

IBM + Cloudera: Better together

Overcome enterprise data
challenges with one solution

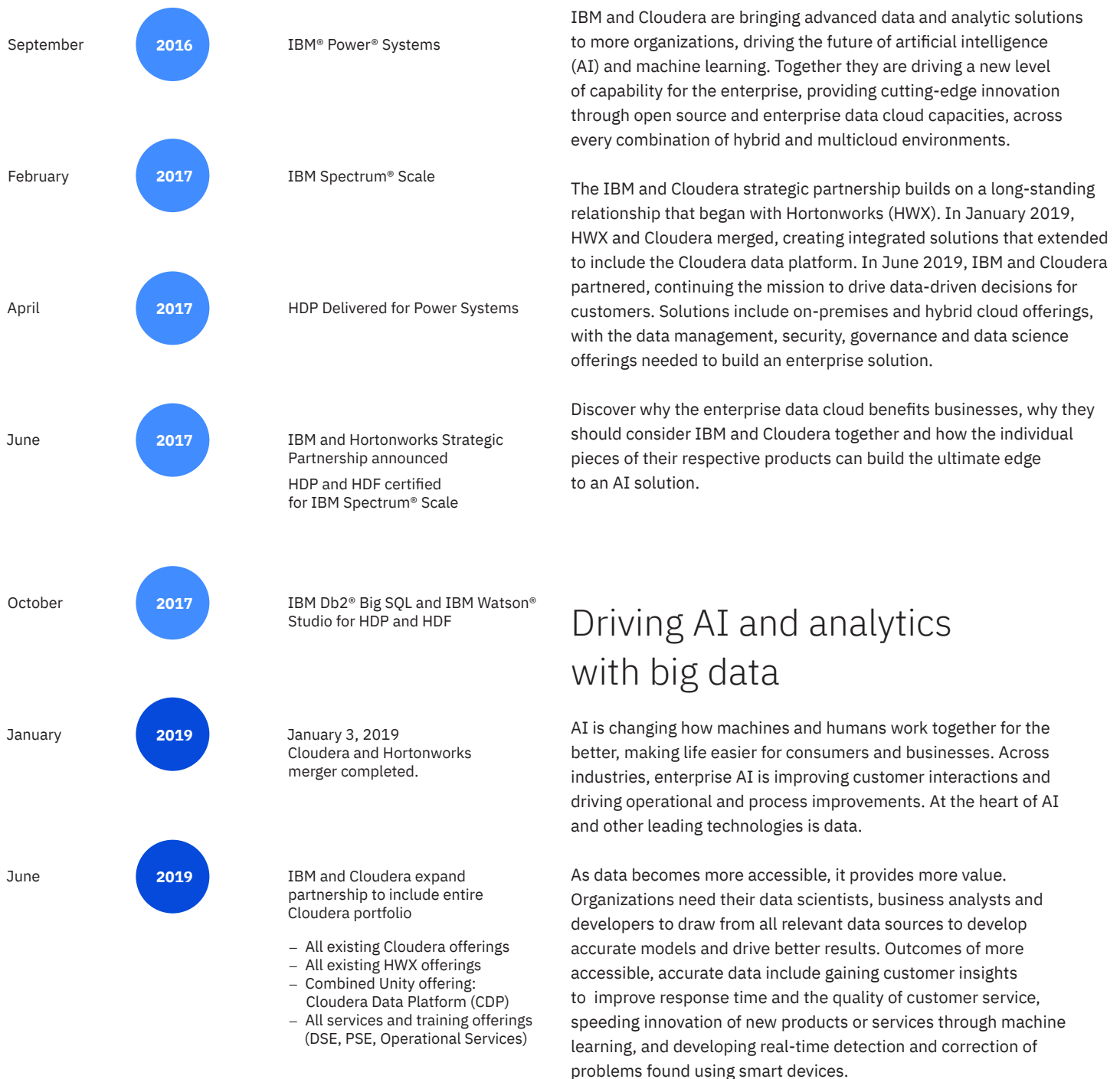


Figure 1: A timeline of IBM and Cloudera's partnership

Industry use cases

Clients are realizing the value of AI for injecting new levels of analytic decisions, productivity and efficiency across multiple industries:



Manufacturing

The ability to capture and analyze data in real time improves detection of defects in manufacturing processes, helping to ensure optimal efficiency. It also drives cost savings with automated tracking and managing of inventory across branches and regions using Internet of Things (IoT) sensors in warehouses and centralized inventory data management.



