

# Practica Hadoop

```
Session Actions Edit View Help
root@kali: ~

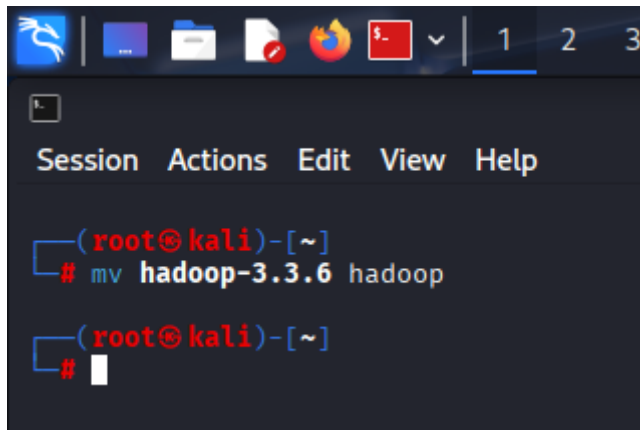
root@kali: ~#
root@kali: ~# curl https://d1cdn.apache.org/hadoop/common/hadoop-3.3.6/hadoop-3.3.6.tar.gz
--2025-10-18 07:27:52-- https://d1cdn.apache.org/hadoop/common/hadoop-3.3.6/hadoop-3.3.6.tar.gz
Resolving d1cdn.apache.org (d1cdn.apache.org) ... 151.101.2.132, 208.114.42.644
Connecting to d1cdn.apache.org (d1cdn.apache.org)|151.101.2.132|:443 ... connected.
HTTP request sent, awaiting response ... 200 OK
Length: 738107476 (699M) [application/x-gzip]
Saving to: 'hadoop-3.3.6.tar.gz'

hadoop-3.3.6.tar.gz          100% [=====>] 696.28M  1.78MB/s   in 6m 37s

2025-10-18 07:24:47 (1.75 MB/s) - 'hadoop-3.3.6.tar.gz' saved [738107476/738107476]

root@kali: ~#
```

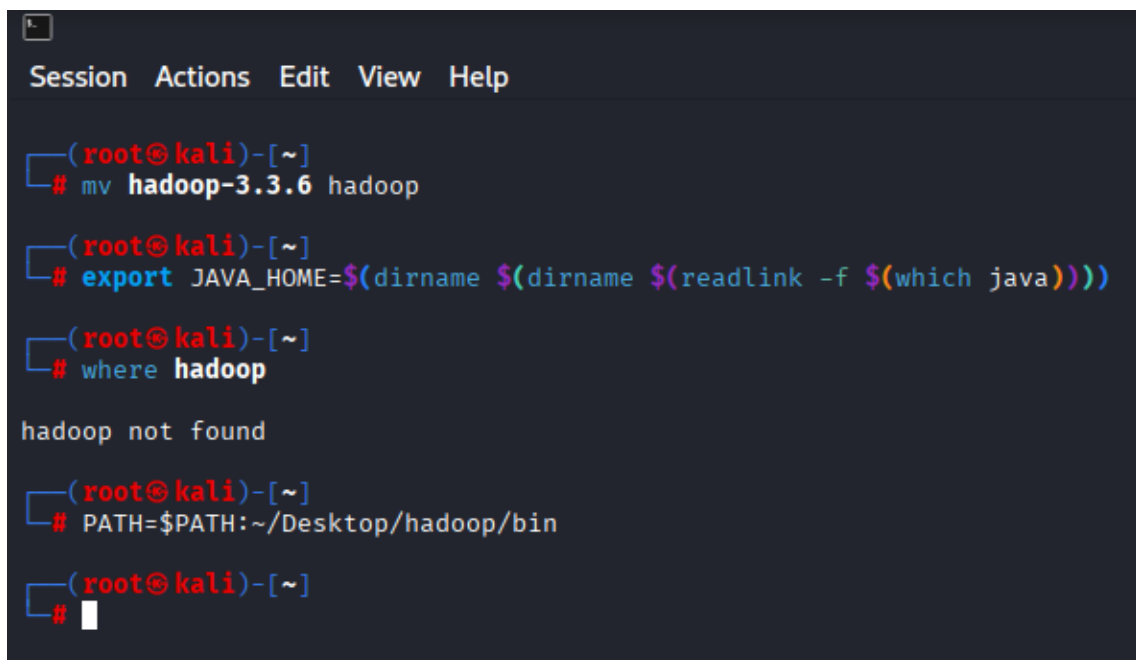
```
(root@kali)-[~]  
# tar -xvf hadoop-3.3.6.tar.gz  
hadoop-3.3.6/  
hadoop-3.3.6/NOTICE-binary  
hadoop-3.3.6/licenses-binary/  
hadoop-3.3.6/licenses-binary/LICENSE-lz4.txt  
hadoop-3.3.6/licenses-binary/LICENSE-zstd-jni.txt  
hadoop-3.3.6/licenses-binary/LICENSE-tree.txt  
hadoop-3.3.6/licenses-binary/LICENSE-azure-cosmosdb.txt  
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-leveldb.txt  
hadoop-3.3.6/licenses-binary/LICENSE-nvd3.txt  
hadoop-3.3.6/licenses-binary/LICENSE-leveldbjni.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
```



A terminal window with a dark background and a menu bar (Session, Actions, Edit, View, Help). The prompt is `(root@kali)-[~]`. The user enters `# mv hadoop-3.3.6 hadoop`. The prompt returns to `(root@kali)-[~]` with a cursor on the next line.

