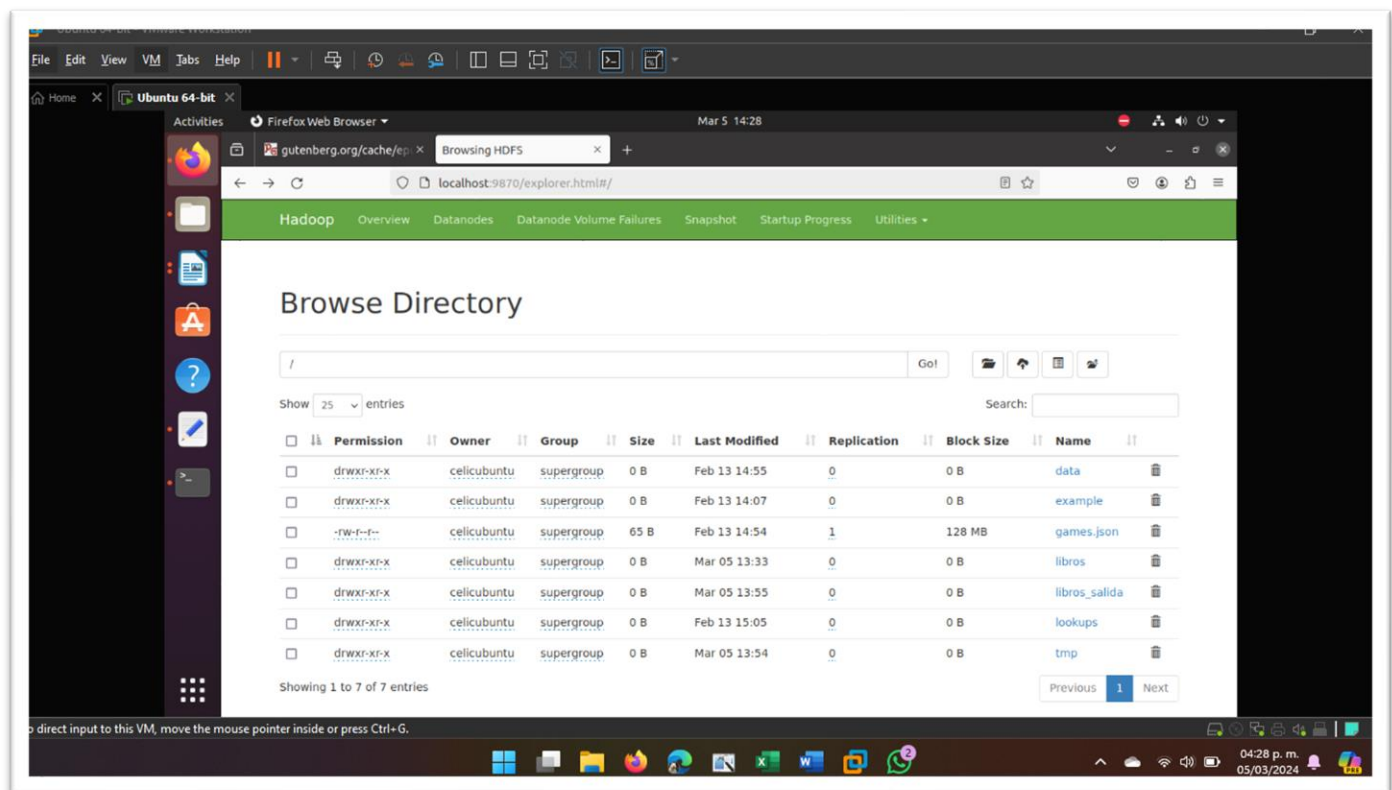
