

Tableau

Module I

Tableau Overview & Architecture

Working with Metadata & Data Blending



Participant's Introduction

- Name
- Professional experience
- Current role
- What do you know about Tableau?
- What do you expect from the session?

LET'S GET TO
KNOW EACH OTHER



Agenda

Introduction to Data Visualization & Power of Tableau

- What is Data Visualization & Data Analytics?
- Power of Tableau & its Overview
- Tableau Products
- Connecting to Data Sources
- Live Vs. Extract Connection
- Tableau Joins
- Tableau Data Types & Data Preparation
- Tableau Field Types & File Types

Tableau Architecture

- Installation of Tableau Desktop
- Architecture of Tableau
- Interface of Tableau
- How to start with Tableau
- Ways to share and export the work done in Tableau

Working with Metadata & Data Blending

- Connection to Excels, PDFs and Cubes
- Managing Metadata and Extracts
- Data Preparation and dealing with NULL values
- Data Joins (Inner, Left, Right and Outer) and Union
- Data Blending
- Cross Database Joins
- Data Extraction



Introduction to Data Visualization & Power of Tableau

What is Data Visualization?



- Data visualization is the process of converting raw data into easily understandable formats of information that enable fast and effective decision-making. Charts, diagrams, pictures, etc. are examples of data visualization.

A	B	C	D	E	F	G	H	I	J	K	
1	iga_id	pop_2006	lga	state	zone	area_sq_km	surveying_ef	pop_density	unique_lga	longitude	latitude
2	1	100180	Abadam	Borno	Northeast	3967.05423	Others	25.2529948	borno_abada	13.3256741	13.4282932
3	2	58444	Abaji	FCT	North-Central	992.361242	113	58.8938761	fct_abaji	6.85341375	5.87756561
4	3	139090	Abak	Akwa Ibom	South-South	189.984563	Others	732.112113	akwa_ibom_	7.7722792	5.01988684
5	4	151723	Abakaliki	Ebonyi	Southeast	584.055167	Others	259.775118	ebonyi_abakaliki	8.21988187	6.2586811
6	5	107488	Aba North	Aba	Southeast	22.7750611	Others	4719.54826	abia_abanorth	7.37164362	5.10690037
7	6	423852	Aba South	Aba	Southeast	49.1665418	Others	8645.74053	abia_abasouth	7.35065009	5.08134569
8	7	201329	Abeokuta Nc Ogun	Ogun	Southwest	808.831493	Others	248.913404	ogun_abeokutanc	3.1735877	7.22300486
9	8	250278	Abeokuta So Ogun	Ogun	Southwest	71.4715348	Others	3501.78572	ogun_abeokutaso	3.35694358	15.1311807
10	9	144802	Abi	Cross River	South-South	281.772825	Others	513.896257	cross_river_abi	8.0266891	5.89308847
11	10	195652	Abio Mbaise Imo	Imo	Southeast	184.48387	Others	1060.53716	imo_abio_mbaise	7.25224758	5.43261623
12	11	282988	Abua Ondu	Rivers	South-South	704.338115	Others	401.778626	rivers_abua_o	6.56209631	4.83853956
13	12	202194	Adavi	Kogi	North-Central	718.503598	Others	281.409864	kogi_adavi	6.44459043	7.67423257
14	13	308621	Ado-Ekiti	Ekiti	Southwest	293.6513	Others	1050.97781	ekiti_adoe	5.27897457	7.61042627
15	14	526565	Ado-Odo/Ot. Ogun	Ogun	Southwest	878.916456	148	599.106998	ogun_adodo	3.08462514	6.61884102
16	15	134173	Afijio	Oyo	Southwest	722.893666	Others	185.656809	oyo_afijio	3.90125509	7.74211826
17	16	156611	Afikpo North	Ebonyi	Southeast	240.157467	Others	652.117972	ebonyi_afikpo	7.9233965	5.87842851
18	17	157072	Afikpo South	Ebonyi	Southeast	377.675196	113	415.891755	ebonyi_afikposouth	7.81363734	5.83719147
19	18	132907	Agale	Niger	North-Central	1904.35471	Others	69.7910949	niger_agale	6.44045921	8.94623535
20	19	115523	Agatu	Benue	North-Central	1014.60843	Others	113.859689	benue_agatu	7.88215837	7.86983845
21	20	57413	Agwara	Niger	North-Central	1539.669	148	37.2891837	niger_agwara	4.52764087	10.7410218
22	21	459939	Agege	Lagos	Southwest	11.214557	Others	41012.6765	lagos_agege	3.31600267	6.62357986
23	22	370172	Aguda	Anambra	Southeast	195.102574	Others	1897.31992	anambra_aguda	7.0880433	5.98862094
24	23	170902	Ahiazu-Mbabi Imo	Imo	Southeast	113.716283	Others	1502.88063	imo_ahiazu	7.26418199	5.56020001
25	24	166747	Ahoada East	Rivers	South-South	341.173834	Others	488.744984	rivers_ahoada	6.6410198	5.06535919
26	25	249425	Ahoada West	Rivers	South-South	403.45939	Others	618.215875	rivers_ahoada	6.51660569	5.05176308
27	26	122321	Ajokuta	Kogi	North-Central	1362.39948	113	89.7835045	kogi_ajokuta	6.5963902	7.51474121
28	27	684105	Ajegun-Ifelé	Lagos	Southwest	12.349309	148	55395.8493	lagos_ajegun_ifele	3.3374815	6.45356193
29	28	174137	Ajingi	Kano	Northwest	713.799255	Others	243.957946	kano_ajingi	9.0690277	11.9731313
30	29	151125	Akamka	Cross River	South-South	5001.71991	Others	30.2146067	cross_river_akamka	8.53398001	5.40945643
31	30	211359	Akinleye	Oyo	Southwest	518.870311	Others	407.344563	oyo_akinleye	3.93139439	7.56243229

Raw Data



Converted to Meaningful Visual Information

Benefits of Data Visualization

- Benefits of data visualization are:



Allows users view several different perspectives of the same data



Makes it possible to interpret vast amounts of data



Offers the ability to note exceptions in the data



Allows the user to analyze visual patterns in the data



Allows to explore trends within a database. It also allows analysts to navigate through data and visually orient themselves to the patterns in the data.

What is Visual Analytics?



Visual analytics is a field in information visualization that focuses on:

Analytical reasoning

Interactive visual interfaces



What is Data Analytics?



Data analytics is the process of examining data to identify the patterns. It is the discovery, interpretation and communication of:

Meaningful information to draw

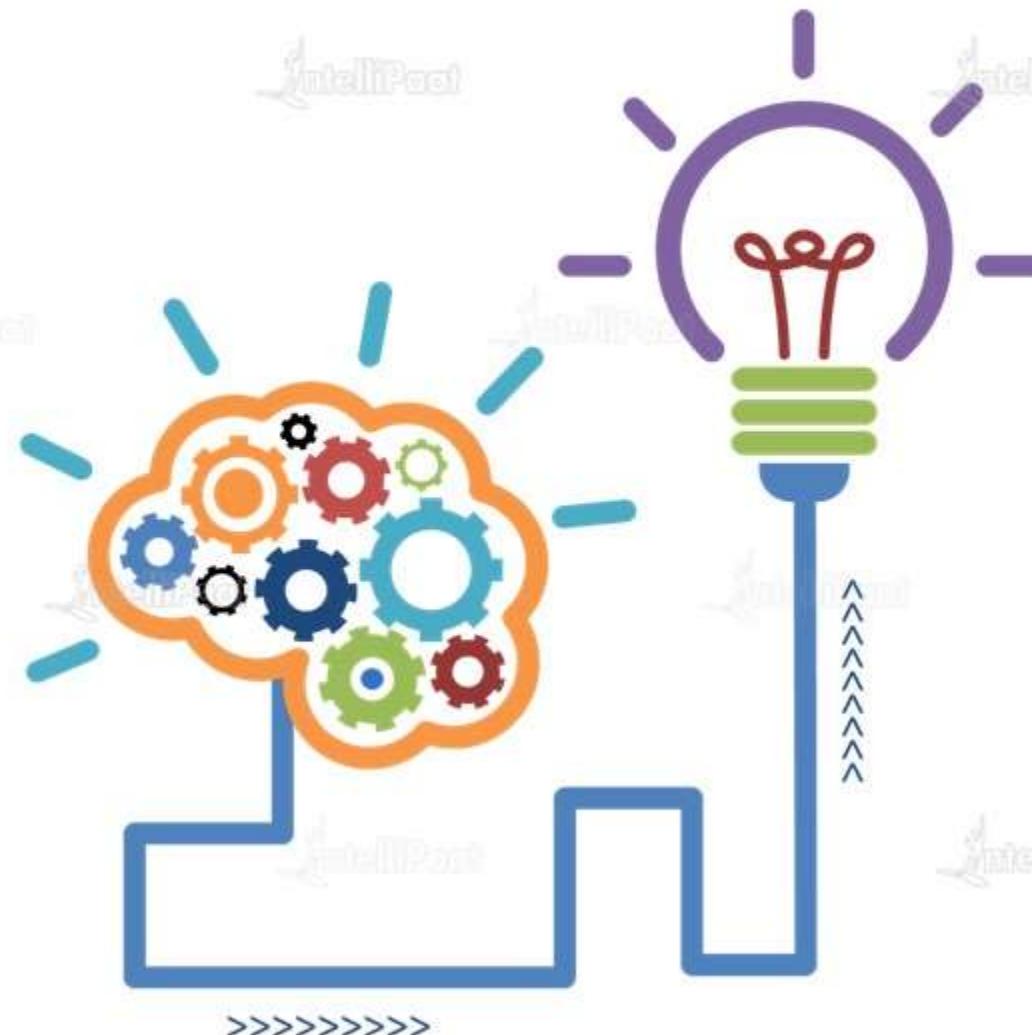
Business decisions usually with the help
of specialized software



A Picture is Worth a Thousand Words



Humans are very
good in solving
complex problems.



Visual representations
enhance the human
cognitive process.

Business Needs



Businesses realize that a visual can communicate:

- More information than a table
- In a much smaller place

DATA VISUALIZATION



Data visualization leads to better:

- Understanding
- Business decisions, accurately and quickly

WE UNDERSTAND
YOUR NEEDS



Appealing and dynamic visuals that:

- Manipulate and interact directly and easily with the data
- Can update automatically
- Facilitate data analytics and data discovery
- Identify trends and patterns
- Act on emerging trends fast
- Help in business decisions

Real-time Scenarios



Data visualization has:

- Replaced manual reporting efforts
- Provided real-time insight into acute hospital services
- Uncovered hospital patterns and trends

Healthcare data has the potential to:

- Reduce costs
- Enhance quality
- Improve the patient experience

The challenge is how to get insights and actions from information.

Tableau empowers everyone in a given healthcare organization to find the right path forward.

