



Facultad de
Ingeniería

Arquitecturas y Sistemas Operativos

Año: 2020

“Introducción a Linux para Bioinformática”

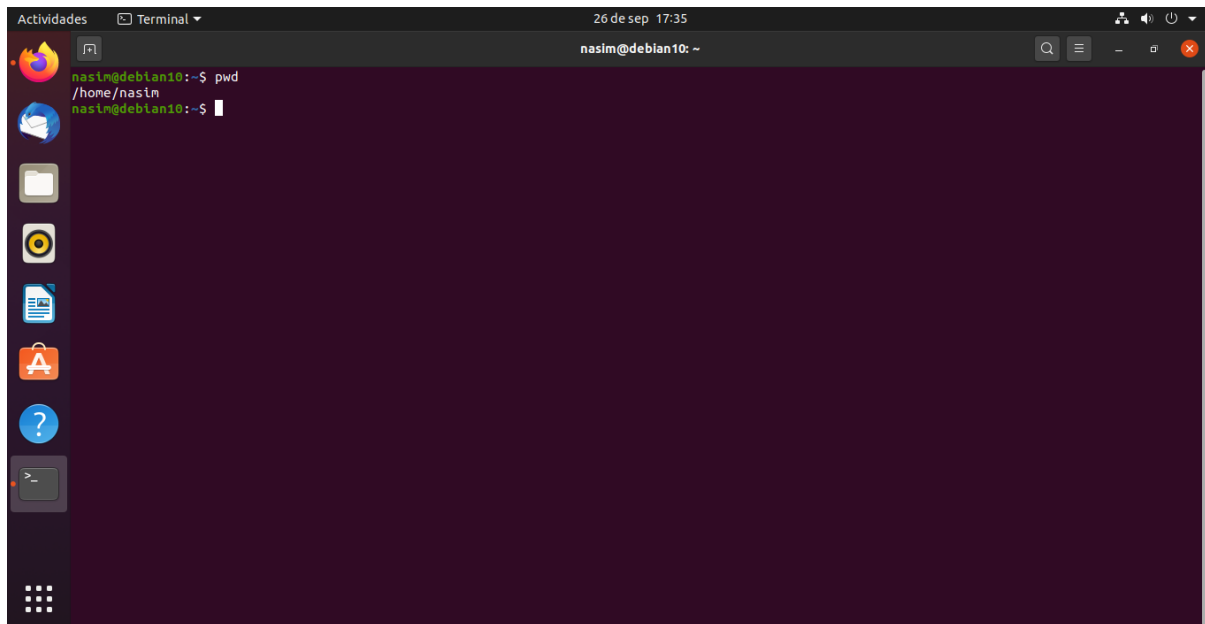
Alumno: Salim Taleb, Nasim Anibal

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.

