

超算互联网白皮书

超算互联网联合体

二〇二四年四月

序 言

过去 20 多年，中国的高性能计算事业取得长足的进步。在“机器、应用、环境”三方面协调均衡发展的方针指导下，自主研发的超级计算机 11 次位居世界超算 TOP500 第一，大规模并行计算应用三次获得世界高性能计算应用最高奖“戈登·贝尔”奖，基于互联网构建的国家高性能计算环境得到广泛应用，支撑了国家重要的科学研究和工程建设项目。

近年来，人工智能和大模型的发展对算力提出了前所未有的新需求，算力水平已经成为衡量一个国家科技与经济水平的重要指标。但另一方面，急剧变化的国际环境和日趋严重的外部限制与封锁，对我国高性能计算的可持续发展提出了严峻挑战。在外部严格封锁的条件下，软硬结合、系统优化、应用为先是打破困境的必然出路，以应用成效论英雄，从机器性能世界领先转向应用成效世界领先，应该成为我国高性能计算发展追求的新目标。

超算互联网正是在这个背景下提出的重要任务。所谓超算互联网，是一种基于互联网理念，借鉴互联网应用的成功经验而发展的计算基础设施。它追求计算应用资源的互通和共享，为最终用户提供优质的计算应用服务，帮助用户用计算解决应用问题。同时，它又具有商业模式，通过资源提供、应用运营、应用开发各方的通力合作，获得自身可持续发展的资金和资源。在科技部、工信部的指导下，已经成立

了国家超算互联网联合体，正在为超算互联网的建设而努力。

本白皮书的撰写和发布恰逢其时。首先，它将使联合体的成员进一步统一思想，凝聚共识，共商大计，明确做什么？怎么做？促进成员间目标和步调的一致。其次，通过论述超算互联网的技术架构、运行机制、应用形态和运营模式，有助于更加科学地开展超算互联网的顶层设计，系统性地安排资源弹性汇聚调度、应用平台和应用商店构建、应用软件开发部署机制等关键支撑技术的研发工作。第三，白皮书所阐述的理念和技术体系将指导超算互联网应用的开发，不断丰富超算互联网的应用资源，对最终用户提供更加优质的服务。最后，白皮书的发布也有助于社会公众了解什么是超算互联网，激发使用超算互联网完成自身业务的兴趣，拓宽各行各业对超算互联网的使用和参与。同时，也有助于潜在的投资主体了解超算互联网的理念和潜力，吸引多方投入共建超算互联网。

真切希望超算互联网白皮书的发布能发挥其在传播理念、凝聚共识、科学规划方面的独特作用，促进超算互联网的技术研发，促使超算互联网理念和愿景的早日落实，推动我国高性能计算技术和应用的水平迈上新台阶。

中国科学院院士、超算互联网总体专家组组长

钱德沛

2024年2月7日

前 言

数字经济时代，算力成为新型生产力。以计算为基础的数值模拟、数字仿真、大数据分析、人工智能等求解问题的方法，正带来新的生产、科研和治理范式。超级计算能力是衡量一个国家或地区科技竞争力和综合实力的重要标志，有赖于机器、应用、环境三位一体协调发展。超算互联网是要用互联网化的理念、思维、模式、技术、平台、机制来运营现有超算基础设施，实现降低算力使用门槛，提高资源利用效率，培育自主软件生态，促进算力泛在应用，推动跨界融合创新，提升超算环境服务能力等目的。

本白皮书在征求、吸收超算互联网联合体众多专家的意见建议的基础上，梳理当前我国超算基础设施与服务环境面临的机遇和挑战，提出超算互联网的基本概念与重要特征，描绘超算互联网的目标愿景与生态价值链，明确超算互联网应当具备的关键能力；描述国内外超算基础设施互联共享的发展现状，梳理超算网格、超算服务化、超算互联网的发展脉络；提出超算互联网的参考模型，梳理各层级功能特征与关键技术问题，明确制订超算互联网标准规范体系的要求；提出超算互联网的参考运营架构，明确超算互联网应具备的运行机制；展望超算互联网创新的使用模式及丰富的应用场景；最后，从技术攻关、应用示范、标准制订、人才培养、数据共享等方面提出行业发展倡议。

本白皮书还有诸多不足，期待各方批评指正，共话未来！

目 录

一、 发展背景	1
二、 基本概念	5
(一) 概念特征	5
(二) 目标愿景	6
(三) 关键能力	8
(四) 参考模型	12
(五) 概念演进	13
三、 发展现状	16
(一) 发达国家高度重视超算设施互联共享	16
(二) 中国推动超算网格向超算互联网升级	21
四、 技术架构	31
(一) 总体参考架构	31
1. 资源层	32
2. 平台层	38
3. 服务层	41
4. 应用层	44
(二) 关键技术问题	45
(三) 标准规范体系	50

五、 运营管理	53
(一) 运行机制	54
(二) 参考运营架构	56
(三) 超算互联网联合体	60
六、 应用展望	62
(一) 创新的使用模式	62
(二) 丰富的应用场景	63
1. 人工智能	64
2. 科学计算	66
3. 工程计算	70
七、 发展倡议	73
八、 附录	76
(一) 附录 1: 缩略语	76
(二) 附录 2: 参考文献	77

图/表目录

图 1 中国超算产业三位一体协调发展策略	3
图 2 超算互联网的重要特征	5
图 3 超算互联网支持多方生态协作	7
图 4 超算互联网应具备的关键能力	9
图 5 超算互联网参考模型	13
图 6 从超算网格到超算互联网的演进关系	14
图 7 EuroHPC (2021-2027) 发展策略框架	19
图 8 日本高性能计算基础设施(HPCI)分布	20
图 9 中国超算基础设施网络化发展历程	21
图 10 “十三五”国家高性能计算环境	23
图 11 超算互联网总体参考架构	31
图 12 超算互联网标准体系	51
图 13 超算互联网主要参与方及其关系	53
图 14 超算互联网运营商参考运营架构	57
图 15 国家超算互联网联合体正式成立仪式	60
表 1 EuroHPC JU 纳管的超级计算机列表	19

一、 发展背景

当今世界，人类社会正循着网络化、数字化、智能化的方向快速发展。人类社会空间、物理空间和信息空间日益紧密融合，以 5G、人工智能、先进计算为代表的新一代信息技术加速向各领域渗透。特别是近年来以 ChatGPT、AlphaFold2、Sora 为代表的新一代人工智能技术，作为变革性力量，带来全新的生产方式、科研范式和治理模式。先进计算成为求解问题、探索未知世界的重要方法，数值模拟、数字仿真、大数据分析、人工智能等手段普及应用，支撑科学研究、技术发明、产品开发、工业制造与社会治理。计算能力与算法、数据、模型等要素相互融合，成为数字经济时代的新质生产力。

近期，工业和信息化部等六部门印发《算力基础设施高质量发展行动计划》，国家发展改革委等部门印发《关于深入实施“东数西算”工程 加快构建全国一体化算力网的实施意见》，明确提出算力是数字经济时代的新型生产力。算力网是支撑数字经济高质量发展的关键基础设施，可通过网络连接多源异构、海量泛在算力，实现资源高效调度、设施绿色低碳、算力灵活供给、服务智能按需。在国家实施“东数西算”战略的背景下，以新型数据中心、超算中心、智算中心为代表的算力基础设施正加速部署。截止目前，我国已启动建设 8 大国家级算力枢纽节点和 10 个数据中心集群，已建成十多个国家超级计算中心，超 40 个城市在建智算中心，25 个国家新一代人工智能公共算力开放创新平台获批建设，全国一体化算力网正加快构建。

超级计算能力是衡量一个国家或地区科技竞争力和综合实力的

重要标志。超级计算已广泛应用于航空航天、工业仿真、气象环保、油气勘探、人工智能、新能源、新材料、生物医药、智慧城市、基础科研等众多领域，在应对重大挑战性问题，促进传统产业转型升级，提高人民生活水平，促进重大科学发现等方面发挥着不可替代的作用。多年来，在中国科技计划的持续支持下，我国超算产业坚持“机器、应用、环境”三位一体协调发展策略，在机器研制、关键应用、服务环境等方面已处于世界前列，“神威”、“天河”、“曙光”系列超级计算机已多次占据全球超级计算机性能 TOP500 领先地位，大气模拟、地震模拟、量子模拟、分子动力学模拟等关键领域大规模并行应用获得戈登·贝尔（Gordon Bell）奖，全国范围建成十多个国家超级计算中心、中国国家网格（China National Grid, CNGrid）等战略性信息基础设施。同时，总体上看，我国高性能计算软件与应用的发展相对不足，应用生态也不够丰富，高性能基础和应用软件大量依赖国外软件，存在“卡脖子”的风险¹。建立发展高性能计算的生态环境，需要形成一个从基础研究、技术突破、产品研发到应用推广的协作共同体，不是简单的链条，而是相互关联的社会网络²。

¹ 钱德沛, 栾钟治, 刘轶. 从网格到“东数西算”: 构建国家算力基础设施[J]. 北京航空航天大学学报, 2022, 48(9):14.

² 李国杰. 发展高性能计算需要思考的几个战略性问题. 中国科学院院刊, 2019. 6

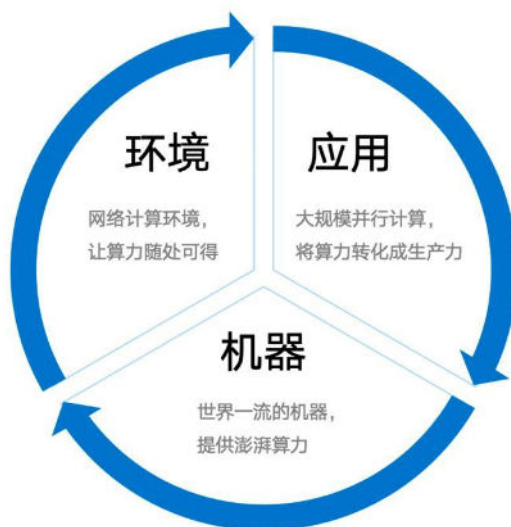


图 1 中国超算产业三位一体协调发展策略

随着数字化转型不断深入以及国际地缘政治形势变化，我国超算产业面临新的机遇与挑战：

- 人工智能带来 AIGC、大模型、AI For Science、AI For Technology 等新兴应用，催生巨大算力需求，并驱动超级计算与人工智能融合发展；
- 云计算、APP Store 等互联网新平台新应用冲击传统的超算资源管理与用户服务模式，超算中心亟需从“卖机时”向“提供应用服务与解决方案”升级；
- 全国范围内不同体系结构、归属不同运营主体的多源异构算力需要互联、集成与统一调度，以实现资源高效利用；
- 随着更多超算中心、智算中心建成，需探索形成更加有效的竞合机制与商业模式，实现计算服务行业可持续发展；
- “东数西算”背景下，大数据洪流(data deluge)带来的数据传输问题备受关注，需升级网络基础设施，优化大规模数据

远程传输效率；

- 中美贸易战、科技战凸显新时期我国解决芯片、基础软件等“卡脖子”问题，加快建设自主软硬件生态的紧迫性。

上述问题与挑战，对我国进一步发展超算基础设施及其服务环境提出了更高要求。

放眼世界，构建具备高效的数据传输、资源共享、算力调配、生态协作能力的超算基础设施网络，也是发达国家塑造高性能计算竞争优势的重要策略，例如美国的能源科学研究网络 ESnet (Energy Sciences Network)、欧盟的高性能计算共同体计划 EuroHPC JU (The European High Performance Computing Joint Undertaking)、日本的高性能计算基础设施 HPCI (High Performance Computing Infrastructure) 等。2023年4月17日，国家超算互联网正式启动部署，并成立了国家超算互联网联合体。超算互联网将在以往中国国家网格 CNGrid 以及业内高性能计算云 (HPC Cloud) 实践的基础上，全面利用互联网化的理念、思维、模式、技术、平台、机制来运营超算基础设施，打造具备互联网理念与特征的新一代高性能计算服务环境，实现降低算力使用门槛，提高资源利用效率，构建自主软件生态，培育科技人才队伍，促进算力泛在应用，推动跨界融合创新，提升超算环境服务能力等目的。

值得一提的是，本白皮书讨论的超级计算，是相对广义的概念，涉及超级计算、高性能计算、智能超算等，包括硬件、软件、算法、应用、产业生态环境。

二、基本概念

(一) 概念特征

在全球新一轮科技革命和产业变革中，互联网与各领域的融合发展具有广阔前景和无限潜力，已成为不可阻挡的时代潮流。超算互联网是具备互联网理念与特征的超算基础设施，是互联网化的高性能计算服务环境，是互联网创新成果与算力基础设施运营的深度融合。超算互联网不仅要在各算力中心之间形成高效数据传输网络，更要构建和完善全国一体的算力调度网络和面向应用的生态协作网络。

超算基础设施深度融合互联网的理念、思维、模式、技术、平台与机制后，将呈现以下重要特征：

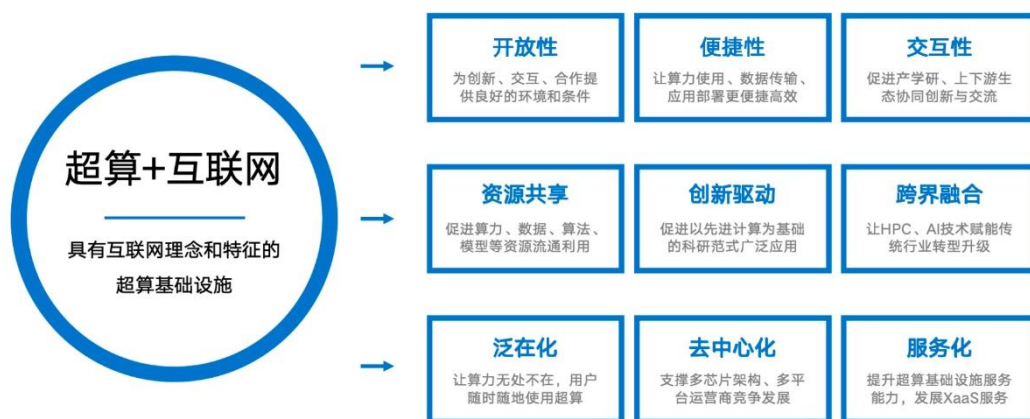


图 2 超算互联网的重要特征

开放性：通过类电商平台、应用商城（APP Store）、开源社区等平台，为创新、交互与合作提供开放包容的环境和条件；

便捷性：通过高速网络、云、XaaS、移动 APP 等技术和服务模式，让算力使用、数据传输、应用服务更加高效便捷；

交互性：通过线上交易、线上社区、线上培训等机制模式，促进供需方、产学研、上下游之间广泛交流与协作；

资源共享：利用网格、云、容器、隐私计算、XaaS、区块链等技术，促进算力、数据、算法、模型等资源共享与流通；

创新驱动：大幅降低超算使用门槛，支持丰富应用场景，让更多用户便捷使用模拟仿真、大数据分析、AI等新手段，进行科学研究和工程技术创新；

跨界融合：通过面向工业、材料、生物、能源等行业的领域应用平台，让HPC、AI技术高效赋能行业转型升级；

泛在化：实现跨区域、跨架构、跨层级、跨运营主体的一体化算力调度，让算力无处不在，用户可按需使用超算；

去中心化：支持多体系架构、多算力中心、多服务平台之间的互联互通与竞合发展，形成包容并蓄、共生共荣的产业生态；

服务化：提供硬件、软件、数据、模型的XaaS化服务，提升超算基础设施综合服务能力，优化用户体验。

(二) 目标愿景

超算互联网以高效安全的数据传输网络为基础支撑，以互联网的思维运营超算基础设施为核心理念，以构建市场化、领域化、服务化、标准化的高性能计算服务环境为发展方向，以供需紧密对接的“类电商”应用服务平台为主要载体，广泛连接算力供给、软件开发、数据交易、模型服务、行业应用等产业链相关各方，促进实现基础设施高效利用、异构算力调度共享、软件硬件协同创新、人才队伍发展壮大、产业生态自立繁荣等多层目标，有效支撑我国超算技术、应用、人才、产业的长期可持续发展。

