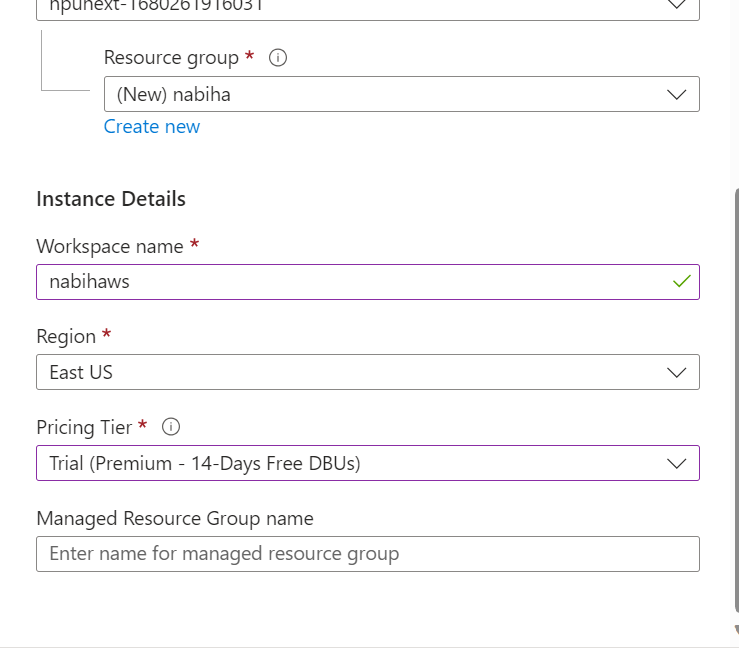
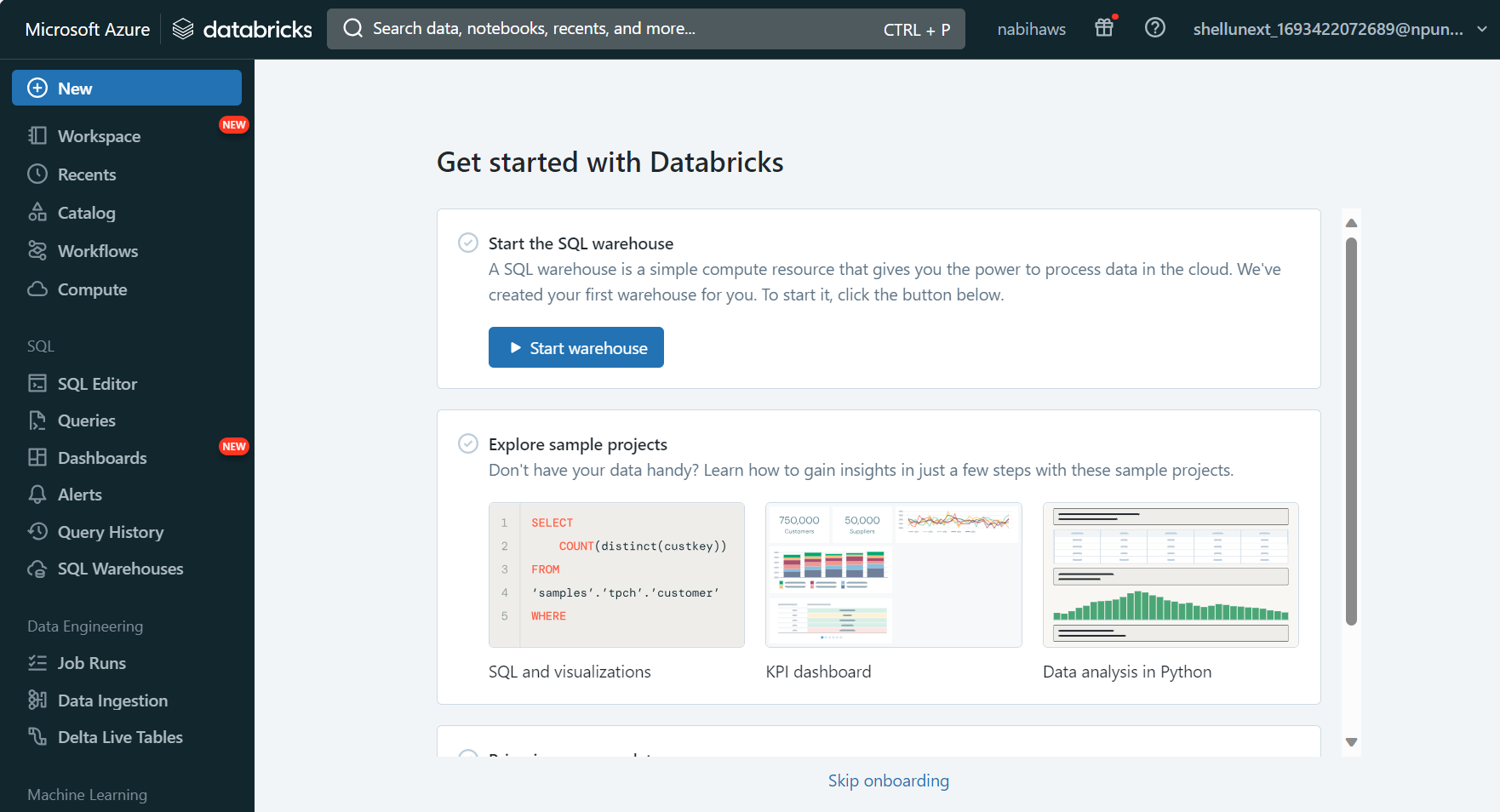
Azure DataBricks

