

# 2020 中国开源年度报告



## 前言

撰写这篇前言，我们的心情一直很复杂，很难用某种词汇来概括。在心里升起的一句话是：“这是最好的时代，这是最坏的时代”。

也许多年以后，我们回看 2020 年，会将其称之为：世界剧变之始。很多朋友在去年，甚至“每一天都在亲历历史，每一天都是活久见！”新冠疫情与中美贸易战，在同一年加剧。而因为疫情和贸易战，世界正在加速发生变化。

在这样的历史大背景下，我们在开源界，也发现了以下三个趋势：

### 1. 开源大发展以及由实向虚进发的趋势

从各种数据，以及我们自己的感觉都能发现：全球开源都出现一个大发展的趋势。GitHub 的活跃代码仓库与活跃用户数在高速增长（35.3% / 21.2%）；Gitee 的代码仓库与用户数在以更加惊人的速度增长（192% / 162%）。

一方面，是开源这么多年一直持续的上升势头。而另一方面，则是我们的一个猜测：疫情以来，越来越多的人开始远程办公，事实上促进了更多的人有机会“斜杠化”，也就是在一台电脑面前方便地切换多种身份，以异步的方式处理多种事务，从而增加了开发者参与开源的时间和机会。

当然，同样由于远程办公，虚拟世界在人类生活中的占比，变得更大了。这样是否更好，还会引发哪些问题？作为站在隧道口的人类，其实是猜不透的。

## 2. 中国开源崛起以及开源世界分裂的趋势

随着越来越多中国开源项目的影响力不断增加、木兰协议的推出、开放原子基金会的成立，Gitee 的超高速增长，CODE China 的新发布，我们已经可以确信，2020 年是中国开源的崛起之年。接下来阅读报告的朋友们，应该会看到很多确切的证据。

但是，另一个值得注意的现象，也可以从下面的数据里看出来。在 GitHub 上最活跃的中国开源项目，与 Gitee 上最活跃的开源项目之间，完全没有重叠。

随着 Gitee 的继续高速发展，可以大胆预测，今后会有越来越多的，高质量的中国开源项目，选择在 Gitee 上开源。曾经有一位朋友提到的“One World, Two Systems”，将会逐渐变成现实。

如果中国的开源，以一种与世界隔离的方式“崛起”，这并不是我们希望看到的未来。

## 3. 开源向善以及我们尚未做好准备

疫情爆发以来，众多的开源人奉献自己的时间、精力、技术和智慧，投入到一个一个的抗疫相关的开源项目之中，也因此诞生了诸多面向医学、防疫、公益、救援的开源组织和开源项目。其中 wuhan2020 就是典型的代表。

因此，在 2020 年的中国开源年会筹办期间，我们毫不犹豫的选择了“开源向善（Open Source for Good）”作为年会的主题，也获得了众多讲师、参会者与赞助厂商的一致认同。

但是，接下来应该怎么发展呢？为了抗击疫情而凝聚起来的人，创造的项目，积累的经验，总结的教训，应该如何“整理、封装、模块化”，并为了今后的各种意外，做好准备呢？这些都是值得长期思考并改进的领域。

作为一份力图完整、客观、全面、丰富的报告，值得探讨的问题当然不止上述这些，也欢迎朋友们阅读下面的报告，并随时与我们交流。

庄表伟，开源社理事长  
2021 年 1 月 16 日

---

### 特邀点评专家：

- 吴晟，Apache 软件基金会 Member，Apache Local Community 联合创始人
- 黄东旭，PingCAP 联合创始人兼 CTO
- 马艳军，百度深度学习技术平台部高级总监
- 蒋涛，CSDN 创始人 & 董事长，极客邦创投创始合伙人

- 高阳, SegmentFault 思否创始人 & CEO
  - 红薯, 开源中国创始人 & CTO
  - 庄表伟, 开源社理事长
  - 陈阳, 开源社副理事长
  - 王伟, 开源社执行长
  - 刘天栋, 开源社理事
- 

## 一、问卷篇

### 1、报告背景

2016 年初, 开源社发布了《2015 年中国开源社区参会调查报告》, 随后的几年中, 持续发布了开发者调查报告, 旨在从多种维度呈现国内的开源发展情况。今年我们再次启程, 结合数据分析手段和调查报告等多种形式, 绘制一份 2020 年中国开源世界的地图。

这份问卷是每年中国开源年报的重要一环, 不基于调研的分析报告不过是纸上谈兵。在往年的基础上, 我们参考了其他现存的主流开发者问卷内容, 并加入了一些新的视角。基于 2020 年 COVID-19 这个不容忽视的大背景, 这份调查问卷进一步将视角发散到开源世界的线上协作, 以及开源能够为世界带来什么等主题。

通过近 60 项左右问题的统计调查与分析, 我们希望能够还原出当前中国开源社区的真实现状, 从而为开源的后来人提供权威的参考。

- 调查对象: 覆盖开发者、社区成员、贡献者、学生、政府企业管理人员
- 调查内容: 主要涵盖个人信息、工作状况、开源社区以及开发者技术
- 调查方法: 以在线问卷方式搜集样本和数据, 交叉对比法分析数据
- 推广方法: 线上社交媒体、博客、开源社、开源中国网站
- 问题数量: 59
- 问题类型: 单选、多选、开放性
- 样本量: 236

### 2、重要发现

由于疫情, 今年的问卷全部通过线上的方式传播, 因此收集到的样本量相比往年较少, 但仍旧能够管中窥豹, 观察到中国开源世界的悄然变化。通过分析 2020 年的统计数据, 并对比往年数据和其它一些公开发布的统计报告, 我们有如下一些重要发现:

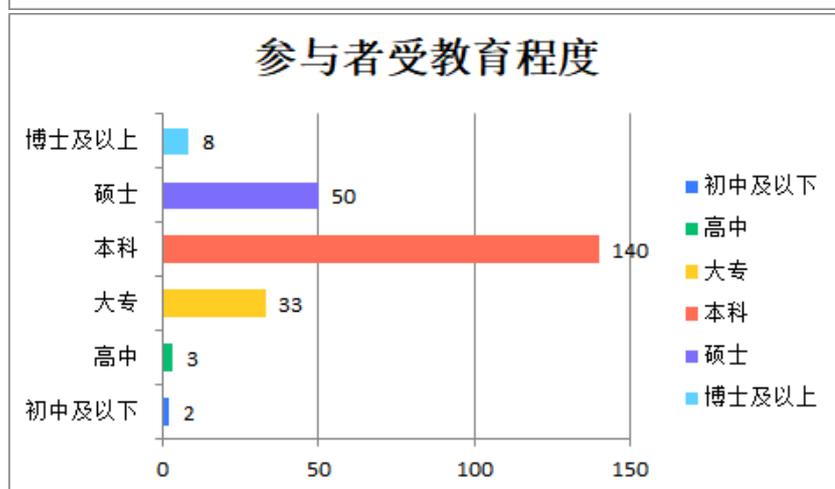
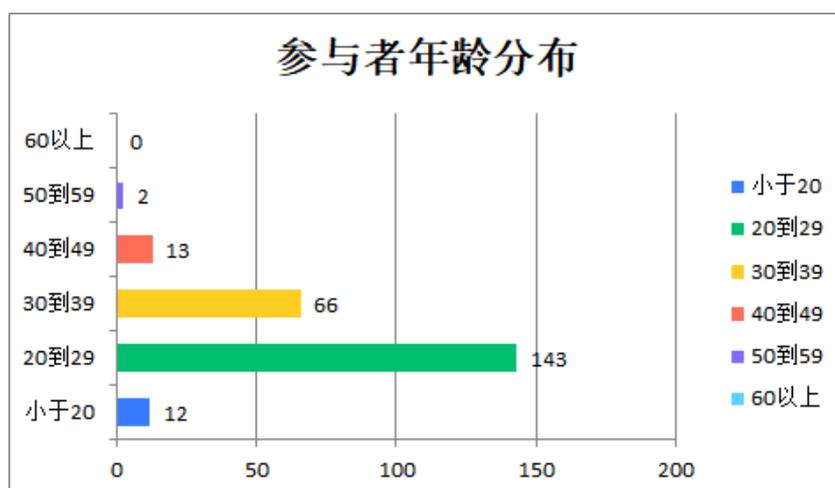
- 开源参与者的主力依旧是 20-39 岁人群, 相比去年比例仍在上升, 而在开源社区的年龄分布中, 20-30 岁人群占据了 64%, 说明开发者的年龄仍在趋于年轻化;
- 相较于 2019 年, 今年参与调研的开发者中, 还未工作的人群占了绝大部分, 从参与者从事领域和职位分布也可以看出, 这些“还未工作”人群多数指向学生群体, 这和当前社会普遍追求更高学历深造有关, 和本次问卷传播形式有关, 但是同时也能说明开源在教育领域取得了不错的进展;
- 开发者所从事的技术方向仍以前后端为主, 其中后端开发占比约 31%, 其次是 Web 前端、AI 以及大数据分析方向, 占比分别为 10%、9%、9%, 非技术人员以及其他行业的从业者占比均为 7%, 说明开源已经越来越受到社会各界的关注;

- 参与开源的形式由 2019 年以代码和测试为主转变为以代码和文档为主，社区和项目正在意识到文档的重要性，更多开源贡献者投入到了文档撰写中；
- 开源活动的参与频率相较往年有所上升，这与国内愈发繁荣的开源氛围和逐渐丰富的开源活动也有关系。81% 的开发者认为开源活动对促进和推动开源社区至关重要。而对于更倾向于线上还是线下的会议，结果竟惊人地持平；
- 在 2020 年 COVID-19 这个大背景下，各行各业都不可避免受到了冲击和影响；82% 的人认为疫情对于参加开源社区的活动产生了影响，但也具有推动作用，11% 的人认为毫无影响，7% 的人认为疫情的负面影响大于正面影响；
- 在在线文档协作工具的使用上，石墨文档以黑马的形式由第三名一跃成为第一名，压倒了 wiki 和 Google Doc，国内正在涌现出一批优秀的协作工具，并越来越为大家认可；
- 2020 年 9 月 9 日，中国首个开源软件基金会「开放原子开源基金会（OpenAtom Foundation）」官方正式对外发声，大家对此抱有多重期望：推广开源的理念，开源文化教育和社区建设，建立开源生态体系，对开源项目进行早期的孵化和扶持，帮助中国开源社区进行资源的整合和优化，并接轨国际开源社区。

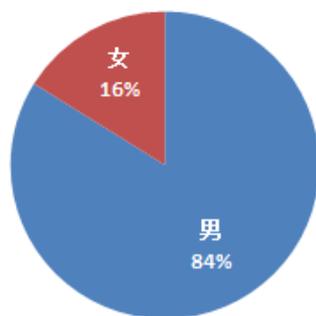
### 3、开发者群体特征

#### 3.1 参与者普遍年轻且学历较高，男性占比八成

参与者的年龄集中在 20-39 岁，受教育程度普遍在本科及以上，其中男性占比约为 84%，女性为 16%，与去年持平。



## 参与者性别分布



### 【专家点评】

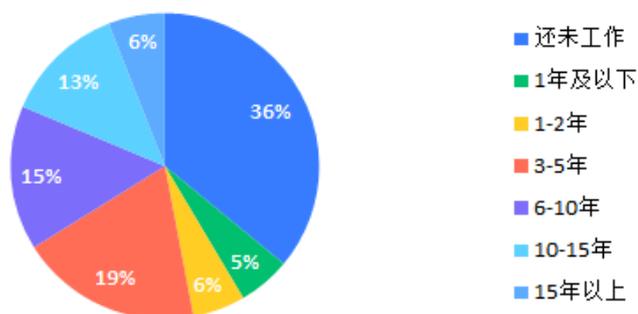
**刘天栋**：女性参与开源的比例这和过去2~3年的调查数据相似，但是和国际平均数据（约10~12%）相比高。期盼未来国内能有更多女性参与开源，形成一道美丽的开源风景线。

**陈阳**：多元化一直是开源社区的口号，全球开源社区早在十几年前就开始发起了“鼓励女性参与开源”的号召；开源社在每一年的中国开源年会 (COSCon) 上，都有举办女性论坛的传统。2020 年的 COSCon 上，我们邀请到了开源社区中的女性领袖，以她们独特能力和视角将整个社区链接起来。GitHub COO Erica Brescia，管理着超过 5 千万程序员的社区；Stormy Peters，从 GNOME 基金会执行总监，到现在负责着微软的开源办公室；来自亚洲越南的 Hong Phuc Dang，从创办 FOSSASIA 开始参与开源，目前出任着 OSI 的董事会 VP。女性参与开源的广度和深度，正在按照我们所期望的方向发展，让这个世界变得有趣和多元化。

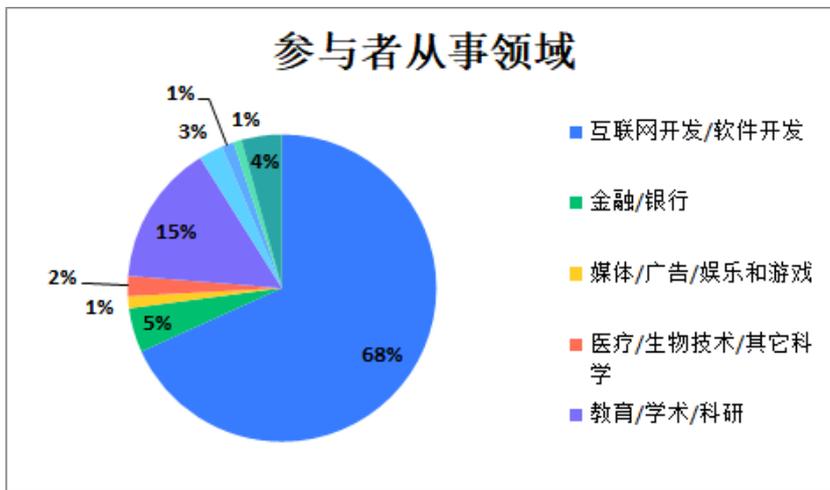
## 3.2 从业时间及行业领域

参与者中还未参加工作的人数最多，占比约 36%，其次则是从业时间为 3-5 年、6-10 年的参与者占比分别为 19%、15%，从业时间在 10 年以上的约 2 成。

## 参与者从业时间

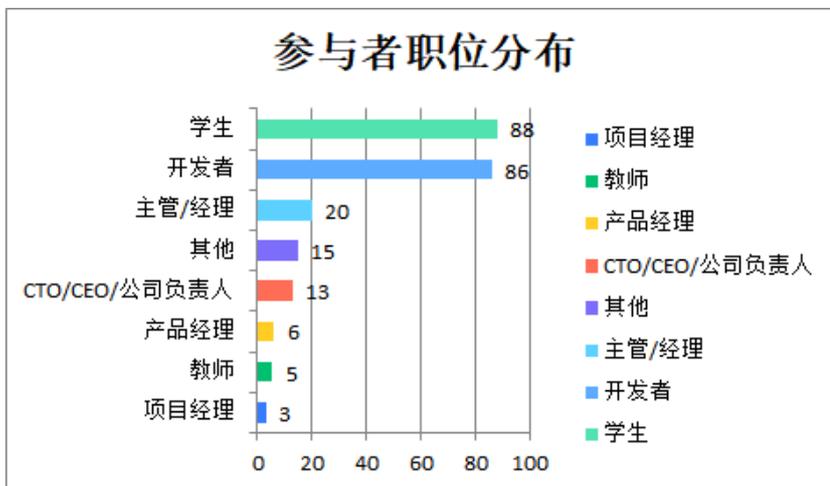


在参与者中，有近七成参与者从事互联网开发/软件开发领域，其次则是教育/学术/科研领域的参与者，占比约 15%，金融银行、媒体广告娱乐等领域也在积极参与开源。



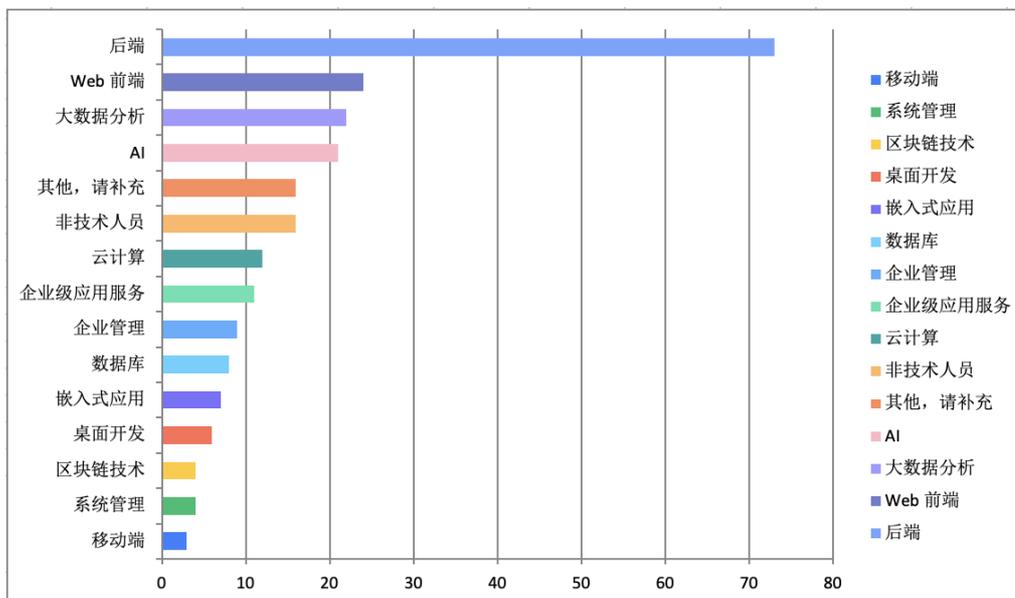
### 3.3 职位分布

参与者中学生和开发者占绝大多数，相比于 2019年，学生占比大大增加，约 37%，开发者占比约 36%。



### 3.4 所从事的技术方向

参与者所从事的技术方向以后端开发为主，占比约 31%，其次是 Web 前端、AI 以及大数据分析方向，占比分别为10%、9%、9%，非技术人员以及其他行业的从业者占比均为 7%，说明开源已经越来越受到社会各界的关注。



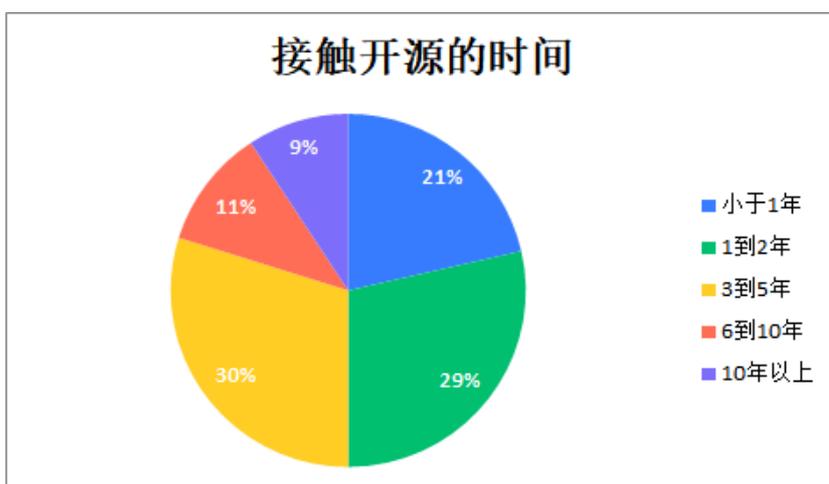
**【专家点评】**

陈阳：开源早期的时候，开源 = Linux。作为当时开源项目的主战场，Linux 操作系统、桌面办公软件（GNOME，OpenOffice）和浏览器（Mozilla）掀起了第一波开源的浪潮；近年来随着互联网的兴起，开源的项目和技术越来越百花齐放，从数据库到中间件，从前端到后端，从编程语言到编译器，从物联网到微服务，从大数据到人工智能，开源的技术和项目越来越丰富和多样化。

