High Level Design (HLD)

Google Play Store Analysis

Document Version Control

Date Issued	Version	Description	Author
06/08/2022	1.0	Abstract, Introduction, General Description	Anil S Adiga

07/08/2022	1.1	Design Detail, KPI, Deployment	Anil S Adiga
08/08/2022	1.2	Final Revision	Anil S Adiga

Contents

Abstract

Google Play is a digital distribution service operated and developed by Google Inc. It serves as the official app store for the Android operating system, allowing users to browse and download applications developed with the Android SDK and published through Google. Google Play also serves as a digital media store, offering music, books, movies, and television programs.

Android is expanding as an operating system. It has captured around 74% of the total market which is a true indicator of the huge amount of population using android. Our goal is to help android developers to know what is the motivating factor for people to download an app. It will also help to find out the factors that affect someone's decision to download an app. I would like to analyze category, reviews, price, ratings and installs for this purpose and find out how they are inter related.

1 Introduction

1.1 Why this High-Level Design Document?

The purpose of this High-Level Design (HLD) Document is to add the necessary detail to the 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

1.2 Scope

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

2 General Description

2.1 Product Perspective & Problem Statement

The goal of this project is to analyze the apps in the google app store and find the relationship between different features such as Category, Reviews, Ratings and Installs of apps and how these factors/features have changed over a period of time and gain insights from the data to know the most popular apps in the play store.

2.2 Tools used

Business Intelligence tools and libraries works such as NumPy, Pandas, Seaborn, Matplotlib, MS-Excel, MS-Power BI, Jupyter Notebook and Python Programming Language are used to build the whole framework.

