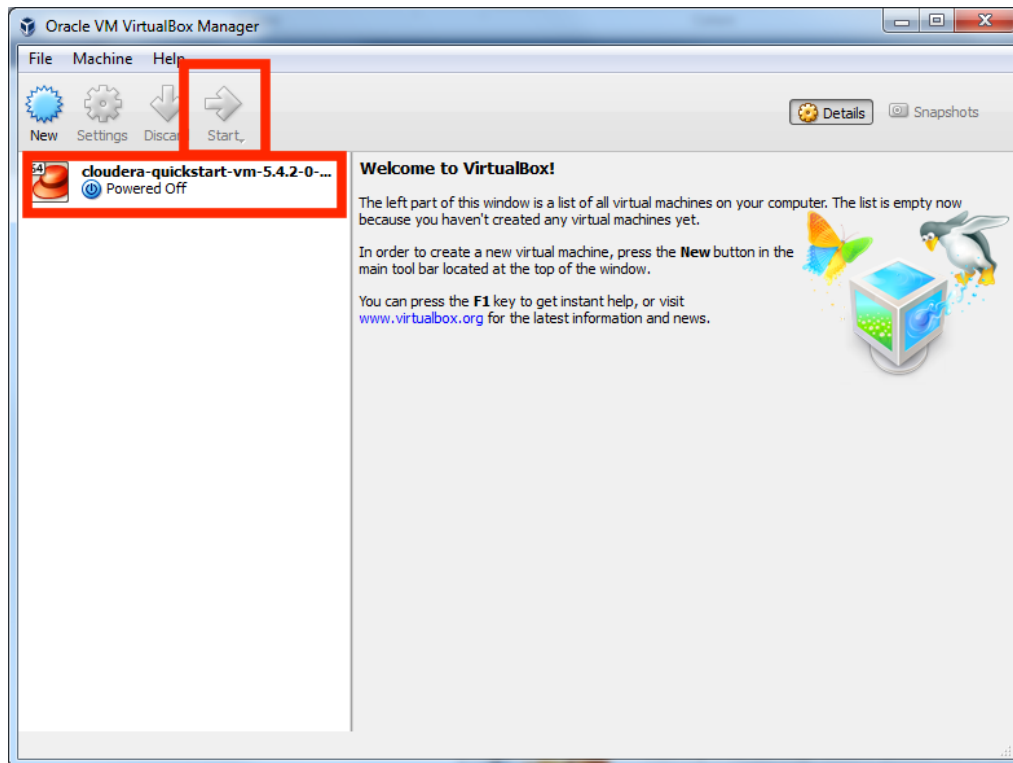


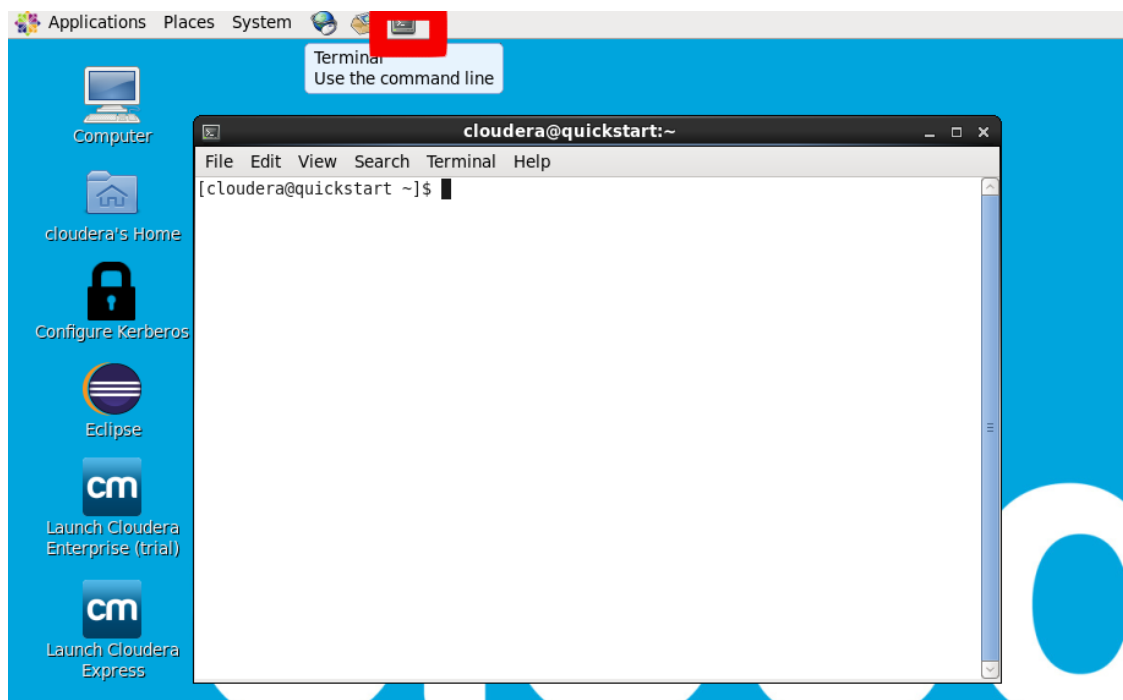
Instrucciones para la instalación de Pyspark

