**awk command or scripting**

awk is a scripting language, and it is helpful when working in the command line. It's also a widely used command for text processing.

When using awk, you are able to select data – one or more pieces of individual text – based on a pattern you provide.

For example, some of the operations you can do with awk are searching for a specific word or pattern in a piece of text given, or even select a certain line or a certain column in a file you provide.

### **Basic Syntax**

The basic syntax of the awk command is:

awk option 'pattern { action }' file\_name

Note: Get sample csv data from site <https://extendsclass.com/csv-generator.html>

1. To print *all* the contents of a file, the action you specify inside the curly braces is **print $0.**

**root@ip-172-31-36-210 ~]# awk -F ":" '{print $0}' /etc/passwd**

**root:x:0:0:root:/root:/bin/bash**

**bin:x:1:1:bin:/bin:/sbin/nologin**

**daemon:x:2:2:daemon:/sbin:/sbin/nologin**

**adm:x:3:4:adm:/var/adm:/sbin/nologin**

**lp:x:4:7:lp:/var/spool/lpd:/sbin/nologin**

**sync:x:5:0:sync:/sbin:/bin/sync**

**shutdown:x:6:0:shutdown:/sbin:/sbin/shutdown**

**halt:x:7:0:halt:/sbin:/sbin/halt**

**mail:x:8:12:mail:/var/spool/mail:/sbin/nologin**

**operator:x:11:0:operator:/root:/sbin/nologin**

**games:x:12:100:games:/usr/games:/sbin/nologin**

**ftp:x:14:50:FTP User:/var/ftp:/sbin/nologin**

**nobody:x:65534:65534:Kernel Overflow User:/:/sbin/nologin**

**dbus:x:81:81:System message bus:/:/sbin/nologin**

**systemd-network:x:192:192:systemd Network Management:/:/usr/sbin/nologin**

**systemd-oom:x:999:999:systemd Userspace OOM Killer:/:/usr/sbin/nologin**

**systemd-resolve:x:193:193:systemd Resolver:/:/usr/sbin/nologin**

**sshd:x:74:74:Privilege-separated SSH:/usr/share/empty.sshd:/sbin/nologin**

**rpc:x:32:32:Rpcbind Daemon:/var/lib/rpcbind:/sbin/nologin**

**libstoragemgmt:x:997:997:daemon account for libstoragemgmt:/:/usr/sbin/nologin**

**systemd-coredump:x:996:996:systemd Core Dumper:/:/usr/sbin/nologin**

**systemd-timesync:x:995:995:systemd Time Synchronization:/:/usr/sbin/nologin**

**chrony:x:994:994:chrony system user:/var/lib/chrony:/sbin/nologin**

**ec2-instance-connect:x:993:993::/home/ec2-instance-connect:/sbin/nologin**

**rpcuser:x:29:29:RPC Service User:/var/lib/nfs:/sbin/nologin**

**tcpdump:x:72:72::/:/sbin/nologin**

**ec2-user:x:1000:1000:EC2 Default User:/home/ec2-user:/bin/bash**

**[root@ip-172-31-36-210 ~]#**

**NR – no of row/record**

**NF – no of field  
$0 – print every things**

**$1 , $2 – etc., represent the first, second, etc., fields in a record.**

**2 ) you would like each line to have a line-number count, you would use the NR built-in variable:**

**root@ip-172-31-36-210 ~]# awk -F ":" '{print NR,$0}' /etc/passwd**

**1 root:x:0:0:root:/root:/bin/bash**

**2 bin:x:1:1:bin:/bin:/sbin/nologin**

**3 daemon:x:2:2:daemon:/sbin:/sbin/nologin**

**4 adm:x:3:4:adm:/var/adm:/sbin/nologin**

**5 lp:x:4:7:lp:/var/spool/lpd:/sbin/nologin**

**6 sync:x:5:0:sync:/sbin:/bin/sync**

**7 shutdown:x:6:0:shutdown:/sbin:/sbin/shutdown**

**8 halt:x:7:0:halt:/sbin:/sbin/halt**

**9 mail:x:8:12:mail:/var/spool/mail:/sbin/nologin**

**10 operator:x:11:0:operator:/root:/sbin/nologin**

**11 games:x:12:100:games:/usr/games:/sbin/nologin**

**12 ftp:x:14:50:FTP User:/var/ftp:/sbin/nologin**

**13 nobody:x:65534:65534:Kernel Overflow User:/:/sbin/nologin**

**14 dbus:x:81:81:System message bus:/:/sbin/nologin**

**15 