

计算机行业深度报告

算力服务：AI 产业晴雨表

方正证券研究所证券研究报告

分析师

张初晨

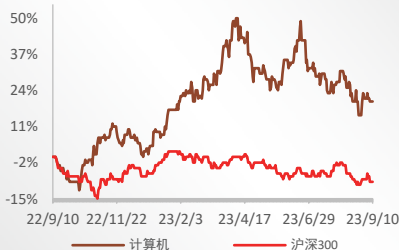
登记编号：S1220523070001

行业评级：推荐

行业信息

上市公司总家数	382
总股本(亿股)	3,445.03
销售收入(亿元)	10,457.87
利润总额(亿元)	-110.15
行业平均 PE	147.32
平均股价(元)	26.04

行业相对指数表现



数据来源：wind 方正证券研究所

相关研究

《算力服务：从“东数西算”到“东推西训”》2023.08.28

《具身智能浪潮将至，重视国内产业链机遇》2023.08.24

《数据库：行业景气百舸争流千帆竞，细分赛道长坡厚雪奋者先》2023.08.03

《AI 赋能之下，电力 IT 投资机会在何处？》2023.07.24

算力服务释放中小实体算力使用需求，国内云计算市场增速超全球平均。中国民营企业上云渗透率为 49.2%，国资企业上云渗透率为 33.7%，相较美国 85%以及欧洲 70%的企业上云率，有较大的发展潜力。云计算很大程度降低算力使用的门槛，释放中小企业数智化转型带来的旺盛算力需求。2022 年中国云计算市场规模约 4550 亿元，同增 40.91%，高于全球 19%的平均增速，预计 2025 年中国云计算市场规模将超万亿元。

政策推动云计算普惠触达，资源、渠道、网络是云计算厂商的重要壁垒。国家算力指数与 GDP 走势呈现显著正相关，算力指数平均每提高 1 点，国家的数字经济和 GDP 将分别增长 3.5%和 1.8%。系列政策大力推动云计算发展，多地发放算力券推动算力普惠触达。当前，运营商及大型科技企业为公有云的主要参与者，公有云 2022 年 CR5 为 71%，包括阿里云、天翼云、移动云、华为云、腾讯云。后续看，云计算厂商的资源丰富度，上下游渠道能力，以及算力网络能力将继续成为其重要壁垒。

模型训练推升短期算力景气度，中期看推理算力需求释放节奏。2023 年初随着海外大模型的催化，国内外人工智能模型数量和参数量快速增长，迭代速度加快，模型训练侧算力需求快速攀升。算力硬件处于供不应求的状态，出现明显的价格上涨和交货周期延长的问题。考虑到人工智能模型仍在快速迭代，至 2023 年底的短期算力需求预计仍将保持景气。中长期看，大模型走向成熟，数量也将收敛，模型推理将贡献更大的算力需求。推理算力需求释放的节奏需要多维度紧密跟踪。

算力服务报价在及时性和客观性上具有优势，可作为判断算力供需节奏的显著指标。订单情况反映的是乐观情况下的需求预期，在关键硬件供给紧张的趋势下，会出现下游需求方报高意向订单量，以保证拿货的优先等级，后续如果下游需求放缓，可能会发生砍单的情况。这种现象在 2019 年-2020 年的比特币等虚拟货币的挖矿热，以及 2020 年-2021 年的智能电动车缺芯潮中，都有明显的体现。随着算力交易商业模式的成熟，算力服务报价将更加具备及时性和客观性上的优势。通过整理跟踪算力服务价格，我们认为横向对比各大云厂商云计算服务的价格孰高孰低的参考性不高，但综合各云厂商的算力服务价格环比变化，预计对判断算力供需有较大参考意义。

风险提示：人工智能技术发展不及预期，企业数字化转型速度不及预期，算力服务行业竞争加剧风险，政策支持力度不及预期，云计算厂商市场拓展不及预期等。

并购家 免费行业报告网

IPOIPO.CN 每日更新 不用注册会员 无广告 永久免费

正文目录

1 算力租赁释放中小实体算力使用需求，政策推动普惠触达	4
1.1 需求侧：算力使用需求更加广泛，中国云计算服务市场快速增长	4
1.2 政策推动算力普惠，促进更广算力触达	6
1.3 供给侧：算力服务提升算力使用效率，资源、渠道、网络是重要能力	7
2 算力服务报价可作为跟踪行业算力供需情况的重要指标	9
2.1 模型训练扩张推升短期算力景气度，中期看推理算力需求释放节奏	9
2.2 当前 AI 算力核心上游硬件处于供不应求的状态，下游订单量较为乐观	12
2.3 算力订单数据存在滞后性及真实性担忧，需要更多维验证	13
2.4 算力服务价格可作为判断行业算力需求节奏的显著指标	15

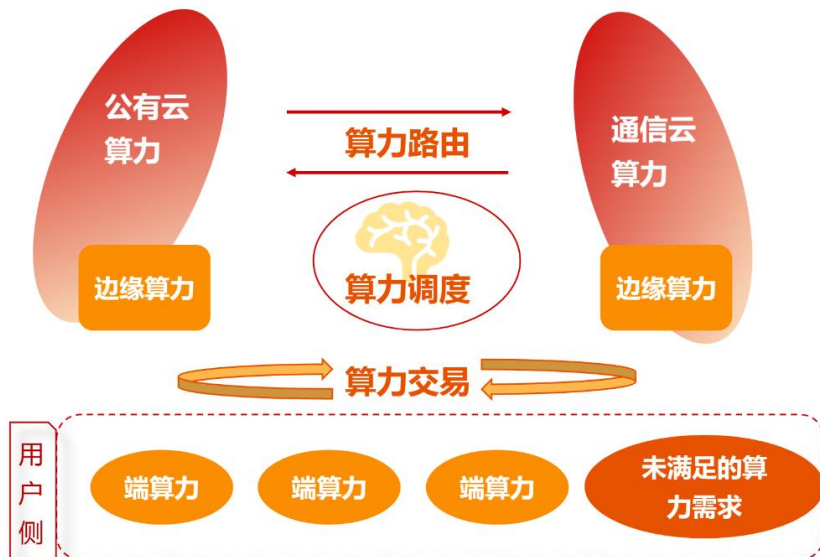
图表目录

图表 1: 算力调度及交易释放更多算力使用需求.....	4
图表 2: 云原生所需能力与特征	4
图表 3: 企业希望导入云端的技术及服务.....	5
图表 4: 企业数智化转型将推升对云计算的需求.....	5
图表 5: 全球云计算市场规模（亿美元）及增速.....	6
图表 6: 中国云计算市场规模（亿元）及增速.....	6
图表 7: 算力指数每提高一个点带动 GDP 增长 1.8%.....	6
图表 8: 中央政策支持云计算建设及发展.....	6
图表 9: 各地方政府纷纷发放算力券支持算力普惠.....	7
图表 10: 云计算供应商选择的主要考量因素.....	8
图表 11: 中国公有云 IaaS 厂商市场份额.....	8
图表 12: 中国公有云 IaaS 市场集中度变化.....	8
图表 13: 三大运营商云计算收入（亿元）.....	9
图表 14: 三大运营商云计算收入增速（%）.....	9
图表 15: 中国计算机网络设备上市企业竞争格局.....	9
图表 16: 华为具有较强的算网建设能力.....	9
图表 17: 讯飞星火大模型给出了明确快速的迭代时间表.....	10
图表 18: 模型参数量及训练所需算力快速增长.....	10
图表 19: 大模型数量仍在较快增长.....	10
图表 20: 国产大模型保持较高的推出及迭代节奏.....	11
图表 21: 大模型将逐步向头部收敛.....	11
图表 22: 人脸识别任务涉及到多个模型的推理.....	12
图表 23: 智算芯片订单及价格相关新闻枚举.....	13
图表 24: 2020.10 到 2021.10 全球芯片平均交付周期（周）.....	14
图表 25: 2021 年缺芯压力导致车厂“减配交付”.....	14
图表 26: 2022 年汽车芯片紧缺程度缓解后出现较明显砍单情况.....	14
图表 27: 算力交易实时反应供需情况.....	15
图表 28: 上海算力交易平台（试运行）上线.....	16
图表 29: 上海算力交易平台对供需撮合有较好的探索.....	16
图表 30: 云厂商提供的服务器类型.....	16
图表 31: 国内头部云计算厂商均可提供一定的全球算力服务能力.....	17
图表 32: 国内头部云计算厂商高性能云计算价格.....	18
图表 33: AZURE 及 AWS 高性能计算云服务价格.....	18

