

《开源软件开发与社区治理》

Kick off

主讲人：王伟



- 开源软件，作为一种媒介，可以帮助我们触及科学精神，帮助我们探索科学精神的实质。
- 科学本身就是一种开源的事业，开源软件开发就是一种科学实践活动。
- 只要您从事过任意一项开源开发项目，您就能体会到科学精神的存在，并且对它的巨大威力所震撼。



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

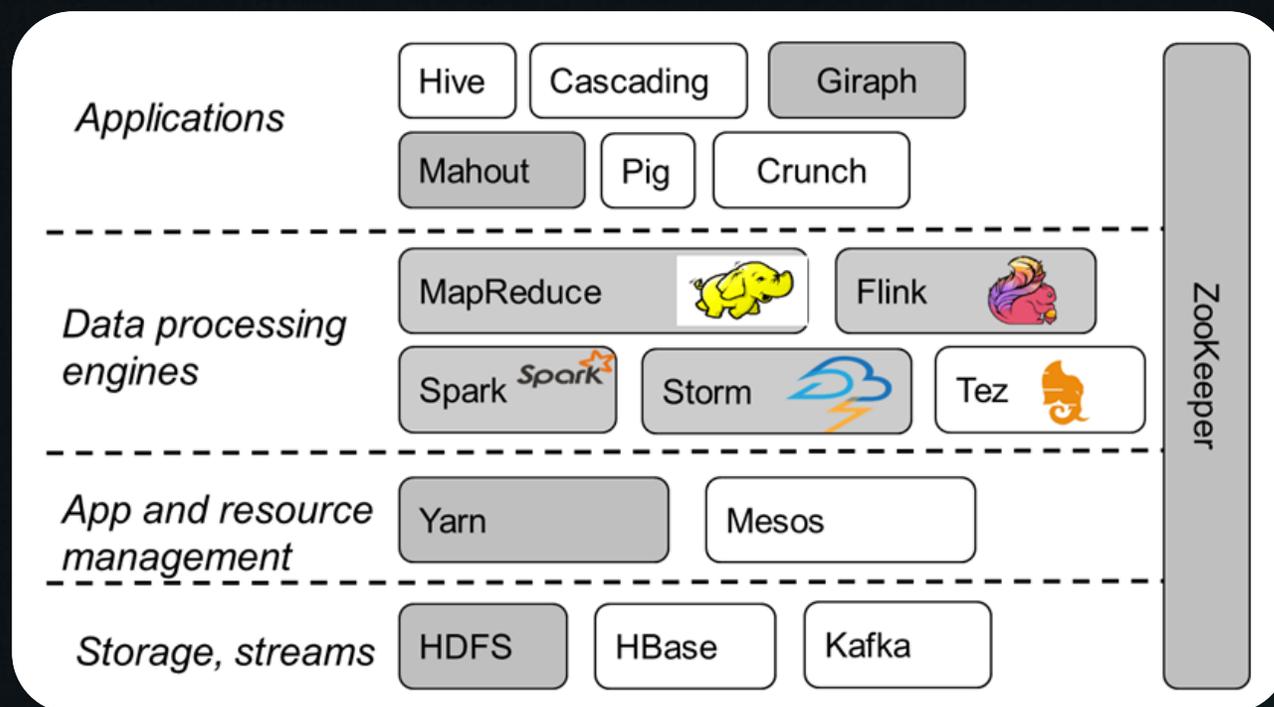
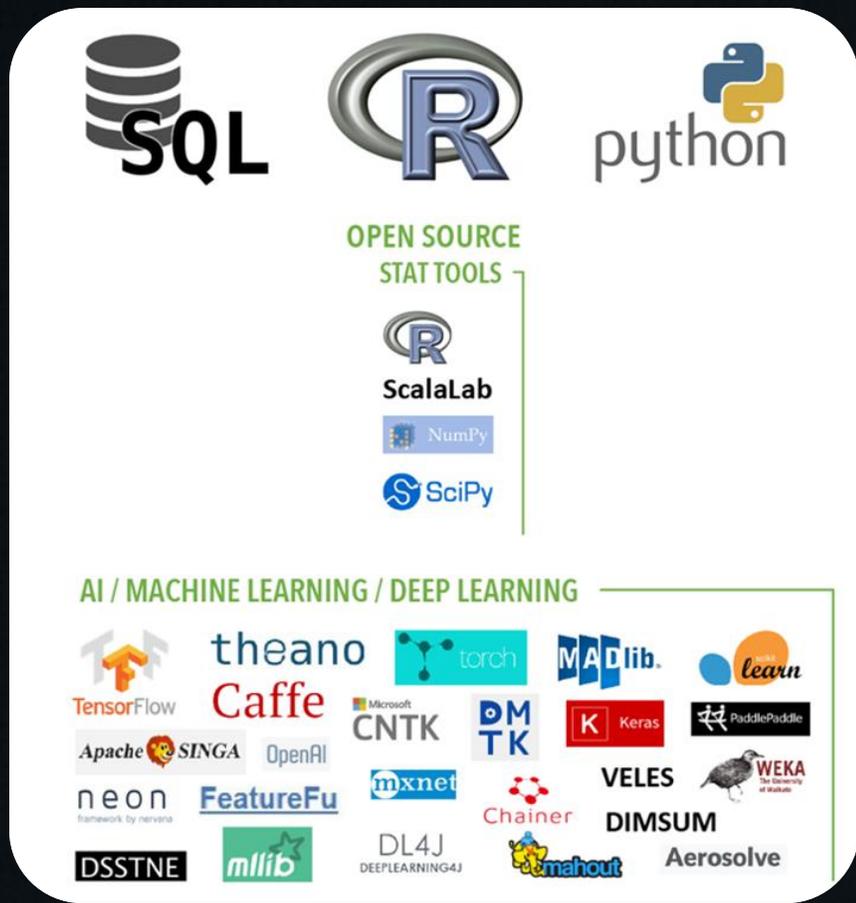


Satya Nadella, 2014



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!

产业的故事：全球 IT 产业竞争与碰撞

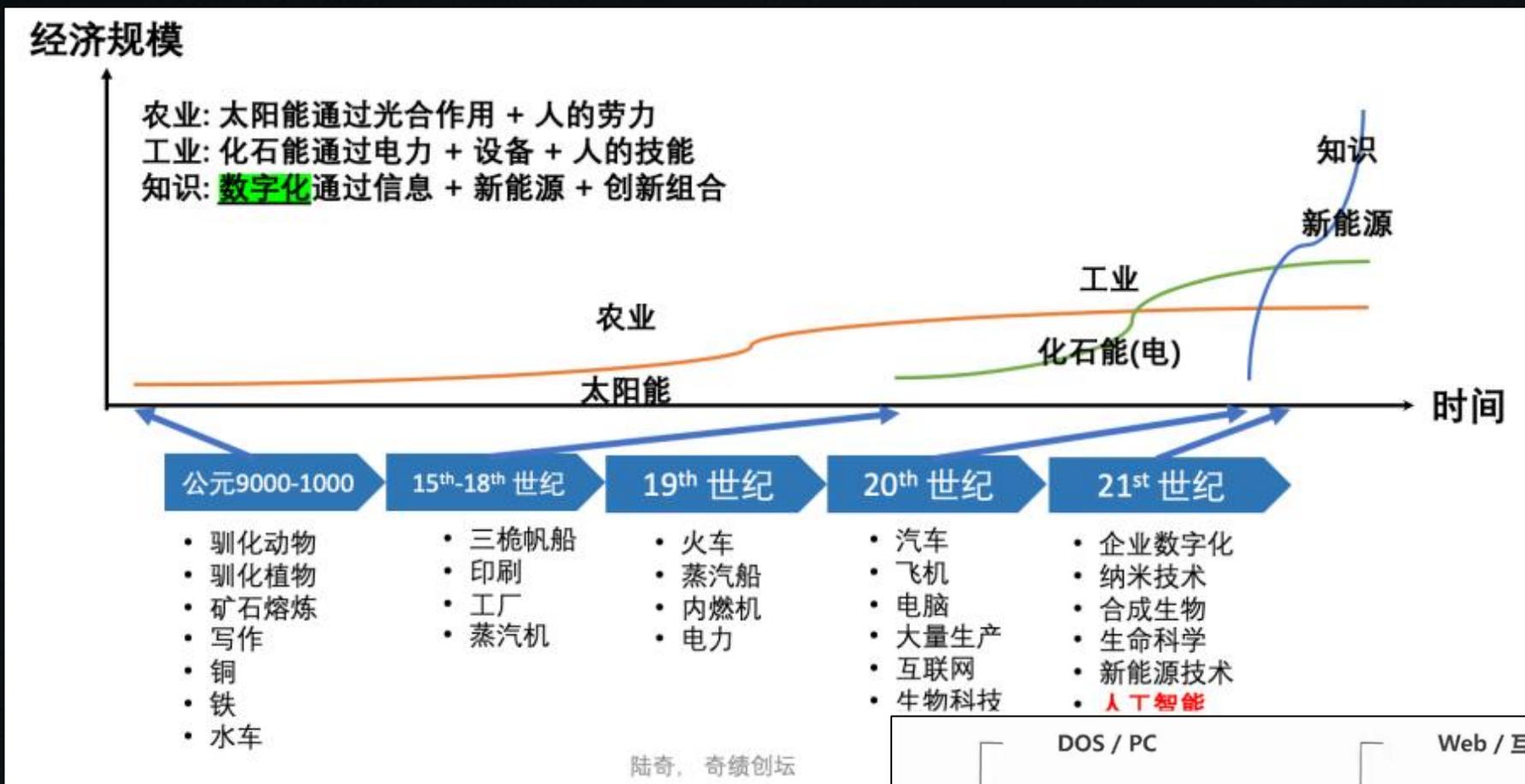
- 突破：国外厂商垄断的“WinTel联盟”
- 华为：基于鲲鹏生态的“第二计算平面”
- 支持：发起建设“第二计算教育平面”

	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



• 数字化核心要素

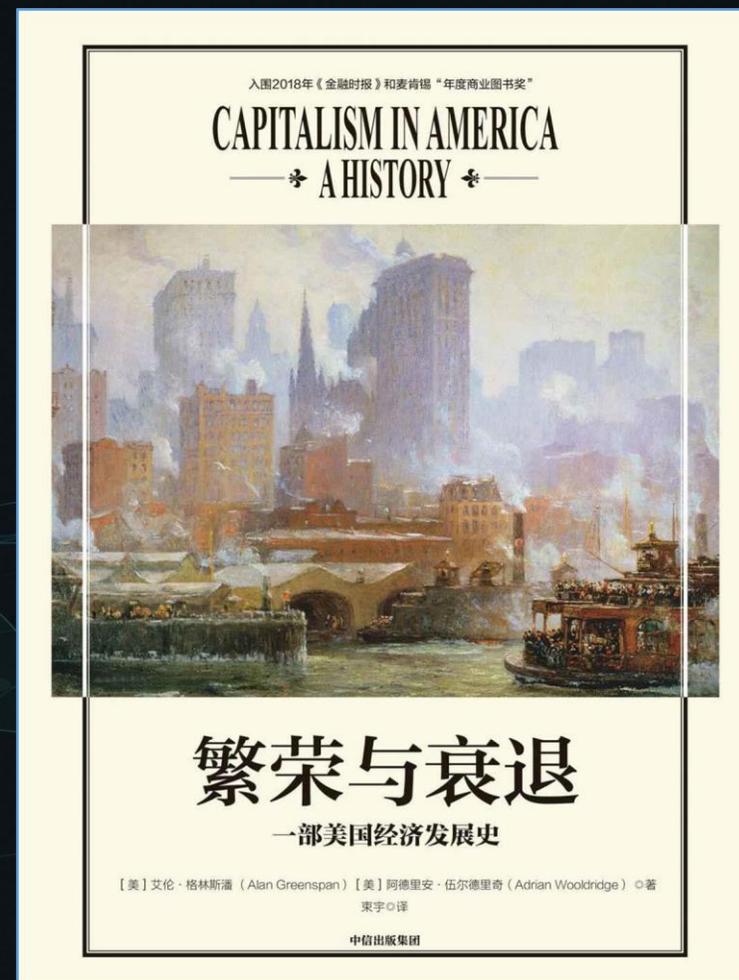
- 获取信息
- 表达信息
- 存储信息
- 传输信息
- 处理信息
- 交付信息

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!

