

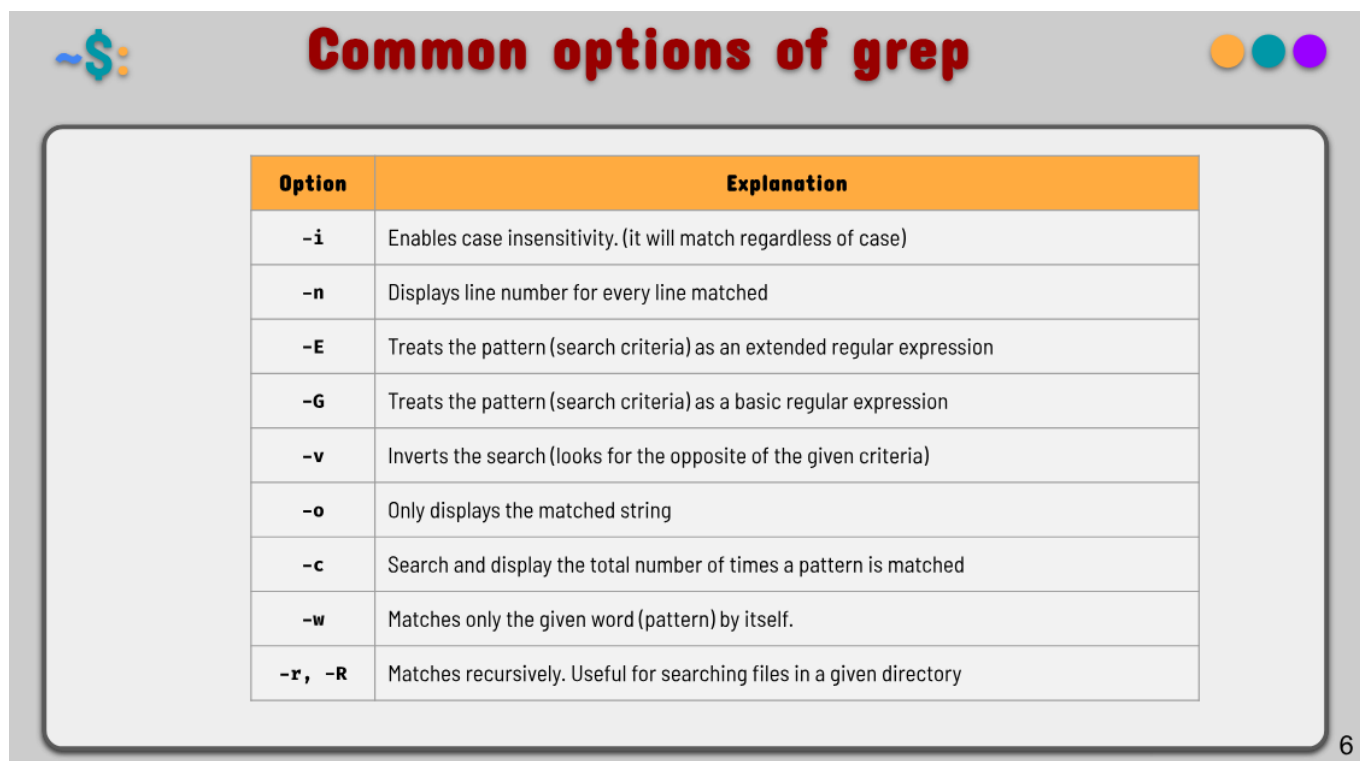
Notes 9

Handling text files part 2

The GREP command

The **grep** command is used to search text in a given file. Grep works by line basis (it matches the search criteria in a line by line basis).

