

# Data Visualisation with PowerBI and Tableau

---

Gazal



# Common Tools

- Anaconda Python 3 Distribution
- Jupyter Notebooks
- Matplotlib and Seaborn
- NumPy and Pandas



# DATA

Data is a collection of various facts, such as measurements, observations, and figures, that have been transformed into a language that computers can understand.

## IMPORTANCE OF DATA

Understanding problems and positions of organization / people

Understanding needs of an organization or people

Knowing nature of customer, employee, market etc....

Decision Making

Solving real world problems

Improving process(fast and less expensive)

Keep track of all

Best utilization of resources available

Finding performance of a company or group of people

# Matplotlib and Seaborn

- Very popular plotting libraries
- Matplotlib API is low-level
- Seaborn wrapper offers high-level graphics
- Adjust final plot output with Matplotlib's syntax

There are a lot of great reasons to use Python to build data visualizations.

## Many types of visualizations

- Boxplots
- Bar graphs
- Line graphs
- Histograms

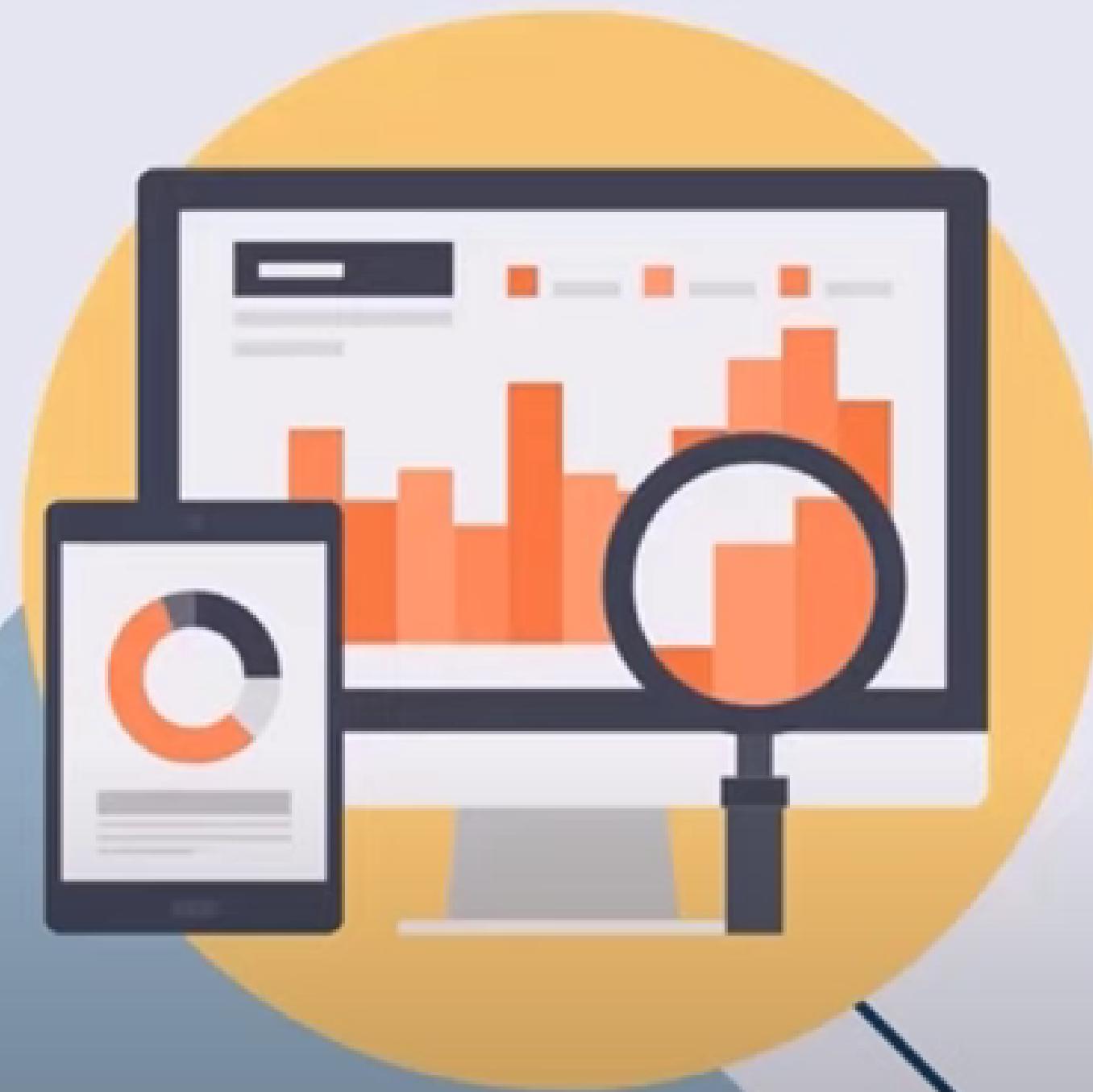
## You can quickly iterate

- Improve a visualization
- Try out many visualizations

# Visualize your data with the power of Tableau.

## Tableau File Types

- **Tableau packaged workbook (.twbx)**
  - Stores extracted data and visualizations for viewing in any full version of Tableau
- **Tableau workbook (.twb)**
  - Stores a visualization without source data
- **Tableau datasource (.tds)**
  - Stores the server address, password, and other information required to access a data source



# Tableau

## Visual Analytics Platform

Founded in 2003 as a result of a computer science project at Stanford that aimed to improve the flow of analysis and make data more accessible to people through visualization.

# PowerBI

## Syllabus

- Program Orientation & Introduction to Data Visualization
- Introduction to Power BI & Power BI Desktop
- Working with Query Editor & Data Cleaning Data Transformation
- Creating Visuals in the Report View
- Live Case Study
- Working with other Data Sources & Creating Custom Visuals (Webscraping)
- Understanding the Data Model: Data & Relationships Live
- DAX Functions Live
- Relational Functions, Context, Measures, Calculate, Iterator
- Live Case Study



# Roles in Data

## Data Administrator

They figures out which data is relevant to be stored in data models. Its less of a technical role and more of a business role with some technical knowledge.

## Data Analyst

Responsible for capturing, cleaning, modifying and transforming data and creating dashboards and reports, using visualization tools.

## Data Engineer

Manages secured flow of structured and unstructured data from multiple sources.

## Business Analyst

Interprets business data that comes from visualization. He knows pulse of business unlike data analyst.

