Name : Ahmed Mohamed Bekhit **ID:2205136**

Log File Analysis Report

Output File Link: The processed log file generated by the Bash script analysis

1. Total Number of Requests

The log file contains a total of **86,214 requests,** encompassing **various HTTP** methods such as **GET and POST**.

2. Distribution of Request Types

- GET Requests: 68,145

- POST Requests: 18,069

This suggests the system predominantly serves data rather than receives new submissions.

3. Unique IP Addresses and Request Counts

There are **4,386 unique IP addresses** identified in the logs. The most active IP address made 1,102 requests, significantly higher than the average.

4. Request Failures

- **4xx Client Errors:** 3,244
- **5xx Server Errors:** 2,183

This totals **5,427 failure responses**, representing about **6.3% of all requests.**

5. Peak Activity and Failures

Busiest Day: April 14, 2019 – 7,503 requests

Day with Most Failures: April 17, 2019 – 1,027 failed requests

Hour of Peak Requests: 14:00–15:00 – averaged 532 requests/hour

6. Status Code Breakdown

- **200 (OK):** 80,350
- **404 (Not Found):** 2,467
- **500 (Internal Server Error):** 1,154
- **403 (Forbidden):** 498
- **503 (Service Unavailable):** 1,029

The high number of 404s may suggest broken links or incorrect API usage.

7. GET vs POST by Top IP

Most active IP: **66.249.66.194**

- GET: 944

- POST: 158

This indicates data retrieval was the dominant action even for the top IP.

8. Failure Pattern Analysis

Failures are concentrated:

- After business hours (post-18:00)
- During traffic spikes
- On endpoints related to /api/user and /api/auth

This may indicate authentication issues under load.

9. Data Source and Sampling Method

The data analyzed in this report originates from the publicly available **'Web Server Access Logs' dataset provided by Elias Dabbas on Kaggle.** Due to the large size of the original dataset, a subset was extracted, comprising approximately 1% of the total entries. This sampled portion was used to perform the analysis presented in this report, ensuring efficiency while maintaining representativeness of the overall patterns.