

An RDKit-centric intro to git and github

Gregory Landrum¹, Nadine Schneider²

¹ KNIME.com, T5 Informatics

² Novartis Institutes for BioMedical Research, Basel

2016 RDKit UGM

26 October 2016

Goal of this talk

- Showing one way of using Git to contribute to the RDKit
- Most important thing about Git: It is very flexible, everyone has to find her/his own way of using it and feeling comfortable
- What this talk isn't meant to be
 - An in-depth general Git tutorial
 - A comparison of SVN or similar and Git
 - there is much of this in the web, books, etc.

Starting point: what's the difference?



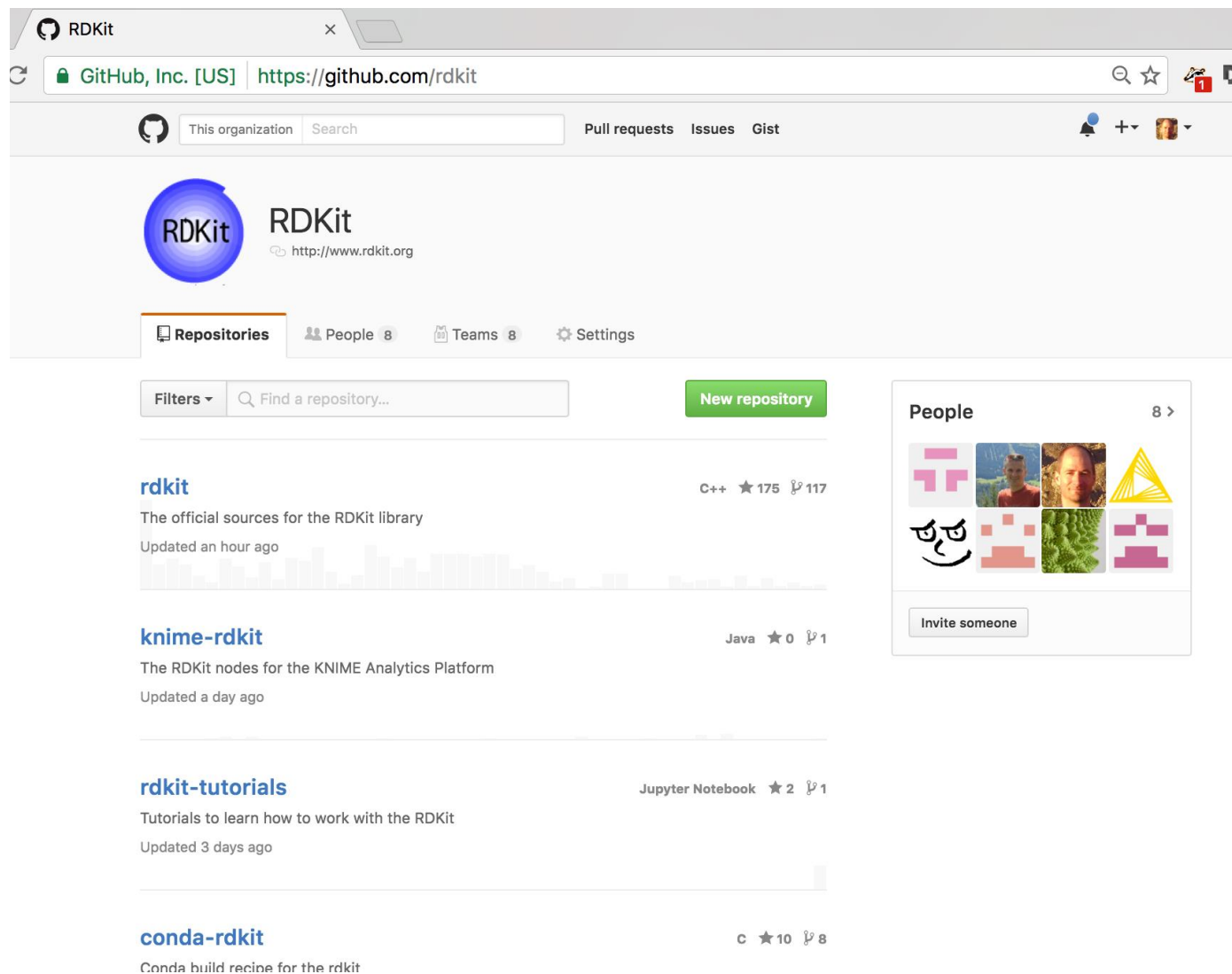
Git: an open-source distributed version control system



Github: a website that hosts git repositories
(among many other things)

A tour of the RDKit github presence

The organization




The screenshot shows the GitHub organization page for RDKit. The browser address bar displays "https://github.com/rdkit". The organization's profile includes a blue circular logo with "RDKit" text, the name "RDKit", and the website "http://www.rdkit.org". Navigation tabs for "Repositories", "People" (8), "Teams" (8), and "Settings" are visible. A search bar and a "New repository" button are at the top of the repository list. The repository list includes:

- rdkit**: C++ language, 175 stars, 117 forks. Description: "The official sources for the RDKit library". Updated an hour ago. A commit history bar is shown.
- knime-rdkit**: Java language, 0 stars, 1 fork. Description: "The RDKit nodes for the KNIME Analytics Platform". Updated a day ago.
- rdkit-tutorials**: Jupyter Notebook language, 2 stars, 1 fork. Description: "Tutorials to learn how to work with the RDKit". Updated 3 days ago.
- conda-rdkit**: C language, 10 stars, 8 forks. Description: "Conda build recipe for the rdkit".

On the right side, the "People" section shows 8 members with their profile pictures and an "Invite someone" button.

A tour of the RDKit github presence

Issues

Filters  is:issue is:open

Labels

Milestones

New issue

☐ ⓘ 155 Open ✓ 440 Closed

Author ▾ Labels ▾ Milestones ▾ Assignee ▾ Sort ▾


☐ ⓘ **Cyclopropenium depiction's dotted triangle has extended edges** **bug**

#1094 opened 2 days ago by shenkin


☐ ⓘ **RDKit cannot create mol from its own canonicalized SMILES**

#1092 opened 3 days ago by shenkin

☐ ⓘ **MolsToGridImage barfs on '&' in labels, at least with useSVG=True** **bug**

#1090 opened 3 days ago by shenkin 🗓 2016_09_1  1

☐ ⓘ **Cant use different fingerprints than RDKit** **bug**

#1085 opened 11 days ago by jaksmid  4


☐ ⓘ **Remove duplicates doesn't work on first column in rdkit.Dbase.DbUtils.GetData** **bug**

