Apache BEAM molecules experiment: technical objectives

1. Understand the differences between Google Dataflow and Dataproc
2. Understand how a pipeline is created
3. Understand the learning task
4. Understand the contents of each script in the pipeline
5. Understand how to use transformations and estimators in Tensorflow
6. Run the pipeline as is locally (`run-local`) and in the cloud (`run-cloud`) (are there any differences in performance?)
7. Vary the `max-data-files` parameter with values 10, 100, 1000
8. Modify this program to include the actual ENERGY of each molecule in the predictions file
9. Modify this program to allow for cross-validation