

AMAZON SALES ANALYSIS



VERSION: 1.0
DATE 14-1-2023

RAVIKUMAR SURAM

DOCUMENT VERSION CONTROL

AMAZON SALES ANALYSIS BUSINESS INTELLIGENCE PROJECT

DATE	VERSION	AUTHOR	CHANGE
14-1-2023	1.0	RAVIKUMAR SURAM	FIRST VERSION OF HLD

Contents:

Document Version Control

Abstract.....	4
1. Introduction.....	5
1.1 Why this High-Level Design Document?.....	5
1.2 Scope.....	5
2. General Descriptipns.....	6
2.1 Product Perspective & Problem Statement.....	6
2.2 Tools used.....	6
3. Design Details.....	7
3.1 Functional Architecture	7
3.2 Optimization.....	9
4. KPIs	10
KPIs (Key Performance Indicators)	11
5. Deployment.....	11

Abstract:

E-Commerce business modules are growing irrespective of the state and country barriers. At the same time it's difficult to maintain the business profitable quarter-on-quarter and year-on-year due to several reasons.

Sales analysis is the important part of the company's growth. Selling the products and making profit out of it, is totally depend on reducing the cost of manufacturing and improved methods of distribution.

Timely analysing the sales and trend is also important key to run the business successfully and predicting the future requirements and preparing for the same is required to run the business smoothly.

Retaining the customers is also a big challenge in the business and for that service to the customer and gaining the trust is the key for success.

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.

2 General Descriptions

2.1 Product Perspective & Problem Statement

The goal of this project is to analyse the sales trend of Amazon in various countries and present it to the Sales management. So that they can take the necessary steps to improve the methods of distribution to reduce cost and to increase profits. The data is provided by the iNeuron. To do EDA and to extract the meaningful insights from the data based on the given information.

2.2 Tools used

To bring meaningful insights from the data we are going to use Python and its libraries such as NumPy, Pandas, Seaborn, Matplotlib to do EDA (exploratory data analysis). Jupyter Notebook (Python Programming Language) is used to build the whole framework. Business Intelligence tool like MS-PowerBI is used to present the data in self-exploratory and interactive visuals.



