Knowledge check

Total points 6

1. Question 1

Which three of the following are features of the Apache Spark application?

1 / 1 point

☑ Distributed execution engine

Correct

The Apache Spark core engine is a distributed execution engine.

☑ 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.

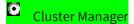☐ Disk-based processing

☑ 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. Question 2

Which one of the following objects is responsible for allocating resources across applications in an Apache Spark pool?

1 / 1 point

◉ Cluster Manager

○ Nodes

○ SparkContext

○ Executors

Correct

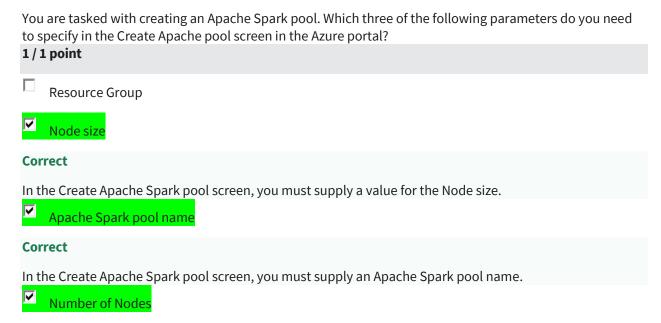The SparkContext can connect to the cluster manager, which allocates resources across applications.

3. Question 3

Apache Spark pools in Azure Synapse Analytics are compatible with which of the following types of storage?

1 / 1 point

☑ Azure Storage

Correct

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

☐ SQL Storage

☑ Azure Data Lake Generation 2 Storage

Correct

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

☐ Azure Data Lake Generation 1 Storage

4. Question 4

You need to manage an end-to-end big data project using one single platform. Which of the following data services is best suited to this task?

1 / 1 point

○ Apache Spark for Azure Synapse

○ Azure HDInsight

◉ Azure Databricks

○ Apache Spark

Correct

Azure Databricks provides the ability to create and manage an end-to-end big data/data science project using one platform.

5. Question 5

Which of the following is an element of an Apache Spark Pool in Azure Synapse Analytics?

1 / 1 point

○ Apache Spark Console

○ Azure HDInsight

◉ Spark Instance

Correct

The definition of an Apache Spark pool is that, when instantiated, it is used to create an Apache Spark instance that processes data.
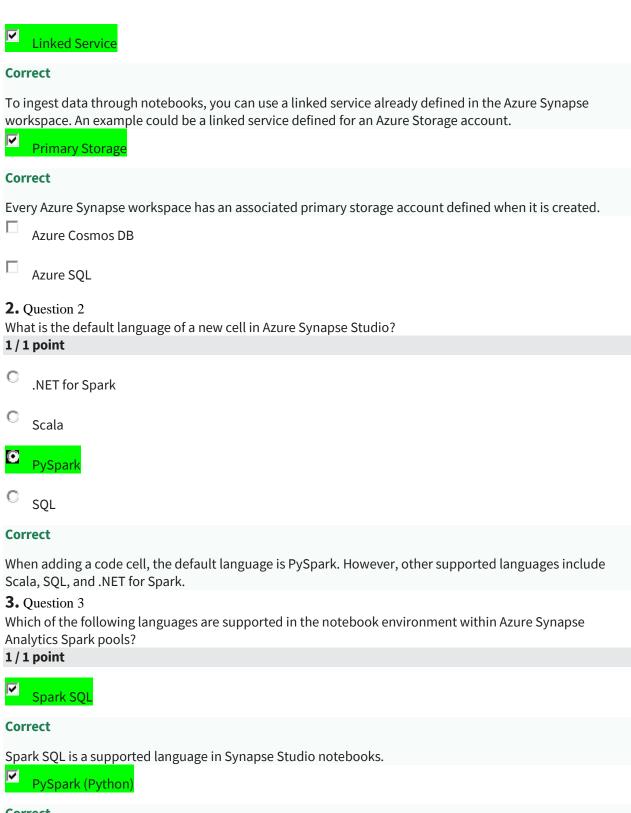
6. Question 6

You are tasked with creating an Apache Spark pool. Which three of the following parameters do you need to specify in the Create Apache pool screen in the Azure portal?

