

开源社区成熟度 研究报告 (2021 年)

中国信息通信研究院云计算与大数据研究所

PingCap

2021 年 7 月

版权声明

本白皮书版权属于中国信息通信研究院和 PingCAP，并受法律保护。转载、摘编或利用其它方式使用本白皮书文字或者观点的，应注明“来源：中国信息通信研究院”。违反上述声明者，编者将追究其相关法律责任。

前 言

开源逐步成为软件领域最重要的发展趋势，开源带来了三个层次的价值，分别是“社会价值、商业价值、创新价值”，开源软件通过激发工程师个体在社群中的创造力，使得大量的个人用户和中小企业可以无门槛使用质量优良的软件产品，带来巨大的社会价值；开源软件公司通过商业模式的创新，形成社区产品和商业化产品的良性互动，推动软件商品化，提升了软件的商业价值；开源的协同方式逐步成为开放创新的引领者，带来创新价值。在这个三大价值方面，开源社区都是基石般的存在，开源社区的成熟度，直接构成社会价值的基础，通过与商业模式融合带来潜在商业价值，随着开源社区的成熟度越来越高，开源社区也成为探索创新模式的支撑。

开源社区是一个由共同身份和集体目标联合起来的群体。一个健康的开源社区有助于开源项目扩大影响力，吸引更多的外部参与者，扩大社区规模，通过“网络效应”营造完整的开源生态。开源社区发展存在不同阶段，如果社区成熟度落后于发展趋势，会导致社区出现“断层”，从而产生大量风险。所以维护者需要度量开源社区所处阶段与成熟度，以保证社区健康可持续性发展。本白皮书聚焦开源社区成熟度，阐述了开源社区成熟度模型与社区构建、度量体系，展望了开源社区的发展趋势。

目 录

一、开源社区概述.....	1
（一）开源社区范畴.....	1
（二）开源社区的重要作用.....	2
二、开源社区构建关键要素.....	3
（一）明确应用场景.....	3
（二）吸引用户.....	4
（三）促进贡献与贡献增长.....	5
（四）建立广泛协作社区.....	6
（五）良性的商业循环.....	7
三、开源社区成熟度模型.....	13
（一）开源社区成熟度曲线与象限图.....	13
（二）数字化场景时代的开源社区发展实践.....	18
四、开源社区度量体系.....	19
（一）社区治理.....	20
（二）社区运营.....	22
（三）社区开发.....	24
（四）基础设施.....	25
五、开源社区未来发展趋势.....	27
（一）基础软件领域开源社区蓬勃发展.....	27
（二）开源社区商业色彩逐渐加深.....	27
（三）最终用户对开源社区的参与更加深入.....	28

附录目录

附录一 中国信通院可信开源社区评估体系.....	29
附录二 Linux 基金会 CHAOSS 项目	30
附录三 CNCF 基金会 CII 最佳实践	31

图 目 录

图 1 开源社区成熟度曲线图	9
图 2 开源社区成熟度象限图	11
图 3 TiDB 飞轮模型	15
图 4 基于开源社区的全球数据库人才生态圈	17
图 5 TiDB 社区运营模式	18
图 6 TiDB 社区治理模式	19
图 7 开源社区度量体系图	20

一、开源社区概述

（一）开源社区范畴

开源社区是一个由共同身份和集体目标联合起来的群体，涉及参与者、项目与基础设施三个重要元素。尽管开源（Open Source）这个概念是 Christine Peterson 于 1998 年才正式提出，但是以开源为理念的社区在很多年前就已经形成。

开源社区由参与者组成。当人们同意共同构建和改进开源软件时，开源社区就形成了。大多数开源软件社区没有正式的成员资格申请或批准流程。通常来说，人们参与项目的方式很简单，比如通过修复软件代码中的错误，为应用程序创建新功能，编写和完善文档，在会议和其他活动中宣传项目等工作。个人参与社区的动机各有不同，一部分参与开源社区是因为软件工作是他们工作的范畴，或者是因为他们的组织雇佣他们来开发软件，又或是因为他们的工作依赖于软件的良好运行；一部分人参与开源社区为了学习和提高个人技能；还有一些人加入是为了社交，或者为了一种认同感和归属感。

开源社区围绕开源项目聚集。开源社区围绕开源项目展开，因此社区发展需要依托一个好的项目，一个有发展潜力的项目可以吸引足够多、足够优秀的人来一同建设。

开源社区依托基础设施协同。开源基础设施涉及代码托管平台、网站、沟通工具等。基础设施影响社区成员的活动方式。如在早期阶段，参与者仅能从文档了解项目，现如今介绍文档已经不仅限于文字形式，参与者还可以通过视频或即时通讯技术直接了解项目的实际情

况。

（二）开源社区的重要作用

开源社区的协作机制激发技术创新。开源社区通过公开透明的方式降低边界成本，对于参与者来讲更易获取项目现有信息及发展轨迹，充分调动个人主观能动性，通过社区协作机制进行思想碰撞，激发技术创新。以 Kubernetes 为例，Kubernetes 以“透明开放”为原则，从 API 到容器中的每一层，都给开发者展示了可扩展的插件机制，鼓励用户通过代码的方式介入每一个阶段。而这一举措在整个容器社区中孵化出了大量基于 Kubernetes API 和扩展接口的二次创新项目，例如 Istio 微服务治理工具、应用部署框架 Operator、云原生存储项目 ROK 等。

开源社区放大工程师红利效应。随着信息化进程加速，各行业进入软件定义的时代，工程师群体数量激增，成效规模化效应，2020 年 GitHub 开发者数量达到 5600 万，预计 2025 年将到达 1 亿。开源社区通过建立社交网络，吸引了潜在人员，有效扩大了工程师群体数量，提升工程师的聚集速度和开发能力。

开源社区推动全球协同发展。开源社区打破单位边界，推动复杂场景与先进技术结合，有效将国内的场景优势和工程师优势与全球的开源运动进行结合，调动全球开发者参与。一方面国内具有充足的开发者群体和使用场景，国内互联网场景在并发性和复杂性上均高于全球平均水平，有利于吸引技术黑客解决具有挑战性的问题，另一方面，

全球开发者在开源社区参与中贡献积极，双方可以通过开源社区共享优秀场景案例，迅速提升项目技术能力。

二、开源社区构建关键要素

（一）明确应用场景

在启动一个新的开源社区或者开源社区规模较小时，社区需要考虑开源项目的发展目标。TiDB 在开源之初的发展方向就确定为打造通用型产品，打造通用产品有利于尽早感知整个系统是否有结构性缺陷。

解决实际问题。开源项目的核心是解决用户问题，需要有效传递给客户解决问题的方式，开源社区要把重点放在如何帮助用户上，用最简洁的语言向项目的潜在用户传递信息。

确定用户群体。开源项目需要明确项目的受众群体，需要明确用户是个人开发者还是企业用户，是否面向特定领域，企业级用户的工作角色，企业级用户是否有商业服务支持能力等信息。

