Capabilities Statement

National Oceanic and Atmospheric Administration (NOAA) Earth Prediction Innovation Center (EPIC)

Solicitation Number: 1305M320RNRMN0002

July 3, 2020



| Prepared for: | General Position |
| --- | --- |
| Red Hat point of contact: | Ben Cushing, Emerging Technology  cush@redhat.com  (202) 640-1730 |

**Treatment of Red Hat’s response**

This response document contains Red Hat trade secrets, including but not limited to sensitive pricing information, and may not be disclosed to a third party without Red Hat’s prior written consent. The exchange of information under this response shall be protected under any applicable agreement between the parties related to the treatment of confidential information. In the absence of such an agreement, the recipient of this response shall protect the information contained herein in the same manner as it protects its own confidential information and may not be shared with any third party.

This is a proposal for a potential business relationship and is neither a contract nor an offer that could be accepted by the other party. If the parties decide to enter into a business relationship, they will memorialize the terms of that relationship in a separate agreement between them or their respective business partners.

Copyright © 2020 Red Hat, Inc. Red Hat, Red Hat Enterprise Linux, the Red Hat logo, and JBoss are trademarks or registered trademarks of Red Hat, Inc. or its subsidiaries in the United States and other countries. Linux® is the registered trademark of Linus Torvalds in the U.S. and other countries.

Contents

[**Executive Summary**](#_1j1d5u5arxi) **3**

[**Overview**](#_octmws13cehb) **4**

[**Functional Areas**](#_ds1307p0n2ok) **5**

[Software performance engineering](#_4unvvgxkmhf) 5

[Cloud based high performance computing base.](#_yam6y7xqmw9v) 6

[Red Hat Enterprise Linux (RHEL) for HPC](#_ca5x5lae04oz) 6

[Why Red Hat Enterprise Linux Server for HPC?](#_j2foddtnudvd) 7

[RHEL delivers tools for container performance](#_ry0cg3d1262i) 8

[RHEL security features](#_9s7gawrn8v9p) 9

[Data Staging/Data at Rest:](#_7ecqpcmtxzdy) 10

[Scientific innovation open data hub.](#_xs5fpptzrp2r) 12

[Data-intensive intelligent applications in a hybrid cloud blueprint](#_gp1li63zinnp) 13

[**Conclusion**](#_m4dtyqrz15v3) **19**

# Executive Summary

The Department of Commerce (DOC), National Oceanic and Atmospheric Administration (NOAA), Office of Oceanic and Atmospheric Research (OAR) Office of Weather and Air Quality (OWAQ) requires a technology partner to help bring to life the Earth Prediction Innovation Center (EPIC). EPIC is slated to become “the world’s most accurate and reliable operational weather forecast model.” [[1]](#footnote-0)

The partner sought will have proven expertise in software engineering, software infrastructure development and experience in delivering support services to government, academic and industry research scientists. The goal is to develop advanced weather modeling capabilities to improve forecast accuracy, protect life and property and enable NOAA to maintain its international leadership role in weather monitoring. Additionally, capabilities for innovation must be provided.

Red Hat has provided multiple science-based organizations with high performance computing (HPC) solutions. One such organization is The Institute of Molecular Biology (IMBA). The IMBA is a leading institute for molecular biology research. With research data volumes growing exponentially, they needed to rapidly deploy high-performance computing environments used by scientists for research. Using Red Hat technology, the institute created a flexible, scalable, software-defined datacenter that has cut installing custom environments from months to hours, helping them to focus on supporting innovative research. [Learn more in this case study.](https://www.redhat.com/cms/managed-files/rh-imba-case-study-f19116wg-201910-en.pdf)

NOAA has reached a point where the data associated with extreme climate variability needs to be accommodated, to produce accurate weather predictions. For over 50 years, NOAA has been using satellites and other sensors to collect environmental data but a big part of prediction accuracy capability is in simulations (artificial intelligence / machine learning [AI/ML]) that will be used for forecasting, in an HPC environment.

Advancing the science behind weather forecasting will be achieved through greater access to data, accelerated ML cycles (fail early, fail often) and integrating more of the weather research community.

Red Hat understands that all phases of AI/ML are demanding on infrastructure. With an interested and qualified integrator / partner Red Hat can provide the modern technology to support the EPIC backbone, to include: 1) software performance engineering, 2) cloud-based high-performing foundation, 3) data staging, 4) scientific innovation, and 5) a forward-looking hybrid cloud strategy that is also applicable to edge computing. Red Hat can take NOAA from its current state to a position that prepares them for a cloud transition. The benefits that a Red Hat EPIC backbone provides to NOAA include:

* DevSecOps lifecycle management, which reduces risk, cost and increased productivity. A DevOps framework that enables NOAA to accelerate workflow, enrich analysis and as a result, enhance the accuracy of predictions.
* Innovation via Data Transparency. The removal of data barriers and silos will enable data scientists and ML engineers to have access to the required data and become more innovative and efficient in their approach.
* A hybrid cloud solution that will enable NOAA to become more agile in integrating data resources and managing multiple environments. In addition, hybrid cloud enables an edge capability with a consistent application and operations experience. Different workload types can have different footprints. The Red Hat hybrid cloud helps to extend work out to the edge devices, so compute power is used most efficiently.

# Overview

Every aspect of scientific research is centered around the collection, distribution, production, analysis, and summary of data. This section of Red Hat’s capability statement covers the stages of a data intensive intelligent application life cycle. From start to finish, every aspect of this life cycle has specific requirements regarding skills, tools, and infrastructure. All the stages need to be connected in today’s dispersed deployment environments.

