

深度信息技术（精品）专辑（第七期）

- 龙芯基础软件生态建设思路
- Deepin 社区和 UOS 桌面操作系统
- OpenCloudOS 操作系统社区
- 我国工业软件发展之困
- 工业大数据软件与开源创新

中国开源软件推进联盟（COPU）

二〇二二年一月十八日

目 录

■ 自主架构生态建设	
胡伟武：龙芯基础软件生态建设思路.....	3
■ Deepin 社区和 UOS 操作系统	
刘闻欢：统信基于 Deepin 社区发展商业版 UOS 桌面操作系 统	13
■ OpenCloudOS 操作系统社区	
OpenCloudOS 社区：OpenCloudOS 操作系统社区介绍.....	21
■ 自主工业软件研发	
中国工业技术软件化产业联盟：我国工业软件发展之困.....	31
本刊编者 ¹ ：2021 中国工业软件年度企业排行 TOP100 发布.....	39
■ 工业大数据	
王建民：工业大数据软件与开源创新.....	42

¹ 本刊编者为：陈伟、鞠东颖。

自主架构生态建设

龙芯基础软件生态建设思路

中科龙芯 胡伟武

构建独立于 Wintel 体系和 AA 体系的安全可控的信息技术体系是国家战略，也是龙芯的发展目标。经过 2001 年以来 20 年的努力，龙芯公司于 2021 年正式推出的龙芯 3A5000 的通用处理性能已经逼近市场主流产品的水平，正在研制的下一代龙芯 3A6000 及 3A7000 系列性能将达到市场主流产品的水平。龙芯自主生态建设的主要瓶颈从 CPU 性能不足转向软件生态不够丰富。本文从指令系统、操作系统、二进制翻译、工控操作系统、开源社区等方面介绍龙芯软件生态建设思路。

一、推出自主指令系统 LoongArch

指令系统是 CPU 所执行的软件指令的二进制编码格式规范，是计算机的软硬件界面，是信息产业的重要基石。一种指令系统承载了一个产业生态，如 X86 指令系统承载桌面计算机和服务器产业，ARM 指令系统承载手持移动终端产业。就像中国人可以用英文写文章，但不可能基于英文发展中华民族文化，中国的信息产业可以基于国外指令系统做产品，但不可能基于国外指令系统构建安全可控的信息技术体系和产业生态。

2021年，龙芯中科基于二十年的CPU研制和生态建设积累推出了龙芯指令系统架构LoongArch，包括基础架构部分、虚拟化、二进制翻译、向量指令等扩展部分，近2000条指令。龙芯中科从2020年起新研制的CPU均支持LoongArch架构，包括3A5000、3B5000、3C5000及后续型号。

龙芯架构具有自主研发、技术先进、兼容生态三方面特点。龙芯架构从整个架构的顶层规划，到各部分的功能定义，再到细节上每条指令的编码、名称、含义，在架构上进行自主设计，具有充分的自主性。龙芯架构摒弃了传统指令系统中部分不适应当前软硬件设计技术发展趋势的陈旧内容，吸纳了近年来指令系统设计领域诸多先进的技术发展成果，不仅在硬件方面易于高性能低功耗设计，而且在软件方面易于编译优化和操作系统、虚拟机的开发。龙芯架构在设计时充分考虑兼容生态需求，融合了各国际主流指令系统的主要功能特性，同时依托龙芯团队在二进制翻译方面十余年的技术积累创新，能够实现多种国际主流指令系统的高效二进制翻译，高效运行X86、ARM等平台上的二进制应用程序，从而达到生态融合的目的。

龙芯自主指令系统的推出，是龙芯中科长期坚持自主研发理念的重要成果体现，是在完成CPU性能补课、全面转向生态建设历史关头的重大技术跨越，标志着龙芯中科在自主信息技术体系和产业生态建设方面从跟随性发展的必然王

国走向了自主发展的自由王国。

二、龙芯基础版操作系统和商业操作系统

操作系统有两种典型的商业模式，一是微软的 Windows 闭源授权模式，二是谷歌的 Android 开源基础版模式。龙芯中科研制并提供基于 LoongArch 的基础版操作系统 Loongnix，支持包括统信、麒麟在内的专业操作系统企业推出品牌操作系统，同时支持整机或系统企业基于 Loongnix 定制操作系统。

Loongnix 操作系统分为面向终端型应用的 Loongnix_Desktop 以及面向服务器应用的 Loongnix_Server，前者基于 Debian 进行龙芯 CPU 的适配优化，后者基于 CentOS 进行龙芯 CPU 的适配优化。龙芯中科基于 Loongnix 平台开展操作系统核心模块的迁移适配、研发和优化，支持硬件兼容的统一系统架构，支持应用兼容的统一编程框架。

在操作系统核心模块方面。完成包括 BIOS、内核、GCC 编译器、LLVM 编译器、GoLang 编译器、Java 虚拟机、JavaScript 虚拟机、.NET 虚拟机、浏览器、媒体播放器、图形库、Spice、KVM、Docker 等操作系统核心模块向 LoongArch 的迁移和优化，联合合作伙伴对版式文件、流式文件、微信等进行迁移和优化。结合应用需求，将大量外设驱动迁移到龙芯平台。

在硬件兼容的统一系统架构方面。通过对 CPU、桥片、BIOS、操作系统内核的升级规范，实现了统一系统结构。在该架构规范下，一个操作系统的二进制可以在 LoongArch 架构的不同 CPU 和不同主板上直接安装。

在应用兼容的统一编程框架方面。应用的兼容性问题是在 Linux 系统普遍存在一个问题，体现在不同的基于 Linux 的操作系统（如 Loongnix、统信、麒麟）应用不兼容以及同一操作系统随开源社区进行版本升级后可能导致原应用不兼容。为了解决上述问题，龙芯公司正在实现 LoongArch 平台上不同厂家、不同版本操作系统的应用兼容框架。该框架支持不同版本的 Linux 平台上应用二进制高效迁移。在此基础上，龙芯公司还将开发自主的编程语言和编程模型，鼓励客户基于龙芯自主编程模型进行应用开发。

三、通过二进制翻译融合软件生态

作为一种新型指令系统，在其原生软件生态成熟之前，LoongArch 将借助二进制翻译实现与 X86、ARM 等指令系统应用的二进制兼容来弥补其早期应用生态的不足。作为 LoongArch 的原生软件生态的补充，需要构建从 X86 到 Loongarch 的体系结构翻译器 LATX (Loongson Architecture Translator from X86)，从 ARM 到 Loongarch 的体系结构翻译器 LATA (Loongson Architecture Translator from ARM)。

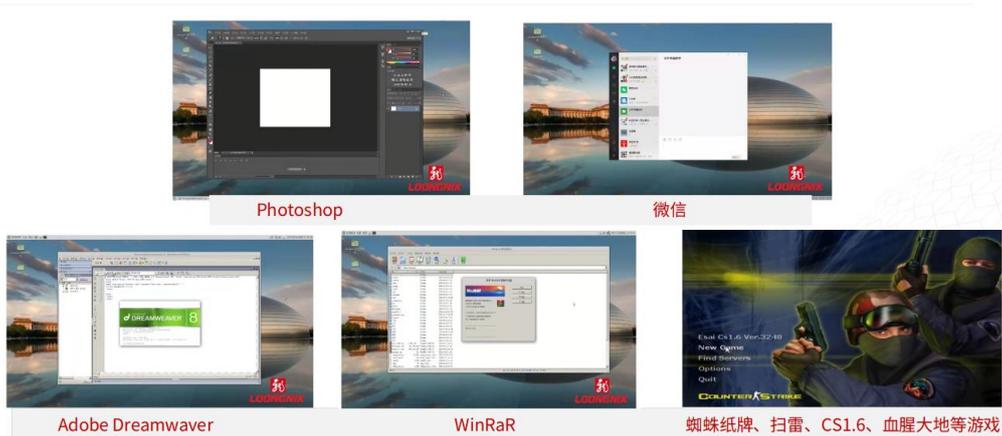
LoongArch 体系结构翻译器和前人工作的区别在于在 LoongArch 定义时就充分考虑多种指令系统二进制翻译的需求，强调更加紧密的软硬件协同，大幅提高了跨指令系统二进制翻译的效率。下图给出了 LoongArch 软件生态的结构。

Loongarch Linux apps	X86 Windows/Linux apps	ARM Android apps
	LAT from X86	LAT from ARM
Loongnix on Loongarch		
Loongarch		

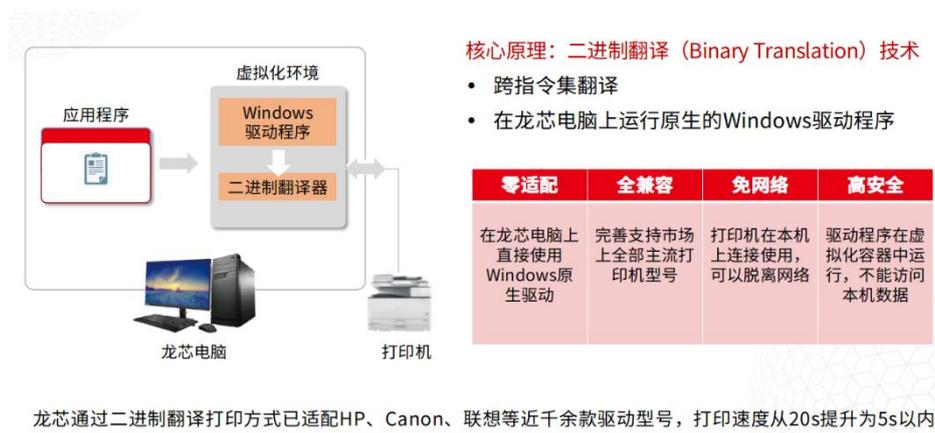
目前,龙芯二进制翻译系统可以支持以下 X86 平台应用。

一是在龙芯平台上支持 X86/Linux 应用。X86/Linux 平台上有较为丰富的工程类软件，如 EDA 工具等。此外，X86/Linux 上办公类软件也相对比较丰富，如腾讯 QQ 等。