构建交流体系。参与度是扩大项目用户群和社区的关键，项目可以通过多种方式与参与者接触，每种方式都需要不同程度的努力，并产生不同的结果。交流方式分为“低接触”、“中等接触”和“高接触”。低接触的方法表示潜在用户和项目社区之间通过间接形式互动，例如网站、文档等；中等接触的方法表示参与者与社区有了直接交流，例如论坛、沙龙活动、邮件等；而高接触的方法表示一对一或一对多的沟通，例如电话、视频会议等。低接触活动有助于提高人们对项目的认识，确保项目网站和其他材料清楚地传达项目的功能，如何帮助用

户，以及贡献者如何尝试并快速参与，因此是至关重要的。中等程度的接触活动非常有助于围绕项目创建“重心”，不仅可以与用户进行交流，还可以让这些用户相互帮助。而高接触活动对于与社区关键用户建立关系、收集社区案例研究、帮助更大的群体提高工作效率并成为项目的倡导者均有着明显的帮助。

分析竞品项目。通过评估一个项目竞品，可以充分分析开源项目的潜在市场。在评估过程中需要梳理清楚主要竞品清单及市场份额；判断竞品的商业战略；分析各竞品间的优缺点；找出竞品对于用户的吸引点。分析竞品可以帮助社区在早期就确定自身发展方向，有助于项目考虑功能优先级并决定如何联系潜在用户。

明确发展目标。开源项目需要明确发展目标，是为了直接盈利还是扩大影响力，是影响市场份额还是了解用户需求。了解项目开源的原因将有助于明确实现目标所需的资源，并在未来的工程、产品和商业模式中保持一致。

（二）吸引用户

吸引用户是扩大贡献者的前提。一个开源社区人数越多，人们参与的兴趣就越大，对项目的贡献量也越大。人口越多样化，社区的规模就越大。很少有专家级别的贡献者突然成为社区贡献者。更多的人首先成为开源项目的用户，然后参与更广泛的用户社区，最后成为一个初始贡献者。

多样性社区更易吸引用户。多样性和包容性是为创新、弹性和长期可持续性提供力量的重要指标。当参与者看到社区有正式的行为准

则，并且领导层拥有不同的企业属性时，参与开源社区安全度极大提升。

优化沟通运营方式吸引用户。随着开源项目的数量与日俱增，项目用来与用户联系以及协调开发人员和贡献者的通信工具也在不断发展。许多大型项目仅通过电子邮件列表进行开发讨论，有关功能设计和项目改进的讨论通过电子邮件进行，并包含在项目的邮件列表存档中。如今沟通方式更为多样，许多项目采用了被认为更加用户友好型或易于访问的服务，例如 Slack、微信等。还有一些社区发现，通过为用户提供在线论坛协作的机会，他们可以取得更好的效果，同时开源软件项目也广泛使用了 Twitter、Facebook、抖音等社交媒体服务，与用户和开发者社区进行互动。社交媒体是项目运营和让受众了解项目发展的绝佳工具。

（三）促进贡献与贡献增长

外在动机与内在动机驱动开发者参与项目。参与开源均为“获利”，利益可能是金钱也可能是个人满足。**物质激励推动参与者扩充**，部分参与者出于理想主义或利他主义的原因为开源项目做贡献，但这些通常不是唯一的动力，特别是在有商业支持的项目中，薪酬和其他工作福利更为重要；**社区认可驱动个人参与**，同行业的认可与声誉是个人参与开源的主要驱动力，虽然不能直接获得金钱利益，但获得同行认可和在新技术领域工作的机会同样是重要动力。

社区应具备多样的贡献方式来保证社区健康成长。有一个常见的误解，即只有代码和文档才是对开源项目的“真正”贡献。事实上，

除了这两种方式之外，还有更多的贡献方式，这些贡献对促进项目成长起着至关重要的作用，贡献方式包括但不限于：代码、文档、制定行为准则、创建与改进社区流程、法律支援、公共服务和项目布道等。

形成稳定的贡献门槛与机制。贡献者是一切开源项目的生命线，但是开源项目如果不断地欢迎新的贡献者，可能会导致项目本身发展混乱，甚至停滞不前。因此任何希望长期可持续性发展的开源项目都应该认真思考自身的贡献机制，精心设计的贡献机制使贡献者更容易为项目做出“有效”贡献。所以社区需要展示给新加入的贡献者，什么样的贡献对开源项目与社区最有价值。例如，一些开源项目已经在他们的源代码库中为“首次贡献者”设置了标签，这个方法可以引导新进入的贡献者解决问题和 bug，同时还可以向新加入者介绍项目的运作机制。

创建辅导机制。在开源项目中，导师制度对社区一致性和认同感有累积效应。它可能开始于两个参与者之间，一个人指导另一个人，而受帮助者可以反过来指导其他参与者，累积的连锁反应极大地增强了参与者对社区的认同感与归属感，最终辅导机制可能成为一种“帮助的回力棒”，当你需要帮助或一些建议时，它就会回到你的身边。所以一个成熟的辅导机制有助于促进更具包容性、协作性和创新性的社区。

（四）建立广泛协作社区

国际化社区虽不是构建开源社区的必备因素，但在开源社区建设之初就将国际化考虑进来有利于推动广泛协作，有利于个激发人创新。

提供本地化服务，在国际化的开源社区上进行交流的主要语言为英语。对于不是以英语为母语的用户和贡献者来说，这一事实可能会成为他们的参与社区的巨大障碍。项目可以采取一些行动来帮助那些英语不是第一语言的人更有效地使用与参与项目。如果项目在全球范围内被广泛采用，并且发展到可以用多种语言进行交流的地步，需要确保在项目网站上突出列出相应资源，同时欢迎社区成员提交非英文资源。**寻求本地化志愿者参与**，寻求帮助的一个关键领域是文档资源的本地化。不管参与者在软件开发实践方面的技能水平如何，社区成员都可以通过翻译文档来积极地发展项目并加以改进，从而使更多的人和更多潜在的贡献者能够更容易地访问项目。

（五）良性的商业循环

拥有足够的市场空间。一个开源软件产品如果功能过于全面，市场匹配度可能非常高，导致用户并不需要商业增值的功能，这也意味着没有能够驱动收入增长的空间。因此，在推动开源项目应用的过程中，社区应该仔细考虑将来可能会商业化的功能，与社区功能进行清晰划分。

明确产品商业模式。社区需要明确思考开源软件提供的价值不仅仅在于它本身的功能，还在于它提供的运维和规模化运行服务。常见的商业模式包括：技术支持和服务模式，开放核心模式与通过 SaaS 提供全套软件托管产品的模式等。商业模式的选择取决于产品能为客户提供什么样的价值，以及如何最好地服务客户。同时商业模式并不是唯一，社区可以根据自身产品的实际业务场景，构建多种模式混合

