**Low-Level Design (HLD)**

**Travel Data Analysis(Air BNB Case Studies)**

Last Revised Date: 31/08/2021

**Document Version Control**

|  |  |  |  |
| --- | --- | --- | --- |
| Version | Date | Author | Comments |
| 0.1 | 25-Aug-2021 | Samrudhi Deshmukh  Manish Kumar  Subhabrat Nath | Abstract,  Scope of HLD |
| 0.2 | 31-Aug-2021 | --- | --- |
|  |  |  |  |

**Reviews:**

|  |  |  |  |
| --- | --- | --- | --- |
| Version | Date | Reviewer | Comments |
|  |  |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Version | Review Date | Reviewed By | Approved By | Comments |
|  |  |  |  |  |

**Approval Status:**

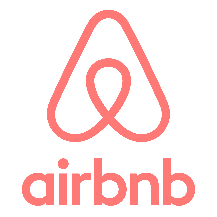
**Index**

|  |  |
| --- | --- |
| **Content** | **Page No** |
| Abstract | 3 |
| 1. Introduction | 3 |
| 1.1 Scope of HLD | 3 |
| 2. Architecture | 4 |
| 3. Architecture Description | 4 |
| 3.1 Database | 4 |
| 3.2 Database Connector | 5 |
| 3.3 Main Component | 5 |
| 3.4 Gateway/Load Balancer | 5 |
| 4. Dataset Description | 5 |
| 4.1 Sample Data | 5 |
| 4.2 Dataset Description | 5 |
| 4.3 Data Transformation/Manipulation | 6 |
| 4.4 Data Insertion into Database | 6 |
| 5. Tableau Server Components and Architecture | 7 |
| 5.1 Gateway | 7 |
| 5.2 Application Server | 7 |
| 5.3 Repository | 7 |
| 5.4 VizQL Server | 7 |
| 5.5 Data Engine | 8 |
| 5.6 Backgrounder | 8 |
| 5.7 Data Server | 8 |
| 5.8 Search and License | 8 |
| 6 Deployment | 8 |

**Abstract**

Observe the AirBNB travel data and find the relation between different attributes and mark the top host those are making more profit and find the particular places where most of the tourists like to visit. Airbnb is an online platform that allows people to rent short-term accommodation. Airbnb Travel Data is describe the listing activity and metrics in San Diego, California for 2019. Our main goal is to find Research Questions from the given data and make a dashboard that stands with our answers.

The major fields are **host id, host\_name, price, numbers\_of\_reviews, calculated\_host\_listing\_coun**t. Along with many columns with multiple tables. The goal is to create a relationship between tables and find the answers to research questions.



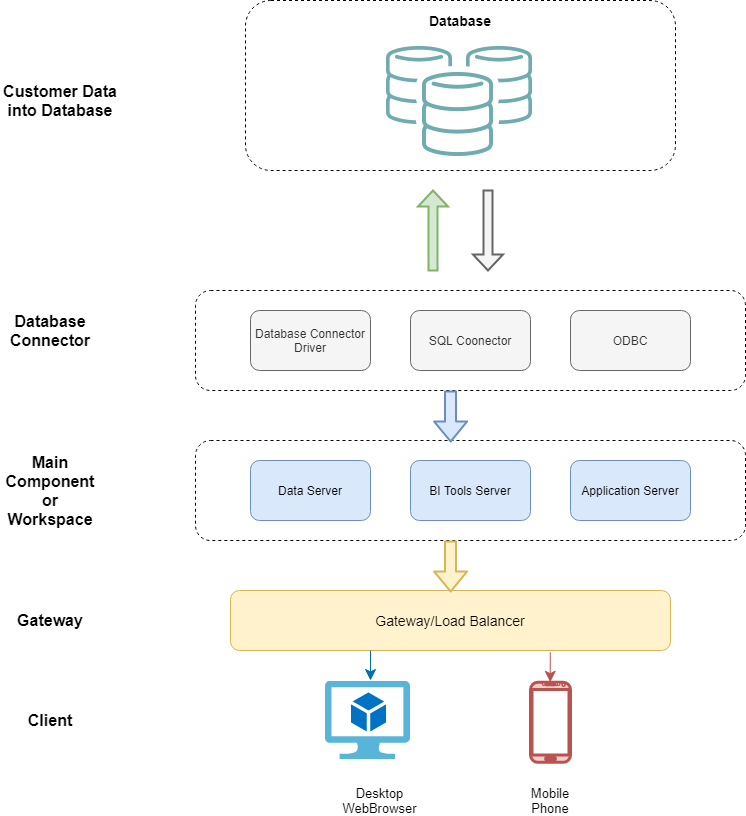
**1. Introduction**

**1.1 Scope of the LLD:**

The LLD system aims to present the internal functionality of the implemented system. The LLD documents will show the Dataset Description. Complete Architecture Description and working methodology of every component and tools are used to analyze AirBNB Travel Data Analysis.

