Amazon Sales Data Analysis

High Level Design (HLD)

Vishal R Nivangune

Revision Number: 1.0

Last date of revision: 10/10/2023

Document Version Control

Date Issued	Version	Description	Author
31-01-2023	1.0	First version of Complete HLD Documentation	Vishal Nivangune

Contents

Abstract

The "Amazon Sales Analysis" project aims to empower sellers with valuable insights into their sales trends on a year, month, and quarter basis. By harnessing the power of data analytics, this project seeks to provide sellers with a comprehensive understanding of their product performance, enabling them to make informed decisions and better cater to customer preferences.

Our project employs a multi-faceted approach to achieve this goal. Beginning with data cleaning and preparation in Excel, we ensure the data is accurate and ready for analysis. The cleaned data is then stored in AWS S3 (a cloud storage platform) for efficient data management.

Derive meaningful insights, we utilize Snowflake, a robust data warehousing platform, to perform data transformations and manipulations as needed. This step is crucial in extracting relevant information and patterns from the extensive dataset, which encompasses profit, revenue, cost, units sold, and more across various regions and countries.

The centrepiece of this project lies in the integration of Power BI with Snowflake. Through direct querying, we seamlessly bring the modified data into Power BI, where we craft dynamic dashboards. These dashboards offer sellers an intuitive and interactive interface to explore their sales trends and product performance, all in real-time.

The impact of this project is profound. Sellers gain the ability to identify their best-performing products promptly, thanks to the insights presented in the Power BI dashboard. Armed with this knowledge, they can strategize effectively, optimize their product offerings, and respond agilely to changing market dynamics. This project empowers sellers to enhance their decision-making process, resulting in improved product and service quality.

In an era marked by technological advancement, the e-commerce industry relies heavily on data analytics to remain competitive. "Amazon Sales Analysis" exemplifies the potential of data analytics in transforming businesses, highlighting the importance of understanding market trends, customer behaviours, and evolving customer preferences. This project serves as a testament to the value of data-driven decision-making in the modern business landscape, and its outcomes hold the promise of delivering greater customer satisfaction and innovation.

1. Introduction

1.1 Purpose of the High-Level Design Document

The High-Level Design (HLD) Document serves a crucial role in our project by providing the necessary depth to the existing project description. Its primary function is to outline a comprehensive model for coding. Moreover, this document acts as an invaluable tool for identifying potential inconsistencies before the coding phase begins, while also serving as a reference manual to illustrate high-level module interactions.

The HLD Document will accomplish the following objectives:

1. **Comprehensive Design Presentation:** It will present an exhaustive overview of all design aspects, offering detailed definitions.
2. **User Interface Description:** We will elucidate the user interface being integrated into the project.
3. **Hardware and Software Interfaces:** The document will delve into the interfaces related to hardware and software components.
4. **Performance Requirements:** It will outline the performance expectations and criteria.
5. **Design Features and Architecture:** The HLD will encapsulate design features and the architectural framework of the project.
6. **Non-Functional Attributes:** These critical attributes encompass:
    - **Security**
    - **Reliability**
    - **Maintainability**
    - **Portability**
    - **Reusability**
    - **Application Compatibility**
    - **Resource Utilization**
    - **Serviceability**

1.2 Scope

The High-Level Design (HLD) documentation covers:

1. **Database Architecture:** How data is structured and managed, explained in non-technical terms.
2. **Application Architecture (Layers):** The separation of components and design patterns, made accessible to non-technical administrators.
3. **Application Flow (Navigation):** Illustrating user journeys and interactions in user-friendly language.
4. **Technology Architecture:** An overview of tech choices and their impact, avoiding complex jargon.

The HLD aims to be administrator-friendly, providing clear insights into system design without requiring deep technical knowledge.

2. General Description

2.1 Product Perspective & Problem Statement:

1. Product Perspective:

Our project is situated within the broader context of the e-commerce industry, where data analytics plays a pivotal role in decision-making. As businesses increasingly rely on data-driven insights to enhance their performance, our project focuses on Amazon sales data analysis, offering valuable tools for sellers to understand their product sales trends.

2. Problem Statement:

Sales management is of paramount importance in the face of growing competition and the need for efficient distribution methods. In this context, our project addresses the need for comprehensive sales data analysis, providing a solution to the following challenges:

- **Limited Visibility:** Sellers often lack a holistic view of their product sales trends over different periods, hindering their ability to make informed decisions.
- **Data Complexity:** Managing and analysing large datasets, such as Amazon sales data, can be overwhelming and require significant technical expertise.
- **Informed Decision-Making:** To optimize sales strategies and product offerings, sellers need access to meaningful insights derived from their sales data.

