



华为 x86 服务器的取舍逻辑

本报记者 李佳师

差不多一年前,华为公布旗下荣耀手机业务的出售方案。一年之后的今天,11月8日,工商登记显示,华为控股的超聚变数字技术有限公司股东已变更为河南超聚能。这意味着此前传闻华为 x86 服务器业务出售一事已有实质性进展,目前华为公司没有发布官方信息。但根据分析,华为或将在全部股东的工商变更完成、业务平稳过渡之后,才会正式发布公告。今年早些时候,坊间传闻华为将打包出售旗下服务器产品线,当时华为内部人员予以否认,表示华为不会放弃服务器业务。在9月的华为媒体见面会上,华为轮值董事长徐直军表示,华为正在考虑出售 x86 服务器业务,并已经与潜在投资者接触。近日,超聚变数字技术有限公司股东变更完成,华为 x86 服务器业务出售一事的“另一只靴子”落地了。

出售或许是最好的选择

工商信息显示,河南超聚能科技成立于2021年8月30日,法定代表人为李亚东,注册资本30亿元,经营范围包含软件开发、大数据服务、网络技术服务等。该公司由河南豫上信联企业管理咨询中心(有限合伙)持股80%,河南信息产业投资有限公司持股20%。其中,河南豫上信联企业管理咨询中心(有限合伙)由吴斌持股约98.04%。

据了解,华为 x86 服务器业务的买家将由产业基金、海外国家主权基金、互联网公司、银行等多方社会资本组成。“此次工商变更只是完成了第一步,后续其他投

资方将陆续完成工商变更。”有关人士透露。等各家投资方完成工商变更后,新公司的组团与运作方式才会逐渐清晰。

华为出售 x86 服务器业务实属无奈,但或许是最好的选择。据悉,在被列入“实体清单”前,华为服务器业务年销售额约在300亿~400亿元左右的规模。在被列入美国“实体清单”之后,华为的 x86 服务器芯片断供,其服务器业务在2020年第三季度跌到了全球第五(IDC数据),今年上半年则跌出了前五;在中国服务器市场今年上半年前五位排名中,同样没有了华为的名字,前五位的厂商排名依次为浪

潮、新华三、华为、戴尔和联想。作为数字信息基础设施重要组成部分的服务器,目前依然增长旺盛,IDC近日发布的报告显示,2021年上半年,中国服务器市场出货量同比增长8.9%,市场规模同比增长12.1%。

面对不断增长的市场需求,卖掉 x86 服务器业务给其一个持续发展的未来之路,这和华为剥离荣耀手机业务如出一辙。与此同时,x86 业务竞争越来越激烈,利润越来越薄,也越来越同质化,在这样的背景下,放弃或许是更好的选择,就像当初IBM卖掉 x86 业务一样。

华为将在时机成熟时退出 x86 服务器业务,并决定未来5年投资30亿元推动鲲鹏处理器生态发展。

进一步做强鲲鹏业务

华为的服务器业务有两条线,左手是 x86 服务器,右手是基于 ARM 的鲲鹏服务器业务。事实上,华为一直在不断强化右手业务。

2019年,华为发布兼容 ARM 的处理器鲲鹏920,推出基于鲲鹏920的TaiShan系列服务器产品和云服务。鲲鹏920采用7nm工艺制造,可以支持32/48/64个内核,主频可达2.6GHz,支持8通道DDR4、PCIe 4.0和100G RoCE网络。

2019年7月23日,徐直军在北京宣布“鲲鹏生态计划”,并表示,华为将在时机成熟时退出 x86 服务器业务,并决定未来5年投资30亿元推动鲲鹏处理器生态的发展。

2B 业务转型加速向云

不再纠缠 x86 服务器,不再纠缠服务器,或许是华为全面走向云计算的更好契机。

2017年华为云业务成立,徐直军在接受媒体采访时表示,华为云关乎华为未来10年将是什么样,华为必须下大力气发展云计算业务。

但必须承认,华为云业务前几年做得并不理想。很多业界人士表示,华为的硬件基因、硬件思维阻碍了云计算业务的发展。

今年,华为对于云的加持力度前所未有。年初,华为“心声社区”发布任正非在企业业务及云业务汇报会上的讲话,要求华为云必须改变卖硬件的思路,转向买服务的思路。

2020年年初,华为公布了鲲鹏+昇腾生态进展。时任华为云与计算BG总裁侯金龙表示,希望用3年的时间让90%的应用都可以跑在鲲鹏上。而从2020年中旬开始,华为鲲鹏事业部的相关技术专家更是多次公开讲解和演示“将代码迁出 x86 架构”的好处与方法。

