



UNIVERSIDAD DE CASTILLA-LA MANCHA
ESCUELA SUPERIOR DE INFORMÁTICA

Planificación y Gestión de Infraestructuras TIC
<Práctica Vagrant y Ansible>

Autores: Miguel de la Cal Bravo y Félix Ángel Martínez Muela

Titulación: Máster en Ingeniería Informática

Fecha: 26/03/2020



ÍNDICE DE CONTENIDOS

1. Introducción	3
2. Objetivos	3
3. Vagrant y Ansible	3
4. Verificación de la instalación	4
5. Problemas encontrados.....	5
6. Conclusiones	6



1. INTRODUCCIÓN

En esta práctica se lleva a cabo el aprovisionamiento de un box de Vagrant, haciendo uso de la herramienta de automatización Ansible, con el software Cloudera Distribution Hadoop, CDH5.

En el presente documento, se detallan los desarrollos y procedimientos realizados, así como una verificación final del proceso de automatización.

2. OBJETIVOS

El objetivo principal consiste en el aprovisionamiento de una máquina virtualizada con Vagrant y VirtualBox, con el software de CDH5 con MRv1.

Otro objetivo trata de la verificación final de la instalación de software necesario, viendo que el proceso de automatización de ha completado correctamente.

3. VAGRANT Y ANSIBLE

Para el aprovisionamiento utilizaremos Vagrant junto con VirtualBox para la virtualización de una máquina Ubuntu 12.04 (Precise) de 64-bits. El nombre del box utilizado es *ubuntu/precise64*, ya que no todos los boxes de Ubuntu sirven en esta práctica.

Partimos del desarrollo de un *Vagrantfile*, que nos automatiza la creación de la máquina virtual. Las opciones más interesantes de este fichero de configuración son las siguientes:

- Red privada, máquina con dirección IP 192.168.100.100
- Ansible como herramienta para el aprovisionamiento
 - Playbook llamado “*cloudera_ubuntu.yml*”
 - Inventario de Ansible “*inventory.ini*”
- Carpeta sincronizada con archivos para el ejemplo de Hadoop.

NOTA: Además, también logramos el aprovisionamiento sobre un box de CentOS 6 en un inicio, aunque no era lo que se pedía (se adjunta el playbook correspondiente en la entrega del trabajo).

En cuanto al playbook de Ansible utilizado, este se encuentra comentado paso a paso, realizando los diferentes pasos de instalación o configuración de forma secuencial.

Una vez tengamos todo configurado, simplemente lanzaremos el siguiente comando en el directorio raíz de nuestro proyecto y empezará el aprovisionamiento de la máquina:

```
$ vagrant up
```

Una vez levantamos la máquina con Vagrant, esta se creará y aprovisionará utilizando Ansible, haciendo uso del playbook e inventario señalados anteriormente.



4. VERIFICACIÓN DE LA INSTALACIÓN

Una vez aprovisionada nuestra máquina, es hora de comprobar que todo se haya realizado correctamente.

Para ello, podemos dirigirnos al navegador, poniendo la IP de la máquina aprovisionada, seguida de :50070, que es el puerto utilizado por el servicio que hemos instalado.