1 算力租赁释放中小实体算力使用需求，政策推动普惠触达

人工智能快速发展，企业数智化转型加快，适用于更多、更广范围用户需要的算力产品和服务也将加快落地。算力租赁的进一步发展对算力的需求方和供给方都会产生明显的价值，降低算力的使用门槛，进一步抬升算力需求的天花板。

图表1:算力调度及交易释放更多算力使用需求

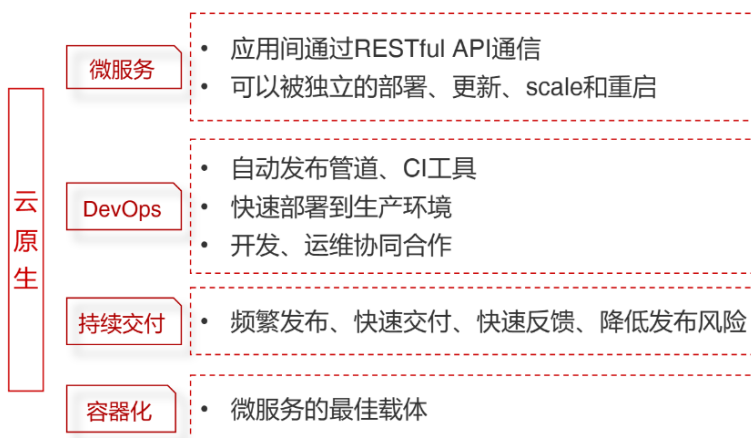


资料来源：华为，方正证券研究所

1.1 需求侧：算力使用需求更加广泛，中国云计算服务市场快速增长

云计算促进软件开发部署模式的创新，成为承载各类应用的关键基础设施。云计算通过集约化的模式为应用提供大规模的算力资源。云原生通过容器、微服务、DevOps、服务网格、Serverless 等核心技术解决了应用的标准化封装和交付、应用间的标准化交互问题，统一了云上应用的底层语言，充分剥离复杂底层环境对上层应用的影响。同时，云计算放大生态合作价值，推动应用的不断丰富和迭代。

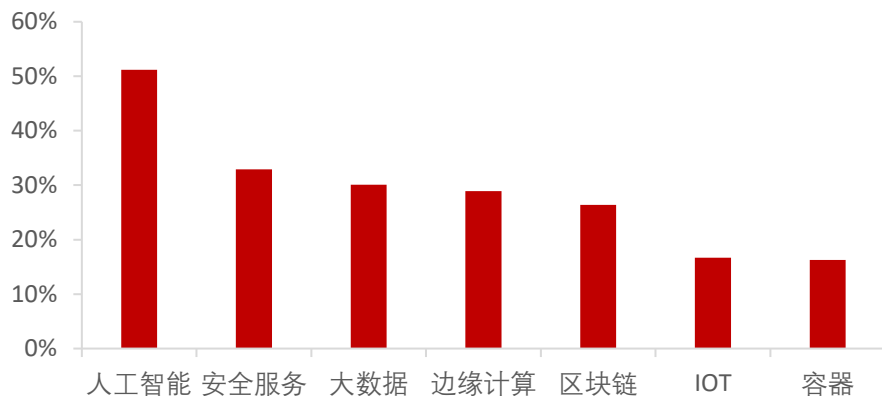
图表2:云原生所需能力与特征



资料来源：腾讯云开发者社区，方正证券研究所

算力服务为大数据、人工智能、物联网等新技术的发展提供基础支撑。人工智能、大数据等技术加速发展，更多企业选择加速融入“上云用数赋智”，将新技术应用到生产、经营、决策中，企业对算力的需求也快速上升。根据亿欧智库的问卷，有超过一半受访企业希望将人工智能技术导入云端，安全服务和大数据紧随其后，分别占比 32.9% 和 30.1%。

图表3:企业希望导入云端的技术及服务



资料来源：亿欧智库，方正证券研究所

企业数智化发展使得算力需求群体由小众扩展为千行百业。当前国内企业上云率仍相较发达国家有明显的差距，根据麦肯锡等研究机构，当前美国企业上云率达 85% 以上，欧盟企业上云率约 70%，据亿欧智库的报告，2021 年中国民营企业上云渗透率为 49.2%，国资企业上云渗透率为 33.7%，有较大的发展潜力。

数智化转型不仅针对大型企业，中小型公司在降本增效、提升运营管理水平、开发新的业务增长方面亦有旺盛的需求。同时，随着业务智能化水平的提高，适用于中小厂商业务的智能化模型的部署需要相应算力的支撑。

图表4:企业数智化转型将推升对云计算的需求

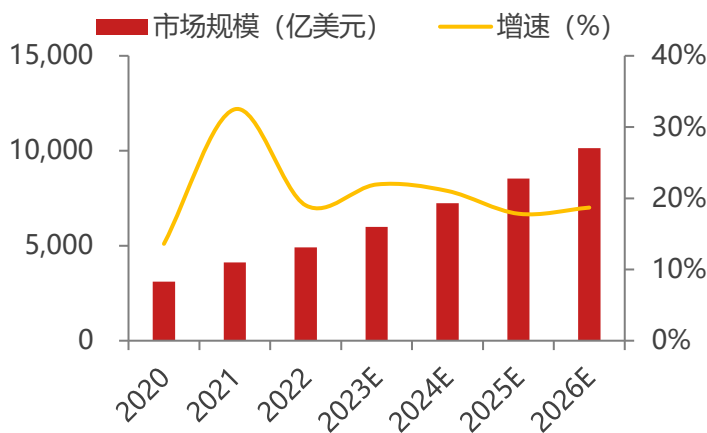


资料来源：亿欧智库，方正证券研究所

全球云计算市场稳定增长，中国保持快速发展。2022 年全球云计算市场规模达 4910 亿美元，增速 19%，未来预计仍将保持稳定增长，至 2026 年预计突破万亿

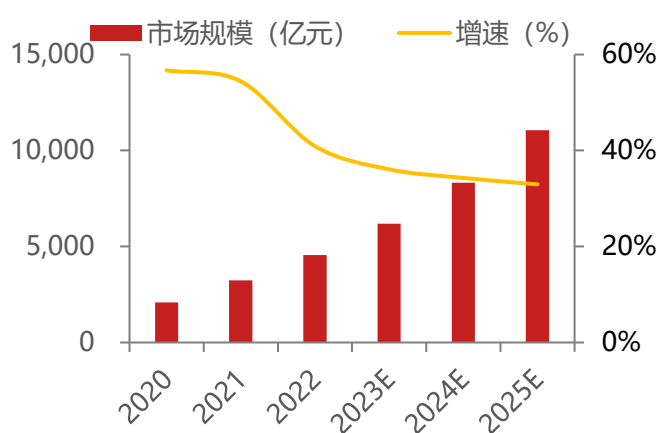
美元。2022 年中国云计算市场规模约 4550 亿元，同增 40.91%，高于全球平均增速，预计 2025 年中国云计算市场规模将超万亿元。

图表5: 全球云计算市场规模 (亿美元) 及增速



资料来源: 中国信通院, 方正证券研究所

图表6: 中国云计算市场规模 (亿元) 及增速



资料来源: 中国信通院, 方正证券研究所

1.2 政策推动算力普惠, 促进更广算力触达

算力是推动经济更快、更高质量发展的基石。根据 IDC 及浪潮信息联合发布的《2021-2022 全球算力指数评估报告》，随着全球数字经济持续稳定增长，数字经济占比预计到 2025 年有望达到 41.5%。国家算力指数与 GDP 走势呈现显著正相关，算力指数平均每提高 1 点，国家的数字经济和 GDP 将分别增长 3.5%和 1.8%。当一个国家的算力指数达到 40 分以上时，国家的算力指数每提升 1 点，其对于 GDP 增长的推动力将增加到 1.5 倍。

