Module 2

Power Query Shaping and Combining Data

Module Overview

- Power BI Desktop queries
- Using databases as a data source
- Shaping data
- Combining data

Lesson 1: Power BI Desktop queries

- The Power Query Editor
- Queries
- APPLIED STEPS
- The Advanced Editor

Queries

- Commands you run against the data source to specify the data to extract:
  - Return entire tables or run a query against the source
  - Use stored procedures against SQL Server databases
  - Only return the data that you need
- Expressions used to transform data:
  - M Query Language:
    - Use in Power Query Editor
    - Generate using menu options or edit query directly
  - DAX:
    - Use in Power BI Desktop
    - Derived from MDX and Excel formulas
    - Straightforward to use but very powerful

The Power Query Editor

- Enables you to load data and apply transformations
- Ribbon comprises of four tabs:
  - **Home**: import data, hide or delete columns, reduce rows, merge and append queries
  - **Transform**: create aggregated columns, transpose, pivot, unpivot, split values
  - **Add Column**: add columns, add indexes, apply functions
  - **View**: show or hide the QUERY SETTINGS pane

APPLIED STEPS

- The Query Editor records all transformations to a query in the APPLIED STEPS setting:
  - All transformation steps are listed in order of creation; Source is first, followed by Navigation if applicable
  - Source contains data source connection information, and Navigation includes select tables and views
  - Can reorder steps if no dependencies exist
  - Can delete steps, but be aware of dependencies
  - Can undo steps, rolling back a previous step
  - Can rename steps

The Advanced Editor

- With the Advanced Editor, you can see the query that Power BI runs against the data source to import the data:
  - Query is written in M Power Query Formula Language
  - To view, on the Home tab, click Advanced Editor
  - The query includes the connection, and connection type; for example, Excel or SQL Database
  - All transformations you apply to your data using Advanced Editor are added to the query code
  - The list of steps are reflected in the query, and in the same order
  - You can edit the query, but use syntax checker

Lesson 2: Using databases as a data source

- SQL Server
- Other data sources
- Query Folding

SQL Server

- SQL Server is a relational database management system (RDBMS):
  - Unlike Access, SQL Server can handle multiple users and transactions
  - Scalable from smallest to largest size databases
  - Cloud and on-premises versions
- Connect from Power BI Desktop:
  - Connect using **Get data**; enter the name of the server instance and optionally the name of database
  - Use a query or select tables and views
  - Load into Power Query Editor or straight into data model

Other data sources

- Connect to a wide range of data sources from Power BI Desktop:
  - More data source connections than the Power BI service
  - Combine data from multiple SaaS providers into one report or dashboard
  - SaaS providers include Bing, Google Analytics, Facebook, Salesforce, Marketo, GitHub, Microsoft Dynamics, and Exchange
  - Supports industry database providers such as Access, Oracle, IBM DB2, MySQL, Sybase, and Teradata
  - Connect to any webpage to scrape structured data
  - Copy and paste from an Excel or text file to create a new table in the dataset

Query Folding Data Sources

- Relational sources
  - SQL Server, Oracle, …
- OData sources
  - Example: SharePoint list and others
- Active Directory
- Exchange
- HDFS,
  - Folder.Files
  - Folder.Contents (for basic operations on paths)

Foldable Transformations

- Must be immediately following Navigation step
  - Add new columns with simple expressions
  - Filtering (on rows or columns)
  - Joins
  - Aggregates and GROUP BY
  - Pivot and unpivot
  - Numeric calculations
  - Simple transformations, such as UPPER, TRIM
- Use View Native Query to see Query Folding
  - Option in Applied Step context menu

Demonstration and Exercise 1

You will see how to:

- Connect to a SQL Server data source
- Eliminate unneeded columns
- Rename columns
- Add new columns with formulas

Lesson 3: Shaping data

- What is shaping data?
- Formatting data
- Transforming data

What is shaping data?

- Shaping data is the process of transforming and formatting data for best presentation in reports:
  - The original data in the source remains unchanged
  - Each shaping step is recorded in the APPLIED STEPS list
- When shaping data:
  - Remove columns and rows that are not needed
  - Rename columns using an obvious naming convention
  - Ensure columns have the correct data types
  - Use date and time functions to create new columns
  - Add columns, and indexes useful for appending data
  - Apply a sort order, or use an index to guarantee order

Formatting data

- Power Query Editor provides many options for creating columns, formatting text, and numbers:
  - General Group:
    - Add custom columns using formulas or duplicate columns
    - Add an index column and move to the front of the table
  - From Text
    - String functions include lowercase, UPPERCASE, Capitalize Each Word, Trim, Clean, Add Prefix, and Add Suffix
    - Merge columns using optional character or space separator
  - From Numbers
    - Add, Multiply, Subtract, Divide columns, or calculate by value
- All formatting uses a query that you can view in the Formula Bar or in Advanced Editor

Transforming data

- Table group:
  - Use Group By to apply aggregations on your table
  - Use First Rows As Headers and use Headers As First Row
  - Transpose to treat columns as rows, and rows as columns
  - Reverse Rows to reverse the order of the data
- Any Column group:
  - Change or detect data types
  - Replace Values and Replace Errors
  - Fill null values in a column
  - Pivot Column and Unpivot Columns
  - Move columns
- Text Column group:
  - Split single column in multiple columns

Demonstration and Exercise 2 and 3

You will see how to:

- Import data from Excel
- Unpivot data into a table
- Apply transformations to the tables

Lesson 4: Combining data

- Adding data from the internet
- Shaping the new data
- Merging data

Adding data from the internet

- Import data from a website that provides data in a tabular structure:
  - Use publicly available datasets, and combine this with your existing data for reporting insights
  - Import using **Get Data**, **Web**, and enter the URL
  - Power BI establishes a connection, and imports the data
  - Use the data just as you would from any other source
  - Preview the table structures that Power BI has detected
  - Load data, or edit in Power Query Editor; data can be refreshed
  - Shape and transform the data as required
- Be aware that the source data could be removed

Shaping the new data

- After importing data from the internet, use shaping and transforming to format and correct
  - All shaping is stored as steps, so will be reapplied each time the query is run, and data can be refreshed
  - Use the data as you would from any other data source
  - Remove columns that you won't use in reporting
  - Ensure the query and columns have names that reflect the content, and are obvious to users and Q&A
  - Make sure columns have the correct data type
  - Apply a sort order if required

Merging data

- Merge columns:
  - Merge one table into another table, using a joining column
  - Choose from join types
  - All columns are initially merged, but use the selector to choose which columns you want to keep
  - Can retain original column names
- Append rows:
  - Adds rows from one or more tables to another table
  - Column data does not have to match
  - Mismatching can result in unclean data and/or nulls
  - Add index to combined table

Demonstration and Exercise 4

You will see how to:

- Import data from the Internet
- Merge table into a single table view