



Intro to Data Model

Using Microsoft Excel 365 <>





Course Outline



Content

- Getting Started
- ER Diagram
- Data Analyst Modern Workflow
- Intro to Data Model

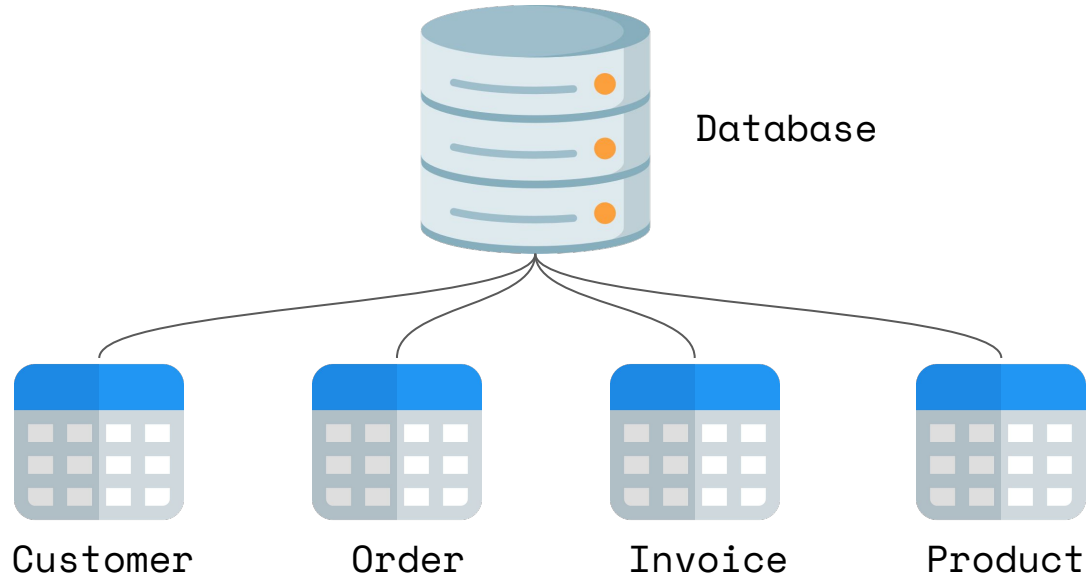




Getting Started

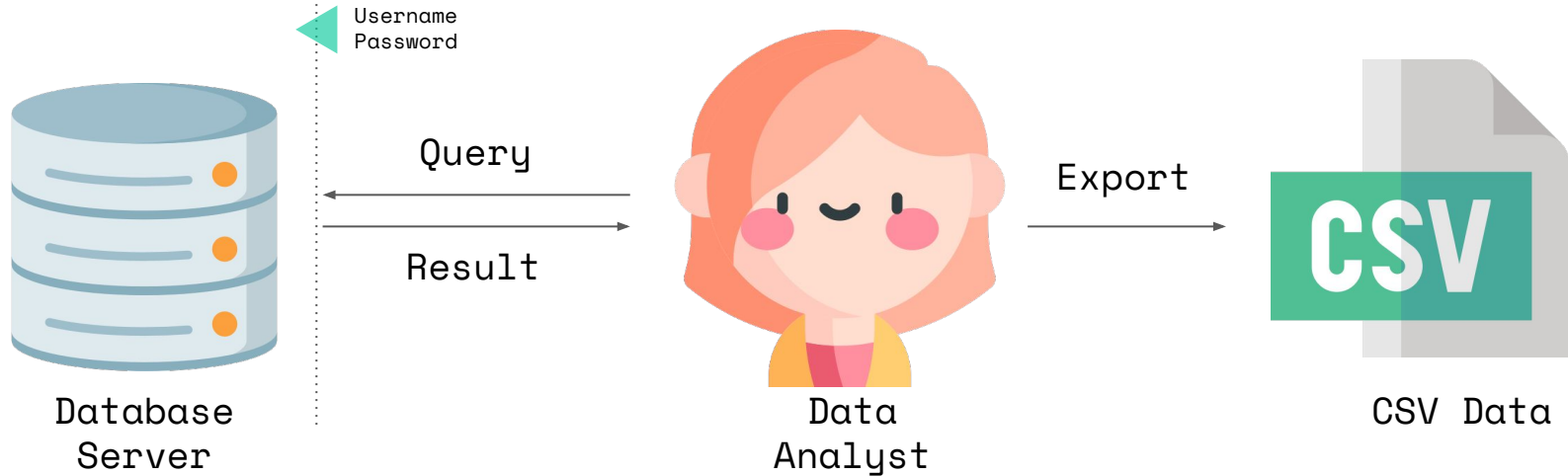


What is Database?