1 / 1 point

☐   Resource Group

☑   Node size

Correct

In the Create Apache Spark pool screen, you must supply a value for the Node size.

☑   Apache Spark pool name

Correct

In the Create Apache Spark pool screen, you must supply an Apache Spark pool name.

☑   Number of Nodes

Correct

In the Create Apache Spark pool screen, you must specify the number of Nodes to use in the Pool.

Knowledge check

Total points 6

1. Question 1

You need to ingest data through Apache Spark notebooks. Which two of the following features can you use to carry out this task?

1 / 1 point

☑ **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.

☑ **Primary Storage**

Correct

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

☐ Azure Cosmos DB

☐ Azure SQL

2. Question 2

What is the default language of a new cell in Azure Synapse Studio?

1 / 1 point

○ .NET for Spark

○ Scala

◉ **PySpark**

○ SQL

Correct

When adding a code cell, the default language is PySpark. However, other supported languages include Scala, SQL, and .NET for Spark.

3. Question 3

Which of the following languages are supported in the notebook environment within Azure Synapse Analytics Spark pools?

1 / 1 point

☑ **Spark SQL**

Correct

Spark SQL is a supported language in Synapse Studio notebooks.

☑ **PySpark (Python)**

Correct

PySpark is a supported language in Synapse Studio notebooks.

☑ **Spark (Scala)**

Correct

Scala is a supported language in Synapse Studio notebooks.

☑ .NET Spark (C#)

Correct

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

☐ JSON

☐ YAML

4. Question 4

Your Azure Studio notebook needs to be able to reference data or variables directly using different languages. Which of the following actions do you need to perform to enable this?

1 / 1 point

○ Use a magic command for that language.

○ You don't need to do anything as you can reference data or variables directly using different languages in an Azure Synapse Studio notebook.

◉ Create a temporary table so that it can be referenced across different languages.

○ Create a new 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 should first create a temporary table so that it can be referenced across different languages.

5. Question 5

Azure Synapse Studio notebooks are based on which one of the following?

1 / 1 point

○ Dedicated SQL pool

◉ Apache Spark

○ Apache Spark pool

Correct

Correct! Azure Synapse Studio notebook is purely Apache Spark based.

6. Question 6

Which one of the following actions should you take to save all notebooks in Azure Synapse studio?

1 / 1 point

◉ Select the Publish all button on the workspace command bar.

○ Select the Publish button on the notebook command bar.

○ Press CTRL + S.

Correct

To save all notebooks in your workspace, select the Publish all button on the workspace command bar.

Knowledge check

Total points 6

1. **Question 1**

Which three of the following are features of DataFrames?

1 / 1 point

☐ DataFrames are a collection of data organized into named Rows.

☑ 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.

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

Correct

DataFrames are a collection of data organized into named columns.

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

Correct

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

2. **Question 2**

You input the following Python code snippet into your workspace:

new_rows = [('CA',22, 45000),("WA",35,65000) ,("WA",50,85000)]

demo_df = spark.createDataFrame(new_rows, ['state', 'age', 'salary'])

demo_df.show()

The variable named *demo_df* above is used to do which of the following?

1 / 1 point

⦿ It uses the *spark.createDataFrame* method referencing the *new_rows* variable in the first parameter. The second parameter defines the column heading names for the DataFrame as state, age, and salary.

○ It uses the *spark.createDataFrame* method to create a variable named new_rows which will store the values '*state*', '*age*', and '*salary*' .

○ It uses the *spark.createDataFrame* method and creates a variable named *new_rows* which creates the data in the code segment to store in the DataFrame.

Correct

The second variable named *demo_df* uses the *spark.createDataFrame* method, which refers to the *new_rows* variable in the first parameter. The second parameter defines the column heading names for the DataFrame as *state*, *age*, and *salary*.

