Low Level Design (LLD)

Analysing Google Apps Store

Contents

Document Version Control

Date Issued	Version	Description	Author
10/12/2022	1.0	Google play store apps analysis	Sakshi Bhalothia

1. Introduction

1.1 What is Low Level Design Document

The goal of the LLD document is to give the internal logic design of the actual program code for the Google Apps Store Analysis dashboard. LDD describes the class diagrams with the methods and relations between classes and programs specs. It describes the modules so that the programmer can directly code the program from the document.

1.2 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 during requirement analysis and then refined during data design work.

2. Problem Statement

The objective of the project is to perform data visualization techniques to understand insights of the data. This project aims to apply various Business Intelligence tools such as Tableau to get a visual understanding of the data.

3. Dataset Overview

Input Dataset Information

We are provided with two datasets for this project – googleplaystore and googleplaystore_user_reviews

Variables Description is as follows:

Googleplaystore dataset:

App: name of the app

Category: category to which the app belongs

Rating: rating given to the app by user

Reviews: number of reviews given to the app

Size: space app will take in device

Installs: number of installs of the app

Type: determines whether the app is free or paid

Price: denotes the price of the app

Content Rating: denotes the audience of the content

Genres: denotes genre of the app

Last Updated: last date when the app was updated

Current Ver: current version of the app

Android Ver: denotes to which android versions the app is compatible with

Googleplaystore_user_reviews dataset:

App: name of the app

Translated_Review: exact review of the app given by user

Sentiment: denotes whether review is positive , negative or neutral

Sentiment_Polarity: denotes polarity of the sentiment

Sentiment_Subjectivity: denotes subjectivity of the sentiment

Data Preprocessing

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

Some columns contain null values and these are handled by some method. Columns which contain null values are Android Ver, Current Ver, Type, Rating etc. We have deleted columns for which null values percentage was very less and handled those for whom null values percentage held significance. We found that there were some outliers in Rating column.

App column also had some duplicate values which were dropped from googleplaystore dataset after careful consideration.

Also, we modified Reviews, Size, Installs, Price columns and changed their datatypes. Datatype of Last Updated was changed to datetime datatype.

For googleplaystore_user_reviews dataset, we dropped rows wherever Translated_Reviews column was empty.

