

Knowledge check

Total points 5

1. Question 1

Which one of the following Azure solutions provides an integrated analytics platform that combines data warehousing, big data analytics, data integration, and visualization into a single environment?

1 / 1 point



Azure Synapse Analytics



Azure Data Lake Storage Gen2



Azure Databricks



Azure Data Factory

Correct

Azure Synapse is a limitless analytics service that brings together enterprise data warehousing and Big Data analytics.

2. Question 2

You have unplanned or ad-hoc workloads to process within Azure Synapse Analytics. Which of the following offers is best suited to assist with this task?

1 / 1 point



Azure Synapse serverless SQL pools



Azure Synapse Pipelines



Azure Synapse dedicated SQL pools

Correct

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

3. Question 3

You need to process large volumes of data for data exploration and analysis. Which resource model in Azure Synapse Analytics is best suited to help you complete this task?

1 / 1 point



Dedicated SQL Pools



Serverless SQL pools.



Pipelines

Correct

The Dedicated SQL pool is the best choice for predictable performance and cost such as when storing data in SQL tables.

4. Question 4

You need to train AI models using AzureML. Which one of the following components of Azure Synapse Analytics can assist you?

1 / 1 point



Azure Synapse Pipelines



Azure Synapse Spark



Azure Synapse Studio

Correct

Use a notebook to ingest and shape data, and then use SparkML and AzureML to train models with it.

Create and administer your Apache Spark pools in an Azure Synapse Analytics workspace.

To integrate an Apache Spark pool with an Azure Machine Learning workspace, link to the Azure Synapse Analytics workspace.

5. Question 5

You need to use your data to answer questions about what will happen in the future. Which one of the following forms of analytics should you deploy?

1 / 1 point



Descriptive analytics



Diagnostic analytics



Predictive analytics

Correct

Predictive analytics helps answer questions about what will happen in the future.

Knowledge check

Total points 5

1. Question 1

An Azure Synapse Analytics workspace stores data in which of the following?

1 / 1 point



Cosmos DB Tables



Azure Data Lake Storage Gen2



Apache Spark Tables



Azure SQL Database

Correct

The workspace stores data in Apache Spark tables

2. Question 2

You need to create and enable SQL or Spark pools. Can you perform these tasks within the Azure Synapse Analytics workspace?

1 / 1 point

☐ No

☒ Yes

Correct

Azure Synapse Analytics enables you to create pools, either SQL pools, or Spark pools within the workspace that can be seamlessly mixed and matched based on your requirements.

It is able to do this through Azure Synapse Analytics shared metadata, which enables the different engines to share databases and tables.

3. Question 3

Do you need to create a Data Lake Gen2 storage account before creating an Azure Synapse Analytics workspace?

1 / 1 point

☒ No

☐ Yes

Correct

When creating a new Synapse Workspace, you can either create a new Data Lake Gen2 storage account or select from an existing account

4. Question 4

In a data warehouse data is retrieved, cleansed, and transformed from a range of source data system, and is then served in a structured relational format. What is this format commonly referred to as?

1 / 1 point

☐ Fact table

☐ Snowflake Schema

☒ Star Schema

☐ Dimension Table

Correct

A star schema is a database organizational structure optimized for use in a data warehouse or business intelligence that uses a single large fact table to store transactional or measured data, and one or more smaller dimensional tables that store attributes about the data.

5. Question 5

You create a dedicated SQL pool to ensure predictable performance and cost by reserving processing power for data permanently stored in SQL tables. Which of the following services are you making use of through this process?

1 / 1 point

☐ Data virtualization

☒ Data warehousing

Correct

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.

Knowledge check

Total points 6

1. Question 1

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

Correct

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

2. Question 2

You need to create a pipeline in Azure Synapse Studio. Which hub should you navigate to?

0 / 1 point

- ☒ Manage hub
- ☐ Data hub
- ☐ Develop hub
- ☐ Integrate hub

Incorrect

Azure Synapse Analytics Studio Manage Hub allows you to manage your workspace within a single dashboard.

It has many options, from creating linked services using connectors such as Power BI and Azure Data Lake to creating SQL pools and Apache Spark pools.

3. Question 3

You need to access the Power BI Workspace from within Azure Synapse Studio. Which hub should you navigate to?

1 / 1 point

- ☐ Data Hub
- ☒ Develop Hub
- ☐ Integrate Hub
- ☐ Manage Hub

Correct

The Develop Hub in Azure Synapse Analytics enables you to write code and define business logic using a combination of notebooks, SQL scripts, and data flows. It also provide access to the Power BI reports node

4. Question 4

You perform analytics over a database system that is seen to provide transactional capabilities without impacting the performance of the system. What is this process referred to as?

