✓ 100 XP ▶

Introduction

1 minute

Large-scale data warehousing and analytics involves two key elements, *data ingestion* and *data processing*. Data ingestion is the process used to load data from various sources into a central data store. Once ingested, the data becomes available for use. Data can be ingested using batch processing or streaming, depending on the nature of the data source. Data processing involves operations on the data to clean, filter, restructure, and prepare the data for analysis.

Data lives in many locations throughout an organization. When you design your cloud data solution, you'll want to ingest your raw data into a data store for analysis. A common approach that you can use with Azure Synapse Analytics is to *extract* the data from where it's currently stored, *load this data into an analytical data store*, and then *transform* the data, shaping it for analysis. This approach is known as **ELT**, for extract, load, and transform.

Azure Synapse Analytics is particularly suitable for this approach. Using Apache Spark and automated pipelines, Synapse Analytics can run parallel processing tasks across massive datasets, and perform *big data* analytics.

> ⓘ **Note**
>
> The term *big data* refers to data that is too large or complex for traditional database systems. Systems that process big data have to perform rapid data ingestion and processing; they must have capacity to store the results, and sufficient compute power to perform analytics over these results.

Another option is to analyze operational data in its original location. This strategy is known as *hybrid transactional analytical processing* (HTAP). You can perform this style of analysis over data held in repositories such as Azure Cosmos DB using Azure Synapse Link.

In this module, you'll explore some of these tools and techniques that you can use to ingest and process data with Azure.

Learning objectives

In this module, you will:

- Describe data ingestion in Azure
- Describe components of Azure Data Factory
- See how to use Azure Data Factory to load data into a data warehouse
- Describe data processing options for performing analytics in Azure
- Explore Azure Synapse Analytics

Next unit: Describe common practices for data loading

Continue >