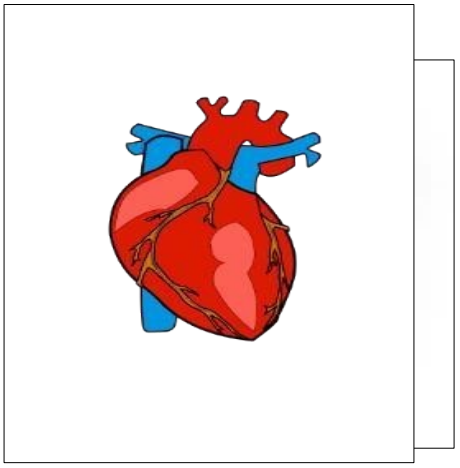
Low-Level Design (LLD)
(LLD)
Heart Disease Diagnostic Analysis



S.Nanthini

Contents

Document Version Control.....2
1. Introduction..... 4
What is Low Level Design Document?.....4
Scope.…..... 4
Project Introduction.....4
2. Problem Statement.....5
3. Dataset Overview.....5
Input Dataset Information.....6
Data Preprocessing.....6
Data Modification in Tableau Desktop.....6
4. Architecture..... 7
5. Architecture Description.....8

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)

1. Data Preprocessing

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

2. 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.

3. no column contains missing data as indicated by the count parameter. But still, we need to check for incorrect data.
4. 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.
5. 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.

6. 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 categoriesin 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

