OCTC AB01—2024

开放计算标准工作委员会
Open Compute Technology Committee

算力工厂建设指南

白皮书

（2024年6月）

开放计算标准工作委员会　发布

开放计算标准
工作委员会
Open Compute Technology
Committee

版权保护文件

版权所有归属于该文件的发布机构，除非有其他规定，否则未经许可，此发行物及其章节不得以其他形式或任何手段进行复制、再版或使用，包括电子版，影印件，或发布在互联网及内部网络等。使用许可可于发布机构获取。

目　　次

前　　言 III
1 概述 1
2 背景 1
2.1 算力发展趋势与挑战 1
2.2 预制化数据中心发展 4
3 算力工厂概述 5
3.1 算力工厂模式定义 5
3.2 算力工厂形态分类 6
3.3 算力工厂核心理念 6
3.4 算力工厂模式优势 7
3.5 算力工厂应用场景 7
3.6 算力工厂收益分析 8
4 算力工厂建设模式 8
4.1 建设模式分类 8
4.2 建设模式推荐 9
5 算力工厂规划建议 10
5.1 算力形式 10
5.2 算力规模 11
5.3 算力应用 11
6 算力工厂建设思路 14
6.1 建设原则 14
6.2 建设流程 15
6.3 算力工厂技术框架 16
7 算力工厂运营模式 25
7.1 算力运营 25
7.2 算力销售 27
7.3 日常运维 27
参 考 文 献 29

开放计算标准
工作委员会
Open Compute Technology
Committee

前　　言

本文件由中国电子技术标准化研究院提出。

本文件由开放计算标准工作委员会归口。

本文件起草单位：中国电子技术标准化研究院、浪潮电子信息产业股份有限公司、中通服咨询设计研究院有限公司、中国建筑西北设计研究院有限公司、西安工程大学、东南大学、西安交通大学、广东新会中集特种运输设备有限公司、金华莘水智能科技有限公司。

本文件主要起草人：张群、陈海、周相峰、朱亮、王涛、刘晓蕾、乌宏亮、朱智国、师宇清、张立功、赵超、李丛洋、张振宇、段晓丽、王太峰、戴蔚、赵明明、孙铁柱、陈忠英、李楠、张水权、盛有海。

开放计算标准
工作委员会
Open Compute Technology
Committee

算力工厂建设指南

1 概述

随着数字经济时代的全面开启，算力已经成为推动各行各业数字化转型的关键力量。在这个时代背景下，我国对于算力基础设施的投入日益加大，算力规模稳步增长，截至2023年底，我国算力总规模达到230 EFLOPS（FLOPS：每秒浮点运算次数）即每秒能完成230百亿亿次浮点运算，算力总规模居全球第二位。

然而，在算力规模高速增长的同时，也面临着诸多挑战。建设周期长、能源利用率低、技术更新换代快、管理运维效率低下以及建设投资高等问题，都在一定程度上制约了算力的进一步发展。为了解决这些问题，急需一种全新的建设模式，以更高效、更经济、更环保的方式推动算力的发展。

算力工厂正是在这样的背景下应运而生。采用预制化集装箱数据中心作为算力底座，相比传统数据中心，这一创新模式能够提升50%以上的交付效率。同时，算力工厂还搭载了计算、存储、网络等IT设备组成的硬件，对外提供算力资源的运营服务。这种模式的出现，不仅极大地缩短了建设周期，降低了建设成本，还提高了能源利用效率和管理运维效率。

算力工厂的核心理念是“打造创新的算力全生命周期服务模式，让算力以更合适的方式更快发生在更亟需的地方”。这一理念贯穿了算力工厂的整个建设过程，从规划、建设到对外提供算力运营服务，为用户提供了一体化全生命周期的服务。算力工厂致力于帮助用户以最快的模式搭建对外提供算力运营服务的平台，从而提高资金使用效率，改善用户整个生命周期的TCO（TCO：总拥有成本）和现金流，助力客户商业成功。

在算力工厂的建设过程中，注重技术创新和模式创新。通过引入先进的IT设备和数据中心技术，不断提升算力工厂的性能和效率。同时积极探索新的商业模式和运营模式，为用户提供更加灵活、多样化的服务选择。

展望未来，算力工厂将在数字经济中发挥越来越重要的作用。算力工厂将秉承创新、高效、环保的理念，不断优化和完善建设和服务模式。在算力工厂建设模式的推动下，我国的算力产业将迎来更加广阔的发展前景。

开放计算标准
工作委员会
Open Compute Technology
Committee

2 背景

2.1 算力发展趋势与挑战

2.1.1 算力的定义与价值

《中国算力白皮书（2022年）》将算力定义为：数据中心的服务器通过对数据进行处理后实现结果输出的一种能力。狭义的概念上，算力是软硬件配合执行某种信息处理需求的能力，一般采用每秒浮点运算次数进行表达，即FLOPS。广义概念上，算力是集信息计算力、网络运载力、数据存储力于一体的新型生产力，主要通过算力中心等算力基础设施向社会提供服务。

在人类历史的演进中，技术革命总是推动着时代的进步。第一次工业革命让人类进入热力时代，第二次工业革命则引领人类步入电力时代。而今天，随着计算机技术的飞速发展，人类迎来了由算力主宰

的第三次工业革命。这一时代变革，不仅标志着人类生产力的又一次飞跃，更预示着全球经济的崭新格局。

随着新一轮科技革命与产业革命的加速演进，算力正成为推动产业数字化转型的核心力量。它如同新时代的引擎，为各行各业的升级换代注入源源不断的新动能。在全球数字经济的浪潮中，算力不仅是发展的加速器，更是国家竞争力的重要标志。越来越多的数据显示，一个国家的算力规模与其经济发展水平之间存在着显著的正相关关系。

在数字经济时代全面开启的背景下，人工智能、云计算、大数据技术等日新月异，传统产业与新兴技术的融合日益加深，数字行业蓬勃发展。算力，作为这些技术的核心驱动力，正在以新的关键生产力形式，重塑着世界经济版图。根据IDC、浪潮信息、清华大学全球产业研究院联合发布的数据，计算力指数的提升与数字经济和GDP的增长呈正相关关系。计算力指数每提高1点，数字经济和GDP将分别增长3.5‰和1.8‰。这一数据充分证明了算力在推动经济增长中的重要作用。

对于中国而言，做强做优做大数字经济，促进数字经济与实体经济融合发展，已成为推动经济高质量发展的重要战略。在这一过程中，筑牢算力基础设施的坚实底座至关重要。只有拥有强大的算力支撑，才能在全球数字经济竞争中占据有利地位，为经济的高质量发展注入新动力。因此，必须加强算力基础设施的建设，推动算力与各行各业的深度融合，为数字经济的蓬勃发展提供坚实保障。

2.1.2 我国算力产业高速发展

随着 5G、人工智能等信息技术的发展，算力需求不断增长，算力产业迎来重要发展机遇。我国不断加大对计算、存储和网络等基础设施的投入，高度重视数据中心、智算中心、超算中心以及边缘数据中心等算力基础设施的高质量发展。

工信部数据显示，截至2023年底，我国在用数据中心机架总规模超过810万标准机架，算力总规模达到了230 EFLOPS，算力总规模居全球第二位。

2023年，人工智能领域的发展达到了重要转折点，在人工智能席卷各个行业的大趋势下，各行业正加速从业务数字化迈向业务智能化，从感知智能到生成式智能，人工智能算力需求快速增长。大模型和生成式人工智能的发展显著拉动了智能算力的增长，为算力产业带来了更大的发展空间。

在适度超前的指导思想下，国家正加大对人工智能算力基础设施的投资。目前，互联网企业、电信运营商，以及各级政府均积极投入到智算中心的建设之中。据《2023-2024 年中国人工智能计算力发展评估报告》中中国智能算力规模测算结果，预计到2027年中国智能算力规模将达1117.4 EFLOPS。同时，该报告测算了中国通用算力规模，预计到2027年通用算力规模将达到117.3 EFLOPS。2022 -2027年期间，中国智能算力规模年复合增长率达33.9%，同期通用算力规模年复合增长率为16.6%。这些数据充分表明，在未来几年内，我国算力基础设施建设将更加注重智算中心的建设，智算产业的发展也将进入一个黄金发展期。这不仅将有力推动人工智能等前沿技术的发展，还将为我国经济的数字化转型和高质量发展提供强大支撑。

2.1.3 算力发展趋势

1）算力需求持续增长

在数字经济时代，算力崭露头角成为全新的生产力，与数据、算法并肩成为这一时代最基础的生产要素。随着云计算、大数据技术的迅猛发展，数字化改革的步伐不断加快，数据呈现爆炸性增长，算法的复杂程度也持续攀升。这种趋势直接促使了对算力规模与能力的迫切需求，算力需求正以前所未有的速度爆发式增长，成为推动数字经济持续繁荣的关键力量。

2）算力类型加速转变

当前的算力领域正经历着一场算力类型的加速转变。传统上，通用算力在算力需求中占据着主导地位，然而，随着AIGC大模型等人工智能技术的迅猛发展，智能算力的占比开始迅速攀升。根据中国信通院发布的《中国综合算力指数（2023年）》报告显示，尽管通用算力目前仍占据74%的算力规模，但智能算力已经以惊人的速度增长，占比达到了25%。更值得注意的是，智能算力的增速同比上年增加了45%，这一增速甚至超过了总体算力的增速。

这一转变不仅体现了人工智能技术在现代社会中的广泛应用和重要性，也预示着未来算力领域的发展方向。随着技术的不断进步和应用的不断扩展，智能算力将继续保持高速增长，成为推动数字经济发展的重要力量。同时，这也对算力基础设施的建设和运营提出了更高的要求，需要不断创新和完善，以满足日益增长的智能算力需求。

3）算力服务泛在流动

云计算崛起之后，算力开始云化，分布化。边缘计算出现，算力还从云端下沉到通信网络的各个层级，遍布于云管端的各个角落。人工智能算力提升了其覆盖规模，实现生成式人工智能推理能力在边缘、终端等位置的部署和应用。随着边缘计算逐步进入稳健发展期，边缘计算与云计算、5G、区块链等其他技术的协同与融合需求进一步增加，推进人工智能在云-边-端的覆盖，满足无处不在的智能化需求。

4）算力设施绿色低碳

伴随人工智能对计算和存储能力要求的不断提升，芯片的功耗正越来越高，发热量也越来越大。在中国，面对大量涌现的人工智能大模型应用以及国家“双碳战略”和“东数西算”的逐步实施，为实现算力建设和能源消耗成本间的有效平衡，对算力基础设施进行功耗控制和改良，是从源头上进行节能减排的有效手段。通过积极探索模块化设计和部署，通过提升散热效率来降低能耗的液冷数据中心受到市场的关注。此外，提高可再生能源的占比，减少化石能源的使用，有助于降低电力成本，从而间接推动算力的绿色发展。

2.1.4 算力发展挑战

开放计算标准
工作委员会
Open Compute Technology
Committee

我国算力基础设施发展仍有提升空间。整体来看，以人工智能大模型为代表的AIGC新应用、新需求的崛起，推动算力规模快速增长、计算技术多元创新、产业格局加速重构。大模型产业井喷式发展也带来算力紧缺、能耗激增等问题。当前我国算力发展仍面临不小挑战：

