

证券研究报告

AI时代的云计算

——供给普惠化与需求扩容带来全新增长空间

2025.06.13

分析师: 苏仪

执业证书编号: S0740520060001

分析师: 王雪晴

执业证书编号: S0740524120003

分析师: 闻学臣

执业证书编号: S0740519090007

分析师: 刘一哲

执业证书编号: S0740525030001



核心观点

- **云计算的本质在于计算资源和服务的调配,边缘计算重要性正在凸显。**云计算的价值体现在计算资源池的统一调度, 能够为下游用户提供灵活、高性能和低成本的计算资源;爆发式的数据处理需求和实时运算需求使得边缘计算越来越 重要。
- AI时代云计算的产业逻辑主要可以概括为供给端云算力普惠化,需求端下游企业需求扩容所带来的整体机会。
- ▶ 从供给端看: 1) 云服务laaS+PaaS/MaaS层正在经历规模化降本。一方面云计算大厂和数据中心正在加速扩容(供给扩张),另一方面云计算底层软硬件和基模突破带来了云服务性能和效率的提升(成本下降)。2) Al+SaaS应用带来软件产品力提升.为下游Software as a Service商业模式打开空间。
- ▶ **从需求端看:** 1)生成式AI驱动需求不断增长。训练端AI模型常常部署在大规模公有云集群;推理端Post-Train和Test-Time Scaling下推理算力需求扩增,驱动大规模公有云、私有云和混合云部署需求; AI+SaaS服务正成为很多中小企业部署的首选。2)我国政策始终重视云计算。伴随AI部署,政务云和金融云等行业云正加速落地。
- 建议关注: 1) 云计算软硬件服务提供商:深信服、金山办公、阿里巴巴、腾讯控股、百度集团、中国联通、中国电信、中国移动、浪潮信息、中科曙光、紫光股份; 2) AI基础模型厂商:阿里巴巴、科大讯飞、商汤; 3) AI应用及SaaS应用厂商:焦点科技、彩讯股份、万兴科技、广联达、明源云: 4) IoT及端侧计算厂商:地平线机器人、中科创达。
- 风险提示:云计算及AI技术落地不及预期、行业竞争加剧、报告信息更新不及时等。



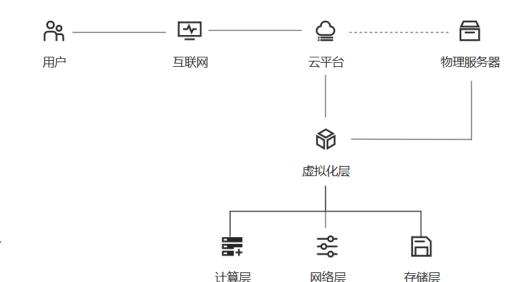


云计算: AI时代应当如何定义?



1.1 什么是云计算

- 云计算:一种通过互联网提供计算资源和服务的技术,允许用户随时随地访问、使用云平台上数据、软件和硬件资源。 在数字化时代,互联网已经成为基础设施。云计算使得数据中心能够像一台计算机一样去工作。通过互联网将算力以 按需使用、按量付费的形式提供给用户,包括:计算、存储、网络、数据库、大数据计算、大模型等算力形态。
- 在云计算中,用户通过互联网连接到云平台,通过云平台来管理和使用云计算资源。云平台由云服务提供商维护和管理、它提供了计算、存储和网络等资源。
- ▶ 虚拟化层:将物理服务器抽象为虚拟的计算资源,使用户可以根据需要动态分配和管理这些资源。虚拟化层允许多个虚拟机在同一台物理服务器上运行,实现资源的共享和高效利用。
- ▶ 存储层:提供了存储空间供用户存储和管理数据。用户可以根据需要 选择适当的存储类型,如对象存储、块存储或文件存储,来满足不同 的存储需求。
- 网络层:云平台中负责网络连接和通信的部分。它负责将用户和云平台之间的数据的传输和通信进行处理和管理,确保数据的安全和可靠传输。
- ▶ 计算层:云平台中负责处理和执行计算任务的部分,包括虚拟机实例、容器服务等,提供了计算能力供用户运行应用程序、处理数据等。
- ▶ 物理服务器:云平台中实际的硬件设备,它们是云平台上虚拟化、存储、网络和计算等层级的基础。云服务提供商通过维护和管理这些物理服务器,为用户提供可靠的计算资源和服务。



资料来源: 阿里云, 中泰证券研究所

图表: 云计算工作原理图示



1.1 从价值的本质理解云计算:实现弹性、多租的增量价值

- 云计算与计算是正交关系,而非替代关系。
- <u>"弹性和多租"是云计算相对于计算的主要增量业务价值。</u>实现云计算的核心关键是统一计算资源池。只有实现计算资源池的统一调度,同时技术层面实现多租安全隔离,才能够真正意义上实现全Region(地域)和全AZ(可用区)的计算资源弹性。

图表: 云计算和计算的关系

 学性
 "计算"演进为"云计算"

 多租
 通用计算 加速计算 (AI) 数据库 大数据

图表: 云计算的精确定义

云计算的精确定义

解释

云计算是按需使用无限量 计算资源

线下IDC是基于本地部署 (On Premise) 供应,需在使用前预测需求,并提前完成服务器/交换机/存储的构建。而云计算服务是按需供应,根据实际业务需求做到计算、存储、网络等资源的极致伸缩。

消除云用户的预先承诺

公共云用户在任何一个区域(Region)或可用区(AZ),只需要付费成功就可以 按量使用计算存储网络的资源,并不需要向云提供商承诺要在哪一个地方消费多少 资源。因此,云服务商必须在全AZ可用区和全Region地域供应各种算力需求。

根据实际需要支付短期使用计算资源的费用

用户不需要承诺购买的计算资源的预期使用时间。对于云计算所使用的容器或者函数,计时单位可以精确到毫秒级。用户购买后即使只用几毫秒也可以。当一个计算资源必须让用户使用四年、五年或者整个生命周期,即说明它不是云计算,而是IDC或者私有云。

通过规模经济显著降低成 木

云服务商希望通过使用超大的数据中心,以规模经济放大供应链效益,摊薄研发成 本。

通过资源虚拟化技术简化 操作并提高资源使用率

云服务商希望通过资源虚拟化,把计算、存储、网络抽象后,达到简化操作,提升 资源利用率。

通过多路复用的方式运行 来自不同组织的负载提高 硬件资源利用率 核心是多路复用。通过多租户使用同一套计算资源大池,削峰填谷,保证公共云的硬件资源利用率显著提升。今天智算中心就面临大任务算力不够用(比如模型预训练生成),小任务跑不满(比如模型推理应用)的情况,本质上是因为它们没有统一的云架构。

资料来源: 阿里巴巴, 中泰证券研究所



1.1 云计算的部署方式:公有云、非公有云与混合云

- 根据部署模式,云计算可以分为公有云与非公有云,主要包括三种类型:
- ▶ 公有云: 由第三方提供商管理维护, 面向公众提供服务, 成本低、可扩展性强, 如亚马逊 AWS。
- ▶ 私有云: 由单个组织或企业内部管理使用, 数据安全性和隐私性高, 适合对数据安全要求高的企业。
- ▶ 混合云: 结合公有云和私有云,数据和应用程序可在两者间流动,企业可根据需求灵活选择,如企业可将非敏感业务放公有云,敏感业务放在私有云。

图表:公有云、私有云和混合云的特点

部署模式	公有云	私有云	混合云	
服务对象	所有用户	某一企业内部用户	部署了私有云,同时对拥有云有需求的 企业用户	
主要客户	中小企业、开发者、个人消费者	大中型政企机构	高校、医院、政府、企业	
成本	低	较高	较高	
灵活性	高	有限	高	
数据安全性	低	高	高	
管理复杂度	低	较高	高	
扩展性	高	有限	较高	
典型产品	阿里云、AWS 等	华为企业云、深信服企业云等	IBM 混合云、天翼混合云等	

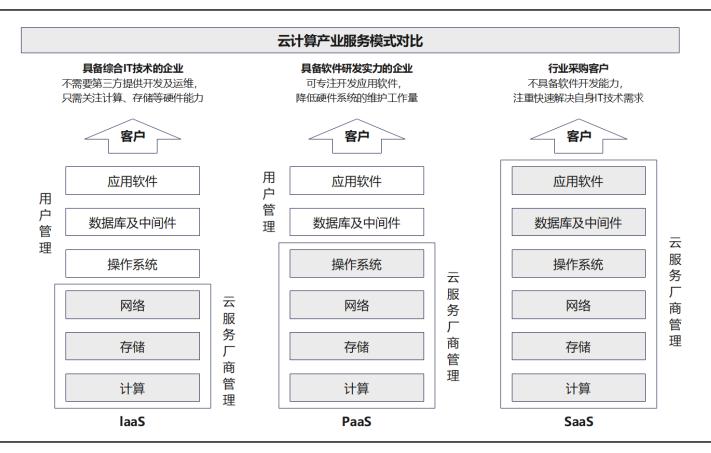
资料来源: 智研咨询, 中泰证券研究所



1.1 云计算的商业模式: laaS、PaaS与SaaS

■ 全球基础设施即服务(laaS)、平台即服务(PaaS)和软件即服务(SaaS)构成了云服务的三大核心模型。

图表:云计算laaS、PaaS、SaaS模式对比



资料来源: 智研咨询, 中泰证券研究所



1.1 AI时代的全新商业模式: MaaS (Model as a Service)

- 智能时代, 大模型得到空前发展和关注, 云计算可以依据业务需求提供以下四种服务模式:
- ▶ 基础设施即服务 (laaS): 提供各类基础设施类能力的服务类别,包括计算、存储、网络等资源服务能力。
- ▶ 平台即服务 (PaaS): 提供应用程序所需要的硬件和软件部署平台的服务类别。
- ▶ 模型即服务 (MaaS): 把 AI 模型当作生产的重要元素,从模型预训练到二次调优,最后到模型部署,围绕模型的生命周期来提供相应的服务。用户可以通过低成本的方式访问、使用、集成模型,提升用户的业务智能化能力。
- ▶ **软件即服务(SaaS)**:提供软件的服务类别,包括协同软件、客户关系管理、企业资源计划、人力资源系统等。用户 无需经过传统的研发流程,而是通过互联网即可使用软件服务,节约了管理基础设施和研发软件的工作。

