Top considerations for building a production-ready AI/ML environment

Red Hat

Get more business value from your data

Data is a critical business asset

The amount of data stored globally is expected to grow to 8.9 zettabytes by 2024.[1] In a digital world, your data can be a competitive advantage. Even so, collecting data is only the starting point—how you use your data is the real differentiator.

Artificial intelligence (AI), machine learning (ML), and deep learning (DL) employ data to deliver business insight, automate tasks, and advance system capabilities. These technologies have the potential to transform all aspects of business—from customers and employees to development and operations. Building AI/ML into your software applications can help you achieve measurable business outcomes:

- Increase customer satisfaction.
- Offer differentiated digital services.
- Optimize existing business services.
- Automate business operations.
- Increase revenue.
- Reduce costs.

AI/ML use cases across industries

Healthcare

- Increase clinical efficiency.
- Boost diagnosis speed and accuracy.
- Improve patient outcomes.

Financial services

- Personalize customer services.
- Improve risk analysis.
- Detect fraud and money laundering.

Telecommunications

- Gain insight into customer behavior.
- Enhance customer experiences.
- Optimize 5G network performance.

Automotive

- Support autonomous driving.
- Predict maintenance needs.
- Improve supply chains.

Insurance

- Automate claims processing.
- Deliver use-based insurance services.

Energy

- Predictive maintenance.
- Optimize field operations and safety.
- Energy trading.

Turn data into a business asset

This e-book discusses several technologies for actionable data analysis.

- **Artificial intelligence** involves machines imitating human behavior to perform tasks that typically require human intervention.
- **Machine learning** is a subset of AI that uses algorithms and statistical models to perform tasks without explicit instructions.
- **Deep learning** is a subset of ML that uses layers to progressively extract high-level features from raw input, similar to a human brain.

Read **An executive's guide to real-world AI** to learn more about the business aspects of AI and ML.

1 *IDC, "IDC's Global StorageSphere Forecast Shows Continued Strong Growth in the World's Installed Base of Storage Capacity," May 13, 2020.* idc.com/getdoc.jsp?containerId=prUS46303920.

Build a production-ready AI/ML environment

Deploying AI/ML in production is an iterative process that extends beyond simply creating AI/ML models. The main steps in the AI/ML life cycle are:

1. Set business goals for your AI/ML initiative and share them with all stakeholders.
2. Gather and prepare the data required for your AI/ML initiative.
3. Develop ML/DL models according to your goals.
4. Deploy ML/DL models in your application development process.
5. Implement intelligent, ML/DL-powered applications and start inferencing.
6. Monitor and manage models for accuracy over time.

An open, adaptable AI/ML architecture will help you execute this process more effectively. This architecture requires several key technologies and capabilities:

- **AI/ML and DevOps tools** allow data scientists, ML engineers, and application developers to create, deploy, and manage ML/DL models and applications.
- **Data pipelines** provide cleaned data to data scientists for creating, training, and testing ML/DL models and to application developers for data management needs.
- A **cloud platform** gives data engineers, data scientists, ML engineers, and application developers access to the resources they need to work rapidly.
- **Compute, storage, and network accelerators** speed data preparation, model development, and inferencing tasks.
- **Infrastructure endpoints** provide resources across on-site, virtual, edge, and private, public, and hybrid cloud environments for all stages of AI/ML operations.

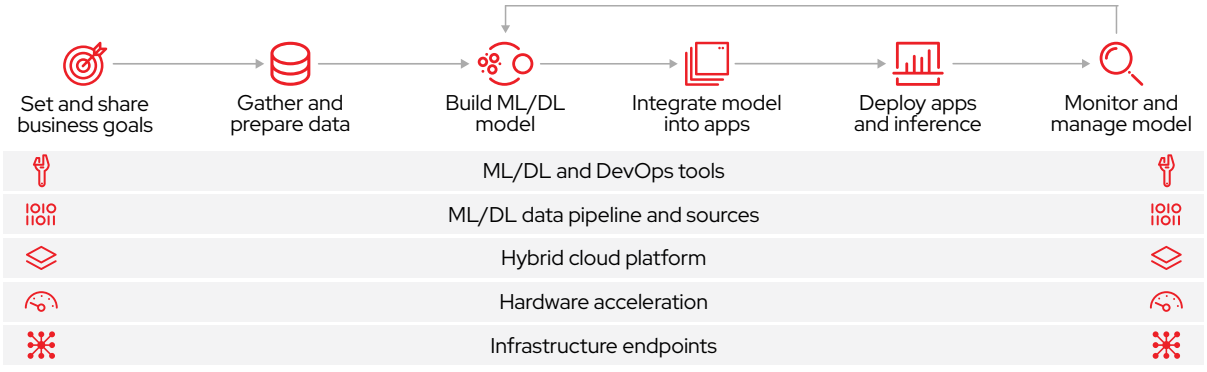
This e-book reviews key considerations for building an effective AI/ML architecture.

AI/ML deployment challenges

Organizations face several challenges when building an AI/ML environment:

- **Talent shortages.** There is a limited amount of AI/ML expertise available, making it harder to find and retain AI/ML staff.
- **Lack of readily usable data.** Businesses collect large amounts of data, but must find and prepare the right data for each AI/ML initiative.
- **Disparate teams.** Slow, manual, and disparate operations can impede AI/ML deployment.
- **Delayed resource availability.** Slow infrastructure and tool delivery can hinder model development, integration, and deployment within applications.

Even so, you can overcome these challenges by applying cloud-native application development approaches to your AI/ML life cycle.



AI/ML environment considerations

Containers and container orchestration

Containers

A **container** is a basic unit of software that packages applications with all of their dependencies. Containers simplify application build processes and allow applications to be deployed across different environments without change.

Why are they important for AI/ML?

Data scientists, ML engineers, and application developers need access to their preferred tools and resources to be most productive. At the same time, IT operations teams need to ensure that resources are up to date, in compliance, and used in a secure manner. Containers let you quickly and easily deploy a broad selection of AI/ML tools across hybrid environments in a consistent way. Teams can iteratively modify and share container images with versioning capabilities that track changes for transparency. Meanwhile, process isolation and resource control improve protection from threats.

Best practices and recommendations

Look for a robust, highly available container platform that includes integrated security features and makes it easy to deploy, manage, and move containers across your environment. Choose an open source platform that integrates with a broad set of technologies to gain more flexibility and choice.

ExxonMobil

ExxonMobil uses an open source container platform to

improve collaboration

between data science teams.[2]

Container orchestration

Container orchestration involves managing the creation, deployment, and life cycle of containers across your environment.

Why is it important for AI/ML?

Once you adopt containers, you need a way to deploy, manage, and scale them efficiently. A container orchestration tool lets you administer the life cycle of your containers in a consistent way. These tools typically centralize access to compute, storage, and networking resources across on-site, edge, and cloud environments. They also provide unified workload scheduling, multitenancy controls, and quota enforcement.

Best practices and recommendations

Select a **Kubernetes**-based container orchestration tool to take advantage of a leading open source technology.

Kubernetes

is the leading container orchestration framework.[3]

2 *Red Hat case study, "**ExxonMobil speeds information sharing, gains agility and productivity**," Accessed June 12, 2020.*

3 *Red Hat, "**Hybrid cloud, enterprise Kubernetes**," Accessed June 12, 2020.*

AI/ML environment considerations

Application management and DevOps

Application life-cycle management

Application life-cycle management involves deploying, scaling, and administering applications that run within containers.

Why is it important for AI/ML?

29% of Red Hat® customers cite managing the compatibility and complexities of an evolving software stack to be a top challenge to their AI/ML initiatives.[4] Container application life-cycle management components integrate with your container orchestration tool to let you directly administer containerized applications, including AI/ML development tools. IT operations teams can automate common life-cycle management tasks like configuration, provisioning, and updates to gain efficiency, speed, and accuracy. Data scientists, ML engineers, and application developers can deploy tools and applications from a preapproved service catalog – without needing to engage IT operations teams. Automation also frees staff from tedious tasks, allowing them to focus on more interesting strategic activities.

Best practices and recommendations

Choose container application life-cycle management tools that include easy-to-use automation and integration with your preferred AI/ML tools. Popular options include **Kubernetes Operators** and **Helm Charts**.

