Tools For C/C++ Developers

Preface:

This document covers the tools that I personally use when developing C/C++ programs. These tools include the GNU C Compiler (GCC), the GNU Debugger (GDB), make and cmake, libcheck, valgrind, and Doxygen. Git is covered in its own document.

Intro:

I’ve said before that knowing a programming language is only half of what makes a good developer. Knowing your tools is equally as important a skill to have. We use tools in our everyday to increase productivity and project maintenance.

The make Command:

This section will be dedicated to GNU Makefiles. Most programming languages have some sort of tooling for build automation. The program most commonly used to automate builds in C is known as make. make is a command in Linux created by Richard Stallman and maintained by the Free Software Foundation (FSF). It searches for a special file known as a Makefile which contains instructions on how to build our executables. Something that makes make unique is that it utilizes timestamps to detect which portions of the program need to be rebuilt, and which don’t. The GNU organization already has great documentation on make which can be found here [https://www.gnu.org/software/make/manual/make.html#toc-Overview-of-make](https://www.gnu.org/software/make/manual/make.html" \l "toc-Overview-of-make). While I encourage you to read through this documentation, it can be a bit daunting if you’re just writing your first Makefile. My aim is to give you a general overview of how we can use make; your experience in the field will fill in the missing gaps as you use make more and more.   
  
Rules:

The idea behind Makefiles is that we write out a series of “rules” which are used to ensure certain conditions are met as we build. Some rules are pre-requisites for other rules. This creates a sort of hierarchy which appears complex on the surface but is really just a sequence of simple rules tied together to form a finished product. Rules follow a special formula so that they can be recognized by make. The general layout of a rule is as follows:

target ... : prerequisites ...

recipe

...

...

Let’s break down these three peices that make up our rule:

**target:** The target is the desired output of processing the rule. In most cases, the target is a file which is created by running certain command line utilities to obtain the desired effect, though the target can also just be a nametag used to specify the rule that we want make to run.  
  
**prerequisites:** The prerequisites are conditions that must be met in order for our rule to be executed. If the prerequisites are not met, the rule cannot be satisfied. Prerequisites are also usually files which we expect to exist in order for our rule to be executed, but may also be other targets.

**recipe:** The recipe is the actual command that we want make to run. There is nothing special about what commands can be put in our recipe; make will execute whatever we add in our recipe in a shell as if we were actually entering those commands into our terminal.

Variables:

As you may expect, make is capable of keeping track of variables. Variables in make are very similar to variables in shell scripts. They are essentially expanded at runtime so that if we assign an entire command to a variable, that command will be executed on the spot. Variables are generally always upper-case as a convention, and they are assigned with the assignment operator (=). In order to expand a variable in make, we use the special operator ‘$’, followed by the variable name enclosed in brackets eg. $(VAR). Again, this is very similar to shell scripts.

Although we are able to define variables however we want in make, there are certain convensions which nearly all programs abide by. Specifically, there are certain variable names that ought to be used for readability. Here is a list of some conventional variable names:

**CC:** In case you needed a refresher, cc stands for c-compiler. In Unix, the compiler was simply called cc, but as you know, Linux was created for the GNU project, and with GNU came the GCC (GNU C-Compiler). Therefore, you would set CC = gcc on a Linux machine.

**CFLAGS:** The CFLAGS variable is used for flags that we wish to add to our compiler. These include the flags that we covered previously, such as -O, -g, -Wflags, etc.

**LDFLAGS:** In Unix, ld stood for “loader”, which at the time, was synonymous with “linker” ie. they were one in the same. Nowadays, the loader is a separate program from the linker, so this name is less relevant. None-the-less, LDFLAGS is used to describe flags that we want to pass to our linker which is usually just whatever libraries we want to link with the -l option.

**OBJS:** A list of object files that are a dependency for the final executable.

**BINS:** The BINS variable is typically used as a list of binaries that are to be made by the “all” target. We will look at a few standard target names momentarily.

**LIBS:** The LIBS variable usually contains the path(s) to a alternative directories that contain important libraries for creating our program that are not already in our system’s PATH variable. These directories can be searched with the -L option in our CFLAGS eg. CFLAGS = -L $(LIBS) where LIBS = /project/lib:/project/other/lib

The Default Goal:

If you run make without specifying a target, make searches for the first target who’s name does not start with ‘.’. This is called the default goal. Usually, we call the default goal ‘all’, which means run all other rules in the Makefile. The behavior of the default goal can be altered using an option when running make from the command line, or by overriding the .DEFAULT\_GOAL special variable. If the default goal depends on other targets, then those get ran first.  
  
Implicit Rules:

PHONY Targets:

Variable Assignment:

Make parses the Makefile in two separate runs (known as first phase and second phase). The first phase is when make reads all of the included Makefiles and internalizes all the variables, their values, implicit/explicit rules, and builds a dependency graph of all the targets and their prerequisites. The second phase uses all of this interalized data to determine which targets need to be updated and run and in what order.

One of the most confusing aspects of make is how it expands variables. There are different assignment operators which determine how the variable will be expanded. There are essentially three options for expansion. Either the variable expansion is immediate (the variable is evalutated during the first phase), deferred (the expansion is either delayed until the variable is referenced by an immediate variable, or until the second phase if it does not get referenced), or it is deferred or immediate (deferred under certain circumstances or immediate under others). Here is a list of the operators:

immediate = deferred

immediate ?= deferred

immediate := immediate

immediate ::= immediate

immediate += deferred or immediate

immediate != immediate

These are each unique, but their use-case is not inherently obvious. The = operator is the most basic. It simply defers the expansion of the variable. The ?= operator is the same as =, but only assigns the value to the variable if the variable does not already contain a value (the = operator overwrites previously assigned values). The := and ::= (simple variable operators) are expanded before variable assignment even occurs. If a variable is assigned with := or ::= and then later reassigned with +=, then it will be immediately expanded. Otherwise, if we only assign a variable with the += operator without previously assigning it with a simple operator, it will be deferred. The append operator (+=) will append a value to a variable. Last but not least, the shell operator (!=) is immediately evaluated and then passed to the shell. The result of the shell is stored in the variable, and the variable becomes a simple variable. The man page for make has additional info on these operators if you need additional guidance.