图表: 以模型为中心的云计算技术体系



资料来源: 阿里巴巴, 中泰证券研究所



1.1 云计算:相比传统IT服务模式优化资源调配方式,提升效率

- 传统 IT 服务模式下,资源固定在某个服务器和存储上,当应用的访问量较低时,由于硬件与应用是绑死的,无法把资源提供给其他应用,服务器和存储资源会大量闲置。
- **云计算模式下**,能够在系统池中动态调配适量的存储、服务和网络资源,自动部署相应的软件资源,使用后可以自动回收。通过云计算,可以把原来分散的数据中心资源包括服务器、存储、网络设备聚集在一起,由专门的服务提供商来运营和维护云计算中心,引用虚拟化技术来突破物理限制、提高资源利用率,实现资源的动态调配和聚合。

图表: 云计算商业模式对传统商业模式的改变



资料来源: 智研咨询, 中泰证券研究所



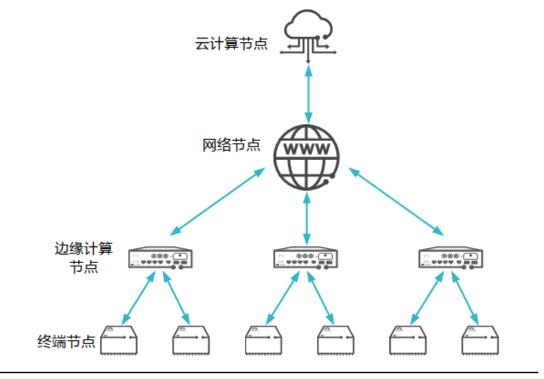
1.2 边缘计算: 物联网时代的终端计算方法

■ 边缘计算是在靠近物或数据源头的网络边缘侧,通过融合网络、计算、存储、应用核心能力的分布式开放平台,就近 提供边缘智能服务。概言之,边缘计算是将从终端采集到的数据,直接在靠近数据产生的本地设备或网络中进行分析, 无需再将数据 传输至云端数据处理中心。

■ 为什么需要边缘计算?

- ▶ 物联网技术的快速发展,使得越来越多具备独立功能的普通 物体实现互联互通,实现万物互联。得益于物联网的特征, 各行各业均在利用物联网技术快速实现数字化转型,越来越 多的行业终端设备通过网络联接起来。
- ➤ 然而,物联网作为庞大而复杂的系统,不同行业应用场景各异,据第三方分析机构统计,到2025年将有超过千亿的终端设备联网,终端数据量将达300ZB,如此大规模的数据量,按照传统数据处理方式,获取的所有数据均需上送云计算平台分析,云计算平台将面临着网络时延高、海量设备接入、海量数据处理难、带宽不够和功耗过高等高难度挑战。
- 为了解决传统数据处理方式下时延高、数据实时分析能力匮乏等弊端,边缘计算技术应运而生。

图表:边缘计算架构



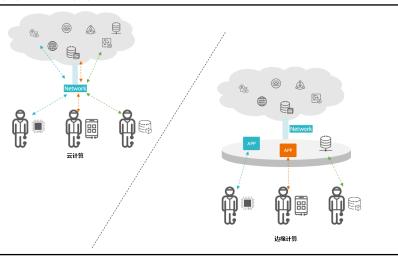
资料来源:华为,Info-Finder,中泰证券研究所



1.2 边缘计算和云计算的联系与区别

- 边缘计算的概念是相对于云计算而言的,云计算处理方式是将所有数据上传至计算资源集中的云端数据中心或服务器处理,任何需要访问该信息的请求都必须上送云端处理。因此,云计算面对物联网数据量爆发的时代,弊端逐渐凸显:
- ▶ 云计算无法满足爆发式的海量数据处理诉求。随着互联网与各个行业的融合,特别是在物联网技术普及后,计算需求 出现爆发式增长,传统云计算架构将不能满足如此庞大的计算需求。
- ▶ 云计算不能满足数据实时处理的诉求。传统云计算模式下,物联网数据被终端采集后要先传输至云计算中心,再通过 集群计算后返回结果,这必然出现较长的响应时间,但一些新兴的应用场景如无人驾驶、智慧矿山等,对响应时间有 极高要求,依赖云计算并不现实。
- 边缘计算的出现,可在一定程度上解决云计算遇到的这些问题。

图表:边缘计算 VS 云计算



图表:边缘计算和云计算的区别

	边缘计算	云计算			
计算方式	分布式计算,聚焦实时、短周期数据 的分析	集中式计算,依赖云端数据中心			
处理位置	靠近产生数据的终端设备或物联网关	云端数据中心			
延时性	低延时	高延时			
数据存储	只向远端传输有用的处理信息,无 冗余信息	采集到的所有信息			
部署成本	低	高			
隐私安全	隐私性和安全性较高	隐私性和安全性相对低,需要高度关 注			

资料来源:华为. Info-Finder. 中泰证券研究所

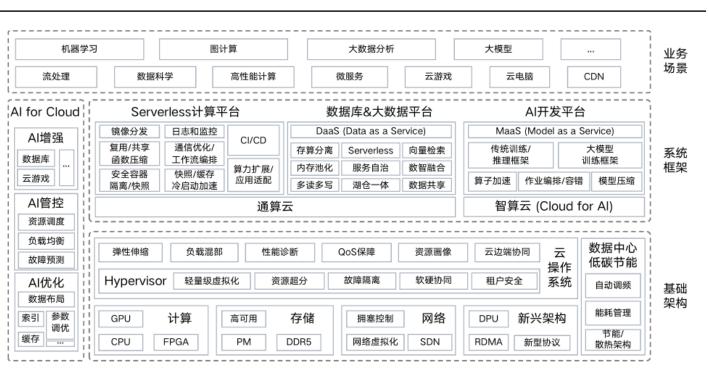
资料来源:华为, Info-Finder, 中泰证券研究所



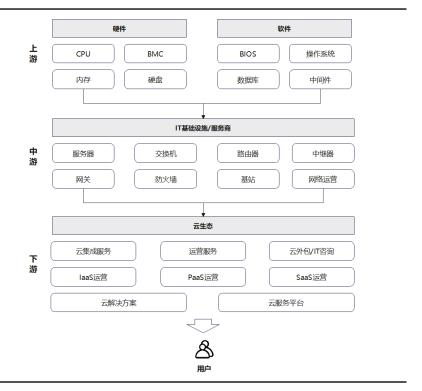
1.3 云计算产业链:综合软硬件、平台服务的完整生态

■ 云计算产业链综合了软硬件与平台服务。产业链上游包括芯片、内存、硬盘等硬件供应商,以及操作系统、数据库等软件供应商,代表企业有Intel、Aspeed、微软等;产业链中游包括服务器、交换机、路由器、中继器等 IT 基础设备制造商,以及网络运营商等各种互联网基础服务商,代表企业有Dell、中国电信、华为、HP等;产业链下游为整个云服务生态,包括云集成服务商及laaS、PaaS和SaaS运营商等,代表企业有华为云、百度云、天翼云等。

图表: 云计算产业和技术分析



图表: 云计算产业结构





1.4 追本溯源:全球云计算产业类别和服务模式逐步完善

- 上世纪八十年代, Sun 公司提出了"网络计算机"概念, 并推出相关的工作站产品; 而后随着虚拟化、格式化、分 布式等技术加速发展, 云计算所依赖的三大底层技术逐个 出现。
- 1996 年左右, Cloud Computing (云计算) 概念正式出 现,在全球范围内掀起了一股云计算产业发展小高潮:
- ▶ 1997 年, InsynQ 基于 HP 的设备上线了按需使用的应用 和桌面服务:
- ▶ 1999 年, MarcAndreessen 创建世界上第一个商业化的 laaS 平台 LoudCloud:
- 自此,全球云计算产业化发展趋势逐渐成形。产业发展至 2005年亚马逊推出 AWS 后,全球云计算行业影响力已 初步显现,开始推动IT产业进行服务模式变革。产业发 展至今,全球云计算不论从类别到服务模式都已十分完善, 服务领域也从企业应用逐渐扩展至政府等公共服务机构。

图表: 全球云计算产业发展历程

√1999年3月, Salesforce成立,最早的SaaS服务出现;同年9月, LoudCloud成立,成为最早的laaS服务商; √2005年, Amazon推出AWS服务; 1999-2006年 √SaaS/laaS云服务出现,并被市场接受。

(成型初期)

√2007年, Salesforce发布Force.com, 即PaaS系统; 2008年4月, Google推出Google APP Engine;

√SaaS、laaS、PaaS三种服务模式逐渐被市场接受, 电信等互联网运 营商纷纷上云。

2006-2008年

(发展阶段)

√2008年,微软发布其公共云计算平台;

√2009年阿里云创立;

√云服务种类日趋完善; 传统IT企业依托客户与合作方优势, 进入云市 场,市面上云服务系统加速涌现。

2008-2015年 (完善阶段)

(成熟阶段)

2015年至今

√产品功能日益健全,市场格局逐渐稳定,行业马太效应显著; √金融机构、政府等公共服务加速上云,云计算市场规模加速扩容; √多云策略成为主流,混合云备受关注。

资料来源: 智研咨询, 中泰证券研究所



1.4 我国云计算的发展历程:启程较晚,正处高速增长期

- 2006年,云计算概念传入我国。与欧美等国家相比,国内云计算市场发展起步相对较晚。随着阿里巴巴于 2008年开始筹备云计算业务、腾讯于 2010年开始跟进云计算,我国云计算市场开始进入成长阶段,行业技术研究逐渐取得新突破、云计算相关应用案例开始逐渐丰富。如 2013年,阿里云宣布突破 5K 测试;同年,腾讯云宣布正式面向全社会开放。
- 近年来,随着数字经济对于国民经济发展推动力日益增强,我国云计算行业发展源动力日益增强,传统 IT 服务企业加速切入云计算领域,产业进入高速增长阶段,全国云计算市场开始逐步走向繁荣。

图表:中国云计算的发展历程

市场引入阶段

(2007-2010年)

- ○云计算概念尚不明确
- ○成功案例十分匮乏
- ○重点厂商各自为阵

成长阶段

(2011-2015年)

- ○応用案例逐渐丰富
- ○用户对云计算已经比较 了解
- ○市面上出现大量功能丰 富的云应用解决方案

成熟阶段

(2015-2017年)