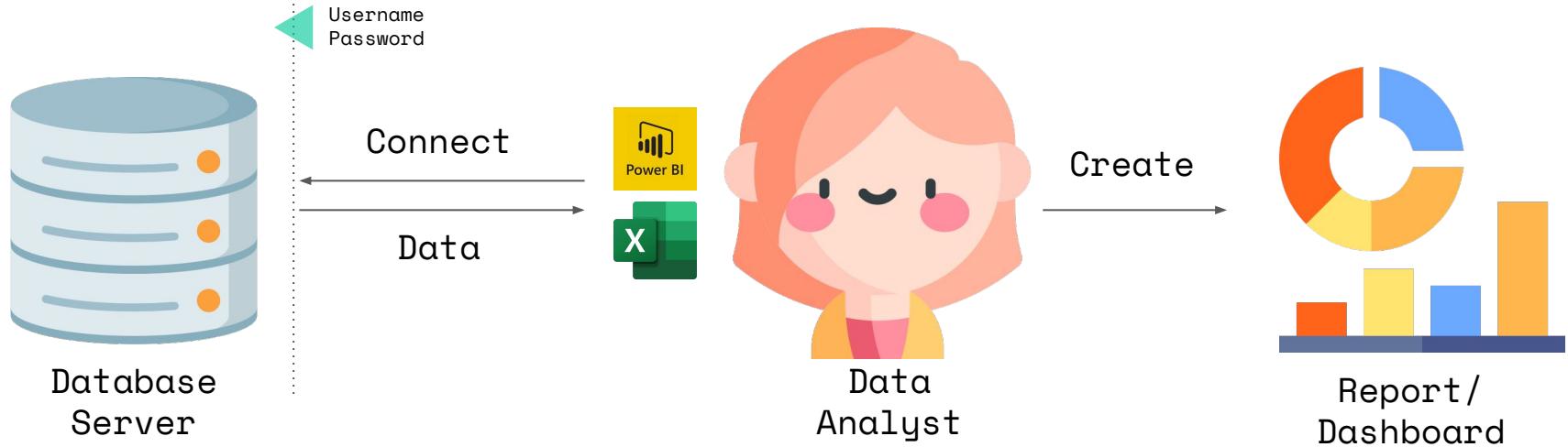
Classic

Data Analyst Workflow



Modern

Data Analyst Workflow V2



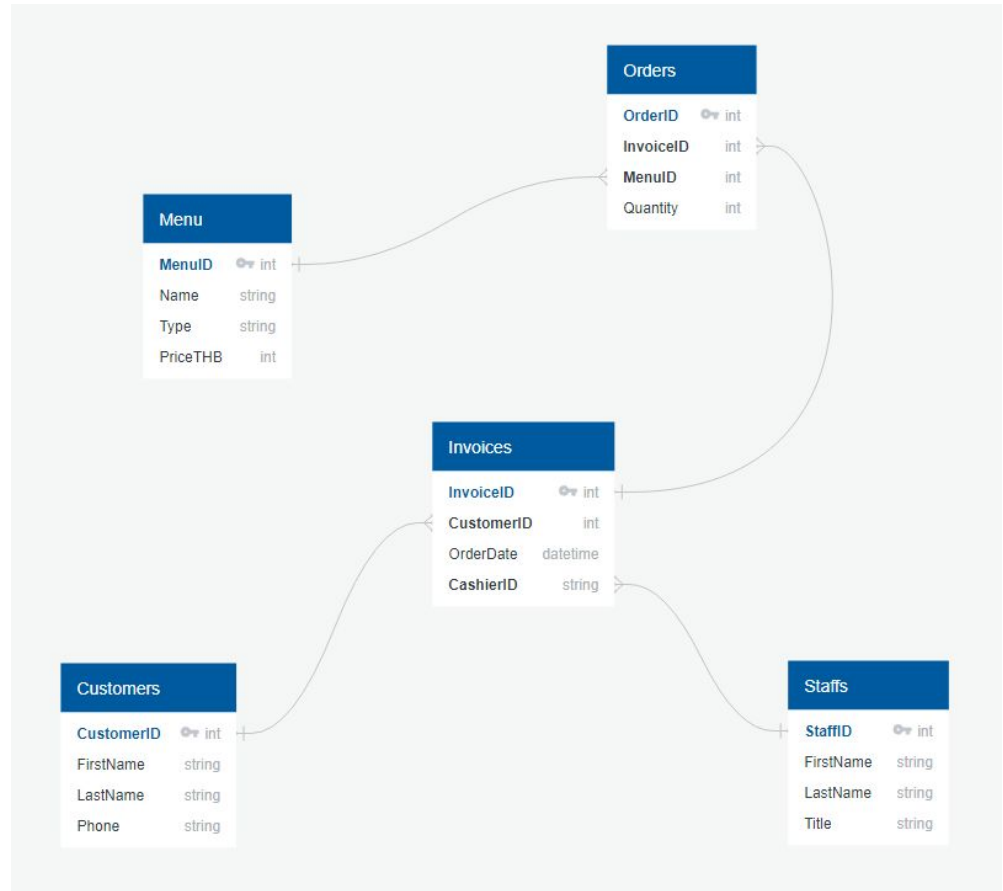


ER Diagram

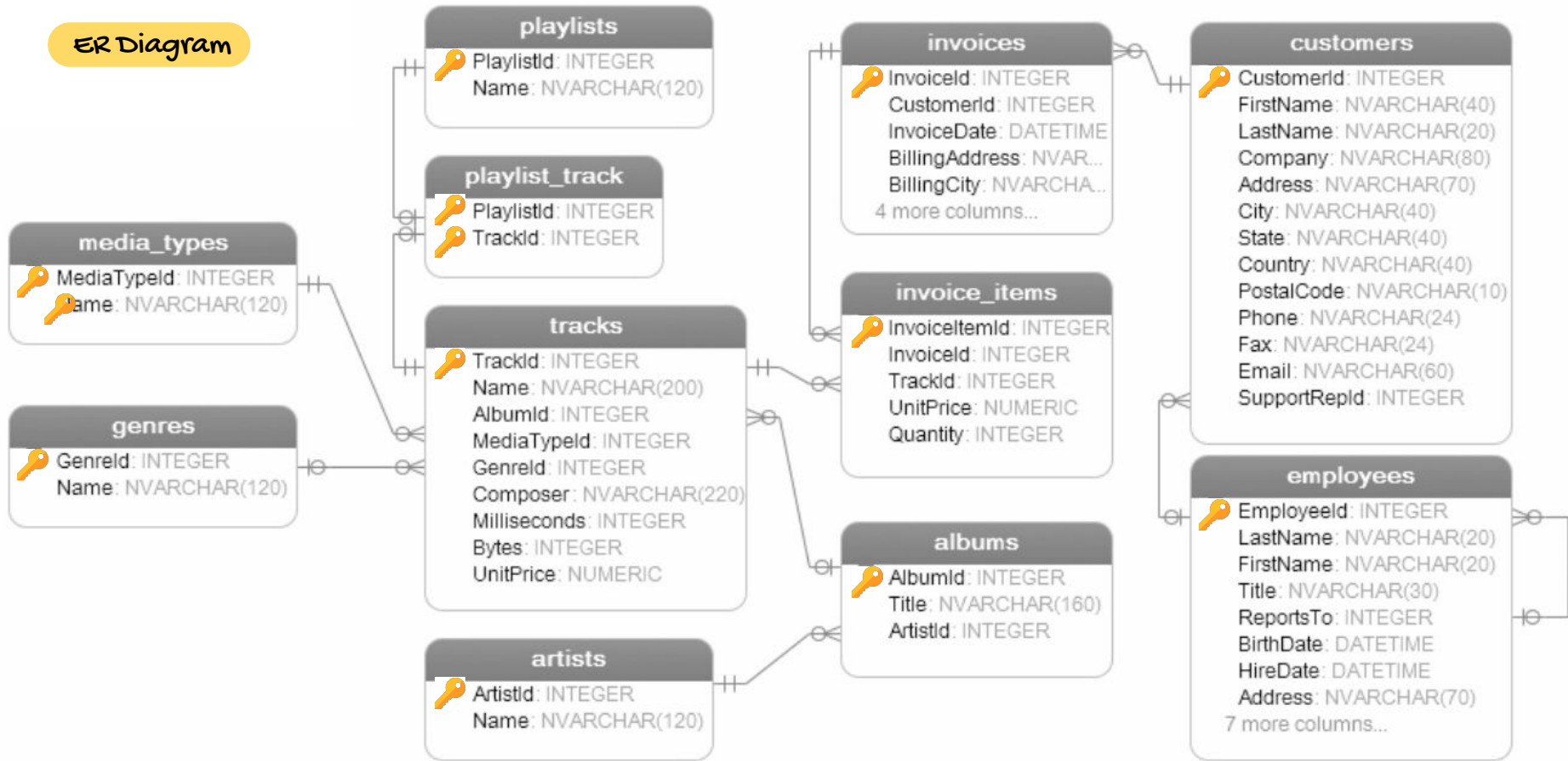


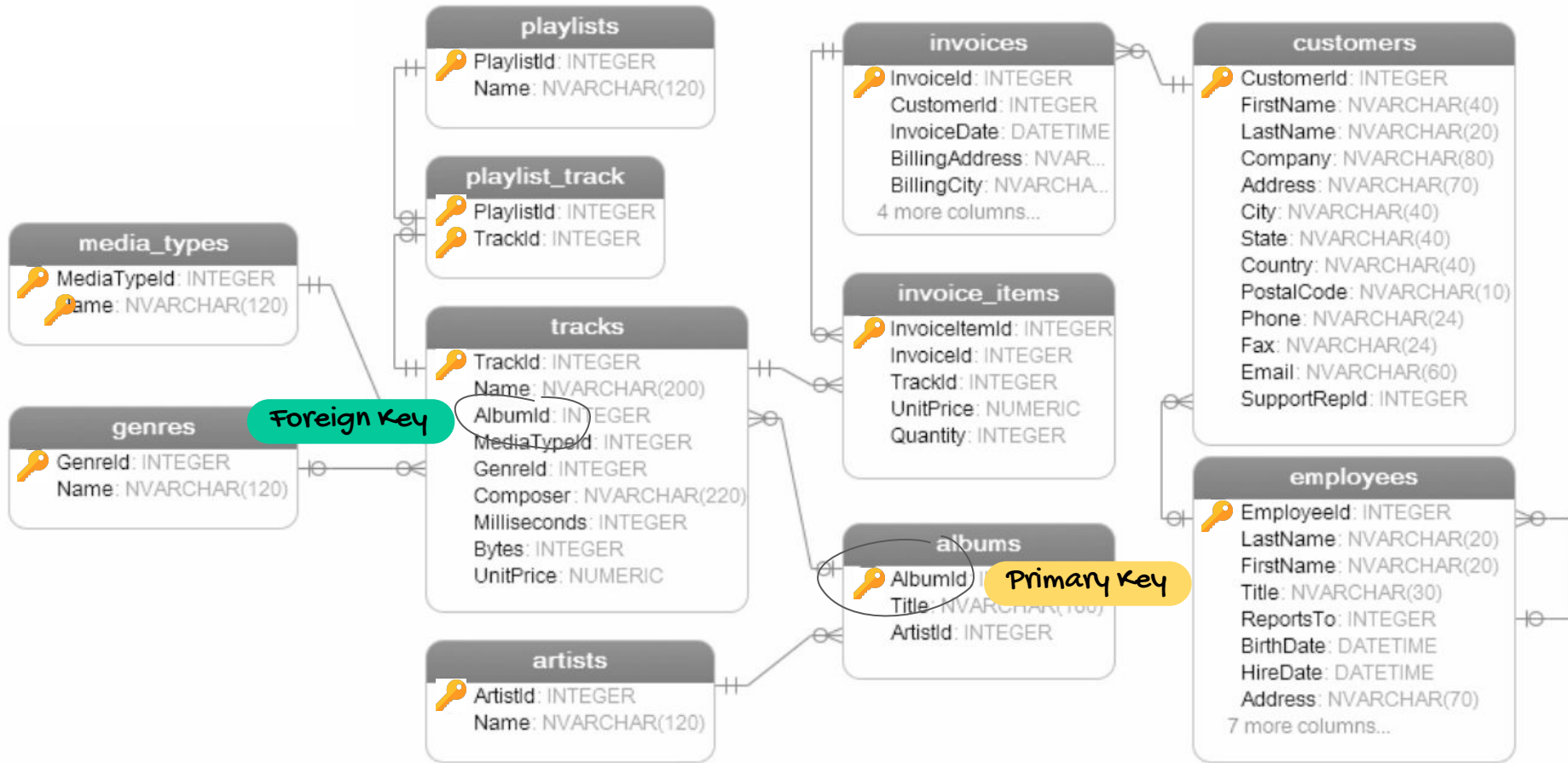
