Microsoft

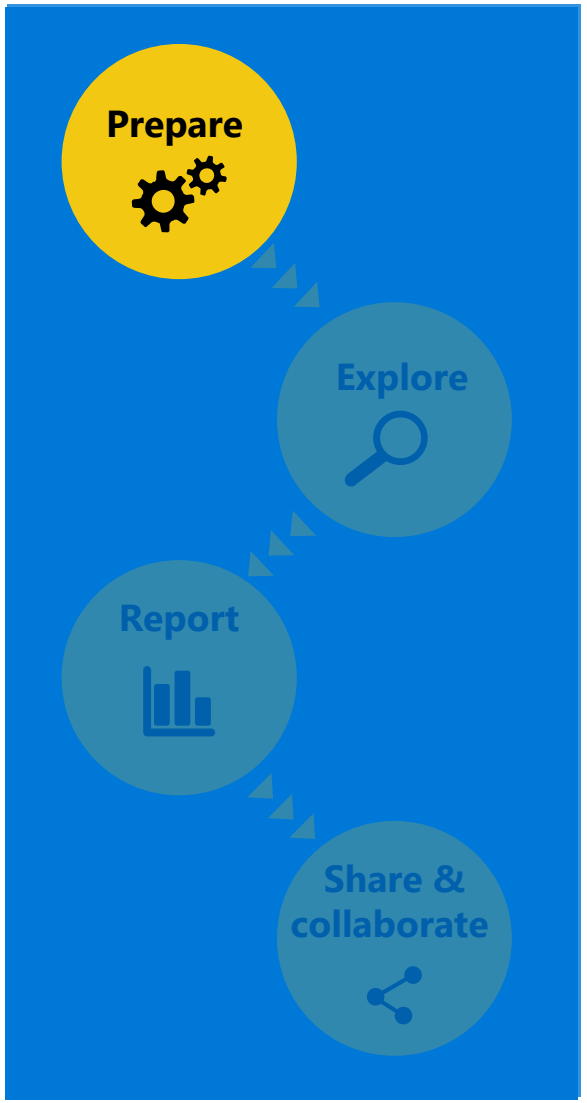
Business Analytics with Power BI

Microsoft Services

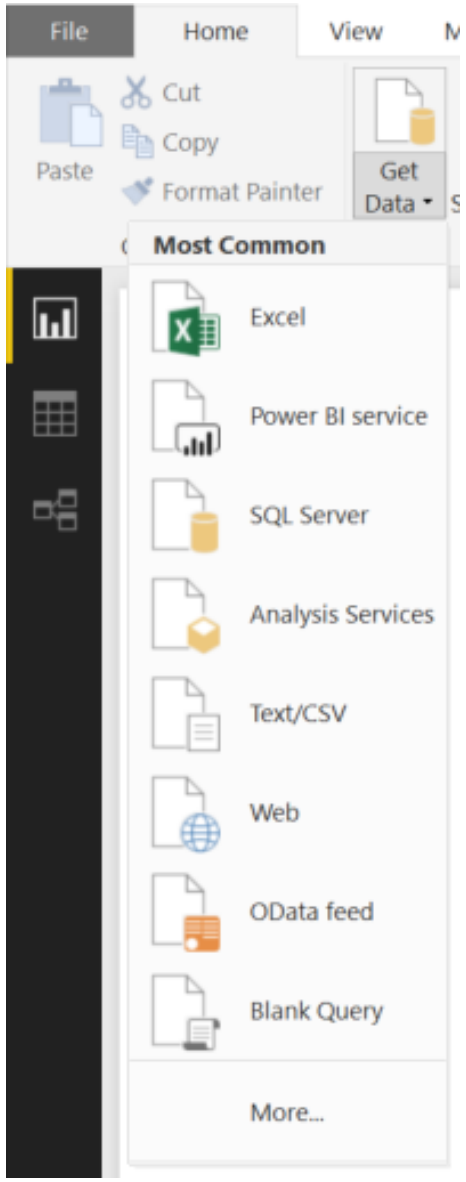
Module 1: Power BI Desktop

Lesson 2: Connecting to Data

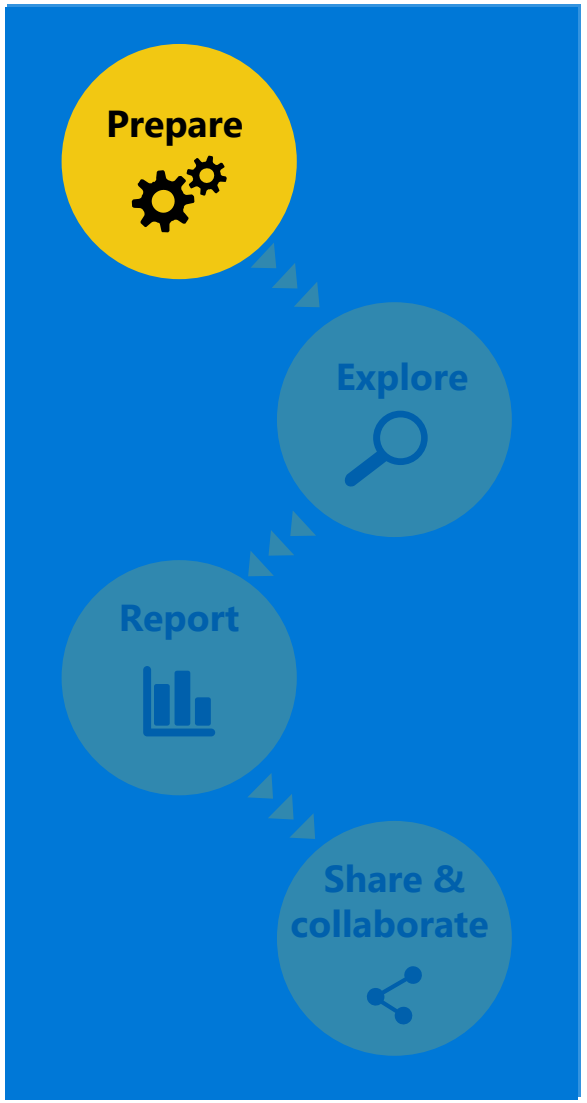
Connecting Data - Prepare



- The "**Get Data**" option allows the definition of **connections** to data sources and the **selection of entities and columns**
- **Authentication** options can be specified including **credential types** such as Windows or others
- Data can be **imported in-memory** or be **accessed live** for some data sources
- **Each selected entity will become a query** that can be **further enhanced** to meet the business requirements