二是在龙芯平台上支持 X86/Windows 应用。为了运行 X86/ Windows 应用,不仅需要通过 LATX 实现对指令的翻译,还需要类似于 Wine 的平台模拟 Windows 环境。目前,基于 LATX+Wine 平台已经可以在龙芯 3A5000 上运行包括微信、PhotoShop 等在内的 X86/Windows 应用,如下图所示。龙芯公司将在兼容性和性能两个方面持续改进和优化 LATX 和 Wine。



三是在龙芯平台上支持 X86/Windows 外设。用户现场的不少老旧打印机未提供 Linux 系统驱动，无法在 Linux 类操作系统上使用。基于二进制翻译技术，在龙芯系统上采用虚拟化方式直接运行 Windows 原生驱动程序。无需再适配，不依赖网络，在隔离本地数据的同时实现了本机打印。目前，基于龙芯基础版操作系统 Loongnix 以及基于龙芯平台的商业操作系统已经开始集成支持此方案。下图给出了在龙芯平台上运行 Windows 平台打印机驱动的原理。



四、建立面向工业控制的生态体系

如果说 Android 改造了服务业（打车、购物、吃饭），我

们需要一个平台改造制造业。LoongOS 是一个以实时操作系统内核为基础的、面向工业互联网和高可靠行业终端的操作系统。该操作系统把控制系统 APP 化，如在 LoongOS 上编写“飞机 APP”来控制飞机，编写“坦克 APP”来控制坦克，编写“高铁 APP”来控制高铁。

LoongOS 是具备精简、高效、实时特征的工控类操作系统，基于通用 Linux 内核，利用 RT-Linux 技术实现实时性。LoongOS 简化了复杂的传统个人计算机图形系统，支持 FrameBuffer、EGL、Wayland 三种图形应用模式，其中 FrameBuffer（二维）和 EGL（三维）都是单窗口应用全屏模式，具备系统轻量、启动迅速、开发便捷等特点；Wayland 模式可以支持多窗口应用，基于 OpenGL 实现了全功能的三维驱动和编程接口，具备简洁的个人计算机图形环境。LoongOS 三种图形模式都支持图形编程开发和应用环境，支持兼容 VxWorks 的 RTAPI 编程接口。针对嵌入式系统的需求，实现了文件系统加固、掉电保护、分区保护、安全隔离等特性。LoongOS 具有丰富的 API 编程接口，并有配套的集成开发环境 LDK，可根据需求灵活定制专用系统。LDK 支持可视化安装、远程调试，支持所见即所得的开发方式。

目前 LoongOS 已经在不少对显示要求很高的显控系统中得到应用。龙芯公司将持续改进和优化 LoongOS，使之成为面向工控类应用的平台型操作系统。

五、龙芯开源社区建设

开源社区聚焦了 CPU 生态的开发者和软件资源。龙芯积极参与社会主流开源软件项目及社区，对 LoongArch 源码进行维护，并“反哺”开源上游社区。

龙芯公司正在积极推动在上游社区建立 LoongArch 分支。在国际开源社区建立与 X86、ARM 等并列的 Loongarch 分支是龙芯生态的根基。目前，LoongArch 已从 GNU 组织获得表征二进制格式的 ELF Machine 编号（258 号），二进制开发工具 binutils 已经并入开源社区，GCC、C 库、内核正在并入开源社区，包括浏览器在内的部分基础软件已并入开源社区。2022 年起，龙芯将每年投入一定的经费资助海外社区开发者的工作。在得到上游开源社区支持后，LoongArch 生态将随社区自动演进，可以极大丰富支持 LoongArch 的软件版本，大幅减少软件迁移适配工作。

龙芯开源社区（loongnix.org）是龙芯 CPU 的基础软件生态建设平台，由龙芯中科组织并联合产业链合作厂商（合作者）、国际开源软件社区及开源爱好者共同参与建设。社区以“开源、开放”的形式进行管理，设置 BIOS、Linux 内核、编译器 GCC、Java 虚拟机、龙芯浏览器(Firefox、Chromium)、媒体解码库、图形桌面、KVM 虚拟机等重要基础软件项目。经过多年建设，信息化应用中的主流编程语言和开发环境都

已经在龙芯平台上完善支持。

Loongnix 操作系统是龙芯开源社区推出的 Linux 发行版，作为龙芯软件生态建设的成果验证和展示环境，集成在内核、驱动、图形环境等操作系统基础设施方面的最新研发成果，以“源码开放、免费下载”的形式进行发布，可直接应用于日常办公、生产、生活等应用环境，同时可供合作厂商、科研机构及爱好者等在龙芯平台上研发其品牌软件或专用系统。

下图描述了 LoongArch 已原生支持的 Linux 主流应用开发环境。



Deepin 社区和 UOS 操作系统

统信基于 Deepin 社区发展商业版

UOS 桌面操作系统

统信软件 刘闻欢

一、统信软件介绍

统信软件技术有限公司（以下简称：统信软件）于 2019 年成立，注册资金 6.27 亿元，现有员工 3000 人。总部设立在北京，在全国共设立了 6 个研发中心、7 大区域服务中心与 3 地生态适配认证中心，包括武汉、上海、广州、南京、成都、重庆、西安、太原、深圳等地，公司规模和研发力量在国内操作系统领域处于第一梯队，技术服务能力辐射全国。

统信软件作为以技术创新为基因的高科技企业，在操作系统研发、行业定制、国际化、迁移适配、交互设计等多方面拥有深厚的技术积淀，现已形成桌面、服务器、专用设备操作系统产品线，以及集中域管平台、企业级应用商店、彩虹平台迁移软件等应用产品及其解决方案，成为中国桌面操作系统解决方案新的标杆，能够满足不同用户和应用场景对操作系统产品与解决方案的广泛需求。

统信软件始终以中国操作系统行业未来发展的视角，拥抱开源，根植中国，放眼世界，积极打造和参与 Linux

上游社区的建设，发展和引领行业生态，迅速成为国产商业操作系统的领军企业。

二、统信推动 Deepin 桌面操作系统上游社区的发展

统信 UOS 开源社区 (deepin) 作为中国最大、最活跃的操作系统社区，积极走出国门，用户遍布全球 100 多个国家与地区，累计发布版本 40 余次，支持语言 34 种，累计下载量超过 8000 万次，全球社区用户超过 200 万，提供开源代码超过 600 万行，是国内屈指可数的国际社区。

其中，统信软件基于 deepin 社区自研了全套深度桌面环境 DDE (Deepin Desktop Environment)，DDE 桌面环境的 GUI 程序皆基于 DTK/Qt 开发完成，有着统一的界面设计，以及精心设计的图标。DDE 具备美观易用、极简操作的特性，主要由桌面、启动器、任务栏、控制中心、窗口管理器等组成，系统中预装了自研特色应用，既能让用户体验到丰富多彩的娱乐生活，也可以满足用户的日常工作需要。

统信软件自主维护了 deepin 发行版和其社区。在 2008-2015 年期间，深度操作系统从 ubuntu 进行分支。自 2015 年 Deepin V15 发布后，开始基于 Debian 进行发展，成为在 Debian 分支下与 Ubuntu 并列的一级发行版。

经过多年的发展，围绕 deepin 发行版的 deepin 社区已成为独立 Linux 社区，在国内外的影响力也得到持续提升，拥有 6 个海外社区，在 40 多个国家拥有 130 多个镜像站点，可为中国操作系统提供国际开源社区支撑。

统信 UOS 的商业版本则基于 deepin 社区版，在社区版本稳定 6 个月后，选择稳定分支开发统信 UOS 的商业版本。从这个角度来说，deepin 社区版类似于 Fedora，而 UOS 商业版类似 RHEL。

Deepin 社区版下一个主要版本发展计划，将直接从上游各源码社区构建顶级国际社区，不再基于 Debian 衍生发展。通过技术与社区维度的提升，确保供应链安全，结合新技术的发展，继续引领国产桌面操作系统再上新台阶。

三、基于 Deepin 社区开发的统信商业版桌面操作系统 V20

当前，国内桌面操作系统正在经历从可用到好用的阶段，已具备初步的生态，可以支撑各项基础性办公、上网、娱乐等需求，尤其是在党政办公的特定领域，已经证明基本具备全面替代的能力。

统信的桌面商业操作系统 V20 基于 deepin 社区版本开发，在稳定性、易用性、安全性等方面有明显的优势，适用于电子政务、金融、电信、电力、军工、交通等关键行业领域。而在操作系统最关键的生态发展方面，仅 2020 年

源等行业都已经通过适配并推广使用，并实现多个单一企业超过万点以上实施替换案例。

倪光南院士曾表示统信 UOS 代表了目前我国自研操作系统的最高水平。据工信部相关数据披露，在国产桌面操作系统领域，统信软件稳居 2020、2021 年中国 Linux 操作系统市场排名第一，市场占有率达到 70% 以上。统信软件已成为当之无愧的国产桌面生态建设的绝对领导者。