#1082 opened 14 days ago by gedeck 🗓 2016_09_1

☐ ⓘ **incorrect image in cookbook** **documentation**

#1079 opened 21 days ago by greglandrum

☐ ⓘ **Consistent formatting of Python code** **enhancement**

#1071 opened 23 days ago by gedeck 🗓 2016_09_1  15

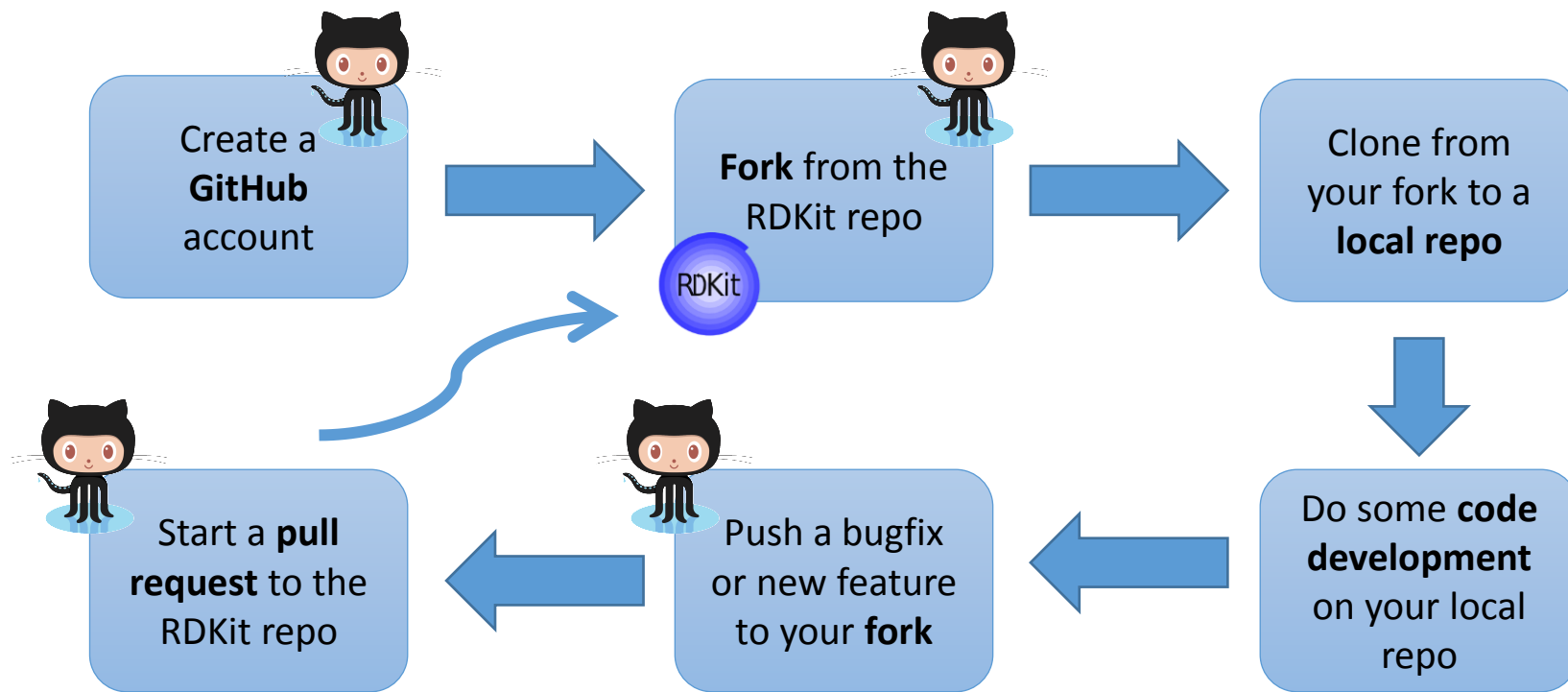
A tour of the RDKit github presence

The repo

The screenshot shows the GitHub repository page for RDKit. At the top, the browser address bar displays the URL `https://github.com/rdkit/rdkit`. The repository name `rdkit / rdkit` is shown with 46 watchers, 175 stars, and 117 forks. Navigation tabs include Code, Issues (156), Pull requests (15), Projects (0), Wiki, Pulse, Graphs, and Settings. A banner states "The official sources for the RDKit library — Edit". Below this, statistics are listed: 5,018 commits, 15 branches, 88 releases, 38 contributors, and the BSD-3-Clause license. Action buttons for "New pull request", "Create new file", "Upload files", "Find file", and "Clone or download" are present. A recent commit by `jandom` is highlighted, with a table of recent activity below it.

File	Commit Message	Time
Code	Conformer GetPos returns a numpy array rather than a tuple of tuples (#...	2 hours ago
Contrib	Issue1071/yapf (#1078)	21 days ago
Data	Issue1071/yapf (#1078)	21 days ago
Docs	Issue1071/vaof (#1078)	21 days ago

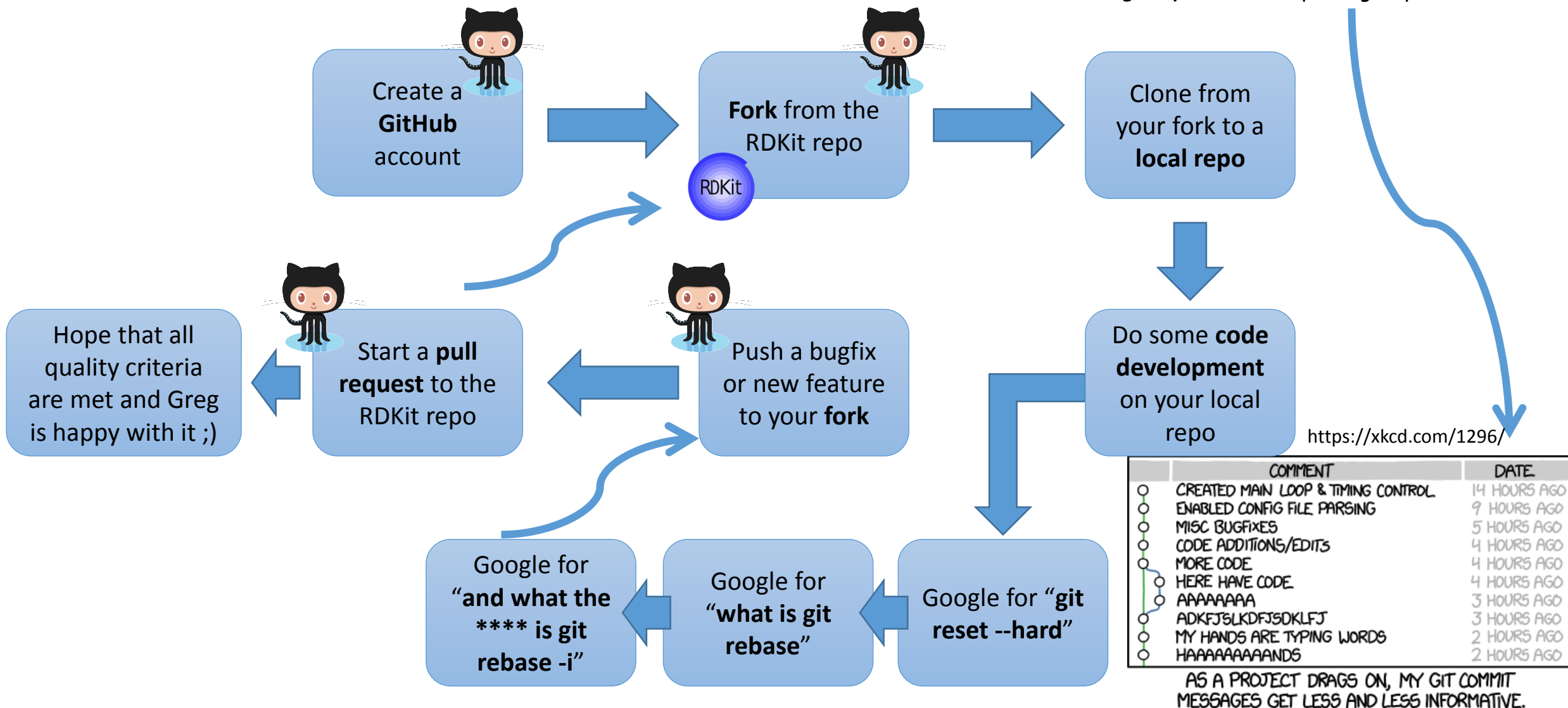
The overall workflow



The ~~overall~~ workflow

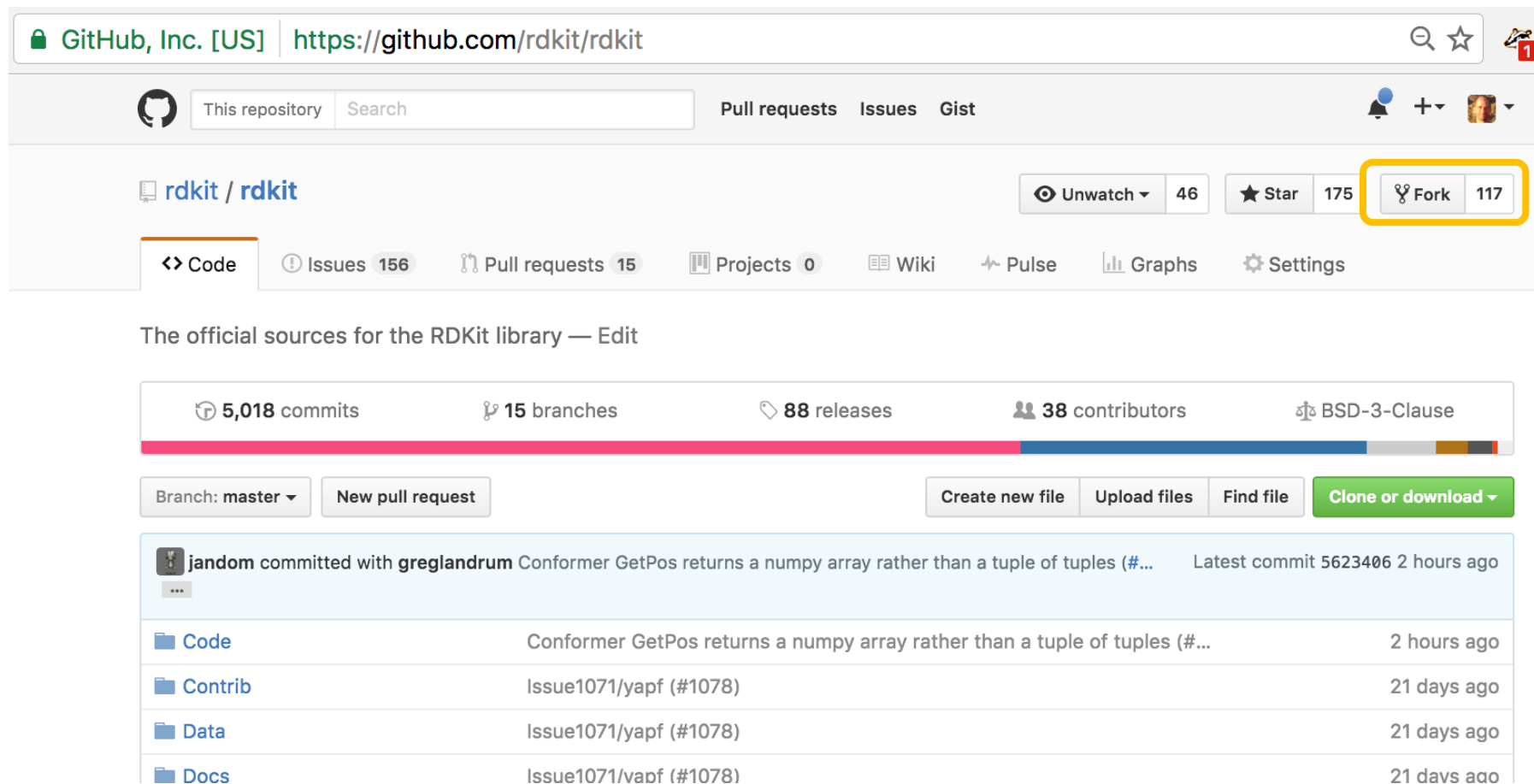
realistic

One cool thing about Git is that you can commit as much as you like in your **local repo**, no one will ever see it if you use **git squash** before pushing to your fork



