**AWK使用**

awk语法介绍

业务日志分析

总结

1.awk语法介绍

运行机理

基本语法

系统变量

常用函数

运行机理

awk 'BEGIN{} {} END{}' file

对于以上命令，可以划分如下三个部分

-- BEGIN{} 读入文件之前

-- {} 每一行的操作

-- END{} 所有行读取之后

以上三个部分都是相互独立的，在实际应用中，根据不同场景，进行不同的代码编写

基本语法

1 .支持多维数组

awk程序编写支持多维数组的写法

AName["name"] = "zhang"

AName["score"] = total

AName["subjectNum"] = len-1

2 .常规语法

awk 支持for, while 循环的使用

(1) for ( i = 2 ; i < NF ; ++i ) {

total += $i

continue

}

(2) do {

break

执行操作

} while (i in score)

系统变量

awk常用的系统变量如下:

FS 文件分割符，默认为 空格 ， 可以在 BEGIN{FS=";"} 中指定

OFS 输出分隔符,默认输出的两个字段间隔为空格，可以用在 BEGIN{OFS=";"}END{ print $1,$2}

NF 对于读入行中， 每一行中 字段总数

NR 对于读入行的 当前行数

常用函数

split() 用于将字符串分割成数组 len = split(str,Array,delimeter);将 str 以delimeter 分割后，赋值给数组 Array,返回值len为 数组长度

substr() 截取字符串，substr(str,m,n)，从m 截取,长度为 n 的字符串

sub() 替换字符串 ，sub(needle,replace,str)

match() 匹配函数match(str,needle)，内置常量RSTART匹配开始位置,RLENGTH匹配长度,例:match($0,/[A-Z]+/){ print($0,RSTART,RLENGTH)}

length() 获取 字符串 或者 数组的长度 len = length(Array)

index() 获取 字符串的问题 position = index("helloworld","w")

2.awk日志分析

(1).分析日志中接口请求次数及耗时

首先根据现有的日志格式，使用awk进行梳理，目前格式有10个字段，$1-$10:

日志字段如下

[yushaolong@10-10-213-219 logs]$ awk '{for (i=1;i<=NF;++i) {print i "\t" $i};exit 4}' info.log.20170304

1 2017-03-04

2 16:56:29

3 [111.200.192.10][58ba813d70aac877][info][cow\_access]

4 INNERACCESS

5 24

6 androidTeacher

7 355749

8 1

9 http://yushaolong.test.knowbox.cn

10 /common/activity/get-ext?source=androidTeacher&version=1.0.0&channel=Knowbox&token=YWcEX6/ifcZQ8Bspq7%2BFzuH1qZz2zQxvzrml30Aq0FgDSPYnA99BMEY0l7QXI3jm&token=YWcEX6/ifcZQ8Bspq7%2BFzuH1qZz2zQxvzrml30Aq0FgDSPYnA99BMEY0l7QXI3jm&kbparam=BB22D30609991E0B3DAF13E625A748D5

编写awk脚本

文件 log.awk

#!/bin/sh

file=$1

awk -v file=$file 'BEGIN{print "statis logs from " file}

/INNERACCESS/{

# split($10,info,"?")

len = index($10,"?")

name=substr($10,1,len-1)

++api[name]["num"]

api[name]["spend"] += $5

}

END{

print "api:\t""num\t""spendTotal\t""spendAvg\t"

for ( i in api ) {

print i "\t" api[i]["num"] "\t" api[i]["spend"]"\t" api[i]["spend"]/api[i]["num"]

}

}

' $file |

sort -nrk 2

运行结果如下：

[yushaolong@10-10-213-219 test]$ ./log.awk /home/yushaolong/wwwroot/cow/cow/runtime/logs/info.log.201703\*

api: num spendTotal spendAvg

/match/teacher-match/setting 21 1352 64.381

/common/activity/get-global-info 20 849 42.45

/unit/unit/class-unit-info 19 1296 68.2105

/match/teacher-match/detail 19 684 36

/word/classword/class-student-learn-info 18 1078 59.8889

/match/teacher-match/list 16 653 40.8125

/common/activity/get-ext 14 371 26.5

/user/teacher/get-info 10 581 58.1

/user/class/teacher-class-list 10 481 48.1

/common/app/log-triger 7 276 39.4286

/common/app/get-resource-list 7 218 31.1429

/common/app/check-version 7 286 40.8571

/user/teacher/passwd-login 5 130 26

/unit/unit/unit-words-learned-status 4 294 73.5

/unit/unit/unit-students-learned-status 4 190 47.5

/user/teacher/login-out 3 17 5.66667

/match/teacher-match/question-preview 3 125 41.6667

/user/teacher/code-login 2 42 21

/user/class/delete-class 2 345 172.5

/user/class/close-class 2 80 40

/word/classword/class-exercise-word-status 1 42 42

/user/teacher/get-login-code 1 46 46

/user/class/update-class-info 1 35 35

/user/class/create-class 1 238 238

[yushaolong@10-10-213-219 test]$

(2).单个接口，某个时段使用统计

脚本文件 log.time.awk

[yushaolong@10-10-213-219 test]$ cat log.time.awk

#!/bin/sh

file=$1

api=$2

interval=$3

awk -v interval=$interval 'BEGIN{

split(interval,logTime,"-")

begin=logTime[1]

end=logTime[2]

}

{

split($2,time,":")

if (time[1] >= begin && time[1] <= end) {

if(index($10,api)){

aim[api]["num"]++

aim[api]["spend"] += $5

}

}

}

END{

if (aim[api]["num"]) {

num = aim[api]["num"]

spend = aim[api]["spend"]

avgspend = spend/num

} else {

avgspend = num=spend=0

}

print "api :\t" api

print "requestNum:\t" num

print "totalSpendTime:\t" spend

print "avgSpendTime:\t" avgspend

}' api=$api $file

运行结果如下：

[yushaolong@10-10-213-219 test]$ ./log.time.awk /home/yushaolong/wwwroot/cow/cow/runtime/logs/info.log.20170304 /common/activity/get-ext 18-21

api : /common/activity/get-ext

requestNum: 5

totalSpendTime: 111

avgSpendTime: 22.2

[yushaolong@10-10-213-219 test]$

书本资料：

awk\_sed.pdf (约32M)

总结

实践是最好的老师！！！