The LLD is the component-level design process that follows a step-by-step refinement process. The process can be used for design 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**



**3. Architecture Description**

**3.1 Database**

In the database, we will store our data and will extract it as per need from the cloud through the database connector.

**3.2 Database Connector**

Database connecter is software or it can be assumed as a driver that creates a connection between cloud database to BI tool where we want to pull the data. It allows the BI tool to extract data directly from the database if the authentication is provided. Then the BI tool has access to the data and we can do analysis on the data.

**3.3 Main Component/Workspace**

This is the main workspace where the data will be extracted and all the major work will be done here. It can be SQL server or Online query editor or data analysis platform also can be BI tools like Power BI in this case. Here we will store the data in memory and perform the analysis.

**3.4 Gateway/Load Balancer**

The cloud platform where the final application will be hosted. We will use Power BI Service for hosting our data.

**4. Dataset Description**

Total 4 datasets are given to perform the case studies and every dataset consists of information from a different aspect. The datasets are **`airbnb price`, `calender`, `listings`, `neighborhoods, `reviews`.**

**4.1 Sample Data**

*  Below is a sample of **`airbnb price`** data.

A Portion of data has been shown The dataset consist of **18724 rows** and **20 columns**. The dataset consists of information of different host id and their price for different hotel rooms along with location information.

* Below is a sample of the `**calender**` dataset.



The `**calendar**` dataset consists of information about `listing\_id`, `date` and it is available or not along with the price. The dataset has **4 columns** and **104,8576 rows**.

* Below is a sample of the `**listings**` dataset.

The `**listings**` dataset is our fact table, most of the answers we have to find here, the data contain **97 columns** and **11923 rows.** Here we can find the hostname, their id, and the listings count with the price of respective listings. Also, we can find location information and booking count.

* Below is a sample of the ` **neighborhoods`**  dataset.



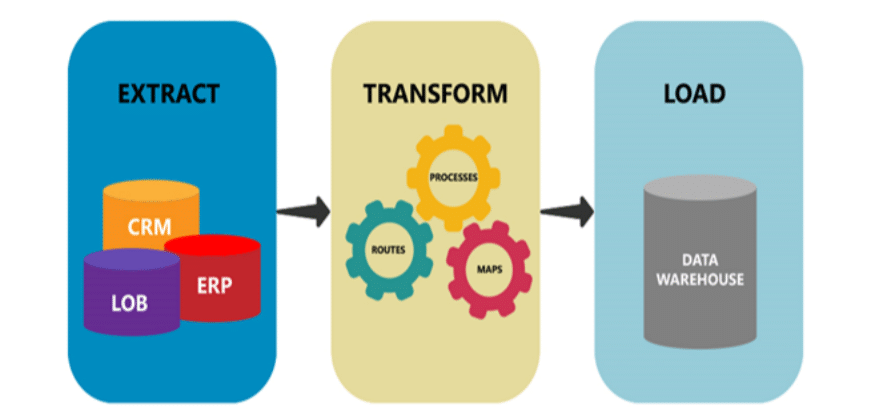
This dataset contains only information about neighborhoods names. Only contain **2 columns** and **109 rows**.

* Below is a sample of the `**reviews**` dataset.

This dataset contains customer reviews or comments regarding their respective host about the host service during the contract time. We can observe the date and host and his\her customer id along with the. In total **6 columns** and **344,405 rows** are in the dataset.

**4.3 Data Transformation/Manipulation**

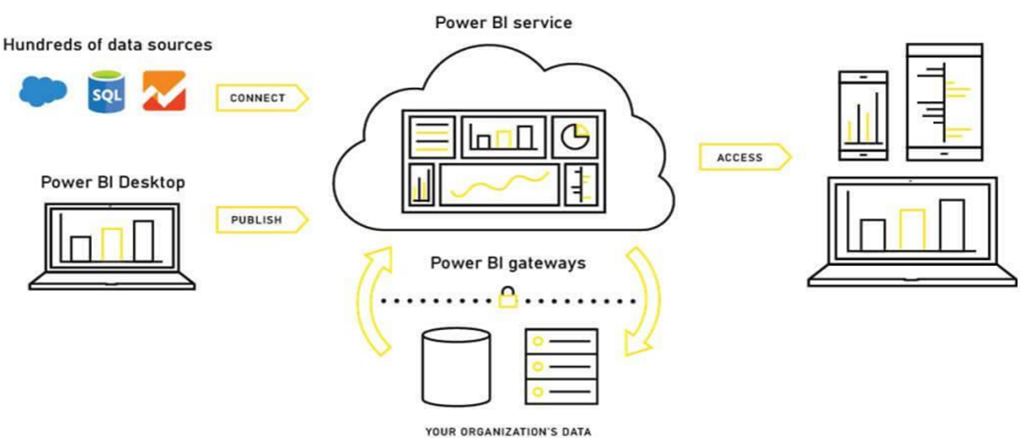
The dataset provided is not suitable to start analysis, Data cleaning and preprocessing are required for further analysis. We will use the **python programming** language and **Power BI** inbuild **ETL** tool also known as **Power Query** for preprocessing the data. We will create an entity diagram for creating a relationship between.