Our project aims to resolve these challenges by offering a user-friendly solution for sales data analysis, allowing sellers to gain deeper insights into their product performance on a yearly, monthly, and quarterly basis.

2.2 Tools Used

1. Microsoft Excel
2. AWS S3 (Simple Storage Service)
3. Snowflake
4. Power BI
5. SQL (Structured Query Language)
6. Documentation Tools (Microsoft Word)



3. Design Details

The design details refer to the specific architectural and functional aspects of the tools and technologies used. These details provide an understanding of how each tool contributes to the project's overall design and functionality.

3.1 Functional Architecture

The functional architecture of the "Amazon Sales Data Analysis" project is supported by the following core tools:

1. Microsoft Excel:

- Microsoft Excel serves as the initial data processing and cleaning tool in the project. It is used to:
    - Remove duplicate records: Excel's built-in features are employed to identify and eliminate duplicate data entries within the Amazon sales dataset, ensuring data accuracy.
    - Manage missing data: Excel's functions are utilized to manage missing or incomplete data by either filling in missing values or making decisions regarding how to handle them.
    - Format data: Excel's formatting capabilities are used to ensure consistency in data presentation, making it suitable for further analysis.

2. AWS S3 (Simple Storage Service):

- AWS S3 acts as the cloud-based repository for storing the cleaned and processed Amazon sales dataset. Its role is to:
    - Ensure data availability: AWS S3 ensures that the dataset is always readily accessible to authorized users and tools.
    - Guarantee data durability: It provides data durability and redundancy, safeguarding against data loss or corruption.
    - Support scalability: AWS S3's scalability allows for the storage of large datasets as the project evolves.

3. Snowflake:

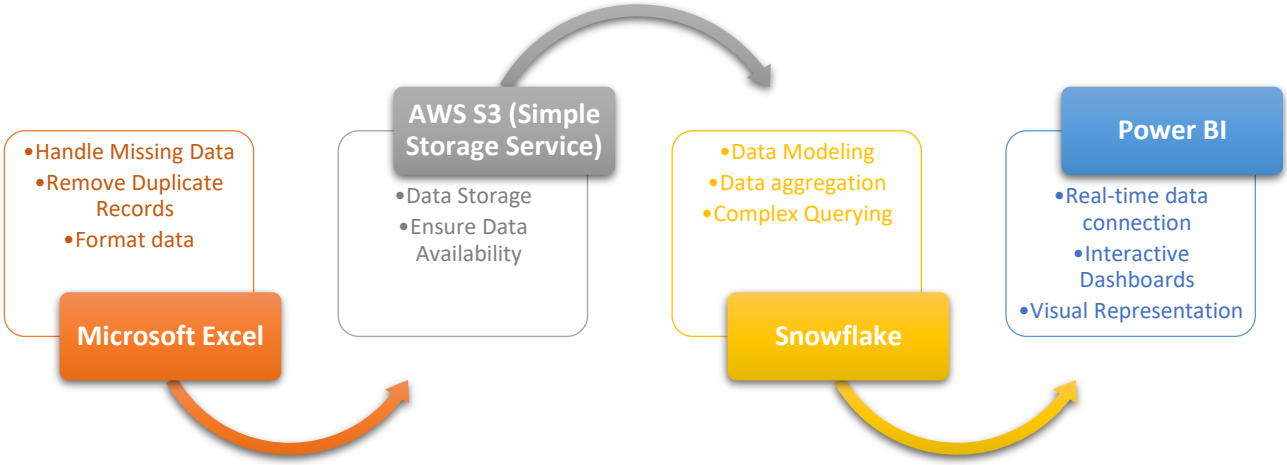
- Snowflake is the core data warehousing platform used for advanced data transformation and manipulation. Its functions include:
    - Data modeling: Snowflake enables the creation of data models that facilitate efficient querying and data organization.
    - Data aggregation: It supports the aggregation of data to calculate metrics, such as sales trends, year-wise and month-wise.
    - Complex querying: Snowflake allows for complex SQL queries to extract meaningful insights from the dataset, facilitating in-depth analysis.

4. Power BI:

- Power BI is the tool for data visualization and dashboard creation. Its primary functions are:
    - Real-time data connection: Power BI connects directly to Snowflake, ensuring that dashboards reflect up-to-date sales data.
    - Interactive dashboards: It provides a user-friendly interface for exploring sales trends, metrics, and relationships within the data.
    - Visual representation: Power BI's visualizations, including charts, graphs, and tables, are used to present sales data in a comprehensible and actionable manner.

