**Project name: Anomalies Detection in API-logs Using Spark**

Project statement: This project entails creating a Spark application for Apache Access logs, aiming to detect anomalies in API logs. It includes features for analyzing response codes, traffic, and content size statistics, covering top endpoints, frequent visitors, and bad requests. The submission involves providing Pyspark code to implement these functionalities.

Project Description: The project focuses on developing a Spark application for analyzing Apache Access logs and detecting anomalies in API logs. The application includes features to extract information from semi-structured parquet logs, analyzing trends based on response codes, traffic, frequent visitors, top endpoints, content flow, etc., to identify anomalies. The use cases cover calculating statistics related to content size, such as top endpoints transferring maximum content, top visited endpoints, daily visited content size, and statistical information like minimum, maximum, and count of content size. The project also involves response code analysis, identifying IP addresses accessing the server more than 10 times, and analyzing bad requests, including the top 10 latest 404 requests with their endpoints and time. The submission includes Pyspark code for implementing these functionalities.

To apply transformations & actions of pyspark on these parameters/features where It includes features for analyzing response codes, traffic, and content size statistics, covering top endpoints, frequent visitors, and bad requests