**Low-Level Design (LLD)**

**Airbnb Data Analysis**

Revision Number: 1.0

Last Date of Revision: 25/03/2024

AYUSH TRIPATHI

**Contents:**

1. Introduction
   * 1.1 What is Low-Level Design Document?
   * 1.2 Scope
2. Architecture
3. Architecture Description
   * 3.1 Data Description
   * 3.2 Data Cleaning and Preprocessing
   * 3.3 Data Analysis
   * 3.4 Data Visualization
   * 3.5 User Interface Elements
4. Implementation Details
   * 4.1 Data Cleaning and Preprocessing Module
   * 4.2 Data Analysis Module
   * 4.3 Data Visualization Module
   * 4.4 User Interface Implementation
5. **Introduction**
   1. **What is Low-Level Design Document?**

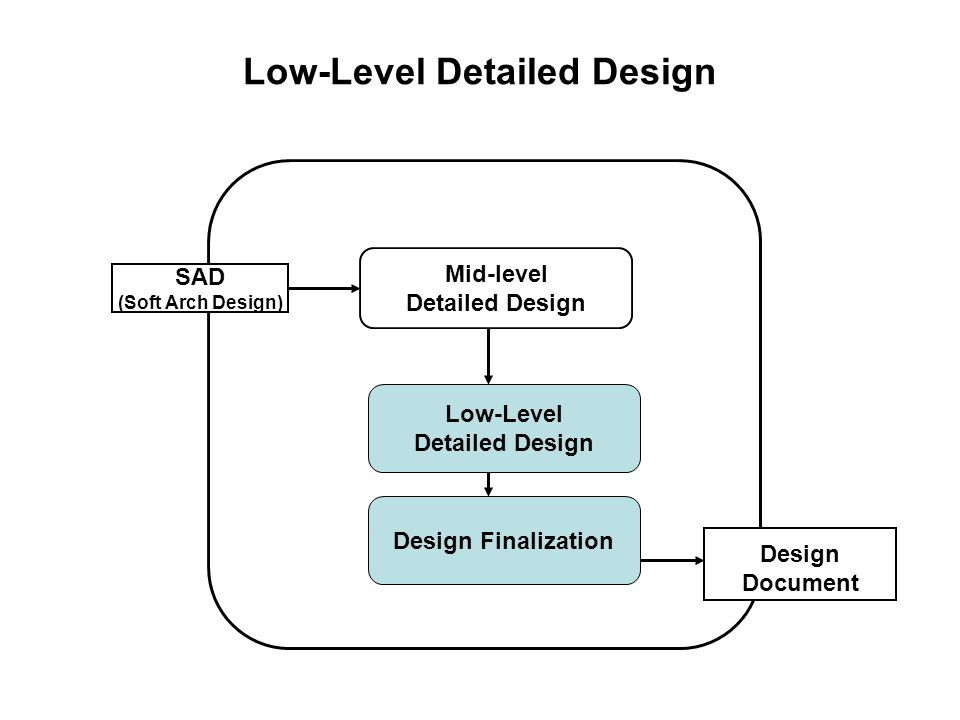
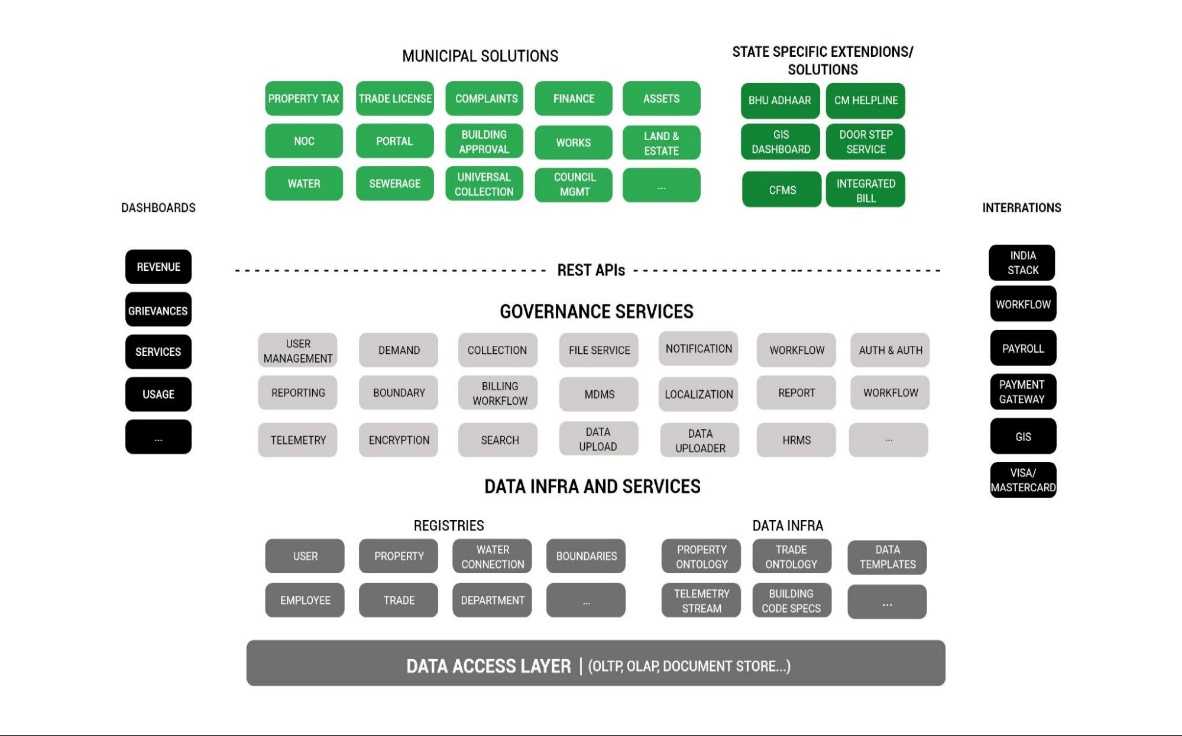
The Low-Level Design Document (LLD) serves as a foundational document that outlines the intricate details of the Airbnb Data Analysis internship project developed on Power BI. Unlike high-level design documents that provide a broader overview, the LLD delves into the internal logic, class diagrams, and specifications essential for translating the project concept into actionable code. It delineates the structure of the program, defining the relationships between classes, methods, and program specifications. Essentially, the LLD bridges the gap between conceptualization and implementation by offering a detailed roadmap for developers to follow during the coding phase.

