

详解 OpenDigger 开源项目

(🐰年版)



2023 年

世界趋势

世界如何变化?

- 全球化 2.0 + 数字化 2.0
- 软件吞噬世界 + 全新的软件供应链
- 数字经济 → 数字商品 → 数字体验

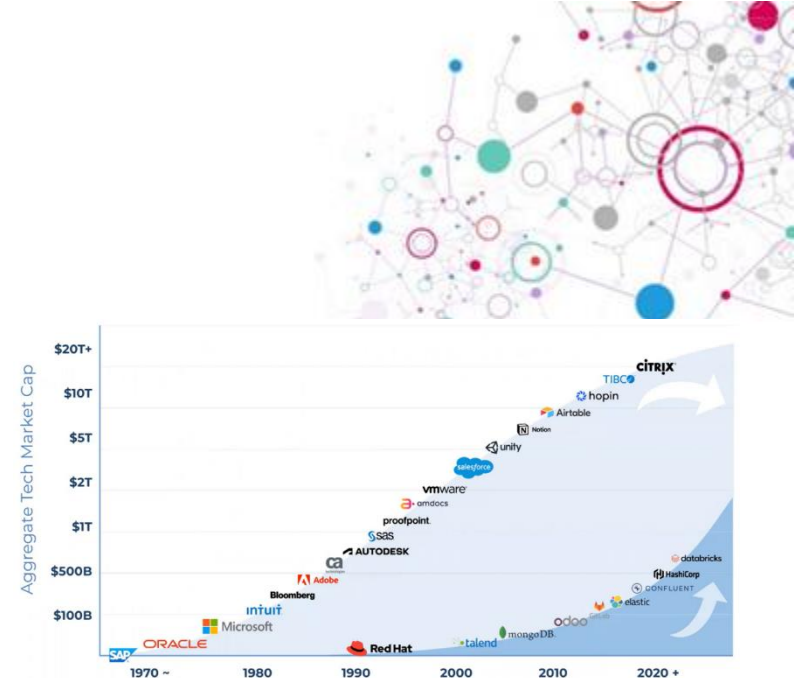


组织如何变化?

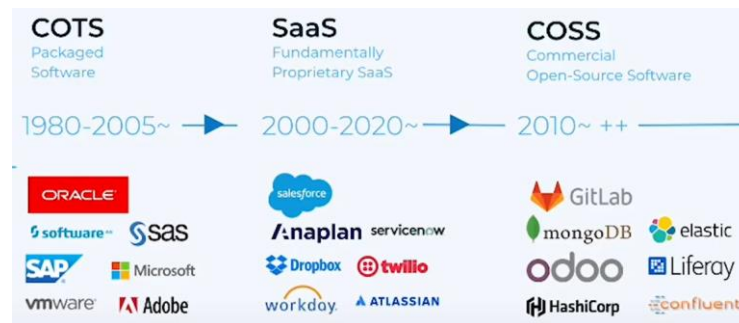
- 所有的公司/组织 → 软件公司/组织 → 研发部门构建软件
- 软件从成本中心变为利润中心 → 软件成为公司/组织的战略核心
- 软件业务 → 数字原生型公司/组织赢得市场份额

个体如何变化?

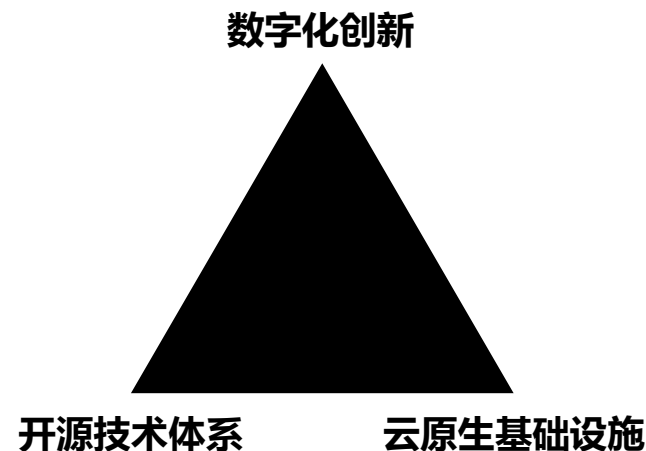
- 开发者 → 成为企业/组织的核心
- 数字化与软件思维方式
- 全新的软件开发与构建过程



COSS 发展趋势



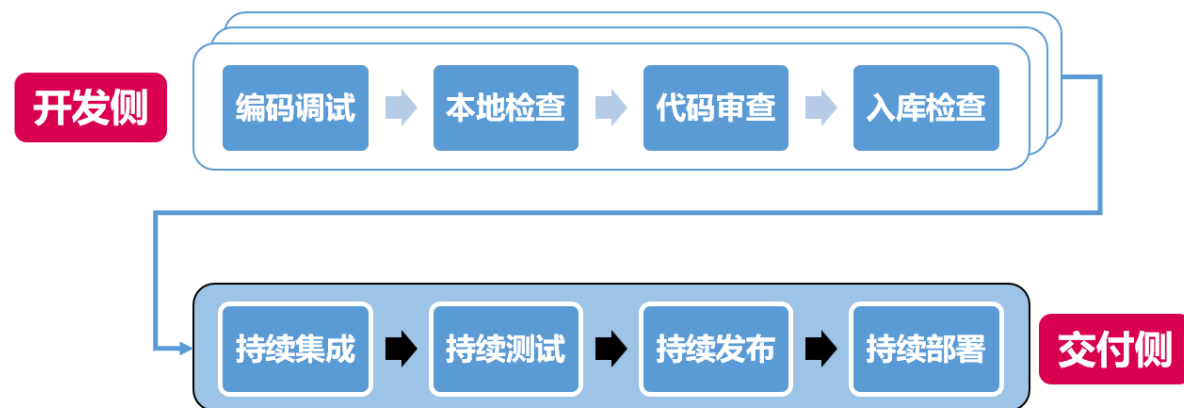
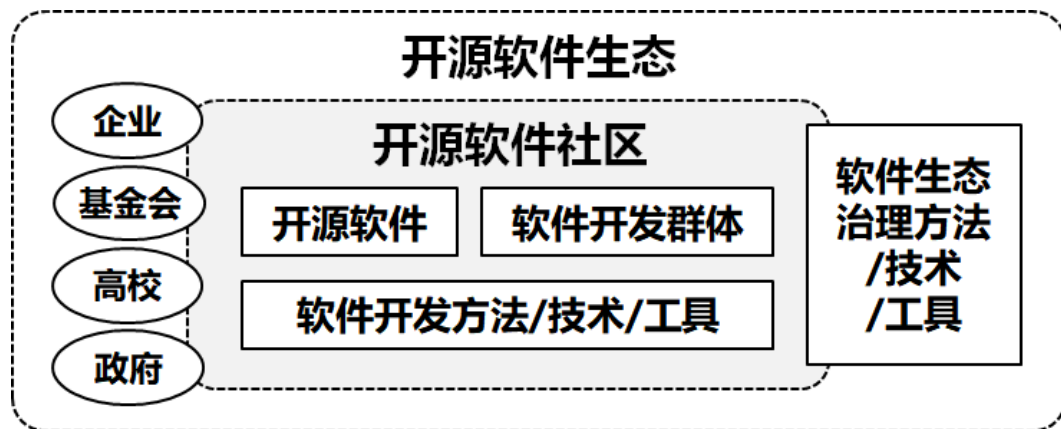
软件发展简史



