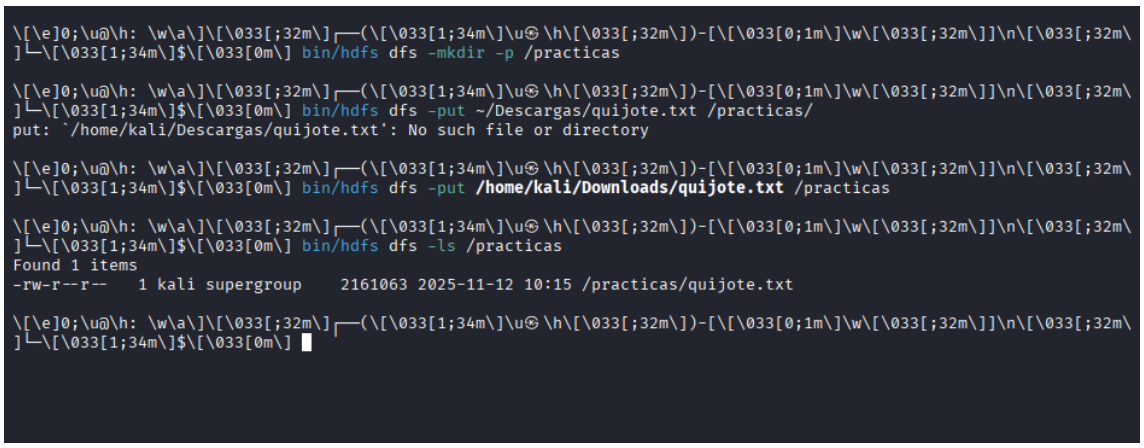
SOLUCION TAREA 5

Descargamos el quijote y lo subimos al HDFS



Contamos cuantas veces aparecen las palabras

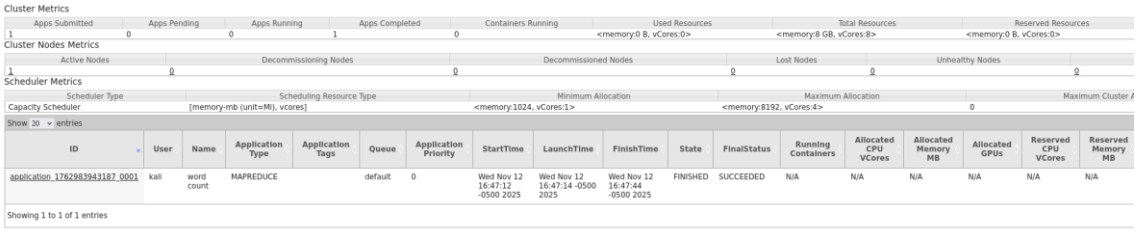
```
\[\e]0;\u@\h: \w\a\]\[\033[;32m\]┌──(\[\033[1;34m\]\u㉿\h\[\033[;32m\])-[\[\033[0;1m\]\w\[\033[;32m\]]\n\[\033[;32m\
/practicas/quijote.txt /practicas/resultado

2025-11-12 16:47:08,135 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at localhost/1
2025-11-12 16:47:08,812 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/stag
2025-11-12 16:47:09,981 INFO input.FileInputFormat: Total input files to process : 1
2025-11-12 16:47:10,603 INFO mapreduce.JobSubmitter: number of splits:1
2025-11-12 16:47:11,570 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1762983943187_0001
2025-11-12 16:47:11,570 INFO mapreduce.JobSubmitter: Executing with tokens: []
2025-11-12 16:47:11,997 INFO conf.Configuration: resource-types.xml not found
2025-11-12 16:47:11,998 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2025-11-12 16:47:12,795 INFO impl.YarnClientImpl: Submitted application application_1762983943187_0001
2025-11-12 16:47:13,261 INFO mapreduce.Job: The url to track the job: http://kali:8088/proxy/application_17629839431
2025-11-12 16:47:13,262 INFO mapreduce.Job: Running job: job_1762983943187_0001
2025-11-12 16:47:26,019 INFO mapreduce.Job: Job job_1762983943187_0001 running in uber mode : false
2025-11-12 16:47:26,020 INFO mapreduce.Job:  map 0% reduce 0%
2025-11-12 16:47:35,980 INFO mapreduce.Job:  map 100% reduce 0%
2025-11-12 16:47:43,411 INFO mapreduce.Job:  map 100% reduce 100%
2025-11-12 16:47:45,465 INFO mapreduce.Job: Job job_1762983943187_0001 completed successfully
2025-11-12 16:47:45,561 INFO mapreduce.Job: Counters: 54
        File System Counters
                FILE: Number of bytes read=605509
                FILE: Number of bytes written=1763629
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=2161171
                HDFS: Number of bytes written=448894
                HDFS: Number of read operations=8
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
                HDFS: Number of bytes read erasure-coded=0
        Job Counters
                Launched map tasks=1
                Launched reduce tasks=1
                Data-local map tasks=1
                Total time spent by all maps in occupied slots (ms)=7597
                Total time spent by all reduces in occupied slots (ms)=5039
                Total time spent by all map tasks (ms)=7597
                Total time spent by all reduce tasks (ms)=5039
                Total vcore-milliseconds taken by all map tasks=7597
                Total vcore-milliseconds taken by all reduce tasks=5039
                Total megabyte-milliseconds taken by all map tasks=7779328
                Total megabyte-milliseconds taken by all reduce tasks=5159936
        Map-Reduce Framework
                Map input records=37861
                Map output records=384260
                Map output bytes=3688599
                Map output materialized bytes=605509
                Input split bytes=108
                Combine input records=384260
                Combine output records=40059
```