## 4、开源工作现状

### 4.1 接触开源的时间

有近三成参与者接触开源的时间在 1-2 年，近八成参与者接触开源时间在一年以上。

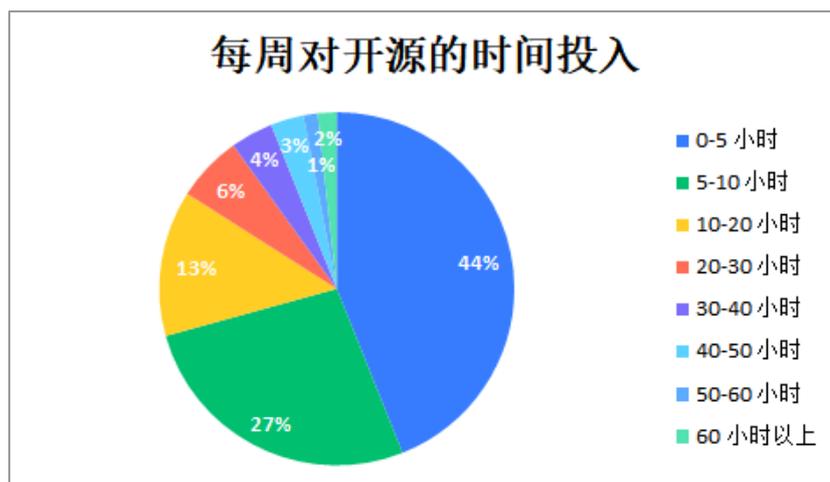


**【专家点评】**

陈阳：开源社区人员非常稳定，有一种类似大家庭的凝聚力。十几年前的开源圈子，活跃于一小群早期的开源人之间。我观察到身边的开源人士大多数都是超过 10 年以上的，并且未来也会持续投入开源。2020 年数据显示有 3 成新参与者，从另外一个角度也印证了开源运动在发展多年之后，呈现出一种出圈的状态。

## 4.2 在开源中的时间投入

约有 44% 的参与者每周在开源上的时间投入不足 5 小时，每周对开源投入时间在 5-20 个小时的参与者约有 40%。

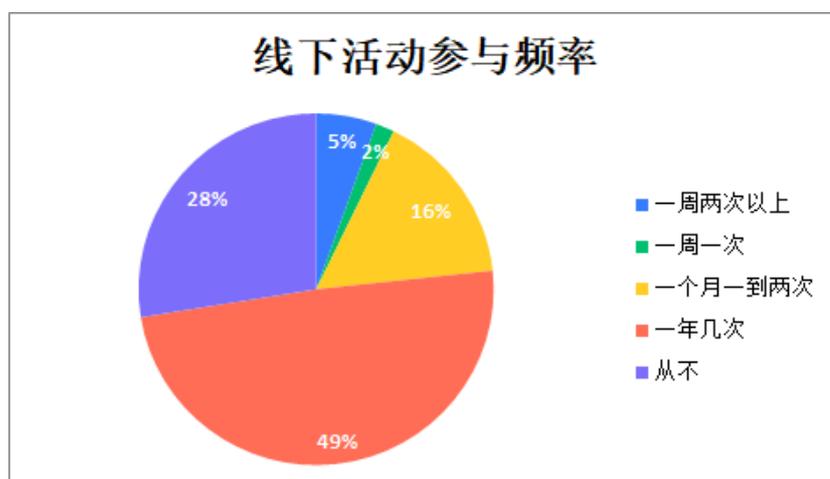


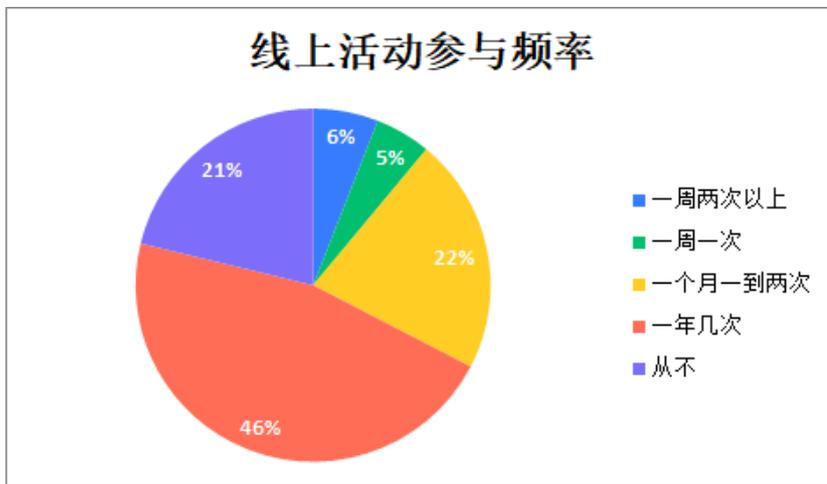
### 【专家点评】

**吴晟：**较低的职业开源开发者（一周超过 20 小时的贡献者）比例，显示出中国企业对于开源项目，依然主要停留在用户层面。多数开源贡献者依然使用业务时间或者工作间隙进行有限的贡献和参与。

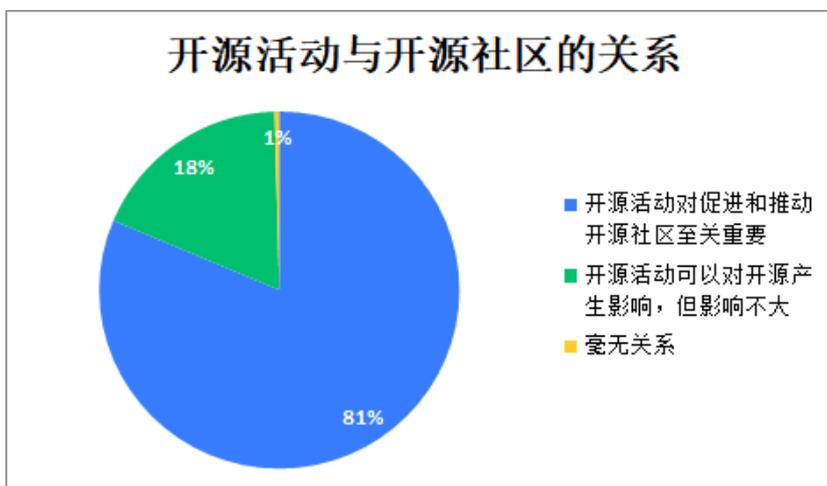
## 4.3 开源活动

绝大多数参与者每年只会参加几次线上/线下的开源活动，每个月参加一到两次线上/线下开源活动的参与者占比约两成。线下的开源活动大多以沙龙、讲座等为主，线上的开源活动以在线会议、邮件列表讨论、PR 活动等为主。





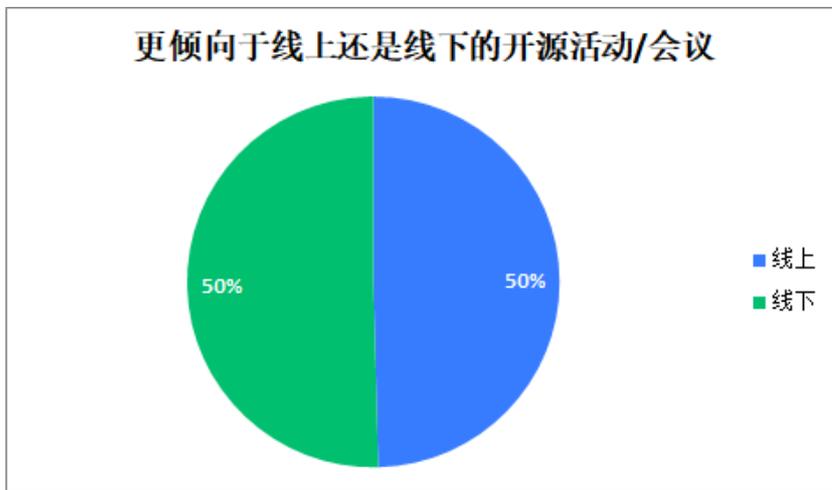
81% 的参与者认为开源活动对促进和推动开源社区至关重要。



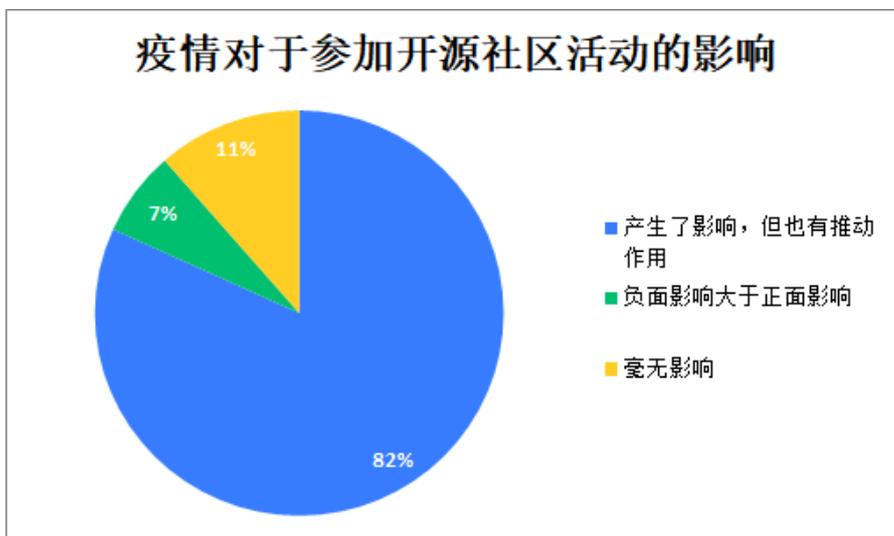
#### 【专家点评】

**高阳：**我们鼓励开发者积极参与到开源活动中来，尤其鼓励大家积极参与线下活动。线下的见面、沟通有助于建立起人与人之间的信任，对于社区的繁荣、人与人之间更高效的协作有重要意义。

关于开展线上还是线下的开源活动/会议，参与者们各抒己见，线下活动可以面对面地沟通交流，交流更有效率，氛围更好，可以有更多走出去看看，结交新朋友的机会，而线上活动则更加安全、便捷，不受时间、地域等限制，成本低廉，且会议内容可以录像，方便会后复查，当然，线下录像也是越来越普遍。



2020 年是十分特殊的一年，COVID-19 这场世界性的疫情给各行各业都带来了不同程度的冲击，82% 的参与者认为，疫情对于参加开源社区的活动产生了影响，但也具有推动作用。

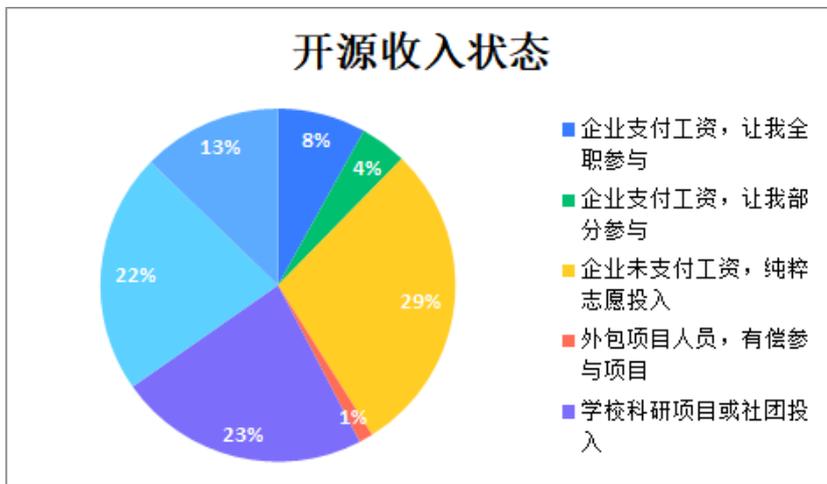


#### 【专家点评】

**刘天栋：**新冠疫情的冲击，加速了更多远程上班的族群参与了开源活动。以开源社在 2020 年 10 月 24-25 日举办的 [第五届中国开源年会 + Apache 中国路演] 为例，有超过百万人次在线参与，这是过去线下活动参与人数的千倍的成长，同时也有比往年更多的国际与国内的开源大咖在线分享。虽然这次大会也同步在五个城市举办了线下聚会，但是我们可以预期未来线上会议将成为常态。

## 4.4 开源收入

可以看到，很多人参与开源并不追求物质回报，有三成参与者没有开源收入，但依然热衷于开源工作，同时，我们还可以注意到有 12% 的参与者表示企业会支付工资支持员工全职/部分参与开源，有 23% 的参与者表示学校科研项目或社团投入方面有在支持参与开源，可见现在企业和学校都开始逐渐关注并重视开源。



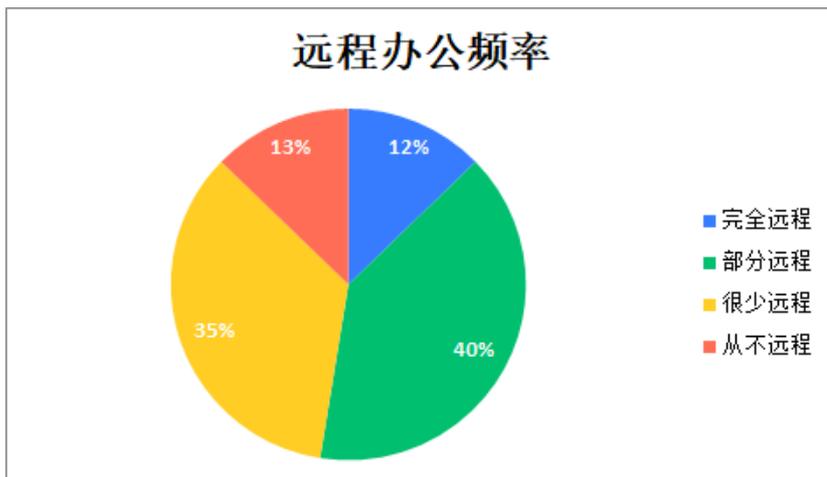
#### 【专家点评】

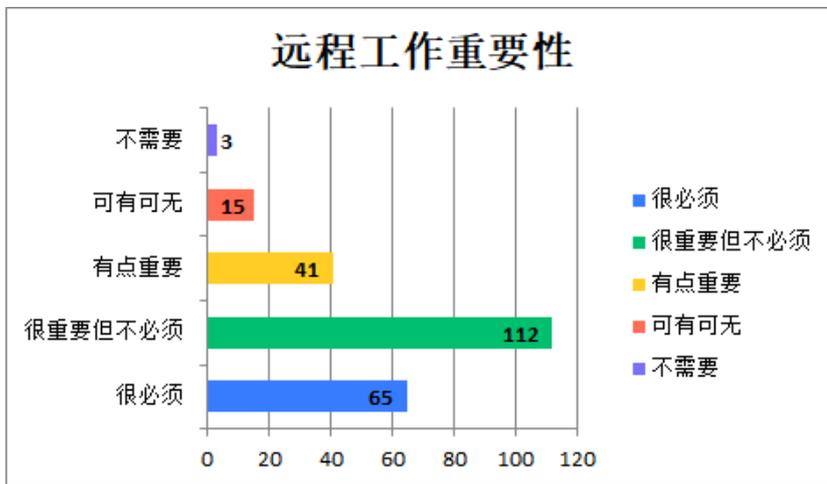
**吴晟：**能够获得开源收入比例和开源参与时间超过 20 小时的比例高度一致，很好的显示出商业价值在开源贡献强度上的显著变化。虽然，贡献时间和收入并不代表开源贡献和开源项目的质量，但高质量的项目需要一定的商业支持才能得到良性循环。

**陈阳：**有12%的参与者从开源中获得兼职或全职收入，开源布道师，开源运营经理，开源开发者等都是2020年的热门职业，这说明开源的生态正在进一步完善，企业开始进行开源人才储备，开源的商业逻辑正在明晰。

## 4.5 远程办公

远程办公是当下非常重要的一种办公方式，在参与者中，有八成认为远程办公是重要的，并且有接近九成的参与者有远程办公的经历。



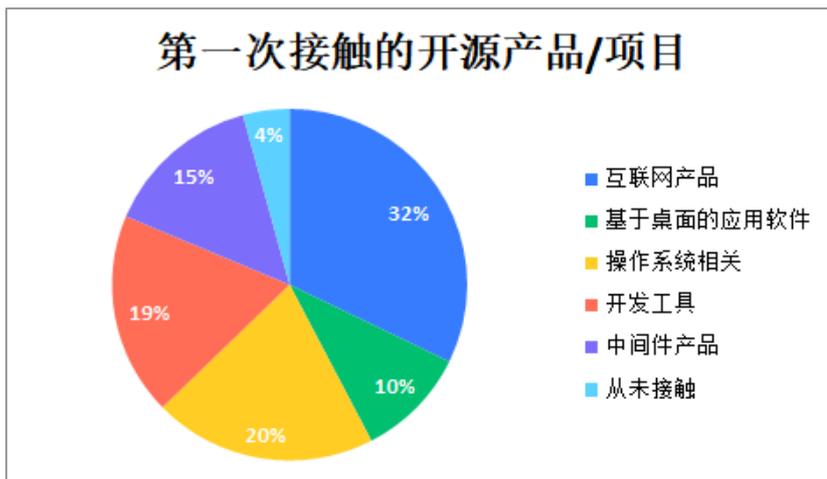


**【专家点评】**

**高阳：**远程办公会变成一种常态化的办公和协作方式融入到我们的工作和生活中，开源运动和分布式、远程协同本来就是天然融合的。

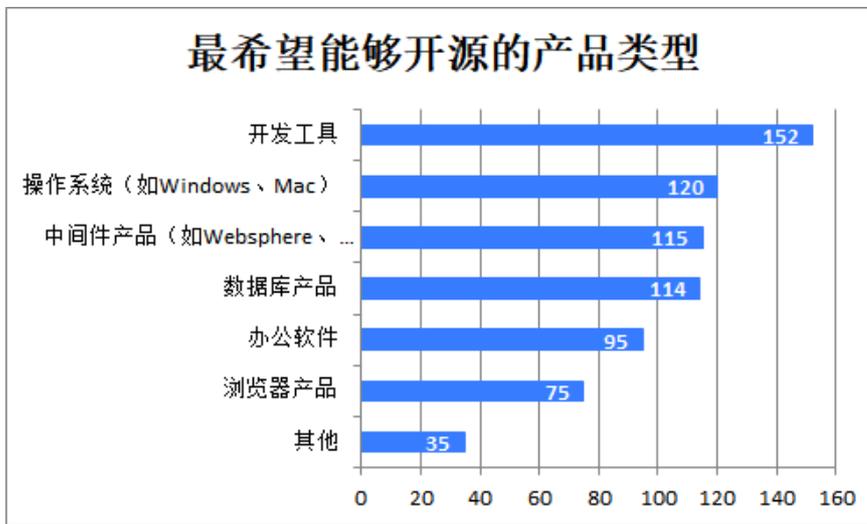
#### 4.6 第一次接触的开源产品

有 32% 的参与者第一次接触的开源产品为互联网产品，紧随其后的便是操作系统相关的产品与开发工具，调查结果与以往基本一致，说明互联网产品和操作系统相关产品仍然是大家了解和接触开源的最初窗口。



