

经典案例

广东省畜禽现代种业产业园基因组测序数据存储分析平台集群建设

■ 项目背景

广东省畜禽现代种业产业园由新兴县政府为申报主体、温氏股份牵头建设，是全国首个获批建设的畜禽种业国家现代农业产业园。该产业园以建成具有国际竞争力的现代畜禽种业产业园为目标，一方面，产业园在加强畜禽遗传资源保护的基础上，不断提升畜禽种业自主创新能力，着力解决种业“卡脖子”问题；另一方面，产业园健全畜禽良种繁育体系，提高核心种源自给率和种业市场竞争力，推动我国畜禽种业振兴。该产业园还建立了国内领先的畜禽育种创新体系、辐射全国的高效良种繁育推广体系等。

■ 客户需求

- ◆ 该产业园拟建家禽全基因组选择及基因工程育种技术服务平台，其中基因组测序数据存储分析平台是重中之重，该平台需要依托于一整套高性能计算集群，用于研究院育种部门的生物信息计算领域相关研究，包括相关生物信息领域的pipeline的集成、平台性能优化、部署与集群利旧工作，从而实现用户业务的加速计算
- ◆ 本项目有部分利旧的、不同品牌的服务器计算主机、硬盘及旧存储设备，需要与新采购的计算节点实现集群统一管理与作业统筹调度；还需要满足现有FPGA设备接入集群平台及存储系统的需求，满足任务作业调度到FPGA节点及存储需求

经典案例

广东省畜禽现代种业产业园基因组测序数据存储分析平台集群建设

■ 项目内容

- ◆ 集群硬件采用双路服务器节点，计算网络采用IB网络互联，实现高带宽、低延迟，提高集群的加速比。采用生物信息领域广泛使用的分布式文件系统架构搭建分布式文件系统存储平台，确保用户数据读写的I/O带宽与数据容量
- ◆ 软件采用集群管理平台CHES，提供集群安全管控功能，支持B/S、C/S架构的图形化管理模式和异构作业调度，实现对生物信息作业脚本及作业流程的广泛支撑，统一监控当前整机和各个节点的状态，支持从不同维度统计CPU核小时、内存使用率、作业相关信息，提供整个集群系统及分布式文件系统的技术支持，随时跟踪基因组测序数据存储分析平台计算集群系统的运行情况，持续优化调度策略和集群系统使用效率

■ 项目亮点

- ◆ 本方案可行的架构设计及实施方案满足了多品牌硬件及现有FPGA设备接入集群平台及存储系统的需求
- ◆ 本项目生物研究产生的数据量庞大，对集群存储容量和计算能力有很严苛的要求，采用专为满足基因组分析研究需求而优化的存储解决方案，实现高吞吐量和高速周转

经典案例

广东省畜禽现代种业产业园基因组测序数据存储分析平台集群建设

■ 客户收益

- ◆ CHESS搭建的集群统一管理平台，帮助用户完成利旧机器、新购机器的整合及统一部署、管理，用一个平台、一套集群解决用户存储瓶颈与集群管理难题，提高用户的集群使用效率，为用户提供了高效的计算能力、海量数据处理能力
- ◆ 平台具有高易用性，集群部署系统、集群资源管理系统、集群作业调度系统，确保硬件计算平台与软件平台的高度契合
- ◆ 协助用户打通海量生物信息数据存储通路，提供新建集群平台的底层硬件、操作系统、应用软件进行调优服务，对目录功能及数据路径进行规划，实现集群性能、存储的最优、最大化利用