Log File Analysis Report

Ziad Ibrahim elbatal
2205075

The purpose of this task is to analyze a web server access log file using a Bash script and extract meaningful insights related to request patterns, failures, and user behavior. This helps in identifying potential issues, usage trends, and improvement opportunities for the server or web service.

Log File Used

https://raw.githubusercontent.com/elastic/examples/master/Common%20Data%20Formats/apache_logs/apache_logs

saved as access.log.

Tools and Environment

- **Operating System**: Kali Linux (running in VMware)
- **Script Language**: Bash
- **Terminal Tools**: awk, grep, cut, sort, uniq, wc, date

Analysis Results

1. Request Counts

Type	**Count**
Total	**10,000**
GET Requests	**9,952**
POST Requests	**5**

---

2. Unique IP Addresses

- **Total Unique IPs: 1,753**

---

3. Failure Requests

- **Failed Requests: 220**
- **Failure Rate: 2.20%**
- **(Failure = status codes in 4xx or 5xx range)**

---

4. Most Active IP

IP Address	Requests
66.249.73.135	482

---

5. Average Requests per Day

- **Daily Average: 2,500.00**
- **Based on 4 days of data**

---

6. Failures by Hour

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

---

7. Requests by Hour

00	361	06	366	12	462	18	478

01 360	07 357	13 475	19 493
02 365	08 345	14 498	20 486
03 354	09 364	15 496	21 453
04 355	10 443	16 473	22 346
05 371	11 459	17 484	23 356

---

8. Request Trend (Day-wise)

Date	Total Requests
17/May/2015	1,632
18/May/2015	2,893
19/May/2015	2,896
20/May/2015	2,579

---

9. Status Code Breakdown

Status Code	Count
200	9,126
206	45
301	164
304	445
403	2
404	213
416	2
500	3

---

10. Most Active IPs by Request Type

Type	IP Address	Count
GET	66.249.73.135	482
POST	78.173.140.106	3

Observations & Recommendations

Failure Patterns

- **Failures are fairly evenly spread, but slightly higher around:**
  - **05:00 – 06:00 and 09:00 – 10:00**
- **404 errors (not found) are the dominant failure type.**

Request Trends

- **Peak traffic: 14:00 – 20:00**
- **Daily peak: 18 and 19 May**

Recommendations

- **Reduce 404s: Audit missing pages or broken links.**
- **Traffic Monitoring: Focus on peak hours**
- **Load Testing: Simulate peak traffic during 14:00–20:00 to ensure server stability.**
- **Security: One IP made nearly 500 GET requests**

Summary

Metric	Value
Total Requests	10,000
GET Requests	9,952
POST Requests	5

Metric	**Value**
Unique IPs	**1,753**
Failed Requests	**220 (2.20%)**
Most Active IP	**66.249.73.135 (482 requests)**
Average Requests per Day	**2,500**

---

1. **Security Insight**:
    - IP 192.168.1.4 made a high number of requests — may need rate limiting.
    - 404s from unusual paths could indicate scanning activity (possible bot).
2. **Improvement Ideas**:
    - Use caching for frequent GET requests.
    - Log more user-agent info for better client understanding.
    - Monitor and alert on rising 500/403 errors in real time.
