DATA SOLUTIONS PLATFORM

DATA ANALYTICS

POWER BI DATA ANALYSIS

With Power BI Expert…. ***Benjamin Taiwo***

TODAY'S CONTENT

- Introduction to our Platform and Set Expectation
- Meet the Trainer
- Why Data Analytics?
- ETL Framework
- Course Outline
- Course Project
- Introducing Power BI
- Prepare the Data
- Data in Analytics Explained
- Constraints and Data Types
- Data Sources
- Connecting to Data
- More Details on the Platform
- Q & A

Introduction to our Platform and Set Expectation

- This is a project-based course, designed for learners looking to build a solid career in the data space, be job-ready for a Power BI Analyst role as well as pass the **Microsoft DP-900 and PL-300** exam

Course resources include:

- End to End Power BI Downloadable PDF eBook
- Weekly Assignments and Hands-On Demos
- At least one paid certification exam
- Practice Test to replicate the DP-900 and PL-300 exam experience

Expectations Set:

- Power Query
- Power BI Desktop
- Power BI Service
- DAX
- SQL – Data Query Language (DQL)

Meet the Trainer – Benjamin Taiwo BSc, MEng

Previous Roles:

- Data Analyst
- Business Intelligence Analyst
- Data Analytics and Reporting Specialist
- Senior Database Modeler

****Microsoft Certified

Industries Experience:

- Manufacturing
- Government
- Health
- Insurance
- Telecom
- Others

Why Data Analytics?

Data in Abundance:

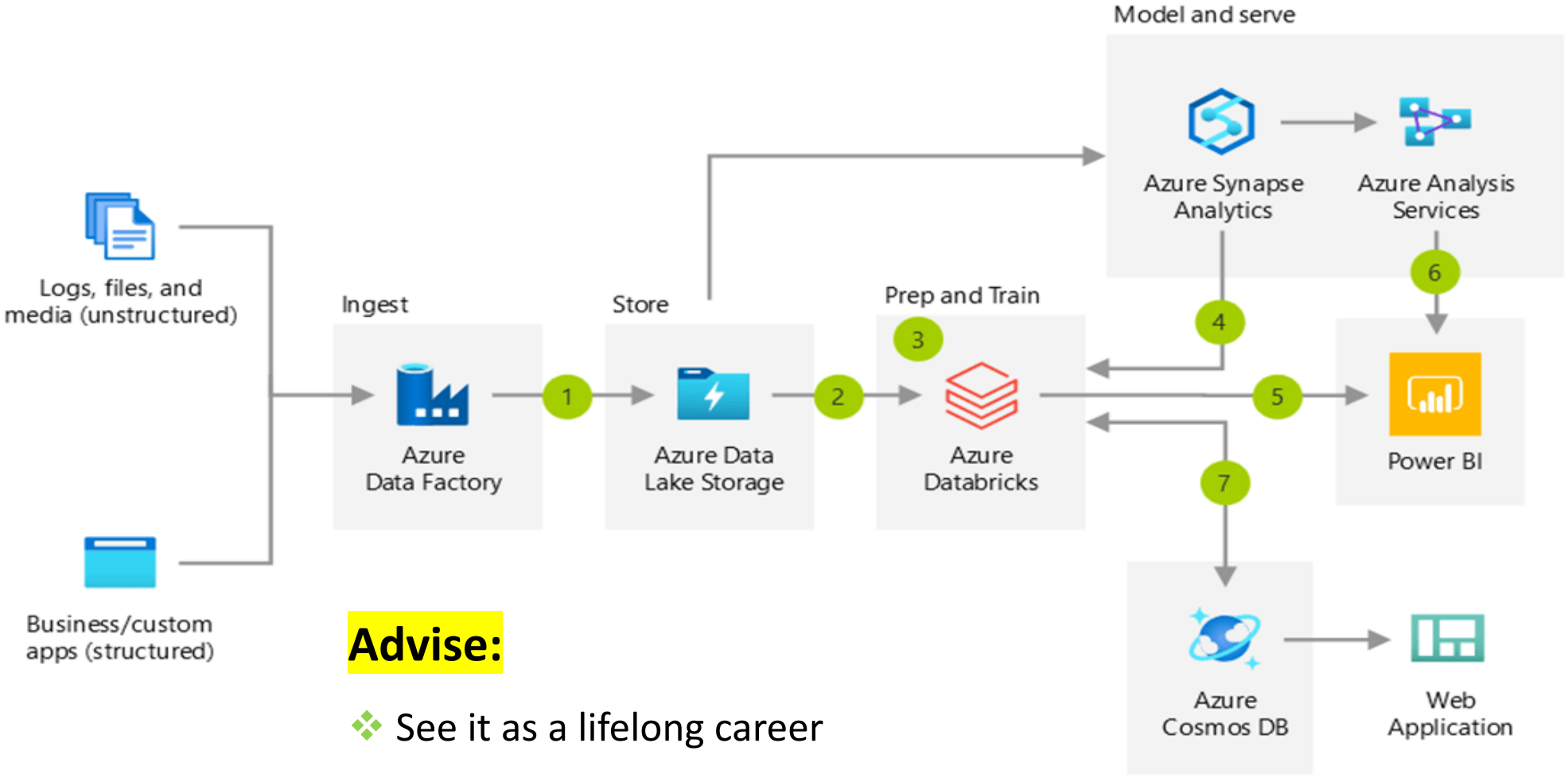
- Between the dawn of time and 2003, five exabytes (1 Bn GB) of data had been created at Google.
- By 2010, this amount of data was being created every **two days**, and by 2021 it was being created every **40 minutes**.
- There are approximately **400,000 bytes** of data for **every grain of sand on earth**
- Companies that make use of customer analytics are 23 times more likely to outperform their competitors in customer acquisition (nine times for retention).
- 75 billion Internet of Things (IoT) devices is estimated to be in the world by 2025

