



高校开源创新的故事（共创版）

—— 以X-lab开放实验室为例

主讲人：王伟

(内容贡献：王伟，赵生宇、周添一、范家宽、吴绍岭、林海铭、王皓月)

Biography (个人经历)

- 华东师范大学数据学院研究员、博士生导师， ACM/IEEE会员、CCF高级会员
- Wisconsin大学Madison分校、Florida大学访问学者
- X-lab开放实验室创始人、开源社理事会执行长、Wuhan2020开源社区理事长
- 研究兴趣： (教育 + 开源) × 数据
 - 计算教育学
 - 开源系统论
 - 复杂信息网络
- E-mail: wwang@dase.ecnu.edu.cn



Microsoft Build

May 19–20, 2020 | Digital event

科技向善，数字普惠

Queue

Na Yang's desktop

Live event

Na Yang (Guest)'s desktop is live

潘正磊
Julia Liuson

微软全球资深副总裁

5M

cross

Visual Studio Co

一个非常重要的

Microsoft

在线技术答疑
Ask-me-anything

请您通过直播窗口右上方的文字对话窗提问，
我们将为您文字解答

Send live

End

Presenters

Annie Yang (Wicres...)

Tianhao Cao (Guest)

ZhaoWencho (Gu...)

Emily Chen (Guest)

Julia Liuson (Guest)

Na Yang

Stop sharing

Mute all

You

Content

My desktop

Jun Han (Guest)

Alex Zhao (Guest)

Zhenwei Su (Guest)

Zihao Zhang (Guest)

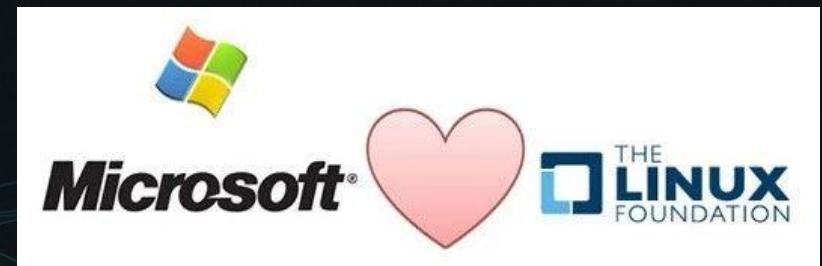
Wei W...

微软的故事

- 2014年，微软宣布了.NET开源，开始在GitHub上建立账户
- 2015年，微软宣布开源Visual Studio Code
- 2016年，微软开源深度学习工具包CNTK
- 2016年，微软发布开源跨平台工具PowerShell
- 2017年，微软开源分布式高性能ML工具LightGBM
- 2018年6月，微软正式收购GitHub
- 2018年9月，微软开源已诞生36年的MS-DOS
- 2018年12月，微软采用Chromium内核开发桌面版Edge浏览器
- 2019年5月，微软开源命令行工具Windows Terminal
- 2019年9月，微软exFAT文件系统开源，无需通过授权就可集成到Linux内核中
- 2019年10月，微软宣布加入OpenJDK参与开发
- 2020年3月，微软首届Linux会议WSLConf线上举行

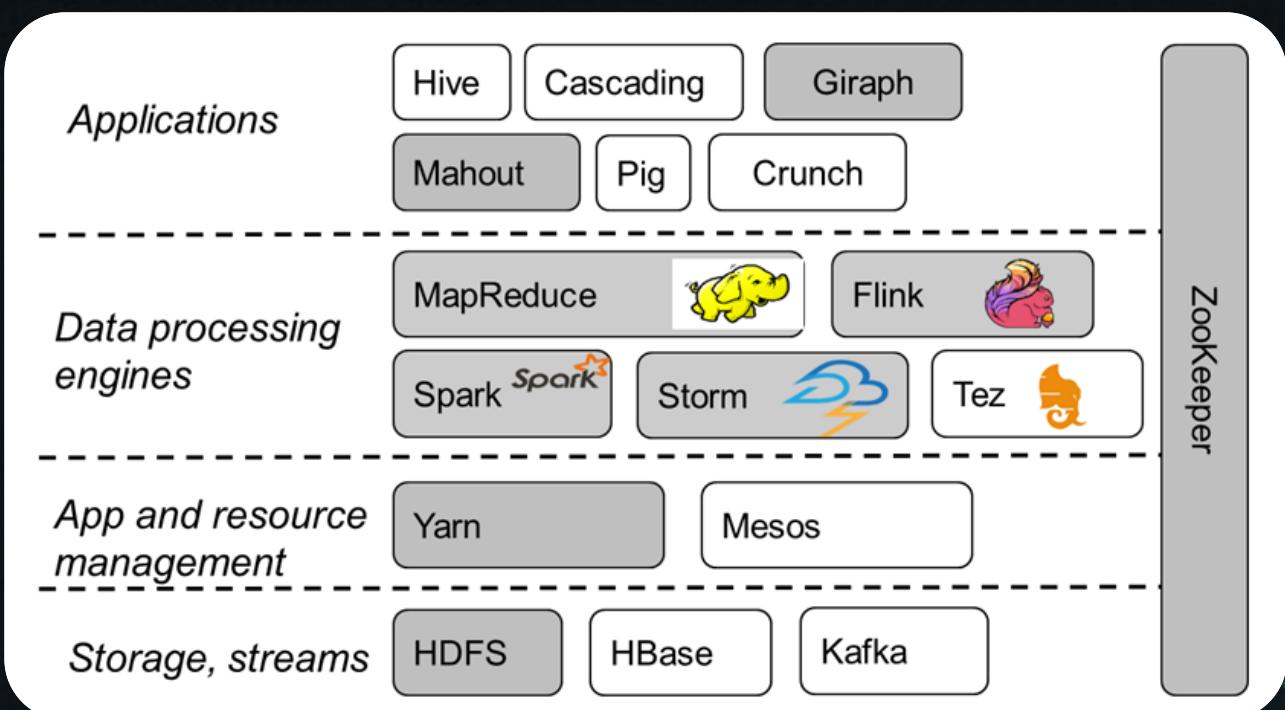


Satya Nadella, 2014



微软已经成为一家开源公司！

高校的故事：新工科下的计算机教育



计算机教育完全依赖于开源！

全球IT产业的故事

- 突破：国外厂商垄断的“WinTel联盟”
- 华为：基于鲲鹏生态的“第二计算平面”
- 支持：发起建设“第二计算教育平面”

	传统计算机教育平面	第二计算教育平面
硬件环境	基于X86的PC机、服务器	鲲鹏芯片PC、泰山服务器
操作系统	Windows	桌面Linux、欧拉OS
数据库	ACCESS、SQL SERVER	GaussDB数据库
编程语言	C/C++、Python、CUDA	Python、C/C++、昇腾编程模型
大数据	Hadoop、Spark	FusionInsight
云计算	Azure、Amazon、阿里	华为云
人工智能	GPU编程平台	Atlas平台、ModelArts平台



开源成为公司和国家战略！

创投的故事：科技推动世界进步，数字化驱动创新发展

经济规模



• 数字化核心要素

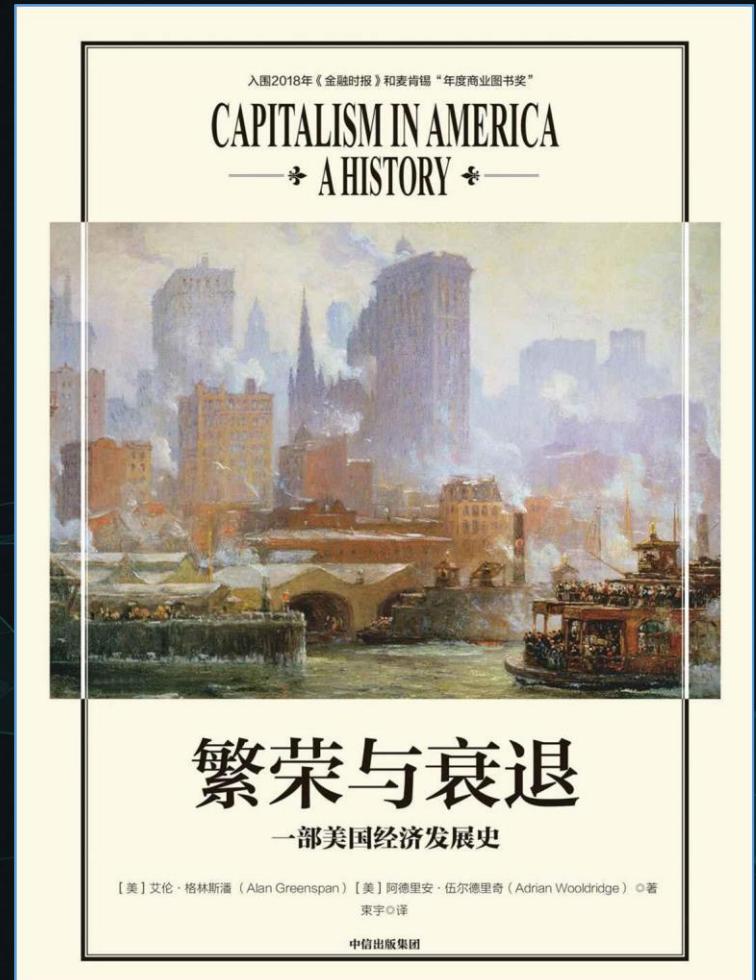
- 获取信息
- 表达信息
- 存储信息
- 传输信息
- 处理信息
- 交付信息



开源在数字化领域的最前沿！

发展的故事：Creative Destruction（创造性破坏）

- 影响社会生产力的主要因素：
 - 资本（包括场地和设备）
 - 劳动力（教育水平 + 熟练程度）
 - 生产率（使用资本和劳动力的方式）
- 提高社会生产力的方法：**创造性破坏**
 - 降低基本经济投入的成本
 - 大力发展大规模教育
 - 更高效的利用生产要素（人 + 物），从而提高生产率
 - 工具创新（技术）
 - 流程创新（管理）
 - 信息创新（认知）
 - 降低风险（政策）



Open source is creative destruction

- 从 软件产业 (**Software industry**) 到 开源产业 (**Open source industry**)
- 生产要素的变迁
 - 场地设备 → 数据 + 代码 + 算力 (生产资料所有权的革命)
 - 物理空间 → 数字空间 (时空扩展)
- 劳动力的变迁
- 使用劳动力方式的变迁
 - 工具: Git
 - 流程: DevOps
 - 信息: Transparent information flow
 - 自动化: RPA (Robotic Process Automation)

