Table of Contents

**Introduction2**

**Design Approach2**

**Design** **Flow3**

**B2B Database** **Structure4**

**B2B Warehouse** **Structure5**

**B2B** **Dashboard6**

**Setup** **Steps7**

**Introduction**

The B2B Warehouse framework manages data for a B2B e-commerce site. The data storage system has three defined data sources that feed the data to the system:

* B2B platform database
* Log data from the webserver
* Marketing lead spreadsheet file

## Design Approach

### Separation of Data Sources, Integration and Warehouse Module Layers

There are three distinct layers in the data warehouse:

1. A layer of source and ETL pipeline-specific data sources, containing Python and SQL code used to transform and rename incoming tables from each source into common formats
2. An Integration layer, containing SQL transformations used to integrate, merge, deduplicate and transform data ready for loading into the main warehouse fact and dimension tables.
3. A warehouse layer made-up of fact and related dimensions

## Design Flow

Process flow in terms of data sourcing, data transformation, data extraction, data flow, error logging and data visualization can be seen in the diagram below.

## 

Data sources are built on **MySQL**, **Text WebLog** and **Spreadsheet**. **Talend** initiates data extraction into staging database from data sources using **Python** **Code** that is built using **Spyder** **IDE**. Then it calls **PostgreSQL** functions to transform and load the data in warehouse. **Power** **BI** tool then imports data from warehouse and generates reports.

## B2B Database Structure

## There are 8 entities in B2B Database. Company contains its catalog of Products in CompanyCatalog table. Supplier is a type of Company having similar attributes and contains its catalog of Products in SupplierCatalog table. Company places order for Customers from multiple Suppliers. Attributes of that order are kept in OrderTable and the Products and Suppliers of that Order reside in OrderItems table.

## 

## B2B Warehouse Structure

## There are 9 dimensions and 1 fact in B2B Warehouse. Sales Fact has granularity level where information of Order is stored for its Products, Company and Suppliers. Company, Customer, Product, Supplier tables from B2B source are kept as is as dimensions. WebLog and Marketing Lead dimensions are created from the other two data sources. Time, Date, Location dimensions are introduced to the warehouse as conformed dimensions.

## 

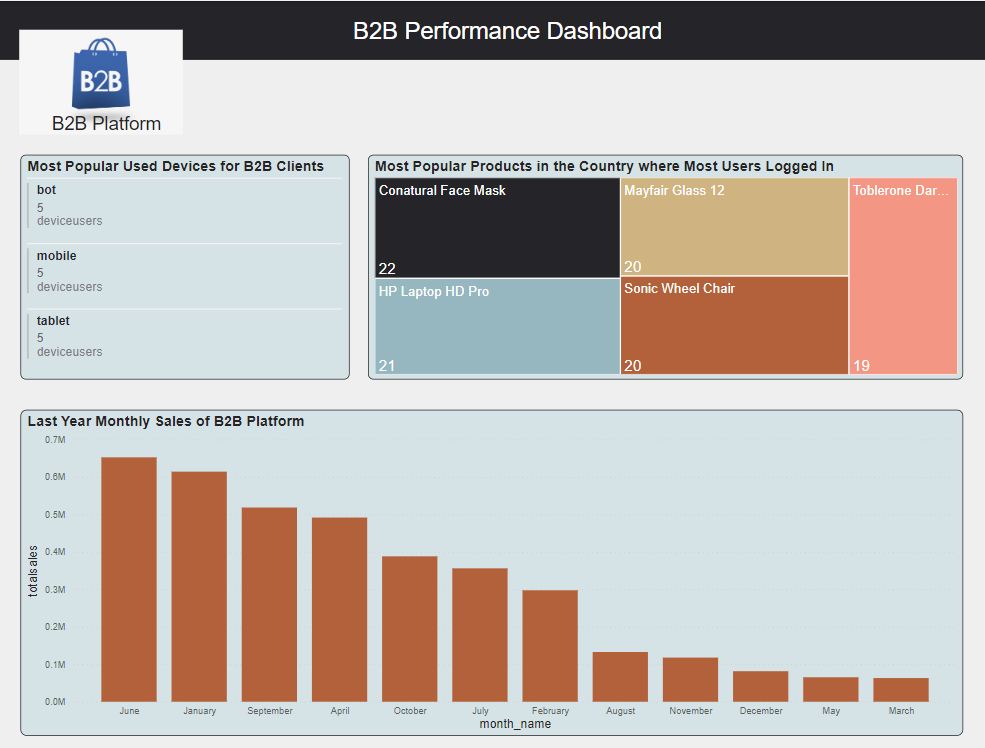
## B2B Dashboard

## B2B dashboard visualizes the following reports:

## Most popular used devices for B2B clients.

## Most popular products in the country from which most users log into.

## All sales of B2B platform displayed monthly for the last year.



## Setup Steps

* Download the Project from GitHub.
* Install the below mentioned tools.
  + MySQL and its dependencies (Visual Studio, etc. depending on your OS)
  + PostgreSQL
  + DBeaver (or any other multi-platform database tool)
  + Python
  + Pip
  + Talend Open Studio
  + Spyder (or any other IDE)
  + Power BI (optional)
* Install the following python libraries.
* Geocoder
* Mysql-connector-python
* User-agents
* Psycopg2
* Getuseragent
* Requests
* Connect to Source Database (MySQL) and Warehouse Database (PostgreSQL) using DBeaver or any other IDE.
* Execute the DDL and DML statements placed in B2B Source Database Setup folder using Source Connection (MySQL). This should set up the structure of the source database system.
* Execute the DDL statements placed in B2B Warehouse Setup folder using Warehouse Connection (PostgreSQL). This should set up the structure of staging db and the warehouse.
* Functions for Data Transformation and Loading into Dimensions and Fact are provided in Warehousing folder. Execute the scripts in this folder to get them created.
* Modify the Talend\_Configs.txt and Configs.json files and change the variables to match your host, database, username, passwords, email configurations and relevant file paths to run Python sub jobs.
* Sample Location Data for Europe is provided in Location\_Data Directory, load it using Load\_Location\_Data.bat file.
* Use Generate\_Weblog\_from\_Website.py file placed in Weblog\_Sourcing Folder as trigger on B2B Website for Logging Device Activities data in Combined Log Format.
* Sample data for Weblog is dumped into Weblog Data.txt.
* Import Talend\_Configs file as Implicit Context and then import B2B\_Warehousing folder in Talend Studio and run B2B\_Warehousing\_Parent Job.
* Run the Queries in Report Queries folder to generate B2B Ecommerce Reports.
* Reporting folder contains Power BI file for reports visualization. Open the file to envision the results of reports.
* Your Warehouse, Reports and Dashboard are Ready. Goodluck!