Congratulations! You passed!
TO PASS 80% or higher

Keep Learning

GRADE
100%

Course practice exam

LATEST SUBMISSION GRADE
100%

1. Can you identify three features of Apache Spark? **1 / 1 point**

- [x] Distributed execution engine

Correct
The Apache Spark core engine is a distributed execution engine.

- [ ] Disk-based processing
- [x] In-memory processing

Correct
Apache Spark is a parallel processing framework that supports in-memory processing to boost the performance of big-data analytics applications.

- [x] Parallel Processing Framework

Correct
Apache Spark is a parallel processing framework that supports in-memory processing to boost the performance of big-data analytics applications.

2. What needs to be created first when building an Apache Spark pool in Azure Synapse Analytics? **1 / 1 point**

- ( ) SQL Database
- (•) Workspace
- ( ) Notebook

Correct
In order create a new Apache Spark pool you will need to have an Azure Synapse Analytics workspace created first.

3. Spark pools in Azure Synapse Analytics are compatible with which two of the following storage types? **1 / 1 point**

- [ ] Azure Data Lake Generation 1 Storage
- [x] Azure Storage

Correct
Spark pools in Azure Synapse Analytics are compatible with Azure Storage.

- [x] Azure Data Lake Generation 2 Storage

Correct
Spark pools in Azure Synapse Analytics are compatible with Azure Data Lake Generation 2 Storage.

- [ ] SQL Storage

4. Which of the following solutions can you utilize to create an embedded Apache Spark capability that can reside on the same platform as data warehouses and data integration capabilities, as well as integrate with other Azure services? **1 / 1 point**

- (•) Apache Spark for Azure Synapse
- ( ) Azure Databricks
- ( ) Azure HDInsight
- ( ) Apache Spark

✓ **Correct**

Apache Spark is an embedded Apache Spark capability within Azure Synapse Analytics residing on the same platform that contains data warehouses and data integration capabilities, as well as integrating with other Azure services.

5. Which two of the following features can you use to ingest data through Spark notebooks? **1 / 1 point**

- [x] Linked Service

✓ **Correct**

To ingest data through notebooks, you can use a linked service already defined in the Azure Synapse workspace. An example could be a linked service defined for an Azure Storage account.

- [ ] Azure SQL
- [ ] Azure Cosmos DB
- [x] Primary Storage

✓ **Correct**

Every Azure Synapse workspace has an associated primary storage account defined when it is created.

6. What are the primary languages available in the notebook environment? **1 / 1 point**

- [x] Spark SQL

✓ **Correct**

Spark SQL is a supported language in Synapse Studio notebooks.

- [ ] YAML
- [x] Spark (Scala)

✓ **Correct**

Scala is a supported language in Synapse Studio notebooks.

- [ ] JSON
- [x] PySpark (Python)

✓ **Correct**

PySpark is a supported language in Synapse Studio notebooks.

- [x] .NET Spark (C#)

✓ **Correct**

.Net Spark is a supported language in Synapse Studio notebooks.

7. Which of the following actions do you need to perform in order to directly reference data or variables in Azure Synapse Studio notebook using different languages? **1 / 1 point**

- ( ) Use a magic command for that language.
- ( ) Create a new Notebook.
- (•) Create a temporary table so that it can be referenced across different languages.
- ( ) Do Nothing. You can reference data or variables directly using different languages in an Azure Synapse Studio notebook.

✓ **Correct**

You cannot reference data or variables directly using different languages in an Azure Synapse Studio notebook. If you wish to do this using Spark, you must first create a temporary table so that it can be referenced across different languages.

8. What are DataFrames? **1 / 1 point**

- [x] DataFrames optimize execution plans on queries that will access the data held in the DataFrame.

✓ **Correct**

DataFrames optimize execution plans on queries that will access the data held in the DataFrame.

- [x] DataFrames enable Apache Spark to understand the schema of the data.

Correct
DataFrames enable Apache Spark to understand the schema of the data.

- [ ] DataFrames are a collection of data organized into named Rows.
- [x] DataFrames are a collection of data organized into named columns.

Correct
DataFrames are a collection of data organized into named columns.

