Linux command line

Priit Adler, Hedi Peterson
15.05.2015

European Union
Regional Development Fund   Investing in your future

elixir
ESTONIA

Course practicalities

- Break after 1h40 minutes (13.40-14.00)

- Mixed theoretical and hands-on parts

- Feel free to leave the room when needed

- Provide feedback before going home

- All materials available at bit.ly/lcl15052015slides

- Online discussion at bit.ly/discussion15052015

Reasons to use Linux
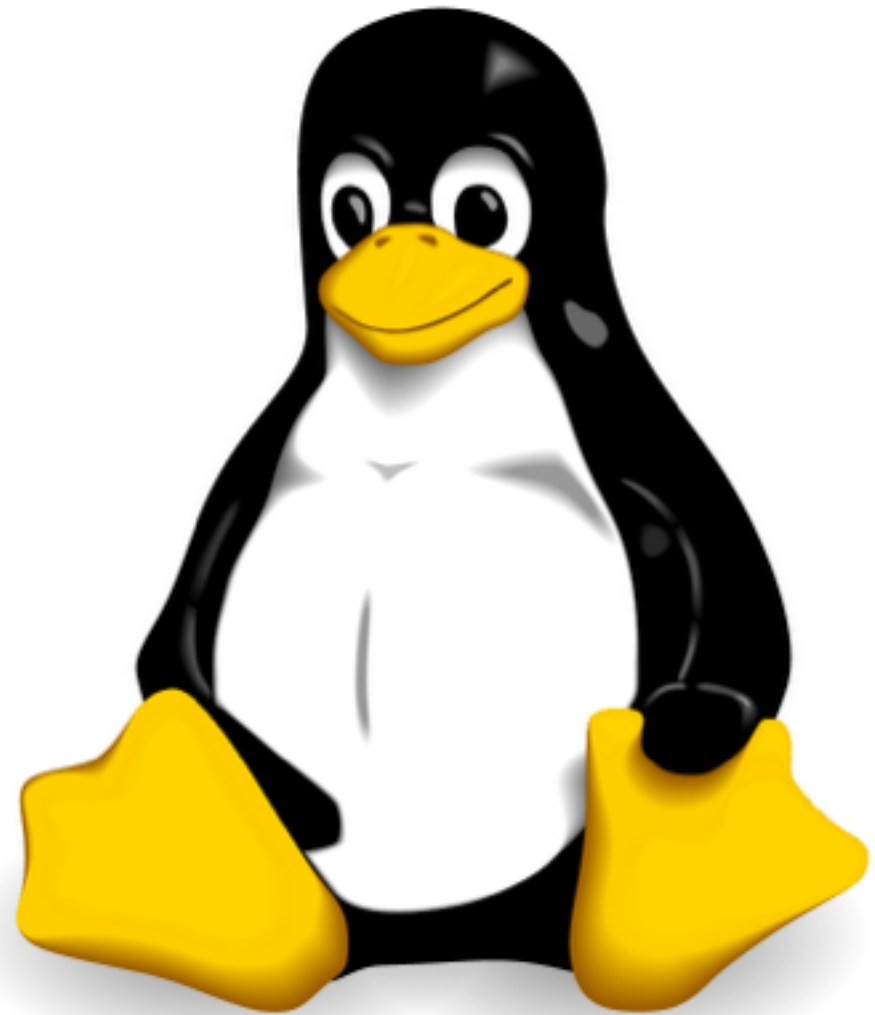
You will learn the following:

- How to work with files

- How to work with file content

- How to control jobs

- And few advanced tricks

Most important command

- man

man(1)

NAME
       man - format and display the on-line manual pages

SYNOPSIS
       man [-acdfFhkKtwW] [--path] [-m system] [-p string] [-C config_file]
       [-M pathlist] [-P pager] [-B browser] [-H htmlpager] [-S section_list]
       [section] name ...

DESCRIPTION
       man formats and displays the on-line manual pages. If you specify sec-
       tion, man only looks in that section of the manual. name is normally
       the name of the manual page, which is typically the name of a command,
       function, or file. However, if name contains a slash (/) then man
       interprets it as a file specification, so that you can do man ./foo.5
       or even man /cd/foo/bar.1.gz.

       See below for a description of where man looks for the manual page
       files.

