Low Level Design
Analyzing Amazon Sales data Analysis

Written by	Karan Singh
Document Version	1.0
Last date of revision	15th July 2023

DOCUMENT CONTROL

Version	Date	Author	Comments
1.0	13th July 2023	Karan Singh	Introduction and architecture defined
1.0.1	15th July 2023	Karan Singh	Architecture Description

Reviews:

Version	Date	Reviewer	Comments
1.0	14th July 2023	Amit	Architecture Description to be added
1.1	15th July 2023	Amit	Completed

Contents

1. Introduction
    1.1 What is Low-Level Design Document?
    1.2 Scope

2. Architecture

3. Architecture Description
    3.1 Data Source
    3.2 Data Description
    3.3 Data Transformation
        a) Data Extraction
        b) Data Cleaning and Preprocessing
        c) Data Transformation
        d) Aggregation and Calculation
        e) Data Visualization and Analysis

1. Introduction

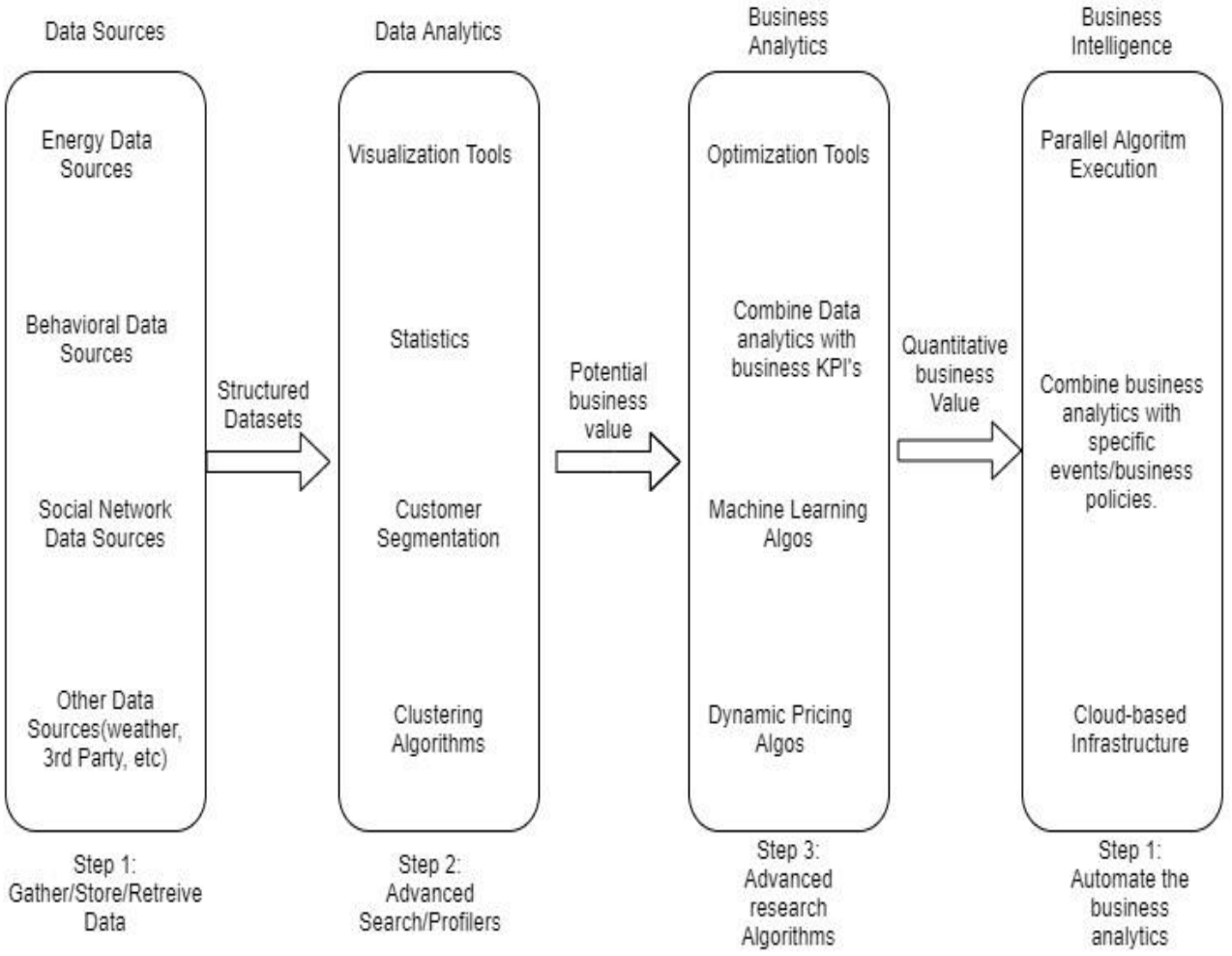
1.1 What is Low-Level design document?

