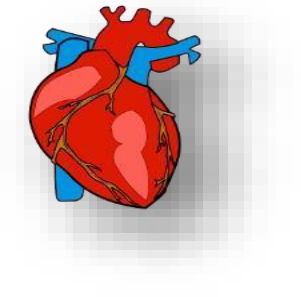
High Level Design (HLD)
Heart Disease Diagnostic Analysis



Contents

Abstract

Health is real wealth in the pandemic time. We all have realized the brute effects of covid-19 on all irrespective of any status. Heart disease is a term covering any disorder of the heart. Located in the chest region of the body, the heart beats around 80 beats per minute. Heart diseases have become a major concern to deal with as studies show that the number of deaths due to heart diseaseshave increased significantly over the past few decades in India it has become the leading cause of death in India.

Thus, preventing heart diseases has become more than necessary. Good data-driven systems for predicting heart diseases can improve the entire research and prevention process, making sure that more people can live healthy lives.

1 Introduction

Why this High-Level Design Document?

The purpose of this High-Level Design (HLD) Document is to add the necessary detail to thecurrent project description to represent a suitable model for coding. This document is also intended to help detect contradictions before coding and can be used as a reference manualfor how the modules interact at a high level.

The HLD will:

- Present all of the design aspects and define them in detail
- Describe the user interface being implemented
- Describe the hardware and software interfaces
- Describe the performance requirements
- Include design features and the architecture of the project
- List and describe the non-functional attributes like:

  -Security
  -Reliability
  -Maintainability
  -Portability
  -Reusability
  -Application compatibility
  -Resource utilization
  -Serviceability

Scope

The HLD documentation presents the structure of the system, such as the database architecture, application architecture (layers), application flow (Navigation), and technologyarchitecture. The HLD uses non-technical to mildly-technical terms which should be understandable to the administrators of the system.

2 General Description

Product Perspective & Problem Statement

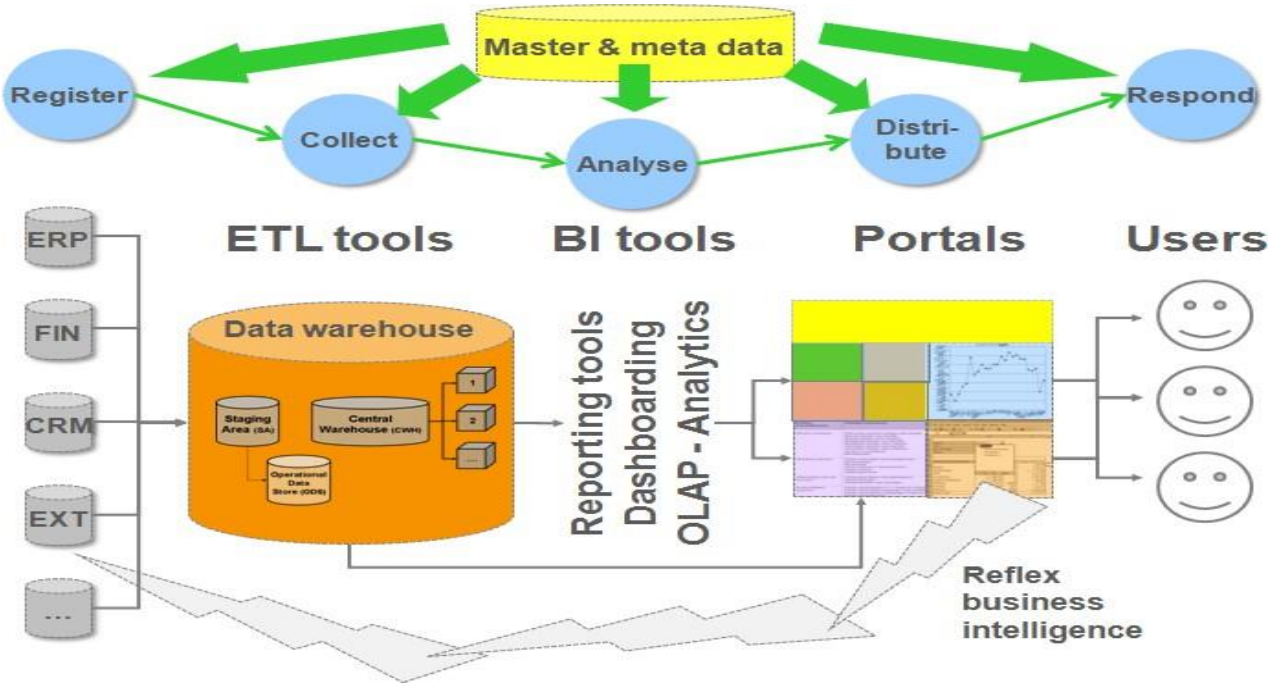
The goal of this project is to analyze to predict the probability of heart disease occurrence, based on a combination of features that describes the disease. To achieve the goal, we used a data set that is formed by taking into consideration some of the information of 303 individuals. The problem is based on the given information about each individual we have tocalculate that whether that individual will suffer from heart disease or not.

