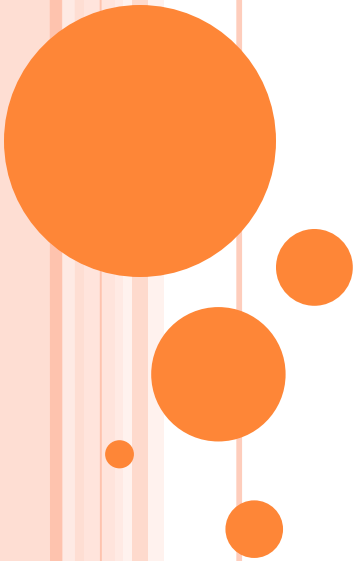
CS5354
UNIX TOOL PROGRAMMING



Awk : An Advanced Filters

- A. Sawarkar

Name.

Made in 1977

Authors: **A**ho, **W**einberger and **K**ernighan (AWK).

Like sed, it combine features of several filters.

Unlike othres filters, it operates at ***filed*** level and can easily access , transform and format individual fileds in line.

It also support extended regular expression(ERE)

In awk we use ? And + for RE.

g+: g,gg,ggg.....

Knowing awk will help you understand perl, which uses most of the awk constructs.

Simple awk filtering

Syntax:

awk *options* ‘ selection_criteria {action}’ file(s)

selection_criteria in **awk** have wider scope than in **sed**

e.g $awk ‘/director/ {print}’ emp.lst

```
adhoc@adhoc:~/Desktop$ awk '/director/ {print}' emp.lst
9876 | jai sharma         | director | production | 12/03/50 | 7000
2365 | barun sengupta     | director | personnel  | 11/05/47 | 7800
1006 | chanchal singhvi   | director | sales      | 03/09/38 | 6700
6521 | lalit chowduryg*   | director | marketing  | 26/09/45 | 8200
adhoc@adhoc:~/Desktop$ awk '/director/ {p}' emp.lst
adhoc@adhoc:~/Desktop$
```

Note: **Printing** is default action of **awk**.

```
adhoc@adhoc:~/Desktop$ awk '/director/ ' emp.lst
9876 | jai sharma         | director | production | 12/03/50 | 7000
2365 | barun sengupta     | director | personnel  | 11/05/47 | 7800
1006 | chanchal singhvi   | director | sales      | 03/09/38 | 6700
6521 | lalit chowduryg*   | director | marketing  | 26/09/45 | 8200
adhoc@adhoc:~/Desktop$
```

Simple awk filtering

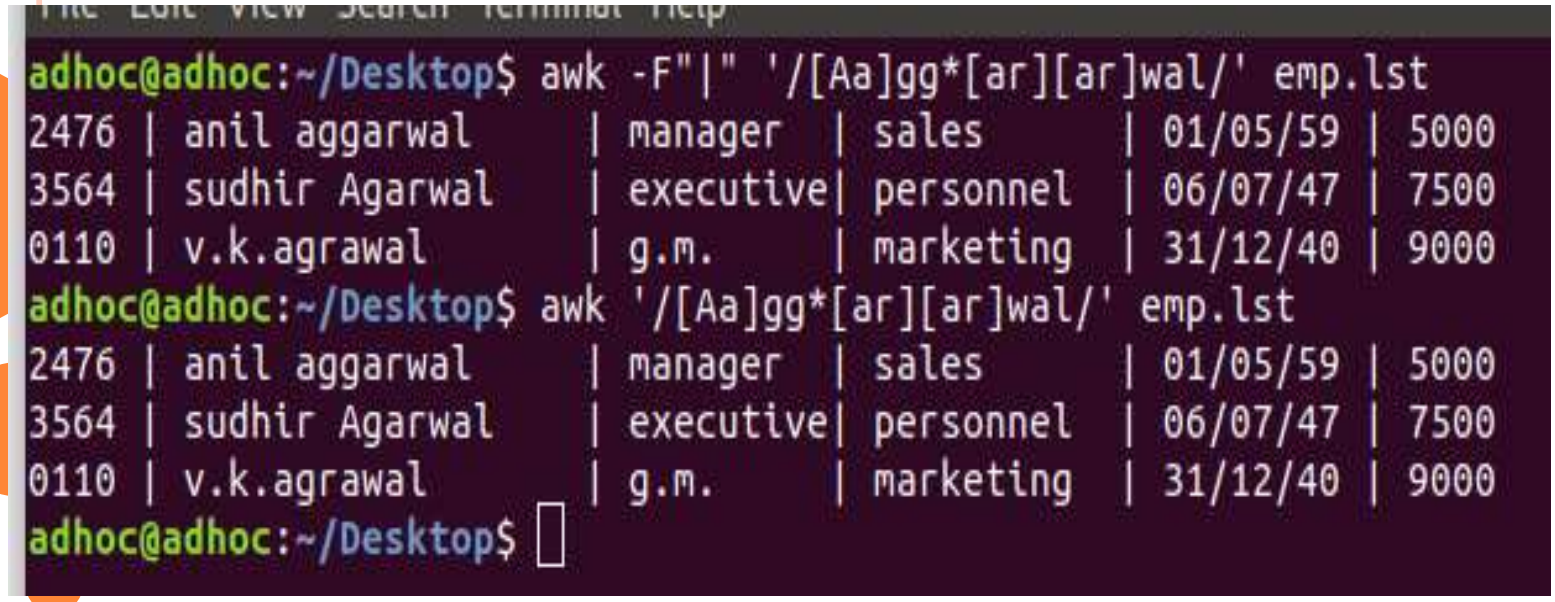
For **pattern** matching **awk** uses regular expression in **sed** style

