算力租赁价格暴涨 玩家涌入赛道拥挤

本报记者 秦枭 北京报道

随着互联网企业以及科技企业纷纷在大模型参数上不断刷新榜单，以及传统企业对于数智化转型需求的增加，导致算力需求大幅上涨，但与此同时，由于高端芯片获取难度大、公共算力建设不足，算力租赁进入新一轮涨价周期。其中，并济科技、中贝通信（603220.SH）、汇纳科技（300609.S2）等公司宣布算力服务收费大幅上涨。

《中国经营报》记者在采访中了解到，从GPU性能、组网能力、软件生态等维度来看，国内的算力建设并没有适配目前算力需求的增长速度，算力供需紧张，算力租赁价格也随之水涨船高。为了解决算力供不应求的问题，国内多地政府出台了政策增加算力供给、或降低算力使用成本。

不过，国盛证券指出，在AI应用加速迭代等各方面因素作用下，此次算力涨价还会持续扩大。

在此背景下，算力供给缺口也被投机者盯上。一位智算中心的负责人向记者表示：“现在很多大张旗鼓准备做算力租赁的企业只是说说而已，想要赚快钱。算力租赁更像是公共服务，其实并不是一门好生意。”

国内的算力建设并没有适配目前算力需求的增长速度。 视觉中国/图

涨价 此次算力集体提价并非短期波动，而是在开启中长期的涨价潮。

“算力租赁”里的“算力”指的是对数据的计算能力，“租赁”就是将算力、存储、网络等资源统一封装，以服务的形式（如API）进行算力交付。区别于算力基础设施，算力租赁可以满足多种类型企业的需求，用户根据自己的需求选择适合自己的服务器或虚拟机来完成大规模的计算任务，无需自己再去花大量时间、成本和精力研发计算模型。

王杰（化名）是一家聚焦智慧交通的企业负责人，平时主要负责视频数据的算法开发，服务于智能交通、城市治理等领域。他对记者表示：“现在很多企业标榜自己是人工智能公司，但本身是没有计算研究能力的，只能租用其他公司的大模型进行计算。还有一种情况是，人工智能公司不只从事一种业务，业务研究过程中需要用到多种计算模型，但这需要消耗的资金等方面的成本也随之变大，对于中小企业来说是无法承受的。我们是以嵌入式硬件为载体的人工智能算法，与P（每秒钟可以进行算力运算的次数，1P等于1024T）的量级相比，仅需要4T和22T的算力，就没有必要再自己建一个模型，租用算力可以大幅降低成本和提升效率。”

“市场上常见的租金计量方式包括，按整台服务器租赁（每台服务器含8张GPU），租金按照每台每月计量；按算力规模租赁，租金按每P每年计量；按单张GPU租赁，租金按照每GPU每小时计量。在5月份的时候大概是6万元/P/年。”王杰向记者介绍。

但进入11月份，算力租赁市场风云突变，价格一路飙升。近日，中贝通信公告显示，公司向北京中新科远提供AI算力技术服务，合同总金额为3.46亿元，单价为18万元/P/年。但根据该公司此前算力服务框架协议显示，单价为12万元/P/年。

11月14日，并济科技公众号发布通知，由于高性能运算设备持续涨价，A100算力资源持续紧张，即日起A100算力服务收费拟上调100%。

汇纳科技在公告中称，由于内嵌英伟达A100芯片的高性能算力服务器算力需求大幅增加，相关高性能运算设备持续涨价，算力资源持续紧张，公司拟将所受托运营的内嵌英伟达A100芯片的高性能算力服务器算力服务收费上调100%。

此外，润建股份（002929.SZ）、青云科技（688316.SH）等算力服务商也都在公共平台表示过近期涨价的意愿。

值得注意的是，近日有消息曝出，某互联网巨头已暂停A100服务器出租业务。虽然该公司没有正面回应，但也同样暴露了国内算力租赁市场供需失衡并不是空穴来风。

国盛证券也在研报中指出，算力涨价潮即将开启，火山引擎、腾讯云、微软Azure的A100租赁零售价相较此前的市场平均价均有不同程度上涨。在AI应用加速迭代等各方面因素作用下，此次算力集体提价并非短期波动，而是在开启中长期的涨价潮。

