

高性能计算发展研究

白香君

中国航空研究院

高性能计算是指利用多处理单元所形成的强大计算能力来解决用单个工作站无法完成的密集型计算任务。高性能计算的发展水平已经成为衡量一个国家综合实力和高科技发展水平的重要标志,美国、欧盟、日本、英国都高度重视高性能计算的发展,并在国家层面设有专门机构负责研究、制定高性能计算发展策略,我国也逐渐将高性能计算的发展提升到了国家战略层面。

高性能计算能够推进人类对诸如星系等太大、原子等太小、核聚变等太快、宇宙等太慢、破坏性试验等太危险或昂贵的问题的研究。高性能计算集群(HPC)拆解复杂问题的能力可以非常显著地缩短突破科技创新瓶颈、解决实际问题的时间,为快速实现科技创新带来机遇,为工业领域取得跨越式发展奠定基础。

本文从我国高性能计算发展现状分析入手,通过中美高性能计算发展多维度对标分析,高性能计算技术管理多方面难点剖析,提出中国高性能计算发展的四大总体策略,并设计出未来工业领域高性能计算发展的架构模型。

材料研发、工业设计、试验研究等领域发挥着重要的作用,对国防建设和国民经济发展具有不可替代的作用。

从2021年中国高性能计算性能(Linpack测试)TOP100排行榜前10名(见表1)来看,中国最强超算仅有20%由国家机构研制并安装在国家超算中心;60%由网络公司研制安装,占比过半。可见,我国高性能计算技术能力十分分散,缺少国家层面的统筹。

从TOP100高性能计算机应用发展趋势(如图1所示)来看,数据分析/机器学习等新兴领域的应用占比由2015年的27%上升到了2020年的55%,已远远超过科学/工程计算的占比,可见我国高性能计算正在由科学计算向应用计算领域发展。

从2021年高性能计算应用领域份额来看(如图2所示),算力服务占比达到46%,人工智能应用占比达到9%,金融、互联网、教育科研、能源/石油、电子商务、工业/制造、电信等应用领域都开始使用高性能计算机,高性能计算正在成为产业发展的重要基础设施。

一、我国高性能计算发展现状分析

当前,我国高性能计算在宇宙探索、气候模拟、

二、中美高性能计算发展对标分析

美国高性能计算发展水平一直处于国际前列,具

表1 2021年中国高性能计算性能TOP100排行榜前10名

排名	研制厂商/单位	型号	安装地点	安装年份	应用领域	CPU核数	Linpack值 (Tflops)	Linpack来源	Linpack峰值 (Tflops)	效率(%)
1	服务器供应商	网络公司主机系统, CPU+GPU异构众核处理器	网络公司	2021	算力服务	285000	125040.0	Q	240000.0	52.1
2	国家并行计算机工程技术研究中心	神威太湖之光, 40960*Sunway SW26010 260C 1.45GHz, 自主网络	国家超级计算无锡中心	2016	超算中心	10649600	93015.0	Q	125436.0	74.2
3	服务器供应商	网络公司主机系统, CPU+GPU异构众核处理器	网络公司	2021	算力服务	190000	87040.0	Q	160000.0	51.2
4	国防科大	天河二号升级系统 (Tianhe-2A), TH-IVB-MTX Cluster +35584*Intel Xeon E5-2692v2 12C 2.2GHz+35584*Matrix-2000, TH Express-2	国家超级计算广州中心	2017	超算中心	427008	61445.0	Q	100679.0	61.0
5	服务器供应商	网络公司主机系统, CPU+GPU异构众核处理器	网络公司	2021	算力服务	120000	55880.0	Q	110000.0	50.8
6	服务器供应商	超算中心主机系统, 992*SW26010Pro异构众核处理器 390C 控制核心2.1GHz 从核2.25GHz, Sunway Network	超算中心	2021	科学计算	386880	12569.0	Q	13913.0	90.3
7	北龙超云/Intel	北京超级云计算中心T6分区, 5360*Intel Xeon Platinum 9242 同构众核处理器 48C 2.300 GHz, EDR	北京超级云计算中心	2021	算力服务	257280	10837.0	C	18935.0	57.2
8	服务器供应商	网络公司主机系统, CPU处理器	网络公司	2021	算力服务	192640	9540.0	C	16644.0	57.3
9	服务器供应商	网络公司主机系统, CPU处理器	网络公司	2021	算力服务	179200	9120.0	C	15482.0	58.9
10	北龙超云/DELL	北京超级云计算中心A6分区, 6000*AMD EPYC 7452 32C 2.350GHz, EDR	北京超级云计算中心	2021	算力服务	192000	4044.0	C	7219.0	56.0