Getting out of things

- type 'q', if that does not work

- 'ctrl-c', if that does not work

- log in with new session, read `man` about the program you are stuck in.

- `kill` the process from another session.

Our server for the course

elixir.cs.ut.ee

user and password given to You

command line ssh or putty to log in

Working with files and directories

Knowing where you are

- **pwd** - gives working directory

```
6un:~ peterson$ pwd
/Users/peterson
```

- **cd** - move to desired directory

```
6un:~ peterson$ cd ~/linux/
6un:linux peterson$
```

```
6un:linux peterson$ cd ../
6un:~ peterson$
```

Relative paths

```
6un:~ peterson$ cd /Users/peterson/linux/
6un:linux peterson$
```

Absolute path

File listing

- **ls** - lists directory contents

```
6un:linux peterson$ ls
HG-U133_Plus2.full.flat biogrid_tab.txt
```

- **ls -l** - lists directory contents with more info

```
6un:linux peterson$ ls -l
-rw-r--r--@ 1 peterson  staff  61632026 14 mai   12:17 HG-U133_Plus2.full.flat

-rw-r--r--@ 1 peterson  staff  16850959 14 mai   12:17 biogrid_tab.txt
```

File listing

- **ls** - lists directory contents

```
6un:linux peterson$ ls
HG-U133_Plus2.full.flat biogrid_tab.txt
```

- **ls -l** - lists directory contents with more info

```
6un:linux peterson$ ls -l
-rw-r--r--@ 1 peterson  staff  61632026 14 mai   12:17 HG-U133_Plus2.full.flat

-rw-r--r--@ 1 peterson  staff  16850959 14 mai   12:17 biogrid_tab.txt
```

Access rights

File listing

- **ls** - lists directory contents

```
6un:linux peterson$ ls
HG-U133_Plus2.full.flat biogrid_tab.txt
```

- **ls -l** - lists directory contents with more info

```
6un:linux peterson$ ls -l
-rw-r--r--@ 1 peterson  staff  61632026 14 mai   12:17 HG-U133_Plus2.full.flat

-rw-r--r--@ 1 peterson  staff  16850959 14 mai   12:17 biogrid_tab.txt
```

Owner

File listing

- **ls** - lists directory contents

```
6un:linux peterson$ ls
HG-U133_Plus2.full.flat biogrid_tab.txt
```

- **ls -l** - lists directory contents with more info

```
6un:linux peterson$ ls -l
-rw-r--r--@ 1 peterson  staff  61632026 14 mai   12:17 HG-U133_Plus2.full.flat

-rw-r--r--@ 1 peterson  staff  16850959 14 mai   12:17 biogrid_tab.txt
```

Group

File listing

- **ls** - lists directory contents

```
6un:linux peterson$ ls
HG-U133_Plus2.full.flat biogrid_tab.txt
```

- **ls -l** - lists directory contents with more info

```
6un:linux peterson$ ls -l
-rw-r--r--@ 1 peterson  staff   61632026 14 mai   12:17 HG-U133_Plus2.full.flat

-rw-r--r--@ 1 peterson  staff   16850959 14 mai   12:17 biogrid_tab.txt
```

File size in bytes

File listing

- **ls** - lists directory contents

```
6un:linux peterson$ ls
HG-U133_Plus2.full.flat biogrid_tab.txt
```

- **ls -l** - lists directory contents with more info

```
6un:linux peterson$ ls -l
-rw-r--r--@ 1 peterson  staff  61632026 14 mai    12:17 HG-U133_Plus2.full.flat

-rw-r--r--@ 1 peterson  staff  16850959 14 mai    12:17 biogrid_tab.txt
```

File modification time

File listing

- **ls** - lists directory contents

```
6un:linux peterson$ ls
HG-U133_Plus2.full.flat biogrid_tab.txt
```

- **ls -l** - lists directory contents with more info

```
6un:linux peterson$ ls -l
-rw-r--r--@ 1 peterson   staff   61632026 14 mai    12:17 HG-U133_Plus2.full.flat

-rw-r--r--@ 1 peterson   staff   16850959 14 mai    12:17 biogrid_tab.txt
```

File name

File listing

- **ls** - lists directory contents

```
6un:linux peterson$ ls
HG-U133_Plus2.full.flat biogrid_tab.txt
```