GNU Debugger (GDB):

The GNU debugger, a.k.a. GDB, is a multi-language debugger maintained by the FSF. GDB currently supports the following languages: C, C++, D, Go, Objective C, OpenCL C, Fortran, Pascal, Rust, Modula-2, and Ada. GDB allows for local and remote debugging over TCP/IP. Most Linux distributions ship GDB with a Terminal User Interface (TUI) option available, which is a bit more visually pleasing than the regular terminal print outs. There also exist 3rd party GUIs for GDB which are not maintained by the FSF.

GDB’s philosophy is to be entirely unintrusive during program execution until the moment that a breakpoint is reached. At this point, program execution is halted and GDB takes over the process’ control flow. GDB can do 4 primary things (an excerpt from their website):

* Start your program, specifying anything that might affect its behavior.
* Make your programs stop on specified conditions.
* Examine what has happened when your program has stopped.
* Change things in your program, so you can experiment with correcting the side effects of one bug and go on to learn about another.

One thing I should mention that wasn’t mentioned in that excerpt is that GDB can analyse core dumps which are produced when a program crashes. Core dumps contain the process’ state up until the moment that the crash occurred, meaning that although we can’t debug the process past the point it crashed, we can debug up until that moment to see what went wrong, which can be extremely helpful in some situations.

Running GDB:

At a basic level, to start GDB, simply run the gdb command from the command line followed by the location of your program. For instance, if my program’s executable is located within the bin directory, I would run gdb bin/program. In order to pass arguments to the program as you normally would from the command line, we give GDB the -o or --option flag followed by a whitespace delimited set of arguments to pass to the program. If you want to start GDB in TUI mode, you can pass GDB the --tui option. The --pid=pid option will tell GDB to attach to an already running process where pid is the process ID of the running process you want to attach to. GDB offers quite a few other interesting options which can be found in its man pages, such as baudrate for serial debugging, the timeout value for remote debugging, the TTY to use for I/O, etc. Note that all of the flags we’ve covered thus far can also be specified *after* GDB has started if you prefer to just boot up gdb and then input setup conditions later.

Debugging With GDB:

Let’s now cover how we actually debug using GDB. If you just ran gdb by itself without any command line arguments, you’ll be greeted with some text regarding the license and version information, followed by an input prompt. As stated above the prompt, we can use the help command to get more information about the commands GDB provides. Note that there are really a ton of internal commands that GDB provides. Don’t be too overwhelmed by this – there are only a few commands that are really important to us, but there are many more for advanced users. Because there are so many commands though, typing help will display a list of categories of commands. In order to see the commands of a particular category, just type help <category>. For example, help running will reveal all the commands related to running the program. You can then inquire more about each individual command by doing help <command>.   
  
 You might notice that GDB has a lot of aliases and shorthands for commands (these are listed in the help output for each command). For example, the continue command can be entered as “continue”, or as “fg”, or as just “c”. GDB is also pretty good at inferring the command that your trying to use. For instance, if I were to type “cont”, GDB can infer that I mean to run the continue command, even though its not an official alias. GDB even provides auto completion if you press tab, so typing “cont” followed by tab will complete to continue.

If you did not specify the program when running GDB from the command line, we can specify it using the file command e.g. file program. This will load any ELF symbols and debug symbols (something I’ll explain in a bit) that are present. In order to run the program, just use the run command. Alternatively, we can use the start command, which works like the run command, but sets a breakpoint at main. The continue command will continue program execution until the program either the next breakpoint is reached or the program terminates. In order to single step the program, we use the step command (alias s). The s command sort of works like a “step into” in normal IDEs, as it will go into other functions. If you want to do more of a “step over”, we use the next command (alias n). This works like s, but steps over function calls. I should mention that for both step and next, we can provide an optional number of lines to step by e.g. step 5. To “step out” of a function, use the finish command (alias fin). This will continue executing the rest of the function that we’re currently in and halt execution after it returns.

In order to set a breakpoint, use the break command (alias b). This command can optionally accept as parameters a file, function, and line. For example I can do b file.c:80 do break at line 80 in the file file.c. I can also just specify a function name if I want e.g. b foo, which will break at the start of function foo(). Use the info breakpoints command to list out the breakpoints you’ve set. These will each have an associated number. We can temporarily disable or enable a breakpoint by doing disable/enable <breakpoint\_no>. If you plan on reusing breakpoints, you can save them via save breakpoints <output\_file>. To print the contents of a variable, use the print (alias p) command, followed by the variable name. The print command can accept the /p, /t, or /x switches to print the value as decimal, binary, or hexadecimal respectively. To set the value of a variable, use the set command. For example, to set an integer n to 9, you would do set n = 9. To list all local variables and their values, we can use info locals. Likewise, info args will display all of the outer function’s arguments and their values. Watches work a bit differently in GDB than they do in other IDEs. Rather than displaying a window with the values of each watchpoint, GDB will halt execution in the case that a watch variable’s value changes, or in the case of a watch statement, when the statement becomes true. So for example, I can say watch a == 4, and the program will halt execution when the variable a becomes 4. Alternatively, watch a will halt execution whenever a becomes any value other than its current value. GDB will print the old value and new value for both variable watches and expression watches. Finally, to quit GDB, just use the quit command.

Advanced Commands:

The basic commands are all that you’ll really need to debug basic programs. More advanced programs and bugs may require using some more advanced commands, however. Two nice ones I sometimes use are info threads and info registers (alias info r). The info threads command prints out the thread ID and stack frame that is being executed by each thread in the program. Likewise, the info registers command displays information about the current state of each of the processor’s registers. The stepi (alias si) and nexti (alias ni) commands allow us to step by single assembly instructions similar to how the normal step and next commands work. These work best with the assembly layout with GDB’s TUI. Speaking of which, let me explain the tui layout command. The tui layout command (alias lay) is capable of displaying an assembly view, a source view, and a split view which shows both. We can cycle between views using tui layout next/prev or just lay n/p. We can also specify the exact layout option. The assembly layout is called asm, source is src, and split is split. For example, lay asm will switch to the assembly view. Sometimes the TUI will get visually corrupted, so to correct this, use the refresh command. If you prefer to not use the tui, the other option is to print out the disassembly using the disassemble command. This command accepts the /m, /r, or /s options (though /m is deprecated in favor of /s). The only difference between /r and /s is that /r will also show the hexadecimal op codes for assembly instructions. We can specify whether to use AT&T or Intel syntax when disassembling by first entering set disassembly-flavor att/intel. By default GDB will use AT&T syntax. To print a backtrace of the programs call stack, we can use the backtrace command (alias bt). This will display each stack frame that has been visited up until the current stack frame. The backtrace command also takes full as an optional parameter, which will display local variables and their states for each stack frame as well. The up and down commands are similar to backtrace. If in TUI mode, up will move the cursor to the calling stack frame and print it. Once we’ve moved up at least one level, we can use down to return to where we were, since up and down are relative to the stack frame that we’ve temporarily moved to. These command don’t execute any code, they just move and print the stack frames within the backtrace. An optional argument can be passed to either command specifying how many frames to move.