3. Question 3

You input the following Python snippet into your code:

from azureml.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?

1 / 1 point

⦿ Return 10 rows of data from the data_df variable

○ Return batches of 10 rows of data from the *data_df* variable until all records are returned.

○ Limit the Dataframe to retrieving 10 rows of data from the *NycTLcYellow* data source.

Correct

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

4. Question 4

Select the correct series of steps to flatten nested structures and explode arrays with Apache Spark

1 / 1 point

⦿

1. Define a function
2. Flatten nested schema
3. Explode Arrays
4. Flatten child nested schema

○

1. Define a function
2. Flatten nested schema
3. Flatten child nested schema
4. Explode Arrays

○

1. Define a function
2. Flatten child nested schema
3. Flatten nested schema
4. Explode Arrays

Correct
This is the correct sequence of steps to flatten nested structures and explode arrays with Apache Spark.

5. Question 5

Which of these actions should you perform to flatten a nested schema?

0.3333333333333333 / 1 point

☐ Explode Arrays.

☐ Load a CSV file.

☑ Create a Parquet file.

This should not be selected
Creating a Parquet file is not required to flatten a nested schema. It is a file format.

6. Question 6

Which two of the following actions do DataFrames perform?

1 / 1 point

☐ Process data only in streaming data architecture.

☐ Process data only in batch data architecture.

☐ Extract large volumes of data from an SQL Database only.

☑ Process data in either batch or streaming data architecture.

Correct
DataFrames can process data in either batch or streaming data architecture.

☑ Extract large volumes of data from a wide variety of data sources.

Correct
DataFrames are designed to extract large volumes of data from a wide variety of data sources.

Knowledge check

Total points 6

1. Question 1
The interoperability between Apache Spark and SQL helps you to directly explore and analyze which three of the following types of files?

1 / 1 point

- ☑ CSV

Correct

The interoperability between Apache Spark and SQL helps you to directly explore and analyze CSV files.

- ☑ TSV

Correct

The interoperability between Apache Spark and SQL helps you to directly explore and analyze TSV files.
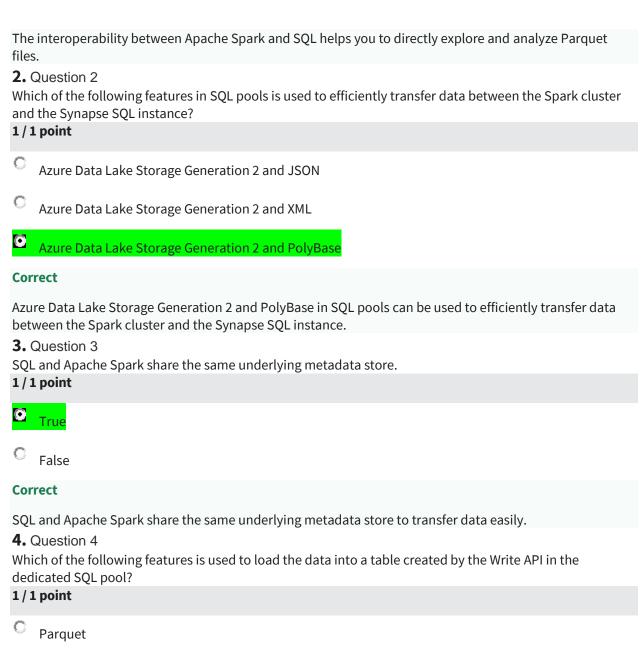
- ☐ YAML

- ☑ JSON

Correct

The interoperability between Apache Spark and SQL helps you to directly explore and analyze JSON files.

- ☑ Parquet

Correct

The interoperability between Apache Spark and SQL helps you to directly explore and analyze Parquet files.

2. Question 2

Which of the following features in SQL pools is used to efficiently transfer data between the Spark cluster and the Synapse SQL instance?