systemd-network:x:192:192:systemd Network Management:/:/usr/sbin/nologin**

**16 systemd-oom:x:999:999:systemd Userspace OOM Killer:/:/usr/sbin/nologin**

**17 systemd-resolve:x:193:193:systemd Resolver:/:/usr/sbin/nologin**

**18 sshd:x:74:74:Privilege-separated SSH:/usr/share/empty.sshd:/sbin/nologin**

**19 rpc:x:32:32:Rpcbind Daemon:/var/lib/rpcbind:/sbin/nologin**

**20 libstoragemgmt:x:997:997:daemon account for libstoragemgmt:/:/usr/sbin/nologin**

**21 systemd-coredump:x:996:996:systemd Core Dumper:/:/usr/sbin/nologin**

**22 systemd-timesync:x:995:995:systemd Time Synchronization:/:/usr/sbin/nologin**

**23 chrony:x:994:994:chrony system user:/var/lib/chrony:/sbin/nologin**

**24 ec2-instance-connect:x:993:993::/home/ec2-instance-connect:/sbin/nologin**

**25 rpcuser:x:29:29:RPC Service User:/var/lib/nfs:/sbin/nologin**

**26 tcpdump:x:72:72::/:/sbin/nologin**

**27 ec2-user:x:1000:1000:EC2 Default User:/home/ec2-user:/bin/bash**

**[root@ip-172-31-36-210 ~]#**

**3) To have the first column printed, you use the command:**

**[root@ip-172-31-36-210 ~]# awk -F ":" '{print $1}' /etc/passwd**

**root**

**bin**

**daemon**

**adm**

**lp**

**sync**

**shutdown**

**halt**

**mail**

**operator**

**games**

**ftp**

**nobody**

**dbus**

**systemd-network**

**systemd-oom**

**systemd-resolve**

**sshd**

**rpc**

**libstoragemgmt**

**systemd-coredump**

**systemd-timesync**

**chrony**

**ec2-instance-connect**

**rpcuser**

**tcpdump**

**ec2-user**

**4) To print more than one column, for example the first and fourth columns, you would do:**

**[root@ip-172-31-36-210 ~]# awk -F ":" '{print $1, $4}' /etc/passwd**

**root 0**

**bin 1**

**daemon 2**

**adm 4**

**lp 7**

**sync 0**

**shutdown 0**

**halt 0**

**mail 12**

**operator 0**

**games 100**

**ftp 50**

**nobody 65534**

**dbus 81**

**systemd-network 192**

**systemd-oom 999**

**systemd-resolve 193**

**sshd 74**

**rpc 32**

**libstoragemgmt 997**

**systemd-coredump 996**

**systemd-timesync 995**

**chrony 994**

**ec2-instance-connect 993**

**rpcuser 29**

**tcpdump 72**

**ec2-user 1000**

**5 ) print the last column**

**[root@ip-172-31-36-210 ~]# awk -F ":" '{print $NF}' /etc/passwd**

**/bin/bash**

**/sbin/nologin**

**/sbin/nologin**

**/sbin/nologin**

**/sbin/nologin**

**/bin/sync**

**/sbin/shutdown**

**/sbin/halt**

**/sbin/nologin**

**/sbin/nologin**

**/sbin/nologin**

**/sbin/nologin**

**/sbin/nologin**

**/sbin/nologin**

**/usr/sbin/nologin**

**/usr/sbin/nologin**

**/usr/sbin/nologin**

**/sbin/nologin**

**/sbin/nologin**

**/usr/sbin/nologin**

**/usr/sbin/nologin**

**/usr/sbin/nologin**

**/sbin/nologin**

**/sbin/nologin**

**/sbin/nologin**

**/sbin/nologin**

**/bin/bash**

**6 ) Search a word - worker**

**[root@ip-172-31-36-210 ~]# awk -F "," '/worker/{print $0}' data.txt**

**101,Cassandra,Atonsah,Cassandra.Atonsah@yopmail.com,worker**

**105,Dolli,Braun,Dolli.Braun@yopmail.com,worker**

**109,Christy,Joeann,Christy.Joeann@yopmail.com,worker**

**110,Feliza,Brackely,Feliza.Brackely@yopmail.com,worker**

**112,Clarice,Blase,Clarice.Blase@yopmail.com,worker**

**113,Gabi,Kunin,Gabi.Kunin@yopmail.com,worker**

**116,Lucy,Christal,Lucy.Christal@yopmail.com,worker**

**120,Dulcinea,Dom,Dulcinea.Dom@yopmail.com,worker**

**122,Basia,Goldina,Basia.Goldina@yopmail.com,worker**

**[root@ip-172-31-36-210 ~]# awk -F "," '/worker/{print NR, $0}' data.txt**

**3 101,Cassandra,Atonsah,Cassandra.Atonsah@yopmail.com,worker**

**7 