Processing and Filtering Text Data with Bash and Z Shell

SEARCHING FOR FILES AND TEXT

Sadequl Hussain
INFORMATION TECHNOLOGIST, DIGITAL CONTENT PRODUCER

@sadequlhussain

Why?

Why spend time learning from this course?

Because

It teaches about a top skill in demand: Linux

Linux User Roles

DevOps

Software

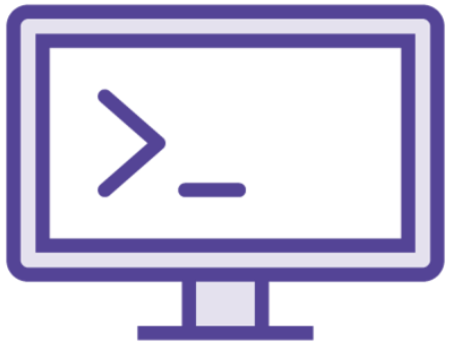
Operations

Data

Course Objective

Learning goal

How to process and filter text data in Bash and Z shell



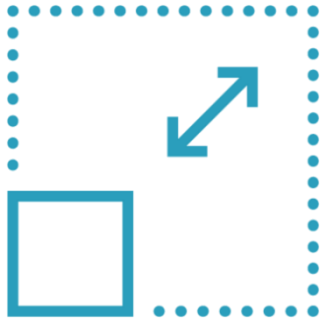
Shells used

Bash and Z shell
(Bash is more common)

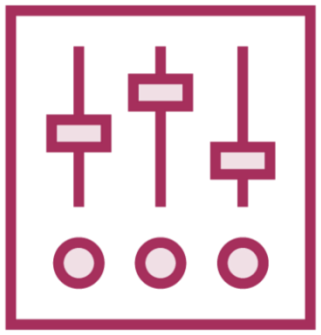
Text Processing Examples

Search logs

Compare files



Transform data



Change output

Participant Experience

Beginner's course – great for anyone starting with Linux (experienced users can brush up their skills too)

Assumes knowledge of running commands, using directories and finding help (absolute beginners can use Pluralsight Linux courses)

Any experience with scripting is a great bonus (not mandatory)

Learning Outcome

Finding files, reading files, searching for text

Basic text handling with sort, cut, transform, merge (and more)

Advanced text processing and filtering with sed and awk

Reading files and writing to files using shell scripts

Linux Shell

Bash or Z shell?
- Any one is fine

Demo systems used
- RHEL 8 (for Bash)
- Ubuntu 16 LTS (for Z)

Almost any Linux distro will have Bash, but not Z shell

Course does not cover how to install Z shell

Exercise files contain links to installation instructions

Sample Files

Data files

Included with exercise files

Unzip

Copy sample files to a folder in Linux server

Suggested location

/tmp/data/
(for consistency)

Do NOT practice or learn
in a production system!

Learning Environment

Build

Download Linux distro and install in VirtualBox (local workstation)

Buy

Virtual Private Server (VPS) from cloud providers (commercial)

Finding Files with find and locate

Linux File Search

This is the first step before we learn text processing.

To open and read a file, or to process its text, we must first find it.

Linux find Command

There are few Linux commands which can search for files

“find” is the most common one: it is a powerful command

“man” page shows “find” has a lot of options

Only important options of “find” are covered in this course

find vs. locate

“find” searches directories in real-time

“locate” searches its internal database

- Depending on size and depth of file system, “locate” can find files quicker
- Needs to keep its database updated
- Daily cron job runs “updatedb”

“locate” is not pre-installed in Red Hat or Ubuntu

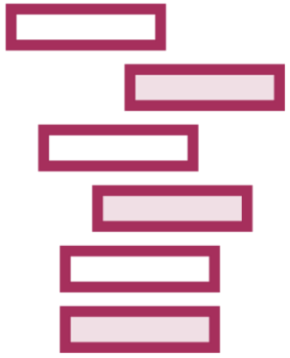
- Red Hat package name: “mlocate”

Reading Files with cat, more and less

Linux cat Command

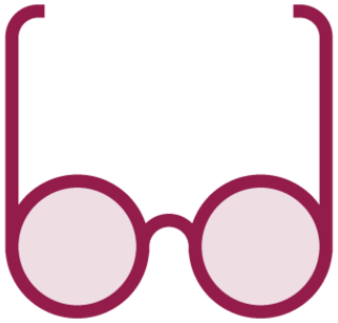
cat

Short for
“concatenate”



Write

Used to combine
multiple files into one

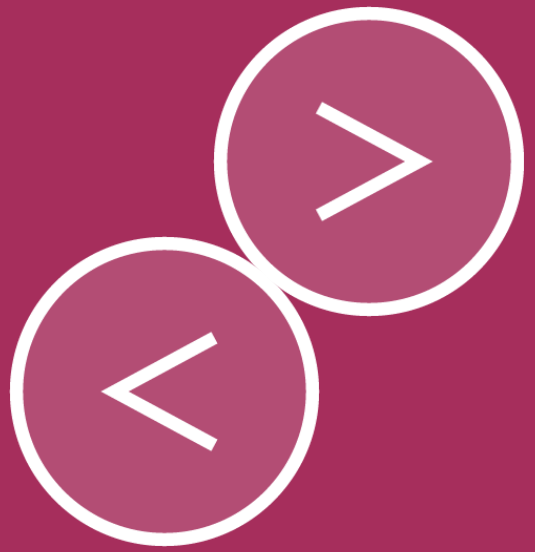


Read

Can also read files

Linux System Log

Known as “syslog”, it’s an important file where Linux records its events. Many apps write here too. Usually this is the first place to start troubleshooting.

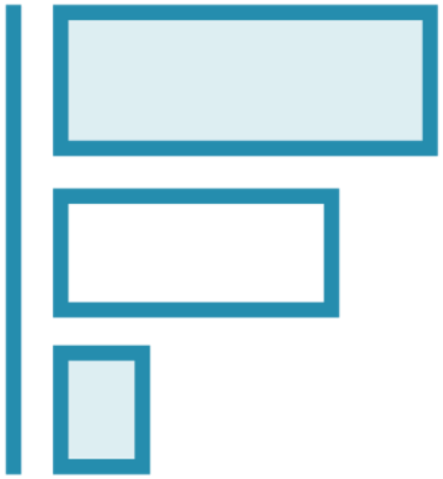


more or less

Linux more or less is used to display one screenful of text at a time. No need to use cat when using either more or less.

Reading Files with head and tail

Linux head and tail Commands



head

Reads a number of lines or bytes from beginning of file

tail

Reads a number of lines or bytes from end of file

“Busy” Files

Files like the system log can be very busy with applications always writing to it.

The “tail” command by itself may not always show the latest data.

Searching Text with grep

Linux grep Command

Global Regular Expression Print (grep)

Powerful command for searching text without opening files

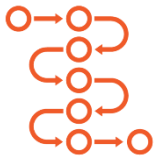
Can also search for text in a command output

Why grep?

Problem: find an event from a set of log files

Open log file in editor (large files can take long time to load)

Search through the file for each occurrence of the event



Repeat the process for every log file

Collate all results to find the answer

```
grep <option_1> <option_2> … <option_n> <search_text_patterns> <file_1> … <file_n>
```

Basic grep Syntax

Numerous options for grep

Summary

Learning recap

- Search for text files
- Read or create text files
- grep (important Linux command)

After the beginning steps

- Start processing file(s) once found

Different commands for processing text

- Next module will cover those