BEGINNER’S GUIDE TO VERSION CONTROL WITH GIT COMMANDS & GITHUB FOR COLLABORATIVE WORK

**Author: Mudit Srivastava**

**March 2021**

Contents

1. Diff ……………………………………………………………………………….01
2. Patch ……...………………………………………………………………………..01
3. Making Copies …...…………………………………………………………….….01
4. Check Git Version …………………………………………………………….……01
5. Basic Configuration ………………………………………………………………..01
6. Making Directory …………………………………………………………………..01
7. Working Tree ………………………………………………………………………02
8. Adding to Staging Area & Commit ……………………………………………...…02
9. Modifying a File …………………………………………………………………...02
10. Log …………………………………………………………………………………02
11. Skipping Staging Area …………………………………………………………..…02
12. House Keeping ……………………………………………………………….…….02
13. Hidden Files ………………………………………………………………….…….03
14. Undoing changes before Committing ……………………………………….……..03
15. Amending Commits ……………………………………………………….……….03
16. Branches …………………………………………………………………………... 03
17. Merge Conflicts ……………………………………………………………..……..04
18. Remote Repositories ……………………………………………………….…..…..04
19. Remote ……………………………………………………………………….…….04
20. Fetching ……………………………………………………………………….…...05
21. Pulling ……………………………………………………………………………...08
22. Push Conflicts ……………………………………………………………………...08
23. Pushing Remote Branches ………………………………………………………....11
24. Merging Remote Branches ………………………………………………………...12
25. Other use of Rebasing ……………………………………………………………..12
26. Forking …………………………………………………………………………….13
27. Pull Request ……………………………………………………………………….13
28. Squashing Changes ………………………………………………………………..13
29. Code Reviews ……………………………………………………………………..14
30. Managing Collaborations ………………………………………………………….14
31. Issue Trackers ……………………………………………………………………..14
32. Continuous Integration (CI) System ……………………………………………....14
33. Continuous Deployment …………………………………………………………..15

Notes ……………………………………………………………………………....16

Good Practices …………………………………………………………………….18

1. **Diff**
   1. $ **diff old\_file.py new\_file.py**
   2. ‘<’ Removed from the file.
   3. ‘>’ Added to the file
   4. ‘c’ Changed.
   5. ‘a’ Added
   6. $ **diff -u old\_file.py new\_file.py** (Unified format=> show diff with more context)
   7. $ **wdiff** (shows changes in words instead of working lines of the code)
   8. Graphical diff examples. **Meld, kDiff3, vimdiff**
   9. $ **diff -u old\_file new\_file > change.diff** (‘>’ Storing result from diff -u to change.diff)
   10. The results of diff could be stored in a diff file or patch file.
   11. $ **git diff** (shows modifications in all the files that **changes before staging**)
   12. $ **git diff -staged** (shows modifications in all the files that are **in Staging area**)
2. **Patch**
   1. $ **patch new\_file.py < changes.diff** (If you want to reflect/apply the changes mentioned in change.diff to your original py file)
3. **Making Copies**
   1. $ **cp original.py original\_unmodified.py**

$ **cp original.py original \_fixed.py**

1. **Check Git version**
   1. $ **git –version (current 2.27)**
2. **Basic Configurations** (Identify user)
   1. $ **git config --global user.email “you email address Here”**
   2. $ **git config --global user.name “Your Name Here”**
   3. $ **git config -l** (to show config)
3. **Making Directory**
   1. $ **mkdir test** (make directory)
   2. $ **cd test** (current directory)
   3. $ **git init** (git initialize)
   4. $ **ls -l** (no. of files)
   5. $ **ls -la** (show file starting a dot)
   6. $ **ls -l .git/** (show more details of the directory)
4. **Working tree**
   1. $ **cp ../file1.py**
   2. $ **ls -l**
5. **Adding to Staging Area & Commit**
   1. $ **git add file1.py**
   2. $ **git status** (commits, to be committed, untracked)
   3. $ **git commit** (commit everything on staging area)
   4. ESC and then :wq (to exit vim)
   5. $ **git commit -m ‘message here’** (Or add commit message)
   6. $ **git commit -m ‘1st msg’ -m ‘2nd msg’** (For multiple lines)
   7. $ **git add -p** (shows modified lines and ‘y’, ’n’ to stage the changes, All tracked files)
   8. $ **git add \*** (Add all the files to staging area)
6. **Modifying a file**
   1. $ **vim file1.py** (open a file in vim)
7. **Log**
   1. $ **git log** (shows history, commit messages)
   2. $ **git log -p** (shows changed lines, p from created patch)
   3. $ **git log --stat** (shows no. of changed lines in all the commits)
   4. $ **git log -2** (shows last 2 commits. Add any no. you like)
   5. $ **git show** (view the log message and diff output the last commit if we don't know the commit ID)
   6. $ **git show insert\_commit\_id\_here** (shows changed lines for specific commit, you can also just use first 6-8 characters of commit id to work)
8. **Skipping Staging Area**
   1. $ **git commit -a**

