Apache Log Analysis Report

Name: shadwa ahmed

Id:2205026

Course: information security management

Script Name: log_analysis.sh

Log File Analyzed: apache_logs.txt

1. Overview

This report presents an analytical summary of access patterns, client behavior, request trends, and potential anomalies based on the contents of the Apache log file apache_logs.txt. The analysis was conducted using a custom Bash script designed to extract and present meaningful insights into web traffic and server responses.

2. Summary of Findings

Metric	Value
Total Requests	10,000
GET Requests	9,952
POST Requests	5
Failed Requests (4xx/5xx)	220
Failure Rate	2.00%
Unique IP Addresses	1,753
Average Requests Per Day	2,500
Most Active IP Address	66.249.73.135 (482 requests)
Most Common Status Code	200 (9126 responses)

3. Top IP Addresses by Request Type

GET Requests (Top 5)

IP Address	Count
66.249.73.135	482
46.105.14.53	364
130.237.218.86	357
75.97.9.59	273

50.16.19.13	113

POST Requests

IP Address	Count
78.173.140.106	3
91.236.74.121	1
37.115.186.244	1

4. Request Failures

- Total Failed Requests: 220

- Status Codes Indicative of Failures:

- 404 Not Found: 213
- 500 Internal Server Error: 3
- 416/403 Errors: 4 combined

Failures by Day

Date	Failures
19/May/2015	66
18/May/2015	66
20/May/2015	58
17/May/2015	30

Failures by Hour

Hour	Count
09	18
05	15
06	14
10	12
13	12
17	12
14	11
11	11
02	10
19	10

5. Hourly Request Distribution

Hour	Requests
00	361
01	360
02	365
03	354
04	355
05	371

06	366
07	357
08	345
09	364
10	443
11	459
12	462
13	475
14	498
15	496
16	473
17	484
18	478
19	493
20	486
21	453
22	346
23	356

6. Status Code Breakdown

Status	Meaning	Count
200	OK	9126
304	Not Modified	445
404	Not Found	213
301	Moved Permanently	164
206	Partial Content	45
500	Internal Server Error	3
416	Range Not Satisfiable	2
403	Forbidden	2

7. Key Observations

- The IP 66.249.73.135 generated the highest traffic, suggesting it may be an automated crawler.

- Failure spikes occurred mainly on May 18–19 and during 9–11 AM and 1–2 PM.

- The number of POST requests is significantly low, indicating low user interaction.

- Over 90% of requests returned a 200 OK status, indicating good overall server responsiveness.

8. Recommendations Based on Analysis

• Reducing Failures: Investigate broken links and misconfigured routes causing 404 and 500 errors. Implement redirects or fix sources of traffic to invalid URLs.

• Critical Days/Times: Focus monitoring and load handling strategies on peak failure periods (May 18–19, and 09:00–11:00 and 13:00–14:00).

• Security Concerns: High-volume IP (66.249.73.135) should be monitored and potentially rate-limited to reduce bot-induced traffic spikes.

