Table of Contents

[Intro to Version Control 1](#_Toc515809565)

[What it is 1](#_Toc515809566)

[Purpose/Benefits 1](#_Toc515809567)

[Types 1](#_Toc515809568)

[Local 1](#_Toc515809569)

[Centralized 1](#_Toc515809570)

[Distributed 2](#_Toc515809571)

[Summary 3](#_Toc515809572)

[Version Control Terminology 3](#_Toc515809573)

[Comparing Version Control software and how to use them 4](#_Toc515809574)

[Local 4](#_Toc515809575)

[Centralized 5](#_Toc515809576)

[Distributed 6](#_Toc515809577)

[An Example of a Popular VC: Git and GitHub in Detail 6](#_Toc515809578)

[Installing Git 6](#_Toc515809579)

[Creating repositories 7](#_Toc515809580)

[Clone an existing repository 7](#_Toc515809581)

[Create a new repository 8](#_Toc515809582)

[Adding, Committing 9](#_Toc515809583)

[Finding differences 11](#_Toc515809584)

[Updating your local repository, fetching 12](#_Toc515809585)

[Reverting, cherry-picking 12](#_Toc515809586)

[Branching, merging 12](#_Toc515809587)

[Remotes 14](#_Toc515809588)

[History 15](#_Toc515809589)

[Using GitHub 15](#_Toc515809590)

[Full Version Control Plan 15](#_Toc515809591)

[If a system goes down 15](#_Toc515809592)

[If a version blows up 15](#_Toc515809593)

[If a new software or version is introduced to your company 16](#_Toc515809594)

[Conclusion 16](#_Toc515809595)

Version Control

# Intro to Version Control

## What it is

Version control systems regulate changes so that developers can keep track of their own work and collaborate with others on the same project at the same time.[[1]](#footnote-1)

## Purpose/Benefits

It allows developers to know who made what changes and when so that everything is organized and controlled. At any time, developers can go back and redo a project or file from any point in ‘history’. Additionally, VCSs allows developers to compare different versions and not override other developer’s code too easily. [[2]](#footnote-2)

## Types

### Local

On each PC there is a directory where the files are saved every so often. This is not ideal because it is only available on one computer and it’s very easy to make mistakes.[[3]](#footnote-3)

### Centralized

Figure 1

Centralized VCSs have one central server with all the versions saved on it. Each developer ‘checks-out’ the versions that they need.[[4]](#footnote-4)

Advantages of a centralized system are that there is communication between developers. Everyone can see what others are doing because there is one centralized server. Additionally, administrators have more control over the project. Lastly, a centralized system is easier to work with.

One disadvantage is that if the central server goes down or is compromised, all the data could potentially be lost. [[5]](#footnote-5)

### Distributed

Figure 2

While using distributed VCS, developers download all the versions (the entire history) when they first get the database onto their local machine. To help developers work together, there is usually a central server that all developers can access. That central server is the version kept most up to date and is where all the developers send their code when they have updated it. Each developer can work locally, make changes, and test their work, and then just send it to the central server for the other developers to access.[[6]](#footnote-6)

There are many advantages to using a distributed system. Firstly, they are much faster to use because everything is stored on the developer’s machine. Secondly, developers don’t need to use the internet or be connected to any network.[[7]](#footnote-7) Additionally, if a server goes down, developers can copy their own code back to the server because each developer has all the code. Lastly, developers can connect to more than one repository at once.[[8]](#footnote-8)

There are not many disadvantages to using a distributed system over centralized or local but, in some situations, distributed systems can take up more space.

### Summary

In summary, local is one copy on one machine, centralized is one copy that can be accessed from many machines, and distributed is many copies on many machines.[[9]](#footnote-9)

# Comparing Version Control software and how to use them

## Local

The standard local VCS is RCS. RCS saves just patches so if a developer wants a specific version the computer combines the patches.[[10]](#footnote-10) This takes up less space than saving each version but can take longer because the computer needs to combine all the patches to produce the desired version.

