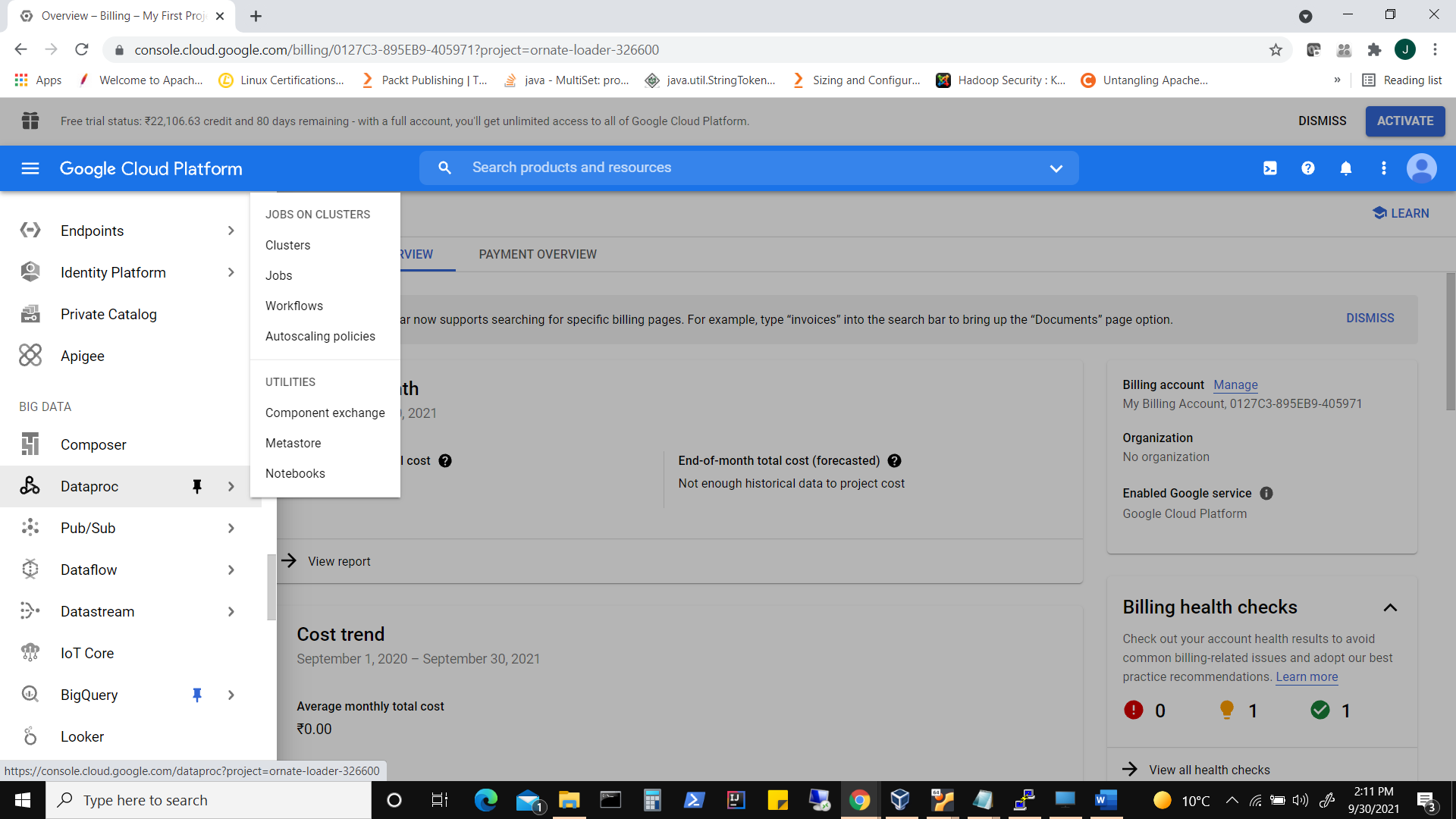
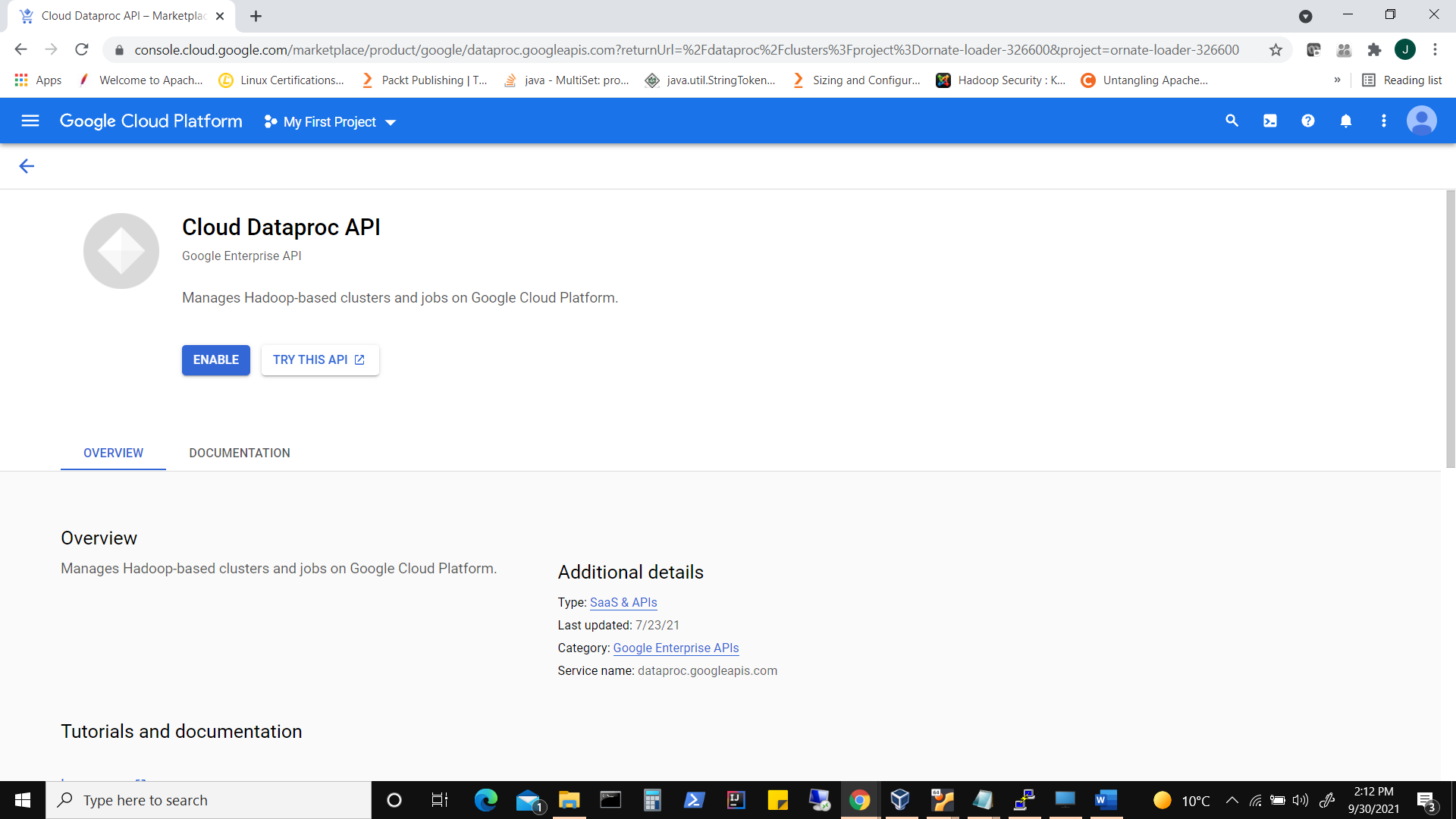
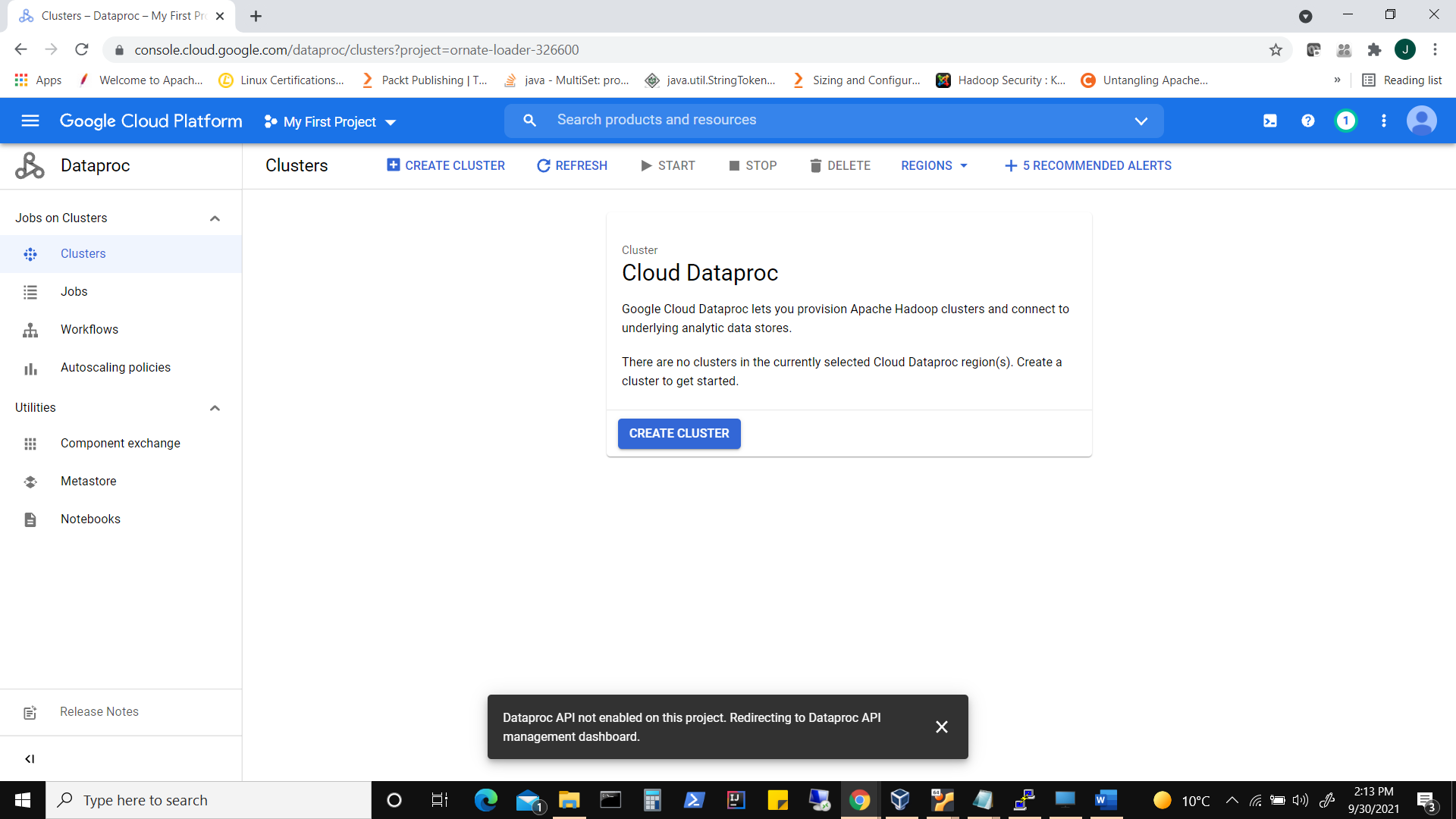
