Congratulations! You passed!

Grade received 80% | Latest Submission Grade 80% | To pass 80% or higher

Go to next item

You're viewing this assessment in its original language
You can switch back to view this content in your preferred translation if you'd prefer. You won't lose any progress if you change languages. Show Vietnamese version

1. In Azure Data Factory you define the work performed as a pipeline of operations. 1 / 1 point

Which of the following statements describe the features of an Azure Data Factory Pipeline?

- ○ You can create pipelines using custom code only.
- ◉ You can create pipelines using the graphical user interface and custom code.
- ○ You can create pipelines using the graphical user interface only.

Correct
You can create pipelines using the graphical user interface provided by Microsoft, or by writing your own code.

2. Azure Data Lake is a repository for large quantities of which form of data? 1 / 1 point

- ◉ Raw data
- ○ Formatted data
- ○ Cleaned data

Correct
A data lake is a repository for large quantities of raw data. Because the data is raw and unprocessed, it's very fast to load and update, but the data hasn't been put into a structure suitable for efficient analysis.

3. Azure Data Lake is a repository for large quantities of raw data. As part of the data processing pipeline, it uses a staging area. What does the term staging area mean? 1 / 1 point

- ○ A staging area is where data is held temporarily after being processed.
- ◉ A staging area is where data is held temporarily before being processed.

Correct
A data lake is a repository for large quantities of raw data. Because the data is raw and unprocessed, it's very fast to load and update, but the data hasn't been put into a structure suitable for efficient analysis. You can think of a data lake as a staging point for your ingested data before it's massaged and converted into a format suitable for performing analytics.

4. Azure Data Lake Storage has which of the following features? 1 / 1 point

Select all options that apply.

- ☑ Directories and subdirectories support

Correct
Azure Data Lake Storage combines the hierarchical directory structure and file system semantics of a traditional file system with security and scalability provided by Azure.

- ☑ Hadoop Distributed File System (HDFS) support

Correct
Azure Data Lake Storage is compatible with the Hadoop Distributed File System (HDFS). Hadoop is a highly flexible and programmable analysis service, used by many organizations to examine large quantities of data.

- ☐ Structured data storage support
- ☑ Support for Role-Based Access Control (RBAC)

Correct
Data Lake Storage supports the Portable Operating System Interface (POSIX) file and directory permissions to enable granular Role-Based Access Control (RBAC) on your data.

5. Azure Databricks is an Apache Spark environment running on Azure, what are some of the features of Azure Databricks? 1 / 1 point

Select all options that apply.

- [x] Streaming

Correct
Azure Databricks is an Apache Spark environment running on Azure to provide big data processing, streaming, and machine learning.

- [x] Big data processing

Correct
Azure Databricks is an Apache Spark environment running on Azure to provide big data processing, streaming, and machine learning.

- [x] Machine learning

Correct
Azure Databricks is an Apache Spark environment running on Azure to provide big data processing, streaming, and machine learning.

- [ ] Software-as-a-Service (SaaS) offering from Azure

6. True or False? 0 / 1 point

PolyBase allows T-SQL queries to join data from external sources to relational tables in an instance of SQL Server. To enable this feature client connection software is required.

- (•) True
- ( ) False

Incorrect
PolyBase enables your SQL Server instance to query data with T-SQL directly from SQL Server without separately installing client connection software.

7. In Azure Data Factory ingestion tasks can be triggered using which of the following components? 1 / 1 point

- (•) Pipeline
- ( ) CSV File
- ( ) Linked service

Correct
Pipelines can be triggered to run activities for ingesting data.

8. The term big data refers to data that is too large or complex for traditional database systems. Systems that process big data have to perform rapid data ingestion and processing; they must have the capacity to store the results, and sufficient compute power to perform analytics over these results. Another option is to analyze operational data in its original location. This is referred to as? 1 / 1 point

- ( ) Hybrid transactional processing (HTP)
- ( ) Hybrid analytical processing (HAP)
- (•) Hybrid transactional analytical processing (HTAP)

Correct
This strategy is known as hybrid transactional analytical processing (HTAP). You can perform this style of analysis over data held in repositories such as Azure Cosmos DB using Azure Synapse Link.

9. You can use HDInsight to analyze data using frameworks such as Hadoop Map/Reduce, Apache Spark, Apache Hive, Apache Kafka, Apache Storm, R, and more. Apache Hive provides which of the following features? 1 / 1 point

- (•) Interactive SQL-like facilities for querying, aggregating, and summarizing data.
- ( ) A clustered streaming service that can ingest data in real-time.
- ( ) A scalable, fault-tolerant platform for running real-time data processing applications.
- ( ) The ability to split a task over a large dataset into a series of smaller tasks.

Correct
Apache Hive provides interactive SQL-like facilities for querying, aggregating, and summarizing data. The data can come from many different sources. Queries are converted into tasks and parallelized.

10. HDInsight can be used to analyze data using a collection of open-source tools. The ability to provide a clustered streaming service that can ingest data in real-time is provided by which service? 0 / 1 point

- ○ Apache Storm
- ◉ Hadoop Map/Reduce
- ○ Apache Hive
- ○ Apache Kafka

⊗ **Incorrect**
Hadoop Map/Reduce uses a simple framework to split a task over a large dataset into a series of smaller tasks over subsets of the data that can be run in parallel, and the results then combined.