e.g

$awk -F"|" '/[Aa]gg*[ar][ar]wal/' emp.lst

-F"|": what filed seprator to use.

Note: Default awk work on line.



Spliting a line into filed

Awk uses **$0**, to indicate the entire line.
$1 onwards is specific field

$awk -F"|" '/sales/ {print $2,$3,$4,$6}' emp.lst

```
3212 | shyam saksena     | d.g.m.    | accounts   | 12/12/55 | 6000
3564 | sudhir Agarwal    | executive | personnel  | 06/07/47 | 7500
2345 | j. b. sexena      | g.m.      | marketing  | 12/03/45 | 8000
0110 | v.k.agrawal       | g.m.      | marketing  | 31/12/40 | 9000
adhoc@adhoc:~/Desktop$ awk -F"|" '/sales/ {print $2,$3,$4,$6}' emp.lst
 a.k.shukla          g.m.       sales        6000
 chanchal singhvi    director   sales        6700
 s.n. dasgupta       manager    sales        5600
 anil aggarwal       manager    sales        5000
adhoc@adhoc:~/Desktop$
```

For selecting lines use built-in variable **NR**(line number)
$awk -F"|" '**NR** == 3, **NR** == 6 {print **NR**, $2,$3,$6}' emp.lst

```
3212 | shyam saksena     | d.g.m.    | accounts   | 12/12/55 | 6000
3564 | sudhir Agarwal    | executive | personnel  | 06/07/47 | 7500
2345 | j. b. sexena      | g.m.      | marketing  | 12/03/45 | 8000
0110 | v.k.agrawal       | g.m.      | marketing  | 31/12/40 | 9000
adhoc@adhoc:~/Desktop$ awk -F"|" 'NR == 3, NR == 6 {print NR, $2,$3,$6 }' emp.lst
3  sumit chakrobarty   d.g.m.     6000
4  barun sengupta          director   7800
5  n.k.gupta       chairman   5400
6  chanchal singhvi    director   6700
adhoc@adhoc:~/Desktop$
```

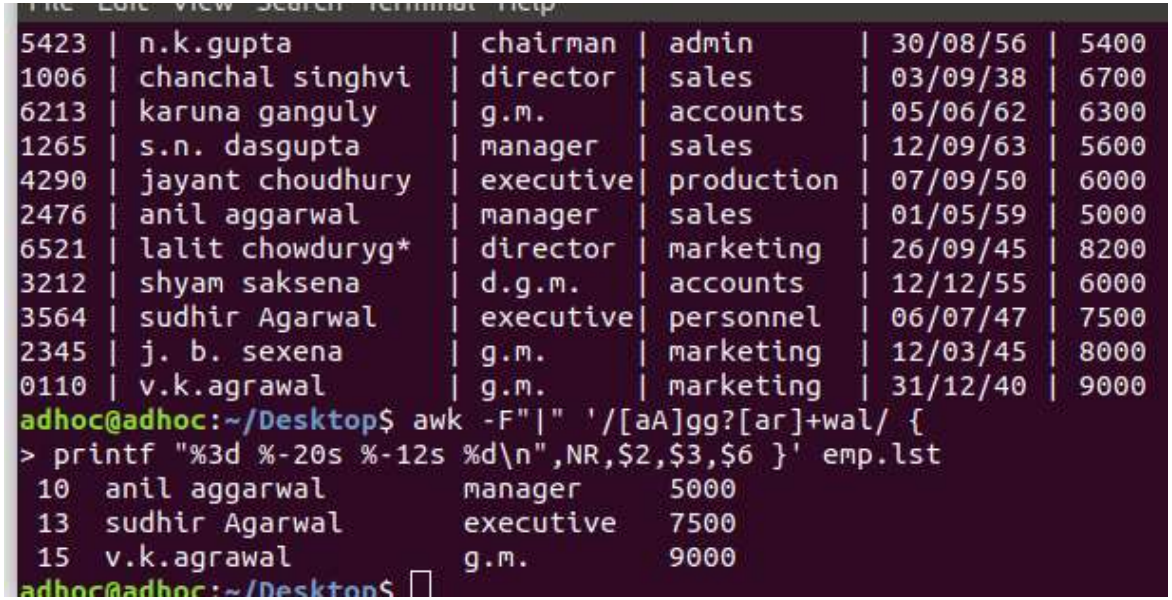
printf: Formatting output

Like a C printf statement , we can use in awk also

%s: format used for string data

%d: format used for numeric data

$awk -F"|" '/[aA]gg?[ar]+wal/ {printf "%3d % -20s % -12s %d\n",
NR,$2,$3,$6 }' emp.lst



Note: (% -20s)- > total space 20 alloacte and -ve means left.

