



Data Visualization with Tableau

Session 5



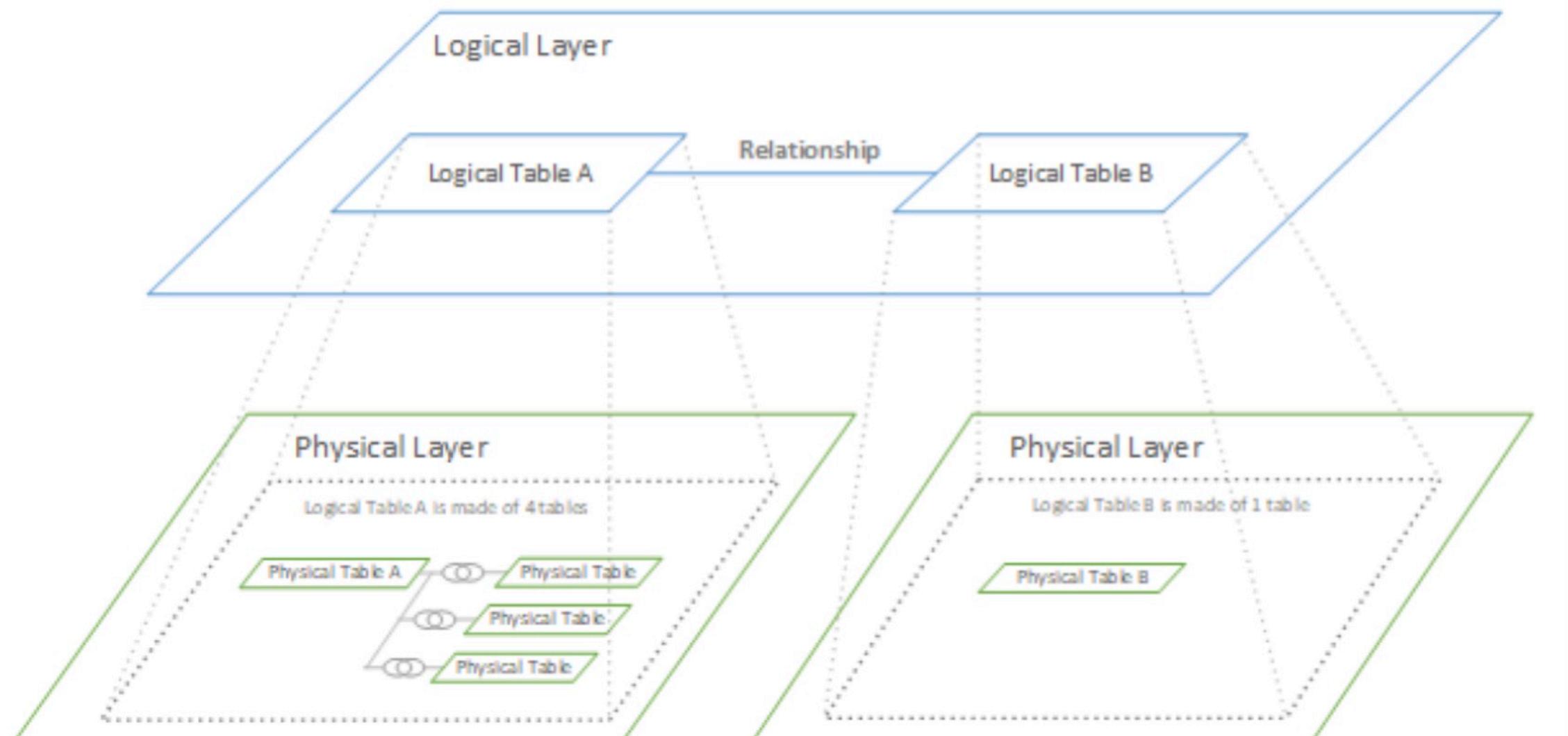
Table of Contents



- ▶ Understanding the data model
- ▶ Join Types and Union
- ▶ Blending Multiple Data Sources
- ▶ Relationships