- Usage:
 - `grep + option + search criteria + file(s)`
- Basic Example:
 - Search any line that contains the word 'Game' in the given file:
 - `grep 'Game' ~/Documents/GameofThrones.txt`
 - Search any line that contains the word 'thrones' regardless of the case.
 - `grep -i 'Thrones' ~/Documents/Books/GameofThrones.txt`
 - Display how many lines contain the matched string:
 - `grep -c 'Game' ~/Documents/Books/GameofThrones.txt`
- More options for `grep`:



Option	Explanation
<code>-i</code>	Enables case insensitivity. (it will match regardless of case)
<code>-n</code>	Displays line number for every line matched
<code>-E</code>	Treats the pattern (search criteria) as an extended regular expression
<code>-G</code>	Treats the pattern (search criteria) as a basic regular expression
<code>-v</code>	Inverts the search (looks for the opposite of the given criteria)
<code>-o</code>	Only displays the matched string
<code>-c</code>	Search and display the total number of times a pattern is matched
<code>-w</code>	Matches only the given word (pattern) by itself.
<code>-r, -R</code>	Matches recursively. Useful for searching files in a given directory

6

- More examples of `grep`
 - Search any line that contains the word dracula regardless of case and with number line:
 - `grep -in 'dracula' ~/Documents/Books/dracula.txt`
 - Search for all the lines that do not contain the word 'war':

- `grep -v 'war' ~/Documents/Books/war-and-peace.txt`
 - Search and display only matched string (pattern)
 - `grep -o 'pride' ~/Documents/Books/Pride&Prejudice'`
 - Display a list of users with the /bin/bash login shell:
 - `grep -i "/bin/bash" /etc/passwd`
 - Display our user's information as stored in the /etc/passwd
 - `grep -i $USER /etc/passwd`
- Expanding on `grep`

~\$:

Even more examples of grep

- Search for a given strings inside files in a given directory.
 - `grep -iR 'conf' /etc/`
- Search and display the total number of times a given word appears in a file
 - `grep -wc '/bin/bash' /etc/passwd`
- The ^ (caret) symbol matches the empty string at the beginning of a line. Search for all the lines that start with a given word.
 - `grep -ni '^dracula' ~/Documents/Books/dracula.txt`
- Search for all the lines that ends with the string "nologin"
 - `grep -n 'nologin$' /etc/passwd`

