

**TERCER EJERCICIO** de las pruebas selectivas para el acceso a la Escala de Ayudantes de Investigación de los Organismos Públicos de Investigación por el sistema de acceso libre, convocadas mediante Resolución del 16 de diciembre de 2022 de la Subsecretaría de Ciencia e Innovación.

**Especialidad “Asistencia en ciencia de datos en investigación” (Tribunal N° 6)**

---

Resuelve por escrito **UNO (1)** de los dos supuestos prácticos propuestos a continuación.

El tiempo de realización de este ejercicio es de **90 MINUTOS (1,5 horas)**.

No olvides **poner el nombre** en todas las hojas.

No podrás ausentarte del aula durante los **primeros 15 minutos** de examen ni cuando **falten 15 minutos para finalizar el mismo**. En caso de ausentarte antes de la finalización del examen no podrás llevarte las preguntas.

---

1.- En un computador con Linux, indica los comandos de la shell que debes usar para completar las siguientes tareas:

- a1) Eres superusuario. Crea un usuario llamado "pepito" y ponle como contraseña "Xy13.17zT"
- a2) Eres ahora el usuario "pepito" y estás dentro de tu carpeta de trabajo ("/home/pepito"). Crea una subcarpeta "datos" y otra "trabajo".
- a3) En la carpeta "datos" tienes cinco ficheros de texto con extensión "txt", copiarlos a la carpeta "trabajo".
- a4) Accede a la carpeta "datos". Borra los ficheros con extensión "txt" que hay en ella.
- a5) Baja al directorio de inicio "/home/pepito" y luego sal del sistema.

b1) Bájate estos dos ficheros de estas direcciones utilizando comandos de la shell:

`https://ftp.test.org/pub/chromosome.I.fa.gz`

`https://ftp.test.org/pub/chromosome.III.fa.gz`

b2) Descomprímelos. Indica que comandos utilizas.

b3) Renómbralos, el primero como "C\_ele\_I.fa" y el segundo como "C\_ele\_III.fa"

b4) Guarda en un fichero nuevo llamado "filas\_40.fa" las 20 primeras filas de "C\_ele\_I.fa" y las 20 últimas filas de "C\_ele\_III.fa"

b5) Elimina de este fichero "filas\_40.fa" todas las filas que comiencen por el carácter ">". Y guarda el resultado en un fichero llamado "letras.fa".

b6) ¿Cuántos caracteres tiene este fichero "letras.fa"? Averígualo usando un comando de la shell.

b7) Ordena alfabéticamente de la A a la Z las filas de "letras.fa" y guarda el resultado en el fichero "letrasor.fa".

b8) El fichero "letras.fa", está formado sólo por las letras de los nucleótidos que forman una cadena de ADN, que son {A, C, G, T}. Genera un nuevo fichero llamado "ficha2.txt", donde la letra "T" es sustituida por la "X", la "A" por la "Y", la "G" por la "Z" y la "C" por la "W".

b9) Mueve todos los ficheros con extensión "fa" a la carpeta "trabajo" que ha generado al principio.

b10) Borra todos los ficheros con extensión "txt".

Utilizando la Shell o con cualquier lenguaje de programación (C, Python, Java, Perl, ...) resuelve estas 3 cuestiones:

c1) Ve a la carpeta "trabajo". En ella construye un script/programa que, utilizando estructuras de control, como por ejemplo *for*, *while-do* o *if-else* genere por pantalla la siguiente salida:

```
1
12
123
1234
12345
123456
1234567
```

c2) Realiza un script/programa que pida por pantalla un número N entre 20 y 99. Que compruebe que el número introducido está en este rango, si no, lo vuelve a pedir. Una vez introducido, que genere N números aleatorios entre 50 y 150 y que lo guarde en un fichero de texto llamado "numeros.txt"

c3) Dado el siguiente fichero de texto "data.txt" que contiene 4 filas de números separados por espacios:

```
13 33 31 7 11 3
90 23 73 63 33
21 13 33 44 55 31
13 21 71 15 93
```

construye un script/programa que lea el fichero y nos dé el número menor, el mayor y el promedio de estos datos.

---

2.- Un centro de supercomputación del CSIC alberga un clúster llamado "sirocco".

a) Necesitas controlar remotamente cuantos procesos están corriendo en la máquina 'sirocco.csic.es' que gestiona un sistema de colas. Describe los comandos de la shell que usarías para, desde una máquina Linux, conectarte remotamente a esta máquina y ejecutar un comando que te permita ver cuantos trabajos hay en curso.

b) Describe los comandos de la shell que usarías para mostrar todos los trabajos que el usuario "pepito" tiene en el sistema de colas y para finalizarlos (como superusuario).

c) Describe los comandos de la shell que usarías para asignarle 8 CPUs y 128 Gb de memoria RAM a un trabajo que ya había sido introducido en el sistema de colas.

d) Un usuario ha bajado el siguiente ejemplo de fichero script PBS. Te pregunta que le expliques para que sirve cada línea del fichero y qué ha de hacer para ejecutarlo. Demuestra que podrías ayudarlo. Describe en detalle lo que ejecuta cada comando del siguiente script que está guardado en un fichero llamado "mis-trabajos.pbs". Y luego escribe el comando que podría ejecutarlo:

```
#PBS -S /bin/bash
#PBS -q batch
#PBS -l nodes=1:ppn=10
#PBS -l mem=64gb
#PBS -l walltime=48:00:00
#PBS -N myscript
#PBS -o myscript.out
#PBS -e myscript.err
#PBS -m abe
#PBS -M d.alonso@csic.es

cd $PBS_O_WORKDIR/
module load R/3.5.1
Rscript --vanilla myscript.R
```

---