# Architecture Design

## Health Analytics on Heart Disease Data

Last Revised Date: 27/08/2021

### Document Version Control

| Version | Date | Author | Comments |
|---|---|---|---|
| **0.1** | 31-08-2021 | Samruddhi Deshmukh<br>Subhabrata Nath<br>Manish Tiwari | |
| **0.2** | | | |
| | | | |

### Reviews:

| Version | Date | Reviewer | Comments |
|---|---|---|---|
| | | | |

### Approval Status:

| Version | Review Date | Reviewed By | Approved By | Comments |
|---|---|---|---|---|
| | | | | |

# Index

# 1. Introduction

## 1.1 What is a BI Architecture design?

- Business intelligence architecture is the framework for the various technologies an organization deploys to run business intelligence and analytics applications.

- It includes the IT systems and software tools that are used to collect, integrate, store and analyze BI data and then present information on business operations and trends.

- BI Architecture aims at using data analysis and reporting to help an organization track-business performance, optimize business processes, identify new revenue opportunities, improve strategic planning and make more informed decisions overall.

- The process is broadly divided into three areas: **data collection, information management, and business intelligence.**

## 1.2 Importance of a BI architecture-

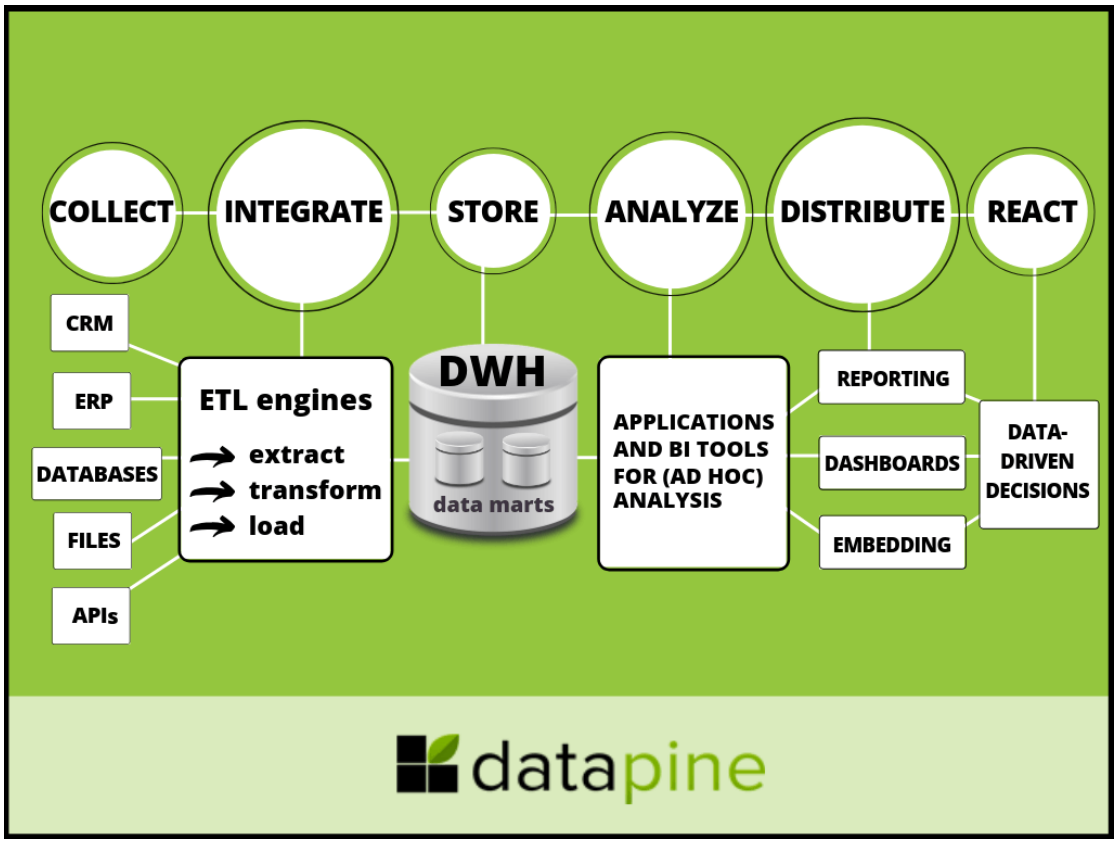
- **A BI architecture serves as a technology blueprint for collecting, organizing and managing BI data and then making the data available for analysis, data visualization and reporting**. A strong BI architecture also incorporates policies to govern the use of the technology components.
- Putting such a framework in place enables a BI team to work in a coordinated and disciplined way to build an enterprise BI program that meets its organization's data analytics needs.
- The BI architecture also helps BI and data managers create an efficient process for handling and managing the data that's pulled into the environment.
- **Enterprises benefit from an effective BI architecture by using the insights generated by business intelligence tools to make data-driven decisions that help increase revenue and profits.**