Comprobamos resultados

```
\[\e]0;\u@\h: \w\a\]\[\033[;32m\]┌──(\[\033[1;34m\]u㉿\h\[\033[;32m\])-[\[\033[0;1m\]\w\[\033[;3
]└─\[\033[1;34m\]$\[\033[0m\] bin/hdfs dfs -ls /practicas/resultado
Found 2 items
-rw-r--r--   1 kali supergroup          0 2025-11-12 16:47 /practicas/resultado/_SUCCESS
-rw-r--r--   1 kali supergroup     448894 2025-11-12 16:47 /practicas/resultado/part-r-00000
```

```
\[\e]0;\u@\h: \w\a\]\[\033[;32m\]┌──(\[\033[1;34m\]u㉿\h\[\033[;32m\])-[\[\033[0;1m\]\w\[\033[;3
]└─\[\033[1;34m\]$\[\033[0m\] bin/hdfs dfs -cat /practicas/resultado/part-r-00000 | head -n 10
!Mal    1
"Al     1
"Cuando 2
"Cuidados       1
"De     2
"Defects,"      1
"Desnudo        1
"Dijo   1
"Dime   1
"Don    1
cat: Unable to write to output stream.
```



Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Used Resources	Total Resources	Reserved Resources
1	0	0	1	0	<memory:0 B, vCores:0>	<memory:8 GB, vCores:8>	<memory:0 B, vCores:0>

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes
1	0	0	0	0

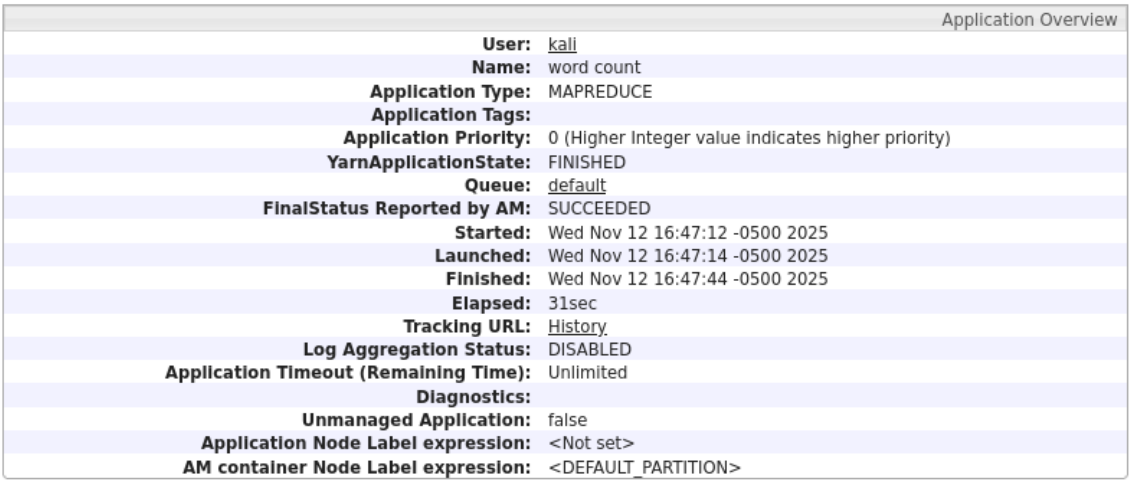
Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum Allocation	Maximum Cluster A
Capacity Scheduler	[memory-mb (unit=Mi), vcores]	<memory:1024, vCores:1>	<memory:8192, vCores:4>	0

