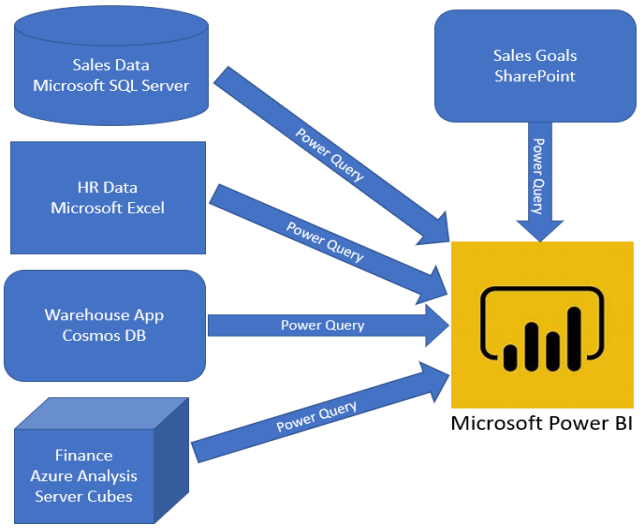
Get Data from Various Sources



Note: Before Starting Lab Make sure you have downloaded all Sample data Files in Folder D:\PowerBI-Learn

Excel Connector:

- Allows to import several worksheets and tables in same excel file.
- Supports xlsx, xlsm extension with less than 1GB.
- The worksheet should not be password protected.
- Use Format as Table option for better import (Required when you use PowerBI Service)

Lab 1: Get Data from Excel

1. PowerBI Desktop→Home→Data Group on ribbon→Get Data→Excel Workbook→
2. **Browse to following path:** D:\PowerBI-Learn\ClassLabs\GetData\CarSales.xlsx→Select Table1→Click on Load

Get Data Options:

Load: Automatically load your data into the Power BI model.

Transform Data: Button to launch the Power Query Editor. Here You can Perform required Transformations on Dataset then Load into PowerBI

Text/CSV:

- Columns are separated by identifier (,tab,;,:)
- Can also have fixed width format.
- (1252,65001) character set is used which supports majority of languages based on Latin alphabets
- Data Type detection is based on the first 200 rows by default.

Lab2: Get Data from Flat File

1. PowerBI Desktop→Home→Data Group on ribbon→Get Data→Text/CSV
2. **Browse to following path:** D:\PowerBI-Learn\ClassLabs\Getdata\Product.csv

Observe the Dialogue Box
It Identifies the delimiter, and allows us to select delimiter. You can select the Character set. (1252/65001) are most options, support majority of languages based on Latin alphabets. It automatically detects Datatype. The default option is to detect based on first 200 rows.

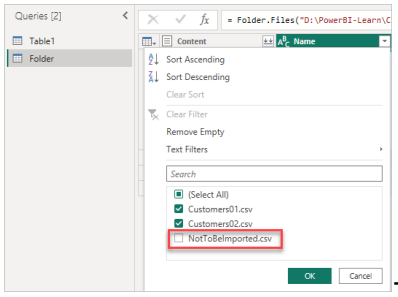
3. Click on Load, to Load Data into PowerBI

Folder Connector:

- Helps to import several files in folder.
- New files added to Folder in letter stages, are also reflected in Report during data refresh.
- Ideally should have same structure and format but PowerBI allows you to select which file should be used as sample.

Lab3: Get Data From Folder

1. Get Data→More→Select Folder→Connect
2. **Browse and Specify Folder path:** D:\PowerBI-Learn\ClassLabs\GetData\Folder
3. Click OK
4. Observe option to combine, Load or Transform Data
5. As you do not want to load one of the file click on transform
6. Power Query editor opens.
7. Click on Name Column drop down arrow →Uncheck NotToBeImported



→Ok

8. For Combining files→Click on content column→in Combine section →Click on Combine Files
9. Under Sample Files→Select Customer01.csv as it contains all columns→Click OK
10. To Load combined data in PowerBI :Click File Menu→Close and Apply