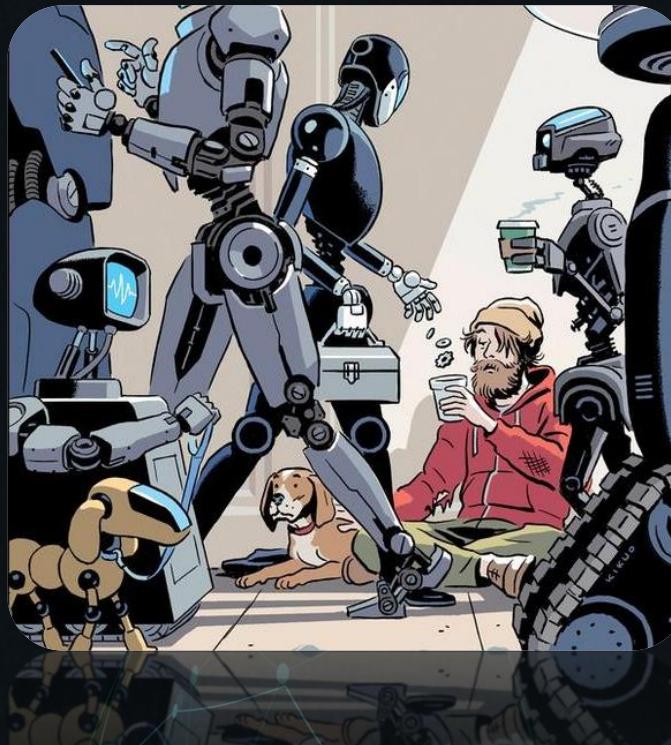


开源是创新与进步的源泉！

技术的步伐不可阻挡

每一个新技术的出现都是为了替代掉使用旧技术的人的劳动力，进而达到解放生产力的目的！

要懂机器
能操纵机器



要理解开源
能玩转开源

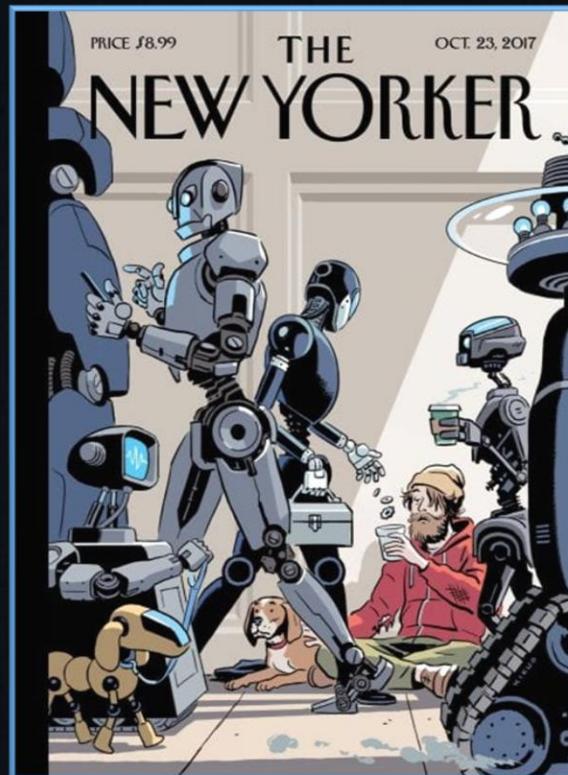
人工智能时代的挑战



ALL Systems Go

At last — a computer program that can beat a champion Go player

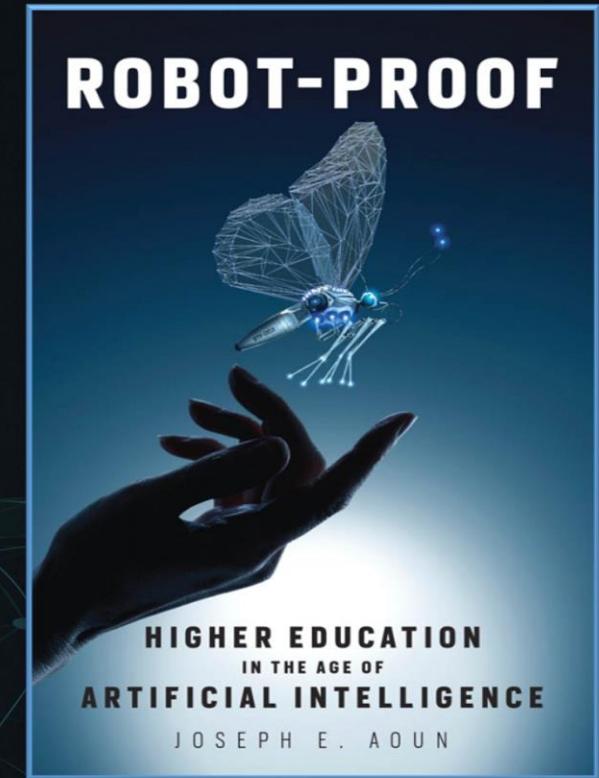
Nature 2016.01



DARK FACTORY

The robotics revolution is changing what machines can do. Where do humans fit in?

The New Yorker 2017.10



Robot-Proof

Higher Education in the Age of Artificial Intelligence

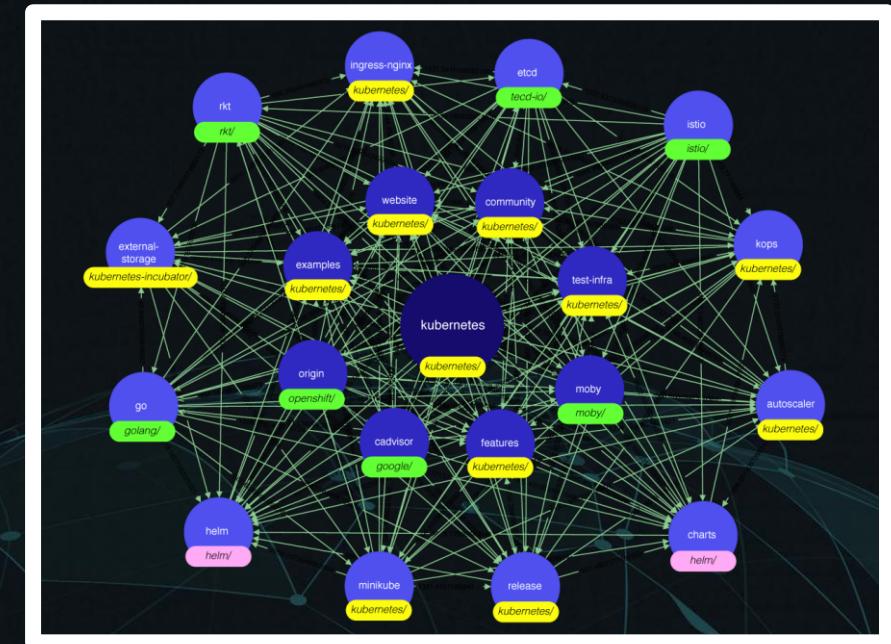
MIT Press 2017.08



智能时代的开源运动

开源时代的智能教育

X-lab 开放实验室方向

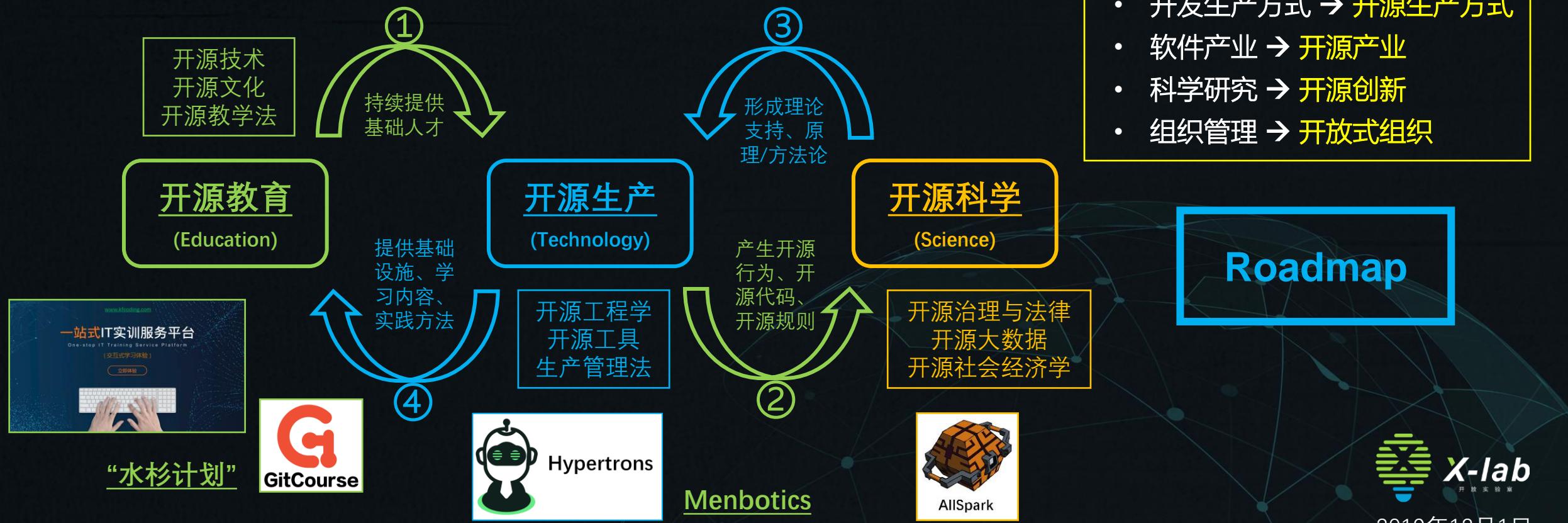


教育

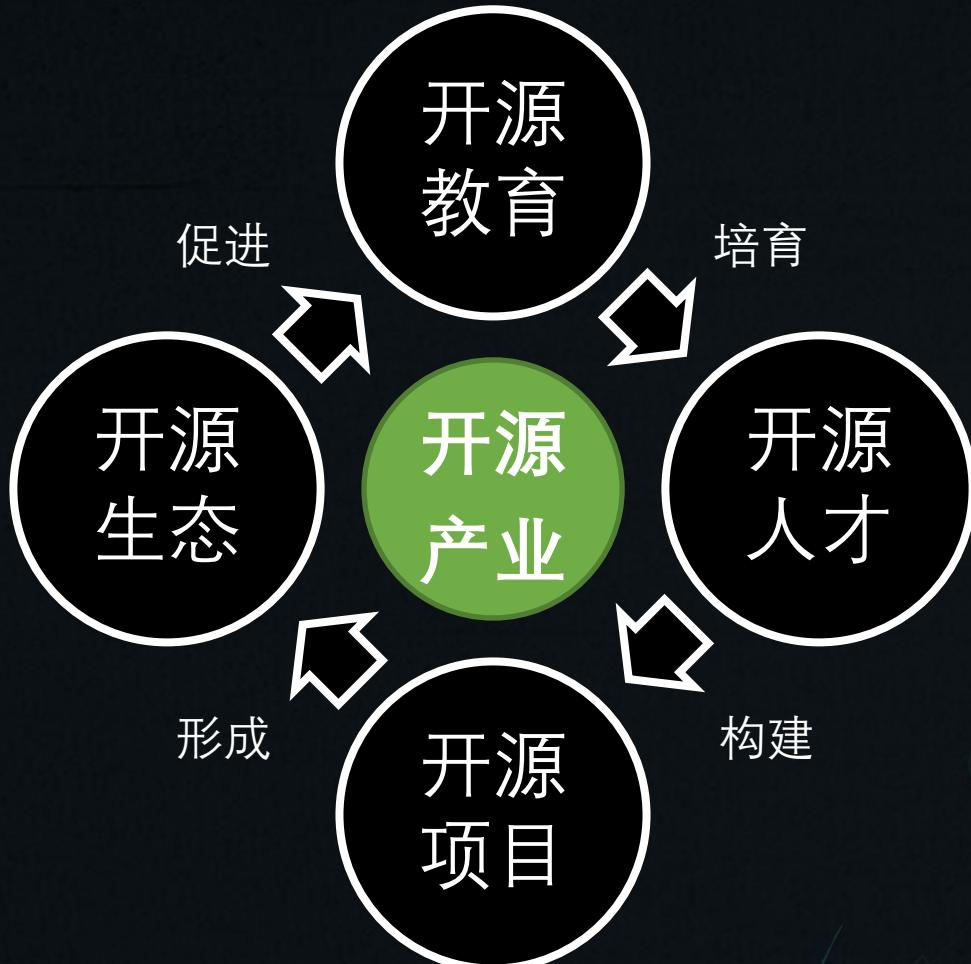
开源

数字社会 | 代码经济 | 开源产业

伟大的发明：语言、标准时间、集装箱、TCP/IP、Git、.....