人工智能时代的挑战



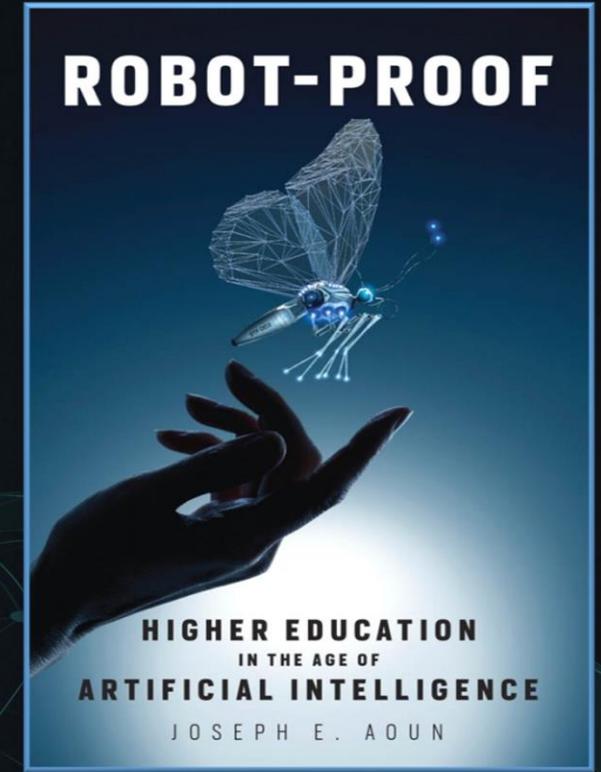
ALL Systems Go

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



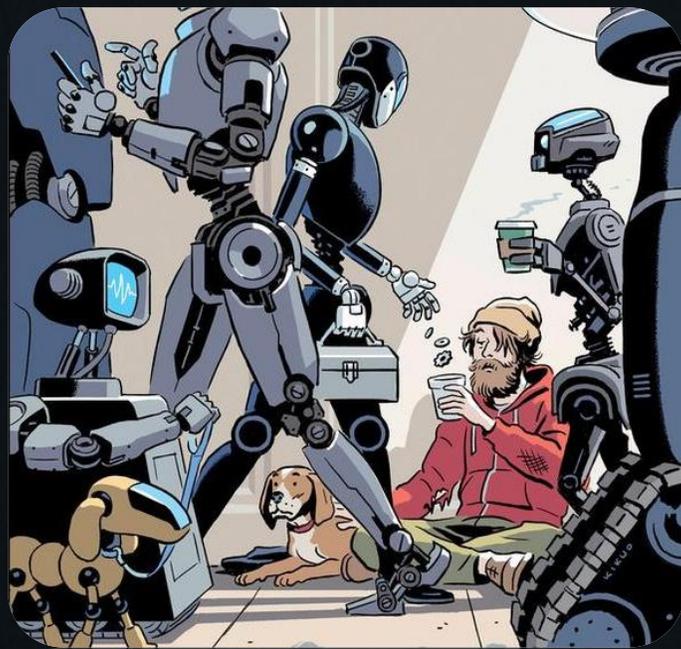
Robot-Proof

Higher Education in the Age of Artificial Intelligence
MIT Press 2017.08

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

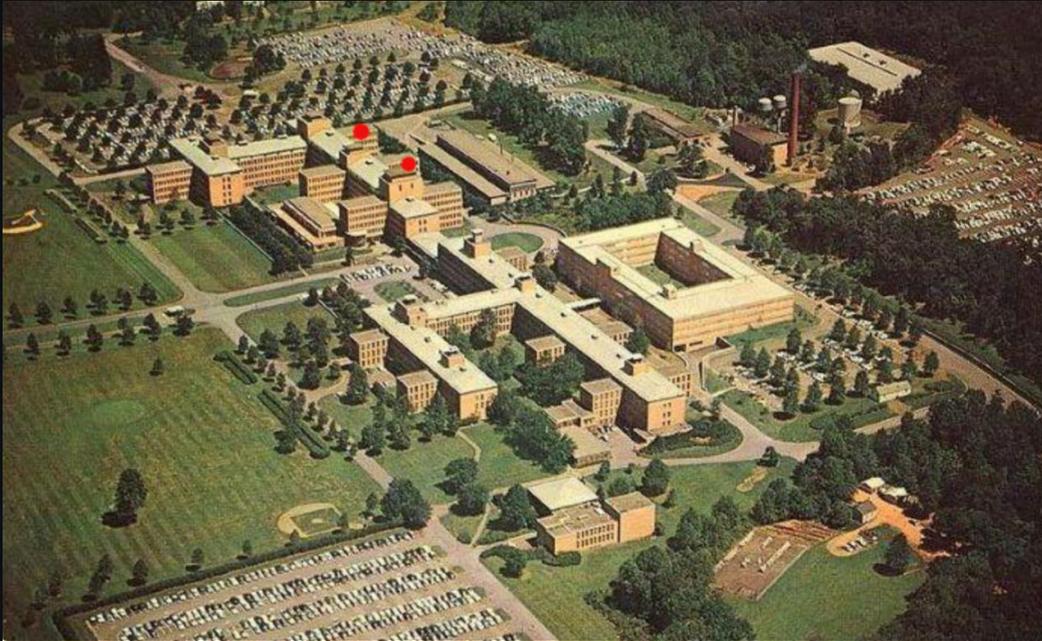


启示

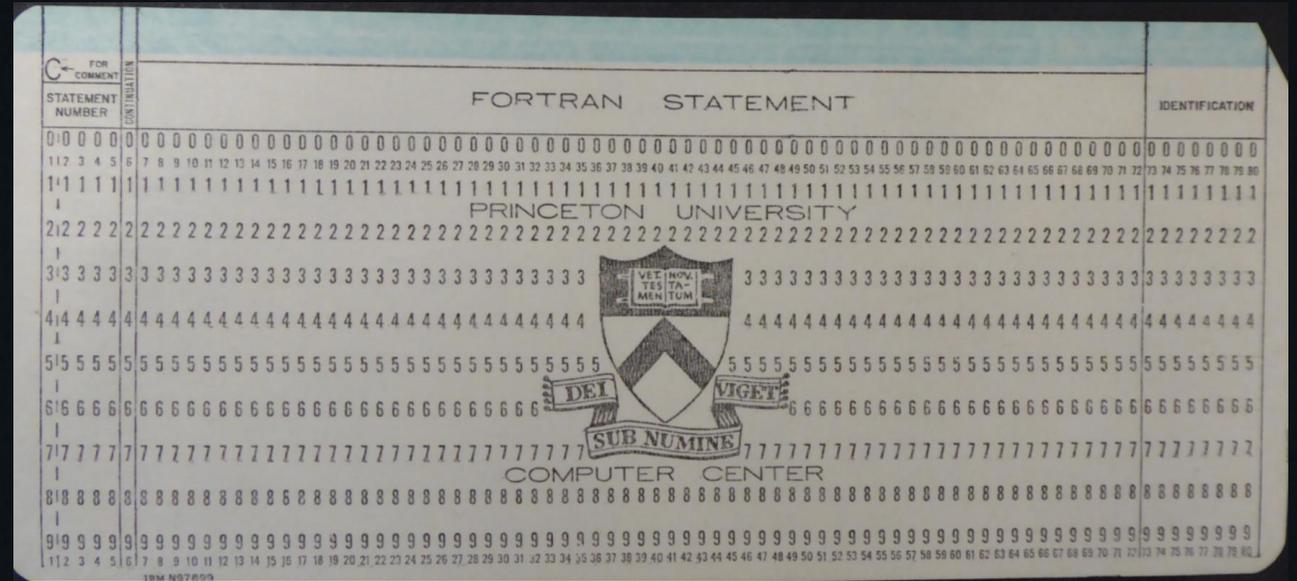
数字时代的**开源运动**

开源时代的**数字教育**

缘起：软件与操作系统



Bell Labs in 1961 (Courtesy of Bell Labs)



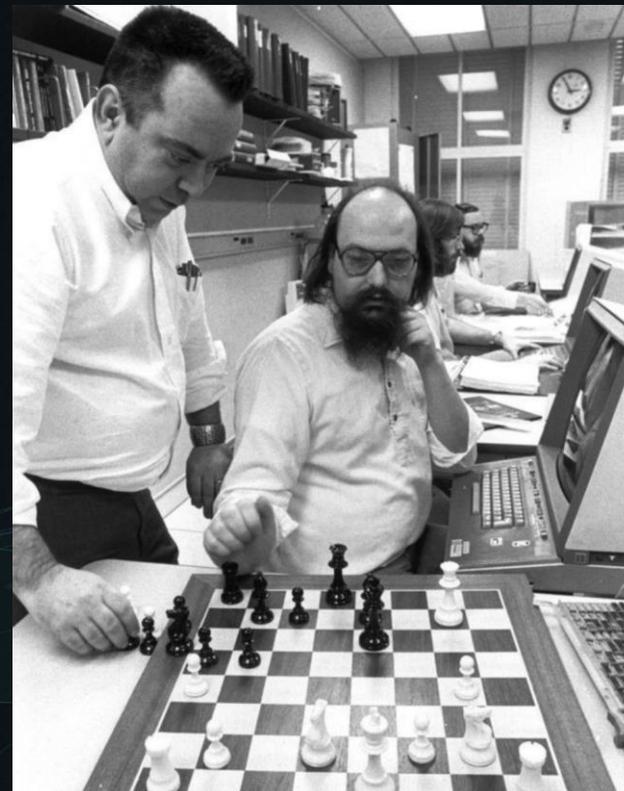
Punch card, 7-3/8 by 3-1/4 in [187.325 mm by 82.55 mm]

1961年的**CTSS** (Compatible Time-Sharing System) 和1965年的**Multics** (MULTiplexed Information and Computing System)

缘起：软件与操作系统



Ken (seated) and Dennis at the PDP-11



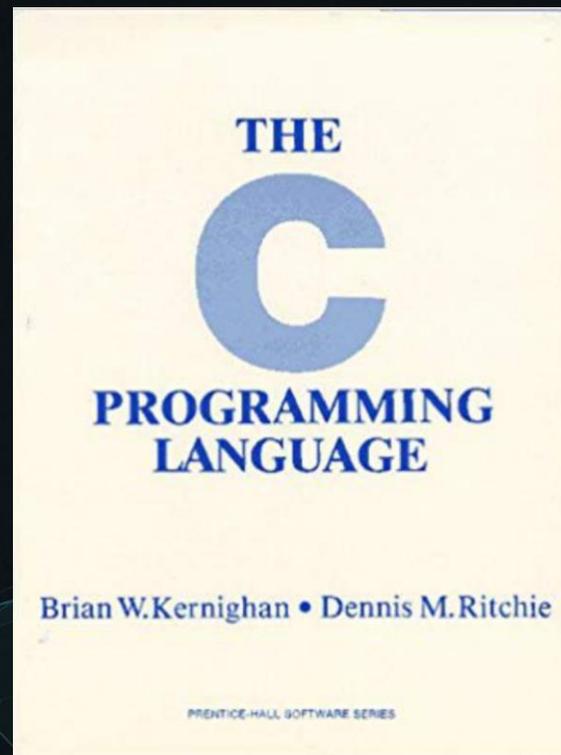
Ken Thompson and Joe Condon

Kenneth Thompson 和 Dennis Ritchie 在 Multics 和 CTSS 的基础上，于 1969 年在小型计算机上开发 UNIX 系统，后于 1970 年投入运行。

