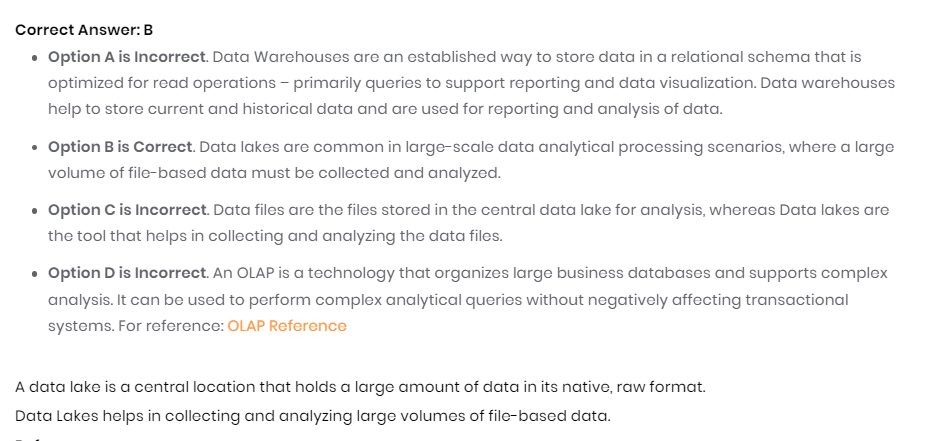
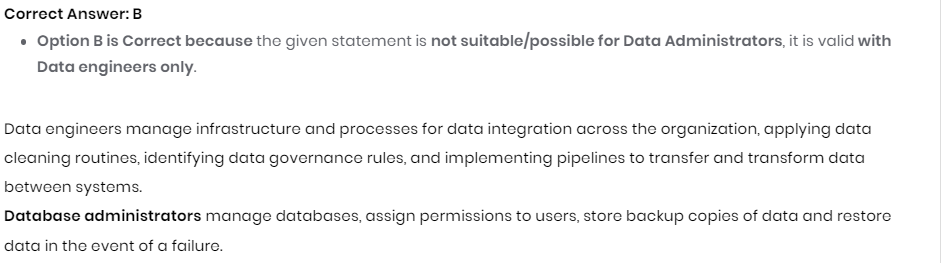
1)



<https://learn.microsoft.com/en-us/training/modules/explore-core-data-concepts/6-analytical-processing>

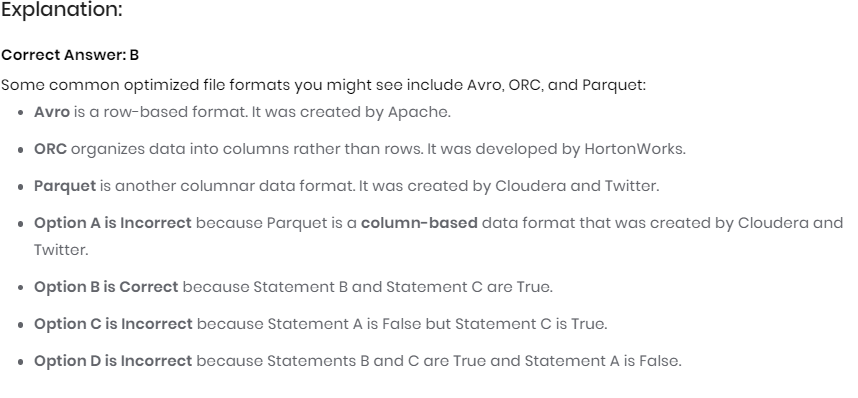
<https://azure.microsoft.com/en-us/solutions/data-lake/>

2)



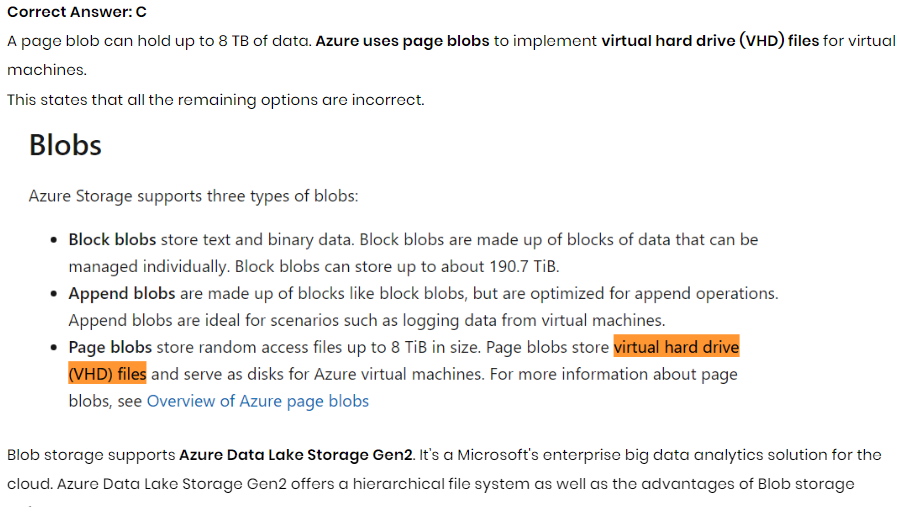
<https://learn.microsoft.com/en-us/training/modules/explore-roles-responsibilities-world-of-data/2-explore-job-roles>

3)



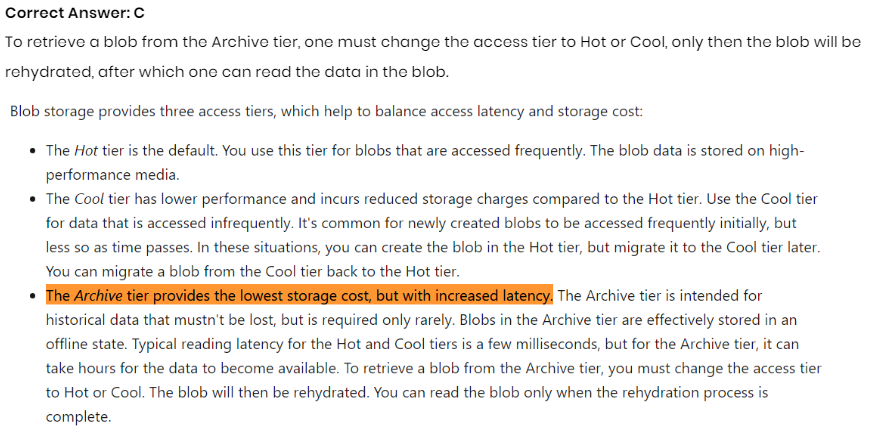
<https://learn.microsoft.com/en-us/training/modules/explore-core-data-concepts/3-file-storage>

4)

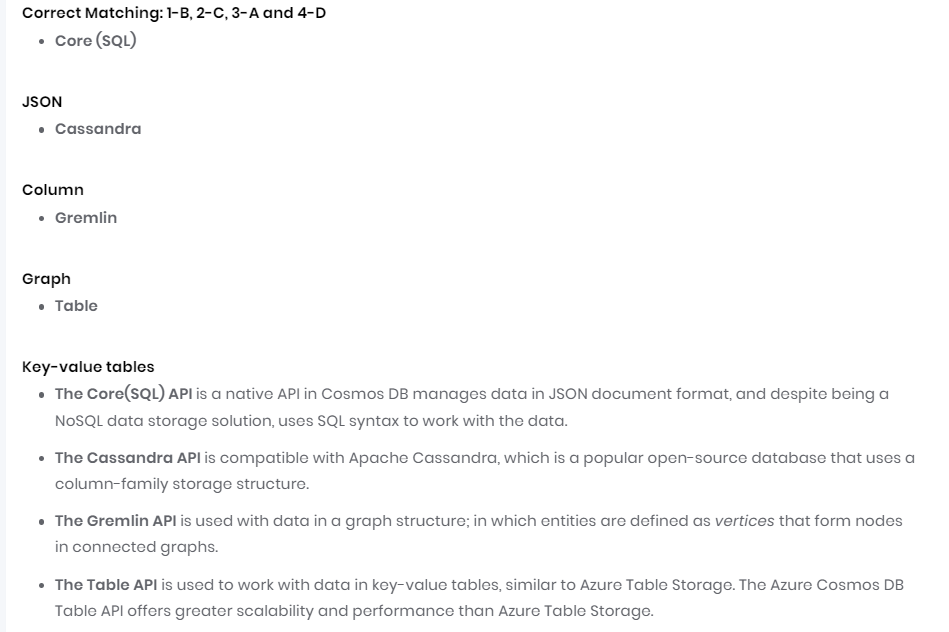


<https://learn.microsoft.com/en-us/training/modules/explore-provision-deploy-non-relational-data-services-azure/2-azure-blob-storage?ns-enrollment-type=learningpath&ns-enrollment-id=learn.wwl.azure-data-fundamentals-explore-non-relational-data>