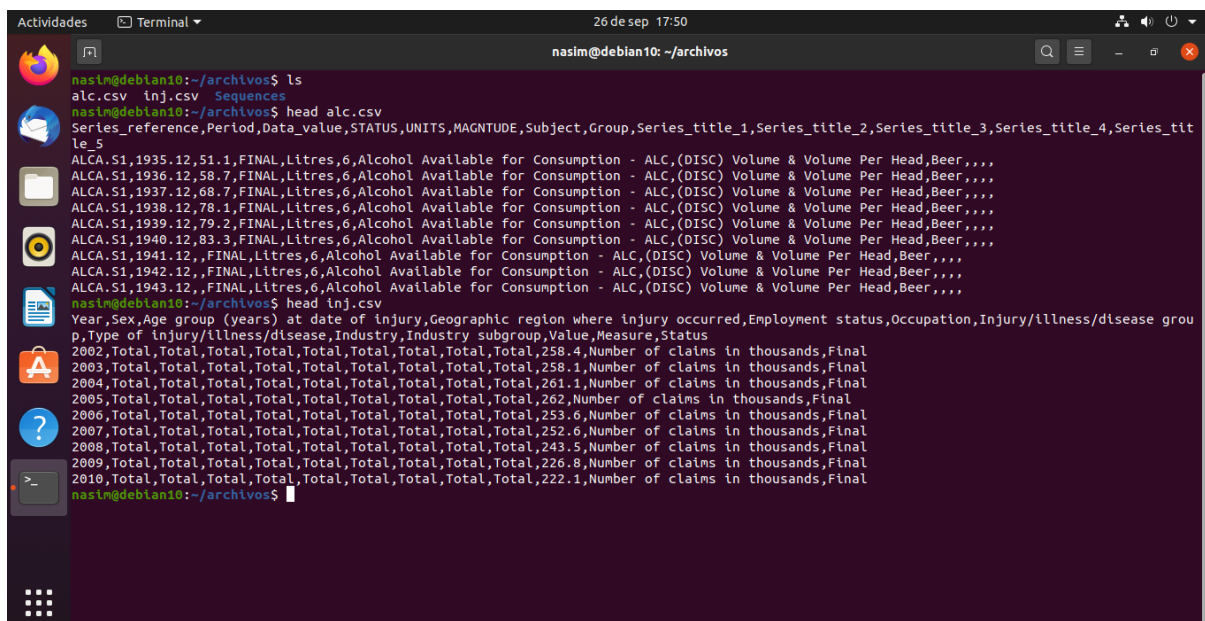


A terminal window titled 'Terminal' with a dark background. The prompt is 'nasim@debian10: ~'. The user has entered the command 'pwd' and the output is '/home/nasim'. The terminal window is part of a desktop environment with a sidebar on the left containing various application icons.

```
nasim@debian10:~$ pwd
/home/nasim
nasim@debian10:~$
```

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.