105,Dolli,Braun,Dolli.Braun@yopmail.com,worker**

**11 109,Christy,Joeann,Christy.Joeann@yopmail.com,worker**

**12 110,Feliza,Brackely,Feliza.Brackely@yopmail.com,worker**

**14 112,Clarice,Blase,Clarice.Blase@yopmail.com,worker**

**15 113,Gabi,Kunin,Gabi.Kunin@yopmail.com,worker**

**18 116,Lucy,Christal,Lucy.Christal@yopmail.com,worker**

**22 120,Dulcinea,Dom,Dulcinea.Dom@yopmail.com,worker**

**24 122,Basia,Goldina,Basia.Goldina@yopmail.com,worker**

**7 ) print only 5th line**

**[root@ip-172-31-36-210 ~]# awk -F "," 'NR==5 {print $0}' data.txt**

**103,Mallory,Sammons,Mallory.Sammons@yopmail.com,developer**

**[root@ip-172-31-36-210 ~]#**

**[root@ip-172-31-36-210 ~]# awk -F "," 'NR==5 {print $5}' data.txt**

**developer**

**[root@ip-172-31-36-210 ~]#**

**8) print line 3 to 7**

**[root@ip-172-31-36-210 ~]# awk -F "," 'NR==3,NR==7 {print NR, $0}' data.txt**

**3 101,Cassandra,Atonsah,Cassandra.Atonsah@yopmail.com,worker**

**4 102,Kristan,Gilbertson,Kristan.Gilbertson@yopmail.com,firefighter**

**5 103,Mallory,Sammons,Mallory.Sammons@yopmail.com,developer**

**6 104,Carilyn,Fiann,Carilyn.Fiann@yopmail.com,firefighter**

**7 105,Dolli,Braun,Dolli.Braun@yopmail.com,worker**

**9) print the line no which is empty**

**[root@ip-172-31-36-210 ~]# awk -F "," 'NF==0{print NR}' data.txt**

**26**

**27**

**[root@ip-172-31-36-210 ~]#**

**10) search multiple words**

**[root@ip-172-31-36-210 ~]# awk -F "," '/Lucy|Robbi|Gabi/{print $0}' data.txt**

**113,Gabi,Kunin,Gabi.Kunin@yopmail.com,worker**

**116,Lucy,Christal,Lucy.Christal@yopmail.com,worker**

**123,Robbi,Ajay,Robbi.Ajay@yopmail.com,firefighter**

**11) ignore case while searching**

**[root@ip-172-31-36-210 ~]# awk -F "," 'BEGIN{IGNORECASE=1}/lucy/{print $0}' data.txt**

**116,Lucy,Christal,Lucy.Christal@yopmail.com,worker**

**[root@ip-172-31-36-210 ~]#**

**12) How to check if a given char is present in column**

**[root@ip-172-31-36-210 ~]# awk -F "," '$2 ~ /e/{print $0}' data.txt**

**id,firstname,lastname,email,profession**

**108,Dennie,Poppy,Dennie.Poppy@yopmail.com,doctor**

**110,Feliza,Brackely,Feliza.Brackely@yopmail.com,worker**

**112,Clarice,Blase,Clarice.Blase@yopmail.com,worker**

**114,Bertine,Eno,Bertine.Eno@yopmail.com,firefighter**

**115,Merrie,Hanshaw,Merrie.Hanshaw@yopmail.com,firefighter**

**117,Florie,Kannry,Florie.Kannry@yopmail.com,developer**

**119,Regina,Agle,Regina.Agle@yopmail.com,doctor**

**120,Dulcinea,Dom,Dulcinea.Dom@yopmail.com,worker**

**121,Marjie,Claudine,Marjie.Claudine@yopmail.com,police officer**

**13) What if a file is having multiple delimeter**

**[root@ip-172-31-36-210 ~]# awk -F [,:] '{print $0}' data2.csv**

**rahul,ranjan:patna-danapur**

**[root@ip-172-31-36-210 ~]# awk -F [,:] '{print $2}' data2.csv**

**ranjan**

**[root@ip-172-31-36-210 ~]#**

**Use Cases Where AWK can be useful**

1. **How to only get Status of service**
2. **How to get list of files**

**Hint NR>1**

1. **How to read logs in range of time**

**awk '$0 >= "2023-01-01 10:00:00" && $0 <= "2023-01-01 12:00:00"' logfile.log**

1. **Get length of a word/row**

**echo "Hi Paul" | awk '{print length($2)}'**

**AWK as Script**

**awk option ‘BEGIN{start\_action}pattern{action}END{end\_action}’ filename**

**awk ‘BEGIN{start\_action}pattern/condition{action}END{end\_action}‘file\_name**

**[root@ip-172-31-36-210 ~]# awk -F ":" 'BEGIN{print "--------USER INFO-------"} {print $0} END{print "---------THE END---------------"}' /etc/passwd**

