

Nombre: Pablo Elías Ramírez Escalante		Matrícula: AL02883894
Infraestructura para Big Data	Nombre del profesor: Miguel de Jesús Martínez Felipe	
Módulo 2	Actividad 7	
Fecha: 06/03/2024		
Bibliografía:		

1.- Ejecutar el commando

\$ jps

```
pablo@Ubuntu-BigData:~$ jps
3010 SecondaryNameNode
2834 DataNode
2698 NameNode
3327 Jps
```

2.- Después ejecutar

\$ stop-dfs.sh

```
pablo@Ubuntu-BigData:~$ stop-dfs.sh
Stopping namenodes on [localhost]
Stopping datanodes
Stopping secondary namenodes [Ubuntu-BigData]
```

3.- Ejecutar

\$ jps

```
pablo@Ubuntu-BigData:~$ jps
3808 Jps
```

4.- Ejecutar el siguiente comando

\$ start-dfs.sh

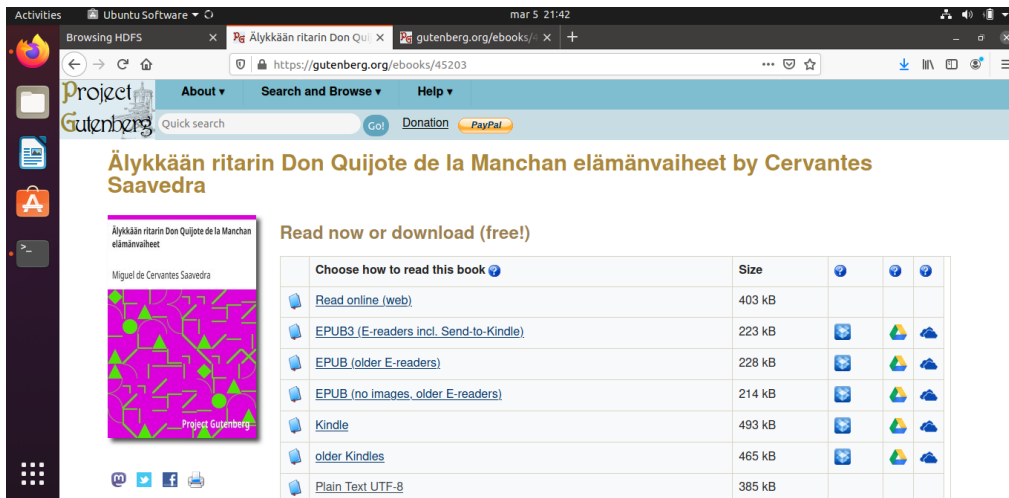
```
pablo@Ubuntu-BigData:~$ start-dfs.sh
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [Ubuntu-BigData]
```

5.- Ejecutar el siguiente comando

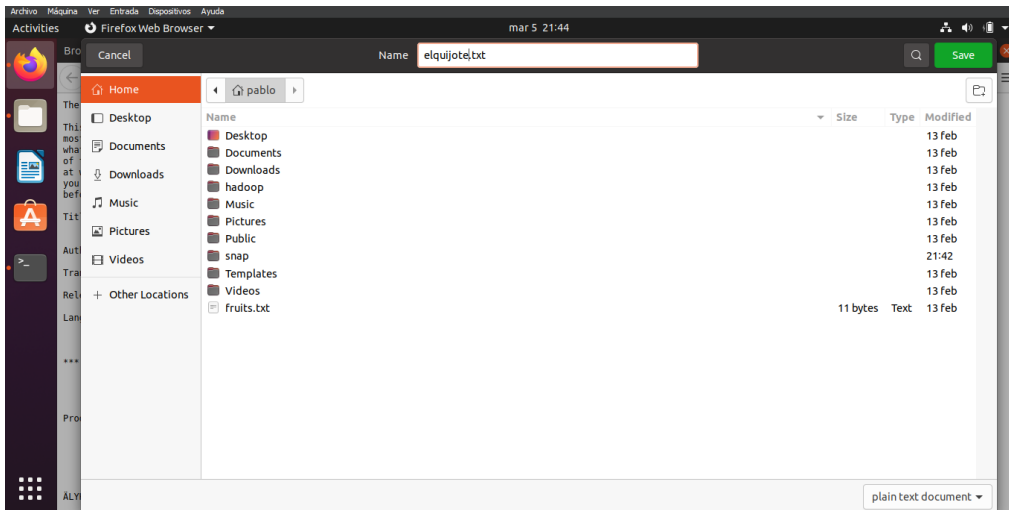
\$ start-yarn.sh

```
pablo@Ubuntu-BigData:~$ start-yarn.sh
Starting resourcemanager
Starting nodemanagers
```

