Ex:No:15

Date:

15.To Analyze the different types of servers using Webalizer tool

Aim

The primary aim of this analysis is to **utilize the Webalizer log analysis tool to study, compare, and understand the web traffic patterns and usage statistics of different types of web servers** (e.g., Apache, Nginx, or servers running on different platforms) based on their access log files.

The secondary objectives are to:

Learn the process of configuring and executing the Webalizer tool on server log files.

Gain proficiency in interpreting key web metrics like **Hits**, **Files**, **Pages**, **Visits**, and **Bandwidth**.

Evaluate the differences in visitor behavior, resource consumption, and error reporting across the analyzed server types.

Theory

1. Webalizer Overview

Webalizer is a fast, free, and robust web server log file analysis program. It processes raw web server log files (such as those in Common Log Format (CLF), Apache Custom Log Format, or W3C Extended Log File Format from IIS/Nginx) and generates detailed, easy-to-read, graphical HTML reports.

2. Log File Analysis

The core theory relies on **Web Usage Mining**, a sub-field of Data Mining, which involves analyzing web access logs to extract insights into user behavior and server performance.

A typical log entry contains critical information logged by the server for every request, which Webalizer parses and aggregates:

Client IP Address: Identifies the user's computer.

Timestamp: The date and time of the request.

Request Line: Specifies the HTTP method (GET, POST), the resource requested (URL), and the HTTP protocol version.

Status Code: The server's response (e.g., **200** for success, **404** for 'Not Found', **500** for 'Server Error').

Transfer Size: The size of the file/data transferred in bytes.

Referrer: The URL the user came from (e.g., a search engine or another website).

User Agent: The software used to access the site (e.g., browser type and operating system).

3. Key Metrics & Interpretation

Webalizer synthesizes the raw log data into the following key metrics:

Metric	Definition	Significance
Hits	Total number of requests made to the server. Includes HTML pages, images, scripts, and all other resources.	Indicates overall **server load** and activity.
Files	The number of hits that resulted in data being sent back (excluding errors or requests for cached items).	A more relevant measure of **actual resource consumption** than Hits.
Pages	Requests for actual web documents (e.g., .html, .php), excluding embedded items like images or CSS.	Represents the number of **page views** or impressions.
Visits	A series of requests from the same IP address within a specified time period (typically 30 minutes).	Represents a single **user session** on the website.
Sites	The count of unique IP addresses or hostnames that made requests.	A rough gauge of the **number of unique visitors**.
KBytes	Total amount of data transferred by the server (in Kilobytes).	Represents the total **bandwidth** consumed.

Export to Sheets

4. Server-Specific Traffic

Webalizer can analyze log files from servers like **Apache HTTP Server**, **Nginx**, and **Microsoft IIS**, provided the logs are in a supported format (CLF or W3C). The main difference in analyzing "different server types" lies in comparing the *traffic characteristics* (e.g., high vs. low traffic, error rates, bot activity) of websites hosted on these servers, as reported by Webalizer.

Observation

(Note: This section requires you to substitute the example data and analysis with the actual results from your Webalizer runs. Assume you analyzed logs from two hypothetical servers: "Server A (Apache)" and "Server B (Nginx)".)

1. Server Configuration and Data Summary

Metric	Server A (Apache)	Server B (Nginx)	Interpretation
Time Period Analyzed	Oct 1 - Oct 31, 2025	Oct 1 - Oct 31, 2025	Consistency in reporting period.
Total Hits	850,000	420,000	**Server A** has twice the overall request load.
Total Visits	25,000	22,000	The number of unique sessions is relatively similar.
Average Hits/Visit	34	19	Visitors on **Server A** interact with significantly more resources per session.
Total KBytes	12.5 GB	5.0 GB	**Server A** consumed much more bandwidth, correlating with higher Hits/Visit.

Export to Sheets

2. Hourly and Daily Usage Patterns

Pattern	Server A (Apache)	Server B (Nginx)	Analysis
Peak Hour	10:00 AM - 11:00 AM	3:00 PM - 4:00 PM	Traffic peaks are different. **Server A** traffic is likely business-hours based, **Server B** may be more consumer-based or international.
Peak Day	Monday	Wednesday	Indicates different user routines and potential marketing campaign impact.

Export to Sheets

3. Referrers and User Agents

Category	Server A Observations	Server B Observations	Interpretation
Top Referrers	Major Search Engines (90%), Direct Traffic (10%)	Direct Traffic (60%), Social Media (30%)	**Server A** relies heavily on Search Engine Optimization (SEO); **Server B** has a strong returning user base and social media presence.
User Agents	Chrome (55%), Firefox (20%), Mobile (25%)	Chrome (70%), Mobile (35%), Very high "Bot" traffic percentage (15%)	**Server B** shows higher mobile and automated/bot traffic, possibly requiring better mobile optimization or improved bot filtering.

Export to Sheets

4. Error Analysis (HTTP Status Codes)

The report on **HTTP Status Codes** is vital for server health.

Server A (Apache):

404 Not Found: Low (0.5% of total hits).

5xx Server Errors: Negligible (0.01%).

Conclusion: Indicates a well-maintained site structure with few broken links.

Server B (Nginx):

404 Not Found: Moderate (3.5% of total hits). Top 404s point to missing icons and old URLs.

503 Service Unavailable: Sporadic spikes correlating with peak traffic.