Debugging Core Dumps:

Sometimes your program will crash with a SIGSEGV signal or something of similar effect. Normally your system does not produce a core dump for this, otherwise your system might accumulate a bunch of files that a normal user would not cleanup, thus wasting a lot of disk space. Linux has a command to specify the maximum size of a core dump file, which is set to 0 by default, meaning that no core dumps will be generated. To alleviate this, we can run ulimit -S -c unlimited. This will set the size to be unlimited, effectively enabling core dumps. All you must do is re-run your crashing program after this and it should output a file called core. This is an ELF file which contains information about the program’s state before the crash occurs. In order to debug this file using GDB, we can use the -c flag from the command line followed by the core file, or GDB’s internal core command (if your file is called core, then in GDB you’d load it as core core). You still need to specify the executable either from the command line as normal or using GDB’s file command. Using the core file means you will not have to pass in program arguments as usual, since the core file retains that information from the time you ran it. If you ran GDB from the command line it will run the where command, which shows where the crash occurred and what caused it, otherwise you can run the where command manually. Since the core dump takes you to the crash site, you can also run backtrace full off the rip instead of having to step through the program manually to get to the crash. Core dumps are not useful to the average programmer, but in production when supporting a customer it can be very helpful to request that the customer send a core dump for you to analyse instead of trying to reproduce the issue yourself. Make sure to run ulimit -S -c 0 to disable core dumps when you’re done.

Remote Debugging:

Remote debugging is achieved with the target remote command. To make a connection with the remote device, we simply enter target remote <url>:<port>. For example: target remote [www.example.com](http://www.example.com/):9898. Alternatively, rather than using ports, which can be a bit of a pain, we can use stdio pipes through ssh. This would look something like the following:

target remote | ssh -T [www.example.com](http://www.example.com/) program - --attach <pid>

The single dash ‘-’ replaces the port number and specifies that we’re using stdio pipes rather than TCP port 9898.

Understanding How GDB Works:

I mentioned earlier about debugging symbols and how they’re required for the TUI mode of GDB. Actually, you won’t get much done even in the normal mode of GDB without debug symbols either. There are two types of symbols that your program will contain. These are the regular ELF symbols such as main, for example, and debug symbols, which are included as a .debug\_info segment within the ELF file. The debug symbols can come in various formats. On Linux systems, it will most likely be in DWARF format, but other formats include stabs, COFF and XCOFF. The readelf command has a -w flag for displaying debug information. For example, we can do readelf -wi program to see the .debug\_info segment’s DWARF information. DWARF format is meant to be extensible and work across different architectures with different languages. DWARF works best with C-like languages, which explains why GDB supports primarily C-like languages, or languages that compile to C as an intermediary step. For example C++, D, and Objective C are all very similar to C. Ada, Rust, and Go all compile to C as an intermediary step, which is why GDB works for those languages. You’ll see if you run the readelf command all of the DWARF types (which each begin with DW\_AT). Each type stores information about a given symbol such as its name, byte size, type, and declaration location (pretty neat). The -g flag in GCC produces debug information in whichever debug format is native to the system. There are other options for debug symbols though, such as -ggdb, and -gdwarf. The -ggdb option produces debug information in whichever format is most expressive (and therefore best for debugging with GDB). The -gdwarf flag attempts to add debug symbols specifically in DWARF format if it’s available on the system.

Version Control With Git

Version Control Systems (VCS):

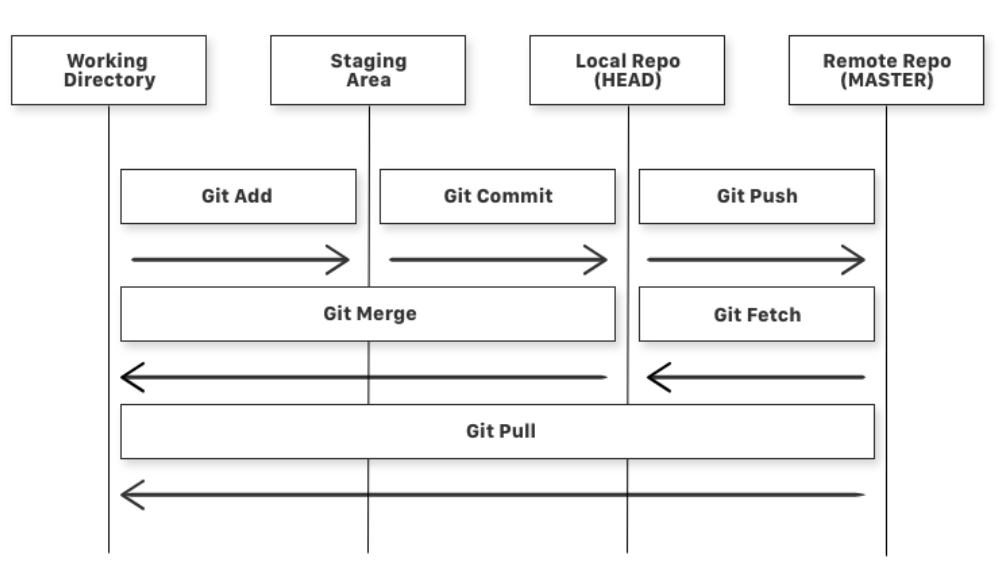
If you are unfamiliar with version control, then buckle up, because we have a lot to discuss! Version control, sometimes called source control, is the practice of tracking and managing changes to code. There are various version control tools which monitor and record changes to our code. This careful monitoring allows us to do very neat things, such as go back in time to a previous commit, create forks of our code into separate projects, or branch our code to work on a new feature without affecting our other branches. Arguably, the most well-known version control systems are Revision Control System (RCS) managed by the GNU project, Subversion (SVN), and Git (created by Linus Torvalds, the creator of the Linux kernel). The order in which I listed those three happen to be the order in which they were created timeline-wise. Git is the most recent, and well known, which is why we will be using it. As a developer, I’m sure you have at least heard of GitHub. Perhaps erroneously you assumed that GitHub and Git were the same thing. For this reason, we will be splitting up this section into 2 parts. The first part will focus on local source control using Git (the bulk of the material), and then later we will look into GitHub, which just happens to be one of many upstream servers for the front-end stuff. Git also encourages collaboration on projects since it allows us to merge changes made by our peers.

