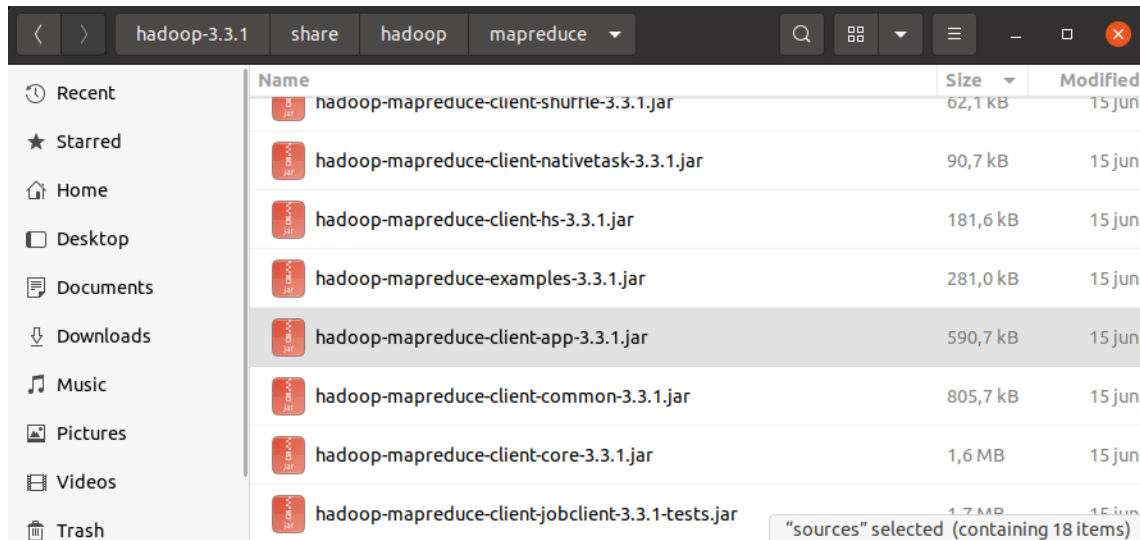


2. Ejemplos ejecución de algoritmo de MapReduce en Monopuesto

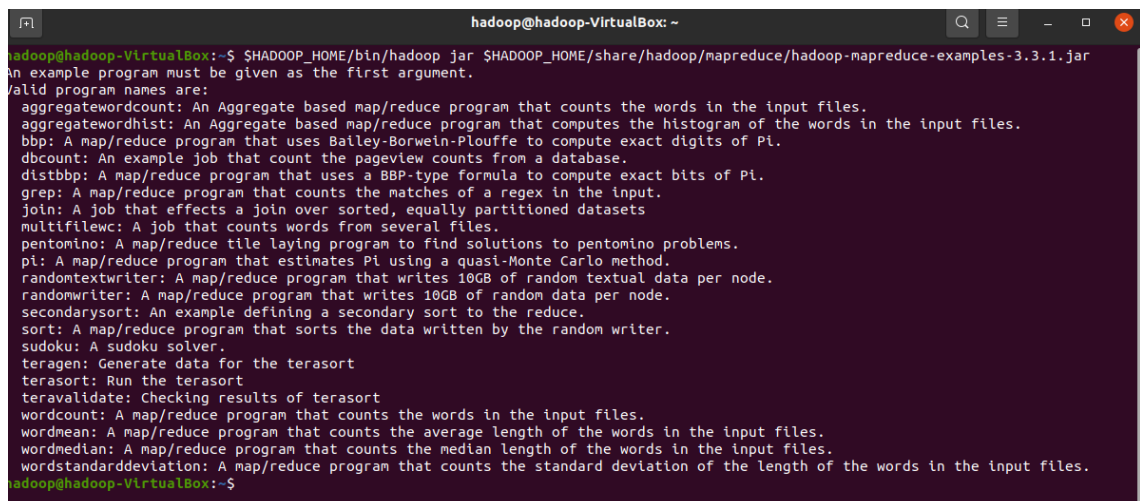
Utilizando el archivo de ejemplo “hadoop-mapreduce-examples-3.3.1.jar” que viene con la instalación, ejecutar varios ejemplos. Basado en la información de <https://hadoop.apache.org/docs/stable/hadoop-mapreduce-client/hadoop-mapreduce-client-core/MapReduceTutorial.html>



