T O P T E C H T E A M

中国顶尖技术团队 CHINA

访谈录

2021 第四季

封面故事

网络安全风云 15 年：没有天才，也没有了江湖

InfoQ

目录

封面故事

重磅访谈

封面故事

网络安全风云 15 年：没有天才，也没有了江湖

采访嘉宾：吴石、蔡军、聂森

采访：Tina、魏星、蔡芳芳

撰稿：Tina

> *这个行业，没有传说中的江湖，没有莫名其妙的绝招，也没有那么多各种莫名其妙的天才。*

一个时代的优秀人才总是成批的涌现，再成批的褪去。中国上一代最优秀的安全人才，有的禁不住诱惑去做了黑产，有的看着安全行业没前途选择了转行，还有一部分在“3Q 大战”中被大企业收编。之前凭着热血和激情人行的“草莽之众”，逐渐大浪淘沙转变为了企业里的正规军。

这些选择进入大企业的人中，不乏当年最顶尖的高手（又被称为“白帽”），他们逐渐成为了这个时代的中坚力量，用自己的一言一行，以及做事的逻辑和方法，为这个圈子培养出了更优秀的“新生代网安人”。科恩就是一支被这样的顶尖人才引导成长起来的队伍，他们更锋利、自信、有实力，比上一代网安人更有力量去实现“改变行业、保护世界”的理想。

安全人才等来了黄金时代

在中国做安全研究的不多，而能称之为顶级选手的人数更是稀少。早期的从业人员，普遍“半路出家”，专业背景包括中文、生物、法律、医学等等五花八门，大多是出于对网络安全技术的热爱而自学成才。

2006 年之后，因为网络游戏的流行，一些早期“黑客”发现可以通过外挂、木马等方式盗取游戏用户信息和虚拟资产进行变现，因此在“3Q 大战”之前，几乎一半的安全人才禁不住诱惑逐渐从事黑产违法活动，人才流失极为严重。

剩下为数不多的网安人，普遍拿着不及 IT 行业平均水平的薪资，拥有一身本领，却始终坚持正义和初心艰苦地熬着日子。

吴石，是这些人中“殿堂级”的大师人物之一。

从复旦大学数学系毕业后，吴石在一家 IT 企业任职，因为兴趣开始利用业余时间查找漏洞。吴石曾在微软 Word 里发现了一个严重漏洞（CVE-2010-3333），这个漏洞导致的最严重的场景是，比如你发送一封邮件给任何一个人，他不用点开邮件打开附件，只要到达服务端，用户的电脑就会被远程控制。在微软正式修复之后，仍有地下黑客通过比对分析得出利用原理，还进行了很长一段时间的攻击活动。因为 0-Day 漏洞极具价值，曾有黑市买家想以十倍于 ZDI 的价钱购买他发现的漏洞，但吴石并不为所动。

到 2010 年《福布斯》报道他时，吴石已经发现并报告了 IE、Safari 和 Chrome 等浏览器中存在的 100 多个严重漏洞。安全专家查理•米勒（Charlie Miller）说：“或许苹果应当聘请吴石来帮助他们，因为他发现的苹果操作系统的漏洞数量是苹果整个安

全团队的两倍还多。”福布斯评论说，苹果很幸运，因为遇到了“像吴石这么厚道的人”。

2012 年，上海碁震云计算科技（Keen Team）在上海成立，当时人数不多，只有三五位在职人员。2013 年吴石以首席科学家的身份加入了 Keen Team，Keen Team 也从此迎来了大发展。

2013 年 11 月 13 日，Keen Team 团队在东京 Pwn2Own Mobile 比赛中攻破 iOS 7.0.3，成为亚洲和国内第一个拿到顶级赛事冠军的团队。

对于上海交大硕士毕业加入 Keen Team 的聂森来说，吴石既是这个行业的领军人物，也是行业的一面旗帜，指引着新人前进道路的方向。因为从小对黑客技术感兴趣，在上海交大读书时，聂森每周会读一两篇行业内的论文，并在微博上记录读后感。作为前辈的吴石是为数不多的、愿意以网友身份无偿地给予点评意见和建议的人。聂森回忆说，当时他正处于瓶颈期，真心能体会到自身技术发展上和国际前沿之间的差距，这些点拨让他有了一个能突破自我的机会。

同时他对吴石充满了钦佩，“在他那个时代做安全，收入不好也没有什么人关注。现在的年轻人可能只看到了大家风光的时候，谁能想象得到这个行业在前一二十年间惨淡的状况。在行业没那么好的时候，还能脱颖而出，这靠的是定力、不断的积累，以及足够的热情……能坚持下来的，只有也唯有真爱。”

2013 年，“斯诺登事件”爆发，从国家层面开始重视安全，奇虎 360、腾讯和阿里巴巴等互联网企业也愿意在安全上做投入，纷纷开始收购市面上不错的安全团队。Pwn2Own 夺冠，使这支一贯低调的团队走进入了公众视野，吸引了包括腾讯在内至少 5 家大型公司的投资意向，COO 任宇昕和腾讯副总裁丁珂甚至带着多位总经理直

接飞到上海与他们聊收购意向。

2014 年 1 月，Keen Team 正式加盟腾讯，2016 年正式成立“科恩实验室”，结合早期的积累，腾讯旋即推出了玄武实验室、云鼎实验室等“七剑下天山”的安全矩阵。腾讯给科恩最初的定位是基于 Keen Team 的漏洞挖掘能力，支持公司的内部产品安全，并不要求帮助公司挣钱，也没有什么 KPI 考核限制。

选择加入腾讯是当时的最优选择，同时也能“使我们当时能够达到比较好的收入水平”，吴石表示。

腾讯对加盟的安全人员也极为重视：整个腾讯 16 级以上的专家只有 3 位，吴石是其中之一。

互联网企业开始收编安全人才，带来一个明显的趋势是“从业人员的待遇越来越好”。科恩副总经理蔡军，在安全行业从业 30 多年，也是早年加入 Keen Team 的老员工，据他回忆，“从毕业生薪资统计来看，连续有几年，网络安全人才薪资福利在 IT 科技行业里都是最高的。”

老一辈的网安人更多是因为个人的兴趣爱好，比如 TK 教主原来是学医出身，他加入腾讯也是行业里的一个标志性事件。

“我们那个时代还有很多也都不是科班出身，但现在我们招收的网络安全专业毕业的科班同事越来越多了，”蔡军补充说，“科恩团队现在 90%以上新加盟的都是经过七八年计算机专业学习、科班出身的年轻人。”大厂在安全上的投入改善的不仅仅是网安从业者的生存现状，更影响了整个行业，“我还听说有很多黑产洗白的故事，不再去做违法乱纪的事情，因为现在正常合法的途径和空间已经很大了。”

没有飞花拈叶的绝招，也没有所谓的江湖

过去十年，网安从业者的形象已经被塑造成了掌管着开启网络世界大门钥匙的人，只需一个动作，便能穿梭屏障来去自如……如果说顶级玩家真的拥有绝招，那么吴石的秘诀是什么？

《福布斯》的报道中提到，吴石掌握了一种独特的“fuzzing”方法，关注的是软件架构，而不是细节。业界还传说他有一套自己的“漏洞数学模型”，能从编译过的二进制文件中，逆向找到软件里的算法逻辑或业务逻辑的问题。

发现微软 Word 软件漏洞（CVE-2010-3333），也是基于这种 fuzzing 分析。微软有一种富文本（RTF）格式，在 2011 年前这个格式基本不公开，但如果用文本编辑器打开 RTF 的一些文本，你会发现它的格式很有规律。通过查看这些样本，吴石手工构造了一些对于 Word 程序来说比较奇怪的样本，再将它们不断“喂”给 Word 程序，Word 很快就崩溃了。通过对崩溃过程的分析，一步一步地找到漏洞。

吴石说，理论上 Word 经过了比较严格的测试，这是一个一般程序员、一般公司都不会犯的软件错误。“漏洞挖掘的过程实际上是在符合规则的前提下，构造了程序员很难想到的一些样本去寻找软件的断点，这是最关键的一点。”

“其次是让构造的样本数据或程序代码尽可能互相关联，这样能较快地进行收敛。一开始目标程序可能是个黑盒，通过投喂精心构造的样本数据了解软件的处理逻辑，能够知道在哪些地方可能有点问题——目标程序就逐渐变成了灰盒。”

微软谷歌这样的大厂产品，实际是很安全的，他们会想出各种方法测试自己的软件，安全问题最主要还是依赖于测试。所以可以理解为漏洞挖掘本质考验的是网安工程师

对软件的理解程度，通过动态分析和静态分析，了解程序在做什么以及是怎么做的，并在测试条件下挑程序员意想不到的地方。所以某种程度上，也是要求开发者不要犯同一个能让别人反复猜到并利用的错误，“只要程序员每次能犯不同的错误，我觉得是可以接受的”，吴石说。

漏洞挖掘过程中的手工分析同时也考验攻防双方的**编程基本功**。编程中任何容易出错的地方，都有可能产生被利用的漏洞。比如如果使用 C 语言，要写得比较安全，得去了解前人总结的几十或上百种不同程序员容易出错的模型，“看完这个才能上岗”。目前开发者比较容易出安全问题的地方包括代码的“边界检查”——尤其是那些容易造成堆栈溢出的逻辑，“老实说，大部分程序员的算数不是很好，有稍微复杂一点的加减乘除运算的地方，甚至包括谷歌和微软的程序员都容易犯错。”另一个值得开发者警惕的是程序执行时的“竞争条件”，典型如不同进程操作同一块数据，非常容易带来各种各样的安全问题，并且一般的测试很难发现这样的问题，是一个需要程序员予以警惕的漏洞模式。

吴石认为，信息安全已经成为了工程技术领域内的一个行业。做得好的话，跟其他的工程技术领域没有什么区别。

“一开始大家觉得这是一个江湖的事情，有很多各种莫名其妙的天才，有各种各样莫名其妙的绝招。**其实没有**。”

“因为经过三十多年的发展，这个领域没有什么东西是你想得到而别人想不到的。2000 年左右，国内开始有一种奇怪的风气，喜欢把网络安全跟武侠文化结合在一起，好像每个人都是有一些不传的秘籍，只要能够一使出来，就可以飞花摘叶、取敌人首级于千里之外。实际上这个领域不存在特别神秘的东西，也不应该有神秘的东西，尤其不需要人为地制造这种神秘感。而且这个领域没有、也不应该有那种江湖侠义或‘黑客精神’。目前它还是会影响很多年轻人，导致他们很崇尚个人主义或名利上一些比较奇怪的追求。”

“我们应该以一种平常心来看待这个行业。要成为一名安全领域的高手，所需要做的

就是不断学习，把过去几十年已有的知识变成自己的能力，这是一个痛苦的、需要不断花精力去不断练习的过程。最重要的就是有自驱动力，踏踏实实地做事，每天进步一点点，日积月累最后终能取得比较大的成就。”

这个领域，对个人能力的依赖性或关联性的确相对于其它领域高一点，安全行业曲折地发展到现在，不缺聪明厉害、非常有个性和能力的人，但崇尚“侠客风范、个人英雄主义”反而有碍于自身发展了。行业在进化，**任何一个大的技术进步都是靠一群默默贡献、踏踏实实的人共同完成的**。个人成就已经和团队紧密相连，需要大家懂得合作。这个行业已经不鼓励独狼行为，而是希望团队至上，能将个人的成功、兴趣爱好叠加到公司需求、行业需求上，包括对国家、对一些技术承载有责任担当，个人才能走得更扎实更远。

“板凳能坐十年冷”是吴石的口头禅，他强调：“安全这个行业最看重的不是天赋、智商，最关键的是要能坚持下去，只要能坚持下去，一定能够比大多数人眼里的聪明人做得更好。”

如何运营一支安全团队

> 安全的软件是经过不断地攻防对抗演化出来的；而网络安全上的竞争，归根到底是人才的竞争。

从 2014 年 1 月被收购，到 2018 年腾讯“9·30 变革”，科恩团队一直相对独立运作，类似学校的研究实验室，针对网络安全做漏洞攻防领域的研究，输出重点是为腾讯的业务保驾护航，作为子公司也需要以人力形式帮客户提供一些软件安全服务。

作为腾讯安全“七剑”之一，吴石给科恩设定的目标之一是成为腾讯安全的一张名片，通过 CTF 比赛或一些比较困难的技术研究，让外界了解到腾讯安全有能力帮助客户做好网络安全服务。

收购之后，科恩代表腾讯安全在国际性比赛中拿了 17 个冠军，赢得了三次世界破解大师（Master of Pwn）的称号。

2016 年，特斯拉风头正劲。它的安全防护技术也是全球领先的，在信息安全技术、包括人力上的投入可能是全球汽车行业里面最多的一个。2016 年 9 月，科恩成功实现无物理接触环境下远程操控特斯拉，这在全球范围是是头一次，同时也向特斯拉报告了多项安全漏洞，马斯克亲自写信致谢，并随信颁发了两块“铁牌牌”——即代表特斯拉安全研究最高荣誉的“特斯拉安全挑战徽章”。2017 年，科恩实验室再次实现了其无物理接触远程攻击，能够在驻车模式和行驶模式下对特斯拉进行任意远程操控。特斯拉也连续两年授予科恩实验室“特斯拉安全研究名人堂”称号。

Keen Security Lab of Tencent
Tower A

Dear Keen Security Lab of Tencent,

Tesla greatly values the research of the security community. We especially appreciate when researchers help us improve the safety and security of our products. Your recent Tesla research has led to valuable improvements to the security of our vehicles.

