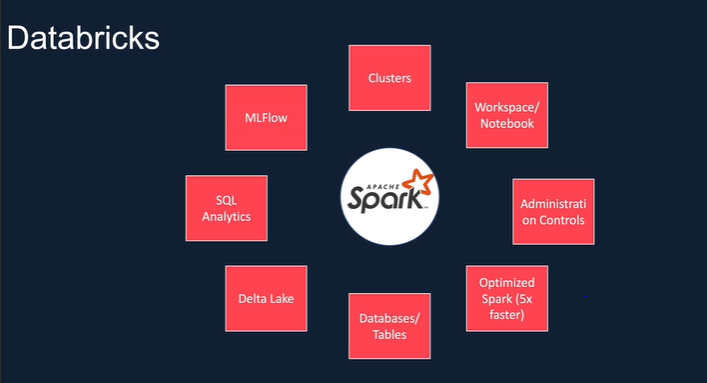
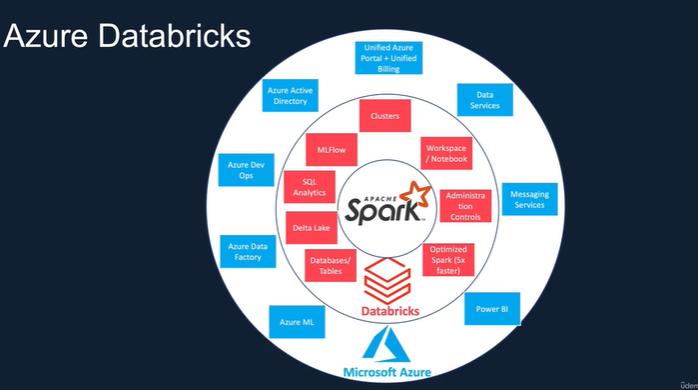
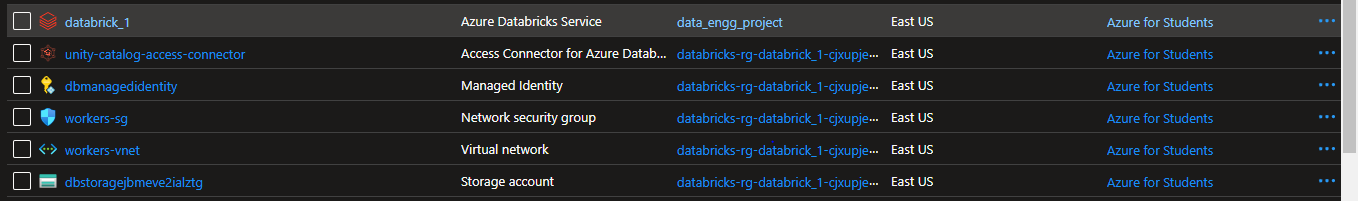
Apache spark provides clusters,workspace/Notebook ,Adminstration control,optmized spark,database/tables,delta lake,sql analysis,MLFlow.