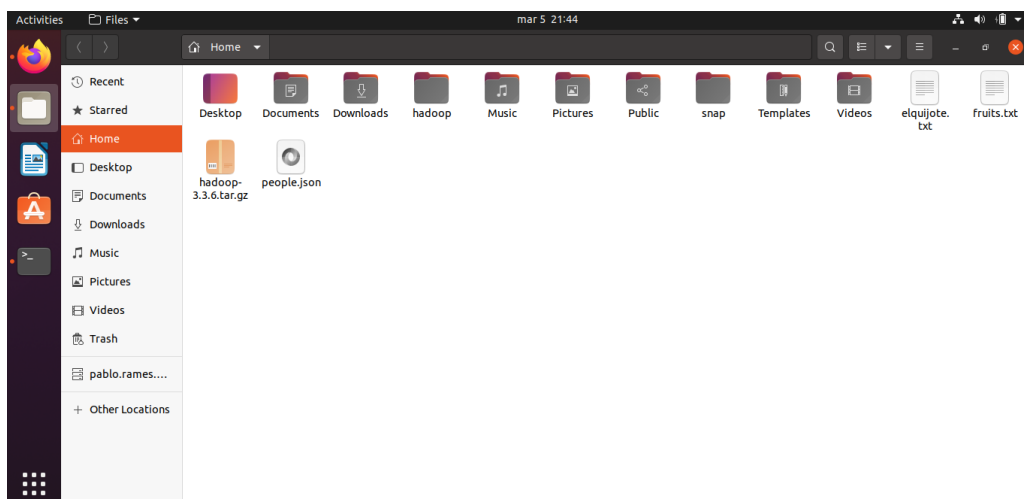
6.- Entrar en <https://gutenberg.org/ebooks/45203>



7.- Descargar Plain Text UTF-8 o Copiar el texto

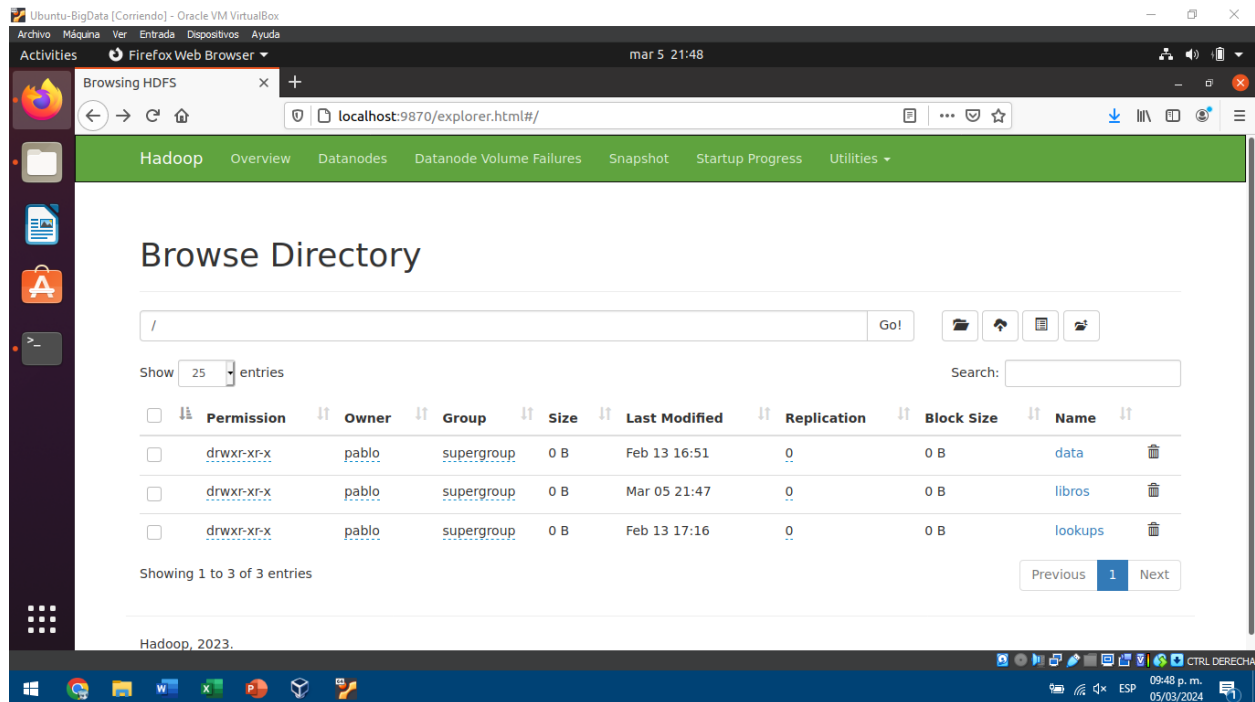


8.- Para guardar el archivo “elquijote.txt”