徐直军在今年举行的华为分析师大会上表示,一旦鲲鹏生态发展起来了,就会有其他企业设计出CPU来,这将给中国提供另外一种选择。徐直军透露,目前已经有12家整机伙伴推出鲲鹏系列服务器和PC产品。2020年,伙伴服务器出货量达到鲲鹏服务器出货的50%以上,今年预计达到80%以上。同时还有8家伙伴发

同样是今年,华为云的掌门人几次更换,其业务组织架构几次大调整,包括4月2日,华为任命张平安为Cloud BU总裁,并撤销原“云与计算BG”。4月9日,华为再次发文对华为云业务进行了新的人事任命,华为轮值董事长徐直军任华为云董事长(兼),华为消费者BG CEO余承东被任命为华为云CEO,张平安为云BU行政管理团队主任。华为频繁调兵换将将进行组织架构的调整,就是希望软硬分离,进一步做强云计算。

走云与硬件分离这一步棋是必需的,如果云和硬件产品捆绑一起卖,很容易回到卖硬件的轨道上,卖硬件和卖服务是完全不同的思维方式。我们看看

布了openEuler操作系统商业版本,6家发布了openGauss数据库商业版本,2000多家合作伙伴的4500多个解决方案通过了鲲鹏兼容性测试认证。可以说,目前鲲鹏在关系国计民生的行业中获得了规模应用。

目前华为的鲲鹏生态处于紧锣密鼓的加速推进中。《中国电子报》记者今年年中在湖南长沙采访湘江鲲鹏和拓维信息时了解到,这两家企业分别是华为鲲鹏生态中的服务器制造商和软件合作伙伴,他们都提到了湘江鲲鹏“39天封顶、70天建成、120天交付”的“长沙加速度”。

华为卖掉 x86,或许是做强鲲鹏的更好契机。

华为将云视为未来,并做好了奋力追赶的准备,不断减持重硬件重资产,是其破釜沉舟的机会。

目前IT业界,那些做服务器、卖服务器的厂商目前都不是排名前5位的公有云厂商,现在前几名的公有云厂商不是做软件出身就是互联网企业。

华为将云视为未来,并做好了用10年时间进行奋力追赶的准备,不断减持重硬件重资产,是其破釜沉舟的机会。

近日发布的华为2021年第三季度财报显示,华为前三季度实现销售收入4558亿元,同比下降约32%。华为轮值董事长郭平表示:“整体经营结果符合预期,To C业务受到较大影响,To B业务表现稳定。我们将继续加强技术创新、研发投入和人才吸引,不断提升运营效率,我们有信心能够为客户和社会持续创造价值。”

“5G+工业互联网”建设按下加速键

本报记者 沈丛

近年来,5G+工业互联网建设按下加速键,多种应用场景不断拓展,呈现新技术赋能产业转型的蓬勃发展态势。站在5G快速发展的当下,日前,紫光展锐举办了5G行业应用研讨会,携手产业链多位5G行业探索先锋,以“聚链凝智 融合共享”为主题,探讨行业发展趋势,分享应用成果,展望5G行业发展的未来。

5G To B 赋能千行百业

随着5G大规模商用的普及,数字世界和物理世界的边界将被打破,5G在To C方面得到广泛应用的同时,也开始在To B方面赋能千行百业。记者在会上了解到,目前,国内5G已经渗透到了工业、农业、建筑、医疗等十多个行业。从全国30多个省市的统计数据来看,5G行业应用已经成为各个省市发展的重点。2021年,5G应用的案例数量已经达到了1.2万多个,可见,5G行业应用已经和各行各业发生了一些化学反应。