- **ls -lh** - lists directory contents with more info

```
6un:linux peterson$ ls -lh
-rw-r--r--@ 1 peterson  staff   59M 14 mai   12:17 HG-U133_Plus2.full.flat

-rw-r--r--@ 1 peterson  staff   16M 14 mai   12:17 biogrid_tab.txt
```

File size for humans

File listing

- **ls** - lists directory contents

```
6un:linux peterson$ ls
HG-U133_Plus2.full.flat biogrid_tab.txt
```

- **ls -a** - show all files in the directory

```
.                    .readme                 biogrid_tab.txt

..            HG-U133_Plus2.full.flat
```

Moving files around

- cp - keeps the original file

- mv - moves the original file

Creating directories

- mkdir - create a directory

```
6un:linux peterson$ mkdir firstdir
6un:linux peterson$ ls
HG-U133_Plus2.full.flat biogrid_tab.txt
firstdir
```

Access rights

- chmod - change user rights for files or folders

```
-bash-4.2$ ls -l packedfile.tar.gz
-rw-r--r--. 1 student1 student1 16860560 11. mai   19:02 packedfile.tar.gz
-bash-4.2$ chmod 755 packedfile.tar.gz
-bash-4.2$ ls -l packedfile.tar.gz
-rwxr-xr-x. 1 student1 student1 16860560 11. mai   19:02 packedfile.tar.gz
```

Creating and removing

- rm - remove / <u>no UNDO in unix</u>

```
6un:firstdir peterson$ ls
biogrid_1000.tab
6un:firstdir peterson$ rm biogrid_1000.tab
6un:firstdir peterson$ ls
6un:firstdir peterson$
```

Disk usage

- **du -h**

```
6un:linux peterson$ du -h
  0B    ./firstdir
 75M    .
```

- **df -h**

```
-bash-4.2$ df -h
```

Failisüsteem	Maht	Kasut	Vaba	Kas%	Haagitud
/dev/mapper/centos-root	47G	2,8G	44G	6%	/
devtmpfs	7,6G	0	7,6G	0%	/dev
tmpfs	7,6G	0	7,6G	0%	/dev/shm
tmpfs	7,6G	8,5M	7,6G	1%	/run
tmpfs	7,6G	0	7,6G	0%	/sys/fs/cgroup
/dev/vda1	497M	227M	271M	46%	/boot

Conserving space packing

```
6un:linux peterson$ du -h
  0B      ./firstdir
 75M      .

6un:linux peterson$ ls -lh HG-U133_Plus2.full.flat
-rw-r--r--@ 1 peterson  staff    59M 14 mai   12:17 HG-
U133_Plus2.full.flat

6un:linux peterson$ tar -zcvf packedfile.tar.gz  HG-
U133_Plus2.full.flat
a HG-U133_Plus2.full.flat

6un:linux peterson$ ls -lh packedfile.tar.gz
-rw-r--r--  1 peterson  staff    16M 14 mai   18:03
packedfile.tar.gz
```

Conserving space unpacking

```
6un:linux peterson$ tar -zxvf packedfile.tar.gz
x ./._HG-U133_Plus2.full.flat
x HG-U133_Plus2.full.flat
```