作为商业操作系统厂商，统信软件建立了完善的售前、售中、售后技术服务体系，负责为来自国家部委、党政机关，国防、教育等行业单位，以及个人用户提供统信 UOS 国产操作系统及定制解决方案产品使用相关的技术咨询、安装部署、培训认证以及运维保障服务；在全国设置 7 个区域服务中心、31 个省级服务网点，实现了省会城市 2 小时、地市（个别偏远地区除外）4 小时的运维服务时效。

统信软件积极承担国家适配认证中心的建设和运营工作，已与近 3000 家生态伙伴深度合作，适配完成 10 万多款软硬件兼容组合，并发起成立“同心生态联盟”，目前已有成员单位 500 余家，有效推动了操作系统生态创新发展。（上述数据持续更新，详见统信软件生态社区网站：www.chinauos.com）

成熟的商业操作系统软件，对系统安全与维护能力尤为重要。统信软件对于从客户或统信产、研部门发现的产品重

大重大漏洞等信息安全问题，采取大客户主动式服务管理流程，从发现，上报，到产研部门分析、修复，再到售后团队按界别通知用户主动修复服务或紧急预案处理，打造标准的服务流程和及时响应能力。

四、基于统信桌面操作系统的解决方案

统信软件在开发操作系统，拓展行业生态的过程中，依据客户实际迁移需求，不断创新，2021年最新升级三大解决方案，成为中国桌面操作系统解决方案新标杆。

1、集中域管平台解决方案

集中域管平台由统信自主研发，提供机构人员维护、统一身份认证及终端安全管控功能，支持对终端的安全策略、系统参数等进行集中配置，适用于大规模终端应用集中管控的业务场景。

2、企业级应用商店解决方案

企业级应用商店提供应用管理、商店运营、安全签名、智能适配、外网应用数据同步等功能，解决内网环境应用分发、合规使用和管理问题，帮助企业实现自主、安全、高效的应用全生命周期管理。

3、彩虹平台迁移软件解决方案

采用融合虚拟技术，面向使用 Windows 7 及以上版本的终端用户，将用户在 X86 终端上的原 Windows 系统、软件及其原生数据平滑、完整、安全迁移至统信桌面操作系统，支持业务连续、数据互通、生态共享。

五、总结

统信软件始终坚持以“打造操作系统创新生态，给世界更好的选择”为愿景，坚持技术创新与开放合作，专注于操作系统的研发与服务，在操作系统研发、行业定制、国际化、迁移和适配、交互设计、咨询服务等各方面拥有深厚的技术沉淀。为政企行业信息化、数字经济建设提供坚实可信的基础支撑，筑牢信息安全基座。

当前，统信软件的桌面操作系统的部署正在从百万级向千万级别迈进。同时统信的服务器操作系统业务也在快速增长，在政府、金融、电信、交通、能源等领域广泛应用，通过对上游社区欧拉和龙蜥的积极参与和贡献，完善商业发行版本，已经成为国内服务器商业操作系统的主要提供商。

未来，统信软件将更加努力，仰望星空，脚踏实地，力争成为全球主流操作系统，产品与解决方案覆盖全应用领域，打破国外厂商垄断格局，建立全球范围的基础软件生态。

OpenCloudOS 操作系统社区

OpenCloudOS 操作系统社区介绍

OpenCloudOS 社区

一、社区概况

2021 年 12 月 22 日，开源操作系统社区 OpenCloudOS 正式宣布成立，包括腾讯及宝德、北京初心、北京红旗、飞腾、浪潮、龙芯中科、OPPO、先进开源、中电科申泰、中科方德、兆芯等 20 余家操作系统生态厂商及用户成为首批创始单位。

OpenCloudOS 社区由腾讯与合作伙伴共同倡议发起，致力于打造一个完全中立、全面开放、安全稳定、高性能的操作系统及生态。完全中立，意味着社区不存在厂商标签，也不由任何一个单独的厂商主导；完全开放，意味着能够聚集起国内外的操作系统研发力量、社区成员共识决策、产学研用共建生态。在成立之初，OpenCloudOS 就决定成为无厂商标签的中立社区，并经过了开放原子开源基金会的 TOC 评议，确认接受社区项目捐赠，未来将成为开放原子开源基金会的一员，由基金会托管和监督，以标准开源社区模式运作，保持中立和开放，社区也将由参与单位共同治理。

OpenCloudOS 操作系统是一款企业级 Linux 服务器操作系统，是采用 Linux 内核的自行设计的 Linux 操作系统。其内核基于 Linux Kernel 社区 LTS 长期维护分支，并根据社

区用户需求进行增强和优化，独立演进和维护；用户态组件部分长短期两条技术路线并行：短期路线(当前 V8 版本)基于 Redhat 软件源重新构建和维护,保持与 CentOS 完全兼容,目标是解决短期 CentOS 替代问题;长期路线(预计于 2023 年初发布 V9 版本)将完全从 Upstream 社区构建和维护所有的软件包,与独立演进的内核一起构成完整的 L1(不依赖其他厂商和下游社区)版本,在保障开源 OS 软件可靠供应的同时,基于腾讯和合作伙伴的海量应用场景验证,保障版本的稳定性,满足企业级生产需求。

二、技术沉淀

作为国产开源操作系统社区,OpenCloudOS 沉淀了腾讯及多家厂商在软件和开源生态的优势,在云原生、稳定性、性能、硬件支持等方面均有坚实支撑,可以平等全面地支持所有硬件平台。

OpenCloudOS 稳定支撑了腾讯及其合作伙伴超过 1000 万节点的大规模场景,在政务、金融、互联网等行业经过长期使用验证,可以为企业提供稳定高可用的服务,满足业务严苛负载需求。

作为 OpenCloudOS 社区的主要发起方之一,腾讯在操作系统领域已有超过 10 年的技术积累。TencentOS Server 是腾讯自研的服务器操作系统,它的诞生最早可以回溯到

2010 年，由于初期使用的开源操作系统无法满足业务日益复杂的需求，同时腾讯有着种类繁多的海量业务生态，从社交，游戏，到金融支付，云，大数据，AI，安全等，对底层操作系统的稳定性，性能，兼容性、安全性等都有更强的要求。腾讯决定结合自身业务的特性需求、性能需求和安全需求自研操作系统，并结合海量业务场景进行系统优化和打磨。在通过 TencentOS 的应用提升产品性能的同时，2019 年，腾讯也正式将 TencentOS 内核开源，向全球开发者全面开放近十年的技术积累。

后续，腾讯也将基于自身的操作系统技术积累和能力，与合作伙伴一起，全力支持 OpenCloudOS 社区的建设，持续向社区投入资金及资源，来支持项目的持续研发，推动社区的繁荣发展。伴随着项目的深入，腾讯还会将其操作系统领域的技术积累，特别是云原生的相关技术优化，如容器基础设施、CPU 调度、内存管理、IO、网络、可维护性、虚拟化等 7 大维度的特性贡献到社区中。

三、技术特色

1、绿色操作系统

中央经济工作会议将“做好碳达峰、碳中和工作”作为 2021 年的重点任务之一，提出我国二氧化碳排放力争 2030 年前达到峰值，力争 2060 年前实现碳中和，并要求抓紧制定

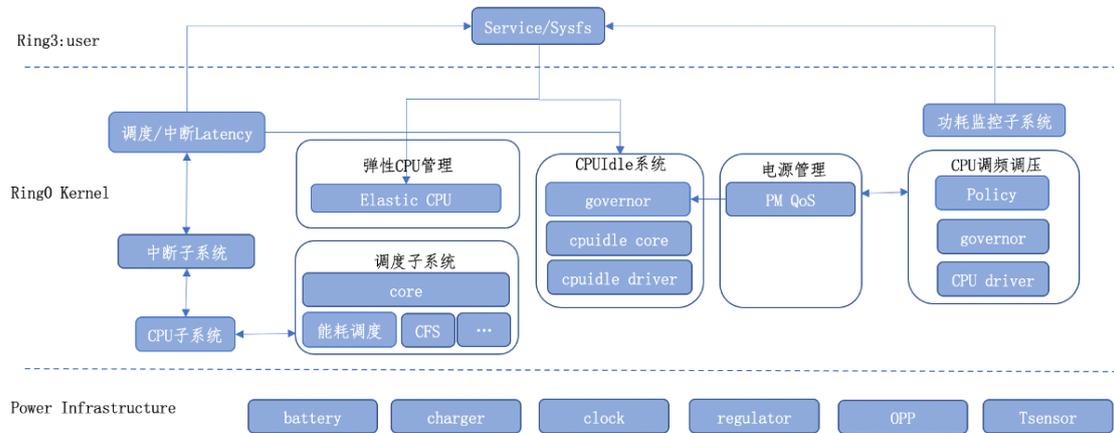
2030 年前碳排放达峰行动方案。2021 年 1 月，腾讯宣布启动碳中和规划，成为国内首批启动碳中和规划的互联网企业之一。

数据中心（IDC）是碳中和的重要主力之一，当前，行业绝大部分 IDC 的日常或者高峰期 CPU 利用率，周平均值小于 15%，能耗和资源利用率都存在巨大的提升空间。OpenCloudOS 从系统级能耗优化和资源利用率提升两个大方向入手，通过系统级的设计和重构，分别提供了两套完整的解决方案，一方面可最大程度优化系统能效，另一方面可大幅提升整机资源利用率，减少单位算力的电能消耗，整体减少 IDC 的碳排放。

