Add a data version to remote storage(s3) and get the url of data version using DVC

1. Create dvc repo

Since dvc version data using git commit, we need to have git initialized or dvc will do it.

dvc init

2. configure remote repo S3

dvc remote add -d -f dvc_repo s3://tiger-mle-playground/hpp/data_dvc

3. Now add the data which creates the local cache

dvc add <data_directory/data_file>

   For instance,

dvc add datasets/

4. Now commit the changes to git

git add .gitignore datasets.dvc

git commit -m "DVC Demo"

5. Push the data to the remote storage

dvc push

6. Now to retrieve the data version, we need to push .dvc and datasets.dvc to git

git push

Sample Code to Retrieve the data version url

import dvc.api

resource_url = dvc.api.get_url('datasets',repo='<git_repo>')

print(resource_url)