Assignment: Creating a cluster in Google cloud

Kamal Nandan

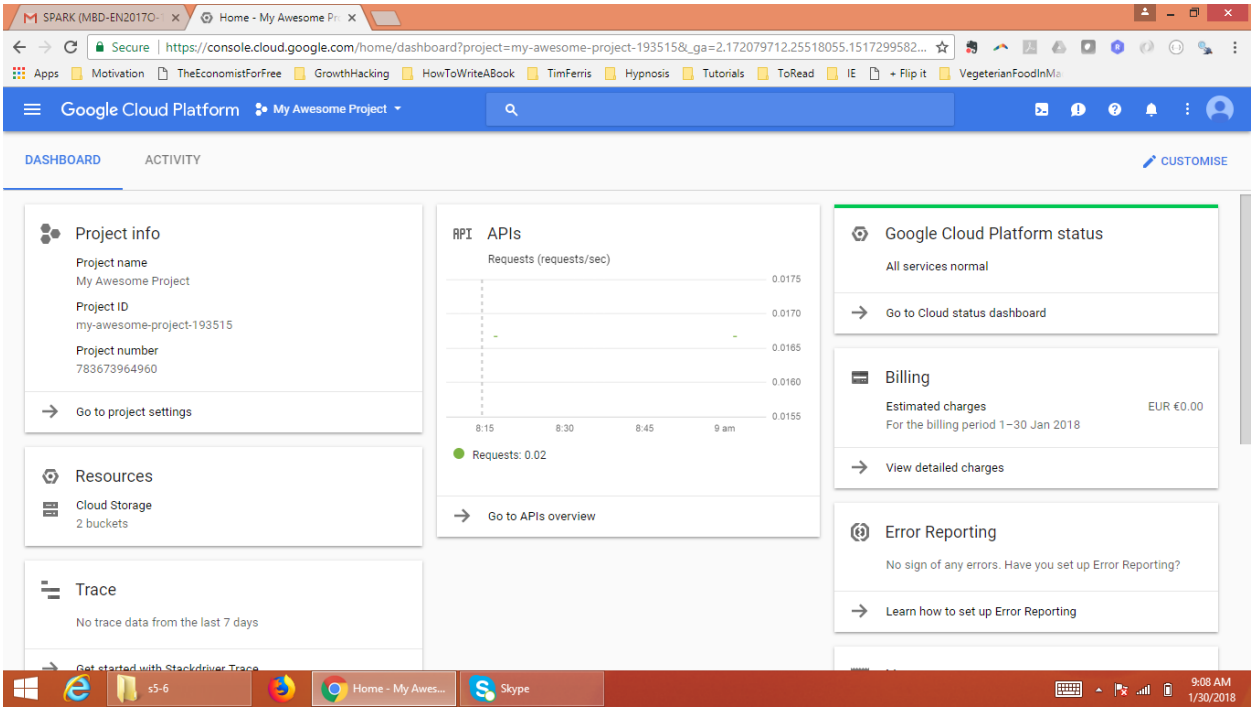
Kamal.nandan@student.ie.edu

(IE-MBD-O1)

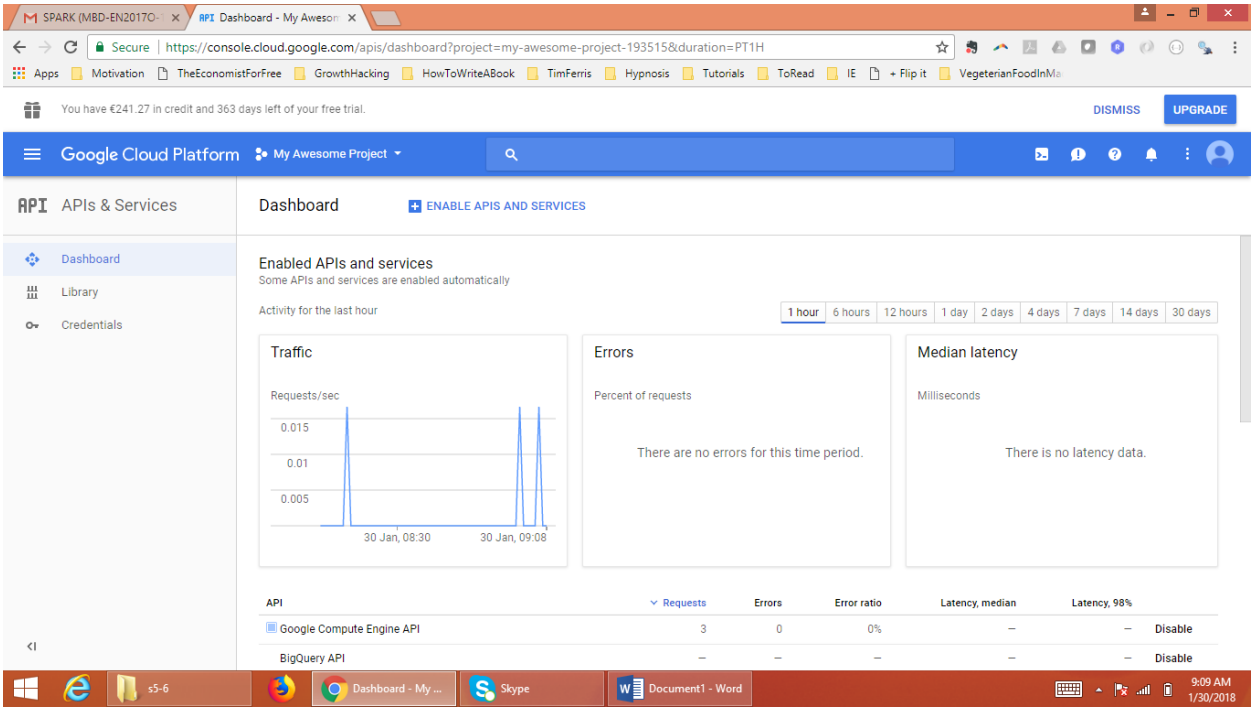
(I had created the cluster already and deleted that, but since I hadn't recorded the screenshots, I am doing it again. Since I had already created the cluster, I wasn't required to enable the APIs again.)