Supply chain

Using IoT to track goods from manufacturing to delivery provides real-time visibility of the bottlenecks in transportation and uncovers product shortages or overages during transport to avoid waste. Streamlining operations and processes is critical to realizing higher profits.



Medical

Healthcare providers can use data insights to help detect patterns and propensities for diseases and diagnoses, for example in scans to detect cancerous tissue. AI is also changing the business of medicine with new IoT devices, including wearable technology that provides real-time visibility and new patient monitoring options, allowing consumer-related businesses, for example, insurers, to more accurately predict need for medical care ahead of time.



Retail and sales

Retailers use data to improve customer interactions and drive operational and process improvements, while detecting and mitigating fraud. Customer analytics, for example, targeting customers with the next best offer, helps reduce attrition and increase sales.

Data's role

AI requires data, the more the better, and there's no shortage with the digital revolution. The challenge for most organizations is capturing and making meaningful correlations from data locked in silos across the organization, data from external sources and data being transmitted into the business in real time. Adding to this challenge is the inability for many organizations to capitalize on new sources of semi-structured, and unstructured data, driven by the extreme growth in device volumes. New sources include social media, streaming audio and video, and call logs just to name a few and each provides valuable untapped insights.

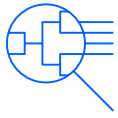
Why AI should be a strategic priority for organizations

AI systems are developed to respond with the intellectual process characteristic of humans; the ability to reason, generalize or learn from experience. In the modern business, this process includes the ability to perform behavior and sentiment analysis for an even better service, to powering comprehensive consumer insights and millions in savings for the company. In short AI can be a catalyst for success.

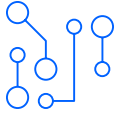
There's growing support of AI as a transformative technology that gives organizations a competitive edge, helping them gain market share. Most enterprises are taking a serious look at the value of AI to their businesses. According to a study by MIT Sloan, 85 percent of respondents view AI as a strategic priority.¹ Based on responses of over 3,000 respondents, the 2019 Gartner CIO survey reported that:

The overwhelming majority of the 500 executives surveyed agreed that data is critical to overall financial performance (90%), growth (88%), improving the customer experience (92%), and improving the employee experience (79%). Additionally, 83% agreed that data will be a competitive differentiator for their company moving forward.

Advanced analytics and AI can unlock the value in IoT data, but organizations need to consider an alternative approach



42% of European organizations are using basic analytics for asset management



Only **5%** are using AI techniques



By 2020, **100%** of effective IoT efforts will be supported by **machine learning** in some form

Figure 2: The right platform for advanced analytics and AI is critical to IoT success²

AI capability will also drive IoT success

With AI capabilities driving key business priorities, such as increased customer service, IoT would be the next step to fully utilizing real-time data. IDC predicts that by 2020, 100 percent of effective IoT efforts will be supported by machine learning in some form, allowing predictive maintenance and asset management use cases, among others, to be solved.

IBM and Cloudera for big data, analytics and AI

The IBM and Cloudera strategic partnership offers customers a direct path to AI and IoT, and is designed to provide easy access to the capabilities, scalability and economy of an advanced data platform. These capabilities include governance, a security-rich environment and the tools for data federation, advanced query and data management. The results are open source data and analytic solutions, ready for the enterprise and its future data growth.

Combined open source solutions for the enterprise

IBM and Cloudera solutions provide an industry-leading open source data platform with a full ecosystem of value adds. Together, they provide the most investment and collaboration within the open source community which, for customers, means the latest innovations are integrated into their solutions, products and services. Features and functions that make these offers enterprise ready are added to take out the experimentation and development time, saving organizations time and money. Companies can deploy, run and manage data and applications in the cloud of their choice in a security-rich environment without the risk of being locked in.

