

Proxy Leader's Spark Cluster

Purpose

Port forward from the leader's spark cluster job queue to your home computer

How-To

1. Login to Datahub and select a CSE255 instance
2. Open a Jupyter Terminal
3. SSH to dsmlp-login.
 - a. `ssh <username>@dsmlp-login`
 - b. `ssh myusername@dsmlp-login`
4. Clone the directory <https://github.com/ucsd-ets/dsc291-spark-cluster>
 - a. `git clone https://github.com/ucsd-ets/dsc291-spark-cluster`
5. Change into the directory
 - a. `cd dsc291-spark-cluster`
6. Run the proxy-spark.sh command supply the leader's username
 - a. `./proxy-spark.sh <username>`
 - b. `./proxy-spark.sh myleadersusername`
7. Open a **local terminal** on your PC
8. Copy the SSH command from Step 3 onto your clipboard

```
[root@dsmlp-login dsc291-spark-cluster]# ./proxy-spark.sh liqiao
=====
=> Successfully connected to the Spark cluster
=> Next create a SSH tunnel from your personal computer using the following command:
    ssh -N -L 127.0.0.1:8080:127.0.0.1:41849 -L 127.0.0.1:4040:127.0.0.1:38087 root@dsmlp-login.ucsd.edu
=> Link to Spark cluster manager UI: http://127.0.0.1:8080
=> Link to Spark job UI: http://127.0.0.1:4040
=====
```

- a.
9. Paste the SSH output into your **local terminal** on your PC. Input your password and keep **this terminal session open for the duration of the proxy session**.
 10. Open a browser and navigate to <http://127.0.0.1:8080>