

1. **True or False:** It is a best practice that companies treat the original, ingested version of data in their data lake as immutable. Any data processing that is done to the original data should be stored as a secondary copy or extra copy of the data, which will then be analyzed.

1 / 1 point

- ☒ True
- ☐ False

✓ **Correct**

Companies can make copies of the data, but the original data that was ingested must remain untouched. With Amazon Simple Storage Service (Amazon S3) Lifecycle policies, companies can move raw data to more cost-effective storage tiers when it becomes more infrequently accessed over time. For more information, see the *Data Prep and AWS Glue Jobs* video in week 4.

2. Which scenario represents AWS Glue jobs as the BEST tool for the job?

1 / 1 point

- ☐ Analyze data in real time as data comes into the data lake.
- ☐ Analyze data in batches on schedule or on demand.
- ☐ Transform data in real time as data comes into the data lake.
- ☒ Transform data on a schedule or on demand.

✓ **Correct**

An AWS Glue job runs extract, transform, and load (ETL) scripts that connect to your source data, process it, and then write it out to your data target. AWS Glue triggers can start jobs based on a schedule or event, or on demand. For more information, see the *Columnar Data Formats and Amazon Athena Optimizations* reading in week 4.

3. A company collects and analyzes large amounts of data daily. Why should the company use a compression strategy for their processes?

1 / 1 point

- ☐ Compressed data uses a row-based data format that works well for data optimization.

- ☐ Compressed data slows the time to process and analyze information.
- ☒ By using compressed data, data-processing systems can optimize for memory and cost.
- ☐ Compressed data increases the risk of losing valuable information.

☒ **Correct**

Most high performance big data technologies copy data to RAM for faster performance. In this case, if companies compress data, they can fit more content into the same memory space. By doing so, companies can save on costs if they use services that charge per usage. For more information, see the *File Optimization* video in week 4.

4. A software developer recently uploaded data logs from their application to Amazon Simple Storage Service (Amazon S3). Who is responsible for encrypting both the data at rest in the S3 bucket and the data in transit to the S3 bucket, according to the AWS shared responsibility model?

1 / 1 point

- ☐ AWS
- ☒ Customer
- ☐ Both AWS and the customer
- ☐ Third-party security company

☒ **Correct**

According to the AWS shared responsibility model, customers are responsible for the security in the cloud. AWS provides many features that customers can use to encrypt data both at rest and in transit. For more information about the correct answer, see the *Security and Compliance* reading.

5. Which statement about data visualization is TRUE?

1 / 1 point

- ☐ Raw data is generally formatted to be read and used by a human eye.
- ☐ Visualization data is always captured in a text editor.
- ☒ If there is more data, making sense of the data will be more difficult without using visualization tools.

☐ A click map is the main reason to invest into data visualization.

☒ **Correct**

Data visualization takes abstract data (or data that is not easily digested or understood) and represents the data in a visual and interactive way. The goal of data visualization is to enhance understanding of the data and to derive insights from it. For more information about the correct answer, see the *Power of Data Visualization* video in week 4.

6. What makes Amazon QuickSight different, compared to other traditional business intelligence (BI) tools?

1 / 1 point

- ☐ The ability to visualize data
- ☐ The ability to create sharable dashboards
- ☒ Super-fast, Parallel, In-memory Calculation Engine (SPICE)
- ☐ Data encryption at every layer

☒ **Correct**

SPICE is a QuickSight feature that is engineered to rapidly perform advanced calculations and serve data. For more information, see the *Introduction to Amazon QuickSight* video in week 4.

7. What is the purpose of the Registry of Open Data on AWS?

1 / 1 point

- ☒ Help people discover and share datasets that are available through AWS resources.
- ☐ Help people discover and share datasets that are available outside of AWS resources.
- ☐ Provide a service that people can use to transform public datasets that are published by data providers through an API.
- ☐ Provide a service that people can use to ingest software as a service (SaaS) application data into a data lake.

☒ **Correct**

The Registry of Open Data on AWS is a collection of publicly available datasets that users can access from AWS resources. For more information, see the *Registry of Open Data* on AWS video in week 4.

8. **True or False:** Amazon QuickSight is a cloud-scale business intelligence (BI) service that developers can use to deliver interactive visualizations and dashboards for data analysis and forecasting.

1 / 1 point

- ☒ True
- ☐ False

☒ **Correct**

QuickSight helps organizations create interactive dashboards from different data sources. For more information, see the *Amazon QuickSight Demo* in week 4.