



哈爾濱工業大學
HARBIN INSTITUTE OF TECHNOLOGY

哈尔滨工业大学软件学院2023年秋季学期

开源软件开发实践

徐汉川

xhc@hit.edu.cn

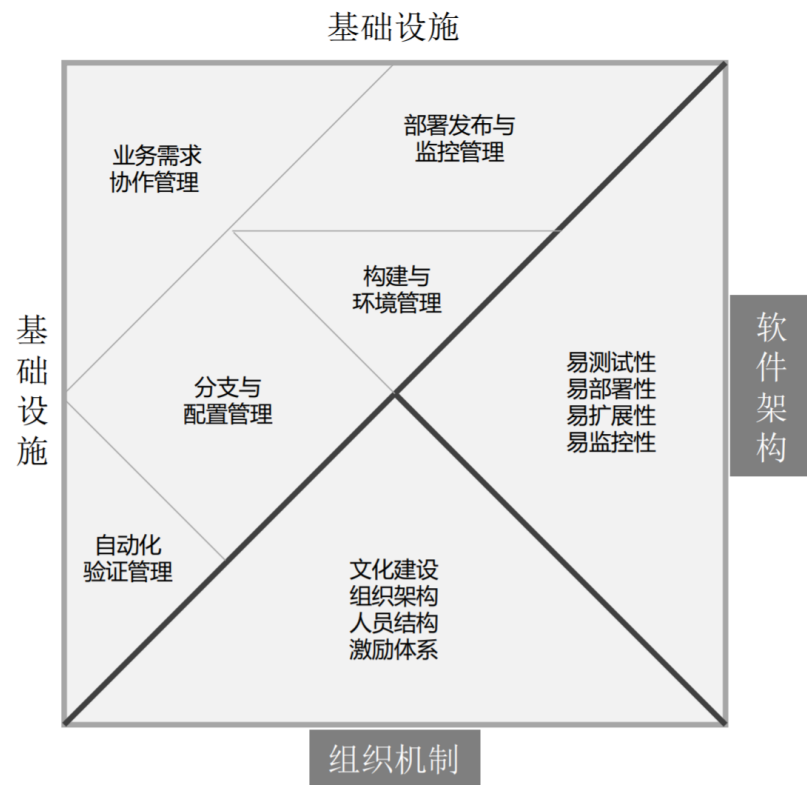
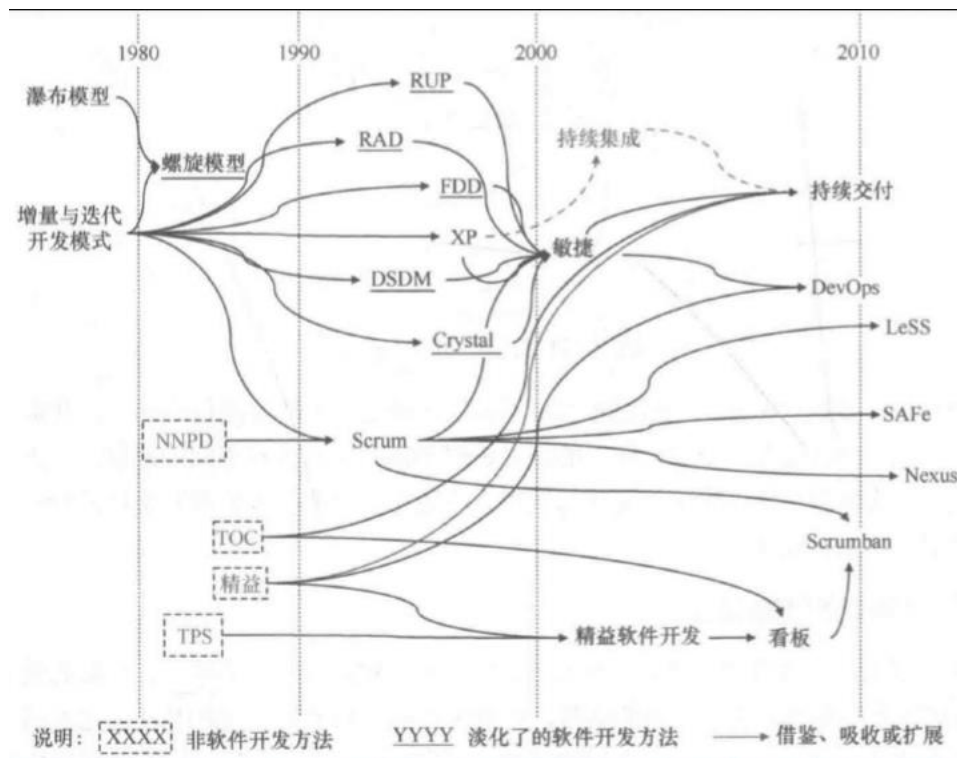
September 29, 2024



