

技术聚焦

智能模拟借助 AI 改善 HPC 结果

赞助方：IBM

Steve Conway、Bob Sorensen、Alex Norton 及 Earl Joseph

2019 年 4 月

执行概要

基于 HPC 的模拟是公认的第三个科学方法分支，它对国家和地区的安全、科学和工程创新，以及人们的生活质量做出了巨大的贡献。在模拟技术的驱动下，全球 HPC 市场价值从 1990 年的 20 亿美元增长到了 2018 年的 280 亿美元，Hyperion Research 预计到 2022 年，该市场价值将增长到 380 亿美元。

在这一预测结果中，增速最快的组件（14.9% 的年复合增长率）是高性能的数据分析（HPDA）技术，即利用 HPC 系统执行数据密集型模拟和分析。预计，人工智能的一个重要的 HPDA 子类别增速更快（26.3% 的年复合增长率），这个类别主要指将 HPC 系统应用于前沿 AI 的研发。Hyperion Research 预测，到 2022 年，HPDA（包括 AI）将为 HPC 服务器市场贡献 47 亿美元，为价值 380 亿美元的 HPC 市场（包括服务器、存储、软件和技术支持）贡献 100 亿美元。

AI 创造了科学方法的第四个分支：通过推理，对理论、试验和传统模拟方法加以补充。AI 学习模型（机器学习和深度学习，即，认知计算）需满足一个重要要求，即，能够推理，基于不完整的证据对现状或未来状况作出智慧的猜测。归根结底，AI 推理是与能够推理的计算机有关，即，反复猜测答案，比人类更快速地找到有用的解决方案。

在运用推理处理模拟任务时，得到的结果是**智能模拟**，即，使用机器学习、深度学习或其他 AI 方法，减少所需的模拟数量，同时提高解决速度和可靠性。借助智能模拟，尤其是通过预测哪些数据和探索性运行行之有效，从而省略其他数据和探索性运行，您有机会在给定的时限内审核更多数据，探索更多问题领域。智能模拟不仅可以应用于科学和工程工作负载，还能应用于由分析驱动的业务运营，比如欺诈检测、商务智能、关联营销、ERP 和销售规划。

本报告探讨了智能模拟的本质和用处，然后介绍了 IBM 的集成式软件解决方案，以满足棘手的 AI 工作负载要求。Summit 和 Sierra 系统是目前全球最强大的两个超级计算机。它们都构建于 IBM 技术之上，有望推动 AI 和 HPC 的进步。

注意：本页特意留白。

状况分析

迅速发展的 HPC 模拟和分析

每年的 HPC 系统销量（主要用于模拟）从 1990 年的不到 1000 套激增至 2018 年的 90980 套，这也使得全球 HPC 市场价值从 1990 年的 20 亿美元增长到了 2018 年的 280 亿美元，我们预计到 2022 年，该市场价值将增长到 380 亿美元。HPC 模拟首先渗透到了政府部门和学术研究机构，帮助他们解决自然科学领域的难题：物理学、化学、生物学、天文学/宇宙学和地质学等领域。很快，HPC 扩散到了私营企业的各个领域，既有全球大型企业，也有只有 25 名员工的中小型企业。就算是在学术领域，现在 HPC 模拟的使用也已经扩展到了多个学科，其中包括文化人类学和考古学、历史语言学以及社会科学。

与基于浮点的模拟不同，HPC 系统主要用于基于整数的数据密集型计算。这种使用趋势一开始出现在 20 世纪 60 年代的情报/国防行业，这也是超级计算机时代的萌芽期，随后在 20 世纪 80 年代这种趋势扩散到了金融服务行业的投资银行。目前，传统的 HPC 用户也开始利用 HPDA 和 AI 方法执行计算，大多数传统 HPC 用户都是利用专用 HPC 数据中心开展上游研发。这些方法也在鼓励越来越多的商业公司将 HPC 系统集成到企业数据中心，以改进业务运营，尤其是欺诈检测和网络安全、商务智能、关联营销、ERP 和销售规划。迫于竞争的压力，尤其是商业公司需要近乎实时地在数据结构中定位更复杂的问题，他们不得不这样做。