ER Diagram

Entity Relationship Diagram



ER Diagram





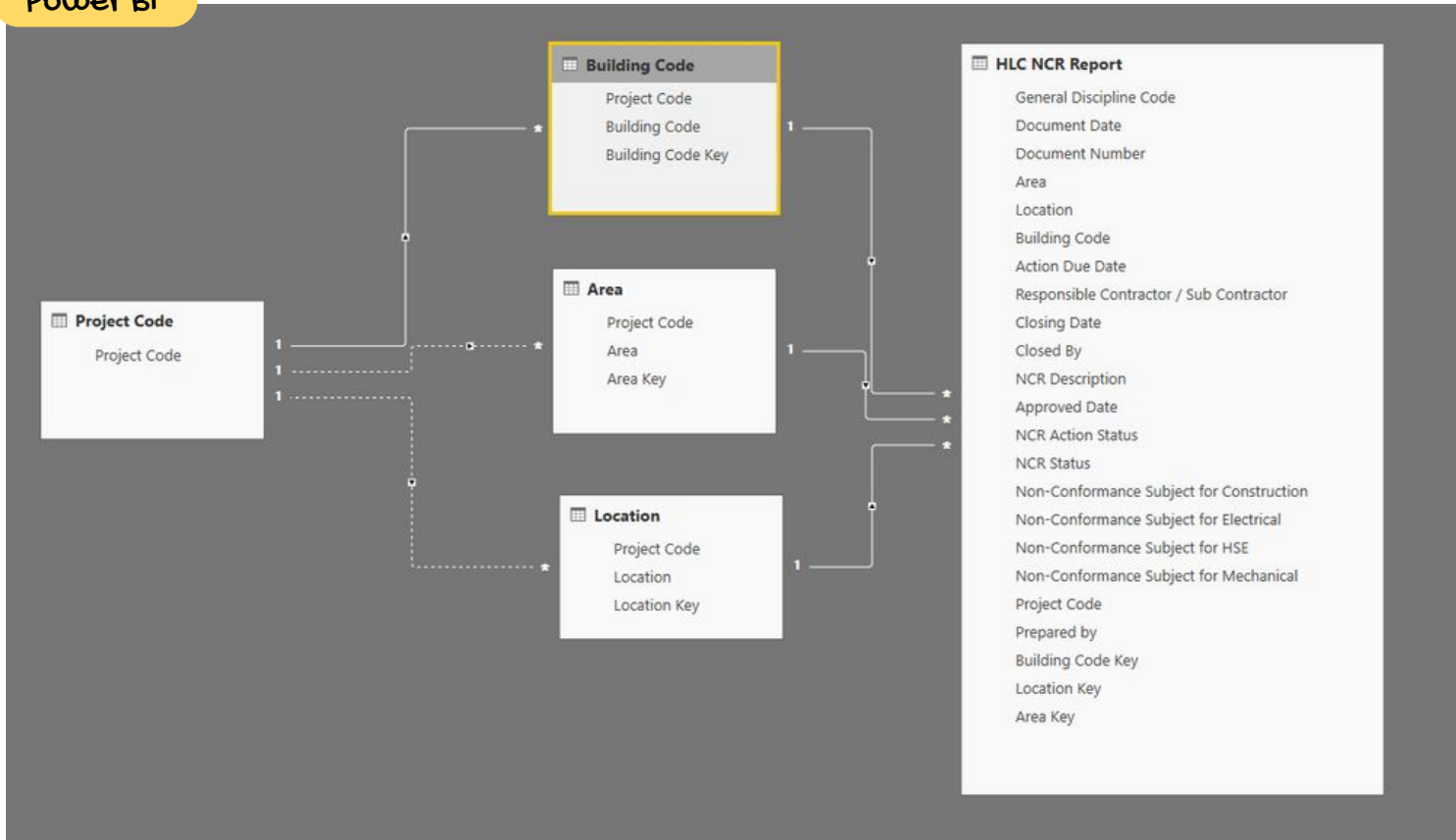
Cardinality

Types of Relationship

- One to One (1 - 1)
- One to Many (1 - *)
- Many to One (* - 1)
- Many to Many (* - *)