3 Design Details

3.1 Functional Architecture

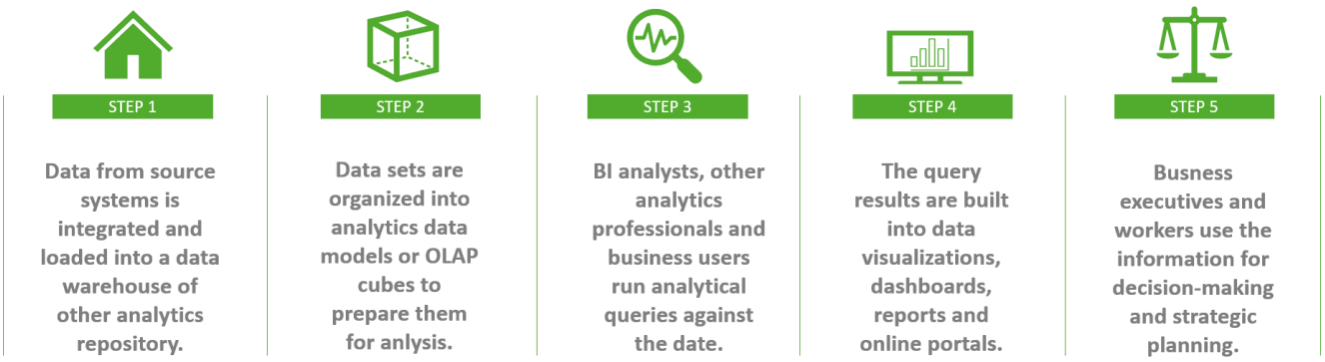
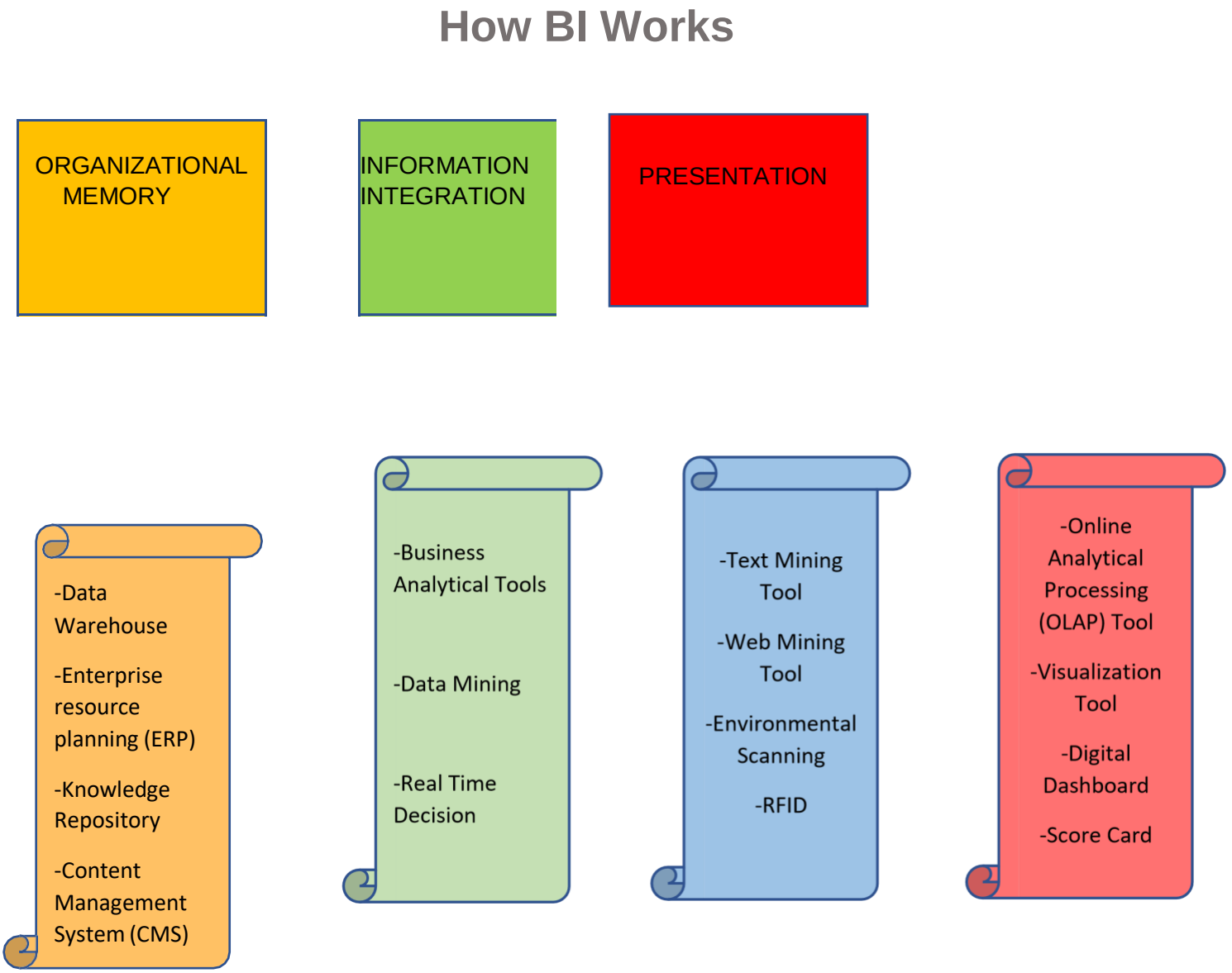


Figure 1: Functional Architecture of Business Intelligence



3.2 Optimization

1. 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

2. 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.

3. 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 dimensionwhile 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).

4. Optimize and materialize your calculations

- Perform calculations in the database • Reduce the number of nested calculations.
- Reduce the granularity of DAX or table calculations in the view. The moregranular the calculation, the longer it takes.

  DAXs - Look at the number of unique dimension members in the calculation.

  Table Calculations - the more marks in the view, the longer it will take tocalculate.

- Where possible, use MIN or MAX instead of AVG. AVG requires more processing than MIN or MAX. Often rows will be duplicated and display thesame 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 KPI

Dashboards will be implemented to display and indicate certain KPIs and relevant indicators 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 factor

4.1 KPIs (Key Performance Indicators)

Key indicators displaying a summary of the google play store and its relationship with different metrics are:

1. Average Rating for all apps of different categories
2. Average Reviews for all apps of different categories and genre
3. Total Apps present in the app store according to the data.
4. Total Apps with ratings of 5 irrespective of any category or genre.
5. Average installs of apps by users over a period of time.

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 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 Power BI at scale, as well as organizing, orchestrating, and unifying disparate sources of data for business users and experts alike to author and consume content.

Power BI prioritizes choice in flexibility to fit, rather than dictate, your enterprise architecture. Power BI Desktop and Power BI Service leverage your existing technology investments and integrate them 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 Power BI to match your requirements