1 / 1 point

◯ Azure Data Lake Storage Generation 2 and JSON

◯ Azure Data Lake Storage Generation 2 and XML

◉ Azure Data Lake Storage Generation 2 and PolyBase

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.

3. Question 3

SQL and Apache Spark share the same underlying metadata store.

1 / 1 point

◉ True

◯ False

Correct

SQL and Apache Spark share the same underlying metadata store to transfer data easily.

4. Question 4

Which of the following features is used to load the data into a table created by the Write API in the dedicated SQL pool?

1 / 1 point

◯ Parquet

◯ ORC

◯ JSON

◉ PolyBase

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.

5. Question 5

You have a requirement to transfer data to a dedicated SQL pool that is outside of the Azure Synapse Analytics workspace. Which form of Authentication can be used to complete this task?

0 / 1 point

○ SQL Authentication Only

○ Azure AD and SQL Authentication

○ None of the above

○ Azure AD only

Incorrect

To establish and transfer data to a dedicated SQL pool that is outside of the workspace without Azure AD, you can use SQL Authentication if AD is not available

6. Question 6

The Azure Synapse Apache Spark to Synapse SQL connector supports which one of the following languages?

1 / 1 point

○ .Net

○ SQL

◉ Scala

○ Python

Correct

The connector uses Scala to integrate Apache Spark pools with dedicated SQL pools in Azure Synapse Analytics.

Knowledge check

Total points 6

1. Question 1

The Apache Spark history server can be accessed directly from which of the following Synapse Studio Tabs.

0.75 / 1 point

☐ Develop

☐ Manage

☑ Monitor

Correct

If you want to navigate to the Apache Spark History server, you can navigate to the Azure Synapse Analytics Studio environment and go to the Monitor tab. In the Monitor tab, you can select Apache Spark Applications and from there select Spark history server.

☑ Data

This should not be selected

To open the Apache Spark History server you can navigate to the Data tab, where if you create a notebook and read a DataFrame you can go to the bottom of the page and find the Spark History Server known as the Spark UI.

2. Question 2

Which of the following features can benefit from query optimization through Catalyst when optimizing Apache Spark jobs in Azure Synapse Analytics?

1 / 1 point

○ Resilient Distributed Datasets (RDDs)

◉ DataFrames

○ Notebooks

Correct

DataFrames provide query optimization through Catalyst.

3. Question 3

You need to specify the minimum number of nodes for an Apache Spark pool with autoscaling. What is the minimum number of nodes allowed?

1 / 1 point

○ 1

○ 2

◉ 3

○ 4

Correct

The minimum number of nodes allowed is 3.

4. Question 4

You scale down Apache Spark pools in Azure Synapse Analytics. What happens to existing nodes?

1 / 1 point

☑ Nodes to be scaled down will be put in a decommissioned state.

Correct

When you scale down Apache spark pools, Nodes to be scaled down will be put in a decommissioned state.

☑ Pending jobs will be in a waiting state and scheduled for execution on fewer nodes.

Correct

When you scale down Apache spark pools, pending jobs will be in a waiting state and scheduled for execution on fewer nodes.

☑ Jobs that are still running will continue to run until completion.

Correct

When you scale down Apache spark pools, jobs that are still running will continue to run until completion.

☐ Nodes to be scaled down will be shut down immediately regardless of current state.

☐ Pending jobs will be lost.

5. Question 5

Which of the following tasks can you perform to optimize an Apache Spark job?

0.6666666666666666 / 1 point

☑ Remove the Apache Spark Pool.

This should not be selected

This is not a valid option as removing the Apache Spark pool removes the capability to process Apache Spark jobs.

☑ Remove all nodes.

Correct

This is not a valid option as removing all nodes would mean there is no compute power to process the Apache Spark jobs.

☐ Use bucketing.

6. Question 6

You encounter a slow performing shuffle job. Which of the following is a possible cause?

1 / 1 point

○ Bucketing

○ Enablement of autoscaling

◉ Data skew

Correct

Correct! The slow performance is due to asymmetry in your job data.