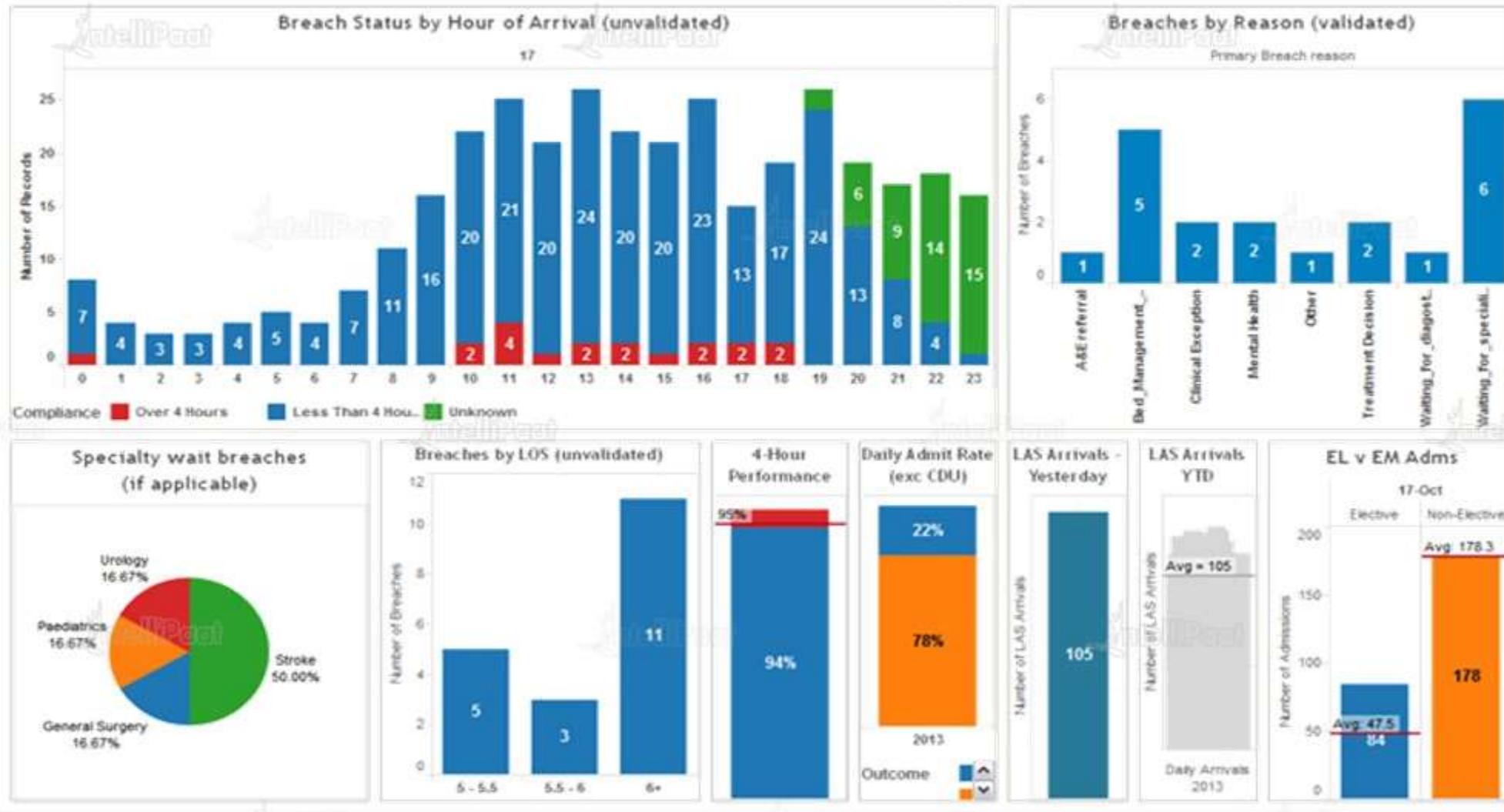


Healthcare Dashboard

IntelliPaat

IntelliPaat

ED Dashboard - Daily





Introduction to Tableau

What is Tableau?



A Tableau is an interactive Business Intelligence (BI) software used for data visualisation and data analytics.

An American computer software company founded in January, 2003, by Chris Stolte developed this software.

A Tableau can:

- Analyze and transform data into interactive visualisations in a few seconds or minutes
- Query a range of data including relational databases, OLAP cubes, cloud databases and spread sheets to generate a number of graph types

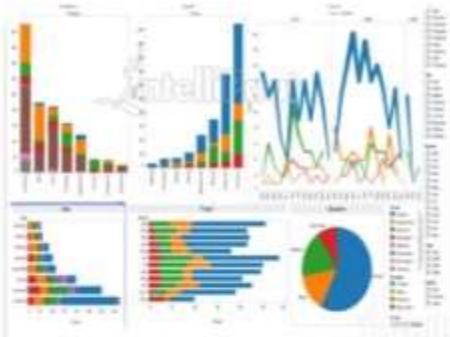


Power of Tableau



- Tableau is an easy-to-learn and easy-to-use tool with the drag-and-drop technology. It is a user-friendly interface that helps generate visual reports in easy-to-understand formats. This encourages business users to create their own reports.
- Tableau stores data in memory (RAM or Flash memory). Therefore, the processing speed is much faster than those that store data in disks.
- More than 10,000 organizations get rapid results with Tableau in the office and on the go.
- Tens of thousands of people use Tableau Public to share data in their blogs and websites.
- Tableau visualizations are automatically streamlined for mobile devices (iPad, Android Tablets, etc.), thus enabling mobility so that data and reports can be accessed from anywhere at any time.

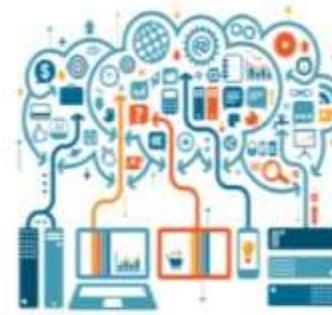
Power of Tableau



Analysis



Source



Integration



Learn



Performance



Mobile



Mobile



Mobile



Tableau Vs. Other Vendors

IntelliPaat

IntelliPaat

- Tableau is an industry leader in Business Intelligence and Analytics.
- Tableau is recognized as a leader in Gartner Magic Quadrant for Business Intelligence and Analytics Platforms for the 8th consecutive year.



Installing Tableau



- To install Tableau:

Go to
<https://www.tableau.com>

Download the
Tableau
Public/Tableau
Desktop software

Install the software

Explore the Tableau
Public Viz of the Day
Gallery

A screenshot of the Tableau website homepage. At the top, there's a navigation bar with links for Products, Solutions, Learning, Community, Support, About, PRICING, SIGN IN, TRY NOW, and a search icon. Below the navigation, there are four main product cards: Tableau Desktop, Tableau Prep, Tableau Server, and Tableau Online. Each card has a brief description and a 'LEARN MORE' button. To the right of these, there's a sidebar titled 'MORE TABLEAU PRODUCTS' containing links for Tableau Public, Embedded Analytics, Developer Tools, and Pricing. The 'Tableau Public' link is highlighted with a red box.

Tableau Products



Tableau Desktop

Create



- Ad hoc analytics, dashboards, reports and graphs
- Explore, visualize and analyze data
- Create dashboards to consolidate multiple views
- Deliver interactive data experiences

Tableau Server

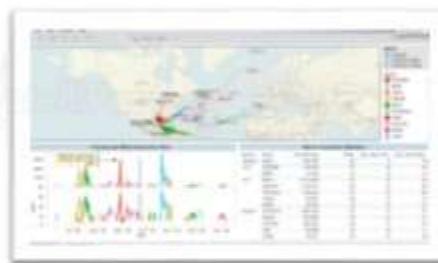
Share - Web



- Used for on-premises deployments.
- Business Intelligence solution scales to organizations of all sizes
- Share visual analytics with anyone with a web browser
- Publish interactive analytics or dashboards
- Secure information and manage metadata
- Collaborate with others

Tableau Reader

Share - Local



- Share visualizations and dashboards on the desktop
- Filter, sort and page through the views
- “Acrobat for Data”
- Free download

Tableau Public

Share - Everyone



- Create and publish interactive visualizations and dashboards
- Embed in websites and blogs
- Free download and free hosting service

Tableau Online Share - Cloud



- Used for cloud based deployments.
- Managed and maintained by Tableau Company.
- High level of security.
- Publish interactive analytics or dashboards
- Secure information and manage metadata
- Collaborate with others

Tableau Prep ETL Tool of Tableau



- Extract-Transform-Load Tool.
- Used for preparation and modification work on data before creating dashboards with Tableau desktop or Public.

How Does Tableau Work?



- Tableau is based on three simple concepts:

Connect

Connect Tableau to any data source that you want to analyze.

Analyze

Analyze the data the way you want, such as filter, sort, perform aggregations on data, summarize and so on.

Share

You can share results with others by:

- Sharing workbooks with other Tableau users
- Pasting results into applications such as Microsoft Office
- Printing to PDF
- Using Tableau Server to publish or embed your views across organization.

Connecting to Data



Click the Tableau icon on the upper left-hand corner of any page.

STEP 1

Under Connect, you can:

- Connect to data stored in a file, such as Microsoft Excel or Access.
- Connect to data that is stored on a server, such as Tableau Server, Microsoft SQL Server or Google Analytics.
- Connect to a data source that you've connected to before.
- Tableau supports the ability to connect to wide ranging data stored in a wide variety of places.

STEP 2

Under Sample Workbooks, view sample dashboards and worksheets that come with Tableau Desktop.

STEP 3

Under Open, you can open workbooks that you have already created.

STEP 4

STEP 5

Under Discover, find additional resources, such as video tutorials, forums or the “Viz of the week” to get ideas about what you can build.



Connecting to Data

Connect

To a File

- Excel
- Text file
- Access
- JSON file
- PDF file
- Spatial file
- Statistical file
- More...

To a Server

- Tableau Server
- MySQL
- Oracle
- Amazon Redshift
- Microsoft SQL Server
- More... >

Saved Data Sources

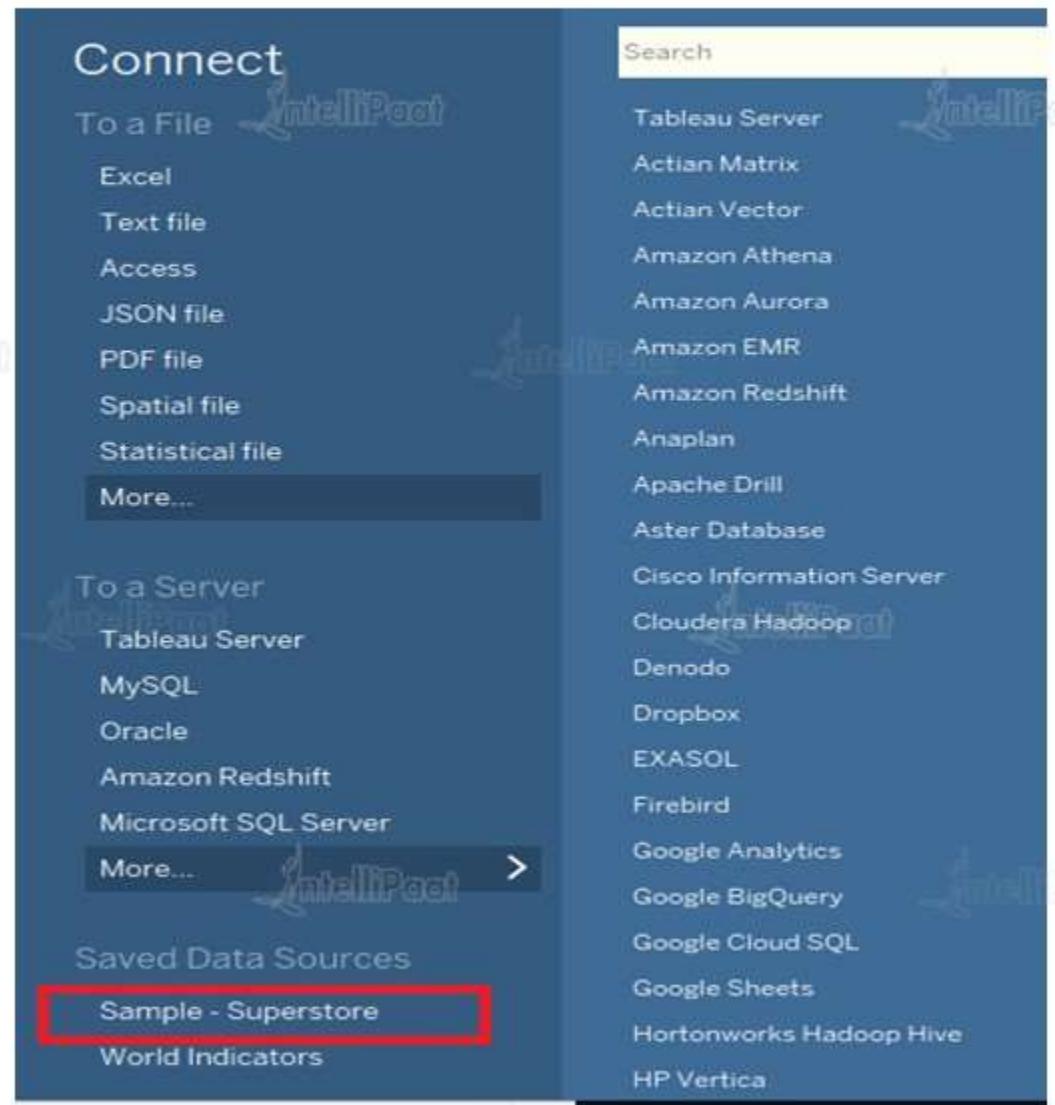
- Sample - Superstore
- World Indicators