- BI is mainly focused on generating business insights, whether operational or strategic efficiency such as product positioning and pricing to goals, profitability, sales performance, forecasting, strategic directions, and priorities on a broader level. The point is to access, explore, and analyze measurable aspects of a business

A solid BI architecture framework consists of:

- Collection of data
- Data integration
- Storage of data
- Data analysis
- Distribution of data
- Reaction based on insights

## 2.Business intelligence architecture components and diagram



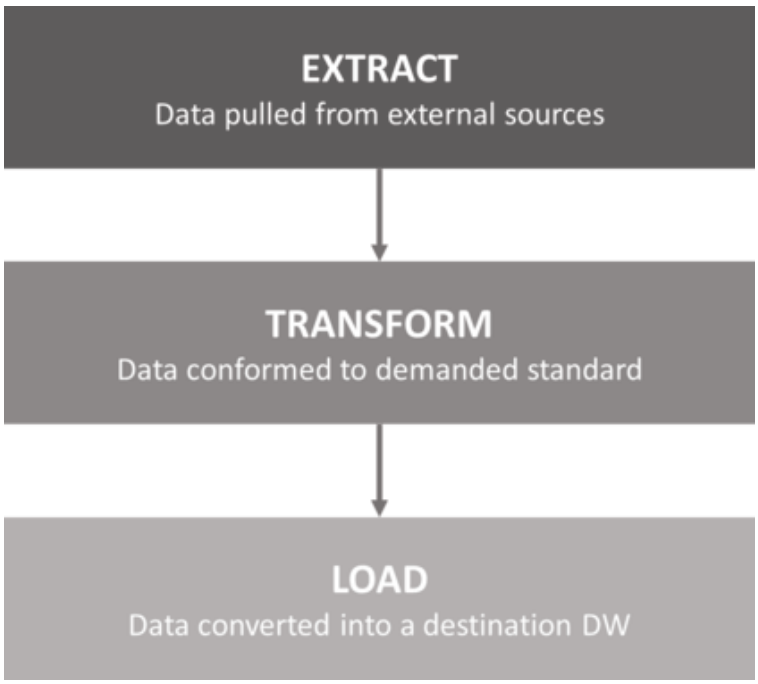
- As shown above the core components include the following items-

## 2.1.Data Collection –

- The first step in creating a stable architecture starts in gathering data from various data sources such as CRM, ERP, databases, files or APIs, depending on the requirements and resources of a company. They can also include secondary sources, such as market data and customer databases from outside information providers. As a result, both internal and external data sources are often incorporated into a BI architecture.
- Modern BI tools offer a lot of different, fast and easy data connectors to make this process smooth and easy by using smart ETL engines in the background.

## 2.2.Data Integration-

- When data is collected through scattered systems, the next step continues in extracting data and loading it to a data warehouse. This process is called ETL (Extract-Transform-Load).



- In extract the data is pulled from external sources. Secondly, data is transformed to the demanded standard i.e clean and prepared to the final stage and finally it is loaded into a data warehouse.

## 2.3. Data Storage-

- A data warehouse (DWH) stores all the company's data from several sources in a single place.
- Data cleansing, metadata management, data distribution, storage management, recovery, and backup planning are processes conducted in a data warehouse while BI makes use of tools that focus on statistics, visualization, and data mining.

## 2.4. Analysis of data-

- Data is analysed using BI tools after it's handled, processed, and cleaned in former steps with the help of data warehouse.

## 2.5. Data distribution-

- Data distribution comes as one of the most important processes when it comes to sharing information and providing stakeholders with indispensable insights to obtain sustainable business development. Distribution is usually performed in 3 ways:

- Reporting via automated e-mails: Created reports can be shared with selected recipients on a defined schedule. The dashboards will be automatically updated on a daily, weekly or monthly basis which eliminates manual work and enables up to date information.

- Dashboarding: Another reporting option is to directly share a dashboard in a secure viewer environment. The users you share with cannot make edits or change the content but can use assigned filters to manipulate data and interact with the dashboard. Another option is to share via public URL that enables users to access the dashboards even if they're outside of your organization.

## 2.6. Conclusions about insights

- CEOs, managers, professionals, coworkers, and all the interested stakeholders can use the power of data to generate valid, accurate, data-based decisions

- The dashboard is the final product on how data warehouse and business intelligence work together. Ultimately, this enables a high-level manager to get a comprehension of the strategic development and potential decisions for creating and maintaining a stable business.

# 3.Tableau Architecture

Tableau operations can be broadly classified into 3 main segments-

- Data sources
- Tableau Server and tableau desktop.
- Clients.

**3.1. Data sources-**
This section is where user can extract data from multiple data source options available like SAP,Oracle, Salesforce, Excel file, MySQL, Teradata etc. A user can establish data connection in two ways; a live data connection which sends instant queries to the data source and gets results instantly. Another way is to extract data from the data source and have a local copy of it as a temporary database. Data can be fetched through a live connection or extraction into both Tableau Desktop and Tableau Server.

