GitHub and Azure Databricks

Contents

[Acquire Personal Access Token (PAT) in GitHub 1](#_Toc173229094)

[Connect Repo In Databricks 2](#_Toc173229095)

## Acquire Personal Access Token (PAT) in GitHub

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

Signed in as 
dfroslie-ldi 
Set Status 
Your profile 
Your repositories 
Your codespaces 
Your projects 
Your stars 
Your gists 
Upgrade 
Feature preview 
Help 
Settings 
Sign out 

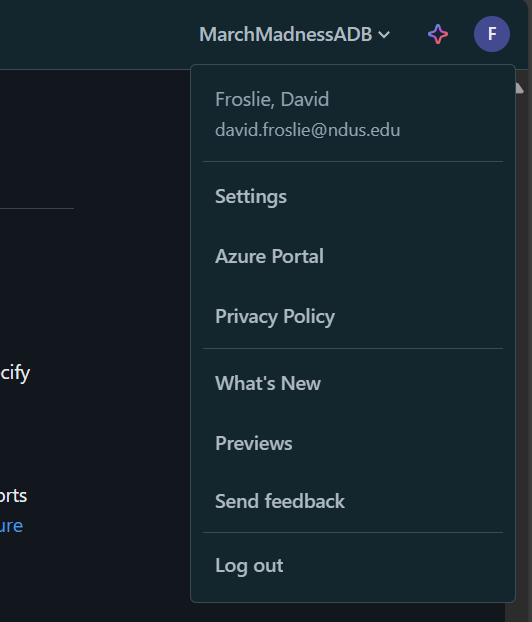
1. Developer Settings (bottom choice on left)
2. Personal access tokens -> Tokens (classic)
   1. Name token
   2. Set expiration to the term of the class
   3. Check 'repo'
   4. Generate token
   5. Copy and save the token for use in Databricks.

Settings 
og GitHub Apps 
OAuth Apps 
Developer Settings 
P 
Personal access tokens 
Fine-grained tokens 
Tokens (classic) 
Beta 
Q Type to search 
Personal access tokens (classic) 
Tokens you have generated that can be used to access the 
GitHub API. 
CSCI 422 Azure Databricks — repo 
Expired on Sun, Apr 9 2023. 
Generate new token 
Last used within the last 10 months 
Revoke all 
Delete 

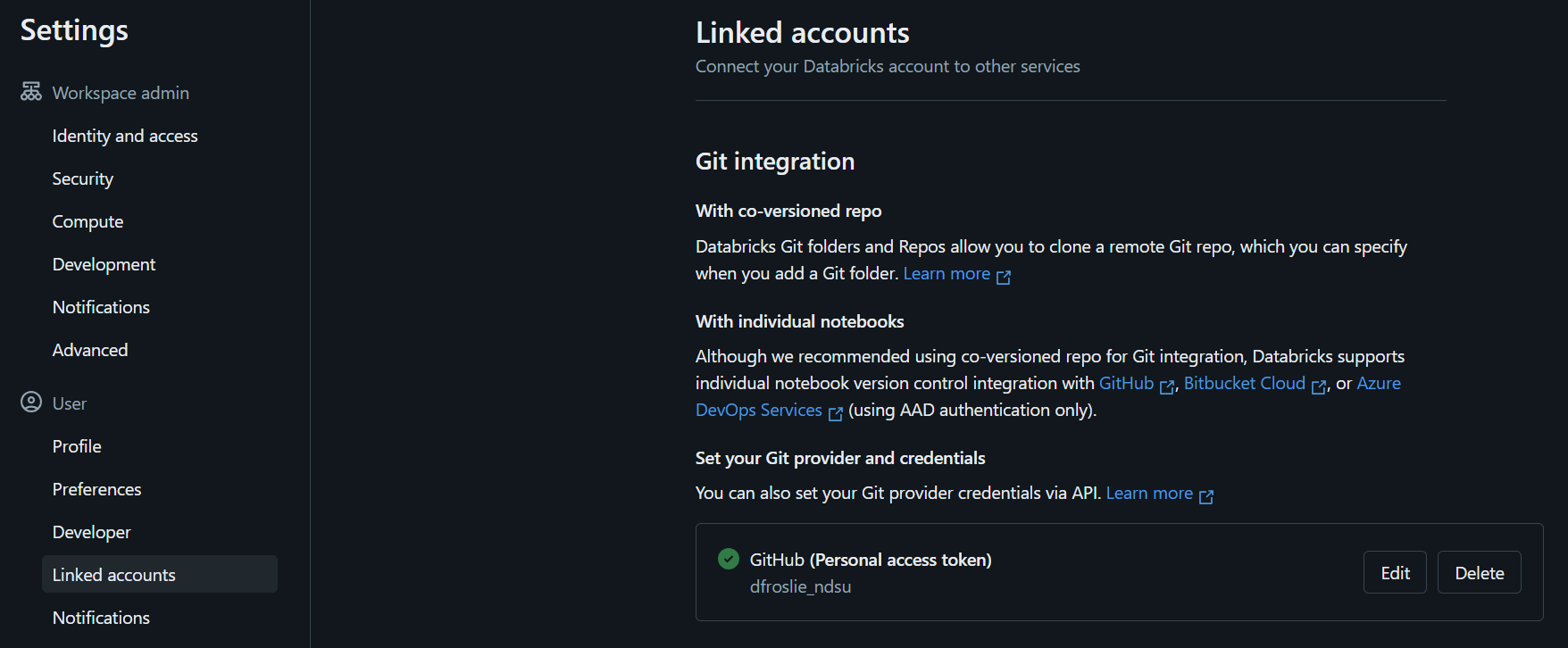
## Connect Repo In Databricks

 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.

A screenshot of a computer

Description automatically generated

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

A screenshot of a computer

Description automatically generated

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

Microsoft Azure @ databricks Q Search 
Data Science & Engi... • 
Repos 
CTRL + P 
Add Repo 
P main 
O New 
Workspace 
Repos 
Recents 
00 Data 
Compute 
Workflows 
david.froslie@ndus.edu 
CSC1422Misc 
P main 
CSC1422Misc 
.gitignore 
ElAPipeline 
IngestSamples 
PySparkExample 
README-.md 
notebook revision from 
Create 
Rename 
Move 
Import 
Export 
permissions 
Copy Link Address 