从 软件产业 到 开源产业



- 用 **开源教育** 培育 **开源人才**；
- 用 **开源人才** 构建 **开源项目**；
- 用 **开源项目** 形成 **开源生态**；
- 用 **开源生态** 促进 **开源教育**。

X-lab开放实验室的开源创新故事

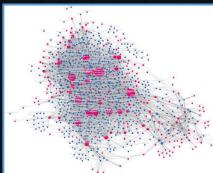
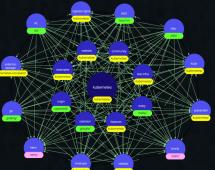
开源教育



开源生产



开源科学



开源公益



故事1：开源启蒙与实践课程



《开源软件设计与开发》课程出品人



华东师范大学 数据学院



开源社



X-lab开放实验室



合作企业

Syllabus

1. 开放源代码介绍
2. CVS、SVN、Git，版本控制系统的演化
3. 开源协作的核心：Diff/Patch/Merge/PR
4. 源代码阅读的艺术（一）：工具篇
5. 源代码阅读的艺术（二）：意义篇
6. 源代码阅读的艺术（三）：实战篇
7. 源代码阅读的艺术（四）：软技能篇
8. 开放源代码需要了解的法律常识
9. 融入和参与上游开源项目所需要的软技能
10. 社区的艺术
11. 开源与商业
12. 开放式开发（一）：以Linux Kernel为例阐明
13. 开放式开发（二）：以Apache HTTPD为例阐明
14. 开放式开发（三）：以FreeBSD为例阐明
15. 开放式开发（四）：以Kubernetes为例阐明
16. 开放式开发（五）：以TensorFlow为例阐明



教育平台：大规模开放在线实训平台（KFCoding）



一站式IT实训服务平台

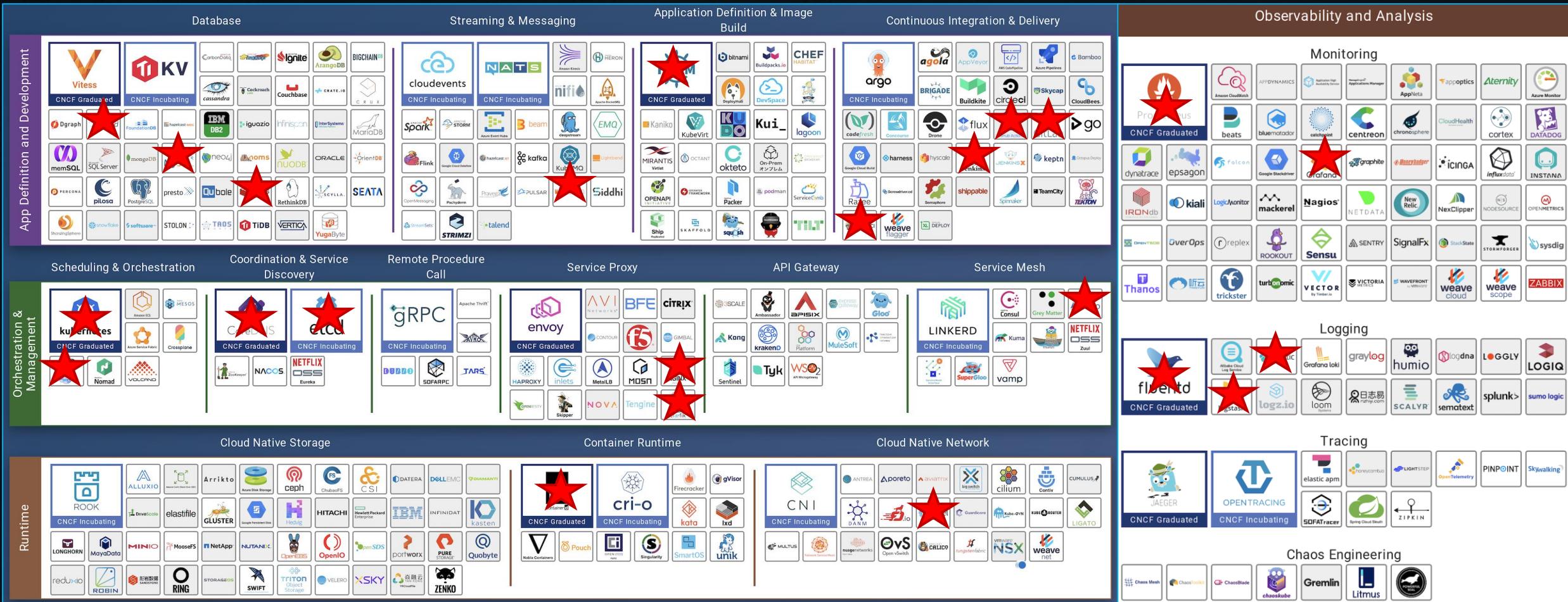
One-stop IT Training Service Platform

立即体验

The screenshot displays several windows from the KFCoding platform:

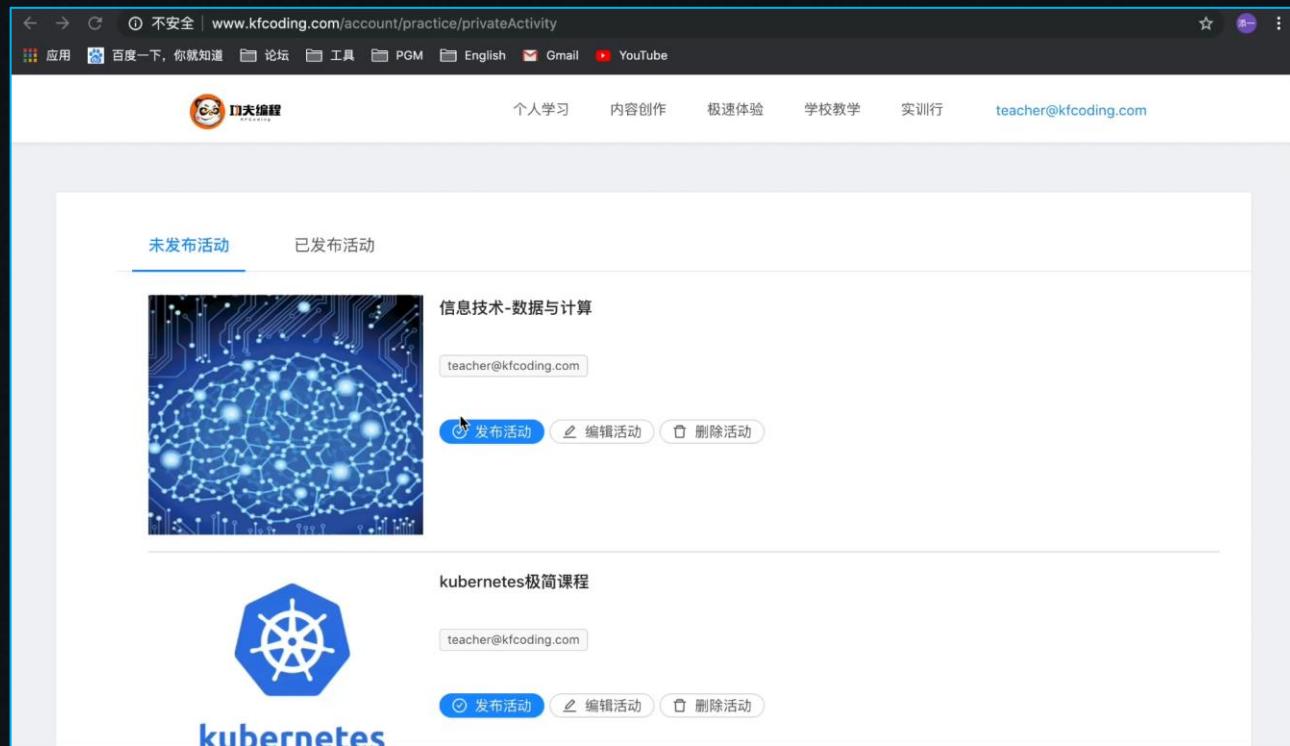
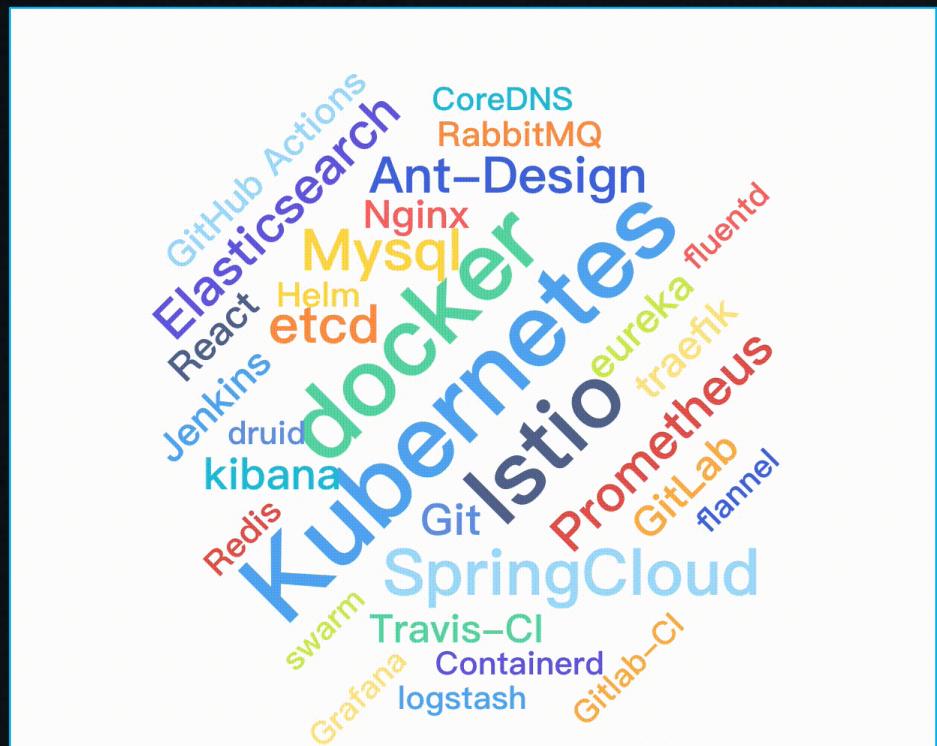
- A terminal window showing command-line code execution.
- A code editor window with C code for a "Hello, world!" program.
- A browser-based course interface titled "Hello World!" with a "Terminal" tab highlighted.
- A Scratch-style visual programming environment.
- A Python code editor window.
- A data visualization tool showing a scatter plot.
- A small video player window showing a street scene.

