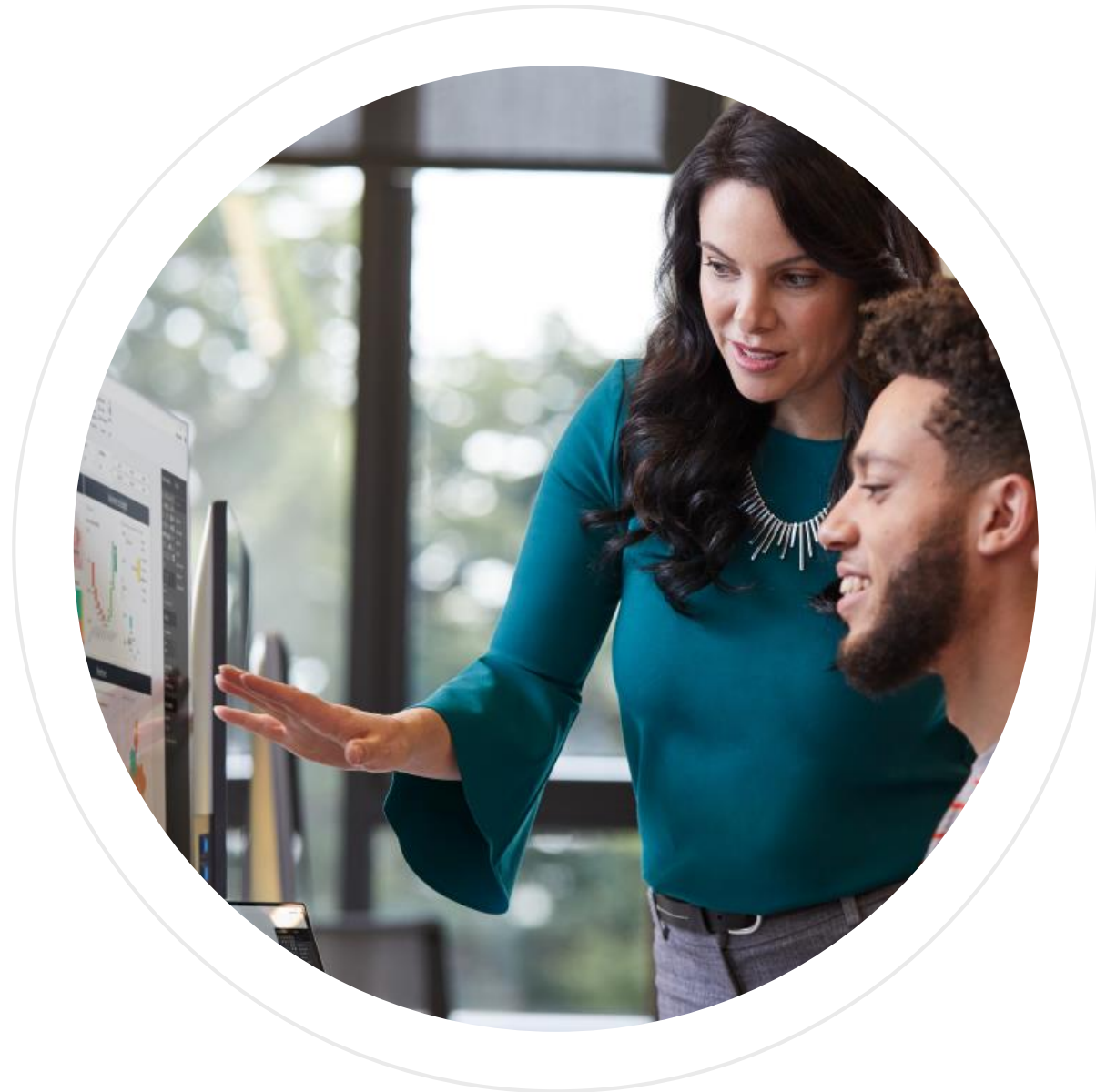


# M02: Prepare Data in Power BI

Mohammed Arif



# Module Agenda



Get data from various data sources



Optimize Performance



Resolve Data Errors

# Lesson 1: Data Analytics and Microsoft



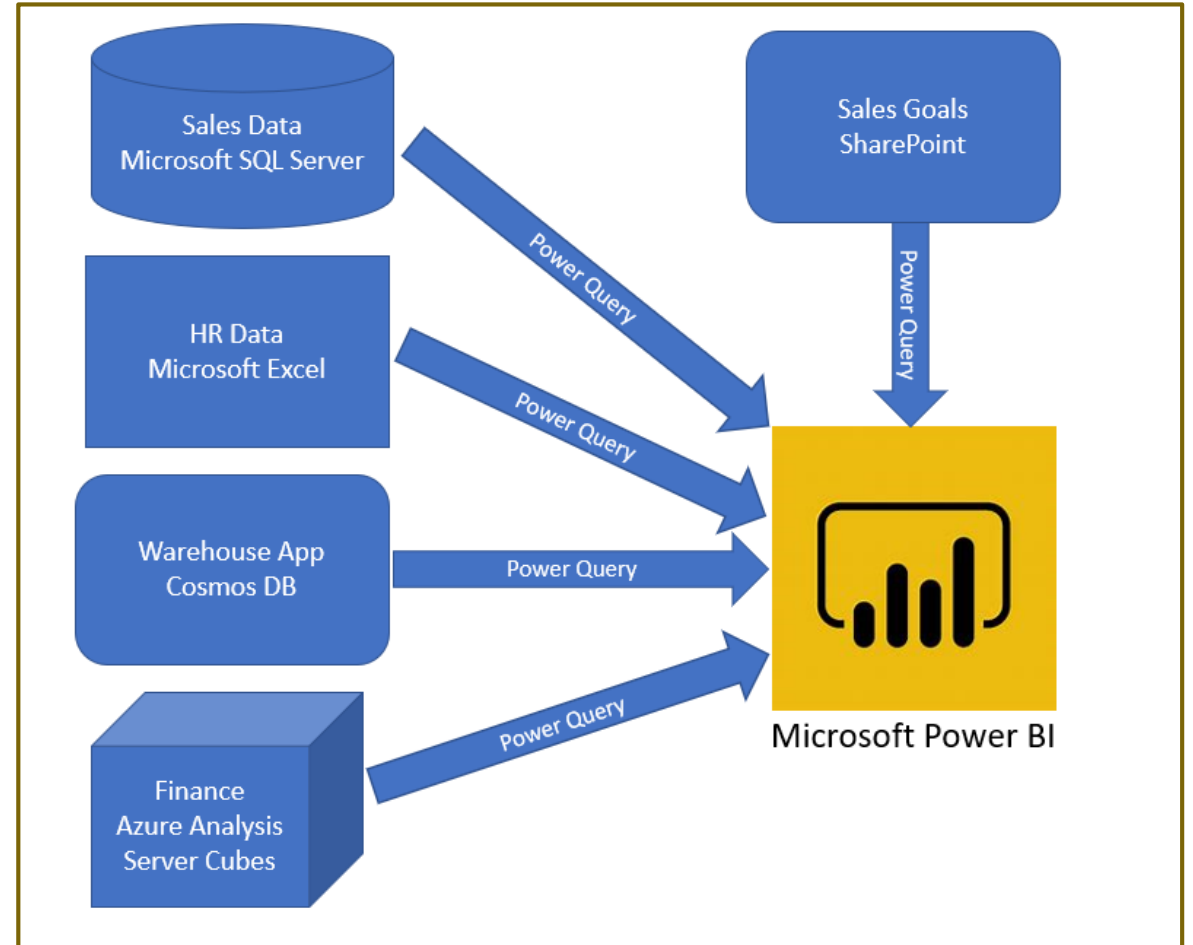
# Learning Objectives

You will learn the following concepts:

- Getting data from various data sources
