

IBM + Cloudera

Drive business outcomes with an enterprise-grade, secure, governed and open data and AI solution across cloud



Table of contents

03	Executive summary
04	A history of collaboration
05	IBM and Cloudera driving the future of Artificial Intelligence (AI)
07	IBM and Cloudera for big data, analytics and AI
08	IBM and Cloudera offerings: <ul style="list-style-type: none">– Hardware essentials– Enterprise data platform– Ingest and process– Govern and secure– Query and analyze– Service and support
13	Why IBM and Cloudera are better together for your business
14	About IBM
14	About Cloudera



Executive Summary

As business leaders navigate a succession of transformation imperatives, data, analytics and artificial intelligence (AI) continue to enable organizations to address urgent and immediate business priorities—quickly and at scale. A data-driven approach can improve a company’s cost base—augmenting human capability to motivate greater and more expansive efficiencies. It also helps enhance or protect top line revenue, experience, and engagement while addressing cybersecurity risks and compliance and governance challenges.

Leading companies are turning decision-making into an industrial process—integrating data gathering, analytics and modeling—to make business decisions with ever-improving accuracy. Connecting and securing this data lifecycle is foundational to scaling data-driven decision-making with AI. IBM and Cloudera make this and more possible with the Cloudera Data Platform that manages and secures the data lifecycle in any cloud or data center and an ecosystem of IBM products, services and multi-vendor support.

IBM and Cloudera deliver an enterprise data platform for hybrid cloud with a data fabric, designed to quickly build data-driven high value applications once, and run them anywhere. Cloudera’s next generation data platform is the industry’s first enterprise data cloud built for hybrid cloud deployments that deliver cloud native agility across public and private clouds, leveraging RedHat OpenShift® as well as public cloud.

A history of collaboration

The IBM and Cloudera strategic partnership builds on a long-standing relationship that began with Hortonworks (HWX). In January 2019, HWX and Cloudera merged, creating integrated solutions that extended to include the CDP. In June 2019, IBM and Cloudera formed a new partnership to continue the mission to drive data-driven decisions for customers.

Current solutions include on-premises and public cloud offerings with the data management, security and governance needed to build an enterprise data and analytic big data solution. Discover how the enterprise data cloud benefits businesses and why they should consider IBM and Cloudera to build the ultimate edge-to-AI solution.