Connecting Data - Prepare

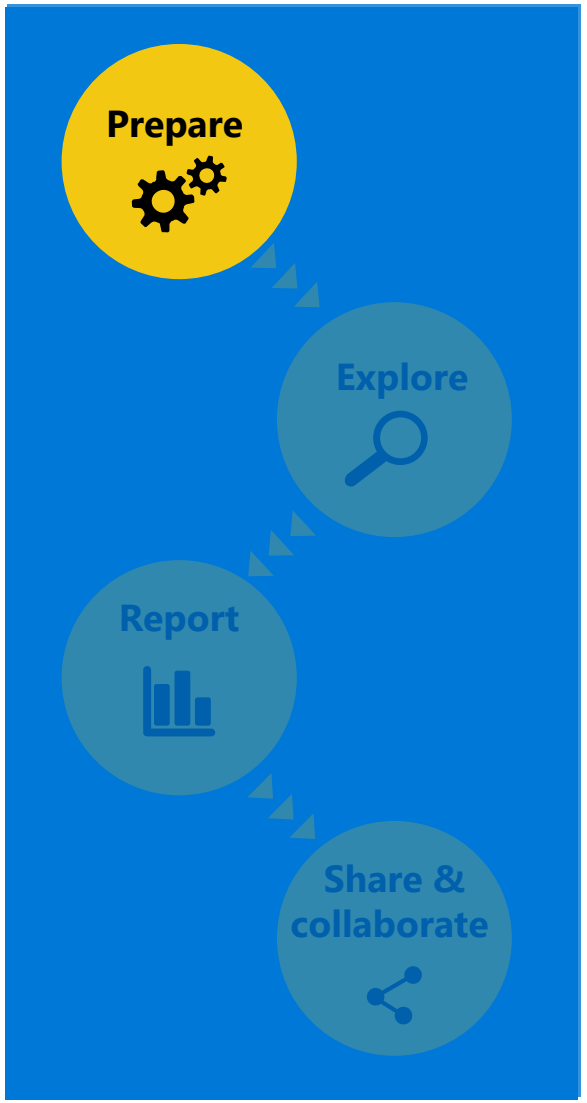


Data sources available through Power BI Desktop (not all listed)

File	Database	Azure	Other/Online Services		Other
• Excel • Text/CSV • XML • JSON • Folder • Sharepoint Folder	• SQL Server **(DQ)** • Access • SQL Server Analysis Services **(LQ)** • Oracle **(DQ)** • IBM DB2 • IBM Informix (Beta) • IBM Netezza **(DQ)** • MySQL • PostgreSQL • Sybase • Teradata **(DQ)** • SAP HANA **(DQ)** • SAP BW **(DQ)** • Amazon Redshift **(DQ)** • Impala (Beta) **(DQ)** • Google BigQuery (Beta) **(DQ)** • Snowflake **(DQ)**	• SQL Database **(DQ)** • SQL Data Warehouse **(DQ)** • Analysis Services (Beta) **(LQ)** • Blob Storage • Table Storage • Cosmos DB (Beta) • Data Lake Store • HDInsight (HDFS) • HDInsight Spark **(DQ)**	• Power BI Service • SharePoint Online List • Exchange Online • Dynamics 365 (online) • Dynamics 356 for Financials (Beta) • Common Data Service (Beta) • Azure Consumption Insights (Beta) • Visual Studio Team Services (Beta) • Salesforce Objects • Salesforce Reports • Google Analytics • appFigures (Beta) • comScore Digital Analytix (Beta) • Dynamics 365 for Customer Insights (Beta)	• Facebook • GitHub (Beta) • Kusto (Beta) • MailChimp (Beta) • Marketo (Beta) • Mixpanel (Beta) • Planview Enterprise (Beta) • Projectplace (Beta) • QuickBooks Online (Beta) • Smartsheet • SparkPost (Beta) • SQL Sentry (Beta) • Stripe (Beta) • SweetIQ (Beta) • Troux (Beta) • Twilio (Beta) • tyGraph (Beta) • Webtrends (Beta) • ZenDesk (Beta)	• Vertica (Beta) • Web • SharePoint List • OData Feed • Active Directory • Microsoft Exchange • Hadoop File (HDFS) • Spark (Beta) **(DQ)** • R Script • ODBC • OLE DB • Blank Query

- The available data sources are **constantly evolving**
- Some of them support **Direct Query (DQ)**. No data is imported but **cannot mix import modes or data sources**
- There are also some **limitations on the transformations** that can be done **with DQ**

