

## Taller de Preparación para Obtener la Certificación de Apache Spark CCA175 – by Datahack

1. Descargar e Instalar Vagrant: <https://www.vagrantup.com/downloads>

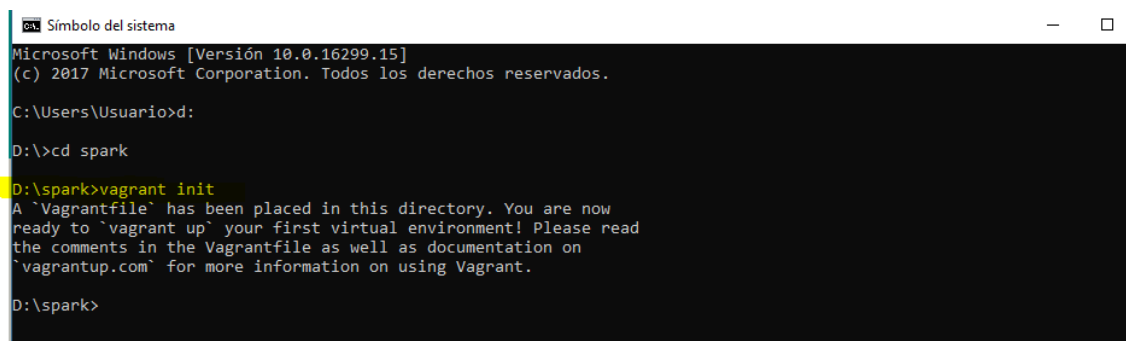
Información de que es Vagrant y para que lo usaremos:

[https://www.youtube.com/watch?v=Ud7cmVCNACE&ab\\_channel=makigas%3Atutorialesdeprogramaci%C3%B3n](https://www.youtube.com/watch?v=Ud7cmVCNACE&ab_channel=makigas%3Atutorialesdeprogramaci%C3%B3n)

2. Descargar e Instalar: <https://www.virtualbox.org/wiki/Downloads>
3. Crear una Carpeta en el directorio D:/spark
4. Usar la consola “CMD” para ubicarte dentro del directorio:



5. Ejecutar el comando “vagrant init”



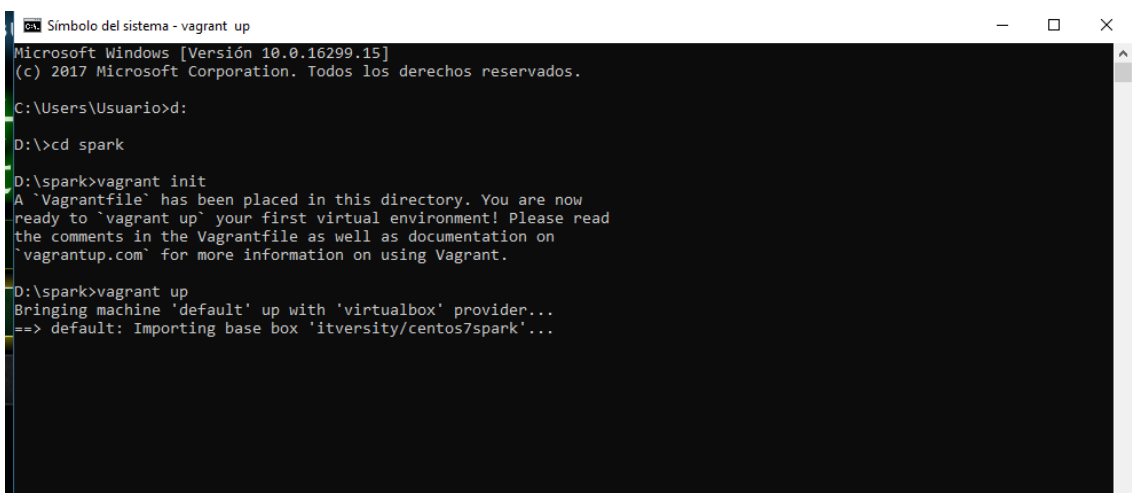
- Dentro de la carpeta D:/spark se creará un file que diga Vagrantfile y lo editaremos con el siguiente texto en su contenido:

```
# -*- mode: ruby -*-  
# vi: set ft=ruby :  
  
Vagrant.configure("2") do |config|  
  config.vm.box = "itversity/centos7spark"  
  config.vm.network "forwarded_port", guest: 8888, host: 8888  
  config.vm.network "forwarded_port", guest: 4040, host: 4040  
  
  config.vm.provider "virtualbox" do |vb|  
    vb.cpus = "2"  
    vb.memory = "4096"  
  end  
end
```



# DataHack

- Dentro de la consola CMD ejecutar: `vagrant up`  
(se descargará todo el contenido de la VM). Esto tardará unos minutos o un par de horas dependiendo de tu velocidad de internet

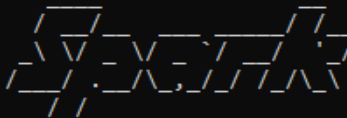


```
Símbolo del sistema - vagrant up  
Microsoft Windows [Versión 10.0.16299.15]  
(c) 2017 Microsoft Corporation. Todos los derechos reservados.  
C:\Users\Usuario>d:  
D:\>cd spark  
D:\spark>vagrant init  
A `Vagrantfile` has been placed in this directory. You are now  
ready to `vagrant up` your first virtual environment! Please read  
the comments in the Vagrantfile as well as documentation on  
`vagrantup.com` for more information on using Vagrant.  
D:\spark>vagrant up  
Bringing machine 'default' up with 'virtualbox' provider...  
==> default: Importing base box 'itversity/centos7spark'...
```

Y se creará una Máquina virtual en VirtualBox

Con esto podemos ingresar a la VM y utilizar el entorno.

```
D:\Jesus\spark>vagrant ssh
Last login: Sat Nov 7 04:55:22 2020 from 10.0.2.2
[vagrant@localhost ~]$ spark-shell
2020-11-07 04:56:17,666 WARN util.Utils: Your hostname, localhost.localdomain res
s to a loopback address: 127.0.0.1; using 10.0.2.15 instead (on interface eth0)
2020-11-07 04:56:17,666 WARN util.Utils: Set SPARK_LOCAL_IP if you need to bind t
other address
2020-11-07 04:56:18,030 WARN util.NativeCodeLoader: Unable to load native-hadoop
ary for your platform... using builtin-java classes where applicable
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel
Level).
Spark context Web UI available at http://10.0.2.15:4040
CSpark context available as 'sc' (master = local[*], app id = local-1604724986302)
Spark session available as 'spark'.
Welcome to

 version 2.4.5

Using Scala version 2.11.12 (OpenJDK 64-Bit Server VM, Java 1.8.0_242)
Type in expressions to have them evaluated.
Type :help for more information.
```

Ya estamos listos para empezar con la práctica.