

红帽助力金融服务领域 加快采用 AI

通过端到端平台缩短
AI/ML 解决方案的
上市时间

AI 模型日益复杂，采用之旅面临更多挑战

金融机构正纷纷寻求采用人工智能（AI），积极把握它所带来的重大机遇。深度学习、对话式和生成式 AI 等领域的快速发展极大地拓展了 AI 解决方案的范围和适用性。与此同时，模型复杂性的增加不仅带来了新的执行挑战，还突显了现已面临的挑战。其中一些挑战包括：

- ▶ **独立的开发流程：**目前，大多数 AI 和机器学习（ML）的开发和训练都是在专用环境中进行，需要采用特殊的资源，例如图形处理单元（GPU）等加速硬件。部署 AI/ML 环境需要很长时间，这是推出基于 AI 的新服务所面临的一大障碍。
- ▶ **可扩展性、灵活性和资源优化：**AI/ML 解决方案需要具有不同资源需求的组件，如中央处理单元（CPU）、内存、磁盘和专用硬件（GPU）、张量处理单元（TPU）和现场可编程门阵列（FPGA）。扩展此类解决方案通常需要混合云方法。
- ▶ **监控和偏移：**AI/ML 模型需要持续监控和定期更新，以检测和纠正偏移。红帽® OpenShift® 提供了一个基于标准的监控基础架构，可以将基于应用的偏移监控与 AI/ML 开发管道连接起来，从而促进模型更新的持续集成。
- ▶ **模型供应链安全防护：**AI/ML 开发人员工具生态系统在很大程度上基于社区驱动的开源框架。在这种环境下，确保软件供应链安全是一项日益严峻的挑战。开发人员需要最新的工具，但企业需要确保这些工具安全可靠、经过安全强化，并且不包含任何易受攻击或恶意的工件。

显著降低复杂性的优势

拟议的 AI/ML 解决方案可为金融机构带来以下优势：

- ▶ 提供用于模型开发、训练和推理的端到端平台。这可以实现一致的公共云和私有云运维，并减少流程不同阶段之间的摩擦。
- ▶ 自助服务功能可缩短 ML 环境的价值实现时间。
- ▶ 一套一致的最新开源 ML 工具和库，以及广泛的开源和合作伙伴支持的技术生态系统。
- ▶ 快速开发和部署 ML 模型，同时利用监控和快速迭代功能，确保所部署的模型始终保持最新状态。



红帽官方微博

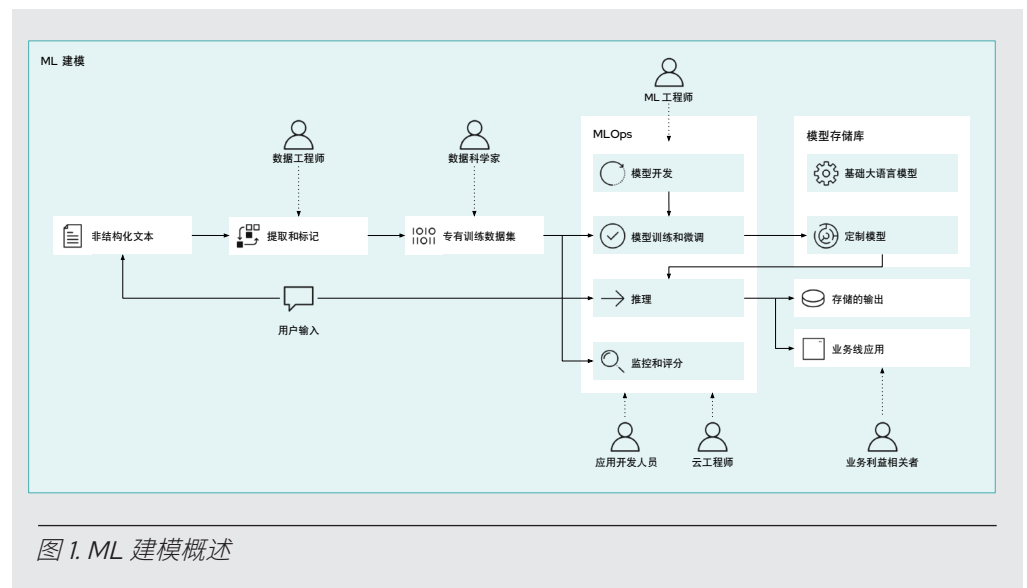


红帽官方微信

案例研究：大语言模型

我们可以通过一个示例来了解金融服务领域所面临的挑战和可获得的优势，即实施基于大语言模型（LLM）的解决方案，如 GPT-4、BLOOM、BART、DOLLY 等。这些类型的解决方案用于以下几个方面：对文件进行数字化处理，以便在客户入驻或了解客户（KYC）流程中使用；分析环境、社会和公司监管数据（ESG）报告；或者实施对话解决方案，如聊天机器人。

