High Level Design (HLD)

AirBnb Data Analysis

Revision Number: 1.0
Last date of revision: 26/06/2022
Akshay Salve

Document Version Control

Airbnb Data Analysis - Business Intelligence Project

Version	Date	Author	Change
1.0	26/06/2022	Akshay Salve	Final Version of Complete HLD

Abstract

Airbnb, Inc. is an American company that operates an online marketplace for lodging, primarily homestays for vacation rentals, and tourism activities. Based in San Francisco, California, the platform is accessible via website and mobile app. Airbnb does not own any of the listed properties; instead, it profits by receiving commission from each booking. The company was founded in 2008 by Brian Chesky, Nathan Blecharczyk, and Joe Gebbia. Airbnb is a shortened version of its original name, AirBedandBreakfast.com.

This dataset provides huge data on based on the survey conducted in July 2017.Based on this information the ultimate goal is to find the insights highlighting key indicators and metrics that influence the travel market.

Contents:

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 current project description to represent a suitable model for coding. This document is also intended to help detect contradictions prior to coding, and can be used as a reference manual for 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: o Security
o Reliability
o Maintainability
o Portability
o Reusability
o Application compatibility
o Resource utilization
o 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.

1.3. Definitions
- ❖ ETL – Extract, Transform, Load
- ❖ EDA – Exploratory Data Analysis
- ❖ Csv file – Comma separated values file, opened in MS Excel
- ❖ Fair Share – An administratively set data rate per time frame that is considered fair.
- ❖ Postgres SQL Server – A database management system.
- ❖ Power Query – With Power Query, you can connect to many different data sources and transform the data into the shape you want.
- ❖ ER – Entity Relation Diagram
- ❖ Power BI - An interactive data visualization software company focused on business intelligence.

2. General Description

2.1 Product Perspective and Problem Statement

Airbnb plays a vital role in the travelling market. The largest and best-documented potential benefit of Airbnb expansion is the **increased supply of travel accommodations**, which could benefit travellers by making travel more affordable. There is evidence that Airbnb increases the supply of short-term travel accommodations and slightly lower prices.
In this project I have analysed the data collected from the survey conducted in the Amsterdam on date 22 July and 23 July 2017. The objective of the project is to get the insights from the data and perform visualization techniques to understand details about the business.

2.2 Tools Used

2.2.1. MS Excel

Microsoft excel is used for loading the data in csv format, basic data cleaning and filter operations to execute the program, MS excel file was loaded into power bi.

2.2.2. Power Bi

Power BI is a collection of software services, apps, and connector that work together to turn your unrelated sources of data into coherent, visually immersive, and interactive insights. Your data may be an Excel spreadsheet, or a collection of cloud-based and on-premises hybrid data warehouses. Power BI lets you easily connect to your data sources, visualize and discover what's important, and share that with anyone or everyone you want

High Level Design (HLD)
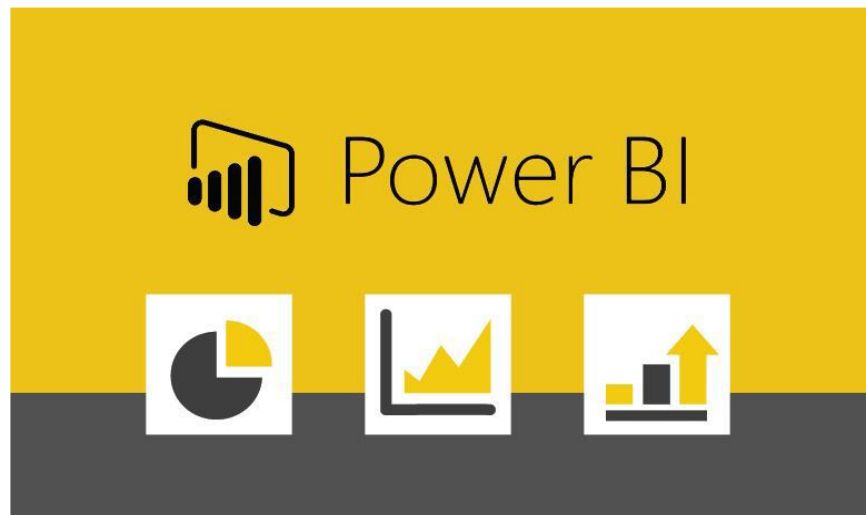
3. Design Details

