

# 案例研究

高性能计算  
英特尔® 至强® 金牌处理器  
英特尔® 傲腾™ 持久内存  
英特尔® 固态硬盘



## Arbutus 为研究人员开启云计算和机器学习 新资源

借助维多利亚大学的新一代云基础设施，科学家利用新兴的高性能计算 (HPC) 技术拓宽科学知识

“当研究人员要求获得某种‘环境’时，这种环境在我们看来也就是他们自己的‘虚拟实验室’，我们会设置相应的网络和硬件来支持他们的工作。无论是否有我们的服务支持，他们都可以在几分钟内就建立自己的虚拟实验室。”

——Belaid Moa 博士，维多利亚大学  
研究计算服务部高级研究计算专家

### 执行概要

维多利亚大学 (简称 Uvic) 位于加拿大温哥华岛，拥有 22,000 多名学生和数百名教职员工/研究人员。其研究计算服务部 (RCS) 通过国际协作，为全国各地的大学研究人员、机构科学家提供高级研究计算 (ARC) 基础设施和服务。该部门托管 Compute Canada 的一个 ARC 数据中心和 Arbutus 云。作为 OpenStack 项目下的云平台，Arbutus 云主要用于托管虚拟机和其他云工作负载。Arbutus 旨在加强传统的大集群 HPC 工作负载，并支持需要不同于传统 HPC 集群能力的研究项目，这些新兴能力包括在线机器学习/人工智能、大数据和协作计算等。Arbutus 基于联想 SR630、SR670 和 SD530 节点而构建，节点搭载了第二代英特尔® 至强® 金牌处理器，并采用了英特尔® 傲腾™ 持久内存和英特尔® 固态硬盘。

### 挑战

2015 年，维多利亚大学与 Compute Canada、WestGrid 和谢尔布鲁克大学合作，启用了 Arbutus 第一阶段，以支持新一代研究人员的研究工作。和过去依赖传统 HPC 集群来运行大规模并行计算任务或大规模仿真工作负载的研究人员不同，新一代的研究有着不同的需求。

“我们原有的 IT 服务无法提供能够满足研究人员高级计算需求的基础设施，”大学计算机系统部门研究计算服务部高级研究计算专家 Belaid Moa 博士说道：“过去，我们也有 HPC 集群，但研究人员迫切需要具备高可用性的协作平台、定制网站、根访问、微服务环境，以及其他云计算服务。这些服务正迅速变得与 HPC 集群同样重要。而且，对许多研究人员来说，这些也是必不可少的 ARC 服务。”

正因如此，Arbutus OpenStack 私有云基础设施应运而生。Arbutus 第一阶段在 250 个节点上搭载了英特尔® 至强® E5-2680 v4 处理器的 7,000 个 CPU 内核，采用节点存储，10 GbE 网络和 1.6 PB 的三重冗余 Ceph 存储 (共 4.8 PB)。Arbutus 使用虚拟化技术来提供“基础设施即服务 (IaaS)”资源，以支持研究人员所需的各种工作负载。

在此后的四年里，新的研究项目陆续启动，其中许多项目开始使用新兴的技术能力和研究环境，如机器学习 (ML)、人工智能 (AI)、JupyterHub 和大数据。这些新的项目，再加上对云服务日益增长的需求，导致平台需要更多的存储、更先进的计算以及更大的



**联想与英特尔**正在携手加速 HPC 和 AI 的融合，共同构建各种规模的解决方案，解锁全新层次的客户洞察。双方在系统和解决方案、软件优化和生态系统支持方面开展合作，旨在加速发现和解决 E 级时代及未来全球具有挑战性的问题。联想服务器是全球超级计算机 500 强的首选系统，由英特尔® 至强® 可扩展处理器和英特尔在存储、内存和软件方面的先进技术提供支持，为快速推动科学和产业进步奠定了创新基础。

内存池，这些共同造就了更大规模的云基础设施和 Arbutus 第二阶段。

## 解决方案

Arbutus 第二阶段于 2020 年初部署。新系统新增了 208 个联想 ThinkSystem SR630、SR670 和 SD530 节点，每节点配备 119 GB 的 ThinkSystem TruDDR4 内存和 1 TB 的英特尔® 傲腾™ 持久内存（英特尔® 傲腾™ PMem）。经过这一扩展，维多利亚大学在其云基础设施中新增了来自第二代英特尔® 至强® 金牌 6248 处理器和英特尔® 至强® 金牌 6130 处理器的 7,968 个内核。借助英特尔® 固态硬盘 S4610，Ceph 平台的固态硬盘存储也实现了 5.7 PB 的扩展。云节点包含两个全新的“数据库即服务”节点，可通过 SQL 提供专用的高性能结构化数据访问。

英特尔在内存、存储和处理器性能方面的创新支持维多利亚大学实现了全新功能。英特尔® 傲腾™ 持久内存可在每节点上支持具备类似 DRAM 性能的超大内存容量。英特尔® 傲腾™ 持久内存存在内存模式下可用作超大内存；在基于 App Direct 的存储模式下则可用作低时延的持久性存储，并实现类似 DRAM 的存储访问。通过采用内存模式下的英特尔® 傲腾™ 持久内存，Arbutus 第二阶段的各个节点既能利用第二代英特尔® 至强® 金牌 6248 处理器所带来的出色计算性能，同时又能具备额外的内存容量。

有了高性能和每节点更大的内存，维多利亚大学可在每台服务器上运行更多的虚拟机，以支持愈发庞大的研究人员群体，尤其是当他们需要 7 天 24 小时全天候持续运行相关工作负载来支持研究项目的时候。英特尔® 至强® 处理器架构内置英特尔® 深度学习加速技术（英特尔® DL Boost），并提供专门面向深度学习的软件，如面向英特尔® 架构优化的 TensorFlow 和英特尔® 分发版 Python。在代码为面向第二代英特尔® 至强® 可扩展处理器编译的版本时，这些技术有助于加速机器学习任务。

