



# DP-200T01: Working with Relational Data Stores in the Cloud



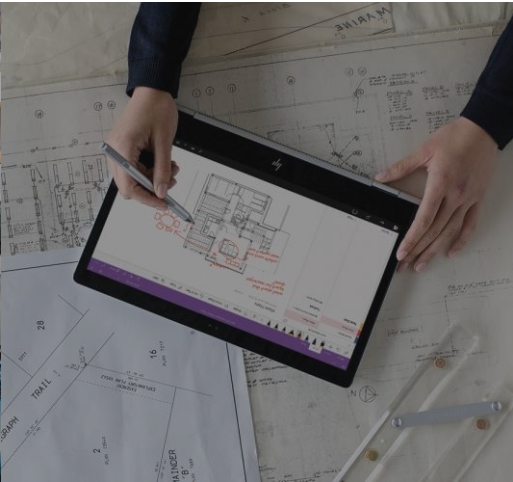
# Agenda

- L01 - Work with Azure SQL Database
- L02 - Work with Azure Synapse Analytics
- L03 - Provision and query data in Azure Synapse Analytics
- L04 - Import data into Azure Synapse Analytics using PolyBase



# Lesson 01

## Azure SQL Database



# Lesson Objectives

- Why Azure SQL Database is a good choice for running your relational database
- What configuration and pricing options are available for your Azure SQL database
- How to create an Azure SQL database from the portal
- How to use Azure Cloud Shell to connect to your Azure SQL database, add a table, and work with data

# Why Azure SQL Database is a good choice

Convenience

Cost

Scale

Security

# Azure SQL Database configuration options

When you create your first Azure SQL database, you also create an *\_Azure SQL logical server\_*. Think of a logical server as an administrative container for your databases.

## DTUs

DTU stands for Database Transaction Unit and is a combined measure of compute, storage, and IO resources. Think of the DTU model as a simple, preconfigured purchase option

## vCores

vCore gives you greater control over what compute and storage resources you create and pay for. vCore model enables you to configure resources independently

## SQL elastic pools

SQL elastic pools relate to eDTUs. They enable you to buy a set of compute and storage resources that are shared among all the databases in the pool. Each database can use the resources they need

## SQL Managed Instances

The SQL managed instance creates a database with near 100% compatibility with the latest SQL Server on-premises Enterprise Edition database engine, useful for SQL Server customers who would like to migrate on-premises servers instance in a "lift and shift" manner

## Create SQL Database

Microsoft

Basics • Networking Additional settings Tags Review + create

Create a SQL database with your preferred configurations. Complete the Basics tab then go to Review + Create to provision with smart defaults, or visit each tab to customize. [Learn more](#)

### Project details

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.

Subscription \* ⓘ

Resource group \* ⓘ  [Create new](#)

### Database details

Enter required settings for this database, including picking a logical server and configuring the compute and storage resources

Database name \*

Server \* ⓘ  [Create new](#)

✘ The value must not be empty.

Want to use SQL elastic pool? \* ⓘ  Yes  No

Compute + storage \* ⓘ **Please select a server first.**  
[Configure database](#)

Create an Azure SQL Database.

Use Azure Cloud Shell to connect to your Azure SQL database

```
sqlcmd -S tcp:contoso-1.database.windows.net,1433  
-d Logistics -U martina -P "password1234$" -N -I 30
```

```
CREATE TABLE Drivers (DriverID int, LastName  
varchar(255), FirstName varchar(255), OriginCity  
varchar(255)); GO
```

```
SELECT name FROM sys.tables; GO
```

```
INSERT INTO Drivers (DriverID, LastName,  
FirstName, OriginCity) VALUES (123, 'Zirne', 'Laura',  
'Springfield'); GO
```



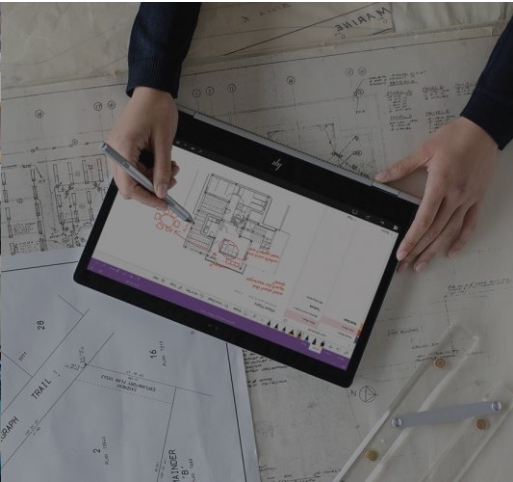
# Review Questions

- Q01 – Who's responsible for performing software updates on your Azure SQL databases and the underlying OS?
- A01 – Microsoft Azure. Azure manages the hardware, software updates, and OS patches for you.
  