3 Design Details

3.1 Functional Architecture

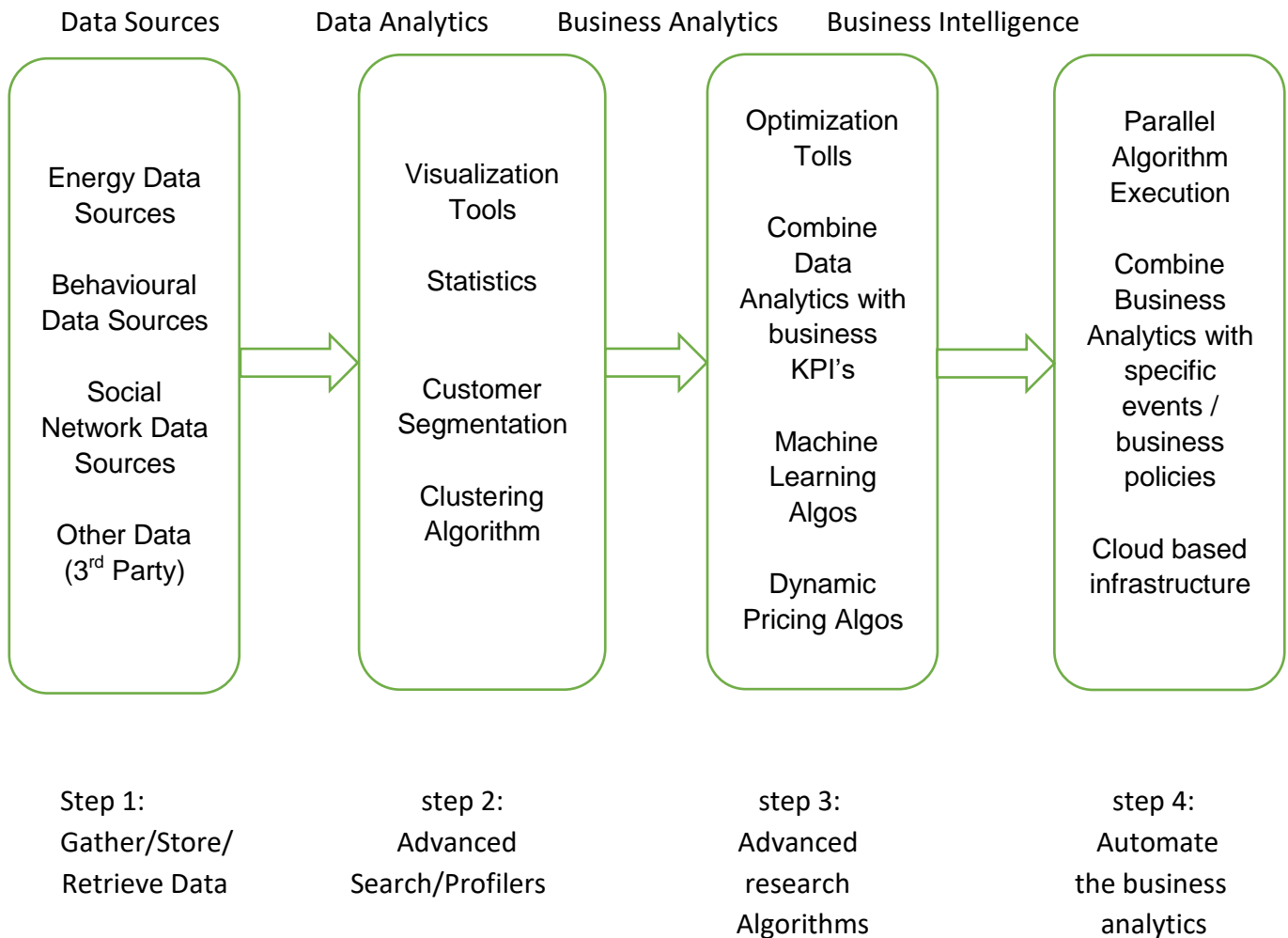


Figure 1: Functional Architecture of Business Intelligence

HOW BI REALLY WORKS

Organisation al Memory

1. Data Warehouse
2. ERP
3. Knowledge Repository
4. CMS
5. DMS

Information Integration

1. Business Analytics
2. Data Mining
3. Real-Time Decision

Insight Creation

1. Text Mining Tools
2. Web mining Tools
3. RFID

Presentation

1. OLAP Tools
2. Data Visualization
3. Digital Dashboard
4. Score card

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.
 - o LODs - Look at the number of unique dimension members in the calculation.
 - o 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>Date Time>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 used in this project are

1. Sales by Year
2. Sales by Month
3. Sales by Year-Month
4. Top selling regions
5. Top selling items

5 Deployments

Prioritizing data and analytics couldn't come at a better time. Your company, no matter what size, is already collecting data and most likely analysing 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 Visualization at scale, as well as organizing, orchestrating, and unifying disparate sources of data for business users and experts alike to author and consume content.

- **Patterns in business operations:** Data visualization techniques help us to determine the patterns of business operations. By understanding the problem statement and identifying the solutions in terms of patterning and applied to eliminate one or more of the inherent problems.
- **Identify business trends and relate to data:** These techniques help us identify market trends by collecting the data on Day-To-Day business activities and preparing trend reports, which helps track the business how influences the market. So that we could understand the competitors and customers. Certainly, this helps to long-term perspective.
- **Storytelling and Decision making:** Knowledge of storytelling from available data is one of the niche skills for business communication, specifically for the Data Science domain which is playing a vital role. Using best visualization this role can be enhanced much better way and reaching the objectives of business problems.
- **Understand the current business insights and setting the goals:** Businesses can understand the insight of the business KPIs, finding tangible goals and business strategy planning; therefore they could optimize the data for business strategy plans for on-going activities.
- **Operational and Performance analysis:** Increase the productivity with the help of visualization techniques the clarity of KPIs depicting the trends of the productivity of the manufacturing unit, and guiding were to improve the productivity of the plant.

Power BI Server - On Premises

- Full control of hardware and software
- Infrastructure and data remain behind your firewall
- Need dedicated administrators to manage hardware and software
- Additional infrastructure needed to access off-network (mobile, external)

Key components

Key components of the Power BI ecosystem comprise:

- **Power BI Desktop**
The Windows-desktop-based application for PCs and desktops, primarily for designing and publishing reports to the Service.
- **Power BI Service**
The SaaS (software as a service) based online service. This was formerly known as Power BI for Office 365, now referred to as PowerBI.com, or simply Power BI.
- **Power BI Mobile Apps**
The Power BI Mobile apps for Android and iOS devices, as well as for Windows phones and tablets.
- **Power BI Gateway**
Gateways used to sync external data in and out of Power BI and are required for automated refreshes. In Enterprise mode, can also be used by Flows and Power Apps in Office 365.
- **Power BI Embedded**
Power BI REST API can be used to build dashboards and reports into the custom applications that serve Power BI users, as well as non-Power BI users.

- **Power BI Report Server**

On-premises Power BI reporting solution for companies that won't or can't store data in the cloud-based Power BI Service.

- **Power BI Premium**

Capacity-based offering that includes flexibility to publish reports broadly across an enterprise, without requiring recipients to be licensed individually per user. Greater scale and performance than the shared capacity in the Power BI service.

- **Power BI Visuals Marketplace**

The marketplace of custom visuals and R-powered visuals.

