GR 5074 Advanced Machine Learning
Recitation 1

Rachel Lee

Columbia University

January 7th, 2022

1. Git - Basic Concepts

1.1. Version Control System (VCS)

Version Control System (VCS) is a software that helps software developers to work together and maintain a complete history of their work. Listed below are the functions of a VCS:

- Allows developers to work simultaneously.
- Does not allow overwriting each other's changes.
- Maintains a history of every version.

Following are the types of VCS:

- Centralized version control system (CVCS)
- Distributed/Decentralized version control system (DVCS)

1.2 Distributed Version Control System

- Centralized version control system (CVCS) uses a central server to store all files and enables team collaboration
- But the major drawback of CVCS is its single point of failure, i.e., failure of the central server
- For DVCS, if the sever goes down, then the repository from any client can be copied back to the server to restore it. Every checkout is a full backup of the repository.

1.3 Advantages of Git

- Free and open source
- Fast and small
- Implicit backup
- Security
- No need of powerful hardware
- Easier branching

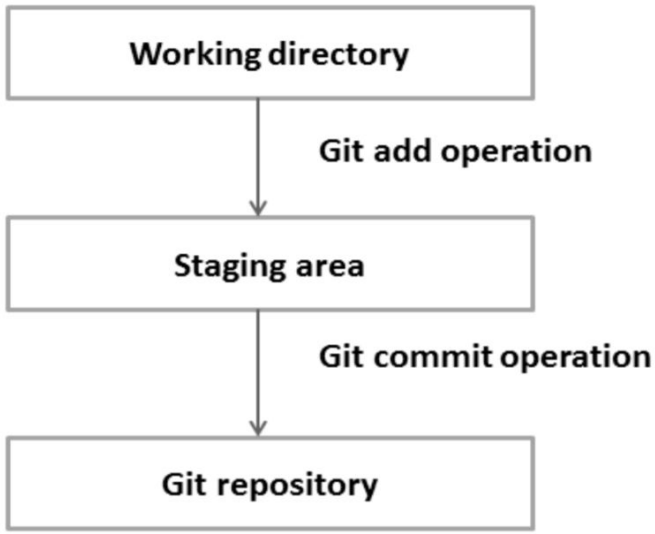
2. DVCS Terminologies

2.1 Working Directory and Staging Area or Index

- In CVCS, developers generally make modifications and commit their changes directly to the repository.
- But Git doesn't track each and every modified file. Whenever you do commit an operation, Git looks for the files present in the **staging area**.
    - Only those files present in the staging area are considered for commit and not all the modified files.

2.2 Git Basic Workflow

- **Step 1**: You modify a file from the working directory
- **Step 2**: You add these files to the staging area
- **Step 3**: You perform commit operation that moves the files from the staging area. After push operation, it stores the changes permanently to the Git repository.



2.3 Example

You modified two files, namely “sort.c” and “search.c” and you want two different commits for each operation. You can add one file in the staging area and do commit. After the first commit, repeat the same procedure for another file.

```
# First commit
[bash]$ git add sort.c

# adds file to the staging area
[bash]$ git commit -m “Added sort operation”
# Second commit
[bash]$ git add search.c

# adds file to the staging area
[bash]$ git commit -m “Added search operation”
```

2.4 Other terminologies

- **Commit**: Commit holds the current state of the repository. Every commit object has a pointer to the parent commit object.
- **Branches**: Branches are used to create another line of development. By default, Git has a master branch. Usually, a branch is created to work on a new feature. Once the feature is completed, it is merged back with the master branch and we delete the branch.
- **Clone**: Clone operation creates the instance of the repository. Clone operation not only checks out the working copy, but it also mirrors the complete repository.
- **Pull**: Pull operation copies the changes from a remote repository instance to a local one.
- **Push**: Push operation copies changes from a local repository instance to a remote one. This is used to store the changes permanently into the Git repository.