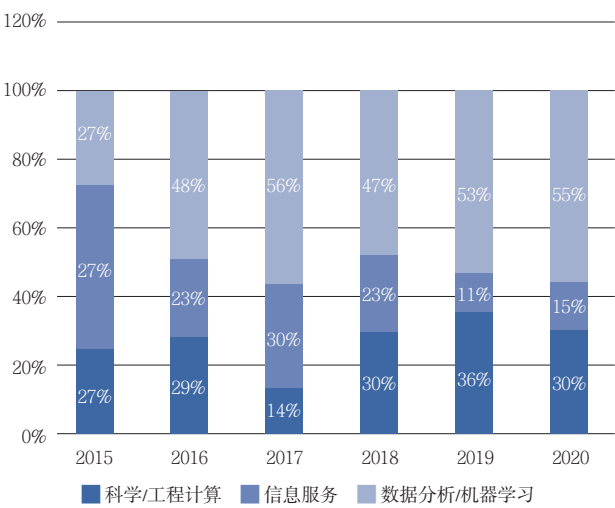


图1 TOP100高性能计算机应用发展趋势图

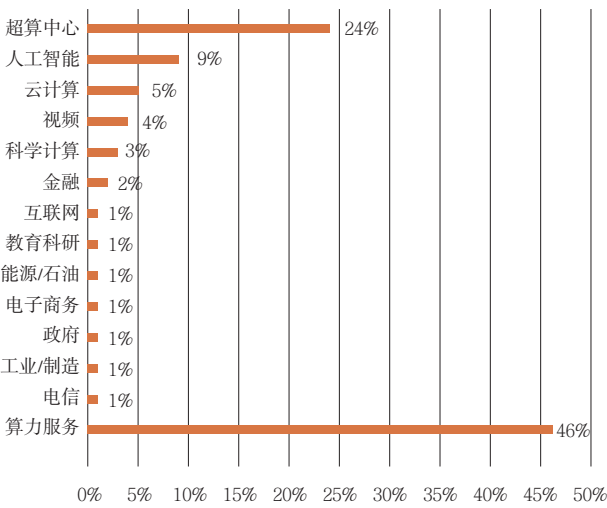


图2 2021年高性能计算应用领域份额

有示范引领作用,本节从战略规划、能力提升、软件研发、高性能计算应用等4个层面,对中美高性能计算发展现状进行对标分析,明确我国高性能计算发展存在的主要短板弱项,为后续提出我国高性能计算发展策略奠定基础。

在战略规划层面,美国从2015年就发布了“国家战略性计算计划”(NSCI),NSCI是比较全面的国家级顶层规划,在此计划下应运而生的HPC4EI计划更是涵盖了制造技术改进、新材料研发、移动系统发展等3个能源创新型子计划。我国通过国家高技术研究发展计划(“863”计划)和国家重点基础研究发展计划(“973”计划)支持过一批高性能计算技术研究和基础建设的项目,但至今没有类似“国家战略计算”的顶层规划。因缺少国家层面持续性的战略性计划支持,多数网络公司利用自身力量难以维持高性能计算领域长足的发展,缺少高层次项目统揽全局,各平行的高性能计算项目之间难以协同创新。

