

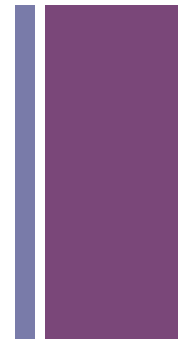


豆瓣 **douban**

How CODE works

Xu Tao
11.15 Beijing
PyCon2014China

+ 说明



- **CODE** 其实包含不少子项目
- 我觉得是一个生态系统了，全靠 **API**
- **CODE** 是一个很多人参与的项目
- 我期望尽量能够把每个项目对应的作者都能说清楚
- **ID** 都是他们的 **GitHub** 的 **ID**
- 当然也期望我能够讲清楚

+ 我与 Pycon 的故事



- 11 年，我曾来上海参加过一次 PyCon
- 那时我还不在豆瓣
- 那时我还不会写 Python
- 第一次听教授 @hongqn 讲“Python在豆瓣的应用”
- 对豆瓣充满了向往
- ...(曲折，略去)
- 13 年初我终于混进了豆瓣
- 14 年，3 年过去了，我也上来讲了，其实你也可以

+ 可以分享的



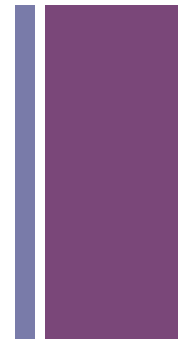
- **GIT – GIT** 相关的
- **Python – Python** 相关的
- **Web** – 从一个协作应用来讲述
- **Workflow** – 日常工作方式
- =人=
- 可以分享（吐槽）的东西太多啦
- (蹉跎中练就一身吐槽技能 → →)



CODE 应用

GIT 那些事儿

+ 概述



- 整体的讲一下 **CODE** 的情况
- 在豆瓣内部 **CODE** 是一个 **DAE** 的应用
- **DAE** 为 **CODE** 提供了大部分的基础设施
- 很遗憾现在的 **CODE** 还不能 **Scale**