1）建设周期长

数据中心的建设需要经过设计、土建、机电安装、调试等多个阶段，整个建设周期长达1～2年以上，这难以满足各行业对算力持续高速增长的迫切需求。

2）能源利用率低，能耗居高不下

算力带来的能耗问题日益显现，据中国通信院数据显示2022年中国数据中心平均PUE为1.52，而国家及多地政府出台PUE值不高于1.25的严格限制，如何实现绿色低碳发展，降低算力的能耗，成为整个行业的重点研究方向。

3）传统机柜供电制冷难以匹配

随着GPU服务器等高性能设备的广泛应用，每台机架的功率可达40kW以上，远超过传统数据中心每机架7kW的功耗水平。这使得数据中心的供电、制冷系统需要全面升级改造，以适应新的算力需求。

4）管理运维效率低下

当前的数据中心通常存在两套独立的监控系统，即动力环境基础设施监控和IT设备监控。这种分离的系统架构使得运维人员难以根据IT设备的实际负载情况进行及时调优，导致整个数据中心的监控运维管理、能耗调优和故障预警效率低下。

5）算力技术更新换代快

随着技术的不断发展，新的算力技术不断涌现，传统数据中心改造扩容难，无法第一时间升级新一代技术，面临建成即落后的窘境。

6）算力需求多维爆炸

算力正在从以DC为中心走向无处不在，走向边缘/超边缘/端，更有效地走进千行百业，这需要构建更加灵活敏捷的算力底座，实现泛在多维立体的算力分布。

7）算力建设的高投资

算力基础设施属于重资产和资本密集型行业，具有前期投入大、技术迭代快、建设门槛高等特点，其建设运营需要消耗巨大的时间成本和资金成本，远远超出了中小微企业的承受范围。

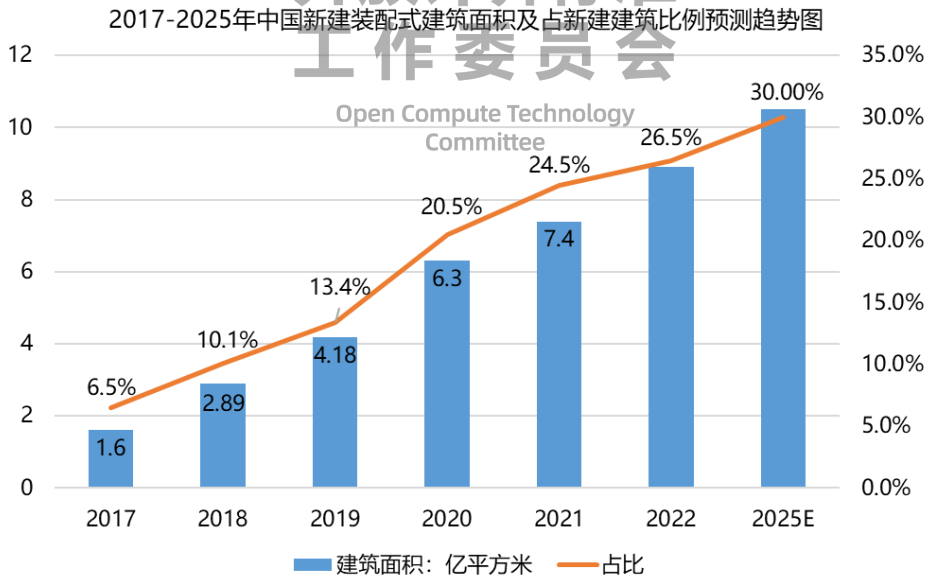
2.2 预制化数据中心发展

预制化数据中心融合了装配式建筑和智能模块化数据中心技术。装配式建筑是指把传统建造方式中的大量现场作业工作转移到工厂进行，建筑所用构件和配件在工厂预制，运输到建筑施工现场，通过可靠的连接方式在现场装配安装而成的建筑。装配式建筑主要包括预制装配式混凝土结构、钢结构、现代木结构建筑等。

装配式建筑符合建筑业产业现代化、智能化、绿色化的发展方向。近几年，一系列政策的颁布加快了我国装配式建筑行业的发展。2016 年是中国装配式建筑开局之年，国务院办公厅《关于大力发展装配式建筑的指导意见》（国办发〔2016〕71 号）明确："推动建造方式创新，大力发展装配式混凝土建筑和钢结构建筑"。2022 年 4 月，《关于进一步释放消费潜力促进消费持续恢复的意见》指出，"推动绿色建筑规模化发展，大力发展装配式建筑"。

受国家政策鼓励发展装配式建筑影响，我国装配式建筑规模持续快速增长。住建部数据显示，2021年全国新开工装配式建筑面积达7.4亿平方米，较2020年增长18%，占新建建筑面积的比例为24.5%。根据住房和城乡建设部发布的《"十四五"建筑业发展规划》，提出到2025年，装配式建筑占新建建筑的比例达30%以上。

开放计算标准
工作委员会
Open Compute Technology
Committee



来源：住建部

图1 2017-2025年中国新建装配式建筑面积及占新建建筑比例预测趋势

近年来，国家和地方政府出台多项政策和意见，指导数据中心的发展和建设。快速上线、绿色低碳、灵活部署等成为新时代数据中心建设的刚性需求，但是传统数据中心建设方案面临建设周期长、施工不

确定性高、能耗高、扩容难等多重挑战，无法满足新时代数据中心的发展要求。而预制化数据中心具有建设周期、质量、性能及能耗等多重优势，成为数据中心发展的重要方向。

同时多个地区对于数据中心采用装配式建筑出台了相应的政策要求，其中东部地区应用较为突出。北京要求数据中心建筑面积大于 5000 ㎡需按照装配式建筑实施；上海要求数据中心项目各幢建筑面积总和大于 10000 ㎡，需按照装配式建筑实施；海南要求数据中心需按照装配式建筑实施；浙江虽未对新建（数据中心）项目强制性采用装配式建筑，但处于逐步推广期。

数据中心预制化技术已有多年发展历史，早期预制化数据中心采用All In One（All In One：单个集装箱体内一体化集成数据中心各子系统）形式设计，满足小规模数据中心快速部署及应急建设要求。在All In One基础上，预制化数据中心逐渐实现了设备区和配电区等核心区域的模块化，但受到可靠性、空间及标准化程度等因素制约，仍为小规模及特定场景的临建级应用为主。随着装配式建筑技术与模块化数据中心融合程度的加深，预制化数据中心呈现主体结构建筑化、空间及内外使用体验楼宇化、功能区域标准化、扩容模块化等趋势，实现高等级、多楼层、大规模集群应用。预制化数据中心从“集装箱”走向永久建筑级应用。

随着国家绿色建筑等政策要求的不断更新，总体建筑将趋向于模块化、标准化方向发展，预制化数据中心将成为数据中心建设的主流模式，有望大规模推广应用。

3 算力工厂概述

面对算力需求的爆发性增长，传统数据中心已无法满足建设周期、能源利用率等严苛的限制要求，而预制化数据中心具备快速交付、绿色极简、柔性扩容等多重优势，能够满足算力的快速投建运营，实现投资的快速变现。

3.1 算力工厂模式定义

算力工厂指快速建设投运，对外提供弹性、高效、安全算力运营服务的算力中心，用户提供从规划、建设到对外提供算力运营服务的一体化算力中心全生命周期服务。算力工厂以不同形态预制化集装箱数据中心作为算力底座，中间层以计算、存储、网络等IT设备组成的硬件资源池作为算力支撑，顶层为算力资源的运营服务。算力工厂旨在为用户以最快的模式搭建对外提供算力运营服务的平台，从而帮助用户降低成本，最大化发挥算力效益，提高资金使用效率，改善用户整个生命周期TCO和现金流，通过算力工厂的模式，降低使用算力的门槛和成本，促进部署和应用，助力客户商业成功。

开放计算标准工作委员会 Open Compute Technology Committee

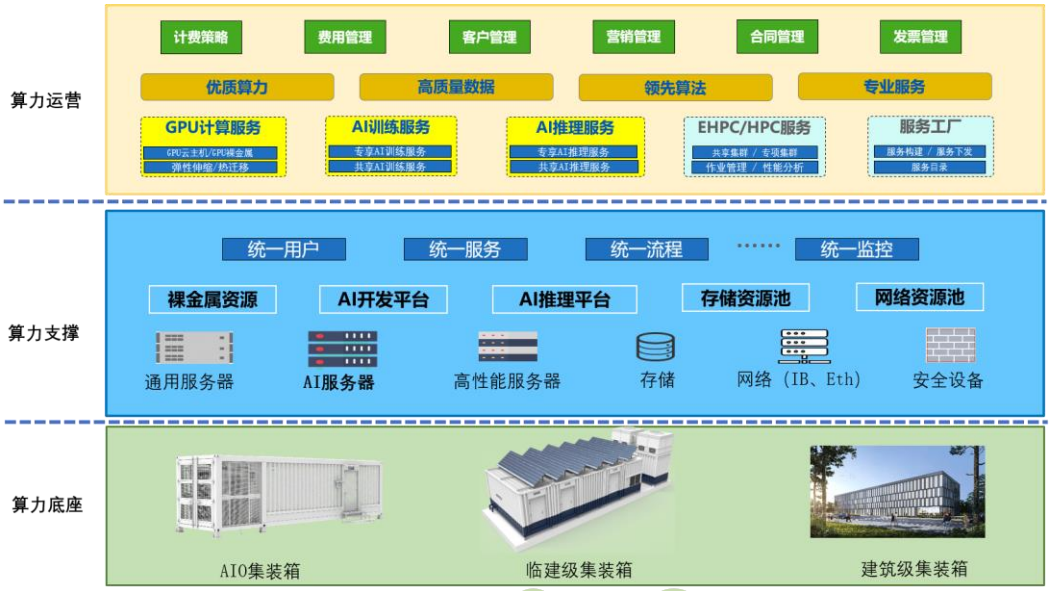


图2 算力工厂架构图

3.2 算力工厂形态分类

按照搭载算力规模的不同，算力工厂可分为以下三大建设形态：

1）All In One集装箱

以40/20英尺集装箱为载体，在单个箱体内集成机柜、配电、空调、监控、消防于一体，一体化快速交付部署。

2）临建级集装箱

以40/20英尺或其他非标准尺寸集装箱为载体，采用拼箱的方式，多个不同功能的集装箱拼接而成的单层临时建筑结构的集装箱数据中心。

3）建筑级集装箱

以40/20英尺或其他非标准尺寸集装箱为载体，40采用拼箱和叠箱的方式，将多个不同功能的集装箱搭建而成的单层或多层具有产权的建筑结构的集装箱数据中心。

3.3 算力工厂核心理念

算力工厂打造创新的算力全生命周期服务模式，让算力以更合适的方式更快发生在更亟需的地方。

1）全流程交钥匙服务模式

算力工厂以用户算力需求为导向，提供定制化的算力规划、建设、运营的全流程交钥匙服务，整合专业的融资租赁平台可为用户解决资金需求问题。

2）敏捷的多形态交付

算力工厂以预制化集装箱数据中心为基础底座，相比传统的土建模式，打造极致快速交付的单箱、拼箱、叠箱多形态算力工厂基座。

3）适配多需求的业务场景

算力工厂提供算力中心的的代建代维服务，面向用户轻重资产的算力需求，提供新建或租赁模式，灵活适配多种需求的业务场景。

开放计算标准工作委员会 Open Compute Technology Committee

3.4 算力工厂模式优势

1）快速建设投运

采用预制化集装箱建设模式，全模块化设计，工厂预制预集成预测试，减少现场施工误差。基础土建与模块工厂生产同步进行，现场乐高式拼装，去工程化交付，相对传统土建机房提升50%交付效率。

