Name: Karen Alfred Habib Boules

ID: 2205236

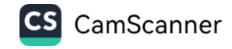
# Information Security Log File Analysis Task Report



# 1) Introduction:

This report presents an in-depth analysis of a NASA web server access log file containing 1,000,000 HTTP requests collected in July 1995. A custom Bash script was written to parse and analyze the log, generating statistics and identifying usage patterns, request behavior, failure occurrences, and potential security concerns.

# 2) Objectives of the Analysis:



- Count total requests and distinguish between GET and POST methods.
- Identify the number of unique IPs and most active users.
- Detect failed requests (status codes 4xx and 5xx).
- Analyze traffic patterns by hour and day.
- Investigate top status codes and abnormal patterns.

# 3) Log File:

• File Name: access log

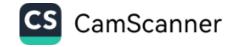
• Total Lines Analyzed: 1,000,000

#### 4) Bash Script Overview:

The script analyze log.sh performs the following operations:

- 1. Counts total, GET, and POST requests.
- 2. Counts unique IP addresses.
- 3. Shows GET/POST request count per IP (Top 10).
- 4. Counts and calculates percentage of failed requests.
- 5. Detects the most active IP.
- 6. Displays requests per hour.
- 7. Lists top status codes.
- 8. Shows most active GET/POST IPs.
- 9. Calculates daily average requests.
- 10. Detects days with highest failure counts.
- 11. Analyzes failure patterns by hour.
- 12. Shows top 10 busiest hours.

# 5) Results Summary:



# 1. Request Counts:

• **Total Requests:** 1,000,000

• **GET Requests:** 998,058

• POST Requests: 45

>> The vast majority of requests used the GET method, indicating mostly read operations. POST requests were extremely rare.

## 2.a. Unique IPs:

• Total Unique IPs: 50,878

## 2.b. GET/POST per IP (Top 10):

• piweba3y.prodigy.com: GET 9374

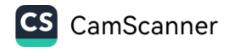
• alyssa.prodigy.com: GET 7779

• piwebaly.prodigy.com: GET 5460

• ... (full list in script output):

piweba3y.prodigy.com *GET: 9374 POST:* alyssa.prodigy.com *GET: 7779 POST:* piwebaly.prodigy.com *GET: 5460 POST:* piweba4y.prodigy.com *GET: 3922 POST:* piweba2y.prodigy.com *GET: 3379 POST:* www-b6.proxy.aol.com *GET: 3097 POST: GET*: 3055 *POST*: news.ti.com 163.206.89.4 *GET: 3046 POST:* bill.ksc.nasa.gov GET: 2881 POST: disarray.demon.co.uk GET: 2783 POST:

>> These IPs represent the most active clients, likely proxies or automated systems.



# 3. Failure Requests:

• Total Failed Requests (4xx/5xx): 5,816

• Failure Rate: 0.58%

#### **Top 5 Days with Highest Failures:**

06/Jul/1995: 656

07/Jul/1995: 583

03/Jul/1995: 543

05/Jul/1995: 520

12/Jul/1995: 488

#### **Failure Patterns by Hour (Top 10):**

10:00 - 388 failures

15:00 - 383 failures

14:00 - 377 failures

16:00 - 376 failures

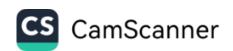
11:00 - 362 failures

>> Failures were slightly more frequent during mid-day hours, possibly due to higher usage.

#### 4. Most Active IP Overall:

• piweba3y.prodigy.com with 9,374 requests

>> This IP was the most active client, heavily using the GET method.



#### 5. Requests Per Hour (Top 5 shown):

14:00 - 64,024 requests 15:00 - 62,881 requests 12:00 - 62,757 requests 11:00 - 62,308 requests 16:00 - 62,055 requests

>> Usage steadily increased from morning and peaked in the early afternoon.

### 6. Status Code Breakdown (Top):

200 OK: 892,291

304 Not Modified: 70,146

302 Found: 29,947

404 Not Found: 5,504

500 Internal Server Error: 54

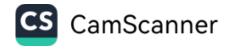
>> Most requests were successful (HTTP 200). There were a significant number of cached responses (304) and redirects (302). Only a small fraction failed.

## 7. Top GET IP:

• piweba3y.prodigy.com with **9,374** GET requests.

## 8. Top POST IP:

• 163.205.1.45 with **21** POST requests



# 9. Daily Request Average:

• Total Days Detected: 25

• Average Requests/Day: 40,000

## 10. Days with Most Failures:

06/Jul/1995: 656 failures

07/Jul/1995: 583 03/Jul/1995: 543 05/Jul/1995: 520 12/Jul/1995: 488

## 11. Failure Patterns by Hour:

 $10:00 \rightarrow 388$  failures

 $15:00 \to 383$ 

 $14:00 \to 377$ 

 $16:00 \to 376$ 

 $11:00 \to 362$ 

# 12. Request Trends (Top Hours):

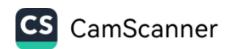
 $14:00 \rightarrow 64,024$ 

 $15:00 \rightarrow 62,881$ 

 $12:00 \rightarrow 62,757$ 

 $11:00 \rightarrow 62,308$ 

 $16:00 \rightarrow 62,055$ 



# 6) Additional Notes:

- Some malformed status codes or corrupted entries were observed; handled gracefully.
- The failure distribution shows peak error activity during midday hours.

# 7) Conclusion & Recommendations:

The log analysis provides deep insights into traffic load, server behavior, and potential issues. Based on findings:

- Optimize high-load hours (14:00–16:00) to balance performance.
- Investigate repeated failures (404 & 500 codes).
- Monitor top active IPs for rate-limiting or suspicious behavior.
- Very few clients used the **POST** method, indicating mostly read-only activity.
- The server handled a **very high volume** of requests efficiently, with a **low failure rate**.

## 8) Files Submitted in GitHub Repository:

- access log (Log file used for analysis)
- analyze\_log.sh (Bash script used)
- Log-report.pdf (This report)

#### **End of Report**