We have enclosed Tesla security challenge coins as a token of our appreciation.

Sincerely,

Elon Musk
CEO, Tesla, Inc.

TESLA

“特斯拉安全挑战徽章”和马斯克亲笔签名感谢信

在这期间，科恩还有一个任务，就是通过组织 CTF 比赛和破解特斯拉等活动来挖掘和选拔人才，将合适的苗子吸收到团队中。找好苗子不是容易的事情，人才培养储备问题是安全行业发展的一大困境。以前安全行业苦，也不挣钱，很难留住人才，大多数高校也没有设立安全专业，“70 后”做安全的屈指可数，“80 后”这一代就已经有了巨大的人才断层。根据调研机构数据，整个信息安全行业总体人员缺口在十几万到几十万之间，做漏洞研究的更加稀缺。随着大家对安全越来越重视，待遇水涨船高，企业间人才争夺也愈加激烈。

利用比赛来“掐尖儿”，科恩储备了不少好苗子，如专注于车联网安全的聂森、拿了 11 个 Pwn2Own 冠军的 zhen、CTF 比赛冠军专业户 jacky、在国家实战演练中拿到冠军的活动负责人 shu……

最开始这是一个研究团队，对大家没有条条框框的限制，天高任鸟飞，无论是移动、PC，还是智能汽车或者电力设备，只要有兴趣都可以随意研究。大家也非常容易沉浸到自己的研究中，聂森就注意到这些优异的人拥有一个共同特质，就是随时遇到问题就有去研究透彻的冲动，能立马将自己关进“笼子里”，“有可能你跟他说着话，或吃着饭的时候，他突然间陷入了一个状态，具备了排斥外界氛围的能力了，多半是大脑里在演练解决思路。”

管理这么多优秀的人，吴石坦言一开始有点“没信心”，但由于单靠个人所取得的成就相对有限，像参加 Pwn2Own、CTF 这些国际性比赛都必须要有团队一起合作。在谈到带团队的感受时他说，“首先，将一群聪明的人拧成一股绳，让大家朝着一个目标去努力并取得比较大的成绩，这个很有成就感。其次，能帮助一些年轻人比较快地成长，少走弯路，这个也很有成就感。”科恩的“战绩”能长盛不衰，吴石自然功不可没。在聂森看来，安全技术以前肯定存在中外技术差距，随着大环境变化，科恩里聪明的年轻人越来越多，且有像吴石这样资深、专业的前辈指导，团队达到国际一流水平已经没有什么阻力。

网安研究团队建设，与一般技术团队不同，行业里没有太多值得大家互相借鉴的成功经验。

由于蔡军岁数相对较大，在团队属于老大哥的角色，不少人会问他如何运营一支网络安全团队。他表示，“这个问题我也想了很多年，我觉得大概分几个方面。”

第一，常在河边走，但是一定不能湿鞋。不碰高压线，不做任何瓜田李下的事情。很重要的是要有一个很好的带头人。

“我一直开玩笑说我们有一个德艺双馨的吴石总，他为人很低调，有非常扎实的功底，厚积薄发。他决定了这个团队的气质和价值观。”

在科恩团队，很多小伙子手上的那些手艺和功夫业界一流，但凡动点歪心思就能获取巨大的利益。面对这种诱惑，对人的道德修养和品德要求是很高的。所以科恩对加入团队的这些人的品质，要求非常高。

“吴石本身是一个很好的典范，他作为一个神一样的存在，天天跟大家一起，树立的榜样就是再能干的人也是要坚持研究和付出的。吴石也四十多岁了，每天晚上也都是看资料看论文，要跟我们讨论，经常都是一两点钟才睡。”

第二，这个团队近几年已经形成一个新老搭配、优势互补的格局。有老人儿，有后起之秀，大家在不同的领域都取得过非常突出的成绩，可以互相碰撞工作研究思路和心得体会。新老搭配，同时又能够互补，这就形成了团队的整体实力。

“我们有非常正向的积极的价值观和团队文化：追求极致，团队至上，勇于担当。这对于新加入的一些同事会有一些潜移默化的影响。”

第三，重视人才的梯队建设，以及对外交流互动。这也是团队过去取得成绩很重要的一个保证。比如在腾讯我们以科恩为技术支持，持续举办的 TCTF 信息安全争霸赛。对于科恩来说，更重要的是想通过这样的赛事在专业领域去发现一些好的苗子和人才，也通过把国内高校战队和国际顶尖 CTF 战队放到一个赛场上竞技，提升国内高校学生的视野，逐步缩小、追平和国外的差距。另外科恩也非常注重跟高校合作，吸引 CTF

战队的同学来这学习和交流。还有一些金融客户、大企业客户的安全团队，他们经常也有些人来实习和交流。

“我觉得科恩的团队建设目前看还是比较成功的，当然在这个过程中也有一些人和我们志趣不同，离开团队，但是我们发现团队的整体实力强，不依赖于一两个天才。”

让安全研究走出实验室

> 黑产利润再大，但也比不上用正大光明的手段挣钱。

2018 年 9 月 30 日，腾讯宣布了重大组织架构调整，成立了新的 CSIG 云与智慧产业事业群，同时提倡科技向善，面向 To B 市场，为政府、企业提供技术支持。科恩团队在这次调整中，从 MIG 换到了云和智慧产业事业群，这个阶段配合着腾讯主体业务发展的要求，科恩的定位也有了变化，除“保驾护航”之外，还有了帮助腾讯安全以及腾讯整个云的业务做“开疆拓土”的要求。

而且任何企业运营都需要一些开销。作为一支高水平的研究团队，科恩也不希望只被“包养”，而是希望自己去创造一些价值，产生营收。吴石给科恩设定的另一个目标是把一些安全能力，即漏洞研究的心得和成果，包装成工具和产品，并将其推向市场。

这也就要求在鼓励大家做自由的安全研究探索之外，还需要有一些特定的方向，其中之一就是车联网。

聂森从 2016 年开始带领一个小团队负责科恩的车联网安全业务。智能汽车系统跟手机不同，是建立在可能几十个单片机系统之上的，比如刹车、娱乐大屏都各有自己的

单机系统，那么它的攻防找的是漏洞链条，整个环节可能涉及至少大大小小 3 到 5 个漏洞，研究周期很长，需要的技术栈也比较复杂。车载系统的破解，相对来说过去缺乏相关的技术积累，不是站在前人的肩膀上，而是一个从 0 到 1 的过程。

这一年，特斯拉的关注度非常高，每一次更新、每一个新功能的发布都受到整个行业的高度关注，所以科恩选择挑战破解特斯拉。他们花了两个月特斯拉进行了实车拆解，通过深入的逆向分析，找到了一个可以让特斯拉车辆主动连入科恩特制 Wi-Fi 的逻辑漏洞，利用浏览器的内存漏洞以及操作系统内核漏洞控制了影音娱乐大屏，随后通过入侵车载网关 FreeRTOS 系统控制了 CAN Bus，进而成功破解了特斯拉的雨刷、车速控制等功能。

在特斯拉的攻防研究的基础上，科恩陆续开展了对宝马、雷克萨斯等品牌的研究，覆盖了德国、日本和美国等全球主流的车联架构。

“我们 2016 年的这次特斯拉研究，包括后续的研究，我能切身感受到它对整个汽车行业的影响，让汽车行业看到这个车竟然可以被这样攻破，那些出现在如《速度与激情 8》电影里的情景可能会变成现实、带来人身和财产安全上的威胁。”聂森说。科恩在车联网安全等细分领域也闯出了自己的地位，不少中外汽车厂商慕名而来。

在此之前，车联网安全漏洞测试主要靠专家进行人工服务，破解特斯拉后，科恩逐渐将这些研究成果沉淀下来，形成了一个通用的工具和平台 SysAuditor。

车联网是一个比较好的安全漏洞研究载体，但实际上还有更广泛的领域，比如说工业互联网或者物联网，如家用的智能电表、骨干网上的路由器、大型电厂里的工控设备、车载的单片机。以摄像头为例，科恩之前的研究证实黑客在摄像头上完全可以做到电

影里的效果：比如黑进监控摄像头的系统，用一段已经录好的视频替换摄像头实时监控的视频。SysAuditor 工具针对的就是这些含嵌入式硬件的行业，以保证 IoT 固件的安全。

此外，科恩还针对 Android 平台发布了 App 漏洞扫描工具 APKPecker，面向大型移动互联网公司、做应用市场的手机厂商等。

以及最近发布的面向全行业的自动分析工具 BinaryAI，可以通过检测编译后的二进制文件分析软件中的漏洞情况。比如软件开发过程中，不断引入第三方代码和组件，这时候安全风险可能就来自这些包含了漏洞的第三方组件。这是一项填补了业界空白的比较特殊的创新产品，目前业内的同类产品，比如美国惠普的 Fortify 等都是分析的软件源代码。

从 2018 年初开始，科恩主动创新，已经产生了千万级别的项目收入。原来作为一支研究团队，科恩并不特别强调“钱”的事情，评判成员贡献主要依据技术突破和研究成果。而现在，在此基础上又增加了一点收入上的权重和比重，如果技术转化或者对外价值输出能够带来收入，也可以作为评判个人贡献的依据。

在未来发展方向上，科恩短期目标是继续引入 AI 和大数据相关要素，完善产品线，打造爆款产品。中期目标是帮助腾讯安全和腾讯云扩展业务。长期目标仍然是坚持一些新的技术前瞻性研究，走在安全行业前面。

写在最后

由于通常不产生明面上的业绩，网络安全行业在中国受重视程度不及其他 IT 行业。欧盟推出 GDPR 后，美国加强安全监管，一般企业对网络安全投入的占比能达到 12% 以上。而在中国，即使是对安全比较重视的金融行业，网络安全投入占比还不足整个 IT 投入的 4%-5%，一般企事业单位可能都达不到 IT 投入的 2%。

最近几年，国家愈加重视信息安全，相继出台了《网络安全法》、《个人信息保护法》、《数据安全法（草案）》以及《网络产品安全漏洞管理规定》。如果网络安全相关工作做的不到位，企业将需要承担相关责任。这些举措对企业起到了监督的作用，同时推动着行业向良性方向发展，网络安全也将踏入黄金时代。

“这个行业一路走过来真的不容易”，在安全一线从业三十年的蔡军感慨，“科恩能发展到现在，我们觉得很难得，很珍惜（当前的机遇）。”

嘉宾简介：

聂森（snie），腾讯安全科恩实验室专家研究员，车联物联安全技术负责人。目前在腾讯负责软件安全的前沿研究工作，研究成果发表在 BlackHat，AAAI，NIPS 等国际会议，并应用于智能网联汽车等安全场景。曾带领团队实现特斯拉、宝马等的远程破解案例，在汽车和安全行业有较大影响力。

蔡军，科恩实验室副总经理，腾讯安全高级专家，负责科恩实验室核心安全研究能力输出和自研产品市场推广，参与领导了特斯拉全球远程破解及成果展示项目，组织协调科恩战队荣获“强网杯”、“网鼎杯”和“护网杯”国家级安全赛事，并全部获得冠军。

吴石，腾讯安全科恩实验室负责人。20 年来一直从事网络安全方面的研究及开发工作。曾在浏览器领域、PC 软件领域的漏洞挖掘取得了系列研究性创新成果。其本人领导的科恩团队专注于移动互联网安全、车联网安全研究，与特斯拉、奥迪、宝马等主流车厂建立了合作关系，为消费者的出行安全做出了较大贡献。吴石还注重人才的培养，先后组建的 Keen Team 安全研究团队、eee CTF 战队，以及现在领导的科恩实验室，培育出了数十位具有世界先进水平的研究员。团队在国内、国际安全大赛均取得了卓越成绩。在世界级的网络安全竞技大赛 Pwn2Own 上斩获了 3 个团体冠军，并在有“黑客世界杯”之称的 DEFCON CTF 上拿到了总冠军。

重磅访谈

金蝶的进击：中国 ERP 厂商的云化进入深水区

作者：罗燕珊

当新集结的研发团队敲下第一行代码时，他们没有料到，这款日后被命名为“苍穹”的平台产品，在五年后会成为金蝶系列云产品的统一基座，以及上云竞争中的利器。

2020 年，已经有超过 400 家大型企业、央国企的数字化系统和应用在这个 PaaS 平台上运转。仅在过去不到一年内，苍穹帮助超过 40 家企业平稳替换了原本运行多年、

甚至长达 20 年的国际管理软件系统，最快的替换周期仅花费不到 3 个月。

但与此同时，大多数人对金蝶的印象，可能还停留在财务软件或 ERP 管理软件时代。实际上 2020 年金蝶的 SaaS 云业务收入就已经超过了营收的 57%，并且比例还在快速扩大。

金蝶凭什么能在短短几年内从软件公司彻底云化，以及研发出云原生的“数字底座”苍穹平台？日前，InfoQ 有幸与金蝶内部的数位专家进行交流，深入了解金蝶在从 ERP 向云 ERP（SaaS 服务）、再到 PaaS 能力构建的转型历程中，是如何思考和实践的，亦希望此文能为正在探索数字化转型的业者带来参考和启发。

