

Statystyki serwera WWW w oparciu o pakiet webalizer

Instalacja

Webalizer w dystrybucji Fedora instalowany jest z gotowej paczki **.rpm**. Instalacja jest więc prosta:
su -c "yum install webalizer"

Uruchomienie

Webalizer to aplikacja, którą należy uruchomić w celu wygenerowania aktualnego raportu. Ścieżka do pliku wykonywalnego to **/usr/bin/webalizer**. Aby aplikacja była uruchamiana cyklicznie należy wykorzystywać mechanizm crona (Unixowy demon umożliwiający wykonywanie określonych poleceń w zadanych przedziałach czasu). Jeżeli webalizer został zainstalowany z paczki rpm w systemie Fedora, wszystkie skrypty są już gotowe.

Zobacz pliki: **/etc/cron.daily/00webalizer**

Zobacz również plik **/etc/httpd/conf.d/webalizer.conf** odczytywany przez serwer Apache.

Znajdziesz tam alias do katalogu z raportami webalizera

```
Alias /usage /var/www/usage
```

oraz sekcję definiującą hosty, z których będzie można odczytać generowane statystyki.

```
<Location /usage>
  Order deny,allow
  Allow from all
  Allow from 127.0.0.1
  Allow from ::1
  Allow from 192.168.10.22
  Allow from 81.168.228.206
  Allow from 81.168.228.211
</Location>
```

Najważniejsze opcje konfiguracyjne pliku **/etc/webalizer.conf**

LogFile */var/log/httpd/access_log* – wskazuje ścieżkę dostępu do pliku z dziennikiem zdarzeń.

LogType *clf* – określa format logów. Domyślnie Webalizer oczekuje logów w formacie CLF lub Combined (dla obu wartości tej zmiennej to *clf*). Możliwe jest również przetwarzanie logów serwera WU-FTPD (wartość *ftp*) oraz Squid (*squid*).

OutputDir */var/www/usage* – definiuje katalog, w którym będą zamieszczane wygenerowane raporty.

HistoryName *webalizer.hist* – ścieżka do pliku przechowującego przetworzone logi serwera w przekroju miesięcznym, wykorzystywane do wygenerowania głównego pliku raportu.

IncrementalName *webalizer.current* – ścieżka do pliku przechowującego informacje wykorzystywane do „przyrostowej” analizy logów.

ReportTitle *tytuł* – określa tytuł raportu.

HostName *nazwa_komputera* – określa nazwę komputera (serwera), dla którego są generowane raporty.

PageType *htm** – wskazuje typ plików traktowanych jako „strona internetowa”, w odróżnieniu od innych elementów, takich jak obiekty multimedialne czy dokumenty. Przykładowe wartości tej zmiennej to *htm** (dla plików HTM i HTML), *php3*, *php*, ale także *cgi* czy *pl* dla skryptów.

Webalizer wyświetla statystyki w formie zestawień zawierających sumaryczne lub uśrednione wartości różnych rodzajów odwołań do serwera. Oto one:

Wywołania (Hits) – liczba odwołań do serwera zarejestrowanych w dzienniku zdarzeń, zakończonych zarówno sukcesem (wyświetlenie strony, przesłanie dokumentu), jak i niepowodzeniem (wywołanie nieistniejącej strony).

Pliki (Files) – liczba odwołań do serwera z żądaniem przesłania pliku (strony internetowej, dokumentu, obiektów multimedialnych).

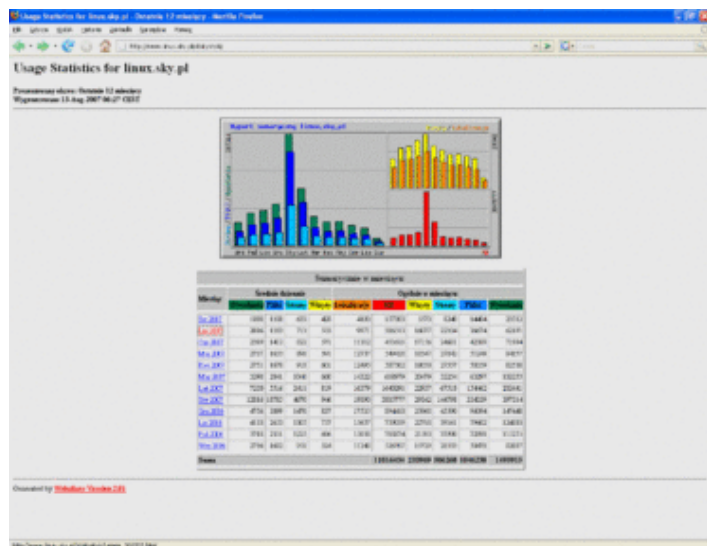
Strony (Pages) – liczba odwołań do plików będących stronami internetowymi, tj. htm, html, php, ale także skryptów cgi, ściśle określonych w pliku konfiguracyjnym Webalizera.