Understanding the data model

Data Model



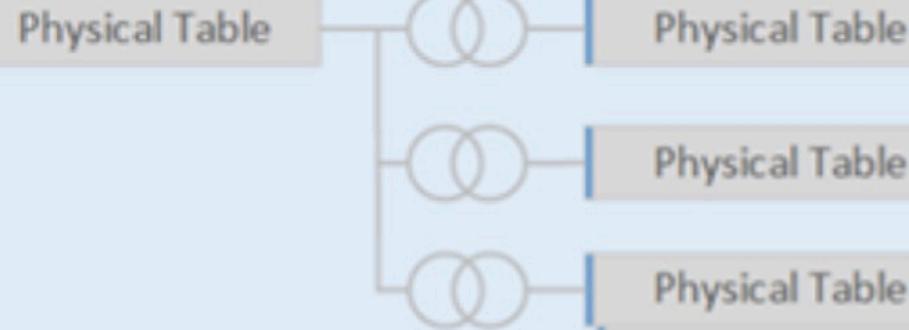
Understanding the data model



Previous versions

Data Source (Previous Tableau)

Data Model

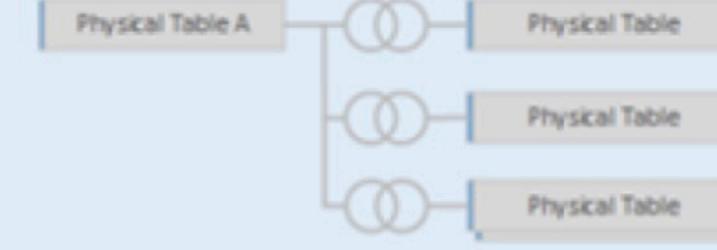


2020.2 and later

Data Source

Data Model

Logical Table A



Logical Table B

Physical Table B

Understanding the data model



- ❖ Tables at the logical layer are **NOT merged** in the data source, they remain **distinct** (normalized), and maintain their **native level of detail**.
- ❖ Logical tables act like **containers** for merged physical tables.
- ❖ A logical table can contain a **single**, physical table. Or it can contain **multiple** physical tables merged together through joins or unions.



Build a new model

- ❖ When you **add one or more tables to the logical layer**, you are essentially **building the data model** for your data source. A data source can be made of a **single**, logical table, or you can drag **multiple** tables to the canvas to create a more complex model.
- ❖ The first table that you drag to the canvas becomes the **root table** for the data model in your data source.
- ❖ After you drag out the root table, you can drag out additional tables in any order. You will need to consider which tables should be related to each other, and the matching field pairs that you define for each relationship.

Build a new model



- ❖ **Deleting a table in the canvas** automatically **deletes its related descendants** as well. If you delete the root table, all other tables in the model are also removed.
- ❖ Each relationship must be made of **at least** one matched pair of fields. Add multiple field pairs to create a compound relationship.
- ❖ Relationships can be based on calculated fields.

Join Types



- ▶ Left Join
- ▶ Right Join
- ▶ Inner Join
- ▶ Outer Join

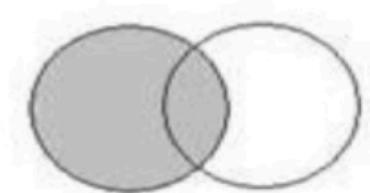
Name	# of Siblings
Taylor	2
Alex	3
Shannon	0
Tracy	1



Name	Eye Color
Taylor	Blue
Alex	Brown
Morgan	Brown

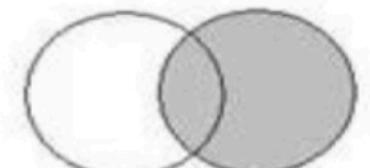
Left Join

Name	# of Siblings	Eye Color
Taylor	2	Blue
Alex	3	Brown
Shannon	0	null
Tracy	1	null



