Computational Science and Scientific Computing Workshop

Elliot S. Menkah, Ph.D
Daniella N. Apeadu

National Institute for Mathematical Sciences, Ghana.
Kwame Nkrumah University of Science and Technology, Ghana.

October 8, 2025

Advanced Shell tools: AWK, SED, etc ..

Linux - Shell Tools: AWK

awk

AWK is a command-line utility that is designed for text processing and typically used as a data extraction and reporting tool.

- it is a tool for manipulating data and generating reports
- it is a filter and cn scan files line by line
- Splits each input line into fields
- Compares input line/fields to pattern and perform action on matches

Syntax & Structure

awk [options] 'selection_criteria action' input-file > output

Linux - Shell Tools: Awk Practice

1. With Ref

> : Output Redirect to new file >> : Output Redirect and append to file

Extracting the first columns of the data file the

```
~ $ awk '{ print $1}' data_output.dat > log0.txt
```

Outcome of awk command

```
#
#
#Step
0.000000
1.000000
2.000000
3.000000
4.000000
...
```

You can redirect the output of one command as the input of awk

```
~ $ cat  data_output.dat | awk '{ print $1}' > log1.txt
```

Both files, log0.txt and log1.txt should contain the same output.

Linux - Shell Tools: Awk Practice

Awk assumes a space as the field separator or delimeter.
The field separator or delimeter can be changed by using the flag -**F**
awk -F 'selection_criteria {action}' input-file > output
Eg. To get the users or username in a given linux system, we can extract it form /etc/passwd using ":"

```
~ $ cat /etc/passwd | awk -F ":" '{ print $1}' > users.txt
```

Outcome of awk command

systlog
_apt
tss
uuidd
tcpdump
...

You can also decide to print multiple columns

```
~ $ cat /etc/passwd | awk -F ":" '{ print $1 " " $3}' >
    users.txt
```

Linux - Shell Tools: Awk Practice

Field separators for both **delimeter field** and **Output Field** can be predefined
awk 'BEGIN{FS=":"; OFS="-"} selection_criteria {action}' input-file > output
Eg. To separate the output by tab spaces:
/etc/passwd using ":"

```
1 ~ $ cat /etc/passwd | awk 'BEGIN{FS=":"; OFS="\t"} { print
    $1, $3}' > users.txt
2
```

Outcome of awk command

systlog 104
_apt 105
tss 106
uuidd 107
tcpdump 108
...

Linux - Shell Tools: Awk Practice

AWK can accept regular expressions to aid filtering
awk 'BEGINFS=":"; OFS="-" selection_criteria {action}' input-file > output
Eg. To get names starting with "ic" from /etc/passwd :

```
~ $ cat /etc/passwd | awk 'BEGIN{FS=":"; OFS="\t"} /^ic/ {print $1, $3}' > users.txt
```

Outcome of awk command

ictptutor 1000
ictpuser 1001

Linux - Shell Tools: Awk Practice

AWK can take let you do some arithmetic
awk 'BEGINFS=":"; OFS="-" selection_criteria {action}' input-file ¿ output
Eg. To divide all values of column 1 by 2.0 :

```
~ $ awk '{ print $1/2.0 }' data_output.dat
```

Outcome of awk command

2
2
2
2
...
...

Linux - Shell Tools: Awk Practice

AWK can accept logicals and conditional statements
awk 'BEGINFS=":"; OFS="-" selection_criteria {action}' input-file > output
Eg. To extract running processes with bash names starting with "ic" from /etc/passwd :

```
~ $ ps -ef | awk '{ if($NF == "bash") print $0 }'
```

Outcome of awk command

ictpuser 9331 9323 0 Oct25 pts/0 00:00:01 bash

Linux - Shell Tools: SED

SED

SED (Stream Editor) is a compact programming language for parsing and transforming text.

- Line Stream matching and extraction
- input is file
- Supports regular expressions
- Supports piping

Syntax & Structure

sed [options] [SCRIPT] input-file > output

Linux - Shell Tools: SED Practice

Eg. To replace the string 'Kinetic' in data_output.dat to 'Total'

```
1 ~ $ sed s/Kinetic/Total/ data_output.dat
2
```

Outcome of awk command

This file was created Tue Oct 11 15:42:37 2022
Created by:
#Step "#Potential" "#Total"
0.000000 1635.648926 331729.281250
1.000000 -10321.562500 347803.593750
2.000000 -18997.654297 370155.781250
3.000000 -24159.796875 398618.187500
. . . . . .

End of Advanced Shell Tools & Programs