

华为鲲鹏920服务器芯片触动谁的神经？

本报记者 李佳师

1月7日，华为宣布推出业界最高性能ARM-based处理器——鲲鹏920(Kunpeng 920)，以及基于鲲鹏920的TaiShan服务器、华为云服务，并宣布推动相关的生态计划。服务器芯片市场是利润丰厚的市场，目前英特尔在X86服务器芯片市场拥有绝对优势，此前包括高通、三星和NVIDIA等都曾跃跃欲试基于ARM架构来研发服务器芯片，但都因种种原因告退。华为此刻进入服务器市场将面临什么挑战？未来服务器芯片市场将呈现出怎样的发展趋势？



在更高性能与更低功耗上有突出表现的服务器芯片，一定还有市场空间。

功耗降低30%是否有竞争力？

据华为介绍，鲲鹏920是目前业界最高性能ARM-based处理器。处理器采用7nm制造工艺，基于ARM架构授权，由华为公司自主设计完成。鲲鹏920主频可达2.6GHz，单芯片可支持64核。该芯片集成8通道DDR4，内存带宽超出现业界主流46%。芯片集成100G

RoCE以太网卡功能，大幅提高系统集成度。鲲鹏920支持PCIe4.0及CCIX接口，可提供640Gbps总带宽，单槽位接口速率为业界主流速率的两倍，有效提升存储及各类加速器的性能。

据悉，鲲鹏920通过优化分支预测算法、提升运算单元数量、改进内

存子系统架构等一系列微架构设计，大幅提高处理器性能。典型主频下，SPECint Benchmark评分超过930，超出业界标杆25%。同时，能效比优于业界标杆30%。鲲鹏920以更低功耗为数据中心提供更强性能。

目前全球的数据中心基础设施需求仍高速增长，而且对功耗的要求

越来越苛刻，此前，IDC的调查数据显示，服务器的费用中有50%~70%是用于服务器的供电和冷却。不少被访者表示，服务器芯片研发大有可为：不管是其功耗方面还是性能方面都还有巨大的潜力，所以能够在更高性能与更低功耗上有突出表现的服务器芯片一定还有市场空间。

华为推出ARM-based处理器是为了满足用户多样性异构计算的需求。

挑战英特尔还是拥抱多样性？

目前英特尔在X86服务器芯片市场拥有绝对优势，而华为又是英特尔X86服务器芯片的大客户，最新数据显示，华为服务器2018年出货量达到91.8万台。华为推出基于ARM的服务器芯片，让华为与英特尔的关系变得异常敏感，未来华为与英特尔是合作伙伴还是对手？

华为在宣布推出鲲鹏920的同时，第一时间将业界所担忧的与英特尔公司关系的敏感窗户纸捅破，明确表示，华为与英特尔仍是亲密战友，推出ARM-based处理器是为了满足用户多样性异构计算的需求。

华为董事、战略Marketing总裁

徐文伟表示：“华为在计算领域围绕客户价值持续创新；我们预见，随着智能社会的到来，未来这一领域还将持续增长。当前业务和数据的多样性带来了异构计算的需求，长期以来，华为与Intel共同合作取得了很好的成绩，为ICT产业发展作出积极贡献，华为与Intel将长期保持

战略合作、持续创新。”徐文伟表示，ARM产业迎来新的发展机会，华为发布鲲鹏920及TaiShan服务器，主要应用于大数据、分布式存储、ARM原生应用等场景。华为希望携手全球合作伙伴，共同促进ARM生态发展，拥抱多样性计算时代。

不少厂商都曾希望基于ARM架构研发服务器芯片，但生态不成熟，是厂商望而却步的关键原因。

生态成熟度将是鲲鹏920最大挑战

异构计算、多样选择的需求并非今天才有，此前三星和NVIDIA等都曾跃跃欲试基于ARM架构研发服务器芯片，但都因种种原因最后偃旗息鼓。而2014年高通宣布基于ARM架构研发服务器芯片，曾一度被认为是打破英特尔服务器芯片一家独大的最有力挑战者，但在2018年，这个信心被高通将放弃ARM服务器芯片的种种行为和传言击破，尽管高通总裁对此进行了否认。

此前，赛迪顾问分析师在接受《中国电子报》记者采访时曾表示，不少厂商都曾希望基于ARM架构研发服务器芯片，但ARM的服务器生态不成熟，是众多厂商望而却步的最为关键原因。

