

Práctica: Configuración de un Pseudoclúster Hadoop

Objetivo:

Configurar un pseudoclúster de Hadoop en una sola máquina. En este tipo de configuración, todos los componentes de Hadoop (NameNode, DataNode, ResourceManager, NodeManager) se ejecutan en la misma máquina, lo que es ideal para pruebas y aprendizaje.

Pasos Generales:

1. Instalar Hadoop:

- Descarga la última versión de Hadoop desde el sitio oficial.
- Extrae los archivos y coloca Hadoop en un directorio adecuado en tu sistema (por ejemplo, /opt/hadoop).
- Configura las variables de entorno (HADOOP_HOME, HADOOP_CONF_DIR, PATH).

2. Configurar core-site.xml:

Este archivo contiene la configuración básica del sistema de archivos distribuido (HDFS). Es importante que configures el fs.defaultFS, que indica la dirección del NameNode, utilizando una dirección de tipo hdfs://localhost:9000 o similar.

3. Configurar hdfs-site.xml:

En este archivo se configura la ubicación de los directorios locales para el NameNode y el DataNode. Estos son los directorios donde Hadoop almacenará sus metadatos y datos distribuidos. Aquí se especifican también otros parámetros como la replicación de los bloques.

4. Crear Directorios en el Sistema Local:

Crea los directorios que has mencionado en los archivos de configuración (namenode y datanode) dentro de una carpeta central (como /datos). Esto es necesario para que Hadoop pueda almacenar los archivos del sistema de archivos distribuido.

5. Formatear el NameNode:

Antes de iniciar el clúster, debes formatear el NameNode. Este paso inicializa el sistema de archivos distribuido de Hadoop.

6. Iniciar los Servicios de Hadoop:

Inicia los servicios de Hadoop (NameNode, DataNode) utilizando el script de inicio correspondiente. Asegúrate de que el NameNode y los DataNodes estén corriendo en tu máquina.

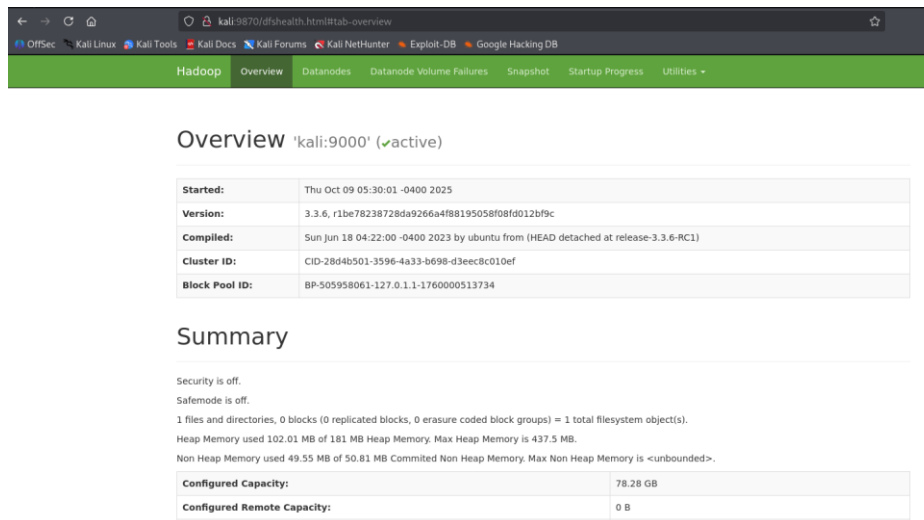
7. Verificar el Estado de los Servicios:

Puedes verificar que los servicios de Hadoop se estén ejecutando correctamente utilizando el comando jps, que te mostrará los procesos activos como NameNode, DataNode, y SecondaryNameNode.

8. Acceder a la Interfaz Web de Hadoop:

Una vez que el clúster esté en funcionamiento, puedes acceder a la interfaz web de

Hadoop a través de tu navegador, normalmente disponible en <http://localhost:9870>. En esta página podrás ver el estado del clúster y de los nodos.



The screenshot shows a web browser window with the address bar displaying `kali:9870/dfshealth.html#tab=overview`. The browser's tab bar includes links for OffSec, Kali Linux, Kali Tools, Kali Docs, Kali Forums, Kali NetHunter, Exploit-DB, and Google Hacking DB. The page has a green navigation bar with tabs for Hadoop, Overview, Datanodes, Datanode Volume Failures, Snapshot, Startup Progress, and Utilities. The main content area is titled "Overview 'kali:9000' (active)".

Overview 'kali:9000' (active)

Started:	Thu Oct 09 05:30:01 -0400 2025
Version:	3.3.6, r1be78238728da9266a4f88195058f08fd012b9c
Compiled:	Sun Jun 18 04:22:00 -0400 2023 by ubuntu from (HEAD detached at release-3.3.6-RC1)
Cluster ID:	CID-28d4b501-3596-4a33-b698-d3eecd8c010ef
Block Pool ID:	BP-505958061-127.0.1.1-1760000513734

Summary

Security is off.
Safemode is off.
1 files and directories, 0 blocks (0 replicated blocks, 0 erasure coded block groups) = 1 total filesystem object(s).
Heap Memory used 102.01 MB of 181 MB Heap Memory. Max Heap Memory is 437.5 MB.
Non Heap Memory used 49.55 MB of 50.81 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.

Configured Capacity:	78.28 GB
Configured Remote Capacity:	0 B

9. Probar la Carga de Archivos en HDFS:

Realiza una prueba básica cargando un archivo al sistema de archivos distribuido de Hadoop (HDFS) y verificando que puedas acceder a él.

Consejos:

- Cada uno de estos pasos puede implicar la modificación de archivos de configuración clave dentro de la carpeta de configuración de Hadoop (`/opt/hadoop/etc/hadoop`).
- Si no estás seguro de algún archivo de configuración, consulta la documentación de Hadoop o busca ejemplos de configuraciones comunes.
- Asegúrate de que los directorios locales sean accesibles y tengan los permisos correctos para el usuario que ejecuta Hadoop.