Amazon Sales Data Analysis

Low level document

**Hypothetical data record – 100 records**

Time Period: 2010-2017



Contents:

|  |  |
| --- | --- |
| Serial No. | Topic |
| 1. | Introduction and scope of the document |
| 2. | Architecture |
| 3. | Working of Power BI architecture |
| 4. | Tool selection advantages |
| 5. | Data Description |
| 6. | Data Transformation |
| 7. | Data insertion into database |
| 8. | Deployment |

|  |  |
| --- | --- |
| Written by: | Anisha A |
| Last Revised Date: | 28-02-2023 |
|  |  |

**Document Control**

**Reviews:**

|  |  |  |  |
| --- | --- | --- | --- |
| Version | Date | Reviewer | Comments |
|  |  |  |  |

**Approval Status:**

|  |  |  |  |
| --- | --- | --- | --- |
| Version | Date | Reviews | Comments |
|  |  |  |  |

# Scope of the document

This document explains the internal logic design of the sales analysis dashboard.

# Architecture

## **Power BI Architecture**

The architecture consists of four steps to explain the process from data sourcing to the creation of reports and dashboards.

* **Sourcing of data**
  + Power BI extracts data from various servers, excel sheets, csv files.
  + 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 1 GB.
* **Transforming the data**
  + Before visualizing the data, cleaning and pre-processing should be done. This means removing useless or missing values from rows or columns. Following that, certain rules will be applied to transform and load the datasets into the warehouse.
* **Report** 
  + 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 charts, donut charts, heat maps and other figures.
* **Create dashboards and Publish**
  + Dashboards are created by pinning individual elements or pages of live reports. Dashboards are created after you have published your reports to the BI service. When the reports get saved, the visual maintains the filter settings chosen so that the user can apply filters and slicers.

# Working of Power BI architecture

The architecture is mainly divided into two parts:

* On-cloud and
* on-premises services.

Below is a Power BI data flow Diagram that helps one understand the flow of data from On-premises to On-cloud server applications.

Data sources such as web browsers, Excel sheets, and other sources feed the information to various Power BI components. Power BI has various data sources, including direct connections, in-house servers, cloud databases, and more. Best practices of Power BI architecture helps in  create stunning reports for better business analytics.

# Why Power BI was used a tool for analysis?

* **We can Access data in different formats:**Power BI can view, analyse, and visualize vast amounts of data in different formats, including Excel, pdf, XML, JSON, CSV, etc.
* **Secure Data Analytics:**Power BI keeps your business data secure by providing features such as sensitive labels, data loss prevention, the oversight of sensitive data with service tags, Azure Private Link etc
* **BI is free to use :**Microsoft provides BI desktop for free and the Power BI Pro version at a very low price so that anyone can access Power BI’s cost-effective tools for their business growth.
* **Interactive AI features to build dashboards:**It helps non-technical individuals as well to build reports and find quick insights into their business from both structured and unstructured data, including text and images.
* **Easy to share:** With the power BI service account, you can share the report with external users also .

# Dataset Description

The dataset contains information in following columns:

1. Region: It broadly gives idea about different regions of the world where products are delivered.
2. Country: It further divides region into countries belonging to that specific region.
3. Item type: Different products of the company
4. Sales channel: mentions two distribution channels ie. Offline and online
5. Order priority: Gives information about critical (C), Medium(M), low(L), high(H)
6. Order Id: mentions distinct order IDs
7. Units sold: gives a number of quantity sales
8. Units price: mentions price per one quantity
9. Units cost mentions cost per one quantity
10. Total revenue: revenue is the multiplication of units sold and units price
11. Total cost: cost is the multiplication of units sold and unit cost
12. Total profit: total revenue -total cost
13. Order date: mentions purchase date in different formats
14. Ship date: mentions ship date in different formats

# Data Transformation

The dataset has been transformed to bring order date and ship date in one single date format so that it is easier to perform further calculations.

# Data insertion into database

A database is created to the name ‘Amazon’. And a connection is established to the relevant database. A table is created and then files are inserted to it.

# Deployment

Once dashboards are created. We publish it (button on the home page). It signs in to the Power BI service that is the cloud server of Power BI. And now it is available for other users to view and take the insights.