- ○云计算厂商竞争格局基 本形成
- 云计算厂商应用解决方 案更加成熟

高速增长阶段

(2017年至今)

- ○云计算市场整体规模 仍旧偏小
- 云计算市场处于高速 增长阶段
- ○SaaS市场缺乏行业领 军企业

资料来源: 智研咨询, 中泰证券研究所





2

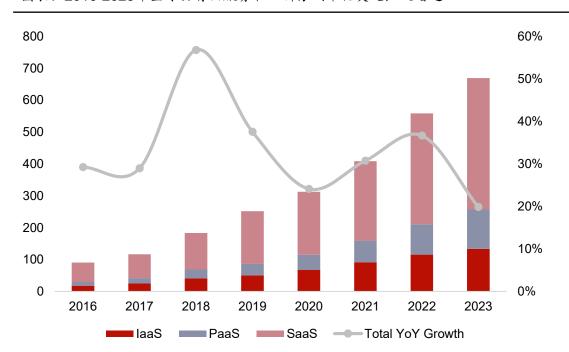
行业现状: laaS+PaaS/MaaS高增, SaaS应用加速拓展



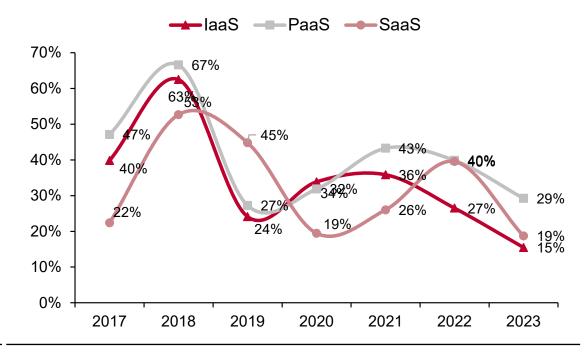
2.1 全球:公有云市场规模快速增长, SaaS占据主导地位

- 从服务模式占比情况看,全球公有云服务市场中SaaS占比最高,2023年实现收入1334亿美元,占市场总额约45%。
- 从增速情况看,2021年PaaS取代laaS和SaaS,成为增速最快的服务方式。2023年PaaS增速达29%。
- IDC预测,2024年全球公共云服务收入将超过8000亿美元,比2023年增长20.5%,预计2025年将实现类似的增长,到2028年全球收入将达到1.6万亿美元。

图表: 2016-2023年全球公有云服务收入拆分(十亿美元)及增速



图表: 2017-2023年全球公有云服务laaS、PaaS、SaaS增速



资料来源: IDC, 中泰证券研究所

资料来源: IDC, 中泰证券研究所



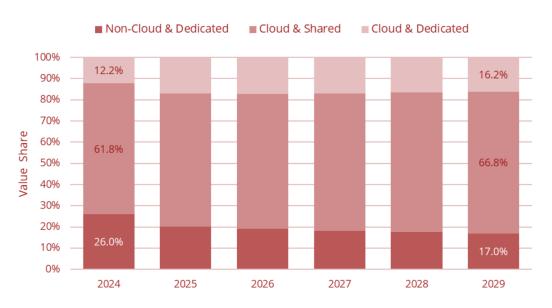
2.1 全球: 预期云基础设施支出将迎快速增长

■ **云基础设施将持续快速增长趋势**。IDC的数据显示,2024年第四季度用于云部署(包括专用和共享IT环境)的计算和存储基础设施产品支出同比增长99.3%,达到670亿美元。长期来看,IDC预测在2024-2029年预测期内,云基础设施支出的复合年增长率(CAGR)将达到17.8%,到2029年将达到4619亿美元,占计算和存储基础设施总支出的83.0%。

图表: 2024E-2034E全球云计算市场规模(十亿美元)



图表:全球企业云基础设施支出拆分预测



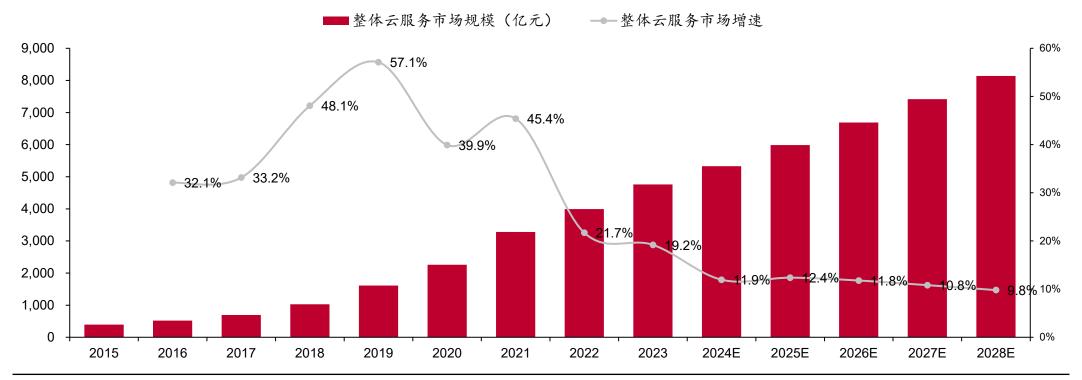
资料来源: IDC, 中泰证券研究所



2.2 我国: 云服务市场增长稳健快速

■ 2023年中国整体云服务市场规模为4758亿元,增速19.2%。以大模型为代表的人工智能相关技术发展,驱动中国整体 云服务市场从"完善综合能力"向"构建融合能力"转型。技术能力革新与商业实践创新成为整体云服务行业关注焦点。

图表: 2015-2028年中国整体云服务市场规模及预测(单位: 亿元)



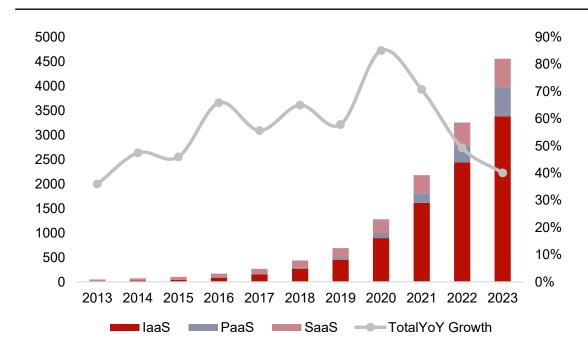
资料来源:艾瑞咨询,中泰证券研究所



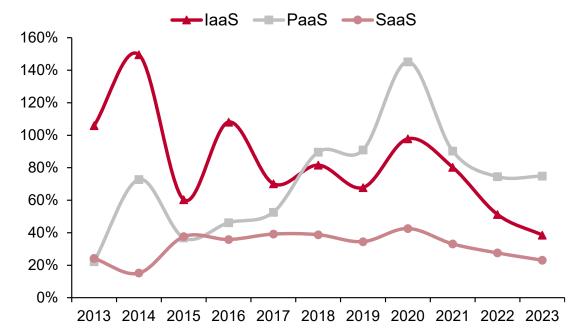
2.3 我国: PaaS成全新增长引擎, laaS依旧占据主导地位

- 从服务方式看,2017年以来laaS始终占据主导地位。2023年我国laaS市场规模达到3383亿元,同比2022年增长39%。
- 从增速情况看,2013-2017年,我国laaS市场增速最快。2018-2023年,PaaS市场增速最快,2023年PaaS市场规模已达598亿元,同比增长75%,超越SaaS市场规模。

图表: 2013-2023年中国公有云市场拆分(亿元)及增速



图表: 2013-2023年中国公有云市场laaS、PaaS、SaaS增速



资料来源: 信通院, 中泰证券研究所

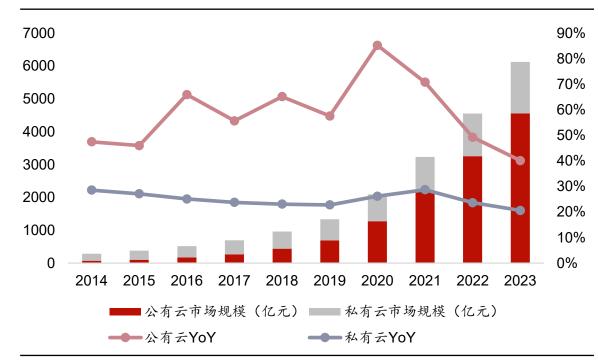
资料来源:信通院,中泰证券研究所



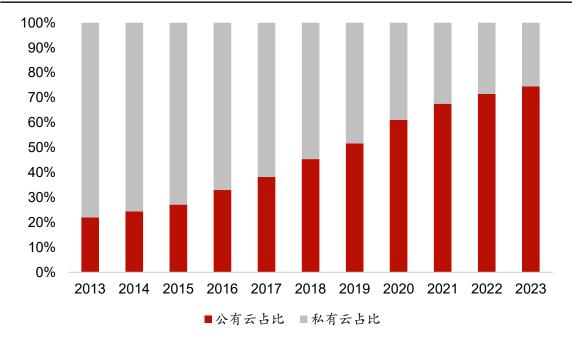
2.4 我国:由私有云主导走向公有云主导,公有云市场规模快速增长

■ 过去10年,我国云服务市场逐渐由私有云主导走向公有云主导。公有云市场快速增长,占比不断提升。2023年我国云计算市场规模达到6165亿元,较2022年增长35.5%,大幅高于全球增速。其中,公有云市场规模4562亿元,同比增长40.1%;私有云市场规模1563亿元,同比增长20.8%。

图表: 2014-2023年中国云服务市场规模及增速



图表: 2013-2023年中国公有云/私有云占比情况



资料来源: 信通院, 中泰证券研究所

资料来源:信通院,中泰证券研究所

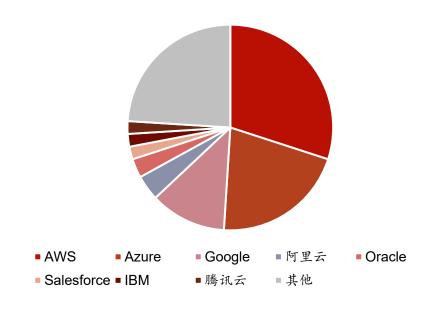


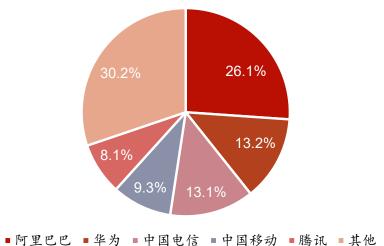
2.5 竞争格局:海外寡头垄断,国内一超多强

- 在laaS+PaaS市场中,海外AWS、微软Azure和Google占据绝对市场规模优势。2024年Q4,全球云基础设施提供商 中AWS占比30%, Azure占比21%, Google Cloud占比12%, 三家厂商合计占比超过60%。
- 国内互联网巨头、软硬件厂商和运营商是laaS市场的主要参与者:互联网背景云厂商占据较大的市场份额,代表性的 阿里云占比26.1%(24H2); 软硬件背景云厂商重视软硬一体的产品研发, 代表性的华为云市占率达13.2%(24H2); 运营商背景云厂商依托集团政策驱动和政企用户的基本盘优势、继续保持市场营收的高速增长。