**3.2.Tableau Desktop and server-**

**Tableau Desktop** is a tool where users create visualizations, workbooks, and dashboards using the data from the data source in Tableau**.** These workbooks or visualizations can be shared to users in many ways like web or mobile clients, onTableau Reader where users can view as well as edit the visualizations. Also, the visualization can be published or loaded into the Tableau Server.

The **Tableau Server** has many components working together as it manages a bunch of important processes. It has components for user and data security, a repository which stores all the visualizations published to the Server, a cache for performance improvement, a manager to manage data loads and schedule updates, a presentation layer which is responsible for all the visualization/presentation related activities.

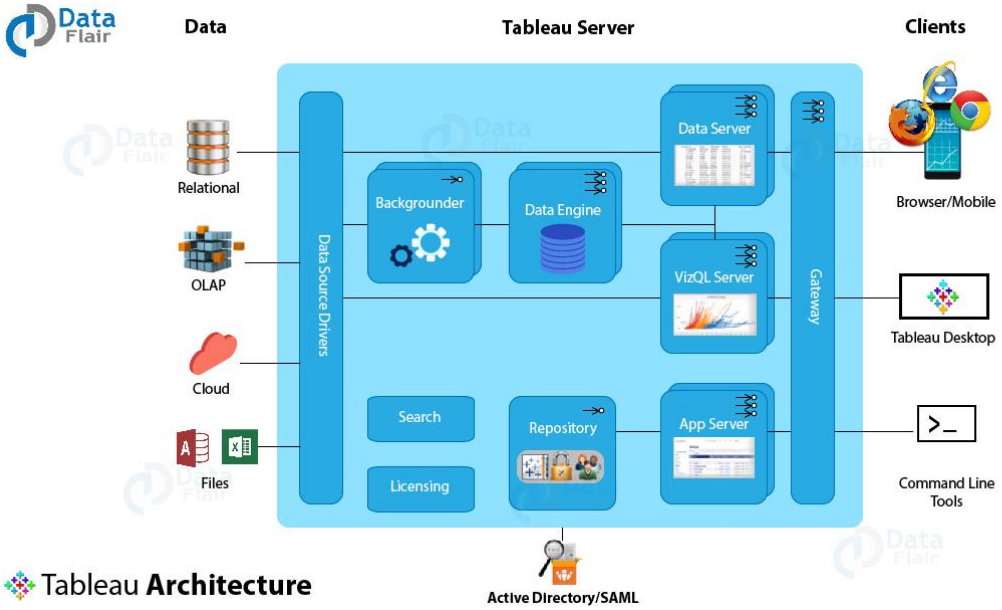
The Tableau Server primarily serves the dynamic user base of the web and mobile customers interacting with the data on Tableau platforms.

**3.3.Clients-**

The clients are the end users using Tableau through a web, mobile devices, on-cloud, on-premise, or on a command-line interface for development. These end-users interact mainly

with Tableau Server for accessing workbooks or visualizations. The server components then work as per the client's request and return the outcome.

# 4. Tableau Server Components and Architecture-




**4.1.Gateway-**
Gateway is a kind of web-server that helps clients communicate to the server via HTTP or https. The server receives incoming client requests and directs them to the appropriate server for action. A gateway handles processes such as load balancing, traffic routing, URL rewriting, serving static files to clients, serving multi-thread processes etc.

**4.2.Application Server-**
The application server deals with login processes, domain authentication, data authorization, user or group permission management, content searches, etc. It works in close association with the server's repository and handles the data access operations.

**4.3.Repository-**

The repository in Tableau Server stores server metadata related to users, permissions, assignments, groups, and projects. Along with the metadata, it stores visualizations in flat files (TWS, TDS), and performance data for auditing.

**4.4VizQL server-** It loads visualizations to work on Tableau. It has an in-built caching for performance improvement and editing tasks. Whenever a user requests a visualization or wishes to update an existing one, the request received by VizQL is first converted into an SQL statement and sent down to the data sources via respective data source drivers. The

requested data sent back from the data source comes to the VizQL server again, where it is processed with some final touches of additional calculations and sent to the user. Any new visualization coming from the data source is cached in the VizQL for further use.

**4.5.Data Engine-**

The data engine stores multiple TDEs (Tableau Data Extract) and can run on multiple servers maximum 2). It also attends multiple requests parallelly at a given point of time. The data engine hosts the piece of data in-memory extracted from the TDE upon getting a request from the user.

**4.6.Backgrounder-**

Backgrounder is a multi-process element that manages schedules for information refreshing and ensures proper functioning of the Tableau Server and Data Engine.

**4.7.Data Server-**

The data server helps in centralizing metadata management, driver deployment, and extract management. It also contributes to access control and serves as a proxy to the data sources. It hosts user queries and requests to prevent users from directly accessing the data source.

**4.8.Search and License**-

Two other important components are search and license. The search component manages the search indexing for the data in the repository. Whereas, the license component is responsible for the licensing and configuration of the Tableau server. Both these services run on the primary server of the Tableau's server cluster.