Show 20 entries

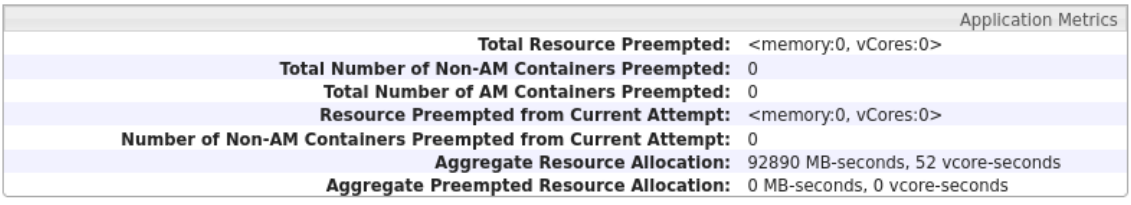
ID	User	Name	Application Type	Application Tags	Queue	Application Priority	StartTime	LaunchTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU VCores	Allocated Memory MB	Allocated GPUs	Reserved CPU VCores	Reserved Memory MB
application_1762983943187_0001	kali	word count	MAPREDUCE		default	0	Wed Nov 12 16:47:12 -0500 2025	Wed Nov 12 16:47:14 -0500 2025	Wed Nov 12 16:47:44 -0500 2025	FINISHED	SUCCEEDED	N/A	N/A	N/A	N/A	N/A	N/A

Showing 1 to 1 of 1 entries



Application Overview

User:	kali
Name:	word count
Application Type:	MAPREDUCE
Application Tags:	
Application Priority:	0 (Higher Integer value indicates higher priority)
YarnApplicationState:	FINISHED
Queue:	default
FinalStatus Reported by AM:	SUCCEEDED
Started:	Wed Nov 12 16:47:12 -0500 2025
Launched:	Wed Nov 12 16:47:14 -0500 2025
Finished:	Wed Nov 12 16:47:44 -0500 2025
Elapsed:	31sec
Tracking URL:	History
Log Aggregation Status:	DISABLED
Application Timeout (Remaining Time):	Unlimited
Diagnostics:	
Unmanaged Application:	false
Application Node Label expression:	<Not set>
AM container Node Label expression:	<DEFAULT_PARTITION>



Application Metrics

Total Resource Preempted:	<memory:0, vCores:0>
Total Number of Non-AM Containers Preempted:	0
Total Number of AM Containers Preempted:	0
Resource Preempted from Current Attempt:	<memory:0, vCores:0>
Number of Non-AM Containers Preempted from Current Attempt:	0
Aggregate Resource Allocation:	92890 MB-seconds, 52 vcore-seconds
Aggregate Preempted Resource Allocation:	0 MB-seconds, 0 vcore-seconds

Application Attempt Overview

Application Attempt State:	FINISHED
Started:	Wed Nov 12 16:47:12 -0500 2025
Elapsed:	31sec
AM Container:	container_1762983943187_0001_01_000001
Node:	kali:45513
Tracking URL:	History
Diagnostics Info:	
Nodes blacklisted by the application:	-
Nodes blacklisted by the system:	-

Total Allocated Containers: 3

Each table cell represents the number of NodeLocal/RackLocal/OffSwitch containers satisfied by NodeLocal/RackLocal/OffSwitch resource requests.

	Node Local Request	Rack Local Request	Off Switch Request
Num Node Local Containers (satisfied by)	1		
Num Rack Local Containers (satisfied by)	0	0	
Num Off Switch Containers (satisfied by)	0	0	2

Show 20 entries Search:

Container ID	Node	Container Exit Status	Logs
No data available in table			

Showing 0 to 0 of 0 entries First Previous Next Last

Respuestas breves

- **Número de map task y reduce tasks.**
  Se ejecutó 1 tarea Map y 1 tarea Reduce

- **HDFS bytes read y HDFS bytes written**
  Unos cientod de miles de bytes desde HDFS y escribió decenas de miles de bytes en salida

- **¿Qué fase tardo más (map/shuffle/reduce) y por qué crees que fue así?**
  Shuffle/Reduce ya que se transfiere los datos desde las tareas Map hasta el Reduce

- **¿Hubo reintentos? ¿En qué tarea?**
  No, no hubo fallos