2.1. Listar los comandos de ejemplo

Ejecutar la instrucción siguiente desde la carpeta de hadoop

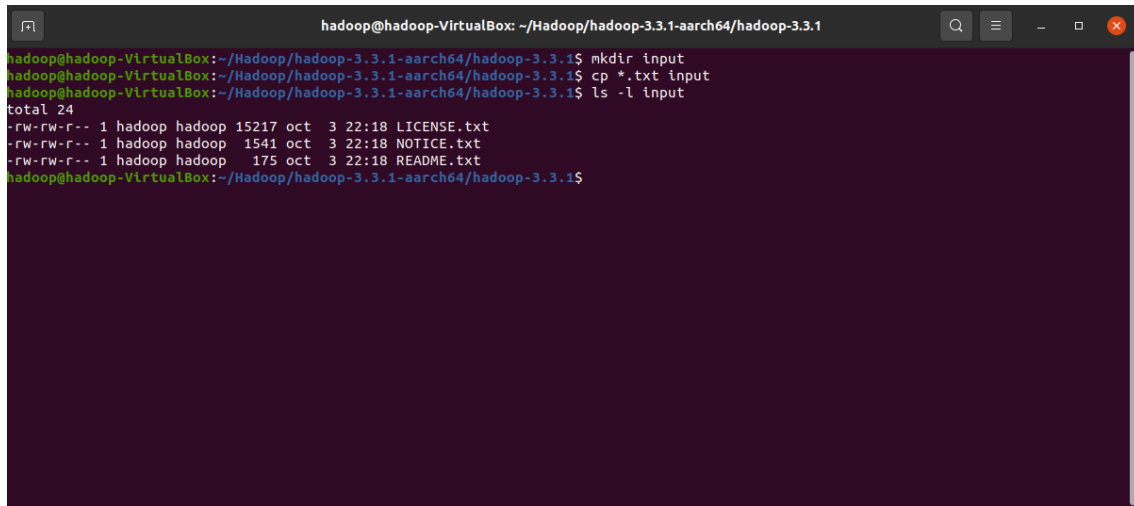
```
$bin/hadoop jar /share/hadoop/mapreduce/hadoop-mapreduce-examples-3.3.4.jar
```



2.2. Ejemplo 1 Contar palabras

Paso 1 Crear una carpeta temporal y copiar los archivos de texto de la carpeta de instalación.