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:

- Speed and agility to innovate through open source technology
- Flexible hybrid cloud, multicloud and on-premises options for deployment
- Software support through one vendor
- Accelerated time to value, with leading industry and technology experts globally
- Vertical specialism providing in-depth knowledge of industry use cases, tailored to specific needs

IBM and Cloudera—better together

- **Leading SQL engine for complex analytic workloads**
- **#1 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**

Key tenants for IBM and Cloudera offerings

Each IBM and Cloudera solution follows 4 key tenants to help ensure that they deliver data solutions at enterprise scale:

1. **Secure and governed.** It must maintain strict enterprise data privacy, governance, data migration, and metadata management requirements across all environments.
2. **Multifunction analytics.** It must allow multiple analytic functions to work together on the same data at its source, solving pressing data challenges in a streamlined fashion.
3. **Open platform.** It must be 100 percent open source, with open compute and open storage, helping to ensure zero vendor lock-in and maximum interoperability.
4. **Hybrid and multicloud.** It must operate with equivalent functionality on and off premises, supporting all major public clouds, as well as private clouds.

Most enterprises today have large-scale data assets on premises while also seeking more agility for the business with cloud deployments. 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 on premises or in the cloud.

IBM and Cloudera offerings

Unlock the value of the open source ecosystem with Cloudera enterprise data cloud, custom Cloudera solutions, and IBM value-add solutions, products and services.

Start with an open source data platform and build a custom solution with Cloudera

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—edge analytics, streaming analytics, data engineering, data warehousing, operational analytics, data science and machine learning—all working together and operating across a choice of security-rich environments—data centers, multiple public clouds and hybrid cloud.

The current Cloudera platforms offer all the key capabilities of an enterprise data cloud—hybrid and multi-public cloud, multifunction analytics, [shared security and governance services \(SDX\)](#), and open source platforms, with choice of compute and storage. Each Cloudera offering is explored in more detail here.

An enterprise data cloud:

- **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 regulatory compliance by design**

[Hortonworks Data Platform \(HDP\)](#) addresses the complete needs of data-at-rest and data-in-motion, powering real-time customer applications and delivering real-time analytics for data scientists, analysts and developers, either on premises or in the cloud.

- Launches apps in a matter of minutes and addresses new use cases for high-performance and machine learning applications
- Enables enterprises to gain value from their data faster and smarter, in a hybrid environment
- Provides cloud storage support to store endless amounts of data in its native format, including Amazon S3 and Microsoft Azure Data Lake Storage
- Helps increase regulatory compliance, including GDPR, through full chain of data custody, as well as fine-grained auditing of events

[Cloudera Enterprise Data Hub \(EDH\)](#) is a modern data management platform that brings together multiple analytic functions, enterprise security and governance, and the flexibility to run in any operating environment—public cloud, on premises or hybrid. Analytics capabilities include stream and batch data processing, data warehousing, operational database and machine learning. [Cloudera SDX](#) spans the full platform and allows users to share data safely and aids security across workloads and environments.

EDH promotes machine learning and analytics with applications for data science, engineering, data warehousing and operational databases. It can be deployed and managed across Amazon Web Services (AWS), Google Cloud Platform, IBM Cloud™, Microsoft Azure and private networks.

- Provides multifunction analytics from the edge to AI on a unified platform that eliminates silos and speeds the development of data-driven applications
- Offers enhanced security by design with encryption, access controls, governance and lineage that provide central IT control and enable end-user self-service on production data
- Leverages hybrid and multicloud environment optimization, helping ensure analytic applications work consistently across environments and acting on data wherever it's stored
- Uses a single pane of glass for cluster administration, automation, management and security

[Cloudera Data Flow \(CDF\)](#) is a highly scalable, open source data-in-motion platform for streaming analytics that ingests, curates and analyzes data for key insights and immediate actionable intelligence. CDF is designed to process large volumes of big data, transporting it from a multitude of sources. It accommodates real-time streaming data, tracks provenance and lineage, and manages edge applications.

CDF also:

- Helps manage and secure data from the edge to the enterprise with a no-code approach to developing sophisticated streaming applications
- Reduces data integration development time with real-time data streaming at high volume and scale
- Provides out-of-the-box compliance with [Cloudera SDX](#)

[Cloudera Data Platform \(CDP\)](#) will combine the best of Hortonworks and Cloudera technologies to deliver the industry's first enterprise data cloud. Built on open source, the enterprise data cloud will support customers' objectives to avoid vendor lock-in and accelerate enterprise innovation.