Search

- Tableau Server
- Actian Matrix
- Actian Vector
- Amazon Athena
- Amazon Aurora
- Amazon EMR
- Amazon Redshift
- Anaplan
- Apache Drill
- Aster Database
- Cisco Information Server
- Cloudera Hadoop
- Denodo
- Dropbox
- EXASOL
- Firebird
- Google Analytics
- Google BigQuery
- Google Cloud SQL
- Google Sheets
- Hortonworks Hadoop Hive
- HP Vertica



About the Dataset



- The Sample - Superstore dataset that comes with every download of Tableau is used to explain the basic concepts in Tableau.

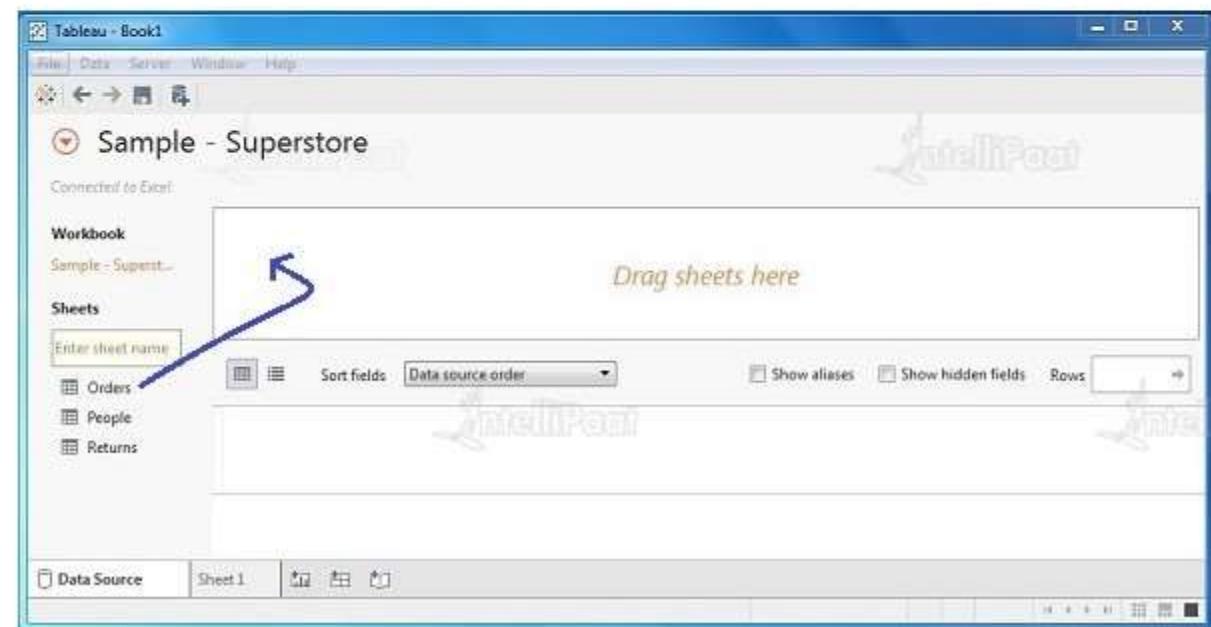


Tableau: Data types



- All fields in a data source have a data type. The data type reflects the kind of information stored in that field, for example, integers (410), dates (1/23/2015) and strings ("Wisconsin"). The data type of a field is identified in the Data pane by one of the icons shown in the image.

Icon	Data type
Abc	Text (string) values
Calender	Date values
Clock	Date & Time values
#	Numerical values
T F	Boolean values (relational only)
Earth	Geographic values (used with maps)

Using Data Interpreter



The screenshot shows the 'Connections' section with two entries: 'little_crime1_excel' and 'little_crime2_excel'. Below it, the 'Sheets' section lists 'little_crime2_excel' with a note: 'Data Interpreter might be able to clean your Excel workbook.' A large orange arrow points from this interface to the next one.

Click on
Review the results

The dialog box shows a checked checkbox labeled 'Cleaned with Data Interpreter' and a link 'Review the results. (To undo changes, clear the check box.)'. It also lists 'little_crime2_excel' and 'little_crime2...cel A1:G10000'. An orange arrow points from this dialog to the final screenshot.

An Excel file is **generated** with the
interpreted null values.
Look at the 'A's!

A	B	C	D	E	F	G
1	ID	Psa	Dc_Key	Location_IUCR_Gen	Text_Gen	Police_Dis
2	0	1	2,02E+11 800 BLOC	600	Thefts	13
3	1	3	2,02E+11 1500 BLOC	600	Thefts	16
4	2	3	2,02E+11 1500 BLOC	600	Thefts	16
5	3 A		2,02E+11 GATE C-18	600	Thefts	
6	61	1	2,02E+11 900 BLOC	600	Thefts	3
7	62	1	2,02E+11 100 BLOC	600	Thefts	20
8	63	2	2,02E+11 5200 BLOC	600	Thefts	20
9	64	1	2,02E+11 N 17TH ST	800	Other Assaults	
10	65 A		2,02E+11 PIA GATE	800	Other Assaults	
11	66	1	2,02E+11 100 BLOC	600	Thefts	20
12	67	4	2,02E+11 1400 BLOC	600	Thefts	16
13	68	2	2,02E+11 5900 BLOC	2600	All Other	15
14	69	2	2,02E+11 1100 BLOC	600	Theft from	11
15	70 A		2,02E+11 PIA GATE	600	Thefts	
16	71 A		2,02E+11 GATE E-8	600	Thefts	
17	72 A		2,02E+11 PIA TERMI	600	Thefts	
18	73 A		2,02E+11 98 PIA WA	600	Thefts	22
19	74 A		2,02E+11 GATE A-7	600	Thefts	
20	75	2	2,02E+11 400 BLOC	200	Rape	4
21	76	2	2,02E+11 1200 BLOC	200	Rape	5
22	77	2	2,02E+11 3200 BLOC	1100	Fraud	7
23	78	2	2,02E+11 2800 BLOC	600	Thefts	16
24	79	2	2,02E+11 5200 BLOC	800	Other Ass	20
25	80	1	2,02E+11 N 410 N 6	300	Robbery No Firearm	

Tableau: Field Types



- When you connect to a new data source, Tableau assigns each field in the data source to either the dimensions area or the measures area of the data pane, depending on the type of data the field contains.
- If a field contains categorical data (such as names, dates or geographical data), Tableau assigns it to the dimensions area.
- If a field contains numbers, Tableau assigns it to the measures area.

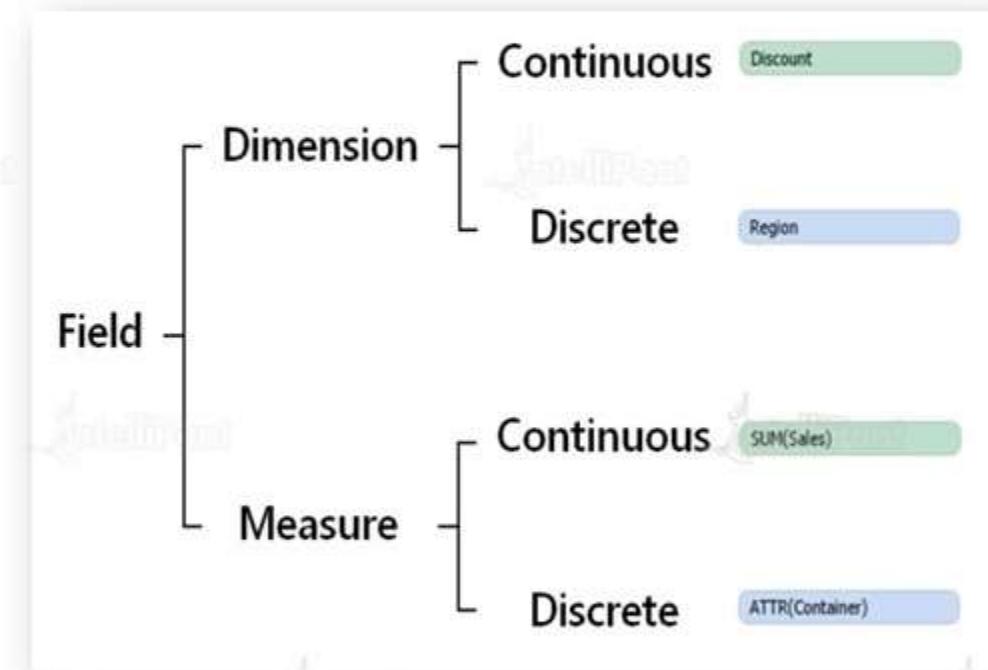


Tableau: Field Types



- By default, dimensions are discrete and measures are continuous. However, all four combinations are possible.
- If you are dragging a field from the dimensions area, the resulting field in the view will be discrete (with a blue background). If you are dragging a field from the measures area, the resulting field will be continuous (with a green background).

discrete dimensions	Product Name
continuous dimensions (possible only with Date dimensions)	QUARTER(Order Date)
discrete measures	SUM(Profit)
continuous measures	SUM(Profit)

A Note on File Types



Workbook (.twb):	Workbooks hold worksheets, dashboards and stories.
Bookmark (.tbn)	Bookmarks contain a single sheet.
Packaged workbook (.twbx)	It is an archive containing a workbook, along with all data sources and files.
Data extract (.hyper)	It is the local copy of a subset or the entire data source to share and improve the performance.
Data source(.tds)	It is the shortcut to frequently-used data sources, containing information and modifications.
Packaged data source (.tdsx)	It is an archive containing a data source file, along with any related files.

Difference between .twb & .twbx



.twb file

- The .twb file alone is not enough to perform any analysis because it only contains Tableau's instructions for interacting with a data source.
- The .twb files are XML files specially tailored to interact with data sources. They are custom-built to make awesome visualizations that Tableau generates. Here, you can see a picture of a .twb file opened in Notepad++.

.twbx file

- A .twbx file is a Tableau package.
- The .twbx files can be considered analogous to specialized zip files, in which these “zip” files contain all the information necessary to work in the Tableau.d Workbook.
- The primary advantage to using .twbx files is that the analysis can be performed without the network/internet connection because your data is already present on your computer in this packaged file.