KFCoding：取之开源



从 CNCF Landscape 看 KFCoding 与开源的关系，标红色星的项目均为 KFCoding 使用到的开源项目。

KFCoding 主要用到的开源项目



故事2：取之开源，用之开源

- **开源大学**：用开源技术来推动开源教育，用开源教育来普及开源精神
 - 提供 100+ 的开源课程，涵盖云计算、大数据、人工智能等领域
 - 提供终端、桌面、VSCode、Notebook 等多种类型的实训环境

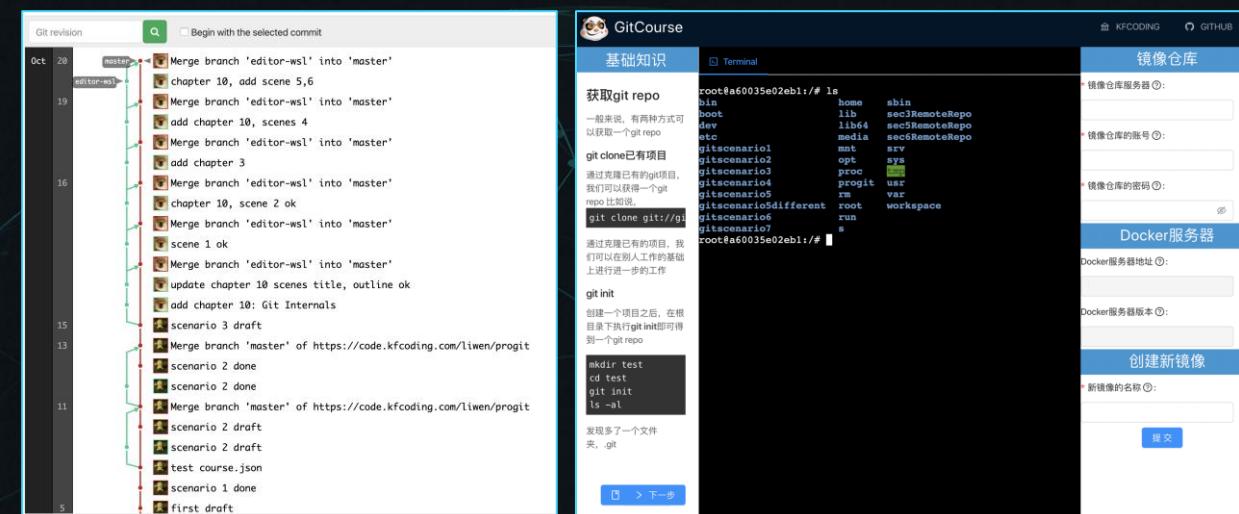


故事3：GitCourse开源交互式情景课程工具



- 基于Markdown的内容渲染
 - 语法简洁灵活，内容可移植
- 基于Git的内容创作体系
 - 实训课程能够像软件代码一样管理
- 基于Docker的实训环境体系
 - 秒级启动，支持用户自定义
- 开放配置
 - 赋予用户最大化自定义交互

GitCourse 设计哲学



<https://github.com/kfcoding/gitcourse>

X-lab实验室开源开发流程



- ① [Feature] We need to extract GitLab and Gitee raw client kind/feature
#281 opened on 27 Feb by frank-zsy  8
- ① [Feature] We need a component to generate contributing guide template kind/feature
#274 opened on 25 Feb by frank-zsy  6
- ① [Community] We need a logo and preview image for hypertrons help wanted kind/community
#259 opened on 20 Feb by frank-zsy  0 of 2
- ① [Doc] Issue and PR template needed area/document
#255 opened on 20 Feb by frank-zsy  1
- ① [feature] Use repold as the unique identifier of the repo kind/feature
#253 opened on 18 Feb by WuShaoling  0
- ① [Discussion] Tracking the code change package-lock.json kind/discussion
#245 opened on 12 Feb by WillemJiang  1
- ① [Refactor] we need onCommand function kind/enhancement
#229 opened on 13 Jan by WuShaoling  1
- ① [Bug] `approve` does not work on gitlab kind/bug
#226 opened on 13 Jan by heming6666  4
- ① [Feature] We need to generate document and update to repo automatically area/document
difficulty/3 kind/feature
#156 opened on 11 Dec 2019 by frank-zsy  v0.3.0  1

以 Issue 和 PR 为中心的项目管理与协作开发流程

Issue:

- **Milestones:** 里程碑
 - **Labels:** 问题分类 (缺陷、特性、文档等)
 - **Assignees:** 任务指派
 - **Discussion:** 讨论
 - **Search:** 沉淀，可追溯

Pull Request:

- **Branch:** 分布式协作开发
 - **Commit:** 代码规范
 - **Code Review:** 代码质量控制
 - **CI:** 单元测试与集成测试
 - **CD:** 持续部署 (staging & production)



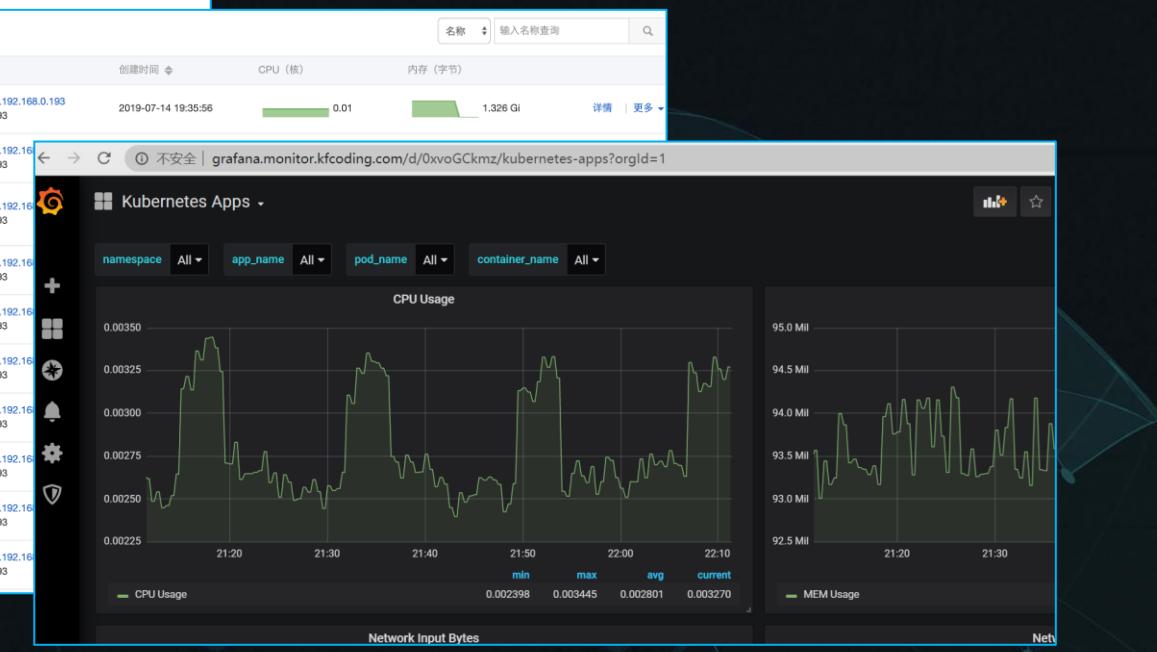
X-lab开源开发基础设施



The screenshot shows the GitHub organization page for X-lab. It features a sidebar with repositories such as X-lab2017.github, grimoirelab-sidecar, and grimoirelab-elk. The main area displays a project management interface with a list of channels and a detailed view of the kf-campus channel. This view includes a Slack-like message history and a table of Kubernetes pods.



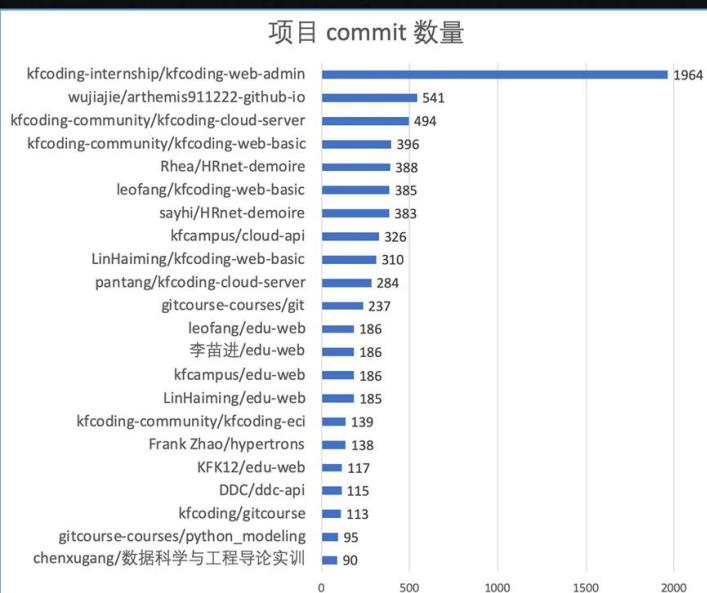
.....