Test prep

Latest Submission Grade 100%

1. Question 1

Apache Spark pools in Azure Synapse Analytics benefit from which four of the following?

1 / 1 point

☑ **Pre-Loaded Anaconda libraries**

Correct

Over 200 Anaconda libraries are pre-installed on the Spark pool in Azure Synapse Analytics.

☑ **REST APIs**

Correct

In order to monitor and submit jobs remotely, you can use Apache Livy as a Rest API for the Spark job server.

Apache Livy is a service that enables interaction with an Spark cluster over a REST interface to enable:

- submission of Spark jobs
- snippets of Spark code
- synchronous or asynchronous result retrieval
- and Spark Context management

☑ **Autoscale**

Correct

Provide support for autoscale, so that pools can be scaled up or down as required, by adding or removing nodes.

☐ Real-time co-authoring (both authors see the changes in real-time)

☑ **Support for third party IDEs**

Correct

Azure Synapse Analytics provides an IDE for IntelliJ to create and submit applications to the Apache Spark pool.

2. Question 2

Which one of the following resources should you create first when building an Apache Spark pool in Azure Synapse Analytics?

1 / 1 point

○ A notebook

○ A SQL Database

◉ **A workspace**

Correct

An Azure Synapse Analytics workspace must be built before creating a new Apache Spark pool.

3. Question 3

Which one of the following resources requires a unique Spark Pool name when creating an Apache Spark pool in Azure Synapse Analytics?

1 / 1 point

◉ **The workspace**

○ The resource group

○ Azure

○ The subscription

Correct

An Apache Spark pool name must adhere to the following specifications:

- Contain letters or numbers only,
- Contain 15 or less characters,
- Start with a letter,
- Contain no reserved words,
- and be unique in the workspace.

4. Question 4

You need to create an embedded Spark capability that can reside alongside, and integrate with, other Azure data warehouse and integration services. Which one of the following solutions can provide these functionalities?

1 / 1 point

○ Azure HDInsight

◉ **Apache Spark for Azure Synapse**

○ Azure Databricks

○ 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. Question 5

You can use the magic command %%spark to execute which one of the following types of query against a spark context?

1 / 1 point

○ SparkSQL

○ Python

○ .Net for C#

◉ **Scala**

Correct

The magic command %%spark runs a Scala query against spark context.

6. Question 6

You are using IDE-style IntelliSense to assist with your code completion. Which two of the following languages does the IntelliSense support for Syntax Code Completion?

1 / 1 point

☐   .Net for Spark (C#)

☑   Spark (Scala)

Correct

The IntelliSense supports Syntax Code Completion for Spark (Scala).

☐   SparkSQL

☑   PySpark (Python)

Correct

The IntelliSense supports Syntax Code Completion for PySpark (Python).

7. Question 7

You run a notebook in Apache Spark pool for the first time. How many minutes will it take for Azure Synapse Analytics to create a new session?

1 / 1 point

○   10 - 15 Minutes

◉   3 - 5 Minutes

○   30 - 60 Seconds

Correct

The first time you run a notebook in a Spark pool, Synapse creates a new session. This can take approximately 3-5 minutes.

8. Question 8

What is a DataFrame?

1 / 1 point

☐   A parquet file

☐   A csv file

☑   A data structure

Correct

A DataFrame creates a data structure and it's one of the core data structures in Apache Spark.

9. Question 9

You need to load data from different file types into an Apache Spark DataFrame. Which four of the following Azure services can you load the data from?

1 / 1 point

☑   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.

☑   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.

☑   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.

☐ Serverless SQL Pool

☑ 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.

10. Question 10

You need to optimize the Apache Spark Jobs in Azure Synapse Analytics. However, first you need to consider the cluster configuration for the workload you're running on that cluster. Which one of the following resources can benefit from query optimization through Catalyst?

1 / 1 point

○ Resilient Distributed Datasets (RDDs)

◉ DataFrames

○ Notebooks

Correct

DataFrames provide query optimization through Catalyst.

