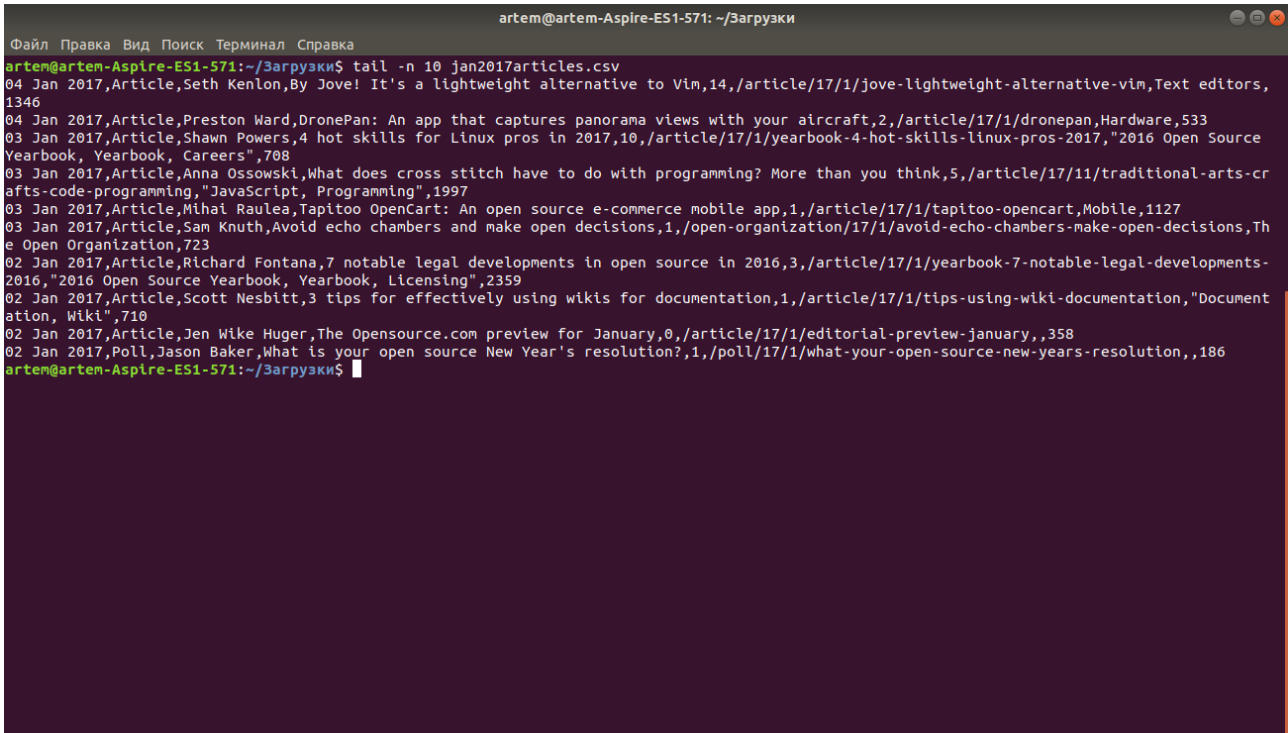# Parallel and distributed systems. Assignment #1.

Artem Reshetnikov, Lawrence Adu-Gyamfi

1. Take a look at the last 10 lines of the file. Which command are you going to use? Modify the command to show just the last line of the file.

Answer: tail -n 10 jan2017articles.csv
tail -n 1 jan2017articles.csv

2. Extract all lines that belong to January 6th from the file and store them in a new file named "reyes.csv". Check that the first line of the new file has the expected values.

Answer: grep "06 Jan" jan2017articles.csv > reyes.csv
- head -1 reyes.csv

3. Use the original csv to find which entries have 0 at the comment count only for those entries from january 25th

Answer: grep "25 Jan" jan2017articles.csv | awk -F "," '$5 == "0"'

4. Now count the number of entries of Q3 and compare with the total number of entries

Answer: grep "25 Jan" jan2017articles.csv | awk -F "," '$5 == "0"'| wc -l
wc -l jan2017articles.csv

5. Now use example.bed file. In this file, we are interested in the exon sizes of each entry. They are located in field number 11. Now you have to get the exon sizes of the first 10 entries of the file.

Answer: head -n 10 example.bed | cut -f 11

6.How would you remove the last comma?