算力券、“打折”并行 各省份也在降低算力使用成本和门槛，算力券成为一种“流行”模式。

为了应对算力资源紧俏，各省份相继出台算力建设目标，以缓解算力供需紧张。据不完全统计，上海、山东、宁夏、天津等地相继出台存力相关指导文件。具体来看，天津发布《关于做好算力网络建设发展工作的指导意见》，山东发布《山东省一体化算力网络建设行动方案（2022—2025）》，宁夏发布《宁夏回族自治区数据中心建设指南》等。

除此之外，各省份也在降低算力使用成本和门槛，算力券成为一种“流行”模式。

具体来看，今年1月12日，成都印发《成都市围绕超算智算加快算力产业发展的政策措施》，提出每年将发放总额不超过1000万元的“算力券”；7月31日，杭州印发《杭州市人民政府办公厅关于加快推进人工智能产业创新发展的实施意见》，提出全市每年设立总额不超过5000万元的“算力券”，重点支持中小企业购买算力服务；7月11日，北京表示将推出首期不低于4000万元的算力券补贴，支持搭模型企业降低训练成本；贵州省则提出每年发放总额不超过8000万元的算力券，支持省内外企业、高校、科研机构等购买贵州算力服务时抵扣一定比例服务费用。

与此同时，也有地区对算力租赁进行了“打折”。上海对租用纳入本市统筹调度的算力进行大模型研发的本市主体，经评估按算力集群规模和成果水平给予最高10%的租用补贴。福建对符合条件的人工智能企业购买算力服务年度总额达到30万元（含）以上的，按照当年实际服务费第一年50%、第二年40%、第三年30%，每年给予最高50万元补助。

一门不错的生意？

看似门槛不高且一片蓝海的算力租赁，吸引了众多玩家涌入。

价格快速上涨，政策的扶持力度提升，让算力租赁看起来是一门不错的生意。

10月18日，恒润股份（603985.SH）公告显示，上海润六尺向供应商A采购75台H800NVLink算力服务器及配套设备（含设备安装调试服务费），合同金额为1.71亿元；采购22台A800NVLinK算力服务器及配套设备（含设备安装调试服务费），合同金额为3080万元。此前，莲花健康（600186.SH）也公告，以6.93亿元的总价采购330台英伟达H800 GPU系列算力服务器，正式宣布跨界算力领域。

一般情况下，每台服务器有8张GPU。记者以此估算，A800单价约为17万元，H800单价约为26万元左右。每片A800算力为0.6P，每片H800算力为2P。2023年10月1P算力价格约为18万元/年，如果统计算力卡折旧、租用机柜费用和人员费用等，基于2023年10月的数据估算，租赁H800毛利率约为30%左右。而此前，毛利率更高。

而且，据相关公司预计到2030年，全球通用计算（FP32）总量将达3.3ZFLOPS，相较2020年增长10倍，AI计算（FP16）总量将达105ZFLOPS，相较2020年增长500倍。

“AI算力有望超越普通算力服务市场，2023年格局分散。AI算力租赁市场规模有望超过普通算力服务市场，规模达到数千亿元。”东吴证券指出。

看似门槛不高且一片蓝海的算力租赁，吸引了众多玩家涌入，数据显示，近十年，我国算力基础设施相关企业呈逐年高速增长态势。2020年、2021年、2022年分别新增47.85万家、76.97万家、80.14万家，同比增长52.62%、60.85%、4.11%。天眼查数据显示，截至目前，今年我国算力基础设施相关企业注册量达86.66万家，已超去年全年注册量，其中今年前10月新增80.27万家，同比增长20.48%。

不仅如此，一些其他行业的上市公司也纷纷宣布跨界，试图在算力租赁行业“分一杯羹”。

但实际上，从事该等业务所需的行政审批和准入资质，包括IDC经营许可证，智算机房大机电和主要产品的质量认证也都需要得到验证。

“如果没有早期的积累，以及政府的扶持，入局算力租赁并不赚钱。”上述智算中心的负责人对记者坦言，“从单纯的成本和营收来看，纸面上确实好看，如果加上后期的运维等事宜，与同等规格的GPU服务器硬件售价大致相当，所以算力租赁并不是一门好生意。因为算力服务更像是公共服务，并不是简单的算力租赁，很多客户都是初创企业，用户更需要的是可行的解决方案。”

而且，上述负责人说道：“目前阶段，大张旗鼓准备做算力租赁的企业，有很大部分只是说说而已。因为各方面的原因，现在GPU太难买到了，而且即便有了卡，你是否有能力组装服务器，服务器是否适配，都是问题，没有专业的团队和雄厚的资金支持，是没有办法在短时间内形成具体业务形态的。”