采访嘉宾：

金蝶中国副总裁、研发平台副总经理、金蝶云苍穹平台部总经理 李帆

金蝶云苍穹平台部副总经理 彭璐

金蝶中国苍穹平台解决方案部总经理 徐昊

从 ERP 到云 ERP 再到 EBC

过去 30 年，金蝶经历了从 DOS 到 Windows、从财务软件到 ERP、再从 ERP 到企业云服务的数次转型。

2019 年，金蝶集团董事会主席兼 CEO 徐少春首次提出企业数字化进入后 ERP 时代，ERP 不再只侧重于“资源”或“计划”上，它正在快速转移到“业务”这个焦点，逐渐发

展成为一种更加广泛的“企业业务能力”（EBC）。

自此，金蝶以 2019 年为分水岭，将这几年概括为“企业数字化从 ERP 到 EBC 时代”，而这段转型历程，如今正演绎到属于企业级 PaaS 平台“苍穹”的重要篇章。

EBC 概念由行业分析机构 Gartner 提出，在数字化转型的大背景下，一方面，它强调了企业的信息化建设不应该只侧重在资源计划上，而是应该侧重在业务能力上；另一方面，它所涉及的范围可以不断外延和扩展。传统 ERP 关注的是企业内部的信息化，不强调与外部的合作和连接，随着数字化持续深入，企业将意识到数字化转型的最终价值在于产品和服务的创新。

在 EBC 的语境下，数字化平台从仅与客户相关的平台衍生为五大业务平台：**面向客户的体验平台、面向员工的信息系统平台、面向万物的物联网平台、面向伙伴的生态平台和数据与智能分析平台**。

理论上的支撑，让金蝶对于自身的发展有了更清晰的战略定位，2019 年，金蝶明确**以“苍穹”为统一开发平台，并以云原生架构开发迭代了多个 SaaS 产品**。2020 年，金蝶发布了“平台+生态”战略，更关注 ISV（独立软件供应商）伙伴和生态构建。

金蝶中国副总裁、金蝶云苍穹平台部总经理李帆表示，2014 年从传统的 ERP 转向云 ERP 是金蝶一次比较大的转型，苍穹则算得上是金蝶全面云化和平台化的标志，并且技术上有了新一轮更迭，同时这也是金蝶向大企业、超大型企业市场进军的转型标志。

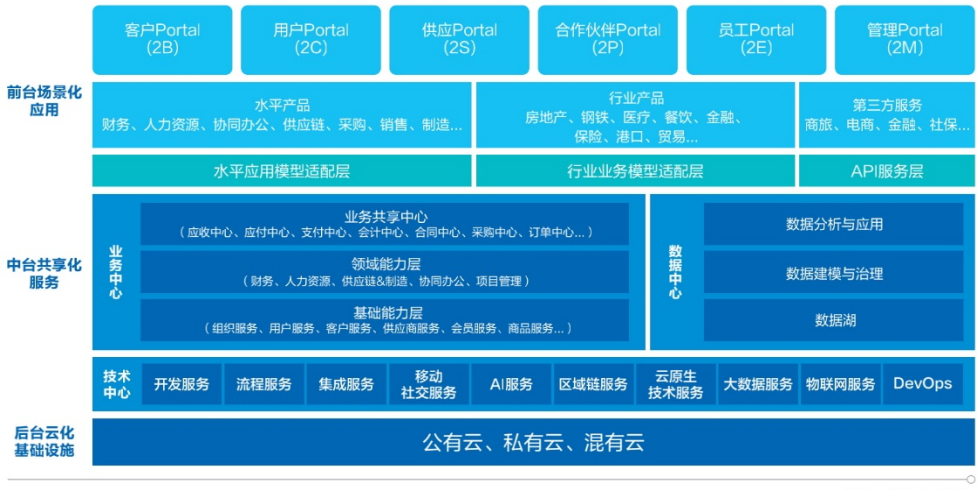
走向云原生

“从技术角度来看，ERP 云化的核心是云原生。”李帆认为，多数企业走向云原生的过程可以分为三个阶段，第一个阶段是云托管，把线下的东西原封不动地往云端托管，而现在有很多所谓的 SaaS 服务其实都是托管模式。

第二个阶段是云的优化阶段，可能会用到一些云的技术，比如容器技术。原来需要托管到主机里面，现在做了一些优化部署到一个容器，那么资源的消耗就降低了。但没有采用云原生的技术栈，仍是单体架构。

到第三个阶段，则是真正实现了多租户的云原生阶段。“比如我布置一套 ERP 给 100 个客户用，以前云托管模式，100 个客户就要在云端布 100 套 ERP。现在多租户架构给 100 个客户可能只需要布一套；云原生还包括微服务化，把大的 ERP 进行服务化拆分，然后每个服务在容器里运行。”

企业级的 PaaS，不只是技术平台，还应该有数据和业务两个核心能力沉淀，让 IT 与业务完全融合，**赋能业务人员，而不只是赋能技术人员**。同时，任何一个平台都应该是开放包容的，所以必须要有一个以 API 管理和治理为核心的开放平台，来对不同软件供应商的应用进行组装。



苍穹中台架构

技术中台是整个中台架构的底层，它为业务中台和数据中台提供了各种各样的大数据、物联网、安全、运维等专业的云计算服务，屏蔽掉技术细节和复杂性，提供简单一致、易于使用的技术基础设施能力接口。

InfoQ 获悉，苍穹从一开始研发就采用云原生技术栈，而不是“改造”已有的产品服务。“在 2016 年真正开始做苍穹的时候，云原生技术主要是像 Google、阿里巴巴等互联网公司在用，企业级的应用软件还很少涉及，金蝶那会对云原生还没有什么感知，所以在技术栈的选择上走过一些弯路。”作为苍穹平台的研发负责人，李帆对早期技术选型踩过的坑仍记忆犹新。

“比如容器技术，我们早期选择的是 Mesos，当时主要看重它的轻量级，考虑到我们企业级应用不像互联网应用规模那么大，所以选择了一个轻量级的来用，但是 Mesos 的开源力量太弱了，后面 Google 的 K8s，也就是 Kubernetes 技术栈成

为了容器的主流开源技术，所以我们后面又必须要转回来。"除此之外，李帆表示微服务框架以及数据库的选型上也有类似情况。比如从 Dubbo 到更主流的 Spring Cloud，默认数据库从 PostgreSQL 到 MySQL 再回到 PostgreSQL，这些技术选型都是在实践中调整、在实践中总结。当下，苍穹已经能兼容多种技术。

目前苍穹整个技术栈中采用开源+自研的比例比较高，这是因为苍穹从研发之初就定下了一个强制要求——**不能用任何商用软件**。理由也很简单，选用商用软件对于做公有云服务的金蝶来说，成本太高。后来，在国家政策层面，自主可控、国产替代成为"关键词"，这也与苍穹"开源+自研"的思路一致。据了解，苍穹已经与华为鲲鹏生态体系全栈技术完成适配。

沉淀"独门技术"

除了云原生，低代码也是如今 PaaS 开发平台的标配。过去，有定制化需求的客户往往需要上门定制应用，而云原生架构下，企业可以自行修改或购买 ISV（独立软件供应商）的模块并获得快速响应。

目前，低代码平台主要分为表单驱动和模型驱动两种技术路径。表单型主要面向非专业开发者，通过简单的"拖拉拽"方式编辑和配置页面、表单和流程。模型驱动的背后则包含许多业务模型、业务组件。

苍穹低代码平台的最大特点是**模型驱动架构**，开放了金蝶动态领域模型 KDDM（Kingdee Dynamic Domain Model）的模型设计标准与开发接口。早在 2000 年，金蝶 EAS BOS 引擎就开始关注和构建类似低代码的平台能力。由于金蝶每年都

会面向成千上万个开发项目进行开发扩展，其中有很多相似和类似的功能和模块，随着积累越来越多，金蝶提炼高频及通用的企业业务场景，将其封装成可复用的功能模块，以元数据和模型驱动设计与开发。

对于 KDDM 中“两个 D”的含义，李帆做了进一步介绍，第一个 D 是“动态”。软件开发有设计时和运行时两种状态，所谓设计时是指“写代码、做设计、编译”等过程，在部署之前都是设计态，真正 run 起来的时候是运行态。而 KDDM 所谓的动态则是指，在做一个应用开发的时候，设计态和运行态是融合在一起的，不需要安装部署编译，直接就可以动态生效，可以极大地提升软件开发的效率。

第二个“D”是领域，更多是指软件采用了 DDD 软件领域驱动设计的思想，然后去做不同领域的划分，包括怎么样去设计每个域。

总的来说，模型驱动的低代码平台包含了底层复杂的架构和模型思想，来支撑企业里面的真正的生产性运营系统，所以在金蝶人看来，KDDM 的“独门”在于模型的设计思想和封装，以及如何兼容二次开发和标准产品。

场景趋同，技术融合

对金蝶来说，进入 EBC 时代，大数据和 AI 这些技术也将发挥越来越重要的作用。作为五大数字化平台之一的数据与智能分析平台，身处其他平台的交叉点，其数据来自于其他四个平台，可以实现实时事件分析和流程调整、提供决策所需的数据和模型、以及自动化决策执行过程的算法。

据金蝶云苍穹平台部副总经理彭璐介绍，在金蝶的产品体系里，数据智能平台主要包括数据平台、AI 平台和 RPA 平台（办公流程自动化），从产品的角度看它们并没有强依赖的关系，“场景是趋同的，技术是融合的。”

数据平台包括：数据的可视化和传统 BI（商业智能）的前端；ETL 工具，相当于数据的后台处理，也是数仓的核心；数据集成。原本这三个方向在金蝶的体系中都是单独的平台，后来逐渐融合到一起。

AI 平台主要面向企业的一些智能应用场景，对企业业务的典型场景主要是洞察、风控、预测。RPA 这个词汇出来得比较晚，其实金蝶内部很早就有类似于工作流、业务流程管理这一类的方案。例如给企业做的“自动结账”功能，有了 RPA 之后就叫结账机器人，类似的还有报税处理，我们后来就叫报税机器人。跟技术型的 RPA 平台有所不同的是，苍穹的 RPA 是以业务为核心，并且是金蝶的企业应用中原来覆盖的一些核心流程。

彭璐表示，大家以前并没有对数据的价值有太多的预期，但现在随着技术和算法的不断发展，能做的事情多了很多，虽然不知道探索的边界在哪，但这里面更多的是机遇。

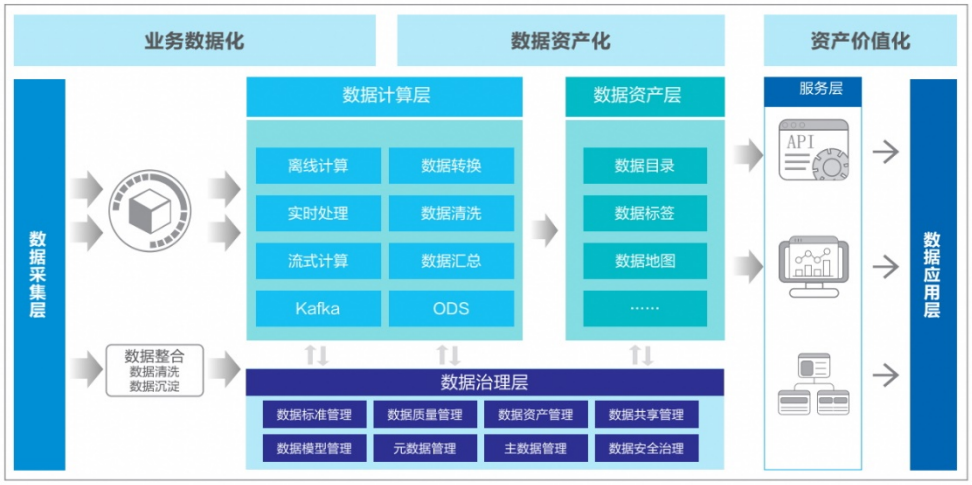


图3.2-2 数据驱动建设
来源：金蝶中国苍穹研发平台

在应用层面上，数据的用户是业务系统，中台的数据以“API”的方式提供数据服务，从而驱动重构业务本身，反哺业务。

苍穹的定位

苍穹平台在服务企业的过程中扮演了一个什么样的角色？

“我认为应该是三个角色。”据金蝶中国苍穹平台解决方案部总经理徐昊阐述，苍穹的第一个角色是“**作为金蝶所有 SaaS 的统一的 PaaS 平台**”，以业务管理为核心，本质还是帮助企业实现业务上的管理，不管上云与否，而平台技术起的是支撑和辅助的作用。

第二个角色是大企业“IT for IT”的企业级 PaaS 平台。针对一些对 IT 能力理解比较深入透彻的大企业，苍穹希望将产品和技术能力输送给他们的专业 IT 人员。

第三个角色是**企业级软件 ISV 生态伙伴的 PaaS 平台**，即软件背后的软件，最后形成一个生态平台。

2019 年底，为了应对复杂多变的产业环境和国际形势，保障业务连续性，华为海洋进行了全面的 IT 系统重构，其中一个重要环节就是替换原有的甲骨文 ERP 系统。华为海洋业务遍及全球，业务是多国、多组织项目运作以及财务结算模式，以项目为核心，对财务、研发、制造、供应链等业务的需求极强，管理难度极大，对 IT 系统的依赖也很重，IT 变革面临重重困难。需要在极短的时间内，替换 184 个涉及方方面面的 IT 系统，且又面临 2020 年 1 月新冠疫情爆发，让原本压力巨大的项目面临更大挑战。