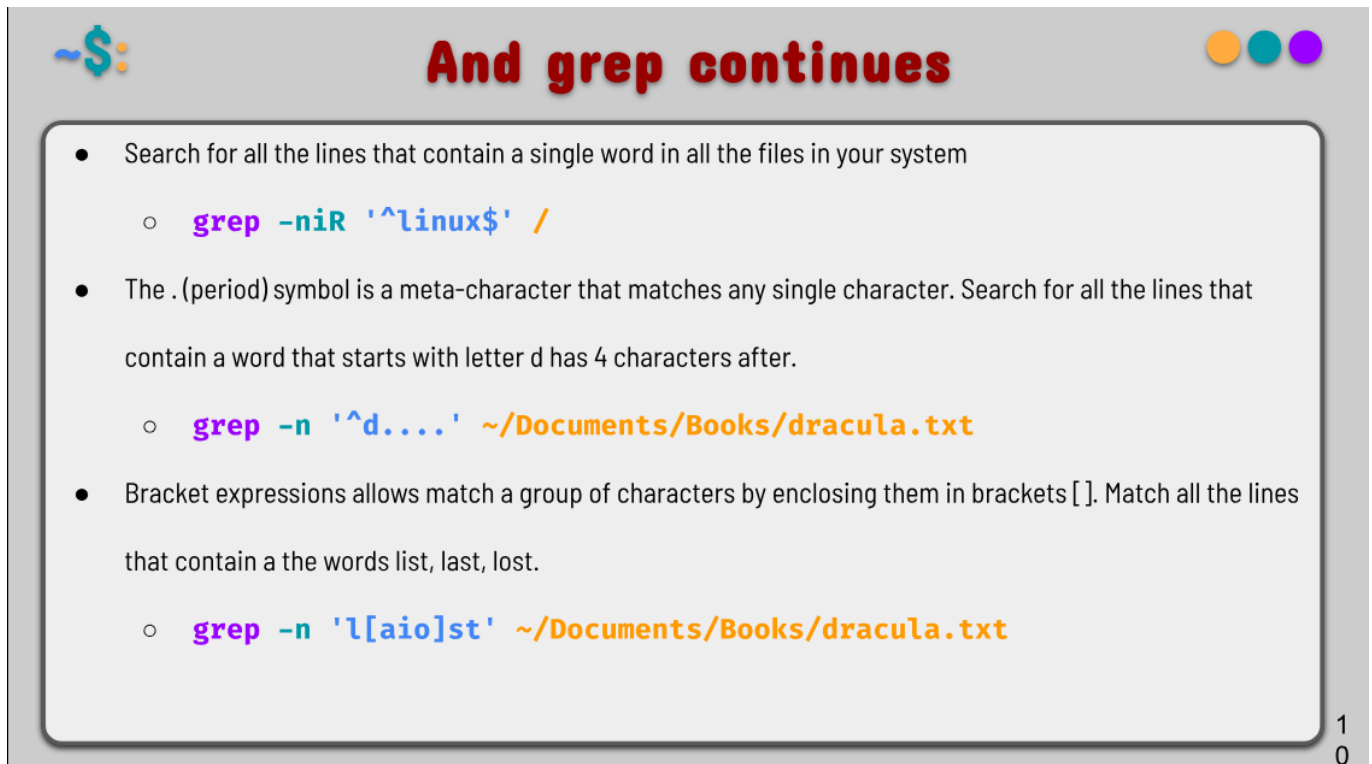
8

~\$:

Is there no end to grep?

- Search for all the lines that start with a capital letter
 - `grep -n '^[A-Z]' ~/Documents/Books/war-and-peace.txt`
- Search for more than one word per line
 - `grep -Ewn 'horror|love|scare' ~/Documents/Books/dracula.txt`
- Match only lines containing IPv4 addresses
 - `grep -E '[[[:digit:]]{1,3}\. [[[:digit:]]{1,3}\. [[[:digit:]]{1,3}\. [[[:digit:]]{1,3}]' Documentation.txt`
- Search all lines that contain a character repeated 3 times
 - `grep -E "A{3}" file.txt`
- Search all lines that contain a phone number of the format 973-111-2222
 - `grep "[[:digit:]]\{3\}[-][[:digit:]]\{3\}[-][[:digit:]]\{4\}" contacts.csv`
- Display only the options of the of any command from its man page
 - `man ls | grep "^[[:space:]]*[[[:punct:]]]"`