It is not an add then commit feature, instead it only directly commits the files those are tracked. Untracked files are not committed.

1. **House keeping**
   1. $ **git rm file\_name\_here** (To delete files that are committed)
   2. $ **git commit -m ‘Deleted file’** (Commit this deletion)
   3. $ **git mv old\_file\_name new\_file\_name** (To rename files)
   4. $ **git commit -m ‘Renamed old file name’** (Commit this renaming)
2. **Creating a hidden file** 
   1. $ **git > .gitignore** (Creates Gitignore type file, here dot means hidden)
   2. $ **git > hello.txt** (Creates a text file)
   3. $ **echo “hello” > hello.txt** (Copy text to hello.txt file)
   4. $ **echo .datfiles > .gitignore** (Copy names of .datfiles to .gitignore)
3. **Undoing Changes before Committing**
   1. $ **git checkout file\_name\_here** (Will restore to its latest commit, restore changes that **are not staged yet**)
   2. $ **git checkout file\_name\_here -p** (Will restore asking each lines changed)
   3. $ **git reset HEAD file\_name\_here** (Will restore changes **in the staging area** to it’s HEAD)
   4. $ **git reset -p** (Asks what changes need to be reset)
4. **Amending Commits**
   1. $ **git commit --amend** (opens the last commit msg along with updated staging area before this and last commit. Overrides last commit)
   2. $ **git revert HEAD** (Revert back the last commit and hence the changes made. It makes a new commit with a msg containing revert details and ID reverted, **Rollback**)
   3. $ **git revert commit\_id\_here** (Revert back any commit and hence the changes made. It makes a new commit with a msg containing revert details and ID reverted)
5. **Branches**
   1. $ **git branch** (list all the branches present, \* shows current branch we’re on)
   2. $ **git branch new\_branch\_name\_here** (Create new branch)
   3. $ **git checkout new\_branch\_name\_here** ( To select this particular branch, \* points to this branch now)
   4. $ **git checkout -b new\_branch\_name\_here** ( To create and select this particular branch, Shortcut)
   5. When switching back to the master branch, the actual files also change to the latest working position of the master branch. Hence the directory changes with the branch currently we are pointing to. Also, the commit history changes too.
   6. $ **git branch -d branch\_name** (Delete this branch. Gives error if unmerged changes with the master, you can force deletion just use **-D** instead of -d)
   7. $ **git merge branch\_name** (Merge this branch with master. Make sure you are currently pointing to master branch and then do this. Head points to now the Master and the merged branch\_name)
6. **Merge Conflicts**
   1. $ **git status** (Use this to locate where merge conflicts are present. Git also automatically add conflict details to the actual file for faster corrections)
   2. $ **git add** (Add the rectification done and then use git status to check if all the conflicts are resolved)
   3. $ **git commit** (automatic commit message as merge branch , you can add resolving merge conflict message too. Automatically merged now)
   4. $ **git log --graph --oneline** (graphical way to represent the merge)
   5. $ **git merge --abort** (Stop the merge if conflict is too complicated for now)
7. **Remote repository (GitHub)**
   1. $ **git clone repo\_url\_here** (Making first local copy from Github repo)
   2. $ **git push** (to push local repo to remote repo)
   3. $ **git push** **-u origin branch\_name** (to push local repo to remote repo if **1st time**)
   4. $ **git config --global credential.helper cache** (To cache the username password for 15 minutes)
   5. $ **git pull** ( update local with new changes on remote repo)
8. **Remote**
   1. $ **git remote -v** (When clone, repo default name **origin**. Fetch/ Push urls)
   2. $ **git remote show origin** (to show origin remote details, where it’s pointing, up to date? )
   3. $ **git remote add remote\_name remote\_url** (Add remote)
   4. $ **git branch -r** (display Remote branches, READ only)
   5. $ **git status** (show additional details like origin (remote) up do date with local branch)

