns the AWS Glue job for any reason, duplicate records are introduced into the Amazon Redshift tables. The company needs a solution that will update the Redshift tables without duplicates. Which solution will meet these requirements?

A. Modify the AWS Glue job to copy the rows into a staging Redshift table. Add SQL commands to update the existing rows with new values from the staging Redshift table.

B. Modify the AWS Glue job to load the previously inserted data into a MySQL database. Perform an upsert operation in the MySQL database. Copy the results to the Amazon Redshift tables.

C. Use Apache Spark's DataFrame dropDuplicates() API to eliminate duplicates. Write the data to the Redshift tables.

D. Use the AWS Glue ResolveChoice built-in transform to select the value of the column from the most recent record.

Answer: A

Explanation: Two step approach involving creating a staging table, followed by using Redshift's merge statement to update the target table from staging table and finally truncate/housekeep the staging table.

QUESTION 143

A company is using Amazon Redshift to build a data warehouse solution. The company is loading hundreds of files into a fact table that is in a Redshift cluster. The company wants the data warehouse solution to achieve the greatest possible throughput. The solution must use cluster resources optimally when the company loads data into the fact table. Which solution will meet these requirements?

A. Use multiple COPY commands to load the data into the Redshift cluster.

B. Use S3DistCp to load multiple files into Hadoop Distributed File System (HDFS). Use an HDFS connector to ingest the data into the Redshift cluster.

C. Use a number of INSERT statements equal to the number of Redshift cluster nodes. Load the data in parallel into each node.

D. Use a single COPY command to load the data into the Redshift cluster.

Answer: D

Explanation:

[https://docs.aws.amazon.com/redshift/latest/dg/t\\_Loading-data-from-S3.html](https://docs.aws.amazon.com/redshift/latest/dg/t_Loading-data-from-S3.html)

QUESTION 144

A company ingests data from multiple data sources and stores the data in an Amazon S3 bucket. An AWS Glue extract, transform, and load (ETL) job transforms the data and writes the transformed data to an Amazon S3 based data lake. The company uses Amazon Athena to query the data that is in the data lake. The company needs to identify matching records even when the records do not have a common unique identifier. Which solution will meet this requirement?

A. Use Amazon Macie pattern matching as part of the ETL job.

B. Train and use the AWS Glue PySpark Filter class in the ETL job.

C. Partition tables and use the ETL job to partition the data on a unique identifier.

D. Train and use the AWS Lake Formation FindMatches transform in the ETL job.

Answer: D

Explanation: AWS Lake Formation provides machine learning capabilities to create custom transforms to cleanse your data. There is currently one available transform named FindMatches. The FindMatches transform enables you to identify duplicate or matching records in your dataset, even when the records do not have a common unique identifier and no fields match exactly. This will not require writing any code or knowing how machine learning works.

QUESTION 145

A data engineer is using an AWS Glue crawler to catalog data that is in an Amazon S3 bucket. The S3 bucket contains both .csv and .json files. The data engineer configured the crawler to exclude the .json files from the catalog. When the data engineer runs queries in Amazon Athena, the queries also process the excluded .json files. The data engineer wants to resolve this issue. The data engineer needs a solution that will not affect access requirements for the .csv files in the source S3 bucket. Which solution will meet this requirement with the SHORTEST query times?

A. Adjust the AWS Glue crawler settings to ensure that the AWS Glue crawler also excludes .json files.

B. Use the Athena console to ensure the Athena queries also exclude the .json files.

C. Relocate the .json files to a different path within the S3 bucket.

D. Use S3 bucket policies to block access to the .json files.

Answer: C

Explanation: Athena does not recognize exclude patterns that you specify in an AWS Glue crawler. For example, if you have an Amazon S3 bucket that contains both .csv and .json files and you exclude the .json files from the crawler, Athena queries both groups of files. To avoid this, place the files that you want to exclude in a different location.

<https://docs.aws.amazon.com/athena/latest/ug/troubleshooting-athena.html>

QUESTION 146

A data engineer set up an AWS Lambda function to read an object that is stored in an Amazon S3 bucket. The object is encrypted by an AWS KMS key. The data engineer configured the Lambda function's execution role to access the S3 bucket. However, the Lambda function encountered an error and failed to retrieve the content of the object. What is the likely cause of the error?

A. The data engineer misconfigured the permissions of the S3 bucket. The Lambda function could not access the object.

B. The Lambda function is using an outdated SDK version, which caused the read failure.

C. The S3 bucket is located in a different AWS Region than the Region where the data engineer works. Latency issues caused the Lambda function to encounter an error.

D. The Lambda function's execution role does not have the necessary permissions to access the KMS key that can decrypt the S3 object.

