Amazon Sales Data Analysis

Low-Level Design Document

Vishal R Nivangune

VERSION:1.0
DATE 11/10/2023

Document Version Control

Date Issued	Version	Description	Author
11-10-2023	1.0	First Version of complete LLD	Vishal R NIvangune

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.

Table of Contents

1. Project Introduction

In today's highly competitive business landscape, effective sales management plays a pivotal role in the success and sustainability of commercial enterprises. With the ever-evolving nature of e-commerce and the constant quest for improved distribution methods, the need for data-driven insights has become more pronounced than ever. This holds particularly true for e-commerce giants like Amazon, where intricate sales data analysis can unlock the door to reduced costs and amplified profits.

The primary objective of this project is to delve into the vast realm of Amazon's sales data. Leveraging the principles of ETL (Extract-Transform-Load), we will embark on a journey to unveil the underlying trends and patterns that govern Amazon's sales dynamics. By meticulously extracting data from diverse sources, transforming it into a more structured and comprehensible format, and loading it into our analytical ecosystem, we aim to uncover invaluable insights that not only reflect the past but also shape the future of Amazon's sales strategy.

1.1 What is a Low-Level Design Document (LLD)?

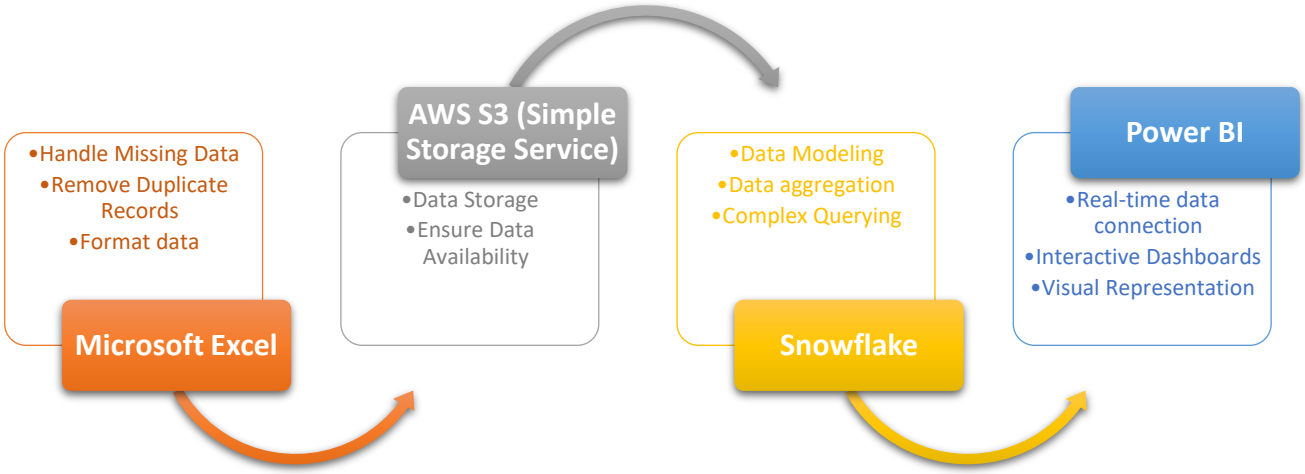
A Low-Level Design Document (LLD) is a detailed technical document that provides a granular explanation of how a specific component or system within a project is designed and implemented. It delves into the nitty-gritty technical details, including data structures, algorithms, code architecture, and other implementation specifics. The LLD is a critical step in the software development process, as it bridges the gap between high-level requirements and actual implementation, offering developers a blueprint to follow. In the context of the Amazon Sales Data Analysis project, an LLD might focus on the technical details of data extraction, transformation, and loading processes, database design, analytics algorithms, and other technical aspects involved in the project's implementation.

1.2 Scope in Terms of Amazon Sales Data Analysis

In the context of Amazon Sales Data Analysis, the scope of a Low-Level Design Document (LLD) would revolve around the technical intricacies of the data analysis and processing components of the project.

The LLD would define and detail the following:

- Data Extraction
- Data Transformation
- Data Loading
- Data Analysis Algorithms
- Database Design
- Analytics Pipeline
- Performance Optimization
- Security Considerations



2. Functional Architecture

We provide a detailed breakdown of the functional architecture and the core tools that support its execution. Here's a brief overview:

1. Microsoft Excel:

- **Role:** Initial data processing and cleaning tool.
- **Functions:**
    - Remove duplicate records.
    - Manage missing data.
    - Format data for consistency.

2. AWS S3 (Simple Storage Service):

- **Role:** Cloud-based data repository.
- **Functions:**
    - Ensure data availability.
    - Guarantee data durability and redundancy.
    - Support scalability for storage of large datasets.