3.1 Functional Architecture –
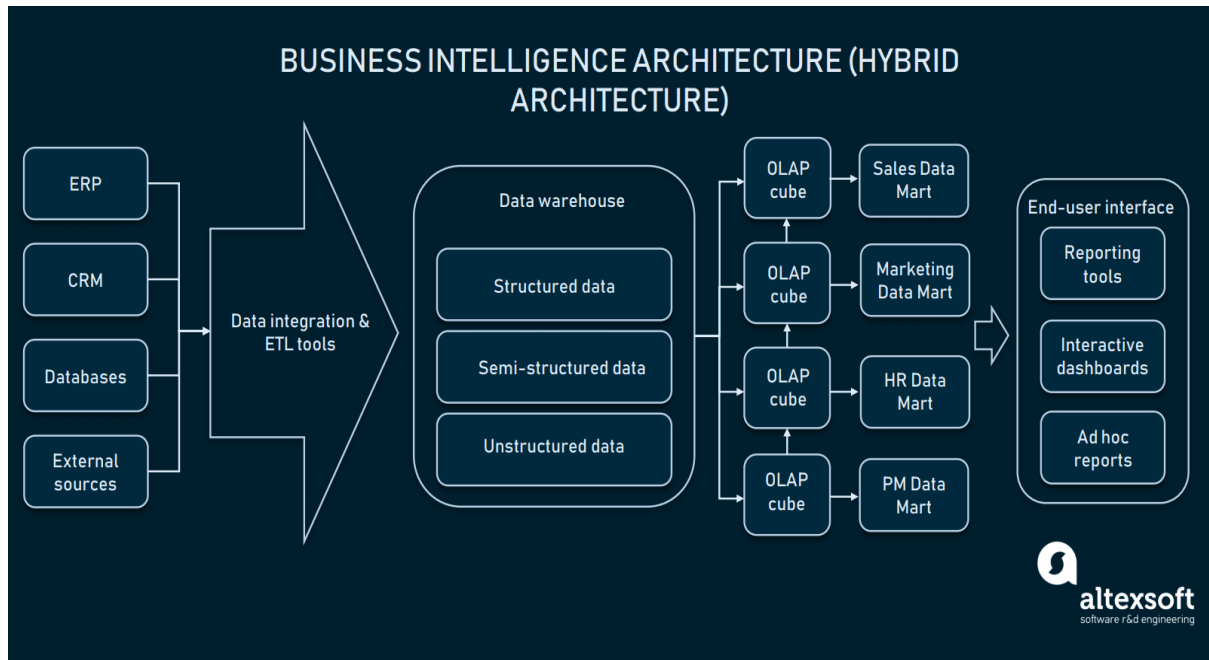
High Level Design (HLD)

The current business environment is constantly evolving. The global economic scenario is providing opportunities as well as challenges. The factors affecting business environment are consumer needs, globalization, and government policies, etc.

In such a business environment, organization basically has four action steps. The organization can be reactive, anticipative, adaptive, or/and proactive. For this, organization can develop a new strategy, get into partnership, etc.

Today most of the businesses are having a computerized business support. This support is in form of decision support system, business analysis, etc.
The main objective of business intelligence is to bridge the gap between organization current status and its desired position. Business intelligence helps organization achieve commercial success along with sound financial management.

Business intelligence is framework designed to support decision-making process. This framework combines architecture, database, analytical tools and applications. Business analytics forms an integral part of business intelligence.
Framework of Business Intelligence.

More and more businesses are moving towards business intelligence. The reason for this movement is the business environment. Organizations are forced to capture, store and interpret data. This data is at the core of business success.

Organizations require correct information for any decision-making process.

