



# 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 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. 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. Use AWS Lambda to publish an Amazon Simple Notification Service (Amazon SNS) notification if the threshold is not met.
- B. Use Amazon Kinesis Data Streams to collect all the data from toll stations. 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. Use AWS Lambda to publish an Amazon Simple Notification Service (Amazon SNS) notification if the threshold is not met. Connect Amazon Kinesis Data Firehose to Kinesis Data Streams to deliver the data to Amazon Redshift.
- C. Use Amazon Kinesis Data Firehose to collect data and deliver it to Amazon Redshift. 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.
- D. Use Amazon Kinesis Data Firehose to collect data and deliver it to Amazon Redshift and Amazon Kinesis Data Analytics simultaneously. 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.

Correct Answer: D

---

### QUESTION 2

A real estate company has a mission-critical application using Apache HBase in Amazon EMR. Amazon EMR is configured with a single master node. The company has over 5 TB of data stored on an Hadoop Distributed File System (HDFS). The company wants a cost-effective solution to make its HBase data highly available.

Which architectural pattern meets company's requirements?

- A. Use Spot Instances for core and task nodes and a Reserved Instance for the EMR master node. Configure the EMR cluster with multiple master nodes. Schedule automated snapshots using Amazon EventBridge.
- B. Store the data on an EMR File System (EMRFS) instead of HDFS. Enable EMRFS consistent view. Create an EMR HBase cluster with multiple master nodes. Point the HBase root directory to an Amazon S3 bucket.
- C. Store the data on an EMR File System (EMRFS) instead of HDFS and enable EMRFS consistent view. Run two separate EMR clusters in two different Availability Zones. Point both clusters to the same HBase root directory in the same Amazon S3 bucket.



D. Store the data on an EMR File System (EMRFS) instead of HDFS and enable EMRFS consistent view. Create a primary EMR HBase cluster with multiple master nodes. Create a secondary EMR HBase read-replica cluster in a separate Availability Zone. Point both clusters to the same HBase root directory in the same Amazon S3 bucket.

Correct Answer: C

Reference: <https://docs.aws.amazon.com/emr/latest/ReleaseGuide/emr-hbase-s3.html>

---

### QUESTION 3

A company has a data warehouse in Amazon Redshift that is approximately 500 TB in size. New data is imported every few hours and read-only queries are run throughout the day and evening. There is a particularly heavy load with no writes for several hours each morning on business days. During those hours, some queries are queued and take a long time to run. The company needs to optimize query performance and avoid any downtime.

What is the MOST cost-effective solution?

- A. Turn on concurrency scaling in the workload management (WLM) queue.
- B. Add more nodes using the AWS Management Console during peak hours. Set the distribution style to ALL.
- C. Use elastic resize to quickly add nodes during peak times. Remove the nodes when they are not needed.
- D. Use a snapshot, restore, and resize operation. Switch to the new target cluster.

Correct Answer: A

---

### QUESTION 4

A financial services company hosts its data warehouse on a Hadoop cluster located in an on-premises data center. The data is 300 TB in size and grows by 1 TB each day. The data is generated in real time from the company's trading system. The raw data is transformed at the end of the trading day using a custom tool running on the Hadoop cluster.

The company is migrating its data warehouse to AWS using a managed data warehouse product provided by a third party that can ingest data from Amazon S3. The company has already established a 10 Gbps connection with an AWS Region using AWS Direct Connect. The company is required by its security and regulatory compliance policies to not transfer data over the public internet. The company wants to minimize changes to its custom tool for data transformation. The company also plans to eliminate the on-premises Hadoop cluster after the migration.

Which solution MOST cost-effectively meets these requirements?

- A. Create a VPC endpoint for Amazon S3. Run a one-time copy job using the DistCp tool to copy existing files from Hadoop to a target S3 bucket over the VPC endpoint. Schedule a nightly DistCp job on the Hadoop cluster to copy the incremental files produced by the custom tool to the target S3 bucket.
- B. Create a VPC endpoint for Amazon S3. Run a one-time copy job using the DistCp tool to copy existing files from Hadoop to a target S3 bucket over the VPC endpoint. Schedule a nightly job on the trading system servers that produces raw data to copy the incremental raw files to the target S3 bucket. Run the data transformation tool on a transient Amazon EMR cluster to output files to Amazon S3.
- C. Create a VPC endpoint for Amazon S3. Run a one-time copy job using the DistCp tool to copy existing files from



Hadoop to a target S3 bucket over the VPC endpoint. Set up an Amazon Kinesis data stream to ingest raw data from the trading system in real time. Use Amazon Kinesis Data Analytics to transform the raw data and output files to Amazon S3.

D. Complete a one-time transfer of the data using AWS Snowball Edge devices transferring to a target S3 bucket. Schedule a nightly job on the trading system servers that produces raw data to copy the incremental raw files to the target S3 bucket Run the data transformation tool on a transient Amazon EMR cluster to output files to Amazon S3.

Correct Answer: A

---

## QUESTION 5

A financial services firm is processing a stream of real-time data from an application by using Apache Kafka and Kafka MirrorMaker. These tools are running on premises to Amazon Managed Streaming for Apache Kafka (Amazon MSK) in the us-east-1 Region. An Apache Flink consumer running on Amazon EMR enriches the data in real time and transfers the output files to an Amazon S3 bucket. The company wants to ensure that the streaming application is highly available across AWS Regions with an RTO of less than 2 minutes.

Which solution meets these requirements?

- A. Launch another Amazon MSK and Apache Flink cluster in the us-west-1 Region that is the same size as the original cluster in the us-east-1 Region Simultaneously publish and process the data in both Regions. In the event of a disaster that impacts one of the Regions, switch to the other Region.
- B. Set up Cross-Region Replication from the Amazon s3 bucket in the us-east-1 Region to the us-west-1 Region. In the event of a disaster, immediately create Amazon MSK and Apache Flink clusters in the us-west-1 Region and start publishing data to this Region.
- C. Add an AWS Lambda function in the us-east-1 Region to read from Amazon MSK and write to a global Amazon DynamoDB table in on-demand capacity mode. Export the data from DynamoDB to Amazon S3 in the us-west-1 Region. In the event of a disaster that impacts the us-east-1 Region, immediately create Amazon MSK and Apache Flink clusters in the us-west-1 Region and start publishing data to this Region.
- D. Set up Cross-Region Replication from the Amazon S3 bucket in the us-east-1 Region to the us-west-1 Region. In the event of a disaster, immediately create Amazon MSK and Apache Flink clusters in the us-west-1 Region and start publishing data to this Region. Store 7 days of data in on-premises Kafka clusters and recover the data missed during the recovery time from the on-premises cluster.

Correct Answer: C

[Latest DAS-C01 Dumps](#)

[DAS-C01 Practice Test](#)

[DAS-C01 Exam Questions](#)