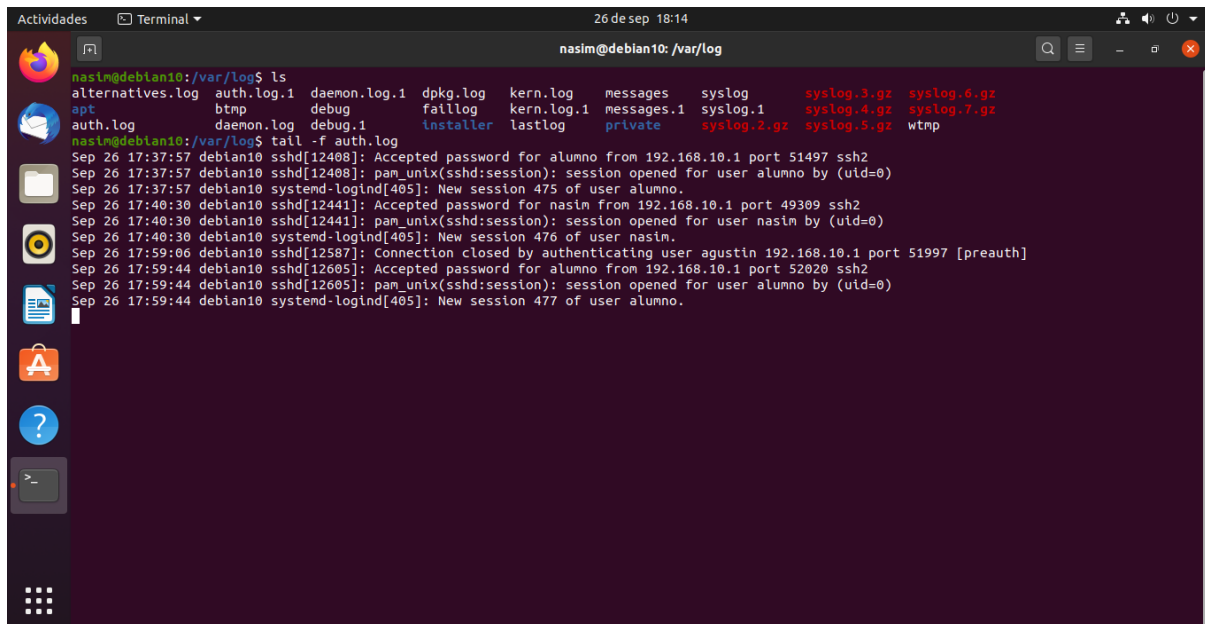


A terminal window titled 'Terminal' with a dark background. The prompt is 'nasim@debian10: ~/archivos'. The user has entered the command 'ls' and the output is 'alc.csv inj.csv Sequences'. The user then enters 'head alc.csv' and the output shows the first few lines of the 'alc.csv' file. The terminal window is part of a desktop environment with a sidebar on the left containing various application icons.

```
nasim@debian10:~/archivos$ ls
alc.csv inj.csv Sequences
nasim@debian10:~/archivos$ head alc.csv
Series_reference,Period,Data_value,STATUS,UNITS,MAGNTUDE,Subject,Group,Series_title_1,Series_title_2,Series_title_3,Series_title_4,Series_title_5
ALCA.S1,1935.12,51.1,FINAL,Litres,6,Alcohol Available for Consumption - ALC,(DISC) Volume & Volume Per Head,Beer,,,,
ALCA.S1,1936.12,58.7,FINAL,Litres,6,Alcohol Available for Consumption - ALC,(DISC) Volume & Volume Per Head,Beer,,,,
ALCA.S1,1937.12,68.7,FINAL,Litres,6,Alcohol Available for Consumption - ALC,(DISC) Volume & Volume Per Head,Beer,,,,
ALCA.S1,1938.12,78.1,FINAL,Litres,6,Alcohol Available for Consumption - ALC,(DISC) Volume & Volume Per Head,Beer,,,,
ALCA.S1,1939.12,79.2,FINAL,Litres,6,Alcohol Available for Consumption - ALC,(DISC) Volume & Volume Per Head,Beer,,,,
ALCA.S1,1940.12,83.3,FINAL,Litres,6,Alcohol Available for Consumption - ALC,(DISC) Volume & Volume Per Head,Beer,,,,
ALCA.S1,1941.12,,FINAL,Litres,6,Alcohol Available for Consumption - ALC,(DISC) Volume & Volume Per Head,Beer,,,,
ALCA.S1,1942.12,,FINAL,Litres,6,Alcohol Available for Consumption - ALC,(DISC) Volume & Volume Per Head,Beer,,,,
ALCA.S1,1943.12,,FINAL,Litres,6,Alcohol Available for Consumption - ALC,(DISC) Volume & Volume Per Head,Beer,,,,
nasim@debian10:~/archivos$ head inj.csv
Year,Sex,Age group (years) at date of injury,Geographic region where injury occurred,Employment status,Occupation,Injury/illness/disease group,Type of injury/illness/disease,Industry,Industry subgroup,Value,Measure,Status
2002,Total,Total,Total,Total,Total,Total,Total,Total,Total,Total,258.4,Number of claims in thousands,Final
2003,Total,Total,Total,Total,Total,Total,Total,Total,Total,Total,258.1,Number of claims in thousands,Final
2004,Total,Total,Total,Total,Total,Total,Total,Total,Total,Total,261.1,Number of claims in thousands,Final
2005,Total,Total,Total,Total,Total,Total,Total,Total,Total,Total,262,Number of claims in thousands,Final
2006,Total,Total,Total,Total,Total,Total,Total,Total,Total,Total,253.6,Number of claims in thousands,Final
2007,Total,Total,Total,Total,Total,Total,Total,Total,Total,Total,252.6,Number of claims in thousands,Final
2008,Total,Total,Total,Total,Total,Total,Total,Total,Total,Total,243.5,Number of claims in thousands,Final
2009,Total,Total,Total,Total,Total,Total,Total,Total,Total,Total,226.8,Number of claims in thousands,Final
2010,Total,Total,Total,Total,Total,Total,Total,Total,Total,Total,222.1,Number of claims in thousands,Final
nasim@debian10:~/archivos$
```

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.

