High Level Design (HLD) Amazon Sales Data Analysis

Revision Number: 1.0

Last date of revision: 15/11/2023

Mr. Tushar Padavi

Miss. Pooja Pawar

Mr. Chetan Manakar

Mr. Dinesh Ingle

# Document Version Control

|  |  |  |  |
| --- | --- | --- | --- |
| **Date Issued** | **Version** | **Description** | **Author** |
| **4th Oct 2023** | 1.0 | First Version of Complete HLD | Pooja Pawar  Tushar Padavi  Chetan Mankar  Dinesh Ingle |

**Contents**

[Document Version Control 2](#_bookmark0)

[Abstract 3](#_bookmark1)

1. [Introduction 5](#_bookmark2)
   1. [Why this High-Level Design Document? 5](#_bookmark3)
   2. [Scope 5](#_bookmark4)
2. [General Description 6](#_bookmark5)
   1. [Product Perspective & Problem Statement 6](#_bookmark6)
   2. [Tools used 6](#_bookmark8)
3. [Design Details 7](#_bookmark9)
   1. [Functional Architecture 7](#_bookmark10)
   2. [Optimization 8](#_bookmark11)
4. [KPIs 9](#_bookmark12)
   1. [KPIs (Key Performance Indicators) 9](#_bookmark13)
5. [Deployment 9](#_bookmark14)

# Abstract

This project involves analyzing Amazon sales data to uncover trends and insights. By examining product sales, customer behavior, and market dynamics, the aim is to optimize strategies for growth and profitability. Statistical techniques and data visualization will be employed to extract actionable recommendations for business improvement. The project seeks to enhance decision-making and drive revenue in the highly competitive e-commerce landscape.

# Introduction

## 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:
      * Security
      * Reliability
      * Maintainability
      * Portability
      * Reusability
      * Application compatibility
      * Resource utilization
      * Serviceability

## 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.

# General Description

## Product Perspective & Problem Statement

Sales management has gained importance to meet increasing competition and the need for improved methods of distribution to reduce cost and to increase profits.

Sales management today is the most important function in a commercial and business

Enterprise.

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

## Tools used

Business Intelligence tools and libraries works such as Numpy, Pandas, Excel, R, Tableau, Power BI are used to build the whole framework.



# Design Details

## Functional Architecture

Figure 1: Functional Architecture of Business Intelligence



## 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

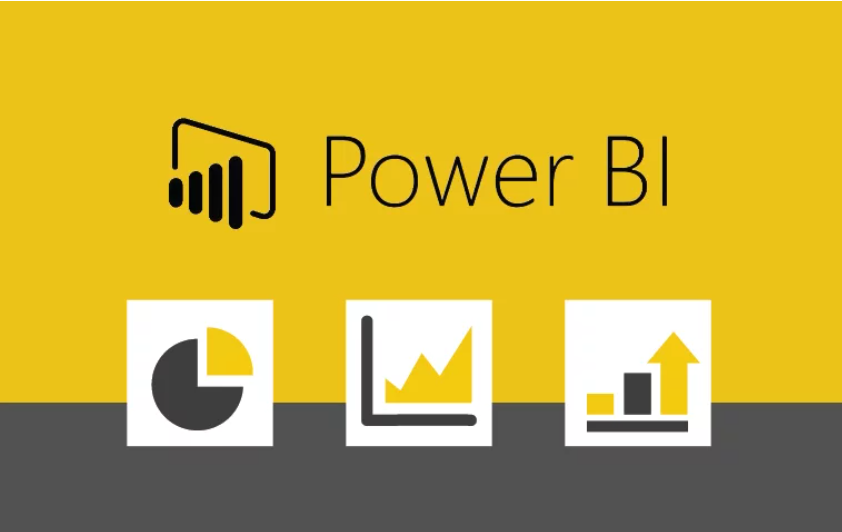
* + - Practice guided analytics. There’s no need to fit everything you plan to show in a single view. Compile related views and connect them with action filters to travel from overview to highly-granular views at the speed of thought.
    - Remove unneeded dimensions from the detail shelf.
    - Explore. 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](http://onlinehelp.tableau.com/current/pro/online/mac/en-us/help.htm#filtering_add_dragfields_dates.html). 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](http://www.tableau.com/learn/tutorials/on-demand/logical-calculations). Computers process integers and Booleans (t/f) much faster than strings.
    - Use [parameters](http://onlinehelp.tableau.com/current/pro/online/en-us/help.htm#parameters.html) and [action filters](http://onlinehelp.tableau.com/current/pro/online/en-us/help.htm#actions.html). These reduce the query load (and work across data sources).