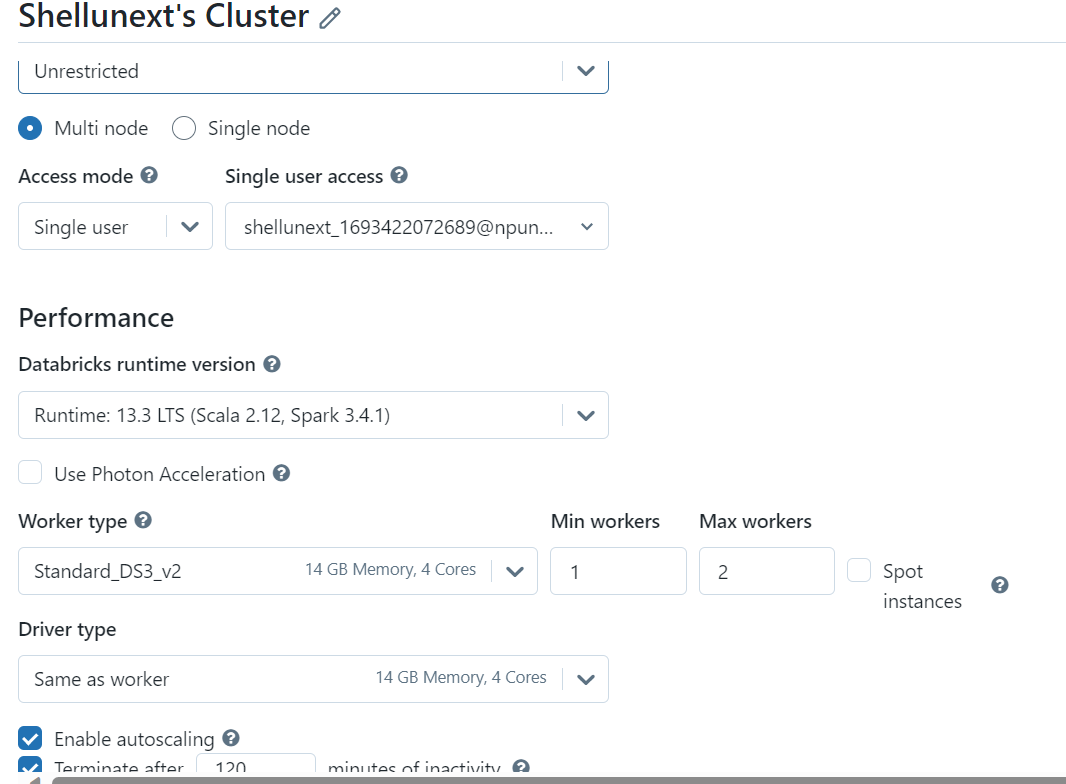
1. Create the databricks resource



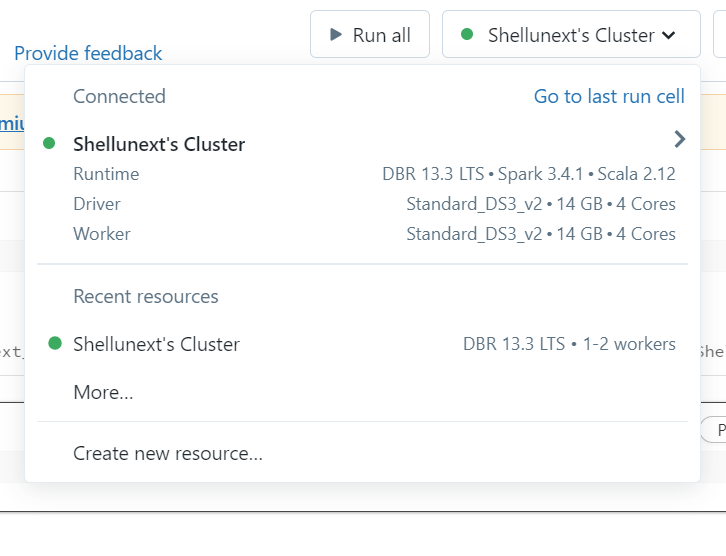
1. Launch the databricks workspace



1. Create Compute



1. Connect workspace to cluster



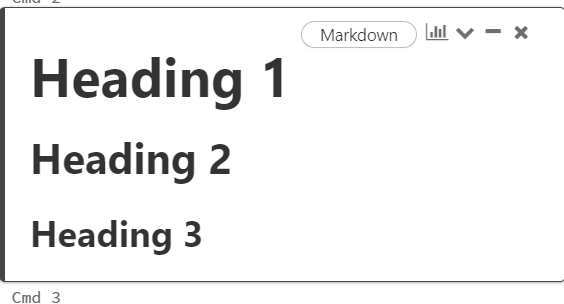
1. Test



1. Magic Command -> change the language



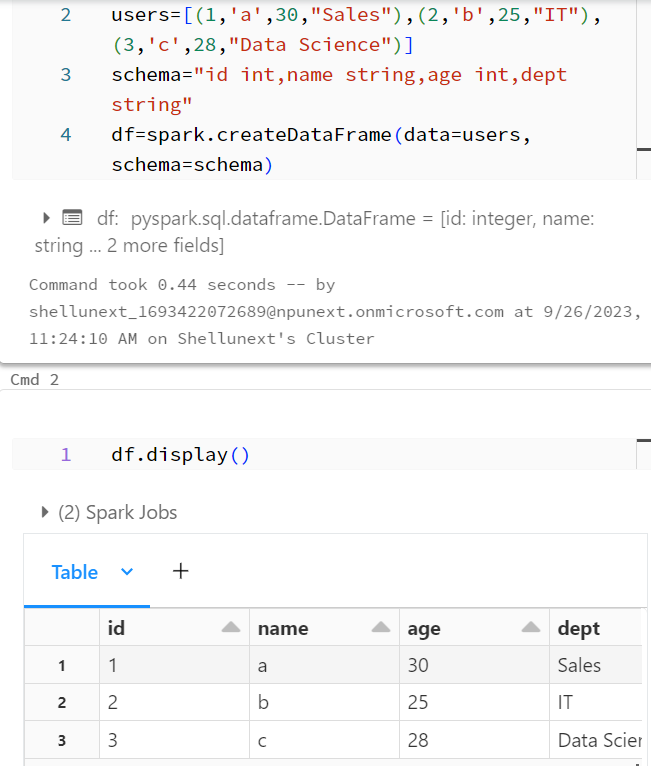
1. Markdown -> md
2. %md
3. #Heading 1
4. ##Heading 2
5. ###Heading 3