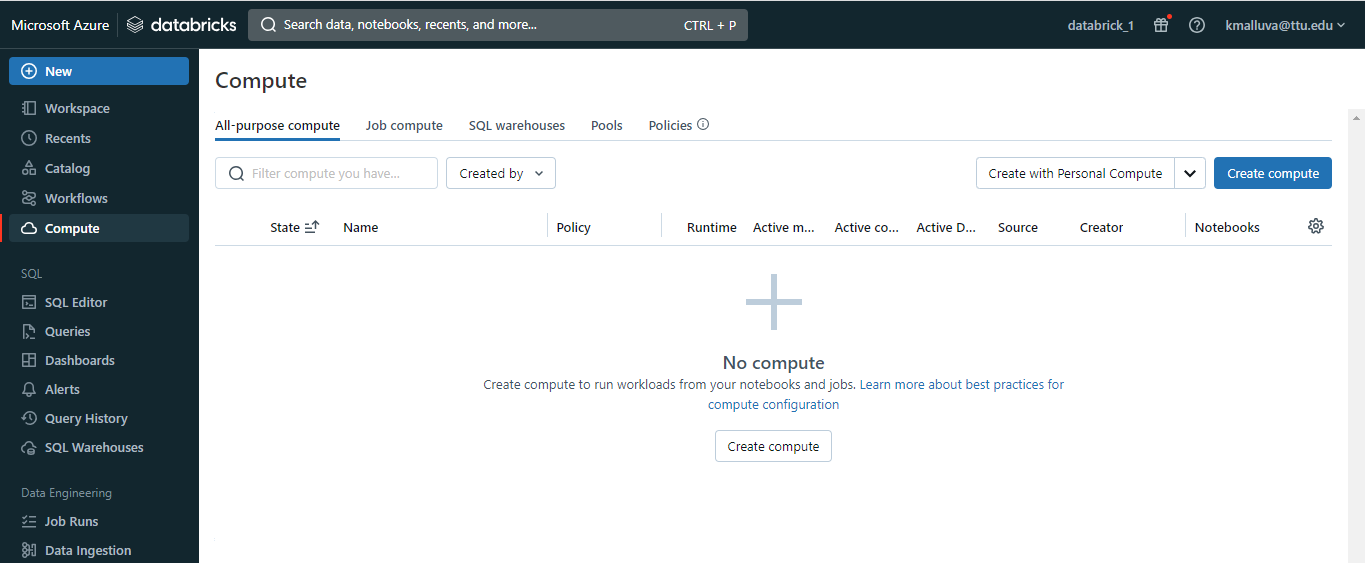


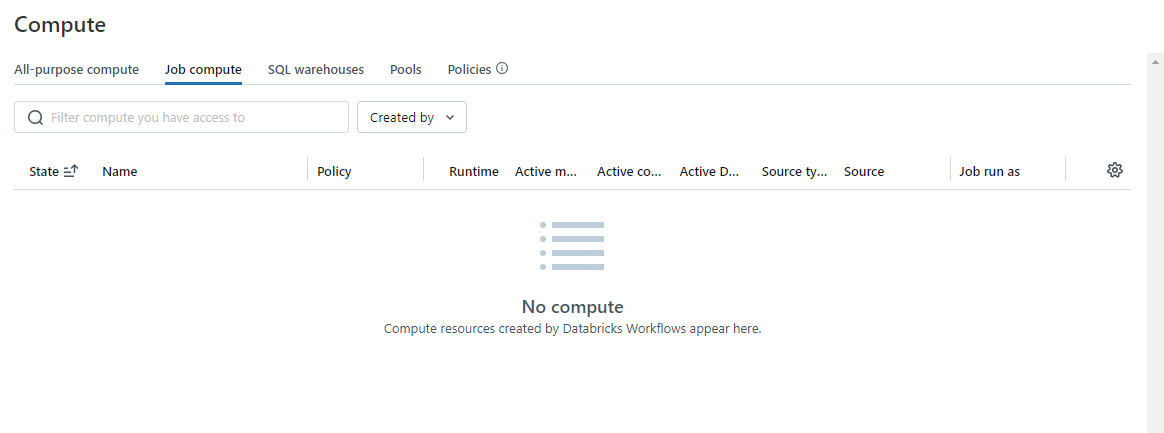


There are two types of cluster

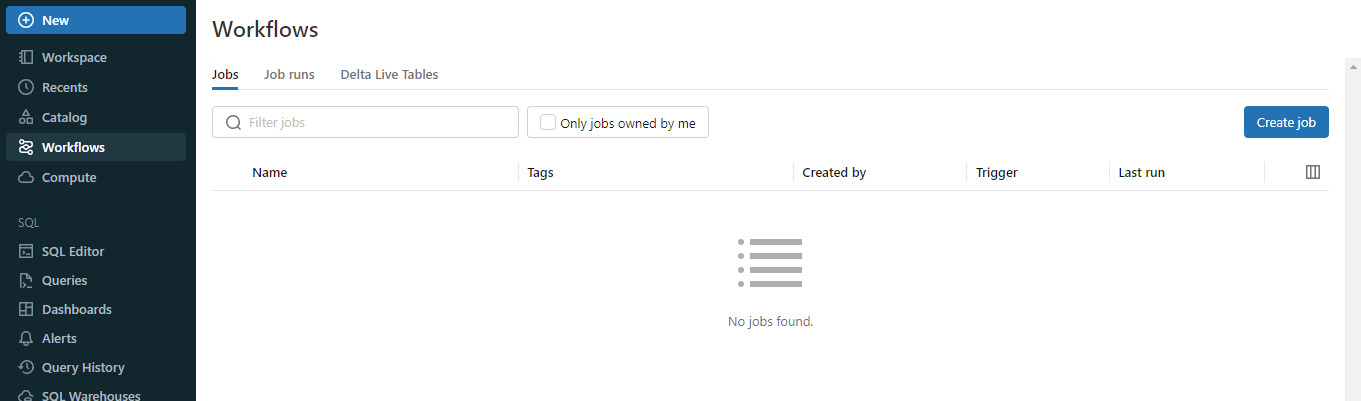
Where all purpose cluster can be created but job cluster cannot be created.

When there a workflow we can create a job cluster ,after end of the job it will get destroied .

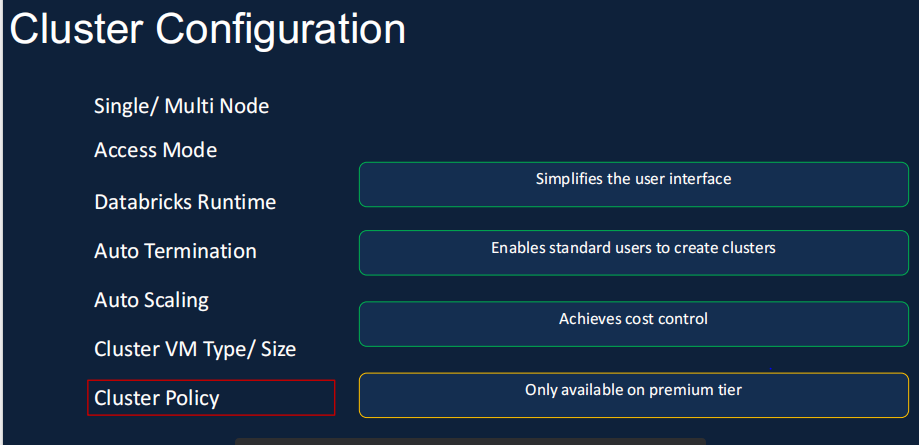




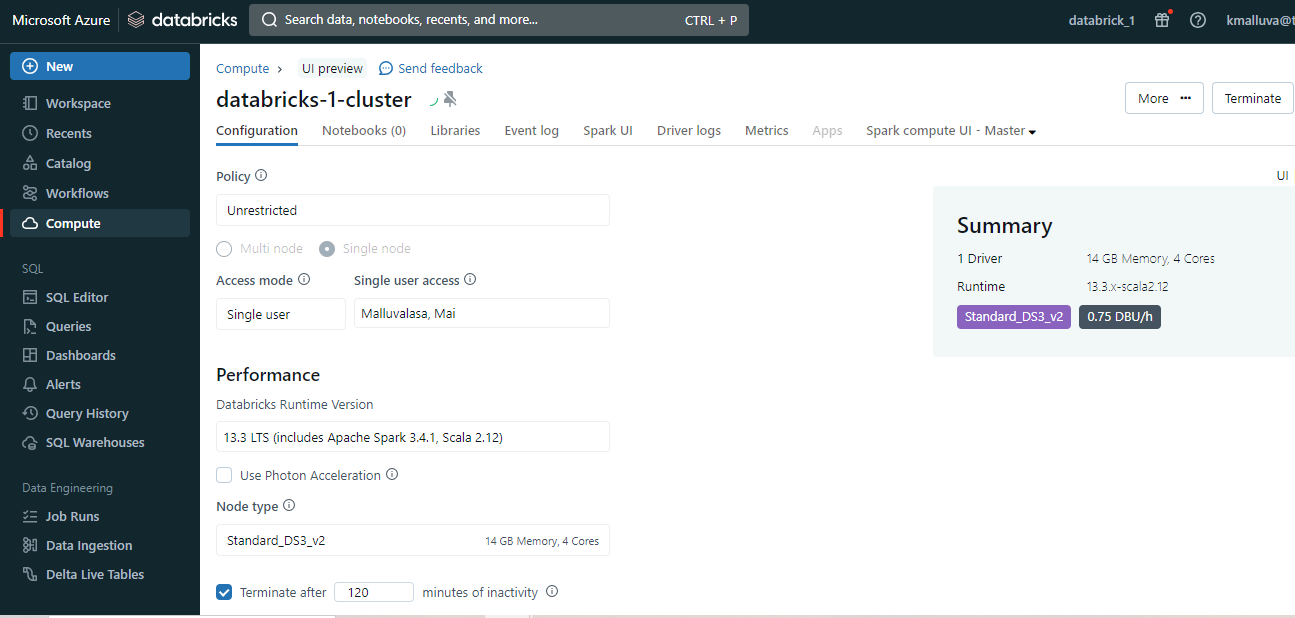
In workflow you can create a job:

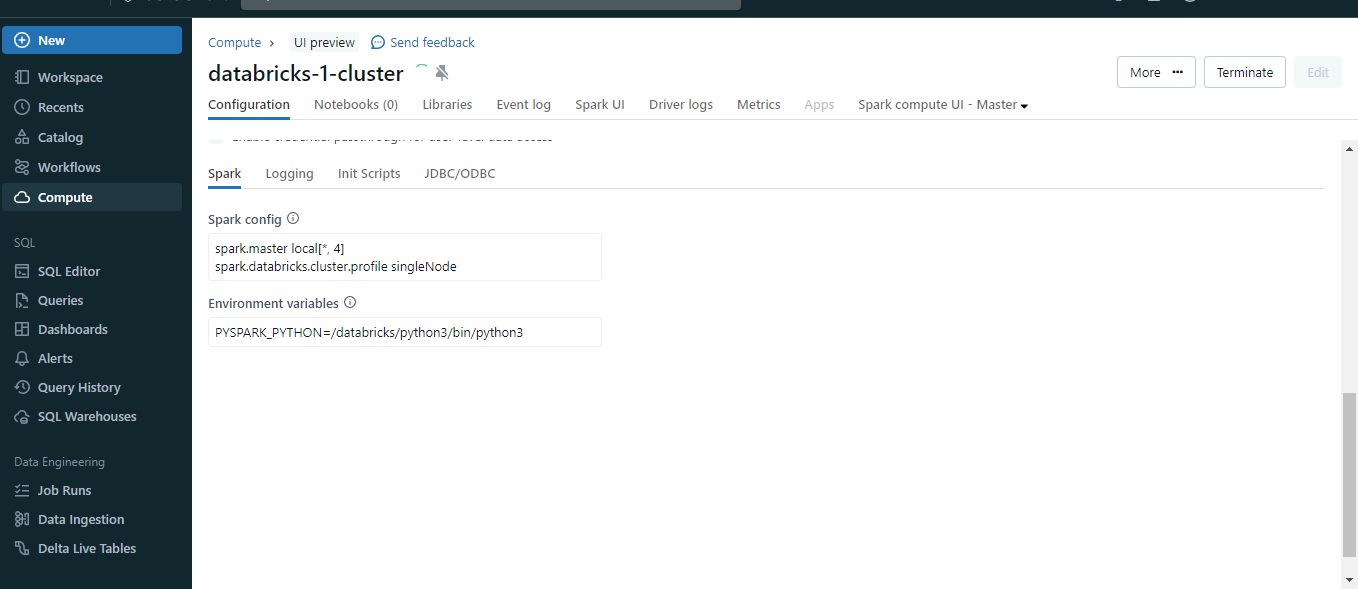


Cluster configuration:

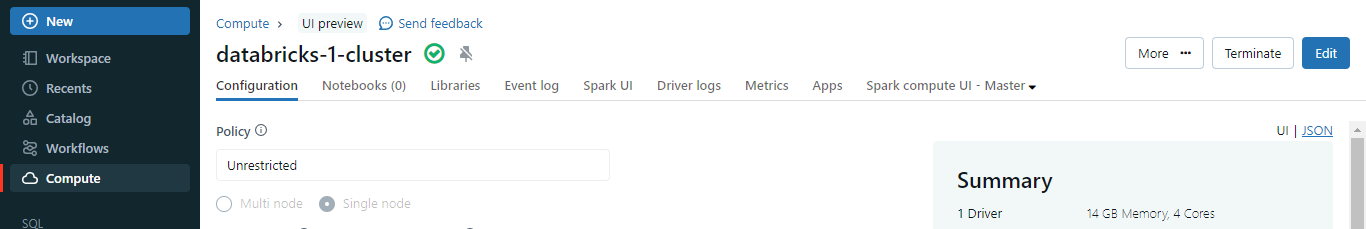


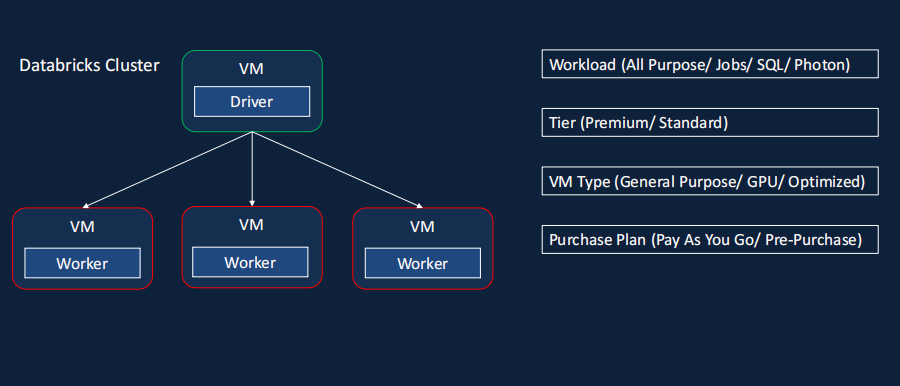
Creating a cluster:

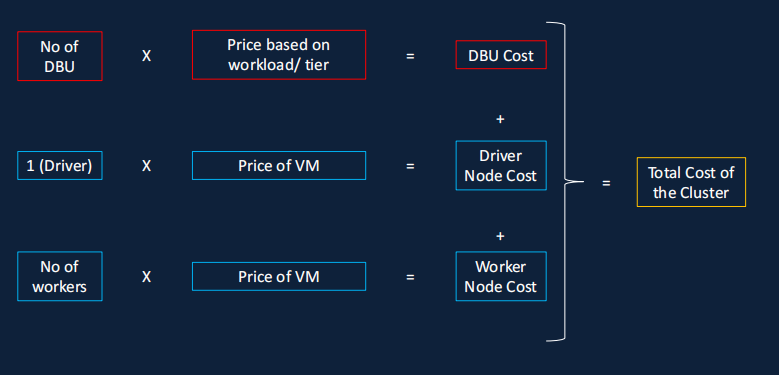




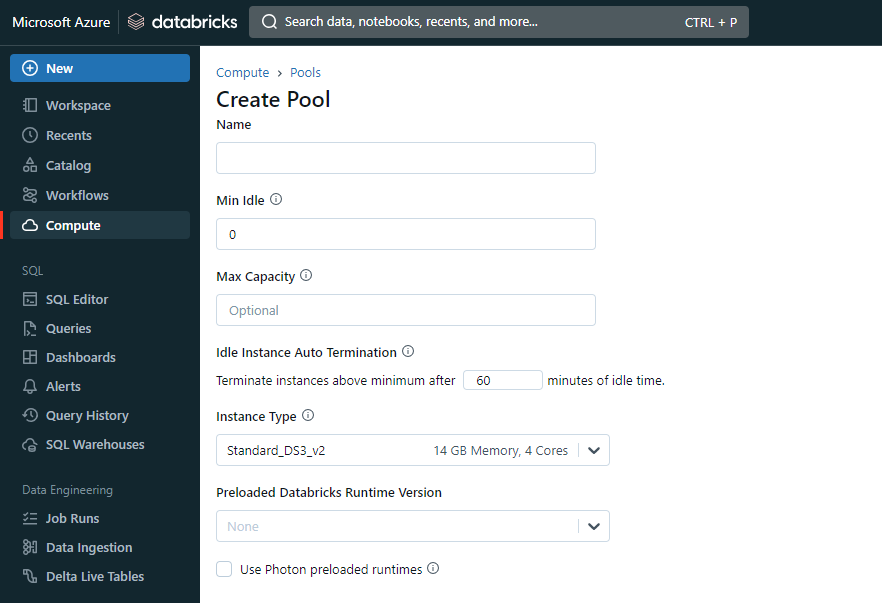
If cluster is created we can see green trick mark

