

ARCHITECTURE DESIGN DOCUMENT FOR AIRBNB DATA ANALYSIS

DOCUMENT VERSION CONTROL

Date Issue	Version	Description	Author
09/03/2023	1	Initial Architecture – V 1.0	Abdul Jaweed
25/06/2023	2	Final Architecture – V 2.0	Abdul Jaweed Aalmeen Khan Safa Ashraf

CONTENT

- DOCUMENT VERSION
- ABSTRACT
- INTRODUCTION
- WHY THIS ARCHITECTURE DESIGN DOCUMENT
- ARCHITECTURE DESCRIPTION
- DATA DESCRIPTION
- DATA STORAGE
- DATA VALIDATION
- DATA TRANSFORMATION
- DATABASE OPERATION
- DATA PREPROCESSING
- MODEL TRAINING
- LOAD PRODUCTION MODEL
- CLOUD SETUP
- PUSH APP TO CLOUD
- DATA FROM CLIENT SIDE FOR PREDICTION PURPOSE
- DATA PROCESSING AND PREDICTION
- EXPORT PREDICTION TO CSV

ABSTRACT

Airbnb is an American company that facilitates an online marketplace for lodging, primarily homestays for vacation rentals, and tourism activities. It basically connects travelers with local hosts who want to rent out their homes with people who are looking for accommodations in that locality. On the other hand, this platform enables hosts to list their available space and earn extra income in the form of rent and it also enables travelers to book unique homestays from local hosts, saving them money and giving them a chance to interact with locals. In the world of rising new technology and innovation, Travel industry is advancing with the role of Data Science and Analytics. Data analysis can help them to understand their business in a quite different manner and helps to improve the quality of the service by identifying the weak areas of the business. This study demonstrates how different analyses help out to make better business decisions and help analyze customer trends and satisfaction, which can lead to new and better products and services. Different analyses performed such as Exploratory Data Analysis and Descriptive Analysis on a variety of use cases to get the key insights from this data based on which business decisions will be taken.

INTRODUCTION

WHY THIS ARCHITECTURE DESIGN DOCUMENT ?

Any software needs the architectural design to represent the design of the 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 help the designer to understand the overall properties of the system

SCOPE

Architecture Design Document (ADD) is an architectural 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

The architecture for an Airbnb data analysis project typically involves several components and steps to process and analyze the data. Here's a high-level description of the architecture for an Airbnb data analysis project:

Data Collection: The first step is to gather the necessary data from Airbnb. This can include data on listings, reviews, bookings, host information, and more. The data can be obtained through Airbnb's public APIs, data scraping, or by downloading publicly available datasets.

Data Storage: Once the data is collected, it needs to be stored in a suitable data storage system. Common options include relational databases (such as MySQL, PostgreSQL, or SQL Server) or NoSQL databases (such as MongoDB or Elasticsearch). The choice of data storage depends on factors such as the volume of data, the complexity of relationships, and the requirements of the analysis.

Data Cleaning and Preprocessing: Before performing analysis, the collected data needs to be cleaned and preprocessed. This involves tasks such as handling missing values, removing duplicates, standardizing data formats, and transforming the data into a suitable structure for analysis.

Data Transformation and Integration: In some cases, additional data sources might be required to enrich the analysis. This could include external datasets such as geographical data, demographic data, or weather data. Data transformation and integration techniques, such as data joins and merges, are used to combine different datasets and create a unified view for analysis.

Data Analysis and Visualization: Once the data is prepared, various data analysis techniques can be applied to gain insights. This can involve statistical analysis, exploratory data analysis, data mining, and machine learning algorithms. Visualization tools, such as Tableau, Power BI, or Python

libraries like Matplotlib or Seaborn, are used to create visual representations of the data and communicate findings effectively.

Deployment and Reporting: The analyzed results and visualizations can be deployed in different ways to share insights with stakeholders. This can include creating interactive dashboards, generating reports, or building web-based applications. The choice of deployment method depends on the intended audience and their requirements for accessing and interacting with the analysis results.

Ongoing Monitoring and Iteration: Data analysis projects are often iterative processes, where insights and findings lead to further questions and refinements. It's important to monitor the performance of the analysis over time, update the data as new information becomes available, and iterate on the analysis to gain deeper insights or address emerging questions.

Overall, the architecture of an Airbnb data analysis project involves data collection, storage, cleaning, preprocessing, analysis, visualization, and deployment stages. The specific tools and technologies used can vary based on the project requirements and the expertise of the data analysis team.

TABLEAU DASHBOARD

Tableau is a leading data visualization and business intelligence tool that empowers users to create interactive and visually compelling dashboards and reports. Here's a Tableau version of the description:

Tableau is a robust data analysis and visualization platform that enables users to explore and present data in a user-friendly and intuitive manner. It offers a wide range of features and capabilities that make it a popular choice for businesses and analysts.

Data Connectivity: Tableau supports connectivity to various data sources, including spreadsheets, databases, cloud-based platforms, and web services. It provides native connectors to popular data sources such as Excel, SQL databases, Salesforce, and Google Analytics, allowing users to seamlessly connect and integrate their data.

Visual Exploration: With Tableau, users can easily drag and drop data fields onto the canvas to create dynamic visualizations. The platform offers a vast library of charts, graphs, maps, and other visual elements that can be customized to convey insights effectively. Users can interact with the visualizations, apply filters, and drill down into the data to explore different perspectives and uncover hidden patterns.

Dashboard Creation: Tableau allows users to combine multiple visualizations into interactive dashboards. Users can arrange and organize their visualizations in a layout that suits their needs, enabling them to create comprehensive and visually appealing dashboards. Dashboard elements can be linked to create interactivity, allowing users to explore the data and make discoveries.

Advanced Analytics: Tableau provides advanced analytics capabilities that empower users to perform calculations, apply statistical functions, and perform complex data modeling. Users can create calculated fields, use parameters for dynamic analysis, and leverage built-in statistical functions for

deeper insights. Tableau also supports integration with popular programming languages such as R and Python for advanced analytics and machine learning.

Sharing and Collaboration: Tableau offers various options for sharing and collaborating on dashboards and reports. Users can publish their visualizations to Tableau Server or Tableau Online, enabling secure access for authorized users within their organization. Dashboards can also be shared via Tableau Public or embedded into websites and applications, allowing for broader dissemination of insights.

Performance Optimization: Tableau provides features to optimize performance when working with large datasets or complex visualizations. Techniques such as data extracts, data blending, and data source filters can be employed to enhance performance. Tableau also offers options for caching, data aggregation, and incremental data updates to improve responsiveness and reduce query times.

Tableau offers a user-friendly interface, powerful data visualization capabilities, and a robust set of features for data exploration and analysis. Its versatility and flexibility make it a valuable tool for individuals and organizations across various industries to derive insights from their data and make data-driven decisions.

AIRBNB DATA ANALYSIS VISUALIZATION DASHBOARD

