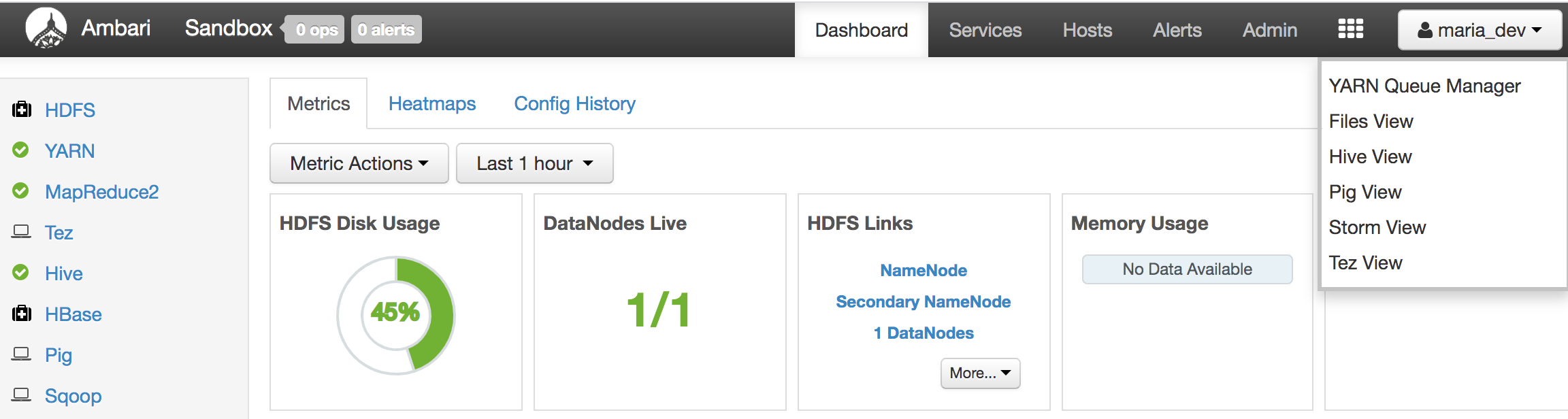
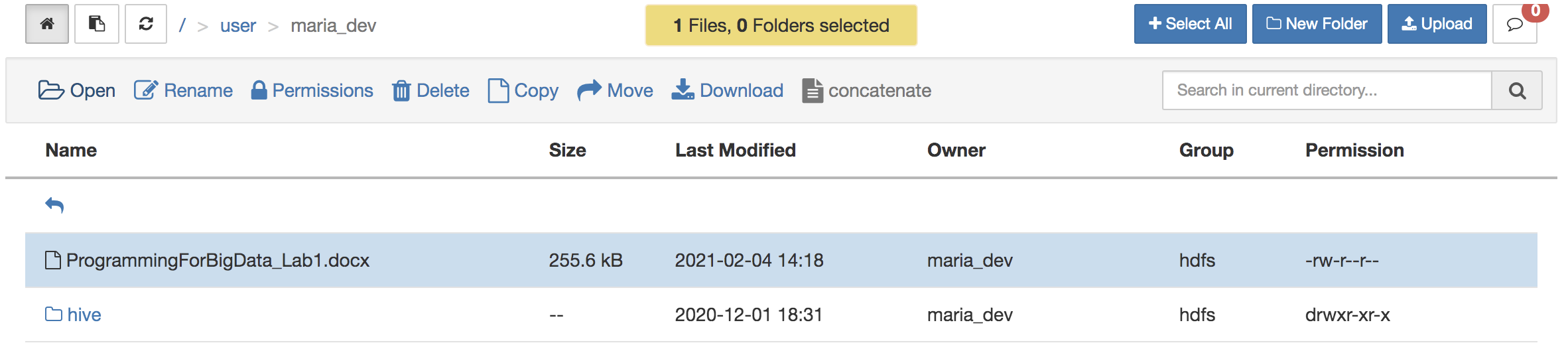


Lab 2

[Click here to Register Attendance](https://forms.gle/pesShFDe1ibVfS3L7)

|  |  |
| --- | --- |
| Name |  |
| Date |  |
| Student No |  |
| Student Email |  |

### **Getting to know HDFS - in the Ambari browser**

1. Go to the UCI Machine Learning Data Repository and explore the datasets available
2. Download a dataset of your choice and extract the files
3. Select Files view of Hadoop in the Ambari browser  
   
4. Go to the User folder and then the maria\_dev folder and upload your chosen dataset, using the Upload button
5. Check out the other file manipulation commands e.g. Rename, Copy, Delete, Download etc  
   

### **Getting to know HDFS - from the Command Line**

1. SSH into your Hadoop cluster (note telnet is on port 2222 by default:  
   **ssh maria\_dev@127.0.0.1 -p 2222**  
     
   Password is also: **maria\_dev**
2. Get a local file listing: **ls**
3. Get a file listing: **hadoop fs -ls**
4. Copy a file from a web source to local folder: **wget https://github.com/marloft/ITSligoIoT/blob/main/BlinkErrors**
5. Upload file from local folder to Hadoop cluster **hadoop fs -copyFromLocal BlinkErrors**
6. Get a file listing: **hadoop fs -ls**
7. Refresh the Ambari browser view of the same folder and see if the file is visible
8. Remove file from Hadoop cluster:  
   **hadoop fs -rm BlinkErrors**