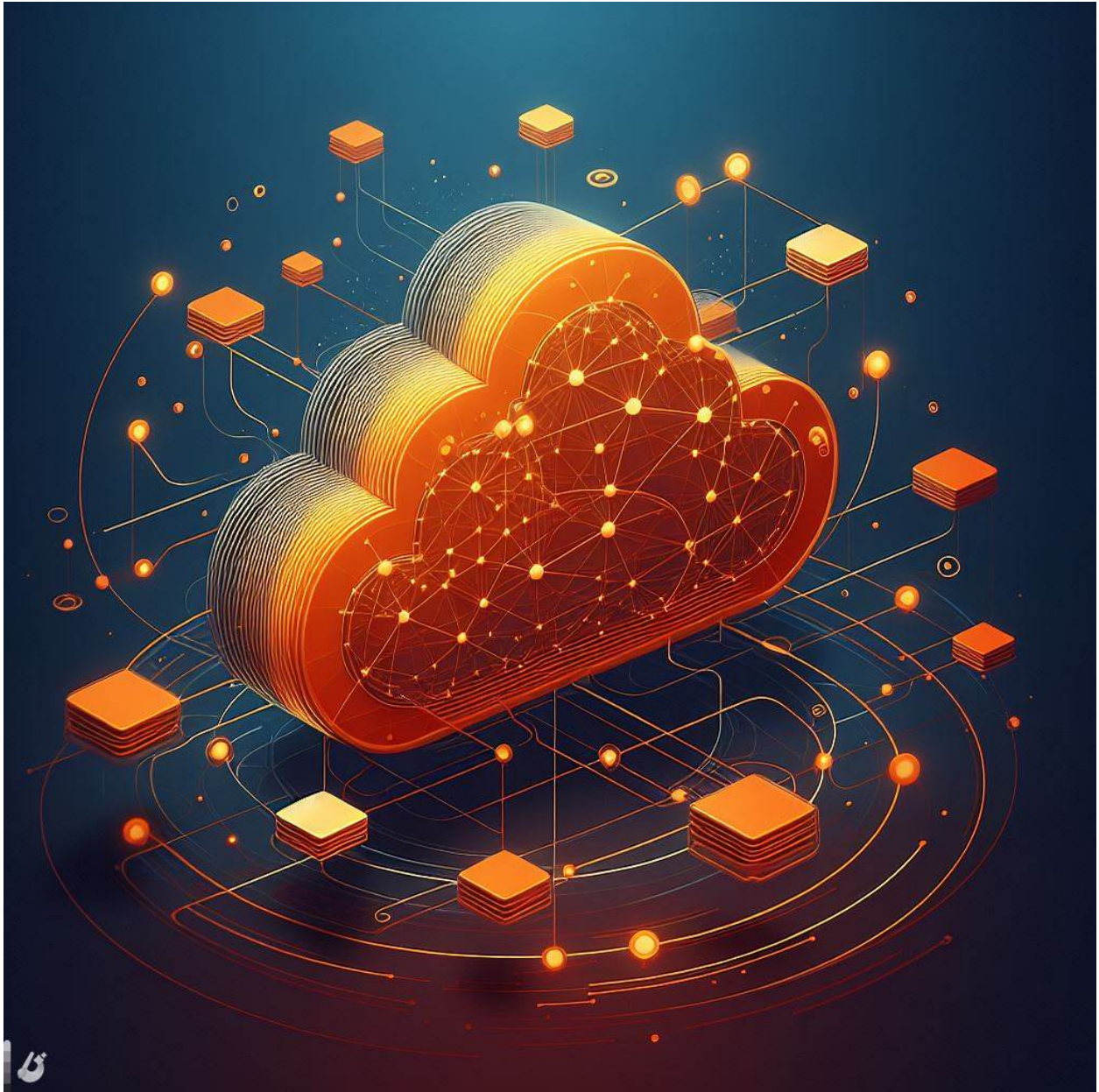


# Modos de funcionamiento de Cloudera

---



---

## Investigación

### 1. ¿Cuáles son los modos de funcionamiento disponibles en Cloudera?

Los modos de funcionamiento disponibles en Cloudera son:

- Standalone
- Pseudo-Distribuido
- Distribuido

### 2. ¿Cuáles son las características clave de cada modo de funcionamiento?

- **Modo standalone:**

- Se ejecuta en un único nodo como un solo proceso Java.
- Es útil para la depuración.
- No se necesita ninguna configuración especial.