Power BI



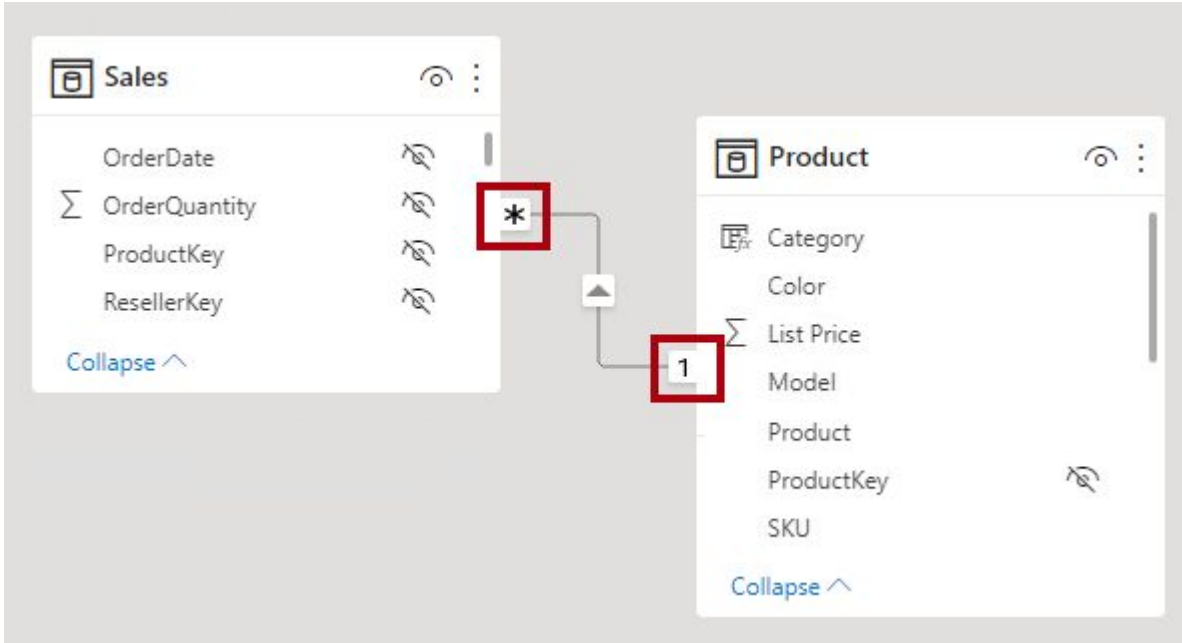


What is the **most common** cardinality in business?

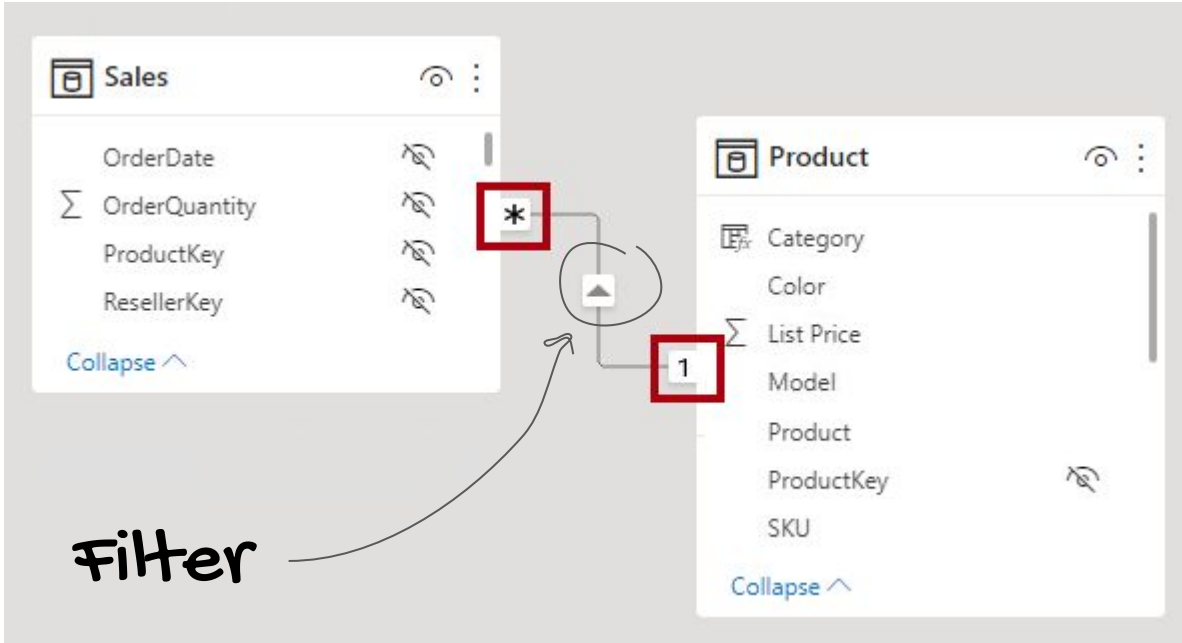
1. One-One
2. One-Many (Many-One)
3. Many-Many



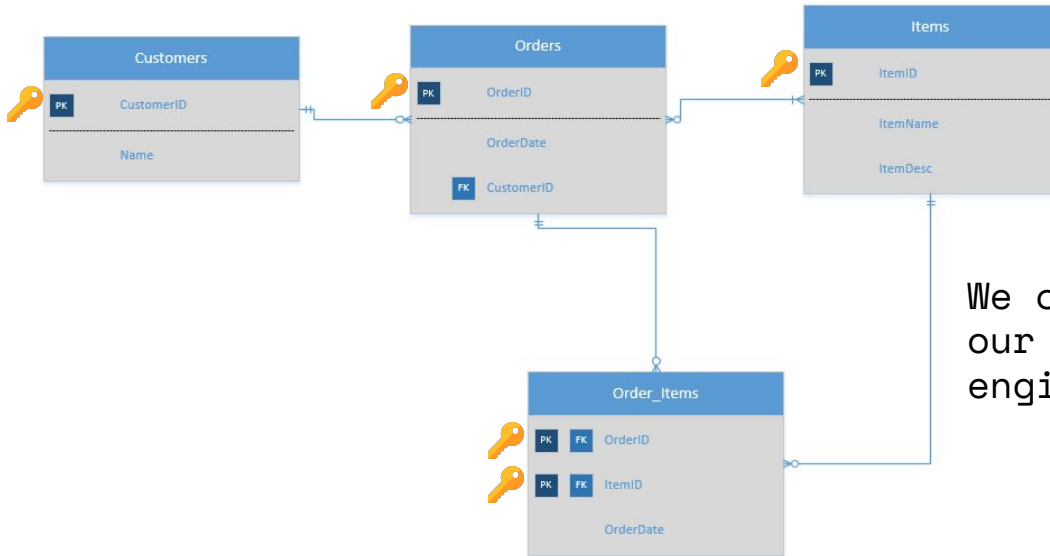
The most common cardinality in business is **(1 - *)** or **(* - 1)**



The most common cardinality in business is **(1 - *)** or **(* - 1)**



Data Model



We can **model relationships** in our data (Usually done by data/engineering team)



A Good Data Model

- Data exploration is **faster**
- Aggregations are **simpler** to build
- Reports are more **accurate**
- Writing reports takes **less time**
- Reports are **easier** to maintain in the future





Our First Model





File > Options

Excel Options

View and manage Microsoft Office Add-ins.

