HIGH LEVEL DOCUMENT (HLD)

AMAZON SALES DATA ANALYSIS

REVISION NO. : 1.0

LAST DATE OF REVISION : 10/10/2024

HARSH DIXIT

Document Version Control

DATE ISSUED	VERSION	DESCRIPTION	AUTHOR
24 SEPT 2024	1.0	First version of complete HLD	Harsh Dixit

Contents

Abstract

This document presents the high-level design of the "Amazon Sales Analysis" system, aimed at analyzing key sales performance metrics and trends. It outlines the architecture, tools used, and key performance indicators (KPIs) that contribute to understanding the business's overall health. The goal is to provide a functional overview, data structure, and the optimizations made to ensure effective data handling and analysis. The document further describes how the analysis system can be deployed and scaled, delivering business-critical insights to improve sales strategies.

1: Introduction

In today's competitive e-commerce environment, data-driven decision-making is essential for success. The "Amazon Sales Analysis" system was developed to address this need by providing real-time insights into sales performance, customer behaviors, product profitability, and other key business metrics. This design document outlines the architecture of the system, the tools and technologies utilized, and the approach taken to build an efficient, scalable analytics solution.

1.1 Why this High-Level Design Document?

This document serves as a guide for stakeholders, developers, and data analysts to understand the design considerations, implementation strategies, and optimization techniques used in the Amazon Sales Analysis system. It provides a holistic view of the system architecture, ensuring a shared understanding of the objectives, functionality, and outcomes expected from the system.

1.2 Scope

The scope of this project is to provide a comprehensive sales analysis platform for Amazon's marketplace data, including sales by product category, geographical trends, customer demographics, and seasonal variations. The system will cater to various stakeholders, including sales teams, marketing strategists, and operational managers, by offering detailed insights through interactive dashboards and reports.

In-Scope:

•Analysis of historical and real-time sales data.
•Insights into customer behavior and purchasing patterns.
•Visualization of KPIs and metrics across various dimensions.
•Sales forecasting and trend analysis.

Out-of-Scope:

•Data beyond the Amazon marketplace.
•Inventory and logistics optimization.
•Vendor and supplier management details.

2: General Description

2.1 Product Perspective & Problem Statement

The "Amazon Sales Analysis" system integrates Amazon's raw sales data with a data visualization platform, providing stakeholders with a clear and interactive view of the business's sales performance. The main problem addressed by this solution is the lack of a consolidated view of sales data and difficulty in identifying actionable insights from raw data. This system transforms large datasets into valuable insights that can guide decision-making.

3: Tools Used

The following tools and technologies were utilized to build the system:

- **Power BI**:
  For building interactive and visual dashboards.
- **Microsoft Excel**:
  For preliminary data cleaning and manipulation.
- **SQL Server**:
  For managing structured sales data and performing advanced queries.
- **Python**:
  For scripting, data processing, and advanced analysis.
- **Amazon Web Services (AWS)**:
  For cloud storage and computing needs.
- **DAX (Data Analysis Expressions)**:
  For defining custom calculations within Power BI.