![Text

Description automatically generated]()

1. **Fetching**

![Text

Description automatically generated]()

(When remote repo is modified without reflecting local remote branch => local out of date)

* 1. ![Text

     Description automatically generated]()$ **git fetch** ( to download remote repository to **current** remote **branch** (local copy) but **not** automatically **merge** to our local branch (main) )
  2. $ **git log origin/main** (notice origin/main points to remote repo (extra) and main (local) points to latest local commit) i.e. local branch is behind origin/main by 1 commit

![Text

Description automatically generated]()

![Text

Description automatically generated]()

* 1. $ **git merge origin/main** (to merge remote origin/main to our local main)

![Text

Description automatically generated]()![Text

Description automatically generated]()![Text

Description automatically generated]()

NOTE: git **fetch** fetches remote updates but doesn't merge; git **pull** fetches remote updates and merges

1. **Pulling** 
   1. $ **git pull** (Automatic Fetch & Merge)

![Text

Description automatically generated]()

* 1. $ **git remote update** ( fetch **all** remote **branches** **without automatic merge** with local branches. Use checkout and merge as needed)

1. **Push conflicts**
   1. ![Text

      Description automatically generated]()$ **git push**

(Since remote repo is modified, first pull to update local repo)

* 1. ![Text

     Description automatically generated]()$ **git pull**

(Merge conflict here as local repo is modified at the same lines as of remote repo)

* 1. ![Text

     Description automatically generated]()$ **git log –graph --oneline --all** (to visualize conflict, Three way merge required here)
  2. ![Text

     Description automatically generated]()**$ git log -p origin/main**

While our local repo has:

![Text

Description automatically generated]()

* 1. $ **vim README.md** (resolve conflict)

![Text

Description automatically generated]()

(Choosing March, 2021, the local repo)

![Graphical user interface, text, application

Description automatically generated]()

* 1. ![Text

     Description automatically generated]()$ **git add README.md** (Now add, commit and push)
  2. $ **git log --graph --oneline –all**

![Text

Description automatically generated]()![Text

Description automatically generated]()

(Automatic commit message generated when committing after resolving push conflict)

1. **Pushing remote branches** 
   1. $ **git push -u origin branch\_name\_here** (Pushing local remote branch to remote repository)
2. **Merging remote branch**
   1. $ git checkout main
   2. $ git pull (all above steps to update main branch )
   3. $ git log --graph --oneline --all (check the common ancestor)
   4. $ **git checkout branch\_name**
   5. $ **git rebase main** (changes the common ancestor to latest main commit, **still need to manually merge**)

![Timeline

Description automatically generated]()![Timeline

Description automatically generated with low confidence]()

(Image source: coursera.org/ google)

**Rebasing** – You can use Merge but then 3 way merge will happen. The problem with this type of merge is that it’s difficult (because of the non-linear commit history) to find the moment any bug was introduced. But by using rebase, a **fast forward merge** will happen and thus it gets easier with finding that moment because of its **linear** nature.

* 1. $ **git checkout master**
  2. $ **git merge branch\_name** ( Merge the new linear branch)
  3. $ **git push --delete origin branch\_name** (delete remote branch)
  4. $ **git branch -d branch\_name** (delete local branch)
  5. $ **git push** (push: since the branch is reflected on main, delete this branch both locally and remotely)