智能模拟

迄今为止，模拟依然是企业解决问题时最常用的 HPC 相关方法。但是，越来越多的企业选择结合利用模拟和 HPDA-AI 数据分析方法来解决。相比单独使用任一方法，结合使用这两种方法能帮助企业更快速、更全面地解决问题。这两种方法的结合被称为智能模拟。

- **分析助力模拟。** 气候研究一直是最棘手的 HPC 模拟问题之一，特别是随着总体模型的扩张，这个问题变得愈加棘手。但是过去十年，研究人员一直在推进基于分析的“气候知识发现算法”，以提供更多洞察力。2008 年围绕这一主题召开的首届 IEEE 研讨会名为“Data Mining for Climate Change and Impacts”。(<https://ieeexplore.ieee.org/document/4733959>).
- **模拟助力分析。** 另一方面，从根本上来说，自动驾驶系统的发展就是一个分析问题，但是专家认为，全自动驾驶汽车需要经过数十亿英里的测试，才能赢得消费者的信任。要在现实世界里完成如此多英里的测试根本不现实，而随着 HPC 模拟成为主流汽车制造商的支柱技术，它能够帮助企业完成这项任务。在这个案例中，虚拟汽车的外观和运行就像是实体汽车的“数字孪生体”。在 AI 模型的指导下，虚拟汽车将经过数百万次试驾，以体现各种各样的实际情况。这些模拟输出结果将持续优化基于分析的 AI 模型和汽车设计。

智能模拟能为 HPC 用户提供诸多竞争优势，包括以下收益：

- **提高准确性**，因为您可以在给定的时限内审核更多数据，探索更多问题领域

- **解决方案识别。**领域特定的认知模型和知识图形能够管理来自不同数据源的海量信息，让您能够利用庞大的专业知识库，快速聚焦最有潜力的调查方向，省略潜力不大的调查方向
- **获得更快速的解决方案，**因为您可以利用智能算法，快速瞄准最有价值的输入数据，从而获取洞察力，实现创新
- **提高成本效益，降低总体拥有成本，**因为您可以提高 HPC 系统的效率，减少未分析的数据
- **提高信任度，**因为您可以分析更多可能性，并提高机器学习/深度学习运营的透明度（以上面提到的自动驾驶系统为例）

通过采用智能模拟，HPC 用户将获得更多竞争优势，而未采用智能模拟的用户则会处于竞争劣势。

IBM、HPC 和 AI

长期以来，IBM 一直是 HPC 领导者。并且 IBM 有 2 台超级计算机在 500 大超级计算机排行榜中名列前两位 (www.top500.org)，借此，他们巩固了自己在 HPC 领域的领导地位。Summit 和 Sierra 这两台超级计算机将支持美国能源部 INCITE 计划下的各种突破性科学和工程研究项目以及其他任务。Summit 是全球最大的超级计算机，位于橡树岭国家实验室。它将利用 HPC 和基于 HPC 的 AI，解决多个极具挑战的问题，其中包括：

- 抗击癌症
- 预测聚变能
- 破译高能物理数据
- 识别下一代材料

在 HPC 环境中运行 AI

假如企业希望探索利用 AI 增强 HPC，提供更真实的结果，缩短建模和模拟解决方案的时间，那么基础架构是他们的一个重要考量因素。企业必须有一个途径来在 AI 训练流程的前端准备好大量数据，并对数据进行管理，这些数据可能来自各种不同的数据源，比如物联网 (IoT) 和传感器。就 AI 工作流本身而言，企业需要根据智能、快速的模型优化与训练，构建、优化和验证机器学习和深度学习的开源框架以及结果的准确性。这些任务相当复杂，而 AI 的成功依赖加速且优化的基础架构。计算能力和存储性能需要与软件工具联系在一起，从而以分布的方式在企业内部和混合云环境内支持 AI 工作流的构建和运行以及现有的 HPC 工作负载。