- Optimizing Performance
- Resolving data errors

# Introduction to getting data

The first step in the data analysis process is identifying and getting data.



# Get data from flat files



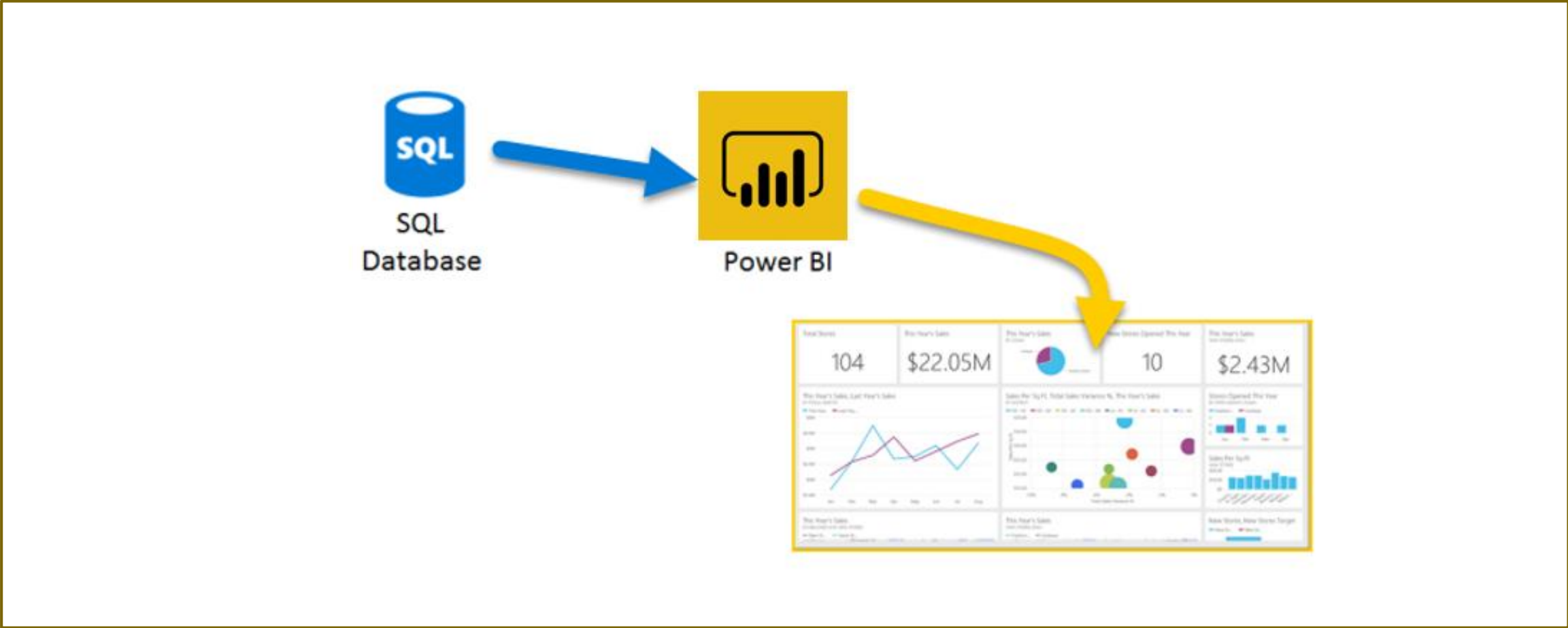
ResellerSales\_202006.csv

File Origin: 1200: Unicode | Delimiter: Comma | Data Type Detection: Based on first 200 rows