故事4：X-lab开源协作的数字化

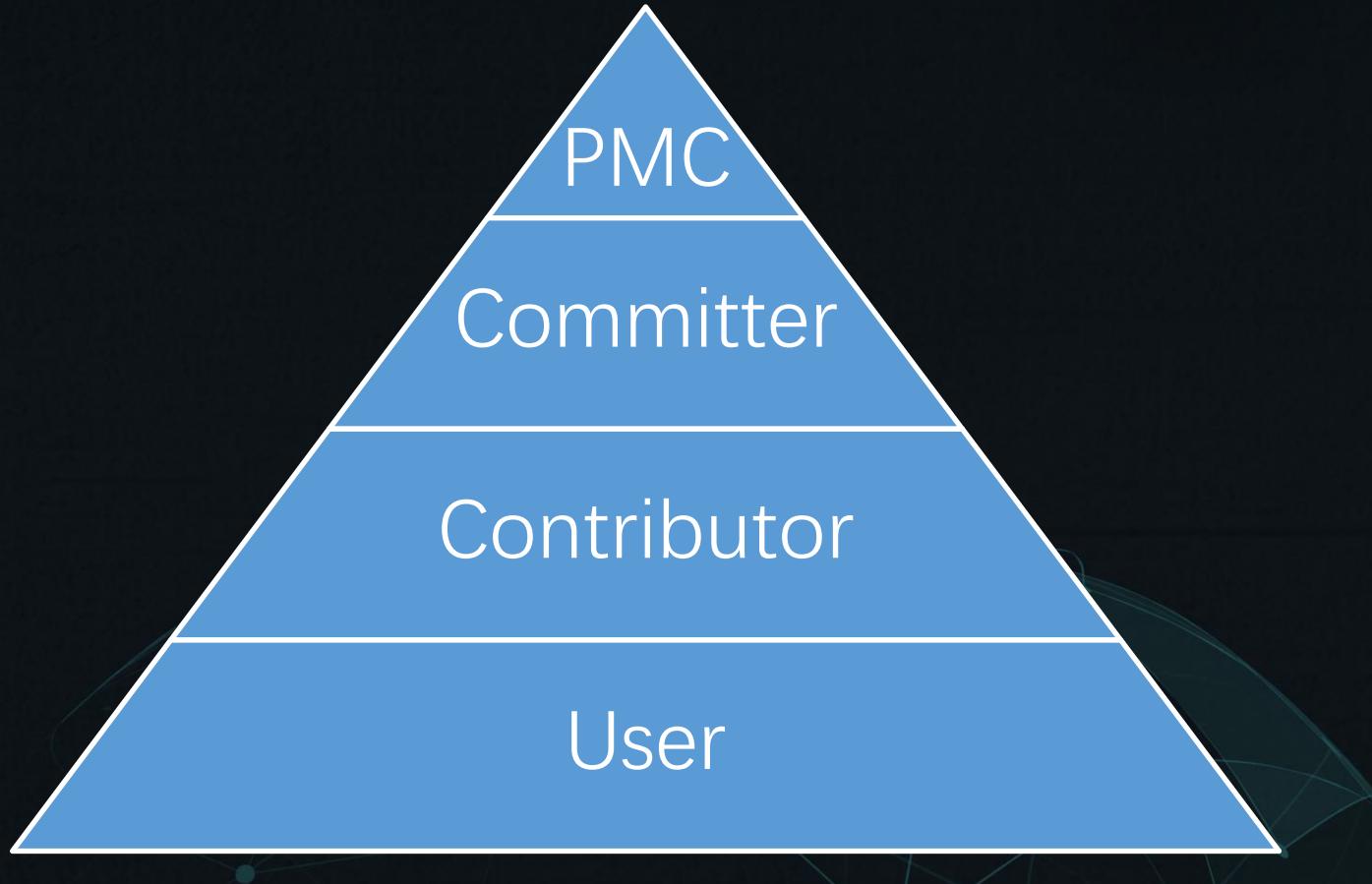
	Project	User	Group	Job	Issue	MR	Milestones	Note
2019.07.22	42	27	10	203	65	4	12	348
2020.01.16	191	74	20	1440+	397	77	23	2481
增长率	355%	174%	100%	609%	510%	1825%	92%	613%



X-lab 2019 年度 Gitlab 平台
最活跃用户: 王伟老师
最活跃项目: kfcampus/edu-web
热词: 开个 issue
新学期热词预测: 提个 PR
总结: GitLab已成为 X-lab 重要的工作学习科研协作平台



开源协作与运营的广泛需求：协作机器人



Open source PRA for open organization

故事5：面向开放组织的流程协作机器人

<http://www.hypertrons.io/>



weekly-report APP 12:00 PM
Weekly Report of hypertrons/hypertrons

This is a weekly report of hypertrons/hypertrons. It summarizes what have changed in the project during the passed week, including pr merged, new contributors, and more things in the future.

1. Repo Overview

Basic data

Basic data shows how the watch, star, fork and contributors count changed in the passed week.

Watch	1
Star	5(-)
Fork	6(-)
Contributors	8(-)

Issues & PRs

Issues & PRs show the new/closed issues/pull requests count in the passed week.

New Issues	Closed Issues	New PR	Merged PR
6	6	21	21

2. PR Overview

Thanks to contributions from community, 6 pull requests was merged in the repository last week. They are:

@GoodMeowing #233 feature: load history component
#242 feature: support gitee
#248 fix: merge config keep origin data
#247 refactor: update npm run test command, update ci Test stage
#250 refactor: update license checking

@WuShaoling

menbotics bot commented on 23 Dec 2019

Weekly Report of hypertrons

This is a weekly report of hypertrons. It summarizes what have changed in the project during the passed week, including pr merged, new contributors, and more things in the future.

Repo Overview

Basic data

Basic data shows how the watch, star, fork and contributors count changed in the passed week.

Watch	Star	Fork	Contributors
1	5 (-)	6 (-)	8 (-)

Issues & PRs

Issues & PRs show the new/closed issues/pull requests count in the passed week.

New Issues	Closed Issues	New PR	Merged PR
6	6	21	21

PR Overview

Thanks to contributions from community, 21 pull requests was merged in the repository last week. They are:

Contributor ID	Count	Pull Requests
@frank-zsy	10	#170 feature: add lua add issue and assign #181 doc: add README #187 feature: add first edition website by docsify #188 refactor: use lua ref instead of global random func name #189 feature: expose installation config #191 refactor: auto catch errors in ts inject functions #192 refactor: change frank login #193 fix: fatal bug

WuShaoling reviewed on 25 Feb

WuShaoling left a comment

/approve

menbotics bot added the pull/approved label on 25 Feb

menbotics bot merged commit 24fe53b into hypertrons:master

PR Review 与 Merge 自动化管理

[Feature] Personal badge for all installed repositories

① Open frank-zsy opened this issue on 1 Apr · 0 comments

frank-zsy commented on 1 Apr

Hypertrons can provide svg image of all contributors in a community and their status in last week and all historical data, then all the data can be seen via the link and the link can be added into a markdown file.

menbotics bot added the kind/feature label on 1 Apr

Slack 专用项目周报

GitHub Issue 项目周报

Issue 自动 Label

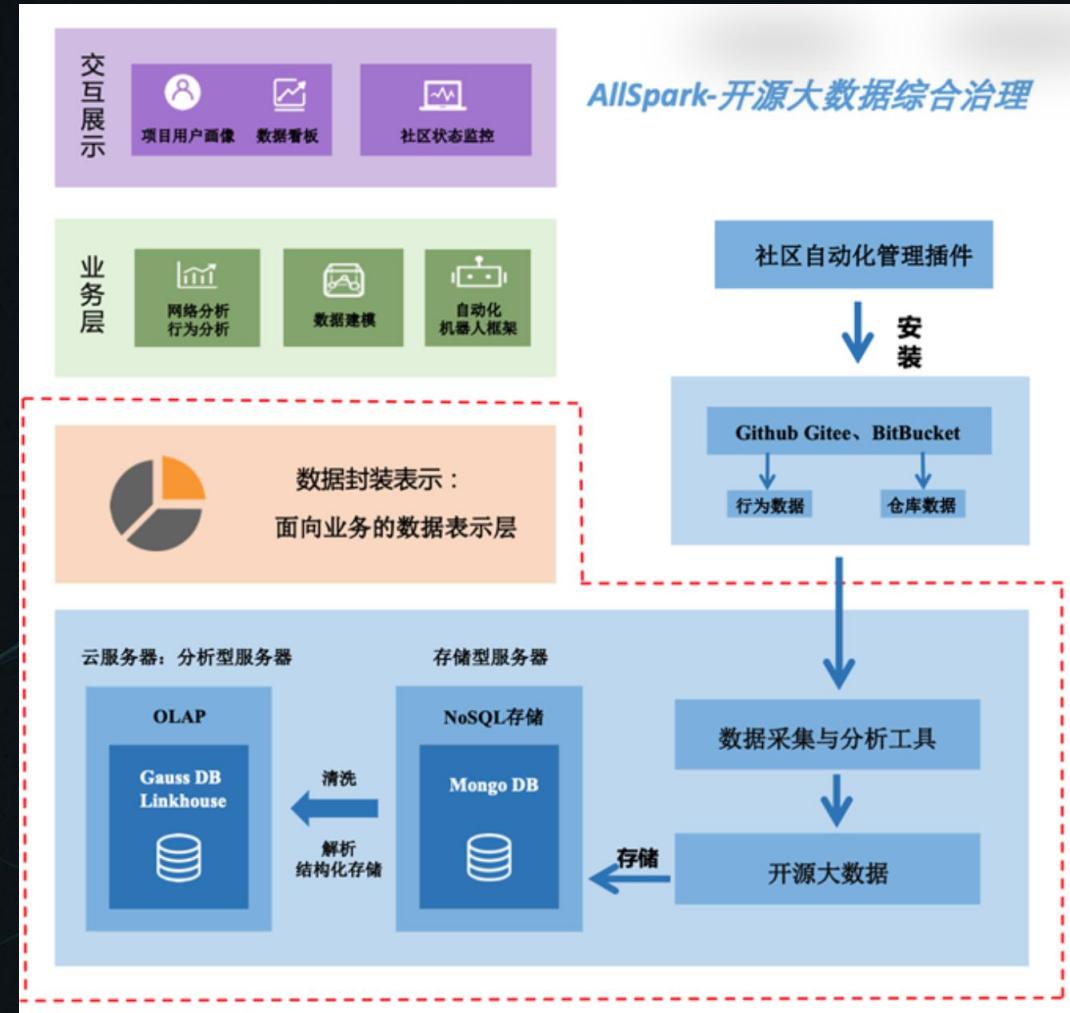
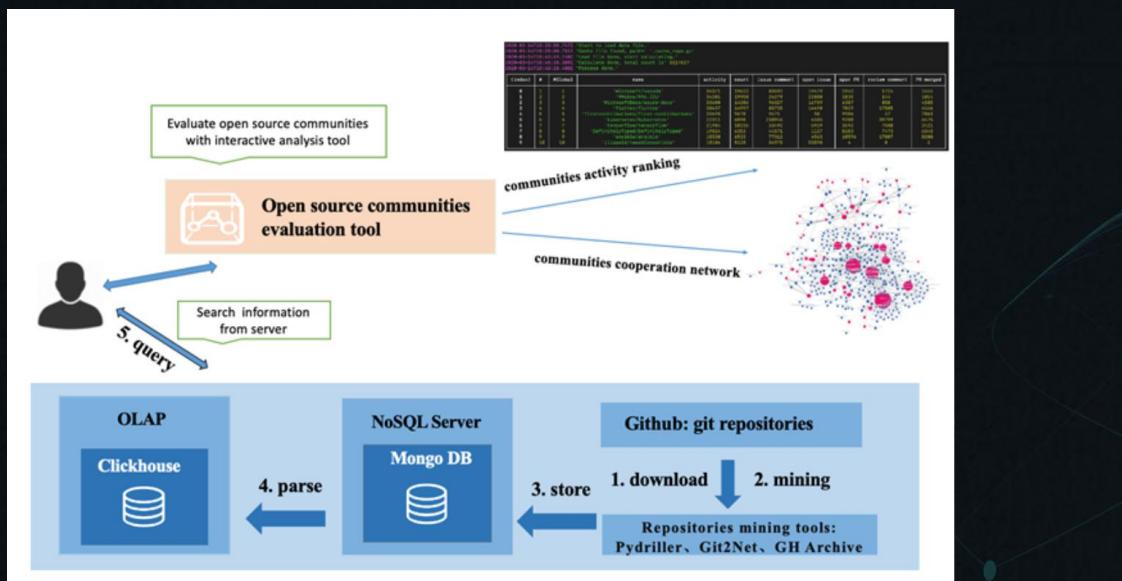
数据驱动的一站式开源研究与治理平台