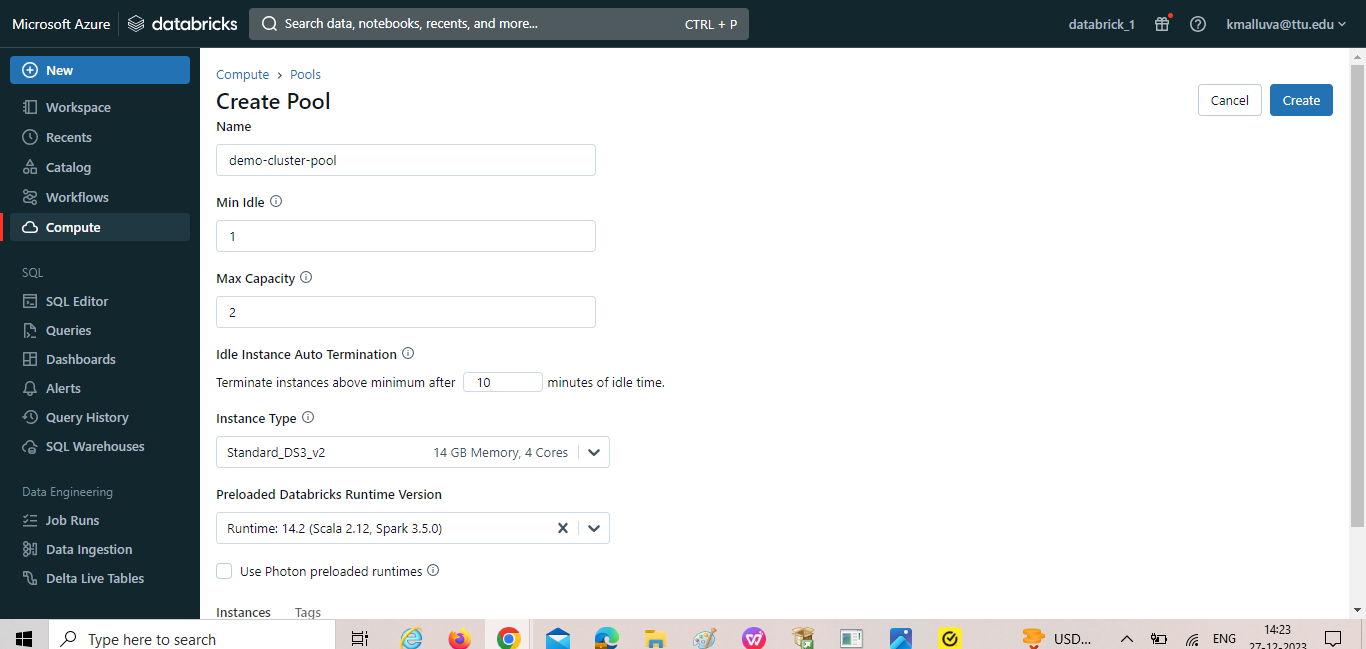


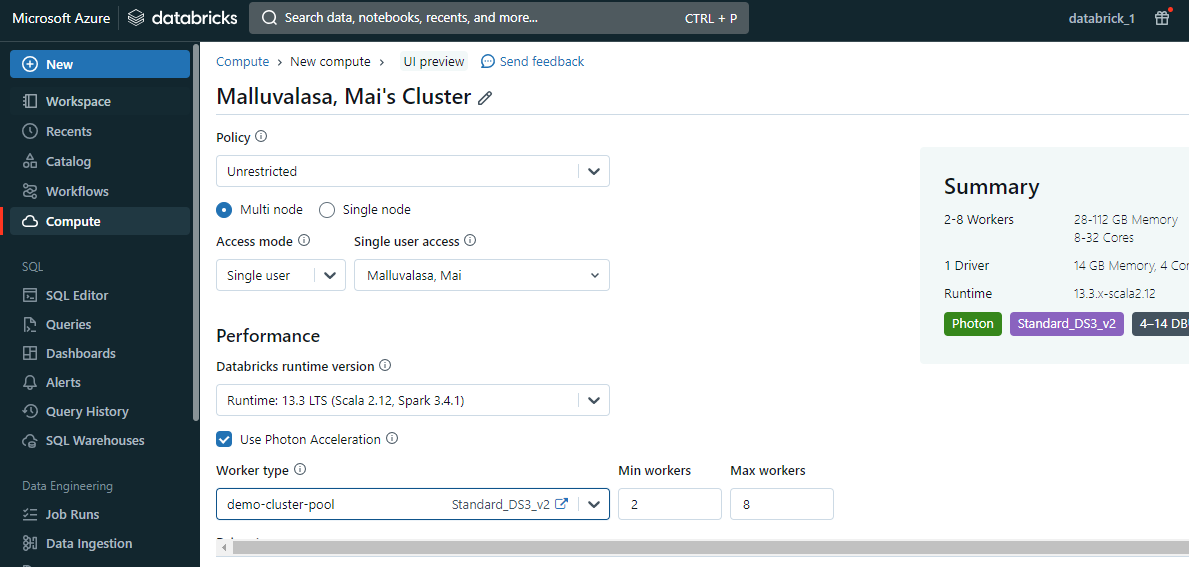




For creating a pool:

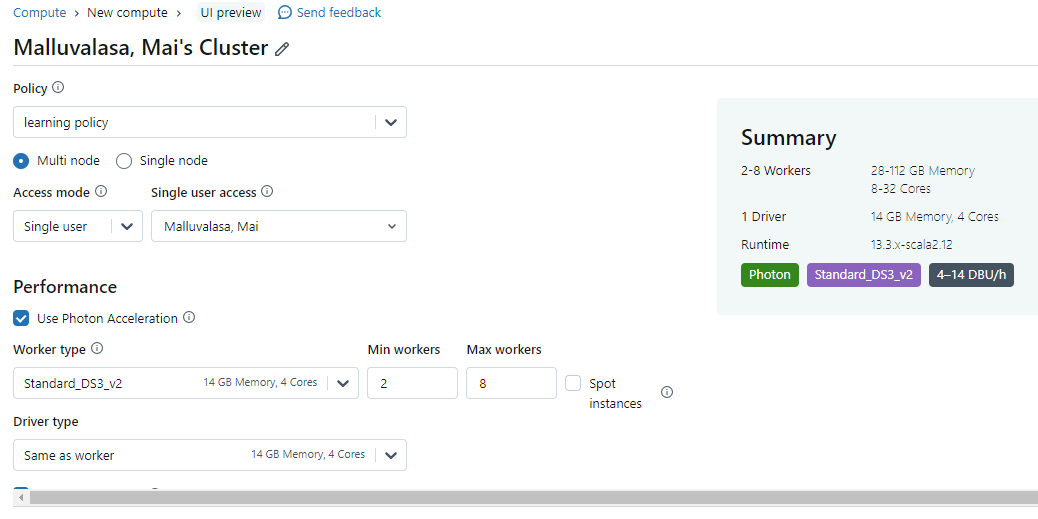
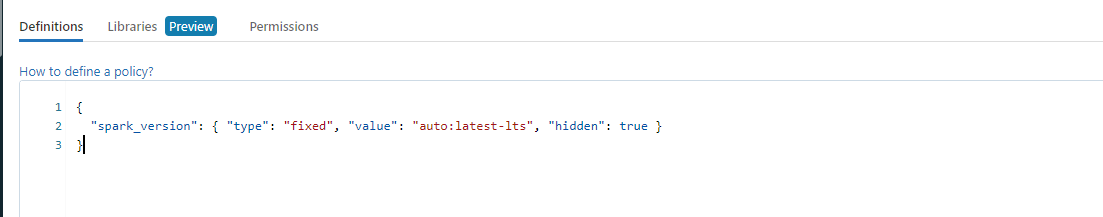
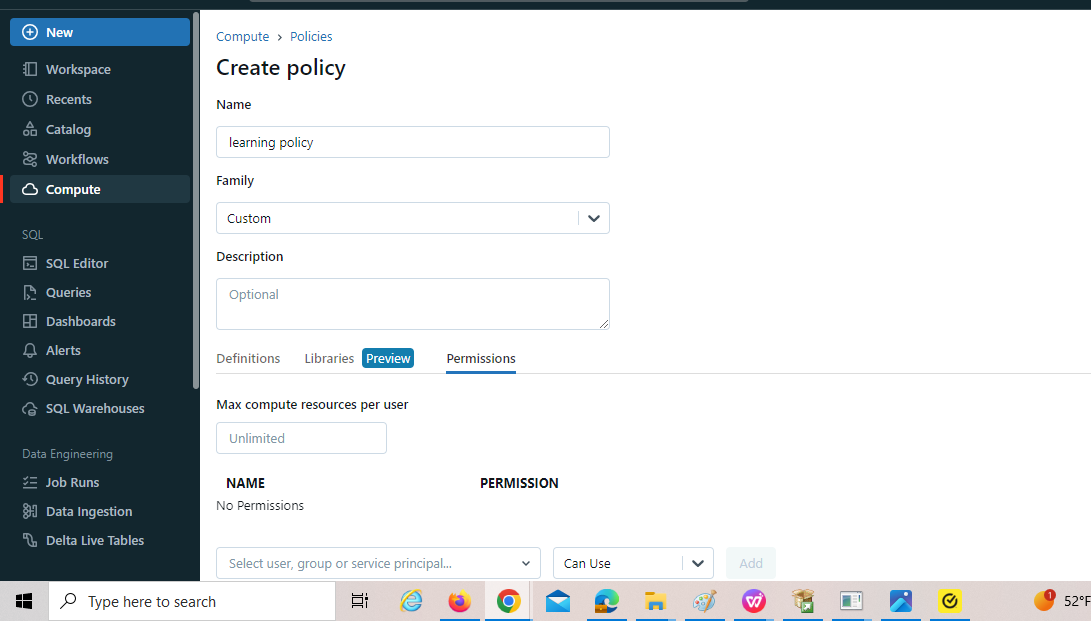
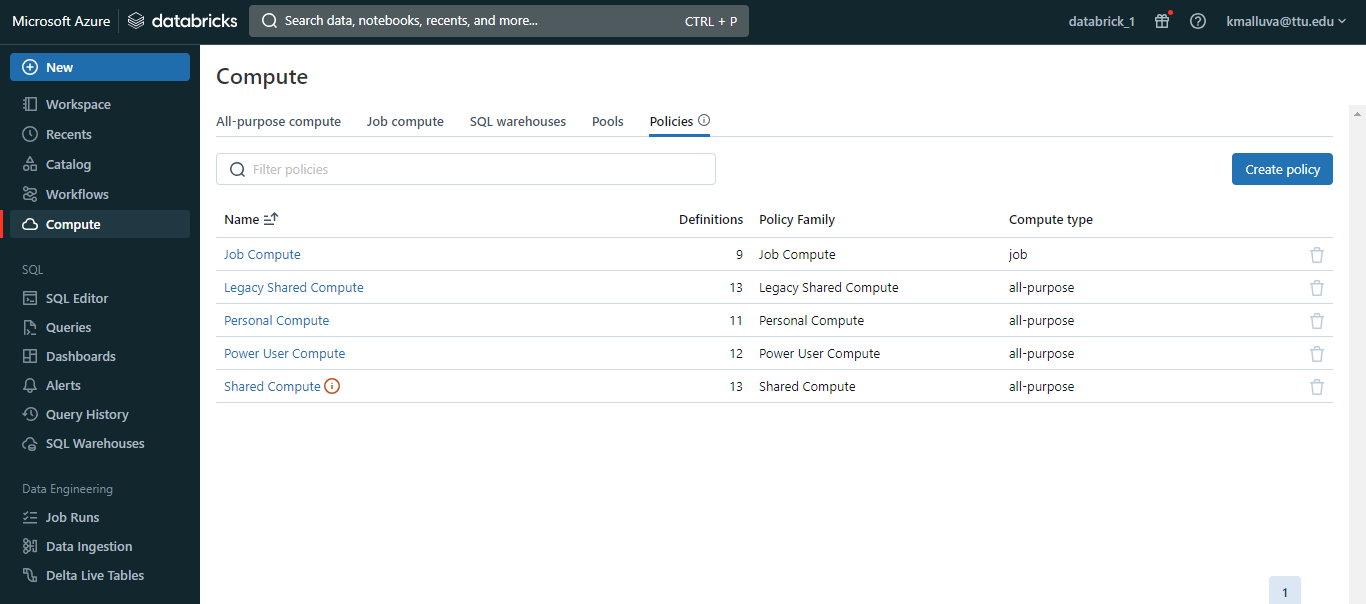




Creating new cluster by adding worker type as demo-cluster 

Where the pool speed up the creation of cluster.