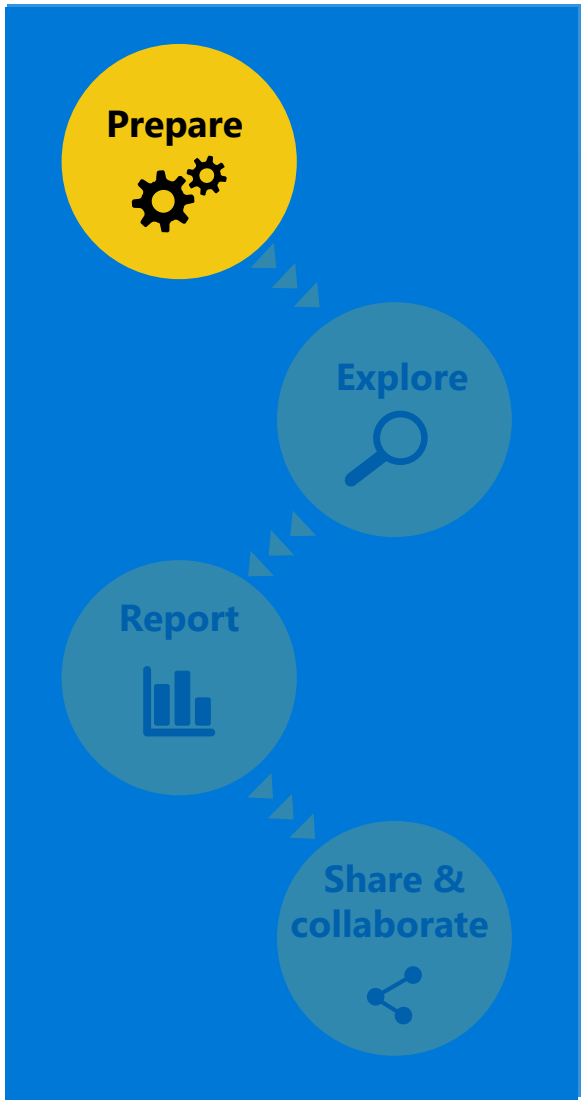
Connecting Data - Prepare



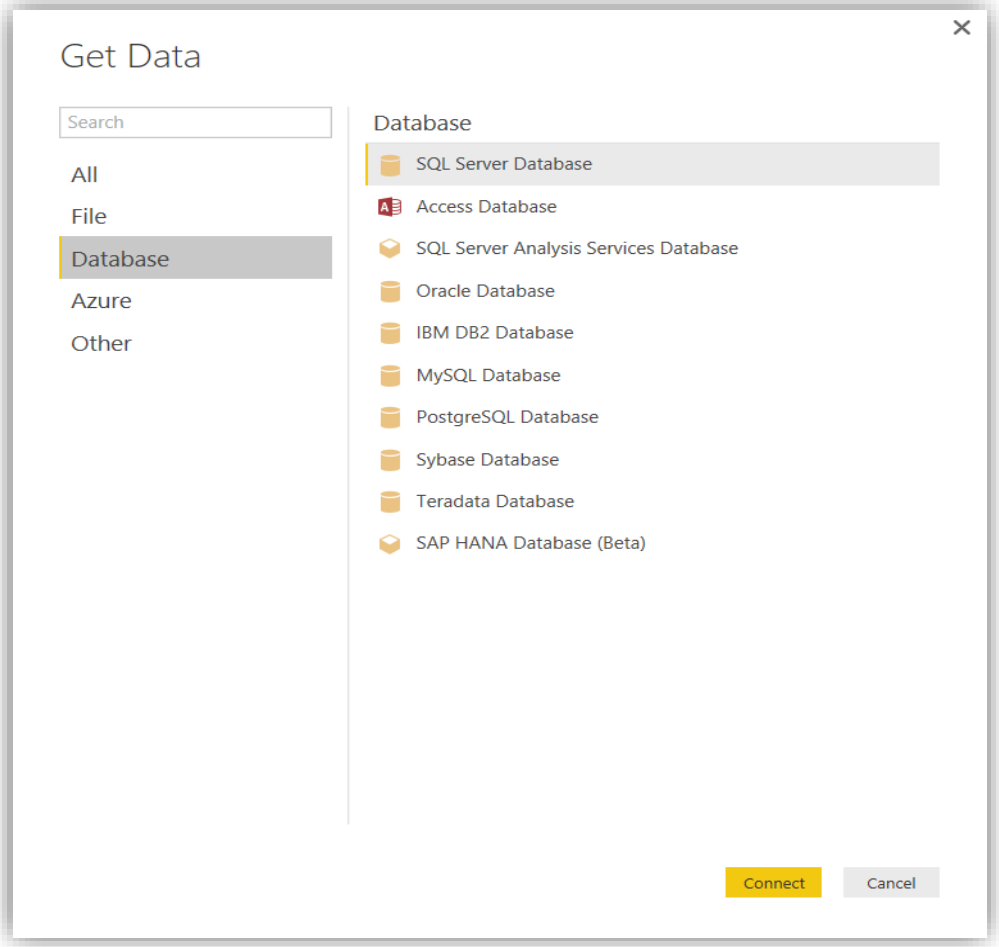
Microsoft Azure Data Catalog (integration under development)

- Azure Data Catalog represents the **evolution of the Data Catalog** that was initially available for **Power BI for Office 365**
- The "**Get Data**" experience should be enhanced with **data discovery** when this service becomes generally available
- **Data sources are published to the catalog** and are **further indexed and described**, so that business users can find useful information with minimum effort
- These **data sources** can be hosted in the **cloud** or **on-premises** and can be from **different providers**
- The **business users** themselves will be **responsible for enriching** the system (crowdsourcing model)

Connecting Data - Prepare

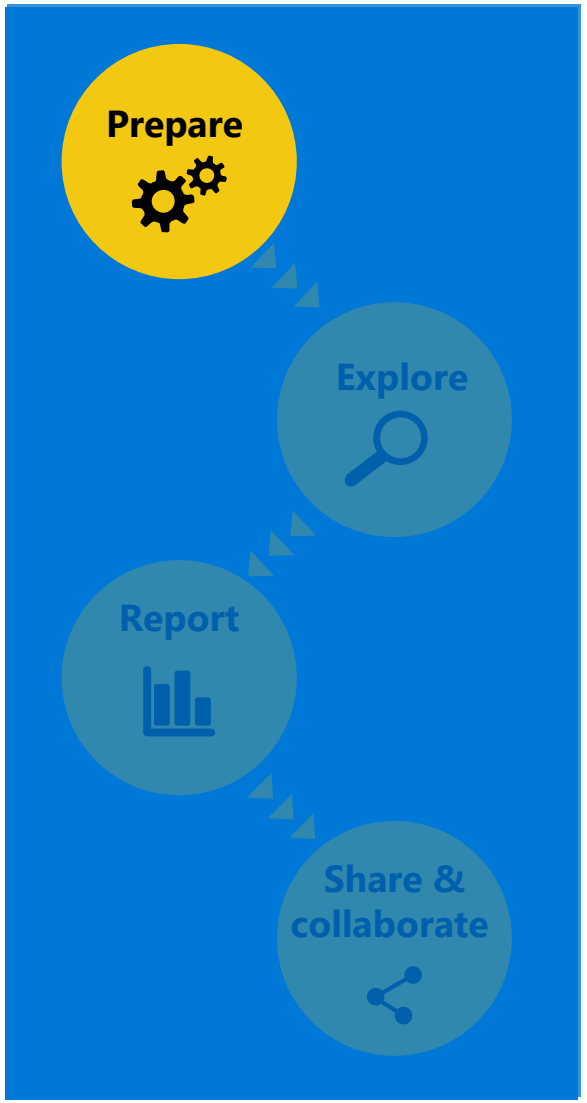


Connecting to SQL Server (an example)

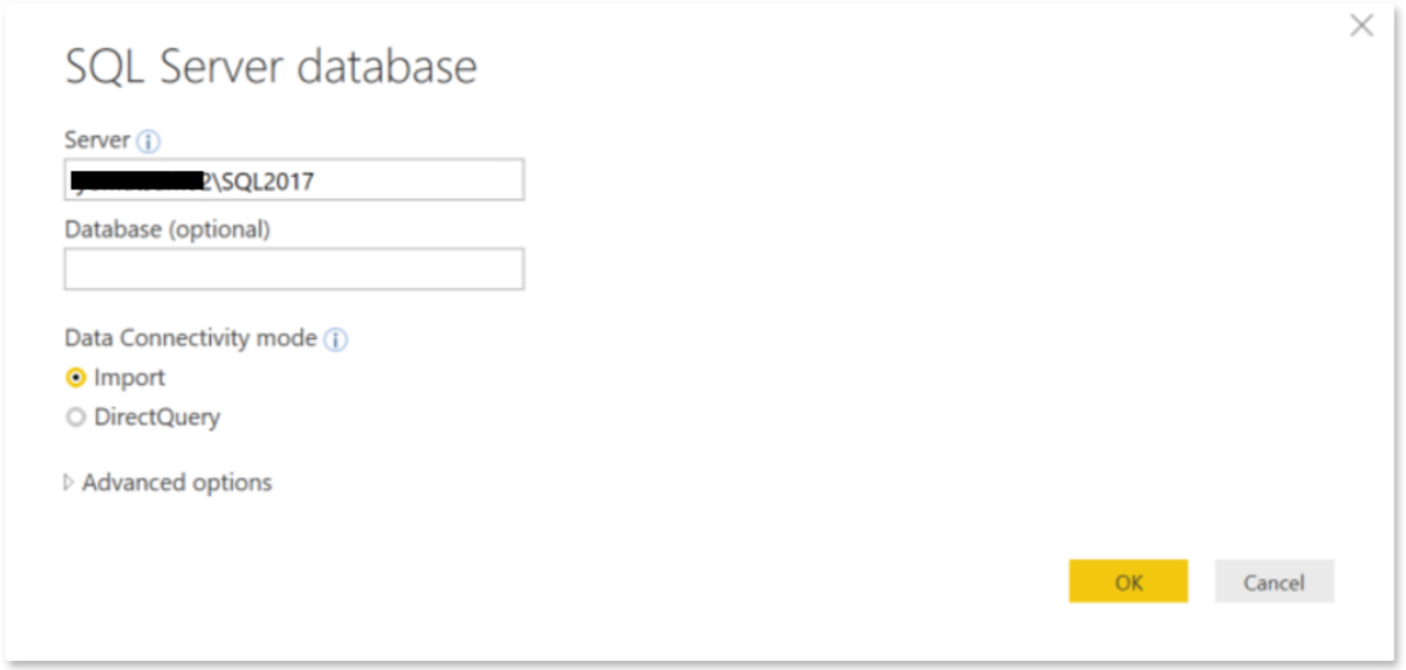


1. Access the **"Get Data"** tab
2. Select **SQL Server Database** from the **"Database"** type of sources
3. And then click **"Connect"**

