

数据密集型网格示范站点建设与应用 取得良好效果*

文/中国科学院高能物理研究所 中国科学院计划财务局
(北京 100039) (北京 100864)

【关键词】中国科学院 知识创新工程重大项目 数据密集型网格

近年来,随着科学研究规模的日益扩大,科学家需要采集海量数据用于研究和解决复杂的科学问题,而科学数据呈数量级增长对计算技术提出了更大的挑战。数据密集型网格平台建立在网格技术基础上,是一种对海量数据进行高效分析处理的有效工具,在物理、化学、生物科学、医学、核能技术、石油勘探、航天工程等众多领域具有广泛的应用前景,并得到了快速推广。

2008年,中科院启动了知识创新工程重大项目“数据密集型网格平台示范站点”(英文名称:Data Intensive Grid Platform),项目以高能物理所为依托单位,联合计算技术所、网络信息中心及研究生院等单位共同承担。

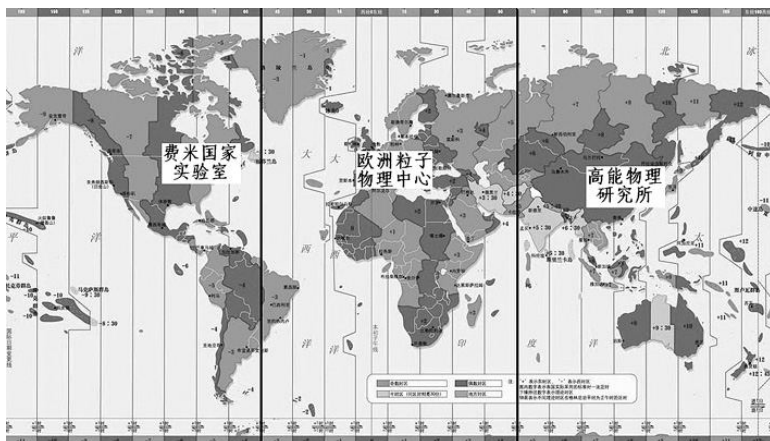
项目以学科实际应用的持续需求为牵引,采用成熟先进的网格中间件搭建和整合数据网格资源,建立了由1760个CPU核、617.3TB磁盘存储组成的计算资源网格环境,已成为国际高能物理



网格平台WLCG(Worldwide LHC Computing Grid)的重要组成部分,并实现了与中国国家网格(CNGrid)的互操作,有效支持了与欧洲粒子物理中心、法国国家核物理与粒子物理研究所计算中心、美国国家费米实验室等国际重要的科研机构以及国内多家研究机构之间的合作。该网格站点运行稳定,任务饱满,在国际高能物理网格系统的200多个站点中被评为Leadership站点。

项目建设的数据密集型网格平台已成功应用于高能物理、地球动力学及生物信息学等若干领域,并成为多领域科学计算的重要支撑系统。在粒子物理计算领域,超环面仪器实验(ATLAS)是运行在LHC(Large Hadron Collider)上最大规模的

* 本文由中科院计划财务局项目管理处茹加(rujia@cashq.ac.cn)、高能物理所科技处王晓辉共同组织撰写
收稿日期:2012年5月22日



的意义,为进一步人工设计和改造蛋白质奠定了一定的理论基础。

此外,项目根据网格平台的需求研究开发了多网格作业管理系统,实现了不同网格平台间的互操作。项目开发了基于异构介质的存储系统、分布式海量数据

实验之一,数据密集型网格平台为中国物理学家参与该大型国际合作项目提供了高性能计算环境。紧凑型缪子螺旋管探测器实验(CMS)则是类似于ATLAS的另一个大型LHC实验项目,数据密集型网格平台同样为中国科学家参与CMS物理分析提供了必要的支撑条件。数据密集型网格平台每年为ATLAS和CMS实验提供约500万CPU小时的机时,完成400余万个计算作业,在希格斯粒子寻找和性质研究、超对称粒子的寻找、超对称破缺共振态的寻找、W和Z玻色子特性研究以及top夸克的特性研究、 $\mu^+\mu^-$ 研究、J/ ψ 、 j 极化研究做出了重要成果。

存储系统、网格运行监控系统等并得到较好的应用。

基于该项目所取得的成果,项目组将进一步致力于推进数据密集型网格平台的实用化和产业化工作。数据密集型网格平台将继续以学科实际应用的持续需求为牵引,与信息技术的最新发展紧密结合,采用成熟的网格中间件建立和整合数据网格资源,形成e-Science示范平台,通过高效合理的运维管理保证网格平台的稳定运行,并为大科学工程、国际合作等科学计算需求提供优质服务。

大连化学物理所利用数据密集型网格平台进行了具有高度序列一致性但不同三维结构的蛋白折叠机理理论研究。该研究进行了大规模的、一系列不同温度下的分子动力学模拟,通过计算找到了蛋白质折叠的过渡态,发现了主导蛋白质折叠规律的重要相互作用,对于认知蛋白质折叠的机理、序列、结构和功能之间的关系具有非常重要

