

ML in der Hybrid Cloud: Entwicklung, Training, Tests und Deployment

Highlights

Schnelle Bereitstellung intelligenter Anwendungen On-Premise oder in der Public Cloud.

Vereinfachte Einführung von KI in Ihrem Unternehmen, erhöhte KI-Nutzung und Flexibilität.

Operative KI/ML-Konsistenz in mehreren Teams durch ein konsistentes Benutzererlebnis, das Data Scientists, Data Engineers und DevOps-Teams eine effektive Zusammenarbeit ermöglicht.

Hybrid Cloud-Flexibilität durch die Möglichkeit, KI/ML-Workloads in der Nähe der Daten in der Cloud, On-Premise oder am Edge zu entwickeln, zu trainieren, bereitzustellen und zu überwachen.

[Open Data Hub](#) ist ein Blueprint für die Entwicklung einer AlaaS-Plattform (AI-as-a-Service) auf [Red Hat OpenShift](#). Er basiert auf Upstream-Projekten wie Jupyter und [Kubeflow](#) und ist die Basis für Red Hat OpenShift AI.

Beschleunigte Deployments von künstlicher Intelligenz und maschinellem Lernen

Künstliche Intelligenz, maschinelles Lernen und Deep Learning (KI/ML/DL) haben sich seit der rapiden und vermehrten Einführung generativer KI (gen KI) schnell zu einem entscheidenden Faktor für Unternehmen und Organisationen entwickelt. Gartner schätzt, dass 85 % der Unternehmen im Jahr 2026 gen KI-APIs (Application Programming Interfaces) verwendet oder gen KI-fähige Anwendungen bereitgestellt haben werden.¹ Das Deployment dieser Technologien kann jedoch kompliziert sein. Data Scientists wird die Entwicklung ihrer Modelle häufig durch die fehlende Abstimmung zwischen sich schnell entwickelnden Tools erschwert. Diese Lücken können sich negativ auf die Produktivität und Zusammenarbeit von Data Science-, Softwareentwicklungs- und IT-Operations-Teams auswirken. Die Skalierung von KI/ML-Implementierungen kann mit begrenzten Ressourcen und einem hohen Verwaltungsaufwand verbunden sein, während gleichzeitig teure GPU-Ressourcen (Graphics Processing Unit) für die Hardwarebeschleunigung und verteilte Workloads für gen KI erforderlich sind. Gängige Cloud-Plattformen ermöglichen Nutzenden zwar eine Skalierung und bieten attraktive Tools, allerdings schränken diese Tools oft die Architektur- und Deployment-Optionen ein.

[Red Hat® OpenShift® AI](#)² basiert auf dem quelloffenen Projekt [Open Data Hub](#) und ermöglicht Data Scientists ein schnelles Trainieren, Testen, Bereitstellen und Überwachen von ML/DL-Modellen, inklusive gen KI. Nutzende können sich sofort auf ihre Modellierungs- und Anwendungsentwicklung konzentrieren, ohne auf die Provisionierung der Infrastruktur warten zu müssen. OpenShift Data AI ist als zusätzliches Add-on zu Red Hat OpenShift entweder als vollständig gemanagter Cloud Service oder als selbst gemanagtes Softwareprodukt erhältlich und kombiniert die Komponenten, Open Source-Software und Angebote der Technologiepartner von Red Hat mit der Flexibilität, Modelle lokal, in einer Cloud oder auf Edge-Infrastruktur zu entwickeln und bereitzustellen.

Red Hat OpenShift AI

Mit OpenShift AI können Unternehmen ein integriertes Set gängiger Open Source- und Drittanbietertools für die KI/ML-Modellierung effizient bereitstellen. Diese Plattform stellt eine Alternative zu den präskriptiven und vorgegebenen KI/ML-Suites dar, die bei einzelnen Cloud-Anbietern erhältlich sind. Anwendende erhalten ein kollaboratives Open Source-Toolset und eine Plattform für die Entwicklung experimenteller Modelle, ohne sich Gedanken über die Infrastruktur oder die Anbieterbindung durch Public Cloud-spezifische Tools machen zu müssen. Sie können diese Basisplattform dann mit Tools von Partnern erweitern, um weitere Funktionen zu erhalten. Modelle können in Hybrid Cloud- und Edge-Umgebungen konsistent in einem containerfähigen Format für Produktivumgebungen bereitgestellt werden. OpenShift AI bietet IT-Operations-Teams eine einfach zu managende Umgebung mit leicht verständlichen Konfigurationen auf einer bewährten, skalierbaren und sicherheitsorientierten Plattform.

