InFoMM Software Management - Version Control with Git

Martin Robinson

October 5, 2015

Why Use Version Control?

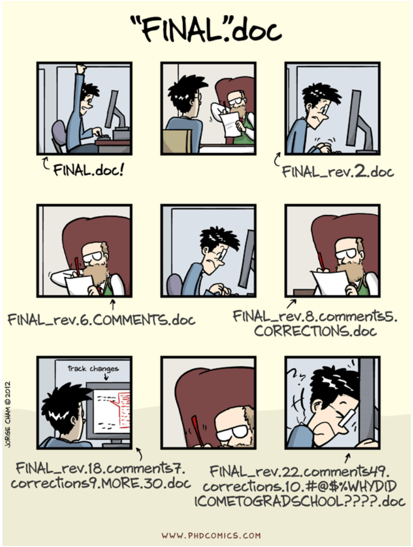

"FINAL".doc
FINAL.doc!
FINAL_rev.2.doc
FINAL_rev.6.COMMENTS.doc
FINAL_rev.8.comments5.CORRECTIONS.doc
track changes
FINAL_rev.18.comments7.corrections9.MORE.30.doc
FINAL_rev.22.comments49.corrections.10.#@$%WHYDID ICOMETOGRADSCHOOL????.doc
JORGE CHAM © 2012
WWW.PHDCOMICS.COM

Figure : "Piled Higher and Deeper" by Jorge Cham, http://www.phdcomics.com

Why Use Version Control?

Version control is better than mailing files back and forth:

- Nothing that is committed to version control is ever lost.
- As we have this record of who made what changes when, we know who to ask if we have questions later on, and, if needed it, revert to a previous version
- the version control system automatically notifies users whenever there's a conflict between one person's work and another's.

Teams are not the only ones to benefit from version control, Version control is the lab notebook of the digital world.

Not just for software: books, papers, small data sets, . . .

Version Control

- Many different VC software packages
  - CVS (1986, 1990 in C)
  - Subversion (2000)
  - Mercurial (2005)
  - Git (2005)
  - . . . many others
- What can you use it for?
  - Text files are best, can see differences between versions
  - Source code is the #1 use case
  - Can also use it for documents or presentations (e.g. latex, beamer, html, markdown)

Git

- Developed in 2005 by the Linux development community for the Linux kernel project
- Features:
  - Branching and merging
  - Fast
  - Distributed
  - Flexible staging area
  - Free and open source
- `http://git-scm.com/`
- `http://git-scm.com/book/en/Getting-Started-Installing-Git`

Installing Git

- Fedora Linux

```
$ yum install git
```

- Debian-based distribution (e.g. Ubuntu)

```
$ apt-get install git
```

- Graphical Mac Git installer at http://sourceforge.net/projects/git-osx-installer/
- MacPorts

```
$ sudo port install git +svn +doc +bash_completion +git
```

- Windows
  - *msysGit* at http://msysgit.github.io. Use provided msysGit shell for command line interface
  - *Cygwin* at http://www.cygwin.com/

Who are you?

Setup your git installation by telling it your name and email

```
$ git config --global user.name "Firstname Lastname"
$ git config --global user.email "example@maths.ox.ac.uk"
```

Creating a Repository

- Initiate an empty local repository. Any files already in the directory need to be added and committed before they included in the repository history.

  ```
  $ git init
  ```

- Clone a remote repository. This can be a directory for a local repository, or a URL for a remote.

  ```
  $ git clone <repo>
  ```

This will download the repository and create a copy on your computer. This is a *separate* local git repository that is linked to the remote

Cloning a repository

for example

```
$ git clone https://server/username/my_git_repo.git
```

This uses the secure `http` protocol, might need to authenticate with a username and password.

```
$ git clone git@server:username/my_git_repo.git
```

This uses the `ssh` protocol, to use this you must have an ssh-key installed on the git server.

