High Level Design (HLD)

Amazon Sales Data Analysis

Revision Number: 1.0
Last date of revision: 30/09/2023

Ankita Gupta
Sanjeev Kumar

Document Version Control

Date Issued	Version	Description	Author
25th July 2023	1.0	First Version of Complete HLD	Ankita Gupta & Sanjeev Kumar

Contents

Abstract

In today's fiercely competitive business landscape, effective sales management is pivotal for cost reduction, profit optimization, and maintaining a competitive edge. This project delves into the analysis of Amazon's sales data, seeking valuable insights for future business enhancements. It will spotlight areas of potential revenue loss and underlying causes. Leveraging Microsoft Power BI, we aim to provide a visually intuitive representation of the data, facilitating the identification and resolution of issues, ultimately driving improved decision-making and business performance.

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:
    - 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

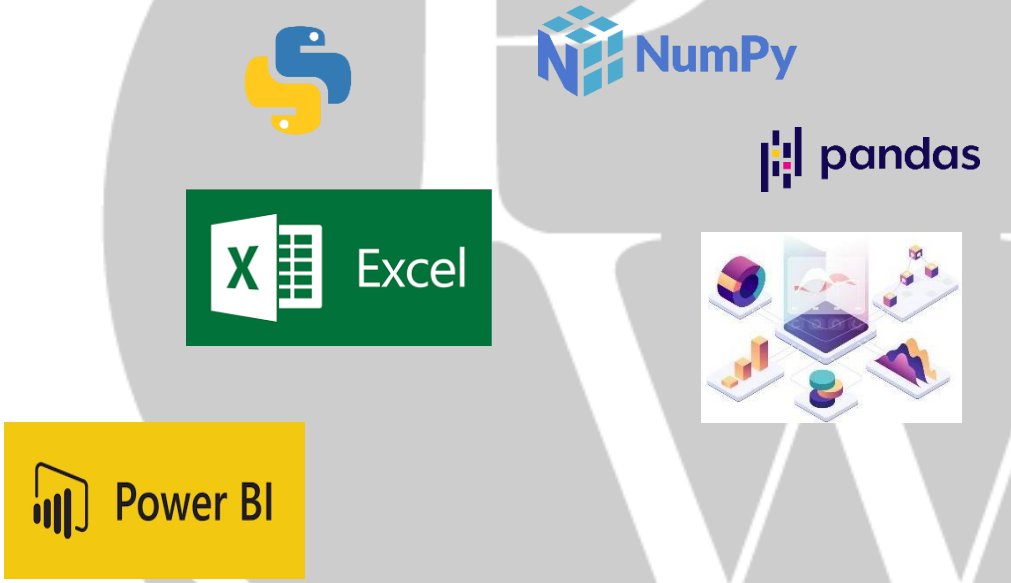
2.1 Product Perspective & Problem Statement

The objective of the project is to Analyze Amazon Sales data to get a substantial data which will help in bringing changes in a business in the future. It will help to reveals flaws in the business model or in the way that one is going about conducting business. Sellers will be able to clearly see where they're losing money, what the problem is, and reduce their losses accordingly.

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.

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.



