

WHITE PAPER

# Cloudera Empowers Enterprise Data, AI, and Analytics on AWS

Cloudera Cloud on AWS Delivers Powerful Capabilities  
to Enterprises Across Industries

By Stephen Catanzano, Senior Analyst  
Enterprise Strategy Group

January 2025



# Contents

<b>Abstract .....</b>	<b>3</b>
<b>Modern Data Platforms Empower Organizations .....</b>	<b>3</b>
<b>Data Empowerment in Organizations Continues to Climb .....</b>	<b>4</b>
<b>The Motivation for Modernizing .....</b>	<b>5</b>
<b>Cloudera .....</b>	<b>6</b>
<b>Benefits of Migrating to Cloudera on AWS .....</b>	<b>7</b>
<b>Your Data Is Your AI Differentiator .....</b>	<b>10</b>
<b>Unique Benefits of Cloudera on Amazon.....</b>	<b>10</b>
<b>Conclusion .....</b>	<b>11</b>

# Abstract

The rapid evolution of data-driven decision-making demands a robust, scalable, and innovative platform to enable enterprises to harness the full potential of their data. This white paper explores the transformative advantages of migrating to Cloudera on Amazon Web Services (AWS). By combining Cloudera's advanced data management, AI, and analytics capabilities with AWS's extensive cloud infrastructure and services, organizations gain access to a robust joint ecosystem that accelerates time-to-insight, enhances operational efficiency, and ensures scalability and reliability.

This paper highlights how Cloudera's hybrid and multi-cloud data platform seamlessly integrates with AWS to deliver a unified environment for modern data architecture, empowering enterprises to optimize workloads, enable advanced analytics, and deploy AI-driven solutions at scale. With a focus on joint capabilities such as enhanced security, compliance, and performance optimization, this paper provides a compelling case for leveraging the synergy between Cloudera and AWS to enable new business agility and innovation. Cloudera not only acts as a platform as a service (PaaS) but also simplifies the migration from on-premises to a public cloud environment. This process resembles an engineered replatforming, eliminating the need for extensive re-architecting, refactoring, or rebuilding a new data platform from scratch—tasks that add no value and consume significant time. With Cloudera, customers and system integrators gain access to a turnkey solution for data and AI use cases, ready to unlock insights and value from customer data.

## Modern Data Platforms Empower Organizations

With data at the core of remaining competitive and aiding the future success of an organization, the data platform becomes increasingly important. Modern data platforms are helping organizations think about their data from an end-to-end perspective. They are integral to building the trusted data source used as the foundation for data-driven organizations. AI and analytics are the intelligence fuel in every organization. Surprisingly, research from Informa TechTarget's Enterprise Strategy Group research found that 75% of organizations reported not having a mature platform, with 58% in progress, 5% beginning, and 12% not started (see Figure 1).<sup>1</sup>

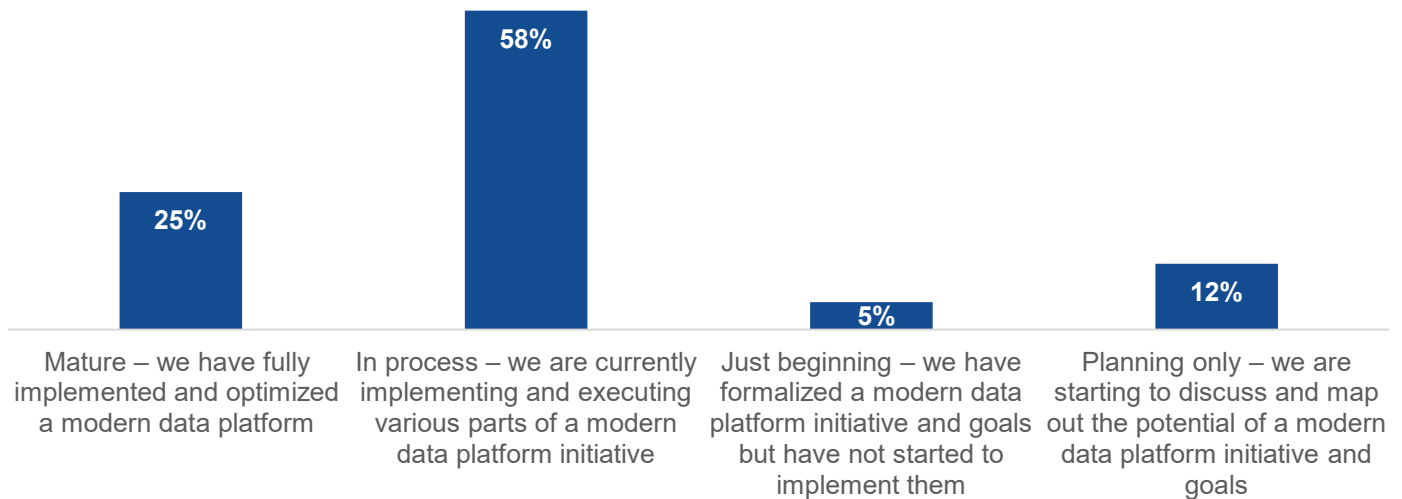
Migrating to a modern data platform such as Cloudera is quickly growing in importance to empower the use of data, AI, and analytics.

---

<sup>1</sup> Source: Enterprise Strategy Group Research Report, [Data Platforms: The Path to Achieving Data-driven Empowerment](#), July 2023.

**Figure 1. Modern Data Platforms**

**Which of the following best describes your organization's modern data platform initiatives? (Percent of respondents, N=354)**



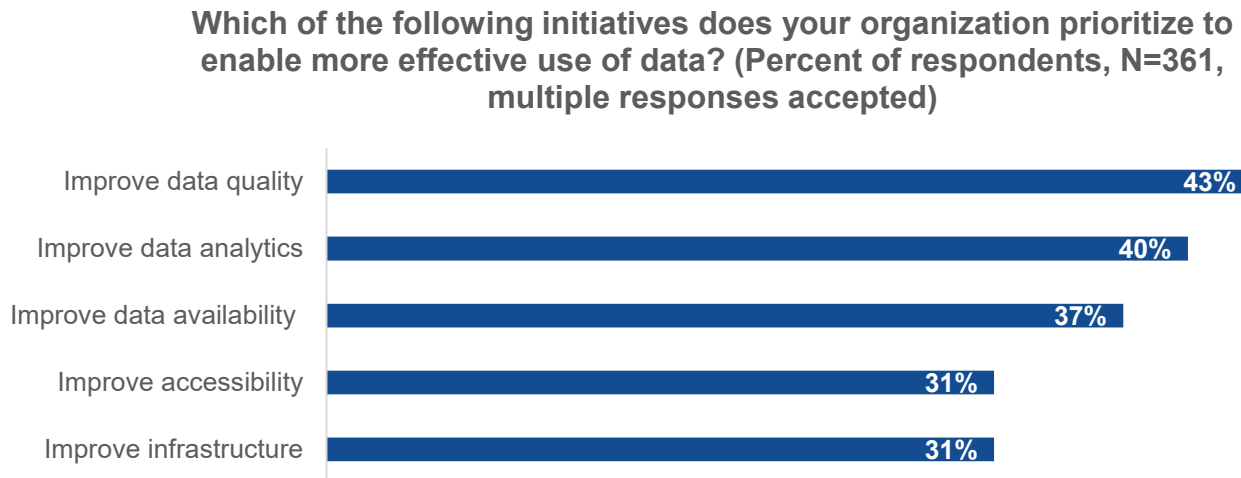
Source: Enterprise Strategy Group, a division of TechTarget, Inc.

## Data Empowerment in Organizations Continues to Climb

Organizations are rapidly attempting to democratize the use of data across their organization and empower their workforce with accurate and trusted data to drive the business forward. Enterprise Strategy Group asked research participants what their data usage priorities were focused on to enable users to be more effective and found that 43% prioritized improving data quality as a top priority. This most commonly cited response was followed by improving data analytics (cited by 40%), improving data availability (37%), improving accessibility of data (31%), and improving infrastructure to manage data (31%) (see Figure 2).<sup>2</sup> Enterprise Strategy Group research also found that 84% of organizations saw an increase in end-user access to their data in the past year and that 44% of line-of-business stakeholders only somewhat trusted the accuracy of the data given to end users for decision-making.<sup>3</sup>

<sup>2</sup> Source: Enterprise Strategy Group Research Report, [The State of DataOps: Unleashing the Power of Data](#), January 2024.

<sup>3</sup> Ibid.

**Figure 2. Top Priorities for Enabling End Users to Use Data More Effectively**


*Source: Enterprise Strategy Group, a division of TechTarget, Inc.*

It's clear that organizations understand that leveraging data from across the organization to empower employees is crucial for long-term success. The goal of being data-driven is clear, but achieving it involves excellence in data management, using AI and generative AI as accelerators, and building a foundation with trusted data. As organizations democratize, trust—including accuracy, timeliness, quality, and governance—becomes part of the foundation of data readiness.

