

High Level Design

AMAZON SALES DATA ANALYSIS

SIDDHARTH S SRIVASTAVA

Document Version Control

Date Issued	Version	Description	Author
29-09-2023	1.0	First version of Complete HLD	Siddharth S Srivastava

Table of Contents

1. Introduction.....	4
Why this High-Level Design Document?	4
Scope.....	5
2. General Description	6
Product perspective and problem statement	6
Tools used.....	7
3. Design Details.....	8
Functional Architecture	8
How BI Works	9
Optimization	10
1. Your data strategy drives performance.....	10
2. Reduce the marks (data points) in your view	10
3. Limit your filters by number and type	10
4. Optimize and materialize your calculations.....	11
4. KPI.....	13
KPIs (Key Performance Indicators.....	13
5. Deployment	15
Power BI Report	16

Abstract

The growths of resell business in most populated cities are increasing and market competitions are also high. The dataset is one of the historical sales of a company named Adventure Works which has records for 3 years. Good data driven systems for analyzing sales can improve the performance of the company and generate more ROI to the stakeholders.

1.Introduction

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

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.

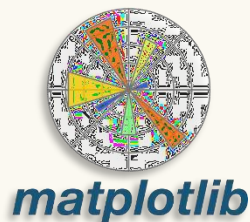
2.General Description

Product perspective and problem statement

The goal of this project is to perform Sales and Cost analysis. It provides insights about the top performing products in terms of Sales and Profit, the general trend of Sales and Cost across different Years, Months and Quarters.

Tools used

Business Intelligence tools and libraries works such as NumPy, Pandas, Seaborn, Matplotlib, MS-Power BI, Jupyter Notebook and Python Programming Language are used to build the whole framework



3. Design Details

Functional Architecture



STEP 1

Data from source system is integrated and loaded into a data warehouse of another analytics repository



STEP 2

Data sets are organised into analytics data models or OLAP cubes to prepare them for analysis



STEP 3

BI analysts, other analytics professionals and business users run analytical queries against the data



STEP 4

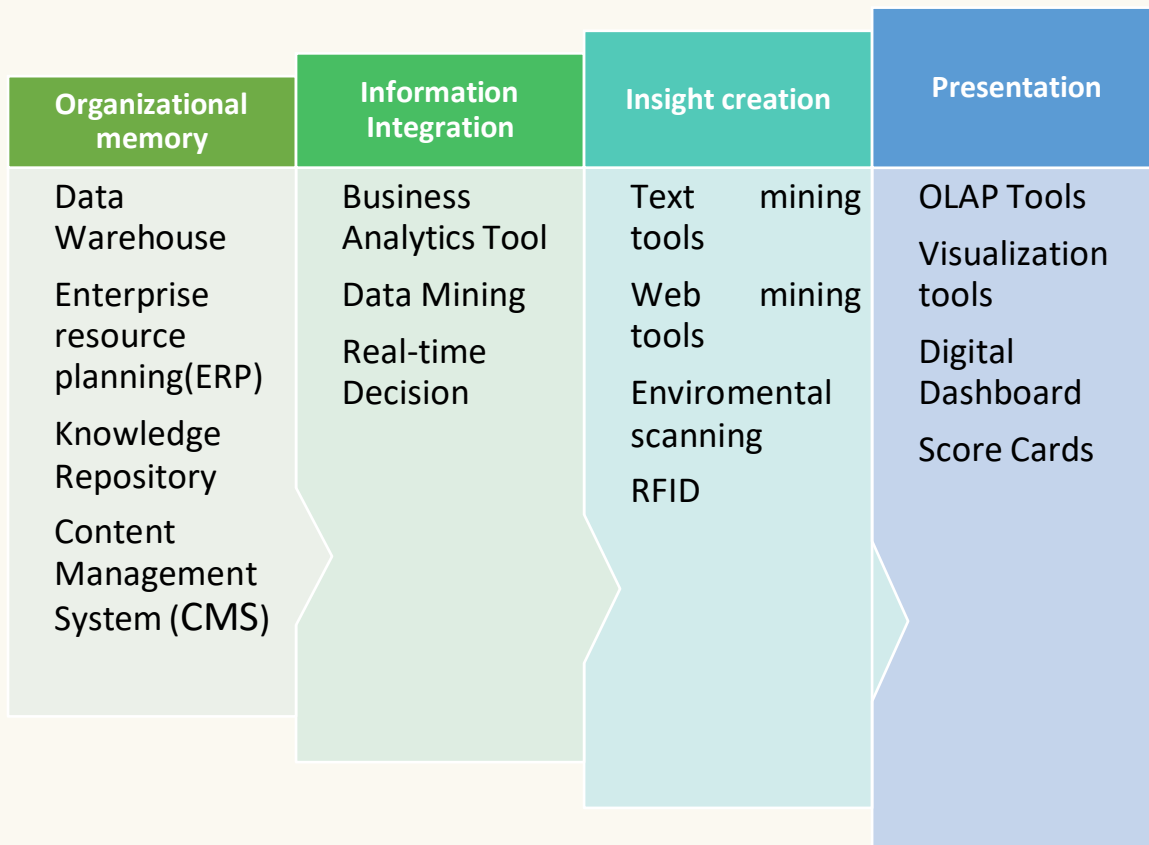
The query results are built into data visualizations, dashboards, Reports and online portals



STEP 5

Business executives and workers use the information for decision-making and strategic planning

How BI Works



Optimization

1. 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

2. 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

3. 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 including 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 data 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).

4. Optimize and materialize your calculations

- Perform calculations in the database
- Reduce the number of nested calculations
- Reduce the granularity of LOD (level of detail) or table calculations in the view. The more granular the calculation, the longer it takes.
- LODs - Look at the number of unique dimension members in the calculation
- 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.KPI

Dashboards will be implemented to display and indicate certain KPIs and relevant indicators for the sales



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

KPIs (Key Performance Indicators)

Key indicators displaying a summary of the sales and profit generation, Total COGS (Cost of goods sold)

1. Total Sales
2. Total COGS
3. Total Profit
4. Total Sales and Total COGS across Years, Months, and Products.
5. Top 7 Sales and Profit generating Products

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 analysing 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 Power BI at scale, as well as organizing, orchestrating, and unifying disparate sources of data for business users and experts alike to author and consume content.
- Power BI Desktop and Power BI Service leverage your existing technology investments and integrate them into your IT infrastructure to provide a self-service, modern analytics platform for your users. With on-premises, cloud, and hosted options, there is a version of Power BI to match your requirements.

5.1 Power BI Report