系统级能耗优化方向。CPU 是服务器功耗的主要来源，通过逐步软件、硬件关部件、各级 cache、降电压使 CPU 进不同深度睡眠，可降低 CPU 的功耗，当前大部分主流 CPU（服务器）支持 CPU 深度睡眠状态（C1-C6），能效提升空间达 5-30%。

OpenCloudOS 通过重新设计系统级能耗管理系统，以 OS 内核作为功能核心，提供全新系统级能耗优化解决方案。系统级能耗优化方案，以渐进式和精细控制为设计原则，同时考虑单机相应、集群建议，通过在操作系统级别 CPU 弹性运行与深度睡眠，结合功耗监控和细粒度的电源管理，达到节能目的。系统级能耗优化系统在 OS 内核中涉及的模块包括：

功耗监控子系统、弹性 CPU 管理、调度子系统、CPUIdle 系统、电源管理、CPU 调频调压、CPU 子系统等，系统架构如下：



当前 OpenCloudOS 的系统级能耗优化方案已逐步在腾讯的 IDC 机房中上线推广中，当前的实测结果显示：能耗收益明显，单机的能耗收益达 5%-20%；同时通过实测典型的业务负载，分别采用 C1-C6 的睡眠状态，实测性能影响很小(小于 0.1%)。预计在腾讯数据中，整体能耗节省 6 亿 KWH/年，减少碳排放 24 万吨/年；从宏观角度看，由于相关技术为通用技术，可推广至几乎所有 IDC 场景，整体减少数据中心 5-30%碳排放，具有巨大的长期及宏观价值。

系统级能耗优化技术作为 OpenCloudOS 的未来核心技术方向之一，OpenCloudOS 在系统层面提供了一套整体解决方案(技术品牌：悟能)，其中 OS 内核作为核心，对现有 Linux 内核中的能耗管理部分进行了系统级重构，以绿色节能作为首要目标，相关方案即将在腾讯内部大规模落地，后续将持续将相关通用技术能力全部输出到 OpenCloudOS 社区中，为更广范围的应用，贡献更大的价值。

全场景混合部署系统方向。当前 IDC 机房利用率偏低的主要限制还在于绝大部分业务具有明显的峰谷特性，须要根据业务峰值进行容量规划，导致非高峰期的低资源利用率。在离线混合部署是业界提升整机资源利用率的一种通用做法，但对操作系统底层的资源隔离性提出了极高的要求，同时也需要从底层内核、容器平台到混部容器调度系统的整体解决方案。OpenCloudOS 针对混部场景设计开发完整的统一资源隔离解决方案(技术品牌：如意 RUE)，并结合腾讯自研并开源的容器服务平台 (TKEShield) 和混部调度系统 (Caelus)，形成整体的全场景(物理机、虚拟机和容器)混部解决方案。其中包含三大腾讯开源核心项目，已全部开源，是业界唯一的全栈开源解决方案。另，TKEShield 项目已开源并捐入开放原子基金会，是开放原子基金会首批孵化的项目；OpenCloudOS 已开源，且即将捐入开放原子基金会，将在基金会中与 TKEShield 项目协调发展。



目前基于 OpenCloudOS 的全栈混部系统在腾讯内部已全上线验证，整体覆盖超过 2000 万核，提供超过 400 万离线算力，每年节省超过 5 万台服务器，资源利用率提升 30%，

能耗减少 4.4 亿 KWH/年，碳排放减少 17.6 万吨/年。

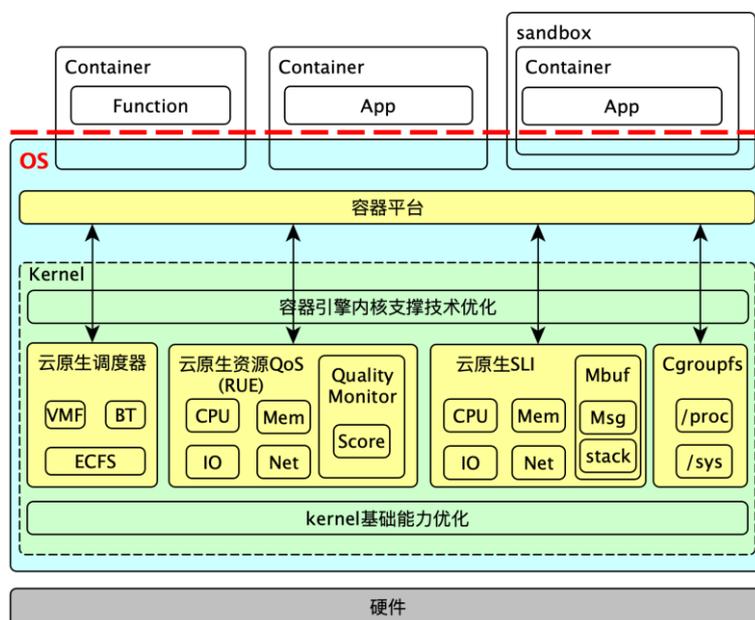
结合系统级能耗优化和全栈混合部署系统的落地成果，OpenCloudOS 预计每年节能超过 10 亿度/年，碳排放减少超过 42 万吨/年。相关通用技术可推广到更广的范围，助力国家的整体碳中和战略，也因为在绿色节能方面的突出表现，相关 OS 系列产品于 2021 年 7 月 15 日源亮相央视《经济半小时》，同时，于 2021 年 9 月 17 日中国信通院主持召开的“云计算开源产业联盟（OSCAR）开源产业大会”，荣获中国信通院授予的“2021 年 OSCAR 尖峰开源项目及开源社区”奖项。

2、云原生最佳底座

在云原生浪潮席卷的行业大背景下，伴随着各种业务架构的快速转身，业务的容器化、微服务化、Serverless 化，对底层的基础设施(包括核心的 OS)提出了新的挑战和要求，OpenCloudOS 也针对云原生场景和需求，进行了深度重构和重新设计，全面拥抱云原生。

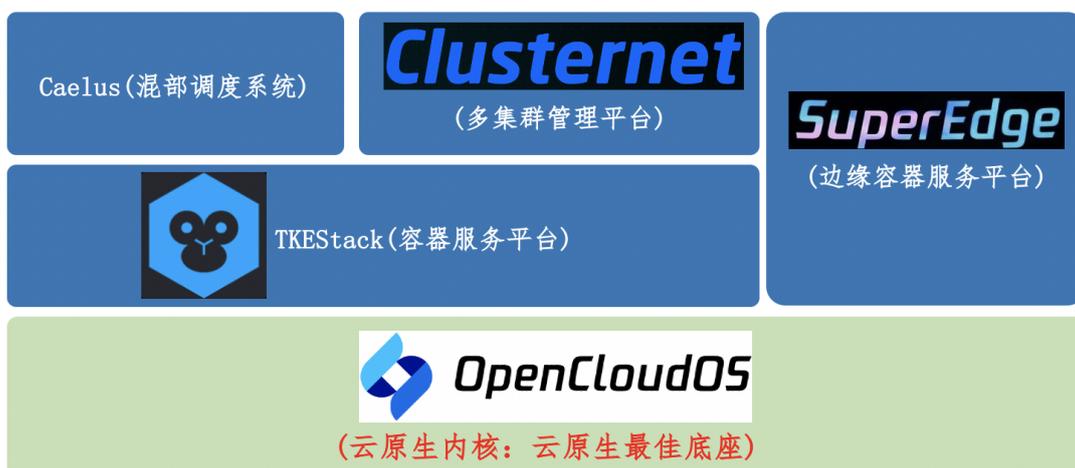
具体来说，在 OS 内核层面，OpenCloudOS 针对云原生场景设计了一整套原创云原生特性，形成云原生内核；包括：云原生调度器(TCNS)、云原生资源隔离系统(RUE)、Quality monitor(服务质量监视器)、云原生 SLI(容器视图的专业指标)、Mbuf(常态化内核关键监控)、Cgroupfs(容器资源视图隔离)。此外，在云原生基础能力方面，对 KVM 进行了深度优

化，腾讯云连续五年入围全球企业 KVM 年度开源贡献榜，KVM 社区贡献国内第一，行业领先；其他还包括大量对容器基础设施的优化。



另一方面，OpenCloudOS 结合腾讯整体在云原生领域持续深耕成果，整体形成了全栈开源云原生解决方案，当前包括 5 大核心开源项目：Clusternet(多集群容器管理平台)、SuperEdge(边缘容器服务平台)、TKESStack(通用容器服务平台)、Caelus(混部调度系统)、OpenCloudOS(最绿色操作系统、云原生最佳底座)，OpenCloudOS 作为整个技术栈中最底层的基石，为云原生全栈的稳定、高效运行提供核心支持。具体来说，云原生调度器(TCNS)为上层的云原生业务调度提供了更高效、更实时的调度能力；云原生资源隔离系统(RUE)为多优先级混合部署场景提供完美的、全面的隔离能力；云原生SLI 为云原生业务提供更精细、更专业的监控系统和指标；

Quality monitor 和 Mbuf 为云原生业务提供更及时、常态化运行的 trace 能力；Cgroupfs 为云原生业务提供独立的资源视图。除云原生内核外，配套的 OpenCloudOS 系统组件，为各开源云原生项目提供最佳载体。



