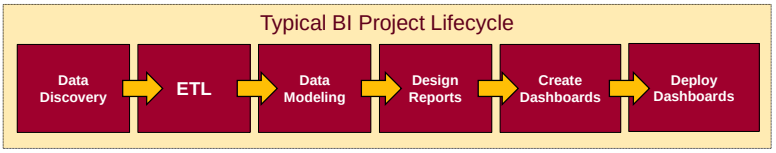
Using the Power Query Features of Power BI Desktop

Agenda

- Getting Started with Power BI Desktop
- Data Sources and Data Discovery
- Power Query Fundamentals
- Working with the Query Editor Window
- Combining Queries
- Importing Data Into a Star Schema

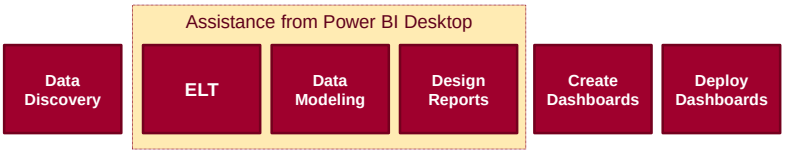
Project Lifecycle for a Custom BI Solution

- Lifecycle of a typical BI project includes...
  - Discover where the data lives
  - Extract, transform and load (ETL) data
  - Model data to create dataset for analytics and reporting
  - Design and implement reports on top of dataset
  - Consolidate reports to one or more dashboards
  - Package project artifacts for deployment
  - Deploy to production environment



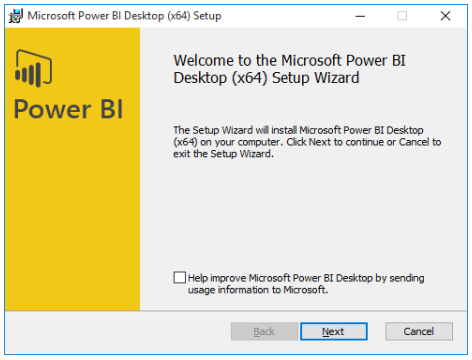
Working with Power BI Desktop

- Power BI Desktop focuses on three phases
  - Power Query features used for ETL
  - Power Pivot and DAX used for data modeling
  - Report creation using the Power BI report designer

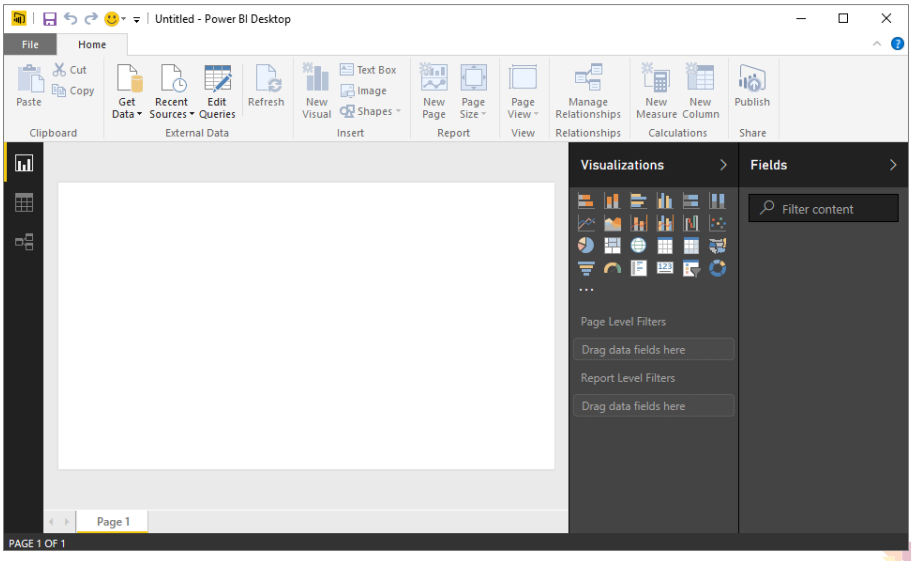


Installing Power BI Desktop

- Power BI Desktop installs as click-once application
  - Can be installed when logged into Power BI service



Launching Power BI Desktop



Projects and PBIX Files

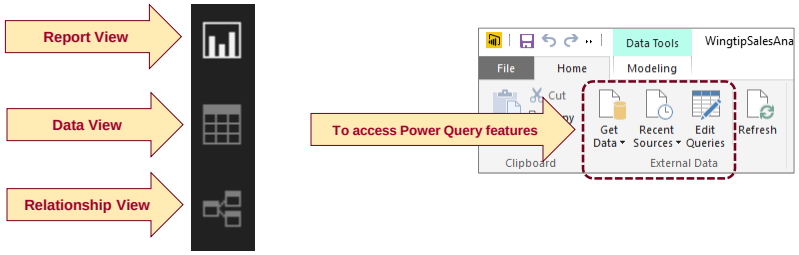
- Power BI Desktop projects saved using PBIX files
  - PBIX file contains queries created with Power Query
  - PBIX file contains data imported using queries
  - PBIX file contains data modeling definitions
  - PBIX file contains reports

Agenda

- ✓ The BI Project Lifecycle
- ➢ Getting Started with Power BI Desktop
- Using Power Query to Import Data
- Using Power Pivot to Model Data
- Publishing Power BI Desktop Projects

