

Tableau Data Modeling:

Each dashboard and chart is based on a strong data model, so having data modeling skills is essential for business intelligence projects.

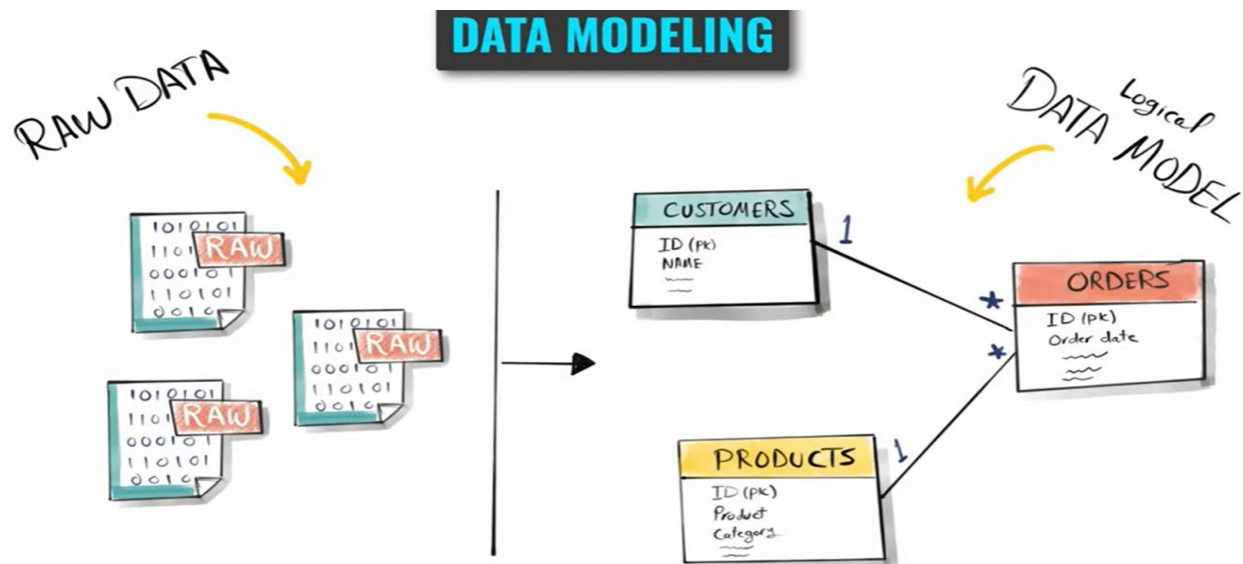
Data Model Basics:

Data is typically stored in data warehouses or data lakes within many different tables. The first step in using any visualization tool is to connect the tables and combine them into a single, large data model.

What is data modelling?

The process of organizing and representing data clearly and understandably.

Each data model has entities, which could be things like customers and products. Inside entities, we have information called attributes. We describe in the data model how those entities are connected and related to each other, and we call it relationships. This data model is a visual representation of data, which makes it easy to understand the data, because it is really important for making decisions and business improvements.



Three Types of Data Models:

Three different types of data models at different levels of abstraction.

1. Conceptual Data Model:

This type is a high-level representation of the data model without going into details on how the data model is implemented. It's like a map that shows the important entities

and the relationships. We usually use this type to explain the data models to business analysts and stakeholders to understand the big picture of the data.

2. Logical Data Model:

In this data model, we go into more detail on how the data is structured and organized. We define in this model the attributes of each entity, and it includes as well constraints and more information about the relationships between entities. This data model is used by database designers and developers as a blueprint for implementation.

3. Physical Data Model:

This represents the actual implementations of the data model. It includes all the technical details about how to store the data, like the data types of attributes, the primary and foreign keys, indexes, and so on. This data model is used by developers to create and manage the databases.