- Q02 - What is an Azure SQL logical server?
- A02 - An administrative container for your databases.
  
- Q03 - Your Azure SQL database provides adequate storage and compute power. But you find that you need additional IO throughput. Which performance model might you use?
- A03 – vCore.



# Lesson 02

## Azure Synapse Analytics



# Lesson Objectives

- Explain Azure Synapse Analytics
- Explain Azure Synapse Analytics features
- Types of solution workloads
- Explain Massively Parallel Processing concepts
- Compare table geometries

# Azure Synapse Analytics

## What is Azure Synapse Analytics

A unified environment by combining the enterprise data warehouse of SQL, the Big Data analytics capabilities of Spark, and data integration technologies to ease the movement of data between both, and from external data sources.

## Data warehouse capabilities.

### SQL Analytics

A centralized data warehouse store that provides a relational analytics and decision support services across the whole enterprise

### SQL Pools

CPU, memory, and IO are bundled into units of compute scale called SQL, determined by Data Warehousing Units (DWU)

### Future features

Will include a Spark engine, a data integration and Azure Synapse Analytics Studio

# Azure Synapse Analytics features

## Workload Management

This capability is used to prioritize the query workloads that take place on the server using Workload Management. This involves three components:

- Workload Groups
- Workload Classification
- Workload Importance

## Result-Set Cache

Result-set caching can be used to improve the performance of the queries that retrieve these results. When result-set caching is enabled, the results of the query are cached in the SQL pool storage.

## Materialized Views

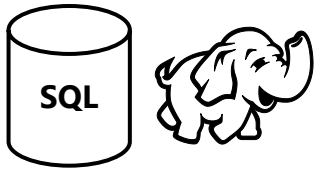
A materialized view pre-computes, stores, and maintains its data like a table. They are automatically updated when data in underlying tables are changed.

## SSDT CI/CD support

Database project support in SQL Server Data Tools (SSDT) allows teams of developers to collaborate over a version-controlled Azure Synapse Analytics, and track, deploy and test schema changes

# Types of solution workloads

The modern data warehouse extends the scope of the data warehouse to serve Big Data that's prepared with techniques beyond relational ETL



## Modern data warehousing

---

"We want to integrate all our data—including Big Data—with our data warehouse"



## Advanced analytics

---

"We're trying to predict when our customers churn"

