

Arquitecturas y Sistemas Operativos

Año: 2020

“Introducción a Linux para Bioinformática”

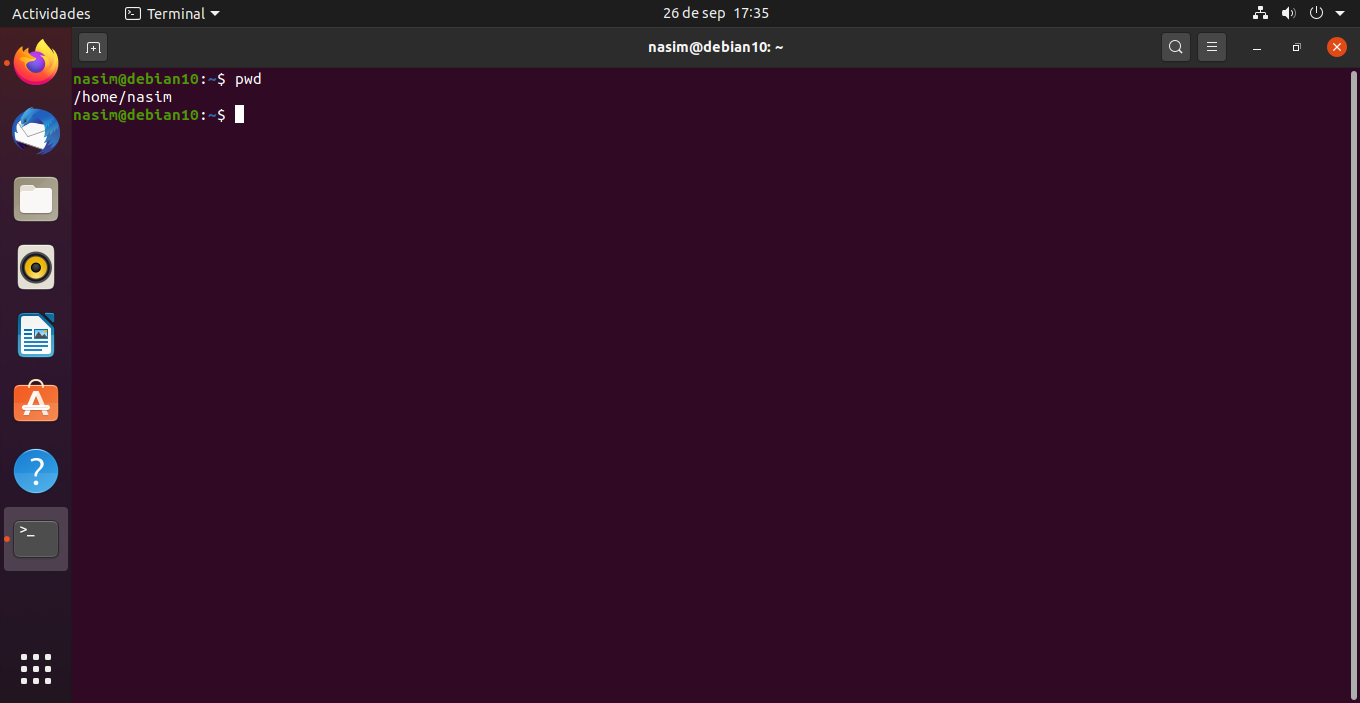
Alumno: Salim Taleb, Nasim Anibal

Docentes: Schneider, Gustavo; Wilka, Daniel