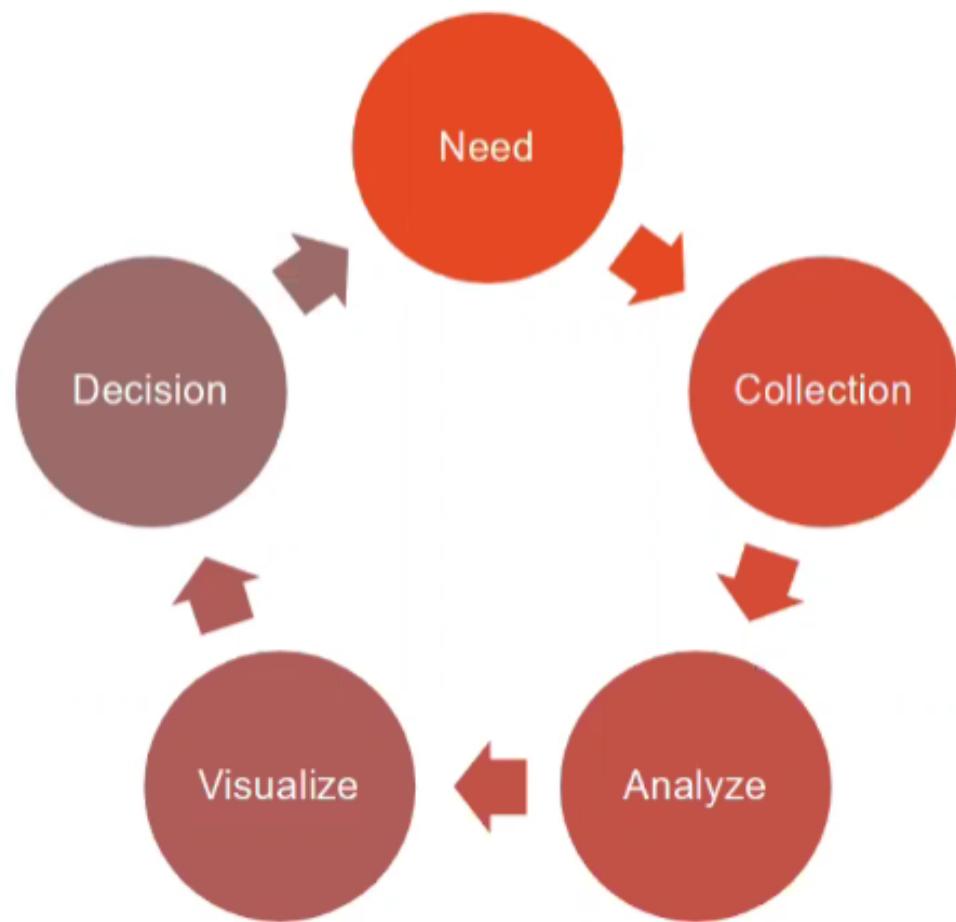
## Data Scientist

Performs advanced analytics that business analyst can't do and extracts value from data. Provides valuable insights using advanced mathematics and technologies like machine learning, deep learning

# JOB ROLES IN POWER BI & CORRESPONDING AVERAGE SALARY

1. Power BI Manager - ₹ 21.5 Lakhs
2. Power BI Project Manager - ₹ 14.1 Lakhs
3. Power BI Administrator - ₹ 8.6 Lakhs
4. Power BI Consultant - ₹ 6.3 Lakhs
5. Power BI Semantic Model Developer - ₹ 5.9 Lakhs
6. Power BI Developer - ₹ 5.8 Lakhs
7. Data Analyst - ₹ 4.3 Lakhs

# Business Intelligence



## Business Intelligence

Business intelligence (BI) refers to a collection of procedures, frameworks, and tools that transform unprocessed data into actionable knowledge that helps businesses operate profitably.

BI encourages the use of historical data to assist fact-based decision making rather than assumptions and intuition.

In order to give users in-depth insight about the nature of the business, BI tools analyze data and produce reports, summaries, dashboards, maps, graphs, and charts.

# Importance of Business Intelligence

- Creating KPIs (Key Performance Indicators) based on historical data
- Determine and establish standards for various procedures.
- Businesses can notice issues that need to be resolved by using BI systems to identify market trends.
- Data visualization, which is aided by BI, improves data quality and, as a result, decision-making quality.
- BI systems can be utilized by SME's as well as large businesses (Small and Medium Enterprises)

# POWER BI DESKTOP

Menu Bar

Sign in

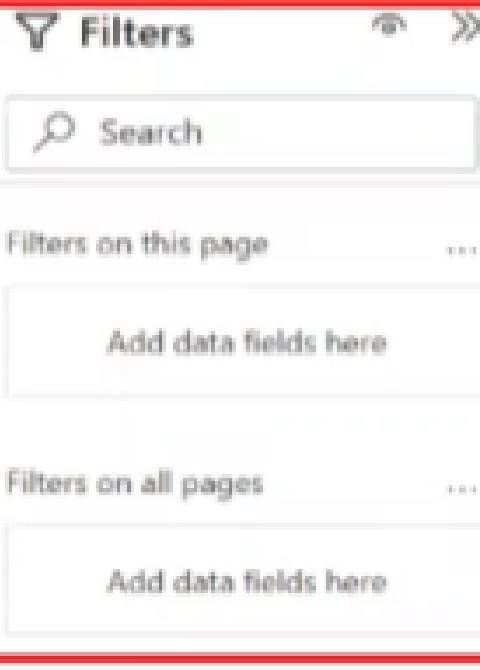
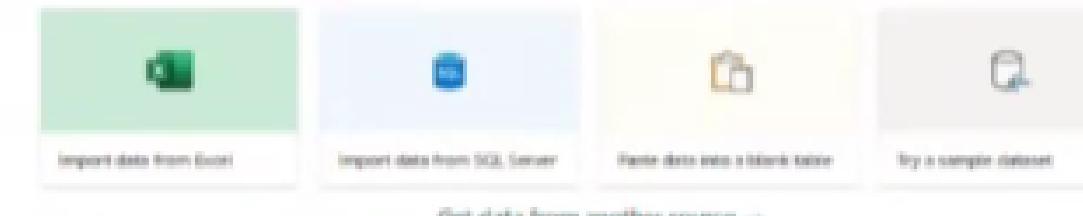
File Home Insert Modeling View Help



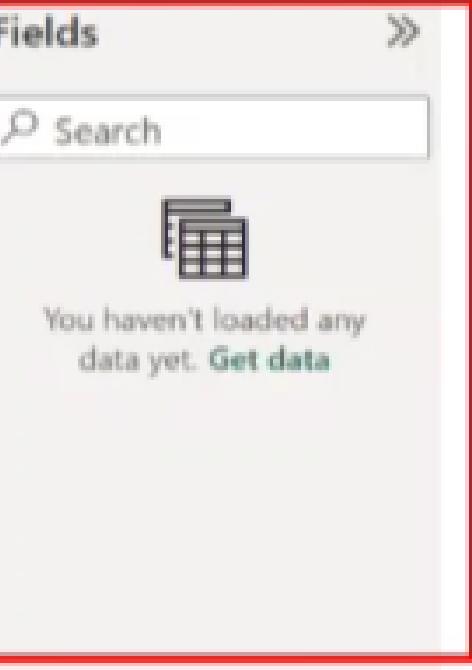
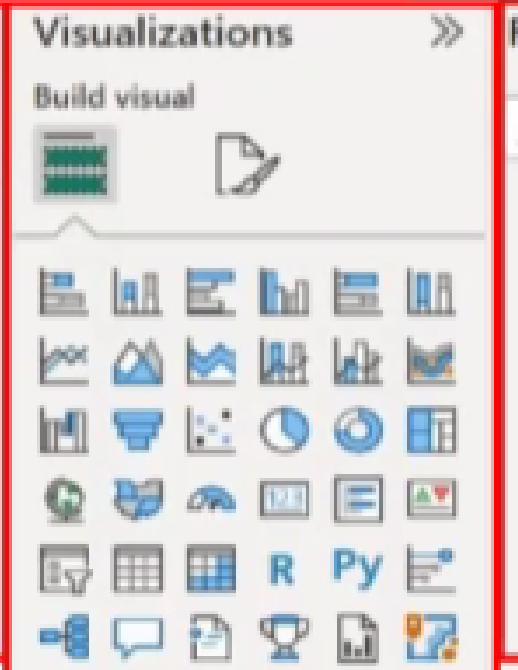
Ribbon