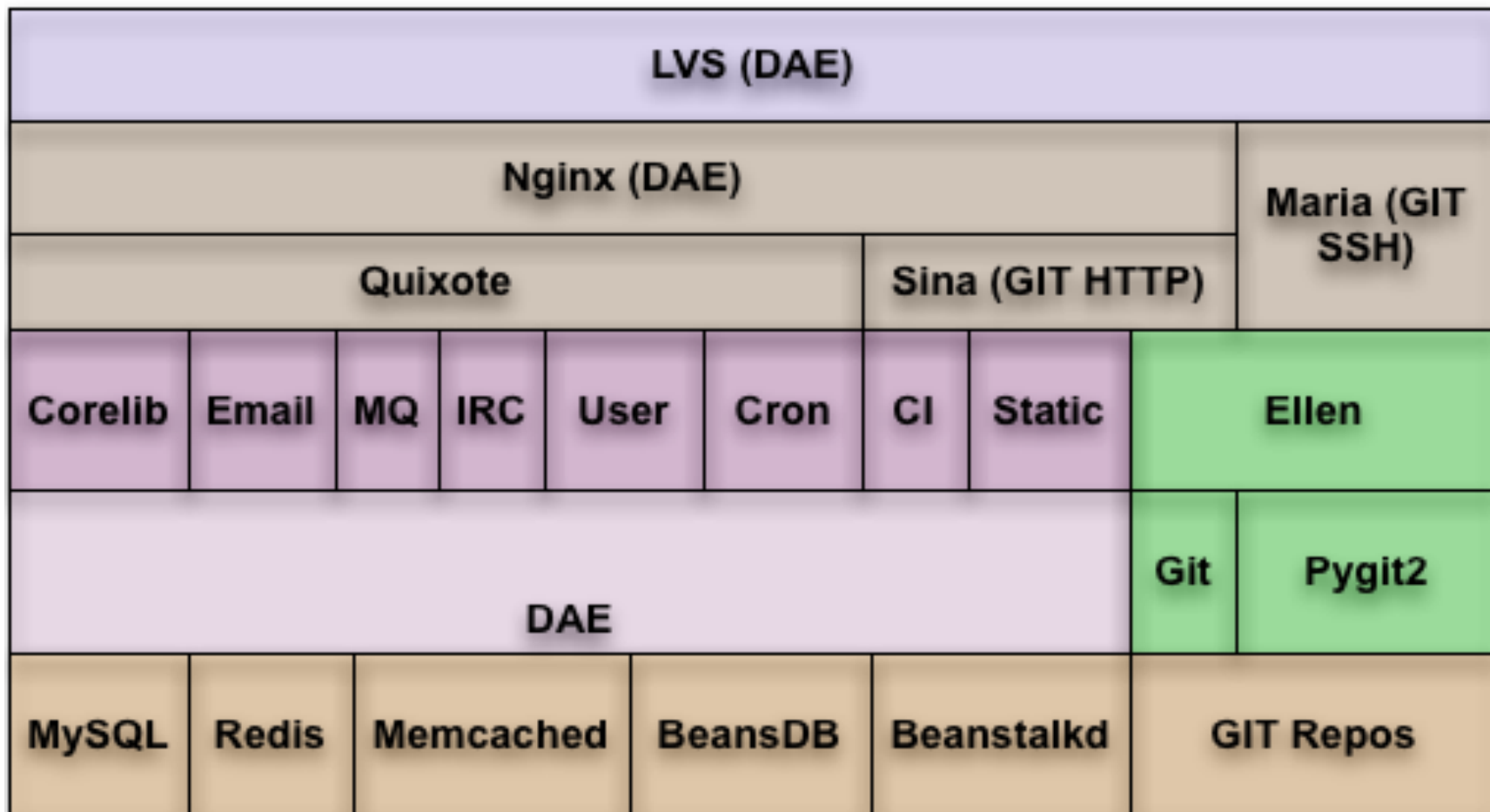
+ 概述 - DAE



DoubanAppEngine

- DoubanAppEngine
- Python PaaS
- 动态 WEB 服务，支持异步/同步模型
- MySQL 数据库
- 永久文件存储系统
- 自动负载均衡
- API: Email, Sentry, IRC, Memcached, Beanstalkd, Redis

+ 概述 - 架构



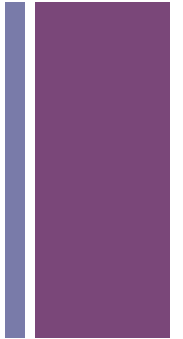
+ 概述 - 前端

- OzJS
 - OzmaJS
 - grunt-ozjs
 - grunt-dispatch
 - grunt-furnace
- JQuery
- Grunt
- Bootstrap 2
- FlatUI

- CodeMirror
- ZeroClipboard.swf
- D3.js
- mustache.js
- socket.io.js
- keymaster.js
- jquery.atwho.js
- jquery.caret.js
- jquery.commits-graph.js
- jquery.forms.js
- jquery.lazyload.js
- jquery.media.js
- jquery.pjax.js
- jquery.scrollTo.js
- jquery.timeago.js
- jquery.zclip.js
- mousetrap.js



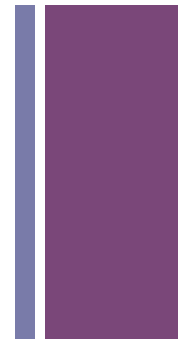
+ GIT



- HTTP – 上传/下载代码 (git clone, pull, fetch, push)
- SSH – 上传/下载代码 (git clone, pull, fetch, push)
- WEB – 代码查看, 代码管理, 代码 Review (commit, blob, tag, reference, diff)