图表7: 算力指数每提高一个点带动 GDP 增长 1.8%



资料来源: IDC, 华为官网, 方正证券研究所

图表8: 中央政策支持云计算建设及发展

时间	发布机构	文件及内容
2022 年 12 月	国务院	《扩大内需战略规划纲要 (2022-2035 年) 》，提出加快建设信息基础设施，推动云计算广泛、深度应用，促进“云、网、端”资源要素相互融合、智能配置。
2023 年 1 月	工信部等六部门	《关于推动能源电子产业发展的指导意见》，提出加快云计算计算推广应用。
2023 年 4 月	工信部等八部门	《关于推进 IPv6 技术演进和应用创新发展的实施意见》，鼓励推动 IPv6 与云计算等技术的创新融合，促进云计算和网络协同发展。

资料来源: 中国政府网, 方正证券研究所

算力普惠方便触及是政策所向，多地算力券政策相继推出。推动算力建设的同时，使得算力更多服务经济发展，使其成为容易获得的社会资源是各级政府思考发力的重点。年初以来，在中央推动数字经济发展的定调下，成都、上海、北京等各地方政府发放算力券可以视为明显的政策支持风向。最终实现如中国移动负责人在 2021 年世界互联网大会上所表述的，实现“网络无所不达、算力无处不在、智能无所不及”的愿景。

图表9: 各地方政府纷纷发放算力券支持算力普惠

城市	时点	文件&政策
成都	2023/1/12	《成都市围绕超算智算加快算力产业发展的政策措施》，每年 1000 万元的算力券，支持算力中介、科技型中小微企业和创客、科研机构、高校等使用成都智算中心算力资源，按算力合同费用的 50%支持
上海	2023/5/17	助力中小企业数字化转型，发放“AI 算力券”，重点支持租用本市智能算力且用于核心算法创新、模型研发的企业，最高按合同费用 20%进行支持
北京	2023/7/14	正在筹划通过算力券等形式支持模型伙伴和模型观察员，首期预计支持不低于 4000 万元算力券，补贴到模型伙伴企业。
杭州	2023/7/13	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。鼓励头部企业、高校院所开展多模态通用大模型研发并向中小企业开放模型应用
河南	2023/8/5	《河南省支持重大新型基础设施建设若干政策》，每年发放总规模不超过 5000 万元的“算力券”，支持企业、科研机构、高校等使用国家超算郑州中心，按照算力资源使用费用的 20%予以奖励

资料来源：中国政府网，新华网，中国证券报，河南政府官网，方正证券研究所

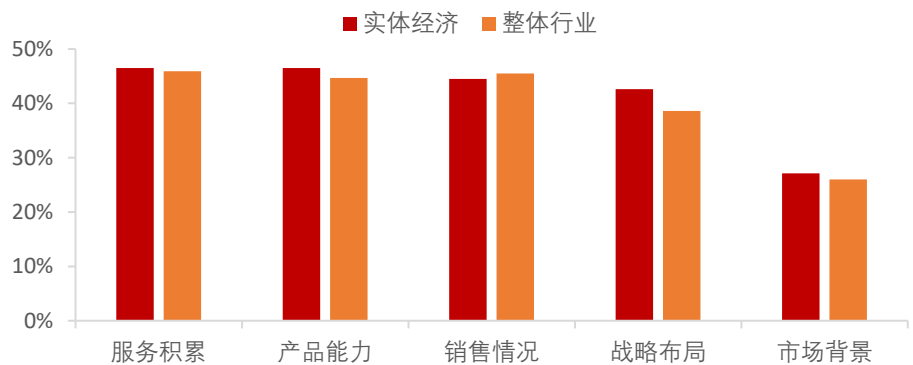
1.3 供给侧：算力服务提升算力使用效率，资源、渠道、网络是重要能力

存量角度：算力使用存在峰谷，算力服务提升算力使用效率。同时，当前已建成的算力基础设施使用率仍较低。根据工信部《全国数据中心应用发展指引》，数据中心利用率约为 50-60%，且随数据中心规模增大而递减。各行各业对算力的使用需求并不是所有时间恒定不变的。以一天的维度看，推理的算力需求较大集中于白天的工作时间，或者晚间休闲娱乐时间，在休息时间需求较少。从中期维度看，一些企业在项目前期搭载的算力资源在项目后期，或者随着企业战略转型可能面临闲置。**将闲置资源变现将为相关企业带来额外收益，减少算力闲置浪费的问题。**

云计算技术推陈出新，满足用户多样性需求。随着云计算用户数智能化升级，计算服务用户更加关注效率、性能、安全等因素，应用现代化、一云多芯、平台工程、云成本优化、系统稳定性、云原生安全等新技术层出不穷。

效率和性能成为云服务商竞争的主要方向。云服务厂商当前主要发力三个方向：**1）注重敏捷迭代，提升效率。**有利于算力用户以较低的学习成本和开发投入，实现快速创新。例如微软的 Power Platform 已经与 Office 365、Dynamics 365 以及 Azure 三大生态打通，形成技术闭环。**2）软硬协同，优化性能。**通过软件平台建设，提供多样化的算力和完善的开发者生态。由 IaaS 向 PaaS 以及 SaaS 延伸。**3）跨域融合，创新发展。**大型云计算厂商推动计算基础设施与大模型、大数据技术的融合，践行“云智一体”发展线路。

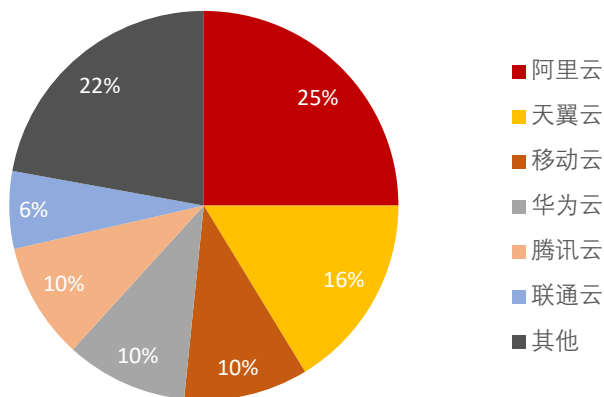
图表10: 云计算供应商选择的主要考量因素



资料来源: 亿欧智库, 方正证券研究所

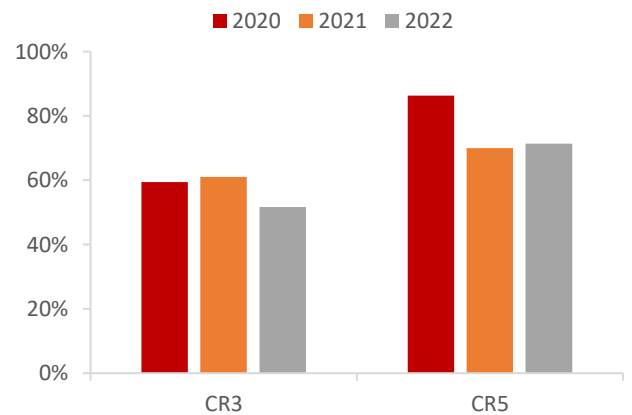
运营商及大型科技企业占据公有云主要份额, 资源优势、渠道优势、算力网络能力占优的大型企业将继续作为建设主体。据通信院统计, 2022 年阿里云、天翼云、移动云、华为云、腾讯云、联通云占据中国公有云 IaaS 市场份额前六。市场集中度有略微的下滑, 但基本保持稳定。云计算厂商后续的竞争力预计仍将主要体现在三个方面: 1) 资源优势。主要体现为数据中心建设用地、政府支持、现金流充裕等方面。2) 渠道优势。在高端计算芯片短缺的背景下, 与主要芯片厂商有紧密的供货关系可视为渠道优势。同时, 互联网云服务厂商在营销渠道和客户触达方面有一定的优势。3) 算力网络能力。通信设备厂商及运营商具有较明显优势。