* 1. **Scope?**

The scope of the Airbnb Data Analysis internship project encapsulates the entirety of the data analysis and visualization process conducted on the Power BI platform. It encompasses data cleaning, preprocessing, analysis, visualization, and the implementation of various user interface elements using Power BI visual tools. By defining the scope upfront, the LLD provides clarity on the project's objectives, deliverables, and boundaries, ensuring alignment with stakeholder expectations and project requirements. Additionally, a well-defined scope serves as a guiding framework for developers, streamlining the development process and minimizing scope creep.

1. **Architecture**

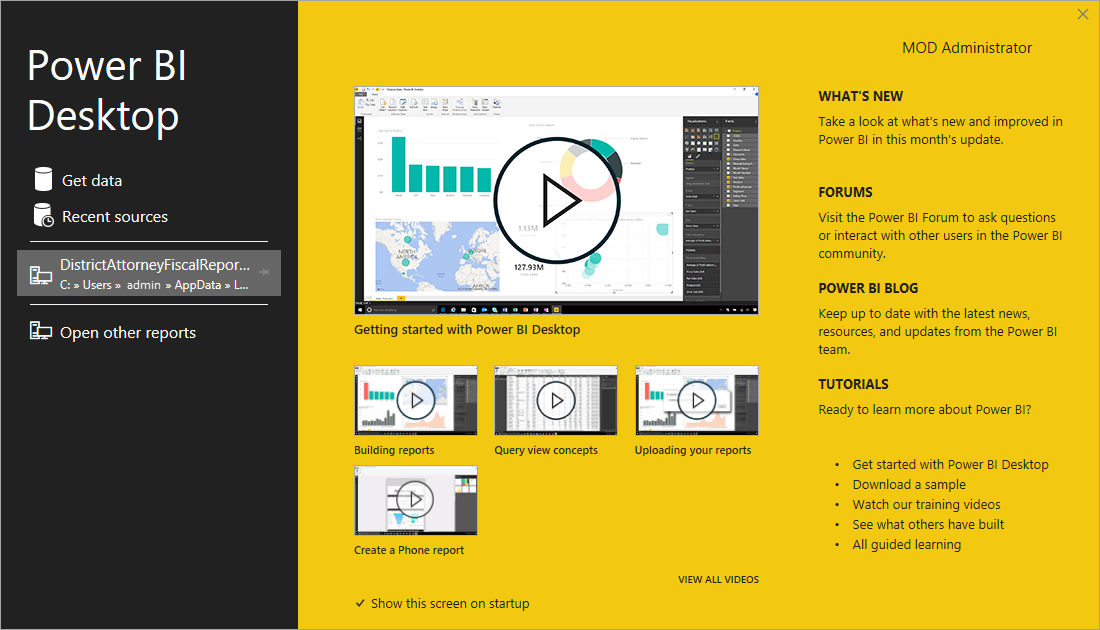
The architecture section of the LLD offers an overarching view of the project's structure, highlighting the key modules and components involved in data analysis and visualization. It provides insights into the system's design principles, architectural patterns, and the interactions between different subsystems. By delineating the architecture upfront, the LLD lays the foundation for a scalable, maintainable, and extensible solution, fostering modularity, flexibility, and reusability in the codebase. Moreover, a well-defined architecture facilitates collaboration among development teams, promotes consistency in code organization, and enables easier troubleshooting and debugging during the development lifecycle.

