ASSIGNMENT 3

1.What is DVC, and why is DVC used

DVC is a command-line tool written in Python. It mimics Git commands and workflows to ensure that users can quickly incorporate it into their regular Git practice. If you haven't worked with Git before, then be sure to check out Introduction to Git and GitHub for Python Developers.

2. How is DVC different from git and GitHub?

Git: Fast, scalable, distributed revision control system. Git is a free and open source distributed version control system designed to handle everything from small to very large projects with speed and efficiency;

DVC: Open-source Version Control System for Machine Learning Projects. It is an open-source Version Control System for data science and machine learning projects. It is designed to handle large files, data sets, machine learning models, and metrics as well as code.

Git and DVC can be primarily classified as "Version Control System" tools.

3.Which command can be used to initialise a DVC project?

init

Initialize a DVC project in the current working directory.

Synopsis

usage: dvc init [-h] [-q | -v] [--no-scm] [-f] [--subdir]

4. In What all use cases DVC can be used?

- track and save data and machine learning models the same way you capture code;
- create and switch between versions of data and ML models easily;
- understand how datasets and ML artifacts were built in the first place;
- compare model metrics among experiments;
- adopt engineering tools and best practices in data science projects;

5. Which command can be used to reproduce the entire pipeline?

repro

Reproduce complete or partial pipelines by executing commands defined in their stages in the correct order.

usage: **dvc repro** [-h] [-q | -v] [-f] [-i]

[-s] [-p] [-P] [-R] [-m]

[--downstream] [--force-downstream]

[--pull] [--dry]

[--glob] [--no-commit] [--no-run-cache]

[targets [<target> ...]]

positional arguments:

targets Stages to reproduce. 'dvc.yaml' by default.

6. Which DVC command can be used to check metrics?

metrics

A set of commands to display and compare metrics: show, and diff.

Synopsis

usage: **dvc metrics** [-h] [-q | -v] {show,diff} ...

positional arguments:

COMMAND

show Print metrics, with optional formatting

diff Show changes in metrics between commits.

7.Can we store a large amount of Data on GitHub? Justify

Every account using Git Large File Storage receives 1 GB of free storage and 1 GB a month of free bandwidth. If the bandwidth and storage quotas are not enough, you can choose to purchase an additional quota for Git LFS.

Using Git LFS, you can store files up to:

Product Maximum file size

GitHub Free 2 GB

GitHub Pro 2 GB

GitHub Team 4 GB

GitHub Enterprise Cloud 5 GB