A tour of the RDKit github presence

The repo



GitHub, Inc. [US] <https://github.com/rdkit/rdkit>

This repository Search Pull requests Issues Gist

rdkit / rdkit Unwatch 46 Star 175 Fork 117

<> Code Issues 156 Pull requests 15 Projects 0 Wiki Pulse Graphs Settings

The official sources for the RDKit library — Edit

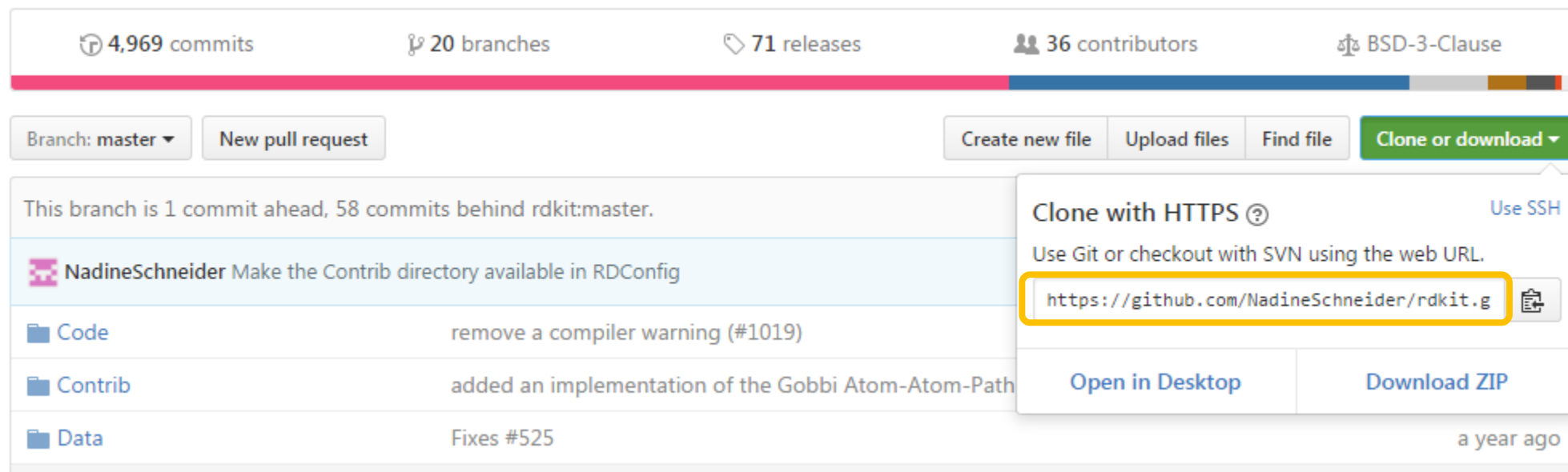
5,018 commits 15 branches 88 releases 38 contributors BSD-3-Clause

Branch: master New pull request Create new file Upload files Find file Clone or download

jandom committed with greglandrum Conformer GetPos returns a numpy array rather than a tuple of tuples (#... Latest commit 5623406 2 hours ago

Code	Conformer GetPos returns a numpy array rather than a tuple of tuples (#...	2 hours ago
Contrib	Issue1071/yapf (#1078)	21 days ago
Data	Issue1071/yapf (#1078)	21 days ago
Docs	Issue1071/vaof (#1078)	21 days ago

Clone from your fork to a local repo



4,969 commits 20 branches 71 releases 36 contributors BSD-3-Clause

Branch: master New pull request Create new file Upload files Find file Clone or download

This branch is 1 commit ahead, 58 commits behind rdkit:master.

NadineSchneider Make the Contrib directory available in RDConfig

Code	remove a compiler warning (#1019)
Contrib	added an implementation of the Gobbi Atom-Atom-Path
Data	Fixes #525

Clone with HTTPS Use SSH

Use Git or checkout with SVN using the web URL.

<https://github.com/NadineSchneider/rdkit.g>

Open in Desktop Download ZIP

a year ago

Copy this to your console:

```
git clone https://github.com/YOUR_USERNAME/rdkit.git
```

Now your contribution to RDKit can begin...

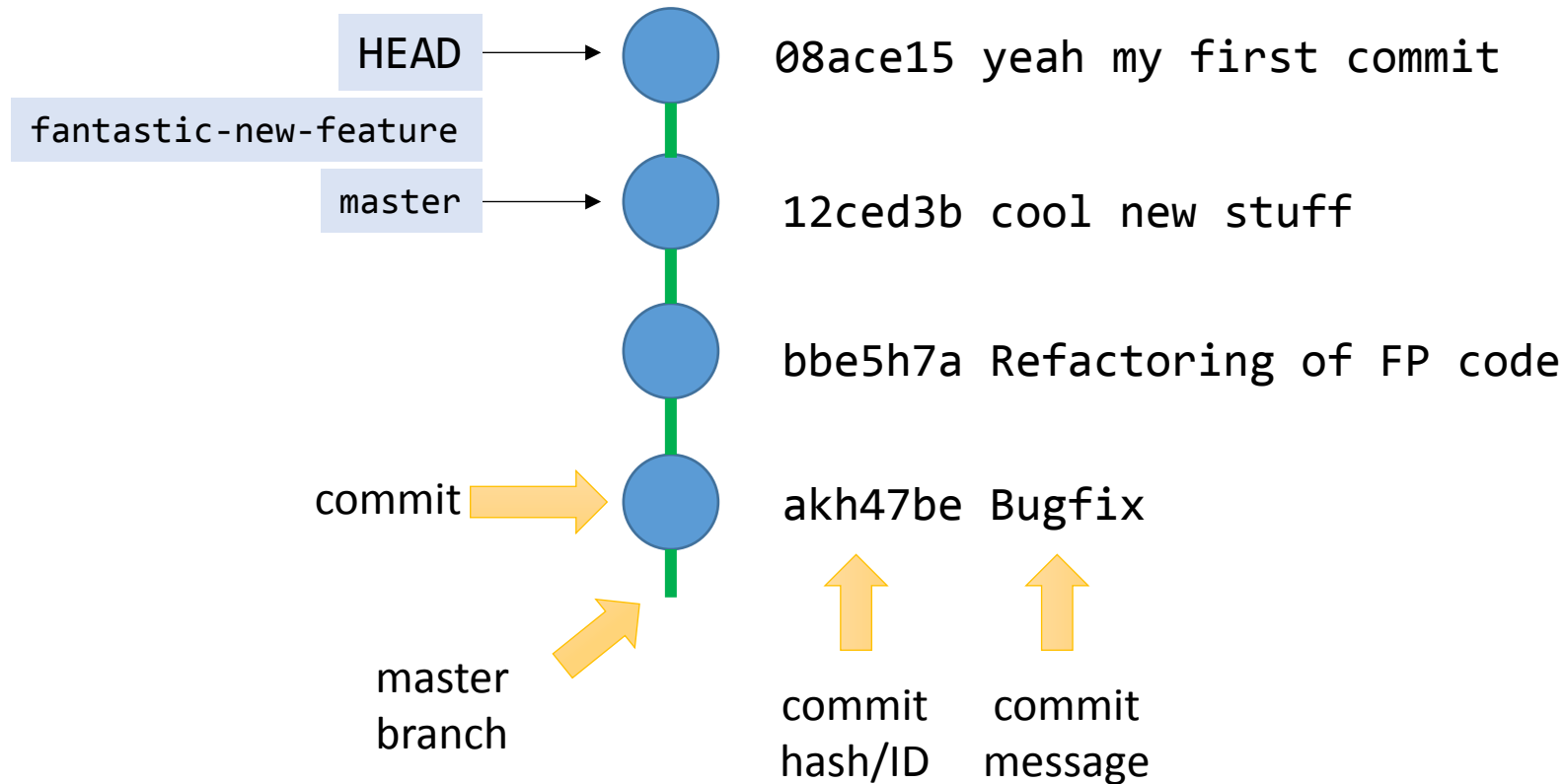
- Let's first checkout a new branch where we develop a new feature:

`git checkout -b "fantastic-new-feature"`

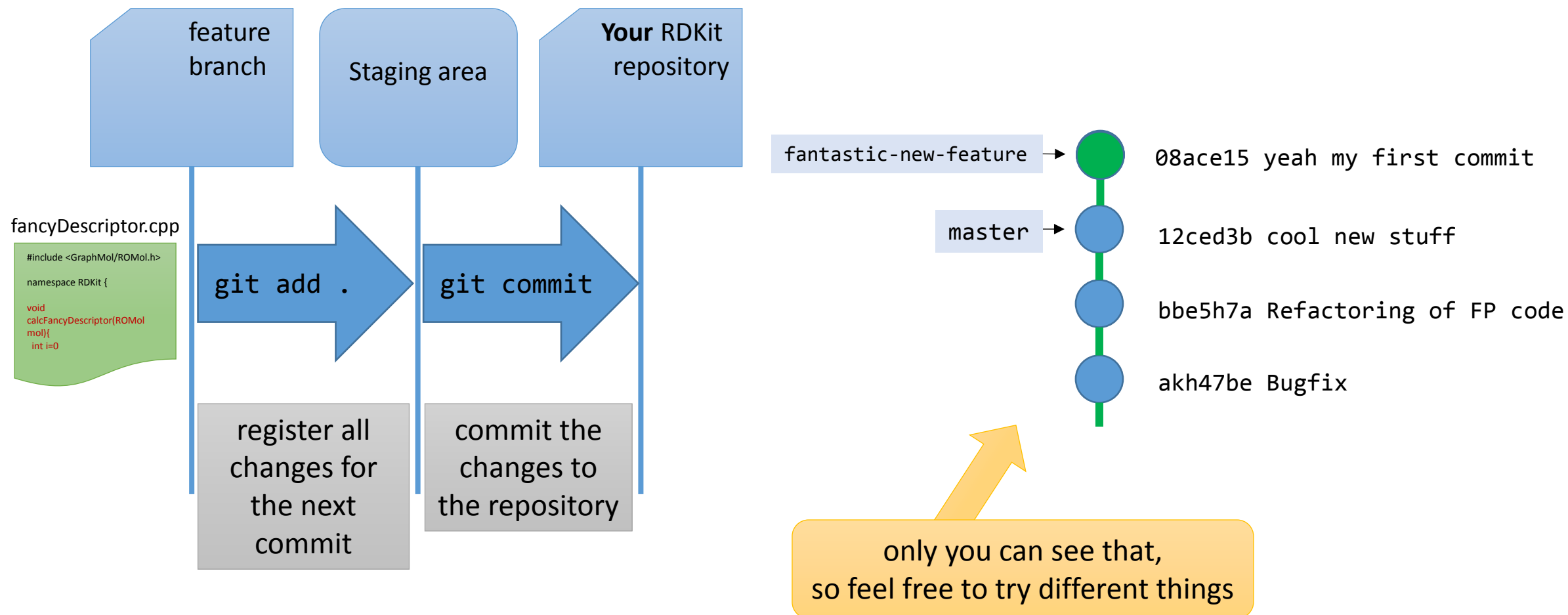


After a few lines of code, let's do a commit

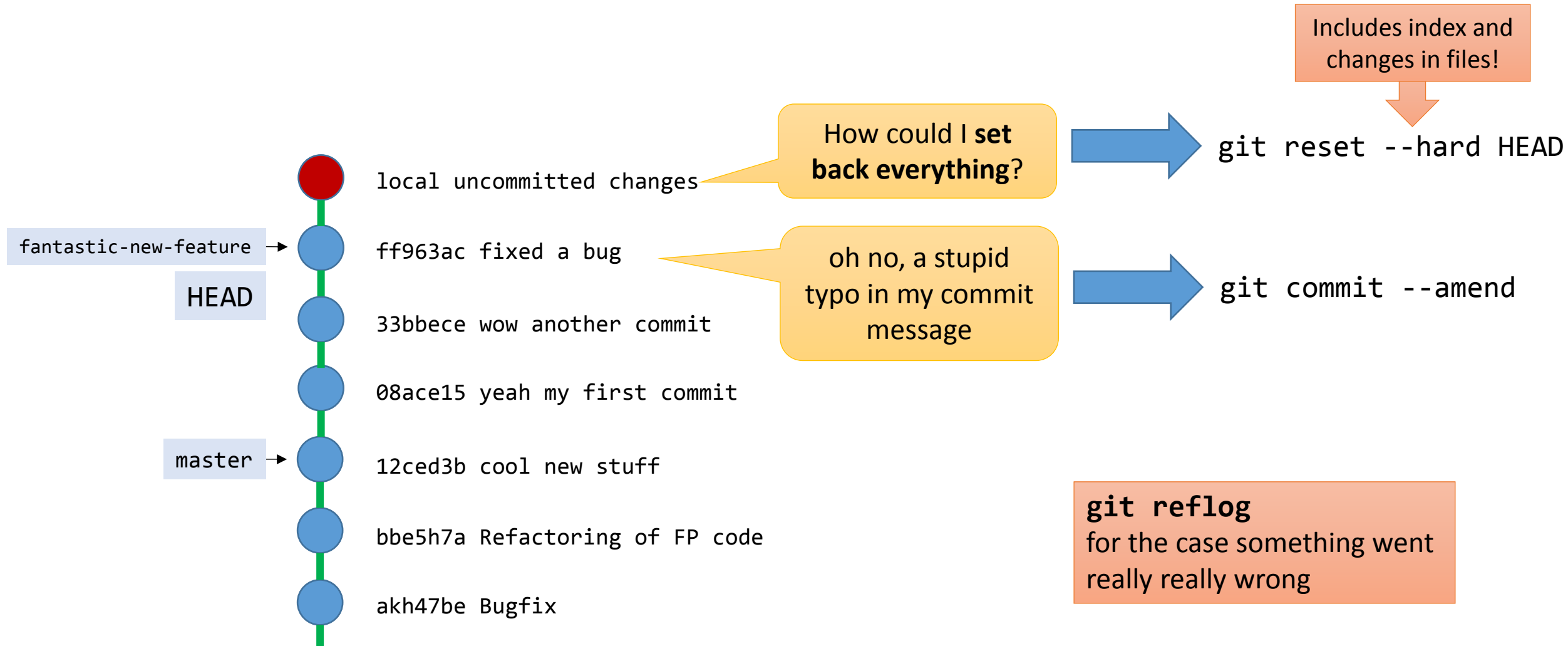
```
git add .  
git commit -m "yeah my first commit"
```



What happens in the back...



Let's continue developing the new feature



Feature finished! Let's put it back on the master

- But first, we have to update our master (RDKit never sleeps 😊)
- For this add a remote to your repository:

```
git remote -v
```

```
origin  https://github.com/YOUR_USERNAME/rdkit.git (fetch)  
origin  https://github.com/YOUR_USERNAME/rdkit.git (push)
```

```
git remote add upstream https://github.com/rdkit/rdkit.git  
git remote -v
```

```
origin  https://github.com/YOUR_USERNAME/rdkit.git (fetch)  
origin  https://github.com/YOUR_USERNAME/rdkit.git (push)  
upstream https://github.com/rdkit/rdkit.git (fetch)  
upstream https://github.com/rdkit/rdkit.git (push)
```

Name of your
remote.
Choose whatever
you like!

Feature finished! Let's put it back on the master

- Now let's get the update from the original RDKit repo

```
git fetch upstream
```

```
remote: Counting objects: 75, done.  
remote: Compressing objects: 100% (53/53), done.  
remote: Total 62 (delta 27), reused 44 (delta 9)  
Unpacking objects: 100% (62/62), done.  
From https://github.com/rdkit/rdkit  
* [new branch]      refactor_FPs -> upstream/refactor_FPs  
12ced3b..eeffa12   master -> upstream/master
```

Yeah we have an update!

- Before we continue, let's see where we are:

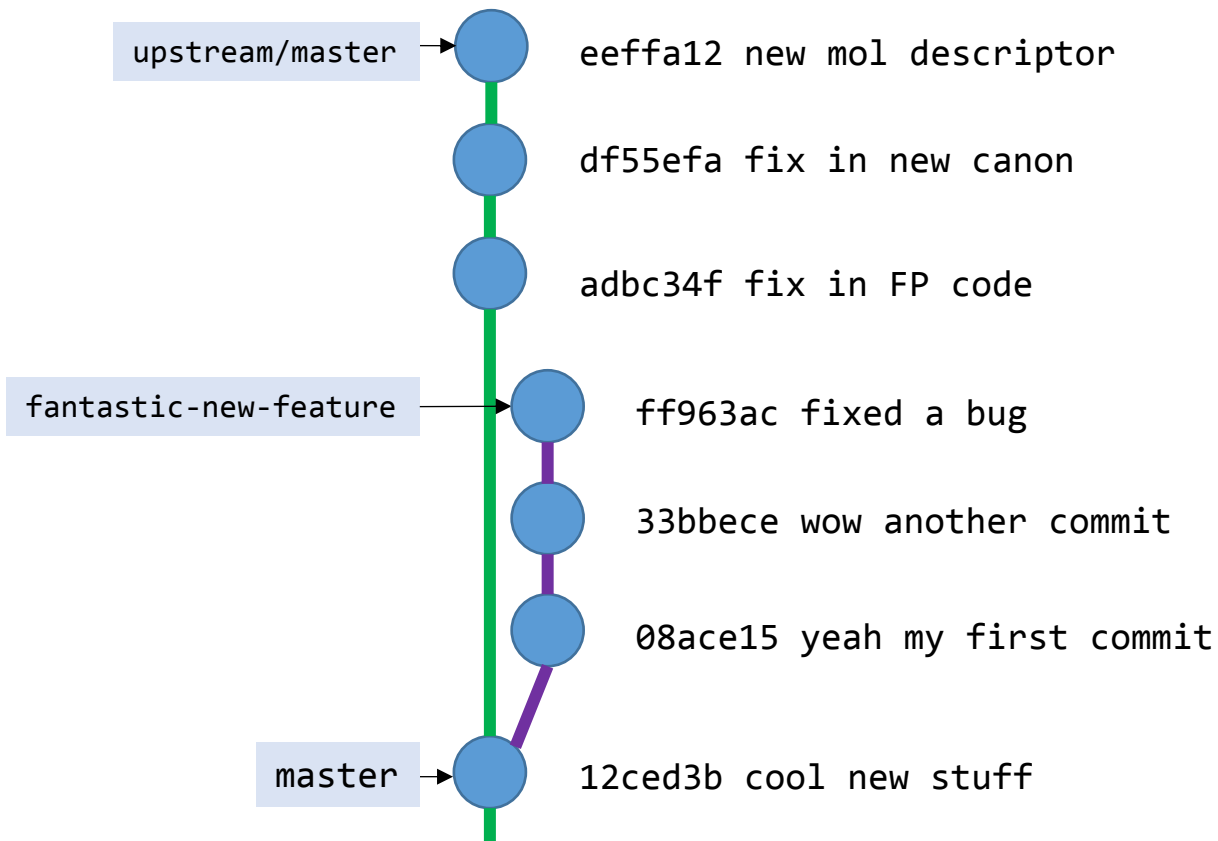
```
git branch
```

```
master  
* fantastic-new-feature
```

We are still on our feature branch

Feature finished! Let's put it back on the master

How the repository looks at the moment



What we can do next:

1 `git merge upstream/master`

or

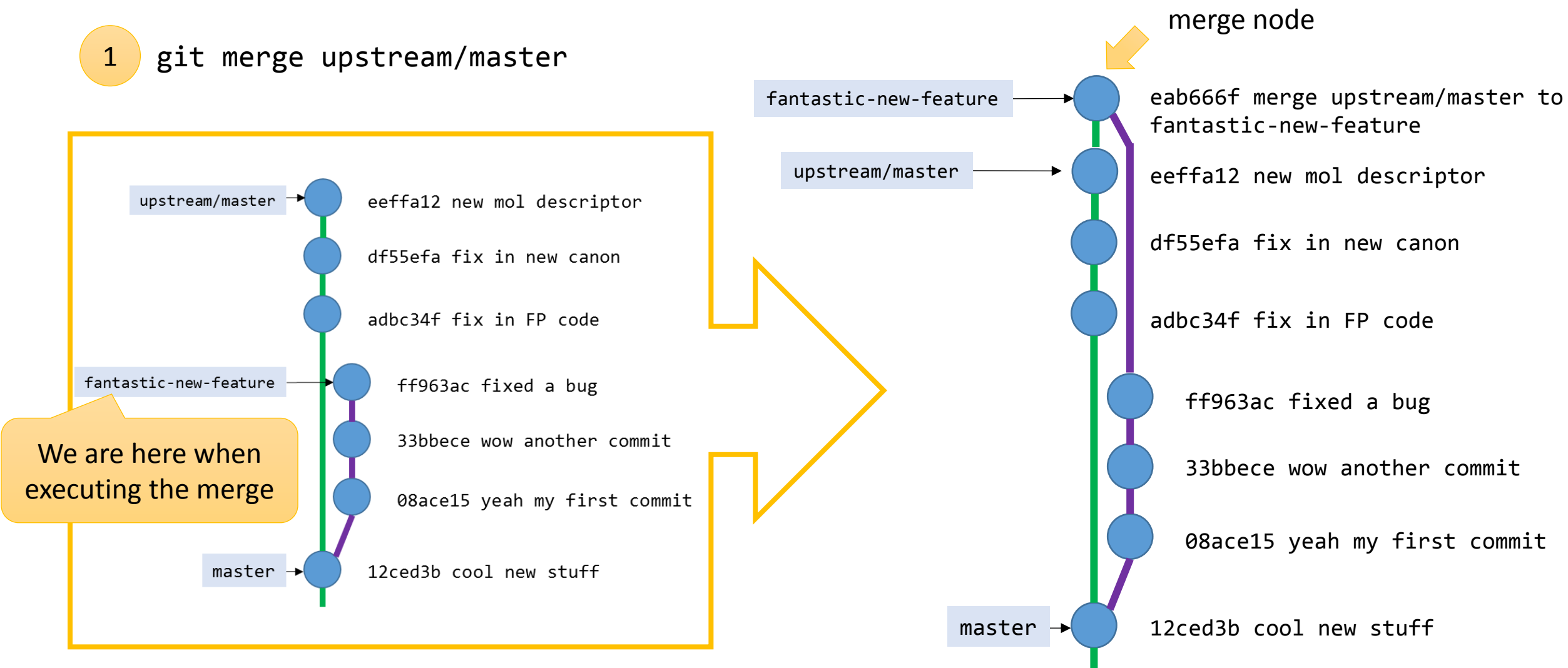
2 `git rebase upstream/master`

or

3 `git checkout -b new-master upstream/master`
`git cherry-pick 08ace15`
`git cherry-pick 33bbece`
`git cherry-pick ff963ac`

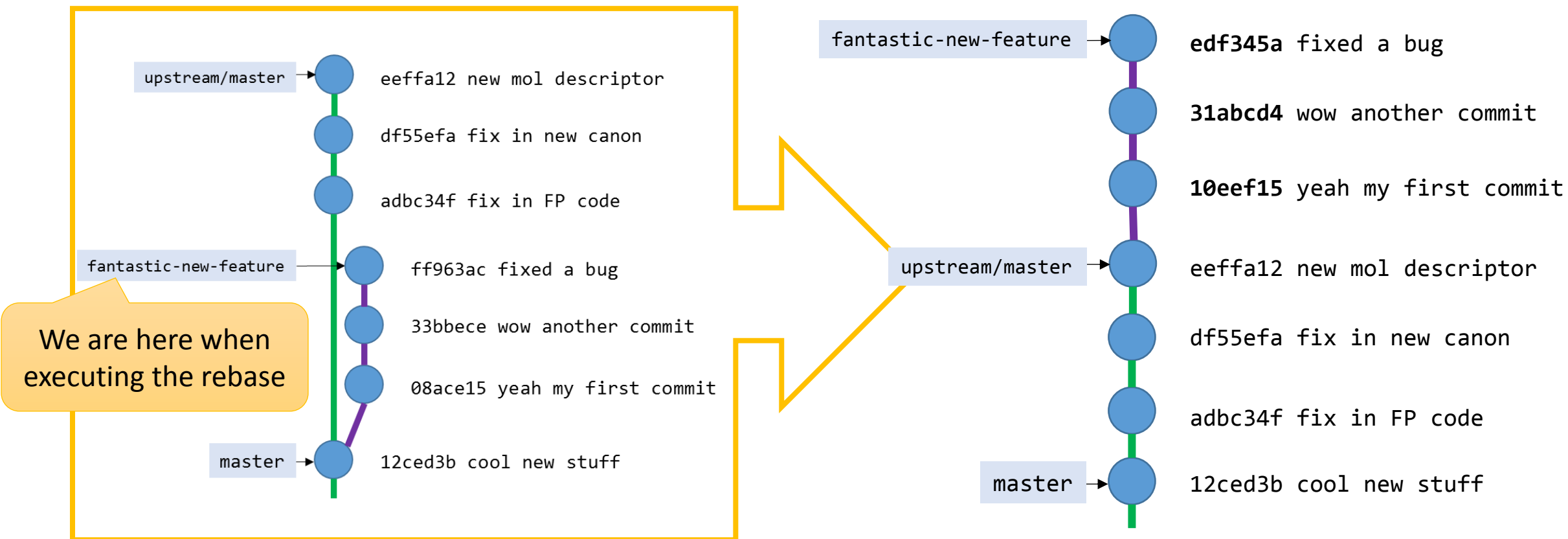
Feature finished! Let's put it back on the master

1 `git merge upstream/master`

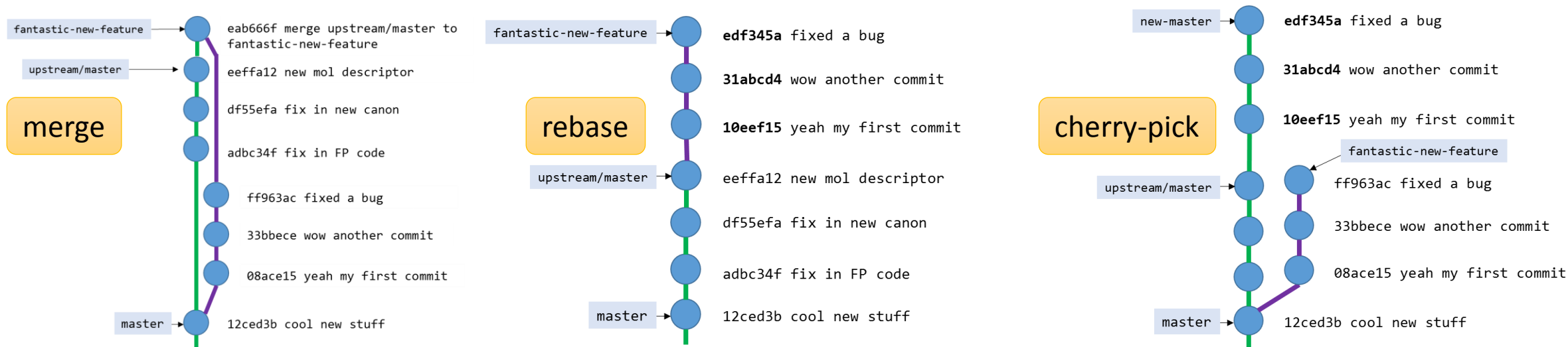


Feature finished! Let's put it back on the master

2 `git rebase upstream/master`



We are ready to push our new feature!