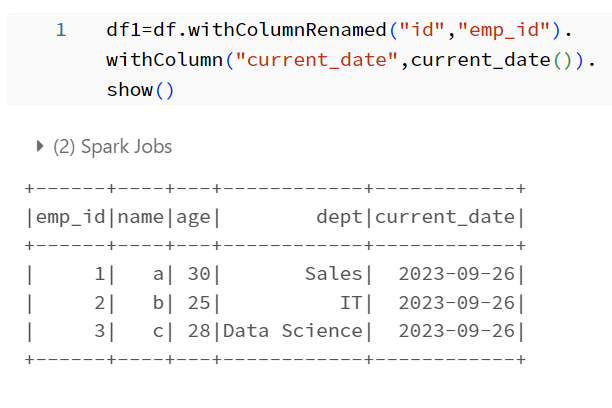
8. Creating tables



9. Creating tables with dataframe

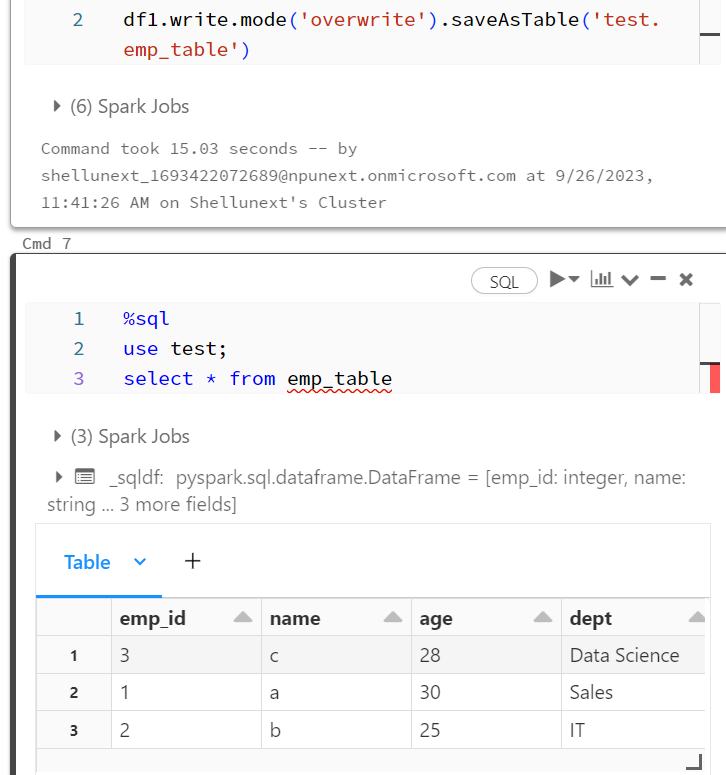


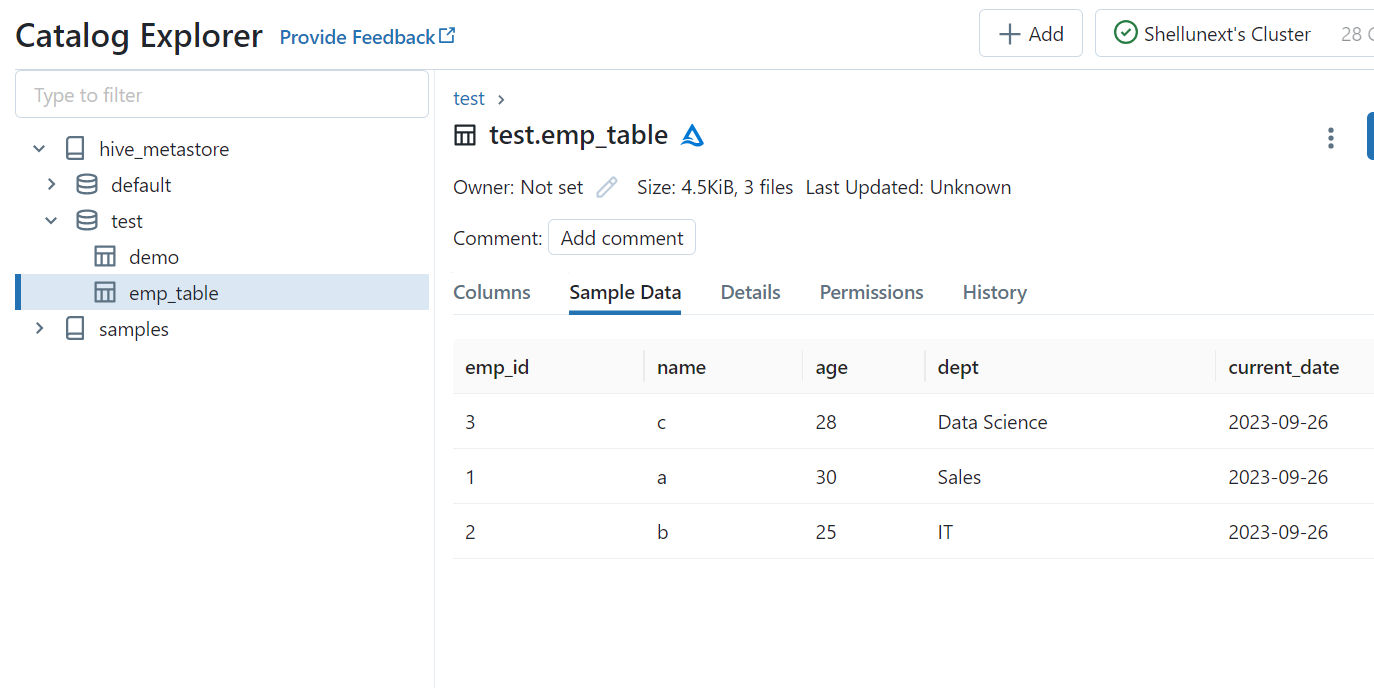
10. New Column



1. Save df1 to new table

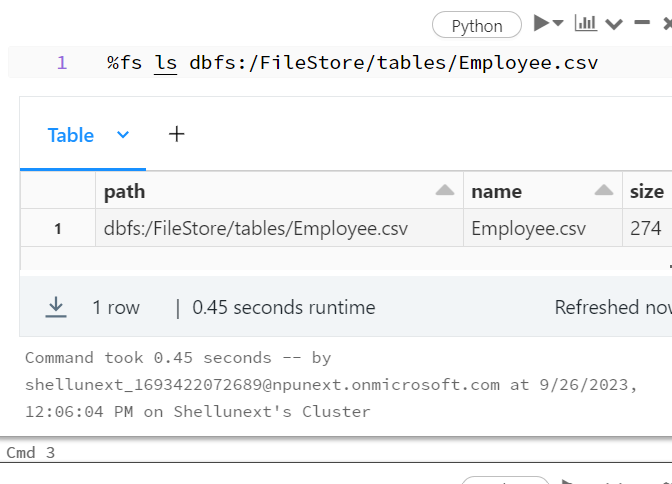




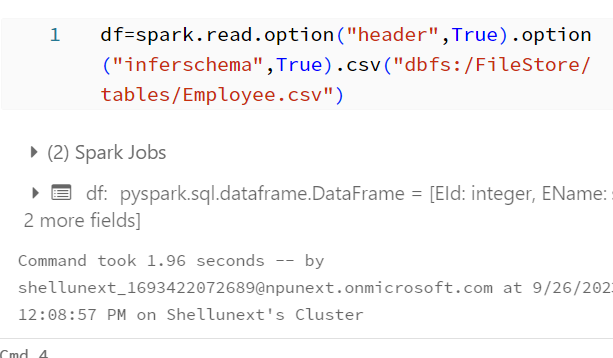


1. Adding files and writing csv to table

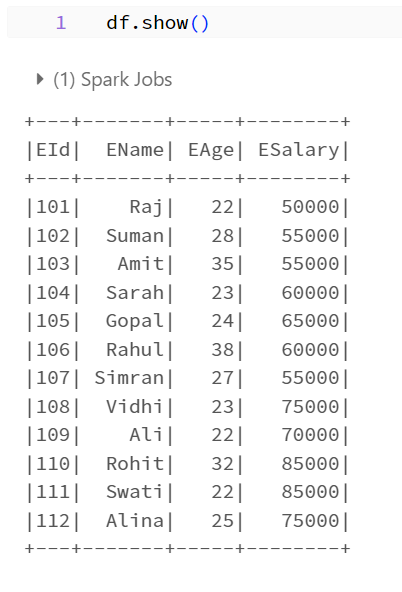
* Go to Catalog Explorer-- Add Data-- DBFS and then upload any csv file