5)

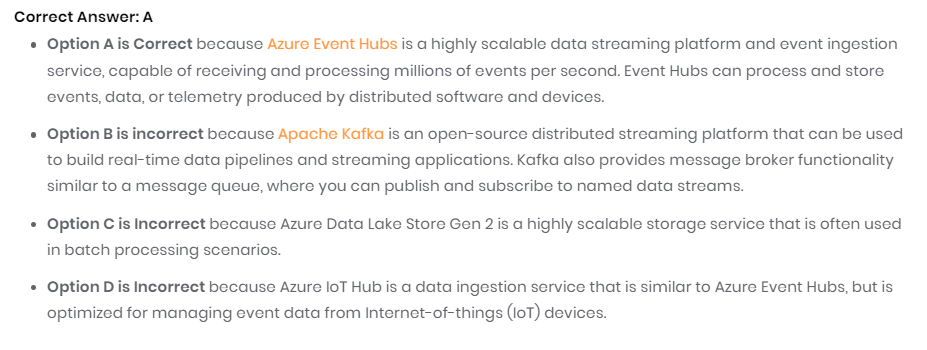


6)



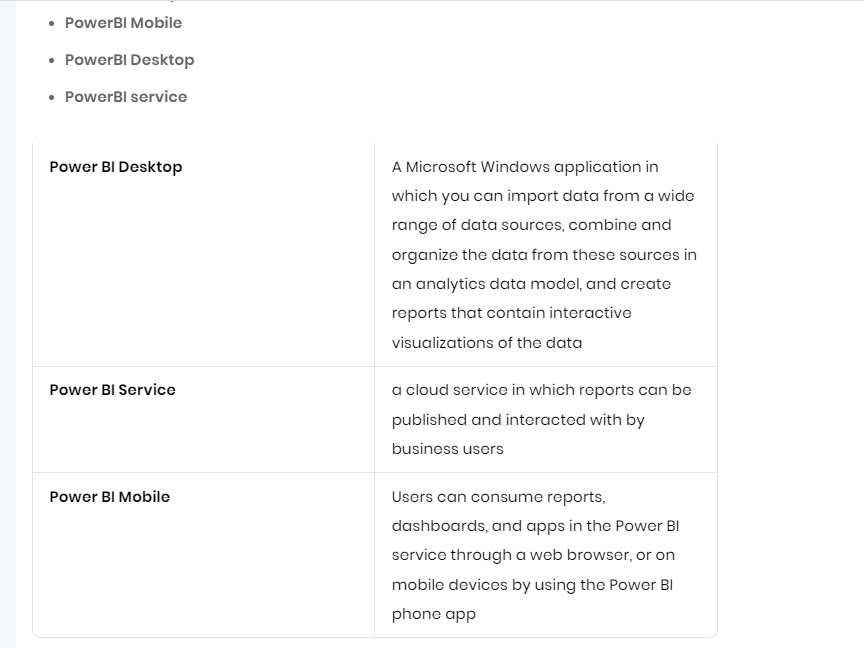
<https://learn.microsoft.com/en-us/training/modules/explore-non-relational-data-stores-azure/3-cosmos-db-apis>

7)



<https://learn.microsoft.com/en-us/training/modules/explore-fundamentals-stream-processing/3-explore-common-elements>

<https://learn.microsoft.com/en-us/azure/architecture/data-guide/technology-choices/real-time-ingestion>

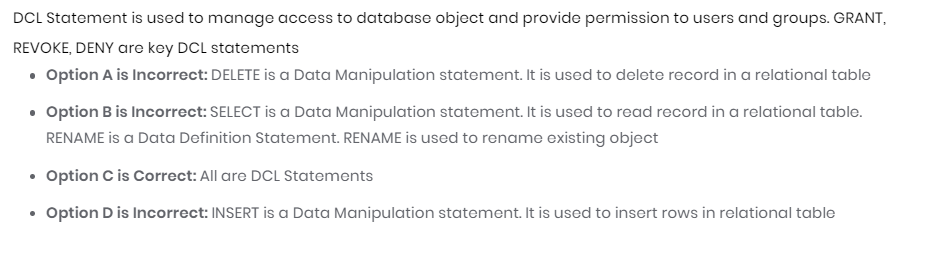
8) 

<https://learn.microsoft.com/en-us/training/modules/explore-fundamentals-data-visualization/2-power-bi>

9) 

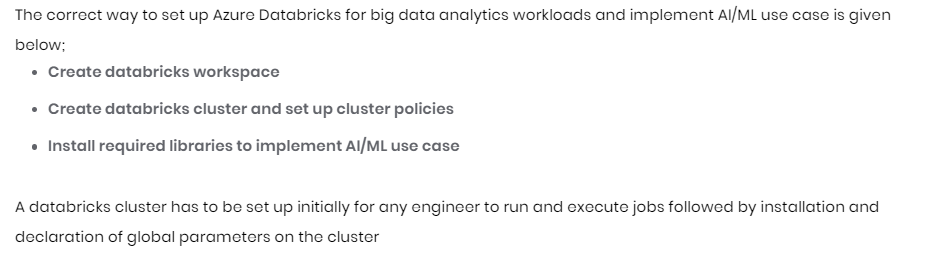


10)



<https://learn.microsoft.com/en-us/training/modules/explore-relational-data-offerings/4-query-with-sql?ns-enrollment-type=learningpath&ns-enrollment-id=learn.wwl.azure-data-fundamentals-explore-relational-data>

11)

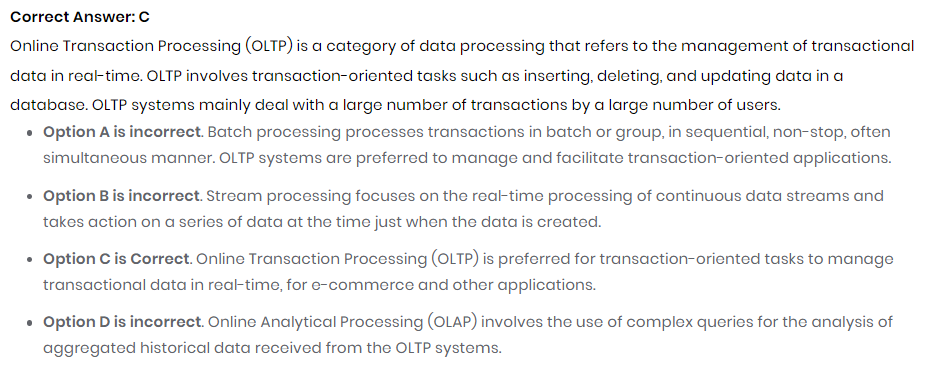


<https://learn.microsoft.com/en-us/azure/databricks/clusters/configure>

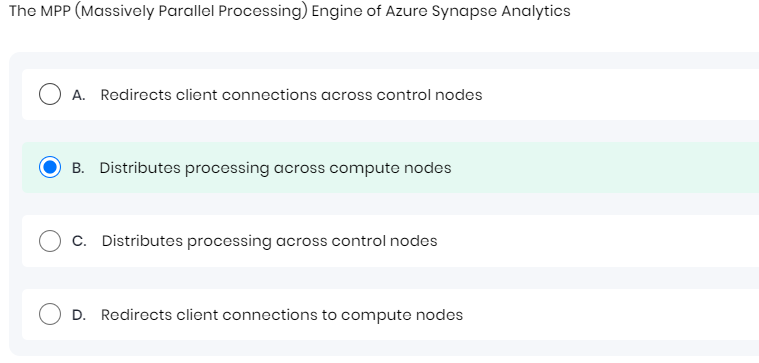
<https://learn.microsoft.com/en-us/azure/databricks/workspace/workspace-assets>

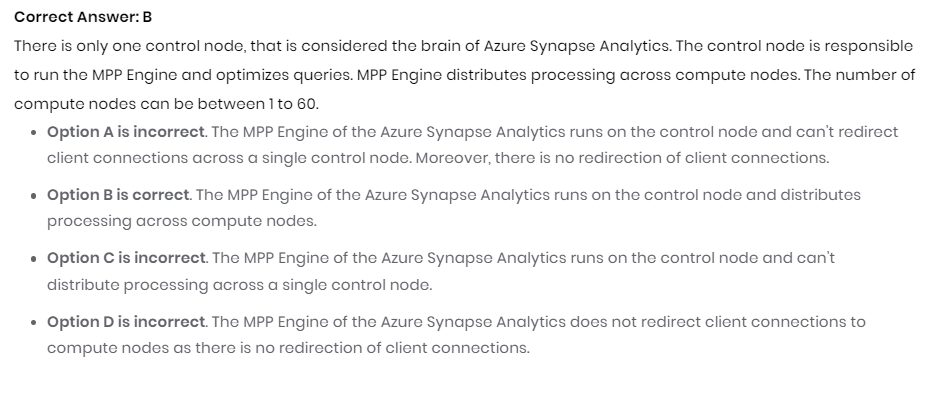
<https://learn.microsoft.com/en-us/azure/databricks/getting-started/?tabs=azure-portal>