Lab 4: Get Data from Json

1. Get Data→More→Select json→Connect→Browse for Sales.json (D:\PowerBI-Learn\ClassLabs\GetData)

Lab5: Get Data from Web

1. Get Data→Web→
2. From Web Dialogue

URL:http://www.boxofficemojo.com/alltime/world/

3. Click ok
4. Select Table2→Load

Rename Tables in PowerBI

5. Data Pane→Table1→Right Click→Rename→CarSales
6. Similarly Rename Table2 as Movies200
7. Save the file as GetDataDemo.pbix

Note: If you want to Load more items from next page use Url:

https://www.boxofficemojo.com/chart/top_lifetime_gross/?area=XWW&offset=200

DataBase Connector and Performance

Understanding Connectivity Modes

- Defines how to retrieve data from DataSource .
- It has implication on performance, Data Freshness and PowerBI Capabilities
- **There are three Options:** Import,DirectQuery,Live Connection

Import

- Default,Popular,Microsoft recommended
- Supported by most Datasources
- Data Loaded in PowerBI(PBIX File).Reports are build on local data.It is Fasteset mode as Data Available Locally
- Gives Full PowerBI Experience
- Cons:
    - Data is fresh.based on Data Refresh
    - Data Refresh:Pro:8 times a day,Premium:42 times a day
    - Size Limit(Pro:1GB,Premium:100-400GB)
    - It is heavy on PowerBI Resources(Memory,Processor)

Direct Query:

- Dynamically Queries the Data at report time.Data resides in Data Source.Database Schema is imported.
- Data is Fresh.You can just use Refresh button on report to get updated data
- No Size Limitation on PowerBI Side
- Less PowerBI Resources are used as many data crunch operations are delegated to data source
- **Cons:**
    - For every visual refresh per session, data must be retrieved from source.
    - Limitation of 1 million row per query

- Not all database supports this option
- Limited PowerBI Experience (Several Transformation, Dax, Reporting capabilities not available)

Connect Live option:

- Data Queried dynamically and Schema is also not imported.
- Many PowerBI Capabilities are disabled(can't do transformation, define relationship or mix with other datasources)
- Data Modeling need to be one in source system, It will be more focused just on reporting
- Using the **Connect live** option helps you keep the data and DAX calculations in their original location, without having to import them all into Power BI.
- Azure Analysis Services can have a fast refresh schedule. Without the need to initiate a Power BI refresh schedule when data is refreshed in Analysis service, Power BI reports will immediately be updated.

Note:

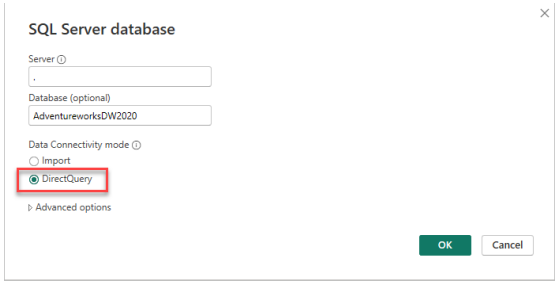
Live connection is available on Azure Analysis Services.

Azure Analysis Services is like the data modelling and storage technology in Power BI.

It uses the tabular model and DAX to build calculations, like Power BI. These models are compatible with one another.

Lab6: Get Data SQL Server:

1. Get Data→SQL Server→Enter Server Name and Database Name.
2. Choose connectivity Mode as DirectQuery→OK



Sign in options:

Windows - Use your Windows account (Azure Active Directory credentials).

Database - Use your database credentials.

Microsoft account - Use your Microsoft account credentials. This option is often used for Azure services.

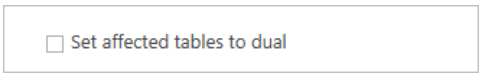
Advanced Options:

SQL statement: You can write an SQL query to import data.

Instead of writing complex SQL Statement in PowerBI it is recommended to create view in source and query view in Power BI.

2. If you get window for credentials, Enter your credentials
3. Select Following Tables From the list

FactInternetSales, DimProduct, DimCustomer, DimDate →Click on Load

4. Go to Table View→Select FactInternetSales →Observe that data can't be shown as it is Directquery.
5. To change the Query Mode
6. Go to Model View→Select FactInternetSales→Go to Properties Pane→Advanced→Observe Three options For Storage Mode→Change it to Import.
7. Observer the Dialogue for storage mode and **Uncheck** 

→Click OK
8. Select FactInternetSales in Table view and Data is Visible

Note: *To Load data based on filter condition you can use SQL Query.*
It is a best practice to avoid writing query directly in Power BI. Instead, consider writing a query like this in a view.
When PowerBI retrieves data using View, it participates in ***Query folding.***

Optimize Performance

Understanding Storage Modes:

Storage mode in Power BI tables is determining where the data of that table is stored, and how queries will be sent to the data source.