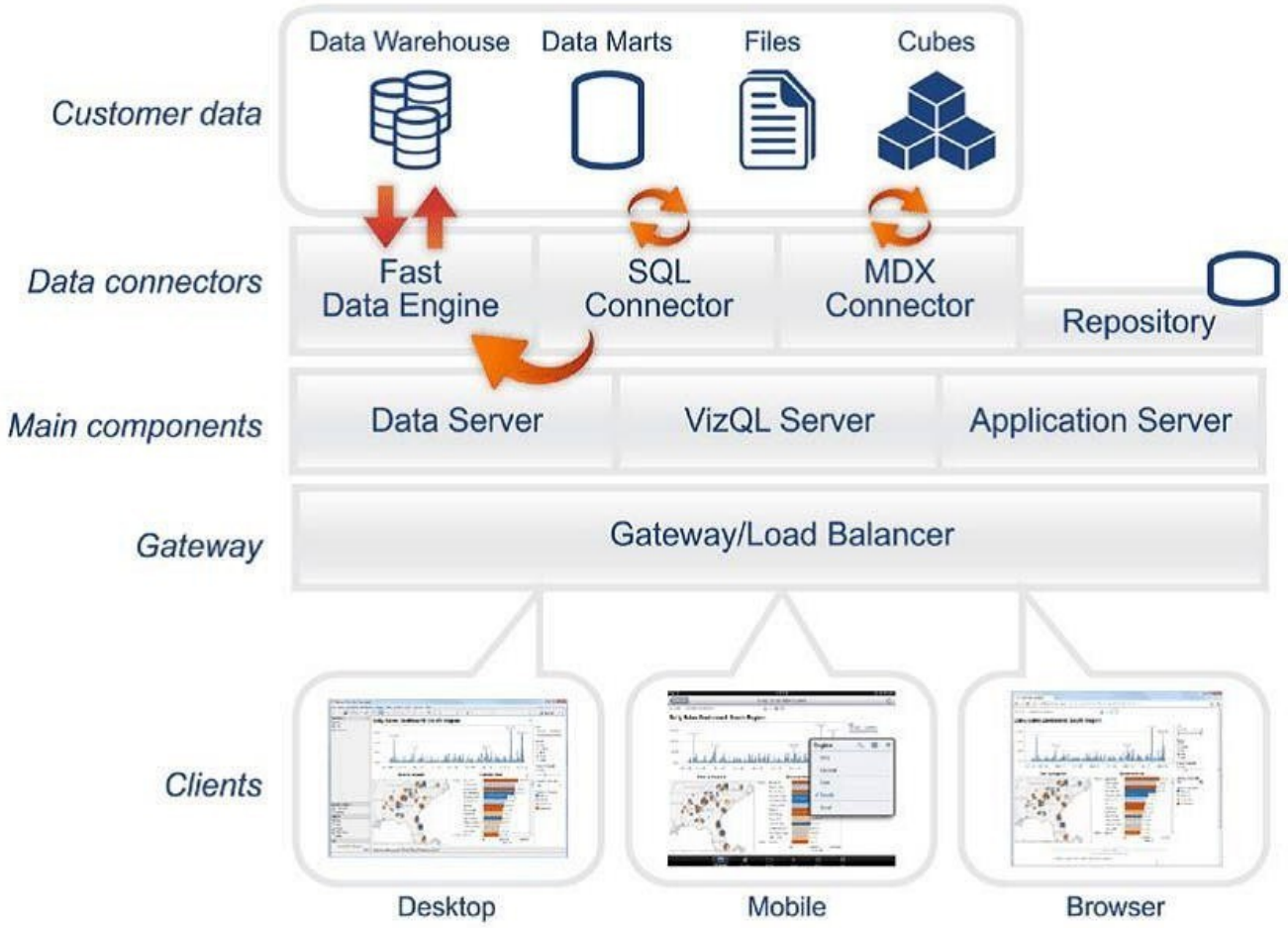
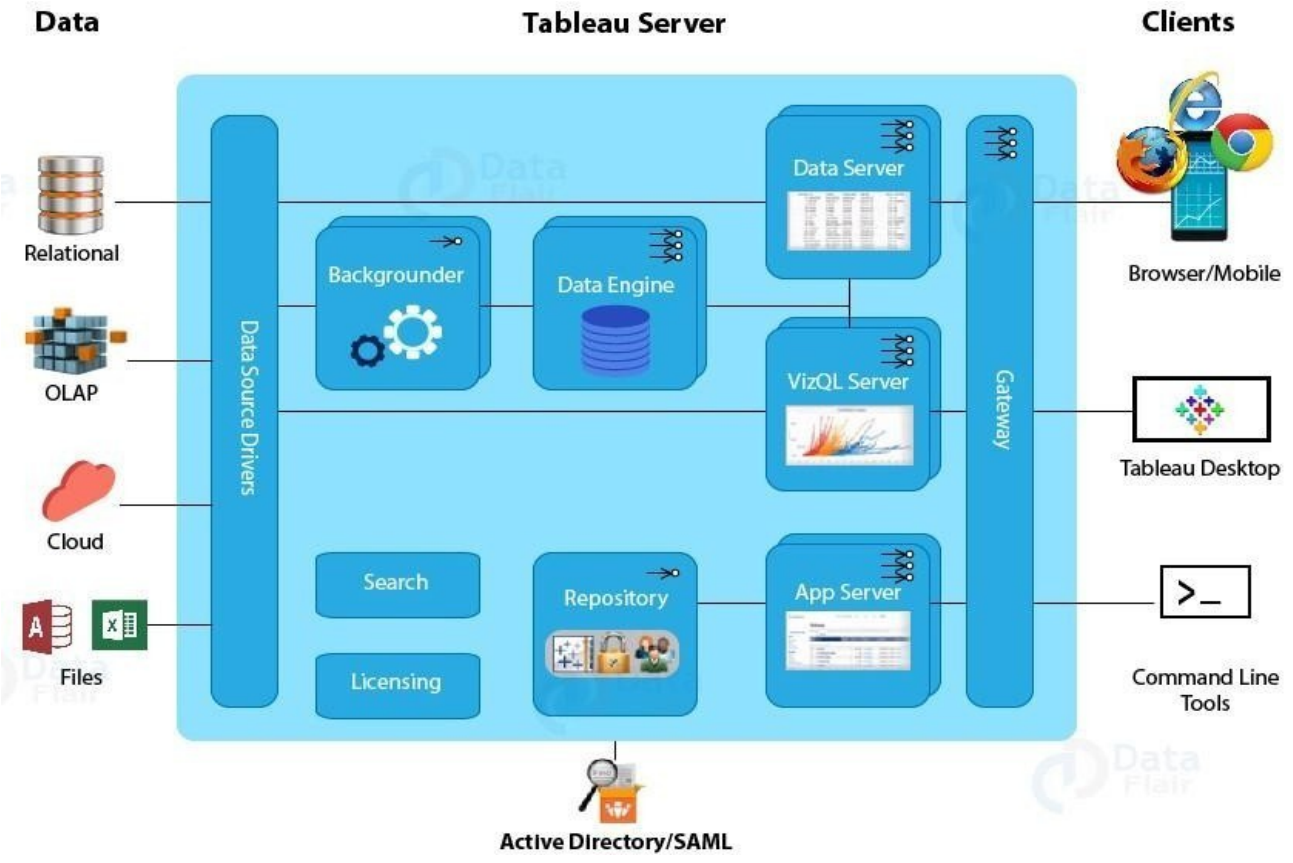


