

跨混合云开发、训练、测试和部署 ML 模型

亮点

在本地或公共云中快速交付智能应用。

简化 AI 在企业中的采用，提高 AI 采用率，且具有灵活性。

通过一致的用户体验在不同的团队之间建立 AI/ML 运维的一致性，让数据科学家、数据工程师和 DevOps 团队能够高效协作。

可在云中、本地或边缘（靠近数据所在位置）构建、训练、部署和监控 AI/ML 工作负载，获得混合云的灵活性。

[Open Data Hub](#) 是一个用于在红帽 OpenShift 上构建人工智能即服务 (AIaaS) 平台的蓝图。它继承了 Jupyter 和 [Kubeflow](#) 等上游项目的工作，并且是红帽 OpenShift AI 的基础。

加快人工智能和机器学习模型的部署

随着生成式 AI (gen AI) 的采用率迅速提高，人工智能、机器学习和深度学习 (AI/ML/DL) 技术已迅速成为企业和组织发展的关键。据 Gartner 预测，到 2026 年，85% 的企业将使用生成式 AI 应用编程接口 (API) 或部署由生成式 AI 提供支持的应用。¹ 不过，这些技术的部署过程可能会很复杂。在数据科学家集中精力构建模型的过程中，他们经常会遇到快速革新的工具之间缺乏一致性的问题。这些问题会对数据科学家、软件开发人员和 IT 运维人员的工作效率和协作性造成不利影响。AI/ML 部署的扩展可能会受到资源的限制，管理起来较为复杂，且需要通过昂贵的图形处理器 (GPU) 资源进行硬件加速以及针对生成式 AI 采用分布式工作负载。主流的云平台具备可扩展性和有吸引力的工具集，但同样，这些工具通常会限制用户在架构和部署方面的选择。

红帽® OpenShift® AI² 基于开源的 [Open Data Hub](#) 项目，可让数据科学家快速训练、测试、提供和监控包括生成式 AI 模型在内的 ML/DL 模型。用户可以立即专注于模型的构建和应用的开发，无需等待基础架构的置备。作为红帽 OpenShift 的附加组件，OpenShift AI 既可作为全托管式云服务，也可作为自管式软件产品，它将红帽组件、开源软件以及由技术合作伙伴提供的产品相结合，具有在本地、云或边缘基础架构中开发和提供模型的灵活性。

红帽 OpenShift AI

借助 OpenShift AI，企业组织可以高效部署构建 AI/ML 模型所需的一系列可集成的常用开源工具和第三方工具。该平台提供了各个云提供商提供的指定和专用 AI/ML 套件的替代方案。采用者可以获得协作式开源工具集和一个用于构建实验性模型的平台，无需担心基础架构或受限于特定公共云工具。然后，采用者可以使用合作伙伴工具扩展该基础平台，从而获得更多功能。可以跨混合云和边缘环境，以一致的方式将模型以容器就绪型格式部署到生产环境中。OpenShift AI 在经过验证、可扩展且安全至上的平台上提供简单明了的配置，为 IT 运维人员提供了一个易于管理的环境。

OpenShift AI 支持热门的生成式 AI 基础模型，您可以根据自己的独特用例及数据进行提示调整、微调以及提供预训练模型。您甚至可以使工作负载分布于多个红帽 OpenShift 集群上，不受集群所在位置的限制。该平台还使硬件加速变得更简单，支持基于中央处理器 (CPU) 和图形处理器 (GPU) 的硬件基础架构，其中包括英伟达 GPU 和英特尔 XPU，且均无需您亲自对数据科学平台进行管理。



红帽官方微博



红帽官方微信

¹ Gartner 新闻稿，“Gartner 表示，到 2026 年，80% 以上的企业将使用生成式 AI API 或部署由生成式 AI 提供支持的应用”，2023 年 10 月 11 日。

² 以前称为红帽 OpenShift 数据科学

红帽咨询团队通过提供与 OpenShift AI Pilot 相关的咨询服务来帮助企业组织开启 OpenShift AI 之旅并将其与现有企业集成。

如果企业组织希望在完成模型实验后制定将模型部署到生产环境中的策略，红帽咨询团队还可提供 [MLOps 基础](#) 咨询服务。

上游开源和商业技术合作伙伴工具

红帽 OpenShift AI 提供了上游 Open Data Hub 项目中的一部分工具（表 1）。企业组织可以在任何云环境、全托管式及自管式红帽 OpenShift 上开发、测试和部署模型并集中监控其性能。红帽会对开源工具（例如 Jupyter、Pytorch 和 Tensorflow）进行定期更新，从而减轻集成、测试和维护负担。该产品还可与多款由技术合作伙伴提供的 AI/ML 产品（表 1）集成。您还可以通过 30 多个已在红帽 OpenShift 上完成产品认证的 AI 技术合作伙伴增加来自技术合作伙伴的商用产品。

表 1. 红帽 OpenShift AI 生态系统

AI/ML 建模和可视化工具	JupyterLab UI ：提供预构建的 Notebook 镜像以及常用的 Python 库和软件包； TensorFlow ； PyTorch 、CUDA；用于管理多个 Notebook 会话的 Kubeflow Notebook 控制器 、 Anaconda （可选用专业版）； 来自英特尔的 AI 工具
数据工程	Starburst （可选择 Galaxy 版或 Enterprise 版）； Pachyderm （可选用）
数据接入和存储	红帽 AMQ （可选用的附加组件）； Amazon Simple Storage Service (S3)
GPU 支持	英伟达 （包含 GPU Operator）、 英特尔 XPU （包括 英特尔至强处理器 、 Habana Gaudi 处理器 以及 英特尔数据中心 GPU Flex 系列 ）
提供和监控模型	提供模型（具有用户界面的 KServe ）、 监控模型 、 OpenShift 源至镜像 (S2I) 、 红帽 OpenShift API 管理 （可选用的附加组件）、 英特尔发行版 OpenVINO 工具套件
数据科学管道	数据科学管道 （Kubeflow 管道）可将数据准备、构建模型和提供模型等流程连在一起



关于红帽

红帽帮助客户跨环境实现标准化，支持他们开发云原生应用，并利用红帽一流的支持、培训和咨询服务，实现复杂环境的集成、自动化、安全防护和管理。



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