抓住AI时代机遇 RISC-V亟须突围强生态

早在2019年，王东升就作出了RISC-V可能成为AI时代原生计算架构的判断。时隔五年，RISC-V在AI等领域的应用前景逐渐清晰。但是，RISC-V在强生态场景，仍然存在缺应用、缺产品、缺体系的痛点。王东升提出的RDI概念可谓恰逢其时，为RISC-V的生态突围提供了抓手和路径。

本报记者 张心怡

“RISC-V是一个新鲜事物，生态还不完善，需要底层芯片、硬件设备、软件系统、场景应用等上下游生态伙伴之间的开放合作、协同创新、共建共赢。为此，我提出RDI这样一个产业概念，即RISC-V Digital Infrastructure，RISC-V数字基础设施。”在7月26日举办的“RDI生态·武汉创新论坛·2024”上，奕斯伟集团董事长王东升首次提出RDI产业概念。

RISC-V在强生态场景仍存挑战

2019年7月，从京东方卸任不足一个月的王东升，应邀加入奕斯伟团队，在集成电路领域开始第二次创业。基于对集成电路领域的长期思考，他以基础材料和计算芯片为着眼点，希望助力产业破解制造端和产品端的难点：从事电子级12英寸大硅片的奕斯伟材料公司主攻基础材料，目前出货量已居国内头部；奕斯伟计算公司则瞄准RISC-V新一代计算架构，主攻底层技术和计算芯片。

“那时候选择RISC-V这个方向是需要勇气的，但在反复的思考和探讨之后，我们在当时认为AI时代的原生架构可能是RISC-V。一方面，它是绿色、开放的，不断演进变化的，可以更好地适配AI时代软件与算法模型不断优化的需求。另一方面，它既能通过移植适配，与原有架构的生态兼容，也可以作为原生架构，支撑不断涌现的新应用、新场景。这意味着它是可以落地的，不是一个悬空的无人区，所以我们把RISC-V作为攻关方向。”王东升对《中国电子报》记者说。

2019年9月，定位为“AI时代以RISC-V为基础的新一代计算架构芯片与方案提供商”的奕斯伟计算公司在北京成立。那一年，RISC-V主要应用于物联感知、微控制器、无线连接等低功耗且生态依赖较弱的场景。在RISC-V国际基金会等行业机构、研究机构、高校，以及奕斯伟计算等国内外企业的共同努力下，RISC-V逐步迈向高性能领域。2019年之后，一系列支持RISC-V高性能扩展的标准陆续颁布，比如RISC-V 矢量扩展规范，可帮助用户使用向量体系结构来优化和加速程序代码；RISC-V 架构 BF16指令集扩展标准化提案，支持BF16这一与FP32有相同数值范围但精度更低的数据类型，有利于提升深度学习效能。另外，H扩展、IOMMU和AIA等规范也达到了稳定的1.0版本，提供了CPU虚拟化、内存虚拟化、中断虚拟化和I/O虚拟化等方面的支持，从而为RISC-V迈向高性能领域提供了坚实的基础。

在此基础上，伴随着生成式人工智能的科技浪潮，RISC-V在AI加速、边缘计算、智能终端等领域的应用也崭露头角，通过矢量以及正在建设中的张量等指令扩展方式，不断完善对AI计算的支撑。在本次论坛展区，记者看到了搭载奕斯伟RISC-V边缘计算SoC EIC7700X的安防场景样例，该系统可以同时解码和识别多路摄像头中的人、车、特定目标物等主体。奕斯伟还提供了以EIC7700X为主控芯片的64位RISC-V开发板，并演示了该开发板支持运行的AI PC电脑桌面和大语言模型应用场景。

“五年前，我们说AI时代原生的计算架构可能是RISC-V，现在我们相信，RISC-V一定能成为AI时代的原生计算架构。”王东升对记者说。

虽然RISC-V在AI领域的天然优势和应用潜力变得越来越清晰，但这种优势不能仅仅通过底层技术来兑现，也要像X86架构的PC处理器、ARM架构的手机处理器一样，能被广大行业用户和消费者感知和使用。而用户触达AI能力的方式，往往是通过各式各样的APP。从生态层面来看，计算平台承载的APP数量越多、类目越丰富，对于生态的要求也就越强，比如手机、笔记本电脑、公有云等，就是典型的强生态场景。人工智能作为底层赋能技术，也将融入大量强生态场景。市场调研机构预估：随着AIGC技术的发展，到2024年全球将涌现出超过5亿个新应用，相当于过去40年间出现的应用数总和。

然而，RISC-V在强生态场景，还处于零星布局的状态。奕斯伟计算高级副总裁、CTO何宁将RISC-V当前的挑战归纳为缺应用、缺产品、缺体系。

所谓缺应用，是指RISC-V软件生态尚不完善，在物联网场景之外的应用相对匮乏，尚无在强生态场景规模化落地案例。缺产品是指RISC-V现有产品矩阵还不够丰富，无法覆盖各类应用场景需求。而缺体系，指RISC-V应用目前还是靠各个企业单打独斗，缺乏统一的旗帜来号召和组织产业链资源，形成体系化、规模化的应用场景。