Tools used

Business Intelligence tools and libraries works such as NumPy, Pandas, Matplotlib, MS-Excel, Tableau, Jupyter Notebook and Python Programming Languageare used to build the whole framework.

3 Design Details

Functional Architecture



How BI Works

ORGANIZATIONAL MEMORY	INFORMATION INTEGRATION	INSIGHT CREATION	PRESENTATION
-Data Warehouse -Enterprise resource planning (ERP) -Knowledge Repository -Content Management	-Real Time Decision -Business Analytical Tools -Data Mining	-Text Mining Tool -Web Mining Tool -Environmental Scanning -RFID	-Online Analytical Processing (OLAP) Tool -Visualization Tool -Digital Dashboard -Score Card

Optimization

Your data strategy drives performance

- Minimize the number of fields
- Minimize the number of records
- Optimize extracts to speed up future queries by materializing calculations, removing columns and the use of accelerated views

Reduce the marks (data points) in your view

- Practice guided analytics. There's no need to fit everything you plan to show in a single view. Compile related views and connect them with action filters to travel from overview to highly-granular views at the speed of thought.
- Remove unneeded dimensions from the detail shelf.
- Explore. Try displaying your data in different types of views.

Limit your filters by number and type

- Reduce the number of filters in use. Excessive filters on a view will create a more complex query, which takes longer to return results. Double-check your filters and remove any that aren't necessary.
- Use an include filter. Exclude filters load the entire domain of a dimension while including filters do not. An include filter runs much faster than an exclude filter, especially for dimensions with many members.
- Use a continuous date filter. Continuous date filters (relative and range-of-date filters) can take advantage of the indexing properties in your database and are faster than discrete data filters.
- Use Boolean or numeric filters. Computers process integers and Booleans (t/f) much faster than strings.
- Use parameters and action filters. These reduce the query load (and work across data sources).

Optimize and materialize your calculations

- Perform calculations in the database
- Reduce the number of nested calculations.
- Reduce the granularity of LOD or table calculations in the view. The more granular the calculation, the longer it takes.
    - LODs - Look at the number of unique dimension members in thecalculation.
    - Table Calculations - the more marks in the view, the longer it will take to calculate.

- Where possible, use MIN or MAX instead of AVG. AVG requires more processing than MIN or MAX. Often rows will be duplicated and display the same result with MIN, MAX, or AVG.
- Make groups with calculations. Like include filters, calculated groups load only named members of the domain, whereas Tableau's group function loads the entire domain.
- Use Booleans or numeric calculations instead of string calculations. Computers can process integers and Booleans (t/f) much faster than strings. Boolean>Int>Float>Date>Date Time>String.

4 Key Performance Indicators

Dashboards will be implemented to display and indicate certain KPIs and relevantindicators for the disease.

As and when the system starts to capture the historical/periodic data for a user, The dashboards will be included to display charts over time with progress on various indicators or factors

KPIs (Key Performance Indicators)

Key indicators displaying a summary of the heart disease and its relationship with different metrics
1. Percentage of People Having Heart Disease
2. Variation of ‘thal’ (Thalassemia type) with ‘sex’
3. Variation of ‘chol’ (Cholesterol), ‘trestbps’ (Resting blood pressure) with ‘fbs’ (Fasting Blood Sugar).
4. Variation of ‘exang’ (Exercise induced angina) with ‘cp’ (Chest Pain type).
5. Variation of ‘num’ (Angiographic disease status) with ‘sex’.
6. Variation of the ‘age’ with ‘chol’ (Cholesterol) and ‘sex’
7. Variation of ‘cp’ (Chest Pain type) with ‘sex’
8. Variation of ‘thalach’ (Maximum heart rate) with ‘age’
9. Variation of ‘restecg’ (Resting electrocardiograph results) with ‘sex’
10.Variation of ‘slope’ (Slope of the peak exercise ST segment), ‘restecg’ (Resting Electrocardiograph results) and ‘oldpeak’ (ST depression induced by exercise relative to rest)