Generating an ssh-key pair

- check if you already have one

```
$ ls ~/.ssh/id_rsa.pub
```

- Generate a public and private key using `ssh-keygen`

```
$ ssh-keygen -t rsa -C "username@maths.ox.ac.uk"
```

- show your public key contents using `cat`

```
$ cat ~/.ssh/id_rsa.pub
```

Normally would copy and paste the contents of this key (starting with “ssh-rsa” and ending with “username@maths.ox.ac.uk”) to the git server web-site

Remote Repositories

- The Mathematical Institute runs its own git server.
  - `https://git.maths.ox.ac.uk`
  - Repositories can be private, internal only or public
- When you don't mind your work being public, hosted git websites can be used. These can have large communities and are useful to attract attention from outside the MI
  - GitHub (`https://github.com/`)
  - Bitbucket (`https://bitbucket.org`)

What is a Repository?

- A git repository is a collection of *commits* arranged in a sequential or branching network

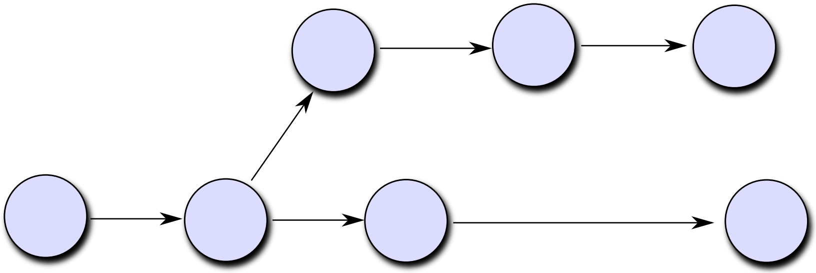


Figure : Series of commits

What is a Repository?

- Each commit contains snapshots of the files that are added, along with timing, author etc. information

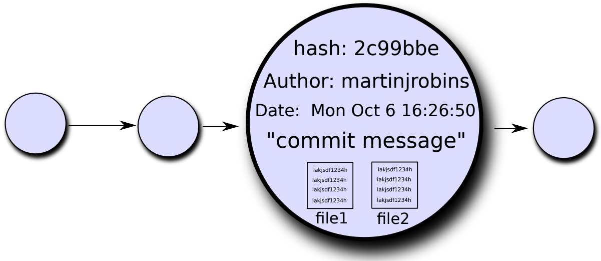
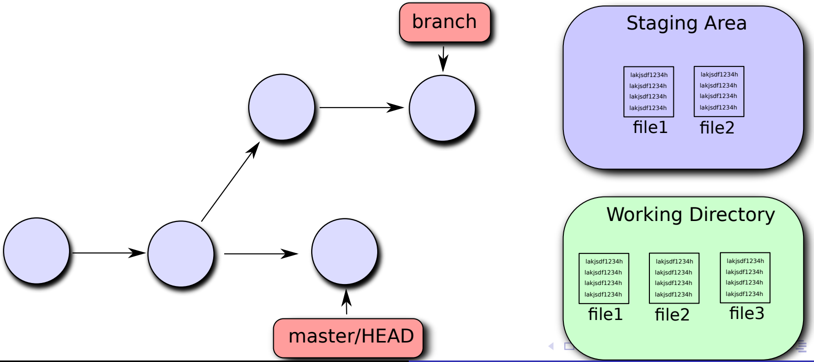


Figure : A commit

What is a Repository?

- As well as the commits, branch pointers show the progress of different branches
- The *working directory* is simply the current set of files in the user's local directory
- The *staging area* is where new edits are added in preparation for creating a new commit



Adding Content

- Use ‘git add’ to add a change in the *working directory* to the *staging area*. All changes or new files need to be added to the staging area before they can be committed (or use the ‘-a’ commit option, see below)

```
$ git add <file>
$ git add <directory>
```

- Commit all changes in the staging area to the project history

```
$ git commit
$ git commit -m "your commit message"
```

- Commit all changes in the working directory

```
$ git commit -a
```

Sending Commits to Remote

- When you want to send your new commits to the remote, use the *push* command.

```
$ git push -u <repo> <branch>
```