Red Hat recognizes that within NOAA’s EPIC scientific research, every global environmental event generates scientific data. This capability statement is centered around the data, and the stages of a data-intensive intelligent application life cycle. The stages need to be connected in today’s dispersed deployment environments, potentially spanning multiple cloud providers, in-house data centers, and edge devices. This detail provides a Red Hat view of the methodology and technology required to build, maintain, and manage these complex hybrid cloud environments.

The following pages describe how the NOAA EPIC - Red Hat solution addresses five functional areas:

1) Software performance engineering - satisfied by a DevOps approach in developing jobs and production and moving to HPC, as well as software engineering lifecycle management, supported by Red Hat OpenShift Container Platform .

2) Cloud-based high-performing foundation - satisfied by RHEL and HPC.

3) Data staging (data at rest) - satisfied by Red Hat OpenShift Container Storage / Red Hat Ceph

4) Scientific innovation - satisfied by Open Data Hub-built on Red Hat OpenShift Container Platform, Ceph Object Storage, and Red Hat Integration, which includes Apache Kafka/Strimzi and a collection of open source projects to enable a data-science-as-a-service platform

5) Hybrid cloud strategy - satisfied by Red Hat AMQ w/AMQ Streams, Open Data Hub, Apache Spark, OpenShift Container Platform, Red Hat Process Automation Manager, Red Hat Decision Manager, Red Hat Runtimes, and Ansible automation.

# Functional Areas

## Software performance engineering

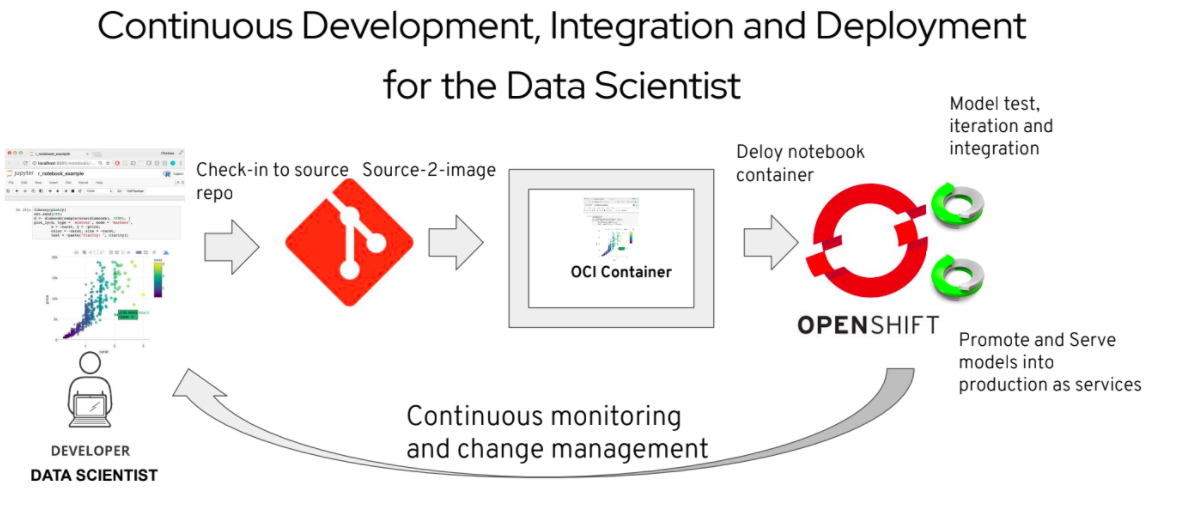
**Supporting a DevOps and continuous deployment approach**

End-to-end data-intensive intelligent programs, such as NOAA EPIC, are composed of a large set of disparate infrastructure and hardware components. These components all need to work together to provide the full value of the solution and deliver intelligent applications that can offer more context-aware and personalized services to NOAA EPIC scientists and ML engineers, as well as to other systems. Additionally, the success of such a solution requires significant and varied skills and expertise. From infrastructure administration and operation to data analytics and science to application development, all these skills are necessary to build, deploy, and operate a highly complex and distributed intelligent application environment.

As stated previously, NOAA EPIC seeks to acquire accelerated ML cycles that will enable ML engineers to focus on new and innovative weather predicting solutions. To that end, Red Hat recommends a DevOps approach. Working within a DevOps environment speeds up the processes by which an idea goes from development to deployment, and enables the efficient and effective operation and evolution of solutions. To support a DevOps approach, NOAA EPIC requires a reliable, automated build, test, and deployment pipeline for all types of workloads of the solution to enable the continuous delivery of new functionality to the production systems, and thus to the data scientists.

NOAA EPIC’s dynamic hybrid cloud environment demands an equally dynamic and scalable build system, one that can be provisioned to ensure reliable and repeatable builds of all the solution’s software components. Red Hat recommends Red Hat OpenShift Container Platform (Red Hat OpenShift) for automated tests and testing toolkits to assess the quality of the built software components individually, as well as the integration of these components with other parts of the system. Red Hat OpenShift includes CI/CD toolkits that support fully automated testing and integration testing environments to check the quality of the individual components and solution as a whole, providing the confidence to rapidly deploy new functionality into production, thereby allowing NOAA developers to fail early and fail often. Additionally, Red Hat OpenShift includes streamlined workflows to help teams get to production faster, including built-in Jenkins pipelines and our source-to-image technology to go straight from application code to container. It is also extensible to new frameworks like Istio (service-mesh) and Knative (serverless).