在此类解决方案中，通常需要依赖拥有数亿或数十亿参数的大型 ML 模型。鉴于创建这些模型的工作量大、复杂性高且需要强大的计算能力，通常会基于预训练或基础模型进行构建。而这些模型通常是针对通用数据集进行训练，因此将它们应用于金融服务的特定场景需要通过微调或迁移学习，就本地的小规模数据集进行额外的领域或公司特定训练。图 1 展示了此类解决方案的示例架构。



功能概述

解决方案架构

红帽提供了一个平台，能够高效且富有成效地托管从开发、训练到推理的整个 AI/ML 生命周期。红帽平台技术可运行于从裸机和本地虚拟化到主流公共云等各种主流基础架构。这意味着采用相同的平台、相同的工具以及相同的 MLOps 流程。

红帽深知开源的意义和保护软件供应链的重要性，因此与上游社区合作开发出出色的新软件并建立相互信赖的良好关系。在此过程中，红帽精心挑选 AI/ML 开发人员所需的大量上游工具，并提供相应支持和认证。红帽可代您参与上游供应链的工作，并为您提供值得信赖的可靠产品和全天候支持服务。

平台组件

操作系统

红帽 AI/ML 架构的基础是红帽企业 Linux®，这是一种操作系统（OS），可在本地、云环境、裸机或虚拟机上的现代部署基础架构上运行。红帽企业 Linux 经认证可在最广泛的硬件生态系统和主流云提供商环境中运行，包括 Amazon Web Services（AWS）、Google Cloud、IBM Cloud for Financial Services、Oracle Cloud 和微软 Azure。该 Linux 平台通过红帽 Ansible® 自动化平台实现安全防护、高性能、全面支持和一流的自动化。此外，红帽企业 Linux 还支持用于 AI/ML 模型开发的专用硬件，包括 GPU 和 FPGA。

容器编排

除了定制版和商业版应用，AI/ML 过程中使用的绝大多数开源工具和库都是容器化的。预训练或生产型 ML 模型也打包为容器镜像。此外，AI/ML 流程涉及多个组件，这些组件必须彼此交互并可灵活扩展。其中包括计算密集型新模型训练、高吞吐量的推理引擎以及数据科学家使用的模型开发环境，这一切都需要一个灵活的弹性平台。红帽 OpenShift 是部署和编排容器化工作负载的行业领导者，它是 Kubernetes 的一个发行版。它也是第三方和开源 AI 开发工具的最受欢迎平台，可确保您的开发团队能够使用所需的 AI/ML 框架来加速实现价值。红帽 OpenShift 还引入了运维人员技术，可自动部署组件，实现自助服务并降低运维成本。

可扩展的安全强化型存储

AI/ML 项目需要大量的训练数据来构建精确的模型。这些数据可以是历史数据，也可以是来自市场数据源、物联网（IoT）和可观测性等来源的实时数据。在任何情况下，数据都必须以对用户友好且可重复访问的方式存储，供开发人员使用。红帽以基于红帽 Ceph® 存储的红帽 OpenShift Data Foundation 的形式支持和集成开源软件定义存储。OpenShift Data Foundation 是一个软件定义存储解决方案，与红帽 OpenShift 集成，可经济高效地扩展到 PB 级及以上。流数据可以通过基于 Apache Kafka 的 AMQ 流来使用，为开发人员提供对流数据的可重放访问。OpenShift Data Foundation 和 AMQ 流都打包在容器中，可通过红帽 OpenShift 进行管理，以便多个开发人员团队以自助方式进行操作。

平台功能

自助服务

使用红帽 OpenShift，开发人员团队和项目可按需加入，并可根据需要扩大或缩小资源规模。此外，昂贵的专用硬件（如 GPU）也可以池化和共享。安全合规措施和软件供应链安全防护已内置于各个层面。

高级监控和可观测性

红帽 OpenShift 包括通过 Prometheus 实现的开源行业标准监控，并提供与第三方监控工具（如 Splunk）的兼容性。这样就可以将 MLOps 管道与灵活的集中式基础架构集成，以便在整个管道中进行监控和警报。跟踪模型性能可自动进行扩展，并针对低准确度水平发出警报。

敏捷性

AI/ML 建模是一个迭代过程。从业者（包括数据工程师和数据科学家）探索数据轨迹留下的路径，而模型开发之旅充满了起点和终点、不可预见的途径、陷阱以及死胡同。典型的挑战包括从数据库、文件系统、数据流、应用编程接口（API）等不同来源获取高质量数据，以及遵守监管义务和安全标准。在工具方面，挑战包括各种库的版本控制，以及更新现有工具和采用新工具。红帽通过在混合云环境中为从业者提供一致的体验，帮助他们简化 AI/ML 管道，从而加速推进 AI/ML 项目。

