

How to notebook and pyspark

1. SSH to notebook host
 - a. `ssh ubuntu@130.238.28.167 -i /Users/Edvin/Downloads/edvinkey.pem \`
 - b. `-L 4040:localhost:4040 \`
 - c. `-L 4041:localhost:4041 \`
 - d. `-L 4042:localhost:4042 \`
 - e. `-L 4043:localhost:4043 \`
 - f. `-L 4044:localhost:4044 \`
 - g. `-L 9870:192.168.2.129:9870`
2. Packages
 - a. `sudo apt update -u`
 - b. `sudo apt install python3-pip`
 - i. `pip3 install pyspark==3.3.2`
 - ii. `pip3 install notebook`
 - c. `sudo apt install net-tools`
 - d. `sudo apt install openjdk-11-jdk-headless`
 - e. `sudo apt install jupyter-notebook`
 - f. `sudo apt install jupyter-core`
 - g. `sudo snap install jupyter`
3. DNS settings
 - a.

```
for i in {1..4};  
do  
    for j in {1..255};  
    do  
        echo "192.168.$i.$j host-192-168-$i-$j" | sudo tee -a /etc/hosts  
    done  
done
```
4. Set hostname of your vm
 - a. `sudo hostname host-$(hostname -I | awk '{ $1=$1 }; 1' | sed 's/\./-/g')`
 - b. `echo "sudo hostname host-$(hostname -I | awk '{ $1=$1 }; 1' | sed 's/\./-/g')"`
`sudo tee -a /home/ubuntu/.profile`
5. Set environment
 - a. `export JAVA_HOME=/usr/lib/jvm/java-11-openjdk-amd64`
`echo JAVA_HOME=/usr/lib/jvm/java-11-openjdk-amd64 | sudo tee -a /etc/environment`
6. Setup jupyter notebook

- a. Get config file
 - i. `jupyter notebook --generate-config`
 1. created at:
 - a. `/home/ubuntu/.jupyter/jupyter_notebook_config.py`
 - b. Set password
 - i. Generate password
 1. `python3 -c 'from notebook.auth import passwd; print(passwd())'`
 - ii. Copy hashed password (hej123)
 1. `argon2:$argon2id$v=19$m=10240,t=10,p=8$Akp6BSOEXTUMiQtEaJH1Rw$+96WY9VrbA/FYLiOyNdDmzG8xDSW9SbQxslM0R92G2A`
 - iii. Edit configurable
 1. `nano ~/.jupyter/jupyter_notebook_config.py`
 2. `c.NotebookApp.password =`
`'argon2:$argon2id$v=19$m=10240,t=10,p=8$bWQ82xXaTbsn/0ZQFrEBcw$6nPfFm82i`
`tOqnlwlla2DL/Dsps1Tw6Jl2ZxtRS98mQc'`
`c.NotebookApp.ip = '*'`
`c.NotebookApp.open_browser = False`
7. Run notebook
 - a. `jupyter notebook --ip=* --no-browser`
 8. Access notebook
 - a. `130.238.28.167:8888`
 9. Accessing HDFS from notebook
 - a. ???