缘起：软件与操作系统

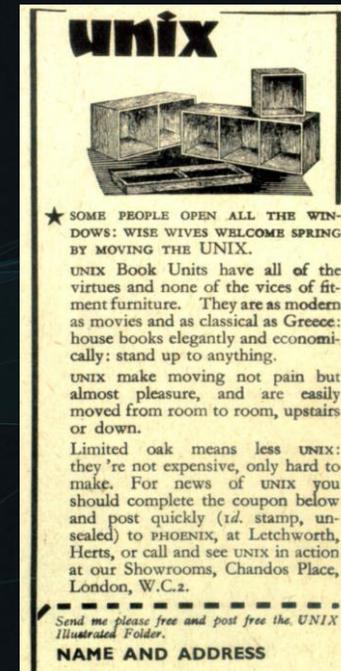
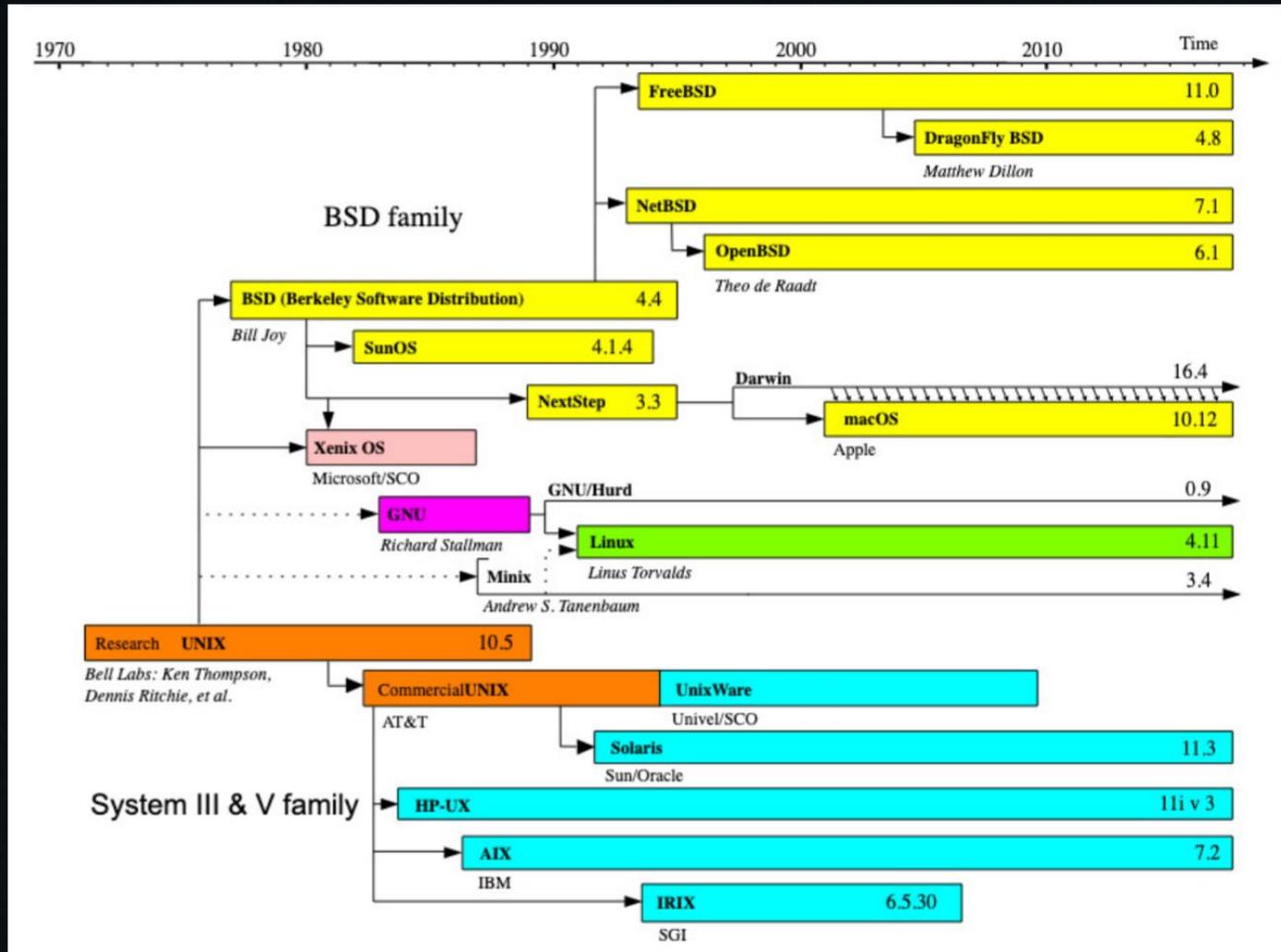


Doug McIlroy and Dennis Ritchie, May 2011



1973年，Dennis Ritchie 开发出 C 语言，用来改写原来用汇编语言编写的 UNIX，由此产生了 UNIX Version V。1974年，Kenneth Thompson 和 Dennis Ritchie 合写的“The UNIX Time-Sharing System”在Communication of ACM上发表，正式向外界披露了UNIX系统。

成功：软件的商业化

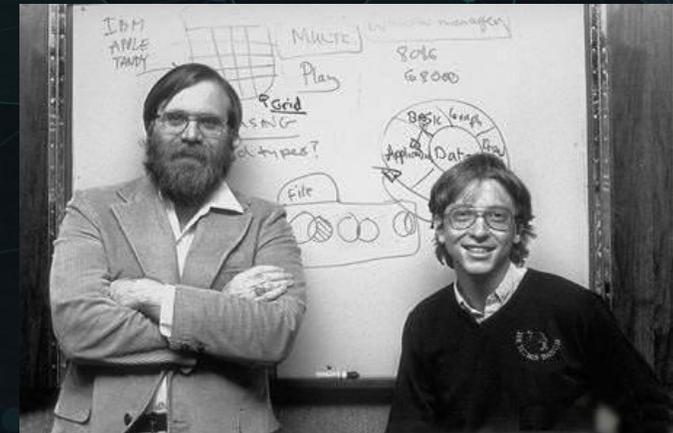
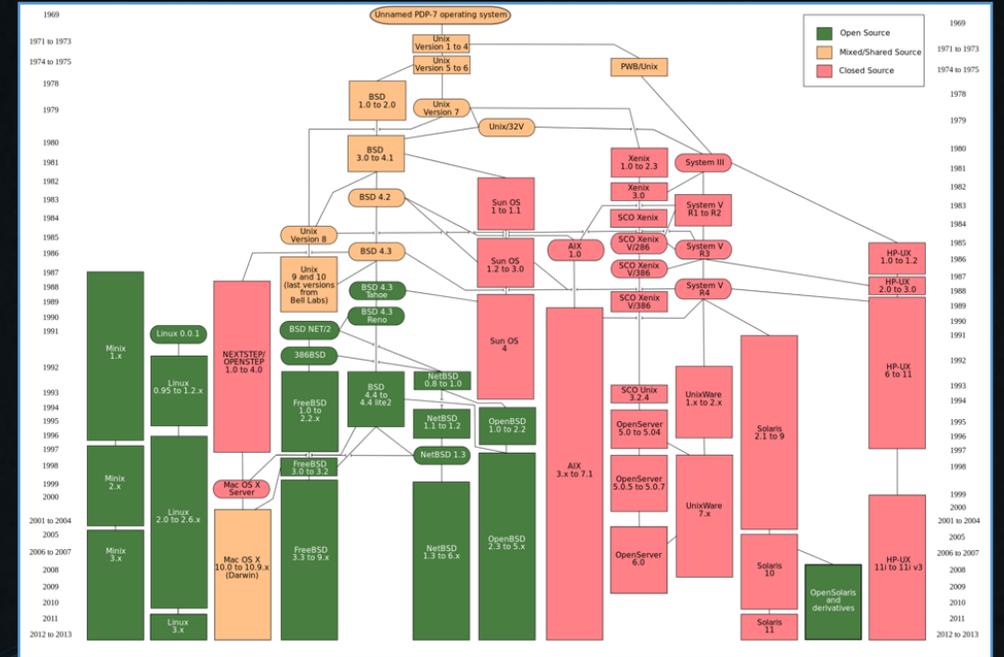


Unix timeline, from Wikipedia

成功：软件的商业化

操作系统之争

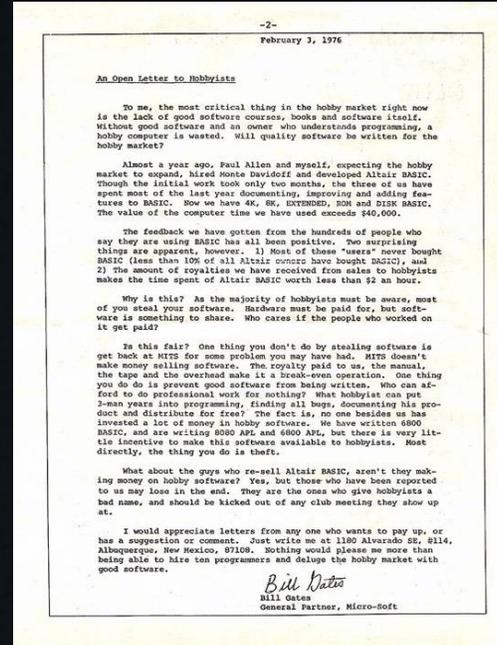
- **Unix**: 1970 Bell lab, AT&T. Free to share and use.
IBM: AIX, HP: HP-UX, Sun: Solaris
- **BSD Unix**: Unix distribution from UC Berkeley
MacOS
- Law case between AT&T and BSD
- **Microsoft**: 1975 Bill Gates & Paul Allen
- **MS-DOS, Windows**



革命：自由软件运动



Richard Stallman 的主要成就包括 Emacs 和 GNU C 编译器。他所写作的GNU通用公共许可证（GPL）是世上最广为采用的自由软件许可证，为 Copyleft 观念开拓出一条崭新的道路。



Linus Torvalds, 芬兰赫尔辛基人，Linux内核的发明人及该计划的合作者，毕业于赫尔辛基大学计算机系。

革命：自由软件运动

GNU & Linux

- 1984.1 Richard Stallman, MIT, GNU(GNU's Not Unix) & Free software movement
- 1985: FSF Free Software Foundation
- 1989: GPL GNU General Public License
- 1991: emacs, gcc, gdb, etc. But not kernel(Hurd)
- 1990: Linus Torvalds, Finland
- 1991: First edition, Linux 0.01
- 1994: Linux 1.0, faster and more stable than Unix
- Distributions: Ubuntu, CentOS, Debian, Fedora, openSUSE...