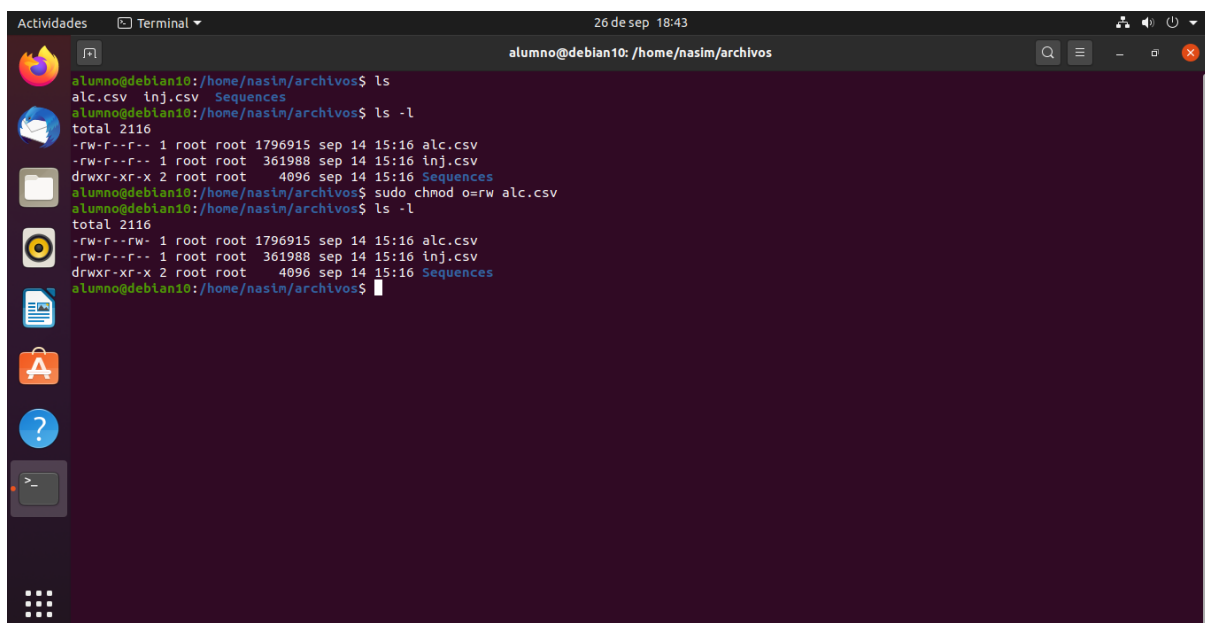
3. El comando tail seguido de -f permite monitorizar en tiempo real el final de un archivo. Puede ser utilizado como en este caso, para visualizar archivos de registros en tiempo real a medida que los usuarios ingresan y salen del sistema. Introduciendo el comando se observa lo siguiente:



```
nasim@debian10: /var/log$ ls
alternatives.log  auth.log.1  daemon.log.1  dpkg.log    kern.log    messages    syslog      syslog.3.gz  syslog.6.gz
apt              btmpt      debug         faillog     kern.log.1  messages.1  syslog.1    syslog.4.gz  syslog.7.gz
auth.log         daemon.log  debug.1       installer   lastlog     private     syslog.2.gz syslog.5.gz  wtmp

nasim@debian10: /var/log$ tail -f auth.log
Sep 26 17:37:57 debian10 sshd[12408]: Accepted password for alumno from 192.168.10.1 port 51497 ssh2
Sep 26 17:37:57 debian10 sshd[12408]: pam_unix(sshd:session): session opened for user alumno by (uid=0)
Sep 26 17:37:57 debian10 systemd-logind[405]: New session 475 of user alumno.
Sep 26 17:40:30 debian10 sshd[12441]: Accepted password for nasim from 192.168.10.1 port 49309 ssh2
Sep 26 17:40:30 debian10 sshd[12441]: pam_unix(sshd:session): session opened for user nasim by (uid=0)
Sep 26 17:40:30 debian10 systemd-logind[405]: New session 476 of user nasim.
Sep 26 17:59:06 debian10 sshd[12587]: Connection closed by authenticating user agustin 192.168.10.1 port 51997 [preauth]
Sep 26 17:59:44 debian10 sshd[12605]: Accepted password for alumno from 192.168.10.1 port 52020 ssh2
Sep 26 17:59:44 debian10 sshd[12605]: pam_unix(sshd:session): session opened for user alumno by (uid=0)
Sep 26 17:59:44 debian10 systemd-logind[405]: New session 477 of user alumno.
```