The combination of a DevOps process with a cloud-native CI/CD toolkit delivers the required set of capabilities to build, test, deploy, provision, monitor, manage, and operate the complex and distributed architecture of data-intensive intelligent applications at scale. Red Hat OpenShift is an example of a fully cloud-native, open source, CI/CD framework (**Figure 1**) that allows organizations to build, test, and deploy across multiple clouds and edge compute nodes in a hybrid cloud environment. Using DevOps, continuous integration/continuous delivery (CI/CD), container, and cloud methodologies enables the rapid creation, provisioning, and deployment of applications in highly distributed environments.



*Figure 1. CI/CD integration and deployment model for Data Science. DevOps, and processes like continuous integration and continuous delivery (CI/CD), software can be shipped in seconds, because the steps involved are automated.*

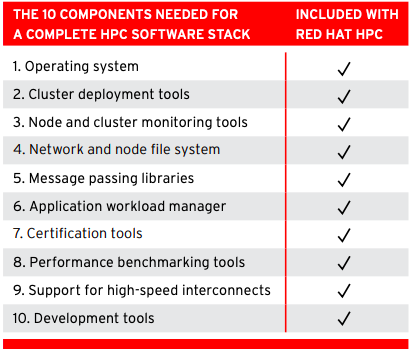
## **Cloud based high performance computing base**

### Red Hat Enterprise Linux (RHEL) for HPC

Analysts, engineers, and scientists are constantly pushing the computational limits of current IT resources. In response, businesses are increasingly adopting and utilizing HPC clusters to gain a competitive edge. As a result, HPC is the fastest growing segment in the server industry today.

For the past two decades, RHEL has served as the standard for building software stacks for supercomputers. Many of our customers, like [Ford Motor Company](https://www.redhat.com/en/resources/ford-motor-company-case-study) and [The Institute of Molecular Biotechnology](https://www.redhat.com/en/resources/imba-case-study) (IMBA) had the need for HPC clusters to deploy clusters of systems to tackle their most challenging missions. The performance of RHEL cannot be surpassed in terms of power, scalability and performance and the HPC add-on only makes it even better, by facilitating workloads that are distributed to many compute nodes.

The RHEL Server for HPC is a fully integrated software stack that enables creation, management, and usage of a high-performance computing cluster, running RHEL. (Figure 2) It removes many of the barriers and complexities associated with integrating an HPC solution, and opens HPC to scientists and workloads.



*Figure 2. This complete solution provides everything needed to get started with HPC: the operating system, device drivers, simple cluster installer cluster management tools, resource and application monitor, interconnect support, and a powerful job scheduler—Platform Lava*

The Red Hat Enterprise Linux for Scientific Computing platform is deployed as an interactive RHEL for high-performance computing (HPC) Head Node and a homogeneous cluster of RHEL for HPC Compute Nodes. RHEL for HPC Head Node is a full-featured platform to be deployed as the interactive system in an HPC cluster. Each RHEL for HPC Compute Node subscription inherits the support level of the RHEL for HPC Head Node that is deployed in the cluster. \*Note that HPC is available only for AMD64 and Intel 64 architectures.

### Why RHEL Server for HPC?

Red Hat offers packages specifically for the scientific HPC engineers and scientists to deploy clusters of systems to tackle the most challenging workloads. Subscriptions for RHEL in clusters are affordable, easy to deploy, and support hardware scale out.

With RHEL for HPC, the NOAA scientists and developers will benefit from:

* Low acquisition and support costs. RHEL for High-Performance Computing is cost-effective for grid-like deployments.
* Streamlined, lightweight deployment. RHEL for High-Performance Computing offers a slim component set that can be customized for each environment to reduce the deployment footprint, simplify management tasks, and minimize points of failure and security vulnerabilities.
* Multi-core affinity and enablement. With its precise ability to detect and optimize hardware configurations, RHEL quickly and easily gets the most from NOAA EPIC team hardware. Red Hat’s deep relationships across the industry translate into a platform optimized for the latest capabilities.
* Validation of proper cluster operation. RHEL provides advanced monitoring capabilities to enhance each cluster’s manageability, provide timely performance information for tuning, and minimize the need for manual system checks.
* Improved I/O performance. Faster I/O performance enables NOAA EPIC to get data into and out of your HPC cluster more quickly, maximizing the cluster’s processing performance.

### **RHEL delivers tools for container performance**

Organizations are deploying containerized applications and services to increase agility and consistency across their IT environments. Containerized applications and services can be written once, then deployed, moved, and scaled across infrastructure as needed to meet changing demand. RHEL delivers advanced, innovative container infrastructure and tools to simplify container development and deployment. It supplies a lightweight, open standards-based container toolkit with everything you need to get started.

Delivered with RHEL, Podman and the accompanying container development tools, Buildah and Skopeo, make it relevant to many HPC environments that have standardized and rely on this operating system (OS).

### **Open Containers Initiative- (OCI) compliant tools**

Based on standards from the Open Container Initiative (OCI) Podman's implementation is rootless (does not require superuser privileges) and daemon-less (does not need constantly running background processes), and focuses on delivering performance and security benefits.

Another important aspect is that Podman shares many of the same underlying components with other container engines, like CRI-O, providing a proving ground for new and interesting features, and maintaining direct technology linkage to Kubernetes and Red Hat OpenShift. The benefits of technology continuity, the ability to contribute and tinker code at the lowest layers of the stack, and the presence of a thriving community, were the fundamental reasons for Red Hat’s investment in Podman, Buildah and Skopeo.