Views



Filter



Visualizatio  
n

Pages

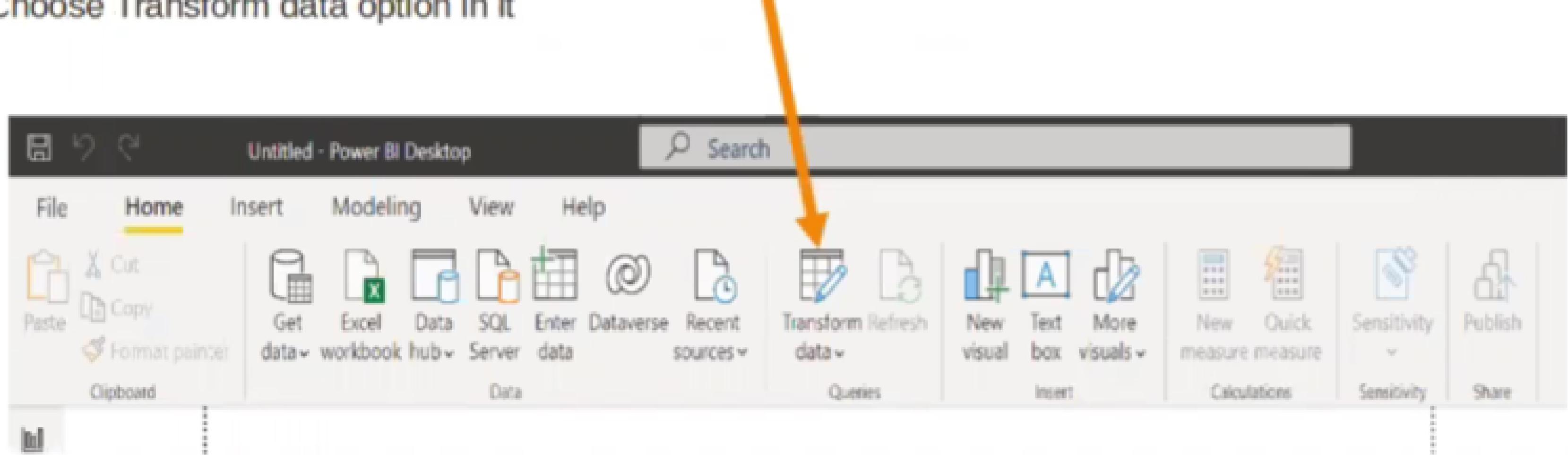
Page 1 +

# PARTS OF POWER BI DESKTOP

- Ribbon - The majority of the controls and options required for creating the report are located in the upper ribbon.
- Canvas – Its primary design area, where visualizations and other elements are applied.
- Page Selector – Used for navigation to different pages of the report.
- Filters - To further filter the data, fields can be added here.
- Visualization – Consists list of available visualizations.
- Fields - The tables and fields that are present in the data model are contained in this section.
- Views – It consists of the model view, the data view, and the report view.

# HOW TO ACTIVATE POWER QUERY?

- Go to PowerBI desktop
- Go to home tab
- Choose Transform data option in it



# COMPONENTS OF POWER QUERY

**Ribbon options**

**Home Ribbon**

**Query Pane**

**Formula Bar** Table.TransformColumnTypes(#"Promoted Headers",{{"For Update List Visit: ",

#	All: For Update List Visit:	All: Column2	All: Column3	All: Column4
1	www.adithyanilcode.com	null	null	null
2	State	District	State Type	
3	Andaman Nicobar	Nicobar	Union Territory	
4	Andaman Nicobar	North Middle Andaman	Union Territory	
5	Andaman Nicobar	South Andaman	Union Territory	
6	Andhra Pradesh	Anantapur	State	
7	Andhra Pradesh	Chittoor	State	
8	Andhra Pradesh	East Godavari	State	
9	Andhra Pradesh	Aldori Ghatamam Raju	State	
10	Andhra Pradesh	Anakapalli	State	
11	Andhra Pradesh	Annamayya	State	
12	Andhra Pradesh	Bapatla	State	
13	Andhra Pradesh	Eluru	State	
14	Andhra Pradesh	Guntur	State	
15	Andhra Pradesh	Kadapa	State	

**Preview**

**Query Settings**

**PROPERTIES**

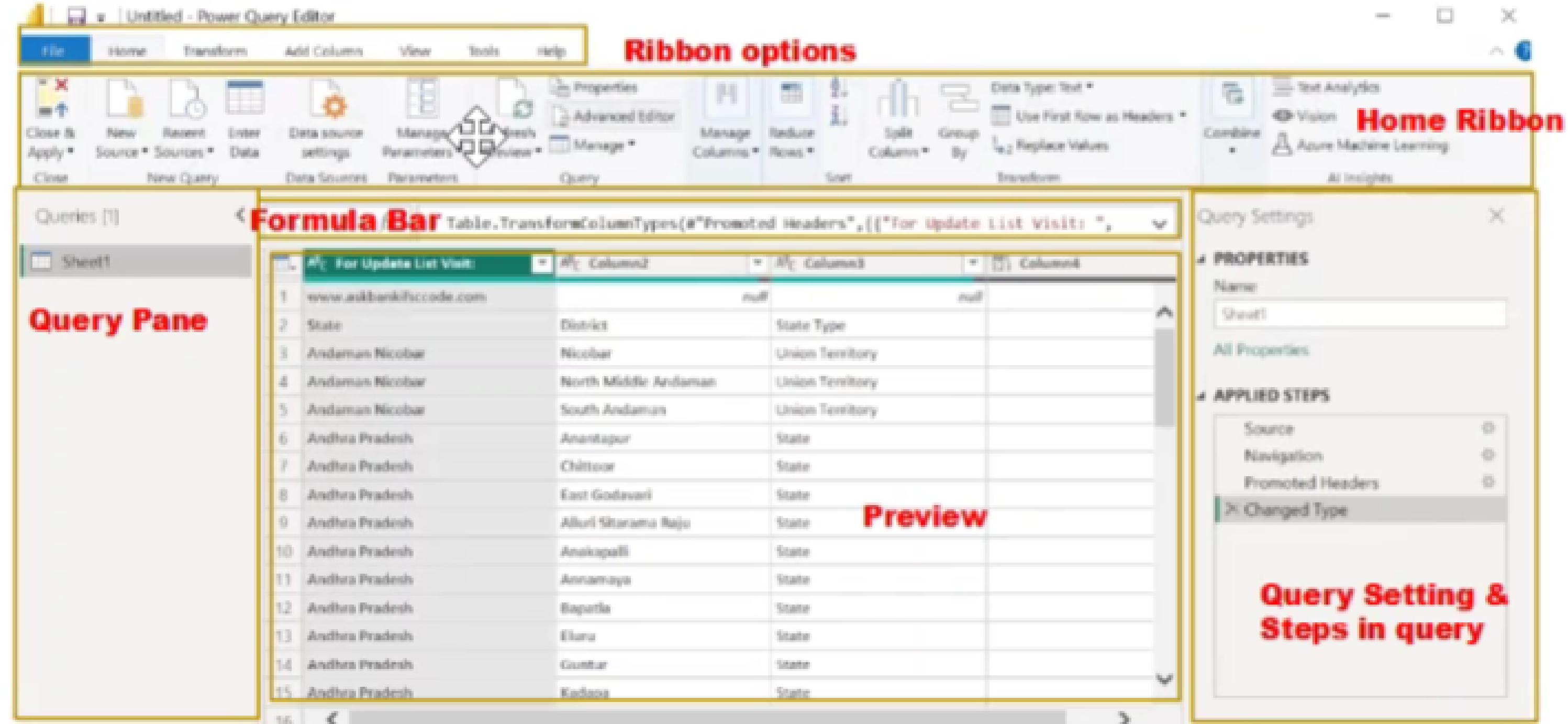
Name: Sheet1

All Properties

**APPLIED STEPS**

Source  
Navigation  
Promoted Headers  
**Changed Type**

**Query Setting & Steps in query**



# COMPONENTS OF POWER QUERY

## Ribbon

- Home – Consists of common items and operations that are frequently used.
- Transform – To change the nature of data
- Add Column - To change shape of data