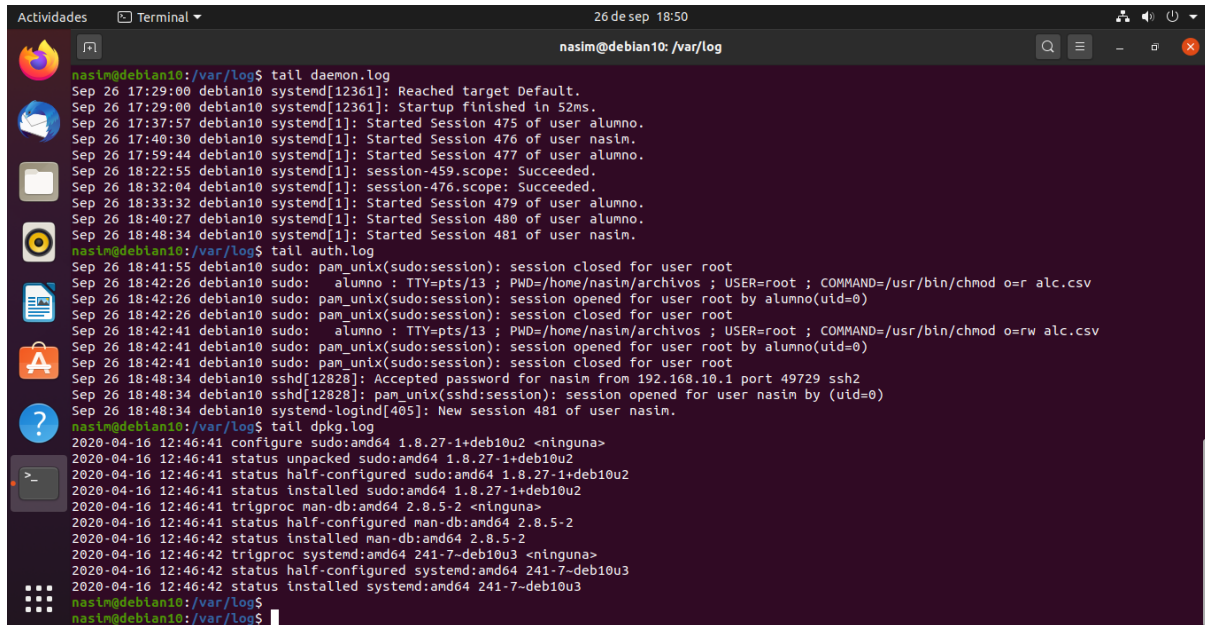
4. La secuencia de comandos para dicha acción será:



```
alumno@debian10: /home/nasim/archivos$ ls
alc.csv  inj.csv  Sequences
alumno@debian10: /home/nasim/archivos$ ls -l
total 2116
-rw-r--r-- 1 root root 1796915 sep 14 15:16 alc.csv
-rw-r--r-- 1 root root 361988 sep 14 15:16 inj.csv
drwxr-xr-x 2 root root 4096 sep 14 15:16 Sequences
alumno@debian10: /home/nasim/archivos$ sudo chmod o=rw alc.csv
alumno@debian10: /home/nasim/archivos$ ls -l
total 2116
-rw-r--r-- 1 root root 1796915 sep 14 15:16 alc.csv
-rw-r--r-- 1 root root 361988 sep 14 15:16 inj.csv
drwxr-xr-x 2 root root 4096 sep 14 15:16 Sequences
alumno@debian10: /home/nasim/archivos$
```

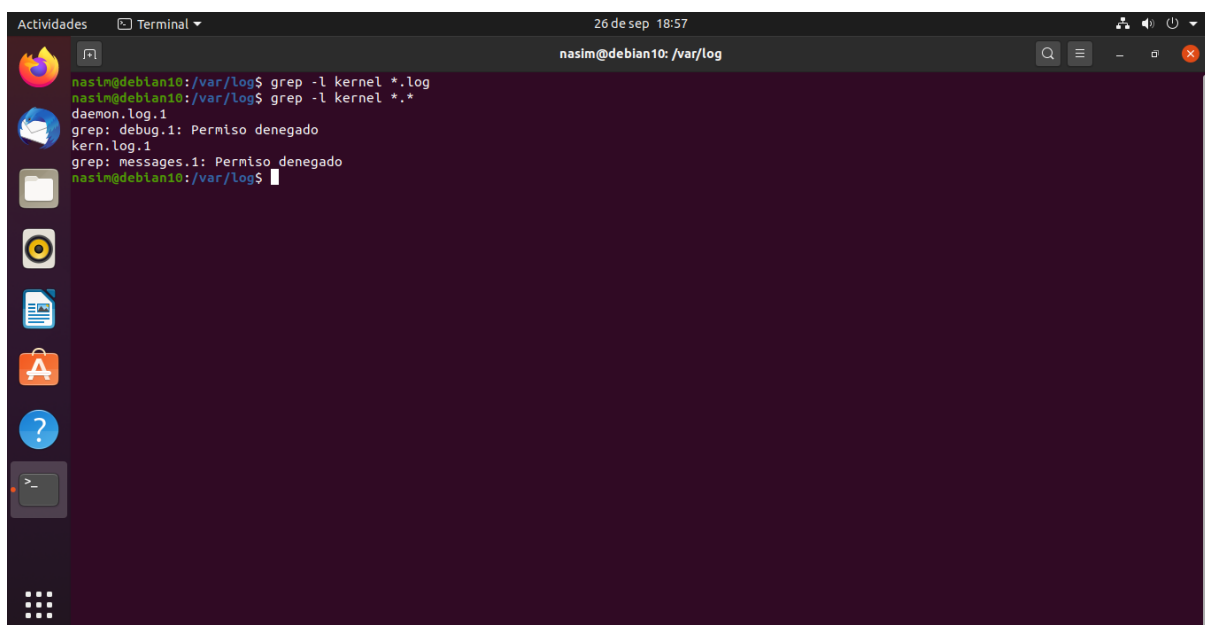
Para este punto se debe ingresar mediante un usuario root ya que el usuario “nasim” no poseía los permisos para realizar la modificación de permisos del archivo. Primero se muestran los archivos y posteriormente se agrega -l para ver en más detalle cada uno, para modificar los permisos usamos chmod, 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:

A terminal window titled 'nasim@debian10: /var/log' showing the output of 'tail daemon.log' and 'tail auth.log'. The 'tail daemon.log' output shows system boot logs and session starts for user 'alumno'. The 'tail auth.log' output shows sudo sessions for 'alumno' and an ssh session for 'nasim'.

```
nasim@debian10:/var/log$ tail daemon.log
Sep 26 17:29:00 debian10 systemd[12361]: Reached target Default.
Sep 26 17:29:00 debian10 systemd[12361]: Startup finished in 52ms.
Sep 26 17:37:57 debian10 systemd[1]: Started Session 475 of user alumno.
Sep 26 17:40:30 debian10 systemd[1]: Started Session 476 of user nasim.
Sep 26 17:59:44 debian10 systemd[1]: Started Session 477 of user alumno.
Sep 26 18:22:55 debian10 systemd[1]: session-459.scope: Succeeded.
Sep 26 18:32:04 debian10 systemd[1]: session-476.scope: Succeeded.
Sep 26 18:33:32 debian10 systemd[1]: Started Session 479 of user alumno.
Sep 26 18:40:27 debian10 systemd[1]: Started Session 480 of user alumno.
Sep 26 18:48:34 debian10 systemd[1]: Started Session 481 of user nasim.
nasim@debian10:/var/log$ tail auth.log
Sep 26 18:41:55 debian10 sudo: pam_unix(sudo:session): session closed for user root
Sep 26 18:42:26 debian10 sudo: alumno : TTY=pts/13 ; PWD=/home/nasim/archivos ; USER=root ; COMMAND=/usr/bin/chmod o=r alc.csv
Sep 26 18:42:26 debian10 sudo: pam_unix(sudo:session): session opened for user root by alumno(uid=0)
Sep 26 18:42:26 debian10 sudo: pam_unix(sudo:session): session closed for user root
Sep 26 18:42:41 debian10 sudo: alumno : TTY=pts/13 ; PWD=/home/nasim/archivos ; USER=root ; COMMAND=/usr/bin/chmod o=rw alc.csv
Sep 26 18:42:41 debian10 sudo: pam_unix(sudo:session): session opened for user root by alumno(uid=0)
Sep 26 18:42:41 debian10 sudo: pam_unix(sudo:session): session closed for user root
Sep 26 18:48:34 debian10 sshd[12828]: Accepted password for nasim from 192.168.10.1 port 49729 ssh2
Sep 26 18:48:34 debian10 sshd[12828]: pam_unix(sshd:session): session opened for user nasim by (uid=0)
Sep 26 18:48:34 debian10 systemd-logind[405]: New session 481 of user nasim.
nasim@debian10:/var/log$ tail dpkg.log
2020-04-16 12:46:41 configure sudo:amd64 1.8.27-1+deb10u2 <ninguna>
2020-04-16 12:46:41 status unpacked sudo:amd64 1.8.27-1+deb10u2
2020-04-16 12:46:41 status half-configured sudo:amd64 1.8.27-1+deb10u2
2020-04-16 12:46:41 status installed sudo:amd64 1.8.27-1+deb10u2
2020-04-16 12:46:41 trigproc man-db:amd64 2.8.5-2 <ninguna>
2020-04-16 12:46:41 status half-configured man-db:amd64 2.8.5-2
2020-04-16 12:46:42 status installed man-db:amd64 2.8.5-2
2020-04-16 12:46:42 trigproc systemd:amd64 241-7-deb10u3 <ninguna>
2020-04-16 12:46:42 status half-configured systemd:amd64 241-7-deb10u3
2020-04-16 12:46:42 status installed systemd:amd64 241-7-deb10u3
nasim@debian10:/var/log$
nasim@debian10:/var/log$
```

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:

A terminal window titled 'nasim@debian10: /var/log' showing the output of 'grep -l kernel *.log' and 'grep -l kernel *.*'. The output shows 'daemon.log.1' and 'kern.log.1' as files containing the word 'kernel'.

```
nasim@debian10:/var/log$ grep -l kernel *.log
nasim@debian10:/var/log$ grep -l kernel *.*
daemon.log.1
grep: debug.1: Permiso denegado
kern.log.1
grep: messages.1: Permiso denegado
nasim@debian10:/var/log$
```

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.

