

专题：国产环境下的大数据处理系统

Big Data Processing System in China's Homemade Computing Environment

客座编辑



陈刚(1961-),男,博士,中国科学院高能物理研究所研究员、博士生导师,高能物理研究所学术委员会主任,国家高能物理科学数据中心主任。长期从事高能物理科学计算及大数据研究工作。推动高能物理分布式计算国际合作,为国际高能物理大规模数据处理和计算做出了重要贡献。任国际高能物理科学计算协调委员会委员、国际高能物理数据长期保存工作委员会委员、中国科学院科学数据中心工作委员会主任委员等职。承担科技部重点研发计划项目,开展在国产超算上面向高能物理的高性能应用软件系统研制工作。现主持国家高能物理科学数据中心工作,推动高能物理科学数据的共享应用。

导读

作为我国科学技术研究以及经济活动的重要资源和技术,大数据正在成为科学发现和社会经济发展的新引擎。尽管我国大数据已经具备了良好的技术基础和独特优势,但仍然面临某些“卡脖子”的技术问题。因此需要在关键领域实现国产化,实现安全自主可控,同时实现国内资源和自主创新的融合。本专题以“国产环境下的大数据处理系统”为主题汇集了相关专家学者的6篇文章。这些文章分别介绍了自主研发的大数据与应用系统。在大数据技术应用方面,其中4篇文章分别介绍了高能物理、空间科学卫星、核反应堆等领域的相关研究与应用。

马福利等人撰写的《面向多星多任务的大数据处理系统设计》设计了高可靠硬件环境方案,提出了基于任务类型感知的统一资源调度系统,并建设了地面大数据处理系统。该系统用于支撑中国科学院的多颗科学卫星数据处理任务。

程耀东等人在《基于国产处理器架构的高能物理数据处理系统》中介绍了基于ARM等架构的国产处理器开展的高能物理数据处理软件移植等相关工作。文章还给出了在国产处理器架构上的典型应用评测结果。

官明等人撰写的《从格点量子色动力学应用看国产超算环境的基础软件》提出了一种适配于大规模高效异构计算和大数据

处理的新编程模型,同时阐述了在多个国产超算平台上开展的LQCD软件框架开发研究工作。

汪岸等人在《数值核反应堆大数据及其应用》中,对数值核反应堆中数据驱动建模和堆内微观现象预测的相关工作进行了讨论。作者总结了面向中国数值核反应堆进行建模优化和科学发现的研究工作,同时讨论了数值核反应堆大数据应用的未来研究方向。

霍建同等人在《广域虚拟数据空间中边缘缓存系统的研究与实现》中提出了利用广域虚拟数据空间系统中的缓存技术将数据缓存在靠近边缘客户端的位置的方案,以提升上层应用访问和共享数据的性能。作者利用测试结果对上述工作进行了验证。

张晨浩等人在《面向大数据处理应用的广域存算协同调度系统》中分析了如何依托我国的高性能计算虚拟数据空间统筹利用广域存储和计算资源的问题,设计并实现了一套广域存算协同调度系统用于提升大数据处理应用的运行性能。

国产环境下的大数据处理系统是一个庞大的工程,需要开展广泛的技术及应用研究。本专题篇幅有限,只涉及该领域的很小一部分工作。希望本专题可以作为一个敲门砖,吸引更多的专家读者关注和投入相关研究,推动大数据技术及国产硬件系统的共同发展。