**Data view:** It looks almost like Excel spreadsheet view.

**Preview** - Data table preview

**Queries** – All queries are listed here

**Formula bar :** We can enable / disable this formula bar from "View ribbon".

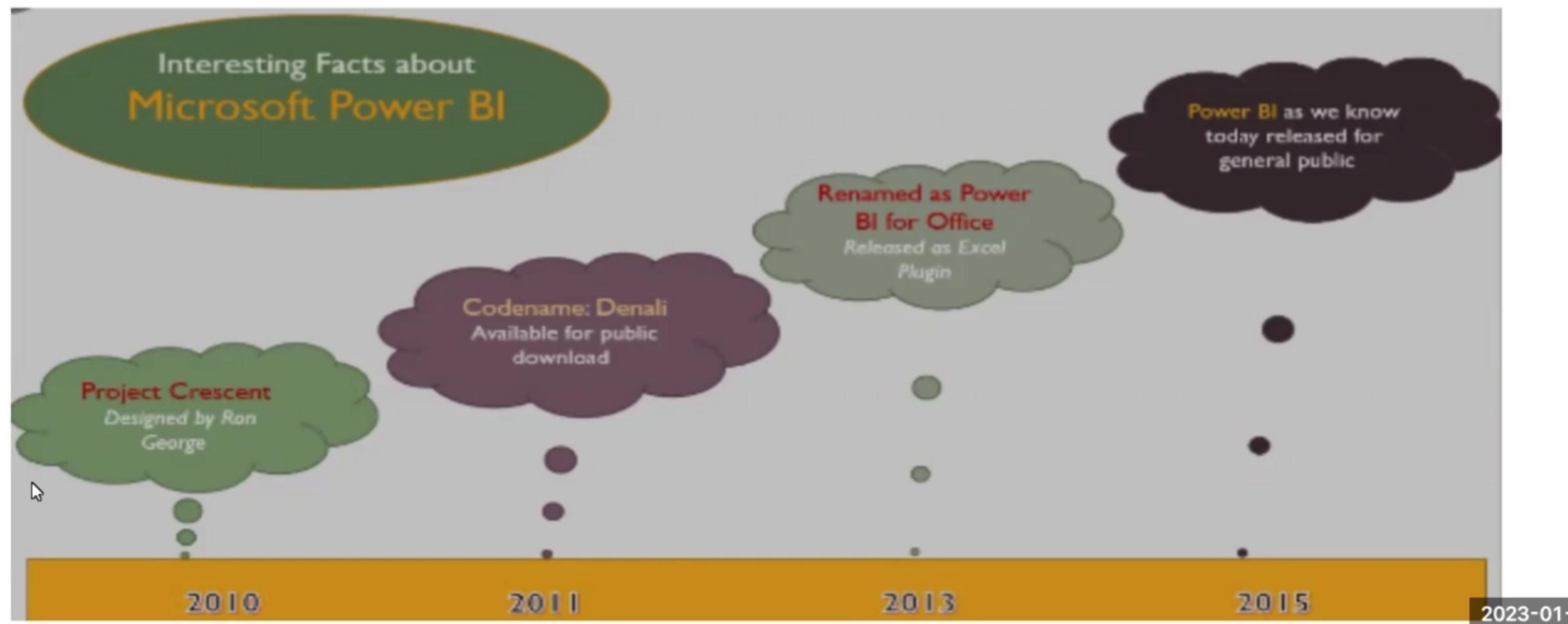
**Close & Apply button:** Used to save all changes made in PowerBI.

## PROPERTIES OF POWER QUERY

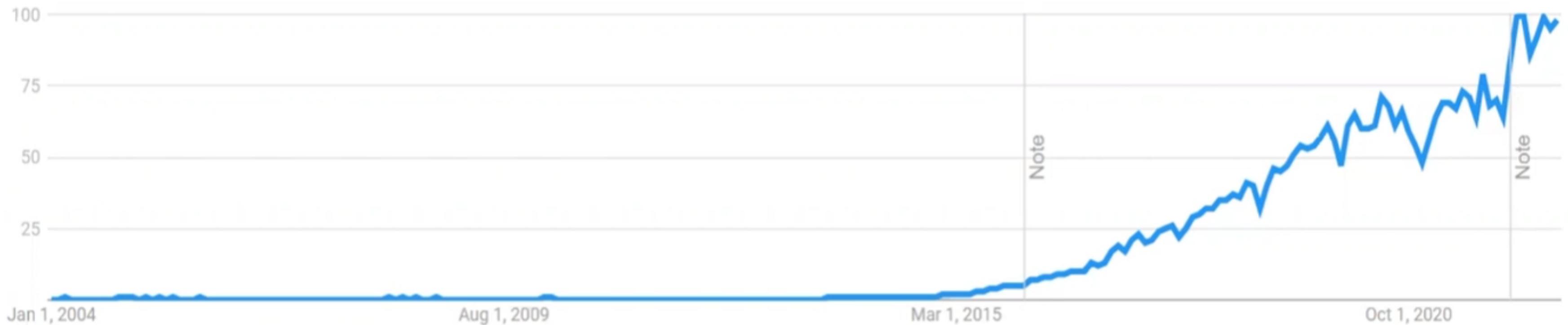
- Its case sensitive
- It's a step recorder & usually one step depends on previous step
- Data source that can be connected – excel, csv, webpage, xml, json, API, databases, cloud sources, folders, pdf etc.....

# FACTS

- It was originally designed by **Ron George** in the summer of 2010 and named Project Crescent.



# OTHER IMPORTANT POINTS OF POWER BI



- 1200 different functions available
- 280 visuals
- Least expensive compared to other BI tools
- Can accept data from 100 different data sources
- Recognized by Gartner Company (which compares all BI products)

# Advantages of using BI tools

- Businesses may generate reports with only one click, which saves a ton of time and resources. Additionally, it enables workers to complete their responsibilities more effectively.
- BI aids in locating any company areas that require attention.
- Through common BI features like dashboards and scorecards, BI systems also assist organizations' decision-makers in gaining a broad perspective.
- By providing predictive analysis, computer modeling, benchmarking, and other approaches, it also automates analytics.
- It enables users, regardless of technical or analytical background, to swiftly gather and process data. This makes it possible for many people to access the power of analytics.

## Disadvantages of using BI tools

- For many small and medium-sized businesses, BI systems are still out of reach and costly.
- Business practices may become stiff when dealing with it because it can be so complex.
- Data warehousing system implementation takes almost one and a half years to complete. Due to this, the process takes a long time.

