

# Knowledge check

Total points 6

## 1. Question 1

The process of building a modern data warehouse typically includes Data Ingestion and Preparation.

You have recently deployed Azure Synapse Analytics. You now have a requirement to ingest data code-free. Which of the following tools can be used to perform this task?

1 / 1 point



Azure Data Factory



Azure Databricks



Power BI

Correct

Data Factory empowers customers to do code-free ETL/ELT, including preparation and transformation.

## 2. Question 2

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

Select all options that apply.

1 / 1 point



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 make data analytics available directly from the staging area



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.



To reduce contention on source systems

Correct

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.

### 3. Question 3

When ingesting raw data in batch from new data sources, which of the following data formats are natively supported by Synapse Analytics?

Select all options that apply.

**1 / 1 point**

☒ ORC

**Correct**

ORC format is natively supported by Synapse Analytics.

☒ Parquet

**Correct**

Parquet format is natively supported by Synapse Analytics.

☐ Scala

☒ JSON

**Correct**

JSON format is natively supported by Synapse Analytics.

☒ CSV

**Correct**

CSV Format is natively supported by Synapse Analytics.

### 4. Question 4

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 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.



Azure Functions

**Correct**

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

**5. Question 5**

Which technology is typically used as a staging area in a modern data warehousing architecture?

**1 / 1 point**



Azure Synapse Spark Pools.



Azure Data Lake



Azure Synapse SQL Pools

**Correct**

Azure Data Lake Store Gen 2 is the technology that will be used to stage data before loading it into the various components of Azure Synapse Analytics.

**6. Question 6**

Which of the following is a Big Data Solution that stores data in a relational table format with columnar storage?

**1 / 1 point**



Azure Synapse SQL Pools



Azure Synapse Spark Pools

**Correct**

SQL Pool is the traditional Data Warehouse. It was formerly known as Azure SQL Data Warehouse before it came under the Synapse Family. It is a Big Data Solution that stores data in a relational table format with columnar storage.

# Knowledge check

Total points 6

## 1. Question 1

Which component enables you to perform code free transformations in Azure Synapse Analytics?

1 / 1 point



Synapse Studio



Synapse Mapping data flow



Synapse Copy activity

**Correct**

You can natively perform data transformations with Azure Data Factory code free using the Mapping Data Flow task.

## 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



Conditional Split



Lookup

**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

Which transformation is used to load data into a destination data store or compute resource?

1 / 1 point



Window



Source



Sink

**Correct**

A Sink transformation allows you to choose a dataset definition for the destination output data. You can have as many sink transformations as your data flow requires.

#### 4. Question 4

True or False

When data is stored in Data Lake Storage Gen2, the file size, number of files, and folder structure can have an impact on performance.

1 / 1 point

☐ False

☒ True

**Correct**

The file size, number of files, and folder structure in Data Lake Storage Gen2 can have an impact on performance. 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.

#### 5. Question 5

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

☐ 1GB to 10GB

☐ 10GB to 100GB

☒ 256MB to 100GB

☐ 256MB to 1GB

**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).

#### 6. Question 6

When building data flows in Azure Synapse you can enable debug mode., When Debug mode is enabled Synapse automatically turns on which of the following?

1 / 1 point

☐ Serverless cluster

☒ Spark cluster

☐ Dedicated SQL Pool

**Correct**

When building data flows in Azure Synapse you can enable debug mode, enabling Debug mode turns on a small interactive Spark cluster

## Knowledge check

Total points 6

### 1. Question 1

A Star schema is a modeling 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 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.

☐ 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.

### 2. Question 2

Which of the following are true in respect of fact tables?

Select all options that apply.

0.5 / 1 point

☒ A fact table describes business entities

**This should not be selected**

A fact table does not describe business entities.

- ☒ A fact table contains numeric measure columns

**Correct**

A fact table contains numeric measure columns.

- ☒ A fact table contains dimension key columns that relate to dimension tables.

**Correct**

A fact table contains dimension key columns that relate to dimension tables.

- ☒ Fact tables store observations or events

**3. Question 3**

Since Synapse Analytics is a massively parallel processing (MPP) system, you need to consider how data is distributed in your table design. What is the recommended distribution option for Fact tables?

**1 / 1 point**

- ☐ Replicate
- ☐ Clustered Index
- ☐ Clustered Columnstore Index

- ☒ Hash-distribution

**Correct**

For Fact tables the recommendation is to use hash-distribution with clustered columnstore index. Performance improves when two hash tables are joined on the same distribution column.

**4. Question 4**

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

Select all options that apply.

**0.6 / 1 point**

- ☒ Star schemas have a high level of Data redundancy.

**Correct**

Star schemas have a high level of Data redundancy.

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

**This should not be selected**

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

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

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

**Correct**

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

☒ Star schema dimension tables are a denormalized data structure.

### 5. Question 5

Examine the following statement and select the missing word with an entry from those supplied below:

A time dimension table is one of the most consistently used dimension tables. This type of table enables consistent \_\_\_\_\_ for temporal analysis and reporting

**1 / 1 point**

☐ uniqueness

☐ indexing

☐ distribution

☒ granularity

**Correct**

A time dimension table is one of the most consistently used dimension tables. This type of table enables consistent granularity for temporal analysis and reporting.

### 6. Question 6

What distribution option would be best for a sales fact table that will contain billions of records?

