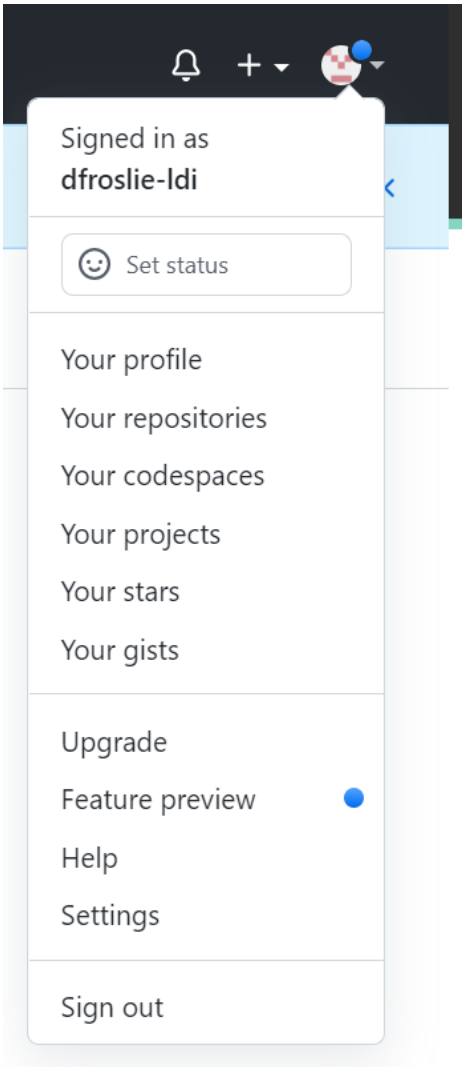
GitHub and Azure Databricks

Contents

Acquire Personal Access Token (PAT) in GitHub 1
Connect Repo In Databricks 2

Acquire Personal Access Token (PAT) in GitHub

1. Log into your GitHub account.
2. Create Personal Access Token
    a. Settings



b. Developer Settings (bottom choice on left)

c. Personal access tokens -> Tokens (classic)

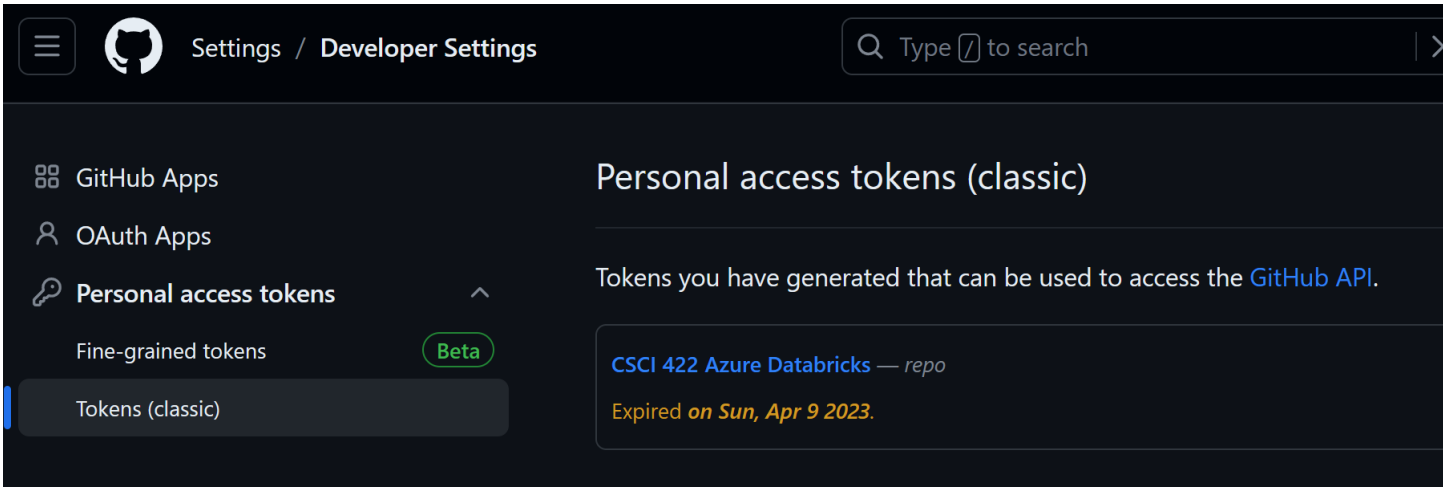
   i. Name token

   ii. Set expiration to the term of the class

   iii. Check 'repo'