****

**4.4 Data Insertion into Database**

We will use Datastax Astra Database for our cloud data storage. We will export our data into a cloud database and will export it into the local system or directly can connect it to the BI tool we will use to Visualize the data for further analysis. The database is a flexible cloud storage solution and can access data from anywhere if having credentials for access. We will import data from the cloud to our BI tool for creating Dashboard and Analysis. The following architecture will describe the relation between clod and Client application for the flow of data.

**5. Power BI Server Components and Architecture:-**



**5.1.Gateway-**

Gateway is a kind of web server that helps clients communicate to the server via HTTP or HTTPS. The server receives incoming client requests and directs them to the appropriate server for action. A gateway handles processes such as load balancing, traffic routing, URL rewriting, serving static files to clients, serving multi-thread processes, etc.

**5.2.Application Server-**

The application server deals with login processes, domain authentication, data authorization, user or group permission management, content searches, etc. It works in close association with the server’s repository and handles the data access operations.

**5.3.Repository-**

The repository in Power BI Server stores server metadata related to users, permissions, assignments, groups, and projects. Along with the metadata, it stores visualizations in flat files (TWS, TDS), and performance data for auditing.

**5.4 VizQL Server-** It loads visualizations to work on Tableau. It has in-built caching for performance improvement and editing tasks. Whenever a user requests a visualization or wishes to update an existing one, the request received by VizQL is first converted into an SQL statement and sent down to the data sources via respective data source drivers. The requested data sent back from the data source comes to the VizQL server again, where it is processed with some final touches of additional calculations and sent to the user. Any new visualization coming from the data source is cached in the VizQL for further use.

**5.5.Data Engine-**

The data engine stores multiple TDEs (Tableau Data Extract) and can run on multiple servers maximum 2). It also attends multiple requests parallelly at a given point of time. The data engine hosts the piece of data in-memory extracted from the TDE upon getting a request from the user.

**5.6.Backgrounder-**

Backgrounder is a multi-process element that manages schedules for information refreshing and ensures proper functioning of the Tableau Server and Data Engine.

**5.7.Data Server-**

The data server helps in centralizing metadata management, driver deployment, and extract management. It also contributes to access control and serves as a proxy to the data sources. It hosts user queries and requests to prevent users from directly accessing the data source.

**5.8.Search and License**-

Two other important components are search and license. The search component manages the search indexing for the data in the repository. Whereas, the license component is responsible for the licensing and configuration of the Tableau server. Both these services run on the primary server of the Tableau’s server cluster.

**5. Deployment**

Power BI service will be used for deployment. It required an official email Id to create a free account. And publishing dashboard for making it publically available. There is some advantage of choosing this server it is required less maintenance, can be host directly form local machine, and the deployment process is also very easy. Deployment can be done from Power BI Workspace only.

**Benefits of Power BI:**

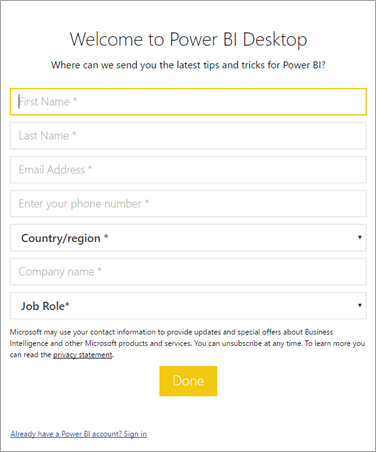
* Easy to use and Free.
* Fast and more reliable.
* No software installation is required for the deployment.
* Easy to share of the globally accessible link.
* The report can be saved locally.

**Drawbacks of Power BI:**

* For Publishing Dashboard required official email id. Regular email id or Gmail Id will not work.
* Computational Expensive software. Required sufficient amount of RAM to run on Local machine.