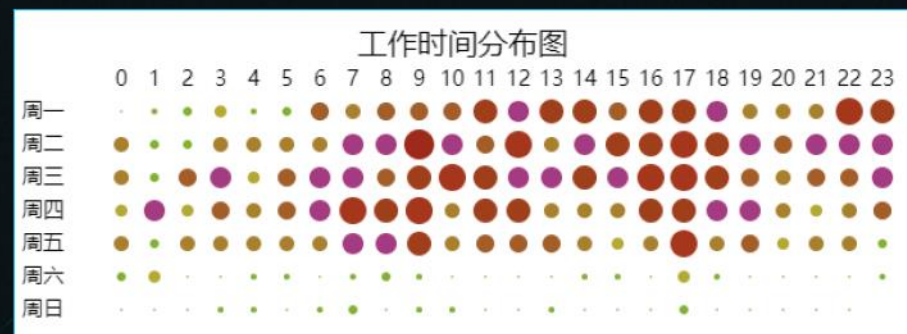
开源软件开发协作流程



软件工程方法的发展



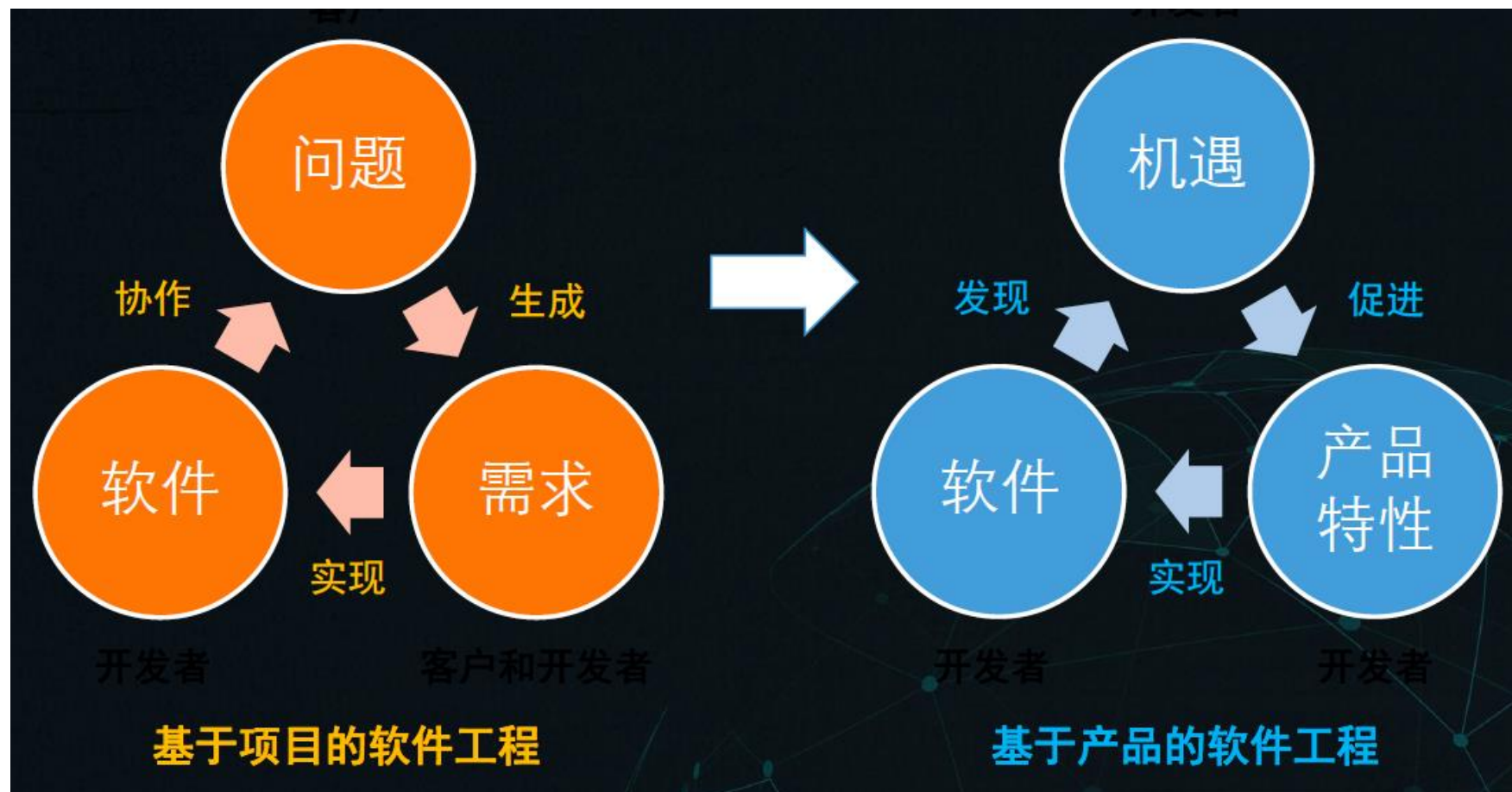
软件开发的全球化趋势



Apache Hadoop-ozone项目的全球化程度

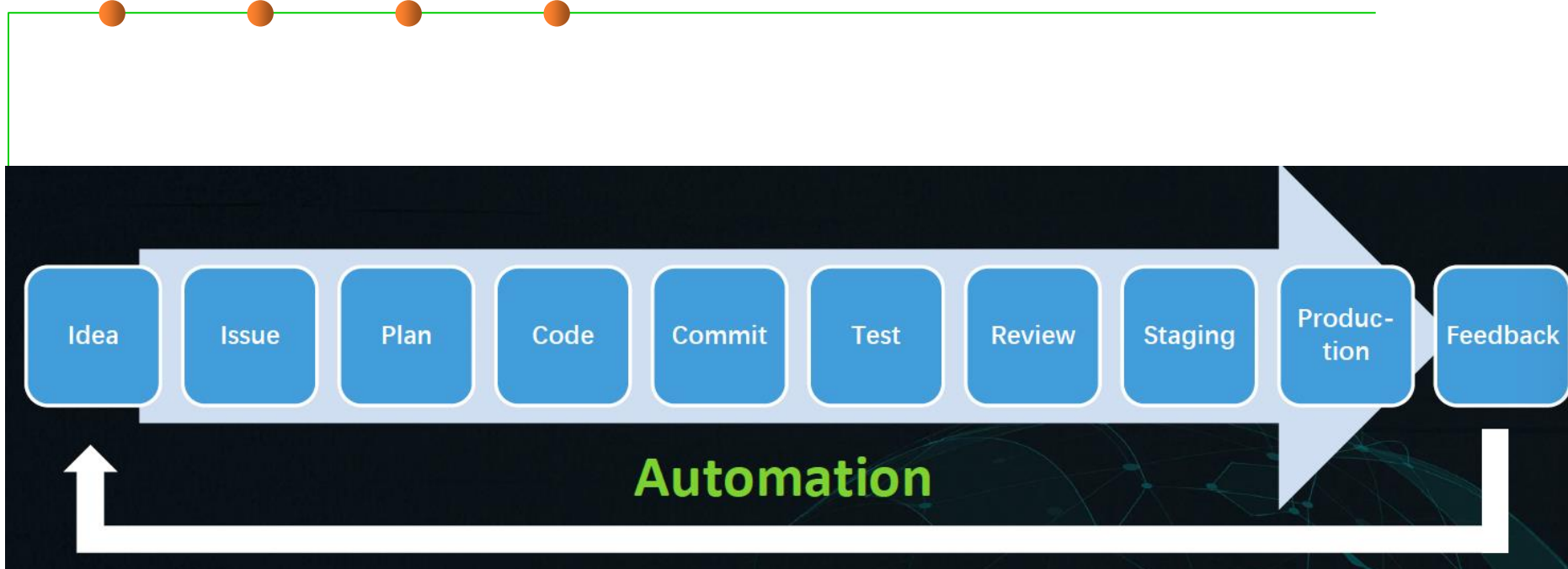
来源：GitHub 2020 数字洞察报告, 2021-02.

从软件项目到软件产品的快速转变



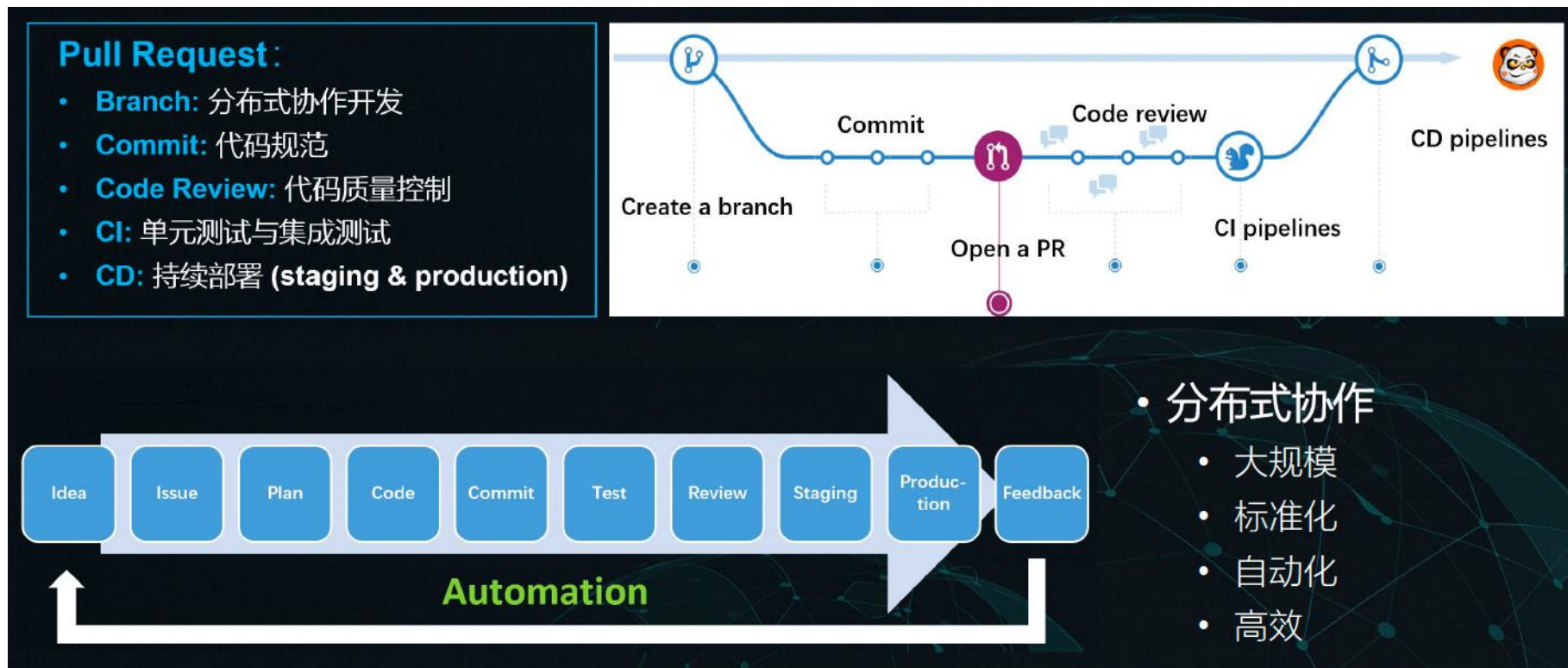
图片来源: <https://github.com/X-lab2017/oss101>