9



And grep continues

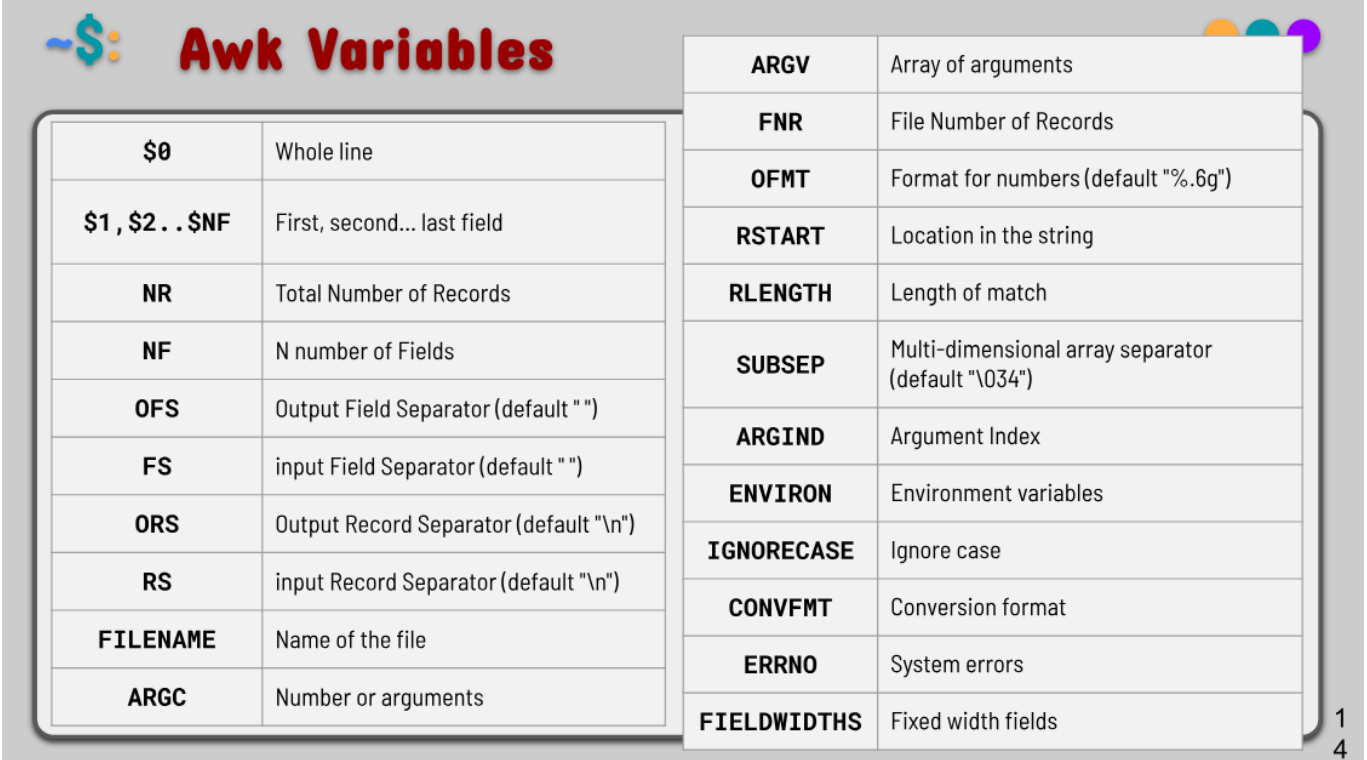
- Search for all the lines that contain a single word in all the files in your system
 - `grep -niR '^linux$' /`
- The . (period) symbol is a meta-character that matches any single character. Search for all the lines that contain a word that starts with letter d has 4 characters after.
 - `grep -n '^d....' ~/Documents/Books/dracula.txt`
- Bracket expressions allows match a group of characters by enclosing them in brackets []. Match all the lines that contain a the words list, last, lost.
 - `grep -n 'l[aio]st' ~/Documents/Books/dracula.txt`

1
0

The AWK command

The **awk** command is a scripting language used for processing and displaying text. Awk can work with a text file from standard output. There are several implementations of Awk, like: nawk, mawk, gawk, and busybox. It performs operations line by line.

- Usage:
 - `awk + options + {awk command} + file + file to save (optional)`
- Basic example:
 - Print the first column of every line of a file:
 - `awk '{print $1}' ~/Documents/Csv/cars.csv`
- Awk Variables:



The graphic features a title 'Awk Variables' in a large, bold, red font with a shadow effect. To the left of the title is a small icon of a dollar sign with a tilde (~) and a colon (:). Below the title are two tables. The first table is on the left and the second is on the right. Both tables have a light gray background and a thin black border. The first table lists variables \$0, \$1, \$2...\$NF, NR, NF, OFS, FS, ORS, RS, FILENAME, and ARGV. The second table lists variables ARGV, FNR, OFMT, RSTART, RLENGTH, SUBSEP, ARGIND, ENVIRON, IGNORECASE, CONVFM, ERRNO, and FIELDWIDTHS. The tables are presented in a clean, organized manner with a slight 3D effect.