**FECHA DE ENTREGA:** 28/09/2020

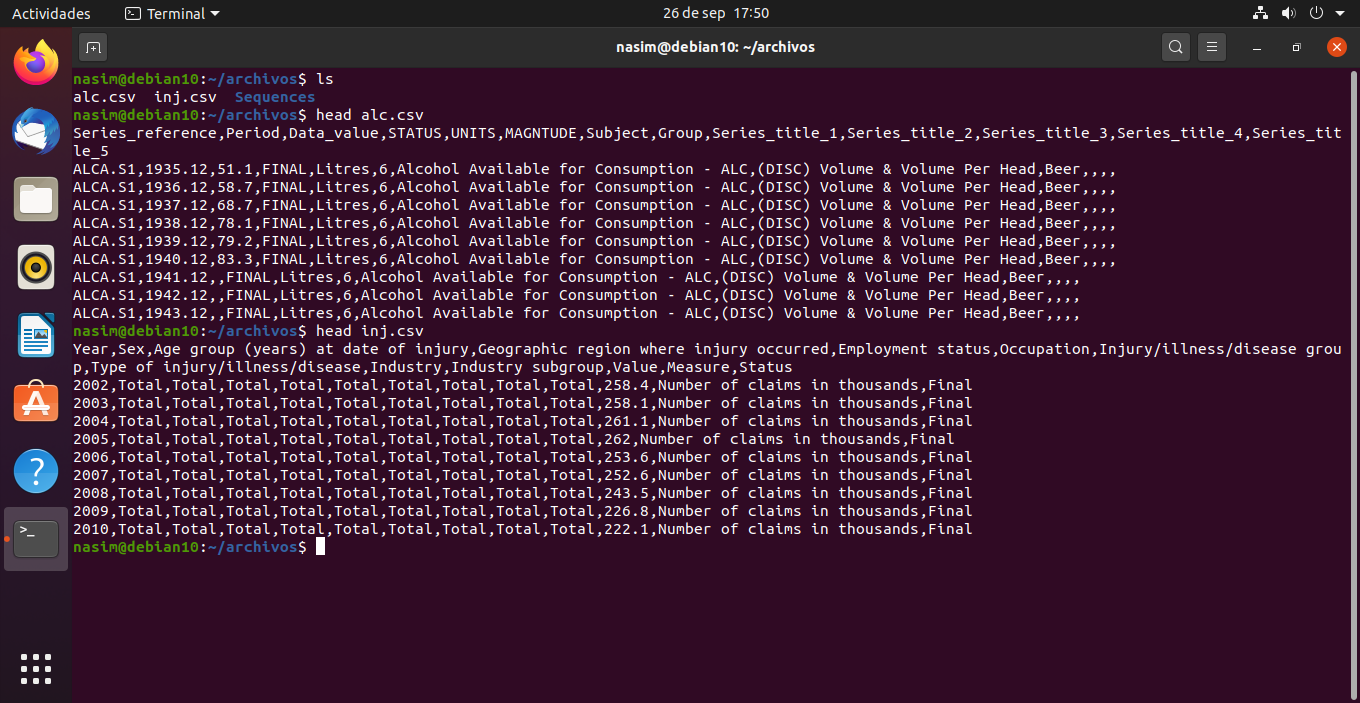
Desarrollo

1. Utilizando el comando pwd, se puede ver el directorio actual, posterior a conectarse al servidor remoto.



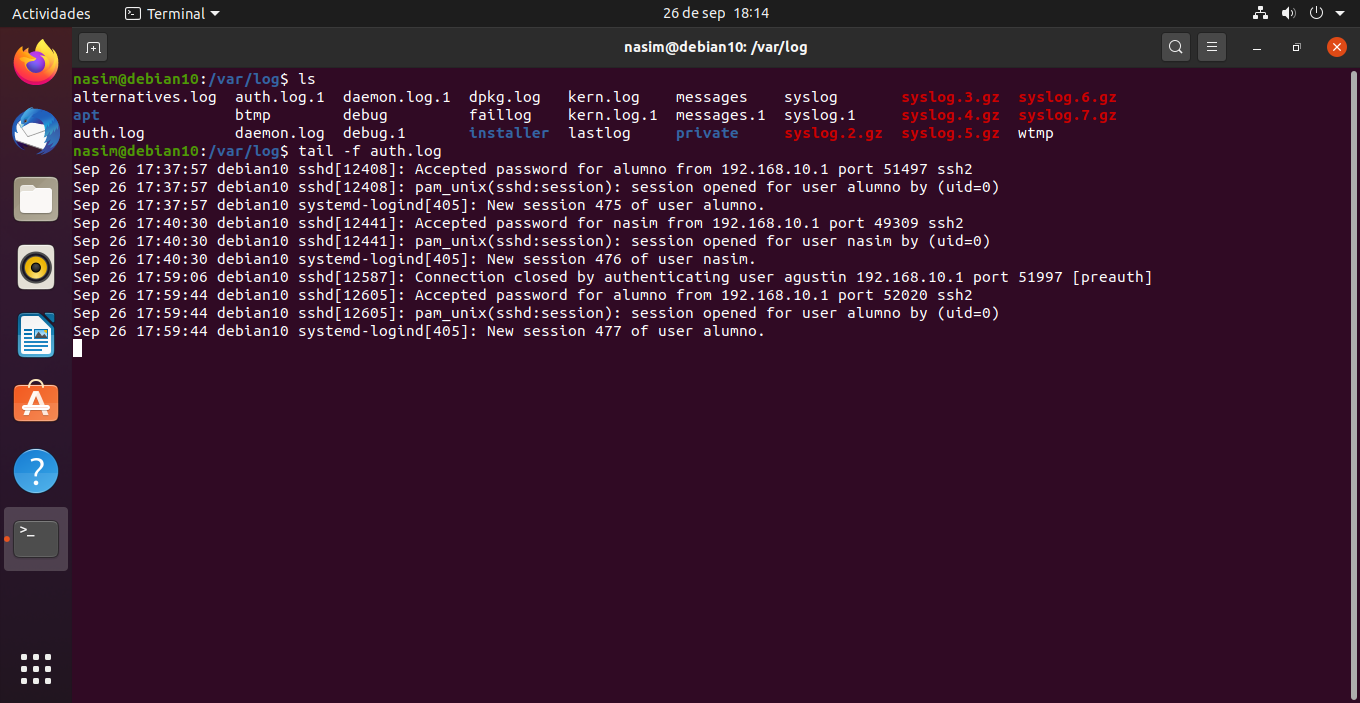
Actualmente el directorio es /home/nasim.