These tools collectively form the functional architecture of the project, each playing a distinct role in processing, storing, transforming, and presenting Amazon sales data to derive valuable insights for sellers.

Functional Architecture design of Amazon Sales Analysis



3.2 Optimization

Optimizing the performance of our "Amazon Sales Data Analysis" project is a critical aspect of ensuring its efficiency and effectiveness. Our data strategy is at the core of this optimization effort, aiming to drive performance improvements. The following strategies and best practices guide our optimization approach:

- **Minimize Fields and Records:** To enhance data processing efficiency, we minimize the number of fields and records within our dataset.
- **Optimize Extracts:** We focus on optimizing extracts to speed up future queries by materializing calculations, removing unnecessary columns, and utilizing accelerated views.
- **Reduce Data Points:** We aim to reduce data points in our views by practicing guided analytics, avoiding overloading single views, and connecting related views with action filters to enable seamless navigation from overview to granular views.
- **Streamline Filters:** We reduce the complexity of queries by removing unneeded dimensions from the detail shelf and limiting the number and type of filters in use. We also prefer include filters for faster performance and use continuous date filters and Boolean or numeric filters to leverage database indexing properties.
- **Optimize Calculations:** We optimize and materialize calculations by performing them in the database, reducing the number of nested calculations, and considering the granularity of Level of Detail (LOD) and table calculations in the view. Additionally, we use MIN or MAX instead of AVG where applicable to expedite processing.

These optimization strategies are integral to ensuring that our project delivers high performance and efficiency while providing valuable insights to our users and stakeholders.

4. KPIs (Key Performance Indicators)

In our "Amazon Sales Data Analysis" project, Key Performance Indicators (KPIs) are pivotal for tracking and visualizing essential sales data. These KPIs offer a concise overview of our sales performance and relationships with various metrics, guiding strategic decision-making and highlighting critical aspects of our sales operations.

4.1 KPIs (Key Performance Indicators)

1. **Total Profit Distribution:** Displays yearly, quarterly, and monthly profit distributions, providing insights into profit trends over different timeframes.
2. **Top Performing Items:** Identifies items that generated the highest revenue and profit, enabling a focus on successful products.
3. **Top 5 Profitable Items:** Highlights the top 5 items generating the highest profit in terms of percentage, aiding in identifying high-margin products.
4. **Regional Profit Analysis:** Examines total profit across various regions and countries, facilitating regional performance assessment.
5. **Revenue Contribution per Item:** Illustrates the contribution of each item to the total revenue, indicating the significance of individual products.
6. **Item-specific Revenue and Cost:** Compares total revenue and total cost by item type, providing insights into profitability per product category.
7. **Year-over-Year Revenue:** Tracks revenue generated year over year, offering a perspective on overall sales growth.
8. **Last Two Years Revenue Distribution:** Compares revenue distribution and profit percentages from the last two years, aiding in trend analysis and forecasting.
9. **Total Units Sold:** Provides an overview of the total units sold across all years, indicating the overall sales volume.

These KPIs will be visualized through interactive dashboards, empowering our team to make data-driven decisions, analyse sales patterns, and enhance our strategies based on comprehensive and real-time insights.

5. Deployment

In the context of our "Amazon Sales Data Analysis" project, deploying our data analytics solution is a pivotal step to extract maximum value from the insights we derive. It is worth noting that prioritizing data and analytics has become increasingly important in today's business landscape. Enterprises of all sizes are not only collecting vast amounts of data but also harnessing analytics to address business challenges, gain a competitive edge, and foster transformative changes.

Our deployment strategy revolves around two core components, Power BI for data visualization and Snowflake for query optimization. By deploying Power BI, we enable self-service data exploration and visualization, empowering our users to interact with and comprehend the data in a user-friendly manner. Power BI provides the tools to create interactive dashboards and reports, enhancing our ability to make informed decisions based on real-time insights.

Concurrently, the integration of Snowflake into our deployment strategy ensures that the data is optimized for efficient querying and analysis. Snowflake's data warehousing capabilities orchestrate and unify disparate data sources, creating a robust foundation for our business users and experts to author and consume content. By operating Snowflake at scale, we aim to provide a high-performance, scalable, and reliable data querying environment that supports our analytical endeavors.

In summary, our deployment approach leverages the strengths of Power BI and Snowflake to provide a comprehensive and user-centric solution for data visualization and analysis. This facilitates data-driven decision-making, improves our understanding of sales trends, and empowers our project to deliver valuable insights for stakeholders and users alike.