Low Level Design (LLD)

Analysing Google Apps Store dataset

Presented By:

Ishita Shetty

Document Version Control

Date Issued	Version	Description	Author
10th June 2022	1.0	Introduction, Problem Statement	Ishita Shetty
12nd June 2022	1.1	Dataset Information, Architecture Deployment	Ishita Shetty
22nd June 2022	1.2	Final Revision	Ishita Shetty

Contents

Abstract

1 Introduction

1.1 Why this High-Level Design Document?

The goal of the LDD or Low-level design document (LLDD) is to give the internal logic design of the actual program code for the House Price Prediction 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 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.

1.3 Project Introduction

In today's scenario we can see that mobile app plays an important role in any individual's life. It has been seen that the development of the mobile application advertise has an incredible effect on advanced innovation. Having said that, with the consistently developing versatile application showcase there is additionally an eminent ascent of portable application designers inevitably bringing about high as can be income by the worldwide portable application industry. Developers and users play key roles in determining the impact that market interactions have on future technology. However, the lack of a clear understanding of the inner working and dynamic of popular app markets impacts both the developers and users. With enormous challenge from everywhere throughout the globe, it is basic for a designer to realize that he is continuing in the right heading.

The Google Play Store is observed to be the biggest application platform. It has been seen that in spite of the fact that it creates more than two-fold the downloads than the Apple App Store yet makes just a large portion of the cash contrasted with the App Store. In this way, I've scratched information from the Play Store to direct my examination on it. Actionable insights can be drawn for developers to work on and capture the Android market!

2 Problem Statement

Technology is the increasing need nowadays and used everywhere. One of the features of Technology is android. Which we all use in our daily life. Android is a mobile operating system based on a modified version of the Linux kernel and other open source software, designed primarily for touchscreen mobile devices such as smartphones and tablets. It is difficult to predict the growth of an app and know the state of the industry

There should be a way which will help app developers predict their number of installs or investors who want to pick out the next big app. Companies may run beta focus groups, or app developers may receive feedback from testers and get certain amounts of reviews. We will use this and some knowledge about the app to predict its success. Knowing the number of installs can help developers and business managers because they can predict the profit. This project's result may show the importance of reviews to apps in the market as it could be one of the determining factors for the number of installs.

3 Data Information

We have two datasets from Kaggle for app reviews; one is the list of apps with information. It has information such as app name, category, rating, and more. And the other is a list of reviews for each app with the sentiment if that particular content of the review was positive, neutral, or negative. Unfortunately, we could not directly use these two files as they are not joined.

App	Category	Ratin	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver
Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19M	10,000+	Free	0	Everyone	Art & Design	07-Jan-18	1.0.0	4.0.3 and up
Coloring book moana	ART_AND_DESIGN	3.9	967	14M	500,000+	Free	0	Everyone	Art & Design;Pretend Play	15-Jan-18	2.0.0	4.0.3 and up
U Launcher Lite 7?FREE Live Cool Themes, Hide Apps	ART_AND_DESIGN	4.7	87510	8.7M	5,000,000+	Free	0	Everyone	Art & Design	01-Aug-18	1.2.4	4.0.3 and up
Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25M	50,000,000+	Free	0	Teen	Art & Design	08-Jun-18	Varies with device	4.2 and up
Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2.8M	100,000+	Free	0	Everyone	Art & Design;Creativity	20-Jun-18	1.1	4.4 and up

App	Translated_Review	Sentiment	Sentiment_Polarity	Sentiment_Subjectivity
10 Best Foods for You	I like eat delicious food. That's I'm cooking foc	Positive	1	0.533333333
10 Best Foods for You	This help eating healthy exercise regular basis	Positive	0.25	0.288461538
10 Best Foods for You	nan	nan	nan	nan
10 Best Foods for You	Works great especially going grocery store	Positive	0.4	0.875
10 Best Foods for You	Best idea us	Positive	1	0.3

Figure 1. Snapshots of the Dataset

Our business case is to locate the best Apps, which we measure by Review check. There are 13 includes that depict each application and an aggregate of 10841 applications. Following variables were initially included:

Variable	Significance
App	Name of the App
Category	Category of the app
Rating	Over all user rating of the app out of 5 on the Play Store
Reviews	Number of user reviews for the app
Size	Size of app
Installs	Number of user downloads/installs for the app
Type	Paid or Free
Price	Cost of the App
Content Rating	Age group the app is targeted at
Genres	An app can belong to multiple genres (apart from its main category)
Last updated	Date when the app was last updated on Play Store
Current Ver	Current version of the app available on Play Store
Android Ver	Minimum required Android Version