OpenShift AI unterstützt beliebte gen KI-Basismodelle, sodass Sie diese vorab trainierten Modelle für Ihre speziellen Use Cases und mit Ihren Daten mit Prompts versehen, feinabstimmen und bereitstellen können. Sie können sogar Workloads unabhängig von Ihrem Speicherort auf mehrere Red Hat OpenShift Cluster verteilen. Die Plattform vereinfacht die Nutzung von KI-Hardwarebeschleunigung und unterstützt CPU- und GPU-basierte Hardwareinfrastruktur (Central and Graphic Processing Units), inklusive Nvidia GPUs und Intel XPUs, ohne dass Sie selbst für die tägliche Verwaltung Ihrer eigenen Data-Science-Plattform sorgen müssen.

¹ Gartner Pressemitteilung. „[Gartner Says More Than 80% of Enterprises Will Have Used Generative AI APIs or Deployed Generative AI-Enabled Applications by 2026.](#)“ 11. Okt. 2023.

² [Ehemals Red Hat OpenShift Data Science](#)

Red Hat Consulting bietet mit [OpenShift AI Pilot](#) eine Beratung für Organisationen an, die OpenShift AI einführen und in ihr bestehendes Unternehmen integrieren wollen.

Für Organisationen, die über Modellexperimente hinaus Strategien für die Bereitstellung von Modellen in der Produktion entwickeln möchten, bietet Red Hat Consulting außerdem einen [MLOps Foundation](#) Consultingservice an.

Tools von Upstream Open Source-Projekten und kommerziellen Technologiepartnern

Red Hat OpenShift AI umfasst einen Teil der Tools aus dem Upstream-Projekt „Open Data Hub“ (Tabelle 1). Unternehmen können Modelle in den verschiedenen Cloud-Umgebungen und in Red Hat OpenShift (entweder vollständig gemanagt oder selbst gemanagt) entwickeln, testen, bereitstellen und ihre Performance zentral überwachen. Red Hat bietet regelmäßige Updates für Open Source-Tools (beispielsweise Jupyter, PyTorch und TensorFlow), wodurch der Integrations- Test-, und Wartungsaufwand entfällt. Das Angebot umfasst auch mehrere Angebote von KI/ML-Technologiepartnern (Tabelle 1). Zusätzliche Angebote von kommerziellen Technologiepartnern können auch von mehr als 30 KI-Technologiepartnern hinzugefügt werden, die ihre Produkte für Red Hat OpenShift zertifiziert haben.

Tabelle 1. Red Hat OpenShift AI Partnernetzwerk

Tools zur KI/ML-Modellierung und -Visualisierung	JupyterLab UI mit vorgefertigten Notebook-Images und gängigen Python-Libraries und -Paketen; TensorFlow ; PyTorch , CUDA ; Kubeflow-Notebook-Controller zur Verwaltung mehrerer Notebook-Sessions, Anaconda (Professional ist optional); KI-Tools von Intel
Daten-Engineering	Starburst (Galaxy und Enterprise sind optional); Pachyderm (optional)
Datenaufnahme und Data Storage	Red Hat OpenShift AMQ (optionales Add-on); Amazon Simple Storage Service (S3)
GPU-Unterstützung	NVIDIA (mit GPU-Operator), Intel XPU s (inklusive Intel Xeon-Prozessoren , Habana Gaudi , und Intel Data Center GPU Flex Series)
Bereitstellen und Überwachen von Modellen	Modellbereitstellung (KServe mit Benutzeroberfläche), Modellüberwachung, OpenShift Source-to-Image (S2I) Red Hat OpenShift API Management (optionales Add-on), Intel-Distribution des OpenVINO-Toolkits
Data Science Pipelines	Data Science Pipelines (Kubeflow Pipelines) verknüpfen Prozesse wie Datenvorbereitung, Entwicklungsmodelle und Bereitstellungsmodelle



Über Red Hat

Red Hat unterstützt Kunden dabei, ihre Umgebungen zu standardisieren, cloudnative Anwendungen zu entwickeln und komplexe Umgebungen mit [vielfach ausgezeichnetem](#) Support, Training und Consulting Services zu integrieren, zu automatisieren, zu sichern und zu verwalten.

[f](#) facebook.com/redhatinc
[t](#) @RedHatDACH
[in](#) linkedin.com/company/red-hat

EUROPA, NAHOST, UND AFRIKA (EMEA)
 00800 7334 2835
 de.redhat.com
 europe@redhat.com

TÜRKEI
 00800 448820640

ISRAEL
 1 809 449548

VAE
 8000-4449549