后来，华为总部、华为海洋、金蝶三方项目团队仅用 9 周时间完成了 184 个系统同时上线，用国产化系统重构了华为海洋的业务能力，覆盖 LTC、PTP、ITR、IPD 四大核心业务流程。这个“惊心动魄”的案例，在某种程度上也是苍穹实力的证明。

关于苍穹 PaaS 服务的使用，徐昊指出有两个常见误区：第一，PaaS 平台并非为了减人，而是提效和赋能，让一些工程师可以更高效地生产；第二，苍穹 PaaS 平台主要解决的还是统一企业 IT 底座、统一 IT 战略规划的问题，如果企业目前的规模没有达到一定的程度，上 PaaS 的意义和价值其实并不大。

徐昊直言，toB 的业务没有偶然性，它是持续稳定的，还是要多年持续地积累，因为在成功那一天之前，你是不知道它会不会成功的。

写在最后

虽然从 2016 年开始研发，但 2018 年 8 月 8 日，是苍穹 PaaS 平台首次面世。

与 toC 行业不同的是，toB 产品的发布需要有成功的客户案例做背书。三年后，苍穹带着多个大客户的实践案例亮相，这些大客户覆盖了包括钢铁、农业、新基建等行业。

在这次对话的过程中，我们似乎还能从采访嘉宾的眼里感受到金蝶人在那一天的兴奋与激动。“2018 年的时候，当时我们很多人，包括老板，都觉得好像跟以前不一样了。”徐昊感慨道。

基础设施云化率已达 60%：海尔集团 IT 架构演进与云化改造的探索实践

作者：蔡芳芳

近几年，数字化转型已经成为所有传统企业的必选项，而不再仅仅是一个可选项。据清华大学全球产业研究院发布的《中国企业数字化转型研究报告》，2020 年企业数字化转型整体成熟度进一步提升，尤其国内传统企业的数字化转型已经从部分行业头部企业的选择性发展，转变为更多行业、更多企业的发展必经之路。而海尔集团在信息化、数字化方面的思考和实践一直走在国内企业的前列。

海尔集团在数字时代的转型，是自我颠覆式的全系统重组，是一场直达终端用户体验的广义“再造”，而 IT 平台在这个过程中起到了核心引擎的作用。本文，InfoQ 采访了海尔集团 IT 的数位嘉宾，以期了解这家多元化企业数字化转型和云化改造背后的故事。正处在数字化转型进程中的海尔集团，其 IT 基础设施架构是如何设计和演进的？基础设施云化改造如何推进？面对庞大的业务体系和多变的业务需求，IT 平台怎么做好支撑？转型过程中有怎样的思考和经验总结？本文将一一解答。

> **采访嘉宾：**
>
> 刘超 海尔集团 CTO
> 郭乾继 海尔集团 IT 平台 基础中心总监
> 陈合 海尔集团 IT 平台 技术中心总监
> 李晓文 海尔集团 IT 平台 安全中心总监

“数字化重生”进行时

InfoQ：海尔集团开始数字化转型的契机是什么？简单介绍下海尔数字化转型的背景？

刘超：当前，整个行业和用户的需求发生了很大变化，用户更看重的是企业能不能快速响应不断变化的用户需求，给用户更好的体验。在这样的大环境下，海尔这两年以数字化转型为契机，让海尔从传统制造业为主的跨国企业转型成为全球物联网生态品牌企业，更好地满足全球不同用户体验需求。

我们能够看到整个社会正在由传统工业化时代向数字化时代迈进，海尔传统工厂规模化制造出的产品已无法适应市场的新需求，可能生产出来就变成库存。数字化时代，规模化定制正在成为现实，比如目前我们的海尔互联工厂应运而生。首先，我们要把从设计研发、生产制造、配送安装、营销服务等各个环节要素打通并连接，同时动态地将客户的需求和生产要素进行匹配，企业的管理人员能够实时知道订单在什么环节，下一步要去哪里，这个订单来自什么用户，这个用户是新客人还是第二次购买；同时企业内部的人员可以在这个数字化平台上看到个性化数据报告，比如产线制造的哪个产品卖得更好，我们的生产工艺需要再进行什么优化等等。在这个过程中海尔通过人单合一商业模式的创新为数字化指引了清晰的方向。

InfoQ：您对数字化以及数字化转型的理解是什么？

刘超：我理解的数字化是能够帮助用户愿意和海尔做生意，用户愿意体验和使用海尔的产品和场景服务，这个过程是更快捷、更安全和更愉悦的。我们的用户获得成功是我们企业存在价值的前提，在这个过程中我们和用户之间实现双赢。而数字化转型可以帮助我们提升用户体验，让用户在接触海尔场景的时时刻刻都能有良好的体验。

数字化转型主要要解决三个维度的问题：一是找到用户，做到以用户体验为中心，当然对我们大型企业来说各个环节上的用户非常多，我们优先满足为企业直接创造价值的一线用户的需求，不仅仅是系统页面简洁和流畅，而且要能帮助我们的用户成长；二是聚焦业务场景，确保我们的数字化转型没有偏离业务目标，让业务朝着设定的目标加速迈进，这个过程中要让我们的业务和运营人员全程参与进来，大家组成链群，共同让海尔各个板块的业务利润得到迅速提升；三是打造平台，沉淀海尔自己的能力，我们打造的这个平台不仅要具备通用的能力，而且要让业务参与进来，共同打造一个有生命力的数字化平台，这个平台应该是动态的，能够随着市场变化而调整服务，共同支撑海尔集团的黑海战略。

InfoQ：您认为 IT 在海尔数字化转型过程中的价值贡献是什么？

刘超：IT 在企业中的角色向来都是成本中心，由传统意义的成本中心向价值中心去转变，最大的验证标志还是反映在业务主干流上的价值指标变化。而围绕 IT 价值创造这个核心，集团 IT 主要基于“SAFE”体系展开数字化能力建设与目标的验证。

“SAFE”目标体系，拆解开来就是集团 IT 数字化转型的五个目标和方向，分别是：体验、效率、成本、质量和安全。

- 体验（Smooth），即给用户带来顺滑自然的体验，保证让用户专注于业务。
- 效率（Agile），即提供敏捷灵活的支撑，保证让业务专注于目标。
- 成本（Fused），即兼容并蓄的体系，保证让实现专注于逻辑。
- 质量（Effective），即高效高质的实现，保证让产品专注于价值。
- 安全（SAFE），保证自始至终的安全。

海尔集团 IT 通过建设以“SAFE”为核心目标的数字化转型支撑体系，统一目标，强化与用户的链接与融合，逐步形成技术驱动型的链群组织，保证人人参与数字化转型、人人专注价值创造。

InfoQ：在数字化转型过程中，不同业务线遇到的痛点和需求各有不同，而且面对越来越多变的业务需求，集团 IT 平台怎么做到快速满足？

刘超：海尔在数字时代的转型，是自我颠覆式的全系统重组，是一场成功直达终端用户体验的广义“再造”。也正是因为海尔正在进行的数字化重构、数字化重生，促使每个业务单元进行根本性变革，而 IT 在里面正是起到核心引擎的作用。

为了支撑产业的数字化变革和灵活多变的商业模式探索，集团 IT 建立整个数字化支撑体系，以“文化、方法论、架构、能力、组织、流程制度”为核心的数字化重生六基石，统一体系内的沟通、基础、运作。在这个基础之上，我们着力数字化技术能力、业务数字化能力、敏捷交付能力三条数字化能力线的建设，通过两大工具体系实现对外的服务赋能：其一是通过业务数字化能力提供敏捷交付能力，服务业务人员；其二是通过一站式整合敏捷交付能力与数字化技术能力服务各领域 IT。其中，以快捷应用 SaaS 为代表的敏捷交付体系，正是为了满足灵活快速地支撑产业探索阶段需求而诞生的数字化工具。

而从整体的 IT 开发模式的支撑来看，我们创新“**面向业务体系的开发模式**”，这是 IT 领域为了更好地适应和支撑企业发展不同阶段而诞生的数字化能力体系，而我们也是通过快捷应用 SaaS 体系将四个阶段的开发模式完全贯穿，结合不同企业阶段的不同特性和对 IT 能力的需求，匹配不同的支撑体系，助力集团不同类型不同规模产业的快速迭代和探索发展。

海尔 IT 架构布局和演进思路

InfoQ：海尔 IT 基础架构演进的思路是怎样的？

陈合：随着技术的发展和积累，技术的垂直领域越来越多，且每个垂直技术领域内的深度也越来越深。整个行业的 IT 演进整体趋势是在向无边界应用时代迈进，海尔 IT 基础架构依托行业趋势演变出了海尔特色演进思路，即从信息化、EAI 时代，到云原生时代，到最终的无边界应用时代。真正的无边界时代是能够打破应用运行边界，连接并打通应用功能，实现顺滑自然的体验。海尔一直在探索和实践如何实现无边界，希望可以真正跨入一个新的时代。

InfoQ：可以简单介绍下海尔 IT 基础架构的布局吗？

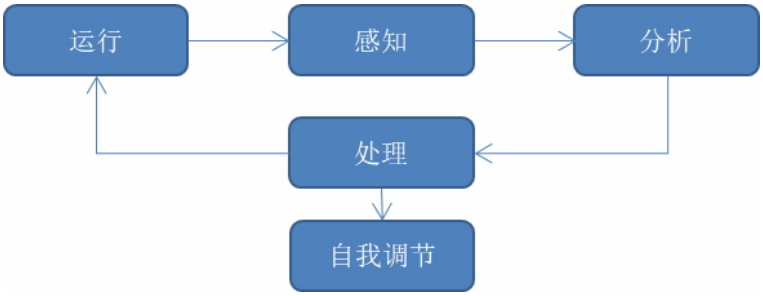
陈合：目前行业大部分企业 IT 系统主要以传统的烟囱式为主，无法响应业务快速变化的需求。海尔依托数字化转型战略，结合竖井式系统现状，采用**三台架构的模式布局**，通过构建 15 大核心科技能力布局前中后台，明确 IT、业务边界，共建实现"稳后台+连中台+活前台"转型，同时通过自主构建去形成海尔独特的数字化核心竞争力。

InfoQ：海尔集团内部业务线非常多，IT 基础架构如何在快速满足业务需求的同时又能保持系统的稳定性？

陈合：海尔集团整体业务体量巨大，内部业务线非常多，人员构成复杂，而当前 IT 设施的重要性就要求我们必须保证系统的稳定性。系统稳定涉及范围非常广，我们大致总结了 4 个方面：①架构设计和代码实现；②开发和运维管理流程；③所有人员的线上意识；④项目质量管理。海尔在这 4 个方面都有比较完备的保障体系。由于时间原因，今天只分享在架构设计上海尔比较有特色的两个点。

第一点是**无缺陷的设计模式**：为了保障系统整体稳定性，我们设计了一套无缺陷的设计模式。这不是说系统完全没有 bug，而是当不管遇到任何已知问题和未知问题的时候，系统在能力输出上都是无缺陷的，能够提供完整的业务能力。为了实现这个目标，我们通过设计让系统在分层、分级、分节点上实现自动感知、自动分析、自动处理，

最终把一个系统从开环状态变成自闭环状态。大致过程如下：



这个模式在 IaaS 层和 PaaS 层的实现相对比较容易理解，所以我以 Nginx 来举一个简单的例子：感知引擎实时感知 Nginx 流量情况，分析引擎根据感知引擎的结果发现当前流量已经超过设定的水位阈值，即交给处理引擎进行扩容处理，从任何用户的角度看系统依然能够完整提供能力。要实现整个设计最重要的点在于关注能力的输出，而不要陷入里面的具体细节。根据这个原则，海尔在能力感知、能力分析、能力处理上形成了一套完善的体系方案。传统的自愈能力都只考虑到在 IaaS 和 PaaS，而我们还深化到 SaaS 和 BaaS 层上。最终在横向业务场景、纵向架构层级上，立体地实现了无缺陷设计。

第二点是**向上统一**：除了无缺陷设计外，另一个有特色的是向上统一。业界都在提数字化转型过程中必须实现“高速路上换轮胎”，而“高速路上换轮胎”是最容易出稳定性问题的，这里就分享一个“高速路上换轮胎”的架构设计。近段时间我们在做账号统一的工作，由于历史原因，海尔内部存在多个账号中心，如果直接建立一个账号中心替换其他账号中心，在实施上涉及多个应用的调整和多套系统数据的迁移等，极容易出现问题。为此我们提出了一种叫向上统一的设计模式，新建一套账号中心只做登录认证。新账号中心作为其他账号体系向上的统一层，和多套旧账号体系实现数据互通，业务

系统和新账号中心对接，使用户流量逐渐汇聚到新账号中心。在实现全产业拉通后，通过治理再把老的业务能力逐渐沉淀到新的账号中心里，这样业务和数据都能平滑过渡，实现“高速公路上换轮胎”。另外，在这种设计模式下，前期账号中心业务单一，实施简单、风险可控，技术架构和代码质量能够比较容易得到保证，从而提升了系统的稳定性。另外由于只做登录和认证，不做用户中心，不入侵老旧业务，实施快，推行起来阻力也比较小。

