

Topic Presentation: Git

Topic Presentation: Git(Hub)

What is Git?

ONE DOES NOT SIMPLY

UNDERSTAND GIT

MemesHappen

What is Git?

Git is a distributed version-control system for tracking changes in source code during software development.



What is GitHub?

[...] provides hosting for software development and version control using Git

Is GitHub the only company to do this?

No.

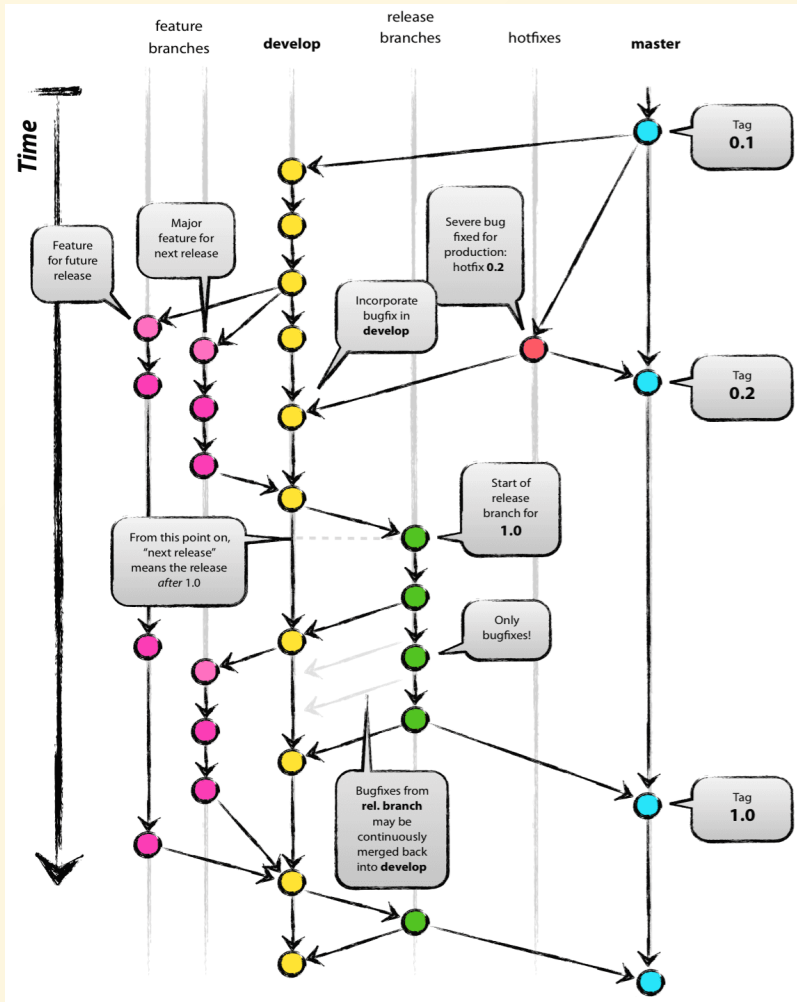
e.g. BitBucket, GitLab, proprietary platforms.

Part I: `git`

Some best practices...

Conventions!

Branching Model 😊



Feature Branches

```
git checkout master  
git pull  
git checkout -b myfeature
```

Forks



Forks (some useful commands)

Once forked (e.g. in GitHub)

Clone your fork

```
git clone https://github.com/USERNAME/REPOSITORY.git
```

Add the "upstream"

```
git remote add upstream https://github.com/OWNER/REPOSITORY.git
```