自主工业软件开发

我国工业软件发展之困

中国工业技术软件化产业联盟

在全球工业进入新旧动能加速转换的关键阶段，工业软件已经渗透并广泛应用于大部分工业领域的核心环节，是现代工业的“灵魂”，是工业强国之重器。习近平总书记在中共中央政治局第三十四次集体学习时强调，要全面推进产业化、规模化应用，重点突破关键软件，推动软件产业做大做强，提升关键软件技术创新和供给能力。工业软件作为关键软件的重要组成部分，要清楚的认识我国工业软件产业还有很长的路要走，还存在着关键技术缺失、高端人才短缺、产业规模较小、核心竞争力较差、发展生态环境脆弱等问题。

一、我国工业软件发展现状

近年来，我国工业软件产品收入的年增长率保持在 15% 左右，截止至 2021 年 11 月，工业软件产品收入达到 2121 亿元²。但是，我国工业软件的市场规模远小于全球市场规模，2020 年我国工业软件市场规模仅占全球市场规模的 6%³。细分领域发展情况各不相同，研发设计类工业软件国内外水平差距较大，求解器、三维建模、网格划分等多项核心技术掌握不足；生产制造类工业软件高端产品供给不足，单项产品

² 数据来源：工信部运行局

³ 数据来源：中国工业技术软件化产业联盟，《工业软件产业白皮书（2020）》

在细分领域小有优势；运维服务类工业软件国内外软件产品有代差，还需要进一步提升产品性能；经营管理类工业软件国内软件占据 70%国内市场，但是高端市场仍被 SAP、Oracle 等国外厂商占据。总体来说，我国工业软件市场大部分被国外工业软件占据，虽然部分产品有单点突破，但是整体实力与国外软件相差较远。

二、我国工业软件面临的问题

（一）国产工业软件缺乏体系化竞争力

一是我国对工业软件和自身工业水平协调发展的重要性认识不到位，导致盲目对标和跟随国外优秀企业和产品，忽视了我国和欧美发达国家工业水平差距较大的客观现实，严重消耗了有限的经费与精力，却未能显著提升国产软件的实用性和竞争力。二是我国软件开发商对国内工业企业“刚需”研究不够，脱离用户实际需求，导致产品和用户需求的脱节。三是国外大型企业已经形成坚固的竞争壁垒，一方面企业通过大量并购加速拓展产品线，逐步形成从研发设计、生产制造到运维管理的全产业链供应闭环，构筑起进入壁垒，导致以点工具为主要发展模式的国内工业软件进入市场更加困难；另一方面因用户路径依赖、转换成本巨大构成“锁定效应”壁垒，用户粘性较强。四是我国贯穿整个制造业过程的生态化部署较少，为用户提供整体解决方案的能力弱，从而造成了“生态效应”壁垒。五是缺乏统一的软件标准，用户企业的产业链上下游采购不同主流厂商的软件，数据格

式、接口标准难以协同，即使通过中性 STEP 格式标准转换也存在重要特征信息丢失的问题，使得国产软件间无法形成合力，难于与国外一体化的平台软件相抗衡；六是目前国内工业软件缺乏科学合理的价值评估和定价机制，软件价值导向得不到强化，且在招投标中盲目压价和中低价择标的现象仍比较突出，恶性竞争影响了软件企业的持续发展。

（二）核心技术缺失，产品竞争力不足

与国外相比，国产工业软件核心技术掌握不足，产品在性能、功能模块数量、平台的稳定性上与国外软件存在较大差距，产品线的完善度不高，且在高端客户中的相关技术积累和市场占有率不占优势。例如，国内 CAX 软件厂商规模较小、研发能力有限，对关键核心技术研发缺少高额度持续性的资金投入，导致国内软件自主核心技术研发进展缓慢；EDA 软件则是对先进工艺支撑能力弱，缺乏全流程的设计平台，与国外软件有较大差距；生产制造类软件面临的问题是，由于对于特定行业的生产工艺和控制机理缺乏深入的研究，存在技术门槛，再加上产品技术水平偏低，产业规模未发展起来，达不到高端行业用户应用需求，导致国内产品线的完善度不高，产品化程度不够，尚不能为用户提供从基础控制、优化控制、生产管理到仿真测试的一站式解决方案；国内 ERP 软件在技术架构、业务模式分层设计、数据分析等方面与国外存在差距，并且软件业务模型积累较少，智能化应用水平有待提高，尤其是与大型企业合作较少，缺乏对大型企业的业务支撑，导致高端市场开拓受阻。

（三）工业软件人才短缺，师资严重流失

目前，我国面临着较为严重的工业软件人才短缺问题，一是工业软件师资队伍严重不足，国内自主工业软件行业长期疲弱，国家支持政策缺乏良好的延续性，开展相关研究的骨干教师或转行或转而研究其他领域，使得教师资源流失严重，致使前人的研究束之高阁，出现较大的研究断层。二是培养工业软件人才的专业设置偏差，绝大部分高校都没有工业软件专业。即使有少量高校开设了专门的工业软件专业，大多是挂靠在其他学院中，每年培养的高端工业软件人才数量屈指可数。三是工业软件人才流失严重，一方面是前些年，国内还没有成熟的吸引人才的政策，大批的人才流向海外；另一方面是其他高薪产业对人才的虹吸效应，由于国内工业软件产业不成规模，工业软件人才的收入与国外有较大差距，金融和互联网等行业的虹吸效应凸显，造成高端人才大量流失。

（四）工业软件知识产权保护意识和力度不足

一是我国工业软件知识产权的保护力度不足。因知识产权保护力度不足，使得国内部分企业、高校和科研院所使用工业软件的成本近乎于零。在不加管制的情况下，知名网站上堂而皇之地发布盗版软件销售广告，盗版软件挤压正版软件市场，特别是国产软件无法在市场中依靠性价比来赢得竞争优势，难以实现零的突破，难以起步和逐步发展。盗版软件扰乱了正常的市场机制，淡化了合理的为知识付费意识，

破坏了良好的工业软件生态。二是我国企业保护工业软件知识产权法制意识比较薄弱。很多企业还未清醒地意识到知识产权所能带来得的相关权利与利益，也没有明确认识到私自使用和仿造他人产品将会承担的法律责任和财产损失。企业没有建立完善的知识产权管理体系，无法很好地应对工业软件创造、运用、保护、管理的需求。

（五）落后的意识观念和信息化程度延缓了产业发展

一是过去我国的信息化、数字化建设有“重硬轻软”的问题，对发展自主可控工业软件重视程度不足，技术研发投入与国外存在明显差距。二是国内厂商设备信息化程度低，数据采集困难，抑制国内工业软件需求。三是我国工业企业广泛应用国外软件，长期忽略对我国工业化进程中产生的许多核心技术、知识经验和工艺数据的有效积累和沉淀，导致优质的工业软件企业缺失。四是相关管理部门对工业软件的技术生态认知不够，往往认为抓自主工业软件聚焦于工业软件本身即可，但是极易忽略工业软件生存所依赖的图形引擎、约束求解器、数学算法、材料知识库、图形交互界面、数据转换标准、软件开发语言甚至软件编译器等工业软件生存的必备条件。

（六）工业软件和工业深度融合不足

一是围绕信息化和工业化所建设的基础要素，是工业赖以生存的基础，而我国工业基础相对薄弱，又因“累积效应”进一步了放大了工业基础薄弱对工业软件发展的影响。二是

国内工业软件与工业应用需求结合不紧密，难以满足复杂多变的工业实际业务与特定场景需求。三是国内软件企业和工业企业缺乏紧密联合机制，应用单位认为国产软件不够成熟不愿意使用，国产软件缺少在应用单位迭代优化成熟的机会，形成恶性循环，国内软件产业化和商业化受阻。

三、推动我国工业软件发展的建议

（一）持续优化政策环境

高度重视工业软件发展。一是针对研发投入不足的问题，统筹现有推进资金，加大对工业软件投入力度；推动成立工业软件发展专项基金，以国家资金撬动社会资金持续投入。二是针对产用脱节，企业不敢用、不愿用问题，探索出台软件首版次政策，通过保险补贴、研发应用双向奖补等政策，推动国产软件在各行业各领域应用。三是针对工业软件人才缺口大的问题，出台工业软件人才专项政策，对于工业软件研发人才，给予医疗保险、子女入学等优惠政策。四是针对企业小散弱、缺乏龙头引领问题，出台鼓励企业兼并重组政策；在国有企业改革框架下，推动成立国有大型工业软件集团，以投资基金促使股权联合，依托国有大型工业企业集团进行产业收购整合，加快工业软件大企业培育。五是针对工业知识积累薄弱问题，建立工业知识搜集、积累、流通使用机制，规范相关标准，促进工业知识流动传承，鼓励工业知识软件化，促进工业 APP 发展。六是加强知识产权保护力度，加大对互联网传播盗版工业软件的治理，依法严厉打击从事

工业软件盗版活动的企业、组织及个人。加强从源头封堵盗版行为，强令关停提供盗版下载的网站。七是加快制定推广软件成本度量和价值评估标准，推动建立科学合理的软件定价机制，在全社会营造尊重软件价值的文化氛围。