3 Design Details

3.1 Functional Architecture

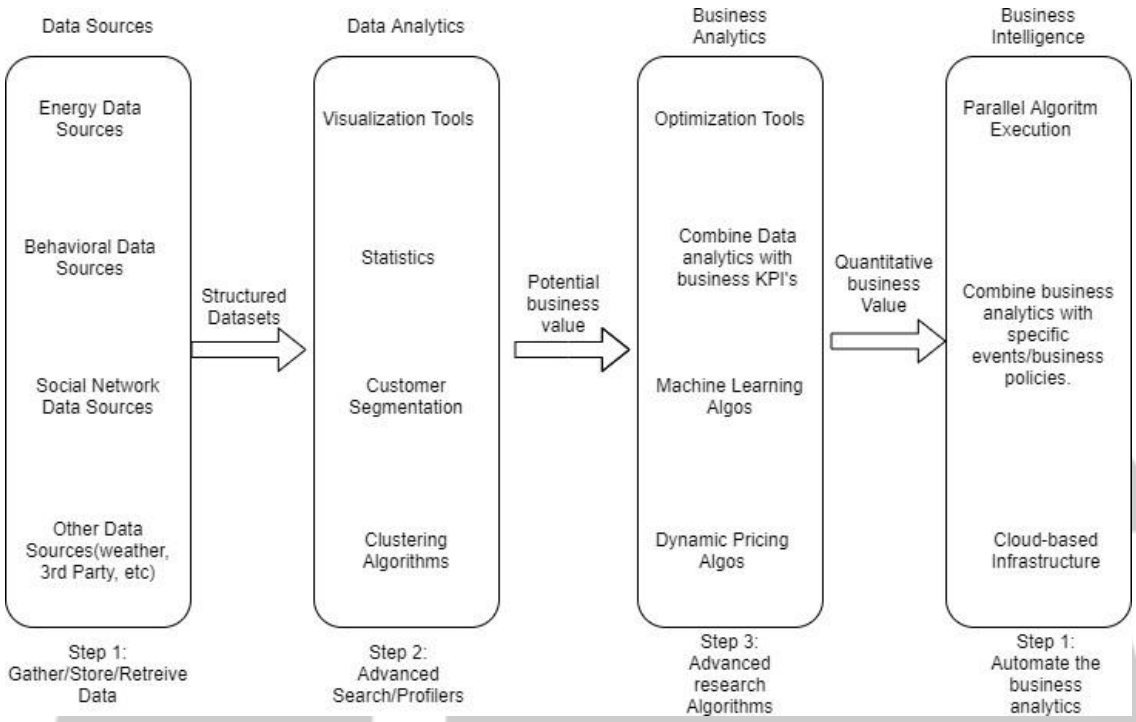
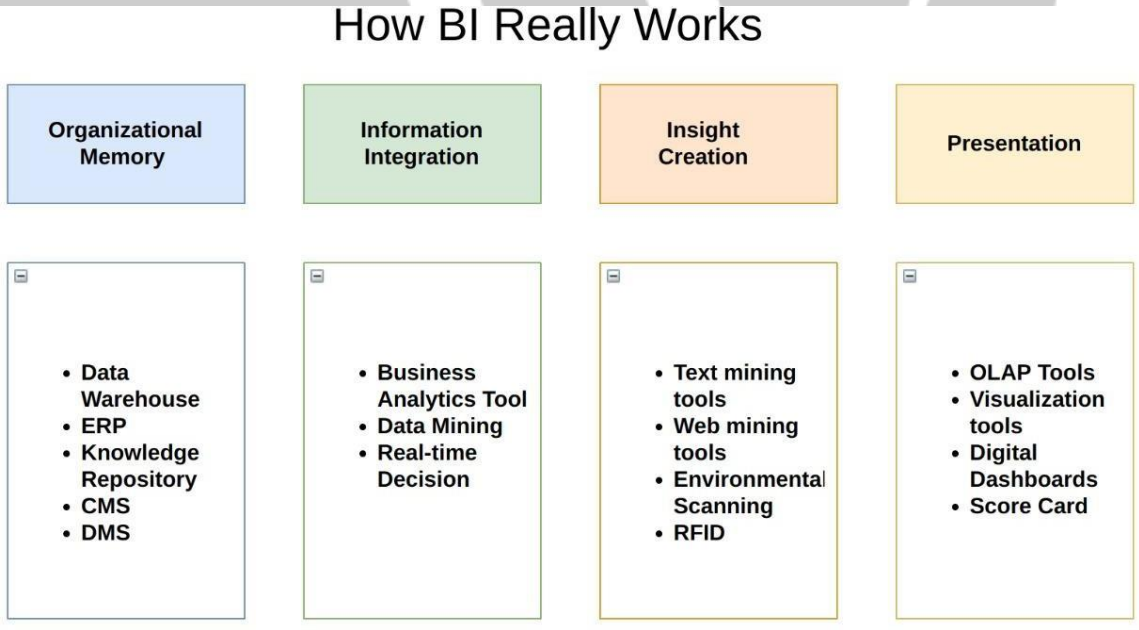


Figure 1: Functional Architecture of Business Intelligence



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.

- Make groups with calculations. Like include filters, calculated groups load only named members of the domain, whereas Tableau's 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 Housing Price and its relationship with different metrics

1. Analyse the Total Revenue of the products.
2. The total profit exceeds the goal by $41.82K, standing at $2.36 million, showcasing a positive performance trend.
3. We have achieved the highest revenue in the Africa region.
4. In 2012, we achieved our highest level of profit, signifying a significant achievement in our overall performance.
5. In the year 2012 stands out as a record-breaking period, with our highest revenue attainment i,e 32 Million

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 offers a range of deployment options, designed to align with your existing enterprise architecture. Whether you choose Power BI Desktop, Power BI Service, or Power BI Report Server, our solutions seamlessly integrate into your IT environment. This ensures that your users can access a self-service, cutting-edge analytics platform that adapts to your specific needs. With on-premises, cloud-based, and hybrid deployment choices, Power BI provides the flexibility to meet your diverse requirements. Below, we compare the three deployment types

TYPE PROS CONS

Power BI - On Premises

- Full control over hardware and software.
- Data and infrastructure are maintained within your organization's network.
- Enhanced security measures can be implemented.
- Ideal for organizations with strict data privacy and compliance requirements.

Power BI - Public Cloud (IaaS)

- Full control of software on managed hardware.
- Positions infrastructure in the same location as data, facilitating cloud migration.
- Flexibility to adjust hardware resources as required.
- Requires dedicated administrators for software management.
- Additional infrastructure is required to enable off-network access, such as for mobile and external use.

Power BI Online (SaaS)

- Fully hosted solution with hardware and software upgrades included
- Quick deployment process.
- User-friendly access for external audiences.
- Operates within a multi-tenant environment with a single-site setup.
- Cubes are not supported in this configuration.
- Guest account access is not available.

Depending on your organizational roles and responsibilities, Power BI deployment should be managed by a systems administrator and the designated Power BI Administrator, working in coordination with relevant IT personnel. For Power BI Online, integration with your existing technology and configuration of site settings are essential steps. The Data & Analytics Survey, conducted by business teams, helps identify and prioritize data use cases, determine audience size, and identify users. This collected information is then used to formulate your deployment strategy, which includes planning for sizing, installation, and configuration of Power BI, or integration and configuration of Power BI Online. Additionally, administrators should plan for the installation of client software such as Power BI Desktop, Power BI Mobile, and Power BI Gateway where applicable.