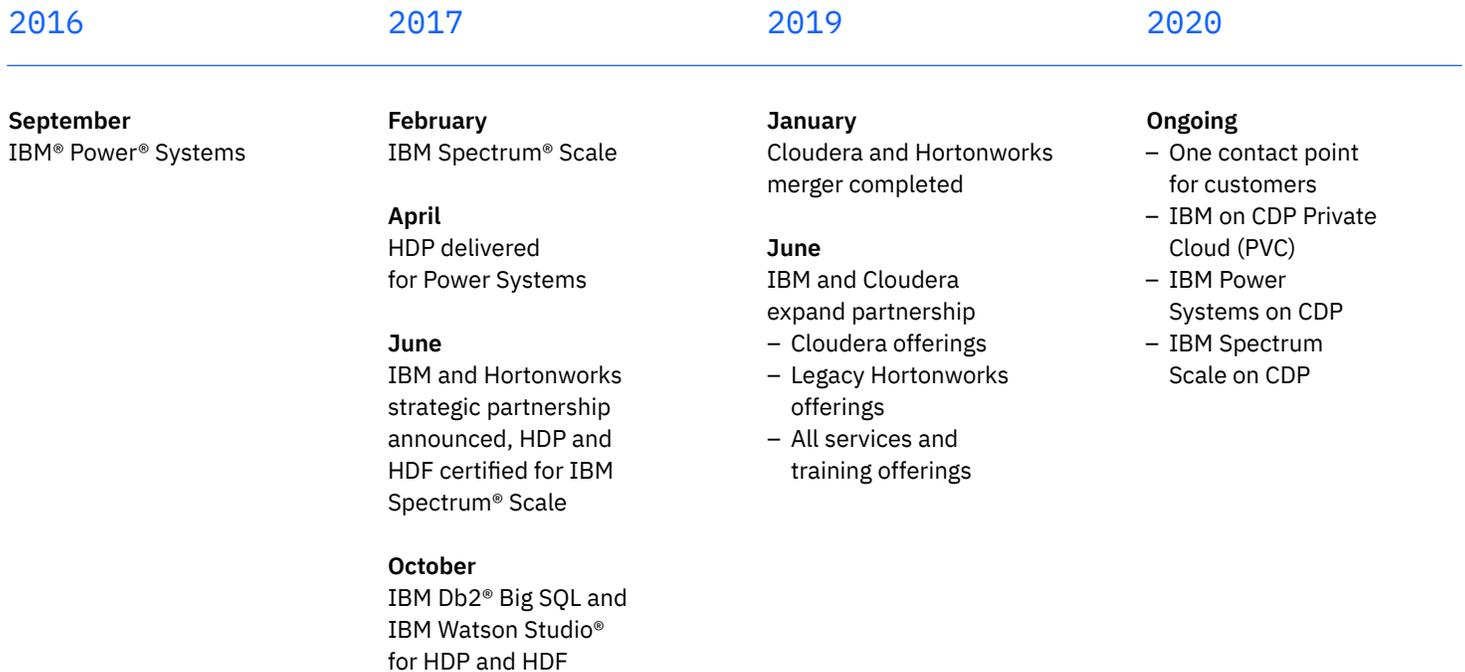


Figure 1: A timeline of IBM and Cloudera's partnership

IBM and Cloudera driving the future of AI

AI systems are developed to respond with the intellectual process characteristic of humans; the ability to reason, generalize or learn from experience. The purpose of AI is to aggregate and process large amounts of unstructured, semi-structured and structured data; use modeling and analytics to drive insights for smarter decision making; and secure and streamline workflows driving new routes to market and revenues.

Businesses are at an inflection point, looking to re-engineer business processes to drive agility through more resilient workflows, which they hope will, in turn, create greater cost efficiencies and meet the needs of virtual customers and staff. There's growing support of AI as a transformative technology. Consider the facts from two studies:

Consider the IBM Global AI Adoption Index¹

Almost one-third of companies reported using AI in their business. For those deploying or exploring AI, they report it is being driven by multiple pressures and opportunities businesses are facing, from the COVID-19 pandemic to advances in the technology that make it more accessible.

A third of global IT professionals report their company plans to invest in both skills and AI solutions over the next 12 months.¹



Industry use cases

IBM clients are seeing the value of AI to inject new levels of analytic decisions, productivity and efficiency across industries:

Financial services

Financial services companies are at the head of the curve in using advanced analytics and AI, experiencing huge benefits in improving the customer experience, creating new products and pursuing new markets, according to Deloitte Insights.² The competitive edge gained by these companies will widen if other companies don't move quickly in their AI adoption. A survey of financial services companies under the aegis of the World Economic Forum reveals that 64% of respondents are planning to employ AI across key areas³ such as:

- Customer lead gen and retention: Acquire new customers and, based on data and trends, provide better personalized service to existing ones through targeted products and services.
- Fraud and risk management: Proactively identify and remediate fraudulent behavior, forgery, credit scams and insider threats.
- Innovation: Drive revenue with companion and next best offers. Provide automated on-line self-service tools, chatbots, and automated messages to quickly meet changing customer needs.

Healthcare

Healthcare companies are facing new and urgent demands. Successful companies are finding new ways of connecting and integrating their data, predicting and driving agility to adapt to new solutions ahead of the curve in areas like:

- Disease tracking: COVID 19 has brought advanced big data and analytics technology for detecting propensities and patterns for diseases and diagnosis to the forefront. Data and analytics are driving the ability to track the spread worldwide, identify clusters/trends, monitor healthcare resources, and treatment results.
- Electronic health records: Healthcare companies are utilizing data and analytics to build models for personalized, predictive and preventive medical care, as well as to lower costs and improve the patient experience.