**5.1 Steps of Deployment**

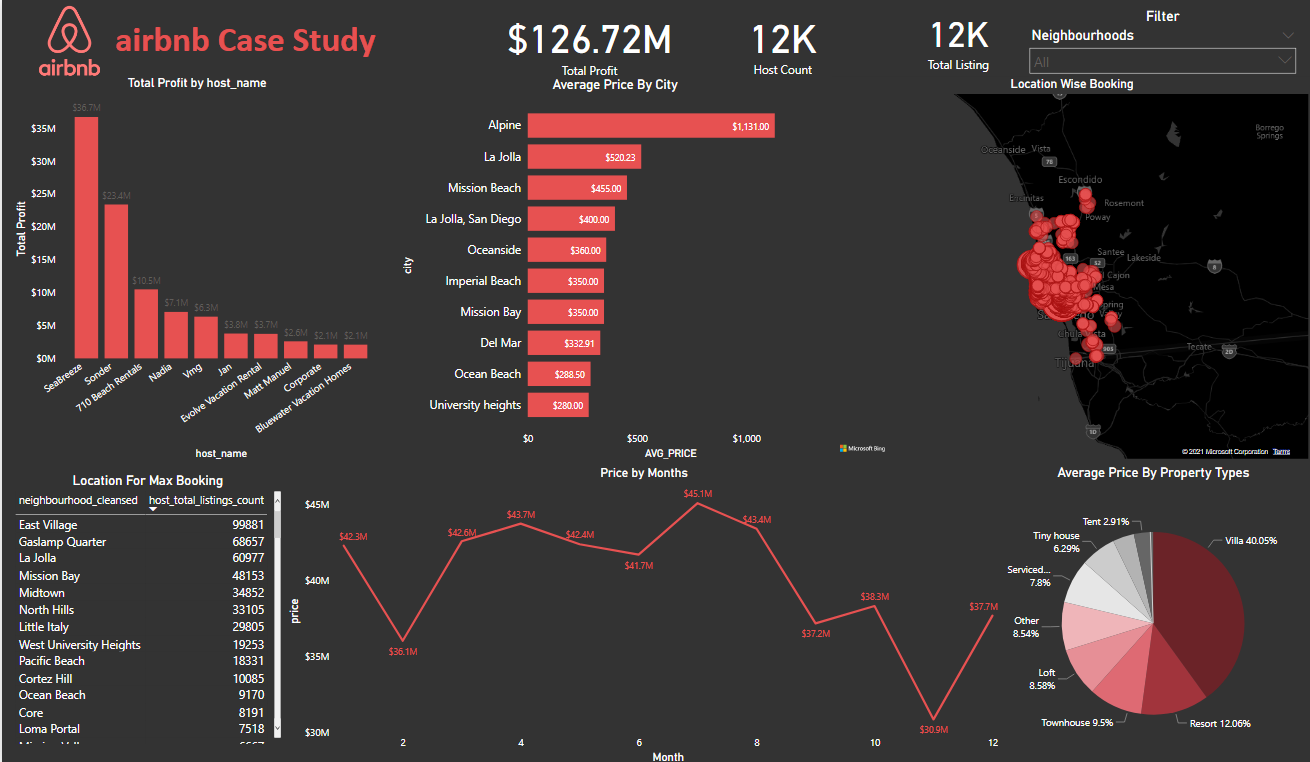
We can deploy our dashboard directly from our workspace, we must have registered on the Power BI service before that, It required an official email id.

**Steps 1 – Login**

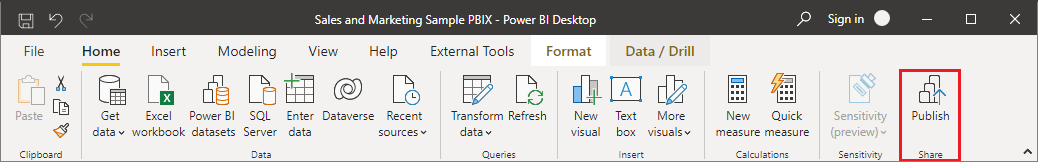
The user has to enter his registered official email address and password for the login.

**Step 2 – Building the Dashboard**

Users need to Build the Dashboard before Deploying on Power BI Services. And have to save the dashboard on the Power BI workspace. We have to select and have to present the desire dashboard or worksheet that we want to deploy.

**Step 3 – Save the Dashboard**

**Step 4 – Publish to Power BI**

In Power BI Desktop, choose File > Publish > Publish to Power BI or select Publish on the ribbon.

Here we have to select the destination. And after selecting the workspace have to click on Publish button. As shown in the above image.

**Steps 5 – Send to Cloud.**

After Selecting the workspace here workspace is our final dashboard. And click on Publish button we will get a confirmation of the successful deployment of our report on the Power BI cloud.