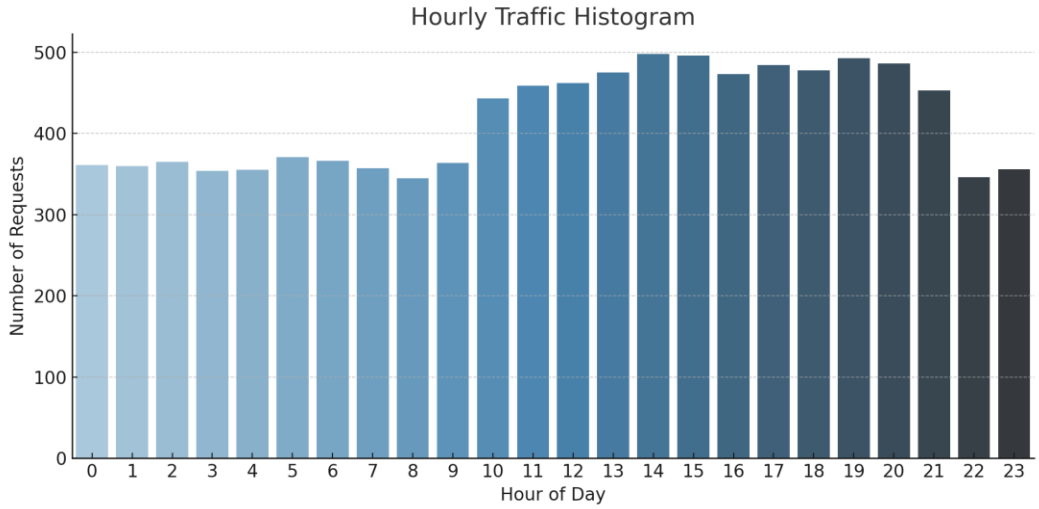
• System Improvements: Add failure alerting, improve error logging, and configure reverse proxy filters to handle bot traffic more efficiently.

9. Visual Data Analysis

To enhance the insights from the log data, this section includes visual representations of traffic and error patterns.

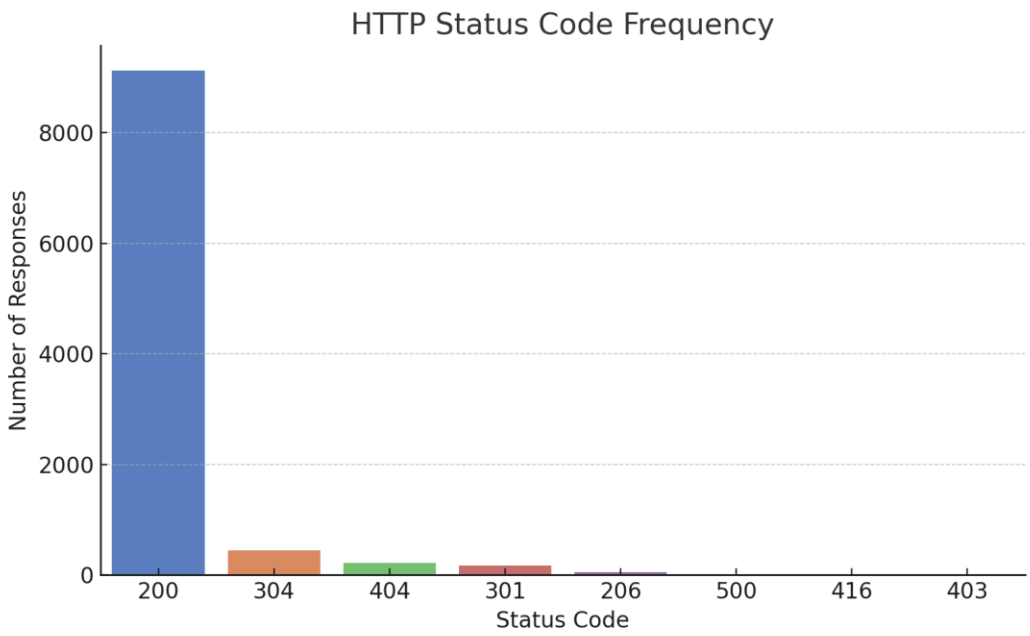
1. Hourly Traffic Histogram

This bar chart illustrates request volume distribution across 24 hours, highlighting peak traffic periods in the early afternoon.