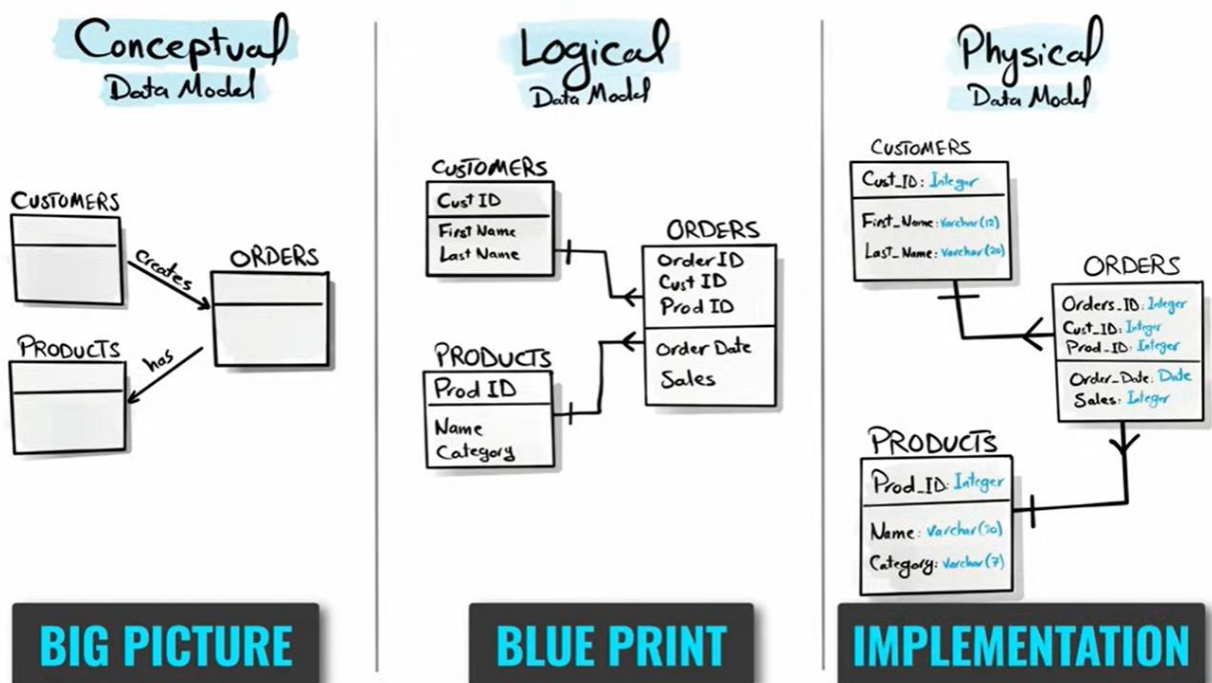


Tableau adopts both logical and physical data models in data sources, but doesn't have a Conceptual data model.

Analytical Data Models:

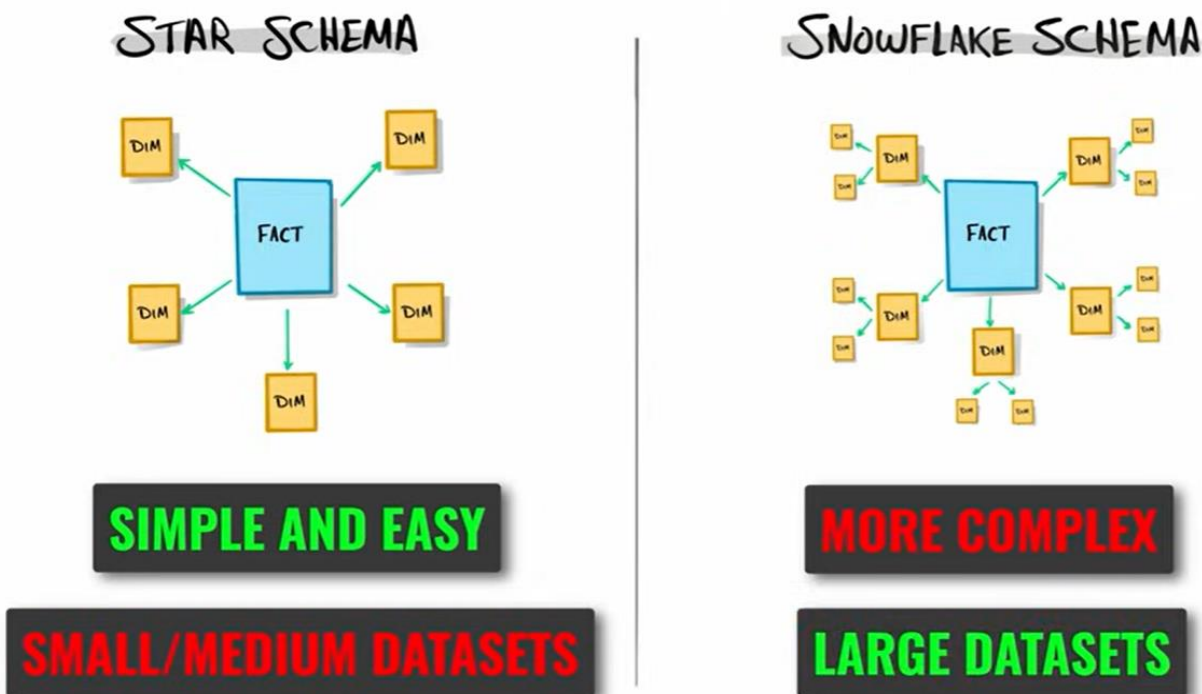
For analytics, Data Warehousing, and business intelligence, we need special data models that are optimized for queries and analytics. It should be flexible and easy to understand. So, for this, we have two data models:

1- Star Schema:

It has a central fact table and is surrounded by dimensional tables. The fact table contains events and dimensions that hold descriptive information. The relationship between fact and dimension tables forms a star shape. That's why we called it star schema.

2- Snowflake Schema:

It is very similar to a star schema, but dimensions here are broken down into sub-dimensions. Normalized tables or dimensions mean that those tables are broken down into small pieces to avoid having big tables or big dimensions, which leads to data duplication and slow performance. The shape of this data model looks like a snowflake.



A star schema is easy to understand and is commonly used for small to medium-sized datasets. Snowflake is a complex schema and is used for large datasets.

Dimension: If you have a table that contains information about a physical person or an object, like Employee, Customers, and Products, then this table is a dimension. And usually they are small tables.

Facts: If you have a table that contains events, such as Sales, Orders, Logs, ATM Transactions, and time. And usually they are huge tables.

Logical and Physical Layers:

In the Tableau data model, we have two layers: the Physical layer and the Logical Layer.

1. Physical Layer:

We might have a couple of physical tables, and we can combine them in Tableau by using two methods, either joining the tables or union between them.

2. Logical Layer:

It is a top-level layer and provides us with an abstract to hide all the details in the physical layer. This is especially nice if we have a lot of tables in the physical layer. So the result of merging the tables using joins and unions in a physical layer is going to be presented in the logical layer with a single table, and we call it a logical table.

So, that means we are going to have two logical tables, the first is gonna present three tables after doing a join, and the second gonna present two tables after doing a union (Join jin tables me ho ga unka alg Logical table aur union walo ka alg).

In Tableau, we have **relationships** to combine those two logical tables. An important thing to understand is that in the logical layer, we cannot merge two tables into one. So, we use relationships to combine them, and everything remains the same. These two layers can be found in the data source. On top of the data source, we built a Visualization. You can start building visuals using the data available in the logical layer.

But sometimes you are working with large projects, you're gonna build another data source with a new data model. So, here important to understand that not all logical tables come from the physical layer; they can directly come from the data source.