华为能够让ARM服务器芯片起飞吗？业内资深人士认为，在很

多领域华为都可以称为市场的“晚到者”，但都后来居上把市场做了起来，所以华为有可能给基于ARM架构服务器芯片市场带来变数。

其一是华为本身就是全球领先的服务器、存储、云业务的系统和服务提供商，自己就可以采用鲲鹏920。在宣布推出鲲鹏920芯片的当天，华为同步推出基于鲲鹏920的TaiShan系列服务器，包括均衡型、存储型和高密型三个机型。TaiShan系列服务器主要面向大数据、分布式存储和ARM原生应用等场景，发挥ARM架构在多核、高能效等方面的优势，为企业构建高性能、低功耗的新计算平台；例如大数据场景，实现了多核高并发和资源调度调优，计算性能提升20%。基于TaiShan服务器，华为云也将提

供弹性云服务、裸金属服务和云手机服务。

其二是华为宣布将在多个层面构建ARM生态。华为宣布将从硬件、基础软件和应用三个层面不断推进产业合作。长期以来，华为与GCC、Linaro、OEHII等产业组织，以及 Hortonworks、Microsoft、Oracle、SAP、SUSE、Ubuntu、中标软件等合作伙伴共同打造开放、合作、共赢的生态环境。在硬件层面，华为已成为Linaro的核心成员；在基础软件层面，华为已成为OpenStack的白金会员、CNCF的Founder成员；在应用层面，华为参与绿色计算产业联盟GCC、GCC发布《绿色计算产业联盟服务器标准白皮书》，构建绿色计算开源开放社区；华为在欧洲参与“OpenEdge+HPCInitiative”。

其三从用户需求和服务器芯片

求变等角度看，需要新元素加入。事实上，服务器芯片市场需要更多的创新，随着云计算与更多应用场景的需求变化，服务器芯片的发展也越来越呈现出多元化的发展趋势。业界各大企业巨头目前都在随着需求的演进而不断调整服务器芯片与系统的发展路线，希望摆脱英特尔服务器芯片一家独大的被动局面。从用户需求的角度看，今天的客户的业务发展要求服务器芯片能够实现更多的定制化。尤其是互联网用户希望构建以数据为中心的基础设施，能够进行更深层次的定制以获得能耗下降与效率上的大幅提升，就像华为所言：多样性计算时代正在到来，多种数据类型和场景驱使计算架构的优化，多种计算架构的组合是实现最优性能计算的必然选择。

腾讯张小龙：微信的原动力从来没变过

保、尽可能少地体现它的设计。

张小龙表示，现在有太多人只关心产品的流量和变现，很少关心自己的产品是不是真的很好。技术的使命应该是帮助人类提高效率；如果是信息资讯类产品，就应该让用户快速找到最有用的信息，而不是一直在里面看个不停。

张小龙表示，微信永远不会将用户停留时长作为目标，反而更加关心用户在这里发布照片、阅读文章、完成支付、找到想用的小程序等行为，能不能快速高效完成。

张小龙说，微信的原动力（初心）可以概括为两点，首先是坚持做一款好的、与时俱进的工具。第二个原动力是让创造体现其价值。“一

一个好的产品是有自己使命的。我很庆幸这么多年过去，微信的原动力从没变过。”张小龙说。

目前，微信事业群(WXG)下辖主要有六大业务板块：公众平台、微信支付、企业微信、小程序、小游戏和微信搜索。

在演讲中，张小龙对全部业务进行了介绍并谈及了其中的满意与不满意。关于小程序，目前有很多科技公司都在做与之类似的产品，但张小龙不认为这些会对微信小程序造成威胁。对于小游戏，张小龙表示，从商业角度看，发展很好，收入很不错，但他并不满意，因为大部分小游戏还是互相拷贝、洗用户流量，并不是真正高质量的原创游

戏。关于微信公众平台，张小龙认为，近几年微信公众平台的内容质量提升并不明显，对用户的贡献程度有所降低。主要是微信没有找到核心问题——内容是否对用户有吸引力。接下来，微信公众平台将重点鼓励内容创作者生产更优质的内容。关于微信支付，张小龙认为，这款产品做到了“润物细无声”，用户很少感知到它的存在。对于朋友圈，张小龙表示不太满意，现在有超过1亿人将朋友圈设置成了3天可见，这是一种压力状态下的反应。关于微信“看一看”中的两个板块：“好看”和“精选”，它们分别代表着两个方向的尝试，前者代表社交推荐，后者代表机器推荐。

阿里达摩院发布2019十大科技趋势

日前，阿里巴巴达摩院发布了“2019十大科技趋势”，涵盖了智能城市、数字身份、自动驾驶、图神经网络系统、AI芯片、区块链、5G等领域。在这次趋势预测中，达摩院认为，AI依然将是科技界最热的方向。2017年，阿里巴巴成立了“达摩院”，并将在未来3年内投入1000亿元用于涵盖基础科学和颠覆性技术创新的研究。2019年第一个工作日阿里达摩院就发布了2019十大科技趋势。