The goal of the LDD or Low-level design document (LLDD) is to give the internal logic design of the actual program code for the House Price Prediction dashboard. LDD describes the class diagrams with the methods and relations between classes and programs specs. It describes the modules so that the programmer can directly code the program from the document.

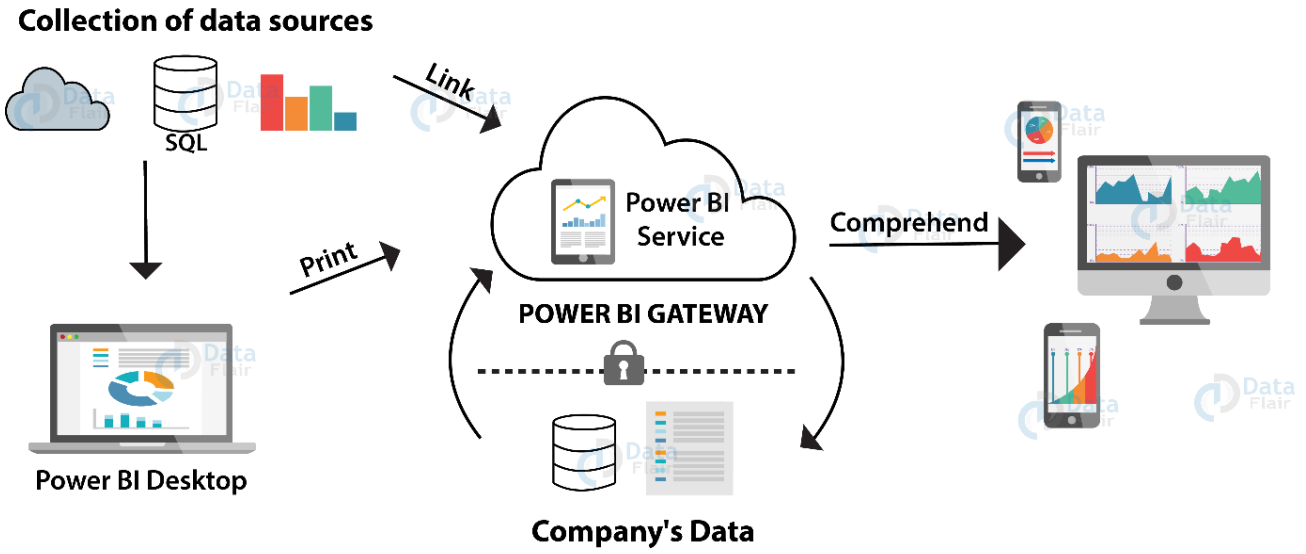
1.2 Scope

Low-level design (LLD) is a component-level design process that follows a step-by-step refinement process. The process can be used for designing data structures, required software architecture, source code and ultimately, performance algorithms. Overall, the data organization may be defined during requirement analysis and then refined during data design work.

2. Architecture



Process Flow of Power BI



3. Architecture Description

3.1 Data Source

To perform the analysis on sales trends, I obtained the dataset from the following link: Dataset Link . The dataset was sourced from [Amazon/any specific source], which provides comprehensive sales data encompassing transaction dates, sales quantities, revenue, and potentially additional relevant information.