2）绿色低碳

光伏+储能+液冷+自然冷源多种绿色技术有机结合统一，PUE可低至1.15以下，降低碳排放25%。集装箱为预制化钢结构建筑，装配率可达90%以上，材料回收率超80%，施工现场无粉尘噪音，建筑垃圾少，绿色环保。

3）超高功率密度

风冷场景单柜设计功率密度40kW以上，液冷场景单柜设计功率密度65kW以上，匹配多种高功率密度算力设备部署上架。

4）智能化运维管理

搭建融合动环监控系统与AI综合运维管理系统统一管控平台，实现数据中心统一纳管；全方位监测基础设施动力环境参数及IT设备功耗、部件温度等重要相关参数，实现机房级与服务器级系统联合动态调优，打造更加高效可控的数据中心，双向故障预警，有效提升安全可靠性。

5）柔性扩容

支持同层水平和多层垂直扩容方式，初期根据资金水平和算力需求按需部署，后期根据算力设备，匹配新型数据中心模式在线柔性升级扩容。

6）多种建设形态

根据算力规模，可提供All In One、临建级和建筑级算力工厂三种模式，满足云计算、边缘计算等多场景需求。

7）算力运营服务

建设前期借助合适的投资方平台，后期为用户提供算力运营服务，彻底解决用户资金成本的后顾之忧。

开放计算标准
工作委员会
Open Compute Technology
Committee

3.5 算力工厂应用场景

算力工厂是一个搭载算力设备的绿色预制化集装箱数据中心，可直接对外提供算力服务。有效破解算力供需难匹配，高能耗挑战，快速变现的算力难题。

面对算力需求的井喷，预制化建设模式保障最大时效的算力供给落地，快速响应客户的需求，敏捷提供业务支撑；对于市场需求的快速变化，算力工厂提供按需供给的模式，IT设备与数据中心基础设施相匹配，模块化部署，无多余无效的投资；面向市场需求的多元化，多样性算力协同，算力工厂提供混合密度灵活可扩展的部署方式，满足不同客户不同算力的建设诉求；高算力设备带动算力中心向高功率密度高能耗方向发展，算力工厂的高单柜功率密度与低PUE设计，适应技术发展的趋势与国家政策的要求。

生成式人工智能有望重构算力服务模式和市场格局。鉴于基础大模型的本地训练成本不菲，企业将更多地使用人工智能就绪的数据中心设施和生成式人工智能服务器集群，从而缩短部署时间，降低设施的长期投资成本。从基础设施层面来说，传统计算基础设施难以满足大模型时代对于算力、存储和网络的高性能需求，因此算力工厂从芯片、处理器、存储、网络、数据库、云原生架构等维度，对算力基础设施进行全面升级，满足用户在超大加速环境中对快速扩展的需求，提供可用、易用、高效的资源供给服务；在算力服务交付的过程中，算力工厂能够提供定制化的基础设施服务能力，并提供全生命周期的算力运营服务。

3.6 算力工厂收益分析

采用预制化集装箱方式相对传统土建可提升机房交付周期50%以上，能够以最快的方式实现算力的落地输出。因此算力工厂的收益主要体现在以最快的建设方式落地，将节省的建设时间转化为算力出售运营的收益。

以华东地区某项目为例：机柜总数500台，单机柜功率20kW，总IT设备用电容量10MW，供电架构2N，备电15min，制冷系统N+1配置，电费0.75元/kWh，对比预制化集装箱方式和传统土建模式下，进行收益对比分析。

表1 预制化集装箱数据中心与传统土建模块化数据中心收益对比分析

对比项		预制化集装箱数据中心	传统土建模块化数据中心	备注
机柜数量（个）		500	500	
功率密度（kW/柜）		20	20	
建设周期（月） 不含论证及报规报建		9	23	预制模块化快14个月
建设总投资	总费用	1.00xA	A	端到端投资基本持平
	IT设备	B	B	
	基础设施	1.0xC	C	
3年收入	租金	2E	E	预制模块化早上线14个月
投资回报期（年）		1.9	3.1	预制模块化快1.2年

注1：1.00x及1.0x表示倍数，x取值区间为1～9。

注2："3年收入"和"投资回报期"均已启动建设时间为起点计算。

注3：以上为某特定应用场景需求明确的前提下收益的测算，仅供参考。

预制化集装箱数据中心由于采用全钢结构，相对传统土建数据中心在数据中心基础设施的造价要高，而IT设备的投资远远高于基础设施的投资，因此在相同规模算力的条件下，预制化集装箱数据中心和传统土建数据中心在建设总投资上端到端基本持平。预制化集装箱数据中心建设周期短，大幅加快业务上线的时间，实现更早的收益，因此在上述规模的算力中心中，土建模式下投资回收期为3.1年，预制化模式下投资回收期为1.9年，提前了1.2年。

开放计算标准工作委员会
Open Compute Technology

4 算力工厂建设模式

4.1 建设模式分类

算力工厂建设模式主要有两大类：设备采购模式和EPC建设模式。设备采购模式，则是业主自行采购或委托承包商采购工程项目所需的设备和物资。EPC模式，即设计采购施工总承包模式，由总承包商负责整个项目的设计、采购和施工等工作。

4.1.1 设备采购模式

设备采购模式则是指业主自行采购所需的设备和物资。业主需要根据项目需要和设计方案，进行设备采购和选型，以确保项目的需求得到满足。同时，设备采购需要考虑设备的质量、价格、交货期等因素，以确保项目的顺利进行。

这种模式的优点在于业主可以更好地控制设备的采购过程和成本，同时可以与供应商建立更紧密的关系。但在这种模式下，业主需要承担更多的管理和风险，同时需要对设备和物资的采购过程进行全程监控。

在EPC模式中，总承包商为了降低风险获得更多的利润，可能通过调整设计方案来降低成本，可能会影响长远意义上的质量。而设备采购模式中，一般用户会对产品设备性能参数品牌等有非常明确的要求。

4.1.2 EPC 建设模式

EPC（Engineering Procurement Construction）模式，即设计采购施工总承包模式，是指企业受业主委托，按照合同约定对工程建设项目的设计、采购、施工、试运行等实行全过程或若干阶段的承包。通常企业在总价合同条件下，对所承包工程的质量、安全、环保、费用和进度负责，项目建设完成后移交给业主，业主向总承包方支付合同规定的项目建设费用。

这种模式具有较高的灵活性，能够根据项目的实际需求和特点进行调整和变动。同时，由于这种模式将设计、采购、施工等环节进行了集成，能够更好地协调和管理项目资源，提高项目效率和质量。较传统承包模式而言，EPC总承包模式具有以下六个方面基本优势：

1）EPC总承包商负责整个项目的实施过程，不再以单独的分包商身份建设项目，有利于整个项目的统筹规划和协同运作，可以有效解决设计与施工的衔接问题、减少采购与施工的中间环节，顺利解决施工方案中的实用性、技术性、安全性之间的矛盾。

2）工作范围和责任界限清晰，建设期间的责任和风险可以最大程度地转移到总承包商。

3）合同总价和工期固定，业主的投资和工程建设期相对明确，利于费用和进度控制。

4）能够最大限度地发挥工程项目管理各方的优势，实现工程项目管理的各项目标。

5）建设工程质量责任主体明确，有利于追究工程质量责任和确定工程质量责任的承担人。

6）可以将业主从具体事务中解放出来，关注影响项目的重大因素上，确保项目管理的大方向。

同时EPC总承包模式也存在以下缺点：

1）业主主要是通过EPC合同对EPC承包商进行监管，对工程实施过程参与程度低，控制力度较低。

2）业主将项目建设风险转移给EPC承包商，因此对承包商的选择至关重要，一旦承包商的管理或财务出现重大问题，项目也将面临巨大风险。

3）EPC承包商责任大，风险高，因此承包商在承接总包工程时会考虑管理投入成本、利润和风险等因素，所以EPC总包合同的工程造价水平一般偏高。

4.2 建设模式推荐

在实际的工程建设项目中，设备采购模式和EPC建设模式并不是互斥的，而是可以相互配合、相辅相成的。在EPC模式中，总承包商需要对整个项目的建设过程进行全面负责，这其中也包括了设备的采购工作。因此，总承包商需要根据具体情况选择合适的设备采购模式，如自主采购、招标采购或委托采购等，以保证设备和物资的质量、成本和供应。而在设备采购模式中，业主可以自行组织采购工作，也可以通过EPC模式将采购环节委托给其他专业公司进行实施，从而减少自身的管理难度和风险。这种模式可以最大限度地发挥各自的优势，提高项目的管理水平和效率。

开放计算标准工作委员会 Open Compute Technology Committee

按照建设规模，算力工厂可分为All In One、临建级和建筑级三大类。All In One算力工厂以单个40英尺集装箱或20英尺集装箱为载体，为临时性建筑结构，宜采用设备采购模式，快速投建投运。临建级算力工厂多以5～10个集装箱拼箱组成，为临时性建筑结构，宜采用设备采购模式，也可采用EPC建设模式。建筑级算力工厂以几十至上百个集装箱多层叠箱组成，为具有产权建筑结构的数据中心，宜采用EPC建设模式。

5 算力工厂规划建议

5.1 算力形式

在算力工厂规划阶段，需首先明确采用哪种算力形式。

通常来说，将CPU、GPU等芯片技术及能力，称为狭义的算力。内存、硬盘相关的存储技术，称为存力。操作系统、数据库、中间件、应用程序等在内的软件技术，称为算法。广义的算力，既包括了狭义的算力，也包括了存力和算法。根据所处理的业务特性不同，广义的算力又可分为通用算力、智能算力、超算算力和边缘算力四种类型。

通用算力以 CPU 芯片输出的计算能力为主；智能算力以 GPU、 FPGA、 Al 芯片等输出的人工智能计算能力为主；超算算力主要以超级计算机输出的计算能力为主；而边缘算力主要以就近为用户提供的实时计算能力为主，是以上三种算力形式的组合。未来算力革命的发展趋势，将是以智能算力为核心的多种算力形式的融合，多场景共荣算力生态正在成型。