9.- Ahora cree un directorio llamado “libros”

```
pablo@Ubuntu-BigData:~$ hdfs dfs -mkdir /libros
```



Ubuntu-BigData [Corriendo] - Oracle VM VirtualBox

Archivo Máquina Ver Entrada Dispositivos Ayuda

Activities Firefox Web Browser mar 5 21:48

Browsing HDFS x +

localhost:9870/explorer.html/#/

Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities

Browse Directory

/ Go! [Icons]

Show 25 entries Search:

<input type="checkbox"/>	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	<input type="checkbox"/>
<input type="checkbox"/>	drwxr-xr-x	pablo	supergroup	0 B	Feb 13 16:51	0	0 B	data	<input type="checkbox"/>
<input type="checkbox"/>	drwxr-xr-x	pablo	supergroup	0 B	Mar 05 21:47	0	0 B	libros	<input type="checkbox"/>
<input type="checkbox"/>	drwxr-xr-x	pablo	supergroup	0 B	Feb 13 17:16	0	0 B	lookups	<input type="checkbox"/>

Showing 1 to 3 of 3 entries

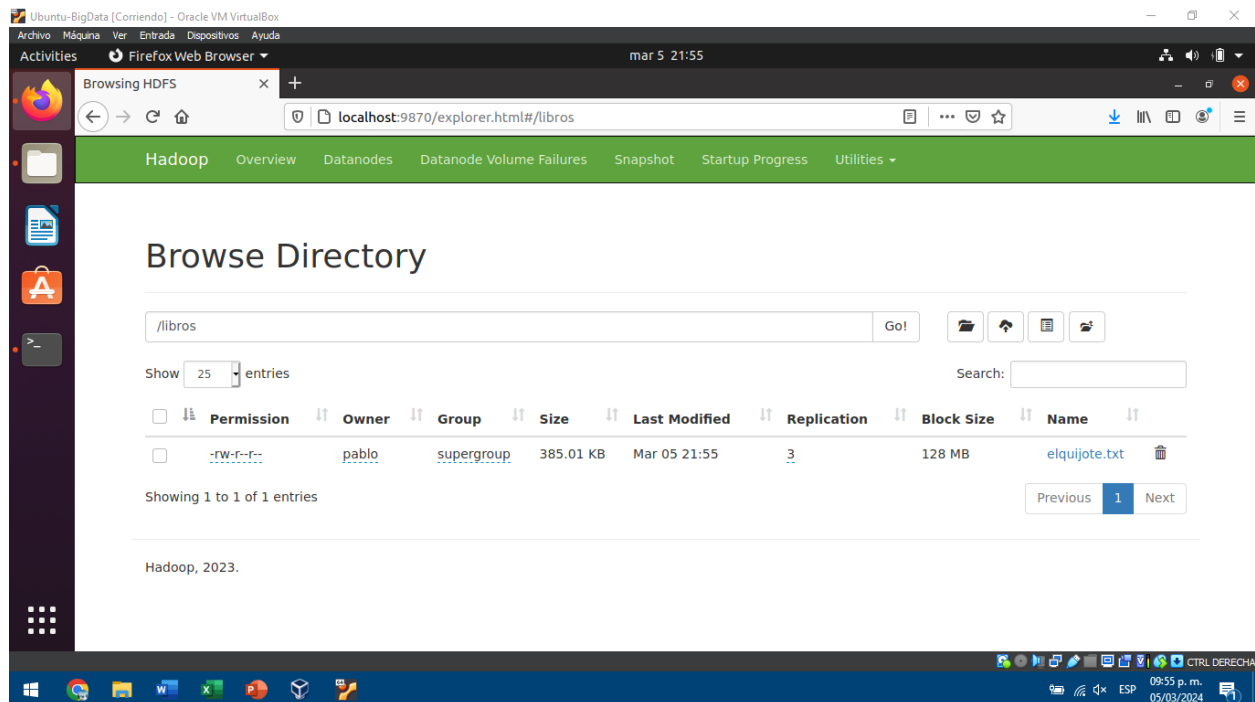
Previous 1 Next

Hadoop, 2023.

