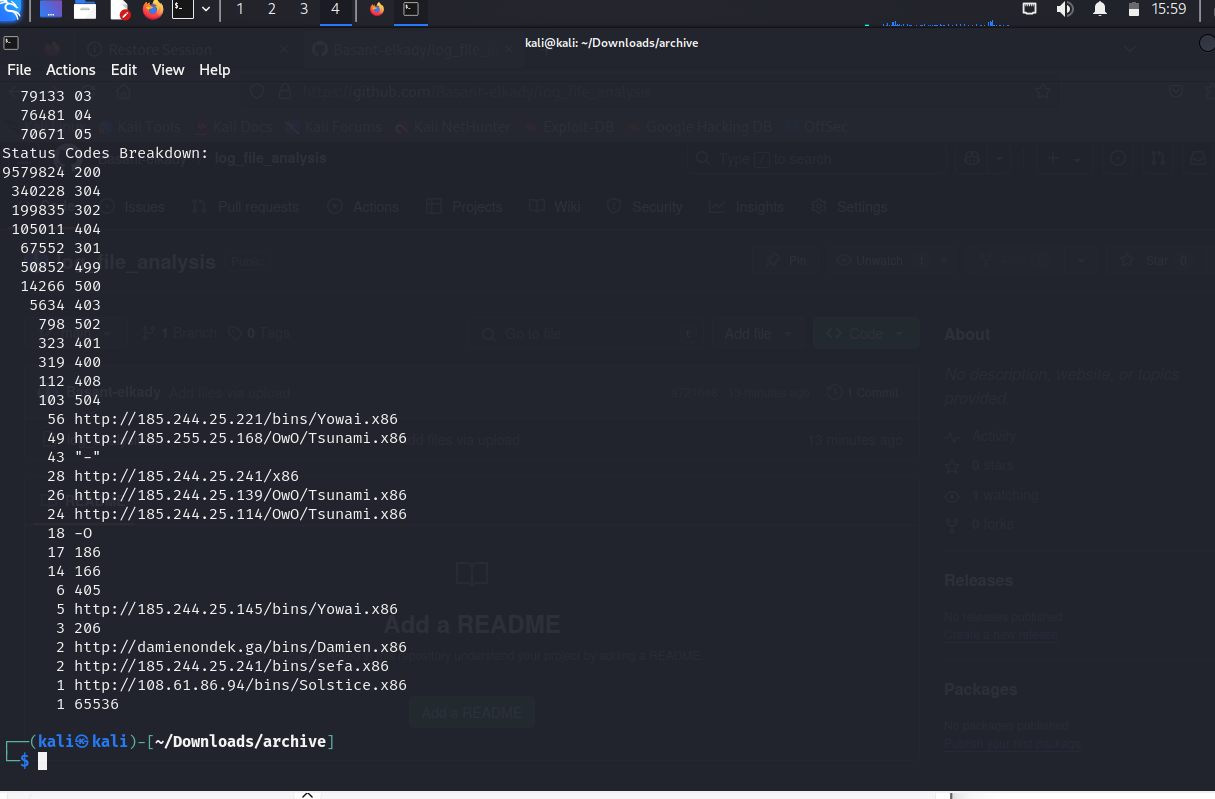
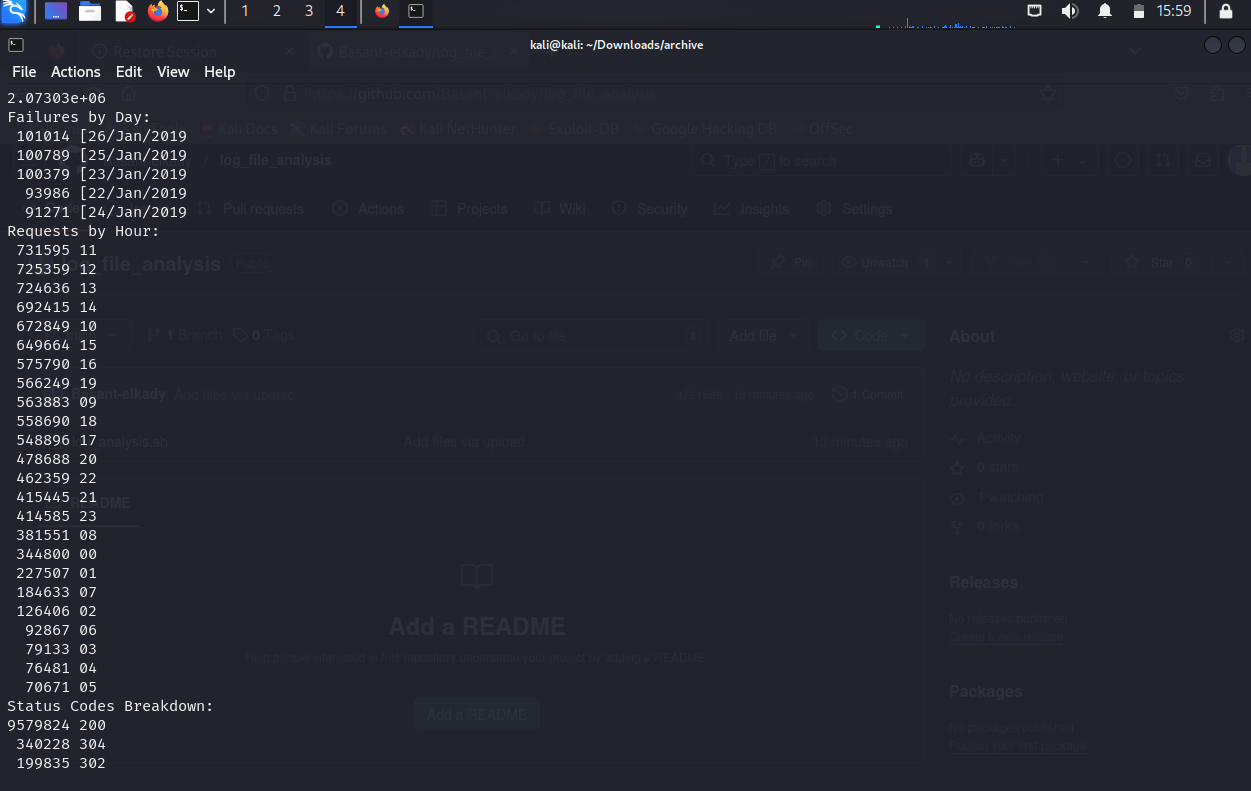