5.1.1 通用算力

通用算力以CPU芯片输出的计算能力为主，能够进行各种数学和逻辑运算、处理文本和图形等数据信息的计算能力。通用算力应用场景极为广泛，几乎囊括了传统服务器提供的全部业务应用，在科学研究、工程设计、数据处理和分析、安全监控、云计算和存储、网络游戏、网络购物、移动视频、移动支付、大数据挖掘、数字李生仿真模拟、图像渲染、视频渲染等领域都有广泛的应用。这些应用场景对实时性要求不高，但是算力强度需求大。随着数字化转型的加速和人工智能技术的不断发展，通用算力的需求也在不断增加。

开放计算标准工作委员会
Open Compute Technology Committee

5.1.2 智能算力

智能算力以 GPU、 FPGA、 Al 芯片等输出的人工智能计算能力为主，具备智能化处理能力的计算设备或计算系统，它们可以实现对大量数据的快速处理、分析和判断，同时能够根据不同的应用场景进行智能化的处理和决策。智能算力的核心在于智能化，它基于人工智能、机器学习等技术，能够自主地学习和适应不同的环境，从而完成各种复杂的任务。

智能算力主要用于图像计算服务、数据推理、强化学习训练等场景。这些应用场景对实时性要求较高，需要大量的智能计算能力。例如，在自动驾驶中，需要对大量的图像数据进行实时分析，以实现车辆的精准控制和避障；在语音识别中，需要快速处理大量的音频数据，以实现实时翻译和回应。

5.1.3 超算算力

超算算力以超级计算机输出的计算能力为主，利用并行工作的多台计算机系统的集中式计算资源，通过专用的操作系统并行计算提高运算速度，解决大规模科技计算和海量并发的数据处理问题。

超算算力主要用于专门承担各种大规模科学计算和工程计算任务，用于科研计算中的流体力学、物理化学、生物信息等科研领域。常见的超算应用场景领域包括气象分析预测、高海拔宇宙线观测、空气动力学、车辆碰撞测试仿真实验、药品实验数据分析等，处理的数据量非常庞大。这些领域的研究需要极高的计算精度和速度，因此需要使用超算中心提供的强大算力支持。

5.1.4 边缘算力

边缘算力主要是以就近为用户提供的实时计算能力为主，是通用算力、智能算力和超算算力三种算力形式的组合，用以解决网络延迟产生的问题。边缘算力主要应用在智能制造、智能交通、智能安防、智能家居、智能农业、无人驾驶、医疗保健、智慧城市等领域，以提供更快速、更高效的计算服务。

5.2 算力规模

在算力工厂规划阶段，应根据算力规模确定采用哪种形式以及何种模式建设。下文以智能算力设备提供的智能算力规模为评估依据。除智能算力外，也可根据用户实际需求，提供通用算力、超算算力、边缘算力的规模评估。

All In One级算力工厂以单个40/20英尺集装箱为载体。风冷场景下，40英尺集装箱最大可提供100kW的IT设备供电和制冷能力，满足10P左右智能算力设备的部署，20英尺集装箱最大可提供50kW的IT设备供电和制冷能力，满足5P左右智能算力设备的部署。液冷场景下，40英尺集装箱最大提供300kW的IT设备供电和制冷能力，满足30P左右智能算力设备的部署，20英尺集装箱最大可提供100kW的IT设备供电和制冷能力，满足10P左右智能算力设备的部署。因此对于5P以内的风冷智能算力需求，建议采用20英尺All In One级形式，对于5P～10P的风冷智算算力需求，建议采用40英尺All In One级形式。对于10P以内的液冷智能算力需求，建议采用20英尺All In One级形式。对于10P～30P的液冷智能算力需求，建议采用40英尺All In One级形式。该种形式宜采用设备采购模式进行建设。

开放计算标准工作委员会 Open Compute Technology Committee

表2 All In One级算力工厂形式

应用场景	风冷		液冷	
集装箱尺寸	20英尺	40英尺	20英尺	40英尺
IT设备供电和制冷能力	50kW	100kW	100kW	300kW
智能算力	5P	10P	10P	30P

临建级算力工厂通常以10个以内的40/20英尺集装箱单层拼箱而成，典型风冷及液冷场景下，可提供1MW左右的IT设备供电和制冷能力，满足150P左右智能算力设备的部署。因此对于10P～200P的风冷或30P～200P的液冷智能算力需求，建议采用临建级形式，宜采用设备采购模式，也可采用EPC建设模式。

建筑级算力工厂通常几十至上百个40/20英尺集装箱多层叠箱而成，典型风冷及液冷场景下，可提供10MW左右的IT设备供电和制冷能力，满足1500P左右智能算力设备的部署。因此对于200P以上的智能算力需求，建议采用建筑级形式，宜采用EPC建设模式。

以上算力规模的分类和建设方式为典型场景化算力工厂建设的推荐方式，可根据客户实际需求进行灵活调整。

5.3 算力应用

在算力工厂规划阶段，还需明确算力应用场景。

为进一步凝聚产业共识、强化政策引导，全面推动我国算力基础设施高质量发展，2023年10月，工业和信息化部、中央网信办、教育部、国家卫生健康委、中国人民银行、国务院国资委等六部门联合印发《算力基础设施高质量发展行动计划》。指出推动算力基础设施高质量发展，更好汇集算力资源、发挥算力价值，构建泛在融合的“算力+”行业应用体系。 围绕工业、金融、医疗、交通、能源、教育等重点领域，各打造30个以上应用标杆，并推动算力赋能千行百业。部分重点行业算力应用模式如下：

5.3.1 互联网行业

在过去的几年中，尽管人工智能在互联网的投资增速有所放缓，但在中国，互联网仍是人工智能技术应用和研发的主战场。从自然语言处理到图像识别，再到智能客服，人工智能的技术不仅提高了用户的体验，也极大地提高了服务质量。如今，互联网行业更为重视人工智能技术。自然语言处理，如语音识别、文本生成和机器翻译等，已被广泛应用。在图像识别方面，深度学习技术使得机器能够识别和理解图像中的各种元素，如人脸识别在安全和身份验证中的应用，以及在智能驾驶中的障碍物识别技术。

互联网企业将各种人工智能技术广泛应用于其业务模块，以优化用户体验并提升商业运营效率：智能客服借助自然语言处理和机器学习技术，为用户带来更为高效和流畅的服务，同时使企业服务效率更高、成本更低；大数据分析与智能决策也逐渐成为互联网企业的核心能力，通过基于智能搜索技术提取有价值的信息，从而助力企业做出更加智慧的决策；在营销领域，智能营销和个性化推荐技术，使得互联网企业能够更准确地洞察用户的兴趣和需求，增加客户粘性，向客户提供更有针对性的帮助。可以说，中国的互联网行业正在经历一个深度的智能化革命，由人工智能、大数据、自动化和其他先进技术共同推动，旨在更好地满足市场和用户的多样化需求。

5.3.2 电信行业

开放计算标准工作委员会 Open Compute Technology Committee

人工智能技术正深入地渗透到电信行业，各大运营商都在积极探索其潜在价值。凭借海量、多样且真实的用户数据，这些运营商已经构建了高价值的人工智能训练数据集，为电信行业未来的智能化发展打下了坚实的基础。一方面，中国运营商加速云数据中心建设，积极部署云上智能化能力，加速数据中心业务发展；另一方面，加速基础设施的智能化建设，支持电信网络的建设和优化。例如，网络智能化已使得资源管理、网络优化和故障诊断更为高效，能够根据实时数据动态调整，以提升网络的性能和用户体验。此外，运营商越来越多地依赖人工智能技术改进自身服务，如在智慧营业厅的建设中，采集的消费者行为数据如停留时间等，都会被智能化地分析并及时做出相应的反应。此外，防诈骗也得到了加强，电信企业利用大数据和人工智能技术大大降低了电信诈骗的可能性。同时，运营商也投入资金到提供智能化算力的平台，以获得针对各种高技术场景的解决方案。至于数据智能化，借助对通话记录和网络行为的深度分析，企业能够更深入地了解用户需求，并为他们提供更加个性化的服务。总的来说，人工智能在电信行业的广泛应用，不仅提升了客户体验，还大幅增加了运营商的业务营收、降低了运营成本，并推动了整体效率的提升。

5.3.3 政府行业

在现代数字化时代，人工智能技术在政府领域的应用日益显著，极大地推动了公共服务和城市治理的进步。例如，数字政府已经利用自然语言处理和机器学习算法，推出了智能客服系统，使公众能够获得快速、准确的在线信息查询与答疑服务。数据挖掘和预测技术也使数字政府能够深入分析海量政府数据，找到潜在的模式，进而为决策提供科学依据。在城市管理方面，结合物联网、大数据和人工智能技

术，政府正朝着智能城市管理的方向发展，优化城市基础设施、交通、环境的监测与管理，确保城市的安全、便捷和可持续性。数字政府还在积极推进办公流程与服务的自动化，通过机器人流程自动化（RPA）等工具，极大提升了工作效率和公众的满意度。值得注意的是，这些应用只是目前数字政府采纳的一部分，随着人工智能技术的不断演进，未来无疑会有更多的创新应用涌现。总的来说，通过这些人工智能应用，数字政府不仅提供了更便捷、高效和智能化的公共服务，还显著提升了公众满意度和城市治理水平。

5.3.4 金融行业

人工智能在金融行业的应用迅猛增长，已经渗透到诸如银行、投资机构及保险和证券等各个领域中，主要应用包括智能客服、实体机器人、智慧网点和云上网点等，为各机构提供了更好的客户体验和高度的便利性。基于人工智能，金融企业利用机器学习算法进行数据分析，根据客户的风险状况进行准确评估，实现了贷款审批流程的自动化，确保贷款的准确性和效率。金融科技应用程序的开发者也正在积极地将更多功能——如EMI计算器和贷款资格自我评估等，集成到人工智能和机器学习技术中。此外，通过引入人工智能技术，金融机构现在可以实时监测交易和市场波动，从而及时制定策略。大量的数据，如交易、用户行为和市场数据，现在都可以通过大数据分析和人工智能技术进行深入挖掘，为风险评估、投资决策和市场预测等提供更强大的支持。为了进一步提供个性化服务，金融机构采用机器学习和推荐系统，为客户提供更符合其需求的投资方案和贷款产品。再者，金融机构也开始利用区块链技术实现更快、更安全、更透明的交易和结算，大大提高了交易效率并减少了中间环节。此外，金融企业与科技公司加速合作，通过技术合作、数据共享和创新孵化等方式，推进智能化进程，推动金融行业向智能化、自动化和个性化方向发展，为整个行业带来了创新和增长的机会。

5.3.5 制造行业

开放计算标准
工作委员会
Open Compute Technology Committee

人工智能正在改变制造业，为其带来工业4.0和工业互联网时代的技术变革。通过结合大数据、物联网和其他先进技术，制造业实现了生产的自动化、智能化和灵活化，进而提高了生产效率。人工智能在制造业中应用包括但不限于：交互界面的智能化，质量管理，维修与生产检测的自动化，以及供应链管理的智能化。无人驾驶技术已进入汽车制造业，人工智能的感知、决策和控制能力促进了车辆间的智能协作，可以提高道路安全和交通效率；工业机器人也在改变汽车制造业，带来智能制造的革新，实现更高效的生产线布局和生产流程。机器视觉技术被广泛应用于产品的智能质检，对产品进行分类和检测；例如，在手机制造中利用图像识别技术，通过亮度、均匀度以及像素纹理等指标对屏幕素质进行检测。对于汽车的维修与保养，智能化系统能够自动检测故障并生成维修方案，以及通过数据分析来预测零部件的使用寿命。制造业也利用大数据分析和人工智能技术来优化生产过程，实现生产的实时监测和分析，以此来优化生产过程，预测设备故障，提高产品质量等。利用物联网和大数据技术，制造业更进一步实现了供应链的智能管理，通过可视化、协同化、智能化，提高了供应链的响应速度，降低了库存成本，大大增加物流管理的灵活度。