9. You enter the following Python snippet into your code: **1 / 1 point**

fromazureml.opendatasets import NycTlcYellow

data =NycTlcYellow()

data_df = data.to_spark_dataframe()

display(data_df.limit(10))

What is the purpose of the *display(data_df.limit(10))* method?

- ( ) Return batches of 10 rows of data from the *data_df* variable until all records are returned.
- ( ) Limit the Dataframe to only retrieve 10 rows of data from the NycTLcYellow data source.
- (•) Return 10 rows of data from the *data_df* variable.

Correct
10 rows of data are returned back from the *data_df* variable using the display method.

10. You need to load data into an Apache Spark DataFrame from several different file types. Which three of the following storage services can you use to complete this action? **1 / 1 point**

- [x] Azure Data Lake Store Generation 2

Correct
You can load data into an Apache Spark DataFrame from different file types stored in an Azure Storage Account, or from data stored in a dedicated SQL pool.

- [x] Primary Storage Account

Correct
You can load data into an Apache Spark DataFrame from different file types stored in an Azure Storage Account, or from data stored in a dedicated SQL pool.

- [x] Azure Storage Account

Correct
You can load data into an Apache Spark DataFrame from different file types stored in an Azure Storage Account, or from data stored in a dedicated SQL pool.

- [x] Dedicated SQL pool

Correct
You can load data into an Apache Spark DataFrame from different file types stored in an Azure Storage Account, or from data stored in a dedicated SQL pool.

- [ ] Serverless SQL Pool

11. The Azure Synapse Apache Spark pool to Synapse SQL connector uses which of the following in SQL pools to efficiently transfer data between the Spark cluster and the Synapse SQL instance? **1 / 1 point**

- ( ) Azure Data Lake Storage Generation 2 and XML.
- (•) Azure Data Lake Storage Generation 2 and PolyBase.
- ( ) Azure Data Lake Storage Generation 2 and JSON.

Correct
Azure Data Lake Storage Generation 2 and PolyBase in SQL pools can be used to efficiently transfer data between the Spark cluster and the Synapse SQL instance.

12. Which of the following role memberships are required to successfully authenticate between two systems in Azure Synapse Analytics? **1 / 1 point**

- [ ] The account used needs to be a member of Storage Blob Data Contributor role in the database or SQL pool from which you to transfer data to or from.
- [x] The account used needs to be a member of the Storage Blob Data Contributor role on the default storage account.

✓ **Correct**
Try going back and reviewing Query pools and manage workloads in Azure Synapse Analytics.

- [x] The account used needs to be a member of db_exporter role in the database or SQL pool from which you to transfer data to or from.

✓ **Correct**
The account used needs to be a member of db_exporter role in the database or SQL pool from which you to transfer data to or from.

13. Which of the following is used to load data into a table created within a dedicated SQL pool using Write API? **1 / 1 point**

- (•) Polybase
- ( ) ORC
- ( ) Parquet
- ( ) JSON

✓ **Correct**
The Write API creates a table in a dedicated SQL pool. It then invokes Polybase to load the data into the table that was created.

14. What is the minimum number of nodes allowed when creating an Apache Spark pool with Autoscaling? **1 / 1 point**

- ( ) 4
- (•) 3
- ( ) 2
- ( ) 1

✓ **Correct**
The minimum number of nodes allowed is 3.

15. What three actions occur within existing nodes in Azure Synapse Analytics when you scale down Apache Spark pools? **1 / 1 point**

- [x] Jobs that are still running will continue to run and finish.

✓ **Correct**
In Azure Synapse when you scale down Apache spark pools Jobs that are still running will continue to run and finish.

- [ ] Nodes to be scaled down will be shut down immediately regardless of current state.
- [x] Nodes to be scaled down will be put in a decommissioned state.

✓ **Correct**
In Azure Synapse when you scale down Apache spark pools Nodes to be scaled down will be put in a decommissioned state.

- [x] Pending jobs will be in a waiting state and scheduled for execution on fewer nodes.

✓ **Correct**
In Azure Synapse when you scale down Apache spark pools pending jobs will be in a waiting state and scheduled for execution on fewer nodes.

- [ ] Pending jobs will be lost.