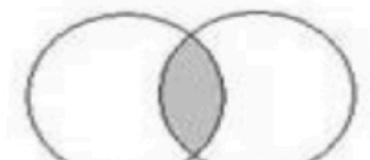
Right Join

Name	Eye Color	# of Siblings
Taylor	Blue	2
Alex	Brown	3
Morgan	Brown	null



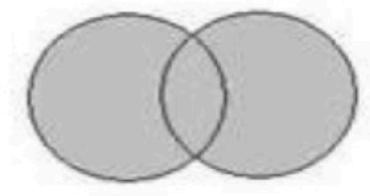
Inner Join

Name	# of Siblings	Eye Color
Taylor	2	Blue
Alex	3	Brown



Outer Join

Name	# of Siblings	Eye Color
Taylor	2	Blue
Alex	3	Brown
Shannon	0	null
Tracy	1	null
Morgan	null	Brown

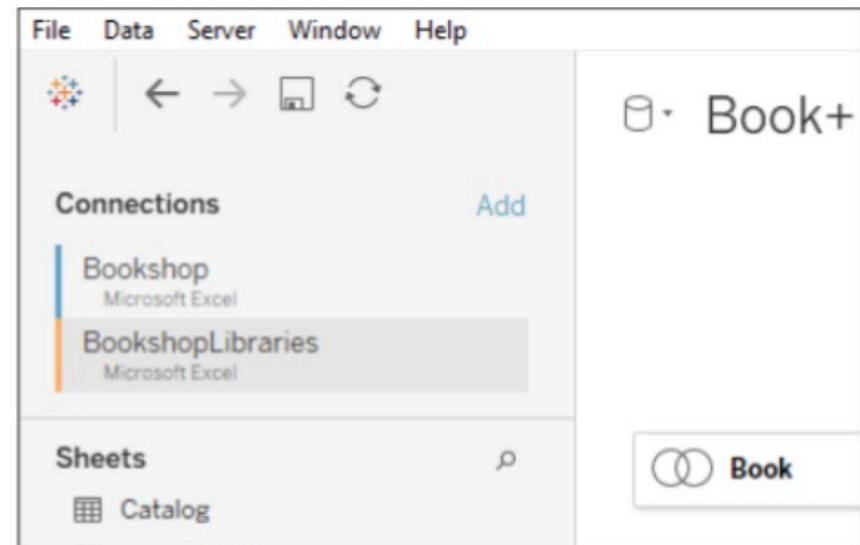
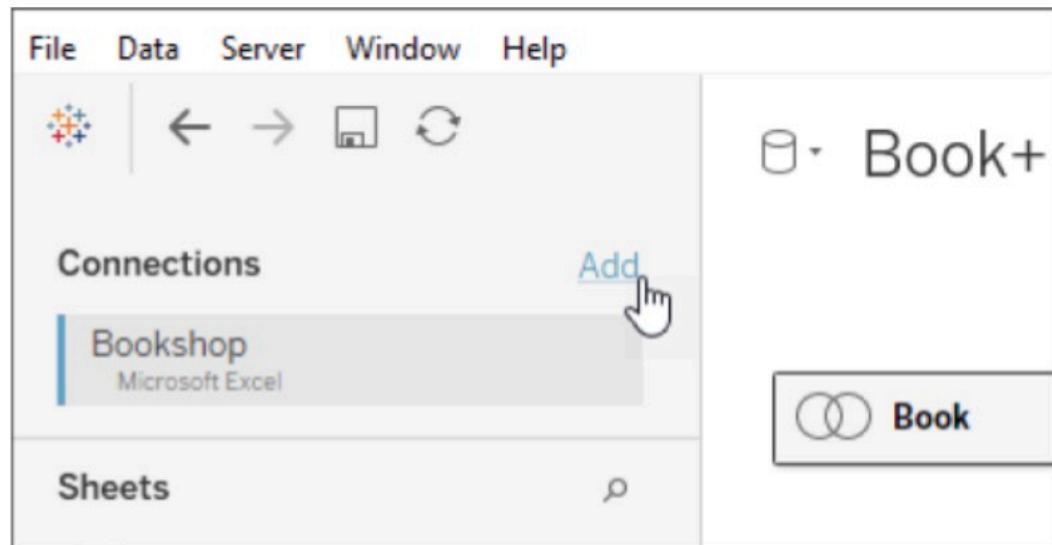


