

Programación distribuida

Diplomatura en Big Data

Instructivo para la inicialización de servicios en la máquina virtual

Introducción:

La máquina virtual (VM) parte de una imagen (curso-hadoop-spark.ova) Oracle Virtual Appliance.

El formato de exportación es compatible con VMWare, pero como la VM fue creada con [Oracle VirtualBox](http://OracleVirtualBox) y probada en virtual box.

La VM contiene:

- Hadoop (HDFS y Yarn)
- Spark (configurado con conexión al HDFS)
- Zookeeper
- Hive
- Java 8 (con maven, dependencias para trabajar con HDFS y zookeeper)
- eclipse

Importación de la VM

Seguir los siguientes pasos para importar en VirtualBox.

Aclaración: el tutorial fue basado en un SO Ubuntu 18.04, por lo que pueden existir leves diferencias en las interfases gráficas.

1. Descargar la imagen de la VM del siguiente link: curso-hadoop-spark.ova
2. Iniciar el proceso de importación de una VM



3. Seleccionar el archivo **.ova** de la ubicación en la que fue guardada

Programación distribuida

Diplomatura en Big Data

Appliance to import

Please choose the source to import appliance from. This can be a local file system to import OVF archive or one of known cloud service providers to import cloud VM from.

Source:

Please choose a file to import the virtual appliance from. VirtualBox currently supports importing appliances saved in the Open Virtualization Format (OVF). To continue, select the file to import below.

File:

Choose a virtual appliance file to import...

4. Siguiendo paso

5. Configurar la importación de la VM

Virtual System 1

Name	curso-hadoop 1
Guest OS Type	Ubuntu (64-bit)
CPU	1
RAM	2048 MB
DVD	<input checked="" type="checkbox"/>
USB Controller	<input checked="" type="checkbox"/>
Sound Card	<input checked="" type="checkbox"/> ICH AC97
Network Adapter	<input checked="" type="checkbox"/> Intel PRO/1000 MT Desktop (82540EM)
Storage Controller (IDE)	PIIX4
Storage Controller (IDE)	PIIX4
Storage Controller (SATA)	AHCI
Virtual Disk Image	curso-hadoop-hive-disk1.vmdk
Base Folder	/home/viking/VirtualBox VMs
Primary Group	/

Machine Base Folder:

MAC Address Policy:

Additional Options: ☐ Import hard drives as VDI

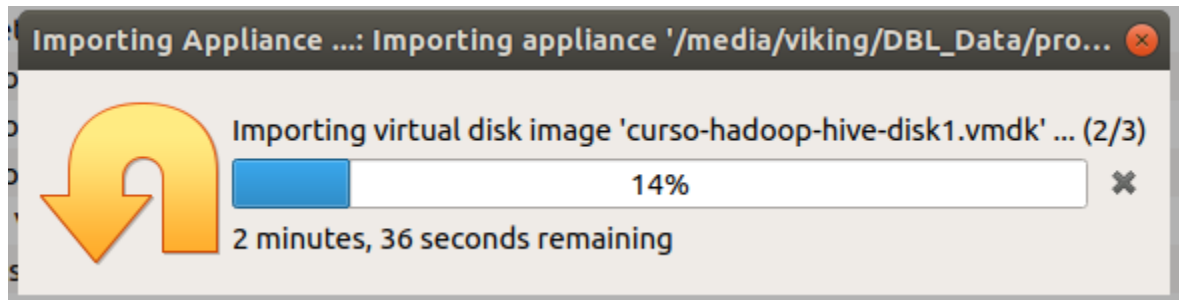
Appliance is not signed

- Es recomendable importar los discos duros virtuales en el formato original.
- Estas opciones pueden ser modificadas una vez importada la VM.
- La cantidad de RAM mínima para ejecutar todos los procesos es de 4GB

Programación distribuida

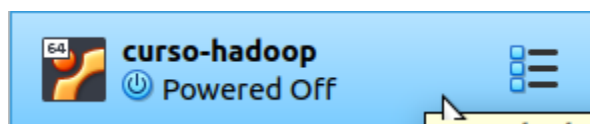
Diplomatura en Big Data

6. Esperar a que finalice la importación

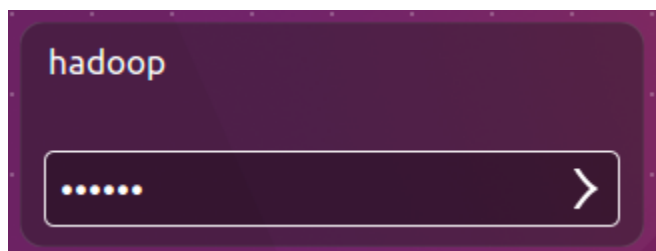


Inicialización de los servicios

1. Inicializar la VM mediante un doble click o pulsando Start

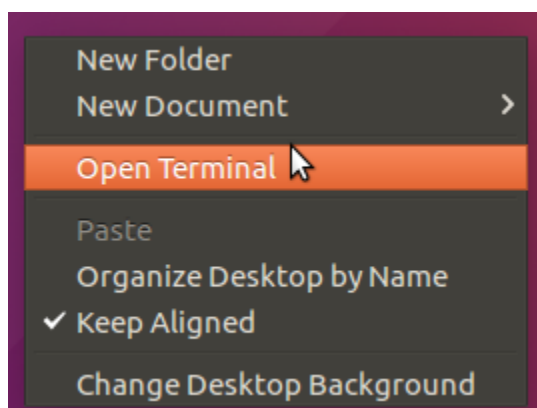


2. Iniciar sesión



- a. User: hadoop
- b. Pass: hadoop

3. Abrir la consola: click derecho



Programación distribuida

Diplomatura en Big Data

4. Dentro de la consola posicionarse en la carpeta **~/Desktop/curso**

```
hadoop@hadoop-ml: ~  
hadoop@hadoop-ml:~$ cd Desktop/curso/
```

```
cd ~/Desktop/curso
```

En ese directorio se encuentran todos los binarios que se utilizarán a lo largo del curso. Dentro de la carpeta **bin** se encuentran 2 scripts encargados de iniciar y finalizar las aplicaciones.

- a. hadoop
- b. yarn
- c. spark
- d. zookeeper
- e. eclipse

Para inicializar una de las aplicaciones ejecutar

```
./bin/start.sh $NOMBRE_APP
```

Por ejemplo, para inicializar **zookeeper** se debe ejecutar

```
./bin/start.sh zookeeper
```

En caso de no especificar el nombre del proceso, se inicializar todos de forma simultanea.

5. Para finalizar la ejecución de la aplicación, utilizar el script **stop.sh**

```
./bin/stop.sh $NOMBRE_APP
```

Algunas aclaraciones:

- Algunos componentes de hadoop no se pueden levantar con menos de 1Gb de memoria configurada. Si suman la memoria máxima configurada de todos los procesos, da mucho más que 4Gb, por lo que si realmente intentan estresar la virtual con 4Gb de memoria alocada, se van a empezar a caer los procesos.
- Para tener los comandos más a mano pueden ejecutar lo siguiente

```
echo export  
PATH=$PATH:/home/hadoop/Desktop/curso/bin:$HADOOP_HOME/bin:$HIVE_HOME/bi
```

Programación distribuida

Diplomatura en Big Data

```
n >> /home/hadoop/.bashrc
```