

生成式 AI 落地企业有 4 个标准

■ 周雅

企业数字化转型过程中，随着转型阶段性的不同，也带来了IT管理的巨大挑战，其中的例子不胜枚举。

“全球有超过90%的企业以某种形式采用了云计算。但是当谈到云原生时，企业的采用速度却慢得多，云原生的增长率在过去几年里只有10%。众所周知，云提供了惊人的弹性、高可用性、可伸缩性，但是大家知道今天32%的云支出都浪费了吗？这主要是由于过度配置云实例、缺乏治理，以及企业在整个云环境中缺乏可观察性。再加上，网络犯罪活动在过去几年中呈指数级增长，仅在2024年CVE(Common Vulnerabilities&Exposures)的数量就高达3万多个……”

这是SUSECON中国2024数字创新峰会期间，SUSE亚太区CTO Vishal Ghariwala用一系列数字指出，无论是现代化应用的集成、分布式架构的管理、安全威胁的演进，以及生成式AI的普及，这些新的变革，对企业IT管理都是一次新的考验。

这些种种迹象表明，当前充满不确定性的市场环境下，企业比以往任何时候都更需要灵活且自主的解决方案。

以开源创新应万变

如果要用一个词来形容企业数字化转型过程中的核心诉求，那就是——韧性。

采访中，Vishal Ghariwala坦言，“据我观察，客户的主要诉求有三方面：第一，如何从数字化的角度打造一个具有

韧性的系统；第二，从网络安全的角度，如何确保也同样有韧性；第三，客户很多时候可能缺乏所需要的IT技能，特别是在网络安全、云原生等领域。”

全球著名汽车制造企业沈阳工厂在云原生领域的成功实践，也再次印证了Vishal Ghariwala的观点：

“每一家制造业都在上云，企业通过上云，可以很方便地获得高质量的基础设施。但是，制造业一旦投资下来，整个基建是非常恐怖的，因为所有的IT系统需要支持整个生产制造的产线系统，仅仅依靠一种管理方式，是不现实的。”

该企业容器平台架构师对“云+边缘”的应用颇有心得，那位架构师进一步表示，“如果在国外，它可以自建一些云的主干网到你的工厂；但在国内，必须在工厂附近建数据中心，因为需要及时处理一些对时延敏感性任务，一些敏感的数据不能上云，所以，弹性的方案对制造业来说非常重要，你既不能只用云，也不能只用边缘，需要‘云+边缘’结合一起使用。”

针对企业的种种挑战，Vishal Ghariwala分享了SUSE四大技术——云原生、边缘计算、企业级Linux、SUSE AI，帮助企业打造一个富有韧性的系统。

第一，在云原生方面，SUSE的云原生虚拟化平台——Harvester产品系列消除了云原生的复杂性，使企业能够在所有设备和环境中标准化云原生操作。比如，最新发布Harvester 1.31，为企业提供了从传统的“基于VMware和其他虚拟机工作负载”向“现代化云原生解决方案”的过渡。

此外，SUSE收购了被誉为“容器

界鹰眼系统”的StackState，将它与SUSE云原生管理平台Rancher Prime整合，以增强云原生可观测性。SUSE通过Rancher Prime实现了虚拟机与容器的统一管理，帮助企业加速关键应用的现代化，并优化核心、云端及边缘的应用交付效率，同时还可快速将生成式AI的能力带入到Rancher Prime中，如提供AI助手等功能。

第二，在边缘计算方面，SUSE Edge平台简化了分布式架构的管理和扩展，确保一致性操作，并为企业提供全面的安全防护，保障边缘应用的高效运行。

其中，SUSE Edge 3.0经过优化，带来了Rancher Prime Kubernetes管理、NeuVector Prime云原生容器安全，以及云原生堆栈的全面功能，可提供边缘设备的完整生命周期管理，以及大规模部署能力，也可在资源受限、互联网连接不稳定的位置运行，非常适合嵌入式设备。

第三，在Linux方面，SUSE Liberty Linux是一个支持各种Linux发行版的企业级解决方案，能实现对各种操作系统的一键纳管、零迁移和持续维护，无缝解决CentOS停更带来的问题。

比如，只需简单配置，不用做任何迁移，你就可以获得CentOS不间断的支持并持续获取安全补丁，并保证业务安全标准；还有，一个SUSE Manager界面就可管理全部跨平台系统，一次性解决分散、复杂的维护难题；此外，还可获得SUSE专业技术支持服务，解决规划、部署、排查等一系列问题，并可为未来系统扩容奠定基础。

第四，在生成式AI方面，为了帮助企业用户构建私有化AI，SUSE发布了SUSE AI解决方案，为各类规模的企业提供一个模块化的、安全的、非供应商绑定、与LLM（大型语言模型）无关的生成式AI平台。