Tableau Architecture

Installing Tableau Desktop



- To install Tableau Desktop:

Step 1

Use the below link to download the Tableau Desktop version

<http://www.tableau.com/products/desktop/download?os=windows>

Step 2

Run .exe file



Installing Tableau Desktop



Step 3

View the License Agreement, check the acceptance box and click on the “Install” button



Installing Tableau Desktop



Step 4

Click on “Start Trial Now”



Installing Tableau Desktop



Step 5

Fill the required details and click on “**Register**”

Activate Tableau

 **Registration**
Please complete all fields for the registered user.

First Name	Last Name	Organization
<input type="text"/>	<input type="text"/>	<input type="text"/>
Email		Phone
<input type="text"/>		<input type="text"/>
		Job Title
		<input type="text"/>
City	Postal Code	Department
<input type="text"/>	<input type="text"/>	<input type="text"/>
Country/Region	State/Province	Industry
<input type="text"/>	<input type="text"/>	<input type="text"/>

Register

Installing Tableau Desktop



Step 6

You can click on the Tableau icon and then on Open to start work with Tableau

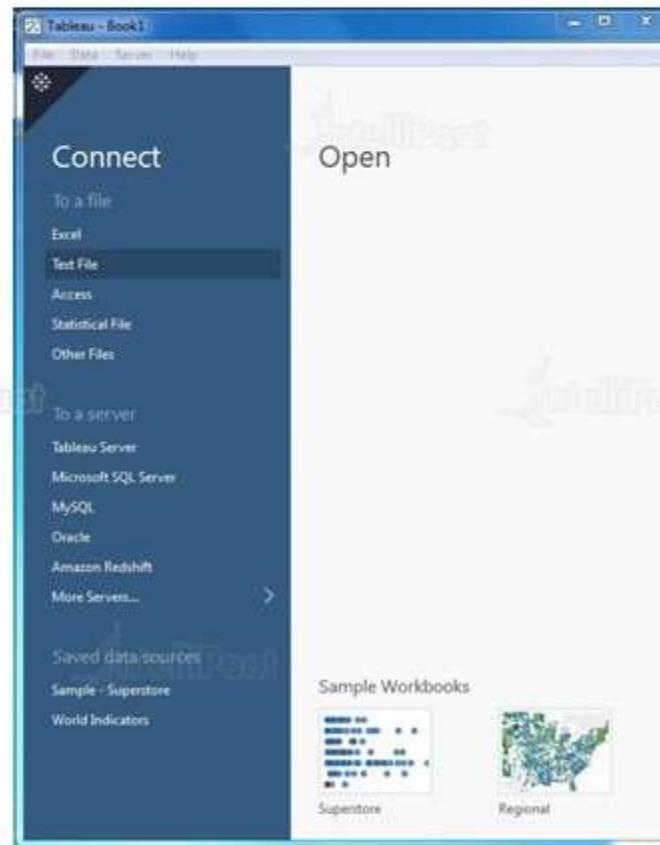


Tableau Architecture

- Tableau has a highly scalable, n-tier client-server architecture that serves mobile clients, web clients and desktop-installed software.
- Tableau Desktop is the authoring and publishing tool that is used to create shared views on the Tableau Server.



Tableau Desktop



Tableau Server



iPad



Mobile Safari



Android



PC Browser

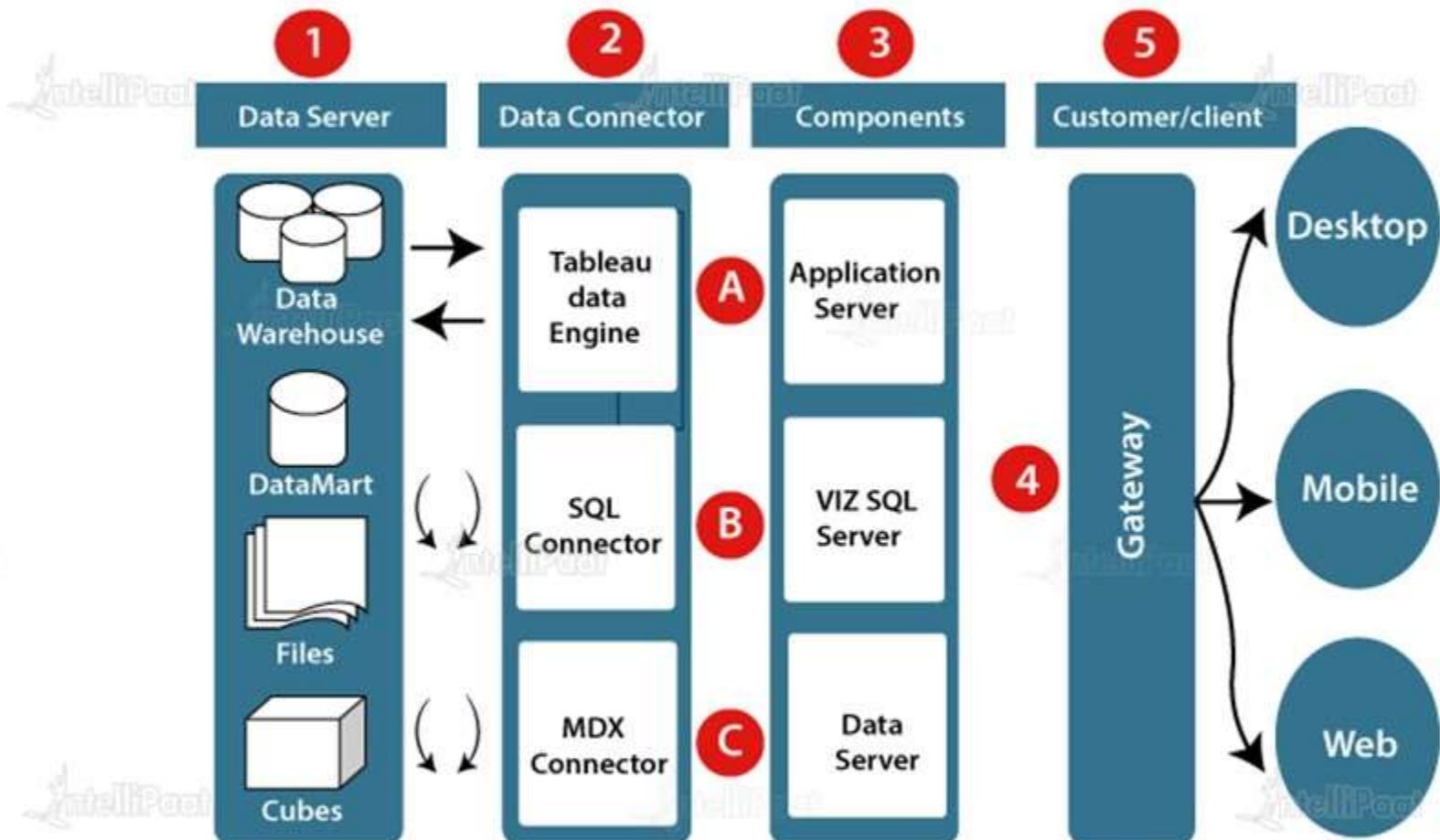
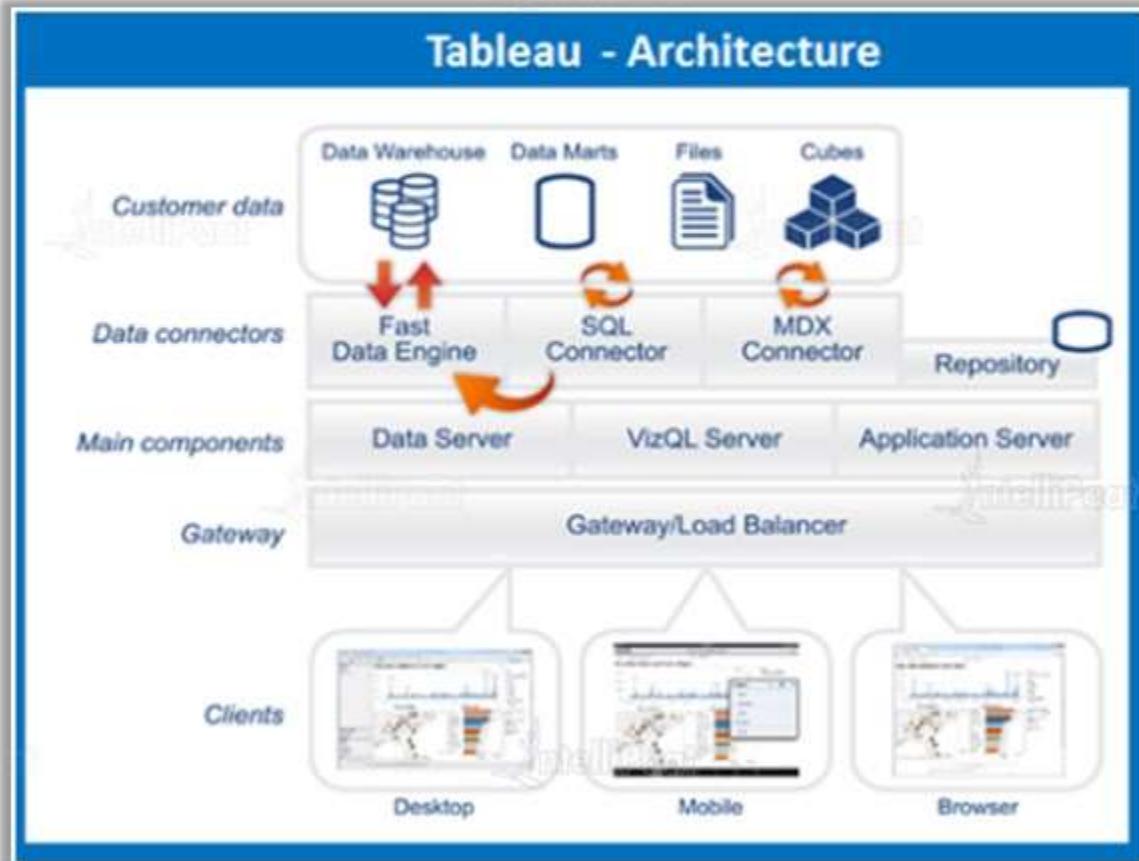


Tableau Architecture



Tableau Server:

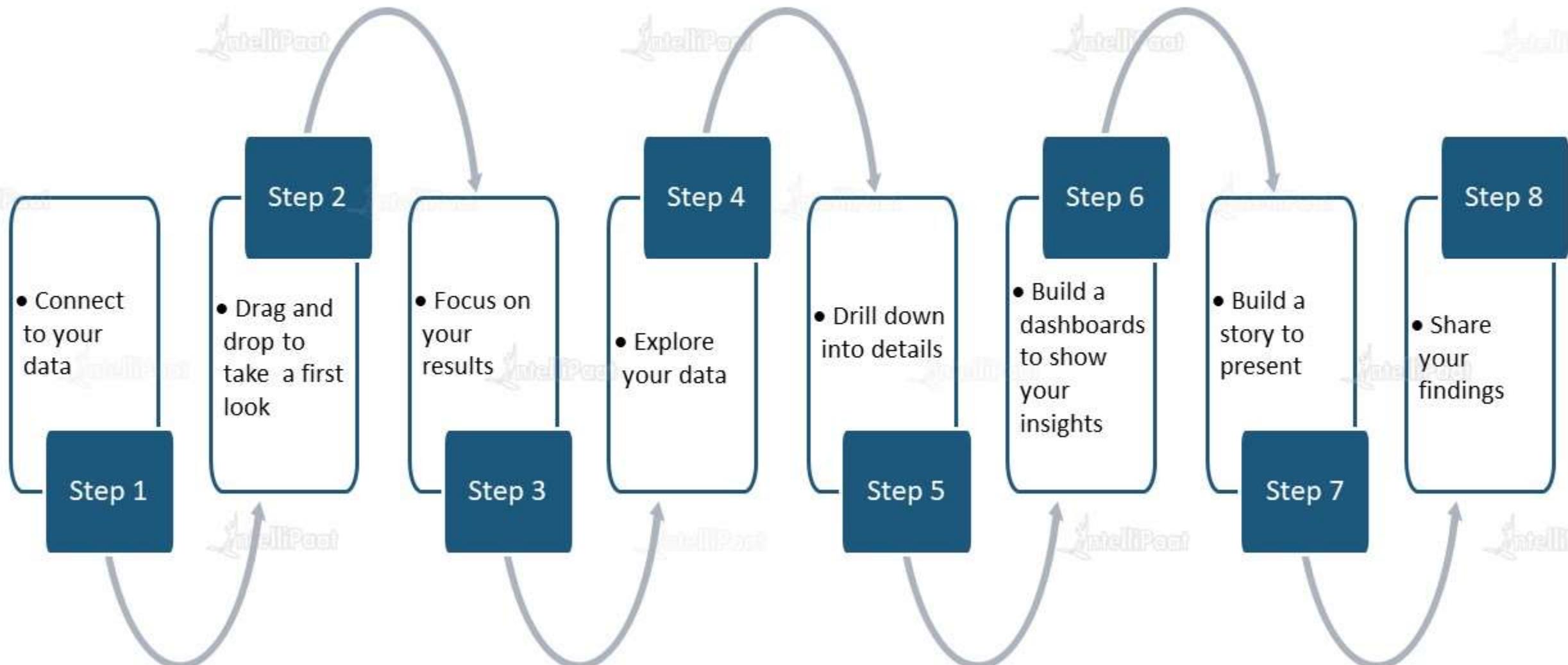
- Scales up is multi-threaded
- Scales out is multi-process enabled
- Provides integrated clustering
- Supports high availability
- Is secure
- Runs on both physical and virtual machines



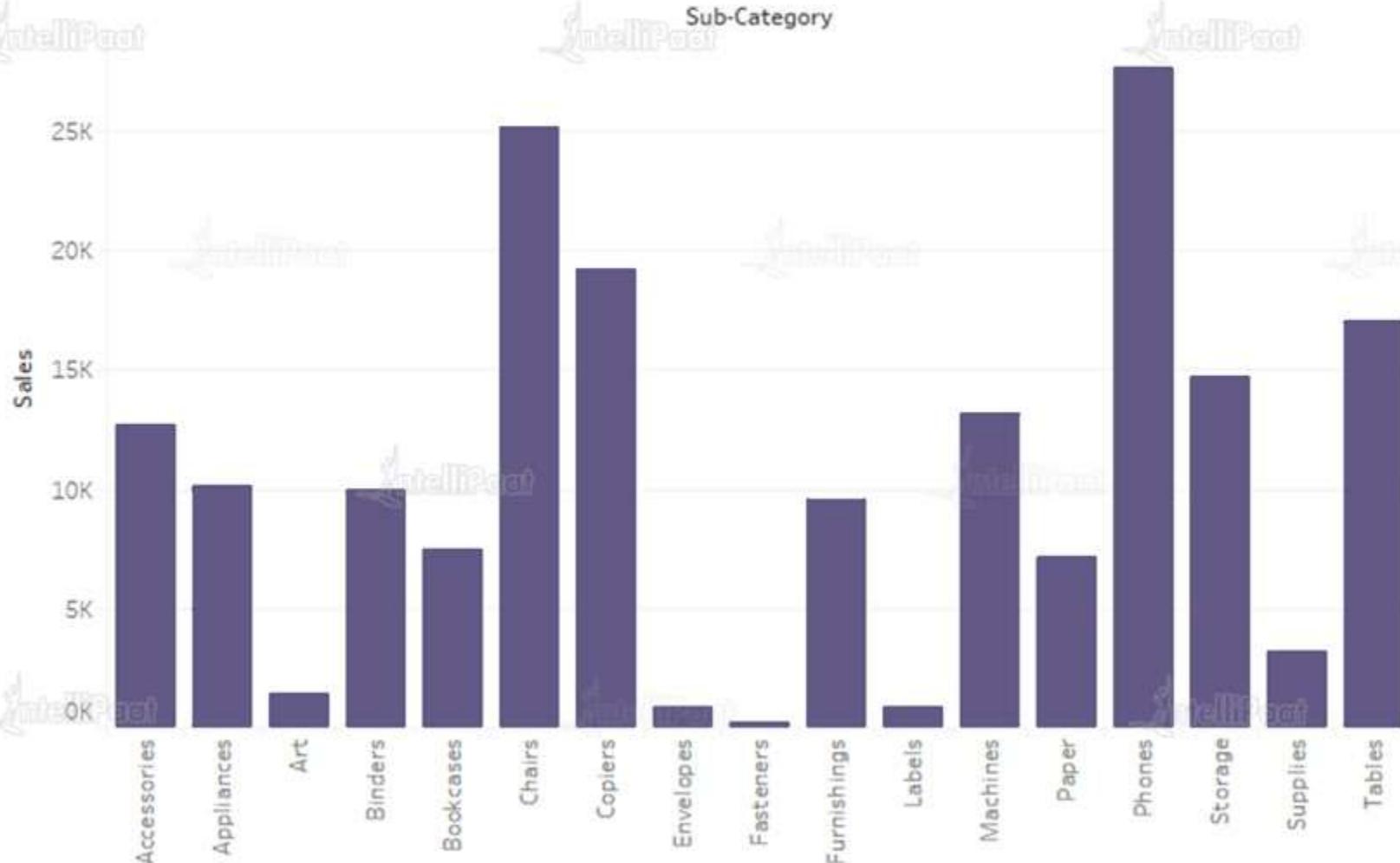


How to start with Tableau?

Starting with Tableau



My First Visualization

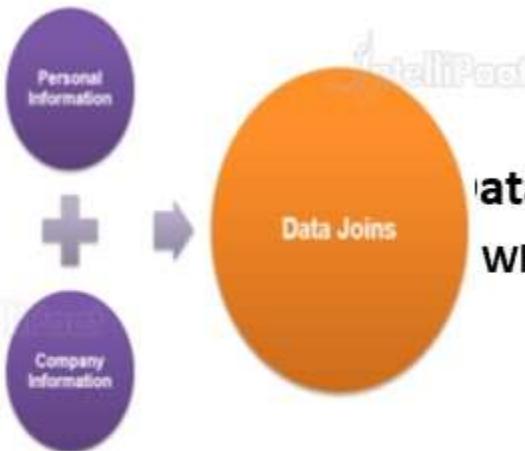


Working with Metadata & Data Blending

Connecting to Multiple Data Sources

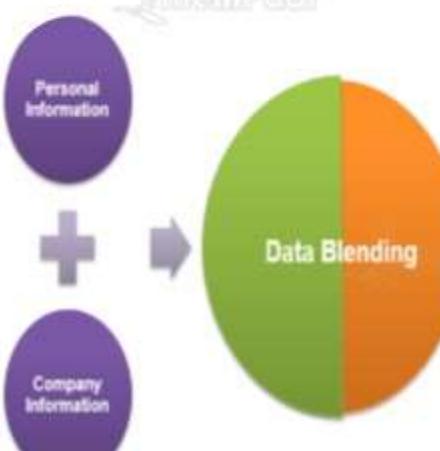


In the case of
multiple sources
use one of the
two mechanisms:



Data Joins

When the source data tables reside in a single source



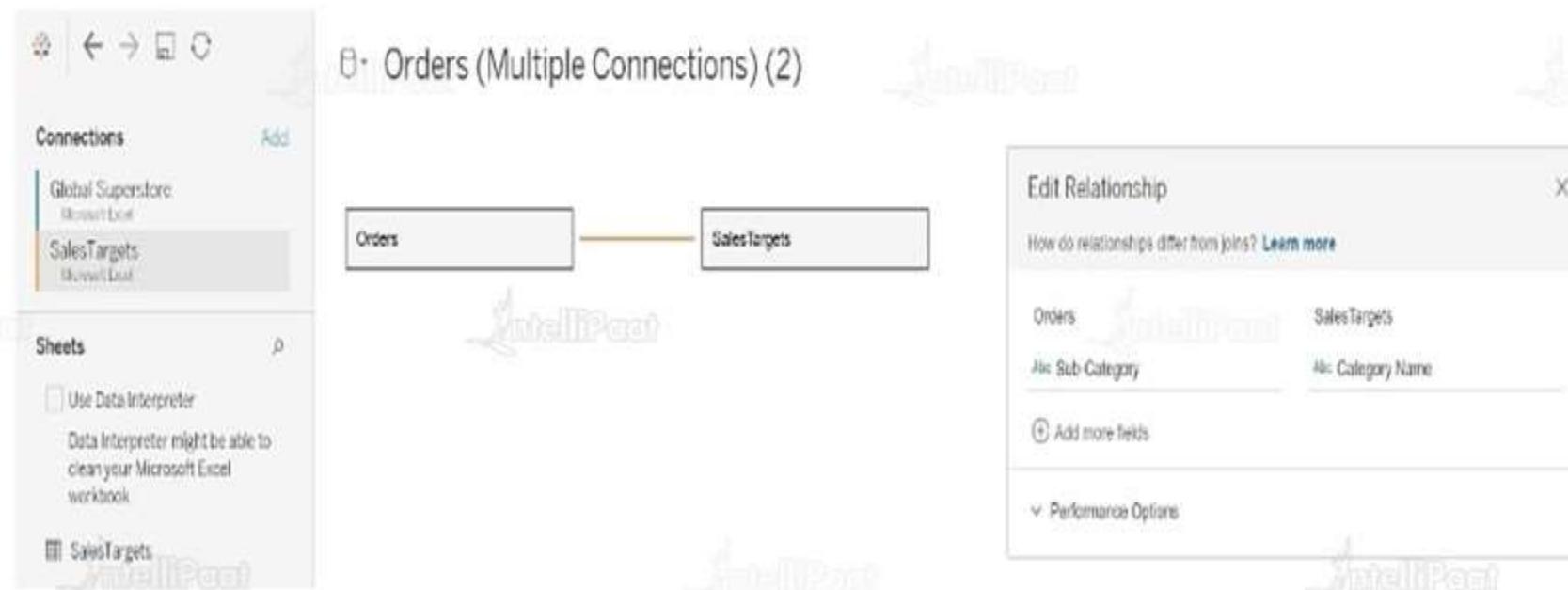
Data Blending

When the data source tables are from different
sources

One data source becomes the primary data source,
and relationships between fields need to be
maintained manually

Relationships

- Apart from Data joins and Data blending, we have this feature called Relationships which was introduced in version 2020.2 of Tableau for connecting multiple data sources.