Policy:

{

  "spark\_conf.spark.databricks.cluster.profile": {

    "type": "fixed",

    "value": "singleNode",

    "hidden": true

  },

  "node\_type\_id": {

    "type": "allowlist",

    "values": [

      "Standard\_DS3\_v2",

      "Standard\_D3\_v2",

      "Standard\_DS12\_v2",

      "Standard\_D12\_v2"

    ],

    "defaultValue": "Standard\_DS3\_v2"

  },

  "autotermination\_minutes": {

    "type": "fixed",

    "value": 20

  },

  "spark\_version": {

    "type": "fixed",

    "value": "auto:latest-lts",

    "hidden": true

  },

  "cluster\_type": {

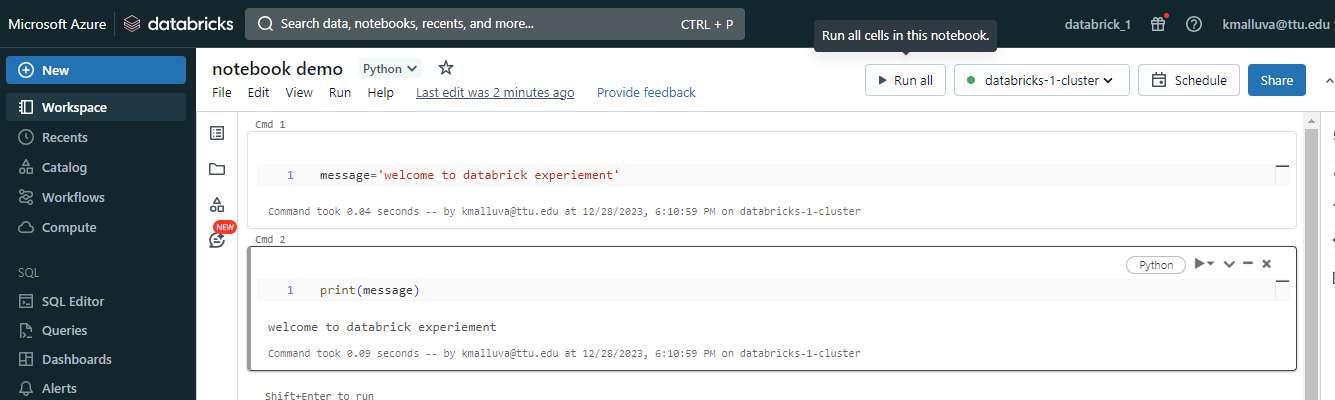
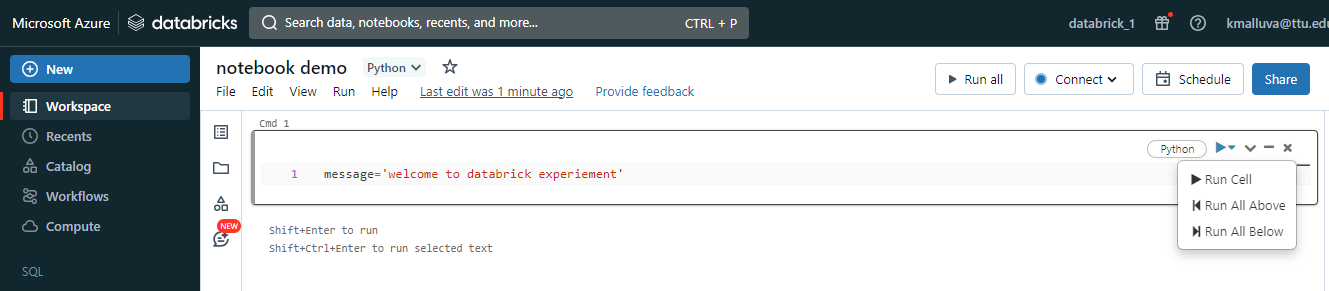
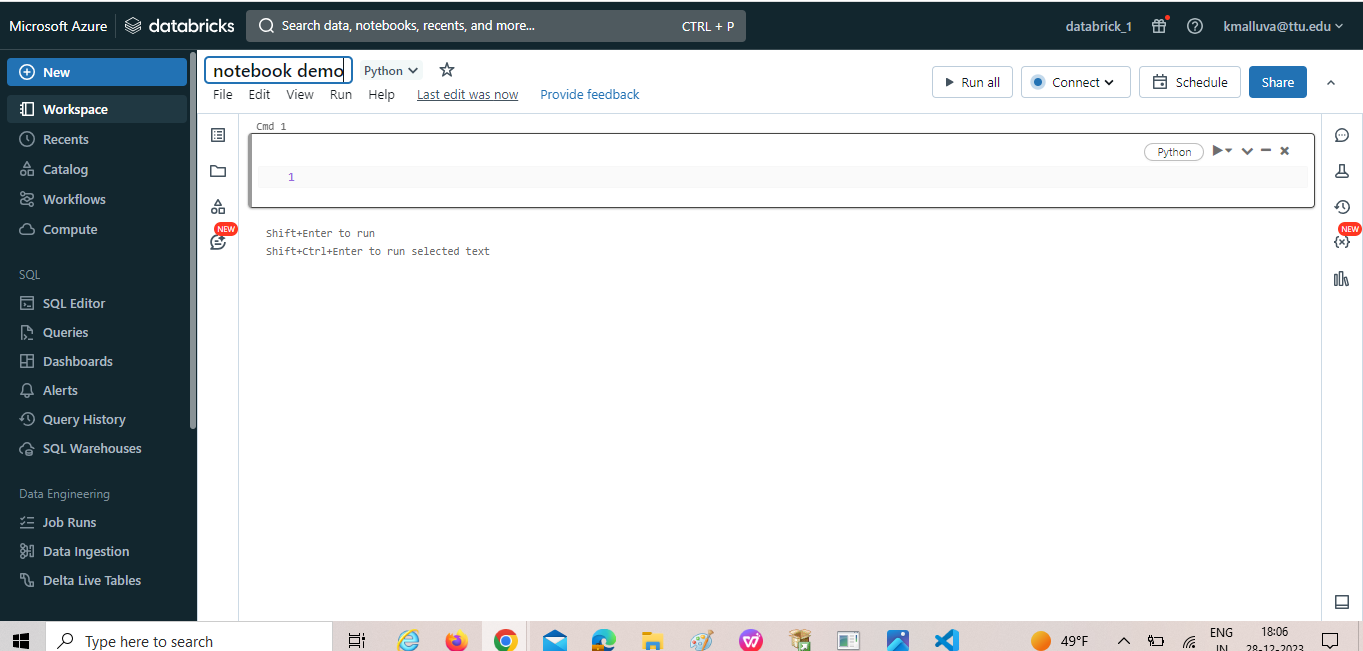
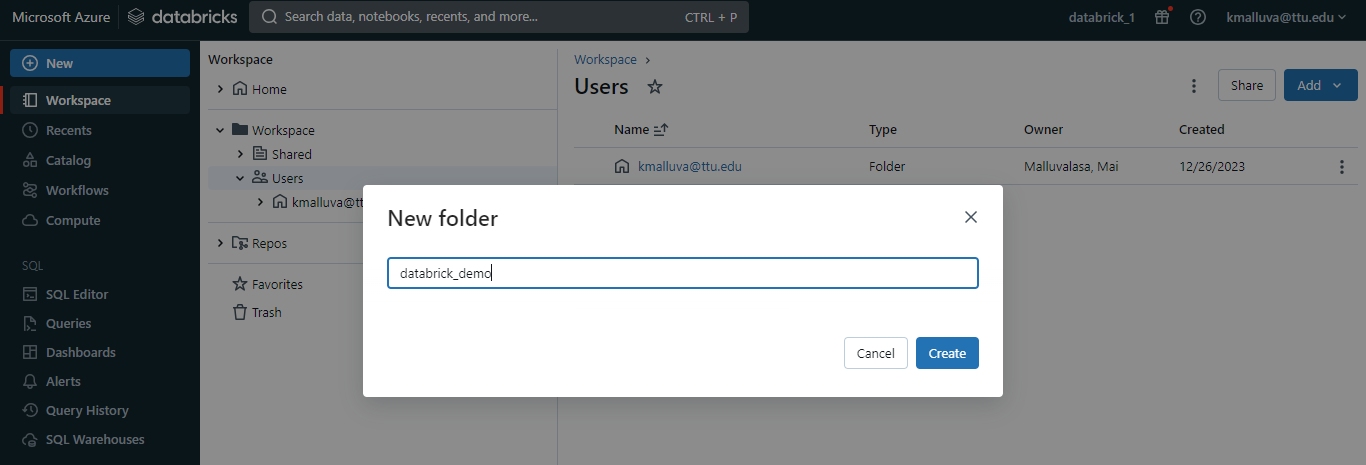
    "type": "fixed",