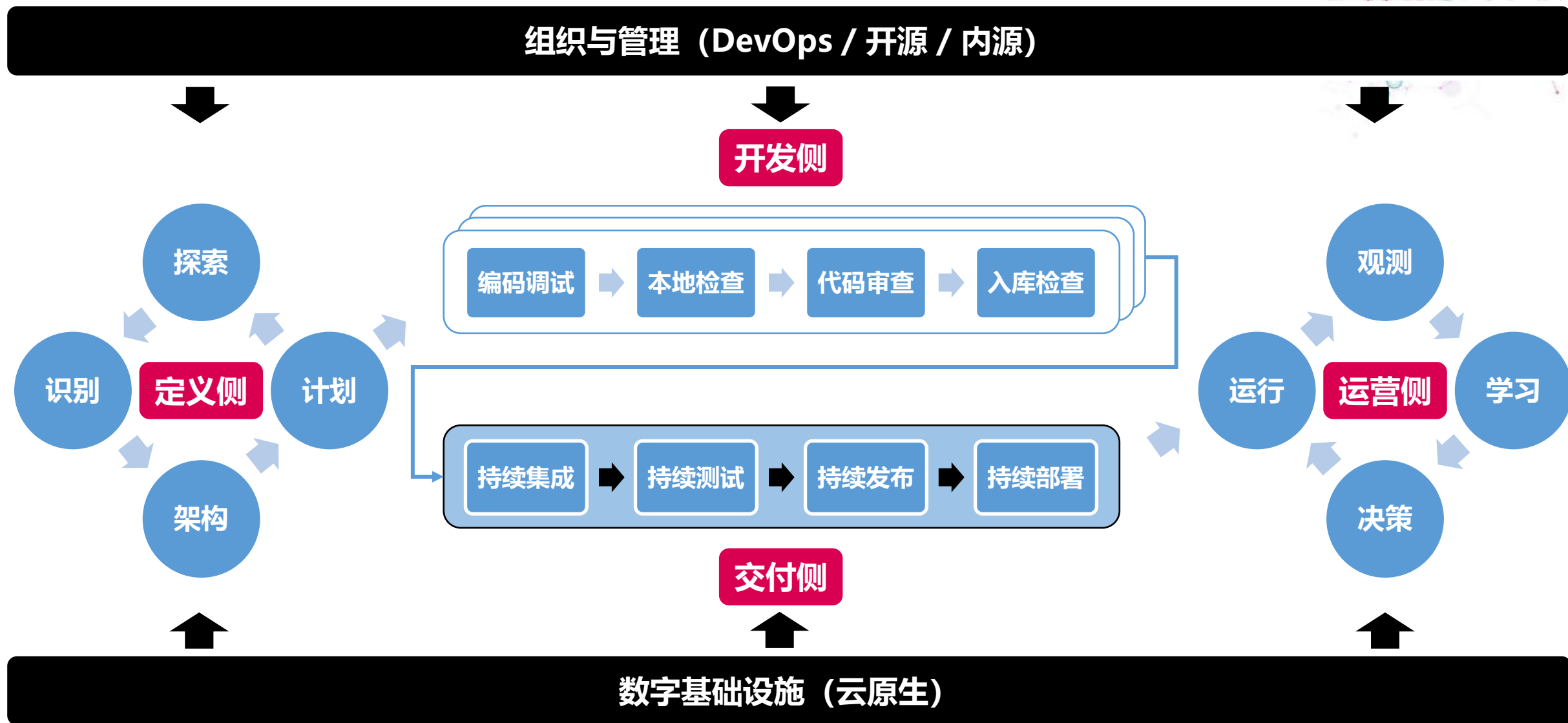
OpenDigger 的分析对象

开源软件开发与生态演化中的活动数据是一个非常宽广的概念，只要是在软件开发、维护、运行、以及生态治理、演化等过程中产生的数据，都可以称作开源软件活动数据，包括但不限于：

- Git/GitHub 日志事件 (Git/GitHub Log Event)
- 源代码 (Source Code)
- 文档 (Documentation)
- 配置文件 (Configuration Files)
- 变更 (Changes)
- 开发流程 (Development Process)
- 开发者 (Developers)
- 包管理平台 (Package Hosting Platforms)
- 社交数据 (Social Data)
- 软件生态网络 (Software Ecosystem Network)



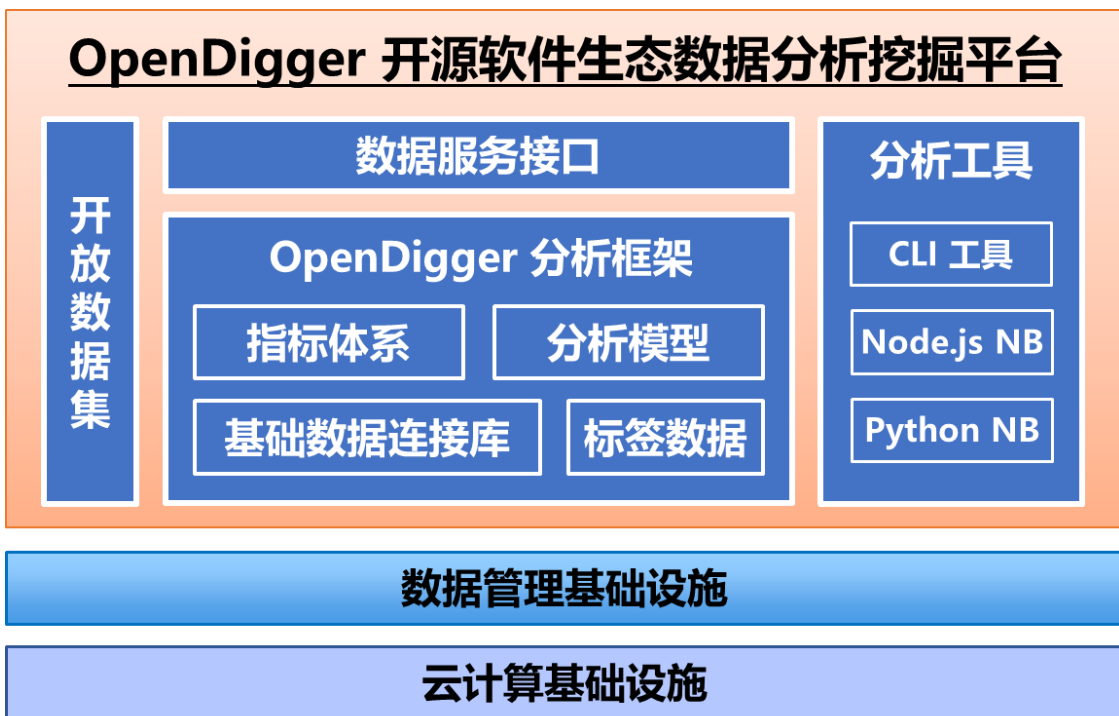
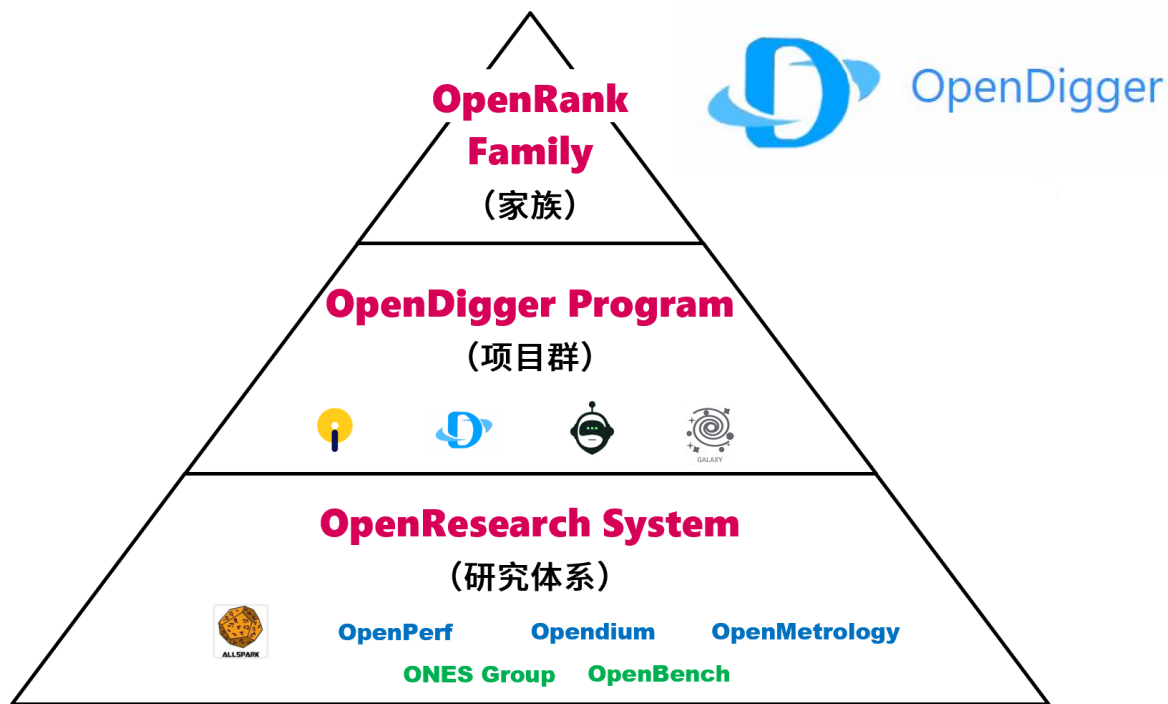
当代软件开发与价值交付流水线