Answer: D

QUESTION 147

A data engineer has implemented data quality rules in 1,000 AWS Glue Data Catalog tables. Because of a recent change in business requirements, the data engineer must edit the data quality rules. How should the data engineer meet this requirement with the LEAST operational overhead?

A. Create a pipeline in AWS Glue ETL to edit the rules for each of the 1,000 Data Catalog tables. Use an AWS Lambda function to call the corresponding AWS Glue job for each Data Catalog table.

B. Create an AWS Lambda function that makes an API call to AWS Glue Data Quality to make the edits.

C. Create an Amazon EMR cluster. Run a pipeline on Amazon EMR that

edits the rules for each Data Catalog table. Use an AWS Lambda function to run the EMR pipeline.D. Use the AWS Management Console to edit the rules within the Data Catalog.

Answer: B

QUESTION 148 Two developers are working on separate application releases. The developers have created feature branches named Branch A and Branch B by using a GitHub repository's master branch as the source. The developer for Branch A deployed code to the production system. The code for Branch B will merge into a master branch in the following week's scheduled application release. Which command should the developer for Branch B run before the developer raises a pull request to the master branch?

A. `git diff branchB master`  
B. `git commit -mB`  
C. `git pull master`  
D. `git fetch -b master`

Answer: C

Explanation: Rebasing In Git, there are two main ways to integrate changes from one branch into another: the merge and the rebase. In this section you'll learn what rebasing is, how to do it, why it's a pretty amazing tool, and in what cases you won't want to use it.

The Basic Rebase If you go back to an earlier example from Basic Merging, you can see that you diverged your work and made commits on two different branches.

QUESTION 149 A company stores employee data in Amazon Redshift. A table named Employee uses columns named Region ID, Department ID, and Role ID as a compound sort key. Which queries will MOST increase the speed of query by using a compound sort key of the table? (Choose two.)

A. `Select *from Employee where Region ID='North America';`  
B. `Select *from Employee where Region ID='North America' and Department ID=20;`  
C. `Select *from Employee where Department ID=20 and Region ID='North America';`  
D. `Select *from Employee where Role ID=50;`  
E. `Select *from Employee where Region ID='North America' and Role ID=50;`

Answer: B E

Explanation: To maximize the speed of queries by using the compound sort key (Region ID, Department ID, and Role ID) in the Employee table in Amazon Redshift, the queries should align with the order of the columns in the sort key.

QUESTION 150 A company receives test results from testing facilities that are located around the world. The company stores the test results in millions of 1 KB JSON files in an Amazon S3 bucket. A data engineer needs to process the files, convert them into Apache Parquet format, and load them into Amazon Redshift tables. The data engineer uses AWS Glue to process the files, AWS Step Functions to orchestrate the processes, and Amazon EventBridge to schedule jobs. The company recently added more testing facilities. The time required to process files is increasing. The data engineer must reduce the data processing time. Which solution will MOST reduce the data processing time?

A. Use AWS Lambda to group the raw input files into larger files. Write the larger files back to Amazon S3. Use AWS Glue to process the files. Load the files into the Amazon Redshift tables.  
B. Use the AWS Glue dynamic frame file-grouping option to ingest the raw input files. Process the files. Load the files into the Amazon Redshift tables.  
C. Use the Amazon Redshift COPY command to move the raw input files from Amazon S3 directly into the Amazon Redshift tables. Process the files in Amazon Redshift.  
D. Use Amazon EMR instead of AWS Glue to group the raw input files. Process the files in Amazon EMR. Load the files into the Amazon Redshift tables.

Answer: B

QUESTION 151 A data engineer uses Amazon Managed Workflows for Apache Airflow (Amazon MWAA) to run data pipelines in an AWS account. A workflow recently failed to run. The data engineer needs to use Apache Airflow logs to diagnose the failure of the workflow. Which log type should the data engineer use to diagnose the cause of the failure?

A. `YourEnvironmentName-WebServer`  
B. `YourEnvironmentName-Scheduler`  
C. `YourEnvironmentName-DAGProcessing`  
D. `YourEnvironmentName-Task`

Answer: D

Explanation: When a workflow fails to run in Amazon MWAA, the task logs (`YourEnvironmentName-Task`) are the most relevant for diagnosing the issue. Task logs contain detailed information about the execution of individual tasks within the workflow, including any error messages or stack traces that can help pinpoint the cause of the failure.

QUESTION 152 A finance company uses Amazon Redshift as a data warehouse. The company stores the data in a shared Amazon S3 bucket. The company uses Amazon Redshift Spectrum to access the data that is stored in the S3 bucket. The data comes from certified third-party data providers. Each third-party data provider has unique connection details. To comply with regulations, the company must ensure that none of the data is accessible from outside the company's AWS environment. Which combination of steps should the company take to meet these requirements? (Choose two.)