12)

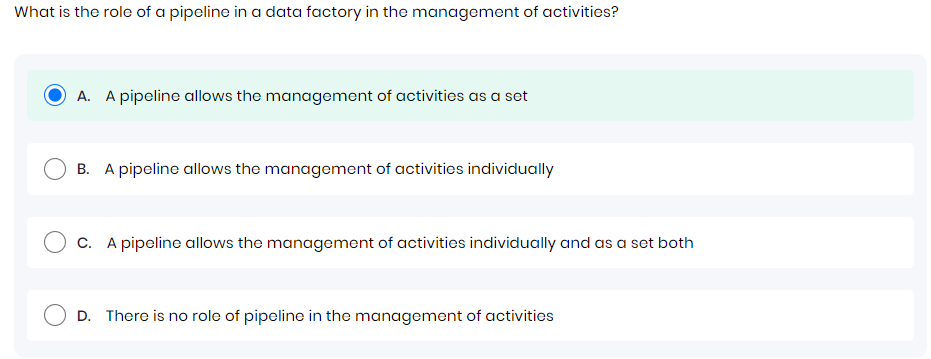


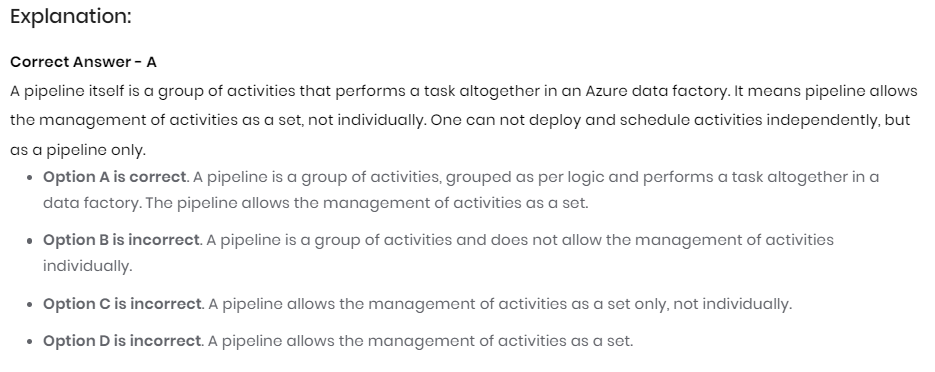
13)





14)

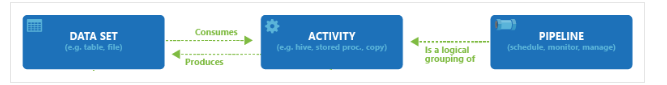




A Data Factory or Synapse Workspace can have one or more pipelines. A pipeline is a logical grouping of activities that together perform a task. For example, a pipeline could contain a set of activities that ingest and clean log data, and then kick off a mapping data flow to analyze the log data. The pipeline allows you to manage the activities as a set instead of each one individually. You deploy and schedule the pipeline instead of the activities independently.

The activities in a pipeline define actions to perform on your data. For example, you may use a copy activity to copy data from SQL Server to an Azure Blob Storage. Then, use a data flow activity or a Databricks Notebook activity to process and transform data from the blob storage to an Azure Synapse Analytics pool on top of which business intelligence reporting solutions are built.

Azure Data Factory and Azure Synapse Analytics have three groupings of activities: [data movement activities](https://learn.microsoft.com/en-us/azure/data-factory/copy-activity-overview), [data transformation activities](https://learn.microsoft.com/en-us/azure/data-factory/transform-data), and [control activities](https://learn.microsoft.com/en-us/azure/data-factory/concepts-pipelines-activities?tabs=data-factory#control-flow-activities). An activity can take zero or more input [datasets](https://learn.microsoft.com/en-us/azure/data-factory/concepts-datasets-linked-services) and produce one or more output [datasets](https://learn.microsoft.com/en-us/azure/data-factory/concepts-datasets-linked-services). The following diagram shows the relationship between pipeline, activity, and dataset:

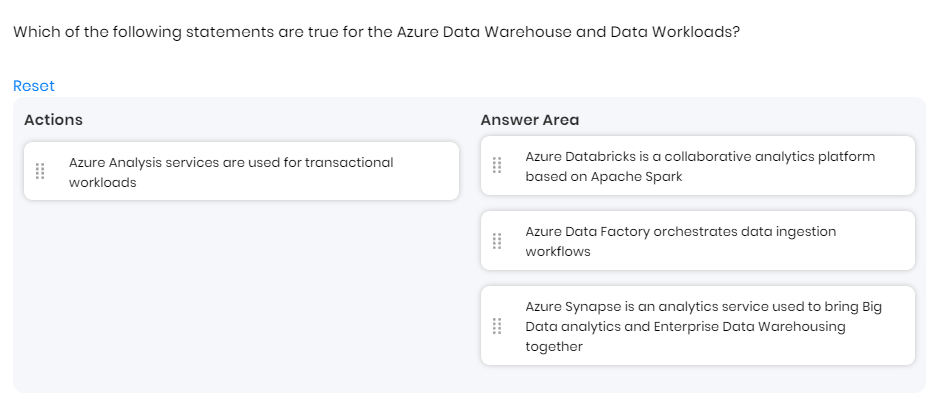


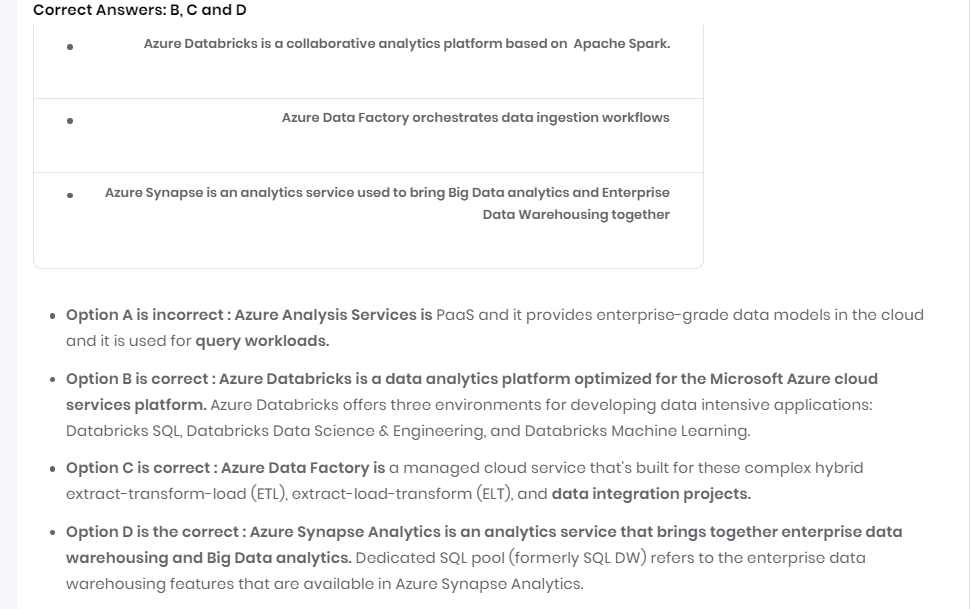
An input dataset represents the input for an activity in the pipeline, and an output dataset represents the output for the activity. Datasets identify data within different data stores, such as tables, files, folders, and documents. After you create a dataset, you can use it with activities in a pipeline. For example, a dataset can be an input/output dataset of a Copy Activity or an HDInsightHive Activity.

**Data movement activities**

Copy Activity in Data Factory copies data from a source data store to a sink data store. Data Factory supports the data stores listed in the table in this section. Data from any source can be written to any sink.

14)





15)

**What is Azure Analysis Services?**

Azure Analysis Services is a fully managed platform as a service (PaaS) that provides enterprise-grade data models in the cloud. Use advanced mashup and modeling features to combine data from multiple data sources, define metrics, and secure your data in a single, trusted tabular semantic data model. The data model provides an easier and faster way for users to perform ad hoc data analysis using tools like Power BI and Excel.

**The right tier when you need it**

Azure Analysis Services is available in **Developer**, **Basic**, and **Standard** tiers. Within each tier, plan costs vary according to processing power, Query Processing Units (QPUs), and memory size. When you create a server, you select a plan within a tier. You can change plans up or down within the same tier, or upgrade to a higher tier, but you can't downgrade from a higher tier to a lower tier.