中国移动上海产业研究院的吴冰表示,目前中国移动已经在18个行业沉淀了100多个可落地的行业应用场景,总共5000多个项目开始在做复制推广。此外,中国移动还提出了“九天人工智能平台”,聚焦在工业、交通、金融、医疗、教育、智慧城市等几大行业,赋能行业数智化转型。

物流方面,京东展锐开展了深度合作。京东物流5G智能物流架构师陈亚述表示,在仓储机器人、视觉CPE等方面都用了基于展锐芯片的一些模块。目前京东物流运营了38个大型“亚洲一号”智能物流园区,并联合合作伙伴发布了“5G全连接智能仓”,涉及从入库搬运、分拣到监控、运维的运营场景。

在医疗方面,紫光展锐与海南联通、联通物联网有限公司等单位共同申报的5G应用创新方案“5G智慧医疗点亮海南健康岛”,是全国首个基于5G技术覆盖全省所有基层机构的远程诊断信息化项目,围绕远程会诊、智慧分级诊疗、智慧医院、急救系统场景开展5G智慧应用。

5G 行业融合 仍面临重重挑战

在5G To B正在不断赋能千行百业的风口下,5G与各个行业之间的融合问题也面临着重重挑战。

陈亚述表示,5G To B主要有四个方面的痛点。首先,5G服务能力的问题。5G To B的服务不够标准化,面向行业还需加强,以满足各行各业的需求。此外,To B相比To C客户会有更高的要求,包括对网络的保障、使用等,需要有更高的安全性与灵活性。因此,5G对于To B方面的服务要求更高。

其次,成本较高。包括运营商在内,许多厂商仍旧是拿To C的网络结构和设施来服务

To B的需求,因此成本难以控制。对此,许多厂商也一直在做改变,包括跟运营商成立联合实验室等。在To B的服务中,可能需要一些轻量化、更低成本的设施,包括低成本的UPF网关等来满足相关的需求。To B的客户比To C的客户更加敏感,因此需要控制5G的成本,来满足“投入产出”的有效计算。

再次,欠缺提升一体化的解决方案。对于不同行业,5G帮助该行业起到降本增效的成果才是客户的刚需。不管是终端,还是网络、平台产品,5G都需要融入到信息化、智能化中,然后形成一体化的解决方案,需要与行业做有效融合,切实满足行业需求。

最后,需要打造“1+1>2”的产品。5G与其他行业的结合,并不仅仅只是一个“面子工程”,也不是简单的“1+1”的结合,而是需要产生“1+1>2”的产品,以更低成本、更高的效能服务更多的行业。对于产品的打造,5G在To B方面依旧有待提高。

打造5G 行业融合创新 可从芯片精简化下手

众所周知,5G通信技术的发展,离不开5G芯片的加持,而在5G从To C到To B转变的过程中,5G芯片的精简化发展,将有效破解5G行业融合创新的种种难题。

尽管每个终端类型的作用都比较有局限性,然而每个终端都有定制化需求。这样对整个模组、芯片,都会提出差异化的需求。因此,这种需求对于传统的面向To C的规模化铺货思路 and 整个推广方式,其实并不适合。因此,为了能够适配行业终端的发展态势,需要在模组、芯片等方面做很多工作,要解决定制化、高成本和研发费用高的难题,从而实现既能满足行业的要求,又能够使行业终端研发或者出货过程中的成本降低。

因此,有专家指出,做芯片的精简化有助于推动5G向To B方向发展,尤其是能有效解决目前行业面临的“终端定制化、终端贵”的问题。如今,轻量化芯片也面临着需要寻找一个规模化市场和规模化应用的问题。这些精简化的功能对于行业应用或者终端而言,必须了解哪些功能是要保留的、性能要达到怎样的程度,同时也要对现有的基于R15、R16的芯片做功能的筛选,或者增加相应的功能,例如,国家电网需要高精度芯片的支持,然而很多功能对于其他行业而言,是不需要的,因此可以在芯片方面适当做一些精简,在提升服务质量的同时,也能降低成本。