的商业模型。

找到潜在的付费用户。通过关注自身项目开源用户，可以了解哪些行业、组织与个人正在使用项目，项目中哪些功能具备更高的吸引力。随后社区可以针对已收集的信息，进行定量与定性分析，考虑是否针对特定细分领域进行重点关注与定制化服务。

三、开源社区成熟度模型

开源社区发展存在不同阶段，开源社区成熟度模型是对开源社区的发展划分为触发器、发展期、协作期、结晶期和流行期，用来帮助社区跟踪自身动态发展趋势。

（一）开源社区成熟度曲线与象限图

开源社区成熟度曲线是对各种开源社区常见发展模式的图形描述，用来帮助社区跟踪自身动态发展趋势。成熟度曲线的横轴为“时间”，表示一项技术随时间发展经历各个阶段。成熟度曲线的纵轴显示在社区发展过程中价值随时间的膨胀和收缩情况，是由市场对社区的预计未来价值的评估决定的。**开源社区成熟度曲线把开源社区分为了五个阶段**，每个阶段不一定是相互独立与无法跨越的，例如某些开源社区在刚刚开源时就积累了众多资源达到结晶期。

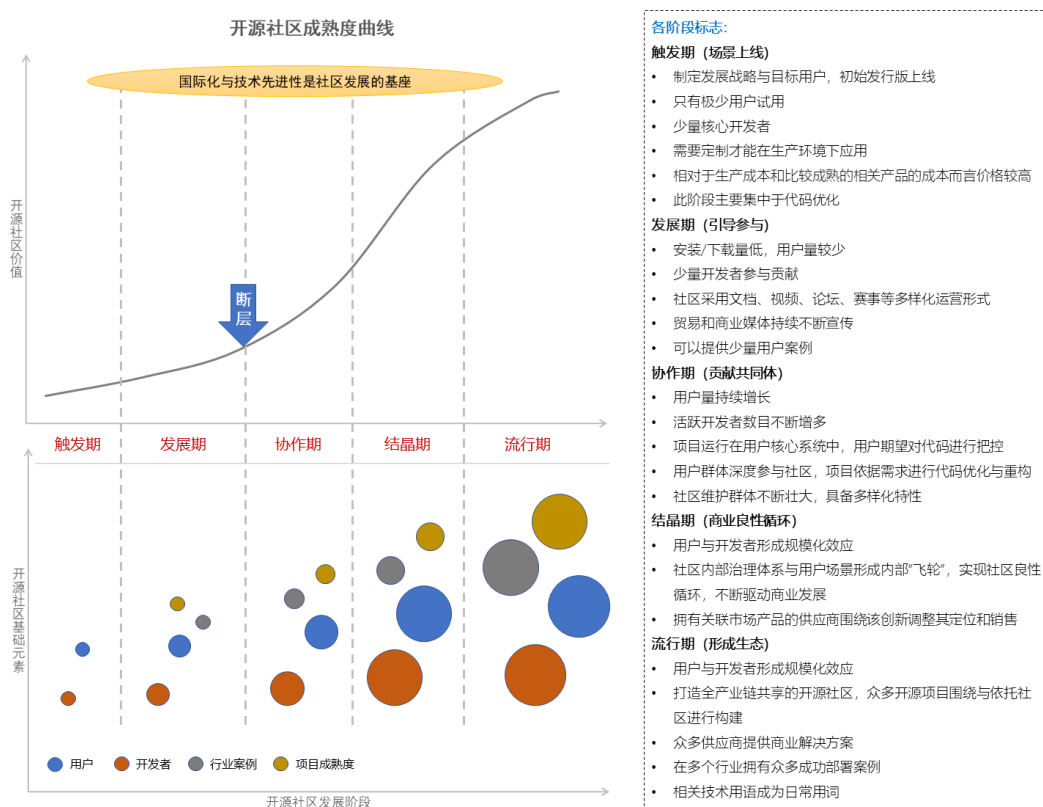


图 1 开源社区成熟度曲线图

触发期为正式开源，吸引用户使用。本阶段开源社区的核心是吸纳外部用户使用。在此之前开源主体需要明确整体的开源发展战略，对开源的投入及产出做出清晰规划。开源社区发展初期以吸纳用户参与为主要目标，需要明确开源项目的主要用户群体，开源项目能够为用户解决哪些问题，如果没有对用户及其用例的理解透彻，项目可能会被朝着不同方向拉扯，导致失去动力。在制定清晰的产品方向后，就可以通过活动及传播吸引目标用户使用，包括文档、视频、论坛等多样化形式，帮助社区参与者能在一定程度下自己解决相关问题。

发展期为引导参与者反馈。社区发展第二阶段需要引导外部使用者结合自身使用场景，对开源社区进行反馈。明确外部反馈机制，同时通过外在与内在激励驱动参与者参与项目，通过各种形式的奖励

(例如赛事奖金、名誉等)吸引人们参与社区,引导社区多样化发展。

协作期为形成共同体,促进贡献。本阶段开源社区需要具备丰富的用户场景,企业如果把开源项目使用在最核心的场景里,客户自然就会希望能对这个项目有源码级别的把控,从而促进企业的技术人员持续对开源社区进行贡献,逐步形成开源共同体,并且通过丰富的用户实践吸引更多的用户参与开源社区。

结晶期为社区完成场景闭环,与商业化形成良性互动。本阶段开源社区已经具备充分的用户案例和完善的贡献机制,同时通过社区内开源治理与运营体系形成飞轮稳定运行,社区具备可持续性发展的特点。同时通过潜在客户开发和销售拓展,找到从用户通向企业买家的路径。

流行期形成完整生态体系。本阶段开源社区可以协调全世界拥有不同文化背景的开发者与用户之间的关系,以及全球各式各样的监管机构、复杂的环境。同时能够疏通产业的上下游关系,打造全产业链共享的开源社区,在上游与下游厂商、社区、开发者到用户的全产业链之间,形成产业正循环,建设商业可闭环的良性生态系统。



图 2 开源社区成熟度象限图

开源社区成熟度象限图是针对特定领域的社区进行成熟度判断。四个象限依次分别为领导者、挑战者、战略者和独特领域者。象限使用二维模型去阐释社区间的差异，是对某一特定领域开源社区的研究总结。同时通过统一的评估准则并汇总至固定的两个维度上，包括**项目流行性**和**商业成熟度**。**项目流行性**：考察用户与活跃开发者数量、项目创新性、项目与配套服务的整体可行性、行业案例、客户体验以及社区的整体运营情况，其实质是揭示开源社区的项目在行业的流行度；**商业成熟度**：考察社区产品定位与战略、企业投入、商业模式、市场情况（供应商与商业解决方案等），其实质是考察开源社区是否有充足的商业资源去支持社区可持续发展。

