Course Practice Exam

Latest Submission Grade 80%

1. Question 1

Which of the following offerings in Synapse Analytics is suitable for unplanned or ad-hoc workloads?

1 / 1 point

○ Azure Synapse Pipelines

◉ Azure Synapse serverless SQL pools

○ Azure Synapse dedicated SQL pools

Correct

For unplanned or ad-hoc workloads, use the always-available, serverless SQL endpoint.

2. Question 2

Which of the following can be used to develop SQL Scripts and Notebooks?

0 / 1 point

◉ Azure Synapse Studio

○ Azure portal

◉ Data Lake

Incorrect

Try going back and reviewing *Azure Synapse Analytics*.

3. Question 3

A data warehouse provides which of the following?

1 / 1 point

○ A central repository of data stored in a nonrelational database.

○ A repository of data distributed across multiple instances of Azure SQL.

◉ A central repository of data stored in relational tables.

Correct

A data warehouse is a central repository of data stored in relational tables.

4. Question 4

Creating dedicated SQL pools to reserve processing power for data permanently stored in SQL tables in order to ensure predictable performance and cost is a type of model applied to which of the following?

1 / 1 point

☑ **Data warehousing**

○ Data virtualization

The dedicated model is referred to as dedicated SQL Pools. It refers to the data warehousing features that are generally available in Azure Synapse Analytics.

Dedicated SQL pools represent a collection of analytic resources that are being provisioned when using Synapse SQL.

When you need predictable performance and cost, creating dedicated SQL pools to reserve processing power for data permanently stored in SQL tables in a data warehouse is the best approach to take.

5. Question 5
A Spark Pool Cluster is created from which two of the following components?
1 / 1 point

☐ Multiple Spark drivers

☑ **Multiple Worker Nodes**

A Spark pool cluster consists of multiple Worker nodes.

☐ A single worker node

☑ **A single Spark Driver**

A Spark pool cluster consists of a single Spark Driver.

6. Question 6
You are working in Azure Synapse Analytics, and you have a requirement to ingest data from a data source in readiness to prepare the data for transformation and analysis. Which feature should you implement to enable the ingestion of the data from a data source?
1 / 1 point

○ SQL Pool

☑ **Linked Service**

○ Datasets

○ Spark pool

Linked services are much like connection strings, which define the connection information needed to connect to external resources. For example, to copy data from Blob storage to a SQL Database, you create two linked services; one for Azure Storage and one Azure SQL Database.

7. Question 7

What term do we use to refer to the ability to perform analytics over a database system that is seen to provide transactional capabilities without impacting the performance of the system itself?

1 / 1 point

○ Data Virtualization

⊙ Hybrid Transactional and Analytical processing (HTAP)

○ Data Visualization

○ Data Warehousing

Correct

Hybrid Transactional and Analytical Processing enables businesses to perform analytics over a database system that is seen to provide transactional capabilities without impacting the performance of the system.

This enables organizations to use a database to fulfill both transactional and analytical needs to support near real-time analysis of operational data to make decisions about the information that is being analyzed.

8. Question 8

Which Azure Synapse Studio hub would you go to create Notebooks?

0 / 1 point

○ Data hub

⊙ Manage hub

⊙ Integrate hub

○ Develop hub

Incorrect

Try going back and reviewing *Explore Azure Synapse Studio*.

9. Question 9

Which three of the following scenarios would be a valid reason for adding a staging area into the architecture of a modern data warehouse?

1 / 1 point

☐ To make data analytical available directly from the staging area.

☑ To reduce contention on source systems.

Correct

Feedback: Source systems typically play an important role in fulfilling business operations and may be mission critical.

Some data warehouse design strategies will involve grabbing data at source, and "dumping" the data into a staging area to minimize the resource usage against the source system.

☑ To enable the ingestion of source systems based on different schedules.

Correct

Staging environments provide a great place to store data from different source systems regardless of the schedule on which the data is ingested.

☑ To join data from different source systems.

Correct

A staging environment provides the opportunity to bring together a single view of data from different source systems.

As the staging area is independent from the source systems and the data warehouse, you have the freedom to perform any work you need without impacting these systems.

10. Question 10

Which four of the following data formats are natively supported by Synapse Analytics when ingesting raw data in batch from new data sources?

Choose all that apply

1 / 1 point

☑ Parquet

Correct

Parquet format is natively supported by Synapse Analytics.

☑ JSON

Correct

JSON format is natively supported by Synapse Analytics.

☑ ORC

Correct

ORC format is natively supported by Synapse Analytics.

☐ Scala

☑ CSV

Correct

CSV Format is natively supported by Synapse Analytics.