超算互联网应致力于构建“平等、互利、协作、创新、共赢”的生态价值链：



图 3 超算互联网支持多方生态协作

最终用户通过超算互联网按需获取计算、存储、软件、数据、模型等各类资源与服务，降低使用门槛与成本，专注自身业务创新；

超算中心通过超算互联网实现互连与资源共享，触达更多用户，提高资源利用率，打造差异化、有竞争力的领域应用服务平台，从机时服务向应用平台服务升级，实现可持续发展；

软件开发者通过超算互联网灵活、便捷地开设应用商店，为更多用户提供软件产品服务，线上直接获取用户反馈，促进自身产品快速迭代，提高自主软件竞争力；

超算研发机构通过超算互联网支持新技术验证与新产品开发，促进不同架构加强协作，推动多平台多架构基于统一接口互联互通；

高校与科研机构通过超算互联网实现算力、数据等资源共享交换，促进算法、模型、软件等科研成果转化，促进产教融合，支持学生参与技术竞赛、众包众筹、开源社区等公益项目；

第三方服务商通过超算互联网为用户提供技术咨询、软件外包、技术培训、数据交易、居间代理等丰富多样的专业性服务，协同超算中心、软件开发商做大计算服务业规模；

地方政府通过超算互联网整合区域内算力、数据、人才、应用等资源要素，构建超算互联网区域枢纽节点，融入全国一体化算力网络，推动产、学、研、用深化协作，促进重点产业转型升级，培育发展具有区域特色优势的先进计算软件与服务产业，提高区域科技创新水平与产业竞争力。

(三) 关键能力

为支撑实现上述价值目标，作为新型超算基础设施，超算互联网应致力于构建和发展以下关键能力：



图 4 超算互联网应具备的关键能力

1. 强大的基础资源服务能力

提供世界一流的软硬件资源和多种国产芯片算力，通过全国一体化算力管理调度核心系统实现资源可测量、可调度，并建立高效的数据传输优化网络，为资源共享与协作提供支持。

- 具有世界一流的软硬件资源，接入、汇聚所有国家级超算中心及具有区域或行业优势的各类算力中心的资源，支持纳管多种国产芯片算力，总聚合算力达到千亿亿次级；
- 建成全国一体化算力管理调度核心系统，拥有全局资源可测量可调度能力，支持跨架构、跨地域、跨层级、跨运营主体的资源管理与任务调度，实现全局统一存储服务，屏蔽超算中心底层硬件差异，提供标准化计算服务；

- 形成全国性大规模数据传输优化网络，超算中心间实现高速安全互连，为大规模资源共享与协作、海量数据传输、“东数西算”、分布式协同计算等任务提供高效网络支撑。

2. 丰富的应用场景支撑能力

为各行业领域用户提供解决方案，并通过应用服务平台促进供需对接，降低超算门槛。

- 支持 HPC、人工智能、大数据、云等丰富应用场景，具备为用户提供解决方案的服务能力，能够帮助各行业领域用户解决实际问题；
- 形成面向最终用户的业务应用服务体系，建成简单、高效、易用的“类电商”应用服务平台，打造应用商城(APP Store)，提供各类应用集成、封装、发布、部署与交易的全流程服务，促进供需广泛对接；
- 面向关键行业、重要领域建成一批领域应用服务平台，支持开发、集成一批领域应用软件，支持用户基于工作流和领域知识快速开发部署业务应用，大幅降低超算门槛。

3. 自主生态体系支撑能力

提供自主生态体系支撑，包括建立高性能计算机软硬件生态、公共软件资源库和统一开发接口等。通过超算互联网行业联合体，形成全产业链合作网络。

- 能够支撑建立我国自主技术的高性能计算机软硬件生态，解决高性能计算机应用的瓶颈问题，能够运行面向国产计算芯片

的基础软件、工具软件和开发环境,具备丰富应用软件资源;

- 建成公共的共性软件资源库,包含数学库、算法库、工具库、求解器、模型库等,为应用开发与集成提供高效支撑,促进高性能并行软件研发,持续提高国产软件的数量和质量;
- 支持研发推广统一编译器、统一数学库、统一开发接口与规范、自动移植工具等,屏蔽底层硬件架构的差异性,降低软件移植成本,提高移植效率,实现软件一次编程后跨中心跨架构平滑运行;
- 持续发展超算互联网行业联合体,探索新型产业协作机制,在产学研之间、软硬件之间、上下游之间、多平台之间形成去中心化的全产业链合作网络。

4. 高水平人才队伍保障能力

人才是超算互联网成功的根本保障,在基础设施高效管理、服务平台稳定运行、并行软件开发推广、应用解决方案研发部署等方面,具备高水平人才队伍保障能力。

- 通过在线知识库、在线培训、并行计算竞赛、开发者大会等多种线上或线下方式,大力推广并行计算编程技术,形成国内活跃的并行计算人才社区;
- 支持产教融合,长期、持续培养面向国产芯片的超算软件开发与复合型应用人才;
- 加强国际合作,通过联合科研、学术会议、超算竞赛等多种方式,促进海内外人才交流与经验分享。

5. 可持续循环发展能力

通过提供优质服务、技术转化和产业投资等方式获取经济效益。支持超算中心向提供应用平台服务转变，并实现用户规模扩张。通过构建应用服务平台和与其他算力网络的互联互通，适应不同超算架构，形成共赢的商业模式。

- 在政府支持的基础上，超算中心能够通过对外提供优质服务、技术成果转移转化、科创产业投资等多种方式获取经济效益，实现自身循环可持续发展；
- 支持超算中心实现从提供机时服务向提供应用平台服务的根本性转变；
- 用户规模不断扩大，数量达到百万级规模；
- 以构建高效运转的“类电商”应用服务平台为抓手，形成超算中心、超算软件开发者和超算用户多方共赢的商业模式；
- 形成适应不同超算架构的超算互联网标准体系；
- 与中国算力网等其他算力网络实现互联互通与协调发展。

(四) 参考模型

本白皮书提出超算互联网的参考架构模型（如图 5），包括资源层、平台层、服务层和应用层，各层遵循统一的运营、安全与标准体系。白皮书第四章详细阐述了该参考架构的具体内容。



图 5 超算互联网参考模型

资源层：以高速网络连接分布各地的算力基础设施，聚合多个中心的软硬件资源，形成可共享调度的基础资源池。

平台层：构建基础核心能力，承上启下提供能力支撑，向下支撑实现资源的接入、管理与调度等功能，向上支撑实现作业管理、模型/数据/工具集管理、用户管理、交易管理和商城管理等功能。

服务层：依托平台层能力，面向科学计算、工程计算、智能计算等领域用户需求，提供算力服务、应用服务、模型服务、数据服务、社区服务等各类服务。

应用层：面向各行业领域用户，支持丰富的应用和应用场景。

运营、安全和标准：构建统一的运营服务体系，保障网络与数据安全，形成统一的标准与规范，并贯通各层。

(五) 概念演进

超算网格是与超算互联网密切相关的重要概念。超算网格可以视作超算互联网的前身，为超算互联网发展提供了一定的支撑条件。

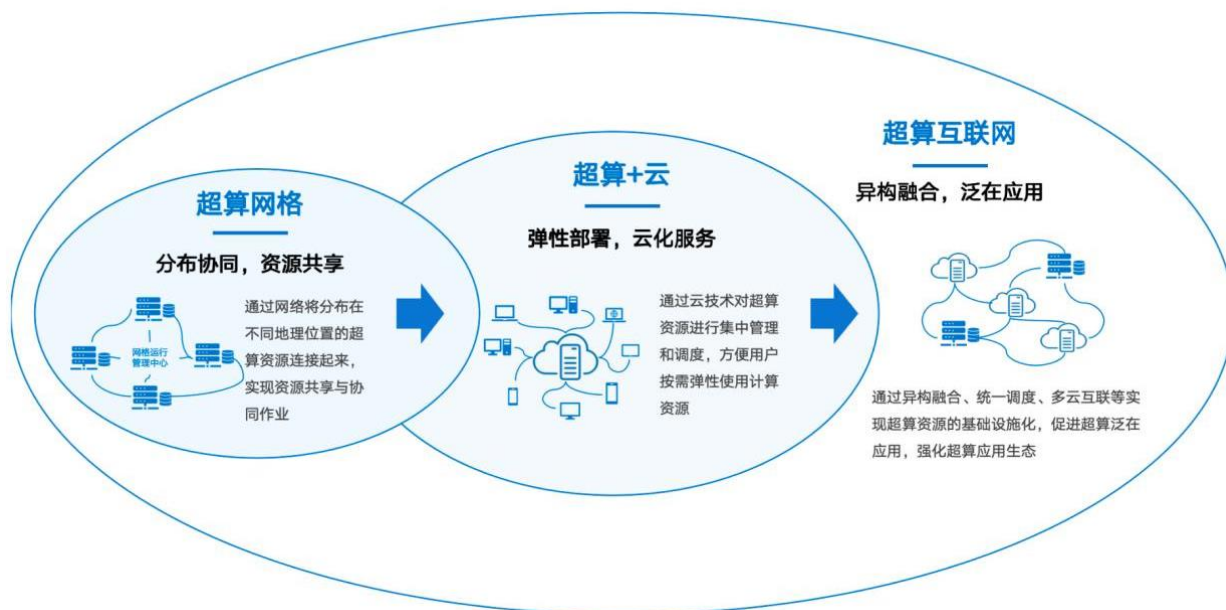


图 6 从超算网格到超算互联网的演进关系

超算网格 (Computing Grid) 的概念出现在上世纪九十年代, 指通过网络将分布在不同地理位置的超算中心连接起来, 建立一个类似电力网 (Power Grid) 的计算资源网络³, 聚合大规模的计算能力, 支持资源共享、任务分发与协同计算。例如, 美国的 TeraGrid、欧盟的 EGEE 和中国的 CNGrid, 都在学术科研领域成功构建了大规模资源共享协作环境。

超算网格在屏蔽不同超算中心差异, 实现统一作业调度等方面, 可为超算互联网提供部分底层设施和技术参考。超算网格主要服务于学术科研领域, 而超算互联网的目标是要将超算能力拓展至千行百业, 因此需要联结更多算力中心和上下游企业, 管理更多类型的资源和作业, 服务更大规模的用户群体, 网络乘数效应更加明显。

高性能计算云 (HPC Cloud) 的概念出现于本世纪初, 肇始于亚

³ 钱德沛, 栾钟治, 刘轶. 从网格到“东数西算”: 构建国家算力基础设施[J]. 北京航空航天大学学报, 2022, 48(9): 14.

马逊等互联网企业在公有云平台上为用户按需弹性提供高性能计算、存储服务。同时近年来大多数超算中心也开始提供云化的 HPC 服务，云计算与高性能计算的结合愈加紧密。超算云可以看成是传统超算的“云化”演进，是以超算中心资源作为底座，基于统一的资源调度平台，通过云计算的服务模式为用户提供高性能计算服务⁴。

超算云一般由单一主体对外提供服务，而超算互联网的核心理念是通过类电商平台联结多个主体，管理调度多源、异构、海量的资源。云计算无疑是超算互联网的关键技术与服务模式之一，但超算互联网的目标并不止于提供云化的机时资源，更需构建上层应用商城，解决用户业务应用问题。同时，超算互联网作为国家级基础设施，需全力支持解决生态适配、软件移植、产学研协作、人才培养等诸多行业共性问题 and “卡脖子” 问题。

⁴ 超级计算创新联盟. 高性能计算云 (HPC Cloud) 白皮书, 2021.9

三、 发展现状

大规模超级计算机造价昂贵，动辄需投入上亿美元，除石油、汽车、互联网等领域少数巨头企业外，中小企业、高校和科研机构等用户无力投资。因此，由政府出资建设，然后委托国家实验室、高校或专业机构运营，成为各国部署超级计算机的主流模式。发达国家或地区长期推进建设超算基础设施网络，促进互联互通、资源共享与业务协作，不断提高资源利用率和投资回报。

（一）发达国家高度重视超算设施互联共享

1. 美国相关情况

目前，美国拥有世界上最强大的超算基础设施互联互通与协作网络，包括商业性较强的亚马逊 AWS 云、微软 Azure 云等，以及政府支持建设的公益性设施网络，如能源科学网络 ESnet6、开放科学网格（Open Science Grid, OSG）等。

早在 1986 年，美国国家科学基金会 (NSF) 构建 NSFNET 网络，允许全美科研人员连接和使用来自 6 家超级计算中心的资源。NSFNET 在 1998 年停止运行后，NSF 又在 2001 年组建了世界上最大规模的分布式计算网格 TeraGrid（2011 年被 XSEDE 计划替代），2005 年又资助建设了开放科学网格（OSG）。

美国能源部（DOE）在 1986 年也组建了其专属科研网络 ESnet，连接下属的国家实验室、大学和科研机构，旨在让数以万计的全国科研人员不必受物理位置限制，高效传输科研数据，共享使用超级计算、仪器设备等各类资源。最新一代网络 ESnet6 已于 2022 年 10 月发布。

本世纪以来，美国政府先后启动全球信息网格计划（GIG）、极大规模科学与工程探索 and 发现环境计划（XSEDE）、国家战略性计算计划（NSCI）、先进科学计算研究计划（ASCR）、先进计算生态系统（ACE）战略计划、研发基础设施国家战略（RDI）、美国国家人工智能研究资源（NAIRR）等一系列战略规划，其重要目的之一就是要充分利用、连接和增强美国的计算与网络基础设施资源，持续发展壮大先进计算生态体系，以保持美国在高性能计算领域的全球领导力和竞争优势。

同时，在全球商业领域，亚马逊、微软等美国企业在各自公有云平台上打造了日臻成熟的高性能计算即服务（HPCaaS）业务模式，引领着全球高性能计算云的发展趋势。以亚马逊为例，其 AWS 高性能计算服务已广泛应用于处理复杂、数据量大的计算任务，包括基因组学、计算化学、地震处理、计算机辅助工程以及深度学习和自动驾驶等诸多应用领域。

2. 欧盟相关情况

欧盟一直致力于在政治、经济、科技等领域推进欧洲共同体的愿景实现。在超级计算领域，欧盟先后实施欧洲先进计算伙伴计划（PRACE）、欧盟地平线（Horizon Europe）计划、欧洲高性能计算技术平台（ETP4HPC）、欧洲极限数据与计算项目（EXDCI）、“开放科学云”行动计划（EOSC）、欧洲高性能计算联合体计划（EuroHPC）等，联合各成员国共同发展资源共享的高性能计算基础设施网络。

在网格计算方面，欧盟曾先后启动欧洲网格（EuroGrid）、数据网格（European DataGrid）、电子科研网格（Enabling Grids for E-

science in Europe, EGEE)、网格基础设施(European Grid Infrastructure, EGI)等一系列项目⁵,旨在建立可持续运维的泛欧计算基础设施,供科研人员全天候使用。

欧洲高性能计算联合体(EuroHPC JU)始建于2018年10月,其宗旨是部署和运行世界级的高性能计算和数据基础设施,推动建设卓越的欧洲高性能计算生态环境。根据欧盟理事会于2021年7月正式通过的“欧洲高性能计算联合体(EuroHPC JU)项目条例”,2021—2027年,欧盟将开发、部署、扩展并维护世界领先的超级计算、量子计算、服务和数据基础设施生态系统;开发并应用需求导向、用户驱动的创新型且具有竞争力的超级计算系统;扩大超级计算基础设施开放范围,公共机构和私营部门均可使用,以支持欧洲科学界和产业界的数字化和绿色化双重转型以及关键技能发展⁶。

⁵ 钱德沛,栾钟治,刘轶.从网格到“东数西算”:构建国家算力基础设施[J].北京航空航天大学学报,2022,48(9):14.

