Log File Analysis

Abdelrahman Elmoghazy

2205032

1 Introduction

This report analyzes an Apache web server log file (apache_logs) using a Bash script to extract key metrics on requests, IPs, and failures. The goal is to identify patterns, detect issues, and suggest improvements for server performance and security. The log covers 10,000 requests over 4 days.

2 Statistics

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

1. Request Counts:
   - Total: 10,000
   - GET: 9,952 (99.52%)
   - POST: 5 (0.05%)
2. Unique IP Addresses:
   - Most Active: 66.249.73.135 (482 GET requests, 4.82%).
3. Failed Requests:
   - 4xx/5xx: 220 • Failure Rate: 2.00%
4. Most Active User:
   - IP: 66.249.73.135 (482 requests).
5. Daily Request Average:
   - 2,500 requests/day (4 days).
6. Failure Analysis by Day:
   - May 19, 2015: 66 failures
   - May 18, 2015: 66 failures

- May 20, 2015: 58 failures

7. Requests by Hour:
   - Highest: 14:00 (498 requests)
   - Lowest: 08:00 (345 requests)
   - Peak: 12:00–20:00.
8. Request Trends:
   - Peak Hour: 14:00 (498 requests)
   - Trend: Afternoon activity.
9. Status Code Breakdown:
   - 200 (OK): 9,126
   - 304 (Not Modified): 445
   - 404 (Not Found): 213
   - 301 (Redirect): 164
   - 206 (Partial Content): 45
   - 500 (Server Error): 3
   - 416 (Range Not Satisfiable): 2
   - 403 (Forbidden): 2
10. Most Active IP by Method:
    - GET: 66.249.73.135 (482)
    - POST: 78.173.140.106 (3)
11. Failure Patterns:
    - Top Hours: 09:00 (18 failures), 05:00 (15), 06:00 (14).

3 Analysis

Key insights from the log:

- Requests: Mostly GET (99.52%), indicating static content. Few POST requests (5) suggest low interactivity.
- Active IP: 66.249.73.135 (likely Googlebot) makes 482 GET requests, typical for crawlers.
- Failures: 2% failure rate; 213 404 errors (broken links) and 3 500 errors (server issues) need attention.
- Daily Patterns: High failures on May 18–19, 2015 (66 each), possibly from maintenance or traffic.
- Hourly Trends: Peak at 14:00 (498 requests), low at 08:00 (345). Afternoon (12:00–20:00) is busiest.
- Failure Times: Early morning (05:00–09:00) sees most failures, likely from scripts or maintenance.

4 Suggestions

Recommendations to improve performance and security:

1. Reduce Failures:
   - Fix 404 errors with wget –spider.
   - Check /var/log/apache2/error.log for 500 errors.
2. Monitor High-Failure Days:
   - Review May 18–19, 2015, logs for issues.
3. Peak Hours:
   - Add resources for 12:00–20:00, especially 14:00.

4. Security:
   - Rate-limit IP 66.249.73.135 if needed (robots.txt).
   - Secure forms with CSRF and CAPTCHA.
5. Improvements:
   - Use caching (Varnish/Cloudflare).
   - Deploy WAF for early morning traffic.
   - Monitor with Prometheus/Nagios.

5 Conclusion

The apache_logs analysis shows a stable server with a 2% failure rate. GET requests dominate, with minimal POST activity. Fixing 404 errors, addressing 500 errors, and optimizing for peak hours will boost reliability. Caching, monitoring, and security measures will enhance performance and protect against issues.