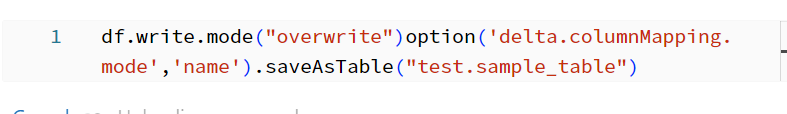
* Write to table



* Display

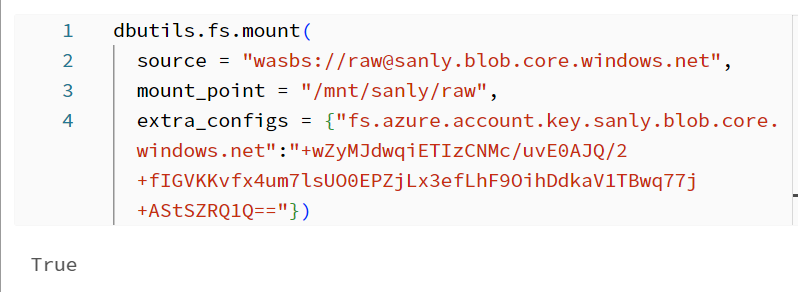


1. Saving to table



1. Mounting data from cloud to databricks

* Access Key

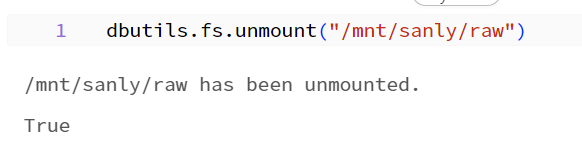






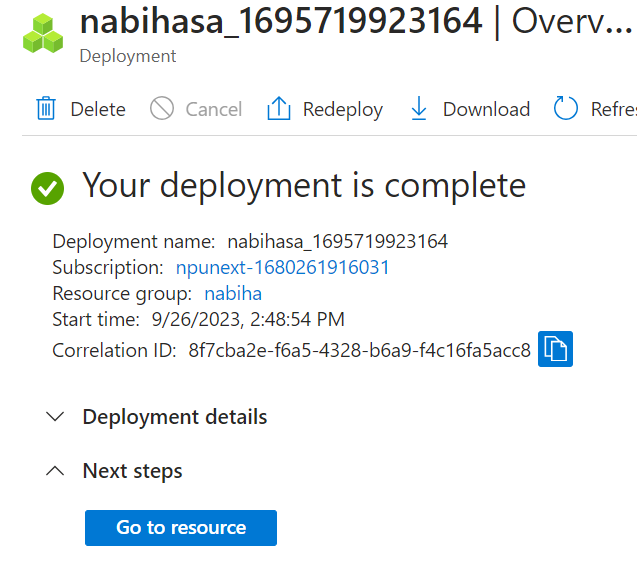
* SAS

Unmount first



* Service Principal

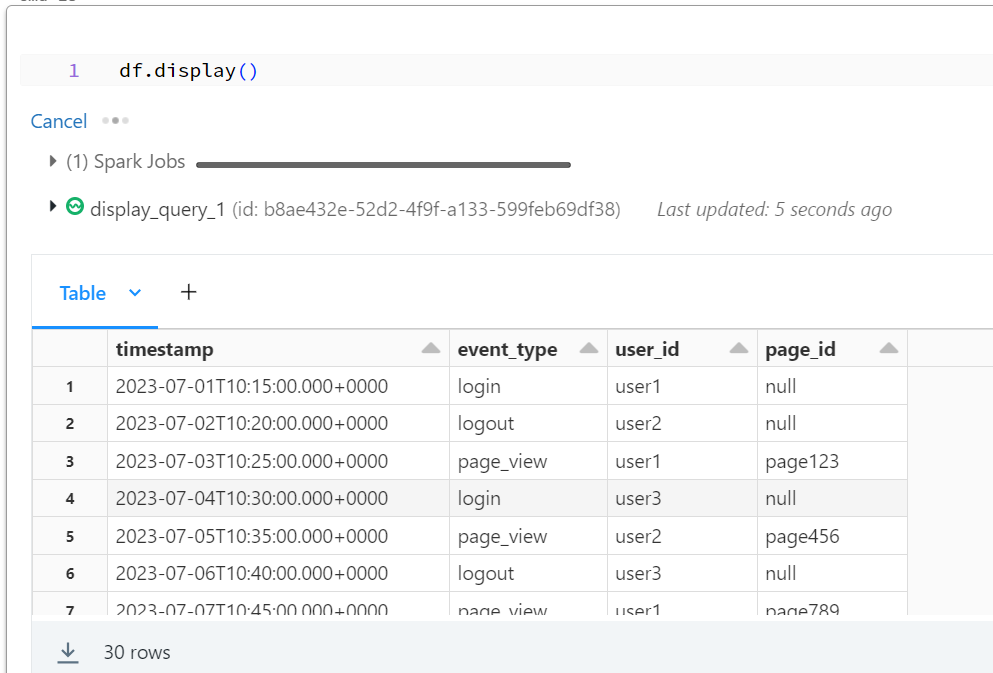
Create storage account

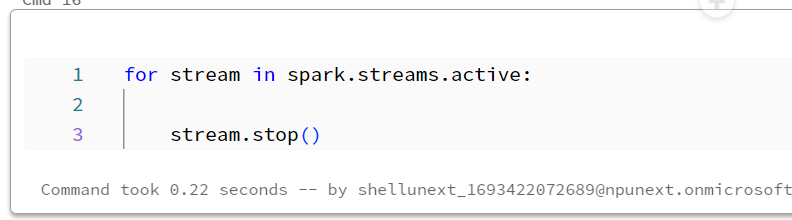


Create container inputfiles > app registrations

Storage acc>access control>role>blob contributor>app>databricks



****

****

DAY – 2

1. Create databricks >cluster & storage acc > container > directory >json upload
2. Mount

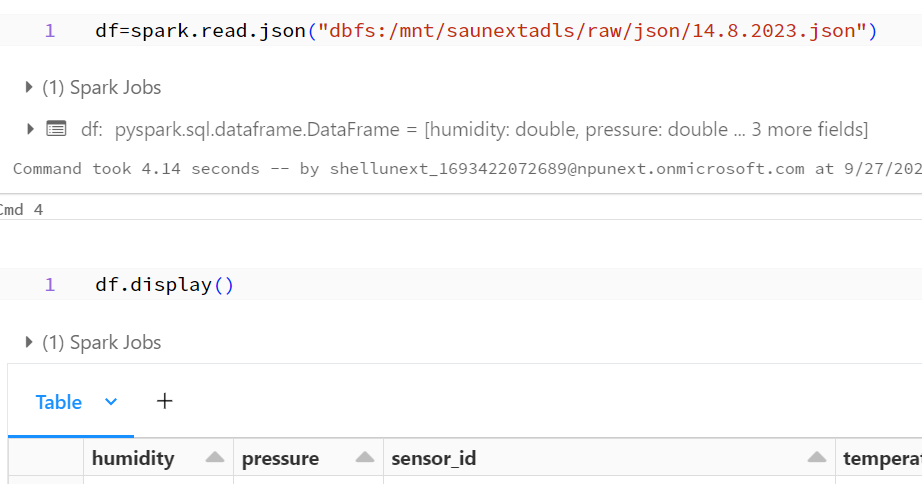
dbutils.fs.mount(

  source = "wasbs://raw@saunextadls.blob.core.windows.net",

  mount\_point = "/mnt/saunextadls/raw",

  extra\_configs = {"fs.azure.account.key.saunextadls.blob.core.windows.net":"DsZWJs7JVVHZz1I7GKyclV8ejCdj0V2UkqMlgAp6QyVOw5rvrHvmVTgwcThdHUymWg7MXon65/0z+AStj4Yiug=="})