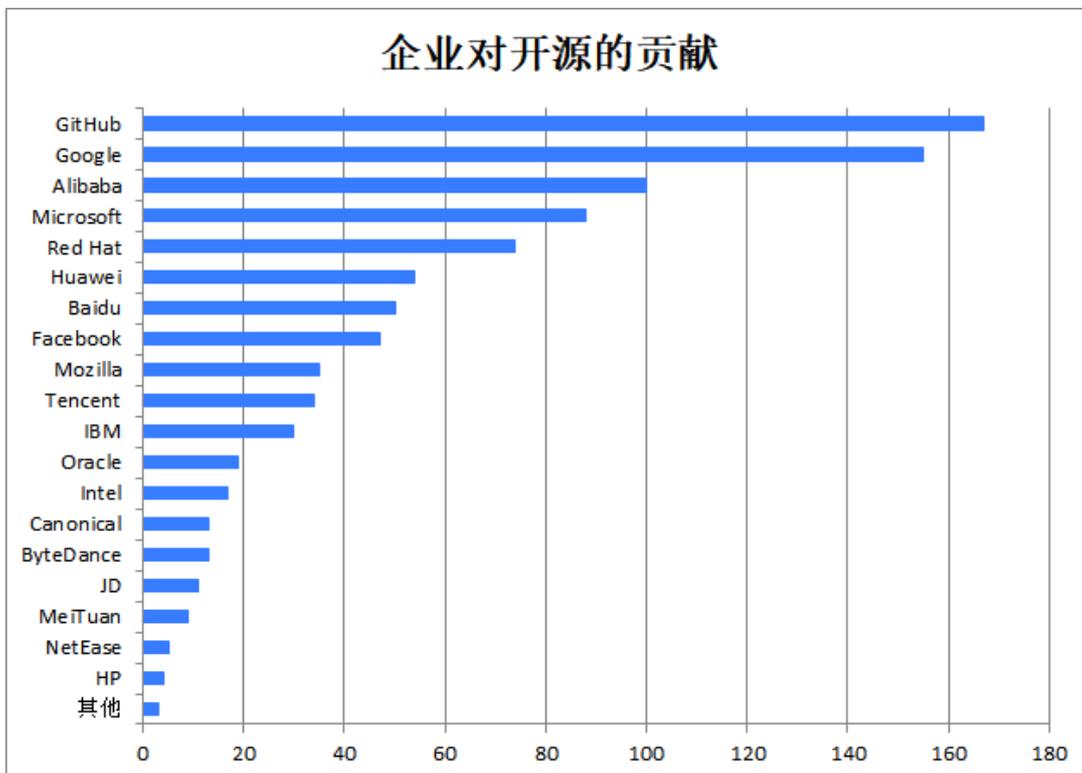
#### 4.7 最希望能够开源的产品

相比于 2019 年，开发工具在今年成为大家最希望能够开源的产品类型第一名，紧接着的是操作系统，数据库产品与中间件产品也是很多人希望能够开源的产品类型。



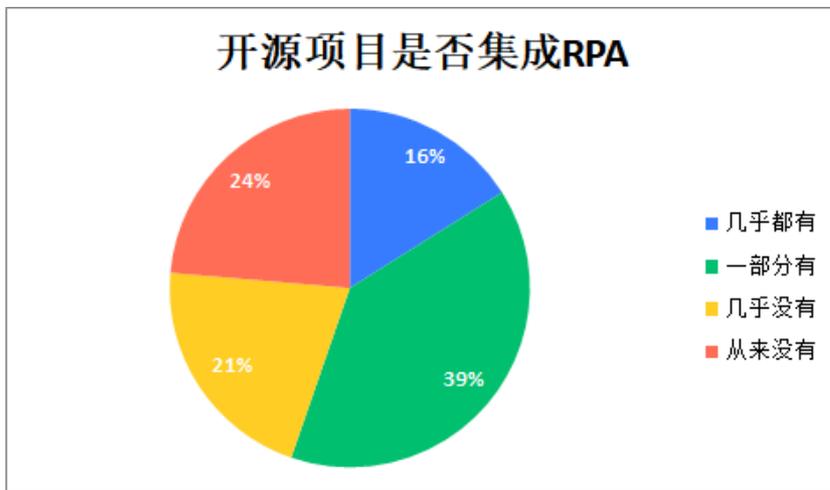
#### 4.8 企业对开源的贡献

在参与者眼中，GitHub 对于开源软件的贡献最大，其次是 Google，而在国内企业中，Alibaba、Huawei 和 Baidu 分别占据前三名。



#### 4.9 机器人流程自动化

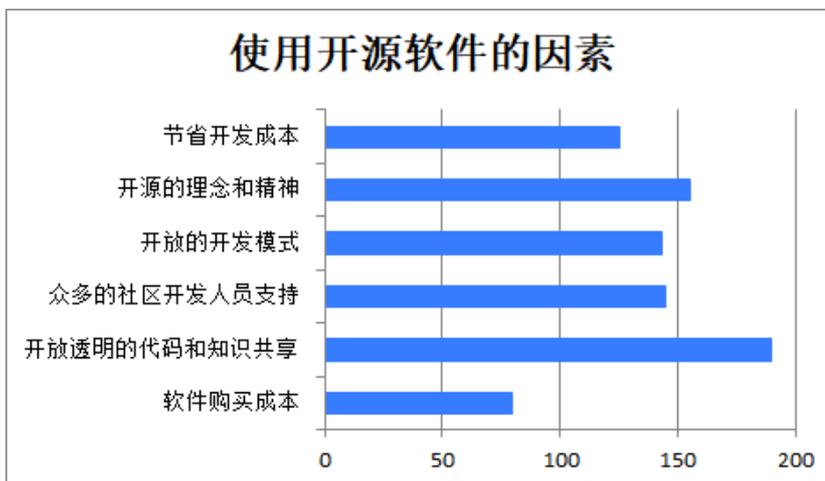
对于开源项目是否继承了 RPA，即机器人流程自动化工具，有四成参与者表示一部分项目集成了 RPA，16% 的参与者表示几乎所有项目都有集成 RPA，同时也有 24% 的参与者表示从来没有在开源项目中集成过 RPA。



## 5、开源社区参与现状

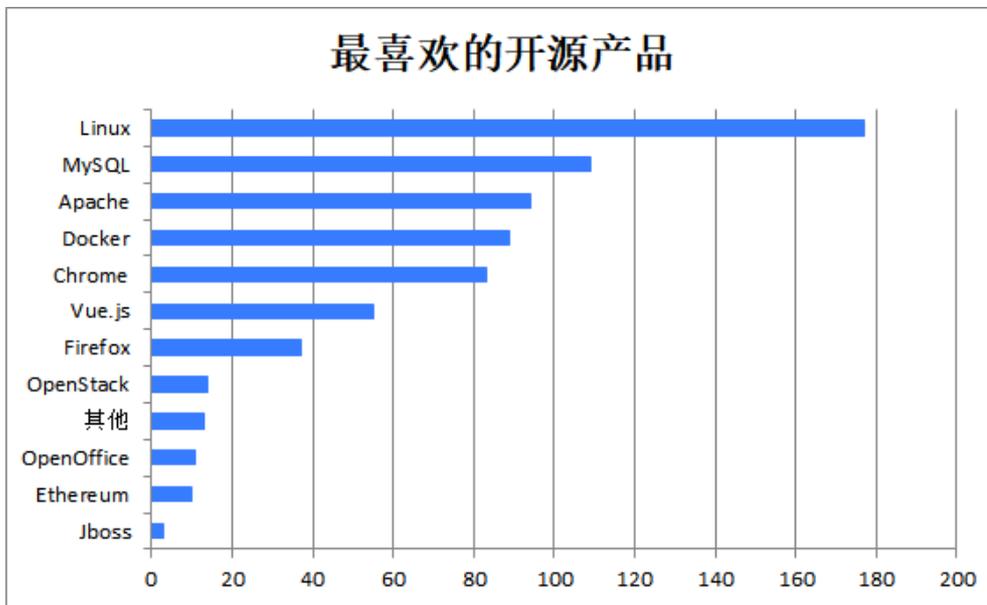
### 5.1 开源最吸引你的因素

开放透明的代码和知识共享，以及开源的精神理念成为开源最吸引参与者的因素，而软件的学习成本却并不是主要的因素。



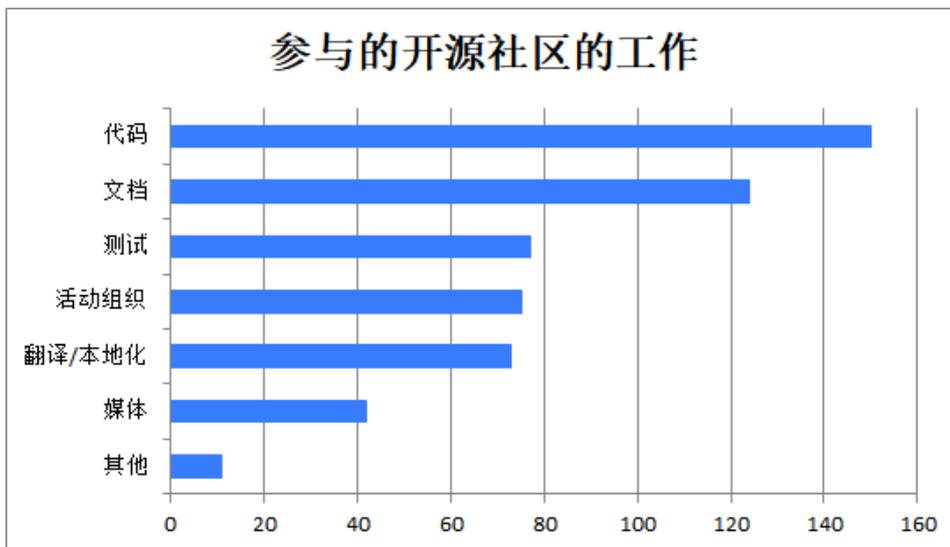
### 5.2 最喜欢的开源产品是什么

Linux 以巨大优势，成为参与者最喜爱的开源产品，MySQL 紧随其后，Apache 和容器界的黑马 Docker 分别位列第三、第四。



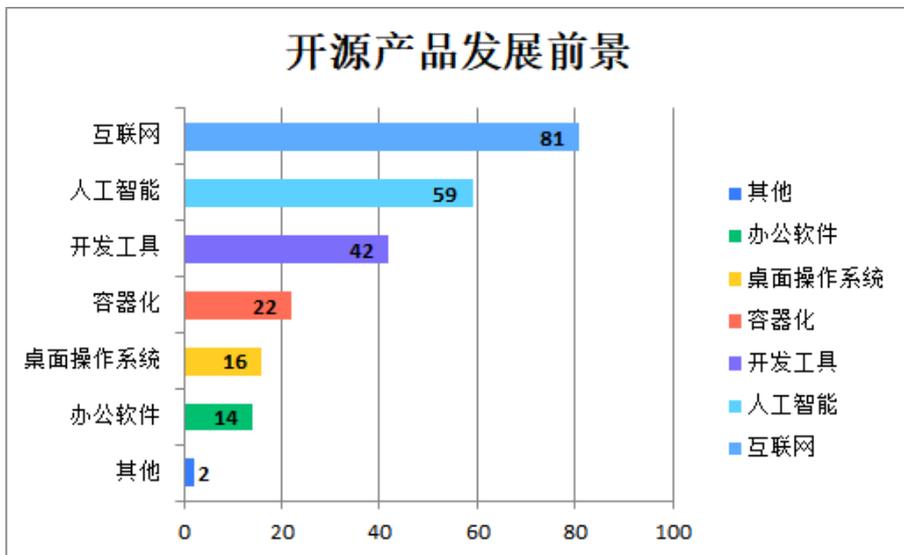
### 5.3 具体参与的社区工作

绝大多数参与者在社区都参与代码或文档撰写的工作，测试、本地化以及活动组织也是很多人在社区会参与的工作。



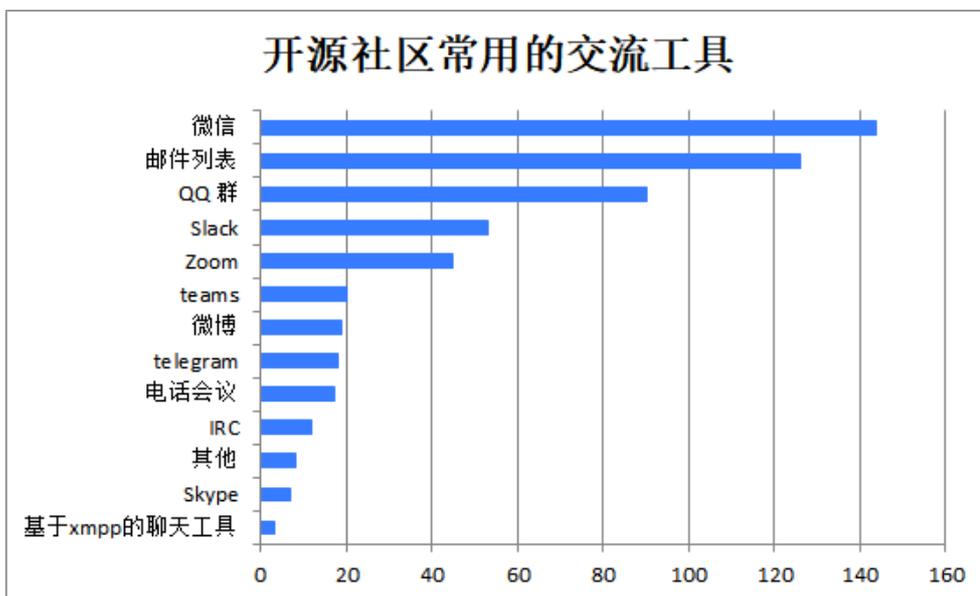
### 5.4 最看好的开源产品

开源产品的前景展望中，互联网产品以 34% 的比例成为参与者最看好的开源产品发展方向。人工智能和开发工具异军突起，成为现在开源产品中的风口和热点。



## 5.5 开源社区的交流方式

在社区的交流方式中，微信和邮件列表目前最为普遍，QQ 也是国内重要的即时交流工具。新锐的交流工具比例上升：Slack 和 Zoom 也是值得关注的现象。



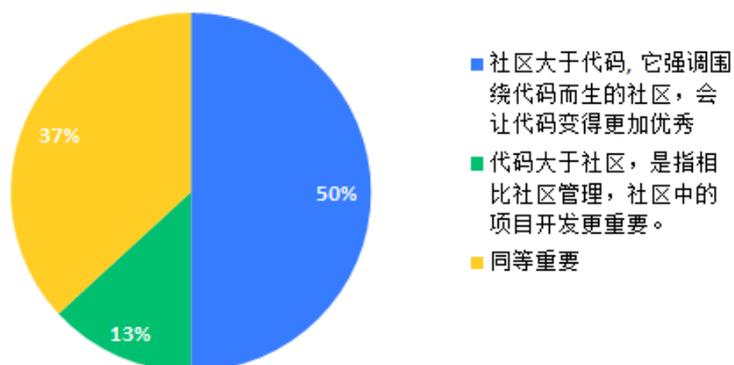
#### 【专家点评】

**吴晟：**微信、QQ 群的高强度使用，仍然显示出开源圈子缺少国际化社交和协作能力的一贯性短板。同样反映出中文依然是绝大部分的开源参与者唯一热衷的交流方式。

## 5.6 社区与代码的关系

有近九成的参与者认为在开源社区中，社区的价值大于或等同于代码，因为围绕代码而生的社区，会让代码变得更加优秀。

## “社区”与“代码”的重要性



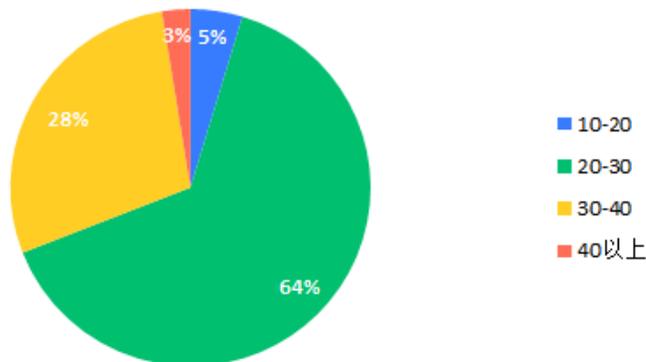
### 【专家点评】

**高阳:** 很高兴看到大家对于社区价值的认同, 在我们看来, 一个健康的社区远比优秀的代码更重要, 只有健康、多样和友好的社区才能推动项目可持续发展, 正如一句话说的“一个人走的很快, 一群人走得很远”。

## 5.7 开源社区年龄分布

开源社区中, 20-30 岁的贡献者是主要组成部分, 占比超过六成。

### 开源社区人群年龄分布



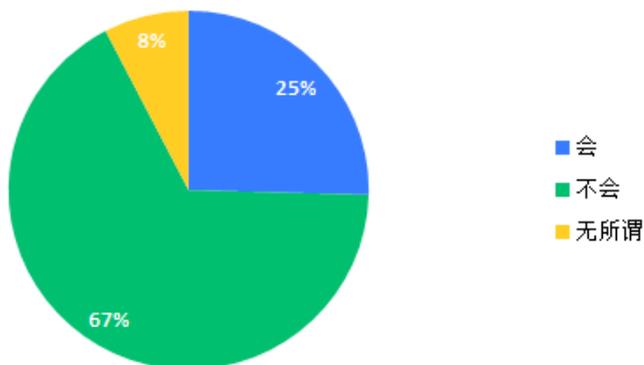
### 【专家点评】

**吴晟:** 开源参与者中年轻研发者较多, 一方面反映了中国对于开源的参与时间较晚, 学生做为可塑性较高的群体, 更容易跟随最近的开源热度, 选择参与到开源。缺少 35-40 岁以上的开发者, 正面反应了中国资深工程师的缺乏。缘起中国的开源项目的大规模崛起, 将十分依赖此部分人的比例提升。

## 5.8 开源软件安全

有近七成的参与者表示不会有不安全感, 但同时也仍有 25% 的参与者表示会有此类担忧, 开源软件的安全问题仍然值得我们关注。

## 使用开源软件是否会有不安全感



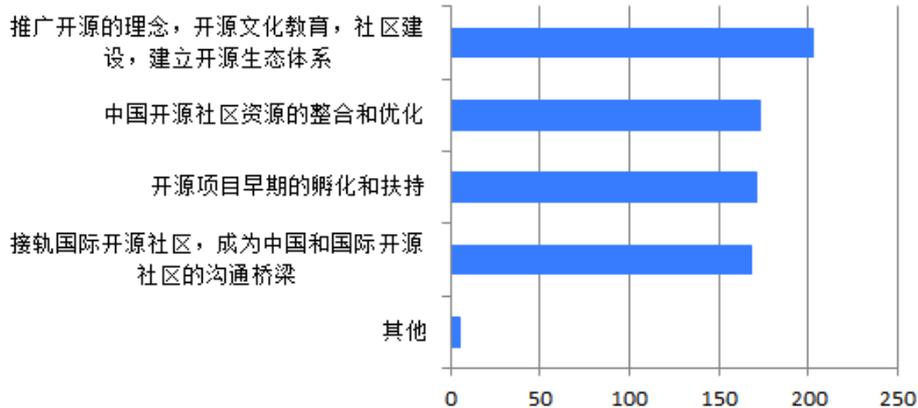
### 【专家点评】

**红薯：**开源软件由于其机制的关系，通过开源社区不断地发现问题并维护，其安全问题可能并不是开发者们的主要关注点。其实除了技术方面的安全问题，开源合规性等许可证方面的安全问题同样需要开发者们重视，许可证冲突问题所带来的法律风险影响非常大且很难被发现，尤其是对于企业来说，检查所使用的开源软件合规性，其重要性不亚于其技术安全。

## 5.9 中国成立开源基金会的作用

参与者们一致认为中国成立开源基金会是非常有意义的，可以推广开源的理念，开源文化教育，社区建设，建立开源生态体系，利于开源项目早期的孵化和扶持，帮助中国开源社区进行资源的整合和优化，也有助于接轨国际开源社区。

## 中国成立开源基金会的作用



### 【专家点评】

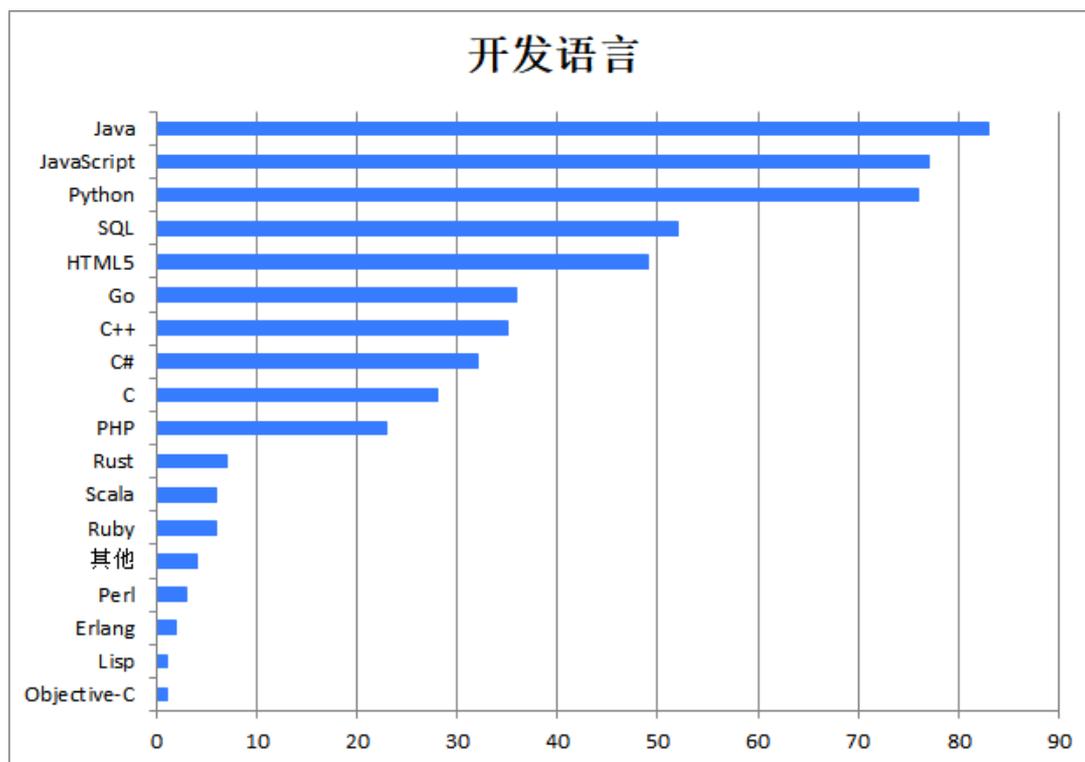
**蒋涛：**在全球科技竞争和中国开源核心技术发展态势下，从开源生态制高点着眼，建设和发展中国开源基金会意义重大。通过中国开源基金会的发展壮大，不断鼓励国内巨头科技企业和社会力量合力共建开源创新生态环境；同时，通过学习借鉴更完善的基金会运行模式、组织机制和法律制度，建设可持续发展的中国开源创新力量。此外，在当前全球开源商业模式越来越成熟的市场环境下，建设开源基金会，还可以引导国内外基金和投资机构孵化支持开源，培育中国开源『独角兽』，最终形成资本、企业、开发人员为主体、共享共惠的中国开源商业生态体系。