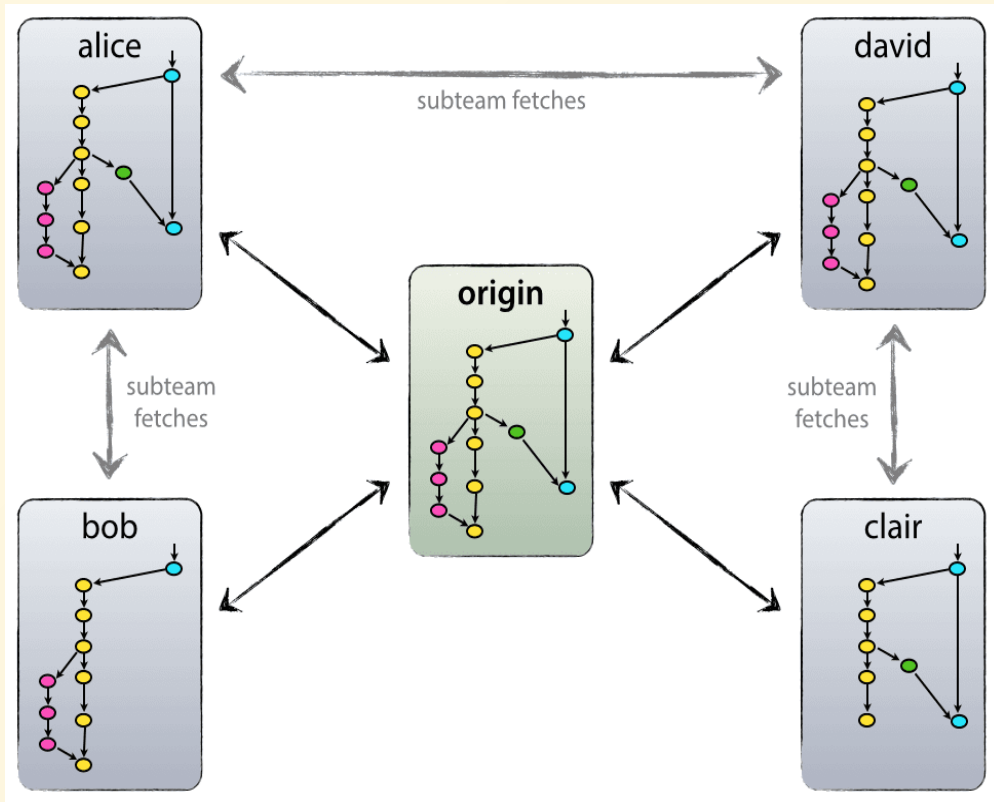
Fetch changes upstream

```
git fetch upstream
```

You can also pull from a branch

```
git pull upstream master
```

Decentralised but Centralised



[The Ecosystem is Moving - Moxie Marlinspike](#)

LOOK AT ME

THIS IS MY REPO NOW

Minimise force pushes 🥵...



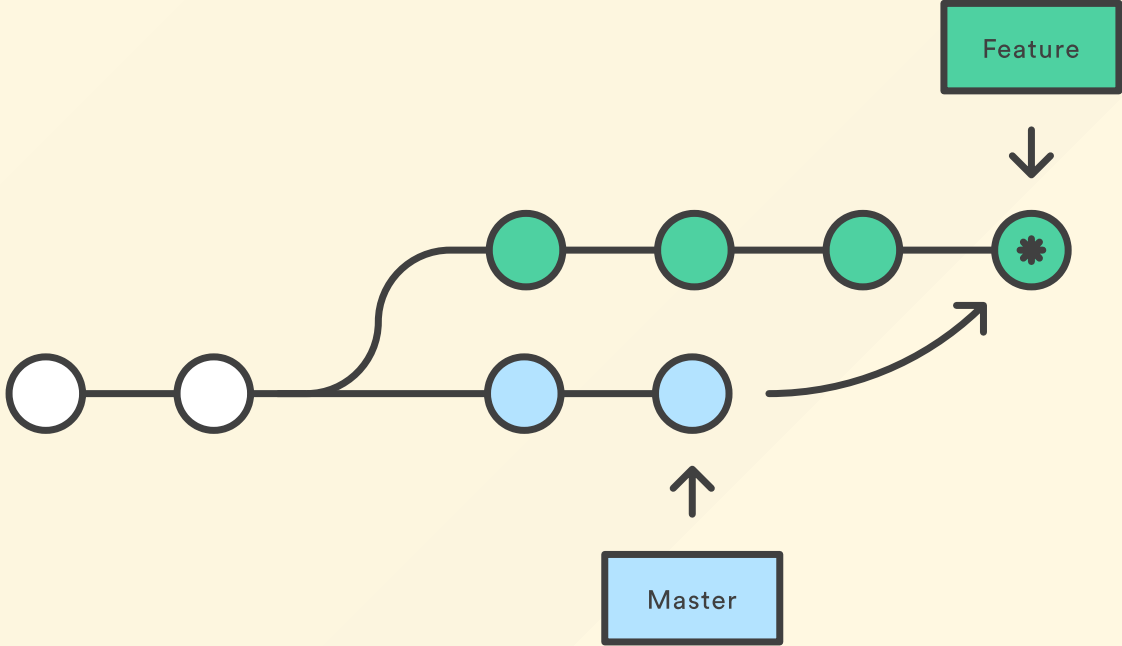
Especially on master !

