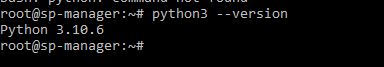
1. Get inside manager container
2. Check if python is installed. Pyspark requires at least python 3.7 or above.

$ python –version

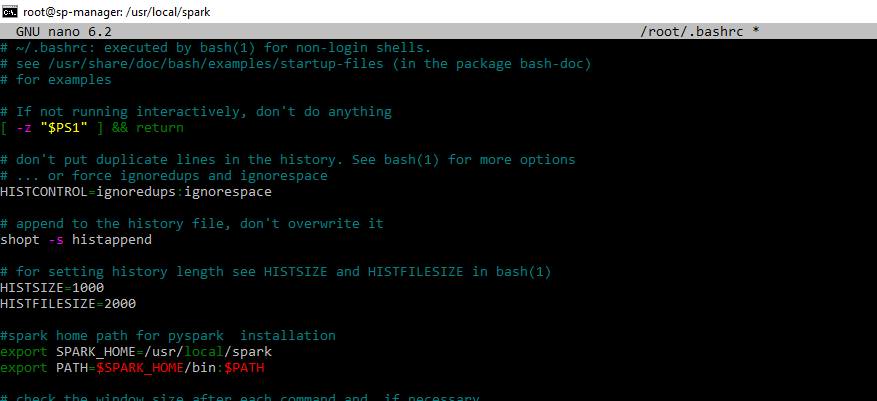


1. As we have python 3.10 , we are good to go with pyspark
2. Add spark home path in ~/.bashrc

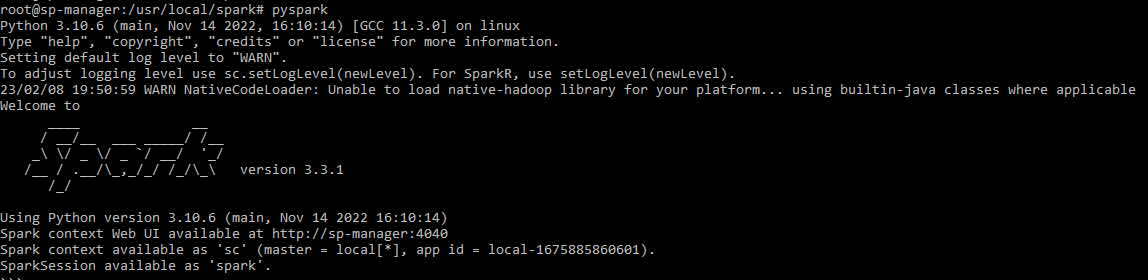
$ nano ~/.bashrc

Add below path in the file and save.

export SPARK\_HOME=/usr/local/spark  
export PATH=$SPARK\_HOME/bin:$PATH



1. Restart the terminal
2. Verify pyspark



1. Install pip and jupyter

$apt-get install pip -y && pip install jupyter -y

1. Launch Jupyter with below command so that I can be accessed from host machine on port 8888

$ jupyter notebook --ip=0.0.0.0 --port=8888 --no-browser --allow-root

Use URL & token shown in the output to open jupyter notebook.

Text

Description automatically generated

1. Use port forwarding command in local machine to open jupyter notebook

$ssh -L 6060:localhost:6060 [gprasad@10.247.62.19](mailto:gprasad@10.247.62.19)

Text

Description automatically generated

1. Now open browser in local machine and copy URL & Token generated in step 8.

Graphical user interface, text, email

Description automatically generated

**PySpark in Jupyter**

1. Update PySpark driver environment variables: add these lines to your ~/.bashrcfile.

export PYSPARK\_DRIVER\_PYTHON=jupyter  
export PYSPARK\_DRIVER\_PYTHON\_OPTS= ‘notebook --ip=0.0.0.0 --port=8888 --no-browser --allow-root’

Text

Description automatically generated

1. Restart terminal and launch PySpark again:

$ pyspark

A screenshot of a computer

Description automatically generated with medium confidence

Graphical user interface, text, application, email

Description automatically generated

1. Execute a test program in Jupyter

Graphical user interface, text, application, email

Description automatically generated

**References**

<https://medium.com/sicara/get-started-pyspark-jupyter-guide-tutorial-ae2fe84f594f>