**高阳**：2020 年中国首个开源基金会开放原子面世，这对于在本土推动开源教育的普及以及开源生态的整体繁荣有重要意义。我们期待看到开放原子可以成为开源的连接器，连接国际开源社区、链接国内外优质资源，帮助开源项目成长与成功。

## 6、开发者技术现状

### 6.1 开发语言

开发语言呈现多超多强的状态，前三甲分别为 Java、JavaScript 和 Python，排名与去年基本保持一致。

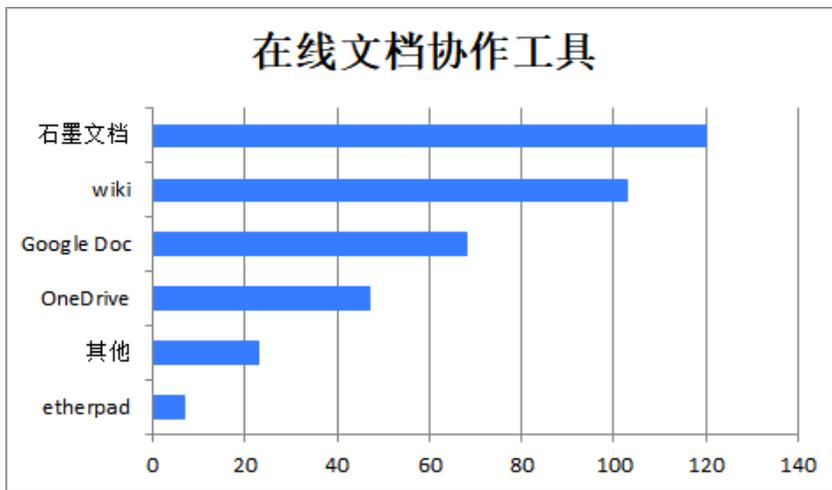


#### 【专家点评】

**王伟**：编程语言是大家喜闻乐见的话题，排名方面没有太大的变化。值得一提的是 SQL 语言，这个平时存在感貌似不是很强，但实用性却非常强。实际上，本报告中的 GitHub 数据篇中的基础数据全部都是用 SQL 跑出来的，是一项开发者必备的重要基础技能。

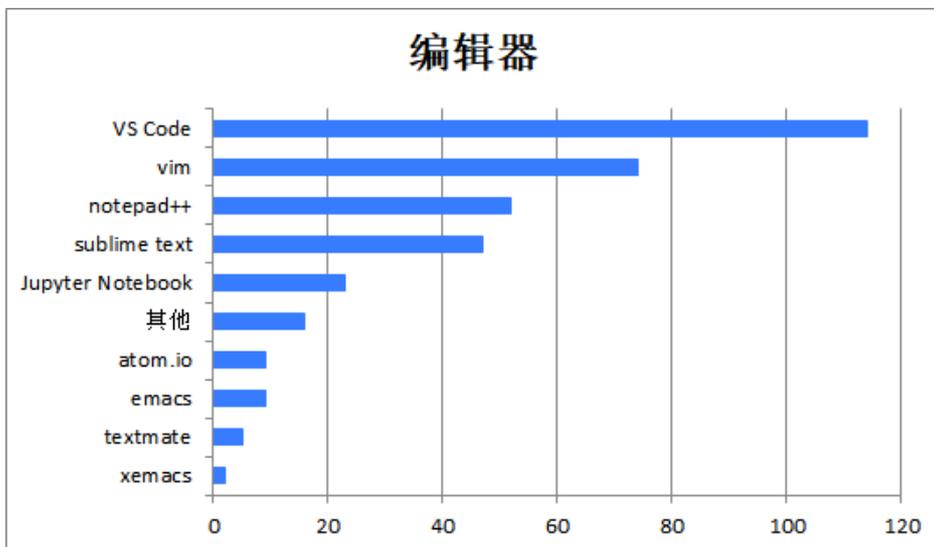
### 6.2 在线文档协作编辑工具

在线文档编辑工具中，石墨文档是目前中国开源社区经常使用的工具，wiki 也依然是非常多参与者使用的协作工具。



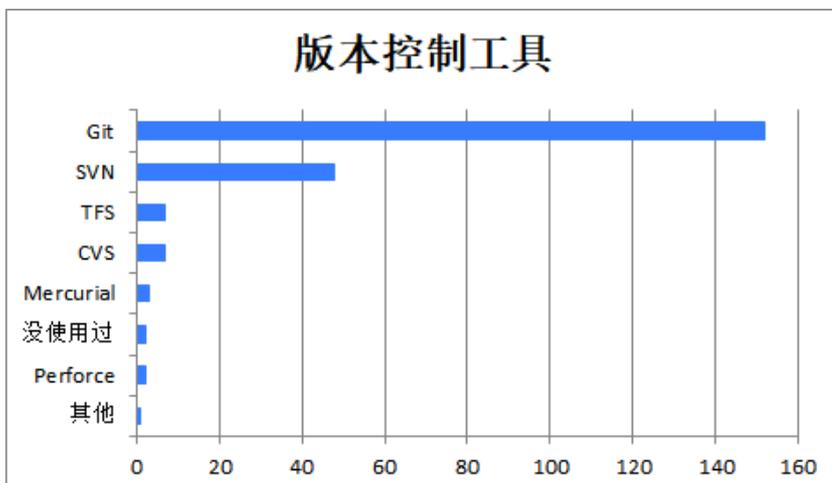
## 6.3 编辑器

VS Code 今年仍然是最受欢迎的编辑器，紧随其后的是 vim 以及 notepad++。



## 6.4 版本控制工具

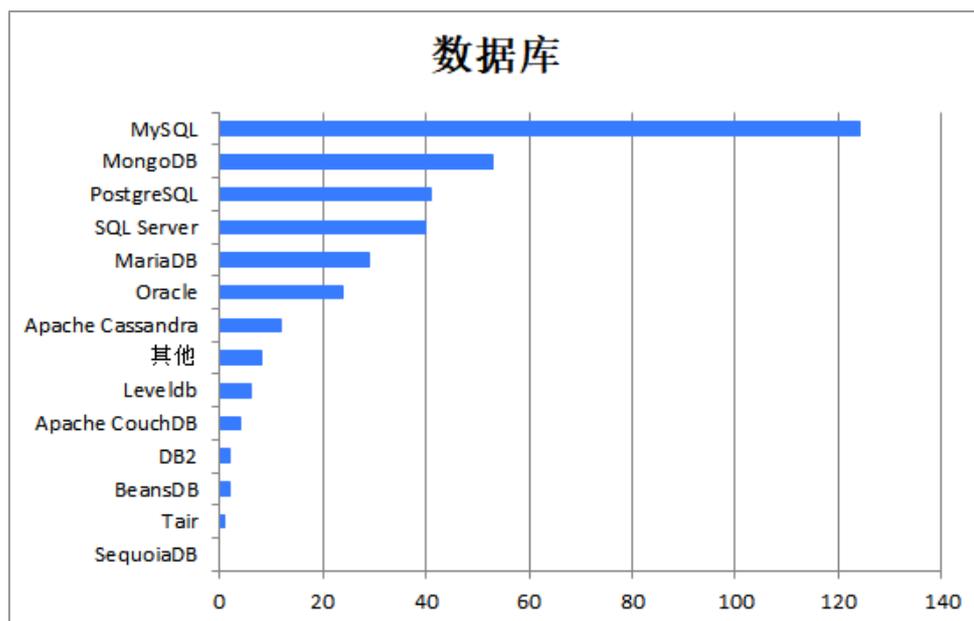
毫无疑问的是，Git 一枝独秀，具有绝对性优势。SVN、TFS、CVS 目前仍有不少参与者在使用。



---

## 6.5 数据库

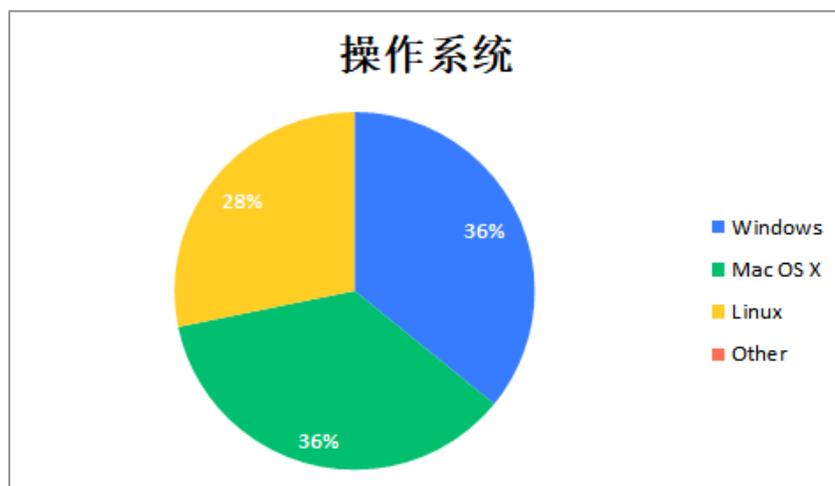
数据库使用情况中，不出意外地，MySQL 以绝对优势遥遥领先，MongoDB 与 PostgreSQL 跟随其后。



---

## 6.6 操作系统

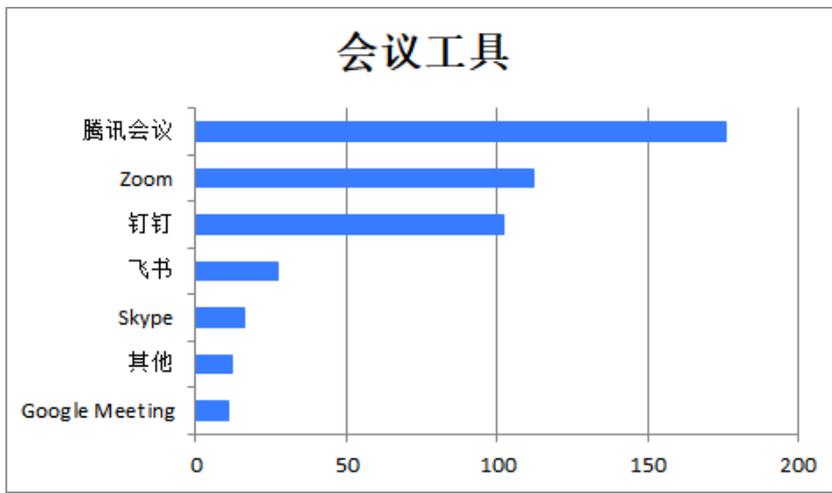
操作系统的使用中，Windows、Linux、MacOS X 三分天下。



---

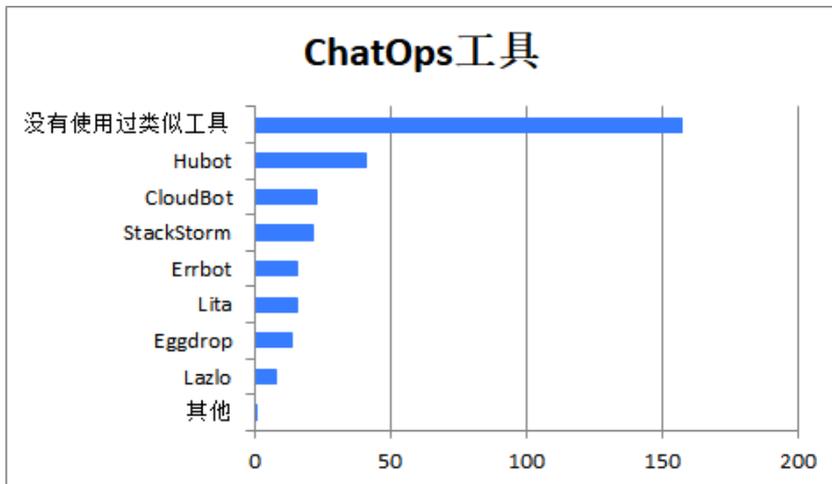
## 6.6 会议工具

腾讯会议是目前参与者们使用最多的会议工具，其次则是 Zoom、钉钉。



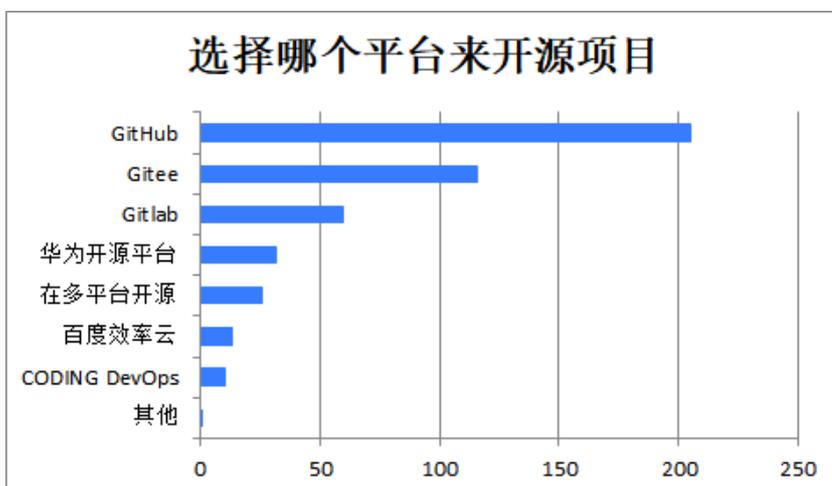
## 6.7 聊天运维工具

目前，绝大多数的参与者都没有使用过聊天运维工具来进行项目的自动化管理，而对于使用过该类工具的参与者来说，Hubot 是他们使用最多的 ChatOps 工具。



## 6.8 开源平台

如果要开源自己的项目，87% 参与者都表示将会在 GitHub 来开源自己的项目，其次则是 Gitee 和 Gitlab。



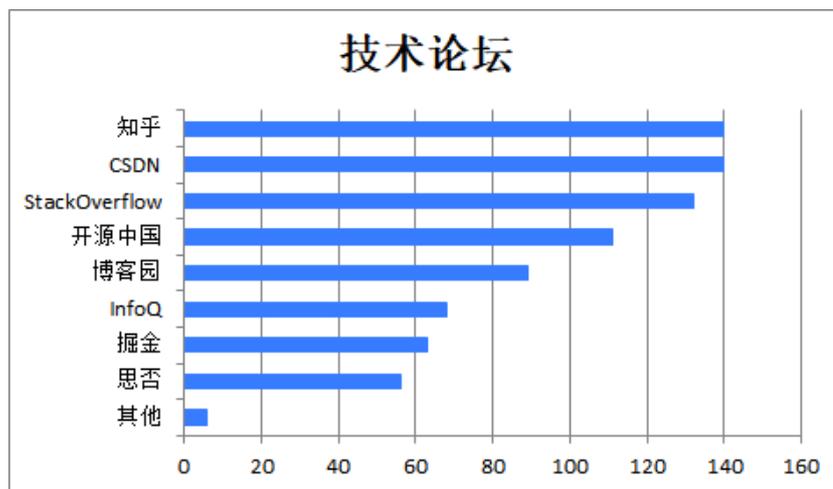
---

### 【专家点评】

**蒋涛**：随着开源规模和商业化的快速发展，开源平台作为开源项目和开发者生态的重要基础及支撑体系，未来将承载更多的服务和应用场景，也将发展出各自特色。2020年新发布的CODE China，作为独立第三方开源平台，将重点围绕AIOT，为更多开源项目和开发者提供运营支持和生态服务。

## 6.9 技术论坛

知乎和CSDN目前是最受参与者们青睐的两个技术论坛，其次则是StackOverflow，开源中国与博客园也受到很多参与者的喜爱。



---

## 7、总结 & 致谢

本次问卷在传播和收集过程中多有坎坷，最终呈现的不过是开源世界的小小一隅，但意义重大。见微知著，不难察觉到中国的开源结构已经在发生变化，开源的星星之火正在燎原，势不可当。希望这份报告能够抛砖引玉，吸引更多的社区和开发者加入。

该报告的问卷问题，报告文档均发布在代码托管平台，开源社官方网站和合作伙伴网站平台，基于 Creative Commons Attribution-ShareAlike 4.0 International (CC BY-SA 4.0) 许可证协议分享。对于问卷的问题设计和报告内容有任何建议和想法，欢迎您在代码托管平台上提交 Patch，对报告进行补充和贡献。您的一小步，整个中国开源社区的一大步。

此报告问卷篇作出贡献的开源社中国开源年度报告小组成员：夏小雅，宁泽欣；社区合作伙伴：X-lab，Gitee，微软 Reactor。感谢特邀的开源领域的专家，更要特别感谢积极参与我们调研的每一位朋友。

---

## 二、GitHub 数据篇

### 1、概述

## 1.1 背景

本部分内容采用 GitHub 全域事件日志进行统计分析（2020 年共计 **8.74亿**条），相较 2019 年的 5.46 亿条增长约 **60%**。并对分析出的部分项目进行人工标注，挑选出中国的个人开发者和企业组织，同时构建了科学合理的数学模型进行分析。

在开发者活跃度与项目活跃度的定义下，本次统计得到 2020 年总活跃项目数量约 **5,373万**个，相较 2019 年的约 3,972 万 增长约 **35.3%**；2020 年 GitHub 总开发者数为 **5,600多万**，总活跃开发者数量 **约 1,446万**，相较 2019 年的约 1,19 万 增加了 **约 21.2%**。

## 1.2 指标解释

指标名称	含义
language	项目开发使用的主要语言
activity	项目活跃度（通过加权计算得到）
developer_count	在活跃度定义下参与该项目的开发者人数
issue_comment	项目所有 Issue 和 PR 在 2020 年获得的评论总数
open_issue	项目 2020 年新增的 issue 数量
open_pull	项目 2020 年新增的 PR（pull request）数量
pull_review_comment	项目 2020 年所有 PR 下的 Review 评论数
merge_pull	项目 2020 年被合并的 PR 数量
pull_commits	项目通过 PR 合入的 commit 数量
pull_additions	项目通过 PR 合入增加的代码行数
pull_deletions	项目通过 PR 合入删除的代码行数

## 2、主要内容与发现

### 2.1 世界开源 Top10 项目

#	name	language	activity	developer_count	issue_comment	open_issue	open_pull	pull_review_comment	merge_pull
1	flutter/flutter	Dart	35108.77	16758	128091	14857	7248	18778	5003
2	microsoft/vscode	TypeScript	27336.95	14162	101450	16366	1816	1778	1383
3	MicrosoftDocs/azure-docs	PowerShell	24021.03	9507	87663	11749	3118	885	1853
4	home-assistant/core	Python	22609.91	8215	77017	5381	7851	31039	6697
5	tensorflow/tensorflow	C++	21637.69	9681	63983	6241	2993	8073	2198
6	kubernetes/kubernetes	Go	19878.4	6111	239791	3669	6869	31345	4817
7	NixOS/nixpkgs	Nix	18571.45	3049	84523	4368	19693	30270	16309
8	pytorch/pytorch	C++	14515.73	4844	69110	4890	11776	39109	336
9	dotnet/runtime	C#	14076.75	3731	83600	7082	7540	41215	6531
10	DefinitelyTyped/DefinitelyTyped	TypeScript	13267.93	4100	53588	538	6426	6529	5414

活跃度最高的项目是来自谷歌的前端跨平台开发框架 flutter/flutter，此外，源于谷歌的深度学习框架 tensorflow/tensorflow，容器编排系统 kubernetes/kubernetes 也分别位于第 5 名和第 6 名，这说明了谷歌在开源上的努力和影响力获得了业内的认可。

微软的跨平台代码编辑器 microsoft/vscode 和微软使用开源的方式来建设其 Azure 云平台的项目 MicrosoftDocs/azure-docs 分别位于第二名和第三名，表明微软在开源上的努力获得了程序员的认可。

microsoft/vscode 和 DefinitelyTyped/DefinitelyTyped 均使用 TypeScript 作为主要的编程语言，这可能与 TypeScript 的受欢迎程度急剧攀升有关。

