

[登录](#) | [免费注册](#) | [忘记密码](#) | [会员登录](#) | [搜索](#) | [帮助](#)

[保留] awk初学之常见问题

[1](#) [2](#) [»](#)[发表主题](#)[发起投票](#)[回复帖子](#)[首页](#) » [论坛](#) » [Shell](#) »[版块跳转](#)[最近访问的版块](#)[有奖讨论：服务器售后那点事](#) | [Power中国用户组：Power技术交流基地！](#) | [2010优秀版主优秀会员评选揭晓！](#) | [买域名送域名 海外主机免备案](#)[\[打印\]](#) [\[收藏\]](#) [\[本帖文本页\]](#) [\[推荐此主题给朋友\]](#)

yinyuemi (SS)



天使

帖子 [1043](#)主题 [25](#)精华 [1](#)

可用积分 4510

专家积分 10 (本版:0)

在线时间 1738 小时

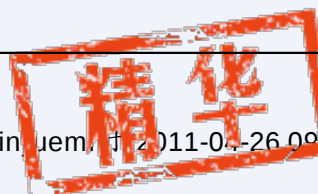
注册时间 2008-06-26

最后登录 2011-04-26

1 楼 发表于 2011-04-26 07:32 | 只看该作者

[\[倒序看帖\]](#)

本帖最后由 yinyuemi 于 2011-04-26 09:44 编辑



从初学awk到现在小有所成，非常感谢CUers的帮助，总结了下自己曾经遇到的问题和犯的错误，供初学者借鉴，因本人非计算机专业，对专业词汇可能有表述不对的地方，还请指正和补充！

1. awk '{code}1' 中的“1”是干什么的？

一个完整的awk语句为：Awk '[patten]{action}.....'，其中pattern缺省为1，action缺省为{print}。

那么awk '1'完整的写法就是awk '{code}1{print}'; 同理，awk '{print}'完整的写法为awk '1{print}'。

2. NR和FNR的区别是啥？

NR: 当前行记录数。

FNR: 当前文件的行记录数。

当awk处理的文件数超过1时，NR和FNR才会有区别。例如：

```
01. cat file
02. a
03. b
04. c
05. d
06. e
07. f
08.
09. awk '{print "NR = " NR " FNR = " FNR, $0}' file
10. NR = 1 FNR = 1 a
11. NR = 2 FNR = 2 b
12. NR = 3 FNR = 3 c
13. NR = 4 FNR = 4 d
14. NR = 5 FNR = 5 e
15. NR = 6 FNR = 6 f
16.
17. awk '{print "NR = " NR " FNR = " FNR, $0}' file file
18. NR = 1 FNR = 1 a
19. NR = 2 FNR = 2 b
20. NR = 3 FNR = 3 c
21. NR = 4 FNR = 4 d
22. NR = 5 FNR = 5 e
23. NR = 6 FNR = 6 f
24. NR = 7 FNR = 1 a
25. NR = 8 FNR = 2 b
26. NR = 9 FNR = 3 c
27. NR = 10 FNR = 4 d
```

28. NR = 11 FNR = 5 e

29. NR = 12 FNR = 6 f

[复制代码](#)

3. Awk怎么引入变量？

有两种方法：

<1>: `awk -v var=$VAR '{code}'`

<2>: `awk '{CODE}'$VAR'{CODE}'`

例如：

01. VAR=XXX

02.

03. `awk -v var=$VAR 'BEGIN{print var}'`

04. XXX

05.

06. `awk 'BEGIN{print ""$VAR}"'`

07. XXX

[复制代码](#)

我推荐使用第一种方法，这样可以避免一些不必要的烦恼。如<http://bbs.chinaunix.net/thread-1835620-1-1.html>

4. 为什么OFS不起作用？

先看一个例子：

01. `echo 'aaa bbb ccc ddd'`

02. aaa bbb ccc ddd

03. aaa bbb ccc ddd

04. `aaa bbb ccc ddd' |awk -v OFS="|" '{print $0}'`

```
05.  aaa bbb ccc ddd
06.  aaa bbb ccc ddd
07.  aaa bbb ccc ddd
08.  aaa bbb ccc ddd
复制代码
```

上面的例子中OFS为什么没有生效呢，原因是OFS指的是输出字段分隔符，所以必须对字段进行操作时OFS才会起作用，正确的方法应该是：

```
01.  echo 'aaa bbb ccc ddd'
02.  aaa bbb ccc ddd
03.  aaa bbb ccc ddd
04.  aaa bbb ccc ddd' |awk -v OFS="|" '{ $1=$1; print $0 }'
05.  aaa|bbb|ccc|ddd
06.  aaa|bbb|ccc|ddd
07.  aaa|bbb|ccc|ddd
08.  aaa|bbb|ccc|ddd
复制代码
```

正如Tim大师所讲的，\$1=\$1这个action，是我们对awk撒的谎，目的就是为了让OFS生效，除此之外，NF+=0也是常有的方法。参考：<http://bbs.chinaunix.net/viewthread.php?tid=1354674&extra=&page=1>

5. 同样的代码，别人运行成功，为什么我运行失败？

这个问题的原因很多，我这里列举两个最常见的，大家可以补充。

<1>: awk版本引起的，如gawk中的一些扩展函数或变量，在nawk中没有，或是不同版本的(g/n)awk也会有差别，这样情况需要重新编写。

<2>: 文本格式的问题，cat-A file查看一下，如果是，dos2unix应该可以解决。

注：书写错误也有可能哦

6. Awk 语句中可以使用{n,m}这样的正则么？

可以，使用方法：`gawk -- re-interval`，其它版本使用方法会有所不同，请大家补充

7. BEGIN 和END 到底是怎么一回事？

有时，对于新手可能也会是个问题。简单说下：

`BEGIN {action}`：读取文本之前进行的操作。要避免类似下面的写法：

```
01. awk 'BEGIN{ filename = FILENAME}' file
02. # or :
03. awk 'BEGIN{FS=":"; for(i=2;i<=NF;i++) print $i}' file
```

[复制代码](#)

如果BEGIN 模块中使用getline函数时，情况会有所不同：

```
01. cat file
02. 1
03. 2
04. 3
05. 4
06. 5
07.
08. awk 'BEGIN{ while (getline <"file") print }' file
09. 1
10. 2
11. 3
12. 4
13. 5
```

END {action}:

它在整个输入文件处理完成后被执行，同样无法对文本进行任何操作，如匹配某个pattern执行action。

8. print，printf 和sprintf？

Print：为一般的打印

print：可以定义打印格式

sprintf：可以完成和printf相同的功能，不同的是sprintf只能输出值，并不能完成打印的功能。

例如：

```
01. awk 'BEGIN{var=123; print "var = " var}'
02. var = 123
03.
04. awk 'BEGIN{var=123;printf "%s %5f\n", "var =",var}'
05. var = 123.000000
06.
07. awk 'BEGIN{var=123;sprintf ("%s %5f\n", "var =",var)}'
08.
09.
10. awk 'BEGIN{var1=123;var2=sprintf ("%5f",var1); print "var2 =" var2}'
11. var2 = 123.000000
```

[复制代码](#)

9. “a==b?c:d”？

这个是一个if语句的简写，即conditional expression1 ? expression2: expression3；完整写法为：

if(a==b) {c} else {d}

10. awk '! a[\$0]++' 怎么理解？

这是一个非常经典的去重复项的awk语句，虽然短小，不过涉及到了不少知识点，下面一一解读：

<1> ："!" 即非。

<2> ：a[\$0]，以\$0为数据下标，建立数组a

<3> ：a[\$0]++，即给数组a赋值，a[\$0]+=1

<4> ：那么组合起来，awk是怎么执行!a[\$0]++的呢？我用一个实际例子来解释：

```
01. cat file
02. 111
03. 222
04. 111
05. 222
06. 333
07.
08. awk '{print a[$0],!a[$0]++,a[$0],[a[$0],$0}' file
09. 1 1 0 111
10. 1 1 0 222
11. 1 0 2 0 111
12. 1 0 2 0 222
13. 1 1 0 333
```

[复制代码](#)

原来，第一个a[\$0]的值为空，所以!a[\$0]++是先作判断，结果为1（非空为真，即为1），再作数组赋值a[\$0]++。这也就是为什么前面的!a[\$0]++并不一定等于后面的!a[\$0]。

awk '++a[\$0]==1' 和上面的代码作用一样，你理解了么？

11. 如何打印单双引号？

```
01. awk 'BEGIN {print "single quote --> \"\"";print "double quote --> \\\\\" }'  
02. single quote --> '  
03. double quote --> "  
复制代码
```

更可靠的的方法如Tim所示：

```
01. awk 'BEGIN {print "single quote --> \047";print "double quote --> \042" }'  
复制代码
```

12. awk 语句中多个{}是怎么执行的？

还是用个例子来说明：

```
01. cat file  
02. 1  
03. 2  
04. 3  
05. 4  
06. 5  
07.  
08. awk '$1==3{printf "\\ " "$0";printf " @@ " "$0";print $0}' file # 这个语句中包含三个action  
09. @@ 11 # 判断$1==3?否；执行 action {printf " @@ " "$0"; 执行 action {print $0}  
10. @@ 22 # 判断$1==3?否；执行 action {printf " @@ " "$0"; 执行 action {print $0}  
11. || 3 @@ 33 # 判断$1==3?是，执行{print "\\ " "$0"; 执行 action {printf " @@ " "$0"; 执行 action {print $0}  
12. @@ 44 # 判断$1==3?否；执行 action {printf " @@ " "$0"; 执行 action {print $0}  
13. @@ 55 # 判断$1==3?否；执行 action {printf " @@ " "$0"; 执行 action {print $0}  
复制代码
```