（二）数字化场景时代的开源社区发展实践

2012 年前后，移动互联网在全球呈现爆发式增长态势，在此背景下各行业相继进入数字化转型时代。代码开发不仅仅成为移动互联网场景暴增的基本支撑，也成为数字化时代各领域的关注点。数字化场景前置成为开源发展的重要契机，通过移动互联网和数字化真实场景的早期采用，进而拉动社区发展，形成场景驱动与技术驱动双轮驱动，持续迭代产品的模式成为当前数字化场景时代诞生的开源社区最大特征。开源社区从最初的个人兴趣驱动演进为用户数字化场景驱动，以 TiDB、Flink 以及 Kylin 为代表的开源社区成为数字化场景时代开源社区的典型代表。本节以 PingCAP 公司与 TiDB 社区发展为例，阐述数字化场景时代开源社区的发展模式和关键特点。

移动互联网催生的数字化场景爆发，逐步催生了三大红利：移动互联网的极致应用场景红利，新一代工程师红利以及数据技术发展红利。以数据库为代表的基础软件发展路径很大程度上取决于对三大红利的趋势把握，开源成为连接三大红利的中枢机制；开源社区也成为“场景-社区-产品”飞轮模式的支撑载体；PingCAP 以“势高则围广”的理念，持续运营以 TiDB 社区为代表的新一代开源社区，打造社区飞轮效应，持续协助用户对数字化场景需求。

1、势高则围广 - 3 个从创业第一天就坚持的理念

全方位开源的战略。PingCAP 从创业初始就将开源当做公司的核心战略来推进，因为 PingCAP 相信开源是当今基础软件领域取得世界范围内成功的最佳路径，TiDB 开源至今短短 6 年时间在 GitHub 上目前已总计获得超过 28300 颗星，拥有约 1300 多位开源代码贡献者。

打造通用产品，实现全场景覆盖。TiDB 从开源之初的发展方向就确定为打造通用型产品，打造通用产品有利于尽早感知整个系统是否有结构性缺陷，验证产品对用户需求的理解是否足够吻合；同时通过互联网+、与金融为主的多种场景可对产品通用性进一步完善。开源社区的运营成为了场景拓展的最佳载体，TiDB 作为一个分布式系统需要有足够全面的场景才能推动产品的迭代速度。目前 TiDB 共积累了 10 个行业的 1500 多个用户。

国际化战略。PingCAP 致力于打造国际化开源项目并坚持走国际化的开源社区和国际化的技术生态，因此配套制定出一系列国际化策略。TiDB 社区经过中国“互联网+”公司和企业用户的全场景打磨之后，在全球范围内发现拥有场景类似的用户，这些用户可以通过国际化的社区很快连接起来，目前经过中国互联网全场景验证的产品正在加速扩展到全球用户。

2、触发期 - 如何冷启动开源社区

技术和场景双轮驱动是社区从 0 到 1 的关键。数据技术发展 50 年，在过去 10 年里，整个数据技术进入快速创新期，PingCAP 顺应数据技术发展方向，将分布式、关系型数据库、大数据、云原生、强一致算法等等技术栈进行了融合和再平衡，将 TiDB 打造成了一款架构与技术续领先数据库产品；另一方面，由于在数字化时代，场景驱动产品迭代的因素增加，比如数据容量、交互效率、数据时效性的突破等，还有市场给产品的研发周期缩短，需要尽快将产品投入到市场去获取目标场景，因此场景驱动和技术驱动推动 TiDB 产品的进化，

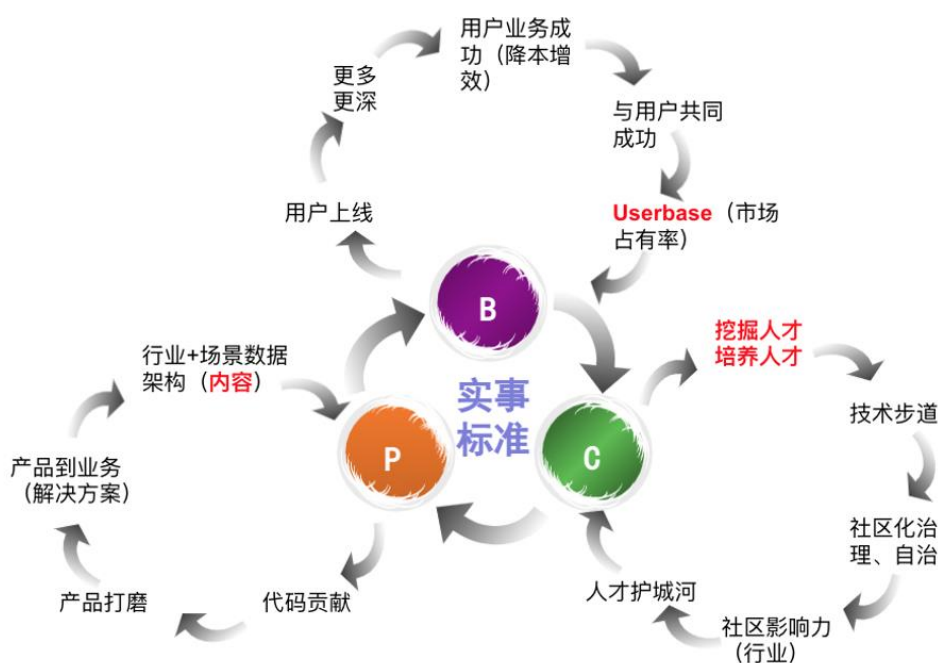
而场景驱动又成为启动这一循环的第一推动力，TiDB 作为一款通用数据库产品，开源之初坚持把真实场景作为第一推动力，解决 MySQL 用户扩容不用分库分表技术难点，形成了早期突破，之后场景驱动和技术驱动双轮驱动形成快速迭代形态。

从量变到质变。在 ToB 的市场战争中，PingCAP 通过从拒绝、到接触、到深度技术交流、到产生技术信服、再到证正满足用户需求的推广模式，拥有零星客户后完成了用户群的扩展和场景范围的扩大，以技术地推方式，形成 DBA 圈和运维圈的圈层运营，是从 1 到 10 的运营特色，TiDB 社区的运营经常通过举办各类活动例如开发者会议、150 余期的技术沙龙、技术大赛等，建立工程师的密切交流平台，吸引用户对项目进行了解，产生认知和兴趣，同时持续进行技术地推。在技术和场景的驱动下，连接了 80% 的超过一亿美金的中国互联网公司并与 300 家进行了面对面技术深度交流。。

3、发展期 - 建立从 1-10 的飞轮体系

开源社区人才战略产生的价值是非线性且多元。社区人才可以产生内容、提供场景、案例、产品测试、完成系统课程，这些内容产出又会吸引更多的人和企业端用户加入社区，企业用户 (B)、人才 (C)、产品 (P)，形成了一个循环和自转的模型，这就是 TiDB 的“飞轮模型”。通过全方位的开源与技术地推，TiDB 获得了早期的企业用户；通过全方位开源与技术宣传，TiDB 吸引了最早期参与项目的工程师，主要来参与产品的开发贡献；通过开源社区人才对产品不断的打磨与