Windows taskbar: 09:48 p.m. 05/03/2024

10.- Posteriormente suba el archivo “elquijote.txt” al cluster en el directorio /libros

```
pablo@Ubuntu-BigData:~$ hdfs dfs -copyFromLocal /home/pablo/elquijote.txt /libros
```



Ubuntu-BigData [Corriendo] - Oracle VM VirtualBox

Archivo Máquina Ver Entrada Dispositivos Ayuda

Activities Firefox Web Browser mar 5 21:55

Browsing HDFS x +

localhost:9870/explorer.html/#/libros

Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities

Browse Directory

/libros Go! [Icons]

Show 25 entries Search:

<input type="checkbox"/>	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	<input type="checkbox"/>
<input type="checkbox"/>	-rw-r--r--	pablo	supergroup	385.01 KB	Mar 05 21:55	3	128 MB	elquijote.txt	<input type="checkbox"/>

Showing 1 to 1 of 1 entries

Previous 1 Next

Hadoop, 2023.

Windows taskbar: 09:55 p.m. 05/03/2024

11.- Posteriormente ubicarse en el directorio:

```
$ cd /hadoop-3.3.6/share/hadoop/mapreduce
```

```
pablo@Ubuntu-BigData:~$ cd /hadoop-3.3.6/share/hadoop/mapreduce
pablo@Ubuntu-BigData:~/hadoop-3.3.6/share/hadoop/mapreduce$
```

12.- Observar los comandos dentro de los archivos .jar, para ello escribir en comando de línea

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

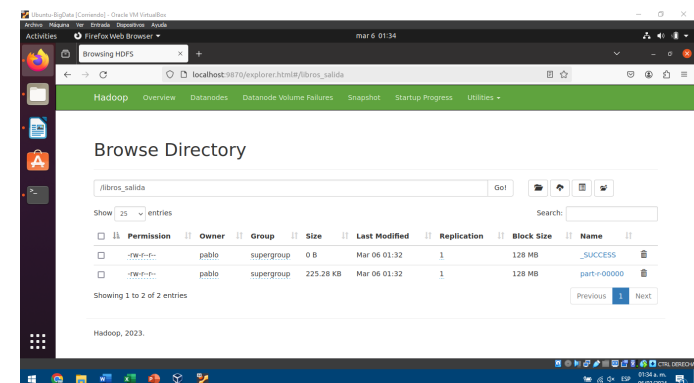
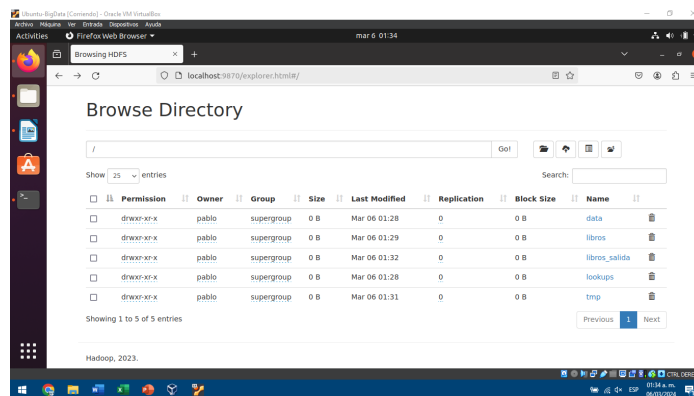
```
pablo@Ubuntu-BigData:~/hadoop-3.3.6/share/hadoop/mapreduce$ hadoop jar hadoop-mapreduce-examples-3.3.6.jar
```

```
An example program must be given as the first argument.
Valid program names are:
aggregatewordcount: An Aggregate based map/reduce program that counts the words in the input files.
aggregatewordhist: An Aggregate based map/reduce program that computes the histogram of the words in the input files.
bbp: A map/reduce program that uses Bailey-Borwein-Plouffe to compute exact digits of PI.
dcount: An example job that counts the pageview counts from a database.
distbbp: A map/reduce program that uses a BBP-type formula to compute exact bits of PI.
grep: A map/reduce program that counts the matches of a regex in the input.
join: A job that effects a join over sorted, equally partitioned datasets
multifilewc: A job that counts words from several files.
pentomino: A map/reduce tile laying program to find solutions to pentomino problems.
pi: A map/reduce program that estimates PI using a quasi-Monte Carlo method.
randomtextwriter: A map/reduce program that writes 10GB of random textual data per node.
randomwriter: A map/reduce program that writes 10GB of random data per node.
secondarysort: An example defining a secondary sort to the reduce.
sort: A map/reduce program that sorts the data written by the random writer.
sudoku: A sudoku solver.
terasort: Generate data for the terasort
terasort: Run the terasort
teravalidate: Checking results of terasort
wordcount: A map/reduce program that counts the words in the input files.
wordmeans: A map/reduce program that counts the average length of the words in the input files.
wordmedian: A map/reduce program that counts the median length of the words in the input files.
wordstandarddeviation: A map/reduce program that counts the standard deviation of the length of the words in the input files.
```