**1 / 1 point**

☐ DISTRIBUTION = HEAP

☐ DISTRIBUTION = REPLICATE

☒ DISTRIBUTION = HASH([SalesOrderNumber])

**Correct**

Hash distribution provides good read performance for a large table by distributing records across compute nodes based on the hash key.



# Knowledge check

Total points 6

## 1. Question 1

Which Workload Management capability manages minimum and maximum resource allocations during peak periods?

1 / 1 point



Workload Importance



Workload Isolation



Workload Containment

Correct

Workload Isolation assigns maximum and minimum usage values for varying resources under load. These adjustments can be done live without having to take the SQL Pool offline.

## 2. Question 2

Select from the options below to complete the missing text in the following statement.

A data warehouse that is built on a Massively Parallel Processing (MPP) system are built for processing and analyzing large datasets. As such they perform well with \_\_\_\_\_ that can be distributed across compute nodes and storage.

Select from the options to complete the missing text.

1 / 1 point



Fewer and larger batch sizes



Multiple small batch sizes

Correct

A data warehouse that is built on a Massively Parallel Processing (MPP) system are built for processing and analyzing large datasets. As such they perform well with larger batch type loads and updates that can be distributed across the compute nodes and storage.

## 3. Question 3

Resource classes are pre-determined resource limits in Synapse SQL pool that govern compute resources and concurrency for query execution. Resource classes can help you configure resources for your queries by setting limits on the number of queries that run concurrently and, on the compute-resources assigned to each query.

Which of the following statements are correct?

Select all options that apply.

**1 / 1 point**

- ☐ Larger resource classes increase concurrency, but reduce the maximum memory per query,
- ☒ Larger resource classes increase the maximum memory per query but reduce concurrency.

**Correct**

Larger resource classes increase the maximum memory per query but reduce concurrency.

- ☒ Smaller resource classes reduce the maximum memory per query but increase concurrency.

**Correct**

Smaller resource classes reduce the maximum memory per query but increase concurrency.

- ☐ Smaller resource classes reduce the concurrency but increase maximum memory per query.

#### 4. Question 4

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

Select all options that apply.

**1 / 1 point**

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

- ☒ 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.

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

#### 5. Question 5

In Synapse SQL pools workload importance influences the order in which a request gets access to resources. There are five levels of importance. Which of the following are valid levels of importance?

Select all options that apply.

**1 / 1 point**

☒ below\_normal

**Correct**

Below\_normal is a valid level of importance.

☒ low

**Correct**

Low is a valid level of importance.

☒ above\_normal

**Correct**

Above\_normal is a valid level of importance.

☒ high

**Correct**

High is a valid level of importance.

☐ very high

☒ normal

**Correct**

Normal is a valid level of importance.

## 6. Question 6

How does splitting source files help maintain good performance when loading into Synapse Analytics?

**1 / 1 point**

☒ Compute node to storage segment alignment

☐ Optimized processing of smaller file sizes

☐ Reduced possibility of data corruptions

**Correct**

SQL Pools have 60 storage segments. Compute can also scale to 60 nodes and so optimizing for alignment of these two resources can dramatically decrease load times.

# Knowledge check

Total points 6

## 1. Question 1

Azure Synapse Analytics is a high performing Massively Parallel Processing (MPP) engine that is built with loading and querying large datasets in mind.

You have received calls from users reporting that the data in the reports they are producing appears to be out of date. Which of the following is the most likely to cause of out-of-date information being presented in user reports?

0 / 1 point



Low concurrency



Poor query performance



Poor load performance

**Incorrect**

You may receive reports from your users that they may be unable to connect to the data warehouse to execute reports or queries due to low concurrency.

## 2. Question 2

What are the three main table distributions available in Synapse Analytics SQL Pools called?

Select all options that apply.

1 / 1 point



Round robin distribution

**Correct**

Round robin distribution is the default distribution created for a table and delivers fast performance when used for loading data.



Replicated tables

**Correct**

A replicated table provides the fastest query performance for small tables.



Hash distribution

**Correct**

This distribution can deliver the highest query performance for joins and aggregations on large tables.



Block Distribution

### 3. Question 3

True or False

When a table is created, by default the data structure has no indexes and is called a heap

1 / 1 point

☒ True

☐ False

**Correct**

When a table is created, by default the data structure has no indexes and is called a heap.

### 4. Question 4

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

☐ Non-clustered

☒ Clustered columnstore

☐ Clustered

**Correct**

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

### 5. Question 5

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

☒ Dedicated SQL Pool

☐ Serverless SQL Pool

**Correct**

Materialized views are prewritten queries with joins and filters whose definition is saved and the results persisted to a dedicated SQL pool.

### 6. Question 6

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.

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

**1 / 1 point**

- ☐ 24 Hours
- ☐ 36 Hours
- ☒ 48 Hours
- ☐ 12 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.

## Knowledge check

**Total points** 6

### 1. Question 1

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

Select all options that apply.

**1 / 1 point**

- ☐ YAML
- ☒ Parquet

**Correct**

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

☒ TSV

**Correct**

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

☒ JSON

**Correct**

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

☒ CSV

**Correct**

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

## 2. Question 2

The Azure Synapse Apache Spark pool to Synapse SQL connector is a data source implementation for Apache Spark. Which of the following 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 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.

## 3. Question 3

True or False

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

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**

- ☐ ORC
- ☒ PolyBase
- ☐ JSON
- ☐ Parquet

**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**

In what language can the Azure Synapse Apache Spark to Synapse SQL connector be used?

