**Topic 4: Azure Data Warehouse Technologies**

**Introduction to Azure Data Warehouse Technologies**

## Introduction to Azure Data Warehouse Technologies

Azure is a popular platform choice for building data warehouses in the Cloud.

* **Data ingestion** is often accomplished using **blob storage** or similar technologies.
* **Data storage** can be accomplished using the warehouse platform, **Azure Synapse**.
* **Data analysis** can utilize **Azure Analysis Services**.
* Finally, **data visualization** is made possible using **Microsoft Power BI**.

A picture containing text, font, diagram, screenshot

Description automatically generated

# Lesson Outline

In this lesson, you'll learn about specific Azure data warehouse technologies.

In the first part of this lesson, we will discuss Azure Data Warehouse Solutions.

* Common architectures for end-to-end solutions
* Azure components and services that drive the solution, including
  + Azure Blob storage
  + SQL DB and Cosmos DB
  + Data Factory
  + Azure IoT Hub
  + Azure Data Lake Gen2
  + Synapse Analytics platform.

In the last part of the lesson, we will cover some of the considerations for making choices about these solutions and components. We will focus on Azure Synapse followed by an overview of Databricks Solutions.

By the end of this lesson, you will be able to

* configure some of the Azure components mentioned, including key features of those components.
* identify different Azure data engineering solutions.
* and make informed choices about when to use these components and solutions.

**Expert Perspective: Data Warehouse on Azure**

## How Do Experts Think About Data Warehouses in Azure?

Microsoft is a Gartner Magic Quadrant leader in Cloud Database Management tools.

The strengths of Azure in this space include:

* A comprehensive vision for their data tools
* A clear migration path from on premise as well as other tools
* Integrated multimodal capabilities for multiple data platforms.

Cautions in this space include:

* Inconsistent product maturity levels
* Pricing and cost challenges
* Data architecture challenges for hybrid on-premises and cloud solutions

Overall, Microsoft's Azure platform offers a comprehensive suite of tools for providing end-to-end cloud data warehousing solutions.

**Azure Data Warehouse Solutions**

## Azure Solutions

Microsoft Azure tools to create Cloud Data Warehouse solutions include:

**Data Warehouse Analytics Solutions**

* Azure Synapse for comprehensive, integrated data warehousing and analytics
  + Synapse integrates with a broad range of data storage and pipelines technologies within the Azure environment.
  + Synapse also provides analytics capabilities on top of the selected architecture.
* Azure Databricks for analytics built on Apache Spark
  + Microsoft Azure provides Databricks optimized for that Azure environment.
  + There are three flavors of Databricks optimized in turn for:
    - SQL
    - Data Science and Engineering
    - Machine Learning

Microsoft Azure tools to support Cloud Data Warehouse solutions include:

**Cloud Data Storage**

* Azure Data Warehouse Gen 2 for traditional data warehouse architectures
* Azure Dedicated SQL Pools for relational data storage
* Blob storage for file-based storage
* CosmosDB for NoSQL solutions such as column-oriented or document databases

**ETL / ELT Pipelines**

* Azure Data Factory for creating intelligent data integrations and data flows for multiple services.
* Azure Databricks for utilizing Spark to create ETL pipelines Azure Polybase for using TSQL to query blob storage in support of ELT scenarios.

#### Additional Resources

The following are links to the official Microsoft documentation.

* [Azure Synapse Overview](https://azure.microsoft.com/en-us/services/synapse-analytics/#overview)
* [Azure Databricks Overview](https://docs.microsoft.com/en-us/azure/databricks/scenarios/what-is-azure-databricks)

A screenshot of a quiz

Description automatically generated with low confidence

A screenshot of a computer

Description automatically generated with medium confidence

**Azure Data Warehouse Components**

Microsoft Azure components, technologies, and services that support Cloud Data Warehouse solutions include:

**Cloud Data Storage**

* Azure Data Warehouse Gen 2 for traditional data warehouse architectures
* Azure Dedicated SQL Pools for relational data storage
* Blob storage for file-based storage
* CosmosDB for NoSQL solutions such as column-oriented or document databases

**ETL / ELT Pipelines**

* Azure Data Factory for creating intelligent data integrations and data flows for multiple services
* Azure Databricks for utilizing Spark to create ETL pipelines Azure Polybase for using TSQL to query blob storage in support of ELT scenarios

### Additional Resources

The following are links to the official Microsoft documentation on various data warehouse components.

* [Azure Data Warehouse Gen 2 Overview](https://docs.microsoft.com/en-us/azure/storage/blobs/data-lake-storage-introduction)
* [Azure Dedicated SQL Pool Overview](https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-overview-what-is)
* [Azure Blob Storage Overview](https://docs.microsoft.com/en-us/azure/storage/blobs/storage-blobs-introduction)
* [Azure Cosmos DB Overview](https://docs.microsoft.com/en-us/azure/cosmos-db/introduction)
* [Azure Data Factory Overview](https://docs.microsoft.com/en-us/azure/data-factory/introduction)
* [Azure Databricks Overview](https://docs.microsoft.com/en-us/azure/databricks/scenarios/what-is-azure-databricks)

A diagram of a storage components

Description automatically generated with low confidence

A picture containing text, screenshot, font, diagram

Description automatically generated

**Azure Components Quiz**

A screenshot of a computer

Description automatically generated with medium confidence

**When To Use Azure Data Warehouse Components**

Azure Data Warehouse components are similar to offerings by both AWS and Google. Here are sometimes when you should use Azure Data Warehouse components to build a data warehouse solution:

* When the data infrastructure already contains Microsoft technologies such as Microsoft SQL Server
* When an on-premises solution needs to be moved to the cloud for scaling
* When you have large amounts of data that need an ELT solution to quickly ingest data from a wide variety of sources

**Lesson Review**

In this lesson, we investigated Azure data warehouse solutions. This included looking at the components involved in the solutions, and the service choices available for each of them. These components include

* Azure Blob storage
* SQL DB and Cosmos DB
* Data Factory
* Azure IoT Hub
* Azure Data Lake Gen2
* Synapse Analytics platform.

In addition to these components, we covered circumstances and requirements that might be ideal for making choices about them.