**Developer tier**

This tier is recommended for evaluation, development, and test scenarios. A single plan includes the same functionality of the standard tier, but is limited in processing power, QPUs, and memory size. **Query replica scale-out is not available for this tier**. **This tier does not offer an SLA.**

| **Plan** | **QPUs** | **Memory (GB)** |
| --- | --- | --- |
| D1 | 20 | 3 |

**Basic tier**

This tier is recommended for production solutions with smaller tabular models, limited user concurrency, and simple data refresh requirements. **Query replica scale-out is not available for this tier**. Perspectives, multiple partitions, and **DirectQuery tabular model features are not supported in this tier.**

| **Plan** | **QPUs** | **Memory (GB)** |
| --- | --- | --- |
| B1 | 40 | 10 |
| B2 | 80 | 16 |

**Standard tier**

This tier is for mission-critical production applications that require elastic user-concurrency, and have rapidly growing data models. It supports advanced data refresh for **near real-time data model updates, and supports all tabular modeling features.**

Total cost depends on a number of factors. For example, your chosen region, tier, query replicas, and pause/resume. Use the [Azure Analysis Services Pricing](https://azure.microsoft.com/pricing/details/analysis-services/) calculator to determine typical pricing for your region. This tool calculates pricing for a single-server instance for a single region. Keep in mind, query replicas are billed at the same rate as the server.

16) **What is Azure Databricks?**

Azure Databricks is a unified, open analytics platform for building, deploying, sharing, and maintaining enterprise-grade data, analytics, and AI solutions at scale. The Azure Databricks Lakehouse Platform integrates with cloud storage and security in your cloud account, and manages and deploys cloud infrastructure on your behalf.

**What is Azure Databricks used for?**

Our customers use Azure Databricks to process, store, clean, share, analyze, model, and monetize their datasets with solutions from BI to machine learning. Use the Azure Databricks platform to build and deploy data engineering workflows, machine learning models, analytics dashboards, and more.

The Azure Databricks platform architecture comprises two primary parts:

* The infrastructure used by Azure Databricks to deploy, configure, and manage the platform and services.
* The customer-owned infrastructure managed in collaboration by Azure Databricks and your company.

**What are common use cases for Azure Databricks?**

Use cases on Azure Databricks are as varied as the data processed on the platform and the many personas of employees that work with data as a core part of their job. The following use cases highlight how users throughout your organization can leverage Azure Databricks to accomplish tasks essential to processing, storing, and analyzing the data that drives critical business functions and decisions.

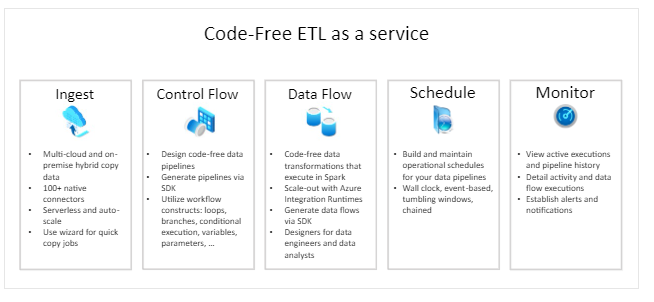
**17) What is Azure Data Factory?**

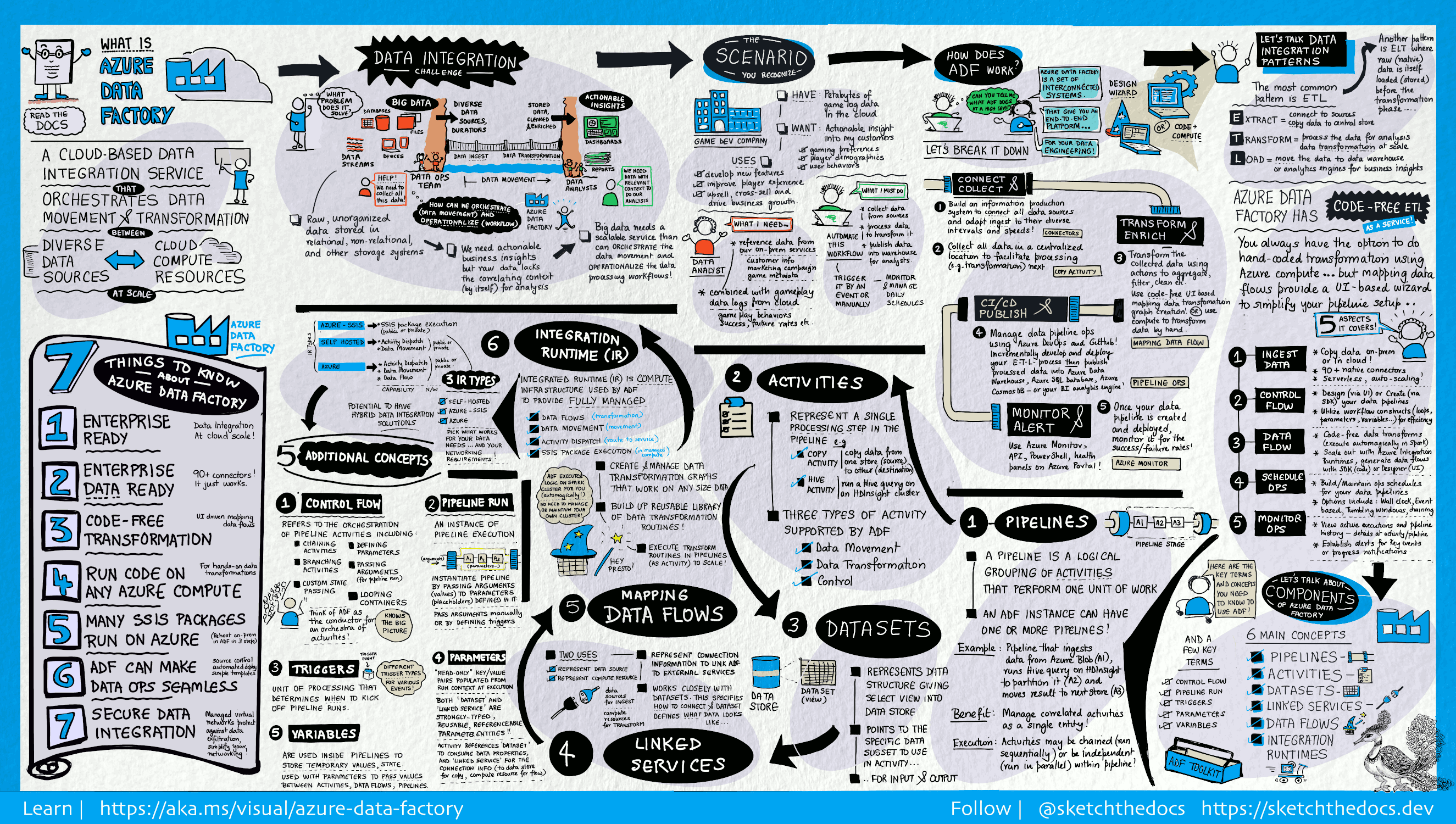
In the world of big data, raw, unorganized data is often stored in relational, non-relational, and other storage systems. However, on its own, raw data doesn't have the proper context or meaning to provide meaningful insights to analysts, data scientists, or business decision makers.

Big data requires a service that can orchestrate and operationalize processes to refine these enormous stores of raw data into actionable business insights. Azure Data Factory is a managed cloud service that's built for these complex hybrid extract-transform-load (ETL), extract-load-transform (ELT), and data integration projects.

**How does it work?**

Data Factory contains a series of interconnected systems that provide a complete end-to-end platform for data engineers.

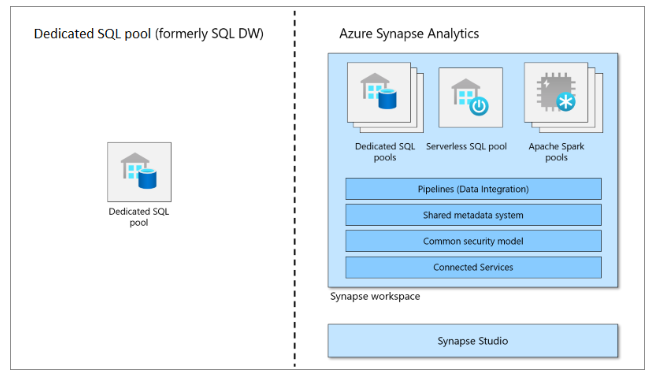




<https://learn.microsoft.com/en-us/azure/data-factory/introduction>

# 18) What is dedicated SQL pool (formerly SQL DW) in Azure Synapse Analytics?

Azure Synapse Analytics is an analytics service that brings together enterprise data warehousing and Big Data analytics. Dedicated SQL pool (formerly SQL DW) refers to **the enterprise data warehousing features** that are available in Azure Synapse Analytics.