InfoQ：我们了解到，自主可控是海尔 IT 数字能力建设的一大方向，从依赖外部到实现自主可控，技术上做了哪些努力？

陈合：海尔 IT 数字能力建设的重要目标肯定是要实现能力自主可控。我们首先规范架构指导，制定并发布集团级架构标准，为海尔架构夯实底座，并以《海尔集团 IT 架构白皮书》为架构评审依据统一架构规范，赋能全集团架构回正；其次，我们自主构建了 15 大科技能力，涉及权限、账号、数据、集成等，完全自研自开发；最后针对科技能力进行专利布局，目前已经申请完成科技专利三十多个。总结下来，主要就是通过标准规范、自建能力、专利保护三个方面来推动实现海尔数字化能力的自主可控。

InfoQ：未来海尔 IT 技术与架构的发展方向是什么？

陈合：毫无疑问，海尔 IT 未来的发展还是要紧紧围绕数字化、自主可控、核心能力等几个方面建设，通过构建具有海尔特色的架构模式，支撑集团黑海战略，同时依托 IT 行业演进趋势，构建账号等六大统一，推动无边界应用早日实现。

为什么要做基础设施云化改造？如何推进？

InfoQ：海尔为什么要做基础设施云化？

郭乾继：数字化转型的大背景下，系统结构发生裂变，业务切分得越来越细，相应的复杂度也呈几何倍数增加，已经超出了人力所能管理的限度。**传统架构基础设施面临以下几个痛点：私有云或托管运维投入巨大，多厂家运维，端到端 SLA 无法保证；IT 技术发展迅速（容器、大数据、EI、区块链），传统私有云不能灵活快速演进**。在这种变化下，需要有更加稳定和敏捷的基础设施平台来支撑业务的高速发展，为业务创新提供坚实的技术底座。同时，基础设施云化可以更合理地使用好资源，比如实现更高的资源使用率、更好的高可用性设计、更灵活的交付方式以及成本节约。

InfoQ：能否介绍下海尔基础设施云化改造的整个历程？分为哪几个阶段？关键阶段遇到的难点和挑战是什么？

郭乾继：云化改造和项目方式没有什么太大的区别，唯一不同就是**云化是一个没有终点的过程**，总结下来主要就是建和改。海尔云化架构采用了两地三中心混合部署的模式，对公有云、专属云、私有云做了不同的场景定位，综合考虑用户体验、信息安全、成本及国家法律法规的要求，来确定业务应该部署到哪朵云、是分布式部署还是部署到单云。改的过程中确实存在一些难点，新业务基本上采用容器云，有挑战的主要是一些老旧系统需要做一些云的适配和改造。我们的解决思路是利用“绞杀模式”，过渡期间暂时采用提供云主机的方式来云化，随着系统完成微服务改造后，再最终容器化。另外，还有一些系统例如 SAP 系统短期内是无法进行云化改造的，就必须制定长期解耦的计划，这也是挺难的一件事。

InfoQ：根据资料，目前海尔 IT 基础设施云化率已经达到了 60%，能具体讲下云化前后的变化吗？

郭乾继：云化带来变化还是非常大的，目前交付时间基本上可以做到分钟级，成本可以节省 20%以上；借助云化技术，故障发生时系统可以做出自动漂移，用户基本上没有感觉，可用性可以达到 99.98%以上。很多以前需要从零开始部署的过程都可以简化为自动交付，大大减少了管理系统的工作量，解放了这些以前不创造价值的工作时间，团队也有了更多的时间参与创新性工作，例如向智能运维方向转型。

InfoQ：目前海尔 IT 基础架构中是公有云和私有云共存吗？你们如何做好混合云的管理？

郭乾继：是的。对于混合云的管理主要有两个层面：管和用，其中**如何用好混合云更有意义**。

管，我们有自研的混合云平台，现在正逐步演变成一站式赋能平台，通过打通公有云和私有云资源，夯实云原生底座，整合了基础平台服务能力和开发平台服务能力。其中基础平台服务能力包括容器运行环境、CICD、高可用等，开发平台服务能力则包括消息队列、缓存、配置中心、注册中心等。由于事先根据业务场景建立了资源配置库，业务人员可以根据场景描述快速选择配置类型，实现定义好的应用架构部署模式，即自动化交付，提升 IT 交付效率，改善用户申请和使用 IT 资源的复杂度，实现一站式服务体验；运行期间则交给感知网。

感知网是我们利用开源技术打造的一套全栈全链路自主可控的监控工具，实现了集团原有商业软件的替代。它可以帮助管理人员图示化地看到整个系统的运行情况，同时，

这个平台也自动连接了我们的运维流程，可以把人、系统结合起来，出了问题，如果系统无法自愈，就会自动通知到相关的负责人来做闭环处理。利用感知网可以实现从底层硬件、操作系统、数据库、中间件，到应用 Web 服务器的实时运行状态监测，达成秒级告警和故障定位，帮助我们掌控 IT 运行状态。感知网上线后，成功支撑了日日顺物流 618、双十一等重大促销活动。

物联网生态模式下的安全保障体系构建

InfoQ：海尔集团数字化转型带给安全什么样的机遇和挑战？

李晓文：随着海尔集团 IT 架构不断演进，云化改造和数字化转型也在加速推进，安全的边界在互联世界里逐渐消退，攻击面也随之扩大，数字化业务将面临更多的安全风险，因此安全保障需要以更快的速度来适配业务的数字化能力，引进新思路、新框架、新技术、新模型帮助数字化生态系统在快速安全构建的同时，更加稳健、有序，以打造数字化业务的韧性。

InfoQ：海尔物联网生态模式下，如何更好地构建安全保障和能力？

李晓文：桥梁越大内部结构越重要。在海尔物联网生态模式下，安全团队需要一套与数字化业务能力相适配的打法，以构建全方位内生安全保障体系。主要包括以下几个层面：

- **风险管理**：网络安全的本质是风险管理，需要基于大数据分析来实现，在海尔的生态体系下，我们基于 Cyber Security Mesh 风险管理思路，快速敏捷地评估业

务中的风险，构建了动态感知和洞察风险的可视化管理能力；

- **将安全文化提升至业务决策层面**：在业务决策时，将网络安全作为一个业务问题而非技术问题，在数字化业务案例预研之时嵌入安全属性，将风险的处置能力左移至业务侧，在数字化过程中，确保安全能力被评估、识别和适配，使得安全投入有最佳的产出比；
- **采取主动防御措施对抗攻击者**：数字化转型的过程中，由于 IT 架构持续演进，基于新的 IT 架构，安全在容器、主机、网络、应用层面的基础能力需持续延伸和夯实；与此同时，需要构建先进的“雷达”能力，通过内部部署的 Sensor 收集并识别潜在威胁，以人工+自动化的方式确保拥有最新情报，通过增强 SOC 的大数据分析能力，积极防范和对抗更广泛的攻击者，并通过对重要资产主动进行持续的威胁监控，安全团队的“狩猎”团队能够转移攻击者对有价值的资产的攻击，具备能力在内部网络中与“敌人”过招；
- **安全的交付能力**：在安全能力建设过程中，夯实基础，沉淀安全的各项能力，打造高弹性组件化的安全一体化体系和能力，将安全的能力以组件化的方式快速交付至海尔不同的业务生态圈，适配不同业务的安全保障需求；
- **建立多方位的生态系统**：新冠肺炎疫情推动数字化加速转型，网络安全不再是安全产品简单的实施和叠加，全球网络互联互通，某一个组织或者企业亦无法独善其身，需要携手供应商、客户、技术供应商、监管机构、标准机构和行业协会共建，打造立体式主动防御的生态体系。

长在云原生架构上的小红书

作者：褚杏娟

采访嘉宾：张雷、高飞

“内容社区”已经成为小红书最被外界认可的“标签”，而作为一家互联网公司，支撑起社区运营、有着很多员工的技术团队更是不可被忽视的存在。小红书的技术团队又细分为后端基础架构、SRE、大数据、AI 算法、端技术、音视频技术和安全技术等团队。其中业务技术团队直接承接业务需求，帮助业务快速奔跑；中台技术团队则侧重更基础和长远的技术研发，并将新技术融合进业务当中。

正是小红书技术团队的多年努力，造就了当前小红书“土生土长”的云原生架构和更为先进的多样化内容分发算法。本文，InfoQ 专访了小红书技术负责人张雷和系统架构负责人高飞，揭开小红书技术团队的神秘面纱。

长在云原生架构上

小红书的云原生架构历史可以追溯到这个公司成立之前。

2013 年，Pivotal 公司的 Matt Stine 正式提出了云原生的概念，云原生开始大规模出现在公众视野，市场上也有了可用的 IaaS 产品。同年 6 月，小红书在上海成立。成立之初的小红书在搭建系统架构时便选择了同样刚刚发展起来的云原生，这也成为了小红书架构与众不同的地方。

“创业之初，选择云其实是很自然的事情。”小红书技术负责人张雷说道。

从成本角度考虑，起步阶段自建 IDC 太重。自建机房、自己管理及其高昂的运维成本，这些对于刚刚成立的企业来说是很大的负担，而云可以帮助企业省去这些费用。从业务角度看，虽然当时小红书并未上线电商业务，但电商却是计划中必定要做的事情。电商业务的特点之一就是对系统的弹性要求很高。对一家创业公司来说，为了满足促销活动带来的临时性资源需求而购买机器是不合理的，云却可以很好地解决这个问题。

如今，小红书的基础技术部会承担很多云原生技术的研发，这个部门包含中间件、存储、缓存、DB、SRE 和质量保障等不同团队。土生土长在云上的架构让小红书拥有很好的先发优势，也使得团队在对新技术的采用上少了很多后顾之忧。

经过 8 年不断地发展和升级，小红书整体架构的容器化率已经达到 80%，架构整体迭代效率大幅提高。

云原生架构带来的不止是自身迭代速度的提高，对于小红书这样对算法要求较高的企业来说，其算法模型从实验到上线的速度也得到了极大提升。

首先，得益于云原生架构，任何基于容器化的技术都可以实现硬件资源隔离。算法的模型训练和线上服务大量采用容器化技术后，研发人员可以不用太关心开发环境问题，而是更专注在算法迭代上。

其次，合理的架构与硬件的有机结合可以释放更多算力。随着 GPU、分布式计算架构等利用率的提高，算力也得到极大提高。这些算力无论在训练阶段还是线上服务阶段都可以提供更大的发挥空间。

最后，实时流式数据能够更好地支持算法的时效性。之前传统架构上的算法模型的更新时效性要以“天”为单位，但小红书基于流式数据，通过使用 Kafka 和 Flink，算法模型的更新时效性达到了“分钟”级别。

自研多样化内容分发算法

作为一个多元化的生活方式分享平台，如何做好内容分发是小红书技术团队面临的重要考验。

通常情况下，内容分发流程是这样的：用户上传内容，之后平台做内容理解，理解后做内容审核，审核合格的内容进入内容分发系统，分发系统进行推荐、搜索等操作，这些内容会得到曝光并被用户看见，用户再与这些内容产生交互，如点赞、分享或者评论等。系统通过捕捉到的交互行为，对用户喜好进行分析，之后再优化其内容分发策略。张雷表示，小红书与其他公司不同的地方在于，小红书在每个步骤上都更加重视多样化内容的曝光。

在张雷团队看来，生活方式有成千上万种，但与商品不同，生活方式不会存在“爆款”。因此，小红书拥抱多样化的内容分发是顺理成章的事情。“如果只对头部内容进行分发，那么曝光量多的内容会得到更多的曝光，而用户能感知到的多样性内容就会变少。中长尾内容的多样化分发对于生活方式平台是非常关键的。”

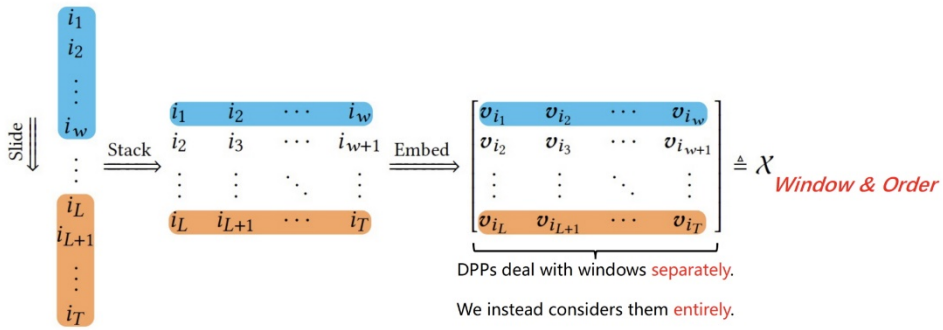
但是，整个过程中存在两个主要的技术难点：

第一，对于中长尾内容的表示和相似性度量。中长尾内容跟头部内容相比，用户交互

数据更稀疏。纯基于内容（CB, Content Based）的相似性度量依赖大量标注数据且不一定能反映用户感知的相似性。而稀疏的用户交互数据使得基于协同过滤（CF, Collaborative Filtering）的相似性度量方法也不太准确。