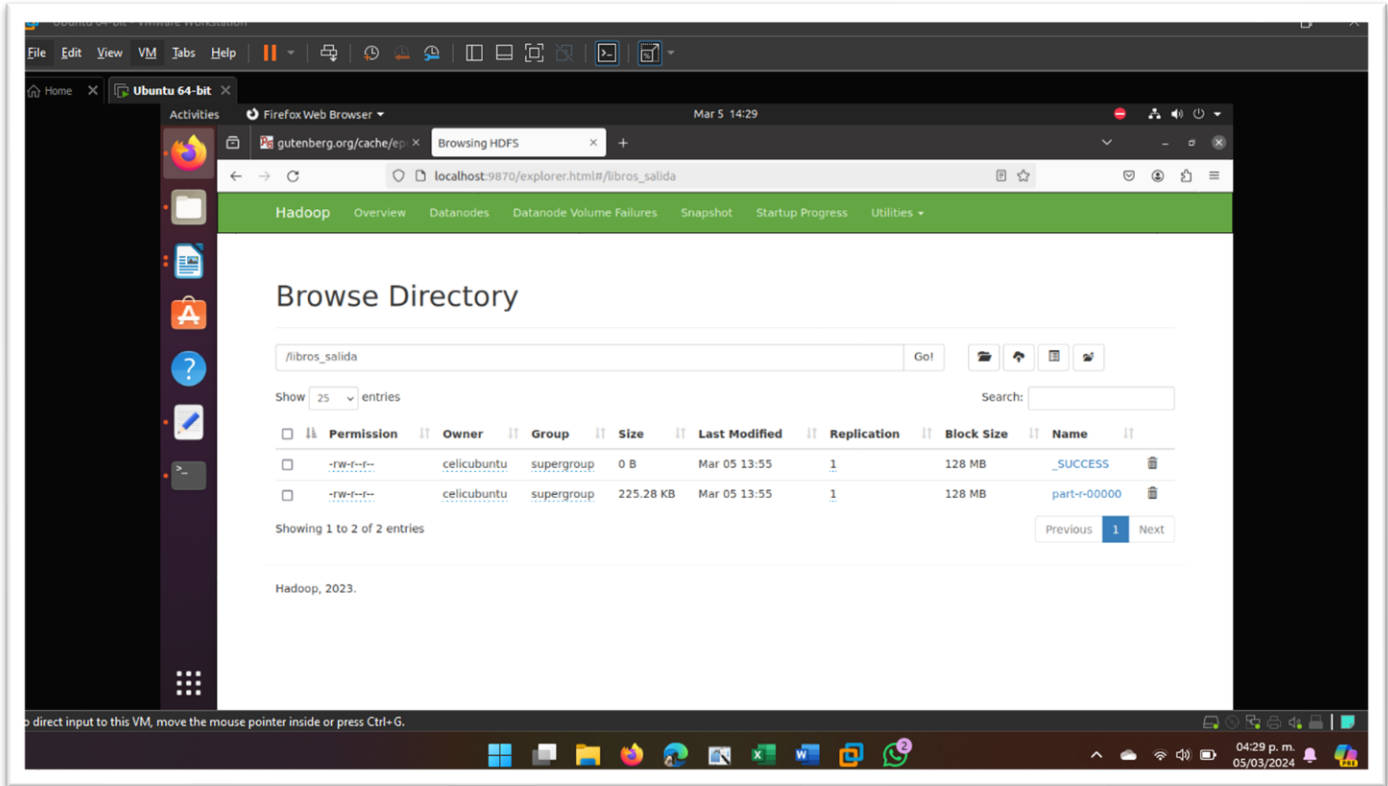
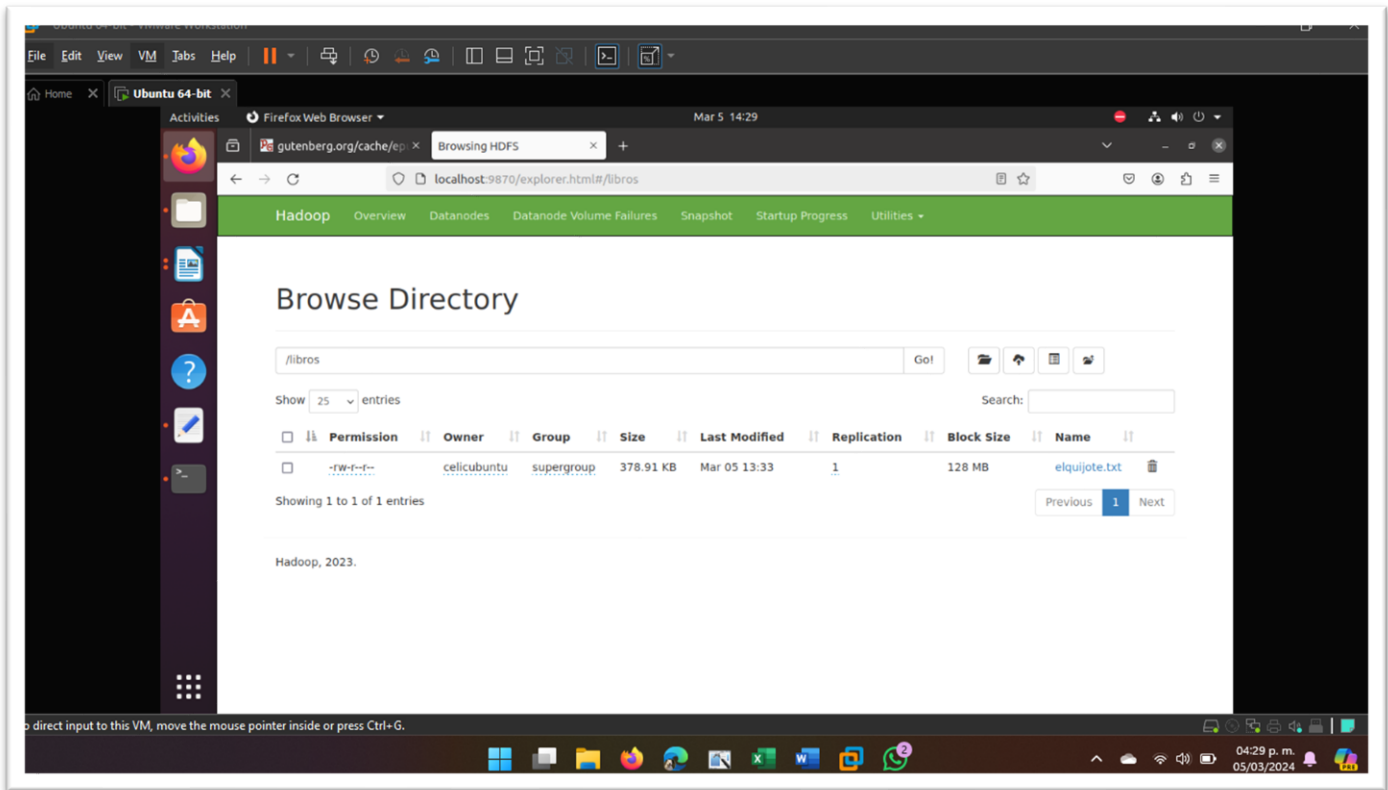
