




# E – Commerce Data Analysis



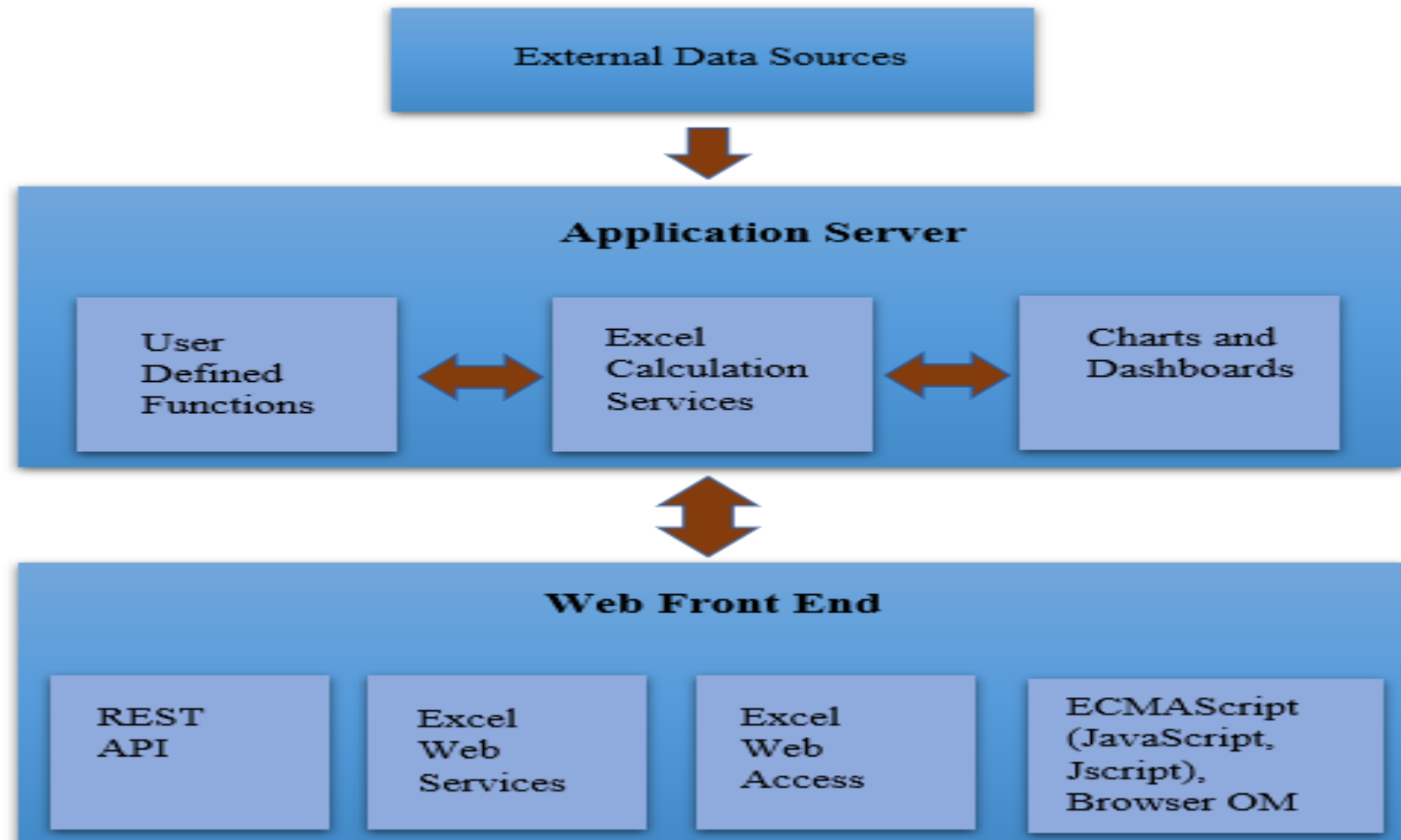
## **Objective:**

Development of a Sales dashboard for different product categories.

