LOW LEVEL DESIGN DOCUMENT

Heart Disease Diagnostic Analysis

BY
VIKASH KUMAR MAHAPATRA
Revision Number & Date :- 1.2 31/08/2022

Document Version Control

Date Issued	Version	Description	Author
29 JULY 2022	1.0	Created Document And Add Document Version Control, Content Page, Introduction Page	Vikash
05 AUG 2022	1.1	Created Architecture Page ,Architecture Description Page & Add page number to contents section	Vikash
31 AUG 20222	1.2	Final Revision	Vikash

CONTENTS

INTRODUCTION

WHY THIS LOW LEVEL DESIGN DOCUMENT

The purpose of this Low-Level Design (LLD) document is to give the internal logic design of the actual program code for Foreign Direct Investment Analysis .LLD document describes the class diagrams with the methods and relations between classes and program specs. It describes the modules so that the programmer can directly code the program from the document.

SCOPE

The LLD document 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 performance algorithms. Overall, the data organization may be defined during requirement analysis and then refined during data design work.

PROJECT INTRODUCTION

Heart disease is a general term that means the heart isn't working normally. Babies can be born with heart disease. This is called congenital heart disease. If people get heart disease later, it is called acquired heart disease and most of them are acquired. The most common types of acquired heart disease are CAD (Coronary Artery Disease), CHF (Congestive Heart Failure), Bad Heart Rhythms. India has one of the highest burdens of CVD worldwide. According to a report the annual numbers of deaths from CVD in India rise from 2.26million(1990) to 4.77million(2020).Coronary heart disease prevalence rate in India have been estimated over the past several decades & have ranged from 1.6% to 7.4% in rural populations and from 1% to 13.2% in urban population. The prevalence rate increased with age from 22% in 45-54 to 38% in age 70 and above as per report. Cure of heart disease has become more than necessary. Informative-data can systems help predicting early heart diseases, improve the research and make prevention process more effective as well as reduction of any future mishappenings.

PROBLEM STATEMENT

The problem statement aims to analyze certain parameter from given three hundred three individual information come to conclusion that whether the individual will suffer from heart disease or not. To achieve the goal, we used a data set that is given and analyze most important parameter that are responsible for heart disease. Also get some meaningful insights from the given information regarding Heart Disease.

Do ETL (Extract-Transform-Load) the dataset and find for me some information from this large data. This is form of data mining. What all information can be achieved by mining this data, would be brainstormed by interns. Find key metrics and factors and show the meaningful relationship between attributes. Do your own research and come up with findings.

ARCHITECTURE

Microsoft Power BI architecture consists of four major steps that explain the whole process from data sourcing to creation of reports and dashboard. Various technologies and process work together to get required results with extreme precision. Let us see those steps in pictorial format.

POWER BI ARCHITECTURE
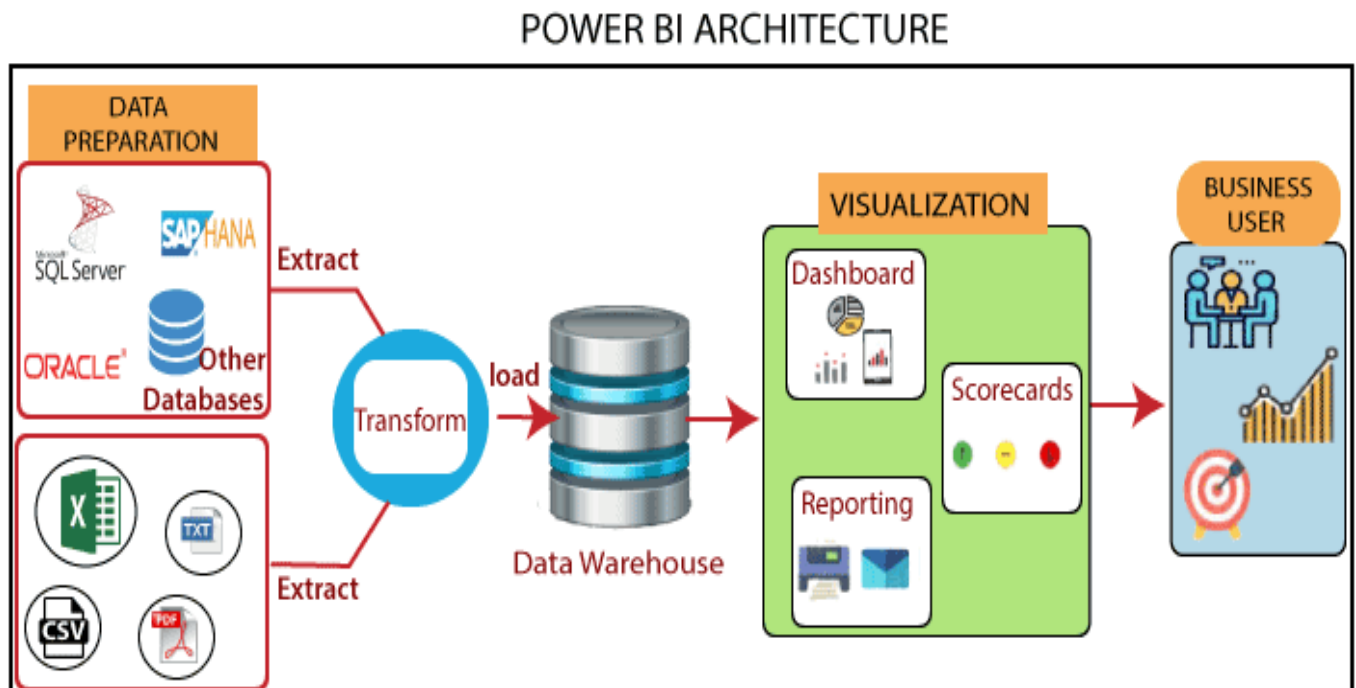
Sourcing data: Power BI extracts data from various servers, Excel sheets, CSV files and databases. The extracted information can be directly imported to Power BI or a live service link is established to receive it. If you directly import the data in Power BI, it will only be compressed up to 1GB post that you can only run live queries on your chunky datasets.