IBM Spectrum LSF 工作负载调度程序得到了广泛普及，迄今上市已经超过 25 年。IBM 表示，当前的 IBM Spectrum LSF 提供多个高级功能，支持企业构建灵活的基础架构，来运行机器学习和深度学习工作负载以及传统的 HPC 工作负载。具体包括：

- 支持容器化工作负载 (Docker、Shifter、Singularity)
- 为 NVIDIA GPU 工作负载提供高级支持
- 处理器、内核和 GPU 核关联性

- 用于运行 Apache Spark、Hadoop 工作负载的连接器
- 高级并行调度功能

如果企业在特定的 AI 框架上实现了标准化，那么他们可以通过 IBM Spectrum LSF 运行 AI/机器学习/深度学习训练工作负载。与此同时，IBM Spectrum LSF 会提供精细的 GPU 访问控制和并行调度功能，让企业受益匪浅。IBM 在下列视频中演示了如何通过 IBM Spectrum LSF 运行 Horovod 和 Distributed Tensorflow：

- 在 IBM Spectrum LSF 套件 v10.2 集群中运行 Tensorflow Deep Learning 案例
<https://www.youtube.com/watch?v=wxeiPBEItJ4>
- 在 IBM Spectrum LSF 套件 v10.2 集群中运行 Horovod Tensorflow Benchmarks
<https://www.youtube.com/watch?v=wxeiPBEItJ4>

通过在现有的 HPC 基础架构上引入 AI 工作负载，您可以消除计算孤岛，更有效地利用计算资源。此外，IBM Spectrum LSF 旨在支持您构建动态的混合云环境，从而根据需要在云端减少和传输计算资源，满足任意类型工作负载的需求。

IBM 表示，IBM PowerAI 结合利用热门的优化型深度学习框架、支持库和开发工具，帮助您更快地准备好 AI 环境，同时为深度学习和机器学习训练提供更高的性能。IBM 表示，目前已经有很多客户通过 IBM Spectrum LSF 运行 IBM PowerAI 工作负载与传统的 HPC 建模和模拟工作负载。

目前，在整个 AI 工作流中，AI 模型训练是计算密集程度最高的流程。因此，利用分布式方法训练 AI 模型变得越来越普遍，因为这有助于加速完成这些任务。但是，您需要考虑整个 AI 工作流中的多个关键环节。比如，数据准备工作占用了数据科学家 80% 的精力¹，并且在构建最优 AI 模型时，超参数优化也是一个相当耗时的必要流程。

为了兼顾数据科学家的需求，构建处理 AI 工作流的解决方案，IBM 提供了 IBM Watson Machine Learning Accelerator（前身为 IBM PowerAI Enterprise）。IBM 表示，它提供端到端的解决方案用于创建分布式深度学习环境，同时提供各种功能用于简化 AI 模型的训练、调优和部署。IBM 指出，企业还可以在现有的 IBM Spectrum LSF 集群上运行 IBM Watson Machine Learning Accelerator 实例。

IBM Watson Machine Learning Accelerator 的功能

- **弹性分布式训练**功能可以在训练时动态地为模型分配 GPU，进而加速获得结果。您无需停止训练，即可添加或移除 GPU。该功能可帮助您更快速地启动和完成作业，提高模型准确性。
- 附带超参数搜索和优化的加速模型开发功能可帮助您在运行训练时，利用基于建议的逻辑提高准确性。

¹ <https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/#926554a6f637>

- **资源共享**功能支持您智能调度工作负载，以最大化利用率，同时支持您无缝访问异构计算环境，包括企业内部或混合云环境内的 IBM Power Systems 和附带 GPU 的 x86 服务器。
- **存储资源连接器**旨在简化和维护与多个数据源的持续连接。
- **多租户**功能通过在通用的共享计算集群上支持多个数据科学家、框架和应用，减少浪费的时间、成本和管理开支，并支持您访问更大的集群。