CDP will initially be offered as a public cloud service and will later be offered for private cloud. It will offer a full complement of open source data management and multifunctional analytics, with the agility and elasticity of a public cloud-like experience and is designed for ease of use.

- Extends current HDP, Cloudera Data Hub (CDH) and Cloudera Enterprise Data Hub (EDH) deployments with native cloud services on two of the most popular public clouds, AWS and Azure
- Provides a single control pane to manage infrastructure and data and analytic workloads across hybrid and multicloud environments
- Extends SDX shared services to safeguard data privacy, regulatory compliance and cybersecurity threats across all cloud environments

[Red Hat Enterprise Linux](#) provides an open source operating system that lets businesses scale existing apps and roll out emerging technologies across essentially all types of cloud and on-premises environments.

- Offers multiple app versions for various users with Red Hat Smart Management
- Promotes security and compliance with Red Hat Insights, a predictive IT analytics service
- Helps develop apps faster with Red Hat OpenShift on containers' native Linux OS

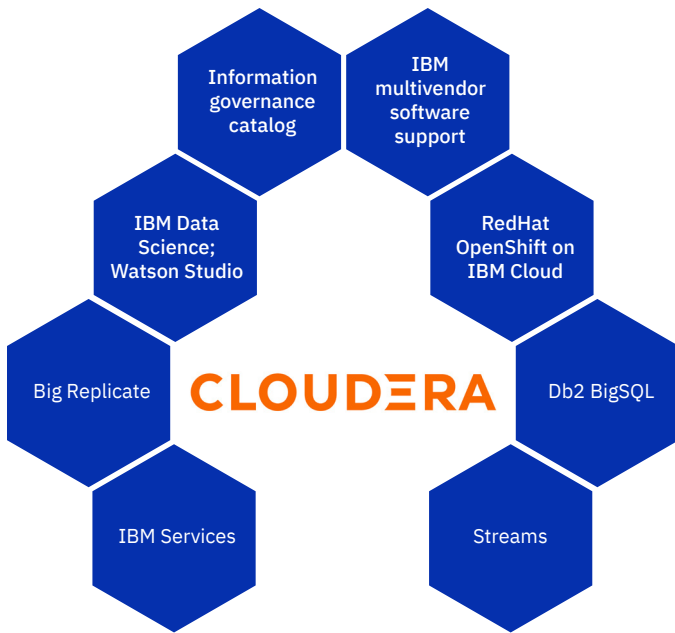


Figure 3: IBM's value-adds for Cloudera

IBM value-adds: An ecosystem of products and services

IBM provides solutions designed to augment and facilitate the enterprise readiness and efficient use of open source data lake offerings, such as those just noted.

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.

With Db2 Big SQL, organizations can connect to disparate sources, such as Hadoop Distributed File System (HDFS), relational database management system (RDMS), NoSQL databases, object stores and WebHDFS. Enjoy low-latency support for ad hoc and complex queries, high performance, robust security, SQL compatibility and federation capabilities to get the most from a data warehouse.

- Provides federated access to many RDMS sources
- Offers large data volume and extremely complex query support
- Supports low-latency, high concurrency workloads

IBM Big Replicate is active-active data replication for Hadoop across supported environments, distributions and hybrid deployments. It replicates big data from lab to production, from production to disaster recovery sites, or from ground to cloud object stores governed by the most demanding business and regulatory requirements.

- Enables real-time, multidirectional data replication between Hadoop clusters
- Offers source application to Hadoop replication
- Provides high availability disaster recovery (HA/DR) with virtually zero RTP
- Replicates from on premises to cloud and cloud to on premises

IBM Watson® Studio drives accelerated machine and deep learning workflows required to tap into data assets and inject predictions needed for AI initiatives. Use Watson™ Studio to build and train AI and machine learning models, prepare and analyze data with automated deep learning and enhanced visual modeling.