So, here are the steps:

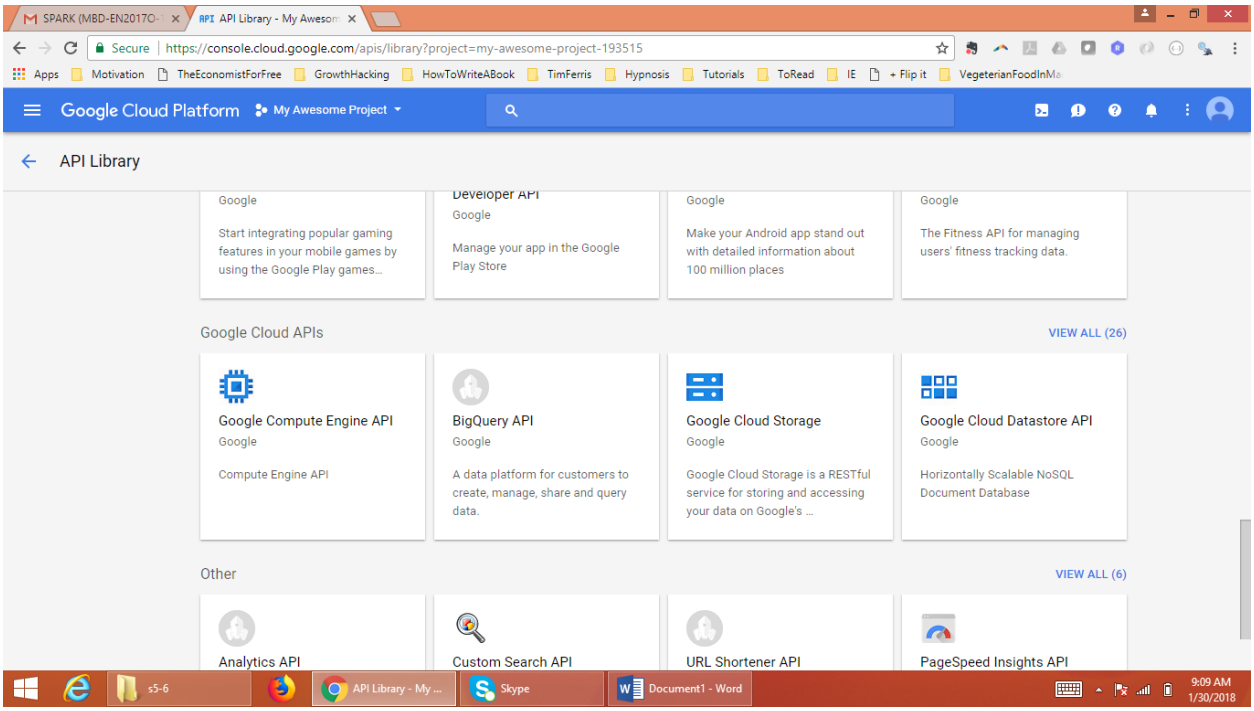
1. **Create "My Awesome Project" and browse to the project.**



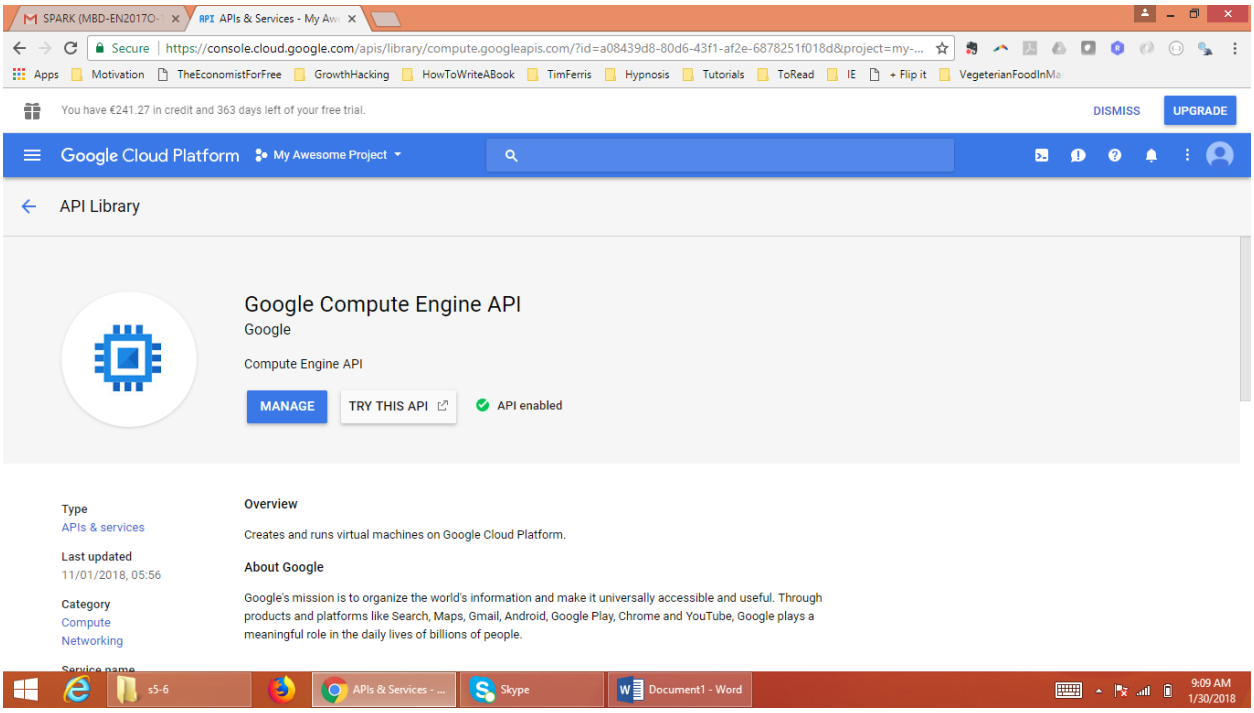
2. **Go to “APIs and Services” and click on “Enable APIs and services”**



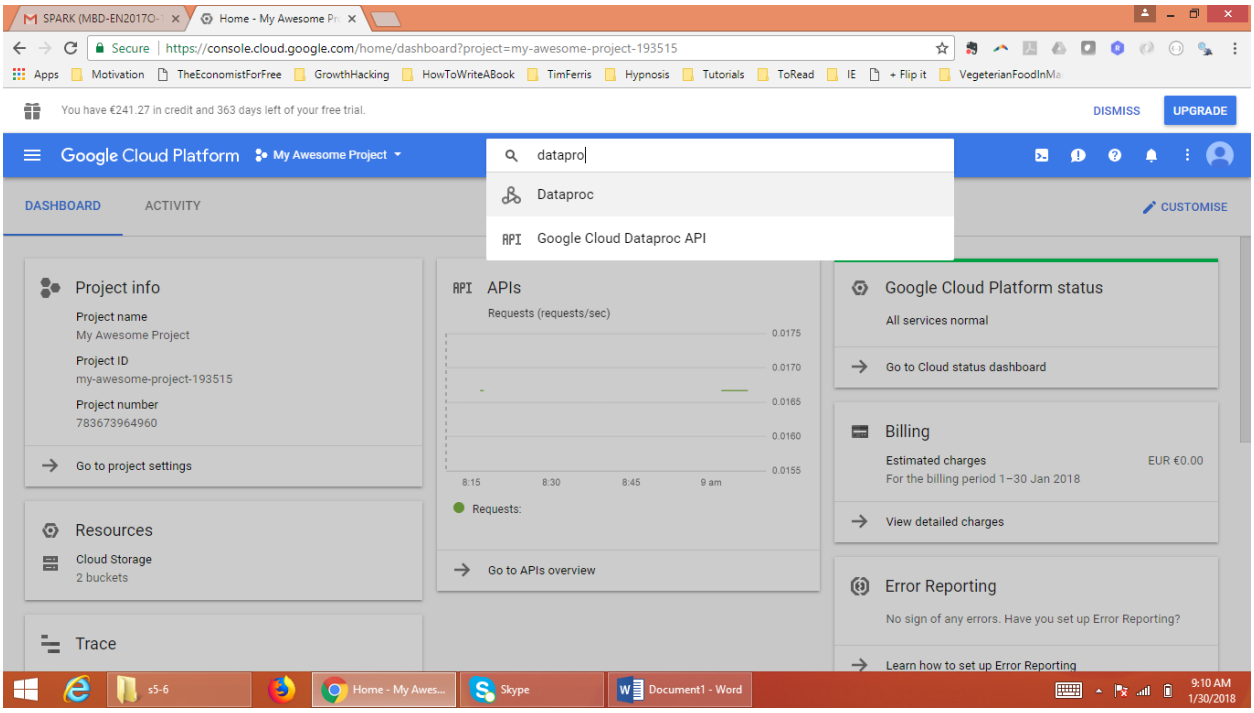
3. Go to “Google Compute Engine API”