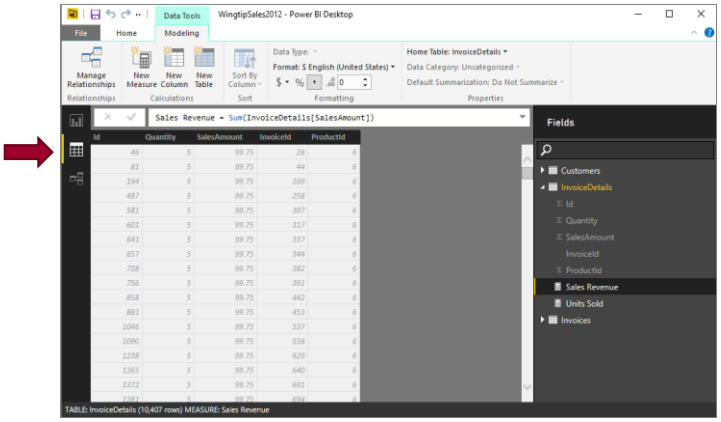
Getting Around in Power BI Desktop

- What do you need to learn to use Power BI Desktop?
  - Power query features for importing data
  - Power Pivot features for modeling data
  - Report designer for creating reports
- Navigating between view modes



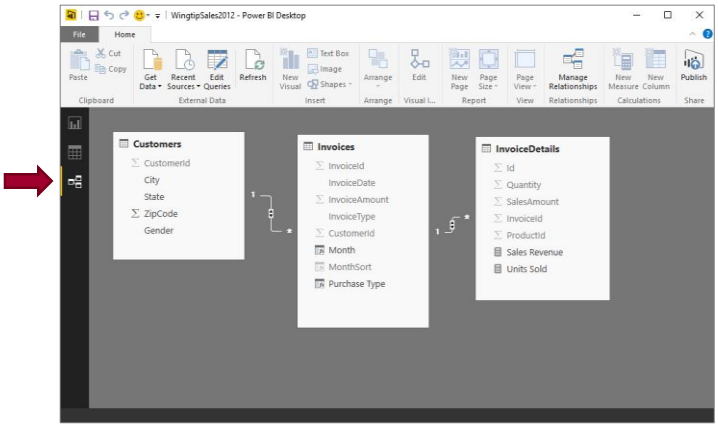
Data View

- Primary view used when data modeling with Power Pivot
  - Data view displays columns and rows for each table
  - You can extend tables with calculated columns and fields



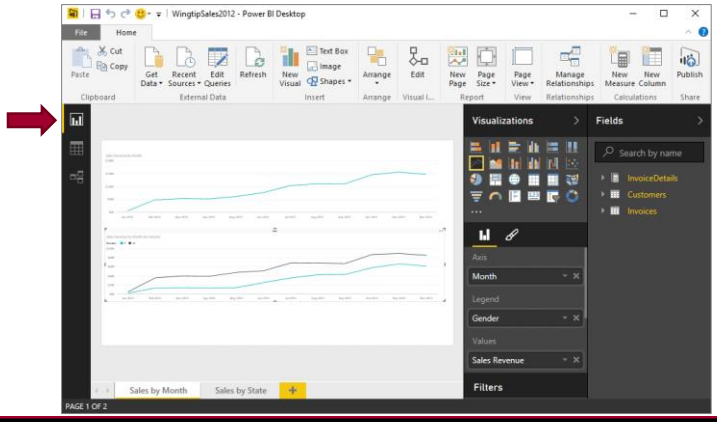
Relationship View

- Displays tables, fields and relationships
  - Used to view tables, fields and relationships in project's dataset
  - Used to create relationships when importing new tables



Report View

- Report view displays report of current project
  - Report designer is the same as in the Power BI service
  - Dataset appears in consumer mode - not author mode



Agenda

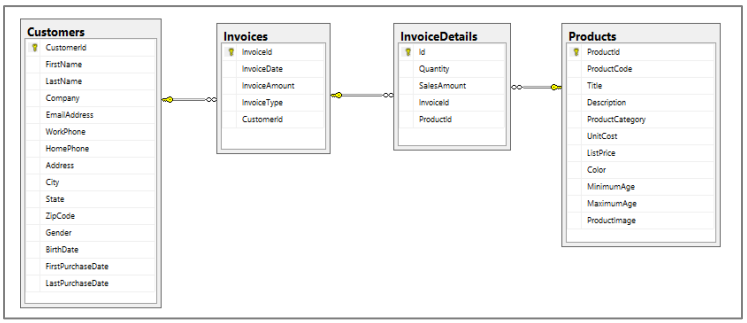
- ✓ Getting Started with Power BI Desktop
- ➢ Data Sources and Data Discovery
- Power Query Fundamentals
- Working with the Query Editor Window
- Combining Queries
- Importing Data Into a Star Schema

Data Discovery

- Data can live in a variety of sources
  - Files (e.g. CSV file, Excel workbook)
  - OLTP Databases
  - OLAP Databases
  - SaaS Applications
  - Web services

An Example OLTP Database

- Online Transaction Processing (OLTP) System
  - Used for real-time data access and transaction-based data entry
  - Optimized for faster transactions (e.g. inserts, updates & deletes)
  - Tables normalized to reduce/eliminate redundancies
  - Table schemas can be hard for business users to understand



Deciding What To Measure

- You Must Determine Measurable Objectives
  - Financial (revenue, expenses, profit margin, etc.)
  - Business processes efficiency
  - Customer Satisfaction Levels

Defining Grain Statements

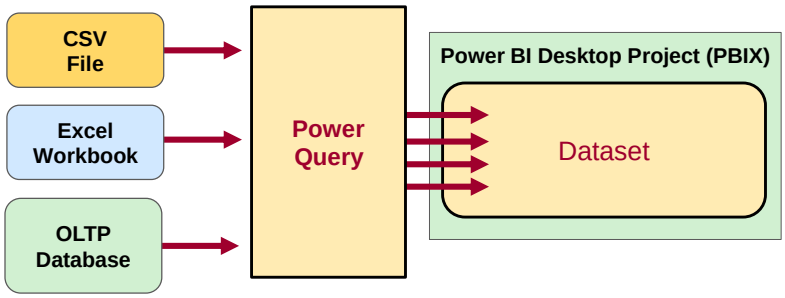
- Grain statements should be defined in initial design phase
  - Grain statements helps determine requirements for BI queries
  - Grain statements can be created & understood by business users
- Example grain statements for BI project at Wingtip Toys
  - What was the total sales revenue over the last 4 years?
  - What was the sales revenue by year, quarter and month?
  - What was the sales revenue by region, state, city and zip code?
  - What was the sales revenue by category, subcategory and product?
  - What was the growth in sales revenue from month to month in 2013?
  - What was profit margin for each product by year, quarter and month?
  - Have their been any products with significantly decreasing profit margin?