To further foster collaboration in the community and enable participants to freely redistribute their applications and containers that encapsulate them, Red Hat introduced the Red Hat Universal Base Image (UBI). UBI is an OS container image that does not run directly on bare metal hardware and is not supported as a stand alone entity, however it offers the same proven quality and reliability characteristics as Red Hat Enterprise Linux since it is tested by the same quality, security and performance teams.

UBI offers a different end user license agreement (EULA) that allows users to freely redistribute containerized applications built with it. Moreover, when a container built with UBI image is running on top of Red Hat platforms, like RHEL with Podman or OpenShift, it can inherit support terms from the host system that it runs on. For many sites that are required to run supported software this seamlessly creates a trusted software stack that is based on a verified OS container image.

RHEL provides several [Open Containers Initiative- (OCI) compliant tools](https://www.opencontainers.org/) to simplify and improve container development, management, and security:

**RHEL containers** are ideal for deploying microservices-based, cloud-native applications. RHEL contains and fully supports Red Hat’s lightweight, open standards-based container toolkit. It also provides several features to simplify container development, management, and security.

[**Buildah**](https://www.redhat.com/en/blog/daemon-haunted-container-world-no-longer-introducing-buildah-10)**-** enables users to build and modify containers without any daemon or docker. It preserves existing dockerfile workflow while allowing detailed control over image layers, content, and commits. Buildah also minimizes container image size by using tools from the container host rather than adding them to the container image.

[**Skopeo**](https://blog.openshift.com/promoting-container-images-between-registries-with-skopeo/) - is a comprehensive tool and library for inspecting, signing, and transferring container images to inspect, verify, and sign image manifests, and move container images between registries. Skopeo takes advantage of the same code base library used by Buildah, Podman, and [CRI-O](https://www.redhat.com/en/blog/red-hat-contributes-cri-o-cloud-native-computing-foundation), a lightweight container engine for Kubernetes.

[**Podman**](https://developers.redhat.com/blog/2018/11/20/buildah-podman-containers-without-daemons/) **-** is a complete, daemon-less container engine for running, managing, and debugging OCI-compliant containers and pods. It lets you manage containers without the daemon dependency and is [docker command-line interface (CLI) compatible](https://developers.redhat.com/blog/2019/02/21/podman-and-buildah-for-docker-users/). It also provides improved integration with systemd. With Podman, you can easily find, build, run, and share containers. Switching from Docker to Podman is very easy and intuitive — two commands are all you need.

**Red Hat Universal Base Images -** containers built from RHEL content and running in Red Hat environments deliver production-grade support, stability, and security features. To extend these benefits to all developers, RHEL introduces OCI compliant Red Hat Universal Base Images (UBIs) for containers. Red Hat UBIs let you take advantage of the reliability, security, and performance of official Red Hat container images, even if you are not a Red Hat customer. Portable application images let you develop once and deploy on any Linux distribution, simplifying container development. You can build a containerized application on a UBI, push it to your choice of registry server, and easily share it with others. Developers can distribute a smaller container image anywhere, while operators can deploy a supportable base image with an enterprise life cycle.

**Application Streams** - is an improved method for delivering multiple versions of userspace packages. Application Streams provides new packages at a cadence that makes sense for each package, rather than combining them into a monolithic distribution. They also offer multiple versions of select packages, giving you more choice. Finally, Application Streams combines all previous distribution channels into a single location. As a result, you can more easily access the latest stable versions of the programming languages, tools, and databases.

### **RHEL security features**

RHEL is a common criteria certified, FIPS 140-2 compliant Linux for server operating systems, VMs and container images, with SELinux security controls and certified DISA STIGs available. It is the underlying operating system upon which all of our products are built. Red Hat has worked closely with various U.S. government agencies, including the U.S. National Security Agency (NSA) on needed certifications to operate on government networks, and is currently in 100% of U.S. government branches. Red Hat OpenShift Container Platform is built on these trusted technologies and extends the security posture of your applications, by enabling tightly integrated automation of compliance, security, and remediation. OpenShift has several PKI components in the platform to support the overall capabilities. All Red Hat government standards and certifications can be found at access.redhat.com/articles/2918071.

## Data Staging/Data at Rest

Historically, manual processes and data that could not be separated from applications resulted in slow implementations. Formatting, filtering, and manipulation of the data, according to the application, was required. Huge amounts of time was spent performing these preparatory tasks. Plus, varied teams of developers required ownership of data, which generated barriers in the form of data ‘silos’, and political data ownership issues compounded the problem. Another data challenge has been, and continues to be in some cases, vendor lock-in. This is where data is locked into a particular vendor’s software, specifically in a cloud platform.

Given the exponential growth in NOAA EPIC data from an increasing number of data sources (satellite imagery, land and ocean based sensors, etc), the challenge of seamlessly storing, integrating, and managing data across on-prem,public and private clouds also expands. Plus, NOAA needs to understand how they will scale the agency’s data connections to account for ongoing growth. Further, for purposes of innovation, easy access to data is one of the most important factors (in addition to innovative technology and a skilled workforce). NOAA requires a scalable solution that will push the boundaries of innovation, predict more accurately and enable EPIC to do more with the data they collect.

According to AI & Machine Learning, Voice of the Enterprises, 451 Research, 2019, companies say that all phases of AI/ML are depending on infrastructure. 39% of the respondents say that data preparation and management was the most infrastructure-intensive phase, followed by AI/ML training (34%) and inference (27%).