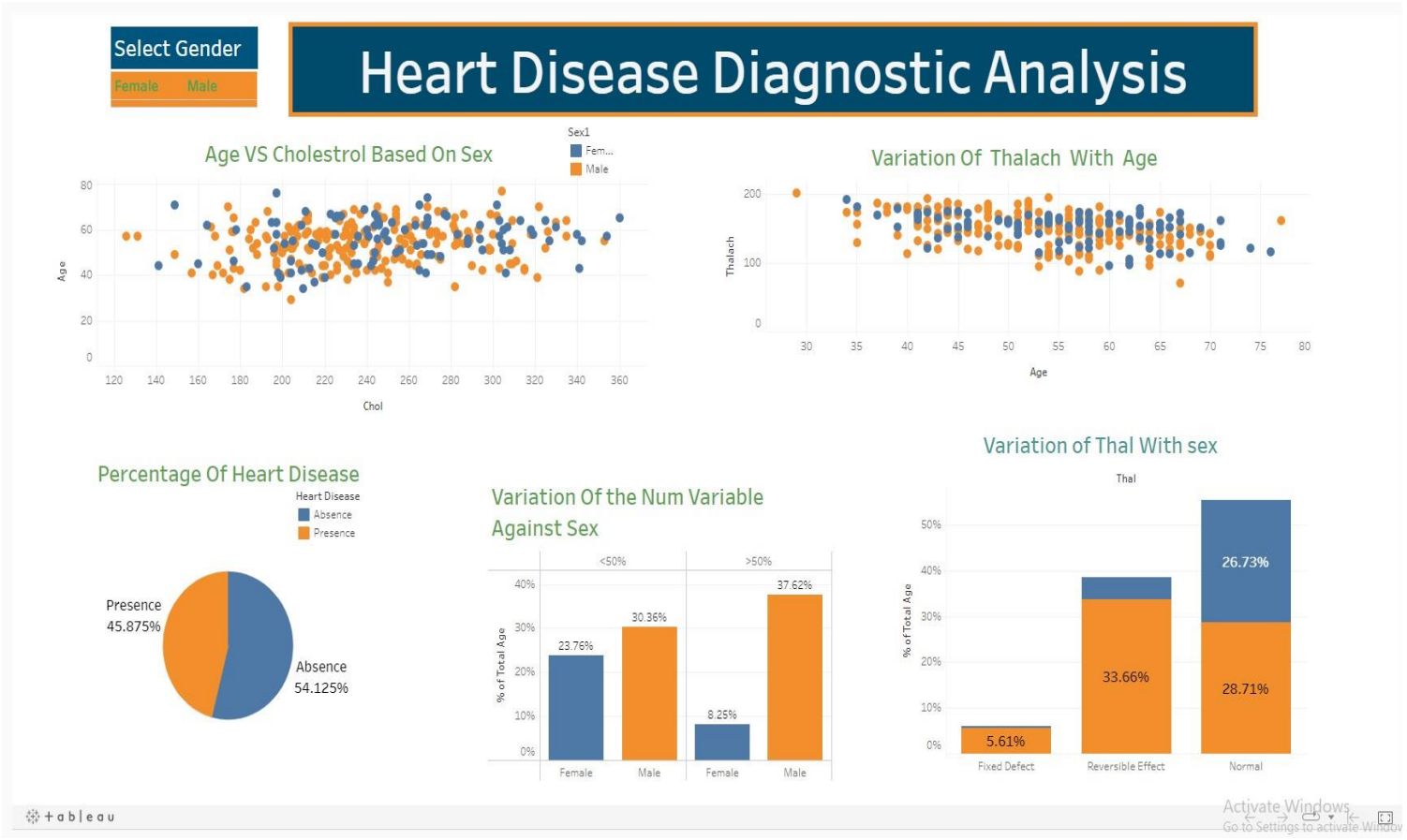
5 Deployment

Prioritizing data and analytics couldn’t come at a better time. Your company, no matter what size, is already collecting data and most likely analyzing just a portion of it to solvebusiness problems, gain competitive advantages, and drive enterprise transformation. With the explosive growth of enterprise data, database technologies, and the high demand for analytical skills, today’s most effective IT organizations have shifted their focus to enabling self-service by deploying and operating Tableau at scale, as well as organizing, orchestrating, and unifying disparate sources of data for business users andexperts alike to author and consume content.
Tableau prioritizes choice in flexibility to fit to the enterprise architecture. Tableau Server and Tableau Online leverages your existing technology investments and integrate into your IT infrastructure to provide a self-service, modern analytics platform for users. With on-premises, cloud and hosted options, there is a version of Tableau to match your requirements. Below is a comparison of the three types:
Types:

A. Tableau Server - On Premises
- Full control of hardware and software
- Infrastructure and data remain behind your firewall
- Need dedicated administrators to manage hardware and software
- Additional infrastructure needed to access off-network (mobile, external)

B. Tableau Server - Public Cloud (IaaS)
- Full control of software on managed hardware
- Puts infrastructure in same place as data (for migration to cloud)
- Flexibility to spin up/down hardware as needed
- Need dedicated administrators to manage software
- Additional infrastructure needed to access off-network (mobile, external)

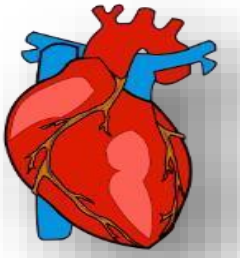
C. Tableau Online (SaaS)
- Fully hosted solution (hardware, software upgrades)
- Fast to deploy


Select Gender
Female Male
Heart Disease Diagnostic Analysis
Age VS Cholestrol Based On Sex
Sex1
Fem...
Male
Age
Chol
Variation Of Thalach With Age
Thalach
Age
Percentage Of Heart Disease
Heart Disease
Absence
Presence
Presence 45.875%
Absence 54.125%
Variation Of the Num Variable Against Sex
<50%
>50%
% of Total Age
23.76%
30.36%
8.25%
37.62%
Female
Male
Variation of Thal With sex
Thal
% of Total Age
5.61%
33.66%
28.71%
26.73%
Fixed Defect
Reversible Effect
Normal
+ableau
Activate Windows


Variation Of ChestPain vs Sex
Female
Male
Cp
Asymptomatic
Atypical Angina
Non-Anginal Pain
Typical Angina
% of Total Count of Age
Asymptoma..
Atypical Angina
Non-Anginal Pain
Typical Angina
Variation Of Exang With Chest Pain
Exang / Cp
No
Yes
% of Total Age
6.27%
15.18%
24.75%
21.12%
1.32%
1.32%
3.63%
26.40%
Typical Angina
Atypical Angina
Non-Anginal Pain
Asymptomatic
Variation Of Slope And Restecg With Old Peak
Restecg
left ventricular hypertrophy
Normal
st-t wave abnormality
Slope / Restecg
Downsloping
Flat
Unsloping
Avg. Oldpeak
4.000
2.500
2.000
2.000
1.171
1.164
0.917
0.732
st-t wave abnormality
left ventricular..
Normal
Avg. Chol VS FBS
True
False
244.0889
243.4070
Avg. Chol
Avg. Trestbps VS Fbs
Fbs
False
True
130.403
139.067
Avg. Trestbps
+ableau
Activate Windows

Low-Level Design (LLD)

Heart Disease Diagnostic Analysis



Contents

1. Introduction

What is Low Level Design Document?

The goal of the Low-level design document (LLDD) is to give the internal logic design of the actual program code for the Heart Disease Diagnostic Analysis dashboard. LLD describes the class diagrams with the methods and relationsbetween classes and programs specs. It describes the modules so that the programmer can directly code the program from the document.

What is Scope?

Low-level design (LLD) is a component-level design process that follows a step-by-step refinement process. The process can be used for designing data structures, required software architecture, source code, and ultimately, performance algorithms. Overall, the data organization may be defined duringrequirement analysis and then refined during data design work.

Project Introduction

Heart disease is a term covering any disorder of the heart. Heart diseases have become a major concern to deal with as studies show that the number of deaths due to heart diseases has increased significantly over the past few decades in India it has become the leading cause of death in India.

Thus, preventing heart diseases has become more than necessary. Good data-driven systems for predicting heart diseases can improve the entire research and prevention process, making sure that more people can live healthy lives.

2. Problem Statement

Health is real wealth in the pandemic time we all realized the brute effects of covid-19on all irrespective of any status. You are required to analyze this health and medical data for better future preparation. A dataset is formed by taking into consideration some of the information of 303 individuals.