Transforming the data: Before visualizing the data, cleaning and preprocessing it should be done. This means useless or missing values from rows or columns. Following that certain rules will be applied to transform and load datasets into the warehouse.

Report and publish: After cleaning and transforming the data, reports will be created based on requirements. A report is a visualization of the data with different filters and constraints presented in the form of graphs, pie-chart, and other figures.

Creating dashboards: Power BI Dashboards are created by pinning individual elements or pages of live reports. Dashboard should be created after you have published your reports to the BI service. When the reports get saved, the visual maintains the filter setting chosen so that the user can apply filters and slicers.

ARCHITECTURE DESCRIPTION

Data Sourcing:

The dataset is in csv (comma separated value) format and Microsoft excel is used to load the data. Dataset is taken from project description document and here goes the drive link https://drive.google.com/drive/folders/165Pjmfb9W9PGy0rZjHEA 22LW0Lt3Y-Q8

Data Description:

The file name heart_disease_dataset.csv contains different columns such as age, chest pain (cp), cholesterol (chol), thalassemia (thal), and fasting blood sugar (fbs),num(represent who have disease and who havenot), etc. . Sample data below.

age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	num
63	1	1	145	233	1	2	150	0	2.3	3	0	6	0
67	1	4	160	286	0	2	108	1	1.5	2	3	3	1
67	1	4	120	229	0	2	129	1	2.6	2	2	7	1
37	1	3	130	250	0	0	187	0	3.5	3	0	3	0
41	0	2	130	204	0	2	172	0	1.4	1	0	3	0
56	1	2	120	236	0	0	178	0	0.8	1	0	3	0
62	0	4	140	268	0	2	160	0	3.6	3	2	3	1
57	0	4	120	354	0	0	163	1	0.6	1	0	3	0
63	1	4	130	254	0	2	147	0	1.4	2	1	7	1
53	1	4	140	203	1	2	155	1	3.1	3	0	7	1
57	1	4	140	192	0	0	148	0	0.4	2	0	6	0
56	0	2	140	294	0	2	153	0	1.3	2	0	3	0
56	1	3	130	256	1	2	142	1	0.6	2	1	6	1
44	1	2	120	263	0	0	173	0	0	1	0	7	0
52	1	3	172	199	1	0	162	0	0.5	1	0	7	0
57	1	3	150	168	0	0	174	0	1.6	1	0	3	0
48	1	2	110	229	0	0	168	0	1	3	0	7	1
54	1	4	140	239	0	0	160	0	1.2	1	0	3	0
48	0	3	130	275	0	0	139	0	0.2	1	0	3	0
49	1	2	130	266	0	0	171	0	0.6	1	0	3	0
64	1	1	110	211	0	2	144	1	1.8	2	0	3	0
58	0	1	150	283	1	2	162	0	1	1	0	3	0
58	1	2	120	284	0	2	160	0	1.8	2	0	3	1

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 following steps:- Handing Missing/Null values , Handling Skewed Data and Outliers Detection and removal.

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 attributes.

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

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)HLD(High Level Document)b)LLD(Low Level Document)c)Architecture Report. d) Wireframe report. e) Detailed Project Report.

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 performing the operations that are need.

Deployment: We create a Power BI Dashboard