- Wearable medical devices: New IoT devices in the form of wearable technology are providing real-time visibility and new patient monitoring options.
- Equalize medical knowledge: Provide AI-based virtual medical decision support with IBM Watson which combines evidence-based drug and disease information to support clinicians at the point of care. Free access for qualified scientist and academics to thousands of peer-reviewed papers and licensed data bases is available to help extract critical COVID-19 knowledge.

Communication service providers (CSPs)

Communication service providers (CSPs) are at the threshold of a major transformation of their businesses and business models. The transition to 5G networks and the proliferation of the Internet of Things (IoT) gives CSPs an opportunity to create a wide range of new revenue streams by enabling compelling use cases such as smart cities, connected factories, self-driving vehicles, data exchanges and many more.

- Innovation: Better understand the need for new products and services by region and demographics.
- Customer lead generation and retention: Enhance customer experience with high performance services, fast feedback and customized offers. Augment internal data with market insights to reduce costs. Create new revenue streams by predicting customer behaviors, preferences and movements. Utilize AI based chat bots to encourage self-service and limit time spent with an agent for greater productivity.
- Forecast network capacity: Proactively address high drop rates and poor-quality call connections. Better measure traffic, network performance and routing to improve service and scalability.

IBM and Cloudera for big data, analytics and AI

A complete solution for the enterprise

IBM and Cloudera's solutions offer massive scalability and governance in a security-rich environment along with the ability to federate data-at-rest and data-in-motion, across the entire organization and from external sources. The solutions are designed for users to easily query data warehouses as well as data lakes, on premises or in the cloud. Users also benefit from self-service data access and the ability to do ad hoc and real-time queries. Ultimately, IBM and Cloudera solutions are built to better support data-in-motion, IoT, ML and data science at enterprise scale. Their individual strengths combine to provide a leading enterprise data cloud platform for businesses, offering:

- Flexible hybrid cloud options, covering multiple public clouds and on-premises deployment
- Software support through one vendor
- Accelerated time to value, with easy-to-use, self-service analytics integrated with leading industry and technology experts globally
- A constantly innovating data platform, based on open source technology
- Vertical specialism providing in-depth knowledge of industry and technology experts globally

IBM and Cloudera's strategic partnership offers customers a direct path to AI with easy access to the capabilities, scalability and economy of an advanced data platform, governance, security and the tools for data federation, advanced querying and data management. The results are inventive data and analytic solutions ready for your enterprise and future data growth.

Most enterprises today have large scale data assets on-premises while also seeking more agility for the business with cloud deployments. They need choice and flexibility depending on their workloads, data types and the need for security and compliance. To be successful with today's AI initiatives they need to obtain clear and actionable insights from complex data anywhere, from the edge to AI. Ultimately, IBM and Cloudera solutions are built to better support machine learning and data science at enterprise scale, both on-premises and in the cloud.

IBM and Cloudera—better together

- Leading SQL engine for complex analytic workloads
- Open source distribution leader in data science
- Leader in on-premises and hybrid cloud solutions
- 2,500+ customers and 3,000 ecosystem partners
- Leading professional services and training
- Top-tier 24x7 customer support
- Simplify data access with a [data fabric](#)

IBM and Cloudera offerings

Unlock the value of the open source ecosystem with Cloudera Data Platform, custom Cloudera solutions, and IBM value-add solutions, products and services.



Hardware essentials

IBM Power Systems

IBM Power® Systems cloud-ready servers are built for the most demanding, data-intensive computing on earth. Unleash insight from your data pipeline—from managing mission-critical data, to managing your operational data stores and data lakes, to delivering the best server for cognitive computing.

- Enterprise cloud-ready to integrate into your organization’s private or hybrid cloud strategy for flexible consumption models.
- Ranked #1 for uptime by ITIC, IBM Power Systems deliver reliable on-premises infrastructure 24/7.⁴
- Deliver both performance and price-performance advantages with superior core performance and memory bandwidth.