⁶ 袁珩,张丽娟.“数字欧洲计划”2021—2022年度工作要点.科技参考,2022,6

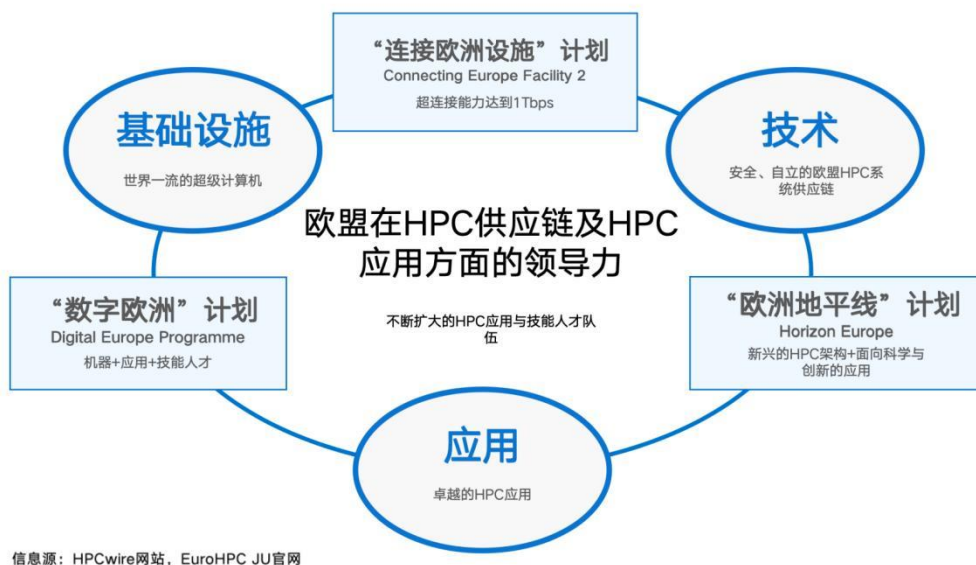


图 7 EuroHPC (2021-2027) 发展策略框架

EuroHPC JU 的具体目标之一，是将欧盟层面和成员国层面的高性能计算资源整合一个通用平台上，向公私部门用户开放，提供云端超算、量子计算及数据基础设施服务，推动高性能计算被各类社群用户广泛使用。截止 2024 年 1 月，EuroHPC JU 已纳管来自 9 个成员国的世界级超级计算机，包括欧洲第一台 E 级超级计算机 Jupiter。

表 1 EuroHPC JU 纳管的超级计算机列表

序号	机器名称	持续性能 petaflops	峰值性能 petaflops	部署国
1	LUMI	386	539.13	芬兰
2	LEONARDO	246.54	313.9	意大利
3	MARENOSTRUM 5	178.3	311.95	西班牙
4	MELUXINA	12.81	18.29	卢森堡
5	KAROLINA	9.59	12.91	捷克
6	DISCOVERER	4.52	5.94	保加利亚
7	VEGA	6.92	10.05	斯洛文尼亚
8	DEUCALION	7.22	10	葡萄牙
9	JUPITER	1000	N/A	德国

信息来源：EuroHPC JU 官网

3. 日本相关情况

日本从上世纪 80 年代中期开始发展超级计算机，在机器性能方面屡次超越美国。2003 年，在日本教育科研网 SuperSINET 基础上，日本文部省启动实施“国家研究网格基础设施”(NAREGI)项目，研制部署面向科学研究的网格基础设施。继 NAREGI 项目之后，日本又开始实施“下一代超级计算机计划”，启动建设日本高性能计算基础设施网络 HPCI⁷。



图 8 日本高性能计算基础设施(HPCI)分布

HPCI 网络由日本信息科学与技术研究组织 (RIST) 负责运营管理，通过将分散在日本高校、研究机构的超算中心计算与存储资源通过高速网络连接起来，形成国家级的共享计算环境，满足各类用户不同需求。目前，HPCI 已经连接日本理化学研究所(RIKEN)、东京大学、筑波大学等 15 个超算中心，其中包括日本最先进的超级计算机“富

⁷ 钱德沛, 栾钟治, 刘轶. 从网格到“东数西算”: 构建国家算力基础设施[J]. 北京航空航天大学学报, 2022, 48(9): 14.

岳” (Fugaku)。HPCI 纳管的计算资源类型异构多样，涵盖通用 x86 CPU、ARM 架构的富士通 A64FX CPU、GPU 以及 NEC 的矢量加速芯片等。

日本政府还同步资助重点工业企业使用高性能计算资源，促进工业仿真模拟技术普及应用。HPCI 网络不仅帮助日本加速科学突破与技术进步，也促进了产业竞争力提升、人力资源开发和用户基础扩大。

(二) 中国推动超算网格向超算互联网升级

与发达国家地区不同，我国超算基础设施的公共服务平台属性更强，普遍要求向社会开放，服务支撑众多领域。20 多年来，以超算中心为基础的高性能计算环境大致经历了超算网格、超算服务化、超算互联网三个发展阶段。各阶段的侧重点不同，其中，超算网格聚焦于基础资源的共享与协同，超算服务化强调通过云计算、领域应用服务平台等方式提升超算基础设施服务能力，超算互联网凸显了新形势下建立新型产业协作机制与自主软件应用生态的紧迫性。

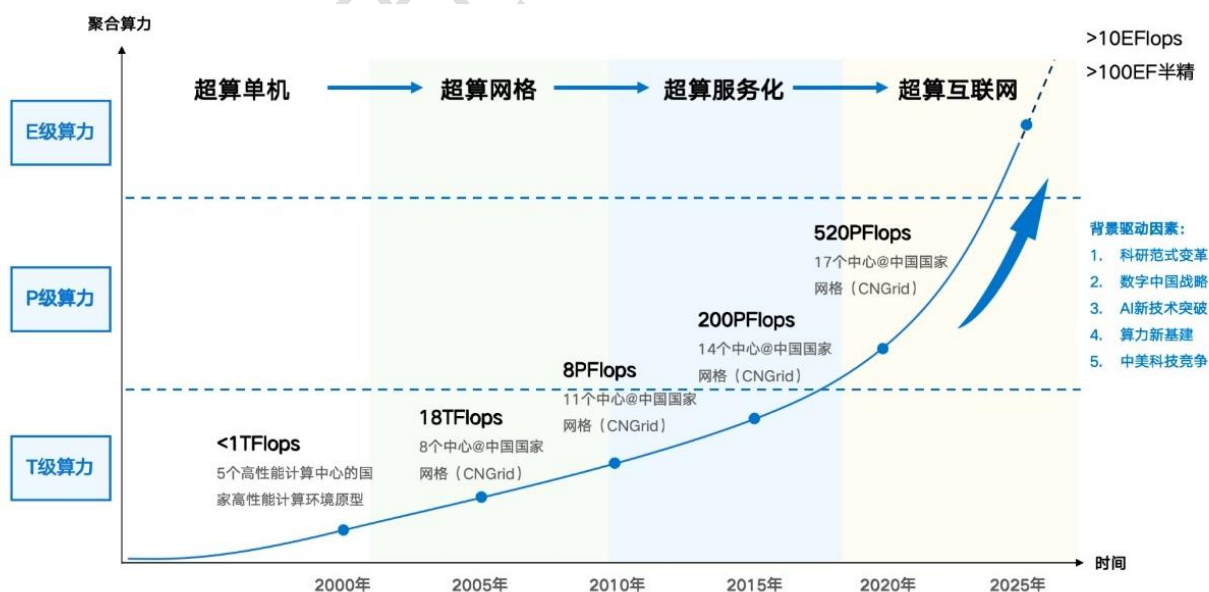


图 9 中国超算基础设施网络化发展历程

1. 超算网格

21 世纪初，网格计算在中国兴起。1999 年-2000 年，863 计划支持建立了由 5 个高性能计算中心构成的国家高性能计算环境原型，形成了超算网格的雏形。2004 年，在教育部支持下，由华中科技大学牵头，启动建设中国教育科研网格，研制网格核心中间件，集成中国教育和科研计算机网(CERNET)上自治的、分布异构的海量资源。到 2006 年，中国教育科研网格集成分布于全国 13 个省市 20 所重点高校的计算、存储、数据、软件等资源，计算能力超 15 万亿次，支持了生物信息学、图像处理、远程教育、流体力学等领域网格应用，成为当时全国高校公共服务体系的重要基础设施。

同期，国家 863 计划启动建设国家高性能计算环境实验床，即中国国家网格 CNGrid，成立中国国家网格运行管理中心。CNGrid 于 2005 年 12 月正式开通运行，首批汇聚中科院超算中心、上海超算中心等 8 个结点，形成 18 万亿次计算能力。在国家科技计划的持续支持下，研发了国家高性能计算环境系统软件 CNGrid GOS 和 CNGrid Suite，实现了资源的互联互通与统一共享、作业的提交与全局调度、数据的全局管理和环境的安全管控。截止 2022 年，CNGrid 已聚合 22 个结点的计算资源 871PF 和存储资源 538PB，部署应用软件和工具超 740 份，研究团队用户数超过 7000 个，支持领域应用社区和业务平台 23 个，覆盖高能物理、计算化学、材料科学、生命科学、地球科学、天文学、工业设计、能源勘探、教育等众多领域。CNGrid 通过资源共享机制，为科学研究、工程设计、产品开发、信息化建设等各类用户提

供高性能计算与数据处理服务，累计支持千余项科研项目。

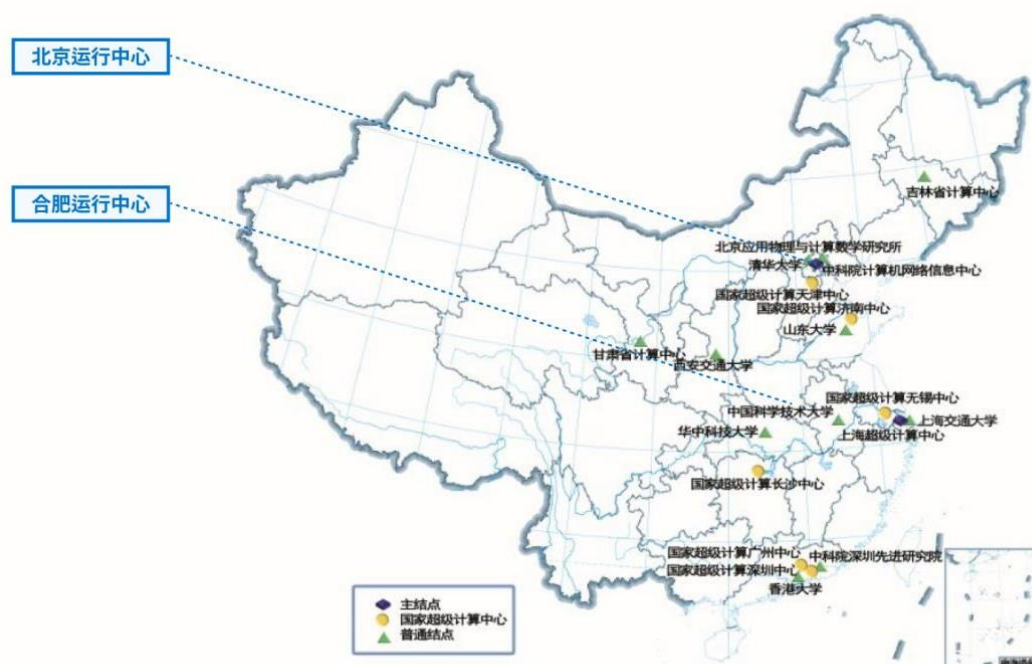


图 10 “十三五”国家高性能计算环境⁸

CNGird 在国家科技计划支持下历经 20 余年发展，已经成为不可或缺的战略信息基础设施，其资源能力与服务水平进入世界前列，在一定程度上促进超级计算机从单机模式向网络计算模式的转变，并在航空航天、石油勘探、生物医药、工业设计等多个关键领域支撑了一批重要应用，为“东数西算”背景下进一步建设超算互联网积累了宝贵经验，提供了技术基础。

2. 超算服务化

随着云计算的蓬勃发展，传统高性能计算开始向云化服务模式转变，超算云化成为行业发展方向。目前，大部分超算中心都已建有云

⁸ 钱德沛. 构建支撑科技创新的新一代计算基础设施[J]. 数据与计算发展前沿, 2020, 2(1): 1-17

资源管理系统或云服务平台，能够为用户按需提供可弹性扩展的机时资源、SaaS 软件等云化服务，降低超算使用成本和门槛。例如：

- 国家超级计算长沙中心自研云服务平台，对不同资源池进行统一监控、管理和调度，实现用户资源的快速弹性扩展，支持多维度角色权限管理，提供租户虚拟化层隔离、租户内部访问控制等功能。
- 并行科技作为独立的超算云服务提供商，将部分超算中心的计算资源整合为统一云计算资源池，构建了“并行超算云”一站式服务平台。
- 甘肃省计算中心的高性能计算云服务系统，涵盖作业调度、资源管理、应用安装、用户管理、系统监控等功能模块。
- 贵安超算中心的 Renderbus 云渲染平台，可对外提供图像渲染云服务。
- 中国航天科工集团智能科技研究院的信创云环境超算服务平台，可以为智能体训练、智能仿真等场景提供基于国产 GPU 的算力服务等。

在超算云基础上，部分超算中心面向行业应用场景需求，结合“互联网+”理念，构建了多个领域应用服务平台，汇聚整合计算、存储、数据、应用软件等多方面资源，在线为用户提供更完整的应用服务或解决方案。例如：

- 中国科学院计算机网络信息中心的生物医药应用服务社区 (biomed.cngrid.org)，汇聚 50 余种主流科学应用软件，涵

盖生物信息、药物靶标鉴定和药物设计等领域，支持基因组分析、转录组分析、蛋白质分析、药物分子设计等应用。

- 上海超算中心的 Xfinity 工业高性能计算服务平台 (xfinity.ssc.net.cn)，于 2010 年 4 月上线，截止 2020 年 9 月，十年间完成作业数超 43 万个，使用机时数超 4790 万核时，在汽车、飞机、钢铁、高端装备设计等领域为上千家企业用户提供仿真模拟计算服务。
- 国家超算天津中心的天河设计与仿真平台 (cae.th-aicloud.cn)，涵盖几何模型设计、求解计算、后处理分析等环节，满足结构、流体、电磁等不同领域的云端设计仿真一体化需求。
- 国家超算天津中心的高通量材料计算平台 (mathtc.nscctj.cn)，集成材料计算软件、前后处理和数据可视化模块，拥有面向材料计算的全链条算法工具集成和统一接口技术，支持多组元材料体系从微观、介观到宏观的结构、物性和服役行为的多尺度计算仿真技术，可实现材料计算设计的全流程自动控制。
- 国家超算无锡中心的神工坊高性能仿真平台 (simforge.cn)，建于 2016 年，是支持工程仿真应用的开发、定制和在线服务的 PaaS+SaaS 平台。神工坊已实现航空发动机等领域极高保真度数值模拟，并行规模达到 100 万核心，网格分辨率达到 10 亿级别，引领行业发展。