Dedicated SQL pool (formerly SQL DW) represents a collection of analytic resources that are provisioned when using Synapse SQL. The size of a dedicated SQL pool (formerly SQL DW) is determined by Data Warehousing Units (DWU).

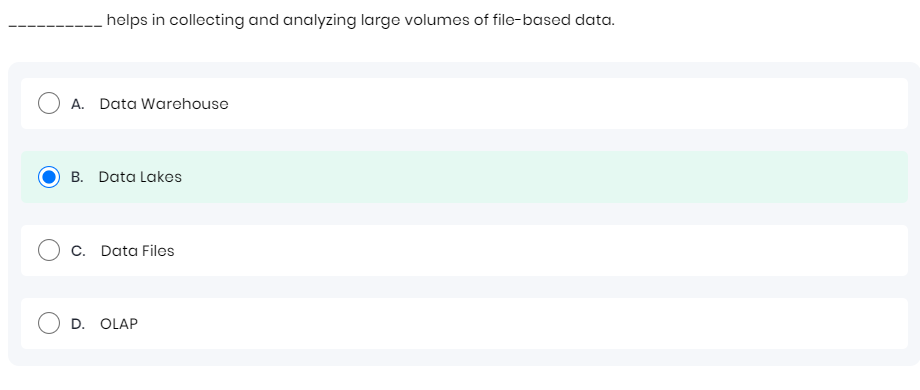
Once your dedicated SQL pool is created, you can import big data with simple [PolyBase](https://learn.microsoft.com/en-us/sql/relational-databases/polybase/polybase-guide?toc=/azure/synapse-analytics/sql-data-warehouse/toc.json&bc=/azure/synapse-analytics/sql-data-warehouse/breadcrumb/toc.json&view=azure-sqldw-latest&preserve-view=true) T-SQL queries, and then use the power of the distributed query engine to run high-performance analytics. As you integrate and analyze the data, dedicated SQL pool (formerly SQL DW) will become the single version of truth your business can count on for faster and more robust insights.

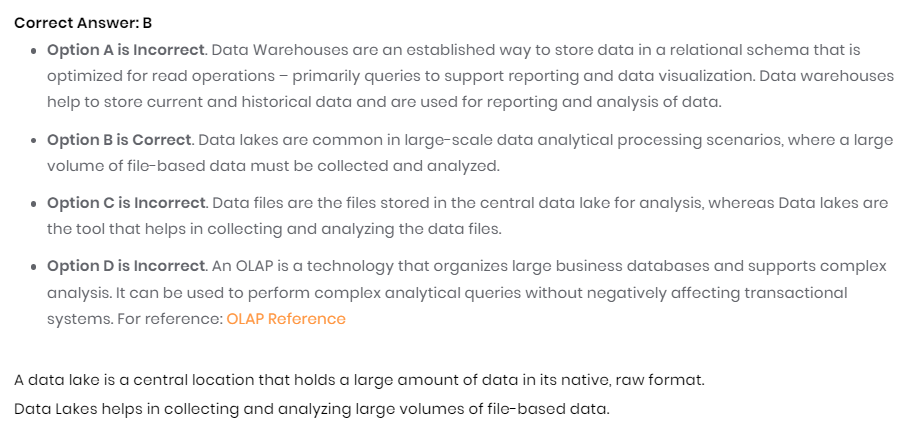
**What is PolyBase?**

PolyBase enables your SQL Server instance to query data with T-SQL directly from SQL Server, Oracle, Teradata, MongoDB, Hadoop clusters, Cosmos DB, and S3-compatible object storage without separately installing client connection software. You can also use the generic ODBC connector to connect to additional providers using third-party ODBC drivers. PolyBase allows T-SQL queries to join the data from external sources to relational tables in an instance of SQL Server.

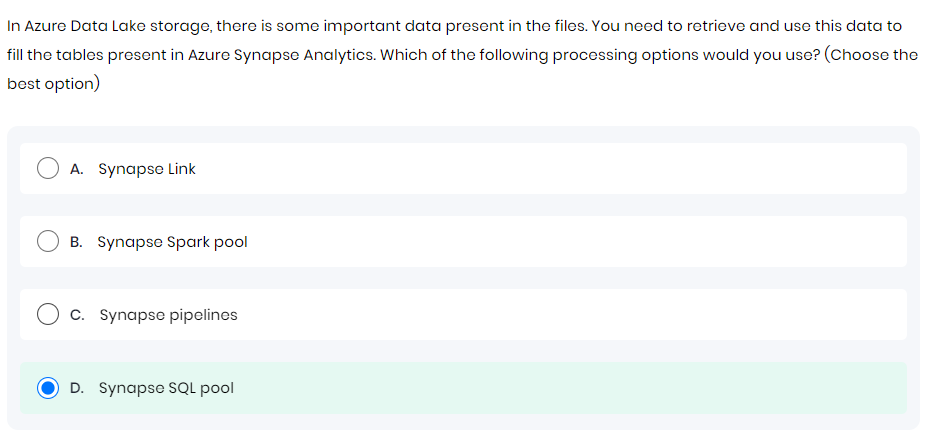
A key use case for data virtualization with the PolyBase feature is to allow the data to stay in its original location and format. You can virtualize the external data through the SQL Server instance, so that it can be queried in place like any other table in SQL Server. This process minimizes the need for ETL processes for data movement. This data virtualization scenario is possible with the use of PolyBase connectors.

20)



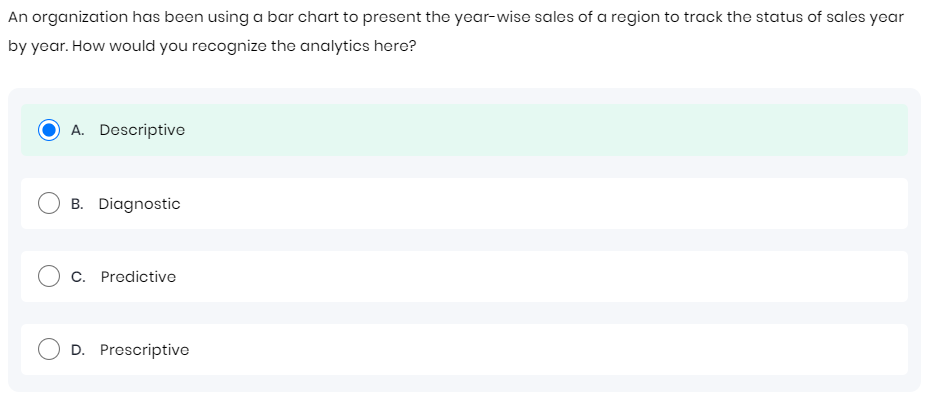


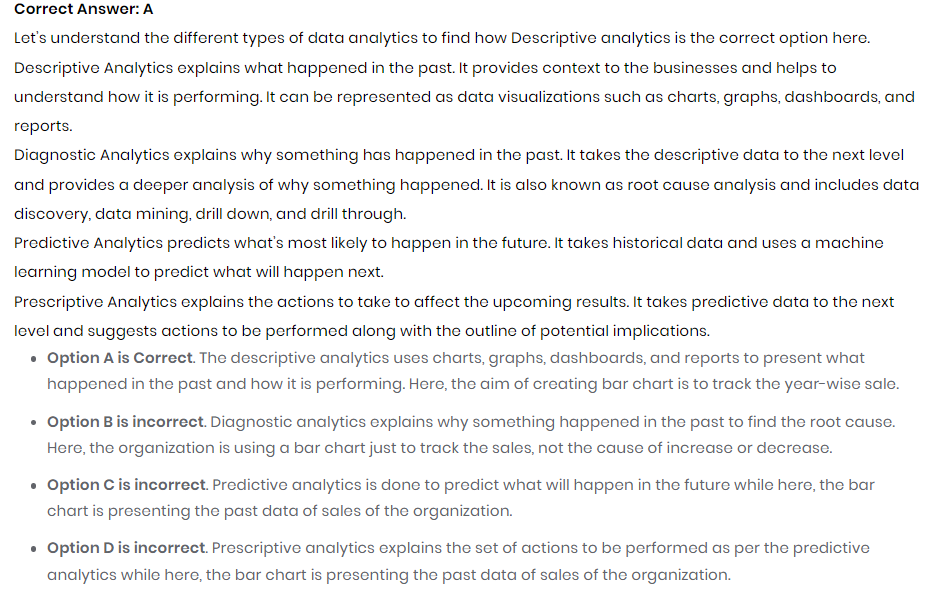
21)