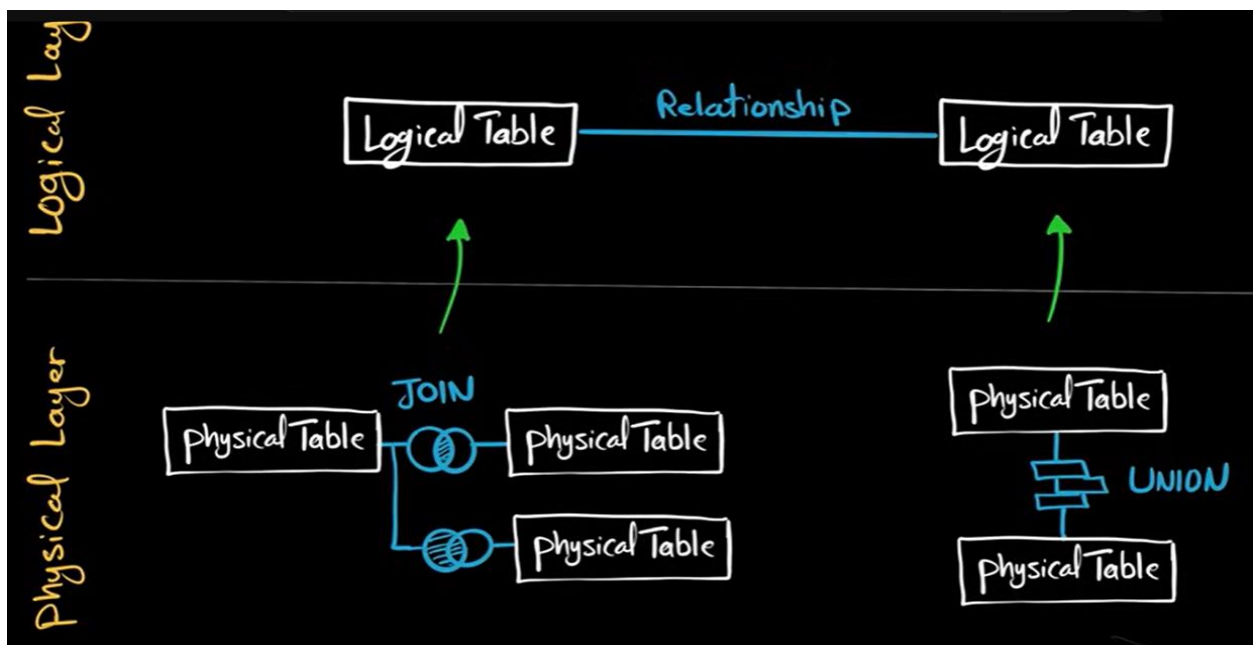
To build one visualization from both of the data sources and models, we have to connect those two data models, and we can do that at the visualization level, where Tableau offers a unique and last method of connecting and combining tables called **data blending**.

- ➔ Tableau offers four different methods to combine and connect tables at different layers. In the physical layer, we have joins and unions. In the logical layer, we have relationships, and at the visualization level, we have data blending.

Navigate through Layers:

In the data source page, by default, we're going to add the logical layer to the data model. It means anything we drag and drop in our data model is considered a logical table. For example: Hm customer ka table drag and drop krty tou wo aik logical table, orders ka krty to we dosra logical table, and Tableau did create a relationship between them.

We can go to the physical layer for this, **double-click on the logical table**, e.g, Customers. Tableau tells here that the customer is made up of only one physical table. AB hm koi b table drag and drop krty data model me tou wo aik physical table consider hta aur tableau unmy aik relationship nhi Joint create krta, and of course we can create union between them too. Ye joint krny k bd hm dekh skty k Logical Table Customer 2 Physical tables se mil k bna.