- 甘肃省计算中心的甘肃省生物医药高性能计算示范平台，将生物医药类计算软件以更友好、更易用的 Web 方式提供给用户，用户无需安装配置环境，即可通过网络访问各种服务。
- 陕西空天超算构建的航空航天领域应用服务平台，可提供结构、流体、燃烧、焊接等高性能计算服务。

超算云和领域应用平台的建设，一定程度上改变了超算中心的传统服务模式，提升了高性能计算环境的综合服务能力，为建设超算互联网提供了样板经验。但目前各平台独自发展，相对分散，如何实现多平台互联互通与协作共赢，成为超算互联网的重要课题。

3. 超算互联网

2017 年 7 月，国家超算济南中心在中国超级计算中心 CEO 联席会议上首次提出“超算互联网”的概念，希望将当时的 6 个国家超算中心联为一体。2019 年 7 月，科技部高新司在中国超级计算事业发展四十年纪念活动上，提出“建设国家超算网络，形成全国超算资源共享新格局”的设想。2023 年 4 月 17 日，科技部高新司在天津启动部署国家超算互联网，提出到 2025 年底形成技术先进、模式创新、服务优质、生态完善的总体布局。同时由 22 家单位发起组成了国家超算互联网联合体，由国家高性能计算机工程技术研究中心任理事长单位。

近年来，清华大学、中国科学院计算机网络信息中心、国家超级计算济南中心、鹏程实验室、中科曙光等多家单位分别从原型体系研究、高速互联网络、大规模数据传输、全国一体化算力调度、区域算

网生态建设、标准研制等方面，开展试验与探索。

1) 超算互联网原型体系研究方面

国家超算无锡中心牵头，联合中科院计算机网络信息中心、国家超级计算天津中心、国家超级计算深圳中心、清华大学等单位，围绕实现超算中心“从提供机时服务转向提供应用平台服务”这一目标，从构建高速互联的核心资源网、面向重点领域的应用平台、应用平台服务体系等三大方面进行了系统性研究，提出“超算中心间网络高速安全互连、跨中心资源管理与任务调度、应用平台基础框架构建与管理、应用程序的移植与优化、应用共性基础软件资源库集成、业务流程管理和应用平台构建”等一系列关键技术问题，同时对以应用商店为代表的的服务模式、机制、评价指标以及多方共赢商业模式等问题也进行了初步研究。

2) 超算互联网高速互联方面

国家超算济南中心在 2022 年的山东算网一期建设中，掌握了长距离无损无限带宽（IB）通信技术，在国内首次实现济南-青岛跨地域 500km 超算间长距 IB 网络的高速互联，将网络延迟从双向 10 毫秒以上降至约 5 毫秒，并在海洋、新材料领域进行了应用示范。江苏未来网络集团有限公司基于未来网络试验设施 CENI，在无锡、青岛、广州和北京之间建成 100G 确定性超算互联网，并通过整合中国网络操作系统 CNOS，提供“分钟级”和“按需定制”的网络切片服务，满足四地超算中心在不同业务场景下的互联需求。中信网络有限公司正在建设的 DCI（Data Center Interconnection）算力网络已经覆盖 20

多个省 100 多个数据中心，可以在 1 分钟内交付数据中心之间的弹性连接。

3) 大规模数据远程高效传输方面

中国移动和国家超算无锡中心联合开展了按实际使用量付费的“数据快递服务”试验，通过弹性云专线+云专网，帮助用户就近接入网络，并在短时间内租用网络大带宽完成数据传送。试验实现了在 3 个小时内完成 4TB 数据搬迁的目标，满足用户低成本、高效率的大规模数据传输需求。为支撑高校用户更便利地使用远程超算中心资源，中国教育和科研计算机网 CERNET 已开通无锡超算中心 100Gbps 接入 CERNET2，并在纯 IPv6 网络上采用两次翻译技术（IVI）实现清华大学与无锡超算 IPv4-IPv6-IPv4、IPv6-IPv6 的链接，为超算中心和用户搭建高速专用数据传输通道。中信网络在 2023 年 11 月开展了北京-广州 100G 链路的广域 RDMA 数据传输测试，结果表明长途链路上容损 LT-RDMA 设备数据传输速率达到 85Gbps，是普通 RDMA 速率 29Gbps、多流 TCP 速率 37Gbps 的 2 倍以上。

4) 全国一体化算力调度方面

中国科学院计算机网络信息中心持续优化建设具有基础设施形态、服务化特征的国家高性能计算环境 CNGrid。中科曙光在 2020 年正式上线全国一体化算力服务平台（ac.sugon.com），实现了对全国十多个算力中心资源的整合与统一任务调度，聚合双精计算能力超过 1000PFlops。中国科学院计算技术研究所、中科南京信息高铁研究院于 2020 年启动“信息高铁综合试验基础设施——算力网”项目，建

设了国内首个云-网-边-端一体化的算力基础设施试验场。鹏城实验室在 2022 年发布“中国算力网 (China Computing NET, C²NET)”计划, 利用新一代通信、网络技术实现国家大型算力高速互联。中国算力网一期已于 2022 年 6 月上线, 以“鹏城云脑” (cloudbrain.pcl.ac.cn) 为枢纽节点, 跨域纳管了 20 余个异构算力中心, 汇聚半精度算力规模超 3000PFlops。

5) 区域超算互联网生态建设方面

国家超算济南中心在济南-青岛间长距离 IB 网络高速互联试验的基础上, 建立了国内首个跨地域的超算互联试验网, 覆盖山东省 16 个地市, 并在 2023 年联合沿黄流域九省区的 11 所计算中心成立“黄河流域算网联盟”, 搭建黄河工业算力调度枢纽平台。合肥先进计算中心、上海超算中心、苏州超算中心、昆山超算中心等 2020 年联合组建了长三角先进计算联盟。国家超算广州中心自成立以来在珠三角地区设立了 15 家分中心, 连通了广州超算到各城市的高速网络专线, 建成超算应用服务网络, 并在 2018 年和 2023 年牵头成立了“粤港澳超算联盟”和“粤港澳超算应用互联网”, 旨在构建超级计算创新应用互联网平台, 为大湾区的科研机构、高校和企业提供计算能力支撑和应用服务。

6) 超算互联网相关标准研制方面

因超算互联网仍处于起步阶段, 相关标准研制主要以企业级、团体级为主, 国家级、行业级标准尚属空白。中国科学院计算机网络信息中心在运行 CNGrid 过程中, 对资源汇聚接入、用户统一访问入口、

异构集群互联互通、领域应用开发接口等方面进行了若干标准化工作；之江实验室联合清华大学、中科曙光等制定《超算互联网 参考架构》标准，并已立项研制资源信息服务接口规范、计算任务接口规范等 3 项标准；中国信息通信研究院联合北京航空航天大学、中国科学院计算机网络信息中心、国家超级计算济南中心等 21 家单位联合发布了《高性能计算云总体架构与技术要求》；中国移动云联合昆山超算中心构建超算并网平台一期试点，正在探索算力并网接口规范与标准等等。

四、 技术架构

超算互联网仍处于起步阶段，其技术路径需要不断探索与持续优化。本白皮书提出超算互联网的总体架构，并梳理其中的关键技术问题，为从业者提供参考与指导。

(一) 总体参考架构

超算互联网汇聚各算力中心的异构算力资源，对资源进行池化管理与统一调度，并协同软件、模型、数据等各类服务商，为各行业领域用户提供应用服务与解决方案，形成上下游协作共赢的超算商业生态。超算互联网采用层次化的设计原则，按资源层、平台层、服务层和应用层进行构建，同时应遵循统一的运营、安全与标准规范。超算互联网总体参考架构如下图所示：

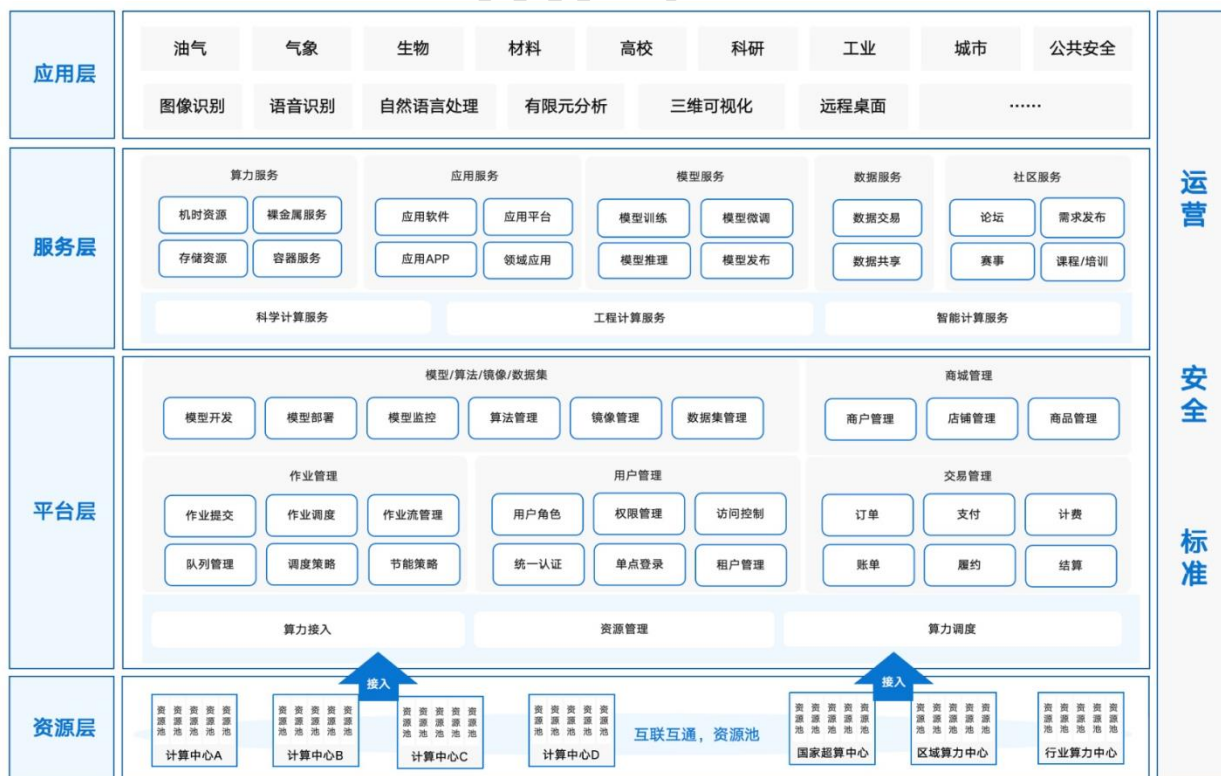


图 11 超算互联网总体参考架构

1. 资源层

资源层通过互联网络连接分布各地的算力中心,实现计算、存储、网络等资源的汇聚与共享。资源层为超算互联网提供算力资源底座支撑,汇聚不同运营主体、不同类型算力中心的异构资源,并通过对计算、存储、网络等资源进行池化,以实现统一管理与调度。

1) 基础资源

基础资源主要包括算力中心提供的计算、存储、网络等资源,呈现芯片多样、体系异构、存储多元、高速互连、绿色低碳等特征,直接或间接影响算力调度系统的设计。

- 芯片多样

随着应用需求和芯片技术的发展,计算芯片呈现多样化发展趋势。算力中心的主要计算芯片包括通用处理器(CPU)和AI处理器,提供FP64、FP32、TF32、FP16、BF16、INT8、INT4等不同精度的计算能力。其中,通用处理器有x86体系结构,如美国的Intel、AMD,中国的海光、兆芯,也有ARM体系结构,如飞腾、华为鲲鹏,还有其他体系结构,如龙芯、申威等。AI处理器有GPU、ASIC、FPGA、NPU等多种实现方式。目前Nvidia的GPU占据最大市场份额,国产芯片发展快速,如寒武纪、海光DCU、华为昇腾、百度昆仑芯、阿里平头哥、沐曦、天数智芯、燧原、摩尔线程、壁仞等等。

- 体系异构

异构计算已成为算力中心的主流架构。根据TOP500全球超级计算机排行榜,2011年6月首次公布的异构系统数量仅有19台,到

2023 年 11 月已增至 186 台。在前十位的系统中，除了日本“富岳”外，其他 9 台都采用了加速器技术。美国的 Frontier、Aurora，我国的天河、曙光等超算系统均采用 CPU+GPU/加速器的节点内异构架构，神威系列超算则采用片内异构众核处理器架构。

对于人工智能计算中心，异构系统更加普遍，除了广泛采用的 CPU+GPU 结构，如 CPU+Nvidia GPU、CPU+海光 DCU 等，还有 CPU+深度学习处理器/加速器的结构，如 CPU+Google TPU、CPU+寒武纪、CPU+华为昇腾等。众多计算芯片的出现以及异构计算的流行，大幅提升算力中心的计算能力，但同时也增加了并行编程的复杂性，提高了算力调度的难度。

- 高效存储

应用的多样性也催生了多种类型的高效存储系统，如传统高性能计算环境常用并行文件系统，而云计算、大数据场景通常使用大容量块存储、对象存储等。

算力中心的存储系统呈现高性能、分层管理、大规模扩展、管理智能化等特征，如：NVMe 等高性能存储技术提供更高读写速度和更低延迟，更好满足数据 IO 访问密集型应用；通过分层存储，满足冷、温、热数据的不同存储需求，兼顾效率与经济性；分布式存储架构结合 IB、RoCE 等高速网络技术，满足大规模扩展存储需求；存储系统具备智能化管理功能，更好满足数据迁移、存储布局优化、预测故障、数据保护等需要。

同时，在超算互联网、多云互联的环境下，数据经常要在不同算

力中心间迁移和传输，因此需要构建覆盖多中心的全局数据资源视图，实现数据全局可视、跨域数据访问与调度、跨中心数据管理等必要功能，常用技术路线包括全局跨域文件系统、全局元数据管理、通过集中共享存储同时向多中心提供存储服务。

- 高速互连

以 CPU+GPU 为代表的异构计算带来计算性能的快速提升，同时也对互连网络提出更高要求。在算力中心建设过程中，内部互连网络的设计是关键因素，不仅需要关注计算节点间的互连，还要考虑 CPU 间、GPU 间、CPU 与 GPU 间的互连等。在节点间互连方面，IB、RoCE 等 RDMA 网络已将带宽提高到 200Gb/s、400Gb/s，并正在向 800Gb/s 及更高速率演进。在节点内部的芯片间互连方面，CPU 间互连主要通过 QPI、GMI 等技术，CPU 与 GPU 间的连接通常采用 PCIe 技术，GPU 间的互连则以 Nvidia 率先推出的 NVLink 为代表。同时近年来 Intel 推出 CXL 规范，试图统一异构部件间的互连协议。

未来，各层级互连网络的性能将不断提升，同时新的技术也会出现，比如通过在交换机上执行 reduce 计算的在网计算(in-network computing)技术正得到越来越多的关注。另外，节点间的互连网络(如 IB)和 GPU 间的互连网络(如 NVLink)已有融合趋势，为算力中心带来更高性能的互连解决方案。

在算力中心之间，互连网络技术也从早期的帧中继、ATM 逐步演进到以全光高速互联、波分复用、SRv6、SDN 和确定性网络等新技术为代表的新发展方向，其趋势主要体现在更高的数据传输速度、更低