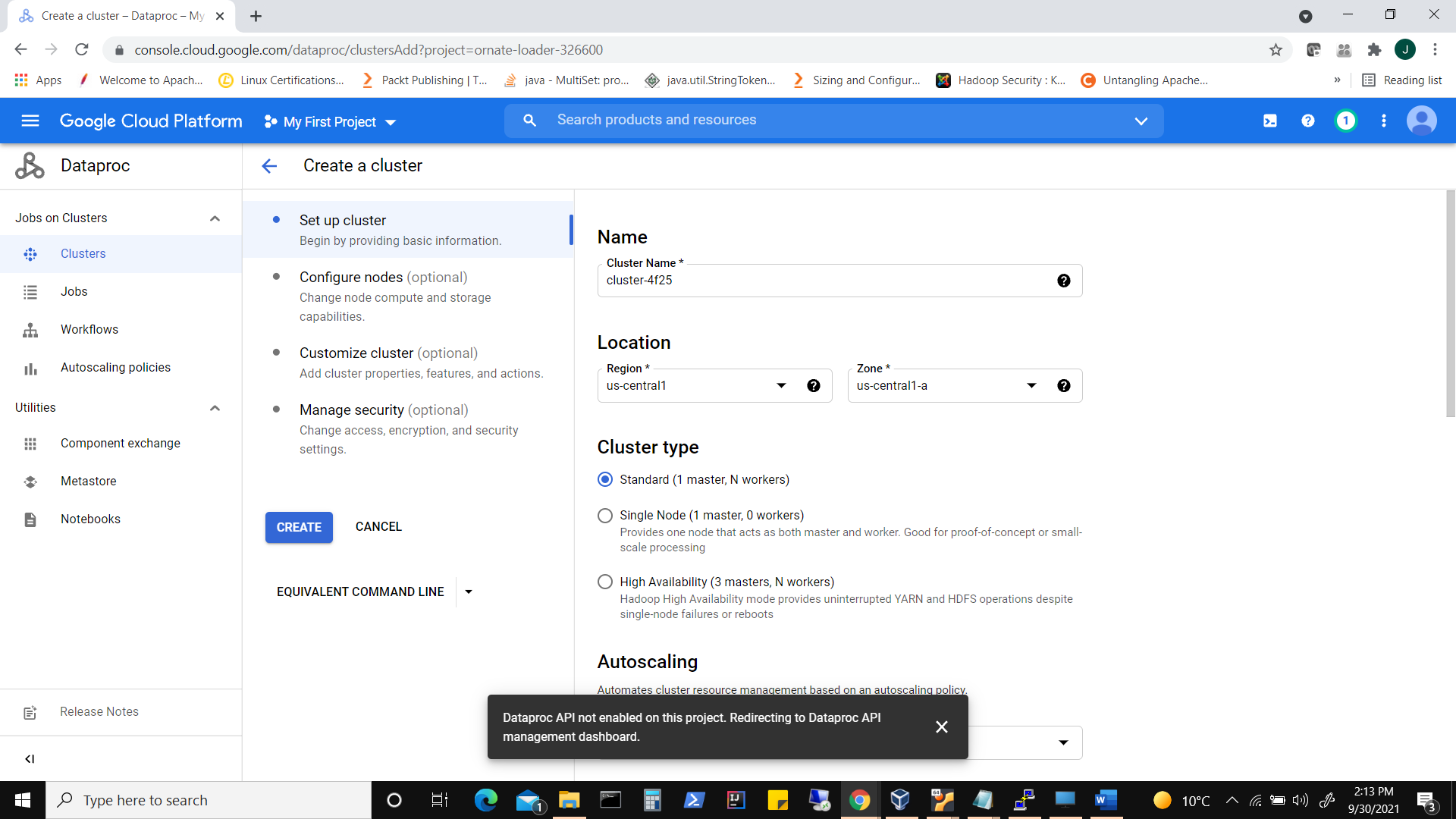
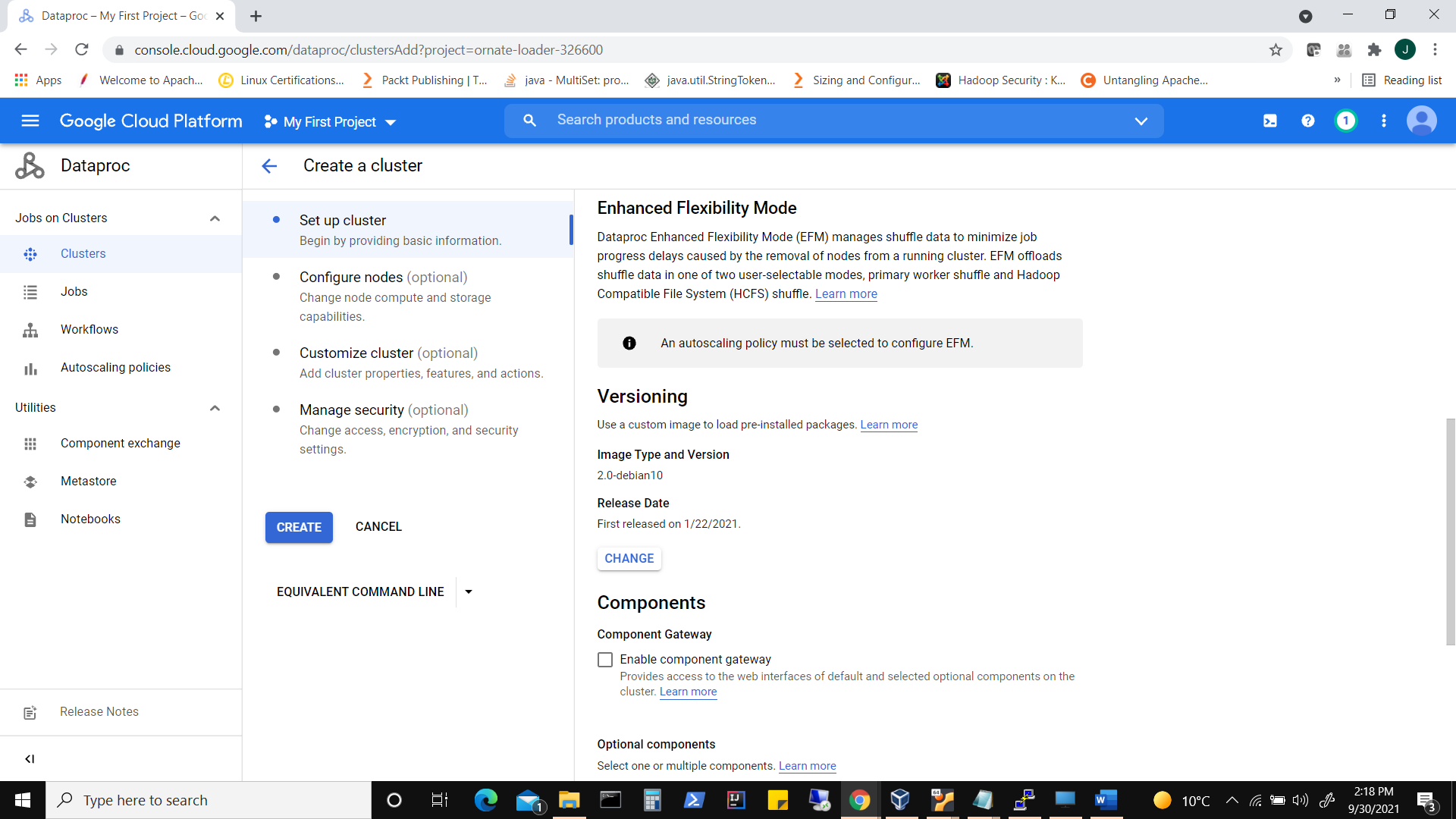
GCP