图表: 2024年Q4全球云基础设施厂商市场份额占比

图表: 2024H2中国公有云laaS厂商市场份额占比





资料来源: Statista. 中泰证券研究所

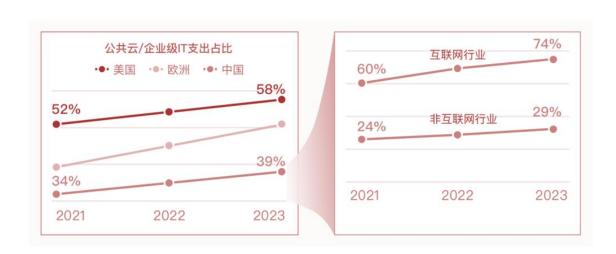
资料来源: IDC. 中泰证券研究所



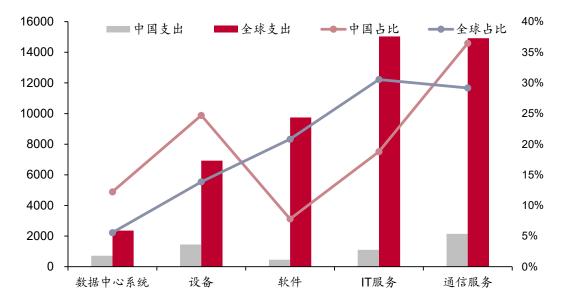
2.6 我国公共云企业级IT支出占比较低,传统行业上云率有待提升

- 云计算在中国发展多年,企业公共云支出占IT支出比例稳步提升,从2021年的34%提升至2023的39%。但与欧美地区的支出占比相比,我国仍存在不小的差距。此外,从行业角度来看,我国互联网行业用云支出占比近四分之三,而非互联网行业用云支出占比则不到三成,有较大提升空间。
- 从IT支出情况看,中国支出主要集中在硬件(数据中心、设备)和通信服务,在软件和IT服务方面支出占比较低,一定程度反映了企业云计算支出还有较大提升空间。

图表: 2021-2023年公共云IT支出占比情况



图表: 2024E中国和全球IT支出(亿美元)及占比情况预测



资料来源: Gartner(2024年7月预测),中泰证券研究所





供给侧规模化降本,需求侧多重驱动, 云计算产业迎来爆发期

5



3供给普惠化+需求多重驱动带来产业爆发

图表: 云计算研究框架

供给侧

技术突破、供给 扩容带来**云算力 普惠化**,同时为 云厂商带来更高 利润



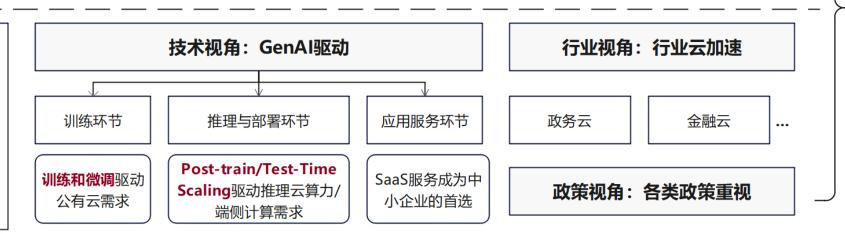
大厂军备竞赛**抢占未来生态**; 各地数据中心战略性部署 软件虚拟化升级/集群构建能力,硬件芯片、服务器等环节 突破,带来**性能和效率的提升**



laaS/PaaS快速扩容; 边缘计算重要性提升; SaaS潜在空间最大

需求侧

云计算的需求本 质来源于**下游企** 业对降本增效和 安全性的追求



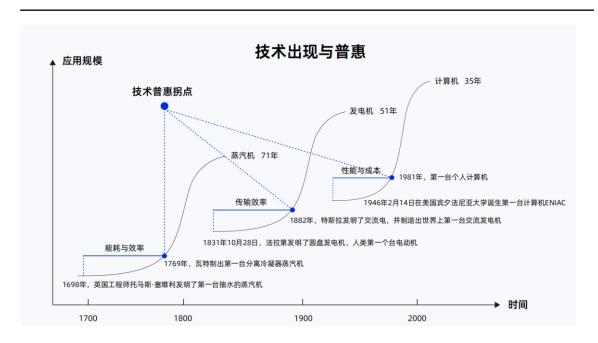
资料来源:中泰证券研究所



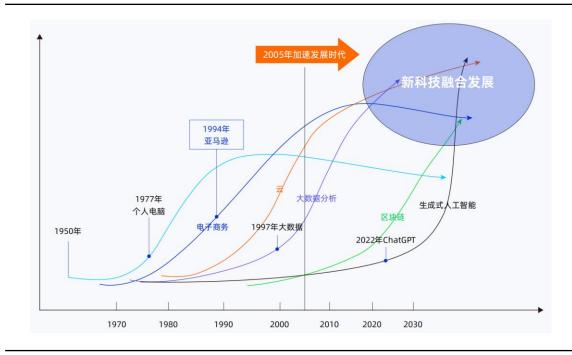
3.1 供给:云算力普惠化是助推AI产业革命的关键

- 回顾历史,从技术突破演进成产业革命,能否快速实现技术的普惠应用,是促成技术扩散和产业体系升级的胜负手。
- AI模型对功耗、内存和通讯的能力需求不断提升,倒逼基础设施厂商进行技术迭代。这种技术提升也将随规模化扩张 成为云厂商的竞争优势和壁垒。正如业内评论所言,没有人会在同一个数据中心训练同一个前沿模型两次。

图表: 历次产业革命都发端于关键技术跨过普惠拐点



图表: 低门槛的技术产品和服务汇聚开启智能革命序幕



资料来源: 阿里巴巴, 中泰证券研究所

资料来源: 阿里巴巴, 中泰证券研究所



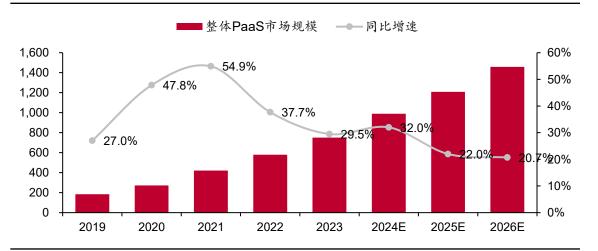
3.1.1 供给:智能化趋势下, laaS+PaaS呈现差异化的服务模式

- 2023年中国laaS市场规模为3121亿元,增速为17.1%; PaaS市场规模为750亿元,增速为29.5%。随着以大模型为代表的相关人工智能技术的发展,有望驱动基础云市场进一步形成"PaaS带动laaS"的联动效应,成为基础云市场新的增长引擎。
- laaS市场: 国产化与智能化,推动基础设施层减弱同质化。相较于传统laaS层在基础算力领域,以价格多样化和算力精细化的商业模式来塑造差异。在大模型时代,政企客户对自研产品和智能算力质量需求显著提升,逐步成为云厂商升级底层laaS能力的重点。
- PaaS市场:在大模型时代,基础云厂商偏重于"基础大模型"与"行业大模型"双管齐下的发展模式。在基础大模型领域,云厂商一方面不断提升自研大模型产品能力,并推出 MaaS概念,构建全链路模型产品和服务能力。在行业大模型领域,云厂商则依托其自研基础大模型,结合行业客户服务经验,生态伙伴能力,聚焦典型业务场景,"以点带面"逐步探索大模型商业落地。

图表: 2019-2026年中国整体laaS市场规模及增速预测(单位: 亿元)



图表: 2019-2026年中国整体PaaS市场规模及增速预测(单位: 亿元)



资料来源:赛迪顾问,中泰证券研究所

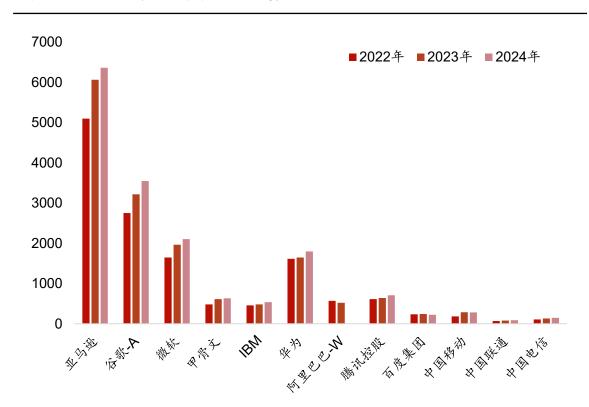
资料来源:赛迪顾问,中泰证券研究所



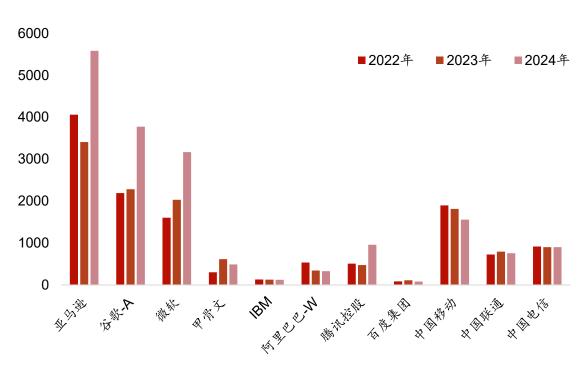
3.1.1 供给:海外厂商CapEx军备竞赛,国内互联网、运营商加大投入

■ 海外云计算巨头亚马逊、谷歌、微软正在进行CapEx投入的军备竞赛。寡头垄断的竞争格局使各家厂商的CapEx投入 变成了博弈论问题,竞争激烈。国内腾讯、阿里等互联网大厂和移动、联通、电信等运营商也在加大资本支出投入。

