Data-intensive intelligent applications in a hybrid cloud blueprint

Introduction

Almost every action we perform generates data. From simple daily tasks like buying groceries or booking a trip, to complex operations like operating manufacturing plants and aircraft, organizations across the globe produce, capture, distribute, and store data. This detail 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, potentially spanning multiple cloud providers, in-house datacenters, 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 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. (For more on this subject, see *Event-driven architecture for a hybrid cloud blueprint*). This kind of information can help organizations make better business decisions, deliver better customer experiences, pre-empt problems, and remain competitive (e.g., fraud prevention, next-best action, clinical decision making). By treating each incoming data point as an event, organizations can apply decision management and machine learning (ML) inference techniques to filter, process, qualify, and combine events to deduce higher-order information. The availability of this information paves the way for the development of intelligent applications that can offer more context-aware and personalized services to end customers, applications, and systems. We refer to these types of applications as data-intensive intelligent applications.

The life cycle of this type of 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

These stages have their own characteristics and challenges.

Characteristics

First, these activities are performed in different parts of the IT landscape. The data ingestion and runtime inference stages are increasingly performed at the edge, close to the generation of the events, and therefore reflect concepts around edge computing. Engineering and analytics are commonly performed within a cloud (private or public), dealing with concepts like data streaming, data lakes, and distributed artificial intelligence (AI) workloads. Business events and insight management address the actions that are taken based on higher-level information. This process includes recording these events in various systems using integration technologies, automated business processes, and human tasks to proactively act on higher-value events.

Second, these activities are performed by different types of users, each with individual skills, tooling, software platforms, and infrastructure requirements. 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 life-cycle 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.

Considerations

To set the context for the rest of this paper, we assume the following:

- Data is the single underlying denominator to intelligent decision making.
- Internet of Things (IoT), devices, networks, systems, and applications are all data-intensive.
- The distributed origin of data helps create more meaningful information.
- Using data to predict, remediate, or crunch logical and statistical models at the point of need yields (business) competitive advantage.
- Communicating, disseminating, and processing higher-value information up to and in business systems that may include human interaction is key to reaping its benefits in order to remain competitive, make better business decisions, pre-empt potential problems, and provide better user experiences.
- 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.

Common challenges

We identify the following challenges in the context of data-intensive intelligent applications:

- Data is stored in disparate, distributed systems and stores, and mostly in different formats.
- Data engineers, data scientists, developers, and operations all use different languages, technologies, and platforms.
- ML workloads require on-demand availability of extensive resources.
- Runtime decisioning and inference require a highly available, high-throughput, low-latency environment.

- Constant optimization of models requires the continuous deployments of workloads, introducing operational risk.
- Massive data generation from IoT devices, networks, systems, and their correlation, requires flexible, scalable, reliable, and secure storage.
- Higher-value information needs to be communicated, disseminated, and processed by separate lines of business that may be isolated units within an organization.

To build data-intensive intelligent applications within this context, we can identify some key ingredients:

- **Ingestion:** Data created by systems and devices needs to be ingested into the overall solution.
- **Filtering:** Data can be filtered by relevance at or near the source. Smart filtering based on automated decisions allows complex filtering logic to prevent network overload with potentially less relevant data.
- **Transport:** The solution needs to provide a high-throughput, elastic transport mechanism that connects its various parts. Connection of edge to cloud, as well as cloud to cloud, is required to provide a homogenous platform experience.
- **Storage:** Massively scalable data lakes can store high-volume streams of data and events. Elastic storage guarantees the expandability of the solution. High-performance in-memory datastore solutions are essential for fast data access needed for data analysis and/or model training.
- **ML modeling:** High-performing, on-demand compute resources are required to run CPU-intensive workloads like training, testing, and selection of the model with the highest prediction accuracy.
- **Model deployment and inference:** Decisioning and predictive logic need to be compiled, packaged, distributed, and deployed in an automated way across the solution, bringing the intelligence close to the physical location of the user and/or source of data.
- **Business events and insights management:** As higher-value information and events are generated by the application and AI systems, they need to be recorded in business systems and propagated to businesses processes, which may include human participation, for their consumption and processing.
- **DevOps and continuous delivery:** An agile, automated process connects the various life-cycle stages of data-intensive intelligent applications. It enables the different disciplines to collaborate on a standard platform, with common tooling for building, distributing, deploying, and serving the various workloads.