# Gartner Magic Quadrant

Gartner's IT consulting company produces a series of market research studies called Magic Quadrant (MQ) that use exclusive qualitative data analysis techniques to show industry trends like direction, maturity, and participation. Their studies are carried out for several distinct technology businesses, and they are updated every one to two years.

Vendors are rated by Gartner based on : *completeness of vision and ability to execute.*

*Completeness of vision* - The completeness of the vision reflects the originality and innovation of the seller and if the vendor is leading or following the market.

*Ability to Execute* - Highlights elements including the vendor's financial stability, responsiveness to the market, product development, sales channels, and client base.

# BI tools in 2015



# BI tools in 2016



# BI tools in 2017

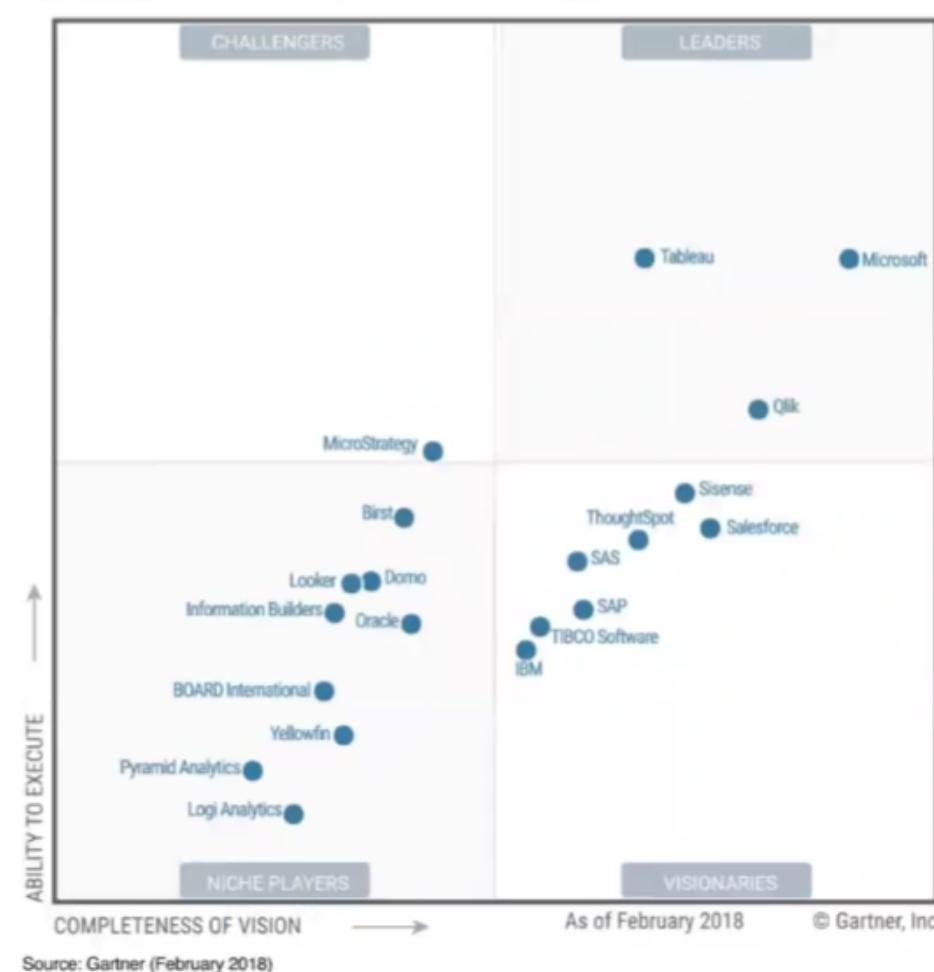


# BI tools in 2018

# BI tools in 2019

# BI tools in 2020

Figure 1. Magic Quadrant for Analytics and Business Intelligence Platforms



# BI tools in 2022



Power BI has been recognized by Gartner as the market's most complete and potent platform for the fourth year in a row.

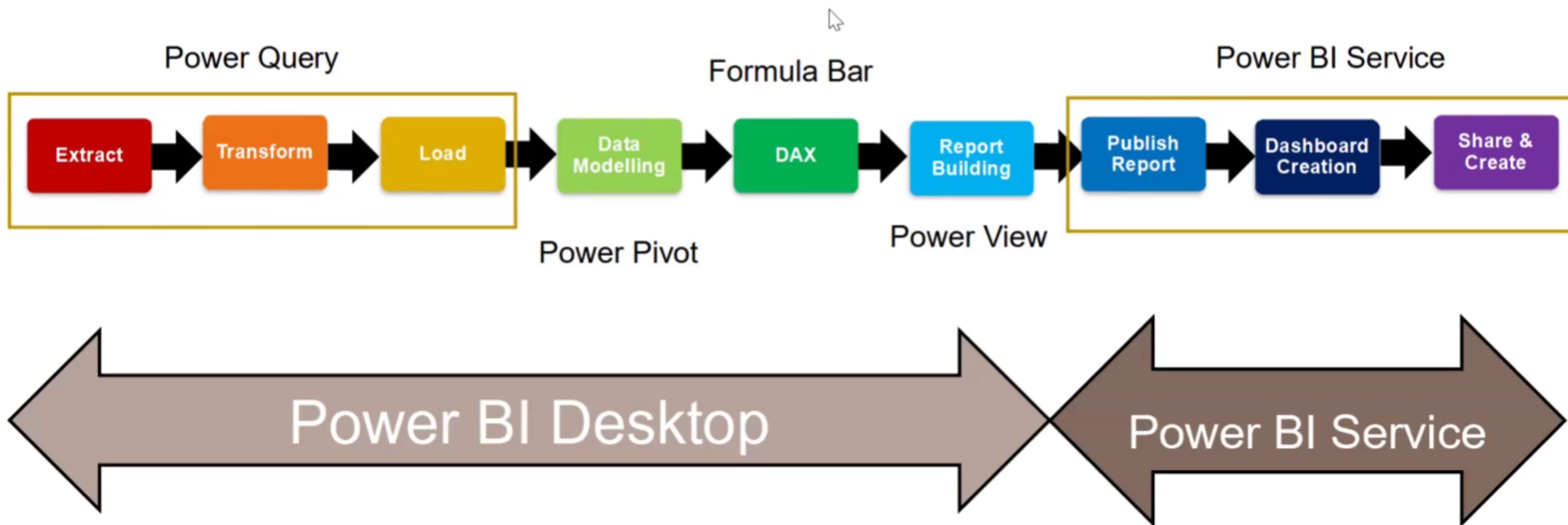
# POWER BI

- Power BI – Business intelligence tool, used to clean and convert data into visualization format like dashboards and reports
- Its business analytics service provided by Microsoft
- Power BI Desktop - Application where we do visualization

## Power BI tools

- Power Query – Cleans data
- Power Pivot – Data modelling, create relation between multiple data sources
- Power View – To create 250 + charts
- Power BI Service – To share files and dashboards





# ODATA

- The SAP (**System Analysis Program**) Data Services program delivers reliable, timely, and relevant information to improve business outcomes. It also integrates and transforms data.
- O Data – Open Data Protocol
- O Data is a protocol that SAP is using to make SAP data accessible to the world so that even developers who don't know SAP lingo can use this data and develop Web applications, websites and Mobile Apps.
- The Open Data Protocol (OData) is a web-based data access protocol that is based on established standards like HTTP.
- The location of the Northwind OData feed : <https://services.odata.org/v2/northwind/northwind.svc/>



# INTRODUCTION

- A collection of one or more tables and, optionally, relationships makes up a Power BI data model.
- A well-designed data model makes it possible for business users to comprehend, examine, and get insights from their data.
- This action should be done by loading your data and defining the relationships between tables before you build any visualizations.
- Data modeling frequently happens in the initial stages of creating a Power BI report.

