LOW LEVEL DESIGN DOCUMENT

Amazon Sales Data Analysis

BY
VIKASH KUMAR MAHAPATRA

Revision Number & Date :- 1.2 26/09/2022

Document Version Control

Date Issued	Version	Description	Author
21 SEPT 2022	1.0	Created Document And Add Document Version Control, Content Page, Introduction Page	Vikash
23 SEPT 2022	1.1	Created Architecture Page ,Architecture Description Page & Add page number to contents section	Vikash
26 SEPT 20222	1.2	Final Revision	Vikash

CONTENTS

INTRODUCTION

WHY THIS LOW LEVEL DESIGN DOCUMENT

The purpose of this Low-Level Design (LLD) document is to give the internal logic design of the actual program code for Foreign Direct Investment Analysis .LLD document describes the class diagrams with the methods and relations between classes and program specs. It describes the modules so that the programmer can directly code the program from the document.

SCOPE

The LLD document 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 performance algorithms. Overall, the data organization may be defined during requirement analysis and then refined during data design work.

PROJECT INTRODUCTION

As different Organizations under the e-commerce industry seek to increase market footprint by creating and sustaining a unique process to collect personal information about customers and their purchasing trends. The report critically evaluates how a service-based organization Amazon uses management information systems as a vibrant tool in attaining competitive advantage through efficient management and acquisition of information. As in today’s market without proper sales management, it’s very hard to predict how the business is running and how it will be in future. Many companies with proper sales management have shown better growth as they already know which item they have to focus on, which product needs some improvement etc. Sales management helps in maintaining its customer base for a longer time by providing them attractive offers, as they already have the information’s like who are their top customers, whom they have to focus on etc. Sales management so helps in minimizing the losses. Also, the competition is increasing day by day as many new companies are coming with better management systems and giving tough

competition due to that it is now very important to have a proper sales management to run any business and to compete with these companies.