GIT HTTP



- `git_http_backend.py`
 - 扩展起来比较头疼
 - `SubprocessIO`，为了实现 `Subprocess` 的 `Chunk`
 - `Subprocess` 完全不兼容 `Gevent` (通信数据量比较大的情况)
 - 尽管 `Gevent` 有一个 `monkey_patch`
 - 所以对待大量 `Fork` 进程的调用方式，很难事件化
 - 最好的方式仍然是独立处理 `Fork` 出来的进程，与 `Web` 进程脱离
 - `GIPC` 尝试为 `Gevent` 管理子进程，但仍然有限制
- `Sina` (GitHub: [douban/sina](https://github.com/douban/sina))
 - 仍然没有解决异步化
 - 代码清爽了，方便做扩展

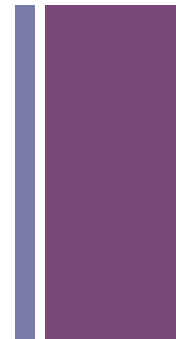
+ GIT SSH



- Maria (GitHub: CMGS/ maria)
 - SSH 支持做的比较晚
 - 基于 Gevent 和 Paramiko
 - 著名的 Python 攻城师 @CMGS 出品

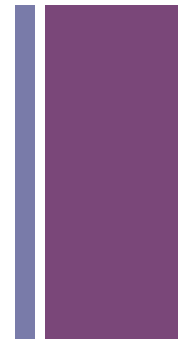


GIT WEB



- **Pygit2 (GitHub: libgit2/pygit2)**
 - 不是很成熟
 - 提交了不少 PR
 - 我们做了一点点 hack
 - @alexband 紧跟 libgit2 的最新版
- **Libgit2 (GitHub: libgit2/libgit2)**
 - 不是很成熟
 - 提交过补丁
 - 不过近期没有跟进
- **Ellen (GitHub: douban/ellen)**
 - 封装了 CODE 的常用命令
 - 混杂着 GIT 命令和 Pygit2

+ GIT 分布式



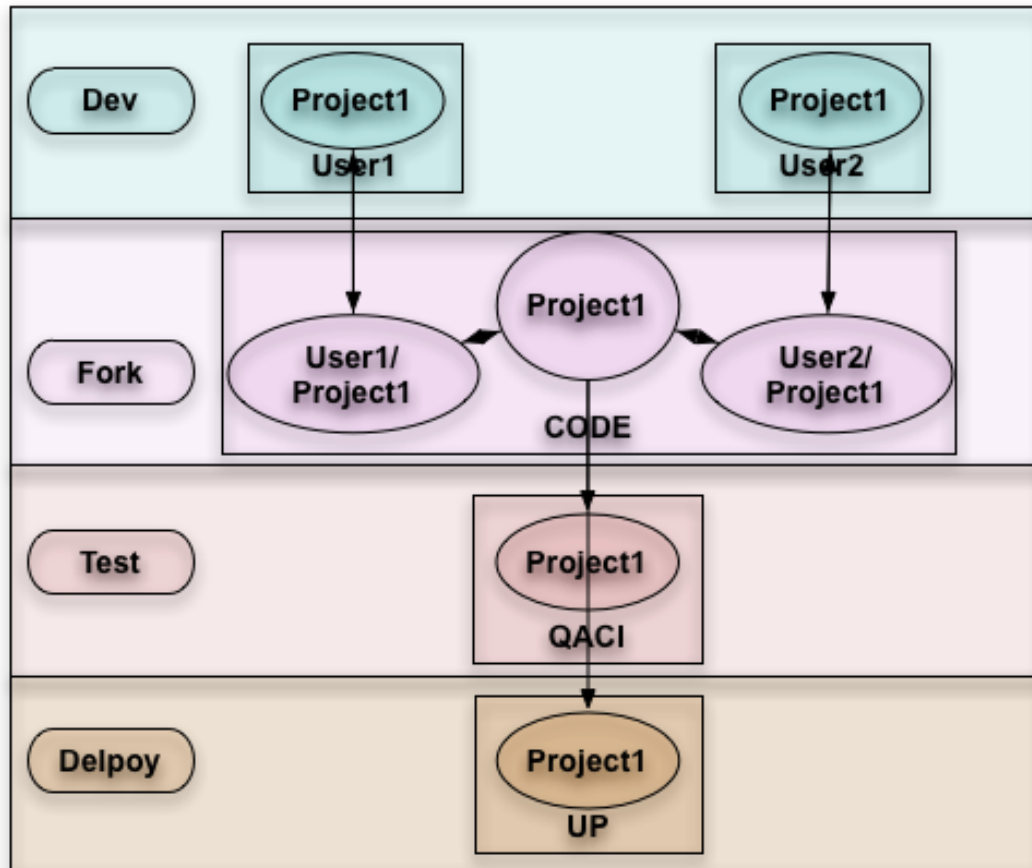
- @tclh123 同学曾经做过一些工作
- 很遗憾，现在未能继续下去，因为一些历史包袱原因
- 目前只是个 DEMO 的状态
- Tproxy + Thrift + Ellen
- 期望后续能够完善这项工作
- GitHub: [tclh123/code-distribute](https://github.com/tclh123/code-distribute)

