Top considerations for cloud-native databases and data analytics

Accelerate delivery with containers and Kubernetes

Red Hat

Data powers cloud-native applications

Across industries, cloud-native applications can help organizations differentiate themselves to gain a competitive advantage. Organizations use cloud-native applications to deliver mobile services, support e-commerce, analyze business metrics, and apply artificial intelligence (AI) and machine learning (ML) to business operations.

All of these use cases rely on data to deliver business value, and databases and data analytics capabilities are an integral part of cloud-native applications. Consequently, databases and data analytics workloads must share many of the same characteristics as the applications they support. They must:

- Be designed in a manner that allows rapid deployment, updates, and changes.

- Scale elastically to meet changing demand.

- Run consistently across datacenter, cloud, and edge IT infrastructure.

Database and data analytics use cases across industries

Healthcare
- Mobile applications for clinicians and patients
- Medical records management
- Hospital equipment management

Financial services
- Mobile banking services
- Fraud detection
- Risk analysis

Government
- Mobile applications for public services
- Web portals for public services
- Geospatial analysis for population-related trends

Energy
- Commercial applications
- Customer needs forecasting

Manufacturing and logistics
- Asset and inventory management
- Shipping route analysis

What are databases and data analytics?

Databases and data analytics are key parts of modern, cloud-native applications.

Databases are used to store important information like analytical, streaming, and transactional data. Relational databases like Microsoft SQL Server and MySQL store data in tables with defined relationships between rows, columns, and cell locations.
In contrast, non-relational databases like MongoDB, Couchbase, and Cassandra store data in a flexible, non-tabular form that supports more diverse types of data.

Data analytics refers to the set of technological capabilities for ingesting, processing, transforming, and analyzing data from multiple sources. Popular data analytics tools include Apache Kafka, Apache Nifi, Red Hat AMQ Streams, Presto, and Apache Spark.

Speed database and data analytics workload development and deployment

Through agile deployment, management, and scaling, containers and Kubernetes can help you accelerate cloud-native development of key architecture components, including databases and data analytics.

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.

Kubernetes is an open source, extensible container orchestrator. Container orchestration involves managing the creation, deployment, and life cycle of containers across your environment. Self-service capabilities let developers easily and quickly provision the resources and services they need to build cloud-native applications.

Together, these technologies provide the agility, scalability, and portability cloud-native database and data analytics workloads need.

Agility
Rapidly deploy, test, update, and manage databases and data analytics workloads to respond to evolving market requirements.

Scalability
Dynamically and elastically scale compute resources to meet fluctuating demands from database and data analytics workloads.

Portability
Deploy containerized database and data analytics workloads across your environment and easily move them as needs change.

55%
of organizations run stateful applications in containers in production.[1]

6 of the top 10
commercially available images running in containers are databases and data analytics workloads.[2]

85%
of enterprises use or plan to use Kubernetes.[3]