\$0	Whole line
\$1, \$2...\$NF	First, second... last field
NR	Total Number of Records
NF	N number of Fields
OFS	Output Field Separator (default " ")
FS	input Field Separator (default " ")
ORS	Output Record Separator (default "\n")
RS	input Record Separator (default "\n")
FILENAME	Name of the file
ARGC	Number of arguments

ARGV	Array of arguments
FNR	File Number of Records
OFMT	Format for numbers (default "%.6g")
RSTART	Location in the string
RLENGTH	Length of match
SUBSEP	Multi-dimensional array separator (default "\034")
ARGIND	Argument Index
ENVIRON	Environment variables
IGNORECASE	Ignore case
CONVFMT	Conversion format
ERRNO	System errors
FIELDWIDTHS	Fixed width fields

1
4

- Examples on awk:
 - Print first field of /etc/passwd file
 - `awk -F: '{print $1}' /etc/passwd`
 - Print the last field of the /etc/passwd file
 - `awk -F: '{print $NF}' /etc/passwd`
 - Print the first and last field of the /etc/passwd
 - `awk -F: '{print $1, " = ", $NF}' /etc/passwd`
 - Print the first 3 and 3 field with line numbers
 - `awl -F: '{print NR,$1,$3}' /etc/passwd`
 - Print the first and 4th field with a different field separator
 - `awk -F: '{OFS="="}{print $1,$4}' /etc/passwd`
 - Start printing a file from a given line (exclude the first 2 lines)
 - `awk 'NR > 3 { print }' /etc/passwd`
- More examples of awk:

~\$ Even more examples of awk

- Convert the first field to upper/lower case
 - `awk -F: '{print toupper($1)}' /etc/passwd`
- Prints the length of a line(record)
 - `awk '{print length($0)}' /etc/passwd`
- Print specific fields based on a command output. For example, the size and file name
 - `ls -lhF Documents/ | awk 'BEGIN { printf "%s\t%s\n", "Size", "Name" } {print $5, "\t", $9}'`
 - BEGIN block is executed once at the start
- Print specific fields with a head of the /etc/passwd file
 - `awk -F: 'BEGIN { printf "%s\t\t%s\n", "User", "Shell" } {print $1, "\t", $7}' /etc/passwd`

1
6

- Interesting example of awk:

~\$ A more interesting example of awk

```
ls -lhd --time-style=+%D.%p | awk -v OFS='\t' 'BEGIN { printf "%s\t%s\t%s\t%s\t%s\t%s\n", "Permissions", "Links", "User", "Group", "Size", "Date Modified", "Name" } {print $1,$2,$3,$4,$5,$6,$7}'
```

Permissions	Links	User	Group	Size	Date Modified	Name
drwxr-xr-x	3	adrian	adrian	4.0K	03/02/22	./Applications
drwxr-xr-x	3	adrian	adrian	4.0K	04/03/22	./Calibre
drwxr-xr-x	2	adrian	adrian	4.0K	04/06/22	./challenge-lab6
drwxr-xr-x	2	adrian	adrian	4.0K	03/01/22	./Desktop
drwxr-xr-x	5	adrian	adrian	4.0K	04/12/22	./Documents
drwxr-xr-x	1	adrian	adrian	28	03/03/22	./dotfiles
drwxr-xr-x	10	adrian	adrian	24K	04/12/22	./Downloads
drwxr-xr-x	2	adrian	adrian	4.0K	04/03/22	./files-awk-examples
drwxr-xr-x	2	adrian	adrian	4.0K	04/02/22	./games
drwxr-xr-x	27	adrian	adrian	60K	04/10/22	./gdrive
drwxr-xr-x	9	adrian	adrian	4.0K	04/02/22	./gems
drwxr-xr-x	2	adrian	adrian	4.0K	04/12/22	./json-files
drwxr-xr-x	7	adrian	adrian	4.0K	04/02/22	./labs
-rw-rw-r--	1	adrian	adrian	4.1M	04/02/22	./labs.zip
drwxr-xr-x	2	adrian	adrian	4.0K	04/11/22	./lab6
drwxr-xr-x	2	adrian	adrian	4.0K	04/02/22	./movies
drwxr-xr-x	3	adrian	adrian	4.0K	04/04/22	./music
drwxr-xr-x	3	adrian	adrian	4.0K	03/05/22	./pictures
drwxr-xr-x	2	adrian	adrian	4.0K	03/01/22	./public
-rw-rw-r--	1	adrian	adrian	0	04/04/22	./song.mp3
drwxr-xr-x	3	adrian	adrian	4.0K	03/02/22	./steam
drwxr-xr-x	2	adrian	adrian	4.0K	03/01/22	./templates
-rw-rw-r--	1	adrian	adrian	108	04/09/22	./todo.md
drwxr-xr-x	1	adrian	adrian	511	04/09/22	./uris.sh
drwxr-xr-x	2	adrian	adrian	4.0K	04/07/22	./videos
drwxr-xr-x	7	adrian	adrian	4.0K	03/21/22	./VirtualBox