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

(root@kali)-[~]
#
```



A terminal window with a dark background and a menu bar (Session, Actions, Edit, View, Help). The prompt is `(root@kali)-[~]`. The user enters `# mv hadoop-3.3.6 hadoop`. The prompt returns to `(root@kali)-[~]`. The user enters `# export JAVA_HOME=$(dirname $(dirname $(readlink -f $(which java))))`. The prompt returns to `(root@kali)-[~]`. The user enters `# where hadoop`. The output is `hadoop not found`. The prompt returns to `(root@kali)-[~]`. The user enters `# PATH=$PATH:~/Desktop/hadoop/bin`. The prompt returns to `(root@kali)-[~]` with a cursor on the next line.

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

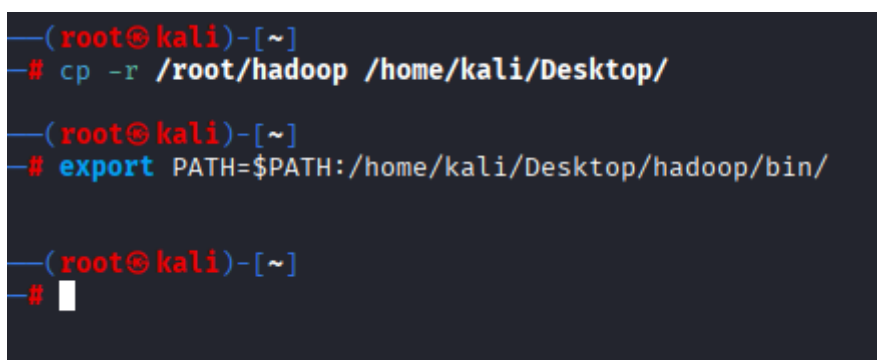
(root@kali)-[~]
# export JAVA_HOME=$(dirname $(dirname $(readlink -f $(which java))))

(root@kali)-[~]
# where hadoop

hadoop not found

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

(root@kali)-[~]
#
```

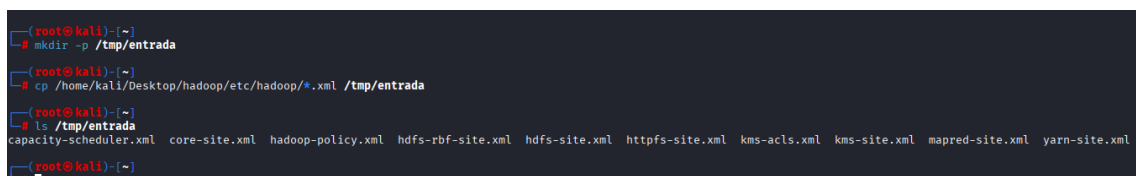


A terminal window with a dark background and a menu bar (Session, Actions, Edit, View, Help). The prompt is `(root@kali)-[~]`. The user enters `# cp -r /root/hadoop /home/kali/Desktop/`. The prompt returns to `(root@kali)-[~]`. The user enters `# export PATH=$PATH:/home/kali/Desktop/hadoop/bin/`. The prompt returns to `(root@kali)-[~]`. The user enters `#`. The prompt returns to `(root@kali)-[~]` with a cursor on the next line.

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

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

(root@kali)-[~]
#
```



A terminal window with a dark background and a menu bar (Session, Actions, Edit, View, Help). The prompt is `(root@kali)-[~]`. The user enters `# mkdir -p /tmp/entrada`. The prompt returns to `(root@kali)-[~]`. The user enters `# cp /home/kali/Desktop/hadoop/etc/hadoop/*.xml /tmp/entrada`. The prompt returns to `(root@kali)-[~]`. The user enters `# ls /tmp/entrada`. The output is `capacity-scheduler.xml core-site.xml hadoop-policy.xml hdfs-rbf-site.xml hdfs-site.xml https-site.xml kms-acls.xml kms-site.xml mapred-site.xml yarn-site.xml`. The prompt returns to `(root@kali)-[~]` with a cursor on the next line.