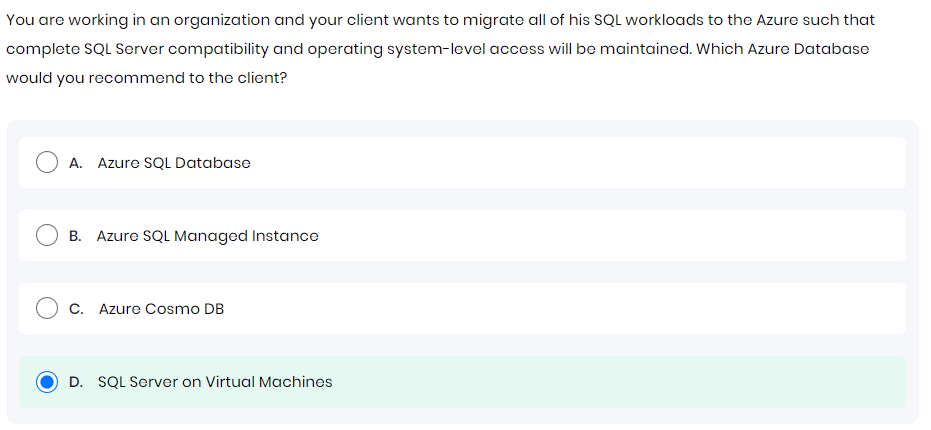


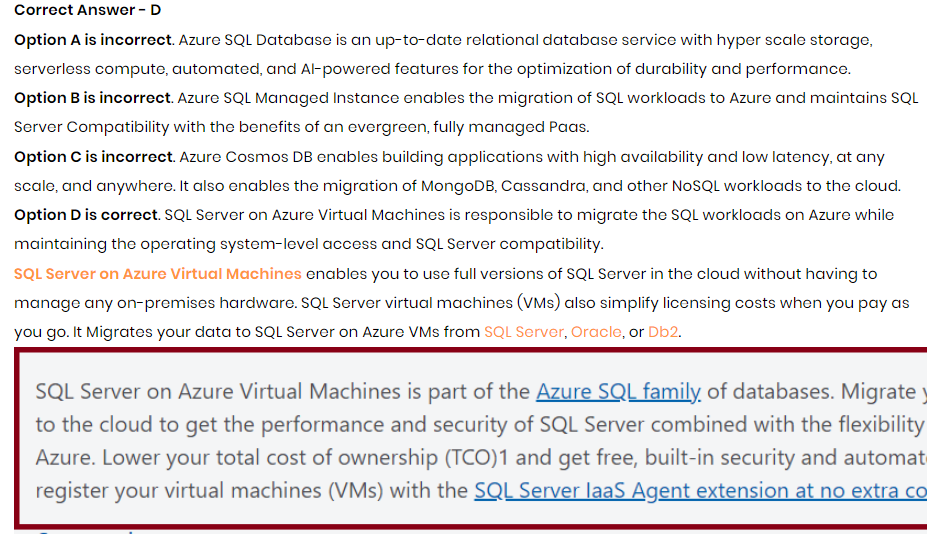
22)

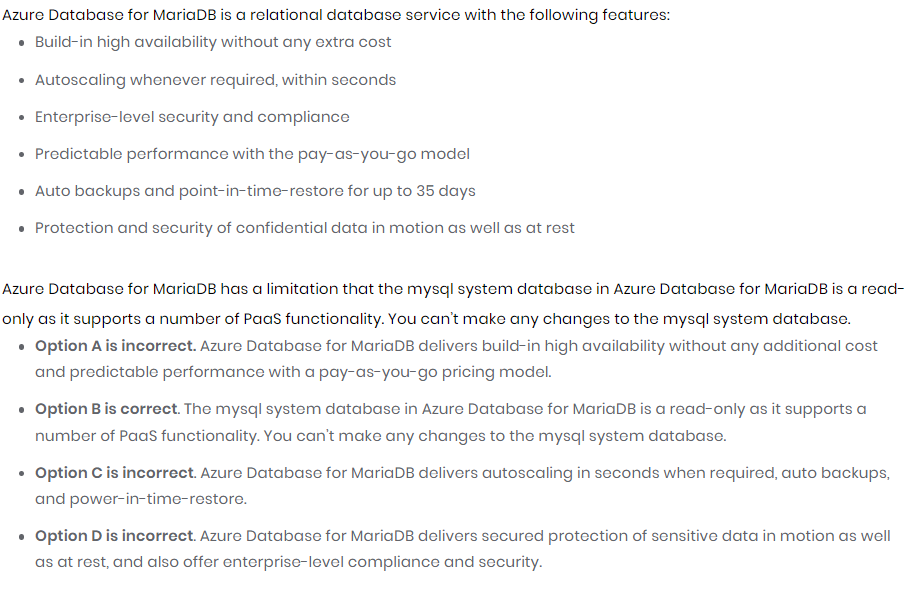




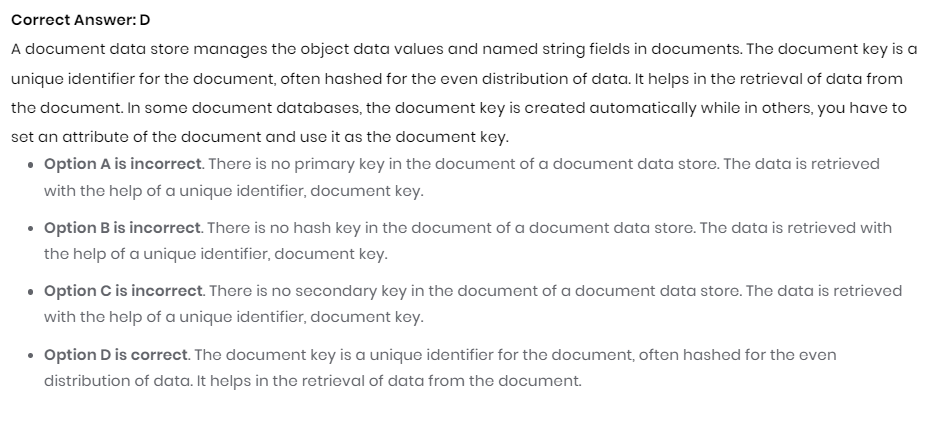
22)





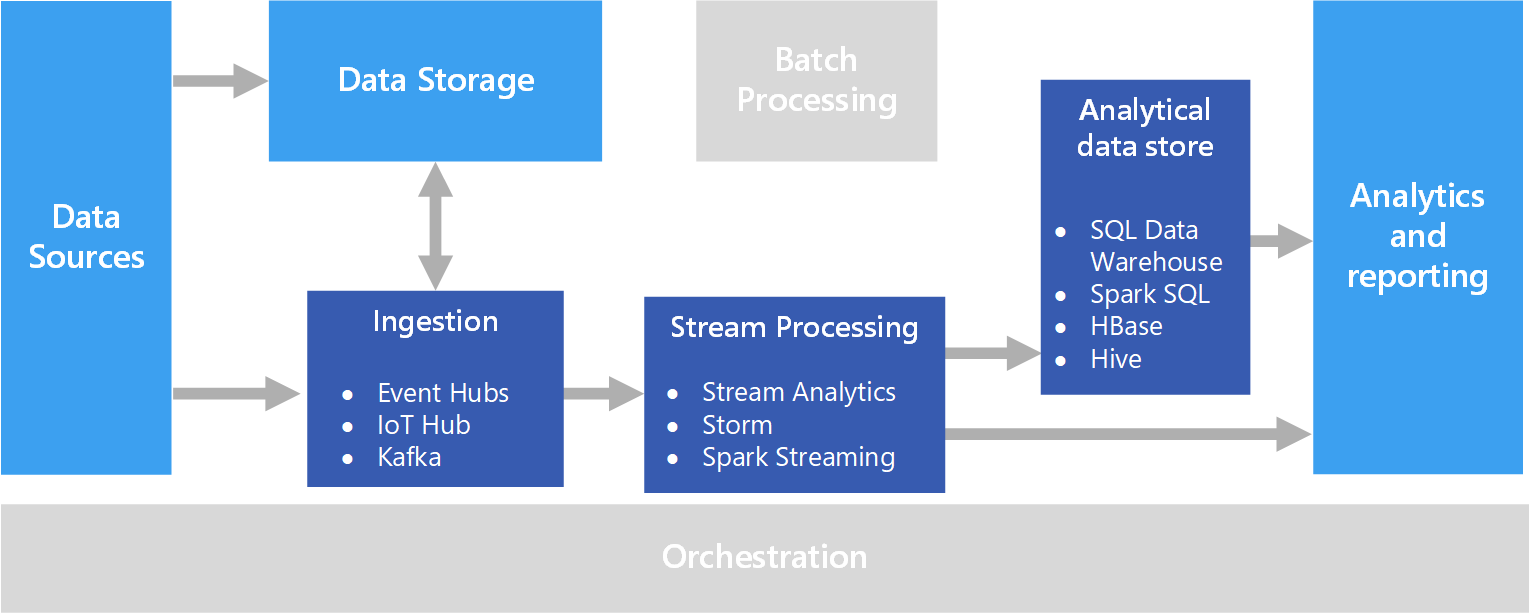
24) 

25)



# 26) 27) Real-time processing

Real time processing deals with streams of data that are captured in real-time and processed with minimal latency to generate real-time (or near-real-time) reports or automated responses. For example, a real-time traffic monitoring solution might use sensor data to detect high traffic volumes. This data could be used to dynamically update a map to show congestion, or automatically initiate high-occupancy lanes or other traffic management systems.