**1 / 1 point**

- ☐ Python
- ☒ Scala
- ☐ .Net
- ☐ SQL

**Correct**

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

**6. Question 6**

When is it unnecessary to use import statements for transferring data between a dedicated SQL and Apache Spark pool?

**1 / 1 point**

- ☒ When using the integrated notebook experience from Azure Synapse Studio.
- ☐ When using the PySpark connector.
- ☐ Use token-based authentication.

**Correct**

Import statements are not needed since they are pre-loaded in case you use the Azure Synapse Studio integrated notebook experience.



## Knowledge check

Total points 6

### 1. Question 1

The Develop hub in Azure Synapse Studio is an interface you can use for developing a variety of solutions against an Azure Synapse Analytics instance. In this area, you can create which of the following objects?

Select all options that apply.

1 / 1 point

☒ Power BI datasets and reports

Correct

You can create Power BI datasets and reports.

☒ SQL Scripts

Correct

You can create SQL Scripts.

☒ Notebooks

Correct

You can create Notebooks.

☐ Synapse Workspace

☒ Azure Synapse Pipelines

Correct

You can create Synapse Pipelines.

### 2. Question 2

Visual Studio 2019 SQL Server Data Tools (SSDT) has which of the following features?

Select all options that apply.

0.75 / 1 point

☒ integrate with source control systems

**Correct**

SSDT is a single tool and can integrate with source control systems.

- ☐ Native integration with Azure DevOps
- ☐ Create Database projects within Serverless SQL pools

☒ Create Database projects within dedicated SQL pools

**Correct**

SSDT is a single tool that allows users to create Database projects within dedicated SQL pools.  
You didn't select all the correct answers

### 3. Question 3

Azure Synapse Analytics supports querying both relational and non-relational data using Transact SQL. The Azure Synapse SQL query language supports different features based on the resource model being used. Which of the following T-SQL Statements are supported on both Dedicated and Serverless Pools?

Select all options that apply.

**0.75 / 1 point**

☐ Cross database queries

☒ Transactions

**Correct**

Azure Synapse Analytics supports transactions on both Dedicated and Serverless Pools.

☒ SELECT statement

**Correct**

Azure Synapse Analytics supports the Select statement on both Dedicated and Serverless Pools.

☒ INSERT statement

**This should not be selected**

Azure Synapse Analytics does not support the Insert statement on Serverless Pools.

### 4. Question 4

Examine the following statement and select from the options below to complete the sentence.

Synapse dedicated SQL Pools supports JSON format data to be stored using standard \_\_\_\_\_ table columns

**1 / 1 point**

☐ VARCHAR

☒ NVARCHAR

☐ CHAR

☐ TEXT

**Correct**

Synapse dedicated SQL Pools support JSON format data to be stored using NVARCHAR table columns.

**5. Question 5**

What Transact-SQL function is used to perform a HyperLogLog function?

**1 / 1 point**

☐ COUNT\_DISTINCT\_APPROX

☐ COUNT

☒ APPROX\_COUNT\_DISTINCT

**Correct**

The APPROX\_COUNT\_DISTINCT function is used to perform a HyperLogLog function.

**6. Question 6**

In Azure Synapse Studio Develop hub you can define Spark Job definitions. Which of the following languages can be used to define job definitions?

Select all options that apply. <https://docs.microsoft.com/en-us/azure/synapse-analytics/spark/apache-spark-job-definitions>

**0.5 / 1 point**

☒ Transact-SQL

**This should not be selected**

Transact-SQL is a language used to query modify databases and data.

☒ Scala

**Correct**

You can define Spark Job definitions using Scala.

☒ .NET Spark

**Correct**

You can define Spark Job definitions using .Net Spark.

☒ PySpark

**Correct**

You can define Spark Job definitions using PySpark.

## Knowledge check

**Total points** 6

### 1. Question 1

In a dedicated SQL pool in Azure Synapse Analytics a distributed table appears as a single table with rows spread across multiple distributions. Across how many distributions are rows stored?

**1 / 1 point**

☐ 40

☐ 20

☐ 80

☒ 60

**Correct**

There are always 60 distributions.

### 2. Question 2

Examine the following statement and select from the listed options to complete the sentence.

A columnstore index scans a table by scanning column segments of individual rowgroups. Maximizing the number of rows in each rowgroup enhances query performance. For best query performance, the goal is to maximize the number of rows per rowgroup in a columnstore index.

Columnstore indexes achieve good performance when rowgroups have at least \_\_\_\_\_ rows.

<https://ellynguyen.wordpress.com/2019/02/12/sql-server-column-store-indexes/>

0 / 1 point

☒ 100,000 rows

☐ 10,000 rows

☐ 1,048,576 rows

☒ 1,000,000 rows

**Incorrect**

1,000,000 rows are not the optimum number of rows.

### 3. Question 3

SQL pool in Azure Synapse supports standard and materialized views. Which of the following are features of Materialized views?

Select all options that apply.

1 / 1 point

☐ Speed to retrieve view data from complex queries is Slow.

☒ View content is pre-processed and stored in SQL pool during view creation. The view is updated as data is added to the underlying tables.

**Correct**

View content is pre-processed and stored in SQL pool during view creation. The view is updated as data is added to the underlying tables.

☐ View content is generated each time the view is used.

☒ Speed to retrieve view data from complex queries is Fast.


**Correct**

Speed to retrieve view data from complex queries is Fast.