+ GIT 未来

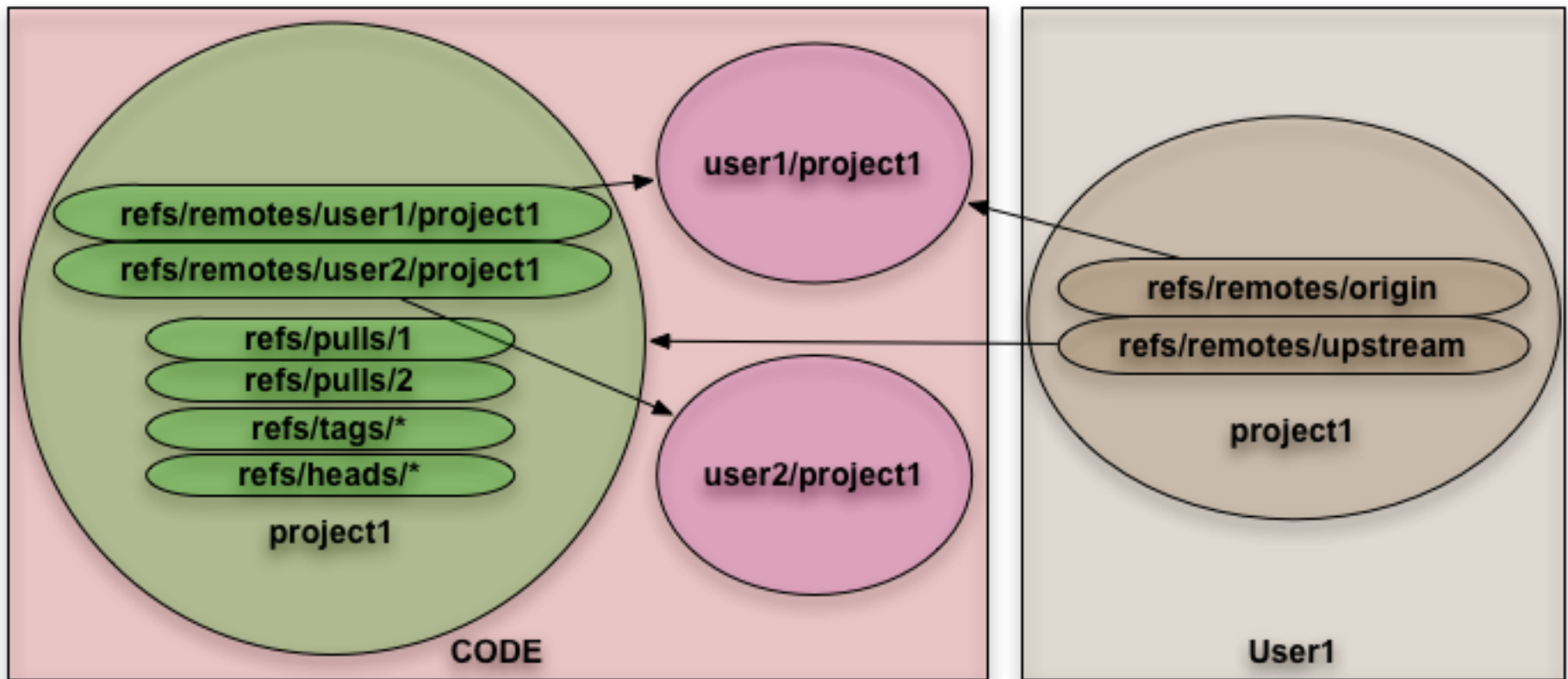


- 异步化，依赖 `libgit2`
- 后端存储，`NoSQL`，`RAM Storage`
- **GIT** 其实也有很多不足的地方，但是他很好用
- 解决了很多开发中经常发生的问题
- 关于 `Libgit2` 的想法 [GitHub: xtao/git-future-slides](https://github.com/xtao/git-future-slides)

+ Flow

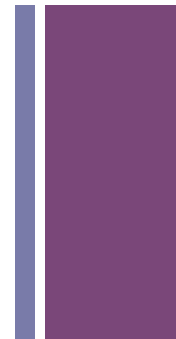


+ Pull Request 的实现





GIT Hooks



- 客户端每次 **Push** 会触发服务器端仓库的 **Hook (post-receive)**
- 入口(**Sina, Maria**) 通过环境变量传递用户名给 **Hook (post-receive)**
- **Hook (post-receive)** 触发 **CODE** 的 **API**
- **CODE** 收到仓库更新的消息

+ Web 应用数据存储原则

- 数据缓存使用 **Memcached**
- 大的文本数据，需要持久化的使用 **Beansdb**
- 结构化数据使用 **MySQL**
- **SQL** 禁用 **Join** 等跨表语句，尽量使用缓存，减少 **MySQL** 服务器的压力
- **The Missing Data Manager In Douban**