Working with file content

Creating content

- -bash-4.2$ **emacs** test.txt

Save file - CTRL+x, CTRL+x
Quit file - CTRL+x,CTRL+x
Write file - CTRL+x,CTRL+w
Get to line end - CTRL+e
Get to line start - CTLR+a
Remove full line - CTRL+k
Remove one character -CTRL+d
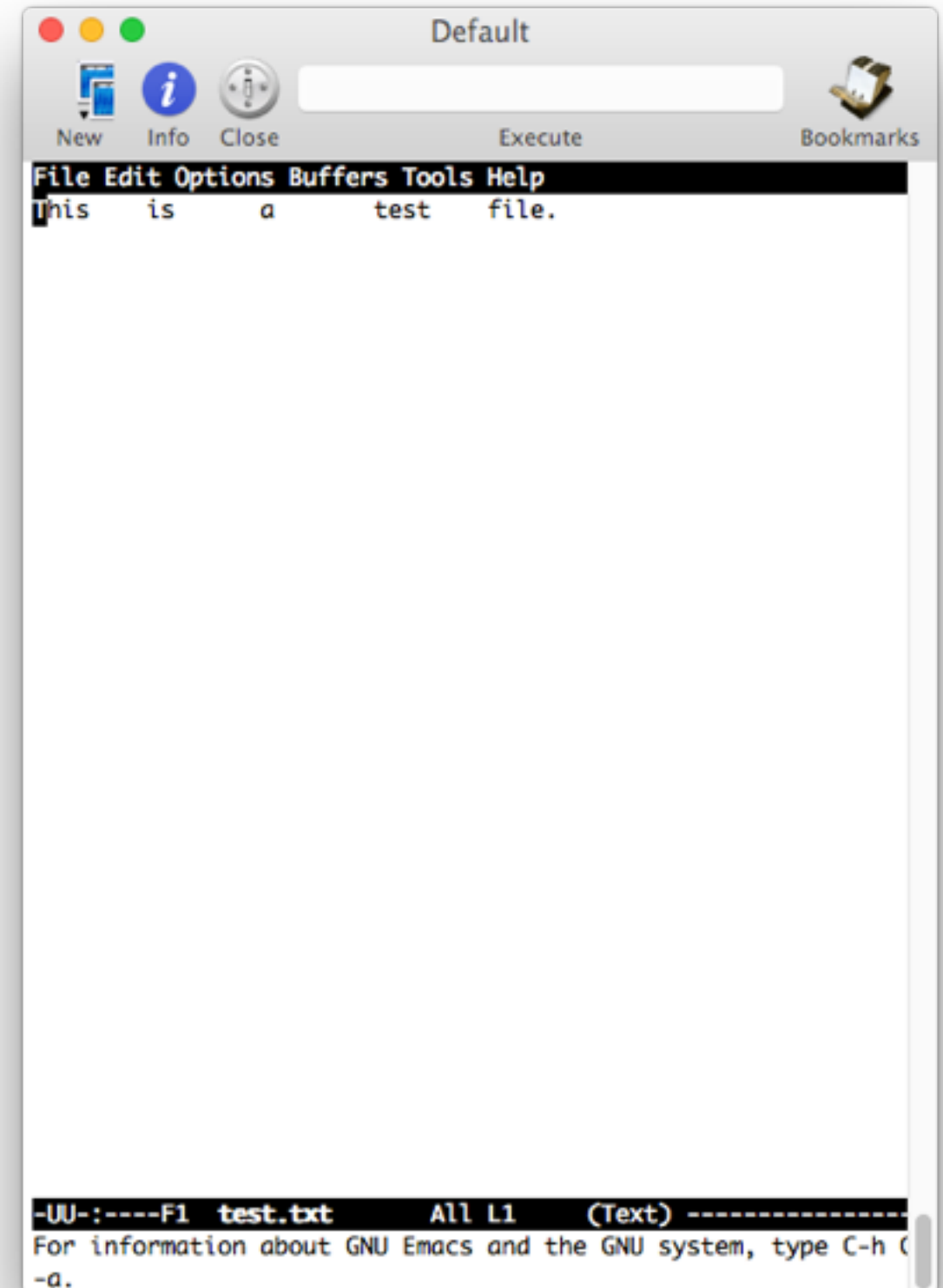
Creating content