OpenDigger 整体技术架构

OpenDigger 是一个面向开源软件生态数据的一站式分析挖掘平台，目标是构建开源领域的数据生态，成为促进开源生态持续发展的数据基础设施开源项目。

- 开源项目群： OpenDigger + OpenLeaderboard + Hypercrx + OpenDashboard +



OpenDigger 中有些啥?

指标

For repos

- [标准院/X-lab] **activity**
- [标准院/X-lab] **openrank** 
- [标准院/X-lab] **attention**
- [标准院/X-lab] **stars**
- [标准院/X-lab] **issue_comments**
- [标准院/X-lab] **participants**
- [LF/CHAOSS] **technical_fork**
- [LF/CHAOSS] **issues_new**
- [LF/CHAOSS] **issues_closed**
- [LF/CHAOSS] **code_change_lines_add**
- [LF/CHAOSS] **code_change_lines_rem**
- [LF/CHAOSS] **code_change_lines_sum**
- [LF/CHAOSS] **change_requests**
- [LF/CHAOSS] **change_requests_accept**
- [LF/CHAOSS] **change_requests_reviews**
- [LF/CHAOSS] **bus_factor**

For users

- [标准院/X-lab] **activity**
- [标准院/X-lab] **openrank**

数据

GitHub Event Log、libraries.io 全域数据

Sample datasets

- [2020 full](#), 855 million, 802 GB
- [2015 2021 top 50 year](#), 168 million, 117 GB
- [second sample](#), 62 million, 57 GB
- [label 2015](#), 3.5 million, 2.9 GB
- [paddle hackathon](#), 803 thousands, 736 MB

Label datasets

- **Communities**: Mulan, openmmlab, X-lab
- **Companies**: #117
- **Foundations**: Linux_foundation, Apache
- **Regions**: America, China
- **Technology**: bigdata, cloud, database, OS
- **COSS**: Global, China

工具

- **Command line interface (CLI) tool**
- **Node.js kernel notebook**
- **Python kernel notebook**

消费接口

- **OpenLeaderboard**
- **Hypercrx**
- **Hacking Force China**
- **China Open Source Report**
- **China Open Source Blue Paper**
- **Supply China Report**
- **COSS Report**
- **XSOSI Dashboard**
- **Mulan Dashboard**
- **Apache ALC Dashboard**