Lokalizacje (Sites) – liczba odwołań z unikatowych adresów IP.

KB (KBytes) – liczba przesłanych przez serwer danych, wyrażona w kilobajtach.

Wizyty (Visits) – liczba odwiedzin strony internetowej, rejestrowana na podstawie odwołań z konkretnych adresów IP, przy czym dane te zliczane są także w przypadku powtarzających się adresów, o ile odwołania nastąpiły w zadanym odstępnie czasu, domyślnie 30 minut.

W widoku sumarycznym statystyki prezentowane są w przekroju miesięcznym z rozróżnieniem całkowitej, miesięcznej liczby odwołań oraz dziennym uśrednieniu.

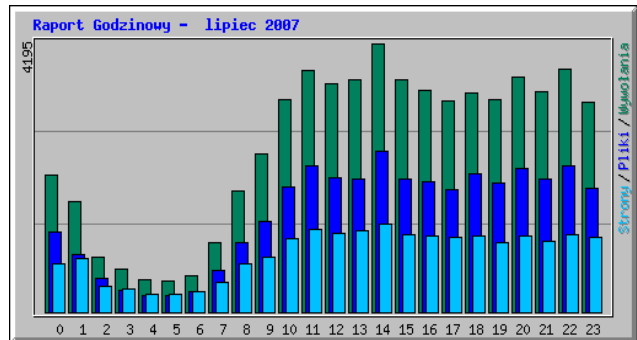
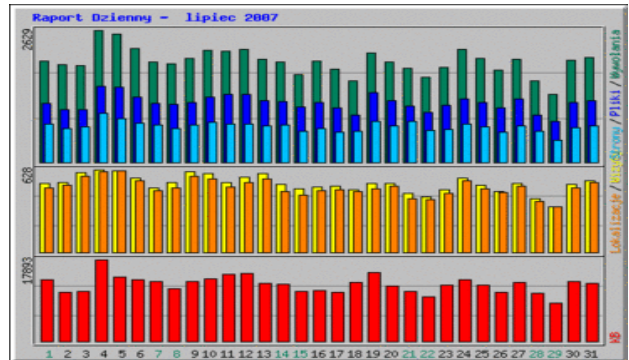


Wystarczy jednak kliknąć nazwę miesiąca, aby bliżej przyjrzeć się bardziej szczegółowym statystykom, wygenerowanych na podstawie danych miesięcznych. W raporcie znajdziemy informację o całkowitej liczbie wywołań, żądanych plików, unikatowych lokalizacji, przeglądarek czy odnośników. Dane te są prezentowane są także w formie uśrednień w przekrojach godzinowych i dziennych.

Monthly Statistics for May 2010		
Total Hits	643927	
Total Files	548588	
Total Pages	105541	
Total Visits	23602	
Total KBytes	4865971	
Total Unique Sites	11008	
Total Unique URLe	4065	
Total Unique Referencje	1195	
Total Unique Użytkownicy	1	
Total Unique User Agents	1741	
	Avg	Max
Hits per Hour	865	4103
Hits per Day	20771	42205
Files per Day	17694	27875
Pages per Day	3404	5615
Visits per Day	761	1085
KBytes per Day	156987	231424

Podążając dalej przez raport, napotkamy raport dzienny oraz godzinowy – oba w formie wykresu i tabeli. Statystyki dzienne dostarczają danych na temat odwołań do serwera w poszczególnych dniach miesiąca, te drugie natomiast w przekroju godzinowym.