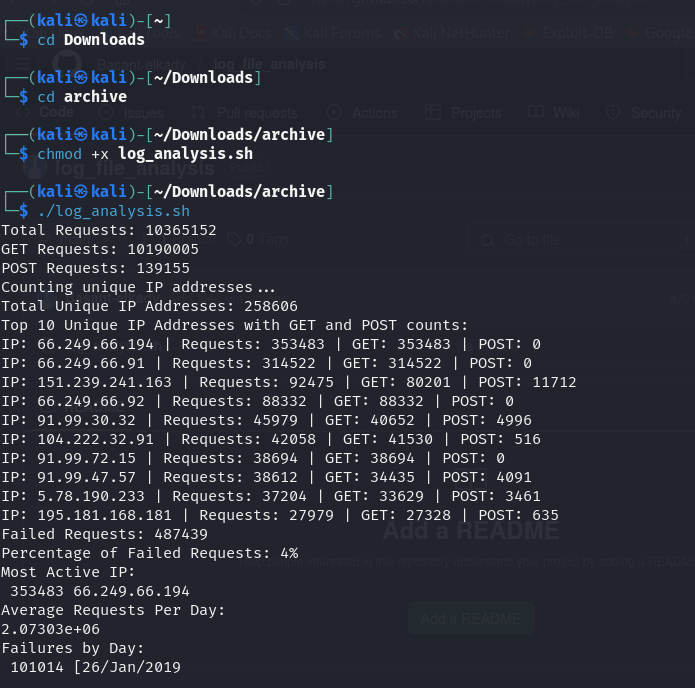
**Repot about log file analysis**