## Centralized

Centralized VCS are like RCS in the way that they store data. They are file based and therefore save each file change using patches. In centralized systems, developers only use <version system> add to add a file the first time. After that, they just commit[[11]](#footnote-11) This is because the files are only saved once (when they are ‘added’) and only patches are saved after that (by ‘commiting.’)

CVS was one of the first centralized VCSs. Developers can ‘rollback’ and completely delete past commits.[[12]](#footnote-12) This is beneficial if a change was made that needs to be undone, but is much more hazardous than other systems because once a developer rolls-back, he can never get back that version. It is removed from the history.

One con of CVS is that it does not allow easy storage of any type of file and metadata[[13]](#footnote-13)

Subversion, also known as SVN, is very commonly used today. One drawback of SVN is what was mentioned earlier, that a developer can’t ‘rollback’ meaning he can change code back but it will always be in the history.[[14]](#footnote-14)

Perversion is a closed-source software. It has many benefits including that it can store all types of files without the user specifying the type and can even store metadata.[[15]](#footnote-15) It allows for easier tracking and has better revision numbering than other centralized systems. Using Perversion, a developer can list branches, delete branches, and if necessary revert changes. Developers also have a copies of the repository that is under control on their personal PC. This is very beneficial but can take up a lot of sometimes unnecessary space on the user’s computer.[[16]](#footnote-16)

## Distributed

Git has become very popular over the last few years. One of its many benefits is that it is open-sourced. Another benefit is that Git never deletes history, it is unique in that it uses pointers and moves the pointers to a new place.

Git takes snapshots of the repository at given times which saves the entire repository, not just patches. This is done at every commit. git add is done whenever developers make changes. Unlike in centralized systems, the command add in Git means that the developer wants a new change to be included in the next commit/ snapshot. git commit produces new snapshots with different hash codes for each new snapshot. In Git, developers can compare different versions of the same file because they are comparing snapshots.

Lastly, when Git sends files, it stores the files as changes, deltas, not as whole files. This minimizes the content being sent over networks thereby maximizing speed and efficiency.[[17]](#footnote-17)

* 1. Mercurial
  2. Bazaar

# An Example of a Popular VC: Git and GitHub in Detail

## Installing Git

Git can be installed in a few ways, I will be using the command prompt which can be accessed in all three installation methods.

Developers can use Git Bash, a version of the command prompt with git already preconfigured.

Git, the actual source code which can we used from the command prompt is another good option.

GitHub Desktop is a graphical interface that works with your local repository to track changes and differences between your local repository and the repository stored on GitHub. It is easy to use and comes preconfigured with a command prompt that has git installed.

When you have a command prompt with git installed open, set up your name and email by using the following,

1. git config--global user.name "your name"
2. git config--global user.email "your email"

## Creating repositories

Once you have git installed, you want to start keeping track of files. You can either clone an existing repository or create a new one.

### Clone an existing repository

GitHub is a website that stores repositories, allows the developers to do all git functions, and much more. Developers use GitHub like a central repository, in that it is the most up to date branch. All developers can access it and update their local repositories from it when necessary. They also merge their own updates to the repository stored on GitHub.

To clone an existing repository, open the command prompt and move to the directory where you want your local repository to exist.

1. cd <local/repository/URL>

Clone the repository by

1. git clone <URL>

You should get back something like the following:

1. Cloning into '<Name of Repository>'...  
   remote:  Counting objects:  97, done.  
   remote:  Compressing objects:  100 % (10 / 10), done.  
   remote:  Total 97 (delta 1), reused 0 (delta 0), pack - reused 87  
    Unpacking objects:  100 % (97 / 97), done.

### Create a new repository

To create a new repository instead cloning, open the command prompt and move to the directory where your new repository is stored. As a reminder you need to use the “cd” prompt.

Then git init to initialize the repository as a git-controlled repository. This will produce a .git folder within your repository.

1. git init

Add and commit all the existing files to your repository. See adding and committing later. (git add . and git commit). Now you should have a version-controlled repository on your local computer.

To clone your new repository to GitHub, open the command prompt and cd into your local repository.

Create a completely empty repository in GitHub.