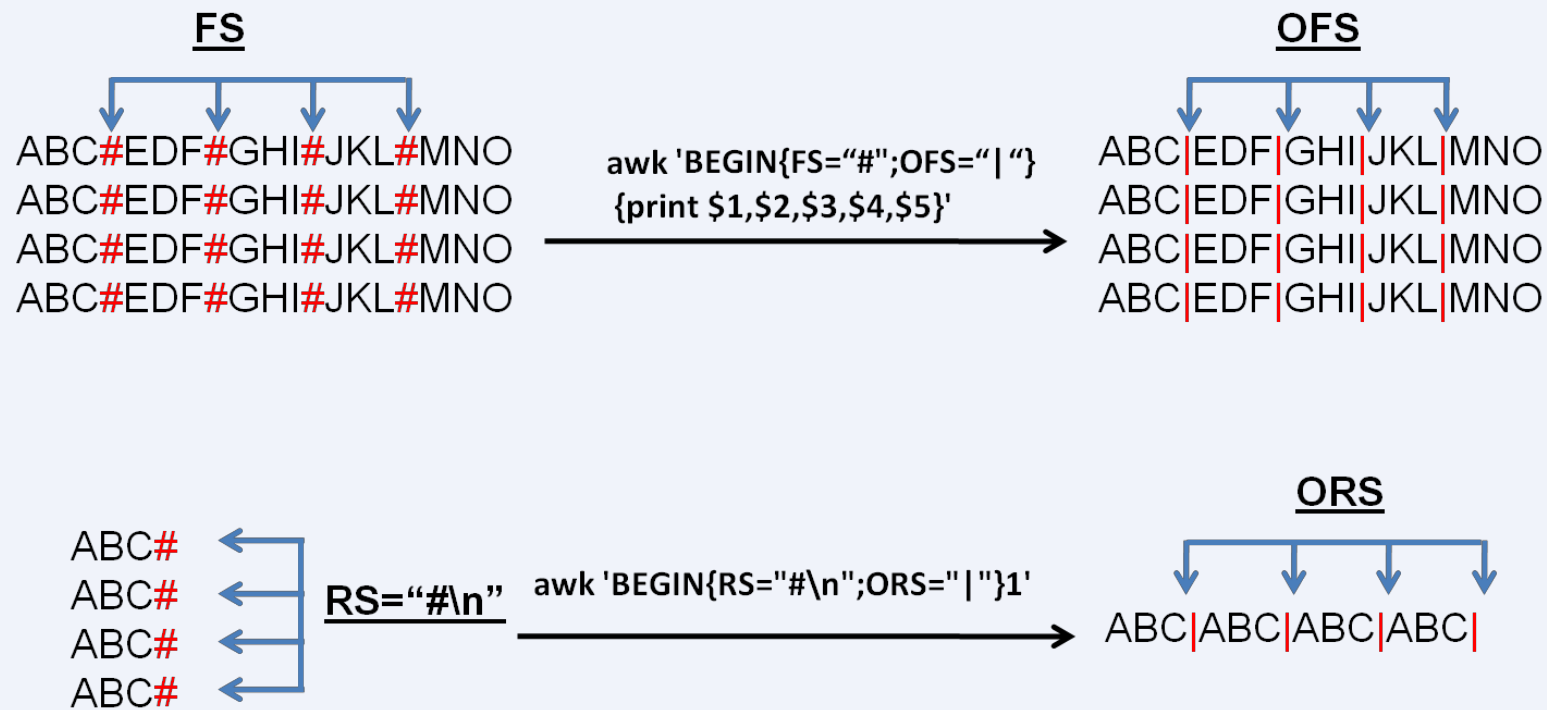

这样可以清楚的看出，awk是一行一行读取的，按照代码的前后顺序，从头到尾执行。但如果action中包含next或exit时，有所不同：

```
01. awk '$1==3{printf "|| "$0;next}{printf "@@ "$0}{print $0}' file
02. @@ 11
03. @@ 22
04. || 3@@ 44
05. @@ 55
06.
07. awk '$1==3{printf "|| "$0;exit}{printf "@@ "$0}{print $0}' file
08. @@ 11
09. @@ 22
10. || 3
    复制代码
```

13. FS, OFS, RS, ORS?

最后用图解的方式说明一下这四个变量：

Picture2.png (38.81 KB)



4

评分人数



xiaopan3322: 很不错，学习..... 可用积分 + 5



expert1: a 可用积分 + 10



ywlscpl: 总结的很不错 可用积分 + 6



ly5066113: 善于总结是一个好习惯，再接再厉！ 可用积分 + 5

本主题由 expert1 于 2011-04-26 09:58 鉴定为 精华

思考是一种习惯
□

UnixWolrd大会&CU网友AIX沙龙 | 2010优秀版主优秀会员评选揭晓！ | 2011数据库技术大会PPT下载 | C语言基本概念一网打尽！！

yinyuemi (SS)



天使
☆☆☆☆☆
帖子 1043
主题 25
精华 1
可用积分 4510
专家积分 10 (本版:0)
在线时间 1738 小时
注册时间 2008-06-26
最后登录 2011-04-26
状态：...保密...
[微博][博客][短信]

2 楼 发表于 2011-04-26 08:41 | 只看该作者



多谢Tim哥支持👍

思考是一种习惯
□

UnixWolrd大会&CU网友AIX沙龙 | 2010优秀版主优秀会员评选揭晓！ | 2011数据库技术大会PPT下载 | C语言基本概念一网打尽！！

ly5066113 (Tim)



法师

3 楼 发表于 2011-04-26 08:44 | 只看该作者



有几个地方感觉不是很准确。

QUOTE:



帖子 6606
主题 56
精华 2
可用积分 18711
专家积分 857 (本版:786)
来自 长春
在线时间 4416 小时
注册时间 2006-09-13
最后登录 2011-04-26
状态：...当前离线...
[\[微博\]](#)[\[博客\]](#)[\[短信\]](#)

6. Awk 语句中可以使用{n,m}这样的正则么？
可以，不过是一种GNU的扩展，使用方法：gawk -- re-interval

{n,m}这样的正则是什么版本的awk里都可以使用，并不是GNU的扩展，只不过使用的方式有所不同。

QUOTE:

7. BEGIN 和END 到底是怎么回事？
有时，对于新手可能也会是个问题。简单说下：
BEGIN {action}：读取文本之前进行的操作，所以不可以妄想在BEGIN的{action}中，对文本进行任何操作，或获取任何文本的信息。

这个说法太绝对了，在BEGIN中可以使用getline函数来操作文本。

QUOTE:

11. 如何打印单双引号？

这个我更倾向与用八进制数来表示，利用shell引号的解析来做看起来很混乱，就像变量的引用一样

01. awk 'BEGIN {print "single quote --> \047";print "double quote --> \042" }'
[复制代码](#)

记住该记住的，忘记该忘记的。改变能改变的，接受不能改变的。

UnixWolrd大会&CU网友AIX沙龙 | 2010优秀版主优秀会员评选揭晓！ | 2011数据库技术大会PPT下载 | C语言基本概念一网打尽！！

ywlscpl (白云苍狗)



精灵使
☆☆☆☆☆
帖子 4577
主题 68

4 楼 发表于 2011-04-26 08:45 | 只看该作者



支持

我们要用纯洁的眼光来看世界

一如我们要相信陈老师是真的只是单纯地想给阿娇拍张照。

精华 0
可用积分 37850
专家积分 140 (本版:140)
在线时间 2531 小时
注册时间 2005-12-11
最后登录 2011-04-26
状态: ...当前在线...
[微博][博客][短信]



UnixWolrd大会&CU网友AIX沙龙| 2010优秀版主优秀会员评选揭晓！| 2011数据库技术大会PPT下载| C语言基本概念一网打尽！！

yinyuemi (SS)



天使
★☆☆☆☆

帖子 1043
主题 25
精华 1

5 楼 发表于 2011-04-26 08:46 | 只看该作者



QUOTE:

有几个地方感觉不是很准确。

{n,m}这样的正则是什么版本的awk里都可以使用，并不是GNU的扩展，只不过使 ...
ly5066113 发表于 2011-04-26 08:44

<div>可用积分 4510</div> <div>专家积分 10 (本版:0)</div> <div>在线时间 1738 小时</div> <div>注册时间 2008-06-26</div> <div>最后登录 2011-04-26</div> <div>状态：...保密...</div> <div>[微博][博客][短信]</div>	<div>好的，马上改正🙏</div> <div></div> <div>思考是一种习惯</div> <div></div> <div>UnixWolrd大会&CU网友AIX沙龙 2010优秀版主优秀会员评选揭晓！ 2011数据库技术大会PPT下载 C语言基本概念一网打尽！！</div>
<div>unixlinuxsys (幽灵)</div> <div>linux小学5年级</div> <div></div> <div>法师</div> <div></div> <div>帖子 6579</div> <div>主题 272</div> <div>精华 0</div> <div>可用积分 6175</div> <div>专家积分 10 (本版:0)</div> <div>在线时间 717 小时</div> <div>注册时间 2009-04-02</div> <div>最后登录 2011-04-26</div> <div>状态：...当前在线...</div> <div>[微博][博客][短信]</div>	<div>6 楼 发表于 2011-04-26 09:03 只看该作者</div> <div>好筒子</div> <div></div> <div>18688829392@wo.com.cn</div> <div>Cisco & Polycom & Tandberg /Solutions Expert</div> <div>网络、音视频通信专家</div> <div>-----“如果你想在硬件市场立足，你必须在网络上有所竞争力。”-----</div> <div>UnixWolrd大会&CU网友AIX沙龙 2010优秀版主优秀会员评选揭晓！ 2011数据库技术大会PPT下载 C语言基本概念一网打尽！！</div>
<div>blackold (黑哥)</div> <div></div>	<div>7 楼 发表于 2011-04-26 09:09 只看该作者</div> <div>这种分享的精神绝对要支持！</div>



大法师
★☆☆☆☆
帖子 11072
主题 87
精华 0
可用积分 12738
专家积分 369 (本版:369)
来自 南宁
在线时间 4477 小时
注册时间 2007-10-22
最后登录 2011-04-26
状态: ...当前离线...
[\[微博\]](#)[\[博客\]](#)[\[短信\]](#)

有些地方可能值得商榷，如：

10.