图表: 2022-2024年各云计算大厂研发费用(亿元)



图表: 2022-2024年各云计算大厂资本支出(亿元)

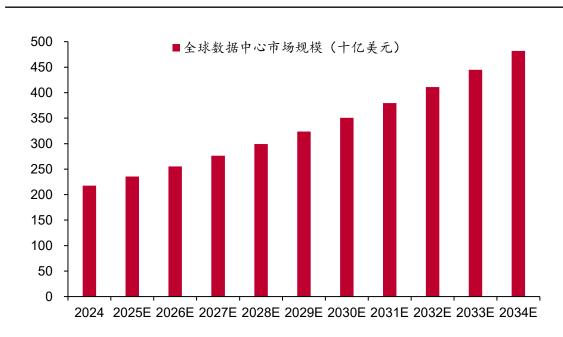




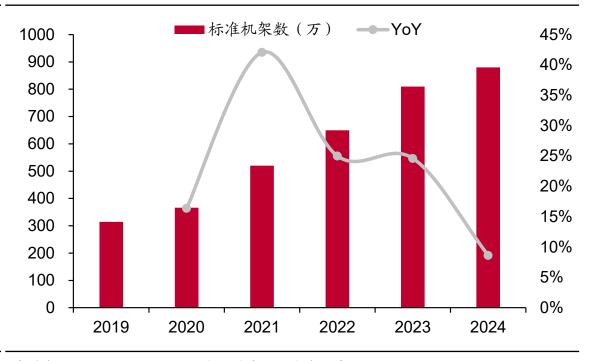
3.1.1 供给:全球数据中心快速扩容,我国正加速建设

- 全球数据中心市场规模不断扩容。据Cervicorn Consulting预测,2025年全球数据中心市场规模预计为2356.5亿美元,预计到2034年将达到约4817.3亿美元,2025年至2034年的复合年增长率为11.4%。
- 我国数据中心建设不断加速。统计显示,截至2024年底全国在用算力中心机架总规模超过880万标准机架,有效承载人工智能、大模型等高密度算力需求。

图表: 2024-2034E全球数据中心市场规模



图表: 2019-2024年我国数据中心标准机架数(万)及同比增长





3.1.1 供给:端侧推理芯片能力增强,端侧模型加速落地

- 端侧 AI 芯片不仅需要满足高算力需求,还要在有限的功耗和散热条件下提供稳定性能。端侧AI不仅让硬件设备能更快响应用户指令,还能在处理敏感数据时保持隐私安全。
- 当下各大厂商都在推出适用于终端AI推理的芯片平台, Meta、Google、Microsoft、DeepSeek、阿里等大厂也纷纷推出8B以下的模型。

图表: 各大厂部分端侧模型

模型名称	公司	规模		
Phi-4 mini	Microsoft	3.8B		
Llama 3.2	Meta	1B/3B		
Gemma 3	Google	1B/4B		
Qwen3	阿里	0.6B/1.7B/4B/8B		
Qwen3 (Reasoning)	阿里	0.6B/1.7B/4B/8B		
DeepSeek R1 Distill Qwen 1.5B	DeepSeek	1.5B		
DeepSeek R1 Distill Llama 8B	DeepSeek	8B		

图表: 各大厂端侧芯片

芯片	企业	应用	
Jetson Orin Nano	NVIDIA	机器人	
Smapdragon X Elite	Qualcomm	PC	
天玑 9300	MTK	手机、平板	
锐龙8040	AMD	PC、平板	
Lunar Lake	Intel	PC、平板	
J5	地平线	无人车	
M4	Apple	平板、PC	

资料来源: Artificial Analysis, 中泰证券研究所

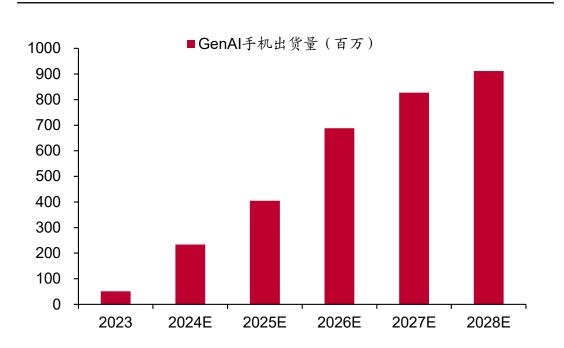
资料来源: 云天励飞, 中泰证券研究所



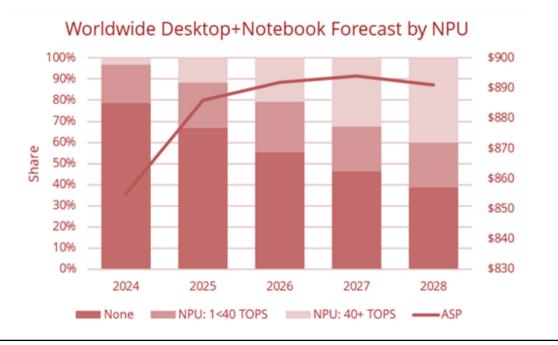
3.2.1 供给: 消费级终端计算设备能力提升, 边缘计算载体升级

■ 边缘计算趋势下,消费设备也将随之升级到人工智能PC/智能手机/新形态设备(例如Meta眼镜、Humane Pin、Rabbit R1)。 **GenAl手机出货量预计将快速提升**。Canalys预计,2024年全球智能手机出货总量中约5%会搭载端侧Al运算能力。IDC预计,GenAl智能手机市场出货量将在2024年达到2.342亿部,较2023年同比增长363.6%。 PC依旧是最核心的生产力工具,随着端侧模型性能提升和处理器芯片迭代,Al PC出货量、渗透率都将快速提升。

图表: 2023-2028E年GenAl手机出货量预测



图表: 2024-2028年AI PC预测



资料来源: IDC, 中泰证券研究所



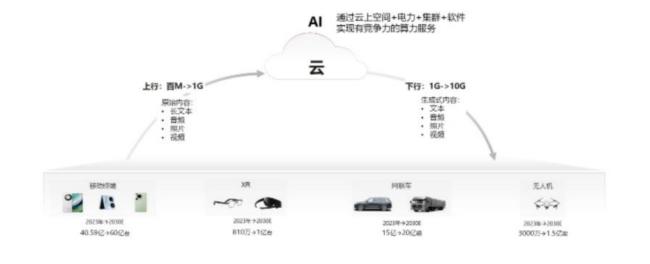
3.1.1 供给:终端算力增强、通信基础设施完备,端云协同趋势明确

■ 端云协同应用将迎来爆发。随着终端设备智能化程度的提升,XR/自动驾驶超低时延场景的普及,终端将产生巨大的计算和功耗需求。端云协同可以释放终端算力、电力压力,百倍增强端侧智能,实现极致低时延,驱动端云协同应用的爆发。据IDC统计,2024上半年,中国边缘云市场规模总计54.8亿元人民币,同比增速达到21.0%。

图表:基于云边端协同的物联网边缘计算框架



图表:端云协同示意



资料来源: 信通院, 中泰证券研究所

资料来源: 华为云计算2030. 中泰证券研究所

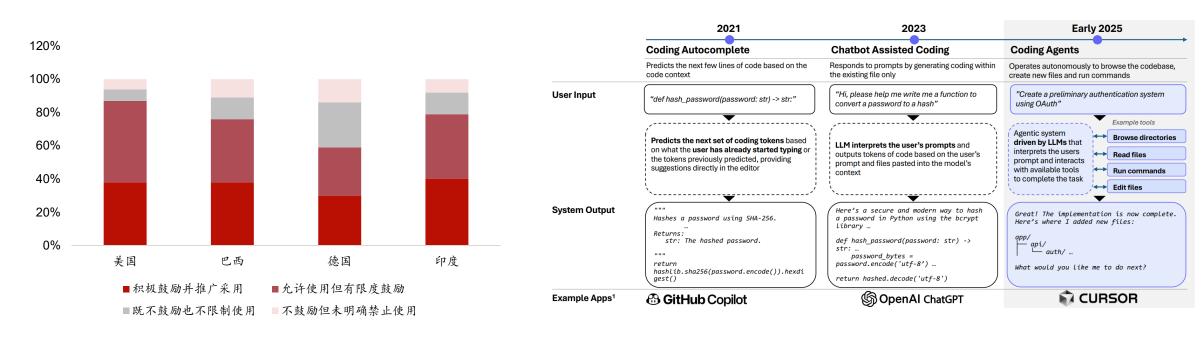


3.1.2 供给: AI原生趋势下SaaS能力涌现, 低成本定制化交付成为可能

■ AI原生趋势下软件开发能力大幅提升, SaaS产品有望解决低成本定制化开发的最大痛点。SaaS软件是标准化的服务, 但企业往往需求定制化的产品。一方面AI工具可以赋能开发, 加速软件迭代和交付; 另一方面有通用能力的Agent能力持续提升, 甚至走向通用, 能够满足不同的定制化需求。这些都将为SaaS厂商带来新的增长空间。

图表: 各国企业AI Coding工具使用调研结果

图表:Coding Agents演变趋势



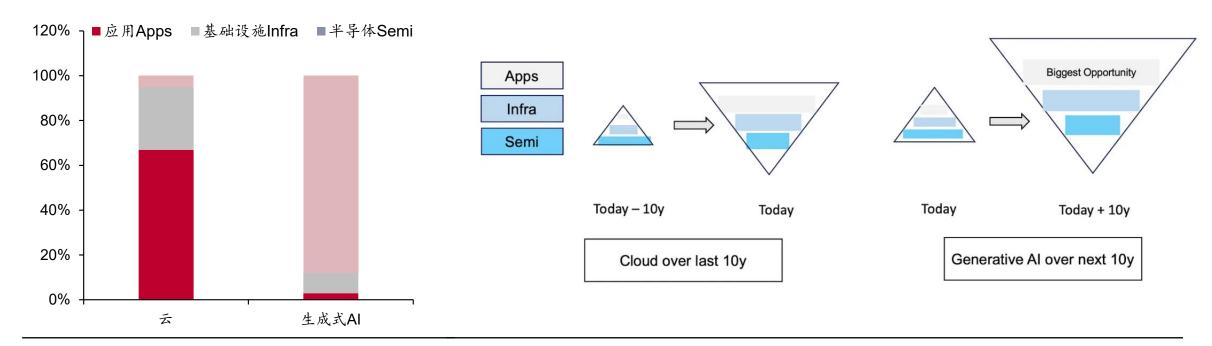


3.2.1 供给:未来应用层(SaaS Apps)价值链占比将提升

- 当下产业正处于转型的早期阶段,半导体(如NVIDIA)占据了产业链上大部分价值。
- 可以预计AI时代未来通用人工智能目前的倒金字塔形收入结构不会永远保持下去,应用层在价值链上的占比将逐渐提升,SaaS应用有望迎来爆发期。

图表:云和生成式AI各层公司毛利润占比

图表:未来10年AI Stack将如何发展?