What Makes Git Special?:

Perhaps you’re wondering about some of the reasons that Git took off in popularity over its predecessors. Well one nice thing about it is that it is a Distributed Version Control System. This means that Git does not rely on a central server to store all the versions of a project’s files. Instead, Git will store the metadata locally in a hidden folder (called .git). When the code is cloned by someone else, they are downloading the source files, and also that .git folder containing all past changes that have been committed.

Understanding the Git Workflow:

It is imperative that you grasp this next concept, because it will ultimately influence the way that we interact with Git. I am totally plagiarizing these notes from freeCodeCamp’s website, including the following image, so check out their site as well. Here we see the four fundamental elements of the Git workflow (Working Directory, Staging Area, Local Repository, and Remote Repository):



Let’s quickly give a briefing of each. The working directory is your project directory. We run the command git init to initialize our project directory. This creates the .git folder I mentioned earlier and prepares it to be tracked by git. The staging area is invisible to us (well it’s actually a file that we can open, but we don’t typically access it directly). It let’s Git know which files we actually want to track. Running git status gives us an overview of which files are and aren’t staged. One other thing to note about the staging area is that files which belong to it will have their changes recorded. If a file is being tracked (i.e. it belongs to the staging area) and we modify it, it will immediately become committable, even if it had already been committed. In the next “phase” (the HEAD) we have files which have been committed. When files are committed, we essentially create a save state or backup point that we can revert to if needed. When a file has been committed, it’s state is actively preserved, unlike files which are staged, which have modifications tracked, but not preserved. Finally, we have MASTER, which is the remote repository (repo for short). This could be GitHub, GitLab, Savannah, etc. From the remote repo, others can clone (download) our code to their own local repo.

What’s very nice about the graph above is that it shows us which git subcommands correspond to which action. Here is a summary of each:

**git add:** This command adds files to the staging area. Typically you need only do git add \* to add all files.

**git commit:** Pretty self-explanatory, the git commit subcommand takes all of our staged/tracked files and commits them. This command is almost always used in conjunction with the -m option which specifies the commit message (all commits require a commit message i.e. a summary of why the commit is being created).

**git push:** Git push pushes our committed files to the remote repo. Usually you would do something like git push -u origin master, where the -u means upstream i.e. MASTER (in our case the master’s name is origin), followed by the name of our local branch (in this case, master). Kind of confusing that our local branch is named master and not the remote repo, but that’s just how it is…

**git fetch:** This one often confuses newcomers. Git fetch just pulls files that exist in the remote repo to the local repo. If files exist on the remote repo that don’t exist on the local repo, you will not be able to push your changes until the local repo is up to date with the remote. You can either do git fetch, or git --force push to overwrite the remote repo with the contents of the local one.

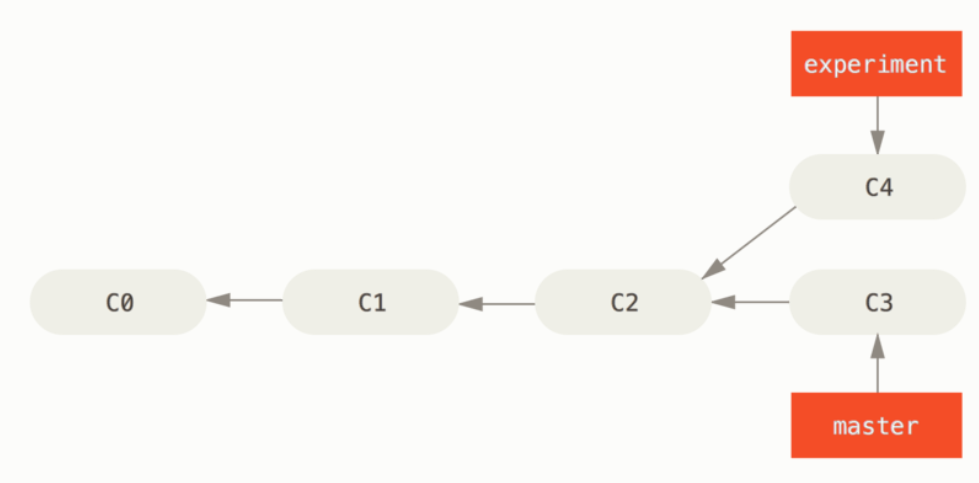
**git merge:** This one is similar to git fetch, except that git merge pulls files into our working directory. Git merge will fail if a file from the local repo (HEAD) has changed that do not exist in the working directory.

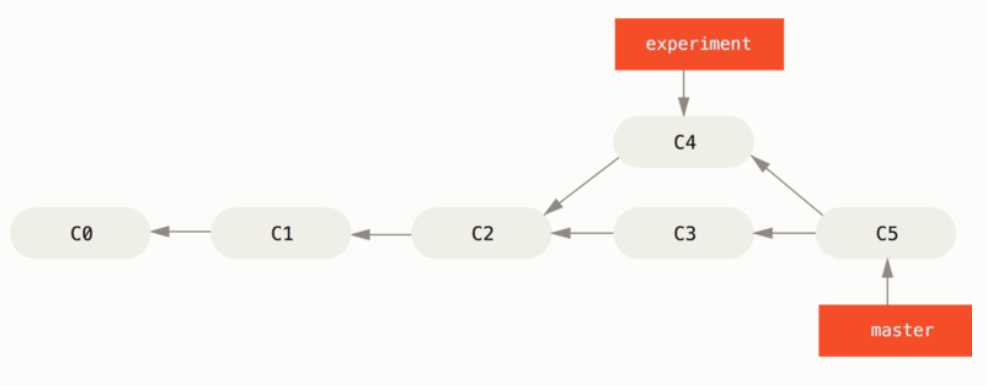
**git pull:** Git pull is equivalent to doing git fetch, followed by git pull. This will essentially pull files that exist on the remote repo into both HEAD and the working directory.