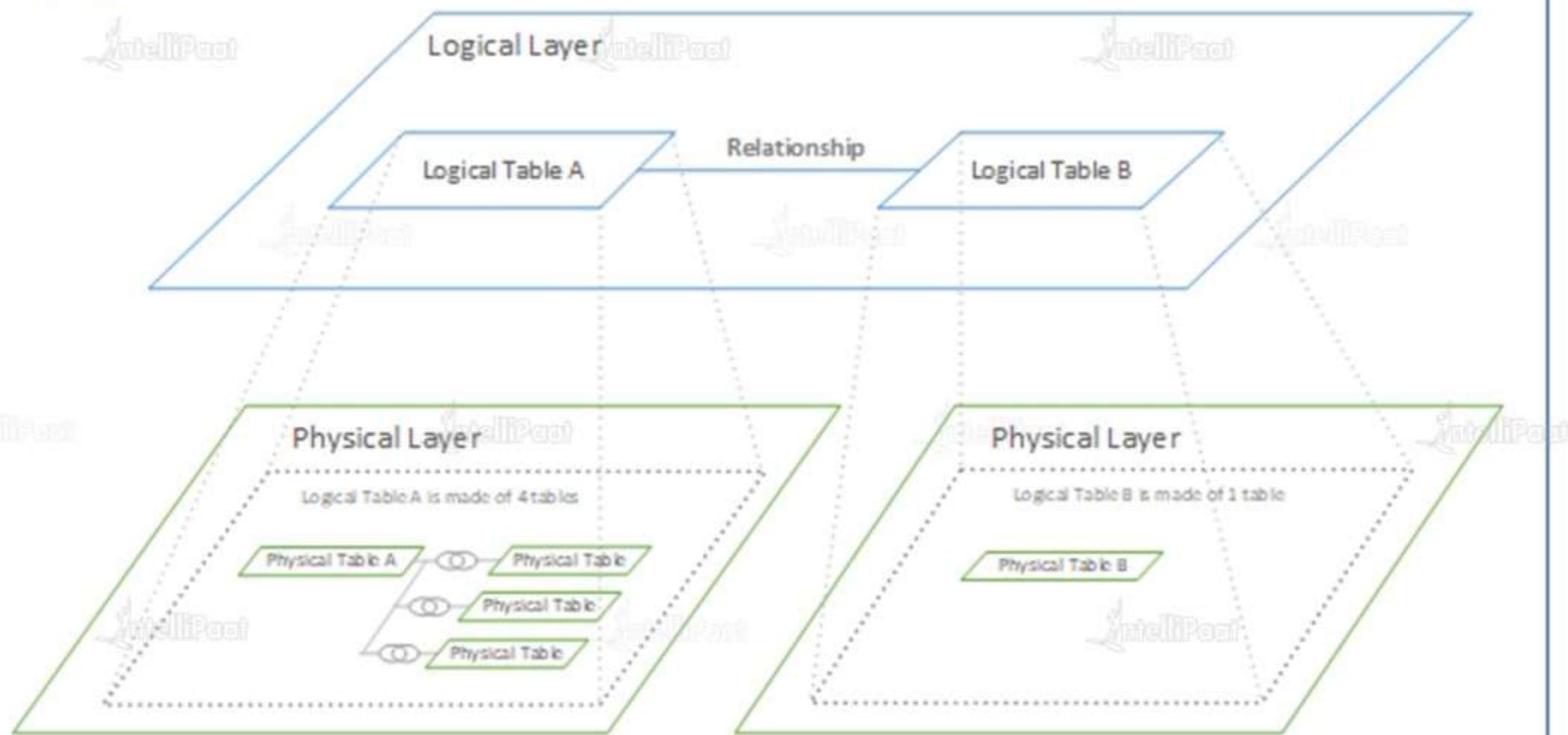


The screenshot shows the Tableau interface with the following components:

- Connections:** A sidebar on the left lists "Global Superstore" and "SalesTargets".
- Sheets:** A section below the sidebar shows a preview of the "SalesTargets" sheet.
- Relationships:** A central panel titled "Orders (Multiple Connections) (2)" displays a relationship diagram where the "Orders" box is connected to the "SalesTargets" box by a yellow line.
- Edit Relationship dialog:** A modal window titled "Edit Relationship" is open, showing the connection between "Orders" and "SalesTargets" based on "Sub-Category" and "Category Name". It also includes sections for "Add more fields" and "Performance Options".

DATA MODEL FROM VERSION 2020.2

Data Model



	RELATIONSHIP	CROSS DATABASE JOIN	DATA BLEND
FEATURE	INTRODUCED IN 2020.2 VERSION OF TABLEAU DESKTOP	ALREADY IN USE	ALREADY IN USE
CONNECTING 2 nd , 3 rd or N th data set	USING 'ADD' BUTTON IN THE DATA SOURCE PAGE	USING 'ADD' BUTTON IN THE DATA SOURCE PAGE	USING 'NEW DATA SOURCE' TOOL IN THE WORKSHEET
LAYER IN WHICH THIS FEATURE IS USED	LOGICAL LAYER	PHYSICAL LAYER	IN THE WORKSHEET
DATA SOURCE PAGE	ONE COMMON DATA SOURCE PAGE – FOR ALL THE DATASETS CONNECTED	ONE COMMON DATA SOURCE PAGE – FOR ALL THE DATASETS CONNECTED	EVERY DATA SET CONNECTED HAS A SEPARATE DATA SOURCE PAGE
DATA IS JOINED	DURING VISUALIZATION ONLY	IN THE DATA SOURCE PAGE ITSELF	DURING VISUALIZATION ONLY

TYPE OF JOIN	FLEXIBLE – DECIDED BY TABLEAU BASED ON THE NATURE AND SHAPE OF OUR DATA OR CAN BE MANUALLY DECIDED USING ‘PERFORMANCE OPTIONS’	4 TYPES 1) INNER JOIN 2) LEFT OUTER JOIN 3) RIGHT OUTER JOIN 4) FULL OUTER JOIN	ONLY LEFT OUTER JOIN
REPRESENTED USING	A FLEXIBLE NOODLE LINE BETWEEN TABLES	JOINTYPE SYMBOLS	BLUE AND ORANGE TICKS IN THE DATA PANE
ORDER OF EXECUTING JOINS AND AGGREGATION	FLEXIBLE – BASED ON THE DATA, DATA IS JOINED, THEN AGGREGATED OR DATA IS AGGREGATED, THEN JOINED	DATA IS JOINED AND THEN AGGREGATED	DATA IS AGGREGATED AND THEN JOINED
DIFFERENCE IN LEVELS OF HIERARCHY – DUPLICATION OF RECORDS	HANDLED AUTOMATICALLY – DUE TO WHICH DUPLICATION IS AVOIDED	DUPLICATION OCCURS	DUPLICATION IS AVOIDED

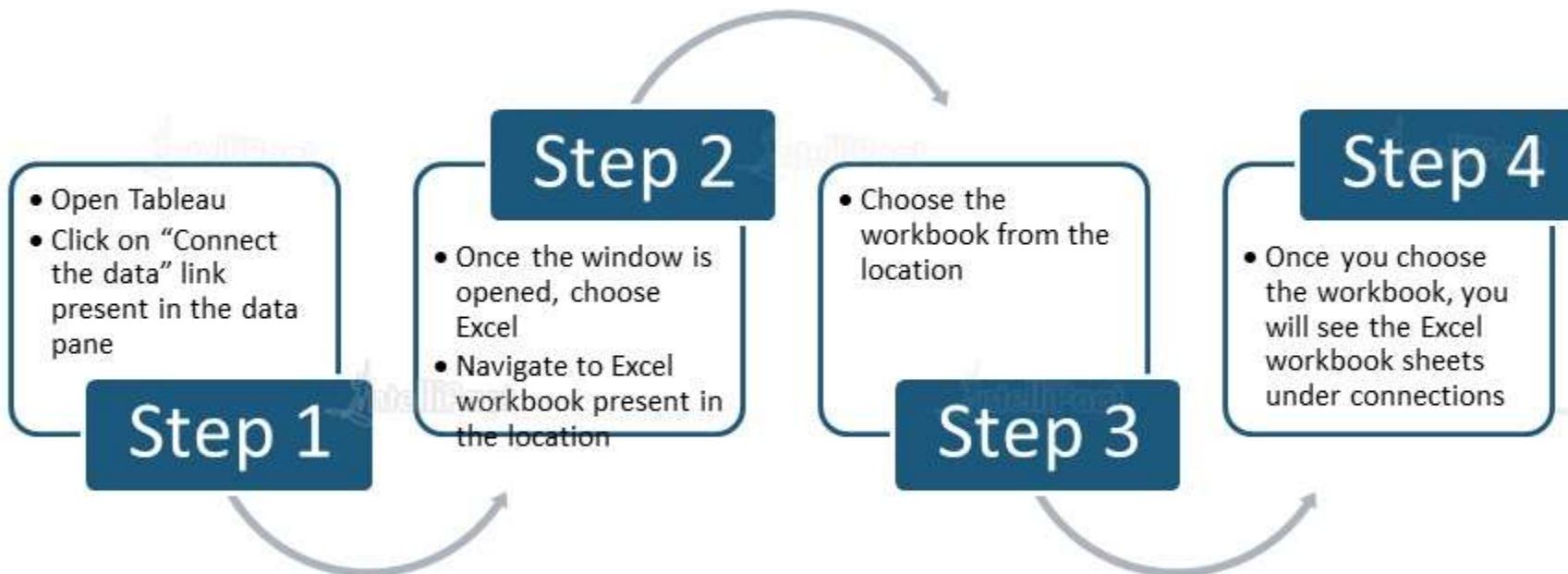
REQUIREMENTS	AT LEAST ONE PRIMARY KEY	AT LEAST ONE PRIMARY KEY	AT LEAST ONE PRIMARY KEY
CHOOSING PRIMARY KEY	TABLEAU DOES IT AUTOMATICALLY OR CAN BE DECIDED MANUALLY AS WELL – IN THE EDIT RELATIONSHIP DIALOG BOX (DATASOURCE PAGE)	TABLEAU DOES IT AUTOMATICALLY OR CAN BE DECIDED MANUALLY AS WELL – BY CLICKING ON THE JOIN TYPE SYMBOL (DATA SOURCE PAGE)	TABLEAU DOES IT AUTOMATICALLY OR CAN BE DECIDED MANUALLY AS WELL – BY USING 'EDIT BLEND RELATIONSHIPS' OPTION UNDER 'DATA' IN THE MENU BAR
JOIN CALCULATION	CAN BE USED	CAN BE USED	CANNOT BE USED
FLEXIBILITY OF DECIDING THE JOIN TYPE	FLEXIBLE	JOIN TYPE SHOULD BE DECIDED IN THE DATA SOURCE PAGE ITSELF	FLEXIBLE
DATA PANE	TABLES IN THE DATA PANE ARE ARRANGED INDIVIDUALLY FOR ALL THE LOGICAL	TABLES ARE ARRANGED TOGETHER IN THE DATA PANE FOR ALL PHYSICAL	EVERY DATA SET HAS ITS OWN DATA PANE IN WHICH CORRESPONDING DATA SET'S

WHEN TO USE	MOST RECOMMENDED BECAUSE OF ITS FLEXIBILITY	AS REQUIRED – DEPENDING ON YOUR DATA	AS REQUIRED – DEPENDING ON YOUR DATA – ESP WHEN WORKING WITH PUBLISHED DATA SOURCES AS RELATIONSHIPS CANNOT BE APPLIED ON THEM
-------------	---	--------------------------------------	--

Connection to Excels, PDFs & Cubes



- Tableau can be connected to Excel by following the below steps:



Connection to Excels, PDFs & Cubes



- You can find individual sheets under “Connections”

The screenshot shows the Tableau interface with the following details:

- Connections:** A sidebar on the left lists "Webspeed Updated" as the active connection.
- Sheets:** Under the "Connections" section, there is a "Data Interpreter" note and a list of sheets: "Sheet1", "Sheet2", "Sheet3", and "Website Speed Data".
- Join:** A modal dialog titled "Join" is open, showing two data sources: "Website" and "Website Speed Data". It includes options for "Inner", "Left", "Right", and "Full Outer" joins.
- Data View:** The main workspace displays a table with the following columns and data:

Website	Load Speed	Page Size	Website (Website ...)	URL	Load Speed (Webs...	Page Size (Websit...	Google Score	Website Speed Data
Amazon	1.0100	4.20000	Amazon	https://amazon.com	1.0100	4.20000	57	
ABC News	6.9400	7.40000	ABC News	http://abcnews.go.co...	6.9400	7.40000	35	
Ebay	1.7700	2.20000	Ebay	http://www.ebay.co...	1.7700	2.20000	87	
Facebook	1.2400	0.43200	Facebook	http://facebook.com	1.2400	0.43200	49	
NBC News	9.9300	1.90000	NBC News	null	9.9300	1.90000	73	
New York Times	1.2400	2.80000	New York Times	null	1.2400	2.80000	80	
Wikipedia	1.1200	0.33600	Wikipedia	https://en.wikipedia...	1.1200	0.33600	83	
Twitter	2.3700	4.10000	Twitter	https://twitter.com/	2.3700	4.10000	23	
YouTube	1.7800	2.10000	YouTube	http://youtube.com/	1.7800	2.10000	74	
Netflix	2.6300	1.00000	Netflix	https://www.netflix.c...	2.6300	1.00000	83	

Once you are connected to Excel, you will need to drag table data sheets from Excel into the window.