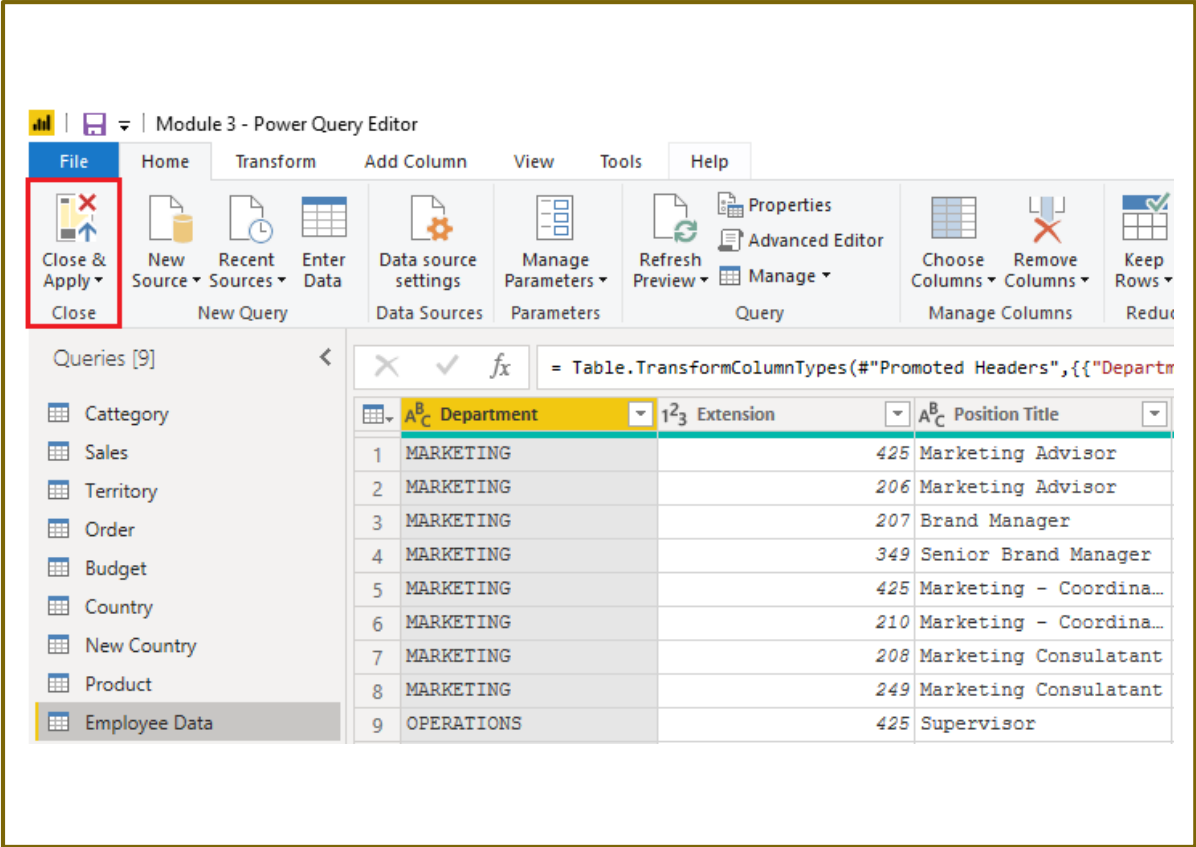
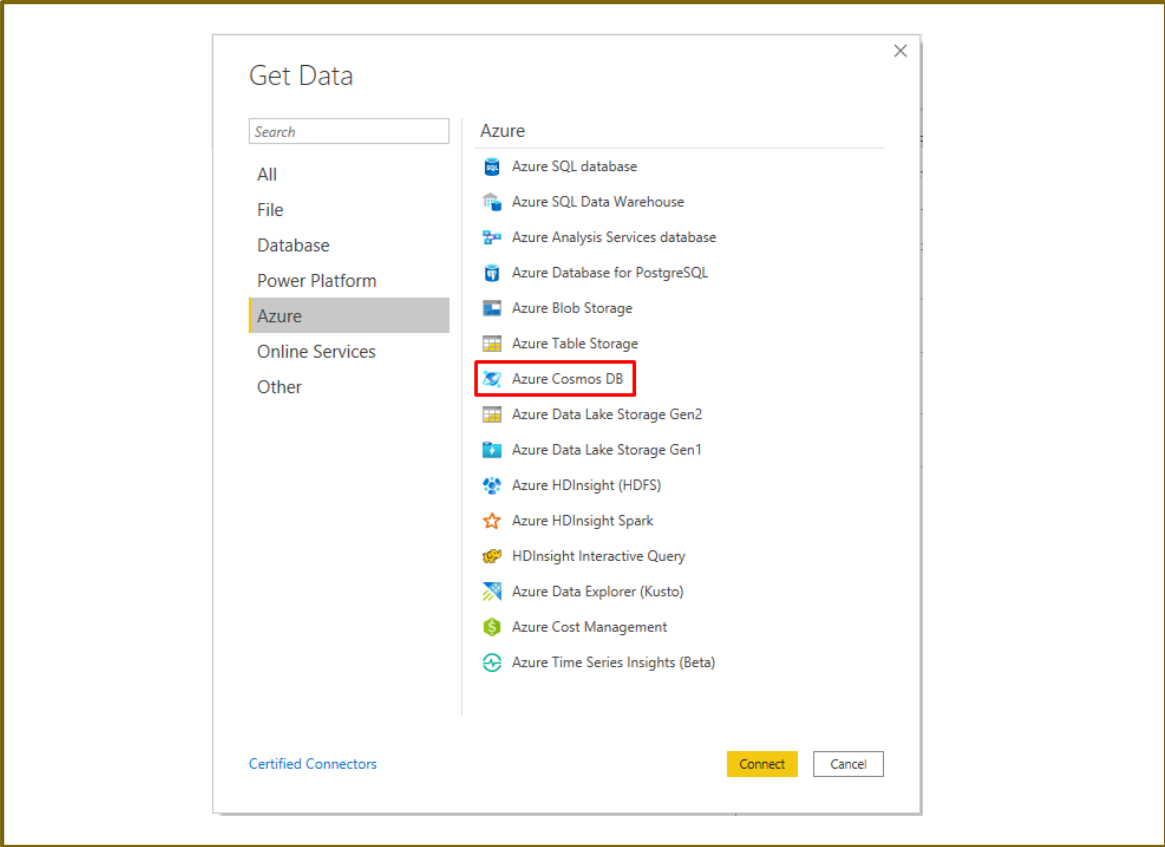
SalesOrderNumber	SalesOrderLineNumber	OrderDate	DueDate	ShipDate	ProductKey	ResellerKey	PromotionKey
S071691	2	6/1/2020	6/11/2020	6/8/2020	434	104	
S071691	4	6/1/2020	6/11/2020	6/8/2020	222	104	
S071774	1	6/1/2020	6/11/2020	6/8/2020	436	609	
S071774	2	6/1/2020	6/11/2020	6/8/2020	418	609	
S071775	1	6/1/2020	6/11/2020	6/8/2020	573	595	
S071775	2	6/1/2020	6/11/2020	6/8/2020	555	595	
S071775	3	6/1/2020	6/11/2020	6/8/2020	490	595	
S071776	1	6/2/2020	6/12/2020	6/9/2020	514	106	
S071777	1	6/2/2020	6/12/2020	6/9/2020	408	128	
S071777	2	6/2/2020	6/12/2020	6/9/2020	436	128	
S071778	1	6/2/2020	6/12/2020	6/9/2020	467	557	
S071778	2	6/2/2020	6/12/2020	6/9/2020	566	557	