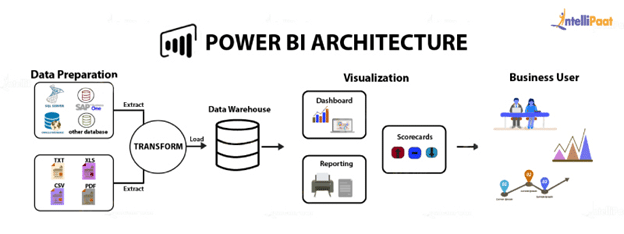
Power BI Server Architecture

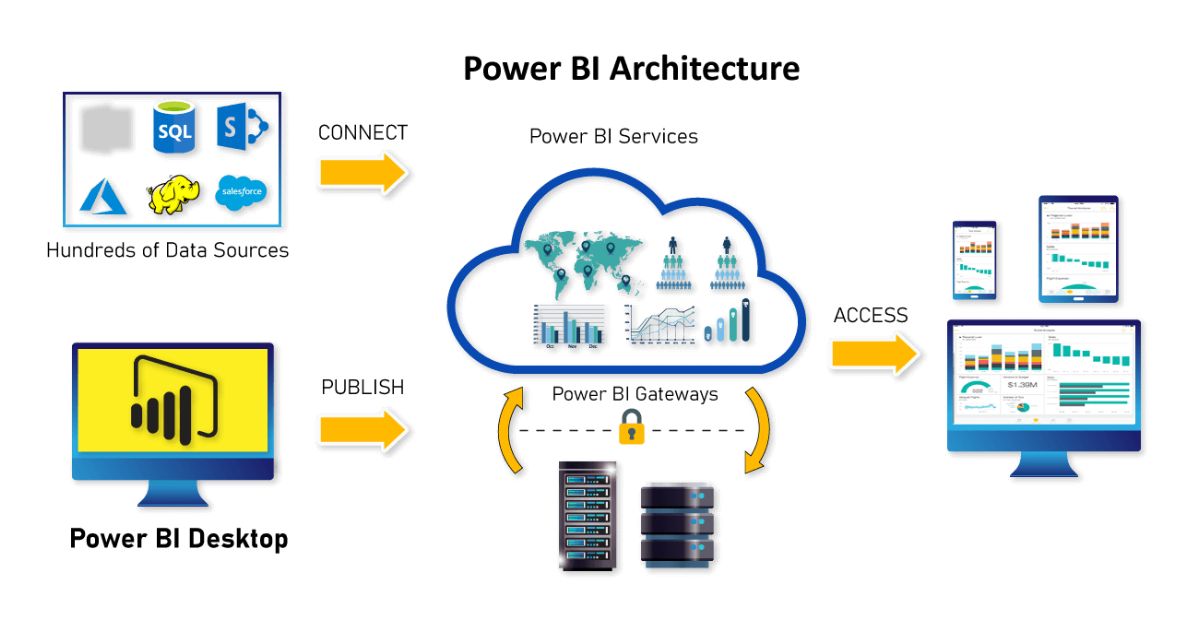
Power BI is a powerful business intelligence tool developed by Microsoft, offering a suite of features for data analysis, visualization, and reporting. Its architecture is designed to facilitate seamless data integration, transformation, and visualization, empowering users to derive actionable insights from their data. Below is an overview of the key components and layers of the Power BI architecture:

1. **Data Sources:**
   * Power BI connects to various data sources, including databases, files, cloud services, and online services, to ingest data for analysis.
   * Supported data sources include SQL Server, Excel, SharePoint, Dynamics 365, Salesforce, Google Analytics, Azure services, and many more.
2. **Data Ingestion:**
   * Once connected to data sources, Power BI performs data ingestion, extracting data from the source systems and loading it into memory for analysis.
   * It supports both direct query mode, where queries are executed live against the source systems, and import mode, where data is imported into Power BI's internal data model.
3. **Data Modeling:**
   * Data modeling is a crucial aspect of Power BI architecture, involving the transformation and shaping of data to create a structured data model.
   * Power BI provides a robust data modeling engine that enables users to define relationships between tables, create calculated columns and measures, and perform data shaping operations.
4. **Data Storage:**
   * Power BI stores data in a proprietary in-memory data model called the Power BI Desktop file (.pbix), which can be published to the Power BI service for sharing and collaboration.
5. **Power BI Desktop:**
   * Power BI Desktop is a desktop application used for authoring and designing Power BI reports and dashboards.
   * It provides a rich set of data visualization tools, including charts, graphs, maps, and tables, as well as advanced features for data exploration and analysis.



1. **Power BI Service:**
   * The Power BI service is a cloud-based platform for publishing, sharing, and collaborating on Power BI reports and dashboards.
   * It allows users to publish Power BI Desktop files, schedule data refreshes, create data-driven alerts, and collaborate with colleagues in real-time.
2. **Gateway:**
   * The Power BI Gateway acts as a bridge between on-premises data sources and the Power BI service in the cloud.
   * It facilitates secure data transfer between the on-premises data sources and Power BI, enabling users to access and refresh data from their local environments.
3. **Data Refresh:**
   * Data refresh is a key feature of Power BI that enables users to keep their reports and dashboards up-to-date with the latest data.
   * Users can schedule data refreshes in the Power BI service, specifying the frequency and refresh settings for each dataset.
4. **Security and Compliance:**
   * Power BI provides robust security and compliance features to ensure data confidentiality, integrity, and availability.
   * It supports role-based access control (RBAC), row-level security (RLS), encryption at rest and in transit, and compliance certifications such as GDPR and HIPAA.
5. **Integration with Other Microsoft Products:**
   * Power BI seamlessly integrates with other Microsoft products and services, including Office 365, Azure, Dynamics 365, and SharePoint, enabling users to leverage existing investments and ecosystems.





1. **Architecture Description**
   1. **Data Description**
2. room\_id: This attribute serves as a unique identifier for each room listed in the dataset. It helps distinguish one room from another and enables tracking and referencing specific rooms in the data analysis process.
3. survey\_id: The survey\_id attribute identifies the survey or data collection event from which the information about the rooms was gathered. It helps in understanding the context or source of the data and may be useful for tracking the survey's details.
4. host\_id: This attribute represents a unique identifier assigned to each host or property owner in the dataset. It allows grouping rooms based on the host, enabling analysis at the host level, such as evaluating the performance of individual hosts.
5. room\_type: The room\_type attribute specifies the type of room available for rent, such as an entire home/apartment, private room, or shared room. It categorizes the rooms based on their accommodation arrangements, providing insights into the diversity of rental options.
6. country: Country indicates the country where each property is located. It helps segment the data based on geographical regions, allowing for country-level analysis or comparison across different countries.
7. city: The city attribute denotes the specific city where each property is situated. It enables analysis at the city level, facilitating comparisons between different cities and understanding variations in rental markets.
8. borough: This attribute refers to the borough or district within a city where each property is located. It provides additional granularity in geographical segmentation, particularly in cities with multiple boroughs or districts.
9. neighborhood: Neighborhood represents the specific neighborhood or area within a borough where each property is situated. It offers localized insights into rental markets and allows for analysis at the neighborhood level.
10. reviews: The reviews attribute indicates the number of reviews received for each property. It reflects the level of feedback or engagement from guests and can serve as a measure of the property's popularity or reputation.
11. overall\_satisfaction: This attribute represents the overall satisfaction rating of each property, aggregated from guest reviews. It provides a quantitative measure of guest satisfaction and helps assess the quality of rental accommodations.
12. accommodates: Accommodates specifies the maximum number of guests that each property can accommodate. It influences rental suitability and helps match properties with guests' accommodation needs.
13. bedrooms: Bedrooms denotes the number of bedrooms available in each property. It indicates the property's capacity for sleeping arrangements and is a key factor in determining rental suitability for groups or families.
14. bathrooms: The bathrooms attribute indicates the number of bathrooms available in each property. It contributes to the property's overall amenities and comfort level, influencing guest satisfaction and rental value.
15. price: Price represents the rental price listed for each property. It serves as a crucial factor for both hosts and guests, influencing rental decisions and market competitiveness.
16. minstay: Minstay denotes the minimum length of stay required for booking each property. It provides information on booking policies and helps guests plan their stays accordingly.
17. name: The name attribute refers to the name or title of each property listed for rent. It offers descriptive information about the properties and may influence guests' booking decisions.
18. last\_modified: Last\_modified indicates the date and time when each property listing was last modified. It helps track updates or changes to property listings over time and ensures data freshness.
19. latitude: Latitude represents the geographic latitude coordinate of each property's location. It enables spatial analysis and visualization, facilitating mapping and geographical insights.
20. longitude: Longitude denotes the geographic longitude coordinate of each property's location. Like latitude, it supports spatial analysis and visualization, enabling precise mapping and geographical insights.
21. location: Location provides additional location information or address details for each property. It may include specific landmarks, street addresses, or descriptive location information to help identify the property's whereabouts.
    1. **Data Cleaning and Preprocessing**