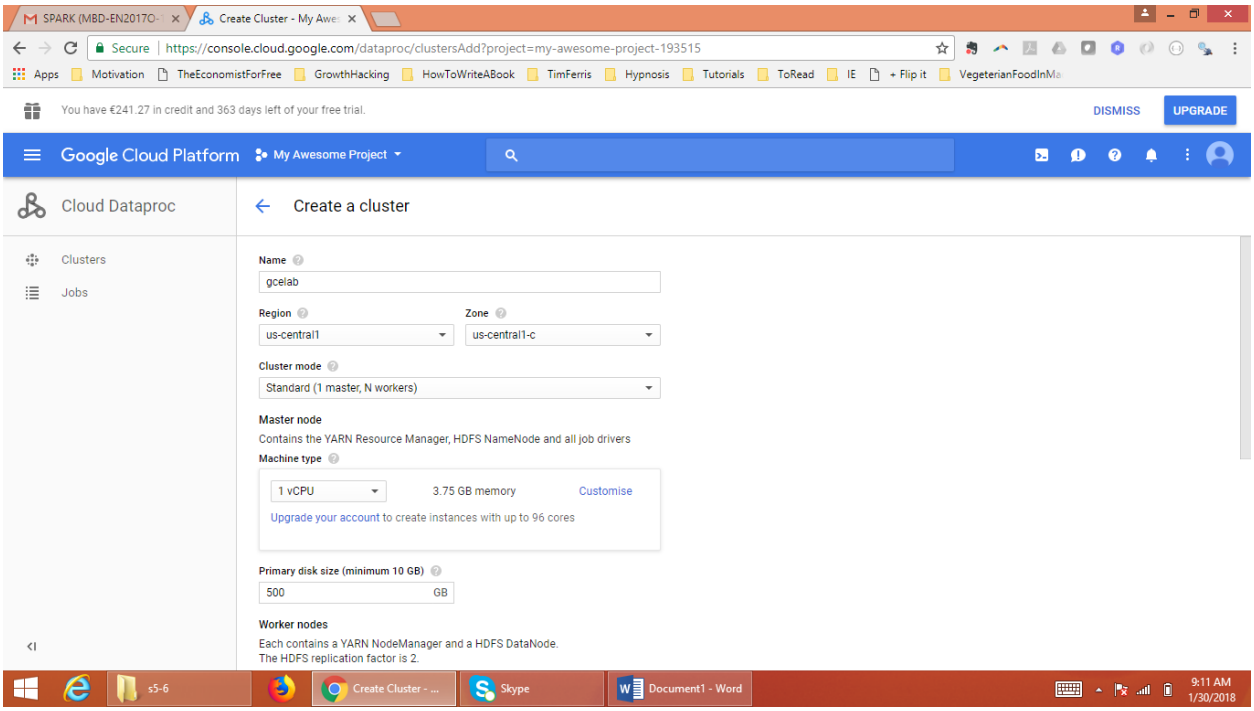
4. Enable “Google Compute Engine API”



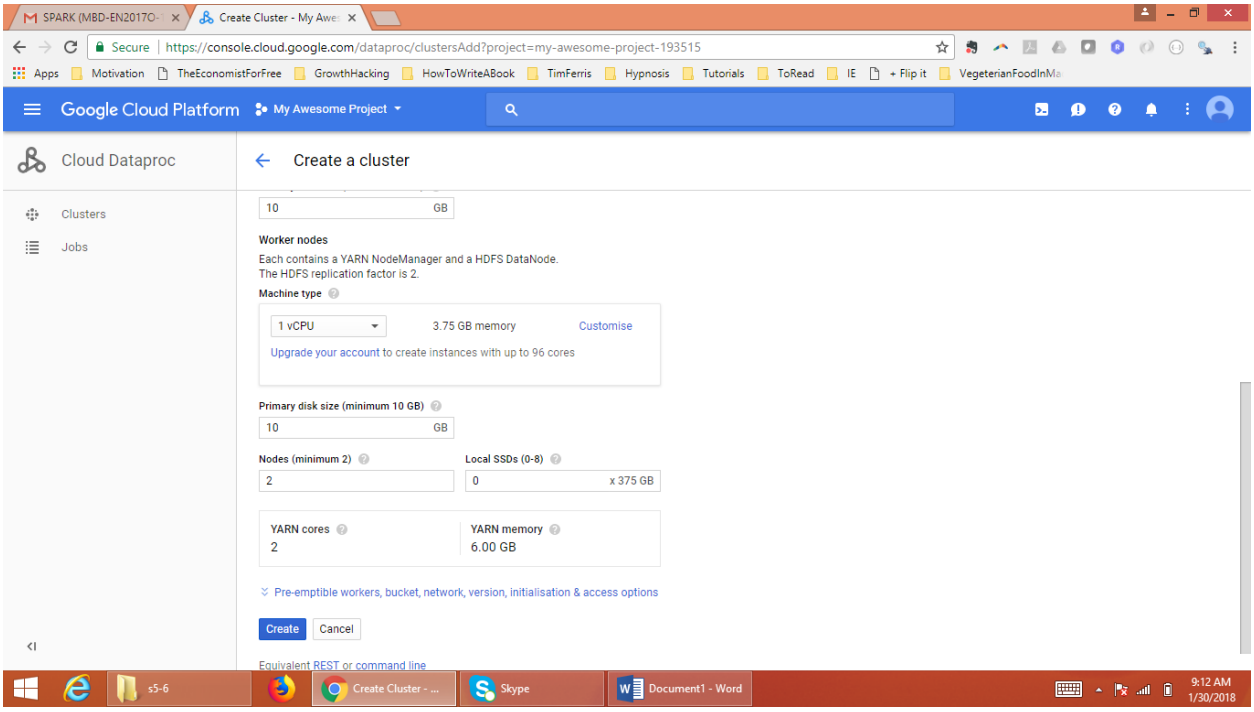
5. Go to “DataProc”