（二）夯实产业发展基础

一是建立关键共性技术突破机制。整合国内产学研用优势力量，建设协同开发与集成验证环境，加强计算物理、计算数学、并行算法等基础研究，推进三维几何建模引擎、求解器等关键技术联合攻关。在超算、异构数据处理等领域推进企业与高校、科研院所的能力合作，完善知识产权和科技成果转化机制。二是建立国家级公共知识库共享平台。针对CAD、CAE等大型工业软件研发共性需求，布局基本求解算法库、标准零部件库、基础数据库、知识库，通过开源等方式，促进企业资源开放共享和开发利用。面向不同行业工业研发设计需求，布局行业通用的结构、功能、性能模型库，降低企业重复研发成本。三是建立数据模型和接口标准体系。基于市场对国际主流工业软件的模型读取转换、异构协议转换、软件二次开发等兼容性和定制化要求，发展和定义自主数据模型和接口标准，形成覆盖设计、建模、仿真、制造、运维等不同环节的自主数据表达和交换能力。

（三）推动关键技术攻关

一是建设国家级工业软件发展基地。建设软硬件适配中心，国家级产学研用联合攻关平台，构建运行有效的产学研

用联动机制。二是**建立工业软件关键技术攻关平台**，选择相关高校科研院所为主体，协调研发设计类工业软件企业参与，突破工业软件关键技术，引导供给侧工业软件企业使用自主内核并实现典型应用示范。三是**设立工业软件重大攻关项目**。围绕三维几何引擎、约束求解器等“卡脖子”技术，设立工业软件重大攻关项目，选择在工业软件领域具有良好基础和积淀的高校科研院所作为主体，联合工业软件龙头企业共同开展工业软件核心技术攻关。

（四）稳步推进工业软件应用

一是开展工业软件产品成熟度评估。聚焦船舶、汽车、航空等重点领域应用需求，编制工业软件产品目录。二是**推动工业软件在重点行业领域典型示范应用**。面向船舶、航空航天、汽车、轨道交通等重点领域，开展国产工业软件产品试点示范，明确示范应用的思路、方法、技术、结果和目标，打造一批优秀行业解决方案，促进一批重大项目立项和建设，逐步推进工业软件在更多领域拓展应用。围绕产品主价值链的系统性软件产品应用，形成整体解决方案。三是**依托工业互联网建设**，推动工业软件普及应用，借助国家大力发展工业互联网，发展基于工业互联网平台的工业软件服务模式。探索利用云化软件、订阅等模式，引导中小企业在工业互联网平台上使用新型工业软件。四是**借力工业 APP**，破解国内工匠不足难题，推进两化深度融合，搭建工业互联网平台应用生态。

2021 中国工业软件年度 企业排行 TOP100 发布

本刊编者⁴

工业软件在产品设计、成套装备设计、厂房设计、工业系统设计中起着非常重要的作用，可以大大提高设计效率，节约成本，实现可视化管理。

目前，在工业软件领域国外企业占据了领导性的市场地位，Gartner 数据显示，2020 年全球工业软件市场规模约为 4332 亿美元，折合人民币约 29891 亿元，而我国工业软件市场规模占全球比例仅约为 6.6%。

工业软件是现代制造业的基础与核心，没有自主可控的工业软件，迈向工业强国的征程就始终缺少坚实的基础。国家政策支持、开放式的创新精神、扬长补短的发展策略，既修内功、也练外力，才能逐步破解受制于人的难题。近年来受到国家政策的鼓励，产业发展迅速，涌现了大量优秀企业。

近日，互联网周刊发布了 2021 中国工业软件年度企业排行 100，榜单显示，国电南瑞位居榜首，宝信软件、中望软

⁴ 本刊编者为：陈伟、鞠东颖。

件排名第二和第三，广联达、中控技术、柏楚电子、华大九天、和利时、神软、用友网络进入前十。

工业大数据

工业大数据软件与开源创新

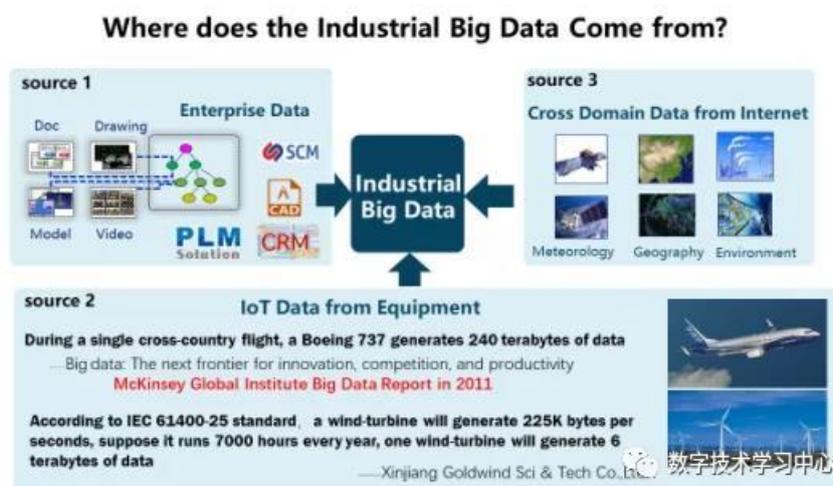
清华大学软件学院 王建民

编者按：本文是王建民教授于 2021 年 8 月在 ApacheCON Asia 2021 年会上做的报告（2021 年 12 月 21 日，受王建民教授委托，乔嘉林博士在 COPU 例会上就该报告内容进行了节选汇报）。

很荣幸与各位分享我们在清华大学工业大数据软件研究和开源实践方面的工作。

如今，大数据已经不是一个新词了，中国有很多互联网行业的领导者，比如阿里巴巴、腾讯、百度、华为，他们都是大数据领域的强势参与者，并且大多数是面向消费者的。但是，如果我们仔细观察中国经济，就会发现中国在大数据应用的一些主要领域内仍存在滞后——例如制造业、建筑业、运输业等。当下，这些行业面临着两大挑战：缺乏对先进大数据技术有深刻理解的人才，以及现今提供的技术水平不足以解决他们手中的具体问题。然而，大数据同样拥有许多新的关注领域，如人工智能、机器学习、数据科学等。

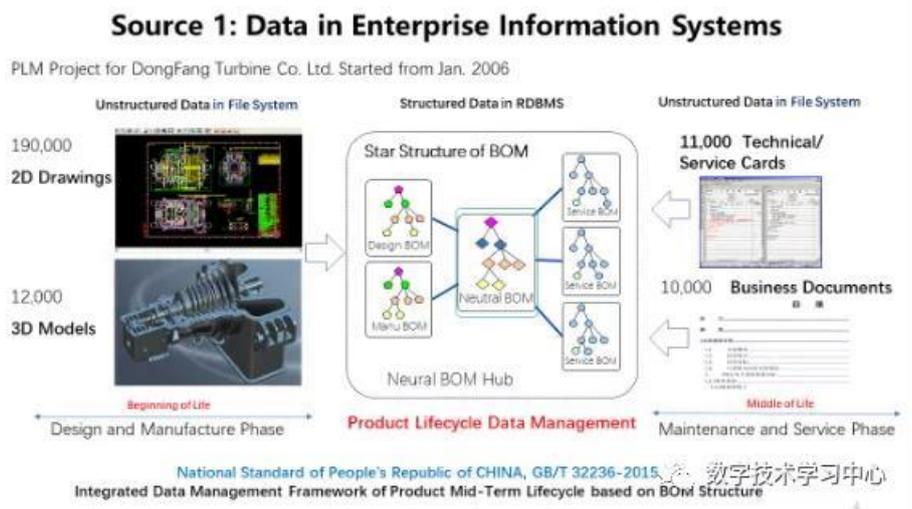
我们的使命是为这些行业进行大数据技术和应用方面的创新。根据 2011 年麦肯锡全球机构发布的大数据报告显示，制造业的数据量甚至超过了金融业。工业大数据从何而来？第一个数据源是企业信息系统，如 CAD 系统、PDM 和 PLM 系统、ERP 和 CRM 系统等（这些系统在上世纪六十年代就开始被企业所使用了）。第二个数据源是始于 21 世纪之初的工业物联网，如飞机、风力发电机等。工业物联网数据又称机器设备数据或工况数据，构成了工业大数据的主体。第三种数据源是来自互联网的跨领域数据，如气象、地理和环境数据，这些数据在当今的人工智能时代很容易得到。



工业大数据的第一数据源是企业信息系统。企业信息系统中的数据包含非结构化数据，如二维工程图纸、三维零件模型、服务卡、业务文档等，它们通常存储在文件系统中；与此同时也包含结构化数据，如物料清单和零件项、产品非