迭代，逐步完成了 TiDB 框架与内容；而早期的企业用户培养了早期的技术使用者和布道者，早期的技术使用者丰富了产品的案例及应用场景；最终又吸引了更多企业用户的使用，同时用户的使用也更深度化，这就有效地形成了有广度和深度的用户基础。最后通过 TiDB 开源社区的治理与运营，包括逐步完成开发者与使用者的人才培养，形成了一个基于开源社区的人才生态圈。



来源：TiDB，2021 年 7 月

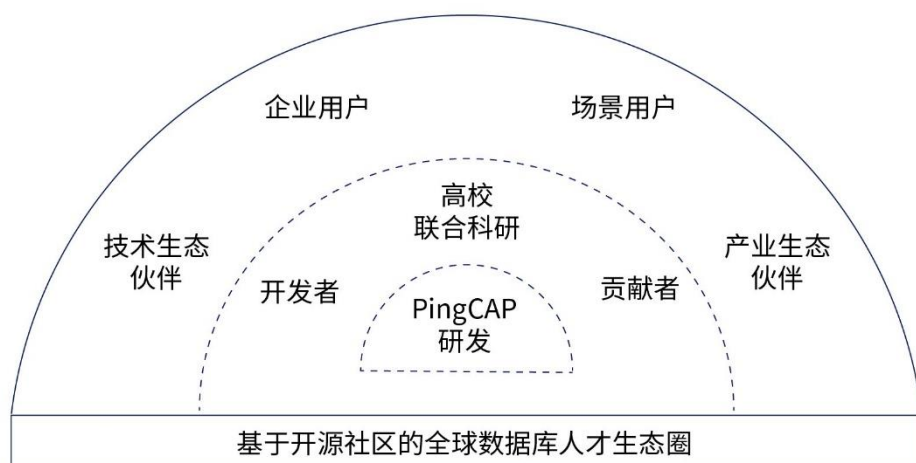
图 3 TiDB 飞轮模型

4、结晶期 - 形成层次化的人才生态

经过导入期和发展期的积累，TiDB 目前成为了具备完整运营体系、人才体系、生态体系的成熟社区，并且在商业化，国际化，企业级等传统开源社区短板领域都有长足的进步。形成了以全球人才生态

圈、全球化技术生态、社区治理与运用体系、国际化开源项目四个方面为支撑的完整体系，体现了数字化场景时代社区的整体性。

打造开源社区的人才生态圈。PingCAP 社区已经形成了一个全球数据库的人才生态圈，拥有数百人的研发团队，专注于 TiDB 的内核研发，同时参与社区众多项目的联合开发，以下图为例，在二环中拥有数以千计的开发者、社区 1600 多位贡献者、来自知名高校与科研机构和世界各地的顶级技术大咖，例如在 TiDB 项目的早期就有 Rust 语言的前核心成员参与，且与全球数据库的顶级学府威斯康星大学，CMU 的合作都建立了全球顶级人才圈层的连接；在三环中展示的是 TiDB 的全球用户与合作伙伴，目前覆盖整个数据技术生态，影响了上万个数据库工程师和运维工程师。这些用户与合作伙伴把各种场景带入到开源社区，与其他贡献者们一起，实现了基于真实场景的产品研发和快速迭代机制，并且不断的把这些场景贡献给企业用户，加速企业用户的场景拓展与业务创新。



来源：TiDB，2021 年 7 月

图 4 基于开源社区的全球数据库人才生态圈

开源社区是采用以用户为中心的社区运营模式。此模式的目的是让目标用户尽早认识产品技术价值，并对其产生兴趣，进而评估场景匹配度和需求匹配度，后续才会进入到场景延伸及商业化，最后用户在使用产品后可以对外分享，与社区其他用户不断的为社区进行贡献。TiDB 社区通过举办 DevCon 技术峰会、Infra Meetup、Hackathon、TechDay、性能竞赛以及 TUG 企业行最大程度上促进客户、社区用户、开发者交流。

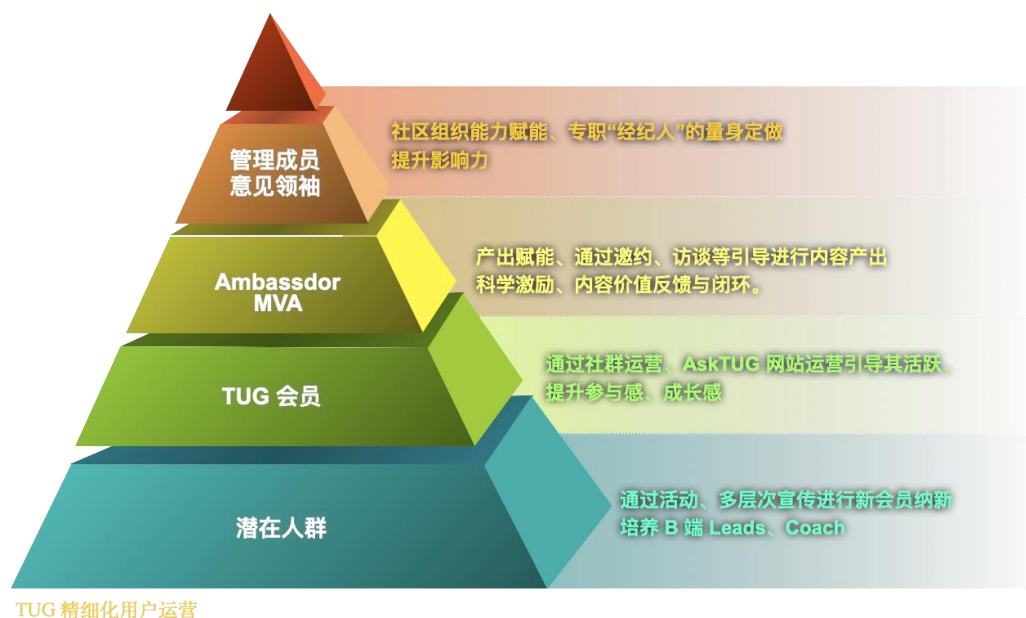


来源：TiDB，2021 年 7 月

图 5 TiDB 社区运营模式

开源社区治理借鉴经验与本地化探索相结合。PingCAP 在开源项目治理中借鉴了很多优秀开源社区的项目治理经验，比如 Apache、Kubernetes、Linux 等，但是在面对国内开发者群体时，还是会根据国内的人群特有的习惯来做一些本地化。比如 PingCAP 对社区人才进行分类，同时成立官方技术培训、社区用户问答平台、社区内容平台，