## **Benefits:**

- Gives region wise Sales
  - Trend of Sales and Profit according to Month
  - Helps in finding frequency of orders in terms of aging
  - Slicer for different product categories
- 

# Architecture





# Data Validation & Data Transformation

- Name Validation – Validation of file name as per DSA. We have created a regex pattern for validation. After it checks for the date format and time format if these requirements are satisfied, we move such files to “Good\_Data\_Folders” else “Bad\_Data\_Folders”.
- Number of Columns – Validation of number of columns present in the files, and if it doesn't match then the file is moved to “Bad\_Data\_Folders”.
- Name of Columns – The name of the columns is validated and should be the same as given in the schema file. If not, then the file is moved to “Bad\_Data\_Folders”.
- Data type of columns – The data type of columns is given in the scheme file. It is validated when we insert the files into database. If the datatype is wrong, then the file is moved to “Bad\_Data\_Folders”.
- Null values in the columns – If any of the columns in a file have all the values as NULL or missing, we discarded such a file and move it to “Bad\_Data\_Folders”.



# Q & A

Q1) What's the source of data?

The data is provided by an online E – Commerce Company.

Q2) What was the type of data?


The data was the combination of numerical and categorical values.

Q3) What's the complete flow you followed in this project?

Refer slide 4 for better understanding.

Q4) After the File validation what you do with incomplete files or files which didn't pass the validation?

Files like these are removed to the archive folder and a list of those files has been shared with the client and we removed the bad data folder.



Q5) How logs are managed?

We are using different logs as per the steps that we follow in validation and modelling like File Validation, Data insertion, Dashboard Creation etc.

Q6) What techniques were you using for data pre-processing?

- Removing unwanted attributes
- Visualizing relation of independent variables with each other and output variables
- Checking distribution of continuous values
- Cleaning data and inputting if null values are present
- Converting categorical data into numeric values
- Scaling the data

Q7) What are the different stages of deployment?

Once the dashboard is completed, we can share the dashboard through various sources like google sheets, Excel, Web Pages etc.





# Findings

## Cards:

- From the cards we can see that the total sales is 80,23,381 where the highest sales is in Fashion product category with 52,12,097 sales and the lowest in Electronics with 394738.
- From the cards we can see that the quantity sold is 1,45,521 where the highest quantity sold is in Fashion product category with 92071 quantity and the lowest in Electronics with 8211.
- From the cards we can see that the total profit is 37,29,903 where the highest profit is in Fashion product category with 2483853 profit and the lowest in Electronics with 174175.



# Findings cont..

## **Histogram:**

- From the histogram it is obvious that most number of orders are in the category 2 – 5 days with 19646 orders and lowest number of orders are in the category >10 with 4891 orders.

## **Monthly Sales & Profit:**

- The maximum sales and profit was obtained in the month December with 693037 sales and 523402 profit.
- The minimum sales and profit was obtained in the month February with 610240 sales and 286103 profit.
- When we look in the product category wise:
  - ❖ Electronics: Maximum sales (39240) & profit (17419) is in October and minimum sales (29644) and profit (12679) is in July and February respectively.



# Findings cont..

- ❖ Fashion: Maximum sales (452066) & profit (216460) is in December and minimum sales (397420) & profit (191681) is in February.
- ❖ Home & Furniture: Maximum sales (119095) & profit (53137) is in July and minimum sales (97361) & profit (42989) is in February.
- ❖ Auto & Accessories: Maximum sales (97347) & profit (43305) is in October and minimum sales (85683) and profit (38238) is in February and August respectively.

## Region wise Sales:

- The maximum sales was obtained in the Central region with 1735900 sales.
- The minimum sales was obtained in the Canada region with 60003 sales.
- When we look in the product category wise:

Product Category	Maximum Sales	Minimum Sales
Auto & Accessories	227929 (Central)	10382 (Canada)
Electronics	82750 (Central)	4602 (Canada)
Fashion	1146920 (Central)	33606 (Canada)
Home & Furniture	278301 (Central)	11413 (Canada)