



# DAS-C01<sup>Q&As</sup>

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.passapply.com/das-c01.html>

100% Passing Guarantee  
100% Money Back Assurance

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

-  **Instant Download** After Purchase
-  **100% Money Back** Guarantee
-  **365 Days** Free Update
-  **800,000+** Satisfied Customers





### QUESTION 1

A company is migrating from an on-premises Apache Hadoop cluster to an Amazon EMR cluster. The cluster runs only during business hours. Due to a company requirement to avoid intraday cluster failures, the EMR cluster must be highly available. When the cluster is terminated at the end of each business day, the data must persist.

Which configurations would enable the EMR cluster to meet these requirements? (Choose three.)

- A. EMR File System (EMRFS) for storage
- B. Hadoop Distributed File System (HDFS) for storage
- C. AWS Glue Data Catalog as the metastore for Apache Hive
- D. MySQL database on the master node as the metastore for Apache Hive
- E. Multiple master nodes in a single Availability Zone
- F. Multiple master nodes in multiple Availability Zones

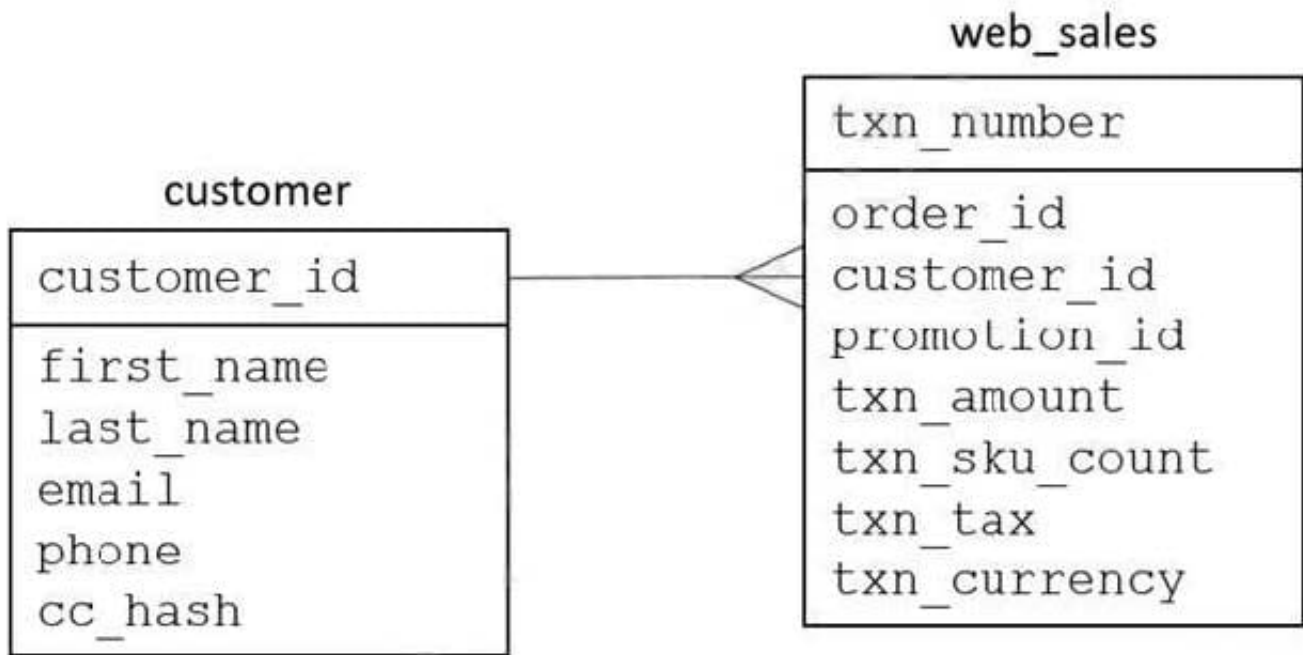
Correct Answer: BCF

---

### QUESTION 2

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

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



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

- A. In AWS Lake Formation, grant the business\_analyst group SELECT and ALTER permissions for the web\_sales table.
- B. In AWS Lake Formation, grant the business\_analyst group the SELECT permission for the web\_sales table.
- C. In AWS Lake Formation, grant the business\_analyst group the SELECT permission for the customer table. Under columns, choose filter type "Include columns" with columns first\_name, last\_name, and customer\_id.
- D. In AWS Lake Formation, grant the business\_analyst group SELECT and ALTER permissions for the customer table. Under columns, choose filter type "Include columns" with columns first\_name and last\_name.
- E. Create users under a business\_analyst IAM group. Create a policy that allows the lakeformation:GetDataAccess action, the athena:\* action, and the glue:Get\* action.
- F. Create users under a business\_analyst IAM group. Create a policy that allows the lakeformation:GetDataAccess action, the athena:\* action, and the glue:Get\* action. In addition, allow the s3:GetObject action, the s3:PutObject action, and the s3:GetBucketLocation action for the Athena query results S3 bucket.

Correct Answer: BCF

Reference: <https://docs.aws.amazon.com/athena/latest/ug/querying.html>

### QUESTION 3

A company owns facilities with IoT devices installed across the world. The company is using Amazon Kinesis Data Streams to stream data from the devices to Amazon S3. The company's operations team wants to get insights from the IoT data to monitor data quality at ingestion. The insights need to be derived in near-real time, and the output must be logged to Amazon DynamoDB for further analysis.



Which solution meets these requirements?

- A. Connect Amazon Kinesis Data Analytics to analyze the stream data. Save the output to DynamoDB by using the default output from Kinesis Data Analytics.