研发工作流的标准化与规范化

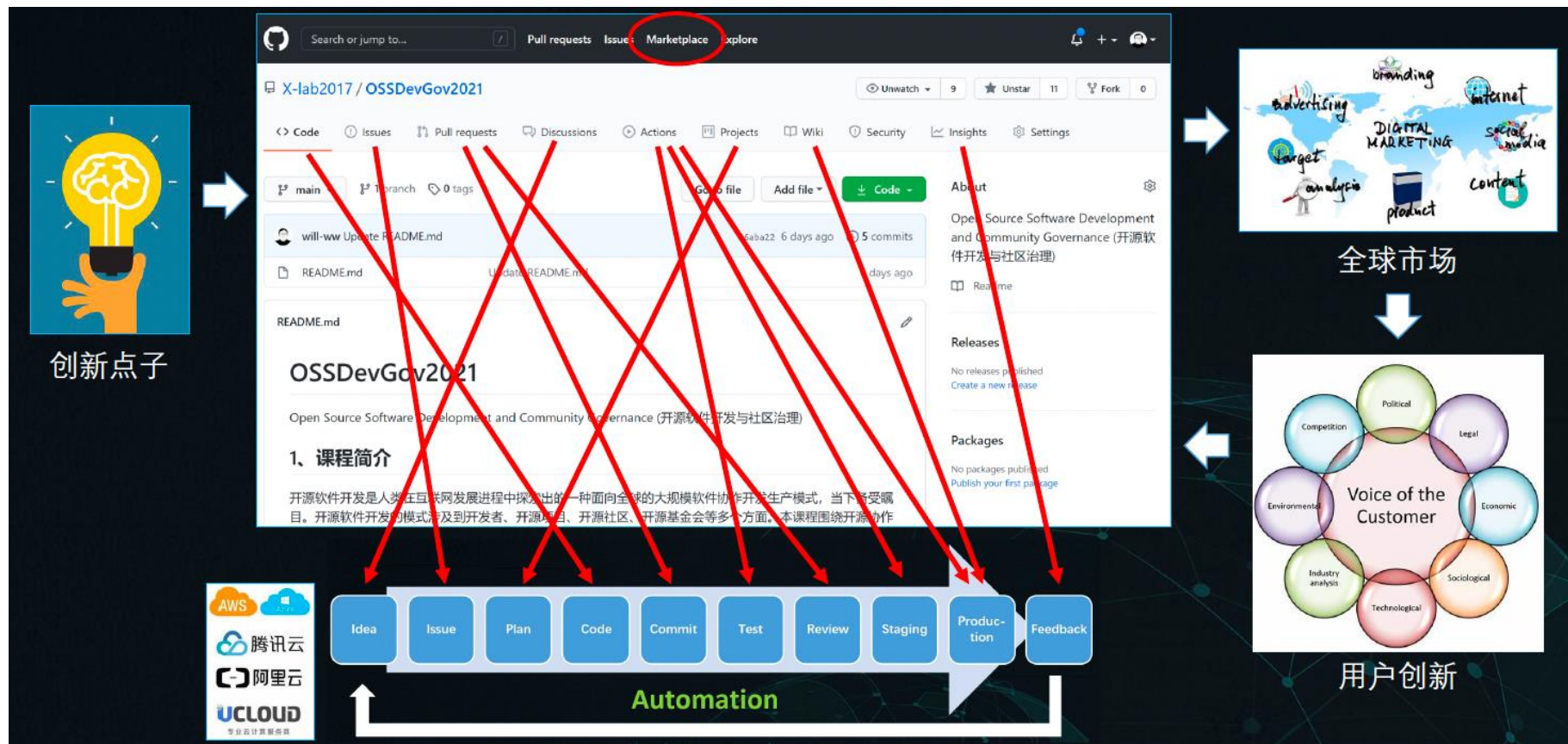


图片来源: <https://github.com/X-lab2017/oss101>

开源软件开发过程是一种最佳工程实践



开源软件开发的协作与平台化



图片来源: <https://github.com/X-lab2017/oss101>

开源协作

- 开源
 - 知识成果的开放、共享、创新
- **Linux**早期协作
 - 1991-2002 邮件列表+Patch
 - 2002-2005 BitKeeper
 - 2005 Git
- 现代开源协作
 - Git
 - 中心化源代码托管平台
 - 1999 Source Forge (SVN, CVS)
 - 2005 Google Code (2014停止新项目创建)
 - 2008 GitHub (Issue + Pull Request)

Git基本操作

■ Git 是一个分布式版本控制系统

Fork repo to self account

```
git clone git@github.com:frank-zsy/opensource.git
```

```
git checkout -b feat/add-frank
```

```
git add .
```

```
git commit -sm 'feat: add frank name into file'
```

```
git push origin feat/add-frank
```

```
git checkout master
```