Join Types



Cross-Database Join

- ❖ Tableau allows joins from tables in different data sources, albeit with some limitations from the database side on which platforms are compatible.
- ❖ Cross-database joins require a multi-connection data source—that is, you create a new *connection* to each database before you join the tables.



Join Types

► Cross-Database Join

- ❖ Note-1: If the connector you want is not available from the **Connect** list when you're trying to add another connection, cross-database joins are not supported for the combination of sources that you want to join. This includes connections to cube data (e.g., Microsoft Analysis Services), most extract-only data (e.g., Google Analytics and OData), and published Tableau Server data sources.
- ❖ Note-2: Typically, joining tables from the same database yields better performance. This is because querying data that is stored on the same database takes less time and leverages the native capabilities of the database to perform the join.



Union

► Union

- ❖ To union your data, the tables must come from the **same connection**.
- ❖ For best results, the tables that you combine using a union must have the same structure. That is, each table must have the **same number of fields**, and related fields must have matching **field names** and **data types**.
- ❖ When field names in the union do not match, fields in the union contain null values. You can merge the non-matching fields into a single field using the merge option to remove the null values.



+



Union

► Working with Unions

- ❖ A unioned table can be used in a join.
- ❖ A unioned table can be used in a join with another unioned table.
- ❖ The fields generated by a union, **Sheet** and **Table name**, can be used as the join key.
- ❖ The field generated from a merge can be used in a pivot.
- ❖ The field generated from a merge can be used as a join key.
- ❖ The data type of the field generated from a merge can be changed.
- ❖ Union tables from within the same connection. That is, you cannot union tables from different databases.

Blending



- ❖ Data blending is a method for combining data from multiple sources. Data blending brings in additional information from a secondary data source and displays it with data from the primary data source directly in the view.
- ❖ Blends, unlike joins, never truly combine the data. Instead, blends query each data source independently, the results are aggregated to the appropriate level, then the results are presented visually together in the view.
- ❖ Because of this, blends can handle different levels of detail and work with published data sources.

Blending



- ❖ Blends are also **established individually on every sheet** and can **never be published**, because there is no true “blended data source”, simply blended results from multiple data sources in a visualization.
- ❖ Data blending is particularly useful when the blend relationship—**linking fields**—need to vary on a sheet-by-sheet basis, or when combining published data sources.
- ❖ Adding another connection to the first data source enables **relationships and joins** on the **Data Source page**. **Blending** requires two or more distinct data sources, listed independently in the **Data pane**.

Blending

► Difference between Joins & Blending

- ❖ **Data blending** simulates a traditional **left join**. The **main difference** between the two is when the **aggregation** is performed.
 - A join combines the data and then aggregates.
 - A blend aggregates and then combines the data.

Blending



❖ Note :

- Aggregating measures is straightforward—we can take the sum, average, maximum, or other aggregation of a number with ease. Measure values are aggregated based on how the field is aggregated in the view. However, all fields from a secondary data source must be aggregated.
- How does that work for dimensions? Dimension values are aggregated using the **ATTR** aggregate function, which returns a single value for all rows in the secondary data source. If there are multiple values contained in those rows, an asterisk (*) is shown. This can be interpreted as "there are multiple values in the secondary data source for this mark in the view".