- Enhances collaboration with community and social features
- Combines the best of open source and IBM value-adds to create state-of-the-art data products
- Offers built-in learning and advanced tutorials

IBM unified governance and integration helps ensure that data lakes contain clean, relevant data from structured and unstructured sources that can easily be found, accessed, managed and protected. The platform the data resides on is security-rich and reliable. Data that comes into the data lake is properly cleaned, classified and protected in timely, controlled data feeds that populate and document it with reliable information assets and metadata.

- Ingests structured and unstructured data from essentially all data sources in real time
- Uses an intelligent metadata catalog and allows users to better understand the data visually
- Protects the integrity and reliability of data through governance policies, helping to keep it compliant and audit ready

Why IBM and Cloudera are better together for businesses

The partnership between IBM and Cloudera promises to be a more significant boon to organizations' data management than either company could have provided on its own. A few of the most important benefits of a combined IBM and Cloudera solution are listed here.

- **Drive speed of innovation with the flexibility of open source**
IBM, with the acquisition of Red Hat and in partnership with Cloudera, leads the pack with the largest number of contributions to the open source community. IBM and Cloudera are dedicated to investing in large open source projects and installs, driving enterprise-grade solutions with added security, increased availability and with interoperability across vendors for freedom of choice.
- **Benefit from freedom of choice with on-premises, hybrid and multicloud options**
Depending on the types of workloads and applications supported, today's enterprises need choice and integration between data platforms. Through the IBM and Cloudera partnership, including the enterprise data cloud, they drive integrated options that include building and modernizing existing on-premises infrastructure, as well as offering next-generation hybrid cloud and multicloud platforms.
- **Save time, money and frustration with one source for open source software support**
Whether using community editions, commercial products, individual packages or a complex software stack, IBM and Cloudera provide multivendor support. A [new Forrester study](#) confirms clients can save as much as 25 percent using multivendor support services through one vendor.
- **Speed time to value through global services with over 100,000 experts and consultants**
IBM and Cloudera's services cover business, technology, industry design and redesign of processes, applications, and on-premises and cloud infrastructures. Their end-to-end capabilities transform companies whether they adopt AI and machine learning; build, govern and driving analytic insights from data lakes; or connect clouds with traditional infrastructures.

- **Benefit from IBM and Cloudera's comprehensive knowledge of industries and technology**
IBM and Cloudera have experience with enterprise organizations across virtually every industry. They can build an integrated vision and plan for organizations, designed to solve vertical-specific challenges. In addition, their breakthrough technologies are transforming industries with new growth opportunities and strategies to compete and win.
- **Drive value from IBM and Cloudera's innovation and IP creation**
IBM inventors received a record 9,100 patents in 2018, marking the company's 26th consecutive year of US patent leadership and crossing the 110,000-patent milestone. IBM led the industry in the number of AI, cloud computing, security and quantum computing-related patent grants.

Why IBM?

IBM builds data management products with the AI, hybrid, and multicloud future in mind. Its enterprise-grade solutions are designed for robust integration across security-rich environments. To assist with your data lake needs, IBM provides value-added offerings that bring these same benefits to Hadoop implementations like those created by Cloudera.

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 and Cloudera partner page](#).



© Copyright IBM Corporation 2019

IBM Corporation
New Orchard Road
Armonk, NY 10504

Produced in the United States of America
October 2019

IBM, the IBM logo, ibm.com, Db2, IBM Cloud, IBM Services, IBM Spectrum, IBM Watson, Power, and Watson 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 trademarks is available on the web at "Copyright and trademark information" at www.ibm.com/legal/copytrade.shtml.

Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.

Microsoft and Azure are trademarks of Microsoft Corporation in the United States, other countries, or both.

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.

Statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.

1 Sam Ransbotham, David Kiron, Philipp Gerbert and Martin Reeves. "Reshaping Business with Artificial Intelligence." MIT Sloan Management Review, September 6, 2017. <https://sloanreview.mit.edu/projects/reshaping-business-with-artificial-intelligence/>

2 According to an IDC Infobite series sponsored by Cloudera. September 2018. <https://2xbbhxc6wk3v21p62t8n4d4-wpengine.netdna-ssl.com/wp-content/uploads/2014/08/Hortonworks-Infobites-EMEA44205718-V6.pdf>