**--------USER INFO-------**

**root:x:0:0:root:/root:/bin/bash**

**bin:x:1:1:bin:/bin:/sbin/nologin**

**daemon:x:2:2:daemon:/sbin:/sbin/nologin**

**adm:x:3:4:adm:/var/adm:/sbin/nologin**

**lp:x:4:7:lp:/var/spool/lpd:/sbin/nologin**

**sync:x:5:0:sync:/sbin:/bin/sync**

**shutdown:x:6:0:shutdown:/sbin:/sbin/shutdown**

**halt:x:7:0:halt:/sbin:/sbin/halt**

**mail:x:8:12:mail:/var/spool/mail:/sbin/nologin**

**operator:x:11:0:operator:/root:/sbin/nologin**

**games:x:12:100:games:/usr/games:/sbin/nologin**

**ftp:x:14:50:FTP User:/var/ftp:/sbin/nologin**

**nobody:x:65534:65534:Kernel Overflow User:/:/sbin/nologin**

**dbus:x:81:81:System message bus:/:/sbin/nologin**

**systemd-network:x:192:192:systemd Network Management:/:/usr/sbin/nologin**

**systemd-oom:x:999:999:systemd Userspace OOM Killer:/:/usr/sbin/nologin**

**systemd-resolve:x:193:193:systemd Resolver:/:/usr/sbin/nologin**

**sshd:x:74:74:Privilege-separated SSH:/usr/share/empty.sshd:/sbin/nologin**

**rpc:x:32:32:Rpcbind Daemon:/var/lib/rpcbind:/sbin/nologin**

**libstoragemgmt:x:997:997:daemon account for libstoragemgmt:/:/usr/sbin/nologin**

**systemd-coredump:x:996:996:systemd Core Dumper:/:/usr/sbin/nologin**

**systemd-timesync:x:995:995:systemd Time Synchronization:/:/usr/sbin/nologin**

**chrony:x:994:994:chrony system user:/var/lib/chrony:/sbin/nologin**

**ec2-instance-connect:x:993:993::/home/ec2-instance-connect:/sbin/nologin**

**rpcuser:x:29:29:RPC Service User:/var/lib/nfs:/sbin/nologin**

**tcpdump:x:72:72::/:/sbin/nologin**

**ec2-user:x:1000:1000:EC2 Default User:/home/ec2-user:/bin/bash**

**---------THE END---------------**

**[root@ip-172-31-36-210 ~]#**

**2)**

**[root@ip-172-31-36-210 ~]# awk -F ":" '{print $1, $3, $4}' /etc/passwd**

**root 0 0**

**bin 1 1**

**daemon 2 2**

**adm 3 4**

**lp 4 7**

**sync 5 0**

**shutdown 6 0**

**halt 7 0**

**mail 8 12**

**operator 11 0**

**games 12 100**

**ftp 14 50**

**nobody 65534 65534**

**dbus 81 81**

**systemd-network 192 192**

**systemd-oom 999 999**

**systemd-resolve 193 193**

**sshd 74 74**

**rpc 32 32**

**libstoragemgmt 997 997**

**systemd-coredump 996 996**

**systemd-timesync 995 995**

**chrony 994 994**

**ec2-instance-connect 993 993**

**rpcuser 29 29**

**tcpdump 72 72**

**ec2-user 1000 1000**

**[root@ip-172-31-36-210 ~]#**

**If you want to add a header line to the output to clearly indicate what each column represents, you can modify the awk command as follows:**

**[root@ip-172-31-36-210 ~]# awk -F ":" 'BEGIN {print "UserID UID GID"} {print $1, $3, $4}' /etc/passwd**

**UserID UID GID**

**root 0 0**

**bin 1 1**

**daemon 2 2**

**adm 3 4**

**lp 4 7**

**sync 5 0**

**shutdown 6 0**

**halt 7 0**

**mail 8 12**

**operator 11 0**

**games 12 100**

**ftp 14 50**

**nobody 65534 65534**

**dbus 81 