Course practice exam

Latest Submission Grade 100%

1. Question 1

Processing data that arrives in real-time/near real-time is also referred to as streaming data processing. Azure offers purpose-built stream ingestion services such as Azure IoT Hub and Azure Event Hubs. To collect messages from these or similar services, and process them you can use which of the following features?

Select all options that apply.

1 / 1 point

☑ Azure Functions

Correct

In your data pipeline, you can collect messages from these services, and process them using Azure Functions.

☑ Azure Stream Analytics

Correct

In your data pipeline, you can collect messages from these services, and process them using Azure Stream Analytics.

☐ Azure IoT Central

☑ Azure Databricks

Correct

In your data pipeline, you can collect messages from these services, and process them using Azure Databricks.

2. Question 2

Which transformation in the Mapping Data Flow is used to routes data rows to different streams based on matching conditions?

1 / 1 point

○ GetMetadata activity

○ Lookup

◉ Conditional Split

Correct

A Conditional Split transformation routes data rows to different streams based on matching conditions. The conditional split transformation is similar to a CASE decision structure in a programming language.

3. Question 3

When working with Data Lake Storage Gen2 many small files can negatively affect performance. The recommended file size for Data Lake Storage Gen2 is between which of the following sizes?

1 / 1 point

○ 256MB to 1GB

◉ **256MB to 100GB**

○ 10GB to 100GB

○ 1GB to 10GB

Correct

If you store your data as many small files, this can negatively affect performance. In general, organize your data into larger-sized files for better performance (256 MB to 100 GB in size).

4. Question 4

What is the difference between a star schema and a snowflake schema?

1 / 1 point

○ A star schema uses surrogate keys while a snowflake schema uses business keys.

○ All dimensions in a star schema are normalized while all dimensions in a snowflake schema join directly to the fact table (denormalized).

○ A star schema has one fact table while a snowflake schema has multiple fact tables.

◉ **All dimensions in a star schema join directly to the fact table (denormalized) while some dimension tables in a snowflake schema are normalized.**

Correct

A star schema is highly denormalized so that the fact table joins directly to dimensions; a snowflake schema normalizes some dimensions into multiple tables such as DimProduct, DimProductSubcategory, and DimProductCategory.

5. Question 5

Which of the following statements are true in respect of a Star schema?

Select all options that apply.

1 / 1 point

☐ In a Star schema Cube processing might be slow because of the complex join.

☑ **Star schema dimension tables are a denormalized data structure.**

Correct

Star schema dimension tables are not normalized; snowflake schemas dimension tables are normalized.

☑ Star schemas have a high level of Data redundancy.

Correct

Star schemas have a high level of Data redundancy.

☑ A Star schema contains a fact table surrounded by dimension tables.

Correct

A Star schema contains a fact table surrounded by dimension tables.

☐ A Star schema will have a fact table surrounded by dimension tables which are in turn surrounded by dimension tables.

6. Question 6

A Star schema is a modelling approach widely adopted by relational data warehouses. It requires modelers to classify their model tables as either dimension or fact. Which of the following are features of dimension tables?

Select all options that apply.
1 / 1 point

☐ A dimension stores numeric measure columns

☑ A dimension table contains a key column (or columns)

Correct

A dimension table contains a key column (or columns) that acts as a unique identifier, and descriptive columns.

☑ A dimension table describes business entities

Correct

Dimension tables describe business entities—the things you model. Entities can include products, people, places, and concepts including time itself.

7. Question 7

In Azure Synapse Analytics configured with a dedicated SQL pool, classifiers assign incoming requests to a workload group. Classifiers are evaluated with every request submitted. If a request is not matched to a classifier, it is assigned to the default workload group. The default workload group in is the _____ resource class.
1 / 1 point

○ largerc

○ xlargerc

○ mediumrc

⊙ smallrc

Correct

The default workload group in an SQL pool is the smallrc resource class.

8. Question 8

SQL Pools have the concept of concurrency slots, which manage the allocation of memory to connected users. Which of the following should you consider to optimize the load execution operations?

Select all options that apply.

1 / 1 point

☑ Assigning higher resource classes that reduce the number of active running tasks.

Correct

To optimize the load execution operations, you should consider assigning higher resource classes that reduce the number of active running tasks.

☐ Increase the number of simultaneous load jobs that are running.

☐ Assigning lower resource classes that reduce the number of active running tasks.

☑ Reducing or minimizing the number of simultaneous load jobs that are running.

Correct

To optimize the load execution operations, you should consider reducing or minimizing the number of simultaneous load jobs that are running.

9. Question 9

Which of the following hubs would you use to create a workload classifier in Azure Synapse Studio?

1 / 1 point

○ Data

⊙ Develop

○ Manage

○ Integrate

Correct

The Develop hub is used to create a workload classifier in Azure Synapse Studio.