图表11: 中国公有云 IaaS 厂商市场份额



资料来源: 中国信通院, 方正证券研究所

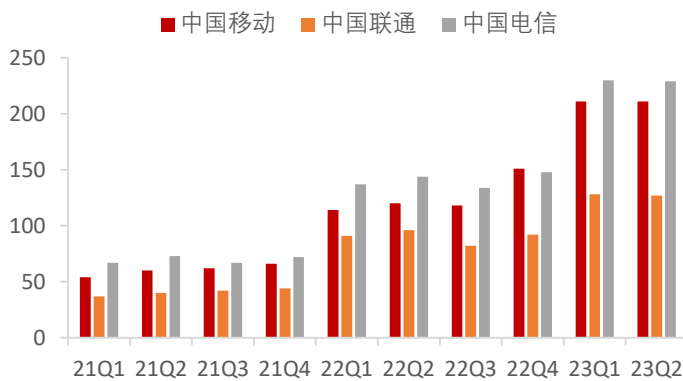
图表12: 中国公有云 IaaS 市场集中度变化



资料来源: 前瞻产业研究, 中国信通院, 方正证券研究所

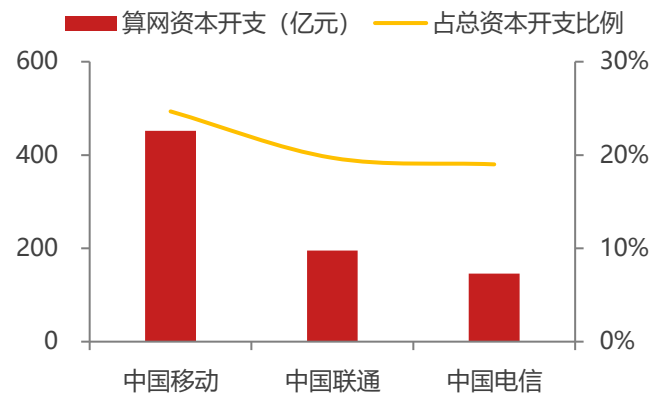
运营商云业务增速超行业平均。运营商云业务 2022 年营收增长均超 100%。2023 年上半年, 移动云 422 亿、同比+80%; 电信云收入 459 亿、同比 63%, 联通大数据收入 29 亿、同比+54%。增速超过行业平均。

图表13:三大运营商云计算收入(亿元)



资料来源:公司公告,方正证券研究所

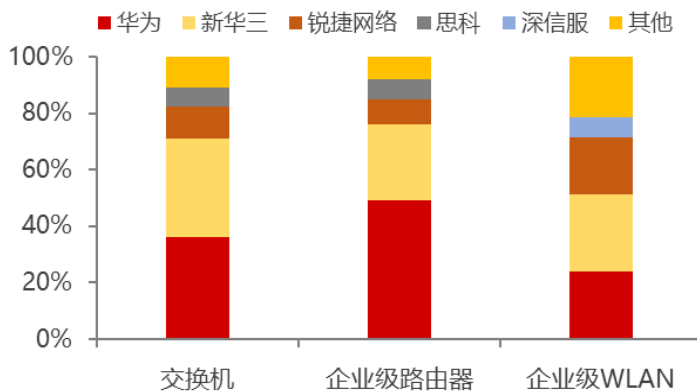
图表14:三大运营商云计算收入增速(%)



资料来源:新浪财经,方正证券研究所

华为在云、计算、存储、网络、能源方面具有综合优势,计算服务能力较强。华为出身通讯设备厂商,在先进网络通信方面持续引领创新,稳稳处于国内计算机网络设备的第一梯队。根据前瞻产业研究院 2023 年 8 月发布的报告,中国计算机网络设备上市公司市场格局看,华为在交换机及企业级路由器行业国内市占率处于第一位,企业级路由器市占率 49%,企业级 WLAN 市占率与新华三接近。华为在算网方面全面的布局,及技术领先性保障了华为总体解决方案的竞争力。在 2023 年人工智能大会上,华为发布业内首个万卡 AI 计算集群解决方案。

图表15:中国计算机网络设备上市企业竞争格局



资料来源:前瞻产业研究,方正证券研究所

图表16:华为具有较强的算网建设能力



资料来源:搜狐新闻,方正证券研究所

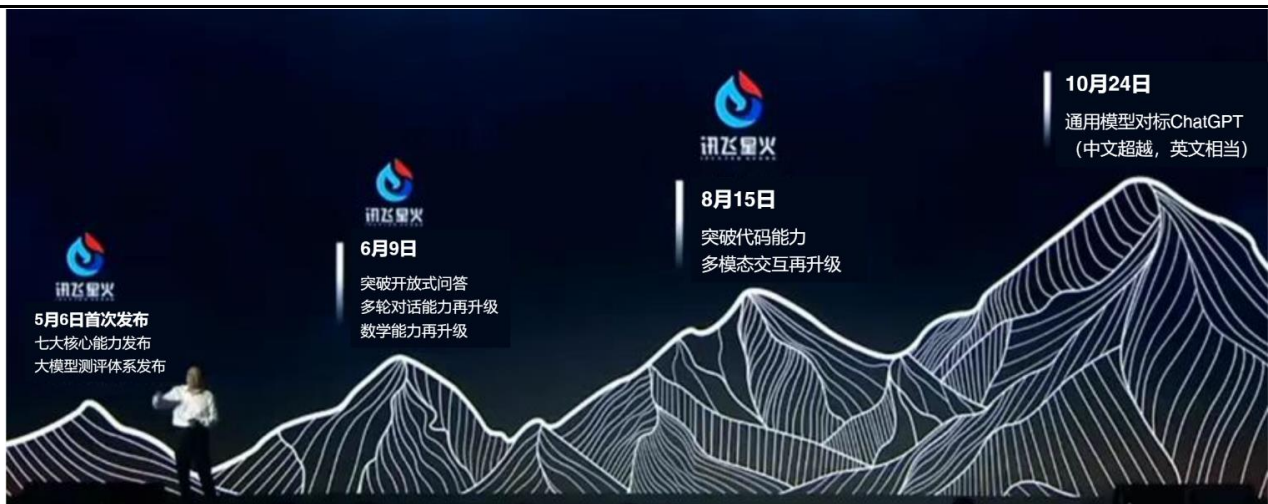
2 算力服务报价可作为跟踪行业算力供需情况的重要指标

落实到产业景气度研究方面,在智能算力发展的第一阶段,市场主要依赖于追踪国内及海外算力硬件龙头的订单情况,亦或是下游主要厂商的资本支出指引,来判断行业总体需求的景气度。但由于订单及资本支出指引可能会存在一定滞后性以及受上下游厂商主观因素的影响,我们认为算力服务报价在及时性和客观性上的优势,市场可以将其作为跟踪行业算力供需情况的重要指标。

2.1 模型训练扩张推升短期算力景气度,中期看推理算力需求释放节奏

当前算力需求主要集中在训练侧，支撑短期算力增长的较好预期。当前国内外的通用大模型仍在不断迭代推出的过程中，在模型参数量、智能化程度方面仍有较大的提升空间。国内各个大模型厂商以海外领先的大模型为追赶方向，正加速迭代。以国内 AI 应用领先企业科大讯飞为例，自 5 月 6 日首发星火大模型以来，持续加大通用大模型的研发，在 6 月及 8 月，分别进行了 1.5 和 2.0 版本的快速迭代，同时计划在 10 月 24 日发布的通用模型实现对标 ChatGPT 的能力。