Add-ins

Name	Location	Type
Active Application Add-ins		
Analysis ToolPak	C:\...Office16\Library\Analysis\ANALYS32.XLL	Excel Add-in
Microsoft Power Map for Excel	C:\...p Excel Add-in\EXCELPLUGINSHELL.DLL	COM Add-in
Microsoft Power Pivot for Excel	C:\...l Add-in\PowerPivotExcelClientAddIn.dll	COM Add-in
Inactive Application Add-ins		
Analysis ToolPak - VBA	C:\...ice16\Library\Analysis\ATPVBAEN.XLAM	Excel Add-in
Date (XML)	C:\...s\Microsoft Shared\Smart Tag\MOFL.DLL	Action
Euro Currency Tools	C:\...root\Office16\Library\EUROTOOL.XLAM	Excel Add-in
Microsoft Actions Pane 3		XML Expansion Pack
Microsoft Data Streamer for Excel	C:\...softDataStre	
Solver Add-in	C:\...Office16\Lib	
Xlstat	C:\...gram Files\A	
Xlstatrib	C:\... Files\Addin	
Document Related Add-ins		
No Document Related Add-ins		
Add-in:	Analysis ToolPak	
Publisher:	Microsoft Office	
Compatibility:	No compatibility information available	
Location:	C:\Program Files\Microsoft Office\root\Office	
Description:	Provides data analysis tools for statistical and engineering analysis	

Manage: COM Add-ins Go...

OK Cancel

AutoSave Off Book1 - Excel Search (Alt+Q)

File Home Insert Draw Page Layout Formulas Data Review View Developer Help **Power Pivot**

Manage Measures KPIs Add to Data Model Detect Settings

Data Model Calculations Tables Relationships





Star Schema



Star Schema

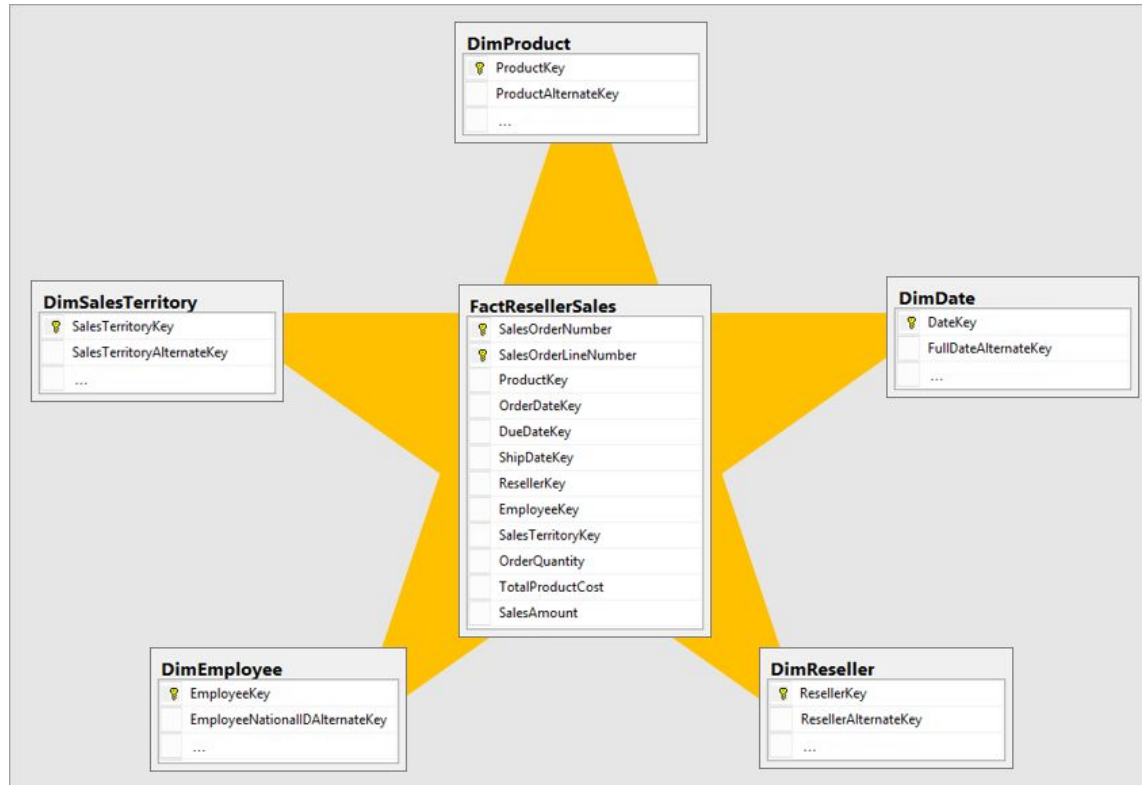
Star schema is a **mature modeling** approach widely adopted by relational data warehouses. It requires modelers to classify their model tables as either **dimension** or **fact**.