Agenda

- ✓ Getting Started with Power BI Desktop
- ✓ Data Sources and Data Discovery
- ➢ Power Query Fundamentals
- Working with the Query Editor Window
- Combining Queries
- Importing Data Into a Star Schema

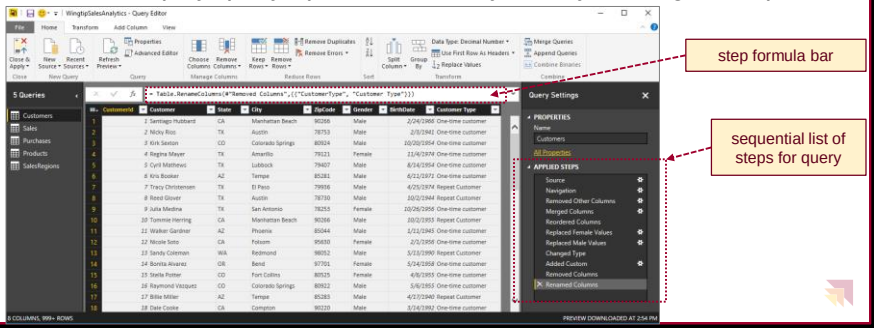
Power Query is an ETL Tool

- ETL process is essential part of any BI Project
  - **Extract** the data from wherever it lives
  - **Transform** the shape of the data for better analysis
  - **Load** the data into dataset for analysis and reporting



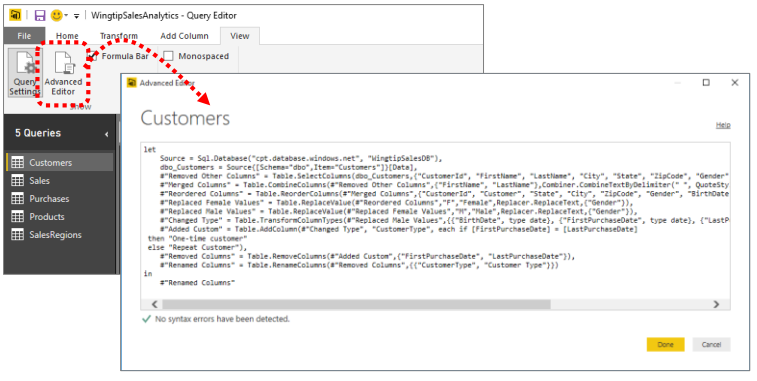
Query Steps

- A query is created as a sequence of steps
  - Each step is a parameterized operation on the data
  - Each step has formula which can be viewed/edited in formula bar
  - Query starts with Source step to extract data from a data source
  - Additional steps added to perform transform operations on data
  - You can replay query operations one by one by clicking on steps



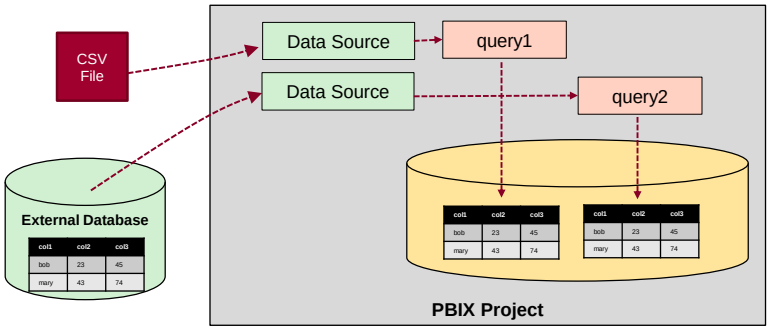
Advanced Editor

- Power Query based on "M" functional language
  - Query in Power Query saved as a set of M statements in code
  - Query Editor generates code in M behind the scenes
  - Advanced users can view & modify query code in Advanced Editor



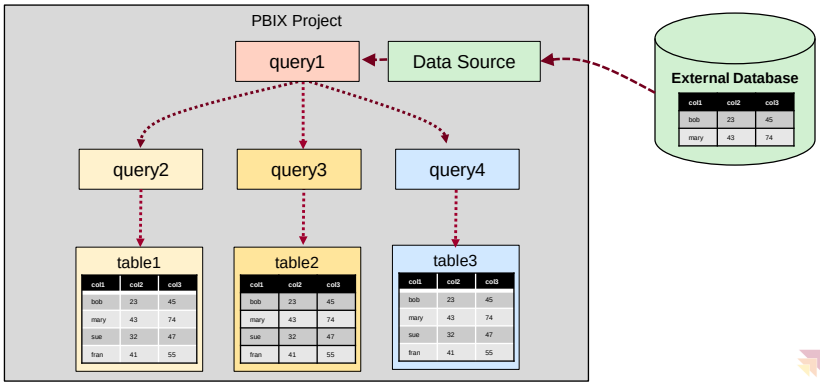
Understanding Query Input and Output

- PBIX project is container for data sources and queries
  - Queries created and saved within scope of Power BI project
  - Queries can pull data from local files
  - Queries can pull data from external content sources
  - Queries main purpose is to load imported data into data model