	Linux	Windows
Desktop	~3%	~86%
Web Server	~38%	~32%
Supercomputer	100%	-

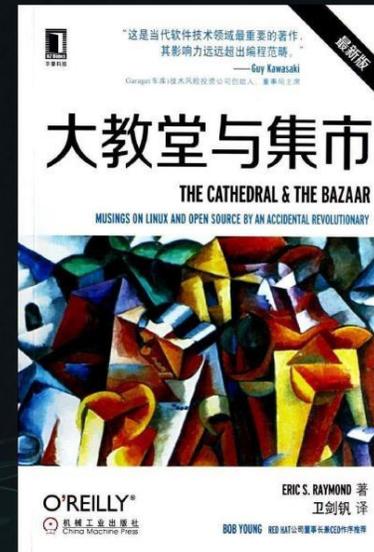
革命：自由软件运动

Open Source Software & Free Software

- 1997: The Cathedral and the Bazaar - Eric Raymond
- 1998: Open Source Initiative, OSI
- 1998: Bruce Perens, Open Source Definition
- 1998: Netscape open source Netscape Communicator
Mozilla & Firefox
- 1999: Acquisition of Netscape by AOL

Comparison

Free Software	Open Source
Philosophy Hacker culture Idealism	Pragmatic Moral neutral Business friendly



发展：开源的成功典范

Linux

- 1991: First edition Linux 0.01, release every week, accept feedback and contribution
- 1992: Linux 0.96, 40,000 LOC, 1,000 users
- 1993: Linux 0.99, 100,000 LOC, 20,000 users
- 1995: Linux 1.2, 250,000 LOC, 500,000 users
- 1997: Linux 2.1, 800,000 LOC, 3.5M users
- 1998: Linux 2.11, 1.5M LOC, 7.5M users
- 2020: Linux 5.7, ~28M LOC

Apache HTTP Server

- 1995: 'a patchy' to NCSA
- 1999: 57% web service
- 2005: 70% web service
- 2020: ~37% web service
- Linux + Apache vs. Windows NT + IIS

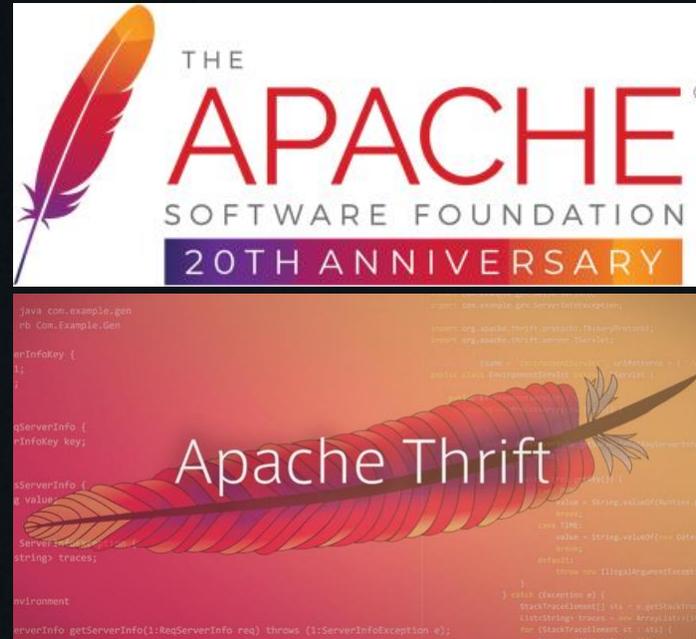
Revolution OS(2001)



发展：开源的成功典范

The Apache Software Foundation (ASF)

- 1999
- 350+ projects, 8000+ committers
- 813 individual members
- All volunteer community
- Apache way: Community over code



进步：开源的繁荣

The Linux Foundation

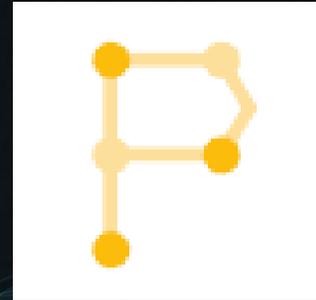
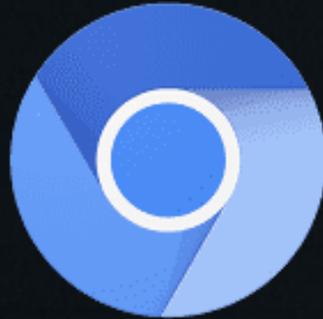
- 2000
- 1000+ members
- 100+ projects
- CNCF, FLAI, Hyperledger, RISC-V
- KubeCon



进步：开源的繁荣

Google

- Very open to open source
- 2000 + open source projects



进步：开源的繁荣

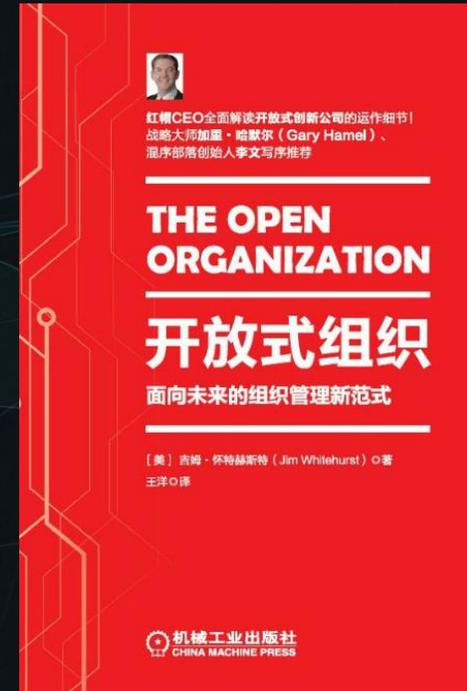
RedHat

- 1995, Bob Young
- 1999, Go public on Nasdaq
- Red Hat Enterprise Linux
- CentOS
- The Open Organization - Jim Whitehurst, CEO



PRESS RELEASE

IBM Closes Landmark Acquisition of Red Hat for \$34 Billion; Defines Open, Hybrid Cloud Future



进步：开源的繁荣



2018年3月
收购：50亿美金



2018年7月
收购：25亿美金



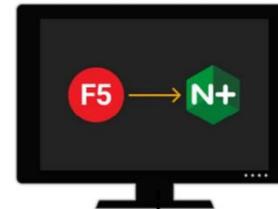
2018年8月
好莱坞 + 基金会



2018年10月
收购：340亿美金



2019年3月
收购：6.7亿美金



2018年1月
收购：2.5亿美金



2018年6月
收购：75亿美金



2018年10月
上市：50亿美金



2018年10月
合并：52亿美金

机遇：国内的开源

China was open source consumer before 2015

- **Apache Kylin**: 2014.11, ebay Shanghai research center.
- **JStorm**: 2015.11 JStorm to Apache Storm and Alibaba join ASF
- ASF: 19 projects from China, 12 graduated projects and 7 incubating projects
 - Graduated: CarbonData, Dubbo, Eagle, Griffin, HAWQ, Kylin, RocketMQ, ServiceComb, ShardingSphere, Skywalking, APISIX, IoTDB.
 - Incubating: brpc, DolphinScheduler, Doris, eCharts, MesaTEE, TubeMQ, Weex.
- The Linux Foundation: Harbor, TiKV, Tars, Milvus, etc.

机遇：国内的开源

- **Alibaba**: Apache Dubbo, Apache RocketMQ, Apache Weex, Ant-design, Arthas, Nacos
- **Baidu**: Apache brpc, Apache echarts, Apache Doris, Apache MesaTEE, PaddlePaddle
- **Tencent**: Tars, WeUI, Angel, Mars, WePY
- **Huawei**: Apache ServiceComb, Apache CarbonData, Harmony OS, MindSpore, openEuler
- **PingCAP**: TiDB, TiKV



GitHub 2020 中国项目活跃度 Top 20

#	项目名	活跃度	参与开发者数量	issue comment	open issue	open pull	pull review comment	merge pull	star	fork
1	pingcap/tidb	210.1	5,831	53,022	2,801	4,969	10,928	3,459.2	4,862	1,052
2	ant-design/ant-design	193.3	23,620	32,026	4,836	3,131	3,320	2,130.7	12,709	8,052
3	PaddlePaddle/Paddle	127.4	4,842	15,329	2,256	5,656	9,625	3,478.2	3,574	786
4	tikv/tikv	81.7	2,593	17,817	997	2,019	5,547	1,279.9	2,129	434
5	apache/shardingsphere	75.3	5,267	9,055	1,713	3,235	1,858	2,539.5	3,834	1,443
6	apache/incubator-tvm	70.4	2,148	7,961	437	2,112	8,506	1,540.1	1,454	662
7	pingcap/docs-cn	65.1	532	8,202	96	2,965	6,959	2,315.9	140	320
8	apache/incubator-echarts	64.2	11,638	7,650	1,620	324	346	194.5	6,664	4,463
9	pingcap/pd	60.9	437	13,325	667	1,667	4,972	1,297.7	214	224
10	alibaba/nacos	59.9	9,956	7,042	1,640	706	827	410.0	6,347	3,450
11	NervJS/taro	54.7	7,469	9,339	2,231	917	135	551.5	5,250	1,012
12	youzan/vant	54.2	9,806	4,897	1,661	715	201	554.4	4,672	4,502
13	pingcap/docs	53.9	314	7,014	64	2,736	5,226	2,257.8	90	164
14	ElemeFE/element	52.7	11,749	4,993	1,762	297	10	33.3	6,853	3,411
15	apache/skywalking	51.9	5,556	6,783	1,084	860	3,455	583.4	4,201	1,471
16	PaddlePaddle/PaddleOCR	47.9	9,394	4,039	1,033	573	622	420.0	8,430	1,664
17	apache/incubator-dolphinscheduler	47.1	2,588	9,364	1,269	1,407	730	902.7	1,835	909
18	apache/apisix	45.4	2,923	5,855	1,109	1,029	3,383	715.0	2,496	579
19	seata/seata	45.1	7,339	3,754	785	517	1,805	313.5	5,261	2,296
20	pingcap/tidb-operator	45.1	425	8,627	703	1,498	3,683	1,172.1	240	140

GitHub 2020 中国企业开源项目活跃一览

#	公司	活跃度	项目数量	issue comment	open issue	open pull	pull review comment	merge pull	star	fork
1	Alibaba	1,571.1	1,496	130,558	33,947	29,097	22,615	17,471.6	216,980	68,864
2	PingCAP	778.4	151	139,255	8,138	25,401	61,538	18,880.4	18,008	5,058
3	Baidu	671.2	540	55,265	12,592	20,720	23,380	13,475.9	70,960	22,148
4	Tencent	432.3	388	21,446	8,599	10,264	2,870	7,088.8	69,198	19,348
5	JD	153.0	74	20,126	4,504	4,483	2,043	3,214.2	13,119	3,316
6	Huawei	101.8	200	10,322	1,709	2,930	3,867	2,005.2	8,758	3,168
7	DiDi	89.4	63	3,114	1,290	827	207	508.7	20,489	3,907
8	Youzan	88.6	58	7,259	2,760	1,409	634	1,068.1	9,509	5,820
9	Bytedance	59.2	85	1,973	645	785	659	514.9	14,034	1,671
10	WeBank	57.9	59	2,197	718	3,501	596	2,411.9	5,225	1,902
11	Xiaomi	50.4	98	1,767	1,604	1,007	3,001	691.6	5,823	1,760
12	Meituan	46.9	68	1,356	564	305	17	147.0	10,879	2,573
13	Bilibili	42.7	51	1,306	446	132	52	66.4	10,278	2,295
14	360	39.8	147	1,769	810	441	40	231.7	8,105	1,914
15	Juejin	39.5	26	3,866	578	661	3,624	546.9	4,208	810
16	CTrip	36.9	25	2,346	537	216	276	130.5	6,196	2,562
17	Linux China	34.1	16	226	10	3,862	11	3,123.3	482	302
18	Netease	25.0	119	1,603	777	313	32	149.6	3,880	1,445
19	Deepin	18.6	267	2,555	931	326	21	132.7	1,339	821
20	Qunar	7.1	43	113	54	56	10	9.9	1,653	478
21	Vipshop	7.1	14	112	127	66	0	14.0	1,604	421
22	Douban	3.7	41	98	43	158	58	128.6	508	99

来源：GitHub 2020 数字洞察报告, 2021-02.