IBM Spectrum Scale

IBM Spectrum Scale™ provides advanced object storage and management of unstructured data for cloud, big data, analytics, objects and more. It supports both new-era big data and traditional applications with security, reliability and high performance. IBM Spectrum Scale is a high-performance solution for managing data at scale with the distinctive ability to perform archiving and analytics in place.

- Leverage no-bottleneck architecture to scale performance for extreme throughput and low-latency access.
- Grow and share your storage infrastructure while automatically moving file and object data to the optimal storage tier.
- Take advantage of authentication, encryption, security and replication options to meet your business and regulatory requirements.

Enterprise data platform

Cloudera offers the breadth of data analysis disciplines needed to solve the most demanding business use cases. These disciplines span from the edge to AI, streaming analytics, flow management, data engineering, data warehousing, operational analytics, data science and machine learning—all working together and operating across a choice of security-rich environments including data centers and multiple public, hybrid clouds.

Cloudera Data Platform supports the following capabilities:

- Hybrid cloud flexibility across on-premises and multiple public clouds
- Easy-to-use, multifunction analytics
- Shared services for security, governance and insight (SDX)
- An open source foundation with a choice of compute and storage

Each Cloudera offering is explored in more detail in this document.

Cloudera Data Platform (CDP)



Integrated data management and analytics platform

- Empowers companies to get clear and actionable insights from complex data virtually anywhere, from the edge to AI
- Provides the flexibility to run modern analytic workloads anywhere—regardless of where data resides
- Offers the ability to move those workloads to different cloud environments—public or private—to avoid vendor lock-in
- Has the agility, elasticity and ease of use of public clouds and a common security and governance framework to enable data privacy and regulator compliance by design

Cloudera Data Platform (CDP) is an Integrated data management and analytics platform that is easy to deploy, manage, and use. By simplifying operations, CDP reduces the time to onboard new use cases across the organization. It uses machine learning to intelligently auto scale workloads up and down for more cost-effective use of infrastructure.

CDP manages data in any environment, including multiple public clouds, bare metal, private cloud, and hybrid cloud. With Cloudera’s Shared Data Experience (SDX)—the capabilities in CDP providing security, governance, and insight—IT can confidently deliver secure analytics running against data anywhere. CDP is a new approach to enterprise data, anywhere from the edge to AI. CDP is available as a public cloud service, as an on premise and as a cloud option.

Cloudera Data Platform—Public Cloud

Create and manage secure data lakes, self-service analytics, and machine learning services without installing and managing the data platform software. **CDP Public Cloud** services are managed by Cloudera; however, unlike other cloud services, your data will always remain under your control in your VPC. CDP runs on AWS, Azure and Google Cloud Platform. CDP Public Cloud lets you:

- Control cloud costs by automatically spinning up infrastructure to meet the needs of workloads and suspending their operation when complete
- Isolate and control workloads based on user type, workload type, and workload priority
- Combat proliferating silos and centrally control customer and operational data across multi-cloud and hybrid environments

Cloudera Data Platform—Data Center and Private Cloud

Take the data architecture built for cloud technologies and run it on-prem, in your data center with Red Hat OpenShift. Instead of thinking about clusters, you can now focus on the data and the needs of your users to ingest, transform, query, serve and predict. [Cloudera Data Platform](#) provides you the ability to:

- Quickly and easily on-board new workloads and users to meet the needs of the organization
- Isolate “noisy” workloads ensuring critical workloads meet SLAs
- Customize environments to meet the specific needs of all your users while maintaining centralized control
- Improve utilization of infrastructure with better efficiency

RedHat OpenShift

[Red Hat® OpenShift®](#) provides the framework for CDP on-premises cloud. The RedHat platform is based on open technologies such as Linux and Kubernetes for securely deploying, running and managing data and applications on premises—without the risk of being locked in. Red Hat® OpenShift® on-premises improves velocity, market responsiveness, scalability and reliability.