1   Cloud Native Computing Foundation (CNCF). "**CNCF Survey 2020**," 2020.

2   Datadog. "**11 Facts About Real-World Container Use**," November 2020.

3   Flexera. "**2020 Flexera State of the Cloud Report**," April 2020.

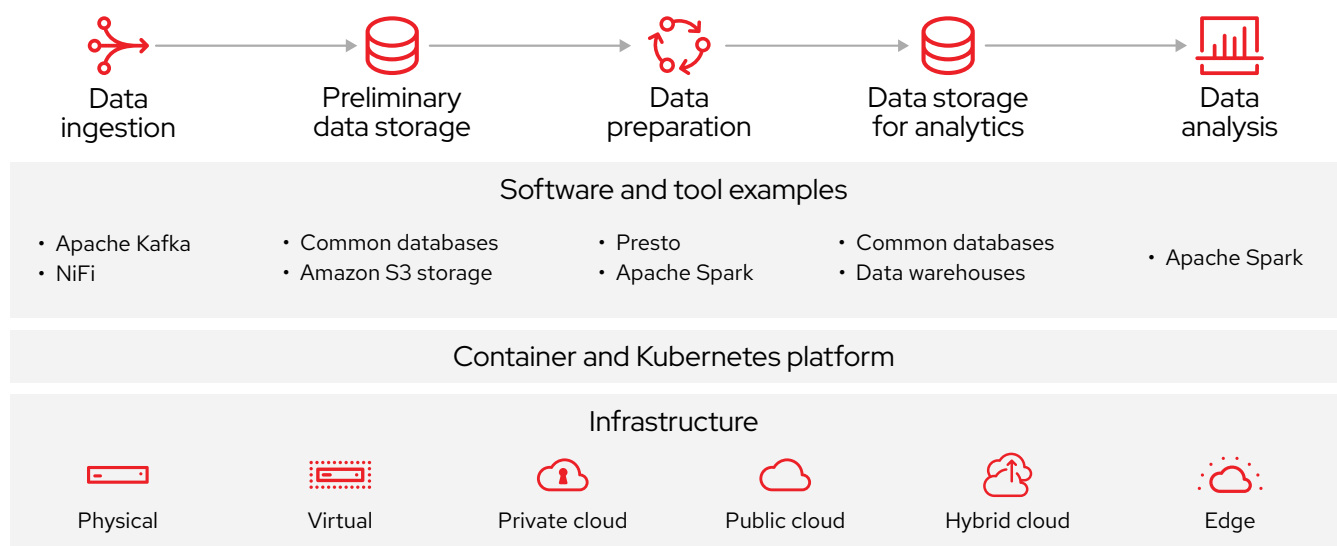# Create an efficient data life cycle using containers and Kubernetes

Implementing databases and data analytics within cloud-native applications involves several steps and tools. The main steps of the data life cycle are:

1. **Data ingestion.** Collect data from multiple sources and devices.

2. **Preliminary data storage.** Store the ingested data in operational databases or data lakes.

3. **Data preparation.** Process the stored data to discard unneeded items and convert the desired data into your preferred format.

4. **Data storage for analytics.** Store the processed, cleaned data in a separate database or data warehouse for further analysis.

5. **Data analysis.** Apply data analysis techniques and tools to generate insight.

An open, adaptable architecture will help you execute this process more effectively. This architecture requires several key technologies.

- **Data management tools** ingest, process, and analyze data from a variety of sources to deliver insights.

- **Databases, data lakes, and data warehouses** store data throughout the journey from collection to use in cloud-native applications.

- **Container and Kubernetes platforms** provide a consistent foundation for deploying databases, data analytics tools, and cloud-native applications across infrastructure, as well as self-service capabilities for developers and integrated compute acceleration.

- **Software-defined infrastructure** provides resources across hybrid cloud environments — including datacenter, edge, and public cloud infrastructure — for all stages of cloud-native application development and deployment.

Data life cycle and conceptual architecture

Data ingestion	Preliminary data storage	Data preparation	Data storage for analytics	Data analysis
Software and tool examples				
• Apache Kafka • NiFi	• Common databases • Amazon S3 storage	• Presto • Apache Spark	• Common databases • Data warehouses	• Apache Spark
Container and Kubernetes platform				
Infrastructure				
Physical	Virtual	Private cloud	Public cloud	Hybrid cloud / Edge

Choose the right platform for your organization's needs

Organizations often have several concerns when considering whether to deploy containerized databases and data analytics tools within cloud-native development and deployment workflows.

- **Will I experience downtime and data loss?**
  Adopting new architecture and approaches can carry the risk of downtime due to failures and errors due to lost or corrupted data.

- **Will this add to our operational complexity?** Cloud-native application development and deployment can be more complicated than your more traditional development operations. Containerized databases and data analytics tools can add to this complexity.

- **Will this support my chosen third-party software and tools?** Your existing development and data management tools may not be certified to run on all container platforms, requiring you to choose and learn new tools.

- **Can my organization deploy this successfully?** Container and Kubernetes deployment is not a trivial task and many organizations do not have the time, resources, or experience to successfully implement a containerized environment from disparate technologies.

What to look for in your container and Kubernetes platform

There are many container and Kubernetes platforms available, each with different capabilities, features, and support. Choosing the right platform can help you accelerate your data life cycle as well as cloud-native application development. Look for container and Kubernetes platforms that offer:

- Production-grade availability, reliability, and scalability.
- Simple container deployment, management, and portability.
- Integration with a broad set of third-party products.

- Simple adoption and intuitive use for developers and IT staff.
- Easy-to-use self-service and automation capabilities.
- Integrated security features and data protection.

Implement containerized workloads with an industry leader

As an open source leader, Red Hat provides a complete technology portfolio, proven expertise, and strategic partnerships to help you achieve your cloud-native development goals. With a broad ecosystem of partner technologies, the company delivers a foundation for implementing data-driven cloud-native application development and deployment, as well as services and training for rapid adoption.

- **Proven expertise.** Red Hat has extensive experience helping global organizations containerize workloads in open hybrid cloud and multicloud environments.

- **Comprehensive portfolio.** Red Hat offers a complete software portfolio for cloud-native application architecture and development.

- **Powerful partnerships.** Red Hat fosters strategic and collaborative engineering partnerships with leading technology, software, hardware, and service vendors, including database and data analytics software providers.

- **Open source advocacy.** Red Hat is a trusted provider of enterprise open source software and works with upstream communities to help you innovate faster.

Red Hat® OpenShift® is an enterprise-ready Kubernetes platform that provides an ideal foundation for containerized database and data analytics workloads that support cloud-native application development. On-demand compute resources and consistency across hybrid cloud environments — including datacenter, edge, and public cloud infrastructure — deliver the speed and flexibility you need for data- and compute-intensive workloads. Self-service provisioning allows development, database, and data analytics teams to provision and scale resources without continual IT engagement. Consistency throughout the development life cycle accelerates time to market for applications while improving deployment reliability. And collaboration features let developers and data administrators create and share containerized applications and data services with peers in a consistent manner.

Red Hat OpenShift also integrates with other Red Hat products to provide a complete foundation for cloud-native operations. Key products include:

- **Red Hat Data Services**: Software-defined storage for containers and petabyte-scale deployments.

- **Red Hat Application Services**: Frameworks, integration solutions, process automation, runtimes, and programming languages for cloud-native application development and deployment.

- **Red Hat Enterprise Linux®**: A robust and security-focused foundation for running all types of applications in — and out of — containers.

Finally, **Red Hat Marketplace** offers a variety of databases and data analytics tools for streamlined procurement and installation on Red Hat OpenShift across environments.

> "Previously, it took us over a month to add new features, but now it takes as little as two weeks. We can react much more quickly to changing consumer demands and give them the features they expect, with less time invested."
>
> **Thanussak Thanyasiri**
> Senior Delivery Manager, Kasikorn Business-Technology Group (KBTG)

Deploy a foundation for cloud-native databases and data analytics

Red Hat OpenShift delivers key capabilities and benefits for building and deploying database and data analytics workloads and cloud-native applications across hybrid cloud environments.

Automated operations

Red Hat OpenShift supports automated deployment and operation of containerized software through **Kubernetes operators** and Helm charts. These technologies simplify initial installation and ongoing life-cycle management of containerized databases and data analytics workloads as well as many tools and software from certified Red Hat partners.

Result: Streamline operations through automation and self-service operations.

Database and data analytics partner ecosystem

Red Hat OpenShift integrates with a large selection of database and data analytics products from certified ecosystem partners like Couchbase, Crunchy Data, Cloudera, Microsoft, and MongoDB, as well as open source technologies like Apache Kafka and Apache Spark. Many of these partners also offer certified Kubernetes operators to simplify deployment and life-cycle management.