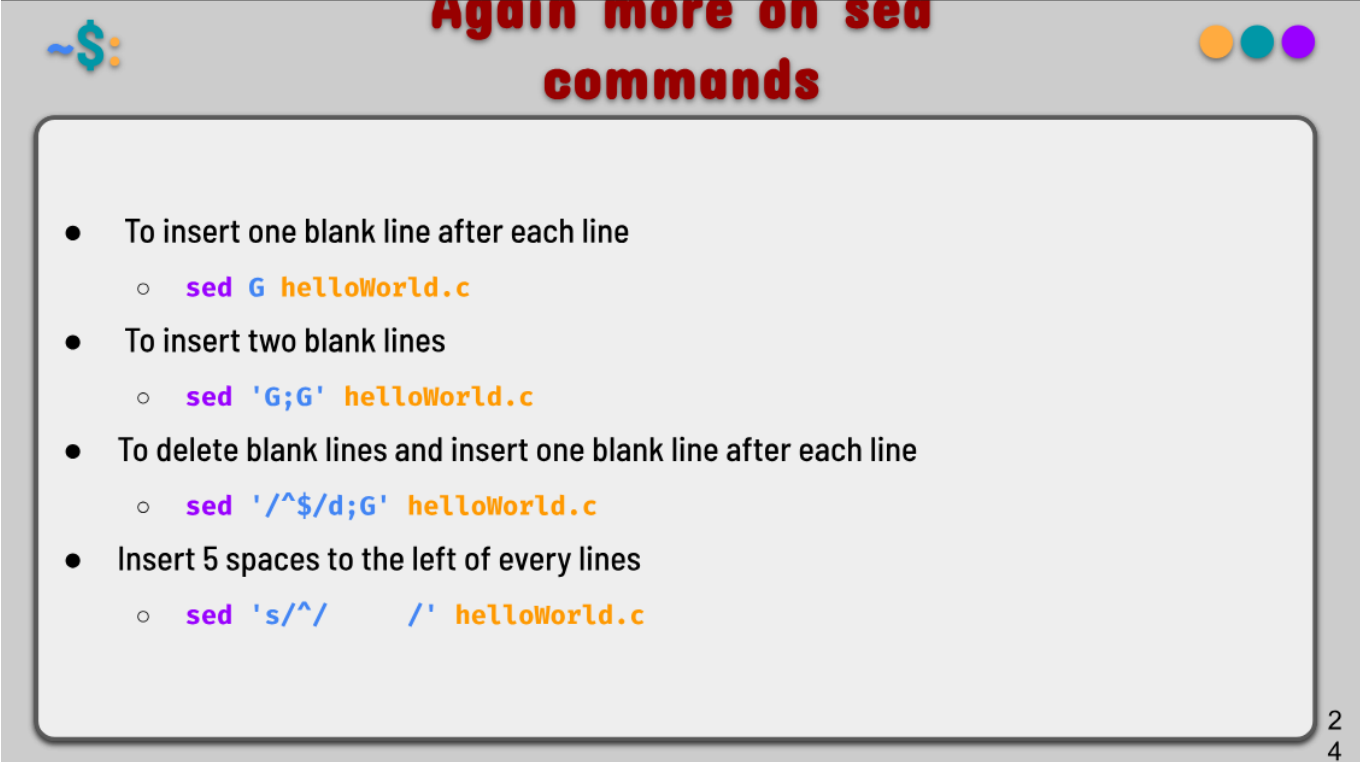
1
7

The Sed command

The **sed** command is a stream editor that perform operations on files and standard output. For instance, it can search, find and replace, insert, and deletion. By using SED you can edit files without opening them.