Tableau Server Architecture

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

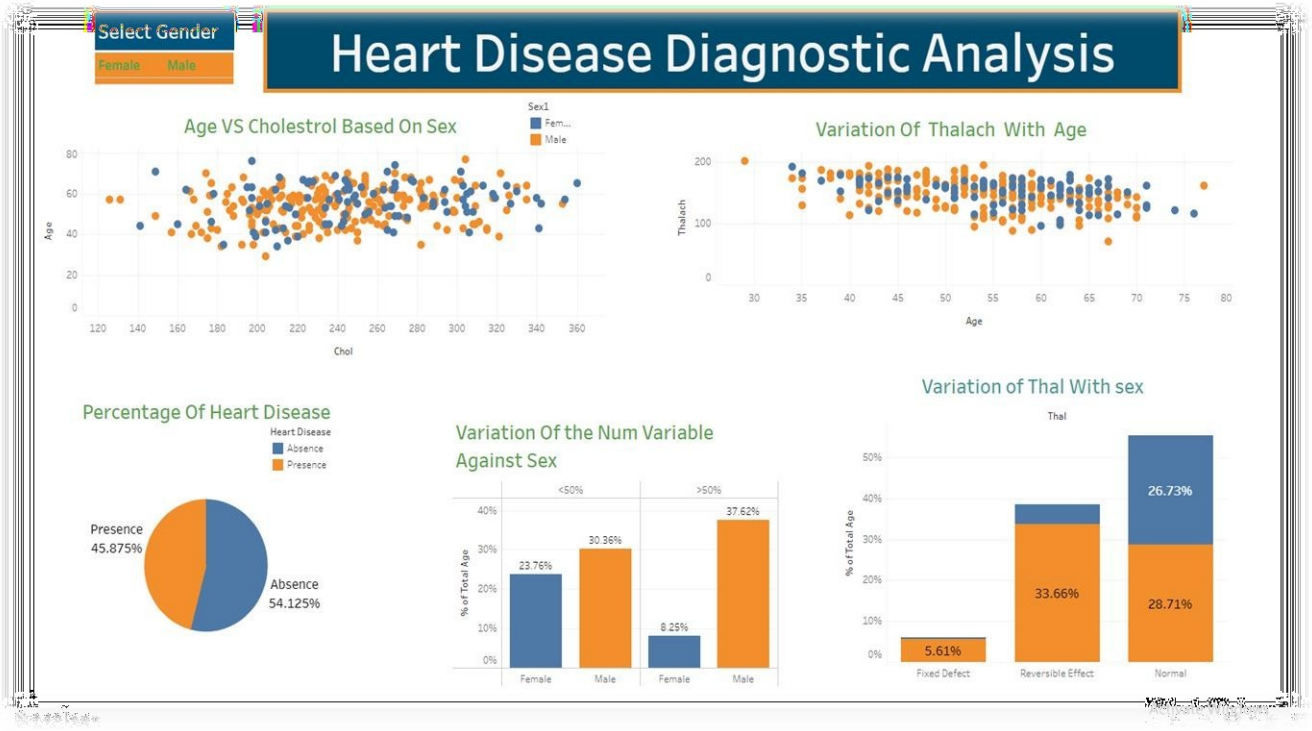
   Data Modelling is the process of analyzing the data objects and their relationship to the other objects. It is used to analyze the data requirements that are required for the business processes. The data models are created for the data to be stored in a database. The Data Model's main focus is on what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

Link to tableau

(https://public.tableau.com/app/profile/nanthini.s8783/viz/HeartDiseaseDiagnosticAnalysis_16924221730460/HeartDiseaseDiagnosticAnalysis)


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%
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
Slope / Restecg
Downsloping
Flat
Unsloping
Restecg
left ventricular hypertrophy
Normal
st-t wave abnormality
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
left ventricula..
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
Go to Settings to activate Win