Buttons: Load, Transform Data, Cancel

# Get data from relational data sources

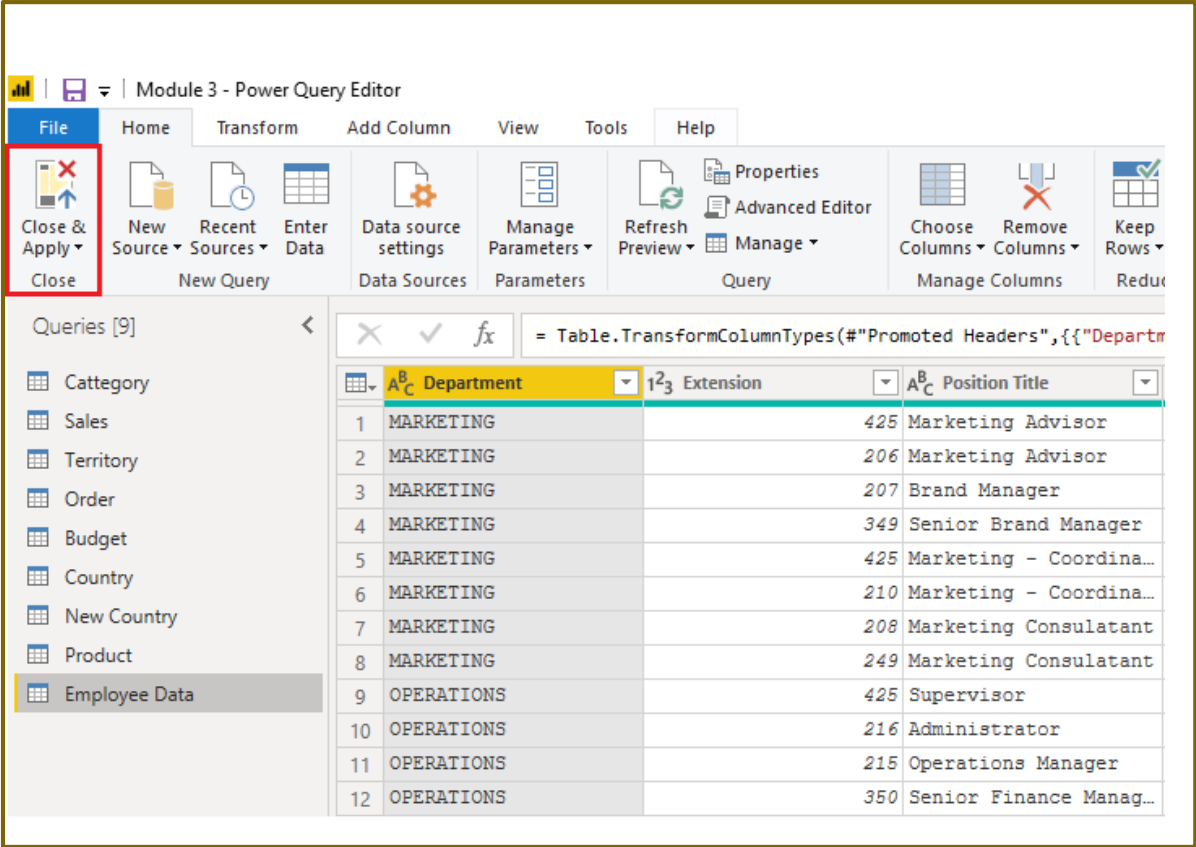
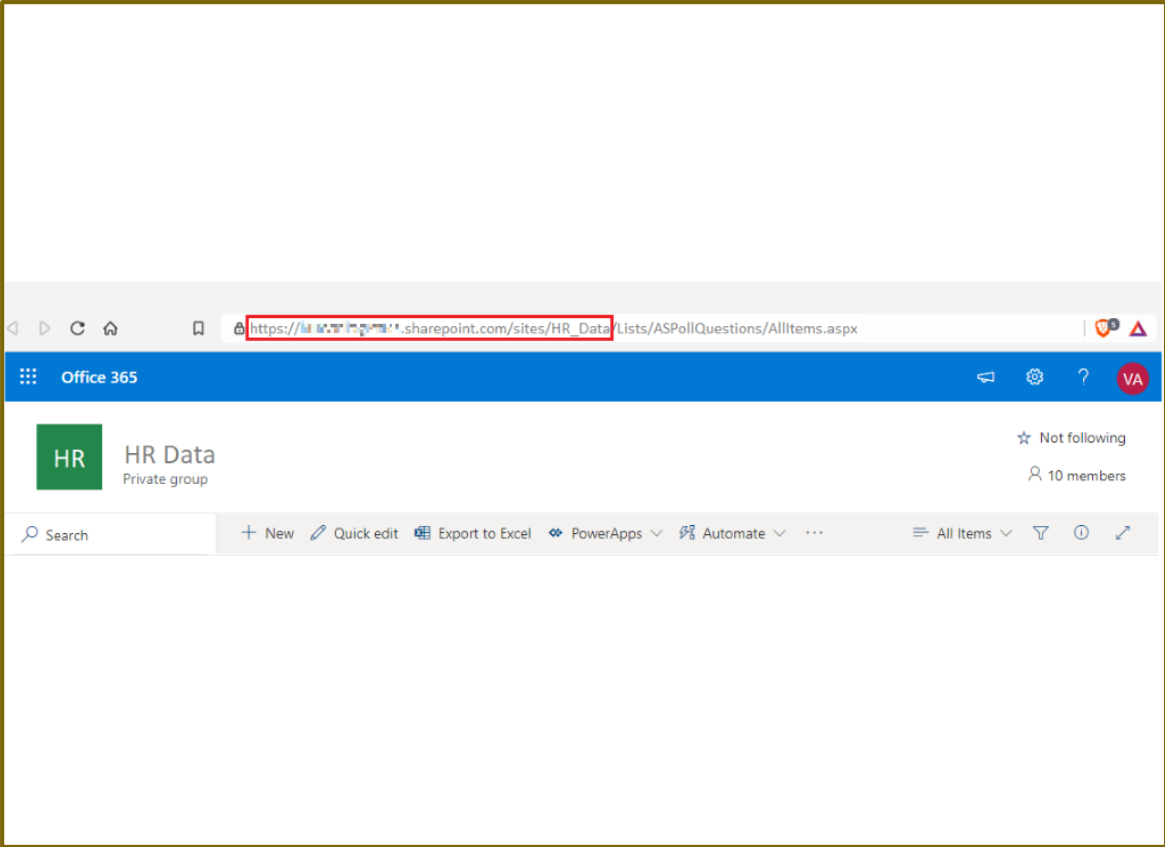


