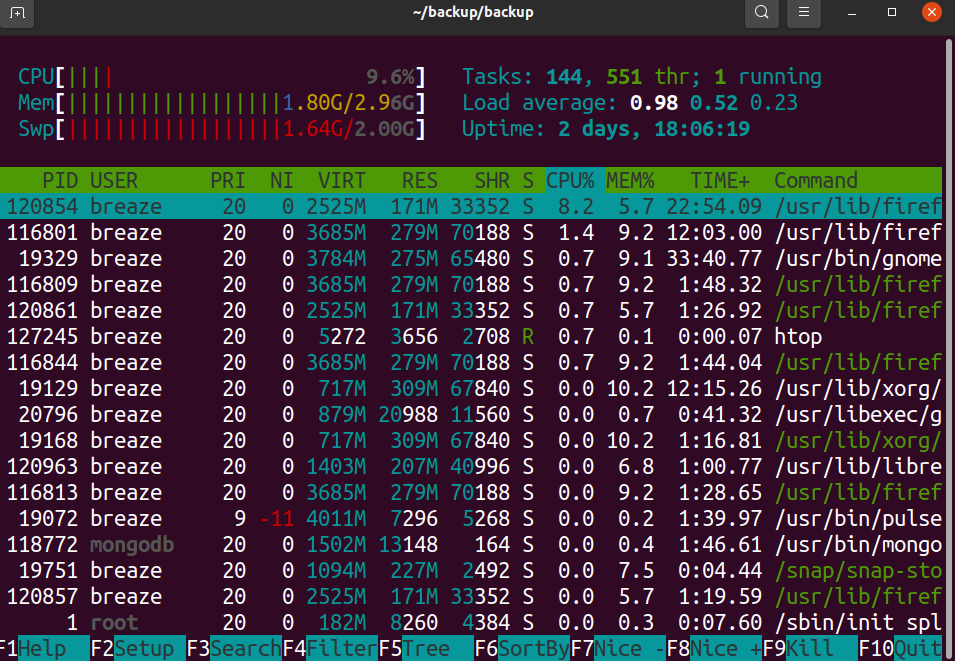
Willington Londoño

Juan Carlos Arango

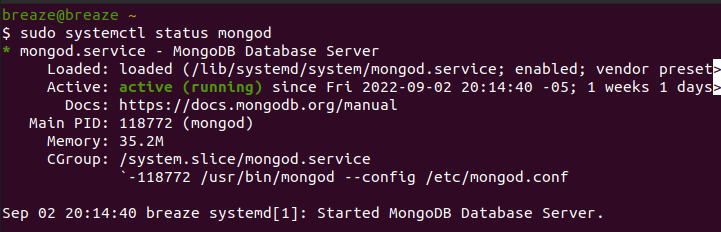
Simón Gaviria

Repositorio de la actividad: <https://github.com/breaze/biological_databases>

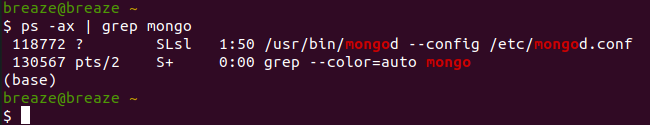
1. htop instalado

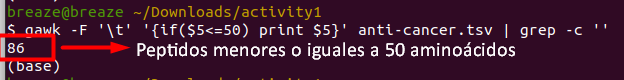


1. Mongo status:



ps -ax | grep mongo:



1. Información de proteínas y extracción:
2. Se encontraron 6186 proteínas
3. 196 están respaldadas por una revisión
4. Se encontraron 86 secuencias con péptidos menores o iguales a 50 aminoácidos, se adjunta comando con el cual se realizó el conteo

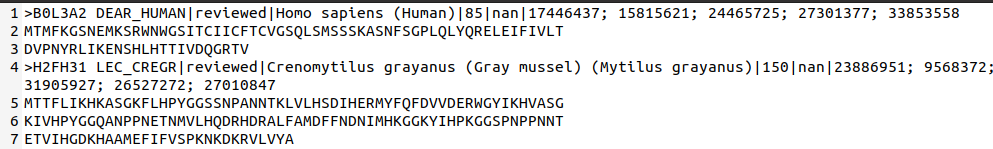
Se adjunta el script con el cual se genera el archivo fasta solicitado con todos los resultados de la búsqueda

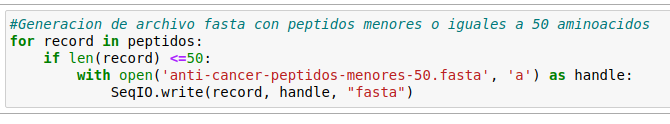


Este script puede ser reemplazado por el siguiente comando de awk:



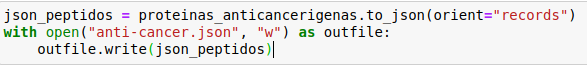
Ejemplos del archivo fasta generado con todos los resultados de la búsqueda:



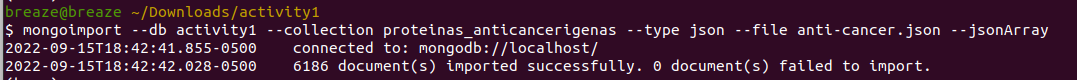
Se adjunta el script con el que se genera un archivo fasta con los péptidos menores o iguales a 50 aminoácidos, este archivo tiene la misma estructura que el archivo original:

1. Generar JSON e importar a mongodb

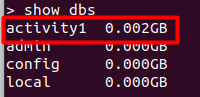
Se adjunta script para la creación de archivo json en formato de interés



Comando utilizado para la importación del json a la bd y colección específicas:



Podemos visualizar que la colección ahora cuenta con datos:



Algunas consultas útiles:

1. Consultar proteínas que han tenido algún soporte o revisión: 
2. Consultar péptidos con 50 o menos aminoácidos que que no pertenezcan a humanos:



1. Consultar todos los PubMed ID de proteínas relacionadas con humanos :



Finalmente se adjunta el archivo requirements.txt por si se desea replicar el ambiente de python para la ejecución de los scripts.

5. Presentación HDF5 vs PostgreSQL

<https://prezi.com/view/32ygK1t8NGUttETxdDUJ/>