



Red Hat OpenShift Data Science:

Putting Open Source ML to work
with Intel oneAPI AI Analytics
Toolkit and Intel OpenVINO

EXECUTIVE SUMMARY

If organizations around the world aren't betting their entire futures on the promise of AI and machine learning, they are certainly betting large amounts of their budgets for this new and emerging technology.

However, far too many data science and AI projects fail to make it out of the experimental or pilot phase and into production.

This whitepaper outlines what is at stake, the reasons why organizations are failing to "operationalize" AI and machine learning, and how recent trends in mainstream software development and operations can also benefit data scientists and AI developers.

It explains how Red Hat OpenShift Data Science platform, in concert with the Intel® oneAPI AI Analytics Toolkit and OpenVINO™ Pro for Enterprise, provides a vendor neutral, cloud native platform and workflow across the AI project lifecycle. This enables data scientists and AI developers to focus on what they do best: develop AI and ML models and intelligent applications with their preferred tools and frameworks.

INTRODUCTION

TRENDS AROUND AI/ML/DATA SCIENCE

There's no doubt that AI and machine learning has captured the imagination of enterprises and other organizations worldwide. In August 2021, IDC forecast that total spending on AI solutions¹ – including software, hardware and services – would hit \$342bn in 2021 and reach the \$500bn mark by 2025. This equates to roughly a third of all spending on what the researcher defines as "new technologies".

There are a myriad of reasons why companies are pursuing AI and ML initiatives so enthusiastically and are prepared to spend such huge sums. While there are clearly organizations looking to operate at the interface of technology and science fiction, the most cited reasons for investing in AI and machine learning, according to the IDC report, *Why Adopt Open Source Artificial Intelligence Platforms for Enterprise Business Transformation*², are all too familiar to technology professionals.

Companies and other organizations are looking to AI to help them deliver a better customer experience (50 percent) and improve employee productivity (47 percent). Other key drivers are accelerating innovation, which at 45 percent, ties with speeding up new product development.

(1) IDC Press Release. "[IDC Forecasts Companies to Spend Almost \\$342 Billion on AI Solutions in 2021](#)." Document #US48127321, August 2021.

IDC Spending Forecast. "[IDC - Global ICT Spending](#)." Forecast 2020 – 2023.

IDC Press Release. "[Investment in Artificial Intelligence Solutions Will Accelerate as Businesses Seek Insights, Efficiency, and Innovation, According to a New IDC Spending Guide](#)." Document #US48191221, August 2021

(2) IDC White Paper, sponsored by Red Hat. "[Why Adopt Open Source Artificial Intelligence Platforms for Enterprise Business Transformation](#)." Doc. #US47629021, May 2021.

The explosion in AI/ML initiatives parallels broader trends in how companies are recasting their broader technology operations and infrastructure and moving towards cloud deployments. The same IDC research shows cloud already accounts for 57 percent of deployment locations for AI and machine learning (on/off prem private cloud accounts for 19.2 percent, public cloud comes in at slightly less at 17.3 percent, while hybrid cloud and multi-cloud comprise 11.6 percent and 9 percent respectively).

ISSUES AND CHALLENGES

Although commercial and non-commercial organizations are enthusiastically embracing AI/ML, there are obvious challenges in making the leap from experimentation to full blown enterprise applications. IDC research shows that simple cost is the most widely cited challenge, at 55 percent, but the lack of MLOps tools is close behind, at 52 percent, and the lack of responsible AI tools is cited by 49 percent.

This translates into a problem getting AI projects into production. In 2020, IDC research reported that 28 percent of AI/ML projects failed. “Lack of staff with necessary expertise, lack of production-ready data, and lack of an integrated development environment are reported as primary reasons for failure,” the researchers said.

“Increasing the number of successful projects is dependent on better operationalization of AI platforms.”

Gartner has reported an even higher failure rate³, with only half of AI projects making it from pilot to production. Even those “successful” projects took an average of nine months to realize. The research firm said that increasing the number of successful projects is dependent on better operationalization of AI platforms so that “AI solutions can be relied upon to solve enterprise-wide problems”, as well as efficient use of data, model and compute resources.

This leaves developers with a dilemma. They can turn to all-inclusive solutions from cloud service providers and face the possibility of vendor lock-in, a prescriptive choice of tooling and frameworks and less control over costs.

Or they can pursue a similar path to mainstream enterprise tech teams and leverage cloud native technologies such as containerization.

Indeed AI and machine learning developers are already taking advantage of Kubernetes, the container orchestration and management platform. Just over half of respondents to a recent Pulse survey⁴ said

(3) Gartner Press Release, [Gartner Identifies Four Trends Driving Near-Term Artificial Intelligence Innovation](#), September 7, 2021. Gartner is a registered trademark and service mark of Gartner, Inc. and/or its affiliates in the U.S. and internationally and is used herein with permission. All rights reserved.

(4) Red Hat: State of Workloads Adoption
<https://www.redhat.com/en/resources/state-workloads-containers-kubernetes-analyst-paper>

they were deploying AI/ML workloads on Kubernetes. The top workload was databases or data cache, with 80 percent of respondents, while 66 percent were using Kubernetes for data ingestion, cleansing and analytics. It's also significant that 37 percent were using Kubernetes for custom applications.

However, the cloud native/Kubernetes route raises its own challenges. Raw Kubernetes can be a complex technology in its own right. Similarly, open source tooling presents issues around patching and updating. And developers still face the hurdle of ensuring their models and applications are production ready and scalable in the eyes of the broader organization.

RED HAT OPENSIFT DATA SCIENCE

Red Hat OpenShift is the open source vendor's hybrid cloud, enterprise Kubernetes application platform. It allows developers to containerize and deploy enterprise workloads on Kubernetes, whether on prem or in the cloud. Central to Red Hat OpenShift is the concept of "operators", which aim to "take human operation knowledge and encode it into software", easing the management of distributed apps on Kubernetes.

Red Hat OpenShift Data Science is an add-on to Red Hat OpenShift cloud services, which aims to bring the same philosophy to the development, training and testing of machine learning and AI models, and to their deployment. It enables the sort of container-based DevOps workflows and management that have revolutionized enterprise software development for data science developers.

The last decade has seen an explosion of tools and frameworks, often open source, which have taken ML and AI out of the domain of academia and made it a realistic proposition for mainstream organizations.

However, as we have seen, developers face a dilemma. They can construct their own tool chains using open source implementations, but face the headache of handling integration, and ensuring each element is up to date and secure, and potentially repeating this process each time they look to exploit a different architecture.

Alternatively, they can use a commercial cloud service provider offering, knowing it will "just work" but potentially locking them into a limited or proprietary set of tools, and having limited – and potentially expensive - options when it comes to deployment, especially in environments that require more of a hybrid cloud approach.

Red Hat itself describes OpenShift Data Science as a "sandbox environment" where "data scientists and developers can rapidly develop, train, test, and iterate ML and DL models ... without waiting for infrastructure provisioning".

So, for example, OpenShift Data Science supports Jupyter Hub, and offers predefined notebook images, as well as TensorFlow and Pytorch, along with optional, integrated partner offerings from Anaconda for package management and IBM Watson Studio for a host of capabilities including AutoAI. OpenShift Data Science also offers optional support for Starburst Galaxy, for data engineering, and RedHat OpenShift Streams for Apache Kafka for data intake engineering.

For model serving it offers the OpenShift Source to Image tool, which builds source code into container images, meaning models can be containerized and deployed as applications across hybrid infrastructure. Seldon Deploy can optionally be added to serve models and manage their deployments.

“This allows data scientists and developers to evolve and train their models and deploy them at scale to deliver inference and insight, without having to get entangled in the low-level details of open source models and frameworks.”

This allows data scientists and developers to evolve and train their models and deploy them at scale to deliver inference and insight, without having to get entangled in the low-level details of open source models and frameworks. It also ensures they are using the most up to date and secure versions of their chosen tools.

Data scientists can work on the same platform as their mainstream development colleagues, making it easier to construct cross functional teams and workflows, and ensure data science, and the insights it can deliver, is embedded in the organization's software strategy.

Similarly, ops professionals can help their data scientist counterparts get AI models developed and deployed and ensure that systems are able to scale up to support them appropriately, whether at the training or inference stage. And do so without having to worry about the security of their data scientist colleagues' tooling or frameworks.

As industry analysts Omdia declared, with OpenShift Data Science, Red Hat has chosen not to “build a solution-complete MLOps platform” but, rather, has built “an ML ‘workflow’ platform that enables users to assemble their MLOps solutions using the tools with which they are most familiar.”

And because Red Hat will certify that software provided through Red Hat Marketplace will work on OpenShift, this “future proofs customer investments in terms of portability, building trust that their software can deploy this software on any platform equipped with OpenShift, whether on-premises or in the cloud.” This adds up to a “unique approach” to closing the chasm between data scientists and IT professionals.

RED HAT OPENSIFT DATA SCIENCE IN ACTION - working with Intel oneAPI AI Analytics Toolkit

The rapid development of AI and machine learning over the last decade or so has been largely powered by open source tools and the use of hardware accelerators, in particular GPUs.

This heterogeneous approach has fueled experimentation but also contributes to the increasing challenges data scientists face developing their models, and of operationalizing and deploying AI/ML in the enterprise, by forcing them to rework their code base as they move between different platforms and architectures.

The oneAPI specification is a cross industry effort, backed by Intel, that aims to produce a standards-based unified programming model, and single developer framework, across multiple accelerator architectures.

The Intel® oneAPI AI Analytics Toolkit (AI Kit), which is available as an OpenShift operator, offers optimizations for key AI frameworks on Intel architectures, including drop-in accelerations for TensorFlow and PyTorch, and Intel® Extension for Scikit-Learn. Other features include Intel® Distribution of Modin, a distributed DataFrame library to accelerate Pandas workflows, and Intel® Distribution for Python, which provides even more performance accelerations powered by the Intel oneAPI libraries.

It also includes Model Zoo for Intel® Architecture, which contains pre-trained models, scripts and best practices and tutorials, again all optimized for Intel architectures, as well as the Intel® Neural Compressor, a low precision optimization tool which provides a unified low-precision interface across multiple deep learning frameworks.

The oneAPI AI Analytics Toolkit enables developers to take full advantage of the Intel platforms they are working on. For example, the base version of Pandas operates on a single core – which, when modern processors can have dozens of cores, presents an obvious under-utilization issue. In this case, the addition of a few lines of code in the Intel Distribution of Modin in the AI Kit, allows the packages to access multiple cores while still using base Pandas. The result is improved performance and vastly improved utilization of the underlying hardware.

Similarly, it offers developers tools for better vectorization, memory management, and scale scientific computation across a cluster. This has implications not just at the CPU or GPU level. By better utilizing both the processor, and memory, this can result in more efficient use of other infrastructure components, such as storage. This ensures the more efficient use of expensive infrastructure, whether on-prem or in the cloud.

AI Kit on Red Hat OpenShift Data Science eases the decision-making process of what infrastructure to use and where - which can be a time-consuming and stressful process for data scientists and AI developers. The Intel oneAPI AI Analytics Toolkit seamlessly allows data scientists to experiment and develop through simple drop-in accelerations without the concerns of coding low-level optimizations or choosing the right proprietary services - and these optimizations are all available as an operator for the OpenShift Data Science platform.

RED HAT OPENSIFT DATA SCIENCE IN ACTION - working with Intel's distribution of OpenVINO

Open Visual Inference and Neural Network Optimization (OpenVINO) is a toolkit for optimizing and deploying deep learning models onto Intel hardware. There is an open source version, and an Intel-backed version, both of which help developers generate results from their trained model – i.e., inference.

OpenVINO builds on the same optimizations that underpin the Intel oneAPI AI Analytics Toolkit and uses them to speed up and ease model deployment. OpenVINO supports key deep learning frameworks, such as TensorFlow and PyTorch as well as Keras, Caffe, and others. Using their own

trained models, developers can access Intel's Open Model Zoo with pretrained and optimized models which they can put to work instantly.

With OpenVINO, developers can run their model through its Model Optimizer and Post-Training Optimization Tool, to analyze their model for performance and accuracy, and gain inference performance without the need for fine-tuning or retraining.

They can then deploy their model using the Inference Engine, which is a set of libraries providing a common API to run inference across the gamut of Intel platforms, including CPU, iGPU, Intel® Movidius™ Vision Processing Unit (VPU), and Intel® Gaussian & Neural Accelerator.

Intel OpenVINO is available as an operator for OpenShift, meaning the model or application can be containerized and scaled up – or down – using Kubernetes, across private or public cloud infrastructure.

It might be tempting to envision OpenVINO fleshing out the production deployment side of the AI project lifecycle workflow. However, together with OpenShift, it can also play a role in refinement of models earlier in the development process. For example, if a developer is achieving the accuracy they need, but not the performance, they can seamlessly use OpenShift Data Science to open a Jupyter notebook and use the OpenVINO tools to tweak precision and finetune their model.

CONCLUSION

As we've seen, data scientists and ML/AI developers face a number of problems in building workflows that can shepherd their projects through from inception to deployment.

At the initial data ingest and model development stage they are often side-tracked by the need to reinvent their underlying architectures, while much of the tooling they rely on doesn't actually fully utilize their available hardware resources.

When they have developed and trained their model, they have the headache of getting it into production, often having to balance accuracy against performance.

Throughout the process they, and their colleagues in mainstream software development and ops, face uncertainty over who should be in control at each stage. This includes the burden of patching, security, integration, and optimization for the array of tools and frameworks AI developers and data scientists may be drawing on.

To date, AI and data science developers have faced the choice of taking on these burdens themselves, potentially causing friction with other parts of the enterprise, and seeing projects potentially fail when they arrive at the deployment phase. Or they have had to turn to full-solution platforms, giving up autonomy in their choice of tools and deployment platforms, and paying handsomely for the privilege.

These are issues that traditional software developers have addressed, and to a great extent solved, through the adoption of containerization, Kubernetes and DevOps.

“Red Hat has put the same container orchestration tooling into the hands of data scientists, producing an open AI platform that enables them to use the tools they want while relieving them of much of the admin and deployment burdens.”

With the debut of Red Hat OpenShift Data Science, Red Hat has put the same container orchestration tooling into the hands of data scientists, producing an open AI platform that enables them to use the tools they want while relieving them of much of the admin and deployment burdens they would otherwise face.

With Red Hat OpenShift Data Science’s close integration with Intel’s OpenVino and oneAPI AI Analytics Toolkit, data scientists and AI developers are promised not just the removal of administrative and operational headaches. They also gain an optimized pathway from data collection, to development, to deployment across multiple architectures – without the need to tie themselves to all-in vendors.

In short, the platform offers them a cloud native, vendor neutral solution, which allows them to focus on data science, machine learning and AI.



ABOUT RED HAT

Red Hat is the world's leading provider of enterprise open source software solutions, using a community-powered approach to deliver reliable and high-performing Linux, hybrid cloud, container, and Kubernetes technologies. Red Hat helps customers develop cloud-native applications, integrate existing and new IT applications, and automate and manage complex environments. A trusted adviser to the Fortune 500, Red Hat provides award-winning support, training, and consulting services that bring the benefits of open innovation to any industry. Red Hat is a connective hub in a global network of enterprises, partners, and communities, helping organizations grow, transform, and prepare for the digital future.

RED HAT AND INTEL PARTNERSHIP

Spanning more than 25 years, Intel and Red Hat's long history of engineering achievements includes advanced software-defined infrastructure and industry-standard platforms that improve data center agility and flexibility. Together, Red Hat and Intel provide innovative and secure enterprise-level solutions that help our customers gain a competitive advantage.