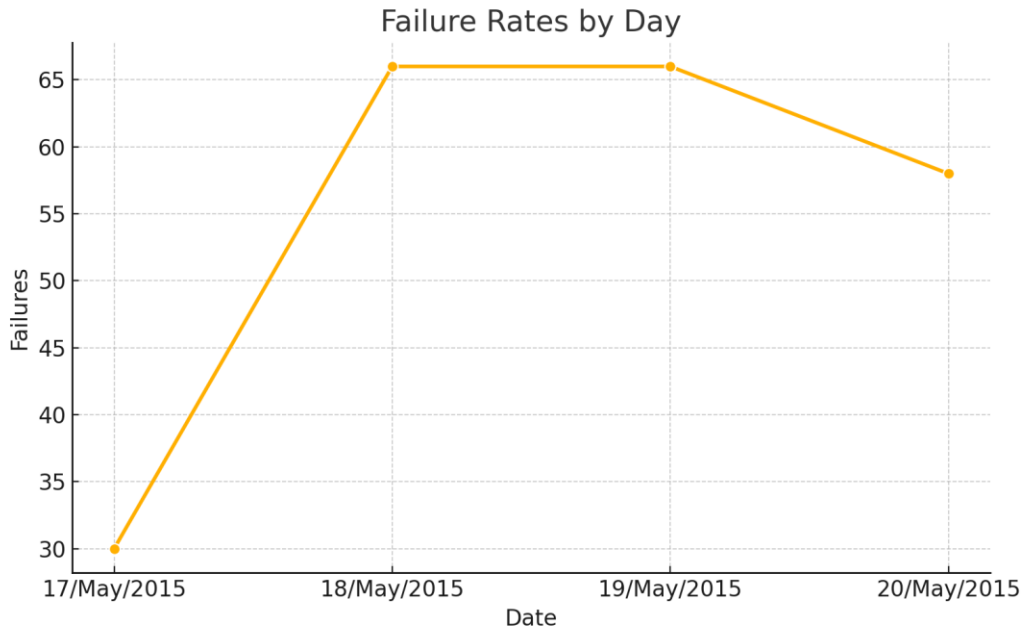
2. Status Code Frequency

This bar chart shows the frequency of each HTTP status code, making even low-frequency errors like 500 easily visible.



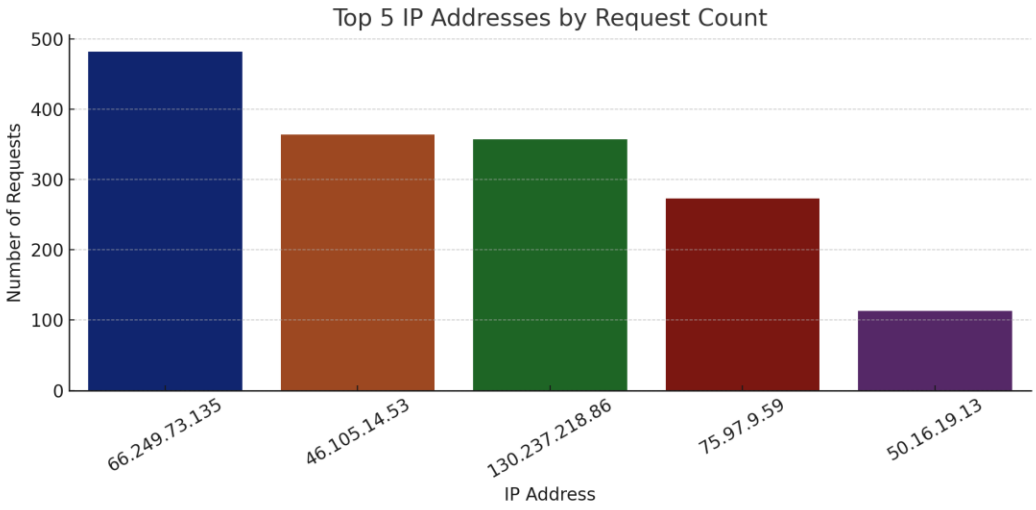
3. Failure Rates by Day

This line chart identifies which days had the most errors, especially May 18 and 19.



4. Top IPs by Request Count

This chart displays the five most active IP addresses, suggesting potential bots or crawlers.



5. Failure Heatmap by Hour

This heatmap visualizes which hours of the day are most error-prone across the dataset.