# Get data from NoSQL



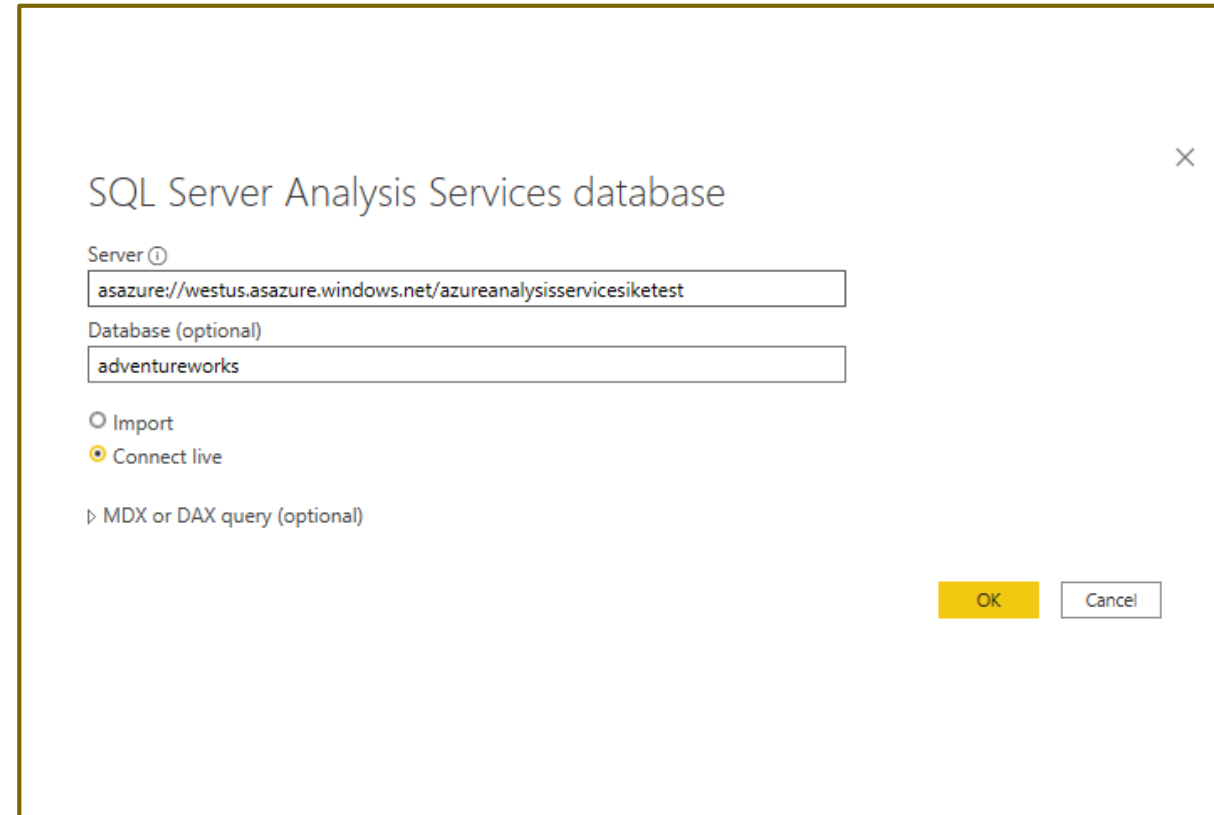


# Get data from applications



# Get data from Analysis Services

An analytical data engine that lets you digest data from multiple data sources and create calculations on the fly.



SQL Server Analysis Services database

Server

Database (optional)