Dzienna statystyka - lipiec 2007												
Data	Wyświetlenia	Strony	Wizyty	Unikalni odwiedzający	IP	...						
1	2007	3.23%	1170	3.26%	753	3.02%	491	4.32%	13439	3.48%		
2	1947	3.23%	1054	2.87%	467	3.02%	529	2.28%	507	3.88%	10770	2.79%
3	1921	3.08%	1090	2.88%	710	3.23%	682	3.76%	876	3.76%	10843	2.88%
4	2629	4.23%	1520	4.26%	982	4.46%	628	3.03%	609	6.12%	17893	4.63%
5	2551	4.23%	1500	4.08%	869	3.03%	620	3.88%	615	4.27%	14086	3.48%
6	2270	3.63%	1297	3.26%	774	3.08%	562	3.08%	848	3.47%	12468	3.48%
7	1995	3.23%	1167	3.08%	738	3.08%	491	3.08%	470	4.27%	12187	3.48%
8	1953	3.26%	1158	3.26%	670	3.08%	531	3.08%	485	4.88%	11451	2.88%
9	2066	3.33%	1187	3.26%	737	3.08%	613	3.08%	879	3.87%	13076	3.38%
10	2232	3.33%	1300	3.26%	799	3.08%	597	3.26%	554	3.38%	13099	3.38%
11	2289	3.33%	1353	3.26%	762	3.08%	628	3.08%	497	4.88%	14718	3.38%
12	2290	3.63%	1347	3.63%	761	3.46%	570	3.58%	532	3.34%	14898	3.46%
13	2054	3.33%	1233	3.26%	733	3.26%	594	3.46%	559	3.62%	12689	3.26%
14	1999	3.23%	1206	3.26%	748	3.08%	513	3.08%	463	4.62%	12578	3.26%
15	1744	2.88%	1094	2.88%	614	2.76%	481	2.88%	436	4.37%	10955	2.88%
16	2009	3.23%	1191	3.23%	662	2.88%	493	3.07%	471	4.27%	11130	2.88%
17	1846	2.87%	1090	2.87%	601	2.76%	503	3.03%	478	4.76%	10741	2.76%
18	1616	2.88%	929	2.88%	619	2.88%	476	2.88%	462	4.62%	12822	2.88%
19	2179	3.08%	1283	3.08%	821	3.26%	523	3.26%	481	4.62%	15169	3.08%
20	1980	3.26%	1225	3.26%	718	3.26%	525	3.26%	563	3.64%	12114	3.26%
21	1874	3.03%	1112	3.03%	607	3.03%	445	2.77%	407	4.88%	10882	3.03%
22	1693	2.76%	986	2.69%	439	2.69%	421	2.69%	398	3.88%	9762	2.69%
23	1884	3.08%	1140	3.23%	649	2.88%	474	2.88%	448	4.48%	12342	2.88%
24	2247	3.63%	1297	3.63%	758	3.63%	560	3.63%	541	3.63%	13423	3.63%
25	2098	3.33%	1195	3.26%	707	3.26%	609	3.26%	481	4.62%	12232	3.26%
26	1833	2.88%	1082	2.88%	605	2.76%	465	2.88%	455	4.38%	10750	2.76%
27	2049	3.26%	1268	3.63%	725	3.26%	522	3.26%	504	3.63%	12887	3.26%
28	1627	2.63%	933	2.63%	623	2.63%	409	2.63%	387	3.88%	10834	2.63%
29	1353	2.88%	815	2.23%	437	1.88%	342	1.23%	342	3.63%	8405	1.23%
30	2027	3.26%	1183	3.26%	689	3.26%	514	3.26%	491	4.62%	13065	3.26%
31	2081	3.33%	1229	3.33%	727	3.33%	546	3.33%	531	3.33%	12766	3.33%



Dalsze statystyki dostarczają informacji o najczęściej wywoływanych URL-ach (czyli najczęściej oglądanych stronach) uszeregowane według liczby wywołań lub rozmiaru przesłanych danych, ale także o stronach, z których nastąpiło wejście do serwisu i wyjście z niego.