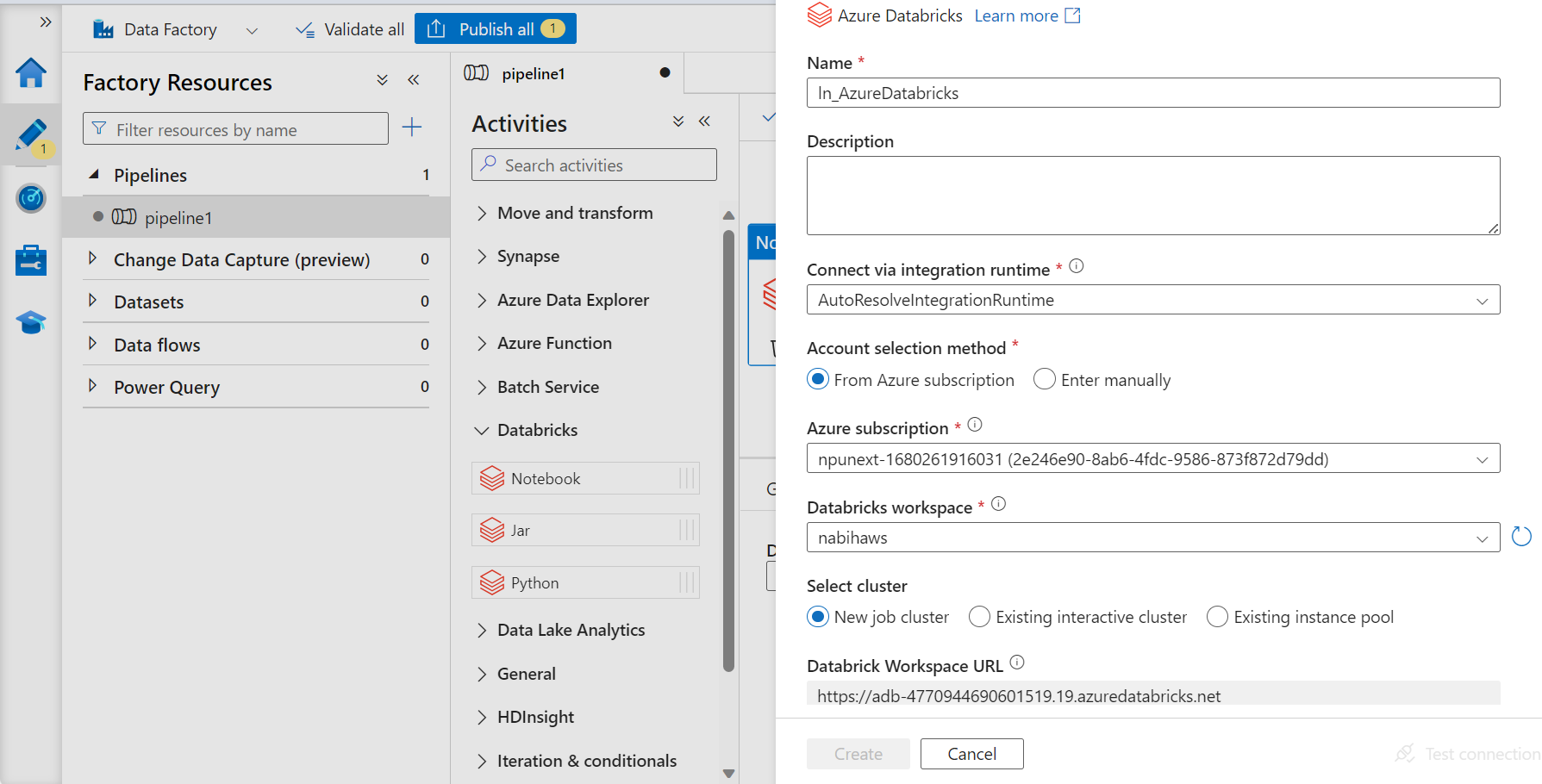


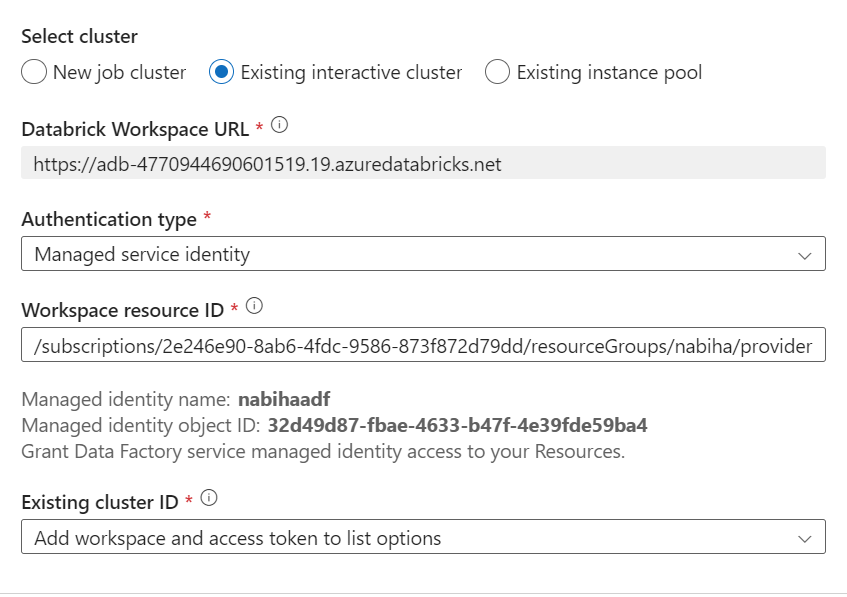




1. Create adf > linked service – connecting to databricks>

Pipeline>databricks> azure data bricks> new



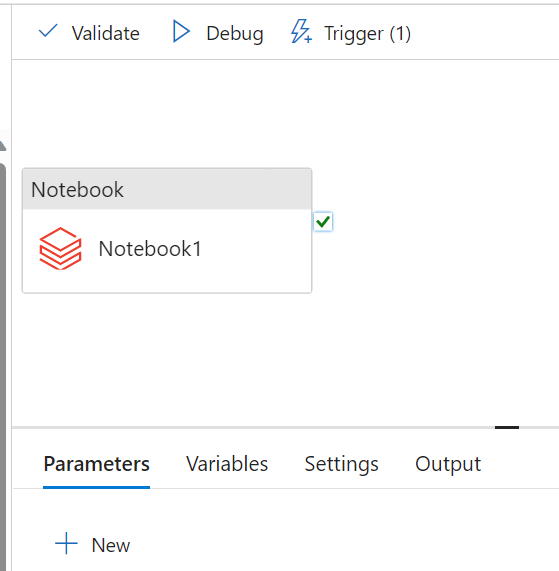


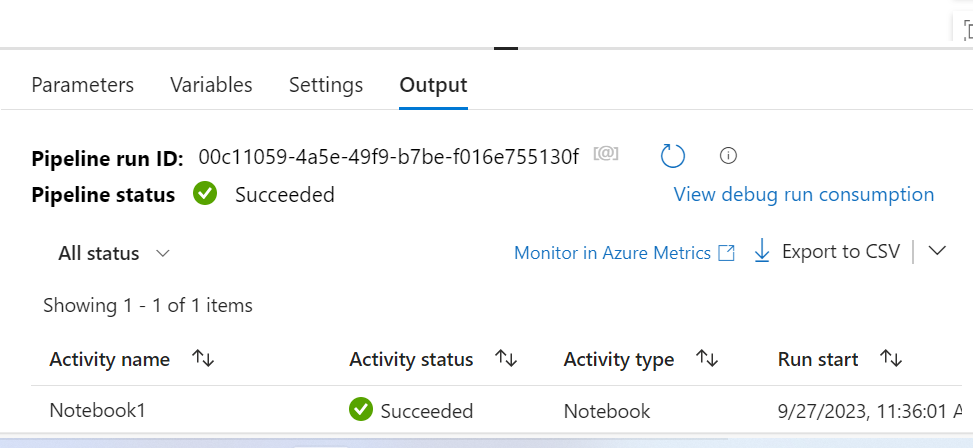
For cluster id – cluster> json>cluster id – **FAILED**