3. Dataset Overview

Input Dataset Information

The dataset used for analysis is the heart disease dataset provided by the UCI Repository. It contains 76 attributes out of which only 14 are used. We will be using the Cleveland dataset.

Dataset source: https://archive.ics.uci.edu/ml/datasets/Heart+Disease

Dataset Variables Description is as follows:

age: The person's age in years

sex: The person's sex (1 = male, 0 = female)

cp: The chest pain experienced (Value 1: typical angina, Value 2: atypical angina,Value 3: non-anginal pain, Value 4: asymptomatic)

trestbps: The person's resting blood pressure (mm Hg on admission to the hospital)

chol: The person's cholesterol measurement in mg/dl

fbs: The person's fasting blood sugar (> 120 mg/dl, 1 = true; 0 = false)

restecg: Resting electrocardiographic measurement (0 = normal, 1 having ST-Twave abnormality, 2 = showing probable or definite left ventricular hypertrophy by Estes' criteria)

thalach: The person's maximum heart rate achieved

exang: Exercise induced angina (1 = yes; 0 = no)

oldpeak: ST depression induced by exercise relative to rest

slope: the slope of the peak exercise ST segment (Value 1: upsloping, Value 2: flat,Value 3: downsloping)

ca: The number of major vessels (0-3)

thal: A blood disorder called thalassemia (3 = normal; 6 = fixed defect; 7 =reversable defect)

num: heart disease (0 = no, 1 = yes)

Data Preprocessing

We will be using Python Numpy and Pandas to perform Preprocessing on the dataset. Observations after Exploratory Data Analysis:

1. The data in the dataset showed that there are some columns in the dataset which are categorical variables but when loaded into Tableau behave as numerical variables. Even though they contain numeric data the values in them are repeating and only limited to a few numbers which means they have been encoded to represent some specific class/category under the variable. The columns include – sex, cp, FBS, restecg, exang, slope, ca, thal and num.
As a part of preprocessing, the data type of these columns will be changed From integer to categorical.
2. no column contains missing data as indicated by the count parameter. But still, we need to check for incorrect data.
3. There are a few columns wherein there are unusual values/outliers. After observation based on the summary statistics of these columns, it is somewhat clear that these are outliers. Maybe it is an outlier. The columns are – cp and thal.
4. The old peak column has many values as zeros and the data is also skewed. As a part of the transformation, we will impute the zeros with either mean/median of the column values because the old peak values cannot be zero for a human being and also remove skewness by using Logarithmic transformation.
5. Remaining numerical columns seem normally distributed

Data Modification in Tableau Desktop

The exported .csv dataset file – preprocessed_heart_disease_dataset.csv will be imported into Tableau Public Desktop.
Since the dataset contains many categorical columns which store the categories

in the form of integers, we will convert these numbers into meaningful phrases which will be understandable to the viewer and also easy to understand the terms used in the visualizations.
Since the dataset contains many categorical columns which store the categories in the form of integers, we will convert these numbers into meaningful phrases which will be understandable to the viewer and also easy to understand the terms used in the visualizations.