Import

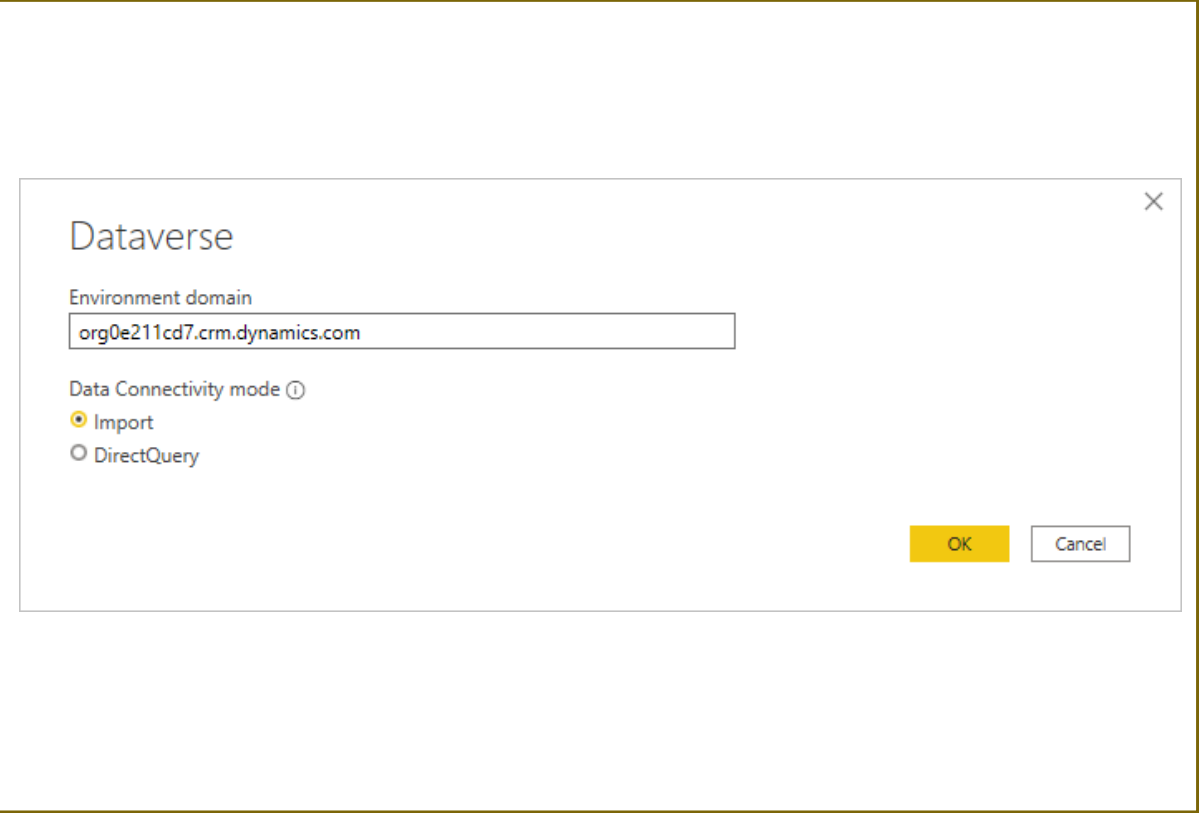
Connect live

▸ MDX or DAX query (optional)

OK Cancel

# Get data from Microsoft Dataverse

A cloud-based, low-code data service and app platform, which allows you to leverage the security and connectivity of Microsoft services. Dataverse was built for powerful, scalable solutions.



# Get data from a dataflow

**Add entities to start creating your dataflow**

**Define new entities**  
Choose a data source to define the entities for your dataflow. You can map your data to standard Common Data Model entities, or define custom entities instead.  
[Learn more](#)  
**Add new entities**

**Link entities from other dataflows**  
Linking to entities from other dataflows reduces duplication and helps maintain consistency across your organization.  
[Learn more](#)  
**Add linked entities**

**Power BI - Choose data source**

All categories: File, Database, Power BI, Azure, Online services, Other

- Acquire File
- Excel File
- FILE
- Table/CSV File
- XML File
- BI4 DES database Database
- Oracle database Database
- Power BI dataflows Power BI
- Azure SQL database Azure
- Azure SQL Data Warehouse Azure
- Microsoft Exchange Online Online services
- Azure Tables Azure
- Common Data Service for Apps Online services
- SharePoint Online list
- Salesforce objects
- Salesforce reports

© Copyright Microsoft Corporation. All rights reserved.

# Review Questions

**Q01 – Which query language do you use to extract data from Microsoft SQL Server?**

- A01 – T-SQL

**Q02 – You're creating a Power BI report with data from an Azure Analysis Services Cube. When the data refreshes in the cube, you would like to see it immediately in the Power BI report. How should you connect?**

- A02 – Connect Live

**Q03 – What can you do to improve performance when you are getting data in Power BI?**

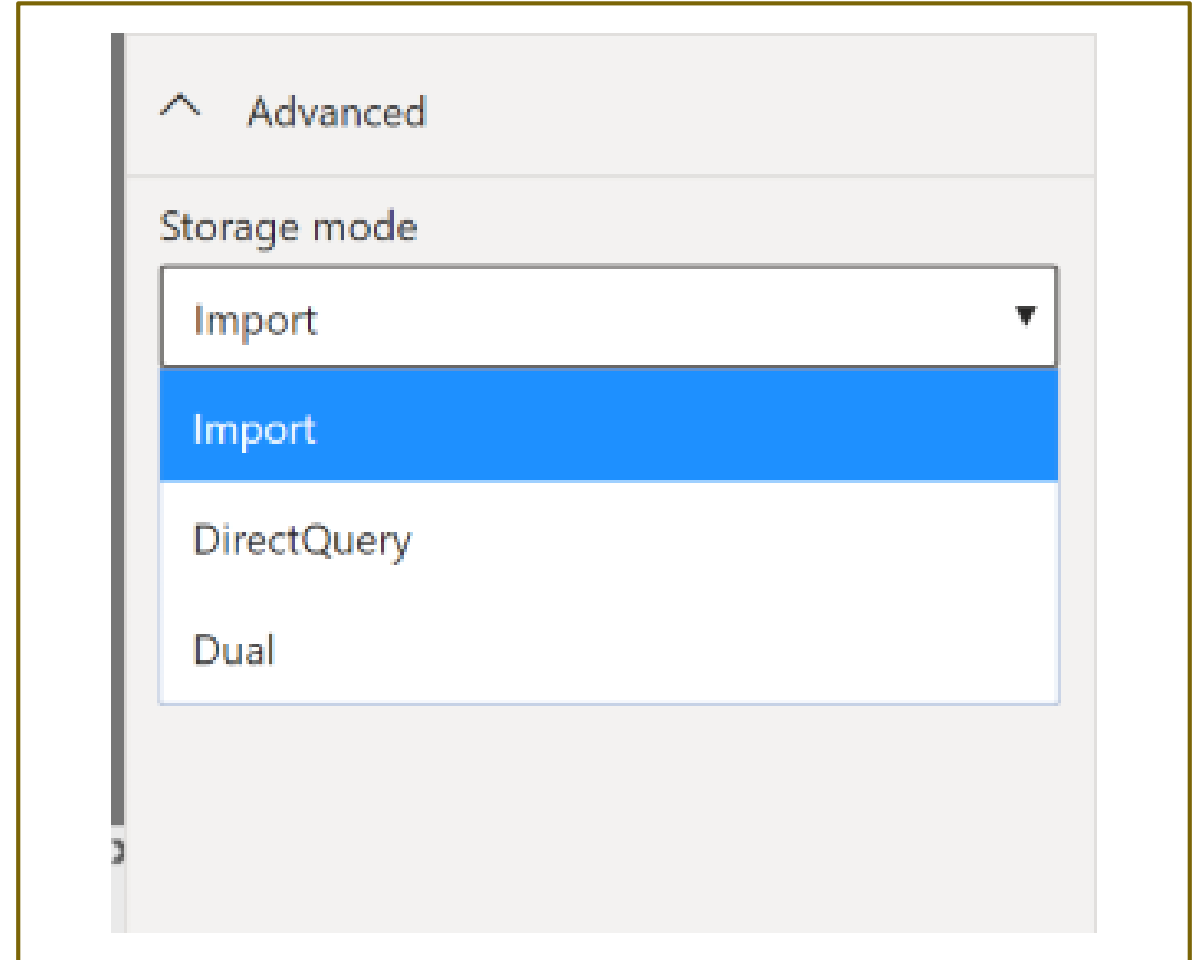
- A03 – Do some calculations in the original data source.