- Independent of your workflow the branch is now up to date and ready for a push on your github repo so that a pull request can be started

```
git push origin fantastic-new-feature:new-feature-for-rdkit  
                        (new-master)
```

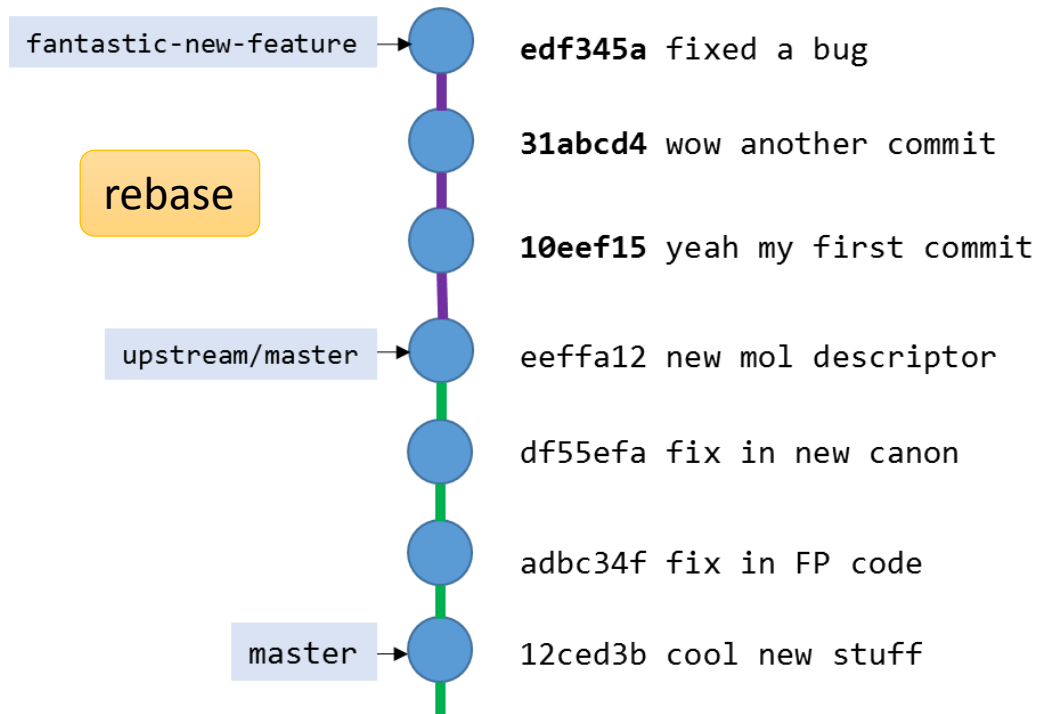
remote

local branch name

remote branch name

Done!!
Your changes
are online. You
can start a pull
request!

One last word about squashing...



Would you really like to share all of your “experimental commits and messages” with the rest of the world?

NOPE!

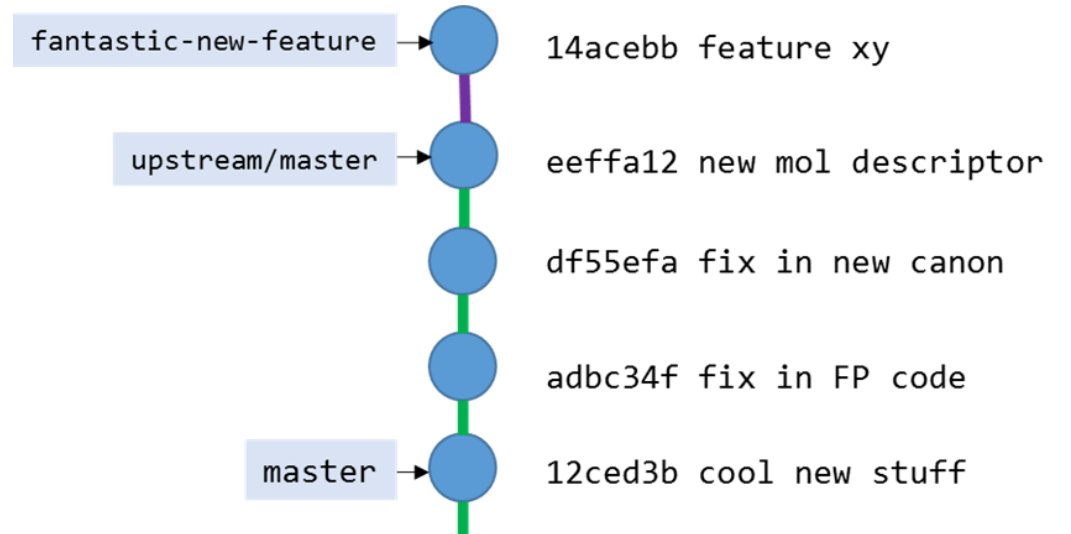
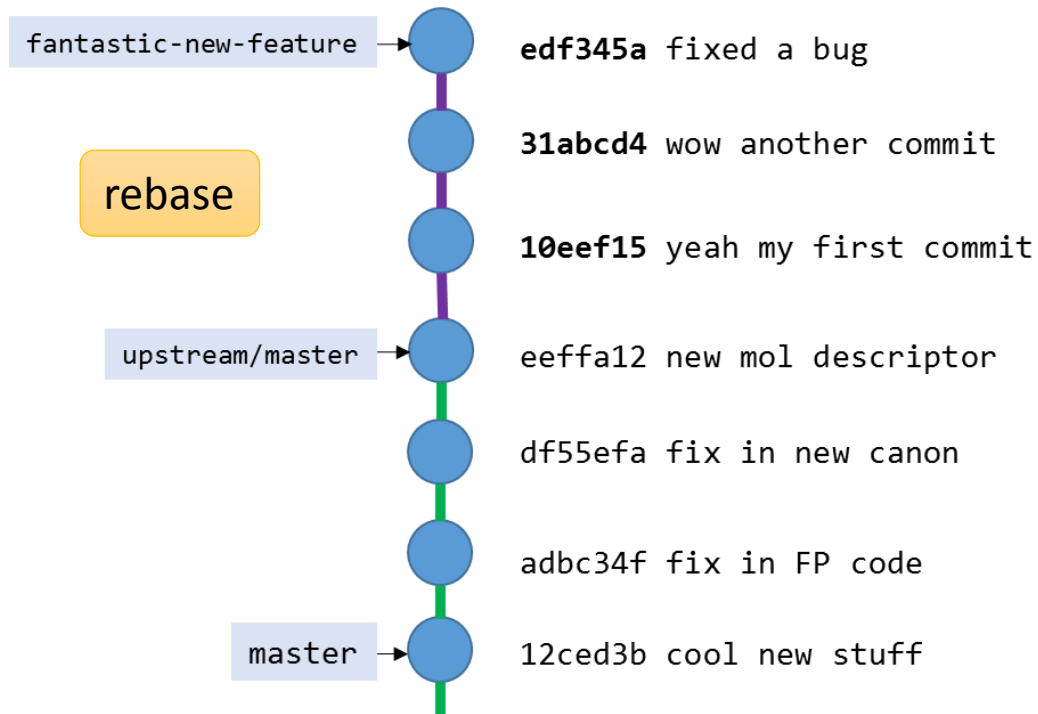
Let's try

```
git rebase -i eeffa12
```



Allows you to consolidate single commits to one meaningful commit

One last word about squashing...give it a try!



