Microsoft Power BI

Power BI is widely used for data analysis, reporting, and data visualisation, making it an essential tool for businesses looking to harness the power of their data for better decision making.

It allows users to connect to various data sources, transform and visualise data, and create interactive reports and dashboards.

Users need to be able to access their data in Power BI to achieve this. There are several methods for importing data into Power BI.
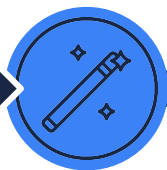
Database connection workflow

Connecting to a database in Power BI involves a series of steps that ensure data is extracted, transformed, and loaded into the data model.

Data source connection

Select the appropriate data **source connector**, provide the **connection details**, and ensure the necessary permissions.

Data transformation

Clean the data by removing duplicates, filtering out irrelevant data, and addressing data quality **issues.** Define relationships between tables, **create calculated columns**, and **establish hierarchies** if necessary.
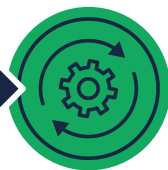
Data import/DirectQuery

Decide whether to **import** the data into Power BI or use **DirectQuery**. Loading data is suitable for **small to medium-sized datasets**, while DirectQuery allows **large datasets** to remain in the source.
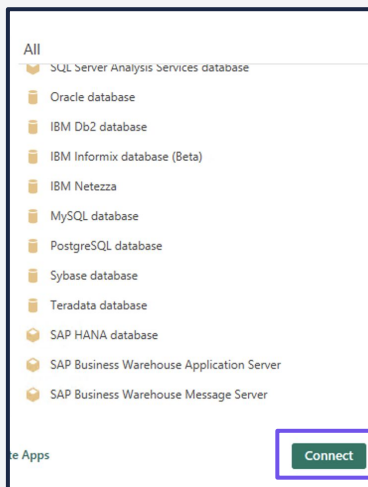
Report and visualisation

Use **Power BI Desktop** to generate **reports** and **visualisations**, using the designed data model. **Maintain synchronisation** with the source database periodically.
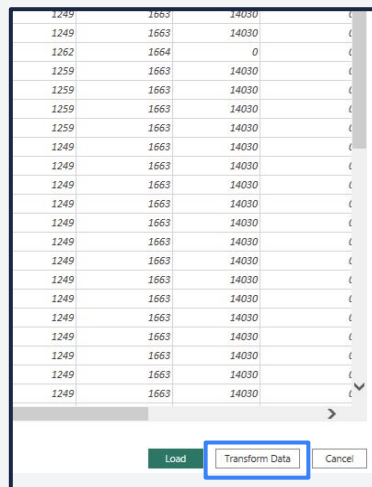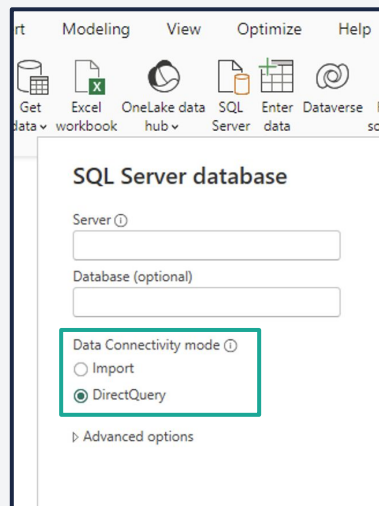
Database connection workflow
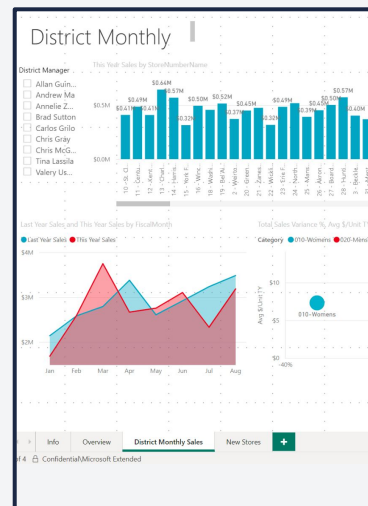

Data source connection

Data transformation

Data import/DirectQuery

Report and visualisation

Importing and connecting to data

One of the most essential processes when working with Power BI is the ability to **ingest data quickly, reliably, and frequently**.

Before digging into how to import data into Power BI, it's important to understand the **difference between importing and connecting data** into Power BI.

| **Importing data**	Involves **copying data from the source** into the Power BI file. The data is stored within the Power BI model, and we can create transformations and calculated columns within Power BI. Importing is suitable for **smaller datasets** or when we need **to combine data from multiple sources.**

Connecting data	DirectQuery is a method where **data remains with the source** (no data is imported into Power BI), and **Power BI connects to it in real-time**. This approach is beneficial for larger datasets, frequently updated data, or when data should not be duplicated within Power BI.

Importing data: Import mode

Power BI supports a wide range of formats, including flat files such as **Excel, CSV**, and **text files**.

Power BI offers connectors to various relational databases like **SQL Server** and **MySQL** as well as **NoSQL** databases. We can also use SQL queries to transform data before importing it.