## 2.2 中国开源 Top 50 项目

#	name	language	activeness	developer_count	issue_comment	open_issue	open_pull	pull_review_comment	merge_pull
1	ant-design/ant-design	TypeScript	8090.1	3746	32882	3341	2844	3315	2416
2	PaddlePaddle/Paddle	Python	5236.1	960	15226	1817	5487	9517	4083
3	pingcap/tidb	Go	4339.4	482	53550	2727	5041	11117	4097
4	NervJS/taro	JavaScript	3649.4	1621	9595	1768	805	132	602
5	apache/shardingsphere	Java	2858.7	786	9332	1581	3234	1851	2990
6	ant-design/ant-design-pro	TypeScript	2784.2	1362	6625	1383	228	118	118
7	alibaba/nacos	Java	2760.6	1204	7037	1233	635	764	491
8	apache/incubator-tvm	Python	2699.6	355	8255	363	2106	8526	1876
9	ElemeFE/element	Vue	2507.7	1694	5016	567	117	7	19
10	apache/incubator-echarts	TypeScript	2307.4	1183	7622	962	301	328	230
11	tikv/tikv	Rust	2184.8	259	17964	987	2023	5588	1531
12	apache/skywalking	Java	1958.6	525	6904	892	838	3528	725
13	youzan/vant	JavaScript	1952.8	970	4978	899	594	181	565
14	apache/incubator-dolphinScheduler	Java	1929.4	458	9439	1138	1440	722	1135
15	pingcap/docs-cn	TeX	1915.3	172	8257	68	2903	6746	2731
16	seata/seata	Java	1720.6	629	3824	587	495	1810	354
17	neoclide/coc.nvim	TypeScript	1670.1	794	4276	684	214	44	146
18	PaddlePaddle/PaddleOCR	Python	1617.5	564	3896	849	545	602	491
19	ApolloAuto/apollo	C++	1589.3	379	3306	703	1732	1477	1503
20	apache/dubbo	Java	1573.8	668	2537	584	558	165	329
21	PaddlePaddle/PaddleDetection	Python	1573	439	4315	903	786	1174	650
22	apache/apisix	Lua	1539.9	300	5874	1030	1000	3363	837
23	pingcap/pd	Go	1481.4	121	7400	665	1682	4653	1538
24	xitu/gold-miner		1368.3	169	3857	459	618	3635	570
25	ant-design/pro-components	TypeScript	1285.8	459	4040	735	401	238	374
26	pingcap/docs	Shell	1262.6	113	7005	44	2686	4999	2622
27	antvis/G6	TypeScript	1236.8	461	3408	798	355	188	314
28	apache/incubator-doris	C++	1221.4	149	2327	1055	1410	3543	1227
29	consenslabs/token-profile	TypeScript	1188.1	332	2219	70	753	19	386
30	kubeedge/kubeedge	Go	1187.5	255	6175	357	670	1673	511
31	pingcap/tidb-operator	Go	1139.6	121	8738	697	1531	3683	1437
32	antvis/G2Plot	TypeScript	1108.5	275	3483	603	913	1515	811
33	youzan/vant-weapp	JavaScript	1105.3	561	2211	504	278	31	269
34	Tencent/mxnet	C++	1104.4	419	2665	532	281	205	227
35	dotnet/docfx	C#	1098.2	246	9972	129	1290	2068	1152
36	alibaba/canal	Java	1095.4	607	1789	389	44	3	12
37	PaddlePaddle/Paddle-Lite	C++	1069.5	205	1364	374	1852	1772	1488
38	alibaba/spring-cloud-alibaba	Java	1040	581	1738	365	157	50	98
39	PaddlePaddle/models	Python	1028.1	301	1410	253	502	500	375
40	antvis/G2	TypeScript	982.21	345	2384	534	507	516	461
41	alibaba/MNN	C++	972.79	427	1916	518	82	17	54
42	alibaba/Sentinel	Java	946.34	372	2143	322	179	569	104
43	recharts/recharts	TypeScript	923.84	629	1562	178	34	33	34
44	PaddlePaddle/PaddleHub	Python	898.05	290	1985	408	331	759	266
45	alibaba/fastjson	Java	870.76	519	1244	213	90	11	54
46	eggjs/egg	JavaScript	853.27	435	1973	274	29	28	24
47	ctripcorp/apollo	Java	835.06	404	2139	300	94	220	79
48	apache/rocketmq	Java	825.41	280	1965	396	254	205	106
49	Tencent/bk-ci	Kotlin	789.93	160	1770	1399	1189	292	1073
50	alibaba/easyexcel	Java	729.58	403	1250	289	42	5	20

Top50 列表中，从项目活跃度来看，第 1 名与第 50 名的项目活跃度相差超过 10 倍，说明中国开源项目之间的差距较大。

从这个列表中，我们发现，阿里在开源领域中的成绩显著。Top10 项目中上榜的项目有 4 个，除了蚂蚁金服采用 React 封装的一套组件库 ant-design/ant-design，基于 Ant Design 框架搭建的中后台管理控制台的脚手架 ant-design/ant-design-pro 和致力于配置和管理微服务的特性集 alibaba/nacos 外，饿了么（已被阿里收购）的开源项目 ElemeFE/element 作为前端组件，表现也十分突出。

PingCAP 在开源领域的表现也是非常地亮眼。Top50 项目中上榜的项目有 6 个，包括由其自主设计、研发的开源分布式关系型数据库 pingcap/tidb，分布式事务型的键值数据库 tikv/tikv，文档项目 pingcap/docs-cn、pingcap/docs 等，表明 PingCAP 很重视项目文档的建设。值得关注的项目是 pingcap/tidb，该项目的 issue\_comment，open\_issue，pull\_review\_comment，merge\_pull 等属性是 50 个项目中最高的，达到惊人的 52,871 和 10,981，相对比之下，issue\_comment 属性下排名第二的 ant-design/ant-design 只有它的 61.5% 左右。而其 480 的开发者参与数量，与其他前端项目相较甚远，但却有着如此高的活跃度，也可以看出其社区的活跃程度。

百度在人工智能领域的表现非常不错，其深度学习平台 PaddlePaddle 占据了 6 个项目，分别是核心框架 Paddle 以及相关工具库、扩展版本和模型库 Models；此外开放自动驾驶平台

ApolloAuto/apollo 也榜上有名。

中国的 Top50 项目列表中，包括阿里的 Ant-Design 组件库，京东基于 React 前端框架的开发框架 taro，由饿了么（已被阿里收购）前端团队开源的 Vue UI 组件库 Element 等等，这说明在国内，前端群体在开源社区更为活跃；另外前端代码一般也不太涉密，因此公司在心态上更开放一些。不过这其中也有一点需要引起注意，上榜的前端项目组件库居多，但是缺少核心项目。

属于人工智能领域的项目 PaddlePaddle/Paddle、apache/incubator-tvm、Tencent/ncnn、alibaba/MNN 均在中国 Top50 项目列表中占据一席之地，这说明了国内各大公司都在大力发展人工智能领域。

## 2.3 中国开源企业情况分析

各大开源项目的背后，基本都有科技公司的支持，我们计算出了科技公司所维护的开源项目在 2020 年的活跃情况，结果如下表所示：

#	name	activeness	repo_count	developer_count	issue_comment	open_issue	open_pull	pull_review_comment	merge_pull
1	Alibaba	58353.73	757	25361	132236	21644	22144	19639	16347
2	Baidu	24868.17	179	7164	55330	9508	17921	21752	14400
3	PingCAP	22113.01	106	2788	131811	7696	24710	56503	22013
4	Tencent	11285.7	204	4345	21583	5927	5932	2440	5051
5	JD	7560.98	34	2982	20843	3652	4210	2045	3695
6	Huawei	3648.04	84	942	10312	1288	2484	3673	2120
7	Youzan	3298.79	21	1612	7437	1466	1089	539	1022
8	DiDi	1878.85	38	717	3172	740	576	185	438
9	Deepin	1560.63	74	819	2556	543	173	10	79
10	Xiaomi	1474.32	39	462	1776	787	902	3071	753
11	WeBank	1453.54	44	301	2170	467	1702	407	1414
12	Juejin	1452.99	7	196	3941	465	641	3711	587
13	Bytedance	1220.59	37	415	1996	381	445	594	360
14	360	1117.96	55	519	1821	401	288	37	180
15	Netease	1074.96	37	551	1656	392	113	27	49
16	Bilibili	991.19	19	621	1250	201	56	36	36
17	Meituan	986.89	26	520	1408	295	200	16	137
18	CTrip	976.78	9	457	2390	328	165	250	119
19	Linux China	582.7	5	50	230	8	3554	10	3213
20	Douban	142.07	13	30	99	28	116	56	112
21	Vipshop	109.35	7	44	121	60	60	0	14
22	Qunar	102.46	10	49	117	28	32	5	2

在国内企业的开源数据中，我们看到阿里巴巴在除 pull\_review\_comment 指标外之外各项指标得分均排名第一，单项得分上，Repo\_count 和 Developer\_count 是其他公司之和，Open\_issue 的数量也比其他公司高出了一个数量级。

AI 是百度开源最鲜明的竞争力，比如国内开源开放的深度学习平台 PaddlePaddle 和自动驾驶平台 Apollo。腾讯 Tencent 在 GitHub 上发布的 192 个 repo，项目范围主要聚集在云原生、大数据、AI、移动开发和 Web 开发五个技术领域。华为在开源领域的投入有目共睹，但从 GitHub 上的数据未能充分展现其实力，中美宏观贸易战的影响在此略见一斑。WeBank 微众银行作为新兴的互联网金融公司，已开源 27 个项目，44 个 repo，覆盖人工智能、区块链、云计算和大数据等领域，也是公司早期就建立开源办公室的金融公司。滴滴开源势头很猛，成立开源委员会，不仅积极参与业界开源项目，还在 GitHub 上积极开放出 38 个 repo。Deepin 的桌面环境 DDE 受到国内外用户喜爱，软件生态除了自研还在其应用商店中收录了上千款桌面应用，作为国内民间出品的 Linux 操作系统，还是可圈可点的。

PingCAP 在 2020 年下半年宣布完成 2.7 亿美元的 D 轮融资，创造了全球数据库历史新的里程碑，同样，PingCAP 今天在开源方面的表现也是非常亮眼，已经超越百度跃居排行榜第二位。其中，pull\_review\_comment 的数量更是超过了阿里，但开发者的数量却不足阿里的 1/10，可见 PingCAP 的开源社区非常的活跃。

有赞的排名上升得非常快，这可能得益于其开源项目 youzan/vant 的优秀表现，该项目是轻量级的移动 UI 组件。

值得注意的是 Bilibili、douban 和掘金等以用户生产内容（User Generate Content）为主的新兴社交媒体企业也在积极使用开源技术。

可以看出，中国开源领军企业近年来均在不断加大开源社区生态的投入和建设，互联网、操作系统、社交网站、金融、云计算和电商等各个领域的企业都积极参与，呈现出百花齐放的开源盛况。

## 2.4 Apache 软件基金会之源自中国的开源项目

Apache 软件基金会（ASF）成立于 1999 年，基金会致力于帮助独立个体和组织去理解开源是如何在一个激烈竞争的市场中发挥优势的。其重点不是生产软件，而是指导生产软件的社区。Apache 之道对于开源社区的持久发展具有显著的优势：我们维护的一切都是开源的，所有的用户都可以从中受益。Apache 目前有 14 个源于中国的顶级项目，还有 7 个孵化器项目正在努力成为顶级项目。

2020 年，Apache 软件基金会源自中国地区活跃的开源项目共有 21 个，其中，有 9 个项目入围中国开源项目排行榜 Top50。

Apache 软件基金会下属的中国开源项目中，最活跃的项目是 apache/shardingsphere，Apache ShardingSphere 是一套开源的分布式数据库中间件解决方案组成的生态圈，它由 JDBC、Proxy 和 Sidecar（规划中）这 3 款产品组成，该项目已于 2020 年 4 月 16 日成为 Apache 基金会的顶级项目。

apache/incubator-echarts 和 apache/skywalking 的表现也非常活跃，分别居中国开源排行榜第 10、12 名。Apache ECharts 是一个免费的，功能强大的图表和可视化库；SkyWalking 是一个观测平台和 APM 工具，它可以选择性的与 Service Mesh 协同工作，为微服务、云原生和基于容器的应用提供自动的指标，目前为阿里、华为、腾讯等国内大厂提供监测服务。ECharts 和 SkyWalking 同样也是 Apache 软件基金会的顶级项目。

从这些数据，我们看到，Apache 作为世界上最活跃的基金会之一，对中国开源社区的贡献是巨大的。一方面，Apache 作为顶级基金会，其优秀的开源项目和和谐的社区氛围吸引着越来越多的中国开源项目加入其中；另一方面，越来越多中国开源项目的参与也为 Apache 社区中加入了越来越多的中国色彩，期待中国开源项目早日冲入世界开源项目 Top10!

### 【专家点评】

**陈阳：**从时间这个维度，我们看到中国开源从早期的跟随者、参与者，逐渐变成影响者和创作者，中国开源的力量正在崛起。中国开始涌现出大量开源项目的创作者，中国企业的开源项目开始捐献给基金会孵化，中国开始定义自己的开源许可证协议（木兰宽松协议）并被 OSI 接受，中国开始成立中国的开源基金会（开放原子基金会），这些都是中国开源开始成熟的很重要的标志。

**刘天栋：**2015 年开源社与 ASF 联合举办 Apache 中国路演，当时在 ASF 里，源自中国的开源项目只有 3 个（Ebay 中国研究院的 Kylin、Eagle、Griffin），而在 6 年时间内，就成长

到了 21 个（其中由 ASF 孵化器毕业成为顶级项目达到了 14 个）。筚路蓝缕，酸甜苦辣点滴在心头。立足中国、贡献全球的愿景正在路途中！

## 2.5 中国顶级开源项目访

### 马艳军（百度深度学习技术平台部高级总监，PaddlePaddle/Paddle）

很高兴看到 PaddlePaddle（飞桨）在 2020 年延续了 2019 年在中国活跃项目排行榜上高度活跃。

深度学习框架是人工智能开源生态的核心，技术复杂度非常高，同时需要结合应用用慢工持续打磨。飞桨作为我国最早开源、功能最为完备的深度学习平台，一直秉持开放透明的理念进行开源开发。飞桨结合 AI 产业发展需要，在框架整体设计以及全流程开发工具方面保持了前瞻性整体设计，保持对工程质量的极致追求，并通过社区机制进行质量的有效保障，凭借项目质量在广大 AI 开发者中形成了良好的口碑。

飞桨一直非常关注开发者对社区的贡献和认同感。飞桨目前除了已经有 5,000 多位开源开发者通过 PR 或者提 issue 的方式做贡献之外，还鼓励更多开发者通过攻略、社群交流其他方式参与贡献。所有开发者在使用过程中的问题都会得到社区的快速响应，也都纳入到新版本的发布中来，通过这样持续闭环的打磨，项目质量得到保障和持续提升，更好地满足广大 AI 开发者的使用需求，也让广大开发者形成了更强的社区认同感。

深度学习框架在 AI 技术栈中处于承上启下的位置，向下对接芯片，向上承接应用，与芯片广泛适配、深度融合优化至关重要。因此飞桨跟硬件厂商建立了紧密合作关系，很多芯片厂商直接在飞桨社区贡献代码，为飞桨生态发展做了大量贡献。飞桨也持续与各大开源组织、AI 社区深度合作，并得到启智（OpenI）社区的支持成为其中重要一员。通过 PaddlePaddle Developer Experts（PPDE）、SIG 等组织形式，推动与开发者的深入交流互动，与社区开发者共同成长。

### 吴晟（Apache/skywalking）

Apache SkyWalking 今年呈现爆发式的发展，语言探针覆盖了所有主流编程语言，Java、.NetCore、Golang、PHP、NodeJS、Python、C++、LUA for Nginx 全面铺开。Star 数量、贡献者数量、PR 数量等常见社区活跃度指标，都相比 2019 年翻了一倍。

SkyWalking 用户厂家覆盖国内几乎所有的大厂，正在形成自己的标准体系。各大云厂商，阿里云和腾讯云的云 APM 服务都全面支持 SkyWalking 传输协议。同时，SkyWalking 也完成了和 OpenTelemetry、Prometheus、OpenCensus 等主要监控生态的无缝对接。

今年的 SkyWalking 社区和暑期 2020 活动的合作十分成功，孵化了 2 位在线研究生成为正式 Committer。活动中这两位学生的表现，展现了学生团体在顶级项目中的潜力。这是相对于 2019 年，全新的面貌。也展现了系统性的学生孵化计划的价值和意义。希望在今后，我们能看到包括已经发布的暑期 2021 之外，更多更好的学生定向孵化和合作计划。

在全球化的过程中，在 Apache 厂商中立的框架下，SkyWalking 项目在跨地域、跨时区异步多元化的协作模式中顺利前行。SkyWalking 项目已经成为多家国内外公司（亚洲、欧洲、北美）的商业化项目的核心组件，更多的职业开发者加入了项目。为项目生态带来了巨大的活性

和迭代速度。SkyWalking 在按照自己的节奏，和国内外的商业化厂商、个人开发者、公司二次开发团队一起，迅速地成熟和壮大。

## 黄东旭 (PingCAP CTO, pingcap/tidb, tikv/tikv)

TiDB 今年社区运营的重点可以分为：用户、开发者两块。从业务角度来看，可以分为项目打磨和完善，人才的培养和生态建设，用户场景的挖掘和业务成功推进。所以，运营一个好的开源社区，最重要的是以人为中心，与他们共同成长。

对于 TiDB 社区来说，过去的一年发展还算迅速，去年 TiDB 4.0 版本正式推出了 TiFlash 列式引擎。在 Real-time HTAP 的道路上迈出了坚实的一步。从我们的数据上来看，接近 1/3 的 4.0 集群用上了 TiFlash。说明直接在 TP 的数据上作实时的数据洞察是普遍的需求。

其中可以感受到国内的开源氛围及开发者的活跃度也在逐步提升，我们从 2019 年底的 500+ contributor 到 2020 年底数字已经达到了 1,200+，大家对于开源项目的参与度与理解也越来越深，越来越多的开发者在 TiDB 社区汇聚，通过 TiDB 连接行业上下游，共享创新成果。从 TiDB 的运营角度来看，开源社区的繁荣首先有一些基本的原则：透明、公开、分享等等，举个例子：我们的所有讨论的文档、发展方向、投票、选举这些都是公开的、透明的，所有社区的贡献者都可以参与进来；其次，我们也制定了一些基础的社区治理的规则、架构，在基础设施方面也做了一些自动化的 Bot 服务，帮助更多的开发者在社区获得更好的体验和感受；最后再通过一些公开透明的激励和反馈，吸引更多的开发者能够积极地参与进来。

## 3、案例分析——ASF

### 3.1 简介

Apache 软件基金会 (ASF) 成立于 1999 年，是一个依据 501(c)(3) 在美国成立的非营利性公共慈善组织。基金会致力于：

- 提供基础设施：为开源项目提供硬件，沟通交流以及项目治理等方面的基础设施；
- 提供用于捐赠的法律实体：建立一个独立的法人实体，公司和个人可以向其捐赠资源，并确保这些资源将用于公共利益；
- 提供法律保护框架：为个人志愿者提供法律保护框架，避免针对基金会项目的法律诉讼；
- 提供 Apache 商标保护：保护其软件产品的“Apache”商标，防止被其他组织滥用。

Apache 软件基金会的使命是为公共利益而提供软件。基金会帮助独立个体和组织去理解开源是如何在一个激烈竞争的市场中发挥优势的。其重点不是生产软件，而是指导生产软件的社区。被称为“Apache 之道”的精英管理流程，超过 800 位个人会员和 7,000 位提交者成功合作开发了免费的企业级软件，使全球数百万用户受益。Apache 是所有人的 Apache。

### 3.2 阿帕奇之道

“Apache 之道”是 ASF 制定的一套行为或惯例，初步通过专注于稳定的治理并鼓励新的贡献者来促进长期成功的项目。所有的 Apache 项目都必须遵循以下基本原则：

- 一个健康、多样、包容的社区，可以促进项目不断增长，可持续发展。社区大于代码：ASF 一直坚信好的软件是由强大的社区重建出来的。
- 挣得权威：人人都有机会参与，他们的影响力是基于公开获得的功绩——即他们为社区做出的贡献。功绩属于个人，不会过期，不受就业状况或雇主的影响，同时不可转让。

