Airbnb Data Analysis
High Level Design

Project On:
Title : Airbnb Data Analysis
By: Rushikesh Khandare

High Level Design (HLD)

Contents

Abstract

Since 2008, guests and hosts have used AirBNB to expand on travelling possibilities and present more unique, personalized way of experiencing the world. This dataset describes the listing activity and metrics in San Diego, California for 2019. So, in this project we are going to analyse some key features. For this purpose, we will get our data from shared drive ink. I did Extract-Transform-Load the dataset and find for me some information from this large data. This is form of data mining. All information can be achieved by mining this data found key metrices and factors and showed the meaningful relationships between attributes.

Keywords: room_id, host_id, room_type, reviews , price, etc.

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.

2 General Description

2.1 Product Perspective & Problem Statement

No business can survive in this competitive market without managing their cost. It does not matter if revenues are high but if cost is higher; it is a red flag. So you are tasked to help management in creating and establishing new structure and models to reduce cost.

The objective of the project is to perform data visualization techniques to understand the insight of the data. This project aims to apply various Business Intelligence tools such as Tableau or Power BI, Python, MS Excel & SQL to get a visual understanding of the data.

2.2 Tools used

Business Intelligence tools and Python libraries such as NumPy, Pandas, Pandas Profiling, Matplotlib, Seaborn, Excel, SQL, Snowflake, Tableau or Power BI are used to build the whole framework.
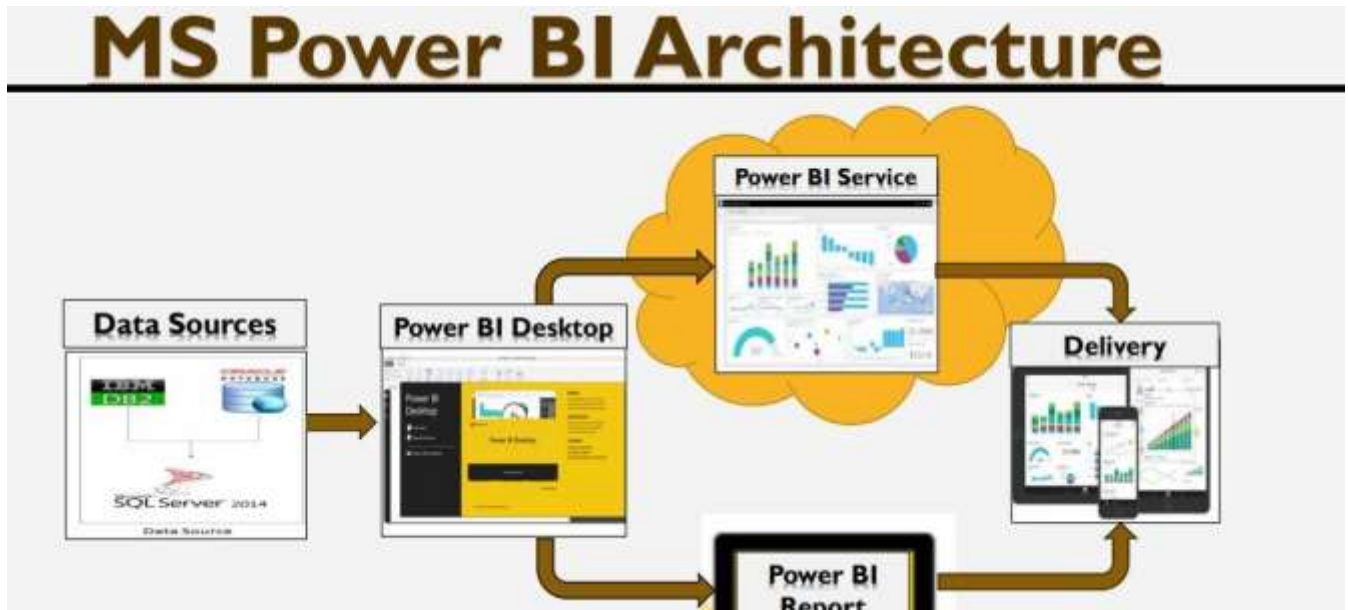
3 Design Details

3.1 MS Power BI Architecture

Fig: Architecture of Power BI

3.2 Optimization

Optimization guide for Power BI

You can optimize your solution at different architectural layers. Layers include:
- The data source(s)
- The data model
- Visualizations, including dashboards, Power BI reports, and Power BI paginated reports
- The environment, including capacities, data gateways, and the network.

1. Optimizing the data model

The data model supports the entire visualization experience. Data models are either external-hosted or internal-hosted, and in Power BI they are referred to as *datasets*. It's important to understand your options, and to choose the appropriate dataset type for your solution. There are three dataset modes: Import, Direct Query, and Composite.

For specific dataset mode guidance, see:

- Data reduction techniques for Import modelling
- DirectQuery model guidance in Power BI Desktop
- Composite model guidance in Power BI Desktop

2. Optimizing visualizations

Power BI visualizations can be dashboards, Power BI reports, or Power BI paginated reports. Each has different architectures, and so each has their own guidance.

1) Dashboards

It's important to understand that Power BI maintains a cache for your dashboard tiles—except live report tiles, and streaming tiles.

2) Power BI reports

The Power BI Desktop includes the Report View. Select the fields you want, add filters, choose from dozens of visualizations, format your reports with custom colors, gradients and several other options. The Report View gives you the same great report and visualizations tools just like when creating a report on PowerBI.com

3) Optimizing the environment

You can optimize the Power BI environment by configuring capacity settings, sizing data gateways, and reducing network latency.

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 Tableau 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 Server and Power BI Online leverage your existing technology investments and integrate 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. Below is a comparison of the three types:

Power BI Types:

1. 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)

2. Power BI Desktop

• Full control of software on managed hardware
• Puts infrastructure in same place as data (for migration to cloud)
• Flexibility to spin up/down hardware as needed
• Need dedicated administrators to manage software
• Additional infrastructure needed to access off-network (mobile, external)

Novypro:

Novypro is the most creative Business Intelligence portfolio that can host Power BI, Tableau, Astrato, or Excel projects and dashboards. novyPro.com is Free to use.

Depending on your organizational roles and responsibilities, Power BI Desktop should be installed by a systems administrator and the designated Power BI Server Administrator in coordination with the appropriate IT roles. For Power BI Online, you will integrate with your existing technology and configure the site settings. The Data & Analytics Survey, completed by business teams, identifies and prioritizes data use cases, audience size, and users. You will use the information collected in both surveys to plan your deployment strategy, including sizing, installation, and configuration of your Power BI Server or integration and configuration of Power BI Online with login through Microsoft Developer Account. We can create dashboards by using Power BI Desktop, Power BI service and publish on **Novypro** Power BI service and Power BI Mobile.