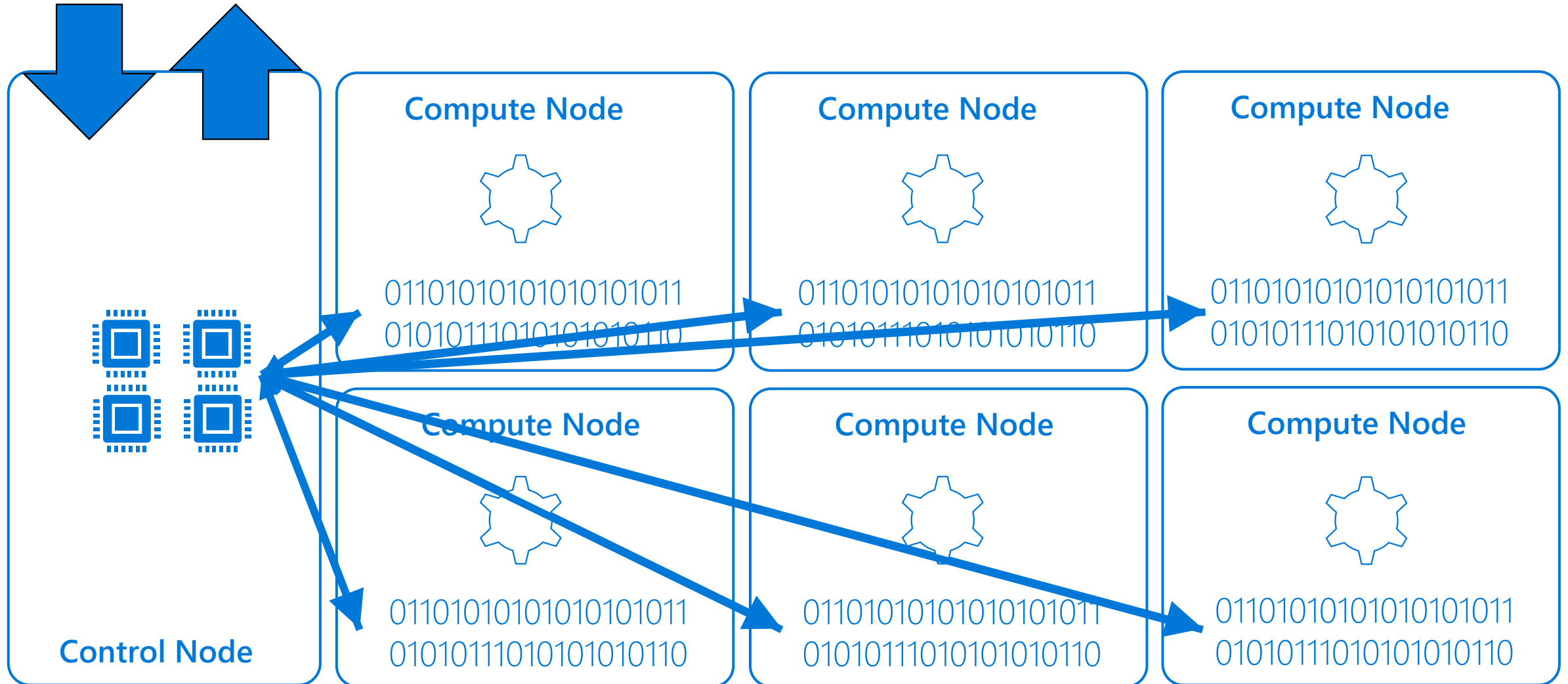


## Real-time analytics

---

"We're trying to get insights from our devices in real-time"

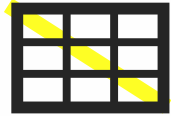
# Massively Parallel Processing (MPP) concepts



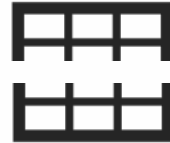
# Table geometries

## Table distribution

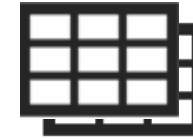
Round Robin  
Tables



Hash Distributed  
Tables



Replicated  
Tables





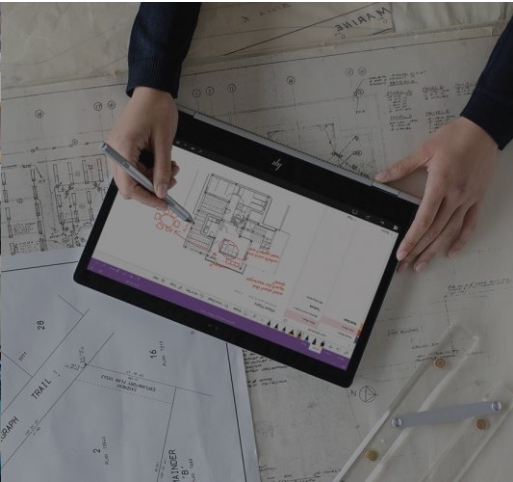
# Review Questions

- Q01 - Which of the following terms refer to the scale of compute that is being used in an Azure Synapse Analytics database?
- A01 – DWU.
  
- Q02 - You have an Azure Synapse Analytics database, within this, you have a dimension table named Stores that contains store information. There is a total of 263 stores nationwide. Store information is retrieved in more than half of the queries that are issued against this database. These queries include staff information per store, sales information per store and finance information. You want to improve the query performance of these queries by configuring the table geometry of the stores table. Which is the appropriate table geometry to select for the stores table?
- A02 – Replicated Table.



