Low Level Design

Airbnb Data Analysis

Written By	Mohammad Tippu Sohail, Mohammad Salman
Document Version	1
Last Revised Date	

Document Version Control

Change Record:

VERSION	DATE	AUTHORS	COMMENTS
1.0	Nov 25, 2023	Mohammad Tippu Sohail & Mohammad Salman	Prepared Low Level Document for Airbnb Data Analysis

Reviews:

VERSION	DATE	REVIEWER	COMMENTS

Approval Status:

VERSION	REVIEW DATE	REVIEWED BY		APPROVED BY	COMMENTS

Contents

1.Introduction

What is Low-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.

2. Architecture

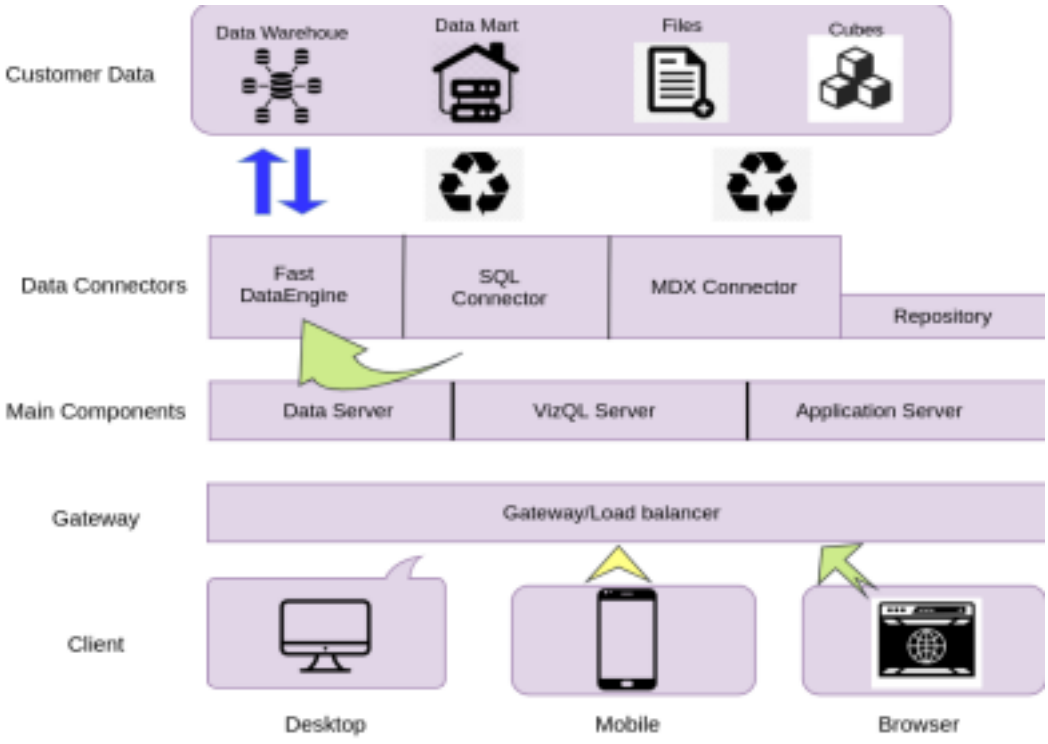


Tableau Server Architecture

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.

The following diagram shows Tableau Server's architecture:

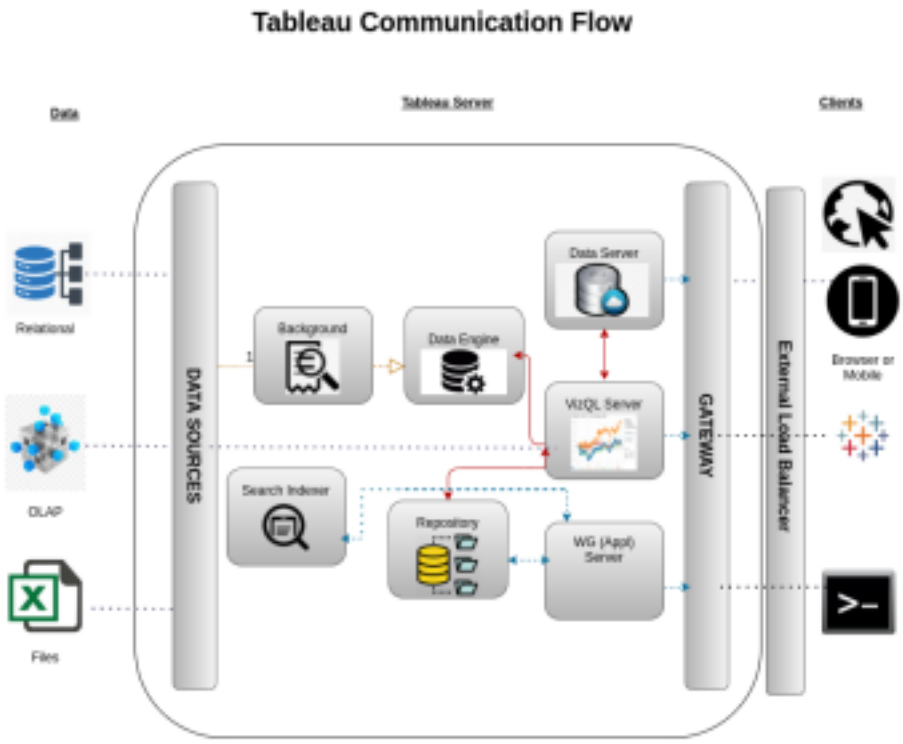


Tableau Server is internally managed by the multiple server processes.