- 一个豆瓣的最佳实践，很遗憾 **@codetender** 因工作关系，没有继续开发下去
- **GitHub: douban/douban-orz**

+ Quixote



- Don Quixote
- 基于目录的 **URL** 分发
- 在目前这个正则匹配的时代显得有些过时
- 豆瓣的主站仍然用的是这个

+ gevent



- 神坑
- **CODE** 因为使用了大量 **fork** 操作，触发了 **gevent** 的 **bug**
- 所以这种情形下还是不要用 **gevent** 了
- 本身 **DAE** 用了 **greenify** 异步化了 **MySQL** 等底层 **IO** 操作
- **Libgit2** 最快的异步化方案，配合 **gevent**，还没有这么做 ...
- **@windreamer** 神作，虽然有那么一点 **hack** ...
- **GitHub: douban/greenify**



CODE 衍生项目

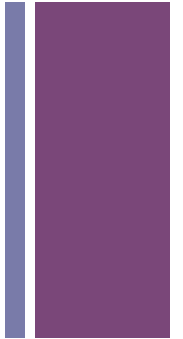
在什么地方用了什么

+ CodeCLI



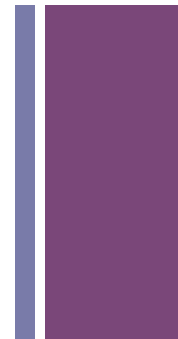
- @hongqn 教授一早参与了 **CODE** 的开发
- 在做完原型就脱离了 **CODE** 的开发
- 不过他写了一个客户端给 **CODE** 用
- 极大方便了工程师的日常工作
- 尽管仍有工程师没有使用的这个
- 期望他们看到这个**PPT**的时候能够意识到 ...
- **GitHub: hongqn/codecli**