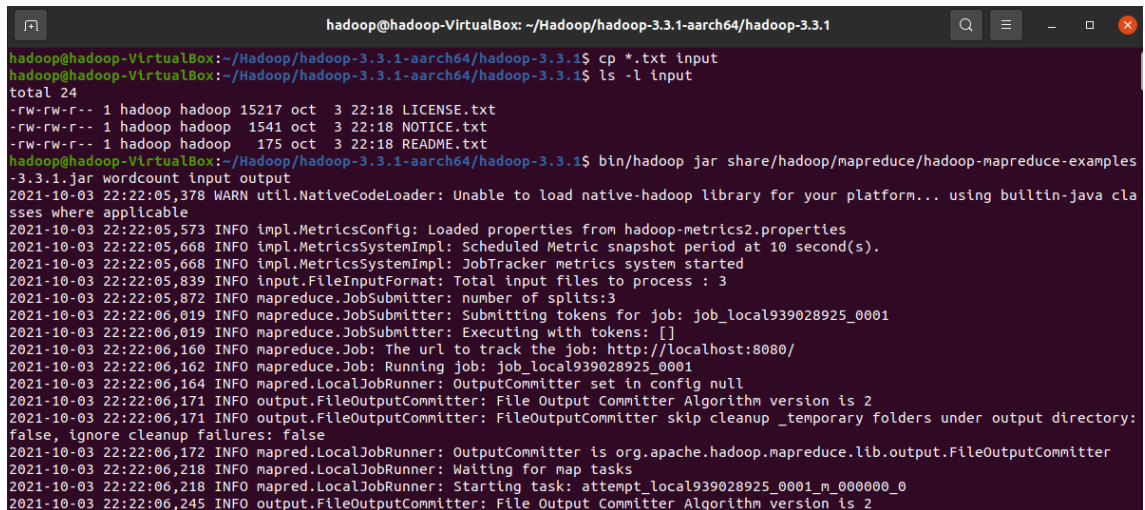
```
$ mkdir input
$ cp *.txt input
$ ls -l input
```



```
hadoop@hadoop-VirtualBox: ~/Hadoop/hadoop-3.3.1-aarch64/hadoop-3.3.1
hadoop@hadoop-VirtualBox:~/Hadoop/hadoop-3.3.1-aarch64/hadoop-3.3.1$ mkdir input
hadoop@hadoop-VirtualBox:~/Hadoop/hadoop-3.3.1-aarch64/hadoop-3.3.1$ cp *.txt input
hadoop@hadoop-VirtualBox:~/Hadoop/hadoop-3.3.1-aarch64/hadoop-3.3.1$ ls -l input
total 24
-rw-rw-r-- 1 hadoop hadoop 15217 oct  3 22:18 LICENSE.txt
-rw-rw-r-- 1 hadoop hadoop  1541 oct  3 22:18 NOTICE.txt
-rw-rw-r-- 1 hadoop hadoop   175 oct  3 22:18 README.txt
hadoop@hadoop-VirtualBox:~/Hadoop/hadoop-3.3.1-aarch64/hadoop-3.3.1$
```

Paso 2 Ejecutar el proceso contando las palabras de los archivos de la carpeta input

```
$ bin/hadoop jar share/hadoop/mapreduce/hadoop-mapreduce-examples-3.3.4.jar wordcount input output
```



```
hadoop@hadoop-VirtualBox: ~/Hadoop/hadoop-3.3.1-aarch64/hadoop-3.3.1
hadoop@hadoop-VirtualBox:~/Hadoop/hadoop-3.3.1-aarch64/hadoop-3.3.1$ cp *.txt input
hadoop@hadoop-VirtualBox:~/Hadoop/hadoop-3.3.1-aarch64/hadoop-3.3.1$ ls -l input
total 24
-rw-rw-r-- 1 hadoop hadoop 15217 oct  3 22:18 LICENSE.txt
-rw-rw-r-- 1 hadoop hadoop  1541 oct  3 22:18 NOTICE.txt
-rw-rw-r-- 1 hadoop hadoop   175 oct  3 22:18 README.txt
hadoop@hadoop-VirtualBox:~/Hadoop/hadoop-3.3.1-aarch64/hadoop-3.3.1$ bin/hadoop jar share/hadoop/mapreduce/hadoop-mapreduce-examples-3.3.1.jar wordcount input output
2021-10-03 22:22:05,378 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2021-10-03 22:22:05,573 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2021-10-03 22:22:05,668 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2021-10-03 22:22:05,668 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2021-10-03 22:22:05,839 INFO input.FileInputFormat: Total input files to process : 3
2021-10-03 22:22:05,872 INFO mapreduce.JobSubmitter: number of splits:3
2021-10-03 22:22:06,019 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local939028925_0001
2021-10-03 22:22:06,019 INFO mapreduce.JobSubmitter: Executing with tokens: []
2021-10-03 22:22:06,160 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
2021-10-03 22:22:06,162 INFO mapreduce.Job: Running job: job_local939028925_0001
2021-10-03 22:22:06,164 INFO mapred.LocalJobRunner: OutputCommitter set in config null
2021-10-03 22:22:06,171 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2021-10-03 22:22:06,171 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory: false, ignore cleanup failures: false
2021-10-03 22:22:06,172 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
2021-10-03 22:22:06,218 INFO mapred.LocalJobRunner: Waiting for map tasks
2021-10-03 22:22:06,218 INFO mapred.LocalJobRunner: Starting task: attempt_local939028925_0001_m_0000000_0
2021-10-03 22:22:06,245 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
```