College students turned OSS leaders



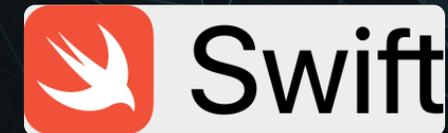
Sage Weil (University of California, Santa Cruz)
Know for: Ceph
Now: Red Hat



Matei Zaharia (UC Berkeley)
Know for: Apache Spark and Apache Mesos
Now: Stanford



Chris Lattner (UIUC)
Know for: LLVM, Clang, Swift
Now: Google Brain



国内代表人物



李沐 (CMU)
Know for: MXNet
Now: Amazon



陈天奇 (University of Washington)
Know for: XGBoost
Now: CMU



贾扬清 (UC Berkeley)
Know for: Caffe
Now: Alibaba



What is open source ?

The Open Source Definition (OSD) is defined by **Open Source Initiative (OSI)**

Distribution terms must comply with the following criteria:

1. Free redistribution
2. Source code
3. Derived work
4. Integrity of the author's source code
5. No discrimination against persons or groups
6. No discrimination against fields of endeavor
7. Distribution of license
8. License must not be specific to a product
9. License must not restrict other software
10. License must be technology-neutral

Derived from the Debian Free Software Guidelines (DFSG) by Bruce Perens.



开源究竟是什么？

- **开源软件技术**：开源大数据技术、开源人工智能技术
- **开源软件产品**：Hadoop、Docker、Tensorflow
- **开源软件工程**：软件开发生命周期、Git、开源协作 workflow

开源软件开发

- **开源社区型组织**：开源社区、开源共同体、社区运营与治理
- **开源生态与战略**：开源供应链、开源商业模式、开源法规
- **开放式创新方法**：科学方法与文化、用户创新、大众创新

开放组织管理

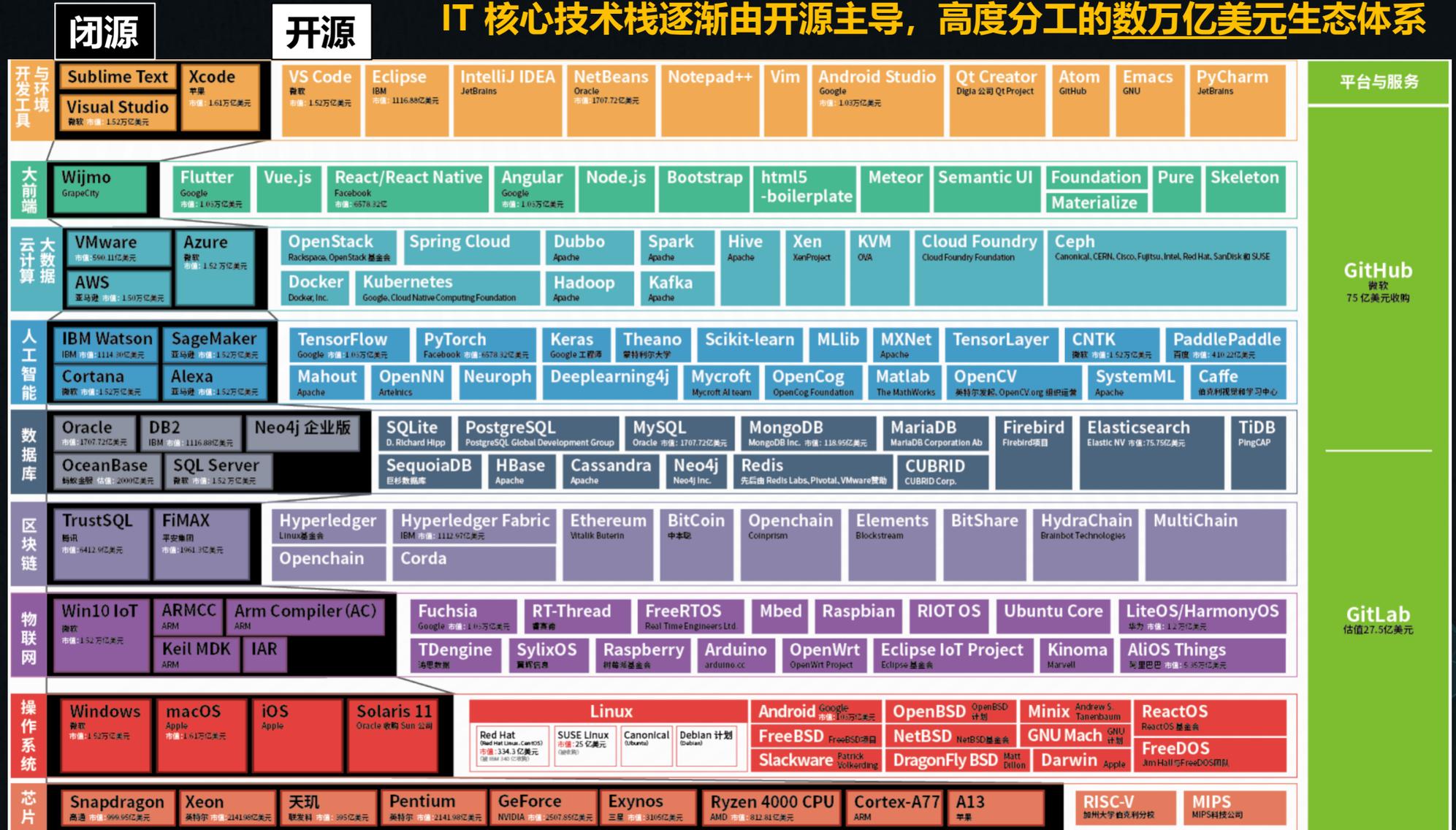
1、开源软件技术

IT 核心技术栈逐渐由开源主导，高度分工的数万亿美元生态体系

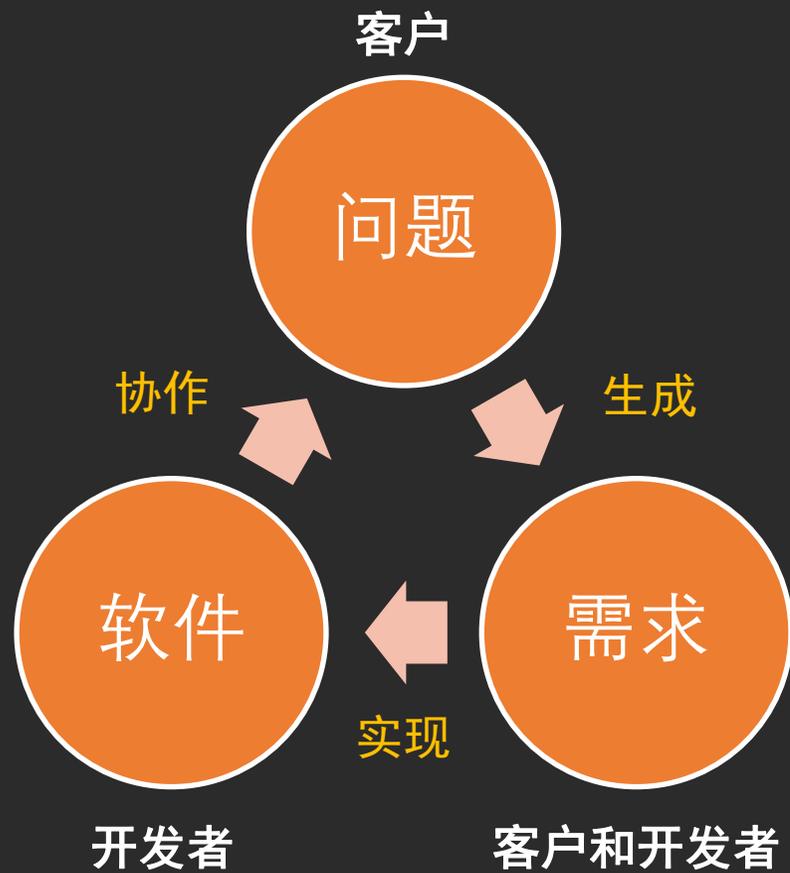
据统计，拥有1000行以上代码的工程中，**99%**的软件项目至少有一个开源依赖项。