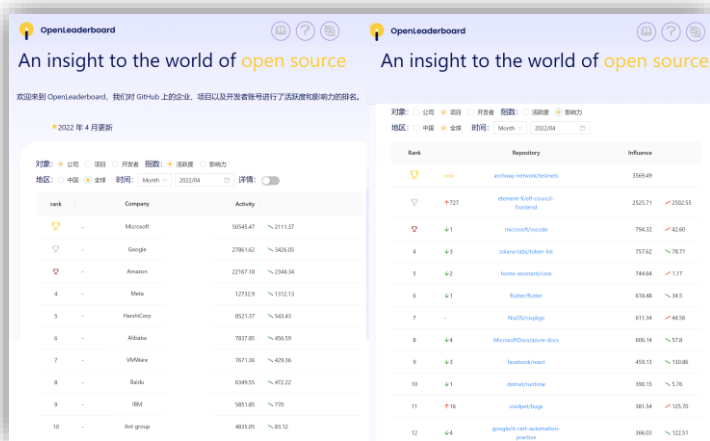
OpenDigger Timeline

- 报告与论文类
- 工具与应用类
- 大屏与看板类
- 模型与算法类



OpenLeaderboard

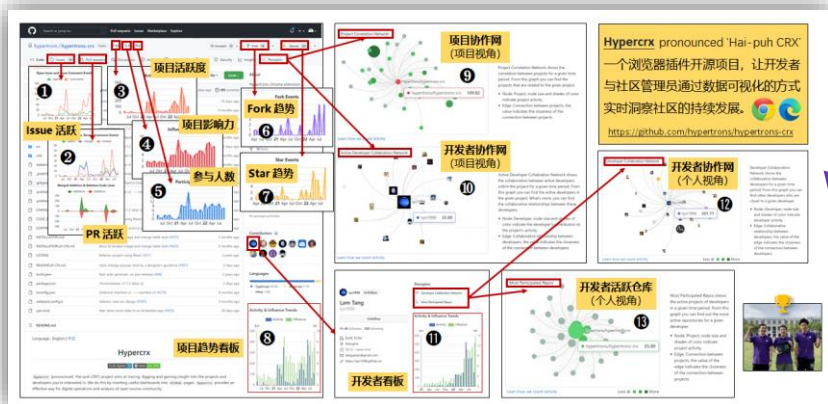
- 基于全域数据的指标排行榜



<https://open-leaderboard.x-lab.info>

Hypercrx 插件项目

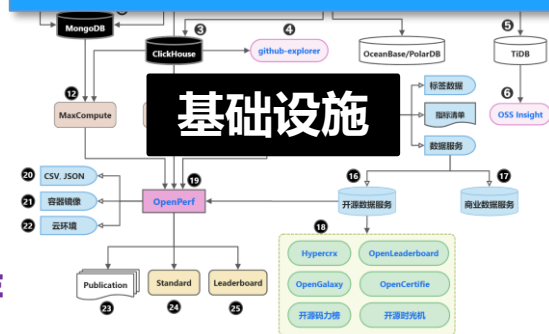
- 基于浏览器插件的展示增强工具



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

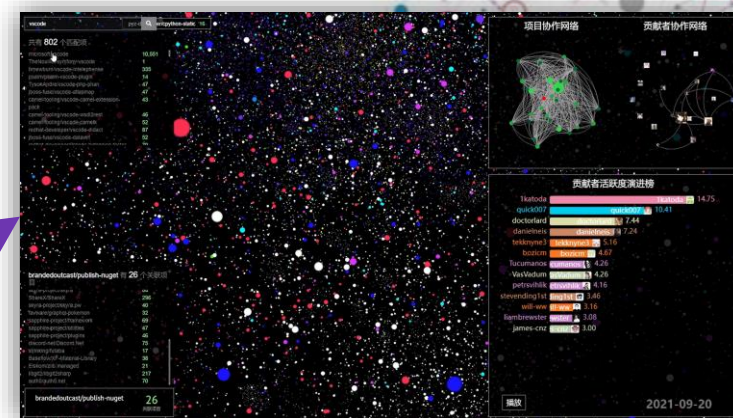
OpenDigger 生态项目群

OpenDigger OpenRank 系列算法



OpenGalaxy

开源星系遨游 3D 项目



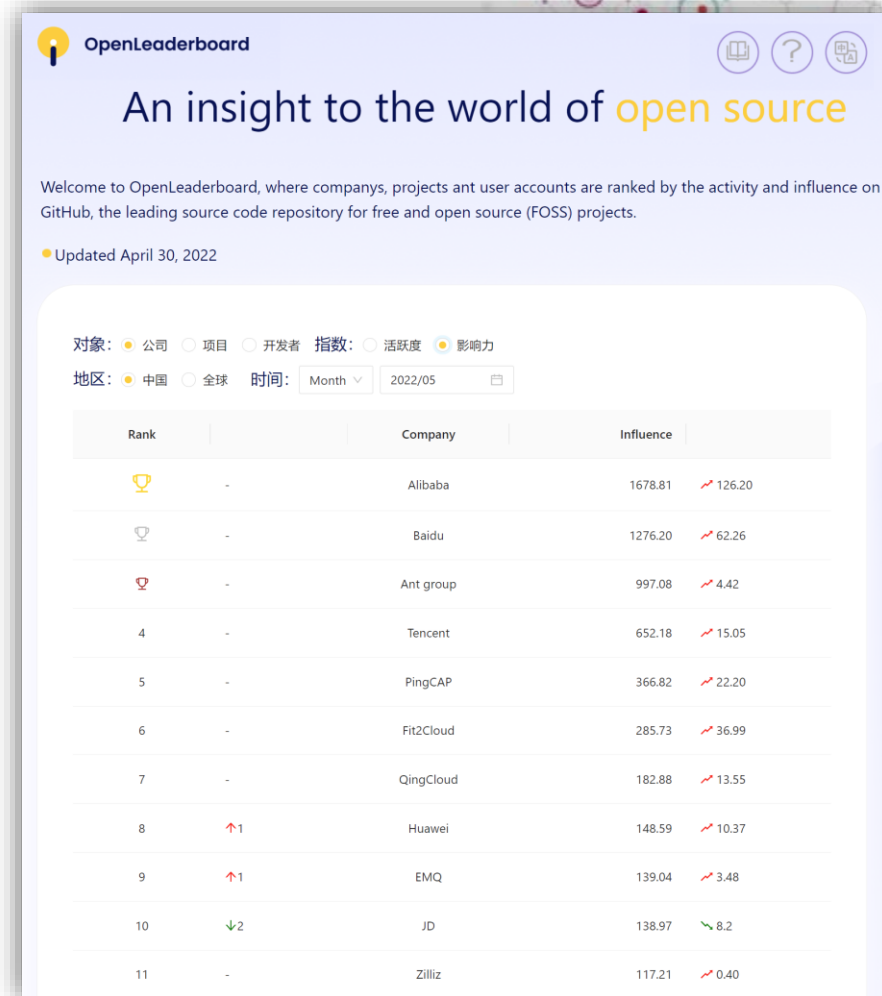
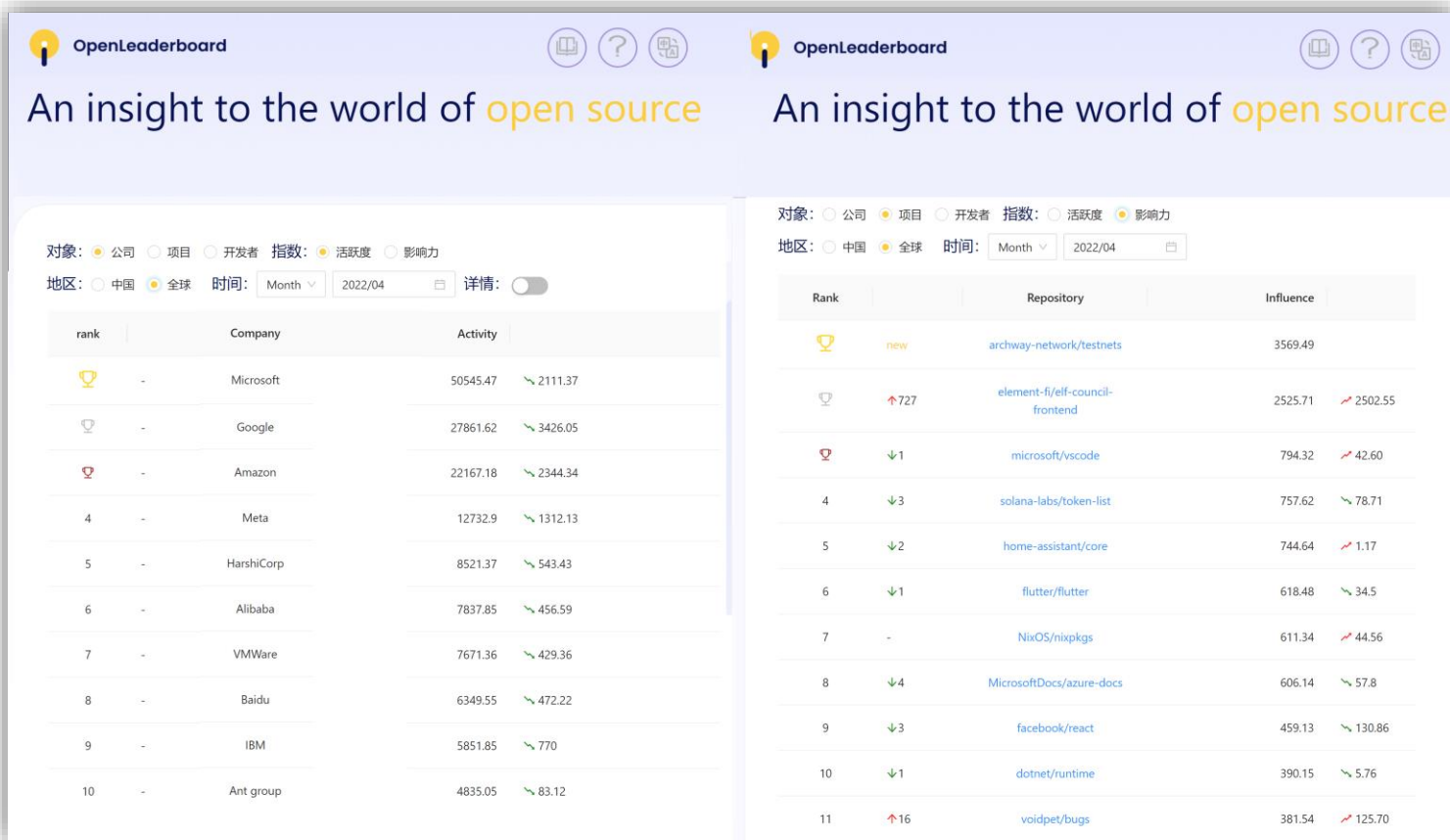
<https://github.com/X-lab2017/open-galaxy>