Test prep

1. Question 1

The Azure Synapse Apache Spark to Synapse SQL connector is designed to efficiently transfer data between which of the following?

1 / 1 point

○ Dedicated Apache Spark pools and Serverless SQL pools in Azure Synapse

○ Serverless Apache Spark pools and Serverless SQL pools in Azure Synapse

◉ Serverless Apache Spark pools and Dedicated SQL pools in Azure Synapse

Correct

The Azure Synapse Apache Spark to Synapse SQL connector is designed to efficiently transfer data between serverless Apache Spark pools and dedicated SQL pools in Azure Synapse.

2. Question 2

You need to develop and execute transformation pipelines within the Azure Synapse Studio notebook experience. Which three of the following languages will help you in your task?

1 / 1 point

☐ JSON

☑ Scala

Correct

You can attach a SQL or Apache Spark pool, and develop and execute transformation pipelines using Spark (Scala).

☑ SparkSQL

Correct

You can attach a SQL or Apache Spark pool, and develop and execute transformation pipelines using SparkSQL.
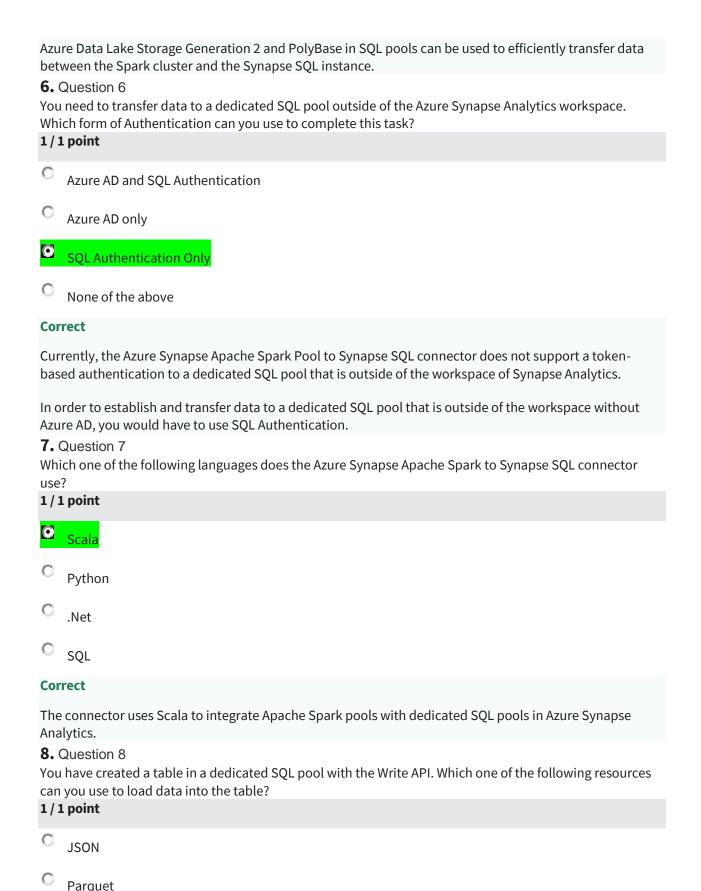
☑ Python

Correct

You can attach a SQL or Apache Spark pool, and develop and execute transformation pipelines using Python.

3. Question 3

Which two of the following prerequisites are necessary to successfully authenticate an account between two different systems?

1 / 1 point

☑ The account used needs to be a member of the Storage Blob Data Contributor role on the default storage account.

Correct

The account used needs to be a member of the Storage Blob Data Contributor role on the default storage account.

☐ 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.

☑ 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.

☐ The account used needs to be a member of the db_exporter role on the default storage account.

4. Question 4
In which of the following scenarios are you less likely to require import statements when transferring data between a dedicate SQL and Apache Spark pool?
0 / 1 point

◉ When relying on the use of token-based authentication.

○ When using the Azure integrated notebook experience.

○ When using the Apache Spark PySpark connector.