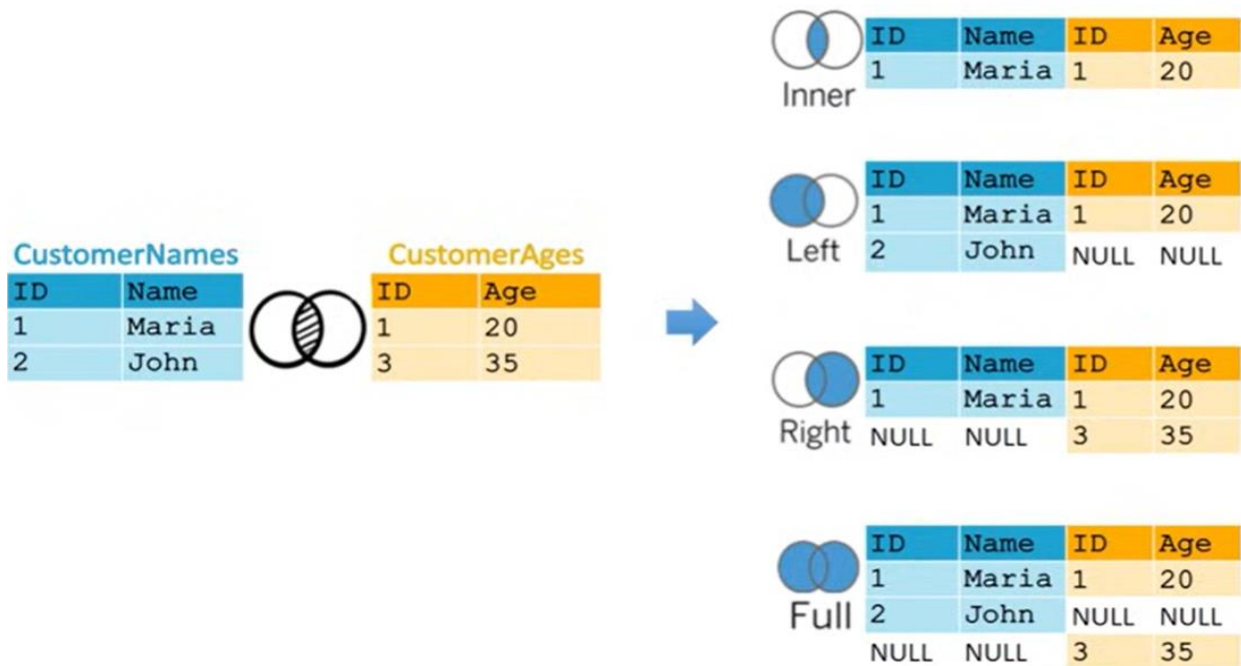


Joins:

For joins we need one common column in both tables and join type.

- 1- Inner : Matching rows from both tables
- 2- Left : All the rows from left table, and only matching rows from right table.
- 3- Right : All the rows from right table, and only matching rows from left table.
- 4- Full : Everything from both tables.

JOINS



Join in Tableau:

Tableau me by default logical alayer hti, Customers k table pr click kry gy physical layer pr a jaye gy , orders k column ko drag and drop krein gy tou wo customers k sath inner join bna ly ga, isy hm change b kr skty, join pr click kry gy, common column Customer id hy agr change krna tou ho jaye ga ya koi second column b add kr skty. Jesy hi back kry gy logical layer pr tou wo show kry ga join between customers and orders.

Unions:

The two things are necessary -> Same Number of fields, Same data types.

Suppose hmary pas 2 tables hy orders 2022, and orders 2023 hmy aik table bnana orders k name se. tou union use kry gy orders me phly 2022 ki rows aye gi phr usky neechy 2023 ki.

Unions in Tableau:

- ➔ One way, Orders table ko logical table me drag and drop krna, double click kr k Physical layer me haye gy, Hm ny Orders and orders_archives ko union krna tou orders_archives

ko drag kr k parallel nhi lana orders k blky is y orders k neechy lana tou union me jaye ga orders k neechy 2 grey lines bn jaye gi. Hm neechy result table me dekh skty k sb se last me column Table name hy jis me orders and orders_archives show ho rha hna.

➔ Second way, data source me left side pr New union ka option hna , usy click krna aik Dialogue box open ho ga, two options hny Manual and Automatic. **In Manual** hm Orders and Orders_archive ko drag kry gy and then click ok. We get the same results without going to physical layer. **In Automatic**, suppose hmary pas Orders ki 100 files hy from 2021-2025 tou sb ko drag krna difficult task hy, Isy bchny kaliye hm tableau kaliye rule define kry gy aur ye hmary liye files dhoondhy ga and union kry ga. Hm automatic ma jaye gy-> Files me-> Include -> Orders k name se start ho rhy sb tables (Orders_2000, Orders_2001) tou hm Likhy gy Orders*-> folder poth dena k kaha search kry. Then ok. Result table k sb se last me new column hna path , usy sheet me ja k drag kry gy tou hm dekh skty k Orders and Orders_archive ka path hy.