81**

**systemd-network 192 192**

**systemd-oom 999 999**

**systemd-resolve 193 193**

**sshd 74 74**

**rpc 32 32**

**libstoragemgmt 997 997**

**systemd-coredump 996 996**

**systemd-timesync 995 995**

**chrony 994 994**

**ec2-instance-connect 993 993**

**rpcuser 29 29**

**tcpdump 72 72**

**ec2-user 1000 1000**

**This command includes a BEGIN block to print the header before processing the rest of the file.**

**3) To align the fields in the output, you can use printf within awk to format the columns with a fixed width. Here's how you can modify the command to achieve this:**

**[root@ip-172-31-36-210 ~]# awk -F ":" 'BEGIN { printf "%-15s %-10s %-10s\n", "UserID", "UID", "GID" } { printf "%-15s %-10s %-10s\n", $1, $3, $4 }' /etc/passwd**

**UserID UID GID**

**root 0 0**

**bin 1 1**

**daemon 2 2**

**adm 3 4**

**lp 4 7**

**sync 5 0**

**shutdown 6 0**

**halt 7 0**

**mail 8 12**

**operator 11 0**

**games 12 100**

**ftp 14 50**

**nobody 65534 65534**

**dbus 81 81**

**systemd-network 192 192**

**systemd-oom 999 999**

**systemd-resolve 193 193**

**sshd 74 74**

**rpc 32 32**

**libstoragemgmt 997 997**

**systemd-coredump 996 996**

**systemd-timesync 995 995**

**chrony 994 994**

**ec2-instance-connect 993 993**

**rpcuser 29 29**

**tcpdump 72 72**

**ec2-user 1000 1000**

**[root@ip-172-31-36-210 ~]#**

### **Explanation**

* **printf is used instead of print to format the output.**
* **%-15s specifies a left-aligned string with a minimum width of 15 characters.**
* **%-10s specifies a left-aligned string with a minimum width of 10 characters.**
* **The BEGIN block prints the header with formatted columns.**
* **The { printf ... } block prints each record with formatted columns.**

**This ensures that the UserID, UID, and GID columns are properly aligned, making the output more readable. Adjust the field widths (%-15s, %-10s, etc.) as necessary to accommodate longer values if needed.**

**Homework**

1. **To kill all Java processes running on a system, you can use the ps, grep, awk, and xargs commands in combination.**

**ps -ef | grep java | grep -v grep | awk '{print $2}' | xargs kill**

### **Safety Precautions**

* **Test the Command: Before killing the processes, you can test the awk and xargs pipeline without kill to ensure it selects the correct PIDs.**

**ps -ef | grep java | grep -v grep | awk '{print $2}' | xargs echo**

**Kill with Confirmation: For a safer approach, you can use xargs with the -p option to prompt for confirmation before killing each process.**

**ps -ef | grep java | grep -v grep | awk '{print $2}' | xargs -p kill**

1. **Find and delete all files larger than 100MB in a specific directory.**
2. **Archive log files older than 30 days.**
3. **Summarize disk usage of each directory in the current directory.**