

Prácticas de HDFS. Interfaz Web.

Descripción

El alumno deberá familiarizarse con la interfaz web.

Solo se requieren los conceptos mostrados en la sesión teórica para completar el ejercicio.

Sigue las instrucciones del profesor para inicializar el entorno.

Explorar el interfaz web

Existen varias opciones web disponibles para monitorizar nuestro HDFS. Vamos

a familiarizarnos con algunas de ellas.

Interfaz legacy

Si miramos la configuración de hadoop, en /etc/hadoop/conf, veremos el fichero hdfs-site.

xml.

En él encontraremos una propiedad de configuración llamada dfs.namenode.http-address.

Ahí se indica por qué puerto está dando servicio el interfaz web del namenode.

[**Nota:** Como hemos visto, en este fichero se sobrescriben propiedades definidas en hdfs-default.xml. En el caso de no haber personalizado el puerto, puede consultarse el valor por defecto de la propiedad en [hdfs-default.xml](#): 0.0.0.0:50070]

```
...  
<property>  
  <name>dfs.namenode.httpaddress</name>  
  <value>quickstart.cloudera:50070</value>  
</property>  
...
```

En esa dirección se encuentra una sencilla consola web de monitorización que vamos a tener disponible con cualquier versión de hadoop.

Las más antiguas, al acceder al puerto 50070 del namenode, mostrarán este interfaz (llamado ahora legacy):

<http://quickstart.cloudera:50070/dfshealth.jsp>

1. Comprueba la capacidad configurada en HDFS, qué cantidad se está utilizando, qué cantidad está aún disponible (en valores absolutos y porcentuales).
2. Consulta cuántos nodos de trabajo tiene el cluster. Cuántos están vivos y cuántos muertos.

3. Consulta en qué ruta local al namenode se están escribiendo los metadatos.
4. Accede a la lista de nodos para verlos uno a uno.
5. Accede a la lista de logs del namenode (comprobarás, al estar en una instalación pseudo-distribuida, que no solo aparecen los logs del namenode, sino otros muchos).
6. Navega por el File System.

Interfaz legacy

En la versión instalada, el interfaz de esta consola web es algo diferente. Para acceder basta con ir a la raíz de la web.

<http://quickstart.cloudera:50070/>

Comprueba que están disponibles las mismas opciones que en el interfaz legacy, aunque ligeramente recolocadas.

7. Comprueba la capacidad configurada en HDFS, qué cantidad se está utilizando, qué cantidad está aún disponible (en valores absolutos y porcentuales).
8. Consulta cuántos nodos de trabajo tiene el cluster. Cuántos están vivos y cuántos muertos.
9. Consulta en qué ruta local al namenode se están escribiendo los metadatos.
10. Accede a la lista de nodos para verlos uno a uno.
11. Accede a la lista de logs del namenode (comprobarás, al estar en una instalación pseudo-distribuida, que no solo aparecen los logs del namenode, sino otros muchos).
12. Navega por el File System.

HDFS Manager de HUE

HUE centraliza el acceso a muchas funciones del ecosistema hadoop, si bien es cierto que podría no estar disponible en función de la distribución, la instalación y la versión de hadoop.

En este caso si la tenemos disponible y también conviene familiarizarse con ella.

Nos centramos en la parte que atañe a esta asignatura, y accedemos a la funcionalidad "Manage HDFS" disponible en un icono hacia la esquina superior derecha de la página principal o en la URL:

<http://quickstart.cloudera:8888/filebrowser/>

Comprueba cómo esta herramienta permite más opciones que las que permitía "Browse File System" en la consola por defecto.

1. Navega por el file system a través del menú de migas.
2. Pre visualiza alguno de los ficheros existentes. Pagina sobre ellos.
3. Comprueba que puedes descargarlos a local utilizando el botón download.
4. Comprueba en un directorio donde haya varios ficheros como puedes restringir los ficheros mostrados utilizando la caja de búsqueda "Search for file name".
 1. Entra en /user/cloudera/data
 2. Escribe quijote en la caja de texto
5. Haz una copia de algún fichero usando el botón copy. ¿Qué ocurre si no tienes permiso para ello? Lo verás si olvidas poner como destino una ruta correcta.
 1. Entra en /user/cloudera/data
 2. Selecciona el fichero access_log
 3. Pulsa el botón Copy
 4. Introduce Destination: /user/cloudera/data/access_log_copy
6. Elimina el fichero anterior usando el botón "Move to Trash".
7. Comprueba el contenido de la papelera pulsando el botón "Trash".
8. Restaura el fichero borrado usando la opción "Restore".
9. Elimínalo para siempre (sin usar la papelera) con el botón desplegable junto a "Move to Trash", eligiendo la opción "Delete forever".