# NORMALIZED DATA

- *Normalization* is the term used to describe data that's stored in a way that reduces repetitive data.
- A sales table is considered normalized when it stores only keys, like the product key.

OrderNumber	ProductKey	Sum of OrderQuantity
SO48795	375	1
SO48796	375	1
SO48797	385	1
SO48798	369	1
SO48799	352	1
SO48800	342	1
SO48801	326	1
SO48802	383	1
SO48803	383	1
SO48804	356	1
SO48805	330	1
SO48806	340	1
SO48807	324	1
SO48808	368	1
SO48809	369	1
SO48810	369	1

# DENORMALIZED DATA

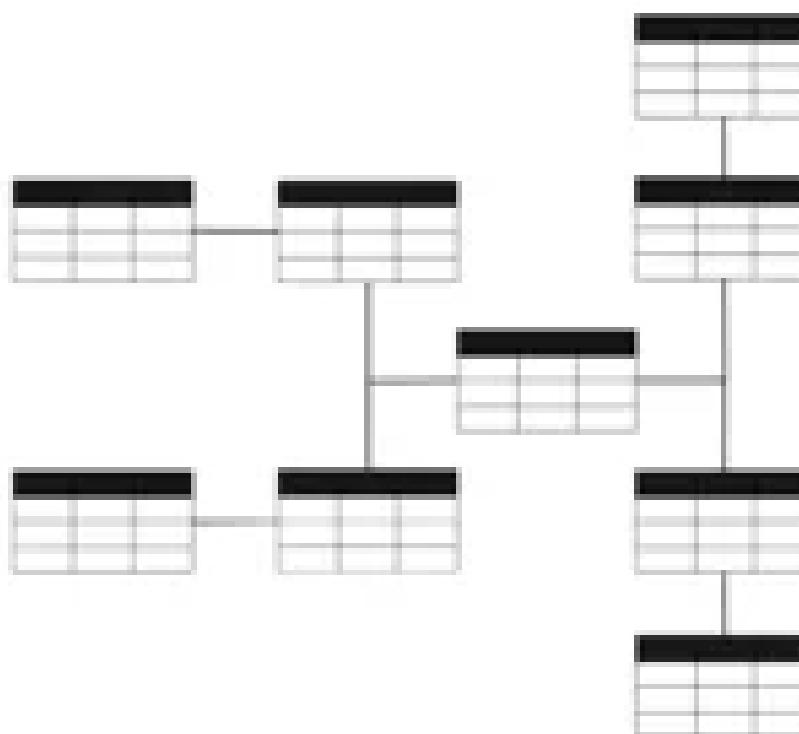
- But if the sales table stores product details beyond the key, it's considered *denormalized*.

OrderNumber	ProductKey	Sum of OrderQuantity	ProductName	ProductSubcategoryKey	ProductDescription
SO48795	375	1	Road-250 Black, 48	2	Aluminum-alloy frame provides a light, stiff
SO48796	375	1	Road-250 Black, 48	2	Aluminum-alloy frame provides a light, stiff
SO48797	365	1	Road-550-W Yellow, 42	2	Same technology as all of our Road series b
SO48798	369	1	Road-250 Red, 48	2	Aluminum-alloy frame provides a light, stiff
SO48799	352	1	Mountain-200 Silver, 38	1	Serious back-country riding. Perfect for all k
SO48800	342	1	Road-650 Black, 52	2	Value-priced bike with many features of our
SO48801	326	1	Road-650 Red, 44	2	Value-priced bike with many features of our
SO48802	363	1	Road-550-W Yellow, 40	2	Same technology as all of our Road series b
SO48803	363	1	Road-550-W Yellow, 40	2	Same technology as all of our Road series b
SO48804	356	1	Mountain-200 Silver, 46	1	Serious back-country riding. Perfect for all k
SO48805	330	1	Road-650 Red, 52	2	Value-priced bike with many features of our
SO48806	340	1	Road-650 Black, 48	2	Value-priced bike with many features of our
SO48807	324	1	Road-650 Red, 62	2	Value-priced bike with many features of our
SO48808	368	1	Road-250 Red, 44	2	Aluminum-alloy frame provides a light, stiff
SO48809	369	1	Road-250 Red, 48	2	Aluminum-alloy frame provides a light, stiff
SO48810	369	1	Road-250 Red, 48	2	Aluminum-alloy frame provides a light, stiff
SO48811	334	1	Road-650 Black, 60	2	Value-priced bike with many features of our
SO48812	356	1	Mountain-200 Silver, 46	1	Serious back-country riding. Perfect for all k
-----	---	-----	-----	-----	-----