该平台基于SUSE行业领先的开源企业级SUSE Linux、Rancher Prime Kubernetes管理平台和Rancher NeuVector安全产品，使企业可以在内部或云端运行私有的GenAI解决方案，并选择他们喜欢的LLM，确保数据隐私、合规性和成本控制。

SUSE的开源解决方案有很强的灵活性，且可以支持不同供应商的系统，这也意味着，如果其中一个系统出现问题，我们可以很容易迁移到其他技术，这样的灵活性，就能够帮助企业打造非常具有韧性的系统；此外，SUSE给了用户和企业更多的选择，企业的系统也能因此具有更强的韧性。”Vishal Ghariwala补充说道。

以德意志银行为例。作为世界领先的金融企业，德意志银行的业务网络遍及全球。德意志银行选择了SUSE Liberty Linux来维护和支持其全球Linux资产，这包括数千台SUSE Linux Enterprise和Red Hat Enterprise Linux服务器。通过选择SUSE Liberty Linux解决方案，德意志银行可以统一管理全球业务网络的操作系统，降低维护、更新难度，实现标准化管理。更重要的是，SUSE Liberty Linux可以无缝接管已经停更的CentOS系统，用户无需做任何迁移，就能使CentOS持续获得安全补丁，满足金融业的安全审计要求。

用SUSE大中华区总裁陈毅威的话来说：“SUSE通过开源创新，赋予企业更多选择，避免供应商锁定。”

边缘+AI，下一个增长点

随着物联网和数字化转型的加速，云原生与边缘计算的新时代已经到来。在集群管理需求下，边缘部署变得越发重要。

Vishal Ghariwala在采访中透露，他看到中国市场有两个商业机会：一个是边缘计算，中国的边缘计算服务器市场到2028年将达到132亿美元（根据IDC数据）；另一个是网络安全，尤其是AI驱动的网络安全创新，存在大量机遇。

但是机遇与挑战总是并存。Vishal Ghariwala指出，“有很多需要边缘计算

的用例，这些用例其实是有非常独特的要求，它和传统的基于云以及基于数据中心的解决方案相比是不一样的。”

具体而言，边缘解决方案非常依赖三要素：极低的时延、极好的性能，以及尽可能在减少使用算力的基础上运行。在这个过程中，也存在三个挑战：第一，要管理各种不同的海量设备，并覆盖设备的全生命周期，且多数时候需要远程管理；第二，由于边缘端缺乏足够和强大的算力，所以要有轻量化的基础架构；第三是安全性，要确保它不会受到网络攻击。

“SUSE坚信，把在这些远程位置的边缘解决方案和AI结合在一起，将会是下一个信息技术领域的增长点。”Vishal Ghariwala强调，SUSE Edge平台可以帮助企业打造这样一套系统，不管是普通的工作负载还是AI的工作负载，都可以在边缘端能够很好的、安全的运行。

如果生成式AI未来得到普及，那么无论它输出什么结果，有四个标准是不变离其宗的：

第一，它输出的内容应该是公平的，且没有任何的偏见和歧视；第二，结果要有透明度和来源；第三，系统本身要能够对数据进行保护，来确保数据不会被他人或软件滥用，从而扭曲了最终的输出结果；第四，输出的结果要有包容性，要考虑到更广泛人群。

“如果想要真正达到生成式AI这四个关键标准，唯一的模式，就是通过开源的方式。”Vishal Ghariwala表示。

而SUSE AI解决方案的理念是，它不仅提供了各种不同的开源项目或者技术（包括Open WebUI、Ollama等），更是打造了一个开放的环境，让客户有更多的选择，放在数据中心、云、边缘都可以，以客户自己想要部署的方式去进行。”

作为来自德国的全球开源解决方案领导者，2024年是SUSE扎根中国的20周年，20年来，SUSE始终致力于为中国本土企业提供从基础架构、云原生到边缘端的全面解决方案，帮助企业应对数字化转型中的复杂挑战。

正如SUSE CEO Dirk-Peter van Leeuwen在活动开幕致辞中提到，“我们致力于为企业带来开源的无限潜力，为您提供技术选择的自由。”■

（文章来源：科技行者techwalker.com）

AI 机器人画作 拍出逾百万美元

■ 荆晶

苏富比拍卖行8日以高达108.48万美元（约合779万元人民币）拍出人工智能（AI）机器人“艾达”的画作《AI之神》。这一价格远超先前的预估价格。

据美国有线电视新闻网报道，《AI之神》为数学家艾伦·图灵的肖像，长162.5厘米，高230厘米。图灵是英国数学家和二战英雄，主要从事数理逻辑、纯数学、计算机科学、理论生物学方面的研究，被誉为“计算机科学和人工智能之父”。

这是全球首次有机器人“艺术家”获得大型拍卖行青睐，因此吸引多方关注。《AI之神》开拍后获得27次竞价，最后被一名没有公开姓名的买家拍得。

“艾达”是一个人形机器人，外表是女性形象，短发、大眼睛，眼睛里装有摄像头。“她”由英国艺术家艾丹·梅勒与英国一家机器人公司合作制造，2019年问世，可以使用大语言模型与外界交流，曾经创作多幅画作，赢得赞誉。