资料来源: Altimeter, 中泰证券研究所

资料来源: Internal Analysis, 中泰证券研究所



3.2 需求: IT支出不断提升,数据中心增速最快

■ 根据Gartner的最新预测,2025年全球IT支出预计将达到5.62万亿美元,比2024年增长9.8%。中国IT支出2024年预计达到5890亿美元,同比增长8.5%,其中数据中心同比增长预计将达40%。

图表:全球2024-2025E年IT支出拆分预测

图表:中国2023-2024E年IT支出拆分预测

	2024年支出(亿 美元)	2024年增长率 (%)	2025年支出(亿 美元)	2025年增长率 (%)		2023年支出(亿 美元)	2023年增长率 (%)	2024年支出(亿 美元)	2024年增长率 (%)
数据中心系统	3291.3	39.4%	4055.1	23.2%	数据中心系统	518.8	17.3%	720.5	40.0%
设备	7341.6	6.0%	8102.3	10.4%	设备	1432.7	-1.6%	1455.1	2.4%
软件	10915.7	12.0%	12468.4	14.2%	软件	399.7	16.4%	459.3	15.8%
IT服务	15881.2	5.6%	17314.7	9.0%	IT 服务	996.0	9.6%	1105.9	11.9%
通讯服务	13717.9	2.3%	14237.5	3.8%	通信服务	2124.0	3.5%	2149.2	2.0%
总计IT支出	51147.7	7.7%	56178.0	9.8%	总计IT支出	5471.2	5.1%	5890.0	8.5%

资料来源: Gartner, 中泰证券研究所(2025年1月)

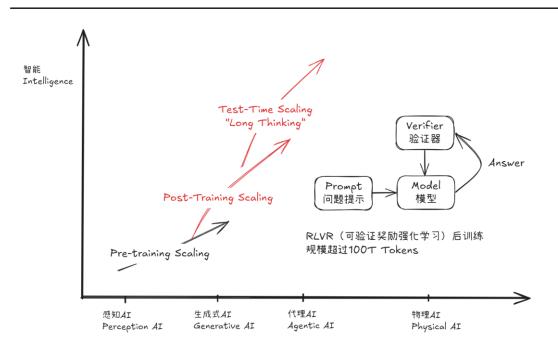
资料来源: Gartner, 中泰证券研究所(2024年7月)



3.2.1 GenAI驱动需求:行业走入后训练和长思考阶段,训练算力需求不断提升

■ 大语言模型对训练算力的需求不断上升,直接带来算力集群的扩容需求。CES 2025上黄仁勋提出了Scaling Law的三 个阶段: 预训练, 即人工智能模型从大量数据中学习模式的初始训练阶段; 后训练, 即使用人工反馈等方法微调人工 智能模型的答案;以及推理计算,让人工智能模型在回答每个问题后有更多时间"思考"。产业正在迎来后训练和场 推理的Scaling Law阶段,再次提升对算力的需求。

图表: NVIDIA提出的三阶段Scaling Law



图表: 2018年至今部分大模型的训练算力规模



资料来源: epoch AI, 中泰证券研究所



3.2.1 GenAl驱动需求: Al为云计算软硬件带来性能和效率的多重挑战

■ AI对单卡和算力集群的功耗、内存传输速度、通讯网络带宽提出了更多挑战。

图表: 计算视角下AI对计算体系结构带来挑战

突破功耗墙

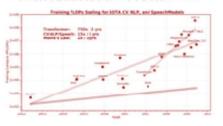
2017年: V100单卡功耗<250w 2020年: A100单卡功耗<400w 2023年: H800单卡功耗<750w

2026年:??

随着算力的成倍的增加,芯片制程与散

热问题将面临挑战

大语言模型算力在2年间增长750x



突破内存墙

AI大语言模型在显存受限的情况下,对内存带宽提出了极高要求。

PCIe Gen5的带宽已经逐渐成为计算密集

场景下GPU与内存交换的瓶颈。

CPU-GPU 联合架构成为趋势



NVIDIA: Grace-Hoper NVLink®-C2C 提供 900GB/s, 7x Faster PCle Gen5 带宽

Intel: Falcon Shores CPU-GPU Combo

Architecture (XPU)

AMD: MI300 CPU-GPU Combo

Architecture (APU)

突破通讯墙

AI分布式训练千卡规模网络带宽主流规模: A100 - 800Gbps、H100- 3.2Tbps

高性能网络与通信库成为分布式训练的关

键路径

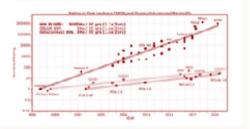
网络拥塞控制

多轨道/多路径能力

通讯库算子优化



通讯带宽的增长远远落后与算力的增长



资料来源: 阿里巴巴, 中泰证券研究所



3.2.1 GenAl驱动需求:顶尖Al模型基于公有云算力集群训练

■ 基础大模型的性能决定AI应用的水平,模型训练要求高性能、稳定和普惠的并行计算支持,全球领先的基础大模型优先基于公有云完成训练。在公有云上部署大模型,不仅能让用户及时获取大模型的最新功能和应用,还可以通过其PaaS(Platform as a Service,平台即服务)层和SaaS(Software as a Service,软件即服务)层为用户提供丰富的工具、组件和应用,支撑千行百业智能化应用。

图表:部分基础大模型训练芯片规模及部署方式

云	基础大模型	GPU型号及数量	部署方式
OpenAl	GPT-4、GPT-4.5、GPT-5	2.2万张 A100、预计几万张H100卡、甚至10 万卡	微软公共云计算集群
亚马逊	Anthropic Claude	1.6万张H200	亚马逊公共云集群
谷歌	Gemini、Gemini Ultra	TPUv5(是GPT-4的5倍)	公共云,支持客户模型训练;跨多个数据中心的大量TPUv4;利用AWS的 云
Meta	LLaMA、LLaMA3	2048张A100(未公布,估计至少2万张A100)	拥有1.6万AI超级计算机集群,用于训练和推理,服务全球数十亿用户;预 计到2024年底,将拥有35万张H100,公司算力总水平相当于近60万块 H100
xAI	Grok 1	数万个GPU集群	据Techcrunch报道,可能是由Oracle提供(财报显示,Oracle是xAI的云服务商)

资料来源: 阿里巴巴, 中泰证券研究所

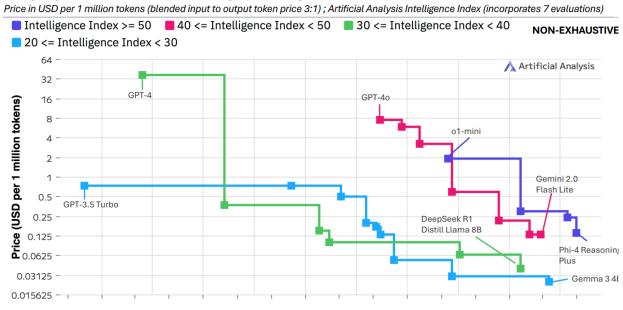


3.2.1 GenAl驱动需求:模型推理成本快速下降,推动应用生态繁荣

■ 随着模型小型化和智能化、推理效率提升以及新硬件迭代推动计算成本下降,推理成本显著降低。据Artificial Analysis统计,高智能模型(左图中智能指数>50)的推理成本自2024年9月起已下降约32倍;与原始GPT-4同水平的智能成本自2023年3月发布以来下降超1000倍。推理成本的快速下降有望推动生态应用的繁荣。

图表:语言模型推理成本快速下降

Language Model Inference Pricing by Intelligence Class, Over Time



Nov '22 Jan '23 Mar '23 May '23 Jul '23 Sep '23 Nov '23 Jan '24 Mar '24 May '24 Jul '24 Sep '24 Nov '24 Jan '25 Mar '25 May '25

Release Date



3.1.2 需求模式: Service-as-a-Software新商业模式将带来价值量再分配

■ Al为云厂商带来Service-as-a-Software的新模式,即服务作为软件产品。Agent等全新Al产品模式正在迭代,用户可以接受按照创造的价值去付费或分成。在这种模式下,laaS/PaaS层的规模化降本能够同时为云厂商和下游用户创造利润,形成生态下的共赢局面。从收入模型看,过去SaaS产品往往基于ARR和工具,按人、月、时间定价;未来可能会基于结果和效果,依节省成本、增加收入情况收费。

图表: 未来新商业模式的出现

	1950-70s 大型机时代	1980-2000s PC时代	2000s至今 云计算时代	未来 AI+SaaS时代
	定制开发应用	通用软件产品	软件即服 务 (SaaS)	服务即软件 (Software as a Service)
典型产品	1962年IBM与美国航空联合打造SABRE大型订票系统	~2000年,SAP ERP产 品,覆盖企业客户所有管理 流程	2001年推出第一款SaaS CRM产品	2024年发布Al Agent产品来提供线上客户体验服务
商业模式	卖定制服务	卖授权	卖服务	卖效果
	为大型企业开发定制软件,如SABRE用了200名程序员 开发了5年以上	每服务器/每客户一个"访问许可证",用户简单安装激活即可使用	订阅制,用户按月/年支付 费用以获得开箱即用的软件 服务	只有Al Agent独立完成问答任务、不需要员工介入时,才向企业收费

