

Practica Hadoop

```
Session Actions Edit View Help
└─# wget https://dlcdn.apache.org/hadoop/common/hadoop-3.3.6/hadoop-3.3.6.tar.gz
2025-10-18 07:27:52  https://dlcdn.apache.org/hadoop/common/hadoop-3.3.6/hadoop-3.3.6.tar.gz
--2025-10-18 07:27:52-- https://dlcdn.apache.org/hadoop/common/hadoop-3.3.6/hadoop-3.3.6.tar.gz
Connecting to dlcdn.apache.org (dlcdn.apache.org)|151.101.7.132|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 730107476 (699MB) [application/x-gzip]
Saving to: "hadoop-3.3.6.tar.gz"

hadoop-3.3.6.tar.gz          100%[=====] 696.28M  1.70MB/s   in 6m 37s

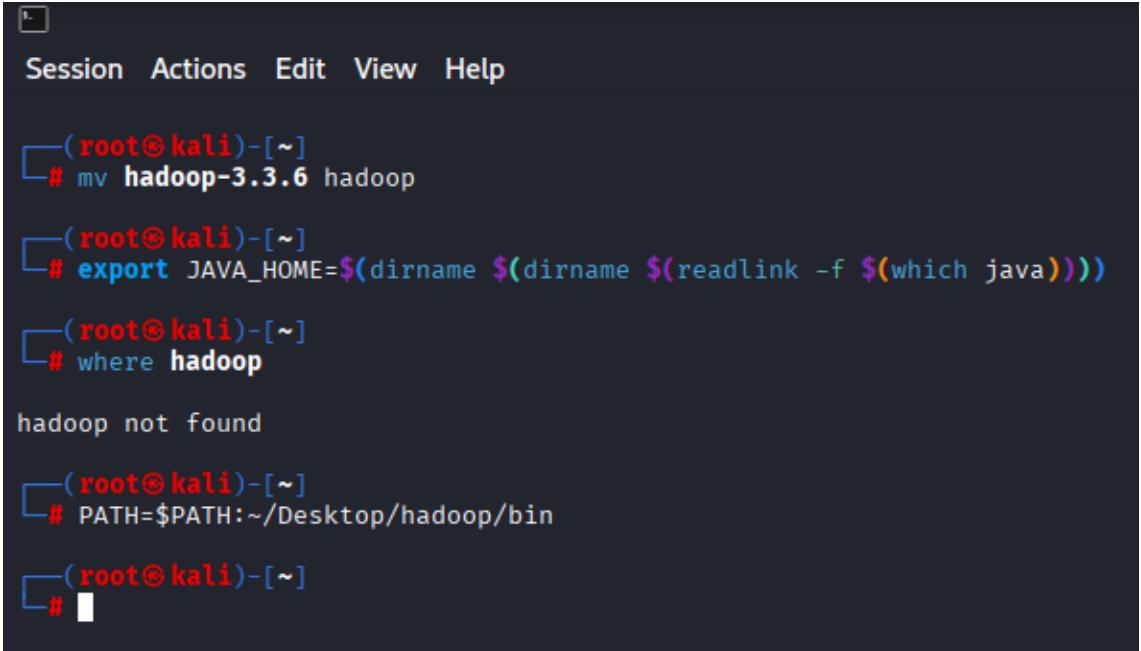
2025-10-18 07:34:47 (1.75 MB/s) - 'hadoop-3.3.6.tar.gz' saved [730107476/730107476]

└─#
```

```
[root@kali)~]# tar -xvf hadoop-3.3.6.tar.gz
hadoop-3.3.6/
hadoop-3.3.6/NOTICE-binary
hadoop-3.3.6/licenses-binary/
hadoop-3.3.6/licenses-binary/LICENSE-lz4.txt
hadoop-3.3.6/licenses-binary/LICENSE-zstd-jni.txt
hadoop-3.3.6/licenses-binary/LICENSE-tree.txt
hadoop-3.3.6/licenses-binary/LICENSE-azure-cosmosdb.txt
hadoop-3.3.6/licenses-binary/LICENSE-moment.txt
hadoop-3.3.6/licenses-binary/LICENSE-jline.txt
hadoop-3.3.6/licenses-binary/LICENSE-hsql.txt
hadoop-3.3.6/licenses-binary/LICENSE-junit.txt
hadoop-3.3.6/licenses-binary/LICENSE-ojalgo.txt
hadoop-3.3.6/licenses-binary/LICENSE-paranamer.txt
hadoop-3.3.6/licenses-binary/LICENSE-freebsd.txt
hadoop-3.3.6/licenses-binary/LICENSE-jakarta.xml.bind-api.txt
hadoop-3.3.6/licenses-binary/LICENSE-bloomfilter.txt
hadoop-3.3.6/licenses-binary/LICENSE-d3.txt
hadoop-3.3.6/licenses-binary/LICENSE-jaf.txt
hadoop-3.3.6/licenses-binary/LICENSE-dust.txt
hadoop-3.3.6/licenses-binary/LICENSE-jdom.txt
hadoop-3.3.6/licenses-binary/LICENSE-datatables.txt
hadoop-3.3.6/licenses-binary/LICENSE-azure-data-lake-store.txt
hadoop-3.3.6/licenses-binary/LICENSE-protobuf.txt
hadoop-3.3.6/licenses-binary/LICENSE-hamcrest.txt
hadoop-3.3.6/licenses-binary/LICENSE-bouncycastle.txt
hadoop-3.3.6/licenses-binary/LICENSE-cddl-gplv2-ce.txt
hadoop-3.3.6/licenses-binary/LICENSE-jsr305.txt
hadoop-3.3.6/licenses-binary/LICENSE-go.txt
hadoop-3.3.6/licenses-binary/LICENSE-leveledb.txt
hadoop-3.3.6/licenses-binary/LICENSE-nvd3.txt
hadoop-3.3.6/licenses-binary/LICENSE-leveledbjni.txt
hadoop-3.3.6/licenses-binary/LICENSE-checker-framework.txt
hadoop-3.3.6/licenses-binary/LICENSE-asm.txt
hadoop-3.3.6/licenses-binary/LICENSE-jstree.txt
hadoop-3.3.6/licenses-binary/LICENSE-jquery.txt
hadoop-3.3.6/licenses-binary/LICENSE-re2j.txt
hadoop-3.3.6/licenses-binary/LICENSE-gtest.txt
hadoop-3.3.6/licenses-binary/LICENSE-dnsjava.txt
hadoop-3.3.6/licenses-binary/LICENSE-angular-nvd3.txt
hadoop-3.3.6/licenses-binary/LICENSE-mssql-jdbc.txt
hadoop-3.3.6/licenses-binary/LICENSE-azure-keyvault.txt
hadoop-3.3.6/licenses-binary/LICENSE-jsch.txt
hadoop-3.3.6/licenses-binary/LICENSE-xtermjs.txt
hadoop-3.3.6/licenses-binary/LICENSE-slicing-by-8.txt
hadoop-3.3.6/licenses-binary/LICENSE-slf4j.txt
```



```
(root㉿kali)-[~]
# mv hadoop-3.3.6 hadoop
```



```
(root㉿kali)-[~]
# mv hadoop-3.3.6 hadoop

(root㉿kali)-[~]
# export JAVA_HOME=$(dirname $(dirname $(readlink -f $(which java))))
```