# Lesson 03

## Creating and Querying an Azure Synapse Analytics



# Lesson Objectives

- Create an Azure Synapse Analytics sample database
- Query the sample database with the SELECT statement and its clauses
- Use the queries in different client applications such as SQL Server Management Studio, and PowerBI

# Create an Azure Synapse Analytics

Home > New > Azure Synapse Analytics (formerly SQL DW) > SQL Data Warehouse

## SQL Data Warehouse

Microsoft

Welcome to Azure Synapse Analytics (formerly known as Azure SQL Data Warehouse). [Learn more!](#)

Basics • **Additional settings\*** Tags Review + create

Create a SQL data warehouse with your preferred configurations. Complete the Basics tab then go to Review + Create to provision with smart defaults, or visit each tab to customize. [Learn more](#)

### Project details

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.

Subscription \* ⓘ

chtestao

Resource group \* ⓘ

Select existing...

[Create new](#)

### Data warehouse details

Enter required settings for this data warehouse, including picking a logical server and configuring the performance level.

Data warehouse name \*

Enter data warehouse name

Server \* ⓘ

Select a server

[Create new](#)

✘ The value must not be empty.

Performance level \* ⓘ

Please select a server first.

[Select performance level](#)

# Perform Azure Synapse Analytics Queries.

## **SELECT Query Basics**

```
SELECT <select_list>  
[FROM <optional_from_specification>]  
[WHERE <optional_filter_condition>]  
[ORDER BY <optional_sort_specification>]  
[JOIN <optional_join_specification>]
```

## **Examples**

```
SELECT *  
FROM Products p WHERE p.id ="1"
```

```
SELECT p.id, p.manufacturer, p.description  
FROM Products p WHERE p.id ="1"
```

```
SELECT p.price, p.description, p.productId  
FROM Products p ORDER BY p.price ASC
```

```
SELECT p.productId  
FROM Products p JOIN p.shipping
```

# Perform Azure Synapse Analytics Queries.

## Create Table as Select (CTAS)

Used in parallel data loads

```
CREATE TABLE
[ database_name . [ schema_name ] . |
schema_name. ] table_name
    [ ( { column_name } [ ,...n ] ) ]
WITH ( DISTRIBUTION =
    { HASH( distribution_column_name )
      REPLICATE | ROUND_ROBIN }
    [ , <CTAS_table_option> [ ,...n ] ]
)
AS <select_statement> [;]
```

### Example

```
CREATE TABLE FactInternetSales_Copy
```

```
WITH
```

```
(DISTRIBUTION = HASH(SalesOrderNumber))
```

```
AS SELECT * FROM FactInternetSales
```

Querying with different client applications.

Connect to Server

## SQL Server

Server type: Database Engine

Server name: sqlservercto.database.windows.net

Authentication: SQL Server Authentication

Login: ctesta-oneill

Password: .....

Remember password

Connect Cancel Help Options >>

# Review Questions

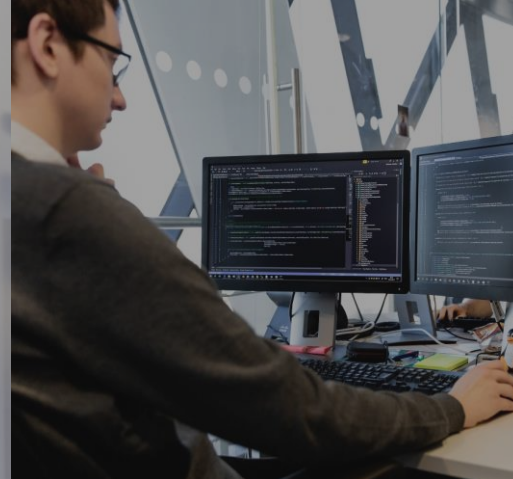
- Q01 – What is the default port for connecting to an Azure SQL Server Data Warehouse?
- A01 – TCP port 1433
- Q02 - The following query is to retrieve the sales by business reseller, but the performance of the query is slow. The query is as follows:

```
SELECT  
S.[SalesAmount] AS [Sales],  
R.[BusinessType],  
R.[ResellerName]  
FROM [FactResellerSales] AS S  
JOIN [DimReseller] AS R  
ON S.[ResellerKey] = R.[ResellerKey].
```

The tables referenced within the query are configured with a distribution of Round\_Robin with a clustered columnstore index. The Data Engineer wants to improve the performance of the query. What operation can be used to improve the performance of the query?