- This command pushes the current branch to the remote *repo/branch* branch.
- The *-u* option sets up the current branch to track the *repo/branch* branch.
- Future pushes do not need extra arguments (even if you clone the repository again)

```
$ git push
```

Guided exercise

We will create a new repository on the MI git server `https://git.maths.ox.ac.uk` and clone it to your local computer

1. Install and setup git on your computer (remember to set your name/email)
2. Create an account (or login) to the GitLab server at `https://git.maths.ox.ac.uk`
3. Generate a ssh-key and add it to the GitLab server
4. Create a new remote git repository on the server
5. Clone this new repository to your computer
6. Make some commits and push them back to the server

Git at its simplest

```
$ git clone git@server:usename/my_git_repo.git
$ cd my_git_repo
$ edit file1.m
$ git add file1.m
$ git commit -a -m "this is the first commit"
$ edit file1.m
$ git commit -a -m "this is the second commit"
$ edit file1.m
$ git commit -a -m "this is the third commit"
$ ...
```

Examining the history

- View the repository commit history using the *git log* command

```
    $ git log
commit 0d5ef743e3c0e58ea92154016ed301c80ab03428
Author: martinjrobins <martinjrobins@gmail.com>
Date:   Mon Oct 6 16:47:06 2014 +0100

    this is the third commit

commit a0687f67bc59aadde572ca1395bae2dc1ea462b2
Author: martinjrobins <martinjrobins@gmail.com>
Date:   Mon Oct 6 16:46:48 2014 +0100

    this is the second commit

commit c7245bbb67c23eec849e9bb2097b45e9c4986149
Author: martinjrobins <martinjrobins@gmail.com>
Date:   Mon Oct 6 16:46:33 2014 +0100
```

Examining the history

- For a brief summary use *–oneline*

```
$ git log --oneline
```

- For detailed information of differences between commits use -*p* option

```
$ git log -p
```

- Use the *diff* command to see the differences between the working directory and the staging area.
  - The *–staged* option shows changes between the staging area and the last commit.
  - Use the *HEAD* pointer to see the changes between the working directory and the last commit

```
$ git diff <file>
$ git diff --staged
$ git diff HEAD
```

Examining the staging area

- Use the *git status* command to get details of the staging area and working directory

```
$ git status
```

```
# On branch master
# Changes not staged for commit:
#   (use "git add <file>..." to update what will be committ
#   (use "git checkout -- <file>..." to discard changes in
#
#   modified:   file1.m
#
no changes added to commit (use "git add" and/or "git comm
```

Undoing Changes

- Reset (not yet added to staging area) to the last committed version:

```
$ git checkout <file>
```

- You can amend the current commit (e.g. change commit message, commit new changes etc)

```
$ git commit --amend
```

- You can remove a file from the staging area (i.e. after using *git add*)

```
$git reset HEAD <file>
```

Altering your history - Reset versus Revert

- You can reset your history (soft reset) and optionally your working directory (hard reset) to a specified :

```
$ git reset <commit>
$ git reset --hard <commit>
```

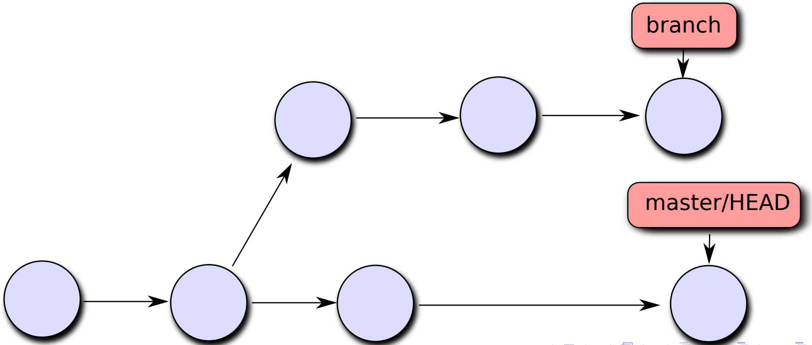
- You can remove a specified commit from the history (a new commit is made with the necessary changes)

```
$ git revert <commit>
```

Altering your history is best avoided unless you know what you are doing.