- **Modo pseudo-distribuido:**

- Se ejecuta en un único nodo.
- Cada demonio Hadoop se ejecuta en un proceso Java separado.
- Este modo es útil para el desarrollo y para probar la funcionalidad de Hadoop.

- **Modo distribuido:**

- Este es el modo de producción.
- Hadoop se ejecuta en un clúster de múltiples nodos.
- Permite que múltiples tipos de cargas de trabajo analíticas se lleven a cabo con los mismos datos a la vez, a gran escala.

### 3. ¿Para qué tipos de cargas de trabajo se recomienda cada modo?

- **Modo standalone**, ideal para la depuración y el desarrollo de aplicaciones.
- **Modo pseudo-distribuido**, útil para el desarrollo y las pruebas de funcionalidad de Hadoop.

- 
- **Modo distribuido**, es el modo de producción y es capaz de manejar grandes volúmenes de datos y múltiples tipos de cargas de trabajo analíticas simultáneamente.

4. ¿Cuáles son las ventajas y desventajas de cada modo?

- **Modo standalone:**
  - **Ventajas:** Es útil para la depuración y el desarrollo de aplicaciones. No se necesita ninguna configuración especial.
  - **Desventajas:** No es adecuado para el acceso a datos de baja latencia. No puede almacenar una gran cantidad de archivos pequeños de manera eficiente.
- **Modo pseudo-distribuido:**
  - **Ventajas:** Útil para el desarrollo y las pruebas de funcionalidad de Hadoop. Cada demonio Hadoop se ejecuta en un proceso Java separado.
  - **Desventajas:** No es adecuado para el acceso a datos de baja latencia. No puede almacenar una gran cantidad de archivos pequeños de manera eficiente.
- **Modo distribuido:**
  - **Ventajas:** Este es el modo de producción y es capaz de manejar grandes volúmenes de datos y múltiples tipos de cargas de trabajo analíticas simultáneamente. Hadoop distribuye datos a través de los clústeres de computadoras disponibles para completar las tareas de almacenamiento y cálculo. Estos clústeres se pueden expandir fácilmente a miles de nodos y tienen una alta escalabilidad.
  - **Desventajas:** Hadoop no es adecuado para el acceso a datos de baja latencia. No puede almacenar una gran cantidad de archivos pequeños de manera eficiente.