# **Log File Analysis Report**

**Student Name:** Basant Ahmed Mahmoud Elkady  
 **Student ID:** 2205193  
 **Date:** May 10, 2025  
 **Log File Analyzed:** access.log  
 **Total Log Entries:** 10,365,152

## **Introduction: What is a Log File?**

A **log file** is a file automatically generated by a web server to record all the requests it receives from users or external systems. Each line in this file typically contains:

* The visitor’s IP address
* The request type (GET or POST)
* The requested URL
* The date and time
* The server’s response code (e.g., 200, 404, 500)

### **Why Analyze a Log File?**

Analyzing log files helps website administrators and developers to:

* Monitor performance and usage trends
* Identify server errors or broken links
* Detect suspicious or malicious behavior
* Optimize website content and structure
* Improve user experience and resource management

## **Objectives of the Analysis**

This analysis aims to extract valuable insights from the access.log file to evaluate server performance, recognize problem areas, detect security threats, and propose enhancements.

## **Assignment Questions & Analysis Results**

### **1. Request Counts**

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

The vast majority of requests are GET, which typically means page views or resource access. POST requests, used for data submission (e.g., forms), are significantly fewer.

### **2. Unique IP Addresses**

* **Total Unique IPs:** 258,606

#### **Top 10 IPs with GET and POST Requests:**

|  |  |  |  |
| --- | --- | --- | --- |
| **IP Address** | **Total Requests** | **GET Requests** | **POST Requests** |
| 66.249.66.194 | 353,483 | 353,483 | 0 |
| 66.249.66.91 | 314,522 | 314,522 | 0 |
| 151.239.241.163 | 92,475 | 80,201 | 11,712 |
| 66.249.66.92 | 88,332 | 88,332 | 0 |
| 91.99.30.32 | 45,979 | 40,652 | 4,996 |
| 104.222.32.91 | 42,058 | 41,530 | 516 |
| 91.99.72.15 | 38,694 | 38,694 | 0 |
| 91.99.47.57 | 38,612 | 34,435 | 4,091 |
| 5.78.190.233 | 37,204 | 33,629 | 3,461 |
| 195.181.168.181 | 27,979 | 27,328 | 635 |

A large number of requests from a few IPs, especially those starting with 66.249, suggest they are bots or web crawlers (e.g., Googlebot).

### **3. Failed Requests**

* **Total Failed Requests (4xx and 5xx):** 487,439
* **Failure Rate:** 4%

A 4% failure rate is relatively moderate, but still worth investigating. Most common errors include missing pages (404) and server errors (500).

### **4. Most Active IP**

* **IP Address:** 66.249.66.194
* **Total Requests:** 353,483

### **5. Daily Average Requests**

* **Average per Day:** Approximately 2,073,030 requests/day

Indicates high usage. The infrastructure should be scaled properly to handle this level of traffic.

### **6. Failure Analysis by Day**

|  |  |
| --- | --- |
| **Date** | **Failed Requests** |
| 26/Jan/2019 | 101,014 |
| 25/Jan/2019 | 100,789 |
| 23/Jan/2019 | 100,379 |
| 22/Jan/2019 | 93,986 |
| 24/Jan/2019 | 91,271 |

Most errors occurred over a 5-day span, possibly indicating a service issue or external attack.

### **7. Requests by Hour**

|  |  |
| --- | --- |
| **Hour** | **Requests** |
| 11 | 731,595 |
| 12 | 725,359 |
| 13 | 724,636 |
| 14 | 692,415 |
| 10 | 672,849 |
| ... | ... |
| 05 | 70,671 |

Peak activity occurs between 10:00 AM and 3:00 PM, likely reflecting standard business hours.

### **8. Status Code Breakdown**

|  |  |  |
| --- | --- | --- |
| **Status Code** | **Meaning** | **Count** |
| 200 | OK | 9,579,824 |
| 404 | Not Found | 105,011 |
| 500 | Server Error | 14,266 |
| 403 | Forbidden | 5,634 |
| 304 | Not Modified | 340,228 |
| 301 | Moved Permanently | 67,552 |

Most responses were successful (200), but there’s a significant number of 404 and 500 errors, which should be addressed.

### **9. Most Active IP by Request Method**

* **Most GET Requests:** 66.249.66.194 (353,483 requests)
* **Most POST Requests:** 151.239.241.163 (11,712 requests)

### **10. Failure Patterns**

* High failures consistently between January 22–26, 2019
* Nighttime to early morning hours have fewer requests, but can include unusual activity or errors

## **Findings & Recommendations**

### **Key Observations:**

* Heavy activity from a few IPs may indicate automated crawling.
* Five consecutive days with high failures may reflect a technical issue or attack.
* Peak usage occurs mid-day, which is typical for business-oriented sites.

### **Recommendations:**

1. **Investigate Failure Peaks:**  
    Review logs and system behavior during Jan 22–26 to check for service disruptions or security issues.
2. **Security Hardening:**  
    Monitor high-frequency IPs and apply rate-limiting or bot-blocking mechanisms if necessary.
3. **Fix 404 Errors:**  
    Audit broken links and either fix them or implement proper redirects.
4. **Nighttime Monitoring:**  
    Ensure that server processes and monitoring systems remain active during low-traffic hours to catch anomalies.
5. **Optimize Performance:**  
    Use caching strategies or a CDN to improve speed and reduce server load during peak hours.