Building Software Systems

Lecture 2.2

Introduction to Git

SAURABH SRIVASTAVA

ASSISTANT PROFESSOR

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

IIT (ISM) DHANBAD

What is Git?

A version control system

- You can keep multiple versions of files and folders
- Used primarily for storing versions of an application, but can be used for any type of files (text or binary)

You start by creating a *repository* or *repo* in short

- Files in a repo are “tracked” for changes
- Git uses Hashing to find a signature of every file
- The Hashes of a directory is computed by combining the hashes of its files and subdirectories
- Usually, we configure everything in the directory to be tracked of the repo – including files and subdirectories
- You can, although, explicitly ask Git to not track certain files or subdirectories

You can have branches in a repo

- A branch represents a different direction of tracking
- For instance, if you want to maintain two different versions of the same repo simultaneously …
- … say for building a *product line*, you can have multiple branches for the same

Some common terms

Commit

- Refers to a checkpoint/milestone in the timeline of the repository (e.g., “v1”, “v1.2” etc.)
- Git stores a message associated with each commit along with the commit timestamp

Head

- Usually refers to the “last” commit state of the repository
- The head may be “detached” to point to a previous commit in the timeline as well

Origin

- If you “cloned” the repository from a remote location, origin refers to the remote repo
- It is useful if you wish to “sync” your local work with the remote repository over a period of time
- Typically, you may either want to “pull” the current state of remote repository on your local machine …
- … or you may want to “push” your work to the remote respository

Git Branches

Branches are a fundamental feature that allows you to diverge from the main line of development

- The main branch is usually called `master` or `main`
- Different branches may be created for different products or to have dev/test codes

Too many branches can be problematic

- Usually, you should keep only one – specially, on your “remote” side
- Local branches are fine !!

You may merge branches with other branches (e.g., merging with `main` or `master`)

What is github.com?

It is a popular public website which can be used to host Git repositories

- You can create your own public/private repositories and ”push” your work to the same
- It provides advanced options to developers for setting up “CI/CD pipelines” (will be discussed later in course)
- You can create an account for free and start storing your repos to prevent loss in case of local failures

There are other such options as well

- BitBucket is another popular option by Atlassian
- AWS CodeCommit is another option, usually used by users of AWS

Check out the Video Recording of the Lecture

https://www.youtube.com/watch?v=BZs9QV_6B3k