Blending



User ID	District	Level	Type
1	2	3	G
2	3	4	J
4	5	6	M
1	2	3	W



Branch	Patron ID	District	Level	Fines
A001	1	2	3	10.00
B001	2	3	4	20.00
C001	1	2	3	30.00



Branch	Patron ID	District	Level	Fines
*	1	2	3	40.00
B001	2	3	4	20.00
*	1	2	3	40.00

User ID	District	Level	Type	Branch	Fines
1	2	3	G	*	40.00
2	3	4	J	B001	20.00
4	5	6	M	null	null
1	2	3	W	*	40.00



Blending

► Data Blending at a glance

- ❖ Data blending occurs on a sheet-by sheet basis.
- ❖ The order in which fields are used determines which data source is the primary versus the secondary data source.
- ❖ The primary data source is indicated with a blue check mark, any secondary data sources and fields from secondary data sources have an orange check mark.
- ❖ Linking fields can be automatically determined based on shared field names or the blend relationship can be manually created.
- ❖ Data blending behaves similarly to a left join, which may result in missing data from the secondary data source.
- ❖ Asterisks (*) may appear. This indicates multiple dimension values in a single mark, because data blending takes aggregated results and combines them in the view.

Relationships



- ❖ Relationships are a dynamic, flexible way to combine data from multiple tables for analysis.
- ❖ A relationship describes how two tables relate to each other, based on common fields, but does not merge the tables together.
- ❖ When a relationship is created between tables, the tables remain separate, maintaining their individual level of detail and domains.
- ❖ Think of a relationship as a contract between two tables. When you are building a viz with fields from these tables, Tableau brings in data from these tables using that contract to build a query with the appropriate joins.
- ❖ It's recommended using relationships as your first approach to combining your data because it makes data preparation and analysis easier and more intuitive.

Relationships



RELATIONSHIPS	JOINS
Defined between logical tables in the Relationship canvas (logical layer)	Defined between physical tables in the Join/Union canvas (physical layer)
Don't require you to define a join type	Require join planning and join type
Act like containers for tables that are joined or unioned	Are merged into their logical table
Only data relevant to the viz is queried. Cardinality and referential integrity settings can be adjusted to optimize queries.	Run as part of every query
Level of detail is at the aggregate for the viz	Level of detail is at the row level for the single table

Relationships



RELATIONSHIPS	JOINS
Join types are automatically formed by Tableau based on the context of analysis. Tableau determines the necessary joins based on the measures and dimensions in the viz.	Join types are static and fixed in the data source, regardless of analytical context. Joins and unions are established prior to analysis and don't change.
Rows are not duplicated	Merged table data can result in duplication
Unmatched records are included in aggregates, unless explicitly excluded	Unmatched records are omitted from the merged data
Create independent domains at multiple levels of detail	Support scenarios that require a single table of data, such as extract filters and aggregation

Relationships



RELATIONSHIPS

Defined in the data source

Can be published

All tables are equal semantically

Support full outer joins

Computed locally

Related fields are fixed

BLENDS

Defined in the worksheet between a primary and a secondary data source

Can't be published

Depend on selection of primary and secondary data sources, and how those data sources are structured.

Only support left joins

Computed as part of the SQL query

Related fields vary by sheet (can be customized on a sheet-by-sheet basis)

Wrap-up

- Understanding the data model
- Join Types with Union
- Blending
- Relationships





Supported data model schemas

- ❖ The data modeling capabilities introduced to Tableau in 2020.2 are designed to make analysis over common multi-table data scenarios—including star and snowflake data models—easy.



Supported data model schemas

- **Single-table**

Analysis over a single logical table that contains a mixture of dimensions and measures works just as in Tableau pre-2020.2. You can build a logical table using a combination of joins, unions, custom SQL, and so on