Result squashing using

```
git rebase -i eeffa12
```

A tour of the RDKit github presence

Pull requests












Filters ▾

is:pr is:open

Labels

Milestones

New pull request

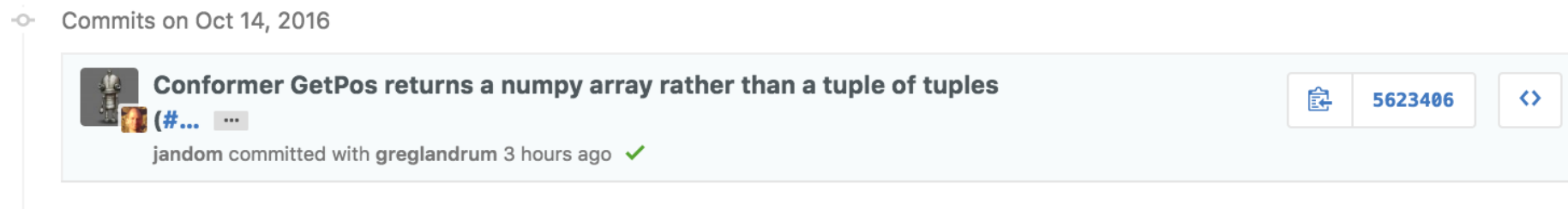
<input type="checkbox"/>	 14 Open	<input checked="" type="checkbox"/> 501 Closed	Author ▾	Labels ▾	Milestones ▾	Assignee ▾	Sort ▾
<input type="checkbox"/>	 Dev/reaction enumeration ✓						 32
	#1111 opened 6 days ago by bp-kelley						
<input type="checkbox"/>	 Morse & RDF initial implementation						 2
	#1107 opened 9 days ago by thegodone						
<input type="checkbox"/>	 Faster version of Pharm2D.Utils.GetUniqueCombos ✓						 6
	#1080 opened on Sep 24 by mitchondra						
<input type="checkbox"/>	 Adds a FreeSASA adapter ✓						 25
	#1058 opened on Sep 9 by bp-kelley						
<input type="checkbox"/>	 Preserve bond direction in fraamentOnBonds ✓						 2

Reviews of pull requests

- We review all pull requests (PRs) for quality, compliance with the coding standards (such as they are), and to ensure that they make sense
- Reviews are open (in github) and anyone is welcome to participate
- We are starting to use the new (as of Sept 2016) formal review mechanism in github, but haven't completely adopted that yet
- After the review, the PR is either merged by a maintainer or closed without merging it.

Why can't I just email you my changes?

- Well... you can, if that's ***really*** necessary
- But then *I* need to create the branch and do the PR
- A real benefit of using the standard route is that you get credit for the contribution:



(this will also show up correctly in the release notes)

Acknowledgements

Brian Kelley

Nikolaus Stiefl

Nikolas Fechner

Nico Pulver

Anna Pelliccioli



<http://xkcd.com/1597/>

Any questions?

Still perplexed?

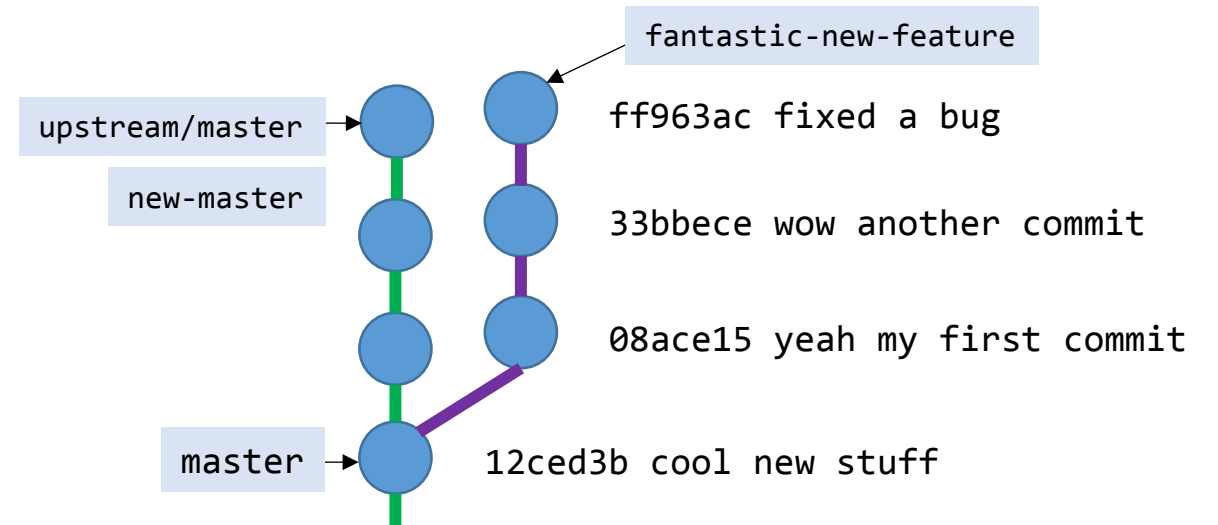
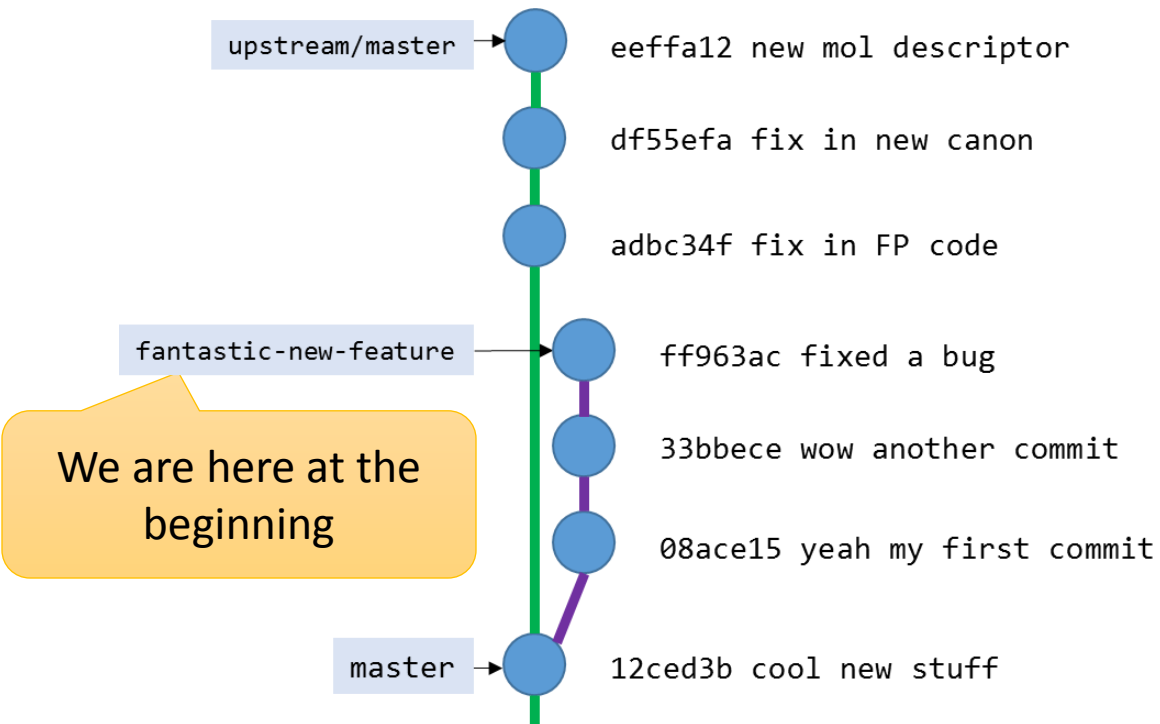
Some good cheat sheets:

<https://services.github.com kit/downloads/github-git-cheat-sheet.pdf>

<https://zeroturnaround.com/wp-content/uploads/2016/05/Git-Cheat-Sheet-by-RebelLabs.png>