Connecting Data - Prepare

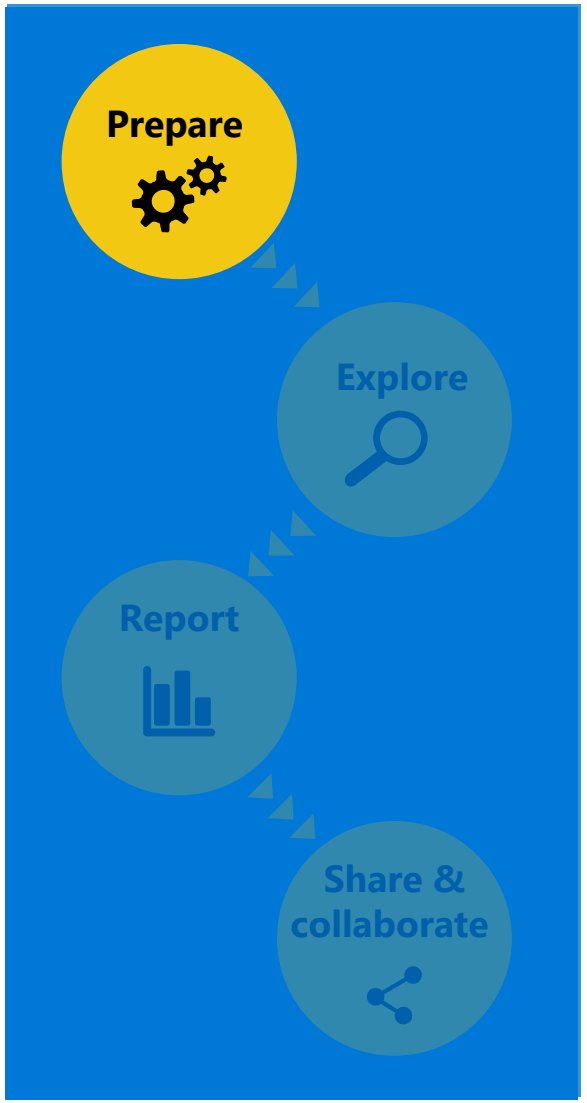


Connecting to SQL Server (an example)

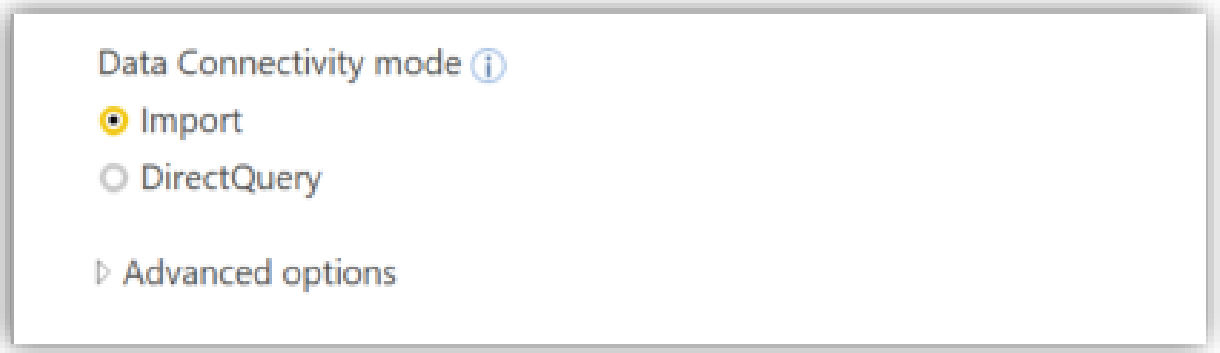


- Specify an **instance** to connect to
- Optionally, specify a **database** – if not, a list will be presented
- Specify **Data Connectivity mode** – further details in next slide.
- Advanced options includes **Timeout, query, failover support option**, etc.

Connecting Data - Prepare

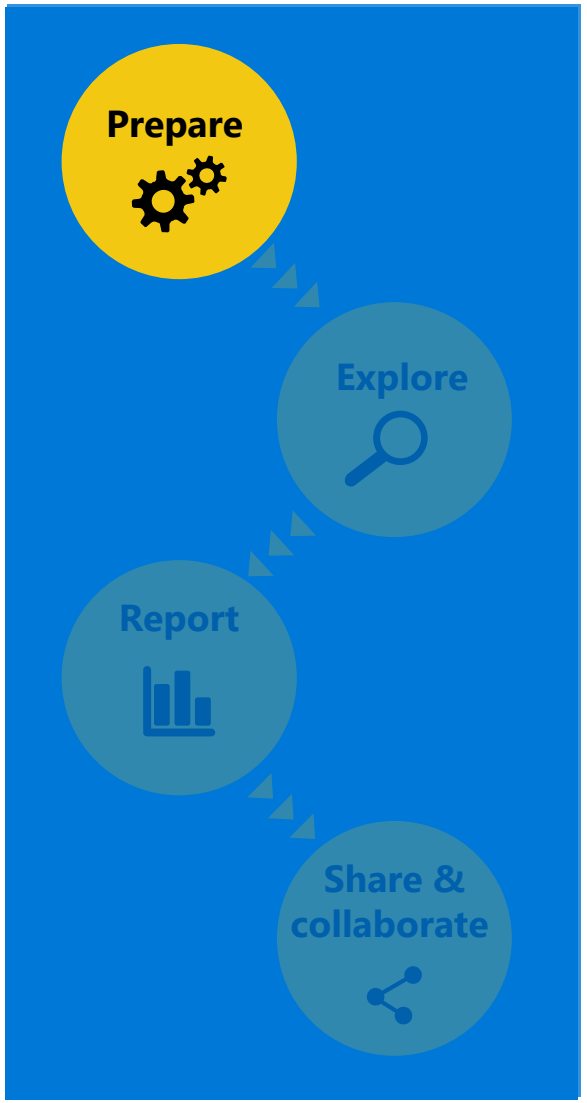


Connecting to SQL Server (an example):



- A **DirectQuery** enabled data source will include the above option in connection setting dialog
- DirectQuery allows **querying very large datasets** and with a **live connection** (no data is imported)
- Supports **additional calculations** and all data either is **Imported** or in **DirectQuery** mode (expect this to be improved in the future)
- **Import** brings all the data into the Desktop