The image shows a Linux desktop environment with a terminal window open. The terminal title bar indicates the date and time as '26 de sep 19:17'. The prompt is 'nasim@debian10: ~'. The terminal output displays system statistics and a list of running processes.

System Statistics:

```
top - 19:17:34 up 12 days, 22:14, 16 users,  load average: 0,00, 0,00, 0,00
Tasks: 154 total,   1 running, 152 sleeping,   1 stopped,   0 zombie
%Cpu(s):  0,0 us,  0,3 sy,  0,0 ni, 99,7 id,  0,0 wa,  0,0 hi,  0,0 si,  0,0 st
MiB Mem : 3947,0 total, 3450,3 free,   145,7 used,   350,9 buff/cache
MiB Swap: 4094,0 total, 4094,0 free,    0,0 used. 3568,2 avail Mem
```

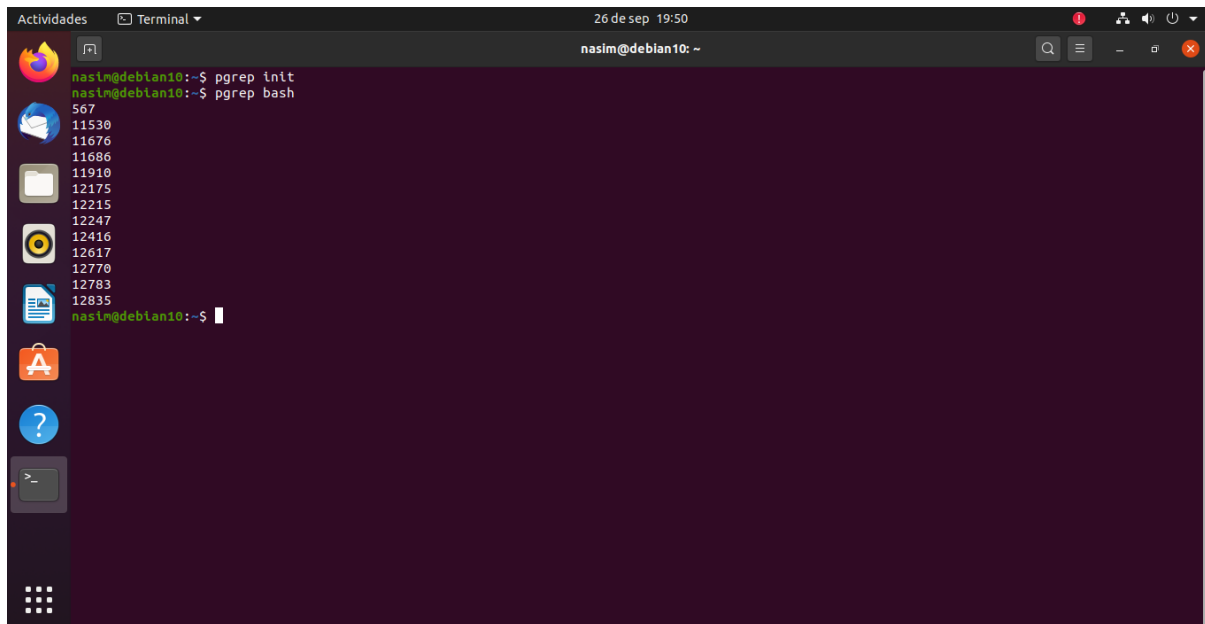
Process List:

PID	USER	PR	NI	VIRT	RES	SHR	S	%CPU	%MEM	TIME+	COMMAND
12361	nasim	20	0	21024	8384	7256	S	0,0	0,2	0:00.03	systemd
12363	nasim	20	0	170784	2464	4	S	0,0	0,1	0:00.00	(sd-pam)
12371	nasim	20	0	16784	5152	4060	S	0,0	0,1	0:00.02	sshd
12372	nasim	20	0	7652	4452	3184	S	0,0	0,1	0:00.04	bash
12834	nasim	20	0	17472	6796	4996	S	0,0	0,2	0:00.56	sshd
12835	nasim	20	0	7924	4784	3500	S	0,0	0,1	0:00.15	bash
12925	nasim	20	0	11092	3648	3088	R	0,0	0,1	0:00.03	top

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:



```
nasim@debian10:~$ pgrep init
nasim@debian10:~$ pgrep bash
567
11530
11676
11686
11910
12175
12215
12247
12416
12617
12770
12783
12835
nasim@debian10:~$
```