# 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

## KPIs (Key Performance Indicators)

Key indicators displaying a summary of the Housing Price and its relationship with different metrics

1. Impact of Unit Sold on Product Price across the Countries
2. Impact of Total Profit and Unit Sold on Month across the Region
3. Influence of Total Profit parameter on Countries
4. Influence of Total Profit parameter on Item Type
5. Influence of Total Revenue parameter on Countries and Year
6. Influence of Total Profit and Total Cost on Year

# 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 Power BI 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:

TYPE PROS CONS

### Power BI Report Server - On Premises

**1] On-Premises Hosting:** Power BI Report Server allows organizations to host their Power BI reports and dashboards on their own servers or private cloud, ensuring data remains on-premises for data security and compliance reasons.

**2] Report Creation and Publishing:** Reports are created using Power BI Desktop, and once designed, they can be published to Power BI Report Server. This allows users to access and interact with reports through a web portal within the organization's network.

**3] Data Source Connectivity:** Reports in Power BI Report Server can connect to a variety of data sources, including on-premises databases and cloud-based sources, enabling organizations to integrate data from different locations.

**4] Scheduled Data Refresh**: Power BI Report Server supports scheduled data refresh, ensuring that reports remain up-to-date with the latest information from on-premises and cloud data sources.

**5] Security and Governance:** The platform provides robust security and governance features, allowing organizations to manage user access and permissions for report content, helping to maintain data integrity and access control.

Regenerate

Top of Form

Bottom of Form

### Power BI Server - Public Cloud (IaaS)

**1] Virtual Machines (VMs) in a Public Cloud:** You can set up virtual machines (VMs) within a public cloud environment such as Microsoft Azure or AWS and install Power BI Report Server on those VMs. This allows you to host Power BI Server in a cloud-based IaaS environment. It provides more flexibility and scalability compared to traditional on-premises deployment.

**2] Data Sources and Connectivity:** You can connect your Power BI Server hosted in the cloud to on-premises and cloud-based data sources, just as you would with an on-premises Power BI Server. This ensures that you can access and analyze data from various sources while keeping your reports within the cloud infrastructure.

**3] Scalability:** Cloud-based VMs offer scalability options, so you can easily adjust the resources (CPU, memory, storage) allocated to your Power BI Server to meet your organization's changing needs. This makes it easier to handle increased workloads or user demands.

**4] Data Security and Compliance:** When deploying Power BI Server in a public cloud IaaS environment, you can leverage the cloud provider's security features and compliance certifications to ensure that your data remains secure and compliant with industry standards.

**5] Access Control:** You can control access to the Power BI Server in the cloud by configuring network security groups, firewalls, and identity and access management (IAM) features provided by your cloud provider. This allows you to manage user access and permissions effectively.

### Tableau Online (SaaS)

**1] Cloud-Based Access:** Power BI Online is accessible through a web browser, making it a convenient and accessible choice for users across an organization. There is no need to install or maintain software on individual machines.

**2] Data Integration:** It allows you to connect to a wide range of data sources, both on-premises and in the cloud. You can ingest, transform, and model data from various sources to create comprehensive reports and dashboards.

**3] Self-Service Analytics:** Power BI Online empowers non-technical users to create their own reports and dashboards with user-friendly drag-and-drop interfaces. It provides a user-friendly experience for data exploration and visualization.

**4 Collaboration and Sharing**: Users can easily share reports and dashboards with colleagues and stakeholders within and outside the organization. Collaboration features include real-time co-authoring and commenting, making it easy to work on reports together.

Depending on your organizational roles and responsibilities, Power BI Server 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. In addition to installing Power BI Server or configuring Power BI Online, administrators will also need to plan for the client software installation of Power BI Prep Builder, Power BI Desktop, Power BI Mobile, and Power BI Bridge for Power BI Online where applicable.