传统应用开发和 AI/ML 应用开发之间的一个区别在于，对应用本身或应用核心的 AI 模型进行更新的显著需求。AI/ML 技术不仅可以通过 ML 对模型进行初始训练，还可以对模型进行持续更新。这使得模型能够提供传统应用所不具备的优势，但也意味着需要定期“闭环”并更新模型以提高性能。红帽 OpenShift 支持应用团队透明地扩展和缩减 MLOps 工具链的组件。当应用需要更新模型时，可以手动分配和扩展配备 GPU 或其他专用组件的训练资源（成本较高）。更新完成后，红帽 OpenShift 可以将这些资源重新分配到最需要的地方。

训练和推理的可扩展性和弹性

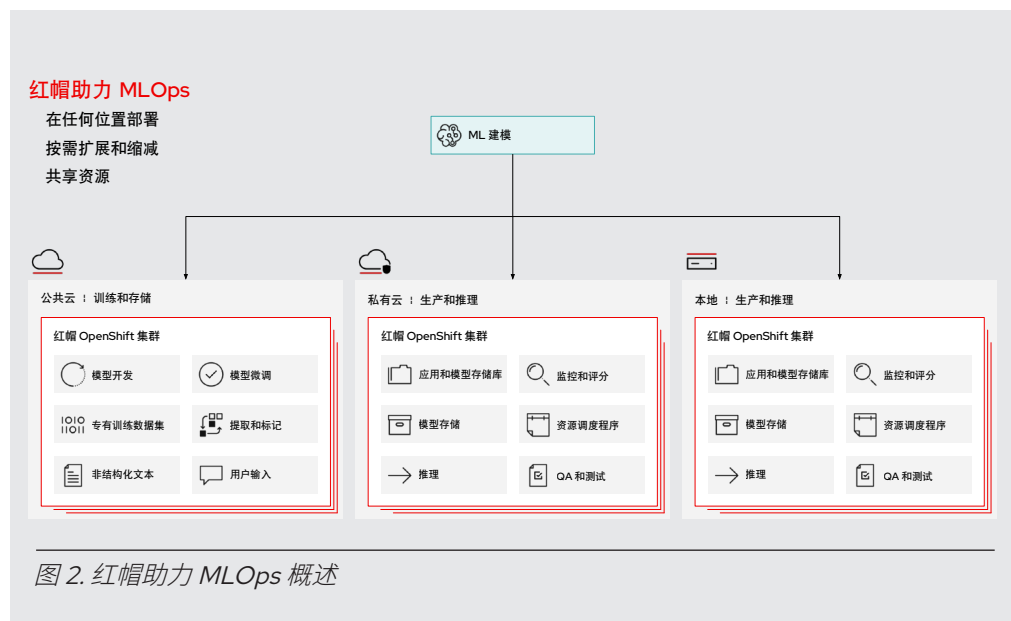
AI/ML 建模的训练阶段是 MLOps 管道中资源最密集的操作之一。这是 AI/ML 工具最大规模扩展实例的所在之处，也是 Nvidia 等公司对专用硬件（如 GPU、TPU 和 FPGA）需求最大的阶段。各个项目和团队都希望能够访问自己的环境来进行训练。红帽 AI/ML 架构能够提供共享基础架构，在效率和经济性方面具有显著优势。红帽 OpenShift 为开发人员提供对整个集群的虚拟按需访问，而不是以高昂的成本囤积专用资源。Kubernetes 则可以编排和协调此访问，以确保在业务需要的时间和地点提供此类资源。

开放生态系统

与所有红帽产品一样，红帽 AI/ML 平台是完全开源的。AI/ML 从业者可用的开源工具和技术生态系统包括：

- ▶ ML 库。
- ▶ AI/ML 生命周期管理。
- ▶ 数据访问、数据质量和元数据管理。
- ▶ 偏差检测和可解释性。
- ▶ 预训练模型。

得益于生态系统的开放性和平台的灵活性，这些工具可以根据解决方案的需要，以各种组合方式一起使用。此外，开放式平台还允许将新兴技术、工具和模型不断接入解决方案，从而支持持续创新。



关于红帽

红帽是世界领先的企业开源软件解决方案供应商，依托强大的社区支持，为客户提供稳定可靠且高性能的 Linux、混合云、容器和 Kubernetes 技术。红帽致力于帮助客户开发云原生应用，集成现有和新的 IT 应用，并实现复杂环境的自动化和管理。作为深受《财富》500 强公司信赖的技术顾问，红帽旨在提供一流的支持、培训和咨询服务，努力将开放创新的优势赋能于各行各业。红帽作为全球企业、合作伙伴和社区网络的互连枢纽，致力于帮助企业发展、转型，并拥抱数字化未来。



红帽官方微博



红帽官方微信

销售及技术支持

800 810 2100
400 890 2100

红帽北京办公地址

北京市朝阳区东大桥路 9 号侨福芳草地大厦 A 座 8 层 邮编: 100020
8610 6533 9300