Git Branches:

An important feature of Git is its ability to create branches. Recall that we can create commits which are just snapshots of a period in time where our file contents are preserved. A Git branch is essentially a pointer to one of these commits which can be moved to other commits. The default branch is called master. You can reassure yourself of this fact by entering the subcommand git branch, which will display a list of the created branches with an asterisk next to the current branch. This should look like \*master by default. As we create new commits, the HEAD keeps up to date with the latest one. Once again, this can be confirmed by running git log --oneline. Running this command will show you which branch HEAD is pointing to, as well as which commit it is currently on. For example, it might look something like the following: 4960843 (HEAD -> master) Initial commit. Here we see the commit hash, 4960843. This is a unique identifier given to each commit. We see that HEAD indeed points to the master branch, as that is the one we are currently on. Finally, we see the commit message created using git commit.

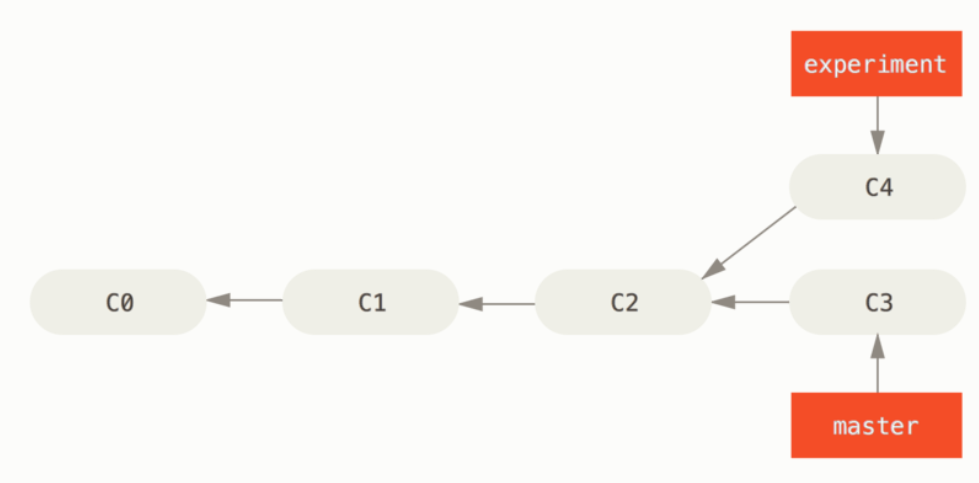
So when should we create a new branch? Well, typically you create a new branch when you want to add a new feature to your program that you don’t want to affect your main branch. Typically, the main branch is referred to as the “mainline” branch, and then you may have a “dev” branch for development, and possibly branches for specific features or versions that are known to be working well. In order to create a new branch, we use the git branch subcommand once again, except this time, supply it with the name of the branch you want to create. For example, git branch new. Once this is created, running git branch by itself will display that we now have a branch called master, and a branch called new. Note that the asterisk is still on master, meaning that HEAD is still looking at master. In order to switch over to our new branch, we use the checkout subcommand e.g. git checkout new. Now, when we run git log --oneline, we see a bit of a change. The log now appears as 4960843 (HEAD -> new, master) Initial commit. This indicates that HEAD points to new, and then the comma followed by master indicates that new was branched off from master. From our new branch, we can continue development and rest assured that master will remain at the commit that we branched from.   
  
 Now let’s say that we finish developing a feature in our feature branch, and we now want to get the mainline branch up to date with the feature branch so that it will implement the changes. There are two ways that we can do this: git merge, and git rebase. Git merge is the easier of the two. It will create an intermediary commit which both the mainline and feature branch can agree to join up at. Here are some illustrations:





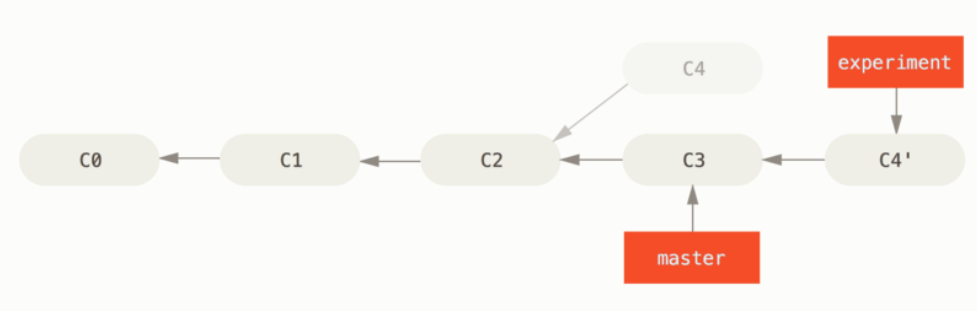
In the first image, we created an experiment branch from commit 2 (C2). We then did some work on both the experiment branch, and the master branch, and created another commit on both (C3 and C4). In order to get these up to date, we create a new intermediary commit C5 using git merge.

The other method of getting these up to date is using git rebase. Rebase functions a bit differently. Rather than creating a new commit which joins the two branches, we take all of the changes that were made on one of the branches and replay them on another branch as if they had occurred on that branch in the first place. Let’s look at how this works using the same starting illustration.