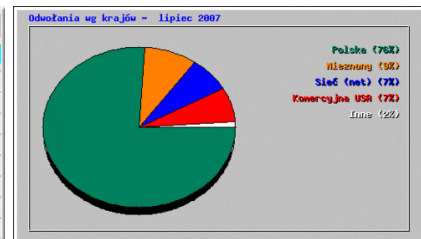
Pierwsze 30 z ogólnej liczby 195 URLi					
#	Wyświetlenia	Strony	Wizyty	URL	
1	2071	3.23%	28664	7.42%	/
2	2042	3.26%	3474	0.98%	/maczek/instalacja.html
3	1795	2.88%	20751	5.57%	/karty/instal.html
4	1321	2.22%	43299	11.38%	/karty/kernel.html
5	827	1.33%	14768	3.92%	/karty/instalacja.html
6	783	1.26%	11100	2.97%	/karty/adman.html
7	760	1.22%	16670	4.32%	/karty/instalacja.html
8	700	1.13%	8972	2.32%	/karty/serwer.html
9	558	0.86%	11741	3.08%	/karty/bash.html
10	530	0.83%	3696	0.98%	/karty/irc.html
11	468	0.77%	4879	1.28%	/karty/serwer.html
12	391	0.63%	241	0.06%	/forum
13	360	0.56%	1272	0.33%	/karty/irc.html
14	336	0.54%	1258	0.33%	/karty/instalacja.html
15	329	0.53%	6075	1.57%	/karty/irc.html
16	328	0.53%	3073	0.80%	/karty/instalacja.html
17	324	0.52%	4481	1.18%	/karty/irc.html
18	320	0.51%	1133	0.29%	/karty/instalacja.html
19	313	0.50%	2471	0.64%	/karty/irc.html
20	310	0.50%	4454	1.17%	/karty/instalacja.html
21	307	0.49%	9163	2.37%	/maczek/instalacja.html
22	290	0.47%	4534	1.17%	/karty/instalacja.html
23	268	0.43%	2738	0.72%	/karty/irc.html
24	243	0.39%	1748	0.46%	/karty/irc.html
25	243	0.39%	3931	1.02%	/karty/instalacja.html
26	240	0.39%	2355	0.61%	/karty/instalacja.html
27	239	0.39%	1482	0.39%	/karty/instalacja.html
28	238	0.39%	1731	0.46%	/karty/instalacja.html
29	228	0.37%	1813	0.48%	/maczek/instalacja.html
30	227	0.37%	1287	0.34%	/karty/instalacja.html

Pierwsze 10 z ogólnej liczby 77 stron wejściowych				
#	Wyświetlenia	Wizyty	URL	
1	2071	3.23%	1656	10.42%
2	1795	2.88%	1414	8.98%
3	1321	2.22%	1129	7.18%
4	827	1.33%	739	4.61%
5	760	1.22%	629	3.96%
6	700	1.13%	604	3.88%
7	783	1.26%	557	3.50%
8	558	0.86%	503	3.16%
9	530	0.83%	472	2.97%
10	2042	3.26%	469	2.93%

Pierwsze 10 z ogólnej liczby 78 stron wyjściowych				
#	Wyświetlenia	Wizyty	URL	
1	1795	2.88%	1525	9.60%
2	2071	3.23%	1279	8.03%
3	1321	2.22%	1087	6.84%
4	827	1.33%	708	4.48%
5	760	1.22%	651	4.10%
6	783	1.26%	625	3.93%
7	700	1.13%	570	3.59%
8	558	0.86%	483	3.04%
9	530	0.83%	469	2.93%
10	2042	3.26%	468	2.93%

Interesujące wydają się także statystyki najczęściej używanych przeglądarek oraz krajów, z których nastąpiło odwołanie do serwera.

Pierwsze 15 z ogólnej liczby 157 przeglądarek			
#	Wyświetlenia	Przebieg	
1	3710	59.77%	Mozilla/5.0
2	5001	0.04%	Opera/9.2
3	4866	7.82%	MSIE 6.0
4	3734	6.00%	Yahoo! Slurp
5	2182	3.51%	MSIE 7.0
6	1145	1.84%	Konqueror/3.5
7	702	1.13%	Goobot/2.1
8	497	0.80%	Opera/9.1
9	480	0.77%	Yahoo! Slurp China
10	409	0.66%	Opera/9.0
11	392	0.63%	msnbot/1.0 (+http://search.msn.com/msnbot.htm)
12	366	0.59%	Crawlpl-1/Nutch-0.9 (bot, www.pjwstk.edu.pl/~msyd/crawlPIP)
13	312	0.50%	MSIE 4.0
14	223	0.36%	holmes/3.11 (OnetSmukaj/5.0; +http://smukaj.onet.pl)
15	218	0.35%	CazoodleBot/Nutch-0.9-dev (CazoodleBot Crawler, http://www.ca



Dostęp do lokalnego katalogu serwera Apache na hasło

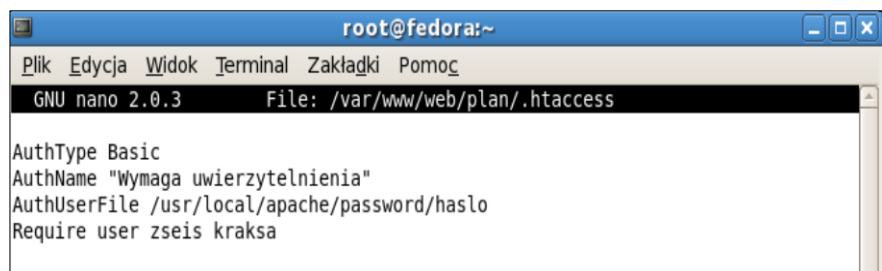
Często się zdarza, że administrator udostępnia wybrane zasoby serwera WWW wybranym użytkownikom na hasło. Wymaga to jednak dodatkowej konfiguracji serwera Apache oraz utworzenia nowych kont użytkowników.