结构化数据的元数据及其文件路径，它们被存储在关系 DBMS 中。

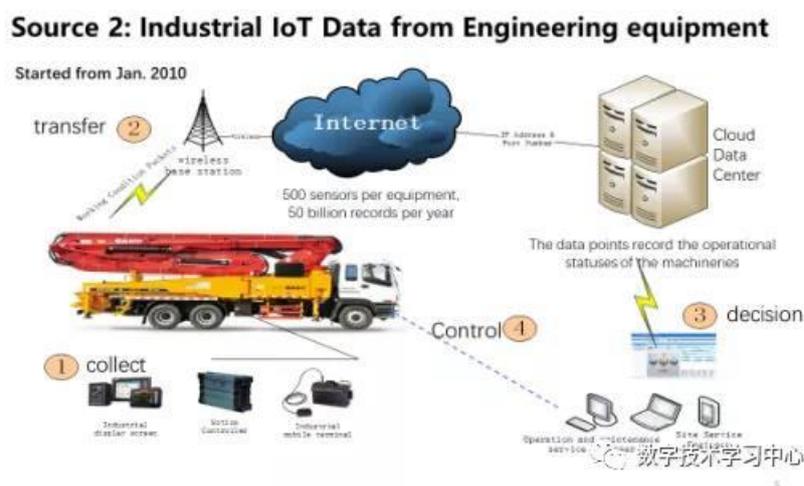
依据 PLM (Product Lifecycle Management, 产品生命周期管理) 理论, 产品的设计和制造阶段也称生命周期的开始阶段 (Beginning of the Life, 下文简称 BOL), 产品的维护和服务阶段也称生命周期的中期阶段 (Middle of the Life, 下文简称 MOL)。为了满足 BOL 数据与 MOL 数据之间的双向连接要求, 我们引入了中性 BOM (Bill of Materials, 产品结构清单) 结构, 将设计制造阶段的 BOM 与服务阶段 BOM 有效协调在一起。中性 BOM 降低了不同生命周期 BOM 之间的关联复杂度, 在企业中得到了广泛应用 (如东方汽轮机有限公司等), 并被作为国家标准进行发布。



工业大数据的第二个数据源是来自工程设备或机械设备的工业物联网数据。为了使设备高效运行, 我们需要尽可

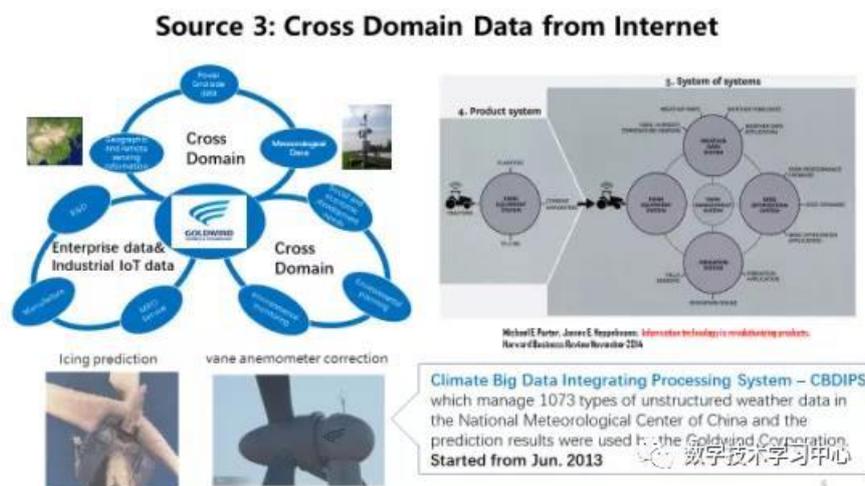
能多地收集、存储和分析工况数据。原设备制造商(如三一重工、中联重科)在机器中嵌入了许多传感器。以挖掘机为例，其在建筑工地工作时，传感器收集数据并通过 WiFi 和 5G 网络将数据发送到云数据中心。这些数据记录了机器设备的运行状态。例如，当设备从一个工地移动到另一个工地时，我们会收集它们的速度、位置和燃料消耗量等数据；当设备工作时，我们收集底盘角度、泵压力等数据。假设一台设备平均有 500 个传感器，那么这 10000 台工程设备每年就会产生 500 多亿条记录。

如今，物联网数据已经成为工业大数据的主体，未来也依旧会在工业大数据的总量中占据主导地位。



第三种数据源是来自互联网和第三方的跨域数据。根据迈克尔·波特的文章，信息技术正在革新工业产品。在未来，大部分产品将接入互联网，成为智能产品。例如，农业设备系

统将与天气数据系统、种子优化系统和灌溉系统协同工作，这就意味着来自互联网和第三方系统的数据将与企业数据和工业物联网数据进行集成和聚合。2013年6月起，我们与中国国家气象中心合作建立了气候大数据集成处理系统。目前，它管理着1073种非结构化实时天气数据，其预测结果已被金风公司所使用——通过预测结冰时间并校正风机叶片对风角的数值，让风力发电机得以平稳运行并产生更多的电力。



正如我们所看到的，工业大数据有三个来源：企业信息
系统、工业物联网和互联网，可用于四种应用场景：

场景一，监控和告警。工业物联网数据和跨域数据可用于设备工作状态和社会事件的监控，并可监控异常报警，甚至对其进行闭环控制。

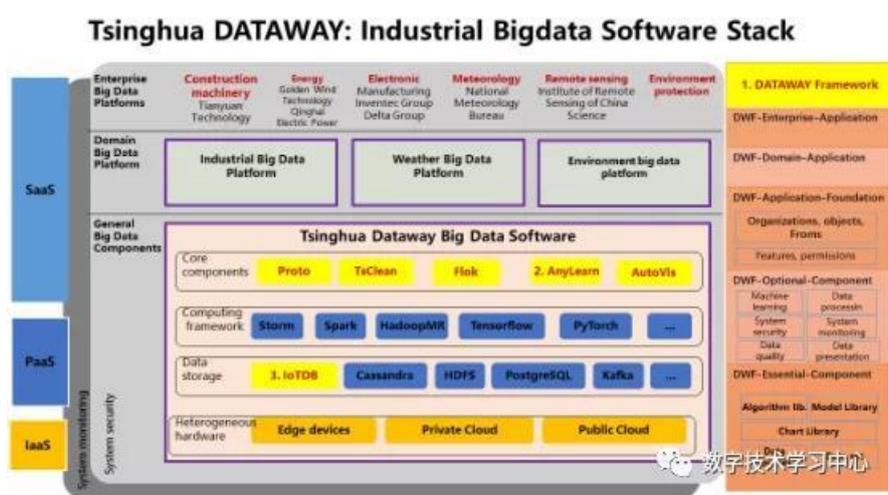
场景二，查询和搜索。ERP、PLM、SCM 等企业信息系统中积累的数据具有较高的价值密度，是工业大数据的主数据。一方面，这些数据用于企业日常运营中的数据查询和搜索任务。另一方面，在企业信息系统数据作为主数据的基础之上，工业物联网数据和跨域数据被组织在一起，形成工业数据湖。

场景三，加工与报表，也就是商业智能（Business Intelligence，下文简称 BI）应用。存储在数据湖的工业原始数据需要经过加工转换（将一个数据集转换为另一个数据集），通常被转换为关键绩效指标（KPI）。处理结果将作为报告进行交付，这就是典型的 BI 应用场景。

场景四，决策与预测、人工智能应用。如果说 BI 应用只是完成了数据集之间的转化工作，那么 AI 应用所做的就是从数据集中提取知识——尤其是训练数据集。如今，机器学习是人工智能应用的主流。因此，深度学习或迁移学习生成的神经网络可以应用于决策和预测工作。

在以上的四大应用场景中，工业大数据生命周期可被划分为收集、管理、处理、分析和应用五个阶段。

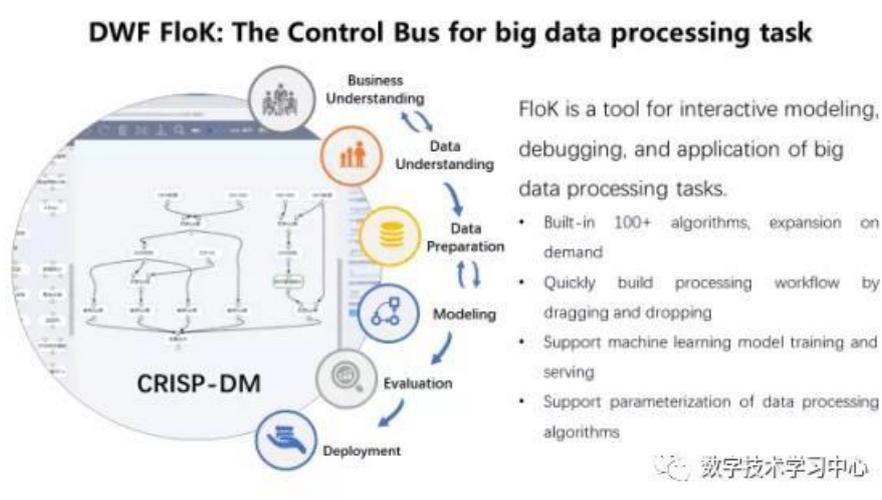
而在具体的工业大数据应用中，这五个阶段可能相互交错的。工业大数据软件栈与大数据生命周期和 DIKW 金字塔（Data-to-Information-to-Knowledge-to-Wisdom Model，即数据-信息-知识-智慧）是一致的。考虑到数据五个阶段的生命周期以及在四个场景下的应用，我们提出了一种新的工业大数据软件架构——清华数为（Tsinghua Dataway）。其中，我们研发了一些针对工业大数据需求的项目（下图中的浅黄色方框），如 IoTDB、TsClean、Flok、AnyLearn、AutoVis 等等。由于时间有限，我将和大家分享其中的三个。



第一个是数为框架（DWF），这是一个用于数据密集型应用程序的快速开发平台。它有两个目的：第一个是快速开发，我们采用模型驱动的架构来改变我们实现应用程序的方式——从传统的硬编码升级到轻量级配置，使得初级工程师也能够以低代码的方式来创建应用程序；第二个是大数据密集型应用程序，这意味着它有一个能让不同的大数据组件（比

