解决方案简要介绍

面向基因组分析的英特尔® 精选解决方案

intel SELECT SOLUTIONS

挑战

当前，相关机构在竞相寻找癌症和新冠等疾病的治疗方法。

构建和扩展用于基因组分析的计算集群面临许多挑战，这可能会给基因组研究人员的工作造成瓶颈。例如，专家预测在几年内，基因组学领域每年可能产生高达 40 EB 的数据。[1] 在这个瞬息万变的行业中，如何获得出色的可扩展性并处理如此多的数据，是一个重大的计算问题。

如今，研究人员需要：

- 能运行多种分析程序，而这些程序对计算系统的需求不尽相同
- 通过多种方式处理不断增加的数据集和复杂性
- 更快地获得结果
- 降低总体拥有成本（TCO）

使用案例：

基因组测序 | 群体扩展 | 癌症/罕见病研究 | 农作物研究 | 濒危物种保护 | 新冠疫情研究

基因组分析

概述 + 优势

为基因分析中的比对与变体识别工作创造了更大价值。英特尔利用英特尔架构优势对 Broad Institute 的 GATK 基因组分析工具套件进行调优，从而优化了上述工作的性能。

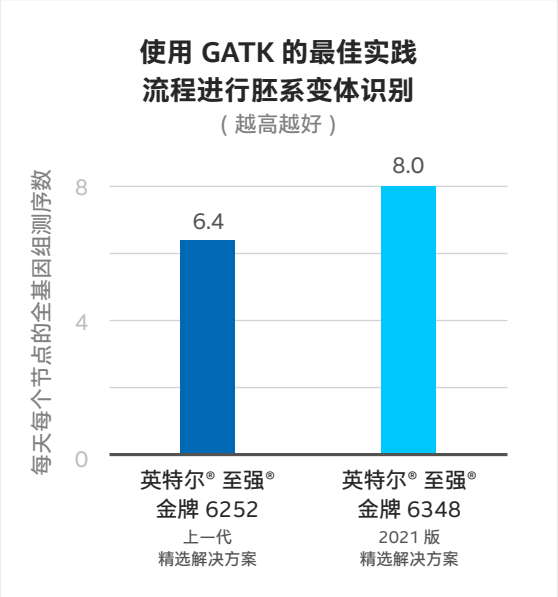
英特尔和 Broad Institute 携手合作，致力于实现三个目标：

1. 开发**优化的软件栈**，更快、更大规模地分析基因组。
2. 赋能和验证**交钥匙方式的计算基础设施的配置、构建和部署**，用于运行基因组分析。
3. **利用重要技术**分析不断增长的数据集，以实现科学和医疗突破。

面向基因组分析的英特尔® 精选解决方案用于从小规模计算集群（最少 4 个节点）扩展到超大型集群超级计算机。定制的系统可以快速、动态地配置，以满足企业随着工作负载增长而扩展的特定需求。此外，我们还提供用于基因组分析的端到端优化硬件和软件解决方案。

- 经过优化的 HPC 集群，可支持多种基因组分析流程以及其他 HPC 工作负载
- 使用英特尔® 技术：第三代英特尔® 至强® 处理器和英特尔® 固态盘（英特尔® SSD）
- 工作流程描述语言（WDL）脚本支持用户快速轻松地复用 GATK 的最佳实践流程来创建自己的分析流程

证据点



胚系变体识别流程可以受益于增加的内存通道数量和对 3200MHz DRAM 的支持，以及更高的内核数量（相比第二代英特尔® 至强® 可扩展处理器）中获益。

大容量英特尔® 固态盘数据中心家族产品（基于 NAND 或英特尔® 傲腾™ 固态盘）也有助于大幅提升解决方案的性能。

性能因用途、配置和其他因素而异。更多信息请见 www.Intel.com/PerformanceIndex

性能测试结果基于配置信息中显示的日期进行测试，且可能并未反映所有公开可用的安全更新。详情请参阅配置信息披露。没有任何产品或组件是绝对安全的。

您的成本和结果可能有所差异。

英特尔技术可能需要支持的硬件、特定软件或服务激活。

© 英特尔公司版权所有。英特尔、英特尔标识以及其他英特尔商标是英特尔公司在美国和/或其他国家的商标。其他的名称和品牌可能是其他所有者的资产。

为何选择面向基因组分析的英特尔精选解决方案

英特尔® 技术集成到英特尔® 至强® 可扩展处理器中，可进一步提高性能和可靠性：

- **英特尔® 高级矢量扩展指令集 512（英特尔® AVX-512）：** 提升要求最苛刻的计算工作负载的性能，相比前代英特尔® 处理器，可将每时钟周期每秒浮点运算数（FLOPS）提升高达两倍。
- **英特尔® Cluster Checker:** 检查与集群状态相关的 100 多种特性。英特尔® Cluster Checker 可在节点和集群级别对系统进行检查，以确保所有组件能够协同工作并获得最佳的性能。
- **集群管理软件堆栈:** 提供部署与管理 Linux HPC 集群所需的软件堆栈。该堆栈包括配置和开发工具、资源管理、I/O 客户端和科学库。Bright Cluster Manager、Warewulf 和 xCAT 等资源管理工具支持软件堆栈。
- **英特尔® Cluster Checker Runtimes:** 提供每个集群所需的关键软件运行时元素，以确保为应用提供最佳性能路径。英特尔® 运行时性能库，例如英特尔® 数学核心函数库和英特尔® MPI 库，可在英特尔® oneAPI 工具套件中获取。该工具套件可帮助提供针对基于英特尔® 架构的集群进行优化的卓越性能。它们还可以简化以数据为中心的工作负载在 CPU、GPU、FPGA 和其他加速器中的开发和部署。