### 4. Question 4

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

1 / 1 point

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

**Correct**

DBCC PDW\_SHOWSPACEUSED returns the number of table rows that are stored in each of the 60 distributions.

**5. Question 5**

To achieve improved query performance, which of the following would be the best data type for storing data that contains less than 128 characters?

**1 / 1 point**

- ☐ NVARCHAR(128)
- ☐ VARCHAR(MAX)

 VARCHAR(128)

**Correct**

Limiting the size of the data type and not using size variability will provide the best performance.

**6. Question 6**

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

**1 / 1 point**

- ☐ Increased resiliency benefits
- ☐ Increased high availability

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

**Correct**

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

## Knowledge check

Total points 6

### 1. Question 1

In a dedicated SQL pool in Azure Synapse Analytics a distributed table appears as a single table with rows spread across multiple distributions. Across how many distributions are rows stored?

1 / 1 point

- ☐ 40
- ☐ 20
- ☐ 80

☒ 60

Correct

There are always 60 distributions.

### 2. Question 2

Examine the following statement and select from the listed options to complete the sentence.

A columnstore index scans a table by scanning column segments of individual rowgroups. Maximizing the number of rows in each rowgroup enhances query performance. For best query performance, the goal is to maximize the number of rows per rowgroup in a columnstore index.

Columnstore indexes achieve good performance when rowgroups have at least \_\_\_\_\_ rows.

0 / 1 point

☒ 100,000 rows

- ☐ 10,000 rows
- ☐ 1,048,576 rows

☐ 1,000,000 rows

Incorrect

1,000,000 rows is not the optimum number of rows.

### 3. Question 3

SQL pool in Azure Synapse supports standard and materialized views. Which of the following are features of Materialized views?

Select all options that apply.

1 / 1 point

☐ Speed to retrieve view data from complex queries is Slow.

☒ View content is pre-processed and stored in SQL pool during view creation. The view is updated as data is added to the underlying tables.

Correct

View content is pre-processed and stored in SQL pool during view creation. The view is updated as data is added to the underlying tables.

☐ View content is generated each time the view is used.

☒ Speed to retrieve view data from complex queries is Fast.

Correct

Speed to retrieve view data from complex queries is Fast.

### 4. Question 4

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

1 / 1 point

☒ 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.

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

Correct

DBCC PDW\_SHOWSPACEUSED returns the number of table rows that are stored in each of the 60 distributions.

### 5. Question 5

To achieve improved query performance, which of the following would be the best data type for storing data that contains less than 128 characters?

1 / 1 point

☐ NVARCHAR(128)



☐ VARCHAR(MAX)

☒ VARCHAR(128)

**Correct**

Limiting the size of the data type and not using size variability will provide the best performance.

**6. Question 6**

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

**1 / 1 point**

☐ Increased resiliency benefits

☐ Increased high availability

☒ Reducing the execution time for complex queries with JOINS and aggregate functions.

**Correct**

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

## Knowledge check

**Total points 6**

**1. Question 1**

What features are provided when using a managed workspace virtual network?

Select all options that apply.

**0.75 / 1 point**



You will need to create a subnet for your Spark clusters based on peak load.

☒ You don't have to configure inbound NSG rules on your own Virtual Networks to allow Azure Synapse management traffic to enter your Virtual Network.

**Correct**

When using a managed workspace virtual network you don't have to configure inbound NSG rules on your own Virtual Networks.

☐ Management of the virtual network is offloaded to Azure Synapse.

☒ Your workspace is network isolated from other workspaces.

**Correct**

When using a managed workspace virtual network your workspace is network isolated from other workspaces.

You didn't select all the correct answers

## 2. Question 2

Azure Synapse Analytics enables you to connect to its various components through endpoints. You can set up managed private endpoints to access these components in a secure manner known as private links. Which of the following statements are true in respect of Private Endpoints?

Select all options that apply.

**0.75 / 1 point**

☐ You must have an Azure Synapse workspace with a Managed workspace Virtual Network.

☒ When you use a private link, traffic between your Virtual Network and workspace traverses entirely over the Microsoft backbone network.

**Correct**

Traffic between your Virtual Network and workspace traverses entirely over the Microsoft backbone network.

☒ You can manage the private endpoints in the Azure Synapse Studio manage hub.

**Correct**

You can manage the private endpoints in the Azure Synapse Studio manage hub.

☐ When you use a private link, traffic between your Virtual Network and workspace traverses over the public Internet network.

You didn't select all the correct answers

## 3. Question 3

When can you choose to enable managed virtual networks?

**1 / 1 point**

☐ When creating a new Workspace or modifying an existing Workspace.

☒ Only when you are creating a new Azure Synapse Workspace.

☐ At any time for an existing Workspace.

**Correct**

You can only choose to enable managed virtual networks as you are creating the Azure Synapse Workspaces.

**4. Question 4**

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**

☒ IP address information

**Correct**

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

☒ 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.

☐ Multi Factor Authentication

☒ User or group membership names

**Correct**

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

☒ 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.

**5. Question 5**

You work at a bank as a service representative in a call center. Due to compliance, any caller must identify themselves by providing several digits of their credit card number. In this scenario, the full credit card number should not be fully exposed to the service representative in the call center. To limit visibility, so

that you would have a query that only gives as a result the last four digits of the credit card number, which of the following would you implement in Azure Synapse Analytics?

1 / 1 point



Dynamic Data Masking



Column Level Security



Row Level Security (RLS)

**Correct**

Dynamic Data Masking ensures limited data exposure to non-privileged users, such that they cannot see the data that is being masked.