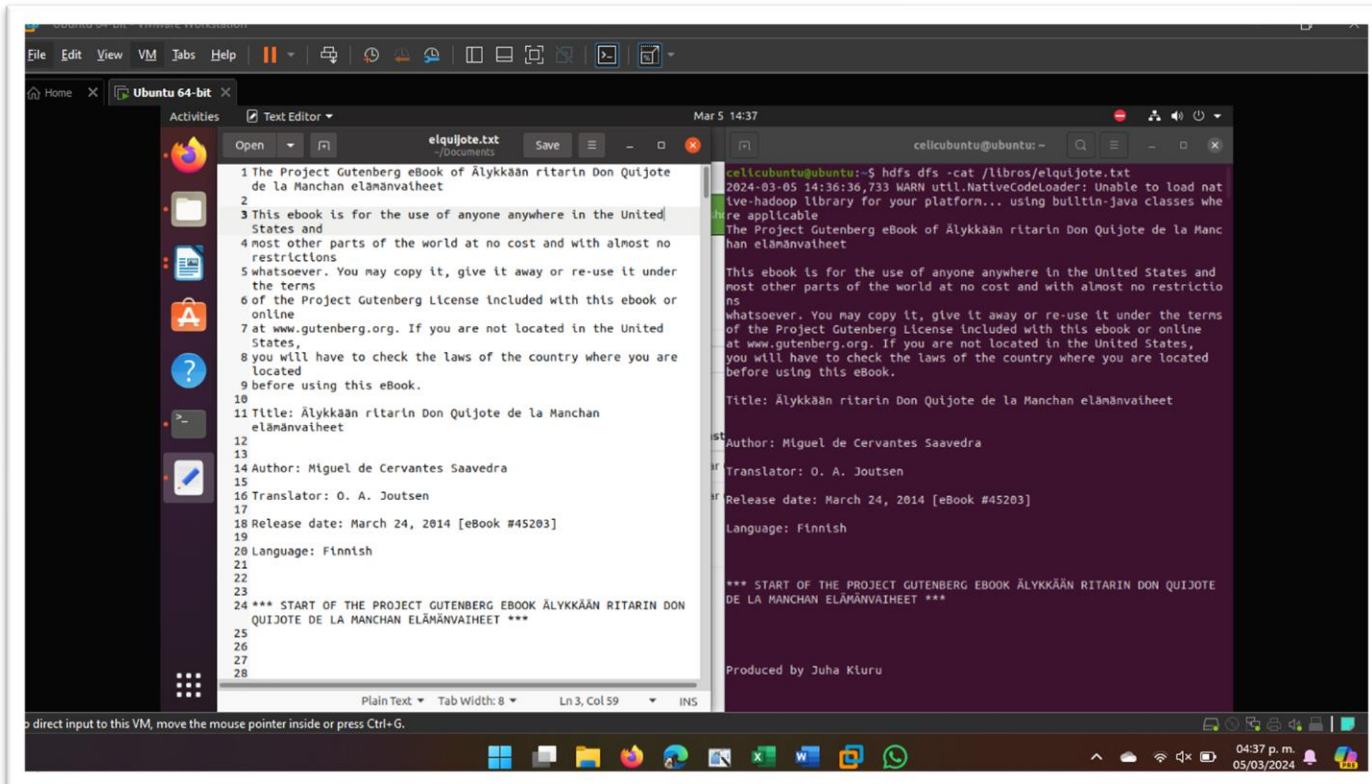
