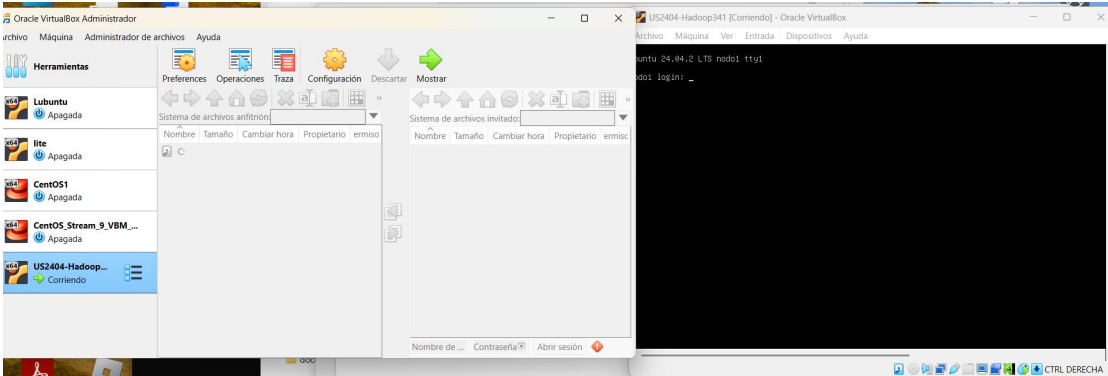


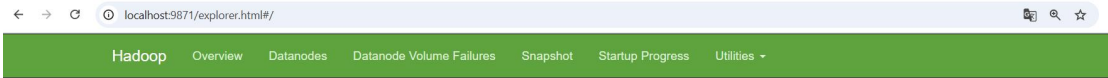
Parte 1: Preparación del Entorno

Configuración de la Máquina Virtual

1. Usa VirtualBox o VMware para crear una máquina virtual con Ubuntu Server.
2. Asigna al menos 4GB de RAM y 2 núcleos de CPU.
3. Configura una interfaz de red en modo "Bridge" o "NAT" con acceso a Internet.



Instala Hadoop y configúralo en modo pseudo-distribuido. Asegúrate de que los servicios de Hadoop estén en ejecución.



Browse Directory

/

Go!

Show

25

entries

Search:

<input type="checkbox"/>		Permission		Owner		Group		Size		Last Modified		Replication		Block Size		Name	
<input type="checkbox"/>		drwxr-xr-x		hadoop		supergroup		0 B		Mar 31 13:46		0		0 B		libros	
<input type="checkbox"/>		drwxr-xr-x		hadoop		supergroup		0 B		Apr 01 12:44		0		0 B		libros_salida	
<input type="checkbox"/>		drwxr-xr-x		hadoop		supergroup		0 B		Mar 31 12:57		0		0 B		prueba	
<input type="checkbox"/>		drwxr-xr-x		hadoop		supergroup		0 B		Mar 31 12:34		0		0 B		prueba2	
<input type="checkbox"/>		drwxr-xr-x		hadoop		supergroup		0 B		Apr 01 09:21		0		0 B		tmp	

Showing 1 to 5 of 5 entries

Previous


1

Next

Hadoop, 2024.

← → ↻

localhost:8089/cluster



All

Cluster

About

Nodes

Node Labels

Applications

NEW

NEW_SAVING

SUBMITTED

ACCEPTED

RUNNING

FINISHED

FAILED

KILLED

Scheduler

Tools

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Used F
1	0	0	1	0	<memory:0 B, vCores