**6. Question 6**

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 conditional access policy has to be defined first.



A managed virtual network has not been created.



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

**Correct**

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

**Test prep**

**Latest Submission Grade 100%**

### 1. Question 1

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

Select all options that apply.

1 / 1 point

☒ To reduce contention on source systems

Correct

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 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.

☐ To make data analytical available directly from the staging area

☒ 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.

### 2. Question 2

When ingesting raw data in batch from new data sources, which of the following data formats are natively supported by Synapse Analytics?

Select all options that apply.

1 / 1 point

☒ ORC

Correct

ORC format is natively supported by Synapse Analytics.

☒ Parquet

Correct

Parquet format is natively supported by Synapse Analytics.

☒ JSON

Correct

JSON format is natively supported by Synapse Analytics.

☒ CSV

**Correct**

CSV format is natively supported by Synapse Analytics.

☐ Scala

### 3. Question 3

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 IoT Central

☒ Azure Stream Analytics

**Correct**

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

☒ Azure Databricks

**Correct**

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

### 4. Question 4

Which technology is typically used as a staging area in a modern data warehousing architecture?

**1 / 1 point**

☐ Azure Synapse SQL Pools

☒ Azure Data Lake

☐ Azure Synapse Spark Pools.

**Correct**

Azure Data Lake Store Gen 2 is the technology that will be used to stage data before loading it into the various components of Azure Synapse Analytics.

**5. Question 5**

Which of the following is a Big Data Solution that stores data in a relational table format with columnar storage?

**1 / 1 point**



Azure Synapse SQL Pools



Azure Synapse Spark Pools

**Correct**

SQL Pool is the traditional Data Warehouse. It was formerly known as Azure SQL Data Warehouse before it came under the Synapse Family. It is a Big Data Solution that stores data in a relational table format with columnar storage.

**6. Question 6**

Which component enables you to perform code free transformations in Azure Synapse Analytics?

**1 / 1 point**



Synapse Mapping data flow



Synapse Studio



Synapse Copy activity

**Correct**

You can natively perform data transformations with Azure Data Factory code free using the Mapping Data Flow task.

**7. Question 7**

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

**1 / 1 point**



Lookup



GetMetadata activity



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.

**8. Question 8**

Which transformation is used to load data into a destination data store or compute resource?

**1 / 1 point**

☐ Source

☒ Sink

☐ Window

**Correct**

A Sink transformation allows you to choose a dataset definition for the destination output data. You can have as many sink transformations as your data flow requires.

**9. Question 9**

True or False

When data is stored in Data Lake Storage Gen2, the file size, number of files, and folder structure can have an impact on performance.

**1 / 1 point**

☐ False

☒ True

**Correct**

The file size, number of files, and folder structure in Data Lake Storage Gen2 can have an impact on performance. If you store your data has many small files, this can negatively affect performance. In general, organize your data into larger-sized files for better performance.

**10. Question 10**

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**

☐ 1GB to 10GB

☒ 256MB to 100GB

☐ 10GB to 100GB

☐ 256MB to 1GB

**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).



## Test prep

Latest Submission Grade 87.5%

### 1. Question 1

Examine the following statement and replace the missing words with your choice from the supplied options.

A fact table contains dimension key columns that relate to dimension tables. Dimension key columns determine the \_\_\_\_\_ of a fact table, while the dimension key values determine the \_\_\_\_\_ of a fact table.

1 / 1 point

☐ granularity, dimensionality

☒ dimensionality, granularity

Correct

A fact table contains dimension key columns that relate to dimension tables, and numeric measure columns. The dimension key columns determine the dimensionality of a fact table, while the dimension key values determine the granularity of a fact table.

### 2. Question 2

What type of system is Azure Synapse Analytics?

1 / 1 point

☐ Symmetric Multiprocessing (SMP) system

☐ Online Transactional Processing (OLTP) Database

☒ Massively Parallel Processing (MPP) system

Correct

Azure Synapse Analytics is a massively parallel processing (MPP) system.

**3. Question 3**

True or False

A Snowflake Schema is an extension of a Star Schema, and it adds additional dimensions. The dimension tables are normalized which splits data into additional tables.

**1 / 1 point**

☐ False

☒ True

**Correct**

A Snowflake Schema is an extension of a Star Schema, and it adds additional dimensions. The dimension tables are normalized which splits data into additional tables.

**4. Question 4**

To create a snowflake schema in Azure Synapse Studio you navigate to which of the following hubs?

**1 / 1 point**

☐ Integrate

☐ Develop

☐ Manage

☒ Data

**Correct**

To create a snowflake schema in Azure Synapse Studio you navigate to the Data hub.

**5. Question 5**

What distribution option would you use for a product dimension table that will contain 1,000 records in Synapse Analytics?

**0 / 1 point**

☐ DISTRIBUTION = REPLICATE.

☒ DISTRIBUTION = ROUND\_ROBIN.

☐ DISTRIBUTION = HASH([ProductId]).

**Incorrect**

Try going back and reviewing the Design and implement data stores in a modern data warehouse lesson.

**6. Question 6**

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

**1 / 1 point**

- ☐ 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 uses surrogate keys while a snowflake schema uses business keys.
- ☐ 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.

**7. Question 7**

Which of the following are true in respect of fact tables?

Select all options that apply.

**0.75 / 1 point**

☒ A fact table contains dimension key columns that relate to dimension tables.