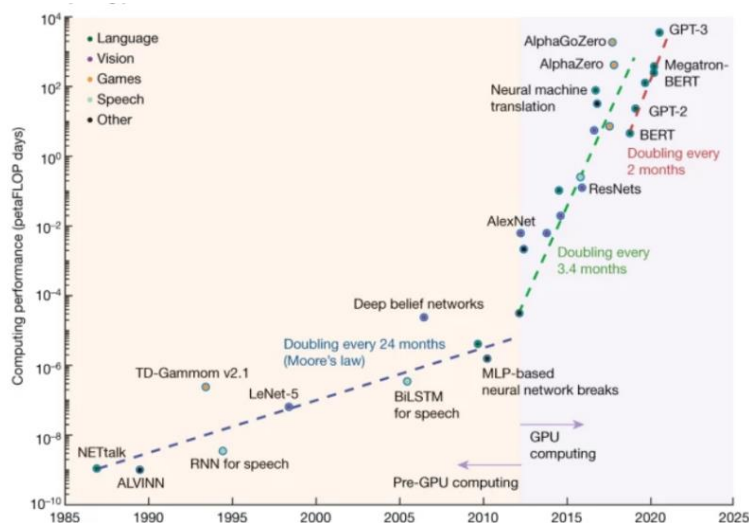
图表17: 讯飞星火大模型给出了明确快速的迭代时间表



资料来源：讯飞星火发布会，方正证券研究所

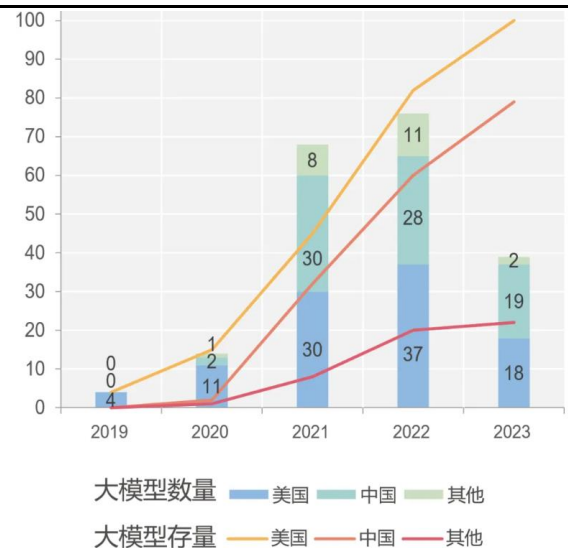
据科技部《中国人工智能大模型地图研究报告》，截至 2023 年 5 月，中国人工智能大模型已发布 79 个，数量排名全球第二，仅次于美国。因此，在 GPT 鲑鱼效应的催化下，国内外有较大数量的大模型处在不断迭代的过程中，总体的参数量仍将扩大，预计至明年初，整体训练端的算力需求仍将有较好的表现。

图表18: 模型参数量及训练所需算力快速增长



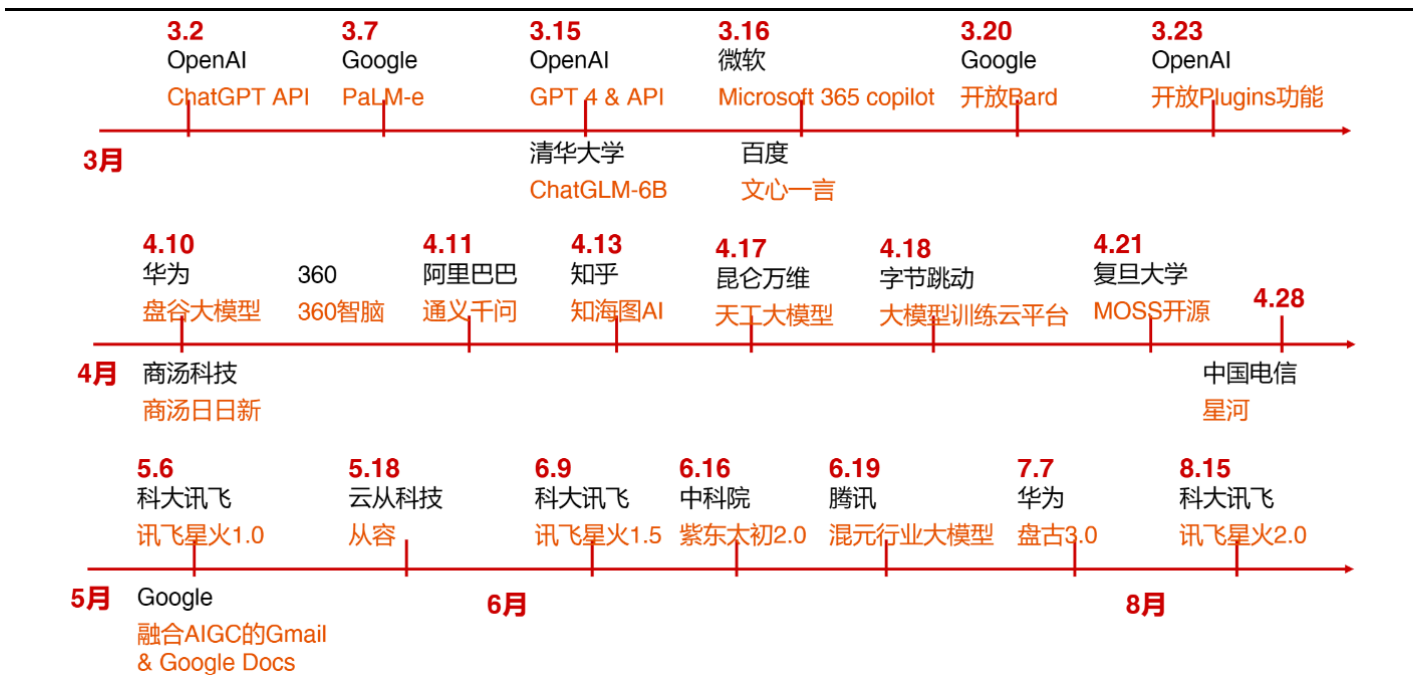
资料来源：Nature《Compute trends across eras of machine learning》，方正证券研究所

图表19: 大模型数量仍在较快增长



资料来源：《中国人工智能大模型地图研究报告》，腾讯新闻，方正证券研究所

图表20: 国产大模型保持较高的推出及迭代节奏



资料来源: 各公司官网, 腾讯新闻, 新华网, 澎湃新闻, 方正证券研究所

随着基础模型的成熟, 预计大模型的格局将演化为“基础模型+专业模型”的模式, 通用大模型的训练算力需求逐步走向收敛。通用大模型长期将演化为类似于PC时代的操作系统, 扮演一个模型底座的作用, 应用需求将逐步向头部几个通用大模型集中。因此, 总体上的通用大模型训练所需算力预计也将随之收敛。

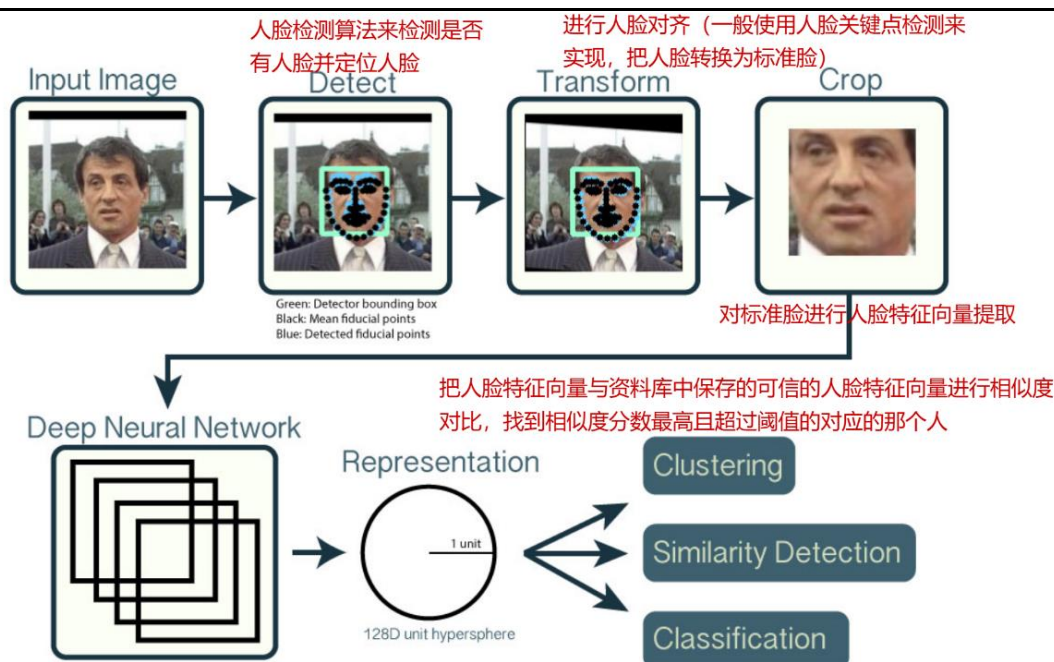
图表21: 大模型将逐步向头部收敛



资料来源: 《通信世界》总第925期《国内外大模型产业如何发展?》, 方正证券研究所

随着训练侧算力增长逐渐平滑, 需要推理侧的算力需求接力。推理侧算力需求的市场空间相较训练侧更大, 将是中长期算力需求增长的主要来源。推理运算作为将人工智能模型落地终端的环节, 推理侧算力需求与下游应用需求的放量紧密相关。在此过程中, 亦可能会有由于训练与推理两端需求起量节奏异步, 导致总体算力需求短期下滑的风险。