Data Science Careers:

- Data science was identified as the skill with the **largest skill gap**, according to a 2021 report by the World Economic Forum.
- In 2020 the number of data science job listings outstripped the number of people searching for such jobs by a factor of **3 to 1**.
- By 2026, the US Bureau of Labor Statistics estimates data science will create around 11.5 million job openings
- The average salary for a data scientist is **$100,000** CAD according to the Bureau of Labor Statistics, and that of an analyst is **$70,000** CAD.

https://www.datacamp.com/blog/25-interesting-facts-about-data-science

ETL Framework



Advise:

- See it as a lifelong career
- Learn something new daily
- Integrate the skills into your lifestyle

Course Outline

Prepare the Data

- Get data from different sources
- Clean, transform, and load data

Model the Data

- Design a data model
- Develop a data model
- Create model calculations with DAX
- Optimize model performance

Visualize & Analyze the Data

- Design Create reports
- Create dashboards
- Enrich reports for usability
- Enhance reports for usability & storytelling
- Identify patterns & trends

Deploy & Maintain Assets

- Manage files & datasets
- Manage workspaces

Data Analysis Expressions (DAX)

- Syntax
- Operators
- Aggregate Functions
- Logical Functions
- Text Functions
- Rounding Functions
- Filter Functions
- Date & Time Functions

Structured Query Language (SQL)

- Data Query Language (DQL)

THE COURSE PROJECT

THE PROJECT

Congratulations! You've just been hired as the Business Intelligence Analyst for **PassThru Corp.**, a national manufacturing company.

PROJECT SCOPE

The company's senior management needs some measures to make informed decision on how to grow their revenues, make larger profits and increase customer retention rate.

Your assignment is to employ the **complete Microsoft Power BI tools** to design, develop and deploy reports and dashboards for this Team

Data Source: A folder with csv files containing information about sales, products, customers, and store locations

OBJECTIVES

- Prepare the Data
- Model the Data
- Visualize & Analyze the Data
- Deploy & Maintain Assets

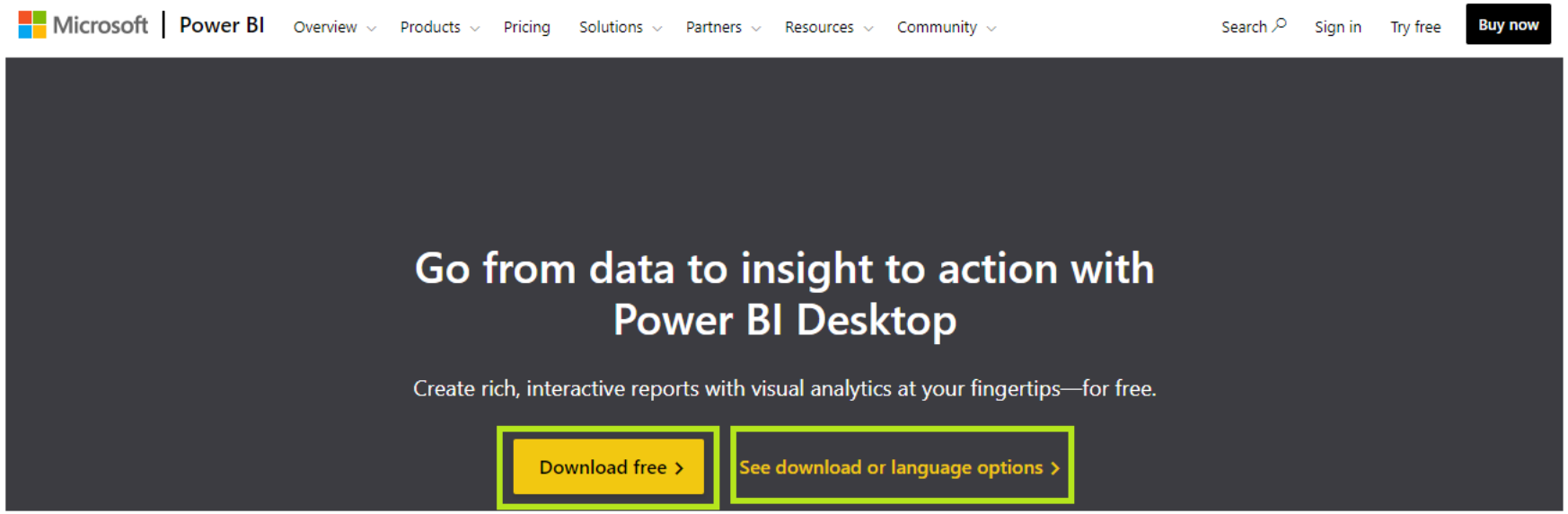
**This data is provided by Microsoft for informational purposes only as an aid to illustrate a concept.*

Introducing Power BI

- A component of Microsoft Power Platform
- Power Platform include Power BI, Power Apps, Power Automate, Power Automate Desktop (PAD), Power Virtual Agents

To Download Power BI:

- Go to **powerbi.microsoft.com/desktop** and click "**Download free**"
- *Use **See download or language options** to update Power BI Desktop*