平均一个软件项目工程有**150**多个开源依赖项。

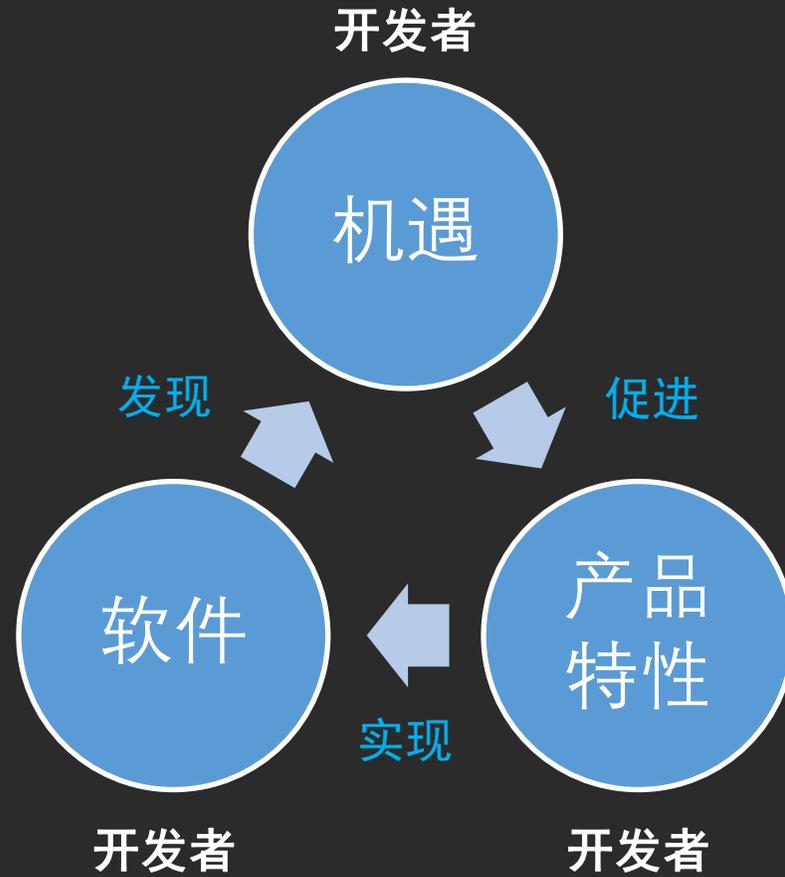
现代软件供应链中，**80%~90%**的代码来自开源。



软件项目与软件产品的区别

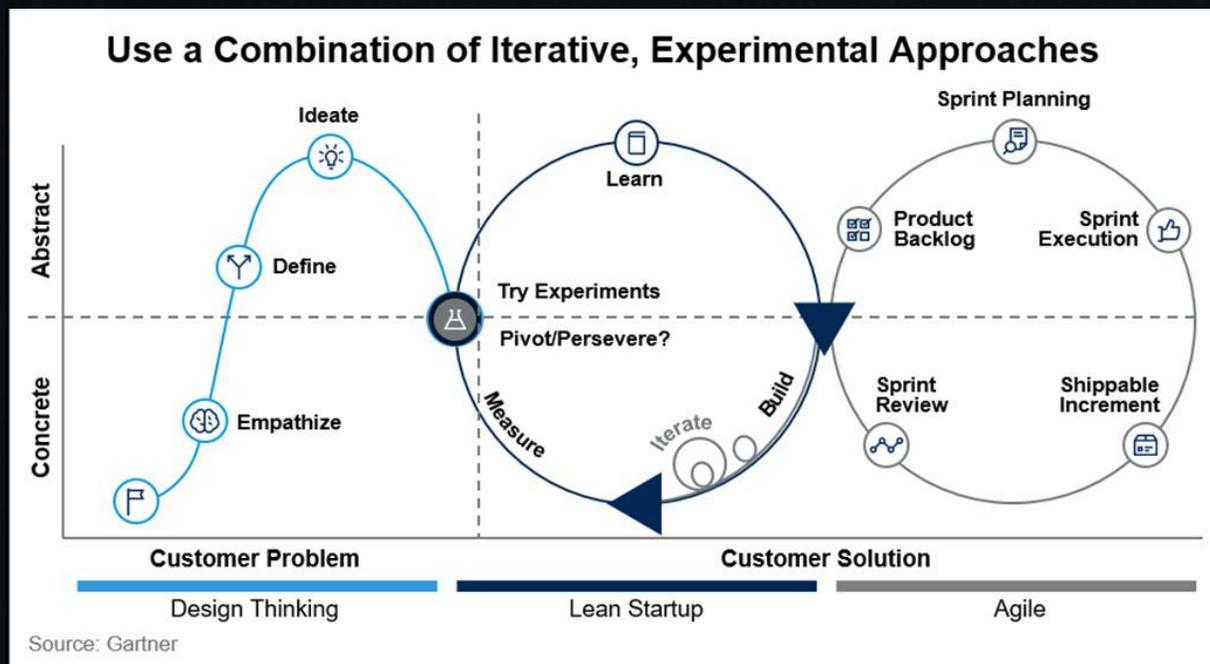


基于项目的软件工程



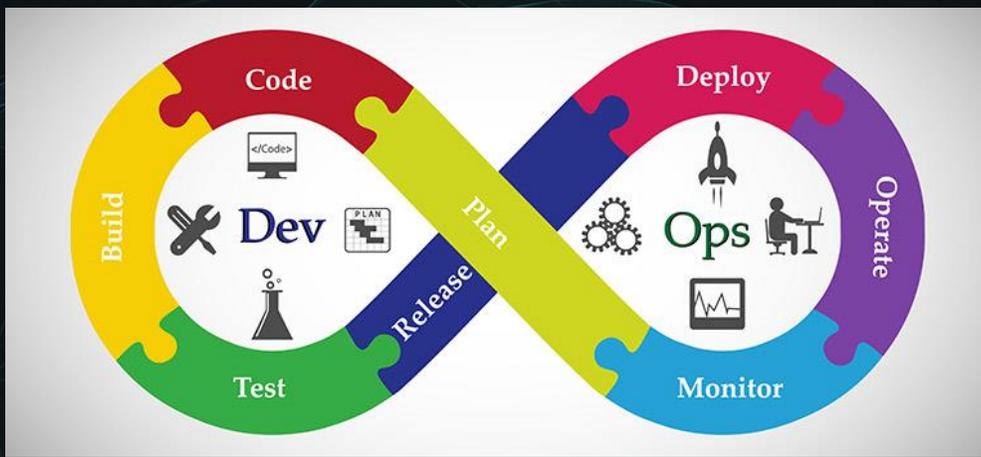
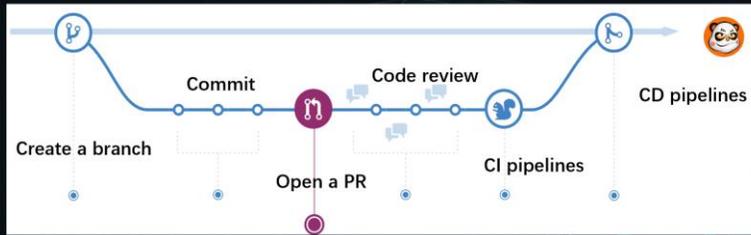
基于产品的软件工程

2、开源软件工程



- Design thinking
- X-Ops
 - DevOps / DataOps / ResOps
- 从代码化到自动化
- 从数据化到智能化

- 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)

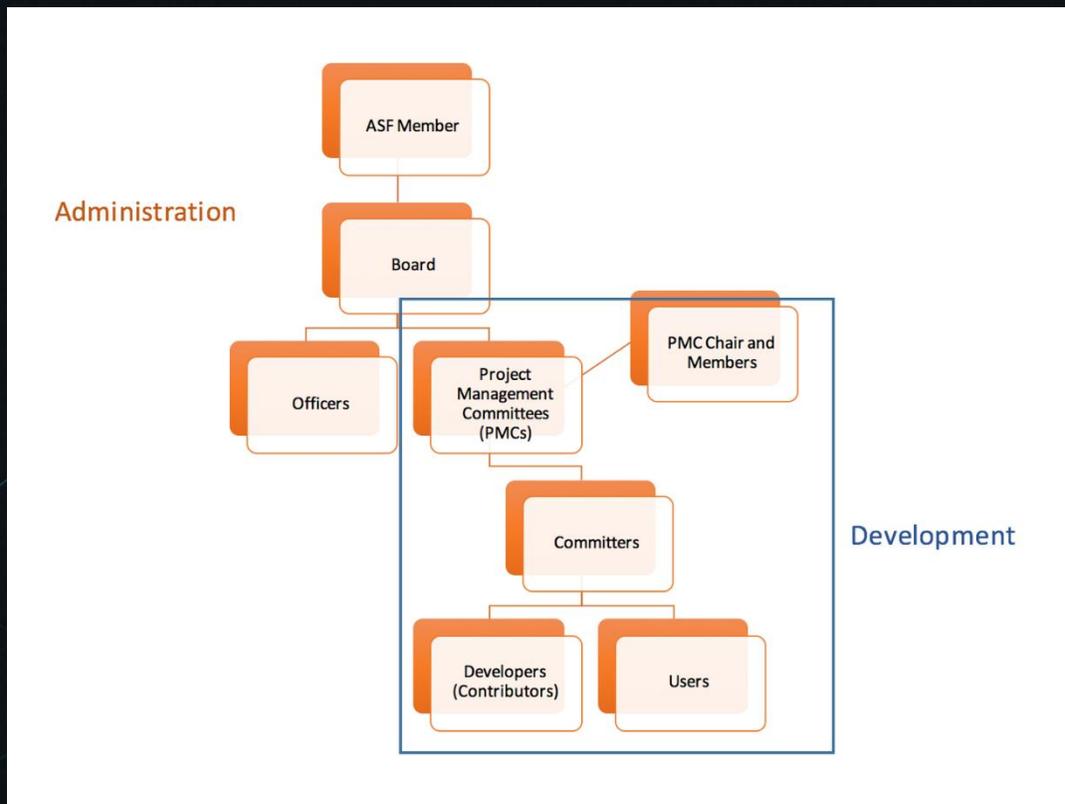
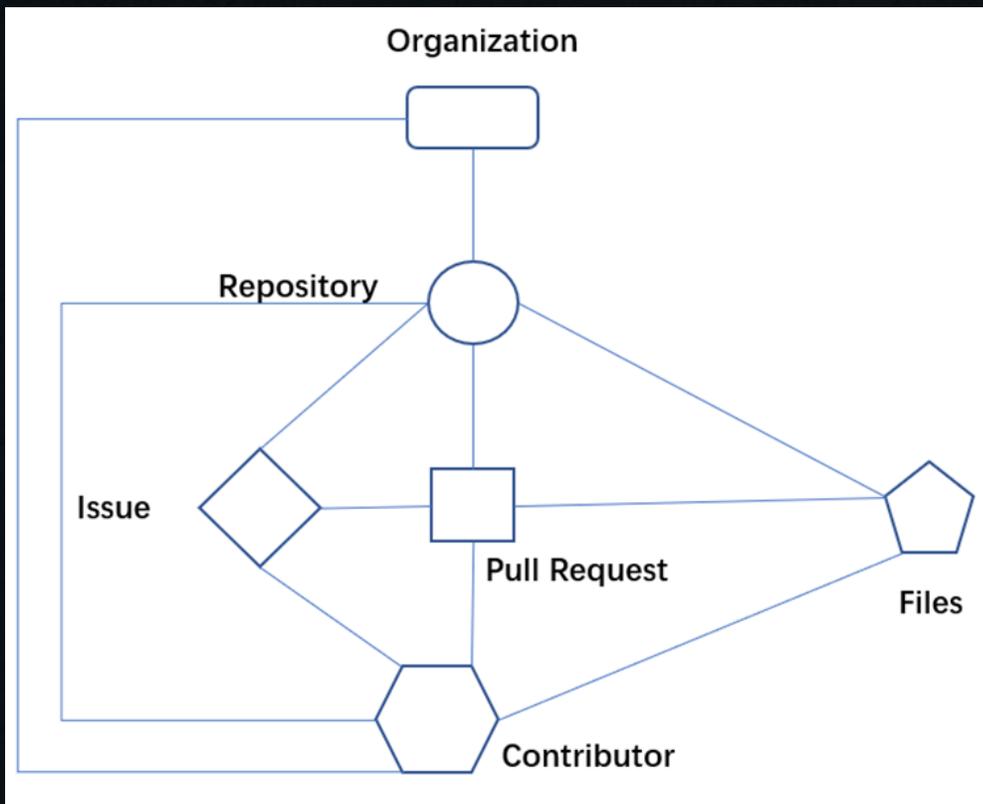