+ User GIT Flow



Clone	code fork project user/project	git clone http://project.git project
		git remote add upstream url
Edit	code start dev	git checkout -b dev upstream/master
	vim xxx	vim xxx
	git commit -a	git commit -a
PR	code pr	git -u push origin dev
Pull	code sync --rebase	code pull upstream/master dev

+ Page & Doc



- @guibog 给 CODE 添加了一个类似 readthedocs 的功能
- 给 CODE 做了很多重构
- 然后就隐去了
- CODE 是支持 Page 和 Doc 的
- 像 DAE 的文档就是 rst, 用的 Doc 功能
- 很方便

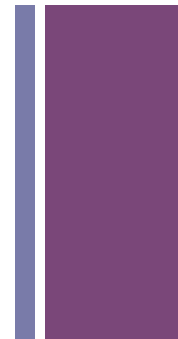
+ Linguist

- 作者 @liluo 为人低调不愿意出来讲，不过这里面还是有不少故事的
- <http://liluo.org/blog/2013/12/douban-linguist/>



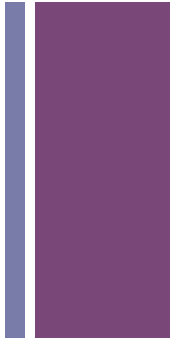
Github 现在的版本

+ Scanner



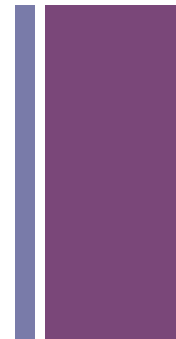
- Linguist 周边之一
- PHP 和 Ruby 都是用的鬼车（oniguruma）
- 做了个 binding
- 提升 linguist 的正则匹配速度
- GitHub: [cuteio/scanner](https://github.com/cuteio/scanner)

+ Mime

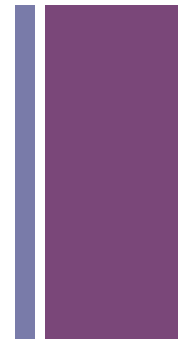


- **Linguist** 周边之二
- 检测文件类型
- 移植 **Ruby** 到 **Python**
- **GitHub**: [liluo/mime](https://github.com/liluo/mime)

+ Misaka & Mikoto



- Markdown 渲染
- 这里有个故事
- 开始是 `vmg/sundown`，然后 deprecated 2 年，终于 `jgm/CommonMark`
- 一开始叫 `standard markdown`，`markdown` 作者不干了，只好改名 `common markdown`
- <http://talk.commonmark.org/t/standard-markdown-is-now-commonmark/434>
- GitHub: `qingfeng/mikoto`
- GitHub: `qingfeng/misaka`



- @qingfeng 发起的项目
- 文件存储，主要用来存图片
- **CODE** 上的图片都是存储 **P** 上
- 通过 **Markdown** 渲染图片链接
- **CODE** 本身没有图片存储功能
- Eg: ![图片](<http://p.dapps.douban.com/i/xxx>)
- GitHub: qingfeng/p



CODE 总结

一些感想

+ 一些数字

Project

TOTAL

7,345

FORK

4,902

Gist

TOTAL

1,232

FORK

61

Pull Request

TOTAL

58,804

COMMENT

99,545

OPEN

140

CLOSED

58,664

Issue

TOTAL

16,365

COMMENT

18,043

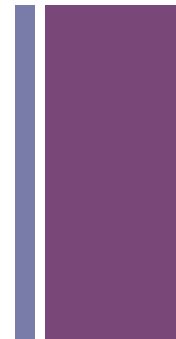
OPEN

12,951

CLOSED

3,414

+ Who

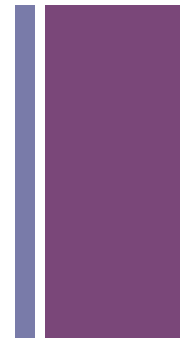


■ 谁在用 CODE ?

- 工程师们
- HR 有仓库哦
- PM 偶尔会冒个泡
- 其实大家都可以用的

■ 谁写的 CODE ?