Within Red Hat OpenShift, applications are produced *independent of* data sets and utilize a shared data solution - Ceph Storage - to promote collaboration in a HybridDevOps environment. Red Hat OpenShift Container Storage includes the NooBaa operator for multi-cloud environments. The NooBaa operator integrated with OpenShift Container Storage is a multi-cloud object storage gateway to any cloud or on-prem S3 bucket. This technology allows replication of telemetry and imaging data across multiple environments to ensure redundancy in availability of the data, while the hosted application is presented with the same standards-based Object Storage API, regardless where the data may reside. OpenShift Container Storage itself offers object based storage to tenant applications as well, meaning applications destined for the NOAA EPIC OpenShift Container Platform environment can take advantage of any object storage during development, and be assured the same storage interface is available once deployed into NOAA EPIC .

Red Hat science-based customers that have successfully integrated a Ceph solution include: [Massachusetts Open Cloud](https://www.redhat.com/en/success-stories/massachusetts-open-cloud), [Cloud Infrastructure for Microbial Bioinformatics (CLIMB)](https://www.redhat.com/en/resources/climb-case-study), and [University of Alabama at Birmingham (UAB)](https://www.redhat.com/en/resources/university-of-alabama-birmingham-case-study). All these customers struggled with data-intensive work, such as engineering and molecular modeling, due to limited storage availability and opportunities for collaborative work. Another great example is the European Organization for Nuclear Research, known as CERN. CERN has been a long-time Ceph user and active community member, running one of the largest production OpenStack clouds - backed by Ceph. They are also [using Ceph for other storage use-cases](https://ceph.io/community/new-luminous-scalability/), backing a range of high energy physics experiments.

**Additional Ceph features**

Ceph Storage Cluster: The Ceph Storage Cluster supports Linux Unified Key Setup or LUKS encryption of OSDs and their corresponding journals, write-ahead logs, and metadata databases. In this scenario, Ceph will encrypt all data at rest irrespective of whether the client is a Ceph Block Device, Ceph Filesystem, Ceph Object Storage cluster or a custom application built on librados.

Administrative API: Ceph provides an administrative API that can be used to perform system management operations like user management, audit logging, and other related operations. All system and audit logs are captured on the Ceph system and available for review. Ceph objects can be tagged with unlimited metadata for use by other third-party metadata management tools such as Informatica.

## **Scientific innovation with an** O**pen** D**ata** H**ub**

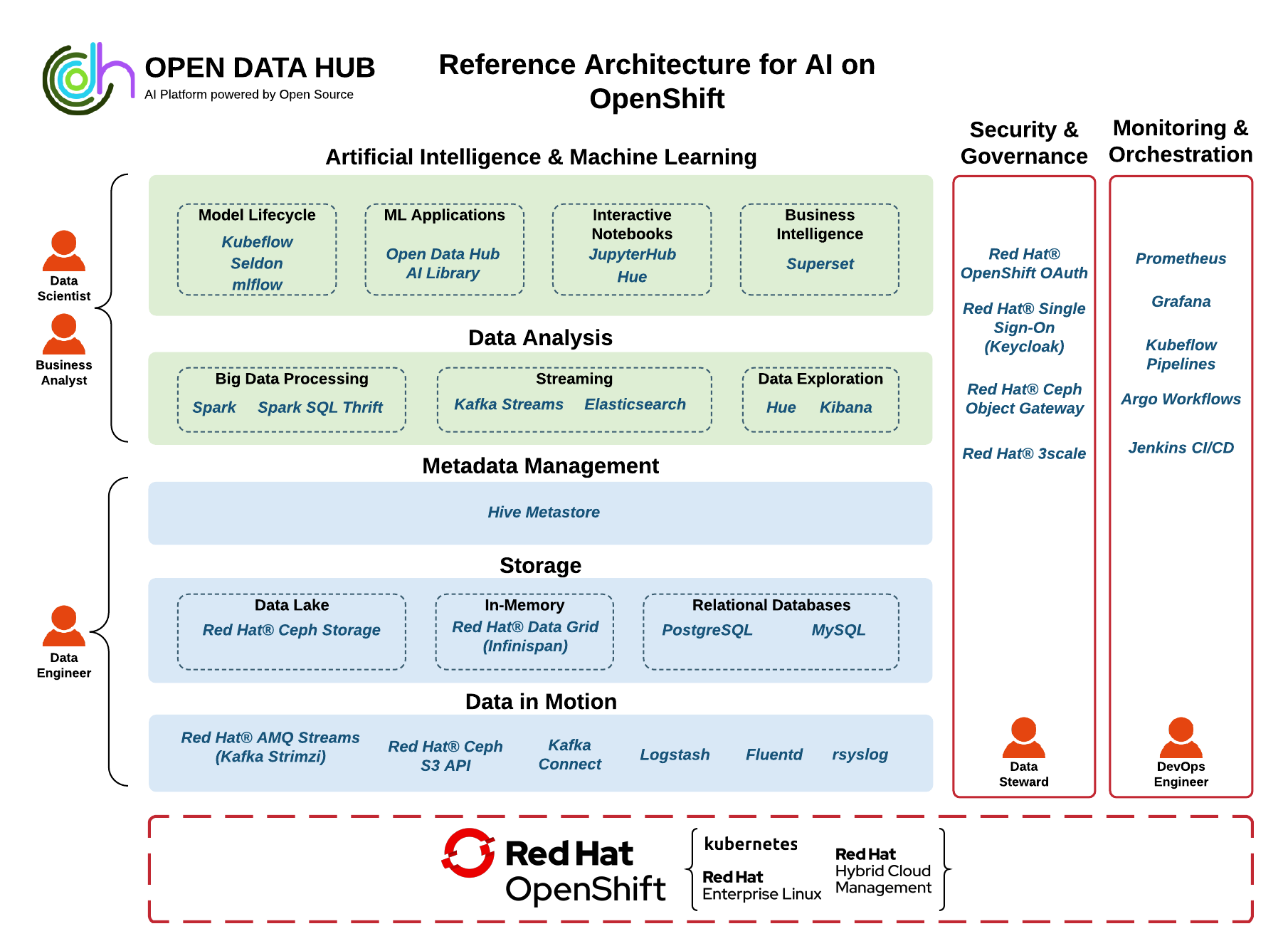
As a complement to using a HPC solution, NOAA EPIC can implement data science-as-a-service via Open Data Hub. Data science-as-a-Service treats data as a product with its own life cycle. The prospect of true machine learning is a tangible goal for NOAA EPIC data scientists, ML engineers and researchers. It has been long known that the ML app platforms have to be fast and hyper-efficient, so that learning can be that much faster. This is the motivation for Red Hat engineers in the Office of the CTO who are applying these principles to the Open Data Hub platform.