Esta página le guiará a través de la instalación y configuración de Pyspark en la MV_Cloudera.

1. Abra VirtualBox y arranque la MV de Cloudera, tal y como hizo durante el proceso de instalación de la misma.

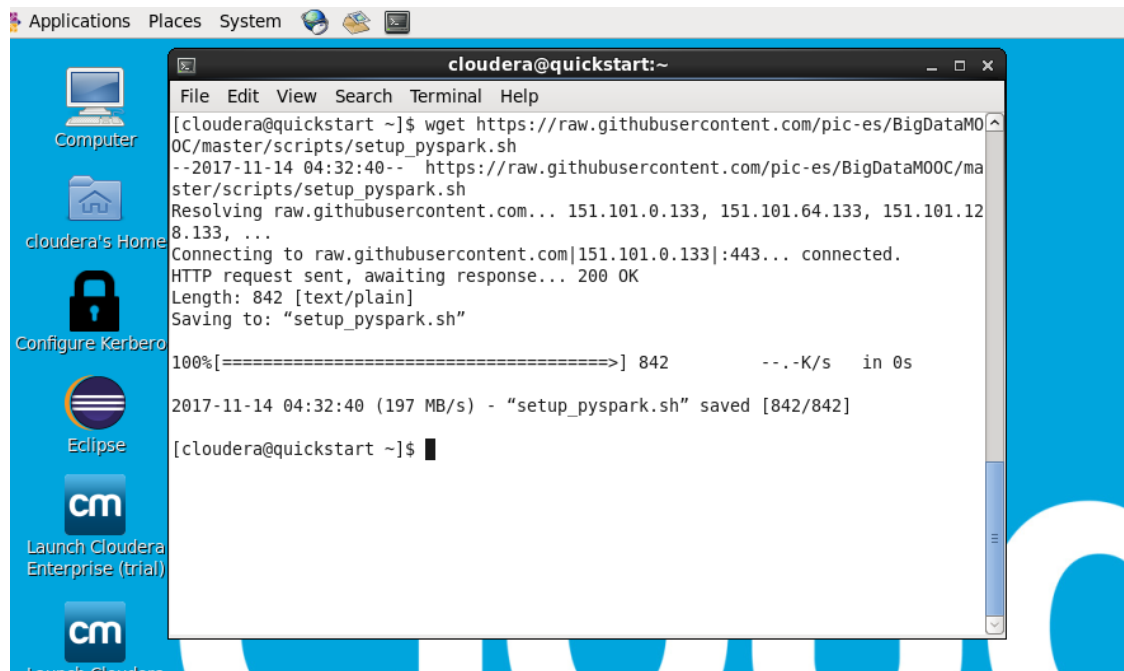


2. Abra un terminal.



3. Descargue el *script* de instalación, ejecutando el siguiente comando en el terminal:

wget https://raw.githubusercontent.com/pic-es/BigDataMOOC/master/scripts/setup_pyspark.sh