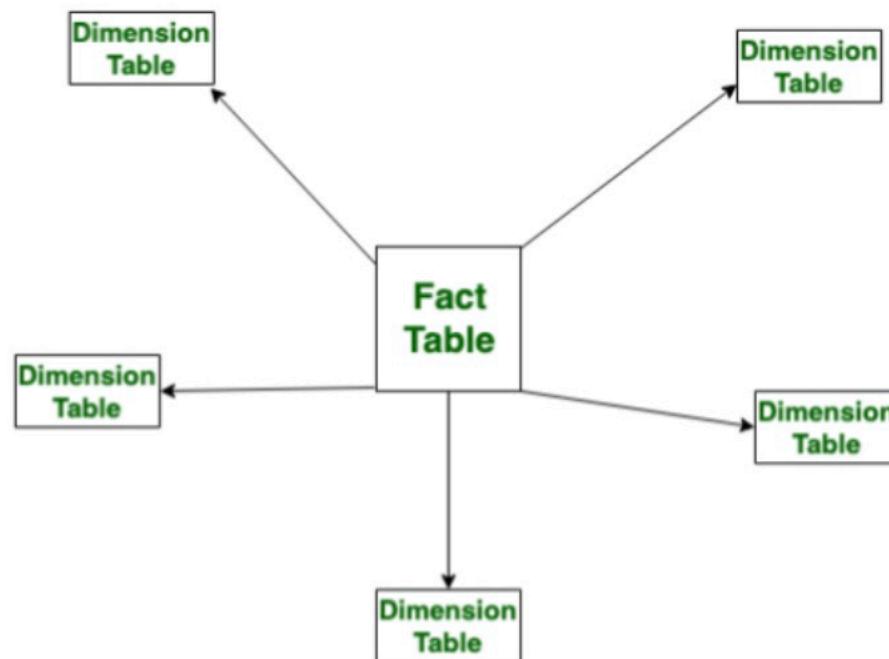




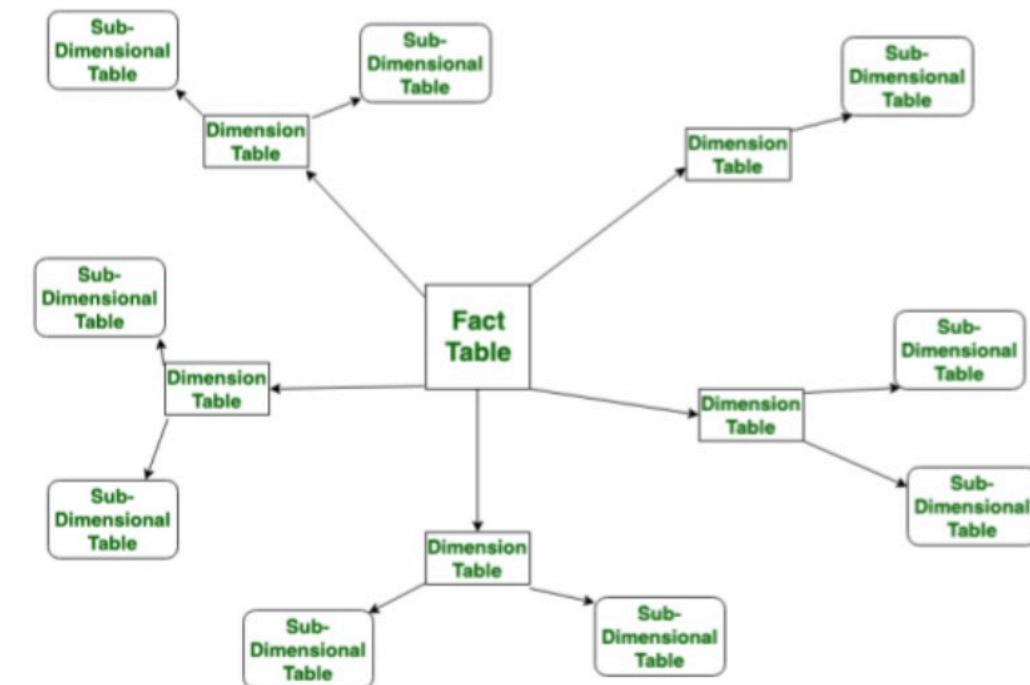
Supported data model schemas

- **Star and snowflake**

In enterprise data warehouses, it is common to have data structured in star or snowflake schemas where measures are contained in a central fact table and dimensions are stored separately in independent dimension tables.



star schema



snowflake schema