En nuestro caso, nos dirigimos a la URL: <http://192.168.100.100:50070>

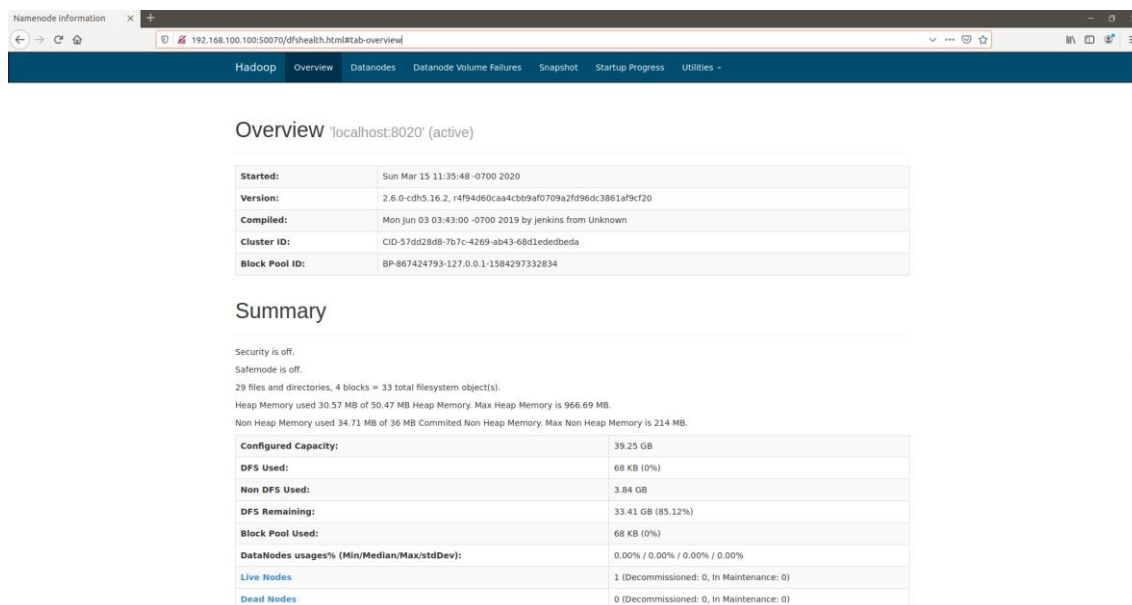


FIGURA 1. INTERFAZ WEB DE HADOOP EN EL PUERTO 50070

Como podemos observar en la Figura 1, tenemos acceso a la interfaz web de Hadoop, habiendo completado el proceso de aprovisionamiento satisfactoriamente.

Otra manera de verificar la instalación es mediante el siguiente comando, que nos devuelve si el paquete de Hadoop ha sido instalado correctamente:

```
$ dpkg -L hadoop-0.20-conf-pseudo
/.
/etc/etc/hadoop
/etc/hadoop/conf.pseudo.mr1
/etc/hadoop/conf.pseudo.mr1/README
/etc/hadoop/conf.pseudo.mr1/core-site.xml
/etc/hadoop/conf.pseudo.mr1/hadoop-metrics.properties
/etc/hadoop/conf.pseudo.mr1/hdfs-site.xml
/etc/hadoop/conf.pseudo.mr1/mapred-site.xml
/etc/hadoop/conf.pseudo.mr1/log4j.properties
/etc/hadoop/conf.pseudo.mr1/fair-scheduler.xml
```



```
/etc/hadoop/conf.pseudo.mr1/hadoop-env.sh
/usr
/usr/share
/usr/share/doc
/usr/share/doc/hadoop-0.20-conf-pseudo
/usr/share/doc/hadoop-0.20-conf-pseudo/copyright
/usr/share/doc/hadoop-0.20-conf-pseudo/changelog.Debian.gz
```

LISTADO 1. VERIFICACIÓN ALTERNATIVA DE LA INSTALACIÓN DE HADOOP

Además, también era necesaria la instalación de Java para la posterior ejecución de ejemplos jar, para lo cual podemos comprobar que también ha sido instalado mediante el comando:

```
$ java -version
java version "1.7.0_121"
OpenJDK Runtime Environment (IcedTea 2.6.8) (7u121-2.6.8-1ubuntu0.12.04.3)
OpenJDK 64-Bit Server VM (build 24.121-b00, mixed mode)
```

LISTADO 2. VERIFICACIÓN DE LA INSTALACIÓN DE OPENJDK 7

De esta forma ya tenemos aprovisionada nuestra máquina Ubuntu con todo el software necesario, y funcionando perfectamente.

Hecho esto, estaríamos preparados para realizar trabajos de Hadoop sobre la máquina aprovisionada, evitándonos todos los pasos previos de instalaciones y configuraciones.

5. PROBLEMAS ENCONTRADOS

Uno de los mayores problemas a los que nos hemos enfrentado surgió al principio de esta práctica, cuando tuvimos que elegir un box de Vagrant válida para la instalación de CDH5.

También tuvimos problemas con la instalación de Oracle JDK, ya que se nos pedía iniciar sesión para descargar el software, por lo que finalmente optamos por descargar e instalar la versión libre llamada Open JDK sin tener ningún problema.

En cuanto al uso de Ansible y Vagrant, tuvimos algún problema con el primero de los mencionados, ya que muchos de los comandos para la configuración de Hadoop no tienen un módulo específico en Ansible, por lo que tuvimos que terminar optando por usar los módulos de command o shell según era necesario (entre los cuales también tuvimos algún error de comprensión de uso para diferenciarlos en un inicio).



6. CONCLUSIONES

A modo de conclusión, esta práctica nos ha parecido más complicada de lo que en un inicio pensábamos. Al tratarse de una versión un poco antigua de CDH (ya va por la versión 7.1, en el momento que esta práctica ha sido realizada), algunos requisitos previos nos han dificultado el progreso de la automatización posterior de la instalación del software.

Sin embargo, ha sido interesante automatizar un proceso como éste en el que se realizan diferentes pasos, para finalmente acabar pulsando la tecla “Enter“ y todo se realice con Vagrant y Ansible de una forma muy sencilla y eficiente, pudiendo así empezar a trabajar con Hadoop de inmediato evitando llevar a cabo todos los pasos previos para su correcto funcionamiento.