PREPARING THE DATA

Get data from different sources

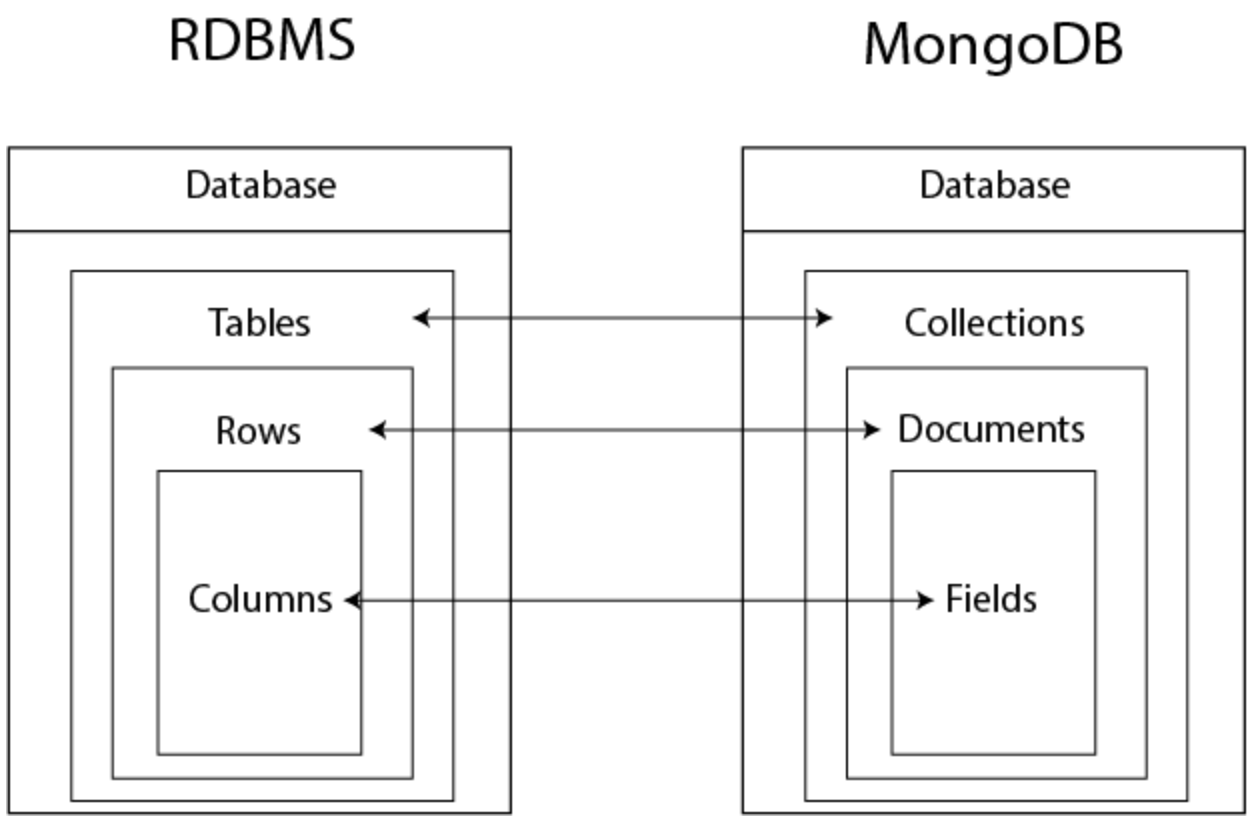
- Connecting to Data
- Data Sources
- Storage Modes
- Parameters
- Data Profiling

Clean, transform and load data

- Cleaning Data
- Identifying Keys
- Shaping Tables
- Combining Queries
- Modifying M Code

Data in Analytics Explained

- Data is a collection of facts such as numbers, descriptions and observations used in decision making.
- A table is a **collection** of related data held in a table format within a **database**. It consists of columns and rows.
- Analytics is the systematic computational analysis of data or statistics. The five areas are **descriptive analytics,** diagnostic analytics, predictive analytics, prescriptive analytics, and cognitive analytics.



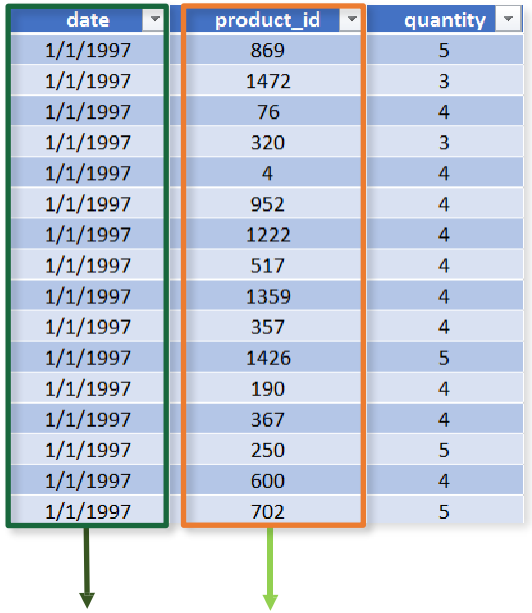
CONTRAINTS AND DATA TYPES

Primary and Foreign Keys

- Primary Key is a column or combination of columns that contain values that uniquely identify each row in the table
- Foreign Key (FK) is a column or combination of columns that is used to establish and enforce a link between the data in two tables