Connecting Data - Prepare



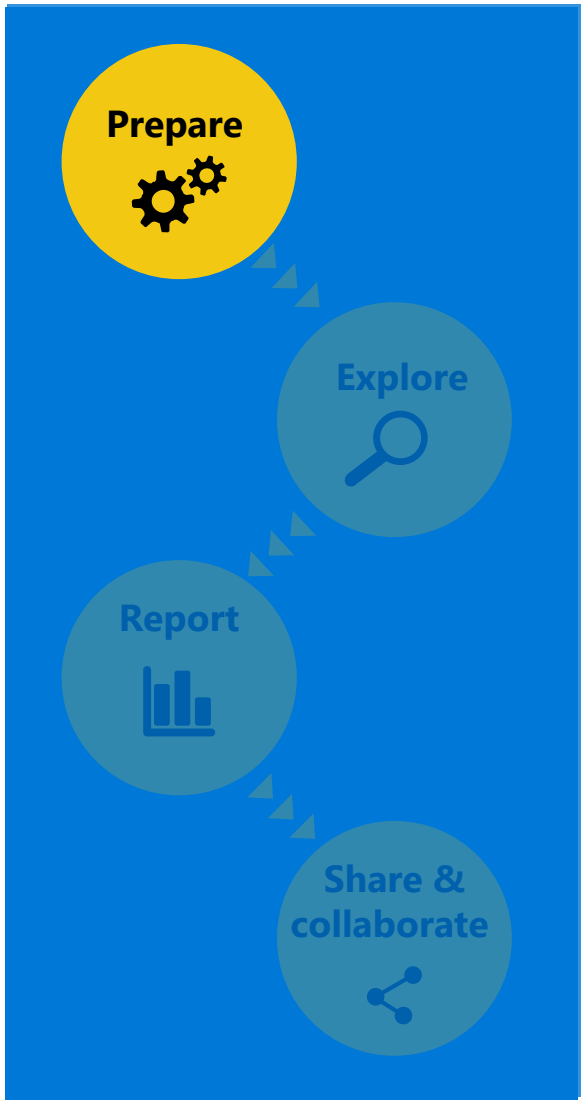
What is Direct Query (DQ)?

- **Connectivity Method** in PowerBI and SSAS where the modeler defines the model but doesn't import any data
- **Any data** needed for visualizations will be retrieved **from the data source directly**

Major Benefits of DQ

- **No need to move any data**, data stays at the source where it is created offering real time updates whenever changes are made.
- Ability to leverage the **security** as it is defined in the data source when using Single Sign On (SSO).
- **You will not hit any memory limits in Power BI** or SSAS so you can see all the raw data (all transactions)

Connecting Data - Prepare



What about pitfalls of DQ?

Data Source
OLTP – not a Datewarehouse

- creating a decent model - more manageability and query complexity (which could lead to perf issues)
- analytical query patterns (e.g. dimensional joins) >>perf issues

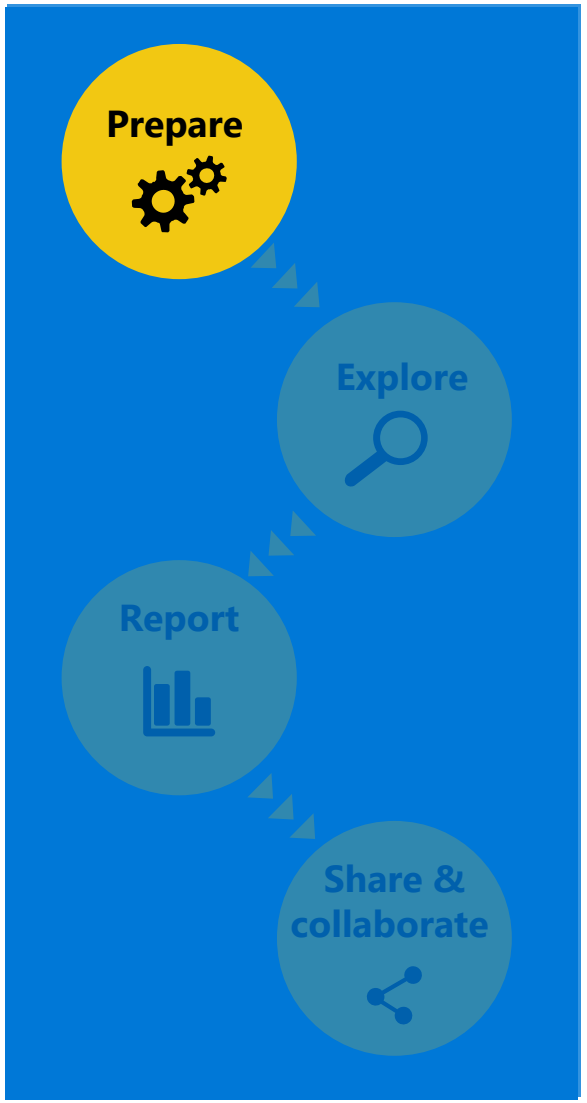
Data Source
MPP \ Big Data solution

- high concurrency might be an issue (the reports are consumed by hundreds of users at the peak time)

Security

- The security as set up in the transactional data source might not be the same security as you expect for reporting

Connecting Data - Prepare



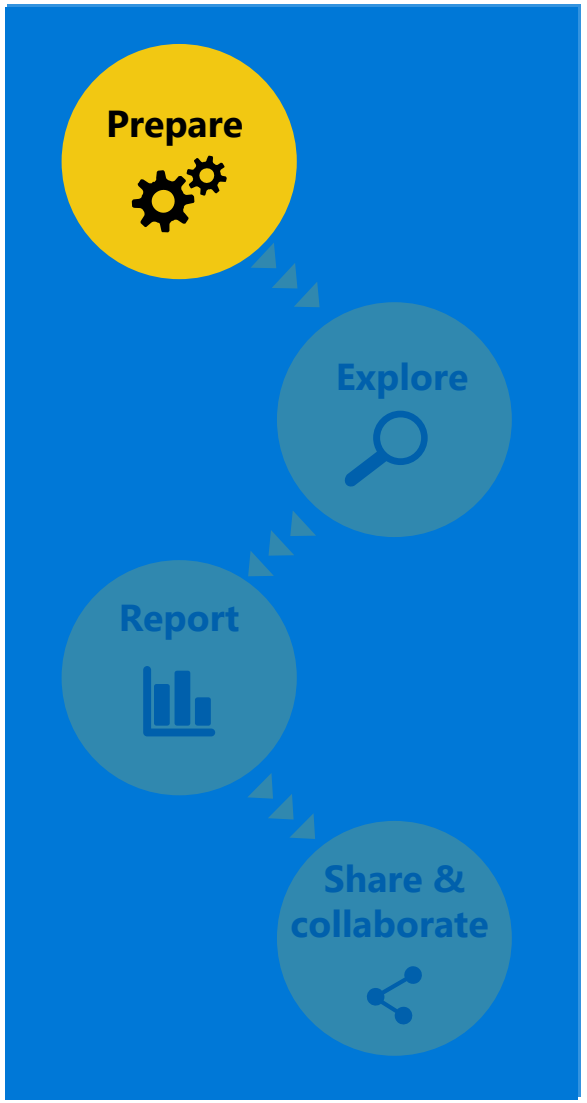
Required Investments to use DQ

- Optimizing your SQL Server data source with in memory indexes designed for the query load the reporting will put on the system.
- Instead of views you create new analytical tables to shape the data and add specific performance features.
- Use clusters specialized for interactive query performance (like serverless pools). Leverage a fast data solution like HANA, SQL DW, Netezza, Spark, etc. In each case test the load together with the amount of concurrency to determine