图表22:人脸识别任务涉及到多个模型的推理



资料来源: AWS, 方正证券研究所

2.2 当前 AI 算力核心上游硬件处于供不应求的状态，下游订单量较为乐观

从上游订单情况来看，算力需求激增致使上游关键硬件及产能紧缺。GPU 芯片凭借其强大的并行计算性能，在智能计算时代成为算力环节中最重要的组成部分。英伟达作为全球 GPU 龙头，市场份额超 80%。年初随着大模型关注度不断提高，算力建设加速，英伟达公司及外部媒体口径的订单需求不断走高，市场亦将其作为判断智能算力产业链景气情况的重要依据。甲骨文预计今年将斥资数十亿美元购买英伟达芯片。根据英国《金融时报》8 月 10 日报道，中国大型互联网公司已向英伟达订购了价值 10 亿美元的共约 10 万张 A800 芯片，将于年内交付，另外价值 40 亿美元的芯片将于 2024 年交付。8 月 15 日媒体指出沙特购买了至少 3000 块英伟达的 H100 芯片，阿联酋同时采购数千块英伟达芯片。在芯片订单激增的情况下，智能芯片上游代工制造及先进封装亦不时被提及产能紧缺的问题。据中国台湾《电子时报》，台积电 CoWoS 产能至 2023 年底订单需求已分配完毕。

从产业链反馈看，出现明显价格上涨及交货周期延长问题。英伟达高端芯片现货价格年初以来由于供应紧张也在不断攀升。据媒体报导，英伟达 A100 价格从 2022 年 12 月开始上涨，截至今年 4 月上半月，其 5 个月价格累计涨幅达到 37.5%，同期 A800 价格累计涨幅达 20.0%。7 月中旬英伟达芯片国内售价达 10 万元/片，单周涨幅超 20%。新华三运营总经理在 MWC 上海展向媒体表示，搭载英伟达高端芯片的服务器供货周期在 30 周左右。根据 Meta，微软及 Google 等 2023Q2 财报交流，下半年 AI 计算的投资预计仍保持明显的增长。

图表23: 智算芯片订单及价格相关新闻枚举

时间	新闻
5月14日	英伟达 A100 价格从 2022 年 12 月开始上涨, 截至今今年 4 月上半月, 其 5 个月价格累计涨幅达到 37.5%; 同期 A800 价格累计涨幅达 20.0%。同时, 英伟达 GPU 交货周期也被拉长, 之前拿货周期大约为一个月, 现在基本都需要三个月或更长。
6月6日	英伟达今年对 CoWoS 的需求将达 4.5 万片, 较年初预估的 3 万片晶圆大幅增长 50%。
6月29日	甲骨文今年将斥资数十亿美元购买英伟达芯片, 以扩展针对新一波人工智能浪潮的云计算服务。此外, 埃里森还透露, 除了英伟达, 他们还将从 Ampere Computing 和 AMD 购买 CPU。
7月3日	英伟达、博通、AMD 陆续上调对台积电订单, 整体下单规模较 2023 年至少再增 2 成以上
7月28日	台积电今年 CoWoS 产能短缺难以缓解, 至年底订单需求已分配完毕
8月7日	英伟达 H100 涨到 4.5 万美元 海外已经有创业企业开始利用 GPU 进行抵押融资
8月11日	百度、字节跳动、腾讯和阿里已向英伟达订购了价值 10 亿美元的共约 10 万张 A800 芯片, 将于年内交付, 另外价值 40 亿美元的芯片将于 2024 年交付。
8月15日	沙特和阿联酋正购买数千颗高性能英伟达芯片。沙特通过阿卜杜拉国王科技大学购买了至少 3000 块英伟达的 H100 芯片; 阿联酋在研发自己大语言模型 Falcon 同时采购数千块英伟达芯片。

资料来源: 金融时报, U. S. News, 财联社, 澎湃新闻, 腾讯新闻, 方正证券研究所

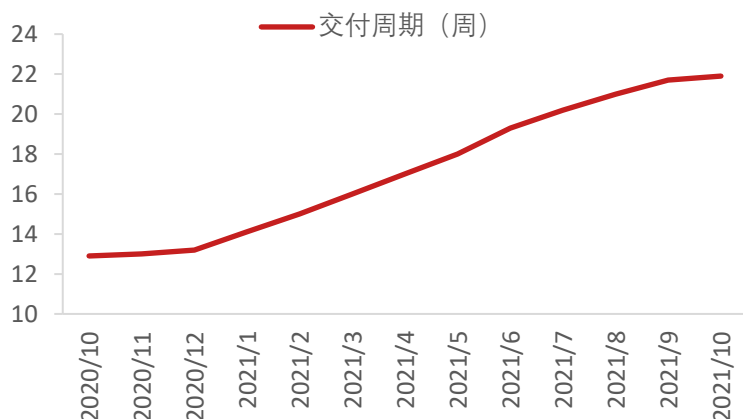
2.3 算力订单数据存在滞后性及真实性担忧, 需要更多维验证

订单情况反应的是乐观情况下的需求预期, 行业增速放缓会发生砍单情况。

2019 年-2020 年的比特币等虚拟货币的挖矿热, 以及 2020 年-2021 年的智能电动车缺芯潮中, 在关键硬件供给紧张的趋势下, 会出现下游需求方报高意向订单量, 以保证上游关键紧缺零部件拿货的优先等级的情况, 后续如果下游需求放缓, 可能会发生砍单的情况。

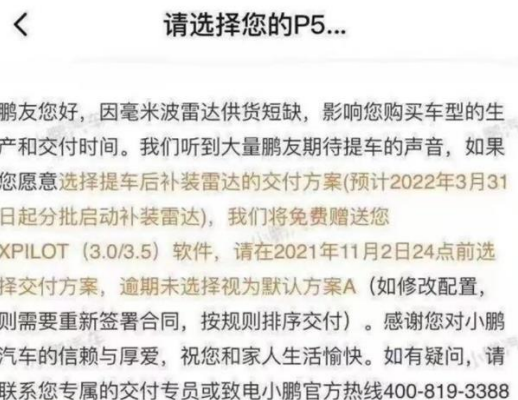
例如, 2021 年全球新冠疫情爆发, 叠加日本地震及瑞萨工厂火灾、美国寒潮、中国台湾缺水电等诸多黑天鹅事件, 半导体全球供应链面临巨大挑战, 芯片平均交货周期由三个月延长至半年至一年。同时, 随着国内外新能源车政策推动, 汽车电动化渗透率加速上升, 智能化从 0 到 1, 成为车厂弯道超车, 差异化竞争的关键。因此, 汽车芯片的下游旺盛的需求与上游供给紧缺的矛盾不断上升, 芯片价格不断上涨。在此背景下, 车厂纷纷调高出货指引, 以争取有限的芯片供给。2022 年以来, 随着疫情等不利影响消除, 汽车价格战以及下游需求放缓背景下, 车厂的智能化硬件投入亦趋缓, 随即出现较多砍单的消息。

图表24:2020.10到2021.10全球芯片平均交付周期(周)



资料来源:盖世汽车,方正证券研究所

图表25:2021年缺芯压力导致车厂“减配交付”