- ASF 的扁平化结构：在 Apache 社区里大家相互尊重，角色平等，每个人的投票具有同等的分量，贡献者也是在志愿者的基础上与其他人享受一样的权利（即使某个组织为他们在 Apache 代码上的工作提供报酬）。
- 大多数 Apache 邮件列表都是存档的，并且可以公开访问，以确保自动协作，这些是一个全球分布的社区所必须的条件。
- 由于无法在任何时候都达成完全共识，此时就可能需要传统而具有约束力的投票或其他协调方式，以帮助消除决策障碍。
- 负责任的监督：ASF 治理模式以信任和授权监督为基础，让项目自治并直接向董事会提供报告。Apache 提交者们互相帮助：相互评估提交，采用强制性安全措施，确保许可合规性，以及保护 Apache 品牌和整个社区免于受到伤害。

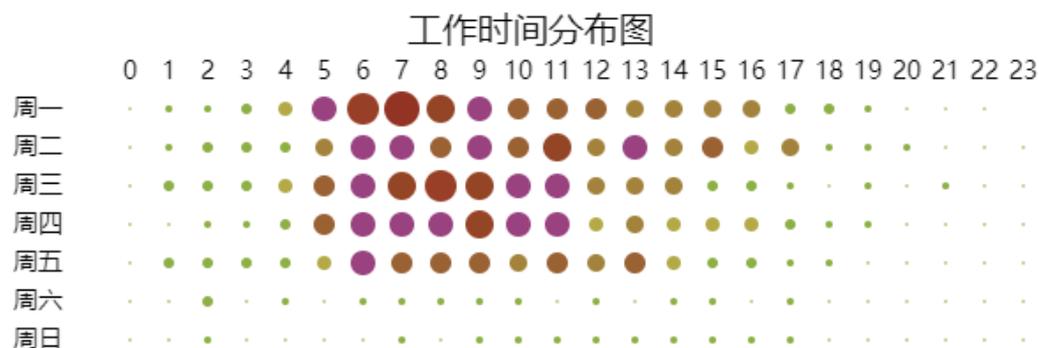
### 3.3 数据分析

我们计算了所有源自中国的共计 21 个 ASF 项目仓库的活动情况，数据如下。

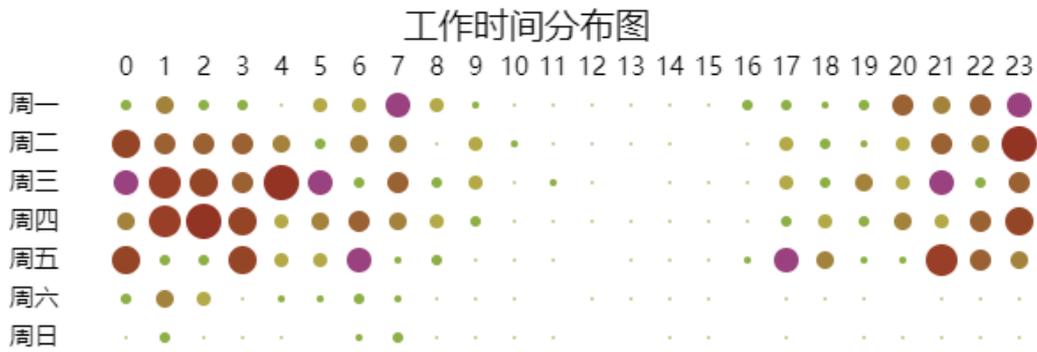
#	name	language	activity	developer_count	issue_comment	open_issue	open_pull	pull_review_comment	merge_pull
1	apache/shardingsphere	Java	2858.72	786	9332	1581	3234	1851	2990
2	apache/incubator-echarts	JavaScript	2307.44	1183	7622	962	301	328	230
3	apache/skywalking	Java	1958.64	525	6904	892	838	3528	725
4	apache/incubator-dolphinscheduler	Java	1929.41	458	9439	1138	1440	722	1135
5	apache/dubbo	Java	1573.76	668	2537	584	558	165	329
6	apache/apisix	Lua	1539.91	300	5874	1030	1000	3363	837
7	apache/incubator-doris	C++	1221.35	149	2327	1055	1410	3543	1227
8	apache/hadoop-ozone	Java	1156.23	70	4270	0	1343	4547	1066
9	apache/incubator-iotdb	Java	996.1	86	4029	116	1485	3467	1268
10	apache/carbondata	Scala	853.36	50	8906	0	518	4767	1
11	apache/rocketmq	Java	825.41	280	1965	396	254	205	106
12	apache/servicecomb-java-chassis	Java	491.5	126	1393	272	326	330	308
13	apache/incubator-brpc	C++	451	170	847	176	70	109	37
14	apache/kyljin	Java	416.18	60	1158	0	495	252	374
15	apache/incubator-weex	C++	291.44	146	449	100	46	8	32
16	XiaoMi/pegasus	C++	222.25	18	161	54	152	1065	137
17	apache/incubator-tubemq	Java	214.9	26	251	0	340	82	309
18	apache/incubator-teaclave	Rust	144.51	24	290	55	180	38	177
19	apache/griffin	Java	48.62	15	114	0	21	20	0
20	apache/hawq	C	11.02	5	13	0	3	1	2
21	apache/eagle	Java	2.65	1	1	0	2	0	0

我们统计了所有源自中国的 ASF 项目仓库的工作时间情况，并为每个仓库绘制了工作时间分布图。在此，我们选择三个工作时间分布有明显特征的项目进行简单分析。

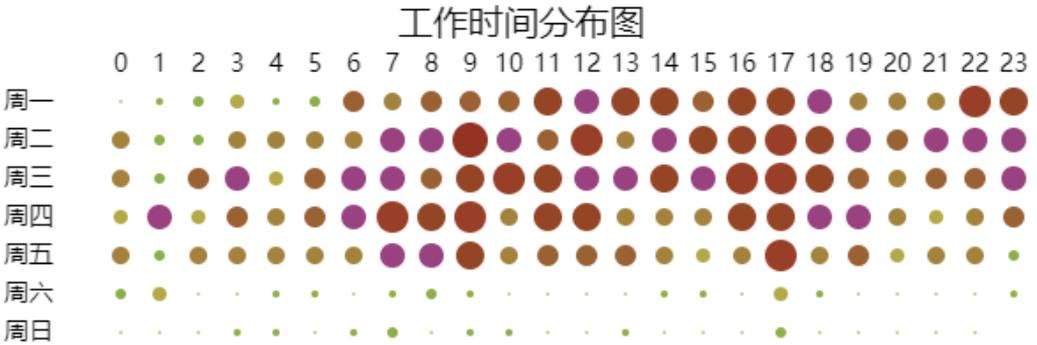
- apache/carbondata, Apache CarbonData 是一种新的融合存储解决方案，利用先进的列式存储、索引、压缩和编码技术提高计算效率，从而加快查询速度，其查询速度比 Petabytes 数据快一个数量级。



- apache/incubator-teaclave, 一个开源的通用安全计算平台，使对隐私敏感的数据的计算安全且简单。



- apache/hadoop-ozone, Ozone 是适用于 Hadoop 的可扩展、冗余和分布式对象存储。



在上述三个项目中，carbondata 明显不在 UTC+8 时区，而在 UTC+4、5 时区，这是因为其主要维护者 Ravindra Pesala 是印度人，与时区匹配；teaclave 的时间分布则完全是美国时间，中国白天几乎都是静默期，虽然是百度捐献的项目，但其维护者 Mingshen Sun 其实位于美国西海岸，而非国内；相比而言，hadoop-ozone 明显是全球化做的最好的一个项目。

## 4、案例分析——CNCF

### 4.1 简介

CNCF 全称为云原生计算基金会（Cloud Native Computing Foundation），是 Linux 基金会（Linux Foundation）的一部分。CNCF 托管着全球技术基础架构的关键组件，汇集了全球顶级开发商、最终用户和供应商。

### 4.2 提案流程

CNCF 对项目提案流程制定了治理政策，提案流程适用于现有要加入 CNCF 的项目和在 CNCF 内形成的新项目。

- 沙箱（sandbox）：所有的例外情况（包括拒绝）均由 TOC（全称为 Technical Oversight Committee，即技术监督委员会的缩写，主要向云原生社区提供技术领导）来处理。项目被拒绝时，可能是“目前不合适”这种情况，并且可以鼓励该项目在解决问题后重新申请。整个过程所需的时间并不固定。目前该阶段有 Artifact Hub、Backstage 等 44 个项目。
- 孵化（incubating）：同沙箱流程，所有的例外情况（包括拒绝）均由 TOC 来处理。目前该阶段的项目有 Argo、Buildpacks、CloudEvents、CNI、Contour、Cortex、CRI-O、Dragonfly、Falco、gRPC、KubeEdge、Linkerd、NATS、Notary、Open Policy Agent、OpenTracing、Operator Framework、SPIFFE、SPIRE、Thanos 等。

- 毕业 (graduated) : 包括提交毕业提案模板、TOC 成员开始为期两周的在 TOC 邮件列表公开评论和 TOC 投票三个步骤。目前该阶段的项目有containerd、CoreDNS、Envoy、etcd、Fluentd、Harbor、Helm、Jaeger、Kubernetes、Prometheus、Rook、TiKV、TUF、Vitess 等。

CNCF 托管已毕业，正在孵化和沙盒中的项目。尽管 CNCF 为所有项目提供了一套共享服务，但并未为沙盒项目提供实质性的营销服务，因为它们是属于早期阶段的项目，需要轻型中立，项目才能自然生长。与孵化和毕业的项目同行相比，沙盒项目的项目服务优先级较低。

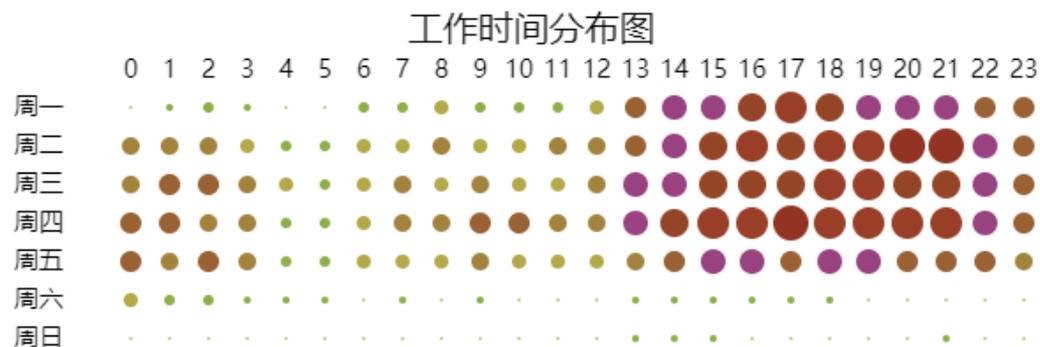
### 4.3 数据分析

我们计算了所有 CNCF 已毕业和正在孵化的项目仓库的活动情况，数据如下。

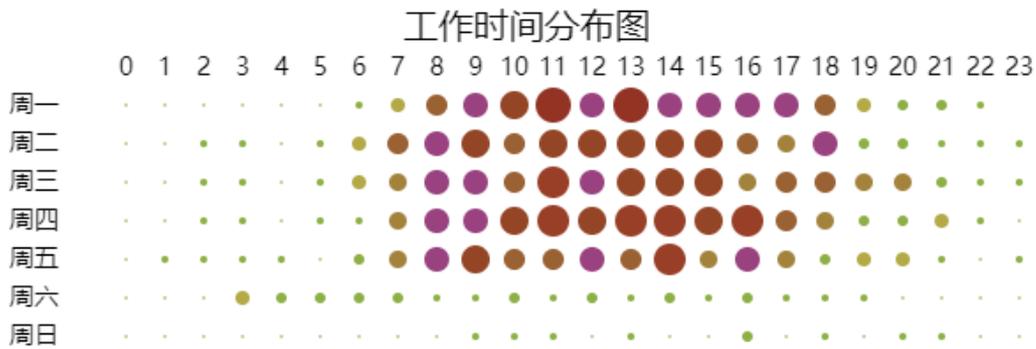
#	name	language	activity	developer_count	issue_comment	open_issue	open_pull	pull_review_comment	merge_pul
1	kubernetes/kubernetes	Go	19878.4	6111	239791	3669	6869	31345	4817
2	envoyproxy/envoy	C++	5859.13	1114	23779	1659	3180	23119	2625
3	helm/helm	Go	4086.31	2047	10369	932	622	1043	421
4	grpc/grpc	C++	3257.29	1194	9952	909	2219	3756	1782
5	goharbor/harbor	Go	2907.22	1153	8621	1751	1529	2618	1294
6	prometheus/prometheus	Go	2564.29	782	9889	541	1032	5790	736
7	rook/rook	Go	2526.45	717	6935	757	1438	7075	1288
8	argoproj/argo	Go	2489.92	738	10242	1406	1356	3742	1126
9	tikv/tikv	Rust	2184.8	259	17964	987	2023	5588	1531
10	thanos-io/thanos	Go	2122.38	576	7071	637	985	4930	784
11	operator-framework/operator-sdk	Go	2019.98	487	6808	686	1213	7894	1019
12	etcd-io/etcd	Go	1568.26	617	4293	313	539	788	389
13	cri-o/cri-o	Go	1448.69	256	36343	194	1210	2069	952
14	linkerd/linkerd2	Go	1433.24	396	4976	570	843	2513	725
15	cortexproject/cortex	Go	1371.17	217	3845	448	1206	6184	1112
16	containerd/containerd	Go	1335.75	413	5020	195	684	999	572
17	kubeedge/kubeedge	Go	1187.46	255	6175	357	670	1673	511
18	jaegertracing/jaeger	Go	1083.09	324	3081	306	338	1942	280
19	vitessio/vitess	Go	1004.34	160	2150	435	1064	2132	911
20	open-policy-agent/opa	Go	913.7	257	2201	450	530	1153	474
21	coredns/coredns	Go	912.9	316	3041	236	489	549	358
22	projectcontour/contour	Go	833.89	182	3608	437	684	2113	623
23	falcosecurity/falco	C++	744.34	253	4132	229	276	425	231
24	spiffe/spire	Go	568.48	73	1139	222	450	1697	409
25	fluent/fluentd	Ruby	535.25	232	1043	184	171	236	164
26	buildpacks/pack	Go	505.06	97	1322	229	311	1217	269
27	nats-io/nats-server	Go	462.93	106	1108	130	390	1920	360
28	dragonflyoss/Dragonfly	Go	403.02	92	1630	125	191	358	148
29	cloudevents/spec	Shell	385.75	76	750	72	110	780	91
30	theupdateframework/tuf	Python	237.41	29	1078	113	171	664	132
31	containernetworking/cni	Go	128.55	44	248	21	25	56	15
32	theupdateframework/notary	Go	112.35	51	185	25	22	32	9
33	opentracing/opentracing-go	Go	19.95	10	21	1	3	6	3

我们统计了 CNCF 项目仓库的工作时间情况，为每个仓库绘制而成的图表如下。

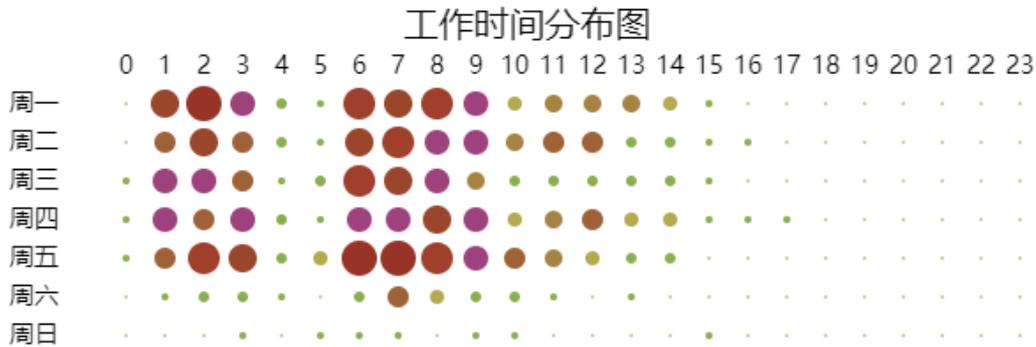
- kubernetes/kubernetes, Kubernetes 是用于自动部署，扩展和管理容器化应用程序的开源系统。它将组成应用程序的容器组合成逻辑单元，以便于管理和服务发现。



- thanos-io/thanos, Thanos 是一组组件，可以组成具有长期存储功能的高可用性 Prometheus 设置。其主要目标是简化操作，保留 Prometheus 的可靠性。



- kubeedge/kubeedge



上述三个项目的开发者工作时间分布图，都具有明显的时区聚集性。其中，kubernetes 的开发者大部分位于 UTC-5 时区附近，即美洲地区，而 thanos 的开发者中则大部分是欧洲的开发者，kubeedge 的开发者则以亚太地区开发者为主。在 kubeedge 的工作时间分布中，我们可以发现，开发者们有午休的习惯，在 UTC 时间 4-5 点，当地时间 12-13 点左右，开发者的工作量锐减。

## 5、案例分析——LF AI&Data

### 5.1 简介

LF AI&Data 是 Linux Foundation 下的一个综合型基金，支持人工智能、机器学习、深度学习和数据方面的开源创新。创建 LF AI&Data 的目的是支持开源人工智能、机器学习、深度学习和数据，并创建一个可持续的开源人工智能生态系统，使得能够使用开源技术轻松地创建人工智能和数据产品与服务。它除了一些扶持性服务，还包括成员资格和资金管理、生态系统发展、法律支持、公关/营销/沟通、活动支持和合规扫描，以及为多样化和蓬勃发展的社区中的开放发展项目提供支持。

目前，已经从 LF AI&Data 基金会毕业的项目有 [Acumos](#) [Angel-ML](#) [Egeria](#) [Horovod](#) [ONNX](#)。

正在孵化的项目有 [Adlik](#)、[Adversarial Robustness Toolkit](#)、[AI Explainability 360 Toolkit](#)、[AI Fairness 360 Toolkit](#)、[Amundsen](#)、[DataPractices](#)、[DELTA](#)、[Elastic Deep Learning \(EDL\)](#)、[Feast](#)、[ForestFlow](#)、[JanusGraph](#)、[Ludwig](#)、[Marquez](#)、[Milvus](#)、[NNStreamer](#)、[OpenD](#) [S4All](#)、[Pyro](#)、[SOAJS](#)、[sparklyr](#) 等。

### 5.2 提案流程

要想进入 LF AI&Data 基金会，需要一个提案。对于试图进入 LF AI&Data 基金会的现有项目，以及将在 LF AI&Data 基金会内形成的新项目，项目提案过程都是相同的。

项目必须通过 GitHub 提出，并向 [info@lfaidata.foundation](mailto:info@lfaidata.foundation) 发送主题为“PROPOSAL [项目名称]”的电子邮件来通知 LF AI&Data 成员。

提交项目提案书后，批准过程分为四个步骤。

- 第 1 步：
  - 项目根据提供的模板准备提案，并通过 GitHub 提交，并附上一封简短的电子邮件至 [info@lfaifoundation](mailto:info@lfaifoundation)，主题是“PROPOSAL [项目名称]”。
  - 由公司代表的项目必须签署并提交商标和帐户转让协议的副本；该协议的目标是将项目商标的所有权和管理权转移到 Linux 基金会的项目帐户中。
  - LF AI&Data 成员会通知项目，提案是否准备好提交给 TAC，并将与项目一起完成提案提交、审查和投票。
- 第 2 步：
  - 该项目将在每两周一次的会议上通过电话会议向 TAC 介绍情况。
  - 演讲者将有 45 分钟的时间发言，阐述相关信息。
  - 演示文稿文件必须在预定的电话会议召开前至少 3 个工作日通过 [info@lfaidata.foundation](mailto:info@lfaidata.foundation) 发送到 LF AI&Data，以确保 TAC 成员有足够的评审时间。
  - TAC 将有 15 分钟的时间与项目代表讨论，然后进行投票，以确定项目的接受程度以及将在 LF AI & Data 中托管项目的阶段。
- 第 3 步：如果该项目在毕业级别获得了 TAC 的批准。
  - 将提案转发给 GB 进行批准。
  - 将邀请项目代表参加下一次 GB 的电话会议，进行简短的演讲（10分钟），然后由 GB 投票以接受毕业级别的项目。
- 第 4 步：公告和加入
  - 通过博客文章或新闻稿宣布 LF AI&Data 的项目托管公告
  - 项目加入——在新项目获得批准后，LF AI&Data 成员将帮助该项目加入基金会，并开始重点领域提供所需的支持。LF AI&Data 成员将为该项目提供每个活动的详细信息和时间安排。