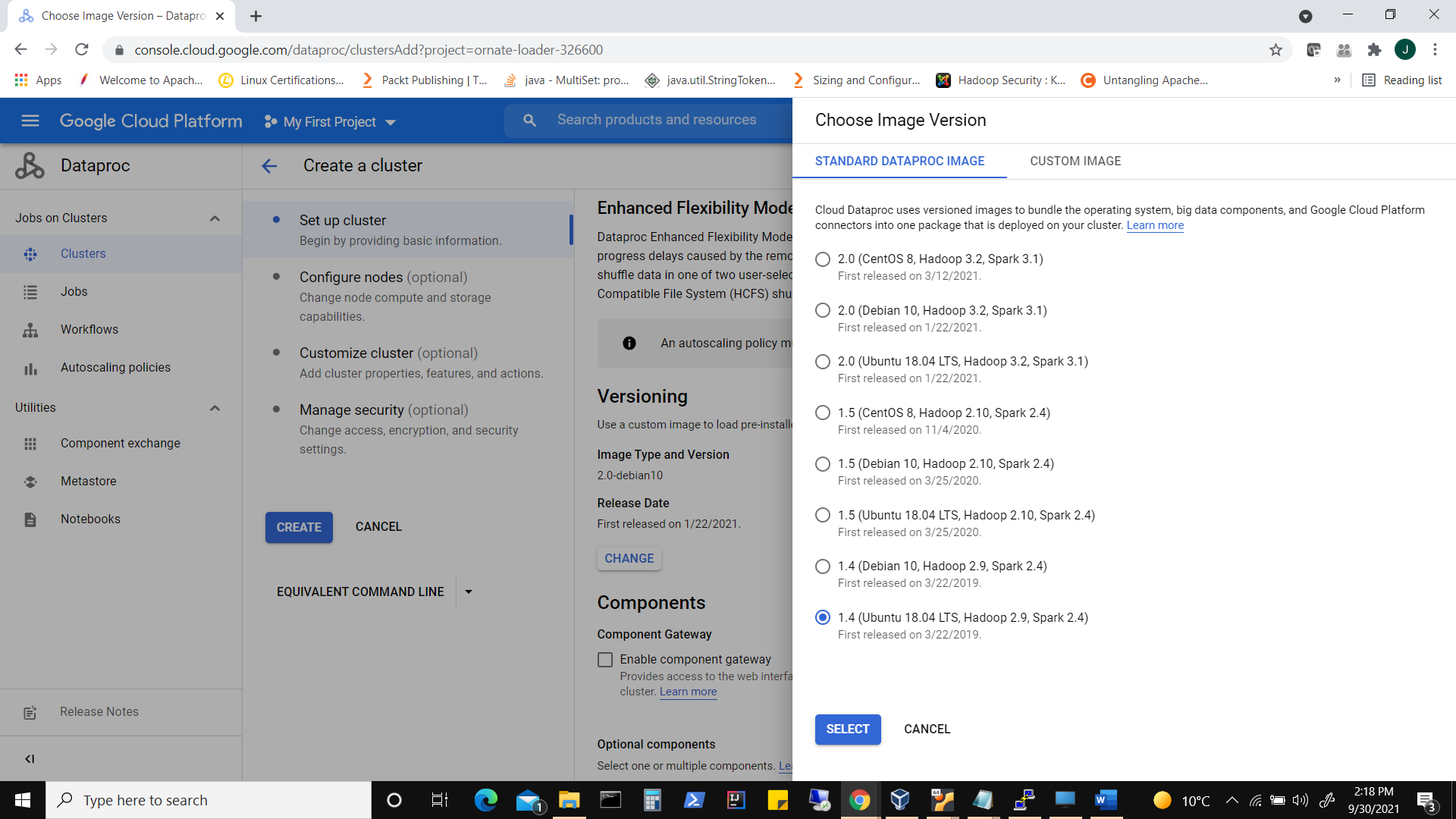


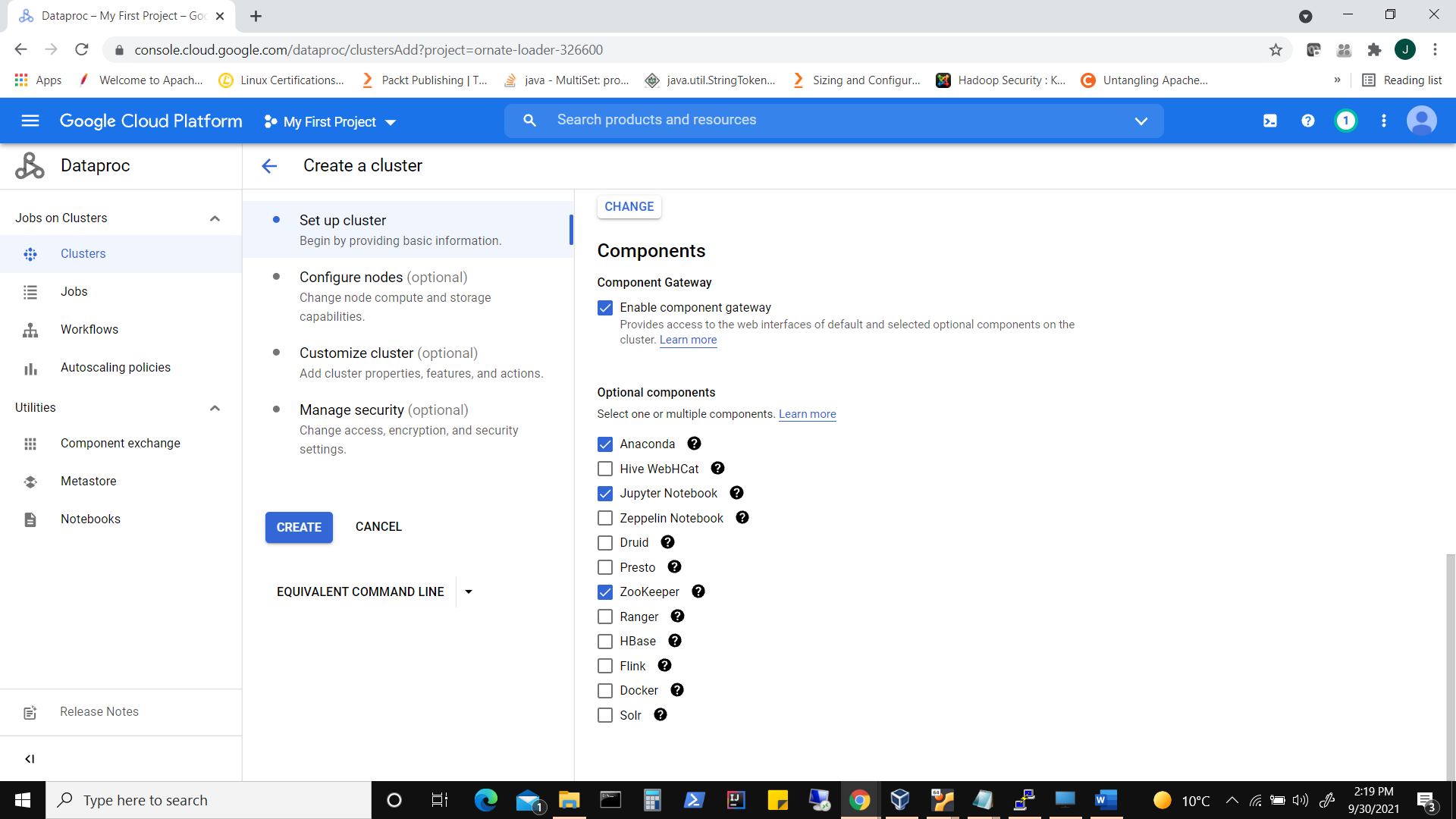


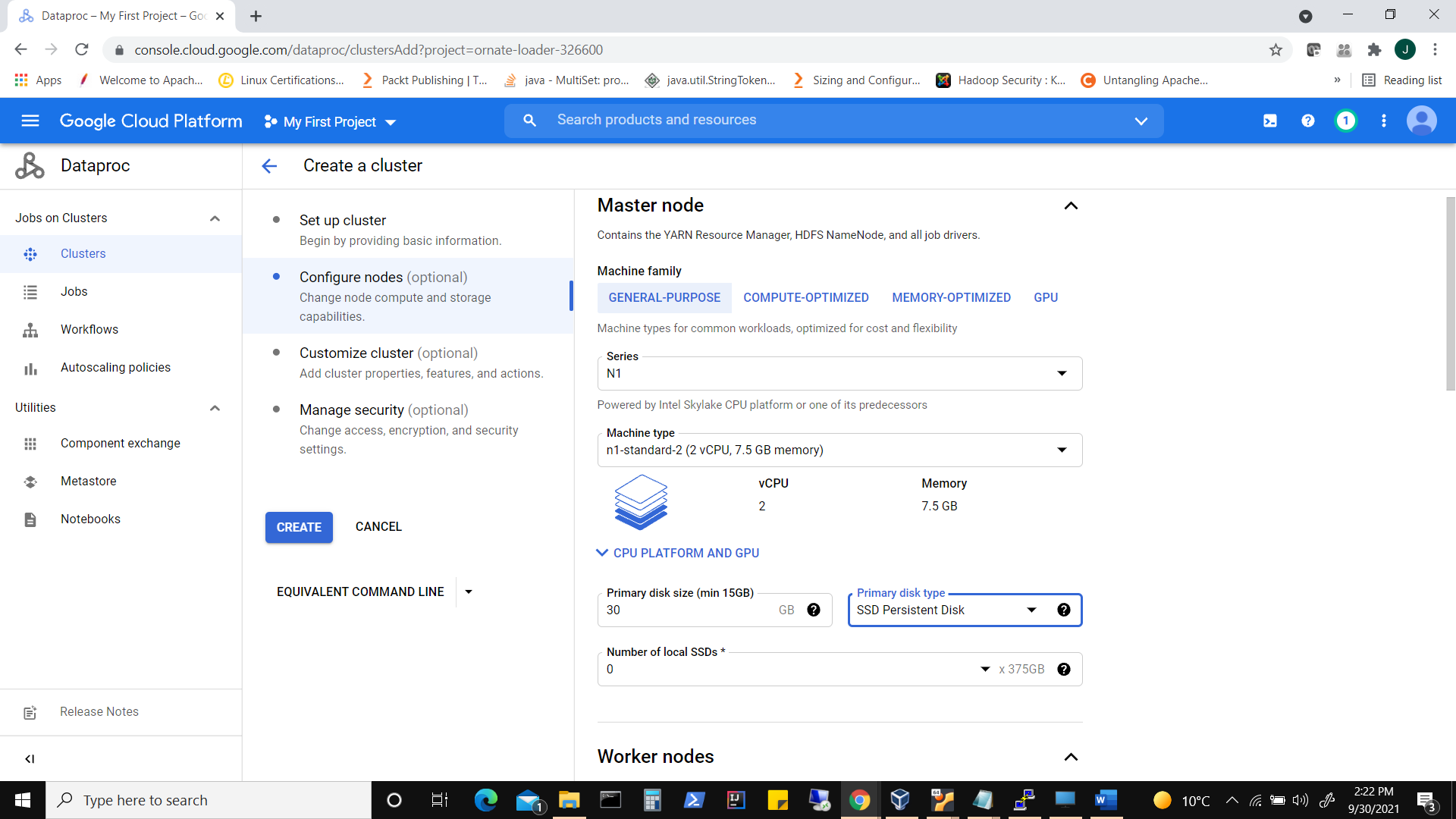