4. Architecture

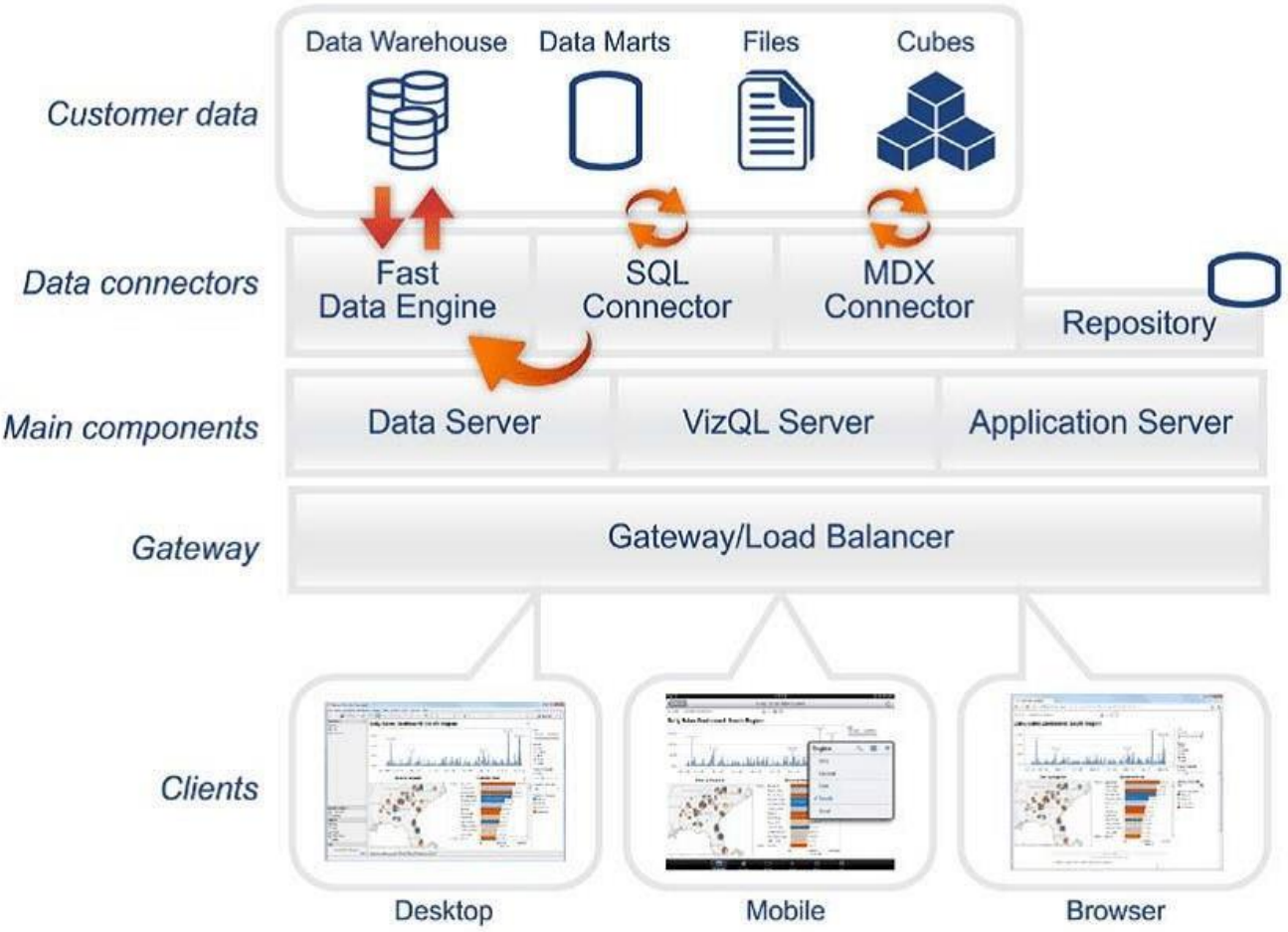
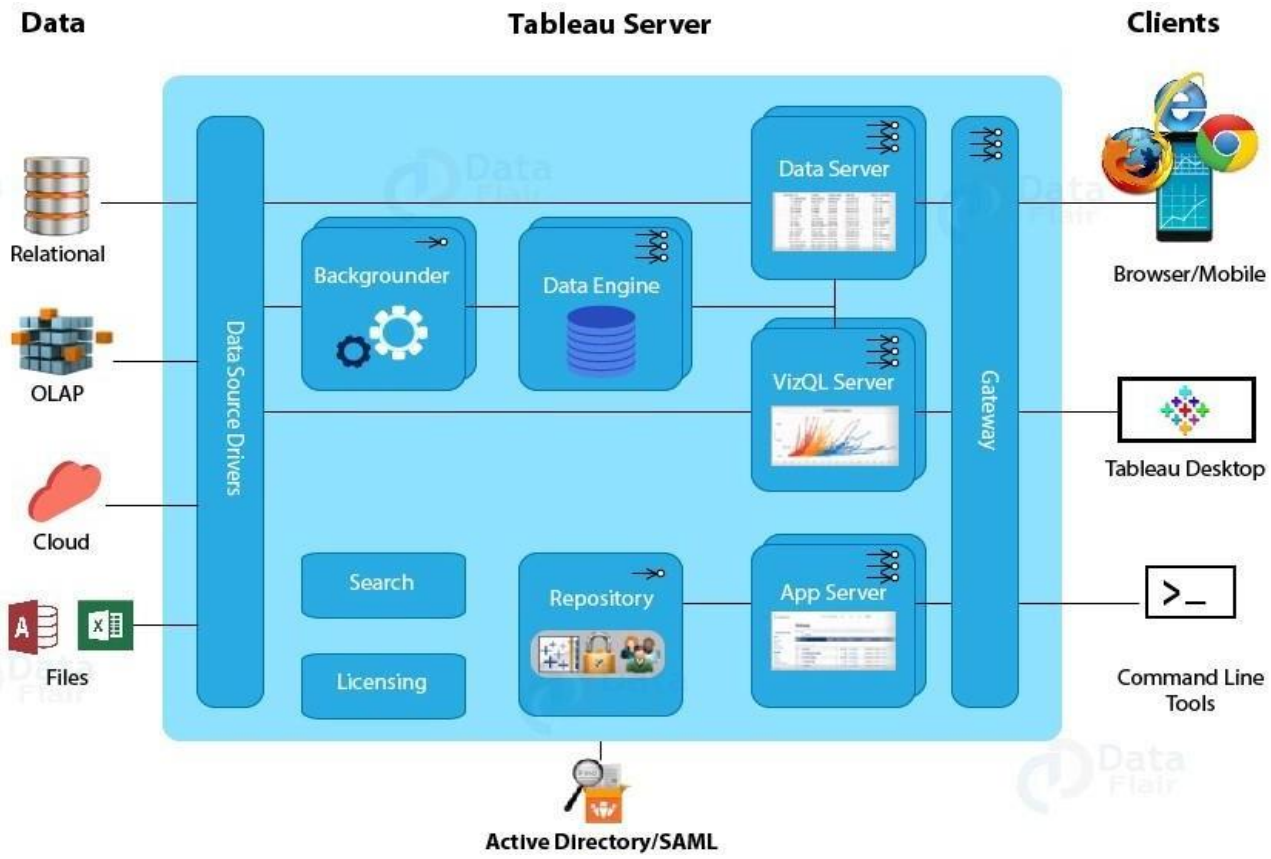


Tableau Server Architecture

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 withthe data on Tableau platforms.



5. Architecture Description

1. Raw Data Collection

The Dataset was taken from iNeuron's Provided Project Description Document and is available in my github repository.

2. Data Pre-Processing

Before building any model, it is crucial to perform data pre-processing to feedthe 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 hypothesesand 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) High-Level Design Document (HLD)
b) Low Level Design Document (LLD)
c) Architecture
d) Wireframe
e) Detailed Project Report

5. 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 forthe 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.

6. Deployment

We created a Tableau Dashboard