Paso 3 Verifico el resultado

```
$ cat output/*
```

```
hadoop@hadoop-VirtualBox: ~/Hadoop/hadoop-3.3.1-aarch64/hadoop-3.3.1
hadoop@hadoop-VirtualBox:~/Hadoop/hadoop-3.3.1-aarch64/hadoop-3.3.1$ cat output/*
AS 2
Contribution" 1
Contributor" 1
Derivative 1
Legal 1
License" 1
License"); 1
Licensor" 1
NOTICE" 1
Not 1
Object" 1
Source" 1
Work" 1
You" 1
Your") 1
[]" 1
control" 1
printed 1
submitted" 1
50%) 1
BIS), 1
Don't 1
ECCN) 1
TSU) 1
a) 1
an 1
```

2.3. Ejemplo 2 Filtrar palabras

Paso 1 Crear carpeta input2 y copiar datos de ejemplo

```
$ mkdir input2
$ cp etc/hadoop/*.xml input2
$ ls -l input2
```

```
hadoop@hadoop-VirtualBox: ~/Hadoop/hadoop-3.3.1-aarch64/hadoop-3.3.1
hadoop@hadoop-VirtualBox:~/Hadoop/hadoop-3.3.1-aarch64/hadoop-3.3.1$ mkdir input2
hadoop@hadoop-VirtualBox:~/Hadoop/hadoop-3.3.1-aarch64/hadoop-3.3.1$ cp etc/hadoop/*.xml input2
hadoop@hadoop-VirtualBox:~/Hadoop/hadoop-3.3.1-aarch64/hadoop-3.3.1$ ls -l input2
total 56
-rw-r--r-- 1 hadoop hadoop 9213 oct 3 22:32 capacity-scheduler.xml
-rw-r--r-- 1 hadoop hadoop 774 oct 3 22:32 core-site.xml
-rw-r--r-- 1 hadoop hadoop 11765 oct 3 22:32 hadoop-policy.xml
-rw-r--r-- 1 hadoop hadoop 683 oct 3 22:32 hdfs-rbf-site.xml
-rw-r--r-- 1 hadoop hadoop 775 oct 3 22:32 hdfs-site.xml
-rw-r--r-- 1 hadoop hadoop 620 oct 3 22:32 https-site.xml
-rw-r--r-- 1 hadoop hadoop 3518 oct 3 22:32 kms-acls.xml
-rw-r--r-- 1 hadoop hadoop 682 oct 3 22:32 kms-site.xml
-rw-r--r-- 1 hadoop hadoop 758 oct 3 22:32 mapred-site.xml
-rw-r--r-- 1 hadoop hadoop 690 oct 3 22:32 yarn-site.xml
hadoop@hadoop-VirtualBox:~/Hadoop/hadoop-3.3.1-aarch64/hadoop-3.3.1$
```