Relationships:

We can connect two logical tables in the logical layer using relationships. Ye 2 tables k between aik contract hta hy, tableau isny data leny se phly contract check krta hy. Ye dono table aik dosry se separated rhty isny koi resulted table nhi bnta hy. Agr hm visualization level me aik column table A se lety aur aik Table B se tou Tableau phly relationship check kry ga phr data de ga. Aur is trah data visualization level pr combine ho ga.

Relationships in Tableau:

Data source page (Logical layer) me orders ko drag krna then customers ko tou dono k drmiyan noodle bnny lgni jo k relationship show kr rhi hti. Hm iski details kaliye noodle pr click kr skty aur neechy metaata me a jaye gy -> Dropdown se relationship select krein gy, there are three things that we are going to set up as relationships. First is key, it is a common column in both tables. We can change it or add one more.

In performance options, we have Cardinality and Integrity (* If you leave it here as it is, as the default, nothing is going to go wrong, you will not lose any data, you don't have t change anything here unless you want to optimize the performance*) In Cardinality we defined as Many or one and in Integrity we have some records match and all record match.

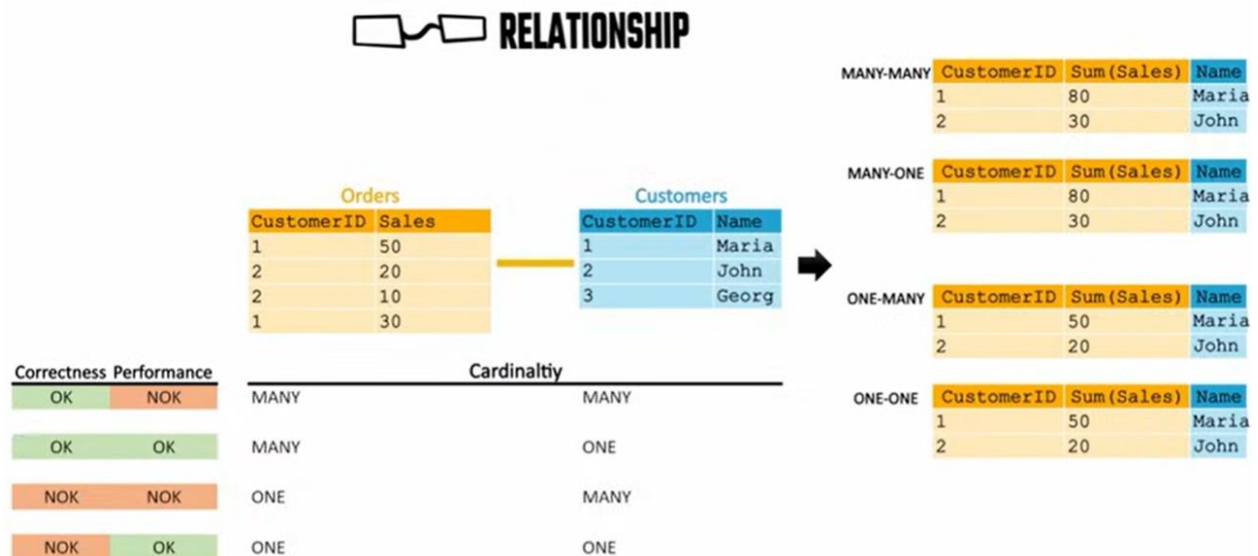
Deep Dive into Cardinality:

Hmary pas 2 tables hy: Customers and orders aur key Customer id hy. In cardinalities, we have two options: either Many or One. To decide which one is the correct we have to do

data profiling -> “The process of examining and investigating the data to understand the content of tables.” After data profiling, many or one ko choose krna asaan ho jata. We have a rule: **Use many when there are duplicate values in the key. One: When the key has unique values.**

Orders me k customer id me unique values nhi hy tou many lgaye gy, Jb many hta tou sary table ko scan krta. Customer table has unique values in the key, so One lgaye gy.