**Correct**

A fact table contains dimension key columns that relate to dimension tables.

☒ A fact table describes business entities.

**This should not be selected**

Try going back and reviewing the Design and implement data stores in a modern data warehouse lesson.

☒ A fact table contains numeric measure columns.

☒ Fact tables store observations or events.

**Correct**

A fact table stores observations or events.

**8. Question 8**

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

Select all options that apply.

**1 / 1 point**

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

☒ Star schema dimension tables are a denormalized data structure.

**Correct**

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

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

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

**Correct**

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

☒ Star schemas have a high level of Data redundancy.

**Correct**

Star schemas have a high level of Data redundancy.

### 9. Question 9

What distribution option would be best for a sales fact table that will contain billions of records?

**1 / 1 point**

☐ DISTRIBUTION = REPLICATE

☐ DISTRIBUTION = HEAP

☒ DISTRIBUTION = HASH([SalesOrderNumber])

**Correct**

Hash distribution provides good read performance for a large table by distributing records across compute nodes based on the hash key.

### 10. Question 10

A Star schema is a modeling 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.

## Test prep

**Latest Submission Grade 87.5%**

### 1. Question 1

Which Workload Management capability manages minimum and maximum resource allocations during peak periods?

**1 / 1 point**



Workload Isolation



Workload Containment



Workload Importance

**Correct**

Workload Isolation assigns maximum and minimum usage values for varying resources under load. These adjustments can be done live without having to take the SQL Pool offline.

### 2. Question 2

Azure Synapse Analytics dedicated SQL pools can support a maximum of how many compute nodes?

**1 / 1 point**



40



80



60



20

**Correct**

Azure Synapse Analytics dedicated SQL pools supports a maximum of 60 MPP compute nodes within the highest performance configuration.

### 3. Question 3

Select from the options below to complete the missing text in the following statement.

A data warehouse that is built on a Massively Parallel Processing (MPP) system are built for processing and analyzing large datasets. As such they perform well with \_\_\_\_\_ that can be distributed across compute nodes and storage.

Select from the options to complete the missing text.

1 / 1 point



Fewer and larger batch sizes



Multiple small batch sizes

Correct

A data warehouse that is built on a Massively Parallel Processing (MPP) system are built for processing and analyzing large datasets. As such they perform well with larger batch type loads and updates that can be distributed across the compute nodes and storage.

### 4. Question 4

In Synapse Analytics resource classes are pre-determined resource limits in Synapse SQL pools that govern compute resources and concurrency for query execution. Dynamic Resource Classes allocate a variable amount of memory depending on the service level.

Which of the following are valid Dynamic Resource Classes?

1 / 1 point



mediumrc

Correct

Mediumrc is a valid dynamic resource class.



xxlargerc



xlargerc

Correct

Xlargerc is a valid dynamic resource class.



smallrc

Correct

Smallrc is a valid dynamic resource class.



largerc

Correct

Largerc is a valid dynamic resource class.

☐ tinyrc

### 5. Question 5

Resource classes are pre-determined resource limits in Synapse SQL pool that govern compute resources and concurrency for query execution. Resource classes can help you configure resources for your queries by setting limits on the number of queries that run concurrently and, on the compute-resources assigned to each query.

Which of the following statements are correct?

Select all options that apply.

1 / 1 point

☒ Smaller resource classes reduce the maximum memory per query but increase concurrency.

Correct

Smaller resource classes reduce the maximum memory per query but increase concurrency.

☐ Larger resource classes increase concurrency, but reduce the maximum memory per query,

☐ Smaller resource classes reduce the concurrency but increase maximum memory per query.

☒ Larger resource classes increase the maximum memory per query but reduce concurrency.

Correct

Larger resource classes increase the maximum memory per query but reduce concurrency.

### 6. Question 6

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.

0 / 1 point

☐ mediumrc

☒ smallrc

☐ xlargerc

☐ largerc

Incorrect

Try going back and reviewing the Use data loading best practices in Azure Synapse Analytics lesson.

<https://www.purplefrogssystems.com/blog/2021/06/azure-synapse-series-what-is-workload-management-part1/>

### 7. Question 7

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.

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

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

☒ 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.

### 8. Question 8

Dedicated SQL pool workload management in Azure Synapse consists of three high-level concepts which give you more control over how your workload utilizes system resources

Which of the following are valid concepts?

**0.75 / 1 point**

☒ Workload Classification

☒ Workload Importance

**Correct**

Workload Importance is one of three high-level concepts.

☐ Workload partitioning

☒ Workload Isolation

**Correct**

Workload Isolation is one of three high-level concepts.

You didn't select all the correct answers

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-workload-management>

### 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

Which T-SQL Statement loads data directly from Azure Storage?

1 / 1 point

- ☐ LOAD DATA
- ☒ COPY
- ☐ INSERT FROM FILE

**Correct**

The T-SQL COPY Statement reads data from Azure Blob Storage or the Azure Data Lake and inserts it into a table within the SQL Pool.

## Test prep

Latest Submission Grade 100%

**1.** Question 1

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.

2. Question 2

Which of the following table distributions available in Synapse Analytics SQL Pools creates a hash function to deterministically assign each row to a distribution and has a column designated as the distribution column?

1 / 1 point

- ☐ Round robin distribution
- ☒ Hash distribution
- ☐ Replicated tables

Correct

This distribution can deliver the highest query performance for joins and aggregations on large tables. To shard data, a hash function is used to deterministically assign each row to a distribution. In the table definition, one of the columns is designated as the distribution column.