In the command prompt, add the GitHub remote repository as a remote repository and name it. In this example origin is the name of your new remote repository.

1. git remote add origin git@ github.com: < username > /<repository name>.git

Push your current repository (here master) to the remote repository (called origin)

1. git push origin master

You will be asked for your username and password of your GitHub account. Keep in mind that when entering your password the command prompt will appear as if nothing is being entered. The command prompt will look very similar to this:

1. fatal:  HttpRequestException encountered.  
     An error occurred
2. while sending the request.  
   Username
3. for  'https://github.com': < username >   
   Password
4. for  'https://<username>@github.com':   
   Counting objects:  53,  done.  
   Delta compression using up to 4 threads.  
   Compressing objects:  100 %  (51 / 51),  done.  
   Writing objects:  100 %  (53 / 53),  55.79  KiB  |  5.58  MiB / s,  done.  
   Total  53 (delta 2), reused 0 (delta 0)  
    remote:  Resolving deltas:  100 %  (2 / 2),  done.  
   To https: //github.com/elishevastrauss1/SchoolWork.git  
   \* [new branch]      master -> master

As we mentioned earlier, when cloning and passing files, git compresses them by saving the changes (called Deltas) to maximize speed. As shown in the previous lines that are reiterated below.

1. Delta compression using up to 4 threads.  
   Compressing objects:  100 %  (51 / 51),  done.

## Adding, Committing

Figure 3

There are many steps to change and save changes to a git repository.

First find out what is the status of each file in the repository to find out what was changed.

1. git status

The output will be:

1. On branch master  
    Your branch is up to date with  'origin/master'.  
     
   Changes  not staged
2. for commit:    (use  "git add <file>..."
3. to update what will be committed)  
     (use  "git checkout -- <file>..."
4. to discard changes in working directory)  
     
            modified:    < file name >

The first two lines are saying that your local repository, master is “up to date” with your remote repository, origin. If the other developers changed the remote repository, you would have to update your local repository before committing. See updating your local repository.

Add those files to the pending list to be committed.

1. git add <file name>

or

1. git add .

“git add” is called staging a commit. Those files are now staged and ready for committing. They have not been saved to your local repository and you can still change the files. “git add <file name>” stages just one file, while “git add .” stages all changed files to be committed.

Additionally, you can do git add -p which will allow you to choose which items to commit in a given file. This allows just parts of files to be commited.

Commit the changes to your local repository.

1. git commit

You can add all files and commit at the same time by

1. git commit - a

A message editor will pop-up to write a detailed message about your commit. Save the message and close the editor.

Another option would be to just add the message after the commit command by saying

1. git commit - m "<message>"

This will not open an editor and will just continue as if the editor was opened, saved, and closed.

Once the editor is closed, the command prompt will say something like:

1. [master(root - commit)  < Commit Log Number > ]  < Your Message >   
   40 files changed,  1136 insertions(+)  
    create mode 100644 .classpath  
    create mode 100644 .project  
    create mode 100644 .settings / org.eclipse.jdt.core.prefs  
    create mode 100644  MusicLessons.html  
    create mode 100644  MusicNotes.jpg  
    create mode 100644  StudentLists.txt  
    create mode 100644 bin / dataStructures / LinkList.class

To unstage a file that was already added to the pending list:

1. git reset HEAD < file name >
   * 1. To undo all changes since the last commit

## Finding differences

To list all differences between working directory and your local git repository:

* + 1. git diff <file name>

## Updating your local repository, fetching

* + 1. git remote update
    2. git fetch
    3. git pull

## Reverting, cherry-picking

To undo changes that were not yet committed to the local repository, you can just checkout the specified file from the repository. This will get rid of all the working directory changes which were not committed and there is no way to get them back.

To change back to a previous version, developers need to first find out what the hash of the specific snapshot is. They can find out by doing the command git log or git relog. Both commands display the previous commits and their hash codes. git relog just tells more information about what was done in history.

To revert, users need to copy the log number (hash code) and then git revert <log number>. (You only need to use the first eight numbers of the hash.) This will make a new commit that does the opposite of the specified commit. In the diff, this means that all the ‘+’ become ‘–‘ and visa-versa. This only reverts that specific commit.