Answer: head -n 10 example.bed | cut -f 11 | sed -e 's/,*\r*$//'

7. How would get the smallest size from each of the records? The result should provide a number for each line of the input

Answer 1 : head -n 10 example.bed | cut -f 11 | sed -e 's/,*\r*$//'| gawk ' {split($0, a, ","); asort(a); printf("%s ", a[1]); printf("\n"); }'

Answer 2: head -10 example.bed | cut –f 11 | sed -e 's/,*\r*$//' | awk –F "," '{min=$1; for(i=2; i<=NF; i++) if($i<min) min=$i; print min}'
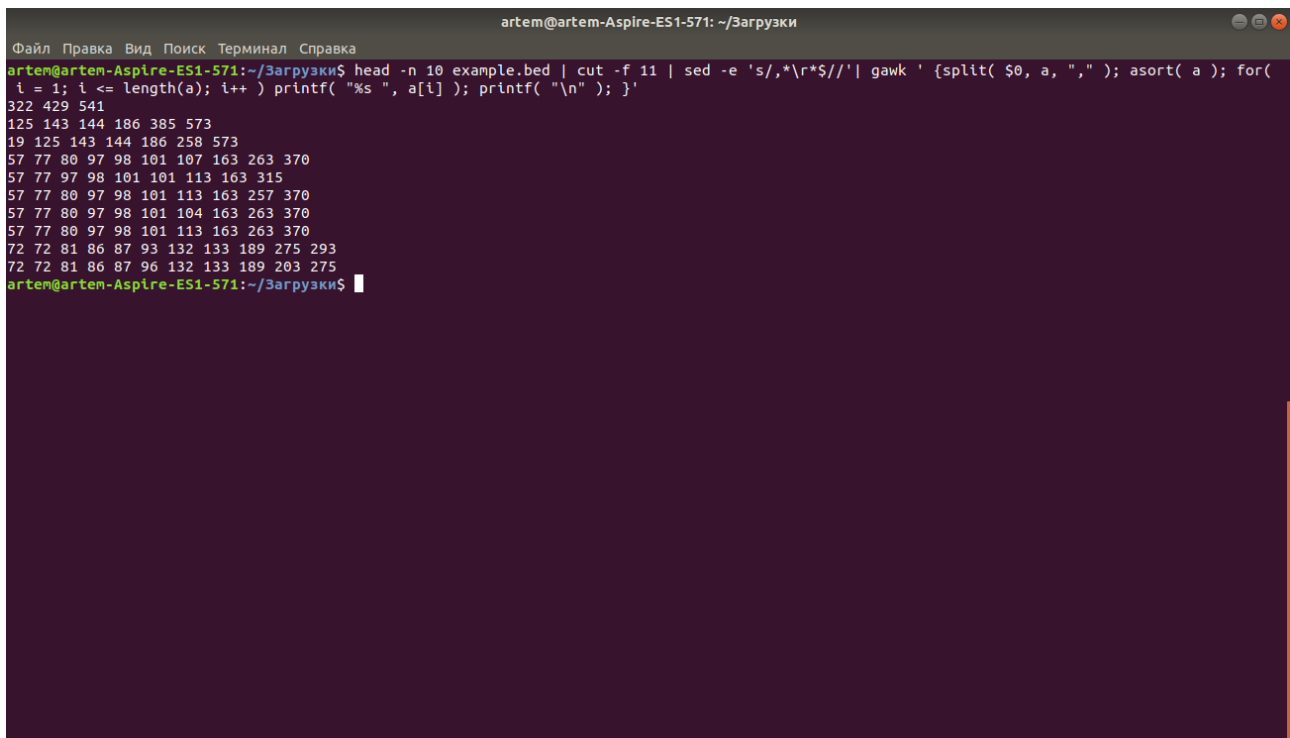
8. How would you now sort the records so that the first number shown is the smallest exon size? Again, the answer must provide a sorted list of numbers for each line of the input

Answer 1: head -n 10 example.bed | cut -f 11 | sed -e 's/,*\r*$//'| gawk ' {split($0, a, ","); asort(a); for(i = 1; i <= length(a); i++) printf("%s ", a[i]); printf("\n"); }'

Answer 2: head -10 example.bed | cut -f11 | sed 's/,$//g' | awk -F ',' -v OFS=',' '{j=0; while (j<NF) {for (i=1; i<NF;i++) if ($(i+1) < $(i)) {temp=$(i+1); $(i+1)=$i; $i=temp;} j++} print $0}'

9. Now Get the 10 largest exons of chr1 stored in example.bed.

Answer 1: grep "chr1" example.bed | cut -f11 | sort -n -r | cut -d "," -f1 | head -n 10

Answer 2: grep -i 'chr1' example.bed | cut -f11 | awk -F ',' '{max=$1; for (i=1; i<NF; i++) if ($i > max) max = $i; print max}' | sort -rn | head -10

10. Now modify Q9 script to receive as a parameter the number of exons to search for

Answer: *Text of script*
#!/bin/bash
n=$1
grep "chr1" example.bed | cut -f11 | sort -n -r | cut -d "," -f1 | head -n $n