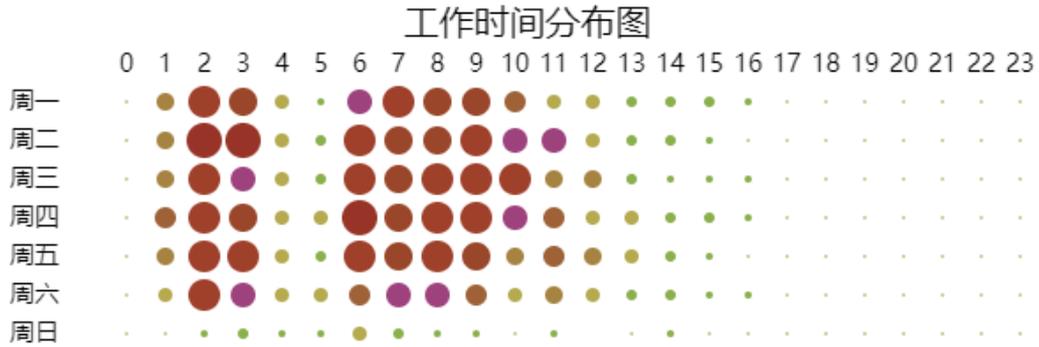
## 5.3 数据分析

我们计算了所有 LF AI&Data 的项目仓库的活动情况，数据如下：

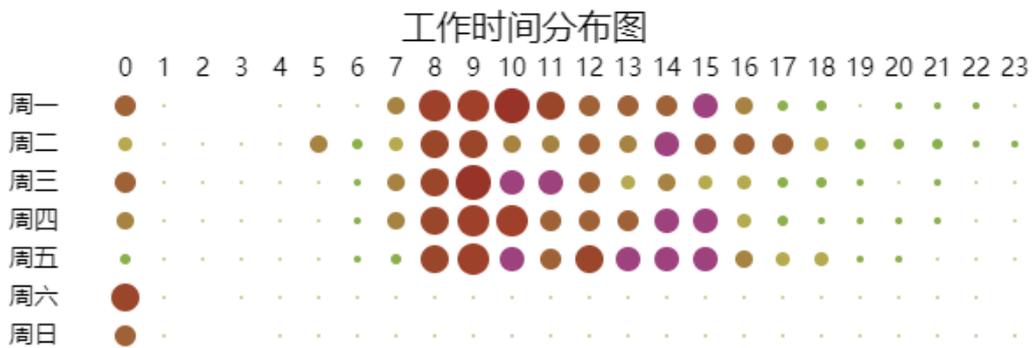
#	name	language	activity	developer_count	issue_comment	open_issue	open_pull	pull_review_comment	merge_pull
1	horovod/horovod	Python	1265.29	446	3579	453	414	947	360
2	milvus-io/milvus	C++	1241.02	240	4716	1551	1870	311	1686
3	onnx/onnx	PureBasic	1139.81	386	2305	217	301	1567	232
4	feast-dev/feast	Java	751.49	94	8129	188	646	2377	558
5	odpi/egeria	Java	620.32	43	4528	711	1298	1071	1047
6	nnstreamer/nnstreamer	C	549.02	39	3984	239	765	2466	735
7	amundsen-io/amundsen	Smarty	502.65	141	1170	212	393	196	361
8	JanusGraph/janusgraph	Java	500.04	136	1231	121	244	970	183
9	uber/ludwig	Python	459.77	119	1786	120	304	620	265
10	IBM/adversarial-robustness-toolbox	Python	448.39	87	884	205	309	1619	269
11	pyro-ppl/pyro	Python	414.42	87	983	93	380	1226	336
12	sparklyr/sparklyr	R	387.19	130	1431	211	417	93	372
13	IBM/AIF360	Python	126.96	36	132	17	38	97	33
14	Adlik/Adlik	C++	122.53	11	820	95	228	103	0
15	MarquezProject/marquez	Java	102.82	15	241	40	145	66	138
16	Angel-ML/angel	Java	99.63	33	114	37	82	0	79
17	IBM/AIX360	Python	74.24	24	109	25	26	0	19
18	odpi/OpenDS4All	HTML	57.43	10	46	3	84	0	78
19	elasticdeephlearning/edl	Python	51.78	5	11	3	86	72	71
20	didi/delta	Python	38.89	8	50	8	27	14	24
21	ForestFlow/ForestFlow	Scala	18.01	2	4	3	22	18	19
22	datadotworld/data-practices-site	CSS	5.74	1	3	0	10	0	0
23	soajs/soajs	JavaScript	2	1	1	0	1	0	0

我们统计了 LF AI&Data 项目仓库的工作时间情况，为每个仓库绘制了图表。在此，我们选择 3 个项目进行分析。

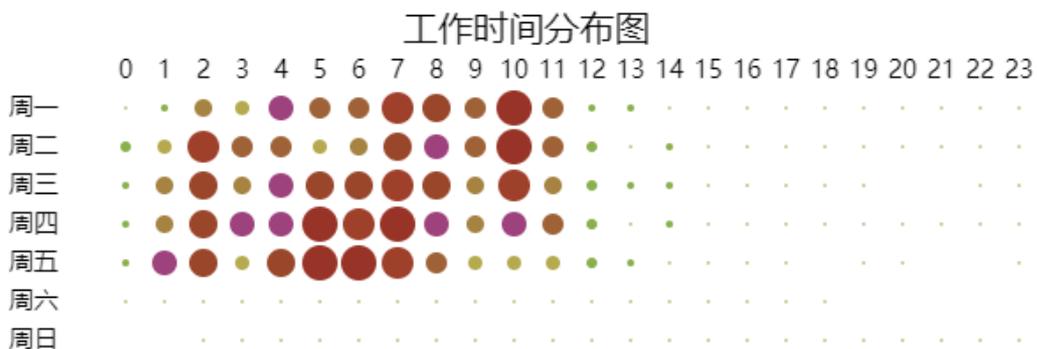
- milvus-io/milvus, Milvus 为海量特征向量的近似最近邻搜索 (ANNS) 而设计。相比 Faiss 和 SPTAG 这样的算子库, Milvus 提供完整的向量数据更新, 索引与查询框架。Milvus 利用 GPU (Nvidia) 进行索引加速与查询加速, 能大幅提高单机性能。



- odpi/egeria



- nnstreamer/nnstreamer, NNStreamer 是一组 Gstreamer 插件, 为采用神经网络模型的 Gstreamer 开发人员和管理神经网络管道及其过滤器的神经网络开发人员提供了方便和高效的支持。



可以发现, 三个项目的开发者工作时间分布情况是存在差异的。milvus-io/milvus 开发者在周一~周六都处于活跃状态, 工作时间主要分布在 UTC+8 时区, 表明该项目的开发者以亚洲开发者居多; odpi/egeria 的开发者主要集中在 UTC+0 时区, 即大多数为欧洲的开发者, 但 0 时无论周末还是工作日, 都有大量事件产生, 几乎可以肯定这个项目使用了自动化协作机器人, 而且很多定时任务是在 0 时执行的; nnstreamer/nnstreamer 开发者的工作时间主要集中在周一~周五, 位于 UTC+8 时区, 表明该项目的开发者以亚洲开发者居多, 且开发者周末是双休的。

## 6、案例分析——Wuhan2020

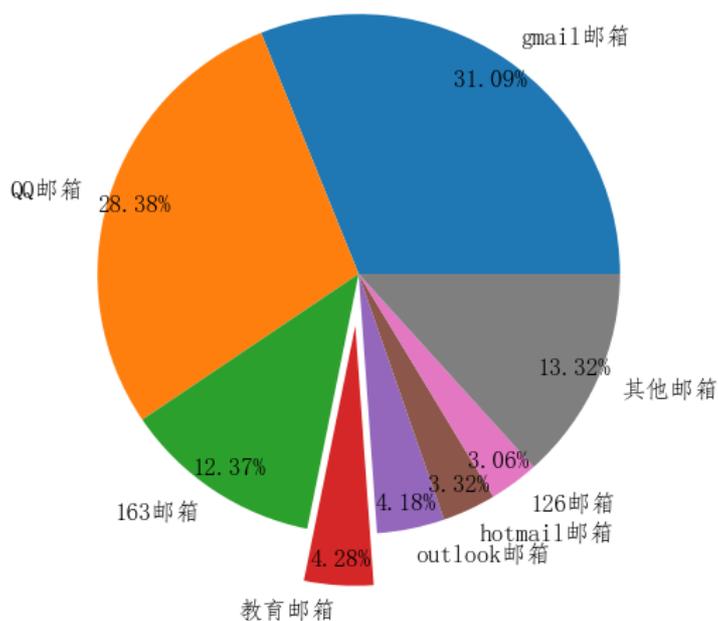
### 6.1 简介

Wuhan2020 是 COVID-19 疫情期间出现的较大规模自组织的代表，也是典型的技术型自组织社区之一。Wuhan2020 成立不到 3 个月，就以开放源代码的方式在互联网上开展志愿服务和合作。在线志愿者合作建立了 5 个网页/网站，在 GitHub 上建立了 23 个代码库，并收集了 4,394 项各种与抗击新冠疫情有关的一手及二手信息。

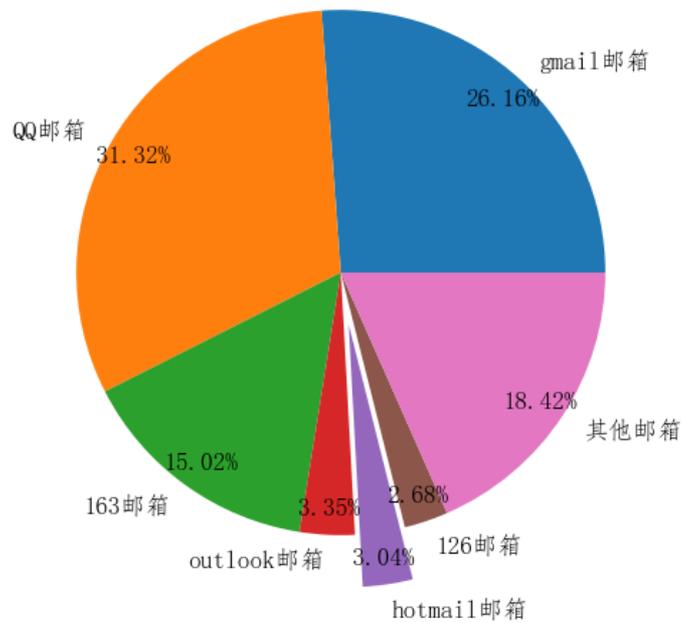
据自组织机构的不完全统计，截至 2020 年 5 月 20 日，Wuhan2020 的 5 个页面/网站的点击量超过 30 万次。其允许公开读取使用的数据信息集合促成了牛津剑桥校友援助会提供的武汉社区及周边地市医院和社区数据对接，硅谷的 Gliding Eagle 团队及美国一家全球医疗信息服务平台的医疗物资援助等案例达成。至此，Wuhan2020 已成为抗击新冠肺炎疫情迄今为止少数几个宣布常态化并完成永久性组织结构建设的自组织代表之一。

### 6.2 成员分析

截至 2020 年 5 月 20 日，Wuhan2020 开源社区通过统计参与者的邮箱确定的参与人数为 4,095 人。随后，通过志愿者邮箱发放志愿者证书，共有 1,942 人领取了 Wuhan2020 志愿者证书。



【社区全部邮箱种类，总数 4,095】

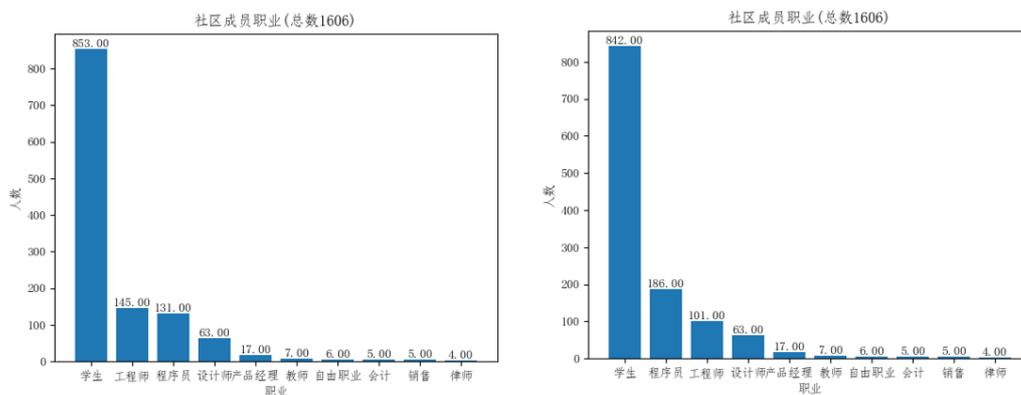


【已领取志愿者证书的邮箱种类，总数 1,942】

可以通过邮箱种类，结合其他数据，粗略估计社区成员的职业以及年龄比例。比如：教育邮箱是参与 Wuhan2020 的老师所使用的，QQ 邮箱使用者多为 95 后。

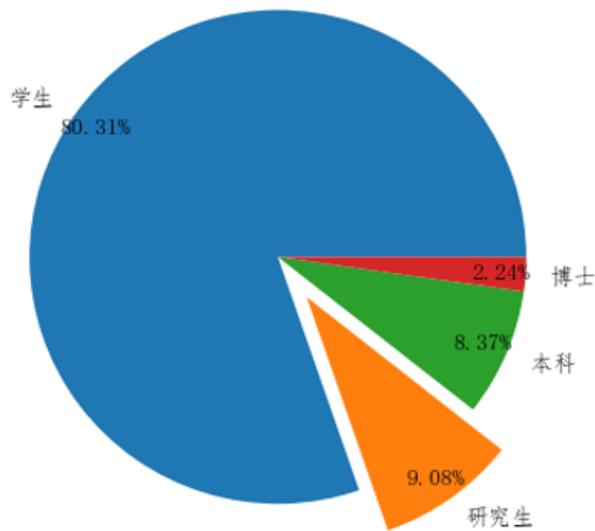
另外发现有 16 个阿里邮箱，5 个流利说邮箱，4 个京东邮箱，以及石墨，普华永道等，这些是企业人员使用的邮箱。

Wuhan2020 项目初期收集了志愿者的职业以及学历信息共计 1,606 条，根据收集的志愿者信息，以这部分志愿者为样本，进行成员分析。



【社区成员职业分析】

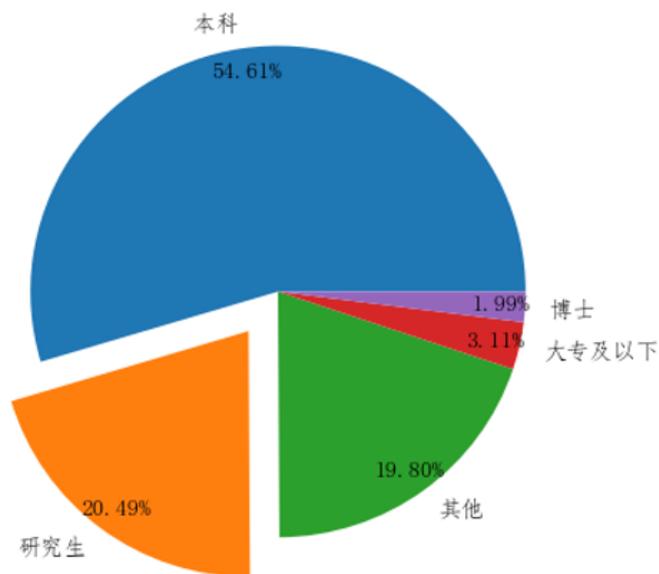
上面两张图的区别在于，横坐标的职业顺序不同，此设计的目的是为了区别有多个职业身份的社区成员。使用左图中某一职业类别减去右图中某一职业类别即为多个身份的成员个数。



【在读学生学历细分，总数 853】

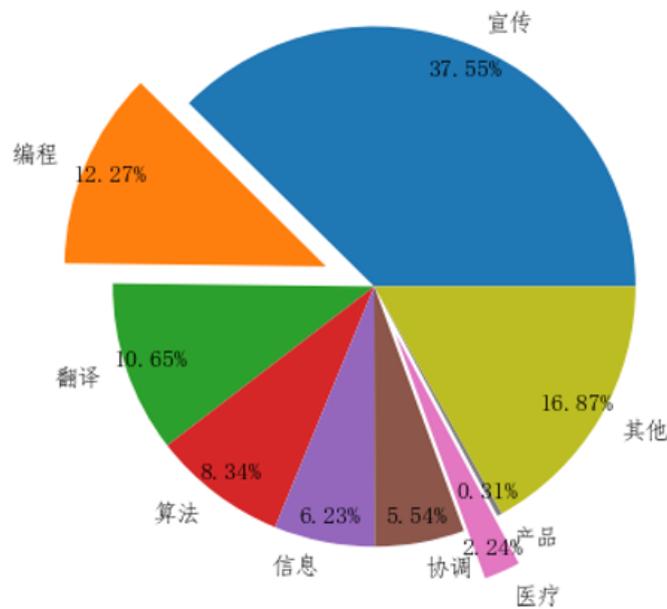
由于在填写职业的时候，更多的人填写是学生，所以只能对填写了除学生之外的具体学历者进行分析，只作为抽样，并不一定具有代表性。

可以进一步，通过社区成员的学历进行细分，包括已经毕业的。



【社区成员学历，总数 1,606】

通过上图可以看到社区成员的学历最多的还是本科学历，可以结合之前的社区成员的职业信息，以及邮箱信息，可以推断出，Wuhan2020 开源社区参与者中，大多数是 95-05 年龄段的学生。使用 Gmail 邮箱的有相当一部分来自于海外，这部分成员领取志愿者证书的比例相对较少等。



【社区成员擅长领域饼状图，总数 1,606】

从社区成员擅长领域的分析结果来看，社区成员在选择自己所擅长领域时，更多地会选择宣传或者设计类的领域，这也解释了为什么 Wuhan2020 宣传组成员严重过剩，而且能有很多宣传组成员在活动结束后仍然留在社区当中。这当然也跟数据来源大多来自于微信而不是 Slack 有关，在 GitHub 和 Slack 当中的更多的是程序员或工程师，会直接参与到项目当中，而不会填写表单。

## 7、总结 & 致谢

本部分所有数据的数据、分析方法与分析结果由 X-lab 开放实验室所支持，内容撰写方面做出贡献的小组成员包括：王皓月、朱香宁，“案例分析——Wuhan2020”部分为 Wuhan2020 开源社区执行长李扬撰写。感谢特邀的开源领域的专家，更要特别感谢积极参与我们调研的每一位朋友。

## 三、Gitee 数据篇

### 1. 概述

Gitee 是开源中国旗下的代码托管平台，至今已有超过 600 万开发者用户。我们对托管在 Gitee 的开源项目进行了统计分析，梳理和解读编程语言、功能分布的变化趋势，分析开发者在 Gitee 参与开源的情况，以期观察国内开源的演进提供一个“本土平台”视角。

### 2. 主要内容与发现

#### 2.1 总体趋势

2020 年总体项目趋势包括：

- 2020 年 Gitee 上托管的代码仓库超过了 **1500 万**。

- Gitee 上的开源项目数量相比 2019 年提升 **192%**，是 2013–2018 年 Gitee 上开源项目的总和。
- 2020 年 Gitee Star 数破千项目的增长率提升 **132%**，是 2019 年 Star 数破千项目总量的 **1.53** 倍。

2020 年总体用户趋势包括：

- 2020 年 Gitee 用户总量超过 **600 万**。
- 2020 年在 Gitee 参与开源的用户数增长了 **162%**，其中 **38%** 是首次参与开源的用户。

## 2.2 总体语言趋势

排名	语言	占比	排名变化
1	Java	38.44%	–
2	JavaScript	13.32%	–
3	PHP	8.82%	↑ 1
4	Python	6.62%	↓ 1
5	C#	5.18%	–
6	Android	4.21%	↑ 4
7	Objective-C	3.79%	↑ 5
8	Go	3.22%	↑ 1
9	C++	2.89%	↓ 3
10	HTML	2.44%	↓ 3