而RDI这一产业概念，为RISC-V的生态建设，提供了抓手和路径。

以RDI为旗帜 推动RISC-V走向显性

早在2022年，仅仅问世12年的RISC-V就达到了百亿颗内核出货量的里程碑。但是，这些内核较为零散地分布在各类芯片中，难以形成显性可见的应用成果。

RDI的提出，整合了RISC-V的产品体系，推动RISC-V在行业场景的体系化、规模化落地，将RISC-V的优势和作用显性地体现出来，既提升了各界对RISC-V的认知度和接受度，也能够吸引更多从业者参与RISC-V的生态建设。

在王东升看来，RDI对产业有两个利好：一是将RISC-V的蛋糕做大，二是帮助所有从业者开拓更广阔的应用场景。“RDI就像海上的冰山，海平面以上的部分是透明、开放的，海平面以下的部分是企业各自的能力。对于海平面以上的部分，我们要完善生态，把蛋糕做大。海平面以下的部分，则需要IP层到系统层的生态链伙伴持续提升各自的技术实力，再一起反哺RISC-V生态。”王东升向《中国电子报》记者表示。

对于“海平面以上的部分”，奕斯伟正在建设包含IP层、芯粒层、芯片层、板卡层和系统层的开放平台，通过开放的软硬件接口，让更多的企业参与RISC-V产业生态建设，并在开放合作的过程中获得商业回报。

垂直行业场景是RDI部署落地的切口。本次RDI生态论坛举办了“武汉RISC-V生态创新中心”揭牌仪式。该中心是在武汉市委市政府、东湖高新区的支持下，由武汉数据集团、武汉光谷金控集团、聚力开源公司（由奕斯伟计算等众多RISC-V生态链企业联合发起）在2023年10月共同组建，其任务是携手RISC-V生态链伙伴遴选垂直行业场景，通过组建产品矩阵、形成系统方案、开展测试验证、实现应用示范及推广，推动RDI生态建设和产业发展。成立至今，该中心牵引元石智算等生态链伙伴开发了RISC-V服务器及跨架构云原生服务器集群。

据武汉RISC-V生态创新中心总经理周宇虹介绍，在私有云的使用场景中，客户一般以集群的方式进行算力部署，根据使用场景的不同进行服务器数量和种类配置，并通过统一的云管平台进行算力调度。“在RISC-V生态成熟的应用场景中，可以大量使用RISC-V服务器。在RISC-V生态尚未完善的场景中，可以将部分应用部署在X86的服务器上，进行跨架构集群。目前RISC-V跨架构云原生服务器集群已经有若干原型和部署实例。”

RDI概念的提出，也对应了何宁提到的RISC-V的三个缺项，将推动RISC-V从隐形走向显性的应用。

面对缺应用的问题，RDI以行业应用作为突破口，逐步向生态需求更强的场景渗透。相比笔记本电脑、手机等应用繁多的强生态场景，垂直行业应用呈现出限定的生态环境。RISC-V可以优先在垂直行业的限定生态场景落地，再逐步向强生态场景突围。面对缺产品，RDI以应用示范项目作为牵引，以创新中心等生态平台组织作为桥梁，围绕实际需求，有组织、有计划、有目的地完善产品矩阵。面对缺体系的问题，产业可以以RDI为旗帜，以领军企业为链主，构建体系化的行业解决方案，推动RISC-V进入显性的应用阶段。“当我们聚集在RDI这面旗帜之下，聚焦行业应用，构建丰富的产品矩阵，RISC-V一定能进入到显性化的应用阶段。进入这个阶段之后，我相信大家不会再问为什么RISC-V出货了这么多，却看不到应用在哪里。”何宁说道。

中国电子报 全媒体
权威性高 传播力强 覆盖面广 影响力大

融媒体服务

- 报纸出版
- 官方网站（电子信息产业网www.cena.com.cn）
- 官方微信（公众号cena1984）
- 官方微博（http://weibo.com/cena1984）
- 视频平台
- 视频服务（视频制作、在线直播、在线会议等）
- 平台推广
- 内参专报
- 行业报告
- 图书出版

会赛展服务

- 会议活动
- 专业大赛
- 展览展示
- 专业培训
- 政府服务
- 指数发布
- 编辑推荐
- 产品评测
- 企业定制
- 舆情监测
- 数据营销
- 招商引资

立足电子信息业
服务新型工业化

中国电子报社创建于1984年。目前拥有集报纸、网站、微信、微博、音视频、第三方平台等全媒体服务，集会议活动、展览展示、专业大赛、定制服务等会赛展训服务于一体的立体化、多介质系列产品，是促进行业高质量发展的“喉舌”与“纽带”。

《中国电子报》是具有机关报职能的权威媒体。《中国电子报》全媒体面向工业和信息化领域，聚焦集成电路、新型显示、智能终端、信息通信、人工智能、物联网、工业互联网、移动互联网、大数据、云计算、区块链、应用服务等电子信息完整产业链。

《中国电子报》全媒体日均触达用户量超过200万。

国内统一连续出版物号：CN11-0005
邮发代号：1-29

官方微信 官方网站

在这里让我们一起把握行业脉动
www.cena.com.cn

地址：北京市海淀区紫竹院路66号赛迪大厦18层
电话：010-88558808/8838/9779/8853
传真：010-88558805

广告