通过与企业合办活动等方式进行赋能与激励。PingCAP 把社区治理的权利交回给社区，目标是打造一个自治、活跃、全球化大数据与数据库技术社区。



来源：TiDB，2021 年 7 月

图 6 TiDB 社区治理模式

国际认可及生态合作。高度活跃的开源社区为 TiDB 产品发展带来了正向反馈闭环，TiDB 的研发能力、工程质量、迭代速度都已处于世界一流水平，成为 PingCAP 最核心的、具有加速度的竞争力和护城河：2020 年 9 月，由 PingCAP 创立的分布式存储引擎 TiKV 正式从云原生基金会（CNCf）毕业；2020 年 7 月，由 PingCAP 创立的混沌工程测试平台 Chaos Mesh® 正式成为 CNCf 托管项目；CNCf2019 年年度报告提到，PingCAP 在 2019 年 CNCf 全球代码贡献排行榜中名列第六，PingCAP 以开源社区为依托，积极打造全球产业生态，已与多家行业领军企业建立了合作伙伴关系，涵盖产品研发、解决方案、技术服务、上下游生态、云服务等多个领域。

四、开源社区度量体系

开源社区度量辅助社区摸清所处进程。通过开源社区的度量能够了解开源社区发展关键要素的发展进程，对应了解开源社区在成熟度曲线所处的位置，找到社区发展的的问题，推动社区改进进行正向发展。在国内外现有的开源社区度量体系已经逐步形成，包括中国信通院可信开源社区评估体系、Linux 基金会 CHAOSS 项目、CNCF 基金会 CII 最佳实践等。

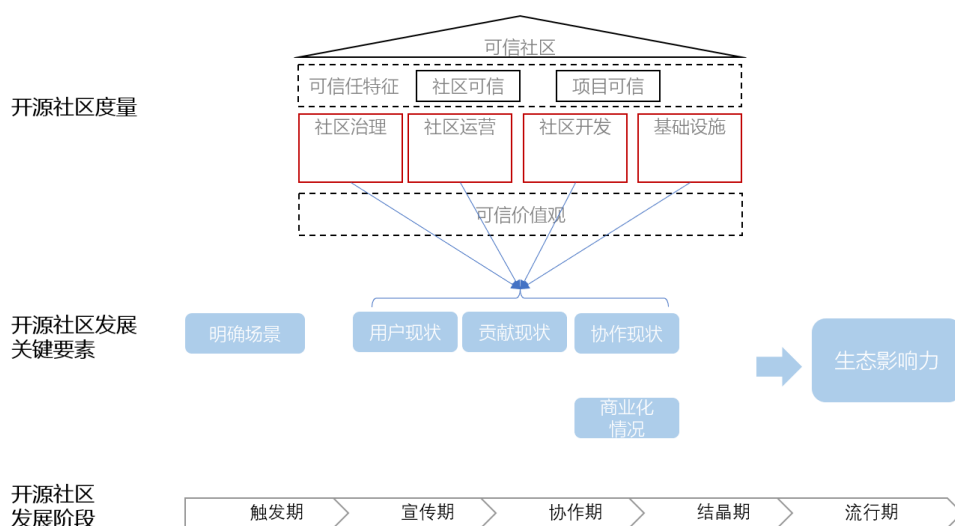


图 7 开源社区度量体系图

四个关键性指标度量开源社区。就目前国内外开源社区的发展，我们不难总结出当下开源社区健康度的关键性指标，概括起来就是社区治理、社区运营、社区开发以及基础设施这四个关键性的指标，从这四个方面着手，可以衡量出社区的健康度情况，梳理出开源社区应关注的内容，聚焦于如何构建活跃的用户、开发者生态与可信的开源社区。

（一）社区治理

成员管理是社区治理的首要举措。社区首先应当了解社区内组织或个人对于项目的参与度,包括何时参与,何种形式,以及做得如何。只有了解组织和个人对开源项目的参与情况后,才能够进行有效分析。同时围绕所了解的信息与情况,制定多样化的,有明确、包容的行为准则来规范社区维护者和参与者的行为,帮助不同背景的参与者融入社区,同时也应提供违反行为准则者的约束措施。保证社区治理的公平公正公开,走好社区治理的第一步。

文档管理是社区治理的重要参考。社区内重要项目的文档应完备不遗漏,并且确保不同的用户都能理解使用。社区应具有高可用性的最终用户文档来提高开源软件的可靠性与易用性。对于社区内项目文档的管理应当有分布式公共版本控制系统,包括版本之间的变化,帮助贡献者与用户沟通版本差异。与此同时,社区应当允许提交、存档和跟踪错误报告,确认或响应错误和增强请求,而不是忽略它们。管理好每一份文档,帮助贡献者与用户参与项目,成为社区治理的重要参考。

组织架构是社区治理的主体框架。开源社区应具备明确的开源组织架构与分工,其职责应至少包括项目管理、代码审查、法律合规审查、安全管理。在项目管理上,应当了解社区围绕或支持给定项目的活跃程度;在代码审查上,应当负责审核代码的质量和安全性、pull request 评审、更新维护版本等;在法律合规审查上,开源社区应当有明确指定 FLOSS 许可证,项目网站支持 HTTPS。在安全管理上,应当接收、调查和披露社区相关的安全漏洞,了解与给定项目的使用相

关的潜在知识产权 (IP) 问题。合理分配好开源组织架构的分工，为开源社区治理搭建起一个良好的主体框架，推动社区后续可持续性发展。

流程规范是社区治理的关键手段。开源社区应具有清晰的社区流程规范，包括社区决策制度、投票机制、审批制度、反馈机制、辅导机制、发布机制等，保证社区领导力的健康发展，促使社区成员明确相关制度的存在并依照制度规定执行。同时，社区内的成员应当了解与软件开发相关的安全流程和程序，使得社区治理更加安全可靠。严格遵守社区内部的流程规范，不仅仅是对于社区文化的尊重与赞同，更是能在社区治理中起到关键性的作用。

（二）社区运营

多样化的活动带动社区运营。开源社区内项目所在地以及社区活动参与应具有多样性和包容性。运营期间，社区应定期举办线上线下的会议活动来推动内部/外部贡献者积极参与社区贡献。除会议活动外，还应进行各类市场活动和项目布道，与不同的组织进行合作，如国内外高校、公司企业、其他开源社区等。TiDB 社区定期举办各类技术峰会、Infra Meetup、Hackathon、TechDay、性能竞赛等。

挖掘项目潜在价值促进社区运营。开源社区内项目的重要性毋庸置疑，只有充分挖掘出项目自身的价值，才能更好的促进社区进步。关注项目的公共价值，确定项目对其用户社区（包括下游项目）或贡献者的价值程度；就个人价值而言，应该确定项目对作为个人用户或贡献者的价值的程度；从组织的角度来确定项目的货币价值程度。优质的项目才是社区发展的核心，它能吸引更多的开发者与用户，促进

社区的运营进入一个良性的循环。

开发者是社区运营的中坚力量。开源社区应具有建立匹配社区的开发者培训认证体系，增加社区开发者的储备厚度，更加系统性的培养人才。同时也要注意降低开发者门槛来吸引更多的开发者参与，通过利益驱动的方式鼓励参与者积极贡献。只有构建了这样良好的运营机制，社区内的人才储备才会逐渐丰厚，利用人才的优势来反哺社区的运营，使得社区的力量愈发坚实。