- -bash-4.2$ **vim** test2.txt

Edit file - :e
Save file - :w
Quit file - :q
Save and quit file - :x
Write file - :w newfile
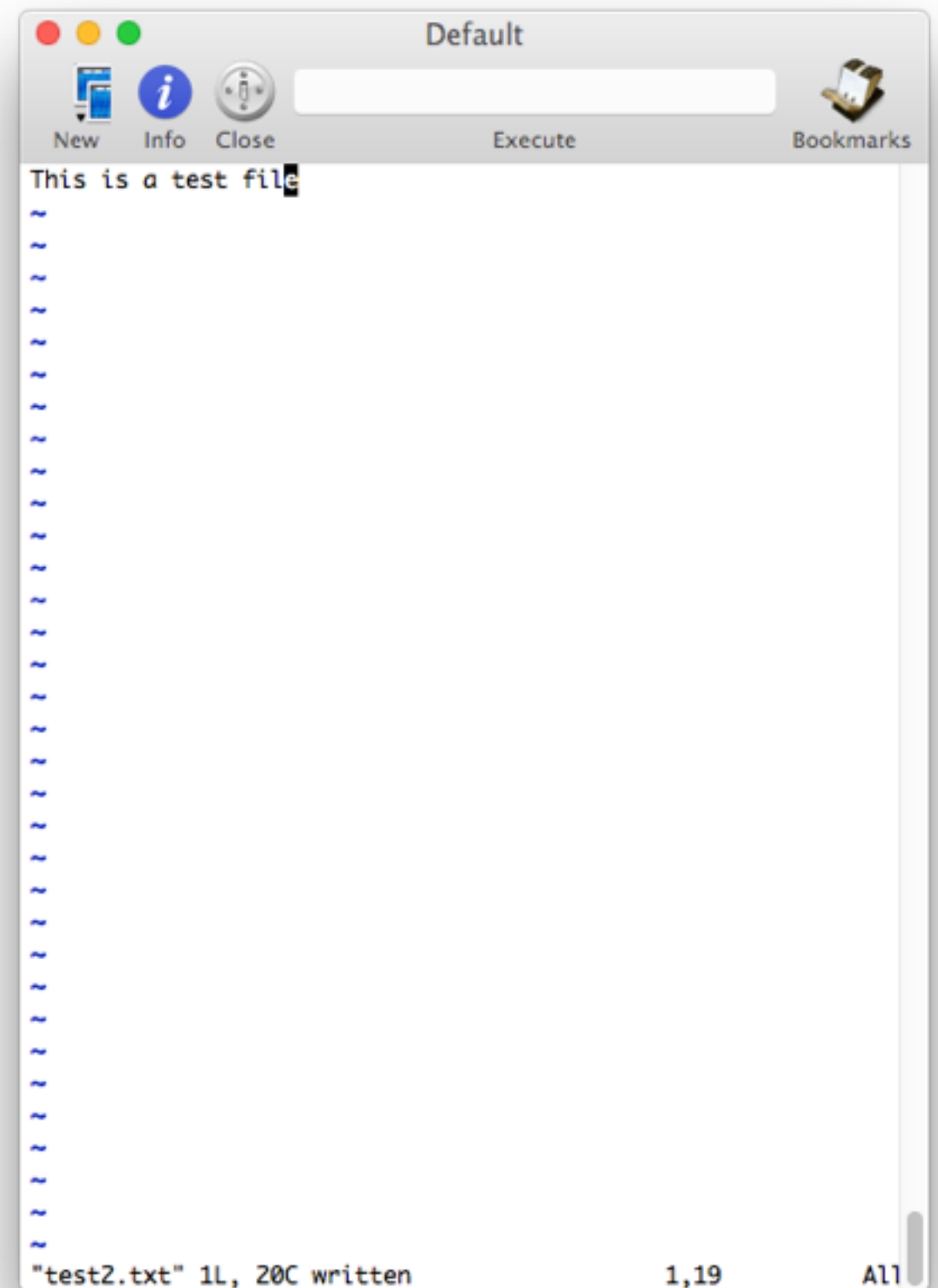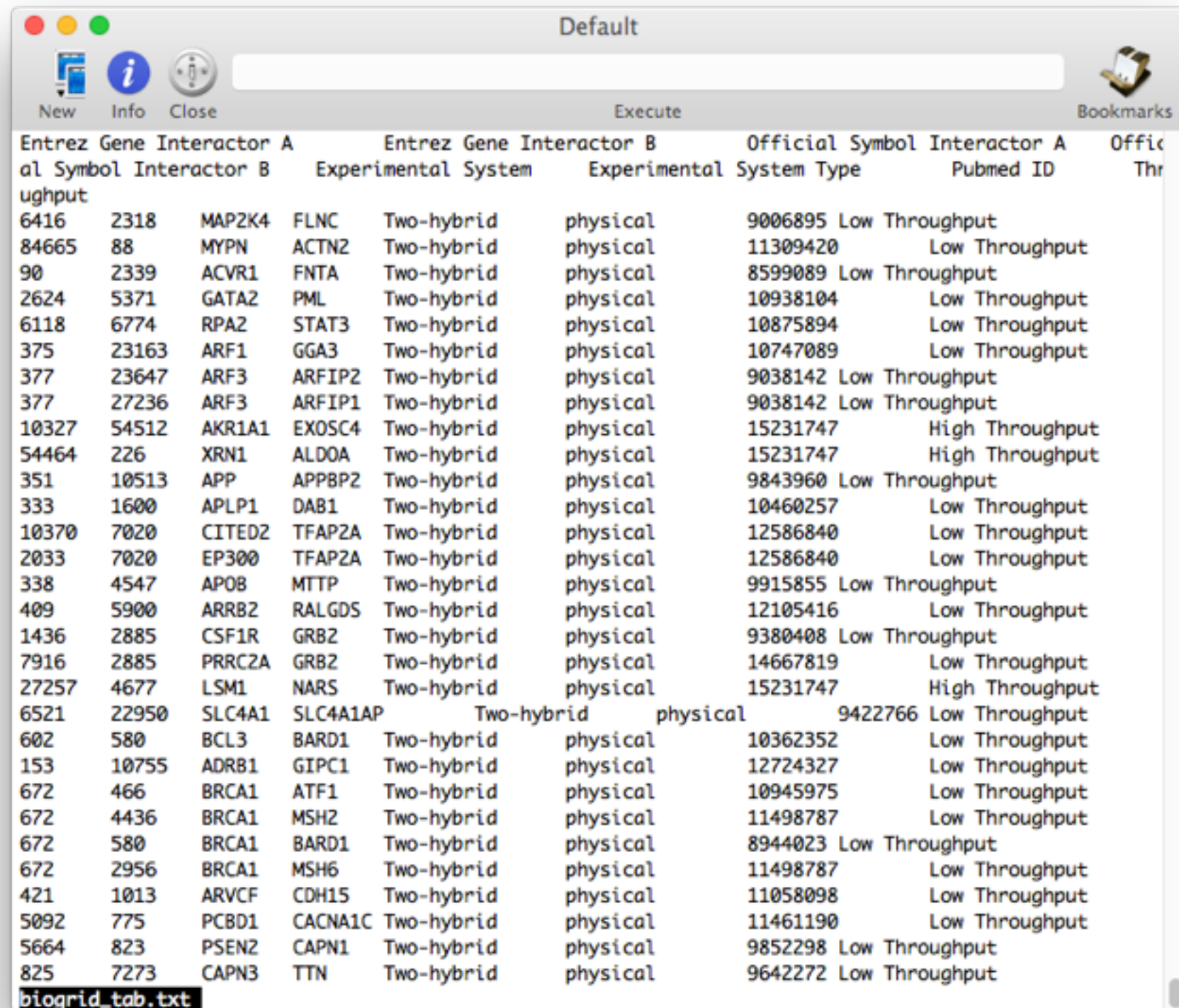
Getting what you want
from files