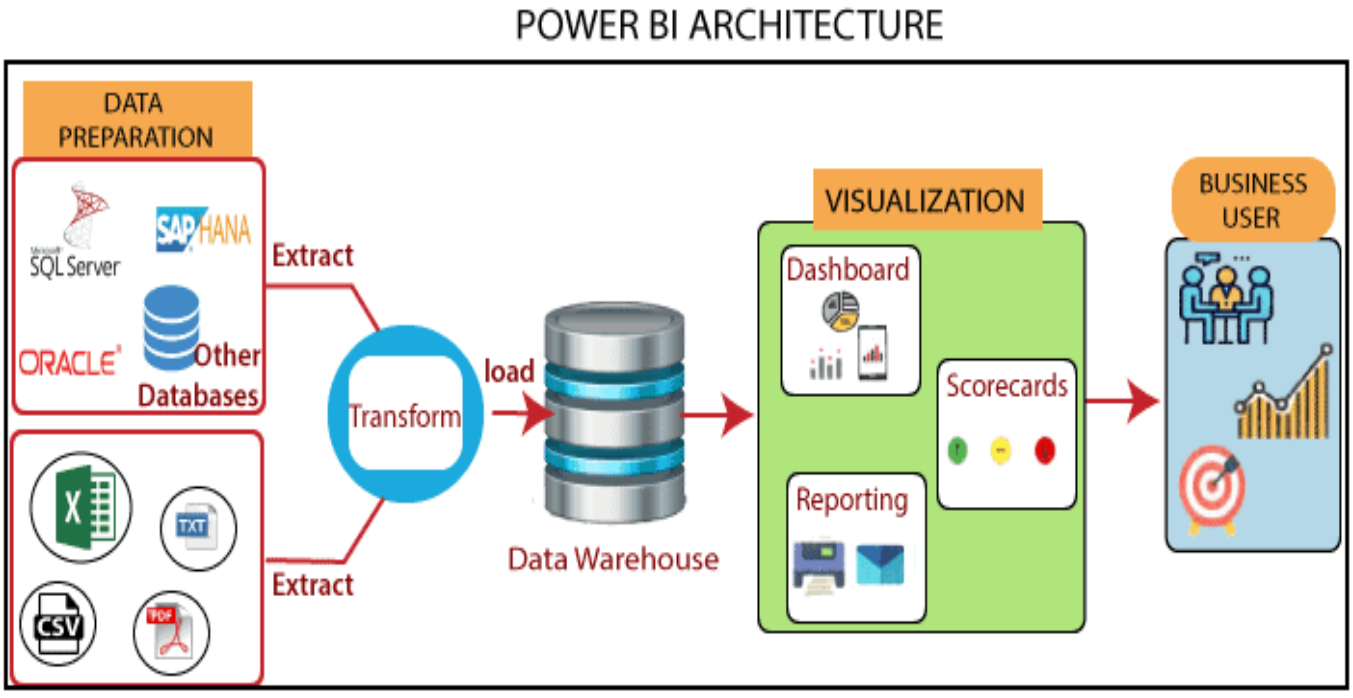
PROBLEM STATEMENT

The problem statement aims to analyze amazon sales from 2017-2019 to know more in depth information about the items that are in high demand, high profit so that ways are find out to improve methods of distribution to reduce cost ,stay ahead in the competition, have progressive growth as well as further increase in the sales. As sale management today is the most important function in a commercial and business enterprise. To achieve the goal, we used a data set that is given and analyze most important parameter that are responsible for it. Also get some meaningful insights from the given information regarding Sales Data.

Do ETL (Extract-Transform-Load) the dataset and find for me some information from this large data. This is form of data mining. What all information can be achieved by mining this data, would be brainstormed by interns. Find key metrics and factors and show the meaningful relationship between attributes. Do your own research and come up with findings.

ARCHITECTURE

Microsoft Power BI architecture consists of four major steps that explain the whole process from data sourcing to creation of reports and dashboard. Various technologies and process work together to get required results with extreme precision. Let us see those steps in pictorial format.



Sourcing data: Power BI extracts data from various servers, Excel sheets, CSV files and databases. The extracted information can be directly imported to Power BI or a live service link is established to receive it. If you directly import the data in Power BI, it will only be compressed up to 1GB post that you can only run live queries on your chunky datasets.

Transforming the data: Before visualizing the data, cleaning and preprocessing it should be done. This means useless or missing values from rows or columns. Following that certain rules will be applied to transform and load datasets into the warehouse.

Report and publish: After cleaning and transforming the data, reports will be created based on requirements. A report is a visualization of the data with different filters and constraints presented in the form of graphs, pie-chart, and other figures.

Creating dashboards: Power BI Dashboards are created by pinning individual elements or pages of live reports. Dashboard should be created after you have published your reports to the BI service. When the reports get saved, the visual maintains the filter setting chosen so that the user can apply filters and slicers.