13.- Ahora probar los comandos dentro del examples jar “wordcount”

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

```
pablo@Ubuntu-BigData:~/hadoop-3.3.6/share/hadoop/mapreduce$ hadoop jar hadoop-mapreduce-examples-3.3.6.jar wordcount /libros /libros_salida
```



14.- Visualice el archivo de salida /libros_salida

\$ hdfs dfs -ls /libros_salida

```
pablo@Ubuntu-BigData:~/hadoop-3.3.6/share/hadoop/mapreduce$ hdfs dfs -ls /libros_salida
Found 2 items
-rw-r--r--  1 pablo supergroup          0 2024-03-06 01:32 /libros_salida/_SUCCESS
-rw-r--r--  1 pablo supergroup    230691 2024-03-06 01:32 /libros_salida/part-r-00000
```

15.- Ahora obtenga el archivo /libros_salida/part-r-00000 en la dirección /home/TECMILENIO/hadoop-3.3.6/share/hadoop/mapreduce

\$ “Escriba su código”

```
pablo@Ubuntu-BigData:~$ hdfs dfs -copyToLocal /libros_salida/part-r-00000 /home/pablo/hadoop-3.3.6/share/hadoop/mapreduce
pablo@Ubuntu-BigData:~$ cd /home/pablo/hadoop-3.3.6/share/hadoop/mapreduce
pablo@Ubuntu-BigData:~/hadoop-3.3.6/share/hadoop/mapreduce$ ls
hadoop-mapreduce-client-app-3.3.6.jar      hadoop-mapreduce-client-jobclient-3.3.6.jar  hadoop-mapreduce-examples-3.3.6.jar
hadoop-mapreduce-client-common-3.3.6.jar   hadoop-mapreduce-client-jobclient-3.3.6-tests.jar  jdiff
hadoop-mapreduce-client-core-3.3.6.jar     hadoop-mapreduce-client-native-task-3.3.6.jar  lib-examples
hadoop-mapreduce-client-hs-3.3.6.jar       hadoop-mapreduce-client-shuffle-3.3.6.jar  part-r-00000
hadoop-mapreduce-client-hs-plugins-3.3.6.jar hadoop-mapreduce-client-uploader-3.3.6.jar  sources
```

16.- Posteriormente escriba para visualizar el conteo:

\$ more part-r-00000

```
pablo@Ubuntu-BigData:~/hadoop-3.3.6/share/hadoop/mapreduce$ more part-r-00000
```

```
"Aseihin, 1
"Astuhan 1
"Ava 1
"Ava 1
"Avatkaat 1
"Don 2
"Dulcinea 1
"Eikaa 1
"Ekäpä 1
"Ei 11
"Ei", 1
"Ei, 4
"Eihän 1
"Eikä 1
"Eikö 2
"Eipähän 1
"Eivät 1
"Elkää 2
"Elkää 1
--More--(0%)
```

17.- Otra opción es con hadoop, entonces utilice el comando cat

\$ “Escriba su código”

```
pablo@Ubuntu-BigData:~/hadoop-3.3.6/share/hadoop/mapreduce$ cat part-r-00000
```

```
äarettömällä 1
ääriin 1
öillä, 1
öin 2
öitteni 1
öljyä 2
öljyä; 1
"AS-IS", 1
"Defects," 1
"Information 1
"Plain 2
"Project 5
"Right 1
• 4
```

Reflexión:

La exploración de MapReduce desde la gestión de grupos hasta la ejecución del ejemplo de "cantidad de palabras" ha proporcionado una comprensión completa de la manipulación de datos a gran escala de manera efectiva. La comprensión de comandos en archivos.jar y la visualización de resultados hacen que MapReduce sea más versátil. Esta experiencia no solo mejora las habilidades prácticas en sistemas distribuidos, sino que también mejora la capacidad de extraer conocimientos importantes de conjuntos de datos de gran tamaño. Este método me ayuda a mejorar mi preparación en ciencia de datos y sistemas al proporcionarme herramientas para abordar de manera efectiva los desafíos del análisis de datos masivos.