Branching

- Branching can be used to:
  - create/try out a new feature
  - provide conflict-free parallel editing for multiple team members
  - separate editing into “stable” and “in development” branches
- A branch is simply a pointer to a commit. The current branch is pointed to by *HEAD*



Branching

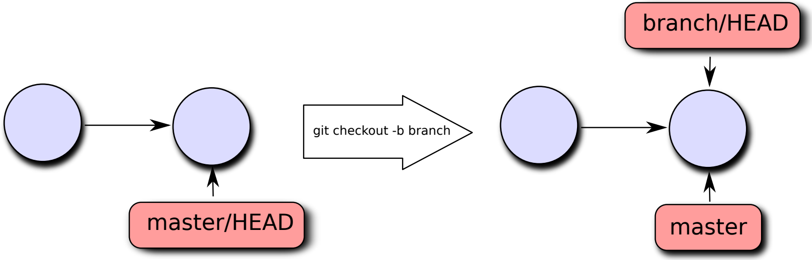


Figure : Creating a new branch

- You can create a new branch and switch to it using the *checkout* command

```
$ git checkout -b <branch>
```

Branching

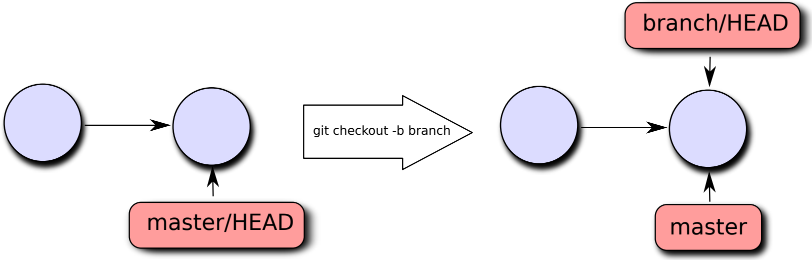


Figure : Creating a new branch

- This is shorthand for

```
$ git branch <branch>
$ git checkout <branch>
```

Branching

- Make a few new edits to your new branch

```
$ edit file1.m
$ commit -a -m "added wow new feature"
$ edit file1.m
$ commit -a -m "fixed bugs in new feature"
```

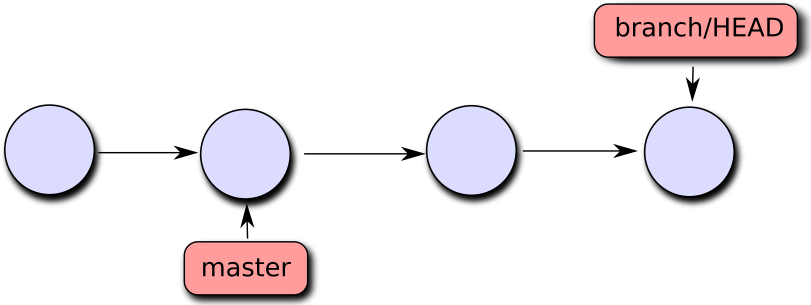
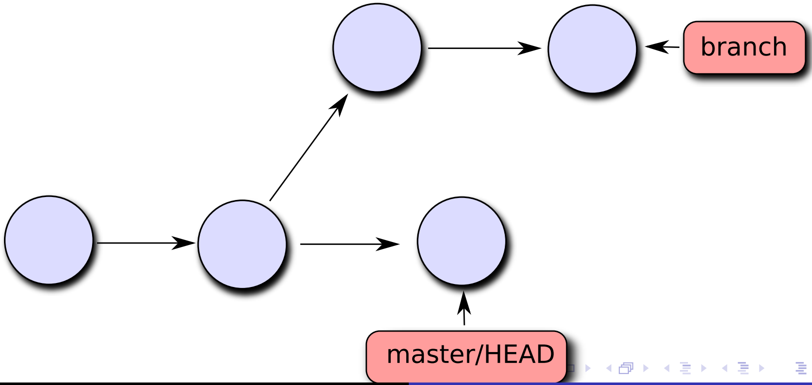