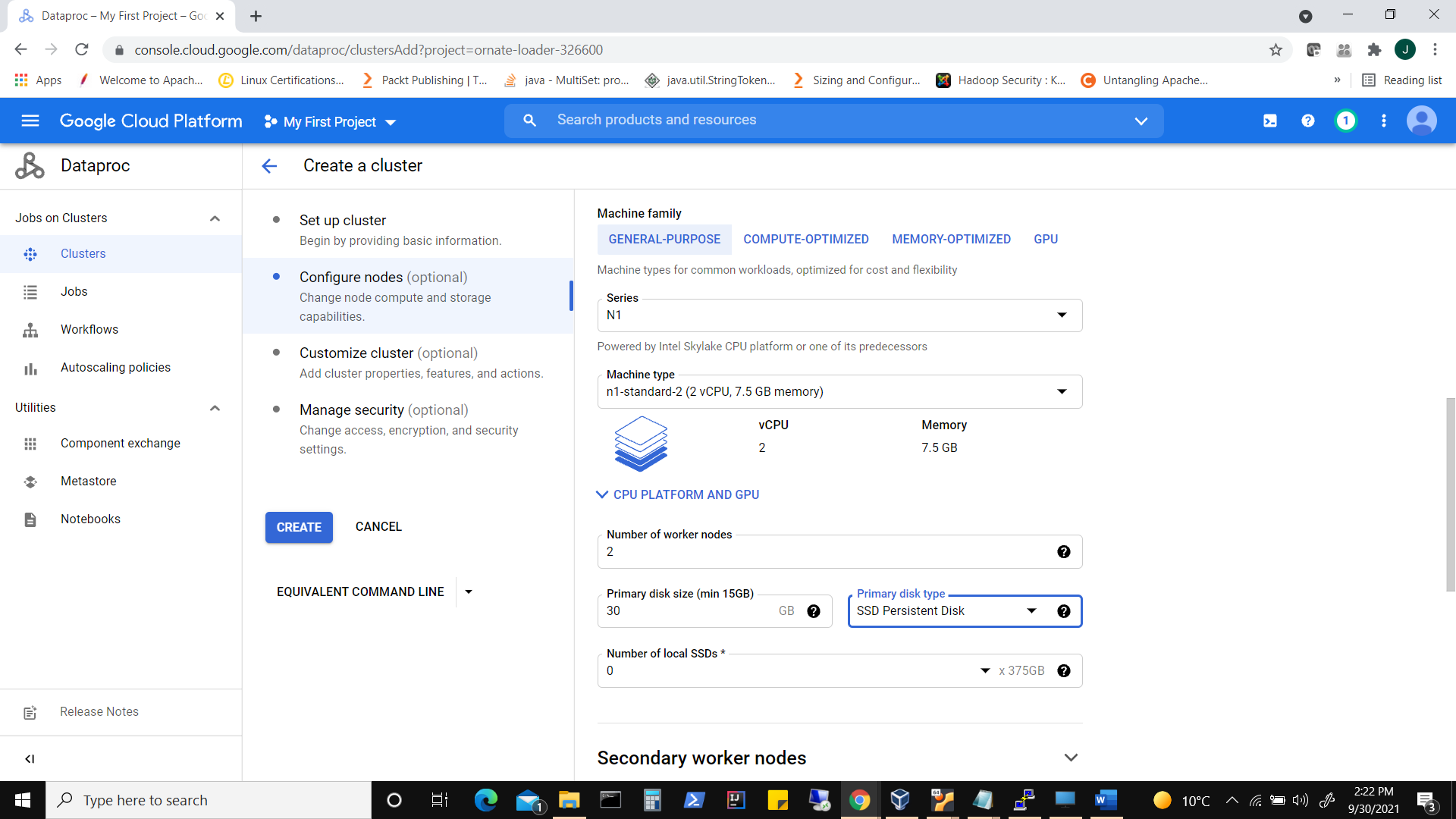


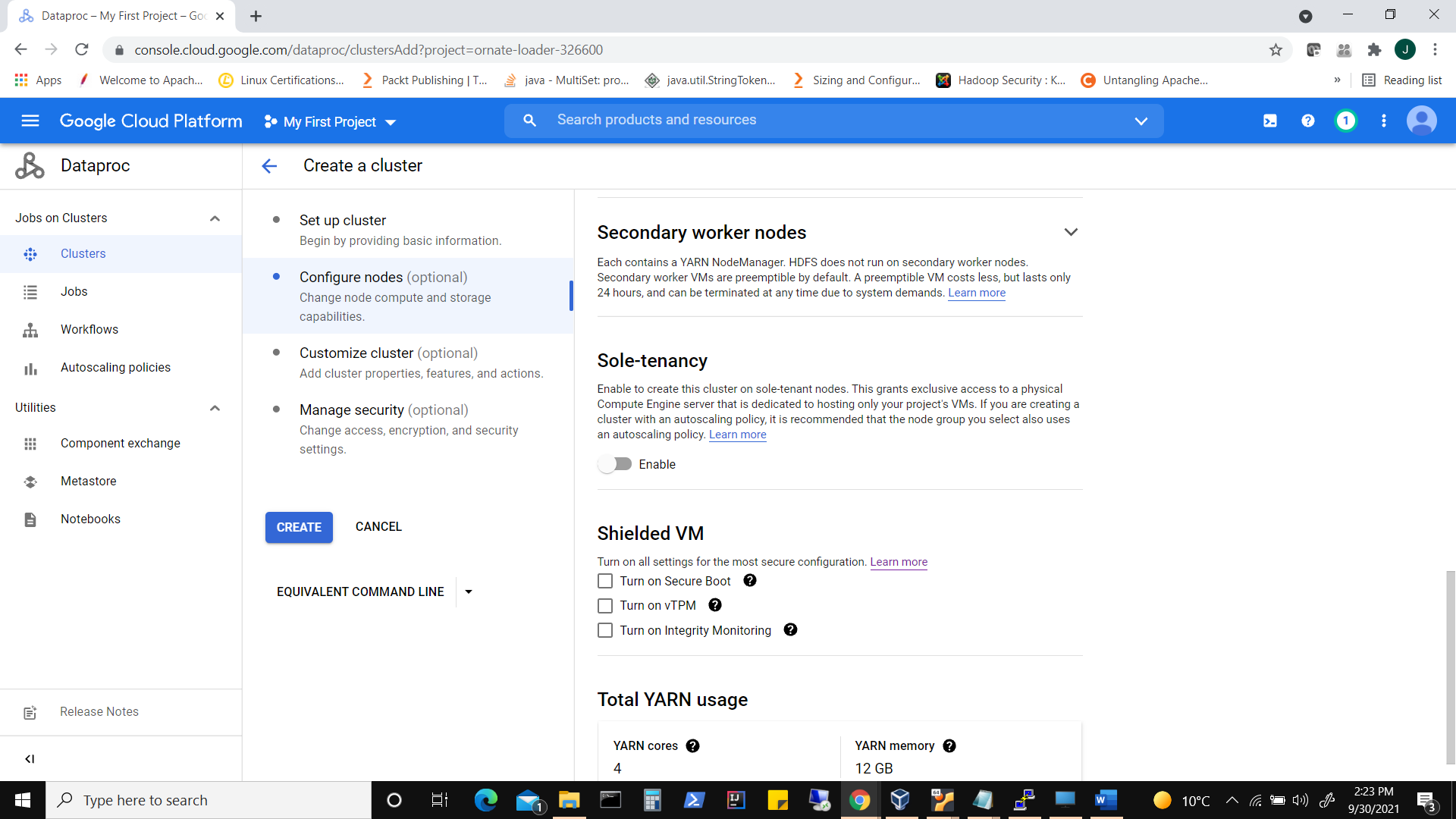
Click on Change to choose distribution and versions of packages