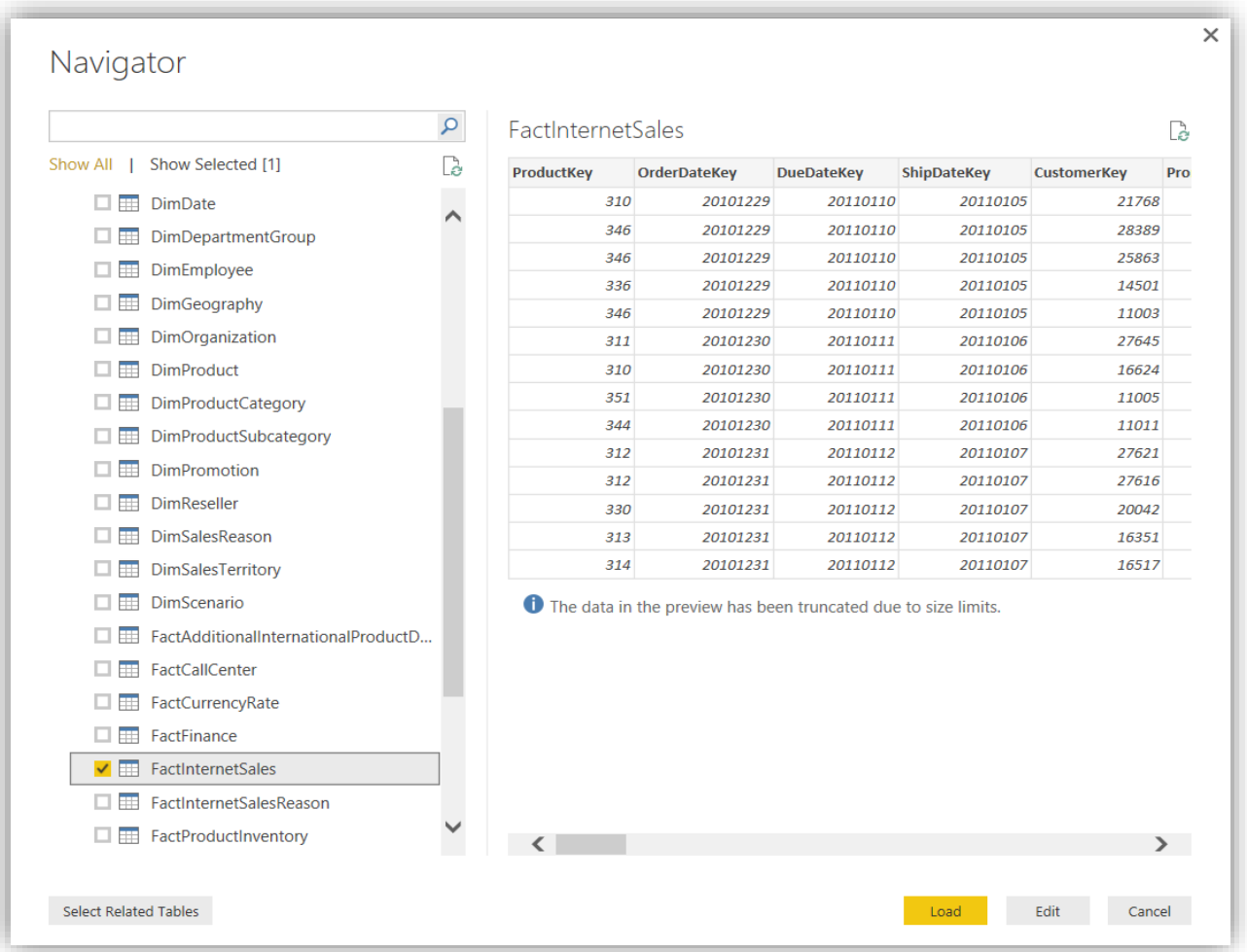
Tradeoffs – DQ or Import?

- Data Refresh – SSAS partitioning
- Memory Limit – smart compression
- Security

Connecting Data - Prepare

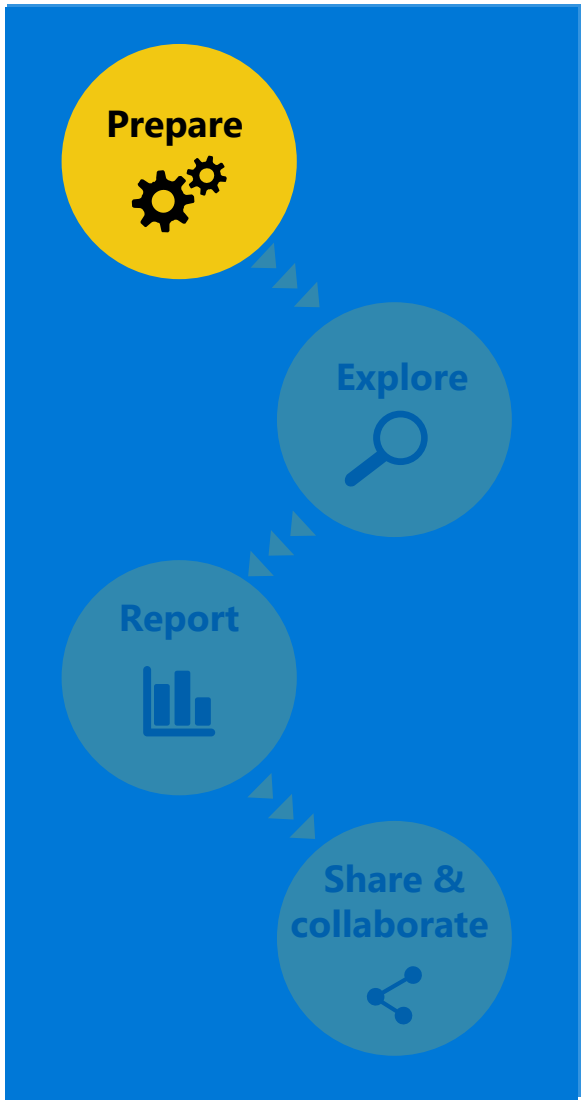


Connecting to SQL Server (an example)