5.3.6 其他行业

随着科技创新的不断涌现，算力的新应用场景也日益丰富，需求量将持续激增。

在交通行业，人工智能技术也正在带来深刻的变革。以物流为例，该行业已经大量采用了智能技术来提高效率和用户体验。智能物流已经成为了发展的趋势，包括智能仓储和智能配送。在智能仓储中，货物的自动存储、拣选和库存管理的自动化，都极大地提高了仓库的运营效率，同时降低了人工成本和

管理的难度。智能配送方面，则采用了多模式运输规划，无人机和机器人技术，以实现物流运输路线的优化，确保货物能够在最短的时间内，以最低的成本送达目的地。此外，汽车制造商为了增强驾驶安全和便捷性，已经开始在其产品中实施半自动驾驶功能，例如先进的驾驶辅助系统（ADAS），从而帮助用户提高使用体验，如智慧入库停车、恶劣天气控制车辆避免碰撞等。除了私人用车方面的应用，人工智能技术在公共交通管理和道路监控方面也扮演了关键的角色，确保道路交通得到智能化管理，从而大幅提升整体交通效率。

在教育行业，科研院所和高等学府正在深度融入智能化技术以提升其教育与管理质量，提高科研创新能力，通过机器学习、深度学习、大模型等人工智能技术，基于对模拟和真实科研数据的分析，加速计算推演，帮助基础学科和应用学科更快、更准确地进行假设的验证和应用。

在医疗行业，尽管当前只有少数医疗人员参与人工智能的开发与应用，这主要是由于相关标准与规范尚处在完善阶段，但随着中国政府对人工智能法规的日益完善，未来五年中，其应用势必将得到快速的扩展。医疗数字化转型中，传统模式向互联网医疗模式的转型趋势显而易见。如人工智能和大数据技术等，已经促使疾病的诊断与治疗模式从单一领域扩展到多个领域，从而推动了医疗信息化的整体升级。预计在未来，受到政策支持和市场需求的双重推动，生物方向的人工智能应用将展现出巨大的发展潜力。

在能源行业，利用人工智能的方法、算法和工具改进和优化能源的生产、分配、消费和管理过程。这些应用涵盖了从能源勘探、生产、传输到消费和管理的整个能源链。随着技术的不断进步和应用场景的扩展，人工智能在能源行业的应用将越来越广泛和深入，为能源行业的可持续发展提供有力支持。

6 算力工厂建设思路

6.1 建设原则

1）安全可靠性

算力工厂的基础设施、网络设施、计算与存储资源等各系统应具有网络和系统的全方位安全性保护部署。应合理规划物理安全、网络安全和设备安全措施，确保数据的机密性、完整性和可用性。避免出现单点故障。在关键设备采用硬件备份、冗余等可靠性技术的基础上，采用相关的软件技术提供较强的管理机制、控制手段和事故监控与安全保密等技术措施来提高安全性。

开放计算标准工作委员会
Open Compute Technology Committee

2）先进成熟性

算力工厂系统设计应以满足用户业务未来5年以上的发展需求为重点，统一规划基础设施、算力资源、统一规划网络、统一安全要求等系统设计，应充分采用符合国际标准的、先进并且成熟的数据机房系统、计算机系统、网络系统、存储系统、集群相关软件系统等先进技术和产品。同时应根据实际使用需求，分析应用特征，避免盲目追求高精配置。

3）灵活扩展性

算力工厂建设应考虑到未来业务发展的需求，为方便实现数据中心规模的扩展，系统的设计应具有灵活的扩展能力，包括网络端口的扩展、带宽容量的扩展、处理用户访问的能力的扩展、计算能力的扩展、存储能力的扩展、基础设施的扩展等。扩展应能在线进行，要在确保在不中断服务的基础上，方便地对系统进行平滑升级与扩容。

4）绿色节能性

算力工厂系统规模大，耗电量大，节能环保不但能够大大降低用户运维成本，同时也是在相应国家节能减排、绿色低碳的号召。为降低集群整体能耗，宜采用冷板式液冷服务器，节点冷板覆盖CPU、GPU、内存等核心发热元件，节省了空气换热环节，提高末端供水温度，降低制冷系统PUE，提高计算密度。

6.2 建设流程

6.2.1 设备采购模式建设流程

设备采购模式多用于All In One算力工厂和临建级算力工厂，由于两种模式均属于临时建筑结构，因此不需要对外的报批报建流程。用户可根据实际需求确认选址、设计方案及产品规格，经内部立项后发起设备采购，由中标厂家负责深化设计、设备生产、设备交付、调试运行及交付验收。算力设备宜与预制化数据中心打包采购。相对EPC建设模式，本模式更加灵活方便快捷。

6.2.2 EPC 建设模式流程

EPC建设模式多用于建筑级算力工厂的建设，主要参照数据中心的建设流程，其建设流程一般包括以下几个环节：

1)内部立项：项目单位根据市场调研情况取得内部立项审批。

2)可行性研究：进行数据中心建设的可行性研究，编制可研报告，确定建设内容和规模。

3)选址：考虑项目所在地自然条件、气温、地址条件、自然灾害等，例如是不是地震带上，有没有发生水灾的可能性，气温也尤为重要，设计到PUE值大小。另外要考虑到项目所在地网络带宽资源是否满足，至少三大运营商网络光纤连接主干网络。

4)外部立项：根据可行性研究的结果及选址，进行项目外部立项，明确建设目标和方案，取得政府批复。

5)办理相关手续：在立项后，需要办理各种手续，包括能评、环评、稳评、安评、规划许可、施工许可、供电方案、消防等。其中能评需要第三方编制能评报告，环评需要编制环境评估报告。

6)设计：进行数据中心的方案设计和施工图设计。

7)土建和机电安装：进行数据中心的土建和机电安装工作。

8)设备调试：对数据中心内的设备进行调试，确保正常运行。

9)验收交付：对数据中心进行验收，确保符合建设要求，并进行交付使用。

开放计算标准工作委员会 Open Computer Technology Committee

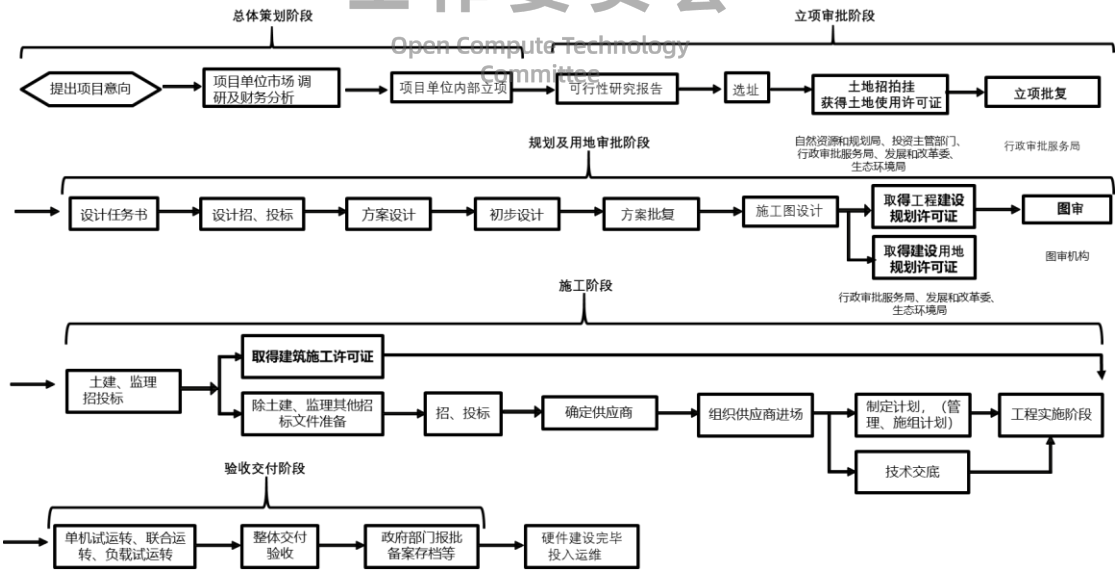


图3 建设流程

请注意，不同地区对数据中心建设的具体要求可能有所不同，实际流程可能需结合当地的具体规定和要求进行调整。

在预制化数据中心交付验收后，按需上架算力设备即可。算力设备的采购宜规划在EPC总包中，以充分发挥EPC模式的优势。

6.3 算力工厂技术框架

6.3.1 预制化数据中心技术框架

6.3.1.1 建设方案分类建议

预制化数据中心以预制化、模块化、标准化为建设理念，融合数据中心土建工程（L0）及机电工程（L1），以集装箱为载体，集成机柜系统、配电系统、制冷系统、监控系统、消防系统、照明系统、防雷接地、综合布线等系统于一体，具备快速部署、柔性扩容、高效节能等特点，是一种新型数据中心建设模式。

预制化数据中心规划建设应参考以下标准及规范

（当依据的标准规范修订或有新的标准规范出版时，应对相关内容进行复核。）

GB 50174-2017《数据中心设计规范》
GB 50054-2011《低压配电设计规范》
GB 50015-2009《建筑给排水设计规范》
GB 50016-2014《建筑设计防火规范》
GB 50140-2005《建筑灭火器配置设计规范》
GB 50013-2018《室外给水设计规范》
GB 50014-2006《室外排水设计规范》
GB 50370-2005《气体灭火系统设计规范》
GB 50057-2010《建筑物防雷设计规范》
GB 50736-2012《民用建筑供暖通风与空气调节设计规范》
GB 50034-2013《建筑照明设计标准》
GB 50116-2013《火灾自动报警系统设计规范》
GB 50343-2012《建筑物电子信息系统防雷技术规范》
GB 50348-2018《安全防范工程技术规范》
GB 50395-2007《视频安防监控系统工程设计规范》
GB 50243-2016《通风与空调工程施工及验收规范》
GB 50222-2017《建筑内部装修设计防火规范》
GB 50050-2017《工业循环冷却水设计规范》
GB/T 36448-2018《集装箱式数据中心机房通用规范》
GB/T2887-2011《电子计算机场地通用规范》
GB/T50311-2016《综合布线系统工程设计规范》
YD 5098-2005《通信局（站）防雷与接地工程设计规范》
YD 5059-2005《电信设备安装抗震设计规范》
YD/T 585-2010《通信用配电设备》
YD/T 2063-2009《通信设备用电源分配单元(PDU)》
YD/T 1173-2016《通信电源用阻燃耐火软电缆》
其他现行各种相关的国颁、部颁、通信管理局颁发的规程、规范、技术标准等。