IBM 支持的智能模拟真实案例

IBM 正与各行各业的客户合作，这些客户在运行传统 HPC 工作负载的基础架构上运行 AI 建模和进行推理。越来越多有 HPC 体系的企业利用 HPC 基础架构，支持从上游研发到日常业务运营的所有企业事务，借此消除孤岛。同步运行 AI 工作负载与现有 HPC 工作负载，并利用模拟和高级分析处理同一问题以提高洞察力和创新能力，这变得越来越普遍。IBM 软件解决方案致力于帮助企业基于传入工作负载的本质和优先级，分配资源。

下面我们提供几个智能模拟示例，以展示面向 AI 和 HPC 的 IBM 解决方案的用途。

电子设计自动化 (EDA)

一家半导体制造商在开展物理设计工作的基础架构上运行晶片检测模型。这样，他们不仅减少了总体成本，还提高了灵活性。因此，一旦晶片质量出现大问题，晶片检测工作负载的优先级会提高，从而占用大量计算资源。

汽车设计

- 很多汽车和零部件设计都在以前的设计模式基础上实现了进化。有数百万种潜在的变更可能可以改进汽车的空气动力学、结构完整性、环境友好性和乘客舒适度。我们不可能全面模拟每一种可能性，因此，我们可能与一些有价值的优化失之交臂。但是，我们已经针对现有设计建立了一个庞大的知识体系。我们能否利用这个知识体系训练模型，评估大量潜在变更，识别能够带来收益的变更？这种模型能够全面探索设计空间，规避死胡同，进而以传统方式模拟多种易驾驭且有潜力的可能性。许多汽车企业已经投资了该领域。比如，2018 年 9 月，福特 (Ford) 在密歇根州迪尔伯恩召开的 HPC User Forum 上谈到了这一点。他们坦言，在处理 2D 模型时，这种方法的准确性比单纯的传统模拟方法高 95%，速度快 300 倍。

计算化学

- 贝叶斯优化方法应用于多个领域。比如，IBM 和英国的 Hartree Centre 发布了一个基于 IBM Spectrum LSF 的化学药剂用例。化学品（比如洗发水或牙膏中的润滑剂）比例的微小变化可能会导致产品的物理属性发生巨大变化，产品可能越变越好，也可能越变越差。为了在湿实验室里测试 1 毫升的所有可能的组合，您可能需要做 500000 次实验。但是这方面的计算工作相当耗时。通过使用贝叶斯优化和深度学习，用户能够快速构建模型，迅速预测化学品的混合方式，从而大幅节省成本。

除了智能模拟外，您还能以其他方式在 HPC 环境中使用 AI，提高效率：

认知发现

- 数据奠定了现代研究的基础。这些现代方法（包括 AI 和深度学习）的数据准备工作是一项艰巨的任务。如今在启动训练之前，数据科学家投入了很多时间在数据准备工作上。数据可能来自无数个数据源，其中包括物联网传感器数据、业务数据和传统数据。这些数据必须转化成常用的分析格式。意识到这一点后，IBM 展示了一款名为 IBM RXN 的工具，它能够预测有机化学反应的结果。通过使用基于大量摄入的相关数据的知识图形的这种认知发现方法能充当您的专家顾问，帮助您预测结果。下面的视频详细介绍了面向化学的 IBM RXN。
<https://www.youtube.com/watch?v=7DVu9ZKPUws&feature=youtu.be>

不确定性量化

- 不确定性量化 (UQ) 是计算流体力学 (CFD) 领域常用的一种方法。它通过量化模型中不确定性参数的影响，提高模拟结果的可靠性。过去，UQ 依赖的是 Monte Carlo 分析，后者属于计算密集型工作，需要进行大量模拟，用不同的输入参数集训练模型。IBM 研究部门正在探索如何将 AI 方法应用于 UQ 作为 UQ “即服务” 架构的一部分，从而生成更好的 UQ 参数集，目标是缩短解决方案总体周期。有关该工作的更多信息，请参见研究报告《不确定性量化即服务》：
<https://dl.acm.org/citation.cfm?id=3291328>