用户反馈是社区运营的良性激励。社区应在运营过程中寻找潜在的行业用户，通过在生产环境下获得反馈，优势场景研究和输出，最终产出用户最佳实践并进行广泛传播。同时，社区也需要持续提升产品的易用性，例如给用户提供好用的操作流程与 API，产品可以持续稳定升级等方法，这样既降低了社区产品的使用门槛，又可以吸引更多的使用者去使用体验。社区的运营需要得到及时的反馈，通过反馈来改良社区自身的运营方式，促进社区不断进步，而用户的反馈则是社区运营最好的激励方式，注重这一环节，势必可以让社区健康向上的发展。

社区活跃度监测保证社区运营平稳良好。开源社区应持续监测社区的活跃健康度，监测指标包括：项目状态，即 Pull request 数、Commit 数、Issue 提交数与评论数、Star 与 Fork 数等；贡献者趋势，也就是贡献者活跃时间、特定时间内贡献者的增加或减少量，特定时间段内有多少贡献者停止贡献的指标等；用户数量；社区文章阅读量；用户案例趋势；企业或组织劳动力投入；社区内会议活动的参与度分析等。

监测指标越详细，对于社区的了解程度就越深，就可以及时发现并解决问题，保证社区运营的平稳良好。

（三）社区开发

代码管理是社区开发的核心基础。社区开发的核心就是代码，要注重社区项目的源代码随时间变化情况以及明确代码管理机制。了解用于改进或审查代码质量的流程（例如：测试、代码审查、标记问题、标记发布、响应时间等）。除此之外，社区应当拥有涵盖大部分代码或功能的自动化测试套件，并正式要求对新代码进行新的测试。保证代码有着良好健康的管理方式，为社区的开发注入一针强心剂。

构建管理使得社区开发有效可靠。社区开发应当使用标准的开源工具作有效的构建，使用公共协议或算法。保证构建环境的可信，构建执行过程的可信以及构建数据源的可信。同时，社区应对各分支代码分别进行管理，保障各分支均明确负责人员，对每个分支进行同样的管理流程和测试步骤，标记并记录每个分支与主版本差异部分。具体来看，还应启用（并修复）编译器警告和类 lint 的检查，执行静态分析工具并发现修复可被攻击利用的问题漏洞。合理的构建管理才能保证社区开发的每一个环节都有效可靠，从而使社区进入良性的发展。

漏洞管理为社区开发做好后勤保障。社区开发中对于漏洞的管理应当是严谨认真的，对于漏洞社区应该有自己的安全软件进行感知，同时也需要让每一个开发者都了解社区内部的安全检测机制，能够甄别常见的漏洞错误。发现漏洞后能够及时记录并报告社区漏洞。同时社区应根据漏洞影响分析报告，组织编写漏洞公告并评审，制订完整

的社漏洞修复方案，通过黑盒、白盒、灰盒、软件包成分扫描分析等方式验证版本漏洞修复情况。最后，社区还应有明确的漏洞与补丁发布机制，对之前发现的漏洞进行修补或应对紧急情况。对于未知的风险应该有紧急的处理方案，对于已发现的风险应该有合理的解决方案，加强漏洞管理，为社区开发做好后勤保障。

（四）基础设施

开源社区应使用自动化工具平台来协助相关人员进行规范的开源管理，工具应具有明确的权限管理与安全机制，为整个社区搭建一个稳定有效的平台，将社区的治理、运营与开发连接起来。

网站是社区的直观体现。开源社区应当有一个稳定的网站，同时能够说明项目可以做什么，应该如何获得，如何提供反馈，如何贡献以及如何首选风格。

代码仓库是社区的核心部分。开源社区对在开发中需要修改的源代码应建立源码仓库存储，同时设置开源软件白名单，对需要引入的开源代码进行开源许可证、开源漏洞等方面审查，通过后方可使用。

发布平台与测试平台是社区的必要保障。开源社区应对软件交付物进行管理，对社区提供给用户的交付物进行开源软件识别和评估，保障交付物符合用户需求和开源合规要求。同时开源社区内测试平台应进行开发者测试，拥有覆盖大部分代码或功能的自动化测试套件，并自动拉取新代码运行测试，应用动态检查，包括但不限于执行内存/行为分析，模糊测试，web 安全扫描等。

代码扫描工具是社区的可靠保证。开源社区应运用代码扫描工具

发现代码漏洞等安全问题，并及时作出相应处理。同时应通过工具识别代码中的开源组件及开源许可证，确保开源许可证合规使用，如有问题应及时作出相应处理。

随着开源技术的广泛应用，开源社区的各项建设也越来越完善，围绕社区治理、社区运营、社区开发、基础设施这四个指标进行社区构建与维护，打造一个健康可信的开源社区，加快推动社区健康度分析以及开源软件的建设工作，是当今国内外现有的开源社区所带来的宝贵经验。未来社区的发展与维护，既要按照现有体系标准，脚踏实地地进行开源社区的建设与发展，又要在已有体系的基础之上寻求突破，创新发展模式，推动社区向更好更健康的方向发展。

五、开源社区未来发展趋势

（一）基础软件领域开源社区蓬勃发展

开源在未来将成为引领基础软件发展的决定性力量。操作系统领域，Linux 社区约 1300 家公司 14000 开发者为内核做出贡献，根据 Linux 年度报告，在操作系统领域，Linux 分别占据 100% 的超级计算机市场和 82% 的智能手机市场，桌面操作系统市场排名第二；数据库领域，根据 DB-Engines 调研，2021 年 1 月，开源数据库流行度首次超过商业数据库，同时在 Top 5 数据库排行中，开源占据三个席位，代表着开源拥有充足的潜力去支撑已经成为支持数字技术创新和数字经济发展的的重要支撑力量。

