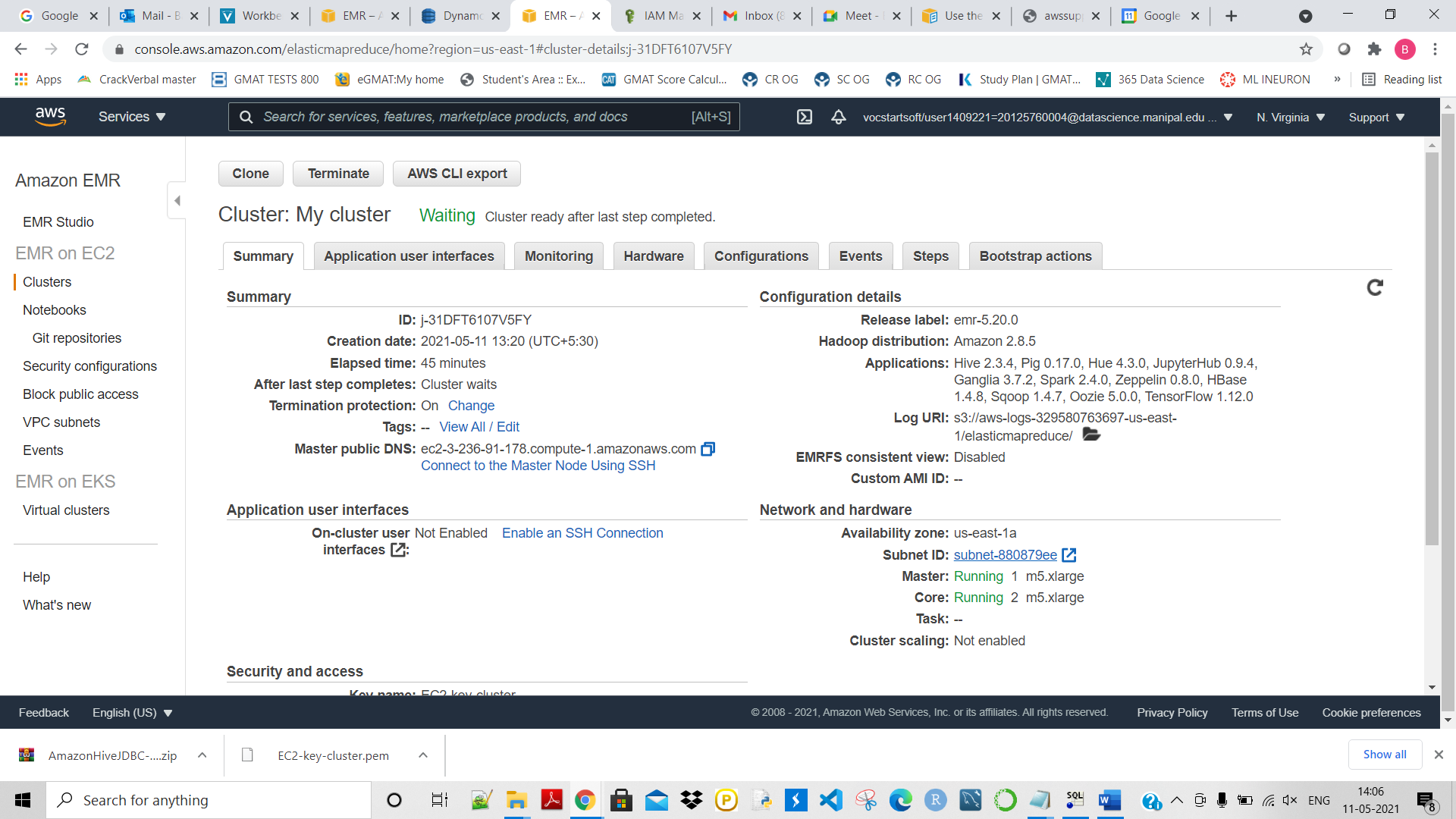
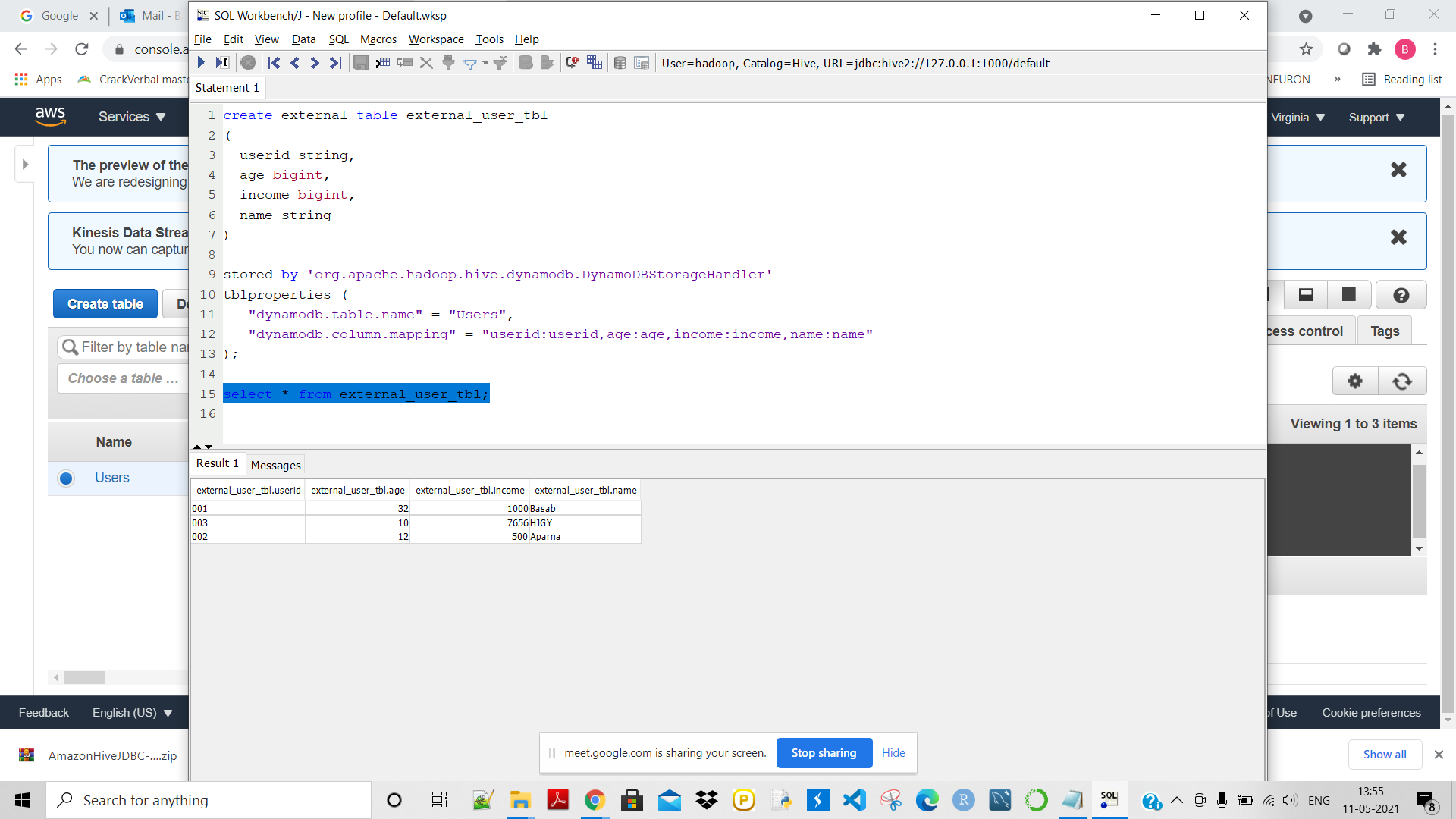
EMR Cluster created:

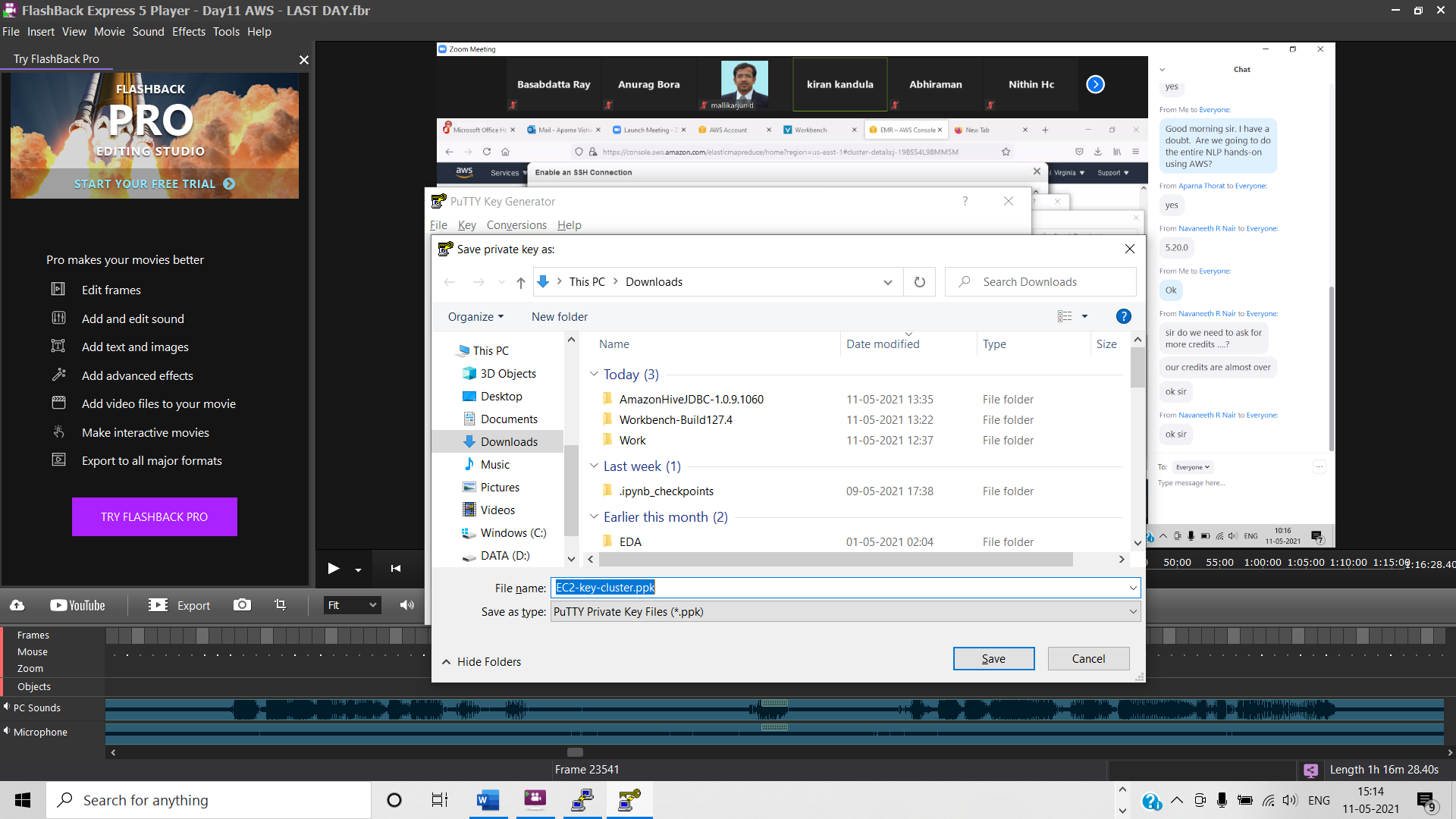


The above cluster is created then connected to the SQL workbench:

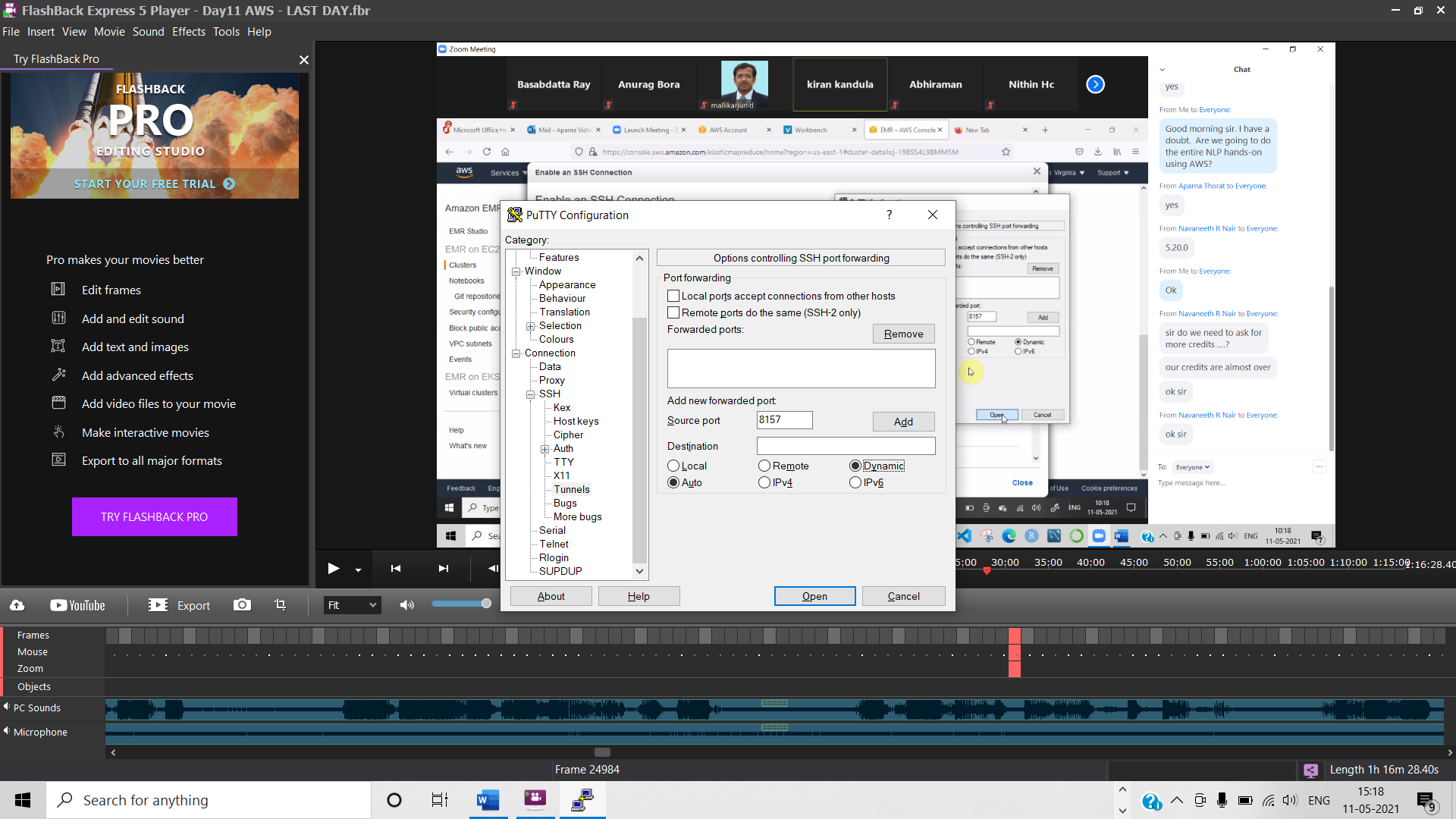


Next we will use tunnelling in order to access the Hive, Jupyter notebook etc:

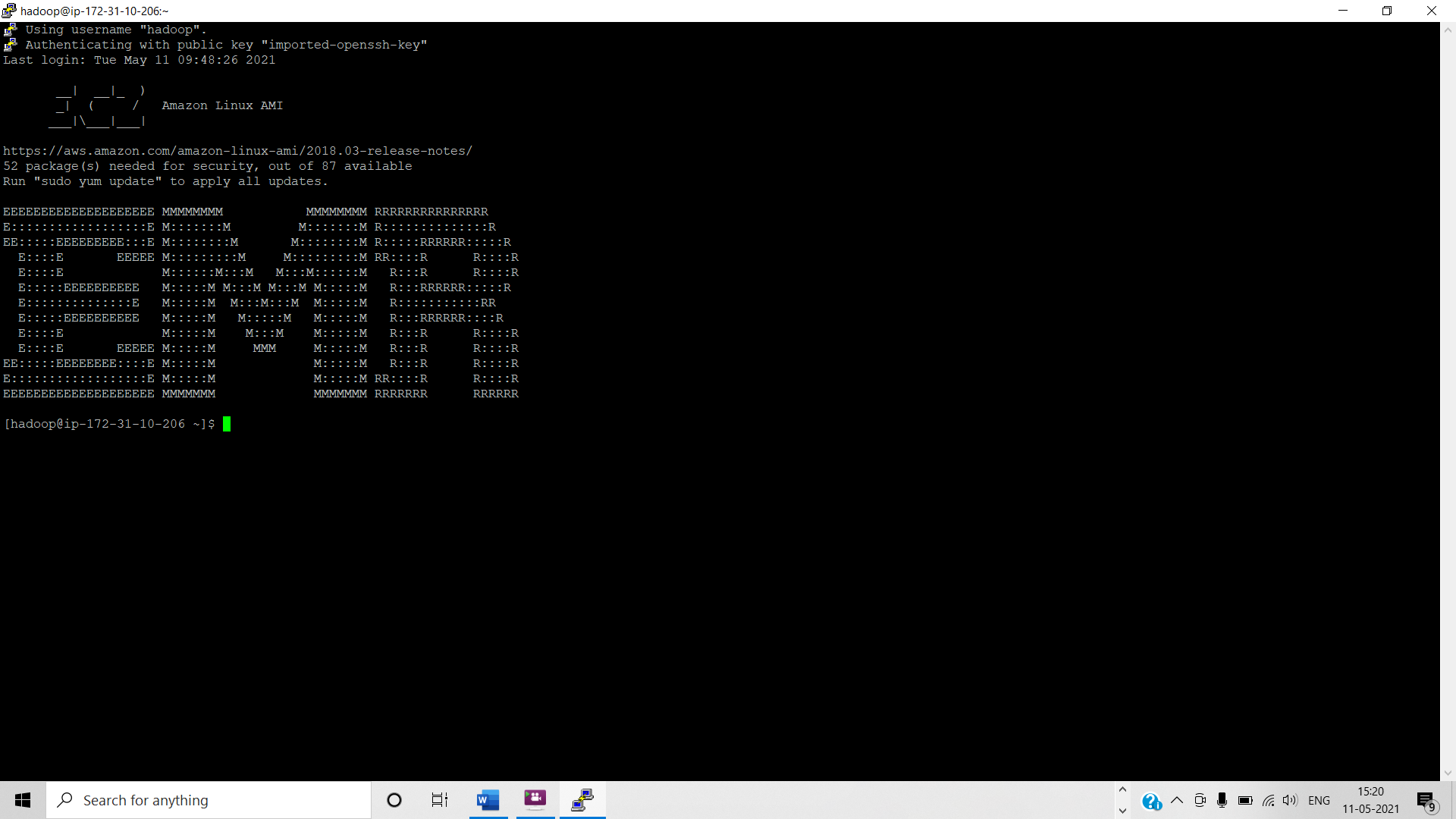
Download the putty, then save the the key as Private key using puttygen:



Then use ‘tunnel’:



Connected to the EMR succeccfully:

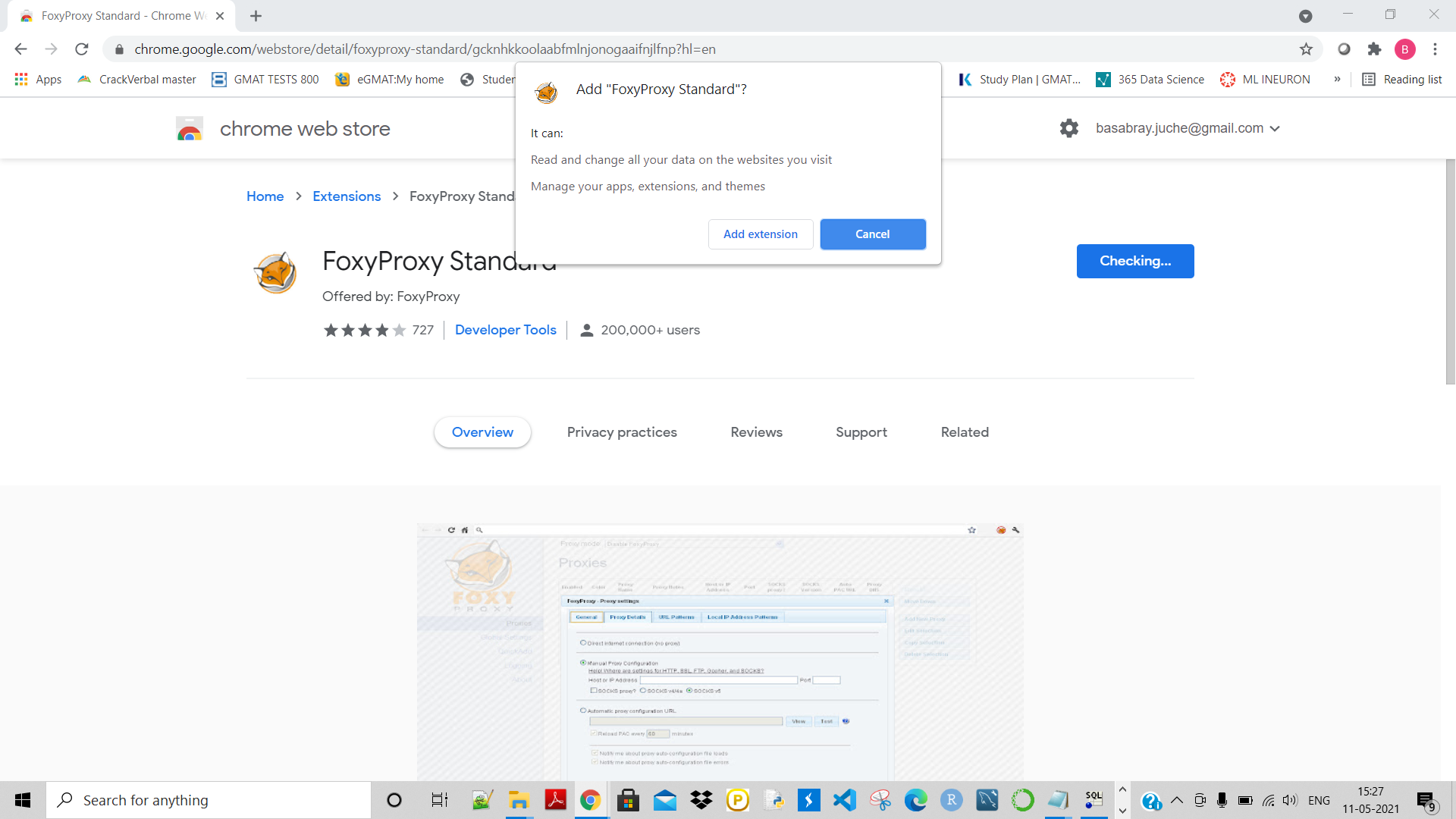


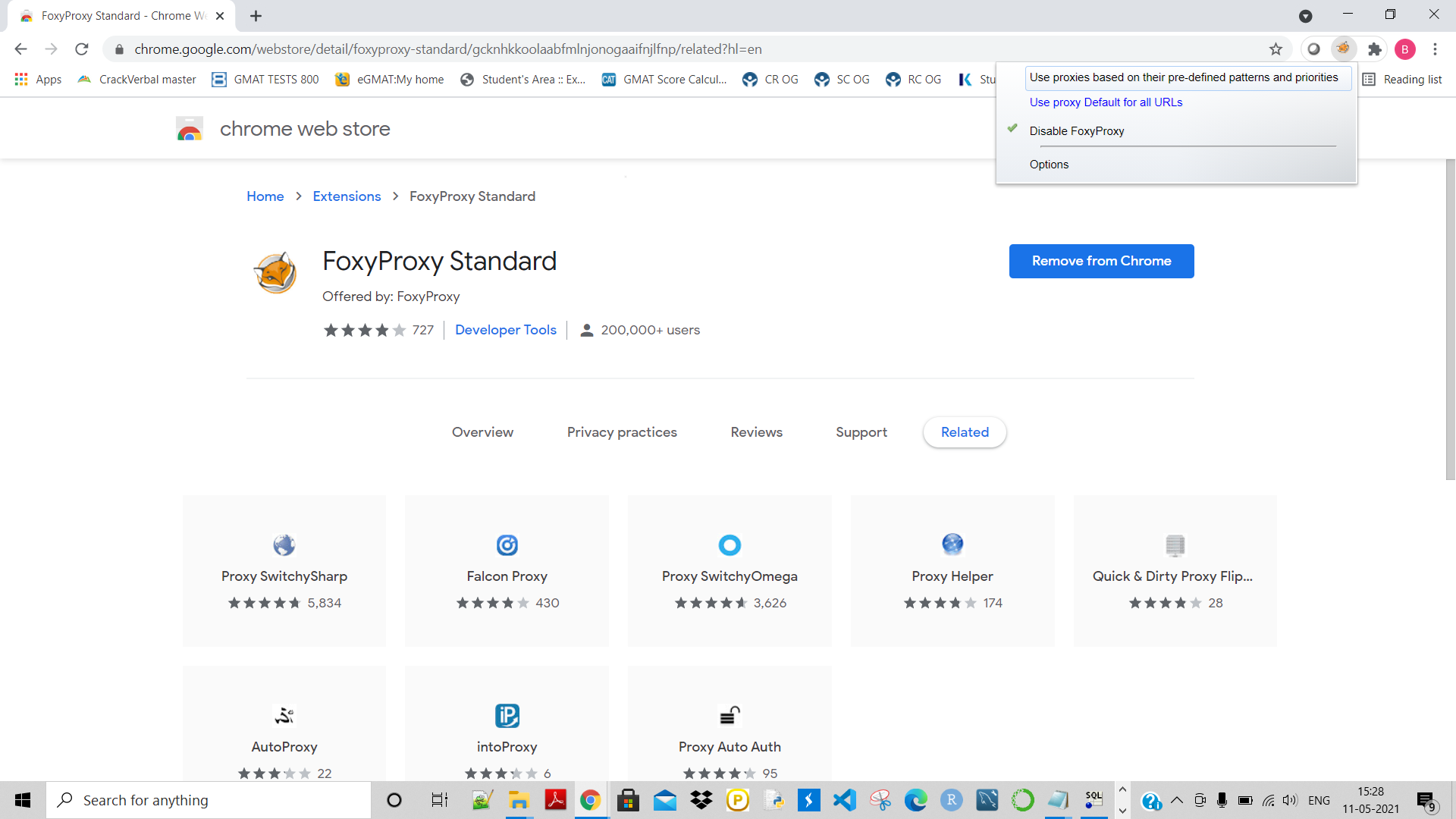
So the tunnelling to the EMR Cluster has been done.

Now using this tunnel to the EMR Cluster we will be able to connect to the services like Jupyter notebook, Hive, Spark etc.

Next we will add different proxy (foxyproxy) and access all these services from EMR:

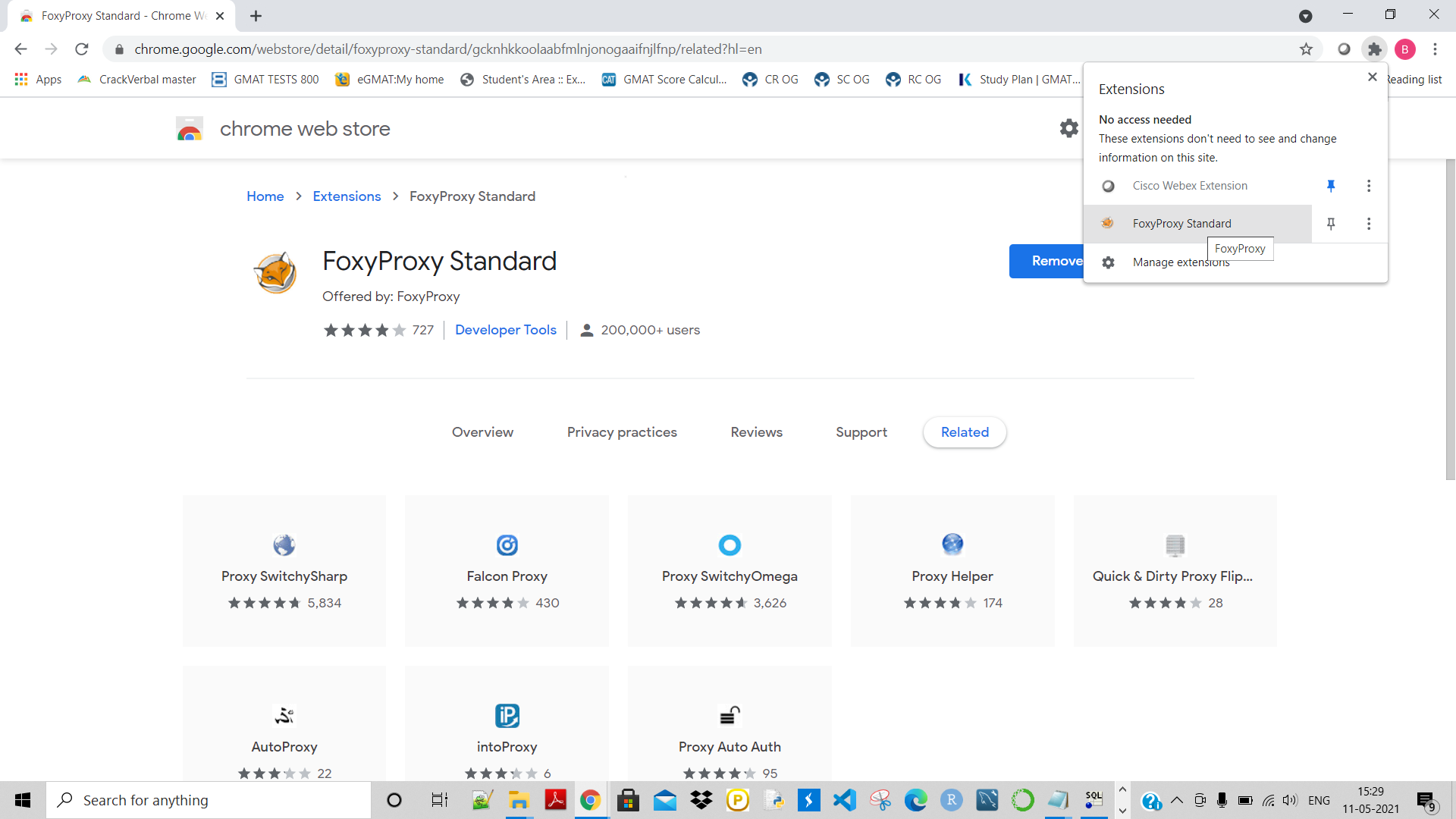
Look for foxyproxy chrome extension and add to connect:

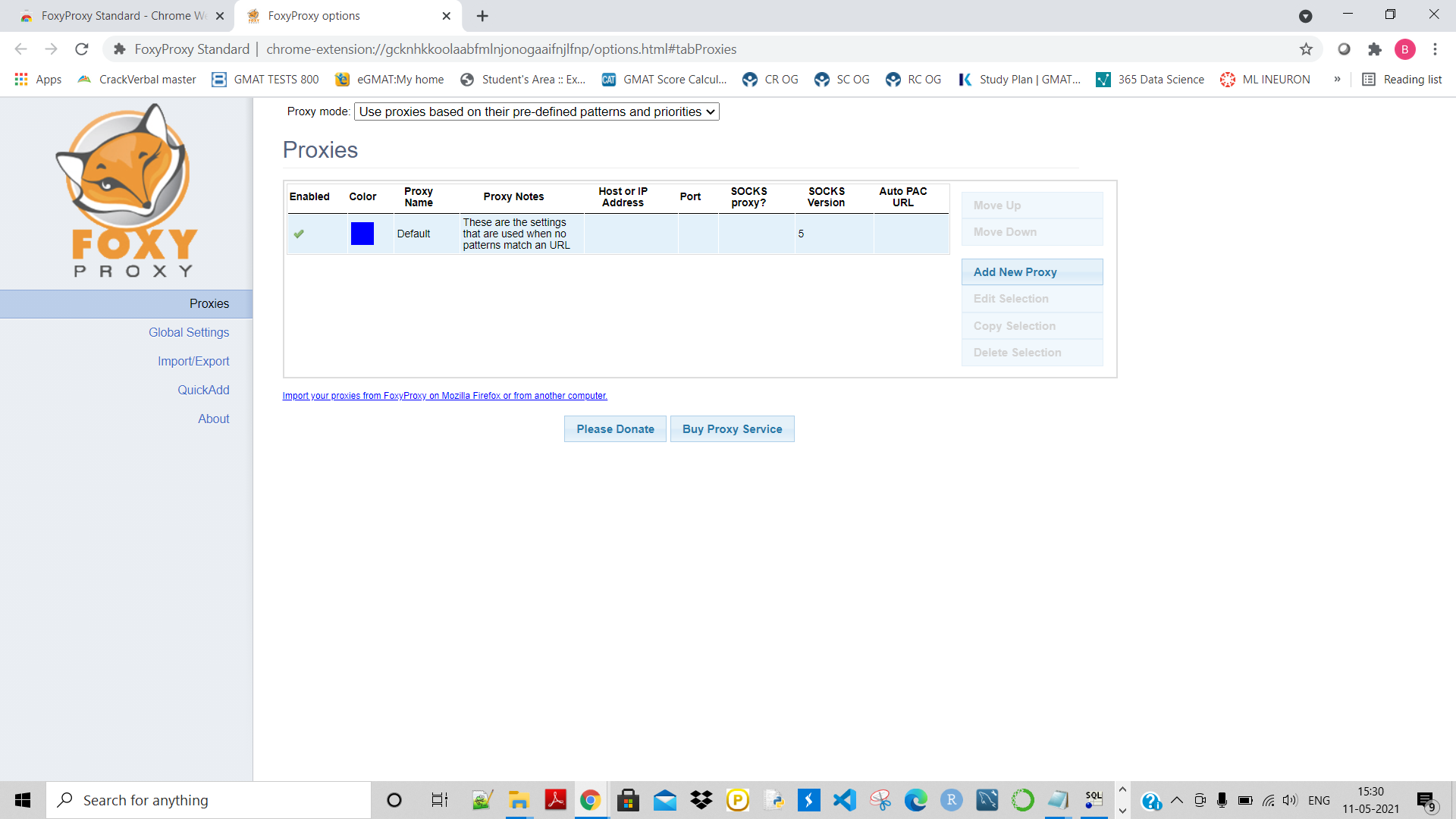




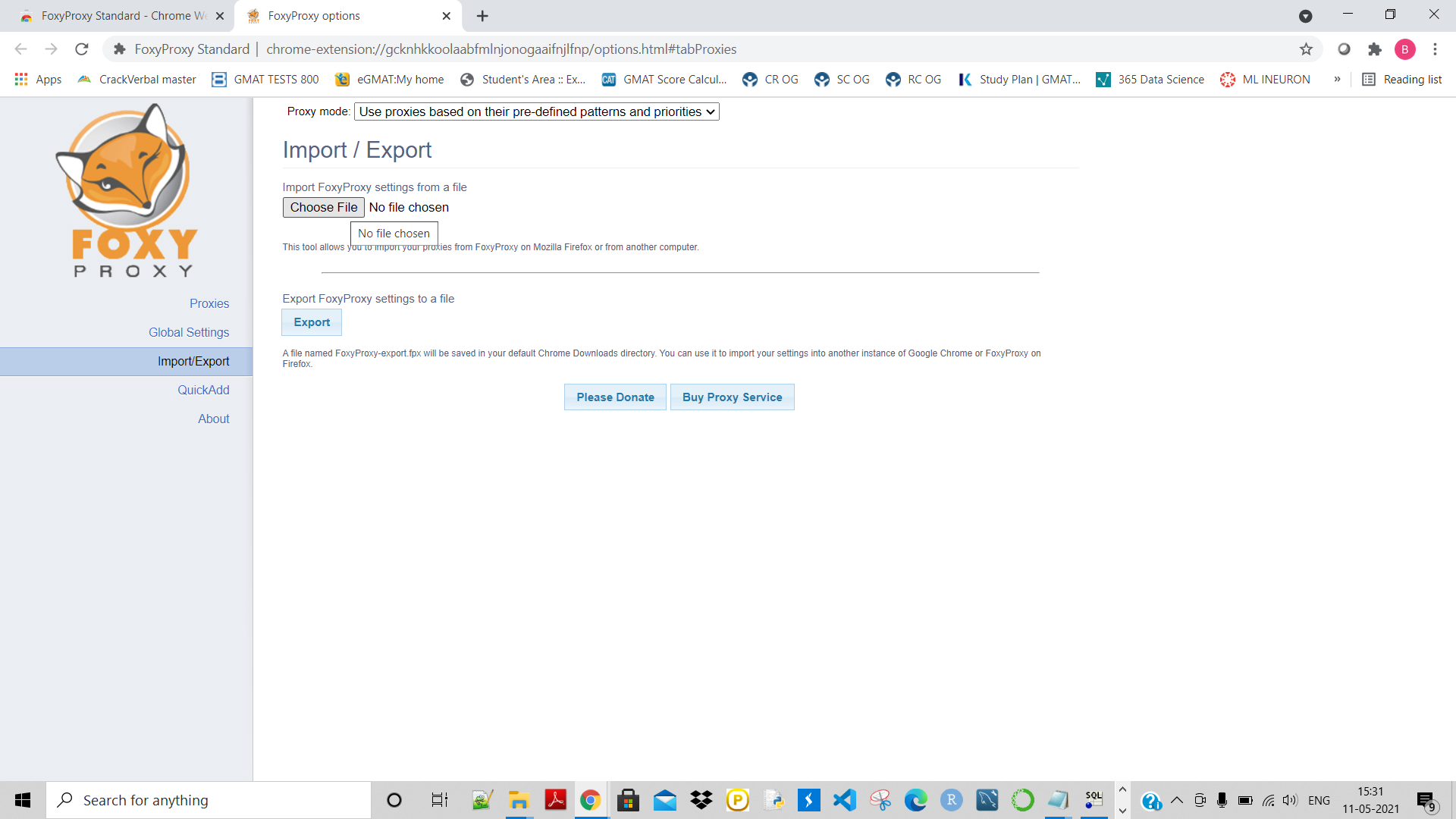
Click on ‘Used proxies based on their pre-define patterns and priorities:

Then clisk on foxyproxy standard and then clisk on Options:

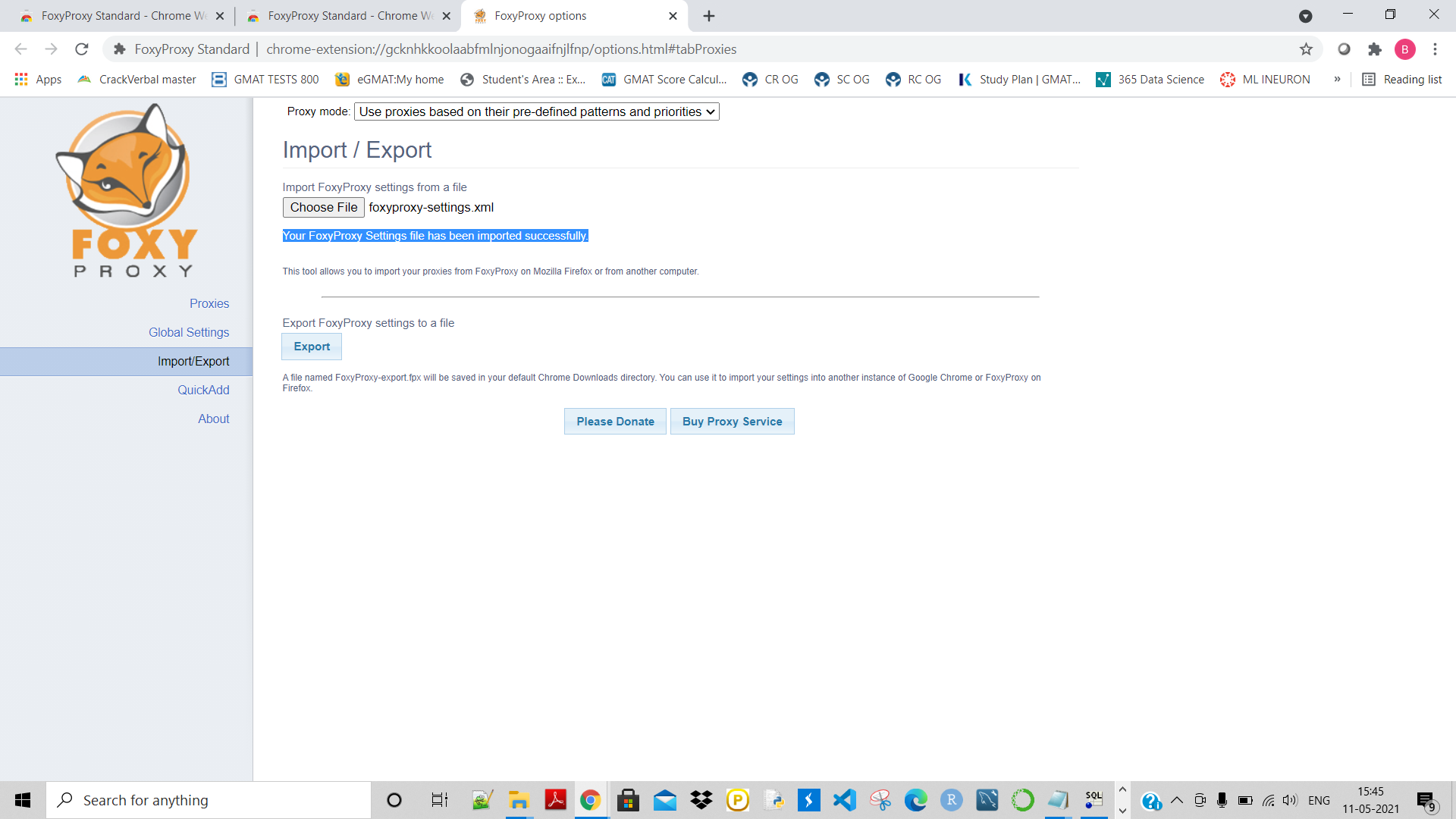




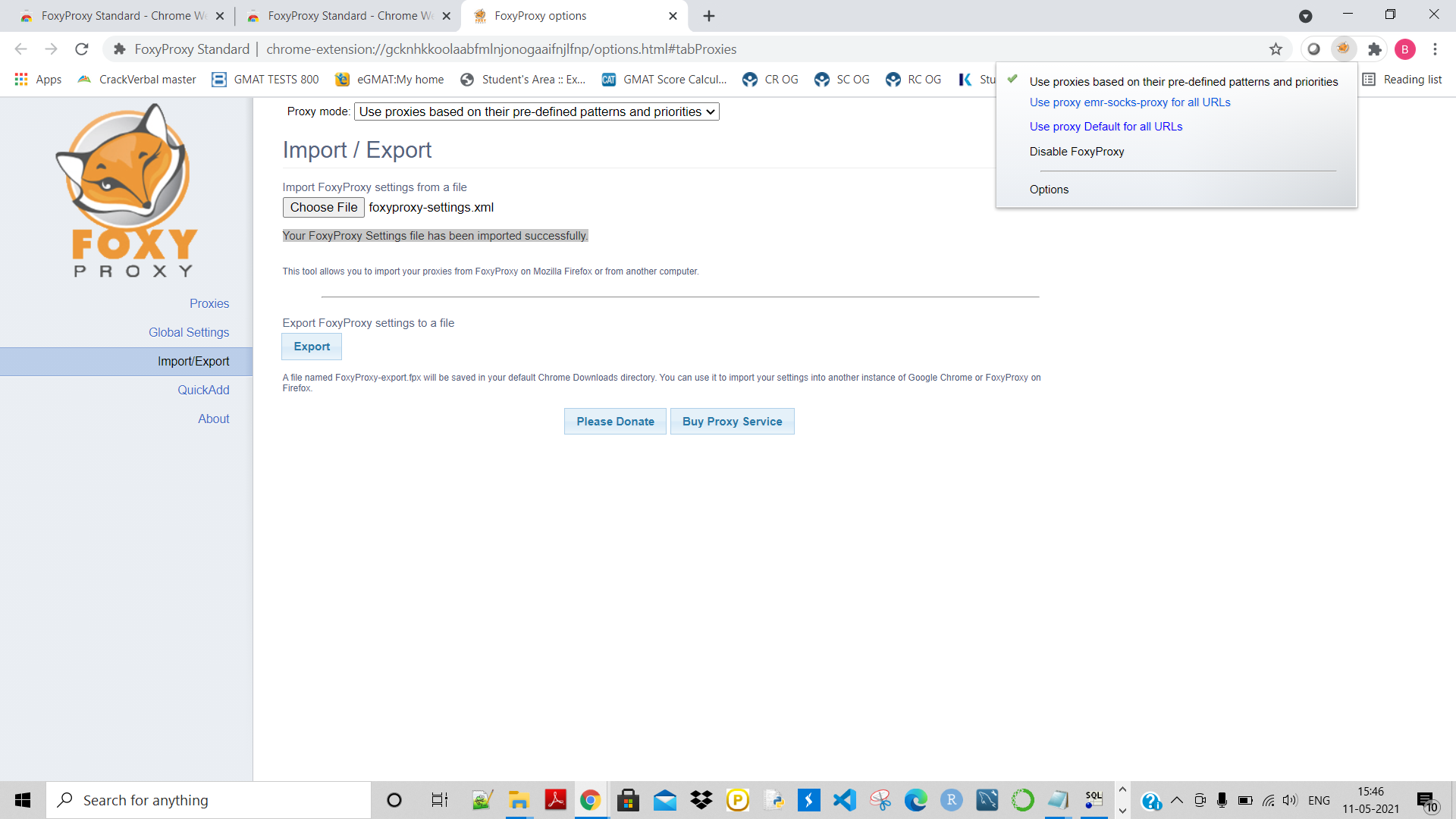
Now go to Import/Export and click on Choose file:



Here after selecting the file we can see that the foxyproxy settings are imported successfully:



Now it is enable:

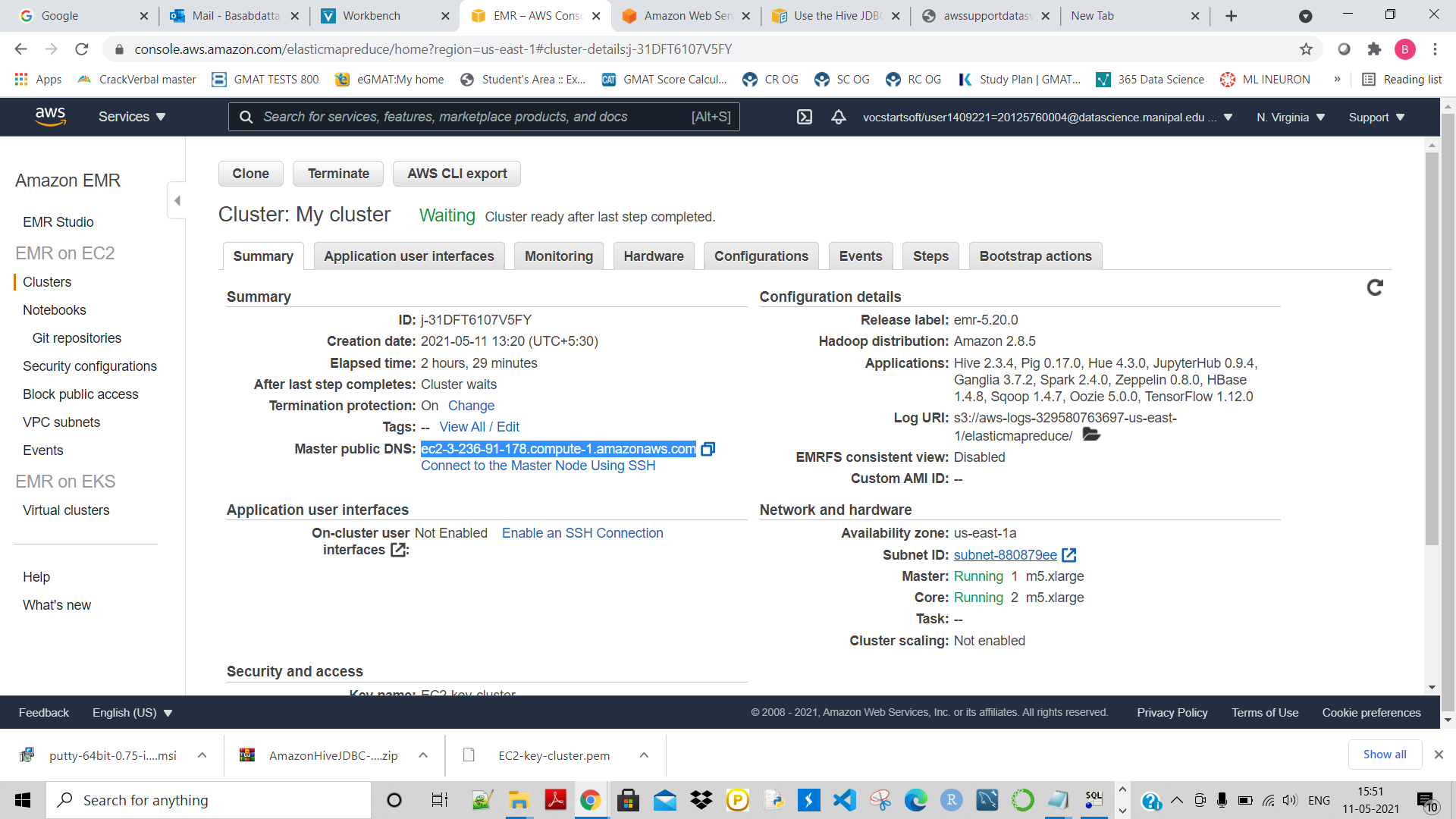


Now we are ready to accesx the services using the browser:

Access ganglia:

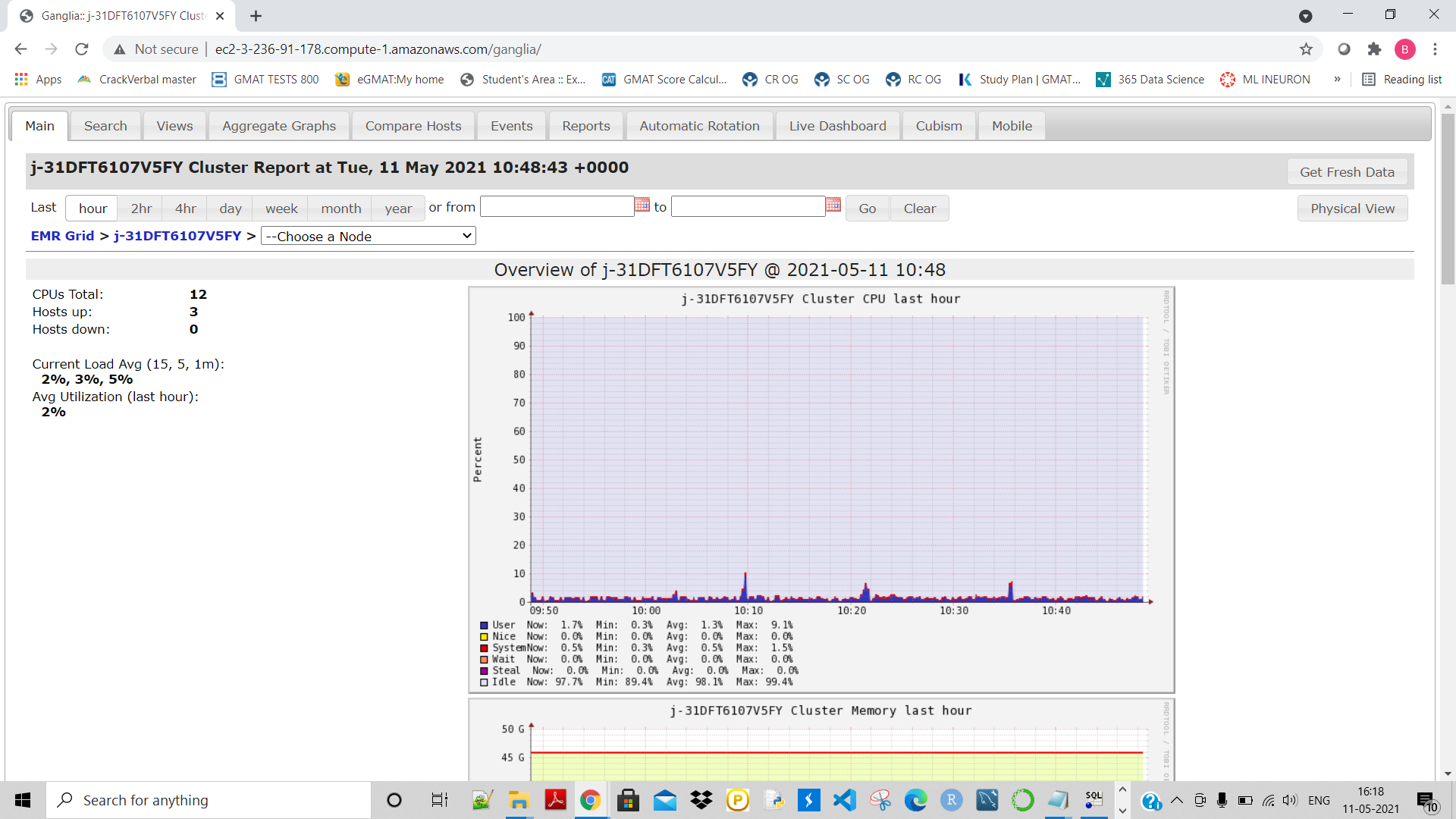
<http://ec2-3-87-209-11.compute-1.amazonaws.com/ganglia/>

here given my Master dns name from here:

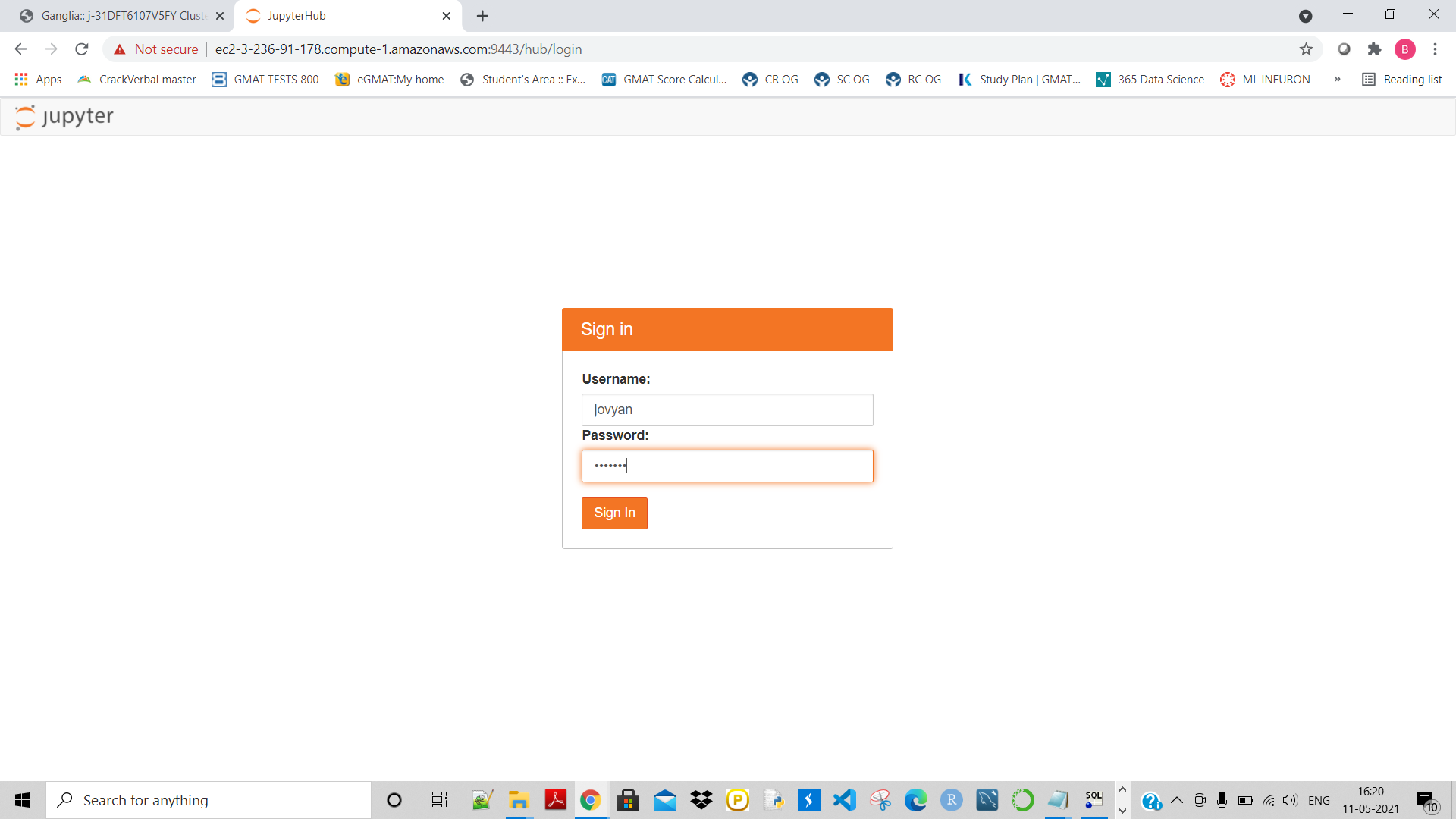


ec2-3-236-91-178.compute-1.amazonaws.com

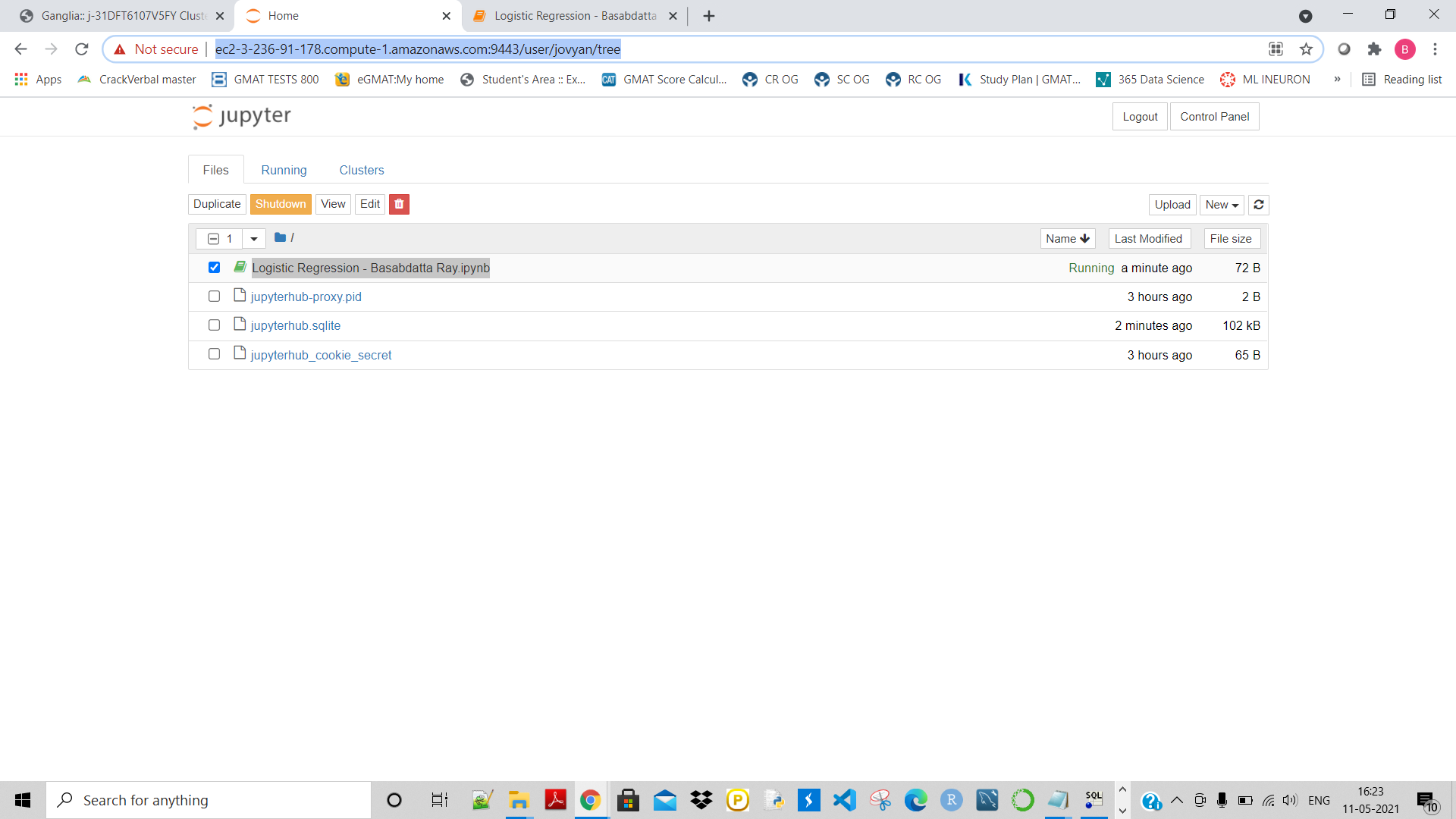
Now access ‘ganglia’:



Now access jupyter notebook:

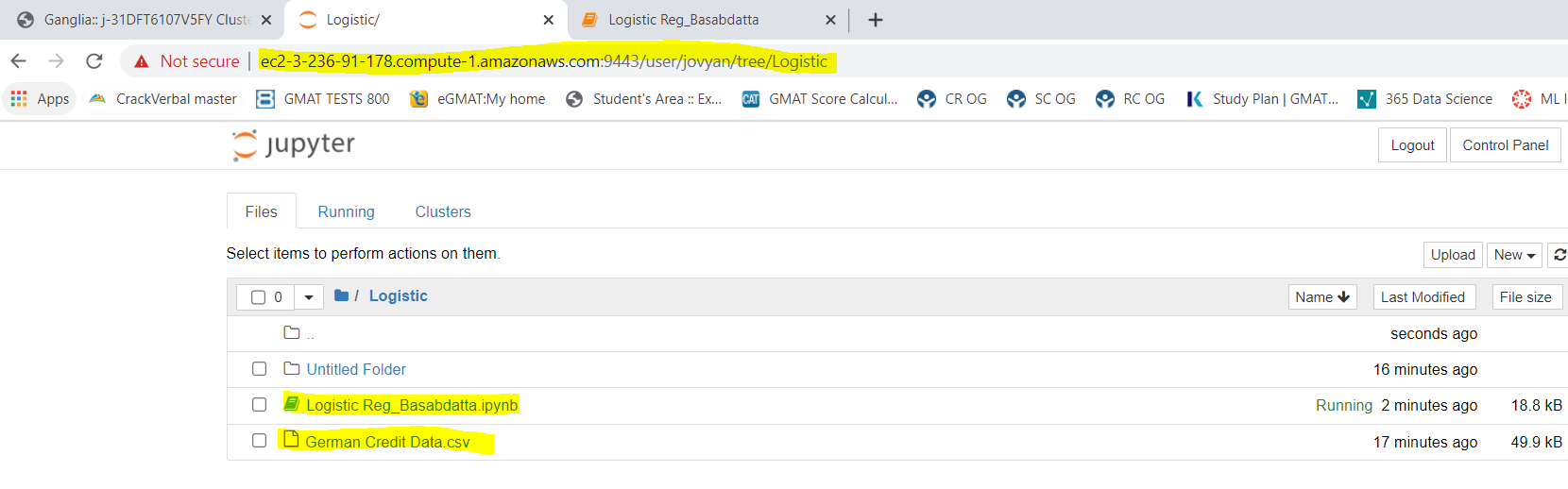


Connected succefully to jupyter notebook:



Created a folder named : ‘Logistic’

And then created a simple Logistic regression model using German credit data:



The Model creation:

