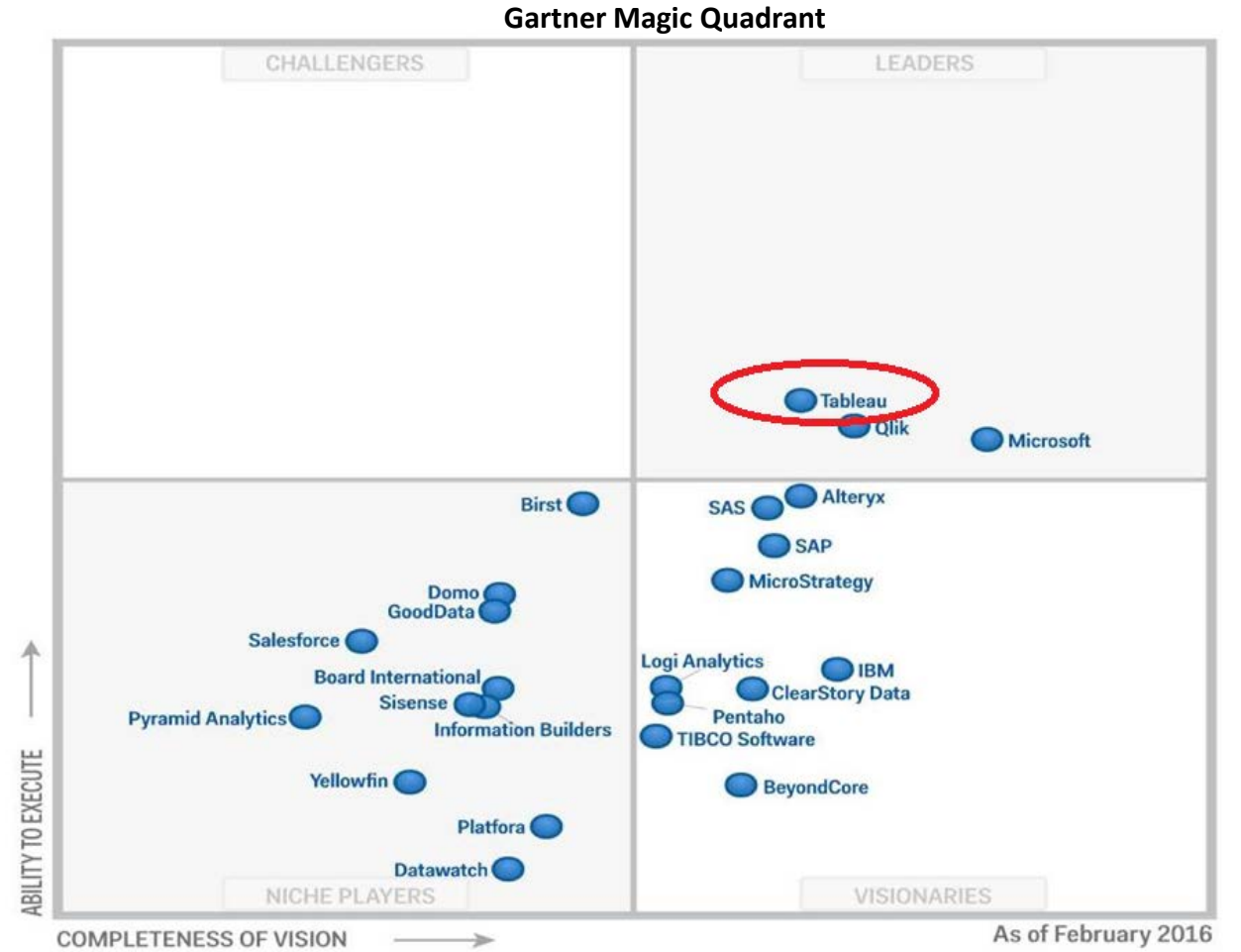


*Stay Ahead*

# **Class 1 – Tableau Architecture**



- Tableau is a Business Intelligence tool for visually analyzing the data.
- Users can create and distribute interactive and shareable dashboards which depict the trends, variations and density of the data in form of graphs and charts.
- Tableau can connect to files, relational and Big data sources to acquire and process data.
- The software allows data blending and real time collaboration, which makes it very unique.
- It is used by businesses, academic researchers and many governments to do visual data analysis.
- It is also positioned as a leader Business Intelligence and Analytics Platform in **Gartner Magic Quadrant**.