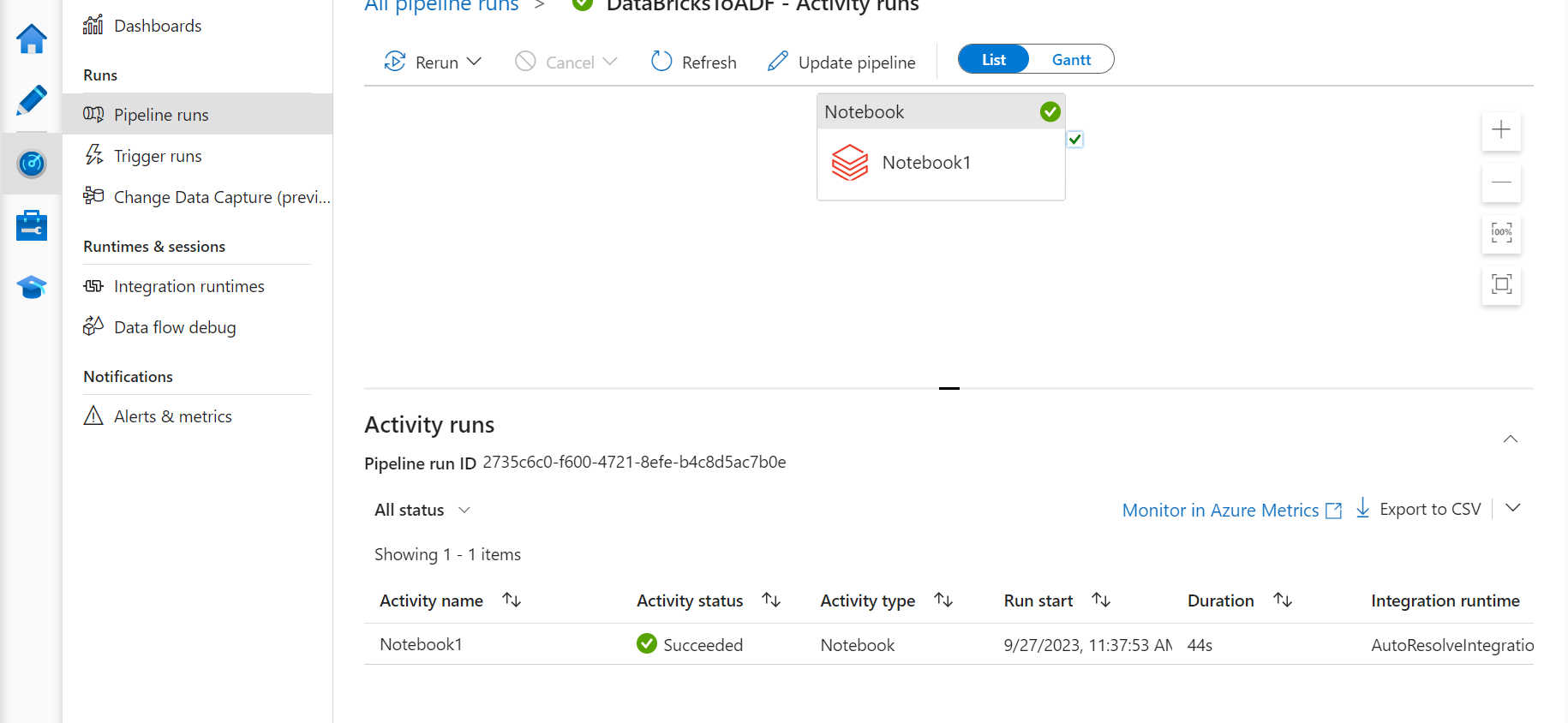
Access Token

Databricls>usersetting>developer>access token>generatre new> copy paste

Go to settings in adf pipeline> validate> add new trigger> validate

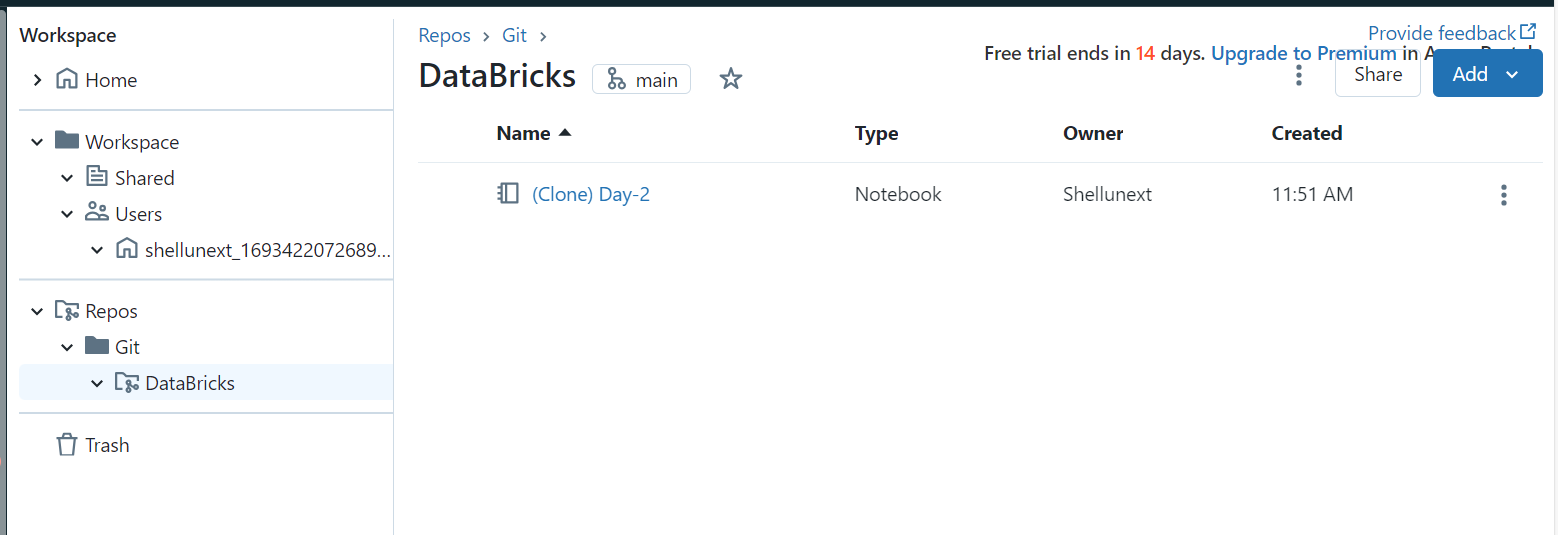




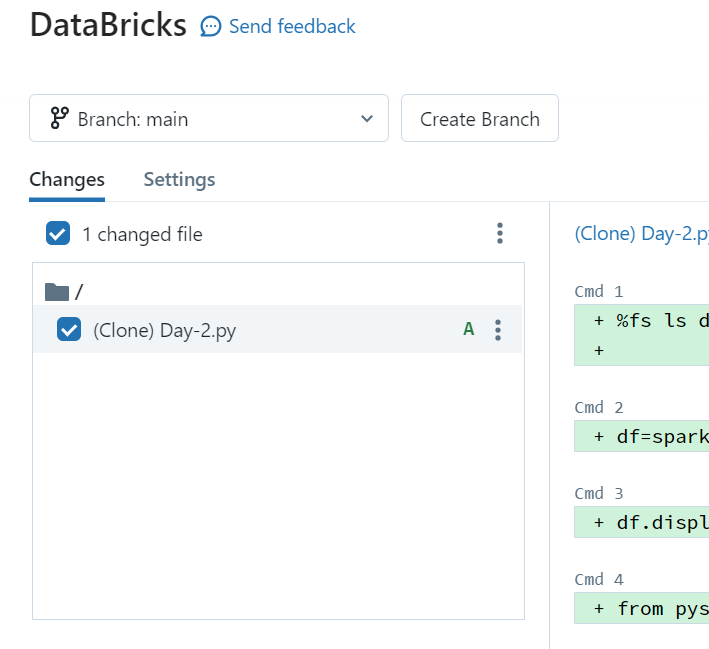


1. Pushing to github

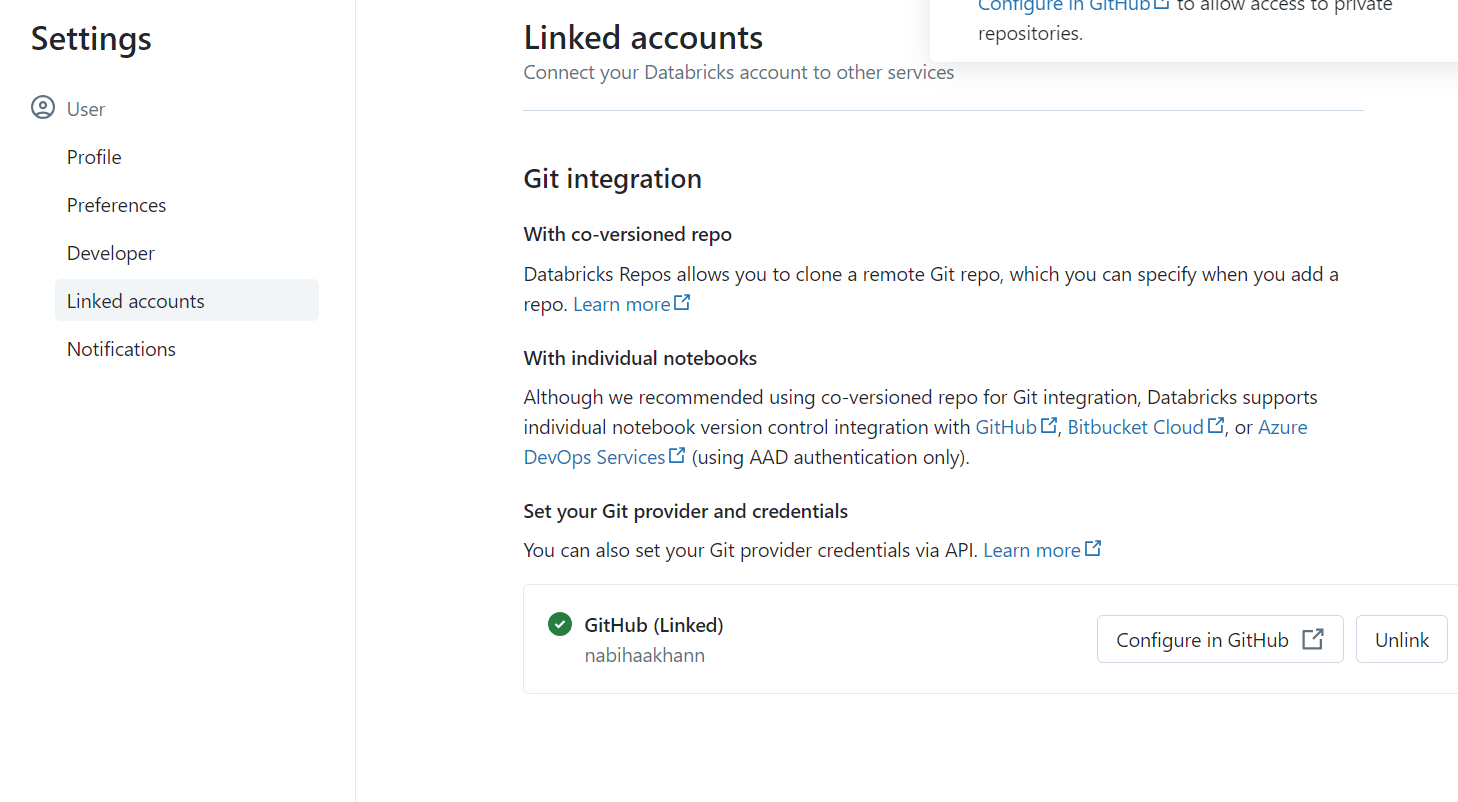
Make a repo in github and copy the url> go to databricks and create a new repo>clone the workspace into the repo