Za autoryzację w wybranych zasobach serwera (katalogach) odpowiada:

- dyrektywa **AllowOverride**.
Zobacz na rysunku obok
Dyrektywa **AllowOverride AuthConfig** powinna się znaleźć w sekcji katalogu, do którego chcemy zrobić autoryzację

```
# Options FileInfo AuthConfig Limit
#
# AllowOverride None
# AllowOverride AuthConfig
```

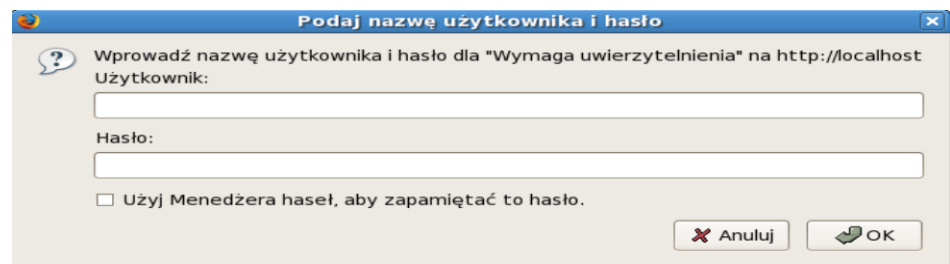
- Plik **.htaccess**
Definiuje on typ autoryzacji, ścieżkę do pliku z hasłami oraz komunikat na oknie logowania.



```
root@fedora:~
Plik Edycja Widok Terminal Zakładki Pomoc
GNU nano 2.0.3 File: /var/www/web/plan/.htaccess

AuthType Basic
AuthName "Wymaga uwierzytelnienia"
AuthUserFile /usr/local/apache/password/hasło
Require user zseis kraksa
```

- Należy ten plik umieścić w katalogu, do którego dostęp będzie wymagał autoryzacji



Podaj nazwę użytkownika i hasło

Wprowadź nazwę użytkownika i hasło dla "Wymaga uwierzytelnienia" na http://localhost

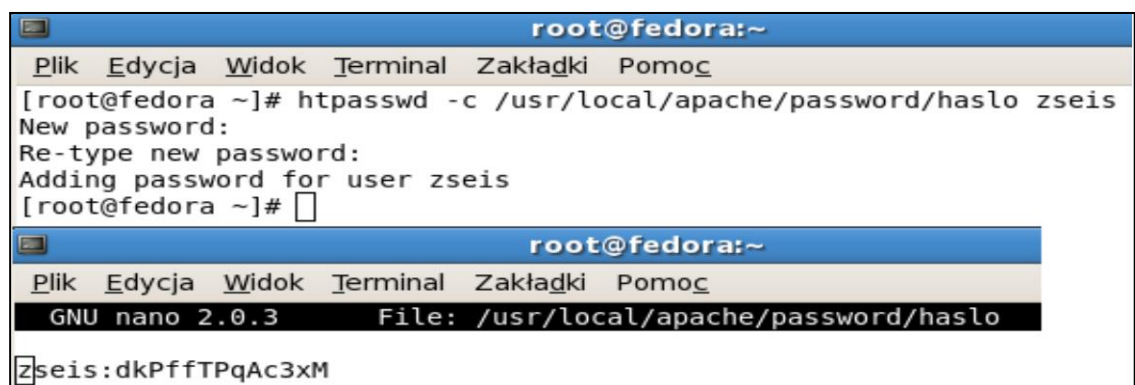
Użytkownik:

Hasło:

Użyj Menedżera haseł, aby zapamiętać to hasło.

Anuluj OK

- Plik z utworzonymi hasłami dla użytkowników tworzonych za pomocą polecenia **htpasswd**



```
root@fedora:~
Plik Edycja Widok Terminal Zakładki Pomoc
[root@fedora ~]# htpasswd -c /usr/local/apache/password/hasło zseis
New password:
Re-type new password:
Adding password for user zseis
[root@fedora ~]#

root@fedora:~
Plik Edycja Widok Terminal Zakładki Pomoc
GNU nano 2.0.3 File: /usr/local/apache/password/hasło
zseis:dkPffTPqAc3xM
```

UWAGA: Nie zapomnij na koniec zrestartować serwera Apache.