hadoop not found

```
(root㉿kali)-[~]
# PATH=$PATH:~/Desktop/hadoop/bin
```

```
(root㉿kali)-[~]
# cp -r /root/hadoop /home/kali/Desktop/
```

```
(root㉿kali)-[~]
# export PATH=$PATH:/home/kali/Desktop/hadoop/bin/
```

```
(root㉿kali)-[~]
#
```

```
(root㉿kali)-[~]
# mkdir -p /tmp/entrada
```

```
(root㉿kali)-[~]
# cp /home/kali/Desktop/hadoop/etc/hadoop/*.xml /tmp/entrada
```

```
(root㉿kali)-[~]
# ls /tmp/entrada
capacity-scheduler.xml  core-site.xml  hadoop-policy.xml  hdfs-site.xml  httpfs-site.xml  kms-acls.xml  kms-site.xml  mapred-site.xml  yarn-site.xml
```

```
(root㉿kali)-[~]
#
```

```

root@kali:~]
# cd /home/kali/Desktop/hadoop/share/hadoop/mapreduce/
[ root@kali:~/home/./hadoop/share/hadoop/mapreduce]
# ls
hadoop-mapreduce-client-app-3.3.0.jar      hadoop-mapreduce-client-hs-3.3.0.jar      hadoop-mapreduce-client-jobclient-3.3.0-tests.jar   hadoop-mapreduce-client-uploader-3.3.0.jar      lib-examples
hadoop-mapreduce-client-common-3.3.0.jar    hadoop-mapreduce-client-hs-plugins-3.3.0.jar  hadoop-mapreduce-client-nativetask-3.3.0.jar     hadoop-mapreduce-client-examples-3.3.0.jar      sources
hadoop-mapreduce-client-core-3.3.0.jar       hadoop-mapreduce-client-jobclient-3.3.0.jar   hadoop-mapreduce-client-shuffle-3.3.0.jar      jdiff

[ root@kali:~/home/./hadoop/share/hadoop/mapreduce]
# jar tvf hadoop-mapreduce-examples-3.3.0.jar
  Sun Jun 18 08:51:36 EDT 2023 META-INF/MANIFEST.MF
  0 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/
  0 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/
  0 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/
  0 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/dancing/
  0 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/terasort/
  0 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/pi/
  0 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/pi/math/
15217 Sun Jun 18 08:51:36 EDT 2023 META-INF/LICENSE.txt
1541 Sun Jun 18 08:51:36 EDT 2023 META-INF/NOTICE.txt
9581 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/dancing/DancingLinks.class
435 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/dancing/DancingLinks$ServerAcceptor.class
6730 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/dancing/Pentomino$Pentomino.class
1377 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/dancing/PentominoSolutionCategory.class
904 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/dancing/Sudoku$RowConstraint.class
1938 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/dancing/DancingLinks$Node.class
3168 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/dancing/Pentomino$Piece.class
2459 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/dancing/OneSidedPentomino.class
6300 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/dancing/Pentomino$Column.class
1082 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/dancing/DancingLinks$ColumnHeader.class
605 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/dancing/Pentomino$Point.class
9618 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/dancing/Pentomino$Column.class
1519 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/dancing/Pentomino$PointPrinter.class