如 Hadoop 和 Spark) 易于协作的底层模型，并将这些组件整合到应用程序中，因此用户可以将这个框架作为数据总线、控制总线和交互总线来使用。

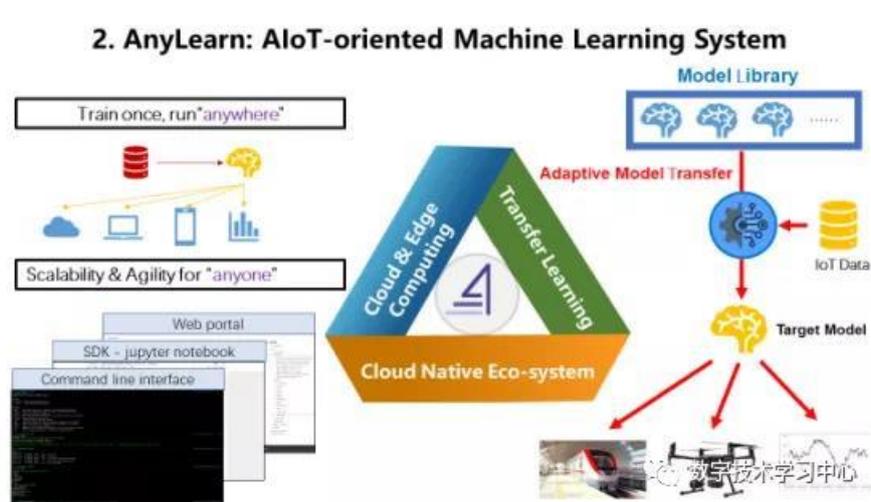
DWF FloK 是大数据处理的控制总线，负责管理大数据软件组件之间的工作流程。众所周知，CRISP-DM 作为公认的工业大数据分析范式共包含有 6 步：业务理解、数据理解、数据准备、建模、评估、部署。FloK 支持通过拖拽操作符快速构建（依照上述分析范式的）数据处理工作流，并拥有 180 多种内置算法。



第二个是 AnyLearn，它是一个面向人工智能的机器学习系统，也是一个云原生的系统。Anylearn 是为工业领域的专家用户构建的，他们拥有丰富的行业知识，但不是机器学习方面的专家。Anylearn 有多种用户友好的 GUI 界面，如交互式 Web 界面，Jupyter Notebook 和命令行。此外，论是在云

上还是在边缘上, Anylearn 都能让生产环境中部署模型变得很容易。最后, Anylearn 将迁移学习能力作为其固有功能, 适用于行业中众多类似场景。

Anylearn 提供了面向不同领域的算法库, 比如天气预报、风力预测以及我们团队研发的迁移学习框架。Anylearn 还为工业物联网场景提供了可在安卓和实时 Linux 操作系统上运行的 Anylearn 边缘推理引擎, ML 模型的推理结果可以与状态图合并 (例如, 将业务规则和来自实时监控的传感器数据进行合并)。我们在实验室搭建了一个由 Anylearn 和 IoTDB 支持的风力预测试验台, 传感器就安装在我们学院东配楼楼顶。它通过一个由太阳能供电的树莓派用 IoTDB 收集传感器数据, 利用 Anylearn 作为引擎提供风力预测服务。风力预测的机器学习模型是在云端训练的。



第三个是一个时序数据库管理系统——IoTDB。这个数据库有三个不同的使用场景。首先，它可以用作终端设备中的一个数据文件，我们提供了高压缩率和简单的写读系统。其次，它可以作为车间级或工厂级的数据库来使用。用作集中控制场景时功能更强大，如资产监控和处理。最后，IoTDB原生支持 Spark 和 Hadoop 等大数据分析框架，使得行业更容易开展产业大数据分析，尤其是基于云计算的行业互联网应用。

IoTDB 项目于 2011 年启动，起源于帮助三一重工升级其三一企业控制中心系统(ECC)，该系统监控全球超过 100,000 台设备。当时 ECC 系统将设备数据存储于关系数据库中。然而，系统的性能不满足如车载混凝土泵锁定和柴油被盗检测等应用的需求。在分析研究了该应用后，我们发现有以下 3 个主要挑战，1) 在工业物联网应用场景中，时间序列的元数据是由终端设备定义的，即新的时间序列可能不需要经过后端注册即可出现。2) 在工业物联网场景中，我们应该尽可能接近现场处理数据，这与 L0 到 L4 工厂层次结构一致。3) 来自于工业互联网的数据通常是关于机器运行状况的数据，信号处理功能会被频繁应用于这类物联网数据集。

3. IoTDB: An Timeseries DBMS for Industrial IoT

- IoTDB is a time series data management system which supports data acquisition, storage, query and analysis. It can be deployed on the edge, the computer, and in the cloud, for individual equipment management, monitoring devices in a factory and timeseries data analytics cloud service.



Manage several time series, support high compression ratio, simple write/query in TsFile format.

A "zip file" of time series



Manage multiple time series, support CRUD and advanced query, e.g., aggregation and subsequence matching.

A "database" of time series

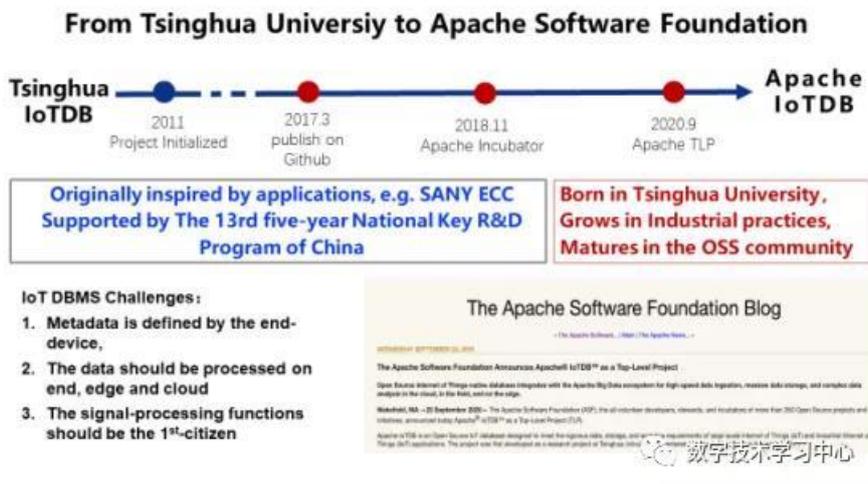


Manage millions of time series and TBs of data, support distributed analysis framework.

A "data warehouse" of time series

数字技术学习中心

从 2015 年开始，我们正式启动 IoTDB 的新版本开发；2017 年我们在 GitHub 上开源了代码；2018 年 11 月，Apache 软件基金会接受 IoTDB 作为孵化器项目；20 个月后，Apache IoTDB 成为 ASF 的顶级项目。



为什么要开源？如今，开源已经成为了软件行业及研究领域的创新范式。我们可以回想谷歌研发的安卓——一个移动操作系统，TensorFlow——一个端到端的机器学习开源平台，前者改变了移动互联网世界，而后者使模型易于搭建和部署。此外，开源是高校对外进行技术输出的一种有效手段，

例如 Spark——用于大规模数据处理的统一分析引擎，Ray——高性能分布式执行框架，它们都来自加州大学伯克利分校。最后，开源对新一代软件人才来说是一种教育平台，在这里他们可以接触到来自现实生活应用的需求，培养出他们在学校里学不到的开发技能。

说到开源的选择，为什么选择 Apache？因为如果只是把代码公布在 Github 上，可能最多就是自己玩一玩这个项目。众所周知，Apache 软件基金会在过去 20 年中获得了诸多成就，然而对我们更重要的是 Apache 的文化。比起代码，他们更相信的是社区——这意味着一个健康的社区比良好的代码更重要。而且，AFS 有很好的行为规范，尊重社区中的每一个人。

下面我来介绍一下我们的开源实践，我负责教授高年级本科生的数据库课程已经有 20 年的时间，这个课程的目标是理解关系数据模型、SQL 语言、数据库设计方法、DBMS（数据库管理系统）结构及其实现。本课程的项目作业之一是开发一个小型数据库管理系统。我鼓励学生从开源项目（以前在使用 HSQL）中学习，并尽最大努力为他们所学习的开源项目做出贡献。此外，在我们学院，软件工程是一门非常重要的学科，这需要较强的实践能力和实践经验。学生参与开源

项目之后，会很好地理解敏捷开发的本质和 SCRUM 项目管理理念。他们还通过单元测试、集成测试和持续集成测试，积累了有关测试驱动的软件开发框架的经验。当他们把代码贡献给社区时，还会用 Sonar 等检查程序，并在源代码中找出问题。

开源平台是一个非常重要的软件工程培训环境。在清华大学软件学院，我们鼓励学生和教师为开源软件项目做出贡献。自 2018 年起，我们学生奖学金的评选标准有了变化：不仅强调论文发表，还考虑学生对开源项目的贡献，如为项目提交代码（pull requests）、bug 修复、提出新想法并被项目采纳等。此外，我们还通过举办开源会议及分享会、高校间项目合作及演讲等方式，积极宣传开源文化。

在演讲的最后，我想强调的是：我们的使命是创新工业大数据技术和软件工具以快速方便地创建应用。我们相信无论在中国还是全世界，工业大数据软件及其应用在未来都是一项长期的工作；清华数为软件栈是我们对这个方向的初步探索，也是我们开源之旅的起点。

最后，我希望大家关注我们的“清华数为”项目，并邀请大家参与到 Apache IoTDB 项目的开发建设中来。