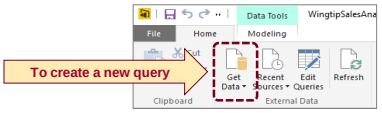
Query Composition

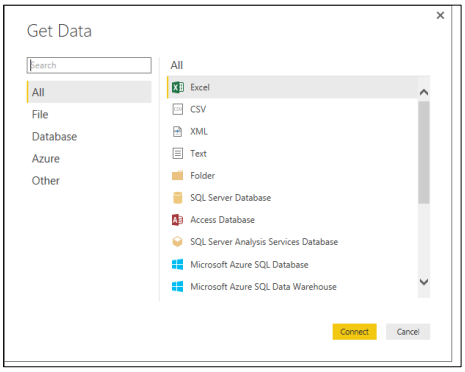
- Query can serve as source for other queries
  - Allows for creation of reusable base queries & query composition
  - Complexity can be hidden in base queries
  - Reference command creates new query based on another query



Creating a New Query

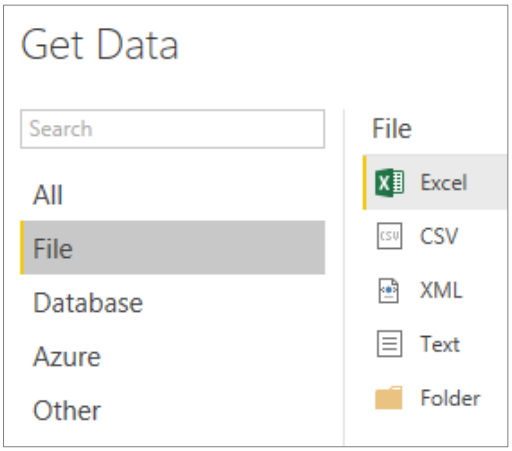
- New query created using **Get Data** menu on ribbon
  - **Get Data** menu has many commands for creating new queries
  - You can create new query based on large variety of data sources
  - New query is opened in Power Query's Query Editor window





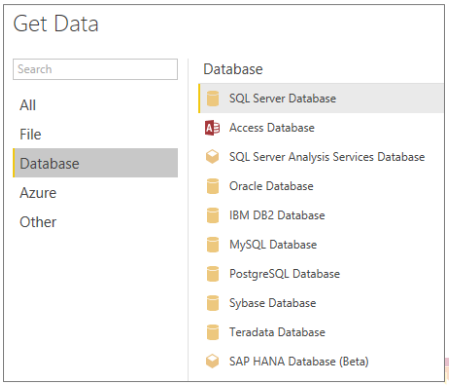
Supported File-based Data Sources

- Power Query supports common file types
  - Excel workbooks
  - CSV files
  - XML files
  - Text files
  - Folder structure



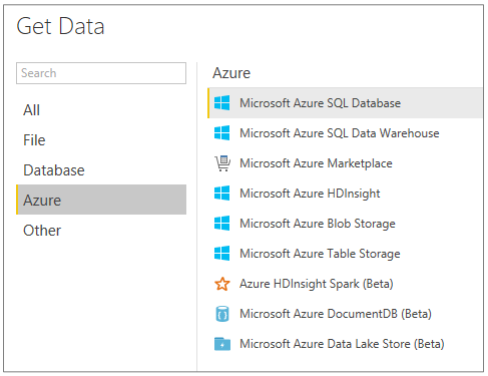
Supported Databases

- Power Query supports many database systems
  - SQL Server
  - SQL Server Analysis Services
  - Access
  - Oracle
  - DB2
  - MySQL
  - PostgreSQL
  - Sybase
  - Teradata