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)

```
Actividades Terminal 26 de sep 20:38
nasim@debian10: ~/archivos/Sequences
nasim@debian10:~/archivos/Sequences$ grep SARS-CoV-2 *.*
Seq41.fasta:MT880778.1 Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/RUS/non-isolated virus/2020 surface glycoprotein (S) gene, partial cds
Seq42.fasta:MT880777.1 Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/RUS/non-isolated virus/2020 ORF1ab polyprotein, RNA-dependent RNA polymerase region, (ORF1ab) gene, partial cds
Seq43.fasta:MT544615.1 Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/IRN/SARS-COV2:2415KERMAN/2020 nucleocapsid phosphoprotein (N) gene, partial cds
Seq44.fasta:MT544616.1 Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/IRN/covid-kerman-1524/2020 surface glycoprotein (S) gene, partial cds
Seq45.fasta:MT544614.1 Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/IRN/wuhan-5-2-1399-205/2020 ORF1ab polyprotein, ns3 region, (ORF1ab) and ORF1a polyprotein, ns3 region, (ORF1ab) genes, partial cds
Seq46.fasta:MT544308.1 Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/IRN/wuhan-12-1-1399-458/2020 ORF1ab polyprotein, 2'-O-ribose methyltransferase region, (ORF1ab) gene, partial cds
Seq47.fasta:MT544307.1 Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/IRN/covid-kerman99/2020 ORF1ab polyprotein, ns3 region, (ORF1ab) and ORF1a polyprotein, ns3 region, (ORF1ab) genes, partial cds
Seq48.fasta:MT544306.1 Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/IRN/wuhan-12-1-1399-458/2020 ORF1ab polyprotein, leader protein region, (ORF1ab) and ORF1a polyprotein, leader protein region, (ORF1ab) genes, partial cds
Seq49.fasta:MT447189.1 Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/UZB/UZB-6/2020 nucleocapsid phosphoprotein (N) gene, partial cds
Seq50.fasta:MT447188.1 Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/UZB/UZB-10692-N/2020 nucleocapsid phosphoprotein (N) gene, partial cds
Seq51.fasta:MT396266.1 Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/mink/NLD/1/2020, complete genome
nasim@debian10:~/archivos/Sequences$ grep -l SARS-CoV-2 *.* | wc -l
11
nasim@debian10:~/archivos/Sequences$ ls | wc -l
60
nasim@debian10:~/archivos/Sequences$
```

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)

```
Actividades Terminal 27 de sep 14:06
nasim@debian10: ~
nasim@debian10:~/archivos/Sequences$ ls
Seq10.fasta Seq1.fasta Seq29.fasta Seq38.fasta Seq47.fasta Seq56.fasta
Seq11.fasta Seq20.fasta Seq2.fasta Seq39.fasta Seq48.fasta Seq57.fasta
Seq12.fasta Seq21.fasta Seq30.fasta Seq3.fasta Seq49.fasta Seq58.fasta
Seq13.fasta Seq22.fasta Seq31.fasta Seq40.fasta Seq4.fasta Seq59.fasta
Seq14.fasta Seq23.fasta Seq32.fasta Seq41.fasta Seq50.fasta Seq5.fasta
Seq15.fasta Seq24.fasta Seq33.fasta Seq42.fasta Seq51.fasta Seq60.fasta
Seq16.fasta Seq25.fasta Seq34.fasta Seq43.fasta Seq52.fasta Seq6.fasta
Seq17.fasta Seq26.fasta Seq35.fasta Seq44.fasta Seq53.fasta Seq7.fasta
Seq18.fasta Seq27.fasta Seq36.fasta Seq45.fasta Seq54.fasta Seq8.fasta
Seq19.fasta Seq28.fasta Seq37.fasta Seq46.fasta Seq55.fasta Seq9.fasta
nasim@debian10:~/archivos/Sequences$ grep -l SARS-CoV-2 *.fasta
Seq41.fasta
Seq42.fasta
Seq43.fasta
Seq44.fasta
Seq45.fasta
Seq46.fasta
Seq47.fasta
Seq48.fasta
Seq49.fasta
Seq50.fasta
Seq51.fasta
nasim@debian10:~/archivos/Sequences$ cat Seq41.fasta Seq42.fasta Seq43.fasta Seq44.fasta Seq45.fasta Seq46.fasta Seq47.fasta Seq48.fasta Seq49.fasta Seq50.fasta Seq51.fasta > /home/nasim/SeqSARS-CoV-2.multifasta
nasim@debian10:~/archivos/Sequences$ cd /home/nasim
nasim@debian10:~$ ls
archivos  busquedaKernel.txt  SeqSARS-CoV-2.multifasta
nasim@debian10:~$
```

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/>