There are three storage modes:

- Import
- DirectQuery
- Dual

Import:

- Data is stored in the Power BI file, creating create a local Power BI copy of your datasets .
- It gets published along with the Power BI reports as a dataset.
- You can use all Power BI Service features including Q&A and Quick Insights.
- Data refreshes must be done manually
- The data of that table will be stored in the in-memory storage of the Power BI server (the machine that runs the Power BI engine), and every query to the data, will be a query to the in-memory structure, not to the data source.

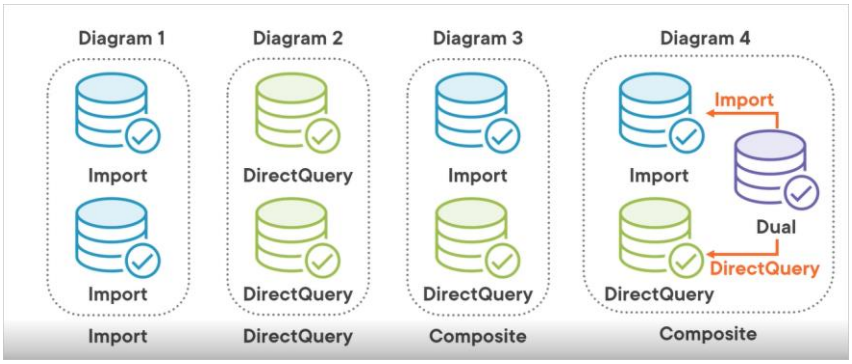
Direct Query:

- Tables with the DirectQuery storage mode will keep the data in the data source.
- It ensures you are always viewing the most recent version of the data.
- If we have a visualization from a table with this storage mode, Power BI will send a query to the data source directly and get the result back.

- You can use this when there is security requirements or datasets may too large and would take too long to load into PowerBI.

Dual:

- With Dual storage mode, one table can act either as DirectQuery or Import respective to the relationship to other tables.
- A table configured as Dual storage mode is both Import and DirectQuery
- Using the Dual mode allows Power BI to choose the most efficient form of data retrieval.



Access storage Modes:

Navigation:

Model View→Select Data Table→Properties→Advanced →Storage Mode:

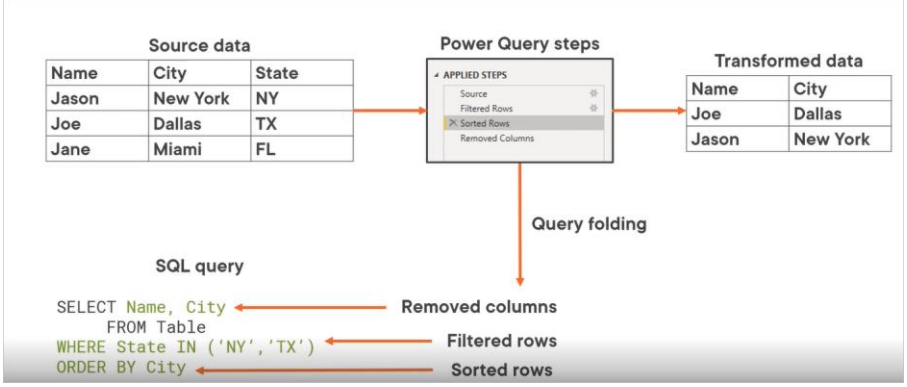
Note: You can change the Mode from Direct Query To Import but not the other way around.

Storage Mode Recommendation:

- Whenever possible use Import.
- For smaller and more static data tables (Dimensions) use Import
- For large and frequently updated tables (Fact Tables)) you can use Direct Query
- Use Dual for tables which are connected to Both types of tables (DirectQuery and Import)

What is Query Folding?

