1. True or False: AWS Lake Formation is a centralized repository, such as a data lake, that stores structured and unstructured data at any scale.

○ True

◉ False

> ✓ **Correct**
> AWS Lake Formation is a service that automates many of the manual steps that are needed to create data lakes. The steps include collecting, cleansing, moving, and cataloging data. The steps also include securely making that data available for deriving insights. For more information, see the *Introduction to AWS Lake Formation* video in week 2.

2. What is the structure of the AWS Glue Metadata Catalog?

◉ The AWS Glue Metadata Catalog consists of tables. Each table has a schema, which outlines the structure of a table, including columns, data type definitions, and more. The tables are organized into logical groups that are called databases.

○ The AWS Glue Metadata Catalog consists of file systems or databases for any applications that require fine, granular updates and access to raw, unformatted, block-level storage.

○ The AWS Glue Metadata Catalog contains buckets with different types of storage options. AWS Glue Metadata Catalog stores data as objects in these buckets.

○ The AWS Glue Metadata Catalog is the storage that is associated with automated database backups and any active database snapshots. It consists of the General Purpose SSD, Provisioned IOPS SSD, Throughput Optimized HDD, and Cold HDD volume types.

3. True or False: Customers can use Amazon API Gateway to ingest real-time data in a RESTful manner through the creation of an HTTP-based API, which acts as the front door or interface to ingestion logic or data lake storage on the backend.

1 / 1 point

⦿ True

◯ False

4. Which statement about AWS Lake Formation is true?

0 / 1 point

◯ AWS Lake Formation runs big data frameworks, such as Apache Hadoop.

◯ AWS Lake Formation registers the Amazon Simple Storage Service (Amazon S3) buckets and paths where the data lake will reside.

◯ AWS Lake Formation ingests, cleanses, and transforms the structured and organized data.

⦿ AWS Lake Formation deploys, operates, and scales clusters in the AWS Cloud

5. Which service is commonly used for real-time data processing when Amazon Kinesis Data Streams is used for data ingestion?

1 / 1 point

- ⦿ Amazon Kinesis Data Analytics
- ◯ Amazon Athena
- ◯ Amazon EMR
- ◯ AWS Glue Jobs

✓ **Correct**
Amazon Kinesis Data Analytics processes data streams and generates real-time dashboards. For more information, see the *AWS Services for Analytics* video.

6. Apache Hadoop is an open-source framework that is used to efficiently store and process large datasets. A solutions architect is working for a company that currently uses Apache Hadoop on-premises for data processing jobs. The company wants to use AWS for these jobs, but they also want to continue using the same technology. Which service should the solutions architect choose for this use case?

1 / 1 point

- ⦿ Amazon EMR
- ◯ Amazon OpenSearch Service
- ◯ AWS Lambda
- ◯ Amazon Kinesis Data Analytics

✓ **Correct**
Amazon EMR is a managed cluster platform that simplifies running big data frameworks (such as Apache Hadoop and Apache Spark) on AWS to process and analyze large amounts of data. For more information, see the *AWS Services for Data Processing* video in week 2.

7. A team of machine learning (ML) experts are working for a company. The company wants to use the data in their data lake to train an ML model that they create. The company wants the most control that they can have over this model and the environment that it is trained in. Which AWS ML approach should the team take?

○ Use a pretrained model from an AWS service, such as Amazon Rekognition.

○ Launch an Amazon Elastic Compute Cloud (Amazon EC2) instance and run Amazon SageMaker on it to train the model.

◉ Launch an Amazon Elastic Compute Cloud (Amazon EC2) instance by using an AWS Deep Learning Amazon Machine Image (AMI) to host the application that will train the model.

○ Create an AWS Lambda function with the training logic in the handler, and run the training based on an event.

> ✓ **Correct**
> The team of ML experts will probably use EC2 instances for their compute power on AWS. They can launch an EC2 instance with the AWS Deep Learning AMIs that provide the greatest control over building and managing deep learning models and clusters. For more information, see the *AWS Services for Predictive Analytics and Machine Learning* video in week 1.

8. A solutions architect needs to process and analyze data as it is ingested into a data lake in real time. They want to get timely insights about the streaming data. Which service should the solutions architect use for this use case?

○ Amazon API Gateway

○ Amazon EMR

◉ Amazon Kinesis

○ AWS Lambda

9. Which services can query data that is needed to build reports? (Choose TWO.)

1 / 1 point

☑ Amazon Athena

☐ AWS Lambda

☐ Amazon Glue

☑ Amazon Redshift

☐ Amazon Elastic Compute Cloud