    "value": "all-purpose"

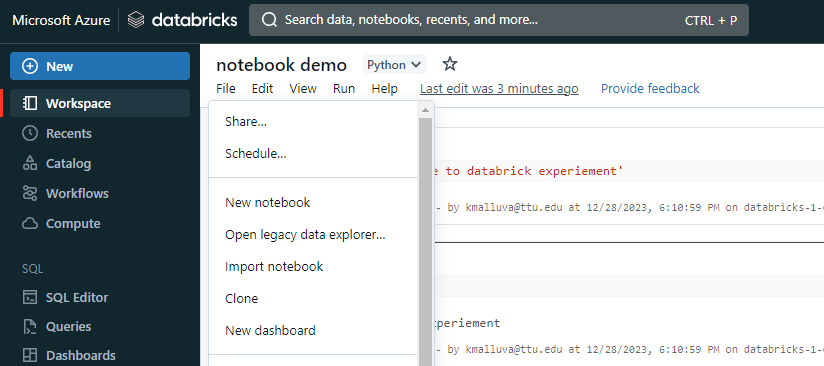
  }

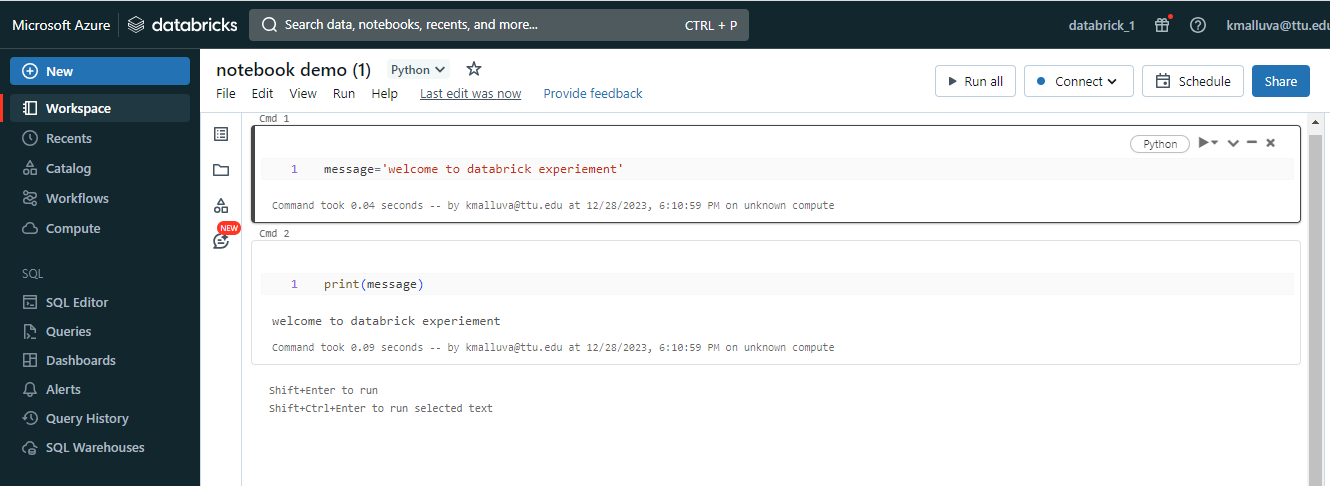
}

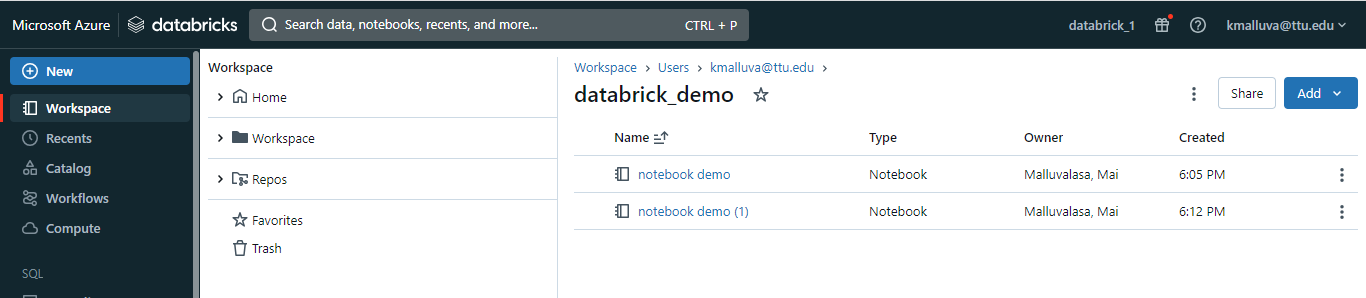
Create notebook:

to duplicate the notebook:

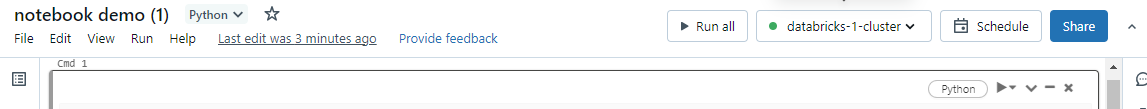
Click on clone

duplicate notebook:



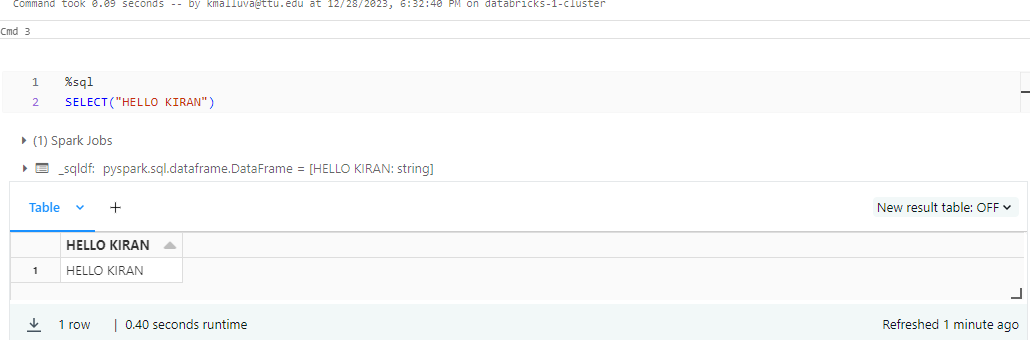


Connected to cluster:

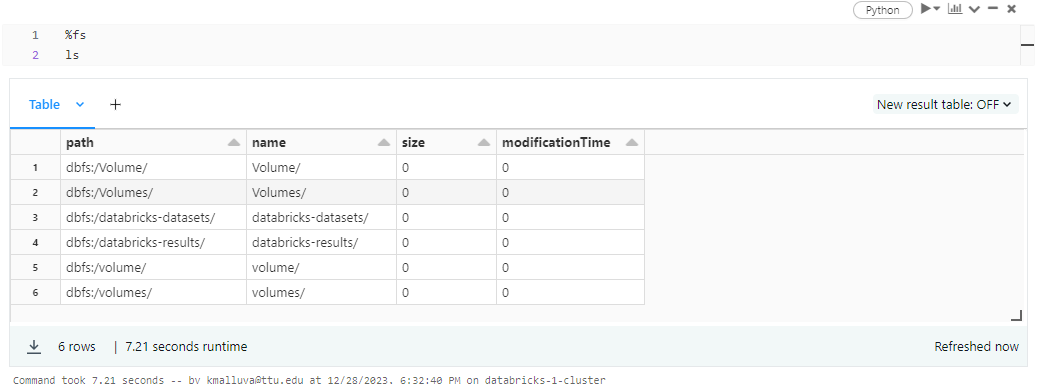


Magic commands:

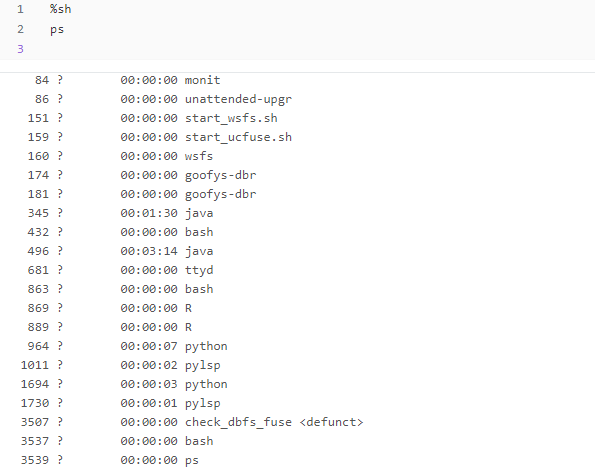
It is used for moving form one lang to another lang in the notebook (eg: we are using python instead of python you want to move to scala lang this command is used.



Magic command for listing out file system:

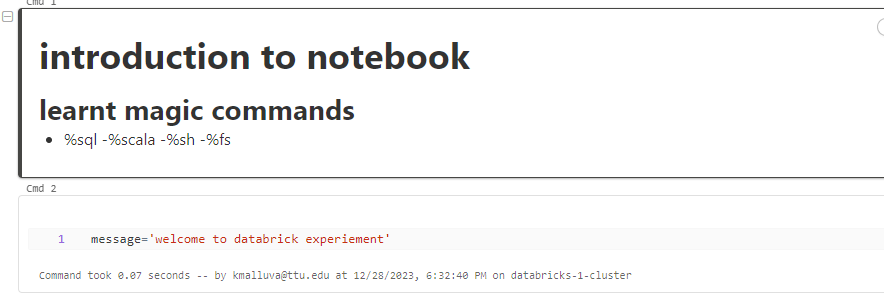


To now what processor are running:

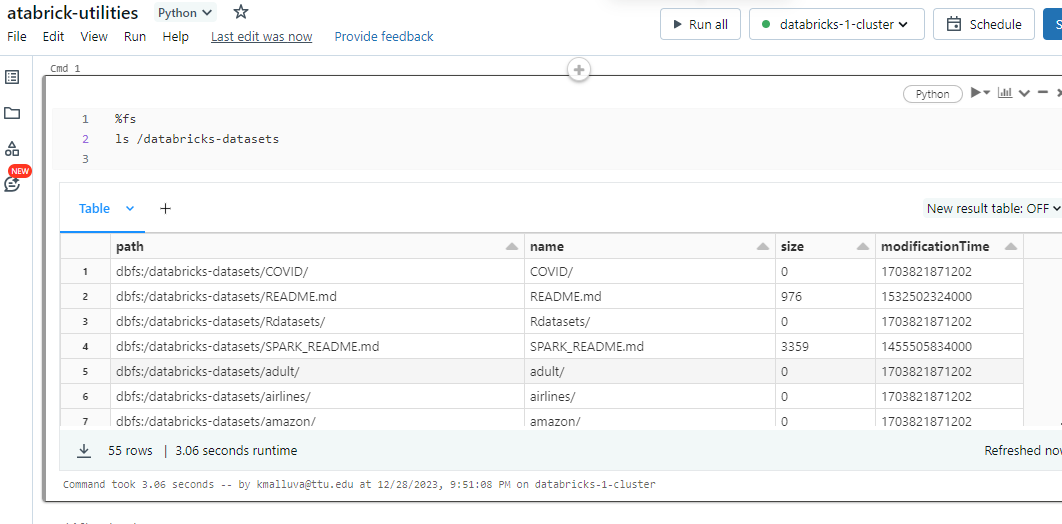


%md

Cell is modified as document



Utilities:



In python:

Dbutils.fs.ls(‘/ ’)

