

ACTIVIDAD PRÁCTICA**“Introducción a Linux para Bioinformática”****Introducción:**

La Bioinformática es una ciencia que se basa en el uso de las computadoras para analizar y resolver problemas biológicos. Estos últimos están caracterizados en la actualidad por su volumen de información y su creciente complejidad. En ese marco, un profesional bioinformático debe lidiar cotidianamente con altos requerimientos computacionales, administrar sistemas e implementar soluciones para el procesamiento de cadenas, y lograr una interconexión de flujos de datos entre diferentes aplicaciones, en tiempo real, que permitan resolver determinadas situaciones. Los sistemas bioinformáticos se caracterizan por disponer fundamentalmente de dos interfaces de usuario típicas, la línea de comandos y plataformas de servicios web. Resulta evidente la necesidad de adquirir las habilidades para comprender su funcionamiento y la práctica para interactuar fluidamente con tales sistemas e interfaces.

Modalidad:

Debido a las restricciones propias de la pandemia de COVID-19, deberá acceder a un sistema Linux remoto que se está ejecutando en la FI-UNER. Para tal efecto, la cátedra ha creado usuarios para cada alumno. Se deberá utilizar la herramienta ssh para conectarse en forma remota e interactuar con la consola del sistema.

Datos de Acceso:

IP: 200.61.251.77 Puerto: 77 Usuario y contraseña a suministrar por la cátedra.

El comando ssh permite establecer una comunicación segura (encriptada) entre dos ordenadores. Desde una terminal en su computadora ingrese:

```
ssh nombreUsuario@200.61.251.77 -p 77
```

Presentación:

Se deberá presentar un informe escrito en formato digital describiendo el desarrollo de las actividades, incluyendo capturas de pantalla, filmaciones breves sin sonido, etc.

Consultas disponibles hasta 24 hs antes de la fecha de entrega.

FECHA DE ENTREGA: 28/09/2020

Enunciados:

1. Identifique en donde se encuentra dentro del árbol de directorio del sistema.
2. Utilice los comandos de visualización de archivos `cat`, `less`, `head` y `tail`. para los archivos existentes dentro la carpeta `archivos`. ¿Qué tipo de información contiene cada archivo? ¿En qué se diferencia cada comando?
3. Explique cómo opera el comando `tail -f` observando el archivo `/var/log/auth.log` ¿Que observa si un usuario ingresa o sale del sistema? ¿En qué se diferencia `tail -f` de los comandos del ítem anterior? ¿En que casos lo utilizaría? * Nota: Puede abrir múltiples sesiones en el sistema con su usuario para experimentar.
4. Modifique los permisos del archivo `alc.csv` de modo que pueda ser escrito por cualquier usuario del sistema. Detalle el comando utilizado.
5. Visualice archivos de Log del sistema.
6. Determine qué archivos de extensión `.log` en la carpeta `"/var/log/"` que contengan la palabra "kernel"
7. Mediante entubamiento, cree un archivo que contenga la información generada en el punto anterior (las líneas de todos los archivos `.log` que contienen la palabra "kernel")
8. ¿Qué comando usaría para analizar los procesos en ejecución? Explique sus parámetros opcionales y la información de salida de los mismos.
9. Determine los procesos en ejecución de su usuario
10. Determine los procesos en ejecución de su usuario, y del usuario `root`
11. Determine los ID de proceso (PID) de los procesos `Bash` e `Init` con el comando `"pgrep"`. ¿Cuál es menor? ¿Es razonable?
12. Indague en la web sobre el formato de archivos `FASTA` y `Multi-FASTA`.
13. Indague en la web sobre el sistema `GenBank`. ¿Qué es? ¿De quién depende? ¿Qué datos podemos obtener del sistema?

14. En la carpeta "archivos/Sequences/" hay 60 secuencias nucleotídicas de 3 especies distintas de virus, en formato FASTA. Es necesario obtener un archivo multi-FASTA que contenga sólo las secuencias de Sars-coV-2 para luego alinearlas con herramientas como T-Coffee o ClustalW en búsqueda de motivos conservados (el alineamiento corresponde a esta actividad). Utilizando alguna/s palabra/s específica/s del encabezado de cada secuencia (primer renglón que comienza con un ">") se pueden obtener las secuencias de una especie particular.

Utilizando comandos de consola, pipes, lenguajes de scripting y/o de programación:

- a) Obtenga un listado de todos los encabezados correspondientes a Sars-coV-2
- b) Obtenga el conteo de todos los archivos que contengan datos correspondientes a Sars-coV-2. Calcule el porcentaje de secuencias Sars-coV-2 respecto del total.
- c) Obtenga un listado de todos los archivos que contengan datos correspondientes a Sars-coV-2
- d) Genere un archivo multi-FASTA que contenga los archivos FASTA correspondientes a Sars-coV-2