梅勒在一份声明中说，每次创作之

前，“艾达”都会与其研发团队“讨论”想画的内容。“创作这幅画作前，我们与‘她’讨论了‘AI向善’的问题，这导致‘艾达’提出，‘她’想画图灵这个人工智能发展史上的关键人物。”

在回答了关于画作风格、内容、色调和质地的问题后，“艾达”用眼部摄像头“观察”图灵的照片并绘制草图，然后根据特定算法对照片产生的不同解读结果创造了15幅不同的图灵肖像，每幅的创作时间都在6至8小时左右。最后，“她”选择了三幅画作组合在一起，并在背景中加入图灵发明的密码破译设备“炸弹机”。

由于“艾达”的机械手臂只能在约30厘米长、42厘米宽的小幅画布上作画，最终成品由3D打印设备打印到更大的画布上。依照苏富比拍卖行说法，打印过程没有改变图像。

梅勒接受美国有线电视新闻网采访时说，在图灵预言计算机和人工智能崛起的大约80年后，他希望“艾达”及其艺术作品能成为“一面镜子，照出人类前进的方向”。■

250亿国产GPU独角兽或筹备上市 字节、腾讯位列股东

■ 申谨睿

半导体涨势还在延续，国产GPU概念股集体高开。

11月11日，东华软件、和而泰、润新科技、威星智能涨停，弘信电子、青云科技涨超10%，卓易信息、麦捷科技跟涨。

消息面上，摩尔线程即将启动IPO辅导。国家企业信用公示系统显示，摩尔线程智能科技（北京）股份有限公司（下称“摩尔线程”）的市场主体类型，在10月28日由其他有限责任公司变更为中国其他股份有限公司（非上市）。同时，公司注册资本由2441.32万元增至33亿元。

另据媒体报道，有接近摩尔线程的人士透露，摩尔线程已完成股份制改造，正在准备上市，预计很快会有新进展，科创板或为其首要目标。时代周报记者就启动IPO等相关问题向摩尔线程致电并发送邮件核实，截至发稿未获回复。

今年启动IPO的国产半导体厂商还有燧原科技和壁仞科技。“二级市场对GPU厂商提供的支持是值得重视的。不过，企业不能过度依赖资本市场。”艾媒咨询CEO兼首席分析师张毅告诉时代周报记者。

国产GPU集体奔赴A股

摩尔线程成立于2020年10月，是一家以全功能GPU（图形处理器芯片）设计为主的集成电路高科技公司。依托MUSA（通用并行计算）架构，摩尔线程建立了从芯片、板卡、集群到软件的全栈AI智算产品线。

摩尔线程已推出MTT S30、MTT S70、MTT S80等多款显卡，其中，瞄准游戏行业推出的高性能游戏显卡

MTTS80，承诺可驱动国产3A大作《黑神话：悟空》的运行。

2023年12月，摩尔线程发布首个全国产千卡千亿模型训练平台——KUAE夸娥智算中心揭幕，宣告国内首个以全功能GPU为底座的大规模算力集群正式落地。2024年，摩尔线程的智算集群解决方案已经从千卡级别大幅扩展至万卡级别。

摩尔线程目前的实控人为原英伟达中国区总经理、英伟达全球副总裁张建中。前五大股东分别为南京神傲管理咨询合伙企业（有限合伙）、张建中、南京华傲管理咨询合伙企业（有限合伙）、深圳市明皓新科技合伙企业（有限合伙）、盛芯启程私募投资基金（上海）合伙企业（有限合伙），分别持股17.63%、13.41%、8.16%、6.86%、5.94%。

天眼查显示，摩尔线程已完成六轮融资，累计融资金额数十亿元，投资方包括深创投、红杉中国、字节跳动、腾讯投资、中关村科学城等。在2022年12月完成15亿元B轮融资时，摩尔线程的估值已达240亿元。2024年4月9日，摩尔线程以255亿元的估值位列《2024·胡润全球独角兽榜》第261名。

今年启动IPO进程的国产GPU公司不止摩尔线程。据中国证监会官网信息，今年8月，燧原科技的上市辅导备案已获受理，辅导机构为中金公司，该公司计划在A股进行IPO；9月，壁仞科技在上海证监局办理辅导备案登记，正式开启IPO之路，辅导券商为国泰君安。

张毅告诉时代周报记者，在消费电子新周期和AI浪潮的双重推动下，市场对GPU的需求呈倍数级增加。在此节点上市，有助于企业拿到理想的估值。张毅表示，虽然上市可以一定程度上缓解企业压力，但企业也不可过度依赖资本，忽略把握市场空间、合理人才调配、产品与营销齐头并进的重要性。■

（文章来源：时代周报）