- Build mission-critical applications once and run them on private clouds with IBM’s next-gen, hybrid, multi-cloud platform.
- Optimize resource usage, saving private cloud costs and licensing fees by preventing the need for future hardware with containerization and hybrid cloud.
- Reduce testing and maintenance with self-service, containerized middleware catalogs and automation.

Ingest and process

Cloudera Data Flow

Cloudera Data Flow (CDF) is a highly scalable, real-time streaming analytics solution that ingests, curates, and analyzes data for key insights and immediate actionable intelligence. CDF is designed to process large volumes of big data, transporting from a multitude of sources. CDF accommodates real-time streaming data, tracks provenance and lineage, and manages edge applications.

- Manage your data from the edge to the enterprise with a no-code approach to developing sophisticated streaming applications.
- Reduce data integration development time with real-time data streaming at high volume and scale.
- Manage and secure data from edge to enterprise CDF makes collecting data of every type simple and accessible from rapidly changing data sources.

Govern and secure

Cloudera Shared Data eXperience

[Cloudera Shared Data eXperience \(SDX\)](#) provides a powerful set of data and metadata shared services for security, governance and insight. Policies are set once and they are automatically enforced across data and analytics in hybrid, private and multi-public clouds across all analytics and machine learning. SDX is used to:

- Achieve and maintain regulatory compliance (e.g. GDPR and CCPA) across all data and environments and lets you identify and manage sensitive data.
- Ensure metadata and policies are part of every data replication or migration, regardless of its purpose, and across on-premises, private and public cloud infrastructures.
- Deliver unified data management with security, governance and control policies that are set once and consistently enforced seamlessly across multiple analytic workloads running against the same diverse data sets on fluid multi and hybrid cloud infrastructures.

SDX creates a trusted, business-ready foundation with consistent governance, protection, and compliance regardless of the location or analysis of the data.

IBM DataOps

A [governed](#) data lake contains clean, trusted data from structured and unstructured sources that can easily be found, accessed, managed and protected. The platform your data resides on is security-rich and reliable. Data that comes into your data lake is properly cleaned, classified and protected in timely, controlled data feeds that populate and document it with reliable information assets and metadata.

- Ingest structured and unstructured data from all your data sources in real-time
- Use an intelligent metadata catalog and better understand the data visually
- Maintain trusted data by protecting the integrity and reliability of your data through governance policies
- Keep data compliant and audit ready

IBM Big Replicate

[IBM Big Replicate](#) provides active/active data replication on-premises, in hybrid cloud environments or for hybrid deployments, including SQL and NoSQL databases. It replicates big data from lab to production, from production to disaster recovery sites, or from ground to cloud object stores supporting governance of the most demanding business and regulatory requirements. Big Replicate minimizes downtime when replicating data across environments.

- Minimize downtime when replicating data across environments
- Provides real-time data replication
- Provides HA/DR with virtually zero RTP
- Replicates data from on-premises to cloud and cloud to on-premises with minimal disruption

Query and analyze big data

Cloudera Data Warehouse (CDW) is an auto-scaling, highly concurrent and cost-effective analytics service that ingests high-scale data anywhere, from structured, unstructured and edge sources. It supports hybrid and multi-cloud infrastructure models by seamlessly moving workloads between on-premises and any cloud for reports, dashboards, ad-hoc and advanced analytics, including AI, with consistent security and governance. Cloudera Data Warehouse offers extremely low query wait times, reduced IT costs and agile delivery. It is used to:

- Optimize legacy data warehouse implementations, by partially or fully migrating difficult workloads from traditional data warehouse infrastructures to Cloudera Data Warehouse
- Analyze large amounts of events and time-series data originating from machine logs, sensors, and other devices at the edge
- Explore and discover new correlations, patterns, and insights to inform your business on what to expect in the future by sifting through vast amounts of unstructured data and correlating them with relational data

CDW analyzes data in modern ways to provide new insights and make better, smarter decisions.