4 Architecture

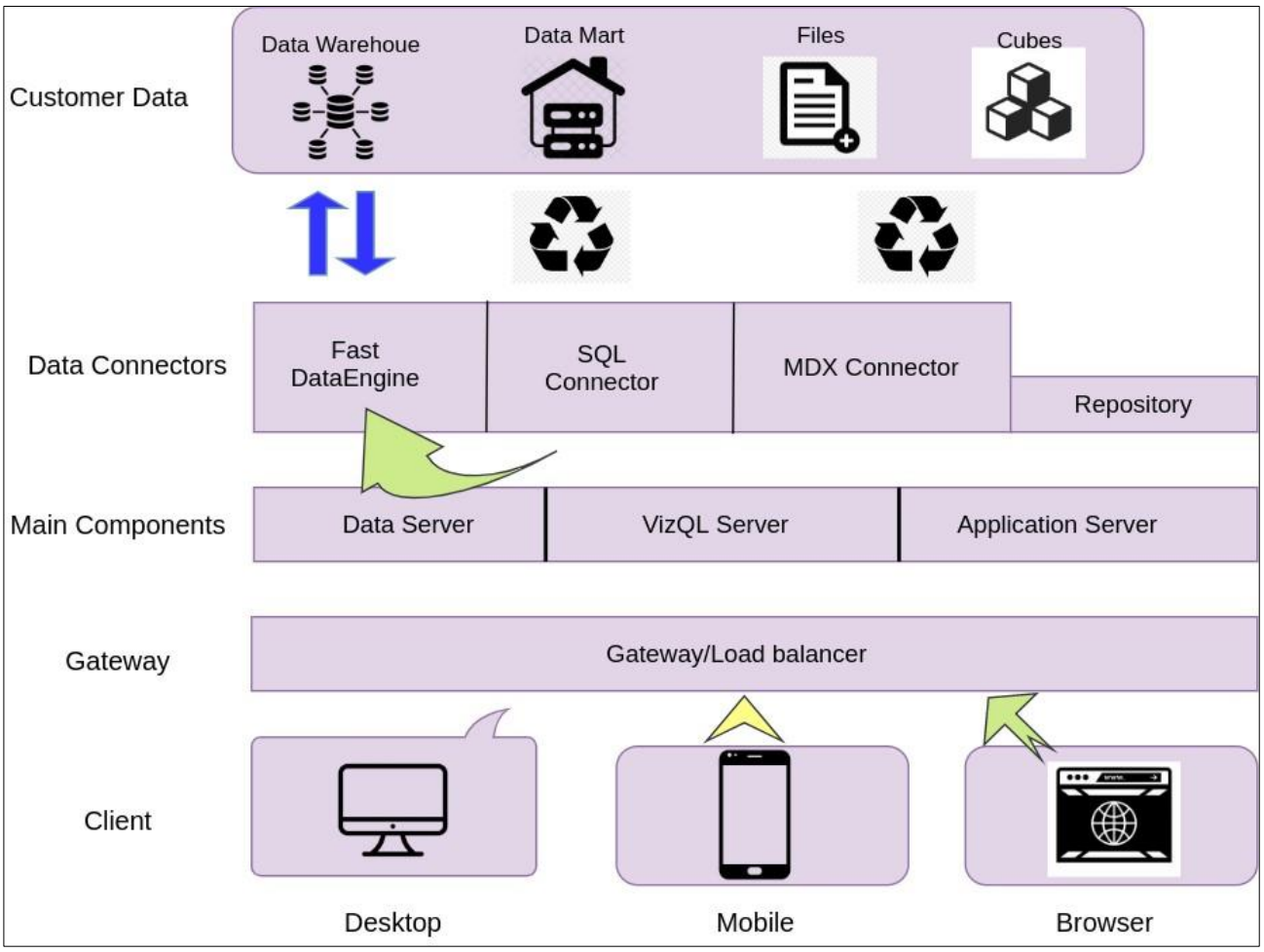

Figure 2. Working of BI

4.1 Working

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 feeded 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 attributes

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 hypothesis and to 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 easy and self-explanatory report because your model will be used by many stakeholders
who are not from technical background.

a) High Level Design Document (HLD)
b) Low Level Design Document (LLD)
c) Architecture
d) Wireframe
e) Detailed Project Report
f) Power Point Presentation

6. Modelling

Data Modelling is the process of analysing the data objects and their relationship to the other objects. It is used to analyse 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 Dashboard on Tableau

4.2 Tableau Server Architecture

Tableau has a highly scalable, n-tier client-server architecture that serves mobile clients, web clients and desktop-installed software. Tableau Server architecture supports fast and flexible deployments.