5112 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/dancing/OneSidedPentomino$OneSidePentomino.class
2109 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/dancing/Sudoku$SolutionPrinter.class
2109 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/dancing/Sudoku$ColumnConstraint.class
912 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/dancing/Sudoku$CellConstraint.class
238 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/dancing/Sudoku$ColumnName.class
70 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/dancing/Sudoku$RowConstraint.class
3681 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/dancing/Distribution$RandomMap$SolutionCatcher.class
239 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/dancing/Pentomino$ColumnName.class
2416 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/WordMedian$WordMedianReducer.class
2258 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/WordCount.class
2880 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/DBScan$PageView$PageviewReducer.class
2602 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/DBScan$SumReducer.class
2459 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/WordMean$WordMeanMapper.class
5198 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/WordMean.class
2134 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/AggregateWordHistogram$AggregateWordHistogramPlugin.class
2935 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/RandomTextWriter.class
2729 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/BlurryBox$BlurryBoxMapper.class
764 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/RandomGauss$RandomGaussConstant.class
1165 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/terasort/TeraGen$Counters.class
4155 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/terasort/TeraValidate$ValidateMapper.class
6938 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/terasort/TeraInputFormat.class
1512 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/terasort/TeraScheduler$Split.class
3534 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/terasort/TeraValidate$ValidateReducer.class
4273 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/terasort/TeraGen$SortMapper.class

[ root@kali:~/home/./hadoop/share/hadoop/mapreduce]
Session Actions Edit View Help
root@kali:/home/kali/Desktop/hadoop/share/hadoop/mapreduce

[ root@kali:~/home/./hadoop/share/hadoop/mapreduce]
# hadoop jar hadoop-mapreduce-examples-3.3.0.jar grep /tmp/entrada /tmp/salida 'dfs[a-z].*'
2025-10-18 09:18:02.764 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2025-10-18 09:18:02.943 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2025-10-18 09:18:03.177 INFO input.FileInputFormat: Total input files to process : 10
2025-10-18 09:18:03.212 INFO mapreduce.JobSubmission: number of splits:10
2025-10-18 09:18:03.503 INFO mapreduce.JobSubmission: Submitting tokens for job: job_local1908251525_0001
2025-10-18 09:18:03.503 INFO mapreduce.JobSubmission: Executing with tokens: []
2025-10-18 09:18:03.622 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