- Usage:

- `sed options + sed script + file`
- Basic Example:
 - Replacing a string in given file globally (replace false or true)
 - `sed 's/false/true/g' ~/Documents/sample_files/Json/joke.json`
- More examples:
- Replacing only the fourth occurrence per line in a file:
 - `sed 's/false/true/4' joke.json`
- Replacing from the given number occurrence to the rest occurrences in a file. Start at the second time the word appears and continue to till the end of the file.
 - `sed 's/false/true/' joke.json`
- Replacing string on a specific line number
 - `sed '3 s/false/true/' joke.json`
- Replacing string on a range of lines
 - `sed '1,3 s/false/true/' joke.json`
- Even more examples on sed:
 - To delete a particular line (line 5)
 - `sed '3d' ~/Documents/sample_files/Code/helloworld.py`
 - To delete the last line
 - `sed '$d' ~/Documents/sample_files/Code/helloworld.py`
 - To delete line from range x to y
 - `sed '2,4d' ~/Documents/sample_files/Code/helloworld.py`
 - To delete from a given number to the last line
 - `sed '3,$d' ~/Documents/sample_files/Code/helloworld.py`
 - To delete pattern matching line in a file
 - `sed '/fav/d' ~/Documents/sample_files/Code/helloworld.py`
- Elaborating on Sed:



Again more on sed commands

- To insert one blank line after each line
 - `sed G helloWorld.c`
- To insert two blank lines
 - `sed 'G;G' helloWorld.c`
- To delete blank lines and insert one blank line after each line
 - `sed '/^$/d;G' helloWorld.c`
- Insert 5 spaces to the left of every lines
 - `sed 's/^/ /' helloWorld.c`

2
4

How to use | the pipe. How to redirect standard output?

The **pipe (|)** allows you to redirect the standard output of a command to the standard input of another.

- Usage:
 - `command_1 | command_2 | command_3 | | command_N`
- Basic examples:
 - Use the grep to look for a string in a particular man page
 - `man ls | grep "human-readable"`
 - Display only the options of the of any command from its man page
 - `man ls | grep "[[:space:]]*[[:punct:]]"`

Other examples of |:

- Display only the ip addresses from the output of the ip command:
 - `ip addr | grep -Eo '[[[:digit:]]{1,3}\. [[[:digit:]]{1,3}\. [[[:digit:]]{1,3}\. [[[:digit:]]{1,3}]'`
- Display only the 2nd line in a file:
 - `head -2 file.lst | tail -1`
- Parse a file with grep and replace a string in the output:
 - `grep -i "honda" cars.csv | sed 's/honda/tesla/g'`

How to save the standard output?

- Usage:
 - `Command output + > + file`
 - Basic Examples:

- Save the output of a command to a file
 - `ls -lA ~ > all-files-in-home.txt`
- Save the error generated by a command to a file
 - `ls -lA downloads/ 2> error-of-ls`
- Save the error to a file and the success to another
 - `ls -lA downloads/ Pictures > success.txt 2> error.txt`
- Save the error and success to the same file
 - `ls -lA downloads/ Pictures &> alloutput.txt`
- Do not display errors. Send errors to the black hole
 - `ls -lA downloads/ 2> /dev/null`

How to append the output of a file

Append means to add more to a file instead of overwriting its content. When we use `>` on a file that already exist and contains data, we overwrite whatever is already inside the file. For example:

- `ls -la > allmyfiles.lst`

In this example, if the file `allmyfiles.lst` had any data prior executing the command, that data will be overwritten by the output of `ls-la`. But, what happens when we don't want to keep the data? Then we use:

- `ls -la >> allmyfiles.lst`

Will add the output of `ls-la` to the end of the file `allmyfiles.lst`