

Module 3

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
- Demonstration: Using APPLIED STEPS

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

Demonstration: Using APPLIED STEPS

In this demonstration, you will see how to:

- Add transformations to a query, and see the steps in APPLIED STEPS
- Rename steps in the APPLIED STEPS list
- See the steps reflected in Advanced Editor
- Delete steps, and change the source table in the Navigation step

Lesson 2: Using databases as a data source

- SQL Server
- Other data sources
- Query Folding
- R script data connector
- Demonstration: Importing data from SQL Server

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

R script data connector

- Run R scripts from Power BI Desktop:
 - Import results of R script into datasets to create reports. Publish to the Power BI service to use in dashboards
 - Must install Microsoft R Open prior to running scripts
 - Write R scripts in local environment, and test to ensure they run successfully before using in Power BI Desktop
 - Limitations include:
 - Only data frames are imported, so include all required data
 - Time-out period is limited to 30 minutes
 - N/A values are converted to NULL values
 - Complex and Vector type columns not imported, error in table
 - Working directory of R script must be full path, and not relative
- Manage your R installations in **Options** dialog box

Demonstration: Importing data from SQL Server

In this demonstration, you will see how to:

- Connect to SQL Server from Power BI Desktop
- Import data into the Power Query Editor

Lesson 3: Shaping data

- What is shaping data?
- Formatting data
- Transforming data
- Demonstration: Transforming data with the Power Query Editor

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: Transforming data with the Power Query Editor

In this demonstration, you will see how to:

- Import data from Excel
- Apply transformations to the table

Lesson 4: Combining data

- Adding data from the internet
- Shaping the new data
- Merging data
- Demonstration: Adding and shaping data from the internet

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 unclear data and/or nulls
 - Add index to combined table

Demonstration: Adding and shaping data from the internet

In this demonstration, you will see how to:

- Import data from the internet
- Shape the data that is imported

Lab: Shaping and combining data

- Exercise 1: Shape Power BI data
- Exercise 2: Combine Power BI data

Logon Information

Virtual machine: **20778C-MIA-SQL**

User name: **ADVENTUREWORKS\Student**

Password: **Pa55w.rd**

Estimated Time: 60 minutes

Lab Scenario

Adventure Works employees are becoming increasingly frustrated by the time it takes to implement managed BI services. The existing managed BI infrastructure, including a data warehouse, enterprise data models, and reports and dashboards, are valued sources of decision-making information. However, users increasingly want to explore relationships with other, currently unmanaged data. It takes too long for the IT department to include these requirements in the corporate BI solution.

Lab Scenario (Continued)

As a BI professional, you have been asked to explore ways in which Adventure Works can empower business users to augment their managed enterprise BI solution with self-service BI.

Lab Review

- Discuss the types of different data in your organization that could be combined using the Power Query Editor. Do you have data stored across locations that could be appended, or lookup data that could be merged into other tables to make it more useful for reporting?

Module Review and Takeaways

- Review Question(s)