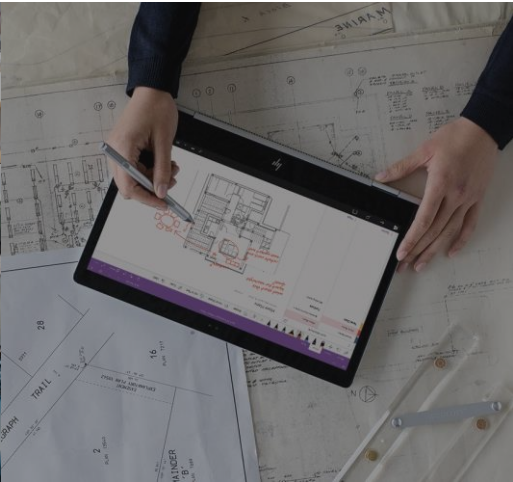
- A02 – Change the Distribution to HASH(ResellerKey) for both tables.





# Lesson 04

## Using PolyBase to Load Data in Azure Synapse Analytics

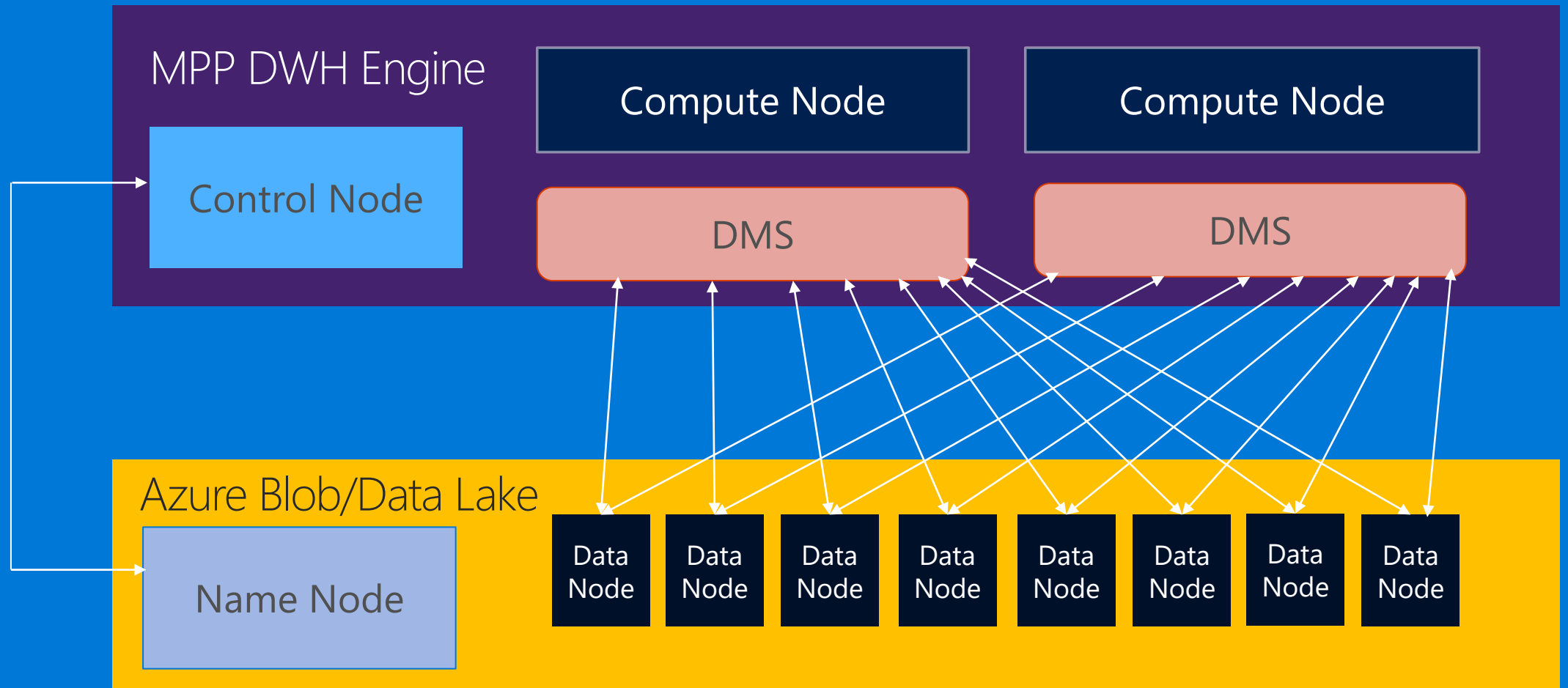


# Lesson Objectives

- Explore how PolyBase works
- Upload text data to Azure Blob store
- Collect the security keys for Azure Blob store
- Create an Azure Synapse Analytics
- Import data from Blob Storage to the Data Warehouse

