

HERRAMIENTAS COMPUTACIONALES

SEMANA 3. UNIX

SECCIÓN-03

2018-2

El archivo final (.sh) debe subirse a SicuaPlus con el nombre del estudiante en el formato `NombreApellido_HW1.sh` antes que termine la clase.

En este taller vamos a trabajar con dos archivos que indican valores de *carat* (peso) y precio en USD de una base de datos de diamantes, con los cuales se busca explorar algunas herramientas de Linux para la edición y exploración de archivos de texto. Estos archivos se encuentran en los siguientes links:

https://raw.githubusercontent.com/ComputoCienciasUniandes/HerramientasComputacionalesDatos/master/data/DistributionData/diamond_carat.csv

https://raw.githubusercontent.com/ComputoCienciasUniandes/HerramientasComputacionalesDatos/master/data/DistributionData/diamond_price.csv

Para el desarrollo del taller, se recomienda utilizar los comandos directamente en consola para comprobar su funcionamiento y luego consignarlos en el archivo .sh. Cabe resaltar que al ejecutar el script este debe realizar cada uno de los puntos del taller, recuerde que puede crear archivos intermedios pero estos deben ser borrados después por el mismo script.

Por último, antes de la ejecución de cada punto el script debe indicar al usuario que punto está resolviendo (**5 puntos**) con un mensaje de la forma **Punto X...** donde *X* corresponde al punto que esté resolviendo. Por ejemplo para el primer punto debe imprimir:

Punto 1...

1. (5 points) Cree una carpeta cuyo nombre corresponda a su nombre y apellido siguiendo el formato **NombreApellido** y trabaje sobre ella.
2. (10 points) Descargue los archivos con la información de los diamantes.
3. (10 points) Cambie el nombre del archivo `diamond_carat.csv` a `diamond_weight.csv`
4. (10 points) Queremos unir la información de ambos archivos en un único catálogo de diamantes con el nombre `diamond_all.csv` el cual debe contener 3 columnas correspondientes a: el número del diamante, el peso del diamante y por último el precio. Para esto existen dos posibles métodos, usted sólo debe realizar **uno**:
 - (a) Utilizando el comando **paste** puede unir los dos archivos en uno, el cual contendrá 4 columnas donde 2 de ellas corresponden al número del diamante. Luego de unir los archivos en uno temporal, utilizando **awk** con la opción **-F** (consulte para que sirve en el manual) imprima las columnas 1,2 y 4 en el archivo final (`diamond_all.csv`) utilizando el operador **>**.
 - (b) Usando **awk** usted puede imprimir las columnas necesarias de cada archivo y pasarlas como entrada al comando **paste** utilizando el operador **<**.
5. (10 points) Imprima el número de diamantes que hay en el catálogo (tenga en cuenta que un diamante corresponde a una línea en el catálogo) con el mensaje:

El total de diamantes es
n5.

Los puntos siguientes deben hacer uso del archivo de texto diamond_all.csv, sí utiliza otro archivo sólo se otorgarán la mitad de los puntos correspondientes.

6. (10 points) Imprima el número de diamantes que tienen un precio **menor** a 400 USD. El resultado se debe ver de la forma:

Hay

n6

Diamantes con precio menor a 400

7. (10 points) Imprima el número de diamantes con peso **mayor** a 1.0. El resultado se debe ver de la forma:

Hay

n7

Diamantes con peso mayor a 1.0

8. (10 points) Imprima el número de diamantes con peso **mayor** a 1.0 y cuyo precio sea menor a 2575USD. El resultado se debe ver de la forma:

Hay

n8

Diamantes con peso mayor a 1.0 y precio menor a 2575

9. (10 points) Imprima el número de diamantes con peso **menor** al resultado de dividir su edad entre 100 y cuyo precio sea menor a la fecha de su nacimiento. El resultado se debe ver de la forma:

Hay

n9

Diamantes con peso mayor a "edad"/100 y precio menor a "año nacimiento"

10. (10 points) Por último borre todos los archivos utilizados y la carpeta de trabajo. Recuerde no borrar el script en el que trabajó.