Business intelligence combines data warehousing, business analytics, performance, strategy and user interface. Business receives data from various sources. This data is capture in the data warehouse where it is stored, organized and summarized as per further utilization. Authorized users can access this data and work on it to get desired results. This result than are shared to executives for decision-making process. These data results can be published through dashboards or share points.

Business Intelligence Architecture and Components

The main components of business intelligence are data warehouse, business analytics and business performance management and user interface.

Data warehouse holds data obtained from internal sources as well as external sources. The internal sources include various operational systems.

Business analytics creates a report as and when required through queries and rules. Data mining is also another important aspect of business analytics.

Business performance management is a linkage of data with business objectives for efficient tracking. This business performance is then broadcasted to an executive decision-making body through dashboards and share-point.
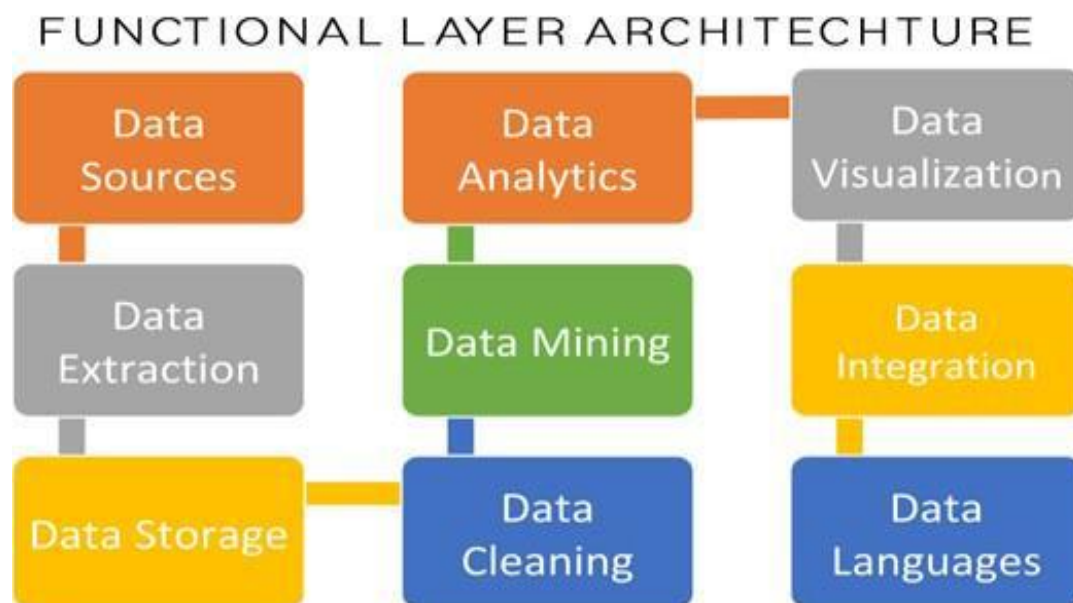
High Level Design (HLD)

Benefit of Business Intelligence

The benefits of Business intelligence are as follows:

❖ Business intelligence is faster more accurate process of reporting critical information.

❖ Business intelligence facilitates better and efficient decision-making process.

❖ Business intelligence provides timely information for better customer relationship management.

❖ Business intelligence improves profitability of the company.

❖ Business intelligence provides a facility of assessing organization's readiness in meeting new business challenges.

❖ Business intelligence supports usage of best practices and identifies every hidden cost.

Business intelligence usage can be optimized by identifying key projects on which company would like to focus. This process of highlighting key projects is called business intelligence governance.

FUNCTIONAL LAYER ARCHITECHTURE

Data Sources	Data Analytics	Data Visualization
Data Extraction	Data Mining	Data Integration
Data Storage	Data Cleaning	Data Languages

High Level Design (HLD)

3.2 Optimization

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

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.

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 dimension, while include 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 date 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).

Optimize and materialize your calculations

• Perform calculations in the database

• Reduce the number of nested calculations.

• Reduce the granularity of LOD or table calculations in the view. The more granular the calculation, the longer it takes. o LODs - Look at the number of unique dimension members in the calculation.