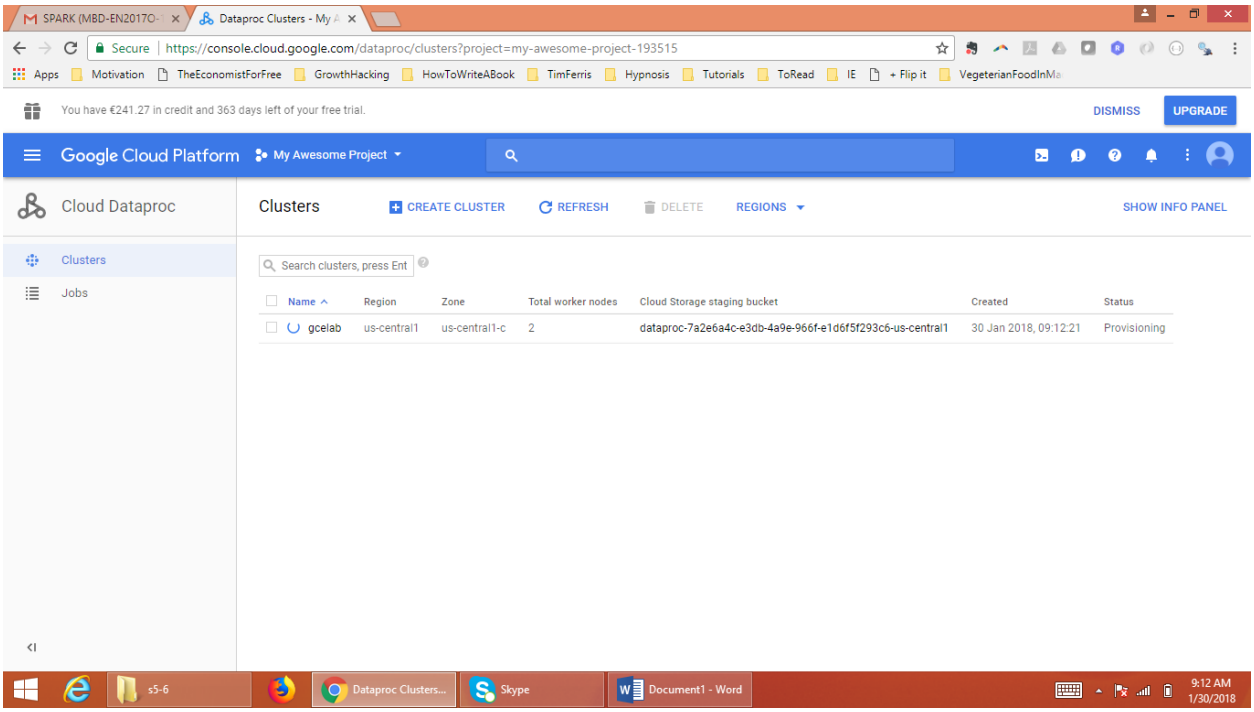
6. **Click on “Create cluster” button and you will reach the “Create a cluster” interface.**



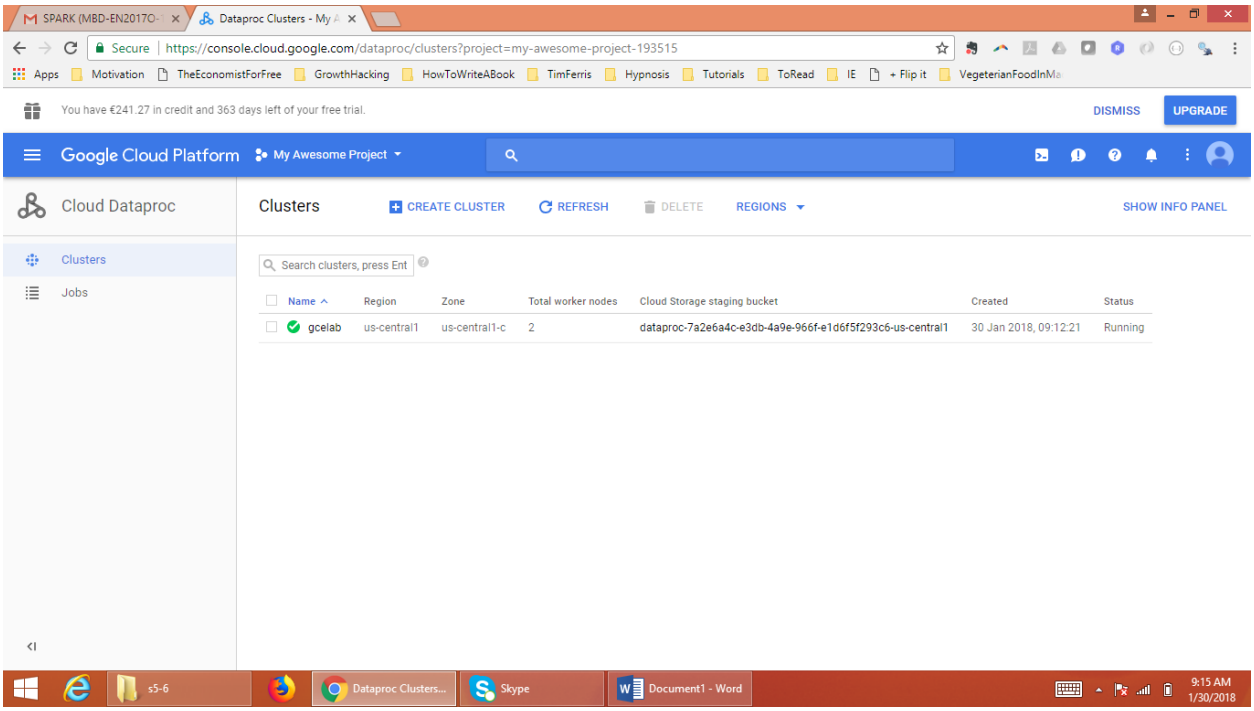
7. **Provide the configuration details. (To keep it cheaper I went for the minimal resources i.e. 1 Master Node and 2 worker nodes; for each node too, I chose the lowest possible configuration as seen in the following screenshot).**



