**Proposed Architecture Diagram**

**Team Members:**

Saptajit Banerjee 20BCE1513

Kunal Gupta 20BCE1560

Prateek Karn 20BCE1771

**Description of proposed architecture diagram:**

**Data Collection:** The Website Firewall collects data from web servers, web applications, and database servers. The data includes web traffic logs, database logs, and other relevant information. This data is then stored in a central location.

**Preprocessing:** The collected data is preprocessed to remove any unnecessary information and to convert it into a format that is suitable for analysis. This includes data cleaning, data transformation, and data integration.

**Feature Extraction:** Relevant features are extracted from the preprocessed data. These features could include HTTP request headers, SQL queries, IP addresses, timestamps, and other relevant data points.

**Machine Learning Model:** The extracted features are then used to train a machine learning model to detect SQL injection attacks. The machine learning model uses a supervised learning model which is logistic regression. Logistic regression is a statistical method used for binary classification problems, where the goal is to predict the probability of an event occurring or not occurring based on a set of input features. In the case of SQL injection attack detection, the binary classification problem is whether a web request is normal or potentially malicious. One advantage of using logistic regression for SQL injection attack detection is that it is a simple and interpretable model that can effectively capture non-linear relationships between input features and the target variable. Another advantage is that it is computationally efficient and can handle large-scale datasets, which is essential for processing web traffic logs in real-time. However, one disadvantage of logistic regression is that it assumes that the input features are independent of each other, which may not always be the case in practice. Additionally, it may not perform well when there are highly imbalanced classes or when the data is not linearly separable. Overall, logistic regression is a suitable choice for SQL injection attack detection in web applications, but its performance depends on the quality of the input features and the characteristics of the dataset.

**Deployment:** The trained machine learning model is then deployed as part of the Website Firewall to detect SQL injection attacks in real-time. The Website Firewall analyzes incoming web traffic and compares it with the trained machine learning model to detect SQL Injection attacks.

**Detection and Blocking:** If the Website Firewall detects any SQL injection attacks, it blocks the attack by not processing the malicious request and loads a page to confirm the detection and blocking of the SQL Injection attacks.