Incorrect

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

5. Question 5
Which of the following SQL pools can you use to efficiently transfer data between the Apache Spark cluster and the Azure Synapse SQL instance with the Azure Synapse Apache Spark pool to Synapse SQL connector?
1 / 1 point

◉ Azure Data Lake Storage Generation 2 and PolyBase

○ Azure Data Lake Storage Generation 2 and JSON

○ Azure Data Lake Storage Generation 2 and XML

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.

6. Question 6

You need to transfer data to a dedicated SQL pool outside of the Azure Synapse Analytics workspace. Which form of Authentication can you use to complete this task?

1 / 1 point

○ Azure AD and SQL Authentication

○ Azure AD only

◉ SQL Authentication Only

○ None of the above

Correct

Currently, the Azure Synapse Apache Spark Pool to Synapse SQL connector does not support a token-based authentication to a dedicated SQL pool that is outside of the workspace of Synapse Analytics.

In order to establish and transfer data to a dedicated SQL pool that is outside of the workspace without Azure AD, you would have to use SQL Authentication.

7. Question 7

Which one of the following languages does the Azure Synapse Apache Spark to Synapse SQL connector use?

1 / 1 point

◉ Scala

○ Python

○ .Net

○ SQL

Correct

The connector uses Scala to integrate Apache Spark pools with dedicated SQL pools in Azure Synapse Analytics.

8. Question 8

You have created a table in a dedicated SQL pool with the Write API. Which one of the following resources can you use to load data into the table?

1 / 1 point

○ JSON

○ Parquet

○ ORC

◉ **PolyBase**

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.

9. Question 9

Which two of the following Azure Synapse Studio tabs can you use to access the Apache Spark history server?

0.5 / 1 point

☐ Manage

☑ Develop

This should not be selected

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

☐ Monitor

☑ Data

Correct

To open the Apache Spark History server, you can navigate to the Data tab, where if you create a notebook and read a DataFrame you can go to the bottom of the page and find the Spark History Server known as the Spark UI.

10. Question 10

Which three of the following actions occur within existing nodes when you scale down Apache Spark pools?

1 / 1 point

☑ Nodes to be scaled down will be put in a decommissioned state.

Correct

Nodes will be placed in a decommissioned state.

☑ Pending jobs will be in a waiting state and scheduled for execution on fewer nodes.

Correct

When you scale down Apache spark pools, pending jobs will be placed in a waiting state and scheduled for execution on fewer nodes.

☐ Nodes to be scaled down will be shut down immediately regardless of current state.

☐ Pending jobs will be lost.

☑ Jobs that are still running will continue to run and finish.

Correct

Running jobs will continue until completion.

Course practice exam

Latest Submission Grade 93.33%

1. Question 1

Can you identify three features of Apache Spark?

1 / 1 point

☑ 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.

☑ Distributed execution engine

Correct

The Apache Spark core engine is a distributed execution engine.

☑ 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.

☐ Disk-based processing

2. Question 2

Which object is responsible for allocating resources across Apache Spark applications in Azure Synapse Analytics?

1 / 1 point

◉ Cluster Manager

○ SparkContext

○ Nodes

○ Executors

Correct

The SparkContext can connect to the cluster manager, which allocates resources across applications.

3. Question 3

Spark pools in Azure Synapse Analytics are compatible with which two of the following storage types?

1 / 1 point

☐ SQL Storage

☑ Azure Data Lake Generation 2 Storage

Correct

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

☐ Azure Data Lake Generation 1 Storage

☑ Azure Storage

Correct

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

4. Question 4

Which of the following solutions can you utilize to create and manage an end-to-end big data or data science project using one single platform?

1 / 1 point

○ Azure HDInsight

○ Apache Spark

◉ Azure Databricks

○ Apache Spark for Azure Synapse

Correct

Azure Databricks provides the ability to create and manage an end-to-end big data/data science project using one platform.

5. Question 5

What is the default language of a new cell in Azure Synapse Studio?

1 / 1 point

○ SQL

○ Scala

○ .NET for Spark

◉ PySpark

Correct

When adding a code cell, the default language is PySpark. However, other supported languages include Scala, SQL, and .NET for Spark.

6. Question 6

You deploy the magic command %%spark. What type of query will this command execute against Spark context?

1 / 1 point

○ Python

◉ Scala

○ SparkSQL

○ .Net for C#

Correct

The magic command %%spark runs a Scala query against spark context.

7. Question 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

○ Create a new Notebook.

○ Do Nothing. You can reference data or variables directly using different languages in an Azure Synapse Studio notebook.

○ Use a magic command for that language.

○ Create a temporary table so that it can be referenced across different languages.

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. Question 8

What are DataFrames?

1 / 1 point

☐ DataFrames are a collection of data organized into named Rows.

☑ 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.

☑ 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 columns.

Correct

DataFrames are a collection of data organized into named columns.

9. Question 9

You input the following Python snippet into your code:

new_rows = [('CA',22, 45000),("WA",35,65000) ,("WA",50,85000)]

demo_df = spark.createDataFrame(new_rows, ['state', 'age', 'salary'])

demo_df.show()

The variable named *demo_df* performs which of the following actions?

0 / 1 point

○ It uses the *spark.createDataFrame* method referencing the *new_rows* variable in the first parameter. The second parameter defines the column heading names for the DataFrame as *state*, *age*, and *salary*.

○ It uses the *spark.createDataFrame* method and creates a variable named *new_rows* which creates the data in the code segment to store in the DataFrame.

◉ It uses the *spark.createDataFrame* method to create a variable named *new_rows* which will store the values *state*, *age*, and *salary*.

Incorrect

Try going back and reviewing Big Data Engineering.

10. Question 10

You need to flatten nested structures and explode arrays with Apache Spark. What series of steps should you perform to complete these tasks?

1 / 1 point

○

1. Define a function
2. Flatten child nested schema
3. Flatten nested schema
4. Explode Arrays

○

1. Define a function
2. Flatten nested schema
3. Flatten child nested schema
4. Explode Arrays

◉

1. Define a function
2. Flatten nested schema
3. Explode Arrays
4. Flatten child nested schema

Correct

The correct order to flatten nested structures and explode arrays with Apache Spark are:

1. Define a function
2. Flatten nested schema
3. Explode Arrays
4. Flatten child nested schema

11. Question 11

You can use the Azure Synapse Apache Spark to Synapse SQL connector to transfer data between which of the following?

1 / 1 point

◉ Serverless Apache Spark pools and Dedicated SQL pools in Azure Synapse.

○ Serverless Apache Spark pools and Serverless SQL pools in Azure Synapse.

○ Dedicated Apache Spark pools and Serverless SQL pools in Azure Synapse.

Correct

The Azure Synapse Apache Spark to Synapse SQL connector is designed to efficiently transfer data between serverless Apache Spark pools and dedicated SQL pools in Azure Synapse.

12. Question 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.

☑ 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.

☑ 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.

13. Question 13

You have a requirement to transfer data to a dedicated SQL pool that is outside of the workspace of Synapse Analytics. Which form of Authentication can you use to complete this task?

1 / 1 point

◉ Azure AD and SQL Authentication.

○ SQL Authentication Only.

○ None of the above.

○ Azure AD only.

Correct

To establish and transfer data to a dedicated SQL pool that is outside of the workspace you can use SQL Authentication or Azure AD if available

14. Question 14

What is the minimum number of nodes allowed when creating an Apache Spark pool with Autoscaling?

1 / 1 point

○ 1

○ 4

◉ 3

○ 2

Correct

The minimum number of nodes allowed is 3.

15. Question 15

What three actions occur within existing nodes in Azure Synapse Analytics when you scale down Apache Spark pools?

1 / 1 point

☑ 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.

☐ Pending jobs will be lost.

☐ Nodes to be scaled down will be shut down immediately regardless of current state.

☑ 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.

☑ 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.