High Level Design (HLD)

Amazon Sales Data Analysis

Revision Number: 1.0

Last date of revision: 04/08/2023

Szimonetta Farkas

iNeuron

Document Version Control

Date Issued	Version	Description	Author
04 August 2023	1.0	First Version of Complete HLD	Szimonetta Farkas

Contents

Abstract

Amazon is an American multinational technology company founded by Jeff Bezos. Amazon has country – specific websites, they sell many product lines, including media, apparel, baby products, consumer electronics, beauty products, gourmet food, groceries, health and personal care products, industrial & scientific supplies, kitchen items, jewelry, watches, lawn and garden items, musical instruments, sporting goods, tools, automotive items, toys and games, and farm supplies and consulting services (1).

The revenue and profit of Amazon increase year by year. The annual revenue for 2022 was $513.983B (2).

Analyzing the given Amazon sales dataset provides us useful insights and helps us to make decisions to reduce cost and increase profit.

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

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.

Do ETL: Extract-Transform-Load some Amazon dataset and find for me Sales-trend -> month wise, year wise, yearly month wise.

Find key metrics and factors and show the meaningful relationships between attributes.

2.2 Tools used

- Excel: the given data was in excel file with impurities
- Python, Pandas, Numpy
    * general view of the raw dataset
    * data cleaning: filling / deleting missing values
    * creating new columns: year, month, day
    * making the dataset ready to build visuals
    * checking the correlation between the features
    * saving the cleaned dataset to excel file

- Power BI

    * importing data, creating visualizations, saving the file in pbix version

3 Design Details

3.1 Functional Architecture

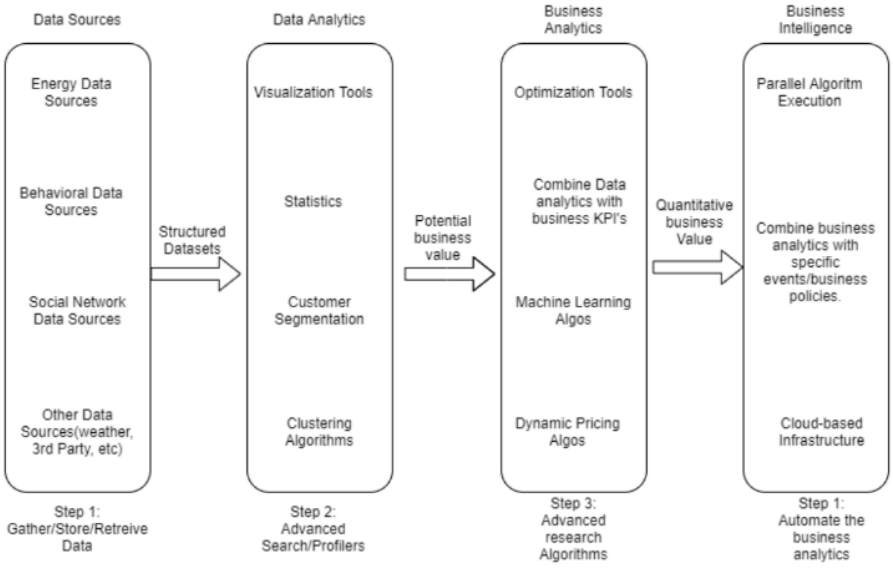
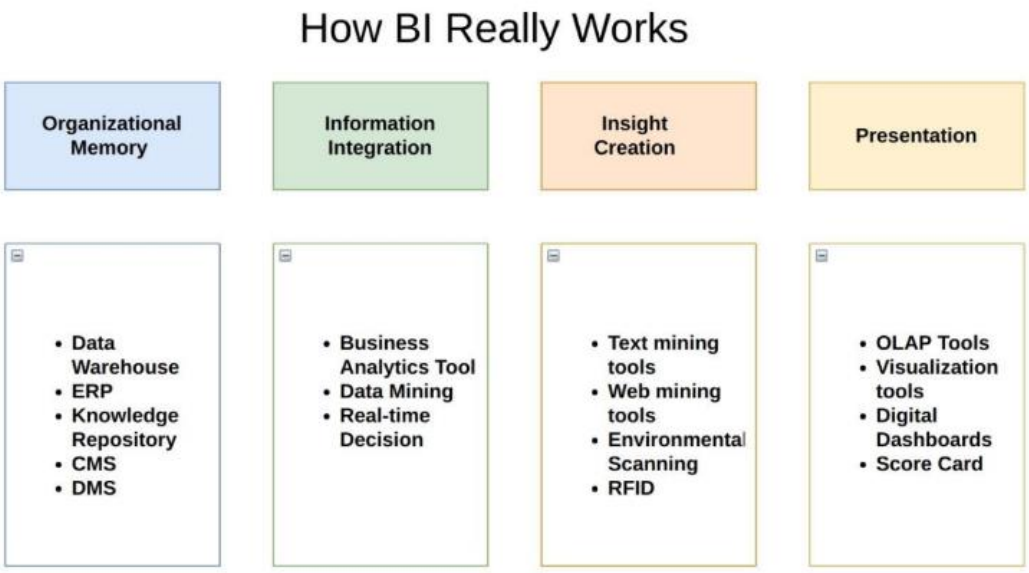


Figure 1: Functional Architecture of Business Intelligence



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.

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>DateTime>String

4 KPIs

Dashboards will be implemented to display and indicate certain KPIs and relevant indicators for the sales and cost. 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 Amazon Sales Data and its relationship with different metrics.

1. Overall Sales
2. Overall Margin
3. Overall Cost
4. Monthly Trend for Sales, Margin and Cost
5. Yearly Trend for Sales, Margin and Cost
6. Frequently Ordered Items, High Margin Items
7. Relationship between Sales-Cost, Sales-Margin, Margin – Cost

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.

Amazon Sales data is cleaned using ETL, Python, Pandas and Numpy tools. It is analyzed with Power BI and key insights are provided to tell a story about the data.

6 References

(1) Amazon (company) - Wikipedia

(2) Amazon Revenue 2010-2023 | AMZN | MacroTrends