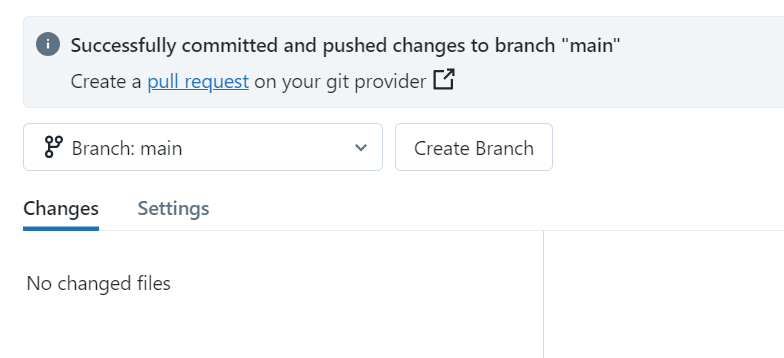


Click on main> new file to be added> add a comment message >



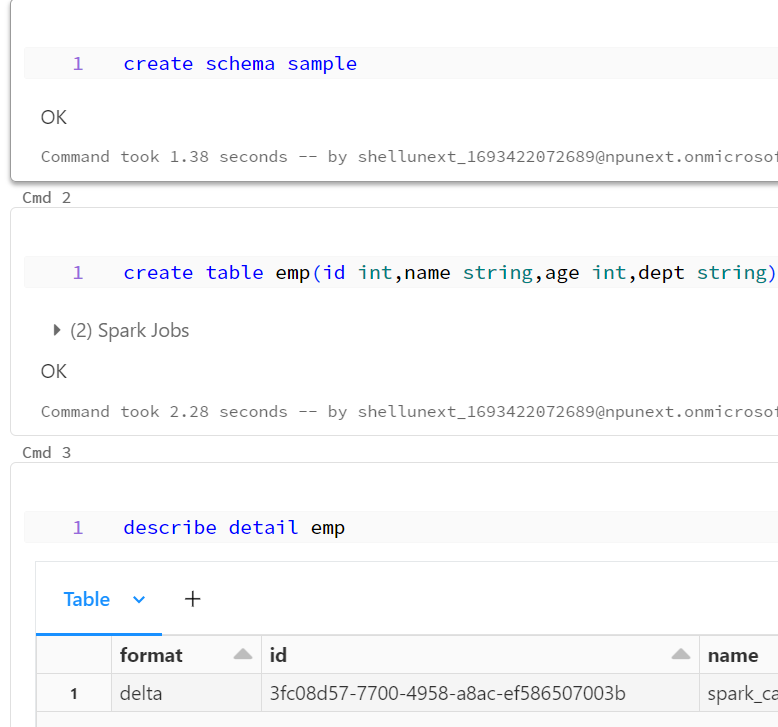
Integrate github> user setting> linked accounts

