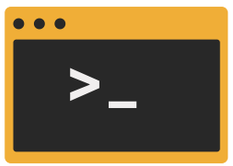
KDnuggets | Visit KDnuggets.com for more cheatsheets and additional learning resources.

Linux CheatSheet



The Linux commands in data science are used for automating tasks, building pipelines, accessing file systems, and enhancing development operations.

ls

It is used to display files and folders in the current directory.

$ ls

pwd

Print the full path of the current directory using pwd.

$ pwd

cd

The cd stands for change directory. The command requires a new directory path.

$ cd /new/directory/path

wget

Download a file or multiple files from the internet by using wget and URL.

$ wget /URL/file.csv

cat

It is used for viewing, creating, and concatenating files. Add a file location to view all the content of the file.

$ cat file.csv

wc

Use wc to get information about word count, character count, and the number of lines.

$ wc file.csv

head

The head is used to display the top n number of lines in a file. The command below is showing the top 5 lines.

$ head -n 5 file.csv

find

Use the find command to find files and folders. The command below will show all the files with the ".dvc" extension.

$ find . -name "*.dvc" -type f

grep

It is used to find data within the file. You can provide a text pattern and it will display all the lines containing that pattern.

$ grep -i "vir" file.csv

zip

Zip a single file or multiple files. It compresses the file size and file package utility. zip requires a zip file name and a list of files that you want to condense.

$ zip data.zip file1.txt file2.txt

unzip

Similarly, unzip is used to uncompress and extract the files from a zip file.

$ unzip data.zip

cp

cp is a copy command. It requires file name and destination directory path. Similarly, you can use mv to move files and rm to remove the files.

$ cp a.txt work

mkdir

Use mkdir to create a new file directory. It requires a directory name or directory path. Similarly, you can use rmdir to delete the directory.

$ mkdir /model

man

Learn about any Linux commands or tools by using the man command and the tool name. The man stands for manual.

$ man echo

diff

Just like `git diff`, you can display line-by-line differences between two files.

$ diff app1.py app2.py

alias

Create shortcuts for your most frequently used command by creating aliases. It requires a shortcut name and command as a string.

You will print "I love KDnuggets" by typing love in the terminal.

$ alias love="echo 'I love KDnuggets'"

Learn more on kdnuggets.com

Abid Ali Awan | 2022