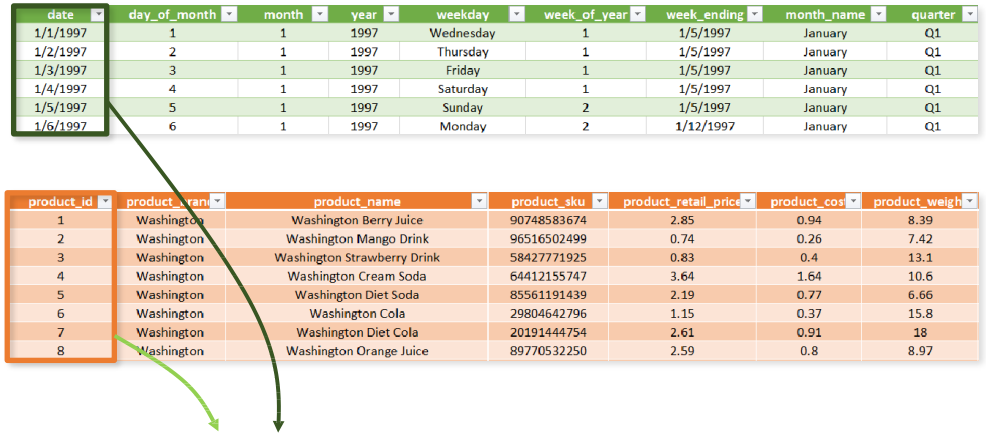
Data Type

- A data type is an attribute that specifies the type of data that the object can hold such as integer data, character data, monetary data, date and time data, binary strings, and so on.



date	product_id	quantity
1/1/1997	869	5
1/1/1997	1472	3
1/1/1997	76	4
1/1/1997	320	3
1/1/1997	4	4
1/1/1997	952	4
1/1/1997	1222	4
1/1/1997	517	4
1/1/1997	1359	4
1/1/1997	357	4
1/1/1997	1426	5
1/1/1997	190	4
1/1/1997	367	4
1/1/1997	250	5
1/1/1997	600	4
1/1/1997	702	5

These columns are **foreign keys**; they contain *multiple* instances of each value, and are used to match the **primary keys** in related lookup tables



date	day_of_month	month	year	weekday	week_of_year	week_ending	month_name	quarter
1/1/1997	1	1	1997	Wednesday	1	1/5/1997	January	Q1
1/2/1997	2	1	1997	Thursday	1	1/5/1997	January	Q1
1/3/1997	3	1	1997	Friday	1	1/5/1997	January	Q1
1/4/1997	4	1	1997	Saturday	1	1/5/1997	January	Q1
1/5/1997	5	1	1997	Sunday	2	1/5/1997	January	Q1
1/6/1997	6	1	1997	Monday	2	1/12/1997	January	Q1

product_id	product_brand	product_name	product_sku	product_retail_price	product_cost	product_weigh
1	Washington	Washington Berry Juice	90748583674	2.85	0.94	8.39
2	Washington	Washington Mango Drink	96516502499	0.74	0.26	7.42
3	Washington	Washington Strawberry Drink	58427771925	0.83	0.4	13.1
4	Washington	Washington Cream Soda	64412155747	3.64	1.64	10.6
5	Washington	Washington Diet Soda	85561191439	2.19	0.77	6.66
6	Washington	Washington Cola	29804642796	1.15	0.37	15.8
7	Washington	Washington Diet Cola	20191444754	2.61	0.91	18
8	Washington	Washington Orange Juice	89770532250	2.59	0.8	8.97

These columns are **primary keys**; they *uniquely* identify each row of a table, and match the **foreign keys** in related data tables