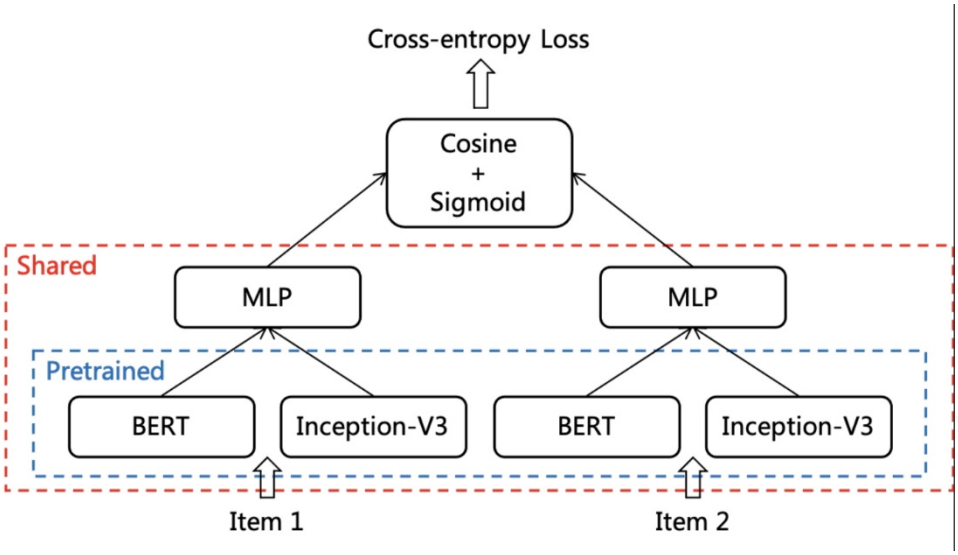
第二，如何去从技术上捕捉和表示用户对内容多样性的感知。之前经典的 DPP 方法是一个基于内容集合的方法，对内容出现的顺序不敏感。但是在 App 实际使用过程中，不同的内容出现顺序会很大程度上影响用户对多样性的感知。

针对上述问题，小红书采用了自己研发的滑动谱分解（SSD, Sliding Spectrum Decomposition）模型，该方法可以捕捉用户在浏览长项目序列时对多样性的感知。与其他多元化推荐方式相比，SSD 将内容序列视为用户观察到的时间序列，在整个序列中组合多个滑动窗口，以此对齐用户在浏览时的感知。



SSD 模型中多窗口堆叠的内容张量

针对中长尾内容相似性度量的问题，小红书研发并采用了通过内容学习行为的方法：CB2CF（Content Based to Collaborative Filtering）。



CB2CF 模式

CB2CF 通过神经网络，从内容本身出发去学习用户协同过滤的交互数据，并依此判断是否内容相似性。模型输入上仅使用内容，这样即使对于新内容和中长尾内容也能依赖模型的泛化能力得到较好的结果。模型目标上学习全体用户的协同过滤的结果，使得模型能够在统计上学习用户感知的相似性。

通过线上 A/B 测试，与 SOTA 的 DPP 模型相比，小红书 SSD 和 CB2CF 模型下的用户浏览时长提高 0.42%、互动率提高 0.81%，而 ILAD（用户浏览笔记之间的平均距离，即曝光多样性）提升 0.32%，MRT（用户平均阅读类目数，即消费多样性）提升了 0.68%。

在张雷看来，多样化内容分发从长远角度考虑会是一种趋势，但是否采用多样化内容分发还要取决于企业具体的业务形态。有的业务需要打造爆款，有的需要多元化，不

同的产品和业务对内容分发方式的偏好是不同的。

去年 2 月，小红书美食类消费 DAU 一度超过美妆，成为小红书社区第一大垂直品类。在内容运营和多元化内容分发机制的共同作用下，其他中长尾内容数量也迅速增长。公开数据显示，教育类同比增长 400%，科技数码类同比增长 500%，体育赛事同比增长 1140%，运动健身增长 300%。过去一年，小红书用户全年笔记总体发布量同比增长超 150%。同时，截至 2020 年 6 月，小红书月活跃用户数已经过亿。

向多云架构转型

随着业务规模的不断增长，小红书已经开启了从单云架构到多云架构的转型之路。

当前小红书对整体架构的目标有三点：第一，架构可以很好地支撑业务快速发展带来的规模的持续扩张，比如能够稳定支撑亿级 DAU 的规模；第二，能够做到较高的可靠性和可用性，这主要表现在跨地域容灾能力和跨云基础设施的容灾设计等方面；第三，架构必须是高效率的，这包括相对低廉的成本和较高的资源利用率。

这三个目标也是小红书做多云架构转型的动力。小红书架构负责人高飞表示，多云可以更加灵活地支撑更大的业务规模。不同的云技术特点不同，小红书可以根据不同云厂商的特点部署不一样的技术，如离线和在线的混布等。另外，多云对资源的冗余要求也更低一些，在容灾上有一定的效率优势。

“先进的架构和理念可以帮助一个起步较晚的企业实现弯道超车。”张雷表示。

据高飞透露，现在小红书团队基本用两个月的时间就可以完成搜索、推荐等核心业务

在另外一个云上的验证，同时小红书很多机器学习模型已经至少在四家云上进行训练。

当然，一旦拥抱多云架构，很多技术挑战也会接踵而至。

首先，多云架构需要统一的资源管理。多云上的资源管理需要做到像单云一样容易管理，否则很难统筹调配。其次，如何保证不同云之间数据的及时同步和一致性也是问题，尤其那些对数据一致性要求较高的业务对此要求更加急迫。最后，多云架构怎么做好稳定性、高可用，做好不同云之间的流量调度也是一个挑战。此外，小红书还有自己的要求：让自己的技术栈做到云独立，即不绑定在特定的云上，业务无论部署在哪朵云上都可以跑得通。

面对这些技术挑战，除了利用现有的开源技术外，小红书也会进行自主研发。

针对资源管理问题，小红书联合华为、工商银行和中国一汽开发了 Karmada 开源项目。Karmada 是一个 Kubernetes 管理系统，可跨多个 Kubernetes 群集和云来运行云原生应用程序，而无需更改应用程序。对于数据一致性问题，小红书会在数据存储和缓存层基于分布式一致性协议，结合不同的业务场景，进行自主架构设计和研发。

为加强多云架构稳定性，小红书使用混沌技术定期进行故障演练，保证一个机房出现故障时可以快速切换到其他机房，同时对不是特别重要的服务进行降级处理。而对于云独立问题，由于要摆脱对单个云厂商的依赖，一些 PaaS 能力必须自研。小红书的 KV 存储、控制面，甚至整个微服务架构都是自研。

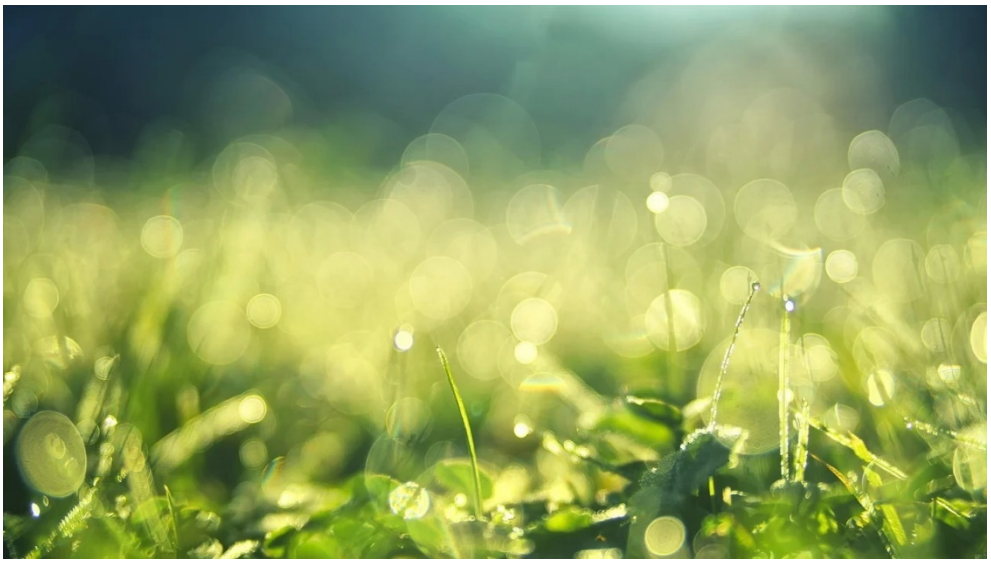
管理层面，小红书由各个专业方向的架构师们组成了技术专业委员会，根据不同技术领域制定相应的技术规划和规范，以此来提高迭代效率并保证产品质量。

对于未来，小红书的目标很清晰：随着业务的增长，多云架构必将登上舞台。真诚的、令人向往的、多元化的内容规划一直没变，更注重 UGC 中长尾内容分发的策略也没变。技术团队的主要任务就是努力完成这些目标。

“一切都在进行中。”张雷说。

Apache Kyuubi PPMC 燕青：为什么说这是开源最好的时代？

作者：凌敏



在大数据领域， Apache Spark 早已成为最炙手可热的计算引擎。随着 Spark 两年磨一剑，正式发布 3.0 版本，带来诸多新特性的 Spark 更是拥有了无限想象空间。不过对于用户而言，平台的技术门槛始终是个不小的挑战。也正因如此，不少项目选择直接建立在 Spark 之上，通过将平台的能力统合，并引入新的特性，从而降低用户使用门槛，实现大数据价值的最大化。

Kyuubi 正是这样一个拥抱 Spark、高性能的通用 JDBC 和 SQL 执行引擎，由网易数帆旗下有数大数据团队开源。Kyuubi 提供标准化的接口，赋予用户调动整个数据湖生态的数据的能力，使得用户能够像处理普通数据一样处理大数据。

6 月 21 日，Apache 软件基金会宣布，Kyuubi 以全票通过的表现，正式进入 Apache 基金会孵化器。这也侧面证明了 Kyuubi 的受欢迎程度。

近日，InfoQ 有幸采访到了网易数帆技术专家、Apache Kyuubi PPMC、Apache Spark Committer 燕青，和他聊了聊 Kyuubi 一路发展背后的故事，以及他对于开源的理解。

“九尾狐”Kyuubi

Kyuubi 的命名源自中国神话《山海经》，意为“九尾狐”。狐会喷火，象征 Spark；狐有九尾，类比多租户。这个命名也体现出了 Kyuubi 系统设计之初的主要目的——在 Spark 上实现多租户。

事实上，在 Kyuubi 之前，市面上已存在具备类似能力的产品，比如 Spark ThriftServer（简称 STS）。这是 Spark 社区现有的、基于 HiveServer2 实现的 Thrift 服务，旨在无缝兼容 HiveServer2。

虽然 STS 的性能极佳，但当前并不完善，尤其在企业场景下存在较多短板。比如，单 Spark 应用实现的 STS 并不能完整支持多租户，因为 STS 本质上是一个 Spark Application，整个 Application 只有全局唯一的用户名，并同时包括 Driver 端和

Executor 端。而对于像网易这样有多条产品线的互联网公司来说，每条产品线的数据在一定程度上是隔离的。因此，只有支持多租户才能满足公司对于数据安全、资源隔离、高可用以及高并发的要求。

这也就促使网易内部开发了 Kyuubi。Kyuubi 在统一接口基础上，拓展了 STS 在多租户模式下的使用场景，并依托多租户概念获得了完善的资源隔离共享能力和数据安全隔离的能力。

在 2018 年上线之初，Kyuubi 的定位只是在 Spark 上实现多租户，再引入一个比较细粒度的权限控制，做一个小而美的系统。因此，Kyuubi 的第一代架构主要面向的是 BI 产品。但是正式上线并开源后，团队发现用户并不在意系统本身的设计初衷以及使用场景是什么，所以当一些用户使用 Kyuubi 做 ETL 等工作时，用起来磕磕绊绊，反馈也不是特别好。

苦苦挣扎一年后，团队决定深入 Spark 社区去贡献，从中找到设计一套比较通用的、面向更多用户的架构方案灵感。这期间，Spark 也正在酝酿一个大版本的跨越式升级，从 2.4 直接来到 3.0 版本。Kyuubi 团队在其中参与了很多工作，并一同推动 Spark 社区发展。

“在这个过程中，我们对 Spark 的内核机制有了一个更加充分的了解。在这个基础上，我们意识到我们对于 Kyuubi 原本的构想是不太可持续的，Kyuubi 的第一代架构视野比较小，应用场景也很少。因此，我们觉得是时候对 Kyuubi 的架构进行一番革新了。”燕青回忆道。

2020 年，团队重新设计了 Kyuubi 第二版的架构，新架构的使用场景更加丰富。具

体来看，Kyuubi 的使用场景主要包括以下三个方面：

1. **替换 HiveServer2，轻松获得 10~100 倍性能提升**。

- Kyuubi 高度兼容 HiveServer2 接口及行为，支持无缝迁移；
- Kyuubi 分层架构，消除客户端兼容性问题，支持无感升级；
- Kyuubi 支持 Spark SQL 全链路优化及再增强，性能卓著；
- 高可用、多租户、细粒度权限认证各种企业级特性统统都有。

2. **构建 Serverless Spark 平台**。

- Serverless Spark 目标绝对不是让用户调用 Spark 的 API、继续写 Spark 作业；
- 通过 Kyuubi 预置的 Engine 模块，用户无需理解 Spark 逻辑，入门门槛极低；
- 用户只需通过 JDBC 及 SQL 操作数据专注自身业务开发即可，资源弹性伸缩，0 运维；
- 支持资源管理器（Kubernetes, YARN 等），Engine 生命周期，Spark 动态资源分配 3 级不同粒度全方位的资源弹性策略；
- 支持 YARN/Kubernetes 多种资源管理器同时调度，保障历史作业安全迁移上云；
- Spark 自适应查询引擎（AQE）及 Kyuubi AQE plus，提供澎湃动力。