1 / 1 point

- ☒ Hybrid Transactional and Analytical processing (HTAP)
- ☐ Data Virtualization
- ☐ 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.

5. Question 5

You need to run near real-time analytics over operational data in Azure Cosmos DB without impacting the performance of transactional workloads on Azure Cosmos DB. Can you perform this task using Azure Synapse Link for Cosmos DB?

1 / 1 point

- ☐ No
- ☒ Yes

Correct

Azure Synapse Link for Azure Cosmos DB is a cloud-native hybrid transactional and analytical processing (HTAP) capability.

It enables you to run near real-time analytics over operational data in Azure Cosmos DB without impacting the performance of transactional workloads on Azure Cosmos DB.

6. Question 6

Which Azure Service is Azure Synapse Pipelines based on?

1 / 1 point

- ☐ Azure Stream Analytics
- ☐ Azure Data Explorer
- ☒ Azure Data Factory

Correct

Azure Synapse Pipelines is based in the Azure Data Factory service

Knowledge check

Total points 6

1. Question 1

Which hub in Azure Synapse Studio is used to view the status of the various integration runtimes that are running, or Apache Spark jobs and data flow debug activities?

1 / 1 point

- ☒ Monitor Hub
- ☐ Develop Hub
- ☐ Integrate Hub
- ☐ Manage Hub

Correct

Use the Monitor hub to view pipeline and trigger runs, view the status of the various integration runtimes that are running, view Apache Spark jobs, SQL requests, and data flow debug activities. If you want to see the status of a job or activity, this is where you want to go.

2. Question 2

You are executing a notebook. Which of the following should you select as its compute target?

1 / 1 point

- ☐ Dedicated Pool
- ☒ Spark Pool
- ☐ SQL Pool
- ☐ Serverless Pool

Correct

When you execute a notebook, you select a Spark pool as its compute target.

3. Question 3

Which of the following would you use to author data flows in Azure Synapse Studio Develop hub?

1 / 1 point



Code-free GUI



PySpark



Parquet



Scala

Correct

Data flows are powerful data transformation workflows that use the power of Apache Spark but are authored using a code-free GUI.

Data flows are visually designed data transformations in Azure Synapse Analytics.

Data flows allow data engineers to develop data transformation logic without writing code.

4. Question 4

You are tasked with creating a Notebook. Which Azure Synapse Studio hub should you access?

1 / 1 point



Manage hub



Integrate hub



Develop hub



Data hub

Correct

The Develop hub is where you manage SQL scripts, Synapse notebooks, data flows, and Power BI reports.

5. Question 5

You want to view the contents of the primary data lake store. Where can you access this content within Azure Synapse Studio?

0 / 1 point



In the workspace tab of the Data hub.



In the Integration section of the Monitor hub.



In the linked tab of the Data hub.

Incorrect

The workspace tab of the data hub is where you can view the contents of the SQL and Spark pools.

6. Question 6

You need to ingest data code-free within Azure Synapse Analytics. Which of the following tools can you use to perform this?

1 / 1 point



Azure Data Factory



Azure Databricks



Power BI

Correct

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

Knowledge check

Total points 5

1. Question 1

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

1 / 1 point



Azure Data Factory



Power BI



Azure Databricks

Correct

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

2. Question 2

Which of the following features does Power BI support? Select all options that apply.

1 / 1 point



Notebooks



Natural Language Query

Correct

Power BI is a powerful tool to build and deploy dashboards in the enterprise, through rich visualizations, and features like natural language querying.



Support for multiple data sources

Correct

Power BI supports an enormous set of data sources, which can be queried live, or be used to model and ingest for detailed analysis and visualization.

☒ Rich visualizations

Correct

Power BI enables customers to build visualizations on massive amounts of data and ensures that data insights are available to everyone across their organization.

3. Question 3

Which of the following data formats are natively supported by Azure Synapse Analytics?

1 / 1 point

☒ Parquet

Correct

Parquet format is natively supported by Synapse Analytics.

☐ Scala

☒ CSV

Correct

CSV Format is natively supported by Synapse Analytics.

☒ ORC

Correct

ORC format is natively supported by Synapse Analytics.

☒ JSON

Correct

JSON format is natively supported by Synapse Analytics.

4. Question 4

Messages can be collected and processed from streaming services using which of the following features?

0.75 / 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 Databricks

☒ Azure Functions

Correct

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

☐ Azure IoT Central

You didn't select all the correct answers

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 pool

☐ 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 Mapping data flow



Synapse Copy activity



Synapse Studio

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



Conditional Split



GetMetadata activity



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



Sink



Source

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

Can the file size, number of files, and folder structure of data stored in Data Lake Storage Gen2 impact on performance?