Open source workflow

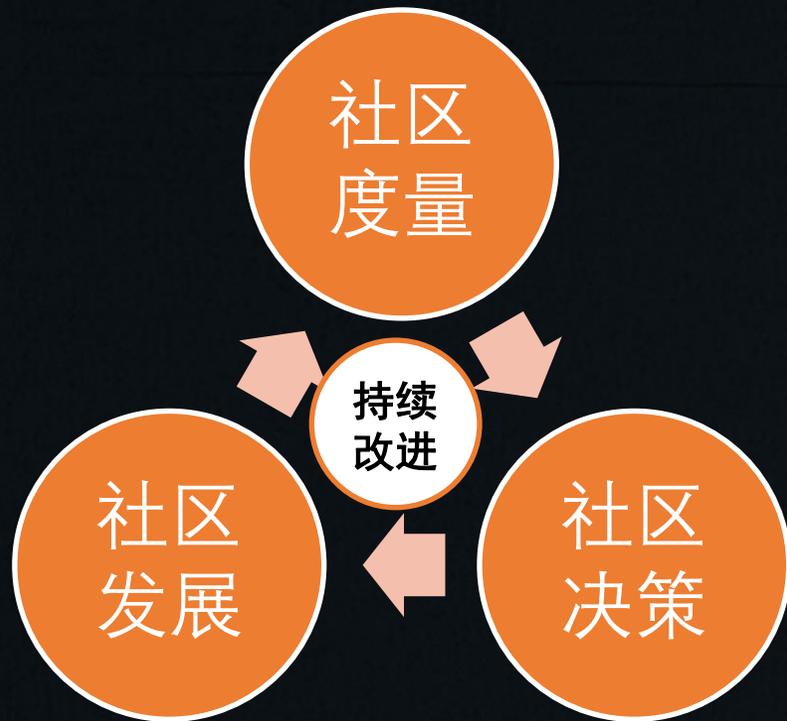


4、开源社区型组织

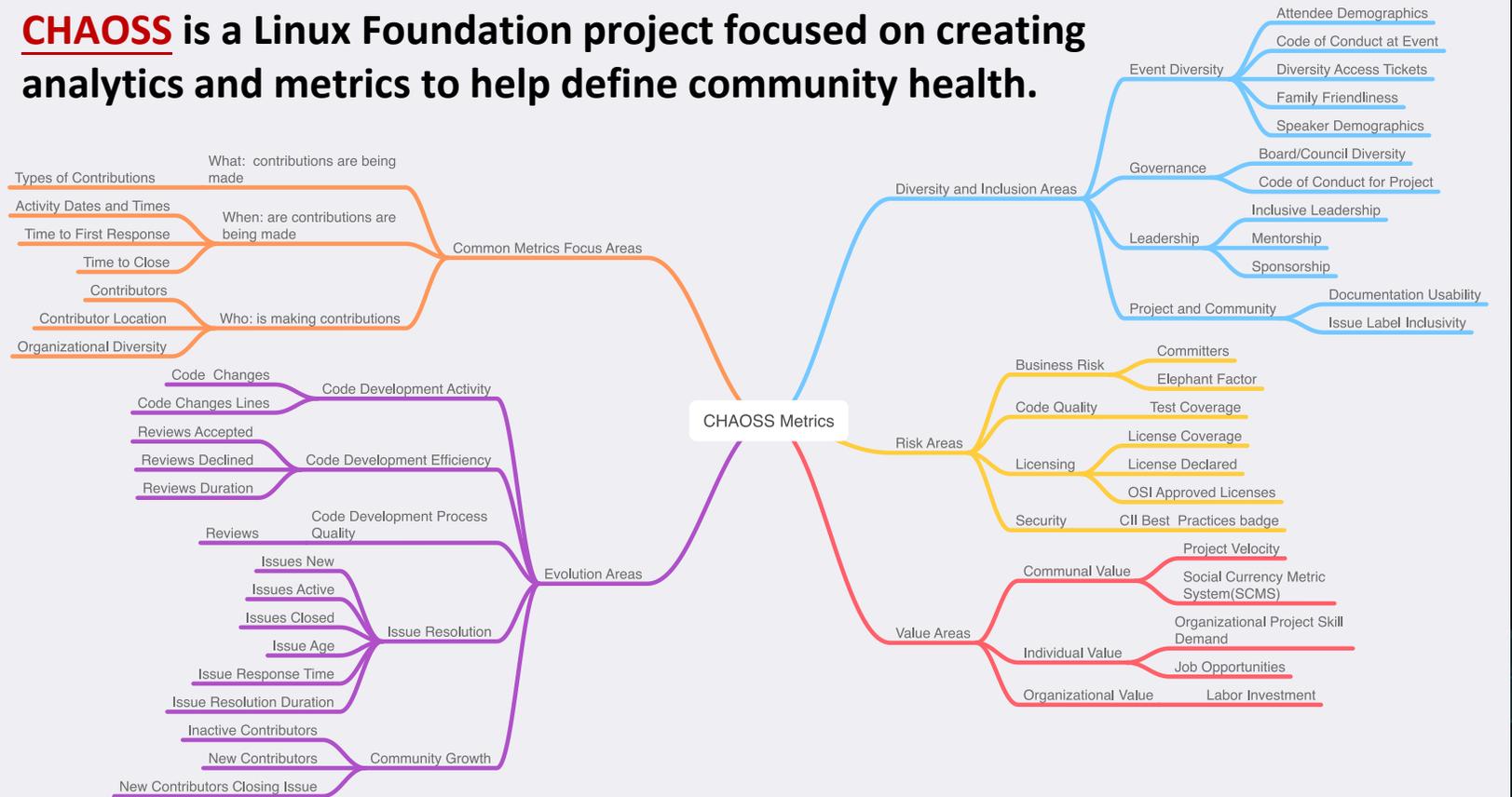
- 1999年非盈利组织Apache软件基金会（Apache Software Foundation，ASF）正式成立，为开源项目提供组织、法务上的支持。
- [社区胜于代码（Community Over Code）](#) 是ASF的核心原则。



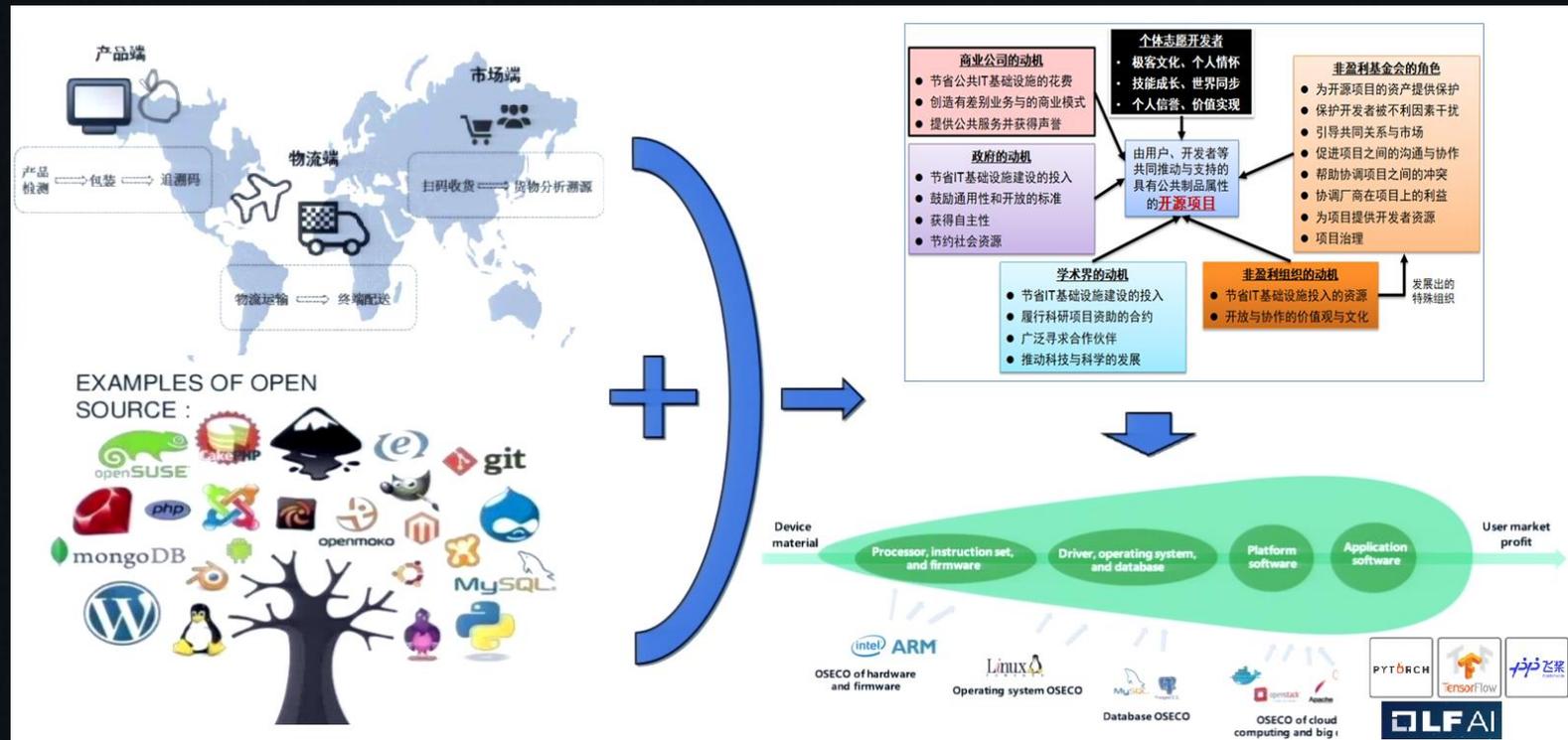
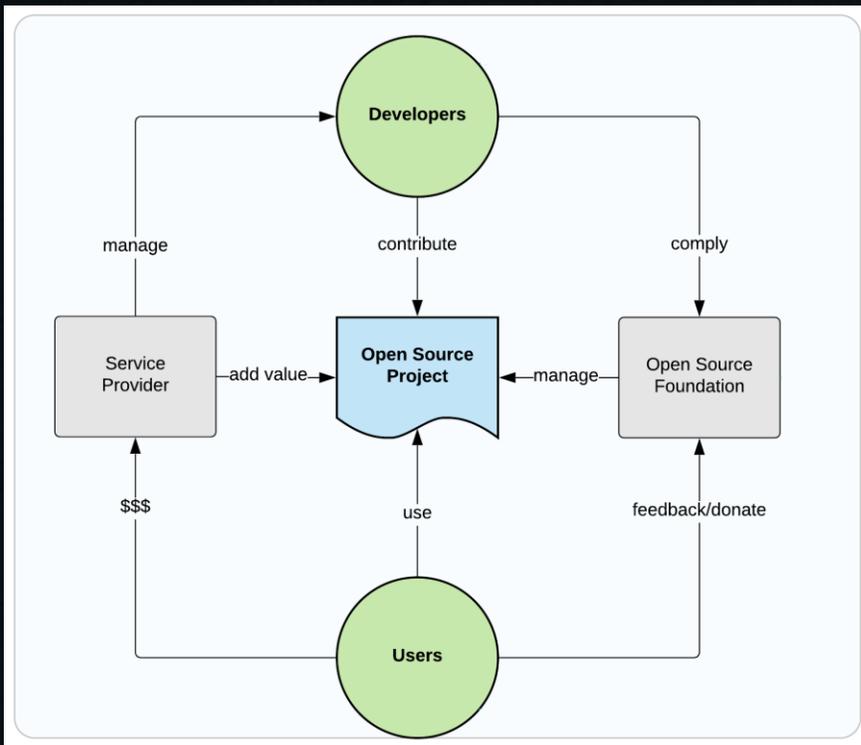
社区度量与社区数字化运营



CHAOSS is a Linux Foundation project focused on creating analytics and metrics to help define community health.