3. Snowflake:

- **Role:** Core data warehousing platform.
- **Functions:**
    - Data modelling for efficient querying and data organization.
    - Data aggregation for metrics calculation.
    - Complex querying for in-depth analysis.

4. Power BI:

- **Role:** Data visualization and dashboard creation.
- **Functions:**
    - Real-time data connection to Snowflake for up-to-date insights.
    - Creation of interactive dashboards for exploring sales data.
    - Visual representation through charts and graphs.

These tools collectively form the functional architecture of the project. Microsoft Excel serves as the initial data processing and cleaning stage, ensuring data accuracy. AWS S3 acts as a robust and scalable data repository, while Snowflake handles advanced data transformation and querying. Power BI brings it all together by providing user-friendly and visually engaging dashboards for actionable insights. This architectural framework enables the comprehensive analysis of Amazon sales data, benefiting sellers in making informed decisions.

3. Architecture Description

3.1 Data Sourcing:

- The dataset is named "Amazon Sales Data Analysis.csv."
- It is available in CSV (Comma Separated Values) format.
- Microsoft Excel is the tool used to load and process the data.
- The dataset is publicly available for research purposes.
- Source: GitHub.

3.2 Data Overview:

- Dataset size: 12.4 KB.
- Format: A single file in ".csv" format.
- Number of records (rows): 100.
- Number of attributes (columns): 14.

3.3 Data Description: The dataset comprises the following attributes:

1. **Region**
    - Description: Name of the Region.
    - Datatype: String.
2. **Country**
    - Description: Name of different Countries.
    - Datatype: String.
3. **Item Type**
    - Description: Different Product type sales on Amazon.
    - Datatype: String.
4. **Sales Channel**
    - Description: Mode of shopping, either online or offline.
    - Datatype: String.
5. **Order Priority**
    - Description: Priority of Sales, ranging from low to high.
    - Datatype: String.
6. **Order Date**
    - Description: Date of the Order.
    - Datatype: Date.
7. **Order ID**
    - Description: Order ID for various Products.
    - Datatype: String.
8. **Ship Date**
    - Description: Date when the product is dispatched.
    - Datatype: Date.
9. **Unit Sold**
    - Description: Number of units sold per product.
    - Datatype: Integer.
10. **Unit Price**
    - Description: Unit Price of the Product.
    - Datatype: Integer.

11. **Unit Cost**
    - Description: Unit cost of the Product.
    - Datatype: String.
12. **Total Revenue**
    - Description: Amount incurred after selling different Products.
    - Datatype: Decimal.
13. **Total Cost**
    - Description: Total Cost incurred by the company for making Products.
    - Datatype: Decimal.
14. **Total Profit**
    - Description: Profit earned by the Company after subtracting all expenses from revenue.
    - Datatype: Decimal.

3.4 Data Transformation:

- Describe the data transformation process, explaining that Snowflake is the core data warehousing platform responsible for data modeling, aggregation, and complex querying to prepare the data for analysis.

3.5 Data Storage and Management:

- Discuss the role of AWS S3 (Simple Storage Service) as a cloud-based repository for storing the cleaned and processed Amazon sales dataset. Emphasize its role in ensuring data availability, durability, and scalability.

3.6 Performance Optimization:

- Highlight performance optimization strategies within Snowflake to ensure efficient querying and analysis of the dataset, which may include data indexing and query optimization.

3.7 Security Considerations:

- Discuss the security measures implemented throughout the process, including access controls, encryption, and data anonymization, considering the sensitivity of the data.

This dataset, "Amazon Sales Data Analysis.csv," is a comprehensive collection of sales data attributes that will be analysed for various insights and trends to aid decision-making and profitability assessments. The dataset provides detailed information about sales activities and financial aspects of the business.

4. Deployment in Power BI Services

The deployment process involves:

1. **Creation and Publishing:** Build the dashboard in Power BI Desktop and publish it to Power BI Service.
2. **Data Refresh:** Configure data refresh to keep the dashboard up-to-date.
3. **Security and Access Control:** Implement access controls and protect sensitive data.
4. **Sharing and Collaboration:** Share the dashboard with team members and stakeholders.
5. **Embedding:** Embed the dashboard in websites or applications.
6. **Mobile Deployment:** Make the dashboard accessible on mobile devices.
7. **Scheduled Email Subscriptions:** Enable users to receive scheduled email notifications.
8. **Version Control:** Manage different iterations of the dashboard.
9. **Governance and Monitoring:** Implement governance rules and monitor usage.
10. **Collaboration Tools:** Encourage collaboration through comments and discussions.
11. **Documentation and Training:** Provide user guidance for effective interaction.
12. **Feedback and Iteration:** Continuously gather feedback and improve the dashboard for better decision-making.