Several Red Hat customers (e.g. [HCA Healthcare](https://www.redhat.com/en/success-stories/hca-healthcare), [ExxonMobil,](https://www.redhat.com/en/success-stories/exxonmobil) [BMW Group](https://www.redhat.com/en/success-stories/bmwgroup), etc.) have deployed containerized ML tool chain and DevOps processes on Red Hat OpenShift and our ecosystem partner products, to operationalize their preferred ML architecture, and accelerate workflows for data scientists.

[**Open Data Hub**](https://opendatahub.io/) includes several open source components which can be individually enabled, including: Apache Airflow, Apache Kafka, Apache Spark, Apache Superset, Argo, Grafana, JupyterHub, Prometheus, and Seldon. Open Data Hub combines the power of AI/ML with the strengths of Red Hat OpenShift Container Platform and the underlying technologies, the agility of the DevSecOps model, and the security inherent in the appropriate utilization of trusted Red Hat platforms.

Open Data Hub is built on Red Hat OpenShift Container Platform, Ceph Object Storage, and Apache Kafka/Strimzi integrated into a collection of open source projects to enable a data science-as-a-service platform. To ensure that contributions to Open Data Hub perform well, Red Hat engineers have taken the step of creating an Internal Data Hub within Red Hat as a testing and proving ground for the learning environment.

The reference architecture for AI on OpenShift (**Figure 3**) illustrates how an organization can use the tooling best suited to their needs, in a cloud-native way, on Red Hat OpenShift. It also demonstrates how workflows can best benefit from cloud-native development that provides a familiar experience to data scientists and engineers when developing machine learning models while maximizing the ability to bring those trained models into production securely.



*Figure 3. Open Data Hub reference architecture for AI running on OpenShift*

## Data-intensive intelligent applications in a hybrid cloud blueprint

Red Hat understands that Scientific Data tends to grow in data analytic and ML environments, and as they get larger it becomes more and more difficult to manage them. The term “data gravity” refers to the desire to have applications and data attract more applications and data on a network. What drives data gravity is the need for low latency and the number of packets that are processed during a specified time period (throughput). Data gravity can become a challenge to NOAA EPIC because it can have expensive consequences. That said, as an alternative to a traditional HPC solution, Red Hat recommends that NOAA EPIC first implement an on-prem storage strategy with a cloud based compute capacity. This Hybrid approach will let NOAA leverage their current capital investment, while taking advantage of the expansive compute capacity available from Cloud Hosts. This approach also reduces the financial burden of cloud data hosting, while novel storage solutions as described in **Section 3** are being explored and proven by NOAA engineers.

By following a storage on-prem first implementation and transitioning into to the cloud, NOAA should expect the following benefits:

**Connectivity** -Improved quality of applications and maximize the value of digital assets by offering multiple options for seamless connectivity between applications and systems - both current and legacy. Red Hat provides a wide range of options to interact with legacy systems and custom-developed in-house applications - including synchronous and event-driven types of exchange - so you can gain access to valuable data.

**DevOps** -Enables microservices developers and DevOps teams to build and deploy applications with access to all the necessary digital assets, without requiring changes to the development schedule, tools, or practices. Red Hat solutions supports deployment to Linux containers instead of a standalone engine that operates outside of cloud native platforms.

**Cloud Ready** - Provides scalability and high availability by using Kubernetes to create, extend and deploy containerized integration services across hybrid cloud environments. Red Hat solutions can be deployed on-prem, in public/private clouds, or as a hosted service.

**Productivity** - Boosts productivity by streamlining the development of integration and messaging capabilities across the enterprise and beyond. Red Hat solutions offers libraries of integration and messaging design patterns and components that have been proven in hundreds of complex projects.

**Time to Deployment** - Get new scientific innovation going faster with current hardware. Red Hat solutions provide seamless integration with automated software development lifecycle tools like authoring, version control, testing, continuous integration, deployment, monitoring, and retirement.

**Life cycle of data-intensive intelligent applications**

While single data points can seem insignificant, the combination of data, or events, from a single source or a collection of event sources can be used to infer higher-value information. This higher-value information helps NOAA better assess high priority / critical atmospheric concerns tracking, and trends, to deliver guidance for global policy. By treating each incoming data point as an event, NOAA can apply decision management and machine learning (ML) inference techniques to filter, process, qualify, and combine events to deduce higher-order information. The availability to identify and extract higher-order information allows for the development of intelligent applications that can offer more context-based services to applications, and scientific systems. These types of applications are referred to as data-intensive intelligent applications.