- ## less - look into file

Getting what you want
from files

- WC - count characters and lines for a file

```
-bash-4.2$ wc biogrid_tab.txt
  220019   2150719 16850959 biogrid_tab.txt
```

Getting what you want
from files

- **head** - get top rows from file

- **tail** - get last rows from file

```
-bash-4.2$ head -n 3 biogrid_tab.txt
Entrez Gene Interactor A        Entrez Gene Interactor B        Official Symbol Interactor A    Official Symbol Inter
ctor B  Experimental System     Experimental System Type        Pubmed ID       Throughput
6416    2318    MAP2K4  FLNC    Two-hybrid      physical        9006895 Low Throughput
84665   88      MYPN    ACTN2   Two-hybrid      physical        11309420        Low Throughput
-bash-4.2$ tail -n 3 biogrid_tab.txt
10537   728689  UBD     EIF3CL  Affinity Capture-MS     physical        23862649        High Throughput
10537   645899  UBD     RPS28P4 Affinity Capture-MS     physical        23862649        High Throughput
10537   646195  UBD     RPS28P7 Affinity Capture-MS     physical        23862649        High Throughput-bash-4.2$
```

Getting what you want
from files

- ### cat - concatenate files together

```
-bash-4.2$ cat test.txt test2.txt
This     is      a        test    file.
This is a test file number two
```