2025-10-18 09:18:03.622 INFO mapreduce.Job: Running Job: job_local1908251525_0001
2025-10-18 09:18:03.630 INFO mapred.LocalJobRunner: OutputCommitter set to null
2025-10-18 09:18:03.635 INFO mapred.LocalJobRunner: OutputCommitter factory defined, defaulting to FileOutputCommitterFactory
2025-10-18 09:18:03.637 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2025-10-18 09:18:03.638 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2025-10-18 09:18:03.674 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
2025-10-18 09:18:03.674 INFO mapred.LocalJobRunner: Waiting for map tasks
2025-10-18 09:18:03.678 INFO mapred.LocalJobRunner: Starting task: attempt_local1908251525_0001_m_0000000_0
2025-10-18 09:18:03.699 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2025-10-18 09:18:03.699 INFO output.FileOutputCommitter: File Output Committer skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2025-10-18 09:18:03.755 INFO mapred.LocalJobRunner: Using ResourceCalculatorProcessTree : [ ]
2025-10-18 09:18:03.825 INFO mapred.MapTask: (EQUATOR) o kvl 26214396(104857584)
2025-10-18 09:18:03.825 INFO mapred.MapTask: mapreduce.task.io.sort.mb: 100
2025-10-18 09:18:03.825 INFO mapred.MapTask: soft limit at 838860800
2025-10-18 09:18:03.825 INFO mapred.MapTask: bufstart = 0; bufvoid = 104857600
2025-10-18 09:18:03.825 INFO mapred.MapTask: Kvstart = 26214396; length = 6553600
2025-10-18 09:18:03.853 INFO mapred.MapTask: MapredLocalJobRunner:
2025-10-18 09:18:03.853 INFO mapred.MapTask: Starting flush of map output
2025-10-18 09:18:03.853 INFO mapred.MapTask: Spilling map output
2025-10-18 09:18:03.853 INFO mapred.MapTask: bufstart = 0; bufend = 17; bufvoid = 104857600
2025-10-18 09:18:03.853 INFO mapred.MapTask: Kvstart = 26214396(104857584); kvend = 26214396(104857584); length = 1/6553600
2025-10-18 09:18:03.873 INFO mapred.MapTask: Finished spill 0
2025-10-18 09:18:03.892 INFO mapred.Task: Task:attempt_local1908251525_0001_m_000000_0 is done. And is in the process of committing
2025-10-18 09:18:03.897 INFO mapred.LocalJobRunner: map
2025-10-18 09:18:03.897 INFO mapred.Task: Task 'attempt_local1908251525_0001_m_000000_0' done.
2025-10-18 09:18:03.913 INFO mapred.Task: Final Counters for attempt_local1908251525_0001_m_000000_0: Counters: 18
File System Counters
FILE: Number of bytes read=294302
FILE: Number of bytes written=921741
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
Map-Reduce Framework
Map input records=275
Map output records=1
Map output bytes=17
Map output materialized bytes=25
Input split bytes=100
Combine input records=1
Combine output records=1
Spilled Records=1
Failed Shuffles=0
Merged Map outputs=0
GC time elapsed (ms)=0
Total committed heap usage (bytes)=202375168
File Input Format Counters
Bytes Read=11765

```

```
[root@kali]~[/home/.../hadoop/share/hadoop/mapreduce]
└─# cd /tmp/salida

[root@kali]~[/tmp/salida]
└─# ls -l
total 4
-rw-r--r-- 1 root root 11 Oct 18 09:18 part-r-00000
-rw-r--r-- 1 root root 0 Oct 18 09:18 _SUCCESS

[root@kali]~[/tmp/salida]
└─# cat part-r-00000
1          dfsadmin

[root@kali]~[/tmp/salida]
└─# █
```