   iv. Generate token

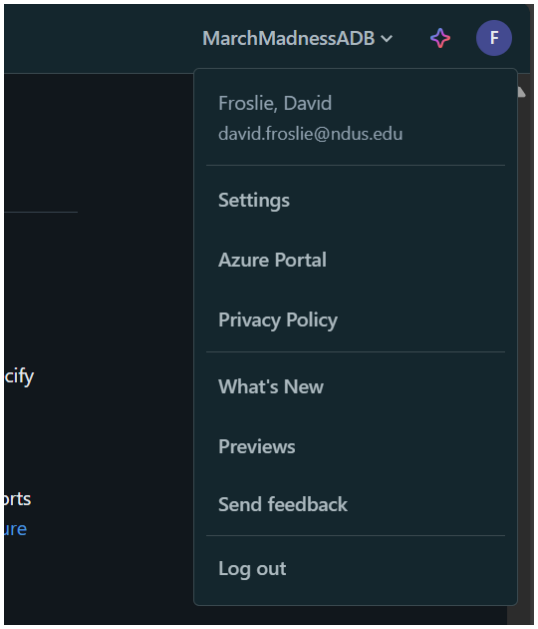
   v. Copy and save the token for use in Databricks.



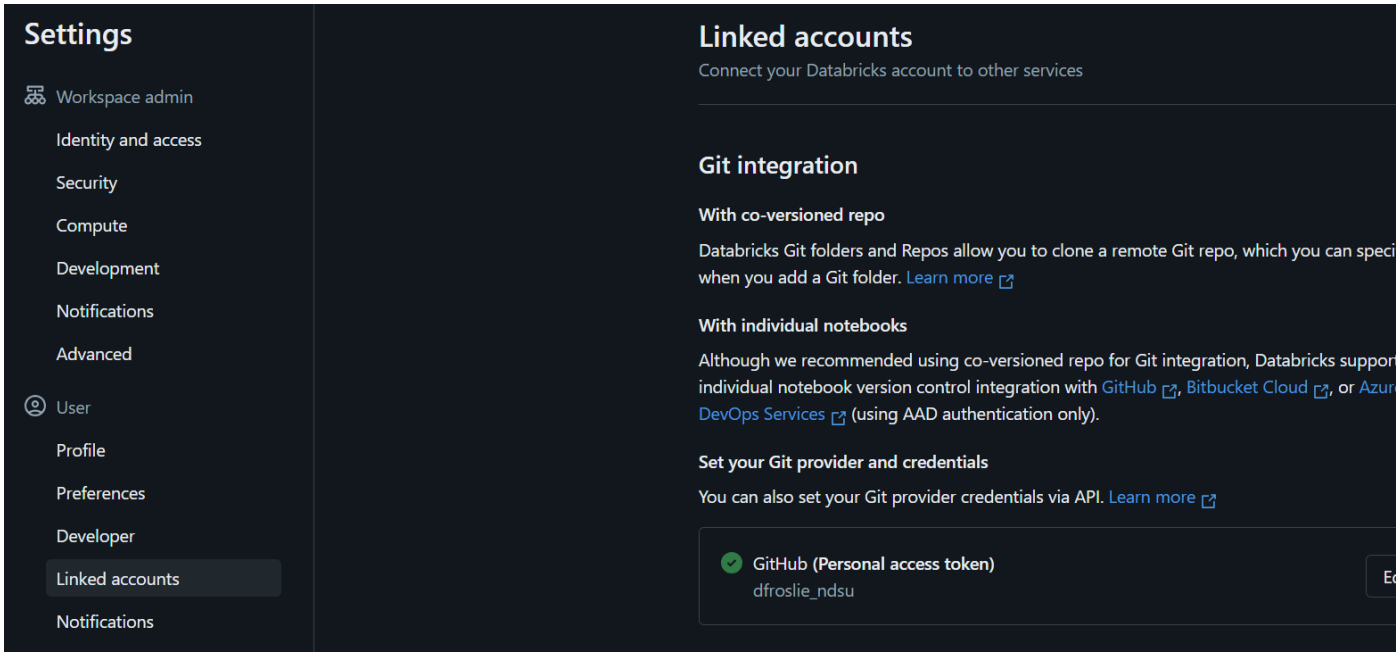
Connect Repo In Databricks

Open your Databricks Workspace.