The data cleaning and preprocessing module are integral components of the Airbnb Data Analysis project, aimed at ensuring data integrity and quality before analysis. This section of the LLD outlines the various steps involved in cleaning and preprocessing the Airbnb dataset, including handling missing values, removing duplicates, standardizing data formats, and resolving inconsistencies. Additionally, it describes the algorithms and techniques employed for data transformation and normalization, preparing the dataset for further analysis.

* 1. **Data Analysis**

The data analysis module is responsible for extracting meaningful insights and patterns from the cleaned dataset, shedding light on key trends, correlations, and predictive models. This section of the LLD elucidates the methodologies and algorithms used for data analysis, ranging from descriptive statistics and exploratory data analysis to advanced machine learning techniques. It discusses the process of feature engineering, model training, evaluation, and validation, emphasizing the importance of robust analytical techniques in uncovering actionable insights from the data.

**3.4 Data** **Visualization**

Data visualization plays a pivotal role in communicating insights and findings effectively to stakeholders, facilitating decision-making and understanding complex datasets. This section of the LLD elaborates on the visualization techniques employed in the Airbnb Data Analysis project, including clustered column charts, stacked column charts, donut charts, clustered bar charts, cards, tables, and slicers. It discusses the principles of visual design, interactivity, and usability, ensuring that the visualizations are intuitive, informative, and visually appealing to end-users.

**3.5 User Interface Elements**

The user interface elements are the interface through which users interact with the analyzed data, exploring insights and making informed decisions. This section of the LLD details the implementation of user interface elements using Power BI, encompassing the design and layout of visualizations, filters, slicers, and navigation controls. It emphasizes the importance of user-centric design principles, accessibility, and responsiveness in creating a seamless user experience, enabling users to explore and interact with the data effortlessly.

1. **Implementation Details**
   1. **Data Cleaning and Preprocessing Module**

The implementation details of the data cleaning and preprocessing module are crucial for ensuring the accuracy and reliability of the analyzed data. This section of the LLD provides a step-by-step guide to implementing data cleaning and preprocessing techniques, including code snippets, functions, and libraries used for handling missing values, data transformation, and standardization. It addresses potential challenges and edge cases encountered during the implementation process, offering insights into best practices and optimization strategies for efficient data preparation.

* 1. **Data Analysis Module**

The data analysis module serves as the core engine driving insights and decision-making in the Airbnb Data Analysis project. This section of the LLD delves into the implementation details of data analysis techniques and algorithms, including statistical analysis, machine learning models, and predictive analytics. It discusses the selection of appropriate analytical tools, libraries, and frameworks, as well as the integration of custom scripts and functions for advanced analysis. Additionally, it outlines the process of model training, evaluation, and validation, ensuring the accuracy and robustness of the analytical results.

* 1. **Data Visualization Module**

The data visualization module is responsible for translating analytical findings into actionable insights, leveraging the power of visual storytelling to convey complex information effectively. This section of the LLD provides insights into the implementation of data visualization techniques using Power BI visual tools, including the creation of interactive dashboards, reports, and visualizations. It discusses the selection of appropriate chart types, color schemes, and layout designs, as well as the integration of interactive elements such as filters, slicers, and drill-down capabilities. Additionally, it addresses considerations for accessibility, responsiveness, and performance optimization to enhance the user experience.

* 1. **User Interface Implementation**

The implementation of user interface elements is critical for facilitating user interaction and engagement with the analyzed data. This section of the LLD outlines the implementation details of user interface elements using Power BI, including the design and layout of dashboards, reports, and interactive visualizations. It discusses the integration of navigation controls, filters, slicers, and tooltips to enhance usability and interactivity. Additionally, it addresses considerations for responsiveness, accessibility, and cross-device compatibility to ensure a seamless user experience across different platforms and devices.