开放计算标准工作委员会 Open Compute Technology Committee

预制化数据中心根据不同的建设需求，提供单箱、拼箱、叠箱级多种形态数据中心建设模式，以单个箱体为独立的功能模块，结合设备具体要求，共分为三大核心模块，十大功能模块。即机房核心模块、辅助核心模块、动力核心模块。其中机房核心模块可细分为IT模块、电力模块、电池模块和办公模块；辅助核心模块可细分为走廊模块、楼梯模块和电梯模块；动力核心模块可细分为集成冷站模块、集成冷源模块和油机模块。为减少现场工作量，所有预制模块在工厂预制、预装、预调测，现场无需大规模土建，简单吊装、乐高式搭建，即可快速完成预制化数据中心的建设及部署。

按照搭载算力规模的不同，算力工厂可分为以下三大建设形态。

1）All In One算力工厂

All In One算力工厂是由单个IT模块组成的预制化数据中心。以40/20英尺集装箱为载体，在单个箱体内集成机柜、配电、空调、监控、消防于一体，作为算力工厂的基础底座，具备快速部署、高效节能、一体化交付的特点。风冷场景下，单个40尺集装箱可提供最大100kW算力设备的供电制冷能力，满足10P算力设备的部署要求。单个20英尺集装箱可提供最大50kW算力设备的供电制冷能力，满足5P算力设备的部署要求。液冷场景下，单个40尺集装箱可提供最大300kW算力设备的供电制冷能力，满足30P算力设备的部署要求。单个20尺集装箱可提供最大100kW算力设备的供电制冷能力，满足10P算力设备的部署要求。作为临时建筑不需要数据中心建设的报批报建流程。但存在无法通过等级保护测评的风险。

图4 20英尺All In One级风冷/液冷算力工厂外观图

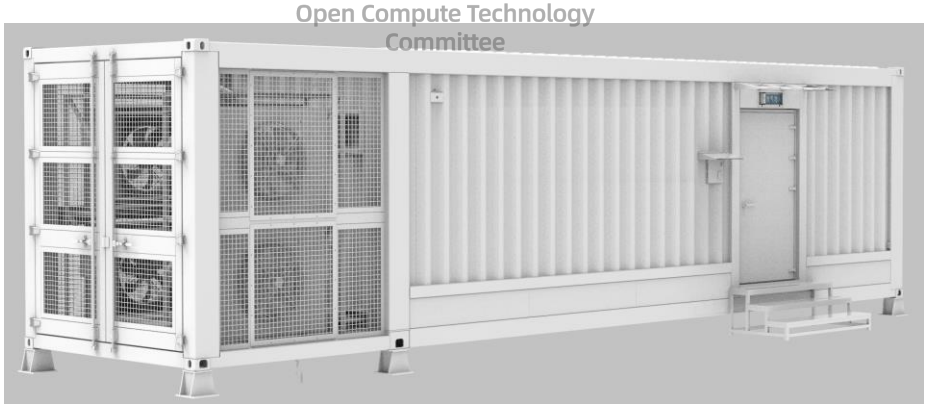


图5 40英尺All In One级风冷/液冷算力工厂外观图

2）临建级算力工厂

临建级算力工厂以40/20英尺或其他非标准尺寸集装箱为载体，采用拼箱的方式，多个不同功能的集装箱拼接而成的单层临时建筑结构的集装箱数据中心，按功能分区分为主机房区、配电区、运维区等，IT模块、电力模块和电池模块作为必选，可选办公模块和油机模块。在风冷及液冷场景下，可提供1MW

开放计算标准工作委员会
Open Compute Technology Committee

左右算力设备的供电制冷能力，满足150P智能算力设备的部署要求。作为临时建筑不需要数据中心建设的报批报建流程。但存在无法通过等级保护测评的风险。

图6 临建级算力工厂外观图

3）建筑级算力工厂

建筑级算力工厂以40/20英尺或其他非标准尺寸集装箱为载体，采用拼箱和叠箱的方式，将多个不同功能的集装箱搭建而成的单层或多层具有产权的建筑结构的集装箱数据中心，按功能分区可分为主机房区、配电区、运维区、制冷区、消防区等，IT模块、电力模块、电池模块、办公模块、走廊模块和集成冷站模块作为必选，可选楼梯模块、电梯模块、集成冷源模块和油机模块。在风冷及液冷场景下，可提供10MW左右算力设备的供电制冷能力，满足1500P算力设备的部署要求。作为永久建筑需要按照上述数据中心建设的报批报建流程执行，而如果拟建设场地使用类型为工业用地，则可按照工业产房的形式进行申报使用，省去其他报批报建的额外审批流程。（请注意，不同地区对数据中心建设的具体要求可能有所不同，实际流程可能需结合当地的具体规定和要求进行调整。）

图7 建筑级算力工厂外观图

6.3.1.2 核心功能模块

1）IT模块

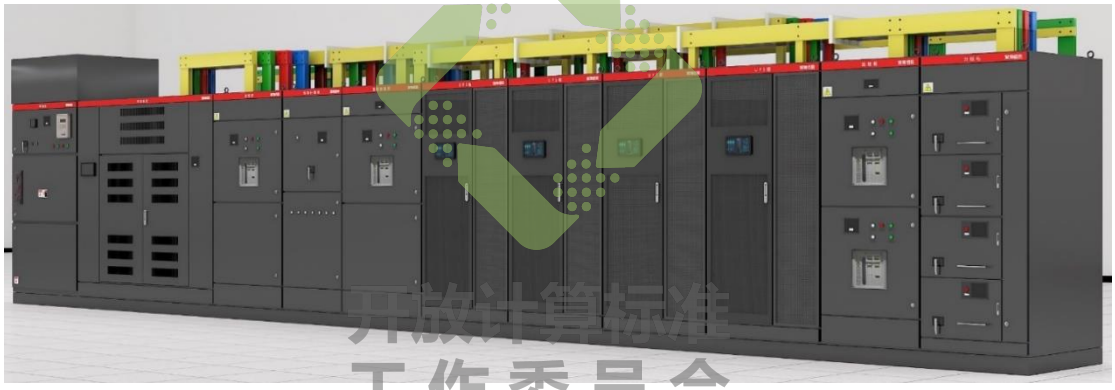
采用40/20英尺集装箱，高度集机柜、配电柜（小母线）、空调、消防、监控、综合布线于一体，充分利用集装箱空间并进行功能分区。支持部署双排封闭冷/热通道微模块及单排封闭冷热微模块，风冷场景下支持单柜最大设计功率40+kW，液冷场景下支持单柜最大设计功率65+kW。为保证IT模块室外

环境长期稳定可靠运行，箱体应通过防火、防水、防风、防霉菌、耐烟雾、9级抗震等严苛认证，保证最高使用年限可达50年。

图8 IT模块

2）电力模块

融合从中压到末端馈线的全功率链路，提供兆瓦级供配电一体解决方案，与集装箱紧密耦合，现场快速拼接，交付效率提升50%。高密部件集成融合、铜排一体连接，节省占地空间40%，全系统智能监测，故障预警，安全可靠。融合高效模块化UPS系统，整体效率提升1%。



开放计算标准
工作委员会
Open Compute Technology
Committee

图9 电力模块

3）集成冷站模块

基于集成模块化的理念，采用设计、生产、装配一体化和机电一体化技术，将冷水机组、板式换热器、冷却塔、冷冻水泵、冷却水泵、水处理装置、定压装置、管路及阀门、配电柜、智能控制柜等与集装箱在工厂集为一体，工厂预生产、预测试、现场吊装、管道对接、即插即用，无需独立机房，结构紧凑，节省60%占地空间。

图10 集成冷站模块

4）集成冷源模块

基于集成模块化的理念，采用设计、生产、装配一体化和机电一体化技术，将一次侧循环水泵、闭式冷却塔、定压补水装置、稳压补水装置、软水处理器、软化水箱、补液箱体、相关管路阀门、智能控制系统、安防环控系统、通风系统等与集装箱在工厂集为一体，工厂预生产、预测试，现场吊装、管道对接、即插即用，无需独立机房，结构紧凑，节省60%占地空间。

集成冷站与集成冷源均采用集成模块化的理念，外观均为集装箱加冷却塔的形式，区别在于集装箱内部集成的设备不同。

图11 集成冷源模块

5）油机模块

当市电停电时，油机模块作为保障算力工厂长时稳定运行后备电源产品，集成一线品牌发动机与发电机，保证供电系统安全可靠。集装箱一体集成+消音器设计，噪音降低15dB。

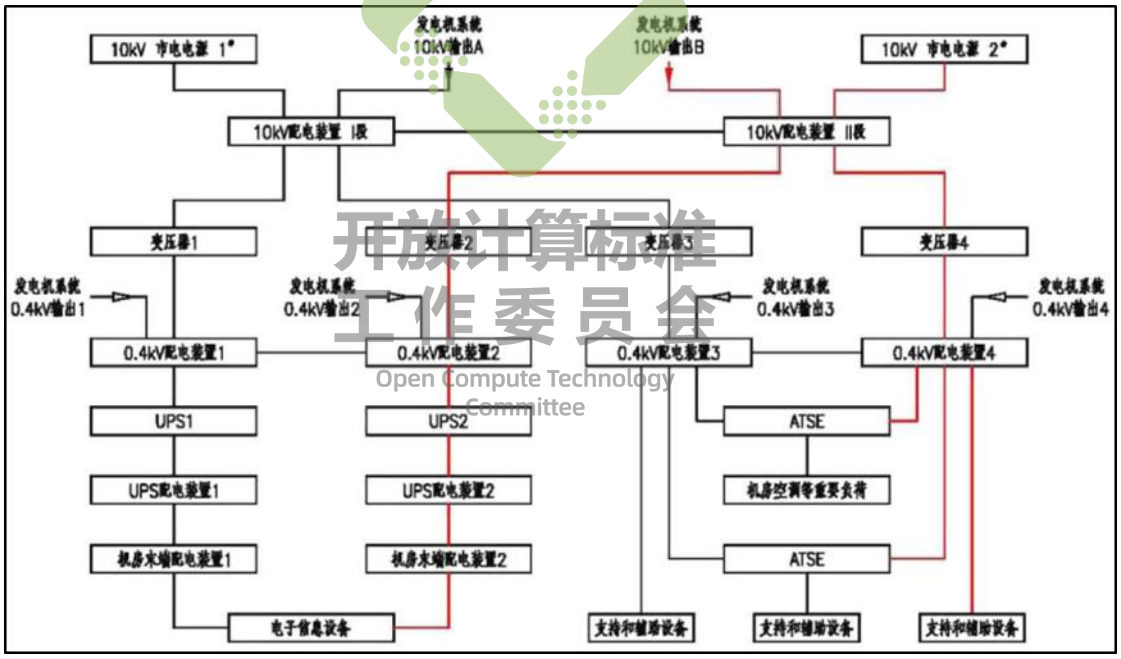
图12 油机模块

6.3.1.3 核心系统要点

1）供配电系统

供配电架构

算力工厂供配电系统宜参照GB50174-2017《数据中心设计规范》中A级数据中心执行。从可靠性及投资两方面考虑，宜采用一路市电+一路（N+1）UPS供电模式，也可采用2N/2（N+1）UPS供电模式，同时兼容B级及C级机房设计。