- ➔ We have two criteria: Correctness and Performance. Correctness is more Important than performance.
- ➔ In the first scenario Output was correct, but the performance was bad. Tableau's right side me full table scan krta hy.
- ➔ In the second, both output and performance were ok, since it was a many – one.
- ➔ In the third (one-to-many), output was not ok, and performance was bad.
- ➔ In the fourth(one-to-one), output was not ok, and performance was ok.
- ➔ Tableau always recommends staying in a many-to-many relationship.



Deep dive into Integrity:

Integrity has two values: Some records match, All records match.



By default, Tableau suggests the first one.

Data Blending:

Visualization level me hmary pas aik visual hy jiskaliye hm data source A se data fetch krty, since this data source was the first one to be queried and to be used, Tableau going to call it a primary data source. In Tableau, anything primary is gonna get the blue color. Ab hm aesi situation k is visual kaliye aik aur data source se b data lena hy tou usy data fetch kry gy aur tableau isy secondary mark kr de Gawith orange icon. Ab jb k hm do Data sources k sath kaam kr rhy tou inhy connect krna ho ga, hm is situation me aik bht hi unique way use krty call Data Blending. Data blending me Tableau left join use krta, and it is fix. It means all the data from the primary data source and only matching records from the secondary. **It is a method of combining data at the visualization level from two different sources using the left join.**

How it works:

Hmary pas 2 data sources hy Products, and Product_prices. Worksheet me jaye gy tou hm phly jis data source se data drag and drop kry gy wo primary ho ga, suppose hm Products se krty hain. Green icon aye ga , aur jb hm Product_prices se kry gy tou orange. And any field we are using is gonna be marked in orange. Suppose hmy lgta k Product ID dono tables me common nhy hy isy change krna tou go to Data-> Edit Blend Relationships -> Click-> New Window and here we have two options automatic and Custom. Agr hm Automatic rkhty tou Tableau khud hi

figure kr leta common key. But hmy change keni tou custom pr jaye gy aur phly se exist key Product id pr double click kry gy tou primary aur secondary dono ki fields appear ho jaye gi. Ab Suppose k hm Product Id common key tou rkhty hi hm aik aik aur b add kr dety jo k common nhi, ok kr wapis jaty tou aik broken chain ka sign show rha ho ga agy. If you want to activate, just click on it. Again, click to deactivate it. Anything we do in data blending is relevant only to this worksheet. If I go to another, it's completely resets. Data Blending offers great Flexibility.

Join combines Fields, Union combines Rows.

Joins first combine the data, then aggregation gonna happen.

Data Blending first performs aggregation, then combines.

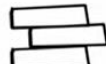









Data Blending: You will not get any duplicates.

Measures can be aggregated. Dimensions cannot be aggregated.