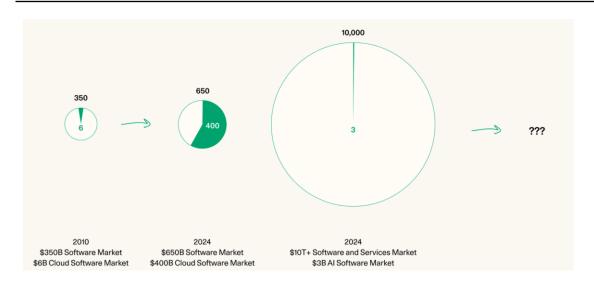
资料来源: 钉钉开放日, 中泰证券研究所



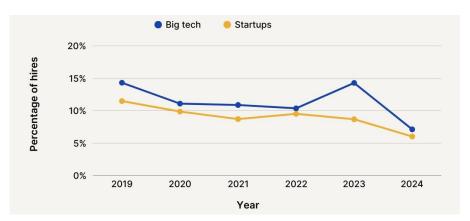
3.2.1 需求模式: Agent取代人力服务,中小企业SaaS市场或将爆发

- GenAl走向Agent时代,SaaS公司正在将服务转化为软件(Service-as-a-Software)。这意味着应用的目标市场不是软件市场,而是以数万亿美元计的服务市场。这种模式下,SaaS厂商为客户交付的内容从软件产品转向Agent等"数字人力"。过去交付服务等,乙方主要与甲方 IT 团队交互;如今 Al 原生企业、SaaS 服务商交付"数字员工",涉及多领域,类似人力资源外包(外包数字人),可能定义流程并创造价值。
- 最重要的是,这种价值创造得到下游客户一致的认同,软件行业下游付费意愿和规模不足的问题有望得到改善。当下AI的人力替代已经成为产业共识,国央企为代表的企业开始进行大规模实质性的AI投入;也有93%中小企业认同AI的价值创造,83%有计划在未来一年内投入AI。

图表: Service as a Software面向的市场规模



图表:科技公司招聘毕业生数量减少



2024年大型科技公司和初创公司招聘毕业生减少, AI的应用可能是原因之一。

资料来源: Sequoia Capital, 中泰证券研究所

资料来源: SignalFire, 中泰证券研究所



3.2.2 政策驱动需求: 云计算服务应用始终是国家政策的关注重点

■ 发展云计算、云服务始终是国家政策的关注重点。

图表: 云计算相关政策梳理

发布时间	发布单位	政策名称	主要内容
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》	打造技术领先优势。开展IPv6演进技术"揭榜挂帅"行动,积极推动新型网络体系、算网融合、真实源地址验证体系结构(SAVA)等下一代互联网关键技术研究。推动IPv6与5G、人工智能、云计算等技术的融合创新,支持企业加快应用感知网络、新型IPv6测量等"IPv6+"创新技术在各类网络环境和业务场景中的应用。
2023年11月	工信部	《关于健全中小企业公共服务体系的 指导意见》	创新服务方式。充分运用大数据、云计算、人工智能等技术,建立贯通国家、省、市、县各级中小企业公共服务机构的全国中小企业服务"一张网",与政务服务平台加强互联互通,为中小企业提供"一站式"服务。
2023年12月	工信部等八部门	《关于加快传统制造业转型升级的指 导意见》	大力推进企业智改数转网联。立足不同产业特点和差异化需求,加快人工智能、大数据、 云计算、5G、物联网等信息技术与制造全过程、全要素深度融合。
2024年4月	商务部	《数字商务三年行动计划(2024- 2026年)》	支持电商平台创新云计算、移动支付等数字化产品和服务,加强远程交付能力,开发海外服务市场。
2024年8月	中共中央办公厅 国务院 办公厅	《关于数字贸易改革创新发展的意见》	大力发展数字技术贸易。加强关键核心技术创新,加快发展通信、物联网、云计算、人工 智能、区块链、卫星导航等领域对外贸易。
2025年1月	金融监管总局 工业和信息化部 交通运输部 商务部	《关于深化改革加强监管促进新能源车险高质量发展的指导意见》	,鼓励保险行业积极运用大数据、区块链、云计算等技术,加快数字化、线上化、智能化转 `型升级,提高对新能源汽车的风险识别和精算定价能力,通过技术创新和优化业务流程推 进降本增效。
2025年3月	工信部等三部门	《关于促进环保装备制造业高质量发 展的若干意见》	提升传统环保装备高端化智能化绿色化水平。充分利用人工智能、大数据、云计算、工业 5 互联网等新一代信息技术,提升环保装备设计、生产、使用、运维等各环节数字化智能化 绿色化水平。
2025年4月	工信部	《云计算综合标准化体系建设指南(2025版)》(征求意见稿)	强化云计算标准化工作顶层设计,加强全产业链标准工作协同,统筹推进标准的制定、实施和国际化,为推动我国云计算产业高质量发展提供坚实的技术支撑。
2025年4月	工信部等七部门	《医药工业数智化转型实施方案 (2025—2030年)》	鼓励建设一批高性能云计算平台、区块链、数据中心、5G行业虚拟专网、物联网等信息基础设施,支撑医药企业"智改数转网联"。

资料来源:各部门官网,中泰证券研究所





行业云正成为企业建云新方式

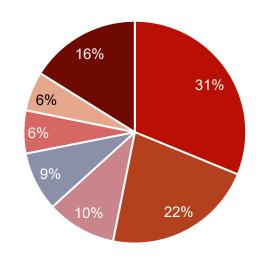


4.1 "云+行业"将成为建云新模式,行业模型带来云服务差异化

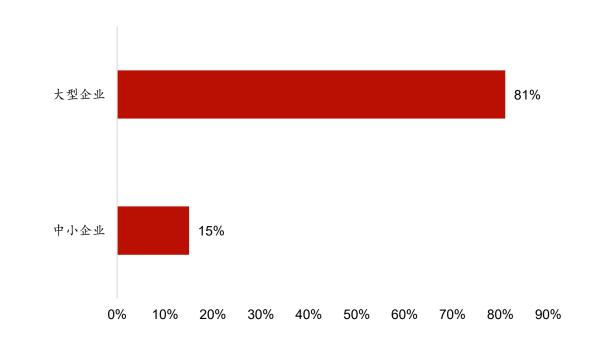
■ "云+行业"方面,行业云平台基于行业技术方案构建,具有共享化、场景化和市场化特征,成为大型企业建云新模式。Gartner在2023年、2024年均将行业云平台列为十大战略技术趋势之一,并预测到2027年超过70%的企业将采用行业云平台加速业务创新。

图表: 2023年行业用云量占比

图表: 2023年不同规模企业上云率



■泛互联网、信息化服务业 ■政务 ■ 金融 ■ 工业制造 ■ 医疗 ■ 交通 ■ 其他



资料来源:信通院,中泰证券研究所

资料来源: 信通院, 中泰证券研究所



4.2 传统行业加速上云: 政务云

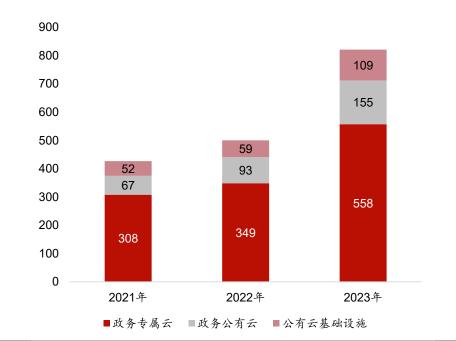
■ 政务行业正在加速上云,行业仍存较大增长空间。2023年政务云整体市场规模为794.5亿元人民币;政务专属云市场规模达557.6亿元人民币,其中专属云基础设施406.6亿元人民币,同比增长16.6%;政务公有云市场规模达154.7亿元人民币,公有云基础设施109.2亿人民币,同比增长18.2%;政务云运营服务市场规模为82.2亿元人民币,同比增长38.7%。据统计,我国政务系统上云率达90%以上的城市占比仅有55%左右,且仍有10%的城市政务系统上云率在50%以下,行业仍存在较大增长空间。

图表: 2023年中国政务云市场拆分

图表: 2021-2023年中国政务云市场规模(亿元)拆分







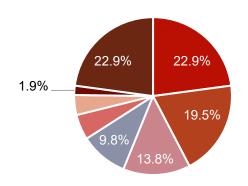
资料来源: IDC, 中泰证券研究所

4.2 政务云竞争格局: 因细分市场服务模式不同, 格局有较大差异

- 政务云竞争格局因服务模式不同有较大差异:
- 在政务专属云基础设施市场,华为、浪潮、新华三/紫光、中国电信、中国移动占比较高;在政务公有云基础设施市场 中,中国电信、华为、阿里、中国移动、浪潮位居前列;在政务云运营服务市场上,则是浪潮、中国移动、中国电信、 华为、中国联通占据优势。

图表: 2023年政务云细分市场竞争格局

中国政务专属云市场份额

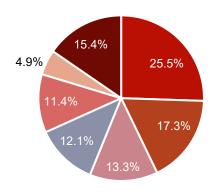


■华为 浪潮

■新华三/紫光 ■中国电信

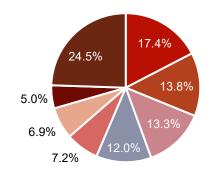
■中国联通 ■中国电子云 ■其他 ■中国移动

中国政务公有云市场份额



■中国电信 ■华为 ■阿里 ■中国移动 ■浪潮 ■中国联通 ■其

中国政务云运营服务市场份额



■ 浪潮

■中国移动

■中国电信 ■华为

■中国联通 ■金山

■新华三/紫光■其他

资料来源: IDC, 中泰证券研究所

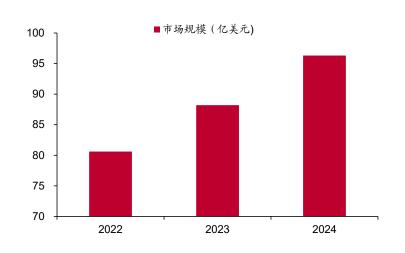


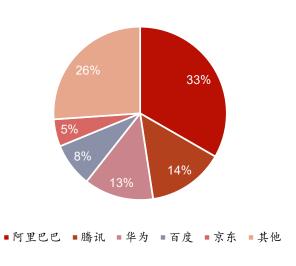
4.2 传统行业加速上云:金融云将迎"基础设施重构+智能应用爆发"双增长

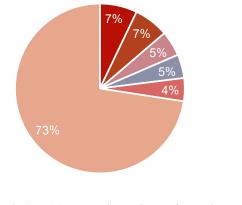
- 随着互联网巨头相继布局云计算,以及传统金融机构围绕新型技术的新一轮 IT 改革,加速了云计算在金融行业的应用 实践。越来越多金融企业开始应用云来承载应用和处理高并发业务,云计算技术正加快与金融行业发展融合。
- 据IDC统计,2024下半年,中国金融云市场规模达到52.3亿美元,同比增长11.0%,市场呈现多元化发展态势。金融企业云化支出比例明显提升,云化需求多样化发展。
- 2025年开始, DeepSeek等开源模型的入场为金融云市场发展的竞争格局引入了新的变化和商机,中国金融云蓄势待发,有望迎来新一轮的"基础设施重构+智能应用爆发"的双增长。