QUOTE:

原来，第一个a[\$0]的值为空，由于“!”的运算级别高于“+”

++与+是不同的操作符，还有++的优先级比!高。

13. 里面的图示是不是弄反了？

LIVE FREE OR DIE! K.I.S.S. DRY

远离shell

[UnixWolrd大会&CU网友AIX沙龙](#) | [2010优秀版主优秀会员评选揭晓！](#) | [2011数据库技术大会PPT下载](#) | [C语言基本概念一网打尽！！](#)

yinyuemi (SS)



8 楼 发表于 2011-04-26 09:26 | 只看该作者 TOP

回复 **7# blackold**

图已更正，黑哥真是火眼👁

++比！级别高的话，!a[\$0]++是先执行a[\$0]++么？那!a[\$0]++不应该是0么？有点晕，黑哥指点下

<div>天使</div> <div>☆☆☆☆☆</div> <div>帖子1043</div> <div>主题25</div> <div>精华1</div> <div>可用积分4510</div> <div>专家积分10 (本版:0)</div> <div>在线时间1738 小时</div> <div>注册时间2008-06-26</div> <div>最后登录2011-04-26</div> <div>状态: ...保密...</div> <div>[微博][博客][短信]</div>	<div></div> <div>思考是一种习惯</div> <div></div> <div>UnixWolrd大会&CU网友AIX沙龙 2010优秀版主优秀会员评选揭晓! 2011数据库技术大会PPT下载 C语言基本概念一网打尽!!</div>
<div>blackold (黑哥)</div> <div></div> <div>大法师</div> <div>☆☆☆☆☆</div> <div>帖子11072</div> <div>主题87</div> <div>精华0</div> <div>可用积分12738</div> <div>专家积分369 (本版:369)</div> <div>来自南宁</div> <div>在线时间4477 小时</div> <div>注册时间2007-10-22</div> <div>最后登录2011-04-26</div> <div>状态: ...当前离线...</div> <div>[微博][博客][短信]</div>	<div>9 楼 发表于 2011-04-26 09:38 只看该作者</div> <div>回复 8# yinyuemi</div> <div>这是前加/后加的区别，前几天还说过。</div> <div>后加: 先使用变量的值，再自加。</div> <div>!a[\$0]++ 对这个表达式的求值，它的值与 !a[\$0] 相同(先使用变量a[\$0]的值)，但对表达式求值后 a[\$0]会自加。</div> <div>图里面的 RS值前后不一致啊。</div> <div>LIVE FREE OR DIE! K.I.S.S. DRY</div>

<div>[微博][阿合][论坛]</div>	<div>远离<div>shell</div></div> <div>UnixWolrd大会&CU网友AIX沙龙 2010优秀版主优秀会员评选揭晓！ 2011数据库技术大会PPT下载 C语言基本概念一网打尽！！</div>
<div>yinyuemi (SS)</div> <div></div> <div>天使 ★☆☆☆☆</div> <div>帖子1043 主题25 精华1 可用积分4510 专家积分10 (本版:0) 在线时间1738 小时 注册时间2008-06-26 最后登录2011-04-26 状态: ...保密... [微博][博客][短信]</div>	<div>10 楼 发表于 2011-04-26 09:47 只看该作者<div>TOP</div></div> <div><div>QUOTE:</div><div>回复 yinyuemi 这是前加/后加的区别，前几天还说过。 后加: 先使用变量的值，再自加。 ... blackold 发表于 2011-04-26 09:38</div></div> <div>我知道问题出哪里了，我之前“困”在那个运算级别的判断了，现在弄明白了，多谢黑哥</div> <div>思考是一种习惯</div> <div>UnixWolrd大会&CU网友AIX沙龙 2010优秀版主优秀会员评选揭晓！ 2011数据库技术大会PPT下载 C语言基本概念一网打尽！！</div>

发表主题

发起投票

回复帖子

首页 » 论坛 » Shell »

感谢所有关心和支持过ChinaUnix的朋友们 转载本站内容请注明原作者名及出处
京ICP证:060528号
北京市公安局海淀分局网监中心备案编号：1101082001

[清除 Cookies](#) - [ChinaUnix](#) - [Archiver](#) - [WAP](#) - [TOP](#)