想要了解更多信息？

什么是面向 HPC 的英特尔® 精选解决方案？

面向 HPC 的英特尔® 精选解决方案可帮助企业充满信心地选择正确的高性能计算产品、硬件配置和软件。

资源

面向 HPC 的英特尔® 精选解决方案

了解关于英特尔® 精选解决方案的更多信息基因组分析解决方案简介

面向基因组的英特尔® 精选解决方案

第三代英特尔® 至强® 可扩展处理器

Broad Institute of MIT and Harvard

基因组分析工具套件

[1] 国家生物技术信息中心，“大数据：天文还是基因组？” https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4494865/

第三代英特尔® 至强® 可扩展处理器配置：截止到 2021 年 8 月 8 日的英特尔测试。一个前端节点和四个计算节点，均使用英特尔® 服务器主板。

M50CYP2SB-003。前端节点配置：双路英特尔® 至强® 金牌 6348 处理器（28 核，2.90 GHz）；英特尔® 超线程技术启用，英特尔® 睿频加速技术启用，总内存 256 GB（16 个插槽/16 GB/3200 MHz）；BIOS 版本：22D08；BMC 2.66，SDR 0.31，CPLD 3p0；uCode：0x0b000280；CentOS Linux installation ISO（最小或完整）8 build 2011；存储 – 启动盘 1 个英特尔® 固态盘 P4610 1.6 TB（3D NAND PCIe 3.1 x4，3D1，TLC）；高性能网络：1 个英特尔® 以太网融合网络适配器 X550-T2（10 GbE），型号 X550T2。计算节点配置：双路英特尔® 至强® 金牌 6348 处理器（28 核，2.60 GHz）；英特尔® 超线程技术启用，英特尔® 睿频加速技术启用，总内存 512 GB（16 个插槽/32 GB/3200 MHz）；BIOS 版本：22D08；BMC 2.66，SDR 0.31，CPLD 3p0；uCode：0x0b000280；存储 – scratch drive：1 个英特尔固态盘 P4610 1.6 TB（3D NAND PCIe 3.1 x4，3D1，TLC）；高性能网络：1 个英特尔® 以太网融合网络适配器 X550-T2（10 GbE），型号 X550T2。

第二代英特尔® 至强® 可扩展处理器配置：截止到 2019 年 11 月 14 日的英特尔测试。1 个应用节点和 4 个计算节点。应用节点配置：2 个英特尔® 至强® 金牌 6252 处理器（24 核，2.10 GHz）；1 个英特尔® 服务器主板 S2600WFT；64 GB（4x 16 GB DDR4 2933 MHz）；1 个 960 GB 英特尔® 固态盘 D3-S4510 系列（2.5 in SATA 6 Gb/s，3D2，TLC）1 个 1.6 TB 英特尔® 固态盘 DC P4610 系列（2.5 in PCIe 3.1 x4，3D2，TLC）；微代码：0x500002c，BIOS：SE5C620.86B.02.01.0009.092820190230；CentOS Linux Installation ISO（最小或完整）7.7 build 1910；英特尔® Cluster Runtimes 2019.4；英特尔® Cluster Checker 2019.3.5；面向 EL7 2018.0 的 RPM 软件包的英特尔® 精选 HPC 解决方案；OpenHPC 1.3.8。4 个计算节点配置：双路英特尔® 至强® 金牌 6252 处理器（24 个内核，2.10 GHz；1 个英特尔服务器主板 S2600WFT；384 GB（12x 32 GB DDR4 2933Mhz）；1 个 960 GB 英特尔固态盘 D3-S4510 系列（2.5 in SATA 6 Gb/s，3D2，TLC）1 个 1.6 TB 英特尔固态盘 DC P4610 系列（2.5 in PCIe 3.1 x4，3D2，TLC）；网络设备：1 个英特尔® C620 系列芯片组以太网连接；英特尔® 以太网适配器 X722 板载 10 GbE；微代码：0x500002c，BIOS：SE5C620.86B.02.01.0009.092820190230；1 个分布式 10 GB Lustre 2.10 ZFS 系统，6 OST，3 OSS，LNet 路由器具有单个 10 GB 链路，用于将所有 I/O 流量客户端连接到 Lustre 服务器，以及 1 个 1.6 TB 英特尔固态盘 DC P4610 系列（2.5 in PCIe 3.1 x4，3D2，TLC）。