IBM Db2 Big SQL

[IBM Db2® Big SQL](#) is a hybrid, highly scalable enterprise-grade SQL engine for Apache Hadoop that delivers easy data querying across the enterprise. It concurrently exploits Hive, HBase and Spark using a single database connection—or even a single query.

IBM Db2® Big SQL, organizations can connect to disparate sources such as HDFS, RDMS, NoSQL databases, object stores and Web HDFS.

Enjoy low latency, support for ad-hoc and complex queries, high performance, robust security, SQL compatibility and federation capabilities to get the most from your data warehouse.

- Understands commonly used ANSI SQL syntax to perform queries on batch and real-time data across Hadoop, object stores and data warehouses
- Shift with hybrid-cloud-ready flexibility within private cloud and on-premises environments
- Federate data natively across the enterprise with Db2
- Run all 99 TPC-DS queries up to 100 TB with high concurrency⁵

Cloudera Machine Learning

[Cloudera Machine Learning \(CML\)](#) provides enterprise data science teams access to business data and the tools and computing resources required for end-to-end machine learning workflows, while giving IT and the business the ability to maintain data governance and control infrastructure costs. Cloudera Machine Learning brings the agility and economics of cloud to self-service machine learning workflows with governed business data and tools that data science teams need, anywhere. It is used to:

- Easily replicate governed business data from on-premises to cloud and deploy new ML workspaces for teams with pre-configured resource consumption guardrails
- Enable batch scoring on large amounts of data such as image or sensor data that resides in cloud storage
- Deploy new ML workspaces for teams with pre-configured resource consumption guardrails that deliver access to the tools and computing resources needed for model training and deployment

CML operationalizes AI throughout the organization with AI models built with trust and transparency.

IBM Data Science—Watson Studio

Simplify and scale data science across any cloud. Prepare data and build models anywhere, using open source codes or visual modeling. Predict and optimize your business outcome to accelerate machine and deep learning workflows required to infuse AI. With its open, flexible multicloud architecture, [Watson Studio](#) provides capabilities that empower businesses to simplify enterprise data science and AI, such as:

- Automating AI lifecycle management with AutoAI
- Preparing, exploring and refining data
- Creating or bringing your own open source notebooks
- Simplifying model development
- Running and training models
- Automating data prep with end-to-end integrated UI

Use Watson Studio to build and train AI and machine learning models, prepare and analyze data with automated deep learning and enhanced visual modeling. Use this single environment to prepare and analyze data, then build and train your AI.

Service and support

IBM Multivendor support

Trust IBM to be your [technical support and maintenance advisor](#). Streamline your maintenance support for hardware, software and your multiple vendor relationships with a single point of contact and put IBM's enhanced analytics, global infrastructure and years of experience to work for you.

- One contract, one call. IBM provides a single point of contact.
- Lifecycle support optimizes the longevity of equipment while enabling more predictable and consistent budgeting.
- Relieve your IT staff from the day-to-day burdens of IT management.

IBM IT Global Infrastructure Services

[IBM IT Global Infrastructure Services](#) help you gain the advantage of the agility, security, speed and flexibility of their IT infrastructure.

Why IBM and Cloudera are better together for your business

Cloudera and IBM are bringing advanced data and analytics solutions to more organizations to promote a digitally connected future with the enterprise data cloud. Together we offer cutting edge solutions through open-source innovation for any data, anywhere, from the edge to AI.

Freedom of choice

Cloudera and IBM are driving integrated options for data workloads in your business, including offering next-gen hybrid and multi-cloud enterprise data platforms. On-premise offerings now also include Cloudera Data Platform available on IBM hardware.

Speed time to innovation

In partnership with IBM and Red Hat, Cloudera drives the largest number of contributions to the open-source community, meaning innovation of open-source products is moving at a faster pace than ever. In addition, the open-source model ensures increased availability and interoperability across all vendors for your enterprise data cloud—uniquely hybrid.

Secure and governed

With the IBM and Cloudera partnership, your data is secured and governed across the enterprise, from edge to AI.

One-stop support

Whether using commercial products, individual packages or a complex software stack, we provide multi-vendor support. IBM clients can [save time and cost](#) by using multi-vendor support with a one-stop support model.