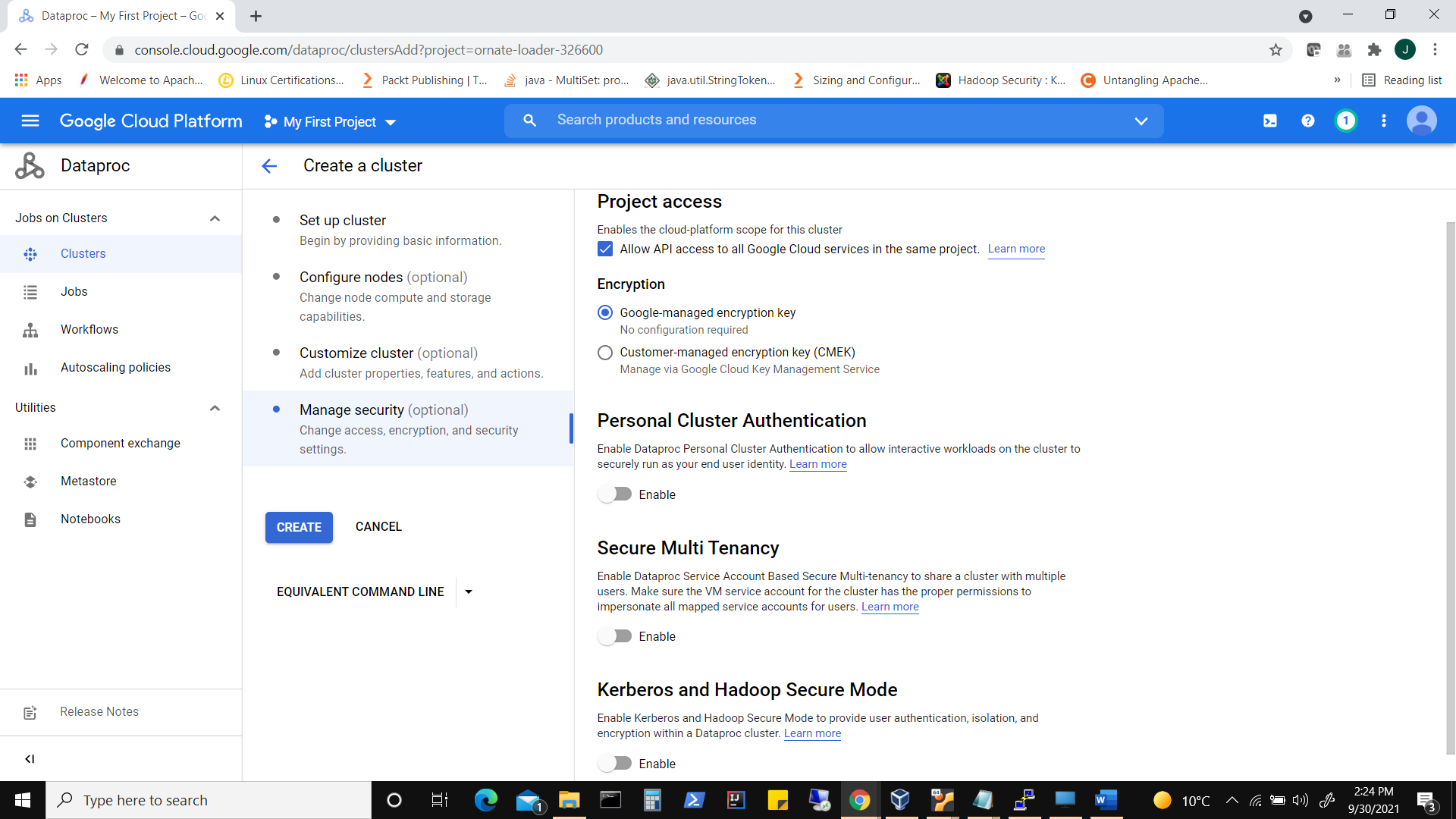


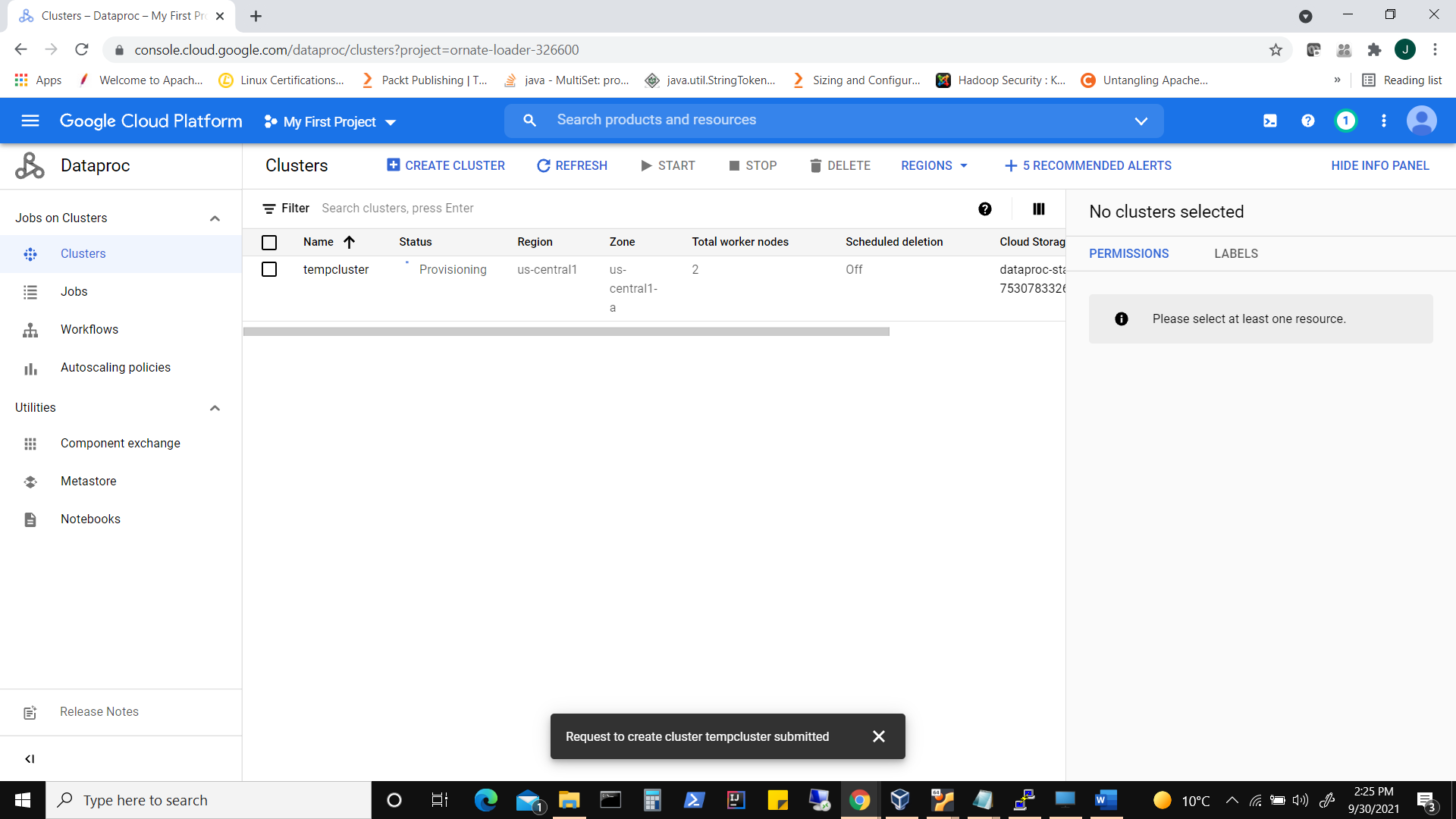






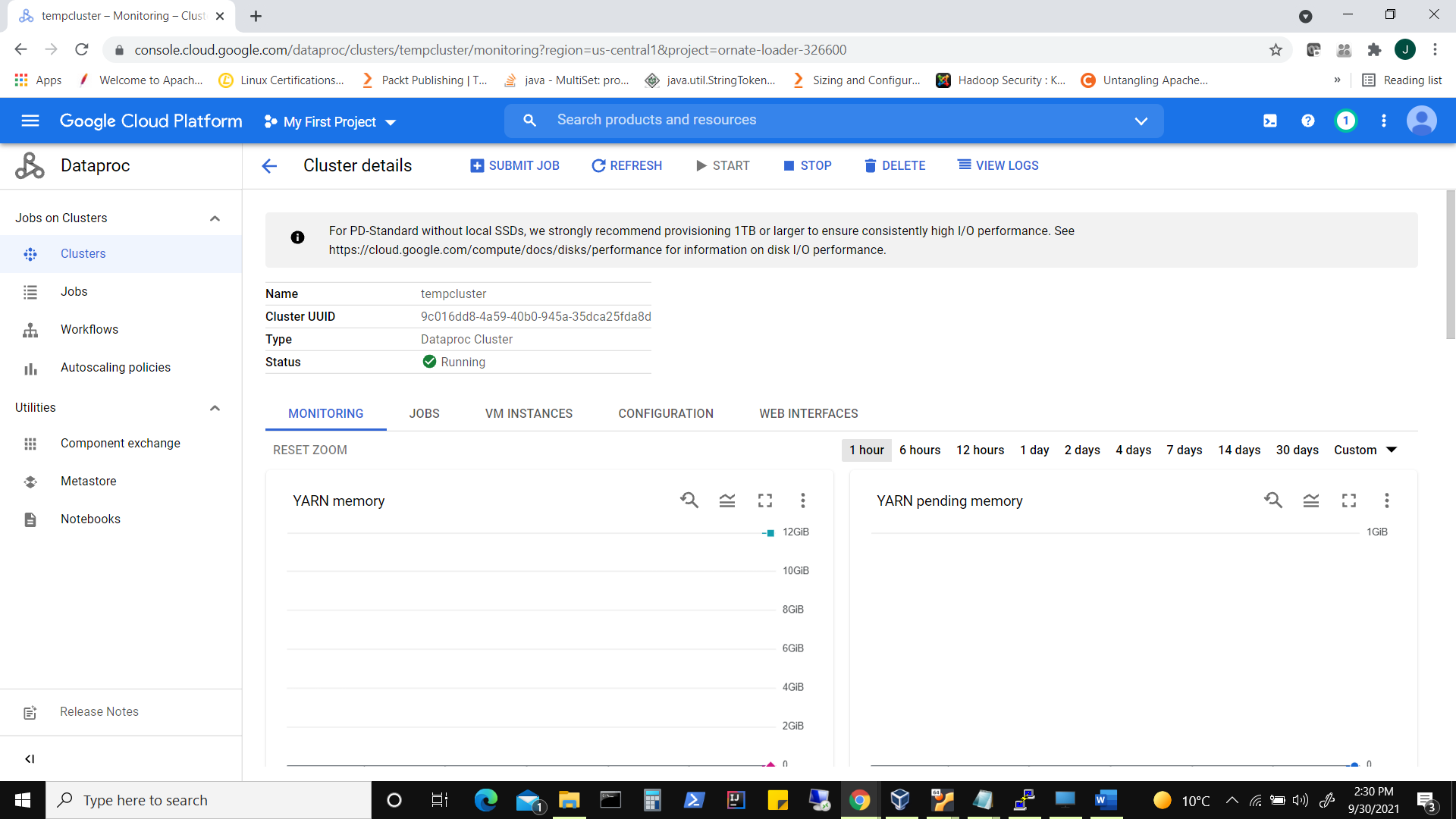


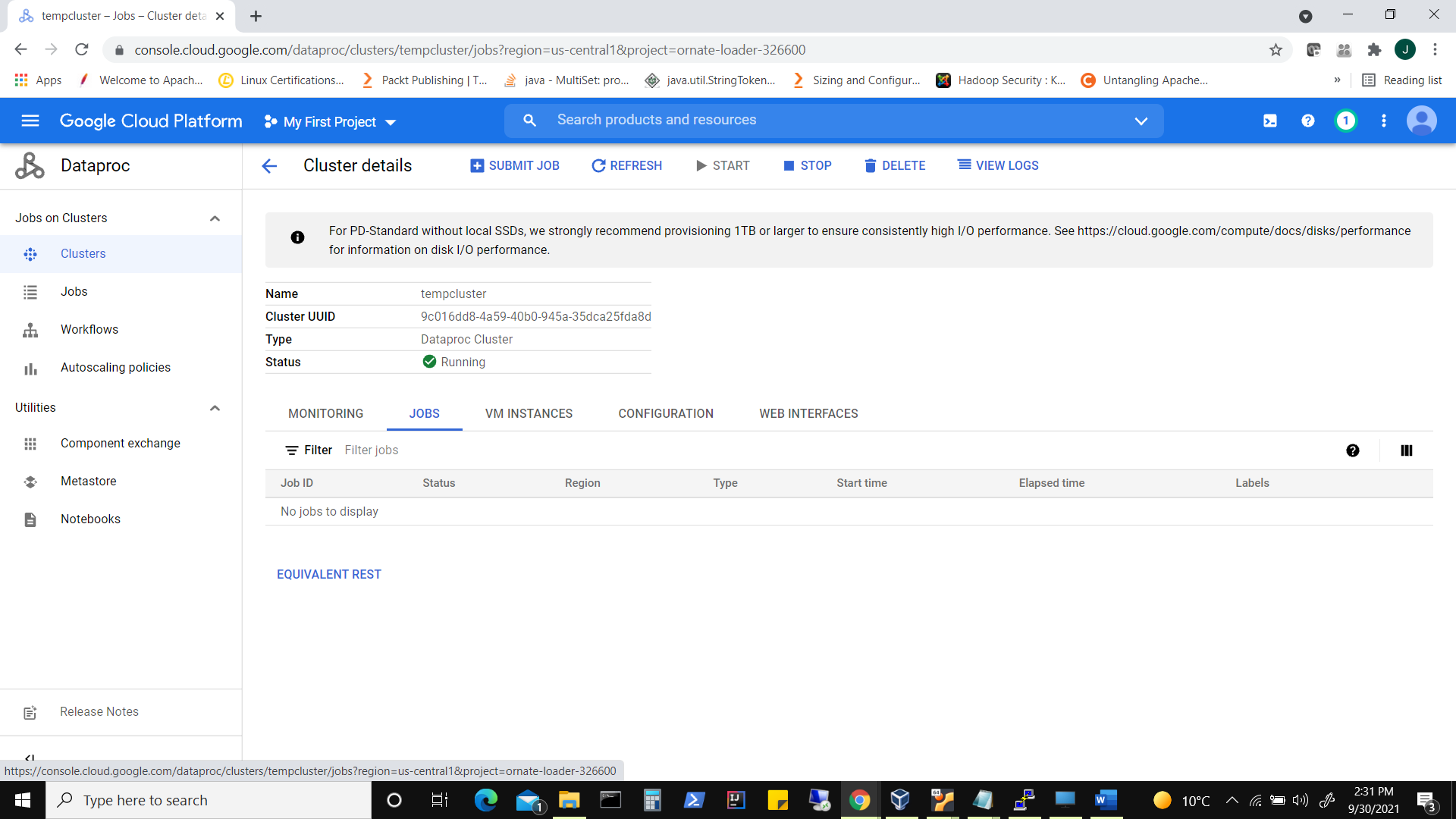


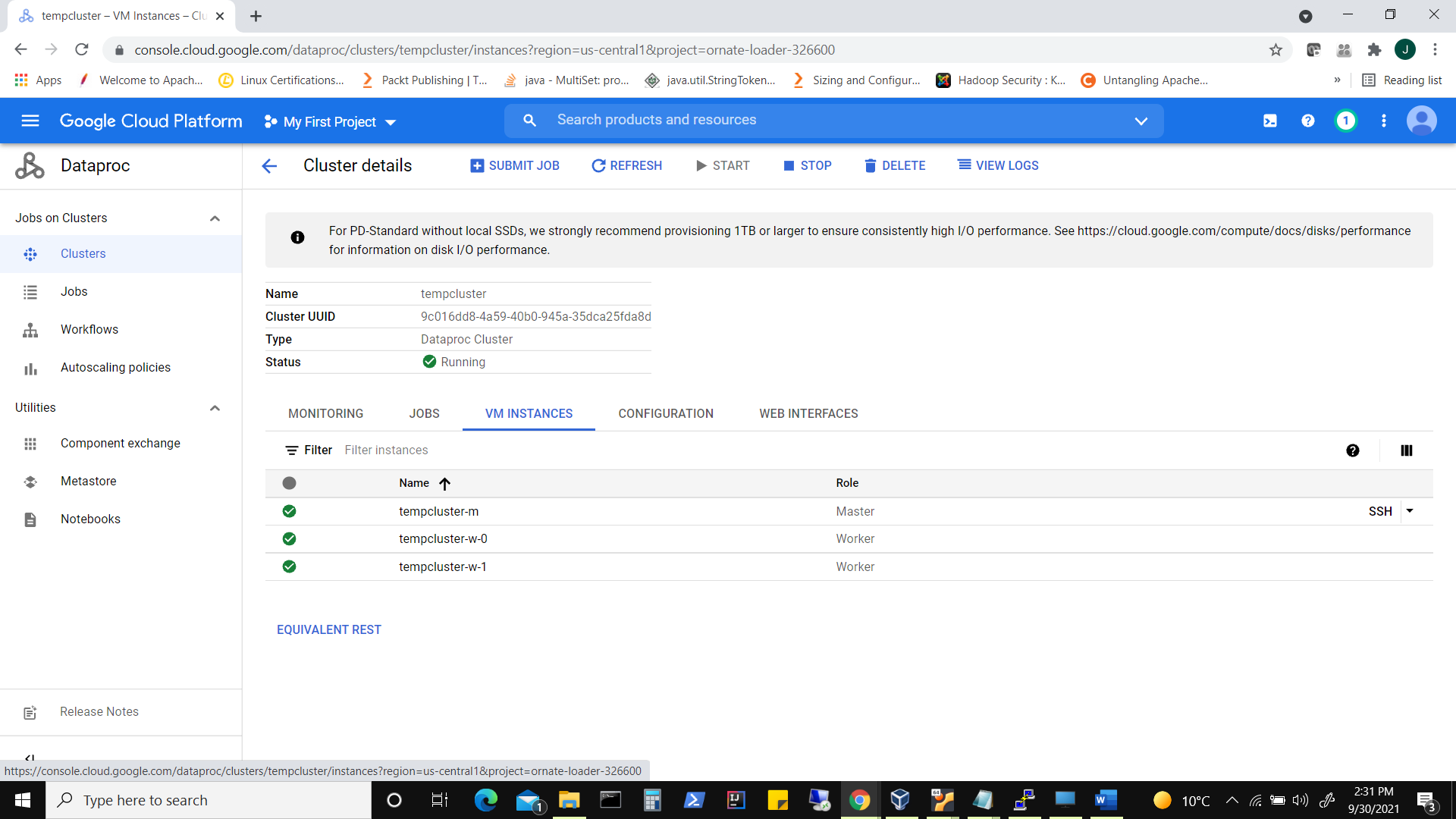


Accessing Cluster

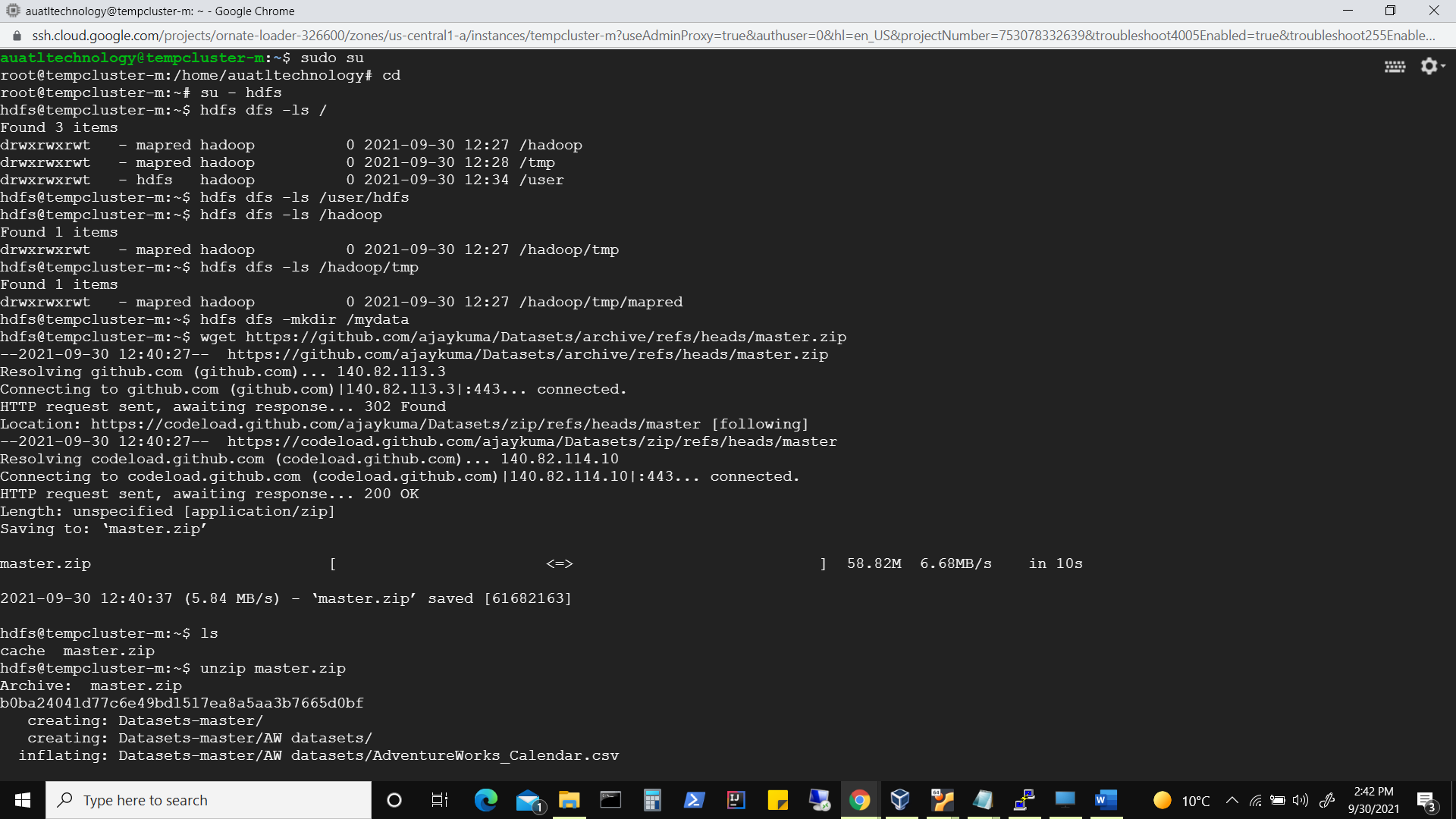