Text of commands:

chmod +x q10.bach
./q10.bash 15

11. Get the first 10 records of jan2017articles.csv with largest number of comments from the original csv file

Answer: sort -t "," -k5 -n -r jan2017articles.csv | head -n 10

12. Modify your previous script to receive a number as a parameter N and then show the top N entries with more comments
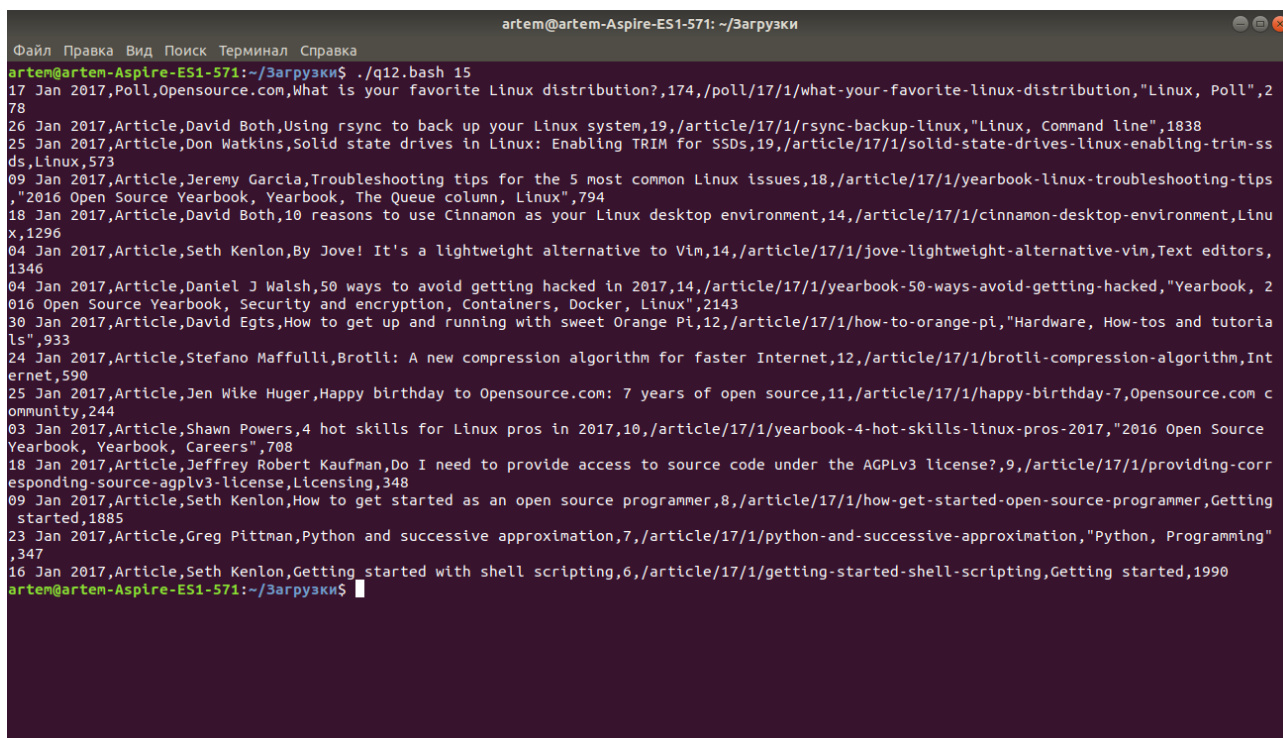
Answer:
Text of script:

```
#!/bin/bash
n=$1
sort -t "," -k5 -n -r jan2017articles.csv | head -n $n
```

Text of commands

```
chmod +x q10.bach
./q12.bash 15
```


13. Now we are going to create a new articles.csv where we get a different output data layout using awk tool INPUT: Post date,Content type,Author,Title,Comm count,Path,Tags,Word count OUTPUT: Title;Comment count;Word count;Post date