1. **Other use of Rebasing** : To rebase the changes (small) made on the main branch by two contributors at the same time to maintain a **linear** history.
   1. $ **git fetch** ( Don’t use git pull as it will perform a 3 way merge here. Instead use fetch so that the remote repo is copied to remote branch. But don’t merge remote branch (origin/main) to our local branch (main) ).
   2. $ **git checkout main**
   3. $ **git rebase origin/main** (attaching linearly the local main branch to origin/main (remote) branch. There might arise some conflicts here)
   4. $ **vim conflict\_file** (solve conflict in the file to continue rebase)
   5. $ **add conflict\_file** (after solving, stage the changes)
   6. $ **git rebase --continue** (complete the rebase process)
   7. $ **git push** (push the new linear structure so that origin/main also points to HEAD-> main)
2. **Forking**
   1. A way of creating a copy of the given repository so that it belongs to our user.
3. **Pull Request**
   1. A commit or series of commits that you send to the owner of the repository so that they incorporate it into their tree.
   2. Since not everyone has a commit access to a repository, pull request is a way to suggest patches or other changes to the team with commit access.
   3. **$** **git push** **-u origin branch\_name** (to push local repo to forked repo if **1st time**)
   4. Before creating a pull request, it's always important to check that the code will merge successfully. GitHub tells us that our change can be automatically merged
   5. Github automatically adds any further commit in the forked repo branch to the existing pull request. In order to stop auto update, make another branch in the forked repo and commit changes to it. Then create a new pull request.
4. **Squashing Changes** (combining all commits for pull request)
   1. Create a single commit that includes both changes and a more detailed description than the one we submitted.
   2. The rule to not push rebase to published commit is not valid for pull request as we will be working with forked repo.
   3. **$** **git rebase -i main** (interactive version of rebase)
   4. An interactive rebase, a text editor opens with a list of all the selected commits from the oldest to the most recent. By changing the first word of each line, we can select what we want to do with the commits.
   5. **Pick**: default action for all the commits here. Rebases the commit to the branch we selected. (Same action as $ git rebase main)
   6. **Squash**: it combines both the commits so to form single commit for pull request. Also let us modify commit descriptions.

* **pick commit1\_id commit1**
* **squash commit2\_id commit2** (Use squash in the text editor of interactive rebase to combine)
* A combine commit message will open in editor where you can modify the descriptions like combining commit messages into a single descriptive message.
* $ **git push** (Git will give warning as merging rebase with published branch and no fast forward merge)
* **$ git push -f** (Since we don’t want to merge but replace old commits with new single commit, hence force this push. The diverging branches will be gone)
  1. **Fixup**: same as squash but discard commit’s log message.

1. **Code Reviews**
   1. Going through someone else's code, documentation or configuration and checking that it all makes sense and follows the expected patterns.
   2. The goal of a code review is to improve the project by making sure that changes are high quality. It also helps us make sure that the contents are easy to understand. That the style is consistent with the overall project. And that we don't forget any important cases.
   3. Code review tools let us comment on someone else's code
2. **Managing Collaboration**
   1. If you're a project maintainer, it's important that you are reply promptly to pull requests and don't let them stagnate. The more time that passes until a pull request gets reviewed, the more likely it is that there's a new commit that causes a conflict when you try to merge in the change.
   2. It's important that you understand any changes you accept.
   3. When it comes to coordinating who does what and when, a common strategy for active software projects is to use an issue tracker.
3. **Issue Tracker**
   1. An issue tracker tells us the tasks that need to be done, the state they're in and who's working on them.
   2. Also used to report bugs by common users.
   3. You can add **Closes #issue\_number\_here** on your commit message while working on the issue. This message will then automatically close the issue identified by the issue no. in the commit message when pushed to the remote branch.
4. **Continuous Integration (CI) system**
   1. Write automated tests to test the code for us and then use a continuous integration or CI system to run those tests automatically.
   2. A continuous integration system will build and test our code every time there's a change. This means that it will run whenever there's a new commit in the main branch of our code. It will also run for any changes that come in through pull request.
   3. Once we have our code automatically built and tested, the next automation step is continuous deployment which is sometimes called **Continuous Delivery (CD)** . Continuous deployment means the new code is deployed often.
5. **Continuous Deployment**
   1. The goal is to avoid roll outs with a lot of changes between two versions of a project and instead do incremental updates with only a few changes at a time. This allows errors to be caught and fixed early.