3. Question 3

Which of the following indexing options are available in Dedicated SQL Pools?

Select all options that apply.

1 / 1 point

- ☒ Non-clustered index

Correct

Dedicated SQL Pools support Non-clustered indexes.

- ☒ Clustered columnstore index

Correct

Dedicated SQL Pools support Clustered columnstore indexes.

- ☒ Clustered index

**Correct**

Dedicated SQL Pools support Clustered indexes.

☐ Table Index

**4. Question 4**

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

☐ Non-clustered

☒ Clustered columnstore

**Correct**

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

**5. Question 5**

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 columnstore

☐ Clustered

**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.

**6. Question 6**

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**

☒ Dedicated SQL Pool

☐ Serverless SQL Pool

**Correct**

Materialized views are prewritten queries with joins and filters whose definition is saved and the results persisted to a dedicated SQL pool.

### 7. Question 7

You can use Materialized Views to improve the performance of either complex or slow queries. As the data in the underlying base tables change, the data in the materialized view will automatically update without user interaction. However, there are certain restrictions you must comply with when defining a materialized view. Which of the following are valid when defining materialized views?

1 / 1 point

☒ Clustered Columnstore index is supported by materialized views.

Correct

Only Clustered Columnstore index is supported by materialized views.

☒ The hash table distribution is supported in the definition.

Correct

The hash table distribution is supported in the definition.

☐ The Replicated tables distribution is supported in the definition.

☒ The round\_robin table distribution is supported in the definition.

Correct

The round\_robin table distribution is supported in the definition.

☐ Clustered index is supported by materialized views.

### 8. Question 8

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.

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

1 / 1 point

☐ 36 Hours

☐ 24 Hours

☐ 12 Hours

☒ 48 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.

### 9. Question 9

Which Index Type offers the highest compression?

1 / 1 point



Columnstore



Heap



Rowstore

**Correct**

This is the default index type created for a table. It works on segments of rows that get compressed and optimized by column.

**10.** Question 10

How do column statistics improve query performance?

1 / 1 point



By caching column values for queries



By keeping track of how much data exists between ranges in columns.



By keeping track of the columns which are being queried.

**Correct**

It tracks cardinality and range density to determine which data access paths return the fewest rows for speed.

## Test prep

Total points 10

**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

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
- ☐ Scala
- ☐ JSON
- ☒ Python

**Correct**

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

## 3. Question 3

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 db\_exporter 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 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.

**4. Question 4**

You have a requirement to transfer data to a dedicated SQL pool that is outside of the workspace of Synapse Analytics. To establish and transfer data to a dedicated SQL pool that is outside of the workspace which form of Authentication can be used?

**1 / 1 point**

- ☐ Azure AD and SQL Authentication

☒ SQL Authentication Only

- ☐ Azure AD 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.

**5. Question 5**

When is it unnecessary to use import statements for transferring data between a dedicated SQL and Apache Spark pool?

**1 / 1 point**

- ☐ Use token-based authentication.
- ☐ Use the PySpark connector

☒ When using the integrated notebook experience from Azure Synapse Studio.

**Correct**

Import statements are not needed since they are pre-loaded in case you use the Azure Synapse Studio integrated notebook experience.

### 6. Question 6

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

☒ PolyBase

☐ JSON

☐ ORC

☐ Parquet

**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.

### 7. Question 7

In Azure Synapse Studio Develop hub you can define Spark Job definitions. Which of the following languages can be used to define job definitions?

Select all options that apply.

1 / 1 point

☐ Transact-SQL

☒ .NET Spark

**Correct**

You can define Spark Job definitions using .Net Spark.

☒ Scala

**Correct**

You can define Spark Job definitions using Scala.

☒ PySpark

**Correct**

You can define Spark Job definitions using PySpark.

### 8. Question 8

Azure Data Studio is a cross-platform tool to connect and query on-premise and cloud data platforms on windows, macOS, and Linux. Synapse Analytics supports using Azure Data Studio for connecting and querying Synapse SQL on which of the following configurations?

1 / 1 point



- ☐ Only dedicated SQL Pool resources
- ☐ Only Serverless SQL Pool resources

☒ Both dedicated and Serverless SQL Pool resources

**Correct**

Synapse Analytics supports connecting and querying Synapse SQL using Azure Data Studio through either Serverless or dedicated SQL Pool resources.

**9. Question 9**

Azure Synapse Analytics supports Approximate execution using Hyperlog accuracy to reduce latency when executing queries with large datasets. Approximate execution is used to speed up the execution of queries with a compromise for a small reduction in accuracy. What percentage accuracy of true cardinality on average will the result return when using Approximate execution?

**1 / 1 point**

- ☐ 6%
- ☐ 1%
- ☒ 2%
- ☐ 4%

**Correct**

It will return a result with a 2% accuracy of true cardinality on average.

**10. Question 10**

What Transact-SQL function verifies if a piece of text is valid JSON?

**1 / 1 point**

- ☐ JSON\_QUERY
- ☐ JSON\_VALUE

☒ ISJSON

**Correct**

ISJSON is a Transact-SQL function that verifies if a piece of text is valid JSON.

# Test prep

Latest Submission Grade 87.5%

## 1. Question 1

Apache Spark pools for Azure Synapse Analytics uses an Autoscale feature that automatically scales the number of nodes in a cluster instance up and down. Autoscale continuously monitors the Spark instance and collects metrics. Which of the following conditions will trigger Autoscale to scale up?