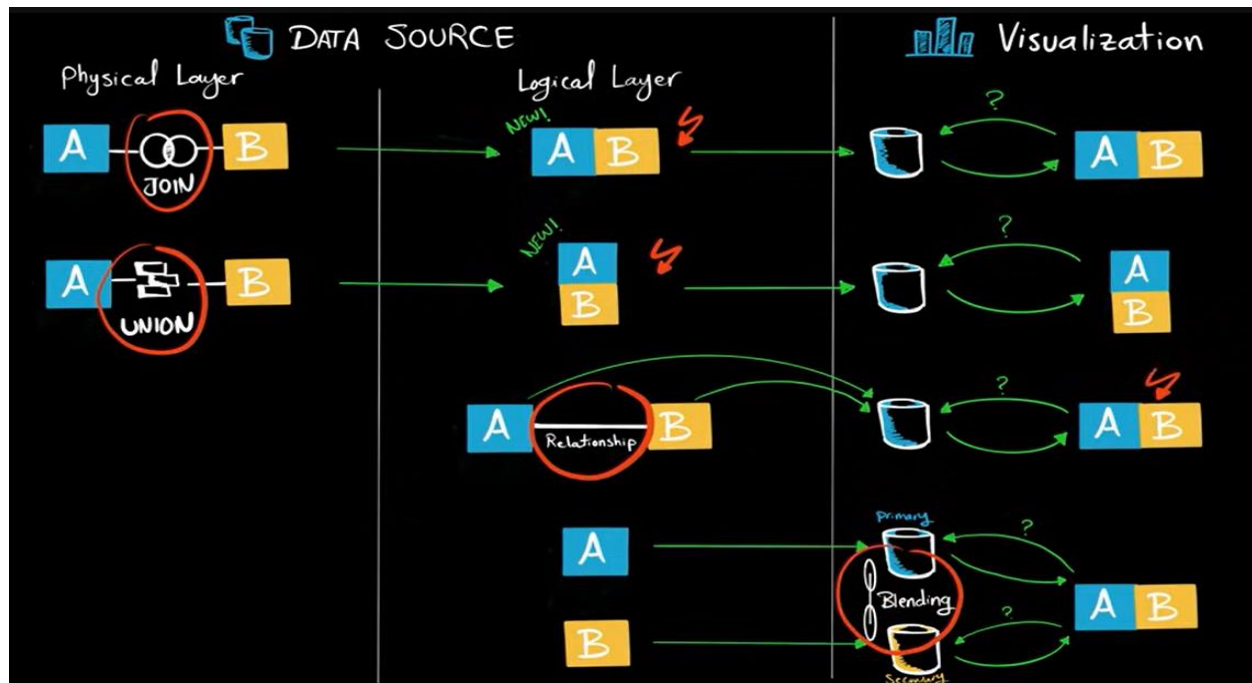
When multiple matches exist, Data Blending adds *

In joins, things are going to get really static, and we might lose data.

In Relationships, we will get more flexibility and will not lose data.

	 UNION	 JOINS	 RELATIONSHIP	 BLENDING
 Layer	Physical Layer	Physical Layer	Logical Layer	Visualization Level
 Multi-Data Sources	✗	✗	✗	✓
 Merge Tables?	✓	✓	✗	✗
 Flexibility	Static	Static	Flexible	Very Flexible (Page by Page)
 Join Types	n/a	Inner, Left, Right, Full	All	Left
 Rank	n/a	3	1 (Default)	2

Summary:



Create a small data source on a small dataset:

Go to Tableau data source-> text file-> small dataset folder -> orders.csv tou sb files load ho jaye gi.. hm scratch se start krein gy. * Darg Fact table First. Then relate the Dimension tables to that Fact.* So our fact table is orders. Phr dimension tables jo k Customers and Products ha ko grap krein gy. Hm Customers ko drag krty tou aik relationship bn jata. Ab hm cardinality dekhy gy, age hmara data ki quality poor ha aur data profiling nhi ki tou aesy hi chor do, otherwise, *Fact side is many and Dimension side is one.* Usually we have unique customers and products. Tou hm cardinality many to one kr dein gy. Integrity stuff as its rkhy gy. Hm sb se important thing Format check krna hta. Agr Fields like Sales, Quantity, orders k oper # bna tou means k wo number hy usky oper click b kr k dekh skty. Lkn agr hm dekhty k for example sakes ki filed string me hy tou means k issue, Hmy ye smjhna ho k some countries like Europe me decimal number me “,” hta jb k Asia me “.” Hta ha. Now, how to fix it -> Data model me orders k table pr right click kr k Text file properities me jana aur local me jana. Wha se matching locale change kr deni. Tableau correct format understand kr le ga.

Second dimension Product ko drag krein gy, Orders k sath many to one relationship banna. Ab hmaar next table Product_details ka hy, yha hmary pas two options ya tou Products k sath hi relationship ya Join. (If Tables refers to same entities/things consider using Joins/unions.)

Ab next table Orders_Archive ka hy. Yha b option union ka hy, kiun k Orders aur Orders_Archive ki same fields aur same data types ha.