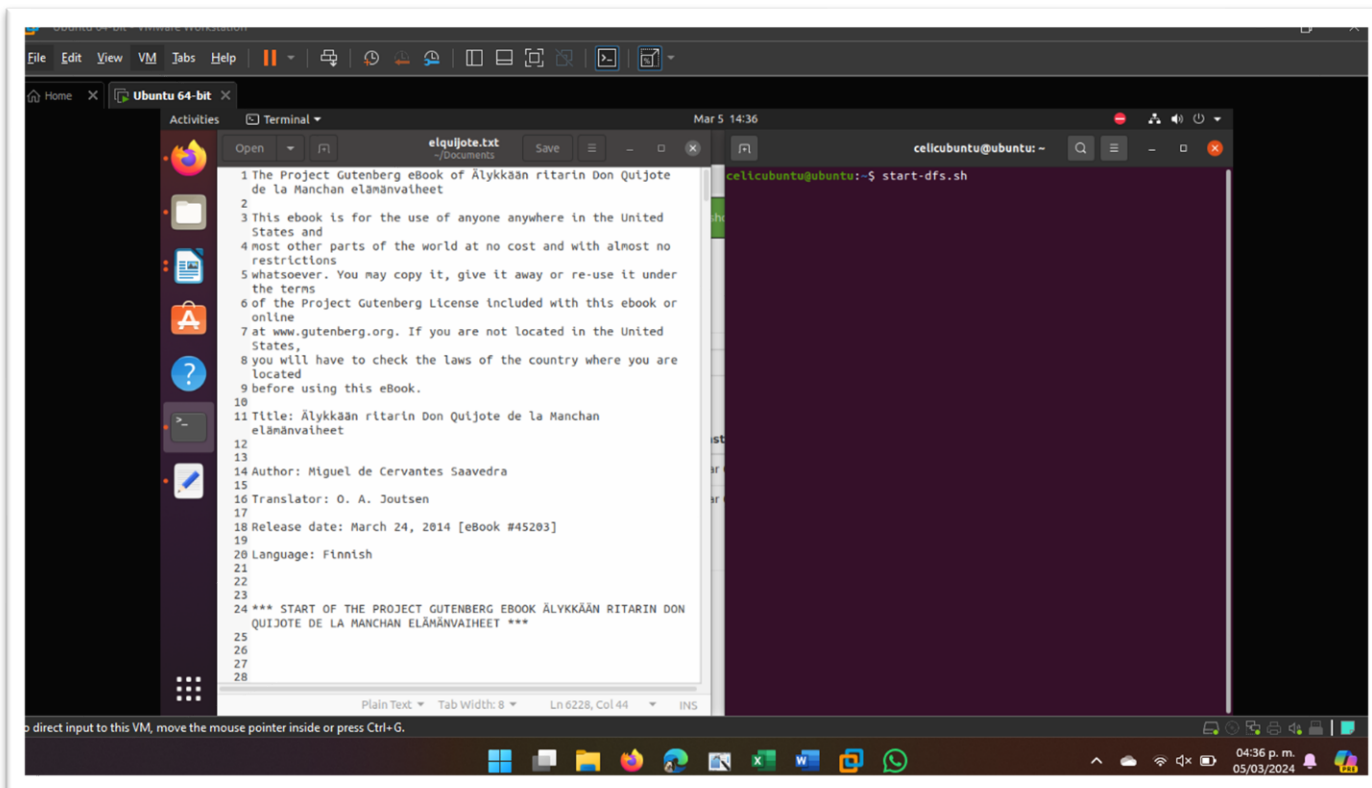