的信号传输延迟、更大的数据传输容量以及更高的可靠性水平。

- 绿色低碳

能耗已经成为制约超级计算机发展的重要瓶颈。根据 2023 年 11 月 TOP500 排行榜，共有 188 台系统披露功耗数据，累计达 388.9 兆瓦，平均每套系统超过 2 兆瓦，排名前两位的 Frontier、Aurora 功耗均超过 20 兆瓦。2022 年全国数据中心耗电量达到 2700 亿千瓦时，占全社会用电量约 3%，预计到 2030 年，将接近 4000 亿千瓦时。巨大的能耗不仅给算力中心运营带来高昂的成本开销，而且对计算机的可靠性和可用性也会造成直接和潜在的不良影响。

目前，业内在研制超级计算机、建设算力中心过程中，已大量使用新的绿色节能技术，以提高能源使用效率，包括处理器动态功耗调节、液冷散热、微模块、余热回收、智能功耗控制、绿电“源网荷储”等等。同时，“东数西算”工程也在推进算力中心向中西部绿色能源丰富地区布局。因此，在超算互联网实现跨域算力调度过程中，如何更高效调度和优先使用绿色算力，也是需要考虑的问题。

2) 资源池化

资源池化是通过队列、容器、虚拟化等技术，将 CPU、内存、磁盘、网络等各类硬件资源变成可动态管理的“资源池”，以简化资源的管理，实现资源的高效利用。资源池化后，可按需调用、弹性扩展，自动匹配满足业务高峰和低谷时的资源需求。其中，队列是作业调度系统组织作业的通道，可将计算资源按一定规则进行分类，调度系统按队列对资源进行统一调度和管理。虚拟化是从逻辑上实现资源实例，

用户无需关注底层物理资源的分配过程。计算资源通过队列、虚拟机、容器等技术实现池化，存储资源通过 NAS、SAN 等技术将多个存储设备聚合成资源池，网络资源通过 SDN（软件定义网络）实现池化。

资源池化应满足以下要求：

- 可隔离

可按租户对其使用的资源进行隔离，保障资源访问的独立性，且可按用户需求实现不同级别的隔离。计算资源可按队列、虚拟机、容器、裸金属等不同颗粒度隔离，存储资源可按目录、分区、介质、存储系统等不同颗粒度隔离，网络资源可按 VPC 等方式进行隔离。

- 可计量

可对用户使用的计算、存储、网络、软件、数据等资源进行计量，且可追溯。计算资源可按核时、卡时、实例数、节点数等单位计量，存储资源可按存储配额、实际实用量等方式计量，网络资源可按带宽配额、实际实用量等方式计量，软件可按功能模块、订阅时长等方式计量，数据可按条目数、订阅时长等方式计量。

- 可调度

可按不同级别实现资源调度，通过 Slurm、LSF、PBS 等作业调度系统实现 CPU 核心级、GPU 卡级的资源调度，通过 Kubernetes 等资源编排管理系统实现容器级的资源调度，通过 Openstack 等云管理系统实现虚拟机级别的资源调度。

3) 互连网络

中心互连网络是指使用高速网络连接多个算力中心，实现中心资

源的互通与共享。跨域数据传输、算力调度、应用迁移等场景都要依靠高速网络。中心互联网络建设应满足以下要求：

- 高带宽

在超算互联网场景下，如石油勘探、天文物理、生物医药、大模型训练等领域，许多超算或智算应用需要处理 TB 级甚至 PB 级的海量数据，同时在多算力中心之间进行算力调度时，需要进行大规模的跨域数据传输和应用迁移，因此对中心之间以及中心与用户之间的网络带宽要求极高。尤其是在碰到瞬时突发流量的情况下，这一需求更加凸显。

- 低时延

随着超算互联网的发展，将出现越来越多的低时延应用场景。例如，在算力跨域调度中，需要在不同算力中心之间进行协同计算；在工业领域的动态仿真设计中，用户与算力中心之间进行实时互动；在 VR/AR、大模型训推一体等应用中，云边端之间进行实时数据交互。这些实时计算交互类业务对网络的时延抖动、数据包丢失率要求较高。

- 高安全

超算互联网连接大量的用户与海量的资源，也带来了更多安全风险。因此需要围绕超算互联网的网络、应用、数据、传输、计算环境、平台系统、运营管理等方面，建立立体的安全体系，保障超算互联网环境安全。应当立足主动防御思维，通过冗余网络线路、安全事件发现、分析研判、通告预警、响应处置、追踪溯源、复盘整改等举措，将技术、管理、流程进行有机整合，提高超算互联网内生安全能力。

2. 平台层

平台层是超算互联网的大脑中枢，提供核心支撑功能。平台层承上启下，衔接资源、服务与应用，向下可屏蔽底层资源差异，实现算力高效接入，并进行资源的跨区域、跨主体、跨架构统一管理和调度；向上可提供作业管理、模型/数据/工具集管理、用户管理、交易管理和商城管理等系列支撑功能。平台层建议使用模块化、微服务的设计架构，以确保其可扩展性、可靠性和安全性。

平台层实现的主要功能包括：

1) 算力接入

算力接入是通过提供统一的标准接口实现各中心资源汇聚，应满足如下要求：

- 实现跨域异构的计算、存储、网络等资源接入；
- 提供多种资源接入方式，包括独占式、代理式、API 式接入等；
- 对异构资源进行抽象，屏蔽资源的个性化差异。

2) 资源管理

资源管理是实现对各中心接入资源的感知、监控、注册、度量等，应满足如下要求：

- 实现资源的感知，多维度采集资源运行状态信息，包括资源的规模、负载、利用率、可用性等，并进行全面的资源监控与健康度分析；
- 实现资源的注册，进行资源的登记、变更、注销等操作，形成动态、一致的资源列表；

- 实现资源的度量，对计算、存储、网络等资源进行统一的抽象描述和建模，形成度量模型。

3) 算力调度

算力调度是针对用户的计算需求，结合各算力中心的资源特征与状态，将计算任务分配到合适的资源池运行，应满足如下要求：

- 识别和分析用户的计算任务需求特征；
- 通过高效的调度策略，实现资源的快速定位与自动匹配；
- 对作业执行过程进行管理和监控，包括作业状态查看、作业日志记录以及作业的挂起、重启、删除、抢占等操作；
- 实现跨域资源的调度，建立跨算力中心的任务协同机制，针对跨域 workflow 进行统一定义；
- 针对同一任务的不同阶段，或同一阶段的多个不同任务进行跨域资源分配，并监督各中心任务执行过程，确保任务完成后及时返回结果；
- 对不同类型的算力任务实现统一融合调度，并提供统一接口。

4) 用户管理

用户管理是对用户、运营者、运维者、第三方商户、开发者等不同角色的管理，形成综合完善的用户管理体系，应满足如下要求：

- 对用户进行分级分类管理、管理权限；
- 实现面向个人用户、组用户、企业用户等的不同规模/类型的用户管理机制；
- 实现多个算力中心的用户账号统一认证，用户一次登录后可

使用多个中心的资源；

- 支持用户扩展，实现用户标签、行为分析等管理机制，更好地为用户服务。

5) 交易管理

交易管理是对应用商城中多种类型资源的交易、支付、结算等过程进行管理，应满足如下要求：

- 构建计费模型，从计量单位、资源包形式、付费周期、优惠活动等多个维度进行定义；
- 提供灵活的计费策略，对机时类、容量类、应用类、服务类等资源进行分类和聚合，提供固定价格、阶梯价格等计费方式；
- 提供线上线下等多种资源购买方式，支持用户在线支付；
- 支持平台与多中心之间、平台与商户之间、平台与个体开发者之间等的结算机制；
- 建立交易结算体系，实现跨地域、跨运营主体的结算通道。

6) 模型算法管理

模型算法管理是指平台层应建立相应的工具集、模型库、算法库、数据集以及镜像仓库等，支撑上层应用服务。模型算法及工具集建设应满足如下要求：

- 支持模型开发，提供相关的开发与测试环境，实现模型的快速开发、测试、上线与部署；
- 支持应用镜像管理，兼容主流镜像库标准格式，便于用户快速构建和发布镜像；

- 支持算法、模型、数据等内容的管理，包括收集、封装、上传、上架、备份等；
- 建立开放的数据集共享环境，支持用户创建、上传和分享。

7) 商城管理

商城管理包括应用的封装与发布、商户的入驻与管理等,应满足如下要求:

- 构建应用仓库，支持商户或开发者上传封装好的应用，支持应用在多中心之间共享和同步，支持应用的存储与自动分发；
- 支持应用封装与发布，兼容多种封装格式，形成丰富的应用模版，满足不同商户快速发布需求；
- 支持商户的入驻与管理，提供快速入驻与认证机制，提供店铺装修、用户引流、品牌宣传、用户和交易分析等运营辅助支持。

3. 服务层

服务层依托平台层的核心能力，整合算力中心、软件提供商、运营商、终端用户及各类服务商的能力，对用户提供服务，包括算力服务、应用服务、模型服务、数据服务、社区服务等。

1) 算力服务

算力服务是将基础资源封装成服务提供给用户使用，包括机时服务、存储服务、容器服务、裸金属服务等。算力服务应满足如下要求:

- 适用于科学计算、工程计算、人工智能等场景，能根据应用需求特征进行分配和调度相应的算力资源；

- 提供处理器、加速器、计算节点等不同形式的计算单元；
- 提供丰富的并行环境、算法库、编译器、深度学习框架等计算环境；
- 按统一规范实现异构算力资源的封装，提供标准化服务输出；
- 支持跨域资源的使用与调度。

2) 应用服务

应用服务是针对不同行业领域的应用软件，进行标准化封装并发布，方便用户一键安装、订阅即享。应用服务应满足如下要求：

- 按统一规范进行应用的标准化封装，封装步骤包括明确环境依赖、定义执行过程、执行测试用例等；
- 支持 Tar、Conda、Docker 等多种应用封装方式；
- 针对不同的硬件环境，进行应用功能与安全性验证，确保其安全性和可用性，同时支持自动推荐或匹配算力资源；
- 面向异地多中心实现应用仓库对应用的统一集中管理；
- 根据应用使用频度和运营策略，支持应用的预装、动态安装等多种策略；
- 支持按需定向发布或公开发布，供用户订阅使用；
- 提供应用分类、快速检索等功能。

3) 模型服务

模型服务包括模型训练、模型推理、模型调优、模型发布等，应满足如下要求：

- 支持主流深度学习框架与常用的模型库，方便用户订阅使用；

- 支持按统一规范进行模型封装，包含定义模型命名、构建模型文件、测试模型实例等步骤；
- 支持用户对其现有模型进行快速调优，或定制开发新的模型，并对外发布模型服务，供第三方调用；
- 模型部署发布时，平台提供容器运行环境、外部访问地址和相应的安全策略；
- 模型使用时可自动推荐或匹配算力资源。

4) 数据服务

数据服务提供数据管理、数据交易、数据共享等服务，以提高数据资源利用率，挖掘数据价值。数据服务应满足如下要求：

- 支持不同行业领域的数据存储管理，满足文本、图片、语音、视频等不同数据格式的存储要求；
- 提供数据封装的标准规范，如统一的命名、大小、格式等，支持按规范对数据进行比较、清洗、转换和加工，方便相关软件和模型高效调用数据；
- 支持按需定向发布或公开发布，供用户购买使用；
- 提供数据安全与隐私保护机制，涉及数据的存储、交换、访问等环节，防止数据泄漏、丢失和被篡改，确保数据的安全性、可用性和完整性。

5) 社区服务

社区服务包括需求发布、项目协作、业务交流、培训、课程、专业赛事等服务，应满足如下要求：

- 支持用户在社区发布各类任务需求,如仿真设计、图像渲染、模型调优、技术培训等任务;
- 支持用户在线交流,探讨行业趋势、技术经验、学术前沿等各类话题;
- 支持用户在线分享案例,发布专业文章,提升行业影响力;
- 提供培训课程、实训环境、竞赛、开发测试环境等相应支持;
- 与行业其他主流社区紧密互动,如用户引流、用户互认证、内容共享、赛事合作等。

4. 应用层

应用层既是超算互联网服务的目标对象,也是超算技术与其他行业领域交叉融合发展的产物。应用领域的持续拓展,用户群体的不断扩大,正是超算互联网建设的核心要义。超算互联网应当面向各行业领域用户,支撑丰富的应用场景,例如:

在人工智能方面,支持基础大模型/行业大模型训练与推理、图像识别、语音识别、自然语言处理(NLP)、计算机视觉(CV)、生成式人工智能(AIGC),以及人工智能驱动的科学研究的(AI For Science),尤其是AI方法在药物研发、基因研究、生物育种、新材料研发等领域的应用。

在科学计算方面,支持大气科学之全球气候海洋模式、空气质量预报,生命科学之药物设计,基础科学之高能物理,计算化学之第一原理计算,计算材料学之材料结构预测,计算宇宙学之N体计算模拟,计算生物学之蛋白质结构预测、脑模拟等。

在工业仿真方面，支持从电子元器件到高端装备的设计、仿真、制造，涉及计算机辅助设计(CAD)、计算机辅助工程(CAE)（包括计算流体力学、空气动力学、结构分析、热分析、电磁分析等）、计算机辅助制造(CAM)等。

在数据密集型计算方面，支持智慧城市之视频数据处理，生命信息领域之基因数据处理、高通量虚拟药物筛选，网络信息安全领域之网络靶场、大数据隐私保护和入侵检测，石油勘领域之海量地震数据处理，气象领域之数值天气预报，遥感领域之卫星遥感数据处理等等。

本白皮书在第六章进一步分析阐述了超算互联网的创新使用模式与支持的若干应用场景。同时，鉴于目前我国自主高性能计算应用软件生态偏弱，需要重点围绕国产自主硬件，加强对软件和应用的高度重视与投入，持续研发和完善包括操作系统、编程语言、编译器、调试及性能分析工具、基础算法库、求解器、各种领域应用软件在内的全栈软件生态，并通过超算互联网等平台不断优化国产软件的用户体验，吸引更多用户使用，进而形成自主软硬件产业的良性循环。

(二) 关键技术问题

本白皮书在借鉴业内相关研究成果的基础上，梳理出以下若干超算互联网关键技术问题，供从业者参考。

1. 资源的统一度量与动态感知

超算互联网环境具有资源分散、多源异构、场景多样等特点。资源类型非常丰富，包括共享资源、独占资源、裸金属、存储、网络、应用、模型、代码、数据等。且不同资源的计量方式有差异，如计算

按核时或卡时计，网络按带宽/流量计，应用、模型等也有不同的计量方式。这对资源服务的高效性、计算的准确性带来了巨大挑战。

需要从计算、网络、存储、软件和数据等多维度构建评估模型，对多样化算力资源信息进行抽象整合与规范描述，关联映射不同来源的异构计算资源，形成一套通用的资源度量与评估体系，实现多方算力资源合理评估、分配和高效调用。

在资源跨域异构调度场景下，需要构建动态更新的资源感知系统，实时获取各中心各类资源的状态，并根据其状态匹配或调整调度策略。建立动态资源感知系统，根据计算性能、存储容量、能效、成本、网络时延、传输带宽和应用特点进行统一度量与评估，实现算力需求和算力资源的统一匹配和高效调度。

2. 跨域异构算力的透明调度

面向跨域异构多中心的算力调度，是超算互联网所面临的重要挑战。用户通过超算互联网平台提交计算任务后，无需关心该任务在哪个算力中心上运行。这一过程的实现其实非常复杂，需要考虑多种因素，解决动态资源感知、资源匹配、数据和环境准备、调度机制等一系列问题。其中，动态资源感知是实现跨域异构算力调度的基础；资源匹配需要同时考虑需求侧和供给侧的因素，包括开发环境、应用需求特征、资源数量、可用性、忙闲状态、资源定价等；数据和环境准备包括系统环境、用户环境、应用环境、输入数据等；调度机制是指建立全局跨中心调度机制，需要考虑资源、用户、权限、优先级、权重等诸多因素。