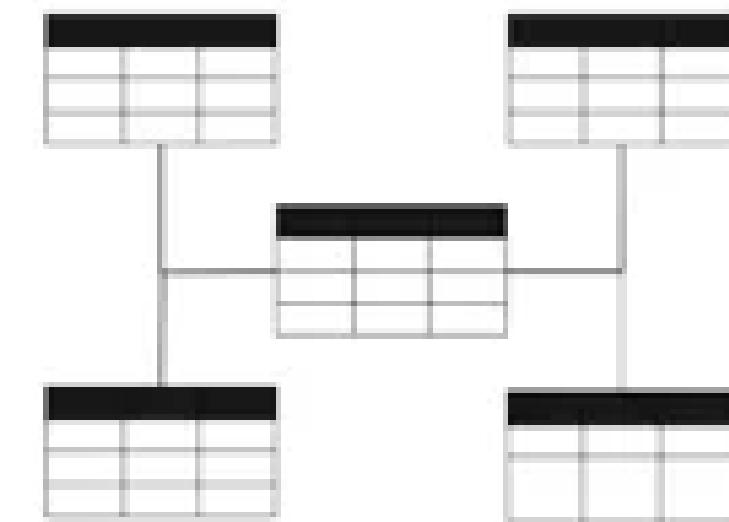
# TYPES OF DATA MODELS

- Flat (fully denormalized) schema
- Star schema
- Snowflake schema

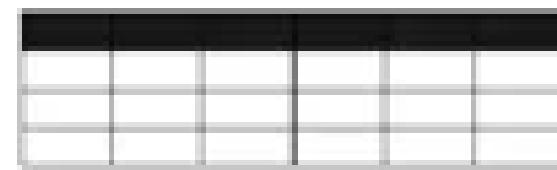
Snowflake Schema



Star Schema

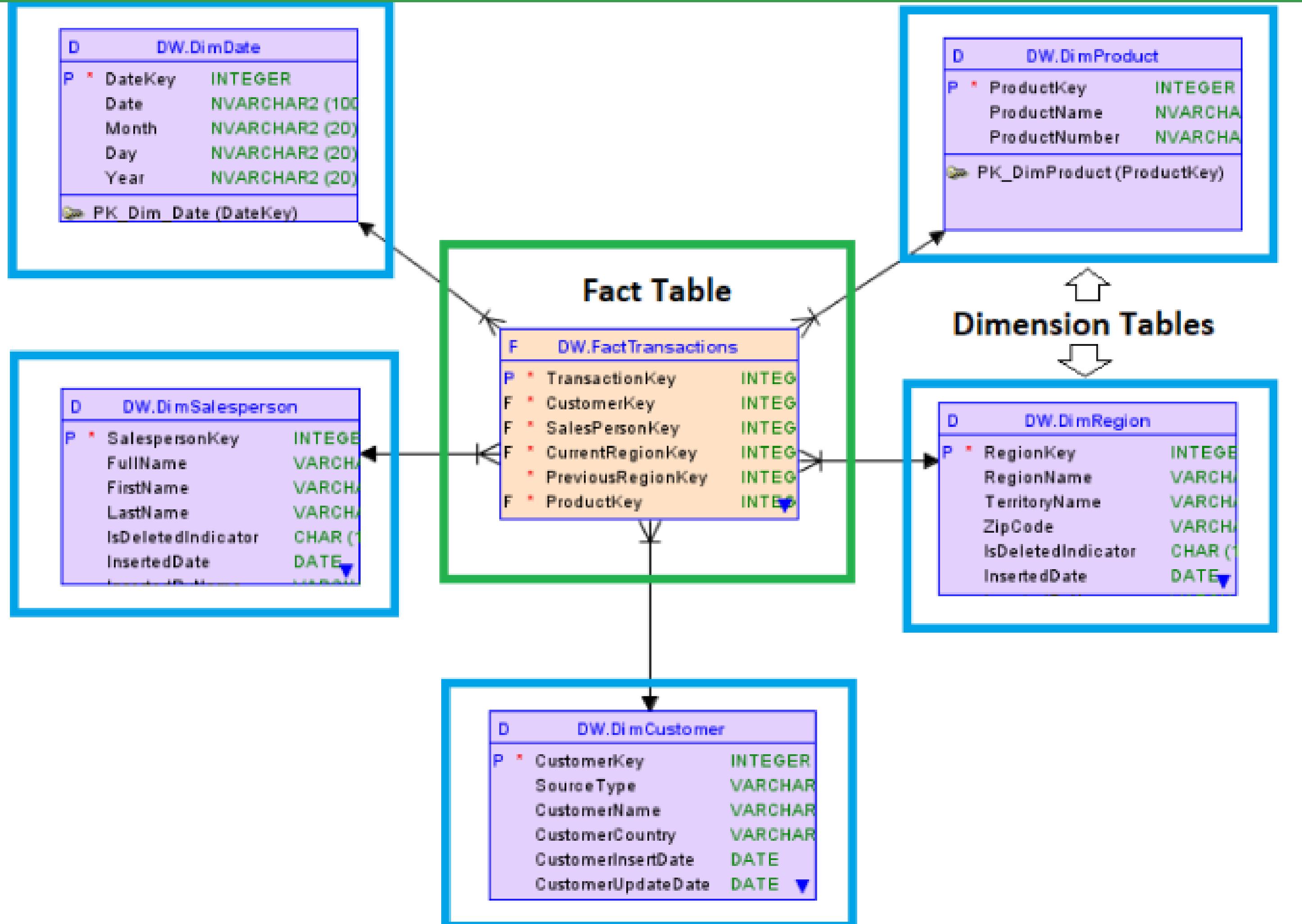


Flat-table Schema



# FLAT SCHEMA

- All characteristics are completely denormalized into a single table in the flat type of data model.
- There are no relationships because there is just one table, and most of the time a key is not required.
- Advantages :
  - Flat schemas are particularly effective in terms of performance.
- Disadvantages:
  - It can be difficult and perplexing to browse a single table.
  - Data and columns might frequently be duplicated, which increases the file size.
  - The complexity of DAX formulae increases as multiple facts are combined.



# DIMENSION TABLE

- You can slice and dice your fact tables using the descriptive features in these tables.
- A key column—the distinctive identifier—and descriptive columns make up a dimension table.
- Dimension tables are also sometimes known as lookup tables
- A dimension table is used to keep track of a business's operations.
- Dimension tables support *filtering* and *grouping*

The distinct set of dimensions, or columns, that specify what a record is the grain of your data.

# FACT TABLE

- The metrics you want to aggregate are in these tables.
- Foreign keys, which are necessary to establish associations with dimensions, and aggregable columns are both features of fact tables.
- Fact tables store observations or events, and can be sales orders, stock balances, exchange rates, temperatures, etc
- Fact tables are also known as data tables.
- It is employed in decision-making and analysis.
- Fact tables support *summarization*
- A common relationship cardinality is *one-to-many* or its inverse *many-to-one*. The "one" side is always a dimension-type table while the "many" side is always a fact-type table.