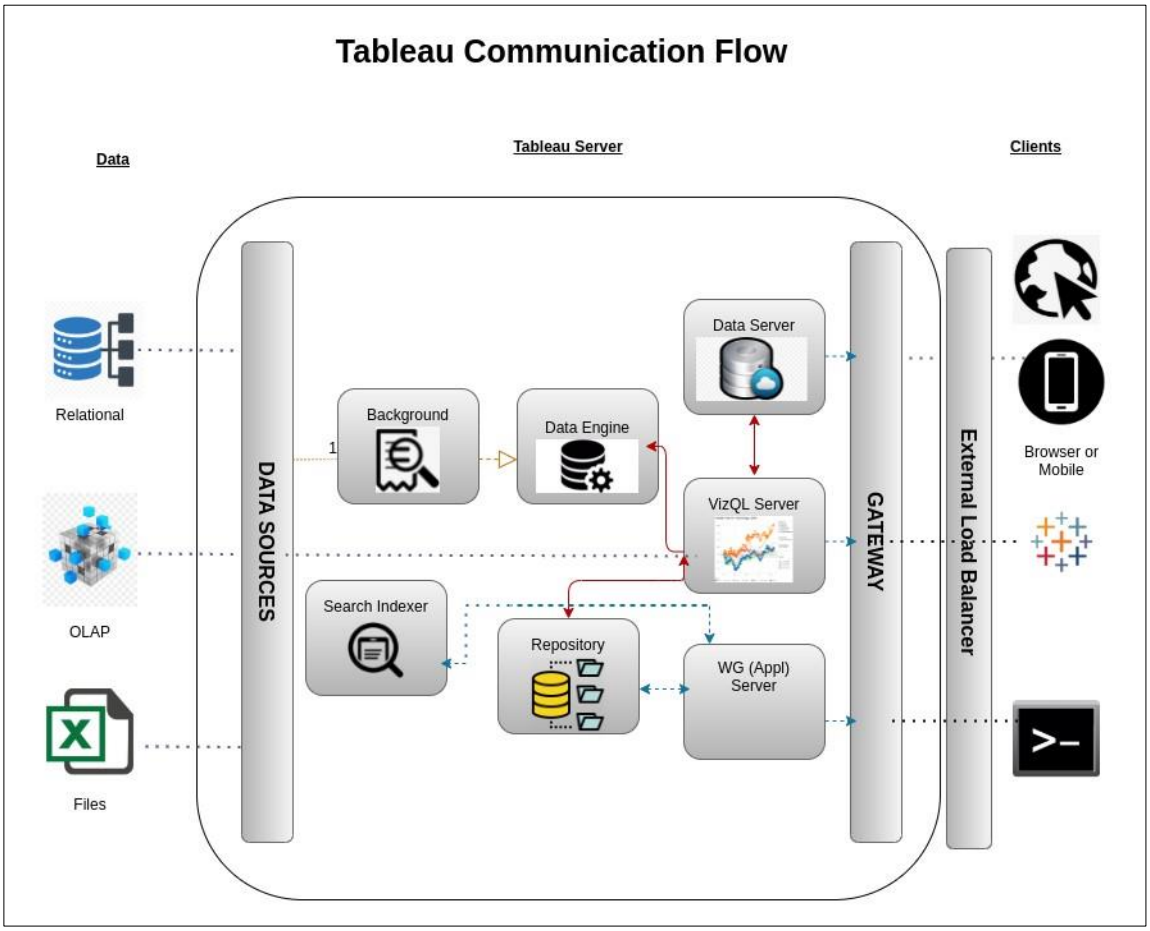

Figure 2. Architecture of Tableau Server

Tableau Server is internally managed by the multiple server processes.

1. **Gateway/Load Balancer:**

It acts as an Entry gate to the Tableau Server and also balances the load to the Server if multiple Processes are configured.

2. **Application Server:**

Application Server processes (wgserver.exe) handle browsing and permissions for the Tableau Server web and mobile interfaces. When a user opens a view in a client device, that user starts a session on Tableau Server. This means that an Application Server thread starts and checks the permissions for that user and that view.