Paso 2 ejecutar el comando

```
$ bin/hadoop jar share/hadoop/mapreduce/hadoop-mapreduce-examples-3.3.4.jar grep input2 output2 'dfs[a-z.]+'
```

```
hadoop@hadoop-VirtualBox: ~/Hadoop/hadoop-3.3.1-aarch64/hadoop-3.3.1
-rw-r--r-- 1 hadoop hadoop 758 oct 3 22:32 mapred-site.xml
-rw-r--r-- 1 hadoop hadoop 690 oct 3 22:32 yarn-site.xml
hadoop@hadoop-VirtualBox:~/Hadoop/hadoop-3.3.1-aarch64/hadoop-3.3.1$ bin/hadoop jar share/hadoop/mapreduce/hadoop-mapreduce-examples-3.3.1.jar grep input2 output2 'dfs[a-z.]+'
2021-10-03 22:38:48,247 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2021-10-03 22:38:48,431 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2021-10-03 22:38:48,508 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2021-10-03 22:38:48,508 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2021-10-03 22:38:48,683 INFO input.FileInputFormat: Total input files to process : 10
2021-10-03 22:38:48,717 INFO mapreduce.JobSubmitter: number of splits:10
2021-10-03 22:38:48,835 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local1473584412_0001
2021-10-03 22:38:48,836 INFO mapreduce.JobSubmitter: Executing with tokens: []
2021-10-03 22:38:48,944 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
2021-10-03 22:38:48,945 INFO mapreduce.Job: Running job: job_local1473584412_0001
2021-10-03 22:38:48,946 INFO mapred.LocalJobRunner: OutputCommitter set in config null
2021-10-03 22:38:48,952 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2021-10-03 22:38:48,952 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory: false, ignore cleanup failures: false
2021-10-03 22:38:48,953 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
2021-10-03 22:38:48,998 INFO mapred.LocalJobRunner: Waiting for map tasks
2021-10-03 22:38:48,998 INFO mapred.LocalJobRunner: Starting task: attempt_local1473584412_0001_m_000000_0
2021-10-03 22:38:49,023 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2021-10-03 22:38:49,023 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory: false, ignore cleanup failures: false
2021-10-03 22:38:49,042 INFO mapred.Task: Using ResourceCalculatorProcessTree : [ ]
2021-10-03 22:38:49,047 INFO mapred.MapTask: Processing split: file:/home/hadoop/Hadoop/hadoop-3.3.1-aarch64/hadoop-3.3.1/input2/had
```

Paso 3 comprobar resultado

```
$ cat output2/*
```

```
hadoop@hadoop-VirtualBox: ~/Hadoop/hadoop-3.3.1-aarch64/hadoop-3.3.1
at org.apache.hadoop.mapreduce.Job$11.run(Job.java:1571)
at org.apache.hadoop.mapreduce.Job$11.run(Job.java:1568)
at java.security.AccessController.doPrivileged(Native Method)
at javax.security.auth.Subject.doAs(Subject.java:422)
at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1878)
at org.apache.hadoop.mapreduce.Job.submit(Job.java:1568)
at org.apache.hadoop.mapreduce.Job.waitForCompletion(Job.java:1589)
at org.apache.hadoop.examples.Grep.run(Grep.java:94)
at org.apache.hadoop.util.ToolRunner.run(ToolRunner.java:76)
at org.apache.hadoop.examples.Grep.main(Grep.java:103)
at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)
at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
at java.lang.reflect.Method.invoke(Method.java:498)
at org.apache.hadoop.util.ProgramDriver$ProgramDescription.invoke(ProgramDriver.java:71)
at org.apache.hadoop.util.ProgramDriver.run(ProgramDriver.java:144)
at org.apache.hadoop.examples.ExampleDriver.main(ExampleDriver.java:74)
at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)
at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
at java.lang.reflect.Method.invoke(Method.java:498)
at org.apache.hadoop.util.RunJar.run(RunJar.java:323)
at org.apache.hadoop.util.RunJar.main(RunJar.java:236)
hadoop@hadoop-VirtualBox:~/Hadoop/hadoop-3.3.1-aarch64/hadoop-3.3.1$ cat output2/*
dfsadmin
hadoop@hadoop-VirtualBox:~/Hadoop/hadoop-3.3.1-aarch64/hadoop-3.3.1$
```