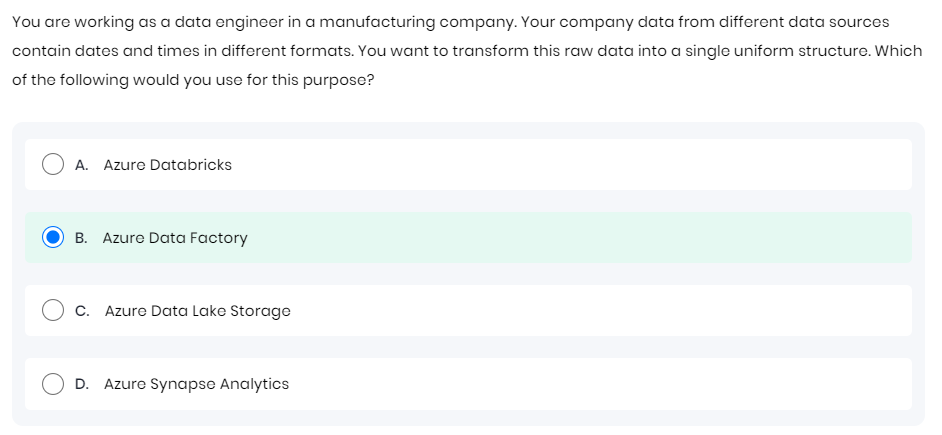


Real-time processing is defined as the processing of unbounded stream of input data, with very short latency requirements for processing — measured in milliseconds or seconds. This incoming data typically arrives in an unstructured or semi-structured format, such as JSON, and has the same processing requirements as [batch processing](https://learn.microsoft.com/en-us/azure/architecture/data-guide/big-data/batch-processing), but with shorter turnaround times to support real-time consumption. Processed data is often written to an analytical data store, which is optimized for analytics and visualization. The processed data can also be ingested directly into the analytics and reporting layer for analysis, business intelligence, and real-time dashboard visualization.

### Real-time message ingestion

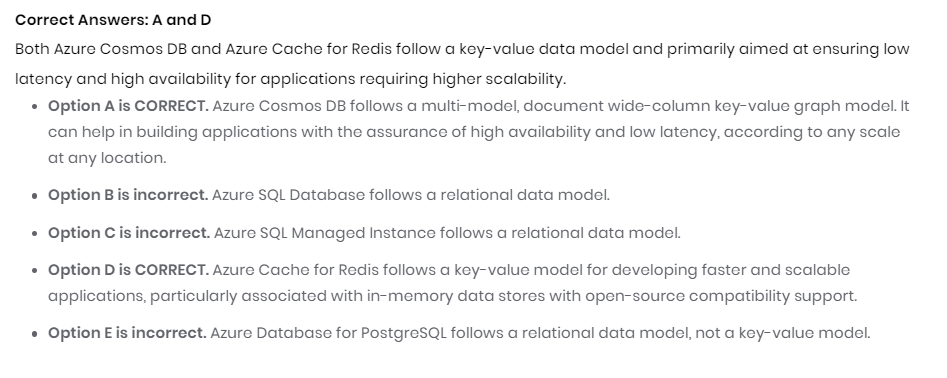
* **Azure Event Hubs**. Azure Event Hubs is a messaging solution for ingesting millions of event messages per second. The captured event data can be processed by multiple consumers in parallel. While Event Hubs natively supports AMQP (Advanced Message Queuing Protocol 1.0), it also provides a binary compatibility layer that allows applications using the Kafka protocol (Kafka 1.0 and above) to process events using Event Hubs with no application changes.
* **Azure IoT Hub**. Azure IoT Hub provides bi-directional communication between Internet-connected devices, and a scalable message queue that can handle millions of simultaneously connected devices.
* **Apache Kafka**. Kafka is an open source message queuing and stream processing application that can scale to handle millions of messages per second from multiple message producers, and route them to multiple consumers. Kafka is available in Azure as an HDInsight cluster type, with Azure Events for Kafka, and also available via ConfluentCloud through our partnership with Confluent.

28)

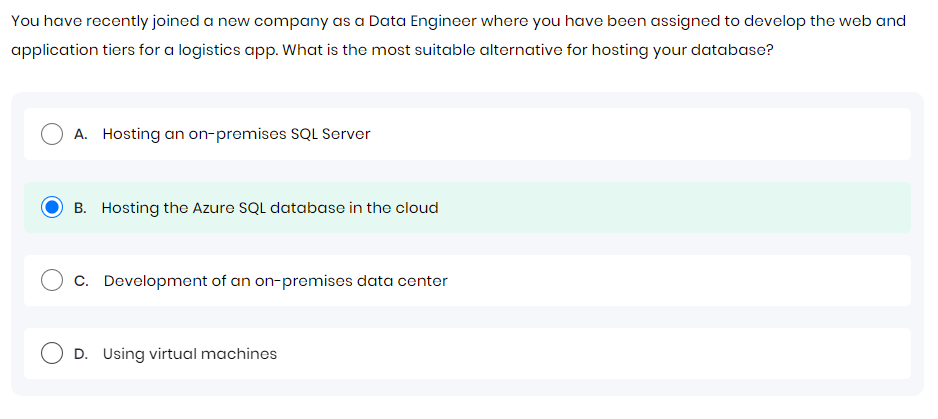


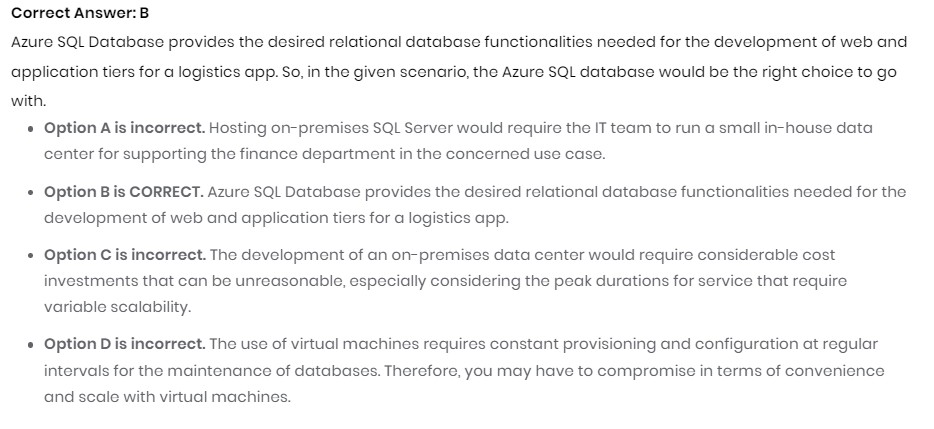


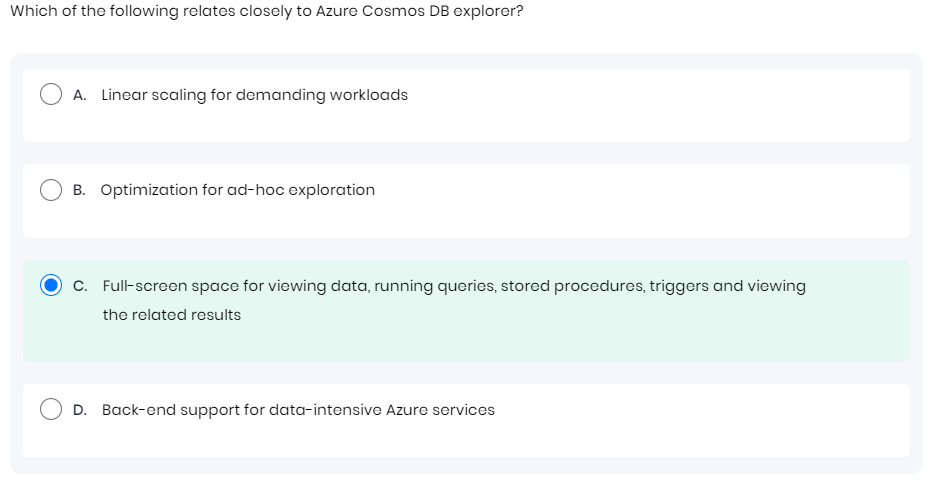
30)