---

# Actividades Prácticas

## Standalone

- Iniciar Cloudera en modo Standalone

```
<configuration>
  <property>
    <name>fs.defaultFS</name>
    <value>file://localhost:8020</value>
  </property>
  ...
</configuration>
```

```
[cloudera@quickstart ~]$ hdfs dfs -ls
Found 50 items
-rw-r--r-- 1 cloudera cloudera 620 2023-12-05 10:12 .ICEauthority
-rw-r--r-- 1 cloudera cloudera 52 2023-11-22 12:09 .bash_history
-rw-r--r-- 1 cloudera cloudera 18 2017-10-23 09:14 .bash_logout
-rw-r--r-- 1 cloudera cloudera 176 2017-10-23 09:14 .bash_profile
-rw-r--r-- 1 cloudera cloudera 176 2017-10-23 09:14 .bashrc
drwxr-xr-x - cloudera cloudera 4096 2023-11-21 09:00 .cache
drwxrwxr-x - cloudera cloudera 4096 2023-11-21 09:06 .config
drwx----- - cloudera cloudera 4096 2023-11-21 09:00 .dbus
-rw-r--r-- 1 cloudera cloudera 16 2023-11-21 09:00 .esd_auth
drwxr-xr-x - cloudera cloudera 4096 2023-11-21 09:00 .fontconfig
drwx----- - cloudera cloudera 4096 2023-12-05 10:12 .gconf
drwx----- - cloudera cloudera 4096 2023-12-05 10:13 .gconfd
drwxr-xr-x - cloudera cloudera 4096 2023-11-21 09:00 .gnome2
drwx----- - cloudera cloudera 4096 2023-11-21 09:00 .gnome2_private
drwxrwxr-x - cloudera cloudera 4096 2023-11-21 09:02 .gstreamer-0.10
-rw-rw-r-- 1 cloudera cloudera 152 2023-12-05 10:12 .gtk-bookmarks
drwx----- - cloudera cloudera 4096 2023-11-21 09:00 .gvfs
drwxr-xr-x - cloudera cloudera 4096 2023-11-21 09:00 .local
drwxrwxr-x - cloudera cloudera 4096 2017-10-23 15:55 .mozilla
drwxr-xr-x - cloudera cloudera 4096 2023-11-21 09:00 .nautilus
drwx----- - cloudera cloudera 4096 2023-11-21 09:00 .pulse
-rw-r--r-- 1 cloudera cloudera 256 2023-11-21 09:00 .pulse-cookie
-rw-r--r-- 1 cloudera cloudera 5 2023-12-05 10:12 .vboxclient-clipboard.pid
-rw-r--r-- 1 cloudera cloudera 5 2023-12-05 10:12 .vboxclient-display.pid
-rw-r--r-- 1 cloudera cloudera 5 2023-12-05 10:12 .vboxclient-draganddrop.pid
-rw-r--r-- 1 cloudera cloudera 5 2023-12-05 10:12 .vboxclient-seamless.pid
-rw-r--r-- 1 cloudera cloudera 928 2023-12-05 10:20 .xsession-errors
-rw-r--r-- 1 cloudera cloudera 1993 2023-12-05 10:05 .xsession-errors.old
drwxrwxr-x - cloudera cloudera 4096 2023-11-21 09:04 Desktop
drwxrwxr-x - cloudera cloudera 4096 2017-10-23 15:55 Documents
drwxr-xr-x - cloudera cloudera 4096 2023-11-21 09:00 Downloads
drwxr-xr-x - cloudera cloudera 4096 2023-11-21 09:00 Music
drwxr-xr-x - cloudera cloudera 4096 2023-11-21 09:00 Pictures
drwxr-xr-x - cloudera cloudera 4096 2023-11-21 09:00 Public
drwxr-xr-x - cloudera cloudera 4096 2023-11-21 09:00 Templates
drwxr-xr-x - cloudera cloudera 4096 2023-11-21 09:00 Videos
-rw-rw-r-- 1 cloudera cloudera 0 2023-12-05 09:43 archivo.txt
-rwxrwxr-x 1 cloudera cloudera 5387 2017-10-23 15:55 cloudera-manager
-rwxrwxr-x 1 cloudera cloudera 9964 2017-10-23 15:55 cm_api.py
drwxrwxr-x - cloudera cloudera 4096 2015-02-19 00:26 eclipse
-rw-rw-r-- 1 cloudera cloudera 53655 2017-10-23 15:55 enterprise-deployment.json
-rw-rw-r-- 1 cloudera cloudera 50515 2017-10-23 15:55 express-deployment.json
-rwxrwxr-x 1 cloudera cloudera 5007 2017-10-23 15:55 kerberos
drwxrwxr-x - cloudera cloudera 4096 2017-10-23 15:55 lib
-rw-rw-r-- 1 cloudera cloudera 0 2023-11-22 12:44 local_file.txt
```

- Crear un directorio en HDFS.

```
[cloudera@quickstart ~]$ hdfs dfs -mkdir standalone
[cloudera@quickstart ~]$ hdfs dfs -ls | grep standalone
drwxrwxr-x   - cloudera cloudera      4096 2023-12-05 10:18 standalone
[cloudera@quickstart ~]$
```

- Cargar un archivo en HDFS.

```
[cloudera@quickstart ~]$ hdfs dfs -copyFromLocal standalone-file.txt standalone/
[cloudera@quickstart ~]$ hdfs dfs -ls standalone
Found 1 items
-rw-r--r--    1 cloudera cloudera      0 2023-12-05 10:21 standalone/standalone-file.txt
```

## Pseudo-Distribuido

- Configurar Cloudera en modo Pseudo-Distribuido.

```
<configuration>
  <property>
    <name>fs.defaultFS</name>
    <value>hdfs://quickstart.cloudera:8020</value>
  </property>
```

- Crear un directorio en HDFS.

```
[cloudera@quickstart conf]$ hdfs dfs -mkdir directorio
[cloudera@quickstart conf]$ hdfs dfs -ls
Found 1 items
drwxr-xr-x   - cloudera cloudera      0 2023-12-05 09:38 directorio
[cloudera@quickstart conf]$
```

- Cargar un archivo en HDFS.

```
[cloudera@quickstart ~]$ hdfs dfs -copyFromLocal archivo.txt directorio/
[cloudera@quickstart ~]$ hdfs dfs -ls directorio
Found 1 items
-rw-r--r--    1 cloudera cloudera      0 2023-12-05 09:45 directorio/archivo.txt
[cloudera@quickstart ~]$
```