Source: Gartner (February 2016)

## Tableau Features

Tableau provides solutions for all kinds of industries, departments and data environments. Below are the unique features which enable tableau handle so many diverse scenarios.

- **Speed of Analysis** - As it does not need high level of programming expertise, any computer user with access to data can start using it to derive value from the data.
- **Self-Reliant** - Tableau does not need a complex software setup. The desktop version which is used by most users is easily installed and contains all the features needed to start and complete data analysis.
- **Visual Discovery** - The user explores and analyses the data by using visual tools like colors, trend lines, charts and graphs. There is very little script to be written as nearly everything is done by drag and drop.
- **Blend Diverse Data Sets** - Tableau allows you to blend different relational, semi-structured and raw data sources in real time, without expensive up-front integration costs. The users don't need to know the details of how data is stored.
- **Architecture Agnostic** - Tableau works in all kinds of devices where data flows. So the user need not worry about specific hardware or software requirements to use Tableau.
- **Real Time Collaboration** - Tableau can filter, sort, and discuss data on the fly and embed a live dashboard in portals like SharePoint site or Salesforce. You can save your view of data and allow colleagues to subscribe to your interactive dashboards so they see the very latest data just by refreshing their web browser.
- **Centralized Data** - The tableau server provides a centralized location to manage all of the organization's published data sources. You can delete, change permissions, add tags, and manage schedules in one convenient location. It's easy to schedule extract refreshes and manage them in the data server. Administrators can centrally define a schedule for extracts on the server for both incremental and full refreshes.