DATA SOURCES

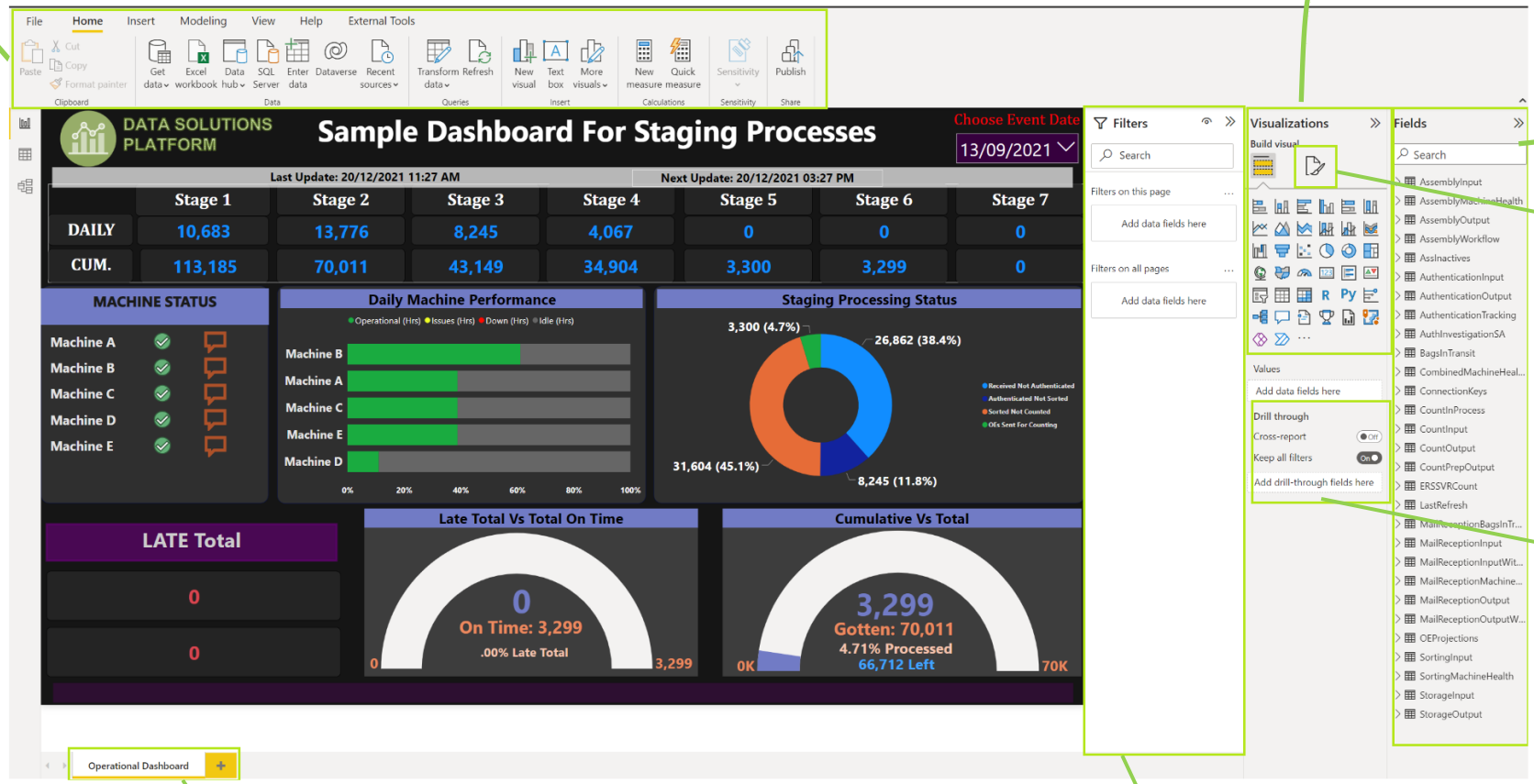
Power BI data sources categories

- Flat files & Folders (csv, text, Excel, JSON etc.)
- Databases (SQL, Access, Oracle, IBM, MySQL, etc.)
- Power Platform (Dataflows, Power BI datasets, Dataverse, etc.)
- Azure (SQL Database, Blob Storage, Cosmos DB, Data Lake Storage, etc.)
- Online Services (SharePoint, GitHub, Dynamics 365, Google Analytics, Salesforce, etc.)
- Others (Web feeds, R scripts, Spark, Hadoop, etc.)

Connecting to Data / Power BI Desktop Overview

Menu/Home Options
Data, Queries, Insert, Calculations, Sensitivity, Share.

VISUALIZATION OPTIONS
Charts, Slicers, Maps, Matrices, etc.



FIELD LIST
Tables, Columns, Measures

FIELDS/FORMAT/ ANALYICTS PANE
Visual-specific configuration & formatting tools

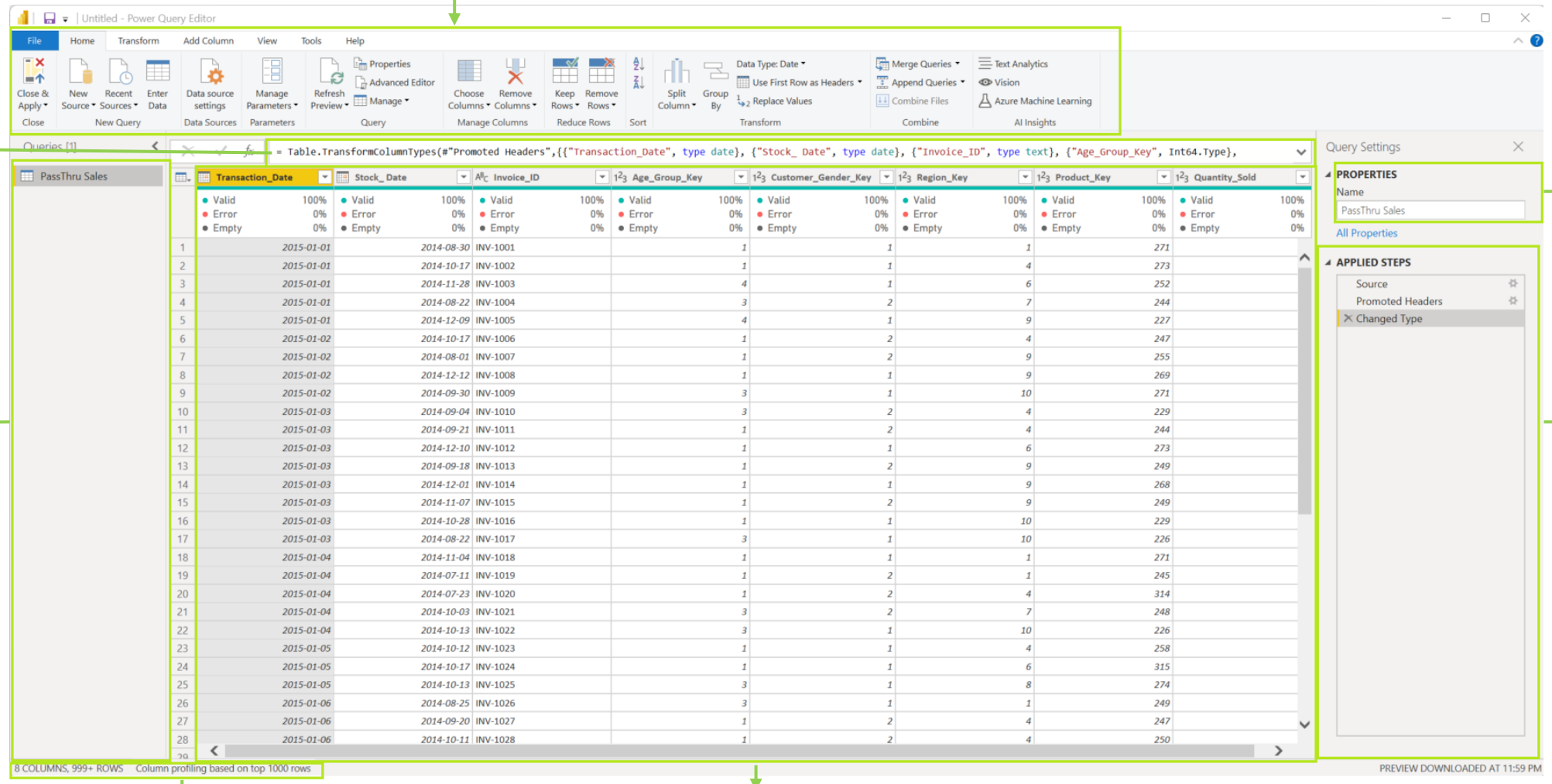
DRILL THROUGH FILTERS
Options for page-level drill through filters

