Project Design Phase-I
Solution Architecture

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

Solution Architecture:

This solution architecture provides a robust and scalable framework for analyzing Walmart's sales data and making informed decisions about the impact of holidays on store sales. It combines data analysis, machine learning, web application development, and cloud deployment to deliver actionable insights to Walmart's retail business.

1. Data Ingestion and Storage:

- Data from 45 Walmart stores, including store information and monthly sales, should be ingested into a data storage system. This could be a relational database, data warehouse, or a data lake, depending on the volume and structure of the data.

2. Data Preprocessing:

- Data preprocessing is a critical step that involves cleaning, transforming, and aggregating the data. This step ensures that the data is ready for analysis and model training.

3. Exploratory Data Analysis (EDA):

-EDA involves using tools like Jupyter notebooks or data visualization libraries to explore the data and discover patterns and trends. EDA helps in feature selection and engineering.

4. Feature Engineering:

- Create new features, such as lag features (previous sales), holiday indicators, and store-specific characteristics, which are relevant for sales forecasting.

5. Machine Learning Models:

- Implement various machine learning models (Random Forest, Decision Trees, XGBoost, ARIMA) to predict sales. These models can be trained on historical data and validated using cross-validation techniques.

6. Model Evaluation:

- Use appropriate evaluation metrics (MAE, MSE, RMSE) to assess the performance of the machine learning models. Select the best-performing model(s) for sales forecasting.

7. Holiday Impact Analysis:

- Analyze the impact of holidays on sales using the trained models. Compare predicted sales during holiday weeks with non-holiday weeks.

8. Web Application:

- Create a web application using Flask to provide a user interface for interacting with the models. Users can input data, such as store information and dates, and receive sales forecasts.

9. Cloud Deployment:

- Deploy the Flask application and machine learning models on an cloud platform. This can be achieved using services like Cloud Foundry or Kubernetes for containerization.

10. Security:

- Ensure data security and access control. Implement authentication and authorization mechanisms to protect sensitive data and the application itself.

11. Documentation and Reporting:

- Create documentation for the solution architecture, data preprocessing steps, model training, and deployment procedures. Share insights through reports and dashboards.

12. Communication and Integration:

- Integrate the solution with Walmart's existing systems and communicate the results and insights to relevant stakeholders.

13. Scalability and Resilience:

- Design the architecture to be scalable to handle increasing data volume and user load. Implement redundancy and failover mechanisms to ensure system resilience.

14. Backup and Recovery:

- Implement regular backup and recovery procedures to protect data and the application against unexpected failures.

15. Feedback Loop:

- Implement a feedback loop to continuously improve the models based on new data and user feedback.

Solution Architecture Diagram: