

# Red Hat OpenShift AI on Dell APEX Cloud Platform for Red Hat OpenShift

Simplify AI deployment and operations

There is incredible enthusiasm around the possibilities of Artificial Intelligence (AI) with 75% of organizations now reporting an increase in budgets to pursue AI initiatives<sup>1</sup>. While AI offers enormous potential, organizations are being challenged to find ways to deliver value with AI for their business. Translating these new capabilities into business outcomes requires a comprehensive plan from IT architecture to data management.

Red Hat OpenShift AI on APEX Cloud Platform for Red Hat OpenShift offers an efficient way to deploy an integrated set of common open source and third-party tools to collaborate on AI/ML model training, serving, monitoring, and lifecycle management. With the jointly engineered validated design, organizations can quickly capitalize on prominent AI use cases.

## Red Hat OpenShift AI

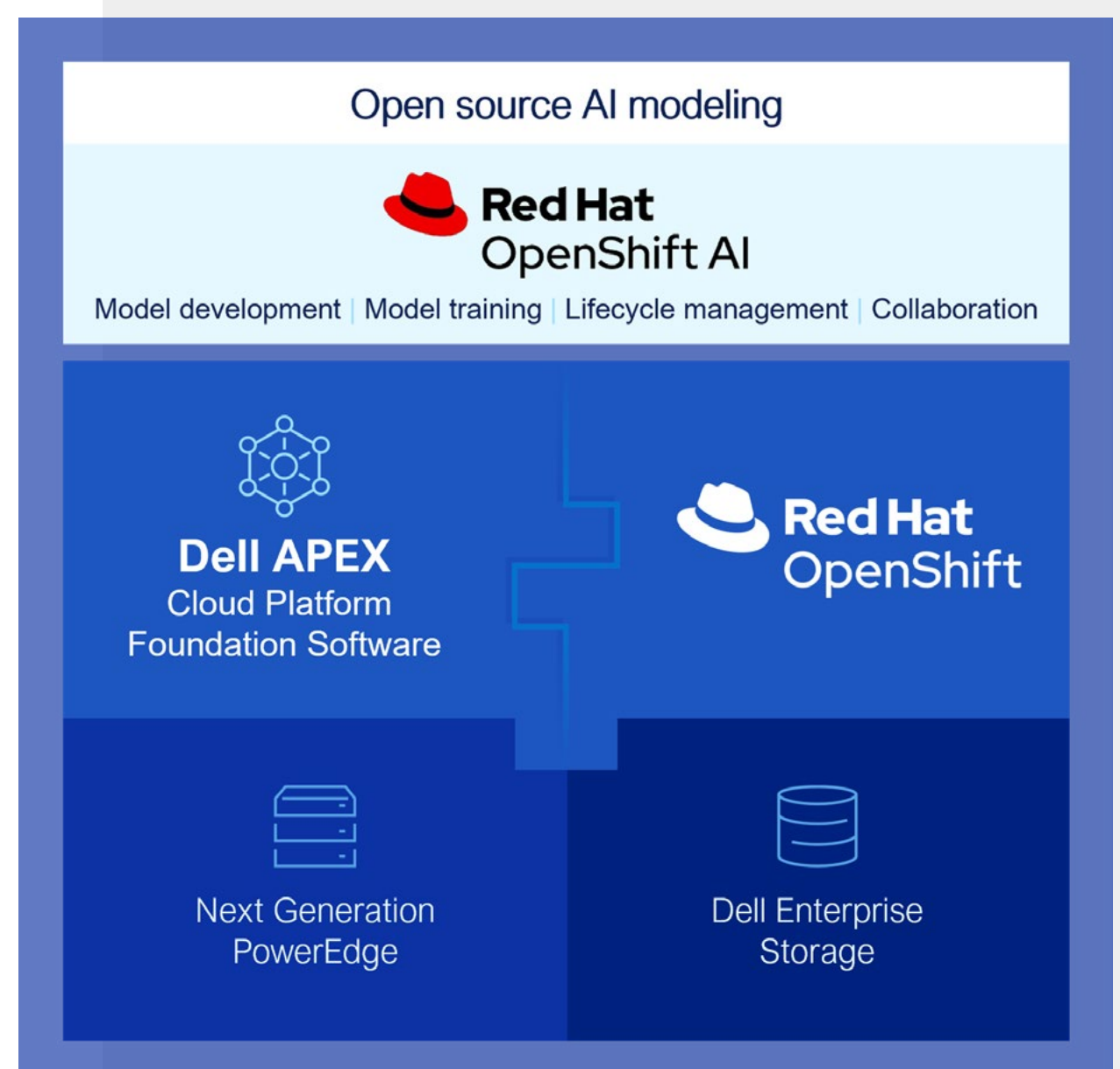
Red Hat OpenShift AI makes it simpler to exploit hardware acceleration, including 5th generation Intel Xeon™ Scalable Processors and NVIDIA GPU supported hardware infrastructure without the need to stand up and perform daily management of Kubernetes on your own.

The platform represents an alternative to prescriptive AI/ML suites available from individual cloud providers. Adopters gain a collaborative open source toolset and a platform for building models without worrying about the infrastructure or lock-in from public cloud-specific tools. They can then extend that base platform with partner tools to gain increased capability. Models can be served and scaled on production environments consistently, across hybrid cloud and edge environments.

## Dell APEX Cloud Platform for Red Hat OpenShift

Dell APEX Cloud Platform for Red Hat OpenShift is designed collaboratively with Red Hat to optimize and extend OpenShift deployments on-premises with an integrated operational experience. **This turnkey platform provides:**

- **Deep integrations and intelligent automation** between layers of Dell and OpenShift technology stacks, accelerating time-to-value and eliminating the complexity of management using different tools in disparate portals.
- **A bare metal architecture** that delivers the performance, security, and linear scalability needed to meet even the most stringent SLAs.
- **Intrinsic multi-layer security rapid availability** of patches and updates, and centralized OpenShift governance to help enterprises maintain a strong security posture.



82%

prefer an on-prem  
or hybrid model  
for Generative AI  
(GenAI)<sup>2</sup>

## Bringing Generative AI to life

With 82% of ITDMs preferring an on-prem or hybrid model for Generative AI (GenAI)<sup>3</sup>, it's more important than ever to have a flexible multicloud architecture that maximizes the value of your data while minimizing complexity. The Dell Validated Designs for Red Hat OpenShift AI on APEX Cloud Platform offer organizations a simple guide on how to quickly deliver real business outcomes with their data for some of the most prominent AI use cases.

[Learn how to deploy a digital assistant](#) utilizing a Large Language Model (LLM) and the Retrieval Augmented Generation (RAG) framework. This integrated multicloud solution illustrates how you can use GenAI with your own data securely on-prem, and implementing an LLM with RAG improves the quality of the information and keeps it current.

Another exciting AI solution is using AI to achieve automated speech recognition (ASR) and text-to-speech (TTS) capabilities. [This design](#) extends the power of Red Hat OpenShift AI on APEX Cloud Platform for Red Hat with NVIDIA Riva, a microservice that builds GPU-accelerated speech AI applications. With this integration you can streamline the deployment of your own Natural Language Processing (NLP) solution.

## World class services accelerate time to value in achieving high value AI use cases

Dell Services bring deep expertise to accelerate time to value of the Red Hat OpenShift AI platform. Our trusted experts will work with you at every stage to drive tangible business value. Rely on Dell's proven methodology from ProConsult Advisory Services to create a winning strategy aligned to high value AI use cases. Leverage Implementation Services to quickly realize a secure, optimized environment. We utilize RAG techniques to customize

models with your data and integrate it into AI Avatars, chatbots or other use cases, providing more relevant and meaningful results. By minimizing time-consuming data preparation and LM training efforts and providing your team with skills, best practices and time to focus on innovation, you will realize maximized ROI now and into the future..

Click to visit our website to learn more about  
Dell APEX Cloud Platform for Red Hat OpenShift →

<sup>1</sup> Dell Technologies Generative AI Pulse Survey, August and September 2023, [www.dell.com/GenAIPulse](http://www.dell.com/GenAIPulse)

<sup>2</sup> Dell Technologies Generative AI Pulse Survey, August and September 2023, [www.dell.com/GenAIPulse](http://www.dell.com/GenAIPulse)

<sup>3</sup> Dell Technologies Generative AI Pulse Survey, August and September 2023, [www.dell.com/GenAIPulse](http://www.dell.com/GenAIPulse)