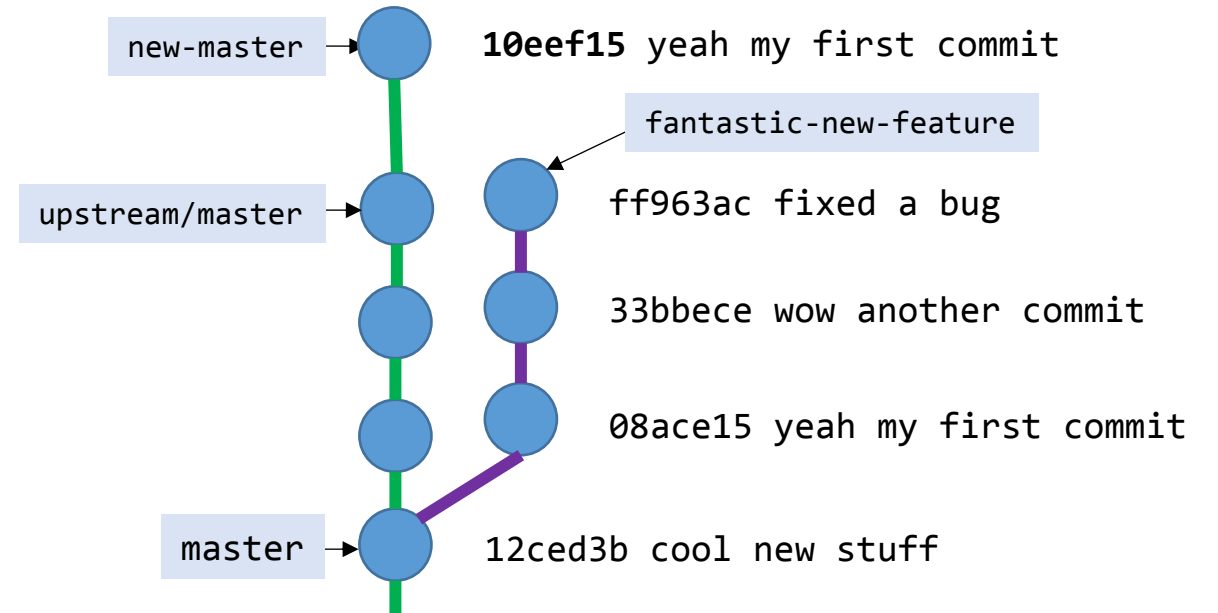
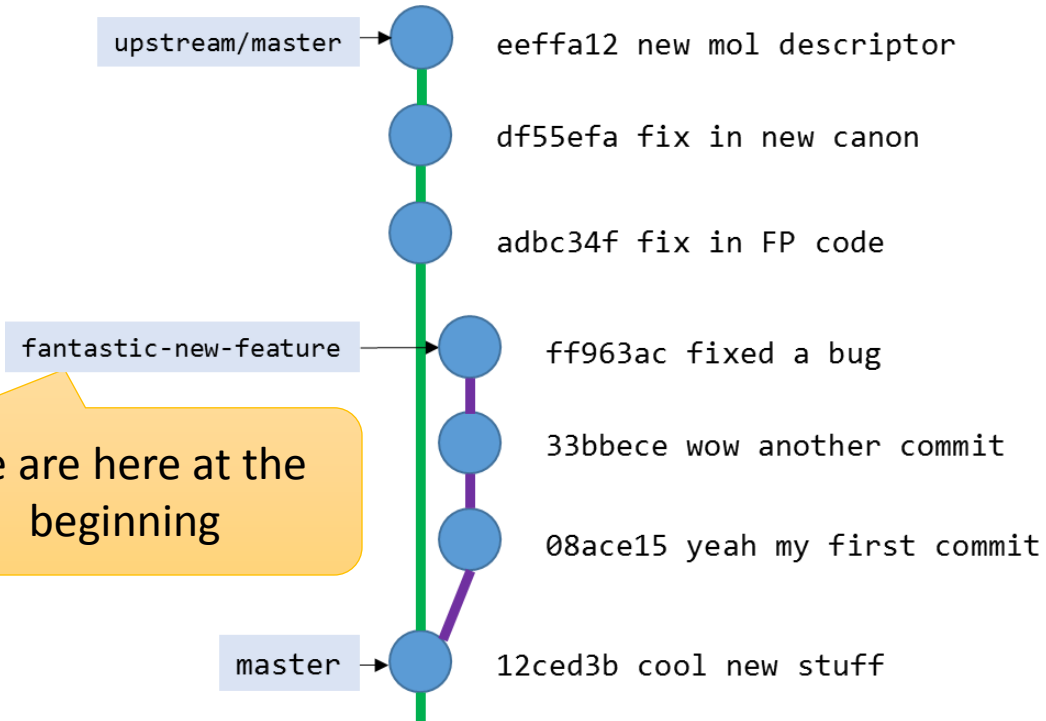
Feature finished! Let's put it back on the master

3 `git checkout -b new-master upstream/master`



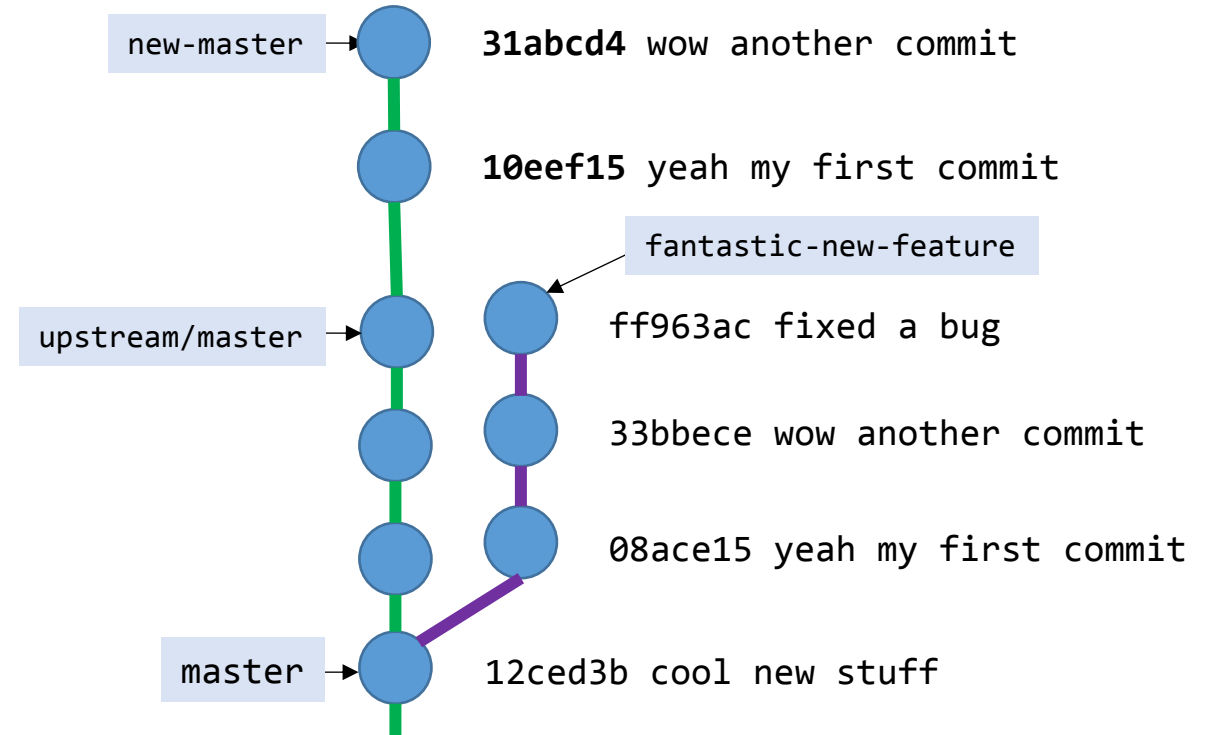
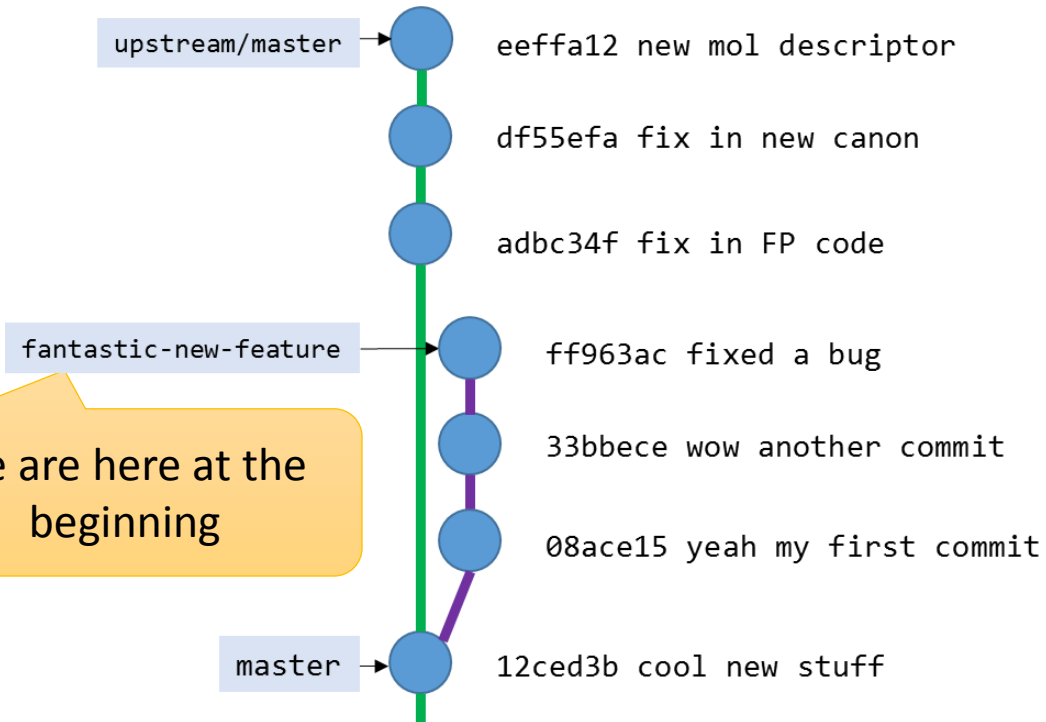
Feature finished! Let's put it back on the master

3 `git checkout -b new-master upstream/master`
`git cherry-pick 08ace15`



Feature finished! Let's put it back on the master

3 `git checkout -b new-master upstream/master`
`git cherry-pick 08ace15`
`git cherry-pick 33bbece`



Feature finished! Let's put it back on the master

3 `git checkout -b new-master upstream/master`
`git cherry-pick 08ace15`
`git cherry-pick 33bbece`
`git cherry-pick ff963ac`