高效是跨域异构算力调度的关键，需要让用户对跨域调度无感，获得与本地计算一致的良好体验。另外还需要协同“东数西算”等工程，探索建立跨区域算力资源调度、网络传输、算电融合、运营服务、交易结算、收益分配等协同机制，制定推广算力调度、算力结算等相关标准规范。

3. 应用的快速开发与高效匹配

异构体系结构显著增加了编程的复杂性。业界厂商纷纷推出支持异构编程的语言和接口，如 Nvidia 的 CUDA、海光的 DTK、申威的 Athread、AMD 的 ROCm/HIP 等，并在此基础上开发了各种基础算法库、求解器、深度学习框架等。近年来也出现一些第三方的编程接口，如 OpenCL、OpenAcc、SYCL 等，但这些接口目前在不同硬件平台上的实现仍然有差异。

超算互联网环境下，异构编程难、移植工作量大等问题变得更加突出。面对各超算中心、智算中心中的不同芯片、不同体系结构、不同厂商的异构硬件系统，如何做到应用“一次编译，到处运行”，成为超算互联网面临的重要挑战。

为应对这一挑战，需要从统一编程语言、自动编译工具、迁移工具、作业调度、应用封装等多方面着手解决，包括：研制独立于硬件平台且可屏蔽硬件细节的编程模型/语言，使程序从源头上适应支撑多种架构；建立统一的数学库，解决目前各超算中心的数学库、函数库存在接口、版本、性能不统一的问题；研究开发异构程序转换迁移工具，实现并行程序在不同硬件平台间的透明转换和自动编译，减少

用户移植难度和工作量；研究并行程序在多样化算力中心的透明调度和运行机制；研究多种应用封装技术，解决应用开发环境与运行环境不一致的问题，实现应用的快速分发部署与资源匹配。

4. 高速互联的新型网络技术

超算互联网在各地算力中心之间实现高速互联，形成一个可以共享和统一调度的基础设施资源池，需要建设高速率、低时延、稳定的网络支撑环境，让用户能够随时、随地、按需地通过无所不在的网络接入无处不在的算力。针对实时计算交互类业务和海量数据传输类业务，网络支撑环境需要在先进性、可靠性、高通量和低成本之间取得平衡，以满足不断变化的业务需求。

因此，网络支撑环境应当满足未来一段时间内的发展需求，并具备持续演进的潜力，持续研究和部署应用新型网络技术，例如：通过全光高速互联技术，满足算力节点 100GE、400GE、800GE 的超宽端口连接需求；采用波分复用系统，支持广域范围的长距离大规模数据传输；基于最新的 SDN 协议与网络操作系统技术，助力网络智能化，为网络管理和控制提供高度的灵活性和效率；研究 SRv6/G-SRv6 协议技术，提供覆盖省网、骨干和数据中心的端到端按需调度能力；研究确定性网络技术，为高实时性业务提供稳定的时延、确定的带宽、高可靠性、精确的时频同步和安全隔离性保障，将网络的性能从“尽力而为”提升到“确保所需”⁹；利用 IPv6 海量地址、确定路由、质量保障、内生安全等优势，支持构建高速、高效、高安全的算力互联等。

⁹ 刘韵洁,范赫男.确定性网络:从“尽力而为”到“确保所需”.网络安全与数据治理,2023.2

同时，需要考虑经济性，网络综合运营成本不应超过当前行业水平，确保超算互联网用户能够以经济实惠的方式获得所需的网络服务。

5. 大规模数据的远距离高效传输

在科学计算、工程计算、人工智能、大数据等应用场景中，需要处理的数据量越来越大。在超算互联网环境中，无论是当用户向算力中心上传数据，还是在各算力中心之间进行跨域的计算任务调度，如何提高数据传输效率，减少数据传输时间，降低数据移动成本，成为一项重要挑战。以往的常用方法是通过人工或快递直接运送存放好数据的硬盘或盘阵。但随着各行业领域产生的数据量越来越大，而且当超算中心互联后，数据在多个算力中心间流转会更加频繁，传统的数据搬运方式无法应对。

因此，需要升级现有网络基础设施。根据国家“东数西算”工程相关规划，到2025年底，1ms时延城市算力网、5ms时延区域算力网、20ms时延跨国家枢纽节点算力网将在示范区域内初步实现，国家枢纽节点间网络传输费用大幅降低¹⁰。另外，也需要持续研究远程数据高效传输技术，例如：FTPS、SCP等数据传输协议优化、广域容损RDMA数据传输技术、数据压缩算法、多路径传输、数据缓存、数据传输感知与网络性能监测等。

6. 超算互联网环境的立体安全防护

与超算中心单体运营模式不同，超算互联网环境将面临一系列新的安全挑战：攻击者可能会利用大规模攻击和高速攻击的方式快速渗

¹⁰ 国家发展改革委等部门印发的《关于深入实施“东数西算”工程 加快构建全国一体化算力网的实施意见》

透和破坏网络安全，从而可能会影响到大量的设备和系统；使用者来自不同地区、不同组织，涉及多个层面的身份认证和访问控制问题；通常要处理敏感的科学计算和数据，一旦发生数据泄露，可能造成更严重的后果，包括知识产权的丧失和机密信息的泄露等；算力中心的异构性也增加了网络安全的复杂性和挑战，如各种不同类型的计算机和操作系统可能存在漏洞和安全隐患等等。

为应对这些挑战，需要采取一系列的网络安全措施，包括加密通信、访问控制、入侵检测和防御系统、漏洞修补和安全培训等。同时也需要加强用户的安全意识和教育，提高网络安全的整体水平。

(三) 标准规范体系

算力中心目前面临算力设施分布不均衡、接口不统一、兼容与互操作标准缺失等问题，需积极制定国家、行业层级标准，以保证超算互联网的开放性和普适性，满足不同角色的接入需求，促进互联互通的生态建设。

在超算互联网标准体系的顶层规划下，体系化、规范化推进超算互联网建设，对支撑超算中心互联互通、行业应用、服务模式及评价等工作，包括但不限于超算互联网参考架构、超算互联网平台技术要求、超算互联网平台接入规范、超算互联网应用服务平台技术规范、超算互联网服务评价指标、超算互联网资源管理技术规范、超算互联网资源调度技术规范等方面的内容。

需要在术语、参考架构、技术要求、接入规范、接口规范、资源管理调度、服务评价和安全保障等方面开展超算互联网标准体系建设，

从顶层规划超算互联网及平台的设计、开发、部署以及应用。

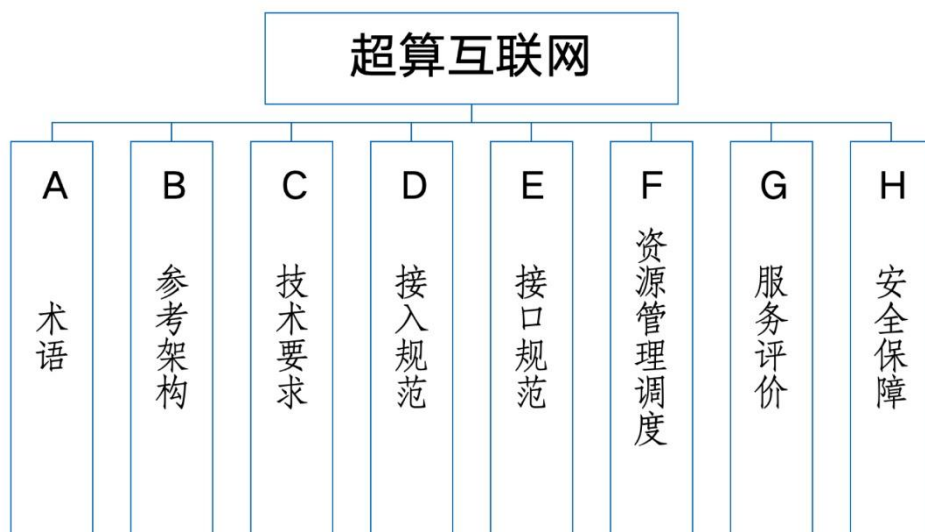


图 12 超算互联网标准体系

- 术语：规范超算互联网术语可以确保信息传递的准确性和一致性，避免歧义，是超算互联网标准体系的基础通用标准。
- 参考架构：规范超算互联网功能、角色和活动等，为超算互联网的设计开发、服务管理和实施应用提供参考。
- 技术要求：规范超算互联网平台总体技术要求、功能分层、业务服务、调度流程等，为超算互联网平台研发及维护提供参考。
- 接入规范：规范超算互联网各算力中心接入超算互联网平台的标准及流程。
- 接口规范：规范超算互联网不同组件、子系统或设备之间的交互方式，促进系统集成，提高互操作性，降低开发和运维成本。

- 资源管理调度：在资源管理和调度层面建立规范，统一资源调度的基本单位、格式、输入，并能实现网上用户和本地用户的统一管理。
- 服务评价：规范超算互联网相关平台的服务及其评价指标，促进算力一体化运营，指导相关服务平台的架构和机制设计。
- 安全保障：规范超算互联网需要遵循的安全保障标准，如防火墙、加密等，以保护计算资源和数据的安全性。

五、运营管理

超算互联网参与方众多，涉及产业链各个环节，涵盖算力中心、运营商、服务商、用户等多种角色。其中，算力中心是超算互联网的主要资源提供方，包括国家超算中心、区域算力中心、行业算力中心及其他算力中心；运营商是超算互联网运行服务的中枢，主要承担应用服务平台的建设与运行维护，提供资源接入、调度、交易与交付等服务；服务商由入驻平台的各类商家构成，包括软件服务商、技术服务商、培训服务商等；用户是超算互联网上各类资源与服务的使用方，包括企事业单位和个人用户。超算互联网联合体是由产业上下游机构联合组成的共同体，属于行业联盟性质的组织，旨在促进行业协作与发展。超算互联网主要参与方及其关系如下图：



图 13 超算互联网主要参与方及其关系

（一）运行机制

超算互联网将推动超算基础设施从以往“孤岛”式的单体运营方式向多方合作的网络化运营方式转变，从横向、纵向维度拉通产业链各个环节，需要灵活高效的运行机制做保障。本白皮书认为，超算互联网应重点围绕平台运营、成果转化、应用服务、生态合作等多方面积极探索新模式，构建新机制，以促进各方紧密协作，实现长期可持续发展。

1. 探索应用驱动多方共赢商业模式

采用一系列市场化机制，持续提升超算基础设施用户服务水平与行业竞争力，扩大超算用户群体，形成多方共赢的新型商业模式。

- 以用户为中心，以解决用户实际问题为出发点，促进超算中心从提供机时服务为主向提供应用服务与解决方案为主转变；
- 以应用为牵引，在超算中心、应用软件开发者和超算用户之间形成共赢的新型商业模式；
- 促进超算领域相关科研成果转化，着重提高软件成果的工程化、产品化、标准化水平；
- 加强用户需求调研与问题反馈改进，不断提升产品服务的可用性与易用性，提高用户体验与满意度；
- 拓展超算应用范围，从以提供科学计算为主的应用，向数字经济各行业领域的计算应用拓展，促进超算在产业升级、民生服务、社会治理等领域的广泛应用；
- 采用互联网运营推广模式，通过在线直播、电商引流、场景

APP、社交媒体、“线上到线下”、数据营销、精准推荐等多种渠道、多种方式推广超算应用服务，全面触达目标用户群体。

2. 探索超算电商平台应用服务机制

通过构建线上应用服务平台、应用商城、开发者社区等，服务相关各方，打造多方交流、交易与协作的平台。

- 在应用服务平台上提供算力、算法、数据、工具、模型等解决方案式服务，完善 XaaS 产品体系，降低超算使用门槛，缩短服务交付时间，提升可用性与易用性；
- 为服务商提供电商服务，解决服务商在资源上线、交易、交付、客户服务、售后过程中面临的系列问题，帮助服务商轻松入驻应用商城；
- 构建开发者开源社区，提供众包、众筹、线上竞赛、线上直播、线上培训等各类服务，促进技术交流、分享与协作。

3. 探索多维度多层次生态协作机制

在超算领域上下游、产学研用之间以及多算力中心、不同超算架构之间形成去中心化、平等协作的网状合作机制。

- 推动超算领域上下游企业之间协作，重点围绕国产芯片构建软件应用生态，促进软硬件之间的协同设计、适配与优化；
- 面向关键行业或重点领域构建领域应用服务平台，促进算力、数据、算法、模型、应用融合创新，面向用户业务场景形成解决方案；
- 推动产学研协作，围绕国家重大战略需求和市场应用需求开

展联合技术攻关，促进科研成果转移转化；

- 促进产教融合，支持企业、高校联合培养复合型超算人才；
- 引入众包、众筹合作机制，促进超算技术开发与应用创新；
- 支持不同架构算力中心之间或多个运营商之间合作，促进资源共享与技术协作；
- 支持不同超算架构之间协作，协同攻关跨架构互联互通的共性技术问题，统一接口标准与开发环境，降低底层硬件异构带来的编程复杂性，促进软件跨架构高效移植与适配。

(二) 参考运营架构

运营商是超算互联网运行服务的中枢，应积极探索互联网化的运营模式，构建 XaaS 化的产品体系，提供领域化的应用服务，实现数智化的管理支撑，形成标准化的运维体系。本白皮书提出一种参考运营架构，如图所示：

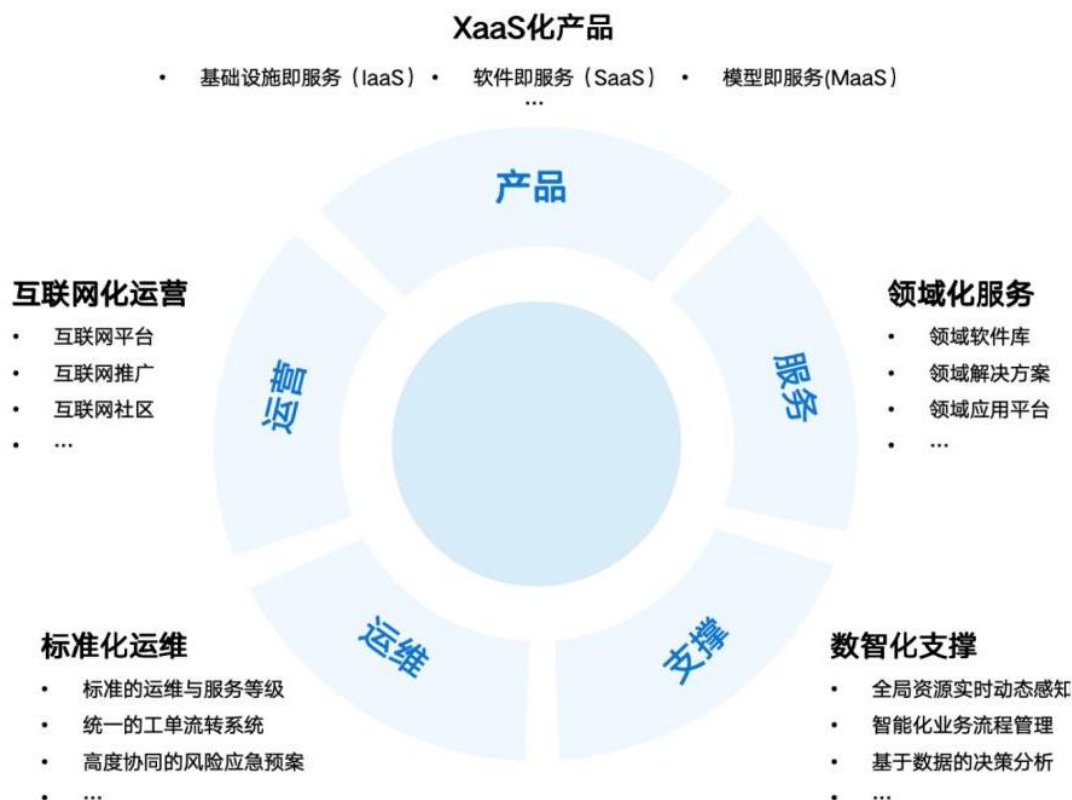


图 14 超算互联网运营商参考运营架构

1. 运营：互联网化

构建类电商应用服务平台，以互联网思维开展平台构建、产品开发、用户服务、运营推广等相关工作，包括但不限于：

- 实现算力、软件、数据、模型等资源上线、入网，提供云化服务，构建应用商城；
- 面向全业务场景、全媒体渠道开展超算应用推广与普及活动，持续扩大超算用户群体；
- 前端需求引领，快速推进平台、商城的技术迭代与功能升级；
- 基于技术社区、众包、众筹等互联网机制，提高用户参与感，促进超算产业生态的大规模交流与协作。