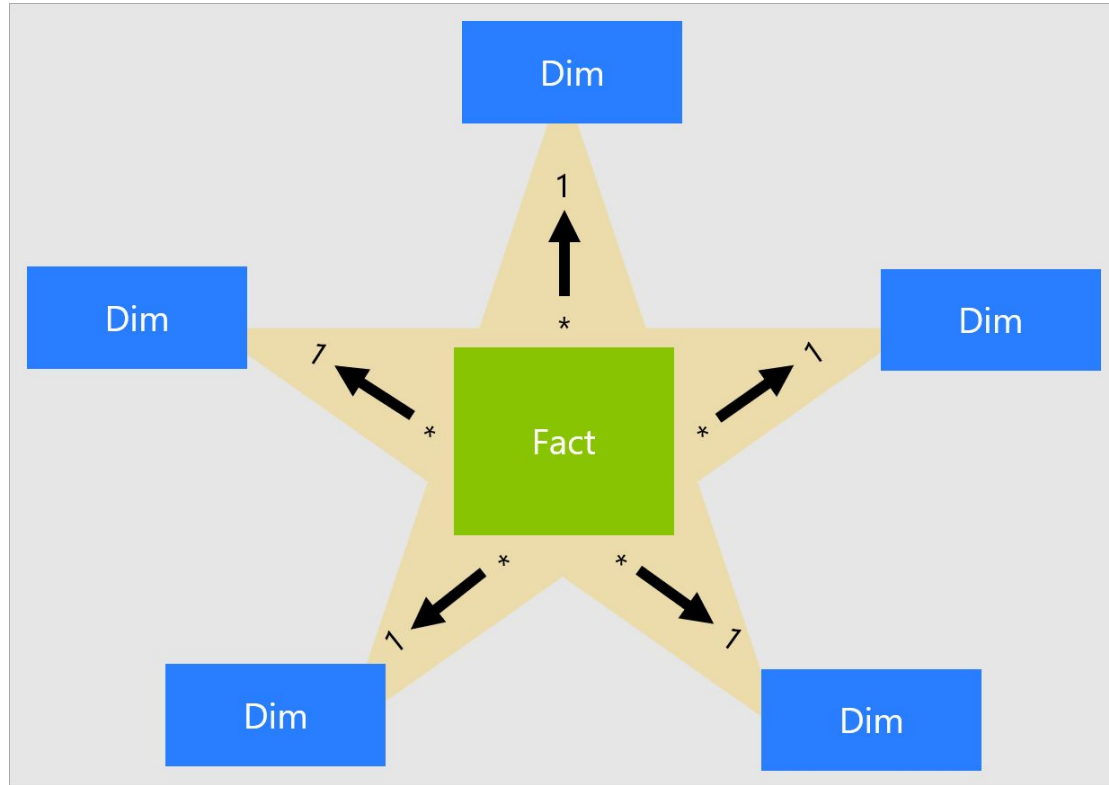


Two Types of Table

- **Fact** → store observations or events, and can be sales orders, stock balances, exchange rates, temperatures, etc.
- **Dimension** → describe business entities—the things you model



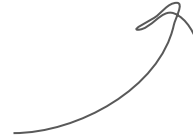




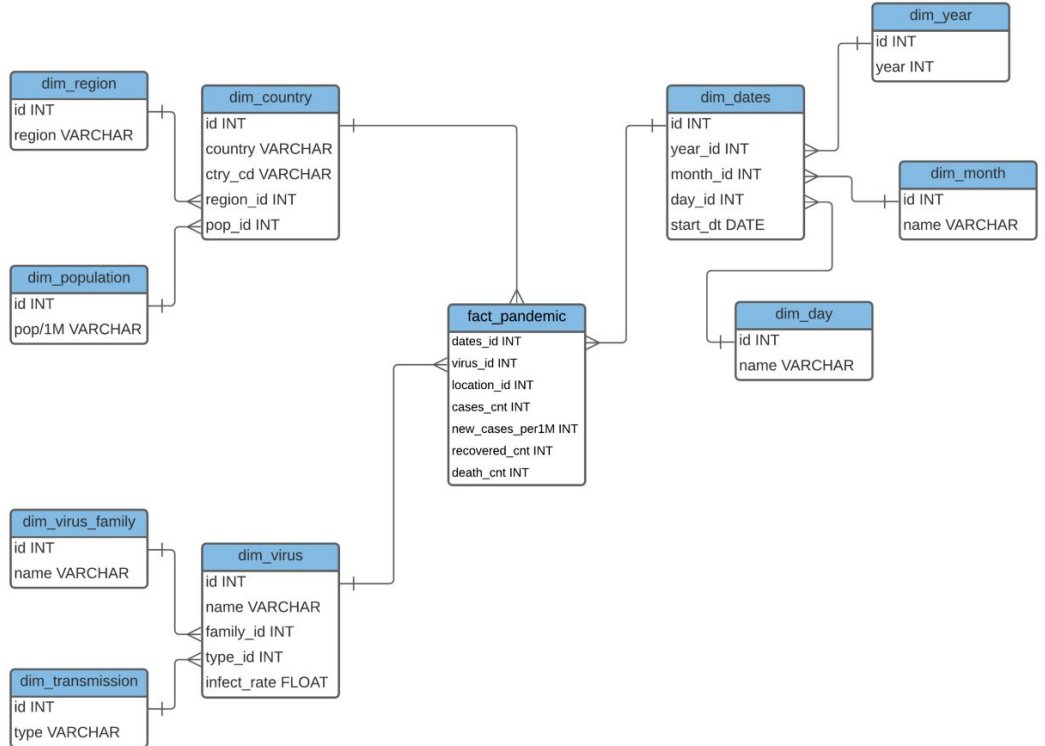


is optimized for **STAR** schema

The best model for
Power BI



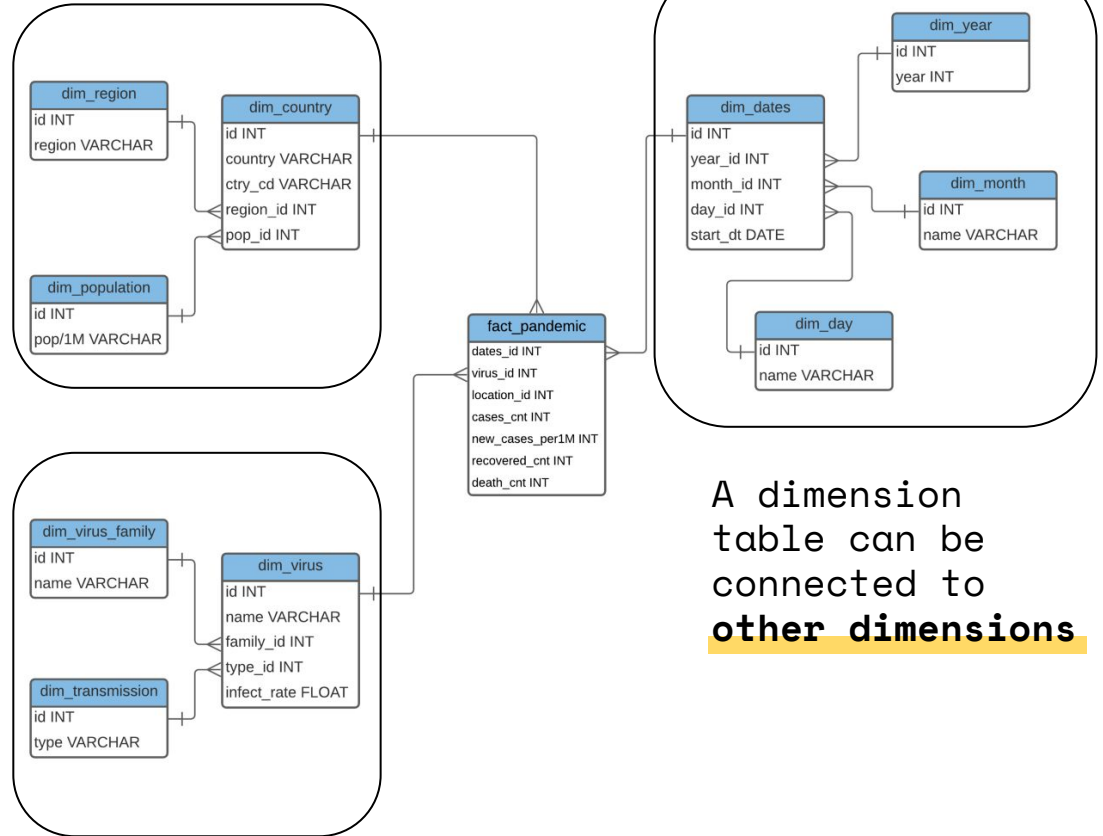
Snowflake



Snowflake



Group of dimension tables



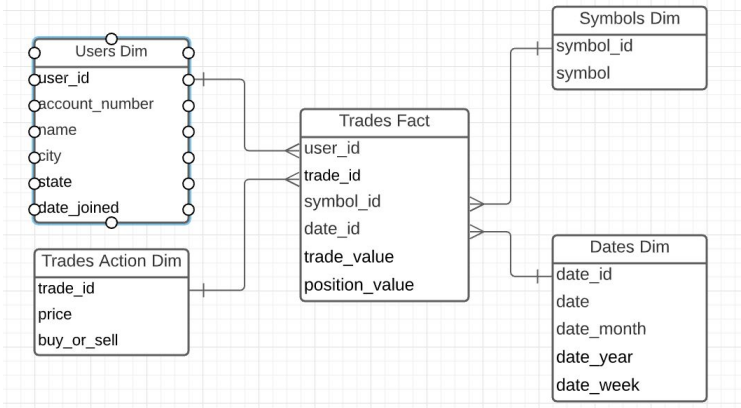
A dimension table can be connected to **other dimensions**