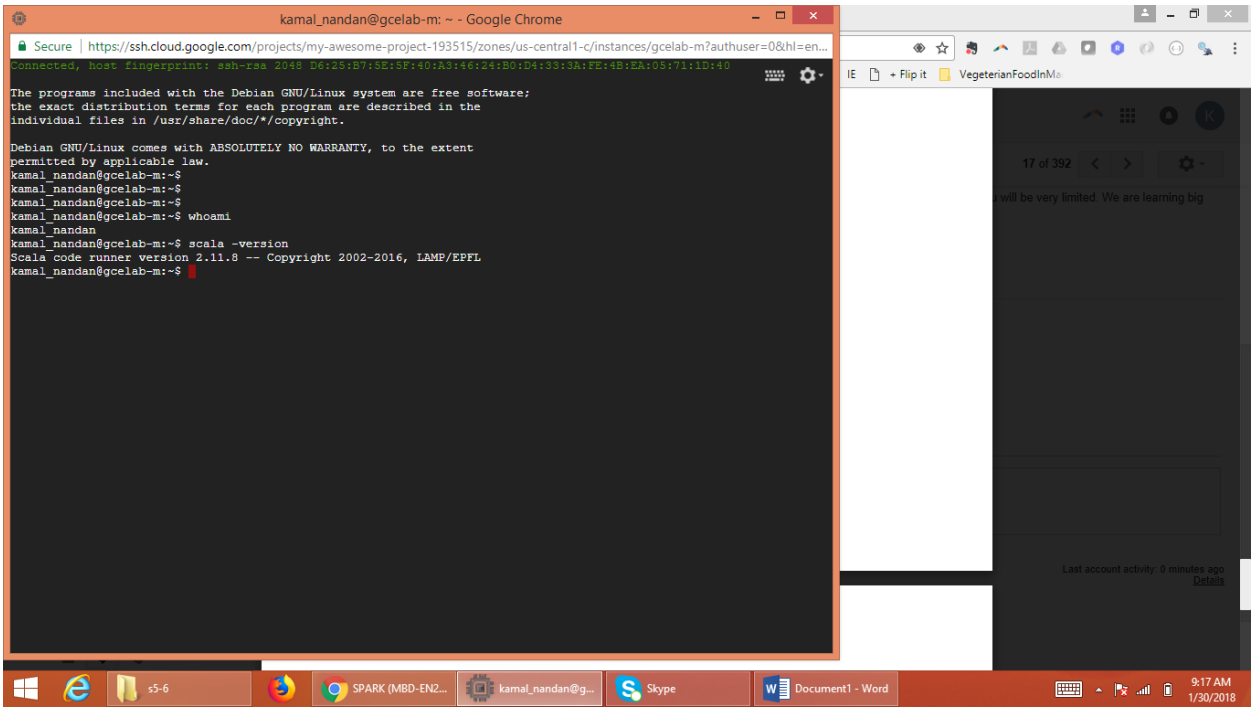
8. **After providing the configuration details, click on Create button and we see that the process of cluster creation has started.**



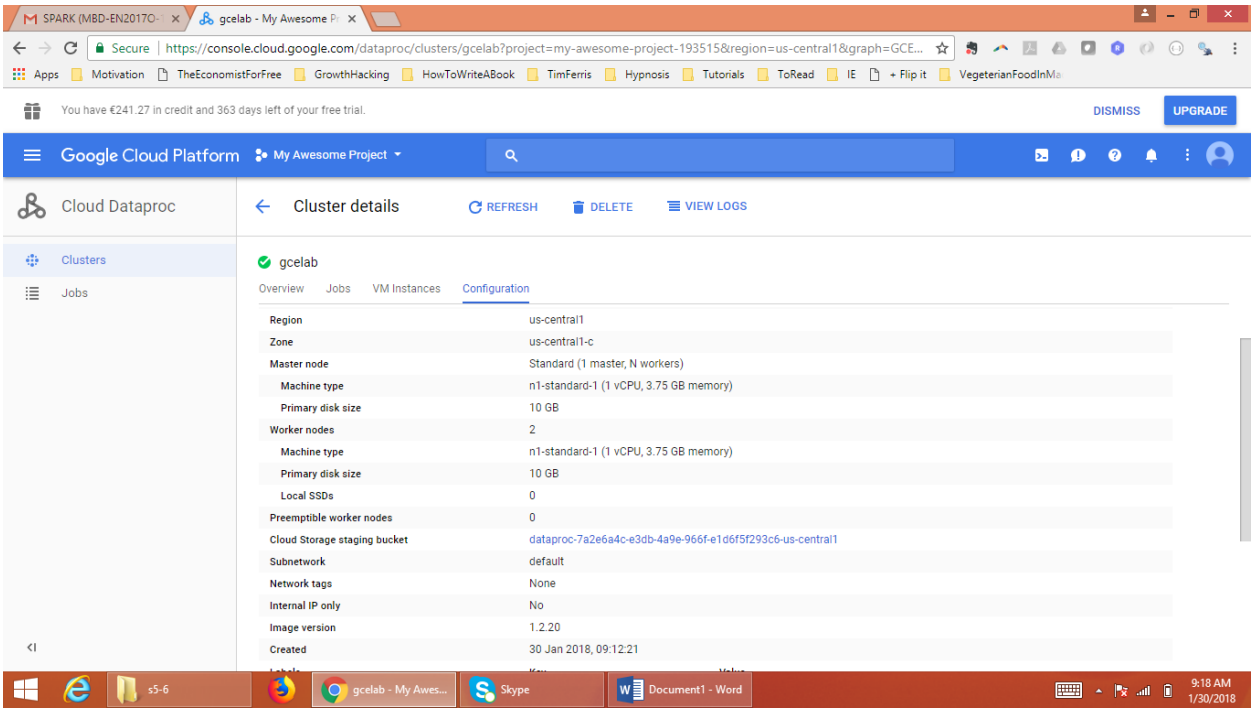
9. Cluster created and its running now.



