

CLOUDERA

EBOOK

Modernize Your Analytics and AI with Cloudera



Table of Contents

A Fragmented Data Landscape	3	Enable Self Service Access to Data —	
Finding Insight in the Data Haystack	3	While Enforcing Governance and Control	10
The True Hybrid Platform: A New Approach to Data Management	4	Take an Integrated Approach to Security and Governance with Shared Data Experience (SDX)	11
Introducing Cloudera	4	One Point of Control	12
Achieve the Full Potential of the Data Lifecycle	5	Replication Manager	12
Manage and Secure the Data Lifecycle	6	Workload Manager	12
		Management Console	12
Cloudera Services for Multi-Function Analytics and AI	7	Accelerate Analytics and AI with a Complete True Hybrid Platform Solution	13
Data Flow and Streaming	7	The Benefits of Cloudera Are Clear	14
Data Engineering	7	Best-in-Class Analytics and AI	14
Data Warehouse	7	Scaling Security Becomes a Non-Issue	14
Operational Database	7	A Holistic View Across Hybrid Environments	14
AI	7	Greater IT Control — and Reduced Complexity	14
Hybrid Cloud for Interoperability Across Environments	8	Resources Optimized for Each Workload	14
Better Control Resources	9	A Foundation Built on Open Source	14
Deploy with Greater Flexibility and Agility	9	Try Cloudera	15
Move Workloads Anywhere with Ease	9		
Ensure Your Workloads Stay Compliant	9		

A Fragmented Data Landscape

Your organization's ability to transform, innovate, and compete depends on how effectively your team utilizes every bit of data at your disposal. By 2022, 90% of organizations will view information as a critical enterprise asset. Analytics will also be considered an essential competency.¹

But while the importance of data is being recognized, accessing and using that data in the myriad of ways you require is far from a simple matter. If your organization is like most, then you:

- Receive a greater amount of data from a rising number of sources
- Are increasingly capturing data at the edge, where it must be ingested and analyzed in real-time
- Keep your data in a myriad of systems that are difficult to integrate
- Are concerned with the security of data accessed across the organization
- Have an average of two public clouds and two private clouds in use²

Finding Insight in the Data Haystack

As long as data remains locked within diverse environments making it difficult to access or consume — then the value of that data is greatly diminished.

That's why overcoming the fragmented data landscape to get the full value of multiple data, analytics, and AI workloads requires a new approach. One that gives you full control of the entire data lifecycle across every environment and system and enables your business teams to make faster, smarter data-driven decisions.

90%

By 2022, 90% of organizations will view information as a critical enterprise asset. Analytics will also be considered an essential competency.¹

¹ "Why Data and Analytics Are Key to Digital Transformation," Christy Pettey, Gartner, March 8, 2019.

² Flexera 2020 State of the Cloud Report.

The True Hybrid Platform: A New Approach to Data Management

The key to eliminating data silos and realizing the full potential of your diverse data sets to drive business value is a true hybrid platform. A true hybrid platform is a big data platform that helps you control and manage the deluge of data across all of your environments (edge, private cloud, and multiple public clouds) and throughout the data lifecycle to gain deeper insights. With the power of a true hybrid platform, you can adopt the most valuable and transformative business use cases by realizing the full analytical potential of all your data.

To see what a True Hybrid Platform can do for you [click here](#).

Introducing Cloudera

Cloudera is a first-of-its-kind true hybrid platform. With its open, hybrid data architecture, it powers data-driven decisions and predictive actions by seamlessly connecting your data across your fragmented IT landscape. It helps you secure and govern the entire data lifecycle — while still providing your users quick, easy access to the critical data needed for predictive insights and guided decision-making, based on a foundation of open source.

With Cloudera you can embrace the seemingly opposing forces (diverse environments, security and governance demands, and evolving business requirements) that otherwise work against each other in the sharing and analyzing of data.

Power Best-in-Class Analytics and Machine Learning.

Cloudera eliminates the complexities of managing petabytes of data and multiple workloads, clearing the way for your organization to collect, enrich, report, serve and model data for deeper insights across the business.

Get the Most from Hybrid Cloud and Multi-Cloud Environments.