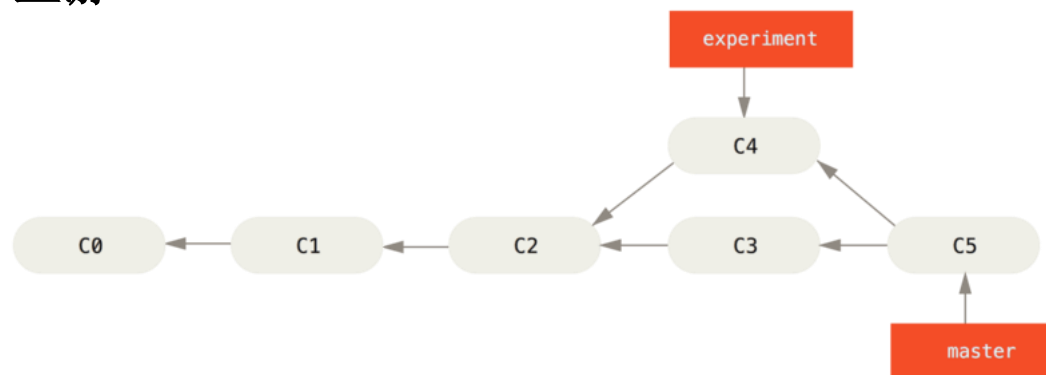
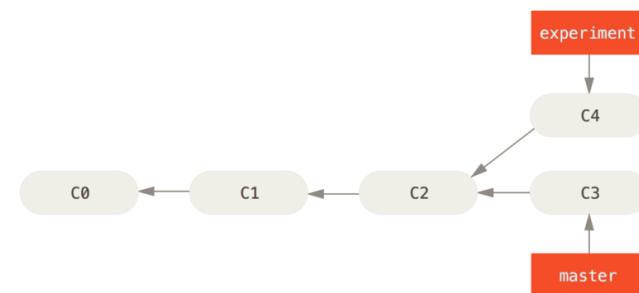
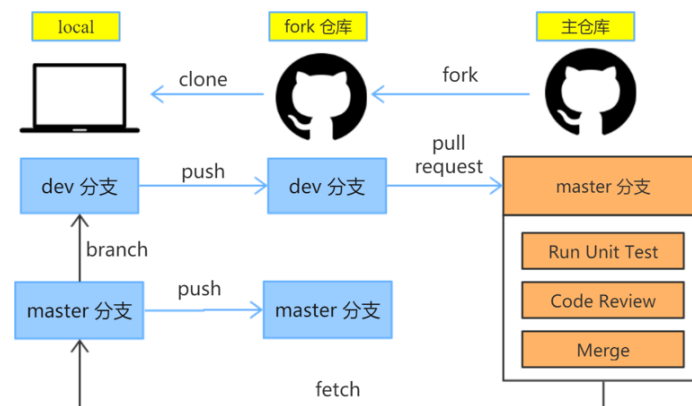
```
git fetch upstream master
```

```
git rebase upstream/master
```

```
git push origin master
```

同步上游

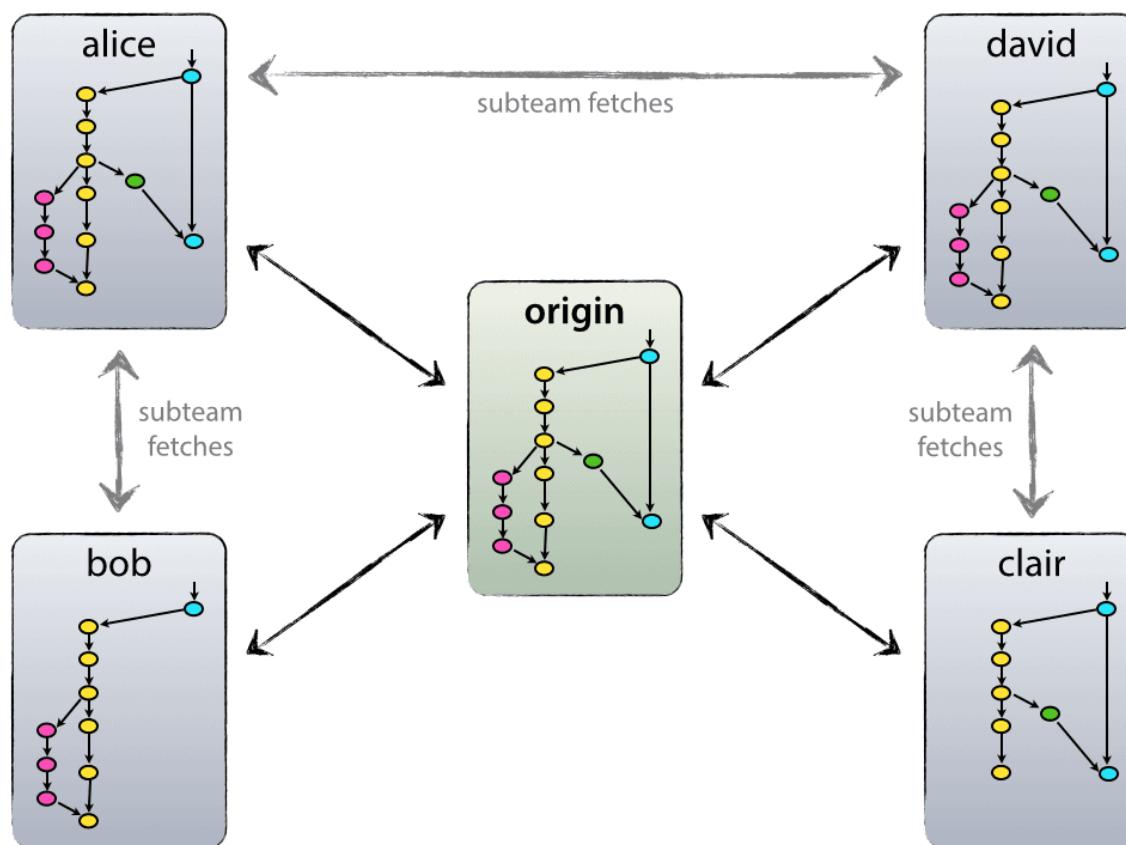
Create pull request on GitHub



Git分布式

- Git 是一个分布式版本控制系统

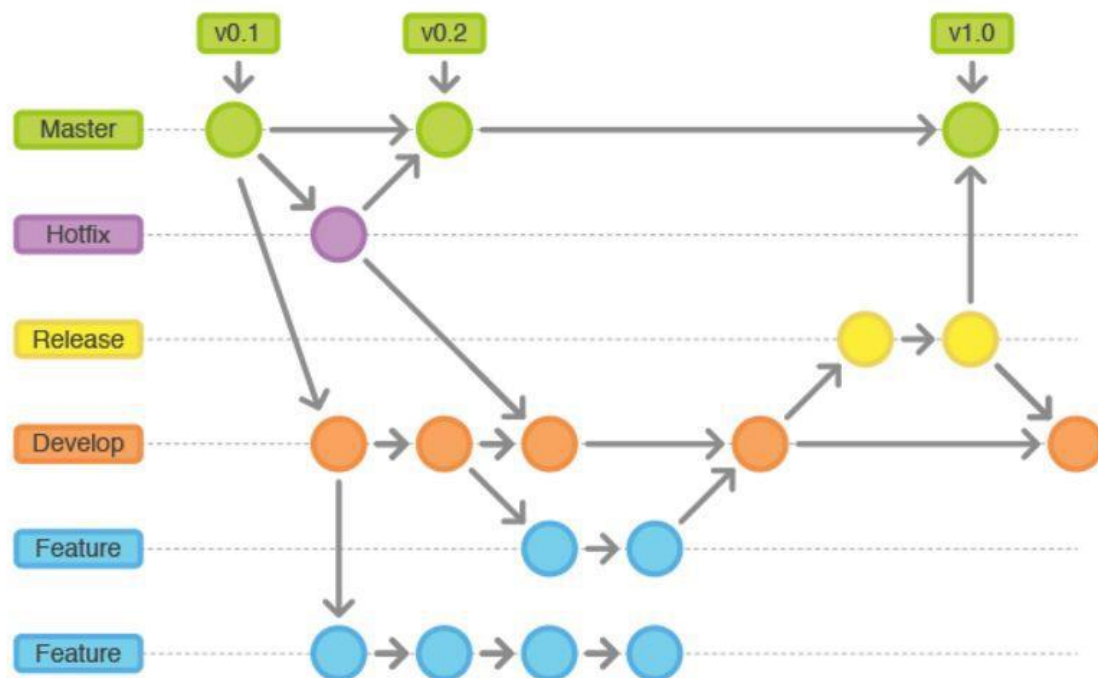
即便GitHub让仓库看起来是中心化的，但本质上依然是分布式的仓库，所有开发者之间可以互相协作，不一定需要依赖原仓库



