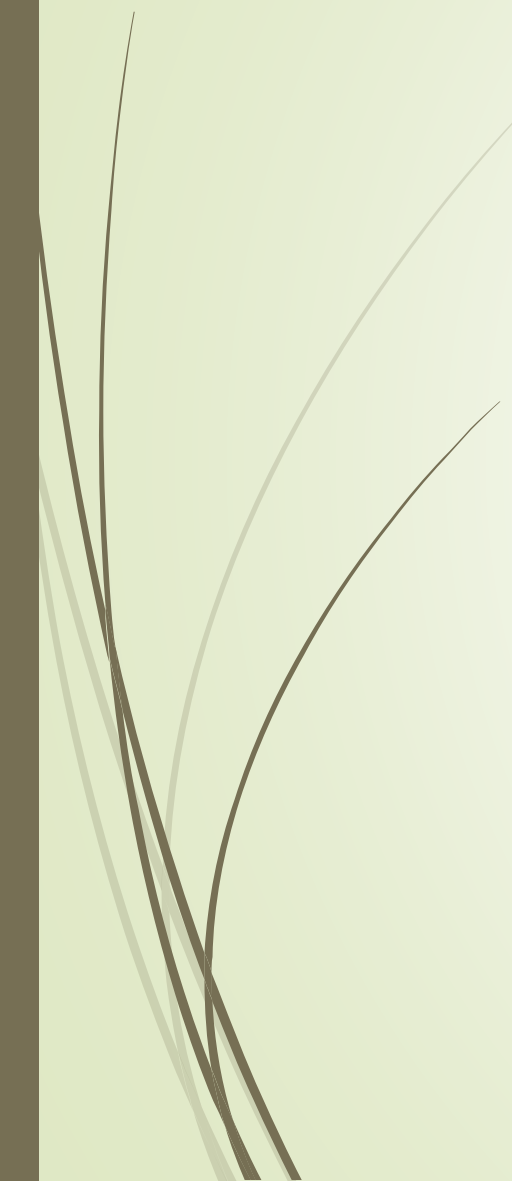# E – Commerce Data Analysis
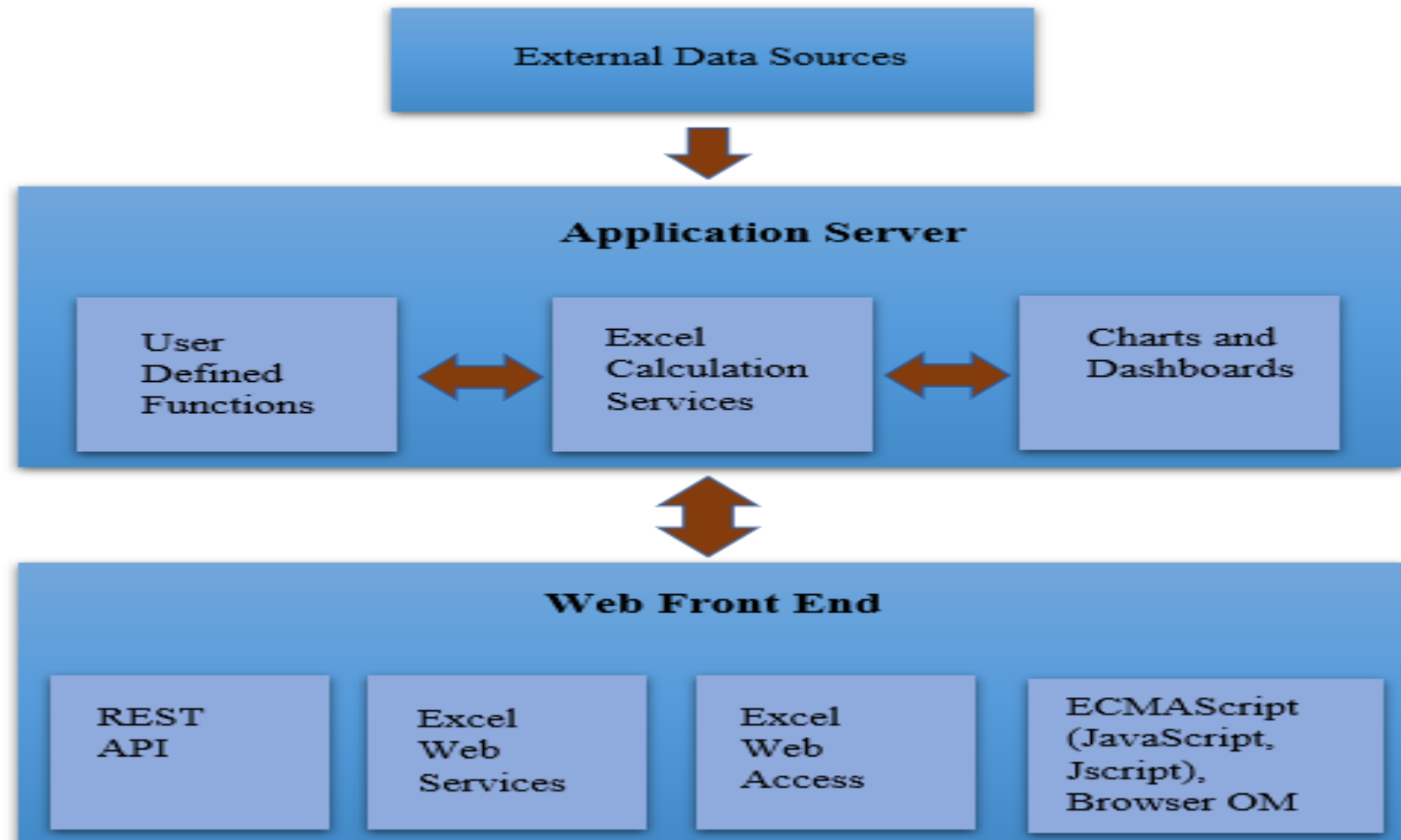
Objective:

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 reges 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, them 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.