REPORT PAGES
Similar to Excel tabs; each is a blank reporting canvas

FILTERS PANE
Visual-Level, Page-Level, and Report-Level Filters

Connecting to Data / Query Editor Overview

QUERY EDITING TOOLS
Table transformations, calculated columns, etc.



FORMULA BAR
This is "M" code

QUERY PANE

TABLE NAME & PROPERTIES

APPLIED STEPS
Like a macro!

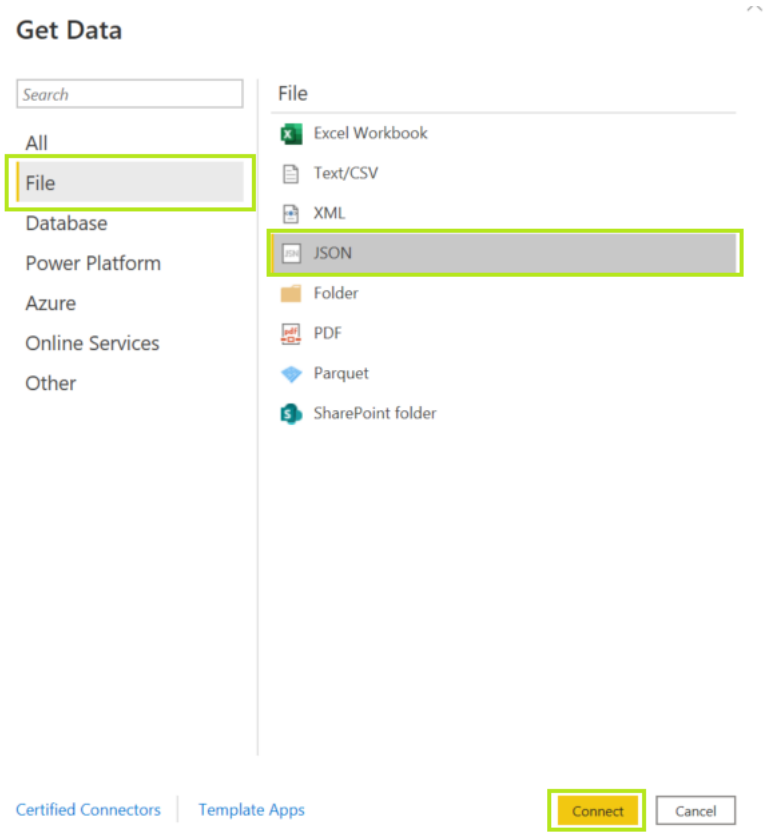
COLUMN PROFILE
Based on first 1,000 rows

DATA VIEW

DATA SOURCE: JavaScript Object Notation (JSON)

Importing a JSON File

- Convert the JSON list to a table
- Expand the attributes in the list to columns
- Change the data type for each column

