# Log File Analysis

Abdelrahman Elmoghazy

2205032

# 1. Introduction

This report examines an Apache web server log file (apache_logs) using a Bash script to extract key metrics on requests, IPs, and failures. The objective is to identify patterns, detect issues, and propose enhancements for server performance and security. The log spans **10,000 requests** over **4 days**.

---

# 2. Analysis Results

The following metrics summarize the log file (**10,000 requests**):

## 2.1 Request Counts

- Total requests: **10,000**
- GET requests: **9,952 (99.52%)**
- POST requests: **5 (0.05%)**

## 2.2 Unique IP Addresses

- Most active IP: **66.249.73.135**, with **482 GET requests (4.82%)**

## 2.3 Failed Requests

- 4xx/5xx errors: **220**
- Failure rate: **2.00%**

## 2.4 Most Active User

- IP: **66.249.73.135**, with **482 requests**

## 2.5 Daily Request Average

- Average: **2,500 requests per day** (over 4 days)

## 2.6 Failure Analysis by Day

- May 18, 2015: **66 failures**
- May 19, 2015: **66 failures**
- May 20, 2015: **58 failures**

### 2.7 Requests by Hour

- Highest: **14:00, with 498 requests**
- Lowest: **08:00, with 345 requests**
- Peak period: **12:00–20:00**

### 2.8 Request Trends

- Peak hour: **14:00, with 498 requests**
- Trend: **Afternoon activity**

### 2.9 Status Code Breakdown

- 200 (OK): **9,126 requests**
- 304 (Not Modified): **445 requests**
- 404 (Not Found): **213 requests**
- 301 (Redirect): **164 requests**
- 206 (Partial Content): **45 requests**
- 500 (Server Error): **3 requests**
- 416 (Range Not Satisfiable): **2 requests**
- 403 (Forbidden): **2 requests**

### 2.10 Most Active IP by Method

- GET: **66.249.73.135, with 482 requests**
- POST: **78.173.140.106, with 3 requests**

### 2.11 Failure Patterns

- Top hours: **09:00 (18 failures)**, **05:00 (15 failures)**, **06:00 (14 failures)**

---

## 3. Analysis and Trends

Key insights from the log:

- **Requests**: Predominantly GET (**99.52%**), indicating static content delivery. Minimal POST requests (**5**) suggest low interactivity.
- **Active IP**: **66.249.73.135** (likely Googlebot) with **482 GET requests**, typical for web crawlers.

- **Failures**: A **2% failure rate**, with **213 404 errors** (broken links) and **3 500 errors** (server issues) requiring attention.
- **Daily Patterns**: High failures on **May 18–19, 2015 (66 each)**, possibly due to maintenance or traffic spikes.
- **Hourly Trends**: Peak at **14:00 (498 requests)**, lowest at **08:00 (345 requests)**. Afternoon (**12:00–20:00**) is the busiest period.
- **Failure Times**: Early morning (**05:00–09:00**) sees most failures, likely from automated scripts or maintenance.

---

# 4. Suggestions

Recommendations to improve performance and security:

## 4.1 Reduce Failures

- Fix **404 errors** using **wget -spider** to identify broken links.
- Investigate **500 errors** in **/var/log/apache2/error.log**.

## 4.2 Monitor High-Failure Days

- Review logs for **May 18–19, 2015**, to identify causes of elevated failures.

## 4.3 Optimize Peak Hours

- Allocate additional resources for **12:00–20:00**, especially at **14:00**.

## 4.4 Enhance Security

- Rate-limit **IP 66.249.73.135** if necessary, using **robots.txt**.
- Secure forms with **CSRF tokens** and **CAPTCHA**.

## 4.5 Performance Improvements

- Implement caching with **Varnish** or **Cloudflare**.
- Deploy a **Web Application Firewall (WAF)** for early morning traffic.
- Monitor server health using **Prometheus** or **Nagios**.

---

## 5. Conclusion

The Apache server log analysis indicates a stable system with a **2% failure rate**. GET requests dominate (**99.52%**), with minimal POST activity. Addressing **404 errors**, resolving **500 errors**, and optimizing for peak hours (**12:00–20:00**) will enhance reliability. Implementing **caching**, **monitoring**, and **security measures** will further improve performance and protect against potential issues.