High Level Design (HLD)

Amazon Sales Data

Revision Number - 1.2
Last Date of Revision –15.10.2022

Arnab De
&
Siddharta Shankar Dhibar

Document Version Control

Date Issued	Version	Description	Author
14/10/2022	1.0	Abstract, Introduction, General Description	Arnab De, Siddharta Shankar Dhibar
14/10/2022	1.1	Design Detail, KPI, Deployment	Arnab De, Siddhartha Shankar Dhibar
14/10/2022	1.2	Final Revision	Arnab De, Siddharta Shankar Dhibar

Abstract

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.

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 before 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

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

The goal of this project is to analyse to predict sales trends over the month, year and year month. Sales trend is basically to analyse whether the sales is going in profit or loss.

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.

2.2 Tools used

Business Intelligence tools and libraries works such as Pandas, Matplotlib, MS-Excel, Power BI, Google colab and Python Programming Language are used to build the whole framework.

- Jupyter Notebook is used as IDE.

- Python is the Programming Language used.

- EDA is done using Pandas.

- Visualizations were done using Matplotlib.

- Power BI is used for dashboard creation & Visualization

# 3 Design Details

3.1 Functional Architecture
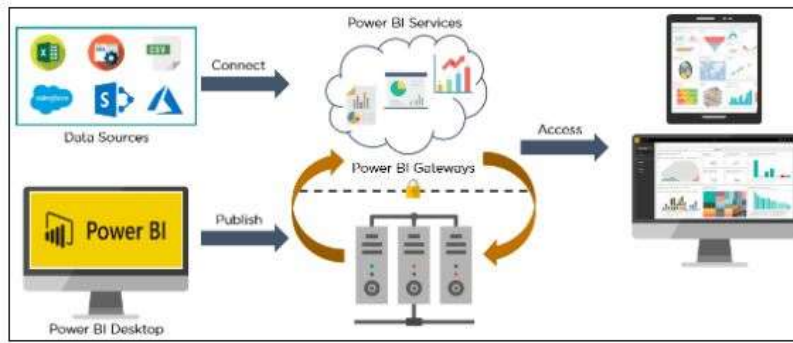
How does Power BI work?

Microsoft Power BI is a suite of tools that supports the process of data processing - from sourcing to suggesting optimal decision.

- Files (e.g. excel .xlsx, .pdf, .csv, .xml, .json)
- Data Bases (e.g. MS SQL, Oracle, SAP HANA, PostgreSQL, MySQL, Amazon Redshift, Snowflake, others)
- Folders (cloud-based e.g. SharePoint and locally-based)
- Azure Services (Azure SQL Database, Azure Blob Storage, Azure Data Lake, others)
- On-line Services (e.g. Microsoft Exchange, Dynamics 365, Salesforce Objects, Google Analytics)
- Other Sources (e.g. websites, OData feed, ODBC, Active Direcory, others)
- Indirect connection via other services (e.g. Microsoft Azure Data Factory)

The data is processed and transformed with Power BI queries (M language) to the so-called data model. Then, with the use of DAX formulas, the data objects formulas, the data objects (e.g. KPI's, targets, etc.) that need to be part of a Power BI report are created (measures, calculated columns, parameters).
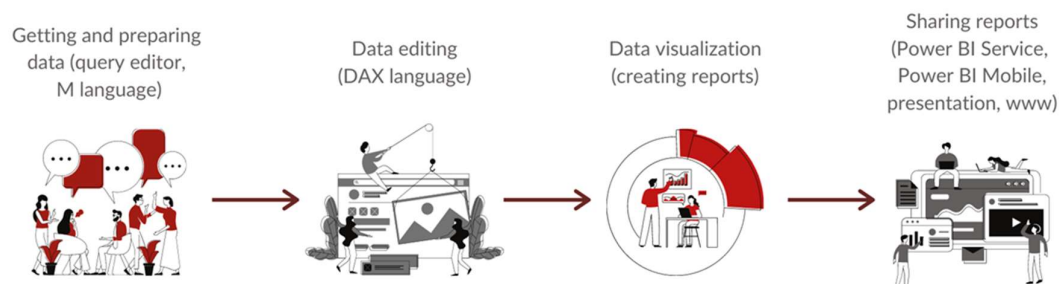
Advantages

There are a few reasons behind the growing popularity of Power BI:

- Reporting Process automation
- Connecting different sources of data in one data model (one source of truth)
- Interactive, multidimensional reports, possibility to track correlation
- Quick and easy analysis
- Standardization and reliability (lower risk of "human error")
- Collaboration and safe report sharing with any member of the organization

For effective deployment of Power BI in your organization, the right strategy (tailored to the organization's size, number of users, and reporting functions) should be chosen.

3.2 Optimization

1. 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

2. 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.

o   Remove unneeded dimensions from the detail shelf.

o   Try displaying your data in different types of views.

3.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 including 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-ofdate filters) can take advantage of the indexing properties in your database and are faster than discrete data 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).

Performance

Investment analytics determines the historic all the previous dataand it should be as accurate as possible. So that it will not mislead to the future investor. Also, model retraining is very important to improve the performance.

Security

Since the investment care analysis consists of years data, the information should be secured.

Reusability

The code written and the components used should have the ability to be reused with no problems.

Resource utilization

When any task is performed, it will likely use all the processing power available until that function is finished.

4 KPI

Dashboards will be implemented to display and indicate certain KPIs and relevant indicators for the Investment.

As and when the system starts to capture the historical/periodic data for a Year, 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 Investment Analysis and its relationship with different metrics

1. Month wise Sales trends
2. Year Wise sales trends

3. Year Month wise sales trends

5 Deployment

Power BI prioritizes choice in flexibility to fit, rather than dictate, your enterprise architecture. Tableau leverage your existing technology investments and integrate them 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 Tableau to match your requirements.

The Dashboard is published on Tableau and an auto-refresh mode has been set so that the dashboard keeps on updating as the real time data loads into the log file.

The Dashboard showcases the multiple insights that has been drawn from the log files as follows:

1. **month wise-**This dashboard shows the sales trend over month
2. **Year wise:** This dashboard shows the sales trend over years.
3. **Year month wise: -**This dashboard shows the sales trend over year months.