## Tableau Architecture

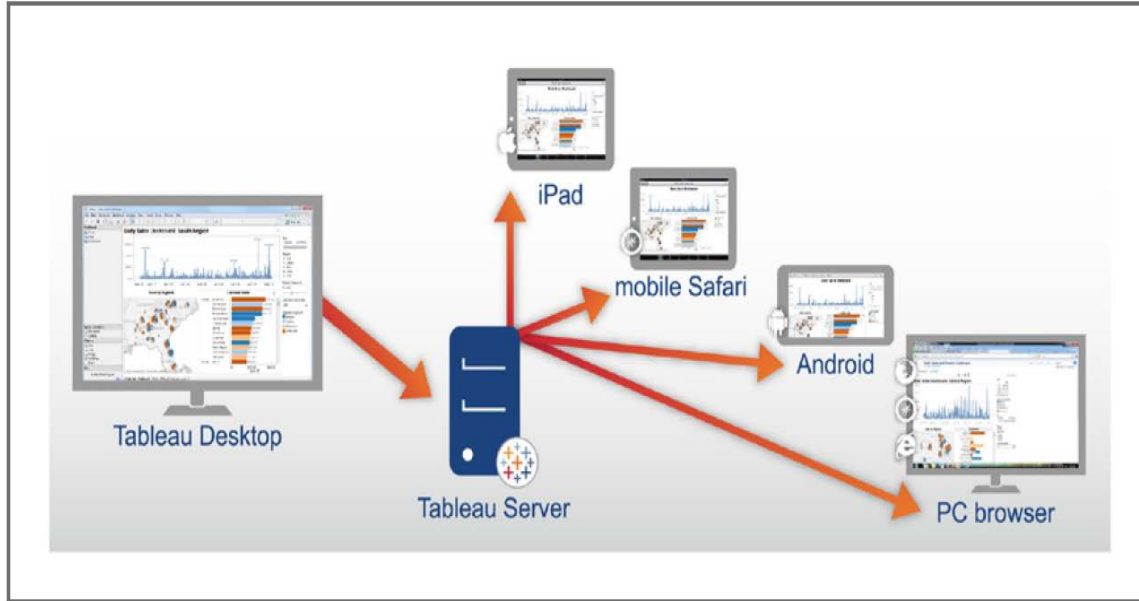
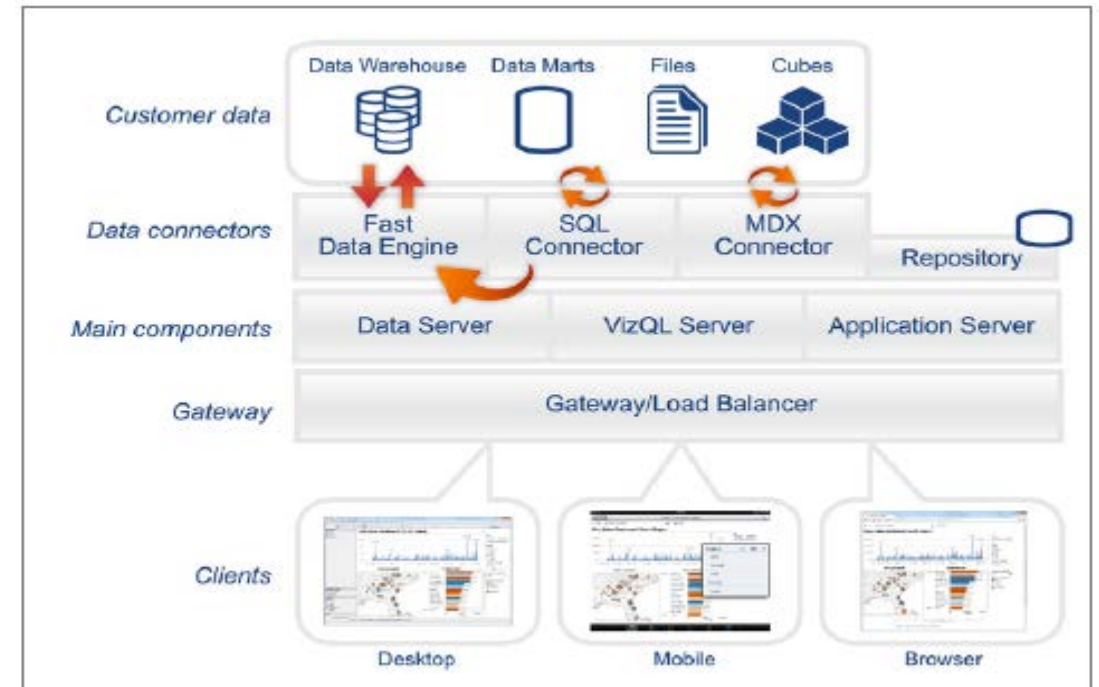


Tableau has a highly scalable, n-tier client-server architecture that serves mobile clients, web clients and desktop-installed software. Tableau Desktop is the authoring and publishing tool that is used to create shared views on Tableau Server.

Tableau Server is an enterprise-class business analytics platform that can scale up to hundreds of thousands of users. It offers powerful mobile and browser-based analytics and works with a company's existing data strategy and security protocols. Tableau Server:

- Scales up: Is multi-threaded (Concurrent programming)
- Scales out: Is multi-process enabled (Distributed programming)
- Provides integrated clustering
- Supports High Availability
- Is secure
- Runs on both physical and Virtual Machines



## Tableau File Types

File Type	File Extension	Purpose
Tableau Workbook	.twb	It contains information on each sheet and dashboard that is present in a workbook. It has the details of the fields which are used in each view and the formula applied to the aggregation of the measures. It also has the formatting and styles applied. It also contains the data source connection information and any metadata information created for that connection.
Tableau Packaged Workbook	.twbx	This file format contains the details of workbook as well as the local data that is used in the analysis. Its purpose is to be share with other Tableau desktop or Tableau reader users assuming it does not need data from the server.
Tableau Data source	.tds	The details of the connection used to create the tableau report are stored in this file. In the connection details it stores the source type(excel/relational/sap etc.) as well as the data types of the columns.
Tableau Packaged Data source	.tdsx	This file is similar to the .tds file with the addition of data along with the connection details.
Tableau Data Extract	.tde	This file contains the data used in a .twb file in a highly compressed columnar data format. This helps in storage optimization. It also saves the aggregated calculations that are applied in the analysis. This file should be refreshed to get the updated data form the source.
Tableau Bookmark	.tbn	These files contain a single worksheet that is shared easily to be pasted into other workbooks.
Tableau Preferences	.tps	This file stores the color preference used across all the workbooks. It is mainly used for consistent look and feel across the users.

