# 목차

# Apache HTTPD 로그 분석

## 분석 스크립트

### 일간 request

```
[root@localhost ~]# awk '{print $4}' access_log | cut -d: -f1 | uniq -c
6095 [20/Jan/2013
7281 [21/Jan/2013
6517 [22/Jan/2013
5278 [23/Jan/2013
```

### 특정 일자 시간별 request

```
[root@localhost ~]# grep "23/Jan" access_log | cut -d[ -f2 | cut -d] -f1 |
awk -F: '{print $2":00"}' | sort -n | uniq -c
200 00:00
417 01:00
244 02:00
242 03:00
344 04:00
402 05:00
522 06:00
456 07:00
490 08:00
438 09:00
430 10:00
357 11:00
284 12:00
391 13:00
163 14:00
```

```
[root@localhost ~]# awk -F"[ :[/]" '{print $7"-"$6"-"$5" "$8}' access_log |
sort | uniq -c
    172 2021-Dec-13 00
    158 2021-Dec-13 01
    109 2021-Dec-13 02
    142 2021-Dec-13 03
    115 2021-Dec-13 04
```

### 특정 일자 분당 request

```
[root@localhost ~]# grep "23/Jan" access_log | cut -d[ -f2 | cut -d] -f1 |
awk -F: '{print $2":"$3}' | sort -nk1 -nk2 | uniq -c | awk '{ if ($1 > 10)
```

```
print $0}'
     14 00:08
     15 00:09
     15 00:15
     25 00:22
     12 00:44
     12 00:45
     12 00:52
     22 01:17
     22 01:26
     16 01:34
     11 01:45
     14 01:52
     13 02:23
```

## Request Query 분석 스크립트

```
#!/bin/bash

###### SETUP ############
LOG_FOLDER=/var/www/vhosts/domain.co.uk/statistics/logs
ACCESS_LOG=$LOG_FOLDER/access_log

HOW_MANY_ROWS=20000



######### FUNCTIONS ##############


function title() {
    echo "
---------------------------------
$@
---------------------------------
"
}

function urls_by_ip() {
    local IP=$1
    tail -5000 $ACCESS_LOG | awk -v ip=$IP ' $1 ~ ip {freq[$7]++} END {for
(x in freq) {print freq[x], x}}' | sort -rn | head -20
}


function ip_addresses_by_user_agent(){
```

```
    local USERAGENT_STRING="$1"
    local TOP_20_IPS="`tail  -$HOW_MANY_ROWS $ACCESS_LOG | grep
"${USERAGENT_STRING}"  | awk '{freq[$1]++} END {for (x in freq) {print
freq[x], x}}' | sort -rn | head -20`"
    echo "$TOP_20_IPS"
}

####### RUN REPORTS #############


title "top 20 URLs"
TOP_20_URLS="`tail -$HOW_MANY_ROWS $ACCESS_LOG | awk '{freq[$7]++} END {for
(x in freq) {print freq[x], x}}' | sort -rn | head -20`"
echo "$TOP_20_URLS"


title "top 20 URLS excluding POST data"
TOP_20_URLS_WITHOUT_POST="`tail  -$HOW_MANY_ROWS $ACCESS_LOG | awk -F"[ ?]"
'{freq[$7]++} END {for (x in freq) {print freq[x], x}}' | sort -rn | head
-20`"
echo "$TOP_20_URLS_WITHOUT_POST"


title "top 20 IPs"
TOP_20_IPS="`tail  -$HOW_MANY_ROWS $ACCESS_LOG | awk '{freq[$1]++} END {for
(x in freq) {print freq[x], x}}' | sort -rn | head -20`"
echo "$TOP_20_IPS"

title "top 20 user agents"
TOP_20_USER_AGENTS="`tail  -$HOW_MANY_ROWS $ACCESS_LOG | cut -d\  -f12- |
sort | uniq -c | sort -rn | head -20`"
echo "$TOP_20_USER_AGENTS"


title "IP Addresses for Top 3 User Agents"

for ((I=1; I<=3; I++))
do
    UA="`echo "$TOP_20_USER_AGENTS" | head -n $I | tail -n 1 | awk '{$1="";
print $0}'`"
    echo "$UA"
    echo "~~~~~~~~~~~~~~~~~~"
    ip_addresses_by_user_agent "$UA"
    echo "
    "
done
```

- Top 20 URLs
- Top 20 URLs exclude POST
- Top 20 Remote IPs

- Top 20 User-Agent