Variables and Expressions

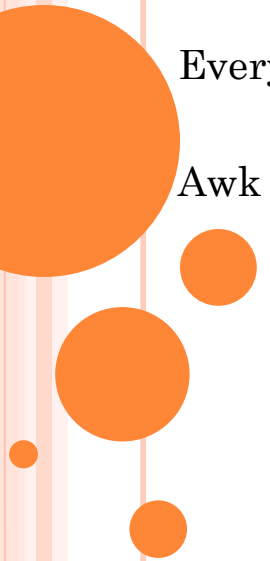
Expressions comprise strings, numners, varaibles and operators.

Expr: (a+b)*c+12

Unlike in programming langauages, awk **doesn't** have char ,int, long ,double primitive data types.

Every expression can be interpreted either as string or numbers.

Awk also support user define varaibles which are case sensitive.



Comparison operators

Select name, designation and salary only for director or chairman

```
$awk -F"|" '$3 == "director" || $3 == "chairman" {
  Printf "%-20s %-12s %d\n", $2,$3,$6 }' emp.lst
```

Select name, designation and salary of other than director and chairman

```
$awk -F"|" '$3 != "director" && $3 !="chairman" {
   Printf "%-20s %-12s %d\n", $2,$3,$6 }' emp.lst
```

~ and !~operator

To locate only **g.m.'s**

$awk -F"|" '$3 ~ *g.m.* { printf "....}' emp.lst

```
adhoc@adhoc:~/Desktop$ awk -F"|" '$3 ~ /g.m./' emp.lst
2233 | a.k.shukla          | g.m.    | sales      | 12/12/52 | 6000
5678 | sumit chakrobarty   | d.g.m.  | marketing  | 19/04/43 | 6000
6213 | karuna ganguly      | g.m.    | accounts   | 05/06/62 | 6300
3212 | shyam saksena       | d.g.m.  | accounts   | 12/12/55 | 6000
2345 | j. b. sexena        | g.m.    | marketing  | 12/03/45 | 8000
0110 | v.k.agrawal         | g.m.    | marketing  | 31/12/40 | 9000
```

To locate only and only g.m.

$awk -F"|" '$3~/^g.m./ {print "....}' emp.lst

```
adhoc@adhoc:~/Desktop$ awk -F"|" '$3 ~ /^ g.m./' emp.lst
2233 | a.k.shukla          | g.m.    | sales      | 12/12/52 | 6000
6213 | karuna ganguly      | g.m.    | accounts   | 05/06/62 | 6300
2345 | j. b. sexena        | g.m.    | marketing  | 12/03/45 | 8000
0110 | v.k.agrawal         | g.m.    | marketing  | 31/12/40 | 9000
adhoc@adhoc:~/Desktop$
```

Number Comparsion

Operator	Significance
<	Less than
<=	Less than or equal to
==	Equal to
!=	Not equal to
>=	Greater than or equal to
>	Greater than
~	Matches a regular exppression
!~	Doesn't match a regular expression

```
$awk -F"|" '$6 > 7500 {
Printf"%-20s %-12s %d\n",$2,$3,$6 }' emp.lst
```

Number Comparsion

Select , either born in 1945 or salary greater than 8000

```
adhoc@adhoc:~$ cd Desktop
adhoc@adhoc:~/Desktop$ awk -F"|" '$6 > 8000 || $5 ~/45$/' emp.lst
6521 | lalit chowduryg*  | director | marketing  | 26/09/45 | 8200
0110 | v.k.agrawal      | g.m.     | marketing  | 31/12/40 | 9000
adhoc@adhoc:~/Desktop$
```

Show only name , designation and salary whose salary greater than 7500.

$awk -F"|" '$6 > 7500 {

Printf "%-20s %-12s %d\n", $2,$3,$6}' emp.lst

```
adhoc@adhoc:~/Desktop$ awk -F"|" '$6 >7500 {
> printf "%-20s %-12s %d\n", $2,$3,$6 }' emp.lst
 barun sengupta              director    7800
 lalit chowduryg*     director    8200
 j. b. sexena         g.m.        8000
 v.k.agrawal          g.m.        9000
adhoc@adhoc:~/Desktop$ awk -F"|" '$6 >7500 {
printf "%-20s %-12s %d", $2,$3,$6 }' emp.lst
 barun sengupta              director    7800 lalit chowduryg*     direc
tor    8200 j. b. seadadhoc@adhoadhoc@adhoadadadhoc@adadadadadadadadadhoad
adhoc@adhoc:~/Desktop$
```

Number Comparsion