A. Replace the existing Redshift cluster with a new Redshift cluster that is in a private subnet. Use an interface VPC endpoint to connect to the Redshift cluster. Use a NAT gateway to give Redshift access to the S3 bucket.  
B. Create an AWS CloudHSM hardware security module (HSM) for each data provider. Encrypt each data provider's data by using the corresponding HSM for each data provider.  
C. Turn on enhanced VPC routing for the Amazon Redshift cluster. Set up an AWS Direct Connect connection and configure a connection between each data provider and the finance company's VPC.  
D. Define table constraints for the primary keys and the foreign keys.  
E. Use federated queries to access the data from each data provider. Do not upload the data to the S3 bucket. Perform the federated queries through a gateway VPC endpoint.

Answer: A C

QUESTION 153 Files from multiple data sources arrive in an Amazon S3 bucket on a regular basis. A data engineer wants to ingest new files into Amazon Redshift in near real time when the new files arrive in the S3 bucket. Which solution will meet these requirements?

A. Use the query editor v2 to schedule a COPY command to load new files into Amazon Redshift.  
B. Use the zero-ETL integration between Amazon Aurora and Amazon Redshift to load new files into Amazon Redshift.  
C. Use AWS Glue job bookmarks to extract, transform, and load (ETL)



load new files into Amazon Redshift.D. Use S3 Event Notifications to invoke an AWS Lambda function that loads new files into Amazon Redshift.  
Answer: D  
QUESTION 154A technology company currently uses Amazon Kinesis Data Streams to collect log data in real time. The company wants to use Amazon Redshift for downstream real-time queries and to enrich the log data. Which solution will ingest data into Amazon Redshift with the LEAST operational overhead?  
A. Set up an Amazon Kinesis Data Firehose delivery stream to send data to a Redshift provisioned cluster table.  
B. Set up an Amazon Kinesis Data Firehose delivery stream to send data to Amazon S3. Configure a Redshift provisioned cluster to load data every minute.  
C. Configure Amazon Managed Service for Apache Flink (previously known as Amazon Kinesis Data Analytics) to send data directly to a Redshift provisioned cluster table.  
D. Use Amazon Redshift streaming ingestion from Kinesis Data Streams and to present data as a materialized view.  
Answer: D  
Explanation: Amazon Redshift supports streaming ingestion from Amazon Kinesis Data Streams. The Amazon Redshift streaming ingestion feature provides low-latency, high-speed ingestion of streaming data from Amazon Kinesis Data Streams into an Amazon Redshift materialized view. Amazon Redshift streaming ingestion removes the need to stage data in Amazon S3 before ingesting into Amazon Redshift.

<https://docs.aws.amazon.com/streams/latest/dev/using-other-services-redshift.html>  
QUESTION 155A company maintains a data warehouse in an on-premises Oracle database. The company wants to build a data lake on AWS. The company wants to load data warehouse tables into Amazon S3 and synchronize the tables with incremental data that arrives from the data warehouse every day. Each table has a column that contains monotonically increasing values. The size of each table is less than 50 GB. The data warehouse tables are refreshed every night between 1 AM and 2 AM. A business intelligence team queries the tables between 10 AM and 8 PM every day. Which solution will meet these requirements in the MOST operationally efficient way?  
A. Use an AWS Database Migration Service (AWS DMS) full load plus CDC job to load tables that contain monotonically increasing data columns from the on-premises data warehouse to Amazon S3. Use custom logic in AWS Glue to append the daily incremental data to a full-load copy that is in Amazon S3.  
B. Use an AWS Glue Java Database Connectivity (JDBC) connection. Configure a job bookmark for a column that contains monotonically increasing values. Write custom logic to append the daily incremental data to a full-load copy that is in Amazon S3.  
C. Use an AWS Database Migration Service (AWS DMS) full load migration to load the data warehouse tables into Amazon S3 every day. Overwrite the previous day's full-load copy every day.  
D. Use AWS Glue to load a full copy of the data warehouse tables into Amazon S3 every day. Overwrite the previous day's full-load copy every day.  
Answer: A  
Resources From: 1. 2025 Latest Braindump2go DEA-C01 Exam Dumps (PDF & VCE) Free Share:

<https://www.braindump2go.com/dea-c01.html>  
2. 2025 Latest Braindump2go DEA-C01 PDF and DEA-C01 VCE Dumps Free Share: <https://drive.google.com/drive/folders/1mDLAMRdq63BtyYPIHBu3zL24ZVenG2kg?usp=sharing>

3. 2025 Free Braindump2go DEA-C01 Exam Questions Download:

[https://www.braindump2go.com/free-online-pdf/DEA-C01-VCE-Dumps\(105-155\).pdf](https://www.braindump2go.com/free-online-pdf/DEA-C01-VCE-Dumps(105-155).pdf)  
Free Resources from Braindump2go, We Devoted to Helping You 100% Pass All Exams!