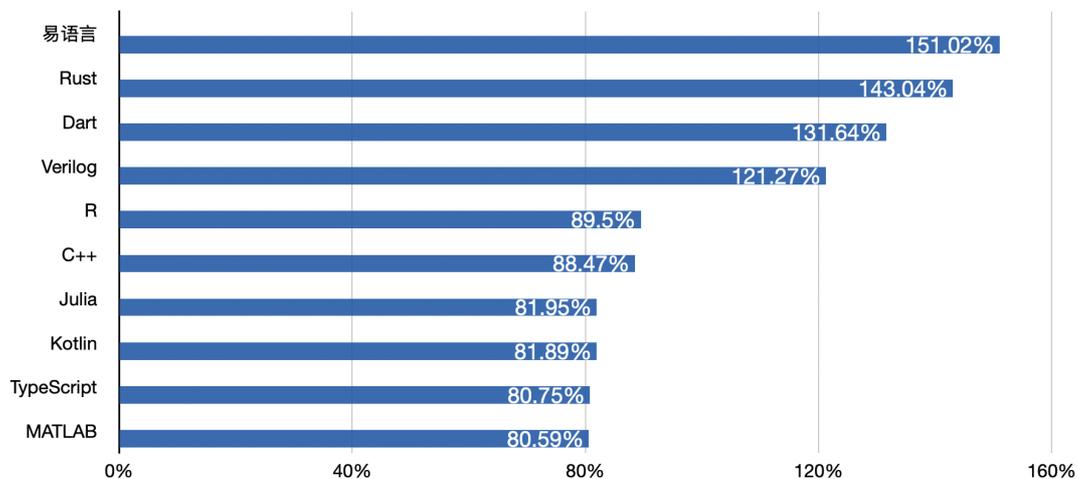
虽然占比较去年（50%）有不小的下跌，但 **Java** 依旧是使用最多的语言（38%），前五名的语言较去年变化较小，**PHP** 超越 **Python** 成为了第三位。

值得一提的是移动端主力语言的占比在 2020 年有较大提升，随着疫情的到来，移动开发似乎又重新被开发者们所关注，跨平台框架在 2020 年的流行也为移动开发注入了新的活力。

## 2.3 增速最快语言

排序	语言	2020增幅
1	易语言	151.02%
2	Rust	143.04%
3	Dart	131.64%
4	Verilog	121.27%
5	R	89.50%
6	C++	88.47%
7	Julia	81.95%

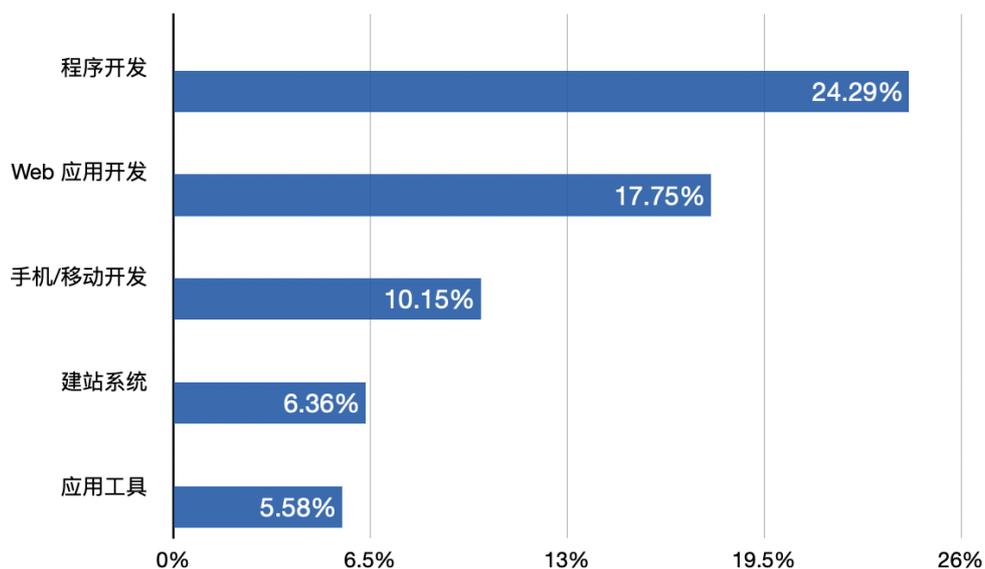
8	Kot lin	81.89%
9	TypeScript	80.75%
10	MATLAB	80.59%



- Rust 和 Dart 连续 2 年保持高速增长。
- 2020 年是操作系统、物联网在国内开源生态中备受瞩目的一年，使得 Verilog 增势强劲。
- 易语言虽登顶增速榜，但其总量与其他语言相比仍非常稀少，与主流编程语言仍有较大的距离。

## 2.4 新增开源项目领域分布

排名	功能分类	占比
1	程序开发	24.29%
2	WEB 应用开发	17.75%
3	手机/移动开发	10.15%
4	建站系统	6.36%
5	应用工具	5.58%



程序开发和 WEB 应用开发依然占据前两位，手机/移动开发领域的开源项目较去年提升了一位排在第三，正如「总体语言趋势」部分所分析，疫情的影响与跨平台开发的流行让移动开发再次受到关注。

## 2.5 新兴影响力项目

排名	项目名	领域	作者	项目地址	2020新增star数 (Gitee年度报告中未展示)
1	BootstrapBlazor	Web 应用开发 - WebUI 组件/框架	Longbow	<a href="https://gitee.com/LongbowEnterprise/BootstrapBlazor">https://gitee.com/LongbowEnterprise/BootstrapBlazor</a>	+2.0k
2	vue-big-screen	大数据 & WebUI 组件/框架	奔跑的面条	<a href="https://gitee.com/MTrun/big-screen-vue-datav">https://gitee.com/MTrun/big-screen-vue-datav</a>	+1.9k
3	AJ-Captcha	程序开发 - 验证码	anji-plus	<a href="https://gitee.com/anji-plus/captcha">https://gitee.com/anji-plus/captcha</a>	+1.8k
4	asyncTool	程序开发 - 常用工具包	京东-平台业务中心	<a href="https://gitee.com/jd-platform-opensource/asyncTool">https://gitee.com/jd-platform-opensource/asyncTool</a>	+1.7k
5	nginxWebUI	DevOps/运维/网管 - 网管工具	陈钊蒙	<a href="https://gitee.com/cym1102/nginxWebUI">https://gitee.com/cym1102/nginxWebUI</a>	+1.7k
6	Plumelog	程序开发 - 日志工具包	陈龙飞	<a href="https://gitee.com/frankchenlong/plumelog">https://gitee.com/frankchenlong/plumelog</a>	+1.6k

新兴影响力项目即 2020 年在 Gitee 获得 Star 数最多的项目。表格中的项目除了 asyncTool 是 2019 年底新建的项目外，其余五个项目均为 2020 年 3-4 月新建。从零开始的开源项目能够在三个季度里获得较高的关注，一方面体现了开源作者们技术水平的精进，越来越多的好项目进入了大家的视野，另一方面也体现出国内关注开源的开发者越来越多，并愿意对优秀的新项目给予肯定。

## 2.6 获得最多捐赠者的项目

项目	作者	领域	项目地址	捐赠者人数
RuoYi	若依	Web 应用开发 - 后台管理框架	<a href="https://gitee.com/y_project/RuoYi">gitee.com/y_project/RuoYi</a>	336
JeeSite 4.2	卓源软件	Web 应用开发 - 后台管理框架	<a href="https://gitee.com/thinkgem/jeesite4">gitee.com/thinkgem/jeesite4</a>	311
Hutool	Hutool	程序开发 - 常用工具包	<a href="https://gitee.com/loool/hutool">gitee.com/loool/hutool</a>	309
FastAdmin	Karson	Web 应用开发 - 后台管理框架	<a href="https://gitee.com/karson/fastadmin">gitee.com/karson/fastadmin</a>	304
zheng	张恕征	Web 应用开发 - 后台管理框架	<a href="https://gitee.com/shuzheng/zheng">gitee.com/shuzheng/zheng</a>	243

pig	lengleg	Web 应用开发 - 后台管理框架	<a href="https://gitee.com/log4j/pig">gitee.com/log4j/pig</a>	153
-----	---------	-------------------	---	-----

2020 年有更多的开发者愿意以「捐赠」的方式来表示自己对于开源项目的支持，从表格中可以看出六个项目中的五个都是后台管理框架类项目。此类整套的后台管理框架更适合企业用户使用，企业使用了更优秀的后台开发框架，得到了快速发展后，也愿意对开源项目捐赠表示感谢，从而形成了「收益于开源-回馈给开源」的良性循环。

## 2.7 GVP – Gitee 指数

GVP (Gitee most Valuable Project) – Gitee 最有价值开源项目计划，是经综合评定优选的开源项目展示平台，截止 2020 年底已有 312 个项目，其中，2020 年有 101 个开源项目入选 GVP。

这些项目的表现，在一定程度上反映了本土优秀开源项目的情况，我们通过 Gitee 指数对其进行了分析。

Gitee 指数是 2018 年底推出的一项功能，从代码活跃度、社区活跃度、团队健康、流行趋势、影响力这 5 个维度对开源项目进行分析，以直观数值的形式来展示对一个开源项目各方位的综合度量。

代码活跃度：与代码提交频次相关

社区活跃度：与项目和用户的 issue、pr 互动相关

团队健康：与团队成员人数和稳定度相关

流行趋势：与项目近期受关注度相关

影响力：与项目的 Star、下载量等社交指标相关

项目名	作者	领域	项目地址	年度 Gitee 指数
MindSpore	MindSpore	深度学习/机器学习框架	<a href="https://gitee.com/mindspore/mindspore">gitee.com/mindspore/mindspore</a>	98.03
kernel	openEuler	Linux 发行版操作系统	<a href="https://gitee.com/openeuler/kernel">gitee.com/openeuler/kernel</a>	86.79
Furion	百小僧	基于.NET 5 的 Web 开发框架	<a href="https://gitee.com/monksoul/Furion">gitee.com/monksoul/Furion</a>	86.46
JEECG BOOT	JEECG开源社区	低代码开发平台	<a href="https://gitee.com/jeecg/jeecg-boot">gitee.com/jeecg/jeecg-boot</a>	86.15
Bootstrap Blazor	Longbow	.NET Web UI 框架	<a href="https://gitee.com/LongbowEnterprise/BootstrapBlazor">gitee.com/LongbowEnterprise/BootstrapBlazor</a>	85.60

## 2.8 案例分析

### (1) OpenHarmony

- 项目简介

OpenHarmony 是开放原子开源基金会 (OpenAtom Foundation) 旗下开源项目，定位是一款面向全场景的开源分布式操作系统。

OpenHarmony 在传统的单设备系统能力的基础上，创造性地提出了基于同一套系统能力、适配多种终端形态的理念，支持多种终端设备上运行，第一个版本支持 128K-128M 设备上运行。

项目地址：<https://gitee.com/openharmony>

- **相关数据（数据选自组织 OpenHarmony）**

鸿蒙是 2020 年 Gitee 上最受关注的项目之一，组织 OpenHarmony 在 Gitee 上的关注数量已经接近 20 万，下表中的数据包含了组织旗下所有的 136 个仓库。

提交次数	Issue 总数量	未解决 Issue	已解决 Issue	PR 总数量	打开 PR 数	已合并 PR 数	已关闭 PR 数	总贡献人数
233818	453	132	321	713	46	458	209	1046

从数据中可以看出，自 2020 年 9 月正式开源后，已经有 1046 名贡献者进行了超过 23 万次的提交，这在全球范围内来说同样是一个惊人的数字。

## (2) 案例分析——RT-Thread

- **项目简介**

RT-Thread 是一个来自中国的开源物联网操作系统，诞生于 2006 年，是一款以开源、中立、社区化发展起来的物联网操作系统。RT-Thread 主要采用 C 语言编写，浅显易懂，且具有方便移植的特性（可快速移植到多种主流 MCU 及模组芯片上）。

项目地址：<https://gitee.com/rththread/rt-thread>

- **相关数据（数据选自仓库 rththread/rt-thread）**

2020 年是物联网备受瞩目的一年，而 RT-Thread 也是国内众多优秀的物联网项目之一，下表中的数据选自其核心仓库 rththread/rt-thread。

提交次数	Issue 总数量	未解决 Issue	已解决 Issue	PR 总数量	打开 PR 数	已合并 PR 数	已关闭 PR 数	总贡献人数
10495	48	36	12	105	7	39	59	409

RT-Thread 并没有设定某个仓库为「主仓库」，而是同时在 Gitee 与 GitHub 接受 Issue 与 PR，这让项目有了更加全面的发展。

## (3) 案例分析——MindSpore

- **项目简介**

MindSpore 是一种适用于端边云场景的新型开源深度学习训练/推理框架。MindSpore 提供了友好的设计和高效的执行，旨在提升数据科学家和算法工程师的开发体验，并为 Ascend AI 处理器提供原生支持，以及软硬件协同优化。

项目地址：<https://gitee.com/mindspore/mindspore>

- **相关数据（数据选自仓库 mindspore/mindspore）**

提交次数	Issue 总数量	未解决 Issue	已解决 Issue	PR 总数量	打开 PR 数	已合并 PR 数	已关闭 PR 数	总贡献人数
16892	4300	695	3605	10897	133	9045	1719	408

作为 Gitee 指数最高的开源项目，MindSpore 始终保持着极高的活跃度。值得关注的是其 Issue 和 PR 的数量，Issue 总数量达到了 4300，而 PR 的数量甚至接近 1.1 万，这个数字对于一个总贡献人数 408 人的团队是非常惊人的。

### 3. 总结 & 致谢

2020 年是国内开源生态快速发展的一年，也是国内开源生态发展过程中意义重大的一年，Gitee 作为本土的代码托管平台，见证了诸多优质国产项目的诞生和崛起，也见证了越来越多的开发者投身开源建设中。国内开源的建设成果已经初见端倪，我们与开发者们接下来要做的就是将其在正确的道路上发展壮大，期待 2021 年成为本土开源崛起的一年。

本部分数据为 Gitee 平台提供，内容撰写方面做出贡献的小组成员包括：刘辰、李泽辰。

## 四、投资篇

国际创业投资及私募资本对开源新创公司及项目的关注与投入已经行之有年，并已获得了极大的回报。而中国国内的商业化开源软件公司获得国内外投资人的关注与投入，正随着源自中国的开源项目的崛起方兴未艾，值得大家关注。

### 1. 全球收入超过一亿美元的商业化开源软件公司（部分列表）

Company	Core Project(s)	OSS License	Company Creation	Project Creation	OSS Origin	VC Raised (M)	Employees	Est. Revenue (M)	Est. Valuation (B)	Exit Value (B)
GitLab	Git	GPL v2	2014	2011	Personal project	\$436	1,300	\$100	\$5.0	
HashiCorp	Many	MPLv2	2012	2010	Personal project	\$350	950	\$150	\$5.1	
Instructure	Canvas	AGPL v3	2008	2011	Company project	\$90	1,420	\$220	\$1.4	\$0.65
JetBrains	IntelliJ	Apache 2.0	2000	2009	Company project	-	800	\$350	\$4.0	
JFrog	Artifactory	AGPL v3	2008	2007	Company project	\$230	460	\$150	\$1.5	
Kaltura	Kaltura	AGPL v3	2006	2009	Company project	\$166.00	480	\$200	\$1.5	
Liberty	Liberty Portal	LGPL v2.1+	2004	2000	Company project		870	\$250	\$2.5	
Magento Commerce	Magento	OSL v3, AFL v3	2007	2008	Company project (Varian)	\$272	700	\$200	\$1.7	\$1.68
Mapbox	Mapbox GL JS	BSD-3	2010	2010	Company project (DevelopmentSeed)	\$275	450	\$100	\$1.5	
Mirantis	OpenStack	Apache 2.0	1999	2010	Academia/Research (NASA)	\$227	500	\$100	\$1.0	
MongoDB (aka 10gen)	MongoDB	SSPL (not OSI)	2007	2009	Spin-out (10gen)	\$311	1,500	\$400	\$9.0	\$1.60
Mozilla Corporation	Firefox	MPLv2	2003	2002	Spin-out (Netscape)	\$22	1,100	\$550	\$5.0	
Microsoft	Microsoft SQL Server	CPAL	2006	2003	Personal project	\$311	1,750	\$700	\$8.0	\$6.50
MySQL AB	MySQL	GPL v2	1995	1995	Company project	\$41	800	\$1,000	\$1.1	\$1.10
Neo4j	Neo4j	GPLv3	2007	2007	Company Project	\$160	300	\$100	\$1.0	
Nicira	Open vSwitch	Apache 2.0	2007	2009	Academia/Research (Stanford)	\$42	100	\$2,000	\$1.3	\$1.26
Odoo	Odoo	LGPL v3	2005	2005	Company project	\$105	950	\$250	\$1.5	
Pentaho	Pentaho	Apache 2.0	2004	2004	Company project	\$75	670	\$150	\$1.0	\$0.60
Pivotal (Now VMware Tanzu)	CloudFoundry	Apache 2.0	2013	2009	Company project (VMware)	\$1,700	2,400	\$800	\$3.5	\$4.00
Puppet Labs	Puppet	Apache 2.0	2005	2005	Company project	\$142	560	\$250	\$2.5	
Red Hat	Linux	GPL v2	1993	1991	Personal project	\$5	13,100	\$3,500	\$34.5	\$34.00
Linden Lab	Second Life	LGPL	1999	2003	Company project	\$19	250	\$100	\$0.5	
Sourcefire	Snort	GPL v2	2001	1998	Personal project	\$40	600	\$500	\$3.0	\$2.70
SugarCRM	SugarCRM	Previously A2.0	2004	2004	Company project	\$146	440	\$175	\$1.0	
SUSE	Linux Kernel	GPL v2	1992	1991	Personal project	-	1,500	\$400	\$2.5	\$2.50
Talend	Talend Data Integration	Apache 2.0	2005	2006	Company project	\$102	1,300	\$200	\$2.0	\$0.55
VA Linux (Geeknet)	Linux	GPL v2	1993	1991	Personal project	\$30	300	\$120	\$0.2	\$0.15
WordPress	WordPress	GPL v2	2010	2003	Personal project	\$291	900	\$200	\$1.0	
						<b>\$10,806</b>	<b>51,470</b>	<b>\$17,165</b>	<b>\$146.6</b>	<b>\$68.49</b>

来源：COSSI: \$100M+ Revenue Commercial Open-Source Software (COSS) Company Index: <https://coss.media/rise-of-the-open-source-ipo/>

### 2. 国内开源项目投资列表（部分）

时间	组织（公司/基金会/个人）	核心开源项目	作者/创始人	最新融资轮次	交易金额	最新投资方
2020年6月	欧若数网	NebulaGraph	叶小萌	Pre-A1轮	800万美元	红点中国 经纬中国
2020年4月	涛思数据	TDengine	陶建辉	A轮	2000万美元	红杉资本中国 GGV纪源资本
2020年1月	上海睿赛德电子科技	RT-Thread	熊谱翔	股权融资	未披露	华强PCB GGV纪源资本
2019年11月	北京易捷思达科技	EasyStack	陈喜伦	D轮	数亿人民币	中国电子
2019年3月	上海硅智科技有限公司	Kylogence	韩卿	C轮	2500万美元	顺为，红点
2018年9月	巨杉数据库	SequoiaDB	王涛	C轮	3000万美元	嘉实投资 启明创投 DCM中国
2018年9月	平凯星辰（北京）PingCAP	TiDB	黄东旭	C轮	5000万美元	晨兴资本 复星集团 云启资本 经纬中国

（来源：CSDN 蒋涛 CODE CHINA -开源趋势发展）

### 【专家点评】

**刘天栋：** 开源不易、创业维艰。如何让开源项目成功地商业化成为开源软件新创公司，国内已经有一些值得大家学习的案例。它山之石可以攻玉，我们推荐大家可以去借鉴如 PingCAP 用心打造开源项目社区，从而提高产品质量、缩短开发周期、凝聚开发者及早期用户向心力的经验。我们期盼未来中国有越来越多的优质开源项目转化为成功的新创开源软件公司。

**陈阳：** 开源呈现出一种出圈的状态。更多的领域开始谈论开源，开始了解开源。近年以来，政府部门开始关注开源，学校开始推广开源的课程和认证，企业们开始考虑成立开源委员会，VC 创投开始垂青开源项目，开源已经成了多轮融资以及上市评估一个非常重要的维度和评价方向。

## 五、全文总结

终于完成了本次中国开源年度报告的数据收集、撰写、专家点评、整理等方面的工作，可谓是一个系统工程，离不开所有贡献者们的密切协作，再次向大家表示感谢。

2020 中国开源年度报告最大的创新莫过于在往年问卷调查的基础上，增加了 GitHub 和 Gitee 两大开源平台的数据分析篇，关于这两块的内容，敬请关注完整的报告：

- 《GitHub 2020 数字年报》，X-lab 开放式实验室，预计 2 月中旬发布
- 《2020 Gitee 开源报告》，Gitee，已正式发布

开源软件已经成为人类数字社会的基石，也是全人类共同努力协作的结晶；我们希望让更多的人认识到开源协作对人类文明的发展所起到的巨大推动作用；谨以此《**中国开源年度报告**》作品形式表达我们对开源事业的热爱！