Git分布式

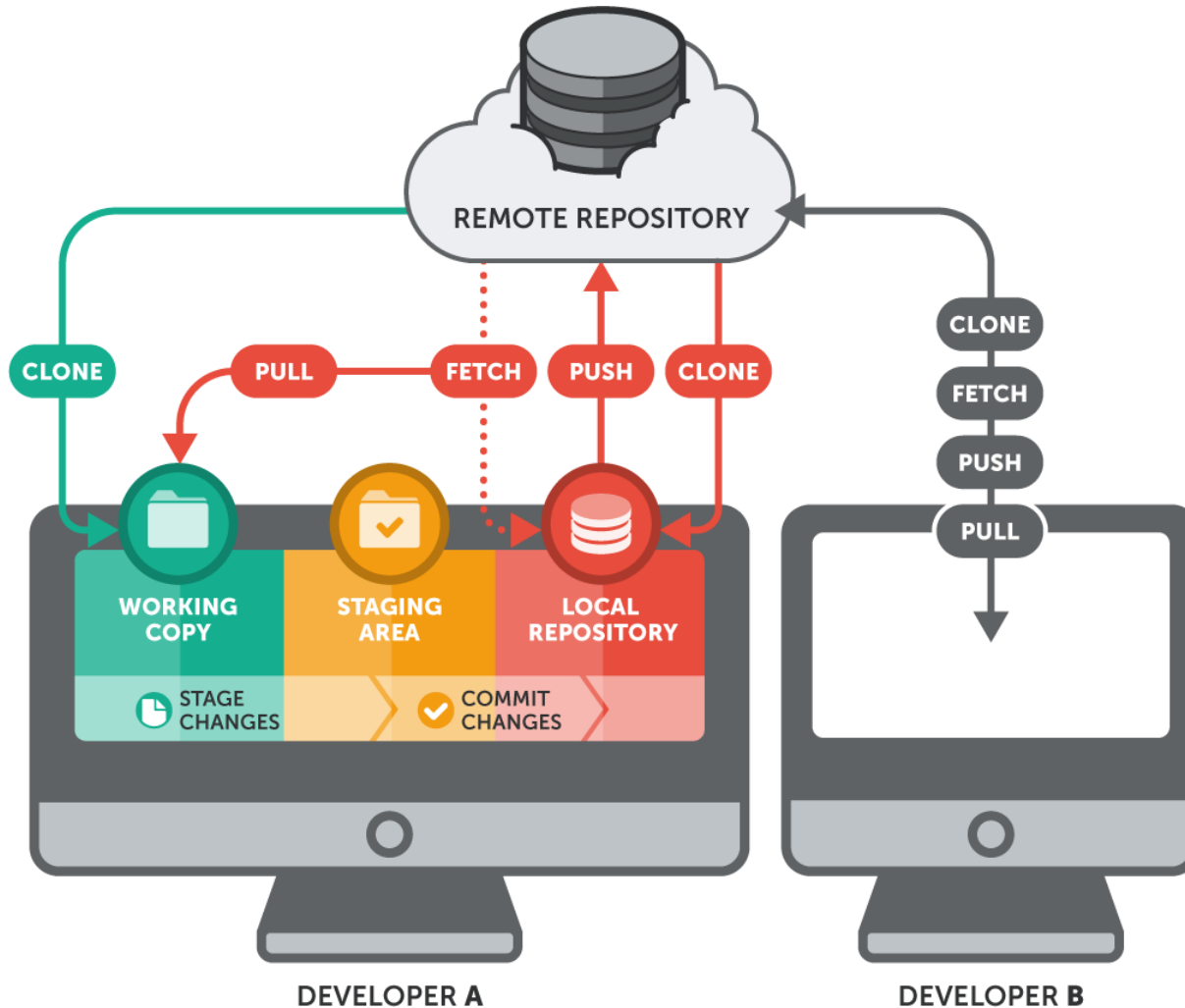
■ 对于分支(branch)的使用

- 每次有新feature开发时都启用一个新分支
- 若均使用master分支开发：
 - feature1开发到一半需要紧急开发feature2



Git supports collaboration

- Local repository and Remote Repository



开源协作方式

- 邮件列表
- 问题跟踪系统: **bugzilla**
- **Wiki**多人协作知识库
- 即时通讯工具
- 博客/论坛
- **Issue**
- **Pull Request**
-

GitHub issue

■ Issue的构成

- 标题Title
- 内容Body
- 评论Comments
- 标签Labels
- 指派者Assignees
- 所属项目看板Project
- 里程碑Milestone

■ Issue的典型功能

- 提问、交流
- 提Bug
- 提需求

container has runAsNonRoot and image will run as root warning should include container name #94898

New issue

Open grosser opened this issue 14 hours ago · 4 comments · May be fixed by #94911



grosser commented 14 hours ago



when deploying a container that has runAsNonRoot but runs as root this error appears, but for pods with lots of containers it can be confusing which container failed, especially if there are 3-rd party webhooks that inject containers please add the container name to the message, making it container "foo" has runAsNonRoot and image will run as root



grosser added the **kind/feature** label 14 hours ago



k8s-ci-robot added the **needs-sig** label 14 hours ago



grosser commented 14 hours ago

Author 🗨️ ⋮

/sig api-machinery



k8s-ci-robot added **sig/api-machinery** and removed **needs-sig** labels 14 hours ago



wawa0210 commented 7 hours ago

Contributor 🗨️ ⋮

/assign

I can try this fix



k8s-ci-robot assigned wawa0210 7 hours ago

Assignees

wawa0210

Labels

kind/feature **sig/api-machinery**

Projects

None yet

Milestone

No milestone

Linked pull requests

Successfully merging a pull request may close this issue.

🔧 Enhance log information of verifyRunAsNo...

Notifications

Customize

🔔 Subscribe

You're not receiving notifications from this thread.

3 participants



GitHub issue

■ 异步沟通

■ 优点:

- 跨时区合作
- 个人开发者时间成本低

■ 挑战:

- 总体沟通时间成本高
- 需要尽量减少沟通的回合数
- 需要尽量在一次沟通中表达尽量多的内容

■ 信息沉淀

■ 优点:

- 所有信息都沉淀下来, 可检索
- 沉淀信息可为后来人提供有效的查询路径



GitHub issue

- 如何降低总体的沟通成本
- 开发者
 - 提Issue之前先搜索一下有无相关历史Issue
 - 提交的内容的尽量详实充分
 - 例如如果是Bug，说明运行环境、参数配置等
 - 最好可以提供日志或最小可复现仓库
- 维护者
 - 维护良好的社区沟通环境CODE_OF_CONDUCT.md
 - 提供有效的沟通指引CONTRIBUTING.md
 - 使用模板指引开发者提供更全面的信息
 - 有效的流程设计使Issue可以被更快的响应
 - 制度与流程、自动化