Getting what you want
from files

- cut - extract sections of lines

```
-bash-4.2$ cut -f3,4,6,8 biogrid_tab.txt|tail -n+2 > pairs.tab
-bash-4.2$ head pairs.tab
MAP2K4    FLNC      physical      Low Throughput
MYPN      ACTN2     physical      Low Throughput
ACVR1     FNTA      physical      Low Throughput
GATA2     PML       physical      Low Throughput
RPA2      STAT3     physical      Low Throughput
ARF1      GGA3      physical      Low Throughput
ARF3      ARFIP2    physical      Low Throughput
ARF3      ARFIP1    physical      Low Throughput
AKR1A1    EXOSC4    physical      High Throughput
XRN1      ALDOA     physical      High Throughput
```

Getting what you want
from files

- **uniq** - find uniq and repeating lines in file

```
-bash-4.2$ wc pairs.tab
 220019 1100554 8192802 pairs.tab
-bash-4.2$ uniq pairs.tab |wc
 214381 1072364 7986583


-bash-4.2$ uniq -cd pairs.tab |head -n5
      2 NR5A1      NRIP1      physical               Low Throughput
      2 TNFRSF1A              PSMD2    physical               Low Throughput
      2 VHL        FLNA       physical       Low Throughput
      2 HTR2C      MPDZ       physical       Low Throughput
      2 AR         RNF4       physical       Low Throughput
```

Getting what you want
from files

- **sort** - sort lines of text files
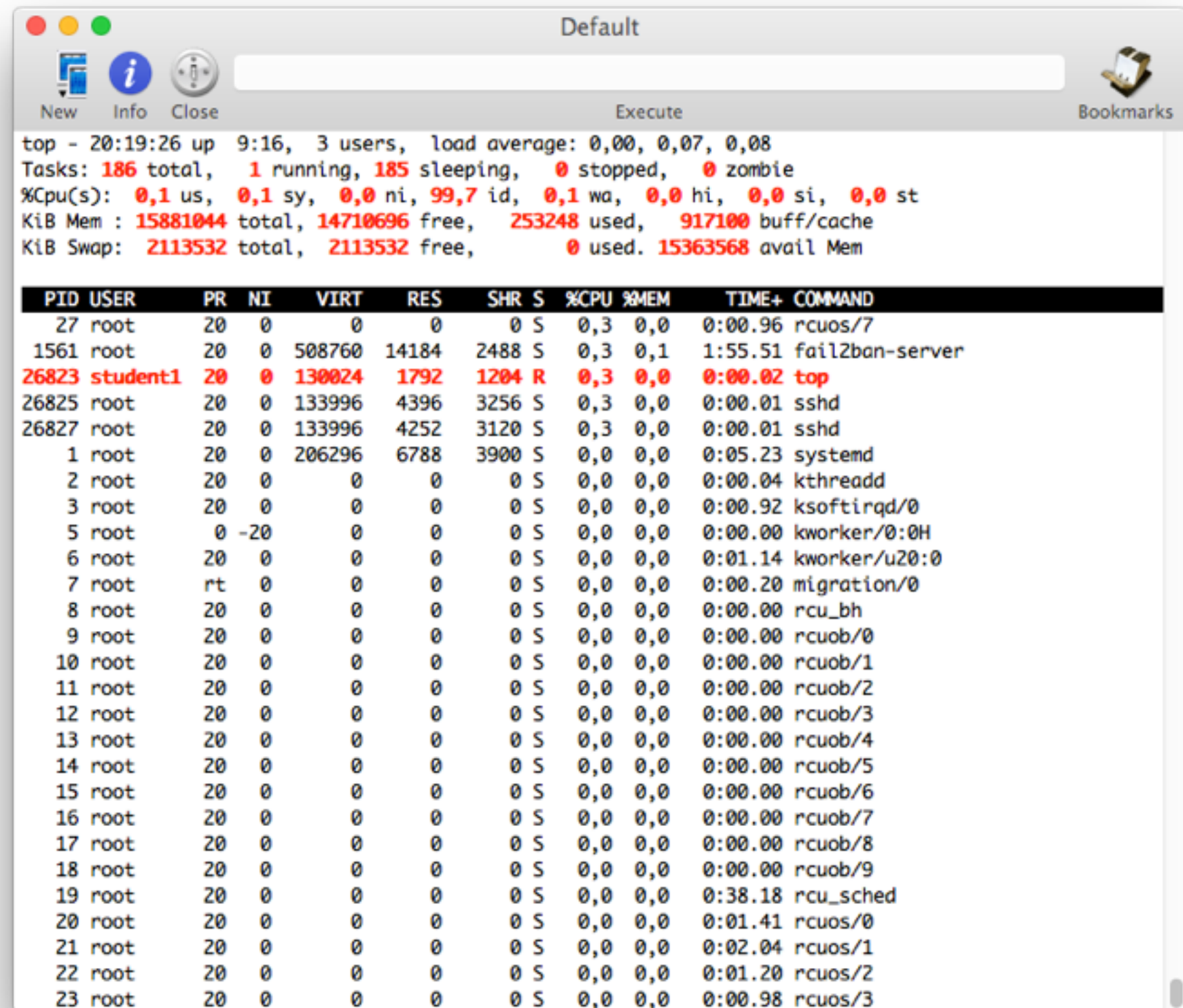
```
-bash-4.2$ cut -f1,2 pairs.tab|sort|uniq -c|sort -k1g|tail
     48 VHL        HIF1A
     54 IKBKG      UBC
     58 MAPK8      JUN
     60 EGFR       CBL
     61 TRAF6      UBC
     73 UBC        TP53
     85 EGFR       UBC
    131 TP53       UBC
    152 TP53       MDM2
    192 MDM2       TP53
```