$ git checkout experiment

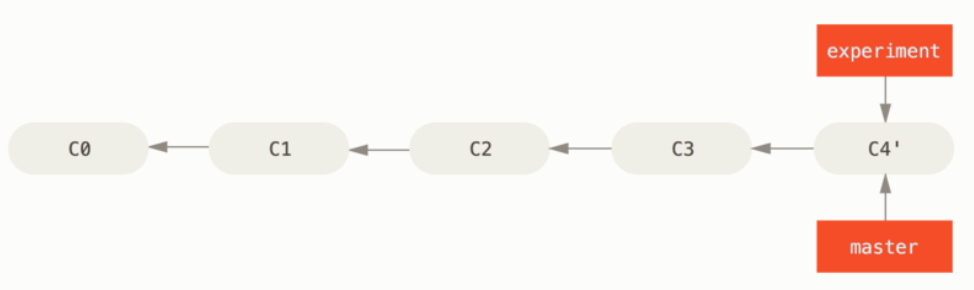
$ git rebase master



Starting from the first image, we have the same scenario. We then checkout to the experiment branch, so HEAD will be pointing to C4. Then we rebase master. What this does is select master as a starting point for replaying the changes made to experiment after it diverged from master. Note though, that master is still behind. We can fix this by using what’s called a fast-forward merge:

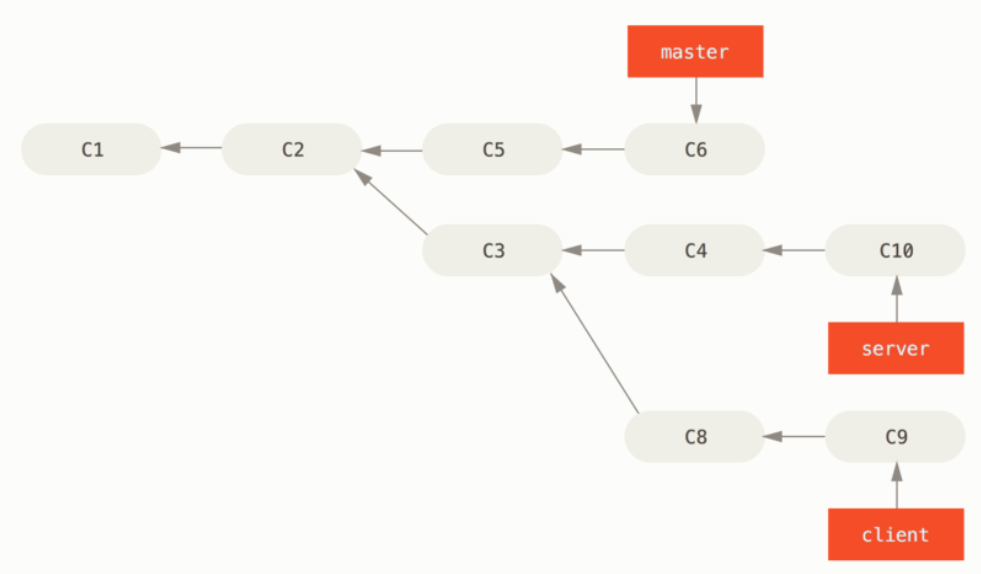
$ git checkout master

$ git merge experiment



This puts us up to date with experiment. I’ve heard the argument that rebase is evil because it deletes history and modifies it to look as if the changes actually occurred in that order even though they didn’t. On the other hand, I’ve heard the argument that rebase makes for a cleaner history and is therefore easier to read. Personally, I haven’t used git enough to have an opinion on this, but I just wanted to bring those two sides to light.

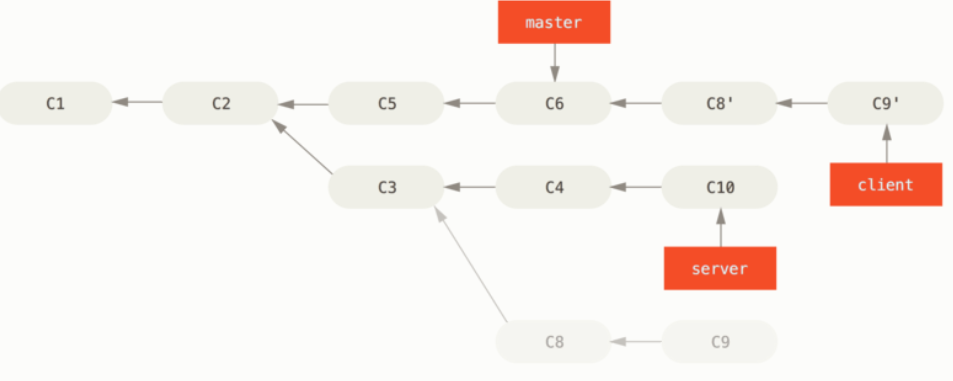
Here is another more complicated example (I know complicated is probably the last thing you want right now, but it is good to look at more realistic examples).



Here we have a more complicated commit history with three active branches. I’m taking this example from Git’s official documentation, so I’ll explain the scenario that they’ve invented. You, the user, have created a new branch for a server application. You also created a branch from that for the client code since they are similar. Now you want to merge the client code into mainline for release, but you want to hold off on the server code until it’s tested further. We can take the changes made for the client code (C8 and C9) and replay them on master using the –onto option of git rebase e.g.

$ git rebase --onto master server client

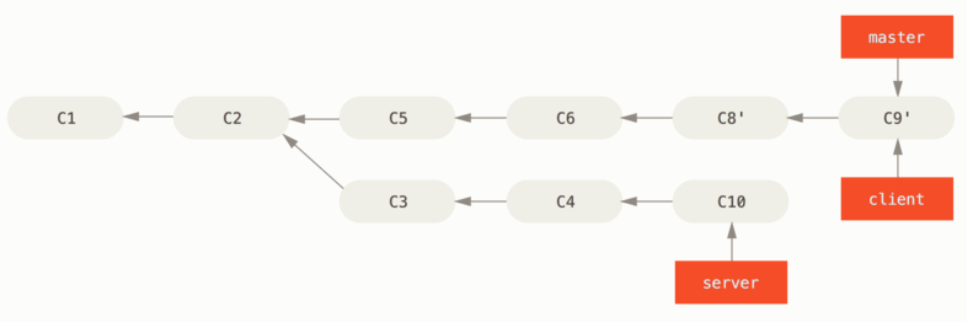
This takes the client branch and pastes it onto the end of master like so:



The branch that comes after the –onto option (master in our case) is the starting point at which we append the commits. We specify both server and client to tell Git to figure out which changes occurred on client since it diverged from server (i.e. after we created the client branch). Once again we can fast-forward merge master to get it up to date with client like so:

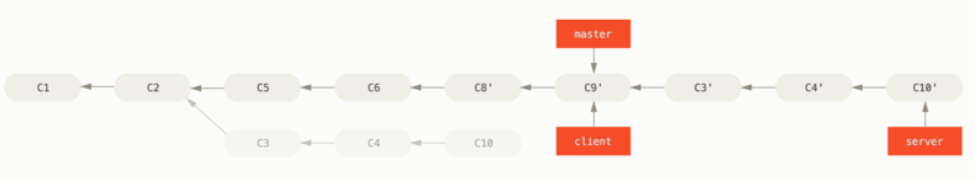
$ git checkout master

$ git merge client



Now in the example, you decide that you’re ready to merge to server. We can rebase server onto the master branch like so:

$ git rebase master server



Once again, this selects master as the start node and then replays changes made to server since it was diverged beginning at master. And once again we can run the following to fast-forward merge master:

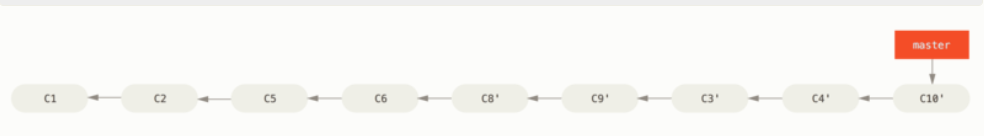
$ git checkout master

$ git merge server

Not only that though, but since we no longer need the client or server branches, we can do the following to delete them:

$ git branch -d client

$ git branch -d server



As mentioned in the Git docs, there is one rule which you must follow when rebasing code: “Do not rebase commits that exist outside your repository and that people may have based work on.” For example, if a co-worker clones the repo and is working on a feature branch, and you rebase the feature branch, your co-worker will have to re-merge their local work into the updated branch structure, which will get messy.

Reverting to Previous Commits:

Sometimes you wish to undo changes to the repos commit history. Subcommands like git checkout and git reset move the HEAD and branch ref. pointers (pointers to existing branches) to a specific commit. git revert is special though, because it does not move ref. pointers to the commit. Instead, it performs the inverse changes made to get to the current commit resulting in a new “revert commit”. So, when we use git revert, it is sort of like a merge where we create an intermediary revert commit that joins HEAD with the commit we want to revert to, this way, we do not delete anything. This is in opposition to git reset which actually destroys subsequent commits in order to go back to a previous one.

Forking Repositories:

Doxygen:

If you are familiar with Javadoc, then you will have a good idea of what Doxygen is. If not, Doxygen is a static site generator that creates HTML/CSS files to neatly display your code’s internal documentation. It does this by reading tags within our code comments. Doxygen is a very useful tool for large projects (e.g. a library) so that developers can navigate the code base and understand what each function does, what parameters it accepts, etc. The tool is relatively easy to use once you’ve learned the basics, which is my aim to teach right now! Be sure to read the official documentation here (created using Doxygen!): <https://doxygen.nl/index.html>

Getting Started With Doxygen:

As always, I am extremely oppressive towards Windows users, so if that includes you then tough luck following along. On Linux, all you must do is install the doxygen package using your native package manager. Once installed, I like to create a directory in my project (usually something like $(project\_root)/res/doc/). Inside of this directory, we want to run the doxygen command with the -g option to generate a default config file. The default name of the file that gets generated is Doxyfile. If, for whatever reason, you don’t like the default name, you can append and alternate name after the -g flag.

Editing the Doxyfile:

Once we’ve generated the default config, we will want to change a few options. And there are many, many, many options. Luckily, we won’t usually care about most of them. Here are a few of the main options and what they do:

**DOXYFILE\_ENCODING:** This sets the encoding format of the Doxyfile. We can pretty much always leave this as UTF-8.  
**PROJECT\_NAME:** The name of your project (surrounded by double quotes).

**PROJECT\_NUMBER:** This represents the revision number of your project. Typically useful for older version control systems such as SVN.

**PROJECT\_BRIEF:** A one line summary of the project.

**PROJECT\_LOGO:** You may optionally include a icon file for your project with max height of 55px and max width of 200px.

**OUTPUT\_DIRECTORY:** This is used if you want to change the location of the output directory for the files which will be generated by doxygen. It can be a relative path or absolute path. If left blank, doxygen assumes the current directory.

**OUTPUT\_LANGUAGE:** This just tells doxygen which encoding scheme to use in the HTML header. For example, it may need a different encoding format for French since it uses a different character set than UTF-8.

**JAVADOC\_AUTOBRIEF:** This one is important if we plan on using Javadoc formatting (my personal preference). The autobrief causes doxygen to interpret the first line of a Javadoc-style comment as the briefing. If this is set to NO (the default), then the briefing must manually be set using the @brief tag. I prefer the latter, but it’s a matter of preference.

**JAVADOC\_BANNER:** Once again, this is only used if you’re using Javadoc-style comments. Setting this to YES will cause doxygen to interpret lines such as /\*\*\*\*\*\*\*\*\* as a Javadoc-style comment “banner”. Otherwise, if set to NO (the default), doxygen just interprets this as the start of a comment.

**QT\_AUTOBRIEF:** This is similar to JAVADOC\_AUTOBRIEF but for QT-style comments. If this is set to YES, then doxygen interprets comments starting with //! or /// as a brief description. I do not personally like this since C++ style comments are only allowed as of C11.

**OPTIMIZE\_\*:** The options beginning with OPTIMIZE\_ are used to generate comments more efficiently for a target language. The options here are for C, Java, FORTRAN, and VHDL. I usually set this to YES since my code is typically written in C, but I doubt it makes a very big difference.

**MARKDOWN\_SUPPORT:** This just lets doxygen use Markdown for generating certain static page elements. I suggest you leave this as YES.

**INPUT:** One of the most important options and it’s way down on line 867! This specifies the files that should be checked for doxygen comments. It can either be a list of files delimited by spaces, or it can be one absolute/relative path to the project directory. Since I like to put my Doxyfile in $(project\_root)/res/doc/ I change INPUT = ../../$(project\_root).