- Select Settings from the drop down menu in the upper right corner



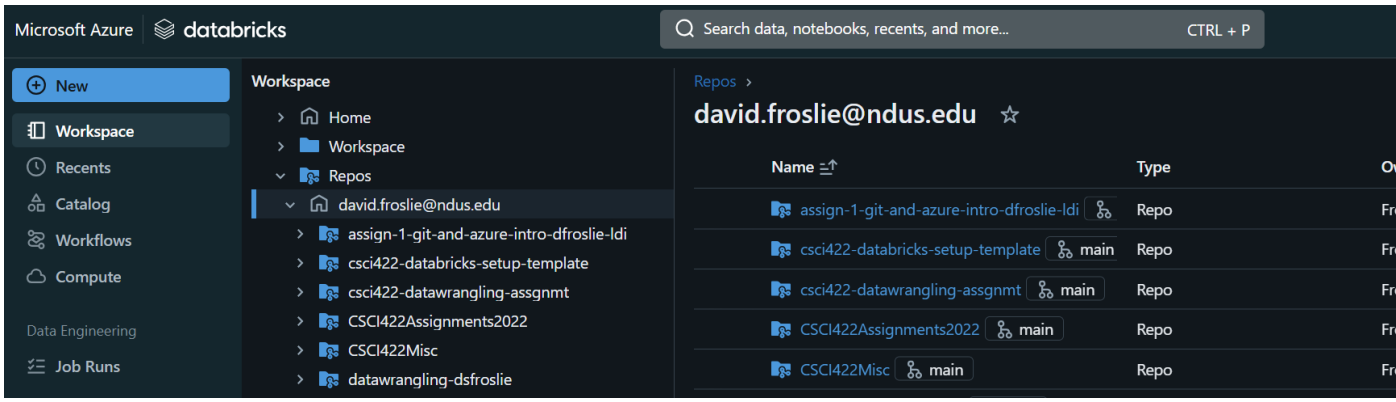
- Go to Linked accounts and Git Integration. Add your personal access token here.



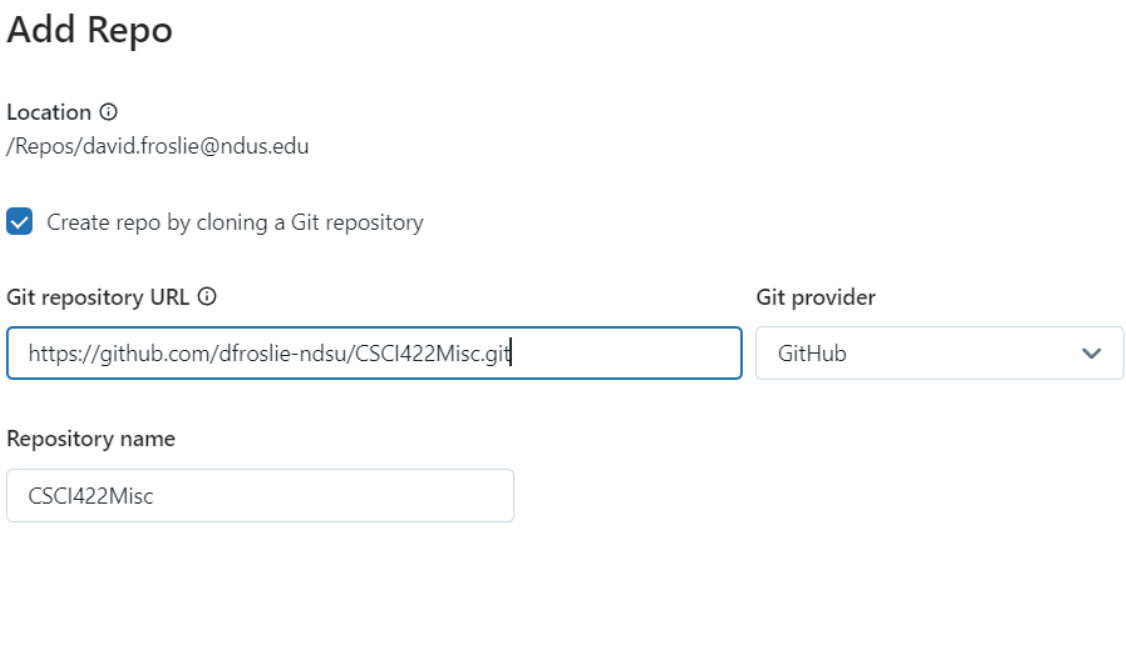
Now to integrate the specific repo:

- Go to Workspaces on left hand menu of Databricks
- Click on your GitHub instance followed by Add->Repo on the right hand side of the page.
    - Note – You can also use ‘Git folder (Recommended)’ and your will find a new folder under your Workspace node. This is a new feature from Databricks that enables private

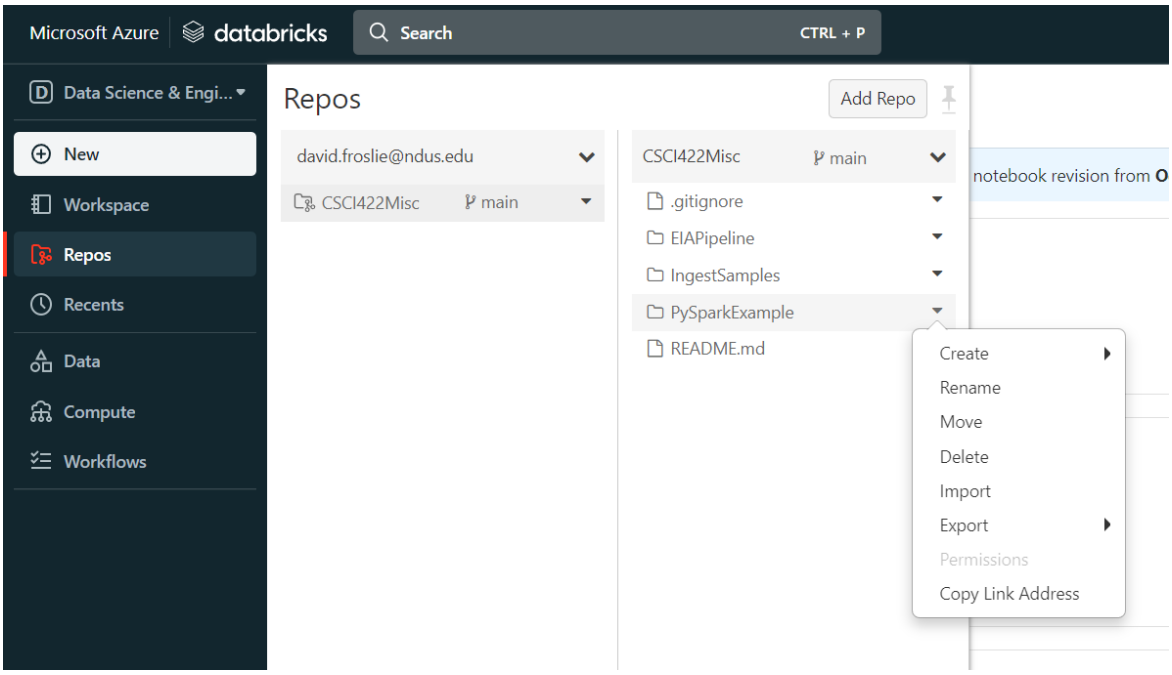
branches associated with specific users in a shared Databricks instance. We won't have that situation in class.



- Get the clone URL for the repo and paste it in. The Git provider and repository name should auto fill



To create new notebooks, do so from the Repo menu:


Microsoft Azure
databricks
Search
CTRL + P
Data Science & Engi...
New
Workspace
Repos
Recents
Data
Compute
Workflows
Repos
Add Repo
david.froslie@ndus.edu
CSCI422Misc
main
CSCI422Misc
main
.gitignore
EIAPipeline
IngestSamples
PySparkExample
README.md
notebook revision from O
Create
Rename
Move
Delete
Import
Export
Permissions
Copy Link Address