Cherry-picking

Going back to a specific snapshot

## Branching

Branches are copies of the repository. Benefits of branching are firstly, to experiment. Branches allow developers to do work that will not affect the master branch or any other work that is being done. Secondly, branches allow developers to do specific tasks separately. This leads to clarity in the code, in the history log, and for all developers.

* + - 1. “Long running topics”

Branching will completely change the developer’s working directory to be identical to the specified branch.

Steps:

To determine what branches exist and what is your current branch:

git branch

To create a new branch:

git branch <new branch name>

To checkout a specific branch that was already created into your working directory:

git checkout <branch>

To make a new branch and switch to it at the same time:

git checkout -b <new branch name>

This is equivalent to git branch <new branch name> and git checkout <branch>

When switching branches, the user’s working directory can’t have any changes that were not added or committed, meaning no pending changes or unstaged changes. To fix this problem, the developer can use the command git stash, which will clean out the working directory and save your unstaged changes.

To get back previously stashed changes use git stash show which will give a list of the branches and files that have stashed changes. Then git stash apply will reapply stashed changes. This is like reverting, which developers need to first see the hash number by seeing the log, and then they can revert the change or apply the stash to the directory.

HEAD is the pointer to the current branch. When the developer commits, the HEAD pointer moves forward in history along with the current branch.

## Merging

Merging is adding your changes from one branch to another branch.

Step:

checkout the branch you want to change.

git merge <branch> will merge the branch listed into your current branch.

Conflicts can occur when trying to merge two files that changed the same area of code. You will be warned that there is a conflict and the file can’t be completely merged. The developer must go to the file and resolve the conflict. It will be found in between “merge conflict markers” (<<<<< and >>>>>). The developer can either open a merge tool by git mergetool that will help you merge the conflict, or he can manually go to the file and resolve the conflict by hand.

## Deleting branches

If the branch was already merged and the developer wants to delete it:

git branch -d <branch name>

If the branch was not merged and the developer wants to forcibly delete it anyways. This can be because the branch is no longer needed or because it failed:

git branch -D <branch name>

## Remotes

A remote is a word that simply means the URL of another server with the specified git repository on it.

When the developer clones, he gets two pointers, master and origin. master is the main local working directory where all the updates get merged to. origin points to where the remote repository was last when the developer cloned or last updated it.

After a developer successfully updates the software and commits it to his local repository, he now needs to push the commits back to the remote to update that repository as well.

There are two scenarios that can occur. Firstly, if the remote’s file has not been changed since you last updated, then the developer can go straight to pushing the commits remotely.

git push <remote url or alias for the remote url> <branch that you’re pushing to in the remote server>

Usually the developer will be pushing his master branch to the origin by the command

git push origin master

This will update the remote repository’s master branch and will move the origin/master pointer to point to where the developer just committed.

When trying to push the commit to the remote, a second situation can occur. If the remote has changed since the developer’s last update, then git will not allow him to push his file changes because they will be overwriting other changes. Therefore, he needs to first fetch the updated repository

git fetch.

Fetching only pulls the changes to the developer’s local repository, not the working directory. Then he needs to merge the updated repository file with what he wants to commit. Usually the command will be

git merge origin/master

Now that the remote’s updated version is in this files history, the developer can push to the remote.

Another option is to use the command, which fetches and then automatically merges with the remote repository to the current branch. This can get confusing sometimes, so it is usually better to first fetch and then merge manually.

## History

git log displays the commit number (also called the sha), who committed each commit, the date of the commit, and the commit message.

git reflog is very similar but displays more information.

## Using GitHub

* + 1. Pull requests

# Full Version Control Plan

## If a system goes down

## If a version blows up

Many times, the developer wants to undo a change he made.

If the changes were not committed yet, the developer can just reset the working directory by checking out the specific files.

If the files were already committed, he has a few options to do.

He can revert the changes, which creates a new commit that does exactly the opposite of the specified commit. If the commit added a line, then the revert will delete that line, and if the commit deleted a line, the revert will add it.

