Architecture Design

Analyzing Google Apps Store dataset in terms of App downloads and Rating

Written By	ANWESHA DAS
Document Version	0.1
Last Revised Date	08-07-2022

DOCUMENT CONTROL

Change Record:

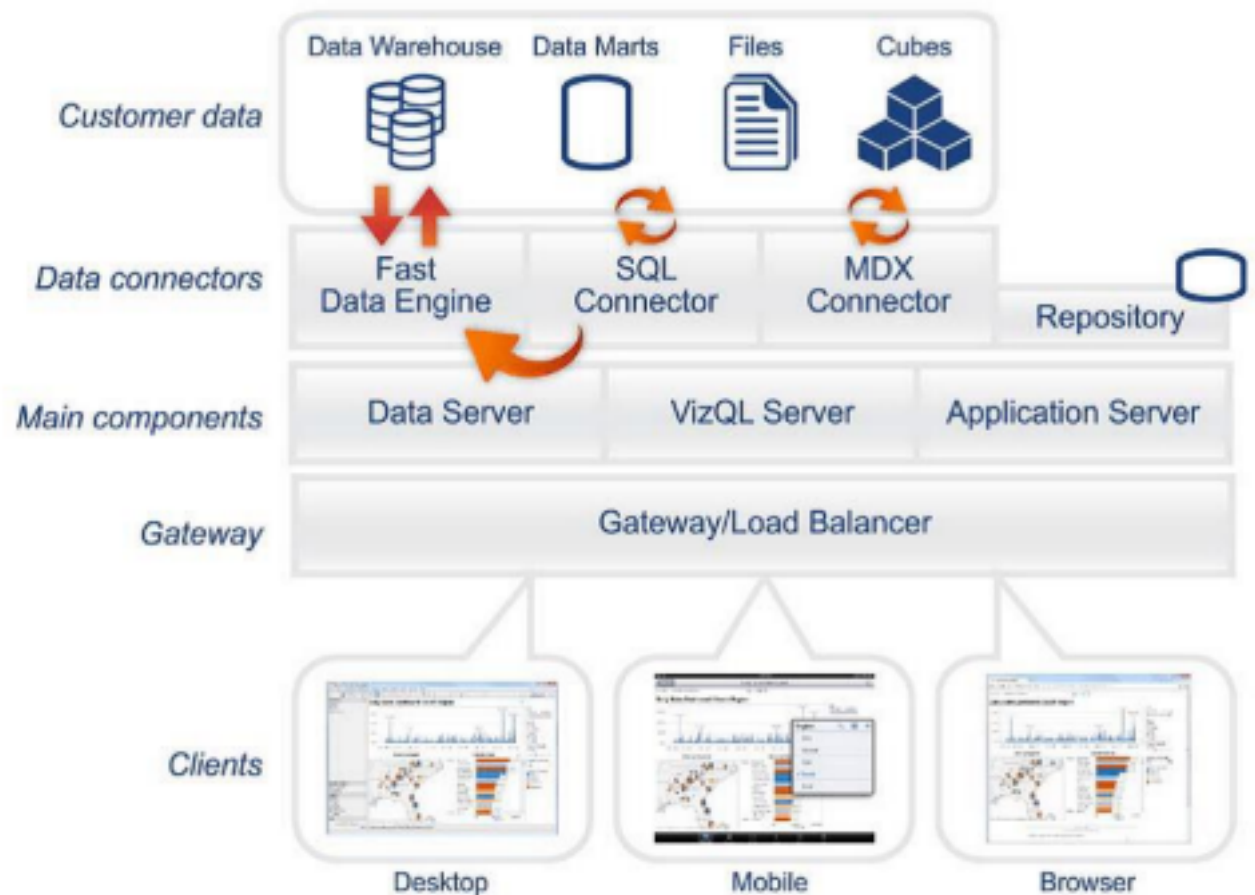| VERSION	DATE	AUTHOR	COMMENTS
0.1	8TH JULY - 2022	ANWESHA DAS	Introduction and architecture defined
			Architecture & Architecture description appended and updated.

Contents

1. Introduction

1.1 What is an Architecture design document?

Any software needs the architectural design to represent the design of software. IEEE defines architectural design as "the process of defining a collection of hardware and software components and their interfaces to establish the framework for the development of a computer system." The software that is built for computer-based systems can exhibit one of these many architectures.

Each style will describe a system category that consists of :

- A set of components (eg: a database, computational modules) that will perform a function required by the system.

- The set of connectors will help in coordination, communication, and cooperation between the components.

- Conditions that how components can be integrated to form the system.

- Semantic models that help the designer to understand the overall properties of the system.

1.2 Scope

Architecture Design Document (ADD) is an architecture 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 design principles may be defined during requirement analysis and then refined during architectural design work.

ARCHITECTURE DESIGN

2. Architecture

2.1 RAW DATA COLLECTION:: The dataset was taken from ineuron's provided project description document .

https://drive.google.com/drive/folders/1FkmFVL8wlJmQWP1z52TD8Plh OJhitTyI

2.2 DATA PREPROCESSING:: Before building any kind of model it is crucial to perform preprocessing to feed the correct data to the model.Model performance depends on the quality of data feeded to the model to train.

The process includes :

1.Handling null values.

2.Handling skewed Data

3.outlier detection and removal

2.3 Data cleaning :: Data cleaning is the process of fixing or removing incorrect , corrupted ,incorrectly formatted ,duplicate ,incomplete data Within data set .

 a. Remove duplicate or irrelevant observation
 b. Filter unwanted outlier
 c. Renaming required attribute

2.4 Exploratory data analysis::: EDA refers to the critical process of performing investigation on data to discover patterns , spot anomalies , test hypotheses and to check assumptions with the help of summary statistics and graphical representation .

2.5 Reporting :: Reporting is a most important and underrated skill of data analytics field.because being a Data Analyst should be good in easy and self explanatory report because model will be used by many stakeholders who are non technical background.

1.High level documentation(HLD)

2.Low level Documentation(LLD)

3.Architecture

4.Wireframe

5.Detailed Project Report

6.Power point presentation

2.6 Modeling:: Data Modeling is the process of analyzing data objects and their relationship to their object.It is used to analyze the data requirements that are required for business processes.The data models are created for data to be stored in the database.The data model 's main aim what data is needed and how we have to organize those data rather than what operation we have to perform.