1. Connect to your virtual wall node via Putty or ssh (as instructed in the Lab3 HowTo)

2. Check if your containers are running **sudo docker ps**

=> spark container should be up

=> if it is not, follow instructions in Lab 3 HowTo, Section 2, to get it running

3. Go to the directory /mnt/notebookdata/ and download the lab's dataset:

=> **sudo wget** [**https://s3-eu-west-1.amazonaws.com/dieterdwdatasets/access-oct-14-mar17.log.bz2**](https://s3-eu-west-1.amazonaws.com/dieterdwdatasets/access-oct-14-mar17.log.bz2)

4. Unzip the dataset and change the ownership to allow access in spark container:

**=> mkdir** **/mnt/notebookdata/ldflogs**

**=> sudo chown -R 1000 /mnt/notebookdata/ldflogs**

**=> mv access-oct-14-mar17.log.bz2 /mnt/notebookdata/ldflogs**

**=> cd /mnt/notebookdata/ldflogs**

**=> bzip2 -d -vv access-oct-14-mar17.log.bz2**

=> unzipping might take about 7-8 minutes (**Do this before the lab starts!!**)

5. In your browser go to: **http://<ip\_virtual\_wall>:8888** and log in with your access token

6. Upload the lab's notebook: LDFLogsWithSpark\_2020\_FILLIN.ipynb

7. Have a look at the research paper showing initial analysis for the LDF logs.

**Good luck! The lab is due on Wednesday 18/3 at 23:59:59**