资料来源:小鹏汽车APP,盖世汽车,方正证券研究所

图表26:2022年汽车芯片紧缺程度缓解后出现较明显砍单情况

时间	新闻
2021年7月	大众集团指出,汽车芯片的持续短缺可能会在未来几个月内加剧,将在2021年第三季度面临非常大的挑战。
2021年8月	从2020年底开始的大半年时间,汽车芯片价格动辄上涨了几十倍、甚至上百倍。丰田计划9月将其全球产量削减40%。通用汽车表示受全球缺芯影响,计划在8月27日/关停负责生产雪佛兰Bolt EV和Bolt EUV的工厂。
2021年10月	奔驰对外宣布,因为芯片短缺问题不再保证为新车提供全部配置。
2021年11月	特斯拉11月交付的Model 3和Model Y汽车,车辆内饰中控和后排位置缺少USB接口,只能看到相应的预留孔位,而中控下方无线充电功能同样被“减配”。此前,因芯片短缺影响,特斯拉Model Y和Model 3在美国的售价已经多次上调。据AutoForecast Solutions,截至2021年11月16日,由于芯片短缺,今年全球汽车市场累计减产已达1009.7万辆。
2022年7月	此前紧缺的博世L9369-TR芯片价格从2021年下半年3500元下跌到671元每颗,跌幅8成。意法半导体STM32F103C8T6芯片售价从200元下降为21.5。经销商现货充足。
2022年11月	台积电对供应链大幅砍单,幅度最高达40%-50%。据钜亨网数据,台积电原定年底3nm月产能可达4.4万片,但如今仅剩1万片左右,降幅超过77%。瑞萨与安森美两大半导体厂已发出砍单令,正在削减第4季的芯片测试订单。
2023年4月	车企的降价潮已经影响到车用芯片产业,部分车厂在针对电源管理IC、金氧半场效应晶体管(MOSFET)、微控制器(MCU)等芯片进行砍单,并要求供应商降价。

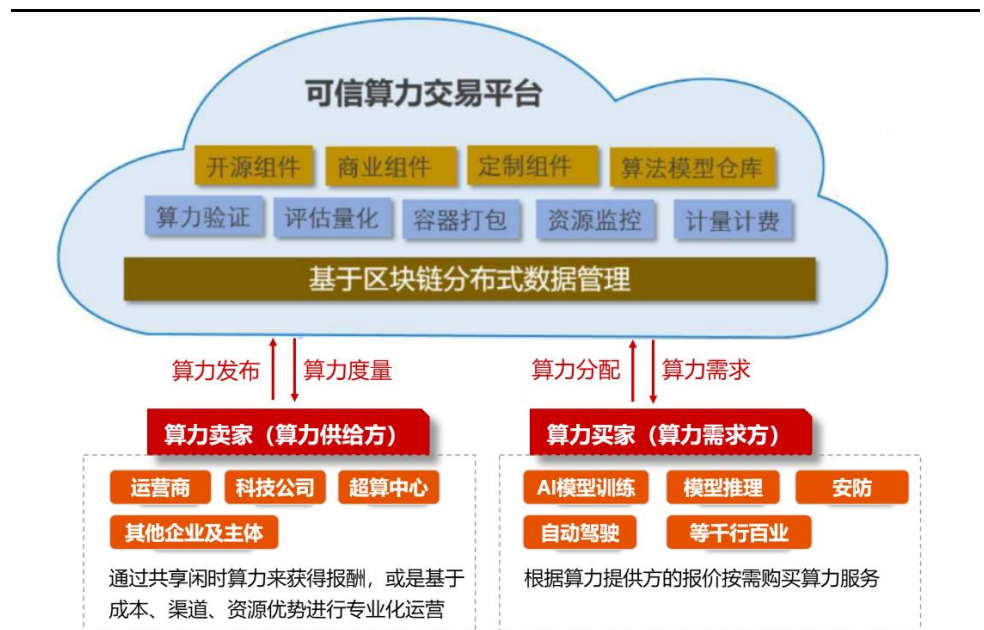
资料来源:腾讯新闻、新浪新闻、中经网,方正证券研究所

订单数据的跟踪相对低频和滞后一些,需要更多维度的算力需求预警指标。订单数据不仅可能存在虚高的可能,而且在时效性上相对滞后。由于订单数据主要依赖上市公司财报披露或定期的自愿交流,因此会较为低频。投资者对能够更及时有效反映市场供需情况的指标有较明显需求。

2.4 算力服务价格可作为判断行业算力需求节奏的显著指标

算力网络应用场景不断丰富，算力网络交易的商业模式逐渐浮现。未来计算场景将无处不在，从更精确的探知未知到准确的模拟现实，从更聪明的个人助理到更高的生产效率，从更高的运营效率到数据驱动的业务创新。下游算力应用场景不断丰富，对算力的需求快速增长也更加多元。算力交易平台作为中间角色连接算力的卖家和买家，维护、纳管、调度算力资源，提供经济、高效、安全、可靠的算力服务。交易平台促进算力拥有方的闲置算力资源变现，提高整个社会的算力使用效率，政策推动与行业需求共振，算力交易的规模及量级也将不断扩大。

图表27: 算力交易实时反应供需情况



资料来源：中国联通《算力网络架构与技术体系白皮书》，方正证券研究所

国内首个算力服务交易平台上线，算力交易实践再进一步。2023年3月，国内首个算力交易平台开始运行，首批将曙光、华为、中兴、阿里云、天翼云、浪潮集团等国内大算力领先的企业，国家信息中心、北京大数据研究院等国内主要的大数据机构，以及商汤、百度等科技企业共计27家纳入平台。2023年4月，“算力浦江”（首届）数字经济发展论坛上，上海算力（试运行）交易平台正式启动，计划到2023年底，初步建成算力调度平台，探索算力交易运营机制；进一步扩大交换流量规模，实现上海市内算力资源调度。算力交易平台实践不断深入，推动算力报价更加透明和及时。

图表28:上海算力交易平台（试运行）上线



资料来源：中国金融信息中心，方正证券研究所

图表29:上海算力交易平台对供需撮合有较好的探索



资料来源：上海算力交易平台网站，方正证券研究所

算力交易模式逐步成熟，云计算服务报价的数据价值将更加凸显。算力作为越来越重要的资源，与其他资源一样，其价格将更多由市场的实时供求关系来决定。由于算力服务报价可以直观刻画算力供给及需求情况，具有实时度高、透明度强的特点，也将为二级市场投资者判断算力市场热度提供重要参考。

各厂商云计算报价相对高频但非标，需要横向及纵向控制变量。算力服务的价格相对非标：1) 不同厂商几个竞争策略会有差异；2) 云计算平台的能力具有差异。这些能力主要体现为算力支持的地域广度，平台服务能力，以及开发者生态丰富程度等。如果考虑 AI 能力输出、模型剪枝蒸馏等，云计算服务价格可能会更高。3) 硬件层面。考虑服务器类型、CPU 及 GPU 型号、内存、带宽、是否独享等多种因素。

图表30:云厂商提供的服务器类型

阿里云	华为云	腾讯云	火山云	Azure	Google Cloud	AWS
通用型	通用计算型	标准型	通用型	通用型	GPU	通用计算型
计算型	通用计算增强型	内存优化型	计算型	网络优化	Cloud TPU	计算优化型
内存型	内存优化型	高 IO 型	内存型	内存优化型	预定义 vCPU 和内存	内存优化型
大数据型	超大内存型	大数据型	本地 SSD 型	存储优化型	自定义 vCPU 和内存	加速计算型
本地 SSD 型	高性能计算型	计算型	大数据型	计算密集型	内存优化 vCPU 和内存	存储优化型
高主频计算型	磁盘增强型	GPU 型	高主频型		共享核心机器	
共享型	超高 I/O 型	FPGA 型	GPU 计算型		高 CPU 机器	
增强型	FPGA 加速型	裸金属	高性能计算		高内存机器	
弹性裸金属	GPU 加速型	高性能计算集群	弹性裸金属		标准机器	
GPU 型	通用入门型		突发型		内存优化机器	
超级计算集群	AI 加速型		共享型			

资料来源：各云厂商官网，方正证券研究所

图表31:国内头部云计算厂商均可提供一定的全球算力服务能力

	阿里云	华为云	腾讯云	火山云	Azure	Google Cloud	AWS
中国	华北、华东、华南、华中、西南、香港	北京、上海、广州、贵阳、香港	北京、上海、南京、成都、重庆、广州、香港	华北、华东、华南	北京、上海、香港	台湾、香港	北京、宁夏、香港
亚洲其他	新加坡、印度、迪拜、日本、印度尼西亚、菲律宾、泰国、马来西亚、韩国	新加坡、泰国、马来西亚、印度尼西亚、阿布扎比	新加坡、印度、泰国、韩国、日本、印尼	柔佛	印度、日本、韩国等	印度、日本、新加坡	印度、日本、新加坡、巴林等
欧洲	英国、德国	法国、荷兰、爱尔兰	德国		英国、德国、法国等	英国、荷兰、芬兰、比利时、德国、瑞士	英国、瑞典、爱尔兰、法国等
北美洲	美国	墨西哥	美国、加拿大		美国、加拿大	美国、加拿大	美国、加拿大等
其他	澳大利亚	南非、智利、巴西、阿根廷	巴西		澳大利亚、巴西、非洲等	澳大利亚、巴西等	澳大利亚、巴西等