3. **Repository:**

Tableau Server Repository is a PostgreSQL database that stores server data. This data includes information about Tableau Server users, groups and group assignments, permissions, projects, data sources, and extract metadata and refresh information.

4. **Data Engine:**

It stores data extracts and answers queries

5. **Backgrounder:**

The backgrounder Executes server tasks which includes refreshes scheduled extracts, tasks initiated from tabcmd and manages other background tasks.

6. **Data Server:**

Data Server Manages connections to Tableau Server data sources It also maintains metadata from Tableau Desktop, such as calculations, definitions, and groups.

5 Deployment

Prioritizing data and analytics couldn't come at a better time. A company, no matter what size, is already collecting data and most likely analysing just a portion of it to solve business 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 and experts alike to author and consume content.

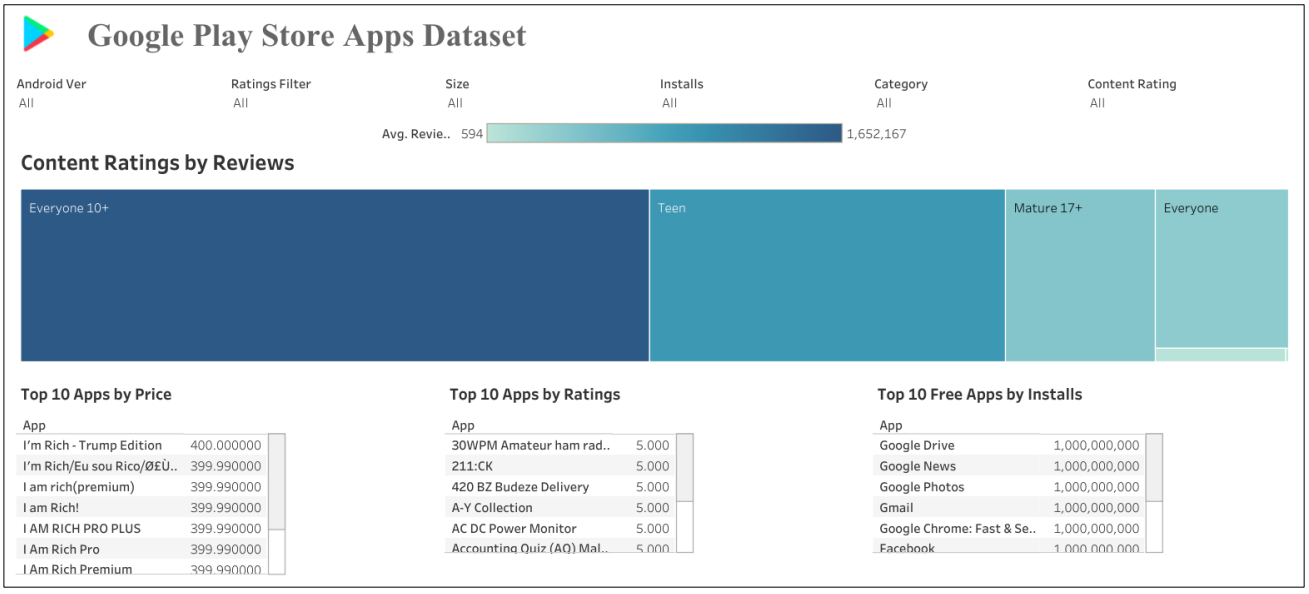

Tableau prioritizes choice in flexibility to fit, rather than dictate, your enterprise architecture. Tableau Server and Tableau Online leverage your existing technology investments and integrate into your IT infrastructure to provide a self-service, modern analytics platform for your users. With on-premises, cloud, and hosted options, there is a version of Tableau to match one's requirements.

6 Unit Test Cases

TEST CASE DESCRIPTION	EXPECTED RESULTS
Android Ver Parameter	When clicked on the slicer, a dropdown should occur which has various versions available.
Ratings Filter	When clicked on the slicer, a dropdown should occur which displays different ratings.
Size	When clicked on the slicer, a dropdown should occur which displays different size of applications.
Installs	When clicked on the slicer, a dropdown should occur which displays different number of installs.
Category	When clicked on the slicer, a dropdown should occur which displays different categories.
Top 10 Apps by Price	Displays the top 10 applications wrt price by the selected filters.
Top 10 Apps by Ratings	Displays the top 10 applications wrt ratings by the selected filters.
Top 10 Free Apps by Installs	Displays the top 10 applications wrt no. of installs by the selected filters.