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes
1	0	0

Scheduler Metrics

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

Show 20 entries

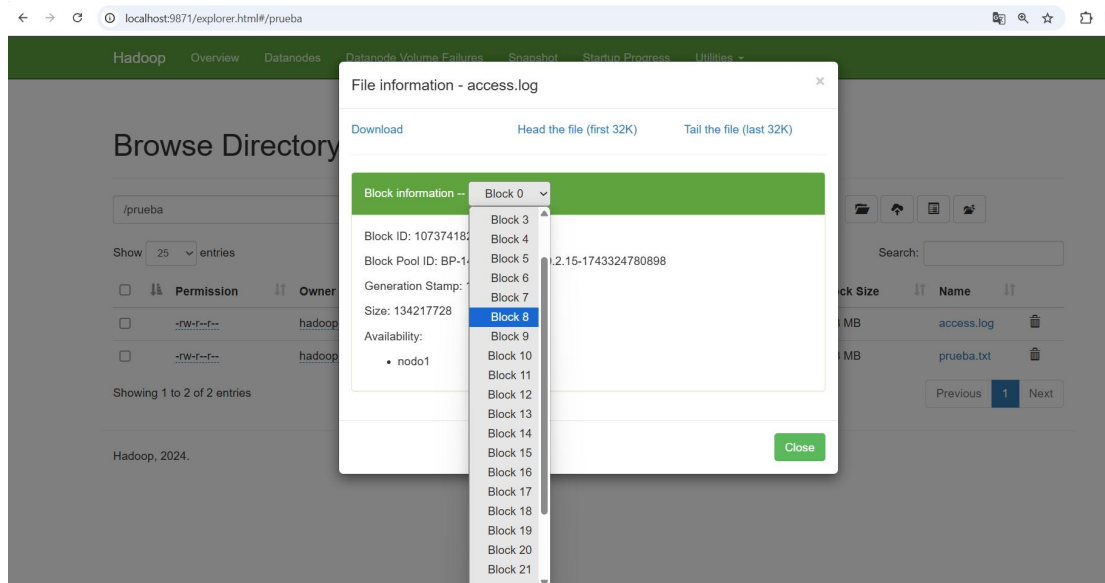
ID	User	Name	Application Type	Application Tags	Queue	Application Priority	StartTime	LaunchTime	FinishTime
application_1743507622987_0001	hadoop	word count	MAPREDUCE		root.default	0	Tue Apr 1 12:43:45 +0100 2025	Tue Apr 1 12:43:47 +0100 2025	Tue Apr 1 12:44:14 +0100 2025

Showing 1 to 1 of 1 entries

Parte 2: Procesamiento de Datos con Hadoop y Herramientas Relacionadas

Cargar y Procesar Datos en HDFS

1. Descarga un dataset grande (por ejemplo, logs de acceso a servidores o un dataset de Open Data).
- Copia los datos a HDFS ejecutando los comandos necesarios.



Ejecutar un Job de MapReduce

1. Ejecuta el programa de ejemplo **wordcount** sobre los datos almacenados en HDFS:

```
hadoop@nodo1: /opt/hadoop/hadoop-3.4.1/share/hadoop/mapreduce$ hadoop jar hadoop-mapreduce-examples-3.4.1.jar wordcount /libros /libros_salida
```

2. Verifica los resultados obtenidos ejecutando..

← → ↺

localhost:9871/explorer.html#/libros_salida

🔍 ⌕ ☆

Hadoop

Overview

Datanodes

Datanode Volume Failures

Snapshot

Startup Progress

Utilities -

Browse Directory

/libros_salida

Go!

📁 ↶ 📄 📧 📎

Show

25

entries

Search:

<input type="checkbox"/>	🔍	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
<input type="checkbox"/>		:-rw-r--r--	hadoop	supergroup	0 B	Apr 01 12:44	1	128 MB	_SUCCESS	🗑
<input type="checkbox"/>		:-rw-r--r--	hadoop	supergroup	220.76 KB	Apr 01 12:44	1	128 MB	part-r-00000	🗑

Showing 1 to 2 of 2 entries

Previous

1


Next

Hadoop, 2024.

← → ↺

localhost:8089/cluster

🔍 ⌕ ☆ 📄 📧 📎



All Applications

Cluster

About

Nodes

Node Labels

Applications

NEW

NEW_SAVING

SUBMITTED

ACCEPTED

RUNNING

FINISHED

FAILED

KILLED

Scheduler

Tools

Cluster Metrics

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

Cluster Nodes Metrics

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

Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum Allocation	Maximum Cluster Application Priority
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
application_1743507622987_0001	hadoop	word count	MAPREDUCE		root default	0	Tue Apr 1 12:43:45 +0100 2025	Tue Apr 1 12:43:47 +0100 2025	Tue Apr 1 12:44:14 +0100 2025	FINISHED	SUCCEEDED	N/A

Showing 1 to 1 of 1 entries