## Tableau Data Types

Data Type	Description	Example
STRING	Any sequence of zero or more characters. They are enclosed within single quotes. The quote itself can be included in a string by writing it twice.	'Hello' 'Quoted' 'quote'
NUMBER	These are either integers or floating points. It is advised to round the floating point numbers while using them in calculations.	3 142.58
BOOLEAN	They are logical values.	TRUE FALSE
DATE & DATETIME	Tableau recognizes dates in almost all formats. But in case we need to force tableau to recognize a string as date then we put a # sign before the data.	"02/01/2015" "#3 March 1982"

## Tableau Data Sources

Tableau can connect to all the popular data sources which are widely used. Tableau's native connectors can connect to the following types of data sources.

- **File Systems** like CSV, Excel etc.
- **Relational systems** like Oracle, Sql Server, DB2 etc.
- **Cloud systems** like Windows Azure, Google BigQuery etc.
- **Other Sources** using ODBC

- **Connect Live**

The connect live feature is used for real time data analysis. In this case Tableau connects to the real time data source and keeps reading the data. So the result of the analysis is up to the second and latest changes are reflected in the result. But on the downside it burdens the source system as it has to keep sending the data to Tableau.

- **In-Memory**

Tableau can also process data in-memory by caching them in memory and not being connected to the source anymore while analyzing the data. Of course there will be a limit to the amount of data cached depending on the availability of memory.

- **Combine Data Sources**

Tableau can connect to different data sources at the same time. For example in a single workbook you can connect to a flat file and a relational source by defining multiple connections. This is used in data blending which is a very unique feature in Tableau.



## An Idea

### Structured Data

Structured data is information, usually text files, displayed in titled columns and rows which can easily be ordered and processed by data mining tools. This could be visualized as a perfectly organized filing cabinet where everything is identified, labeled and easy to access. Most organizations are likely to be familiar with this form of data and already using it effectively.

Databases like Oracle, Teradata, SQL Server, My Sql are some of the databases which are market leaders for structured databases.

### Unstructured Data

***80% of business-relevant information originates in unstructured form, primarily text***

Unstructured data is raw and unorganized and organizations store it all. Ideally, all of this information would be converted into structured data however, this would be costly and time consuming. Also, not all types of unstructured data can easily be converted into a structured model. For example, an email holds information such as the time sent, subject, and sender (all uniform fields), but the content of the message is not so easily broken down and categorized. This can introduce some compatibility issues with the structure of a relational database system.

In case you're still not quite sure what we mean, here is a limited list of types of unstructured data:

- Emails
- Word Processing Files
- PDF files
- Spreadsheets
- Digital Images
- Video
- Audio
- Social Media Posts

### Cloud environment

#### Easy to use

AWS is designed to allow application providers, ISVs, and vendors to quickly and securely host your applications – whether an existing application or a new SaaS-based application. You can use the AWS Management Console or well-documented web services APIs to access AWS's application hosting platform.

#### Flexible

AWS enables you to select the operating system, programming language, web application platform, database, and other services you need. With AWS, you receive a virtual environment that lets you load the software and services your application requires. This eases the migration process for existing applications while preserving options for building new solutions.

#### Cost-Effective

You pay only for the compute power, storage, and other resources you use, with no long-term contracts or up-front commitments. For more information on comparing the costs of other hosting alternatives with AWS, see the AWS Economics Center.

#### Reliable

With AWS, you take advantage of a scalable, reliable, and secure global computing infrastructure, the virtual backbone of Amazon.com's multi-billion dollar online business that has been honed for over a decade.

#### Scalable and high-performance

Using AWS tools, Auto Scaling, and Elastic Load Balancing, your application can scale up or down based on demand. Backed by Amazon's massive infrastructure, you have access to compute and storage resources when you need them.

#### Secure

AWS utilizes an end-to-end approach to secure and harden our infrastructure, including physical, operational, and software measures. For more information, see the AWS Security Center.