Cloudera unifies any on premises, hybrid, and multi-cloud environment into a single platform that can support an unlimited number of concurrent workloads with a consistent experience.

Take Control With Security and Governance.

Cloudera ensures security and governance of your data and analytics lifecycle with built-in policies to manage access, activity, and usage of shared data. Your organization gains greater agility to act quicker on insights while keeping its data safe and ensuring regulatory compliance.

Simplify Management With Faster, Easier Administration.

Cloudera gives you a holistic view of the full data lifecycle through a single pane of glass that simplifies infrastructure management tasks.

The Mission-Critical Advantages of Cloudera

- Best-in-class analytics and AI for unlimited use cases
- Single platform unifying data across hybrid and multi-cloud environments
- Built-in security that supports virtually unlimited data and thousands of users
- Scalability, performance, and cost efficiency for all workloads

Achieve the Full Potential of the Data Lifecycle

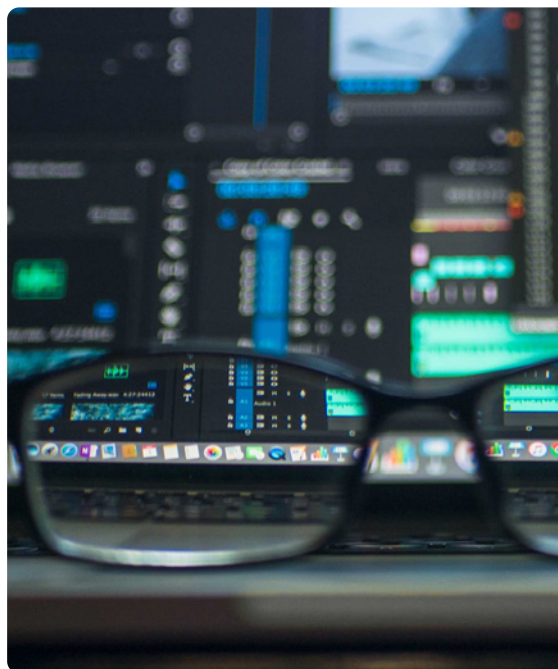
Too often, analytic workloads run independently in their own silos because analytic engines and data science tools weren't designed to work together.

Each silo has its own security, governance, and control needs. They also come with their own schema and data catalog. And the arduous task of keeping these silos in sync falls to the IT team. So, as a result, you get increased operational costs and overhead and your data users are hindered by siloed systems that don't allow multiple analytics workloads to run against the same data set.

The most compelling use cases for data require multiple analytic engines and processes working together, sharing data regardless of where that data lives.

Cloudera runs on any infrastructure — from on premises private cloud to multiple public clouds, or in any combination of these to form a hybrid cloud. It empowers users to work on the same data sets, with potentially different tools, while also adhering to compliance regulations and without compromising security. With self-service

access granted by Cloudera, users and groups navigate the full data lifecycle and apply analytics directly where the data lives, interacting with that data at its source to uncover the value it holds. With Cloudera, your organization transforms into a data-driven enterprise by taking control over the entire data lifecycle, from edge to AI.