- 一些核心开发者 (10个以内)
- 实习生们
- 热心的使用者也会经常贡献代码
- 绝对零散的开发模式
- 只有过一个全职开发者



新兵训练营

基础

- 学习 HTML & CSS
- Shire 之旅
- Python 教学 - 上
- Python 教学 - 下
- MySQL
- DAE 环境
- Quixote 教学
- Introduction SA & DBA
- 如何写出好代码

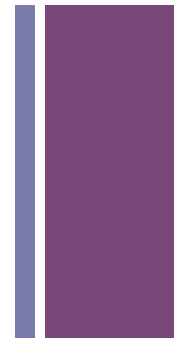
进阶

- 电影代码拆分
- FM 算法实践
- 持续集成在豆瓣读书的应用

Chat Room: [Bootcamp](#)

Projects(41) [Add Project](#)

- chetwood
- 新兵训练营
- yangxiufeng/notepad
- shibingwen/npcc
- notpad.cc 喵了个咪~...
- liutianwei/dailynote
- notepad
- zhuzhaolong/exam
- 基础课考试!!! ...
- zhangjiangge/seekpad
- daiximing/notepad
- notepad
- jiangqijiong/notepad



liwanjin / chetwood forked from chetwood

[New Issue](#)[Watch](#)

1

[Fork](#)

0

[Code](#)[Docs](#)[Pull Requests](#) 0[Issues](#) 0[Graphs](#)

新兵训练营

HTTP

branch: **master****Files**[Commits](#)

Latest commit to the **master** branch

[Edit README.md](#) **liwanjin** authored 22天前

commit 30c8f636f1

liwanjin/chetwood /



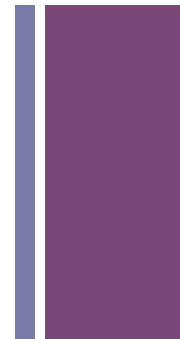
name	age	message
step1	3 months ago	Initial commit. [xutao]
step2	3 months ago	Initial commit. [xutao]
.gitignore	3 months ago	Initial commit. [xutao]

+ Markdown + GIT

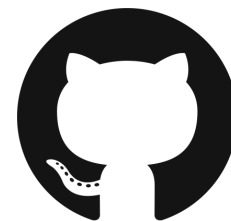


- 大家都可以用
- 文件是需要版本管理的
- 文本文件也需要规定格式
- **Markdown** 和 **GIT** 结合很好的管理了起来
- 只要有一个好的界面，就是一个大家都可以用的 **Wiki** 系统
- **Markdown** 谁都可以用
- 简单的保存，编辑，查看历史也是大家都能学会的操作

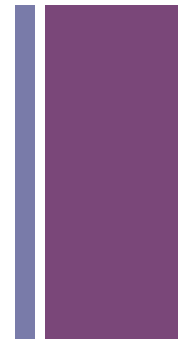
+ 相关工作



- GitHub
- GitHub Enterprise
- Bitbucket
- Stash (Bitbucket)
- Gitlab
- Gerrit (Android)
- GitCafe
- GitCafe Enterprise
- Coding.net



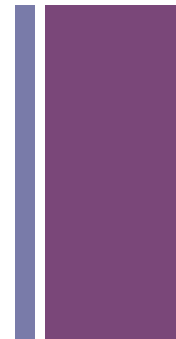
+ CODE 的开源问题



- 客观
 - 人手有限
 - 精力有限
 - 历史包袱有点多
- 主观
 - 尝试了一些过于先进的东西
 - Ember.js
 - Backbone.js
 - 但是我们后端经验更多一些
 - 一开始没有想好
 - 过于激进
 - 核心开发人员散去

+ 但是

- 已经有一个样子
- 脱离 **DAE**
- 重写了很多代码
- 目前在这个架子上填充代码中
- 仍然只是期望，但是近了一些



FAQ

Thanks

+



ruby@douban.com