The screenshot shows a terminal window titled 'cloudera@quickstart:~' with a menu bar (File, Edit, View, Search, Terminal, Help). The terminal output shows the execution of the wget command to download 'setup_pyspark.sh' from a GitHub repository. The download is successful, showing a 200 OK status and a file size of 842 bytes. The file is saved to the current directory. The desktop background is blue with various icons on the left, including 'Computer', 'cloudera's Home', 'Configure Kerberos', 'Eclipse', and 'Launch Cloudera Enterprise (trial)'.

```
cloudera@quickstart:~$ wget https://raw.githubusercontent.com/pic-es/BigDataMOOC/master/scripts/setup_pyspark.sh
--2017-11-14 04:32:40-- https://raw.githubusercontent.com/pic-es/BigDataMOOC/master/scripts/setup_pyspark.sh
Resolving raw.githubusercontent.com... 151.101.0.133, 151.101.64.133, 151.101.128.133, ...
Connecting to raw.githubusercontent.com|151.101.0.133|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 842 [text/plain]
Saving to: "setup_pyspark.sh"
100%[=====] 842 --.-K/s in 0s
2017-11-14 04:32:40 (197 MB/s) - "setup_pyspark.sh" saved [842/842]

[cloudera@quickstart ~]$
```

4. Ejecute el script de instalación y siga la instrucciones que aparecen en la pantalla:

- Ejecute este comando en el terminal, para iniciar el proceso de instalación:

```
[cloudera@quickstart ~]$ . setup_pyspark.sh
```

- El script descargará el paquete Anaconda e iniciará su instalación:

```
% Total      % Received % Xferd  Average Speed   Time    Time
Time Current Dload Upload  Total   Spent    Left   Speed 100
455M 100 455M    0     0 11.1M    0  0:00:41  0:00:41 --:--:--
11.2M
```

Welcome to Anaconda3 4.2.0 (by Continuum Analytics, Inc.)

In order to continue the installation process, please review the license agreement.

Please, press ENTER to continue

>>>

- Pulse ENTER para ver el Acuerdo de Licencia

```
=====
Anaconda License
=====
```

Copyright 2016, Continuum Analytics, Inc.

All rights reserved under the 3-clause BSD License:

Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

* Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

* Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

* Neither the name of Continuum Analytics, Inc. nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

--More--

- Pulse q para salir del Acuerdo de Licencia y proceder a aceptarlo.

Do you approve the license terms? [yes|no]

>>>

- Escriba **yes** y pulse ENTER. Se le preguntará que confirme la ruta de instalación.

Anaconda3 will now be installed into this location:
/home/cloudera/anaconda3

- Press ENTER to confirm the location
- Press CTRL-C to abort the installation
- Or specify a different location below

[/home/cloudera/anaconda3] >>>

- Pulse ENTER para aceptar la ubicación propuesta y la instalación se iniciará. Cuando termine se le preguntará si quiere:

```
installing: yaml-0.1.6-0 ...
installing: zeromq-4.1.4-0 ...
installing: zlib-1.2.8-3 ...
installing: anaconda-4.2.0-np111py35_0 ...
installing: ruamel_yaml-0.11.14-py35_0 ...
installing: conda-4.2.9-py35_0 ...
installing: conda-build-2.0.2-py35_0 ...
Python 3.5.2 :: Continuum Analytics, Inc.
creating default environment...
installation finished.
Do you wish the installer to prepend the Anaconda3 install location
to PATH in your /home/cloudera/.bashrc ? [yes|no]
[no] >>>
```

- Escriba **yes** y pulse ENTER para añadir el directorio de instalación de Anaconda al PATH. Haciendo esto, la versión de Python contenida en Anaconda, será la utilizada por defecto.

```
Prepending PATH=/home/cloudera/anaconda3/bin to PATH in
home/cloudera/.bashrc
A backup will be made to: /home/cloudera/.bashrc-anaconda3.bak
```

For this change to become active, you have to open a new terminal.

Thank you for installing Anaconda3!

Share your notebooks and packages on Anaconda Cloud!
Sign up for free: <https://anaconda.org>

- La instalación ha completado. Escriba **pyspark** y pulse ENTER para comprobar la instalación de Pyspark. Haciendo esto, debería abrirse una ventana de navegador con un Jupyter Notebook.