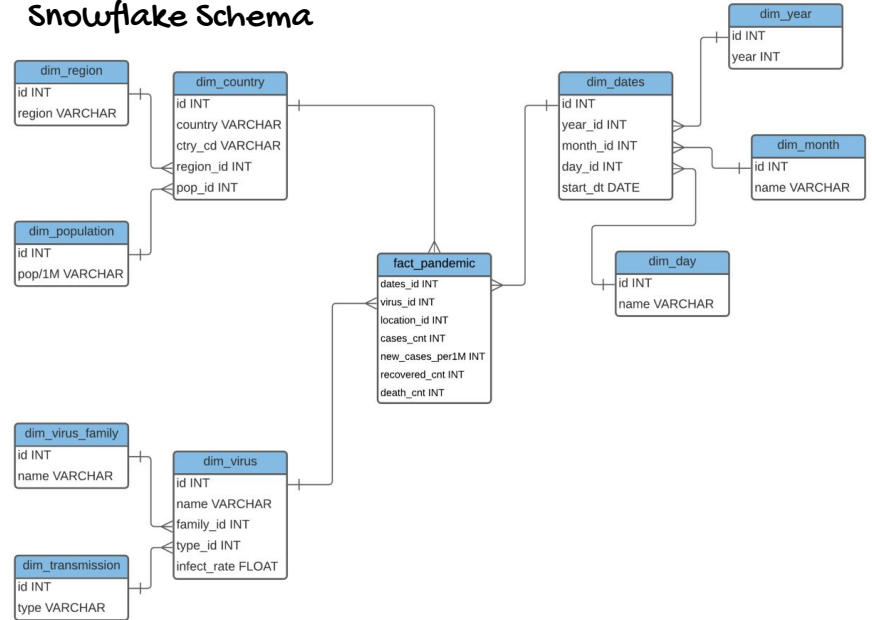


What schema do you think is **better**?

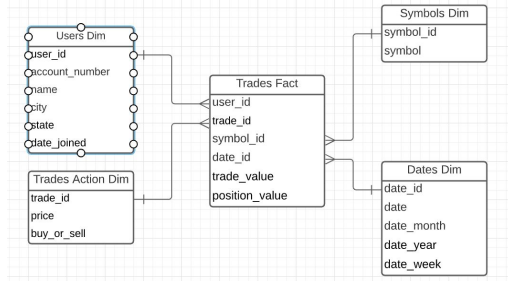
Star Schema



Snowflake Schema

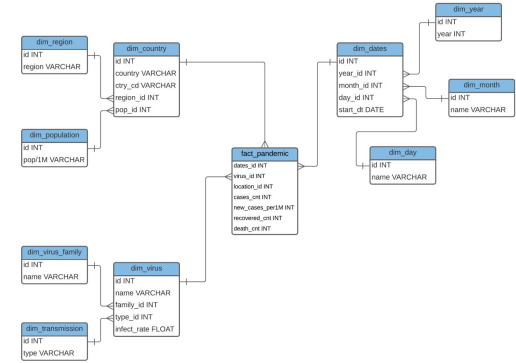


Star Schema



- Fast
- More efficient
- Simple model
- Redundant data

Snowflake Schema



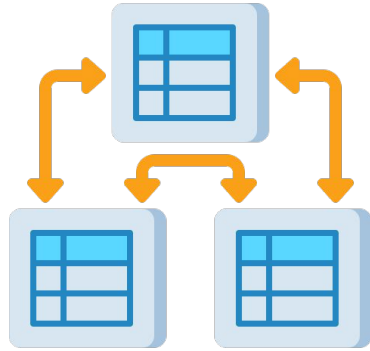
- Remove redundant data
- More complex as needed
- Relatively slower than star schema





Normalization





Normalization is the term used to describe data that's stored in a way that reduces repetitive data.



Sales Table

	SalesOrderNumber	OrderDate	ProductKey	Product	Category	Color	Size	ResellerKey	SalesAmount
1	SO69561	2020-05-31	594	Mountain-500 Silver, 48	Bikes	Silver	48	546	226.00
2	SO69560	2020-05-30	513	ML Mountain Frame-W - Silver, 46	Components	Silver	46	100	218.45
3	SO69560	2020-05-30	594	Mountain-500 Silver, 48	Bikes	Silver	48	100	113.00
4	SO69539	2020-05-28	243	HL Road Frame - Red, 44	Components	Red	44	529	858.90
5	SO69539	2020-05-28	378	Road-250 Black, 52	Bikes	Black	52	529	1466.01
6	SO69541	2020-05-28	594	Mountain-500 Silver, 48	Bikes	Silver	48	661	113.00
7	SO69542	2020-05-28	243	HL Road Frame - Red, 44	Components	Red	44	317	1717.80
8	SO69544	2020-05-28	243	HL Road Frame - Red, 44	Components	Red	44	666	3435.60
9	SO69545	2020-05-28	378	Road-250 Black, 52	Bikes	Black	52	436	5864.04
10	SO69532	2020-05-27	594	Mountain-500 Silver, 48	Bikes	Silver	48	312	113.00
11	SO69532	2020-05-27	513	ML Mountain Frame-W - Silver, 46	Components	Silver	46	312	436.90
12	SO69533	2020-05-27	594	Mountain-500 Silver, 48	Bikes	Silver	48	476	226.00





Sales: Fact Table

	SalesOrderNumber	OrderDate	ProductKey	ResellerKey	SalesAmount
1	SO69561	2020-05-31	594	546	226.00
2	SO69560	2020-05-30	513	100	218.45
3	SO69560	2020-05-30	594	100	113.00
4	SO69539	2020-05-28	243	529	858.90
5	SO69539	2020-05-28	378	529	1466.01
6	SO69541	2020-05-28	594	661	113.00
7	SO69542	2020-05-28	243	317	1717.80
8	SO69544	2020-05-28	243	666	3435.60
9	SO69545	2020-05-28	378	436	5864.04
10	SO69532	2020-05-27	594	312	113.00
11	SO69532	2020-05-27	513	312	436.90
12	SO69533	2020-05-27	594	476	226.00



