Amazon Customer ANALYTICS

High Level Design (HLD)

Revision Number: 1.0

Last date of revision: 19/11/2022

Document Version Control

Date Issued	Version	Description	Author
19th June 2022	1.0	complete HLD	Ayush Bhardwaj

Contents

Abstract

Housing prices are an important reflection of the economy, and housing price ranges are of great interest for both buyers and sellers. Ask a home buyer to describe their dream house, and they probably won't begin with the height of the basement ceiling or the proximity to an east-west railroad. But this playground competition's data-set proves that much more influences price negotiations than the number of bedrooms or a white-picket fence. Accurately predicting house prices can be a daunting task. The buyers are just not concerned about the size(square feet) of the house and there are various other factors that play a key role to decide the price of a house/property.

High Level Design (HLD)   High Level Design (HLD)

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.

High Level Design (HLD)

2 General Description
2.1 Product Perspective & Problem Statement
Our "Domain Sale" process is structured to help potential buyers purchase the domain
they want immediately without the hassle of contacting the seller directly. ... A seller lists a domain for sale at a specific price in our Marketplace. An interested buyer sees this domain for sale and decides to buy it.
 Our main aim is to do the ETL and analyze, Visualize the relationship between Revenue, Sales, and Profit over various features like customers, time period, location, categories of items, and budget.
2.2 Tools used
Business Intelligence tools and libraries works such as Numpy, Pandas, Excel, R, Tableau, Power BI are used to build the whole framework.
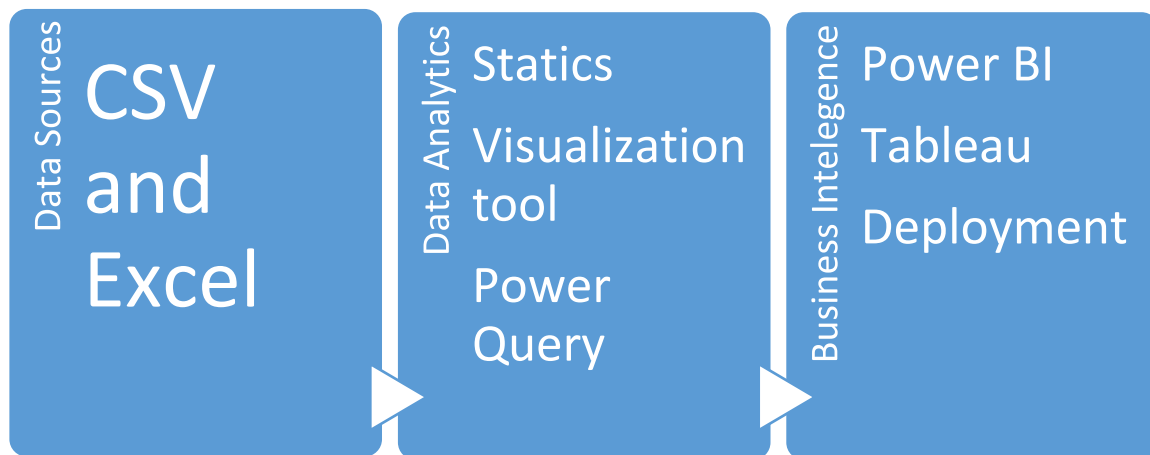High Level Design (HLD)

3 Design Details
3.1 Functional Architecture
Figure 1: Functional Architecture of Business Intelligence
  High Level Design (HLD)

Data Sources	Data Analytics	Business Intelegence
CSV and Excel	Statics Visualization tool Power Query	Power BI Tableau Deployment

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. ○ LODs - Look at the number of unique dimension members in the calculation.

　　○ 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.

High Level Design (HLD)

• • Make groups with calculations. Like include filters, calculated groups load only named members of the domain, whereas Power BI group function loads the entire domain.

• • Use Booleans or numeric calculations instead of string calculations. Computers can process integers and Booleans (t/f) much faster than strings. Boolean>Int>Float>Date>DateTime>String

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 the Sales, Revenue, Profit, and its relationship with different metrics
1. Revenue over different territory
2. Correlation of revenue and list price
3. Revenue, List Price ,Days to manufacture over different category
4. Influence of age, gender, and Marital Status on Income
5. Change of profit over time
6. Price and tax Amount by Sub category and category
7. Budget of 2016 overtime
8. Budget and expenses by category, subcategory and product name

5 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 analyzing 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.
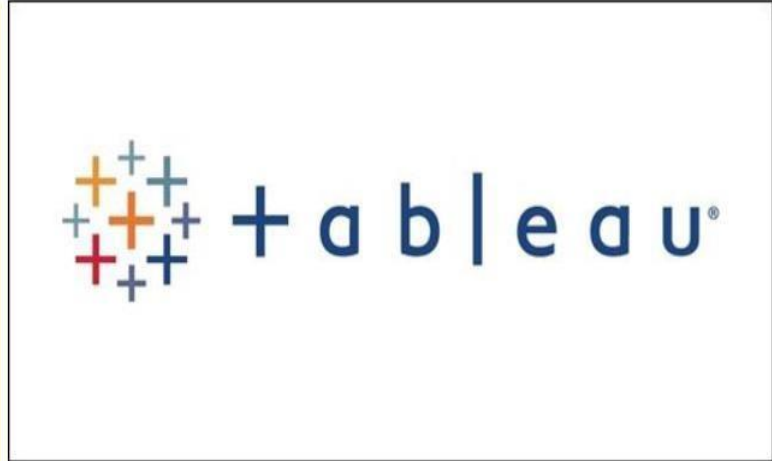
Power BI prioritizes choice in flexibility to fit, rather than dictate, your enterprise architecture. Power BI leverages your existing technology investments and integrate into your IT infrastructure to provide a self-service, modern analytics platform for your users. With onpremises, cloud, and hosted options, there is a version of Power BI to match your requirements. Below is a comparison of the three types:

TYPE PROS CONS

Power BI Desktop

- • Full control of hardware and software
- • Infrastructure and data remain behind your firewall
- • Need dedicated administrators to manage hardware and software

Power BI Service

- • Full control of software on managed hardware
- • Single-site in a multi-tenant environment
- • No guest account access

Power BI mobile

- • Fully hosted solution (hardware, software upgrades)
- • Easy for the external audience to access

Depending on your organizational roles and responsibilities, Power BI service should be installed by a systems administrator and the designated Power BI Administrator in coordination with the appropriate IT roles. For Power BI services, 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 service or integration and configuration of Power BI service. In addition to installing Power BI Service, administrators will also need to plan for the client software installation of Power BI Desktop, Power BI Mobile, and Power BI service where applicable..