OpenDashboard

- 基于大屏的可交互组织数据洞察工具



① 工具集之：OpenLeaderboard



Powered by



OpenDigger



Mulan Community

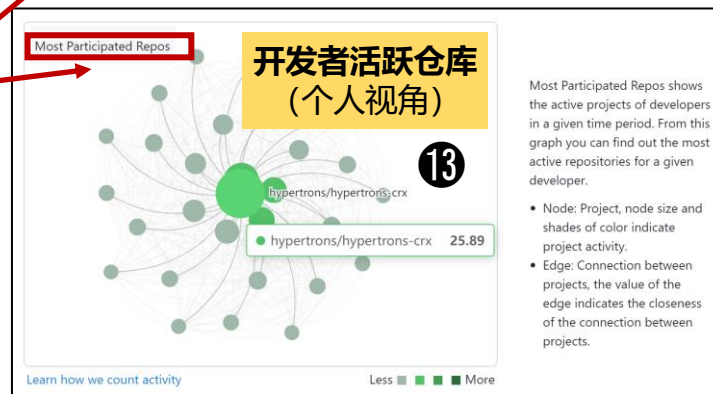
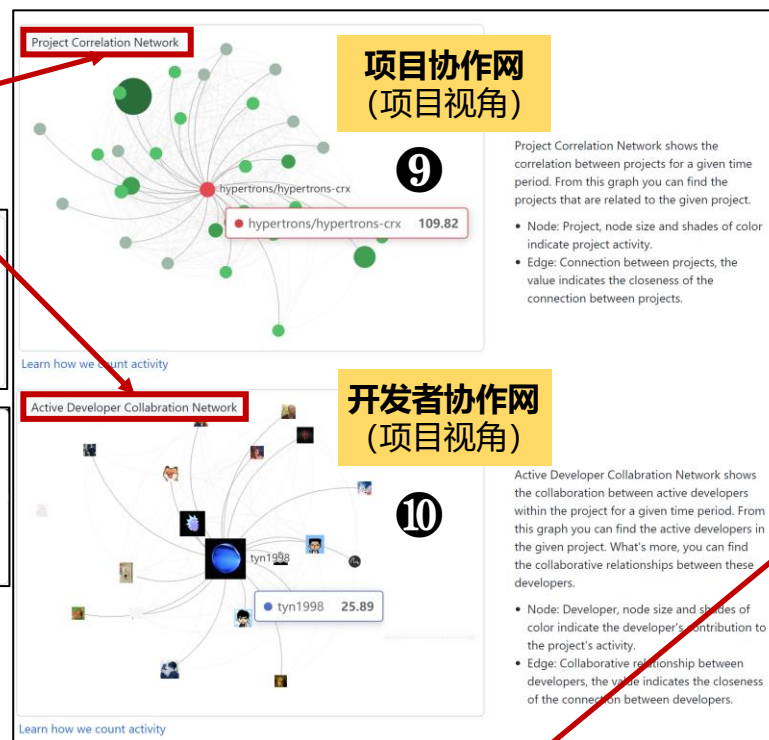
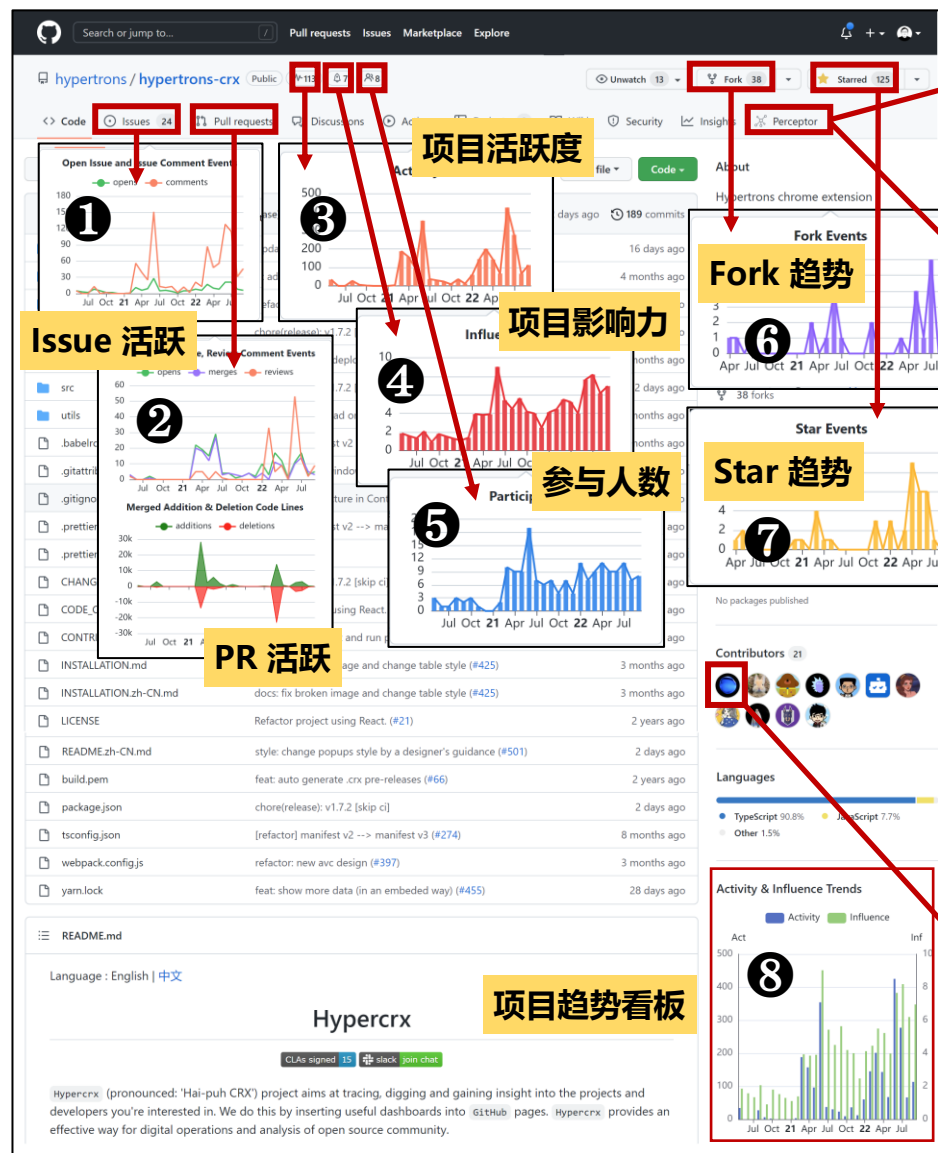