## 结果

借助 Arbutus 第二阶段，维多利亚大学的研究计算服务部能够以规模更小的 HPC 为全国更多的研究人员提供更强大的云计算资源。尽管该部门还在通过其大型 HPC 集群继续支持传统大型的超算工作负载，但研究人员也可在 Arbutus 上同时运行较小的并行作业，这样的启动速度要比等待大型设备的使用窗口更快。

Moa 博士解释道：“当研究人员要求获得某种‘环境’时，这种环境在我们看来也就是他们自己的‘虚拟实验室’，我们会设置相应的网络和硬件来支持他们的工作。无论是否有我们的服务支持，他们都可以在几分钟内就建立自己的虚拟实验室。”

据 Moa 博士介绍，Arbutus 让用户可以在 TensorFlow、PyTorch、Julia、Pandas、scikit-learn 和 Apache Spark 等不同的机器学习环境中进行选择。这些运行环境依赖 Conda 分发版。在使用 numpy、scipy 和 scikit-learn 等软件包时，Conda 分发版采用英特尔® 数学核心函数库（英特尔® MKL）进行低级运算。今后，维多利亚大学会安装面向英特尔® 架构优化的 TensorFlow 和英特尔® 分发版 Python。

Moa 博士总结说：“一些虚拟实验室甚至在运行小规模 HPC 工作负载，比如用于研究 SARS-CoV-2 病毒等问题的分子动力学软件 GROMACS。”

[Dennis K. Hore 教授](#)对 GROMACS 软件和利用云计算进行学术研究都非常熟悉。他是维多利亚大学化学和计算机科学系的研究员和教授，带领着一支 25 人的研究团队，负责 [15 个研究项目](#)。

“过去 15 年里，我的大部分研究项目都围绕着分子与表面的相互作用，”Hore 教授说：“比如，人体内其实使用了很多塑料，例如导管、支架、缝线、人工器官等。我的团队研究蛋白质如何与它们相互作用，希望能获得生物相容性的分子基础。”

但在过去三年里，Hore 教授的团队启动了一个全新项目。该项目结合了化学分析与基于 Arbutus 进行的大数据和机器学习，旨在帮助那些服用非处方街头药物的人们。

Hore 教授补充说：“我们在维多利亚的三地与他们开展了匿名合作，他们自愿为我们提供药品进行分析，随后我们会告诉他们这些药物的成分。我们使用先进的分析仪器对他们提供的样本进行了大量化学分析。随后，我们利用收集到的数据和化学库以及数据库来构建机器学习算法和应用。该项目旨在使人们能够获取药物的成分和强度信息，并基于此做出用药决定。”



温哥华岛药物检测项目与加拿大卫生部和维多利亚大学合作向公众提供药物检测服务。图片来源：Jay Wallace



在红外吸收光谱测量中，将少量样本置于晶体上，红外光通过晶体反射。  
图片来源: Jay Wallace

据 Hore 教授介绍，这些还只是冰山一角。该项目不仅能够造福人类，引入数据和计算机科学，还有利于推进远程医疗的潜在应用。这项研究带来的成果和应用可用于开发能够快速进行交互式化合物分析的便携设备和自助终端机。这种通过在线交互式机器学习实现的远程分析可以告诉人们其样本可能存在什么潜在影响，并为寻求相关分析的人提供指导。

Hore 教授解释说：“该项目源于 4 年前，当时一名投身减少药物伤害的药剂师联系我，要我对他配制的一种处方药进行质检。虽然多年来他都是通过同一家制造商购买的该特定药物，但是他的客户却告诉他，该药物对他们产生的影响与过去不同。这名药剂师希望可以分析该药物及其各成分的浓度。这也是这一科学研究领域的另一种潜在应用。”

这名药剂师的要求促成了如今的“药物检测项目”，该项目采用基于英特尔® 架构的云计算基础设施来支持先进的计算技术，从而帮助构建社会科学和化学领域的新知识。这项研究的成果可用于改善医疗、公共安全等各个领域。

## 解决方案总结

维多利亚大学的研究人员需要更多不同类型的云计算资源，这促使该大学的研究计算服务部对现有的 Arbutus 基础设施进行了扩展。Arbutus 第二阶段额外部署了近 8,000 个来自先进的英特尔® 至强® 可扩展处理器的全新内核，并采用英特尔® 傲腾™ 持久内存和英特尔® 固态硬盘 S4610。规模更大、更先进的 Arbutus 现已用于广泛的计算任务，包括 Web 服务、人工智能/机器学习和大数据等。

## 更多信息

了解有关[维多利亚大学研究计算服务部](#)的更多信息。

了解有关 [Arbutus](#) 的更多信息。

探索[第二代英特尔® 至强® 可扩展处理器](#)（内置用于加速 AI 推理的英特尔® 深度学习加速技术）的多重功能。

探索[联想 HPC 解决方案](#)。

## 解决方案组成部分

- 208 个联想 ThinkSystem SR630、SR670 和 SD530 节点，配备 ThinkSystem TruDDR4 内存
- 第二代英特尔® 至强® 金牌 6248 可扩展处理器和英特尔® 至强® 金牌 6130 处理器的近 8,000 个内核
- 每节点 1TB 英特尔® 傲腾™ 持久内存
- 英特尔® 固态硬盘 S4610



<sup>1</sup> Top500.org, 2020 年 6 月榜单。