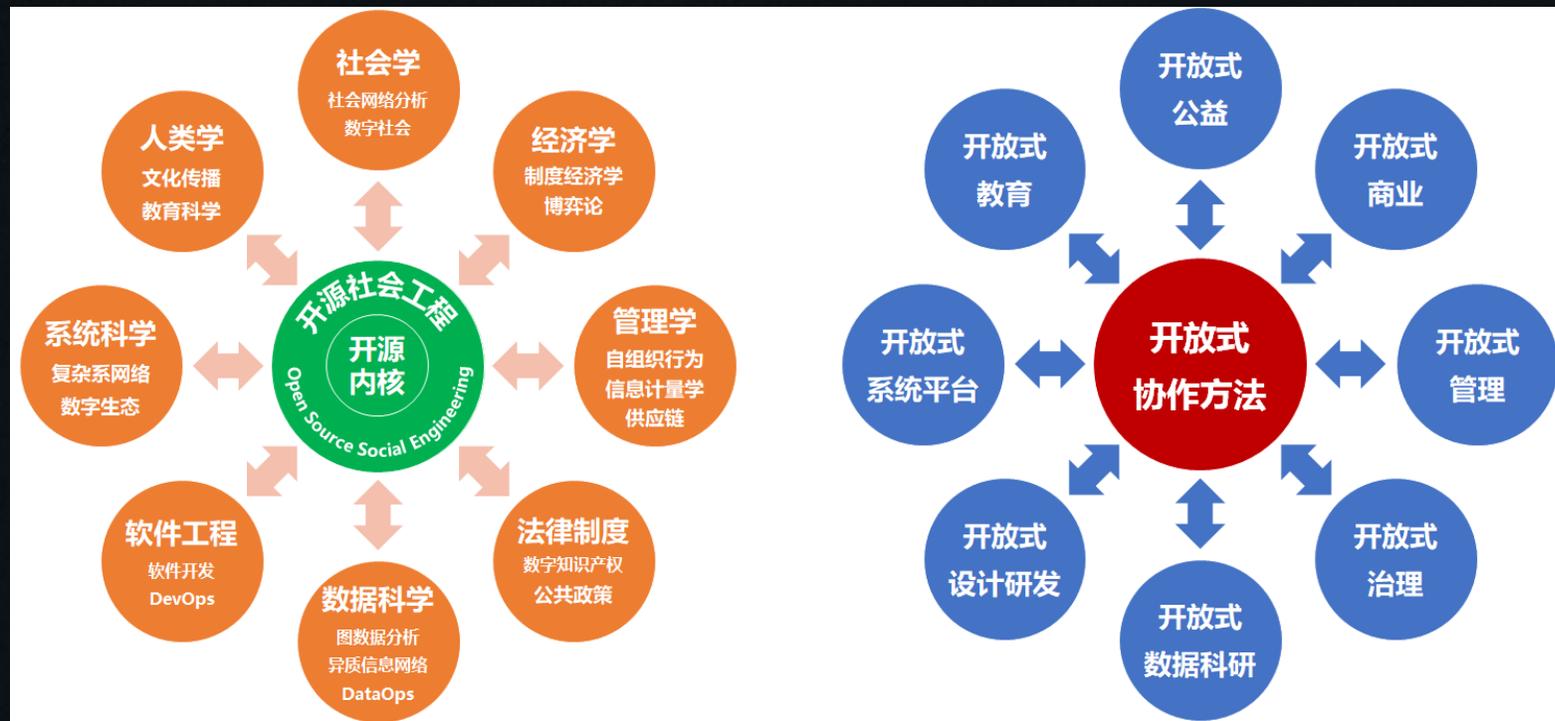
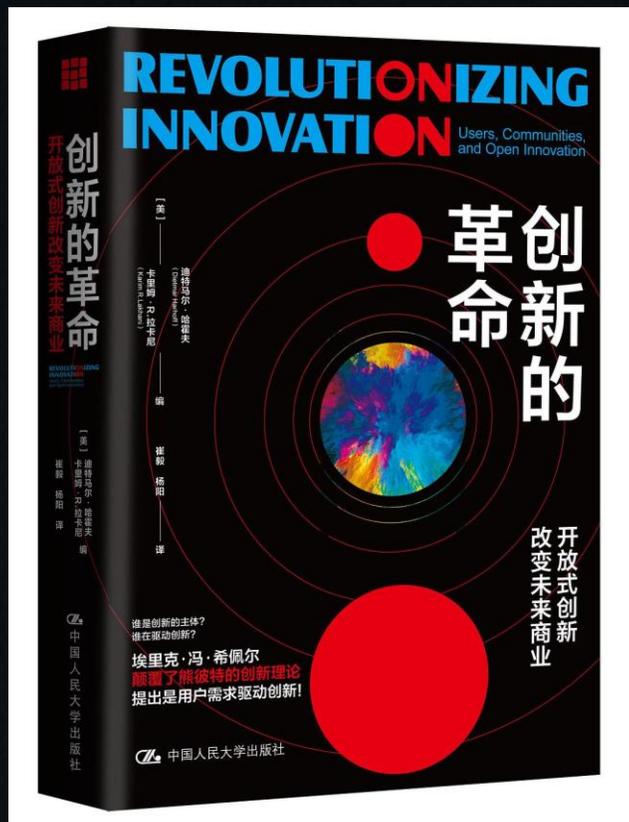


5、开源生态与战略



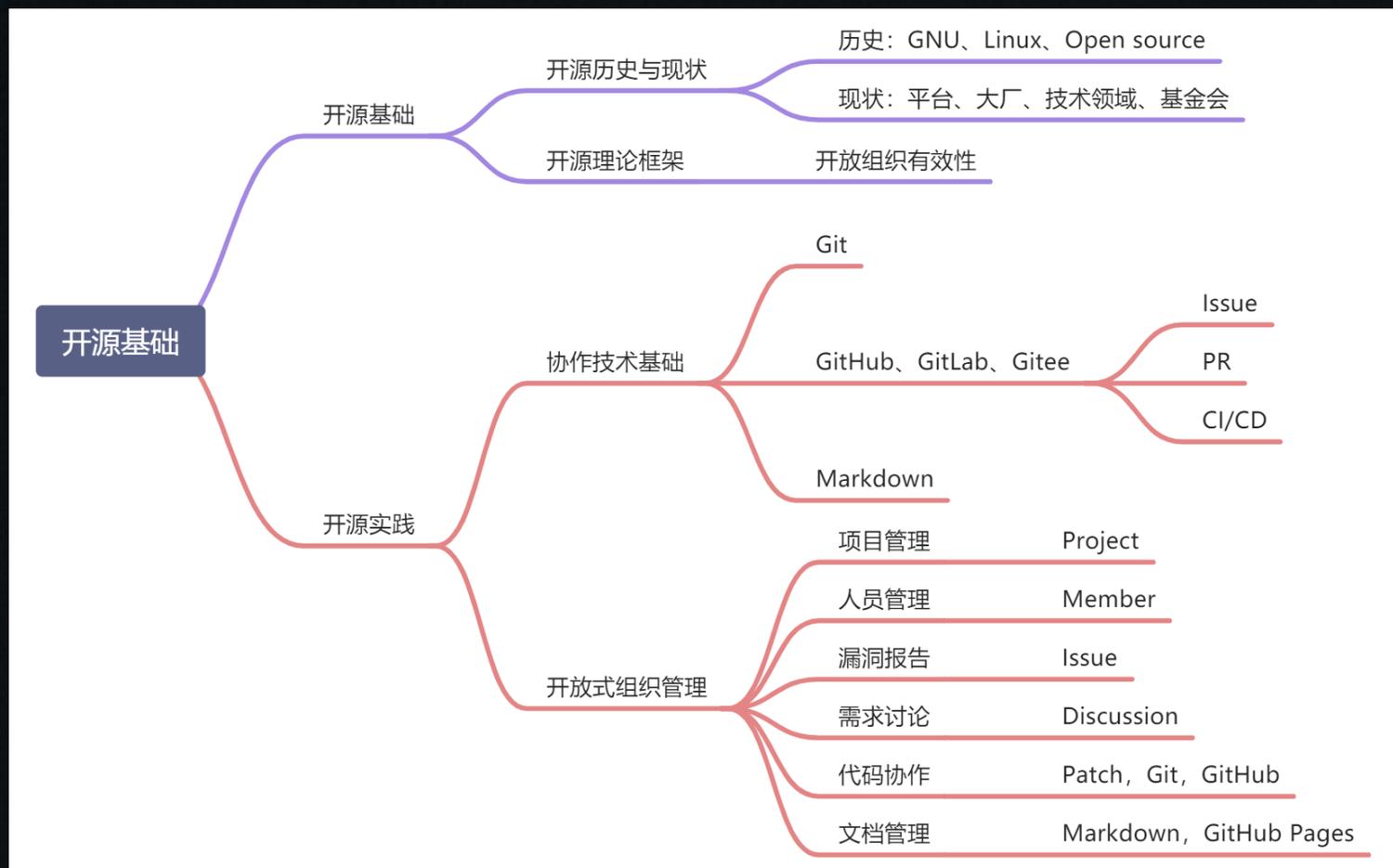
开源软件已经成为构建全球软件产业的核心，围绕开源软件所形成的**全球开源供应链**与**开源软件生态**（开发者、项目、企业、基金会、学术组织等）直接决定了一个国家或地区在数字化时代的核心竞争态势。

6、开放式创新方法

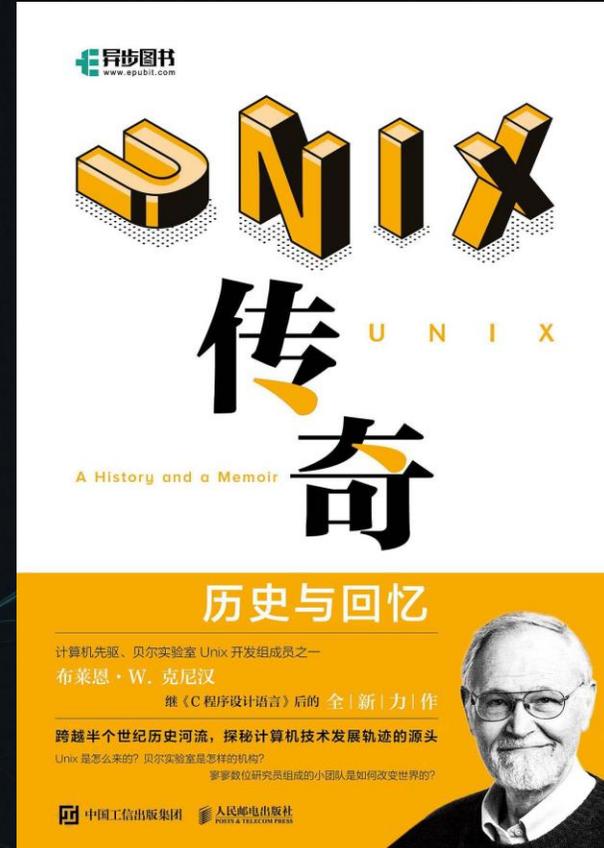
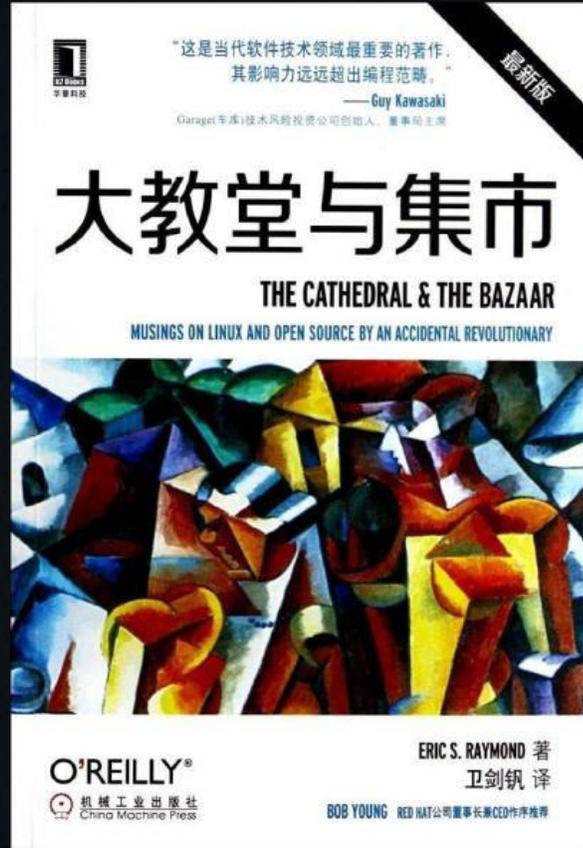


课程目标与内容

- 对开源有全面理性的认识，并掌握参与开源项目的基本技术能力。



延伸阅读



Q & A