Connection to Excels, PDFs & Cubes

- PDF connection:



Connection to Excels, PDFs & Cubes



Cubes connection:

- In Cube data source, the hierarchies and aggregations are created by Cube's designer in advance.
- Cubes can return information faster than a relational data source. This is because, aggregations and hierarchies are pre-built in Cubes.
- The Cube data sources supported in Tableau are:

Oracle Essbase

Teradata OLAP

Microsoft Analysis Services (MSAS)

SAP NetWeaver
Business
Warehouse

Microsoft
PowerPivot

Analytical Views in
SAP HANA

Connection to Excels, PDFs & Cubes



- Cubes connection:

Step 1: Open Tableau, go to “Connect” and click on “Other Files”

Step 2: Choose the cube file that you want to connect and click on “Open”

Step 3: Check a local cube file and click on “Connect”

Managing Metadata & Extracts



- Tableau facilitates in capturing the information details of sources like columns and their information sorts.
- Information is employed to form dimensions and measures. We can also calculate the fields square measure. A number of properties of the information will be modified.
- The metadata in Tableau can be edited and renamed, and also columns can be hidden.

Edit Metadata

- The datatype of a number of fields will be modified if needed. Looking on the character of supply information, generally, Tableau might fail to acknowledge the information sort from the supply. In such a situation, the information sort will be amended manually. The screen below shows the choice.



Live Vs. Extract Connection



Live Connection:

- Connects to your data live. Live is nothing but real-time extract.
- Offers the convenience of real-time updates, with any changes in the data source reflected in Tableau. However, live connections also rely on the database for all queries. Unlike extracts, databases are not always optimized for fast performance. With live connections, your data queries are only as fast as the database itself

Extract (In-memory):

- Presents a fast, in-memory data engine to optimize for analytics. You can connect to your data, and after that, with one click, extract your data to get it in-memory in Tableau.
- Snapshots of data optimized for aggregation and loaded into the system memory to be quickly recalled for visualization. Extracts tend to be much faster than live connections, especially in more complex visualizations with large data sets, filters, calculations, etc.

Data Prep with Text & Excel Files



Data Interpreter

With no columns names and a lot of nulls, **Data Interpreter** strips out them and the columns are properly identified.

Pivot

The **Pivot** merges the information from the original columns and rows into two new columns.

Metadata Grid

It can be a useful view with a large number of files and when tables have been joined.

Split

You can **split** the columns based on the shared delimiter of that hyphen.

Custom Split

You can choose the delimiter and use a hyphen.

Dealing with NULL Values



Null values are **blanks** that often appear in certain database columns.

They can be either intentionally blank (representing zero) or incomplete data.

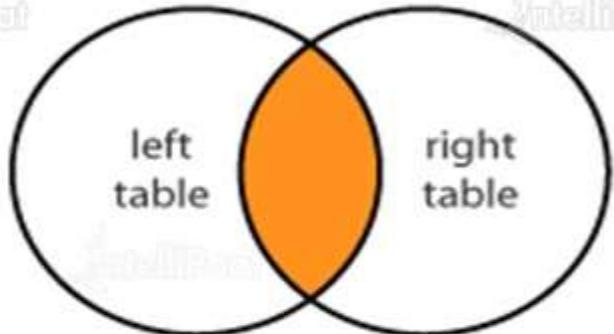
If they are intentionally blank, it's best to convert nulls to zero using the **ZN()** function.

This helps avoid some potential mistakes in your calculations.

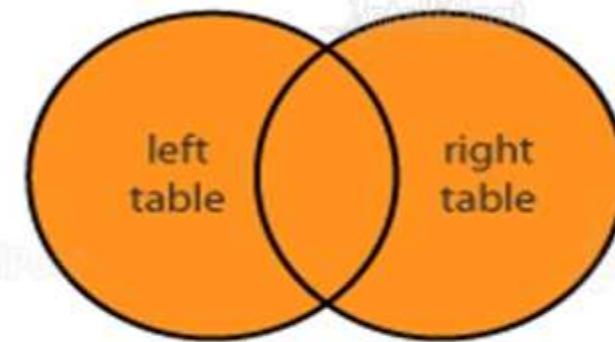
The **ZN()** function returns the number if a number is present but zero if a null is present

Connecting to Multiple Data Sources: Understanding Joins

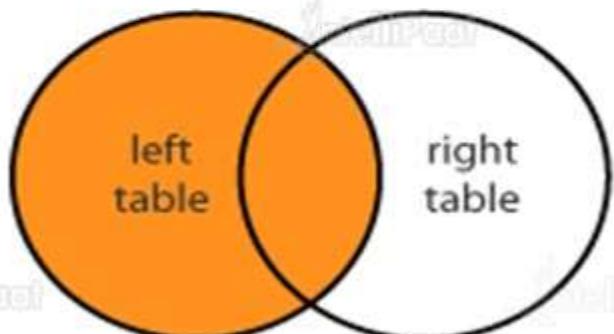
INNER JOIN



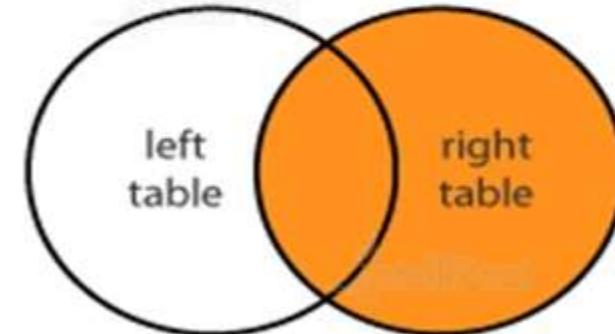
FULL JOIN



LEFT JOIN



RIGHT JOIN



- ❖ How to do a Full Outer Join in Tableau while connecting to an Excel file?

Cross-database Joins



Adding Another Connection

Add new connections only if the data comes from different databases. For example, add new connections to each database if you need to analyze tables from a text file and tables

from Google Sheets: Connecting to Data

On the start page, under "Connect", click a connector to connect to your data. This step automatically creates the first connection in the Tableau data source.

To create a cross-database, you must create a multi-connection Tableau data source. You can do so by adding and then connecting to each of the different databases (including Excel and text files) before you join.

When related data is stored in tables across different databases, you can use a cross-database join to combine the tables.

Data Blending



1

It helps user combine two different datasets that might be related to each other by some common attribute.

2

Tableau automatically tries to link 2 datasets based on the column names in both datasets. Automatic relationship can be edited by going to the data menu and clicking on "Relationship".

3

Data blending should be used when you have related data in multiple data sources that you want to analyze together in a single view.

4

Tableau only supports left join in data blending. You must change the order of data sources for right outer join.

5

For Inner joins, you need to filter nulls.

6

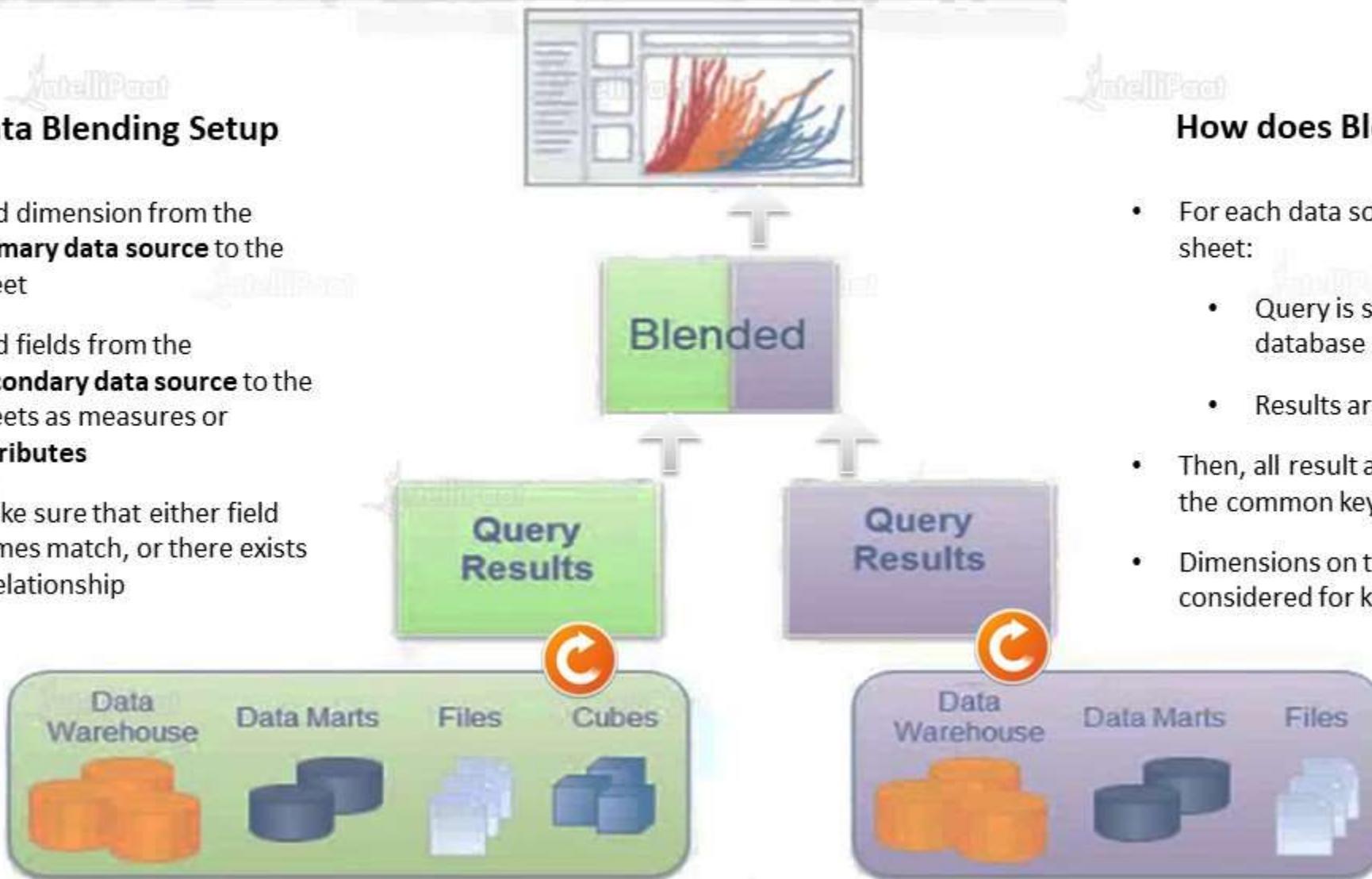
Cubes are not supported in secondary datasets.

Data Blending



Data Blending Setup

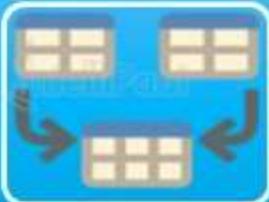
- Add dimension from the **Primary data source** to the sheet
- Add fields from the **secondary data source** to the sheets as measures or **attributes**
- Make sure that either field names match, or there exists a relationship



How does Blend work?