2022年 6 月北美 Open Source Summit OSPO 上正式发布~



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

② 工具集之：Hypercrx



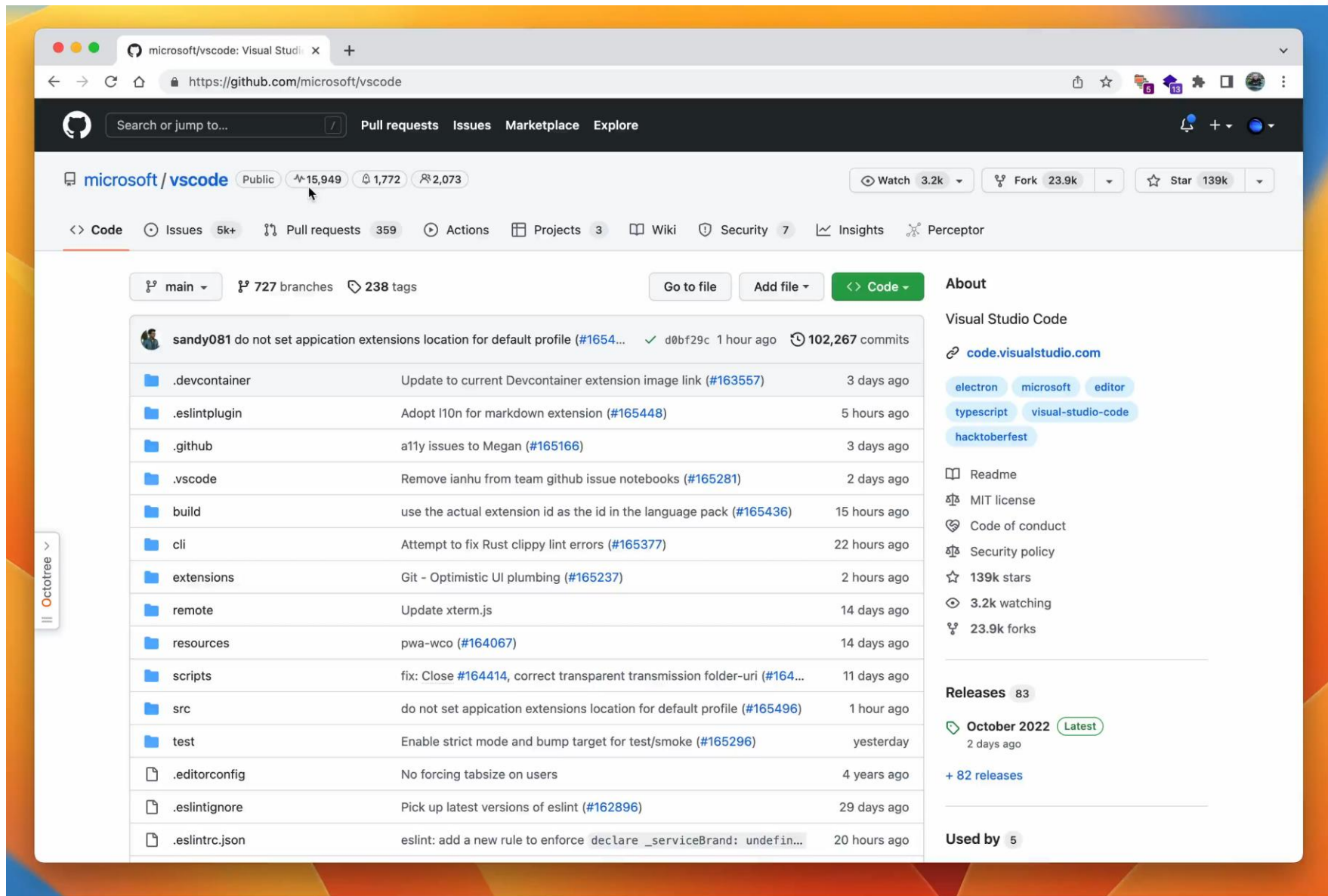
Hypercrx pronounced 'Hai-puh CRX'

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>



Hypercrx 插件项目: <https://github.com/hypertrons/hypertrons-crx>



The screenshot displays the GitHub repository for Microsoft VS Code. The repository is public, with 15,949 stars, 1,772 issues, and 2,073 pull requests. It has 3.2k watchers, 23.9k forks, and 139k stars. The repository is managed by Microsoft and is categorized as a code editor. The main branch is 'main', and there are 727 branches and 238 tags. The repository is licensed under MIT and is part of the Visual Studio Code project. The repository is managed by Microsoft and is categorized as a code editor. The repository is licensed under MIT and is part of the Visual Studio Code project.

Repository: **microsoft/vscode** (Public) 15,949 stars, 1,772 issues, 2,073 pull requests. 3.2k watchers, 23.9k forks, 139k stars.

Navigation: <> Code, Issues 5k+, Pull requests 359, Actions, Projects 3, Wiki, Security 7, Insights, Perceptor.

Repository details: main (727 branches, 238 tags). Go to file, Add file, <> Code.

Recent commits by sandy081:

Commit	Description	Time
do not set application extensions location for default profile (#165496)	do not set application extensions location for default profile (#165496)	1 hour ago
Enable strict mode and bump target for test/smoke (#165296)	Enable strict mode and bump target for test/smoke (#165296)	yesterday
eslint: add a new rule to enforce declare _serviceBrand: undefin...	eslint: add a new rule to enforce declare _serviceBrand: undefin...	20 hours ago
Pick up latest versions of eslint (#162896)	Pick up latest versions of eslint (#162896)	29 days ago
No forcing tabsize on users	No forcing tabsize on users	4 years ago
fix: Close #164414, correct transparent transmission folder-uri (#164...	fix: Close #164414, correct transparent transmission folder-uri (#164...	11 days ago
pwa-wco (#164067)	pwa-wco (#164067)	14 days ago
Update xterm.js	Update xterm.js	14 days ago
Git - Optimistic UI plumbing (#165237)	Git - Optimistic UI plumbing (#165237)	2 hours ago
Attempt to fix Rust clippy lint errors (#165377)	Attempt to fix Rust clippy lint errors (#165377)	22 hours ago
use the actual extension id as the id in the language pack (#165436)	use the actual extension id as the id in the language pack (#165436)	15 hours ago
Remove ianhu from team github issue notebooks (#165281)	Remove ianhu from team github issue notebooks (#165281)	2 days ago
a11y issues to Megan (#165166)	a11y issues to Megan (#165166)	3 days ago
Adopt l10n for markdown extension (#165448)	Adopt l10n for markdown extension (#165448)	5 hours ago
Update to current Devcontainer extension image link (#163557)	Update to current Devcontainer extension image link (#163557)	3 days ago

Repository structure:

- .devcontainer
- .eslintplugin
- .github
- .vscode
- build
- cli
- extensions
- remote
- resources
- scripts
- src
- test
- .editorconfig
- .eslintignore
- .eslintrc.json

About Visual Studio Code

code.visualstudio.com

electron, microsoft, editor, typescript, visual-studio-code, hacktoberfest

Readme, MIT license, Code of conduct, Security policy, 139k stars, 3.2k watching, 23.9k forks