Since the data is **copied and stored within Power BI**, the data will only be updated when we **manually refresh** the dataset or set up a **scheduled refresh**.
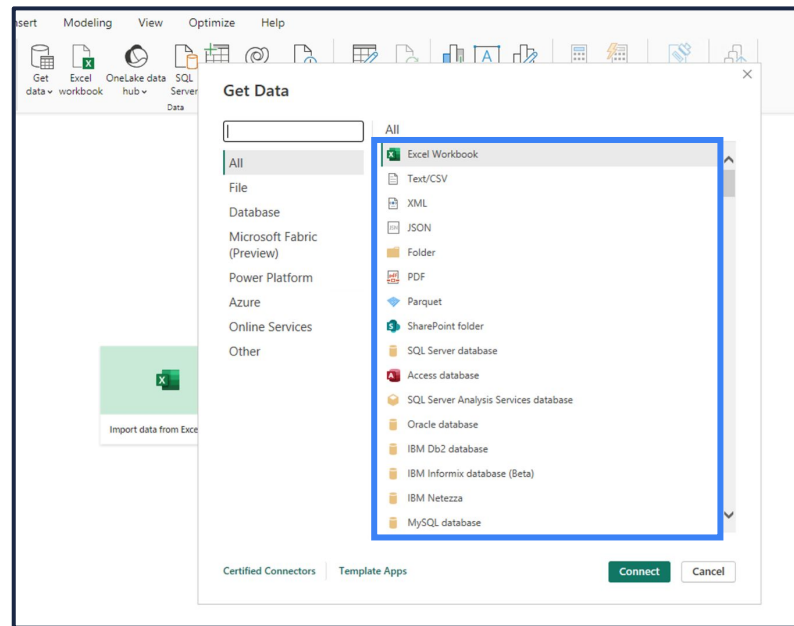
When using Import mode, Power BI will cache the data, providing a snapshot of the data at a specific point in time. All of the interactions and filters will be applied to this compressed cache source rather than the original data source.

Connecting to data: DirectQuery mode

Power BI offers connectors to various online services and applications like **SharePoint**, **Google Analytics**, **Salesforce**, as well as most **relational** and **NoSQL** databases.

DirectQuery mode involves establishing a **live or near-live link** between Power BI and an external data source.

Connecting to applications and services in Power BI typically requires **specific connection details** related to the chosen service, such as **API keys**, **URLs**, or authentication methods.

These connections may require **periodic refreshes**. Power BI allows us to set up **automatic data refresh schedules** to keep reports and dashboards up-to-date.

SQL Server database

Server ⓘ

Database (optional)

Data Connectivity mode ⓘ
○ Import
● DirectQuery

▷ Advanced options

When using DirectQuery mode, the dashboard will query the data source immediately during runtime. Each filter and interaction with the report will result in a series of new queries to the database.

Managing data storage in Power BI

Managing data storage in Power BI is a critical aspect of **optimising report performance** and ensuring **data validity and accessibility**.

Import mode

Advantages

Suitable for relatively **small to medium-sized** datasets that **don't change** frequently.

High performance, **enhanced modeling**, and **offline access** to data, allowing **quick querying** and visualisation, as well as the ability to create calculated columns and hierarchies even **without a live connection** to data.

DirectQuery mode

Advantages

Suitable for **large datasets** or scenarios where **data is frequently changing**, or we **require real-time reporting**.

Doesn't have a **data storage limit**, as it doesn't consume Power BI's memory.

Dual Mode

Advantages

Dual mode allows a **combination** of **real-time** data with **in-memory** benefits, allowing us to create **composite models** that combine data from multiple sources and connections.

Complex transformations and calculations can be done in Import mode while **real-time data** remains in DirectQuery mode.

Managing data storage in Power BI

The choice of which mode to use in Power BI depends on **specific requirements, data volume, update frequency, and available resources**.

Import mode

Disadvantages

Very large datasets can **consume significant memory** therefore **slowing down performance**.

DirectQuery mode

Disadvantages

Modeling capabilities are limited since some complex calculated columns, measures, or hierarchies cannot be built in Power BI and the **query performance** depends on the underlying data source and network speed.

Dual Mode

Disadvantages

Managing a dual model can be **more complex** and a **compromise** between **performance** and **functionality** must be established.

Data import issues and resolutions

Data import issues can occur due to various factors, including **regional settings, data quality, and data volume**. Common data import issues we can expect include:

Regional settings

Issue: Data may not display correctly due to regional date and number formatting differences between the data source and Power BI.

Resolution: Ensure that Power BI's regional settings match the data source's locale. This includes date formats, decimal points, and currency symbols.

Data type mismatches

Issue: Data import errors may occur when data types in the source don't match the data model.

Resolution: Use data type conversion functions to match data types, and handle errors or missing values appropriately.

Data quality and cleanup

Issue: Inconsistent data quality can lead to import errors and incorrect visualisations.

Resolution: Use data transformation and cleaning techniques to address data quality issues, such as handling missing data or correcting anomalies.

Data volume

Issue: Large datasets and complex data models with many tables can lead to slow import times and report performance.

Resolution: Optimise data sources by reducing unnecessary columns and rows and simplify the data model by reducing the number of tables and relationships.