"Seldon's integrations [...] via Kubernetes Operators helps organizations speed up deployment of machine learning models across the hybrid cloud, providing faster roll-outs of AI-powered digital services."[5]

Alex Housley
Founder and Chief Executive Officer, Seldon

DevOps

DevOps is a collaborative approach that brings together people, processes, and technology to speed delivery of high-quality services and applications.

Why is it important for AI/ML?

Organizations need to develop and deploy AI/ML models – and the applications that use them – quickly and efficiently. However, lack of collaboration across teams prevents 87% of data science projects from making it into production.[6] DevOps approaches foster collaboration between AI/ML teams, application developers, and IT operations to accelerate time to production for ML-powered applications. Automation – often in the form of **continuous integration/continuous delivery (CI/CD)** pipelines – makes rapid, incremental, and iterative change possible for faster application development life cycles.

Best practices and recommendations

DevOps is not just about technology – it also involves people and processes. Apply DevOps approaches to your entire AI/ML life cycle. Take advantage of automation in your platforms and tools, as well as open source technologies like Argo, Tekton, Jenkins, and Spinnaker, to create CI/CD pipelines.

DXC.technology

"Without this solution, achieving the right level of analysis and efficiency would take literally millions of years of effort. Red Hat OpenShift makes the deployment of new applications as easy as possible for the entire DevOps team."[7]

Dr. Jochen Thaeder
Chief Architect, High-Performance Data-Driven Development (D3) Platform, DXC Technology

*4 Red Hat, "**2020 Red Hat Global Customer Tech Outlook**," November 2019.*

*5 Red Hat press release, "**Red Hat Accelerates AI/ML Workflows and Delivery of AI-Powered Intelligent Applications with Red Hat OpenShift**," March 24, 2020.*

*6 VentureBeat, "**Why do 87% of data science projects never make it into production?**," July 19, 2019.*

*7 Red Hat case study, "**Global automotive group races to automated driving with data platform**," April 2020.*

AI/ML environment considerations

Hybrid cloud platform and data pipelines

Hybrid cloud platform

A hybrid cloud platform provides a unified software foundation for developing, deploying, and managing tools, applications, and models across on-site, edge, and cloud environments.

Why is it important for AI/ML?

AI/ML models, software, and applications require infrastructure for development and deployment. A consistent hybrid cloud platform allows you to develop, test, deploy, and manage AI/ML models and applications in the same manner across all parts of your infrastructure, giving you more flexibility. It can also provide self-service capabilities to speed resource delivery while maintaining IT control. Finally, a consistent platform supplies a foundation for technology integrations from third-party vendors, open source communities, and any custom-developed tools you may use.

Best practices and recommendations

Select a security-focused platform that supports hardware acceleration, a broad ecosystem of AI/ML and application development tools, and integrated DevOps and operations management capabilities. Choosing an open source platform can provide more integration opportunities and flexibility.

HCA Healthcare℠

"HCA Healthcare and healthcare in general are at the beginning of a digital transformation journey. The platform that we're building lets us gather data from all of our sites in real time, run any algorithm on it, and apply results to bedside care."[8]

Dr. Edmund Jackson
Chief Data Scientist, HCA Healthcare

Data pipelines

Data pipelines provide methods for collecting, preparing, storing, and accessing datasets for AI/ML model development, training, and inferencing.

Why is it important for AI/ML?

Data is a key component in all AI/ML initiatives – it is required for training, testing, and operating models. However, 22% of Red Hat® customers cite getting access to relevant data as a top challenge to their AI/ML initiatives.[9] Data pipelines can connect discrete data sources, prepare data for use, and place it in an accessible repository for AI/ML engineers and application developers. They can also move used data from temporary storage to permanent storage archives.

Best practices and recommendations

Look for technologies that connect to your existing databases, data lakes, and other repositories. Standardized application programming interfaces (APIs) and high-bandwidth, low-latency networking will make it easier to access data throughout the AI/ML life cycle. Integration with open source data streaming, manipulation, and analytics tools like Apache Spark, Kafka, and Presto can help you manage your data more efficiently. You should also select technologies that provide data governance capabilities and integrated security features to protect your business.

BMW GROUP

BMW Group uses a Kubernetes-based data platform to access nearly

230 PB of usable storage

and simulate up to 240 million kilometers of test data.[10]

8 *Red Hat case study, "HCA Healthcare develops predictive analytics using Red Hat software," May 2019.*

9 *Red Hat, "2020 Red Hat Global Customer Tech Outlook," November 2019.*

10 *Red Hat case study, "Global automotive group races to automated driving with data platform," April 2020.*

Build an open, flexible foundation for AI/ML

Red Hat provides a complete technology portfolio, proven expertise, and strategic partnerships to help you achieve your AI/ML goals. We deliver a foundation for building production-ready AI/ML environments as well as services and training for rapid adoption.

Red Hat OpenShift® is an enterprise-ready Kubernetes platform for hybrid and multicloud environments. On-demand compute resources, support for hardware acceleration, and consistency across on-site, public clouds, and edge environments provide the speed and flexibility teams need to succeed. Self-service provisioning allows AI/ML teams to access resources without IT engagement. NVIDIA graphics processing unit (GPU) integration accelerates modeling and inferencing. Collaboration features let data scientists create and share containerized modeling results with peers and developers in a consistent manner. Built-in DevOps capabilities streamline development of intelligent, AI/ML-based applications.

Red Hat OpenShift Data Science is a cloud service that gives data scientists and developers a powerful AI/ML platform, including tools like Jupyter and associated TensorFlow and Pytorch frameworks, for building intelligent applications. Teams can quickly move from experiment to production in a collaborative, consistent environment with their choice of certified tools. Several technology partners, such as Starburst, Anaconda, IBM, Intel, and Seldon, are integrated into the offering.

"Working with great colleagues at Red Hat means we can use new tools like natural language processing and machine learning to develop new insights from that unstructured data that transform healthcare."

Dr. Jonathan Perlin
Chief Medical Officer, HCA Healthcare

Read the success story

The **Red Hat Application Services** portfolio helps you create a unified environment for application development, delivery, integration, and automation. Data integration services help you build effective data pipelines, while runtime services simplify application development. Process automation tools and services can access intelligent applications and ML/DL models to streamline and automate business processes and decisions.

Red Hat platform and storage products – including **Red Hat Enterprise Linux®**, **Red Hat OpenStack® Platform**, **Red Hat OpenShift Data Foundation**, and **Red Hat Ceph® Storage** – provide a scalable, software-defined infrastructure.

The **Red Hat certified partner ecosystem** allows you to integrate your choice of AI/ML and application development tools into this architecture for simple, automated deployment and life-cycle management.

Red Hat actively participates in the **Kubeflow** and **Open Data Hub** open source communities. Open Data Hub is a community project that provides a blueprint for integrating over 20 different open source AI/ML tools into an OpenShift environment. Common data analytic and machine learning tools – such as using Apache Spark, Ceph, Apache Kafka, Kubeflow, TensorFlow, Jupyter notebooks, and Hue – are included in the reference architecture.

ML/DL and DevOps tools	Red Hat Application Services
ML/DL data pipeline and sources	Red Hat Application Services
Hybrid cloud platform	Red Hat OpenShift
Hardware acceleration	intel. NVIDIA
Infrastructure endpoints	Red Hat Enterprise Linux; Red Hat Virtualization; Red Hat OpenStack Platform; Red Hat OpenShift Data Foundation; Red Hat Ceph Storage

Ready to get more from your data?

AI, ML, and DL are transforming nearly every aspect of business. Red Hat can help you build a production-ready AI/ML environment that speeds development and delivery of intelligent applications to support your business goals.

Learn how your business can benefit from effective AI/ML deployment: **cloud.redhat.com/ai-ml**

Explore Red Hat Marketplace

Simplify the trial, procurement, and deployment of Red Hat certified software from our ecosystem partners on Red Hat OpenShift. Visit Red Hat Marketplace to find Red Hat certified AI/ML software to help accelerate your AI/ML projects from pilot to production.

→ Find certified AI/ML software tools with Red Hat Marketplace

Get started faster with Red Hat Consulting

Work with Red Hat experts to jump-start your AI/ML projects. Red Hat offers consulting and training services to help your organization adopt AI/ML faster.

Learn about AI/ML services: **red.ht/ai-consulting-services**

Schedule a complimentary discovery session: **redhat.com/consulting**

redhat.com
#F28603_0621

Red Hat