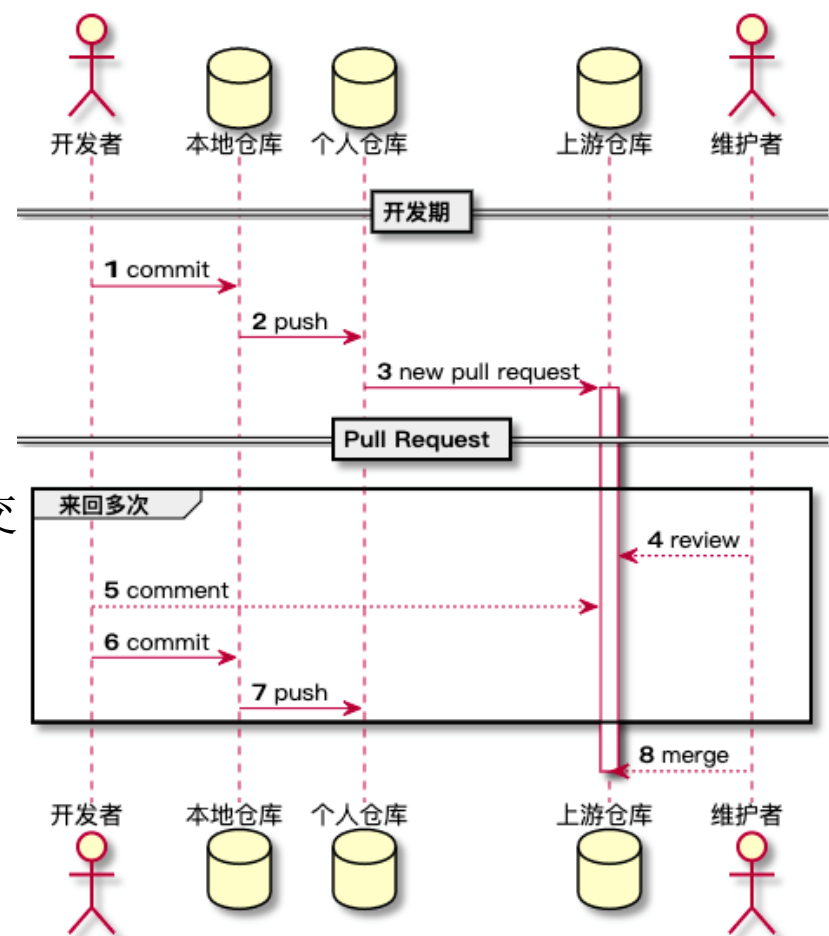
GitHub issue

- 管理的主要目的：
- 在现有条件下，通过合理的组织和配置各种资源，提高生产力水平。
- 异步沟通→减少同步沟通而减少核心维护者的时间成本
- 信息沉淀→减少开发者搜寻解决方案的成本和维护者反复回答的成本
- 流程清晰→减少决策的心智成本同时降低响应时间从而使开发者有更好的体验
- 流程自动化→减少组织的制度成本而使整体项目可以更加顺滑的运转

GitHub Pull Request

PR协作的基本流程

- 1 开发者在本地仓库开发
- 2 本地仓库提交到GitHub个人仓库
- 3 从个人仓库向上游仓库提起PR
- 4 维护者ReviewPR
- 5 开发者与维护者讨论细节
- 6,7 需要修改则开发者添加新的修改并提交
- 8 最终维护者合入PR
- 自动化测试?



GitHub Pull Request

■ PR的提交与Review

■ 开发者

- 提交PR时，除了提交文件commit外
- PR的标题与内容

■ 维护者

- 通过PR的标题与内容了解PR内容
- 确定PR中的文件修改是正确、合理的
- 不正确、不合理之处与开发者进行讨论
- 将PR合入

refactor: modify lua vm #352

 Merged menbotics merged 1 commit into hypertrons:master from frank-zsy:modify-lua

 Conversation 2  Commits 1  Checks 1  Files changed 6



frank-zsy commented 24 days ago

Signed-off-by: Frankzhaopku syzhao1988@126.com

Brief change log

- Refactor Lua VM to support async function in js

Checklist

- Are tests written for added code?
 - ☒ Yes
 - ☐ No because:
- Did you introduce any new dependencies?
 - ☒ No
 - ☐ Yes
 - Dependency:
 - Reason:

16 16


17 + ## Communication

18 +

19 + Click here to join the [slack communication group](https://join.slack.com/t/x-github-analysis/shared_invite/zt-gz8ykv3l-qP53IpajsZIGsx6dFo_8Lg) if you want to know more about the project and join it.

 frank-zsy 18 days ago Member 

Sorry, but I think Slack should be capitalized.

 frank-zsy 18 days ago Member 

And if you use Click here to join, maybe the url should be added on here but not on Slack communication group

 sunshinemingo 18 days ago Author Member 

OK

 Reply...

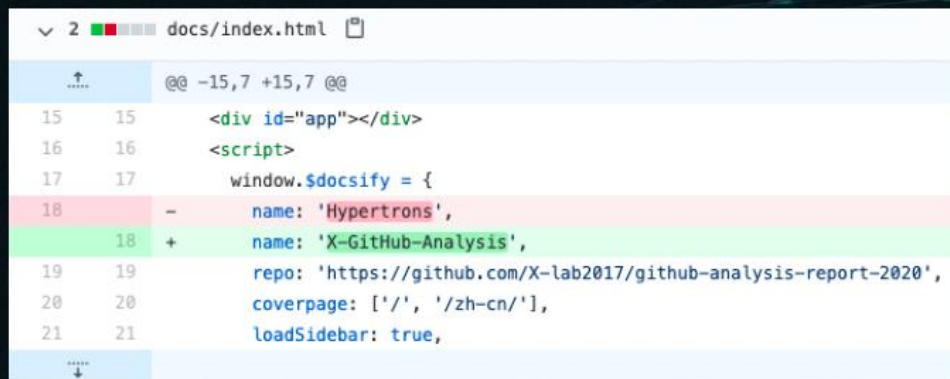
GitHub Pull Request

- **PullRequest** 到底是什么？开源到底是什么？
- 开源是一种大规模分布式异步协作生产软件的生产方式，是一种方法论
- **PR**是开源中由**GitHub**创造，运行在**git**协作流程之上的**Web**协作模式，用于替代**Linux**开发中使用邮件**patch**交付或多**remote**交互的流程。
- 问题1：为什么开源目前只用于软件生产，为什么可以协作生产代码，可以生产其他内容吗？
- 问题2：为什么**PR**流程是有效的，从**svn**到**git**到**PR**，到底是什么决定了一个流程的有效性？
- 问题3：基于问题2，我们是否可以解决问题1？

GitHub Pull Request

问题1：为什么开源目前只用于软件生产，为什么可以协作生产代码，可以生产其他内容吗？

- 文本化为大规模的协作和频繁修改迭代带来了可能
- 代码是软件的文本化表现形式
- 文本的修改是结果是一目了然的，加一字符删一个字符清清楚楚，而二进制的就……



```
docs/index.html
@@ -15,7 +15,7 @@
15 15     <div id="app"></div>
16 16     <script>
17 17         window.$docsify = {
18 -         name: 'Hypertrons',
18 +         name: 'X-GitHub-Analysis',
19 19         repo: 'https://github.com/X-lab2017/github-analysis-report-2020',
20 20         coverpage: ['/', '/zh-cn/'],
21 21         loadSidebar: true,
```



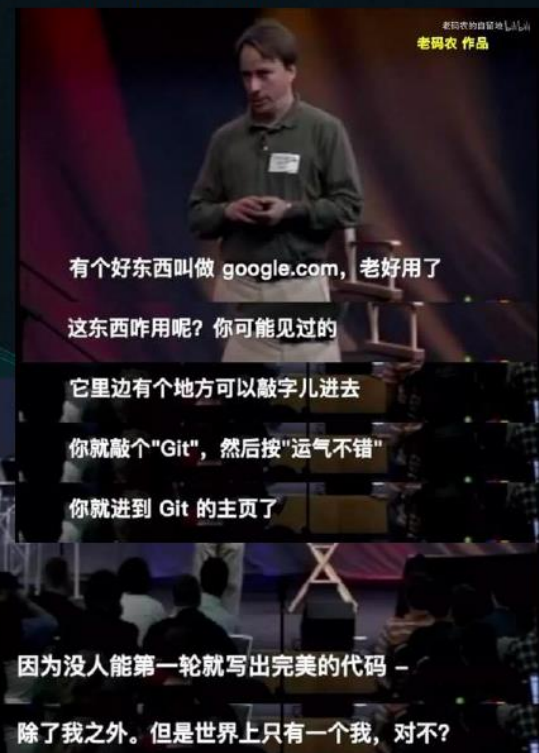
来源：<https://github.com/X-lab2017/oss101>

GitHub Pull Request

问题2: 为什么 PR 流程是有效的, 从 svn 到 git 到 PR, 到底是什么决定了一个流程的有效性?