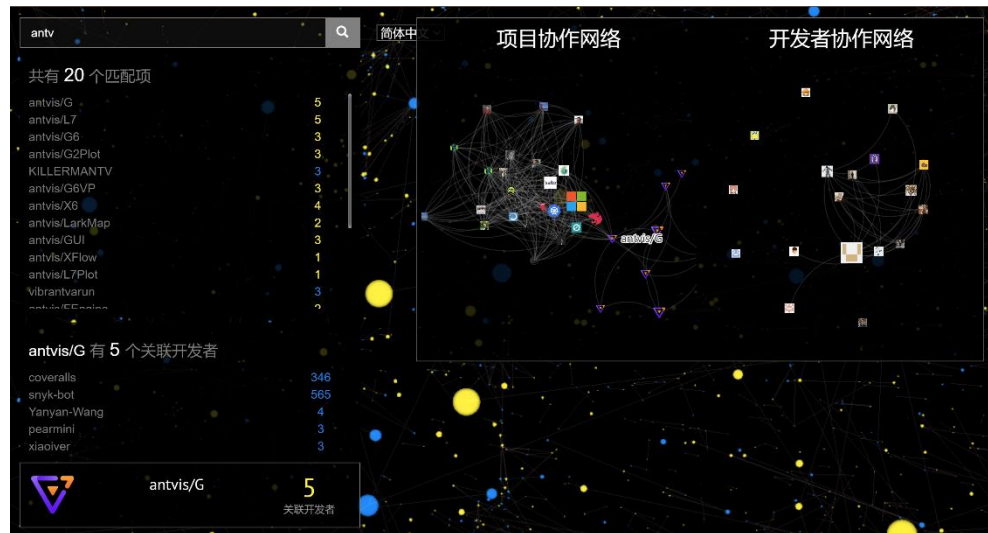
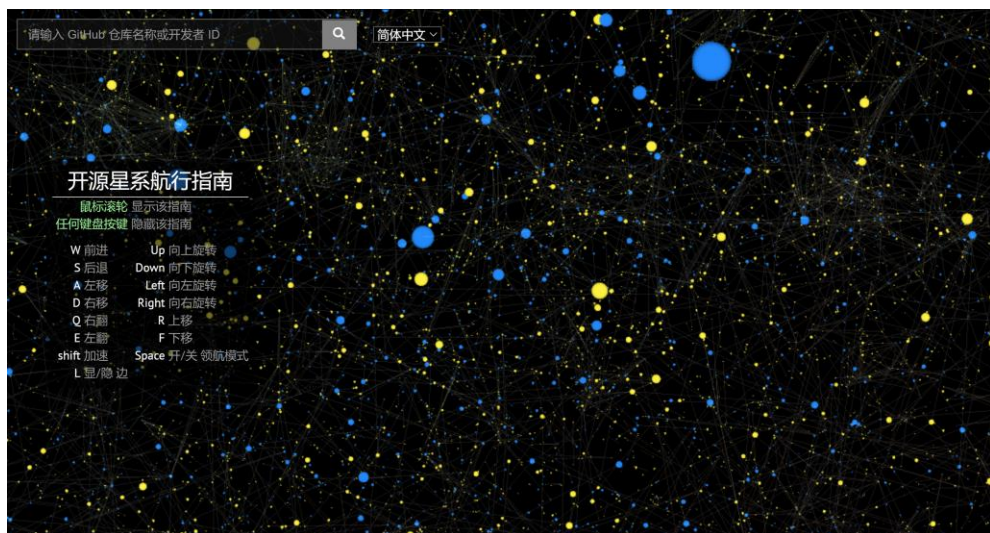
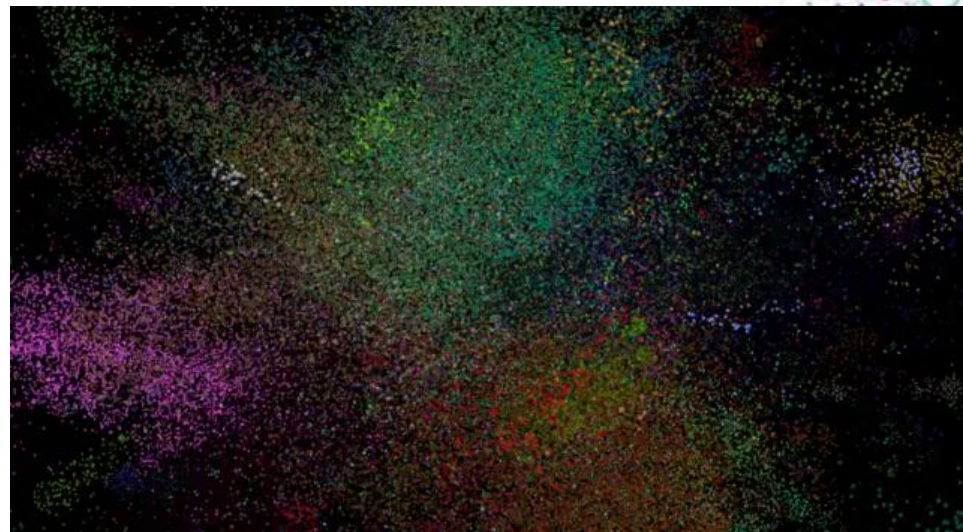
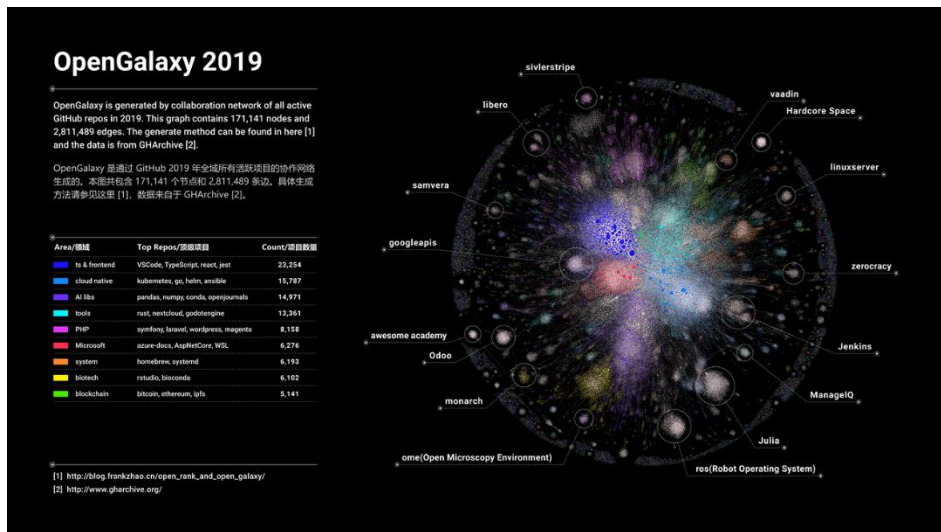
Releases 83

October 2022 (Latest) 2 days ago

+ 82 releases

Used by 5

③ 工具集之：OpenGalaxy



Activity Ranking

#	org	repo	activity
1	ant-design	ant-design	5505.65
2	kata-containers	kata-containers	2386.09
3	ant-design	pro-components	2269.66
4	umijs	umi	1992.36
5	ant-design	ant-design-pro	1474.57
6	ant-design	ant-design-mobile	1316.59
7	seata	seata	1198.16
8	antvis	x6	964.04
9	antvis	52	957.45
10	umijs	qiankun	870.38
11	kata-containers		831.98
12	umijs	umi-next	787.2
13	antvis	56	692.13
14	ant-design	ant-design-charts	643.72
15	moon	layotto	584.18

Ant Group Open Source Dashboard

2022-09-12 17:37:12

企业开源治理大屏

Attention Ranking

#	org	repo	attention
1	ant-design	ant-design	13954
2	seata	seata	3157
3	ant-design	ant-design-pro	2861
4	umijs	qiankun	2117
5	umijs	umi	1932
6	antvis	x6	1923
7	ant-design	ant-design-mobile	1678
8	Conex36	conex36	1594
9	ant-design	pro-components	1445
10	secretflow	secretflow	927
11	umijs	dumi	891
12	oceanbase	oceanbase	886
13	kata-containers	containers	867
14	nails-engine	engine	785
15	eggjs	egg	733

Orgs

27

Repos

1184

Star

386877

Contributor

5264

Repo MoM

18

Repo YoY

134

Repos without commit

org	repo	last commit at
eggjs	egg-dependencies	2016-06-18
ant-design	OTA.mobile.ant.design	2016-06-26
eggjs	egg-dns-client	2016-06-26
ant-design	OTA.mobile.ant.design	2016-10-01
duysj	duysj.com	2016-11-01
macaca-samp	macaca-multi-task-samp	2016-12-08
macaca-samp	macaca-multi-driver-samp	2016-12-08
macaca-samp	macaca-nodejs-boltsjs	2016-12-08
macaca-samp	macaca-test-pyma-samp	2016-12-08
macaca-samp	macaca-netty-sample	2016-12-08
macaca-samp	macaca-uniformer-samp	2016-12-08
macaca-samp	macacaboard	2016-12-08
macaca-samp	macaca-sample-test-chc	2016-12-09
macaca-samp	power-querer-test-cil	2017-01-07
duysj	duy-gui	2017-02-09