By utilizing this dataset, we aimed to extract insights into the sales management domain, focusing on meeting increasing competition, improving distribution methods to reduce costs, and ultimately increasing profits. Sales management is recognized as a pivotal function within commercial and business enterprises, and leveraging this dataset allows us to evaluate the performance and effectiveness of sales strategies.

The dataset was acquired using an ETL (Extract-Transform-Load) process, employing tools such as Python and Power BI. The ETL process ensures that the data is extracted from its source, transformed into a suitable format for analysis, and then loaded into the chosen analysis tool.

Conducting in-depth analysis on the sales dataset enables us to explore sales trends on various time-based dimensions, including month-wise, year-wise, and yearly-month-wise patterns. By identifying key metrics and factors, we aim to establish meaningful relationships between attributes, providing valuable insights to support decision-making and uncover opportunities for sales growth and optimization.

It is important to note that thorough research and analysis were conducted to ensure the accuracy and reliability of the dataset. By utilizing this dataset and employing appropriate tools and techniques, we can derive actionable insights that drive effective sales management strategies and contribute to overall business success.

3.2 Dataset Description

CustKey: Customer identifier.

DateKey: Date identifier.

Discount Amount: Amount of discount applied to the sale.

Invoice Date: Date of the invoice.

Invoice Number: Unique identifier for the invoice.

Item Class: Classification or category of the item.

Item Number: Unique identifier for the item.

Item: Name or description of the item.

Line Number: Line number of the item in the invoice.

List Price: Price of the item as per the list or catalog.

Order Number: Unique identifier for the order.

Promised Delivery Date: Date on which the delivery was promised.

Sales Amount: Total amount of the sale.

Sales Amount Based on List Price: Sales amount based on the list price of the item.

Sales Cost Amount: Cost associated with the sale.

Sales Margin Amount: Difference between sales amount and cost amount.

Sales Price: Actual price at which the item was sold.

Sales Quantity: Quantity of the item sold.

Sales Rep: Sales representative responsible for the sale.

U/M: Unit of measurement for the item.

3.3 Data Transformation

To perform the data transformation and analysis, I utilized Python and Power BI to extract, transform, and visualize the Amazon dataset. Here is an overview of the data transformation steps and techniques used:

a. **Data Extraction:** I accessed the Amazon dataset, which contained sales-related information such as transaction dates, sales quantities, and revenue. The dataset was obtained from a reliable source and stored in a suitable format for further processing.

b. **Data Cleaning and Preprocessing:** To ensure data quality, I performed data cleaning and preprocessing techniques. This involved identifying and handling missing values, removing duplicates, and addressing any inconsistencies or errors in the data. I used Python's Pandas library to apply data cleaning techniques such as removing null values, dropping duplicates, and standardizing data formats.

c. **Data Transformation:** To analyze the sales trends on a month-wise, year-wise, and yearly_month-wise basis, I transformed the dataset accordingly. Using Python, I employed the datetime functionality to extract the month, year, and yearly_month attributes from the transaction dates. This transformation allowed me to categorize the sales data based on these time-related factors.

d. **Aggregation and Calculation:** To calculate key performance indicators (KPIs), I applied various aggregation and calculation techniques. Using Python's Pandas library, I grouped the data by month, year, and yearly_month, and calculated metrics such as sales revenue, sales quantity, average order value, and growth rates. These calculations were performed to derive meaningful insights and evaluate sales performance.

e. **Data Visualization and Analysis:** I utilized Power BI to visualize and analyze the transformed data. Power BI provided a range of visualization options, including line charts, bar charts, and tables, allowing me to represent the sales trends in an interactive and insightful manner. Through visualizations, I identified patterns, trends, and relationships between attributes, enabling a deeper understanding of sales management performance.

Throughout the data transformation process, I ensured data integrity, accuracy, and consistency by performing rigorous quality checks and validations. By leveraging the capabilities of Python and Power BI, I was able to transform the raw Amazon dataset into meaningful information, uncover key insights, and provide valuable recommendations for sales management improvement.