# Lesson 2: Optimize Performance

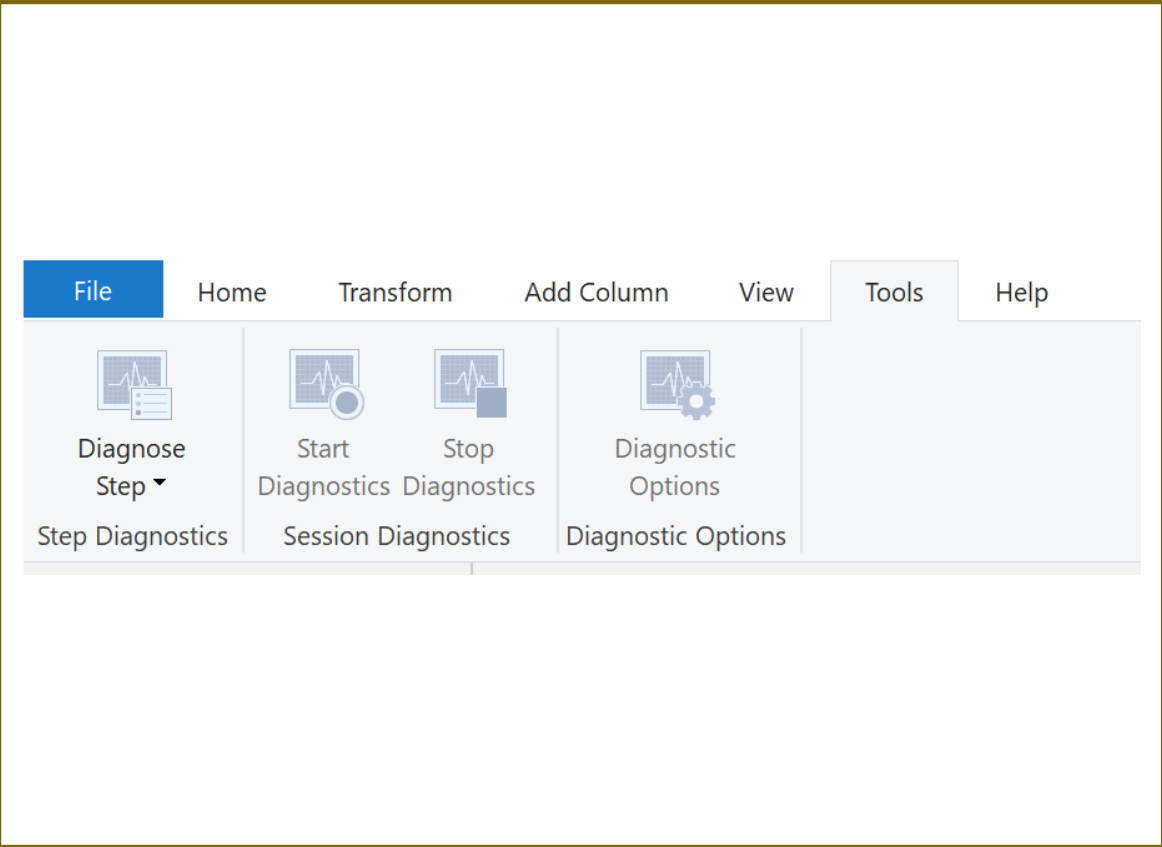


# Select a Storage Mode

- Specifies the storage mode of a table and lets Power BI determine how to cache data for reports.
- Set the storage mode for each table individually.



# Fix Performance Issues



The screenshot shows a table with the following columns: **Id**, **Query**, **Step**, **Exclusive Duration**, **Category**, and **Data Source Kind**. The table contains two rows of data:

Id	Query	Step	Exclusive Duration	Category	Data Source Kind
1.1	Product	Changed Type	0.00:00:00.0457545	Evaluator	null
1.2	Product	Source	0.00:00:01.9567741	Evaluator	null



# Optimize Query Performance



Performance in Power Query depends heavily on the performance at the data source.

---



Follow performance tuning guidelines of the source product.