Nombre: Celic Gabriel Hernández Archundia.		Matrícula: 2877240
Nombre del curso: Infraestructura para Big Data	Nombre del profesor: Miguel de Jesús Martínez Felipe	
Módulo II	Actividad: Actividad 7	
Fecha: 05/03/2024		

Capturas de Ubuntu







Conclusión

En la actividad que hicimos aplicamos el concepto de *MapReduce*; el cual es un modelo de programación y un paradigma de procesamiento de datos que se utiliza especialmente en entornos distribuidos para realizar cálculos sobre grandes conjuntos de datos. En esta ocasión no se nos presenta una gran cantidad de datos, en cambio, utilizamos un escrito: *elquijote.txt*. Para aplicar el concepto de *MapReduce* tuvimos que acceder a nuestra máquina virtual y activar Hadoop por medio de ciertos códigos: `start-dfs.sh` para iniciar el sistema de archivos distribuidos de Hadoop y `start-yarn.sh` para iniciar el *framework* de recursos distribuidos YARN (Yet Another Resource Negotiator); en el que los componentes *ResourceManager* y *NodeManager* son activos, siendo *ResourceManager* el maestro del clúster que gestiona los recursos disponibles y asigna recursos a las aplicaciones que se ejecutan en este.

Posteriormente, por medio de comandos vistos anteriormente en otras entregas, creamos una carpeta en la que se guardaría el archivo de *elquijote.txt* y ejecutamos un comando que nos otorgaría los programas de *MapReduce* proporcionados por la versión 3.3.6 de Hadoop:

```
$ hadoop jar hadoop-mapreduce-examples-3.3.6.jar
```

El archivo que vamos a usar para aplicarle un programa de *MapReduce* fue a nuestro archivo de *elquijote.txt* con ayuda del siguiente comando:

```
$ hadoop jar hadoop-mapreduce-examples-3.3.6.jar wordcount /libros /libros_salida
```

- **Hadoop:** El comando principal utilizado para interactuar con el clúster de Hadoop
- **Jar:** Indica que se va a ejecutar un archivo JAR
- **Hadoop-mapreduce-examples-3.3.6.jar:** El archivo JAR que contiene los ejemplos de programas de *MapReduce*
- **Wordcount:** Este es el nombre del programa de *MapReduce* que se va a ejecutar; el cual cuenta la frecuencia de cada palabra en un conjunto de documentos
- **/libros:** Esta es la ruta de entrada en el sistema de archivos distribuido de Hadoop. Se supone que aquí se encuentran los datos de entrada para el programa *wordcount*.
- **/libros_salida:** Esta es la ruta de salida en el sistema de archivos distribuido de Hadoop. Se supone que aquí se almacenarán los resultados del procesamiento realizado por el programa *wordcount* (en caso de no existir, se creará uno).

Finalmente, obtuvimos el archivo que contenía “/libros_salida” y visualizamos el conteo que aplicamos con *wordcount* con el comando *more* y el nombre del archivo (“part-r-00000”, según recuerdo). Cuando ejecutamos el código pudimos ver el conteo de las palabras que aparecían en el escrito de *elquijote.txt*. Además de lo anterior, utilizamos el comando *cat* para ver todo el contenido del texto de una sola vez.

Hay cosas en esta actividad que aún no terminan de quedarme 100% claras, pero espero que en el futuro pueda practicar y estudiar más al respecto para poder encontrarle el sentido y la utilidad que poseen estos temas.