LAB EXERCISE #1

MSc. Computational Modelling

Module I: Linux scripting

Deliverable 1. Linux Lab Exercices

Toni Espinosa: antoniomiguel.espinosa@uab.cat
Department of Computer Architecture and Operating Systems, UAB

Exercise description:
This work must be done in groups and must submit the source code for each question in the same document

Solutions for these questions must be delivered until November 15th

Download jan2017articles.csv and example.bed files to solve the next questions.

Q1 (1 point): 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.

Q2 (1p): 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.

Q3 (1p): Use the original csv to find which entries have 0 at the comment count only for those entries from january 25th

Q4 (1p): Now count the number of entries of Q3 and compare with the total number of entries

Q5 (1p): 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.

Q5 (1p): How would you remove the last comma?

Q6 (1p): How would get the smallest size from each of the records? The result should provide a number for each line of the input

Q7 (1p): 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

Q8 (1p): Now Get the 10 largest exons of chr1 stored in example.bed.

Q9 (1p): now modify Q9 script to receive as a parameter the number of exons to search for

Q10 (1p): Get the first 10 records of jan2017articles.csv with largest number of comments from the original csv file

Q11 (1p): Modify your previous script to receive a number as a parameter N and then show the top N entries with more comments

Q12 (1p): 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

Q13 (1p): 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.