开放计算标准工作委员会
Open Compute Technology Committee

图13 A级数据中心2N供配电系统架构图

算力工厂应由专用配电变压器或专用回路供电，并考虑可扩展性预留备用容量。A级数据中心的供电电源应按一级负荷中特别重要的负荷考虑， B级数据中心的供电电源按一级负荷考虑， C级数据中心的供电电源应按二级负荷考虑。

对于A级数据中心应由双重电源供电，并应设置备用电源。备用电源宜采用独立于正常电源的柴油发电机组，宜采用油机模块的方式满足快速部署及降噪的要求。油机模块的容量应满足正常电源发生故障时，能承担算力工厂正常运行所需要的用电负荷，输出功率应满足数据中心最大平均负荷的需要。

对于B级数据中心宜由双重电源供电，当有双路电源时，不需要再设置油机模块；当只有一路电源时，应设置油机模块作为备用电源，油机模块的输出功率可按限时500h运行功率选择。

对于C级数据中心，采用一路电源供电，不需要设置油机模块。

为节省占地面积，简化供电链路，提升供电效率，算力工厂宜采用电力模块供配电系统，融合变压器、模块化UPS、输入输出、功率补偿等全链路设备。IT设备配电宜采用配电列头柜或智能小母线，当机柜容量或位置发生变化事，智能小母线能够灵活进行容量和位置调整，即插即用。配电列头柜或智能小母线应配置浪涌保护器和电源监测装置，并提供远程通信接口，接入算力工厂监控运维平台，对IT设备用电进行全方位监测与管理，保障IT设备的正常运行和能耗统计。

模块化UPS

模块化UPS具有可靠性高、可用性强、维护方便、扩展性强、节能环保等优点，因此算力工厂的设备宜采用模块化UPS系统供电。UPS基本容量宜按额定负载的1.2倍选型。IT设备与空调设备应由不同组UPS系统供电。

图14 模块化UPS

开放计算标准工作委员会 Open Compute Technology Committee

蓄电池

铅酸蓄电池具有安全性能好、成本低廉、稳定性高、适用范围广、维护方便、可回收利用等优点，因此算力工厂宜优选采用铅酸蓄电池方案。当前储能结合新能源，处于风口，随着锂电价格的下降，未来会有较大的应用空间，如有项目需求，算力工厂也可支持锂电方案。

2）制冷系统

当前，芯片功耗与服务器功耗逐步上升。高功耗芯片，高密度服务器及单机柜密度的演进，对于传统风冷散热来说，既容易出现局部热点，影响换热性能；又需要不断降低送风温度，增大送风风量，影响制冷能耗。因此，传统风冷制冷模式在换热性能及能耗优化方面逐步受限。液冷技术能够突破风冷局限，解决数据中心高功耗、高密度散热问题。两者都可以通过有效利用自然冷源冷却，其中液冷可实现数据中心PUE低至1.1，更好地促进数据中心减少碳排放。

液冷数据中心制冷解决方案主要有三条技术路线，即冷板式、浸没式和喷淋式。IDC数据显示，2022年，中国液冷服务器市场中，冷板服务器占到了90%。冷板方案在对原有基础设施进行改造的投入和难度方面具有优势，同时具有较高成熟度和较好商用基础。浸没式在散热效率和单机柜功率、空间利用率等方面比冷板式更具优势，但是受限于基础设施改造、建设成本、电子氟化液或其他冷却液的成本及可维护性等因素，目前发展仍相对缓慢。喷淋式与浸没式类似，同样适用于结构承重经过特殊加固的新建项目，不同之处在于：喷淋式方案中目前单机柜最大负载为48KW，应用范围相对狭窄。

液冷数据中心散热架构分为液冷系统和风冷系统。其中液冷系统主要是解决CPU、GPU、NV Switch等主要发热部件的散热，占服务器整体散热量的70%～85%，主要通过CDU实现二次侧冷却液的换热，

并通过一次侧集成冷源模块将热量带走。由于液冷服务器中仍存在一部分非液冷部件，占服务器整体散热量的15%～30%，且算力中心中仍有部分非液冷设备，因此仍需精密空调进行风冷散热，并通过室外机组将热量带到室外。精密空调系统宜采用氟冷列间空调+氟冷室外机或水冷列间空调+集成冷站模块的方式。列间空调采用水平送风背回风的气流组织，对环境具有高精度控制。同时具有高可靠性，节能性等特点。能保证敏感设备、通信设备和计算机等设备拥有一个合理的运行环境。

散热架构示意图如下：

图15 液冷数据中心散热架构图

在液冷需求场景下，冷板式液冷是当前的主流技术方案。冷板式液冷，主要通过冷板（高导热金属构成的封闭腔体，铜材质）将元器件的热量传递给封闭在循环管道中的冷却液体，然后利用冷却液体将热量带走，通过工作流体的传递将中间热量运输到后端进行冷却。

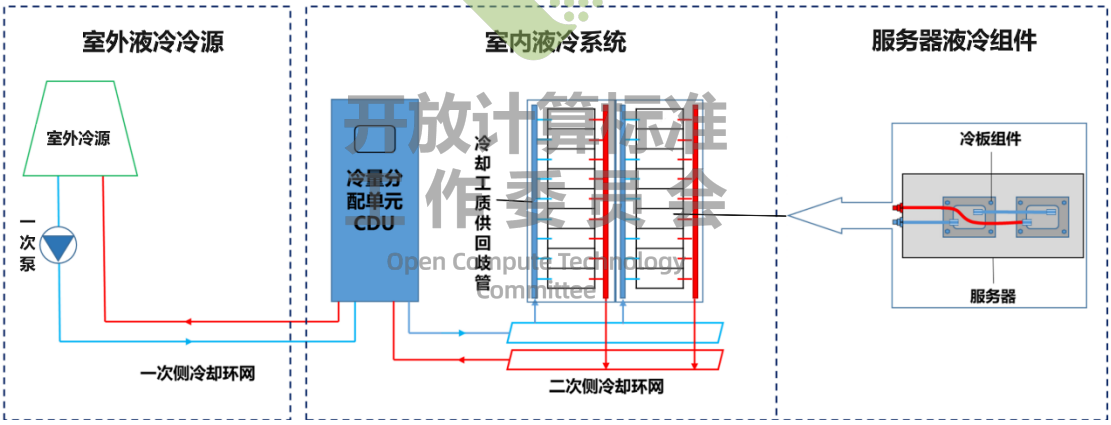


图16 液冷系统结构示意图

液冷系统具有优秀的散热能力，配方水作为液冷板载热材料，其具有空气3500倍的体积载热能力和20倍的导热系数。远高于空气的载热、导热能力，提高了带出服务器热量的品味，使得室外冷源实现了去压缩机化，液冷系统设备的换热方式得以简化、效率大幅提高。

液冷数据中对室外液冷冷源的需求和风冷数据中心也有不同。在换热上都支持冷塔，部分条件下支持干冷器表冷方案。不同点在于冷源温度方面，液冷数据中心支持露点温度以上宽范围调节，大于10℃以上的供回温差，甚至部分地区可支持干冷器自然冷却，相较于冷机方案，冷却系统更加简介高效，方案更加灵活多样。

6.3.2 算力设备技术框架

算力工厂可为用户提供定制化算力设备建设方案，包括智能算力方案、超算算力方案、通用算力方案和边缘算力方案。本文以智能算力场景为代表，介绍设备选型和技术方案。其他算力场景可根据客户实际需求定制，同时满足国产算力与进口算力设备的部署要求。

6.3.2.1 算力设备选型要求

在算力工厂建设时，需选用高性能、高可靠性的服务器设备，以满足海量数据的存储和处理需求。为满足各领域场景和复杂的AI模型的计算需求，AI服务器对计算芯片间互联、扩展性有极高要求。AI服务器内基于特定协议进行多加速器间高速互联通信已成为高端AI训练服务器的标准架构。服务器集群采用模块化方法构建，可以实现大规模的算力扩展。为了满足大模型训练常用的数据并行、模型并行、流水线并行等混合并行策略的通信需求，需要为芯片间和节点间提供低延迟、高带宽的互联。AI集群的构建需要配置面向AI优化的高速存储。通过配置高性能、高扩展、多层级的智能存储，为各种数据访问需求提供优化性能。智能存储具备随需扩展功能，实现高IOPS处理能力，支持RDMA技术，同时实现高聚合带宽。

6.3.2.2 算力设备技术方案

算力工厂的算力设备采用融合架构进行整体设计，在硬件层面，基于最新一代强劲高扩展AI训练平台和多元高弹性AI算力平台，融合多元算力，将不同类型的AI芯片进行统一管理和调度，结合新型超高速内外部互连技术、池化融合、异构存储介质等，形成高效池化的算力中心，实现不同计算资源的高效协同。配置成熟的并行存储系统，支持GPU显存直接存取，为AI计算提供高带宽、低延迟、高扩展性的存储服务；采用高速低延迟的RDMA计算网络，算力模块采用全线速无阻塞的胖树组网架构，可以支撑千亿参数级别的大模型训练并模块化扩展；部署管理节点承载智算中心操作系统，实现算力中心的智能化运维；配置完善的智算集群系统环境部署工具，实现训练、推理环境的快速部署；配置AI算力调度平台，实现多元芯片的统一管理和调度，提高集群利用率；部署算法基础设施，预置常见的模型和大模型算法，同时配置多模型管理平台，实现模型的纳管、评测、输出。

整个AI训练算力系统采用模块化、高性能、高扩展、最佳实践的设计理念，以实现不同规模的算力灵活扩展。每台服务器配置4或8张计算网卡，并分别连接到4个或8个InfiniBand交换机，使每个服务器上同一位置的IB网卡连接到同一个交换机，可以减少延迟跳数，降低通信延迟，提升通信效率。4台或8台IB交换机下的服务器组成一个可扩展单元。由可扩展单元构成的集群支持横向扩展，以满足千亿级参数大模型并行训练的算力和通信带宽需求。AI训练过程中会频繁的进行AllReduce操作，甚至有All2All操作。结合这些计算模式，尤其是针对Transformer大模型并行训练过程常用的AllReduce，采用该组网模式，以最小的跨网络跳步数为优化目标。

除了高速计算网络以外，部署带内管理网络和IPMI硬件监控网络。管理网络和监控网络通常运行基于TCP/IP的数据交换，用来连接所有的节点，进行节点操作系统系统的布署和软件的安装、全部节点硬件监控，实现并行文件系统的数据管理和软件分发功能等。

结合不同技术路线的性能、应用、生态等差异，算力中心建议采用技术先进、多元算力的技术路线，即采用通用GPU+国产化AI芯片的技术架构，满足性能和良好生态的共同需求，配置大部分的通用GPU算力，助力大多市场化AI应用，并配置适量的国产AI算力，作为国产AI应用的创新试验平台。在提供多元算力的同时，保证算力中心的高性能、高可扩展性和高可用性。

