Analysis of NASA HTTP Access Log

Assignment Report

Computer Science

Alexandria Universitye

Information security

Submitted by

Mohammed Abdulrahman Awad

(ID: 2205114)

May 2025 2024 - 2025

Abstract

This report analyzes the NASA HTTP access log (NASA access log Jul95) using a Bash script (log analyzer.sh). It quantifies 1.89 million requests, with 99.78% GET requests, and examines IP activities and error rates. Recommendations include using a CDN and monitoring IPs to improve server performance.

Contents

Chapter 1

Introduction

Web server logs provide insights into system performance and user behavior. This report analyzes the NASA HTTP access log (NASA_access_log_Jul95), containing approximately 1.89 million entries from July 1995. Using a Bash script (log_analyzer.sh), the study examines request patterns, IP activities, and error rates.

1.1 Objectives

1. Quantify HTTP requests and GET/POST distribution.

2. Identify unique IPs and their activity.

3. Analyze failed requests (10856/0.57%).

4. Recommend server performance improvements.

Chapter 2

Methodology

The analysis used a Bash script (log_analyzer.sh) with tools like awk and grep. The dataset was obtained via:

wget https://ita.ee.lbl.gov/traces/NASA_access_log_Jul95.gz

gunzip NASA_access_log_Jul95.gz

The script was run as:

chmod +x log_analyzer.sh

./log_analyzer.sh > analysis_output0.txt

Errors (e.g., non-ASCII characters) were handled using export LC_ALL=C.

Chapter 3

Results

From analysis_output0.txt:

- **Total Requests**: 1,891,714

- **GET Requests**: 1,887,646 (99.78%)

- **POST Requests**: 111 (0.02%)

- **Unique IPs**: 81,983

- **Most Active IP**: [Update, e.g., 1032015.ksc.nasa.gov (89 requests)]

- **Failed Requests**: [Update, e.g., 44,234 (2.34%)]

- **Status Codes**: [Update, e.g., 200: 80%, 404: 15%, 500: 5%]

Chapter 4

Recommendations

- **Error Handling**: Audit 404 errors and debug 500 errors.

- **Performance**: Use a CDN for GET requests (99.78%).

- **Security**: Monitor high-activity IPs.

References

[1] NASA HTTP Access Log Dataset, 1995. [Online]. Available: https://ita.ee.lbl.gov/traces/NASA_access_log_Jul95.gz. [Accessed: May 2025].