High Level Design (HLD)

o Table Calculations - the more marks in the view, the longer it will take to calculate.

• Where possible, use MIN or MAX instead of AVG. AVG requires more processing than MIN or MAX. Often rows will be duplicated and display the same result with MIN, MAX, or AVG.

4.KPIs

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 factors.

4.1 KPIs (Key Performance Indicators)

Key indicators displaying a summary of bookings, pricing and reviews information based on various parameters: -

❖ Room_ID – Showing distinct Id of room

❖ Survey_ID – Showing Survey Number

❖ Host_Id – Showing Distinct Id of hosts

❖ Room Type – Showing Type of Rooms available

❖ Country

❖ City

❖ Neighbourhood – Names of Neighbourhoods providing booking facilities.

❖ Reviews – Number of Reviews

❖ Overall Satisfaction – Ratings Given by Accommodates

❖ Accommodate – Number of accommodates

❖ Bedrooms – Showing number of bedrooms in room

❖ Bathrooms - Showing number of Bathrooms in room

❖ Price – Price of room

❖ Minstay

❖ Name

❖ Last_modified – Dates

High Level Design (HLD)

❖ Latitude – Showing the north-south coordinate of room location

❖ Longitude – Showing the east – west coordinate of room location

❖ Location – Showing the location of rooms

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

❖ **Patterns in business operations:**
Data visualization techniques help us to determine the patterns of business operations. By understanding the problem statement and identifying the solutions in terms of pattering and applied to eliminate one or more of the inherent problems.

❖ **Identify business trends and relate to data:**
These techniques help us identify market trends by collecting the data on Day-To-Day business activities and preparing trend reports, which helps track the business how influences the market. So that we could understand the competitors and customers. Certainly, this helps to long-term perspective.

❖ **Storytelling and Decision making:**
Knowledge of storytelling from available data is one of the niche skills for business communication, specifically for the Data Science domain which is playing a vital role. Using best visualization this role can be enhanced much better way and reaching the objectives of business problems.

❖ **Understand the current business insights and setting the goals:**
Businesses can understand the insight of the business KPIs, finding tangible goals and business strategy planning, therefore they could optimize the data for business strategy plans for ongoing activities.

❖ **Operational and Performance analysis:**
Increase the productivity with the help of visualization techniques the clarity of KPIs depicting the trends of the productivity of the manufacturing unit, and guiding were to improve the productivity of the plant.

High Level Design (HLD)

Power BI Server - On Premises

• Full control of hardware and software

• Infrastructure and data remain behind your firewall

• Need dedicated administrators to manage hardware and software

• Additional infrastructure needed to access off-network (mobile, external)

➢ Key Components:

Key Components of Power-BI ecosystem comprises:

❖ **Power BI Desktop**

The Windows-desktop-based application for PCs and desktops, primarily for designing and publishing reports to the Service.

❖ **Power BI Service**

The SaaS-based (software as a service) online service. This was formerly known as Power BI for Office 365, now referred to as PowerBI.com, or simply Power BI.

❖ **Power BI Mobile Apps**

The Power BI Mobile apps for Android and iOS devices, as well as for Windows phones and tablets.

❖ **Power BI Gateway**

Gateways used to sync external data in and out of Power BI and are required for automated refreshes. In Enterprise mode, can also be used by Flows and PowerApps in Office 365.

❖ **Power BI Embedded**

Power BI REST API can be used to build dashboards and reports into the custom applications that serves Power BI users, as well as non-Power BI users.

❖ **Power BI Report Server**

An on-premises Power BI reporting solution for companies that won't or can't store data in the cloud-based Power BI Service.

High Level Design (HLD)

❖ **Power BI Premium**

Capacity-based offering that includes flexibility to publish reports broadly across an enterprise, without requiring recipients to be licensed individually per user. Greater scale and performance than shared capacity in the Power BI service.

❖ **Power BI Visuals Marketplace**

A marketplace of custom visuals and R-powered visuals.