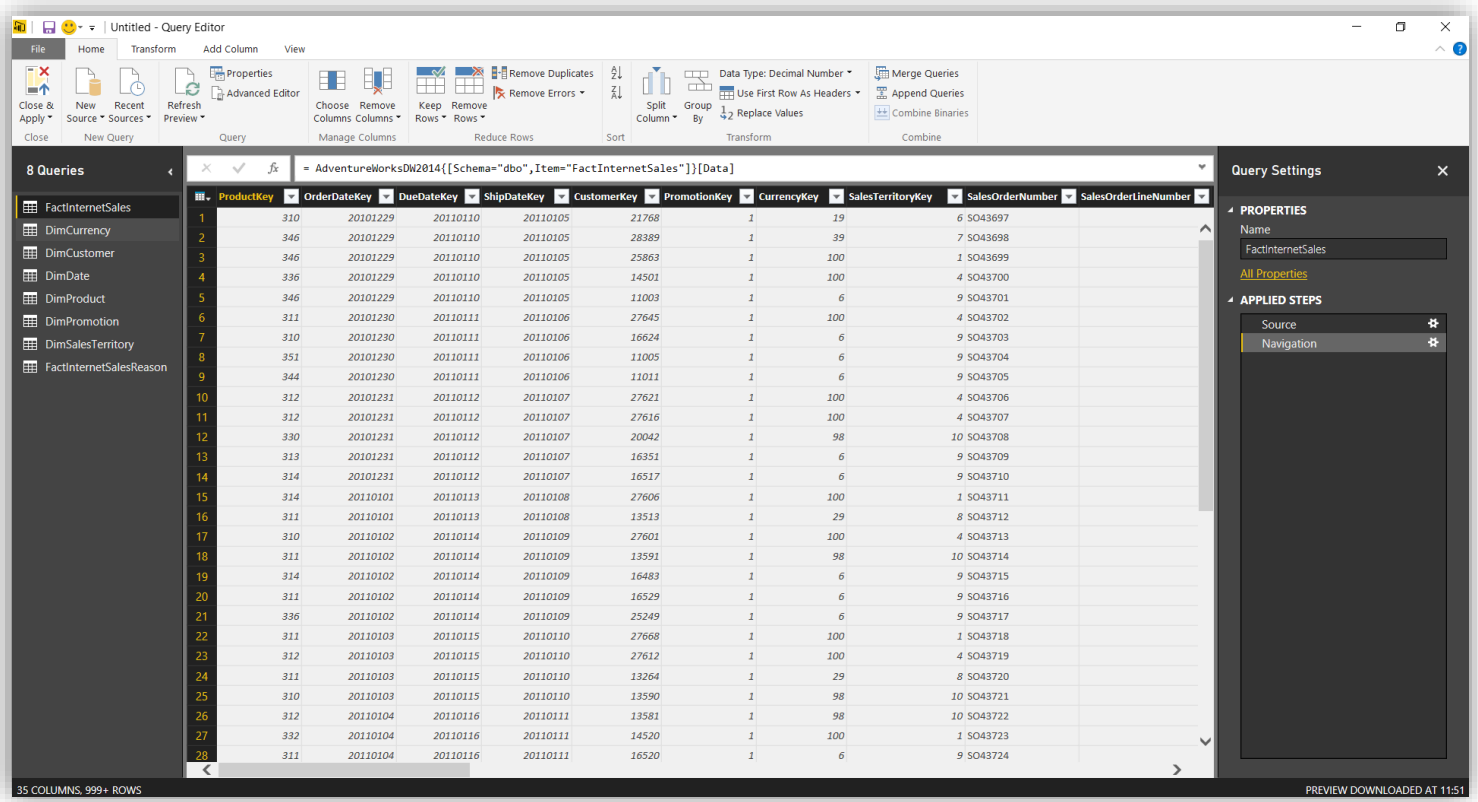
- Now **select** the desired tables
- **A preview** is generated for ease of use
- You can use **relationship detection per-table**
- Choose **“Load”** if the data is ready
- Choose **“Edit”** if the data needs to be transformed

Connecting Data - Prepare

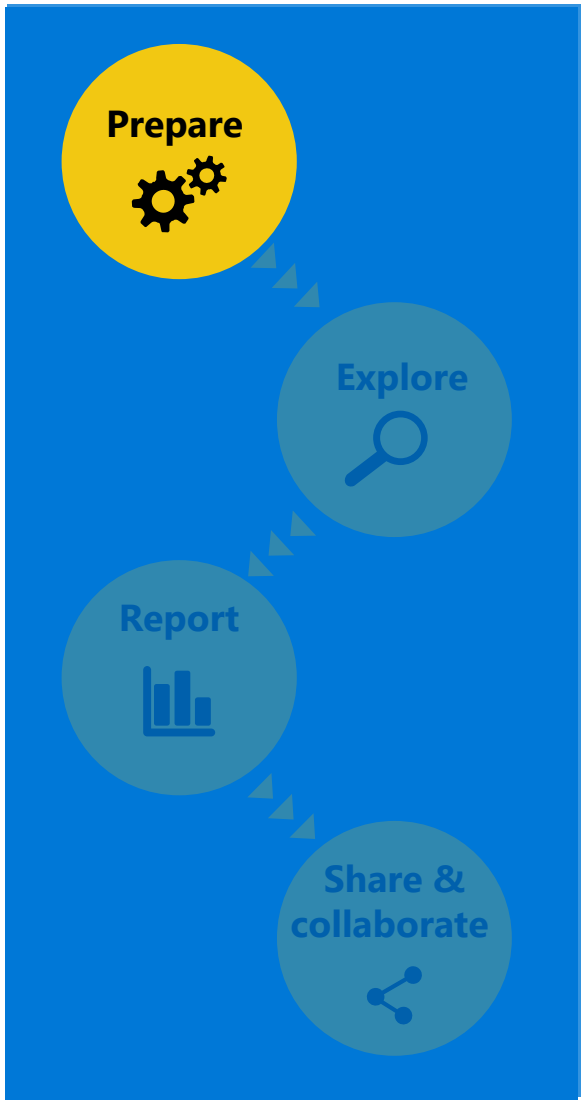


Connecting to Microsoft SQL Server (an example)

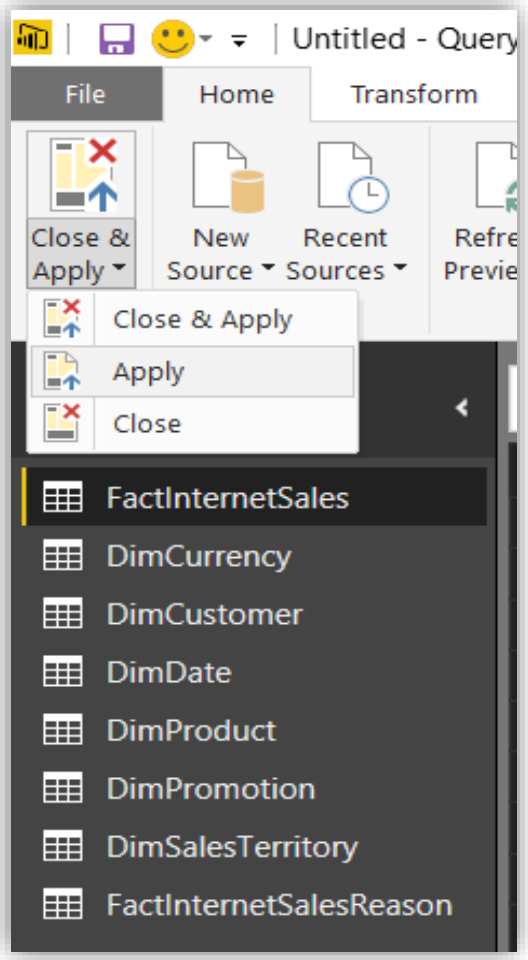
- And we would be presented with the **Query Editor** to further refine our data, if we specified "Edit":



