# CSCI 274 - Intro to Linux OS

## Week 11 - Common Pipeline Utilities

Max Gawason (maxgawason@mines.edu) [A]
Rocco Marchitto (rmarchitto@mines.edu) [B/C/D]

# Piping in Unix/Linux

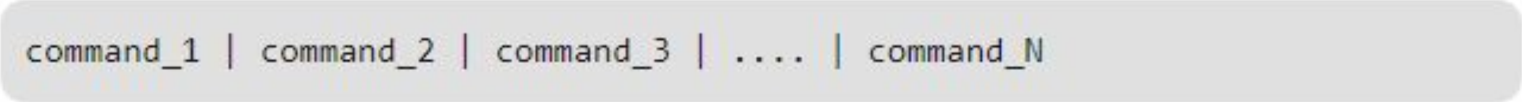
A **pipe** is a form of redirection used in Linux and other Unix-like operating systems.

It is used to send the output of one command/program/process to another command/program/process for further processing.

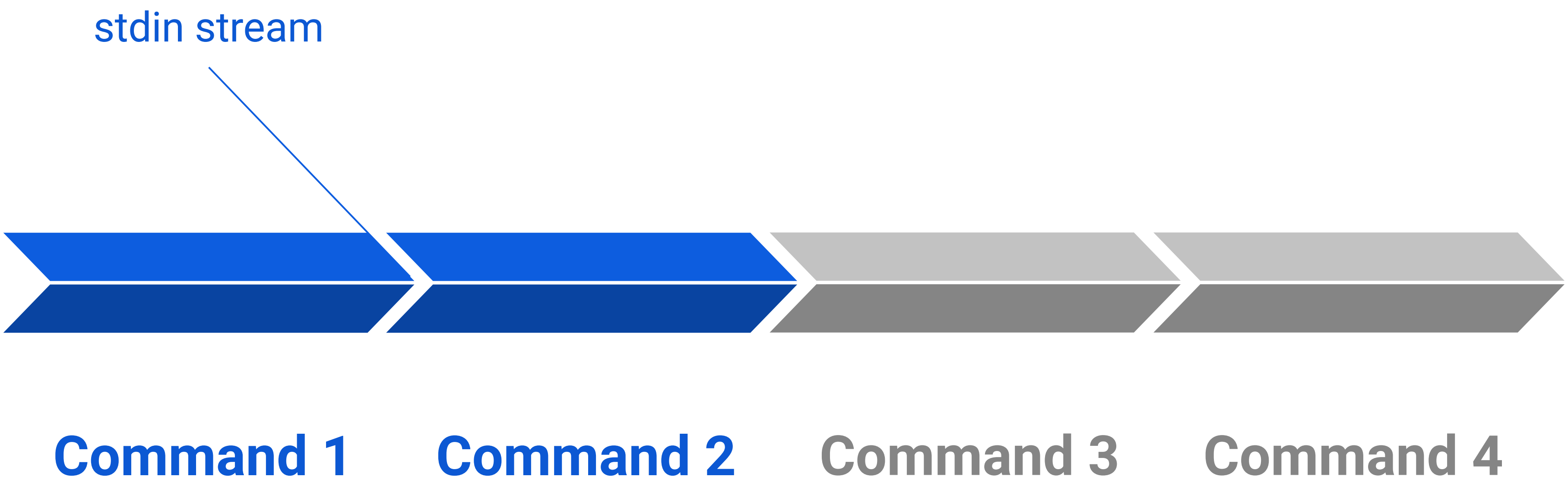
**You can make it do so by using the pipe character '|'.**

- Used to combine two or more commands
- Data flows from left to right through the pipeline