Click on cluster



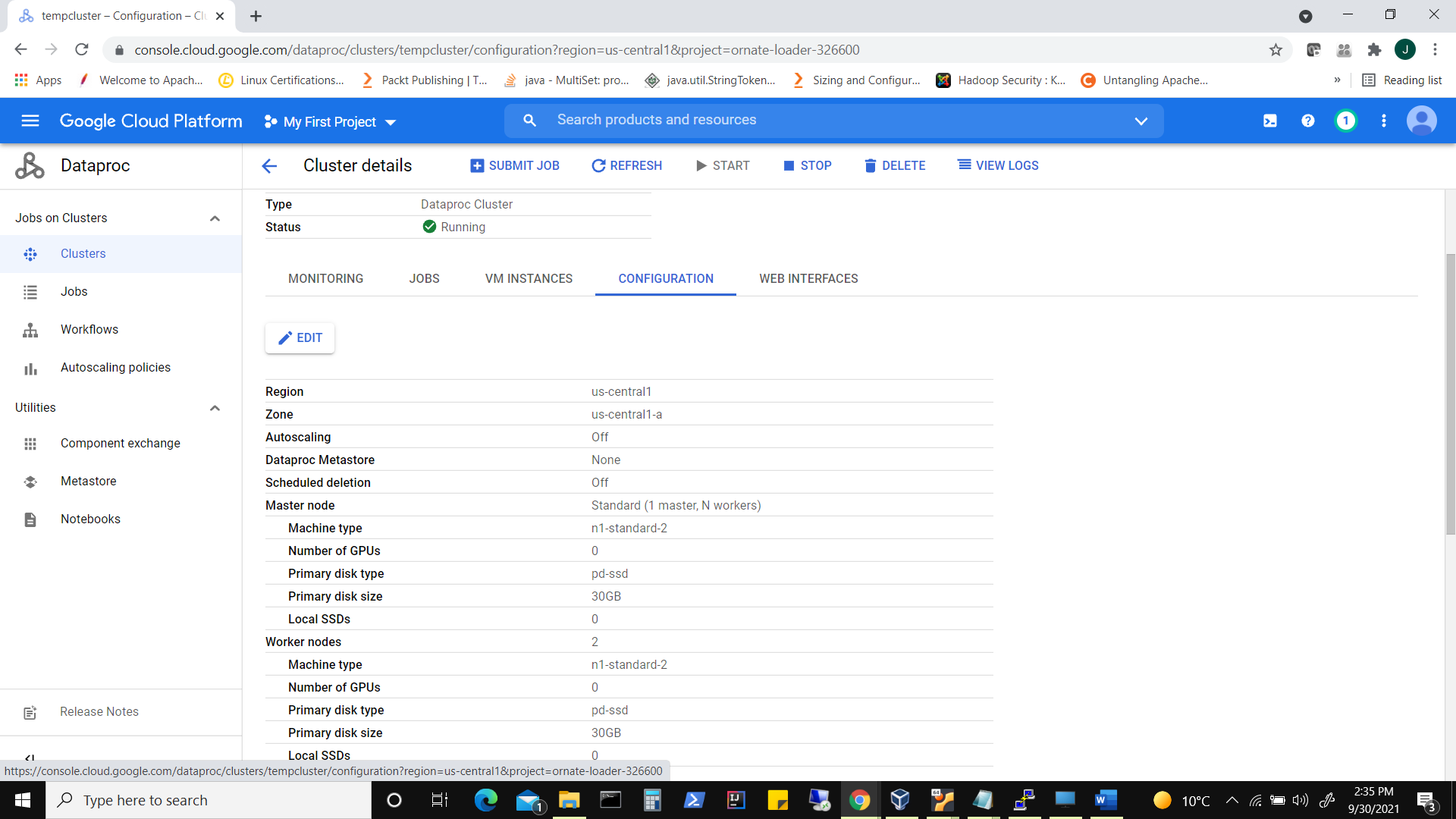


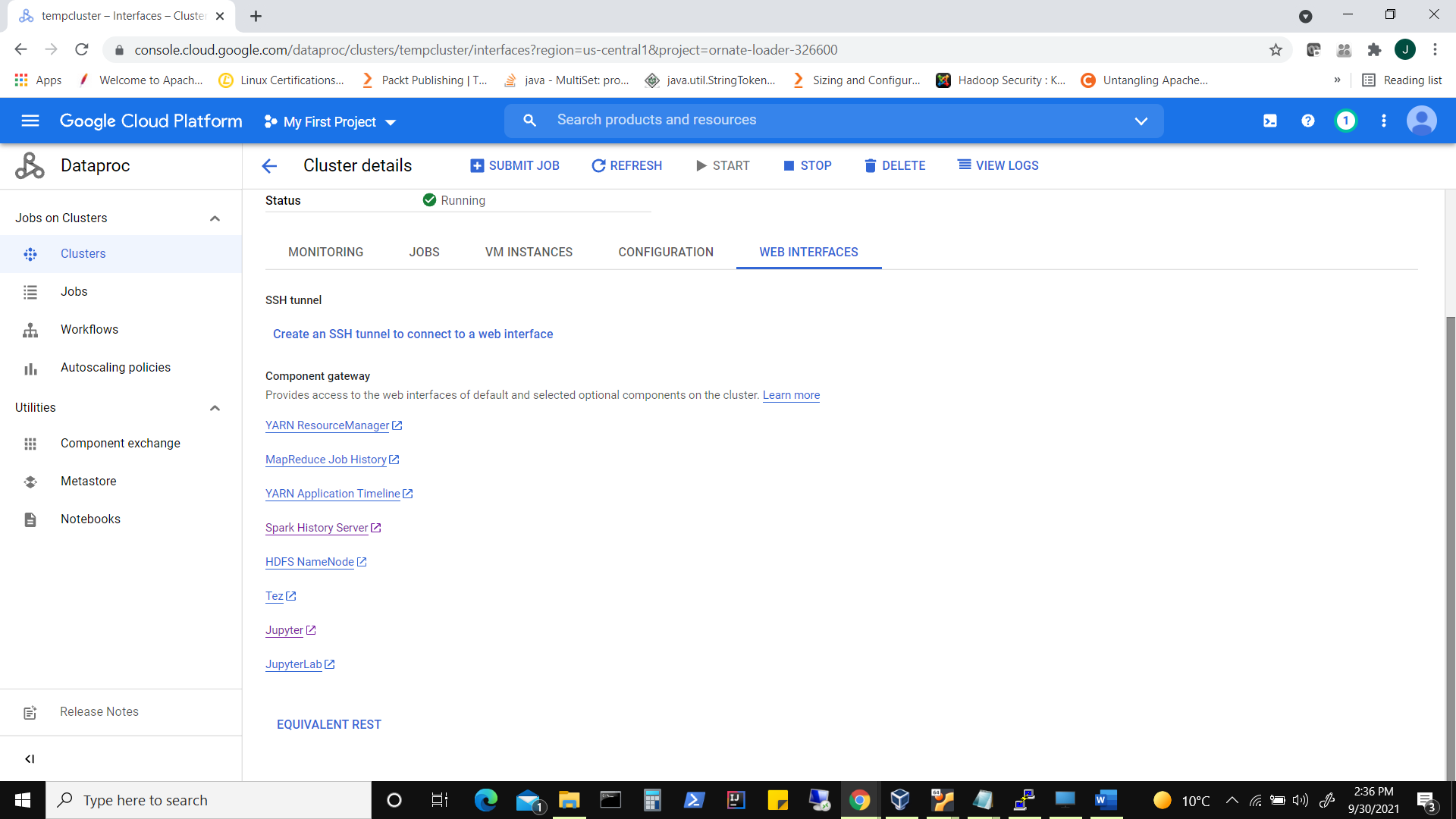


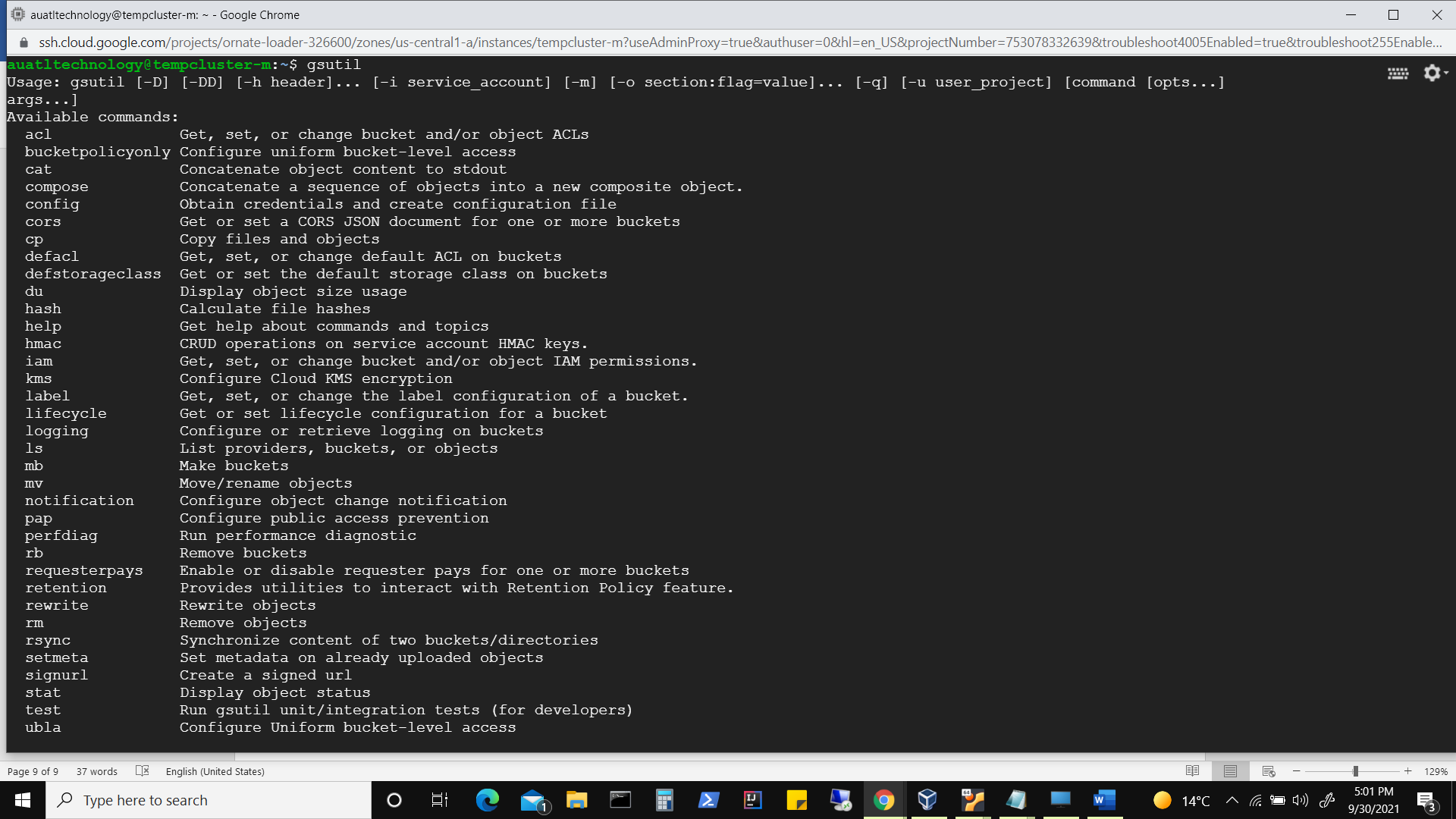
Click on say ‘tempcluster-m’ master node and SSH into it to connect to terminal



Similarly we can access ‘spark-shell’ or ‘pyspark’







$ gsutil ls

To create bucket from terminal

$ gsutil mb gs://data-bucket-frmt-aj

gsutil mb -p *PROJECT\_ID* -c *STORAGE\_CLASS* -l *BUCKET\_LOCATION* -b on gs://*BUCKET\_NAME*

To upload data into bucket

$ gsutil cp /var/lib/hadoop-hdfs/Datasets-master/abc1.txt gs://data-bucket-frmt-aj

To look for content in your bucket

$ gsutil ls gs://data-bucket-frmt-aj

To download content from bucket

gsutil cp gs://data-bucket-frmt-aj/abc1.txt .

(option -m along with copy to let copying happen in parallel)