

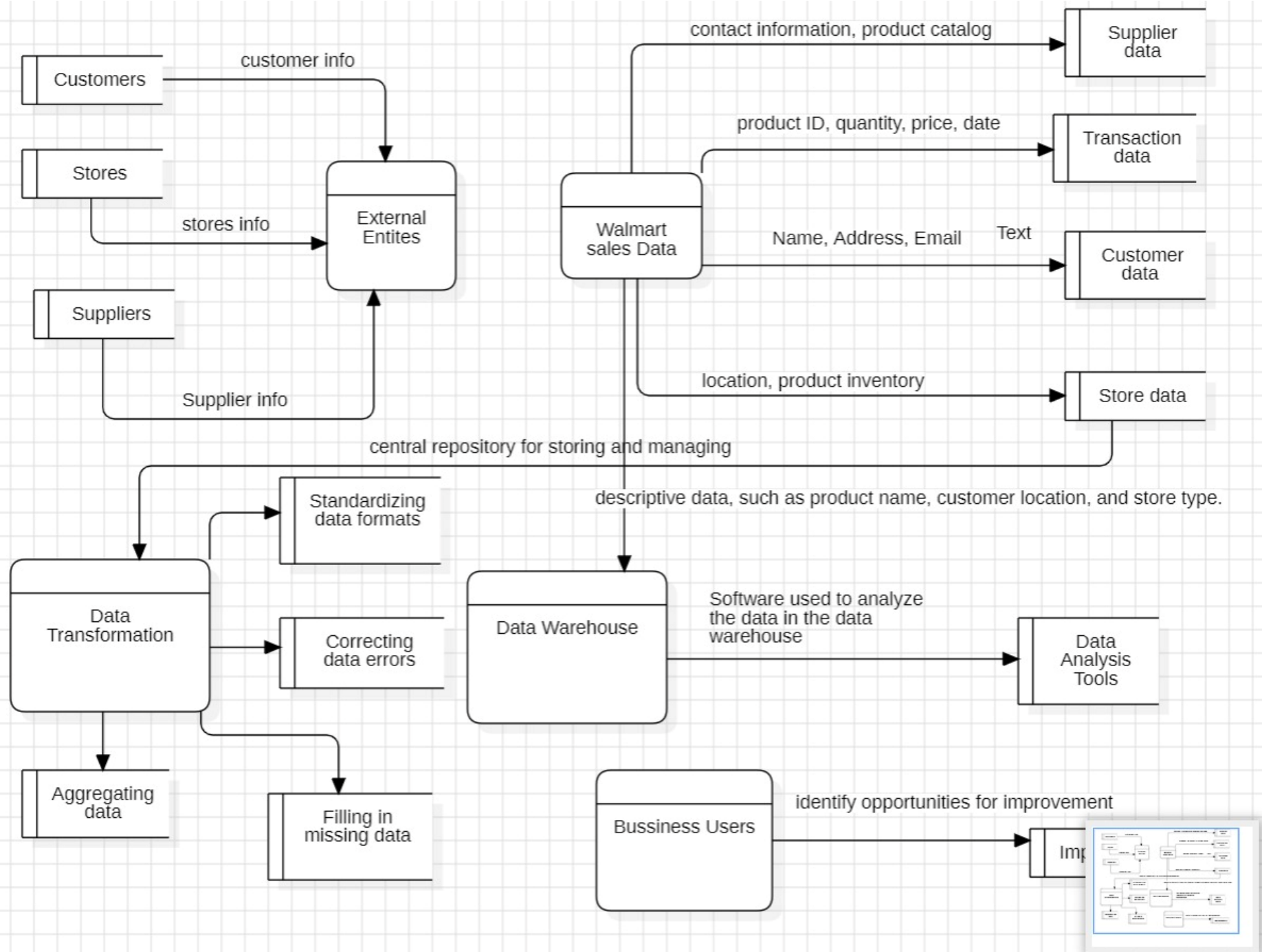
Project Design Phase-II Data Flow Diagram & User Stories

Date	23 October 2023
Team ID	Team-593170
Project Name	Walmart Sales Analysis for Retail Industry with Machine Learning
Maximum Marks	4 Marks

Data Flow Diagrams:

A Data Flow Diagram (DFD) is a traditional visual representation of the information flows within a system. A neat and clear DFD can depict the right amount of the system requirement graphically. It shows how data enters and leaves the system, what changes the information, and where data is stored.

The data flow diagram (DFD) for the Walmart sales analysis project illustrates a multi-level structure. At the top level (Level 0), the "Walmart Sales Analysis System" acts as the core, receiving data from "Walmart Sales Data," processing it, training machine learning models, and providing forecasts through a Flask web application. The machine learning models, at Level 1, interact with data stores for configuration and feature importance. The Flask web application, also at Level 1, communicates with users and passes their inputs for sales forecasting. Additionally, the system supports IBM Cloud deployment, allowing users to retrieve predictions. This DFD provides a concise visual representation of how data flows through various processes, entities, and data stores within the project, facilitating the understanding of the system's architecture and data pathways.



User Stories

Use the below template to list all the user stories for the product.

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Data Analyst	Data Preparation	USN-1	As a data analyst, I want to collect historical sales data for 45 Walmart stores.	The system should be able to gather and store sales data for all 45 stores in a structured format.	High	Sprint-1
		USN-2	As a data analyst, I need to preprocess the collected data, including handling missing values and outliers.	The system should successfully clean and preprocess the data, resulting in a high-quality dataset for analysis.	High	Sprint-1
	Holiday Impact Analysis	USN-3	As a data analyst, I want to identify weeks that include Christmas, Thanksgiving, Super Bowl, and Labor Day.	The system should create a holiday indicator variable that correctly identifies holiday weeks.	Medium	Sprint-1
		USN-4	As a data analyst, I need to analyze the impact of holidays on store sales.	The system should provide statistical insights and visualizations showing the influence of holidays on sales.	High	Sprint-1
	Sales Forecasting	USN-5	As a data analyst, I want to apply machine learning algorithms like Random Forest, Decision Tree, XgBoost, and ARIMA to forecast future sales.	The system should train and test these algorithms, providing accurate sales forecasts.	High	Sprint-2
	Deployment and Integration	USN-6	As a data analyst, I want to integrate the analysis and forecasting models into a Flask web application.	The system should create a user-friendly web interface for stakeholders to access the analysis and forecasts.	Medium	Sprint-3
		USN-7	As a data analyst, I need to deploy the Flask application on IBM Cloud for easy access and scalability.	The system should deploy the Flask application on the IBM Cloud platform, ensuring it is accessible to authorized users.	High	Sprint-3