资料来源：各云厂商官网，速石科技，方正证券研究所

控制常规变量：将计算节点控制为北京（除华为 g6.18xlarge.7 部署节点为贵阳），BGP 多线，带宽上限为 1Mbps，独立公网 IP，50GB 性能型 SSD 系统盘，独享线路，单台服务器一年的价格如下表。

横向对比各大云厂商云计算服务的价格孰高孰低的参考性不高。可以看出，在控制一些通常的指标之后，各个厂商提供的云服务器在具体硬件参数和计算框架方面仍存在差异，华为等主要发力国产软硬件生态，横向对比各家云服务的性价比需要控制的变量较多。1) 从中端云计算服务器看，腾讯云及火山云提供的采用英伟达 A100 和 V100 芯片的 32 核云计算服务器一年的价格在 25-27 万的范围。2) 从高性能服务器来看，当前腾讯云及火山云提供基于英伟达 A100 的计算服务器，8 卡英伟达 A100 服务器一年的价格大约是 110-150 万元。

综合各云厂商的算力服务价格环比变化，预计对判断算力供需有较大参考意义。单一厂商的云服务价格可能会受市场策略等因素影响，控制常规变量情况下，对多厂商算力服务价格边际变化情况的跟踪，将对判断算力总体供需情况提供明显的价值。

图表32: 国内头部云计算厂商高性能云计算价格

型号	vCPU (核)	内存 (GiB)	GPU 型号	内网带宽	网络收发包	费用 (每年)
中端算力云服务器:						
阿里云 ecs.g8ae.8xlarge	32	128	-	-	-	53631.02 元
华为云 pi1.8xlarge.4	32	128	NVIDIA P4 * 4	-	-	19744.4 美元
腾讯云 GT4.8XLARGE192	32	192	NVIDIA A100 * 2	10Gbps	235 万 PPS	274557.36 元
火山云 ecs.g1te.4xlarge	16	52	NVIDIA T4 * 2	-	-	70174.02 元
ecs.g1ve.8xlarge	32	128	NVIDIA V100 * 4	-	-	251137.66 元
高性能算力云服务器:						
阿里云 ecs.gn7r-c16g1.32xlarge	128	512	ARM+GPU 加速 (型号不详)	-	-	5237.31 元
ecs.video-trans.26xhevc	104	192	ASIC 加速 (Intel Xeon (Cascade Lake) Platinum 8269CY)	-	-	4344.67 元
华为云 p2v.16xlarge.8	64	512	NVIDIA V100-SMX2 * 8	-	-	90258.4 美元
g6.18xlarge.7	72	504	NVIDIA T4 * 2	-	-	23217.37 美元 (贵阳节点)
km1.15xlarge.8	60	480	鲲鹏 Km1	-	-	13427.4 美元
腾讯云 G13X.22XLA	90	226	NVIDIA T4 * 2	13Gbps	375 万 PPS	92329.2 元
PNV4.28XLARGE466	112	466	NVIDIA V100 * 4	50Gbps	950 万 PPS	205434.96 元
GT4.41XLA	164	948	NVIDIA A100 * 8	-	-	1136296.56 元
火山云 ecs.pni2.7xlarge	28	490	NVIDIA A100 * 2	-	-	393315.96 元
ecs.pni2.28xlarge	112	1960	NVIDIA A100 * 8	-	-	1571066.04 元

资料来源: 各云厂商官网, 方正证券研究所

通过跟踪微软 AZURE 及亚马逊云服务的价格, 可以看出两家厂商在国内北京节点的报价相较美东节点的高性能算力服务报价更高, 可能由于国外厂商运营成本以及高端芯片进出口成本导致。与国内云计算厂商相比, 国外厂商的云计算服务在中国可以提供的算力性能上限更高, 预计由于海外云计算大厂在高端芯片, 例如 H100 等的采购限制更少。

图表33: AZURE 及 AWS 高性能云计算服务价格

型号	vCPU (核)	内存 (GiB)	费用 (每月)	费用 (每年)
国内定价 (北京节点, 元)				
Amazon EC2 g4dn.8xlarge	32	128	25802.58	309630.96
g4dn.16xlarge	64	256	42523.23	510278.76
x2idn.16xlarge	64	2048	119687.64	1436251.68
Azure 机器学习 D32 v3/DS32 v3	32	128	8548.56	102582.72
E64 v3/E64s v3	64	432	34715.04	416580.48
M128ms	128	3892	145571.04	1746852.48

美国定价（美东节点，美元）

Amazon EC2	g4dn.8xlarge	32	128	3144.11	37729.32
	g4dn.16xlarge	64	256	4827.49	57929.88
	x2idn.16xlarge	64	1024	6815.21	81782.52
AZURE 机器学习	D32 v3/DS32 v3	32	128	971.15	11653.8
	E64a v4	64	512	1990.31	23883.72
	M128ms	128	3800	13440.8	161289.6

资料来源：各云厂商官网，方正证券研究所

分析师声明

作者具有中国证券业协会授予的证券投资咨询执业资格，保证报告所采用的数据和信息均来自公开合规渠道，分析逻辑基于作者的职业理解，本报告清晰准确地反映了作者的研究观点，力求独立、客观和公正，结论不受任何第三方的授意或影响。研究报告对所涉及的证券或发行人的评价是分析师本人通过财务分析预测、数量化方法、或行业比较分析所得出的结论，但使用以上信息和分析方法存在局限性。特此声明。

免责声明

本研究报告由方正证券制作及在中国（香港和澳门特别行政区、台湾省除外）发布。根据《证券期货投资者适当性管理办法》，本报告内容仅供我公司适当性评级为 C3 及以上等级的投资者使用，本公司不会因接收人收到本报告而视其为本公司的当然客户。若您并非前述等级的投资者，为保证服务质量、控制风险，请勿订阅本报告中的信息，本资料难以设置访问权限，若给您造成不便，敬请谅解。

在任何情况下，本报告的内容不构成对任何人的投资建议，也没有考虑到个别客户特殊的投资目标、财务状况或需求，方正证券不对任何人因使用本报告所载任何内容所引致的任何损失负任何责任，投资者需自行承担风险。

本报告版权仅为方正证券所有，本公司对本报告保留一切法律权利。未经本公司事先书面授权，任何机构或个人不得以任何形式复制、转发或公开传播本报告的全部或部分内容，不得将报告内容作为诉讼、仲裁、传媒所引用之证明或依据，不得用于营利或用于未经允许的其它用途。如需引用、刊发或转载本报告，需注明出处且不得进行任何有悖原意的引用、删节和修改。

评级说明：

类别	评级	说明
公司评级	强烈推荐	分析师预测未来12个月内相对同期基准指数有20%以上的涨幅。
	推荐	分析师预测未来12个月内相对同期基准指数有10%以上的涨幅。
	中性	分析师预测未来12个月内相对同期基准指数在-10%和10%之间波动。
	减持	分析师预测未来12个月内相对同期基准指数有10%以上的跌幅。
行业评级	推荐	分析师预测未来12个月内行业表现强于同期基准指数。
	中性	分析师预测未来12个月内行业表现与同期基准指数持平。
	减持	分析师预测未来12个月内行业表现弱于同期基准指数。
基准指数说明		A股市场以沪深300 指数为基准；香港市场以恒生指数为基准，美股市场以标普500指数为基准。

方正证券研究所联系方式：

北京：西城区展览馆路 48 号新联写字楼 6 层

上海：静安区延平路71号延平大厦2楼

深圳：福田区竹子林紫竹七道光银行大厦31层

广州：天河区兴盛路12号楼隼峰苑2期3层方正证券

长沙：天心区湘江中路二段36号华远国际中心37层

网址：<https://www.foundersc.com>

E-mail：yjzx@foundersc.com