AllSpark开源大数据平台



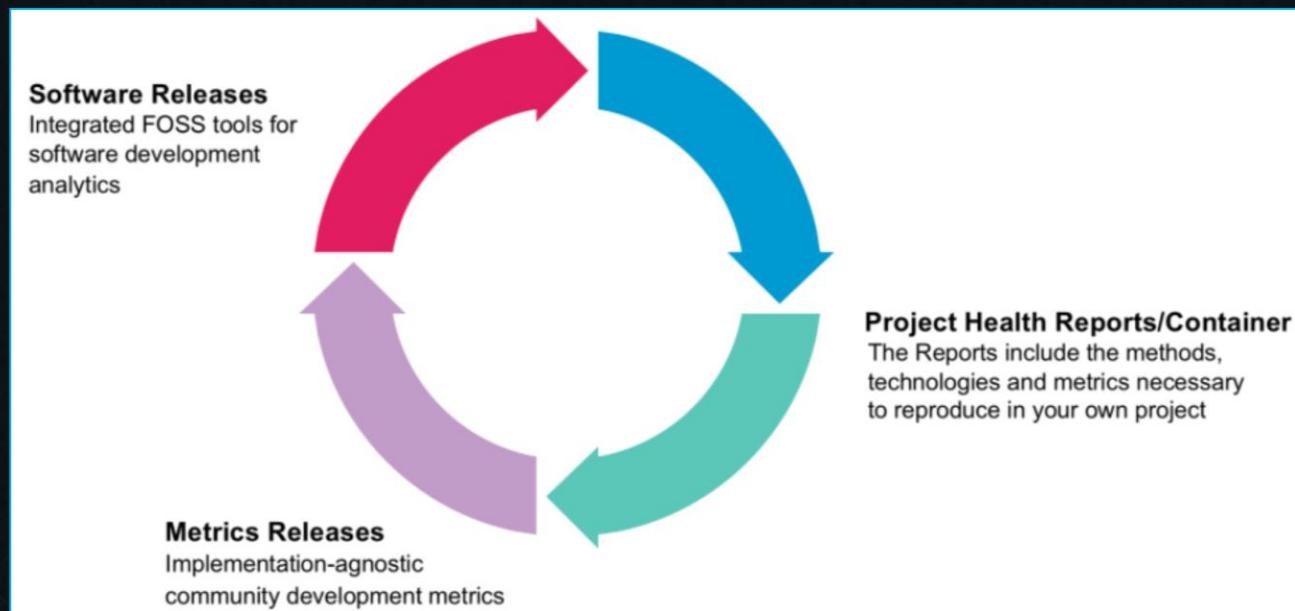
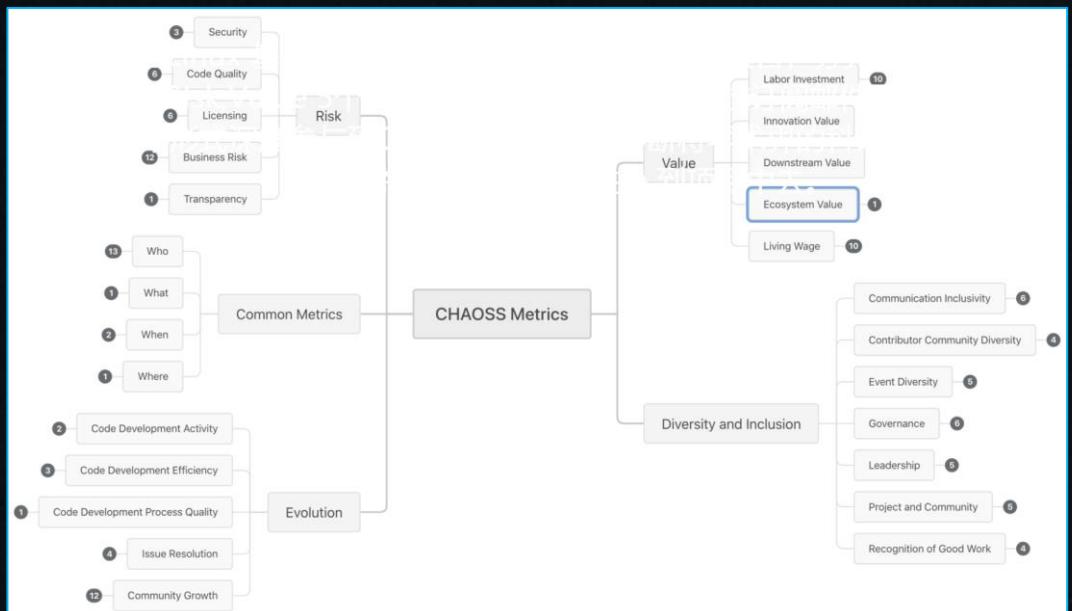
- AllSpark开源大数据平台是一个开源数据管理和分析平台，平台支持自动化的开源大数据采集和存储管理，定义开源数据指标和数据结构，支持高效数据计算分析、建模，呈现开源软件和社区生态的健康状态



CHAOSS上游项目的参与



- **CHAOSS**是由Linux 基金会发起的关于衡量社区健康度的项目，分为包括Common Metric、D&I、Evolution、Risk、Value总共5个Group，实验室同学分工通过订阅邮件列表、issue交流、参与周例会、提交PR等形式深度参与到CHAOSS项目中，不断将学术界的科研成果和产业界的经验和实际需求对接，更好的融入到项目中去。



故事6：实验室成员入选GSOC 2020项目



Ria Gupta Social Currency Metric System	Implementing Social Currency Metric System (SCMS) will be a huge milestone in providing a better and holistic view of project health in the open-source community. By adding social
Tianyi Zhou Large Social Network Analysis and Anomaly Detection with Augur	Augur is a software that collect data for a list of given repositories and provide a variety of CHAOSS metrics to provides open source health and sustainability metrics.Then users can
Sarit Adhikari Machine Learning for Anomaly Detection in Open Source Communities	Open-source software development is a collaborative effort that requires decentralized decision making from different developers and maintainers. In order to measure the progress
Abhinav Bajpai Data Collection Worker and Mapper to bind the responses of GitLab API, Github API & the Augur schema	The project aims to develop a Gitlab collection worker closely tied up with the Github collection worker using mapper files to bind the intended attributes of their API responses
bistaastha Build Workflow process for CHAOSS D&I Badging Project	CHAOSS Diversity and Inclusion Badging program aims to encourage projects and events to obtain badges for good diversity and inclusion practices. This project will be about building a
vchrombie Creating Quality models using GrimoireLab and CHAOSS metrics	GrimoireLab is a powerful open-source platform that provides support for monitoring and in-depth analysis of software projects. It produces a rich set of dashboards, which can be easily
Saicharan Reddy Implementation of GitLab Data Collection Worker & Test Coverage Improvement	The primary goal of this project is to congregate data pertaining to GitLab issues, commits, merge requests amongst other entities & store it into the unified data model ecosystem of
Akshara P Machine Learning for Anomaly Detection in Open Source Communities	Augur is a Flask based prototyping web stack for CHAOSS metrics. It provides structured data mined from various sources like git repositories, mailing lists and issue trackers using a plugin
Pratik Mishra Machine Learning for Anomaly Detection in Open Source Communities	This project will play one of the most vital roles in achieving Augur's goal towards analysis of open source organisation health.It will not only provide visualisation but also offer useful

Google Summer of Code 2020 Projects

Search for a project, organization, or student [VIEW BY ORGANIZATION >](#)

Tianyi Zhou
Large Social Network Analysis and Anomaly Detection with Augur

ORGANIZATION
CHAOSS

MENTORS
Sean Goggins
Jonahz
Gabe Heim

SHARE LINK

Augur is a software that collect data for a list of given repositories and provide a variety of CHAOSS metrics to provides open source health and sustainability metrics.Then users can empirically investigate and uncover useful insights for software engineering, such as understanding the collaborative pattern.

In the current stage, Augur is unable to mining the repositories to generate the co-editing information in the open source software development. This project idea is about to develop this kind of functions to enhance the Augur project. It will end up with a new data worker and analysis tool for fine-grained co-editing network which opens up a massive new source of high-resolution data on human collaboration patterns.

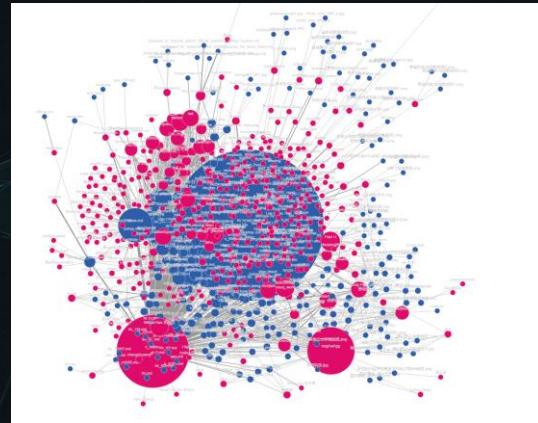
Then I would like to establish the Augur server to mining and monitor the open source ecosystem (with 10000+ repositories or more). The collaboration networks and social trace data of all contributors in the open source ecosystem will be evaluated in terms of social network analysis. It will play a vital role in achieving Augur's goal towards analysis of open source organizations health as well as the goal of the CHAOSS project towards diversity and inclusion.