图表: 2022-2024年我国金融云市场规模情况

图表: 24H2中国金融云 (平台) 解决方案市占率情况 图表: 24H2中国金融云 (应用) 解决方案市占率情况







■中电金信 ■中科软科技 ■百度 ■南天信息 ■腾讯 ■其他

资料来源: IDC. 中泰证券研究所 资料来源: IDC. 中泰证券研究所

资料来源: IDC, 中泰证券研究所





部分相关公司梳理



5.1 深信服:全栈云服务厂商,企业云、专属云、桌面云

■ 深信服从2012年开始布局云计算业务,核心产品主要包括超融合、桌面云、分布式存储、托管云等云计算产品和解决方案等。信服云是深信服旗下的云计算品牌,致力于以领先的技术为用户提供超融合、分布式存储、私有云、桌面云和托管云等产品,以及更简单、更安全的解决方案和服务。信服云拥有广泛的生态合作伙伴,并与他们一起构建了软件定义的数据中心基础设施、私有云、托管云、桌面云、边缘计算等云-网-边-端架构,在优势的安全体系保障下,可以为用户提供不同数字化转型阶段的服务。

图表: 信服云架构全景



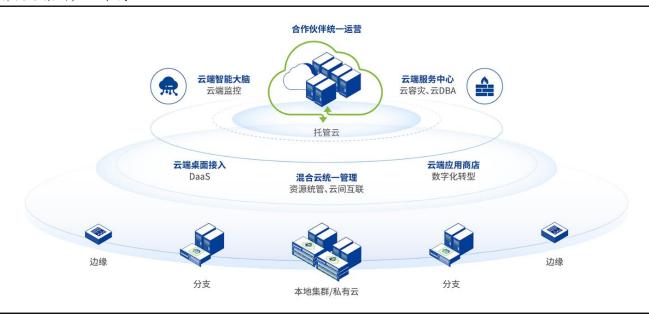
资料来源:深信服官网,中泰证券研究所 48



5.1 深信服: 未来云业务理念——线上线下一朵花

■ 深信服"同架构混合云"解决方案,底层基于架构一致的私有云(线下)和托管云(线上),实现线上线下统一管理、 统一监控运维。借助"云间互联"技术安全便捷地实现线上同一VPC与线下大二层互通,并支持国产X86与ARM集群 与线上互通;云上推出智能大脑,为线下的IT私有云提供远程监控和专家值守,协助用户闭环问题,提高本地数据中 心的可靠性;同时建设云端服务中心,为线下私有云用户提供云上的灾备服务、数据库服务、安全防护服务等等,把 复杂的IT运维交给云端专家,把简单高效留给客户。该方案广泛应用于"本地数据中心延伸"、"业务混合部署"、 "混合云灾备"等场景,为企业数字化转型构筑坚实 IT 底座。

图表:深信服"同架构混合云"解决方案的产品架构



资料来源:深信服官网,中泰证券研究所



5.1 深信服云计算相关业务营收持续增长

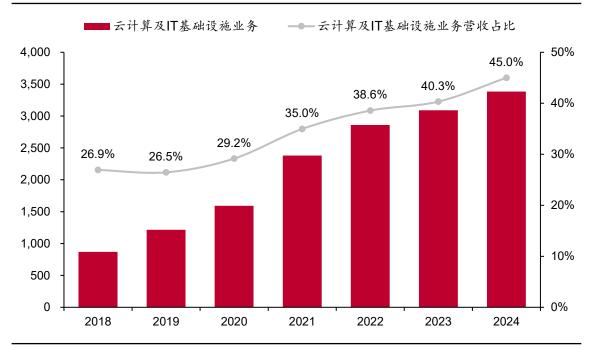
■ 2018年以来,深信服云计算及IT基础设施业务营收持续增长,占深信服总营收比重持续提升。2024年深信服总营收75.20亿元,同比略有下降,但云计算及IT基础设施业务营收达33.84亿元,占公司总营收45.0%,占比较2023年继续提升4.7pcts。

图表: 2015-2024年深信服营收与利润情况(单位: 百万元)



资料来源: Wind, 中泰证券研究所

图表: 2018-2024年深信服云计算业务营收及占总营收比重变化情况(单位: 百万元)



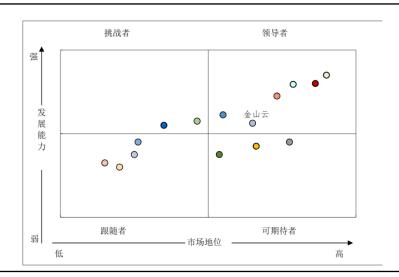
资料来源:深信服年报,中泰证券研究所



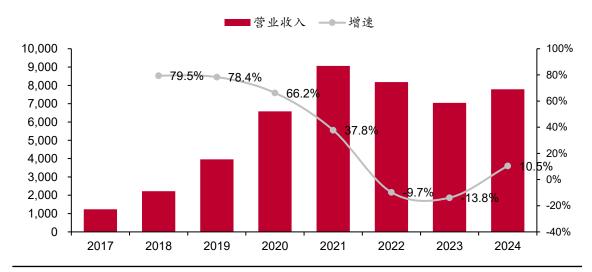
5.2 金山云:独立云服务厂商

- 金山云创立于2012年,作为中国知名的独立云服务商,业务范围遍及全球多个国家和地区。依托金山集团36年企业级服务经验,金山云坚持技术立业,逐步构建了完备的云计算基础架构和运营体系,并提供超过150种适用于互联网、公共服务、数字健康、金融等领域解决方案,为500+优质客户提供高品质云服务。
- ▶ 2023年起,金山云全面拥抱人工智能新时代,构建面向全行业人工智能的全栈公有云基础设施,提升高性能算力和网络、云原生基础设施、人工智能平台能力和行业应用能力的建设。
- ▶ 2024年,赛迪顾问发布《2024中国央国企云市场研究报告》,金山云成功跻身领导者象限。
- 根据公司2024年业绩报告,2024年金山云实现营业收入77.85亿元,同比增长10.5%,其中公有云服务收入50.07亿元,同比增长14.3%。截至2024年12月31日,金山云在国内共计拥有两个数据中心、约10.4万台服务器。

图表: 2023年中国央国企云"laaS+PaaS"市场竞争格局



图表: 2017-2024年金山云营收及增速(单位: 百万元)



资料来源: Wind, 中泰证券研究所

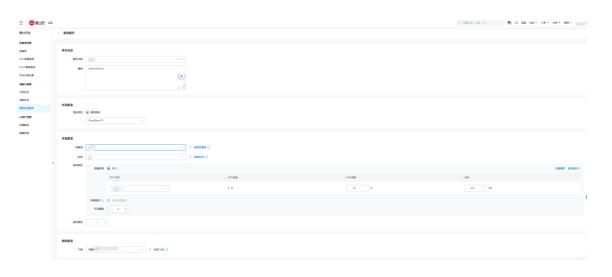


5.2 金山云:多场景支持Deepseek大模型

- 2025年2月,金山云宣布在公有云场景和国资云/政务云场景已支持DeepSeek-R1/V3。金山云已正式发布基于DeepSeek-R1蒸馏模型的多种镜像服务,适用于长文档分析、多轮对话、复杂代码项目理解等场景。本次金山云支持DeepSeek-R1/V3,亦提供较小尺寸的DeepSeek蒸馏版本模型适用于基础公有云、中小型企业内部实时部署,也方便个人用户快速搭建对话服务。
- 2025年3月,金山云宣布完成基于国产芯片的DeepSeek满血版适配,同时金山云星流训推平台也已完成DeepSeek满血版的部署,且支持多机部署实现分布式推理。

图表: 金山云国资云/政务云平台已正式上架 DeepSeek-R1和 DeepSeek-V3模型

 图表:金山云星流训推平台可多机部署





风险提示

- 云计算及AI技术落地不及预期的风险。从研发到实际应用的转化过程里,云计算和AI技术遭遇了诸多阻碍。软硬件技术开发、算法迭代、高质量数据集清晰以及应用场景的适配难题,可能阻碍云计算和AI技术迅速、有效地融入各个行业。
- **行业竞争加剧的风险**。越来越多的企业投身云基础设施领域,可能导致了人才的激烈争夺,存在技术研发投入不断攀升,市场同质化竞争加重的风险。
- 报告信息更新不及时的风险。存在信息更新不及时的风险。



重要声明

- 中泰证券股份有限公司(以下简称"本公司")具有中国证券监督管理委员会许可的证券投资咨询业务资格。 。本公司不会因接收人收到本报告而视其为客户。
- 本报告基于本公司及其研究人员认为可信的公开资料或实地调研资料,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。本公司力求但不保证这些信息的准确性和完整性,且本报告中的资料、意见、预测均反映报告初次公开发布时的判断,可能会随时调整。本公司对本报告所含信息可在不发出通知的情形下做出修改,投资者应当自行关注相应的更新或修改。本报告所载的资料、工具、意见、信息及推测只提供给客户作参考之用,不构成任何投资、法律、会计或税务的最终操作建议,本公司不就报告中的内容对最终操作建议做出任何担保。本报告中所指的投资及服务可能不适合个别客户,不构成客户私人咨询建议。
- 市场有风险,投资需谨慎。在任何情况下,本公司不对任何人因使用本报告中的任何内容所引致的任何损失负任何责任。
- 投资者应注意,在法律允许的情况下,本公司及其本公司的关联机构可能会持有报告中涉及的公司所发行的证券并进行交易,并可能为这些公司正在提供或争取提供投资银行、财务顾问和金融产品等各种金融服务。本公司及其本公司的关联机构或个人可能在本报告公开发布之前已经使用或了解其中的信息。
- 本报告版权归"中泰证券股份有限公司"所有。事先未经本公司书面授权,任何机构和个人,不得对本报告进行任何形式的翻版、发布、复制、转载、刊登、篡改,且不得对本报告进行有悖原意的删节或修改。