Select all options that apply.

1 / 1 point



Total pending CPU is greater than total free CPU for more than 1 minute.

Correct

This will trigger Autoscale to scale up.



Total pending memory is greater than total free memory for more than 1 minute.

Correct

This will trigger Autoscale to scale up.



Total pending CPU is less than total free CPU for more than 2 minutes.



Total pending memory is less than total free memory for more than 2 minutes.

## 2. Question 2

Azure Advisor provides you with personalized messages that provide information on best practices to optimize the setup of your Azure services. Azure Advisor recommendations are free, and the recommendations are based on telemetry data that is generated by Azure Synapse Analytics. Azure Advisor recommendations are checked every how many hours?

1 / 1 point



4 Hours



1 Hour



12 Hours

Correct

Advisor recommendations are checked every 1 Hour.

### 3. Question 3

Which Dynamic Management View enables you to view the active connections against a dedicated SQL pool?

1 / 1 point

- ☐ DBCC PDW\_SHOWEXECUTIONPLAN
- ☒ sys.dm\_pdw\_exec\_requests
- ☐ sys.dm\_pdw\_dms\_workers

Correct

sys.dm\_pdw\_exec\_requests enables you to view the active connections against a dedicated SQL pool.

### 4. Question 4

Examine the following statement and select from the listed options to complete the sentence.

A columnstore index scans a table by scanning column segments of individual rowgroups. Maximizing the number of rows in each rowgroup enhances query performance. For best query performance, the goal is to maximize the number of rows per rowgroup in a columnstore index.

Columnstore indexes achieve good performance when rowgroups have at least \_\_\_\_\_ rows.

1 / 1 point

- ☐ 1,000,000 rows
- ☒ 100,000 rows
- ☐ 10,000 rows
- ☐ 1,048,576 rows

Correct

100,000 rows achieves good performance.

### 5. Question 5

SQL pool in Azure Synapse supports standard and materialized views. Which of the following are features of Materialized views?

Select all options that apply.

1 / 1 point

- ☒ View content is pre-processed and stored in SQL pool during view creation. The view is updated as data is added to the underlying tables.

Correct

View content is pre-processed and stored in SQL pool during view creation. The view is updated as data is added to the underlying tables.

☐ View content is generated each time the view is used,

☒ Speed to retrieve view data from complex queries is Fast,

**Correct**

Speed to retrieve view data from complex queries is Fast,

☐ Speed to retrieve view data from complex queries is Slow,

## 6. Question 6

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

**1 / 1 point**

☒ Using DBCC PDW\_SHOWSPACEUSED to see the number of table rows that are stored in each of the 60 distributions.

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

☐ 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.

## 7. Question 7

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

**1 / 1 point**

☐ Increased resiliency benefits

☐ Increased high availability

☒ Reducing the execution time for complex queries with JOINS and aggregate functions.

**Correct**

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

## 8. Question 8

What features are provided when using a managed workspace virtual network?

Select all options that apply.

**0.5 / 1 point**

☒ You will need to create a subnet for your Spark clusters based on peak load.

**This should not be selected**

You don't need to create a subnet for your Spark clusters based on peak load.

☒ You don't have to configure inbound NSG rules on your own Virtual Networks to allow Azure Synapse management traffic to enter your Virtual Network.

**Correct**

When using a managed workspace virtual network, you don't have to configure inbound NSG rules on your own Virtual Networks.

☐ Management of the virtual network is offloaded to Azure Synapse.

☒ Your workspace is network isolated from other workspaces.

**Correct**

When using a managed workspace virtual network your workspace is network isolated from other workspaces.

**9. Question 9**

Azure Synapse Analytics enables you to connect to its various components through endpoints. You can set up managed private endpoints to access these components in a secure manner known as private links. Which of the following statements are true in respect of Private Endpoints?

Select all options that apply.

**1 / 1 point**

☒ When you use a private link, traffic between your Virtual Network and workspace traverses entirely over the Microsoft backbone network.

**Correct**

Traffic between your Virtual Network and workspace traverses entirely over the Microsoft backbone network.

☐ When you use a private link, traffic between your Virtual Network and workspace traverses over the public Internet network.

☒ You can manage the private endpoints in the Azure Synapse Studio manage hub.

**Correct**

You can manage the private endpoints in the Azure Synapse Studio manage hub.

☒ You must have an Azure Synapse workspace with a Managed workspace Virtual Network.

**Correct**

This can only be achieved in an Azure Synapse workspace with a Managed workspace Virtual Network.

### 10. Question 10

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. Which of the following are common signals used as a basis to determine if conditional access should first be applied?

Select all options that apply.

1 / 1 point

☒ 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.

☐ Multi Factor Authentication

☒ IP address information

**Correct**

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

☒ User or group membership names

**Correct**

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

☒ 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.

### 11. Question 11

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

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

☒ A managed virtual network has not been created.

☐ 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.

**12.** Question 12

Transparent data encryption (TDE) is an encryption mechanism to help you protect Azure Synapse Analytics. The way TDE will do so, is by encrypting data at rest.

If you export a TDE-protected database, the exported content will be in a specific state. Which of the following represents the state of the exported content?

**0 / 1 point**



Will be in an unencrypted state



Will be encrypted using a new TDE Key



Remains in its original encrypted state

**Incorrect**

Try going back and reviewing the Secure a data warehouse in Azure Synapse Analytics lesson.

# 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



**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.