To revert, the developer first checks the log to find the commit that caused the bug. Developers should have written thorough messages which describe the changes made, so that the bug is easy to find. Then the developer reverts that specific commit by

git revert <log number>

Developers can also reset back to a specific commit. This will make a new commit that reverts all the commits until the specified log number. There are two options when resetting, to keep the changes in the working directory after the reset, or completely delete them.

git reset –soft <log number> undoes the changes and saves them in the working directory.

git reset –hard <log number> undoes the changes and does not save them in the working directory.

## If a new software or version is introduced to your company

* + 1. Check all software to make sure it still works
    2. Trainings
       1. Old people
       2. New people

# Conclusion

* 1. Reinforce why we need VC (Card 22)
     1. Track code revisions
     2. Show history
     3. Support multiple users on same code at once
     4. Support branching
     5. Support merging of branches
     6. Reveal conflicts and allow to fix them
     7. Support going back in history

# Glossary

Repository

* 1. A group of files that is under version control
  2. Other older VCSs it means the central code[[18]](#footnote-18), which there is no such thing in distributed systems.

Checkout- checkout a specific branch from central repository

Revert- reverts to initial state and deletes pending changes[[19]](#footnote-19)

Working copy- developers personal copy of the central repository.[[20]](#footnote-20)

Trunk- main branch meaning the most up to date branch where all developers push to and pull from

Branches- divide from the trunk to attempt something new.[[21]](#footnote-21) Branching will completely change your working directory to be identical to your new branch.[[22]](#footnote-22)

Patch- A change done to one version that produces a new version.[[23]](#footnote-23)

Commit-

1. Save new changes to original on your computer[[24]](#footnote-24)
2. CVS- Send changes to central repository[[25]](#footnote-25)

Pull/ Update- update your current branch by ‘pulling’ from the shared central repository. Used when other developers made changes to central repository. Needs to be done before merging changes to central repository so that there is no conflicts.[[26]](#footnote-26) Does fetch and merge together.

Merge- merge working copy to central repository[[27]](#footnote-27)

Fork- branch off the main code to produce entirely new file. Will not be merged back. Now owned by you and does slight or big differences from the original.[[28]](#footnote-28)

Status- new, changed, and deleted files[[29]](#footnote-29)

Log- See who did what[[30]](#footnote-30)



Figure 1 From Pro Git



Figure 2 From Pro Git



Figure 3

1. Source 1 (Card 1) [↑](#footnote-ref-1)
2. 1 (2) [↑](#footnote-ref-2)
3. 1 (4) [↑](#footnote-ref-3)
4. 1 (6) [↑](#footnote-ref-4)
5. 1 (7) [↑](#footnote-ref-5)
6. 11 [↑](#footnote-ref-6)
7. 11 [↑](#footnote-ref-7)
8. 1 (8) [↑](#footnote-ref-8)
9. 3b (13) [↑](#footnote-ref-9)
10. 1 (5) [↑](#footnote-ref-10)
11. 11 [↑](#footnote-ref-11)
12. 9 (26) [↑](#footnote-ref-12)
13. 9 (27) [↑](#footnote-ref-13)
14. 9 (26) [↑](#footnote-ref-14)
15. 9 (27) [↑](#footnote-ref-15)
16. 10 (28) [↑](#footnote-ref-16)
17. 11 [↑](#footnote-ref-17)
18. 3c (16) [↑](#footnote-ref-18)
19. 2 (24) [↑](#footnote-ref-19)
20. 3b (14) [↑](#footnote-ref-20)
21. 3c (16) [↑](#footnote-ref-21)
22. 11 [↑](#footnote-ref-22)
23. 1 (3) [↑](#footnote-ref-23)
24. 3b (14) [↑](#footnote-ref-24)
25. 5 (20) [↑](#footnote-ref-25)
26. 3b (14) [↑](#footnote-ref-26)
27. 3b (14) [↑](#footnote-ref-27)
28. 3b (14) [↑](#footnote-ref-28)
29. 7 (23) [↑](#footnote-ref-29)
30. 2 (24) [↑](#footnote-ref-30)