



业务挑战

目前，最先进、最令人兴奋的数据科学技术大多源自于开源社区。五三银行（Fifth Third Bank）如何放心地将这些工具整合到自己高度安全的 IT 环境中？

转型

五三银行采用 IBM® Data Science Experience 构建了功能强大、易于扩展的环境，不仅支持访问 Jupyter Notebook、R 和 Python 库，还可满足 IT 安全要求。

业务效益：

解锁

潜力，充分利用尖端的机器学习与深度学习方法

保持

安全，一站式控制所有开源工具和数据

扩展

过程无缝衔接、经济高效，有效满足不断增加的分析需求

五三银行 为最先进的数据科学构建高度安全的环境

五三银行是一家多元化的金融服务公司，总部位于俄亥俄州辛辛那提市。截至 2018 年 3 月 31 日，五三银行拥有 1420 亿美元的资产，在俄亥俄州、肯塔基州、印第安纳州、密歇根州、伊利诺斯州、佛罗里达州、田纳西州、西弗吉尼亚州、乔治亚州和北卡罗来纳州运营着 1153 个全业务银行中心及 2459 台自动取款机。

“Data Science Experience 解决方案使我们能够在相同的时间内研究更多模型、进行更多次迭代，帮助我们快速实现业务所需的成果。”

Jeffrey Allard

五三银行决策科学团队首席数据科学家

分享本文



拥抱新一代数据科学工具

在过去几年中，数据科学逐步从统计学范畴中脱离出来，成为许多企业分析战略的核心支柱。硬件的进步，例如高效 GPU 集群的出现，使得企业能够经济高效地训练复杂的神经网络；与此同时，新型软件框架帮助简化模型的设计、开发与部署过程。

从企业的角度而言，真正令人吃惊的是，数据科学工具的许多创新都来自开源社区。在许多情况下，当前毕业于高等院校的数据科学家们，其主要专业技能都是 R 和 Python 之类的开源语言、TensorFlow 和 Keras 之类的框架，以及 Jupyter Notebook 之类的开发工具，而不再是传统的统计软件。

五三银行的决策科学团队 (DSG) 很快认识到，开源软件有能力帮助他们获得竞争优势 — 不仅能优化市场营销分析，还可吸引顶尖的数据科学人才。



当五三银行决定开始重塑数据科学实践时，DSG 团队认识到他们迎来了拥抱开源技术的黄金机会。该团队决定完全摒弃传统的建模平台，不再重写数千行代码以将数据导入传统模型，而是采用 Python 和 R 语言重新构建模型。

然而，确保安全成为他们面临的一大挑战。在许多行业，数据科学家可以自由地下载最新软件框架，并且立刻开始构建模型 — 然而在金融服务环境中，数据监管受到法规的严格约束，因此必须对客户数据的使用实施更为严格的安全保障。

构建安全的数据科学环境

因此，五三银行决定构建一个 Hadoop 集群，有效支持大数据分析和训练模型。通过在内部部署集群，五三银行可确保所有敏感数据始终保持在自己的防火墙内，同时保证符合 IT 安全策略。

决策科学团队首席数据科学家 Jeffrey Allard 表示：“我们研究了不同供应商的各种 Hadoop 产品，但我们对大数据基础架构本身还是不太了解。我们的要求主要集中在数据科学工具及其安全性方面。”

当该团队观看了 IBM Data Science Experience 的演示后，他们敏锐地发现，该解决方案能够满足自己的需求。Data Science Experience 作为一个平台，为数据科学团队提供广泛的开源工具和企业工具，涵盖从数据准备到模型开发的整个数据科学生命周期。Data Science Experience 是一种混合型解决方案，可在本地部署，也可通过公共云或私有云访问。

五三银行首席应用开发师 Brian Robinson 表示：“问题在于，作为银行，我们该如何安全地利用开源软件，构建一流的模型？这需要我们建立受控环境，不仅支持数据科学家访问这些工具，同时仍能满足我们的 IT 安全标准。”

“首先，我们只是在 Data Science Experience 的云版本中设置个人帐户，了解如何使用 Jupyter Notebook 进行模型开发，”Jeffrey Allard 表示，“如果我们能够通过 Data Science Experience Local 在本地部署相同的解决方案，那么就可以建立所需的受控环境，使用真实的银行数据构建模型。”

于是五三银行与 IBM 开展合作，建立了一个具有三个节点的小型 IBM Data Science Experience Local 集群。随着 DSG 团队开始将更多的模型训练工作负载转移至该平台上，集群环境迅速扩展，一开始增加到九个节点，然后是十一个节点。

Brian Robinson 表示：“从 IT 角度来看，由于 Data Science Experience Local 采用横向扩展方法，因此可以实现无缝增长。如果需要更多计算能力，只需添加更多节点即可。”

Data Science Experience 支持五三银行的 DSG 或其他团队在受控环境中访问所需的数据、开源工具以及框架。这样，数据科学家就不能随意将可能已被恶意软件感染的软件包下载到自己的 PC 中，与此相反，每个软件包都要由银行的安全团队集中下载并扫描杀毒后，再提供给用户使用。