- 为什么会有 git? (B站 git + linux) 真·分布式
- 为什么会有 PR? 更简单的可视化协作流程
- 它们在做什么?

协作规模越大则协作成本越高, 需要有流程机制的设计来控制成本



GitHub Pull Request

■ 问题3：基于问题2，我们是否可以解决问题1？

■ 需要控制哪些成本

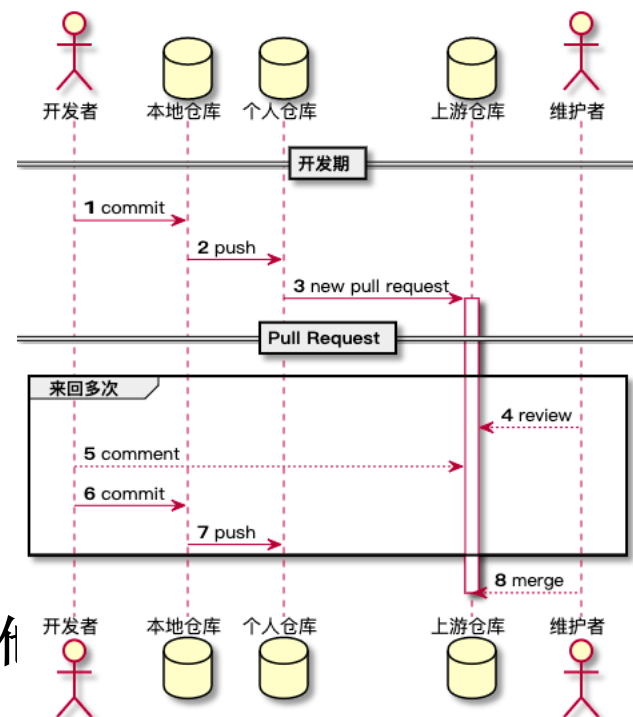
— 代码生产成本

— Review的成本

- 理解和发现修改内容的成本
- 对修改内容的正确性做出评判的成本
- 与开发者沟通的成本

■ 如果可以大幅降低上述成本，是否可以支持其他类型内容生产的协作呢？

答案是肯定的



GitHub Pull Request

- 书写PR遵循的规范
- <https://opensource.com/article/18/6/anatomy-perfect-pull-request>

开源项目的说明文件

■ 开源项目的通用结构

- 每个开源项目的人员组成和项目结构都是不一样的,但是每个开源项目都会遵循一些通用的结构. 了解这些通用的结构可以让你在不同的开源项目中切换的更加游刃有余. 一个开源项目一般都会有以下的文件:
- **README:** README.md 是一个项目总的介绍入口, 会在github的主页中展示, 一般会介绍项目是做什么的, 有什么功能, 如何启动和参与项目等.
- **LICENSE:** 每个开源项目都必须有一个开源协议 open source license. 想了解更多可以看阮一峰老师的 如何选择开源许可证.
- **CONTRIBUTING:** 对于想要参与开源项目的同学来说, 这个文档非常重要, 如何参与开源项目, 方式和途径有哪些这里会找到答案.
- 其他文件: 较为优秀和成熟的开源项目, 都会有tutorials, 使用示例, 详尽的文档等等, 来帮助新人更快的熟悉项目.



哈爾濱工業大學
HARBIN INSTITUTE OF TECHNOLOGY

GitHub 常用工具



GitHub设置

■ 创建SSH连接

<https://docs.github.com/en/authentication/connecting-to-github-with-ssh>

- 1 Open Terminal.
- 2 Paste the text below, replacing the email used in the example with your GitHub email address.

```
ssh-keygen -t ed25519 -C "your_email@example.com"
```

Note:

SSH keys

New SSH key

ssh

This is a list of SSH keys associated with your account. Remove any keys that you do not recognize.

Authentication Keys



MacAir

SHA256:0nCz8+QSMuah3wZkrkrPgX4NFGKN/r8MXoEBVtjRhso

Added on Nov 1, 2023

Last used within the last week — Read/write

Delete

Check out our guide to [generating SSH keys](#) or troubleshoot [common SSH problems](#).

Github desktop

■ Github desktop

The screenshot displays the GitHub Desktop application interface. At the top, the 'Current Repository' is 'OSSDP-Lab2' and the 'Current Branch' is 'main'. A 'Fetch origin' button indicates the last fetch was 3 minutes ago. Below this, a notification states: 'An updated version of GitHub Desktop is available and will be installed at the next launch. See [what's new](#) or [restart GitHub Desktop](#).'

The main interface is divided into three panes. The left pane shows the 'Changes' tab with '0 changed files' and a 'History' tab. The right pane shows the 'History' tab with a list of commits. The selected commit is 'in first 10' by HansXU, 15 hours ago, with 10 changed files (+788, -0). The diff view shows changes to 'Solution1.java' through 'Solution9.java'. The diff for 'Solution1.java' is visible, showing changes to the package imports and the main method.

Commit History:

- questions 11-20 (HansXU • 10 hours ago)
- in first 10 (HansXU • 15 hours ago)
- Initial commit (Hanchuan Xu • 15 hours ago)

Diff View:

```
@@ -0,0 +1,69 @@
1 +import java.util.HashMap;
2 +import java.util.Map;
3 +
4 +/**
5 + * @description:
6 + *
7 + * 给定两个整数，分别表示分数的分子 numerator 和分母 denominator，以 字符串形式返回小数 。
8 + *
9 + * 如果小数部分为循环小数，则将循环的部分括在括号内。
10 + *
11 + * 如果存在多个答案，只需返回 任意一个 。
12 + *
13 + * 对于所有给定的输入，保证 答案字符串的长度小于 104 。
14 + *
15 + * 示例 1:
16 + *
17 + * 输入: numerator = 1, denominator = 2
18 + * 输出: "0.5"
19 + * 示例 2:
20 + *
21 + * 输入: numerator = 2, denominator = 1
22 + * 输出: "2"
```

Good first issues

■ Good first issues


- <https://goodfirstissue.dev> 运行网站
- <https://github.com/DeepSourceCorp/good-first-issue> 代码网站




The screenshot shows the 'goodfirstissue' website interface. On the left, there's an 'ABOUT' section and a 'BROWSE BY LANGUAGE' section with buttons for various programming languages and their issue counts (e.g., Python x 123, JavaScript x 93, Go x 88). Below this is a green 'ADD YOUR PROJECT' button and a note 'A DeepSource initiative'. The main content area on the right lists several projects, each with a title, description, language, star count, last activity, and a 'Good First Issues' count in a pill-shaped button.