● 趋势1：城市实时仿真成为可能，智慧城市诞生

城市公共基础设施的感知数据与城市实时脉动数据流将汇聚到大计算平台上，算力与算法发展将推动视频等非结构化信息与其他结构化信息实时融合，城市实时仿真成为可能，城市局部智能将升级为全

局智能，未来会出现更多的力量进行城市大脑技术和应用的研发，实体城市之上将诞生全时空感知、全要素联动、全周期迭代的智能城市，推动城市治理水平优化提升，中国会有越来越多城市具有大脑。

● 趋势2：语音AI在特定领域通过图灵测试

随着端云一体语音交互模组的标准化、低成本化，会说话的公共设施会越来越多，未来每一个空间都至少会有一个可以进行语音交互的触点。随着智能语音技术的提升，

移动设备上的实时语音生成与真人语音可能将无法区分，甚至在一些特定对话中通过图灵测试。针对这一领域的规则甚至法律会逐步建立，引导行业走向规范化。

● 趋势3：AI专用芯片将挑战GPU的绝对统治地位

当下数据中心的人工智能训练场景下，计算和存储之间数据搬移已成为瓶颈，新一代的基于3D堆叠存储技术的AI芯片架构已经成为趋势。AI芯片中数据带宽的需求会进一步推动3D堆叠存储芯片在AI训练芯片中的普遍应用。而类

脑计算芯片也会在寻找更合适的应用中进一步推动其发展。在数据中心的训练场景下，AI专用芯片将挑战GPU的绝对统治地位。真正能充分体现Domain Specific的AI芯片架构还是会更多地体现在诸多边缘场景。

● 趋势4：超大规模图神经网络系统将赋予机器常识

单纯的深度学习已经成熟，而结合了深度学习的图神经网络将端到端学习与归纳推理相结合，有望解决深度学习无法处理的关系推

理、可解释性等一系列问题。强大的图神经网络将会类似于由神经元等节点所形成网络的人的大脑，机器有望具有理解、认知能力的AI。

● 趋势5：计算体系结构将被重构

无论是数据中心或者边缘计算场景，计算体系将被重构。未来的计算、存储、网络不仅要满足人工智能对高通量计算力的需求，也要满足物联网场景对低功耗的需求。基于FPGA、GPU、ASIC等计算芯片的异构计算架构，以及新型存储器的出现，已经为传统计算架构的演进拉开了序幕。从过去以CPU为核心的通用计算而走向由应用驱动(Application-driven)和技术驱动所体现的Domain-specific体系结构的颠覆性改变，将加速人工智能甚至是量子计算黄金时代的到来。

● 趋势6：5G网络催生全新应用场景

第五代移动通信技术将使移动带宽大幅度增强，提供近百倍于4G的峰值速率，促进AR/VR等沉浸式交互模式的逐步成熟。连接能力将增强至百亿级，带来海量的机器类通信及连接的深度融合。网络向

云化、软件化演进，网络可切片成多个相互独立、平行的虚拟子网络，为不同应用提供虚拟专属网络，加上高可靠、低时延、大容量的网络能力，将使车路协同、工业互联网等领域获得全新的技术赋能。

● 趋势7：数字身份将成为第二张身份证件

生物识别技术正逐渐成熟并进入大规模应用阶段。随着3D传感器的快速普及、多种生物特征的融合，每个设备都能更聪明地“看”和“听”。生物识别和活体技术也将重

塑身份识别和认证，数字身份将成为人的第二张身份证。从手机解锁、小区门禁到餐厅吃饭、超市收银，再到高铁进站以及医院看病，靠脸走遍天下的时代正在加速到来。

● 趋势8：自动驾驶进入冷静发展期

单纯依靠“单车智能”的方式革新汽车，在很长一段时间内无法实现终极的无人驾驶，但并不意味着自动驾驶完全进入寒冬。车路协同技术路线，会加快无人驾驶的到

来。在未来2~3年内，以物流、运输等限定场景为代表的自动驾驶商业化应用将迎来新的进展，例如固定线路公交、无人配送、园区微循环等商用场景将快速落地。

● 趋势9：区块链回归理性，商业化应用加速

在各行业数字化的进程中，物联网技术将支撑链下世界和链上数据的可信映射，区块链技术将促进可信数据在流转路径上的重组和优化，从而提高流转和协同的效率。

随着“链接”价值的体现，分层架构和跨链互联将成为区块链规模化的技术基础。区块链领域将从过度狂热和过度悲观回归理性，商业化应用有望加速落地。

● 趋势10：数据安全保护技术加速涌现

各国政府都会倾向于推出更加严厉的数据安全政策法规，企业在个人数据隐私保护上投入更多力量。未来几年，黑客、黑产攻击不会

停止。跨系统的数据追踪溯源相关的技术，比如水印技术，数据资产保护的技术以及面向强对抗的高级反爬虫技术等将得到更加广泛应用。