# How PolyBase works

## The MPP engine's integration method with PolyBase



# Upload text data to Azure Blob store

## Create storage account

[Basics](#) [Advanced](#) [Tags](#) [Review + create](#)

Azure Storage is a Microsoft-managed service providing cloud storage that is highly available, secure, durable, scalable, and redundant. Azure Storage includes Azure Blobs (objects), Azure Data Lake Storage Gen2, Azure Files, Azure Queues, and Azure Tables. The cost of your storage account depends on the usage and the options you choose below. [Learn more](#)

### PROJECT DETAILS

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.

\* Subscription

\* Resource group

[Create new](#)

### INSTANCE DETAILS

The default deployment model is Resource Manager, which supports the latest Azure features. You may choose to deploy using the classic deployment model instead. [Choose classic deployment model](#)

\* Storage account name ⓘ

\* Location

Performance ⓘ  Standard  Premium

Account kind ⓘ

Replication ⓘ

Access tier (default) ⓘ  Cool  Hot

Review + create

Previous

Next : Advanced >

# Collect the Storage keys


ctoazureblob - Access keys

## Access keys

Use access keys to authenticate your applications when making requests to this Azure storage account. Store your access keys securely - for example, using Azure Key Vault - and don't share them. We recommend regenerating your access keys regularly. You are provided two access keys so that you can maintain connections using one key while regenerating the other.


When you regenerate your access keys, you must update any Azure resources and applications that access this storage account to use the new keys. This action will interrupt access to disks from your virtual machines. [Learn more](#)

Storage account name  
ctoazureblob

**key1** 

Key  
eU7[REDACTED]Cg==

Connection string  
Def[REDACTED]9YrQ...

**key2** 

Key  
NWC[REDACTED]pUgB5w==

Connection string  
Def[REDACTED]s6...

# Create an Azure Synapse Analytics

Home > New > Azure Synapse Analytics (formerly SQL DW) > SQL Data Warehouse

## SQL Data Warehouse

Microsoft

Welcome to Azure Synapse Analytics (formerly known as Azure SQL Data Warehouse). [Learn more!](#)

Basics • **Additional settings\*** Tags Review + create

Create a SQL data warehouse with your preferred configurations. Complete the Basics tab then go to Review + Create to provision with smart defaults, or visit each tab to customize. [Learn more](#)

### Project details

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.

Subscription \* ⓘ

chtestao

Resource group \* ⓘ

Select existing...

[Create new](#)

### Data warehouse details

Enter required settings for this data warehouse, including picking a logical server and configuring the performance level.

Data warehouse name \*

Enter data warehouse name

Server \* ⓘ

Select a server

[Create new](#)

✘ The value must not be empty.

Performance level \* ⓘ

Please select a server first.

[Select performance level](#)

# Create external objects

--STEP 1: Create an external data source

```
-- DROP EXTERNAL DATA SOURCE FXR_TEST_DSRC;
```

```
CREATE EXTERNAL DATA SOURCE FXR_TEST_DSRC
```

```
  WITH ( TYPE = HADOOP
```

```
    , LOCATION = 'hdfs://192.168.210.145:8020'
```

```
    , JOB_TRACKER_LOCATION = '192.168.210.145:8032'
```

```
      ---- defaults:8021 - Cloudera 4.3; 8032 - HDP 2.x on Windows | Cloudera 5.1;
```

```
      ----          8050 - HDP 2.x on Linux; 50300 - HDP 1.3
```

```
  );
```

