**Topic 5: Implementing Data Warehouses in Azure**

**Lesson Introduction**

## Lesson Introduction

In this lesson, we will implement a data warehouse in Azure.

We'll start by looking at Azure data warehouse architecture. Next, we'll see how to ingest data into Azure using linked services, pipelines, and triggers. In the last part of the lesson we will implement ELT in Azure Synapse using staging tables and SQL to SQL, extract, load, and transform.

By the end of this lesson, you'll be able to

* Describe Azure data warehouse architecture.
* configure Azure Synapse Analytics
* ingest data into Synapse and perform extract, load, and transform with SQL.

**Data Warehouses in Azure: A Closer Look**

Creating data warehouses using Microsoft Azure technologies requires an understanding of the purpose of each component. Azure data warehouse pricing models include:

* Batch Data Ingestion
* Streaming Data Ingestion
* SQL Data Storage
* NoSQL Data Storage
* Big Data Analytics
* Log and Telemetry Streaming Analytics

Your data warehouse solution in Azure may include all or just some of these options and the pricing models vary from the amount of data stored/processed to time-based billing.

**Benefits** of the mix and match approach:

* Pay for just what you need.
* Tailor solution to business needs not the platform.

**Cautions** for the mix and match approach:

* Need to understand a variety of billing models.
* Must understand data and time needs for each of the components to understand the total cost for the solution.

A diagram of data storage

Description automatically generated with medium confidence

**Azure Data Warehouse Architectures**

### Enterprise BI Architecture

A typical Enterprise DW Architecture involves:

* A Data Source (MS SQL Server)
* An Ingestion Mechanism (Azure Blob Storage)
* Data Storage (Azure Synapse)
* Analysis (Azure Analysis Service)
* Visualization (PowerBI)

### Automated Enterprise BI Architecture

An Automated Enterprise BI Architecture has the same components with the Azure Blob Storage and Azure Synapse components orchestrated using Azure Data Factory.

### Additional Resource

Microsoft documentation on [Azure data warehouse architecture](https://docs.microsoft.com/en-us/azure/architecture/data-guide/relational-data/data-warehousing).

A diagram of data storage

Description automatically generated with low confidence

The above architecture can be modified. For example, in the figure below, by adding the robust orchestration and integration capabilities of Azure Data Factory to the architecture, the entire data pipeline can be **automated**.

A picture containing text, screenshot, diagram, font

Description automatically generated

**Azure Data Warehouse Architectures Quiz**

A screenshot of a computer

Description automatically generated with low confidence

Each step of the Enterprise Data Warehouse architecture maps to a distinct Azure Solution.

**Ingesting Data at Scale into Azure Synapse**

Ingesting data at scale into Azure Synapse involves: (steps)

* Creating linked services
  + a linked service is where you define your connection information to other services
* Creating a pipeline
  + A pipeline contains the logical flow for an execution of a set of activities
* Using a trigger or a one-time data ingestion
  + You can manually start a data ingestion or you can schedule a trigger

### Learn More

[Ingesting data into Azure Synapse](https://docs.microsoft.com/en-us/azure/synapse-analytics/quickstart-copy-activity-load-sql-pool)

A picture containing text, screenshot, diagram, font

Description automatically generated

A picture containing text, diagram, screenshot, line

Description automatically generated

A screenshot of a quiz

Description automatically generated with medium confidence

**SQL to SQL ELT in Azure**

Doing a SQL to SQL ELT in Azure involves:

* Starting with data ingested into either Blob Storage or Azure Delta Lake Gen 2
* Create EXTERNAL staging tables in the Data Warehouse
* Transform data from staging tables to DW tables

### Dedicated or Serverless SQL Pools?

In Azure Synapse Analytics, there are two types of SQL pools to choose from: dedicated SQL pools and serverless SQL pools. You can use either for ELT pipeline processes and staging data. They offer different functionalities, performance characteristics, and pricing models.

* Serverless SQL Pools are created when you create a Synapse Analytics resource in Azure.
* Dedicated SQL Pools have to be created as a separate resource.

**Azure Dedicated SQL Pools**

Dedicated SQL pools are designed to provide a high-performance, scalable, and cost-effective solution for big data workloads. They utilize a Massively Parallel Processing (MPP) architecture. This architecture enables users to perform queries faster, especially for complex analytical queries. Dedicated SQL pools are provisioned with a fixed amount of resources and are billed based on the resources allocated, regardless of usage.

* In a Synapse Analytics production data warehouse, you likely would use Dedicated SQL Pools.

**Azure Synapse Analytics Serverless SQL Pools**

Serverless SQL pools provide a pay-per-query model, which means you only pay for the resources used by each executed query. They are designed to handle both small and large-scale data processing tasks and can automatically scale resources based on workload requirements. Serverless SQL pools do not require any upfront provisioning or resource allocation.

* For development workloads, ad-hoc querying, or volatile workloads, you can use serverless SQL pools.
* Use Serverless SQL pools if you are using the Udacity Azure Cloud Lab.

### Additional Resource

Documentation on [SQL ELT in Azure](https://docs.microsoft.com/en-us/azure/architecture/data-guide/relational-data/etl) from Microsoft.

A picture containing text, diagram, screenshot, font

Description automatically generated

**Course Summary**

## Course Summary

This course has covered a lot of material related to data warehouses on Azure. The first lesson was an introduction to the course and an overview of:

* course prerequisites
* tools and environments you’ll use to build **cloud data warehouses**
* the project you’ll do at the end of the course
* stakeholders you might encounter as a data engineer in this space
* the history of data warehousing

In the first part of the course, we covered some foundational topics in data warehousing

* What a data warehouse is, and the role they play in providing valuable data to an organization.
* The architecture of a data warehouse.
* **Dimensional data modeling** techniques for data warehouses.
* Storage technologies for optimizing data warehouses.

The next part of the course was about data warehouses in the cloud including:

* Components of a cloud data warehouse solution
* **ELT** and **ETL**
* Storage options for building data warehouses in a cloud environment
* Pipeline solutions
* Cloud data warehouse solutions

The last part of the course covered data warehouse technologies in the Azure cloud platform. You learned about the components of Azure data warehouses and:

* Implementation;
* architecture;
* Ingesting data;
* as well as extracting, loading, and transforming data.