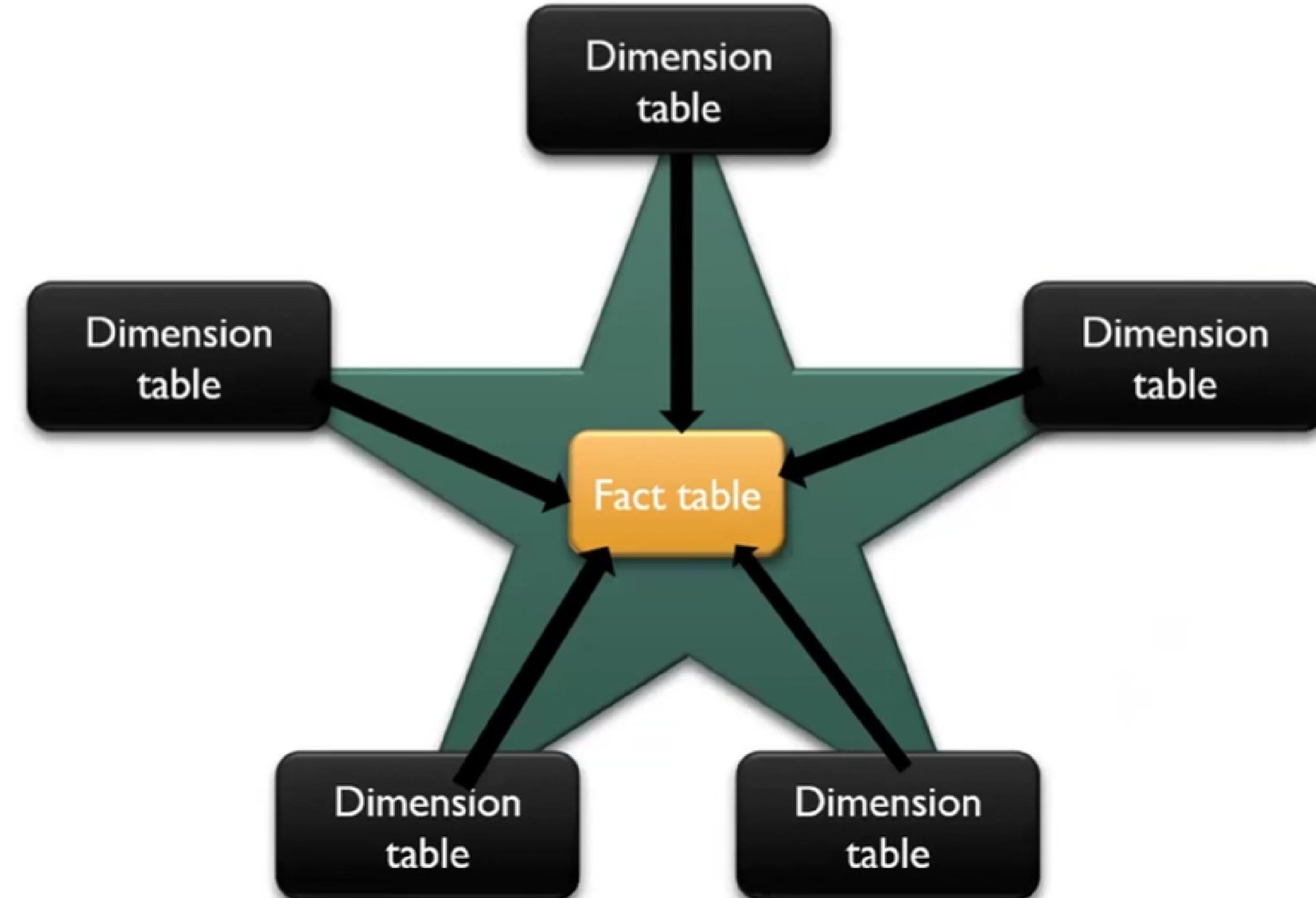
# COMPARISON

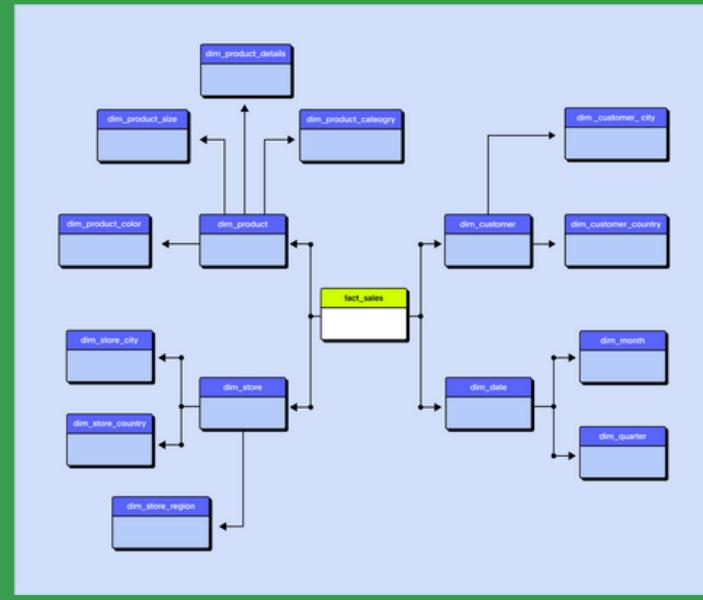
Basis of Distinction	Fact Table	Dimension Table
Design	It's defined at the atomic level.	It must be in-depth, comprehensive, and of the greatest caliber.
Data Type	May contain information on sales in respect to various criteria, such as Date, Product etc.	There are attributes in every dimension table that describe the details of the dimension. For instance, the product ID, category, and other information might be included in the product dimensions.
Key	The primary key of the fact table is transferred to dimensions as a foreign key.	A main key column in a dimension database uniquely identifies it.
Hierarchy	No hierarchy present in it.	Hierarchies are present in dimension table. Location, for instance, may include a nation, state, city, zip code, and more.
Application	Used for reporting and analysis.	Used to collect background data of a company.

# STAR SCHEMA

- The fact table is in the center, and the dimension tables surrounds it, giving the star schema its name.
- A star schema can have more than one fact table while still remaining a star schema.
- In most circumstances, Power BI's preferred data modeling method is the star schema.
- Because the fields are logically arranged, the model is simpler to comprehend.
- Less data is duplicated, which improves storage effectiveness.
- To work with fact tables that have a varied grain, you don't need to create incredibly complicated DAX formulas.

# STRUCTURE OF STAR SCHEMA





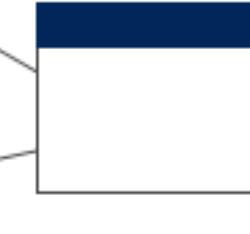
## SNOWFLAKE

## Snowflake Schema

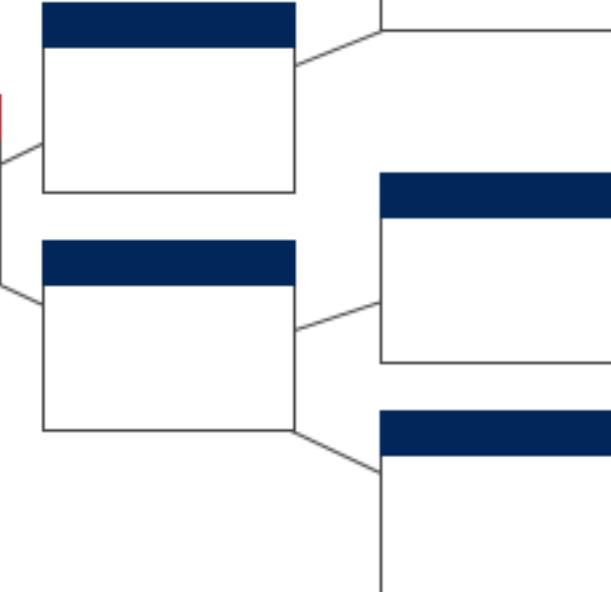
Subdimensions



Dimensions



Dimensions



- The snowflake schema is similar to the star schema, except it can have some dimensions that “snowflake” from other dimensions.
- A **snowflake dimension** is a set of normalized tables for a single business entity.
- Its less efficient from storage and performance perspectives, because it loads more table(which may not be very important).
- Snowflakes is not optimal data modelling regarding powerbi.
- Its better to combine all related table into one table and moving forward with star schema.

# PRIMARY KEY AND FOREIGN KEY

Primary Key	Foreign Key
To guarantee that the data in the particular column is unique, a primary key is employed.	Foreign key is a column or set of columns that creates a connection between the data in two tables.
In a table, only one primary key is permitted.	A table may include several foreign keys.
It consists of unique and non-null values	Duplicate values can be included.

- Having active foreign keys on tables improves data quality but hurts performance of insert, update and delete operations.

## PREFERRED CARDINALITY

- ✓ One to One
- ✓ One to Many or Many to One
- ✗ Many to Many

# ACTIVE AND INACTIVE RELATIONS BETWEEN TABLES

- Active relationships automatically propagate filters to other tables.
- However, an inactive relationship only propagates filters when it is activated by a DAX expression.
- One can only have one active relationship between two tables.
- Rest of the relationships will be inactive
- If two tables are connected with a steady line the relationship is active.
- And if two tables are connected with a dotted line the relationship is inactive.



# INTRODUCTION

- DAX – Data Analysis Expressions
- It's a library of functions and operators that can be combined to build formulas and expressions in Power BI

# DIFFERENCES BETWEEN MEASURES AND CALCULATED COLUMN

Measures	Calculated Columns
New calculated value	New calculated column
Based on filter context	Based on row context(performing calculation for each row)
Can be seen in report view only	Can be seen in both report and data view
Does not increase file size	Increases file size(because it takes some space)  Therefore its formed only when its necessary



# tutorialspoint

## **DAX Basics in Power BI**

DAX Basics in Power BI - In this chapter, you will learn how to use various DAX functions in Power BI.