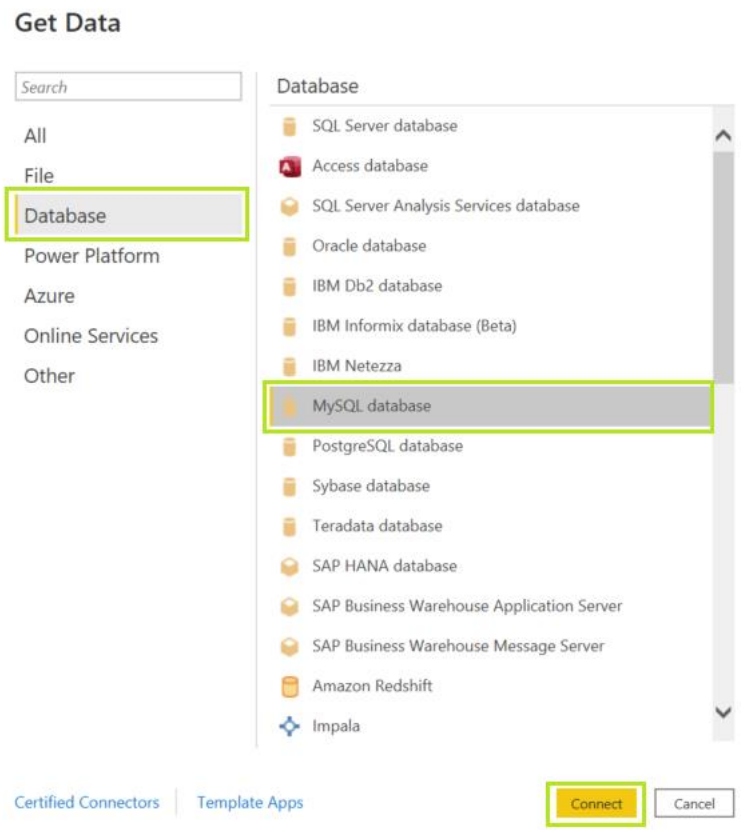


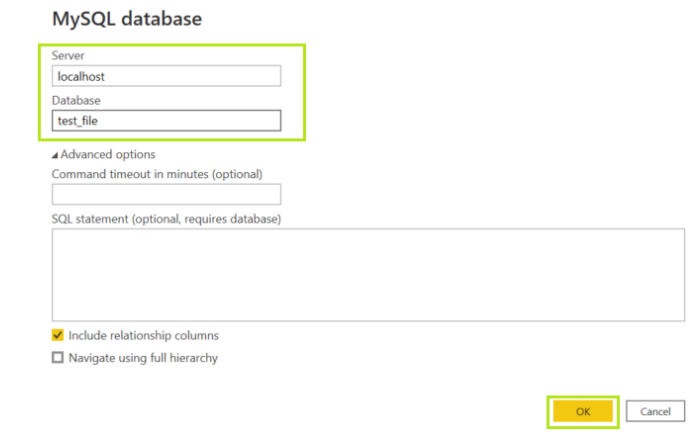
For more detail visit: ***https://docs.microsoft.com/en-us/power-query/connectors/json***

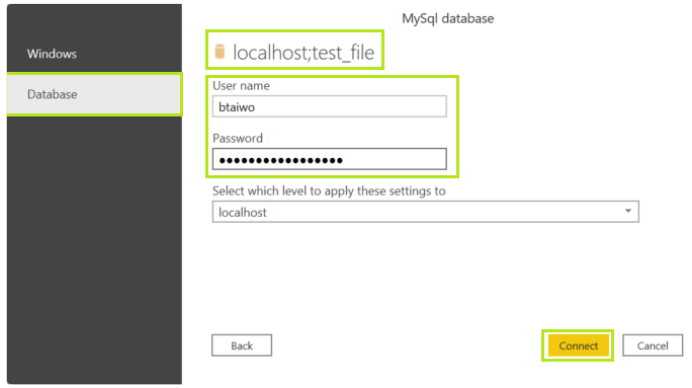
DATA SOURCE: SQL SERVER DATABASE

Importing data from RDB

- Select the right SQL Database
- Add the server and schema information and connect
- Enter the correct user credentials
- Select the interested tables and load into Power BI







For more detail visit: ***https://docs.microsoft.com/en-us/power-query/connectors/json***

STORAGE MODES

Power BI lets you choose between these types of **storage modes** for your data sources:

- **Import:** Tables stored in-memory within Power BI and queries are fulfilled by cached data
- **DirectQuery:** Tables connected directly to the source & queries executed on-demand at the source
- **Dual:** Tables come from in-memory data or by an on-demand query to the data source

Use **DirectQuery** when:

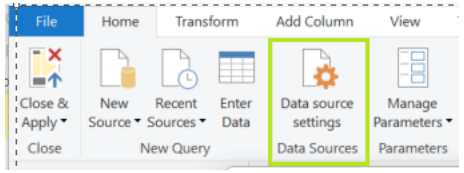
- Dataset is too large to be stored in-memory
- Source data changes frequently & reports must show the most recent data
- Company policy states data can only be accessed from the original source

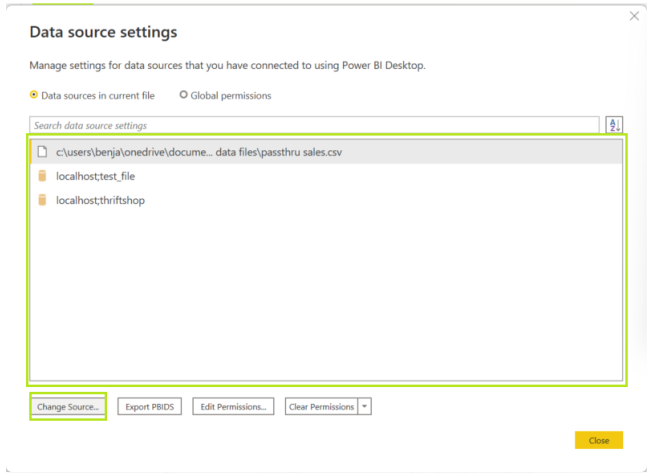
	Imported Data	DirectQuery
Performance	-Best	-Depends on the data source. Generally, slower compared to imported data
Number of Data Sources	-Unlimited	-Unlimited
Data Transformations	-No M transform restrictions	-Limited M transforms -Transforms based on data source language
Data Modeling	-No restrictions on data modeling	-Very restricted -Limited DAX & no calculated tables -Quick Insights not supported
Data Model Size	-Loaded in-memory (increases model size) -Pro: 1GB per dataset -Premium: Capacity based	-Large/frequent volume of data -Does not increase model size -Limited by data source hardware
Data Refresh	-Data only current to last refresh -Pro: 8x per day 30-min intervals -Full refreshes are “expensive”	-Near real-time -Report always shows latest data available
Row-level Security	-User-level role definitions	-User-level role definitions only available for some data sources

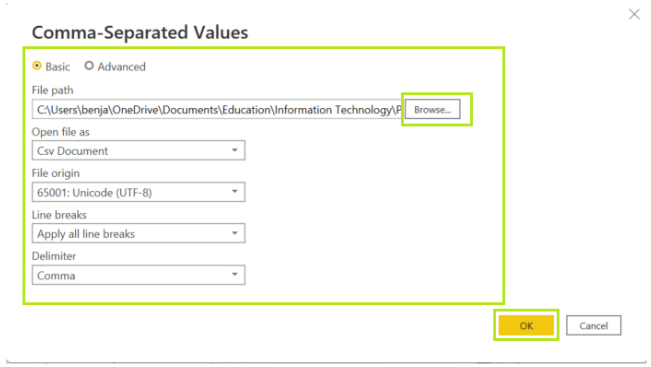
DATA SOURCE SETTINGS

Managing data connections and permissions

- Connections to local files reference the exact path If the file name or location changes, you will need to change the source and browse to the current version