故事7：GitHub 2019 数字年报

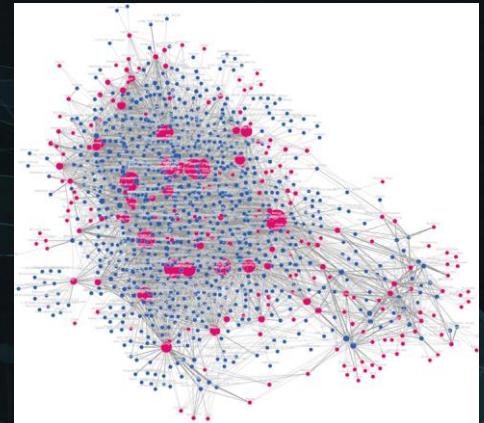
- GitHub 2019 数字年报项目旨在通过分析Github全网的开发者行为日志，通过数据的视角，来观察全球范围内的开源现状、进展趋势、演化特征、以及未来挑战等问题，除了展现目前开源世界全貌之外，我们特别关注中国的开发者和企业组织在整个开源产业中的表现。
- 本次使用 2019 年全年 GitHub 日志进行统计，总日志条数约 5.46 亿条，相较 2018 年的 4.21 亿条增长约 29.7%。在上述开发者活跃度与项目活跃度的定义下，统计得到 2019 年总活跃项目数量约 512W 个，相较 2018 年的约 313W 增长约 63.6%，2019 年总活跃开发者数量约 360W，相较 2018 年的约 303W 增长约 18.8%。

(index)	#	#Global	name	activity	count	issue comment	open issue	open PR	review comment	PR merged
0	1	1	'dependabot[bot]'	3641110	1430942	272213	0	3006606	0	374345
1	2	2	'direwolf-github'	589065	340091	0	0	340105	0	0
2	3	3	'dependabot-preview[bot]'	503923	42846	808143	19595	1675424	0	790491
3	4	4	'pull[bot]'	281141	26699	0	0	1127083	0	1124472
4	5	5	'github-learning-lab[bot]'	262617	86142	372192	142165	43742	12864	33362
5	6	6	'renovate[bot]'	200693	26442	37663	1807	574740	0	455955
6	7	7	'greenkeeper[bot]'	147530	21873	825189	56071	161884	0	81707
7	8	8	'imgbot[bot]'	105982	54184	0	0	58822	0	11459
8	9	9	'autotester-one'	88989	47856	0	89029	0	0	0
9	10	10	'snyk-bot'	88096	26179	1	0	126804	0	10589

(index)	#	#Global	name	activity	count	issue comment	open issue	open PR	review comment	PR merged
0	1	1	'dependabot[bot]'	3641110	1430942	272213	0	3006606	0	374345
1	2	2	'direwolf-github'	589065	340091	0	0	340105	0	0
2	3	3	'github-learning-lab[bot]'	262617	86142	372192	142165	43742	12864	33362
3	4	4	'imgbot[bot]'	105982	54184	0	0	58822	0	11459
4	5	5	'autotester-one'	88989	47856	0	89029	0	0	0
5	6	6	'dependabot-preview[bot]'	503923	42846	808143	19595	1675424	0	790491
6	7	7	'pull[bot]'	281141	26699	0	0	1127083	0	1124472
7	8	8	'renovate[bot]'	200693	26442	37663	1807	574740	0	455955
8	9	9	'snyk-bot'	88096	26179	1	0	126804	0	10589
9	10	10	'greenkeeper[bot]'	147530	21873	825189	56071	161884	0	81707



996 .ICU全年贡献图



TiDB 2019全年贡献图

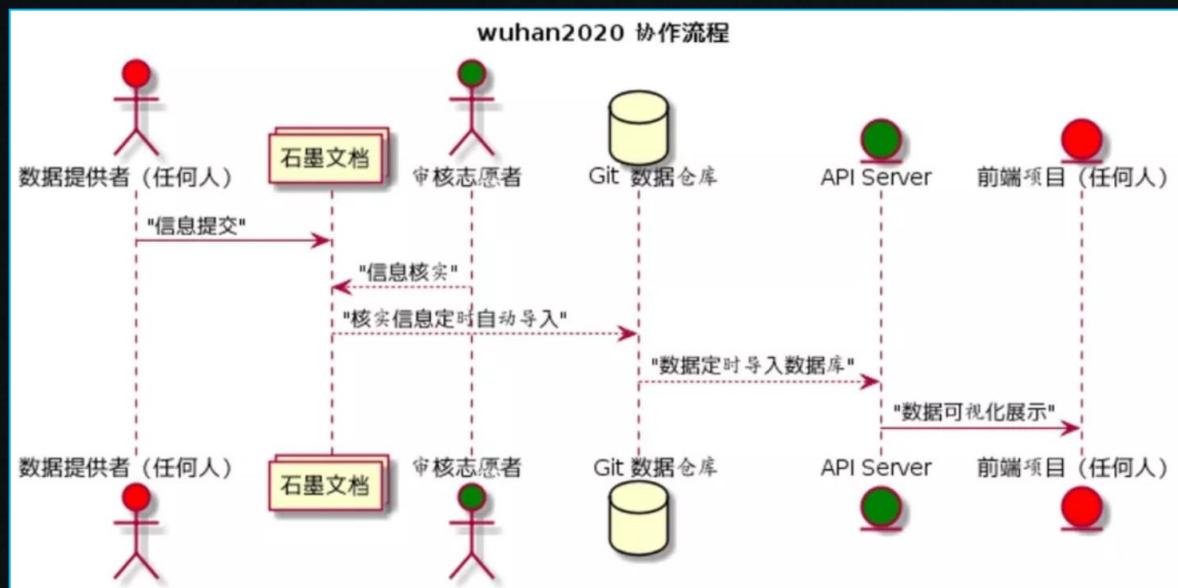
好评如潮！

<https://github.com/X-lab2017/github-analysis-report-2019>

故事8：Wuhan2020开源社区



- 2020年1月25日，春节，X-lab实验室正式发起了《wuhan2020：武汉新型冠状病毒防疫开源信息收集平台》的开源项目，旨在统一收集本次事件中相关事务处理方的信息，并利用开源和分布式协作优势实时更新并通报，提供各方的联系平台。
- 面对突如其来的新型冠状病毒疫情，信息采集与公布平台不一致，信息散乱很难进行有效沟通，大量图片信息不利于实时沟通。作为开发者群体，我们希望利用数字平台优势，让各供需方进行分布式自助对接可大幅提升效率。故发起该项目。



WUHAN
2020

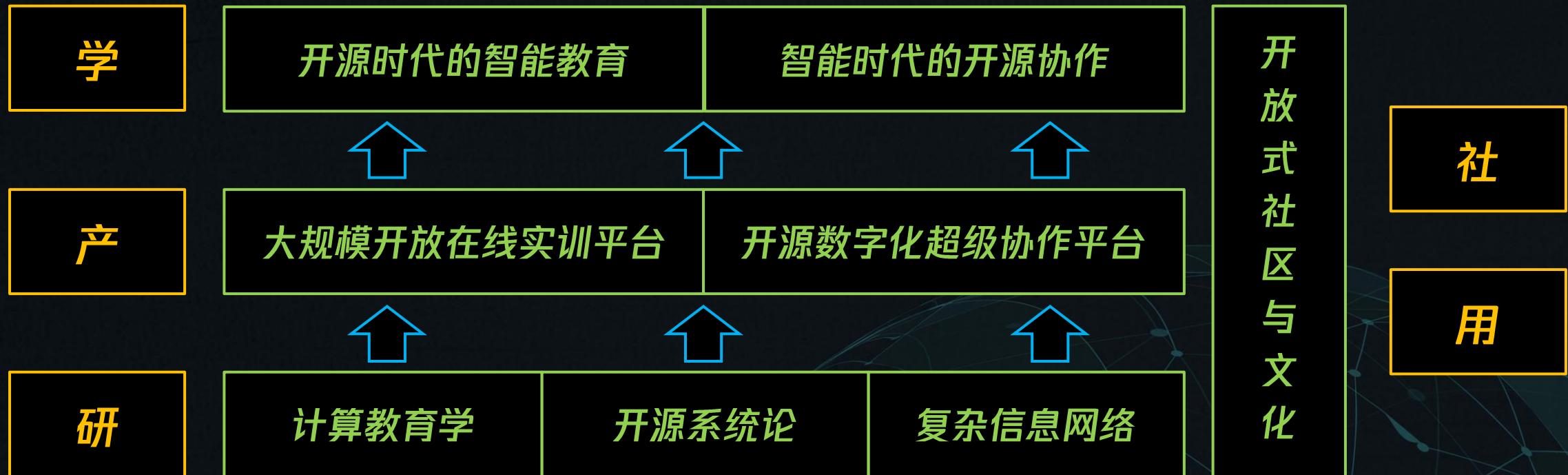
首页 文档 贡献者 博客 社区 项目列表 黑客松 新型冠状病毒防疫信息平台 简体中文

Wuhan2020 社区是做什么的？

我们当前在做的事是：建立一个医院、工厂、采购等信息实时同步的数据服务，同时召集所有希望对这次抗击病毒战役进行贡献的人，让每个具有相关技能的人都可以参与相关主题的开发工作，用开源的社区文化，以自组织协作方式完成。

黑客马拉松 HACK FOR WUHAN
决赛即将开启
时间：3月8日
方式：线上直播

总结一下



研究方向1: 计算教育学

Computational education



The image displays a multi-panel interface for a computational education platform:

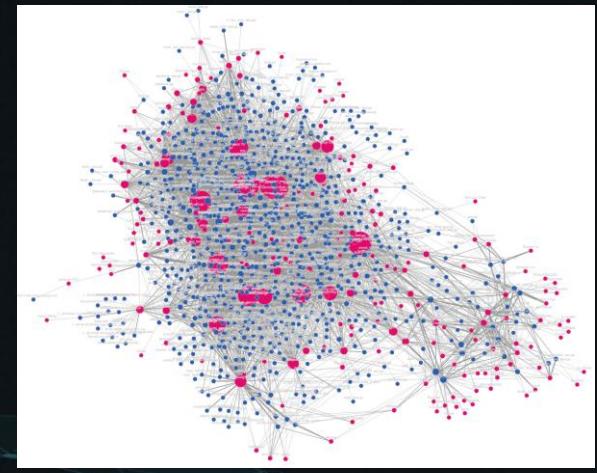
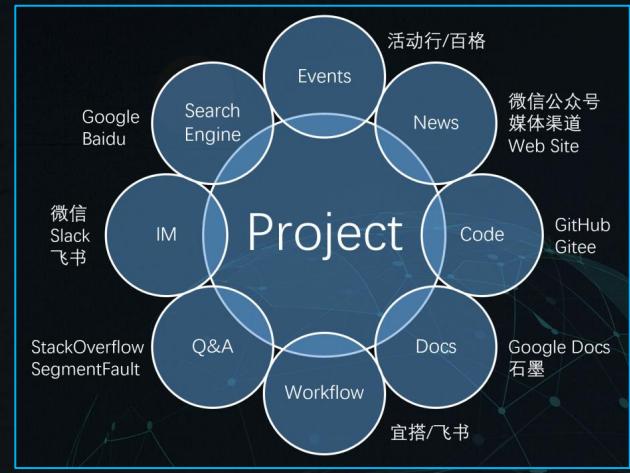
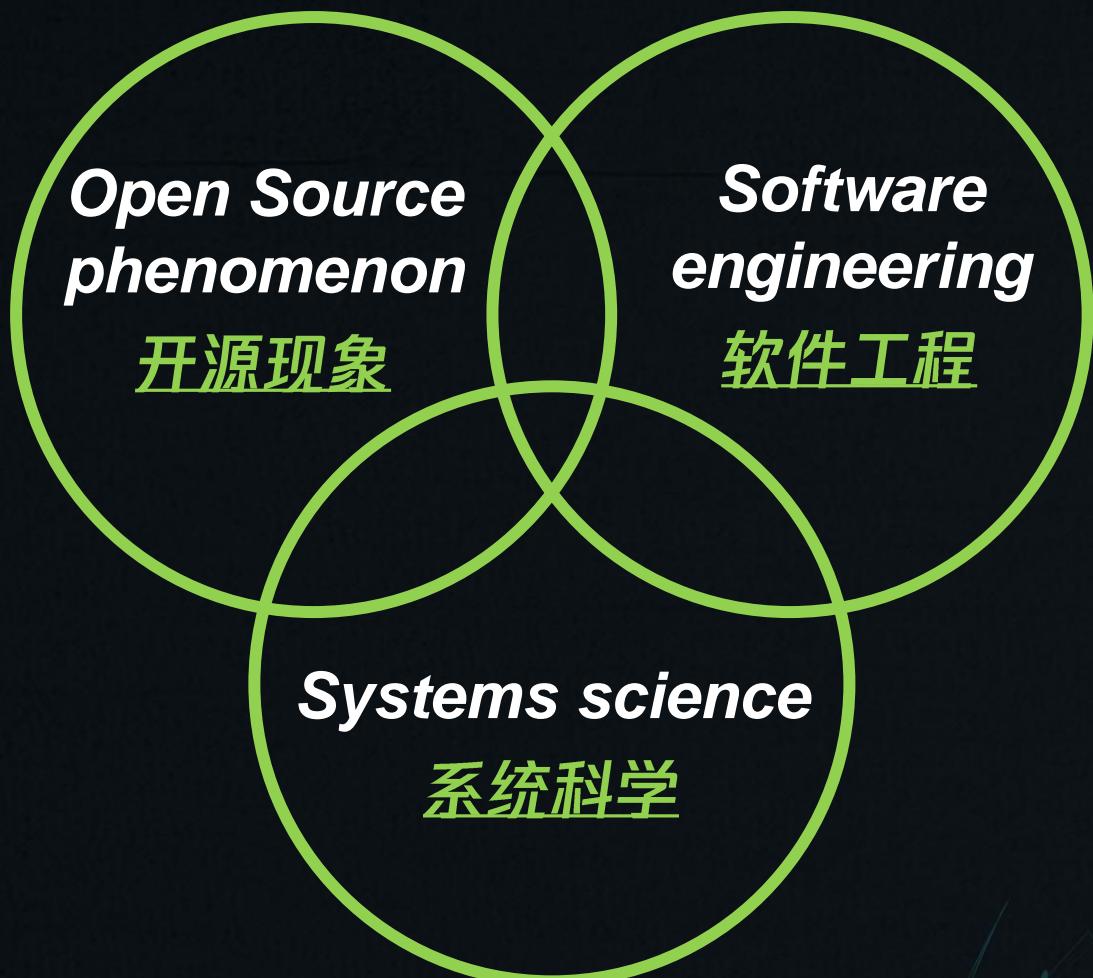
- Top Left Panel:** A course navigation menu for "程序设计与人工智能 王老师班" (Program Design and Artificial Intelligence Class). It includes sections like "0 课程预备", "1 图形界面编程", "2 数据的获取和分析", "HTML简单教程", "高维数据格式", "Python网络爬虫", and "正则表达式".
- Top Right Panel:** A "Hello World!" example in C programming. It shows a code editor with the following code:

```
1 #include<stdio.h>
2 Int main(){
3     printf("hello X-lab !\n");
4 }
5 }
```

The terminal window shows the command: `cd "/home/" && gcc 'main.c' -o 'main' && ./main`. The output is: `hello!`
- Bottom Panel:** A dashboard titled "一站式IT实训服务平台" (One-stop IT Training Service Platform) featuring a "交互式学习体验" (Interactive Learning Experience) button and a keyboard icon with hands. To the right are three sub-panels: (a) showing user activity over time, (b) showing the most active people, and (c) showing the most popular activities.

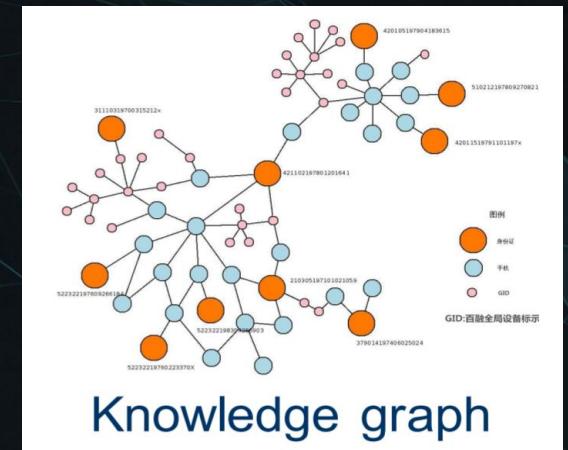
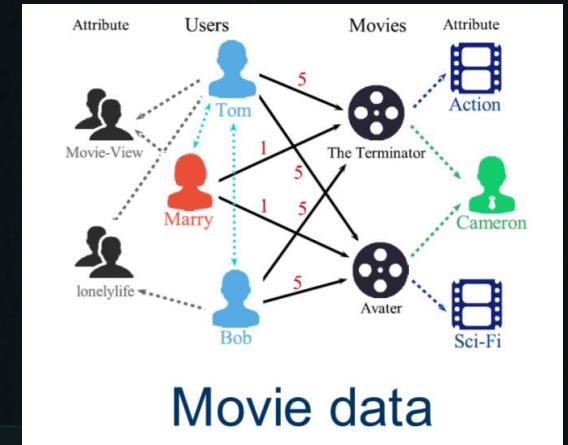
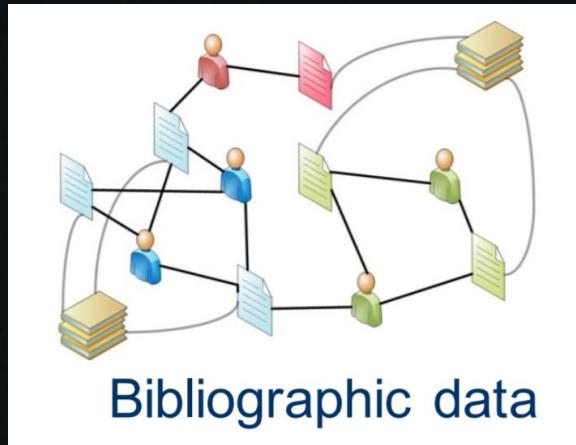
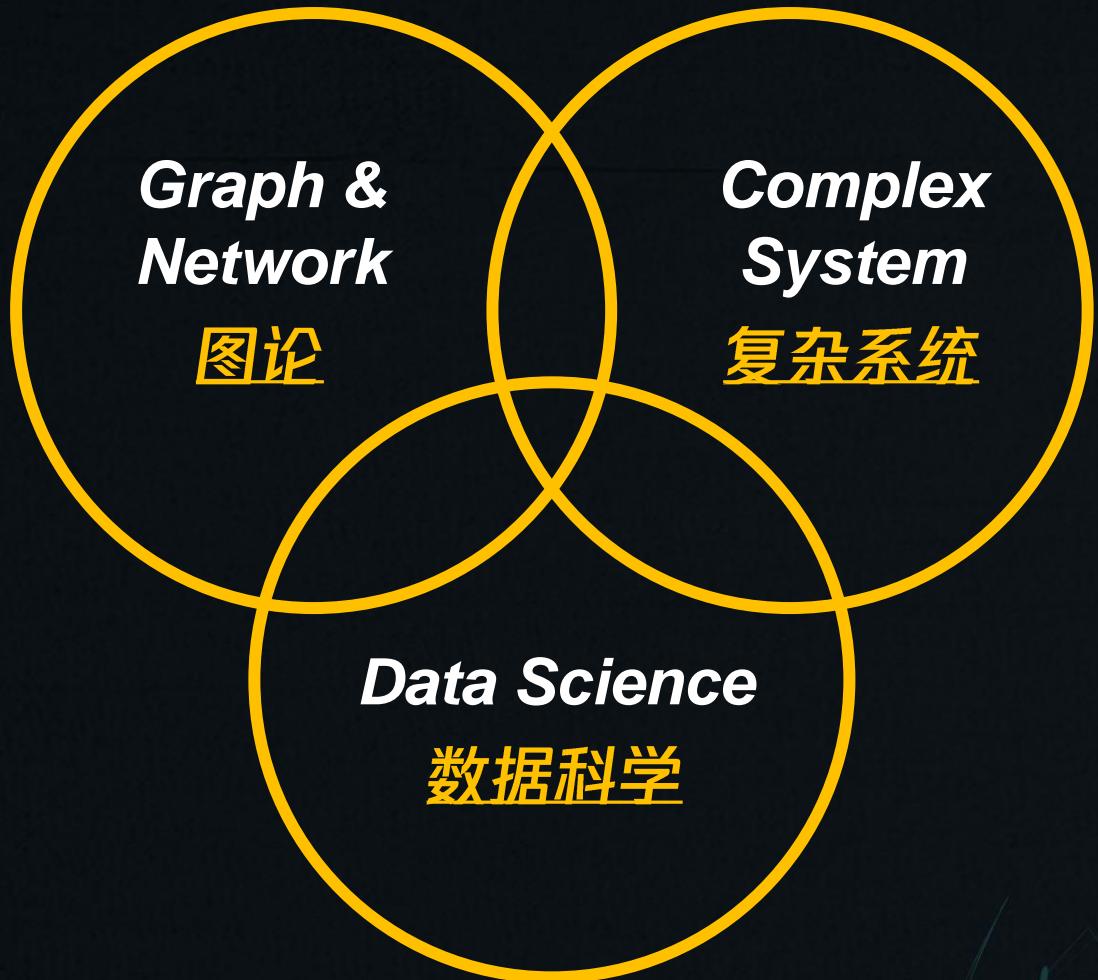
研究方向2：开源系统论

Scientific principles of open source systems



研究方向3: 复杂信息网络

Complex information network



Current members



事求人 Open Source Career in X-lab

- 研究方向：开源教育平台、开源协作机器人、开源系统科学
- Position：硕士、博士、博士后、工程师、访问学者
- 我们的优势：
 - 前沿与深度的技术
 - 开源文化与高度的自由度
 - 学术界、工业界、社区大牛导师
 - 创意激发与垂直成长
- **开源相关经验者优先！**
- 联系邮箱：wwang@dase.ecnu.edu.cn





THANK YOU



X-lab 公众号