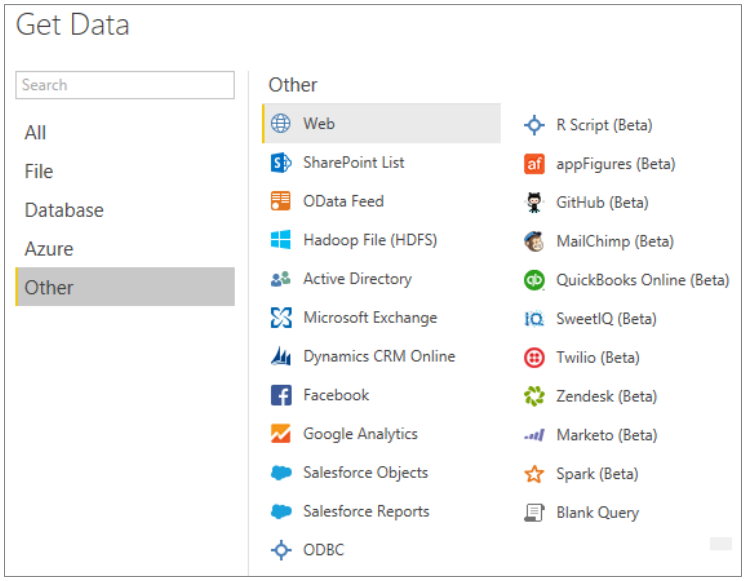


Azure Data Sources

- Power Query supports many azure data sources

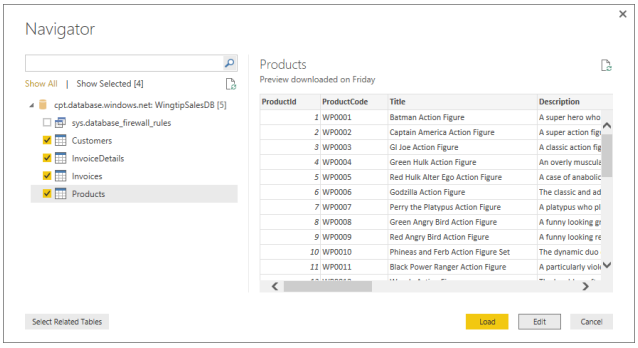


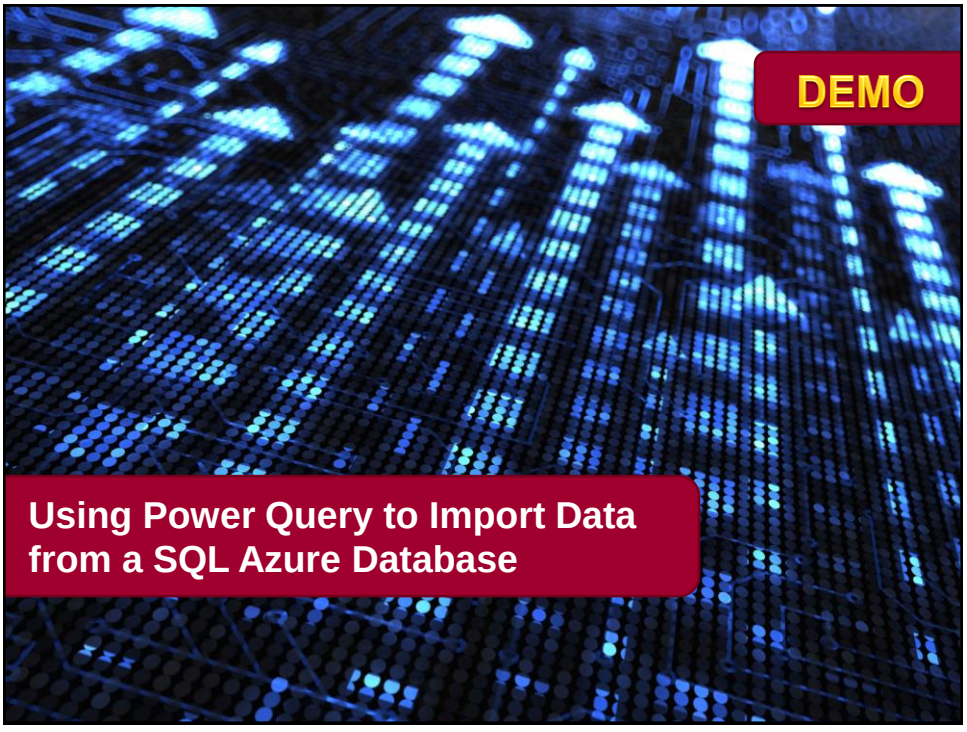
Other Supported Data Sources



Selecting Tables from a Data Source

- Power Query provides Navigator dialog
  - Allows you to select tables
  - Navigator understands existing table relationships
  - Clicking **Load** will run query and import data
  - Clicking **Edit** will open queries in Query Editor window



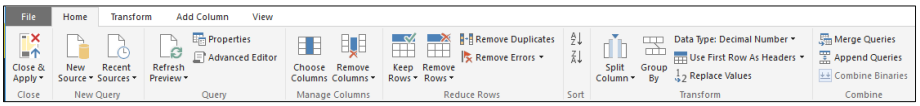


Agenda

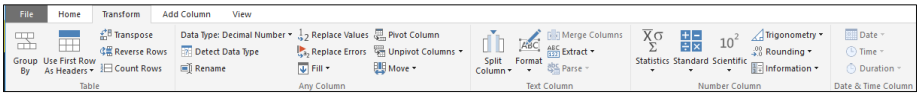
- ✓ Getting Started with Power BI Desktop
- ✓ Power Query Fundamentals
- ✓ Power Query Data Sources
- ➢ Working with the Query Editor Window
- • Combining Queries
- • Importing Data Into a Star Schema

Query Editor Ribbon Tabs

Home tab



Transform tab



Add Columns tab

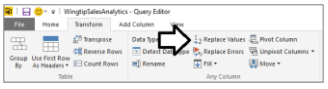
View tab

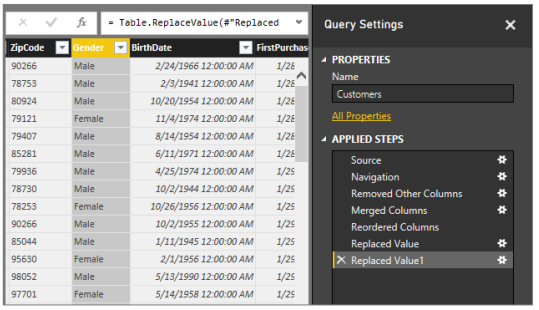
Examples of Basic Power Query Steps

- Rename column
- Convert column type
- Format column values
- Reorder columns
- Replace column values
- Expanding related column
- Merging columns
- Splitting columns
- Adding custom column

Replacing Values

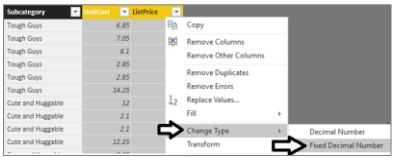
- Used to substitute values during import





Converting Column Types

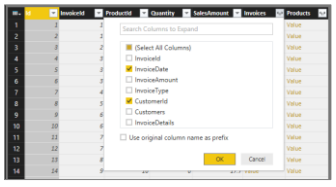
- Transform data to make it more reliable
  - Convert date-time column to date column
- Transform data to make it more efficient
  - Convert decimal to fixed decimal number for currency



- Beware: Conversion can have destructive effect on data

Expanding Related Columns

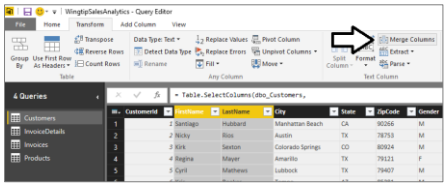
- Used to pull data from related tables
  - Saves you from performing SQL joins or VLOOKUP

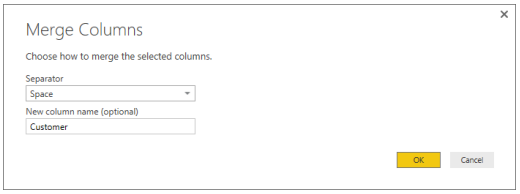


	Id	InvoiceId	ProductId	Quantity	SalesAmount	InvoiceDate	CustomerId	Products
1	1	1	22	4	119.8	1/28/2012 12:00:00 AM	1	Value
2	2	1	22	1	29.95	1/28/2012 12:00:00 AM	1	Value
3	3	2	22	2	59.9	1/28/2012 12:00:00 AM	2	Value
4	4	3	17	8	399.6	1/28/2012 12:00:00 AM	3	Value
5	5	3	18	2	29.9	1/28/2012 12:00:00 AM	3	Value
6	6	3	18	4	59.8	1/28/2012 12:00:00 AM	3	Value
7	7	4	16	1	2.95	1/28/2012 12:00:00 AM	4	Value