2. 产品：XaaS化

推进算力、软件、数据、模型等要素的在线服务化（XaaS），实现资源在线便捷交易，提供解决方案，相关工作包括但不限于：

- 拓展服务品类，提供 IaaS、PaaS、DaaS、MaaS、SaaS 等系列产品，包括基础资源服务、应用服务、数据服务、模型服务等；
- 支持超算中心通过统一接口，在平台上开设算力商店；
- 支持软件、数据、模型等开发者、服务商入驻应用商城，创建行业应用 APP 等；
- 为超算中心、软件服务商、技术服务商等提供在线开店、交易、交付、推广等全流程电商服务。

3. 服务：领域化

在工业、生物、材料、人工智能等行业应用领域，结合领域专业知识，集成行业共性软件，形成各类领域应用平台，面向最终用户的业务场景提供解决方案和支持服务，相关工作包括但不限于：

- 形成领域共性软件资源库，提供领域应用配套的软件工具、数据、模型等资源，为用户提供开箱即用的软件与应用支持服务；
- 配备应用领域技术专家团队，提供领域应用软件的开发、适配与优化支持服务，为用户提供面向业务应用的解决方案；
- 集成现有并行应用软件与应用服务平台，形成一批行业领域重大应用示范，解决卡脖子问题；

- 建立面向领域应用的服务体系和应用服务评价指标。

4. 支撑：数智化

采用数字化、智能化技术手段，构建运营管理支撑体系，优化业务流程管理、用户管理、资源管理、结算管理与运营分析决策，保障平台高效、稳定运行，相关工作包括但不限于：

- 基于用户在线活动数据，进行用户画像，为用户提供精准的个性化、定制化服务，提高用户体验；
- 对全局资源状态进行实时动态感知与精细化管理，提高资源监控管理、作业调度、应用适配等业务效率；
- 持续优化计费与结算模型，提高用户按用量计费以及中心之间结算的效率与准确性；
- 围绕咨询、试用、签约、交付、公益赠送等业务场景，进行数字化业务流程设计与评估更新；
- 通过抓取全量运营数据，进行多维度运营分析，根据分析结果进行迭代与决策，提升运营效率与服务质量。

5. 运维：标准化

围绕资源接入、资源管理、用户服务等方面，建立统一规范的运维服务标准，保障平台高效运转，为用户提供高品质服务，相关工作包括但不限于：

- 超算互联网各资源提供方应遵循统一的运维与服务等级标准，确保资源运维质量的一致性；
- 构建超算互联网全局资源管理与监控系统，为资源提供方和

使用方提供专业、统一、规范的运维保障服务；

- 按照高度协同的风险管理等级与应急预案，提前做好计划内维护、升级的通知与告警；
- 通过统一的工单流转系统和标准的服务流程，及时响应、处理用户反馈的系统相关问题，提升用户体验与满意度。

(三) 超算互联网联合体

超算互联网联合体是在国家相关部委的指导下，本着自愿、互利、共赢的原则，由国家超算、区域超算等算力中心，以及从事超算相关领域的企业、高校和科研机构等单位组建而成，于2023年4月17日在天津成立，首批发起成员单位有22家。截止2024年1月，超算互联网联合体成员单位数量已增至82家。



图 15 国家超算互联网联合体正式成立仪式

超算互联网联合体以强化国内高性能计算产业创新发展能力为目标，以推进各级超算中心互联互通，深化高性能计算应用服务创新为重点，以完善高性能计算服务环境和提升安全保障能力为支撑，助

力打造数据、计算、应用与安全协同发展的产业生态体系。联合体主要任务包括：协助国家相关部门制定超算互联网顶层设计和发展规划；设计超算互联网运营机制；制定超算互联网运营服务的技术标准、服务规范；评测和认证超算互联网运营实体；推广和宣传超算互联网理念、标准和政策，广泛联合对超算互联网的有益资源等。

超算互联网联合体

六、应用展望

超算互联网的应用价值，更多体现在超算使用门槛的大幅降低，应用软件生态的自主自强，以及对众多应用场景的泛在支撑。

（一）创新的使用模式

超算互联网将催生一系列新机制、新模式、新业态，帮助用户降低使用门槛，推动算力中心竞合发展，促进产业生态紧密协作。

1. 更多元的算力服务模式

通常情况下，用户以直接访问的方式使用超算中心的算力资源，这需要具备一定的专业技能。超算互联网为用户带来多样、便捷、场景化的算力使用模式。用户既可使用直接访问方式，也能通过应用服务平台使用 PaaS、SaaS、MaaS 等服务，或通过应用场景 APP 获取算力服务。新模式下，用户按需使用，按使用量付费，无需关注其使用的算力类型和地理位置，甚至无需关注使用的应用软件，从而大幅降低用户的算力使用门槛。

2. 更多样的资源部署模式

当前，中小用户一般倾向于使用公共算力资源，而大型用户要考虑业务连续性、供应稳定性、数据安全性等诸多因素，往往采用自建方式。超算互联网为用户带来更多样的资源部署模式，以满足不同类型用户或同一用户在不同发展阶段的需求，包括：同时使用公共算力和私有算力的混合云部署模式，使用多个算力中心资源的多云部署模式，支持中心训练与边缘推理一体部署的云边端协同模式等。用户可以根据自身业务需求选择合适的资源部署模式，并可动态调整。

3. 更良性的算力中心竞合模式

现有模式下，单个算力中心独立为用户提供服务，往往受自身算力资源规模、资源类型、资源忙闲状态、应用与工具软件丰富度、专业服务能力和地理位置等因素限制，难以很好地满足不同类型用户的多样化、差异化需求。超算互联网模式下，通过一体化算力服务平台，允许用户选择最合适的算力中心或者同时享受多个算力中心的服务。算力中心之间也将从竞争走向竞合，从而弥补自身资源与能力不足，更好满足用户差异化需求。同时，也将催生一批具有领域应用特色优势的专业服务平台，实现错位发展，避免同质化竞争。

4. 更紧密的产业生态协作模式

超算互联网不仅是数据传输网和算力调度网，也是算力产业的生态协作网，其协作范围涉及软硬件之间、上下游之间、不同要素之间、算力中心之间和产学研之间。在超算互联网上，活跃的不仅是算力的提供商和消费者，更有大量的应用软件商、应用服务商、数据提供商、代算商、技术与信息提供商、居间服务商等多元角色。超算互联网平台通过在线一站式配齐算力、软件、数据、模型、技术服务、技术资料、培训等各类要素资源，促进供需高效对接，从而形成算力产业的生态大协作平台。

(二) 丰富的应用场景

超算互联网应具备丰富的应用场景支撑能力，促进超算在各行业领域的广泛应用，尤其是要重点支撑高性能计算与人工智能融合创新。

1. 人工智能

1) AI 大模型

人工智能正在从专用智能走向通用生成式智能。大模型通常拥有百亿以上参数，具有能力“涌现”特征，如 Stable Diffusion、DALL-E 2 和 Midjourney 根据用户的文本提示生成高质量的图像，Sora 视频生成模型成为“世界模拟器”。但大模型依赖大算力，如 GPT-3 训练所需计算资源超过 3640 PetaFlop/s-day。超级计算机可在短时间内处理海量数据，提高模型训练效率，也可以优化算法，提高模型的精度和泛化应用能力。超算互联网服务平台不仅能够为垂直行业大模型训练提供算力资源，也可以为最终用户提供大模型即服务(MaaS)。

2) AI For Science

机器学习、深度学习等人工智能技术解决高维数学问题的强大能力，催生了一种新的科研范式，即 AI For Science，或称科学智能。目前人工智能方法在蛋白质结构预测、芯片设计、天气预报、大规模分子模拟甚至核聚变点火等方面取得了突破性进展。2023 年 3 月，科技部启动“人工智能驱动的科学研究的科学研究”专项部署工作，支持发展一批针对典型科研领域的“人工智能驱动的科学研究的科学研究”专用平台。面对 AI For Science 应用，算力服务平台需要整合超算与智算资源，支持混合精度运算，并实现计算过程任务流和数据流分解，同时通过统一不同平台、不同数据库的接口，实现自动计算 workflows，简化用户的操作流程。

3) 自动驾驶

自动驾驶的实现，需要基于感知融合、虚拟路测、高精地图、车路协同等关键技术，完成数字世界与实体路况的深度融合，这些场景的实现需要大量的数据和计算能力。例如，Tesla 的 L2 级别 FSD 自动驾驶融合感知模型，其训练数据集含有百万量级道路采集视频片段，算力投入约 500PetaFlops/day。自动驾驶等级每增加一级，所需要的算力就会呈现数量级的增长，比如多模态虚拟路测将再带来若干个数量级的算力需求。

4) 智慧城市

新型智慧城市需要融合 5G、物联网、人工智能、区块链等技术，实现对人、车、事件、场景的全面感知、实时汇聚和智能分析，因此对于算力需求日益旺盛。以城市安防场景为例，当前市场主流加速卡支持约 30 路高清视频解析，一个大型城市的视频监控达到 30-100 万路，对应需要约 1-3 万张加速卡提供算力。随着城市信息模型(CIM)、数字孪生、AR/VR 等新技术运用，以及民生服务、城市治理、文化旅游相关 AI 场景的拓展，可以预见算力需求将持续增长。

5) 社会计算

随着全社会数据总量爆发式增长，海量数据分析带来的计算需求正在激增。社会认知大脑是通过实时感知社会大众的网络空间活动，然后对各类感知数据进行分析建模，为化解风险和行动决策提供有效信息支撑。社会认知大脑需要采集海量的互联网社交媒体数据，其数据量每天数以亿计，且以文本、图片、音视频等非结构化数据为主，需要通过 AI 算法模型完成实体抽取、文本分类、情感分析、机器翻

译、OCR 识别等自动化分析处理，需要大量的计算资源。

2. 科学计算

1) 数值天气预报

数值预报系统是现代化气象预报业务的基础，需要通过复杂的数据处理和计算对天气状态进行推演，是典型的计算密集型应用。提升数值预报精度和预报时效一直是气象行业的发展目标，数值预报模式持续向更高时空分辨率、更快计算效率、更准确预报结果以及多模式耦合、集合预报的方向发展，对计算能力的要求越来越高。业内预测，天气预报从当前的 10 公里精度，提升到公里、次公里级，数据规模和算力需求都将提升 100-1000 倍。

2) 计算流体力学

计算流体力学(CFD)通过数值求解各种流体动力学控制方程，达到研究各种流动现象和规律的目的。CFD 在航空航天、汽车、船舶等领域已逐渐成为与理论分析、实验并列的三大主要方法之一。美国国家航天局(NASA)预测，21 世纪，高效能计算机和 CFD 技术的进一步结合将给各类航空航天飞行器的气动设计带来一场革命¹¹。目前，包含非稳态 RANS 紊流模型的整体飞行器的流固耦合的计算规模已经达到 E 级，而使用大涡模拟 LES(Large EddySimulation)方法的噪声模拟需要的计算规模超越 Z 级规模，远远超过今天最快超级计算机的处理能力¹²。

¹¹ 徐传福,车永刚,李大力,等. 天河超级计算机上超大规模高精度计算流体力学并行计算研究进展[J]. 计算机工程与科学, 2020, 42(10 高性能专刊): 1815-1826.

¹² 龚春叶,包为民,汤国建,等. 航天领域高性能并行计算研究进展[J]. J4, 2014, 36(09): 1629-1636.

3) 粒子输运数值模拟

粒子输运理论是研究微观粒子在介质中迁移统计规律的数学理论,在核武器、核电、放射性治疗等领域已产生多款粒子输运数值模拟软件。据统计,结合复杂物理边界条件的粒子输运程序的运行时间占美国能源部高性能计算系统上实际物理模拟运行时间的 50-80%。对于确定性模拟,其计算涉及三维空间、二维角方向、一维时间、一维通量等 7 个维度,全尺度计算中每个时间步需要 10^{17} 到 10^{21} 个自由度,整个计算超出了 E 级计算规模。对于非确定性模拟,计算精度与抽样粒子数的平方根成正比,高精度模拟将超出 Z 级计算规模。

4) 密码分析

密码是信息社会的信任基础。对称密码暴力破解的计算复杂度依赖于密码长度 N ,与数字、大小写字母、特殊符号的排列组合数成正比,约为 96^N 。对当前的超级计算机而言,12 位密码已较难破解,而 16 位密码分析更是远超 E 级计算规模。RSA 是目前最常用的非对称密码体系,其安全性依赖于大整数分解的难度。单个 RSA-768(232 位十进制数)分解约消耗 1752 万核时,1024 位 RSA 整数分解的计算量预计将高千倍以上,而现在常用 RSA 加密是 2048 位,其大整数分解的计算量将超过 Y 级¹³。

5) 新药研发

新药研发面临成本高、周期长、失败风险大等问题。在美国,一

¹³ 李翔谁,穆雨桐,迟利华,等. 数域筛法研究综述. 计算机应用,2018,38(S1):104-107

种新药的上市往往需花费超过 10 亿美元并耗费 10—17 年的时间¹⁴。使用高性能计算机辅助药物设计，可以加速药物研发，降低风险。在超级计算机上，可以快速完成超大规模药物先导化合物筛选与评估，提高药物研发效率；应用分子动力学模拟可大规模搜寻药物潜在靶点，对药-靶结合体系进行评估，为新药发现提供重要依据；采用人工智能与高性能计算结合的方法，可加速筛选甚至从头生成新的苗头化合物。目前可估算的化合物空间已高达 10^{60} ，需要海量的算力支撑才能够完成相关任务。

6) 新材料开发

借助强大的计算能力和先进的数据挖掘算法，新材料研发已由“经验+试错”模式向计算驱动模式转变。通过采用不同时间尺度和空间尺度的数值模拟计算，可揭示材料在不同条件下的微观组织结构和特性，提升新材料的研发效率。例如，目前科研人员已经将合金微结构的相場模拟扩展到了超千万计算核心。近年来科研人员还将材料模拟与人工智能相结合，实现了具有第一性原理精度且更大原子体系和更长时间尺度的大规模分子动力学模拟，使得模拟速度提升 2-3 个数量级。

7) 宇宙学模拟

宇宙学模拟是理解星系形成、暗物质、暗能量等重大科学问题的重要手段，其涉及的物理过程多、动力学范围大、计算方法复杂、计算规模大，一直是反映国际高性能计算发展水平的典型代表。其中最