Answer: awk -F "," '{print $4"\t"$5"\t"$8"\t"$1}' jan2017articles.csv > articles.csv

```
artem@artem-Aspire-ES1-571:~/Загрузки$ awk -F "," '{print $4"\t"$5"\t"$8"\t"$1}' jan2017articles.csv
Title     Comment count   Word count      Post date
Book review: Ours to Hack and to Own    0       660     31 Jan 2017
5 new guides for working with OpenStack 2       How-tos and tutorials" 31 Jan 2017
Be the open source supply chain 1       1668    31 Jan 2017
Developing open leaders 1       768     31 Jan 2017
How to get up and running with sweet Orange Pi  12      How-tos and tutorials" 30 Jan 2017
4 ways to improve your security online right now        3       1242    30 Jan 2017
" WOOTConf 2017: Lockpicking     Willie Nelson developers        /article/17/1/women-open-technology-miniconf     30 Jan 2017
"From hobbyist to professional  new analyst papers      /article/17/1/openstack-news-january-30 30 Jan 2017
How communities in India support privacy and software freedom   0       453     28 Jan 2017
"New Minecraft launcher comes to Linux  Tilt Brush Toolkit      /article/17/1/open-gaming-news-january-28        28 Jan 2017
"Top 5: Solid state drives in Linux     Brotli compression algorithm    /article/17/1/top-5-january-27  27 Jan 2017
Data Privacy Day 2017: Solutions for everyday privacy   5       Security and encryption"        27 Jan 2017
How to successfully enter the FOSS emerging market      0       915     27 Jan 2017
Getting started with my new LulzBot Mini 3D printer     2       1390    27 Jan 2017
A 5-step plan to encourage your team to make changes on your project    5       1160    26 Jan 2017
Search this database for inactive patents that are now in the public domain     0       Patents"        26 Jan 2017
Using rsync to back up your Linux system        19      Command line"   26 Jan 2017
How to join a technical community       1       Community"      26 Jan 2017
Solid state drives in Linux: Enabling TRIM for SSDs     19      573     25 Jan 2017
An executive's guide to containers      4       2032    25 Jan 2017
24 Pull Requests challenge encourages fruitful contributions    0       429     25 Jan 2017
Happy birthday to Opensource.com: 7 years of open source        11      244     25 Jan 2017
Announcing the 2016 Open Source Yearbook: Download now  0       58      25 Jan 2017
Brotli: A new compression algorithm for faster Internet 12      590     24 Jan 2017
School systems desperate for standards-aligned curricula find hope      0       Education"      24 Jan 2017
What your code repository says about you 1       1085    24 Jan 2017
Unpicking the community leader  0       Six Degrees column"     23 Jan 2017
Win a LulzBot Mini 3D printer in our giveaway   1       Getting started"        23 Jan 2017
Shell scripting: An introduction to the shift method and custom functions       4       Programming"    23 Jan 2017
Python and successive approximation     7       Programming"    23 Jan 2017
2017 People's Choice Awards: Cast your vote     3       562     23 Jan 2017
"Tips for instance configuration        creating a new project /article/17/1/openstack-news-january-23 23 Jan 2017
"The new Raspberry Pi Compute Module    FSF's 2017 priorities   /article/17/1/news-january-21   21 Jan 2017
5 ways to expand your project's contributor base        2       690     20 Jan 2017
How to write web apps in R with Shiny   2       218     20 Jan 2017
"Top 5: Shell scripting  the Cinnamon Linux desktop environment /article/17/1/top-5-january-20  20 Jan 2017
```

14. Now create a new article2.csv format where we cut the Title text to 10 characters and we get only the last level of the Path.

Answer: awk -F "," '{split($6,arr,"/"); print substr($4,1,10) "\t" "/"arr[4]"/"arr[5]}'
jan2017articles.csv > article2.csv


```
artem@artem-Aspire-ES1-571:~/Загрузки$ awk -F "," '{split($6,arr,"/");  print substr($4,1,10) "\t" "/"arr[4]"/"arr[5]}' jan2017articles.csv
Title     //
Book revie      /1/review-book-ours-to-hack-and-own
5 new guid      /1/openstack-tutorials
Be the ope      /1/be-open-source-supply-chain
Developing      /1/developing-open-leaders
How to get      /1/how-to-orange-pi
4 ways to       /1/4-ways-improve-your-online-security
" WOOTConf      //
"From hobb      //
How commun      /1/how-communities-india-support-privacy-software-freedom
"New Minec      //
"Top 5: So      //
Data Priva      /1/every-day-privacy
How to suc      /1/cultivating-business-foss-market
Getting st      /1/beginner-LulzBot-Mini-3D-printer
A 5-step p      /1/escape-the-cave
Search thi      /1/making-us-patent-system-useful-again
Using rsyn      /1/rsync-backup-linux
How to joi      /1/how-join-technical-community
Solid stat      /1/solid-state-drives-linux-enabling-trim-ssds
An executi      /1/container-strategy-for-executives
24 Pull Re      /1/24-Pull-Requests-challenge
Happy birt      /1/happy-birthday-7
Announcing      /1/announcing-2016-open-source-yearbook
Brotli: A       /1/brotli-compression-algorithm
School sys      /1/open-up-resources
What your       /1/repo-tells-a-story
Unpicking       /1/community-leadership-matters
Win a Lulz      /1/getting-started-giveaway
Shell scri      /1/shell-scripting-shift-method-custom-functions
Python and      /1/python-and-successive-approximation
2017 Peopl      /1/2017-peoples-choice-awards
"Tips for       //
"The new R      //
5 ways to       /1/expand-project-contributor-base
How to wri      /1/writing-new-web-apps-shiny
"Top 5: Sh      //
```