集群以生产算力、聚合算力、调度算力、释放算力为作业环节，主要为算力中心提供计算资源，管理平台和优化工具。其中硬件资源主要服务算力生产、算力聚合环节，包括计算设备和设施，包括AI服务器、通用计算服务器、高速网络、高速存储以及管理服务器和附属设备等。

开放计算标准工作委员会 Open Compute Technology Committee

生产算力：基于领先的AI服务器为算力机组，支持先进多样的AI芯片，支持成熟丰富的软件生态，形成高性能、高吞吐的计算系统，为AI训练和AI推理生产输出强大、高效、易用的计算力。

聚合算力：基于智能网络和智能存储技术，针对多任务、大规模、高并发、高吞吐的AI应用特点为算力机组集群构建高带宽、低延迟的通信系统和数据平台，提供弹性、可伸缩扩展的算力聚合能力。

调度算力：基于人工智能应用对算力的需求特点，通过虚拟化、容器化等技术，将算力资源池化为标准算力单元，通过适应性策略及敏捷框架对算力进行精准调度配给，保障AI开发和AI业务的高效运行。调度能力是智算中心连接上层应用与底层计算设备的核心能力，将聚合的CPU、GPU、FPGA、ASIC等算力资源进行标准化和细粒度切分，满足上层不同类型智能应用对算力的多样化需求，让上层应用更高效、更便捷地对算力资源进行利用。

释放算力：基于主流人工智能理论算法，采用全流程软件工具，针对不同场景应用需求，通过机器学习自动化的先进方法，产出高质量的AI模型或AI服务，提升AI应用生产效率，促进算力高效释放转化为生产力。

7 算力工厂运营模式

算力工厂的价值在于提供稳定、高效的算力服务。不仅要关注硬件的投入和维护，更应该注重软件能力、运营管理以及服务保障，以满足不同客户的需求。随着AIGC的发展，算力产业将迎来产业升级，服务将取代单纯的算力租赁成为业务主流。未来，算力中心95%以上的需求将来自于服务市场。

数据处理能力和推理服务是算力中心的两大核心功能。数据处理关乎模型训练的质量和速度，而推理服务则直接影响到业务逻辑的实时响应和执行效率。一般来说，算力工厂可以为用户提供以下几种服务：

1）基础设施服务：该服务模式与传统的数据中心托管类似，单应更多关注高效供电、散热以及网络可靠链接等。

2）数据及应用服务：包括训练服务、推理服务，以及数据的采集、清洗、标注、转换、存储等数据处理服务。

3）调度服务：为大模型等智算服务的运行提供弹性的计算、存储、网络等资源，以及服务的编排、保障、监控等功能。

整个服务体系，底层是预制化集装箱数据中心基础设施，中间层是计算和异构的支持，网络/RDMA，计算、存储、网络，顶层是算力调度、算力服务和服务的编排。算力中心的价值不仅在于提供空间、电力乃至算力，更在于将算力转化为推动产业发展的动力，以适应快速变化的技术和市场需求，解决算力服务面临的建设周期长、成本高，回收成本和盈利的问题。

7.1 算力运营

算力工厂采用商业化算力服务模式运营，提供按需扩容的计算服务，以算力服务为基础，以生态建设为目标，提供智能算力、超算算力、通用算力、边缘算力的数据处理和使用、算法模型、训练加速和推理等算力服务，计算资源依据每小时的使用量进行计费。收费标准将参考市场价格，采取灵活、多样、具有弹性的收费机制。国家和地方政府、企业的重大项目给予平台服务费补贴，返还服务费，设置试用窗口期等政策措施支持。

开放计算标准工作委员会 Open Compute Technology Committee

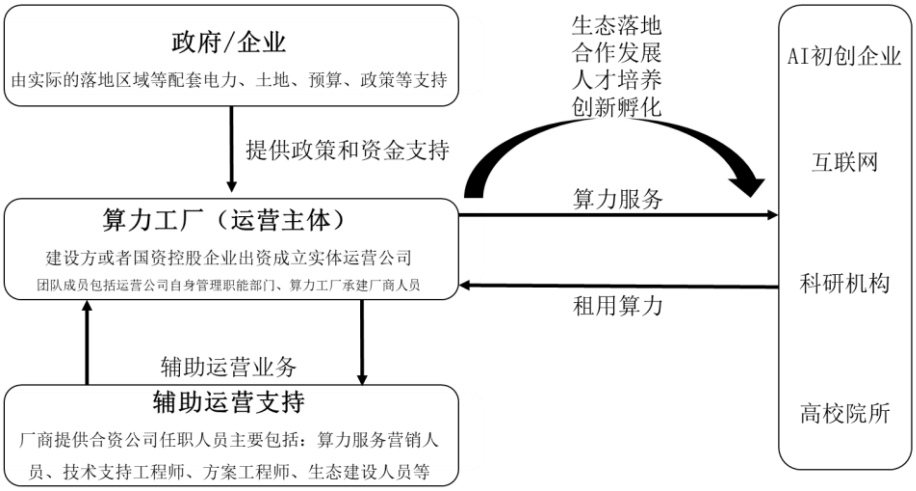


图17 算力运营

7.1.1 责任分工

政府/企业：

1.面向运营公司提供办公场所、政策及资金支持。

2.提供建设人工智能计算中心的配套土地、电力保障。

3.建设人工智能生态创新中心发展本地人工智能产业。

4.投入产业配套资金。

5.配套出台产业发展政策。

算力工厂（运营主体）：

1.完成中心人员招聘与培训。

2.完成中心实验室、办公室、展厅建设，软硬件平台搭建。

3.提供线上训练算力。成果进行产业转化。

4.设计和落地人工智能计算中心商业模式。

5.承担日常运营过程中产生的人力、水、电、物业、带宽等基础运营费用。

辅助运营支撑：

1.生态运营。通过生态适配、企业培训、协同创新、联合营销、资源开放、品牌活动、优质企业应用推广等方式，持续打造人工智能生态，引入系统化人才培养机制，为当地AI产业培养人才。

2.算力运营服务。产品运营、客户运营、伙伴运营、运营支撑。

7.1.2 运营内容

1）总体目标

因地制宜，结合本地实际人工智能发展情况，覆盖政府单位、企业、教育科研等单位及行业，打造符合当地业务需求的人工智能“产、学、研”生态圈，实现经济效益与科研水平的双重驱动。

2）算力运营

产品运营服务：运营商务模式设计，运营服务产品设计与发布，如试算流程设计、定价策略设计、优惠活动设计、增值服务设计等。

客户运营服务：运营目标行业和客户拓展，如客户商机管理、客户拓展活动。行成主要目标场景和客户清单。

开放计算标准工作委员会 Open Compute Technology Committee

伙伴运营服务：渠道和伙伴管理服务。确定渠道伙伴引入标准、管理流程、效果评估等工作。确保产品有效触达目标客户。根据地方产业特点，有目标选择行业伙伴，制定伙伴引入标准、流程，并对伙伴进行赋能，提高伙伴产品竞争力。

3）生态运营

公共算力服务：洞察本地人工智能产业发展现状，调研算力需求，制定算力普惠政策。

应用创新孵化服务：结合本地产业特点，针对人工智能重点应用场景，牵引科技创新成果商用转化、形成重大产品创新和示范应用。

产业聚合发展服务：持续开展技术沙龙、线下线上赋能培训、人工智能创新大赛、初创孵化、技术方案对接等生态活动。

科创人才培养服务：构建多层次、可持续发展的人才生态系统，实现 AI 核心技术突破及战略性发展产业生态体系协同创新。

7.2 算力销售

7.2.1 客户拓展

算力工厂聚焦应用创新孵化，牵引创新成果转化。结合重点应用场景，协助用户开展竞争性和先导性应用开发和场景试验，牵引科技创新成果商用转化、形成重大产品创新和示范应用，进一步带动产业的智能化升级。通过产业发展洞察，调研算力工厂所在地及周边城市算力产业发展现状和算力需求，形成咨询报告并支撑算力普惠政策修订，引导算力资源构建和分配，配合落实算力补贴政策，助力用户解决算法开发贵的难题，降低算法开发成本。面向政府、企业、教科研等用户提供普惠算力服务。

1）政府

支撑政府相关单位的智能化、数字化转型，通过人工智能技术底层赋能，有效提升工作效率。赋能区域产业升级，加强与传统经济的深度融合，助力政府人工智能产业加速，提升当地GDP。

2）企业

赋能当地及周边中小企业和AI初创企业，提供普惠算力，带动AI产业集约化发展。面向不同发展阶段的企业给与对应的资源、技术方面的支撑。强化龙头企业示范效益，引领骨干企业加速发展，扶持中小企业发展壮大。

3）教科研

支撑科研创新和AI人才培养，实现人工智能共性难题的突破。支撑区域内和全国范围内高校、科研院所在人工智能领域开展持续技术攻关和理论研究，高校设立人工智能专业和学科，培育出一批人工智能专业人才，提升当地人工智能教育水平。

开放计算标准工作委员会
Open Compute Technology Committee

7.2.2 计费模式

算力资源支持按需分配和独占使用两种方式，可选卡时计费、整机按月（年）租用计费等方式。

算力服务分为按需和独占两种计费模式。按需模式以（元/卡时）为计费方式，独占模式分为整机以（元/台月）、集群单议两种计费方式。

存储按需，以（TB/月）为计费方式。

数据、算法、服务、平台采用单议的计费方式。

提供免费的试算服务。

7.3 日常运维

为保障算力工厂的正常运行和持续可用性，高效运维服务必不可少。算力工厂可提供以下多元日常运维服务。

7.3.1 环境监控与设备维护

算力工厂环境监控主要包括温湿度、空调、电力、网络等关键设备的监控。定期检查算力工厂内的硬件设备，确保设备正常运行，及时发现并解决潜在隐患。此外，定期对硬件设备进行清洁、除尘，防止尘埃积累导致设备过热，也是环境监控的重要任务。

7.3.2 系统巡检与优化

系统巡检是算力工厂运维工作的重中之重。定期检查服务器、存储、网络设备等关键系统的运行状态，确保系统稳定可靠。同时，根据业务需求，对系统进行优化调整，提高资源利用率，降低能耗。

7.3.3 数据备份与恢复

数据是企业的核心资产，数据备份与恢复是算力工厂运维的关键环节。制定完善的数据备份策略，确保数据安全。同时，定期进行数据恢复演练，提高数据恢复能力，以防数据丢失造成的业务中断。

7.3.4 网络安全防护

网络安全是算力工厂运维的重要内容。密切关注网络安全动态，及时更新安全防护策略，防范各类网络攻击。此外，定期进行安全漏洞扫描和修复，提高算力工厂的安全防护能力。

7.3.5 业务监控与故障处理

对业务系统进行实时监控，发现并处理故障。对于突发故障，要快速响应，及时恢复正常业务。对于潜在的故障隐患，要提前排查，防患于未然。

开放计算标准
工作委员会
Open Compute Technology
Committee

参 考 文 献

[1] 《中国算力白皮书（2022年）》
[2] 《2023-2024 年中国人工智能计算力发展评估报告》
[3] 《中国综合算力指数（2023年）》

开放计算标准
工 作 委 员 会
Open Compute Technology
Committee