

曙光 GPU 集群助力 天坛医院生命科学探索

案例简介

本案例中，天坛医院利用 NVIDIA GPU 作为主要加速器件构建高性能计算集群，用以进行生命科学领域的计算和研究工作。

Case Introduction

In this case, Tiantan Hospital used NVIDIA GPUs as the main acceleration devices to build high-performance computing clusters for computing and research in the field of life sciences.

背景

首都医科大学附属北京天坛医院基于前沿科研需求，及其用于生命科学领域前沿新兴学科的计算模拟需求，对计划构建的计算集群提出了如下的技术要求：

1. 满足天坛医院临床实验与临床研究中心（CTRC）临床研究的临床诊疗信息、组学数据（全基因组、外显子、RNA、蛋白质组数据）及高分辨影像数据的存储、处理和运算的需求（约 1 万人/年）；
2. 满足天坛医院神经系统单基因病诊断中心每年三千例单基因病患者致病突变和药物代谢基因 DNA 检测数据的处理和分析需求；
3. 满足国家神经系统疾病临床医学研究中心对大规模多模态医学影像数据进行数据分析的需求，建立基于医学数据的智能辅助诊断系统，满足相关的深度学习算法开发、平台搭建以及应用的需求；
4. 满足天坛医院未来开展全国性或地区性项目所需的数据收集、上传、共享、管理等需求。

挑战

根据用户科研需求，需要提供一套基于 NVIDIA GPU 的面向高通量基因组学数据处理与智能病理影像识别的高性能计算集群总体解决方案。用于高通量测序数据的生物信息分析、存储、共享，以及精准医疗相关研究中大数据的分析、解读。项目存在如下挑战：

- 1、运算集群需满足高通量基因组学数据处理与智能病理影像识别方面的需求，整合高性能计算系统、存储系统、基础软件与并行环境、集群管理与作业调度系统。构建完整的、面向精准医学的生物信息与智能影像研究平台，并提供有效的软硬件系统及应用层面（包括但不限于软件并行化、数据分析和存储方面）的技术支持服务。
- 2、运算集群应具有较强的扩展能力，可以方便地实现节点增加、系统扩充和升级。计算节点的扩展不受硬件品牌限制。
- 3、运算平台能够兼容研究课题的解决方案和相应的软件系统，例如高通量基因组学数据处理相关软件及流程开发、智能病理影像识别算法移植等，提供并行计算应用分析。
- 4、尽量减少数据中心的设备管理难度，简化架构，底层硬件对用户透明，实现应用简单易用的目的。

方案

中科曙光为天坛医院生物信息分析与基因检测系统项目提供了一套基于 NVIDIA GPU 的面向高通量基因组学数据处理与智能病理影像识别的高性能计算集群总体解决方案，为天坛医院建设了生物信息分析系统、基因组数据管理软件系统和医学影像人工智能平台。

针对用户需求，结合曙光在生命科学领域多年来的技术积累，以 NVIDIA GPU 为主要运算加速器，为用户提供全方位的高通量基因组学数据处理与智能病理影像识别的高性能计算集群。其中 GPU 计算系统包涵 9 台曙光 X740-G30 服务器，共搭载 24 块 NVIDIA Tesla P100 加速卡，以及一台搭载了 8 块 NVIDIA Tesla P100 加速卡 NVIDIA DGX-1 服务器。GPU 集群的总运算能力峰值可达到 598.4 TFLOPS。此外整体系统还包括面向序列对比

和序列拼接的计算平台、面向医疗影像处理的计算平台、面向测序数据及影像数据的海量存储及备份系统、高速网络系统及生物信息应用软件与工具、组学大数据分析流程系统和基因位点数据库系统与可视化平台，同时，针对医院安全等级的要求，曙光安全产品为运算集群提供全方位的安全保障。使用户在生物信息处理，医疗影像识别，智能测序，生物大数据处理等应用上取得了长足的性能提升。

经后续测试验证，在原 CPU 集群上需用几天时间才可处理完成的数据集，借由新构建的 GPU 集群助力，可将时间复杂度降低到以分钟为单位。

影响

通过曙光生物信息分析与基因检测系统的搭建，天坛医院国家神经系统疾病临床研究中心可以依托平台开展单基因病患者致病突变和药物代谢基因 DNA 检测数据的处理和分析、面向大规模多模态医学影像数据的深度学习算法研究及神经系统疾病基因库的建立。

曙光生物信息分析与基因检测系统的搭建可以提高科研质量、缩短科研周期、降低科研成本，助力天坛医院在精准医疗前沿技术成果上取得重大突破。