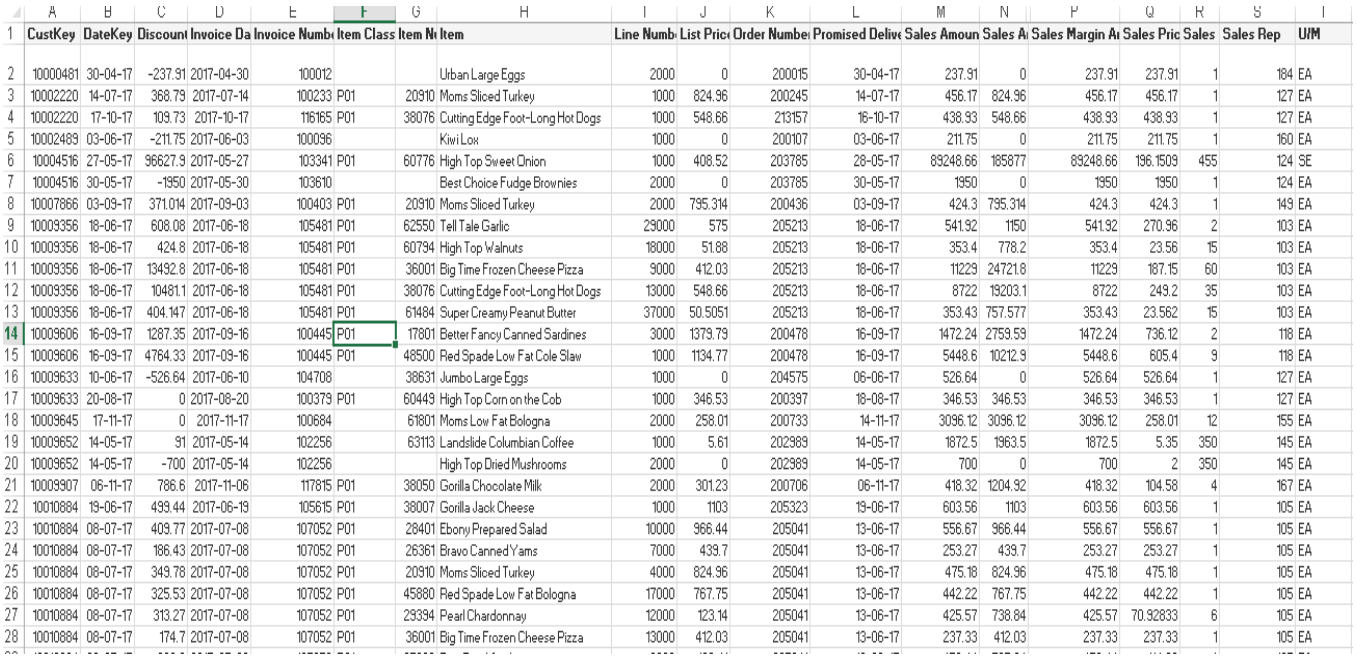
ARCHITECTURE DESCRIPTION

Data Sourcing:

The dataset is in excel format and Microsoft excel is used to load the data. Dataset is taken from project description document and here goes the drive link https://drive.google.com/drive/folders/1FkmFVL8wlJmQWP1z52TD8PlhOJhiTyl?usp=sharing

Data Description:

The file name salesdata.xls contains different columns such as invoice data(day on which invoice generated),discount amount(total discount provided on any item), sales amount (total sales price of an item), sales margin amount (profits), sales cost amount (total cost price of any item),Sales quantity(total quantity sold of any item) and U/M(unit of measure which have values like EA-> Each ,PR->Pair and SE-> kilos or gallons), etc. . Sample data below.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	P	Q	R	S	T
1	CustKey	DateKey	Discount	Invoice Da	Invoice Numb	Item Class	Item Nu	Item	Line Numb	List Pric	Order Numbe	Promised Delive	Sales Amoun	Sales A	Sales Margin A	Sales Pric	Sales	Sales Rep	U/M
2	10000481	30-04-17	-237.91	2017-04-30	100012			Urban Large Eggs	2000	0	200015	30-04-17	237.91	0	237.91	237.91	1	184	EA
3	10002220	14-07-17	368.79	2017-07-14	100233	P01	20910	Moms Sliced Turkey	1000	824.96	200245	14-07-17	456.17	824.96	456.17	456.17	1	127	EA
4	10002220	17-10-17	109.73	2017-10-17	116165	P01	38076	Cutting Edge Foot-Long Hot Dogs	1000	548.66	213157	16-10-17	438.93	548.66	438.93	438.93	1	127	EA
5	10002489	03-06-17	-211.75	2017-06-03	100096			Kiwi Lox	1000	0	200107	03-06-17	211.75	0	211.75	211.75	1	160	EA
6	10004516	27-05-17	96627.9	2017-05-27	103341	P01	60776	High Top Sweet Onion	1000	408.52	203785	28-05-17	89248.66	185877	89248.66	196.1509	455	124	SE
7	10004516	30-05-17	-1950	2017-05-30	103610			Best Choice Fudge Brownies	2000	0	203785	30-05-17	1950	0	1950	1950	1	124	EA
8	10007866	03-09-17	371.014	2017-09-03	100403	P01	20910	Moms Sliced Turkey	2000	795.314	200436	03-09-17	424.3	795.314	424.3	424.3	1	149	EA
9	10009356	18-06-17	608.08	2017-06-18	105481	P01	62550	Tell Tale Garlic	29000	575	205213	18-06-17	541.92	1150	541.92	270.96	2	103	EA
10	10009356	18-06-17	424.8	2017-06-18	105481	P01	60794	High Top Walnuts	18000	51.88	205213	18-06-17	353.4	778.2	353.4	23.56	15	103	EA
11	10009356	18-06-17	13492.8	2017-06-18	105481	P01	36001	Big Time Frozen Cheese Pizza	9000	412.03	205213	18-06-17	11229	24721.8	11229	187.15	60	103	EA
12	10009356	18-06-17	10481.1	2017-06-18	105481	P01	38076	Cutting Edge Foot-Long Hot Dogs	13000	548.66	205213	18-06-17	8722	19203.1	8722	249.2	35	103	EA
13	10009356	18-06-17	404.147	2017-06-18	105481	P01	61484	Super Creamy Peanut Butter	37000	50.5051	205213	18-06-17	353.43	757.577	353.43	23.562	15	103	EA
14	10009606	16-09-17	1287.35	2017-09-16	100445	P01	17801	Better Fancy Canned Sardines	3000	1379.79	200478	16-09-17	1472.24	2759.59	1472.24	736.12	2	118	EA
15	10009606	16-09-17	4764.33	2017-09-16	100445	P01	48500	Red Spade Low Fat Cole Slaw	1000	1134.77	200478	16-09-17	5448.6	10212.9	5448.6	605.4	9	118	EA
16	10009633	10-06-17	-526.64	2017-06-10	104708		38631	Jumbo Large Eggs	1000	0	204575	06-06-17	526.64	0	526.64	526.64	1	127	EA
17	10009633	20-08-17	0	2017-08-20	100379	P01	60449	High Top Corn on the Cob	1000	346.53	200397	18-08-17	346.53	346.53	346.53	346.53	1	127	EA
18	10009645	17-11-17	0	2017-11-17	100684		61801	Moms Low Fat Bologna	2000	258.01	200733	14-11-17	3096.12	3096.12	3096.12	258.01	12	155	EA
19	10009652	14-05-17	91	2017-05-14	102256		63113	Landslide Columbian Coffee	1000	5.61	202989	14-05-17	1872.5	1963.5	1872.5	5.35	350	145	EA
20	10009652	14-05-17	-700	2017-05-14	102256			High Top Dried Mushrooms	2000	0	202989	14-05-17	700	0	700	2	350	145	EA
21	10009907	06-11-17	786.6	2017-11-06	117815	P01	38050	Gorilla Chocolate Milk	2000	301.23	200706	06-11-17	418.32	1204.92	418.32	104.58	4	167	EA
22	10010884	19-06-17	439.44	2017-06-19	105615	P01	38007	Gorilla Jack Cheese	1000	1103	205323	19-06-17	603.56	1103	603.56	603.56	1	105	EA
23	10010884	08-07-17	409.77	2017-07-08	107052	P01	28401	Ebony Prepared Salad	10000	966.44	205041	13-06-17	556.67	966.44	556.67	556.67	1	105	EA
24	10010884	08-07-17	186.43	2017-07-08	107052	P01	26361	Bravo Canned Yams	7000	439.7	205041	13-06-17	253.27	439.7	253.27	253.27	1	105	EA
25	10010884	08-07-17	349.78	2017-07-08	107052	P01	20910	Moms Sliced Turkey	4000	824.96	205041	13-06-17	475.18	824.96	475.18	475.18	1	105	EA
26	10010884	08-07-17	325.53	2017-07-08	107052	P01	45880	Red Spade Low Fat Bologna	17000	767.75	205041	13-06-17	442.22	767.75	442.22	442.22	1	105	EA
27	10010884	08-07-17	313.27	2017-07-08	107052	P01	29394	Pearl Chardonnay	12000	123.14	205041	13-06-17	425.57	738.84	425.57	70.92833	6	105	EA
28	10010884	08-07-17	174.7	2017-07-08	107052	P01	36001	Big Time Frozen Cheese Pizza	13000	412.03	205041	13-06-17	237.33	412.03	237.33	237.33	1	105	EA