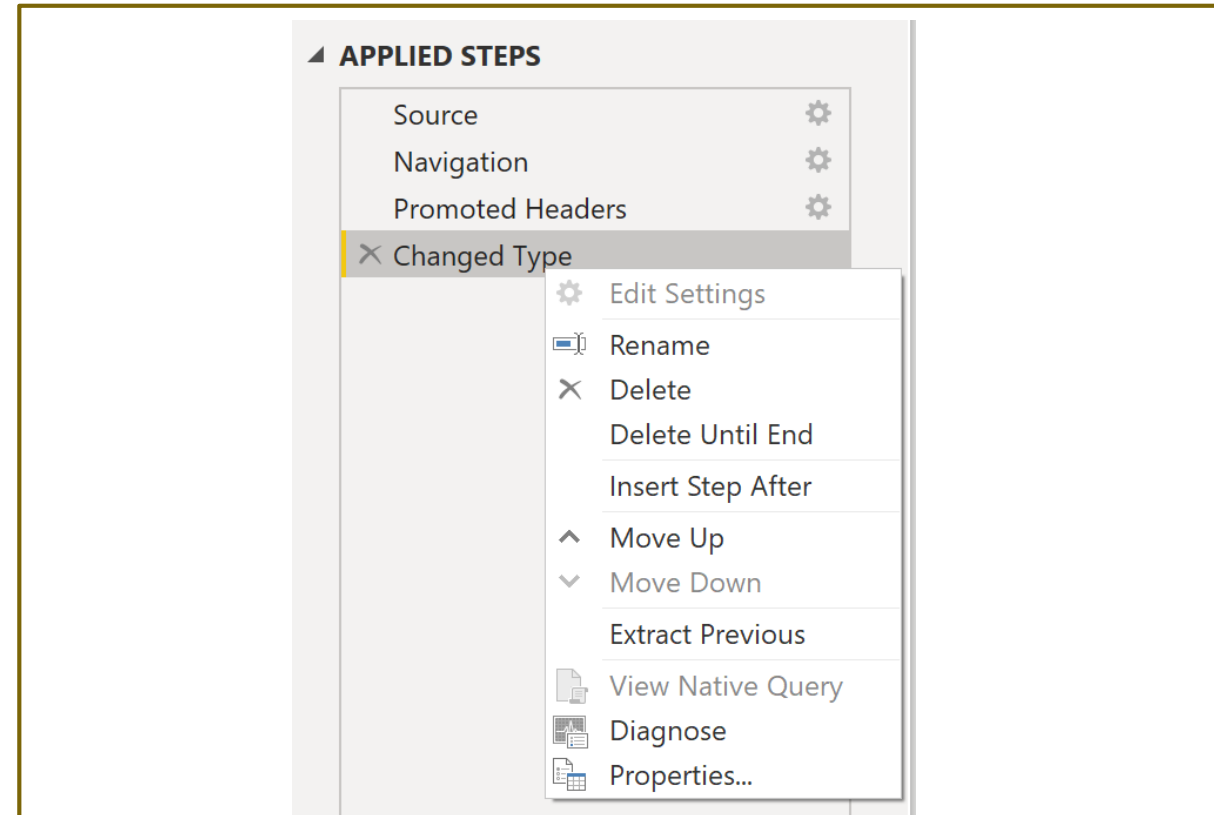
---



Some performance tuning can be done in Power BI.

# Query Folding

The process that lets Power Query generate a single query statement to retrieve and transform source data.



# Review Questions

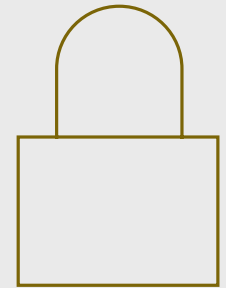
**Q01 – Which storage mode leaves the data at the data source?**

- A01 – DirectQuery

**Q02 – Which technology improves performance by generating a single query statement to retrieve and transform source data?**

- A02 – Query Folding

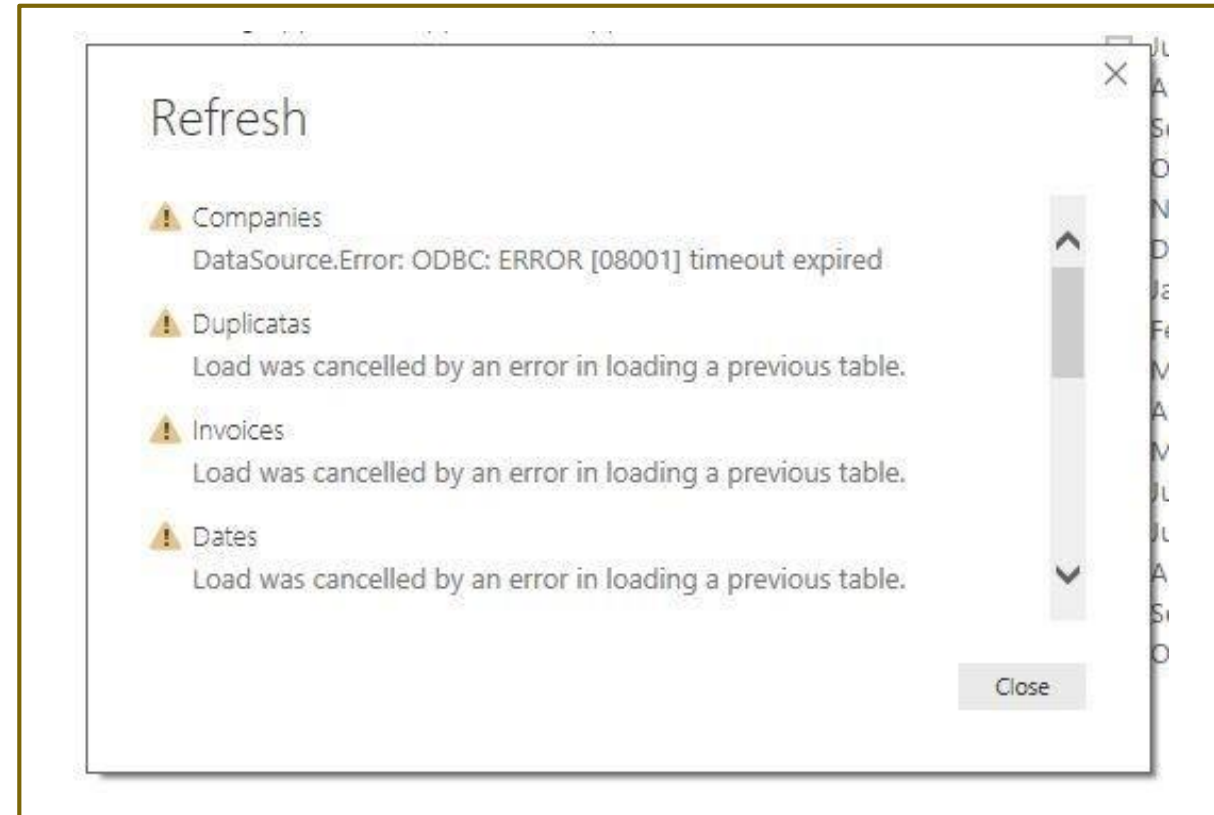
# Lesson 3: Resolve Data Errors



# Identify and Resolve Data Import Errors

## You may encounter the following errors:

- Query Timeout.
- Couldn't find data formatted as a table.
- Could not find file.
- Data type errors.



# Review Questions

Q01 – What type of import error might leave a column blank?

- A01 – Data type error

# Module Overview

## We covered the following concepts:

- Getting data from various data sources
- Optimizing Performance
- Resolving data errors

# Lab: Prepare Data in Power BI Desktop

Lab: Prepare Data in Power BI Desktop





# References

PL-300 Prepare data for analysis

<https://docs.microsoft.com/en-us/learn/modules/get-data/>

PL-300 Clean, Transform, and load data in Power BI

<https://docs.microsoft.com/en-us/learn/modules/clean-data-power-bi/>