1 / 1 point



Yes



No

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

What is the recommended file size for Data Lake Storage Gen 2?

1 / 1 point



256MB to 1GB



1GB to 10GB



256MB to 100GB



10GB to 100GB

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

Enabling Debug mode while building data flows in Azure Synapse 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.

Test prep

Latest Submission Grade 80%

1. Question 1

Which of the following analytics can you use to answer questions about what has happened based on historical data?

1 / 1 point



Descriptive analytics



Diagnostic analytics



Predictive analytics

Correct

Descriptive analytics helps answer questions about what has happened based on historical data. Descriptive analytics techniques summarize large datasets.

2. Question 2

What type of analytics can you use to answer the question of “What is likely to happen in the future based on previous trends and patterns?”

1 / 1 point



Descriptive analytics



Predictive analytics



Prescriptive analytics



Diagnostic analytics

Correct

Predictive analytics can answer the question of “What is likely to happen in the future based on previous trends and patterns?”. Analyzing past data patterns and trends by looking at historical data and customer insights can predict what might happen going forward and, in doing so, inform many aspects of a business.

3. Question 3

Which of the following Synapse Analytics tools can you use to develop SQL Scripts and Notebooks?

1 / 1 point



Data Lake



Azure Synapse Studio



Azure portal

Correct

You can use the Azure Synapse Studio to develop TSQL scripts and notebooks.

4. Question 4

Which of the following languages are supported in Apache Spark within Azure Synapse Analytics? Select the three languages that apply.

1 / 1 point



JSON



PySpark

Correct

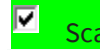
Synapse notebooks supports four Apache Spark languages. PySpark (Python), Spark (Scala), SparkSQL, .NET for Apache Spark (C#).



C#

Correct

Synapse notebooks supports four Apache Spark languages. PySpark (Python), Spark (Scala), SparkSQL, .NET for Apache Spark (C#).



Scala

Correct

Synapse notebooks supports four Apache Spark languages. PySpark (Python), Spark (Scala), SparkSQL, .NET for Apache Spark (C#).

5. Question 5

Which of the following offerings in Azure Synapse Analytics ensures predictable performance and cost?

0 / 1 point



Azure Synapse dedicated SQL pools



Azure Synapse Pipelines



Azure Synapse serverless SQL pools

Incorrect

Try going back and reviewing the *What is Azure Synapse Analytics?* lesson.

Test prep

Latest Submission Grade 100%

1. Question 1

Which three of the following Azure Data Lake file types can be explored and analyzed directly with Synapse Analytics SQL and Spark?

1 / 1 point



JSON

Correct

SQL and Spark can directly explore and analyze JSON files stored in the data lake.

☒ CSV

Correct

SQL and Spark can directly explore and analyze CSV files stored in the data lake.

☐ XML

☒ Parquet

Correct

SQL and Spark can directly explore and analyze Parquet files stored in the data lake.

2. Question 2

You are creating a new Azure Data Lake Storage Generation 2 workspace. In which instance do you need to ensure that the name of the workspace is unique?

1 / 1 point

☐ Within the selected Subscription only.

☒ Globally within Microsoft Azure.

☐ Within the selected Resource Group only.

Correct

When creating an Azure Data Lake Storage Gen2 name, the storage account name must be unique across all existing storage account names in Azure.

3. Question 3

A data warehouse provides which one of the following features?

1 / 1 point

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

☒ A central repository of data stored in relational tables.

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

Correct

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

4. Question 4

The ability to interact with data without the need to understand its type, format, or structure is referred to as...

1 / 1 point

☐ Data Ingestion

☒ Data Virtualization

☐ Data Warehousing

☐ Data Visualization

Correct

Data virtualization allows you to interact with data without the need to understand how without the need to understand its type, format, structure, or the technical specifications of the source data.

This is very helpful when performing diagnostic analytics where the need to access data in a timely manner to answer a question is more important.

5. Question 5

What method can you use to populate permanent tables with data in a data warehouse?

1 / 1 point



Extract, Transform and Load (ETL)



Extract, Load and Transform (ELT)

Correct

Data in a data warehouse is stored in permanent tables that are populated using an extract, transform, and load (ETL) process by services such as Azure Synapse pipelines, or Azure Data Factory.

As a result, you need to understand the data that is stored in the sources systems and how it should arrive within the data warehouse. This in turn dictates how you should cleanse or transform the data.

6. Question 6

Which of these actions does a driver node perform when working with Spark pool clusters?

1 / 1 point



The Driver node pulls data from a specified data source and sends the work to the Worker nodes.



The Driver node sends work to the Worker nodes and instructs them to pull data from a specified data source.

Correct

In a Spark Pool cluster, the Driver node sends work to the Worker nodes and instructs them to pull data from a specified data source.