“

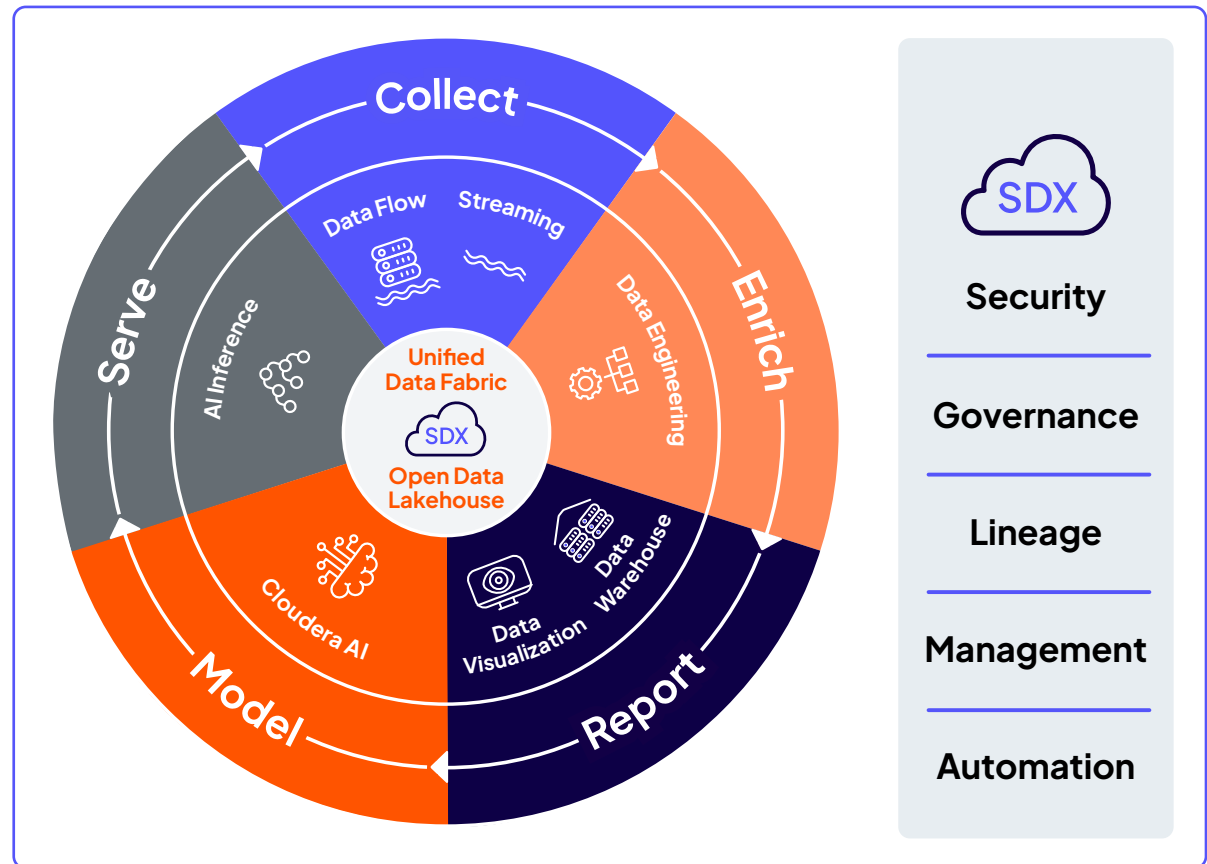
Cloudera running on AWS helps us manage our data, interpret analytics, and put those insights into action with customer segmentation so we can better tailor how our customers discover, buy and use our services.”

Gil Genio
Chief Technology and Information Officer,
Chief Strategy Officer, Globe Telecom

Read more [here](#).

Manage and Secure the Data Lifecycle

Cloudera gives your organization control of its data across the entire data and analytics lifecycle. You can eliminate data silos while securing, governing, managing, and automating the entire lifecycle process.



Cloudera Services For Multi-Function Analytics and AI



Data Flow and Streaming

Cloudera has a highly scalable, real-time streaming analytics engine that ingests, curates, and analyzes data for key insights and immediate actionable intelligence.



Data Engineering

Cloudera enables you to enrich, transform, and cleanse data to easily create, execute, and manage end-to-end data pipelines.



Data Warehouse

Cloudera is auto-scaling, highly concurrent, and cost-effective because it ingests large-scale data from anywhere (structured, unstructured, and edge data).



Operational Database

Cloudera has a real-time, scalable database, giving you the ability to serve traditional structured data alongside unstructured (and semi-structured) data within a unified platform.



AI

Cloudera provides data science teams simple and intuitive access to business data and the tools and computing resources required for end-to-end machine learning workflows.

Cloudera success stories

23PB

of unstructured data
used to bring new drugs
to market

A pharmaceutical company leverages fast, holistic analytics and ML to accelerate FDA approvals and bring new drugs to market faster.

Read more [here](#).

Hybrid Cloud for Interoperability Across Environments

Many organizations are utilizing the benefits of multiple public clouds in conjunction with their private cloud. This combination, a hybrid cloud environment, offers flexibility, agility, performance, and cost-efficiency. It gives you the freedom to decide where each of your workloads will run, based on each workload's needs. For example, you can run intermittent/batch workloads in public cloud while running more consistent workloads in your private cloud.

Oftentimes, mission-critical data can be found in multiple places throughout hybrid cloud environments. And while it's critical to analyze that data, there's a catch: data silos lead to incomplete and inaccurate analysis. This is due in part to the fact that hybrid cloud, on its own, lacks a centralized way to manage workloads across different environments. Ensuring data governance and security across your hybrid cloud is also painstakingly manual and time-consuming, leading to compliance risks.

You get the greatest value from your data when you're able to run multiple analytics and machine learning models against the same diverse data sets for all kinds of business use cases. But workloads need a way to share data sets — and to do so without adding greater complexity to managing your hybrid cloud.



Cloudera success stories

40%

increase in sales force efficiency

A consumer packaged goods company maintains a central repository of multi-structured data across apps and external sources for a single source of insight.

Read more [here](#).

Cloudera makes all of this possible. It removes the complexities of managing petabytes of data and multiple workloads across private and public cloud environments, clearing the way for you to better utilize hybrid cloud for data-centric workloads. With Cloudera, you can:



Better Control Resources

Provision resources to rapidly burst a workload to the cloud if it becomes constrained by on-premises cluster capacity.



Deploy With Greater Flexibility and Agility

Quickly deploy new use cases in public cloud and move to private cloud once performance and growth are understood.



Move Workloads Anywhere with Ease

Migrate select data workloads to different clouds without developers having to rewrite code.



Ensure Your Workloads Stay Compliant

Meet data locality rules and regulations within your data center(s) with a cloud-native data platform.

58%

of enterprises are moving toward a hybrid IT environment.³

³ "The Future Is Hybrid" presentation, 451 Research, 2020.

Regardless of how you need to use hybrid cloud, Cloudera ensures your workloads are able to run at their highest possible performance and cost efficiency.

Start your free test drive at [Cloudera.com](https://cloudera.com).

Enable Self Service Access to Data — While Enforcing Governance and Control

You want to give employees, partners, and customers access to your data for a wide array of business use cases. But without a true hybrid platform, maintaining security, governance, and control across siloed systems is a losing battle. Your private and public cloud environments all have different tools, capabilities, and processes for securing data. However, typical hybrid cloud management doesn't let IT and information security teams apply blanket policies across these environments. Additionally, your IT team is likely forced to work within the confines of legacy systems and proprietary lock-in.

In an effort to grant access to data while working under these conditions, your IT team must create and manage multiple copies of the same data sets and stay ahead of the barrage of access requests. This is complex, manual, and costly. Meanwhile, your IT team still needs to maintain compliance. The only way to do so effectively is to restrict access to the valuable data that users need. The results: delayed business insight, increased operational overhead, and blind spots that only create greater security risks.

If you want to give more users the benefits of analytics and ML for better decision-making — and to do so securely with the appropriate restrictions as to the data they can see and use — you need a consistent, shared governance, and security model that enables a variety of analytics across private and multiple public cloud environments.



Cloudera success stories

10X

growth in data volume
handled with ease

A global logistics uses Cloudera architecture for increased scalability and accessibility while better ensuring compliance and exploring new revenue opportunities.

Read more [here](#).

Take an Integrated Approach to Security and Governance with Shared Data Experience (SDX)

SDX is a fundamental layer of the Cloudera architecture. This always-on service gives you control over the security and governance of every deployment across any environment — all through a single pane of glass.

Apply Once. Enforce Everywhere.

Through SDX, Cloudera lets you define data access policies once, and then enforce those policies seamlessly throughout workloads and the hybrid and multi-cloud environments where they're deployed. This eliminates the need to create multiple copies of the same data, reducing both cost and complexity.

Track and Audit Data Access and Usage.

With SDX, you can quickly identify and manage sensitive data and effectively address regulatory requirements with unified, platform-wide operations. You can also efficiently prove compliance throughout the complete data lifecycle across all analytics and deployments.

Give Analysts Easy Access to Data Sets.

SDX powers a robust data catalog for searching, organizing, securing, and governing data within the true hybrid platform. This gives data analysts a single-pane view, where they can manage and discover data — which is enhanced with metadata — they are authorized to access. By applying operational, social, and business context, SDX creates trusted, reusable data sets that data analysts can find quickly, so they can focus on critical work that much faster.