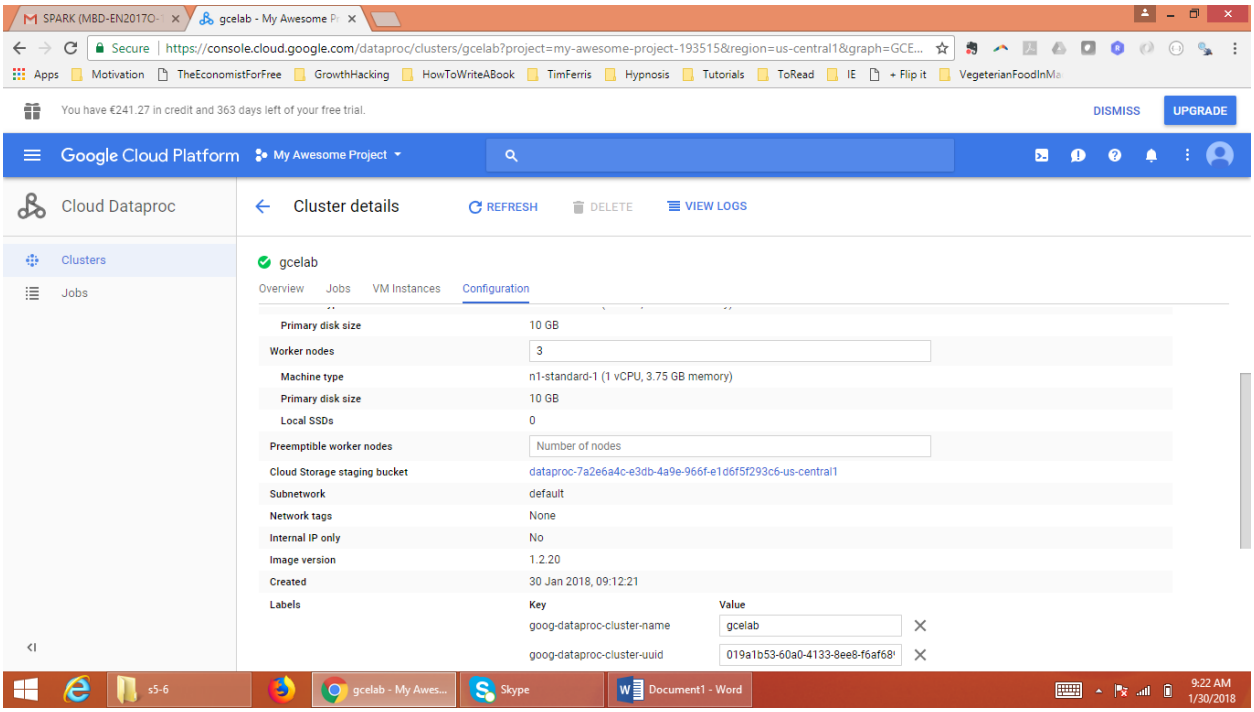
10. **Go to “VM instances” Tab and there we will see the option to connect to our cluster through the SSH interface provided by dataproc itself.**



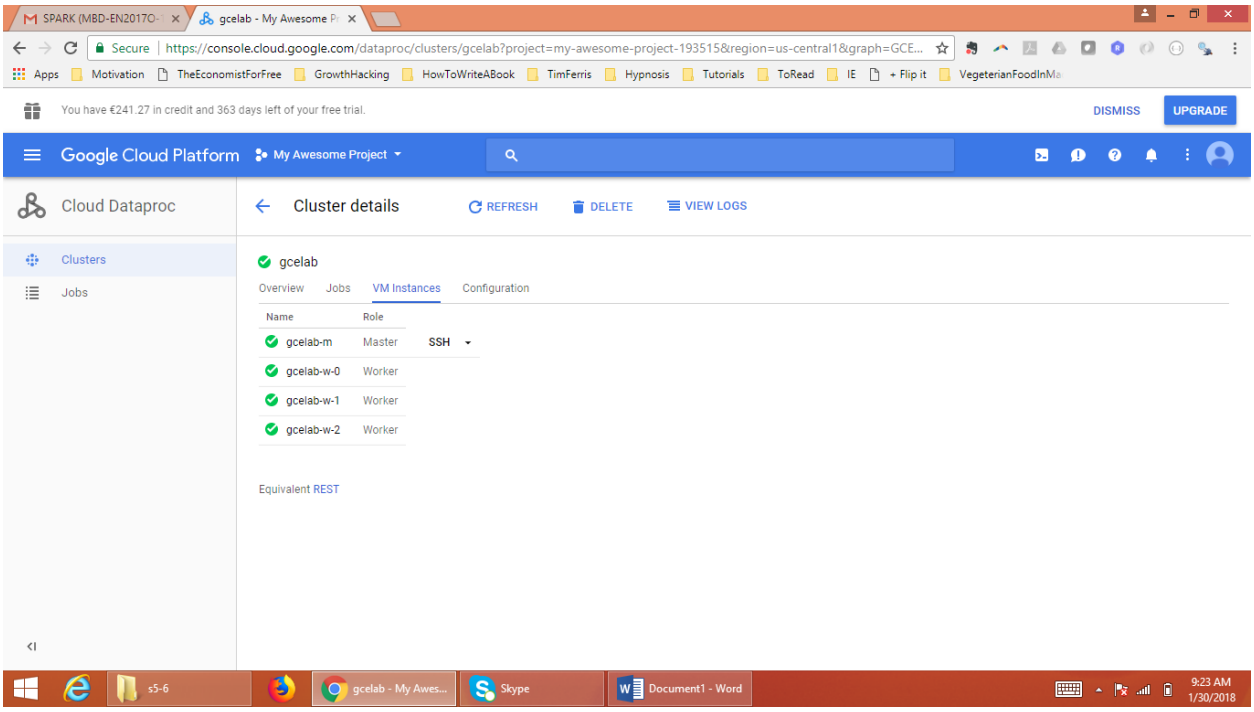
11. Now we will like to test scaling up the cluster – I went from 2 worker nodes to 3 worker nodes



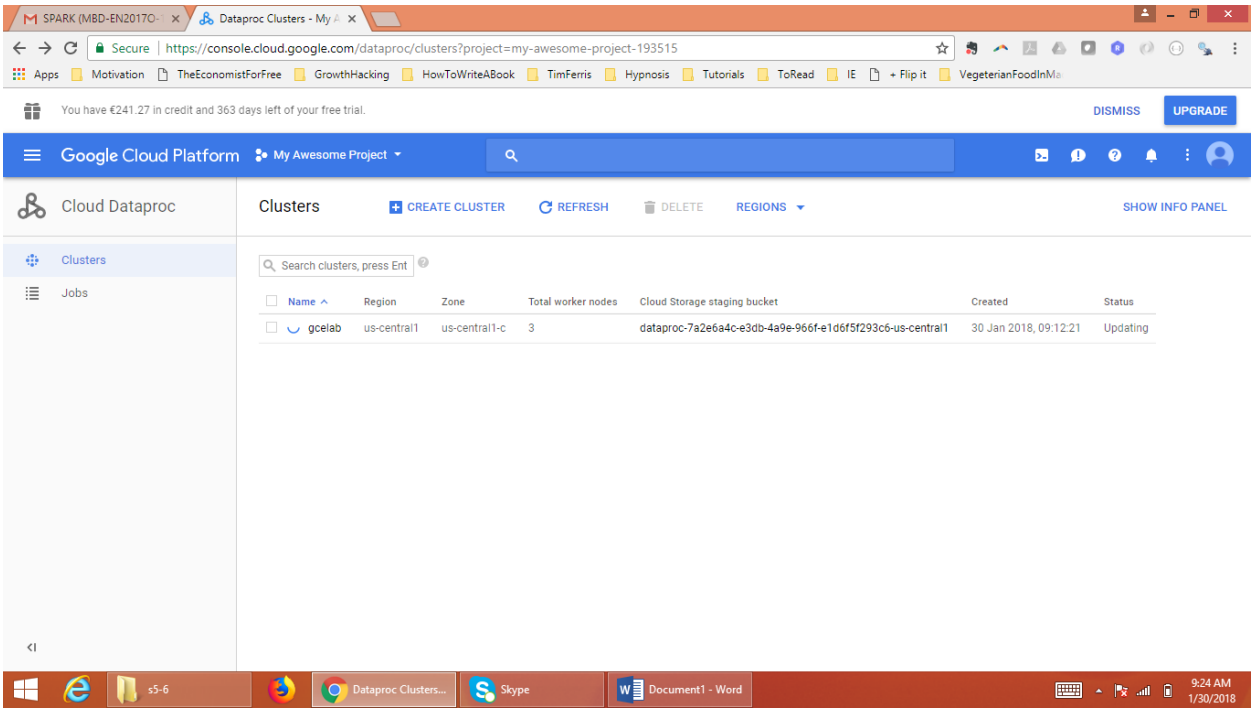
12. **Click on “Edit button” and provide the no. of nodes we want. Provide 3 and click on “Save” button. Cluster scaling-up process would start.**