4. Architecture

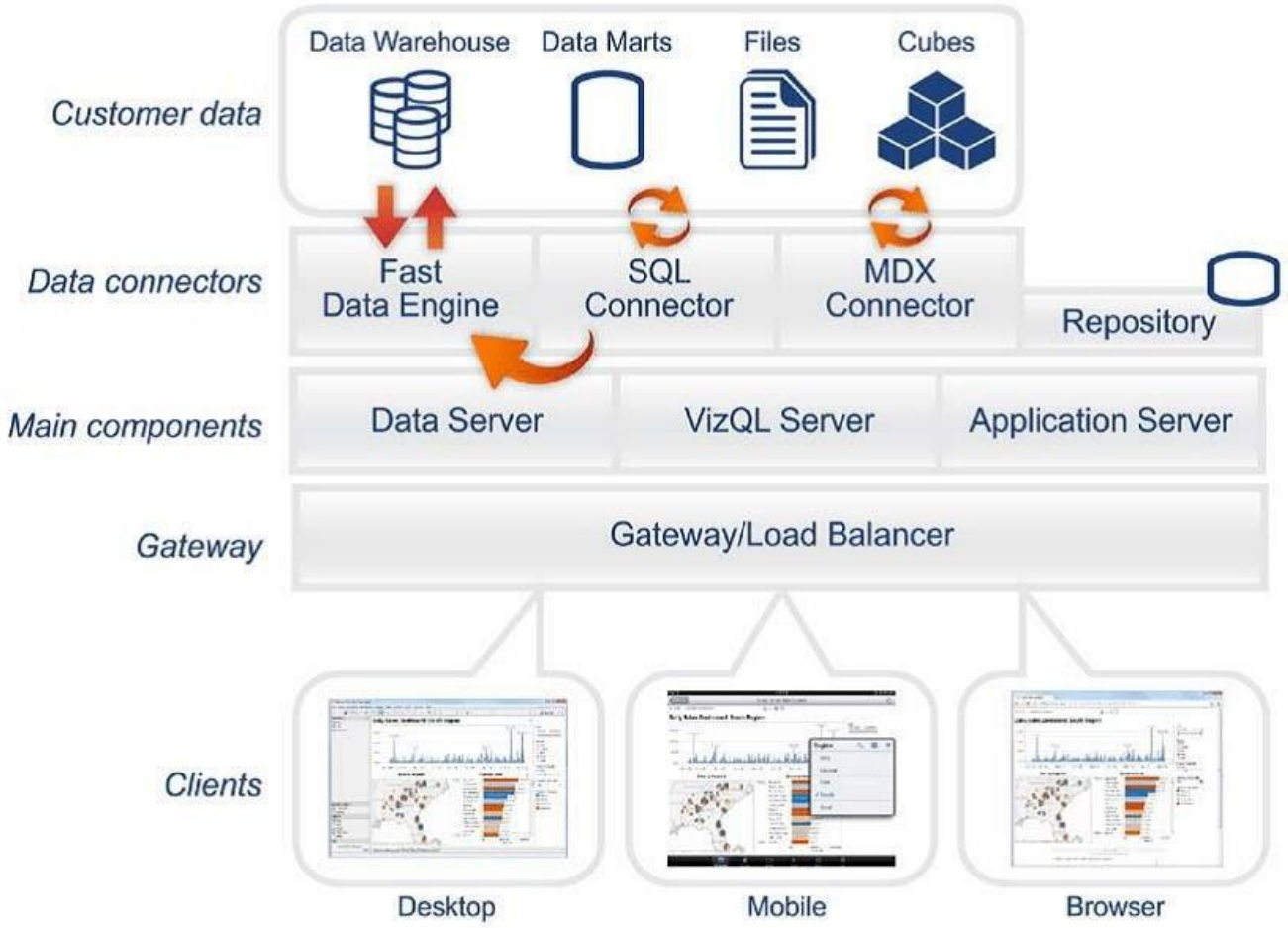
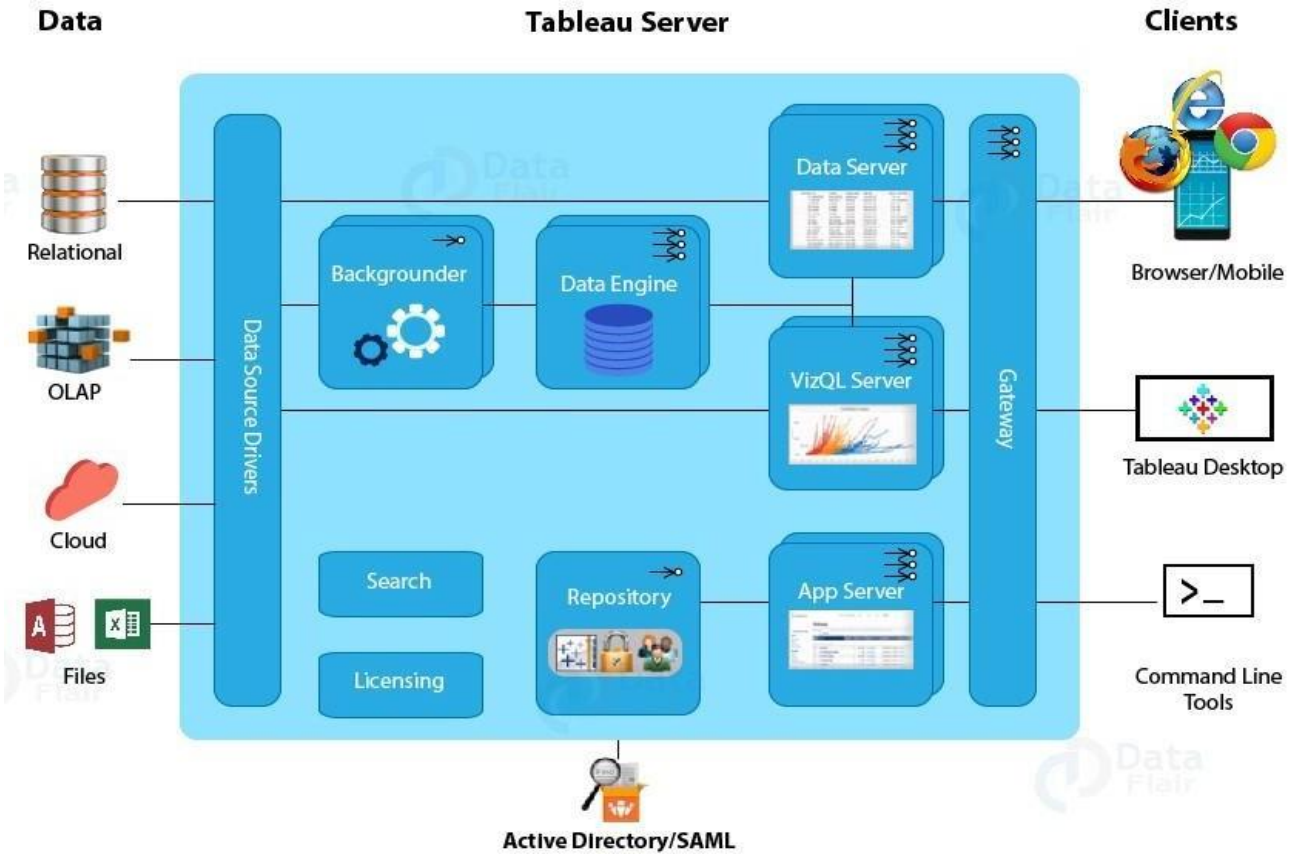