Connecting Data - Prepare

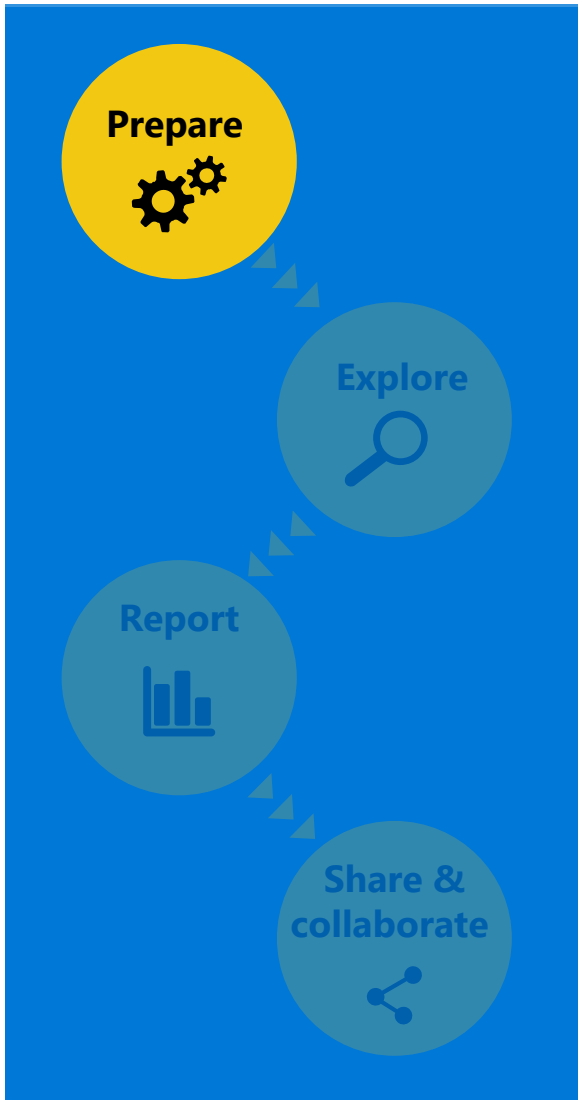


Connecting to SQL Server (an example)



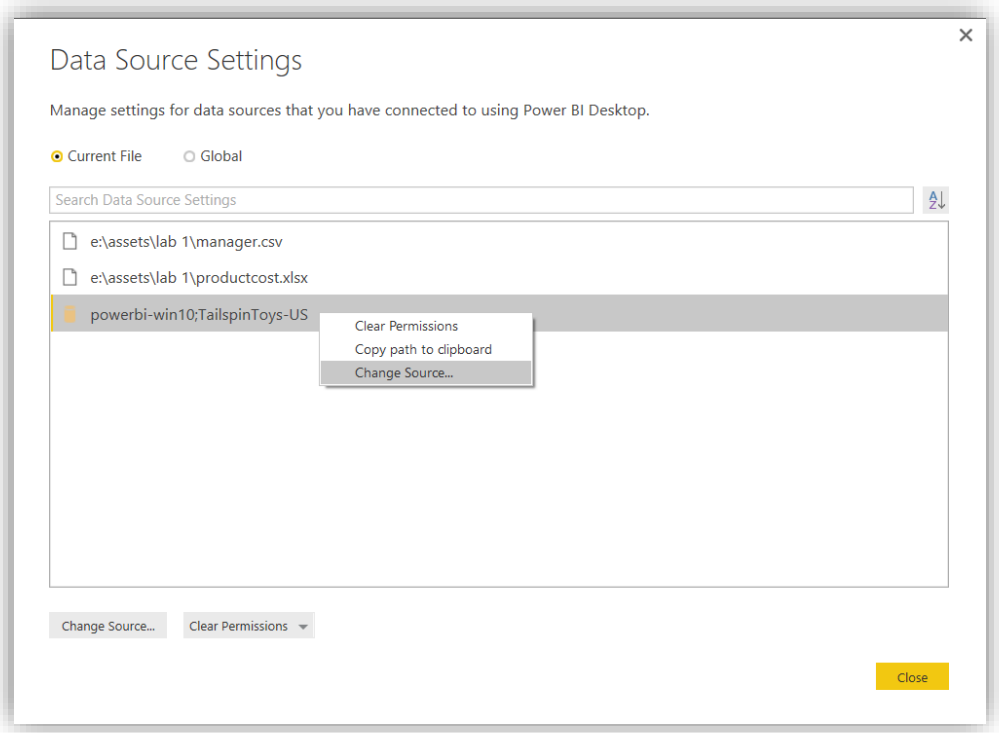
- Finally, we would apply our changes so that we can start exploring the data
- **Close & Apply** – Closes the query editor and loads the data
- **Apply** – Re-loads the data from source
- **Close** – Closes the query editor without applying changes

Connecting Data - Prepare

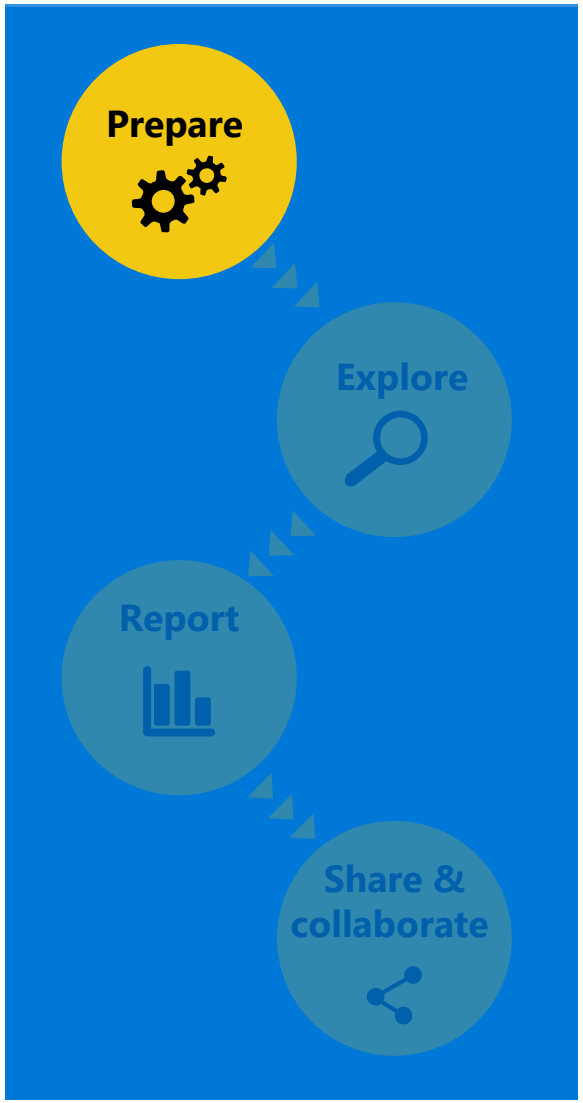


Changing connections

- After the **connections** are established, you can easily **change** them
- For instance, you can change from Test to Production Environment
- **Go to File -> Options and Settings -> Data Source Settings**
- Specified **credentials can be cleared**
- The scope can be the current file or previous created connections



Connecting Data - Prepare



Power BI Custom Connectors

- **Create** your own **data connector** which can be added to the Get Data menu.
- Connect to a **data source that may not be supported yet** by Power BI Desktop
- Use your **own logic**, and have **multiple options for credentials**, such as Windows, basic, API Key, and database authentication; and more.
- Created using the same **M language** used by Power Query; (*Power Query SDK is available)*

Note: Custom data connectors is still in preview for Power BI Desktop as of October 2017