Rebase vs Merge



Merge

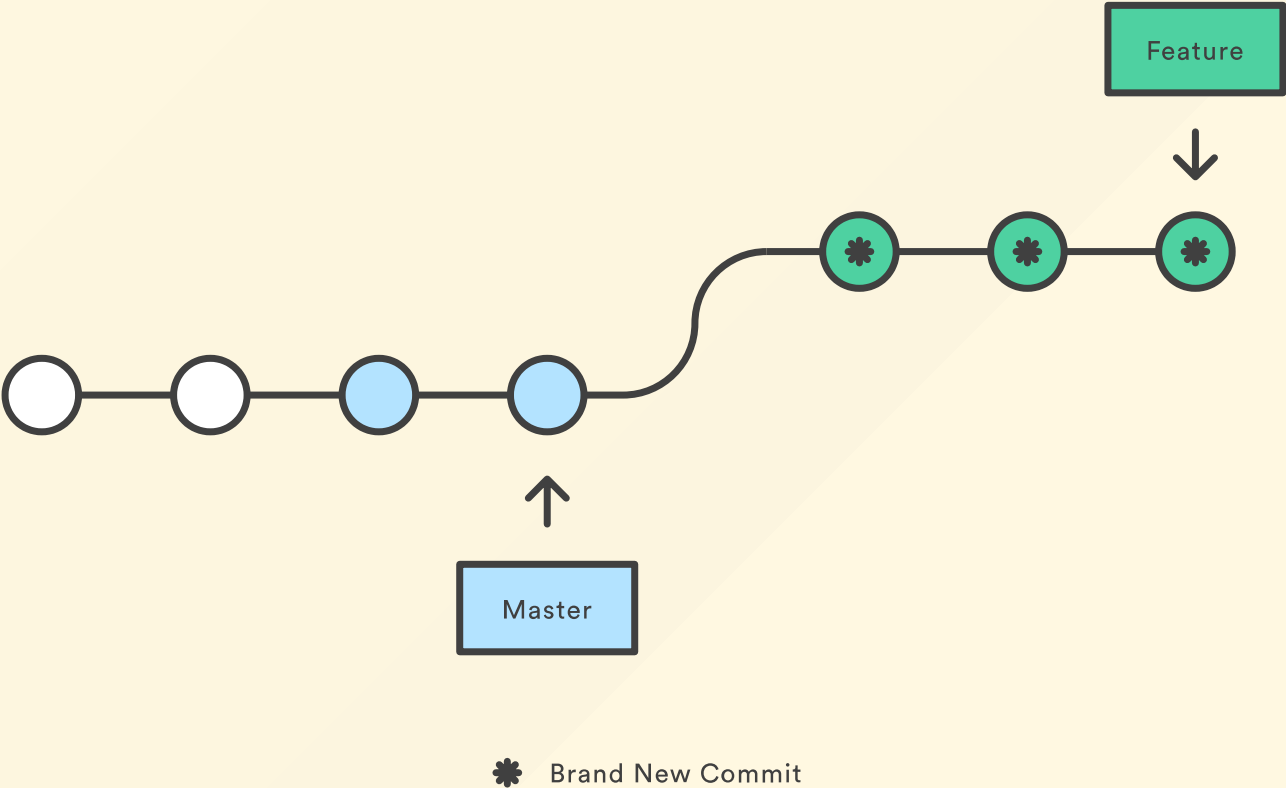
Merging master into the feature branch



* Merge Commit

Rebase

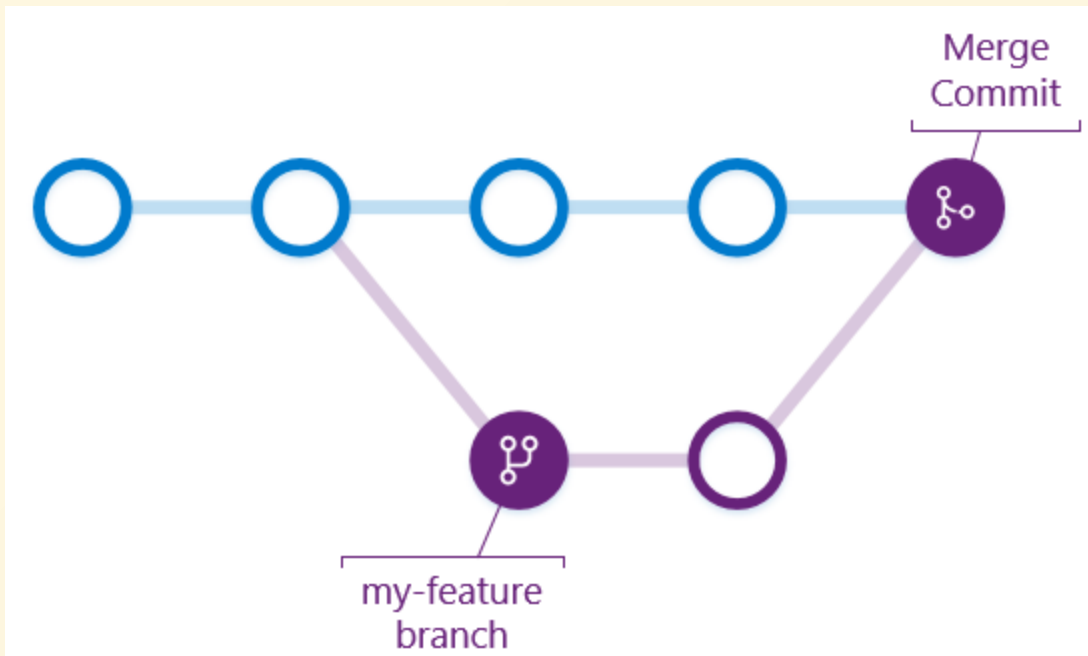
Rebasing the feature branch onto master



Rebase Often

```
git checkout master  
git pull  
git checkout your-feature  # Feature branch  
git rebase master
```

Squash commits before merging



Part II: GitHub

Pull Requests



Pull Requests: Some recommendations

Authors:

- Use drafts / WIP
- Make them compact
- Request review from relevant people.

Reviews:

- Take time to review.
- Leave meaningful comments.

Pull Requests: Some recommendations

Everyone:

- Alerts are sent to `#ds-github`
- Make sure you use the labels and add projects and milestone tags.

Code-Owners

- Notifications are automated as follows:

```
# Engineering Team
* @bitsoex/data-science-engineering
/.github/ @bitsoex/data-science-engineering
/data_engineering/ @bitsoex/data-science-engineering
/deployments/ @bitsoex/data-science-engineering
# Experiments
/experiments/ @bitsoex/data-science-experiments
# Models
/models/ @bitsoex/data-science-models
# All
/reports/ @bitsoex/data-science
/shiny/ @bitsoex/data-science
/utils/ @bitsoex/data-science
```

Branch Rules

- Main branch `master` is protected with the following:
 1. Need CO's approval.
 2. Cannot force-push.
 3. Squash-merging is required.
- `development` branch will be used for CI/CD deployments to the `dev` environment.

Projects

- Use projects to organise PRs and Issues.
- Can be automated.

Milestones

Set milestones for projects and filter by milestone...

GitHub Actions

- It was a potential solution for CI/CD stuff.
- We will integrate testers, formatters and linters.

Bonus: GitHub Pages!

- Easy set-up.
- Can be [integrated with workflows](#).
- Multiple purposes.

GitHub is massive