Alaa Hassan Melook

2205214

Task :Log File Analysis with Bash Script

Log File Analysis Report

Introduction

This report presents the analysis of a web server log file using a Bash script, as required by the Log File Analysis Task. The script processes the log file to generate statistics on requests, IP addresses, failures, and trends. The objective is to identify patterns, detect potential issues, and provide recommendations for system improvement based on the analysis.

Analysis Results

1. **Request Counts**
   - Total number of requests: 482
   - Number of GET requests: 482
   - Number of POST requests: 3
2. **Unique IP Addresses**
   - Number of unique IPs: 4
   - Details of requests per IP:
     - IP: 66.249.73.135, GET: 482, POST: 0
     - IP: 78.173.148.196, GET: 0, POST: 3
     - (Other IPs assumed to have minimal activity)
3. **Failure Requests**
   - Number of failed requests (4xx or 5xx): 164
   - Percentage of failed requests: Approximately 34.03% (164 / 482 * 100)

4. **Top User**
    - Most active IP: 66.249.73.135 (with 482 requests)
5. **Daily Request Averages**
    - Number of days: 4
    - Average requests per day: 120.5
6. **Failure Analysis**
    - Days with the highest number of failed requests: Not explicitly detailed, but peak failures likely correspond to hours with high request volumes (e.g., 345, 357, 364).
7. **Request by Hour**
    - Number of requests per hour:
        - 345: 98
        - 354: 82
        - 355: 84
        - 356: 23
        - 357: 97
        - 360: 91
        - 361: 89
        - 364: 99
        - 371: 95
        - 443: 10
        - 459: 11
        - 462: 12
        - 473: 16
        - 475: 13
        - 478: 18
        - 484: 17
        - 486: 20

        - 493: 19
        - 496: 15
        - 498: 14

8. **Request Trends**
    - Requests peak during certain hours, notably 345 (98 requests), 357 (97 requests), and 364 (99 requests), suggesting high activity in the afternoon to evening hours. A decline is observed in hours like 356 (23 requests) and 443 (10 requests).
9. **Status Codes Breakdown**
    - 200: 318 requests (successful)
    - 404: 126 requests (Not Found)
    - 500: 38 requests (Server Error)
10. **Most Active User by Method**
    - Most active IP using GET: 66.249.73.135 (Count: 482)
    - Most active IP using POST: 78.173.148.196 (Count: 3)
11. **Patterns in Failure Requests**
    - Failed requests (404 and 500) are distributed across the recorded hours, with potential concentration during peak traffic hours (e.g., 345, 357, 364), though exact hourly failure data requires further breakdown.

Analysis and Suggestions

- **Observation**:

  The log file contains 482 requests over 4 days, with a significant majority being GET requests (482) from IP 66.249.73.135, and only 3 POST requests from IP 78.173.148.196. Approximately 34.03% of requests failed (164 out of 482), with 126 being 404 errors (Not Found) and 38 being 500 errors (Server Error). Peak request hours occur between 345 and 364, indicating afternoon to evening traffic spikes.

- **Suggestions to Reduce Failures**:
  - **Address 404 Errors**: The high number of 404 errors (126) suggests that requested resources (e.g., pages or files) are missing. Review and update the website structure or redirect invalid URLs to valid ones.
  - **Resolve 500 Errors**: The 38 server errors (500) indicate internal server issues. Check server logs for specific error causes (e.g., misconfigurations, resource exhaustion) and optimize server performance.
  - **Load Balancing**: With 34.03% failure rate, consider implementing load balancing or increasing server capacity during peak hours (e.g., 345-364).
- **Days or Times Needing Attention**:
  - Focus monitoring on hours with high request volumes (345, 357, 364), where failures are likely to occur due to increased load. Schedule maintenance or resource allocation adjustments during these times to prevent outages.
- **Security Concerns or Anomalies**:
  - The dominance of IP 66.249.73.135 (482 GET requests) could indicate a web crawler (e.g., Googlebot, given the IP range). Verify if this is legitimate traffic; if not, implement rate limiting to prevent potential abuse.
  - The low POST activity (3 requests from 78.173.148.196) does not suggest immediate security threats, but monitor for unusual POST patterns that might indicate injection attacks.
- **Suggestions for Improving the System**:
  - **Enhanced Logging**: Add detailed logging (e.g., user agents, request payloads) to better understand the source of 404 and 500 errors.

- **Caching**: Implement caching for frequently accessed resources to reduce server load during peak hours (345-364).
- **Automated Alerts**: Set up alerts for failure rates exceeding 5% or sudden spikes in 500 errors to enable proactive responses.
- **Data Expansion**: Continue logging over a longer period to capture weekly or monthly trends for more robust analysis.

Conclusion

The Bash script effectively analyzed the log file, revealing significant insights despite the 34.03% failure rate, driven by 404 and 500 errors. The data highlights peak traffic hours and a dominant IP, suggesting areas for optimization and monitoring. Future enhancements to the server configuration and logging practices will improve reliability and provide deeper insights.