Controlling jobs

Process information

- **top** - check how much computer resources are used

Process information

- renice - change priority of existing process


```
-bash-4.2$ renice -n 9 26910
26910 (process ID) old priority 0, new priority 9
```

Process information

- ps - current processes

```
-bash-4.2$ ps -U student1 -h
20057 ?         S       0:01 sshd: student1@pts/2
20058 pts/2     Ss      0:00 -bash
26910 pts/2     S+      0:00 /usr/lib64/R/bin/exec/R
26975 ?         S       0:00 sshd: student1@pts/1
26976 pts/1     Ss      0:00 -bash
27019 pts/1     R+      0:00 ps -U student1 -h
```

Process information

- kill - quit a process

  ```
  -bash-4.2$ kill 26910
  ```

Advanced tricks

Getting along with the server

- screen - virtual terminal; helps to keep programs running while you need to disconnect the computer

- uptime - check for how long server has been continuously running

Awk when cut does not cut!

```
-bash-4.2$ cut -f1,3 sorted_pairs.tab |head -n 5
      1 1810055G02Rik
      1 1810055G02Rik
      1 1810055G02Rik
      1 2810046L04Rik
      1 2810046L04Rik

-bash-4.2$ awk '{print $1,$3}' sorted_pairs.tab |head -n 5
1 MPRIP
1 RAI14
1 TPM1
1 ARF1
1 ARF3
```

Searching

- grep - grep for patterns in files

```
-bash-4.2$ grep BRCA1 sorted_pairs.tab |tail -n5
    14 BRCA1    UBE2D3
    19 BRCA1    BRIP1
    21 BRCA1    RBBP8
    24 BARD1    BRCA1
    48 BRCA1    BARD1
```

Exercise #1

In Your home catalog make directory called `data`. Also choose a text editor and start a new file, where You can write the commands You used to solve exercises as well answers to questions.

Exercise #2

File : `/home/student1/data/mart_export_14_05_15.txt.tar.bz2`

- copy to Your own **data/** catalog, extract the file to plain text.

- make Yourselves familiar with the file, explore, look inside, open in text editor.

  - how many lines does the file have?

  - unique lines, can You find repeated lines? Remove them.

Exercise #3

From the edited file:

- Make new file, where You only have lines matching chromosomes 1 - 22, X and Y.

- What gene has the highest / lowest %GC?

- How many genes have %GC above 70, how many below 30?

- Can You find the longest / shortest gene?

Exercise #4

File: `/home/student1/data/`
`mart_export_variation_14_05_15.tar.gz`

- copy to Your own **data/** catalog, extract the file to plain text.

    - express the size difference in percentages

- Make table of "Consequence Type" frequencies

- Make table of "Variant Alleles" frequencies (top 25)

Exercise #5

File: `/home/student1/data/variation.tar.gz`

- copy to Your own **data/** catalog, extract the archive. Essentially the same as last file, split into chromosomes.

- make the same tables as in last exercise for chromosomes 20 - 22 combined

- Given all the files available to You, think of exercise relevant to You, ask help if needed to solve it!

R basics course
next Thursday
room 202

Provide us feedback

- bit.ly/linux15feedback