7. Question 7

Which three of the following features in Azure Synapse Analytics can you use to create a new Spark pool?

1 / 1 point



Azure Portal

Correct

You can create a new Spark pool in Azure Synapse using the Azure Portal.



Synapse Analytics .Net SDK

Correct

You can create a new Spark pool in Azure Synapse using the Synapse Analytics .Net SDK.



JSON



Azure PowerShell

Correct

You can create a new Spark pool in Azure Synapse using Azure PowerShell.

8. Question 8

Which one of the following components of Azure Synapse Analytics can the different engines use to share the databases and tables between Spark pools and SQL on-demand engine?

1 / 1 point



Azure Synapse shared metadata



Azure Synapse Link



Azure Synapse Spark pools

Correct

The shared metadata allows the different engines to share the databases and tables between Spark pools and SQL on-demand engine.

9. Question 9

Which one of the following Azure Synapse Analytics components can you use to perform Hybrid Transaction and Analytical processing?

1 / 1 point

- ☐ Azure Synapse Studio
- ☒ Azure Synapse Link
- ☐ Azure Synapse Pipeline

Correct

Azure Synapse Link is the component that enables Hybrid Transactional and Analytical Processing.

10. Question 10

Where in Azure Synapse Analytics can you develop TSQL scripts and notebooks?

1 / 1 point

- ☐ Azure Data Lake
- ☐ Azure portal
- ☒ Azure Synapse Studio

Correct

You can develop TSQL scripts and notebooks within Azure Synapse Studio.

Test prep

Latest Submission Grade 100%

1. Question 1

You need to grant access to the Azure Synapse workspace and its resources. Which hub should you use?

1 / 1 point

- ☐ Monitor hub
- ☐ Integrate hub

 **Manage hub**

Correct

You can grant access to Synapse workspace in the Manage hub.

2. Question 2

You are executing a Synapse notebook. Which one of the following pools must be selected as a compute target?

1 / 1 point

 **Spark pool**

- ☐ Dedicated pool
- ☐ SQL pool
- ☐ Serverless pool


Correct

When you execute a notebook, you select a Spark pool as its compute target.

3. Question 3

You are building a modern data warehouse. Which one of the following tools can you use to ingest data code-free?

1 / 1 point

- ☐ Power BI
-  **Azure Data Factory**
- ☐ Azure Databricks


Correct

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

4. Question 4

You need to view the contents of the primary data lake. Which hub within Azure Synapse Studio can you use?

1 / 1 point

- ☐ The Integration section of the Monitor hub.
- ☐ The workspace tab of the Data hub.
-  **The linked tab of the Data hub.**

Correct

The linked tab of the data hub is where you can view the contents of the primary data lake store.

5. Question 5

You need to author Data Flows within Azure Synapse Studio. Which one of the following can you use?

1 / 1 point

- ☐ Parquet files
- ☐ Scala
- ☐ PySpark



Code-free GUI

Correct

Data flows are powerful data transformation workflows that use the power of Apache Spark but are authored using a code-free GUI.

Data flows are visually designed data transformations in Azure Synapse Analytics.

Data flows allow data engineers to develop data transformation logic without writing code.

Test prep

Latest Submission Grade 100%

1. Question 1

You have recently deployed Azure Synapse Analytics and need 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 or ELT, including preparation and transformation.

2. Question 2

Identify three reasons as to why you might need to add a staging area into the architecture of a modern data warehouse.

1 / 1 point



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.



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.

☒ 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.

3. Question 3

You are implementing a modern data warehouse which requires a staging area. Which one of the following technologies would be most suited to creating this staging area?

1 / 1 point

☐ Azure Synapse Spark Pools.

☐ Azure Synapse SQL Pools

☒ Azure Data Lake

Correct

Feedback: 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.

4. Question 4

You are creating a new Mapping Data Flow transformation and need to route data rows to different streams based on matching conditions. Which one of the following data flow activities will give you the ability to route data to multiple streams?

1 / 1 point

☒ Conditional Split

☐ Lookup

☐ GetMetadata activity

Correct

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

5. Question 5

You are working with Data Lake Storage Generation 2 and need to ensure that its performance is not negatively affected by your data’s file size. What is the recommended file size that you should store your data in?

1 / 1 point

☐ 256MB to 1GB

☐ 10GB to 100GB

☒ 256MB 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).

6. Question 6

You enable Debug mode while building Data Flows in Azure Synapse Analytics. Which one of the following features does this automatically create?

1 / 1 point

- ☐ A dedicated SQL Pool.
- ☐ A serverless cluster.
- ☒ A small Spark cluster.

Correct

When building data flows in Azure Synapse you can enable debug mode which turns on a small interactive Spark cluster.