1. Gateway/Load Balancer

It acts as an Entry gate to the Tableu 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) VIZQL Server:

Once a view is opened, the client sends a request to the VIZQL process (vizqlserver.exe). The VIZQL process then sends queries directly to the data source, returning a result set that is rendered as images and presented to the user. Each VIZQL Server has its own cache that can be shared across

multiple user

5) Data Engine:

It Stores data extracts and answers queries.

6) Backgrounder:

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

7) Data Server:

Data Server Manages connections to Tableau Server data sources

It also maintains metadata from Tableau Desktop, such as calculations, definitions, and groups.

3. Architecture Description

3.1 Data Description:

This data file includes all needed information to find out more about hosts, geographical availability, necessary metrics to make predictions and draw conclusions.

- **Host Information:**
  - Host ID: Unique identifier (ID) for each host.

- **Geographical Availability:**
  - Listing Location: Detailed location of each Airbnb property.
  - Neighborhood: Specific neighborhood or district where the property is located.
  - City: City or town where the property is situated.
  - Country: Country in which the property is located.
  - Latitude and Longitude: Geographical coordinates of the property.

- **Necessary Metrics:**
  - Reviews: Number of reviews for each property.
  - Rating: Average rating given by guests for each property.

3.2 Raw Data Collection:

The Dataset was taken from iNeuron's Provided Project Description Document.

https://drive.google.com/drive/folders/1ANkgtAT0Pdp2r86IxFKv9vKYmnsYjJDO?usp=s

3.3. 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

- Handling Null/Missing Values
- Handling Skewed Data
- Outliers Detection and Removal

3.4. Data Cleaning

Data cleaning is the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset.

- Remove duplicate or irrelevant observations
- Filter unwanted outliers
- Renaming required attributes

3.5 Exploratory Data Analysis (EDA)

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

3.6 Reporting

Reporting is a most important and underrated skill in the data analytics field. Because being a Data Analyst you should be good with the easy and self-explanatory reports because your model will be used by many stakeholders who are not from a technical background.

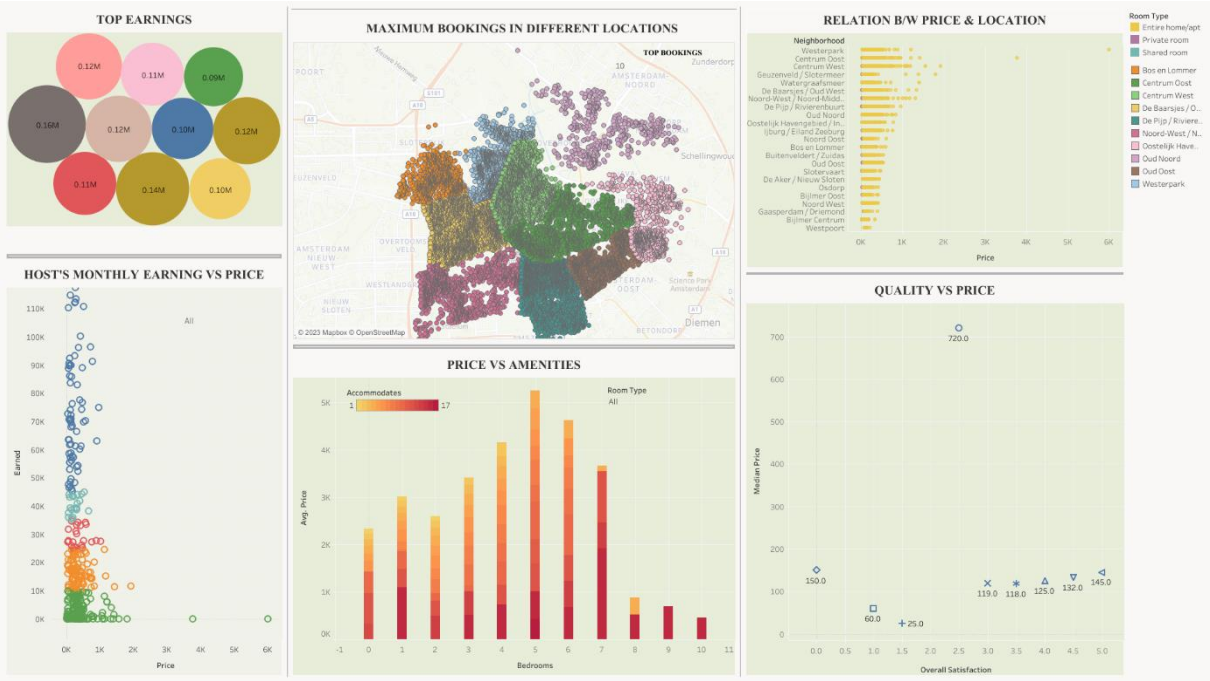
- High-Level Design Document (HLD)
- Low-Level Design Document (LLD)
- Architecture
- Wireframe
- Detailed Project Report
- PowerPoint Presentation

3.7 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 business processes. The data models are created to store the data 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.

3.8 Deployment:

We created a Tableau Dashboard



4. Unit Test Cases

TEST CASE DESCRIPTION	EXPECTED RESULTS
Top Earners Analysis	A visual representation highlighting the top earners based on monthly earnings.
Relationship Between Monthly Earning and Prices	A scatter plot depicting the relationship between monthly earnings and prices.
Maximum Number of Bookings by Location	Geographical visualization showcasing the particular location (neighborhood) with the maximum number of bookings.
Price Relationship with Location	A chart showing the price distribution across different locations.
Relationship Between Quality and Price	A scatter plot illustrating the relationship between customer reviews' quality(overall satisfaction) and median prices.
Price vs Amenities	Bar chart depicting the correlation between prices and the availability of amenities.