在能力提升层面,美国发展高性能计算主要是靠应用牵引,其E级计算机研制成功的标志不仅是Linpack测试,而是25个应用的几何平均值。我国E级计算机研制成功的标志仍然停留在Linpack测试阶段,但随着大数据、人工智能等新兴领域对于高性能计算体系结构的要求越来越高,单纯追求“容量型”高性能计算的计算速度而忽略了“能力型”高性能计算的计算效率,显然不是明智之举。

在软件研发层面,我国常用高性能计算软件主要依靠进口,自主研发软件使用极少,并且在超算经费投入中用于软件研发的费用还不足10%,距离实现软件自主可控差距较大。美国高性能计算常用软件主要依靠自主研发,其每年在软件研发方面投入的经费约为中国的6倍。

在高性能计算应用层面,我国使用高性能计算机较多的是网络公司,制造业普遍使用高性能计算较少,且规模较小。美国汽车、航空航天、电子通信等制造业普遍使用高性能计算机,每家企业都有多个高性能计算中心,美国公司的总体超算规模约为中国的10倍。只有大多数企业使用高性能计算集群,才能加速科技

创新,使国家走向高质量发展的道路。

三、高性能计算技术管理难点剖析

从工业领域来看,当前各单位高性能计算在技术和管理层面均存在一些难点,制约其持续发展。

在技术层面,高性能计算运行维护涉及专业技术广泛,各单位普遍缺乏HPC专业人员;高性能计算软件购买成本高,各单位Licence软件普遍不够用;高性能计算主要支撑复杂计算任务,内存需求量极大,存储容量不足现象频现;同一项目的不同单位人员分别在本单位高性能计算集群进行计算,协同设计十分困难;除此之外,存在计算网格量巨大,本单位高性能计算资源难以支撑;用户个性化定制成本过高,存储数据量利用率极低等问题,严重制约了本单位高性能计算的高效运行。在管理层面,当计算项目较多时,各单位存在阶段性应用需求旺盛、资源不足的现象;当计算项目较少时,又会出现阶段性资源空闲的情况;各单位高性能计算软件研发能力分散,自主创新能力不强,持续性维护成本高,科研投入支撑十分薄弱,持续性维护高性能计算机存在资金困难。

四、我国高性能计算发展总体策略

面对国内、国际高性能计算发展现状及趋势,为加速科技创新,加快构建高性能计算发展的良好生态环境,我国应当从以下4个方面发力。

一是制定高性能计算发展顶层规划,强化国家层面统筹,促进高层次项目间的协同创新,系统提升高性能计算实力。二是注重现有高性能计算资源的统筹,充分利用已有优势资源力量,构建现阶段高性能计算技术发展生态环境。三是注重高性能计算应用牵引,促进大数据、人工智能与高性能计算的深度融合,大力发展“能力型”高性能计算。四是加强高性能计算软件研发投入,提升自研软件应用占比,争取实现软件自主可控。