3. Configuración modo pseudo distribuido en local

3.1. Modificar el archivo de configuración "etc/hadoop/core-site.xml"

```
<configuration>

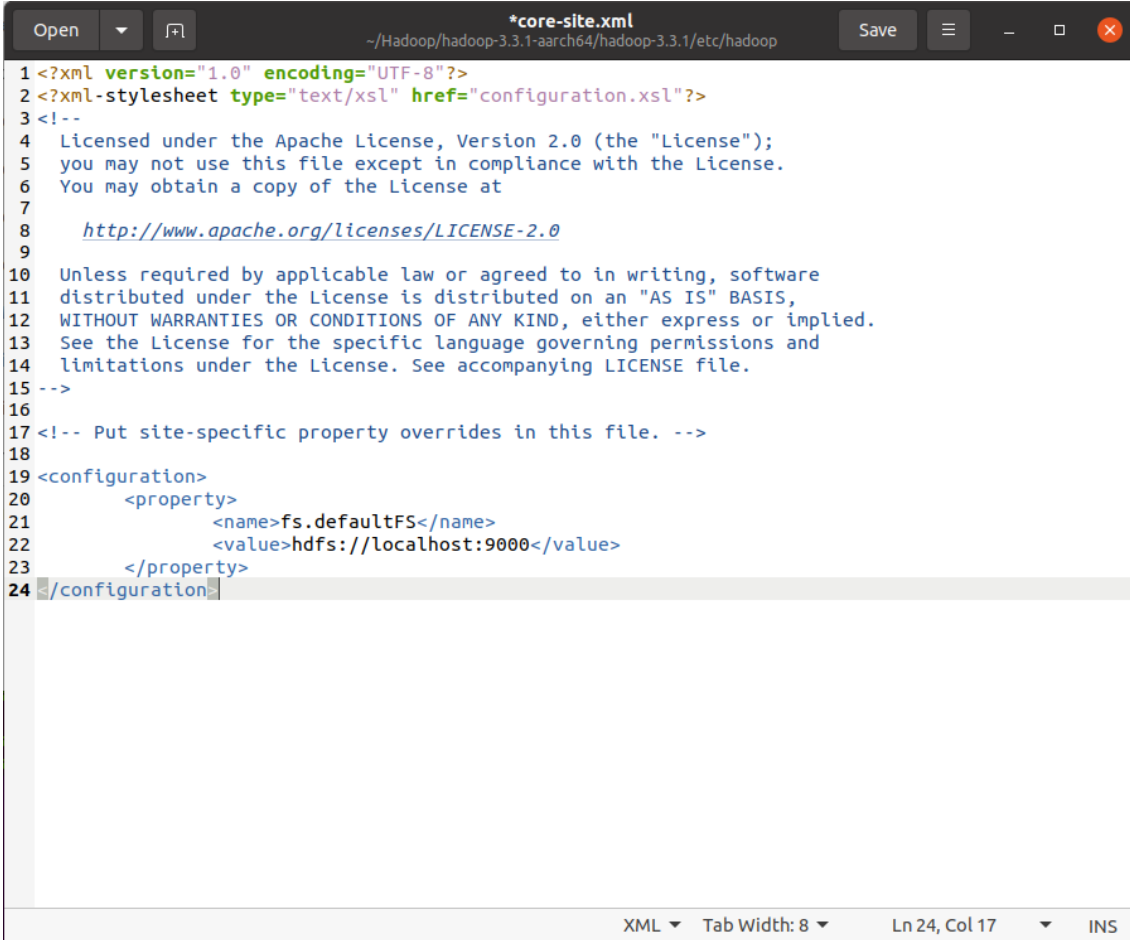
  <property>

    <name>fs.defaultFS</name>

    <value>hdfs://localhost:9000</value>

  </property>

</configuration>
```



```
*core-site.xml
~/Hadoop/hadoop-3.3.1-aarch64/hadoop-3.3.1/etc/hadoop

1 <?xml version="1.0" encoding="UTF-8"?>
2 <?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
3 <!--
4 Licensed under the Apache License, Version 2.0 (the "License");
5 you may not use this file except in compliance with the License.
6 You may obtain a copy of the License at
7
8   http://www.apache.org/licenses/LICENSE-2.0
9
10 Unless required by applicable law or agreed to in writing, software
11 distributed under the License is distributed on an "AS IS" BASIS,
12 WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
13 See the License for the specific language governing permissions and
14 limitations under the License. See accompanying LICENSE file.
15 -->
16
17 <!-- Put site-specific property overrides in this file. -->
18
19 <configuration>
20   <property>
21     <name>fs.defaultFS</name>
22     <value>hdfs://localhost:9000</value>
23   </property>
24 </configuration>
```

3.2. Modificar el archivo de configuración etc/hadoop/hdfs-site.xml

```
<configuration>

  <property>

    <name>dfs.replication</name>
```