- B. Connect Amazon Kinesis Data Analytics to analyze the stream data. Save the output to DynamoDB by using an AWS Lambda function.
- C. Connect Amazon Kinesis Data Firehose to analyze the stream data by using an AWS Lambda function. Save the output to DynamoDB by using the default output from Kinesis Data Firehose.
- D. Connect Amazon Kinesis Data Firehose to analyze the stream data by using an AWS Lambda function. Save the data to Amazon S3. Then run an AWS Glue job on schedule to ingest the data into DynamoDB.

Correct Answer: C

---

#### QUESTION 4

A company uses Amazon Redshift for its data warehouse. The company is running an ETL process that receives data in data parts from five third-party providers. The data parts contain independent records that are related to one specific job.

The company receives the data parts at various times throughout each day.

A data analytics specialist must implement a solution that loads the data into Amazon Redshift only after the company receives all five data parts.

Which solution will meet these requirements?

- A. Create an Amazon S3 bucket to receive the data. Use S3 multipart upload to collect the data from the different sources and to form a single object before loading the data into Amazon Redshift.
- B. Use an AWS Lambda function that is scheduled by cron to load the data into a temporary table in Amazon Redshift. Use Amazon Redshift database triggers to consolidate the final data when all five data parts are ready.
- C. Create an Amazon S3 bucket to receive the data. Create an AWS Lambda function that is invoked by S3 upload events. Configure the function to validate that all five data parts are gathered before the function loads the data into Amazon Redshift.
- D. Create an Amazon Kinesis Data Firehose delivery stream. Program a Python condition that will invoke a buffer flush when all five data parts are received.

Correct Answer: C

---

#### QUESTION 5

A company uses an Amazon EMR cluster with 50 nodes to process operational data and make the data available for data analysts. These jobs run nightly use Apache Hive with the Apache Jez framework as a processing model and write results to Hadoop Distributed File System (HDFS). In the last few weeks, jobs are failing and are producing the following error message:

"File could only be replicated to 0 nodes instead of 1".



A data analytics specialist checks the DataNode logs the NameNode logs and network connectivity for potential issues that could have prevented HDFS from replicating data. The data analytics specialist rules out these factors as causes for the issue.

Which solution will prevent the jobs from failing\?

- A. Monitor the HDFSUtilization metric. If the value crosses a user-defined threshold add task nodes to the EMR cluster
- B. Monitor the HDFSUtilization metric. If the value crosses a user-defined threshold add core nodes to the EMR cluster
- C. Monitor the MemoryAllocatedMB metric. If the value crosses a user-defined threshold, add task nodes to the EMR cluster
- D. Monitor the MemoryAllocatedMB metric. If the value crosses a user-defined threshold, add core nodes to the EMR cluster.

Correct Answer: C

[DAS-C01 PDF Dumps](#)

[DAS-C01 Study Guide](#)

[DAS-C01 Exam Questions](#)