DSG 团队已在 Data Science Experience 中使用 Jupyter Notebook 构建新模型，用以解决一系列广泛的营销和客户分析问题。例如：该解决方案帮助直接营销团队在潜在客户数据库中找到合适的目标；预测新客户账户的余额和收入；以及评估未来五年客户流失的可能性。

“除使用较为传统的统计模型外，我们还在使用机器学习和深度学习技术。”Jeffrey Allard 表示，“例如，我们利用神经网络来进行预测和序列建模，帮助进行客户价值分析和营销优化。

借助 Data Science Experience 中先进的自然语言处理框架，我们可以采用传统平台难以实现的方法对非结构化文本进行分析。”

更加快速、安全、经济高效

与传统平台相比，Data Science Experience 可帮助五三银行实现更准确更快速的模型训练。

Jeffrey Allard 表示：“即使只是在笔记本电脑上训练模型，新的开源工具也能让我们实现显著的性能提升。我们在 Data Science Experience 集群中运行大型训练作业时，不存在资源争用情况 — 这个新平台使我们能够在相同的时间内研究更多模型、进行更多次迭代，帮助我们快速实现业务所需的成果。”

更重要的是，DSG 团队可放心使用新的开源工具和框架，无需担心数据监管和安全隐患。Data Science Experience 中提供的每个软件包都经过了银行安全专家的彻底检查和审计，这对降低风险和推动遵守信息安全策略都十分有利。

五三银行还希望能够在不大幅增加基础架构成本的前提下扩展该环境。Brian Robinson 表示：“集群中的每个节点只是一个虚拟 Linux 服务器，而这些服务器可在商用硬件上运行。因此，与升级支持传统平台的企业级服务器相比，添加新节点更加经济实惠。”

DSG 团队希望未来可以利用 Data Science Experience 的更多功能，如自动化模型部署。该团队还计划向企业中的其他团队提供培训，帮助他们将模型转移至新平台，从而更有效地共享技能和数据。

Jeffrey Allard 总结道：“依托 Data Science Experience 提供的强大平台，我们打造五三银行数据科学的未来。该解决方案支持我们在安全、快速和可扩展的环境中使用最先进的开源工具，帮助我们转变客户分析模式，并吸引新一代数据科学家加入五三银行，携手书写辉煌篇章。”

“借助 Data Science Experience 中先进的自然语言处理框架，我们可以采用传统平台难以实现的方法对非结构化文本进行分析。”

Jeffrey Allard 是五三银行决策科学团队的首席数据科学家

解决方案组件

IBM® Data Science Experience

后续行动

IBM Data Science Experience 是企业数据科学平台，支持团队研究和构建数据科学实践并更快地投入生产环境。有关 IBM Data Science Experience 如何利用数据转变行业和相关职业的更多信息，请访问：

[ibm.com/products/data-science-](https://ibm.com/products/data-science-experience)

[experience](https://ibm.com/products/data-science-experience)。请关注我们的 Twitter

@IBMDataScience 和博客

ibmbigdatahub.com 并参与对话 #DSX。

联系我们



© Copyright IBM Corporation 2018, IBM Corporation, 1 New Orchard Road, Armonk, NY 10504 U.S.A. 美国出品 2018 年 8 月

IBM、IBM 徽标和 ibm.com 是 International Business Machines Corporation 在全球许多司法管辖区的注册商标。其他产品和服务名称可能是 IBM 或其他公司的注册商标。Web 站点 ibm.com/legal/copytrade.shtml 上的“Copyright and trademark information”部分中包含了 IBM 商标的最新列表。

IBM 并不一定在开展业务的所有国家或地区提供所有这些产品或服务。

本文引用的性能数据和客户示例仅用于说明目的。实际性能结果可能因特定配置和运行条件而异。

引述的客户示例都只是表明一些客户使用 IBM 产品的方式及其取得的成果。实际环境成本和性能特征将根据不同的客户配置和条件而变化。欢迎联系 IBM，我们随时为您效力。

评估和验证任何其他产品或程序与 IBM 产品和程序一起运行的情况，则由用户自行负责。

客户负责确保遵守任何适用的法律和法规。IBM 不提供法律建议，也不表示或保证其服务或产品会确保客户遵守任何法律法规。

良好安全实践声明：IT 系统安全性涉及通过预防、检测和响应来自企业内外的不当访问，保护系统和信息的安全。不当访问可能导致信息被篡改、损毁或滥用，也可能导致您的系统遭到损坏或误用，包括用于攻击他人。任何 IT 系统或产品都不应被认为是完全安全的，并且没有任何单一产品、服务或安全措施对于防止不当使用或访问是完全有效的。IBM 系统、产品和服务旨在成为合法而全面的安全方法的一部分，该方法必然会涉及其他操作程序，并可能需要其他系统、产品或服务配合才能获得最好的效果。IBM 不保证任何系统、产品或服务可以免受任何一方的恶意或非法行为的影响，或使您的企业可以免受此类影响。



请回收利用

13017713-CNZH-00

