Congratulations! You passed!

Grade received 100%

Latest Submission Grade 100%

To pass 71% or higher

Go to next item

1. What is the main value proposition of data lakes? **1 / 1 point**

- The ability to store user-generated data, such as data from antennas and sensors.
- The ability to combine multiple databases together to expand their capacity and availability.
- The ability to ingest and store data that could be the answer for future questions when they are processed with the correct data processing mechanisms.
- The ability to define the data schema before ingesting and storing data.

Correct
A data lake is a centralized repository that stores data as-is, without needing to first structure the data and run different types of analytics. The ingested data can be later processed and visualized for specific needs. For more information, see the *Why Data Lakes* video in week 1.

2. **True or False:** Two of the fundamental components of data lakes are data catalog and search. **1 / 1 point**

- True
- False

Correct
Mature data lakes provide efficient data cataloging (otherwise known as indexing) and searching mechanisms to quickly discover what data is stored, and where. For more information, see the *Data Lakes Components* video in week 1.

3. A company sorts and structures data before entering information into a database. They also store unstructured data in another storage location. These two data locations are siloed from each other. How can the company benefit from using a data lake? **1 / 1 point**

- Data lakes mostly process data after it has been stored in the cloud or on-premises.
- A data lake provides the most secure way to store data in the AWS Cloud.
- With a data lake, a company can store structured and unstructured data at virtually any scale.
- A data lake is a direct replacement of a data warehouse.

Correct
Companies can store data as-is, without needing to first structure the data. After analyzing raw data, companies can identify and act upon opportunities for business growth more quickly by attracting and retaining customers, boosting productivity, proactively maintaining devices, and making informed decisions. For more information, see the *Data Lake Characteristics and Components* reading.

4. Which statements about data lakes and data warehouses are true? (Choose TWO.) **1 / 1 point**

- [ ] Data lakes use schema-on-write architectures and data warehouses use schema-on-read architectures.
- [x] Data lakes offer more choices in terms of the technology that is used for processing data. In contrast, data warehouses are more restricted to using Structured Query Language (SQL) as the query technology.

Correct
Data lakes provide more power and flexibility by supporting multiple choices for processing data. For more information, see the *Comparison of a Data Lake to a Data Warehouse* video in week 1.

- [x] The solutions architect can combine both data lakes and data warehouses to better extract insights and turn data into information.

Correct
Some common architectures use data lakes to ingest, store, and clean data. Then, the solutions architect can move that data into a data warehouse for visualization. For more information, see the *Comparison of a Data Lake to a Data Warehouse* video in week 1.

- [ ] The solutions architect cannot attach data visualization tools to data warehouses.
- [ ] Data lakes are not future-proof, which means that they must be reconfigured each time new data is ingested.

5. True or False: Data lakes integrate with analytics tools that can help companies eliminate costly and complex extract, transform, and load (ETL) processes. **1 / 1 point**

- True
- False

Correct
The breadth and depth of analytics services on AWS makes it easier to provision the appropriate resources to run whatever analysis is most appropriate for a specific need. For more information, see the *Data Lake Characteristics and Components* reading in week 1.

6. Which term indicates that a data lake lacks curation, management, cataloging, lifecycle or retention policies, and metadata? **1 / 1 point**

- Data swamp
- Data warehouse
- Data catalog
- Database

Correct
Data swamp is an informal term that represents a data lake with disorganized data. For more information, see the *Data Lakes Components* video in week 1.

7. Which service can be used to run simple queries against data in a data lake? **1 / 1 point**

- Amazon Athena
- Amazon Kinesis Agent
- Amazon Kinesis Data Firehose
- Amazon Simple Storage Service (Amazon S3)

Correct
After the dataset reaches Amazon Simple Storage Service (Amazon S3), Structured Query Language (SQL) queries can be run in Amazon Athena to gain insights on the data. For more information, see the *Discussing Sample Data Lake Architectures* video in week 1.