Faster ROI

Our end-to-end capabilities are transforming organizations adopting AI and machine learning, building, governing and driving analytical insights from data lakes, or connecting clouds with traditional infrastructures—all through our global services offerings.

Industry expertise

Whatever your vertical, IBM and Cloudera have experience with enterprise organizations across virtually every industry. We can build an integrated vision and plan for your organization designed to solve vertical specific challenges, from customer 360 to predictive maintenance and modeling.

About IBM

IBM provides a complete set of AI-enabled solutions that allow organizations to collect data of any type, source and structure to make it simple and accessible across multiple vendors, deployments and workloads. Our enterprise grade, secure solutions support hybrid multi-cloud environments and automation through embedded AI capabilities.

About Cloudera

Cloudera believes that data can make what's impossible today, possible tomorrow. It empowers people to transform complex data into clear and actionable insights. Cloudera delivers an enterprise data cloud for virtually any data, essentially anywhere, from the edge to AI. Powered by the relentless innovation of the open source community, Cloudera advances digital transformation for the world's largest enterprises. Learn more at [Cloudera.com](https://cloudera.com).

To learn more about IBM and Cloudera solutions, please contact your IBM representative or visit the [IBM/Cloudera partner page](#).



Endnotes

1. <https://newsroom.ibm.com/IBMs-Global-AI-Adoption-Index-2021>
2. "AI Leaders in Financial Services: Common Traits of Frontrunners in the artificial intelligence race," Deloitte Insights, August 13, 2019
3. "Transforming Paradigms: A Global AI in Financial Services Survey," World Economic Forum, et al, January, 2020
4. <https://www.ibm.com/it-infrastructure/us-en/resources/power/fivenines-power9/>
5. IBM United States Software Announcement220-141, dated March 24, 2020

© Copyright IBM Corporation 2022

IBM Corporation
New Orchard Road, Armonk, NY 10504
Produced in the United States of America
May 2022

IBM, the IBM logo, IBM Db2, IBM Big Replicate, IBM Power Systems, IBM Spectrum Scale, Watson, and ibm.com are trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademark is available on the Web at "Copyright and trademark information" at [HYPERLINK "http://www.ibm.com/legal/copytrade.shtml"](http://www.ibm.com/legal/copytrade.shtml)www.ibm.com/legal/copytrade.shtml.

Red Hat and OpenShift are rademarks or registered trademarks of Red Hat, Inc. or its subsidiaries in the United States and other countries.

This document is current as of the initial date of publication and may be changed by IBM at any time. Not all offerings are available in every country in which IBM operates.

It is the user's responsibility to evaluate and verify the operation of any other products or programs with IBM products and programs. THE INFORMATION IN THIS DOCUMENT IS PROVIDED "AS IS" WITHOUT ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND ANY WARRANTY OR CONDITION OF NON-INFRINGEMENT. IBM products are warranted according to the terms and conditions of the agreements under which they are provided.

The client is responsible for ensuring compliance with laws and regulations applicable to it. IBM does not provide legal advice or represent or warrant that its services or products will ensure that the client is in compliance with any law or regulation.

Statement of Good Security Practices: IT system security involves protecting systems and information through prevention, detection and response to improper access from within and outside your enterprise. Improper access can result in information being altered, destroyed, misappropriated or misused or can result in damage to or misuse of your systems, including for use in attacks on others. No IT system or product should be considered completely secure and no single product, service or security measure can be completely effective in preventing improper use or access. IBM systems, products and services are designed to be part of a lawful, comprehensive security approach, which will necessarily involve additional operational procedures, and may require other systems, products or services to be most effective. IBM DOES NOT WARRANT THAT ANY SYSTEMS, PRODUCTS OR SERVICES ARE IMMUNE FROM, OR WILL MAKE YOUR ENTERPRISE IMMUNE FROM, THE MALICIOUS OR ILLEGAL CONDUCT OF ANY PARTY.

