High Level Design (HDL)

Heart Disease Diagnostic-Analysis

Revision Version: 1.0

Last Date of Revision:16/02/2023

Manish Jha

Document Version Control

Date Issued	Version	Description	Author
1 Feb 2023	1.0	Abstract, Introduction	Manish Jha
1 Feb 2023	1.1	General Description	Manish Jha
5 Feb 2023	1.2	Design Detail	Manish Jha
10 Feb 2023	1.3	KPIs	Manish Jha
16 Feb 2023	1.4	Deployment	Manish Jha

Contents

Abstract

These days online shopping is the new way to shop and when it comes to an online place which can offer you everything at one place then its none other than Amazon one of the leading E-Commerce which has services across the globe and has everything to shop online which one can think.

But these days Competition exists In every sector hence E-Commerce is also not an exception so in order to stay in the market and keep growing at required pace taking good business decision becomes a necessity and analysis of Data can be a catalyst in this process it can help us to see right patterns and help us to generate good insights.

so this Project we are analysing Amazon sales data for various regions and for various item categories ,and data which is available to us is from 2010 to 2017 .

1 Introduction

1.1 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:

A. Present all the design aspects and define them in detail
B. Describe the user interface being implemented
C. Describe the hardware and software interfaces
D. Describe the performance requirements
E. Include design features and the architecture of the project
F. List and describe the non-functional attributes like:
    - Security
    - Reliability
    - Maintainability
    - Portability
    - Reusability
    - Application compatibility
    - Resource utilization
    - Serviceability

•

1.2 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

2.1 Product Perspective & Problem Statement

These days Competition exists In every sector hence E-Commerce is also not an exception so in order to stay in the market and keep growing at required pace taking good business decision becomes a necessity and analysis of Data can be a catalyst in this process it can help us to see right patterns and help us to generate good insights.

The objective of the project is to perform data visualization techniques to understand the insight of the data. This project apply Business Intelligence tools such as Tableau to get a visual understanding of the data.

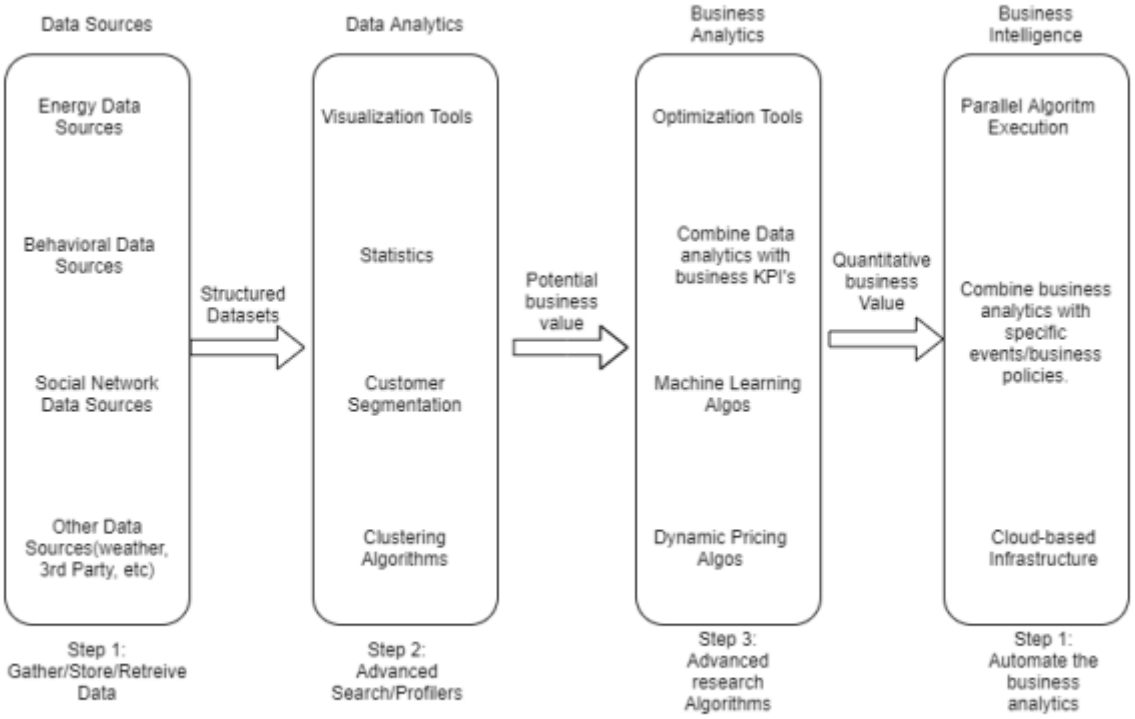
2.2 Tools used

Business Intelligence tools Excel and Tableau are used to build the whole framework.

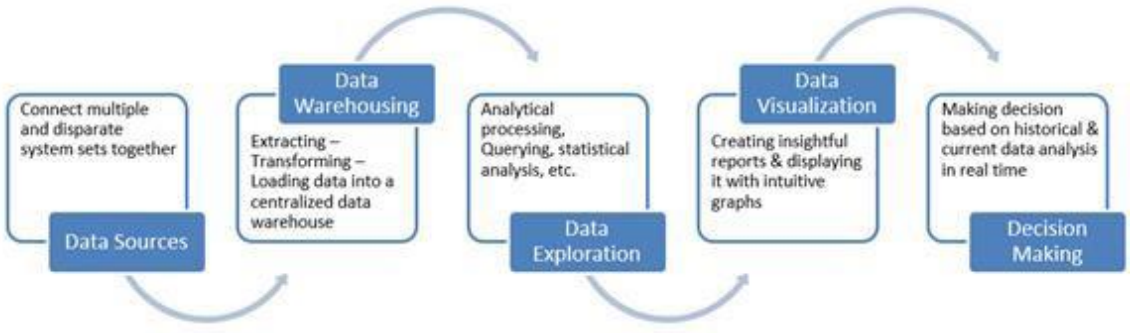
3 Design Details

3.1 Functional Architecture

Functional Architecture of Business Architecture:



How Business Intelligence Work:



3.2 Optimization

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

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.