Data Pre-Processing: Before building any model, it is crucial to perform data pre-processing to feed the correct data to the model to learn and predict. Model performance depends on the quality of data fed to the model to train. This process includes following steps:- Handing Missing/Null values , Handling Skewed Data and Outliers Detection and removal.

Data Cleaning: Data cleaning is the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset. a)Remove duplicate or irrelevant observations. b) Filter unwanted outliers. c) Renaming required attributes.

Exploratory Data Analysis(EDA): Exploratory Data Analysis refers to the critical process of performing initial investigations on data to discover patterns, spot anomalies,

test hypothesis and check assumptions with the help of summary statistics and graphical representations.

Reporting: Reporting is a most important and underrated skill of a data analytics field. Because being a Data Analyst you should be good in the easy and self-explanatory report because your model will be used by many stakeholders who are not from a technical background. a)HLD(High Level Document)b)LLD(Low Level Document)c)Architecture Report. d) Wireframe report. e) Detailed Project Report.

Modelling: Data Modelling is the process of analyzing the data objects and their relationship to the other objects. It is used to analyze the data requirements that are required for the business processes. The data models are created for the data to be stored in a database. The Data Model's main focus is on what data is needed and how we have to organize data rather performing the operations that are need.

Deployment: We create a Power BI Dashboard