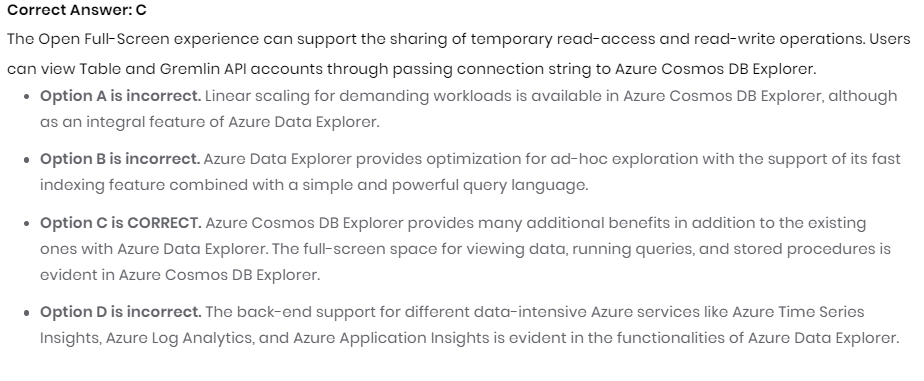


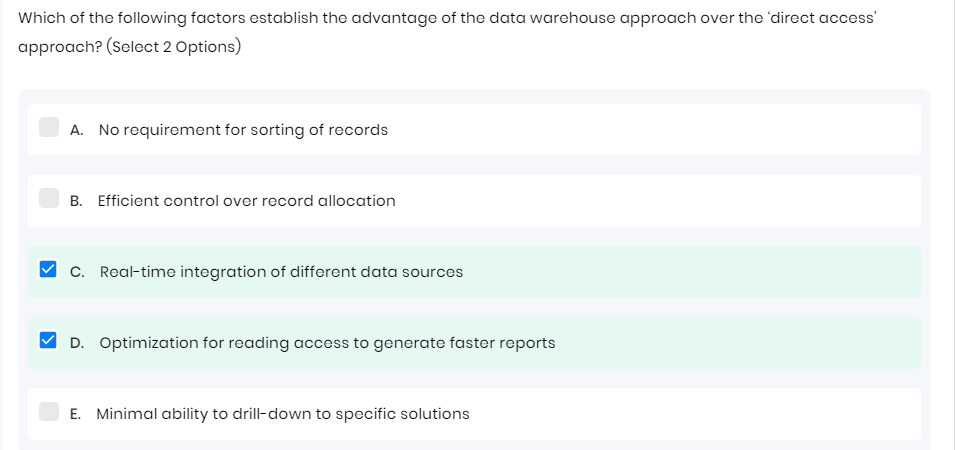
31)





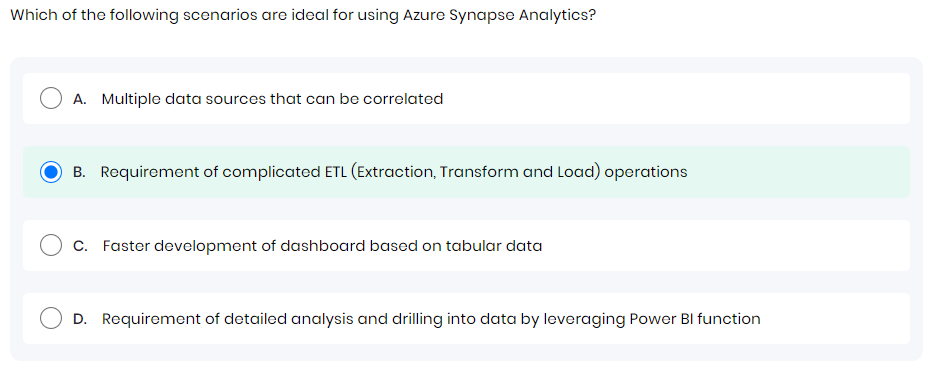
31) 

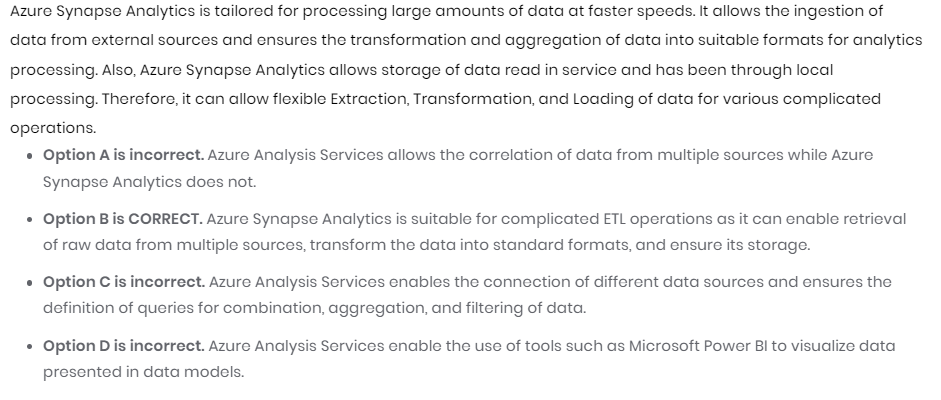
33)



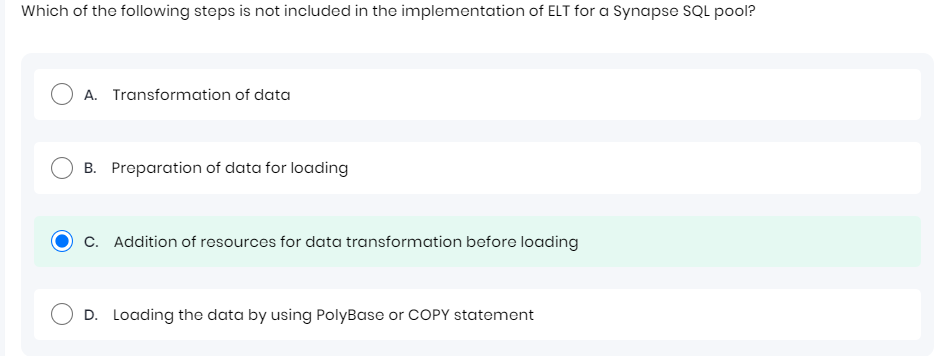


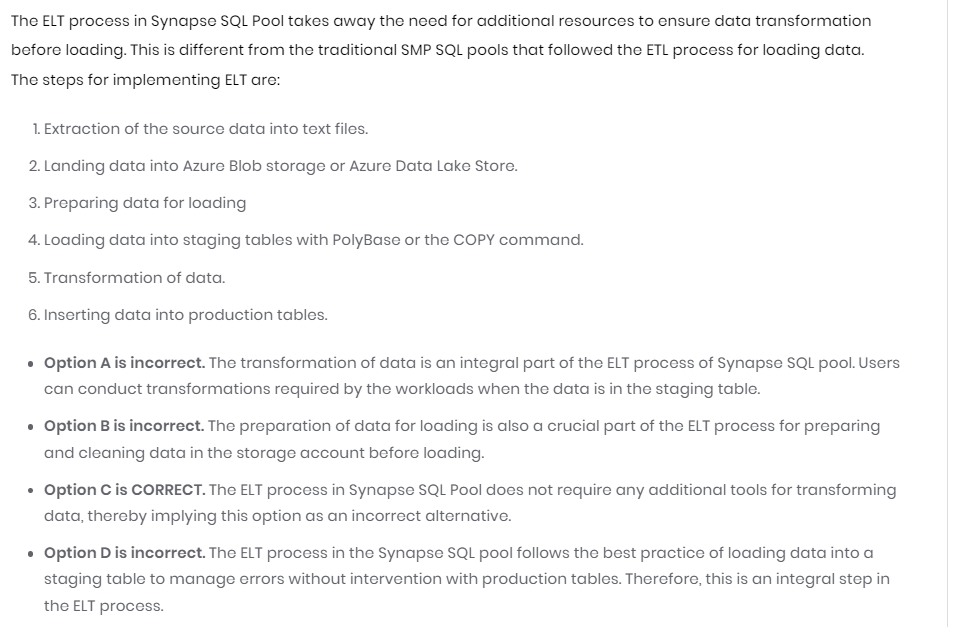
34)





35)





37) 