¹⁴ 金钟, 陆忠华, 李会元等. 高性能计算之源起——科学计算的应用现状及发展思考. 中国科学院院刊, 2019. 6

核心的 N 体问题模拟及其应用先后 9 次获得“戈登·贝尔”奖。基于国产计算系统的天文 N 体计算模拟软件将实施千亿量级及以上规模的高效率宇宙学模拟，为国际大型星系巡天、暗物质、暗能量大型探测计划，以及我国重大科学工程 500 米口径球面射电望远镜(FAST)和空间站巡天望远镜等大科学装置提供必要的数值模拟支撑¹⁵。

8) 生命科学研究

生命科学研究包括基因测序、蛋白质结构预测、脑模拟等，大部分计算任务属于数据密集型。以基因测序为例，需要处理 DNA 序列、基因组结构、变异位点等大量数据。随着测序技术高速发展和国家增加普惠健康领域投入，基因测序人群队列规模越来越大，且包含全面的健康、环境和社会学相关信息，这些数据具备大数据特征，需要更大规模的计算资源。超级计算机还能进行全基因组关联分析，发现以前未发现的影响人类复杂表型的遗传因子，深入理解其遗传结构，针对复杂疾病提供个性化预防与诊疗方案。

9) 类脑模拟

人类大脑大约由 10^{10} 个神经元和 10^{14} 个突触组成。为了在计算机中“重现”人类智能，类脑模拟已成为神经科学研究热点，并逐步从区块模拟向全脑模拟发展。大脑模拟首先面临的是复杂度问题，例如日本超级计算机 K 曾模拟过 1.51×10^9 个神经元和 1.68×10^{12} 个突触，但仅达到人脑规模的 10%¹⁶。完美模拟一个生物学上的真实大脑需

¹⁵ 金钟, 陆忠华, 李会元等. 高性能计算之源起——科学计算的应用现状及发展思考. 中国科学院院刊, 2019. 6

¹⁶ 冯圣中, 李根国, 栗学磊等. 新兴高性能计算行业应用及发展战略. 中国科学院院刊, 2019. 6

要近乎无限的参数。全脑模拟需要整合多尺度数据，进行超大规模、分布式、高复杂度仿真模拟，并从中挖掘出执行特定任务的神经环路，这需极大规模计算能力的支撑。

10) 基础科学研究

高性能计算在基础科学研究领域已经成为不可或缺的重要手段。以第一性原理电子结构计算为例，科研人员借助计算软件，通过数值方法计算材料的电子结构，预测材料的性质及其应用。随着第一性原理向更精确、更大体系发展，需要实现更长时间尺度的分子动力学模拟，计算结果更加接近真实，相应地对计算能力的要求也越来越高。

3. 工程计算

1) 工业仿真

建模与仿真是智能制造的关键技术，在汽车、航空航天、船舶等领域发挥着无可替代的优势和作用。例如，在汽车的研发设计过程中，需要在超级计算机上进行多领域物理系统的建模仿真，涉及刚度、强度、噪声、振动与声振粗糙度（NVH）、机构运动、碰撞模拟、板件冲压、疲劳和空气动力学分析等。据业内测算，未来整车仿真规模将从3000万网格细化到3亿网格，计算需求增加约十倍。超算互联网不仅可以提供所需高性能算力资源，还能通过汇聚各类仿真计算软件，满足结构、流体、电磁、气动等多学科协同仿真设计需求，支持复杂系统工程仿真。

2) 石油勘探

随着三维地震勘探等新技术的应用推广，物探行业正迈入高精度

勘探时代，勘探原始数据量在 10 年内增加了 50 倍，对计算、存储资源提出更高需求。以目前物探领域的常规偏前处理、Kirchhoff 偏移和炮域波动方程偏移算法为例，每字节数据所需浮点算力达 10^4 次到 10^8 次级别。超算互联网通过实现跨区域、跨超算中心的算力调度和全局存储，将有效提升地震数据处理效率，缩短地震数据处理周期。

3) 数字孪生

数字孪生作为一种通用目的技术 (GPT)，能够为跨层级、跨尺度的现实世界和虚拟世界建立起沟通桥梁，在工业化、城市化、全球化进程中有着广泛的应用场景，涉及增材制造工艺仿真、研发设计、智能车间/工厂、智能家居、智能建筑、数字孪生城市、智慧交通、智慧医疗、智能能源、供应链网络、智慧地球等。数字孪生(Digital Twin) 是将物理对象表达为计算机和网络能识别的数字模型，并根据当前状态，通过实测、仿真和数据分析来实时感知、诊断、预测物理实体对象的未来状态，而且数字对象之间，数字对象与其物理对象之间能够实时动态互动，可以通过优化和指令来调控物理实体对象的行为。在这一过程中，需要利用仿真计算、数值模拟等技术进行全周期、全领域的动态仿真，利用大数据、机器学习等技术基于不完整信息和不明确机理推测未来，利用云计算、区块链等技术实现多个数字孪生体之间共享智慧，共同进化。随着数字孪生技术的进化，如大数据、人工智能、机器学习、增强现实等新技术进入研发设计阶段后，研发设计将真正实现“所想即所得”¹⁷。

¹⁷ 安世亚太科技股份有限公司. 数字孪生体技术白皮书, 2019. 12

4) 元宇宙

元宇宙需要在对现实世界进行多模态数据采样的基础上，借助模拟仿真、人工智能技术进行虚拟场景的重建、生成和渲染，在协同创建、高精仿真、实时渲染、智能交互等环节都对计算资源提出巨大需求。据英特尔测算，要实现《雪崩》和《头号玩家》中描绘的持续运行且极具沉浸感的计算体验，并让数十亿用户实时访问，现有算力需要提升 1000 倍。同时，随着技术的发展，元宇宙将从最初的娱乐、游戏场景，逐步拓展到文旅、教育、能源、农业、金融、数字人等更多领域，对多元泛在的算力需求将呈现几何级数增长。

5) 多媒体渲染

多媒体渲染是从模型生成图像的过程，先使用严格定义的语言或数据结构对物体进行三维建模，描述几何、视点、纹理、照明等信息，然后添加位图纹理、程序纹理、照明光线、凸凹纹理映射与相对位置，进而得到人眼所见的完整图像。多媒体渲染已广泛应用于建筑、影视、游戏、动漫、广告、AR/VR 等领域。随着 4k、8k 显示屏的普及，以及各行业对高清内容和实时交互需求的增加，传统的 PC 工作站已无法满足愈加庞大的渲染计算量。超算互联网通过对全局计算资源进行合理分配和调度优化，对作业进行高效管理，支撑大规模渲染任务的快速完成，提高动漫影视等内容制作效率。

七、 发展倡议

(一) 加强理论研究与技术攻关

超算互联网是新形势下适应中国国情提出的一项系统性工程，涉及面广，环节多，问题错综复杂，在国际上没有完全可参照的现成经验。超算互联网的发展在体系结构、关键技术、管理体制、运行机制、商业模式、标准规范等诸多方面面临挑战，需要在政府统筹支持下，通过设立长期科技计划专项，持续推动产业界、学术界紧密协作，不断探索创新，努力前行。

(二) 推进关键领域应用示范

行业应用是超算互联网的服务目标，也是其发展的助推器。支持有能力、有条件的超算中心打造有特色、有优势的领域应用服务平台，提高用户服务能力，同时遴选若干国家重大战略应用，给予重点支持，发挥多超算中心协同优势，吸引软件、数据、模型等生态伙伴共同参与，解决国家亟待解决的重大问题，形成生态协作的标杆示范。

(三) 持续完善超算互联网标准规范

为实现算力按需调度、数据可信流通、应用开箱即用、模型可上网等目的，需要业界共同努力，持续构建和完善超算互联网标准规范体系。在资源接入与输出、资源测量感知与标识、应用封装、数据共享、性能调优与评测、服务质量评价等各方面，逐步形成一套行业广泛认可、行之有效的标准体系，包括统一接口、标准、规范、流程、协议、架构、指南、最佳实践等。

(四) 强化产教融合培养超算人才

超算互联网的发展可能需要几代人的努力，而高水平的人才队伍是根本保障。我国超算领域人才缺口极大，特别是在先进算法、并行编程、大模型、工业软件、交叉研究、行业超算应用等方面，制约着超算应用水平的提升。应当以产业应用需求为引领，面向国产芯片自主生态建设目标，优化产教融合模式，为各行业领域持续培养和输送高水平、高技能的并行计算人才。

(五) 促进数据开放共享与流通

超算互联网的发展离不开成熟、规范、活跃的数据要素市场。在国家“数据要素×”行动计划统一部署下，围绕科技创新、工业制造、应急管理、气象服务、城市治理、医疗健康、金融服务等领域，应按照“开放为常态、不开放为例外”的共享理念，推进各领域数据开放共享与流通，同时保障好数据安全，促进大数据分析、大模型预测、AI For Science、开放科学等新范式的普及应用。

(六) 构建超算互联网区域枢纽节点

超算互联网的建设需要与区域经济社会发展紧密结合。围绕重点区域的资源优势和发展需要，建设超算互联网区域枢纽节点，以线上线下相结合的方式，充分汇聚区域内算力、算法、数据、人才等相关资源要素，形成区域一体化算力服务网络和生态协作网络，支撑区域产业转型升级与智能社会治理。

(七) 与中国算力网等计划协调发展

在国家推进“东数西算”工程的背景下，算力资源的基础设施化、算力网络融合化发展受到广泛重视。目前，国内存在多个算力网络相

关计划，包括鹏城实验室的“中国算力网”计划，中国移动、中国电信、中国联通等运营商的算网融合计划，中国科学院计算所的“信息高铁”计划，国家信息中心的“城市算力网”计划以及中国信息通信研究院牵头建设的算力互联互通验证平台，等等。超算互联网需要与各大工程计划加强沟通协调，着重在高速网络数据传输、异构算力共享调度、“东数西算”应用创新、统一接口规范等方面有效协作，为加快构建全国一体化算力网提供保障支撑。

八、附录

(一) 附录 1: 缩略语

缩略语	英文全称	中文全称
3D	3-Dimension	三维
AI	Artificial Intelligence	人工智能
AIGC	Artificial Intelligence Generated Content	生成式人工智能
API	Application Programming Interface	应用程序编程接口
CAD	Computer Aided Design	计算机辅助设计
CAE	Computer Aided Engineering	计算机辅助工程
CLI	Command-Line Interface	命令行界面
CNGrid	China National Grid	中国国家网格
CPU	Central Processing Unit	中央处理器
ESnet	Energy Sciences Network	美国能源科学资源网
EuroHPCJU	European High Performance Computing Joint Undertaking	欧盟高性能计算共同体计划
GPU	Graphics Processing Unit	图形处理器
HPC	High Performance Computing	高性能计算
HPCI	High Performance Computing Infrastructure	日本的高性能计算基础设施
IaaS	Infrastructure as a Service	基础设施即服务
LSF	Load Sharing Facility	一种分布式集群管理系统软件
MaaS	Model as a Service	模型即服务
PaaS	Platform as a Service	平台即服务
QoS	Quality Of Service	服务质量
SaaS	Software as a Service	软件即服务
SC	Supercomputing	超级计算
SLURM	Simple Linux Utility for Resource Management	一种调度系统

(二) 附录 2: 参考文献

1. 李国杰. 发展高性能计算需要思考的几个战略性问题. 中国科学院院刊, 2019. 6
2. 钱德沛, 栾钟治, 刘轶. 从网格到“东数西算”: 构建国家算力基础设施[J]. 北京航空航天大学学报, 2022, 48(9): 14.
3. 钱德沛. 构建支撑科技创新的新一代计算基础设施[J]. 数据与计算发展前沿, 2020, 2(1): 1-17
4. 孙凝晖. 对信息技术新体系的思考. 中国科学院院刊, 2022, 37(1): 8-14.
5. 历军. 中国超算产业发展现状分析. 中国科学院院刊, 2019. 6
6. 高文, 邹鹏, 田永鸿, 曾炜. 智能超算的技术演进. 鹏城实验室通讯, 2020. 6
7. 袁珩, 张丽娟. 数字欧洲计划 2021-2022 年度工作要点. 科技参考, 2022, 6
8. 超级计算创新联盟. 高性能计算云 (HPC Cloud) 白皮书, 2021. 9
9. 刘韵洁, 范赫男. 确定性网络: 从“尽力而为”到“确保所需”. 网络安全与数据治理, 2023. 2
10. 徐传福, 车永刚, 李大力, 等. 天河超级计算机上超大规模高精度计算流体力学并行计算研究进展[J]. 计算机工程与科学, 2020, 42(10 高性能专刊): 1815-1826.
11. 龚春叶, 包为民, 汤国建, 等. 航天领域高性能并行计算研究进展[J]. J4, 2014, 36(09): 1629-1636.
12. 李翊谁, 穆雨桐, 迟利华, 等. 数域筛法研究综述. 计算机应用, 2018, 38(S1): 104 -107
13. 金钟, 陆忠华, 李会元等. 高性能计算之源起——科学计算的应用现状及发展思考. 中国科学院院刊, 2019. 6
14. 冯圣中, 李根国, 栗学磊等. 新兴高性能计算行业应用及发展战略. 中国科学院院刊, 2019. 6
15. 迟学斌等. 国家高性能计算环境发展报告. 科学出版社, 2018. 12
16. 中国信息通信研究院. 中国算力服务研究报告 (2023 年). 2023. 7
17. 中国信息通信研究院. 中国算力发展指数白皮书 (2023 年). 2023. 9
18. 安世亚太科技股份有限公司. 数字孪生体技术白皮书, 2019. 12

编写委员会

参编机构 (排名不分先后):

国家高性能计算机工程技术研究中心 中国信息通信研究院 国家超级计算天津中心
国家超级计算济南中心 国家超级计算广州中心 国家超级计算无锡中心 国家超级计算深圳中心
国家超级计算长沙中心 国家超级计算郑州中心 国家超级计算成都中心
中国科学院计算机网络信息中心 中国科学院计算技术研究所 江南计算技术研究所
国防科技大学 北京航空航天大学 上海交通大学 清华大学 青岛国实科技集团有限公司
山西太行实验室有限公司 南方科技大学 之江实验室 全国智能计算标准化工作组
中国电信集团有限公司 赛尔网络有限公司 中信网络有限公司 江苏未来网络集团有限公司
中移(苏州)软件技术有限公司 北京世纪互联宽带数据中心有限公司
合肥市大数据资产运营有限公司 无锡恒鼎超级计算中心有限公司
国超(西安)计算技术有限公司 上海超级计算中心 甘肃省计算中心 哈尔滨市超级计算中心
陕西空天超算中心有限公司 雄安(衡水)先进超级计算中心 贵安超级计算中心
航天科工集团智能科技研究院有限公司 大连市大数据运营有限公司 深圳屹良科技有限公司
郑州祺石信息技术有限公司 中关村超互联新基建产业创新联盟

感谢以下人员参与编写 (排名按姓氏拼音不分先后)

曹振南 柴华 陈功纪 陈新海 陈页 崔华 樊党锋 范志华 冯顺磊 高巍 高占普
龚春叶 郭荣祥 何铁宁 何晓斌 洪钊峰 黄聃 黄震春 金能智 李可 李冉 李屹 刘辉辉 刘靖宇
刘杰 陆腾 栾钟治 吕灼恒 马炬 马庆怀 彭本黔 瞿隆 宋平 宋尧 隋剑峰 王斌 王健
王建波 王英龙 魏旗 吴璨 许华俊 许佳立 姚海鹏 杨莉 袁玉峰 张国华 张宏巍
张嘉鹏 张瑞友 张武生 张战炳 赵峰 赵欢 郑亮 郑双双

www.scnet.cn



超算互联网微信公众号



超算互联网官方网站