Syntax:



```
command_1 | command_2 | command_3 | .... | command_N
```



stdin stream

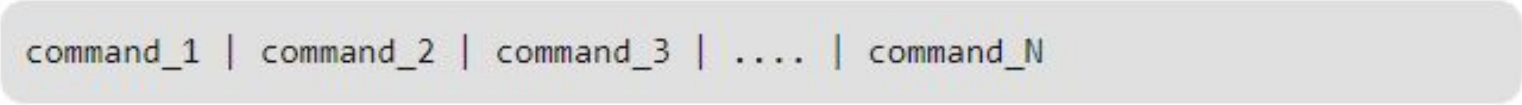
Command 1

Command 2

Command 3

Command 4

**Syntax:**



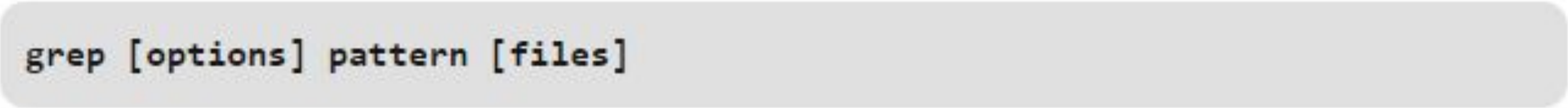
```
command_1 | command_2 | command_3 | .... | command_N
```

# Common Pipeline Utilities

**grep** (aka global regular expression print) - searches the named input file or stdin for lines containing a match to the given pattern

By default, grep matches the given string/pattern even if it found as a substring in a file

Syntax:



```
grep [options] pattern [files]
```

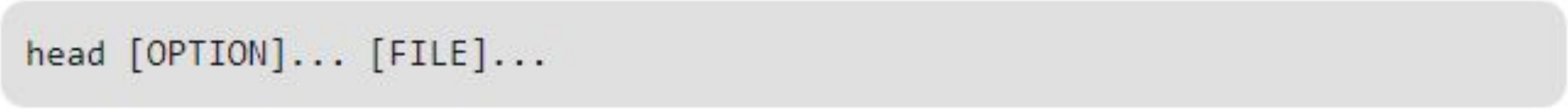
Use the --color option to highlight matches within their lines

# Common Pipeline Utilities

**head** - print the top N number of data of the given input

By default, it prints the first 10 lines of the specified stdin

Syntax:



```
head [OPTION]... [FILE]...
```

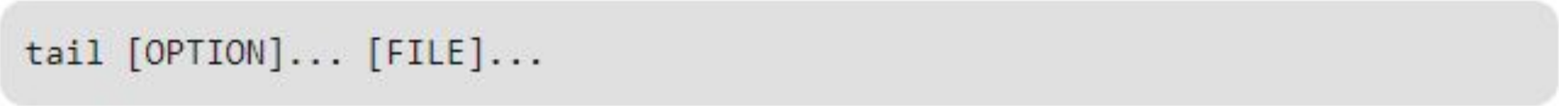
# Common Pipeline Utilities

**It is the complementary of head command**

**tail** - print the bottom N number of data of the given input

By default, it prints the first 10 lines of the specified stdin

Syntax:



```
tail [OPTION]... [FILE]...
```

Newline count as a single character

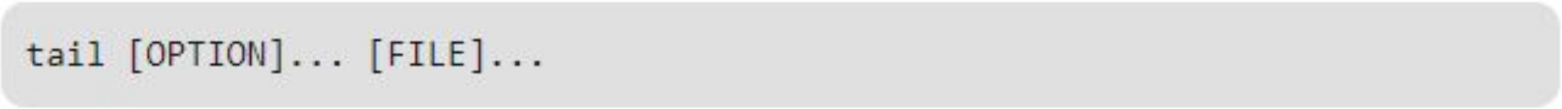
# Common Pipeline Utilities

**It is the complementary of head command**

**tail** - print the bottom N number of data of the given input

By default, it prints the first 10 lines of the specified stdin

Syntax:



```
tail [OPTION]... [FILE]...
```

Newline count as a single character

# Common Pipeline Utilities

**sort** - used to sort a file, arranging the records in a particular order. Follows these features as stated below:

1. Lines starting with a **number will appear before lines starting with a letter**.
2. Lines starting with a **letter that appears earlier** in the alphabet will appear before lines starting with a letter that appears later in the alphabet.
3. Lines starting with a **lowercase letter will appear before** lines starting with the same letter in uppercase.

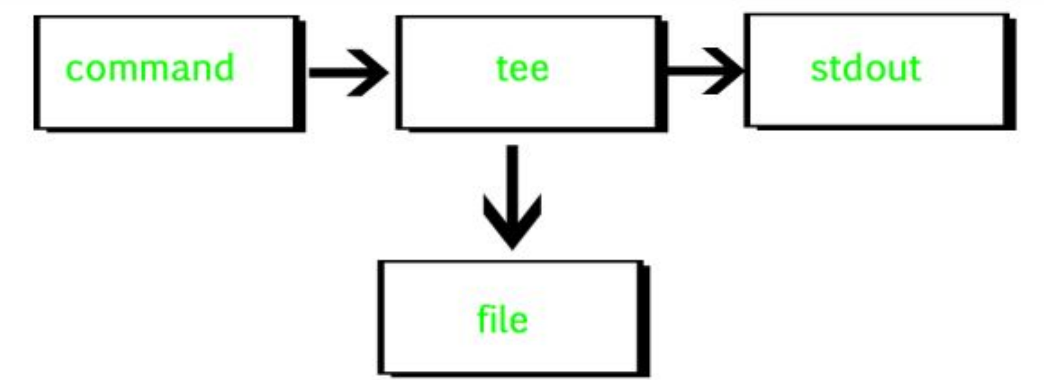
By default, the entire input is taken as sort key. Blank space is the default field separator. Sorts assuming the contents are ASCII

Note that sort **doesn't change the file**. It just prints the sorted version to standard out.

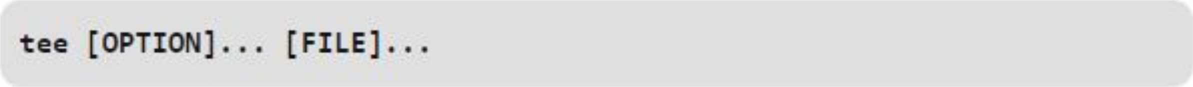
# Common Pipeline Utilities

**tee** - reads the standard input and writes it to both the standard output and one or more files

Breaks the output of a program so that it can be both **displayed and saved in a file**



Syntax:



```
tee [OPTION]... [FILE]...
```

# Common Pipeline Utilities

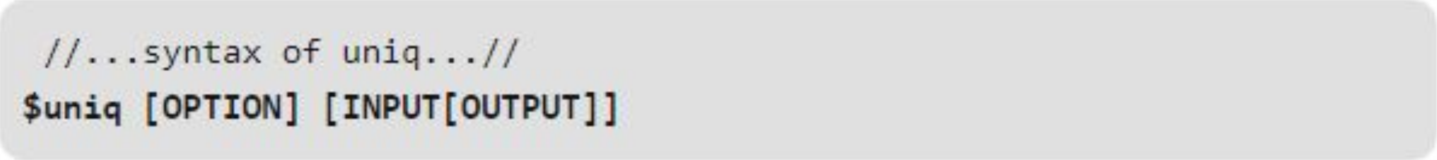
**uniq** - reports or filters out the repeated lines in a file; by default, writes to standard output

helps to detect the adjacent duplicate lines and also deletes the duplicate lines

uniq **isn’t able** to detect the duplicate lines unless they are adjacent

The content in the file must be therefore **sorted before** using uniq

Syntax:

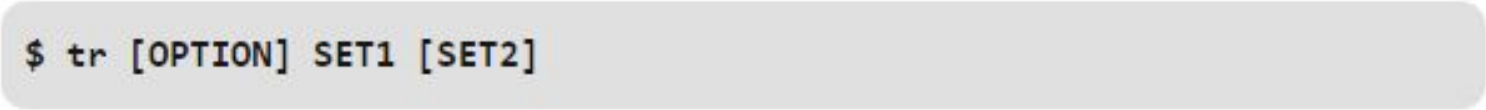


```
 //...syntax of uniq...//
$uniq [OPTION] [INPUT[OUTPUT]]
```

# Common Pipeline Utilities

**tr** (aka translate) - used for translating or deleting characters. Supports a range of transformations including uppercase to lowercase, deleting specific characters, basic find and replace, etc.

Syntax:



```
$ tr [OPTION] SET1 [SET2]
```