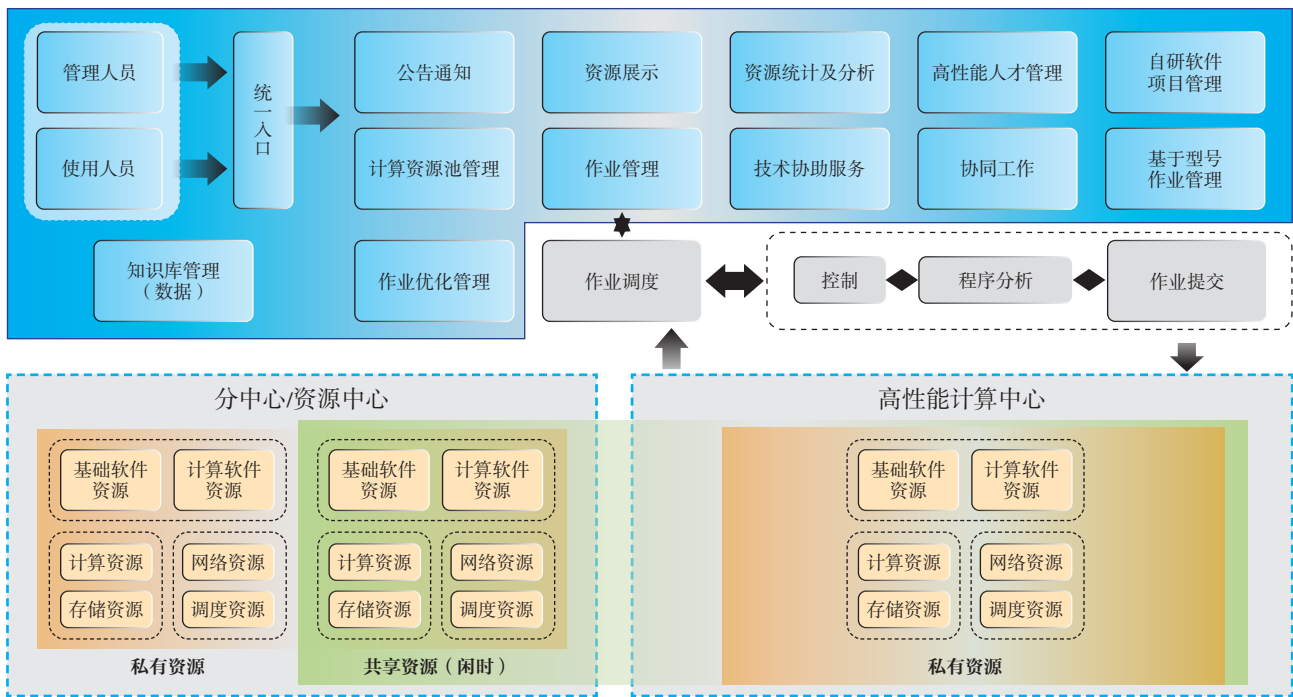


图3 高性能计算云平台架构模型

五、工业高性能计算未来架构模型

面对工业领域高性能计算发展现状及普遍存在的问题，高性能计算云平台将会是工业领域高性能计算未来主要的架构模型。高性能计算云平台能够从技术上解决各单位软硬件资源不足、存储容量不够，协同设计困难等问题；从管理上解决高性能计算维护成本高、自主创新能力不足等问题。

高性能计算云平台架构模型如图3所示。它具体指的是以行业内某单位其中一个高性能计算集群为中心，以各单位高性能计算分中心为节点，连接所有高性能计算中心，整合各单位闲时计算、存储资源，形成一个大的资源池，对计算资源、存储资源进行统筹管理，根据各单位实际使用需求，合理调配、占用其他单位闲时资源，大幅提高计算、存储资源使用效率；利用软件浮动 Licence，实现各单位软件资源的共享，大幅降低行业软件购买成本；联合各单位自研软件研发团队，在线协同设计，促进科技创新，提升自研软件占比；聘请高性能计算专业团队，对高性能计算云平台进行管理，不断优化管理水平，提升自主创新能力。

行业内高性能计算云平台将形成计算资源调度能力，应用软件共享能力，数据管理分析能力，自研软件创新能力，应用发展支撑能力等五大能力，支撑本领域高质量发展。

行业内高性能计算云平台建设第一阶段的主要任务是解决各单位高性能计算资源连接起来、统一调度起来的问题；第二阶段的难点是构建高性能计算的“高速公路”，即解决行业内不同单位高性能计算连接网络带宽、速率的问题。现阶段，部分工业领域可使用5G网络连接高性能计算，但如航空、航天、船舶等军工行业因保密原因，仅能使用行业内专网连接高性能计算资源，且网络带宽严格受限。行业内统一设计、部署实施高速网络是高性能计算云平台建设第二阶段的重点，也是行业走上高质量发展道路的关键一步。高性能计算云平台建设第三阶段应重点关注大数据应用技术的发展，行业高性能计算云平台的一个显著特点是解决的问题均是行业内复杂难点问题，数据存储量大；存储数据均为行业内计算数据，数据类型较统一；利用高性能计算云平台的大数据优势，发展本行业大数据应用技术，对于支撑行业重大决策部署，成为尖端科技发展的引领者具有重大意义。