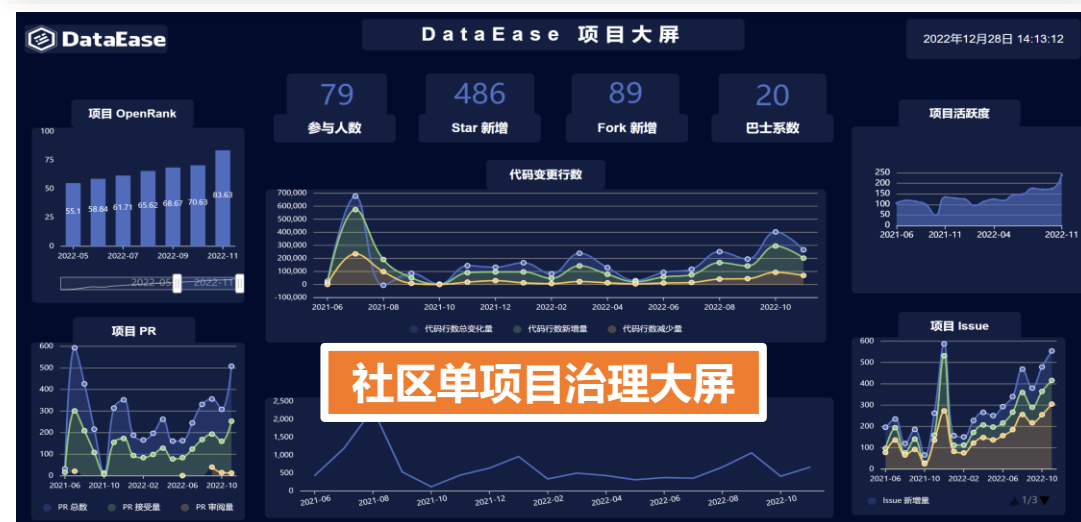
Org distribution

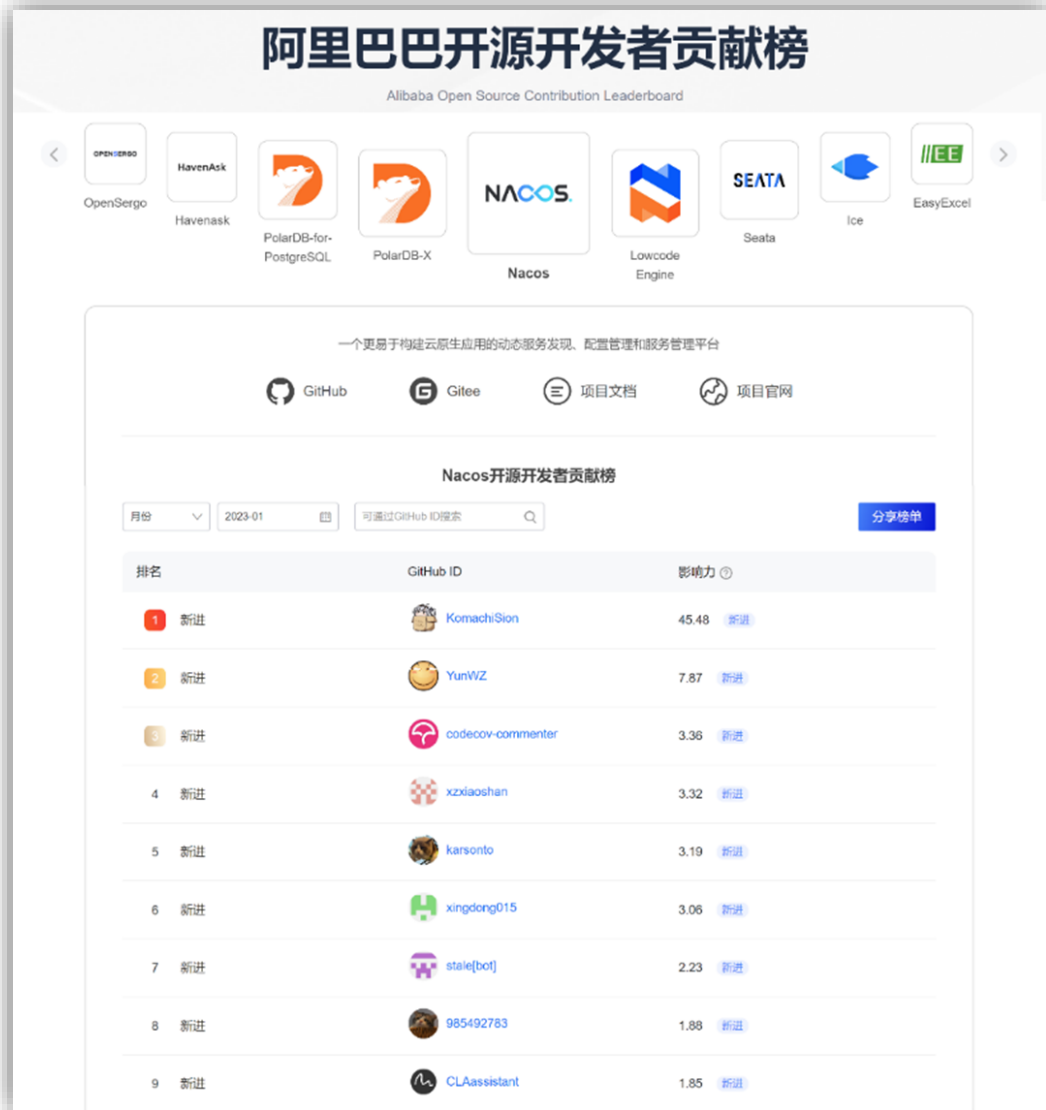
100% ant-design

License distribution

100% MIT License

// Powered by Xlab & DataV //





高校师生参与全球开源社区贡献是当下科技强国的重要举措，如由中央网信办、教育部正在实施的一流“网络安全学院”、“特色软件工程学院”等建设示范项目，拟在试点学校中，开展高校参与开源贡献的评价与激励工作。科学、客观、公正地对开源贡献进行度量评估，是有效引导广大师生、激励开源贡献者的重要抓手，是教育评价改革的“牛鼻子”工程。



OpenDigger 开源软件生态数据分析挖掘平台挑战赛

Open Source Software Ecosystem Data Analysis and Mining Platform Competition, **OpenSODA**



主办方：中华人民共和国工业和信息化部
Ministry of Industry and Information Technology of the People's Republic of China

承办方：



时间	大赛培训课程	挑战赛端
3.10 ~ 4.3 日	第一阶段：拥抱开源	<u>初赛（知识问答）</u>
4.4 ~ 5.8 日	第二阶段：贡献开源	<u>复赛（方案提交）</u>
5.9 ~ 6.26 日	第三阶段：发展开源	<u>决赛（作品提交）</u>
7月初	-	<u>宣布获奖名单</u>

奖项	数量	总奖金额
一等奖	1	10 万
二等奖	5	15 万
三等奖	10	10 万
创新贡献奖	50	15 万

(1) 任务类 (T)

- T1: **OpenDigger 指标的实现与优化**（开发类任务，预估难度：★）
- T2: **命令行交互的指标结果查询子模块**（开发类任务，预估难度：★★）
- T3: **基于机器学习的指标拟合与优化**（数据科学类，预估难度：★★★）

(2) 作品类 (W)

- W1: **可视化艺术作品或洞察报告**（数据创意类作品，预估难度：★）
- W2: **开源领域的可视化大屏**（Dashboard类作品，预估难度：★★）
- W3: **开源协作网络可视化**（网络可视化类作品，预估难度：★★★）

目前，报名系统已经上线，欢迎大家积极报名，也预祝大家取得好的成绩！👉

本次大赛报名采用二步注册机制，全部由队长代表队伍完成：

1. 首先，队长在开放原子大赛官网报名（填写基本信息）
2. 然后，队长在金数据问卷中报名（填写详细信息）

每支队伍由**1~4名成员组成**，详见金数据问卷。如遇到问题，请邮件联系大赛运营组：
opensoda@126.com



大赛详情



报名详情