Project	Description	Language	Stars	Last Activity	Good First Issues
bitwarden / mobile	The mobile app vault (iOS and Android).	C#	4.17K	10 months ago	9 issues
neurodata / hyppo	Python package for multivariate hypothesis testing	Python	137	10 months ago	3 issues
NativeScript / NativeScript	Empowering JavaScript with native platform APIs. ✨ Best of all worlds (TypeScript, Swift, Objective C, Kotlin, Java). Use what you love ❤️ Angular, Capacitor, Ionic, React, Svelte, Vue and you name it compatible.	TypeScript	21.98K	10 months ago	10 issues
thepracticaldev / dev.to	For empowering community 🌱	Ruby	20.06K	10 months ago	2 issues
http4s / http4s	A minimal, idiomatic Scala interface for HTTP	Scala	2.36K	10 months ago	10 issues
matplotlib / matplotlib	matplotlib: plotting with Python	Python	-	-	10 issues

OpenLeader

- <https://open-leaderboard.x-lab.info/>
- <https://github.com/X-lab2017/open-leaderboard>




 OpenLeaderboard

An insight to the world of open source

对象: ☒ 公司 ☐ 项目 ☐ 开发者 指数: ☒ 活跃度 ☐ 影响力

地区: ☐ 中国 ☒ 全球 时间: Month 2022/04 详情: ☐

rank		Company	Activity	
		-	Microsoft	50545.47 ↗ 2111.37
		-	Google	27861.62 ↗ 3426.05
		-	Amazon	22167.18 ↗ 2344.34
4	-	Meta	12732.9 ↗ 1312.13	
5	-	HarshiCorp	8521.37 ↗ 543.43	
6	-	Alibaba	7837.85 ↗ 456.59	
7	-	VMWare	7671.36 ↗ 429.36	
8	-	Baidu	6349.55 ↗ 472.22	
9	-	IBM	5851.85 ↗ 770	
10	-	Ant group	4835.05 ↗ 83.12	

 OpenLeaderboard

An insight to the world of open source

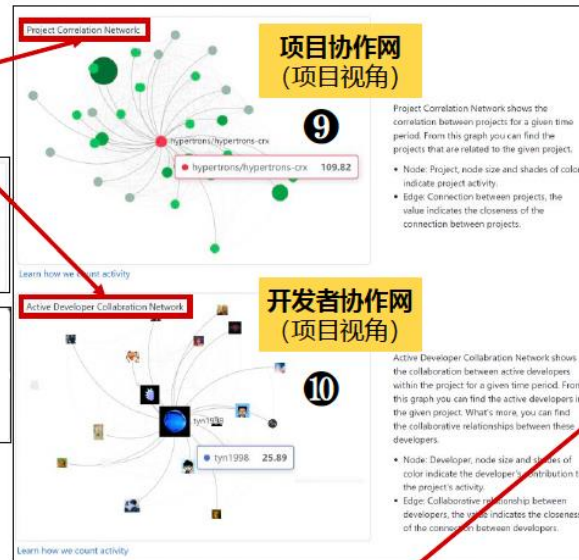
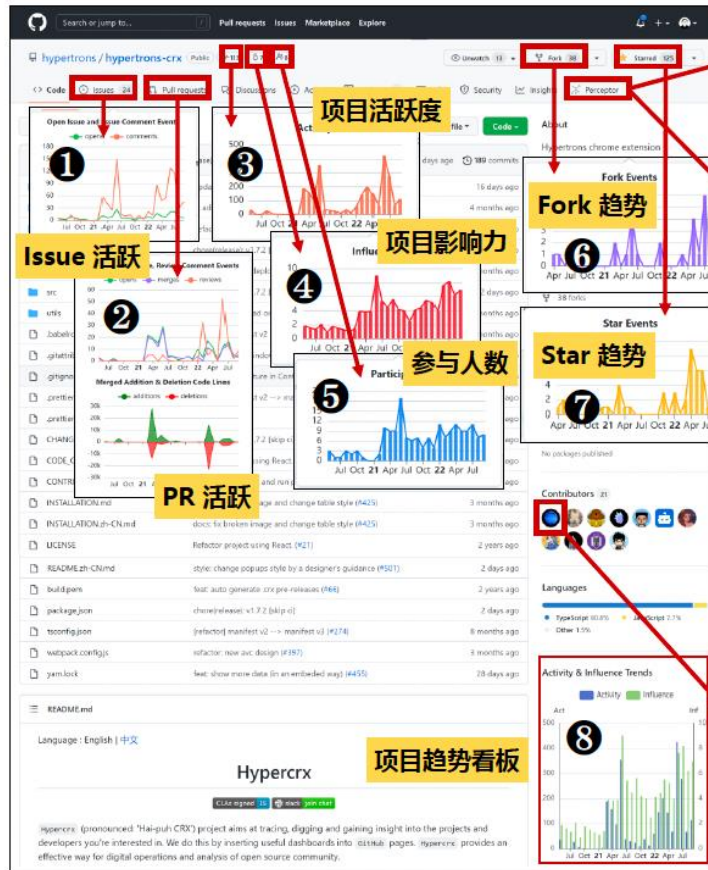
对象: ☐ 公司 ☒ 项目 ☐ 开发者 指数: ☐ 活跃度 ☒ 影响力

地区: ☐ 中国 ☒ 全球 时间: Month 2022/04

Rank		Repository	Influence	
		new	archway-network/testnets	3569.49
		↑727	element-fi/elf-council-frontend	2525.71 ↗ 2502.55
		↓1	microsoft/vscode	794.32 ↗ 42.60
4	↓3	solana-labs/token-list	757.62 ↗ 78.71	
5	↓2	home-assistant/core	744.64 ↗ 1.17	
6	↓1	flutter/flutter	618.48 ↗ 34.5	
7	-	NixOS/nixpkgs	611.34 ↗ 44.56	
8	↓4	MicrosoftDocs/azure-docs	606.14 ↗ 57.8	
9	↓3	facebook/react	459.13 ↗ 130.86	
10	↓1	dotnet/runtime	390.15 ↗ 5.76	
11	↑16	voidpet/bugs	381.54 ↗ 125.70	

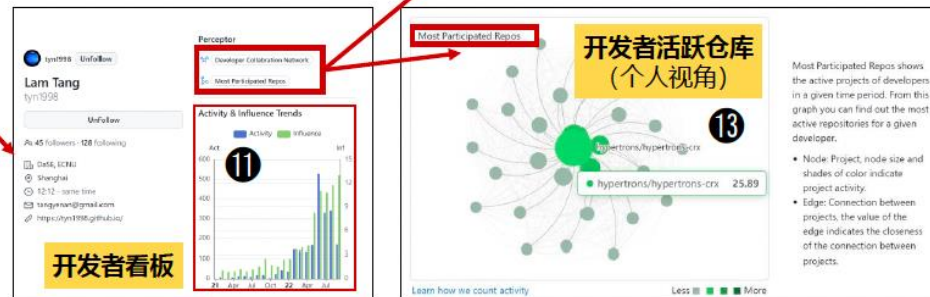
Hypercrx

■ <https://github.com/hypertrons/hypertrons-crx>



Hypercrx pronounced 'Hai-puh CRX' 一个浏览器插件开源项目，让开发者与社区管理员通过数据可视化的方式实时洞察社区的持续发展。

<https://github.com/hypertrons/hypertrons-crx>





哈爾濱工業大學
HARBIN INSTITUTE OF TECHNOLOGY

結束！

September 29, 2024