Essential knowledge to build a *star schema*



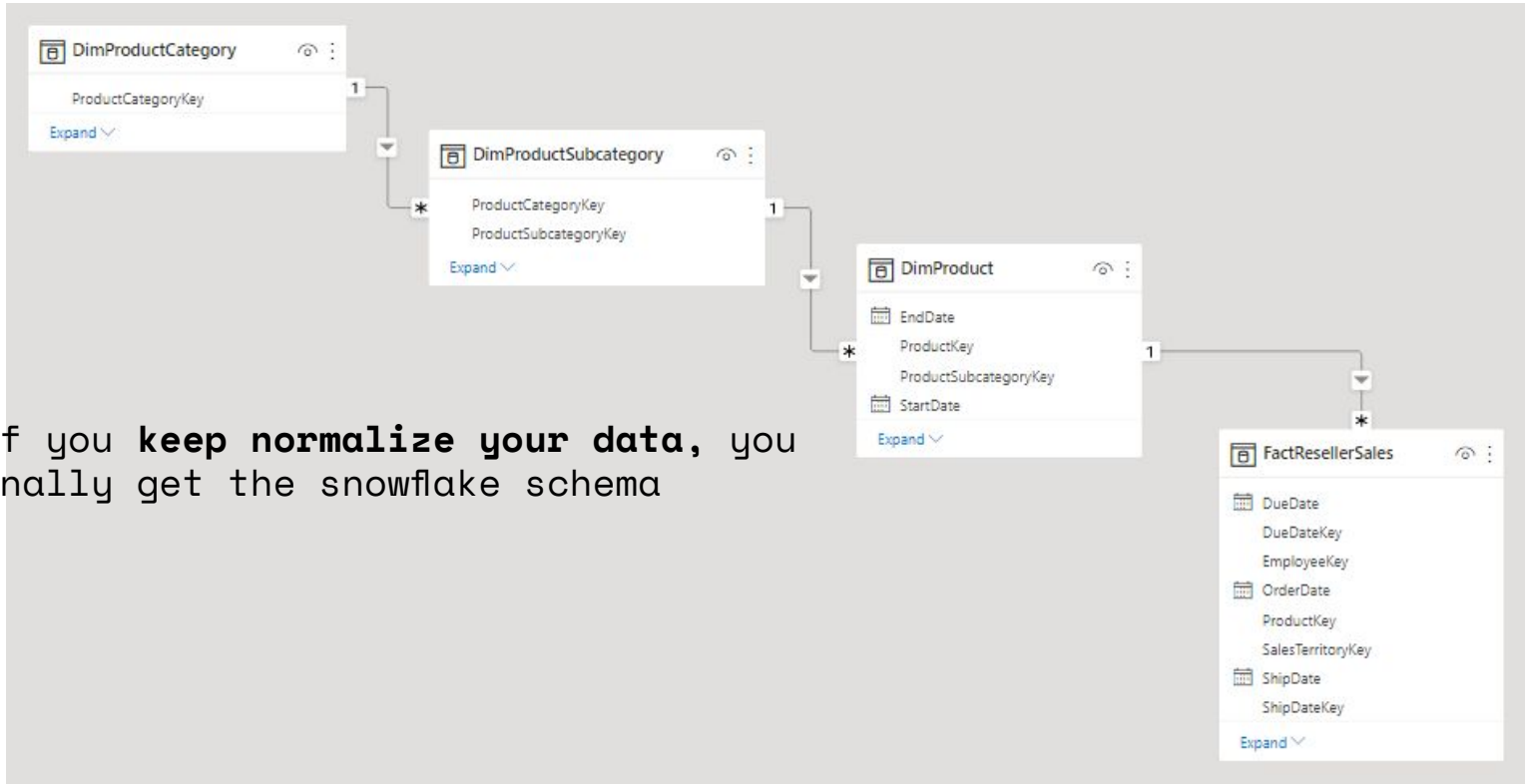
Products: Dimension Table

ProductKey	ProductName	ProductDescription
100	Bikes V100	Mountain bike
101	Bikes V150	Mountain bike upgraded
102	Motorbike T1	A cool looking motorbike
103	Motorbike T2	A redesigned T1 model
..





If you **keep normalize your data**, you finally get the snowflake schema



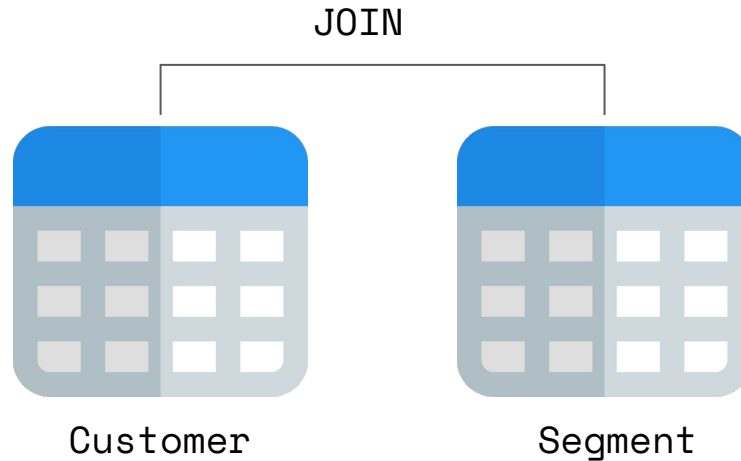


JOIN Multiple Tables



What is JOIN?

Getting data from multiple tables



SQL JOIN = VLOOKUP()

Customer

ID	Name	City
1001	Toy	BKK
1002	Anna	LON
1003	Marry	LON

Segment

ID	SegName	Cust_ID
1	Deal Hunter	1001
2	Price Sensitive	1002
3	Premium	1003

Join PK=FK



Result Set

Cool!

ID	Name	City	SegName
1001	Toy	BKK	Deal Hunter
1002	Anna	LON	Price Sensitive
1003	Marry	LON	Premium

Super Easy!





Common JOINS



INNER JOIN (default)

Customer

ID	Name	City
1001	Toy	BKK
1002	Anna	LON
1003	Marry	LON
1004	Ken	JPN

Segment

ID	SegName	Cust_ID
1	Deal Hunter	1001
2	Price Sensitive	1002
3	Premium	1003

Join PK=FK



Result Set

inner join

Only Matched Rows Return

ID	Name	City	ID	SegName	Cust_ID
1001	Toy	BKK	1	Deal Hunter	1001
1002	Anna	LON	2	Price Sensitive	1002
1003	Marry	LON	3	Premium	1003



LEFT JOIN

Customer

ID	Name	City
1001	Toy	BKK
1002	Anna	LON
1003	Marry	LON
1004	Ken	JPN

Segment

ID	SegName	Cust_ID
1	Deal Hunter	1001
2	Price Sensitive	1002
3	Premium	1003

Join PK=FK



Result Set left join

All rows in left table will be in the result set

ID	Name	City	ID	SegName	Cust_ID
1001	Toy	BKK	1	Deal Hunter	1001
1002	Anna	LON	2	Price Sensitive	1002
1003	Marry	LON	3	Premium	1003
1004	Ken	JPN	NULL	NULL	NULL





If you JOIN two tables,
which can give **lower count**
of records?

1. LEFT JOIN
2. INNER JOIN





Intro to Data Model

Using Microsoft Excel 365 <>