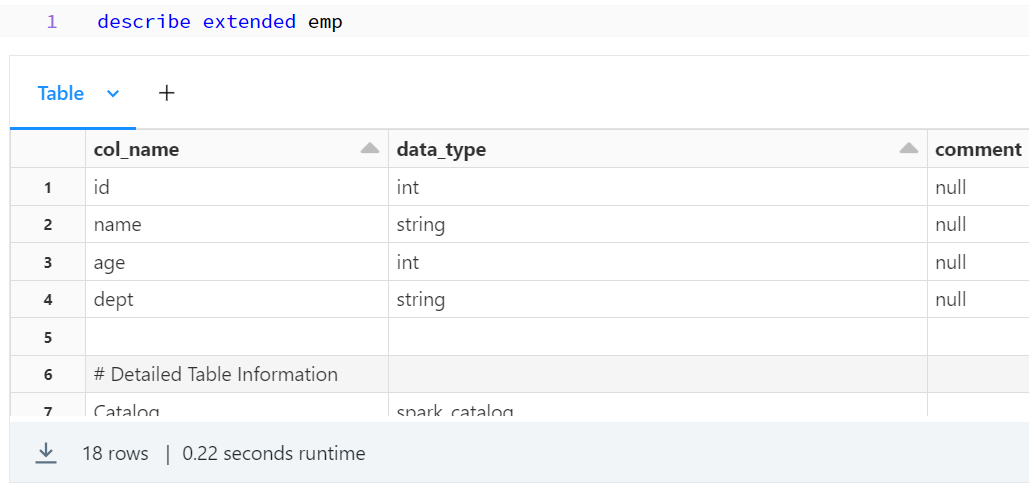




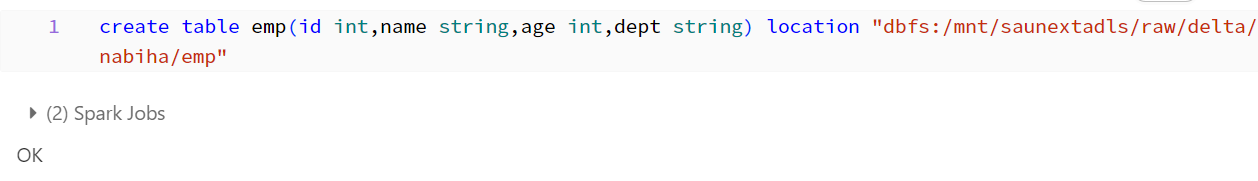
DELTA LAKE



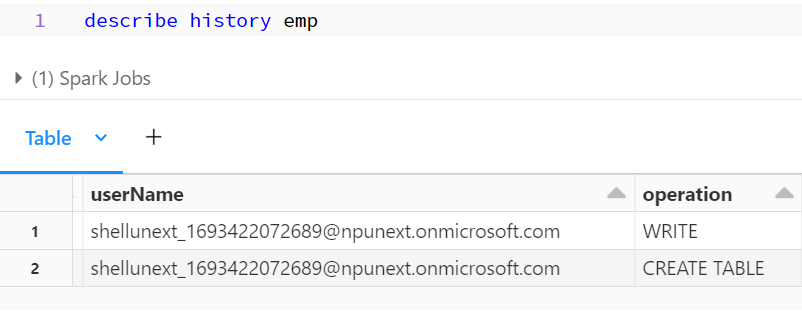




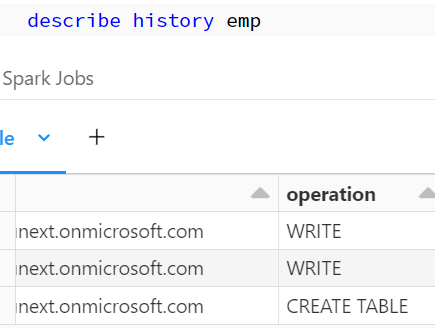
Managed table -> path not mentioned

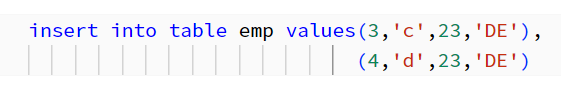


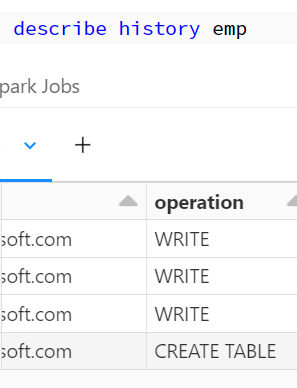




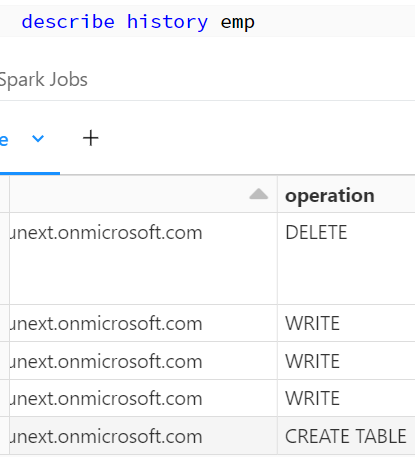


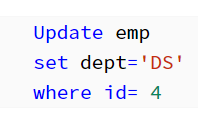


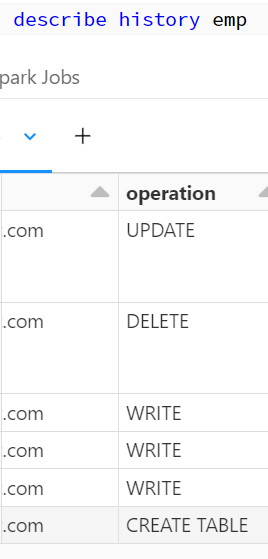






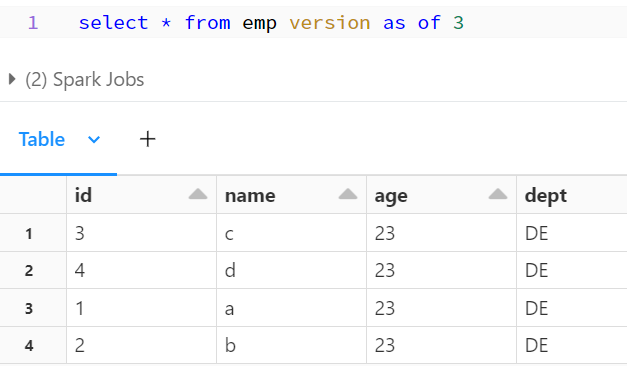




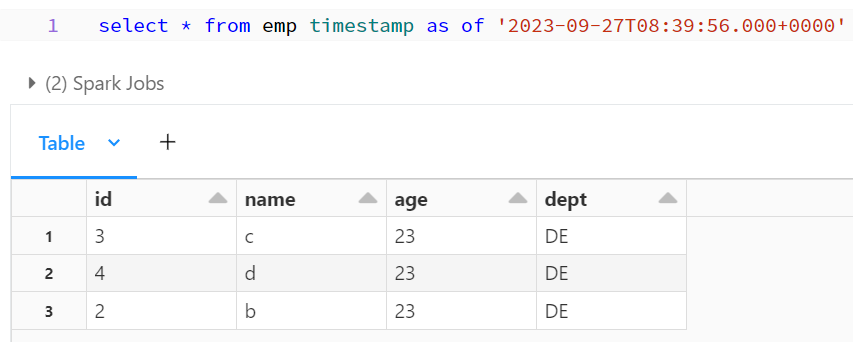


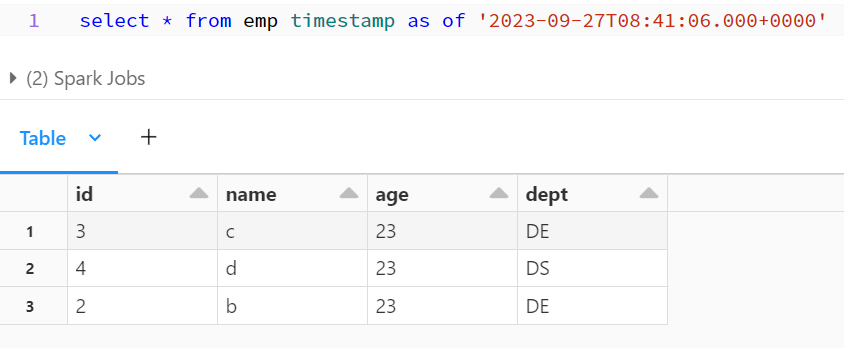
* TimeTravel

Version



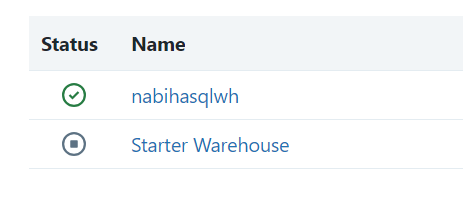
TimeStamp





1. SQL to databricks

Create sql db> server>auth>use both sql and MS> public endpoint>create> configure a connection to sql server databricks> connection strings – hostname> username- > sql warehouse in data bricks> create sql warehouse> 2x-small>create>



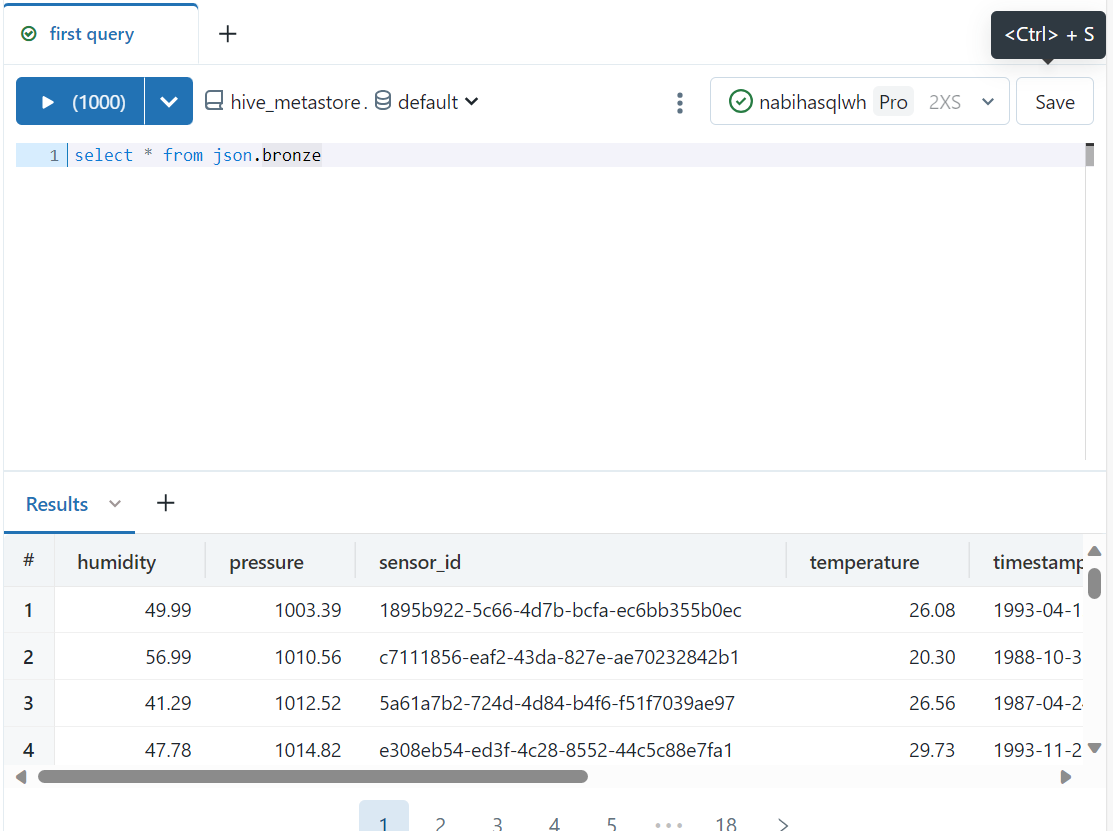
* Query

In sql editor type the query

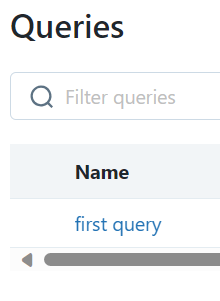
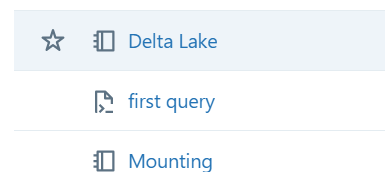
select \* from json.bronze

connect to sqlwarehouse

save the query



Check in queries section & workspace

* Dashboards

Go to dashboards> create new> connect to sql warehouse> add>