- For each data source used on the sheet:
 - Query is sent to the database
 - Results are processed
- Then, all results are left joined on the common key
- Dimensions on the sheet are considered for keys

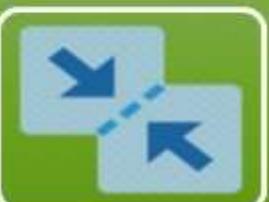
When to Join & When to Blend



If all the data resides in one data source, joining the data tables can improve performance and filtering control.



If new data or table is added in the same data connection, using a join is better.



If the data source contains too many records for a join to be practical, using data blending is preferable.



When both summary and details need to be showcased at the same time, data blending is preferred.

Data Extraction

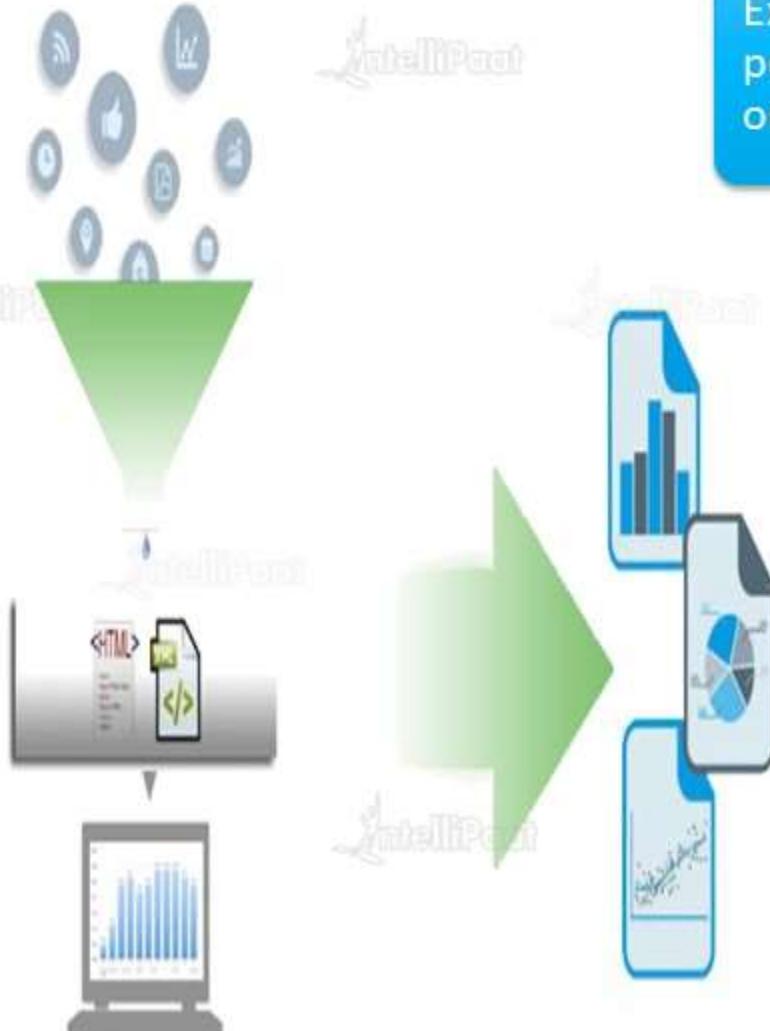


Extracts are the saved subsets of data that you can use to improve performance or to take advantage of Tableau functionality not available or supported in your original data.

When you extract your data to create an extract, you can reduce the total amount of data by using filters and defining other limits.

After you create an extract, you can refresh it with the data from the original data.

When refreshing the data, you have the option to either do a full refresh, which replaces all of the extracted contents, or an incremental refresh, which only adds rows that are new since the previous refresh.



Data Extraction



Extracts can do the following:

They improve performance. For file-based data sources such as Excel or Access, a full extract takes advantage of the Tableau data engine. For large data sources, a filtered extract can limit the load on the server when you only need a subset of data.

They take advantage of Tableau functionality that is not available in the original data source, such as the ability to compute Count Distinct.

They provide offline access to your data. If your data source is not available (for example, because you are traveling), you can extract the data to a local data source.

Data Extraction



Original data: Each record is shown as a separate row. There are 7 rows in your data.

Aggregating data for visible dimensions (no roll up)

Records with the same date and region have been aggregated into a single row. There are 5 rows in the extracted data source.

Aggregating data for visible dimensions (roll up dates to Month): Dates have been rolled up to the month level, and records with the same region have been aggregated into a single row. There are 3 rows in the extracted data

Source:

Refresh Extracts



Full Refresh

By default, extracts are fully refreshed, which means that every time you refresh the extract all the rows are replaced with the data in the underlying data source.

While this kind of refresh ensures you with having an exact copy of what is in the underlying data source, it can sometimes take a longer time and be expensive on the database depending on how big the extract is.

If the extract is not set up for incremental extract, selecting to refresh the extract will fully refresh the extract. If you're publishing the data source to Tableau Server, you can specify the type of refresh in the "Scheduling & Passwords" dialog box. Most data sources support an incremental refresh.

Incremental Refresh



Rather than refreshing the entire extract, you can set it up to only add the rows that are new since the last time you extracted data.

- **For example:** you may have a data source that is updated daily with new sales transactions. Rather than rebuilding the entire extract each day, you can just add the new transactions that occurred that day. Then, once a week, you can do a full refresh just to be sure that you have the most up-to-date data.

Data Extracts



Data extraction in Tableau creates a subset of data from the data source.

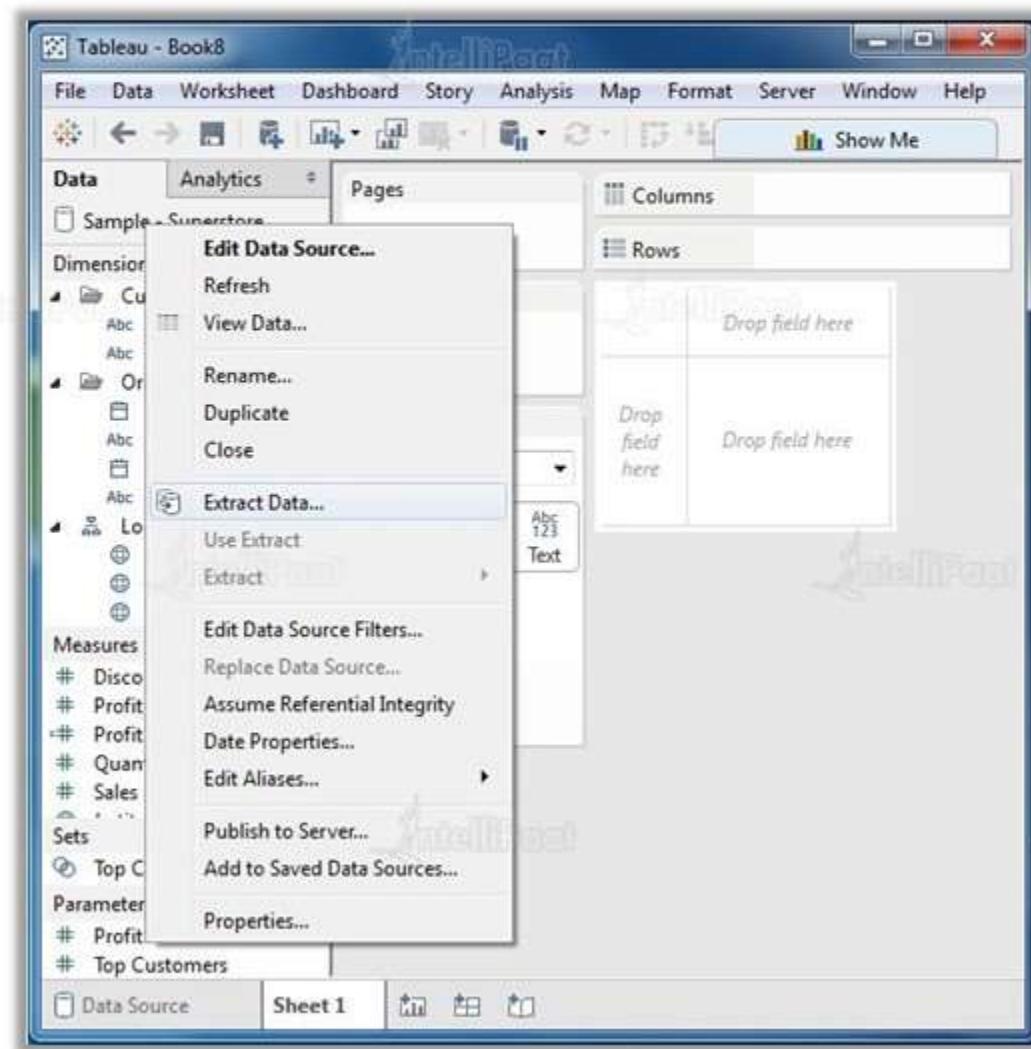
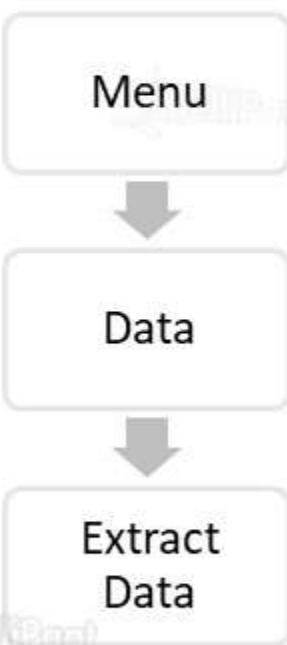
This can be used to increase the performance by applying filter.

This also helps in applying some features of Tableau, that are not available in data sources.

This is mainly used to create an extract that is stored in the local drive for offline access by Tableau.

Data Extraction: Steps

- Extraction of data is done as follows:



Quiz



Q 1. In which year Tableau was introduced?

- a) 2000
- b) 2001
- c) 2010
- d) 2003

A circular orange icon containing the word "QUIZ" in large, bold, orange letters, with a small orange square below the letter "I".

Quiz

Q 2. How many modes of data connecting do Tableau have?

- a) 1
- b) 2
- c) 3
- d) 4



QUIZ

Quiz



Q 3. Which is a Primary Tableau Server?

- a) Data Server
- b) VizQL server
- c) Gateway
- d) Application Server

A circular orange icon with a white center. In the center, the word "QUIZ" is written in orange capital letters, with a small orange square symbol positioned below the letter "I".

Quiz



Q 4. What are MDX connectors essentially used for?

- a) OLAP database
- b) Quickbooks database
- c) JSON database

A circular orange icon with a white center, containing the word "QUIZ" in orange capital letters with a small orange square below the letter "I".

QUIZ

Quiz



Q 5. What does a fast data engine do?

- a) Transfers data
- b) Assigns data interpreter
- c) Imports data quickly to perform analysis
- d) All the above

A circular orange icon containing the word "QUIZ" in large, bold, orange capital letters, with a small orange square at the bottom right corner.

Quiz



Q 1. What is data blending?

- a) Extraction of limited number of data
- b) Establishing connection with database
- c) Identifying the common data field

A circular orange icon with the word "QUIZ" in the center, with a small orange square below it.

Quiz



Q 2. What is concatenation in Tableau?

- a) Combining group of attributes into a single attribute
- b) Splitting of group of observations from a single observation
- c) Duplicating set of fields
- d) Combining group of fields into a single field

A circular orange icon containing the word "QUIZ" in large, bold, orange letters, with a small orange square below it.



Thank You