```
(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-rbf-site.xml  hdfs-site.xml  https-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/  
# ls  
hadoop-mapreduce-client-app-3.3.6.jar      hadoop-mapreduce-client-hs-3.3.6.jar      hadoop-mapreduce-client-jobclient-3.3.6-tests.jar  hadoop-mapreduce-client-uploader-3.3.6.jar  lib-examples  
hadoop-mapreduce-client-common-3.3.6.jar   hadoop-mapreduce-client-hq-plugins-3.3.6.jar  hadoop-mapreduce-client-mativetask-3.3.6.jar      hadoop-mapreduce-examples-3.3.6.jar        sources  
hadoop-mapreduce-client-core-3.3.6.jar     hadoop-mapreduce-client-jobclient-3.3.6.jar  hadoop-mapreduce-client-shuffle-3.3.6.jar          jdifff  
# cd /home/kali/Desktop/hadoop/share/hadoop/mapreduce/  
# jar tvf hadoop-mapreduce-examples-3.3.6.jar  
0 Sun Jun 18 08:51:36 EDT 2023 META-INF/  
186 Sun Jun 18 08:51:34 EDT 2023 META-INF/MANIFEST.MF  
0 Sun Jun 18 08:51:36 EDT 2023 org/apache/  
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$SolutionAcceptor.class  
6738 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/dancing/DistributedPentomino.class  
1377 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/dancing/Pentomino$SolutionCategory.class  
984 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/dancing/Sudoku$RowConstraint.class  
1930 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/dancing/DancingLinks$Node.class  
3160 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/dancing/Pentomino$Piece.class  
2459 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/dancing/OneSidedPentomino.class  
6799 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/dancing/Sudoku.class  
1082 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/dancing/DancingLinks$ColumnHeader.class  
605 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/dancing/Pentomino$Point.class  
9619 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/dancing/Pentomino.class  
1557 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/dancing/Pentomino$SolutionPrinter.class  
5112 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/dancing/DistributedPentomino$PentMap.class  
2109 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/dancing/Sudoku$SolutionPrinter.class  
919 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  
230 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/dancing/Sudoku$ColumnName.class  
975 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/dancing/Sudoku$SquareConstraint.class  
3603 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/dancing/DistributedPentomino$PentMap$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  
2250 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/WordCount.class  
2800 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/DBCount$PageView$PageviewReducer.class  
2406 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/WordCount$IntSumReducer.class  
2459 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/WordMean$WordMeanMapper.class  
5190 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$AggregateWordHistogramPlugin.class  
29535 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/RandomTextWriter.class  
2762 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/BaileysBrewinPluff$BbMapper.class  
764 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/terasort/Random6$RandomConstant.class  
1165 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/terasort/teraGen$Counters.class  
4155 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/terasort/teraValidate$ValidateMapper.class  
6938 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/terasort/terasort/teraInputFormat.class  
1512 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/terasort/terasort/teraScheduler$Split.class  
3534 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/terasort/teraValidate$ValidateReducer.class  
4273 Sun Jun 18 08:51:36 EDT 2023 org/apache/hadoop/examples/terasort/teraGen$SortGenMapper.class  
  
root@kali: /home/kali/Desktop/hadoop/share/hadoop/mapreduce  
Session Actions Edit View Help  
  
(root@kali)~# cd /home/kali/Desktop/hadoop/share/hadoop/mapreduce/  
# hadoop jar hadoop-mapreduce-examples-3.3.6.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 09:18:02,943 INFO impl.MetricsSystemImpl: JobTracker metrics system started  
2025-10-18 09:18:03,217 INFO input.FileInputFormat: Total input files to process : 10  
2025-10-18 09:18:03,217 INFO mapreduce.JobSubmitter: number of splits:10  
2025-10-18 09:18:03,503 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local1908251525_0001  
2025-10-18 09:18:03,503 INFO mapreduce.JobSubmitter: Executing with tokens: []  
2025-10-18 09:18:03,626 INFO mapreduce.Job: The url to track the job: http://localhost:8080/  
2025-10-18 09:18:03,627 INFO mapreduce.Job: Running Job: job_local1908251525_0001  
2025-10-18 09:18:03,630 INFO mapred.LocalJobRunner: OutputCommitter set in config null  
2025-10-18 09:18:03,635 INFO output.PathOutputCommitterFactory: No output committer 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 09:18:03,637 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup_temporary folders under output directory:false, ignore cleanup failures: false  
2025-10-18 09:18:03,638 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 09:18:03,674 INFO mapred.LocalJobRunner: Starting task: attempt_local1908251525_0001_m_0000000_0  
2025-10-18 09:18:03,699 INFO output.PathOutputCommitterFactory: No output committer factory defined, defaulting to FileOutputCommitterFactory  
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: FileOutputCommitter skip cleanup_temporary folders under output directory:false, ignore cleanup failures: false  
2025-10-18 09:18:03,750 INFO mapred.Task: Using ResourceCalculatorProcessTree : [ ]  
2025-10-18 09:18:03,755 INFO mapred.MapTask: Processing split: files/tmp/entrada/hadoop-policy.xml:0+11765  
2025-10-18 09:18:03,825 INFO mapred.MapTask: (EQUATOR) 0 kvi 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 83886080  
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 09:18:03,837 INFO mapred.MapTask: Map output collector class = org.apache.hadoop.mapred.MapTask$MapOutputBuffer  
2025-10-18 09:18:03,853 INFO mapred.LocalJobRunner:  
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 09:18:03,870 INFO mapred.MapTask: Finished split 0  
2025-10-18 09:18:03,892 INFO mapred.Task: Task:attempt_local1908251525_0001_m_0000000_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_0000000_0' done.  
2025-10-18 09:18:03,913 INFO mapred.Task: Final Counters for attempt_local1908251525_0001_m_0000000_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]
#
```