AWK应用：

Linux Web服务器网站故障分析常用的命令

系统连接状态篇：

1.查看TCP连接状态

netstat -nat |awk ‘{print $6}’|sort|uniq -c|sort -rn

netstat -n | awk ‘/^tcp/ {++S[$NF]};END {for(a in S) print a, S[a]}’ 或

netstat -n | awk ‘/^tcp/ {++state[$NF]}; END {for(key in state) print key,"t",state[key]}’

netstat -n | awk ‘/^tcp/ {++arr[$NF]};END {for(k in arr) print k,"t",arr[k]}’

netstat -n |awk ‘/^tcp/ {print $NF}’|sort|uniq -c|sort -rn

netstat -ant | awk ‘{print $NF}’ | grep -v ‘[a-z]‘ | sort | uniq -c

2.查找请求数请20个IP（常用于查找攻来源）：

netstat -anlp|grep 80|grep tcp|awk ‘{print $5}’|awk -F: ‘{print $1}’|sort|uniq -c|sort -nr|head -n20

netstat -ant |awk ‘/:80/{split($5,ip,":");++A[ip[1]]}END{for(i in A) print A[i],i}’ |sort -rn|head -n20

3.用tcpdump嗅探80端口的访问看看谁最高

tcpdump -i eth0 -tnn dst port 80 -c 1000 | awk -F"." ‘{print $1"."$2"."$3"."$4}’ | sort | uniq -c | sort -nr |head -20

4.查找较多time\_wait连接

netstat -n|grep TIME\_WAIT|awk ‘{print $5}’|sort|uniq -c|sort -rn|head -n20

5.找查较多的SYN连接

netstat -an | grep SYN | awk ‘{print $5}’ | awk -F: ‘{print $1}’ | sort | uniq -c | sort -nr | more

6.根据端口列进程

netstat -ntlp | grep 80 | awk ‘{print $7}’ | cut -d/ -f1

网站日志分析篇1（Apache）：

1.获得访问前10位的ip地址

cat access.log|awk ‘{print $1}’|sort|uniq -c|sort -nr|head -10

cat access.log|awk ‘{counts[$(11)]+=1}; END {for(url in counts) print counts[url], url}’

2.访问次数最多的文件或页面,取前20

cat access.log|awk ‘{print $11}’|sort|uniq -c|sort -nr|head -20

3.列出传输最大的几个exe文件（分析下载站的时候常用）

cat access.log |awk ‘($7~/.exe/){print $10 " " $1 " " $4 " " $7}’|sort -nr|head -20

4.列出输出大于200000byte(约200kb)的exe文件以及对应文件发生次数

cat access.log |awk ‘($10 > 200000 && $7~/.exe/){print $7}’|sort -n|uniq -c|sort -nr|head -100

5.如果日志最后一列记录的是页面文件传输时间，则有列出到客户端最耗时的页面

cat access.log |awk ‘($7~/.php/){print $NF " " $1 " " $4 " " $7}’|sort -nr|head -100

6.列出最最耗时的页面(超过60秒的)的以及对应页面发生次数

cat access.log |awk ‘($NF > 60 && $7~/.php/){print $7}’|sort -n|uniq -c|sort -nr|head -100

7.列出传输时间超过 30 秒的文件

cat access.log |awk ‘($NF > 30){print $7}’|sort -n|uniq -c|sort -nr|head -20

8.统计网站流量（G)

cat access.log |awk ‘{sum+=$10} END {print sum/1024/1024/1024}’

9.统计404的连接

awk ‘($9 ~/404/)’ access.log | awk ‘{print $9,$7}’ | sort

10. 统计http status

cat access.log |awk ‘{counts[$(9)]+=1}; END {for(code in counts) print code, counts[code]}'

cat access.log |awk '{print $9}'|sort|uniq -c|sort -rn

10.蜘蛛分析，查看是哪些蜘蛛在抓取内容。

/usr/sbin/tcpdump -i eth0 -l -s 0 -w - dst port 80 | strings | grep -i user-agent | grep -i -E 'bot|crawler|slurp|spider'

网站日分析2(Squid篇）按域统计流量

zcat squid\_access.log.tar.gz| awk '{print $10,$7}' |awk 'BEGIN{FS="[ /]"}{trfc[$4]+=$1}END{for(domain in trfc){printf "%st%dn",domain,trfc[domain]}}'

数据库篇

1.查看数据库执行的sql

/usr/sbin/tcpdump -i eth0 -s 0 -l -w - dst port 3306 | strings | egrep -i 'SELECT|UPDATE|DELETE|INSERT|SET|COMMIT|ROLLBACK|CREATE|DROP|ALTER|CALL'

系统Debug分析篇

1.调试命令

strace -p pid

2.跟踪指定进程的PID

gdb -p pid