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 include 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 date 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).

Optimize and materialize your calculations

- Perform calculations in the database
- Reduce the number of nested calculations.
- Reduce the granularity of LOD 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 KPIs

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

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.

4.1 KPIs (Key Performance Indicators)

Key indicators displaying a summary of the Amazon sales Data-Analysis and its relationship with different metrics:

1. Profit distribution item wise
2. Total Sold units distribution item wise
3. Profit distribution region and country wise
4. Region wise Distribution of cost to produce single unit
5. Calculation of shipping time for different item in different region
6. Sales Channel preference

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 analyzing 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 Tableau at scale, as well as organizing, orchestrating, and unifying disparate sources of data for business users and experts alike to author and consume content.

Tableau prioritizes choice in flexibility to fit, rather than dictate, your enterprise architecture. Tableau Server and Tableau Online leverage your existing technology investments and integrate 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 Tableau to match your requirements.

Below is a comparison of the three types:

TYPE PROS CONS

Tableau Server - On Premises

- Full control of hardware and software
- Infrastructure and data remain behind your firewall
- Need dedicated administrators to manage hardware and software
- Additional infrastructure needed to access off-network (mobile, external)

Tableau Server - Public Cloud (IaaS)

- Full control of software on managed hardware
- Puts infrastructure in same place as data (for migration to cloud)
- Flexibility to spin up/down hardware as needed
- Need dedicated administrators to manage software
- Additional infrastructure needed to access off-network (mobile, external)

Tableau Online (SaaS)

- Fully hosted solution (hardware, software upgrades)
- Fast to deploy
- Easy for external audience to access
- Single site in multi-tenant environment
- Cubes are not supported
- No guest account access

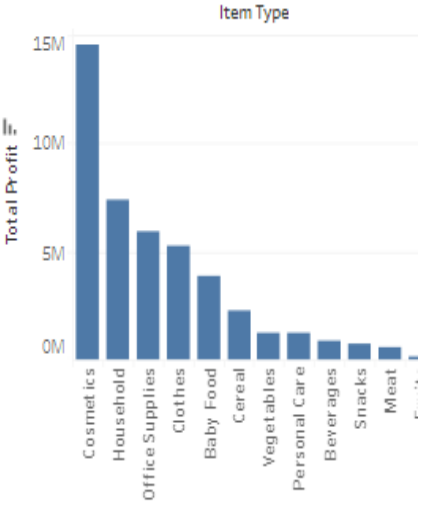
Shiping Time in days for each product for each country and every region

Country	Item Type	Order Date	Ship Date	T
Albania	Clothes	2/2/2010	3/18/2010	4
Angola	Household	4/23/2011	4/27/2011	4
Australia	Beverages	7/7/2014	7/11/2014	4
	Cereal	6/9/2013	7/2/2013	2
	Office Supp..	10/27/2015	11/25/2015	2
Austria	Cosmetics	2/23/2015	3/2/2015	7
Azerbaijan	Cosmetics	2/6/2010	2/25/2010	1
	Office Supp..	6/13/2012	7/24/2012	4
Bangladesh	Clothes	1/13/2017	3/1/2017	4
Belize	Clothes	7/25/2016	9/7/2016	4
Brunei	Office Supp..	4/1/2012	5/8/2012	3
Bulgaria	Clothes	4/23/2012	6/3/2012	4

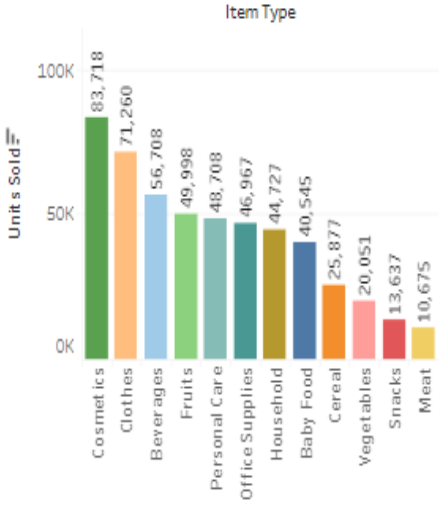
Sales Channel Preference

Sales Chan..	
Offline	50
Online	50

Most Profitable Item



Total Units Sold Per Item Type



Sum of unit price for each item region wise

Item Type	Region	
Baby Food	Australia and Oceania	511
	Europe	1,021
	Sub-Saharan Africa	255
Beverages	Australia and Oceania	95
	Central America and the C..	47
	Europe	95
	Sub-Saharan Africa	142
Cereal	Australia and Oceania	206
	Central America and the C..	206
	Middle East and North Afr..	206
	Sub-Saharan Africa	823
Clothes	Asia	219
	Australia and Oceania	109
	Central America and the C..	109
	Europe	328
	Middle East and North Afr..	219
	Sub-Saharan Africa	437
Cosmetics	Asia	437
	Australia and Oceania	437
	Central America and the C..	437
	Europe	2,186
	Middle East and North Afr..	1,312
	Sub-Saharan Africa	874
Fruits	Asia	9
	Australia and Oceania	19

0K 2K 4K
Unit Price

Total Profit Country Wise

Region	Country
	Samoa
	Solomon Islands
	Tuvalu
Central America and the Caribbean	Belize
	Costa Rica
	Grenada
	Haiti
	Honduras
	Nicaragua
Europe	Albania
	Austria
	Bulgaria

0M 1M 2M
Total Profit