- **Query** is **set of all steps** performed to obtain, transform and Load Data.
- Steps can be filter or order rows etc.
- ***Query folding*** is the process by which the transformations and edits that you make in Power Query Editor are simultaneously tracked as native queries, or simple **Select** SQL statements, while you are actively making transformations.
- When you load your data, the transformations take place independently in the original source, this ensures that performance is optimized in Power BI.



Benefits:

- Transformation will be executed much quicker.
- More efficiency in data refreshes as Power BI does not have to run through each transformation locally.

Note: Power BI Can Use Query Folding on Many Data sources(relational DB, SharePoint lists etc).It cannot be used for source like Files, blobs etc.

Query Folding Best Practices:

- Some transformations like merge, append, custom M script cannot use Query Folding
- Try to move these operations to source system by creating SQL View
- If view can't be created, move this transformation that cannot use query folding as last step in query.

Data Source Settings

You can change information about your Data Sources.

Example: Changing File Path or Database Name etc

Change Data Source Setting from PowerBI Main Interface:

Navigation:

Home→Transform Data→Data Source Settings

OR

File→Options and Settings→Data Source settings

Change Data Source Settings From PowerQuery:

Navigation:

Select Table→Home→Data Sources(Data Source Settings)

Data Source settings Options:

1. Change Source: You can make changes to connection information like file path or in case of database server name ,database name etc.
2. Export PBIDS: Connection information can be saved as **pbids** file which is in JSON format.

Contains one connection per file and can't store credentials.

3. Edit Permissions: You can change permission information.
4. Clear Permissions: Can act as security measure when you need to send the file to someone else.

Using Parameters:

- You can store and manage reusable values which can dynamically change queries.
- Using parameters you can change between environments (Developement, Test,Production)
- Entire report can be changed based on parameters.

Lab7: Create Parameters for Folder Name, Database Name

1. Home Ribbon→Click on Transform Data→It opens Power Query Editor
2. Power Query Editor→Home Ribbon→Manage Parameters→Manage Parameter
3. Create Parameter for Folder Location

   Name: pFolderLocation

   Type:Text

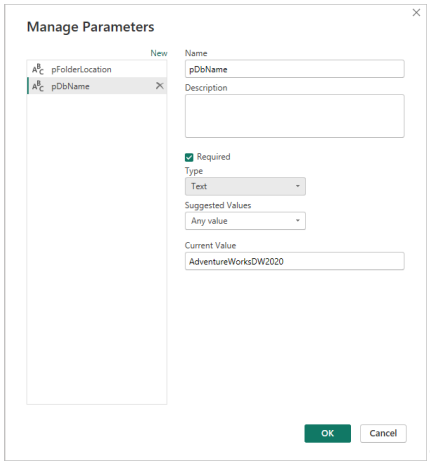
   Current Value: D:\PowerBI-Learn

   Under Manage Parameter→Click on New→

   Name:pDbname

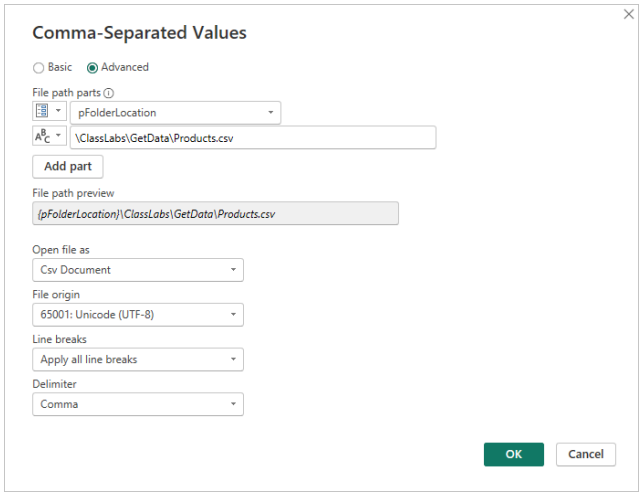
   Type: Text

   Current Value: AdventureWorksDW2020



→Click Ok

4. Use the pFolderLocation for Products Query
5. Click on Products.
6. Applied steps→Source→Click on gear 
7. In Dialogue box select Advanced and change settings as shown:



8. DimCustmer Query→ Applied steps→Source→Click on gear 
9. use pDBName parameter for DimCustomer Table Source