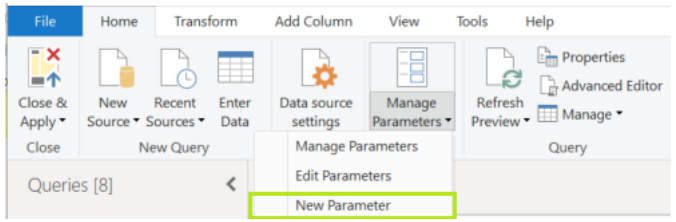


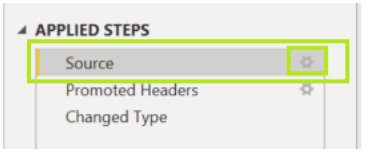


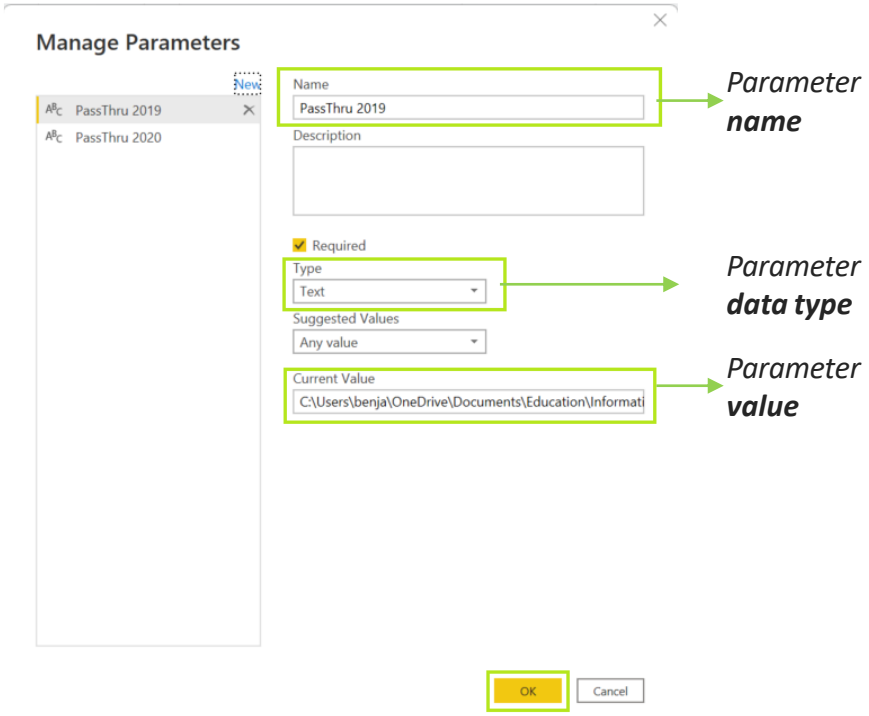


DYNAMIC SOURCES WITH PARAMETERS

Parameters are a useful way to change data source values dynamically in Power Query



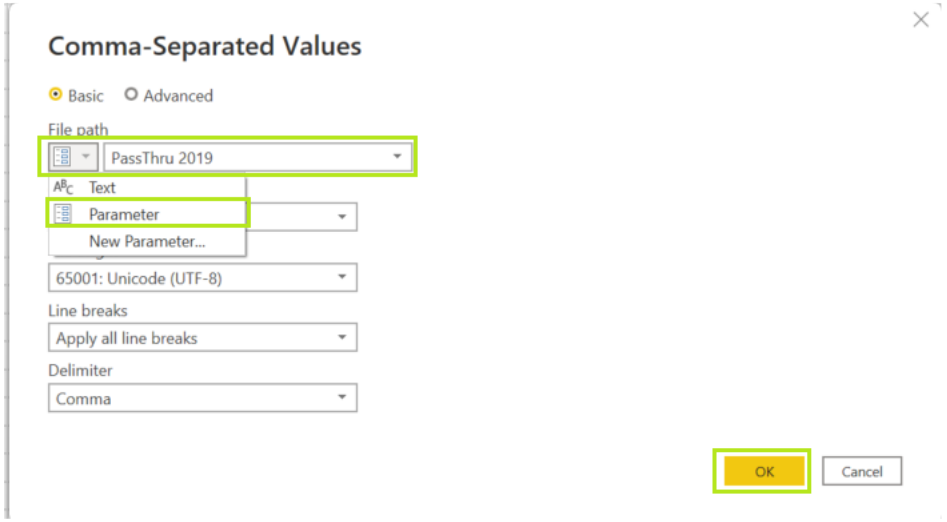




*Parameter **name***

*Parameter **data type***

*Parameter **value***



More Details on the Platform

- Class starts fully in October
- Duration is 8 weeks with ongoing coaching (Your success is our success)
- Cost: $400
- Payment: Interact to datasolutionsplatform@gmail.com
- People in Nigeria – Contact me on WhatsApp for account details.

Next Class Expectation

- DP-900 Exam should be scheduled
- Payment should be completed
- Rolling Calendar Assignment should be completed
- We are working on the PL-300 exam discount.

THANK YOU FOR LISTENING

Q & A