Merging Columns

- Merge two columns into a single column

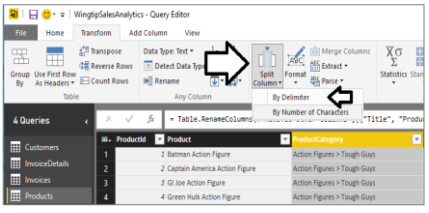


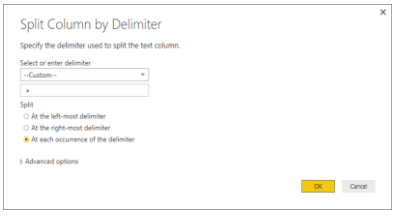


	CustomerId	Customer
1	1	Santiago Hubbard
2	2	Nicky Rios
3	3	Kirk Sexton
4	4	Regina Mayer
5	5	Cyril Mathews
6	6	Kris Booker
7	7	Tracy Christensen
8	8	Reed Glover

Splitting Columns

- Split a single column up into two columns

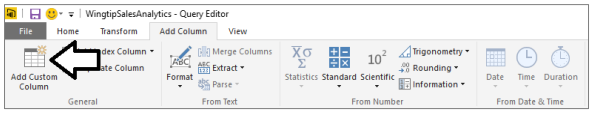


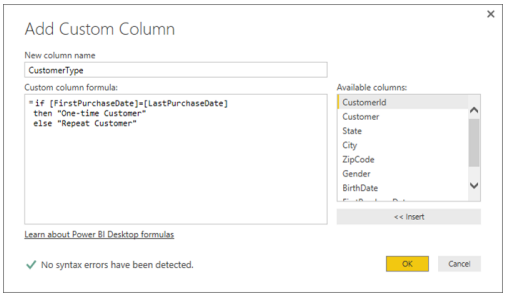


	ProductId	Product	Category	Subcategory
1	1	Batman Action Figure	Action Figures	Tough Guys
2	2	Captain America Action Figure	Action Figures	Tough Guys
3	3	GI Joe Action Figure	Action Figures	Tough Guys
4	4	Green Hulk Action Figure	Action Figures	Tough Guys
5	5	Red Hulk Alter Ego Action Figure	Action Figures	Tough Guys
6	6	Godzilla Action Figure	Action Figures	Tough Guys
7	7	Perry the Platypus Action Figure	Action Figures	Cute and Huggable
8	8	Green Angry Bird Action Figure	Action Figures	Cute and Huggable

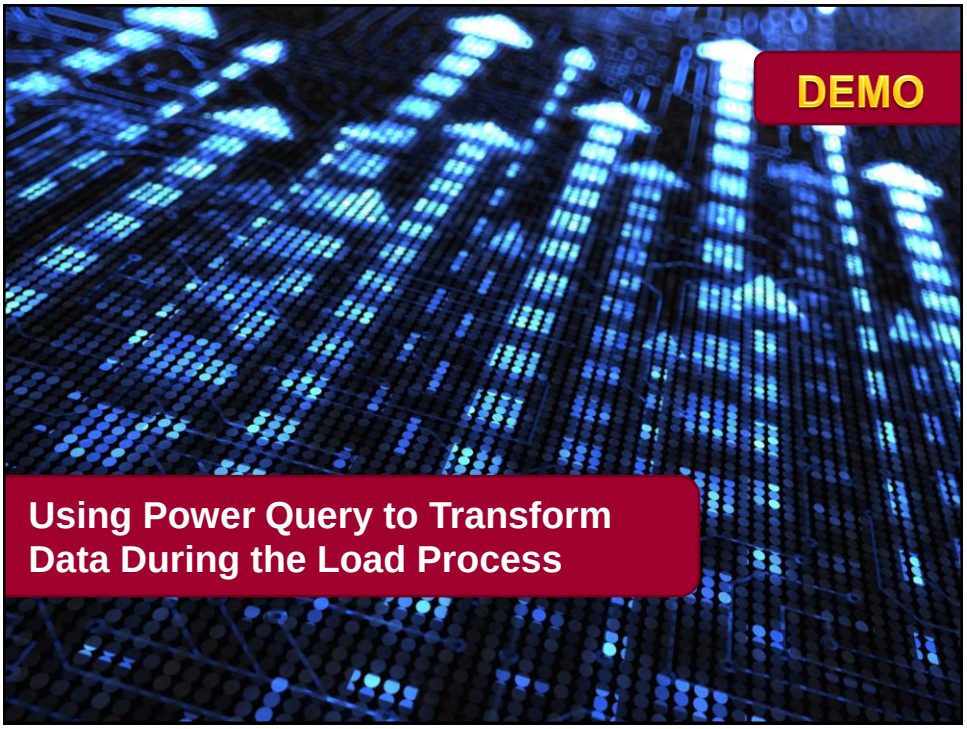
Adding a Custom Column

- Custom column provide custom logic
  - Logic must be written in M programming language





FirstPurchaseDate	LastPurchaseDate	CustomerType
1/28/2012	1/28/2012	One-time Customer
1/28/2012	1/28/2012	One-time Customer
1/28/2012	1/28/2012	One-time Customer
1/28/2012	1/28/2012	One-time Customer
1/28/2012	1/28/2012	One-time Customer
1/28/2012	1/28/2012	One-time Customer
1/29/2012	11/22/2015	Repeat Customer
1/29/2012	10/2/2015	Repeat Customer
1/29/2012	1/29/2012	One-time Customer
1/29/2012	5/6/2015	Repeat Customer
1/29/2012	1/29/2012	One-time Customer

