Setup Your Big Data enviroment

# 1 Access your Virtual Wall Node

1. You will recieve a login, IP and password

2. To connect to our datacenter via:

* **ssh <username>@<ip>**
* **or use putty ssh client**

# 2 Create a docker container for running jupyter + spark

1. More info on the container: https://github.com/jupyter/docker-stacks/tree/master/all-spark-notebook

2. In the shell of your Virtual Wall node

* change the ownership of /mnt (50GB volume)
  + **sudo chown -R <username> .**
  + **mkdir notebookdata**
* Create a container:
  + **sudo docker run -d --name spark -v /mnt/notebookdata:/home/jovyan/work -e GRANT\_SUDO=yes -p 8888:8888 -p 4040:4040 jupyter/all-spark-notebook**
  + (this will take a few minutes)
* You can access the container by attaching a shell:
  + **sudo docker exec -it spark /bin/bash**
* Get your jupyter access token via:
  + **jupyter notebook list**
* Open a browser and go to:
  + http://<your\_ip>:8888/?token=<access token>
* The directory /mnt/notebookdata on the Virtual Wall node is shared with the /home/jovyan/work directory in your container. Save everything here since it's safe against container crashes, which WILL happen.
* ISSUE: fix permissions:
  + **sudo chown -R 1000 /mnt/notebookdata** (uid=1000 is the jovyan user)