调度优化

- 工作负载管理已经成为了 HPC 环境里的一个标准工具。工作负载经理已经习惯于在愈加复杂的异构计算环境里调度工作负载，服务背景不同的用户。通常，HPC 环境用户要求的资源要超过工作负载实际需要的资源。这是为了避免因资源不足导致出现应用故障。但是，这会导致高昂资源无法得到充分利用。以 EDA 为例。大多数应用都是单核的，导致很多作业都在同一个计算机上运行。因此，内存争用也是一个重要的考量因素。此外，对于一个特定的作业需要多少内存，用户往往了解不多。请求的内存不足会导致作业停摆；请求的内存过量又会大幅降低总体利用率和吞吐量。IBM 正在评估机器学习方法，利用 IBM Spectrum LSF Predictor 技术预览，预测工作负载的内存和运行时需求。
 - 此外，IBM 客户 MediaTek 公司构建了自己的深度学习/机器学习模型，用于预测作业内存。他们表示，准确率超过了 95%，利用率和吞吐量也得到了显著提升。
 - 埃森哲 (Accenture) 也投资了该领域：<https://insidehpc.com/2018/11/video-accenture-engineering-compute-takes-hpc-enterprise/>

未来展望

近日，Hyperion Research 对政府部门、学术机构和各行各业的知名 AI 专家展开了全球调研。调研结果证实了 HPC 领域和 AI 领域正在迅速融合。在前沿 AI 研发领域 HPC 不可或缺，它也可用于新的用途，包括用于精准医疗、自动驾驶系统、关联营销、商务智能、网络运营和物联网等等。与此同时，现有的 HPC 用户也在模拟中加入了 AI 方法和工具。Hyperion Research 预测，全球 HPC 市场仍将延续强劲的增长势头，随着 HPC 市场和大规模商用计算市场的日益融合，到 2022 年 HPC 市场价值将高达 380 亿美元，其中超过 100 亿美元来自 HPDA（14.9% 的年复合增长率）和 AI（26.3% 的年复合增长率）收入。

HPC 和 HPDA-AI 买家与用户告诉我们，他们想要进入一个“泳道”，这意味着，HPC 资源能够高效支持 HPC 和高级分析工作负载，尤其是那些能从模拟和分析运行中受益的与日俱增的问题。最终，用户希望 HPC 系统能够近乎实时地集成模拟和分析结果。而智能模拟能满足用户的这一需求。

IBM 面向 HPC 和 AI 的集中式基础架构旨在提供通用的共享平台，用于运行传统的 HPC 工作负载和高级分析（深度学习和机器学习）。IBM 的 Summit 和 Sierra 系统是全球排名前两位的超级计算机。它们能够解决需要突破性性能（即智能模拟）的高难度问题，推动 HPC、HPDA 和 AI 的融合。Hyperion Research 认为，IBM 已做好准备，能够加速 HPC、HPDA 和 AI 的融合。预计不断融合的 HPC-HPDA-AI 市场将持续增长，而 IBM 将继续从该市场大获裨益。

关于 Hyperion Research, LLC

Hyperion Research 针对高性能计算和新兴技术领域的技术、应用与市场提供由数据驱动的研究、分析和建议，帮助全球各地的企业制定有效的决策，把握业务增长机遇。他们的研究包括市场规模和预测、共享跟踪、细分、技术和相关趋势分析，以及适用于 HPC 和 HPDA（高性能数据分析）的多用户技术服务器技术的用户和供应商分析。我们聚焦政府、行业、商业和学术领域的重要市场和技术趋势，为用户、供应商和其他 HPC 社区成员提供思维领导力和实用的指导。

总部

365 Summit Avenue

St. Paul, MN 55102

USA

612.812.5798

www.HyperionResearch.com 或 www.hpcuserforum.com

版权声明

2019 Hyperion Research LLC 版权所有。未经授权，严禁翻录。保留所有权利。有关更多信息，敬请访问 www.HyperionResearch.com。如需了解有关重印、其他副本、Web 权利或引用许可的信息，请拨打电话 612.812.5798 或发送电子邮件至 ejoseph@hyperionres.com。