There are obviously many other options which can be set. Doxygen does a pretty good job of giving explanations of what each one does so you can skim through it more thoroughly if there is something specific that you want to change. I just wanted to cover the basic options to give an idea of what I might change.

Doxygen Tags:

Doxygen provides many tags, which are simply called special commands in the official documentation. These are sort of the typical things that you would expect to find in a Javadoc comment. They can begin with an escape character ‘\’ or with an ‘@’ symbol. I prefer the latter because I like to stick with Javadoc-style, but just pick one and be consistent. Many of the tags are (in my opinion) rather useless. I will only cover the ones that I think are useful but be sure to check out the official docs if you’re looking for something specific.

**@class <name> [header file] [header name]:** Used for C++ or Java code when there is a class. You must provide the class name after @class, but you can optionally also specify the header file that it belongs to, or just the header name as it would be #included in your C++ file. Here is an example of using both the header file and header name:

/\*\*

\* @class Main main.h “includes/main.h”

\*/

**@enum <name>:** Pretty straightforward – gives a description of the enum type/class.

**@example[‘lineno’] <file-name>:** This one is definitely niche, but I could see it being used. It essentially points to a file which provides example code. This is nice if you are writing a library and want to provide the end user with some example files.

**@extends <name>:** Another C++/Java-ism – Defines the base class that the derived class derives from.

**@file [name]:** Defines the current file. You should include this in all of your Doxygen comments. Although not necessary, it is required if you want to document external variables, as they rely on this tag to reference the location in which they are declared.

**@implements <name>:** Same as extends but for interfaces.

**@internal:** This is used for an internal comment that should not be included in the static page generated by Doxygen. The comment will terminate naturally at the end of the comment block, but if you require that it terminates before the end of the comment block, you can use **@endinternal**.

**@mainpage [title]:** This tells Doxygen what the front page should be in the static site.

**@namespace <name>:** Indicates that the comment relates to the namespace indicated by name.

**@overload [function decl]:** Creates documentation for overloaded functions.

**@package <name>:** Indicates that the comment addresses the package indicated by name.

**@private, @protected, @public:** Indicates that a member or function is private, protected, or public respectively. Note that you do not need to include these in OOP languages such as C++/Java, as Doxygen will detect this for you. These are only used for languages that don’t support protection natively (which… why would you use these for those?)

**@static:** Similar to private, protected, and public, static should be detected by languages which natively support it. Judging by the fact that the official documentation uses C as an example of a language which doesn’t natively support static (even though it has the static keyword), I’m assuming Doxygen means a static variable in the OOP sense of the word.

**@struct, @union <name> [header file] [header name]:** Indicates that the comment is in reference of the struct/union indicated by name.

**@author, @authors:** Gives the author(s) name(s). These tags function identically to one another.

**@brief {description}:** Gives a one line briefing. If the brief is split into multiple lines, only the first one is considered as the brief.

**@bug {description}:** Reports a known bug in the code. Multiple @bug tags next to one another are just merged into one bug. The tag ends either when a blank line or other tag is encountered.

**@copyright {description}:** Primarily used for larger corporations if you have some sort of legally binding copyright for the written code.

**@deprecated {description}:** Describes a description as to why the particular entity is deprecated.

**@param ‘[‘dir’]’ <param name> [param description]:** This tag is a bit unique. It declares a function parameter. Square brackets may be added after the @param tag with a value of “in”, “out”, or “in,out” depending on whether or not the parameters are incoming, outgoing, or both. This might look something like the following (taken from the Doxygen docs):

/\*\*

\* @param[out] dest The memory area to copy to.

\* @param[in] src The memory area to copy from.

\*/

**@return, @returns {description}:** Gives a description of what the function returns. The @return and @returns tags are identical.

**@see {references}:** Similar to @example, describes one or more classes, functions, methods, variables, files, or URLs as references to view for more information. Two names joined with :: or # are understood as referring to a class (left-hand side) and one of its members (right-hand side).

**@since {date}:** Declares the date or time that the entity was added to the codebase.

**@exception, @throw, @throws <exception object> {description}:** All equivalent in functionality. They give a description of a throwable exception object within the function or class.

**@todo {description}:** Describes a TODO item. Doxygen is nice enough to create a TODO list with all @todo items. Personally, I would rather keep my TODOs as @internal TODO:, but that is up to you.

**@version {version num}:** Gives the version number of the program/application.

**@note, @warning {description}:** Creates a note/warning message. The paragraph generated will be indented by Doxygen and have a little yellow or red indicator bar next to the comment respectively. Similar to @bug, multiple @note/@warning tags next to one another will be joined into one note/warning.

As you can see, there are many commands/tags available to us in Doxygen, many of which resemble Javadoc, and all of which are intuitive and easy to remember. As I mentioned, there are many more of these tags. I included ones which I think I would reasonably see someone using, though I’m sure there are others that I omitted which might be beneficial under certain circumstances. Once more, I implore you to read the official docs for a better overview.

One Other Detail:

One final detail which I did not include is single line Doxygen comments. I wasn’t sure where to mention that, so I’ll mention it here. If you would like to document a variable on the same line as its declaration, you will need to use a single line Doxygen comment. This is accomplished using the same Javadoc or Qt syntax. The catch is that you must use a ‘<’ symbol to indicate that we’re referencing a variable. Here are a few examples:

int var; /\*!< Multiline description \*/

int var; /\*\*< Multiline description \*/

int var; ///< Single line description

int var; //!< Single-line description

There exist other comment syntaxes for multiline comments, however, I think you’ll agree with me upon seeing them that they are rather quite hideous. These 4 seem to be the most sane ones, the second being my personal favorite because of the consistency between Javadoc, and support on C99 or lower. On C++ or Java, the choice will have even less of an impact.

Compiling Doxygen:

It is probably good that you actually know how to compile your comments into the proper HTML files. To do this, simply run the doxygen command followed by the name of the configuration file (Doxyfile is the default). Note that up until now, I’ve only mentioned support for HTML. In reality, Doxygen supports other output formats such as latex, rtf (MS Word), and a few others. I figure that if you are familiar with these formats, you’re probably smart enough to figure out how to compile to them.