Name:mohamed ayman mohamed

ID: 2205045

Log File Analysis Report

Log File Analysis Report

1. Summary of Key Results

Log file analyzed: `access.log`
Analysis duration: 5 days
Source file link: (
https://www.kaggle.com/datasets/adchatakora/nasa-http-access-logs?resource=download

)

---

2. Request Counts

- Total Requests: 10,365,152
- GET Requests: 10,190,005
- POST Requests: 139,155

---

3. Unique IP Addresses

- Total Unique IPs: 258,606
- Top 5 IPs with Most GET Requests:
    - 66.249.66.194: 353,483
    - 66.249.66.91: 314,522
    - 66.249.66.92: 88,332
    - 151.239.241.163: 80,201
    - 104.222.32.91: 41,530
- Top 5 IPs with Most POST Requests:
    - 151.239.241.163: 11,712
    - 91.99.30.32: 4,996
    - 91.99.47.57: 4,091
    - 5.78.190.233: 3,461
    - 5.117.116.238: 1,095

---

4. Failure Requests

- Total Failures (4xx/5xx): 177,424
- Failure Rate: 1.71%

---

5. Most Active IP Address

- 66.249.66.194 with 353,483 requests

---

6. Daily Request Averages

- Days Covered: 5
- Average Requests/Day: 2,073,030.40

---

7. Days with Highest Failures

- 26/Jan/2019: 45,293
- 23/Jan/2019: 37,237
- 22/Jan/2019: 33,487
- 24/Jan/2019: 32,657
- 25/Jan/2019: 28,750

---

8. Requests per Hour

- Highest traffic at 11:00 AM (731,595 requests)
- Lowest traffic at 03:00 AM (79,133 requests)

---

9. Status Code Breakdown

- 200: 9,579,824
- 304: 340,228
- 302: 199,835
- 404: 105,011
- 499: 50,852
- 500: 14,266
- 403: 5,634
- Others: 6,502 (various)

---

10. Failure Patterns by Hour

- Most Failures:
    - 19:00 – 14,852 failures
    - 18:00 – 13,795
    - 14:00 – 11,181
    - 15:00 – 11,094
    - 12:00 – 10,824

---

11. Request Trends

- Traffic rises from 07:00 and peaks at midday
- Highest failures during business hours (10:00 to 19:00)

---

12. Most Active IPs by Method

- GET: 66.249.66.194
- POST: 151.239.241.163

---

13. Insights and Recommendations

- High request volume from a few IPs (likely crawlers or automated systems). Consider rate-limiting or bot detection.
- Failures peak during working hours—check server capacity, error logs, and backend services around these times.
- Status 404 and 500 errors are relatively high. Improve routing and error handling.
- Monitor IP 151.239.241.163 for possible abuse due to high GET and POST activity.
- Schedule maintenance or deploy updates during low traffic hours (02:00–06:00).

---

14. Suggested Actions

- Implement IP rate limits or bot filters
- Monitor backend service health during 10:00–19:00
- Review logs on 26/Jan/2019 for failure root causes
- Audit routes causing 404 and backend components causing 500

Based on the analysis, the following recommendations address failures, performance, and security:

1. **Reducing Failures**

    - **404 Errors (9,978 occurrences)**: Audit the website for broken links and missing resources. Implement redirects for deprecated URLs and ensure content is properly maintained.

    - **Peak Hour Failures (Hours 12–13)**: Scale server capacity during peak hours (12:00–15:00) using load balancing or cloud-based resources to handle high traffic.

    - **Hour 02 Anomalies**: Investigate high failures during low-traffic Hour 02. This could indicate misconfigured scripts, bots, or maintenance tasks causing errors.

2. **Days/Times Needing Attention**

- **End-of-Month Failures (30–31 August)**: Monitor system performance at month-end, as increased failures suggest higher traffic or system strain. Schedule maintenance outside these periods.

- **Peak Hours (12:00–15:00)**: Optimize server performance during these hours by caching static content and prioritizing critical requests.

3. **Security Concerns and Anomalies**
   - **High Activity from Single IP (edams.ksc.nasa.gov, 6,530 requests)**: Investigate this IP's behavior to determine if it's a legitimate user (e.g., NASA crawler) or a potential bot. Implement rate-limiting for IPs exceeding a request threshold.

   - **POST Request IPs**: Monitor IPs making POST requests (e.g., seabrk.mindspring.com), as these are rare and could indicate form submissions or API interactions. Ensure POST endpoints are secure against abuse.

   - **Unusual Status Codes (e.g., 786, 669)**: Investigate non-standard status codes to confirm they are intentional or identify misconfigurations.

4. **System Improvements**
   - **Content Delivery Network (CDN)**: Deploy a CDN to reduce server load during peak hours and improve response times for global users.

- **Logging Enhancements**: Add more granular logging (e.g., request paths, user agents) to better diagnose 404 errors and anomalous failures.

- **Automated Monitoring**: Implement real-time monitoring for failure spikes and alert administrators during high-failure periods (e.g., Hour 02 or 30–31 August).

The log_file source

https://www.kaggle.com/datasets/adchatakora/nasa-http-access-logs?resource=download