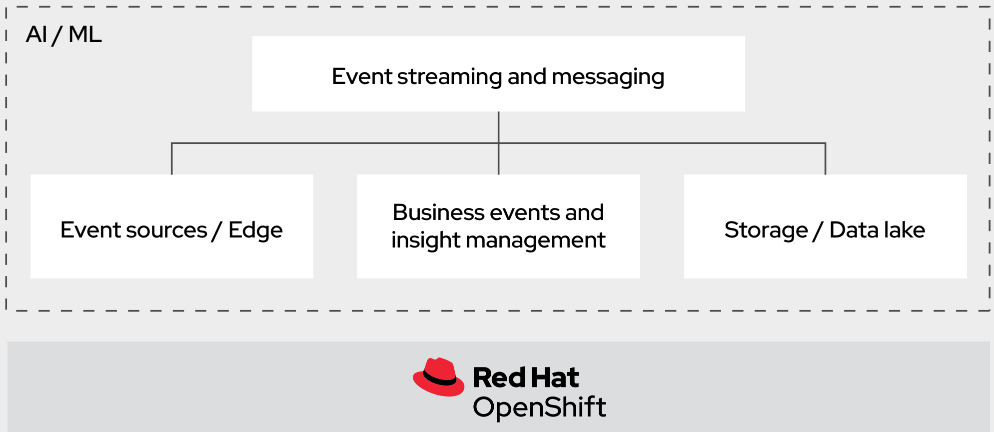

Figure 1. Components of data-intensive intelligent applications

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, which has a high-performing rules engine that can be directly embedded into a Red Hat Runtimes-based application, 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), allows for 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 outlined earlier.

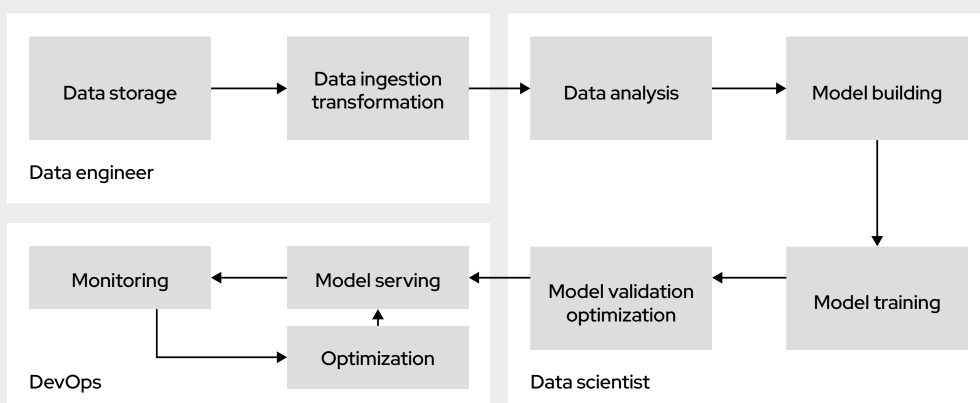

Figure 2. AI workflow and personas

The flow of data and events through a solution is essential to successful data-intensive intelligent applications. Data is ingested at multiple disparate locations from which it needs to be transported and stored. Red Hat AMQ and AMQ Streams provide robust and scalable data transfer capabilities native to Red Hat OpenShift Container Platform that 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. For example, a data lake based on Red Hat OpenShift Container Storage provides scalable, petabyte-scale storage to support AI workloads. The 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. When the cluster is no longer required by the user, the cluster is shut down, providing efficient resource 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. As with all other resources required by such a solution, these notebooks are provisioned on demand and can be torn down when they are no longer required.

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 pre-emptively inform and engage business users to better serve customers is a competitive advantage.

Supporting a DevOps and continuous deployment approach

End-to-end data-intensive intelligent applications are composed of a large set of disparate infrastructure and hardware components. These components all need to work in concert to provide the full value of the solution and deliver intelligent applications that can offer more context-aware and personalized services to end customers, 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. A DevOps approach, which has a goal of speeding up the processes by which an idea goes from development to deployment in a production environment, enables the efficient and effective operation and evolution of such solutions. To support a DevOps approach, a reliable, automated build, test, and deployment pipeline is required for all types of workloads of the solution to allow the continuous delivery of new functionality to the production systems, and thus to end users.

A 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. Automated tests and testing toolkits assess the quality of the built software components individually, as well as the integration of these components with other parts of the system. Fully automated testing and integration testing environments check the quality of the individual components and solution as a whole, providing the confidence to rapidly deploy new functionality into production.

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 Pipelines is an example of a fully cloud-native, open source, CI/CD framework that allows organizations to build, test, and deploy across multiple clouds and edge compute nodes in a hybrid cloud environment.

Conclusion

We have reviewed 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 its own specific requirements regarding skills, tools, and infrastructure. At the same time, these stages need to be connected in today's dispersed deployment environments, potentially spanning multiple cloud providers, in-house datacenters, and edge devices. Red Hat provides the technology and methodology to help enterprises build, maintain, and manage these complex hybrid cloud environments.

Data-intensive intelligent applications are a key part of a digital transformation journey, in which personalized user experiences and intelligent processing of customer data help organizations stay competitive and relevant. Red Hat OpenShift and Red Hat Middleware provide customizable components and capabilities to help your organization achieve its goals in an agile, cloud-native way on a hybrid cloud infrastructure.

facebook.com/redhatinc
@RedHat
linkedin.com/company/red-hat

North America
1 888 REDHAT1
www.redhat.com

Europe, Middle East, and Africa
00800 7334 2835
europe@redhat.com

Asia Pacific
+65 6490 4200
apac@redhat.com

Latin America
+54 11 4329 7300
info-latam@redhat.com