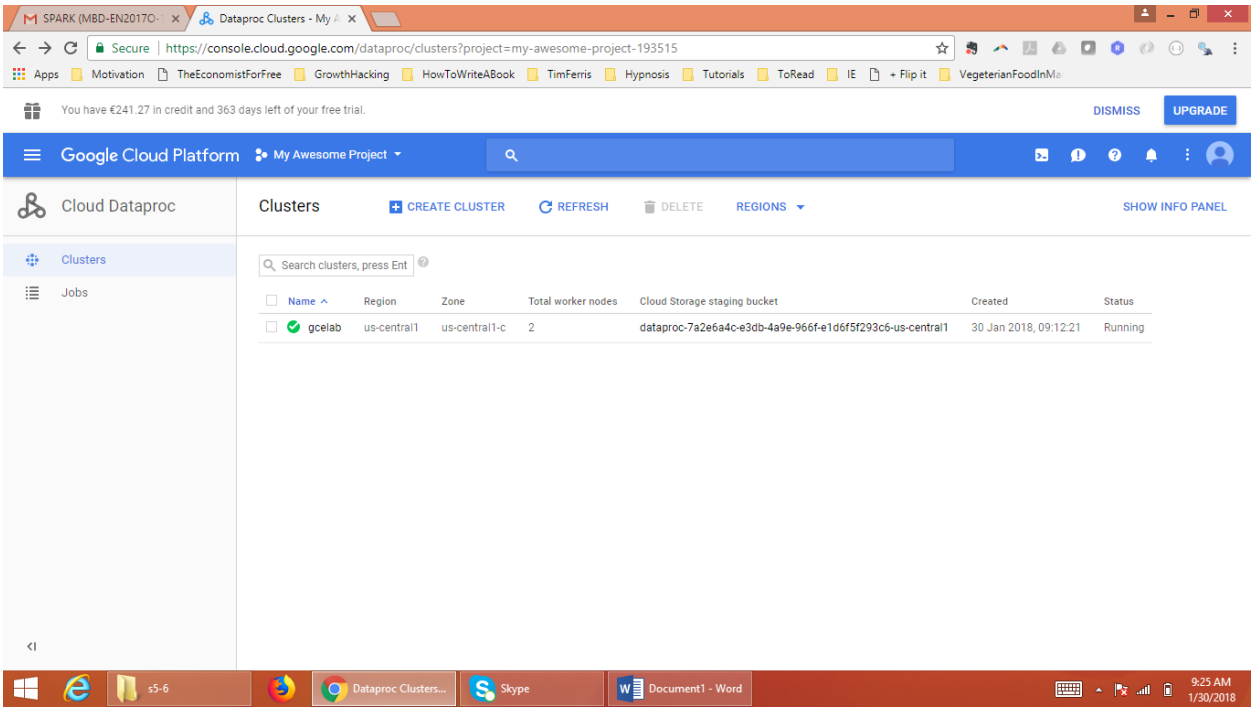
13. **We have successfully scaled up the cluster from 2 to 3 nodes. (It took me 50 seconds in scaling up. But, it is not consistent always – I tried a no. of times and it varied between 35 to 56 seconds)**



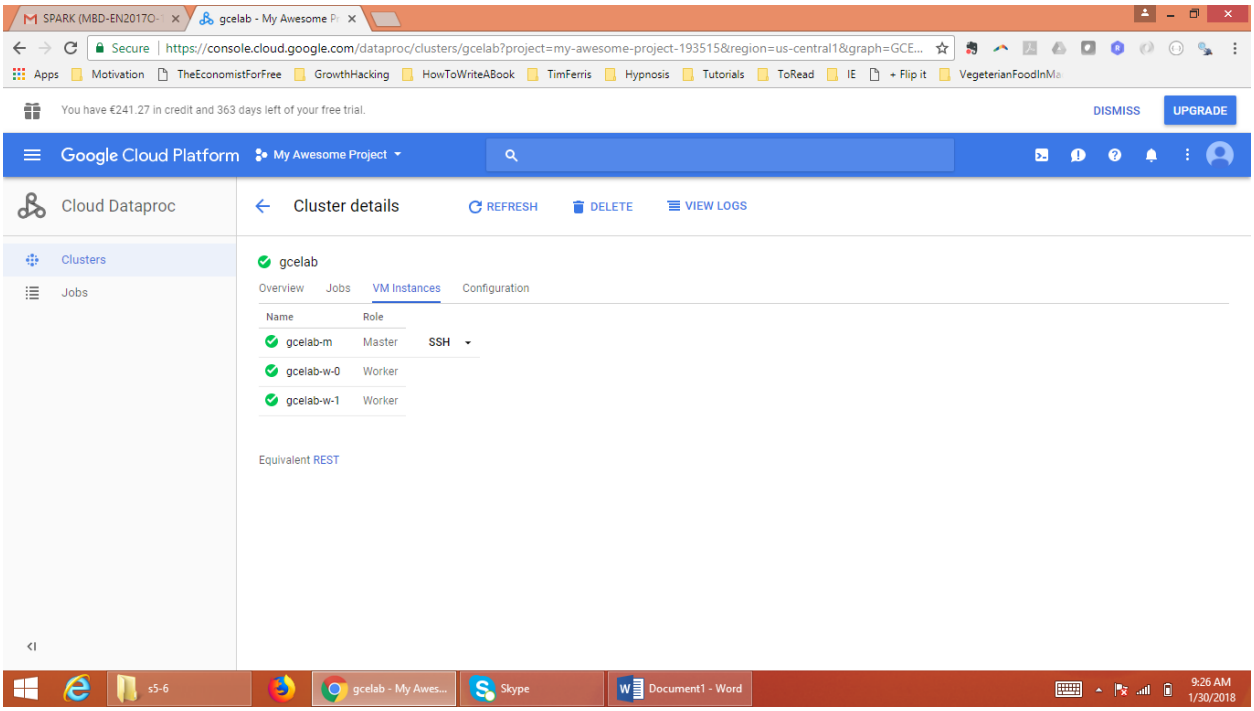
14. **We followed the same steps as we did in scaling up, to scale down the cluster – we scaled it down from 3 to 2 nodes.**