Tableau Server Architecture

Tableau has a highly scalable, n tier client architecture that serves the mobile Architecture that serves the mobile clients web clients and desktop-installed Software

Tableau Server architecture supports fast and flexible deployments.
The below diagram shows Tableau Server's architecture.
The Tableau Server has many components working together as it manages a bunch of important processes. It has components taking care of the user and data security a repository which stores all the visualizations published to the Server, a cache for performance improvement, a manager/automation 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



5. Architecture Description

1. **Raw Data Collection**
   The Dataset was taken from iNeuron's Provided Project Description Document.
   https://drive.google.com/drive/folders/165Pjmfb9W9PGy0rZjHEA22LW0Lt3Y-Q8

2. **Data Pre-Processing**

Before building any model, it is crucial to perform data pre-processing to feed the correct data to the model to learn and predict. Model performance depends on the quality of data fed to the model to train.
This Process includes-

a) Handling Null/Missing Values
b) Handling Skewed Data
c) Outliers Detection and Removal

3. Data Cleaning

Data cleaning is the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset.

a) Remove duplicate or irrelevant observations
b) Filter unwanted outliers
c) Renaming required attribute

4. Exploratory Data Analysis (EDA)

Exploratory Data Analysis refers to the critical process of performing initial investigations on data to discover patterns, spot anomalies, test hypotheses and check assumptions with the help of summary statistics and graphical representations.

5. Reporting

Reporting is a most important and underrated skill of a data analytics field. Because being a Data Analyst you should be good in the easy and self-explanatory report because your model will be used by many stakeholders who are not from a technical background.

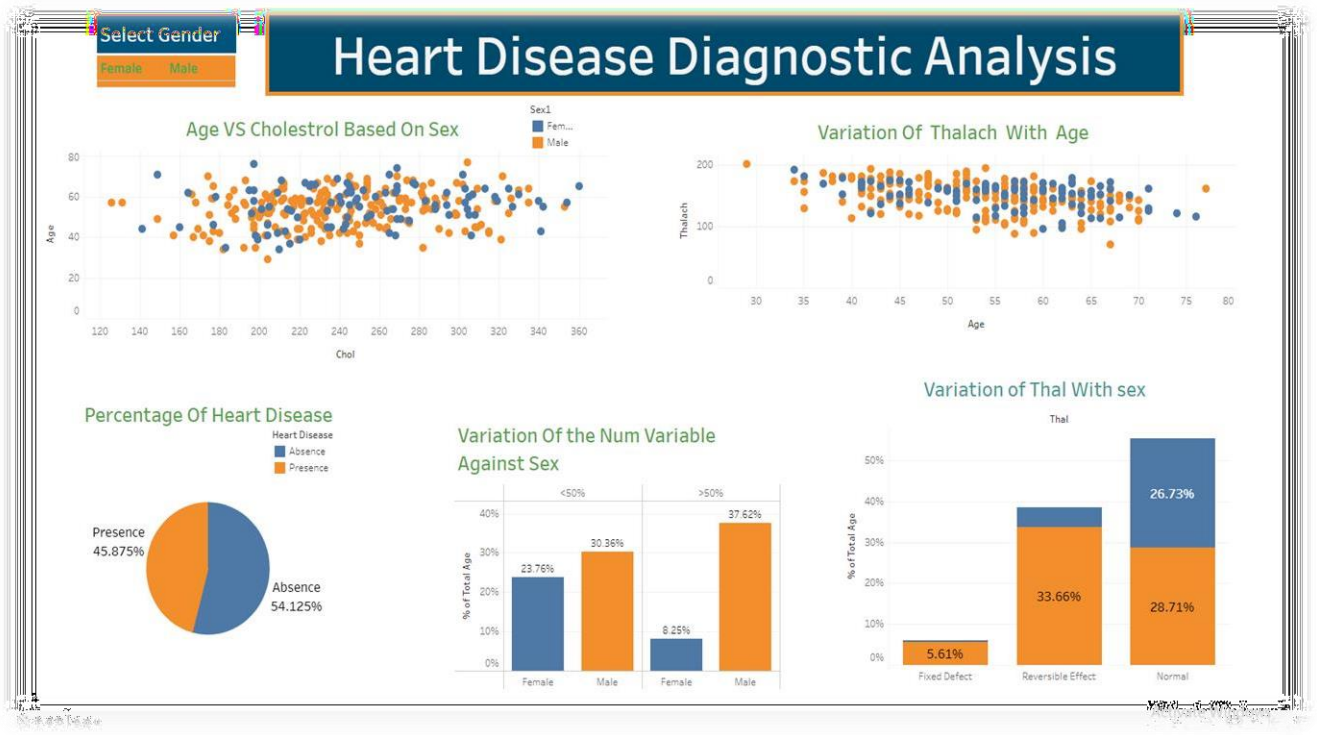
a) High-Level Design Document (HLD)
b) Low Level Design Document (LLD)
c) Architecture
d) Wireframe
e) Detailed Project Report

6. Modelling

Data Modelling is the process of analyzing the data objects and their relationship to the other objects. It is used to analyze the data requirements that are required for the business processes. The data models are created for the data to be stored in a database. The Data Model's main focus is on what data is needed and how we have to organize data rather than what operations we have to perform.

7. Deployment

We created a Tableau Dashboard


Select Gender
Female
Male
Heart Disease Diagnostic Analysis
Age VS Cholestrol Based On Sex
Sex1
Fem...
Male
Age
Chol
Variation Of Thalach With Age
Thalach
Age
Percentage Of Heart Disease
Heart Disease
Absence
Presence
Presence
45.875%
Absence
54.125%
Variation Of the Num Variable
Against Sex
<50%
>50%
% of Total Age
23.76%
30.36%
8.25%
37.62%
Female
Male
Female
Male
Variation of Thal With sex
Thal
% of Total Age
5.61%
33.66%
26.73%
28.71%
Fixed Defect
Reversible Effect
Normal


Variation Of ChestPain vs Sex
Female
Male
Cp
Asymptomatic
Atypical Angina
Non-Anginal Pain
Typical Angina
% of Total Count of Age
Asymptoma..
Atypical Angina
Non-Anginal Pain
Typical Angina
Variation Of Exang With Chest Pain
Exang / Cp
No
Yes
% of Total Age
6.27%
15.18%
24.75%
21.12%
1.32%
1.32%
3.63%
26.40%
Typical Angina
Atypical Angina
Non-Anginal Pain
Asymptomatic
Variation Of Slope And Restecg With Old Peak
Restecg
left ventricular hypertrophy
Normal
st-t wave abnormality
Slope / Restecg
Downsloping
Flat
Unsloping
Avg. Oldpeak
4.000
2.500
2.000
2.000
1.171
1.164
0.917
0.732
st-t wave abnormality
left ventricular..
Normal
st-t wave abnormality
left ventricular..
Normal
left ventricula..
Normal
Avg. Chol VS FBS
True
False
244.0889
243.4070
Avg. Chol
Avg. Trestbps VS Fbs
Fbs
False
True
130.403
139.067
Avg. Trestbps
+ableau