# Log File Analysis Task

Name: Kerlous Nasser Shehata

ID: 2205039

# 1 Introduction

This report analyzes a web server log file using a Bash script, addressing the Log File Analysis Task requirements. It provides statistics on requests, IPs, failures, and trends, with suggestions for improvement.

# 2 Analysis Results

## 2.1 Requirements

1. **Request Counts:**
   - Total Requests: 10,000
   - GET Requests: 9,952 (99.52%)
   - POST Requests: 5 (0.05%)

   Mostly GET requests, showing data retrieval focus.

2. **Unique IP Addresses:**
   - Total Unique IPs: 1,753
   - GET and POST Requests per IP: Most IPs made <10 requests; top IPs:
     - 66.249.73.135: 482 GET
     - 46.105.14.53: 364 GET
     - 130.237.218.86: 357 GET

   Diverse IPs, few heavy users.

3. **Failure Requests:**
   - Failed (4xx/5xx): 220
   - Percentage: 2.20%

   Low failure rate, but needs review.

4. **Top User:**
   - Most Active IP: 66.249.73.135 (482 GET)

   Likely a crawler, impacting load.

5. **Daily Request Averages:**
   - Average/Day: 2,500

   Consistent traffic over 4 days.

6. **Failure Analysis:**
   - Days with Most Failures:
     - 19 May 2015: 66

- 18 May 2015: 66
- 20 May 2015: 58
- 17 May 2015: 30

High failures on 18–19 May need investigation.

## 2.2 Additionally:

- **Request by Hour:** Peaks at 14:00–15:00 (498–496 requests).

Table 1: Requests by Hour

| Hour | Requests |
|---|---|
| 00 | 361 |
| 01 | 360 |
| 02 | 365 |
| 03 | 354 |
| 04 | 355 |
| 05 | 371 |
| 06 | 366 |
| 07 | 357 |
| 08 | 345 |
| 09 | 364 |
| 10 | 443 |
| 11 | 459 |
| 12 | 462 |
| 13 | 475 |
| 14 | 498 |
| 15 | 496 |
| 16 | 473 |
| 17 | 484 |
| 18 | 478 |
| 19 | 493 |
| 20 | 486 |
| 21 | 453 |
| 22 | 346 |
| 23 | 356 |

- **Request Trends:** Requests rise from 09:00, peak at 14:00–15:00, drop after 20:00, indicating workday activity.
- **Status Codes Breakdown:**
  - 200: 9,126
  - 304: 445
  - 404: 213
  - 301: 164

  - 206: 45
  - 500: 3
  - 416: 2
  - 403: 2

  High 404 errors suggest missing resources.

- **Most Active User by Method:**
  - Top IP: 66.249.73.135 (482 GET)

  POST requests too few to analyze.

- **Patterns in Failure Requests:** Failures peak at 09:00 (18), 05:00 (15), 06:00 (14).

Table 2: Failures by Hour

| Hour | Failures |
|---|---|
| 00 | 6 |
| 01 | 10 |
| 02 | 10 |
| 03 | 7 |
| 04 | 9 |
| 05 | 15 |
| 06 | 14 |
| 07 | 7 |
| 08 | 2 |
| 09 | 18 |
| 10 | 12 |
| 11 | 11 |
| 12 | 7 |
| 13 | 12 |
| 14 | 11 |
| 15 | 6 |
| 16 | 8 |
| 17 | 12 |
| 18 | 9 |
| 19 | 10 |
| 20 | 4 |
| 21 | 8 |
| 22 | 8 |
| 23 | 4 |

# 3 Analysis Suggestions

## 3.1 Reduce Failures

- Fix 404s: Audit broken links; add redirects.

- Check 500s: Review logs for server issues on 18–19 May.
- Adjust Maintenance: Move tasks from 05:00–09:00 to 03:00.

### 3.2 Monitor Key Times

- 18–19 May: Investigate high failures.
- 10:00–20:00: Balance load during peaks.
- 05:00–09:00: Set error alerts.

### 3.3 Security Concerns

- Heavy IPs: Limit crawlers like 66.249.73.135 with robots.txt.
- Suspicious IPs: Block 180.76.5.x/6.x if malicious.
- POSTs: Confirm low count; monitor for attacks.

### 3.4 System Improvements

- Cache GETs: Use CDN.
- Enhance Monitoring: Add real-time alerts.
- Scale Servers: Handle 14:00–15:00 peaks.
- Improve Errors: Create user-friendly 404 pages.

## 4 Conclusion

The server is stable (2.20% failures, 2,500 requests/day). Failures spike on 18–19 May and 05:00–09:00. GETs dominate; 66.249.73.135 is top user. Fix 404s, limit crawlers, cache content, and monitor peaks for better performance.