The Data Catalog in SDX is used by data stewards to browse, search, and tag the content of a data lake, create and manage authorization policies (by file, table, column, row, and so on), identify what data a user has accessed, and access the lineage of a particular data set.

90%

By 2023, 90% of the world's top 500 companies will have converged analytics governance into broader data and analytics governance initiatives.⁴

⁴ "100 Data and Analytics Predictions Through 2024," Gartner, March 20, 2020.

One Point of Control

You can manage, monitor, and orchestrate all Cloudera services through a single console for data management, workload analysis, and data movement capabilities.



Replication Manager

Copy, migrate, take snapshots, and restore data between environments within Cloudera. Administrators and data stewards are able to move, copy, backup, replicate, and restore data in or between data lakes. This can be done for backup, disaster recovery, or migration purposes, or to facilitate dev/test in another virtual environment.



Workload Manager

Get the tools needed for an in-depth understanding of your workloads. Workload Manager lets you view relevant statistics about your jobs and information that you can use to either troubleshoot failed jobs or optimize slow ones. This means you can ensure the best possible performance of every workload in Cloudera.



Management Console

Streamline and simplify the administration, management, and provisioning of users, environments, and services all through one dashboard. The Management Console lets you work from a single pane of glass to manage clouds, administer cloud and on premises resources, and govern users and their access.



Cloudera success stories

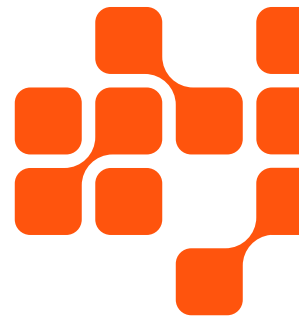
\$200K

expected reduction in
monthly OPEX

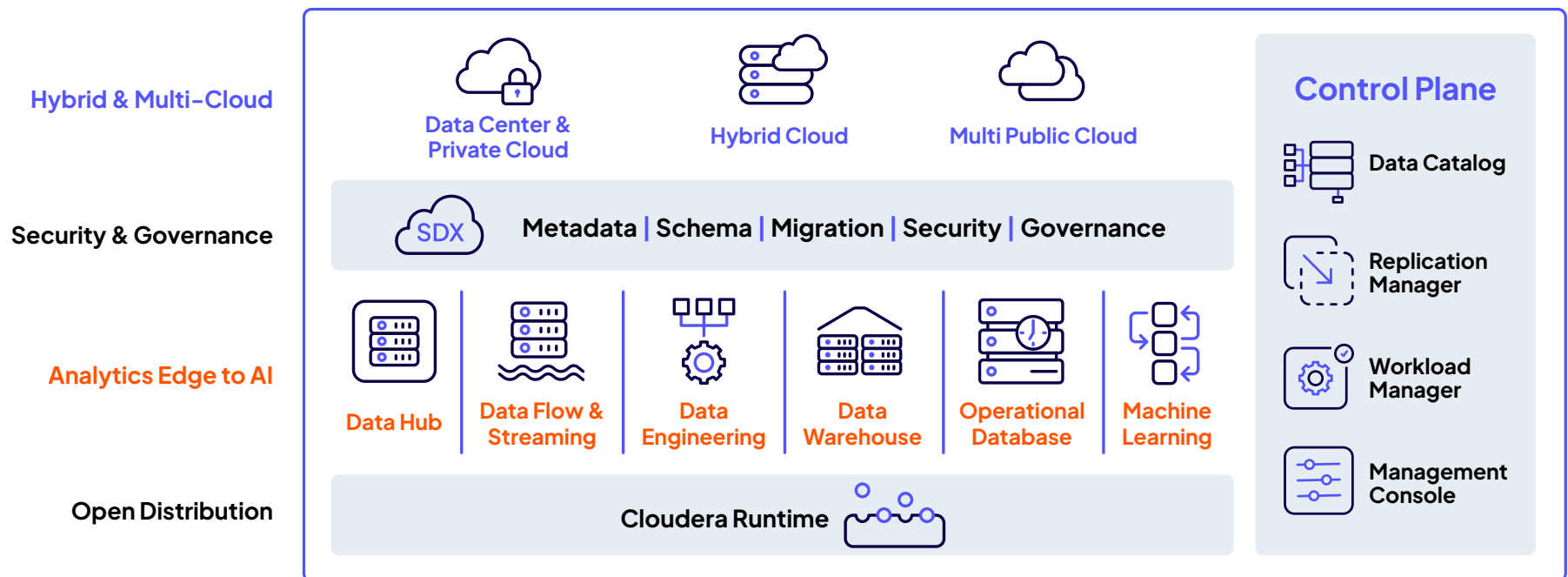
A global consumer credit card company better manages its sensitive credit data while adhering to regulation and minimizing cost.

