



MLS-C01^{Q&As}

AWS Certified Machine Learning - Specialty (MLS-C01)

Pass Amazon MLS-C01 Exam with 100% Guarantee

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

<https://www.passapply.com/aws-certified-machine-learning-specialty.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 Machine Learning Specialist is working for an online retailer that wants to run analytics on every customer visit, processed through a machine learning pipeline. The data needs to be ingested by Amazon Kinesis Data Streams at up to 100 transactions per second, and the JSON data blob is 100 KB in size.

What is the MINIMUM number of shards in Kinesis Data Streams the Specialist should use to successfully ingest this data?

- A. 1 shards
- B. 10 shards
- C. 100 shards
- D. 1,000 shards

Correct Answer: B

1 shard can ingest 1 MB/second or 1,000 records/second. So $100 \text{ KB} * 100 = 10 \text{ MB}$ (10 shards required)

QUESTION 2

A Machine Learning Specialist needs to move and transform data in preparation for training. Some of the data needs to be processed in near-real time, and other data can be moved hourly. There are existing Amazon EMR MapReduce jobs to clean and feature engineering to perform on the data.

Which of the following services can feed data to the MapReduce jobs? (Choose two.)

- A. AWS DMS
- B. Amazon Kinesis
- C. AWS Data Pipeline
- D. Amazon Athena
- E. Amazon ES

Correct Answer: BC

<https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/what-is-datapipeline.html>

Amazon Kinesis (Option B) can be used to process data in near-real time, as it is a real-time data streaming service that can handle large amounts of incoming data from multiple sources. The data can be fed to Amazon EMR MapReduce jobs for processing.

AWS Data Pipeline (Option C) can be used to move the hourly data, as it provides a way to move data from various sources to Amazon EMR for processing.



QUESTION 3

A company wants to create a data repository in the AWS Cloud for machine learning (ML) projects. The company wants to use AWS to perform complete ML lifecycles and wants to use Amazon S3 for the data storage. All of the company's data currently resides on premises and is 40 TB in size.

The company wants a solution that can transfer and automatically update data between the on-premises object storage and Amazon S3. The solution must support encryption, scheduling, monitoring, and data integrity validation.

Which solution meets these requirements?

- A. Use the S3 sync command to compare the source S3 bucket and the destination S3 bucket. Determine which source files do not exist in the destination S3 bucket and which source files were modified.
- B. Use AWS Transfer for FTPS to transfer the files from the on-premises storage to Amazon S3.
- C. Use AWS DataSync to make an initial copy of the entire dataset. Schedule subsequent incremental transfers of changing data until the final cutover from on premises to AWS.
- D. Use S3 Batch Operations to pull data periodically from the on-premises storage. Enable S3 Versioning on the S3 bucket to protect against accidental overwrites.

Correct Answer: C

Configure DataSync to make an initial copy of your entire dataset, and schedule subsequent incremental transfers of changing data until the final cut-over from on-premises to AWS. Reference: <https://aws.amazon.com/datasync/faqs/>

QUESTION 4

Each morning, a data scientist at a rental car company creates insights about the previous day's rental car reservation demands. The company needs to automate this process by streaming the data to Amazon S3 in near real time. The solution must detect high-demand rental cars at each of the company's locations. The solution also must create a visualization dashboard that automatically refreshes with the most recent data.

Which solution will meet these requirements with the LEAST development time?

- A. Use Amazon Kinesis Data Firehose to stream the reservation data directly to Amazon S3. Detect high-demand outliers by using Amazon QuickSight ML Insights. Visualize the data in QuickSight.
- B. Use Amazon Kinesis Data Streams to stream the reservation data directly to Amazon S3. Detect high-demand outliers by using the Random Cut Forest (RCF) trained model in Amazon SageMaker. Visualize the data in Amazon QuickSight.
- C. Use Amazon Kinesis Data Firehose to stream the reservation data directly to Amazon S3. Detect high-demand outliers by using the Random Cut Forest (RCF) trained model in Amazon SageMaker. Visualize the data in Amazon QuickSight.
- D. Use Amazon Kinesis Data Streams to stream the reservation data directly to Amazon S3. Detect high-demand outliers by using Amazon QuickSight ML Insights. Visualize the data in QuickSight.

Correct Answer: A



QUESTION 5

A data scientist wants to use Amazon Forecast to build a forecasting model for inventory demand for a retail company. The company has provided a dataset of historic inventory demand for its products as a .csv file stored in an Amazon S3 bucket. The table below shows a sample of the dataset.

timestamp	item_id	demand	category	lead_time
2019-12-14	uni_000736	120	hardware	90
2020-01-31	uni_003429	98	hardware	30
2020-03-04	uni_000211	234	accessories	10

How should the data scientist transform the data?

- A. Use ETL jobs in AWS Glue to separate the dataset into a target time series dataset and an item metadata dataset. Upload both datasets as .csv files to Amazon S3.
- B. Use a Jupyter notebook in Amazon SageMaker to separate the dataset into a related time series dataset and an item metadata dataset. Upload both datasets as tables in Amazon Aurora.
- C. Use AWS Batch jobs to separate the dataset into a target time series dataset, a related time series dataset, and an item metadata dataset. Upload them directly to Forecast from a local machine.
- D. Use a Jupyter notebook in Amazon SageMaker to transform the data into the optimized protobuf recordIO format. Upload the dataset in this format to Amazon S3.

Correct Answer: A

Target and metadata must be in two files and loaded from S3, based on documentation:
<https://docs.aws.amazon.com/forecast/latest/dg/dataset-import-guidelines-troubleshooting.html>