3. **构建统一数据湖探索分析管理平台**。

- 支持 Spark 所有官方数据源及第三方数据源；
- 支持 Spark DSv2 元数据管理，直观进行数据湖构建及管理；
- 支持 Apache Iceberg/Hudi, DeltaLake 等所有主流数据湖框架；
- 一个接口一个引擎一份数据，提供统一的分析查询、数据摄取、数据湖管理平台；

- 批流一体，支持流式作业（Upcoming）。

当前，Kyuubi 不仅在网易内部承接了大量工作，在业内也有多家大型公司采用 Kyuubi 解决问题。“Kyuubi 新架构做出来之后，很多其他公司的小伙伴也开始活跃地参与到这个项目中来，慢慢地参与的人多了以后，我们就想着是不是可以去 Apache 软件基金会进行孵化。”燕青说道。

独行者速，众行者远

2021 年 3 月份，团队开始正式筹备将 Kyuubi 捐赠给 Apache 软件基金会。事实上，拥抱 Apache 软件基金会的念头一早就扎在 Kyuubi 团队成员心中了。在燕青看来，做下这一决定背后的原因主要有以下 3 点：

- 团队以及公司都有意愿去做这件事情。网易内部本身始终贯行开源开放的策略，所以当团队和公司提出这个想法后，公司内部也给予了一定的支持，比如为项目配套了一些运营力量等等。
- 随着第二版架构的登场，Kyuubi 发展路线日益清晰，也有越来越多的开发者参与到项目中来，社区规模逐步扩大。
- 团队成员发现一些潜在的 Kyuubi 用户或是开发者产生 IP 产权方面的顾虑。此前即便 Kyuubi 已经开源，但依旧是属于网易的产品，将项目捐赠给 Apache 软件基金会后，可以消除这些开发者的顾虑，进一步吸引更多的开发者参与其中。

6 月 21 日，Apache 软件基金会宣布，Kyuubi 以全票通过的表现，正式进入 Apache 基金会孵化器。根据投票结果，Kyuubi 获得了 13 个约束性投票(binding votes)和

8 个无约束性投票(non-binding votes)，投票全部持赞同意见，无弃权票和反对票。

在走向 Apache 软件基金会的过程中，燕青坦言 Kyuubi 很幸运地得到了来自公司内外的不少帮助。

“有些帮助是无形的，有些帮助是有形的”，燕青说，“Kyuubi 能够成功进入 Apache 基金会孵化器，很大程度上是因为那些开源前辈们已经把路铺好，很多前辈在国内开源领域深耕多年，比如姜宁老师等等，他们在前期做的一些努力让我们后人能够乘凉。理论上来说，现在的项目要想进入 Apache 软件基金会孵化，比他们那个年代要更加容易一些。

“另外，姜宁老师发起并创立的 ALC Beijing 也会帮助我们国内开发者或企业去孵化一些本土项目，比如他们会把一些 Apache 软件基金会的文档翻译成中文，这对于英文不太好的开发者来说帮助非常大。”

在这个过程中，Kyuubi 也得到了不少来自 Apache 软件基金会的支持与帮助。“Apache 的指导文档本身就对我们产生很大的帮助。如果没有这个文档，你根本不知道这一步做什么，下一步做什么。Apache 的指导文档写得非常详细，就算没有其他人的帮助，只要你耐心地把这些文档阅读下来，也能更好地完成项目捐献。”

在 Kyuubi 项目官宣进入 Apache 孵化器的那天，Kyuubi 特别感谢了很多给予过自己帮助的人，比如给 Kyuubi 提供指导的 Champion 和 Mentors 姜宁，Mentors 章剑锋、张铎、Akira Ajisaka。也有为 Kyuubi 提出 issue 和建议的伙伴们，以及为 Kyuubi 做出贡献与支持的国内外数十家企业用户。

独行者速，众行者远。而这，也是开源一贯传承的精神。

发展社区的关键在于多倾听开发者的声音

现在，距离 Kyuubi 进入 Apache 大家庭已有月余，至于进入 Apache 之后发生了哪些变化，燕青认为主要体现在项目管理和社区管理两个方面。

在过去，Kyuubi 归属于网易，项目管理者相当于拥有超级管理员的权限，可以直接对项目做决策，发版时间也不固定。进入 Apache 孵化器后，项目有发版等重大决策时，需要在邮件列表里一起讨论，并发起投票。内部沟通讨论结束后，还需要在孵化器邮件列表里面再去发起讨论、投票，此外，Apache 软件基金会方面也会帮忙做项目审查，看下在法律或是其他方面是否合规。相较过去，虽然整个决策链路长了一些，但这样的层层讨论对项目本身来说更负责，借助这些外脑，项目的发展路线也愈加清晰。

在社区管理方面，燕青坦言“以前我们比较随意，现在进入 Apache 软件基金会后，我们定了一个目标就是要把社区建立起来。"在燕青看来，**社区是由人构成的，要想让社区建立并发展起来，需要多倾听社区的声音，多去了解这些个人开发者或是公司的需求**。

至于如何提高社区的活跃度，燕青认为最重要的一点是要**尽量避免犯一些错误**，要以开放友好且包容的态度去接纳开发者。对于社区新人，要更加有耐心，给他们时间去成长。此外，在宣传方面还需要加大力度，社区运营者也需要积极地去配合宣传。

最后在项目上，要更加注重品控的把握。“如果 Release 版本不太稳定的话，从用户层面来说，会劝退很多用户"，燕青解释道。

Kyuubi 的终极目标：让大数据平民化

对于 Kyuubi 的未来，燕青也有着很高的期待：希望建立在 Apache Spark 和数据湖技术之上，统一门户，打造一个理想的数据湖管理平台，让用户处理大数据能像处理普通数据一样轻松。

具体来说，第一要增强 Kyuubi 对 Kubernetes 云原生的支持，让 Kyuubi 提供的服务以及计算资源都可以在容器中进行；第二要增强 Kyuubi 对数据湖的支持，让用户能够更简单地管理、使用和建设数据湖，实现所见即所得；第三要继续深入优化引擎侧，增加流式场景，打造批流一体的平台。

“未来，希望 Kyuubi 可以让 Spark、让大数据平民化”，燕青憧憬道。

“这是开源最好的时代”

除了是 Apache Kyuubi 的作者，燕青还有多个身份，比如 Apache Spark Committer、Apache Submarine Committer。从业多年，燕青始终在开源大数据领域深耕，对他来说，“开源是可以做一辈子的事情”，而他也始终热爱开源，信仰开源。

在燕青看来，开源非常利于个人学习和提升自己。“很多开源大神把自己的代码或 idea 放在 GitHub 或其他开源平台上面，如果你选择在某一技术领域深耕的话，可以在这些平台中找到很多学习资料，边学边做”。

通过在开源社区的不断学习，燕青也实现了从医学信息学博士到顶级开源社区核心贡

献者的身份转变。

提及自己积累下来的经验，燕青觉得**在持续学习之外，也要拥有良好的心态，以及明确且专注的技术方向**。“开源是一个圈子，这个圈子是开放的，但从个人的角度来看，每个人的想法不同，圈子外的人或多或少会感觉到有一堵墙的存在。如果你能够拥有良好的心态，谦虚、友善地去和社区沟通，自然也能换来同样友好的回应，从而形成良好的回路。

此外，从技术角度来讲，方向一定要明确，专注在一个领域内深耕，从而慢慢从不擅长变得擅长。比如我自己就是一直围绕着 Spark 这个项目去深耕，当我在这个项目里做得差不多的时候，我并不是离开它，而是继续在这个项目里去帮助别人，包括我做的 Kyuubi 也是围绕着 Spark 去做的。”

当前，虽然国内开源发展存在进步与乱象并生，机遇与挑战共存的状态，但燕青坚信“这是开源最好的时代”。

在其看来，很多开源前辈已经在这个领域做出了长时间的铺垫和积累，很多热爱开源的开发者也都投入其中，大家都在为开源贡献出自己的一份力量，帮助它朝着更好的方向去发展。

“从某种程度上来讲，开源氛围好与坏是在描述一个圈子，是圈子就会有一堵无形的墙，有时墙里面的人会翻出去把墙外的人引进来，有时墙外的人会主动地进到墙里。其实只要你学会融入它，帮助它去改善，就一定会朝着好的方向去发展。另外，从开源角度来讲，有些项目正确地衰落，其实正说明我们在往更正确的方向走去”，燕青说道。

对于国内开源的未来，燕青期望可以有越来越多的优秀开源项目诞生自高校，让开源氛围更加“年轻化”。监管层面，也希望有一些相应的规范，帮助开源更好地持续发展。

嘉宾介绍：

燕青，Apache Kyuubi PPMC，Apache Spark Committer，Apache Submarine Committer。目前就职于网易数帆有数大数据团队，专注于开源大数据领域。

多媒体内容如何防伪防盗？揭秘阿里安全团队数字水印技术

作者：万佳

如今，数字化媒体内容充斥着人们生活的方方面面，但是却因其易复制、易分发的特点，而饱受盗版问题的困扰。比如，《画皮》、《泰囧》、《捉妖记》、《寻龙诀》和《老炮儿》等多部热门电影都发生了首映场即遭盗录的问题；2019 年春节档，《流浪地球》、《飞驰人生》和《疯狂的外星人》等 8 部国产电影仍在上映阶段，盗版资源就在网上出现，预计共造成了约 7.87 亿票房损失。除了盗版外，篡改、造假、盗用等侵害媒体知识产权行为的门槛也变得越来越低，且方法越来越多样化。以上种种不仅严重打击了创作者的积极性，也给整个媒体产业和市场造成了恶劣的负面影响。

针对上述问题，数字水印技术成为一种行之有效的重要技术解决方案。什么是数字水印技术？它有什么特点？它又是如何解决盗版等问题的？带着这些问题，InfoQ 记者采访了阿里媒体安全技术研究团队负责人屏翰和团队成员越永、渡明、乐仙。

一支“术业有专攻”的技术研究团队

InfoQ：阿里媒体安全技术研究团队成立于什么时候？主要研究方向和领域是什么？

屏翰：团队成立于 2017 年初，从最初的一两个人发展到如今的十几人，其中博士后 2 人（均有访学经历），博士 6 人，团队其他成员均为 985 重点高校硕士。

目前，团队研究方向主要有两大块：一是数字水印，二是多媒体取证。数字水印领域，细分为音频水印、视频水印、图像水印、文档水印、网页水印等；媒体取证领域，细分为原图识别、篡改检测、篡改定位和来源识别等。

InfoQ：阿里媒体安全技术研究团队成立的背景是什么？

屏翰：阿里的业务板块比较多，应用场景复杂，集团内部有很多媒体安全场景的需求，包括图像、文档、音频和视频等。并且，不同 BU 之间的很多需求存在一些共性，所以抽象总结研究通用的技术和解决方案。

但是，在实践过程中，我们发现媒体安全技术在知识产权保护领域有更广的应用，可以发挥更大的作用。举几个例子，像商品图，商家聘请专业模特，找专业摄影师拍摄宣传图，但可能被别人直接盗用；为蒙混过关，有人把核酸检测报告进行 P 图，把阳性改成阴性；商家入驻淘宝或天猫平台，上线产品都要提供一定的资质证明，比如营业执照、许可证，如果销售国外品牌，还要提供总代理资格，如果商家没有，可能通过 P 图造假。这些场景都能用到媒体安全技术。

尤其是数字水印技术，它可以在知识产权保护中发挥重要作用。知识产权保护的核心不仅仅是把东西保护起来，重点是让这个东西能更好地传播和分享、被使用和消费。

InfoQ：作为团队负责人，您如何带领团队开展技术研究和开展跨部门协作？

屏翰：我们团队有十几个人，在专业背景上，团队成员之间互补，我学数学专业，对密码学比较熟悉，而越永的背景是视频编解码，对视频很熟，渡明是音频，乐仙是社交媒体。

数字水印已有二十多年历史，但真正大规模应用不多，因为没有找到好的应用场景。或者有需求时，没有合适的技术。对我们做水印研究的人来说，阿里的一大优势是场景非常丰富。如此，我就知道水印技术可以用在哪，而不同的应用场景对技术的要求不一样。知道应用场景是什么，我就结合应用场景发展技术，这样技术就有不同的能力，提供不同的方案满足应用需求。如此，技术和业务处于正循环，互相促进发展，不仅团队实力可以壮大，而且技术能力也能逐步提高。

从技术能力上讲，如果你不知道这个东西有什么用，或者没有应用场景做检验、验证，那么技术能力就无法得到提高。比如，从学术角度研究，你可能觉得鲁棒性要做得多么好，但是业务方可能希望在满足质量的前提下，大幅提升效率。像商品图，淘宝上一天的商品图过亿，学术界的技术可能没考虑过支持商业的大规模应用，所以学术界跟实际应用离得稍微远一点，而我们结合得更紧密一些。通过真实的业务场景应用，不断检验，提高我们的技术能力。

