

# 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 **6,705 requests**, encompassing various HTTP methods such as **GET** and **POST**.

## 2. Distribution of Request Types

- GET Requests: **6,583 (98.2%)**

- POST Requests: **105 (1.6%)**

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

## 3. Unique IP Addresses and Request Counts

There are **2,884 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. Traffic Patterns

**Most active hours:** 9-11 AM (local time +0330)

**Highest traffic** at 11 AM with 1,604 requests

**Single day of log data** (January 22, 2019)

## 6- User Behavior

**Most active IP:** 66.249.66.194 (**216 requests**) - Google bot

**High number of unique visitors** (2,884) with most making **only 1-2 requests**

**Low POST request count suggests limited form submissions/interactions**

## 7- Failure

**Failure rate:** 1.58% (106 requests)

**Status code breakdown:**

**200 OK:** 6,302 (94%)

302 Found: 142

304 Not Modified: 114

404 Not Found: 60

Other errors: 87 (including 499, 403, 500, 408)

Most errors occurred during peak hours (10-11 AM)

## Recommendations

### Bot Traffic Management:

- Use robots.txt to guide search engines and bots on what they can access.
- Limit how often bots can visit your site (Googlebot was very active).

### Fix Errors:

- Check the 60 "Page Not Found" (404) errors – fix broken links or missing pages.
- Look into 499 errors (users leaving before the page loads) – maybe your site is too slow.
- Fix the 2 "Server Error" (500) errors – something's wrong with the server.

### Make the Site Faster:

- Focus on improving speed during busy hours (9-11 AM).
- Use a CDN (Content Delivery Network) to load images/files faster.
- Enable caching so repeat visitors get pages quicker.

### Better Monitoring:

- Set up alerts for sudden errors or traffic spikes.
- Check where users are located (geographic data).
- Track which browsers/devices people use.

### Improve Log Analysis:

- Check which bots/visitors are accessing your site.
- Find out which pages load slowly.
- Use graphs to see traffic trends.

### Security Upgrades:

- Watch for strange visitor behavior.
- Limit how often an IP can request pages (stop abuse).
- Scan for hacking attempts (like fake login forms or harmful code).

### Understand Users Better:

**Name : Ahmed Mohamed Bekhit**

**ID:2205136**

- See which pages are most popular.
- Track how users move through your site.
- Check if visits lead to sales/sign-ups (if applicable).

## **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.