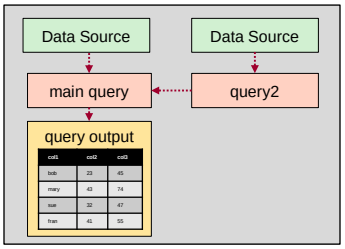


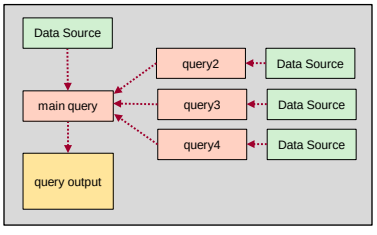
Agenda

- ✓ Getting Started with Power BI Desktop
- ✓ Power Query Fundamentals
- ✓ Power Query Data Sources
- ✓ Working with the Query Editor Window
- ➢ Combining Queries
- • Importing Data Into a Star Schema

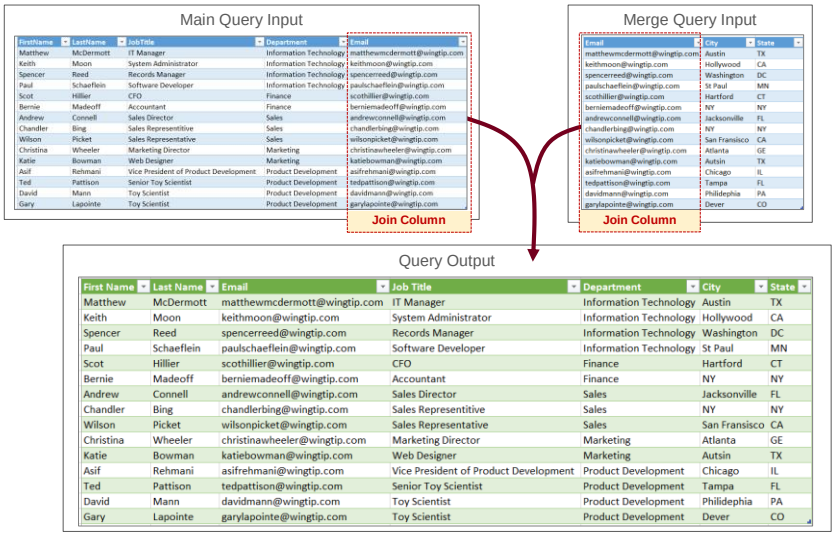
Combining Queries

- Query can be merged or appended with another query
  - Merge operation allows you combine columns from two tables
  - Append operation allows you to combine rows from two tables
- Two queries are combined into single output for loading
  - Load settings of main query determines where output is loaded
  - Secondary query acts as source for main query
  - Secondary query be can created with connection-only load setting



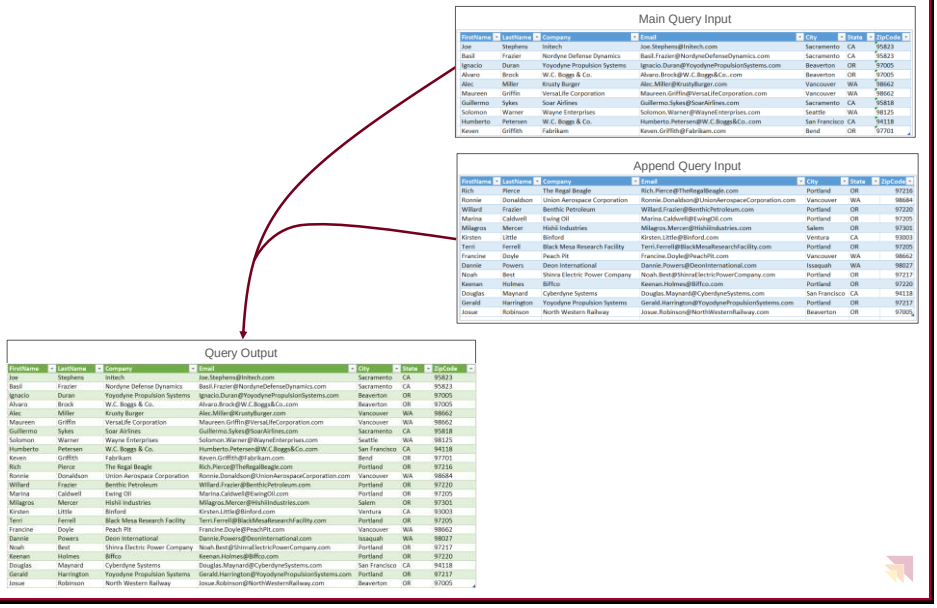


Merging Columns



First Name	Last Name	Email	Job Title	Department	City	State
Matthew	McDermott	matthewmcdermott@wingtip.com	IT Manager	Information Technology	Austin	TX
Keith	Moon	keithmoon@wingtip.com	System Administrator	Information Technology	Hollywood	CA
Spencer	Reed	spencerreed@wingtip.com	Records Manager	Information Technology	Washington	DC
Paul	Schaeflein	paulschaeflein@wingtip.com	Software Developer	Information Technology	St Paul	MN
Scot	Hillier	scothillier@wingtip.com	CFO	Finance	Hartford	CT
Bernie	Madeoff	berniemadeoff@wingtip.com	Accountant	Finance	NY	NY
Andrew	Connell	andrewconnell@wingtip.com	Sales Director	Sales	Jacksonville	FL
Chandler	Bing	chandlerbing@wingtip.com	Sales Representitive	Sales	NY	NY
Wilson	Picket	wilsonpicket@wingtip.com	Sales Representative	Sales	San Fransisco	CA
Christina	Wheeler	christinawheeler@wingtip.com	Marketing Director	Marketing	Atlanta	GE
Katie	Bowman	katiebowman@wingtip.com	Web Designer	Marketing	Autsin	TX
Asif	Rehmani	asifrehmani@wingtip.com	Vice President of Product Development	Product Development	Chicago	IL
Ted	Pattison	tedpattison@wingtip.com	Senior Toy Scientist	Product Development	Tampa	FL
David	Mann	davidmann@wingtip.com	Toy Scientist	Product Development	Philidephia	PA
Gary	Lapointe	garylapointe@wingtip.com	Toy Scientist	Product Development	Dever	CO

Appending Rows



Pivoting Columns

- Pivot column adds its values are new columns
  - Create table layout like PivotTable

City	Year	Population
New York, NY	2010	8175133
New York, NY	2000	8008278
New York, NY	1990	7322564
Los Angeles, CA	2010	3792621
Los Angeles, CA	2000	3694820
Los Angeles, CA	1990	3485398
Chicago, IL	2010	2695598
Chicago, IL	2000	2896016
Chicago, IL	1990	2783726
Houston, TX	2010	2100263
Houston, TX	2000	1953631
Houston, TX	1990	1630553
Philadelphia, PA	2010	1526006
Philadelphia, PA	2000	1517550
Philadelphia, PA	1990	1585577
Phoenix, AZ	2010	1445632
Phoenix, AZ	2000	1321045
Phoenix, AZ	1990	983403

City	2010	2000	1990
New York, NY	8175133	8008278	7322564
Los Angeles, CA	3792621	3694820	3485398
Chicago, IL	2695598	2896016	2783726
Houston, TX	2100263	1953631	1630553
Philadelphia, PA	1526006	1517550	1585577
Phoenix, AZ	1445632	1321045	983403
San Antonio, TX	1327407	1144646	935933
San Diego, TX	1307402	1223400	1110549
Dallas, TX	1197816	1188580	1006877
San Jose, CA	945942	894943	782248

Unpivoting Columns

- Unpivot columns to collapse them into single column
  - Removes PivotTable layout
  - Can be useful to prepare data for charting and analysis

City	2010	2000	1990
New York, NY	8175133	8008278	7322564
Los Angeles, CA	3792621	3694820	3485398
Chicago, IL	2695598	2896016	2783726
Houston, TX	2100263	1953631	1630553
Philadelphia, PA	1526006	1517550	1585577
Phoenix, AZ	1445632	1321045	983403
San Antonio, TX	1327407	1144646	935933
San Diego, TX	1307402	1223400	1110549
Dallas, TX	1197816	1188580	1006877
San Jose, CA	945942	894943	782248

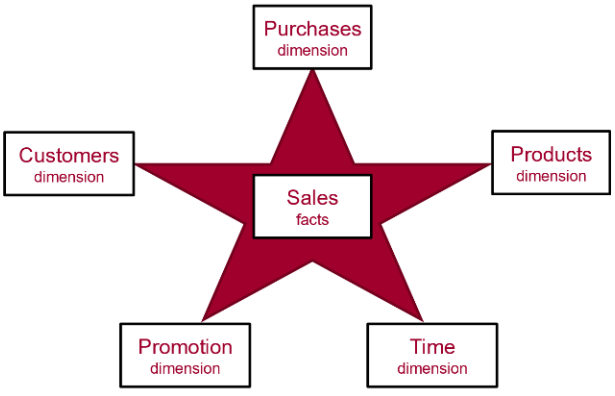
City	Year	Population
New York, NY	2010	8175133
New York, NY	2000	8008278
New York, NY	1990	7322564
Los Angeles, CA	2010	3792621
Los Angeles, CA	2000	3694820
Los Angeles, CA	1990	3485398
Chicago, IL	2010	2695598
Chicago, IL	2000	2896016
Chicago, IL	1990	2783726
Houston, TX	2010	2100263
Houston, TX	2000	1953631
Houston, TX	1990	1630553
Philadelphia, PA	2010	1526006
Philadelphia, PA	2000	1517550
Philadelphia, PA	1990	1585577
Phoenix, AZ	2010	1445632
Phoenix, AZ	2000	1321045
Phoenix, AZ	1990	983403

Agenda

- ✓ Getting Started with Power BI Desktop
- ✓ Power Query Fundamentals
- ✓ Power Query Data Sources
- ✓ Working with the Query Editor Window
- ✓ Combining Queries
- ➢ Importing Data Into a Star Schema

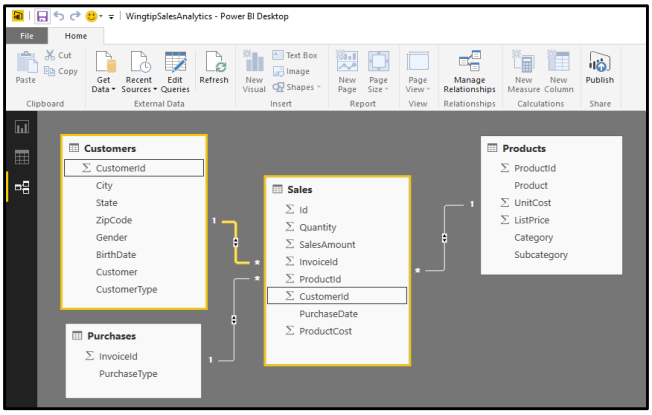
Data Modeling using a Star Schema

- OLAP Modeling often based on Star Schema
  - Tables defined as fact tables or dimension tables
  - Fact tables related to dimension table using 1-to-many relationships



Building a Star Schema with Power Query

- Eliminate need to join three tables
  - Pull CustomerId column into Sales table



Summary

- ✓ Getting Started with Power BI Desktop
- ✓ Power Query Fundamentals
- ✓ Power Query Data Sources
- ✓ Working with the Query Editor Window
- ✓ Combining Queries
- ✓ Importing Data Into a Star Schema