10. Question 10

When data is loaded into Synapse Analytics dedicated SQL pools, the datasets are broken up and dispersed among the compute nodes for processing, and then written to a decoupled and scalable storage layer. What term is used to describe this action?

1 / 1 point

○ Sharing

○ Separating

◉ Sharding

○ Shredding

Correct

This action is termed Sharding.

11. Question 11

Select from the following options to complete the missing word in the sentence.

Dedicated SQL Pools create a _____ index when no index options are specified on a table.

1 / 1 point

◉ Clustered columnstore

○ Clustered

○ Non-clustered

Correct

Dedicated SQL Pools create a clustered columnstore index when no index options are specified on a table.

12. Question 12

Which of the following index types can be defined on a table or view with a clustered index or on a heap?

1 / 1 point

◉ Non-clustered

○ Clustered

○ Clustered columnstore

Correct

A non-clustered index can be defined on a table or view with a clustered index or on a heap. Each index row in the non-clustered index contains the non-clustered key value and a row locator.

13. Question 13

In Azure Synapse SQL, you should enable result-set caching when you expect results from queries to return the same values. This option stores a copy of the result set on the control node so that queries do not need to pull data from the storage subsystem or compute nodes.

Materialized views are prewritten queries with joins and filters whose definition is saved and the results persisted to pools. Which of the following pools are the results for Materialized views persisted to?

1 / 1 point

🔘 48 Hours

⚪ 12 Hours

⚪ 24 Hours

⚪ 36 Hours

Correct

By default, data within the result-set cache is expired and purged by the dedicated SQL pool after 48 hours of not being accessed.

14. Question 14

The Azure Synapse Studio experience provides an integrated notebook experience. Within this notebook experience, you can attach a SQL or Apache Spark pool, and develop and execute transformation pipelines using which of the following?

1 / 1 point

☐ SparkSQL

☑ Python

Correct

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

☐ JSON

☐ Scala

15. Question 15

In Azure Synapse Analytics the authentication process between two systems can be seamless. However, there are some prerequisites. Which of the following role memberships are required to successfully authenticate?

Select all options that apply.

1 / 1 point

☑ The account used needs to be a member of db_exporter role in the database or SQL pool from which you 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 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

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 transfer data to or from.

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

16. Question 16

To write data to a dedicated SQL Pool, you use the Write API. The Write API creates a table in the dedicated SQL pool. Which of the following is used to load the data into the table that was created?

1 / 1 point

○ Parquet

○ ORC

◉ PolyBase

○ 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.

17. Question 17

What would be the best approach to investigate if the data at hand is unevenly allocated across all distributions?

1 / 1 point

○ Grouping the data based on partitions and counting rows with a T-SQL query.

◉ Using DBCC PDW_SHOWSPACEUSED to see the number of table rows that are stored in each of the 60 distributions.

○ Monitor query speeds by testing the same query for each partition.

Correct

DBCC PDW_SHOWSPACEUSED returns the number of table rows that are stored in each of the 60 distributions.

18. Question 18

Which of the following statements is a benefit of materialized views?

1 / 1 point

◉ Reducing the execution time for complex queries with JOINs and aggregate functions.

○ Increased resiliency benefits

○ Increased high availability

Correct

Materialized views help to improve complex query performance. The more complex the query, the higher the potential for execution-time saving.

19. Question 19

Conditional access is a feature that enables you to define the conditions under which a user can connect to your Azure subscription and access services. Conditional access policies use signals as a basis to determine if conditional access should first be applied. Common signals include?

Select all options that apply.

1 / 1 point

☐ Multi Factor Authentication

☑ Microsoft Cloud App Security (MCAS)

Correct

Enables user application access and sessions to be monitored and controlled in real time, increasing visibility and control over access to and activities performed within your cloud environment.

☑ User or group membership names

Correct

Policies can be targeted to specific users and groups giving administrators fine-grained control over access.

☑ IP address information

Correct

Organizations can create trusted IP address ranges that can be used when making policy decisions.

☑ Device platforms or type

Correct

Users with devices of specific platforms or marked with a specific state can be used when enforcing Conditional Access policies.

20. Question 20

You want to configure a private endpoint. You open up Azure Synapse Studio, go to the manage hub, and see that the private endpoint is greyed out. Why is the option not available?

1 / 1 point

⦿ A managed virtual network has not been created.

○ Azure Synapse Studio does not support the creation of private endpoints.

○ A conditional access policy has to be defined first.

Correct

In order to create a private endpoint, you first must create a managed virtual network.

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. <mark>Define a function</mark>
2. <mark>Flatten nested schema</mark>
3. <mark>Explode Arrays</mark>
4. <mark>Flatten child nested schema</mark>

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

◉ <mark>Serverless Apache Spark pools and Dedicated SQL pools in Azure Synapse.</mark>

○ 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.

☑ <mark>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.</mark>

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.