5G产业的蓬勃发展,需要产业链上下游携手共进。若想实现突破,不仅仅是芯片厂商,还需要各行各业携手并进,才可“众人拾柴火焰高”。鼎桥物联网模组产品线总经理陶晗丰表示,整个物联网产业涉及千行百业,大家一起努力才能夯实整个数字经济社会的底座,因此走融合创新是必经之路。只有这样,才能够打造高质量、高稳定的产品和服务。

英伟达 GTC 2021 上展示多项新技术

本报讯 记者沈丛报道:英伟达GTC2021于11月9日在线上举行。在此次GTC大会上,英伟达首席执行官黄仁勋率先发表了主题演讲,并展示了基于英伟达Omniverse平台开发的多项最新技术,彰显了其AI技术的愈发强大。

在此次的GTC大会上,英伟达发布了NVIDIA Omniverse Avatar——一个用于生成交互式AI化身的技术平台。据介绍,Omniverse Avatar集合了该公司在语音AI、计算机视觉、自然语言理解、推荐引擎和模拟技术方面的技术。该平台为创建人工智能助手打开了大门,可以帮助处理数十亿的日常客户服务互动,从而带来更大的商业机会,并提高客户满意度。

据了解,Omniverse平台是一个可轻松加速设计工作流程、实现实时协作的开源云原生平台。专业人士可通过Omniverse创建并模拟一个虚拟3D世界。近期上映的中国探月官方纪录片《飞向月球》第二季中也采用了Omniverse平台,通过运用物理渲染、实时光线追踪、高精度3D扫描等多项黑科技,首次将AI超写实数字人应用在4K科学纪录片中。

随着NVIDIA Omniverse的更新发展,新的特征和功能将赋予更大范围的设计和仿真模拟工作流程。

会上,黄仁勋还宣布,NVIDIA Triton推理服务器迎来了重大升级。据了解,NVIDIA Triton是一款开源软件,对于所有推理模式都可以简化模型在任意框架中以及任何GPU或CPU上的运行方式,从而在生产环境中使用AI。Triton支持多模型ensemble,以及TensorFlow、PyTorch、ONNX等多种深度学习模型框架,可以

很好地支持多模型联合推理的场景,构建起视频、图片、语音、文本整个推理服务过程,大大降低多个模型服务的开发和维护成本。

黄仁勋介绍,升级后的Triton推理服务器将扩大对所有模型、各种框架、多查询类型的推理,面向所有平台、云、本地、边缘和嵌入式系统提供更好的机器学习和深度学习。此外,Nvidia Triton这一引擎可用于所有推理的工作负载,在图像处理、语音AI、自然语音处理、推理系统和强化学习等方面都将展现更出色的性能。未来,对于云服务商而言,Triton将提高其基础设施的利用率和吞吐量,释放其基础设施容量,以满足更高的技术需求。

此外,黄仁勋还宣布,英伟达推出最新Nvidia Clara Holoscan软件定义的可编程影像平台。黄仁勋介绍,Holoscan是英伟达多年的研究成果,打造Holoscan需要充分运用Nvidia的所有技术,其中包括统一的计算框架,以及全新的超高速传感器处理机器人芯片Orin。

黄仁勋表示,基础Holoscan平台由Orin和Connect-7组成,Orin可以在单个芯片中处理整个机器人流水线,即传感器、物理、AI、成像和图形。据了解,Orin SoC在单个芯片上可提供200 TOPs的性能,共有170亿个晶体管,其性能大约是NVIDIA Xavier SOC的7倍。它共有12个ARM Cortex-A78AE内核以及Ampere GPU架构。

黄仁勋介绍,Holoscan是开放的平台,第三方用户可以在Holoscan的界面和API的基础上进行构建。例如,仪器制造商可以将Holoscan集成到他们的解决方案中。此外,Holoscan应用程序还可以完整部署在设备、数据中心等领域。