![Diagram

Description automatically generated]()(Image source: travis-ci.org)

Fig: Integration of Travis with Github Project

**NOTES**

1. Git-scm.com ( SCM – Source Control Management, similar term for VCS)
2. Other VCSs (Ex: Subversion, Mercurial)
3. MinGW64 – Environment in Windows to let us use same command lines as in Linux.
4. IDE- Integrated Development Environments
5. GPL Version 2 License – free license to use and modify
6. .git – Configuration file open with text and .sh file open with bash
7. Git directory – Contains changes
8. Working tree – uncommitted files (current working files)
9. Git project have three areas:
   1. Git directory
   2. Staging Area
   3. Working Tree
10. A file in the directory could be:
    1. Untracked
    2. Tracked
       1. Modified
       2. Staged
       3. Committed
11. Commit without a message is aborted. No commit without message.
12. Good Commit Message :
    1. First line – short description (<50 characters)
    2. Second line – Empty
    3. Follows a paragraph (<72 characters) for detailed description
13. HEAD – points to current commit (could be in different branch, like bookmark)
14. $ clear – to clear screen
15. .gitignore files are used to tell the git tool to intentionally ignore some files in a given Git repository. For example, this can be useful for configuration files or metadata files that a user may not want to check into the master branch.
16. Avoid amending commits that are already public. Good for local commits though.
17. Commit ID : 40 char long string (Hash key generated by SHA1 algorithm)
18. Branch – A pointer to a particular commit (independent line of development)
19. Git Checkout – To check out the latest snapshot for both files and for branches
20. Branch Merge – two different algorithms for this: **fast forward** and **three-way merge**
21. Three-way merge activates when there some kind of divergence between two branches. From their most recent common ancestor, git combines the changes in both the branches. If the changes are in the same part of a file, **merge conflicts** arise.
22. Github, BitBucket, Gitlab are some examples of remote web based repository hosting.
23. .md – Markdown file
24. Remote work order – modify, stage, commit, fetch (automatic merge), push
25. $ git remote -v, Fetch url can be set with READ only using HTTP and Push url can be set with access control given by HTTP or SSH.
26. Remote branches (local copy) – store copy of the data stored in remote repositories.
27. If there’s no Merge conflict, git pull perform **fast forward merge**.
28. git remote update will update all of your branches set to track remote ones, but not merge any changes in.
29. git fetch will update only the branch you're on, but not merge any changes in.
30. git pull will update and merge any remote changes of the current branch you're on.
31. Conflict Markers – highlight the line where conflict arises using >>>>> ======= <<<<<< symbols. Remove these symbols when resolving conflict.
32. CI/ CD Tools – Jenkins, Travis
33. Artifacts -name used to describe any files that are generated as part of the pipeline.
34. Make sure the authorized entities for the test servers are not the same entities authorized to deploy on the production servers.
35. Always have a plan to recover your access in case your pipeline gets compromised.

**GOOD PRACTICES**

1. Always synchronize your branches before starting any work on your own.
2. Avoid having very large changes that modify a lot of different things. Instead, try to make changes as small as possible as long as they're self-contained.
3. Push your changes often and pull before doing any work, you reduce the chances of getting conflict.
4. When working on a big change, it makes sense to have a separate feature branch. This lets you work on new changes while still enabling you to fix bugs in the other branch.
5. Regularly merge changes made on the master branch back onto the feature branch. Thus, you won't end up with a huge number of merge conflicts when the final merge time comes around.
6. If you need to maintain more than one version of a project at the same time, it's common practice to have the latest version of the project in the master branch and a stable version of the project on a separate branch.
7. Merge your changes into the separate branch whenever you declare a stable release. When using these two branches, some bug fixes for the stable version may be done directly on the stable branch if they aren't relevant to the latest version anymore.
8. Don't rebase changes that have been pushed to remote repos.

Whenever we do a rebase, we're rewriting the history of our branch. The old commits get replaced with new commits, so they'll be based on different snapshots than the ones we had before, and they'll have completely different hash sums. This works fine for local changes but can cause a lot of trouble for changes that have been published and downloaded by other collaborators.

Git server will automatically reject pushes that changes branch history.

Don't push the rebase changes to the feature branch, only to the master branch that hadn't seen those changes before.

1. Having good commit messages are important.