2. Utilizando el comando cat [nombre del archivo] se muestra en consola el archivo completo, el cual resulta complicado de leer. El comando less [nombre del archivo] permite ver el archivo de manera que se pueda desplazar en el texto que contenga haciendo más fácil de leer su contenido. El comando head [nombre del archivo] muestra por consola una pequeña cantidad de las primeras líneas del archivo. Finalmente, tail [nombre del archivo] muestra algunas líneas del final del archivo.

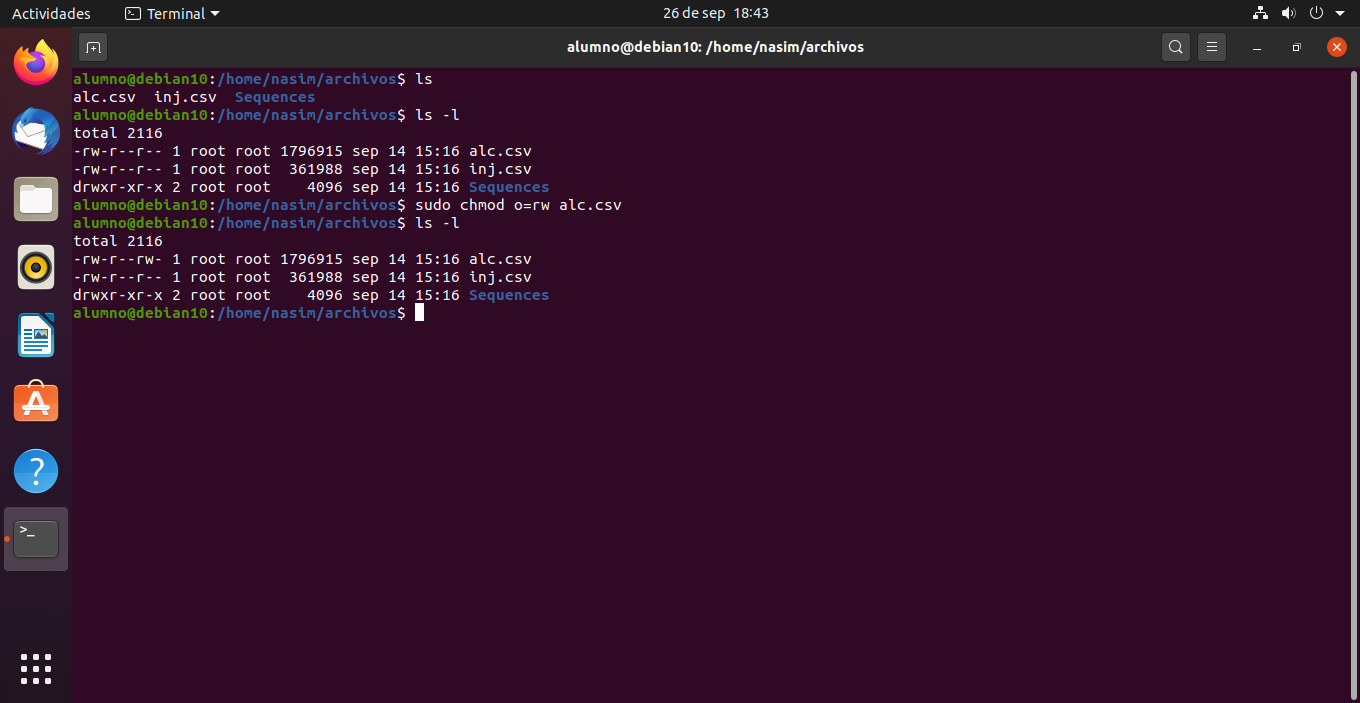


Utilizando el comando ls para ver los archivos en el directorio y head para ver los encabezados, se puede observar que el archivo “alc.csv” contiene información respecto a bebidas alcohólicas. El archivo “inj.csv” parece contener información respecto a accidentes/heridas de personas.

3. El comando tail seguido de -f permite monitorizar en tiempo real el final de un archivo. Puede ser utilizado como en este caso, para visualizar archivos de registros en tiempo real a medida que los usuarios ingresan y salen del sistema. Introduciendo el comando se observa lo siguiente:

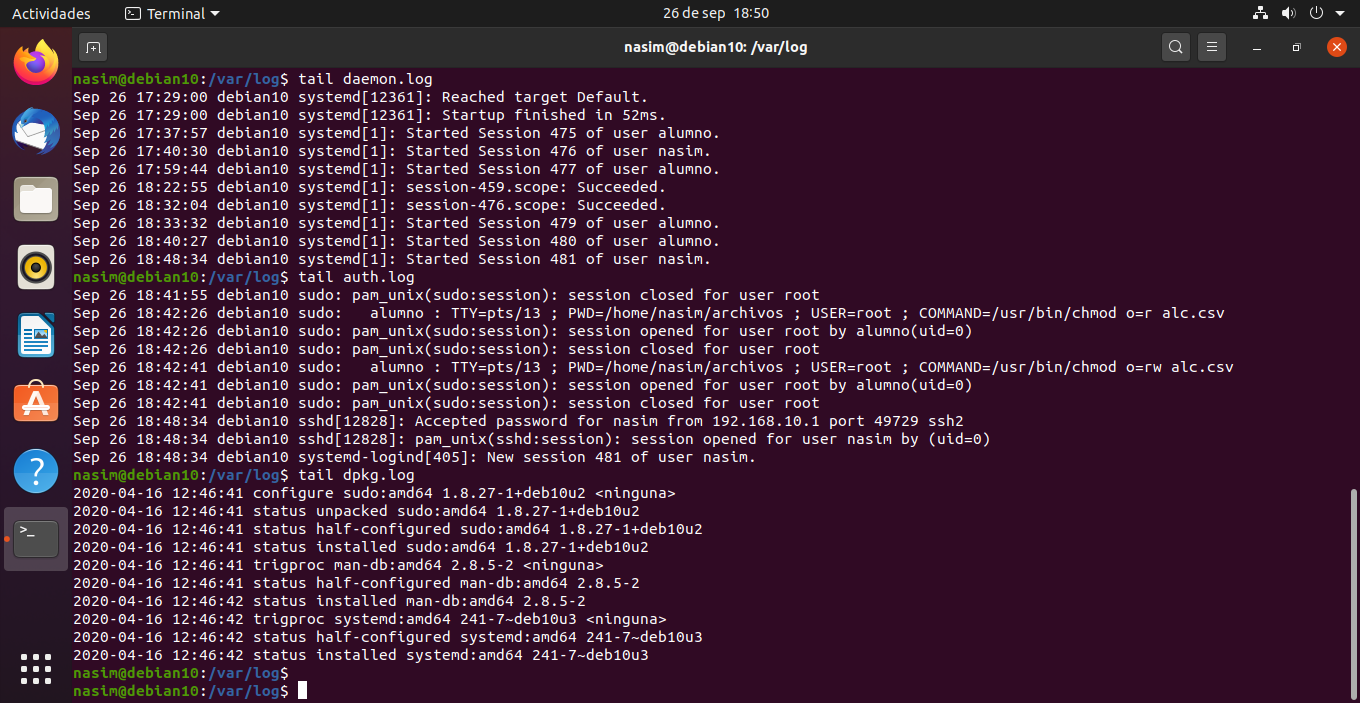


4. La secuencia de comandos para dicha acción será:

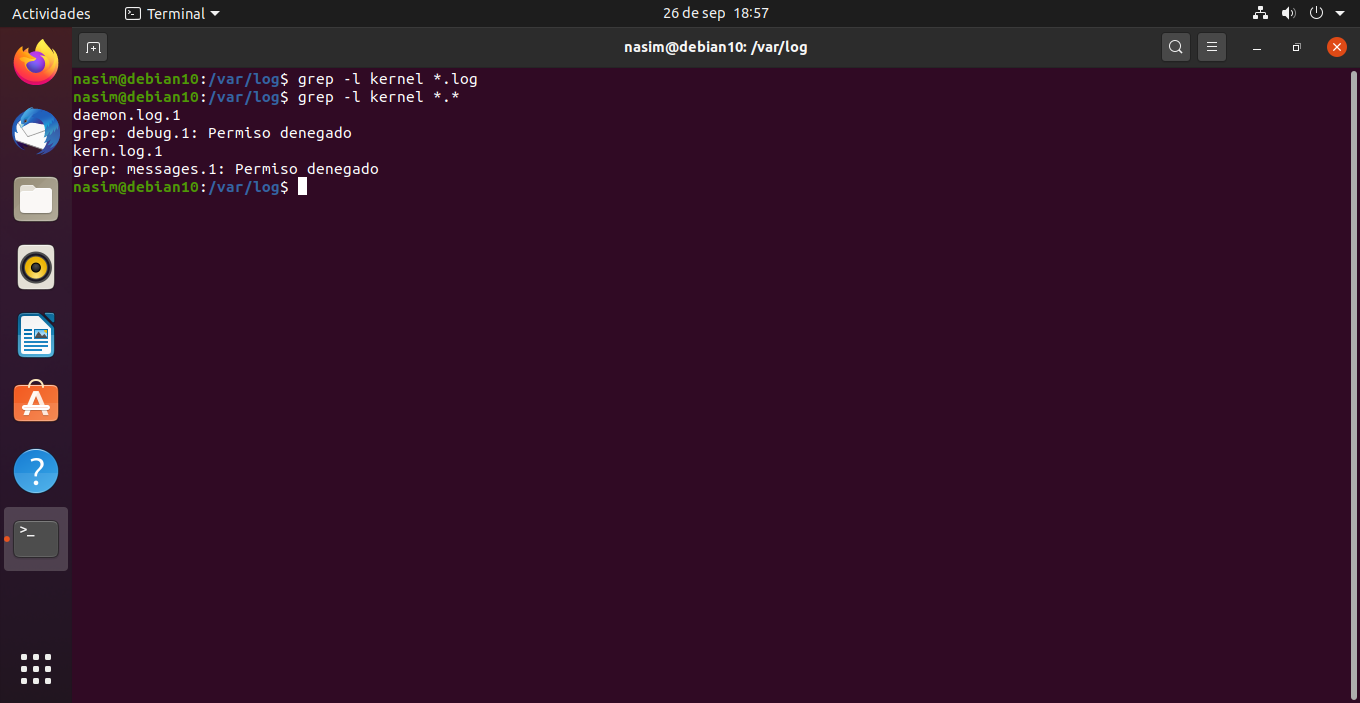


Para este punto se debe ingresar mediante un usuario root ya que el usuario “nasim” no poseía los permisos para realizar la modificación de permisos del archivo. Primero se muestran los archivos y posteriormente se agrega –l para ver en más detalle cada uno, para modificar los permisos usos chmod, precedido por sudo para ejecutar la instrucción como