Figure : new edits to branch

Branching

- You might need to got back to the *master* branch to make some corrections

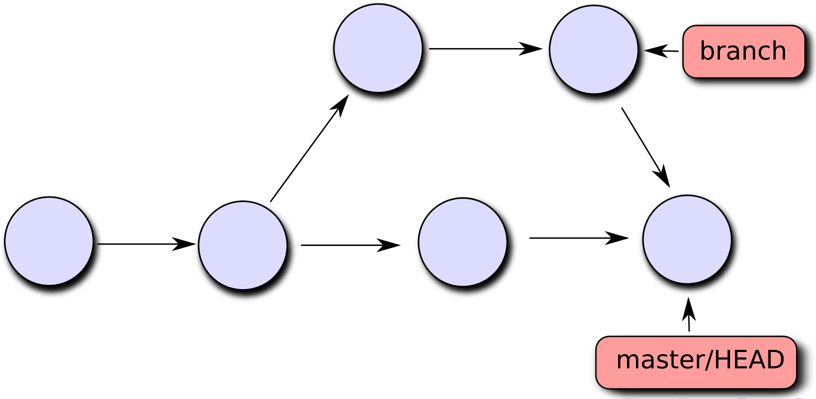
```
$ git checkout master
$ edit file1.m
$ commit -a -m "fixed a major bug"
```



Merging

- Now we have two separate branches, but what if we want to merge them together?

```
$ git checkout master
$ git merge branch
```



Merge Conflict

- If there are conflicting edits to *file1.m* in *master* and *branch*, you might get an error message like so:

```
$ git merge branch
Auto-merging file1.m
CONFLICT (content): Merge conflict in file1.m
Automatic merge failed; fix conflicts and then commit the r
```

Resolving Merge Conflict

- If you open *file1.m*, you will see standard conflict-resolution markers like this:

```
<<<<<<< HEAD
This is the new line in master
=======
This is the new line in branch
>>>>>>> branch
```

- If you want to see which files are still unmerged at any point, you can use *git status* to see the current state of the merge
- Finally, commit the results of the merge

```
$ git commit -a -m "merged branch into master"
```

Adding Remote Repositories

- Inspect currently added remote repositories (if you cloned the repo initially, this will be listed as *origin*)

```
$ git remote -v
```

- You can add a remote repository with a convenient name

```
git remote add <name> <url>
```

- Remote can be a https (`https://host/path/to/repo.git`) or ssh (user@host:/path/to/repo.git) address.

Getting Commits from Remote

- If the remote repository has new commits that are not on your local computer, use the *pull* command to get these and automatically merge them

```
$ git pull
```

- The above command combines a *fetch* and *merge* command, see next slide for details. . .

Getting Commits from Remote

- You can fetch the contents of the remote repository using the *fetch* command.
- The *branch* command will then show both the local and remote branches

```
$ git fetch
$ git branch -a
* master
remotes/origin/master
```

- You can then merge any changes the remote branch (*origin/master*) into the current local branch (*master*)

```
$ git merge origin/master
```

More info

- If there is any command you are unclear about, you can use *git -h* to get more information. Or simply google it...
- Further tutorial can be found online at:
    - Git documentation and book (`http://git-scm.com/doc`)
    - Atlassian tutorials (`https://www.atlassian.com/git/tutorials`)
    - Software Carpentry Foundation (`http://software-carpentry.org/`)
- Acknowledgements: material for this lecture modified from links above.
    - Git book: Creative Commons Attribution-NonCommercial-ShareAlike 3.0
    - Atlassian tutorials: Creative Commons Attribution 2.5 Australia License.
    - Software Carpentry: Creative Commons Attribution Licence (v4.0)

Hands-on practice

- Best way to learn Git is to use it
- Clone these lecture notes at `https://git.maths.ox.ac.uk/robinsonm/infomm-gitlecture.git`.
  - Also includes a handy Git cheat sheet from GitHub.
- Go through the included git exercise sheet (`git_exercises.pdf`)