新一代基础软件领域开源发展机遇来临。第一代开源社区发展初期，分布式场景尚未兴起，最终行业用户数字化程度不高，对于上游

开源社区参与不多。新一代开源社区模式发展的基础软件充分考虑用户需求，同时形成开源社区、商业产品、用户场景的良性互动。

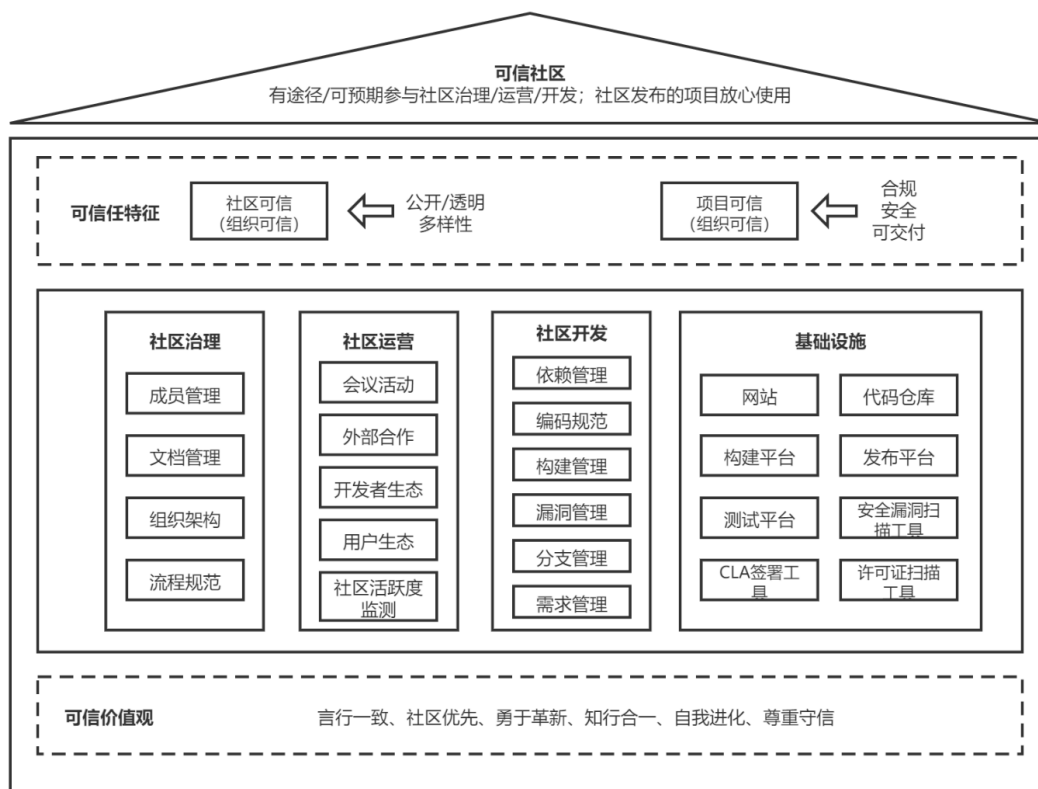
（二）开源社区商业色彩逐渐加深

开源社区的参与方不仅依靠个人情怀，企业带着商业诉求参与开源社区的趋势更为明显，开源社区逐步考虑用户真实场景需求。于此同时，开源的参与方商业模式更易构建，企业形成开源+服务的多种商业模式，随着开源商业模式的不断成熟，开源社区与商业模式将形成良性循环，商业上的成功支撑开源社区不断壮大，从而产生更多的技术创新。

（三）最终用户对开源社区的参与更加深入

软件定义成为趋势，各行业面临海量数据处理场景，开源成为提升企业数字化能力的有效方式。全球数据量飞速增长，根据 Statista 统计 2020 年全球数据量达到 47ZB，人工智能、科学计算、航空航天等领域均面临大数据处理以及复杂平台建设难题。行业用户逐步自发开源，将通用技术平台开源，推动行业协作的方式解决通用问题，石油、航空航天等多个领域探索开源方式解决信息化建设问题。石油加工服务商斯伦贝谢从 2019 年逐步将勘探环境的核心数字平台开源，为油气产业链提供一个可扩展的数字化生态系统；NHERI SimCenter 为自然灾害工程研究和教育界提供数值模拟与仿真软件工具，2019 年发布了三款研究工具全部开源，来支持自然灾害工程建设。

附录一：中国信通院可信开源社区评估体系



开源社区评估体系评估对象为开源社区，开源社区=人+项目+基础设施平台，一个好的开源社区有助于开源项目营造良好的开源生态并扩大影响力。信通院开源社区评估从基础设施、社区治理、社区运营与社区开发这4个角度，梳理开源社区应关注的内容及指标，聚焦于如何构建活跃的开发者的生态与可信的开源社区。

同时梳理出开源社区应该监测哪些健康度指标，包括：a) 项目状态：Pull request 数、Commit 数、Issue 提交数与评论数、Star 与 Fork 数等；b) 贡献者趋势：贡献者活跃时间、特定时间内贡献者的增加或减少量，特定时间段内有多少贡献者停止贡献的指标等 c) 用户案例

增长趋势；d) 社区文章阅读量；e) 技术交流与输出情况（国际论文，讲座等）；f) 企业/组织分布与劳动力投入情况；g) 活动分析：社区内会议活动的频次与参与度情况。

附录二：Linux 基金会 CHAOSS 项目

CHAOSS 是 Linux 基金会下度量开源项目与社区健康度的开源项目，全称是 The Community Health Analytics Open Source Software，致力于推动社区健康度分析开源软件的建设工作，包括全球范围之内社区工作组的建设、社区健康指标体系的制定、开源指标数据采集与开源分析工具等。

目前来说，有三类目标：**建立与实施无关的度量标准**，以衡量 Community 活跃度、贡献以及健康度；**开发和集成相应的开源软件**，来分析 Community 的开发；**构建可复制的项目健康报告**。

CHAOSS 项目区分了两个委员会和五大工作组。委员会主要区分为指标委员会和软件委员会。工作组的目标即完善开源 Community 相关的指标，并确保其可以使用相应的软件可以实现。使用 GitHub 的项目进行组织和管理，具体的有：**多样性和包容性**：目的是在开放源代码项目中引入经验以衡量一致性和包容性，并在可能的情况下获得软件的支撑；**Evolution 指标**（用于软件开发项目）：专注于开源项目的生命周期；**通用指标**：所有的用于 Community 和开源相关的指标均在这个工作组进行；**经济价值**：围绕开源项目和 Community 的相关经济活动；**风险评估**：重点关注与开源风险有关的指标。

附录三：CNCF 基金会 CII 最佳实践

Linux 基金会 (LF) 的核心基础设施计划 (Core Infrastructure Initiative, CII) 最佳实践徽章是 Free/Libre 和开源软件 (FLOSS) 项目用于表明它们遵循最佳实践的一种方式。项目可以自愿进行自我认证, 通过使用此 Web 应用程序来解释他们如何遵循每种最佳实践。CII 最佳实践徽章的灵感来自 GitHub 项目众多可获得的徽章系统。徽章的使用者可以快速评估哪些 FLOSS 项目遵循最佳实践, 因而更有可能生成更高质量的安全软件。

下面列举了一部分 CII 最佳实践指标项: 有一个稳定的网站, 说明项目的功能, 如何获得源代码, 如何提供反馈, 如何进行贡献等; 拥有明确的开源许可证; 文档记录如何安装和运行 (安全地) 以及任何 API; 有分布式公共版本控制系统, 包括版本之间的变化; 允许提交、存档和跟踪错误报告; 有安全的、记录在案的流程来报告漏洞, 并在在 14 天内回复, 在 60 天内修复公开漏洞; 使用标准的开源工具作有效的构建, 包括启用 (并修复) 编译器警告和类似 lint 的检查, 执行其他静态分析工具并修复可被攻击利用的问题; 拥有涵盖大部分代码/功能的自动化测试套件, 并正式要求对新代码进行新的测试。

中国信息通信研究院 云计算与大数据研究所

地址：北京市海淀区花园北路 52 号

邮编：100191

电话：010-62300557

传真：010-62300557

网址：www.caict.ac.cn