Result: Easily deploy and use your preferred tools and software in your cloud-native applications.

Consistency and portability

Red Hat OpenShift offers a consistent way to deploy all cloud-native application components—including databases, data analytics, and other workloads—on a unified platform that works consistently across all development, test, and production environments and application life-cycle phases.

Result: Deploy workloads where it makes sense now and easily move them as needs change.

> "We strongly believe open source provides not only key technology, but an optimal work approach, to create and scale holistic platforms that adapt to existing and future tools and cloud technology."[4]

Joaquín Moraga Gallego
Head of Architecture, Santander Tecnología

---

4 Red Hat case study. "**Spanish financial group speeds delivery with enterprise IT residency**," December 2020.

Take advantage of a broad database and data analytics partner ecosystem

Red Hat fosters a growing ecosystem of certified database and data analytics partners, allowing you to incorporate popular products and technologies to best fit your organization's requirements.

Red Hat database and data analytics certified partner ecosystem

> "Red Hat Marketplace supports the shift we're seeing to Kubernetes, workload portability, and a cloud-first approach. It allows us to expand our reach to Red Hat OpenShift users and gives them ways to make hybrid cloud management easier."[5]

Alan Chhabra
Senior Vice President, Partners and APAC Sales, MongoDB

---

5   Red Hat press release. "Red Hat Marketplace Aims to Accelerate Open Hybrid Cloud Innovation With Certified Software Solutions Ready to Run on Any Cloud," 8 September 2020.

Santander Tecnología

Accelerate innovation with open technologies

Santander Tecnología

Challenge

Santander Tecnología, the in-house technology division of Santander Group, supports the group's digital transformation efforts to improve the customer experience. Santander Tecnología's application development architecture, Darwin, consists of innovative technology solutions, from big data to the AngularJS open source web framework. Darwin is available on-premise and in the cloud, but adapting the architecture for different development teams, changing business demands, and multicloud use required an easier way to customize configurations and allocate resources.

Solution

Santander Tecnología worked with Red Hat Consulting to support its adoption of two new Red Hat technologies: Red Hat OpenShift and Red Hat Data Grid. As part of this technology adoption, the company participated in Red Hat Open Innovation Labs. Over a residency of several weeks, Santander Tecnología's teams worked with Red Hat experts to optimize its adoption of Red Hat OpenShift and adjust processes to deliver desired features and services to customers faster.

Santander Tecnología now runs its microservices-based Home Banking and Web Empresas applications on Red Hat OpenShift, supported by Red Hat Data Grid, MongoDB, and Apache Kafka, with a separate persistence layer.

> "Participating in Red Hat Open Innovation Labs and adopting technology like Red Hat OpenShift has demonstrated our commitment to exploring new ideas. We aim to attract passionate professionals, and we can't do that if we're working with proprietary, closed-source technology using traditional methods."
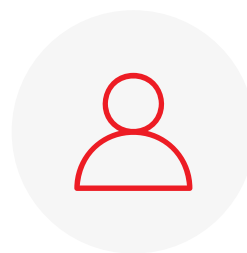>
> **Joaquín Moraga Gallego**
> Head of Architecture, Santander Tecnología

Reduced provisioning time from weeks to 10-12 minutes

Established CI/CD and DevOps processes for faster time to market

Improved ability to attract and retain skilled talent

Ready to deploy data-driven applications?

Databases and data analytics can help you deliver differentiated cloud-native applications and gain a competitive advantage. Red Hat OpenShift gives you a consistent foundation for implementing data-driven cloud-native application development and deployment to support your business goals.

Discover the benefits of containerized databases and data analytics workloads: openshift.com/data

Take your cloud-native development journey further with Red Hat Consulting.

Red Hat experts can help you, your team, and your organization develop the practices, tools, and culture needed to more efficiently modernize existing applications and to build new ones using technologies like containers, Kubernetes, databases, and data analytics.

Assess your organizational maturity: red.ht/rta