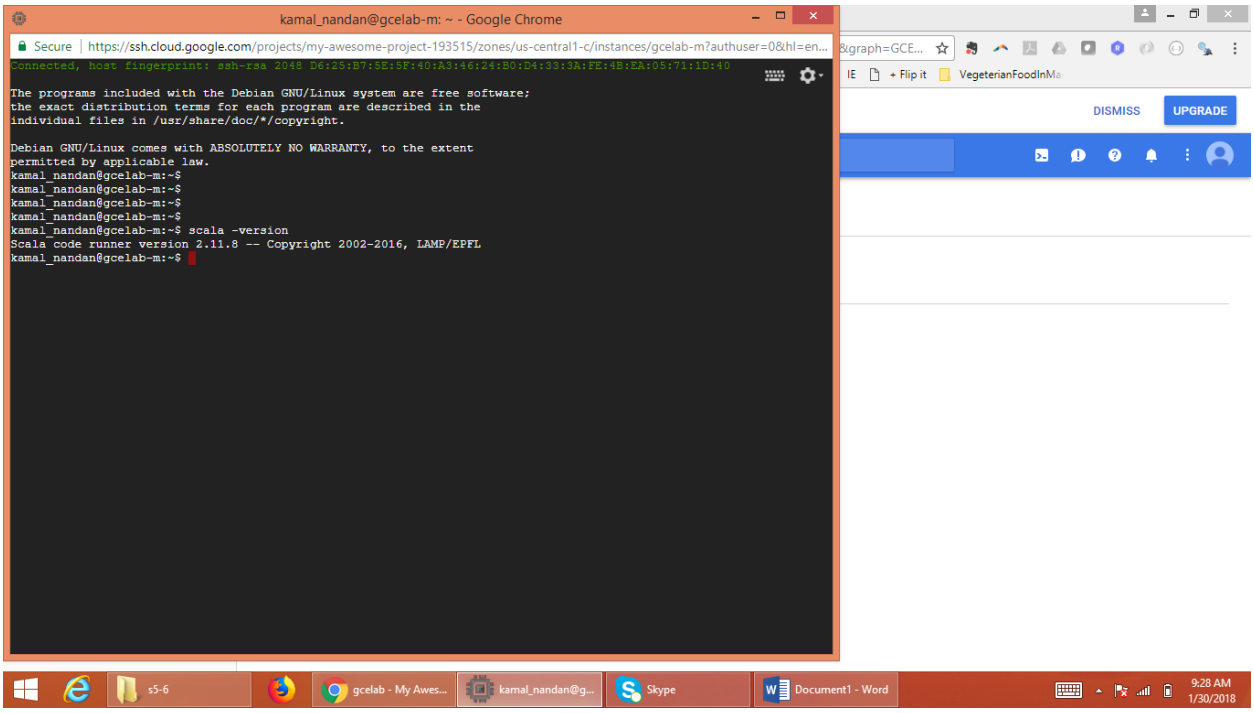
15. **Successfully scaled it down to 2 worker nodes. (Scaling down took 1 min. and 22 seconds and this time was more or less consistent.)**



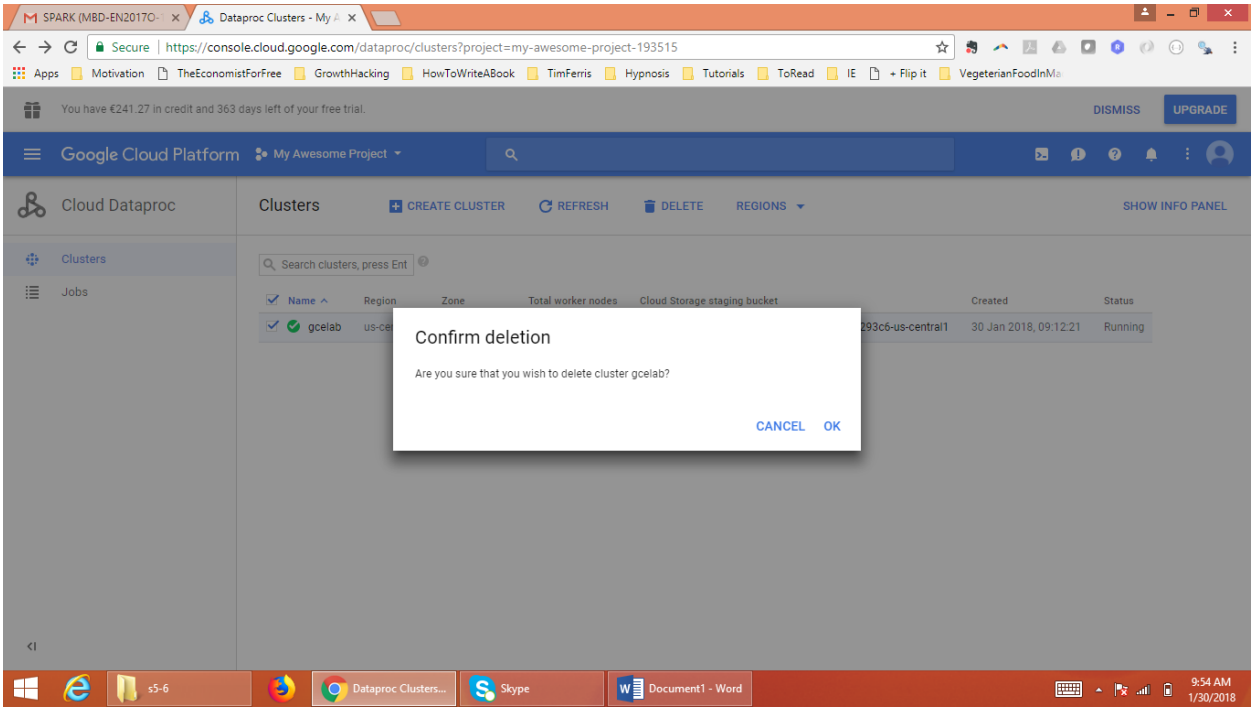
16. **Go to VM instances tab and click on “SSH”, and we would have a browser based SSH client.**



17. Check the scala version – its 2.11.8



18. **Now delete the cluster – we don't need it for now.**



19. **As an extra step, I also tried to login through putty ssh client, which I find more convenient. For this, I had to generate public/private ssh keys and provide my public key to dataproc and then login through putty using my private key. I have not taken the screenshots of this process.**