súper usuario, luego ingresamos a que usuario decidimos modificar (‘o’ en este caso) y le asignamos los permisos, en este caso rw o lectura y escritura. Al final se puede ver que se modificaron los permisos exitosamente.

5. Algunos archivos de registro:

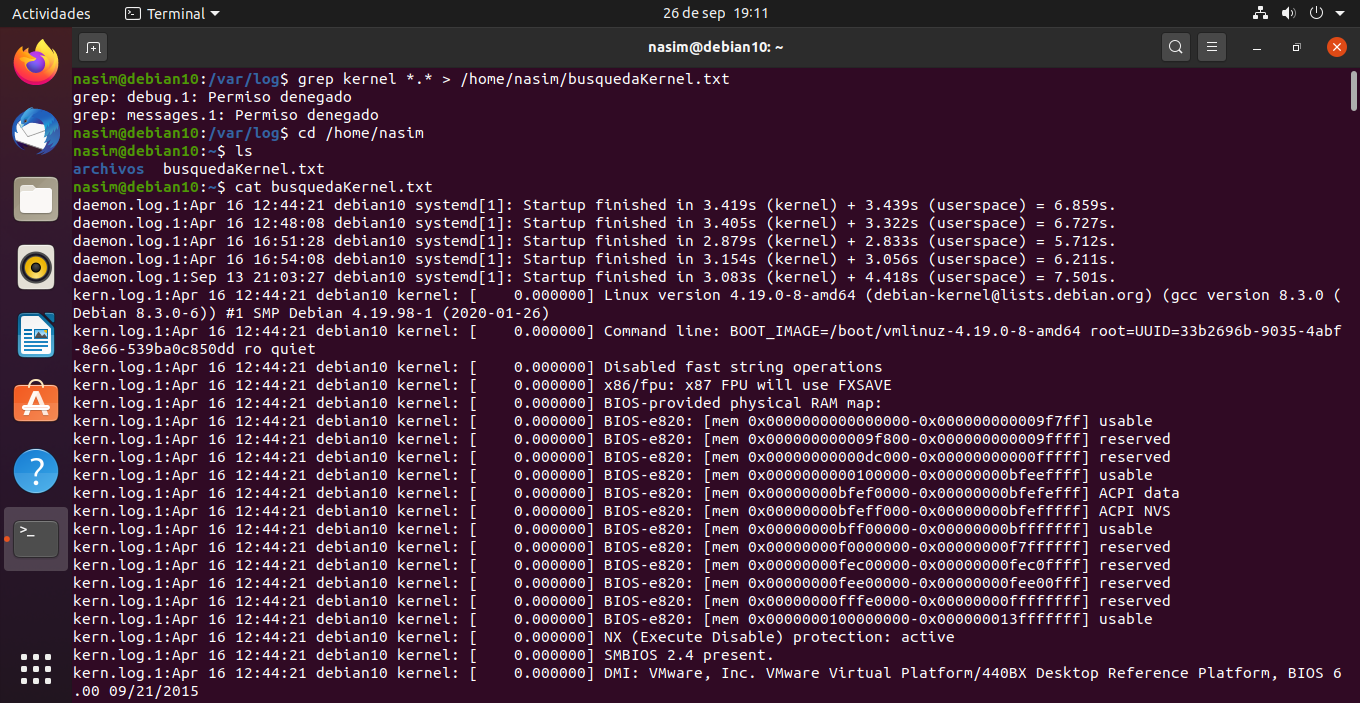


6. Para buscar en los archivos de extensión “.log” la palabra kernel, se utiliza el comando grep acompañado de –l para que solo se muestre el nombre de los archivos. De la siguiente manera:



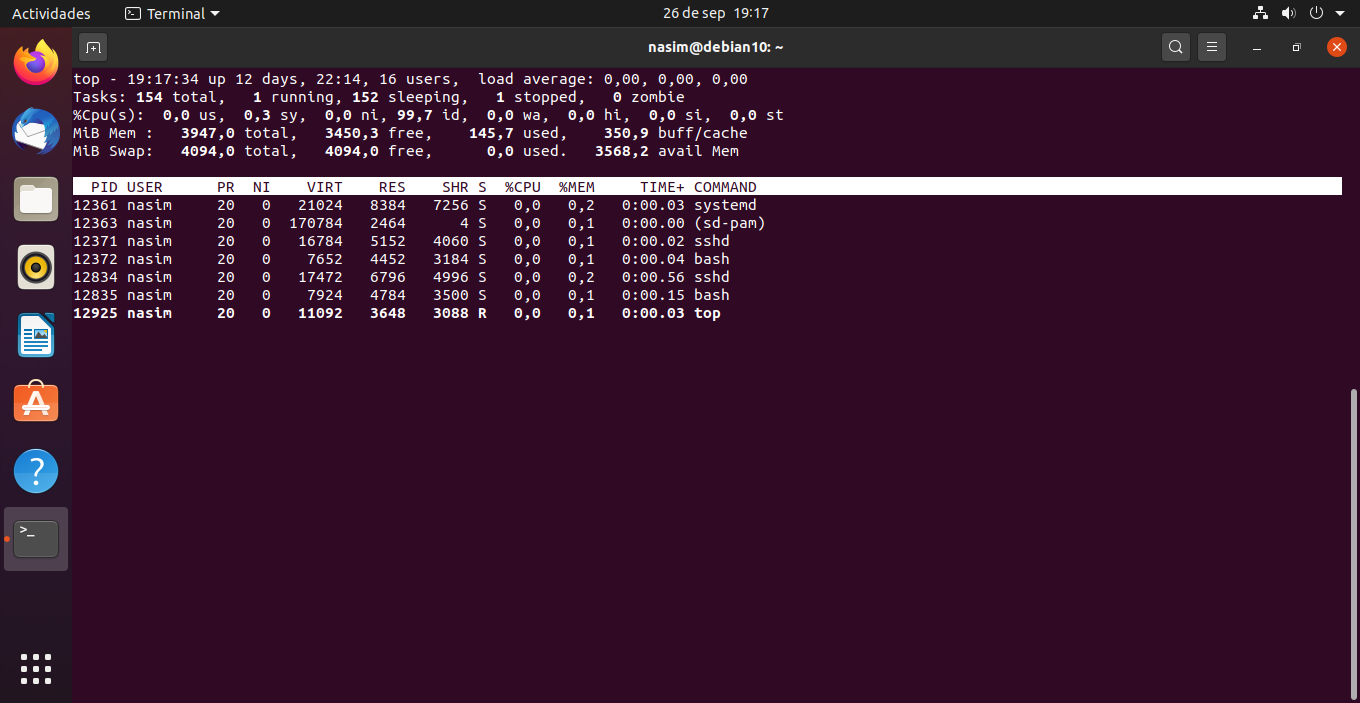
Ya que la búsqueda no devolvió ningún resultado, se probó buscando en todos los archivos, independientemente de su extensión, donde se encontró kernel en “daemon.log.1” y “kern.log.1”, no se pudo acceder a “debug.1” y “messages.1” por no tener permisos suficientes.

7. El procedimiento será el siguiente:



Con el comando grep buscamos las líneas que contengan kernel en todos los archivos del directorio y mediante ‘>’ volcamos la salida en un archivo de texto que se crea en /home/nasim. Luego se muestra el contenido como prueba de que se creó el archivo satisfactoriamente.

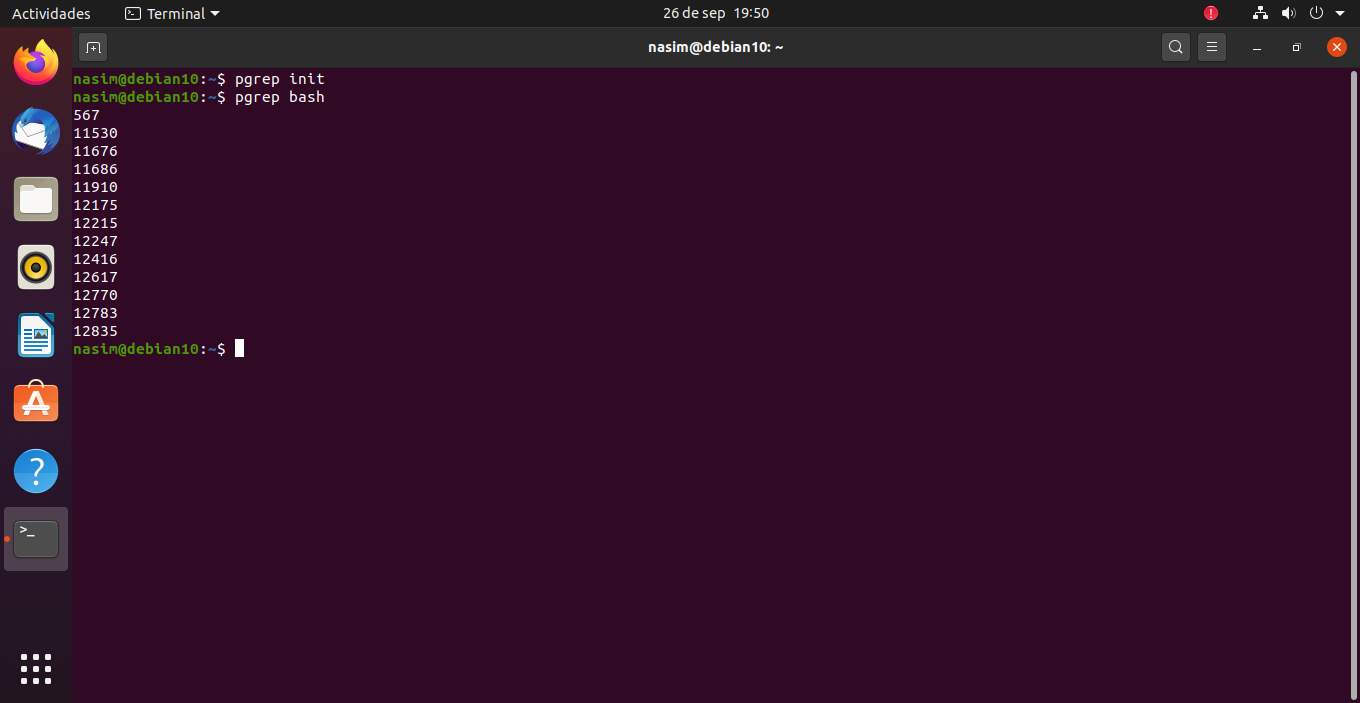
8. Para analizar los procesos en ejecución se utiliza el comando top, el cual puede ir acompañado de –u [usuario] para filtrar y mostrar solo los procesos de un usuario especifico como se muestra en la imagen, también se puede adicionar –p [PID] para monitorear un proceso en específico cuya PID sea la ingresada.



9. Mostrado en 8.

10. El comando top no ofrece una opción para filtrar los procesos de 2 usuarios. Las alternativas son: mostrar los procesos de todos los usuarios, correr un top para cada usuario en un terminal distinto o, mediante entubamiento, buscando en la salida los 2 usuarios que se quiera ver, aunque esto no se actualizaría en tiempo real.

11. Buscando las PID de los procesos dados:

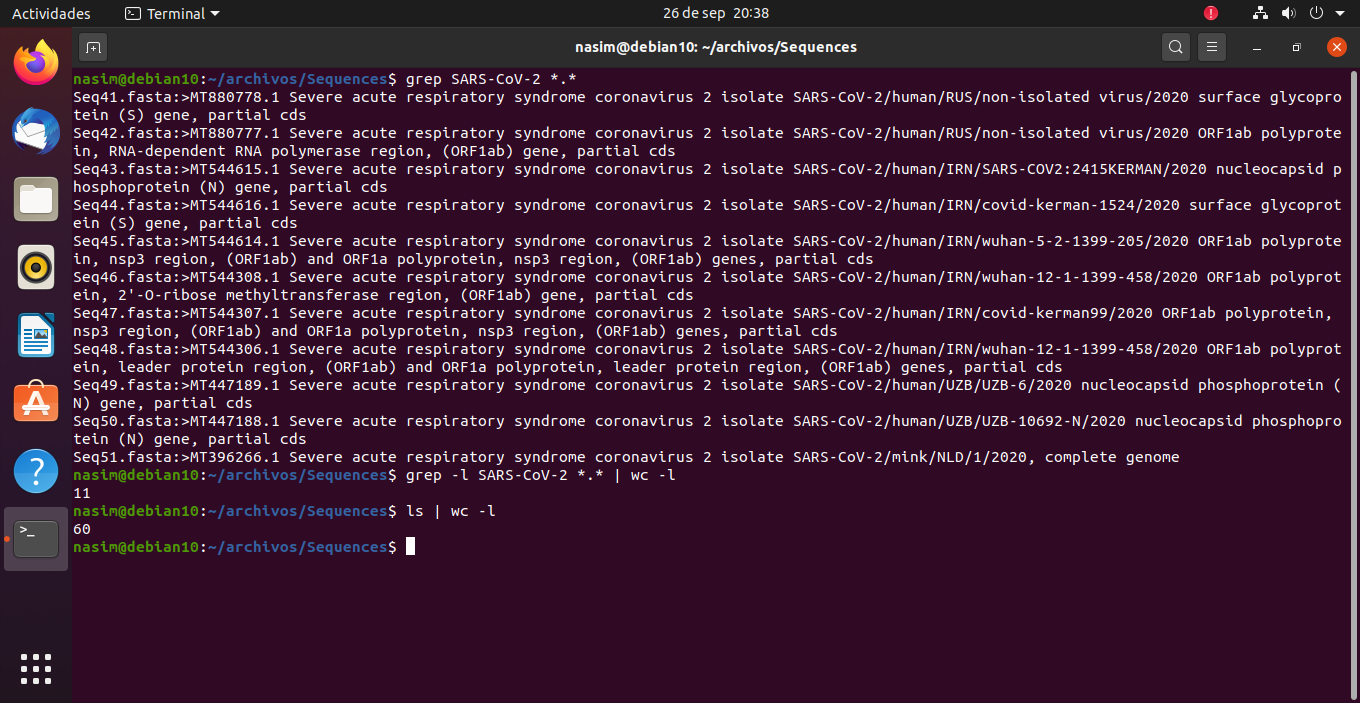


Mientras que el proceso bash muestra diferentes PID, el proceso init no devolvió ningún resultado. Independientemente de lo mostrado, sería razonable que el proceso init tuviera una PID menor, cercana a 1, ya que es uno de los procesos que primero se inician.

12. Los archivos FASTA son archivos que contienen secuencias de aminoácidos y por su estructura son fácilmente manipulables por editores de texto. Contienen un símbolo ‘>’ seguido de información de referencia respecto a la secuencia, lo cual corresponde a la cabecera, luego la secuencia de aminoácidos codificada por sus respectivas letras. Un formato multi-FASTA, no es más que múltiples secuencias FASTA concatenadas.

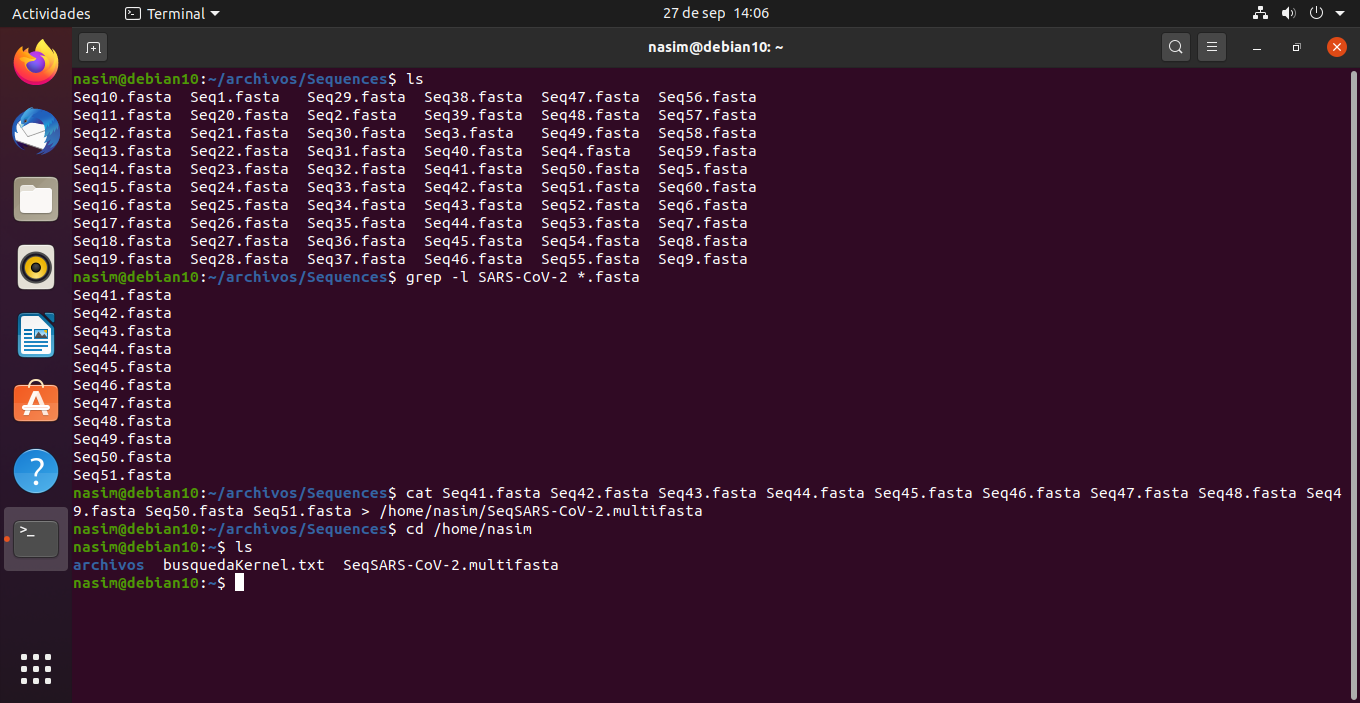
13. Genbank es una base de datos genética, es parte de “International Nucleotide Sequence Database Collaboration”, la cual la integran la base de datos de ADN de Japón, el laboratorio Europeo de Biología Molecular y GenBank.

14.a) b)



La cantidad de archivos que contienen secuencias de SARS-CoV-2 son 11, el total de archivos es 60, por la cual las secuencias de SARS-CoV-2 representan el 18,33%.

c) d)



El proceso para obtener un listado de los archivos que contienen datos de SARS-CoV-2 consiste en un grep con un parámetro –l para que muestre solamente el nombre de los archivos. Posteriormente, a estos archivos se los concatena con el comando cat y se los concatena en un nuevo archivo de extensión multifasta.

# Bibliografía

GenBank. (26 de 09 de 2020). *GenBank*. Obtenido de GenBank: https://www.ncbi.nlm.nih.gov/genbank/