The life cycle of these application includes various stages:

* Data ingestion: Intake, pre-processing, and transportation
* Data engineering: Storage and transformation
* Data analytics: Data analysis and model training
* Runtime inference: Model serving and monitoring
* Business events and insight management: Event management, insights, and process and integration management.

**Challenges and Characteristics of data-intensive intelligent applications**

Data-intensive intelligent applications activities are performed in distributed parts of the network architecture. Data ingestion and runtime stages are performed at the edge, close to the generation of environmental events, and require edge computing capabilities. Engineering and analytics are performed within a public or private cloud that facilitates data streaming, data lakes, and distributed artificial intelligence (AI) workloads. Further, these activities are performed by various users, each with individual skills, tooling, software platforms, and infrastructure requirements. Business events and insight management address the actions that are taken based on higher-level information, this includes recording these events in various systems, using integration technologies, automated business processes, and human tasks to proactively act on higher-value events.

An infrastructure and application development platform that is flexible, adaptable, and elastic is required to fulfill these different needs and provide the connection between these various stages. The hybrid cloud approach, an IT architectural style that provides a homogenous cloud experience across disparate public and private clouds as well as cloud services, enables portability, orchestration, and management of workloads across environments. In the context of data-intensive intelligent applications, this not only means that we can provision compute resources and storage on demand, but it also implies that we can target the optimal cloud environment for a workload.

Given the various lifecycle stages of these types of applications, the hybrid cloud provides the flexibility to optimally provision the data capture and intelligent inference workloads at the edge of an environment, the resource-intensive data processing and training workloads across cloud environments, and the business events and insight management systems close to business users.

**Intelligent distributed events processing**

An event source can be any device or system that generates events. These can be devices and systems at the edge, close to the physical location of either the user or source of data. Or they can be internal systems and services, like enterprise resource planning (ERP) systems, databases, and enterprise applications, that, for example, generate events during online transaction processing.

Edge computing is an important part of the hybrid cloud vision that offers a consistent application and operation experience, and Red Hat offers a full complement of products that enable it. **Red Hat OpenShift Container Platform** provides the foundation to automate and manage infrastructure from the core datacenter to remote edges. It enables a standardized operational model for hybrid workloads that can scale in and out while running a consistent deployment model across both small and large deployments.

**Red Hat Runtimes** provides the runtimes and frameworks for building highly distributed, cloud-native applications. It lets developers deploy their intelligent applications on a vast array of devices, close to the user, providing a highly available, low-latency application experience.

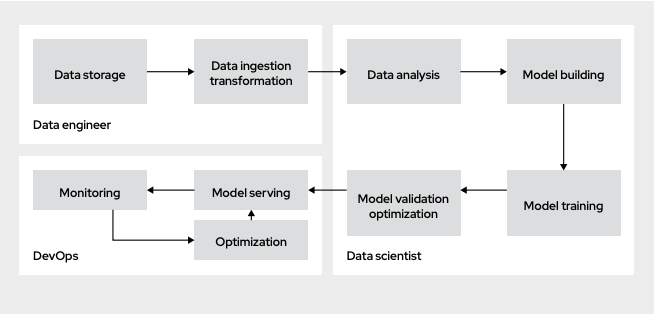
**Red Hat Decision Manager**, has a high-performing rules engine that can be directly embedded into a Red Hat Runtimes-based application, which provides the functionality to declaratively define filtering logic. This capability allows for the separation of application and decision logic, making filtering rules easier to manage and maintain. Changes in data and event filtering requirements can be rapidly implemented and provisioned to the intelligent application without impacting application code and logic.

As devices and systems generate a constant stream of events, they need to be capable of handling and transporting high volumes of data. Red Hat Integration provides the high-performing and elastic messaging solutions required to handle the transportation of high-volume data streams in a reliable and scalable way. **Red Hat AMQ**, which is included in Red Hat Integration and contains AMQ Streams, is a messaging solution that enables cross-cloud connectivity based on open standards and protocols. Red Hat Integration also provides the capabilities to integrate disparate systems in an agile and cloud-native way via its many connectors. In the context of data-intensive intelligent applications, this feature provides the ability to ingest, transform, and transport events from different and distributed event sources to the required destination, such as a data lake.

The serving and inference of AI machine-learned models, which provide the application’s intelligence directly at the edge, is performed by **Red Hat Decision Manager**. The support for the Predictive Model Markup Language (PMML) specification, in combination with Decision Model and Notation (DMN), provides the inference of predictive, machine-learned models from within DMN decision models. This means intelligent applications can provide a better explainability of decisions by allowing opaque models, like neural networks, and their predictions, to be integrated with declaratively defined decisions and rules.

**Artificial intelligence and machine learning workloads in a hybrid cloud**

It is important to identify and visualize an end-to-end AI platform in a robust, hybrid application environment that is equipped to run large AI and ML distributed workloads. The end-to-end application environment should include all phases of applications driven by AI and ML. The flow of data and events through a solution (**Figure 4**) is essential to successful data-intensive intelligent applications. Data is ingested at multiple disparate locations from which it needs to be transported and stored.