--STEP 2: Create an external file format.

```
--DROP EXTERNAL FILE FORMAT FXR_Test_Format;
```

```
CREATE EXTERNAL FILE FORMAT FXR_Test_Format
```

```
  WITH ( FORMAT_TYPE = DELIMITEDTEXT
```

```
    , FORMAT_OPTIONS ( FIELD_TERMINATOR = N';'
```

```
    , USE_TYPE_DEFAULT = TRUE
```

```
    , STRING_DELIMITER = '' )
```

```
  );
```

# Import data into an external table

```
--STEP 3: Create a new external table in SQL Server MPP SQL
-- DROP EXTERNAL TABLE ExternalTest;
CREATE EXTERNAL TABLE ExternalTest
    (name nvarchar(17), startzeitpunkt nvarchar(35),
    endzeitpunkt varchar(35), flms_system_realtime nvarchar(19),
    dummy nvarchar(19) NULL, Counter1DTonDur nvarchar(19),
    Counter1DMileage nvarchar(19), dummy2 nvarchar(2) NULL
    )
WITH
    (LOCATION = '/user/fxr47511/pdwtest'
    , DATA_SOURCE = FXR_TEST_DSRC
    , FILE_FORMAT = FXR_Test_Format
    , REJECT_TYPE = value
    , REJECT_VALUE = 1000
    );
```



# Load data into an Azure Synapse table

--STEP 4: Create a new table.

```
CREATE TABLE [dbo].[StageDate]
WITH (
    CLUSTERED COLUMNSTORE INDEX,
    DISTRIBUTION = ROUND_ROBIN
)
AS
SELECT * FROM [dbo].[ExternalTest];
```

# Review Questions

- Q01 – Mike is the data engineer for Contoso and has a Data Warehouse created with a database named Crystal. Within the database is a table named DimSuppliers. The suppliers' information is stored in a single text file named Suppliers.txt and is 1200MB in size. It is currently stored in a container with an Azure Blob store. Your Azure Synapse Analytics is configured as Gen 2 DW30000c. How can Mike maximize the performance of the data load?
- **A01 – Split the text file into 60 files of 20MB each.**
- Q02 – Mike is the data engineer for Contoso and has a Data Warehouse created with a database name Crystal. He has created a master key, followed by a database scoped credential. What should he create next?
- **A02 – An external data source.**

# Lab: Working with Relational Data Stores in the Cloud



# Lab overview

The students will be able to provision an Azure SQL Database and Azure Synapse Analytics and be able to issue queries against one of the instances that are created. They will be also be able to integrate SQL Data Warehouse with a number of other Data platform technologies and use PolyBase to load data from one data source into Azure Synapse Analytics.

# Lab objectives

After completing this lab, you will be able to:

1. Use Azure SQL Database
2. Describe Azure Data Warehouse
3. Creating and Querying an Azure Synapse Analytics
4. Using PolyBase to Load Data into Azure Synapse Analytics

# Lab scenario

You are the senior data engineer at AdventureWorks, and you are working with your team to transition relational database systems from on-premises SQL Servers to relational database located in Azure. You will begin by creating an instance of SQL Database that will be handed of the junior data engineers to migrate some departmental databases.

You will then provision a SQL Data Warehouse and test that the provisioning of the server is successful by testing a sample database with a series of queries. You will then use PolyBase to load dimension tables from Azure Blob and Azure Databricks to test that the integration of these data platform technologies with Azure Synapse Analytics.

At the end of this lab, you will have:

1. Use Azure SQL Database
2. Describe Azure Data Warehouse
3. Creating and Querying an Azure Synapse Analytics
4. Using PolyBase to Load Data into Azure Synapse Analytics

# Lab review

- Exercise 1 – Can you think how Azure SQL Database can help your organization?
- Exercise 2 – Are there any students in the classroom who would consider SQL Data Warehouse a replacement for an existing solution?
- Exercise 3 – Can you describe the table options that are available in Azure Synapse Analytics?
- Exercise 4 – Why is it best practice to use PolyBase to load data into Azure Synapse Analytics?

# Module Summary >

## In this module, you have learned about:

- Work with Azure SQL Database.
- Work with SQL Data Warehouse.
- Provision and query data in Azure Synapse Analytics.
- Import data into Azure Synapse Analytics using PolyBase.

## Next steps >

After the course, use this [cheat sheet](#) as a reminder of the features of Azure Synapse Analytics.