4: Design Details

4.1 Functional Architecture

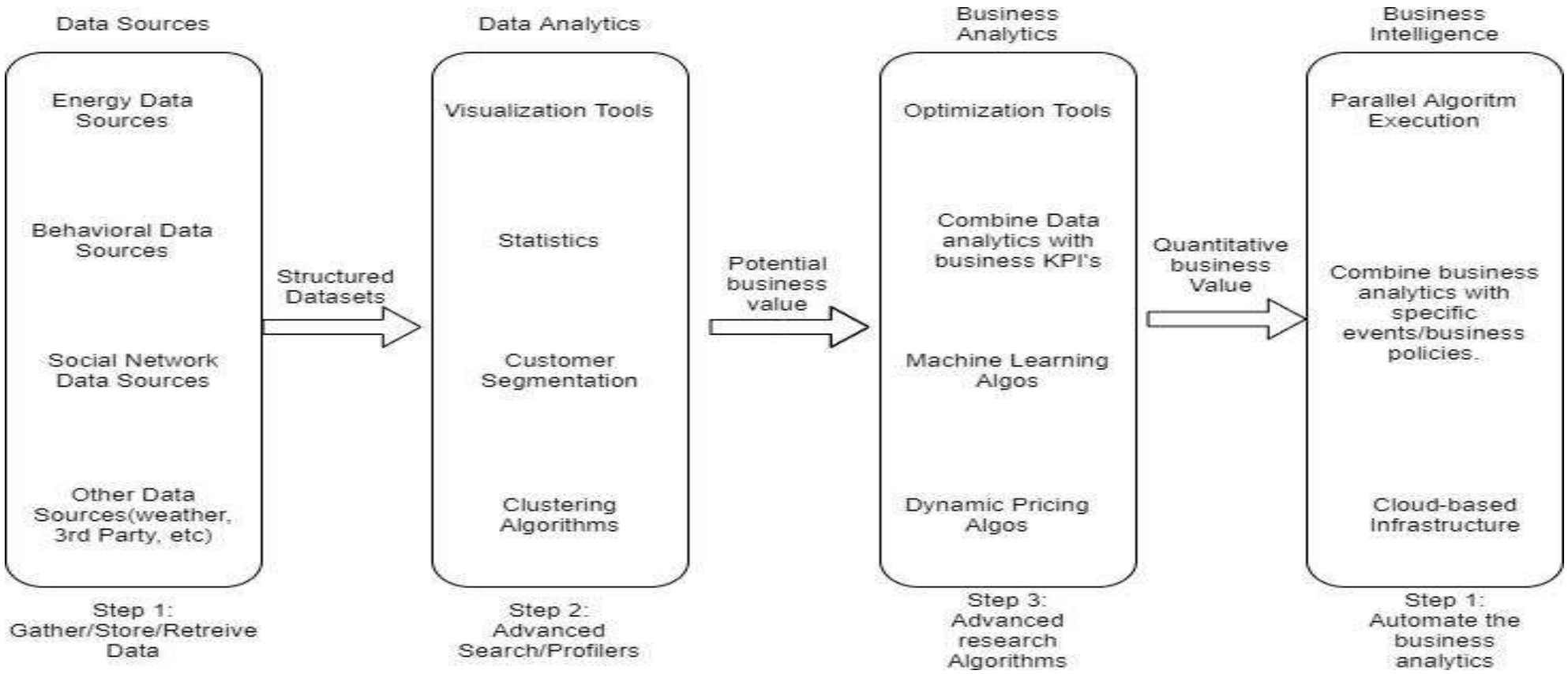

Figure 1: Functional Architecture of Business Intelligence

How BI Really Works

Organizational Memory	Information Integration	Insight Creation	Presentation
• Data Warehouse • ERP • Knowledge Repository • CMS • DMS	• Business Analytics Tool • Data Mining • Real-time Decision	• Text mining tools • Web mining tools • Environmental Scanning • RFID	• OLAP Tools • Visualization tools • Digital Dashboards • Score Card

4.2 Optimization

To optimize the performance of the system:

•Efficient data models were created to reduce data redundancy and enhance query performance.

•Indexing strategies were employed in the SQL database to improve data retrieval speed.

•Power BI's aggregations and hierarchies were leveraged to handle large datasets without impacting report performance.

5: KPIs

Key Performance Indicators (KPIs) were identified to help track and improve business performance. The KPIs used in the system include:

•**Sales Growth Rate**: Tracks the growth in sales over time.
•**Average Order Value (AOV)**: Measures the average dollar amount spent each time a customer places an order.
•**Customer Retention Rate**: Tracks the percentage of customers who continue to purchase over time.
•**Product Return Rate**: Measures the percentage of sold products that are returned by customers.
•**Conversion Rate**: The percentage of website visitors that result in a purchase.

5.1 KPIs (Key Performance Indicators)

To facilitate a detailed analysis, several KPIs are visualized across different dimensions:

•**Category-wise Sales**: Breakdown of sales by product category.
•**Geographic Sales Distribution**: Sales performance across different regions.
•**Seasonal Trends**: Sales performance across different time periods and seasons.
•**Top Products by Sales**: Identifying the top-performing products based on revenue.

6: Deployment

The Amazon Sales Analysis system is deployed using AWS cloud infrastructure, allowing for scalability and easy access to data. The deployment process includes:

- Hosting the SQL Server database on AWS RDS for managed database services.
- Publishing Power BI reports to Power BI Service for easy access and sharing with stakeholders.
- Automating data refresh schedules in Power BI to ensure that reports always reflect the most recent data.
- Ensuring security and access control through AWS Identity and Access Management (IAM) and Power BI's role-based access system.