在协作上，我们与业务方、客户不断在磨合。它们对技术的期待越来越高，这也导致我们的技术能力要与时俱进，支持它们。同时，还有对抗问题，比如与黑灰产作斗争，我们加了水印，它们试图去掉水印，给你搞破坏。在这个过程中，我们的技术能力也在不断发展和进步。

数字水印技术的现状

InfoQ：目前，业界有哪些主流的数字水印技术方案？它们的不同点是什么？

越永：在图像水印领域，Digimarc 公司是图像水印技术的行业标杆。该公司在这个

领域深耕多年，拥有 1200 多项专利，其图像水印技术涵盖数字和实体图像。Digimarc 利用自己拥有专利的数字水印技术和信号处理技术在包装、打印材料和图像中增加不可感知的水印。这项技术拥有充足且可扩展的容量来应对广泛的商业应用需求。据我们了解，Digimarc 公司的主要业务集中在条形码的推广和应用。

在音频水印领域，Verance 公司旗下基于音频水印的内容保护产品 Cinavia 比较有代表性，它是电影领域蓝光标准的一部分，UHD/4K 内容的行业要求，并被混合广播宽带电视（HbbTV）协会在宽带应用规范中采用，也是电影、电视和音乐领域中内容保护的商业标杆。这项技术专注高级影视内容的版权保护，使其在预发行、院线上映期间和电视播放时免受盗版的侵害。

在视频水印领域，Nexguard 的技术在国际上比较流行，他们通过视频水印技术实现多媒体内容的版权保护和反盗版服务。Nexguard 的主要业务围绕付费电视内容保护展开，包括点播电视节目、体育赛事直播、互联网直播服务、互联网电视等。他们的付费电视服务（NexGuard Pay-TV）将特定用户的取证水印嵌入到托管的付费电视客户端设备（机顶盒或智能电视）中，几乎所有的主要芯片组供应商都支持该技术，并且不需要额外的前端视频处理。

而在文档水印上，宇飞的文档水印技术则比较出众。他们的技术包括覆盖于全文档并与页面进行融合的网纹水印；人眼不可识别，但是计算机可识读的离散可见水印；可通过滤光镜看见的非离散实体水印；彩色方案偏光性水印；以及用于验证的水印二维条码。宇飞的水印技术不仅适用于电子文档，还适用于打印（印刷）的实体文档，并且综合考虑了印刷品防复制、防破损、防模糊、快速检测等需求。

InfoQ：能介绍一下你们在数字水印技术方面的研究情况吗？

屏翰：阿里自主研发了一套完整的水印技术体系，从载体上看，我们的水印技术覆盖图像、文档、音视频、网页等几乎全部的数字媒体；从技术角度看，我们的水印在载体和水印预处理、水印嵌入和提取、消息机制等所有操作阶段都拥有相应的技术实现；从应用角度，我们的水印技术包含版权保护、追踪溯源、内容认证等相关解决方案，并且还在寻找新的技术突破点；从业务角度看，我们的水印技术在解决权利纠纷、追踪传播路径、感知安全风险、支持用户增长、防伪、内容认证等问题上已经发挥一定的作用。

数字水印技术的难点与落地

InfoQ：在数字水印技术上，你们认为比较大的挑战有哪些？

屏翰：主要有五个方面：

第一，技术可靠性。存在未知组合攻击、水印的安全性、不可感知性与高鲁棒性的矛盾、水印技术标准和认证不够完备等问题。比如，攻击者可能利用其它未知攻击手段对载体内容进行处理，在保证载体失真较少情况下令水印提取失效；在获取大量嵌入水印前后的样本情况下，攻击者可能估计出水印的嵌入空间，从而进一步对水印去除或替换；抗屏摄的水印技术在能做到无对比图情况下不可感知，但与原图对比仍存在视觉差异。

第二，社会的应用。大众对水印的能力边界不清晰、数据多样性带来的问题、需要专业的开发、部署和运营人员。比如，有些用户希望仅凭借水印就想得到 DRM 的所有功能；文档水印应用中可能出现某些特殊文档无法嵌入水印的情况。

第三，法规的完善。水印的取证环节链路过长、水印的法律边界还不够清晰、水印相关法律法规还在完善中。比如，水印取证需要司法鉴定机构全链路参与，对个人用户成本过高；从盗版内容中提取出水印作为版权所有权证明在法律界尚未完全达成共识。

第四，新媒介、新技术。载体属性的调整、压缩标准的更新、传输信道的变化。比如，视频发展从 1080p 到 2k 到 4k 再到 8k，从 SDR 到 HDR，视频水印技术需要进行适配；图像、视频的压缩标准更新迭代，特别是实时流压缩协议，对已有水印技术的鲁棒性和隐蔽性带来挑战。

第五，新需求、新场景。新应用场景不断涌现、水印性能需要随业务变化进行调整。比如，水印用于媒体桥、第二屏、热词识别、签到打卡等新场景；图像水印用于信息传递需要穿透不同的社交媒体平台，提出新的鲁棒性需求。

InfoQ：数字水印技术在阿里最先落地的业务场景是什么？有什么样的效果？

越永：最早是短视频和影视版权保护，比如版权影视剧的交易和运营。版权商把视频内容卖给运营商，但也会担心运营商泄露影视内容或运营商员工参与盗版，所以需要使用数字水印技术。之前，版权方会在市场上购买水印技术解决方案，但是国外公司报价非常高。

用了我们技术后，它就不用采购国外垄断公司的技术方案，一年可以节省几百万。我们利用数字水印技术解决了两个问题：一是链路上，哪个地方容易被切入，插入盗链，成为盗版的开口。同时，在分发时，我们可以追溯到影视是从哪个渠道或哪些点泄露；二是，在终端上，找到泄露的视频泄露节点。

其次是商品图。以前通过员工来识别图像是否 P 图，或者把这个工作外包出去，但是

无论怎样它都有人工成本，一年大概在千万级左右。同时，如果这张图是假的，但是没识别出来，这张假图就会给我们带来资损成本。针对商品图，数字水印技术已在阿里落地三四年。我们先在内部进行灰度测试，然后开放给商家。如果商家想用，可以加入知识产权保护计划，但我们会提前告知对方，如果使用这个技术，它对图片质量或多或少有点影响。因此，商家会自行评估。一旦加入计划，如果有其他商家盗用商品图，那么他在上传图片时就会被系统自动拦截。

目前，我们的数字水印累计完成 70 余篇专利申请，接入应用数超过 700 个，覆盖办公、业务、娱乐等平台。同时，水印服务累计调用量近 60 亿次，覆盖过亿文件，在内部，文档、网页和图像是调用量最大的，但音频、视频的调用量也不少；针对外部平台，则以图像水印调用为主。

InfoQ：从技术到产品再到落地，你们如何研发出真正有市场价值的数字水印产品？

乐仙：数字水印技术在阿里内部有两种落地方式。一方面，技术驱动，我们研究一项新技术会根据它的性能指标去评估其应用场景和业务价值，然后主动去跟业务方沟通，看它们是不是对我们的技术感兴趣，我们的技术能不能提升它们的业务能力或解决它们棘手的问题。

另一方面，我们会通过内部的技术交流平台宣传我们的数字水印能力，有业务方看到我们的技术，了解了我们的能力后，就会主动联系我们。大家一起探讨如何通过数字水印技术满足业务方的需求。

通过这两种方式，我们从无到有，慢慢将我们的技术在阿里内部铺开。当我们的技术得到应用后，接下来就是性能验证、技术迭代的过程。

渡明：从研究到具体的业务场景，它们之间差别很大，像我们早期研究的数字水印技术，它的 DEMO 或雏形离真实场景有一定的距离。因此，我们要去做适配或调整优化。

越永：先有技术，再有产品，最后落地，这是一个不断迭代的过程。它越往后走，情况越复杂，可能到落地时，每个人都有自己定制化或个性化需求，反映到产品上就是有各种各样的形态。从研究层面，数字水印要解决很多技术维度的问题，包括质量、效率和不可见性，到落地时，业务方可能只关注一点，把这一点做到极致、最好。

回归到产品上，产品每个点都要做到最好，但这种产品不存在。我们需要不断地去打磨，做出取舍和权衡。

InfoQ：从技术到产品再到落地，你们认为哪些环节比较难？

屏翰：一方面，数据多样性问题导致部分特例数据无法嵌入和提取，比如文档水印的应用场景中可能出现某些特殊文档无法嵌入水印。如果应用开发人员的知识或能力不足，可能会误用系统造成内容感知质量过低。另一方面，大众对数字水印技术尚不了解，可能出现适用于这项技术的场景没有使用数字水印，但不适合数字水印的场景盲目接入数字水印。

从我的经验看，两头比较难，一头是研究上，从 0 到 1 去创造全新的技术；另一头是真正落地，满足千变万化的现实场景需求。

先说研究。首先，你自己要有足够的积累，科研是“站在巨人的肩膀上”，你先要爬上巨人的肩膀，这需要不断积累、广泛阅读和大量实验，对前人的工作和业界发展水平了如指掌。其次，就是更难的地方，要研发属于自己的技术，虽然路子很多，但是都

不容易。

比如，如果发现了现有技术的不足，那么你就要有解决这些技术缺陷的能力；或者，你有自己的创新，有一些灵光一现的想法，这就是所谓的“打入了技术无人区”，你会面对很多之前从未遇到的问题，依靠自己和团队的力量一步步解决问题。或者，你发现了许多问题的底层共性难题，解决这个问题能提升一批算法的性能。

落地的难点主要有两方面：一是现实场景确实太多，太复杂，业务需求“千差万别”，另一方面是要跟人打交道，需要能在“同一个频道”里沟通对话。我们要调研应用链路的每个细节，再根据客户要求，评估和调整我们的技术，制定通信和接口模式，尤其是面对外部客户时，还要考虑技术安全问题。并且，我们要面对各种各样的攻击，因为水印技术主要用于知识产权保护，无论是黑灰产，还是泄密者，他们会采取各种各样的方法攻击你的水印技术。

此外，业务方有时并不能理解技术的细节，所以有时会挑战技术的性能极限，比如“你这个东西要达到 99%的准确率，我们才用”，但是这项技术的理论极限可能是 90%，甚至更低。对业务方来说，80%、90%、99%只是一个数字，但是技术越往后，每提升一个点的准确率，背后的代价都是巨大的。

InfoQ：如今，视频很流行，尤其短视频特火，在这个领域，你们的数字水印技术研究情况怎么样？

越永：我们在长视频和短视频领域都有相应的技术能力。先说长视频，其业务形态比较成熟，对应的技术要求也比较成熟。而短视频是新的媒体形态，它有一些新的保护需求，比如长视频切成短视频，短视频里切条抽出来，再从视频里扣出一块区域，它

的技术难度可能更高。

长视频，好莱坞要求 15 分钟提取水印信息，而短视频只有几秒或十几秒的水印提取时间，这是两者的差别。当然，并不是需要的时间越短，技术越难，因为还要考虑质量，长视频是 15 分钟提取出来，必须把视频质量保持得非常好，短视频比如几秒钟把（水印）提取出来，质量相对来说稍微差一点，这是可以接受的。

渡明：我稍微补充一下，因为长视频通常是 PGC 模式，就是专业机构生产，通常其价值都特别大，因为投入了很大成本，可能需要专业团队才能制作，包括电视剧、电影或纪录片，仅仅一集投入的资金就非常高，内容的价值也非常大，一旦被泄露或盗版，对生产者的损失非常大。所以，它们对版权保护的诉求非常高。但是，短视频可以批量生产，投入成本低，制作门槛不高，对其的保护意愿不会特别高，除非有一些精品的短视频。

所以，从水印需求上看，长视频的诉求更强，而短视频中的头部短视频内容可能也有比较强的保护意愿。

InfoQ：做数字水印技术研究经历中，有哪些让你们印象比较深刻的事？

渡明：自己研究或尝试一些新方法的过程中，你可能卡壳了一天或两天，寻找各种问题，一直找不到，就卡在那，陷进去了，回头都回不来，不知道问题在哪，突然有一天灵光一闪，知道了问题所在，很快研究局面就打开了，顺利验证。这种豁然开朗的感觉是令人印象最深刻的。

屏翰：我们的水印技术和取证技术在某个集团关键项目中得到应用，帮助业务方解决了一个很大的痛点，而这个问题困扰了业务方七到八个月。技术上线后，业务方看到

了效果，对我们平台也更有信心。这种“多赢”的效果和技术的真正作用，让我们很高兴，很有成就感和技术人的自豪感。

数字水印技术领域的新尝试

InfoQ：你们目前在数字水印上正在做哪些新尝试？

屏翰：大致七个方面：

第一，水印安全与攻防，包括抗隐写分析的鲁棒水印、水印的盲盒攻击方法。

第二，信息论模型，包括基于信息论研究水印的安全性、鲁棒性等，结合信息论构建完善的水印方法理论框架。

第三，基于深度学习的水印，包括基于深度学习的水印嵌入提取方法、深度学习结合信号处理的水印方法。

第四，神经网络水印，包括网络模型的水印嵌入提取方法，水印对模型性能的影响、不同水印的相互影响。

第五，加密域水印，包括在加密信息中嵌入数字水印，数据全生命周期无死角安全防护，加密信息的标注、验证和防篡改。

第六，数据库水印，包括数据库防篡改水印、数据库溯源水印。

第七，水印结合区块链，比如水印技术和区块链技术相互融合。

扫码关注InfoQ公众号

Geekbang. 极客邦科技 | InfoQ