Read more [here](#).

Accelerate Analytics and AI with a Complete True Hybrid Platform Solution



With Cloudera, you have control over the entire data and analytics lifecycle — in any environment, with complete security and control.



The Benefits of Cloudera Are Clear



Best-in-Class Analytics and AI

Cloudera brings you unmatched data management and analytics capabilities that can be applied everywhere across your entire hybrid environment — without having to compromise ease of use, capabilities, performance, and more anywhere along the way.



Scaling Security Becomes a Non-Issue

Through its SDX layer, Cloudera makes it easy to scale up to thousands of users, all working within different policies as determined by your governance team. This ensures your organization benefits from data-driven decision-making without increasing the risk of non-compliance.



A Holistic View Across Hybrid Environments

Cloudera is a containerized, scalable, and portable service that lets you deploy between different workloads and clouds via a centralized console. You can manage, monitor, and orchestrate all of your deployments from a single pane of glass — whether those deployments are in your private cloud in multiple public clouds. You can also easily create, provision, and destroy services.



Greater IT Control — and Reduced Complexity

As Cloudera advances in its technical capability, it also becomes more intuitive to use and manage. IT teams control hybrid environments more efficiently through consolidated visibility, while users gain self-service access to critical data and use their own choice of tools.



Resources Optimized for Each Workload

Cloudera has been architected to increase efficiency while simultaneously controlling costs. It separates compute and data resources giving you more deployment options and the ability to scale your workloads to petabytes of data, at the exact levels of performance required.



A Foundation Built on Open Source

Cloudera frees you from vendor lock-in and lets you take advantage of an open source ecosystem that can be integrated with your existing systems and applications. You can access both highly customizable services and provide ready-made end-user experiences on a 100% open platform.



Find the Needle in the Data Haystack. Try Cloudera now.

Business transformation demands innovative data analytics and AI. With Cloudera, you can now meet that demand — with a petabyte-scale, hybrid data architecture that can flex to use private and public clouds. Let your data take you farther and faster.

5-Day Free Trial

- Unlock actionable insights in minutes.
- Experience Cloudera through common use cases that also introduce you to the platform's fundamentals and key capabilities.

trycloudera.com

About Cloudera

Cloudera is the only true hybrid platform for data, analytics, and AI. With 100x more data under management than other cloud-only vendors, Cloudera empowers global enterprises to transform data of all types, on any public or private cloud, into valuable, trusted insights. Our open data lakehouse delivers scalable and secure data management with portable cloud-native analytics, enabling customers to bring GenAI models to their data while maintaining privacy and ensuring responsible, reliable AI deployments. The world's largest brands in financial services, insurance, media, manufacturing, and government rely on Cloudera to be able to use their data to solve the impossible — today and in the future.

To learn more, visit [Cloudera.com](https://cloudera.com) and follow us on [LinkedIn](#) and [X](#). Cloudera and associated marks are trademarks or registered trademarks of Cloudera, Inc. All other company and product names may be trademarks of their respective owners.



Cloudera, Inc. | 5470 Great America Pkwy, Santa Clara, CA 95054 USA | cloudera.com

© 2025 Cloudera, Inc. All rights reserved. Cloudera and the Cloudera logo are trademarks or registered trademarks of Cloudera Inc. in the USA and other countries. All other trademarks are the property of their respective companies. Information is subject to change without notice. EB_020_V3 January 27, 2025