*Figure 4. AI workflows and personas for data-intensive intelligent application model. Note that data is ingested at disparate locations and by disparate consumers across the workflows.*

Red Hat AMQ and AMQ Streams provide robust and scalable data transfer capabilities native to Red Hat OpenShift Container Platform that scientists and data engineers can use to transfer required data from multiple sources. Once transported, data needs to be stored and made available for the workloads that operate on this data to deduce information and train predictive models. A data lake based on Red Hat Ceph provides scalable, petabyte-scale storage to support AI workloads. Red Hat OpenShift platform provides multiprotocol support, including block, file, and S3 object application programming interface (API) support. For ultrafast data access, a core requirement for data analysis and model training, Red Hat Data Grid delivers in-memory access to distributed data.

OpenShift operators natively integrate the storage components into the hybrid cloud infrastructure. To support big data processing, the Open Data Hub initiative prescribes the tools needed to run large distributed AI workloads. An Apache Spark operator provisions and manages distributed AI workloads natively on Red Hat OpenShift, delivering an on-demand, cloud-native experience. These Spark clusters are provisioned dynamically and delivered specifically for each user, providing isolation of resource and usage management. An important component of data-intensive intelligent applications is the data science workflow in which data is analyzed and models are built, trained, and validated. To support this workload, organizations can provision interactive Jupyter Notebooks to provide a development workspace for data scientists and business analysts in a cloud environment.

ML environments are built on a complex set of technologies that need to interact with each other. On top of that, these workloads are highly demanding on resources like CPU, memory, and network bandwidth. Setting up such a system in a traditional environment is complex and hard to manage. A cloud environment provides the ability to dynamically provision and intelligently manage compute resources. With the OpenShift operator framework, the provisioning of an ML environment and the integration of its various components can be fully automated and provided on demand. Using an Infrastructure-as-Code (IaC) approach, this automation can be properly structured, versioned, and governed.

**Business events and insights management**

The goal of data-intensive intelligent applications is to help organizations provide a better user experience to end customers, make better business decisions, and pre-empt potential problems. In addition to ingesting and analyzing events and inferring new information, the system needs to provide insight into this information and take appropriate pre-emptive actions. The ability to proactively start automated business processes, automatically inform business users ahead of time, and provide 360-degree views of customer journeys is key in the quest to deliver a personalized experience to end customers.

**Red Hat Process Automation Manager**  delivers an advanced process automation, case management, and decision management platform to build and execute automated business processes and case definitions. Fully supported on Red Hat OpenShift, the platform enables business and IT professionals to design intelligent, cloud-native processes that a business can act on rapidly, consistently, and reliably in a repeatable fashion. The ability to stay ahead of change in an automated way and preemptively inform and engage business users to better serve customers is a competitive advantage.

**Ansible Network Automation** Ansible is a simple, powerful IT automation platform that helps you stream-line and manage complex data center environments, from servers and net-works to applications and DevOps. Ansible provides support for legacy and open network infrastructure devices across multi vendor virtual and physical environments so you can automate your entire network using a single tool. Using a common language, Ansible makes everyday tasks repeatable and scalable so you can run your network more efficiently. Choose to automate where you need it most. Ansible platform supports automation for provisioning systems, deploying applications, orchestrating complex workflows, and managing the configuration of your systems, networks, and applications—all of which helps improve your IT organization’s efficiency. Ansible’s flexible framework embraces incremental change, so you can start small and expand over time. With Ansible, you can manage your network infrastructure throughout the entire production life cycle.

# Conclusion

In review of the various stages of the life cycle of data-intensive intelligent applications, from data ingestion to transportation, filtering, analysis, modeling, deployment, and inference, every aspect of this life cycle has specific requirements regarding skills, tools, and infrastructure. Integration between applications and systems is essential to communicate, and exchange data. Meanwhile, cloud-native applications with microservices architectures are changing the development environment, and the approach for the life cycle of Machine Learning is no different.

The massive volume of climate variability data that is required for NOAA’s EPIC program requires a scalable, robust, high-performance computing solution, for simulation, forecasting and weather prediction. The solution must surpass the accuracy and sophistication of our international weather forecasting peers, so the United States will become qualified collaborators with those supporting other forecasting models.

Red Hat provides the technology and methodology to help NOAA build, maintain, and manage a successful, complex hybrid cloud environment. This is evidenced through our many HPC implementations, some of which have been referenced within. Satisfaction of the five functions that comprise the recommended EPIC backbone: 1) software performance engineering, 2) cloud-based high-performing foundation, 3) data staging, 4) scientific innovation, and 5) a forward-looking hybrid cloud strategy that is also applicable to edge computing, meets the requirements of a robust platform.

NOAA benefits by a Red Hat HPC solution in the following ways:

* DevSecOps lifecycle management, which reduces risk, cost and increased productivity.
* Innovation via data transparency and removal of barriers and silos.
* A hybrid cloud solution that will enable NOAA to become more agile in integrating data resources and managing an environment with multiple environments.
* Lastly, a hybrid cloud enabling an edge capability with a consistent application and operations experience.

With an eye toward the future, [Red Hat OpenShift](https://www.redhat.com/en/resources/openshift-container-platform-datasheet?extIdCarryOver=true&intcmp=701f2000000uHrOAAU&sc_cid=701f2000001Css5AAC) and [Red Hat Middleware](https://www.redhat.com/en/products/middleware) provide customizable components and capabilities to help your organization achieve its goals in an agile, cloud-native way, within a hybrid cloud.

1. [NOAA seeks partnership to help develop world’s best weather model.](https://www.noaa.gov/media-release/noaa-seeks-partnership-to-help-develop-world-s-best-weather-model) [↑](#footnote-ref-0)