## The Motivation for Modernizing

Organizations are under pressure to prepare for the future and build resilient, data-driven companies. Many organizations have shifted from the idea that data is part of their business to it now *being* the business. This shift is driven by market pressures from the board level, consumers, investors, suppliers, and senior-level managers, as the reality of AI is here and no longer a future consideration. However, for effective AI applications, data must be relevant, high-quality, and timely. This is one of the core challenges organizations must meet to achieve the desired outcomes. The good news is that the goals are clear, with the challenges focused on the data, tools, and processes needed to achieve them. Some of the motivations to implement AI across organizations include:

- **Accelerating growth:** Intelligent data drives faster decision-making, resulting in transformative growth.
- **Streamlining operations:** Innovation happens faster when processes are streamlined.
- **Improved customer engagement:** Higher customer satisfaction, increased retention, higher average sales, and loyalty all help to keep businesses afloat in a competitive environment.
- **Competitive advantage:** Staying ahead of the competition is achieved with differentiation as well as innovation.



### Market Insight

79% of organizations said they must use AI in business- and mission-critical processes to compete better.



AI helps to achieve these goals. As organizations strive to modernize their applications, they are looking to integrate AI in order to position themselves for the future. In a recent Enterprise Strategy Group research survey, we asked organizations about their plans for AI, and 79% said they must use AI in business- and mission-critical processes to compete better.<sup>4</sup>

## Cloudera

Organizations that migrate to Cloudera are enabled on a true hybrid platform for data, analytics, and AI (see Figure 3). Cloudera empowers global enterprises to transform diverse data across public and private clouds into actionable insights. Through its open data lakehouse architecture, Cloudera delivers scalable, secure data management and portable cloud-native analytics, enabling organizations to deploy generative AI models while safeguarding privacy and ensuring responsible AI practices. Organizations on Cloudera and AWS have the advantage of leveraging a large ecosystem of joint integrations as a high value-added benefit. A few of the core features of Cloudera that enable organizations to drive actionable insights include:

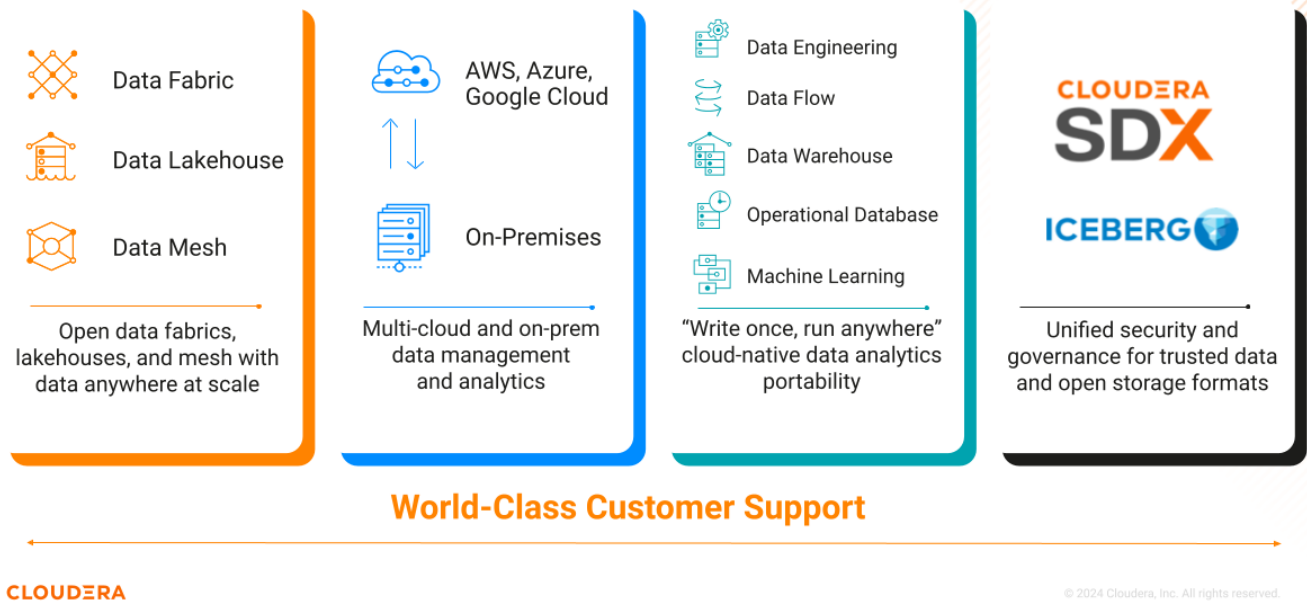
- **True hybrid:** Cloudera provides a unified, consistent, and flexible data platform that seamlessly operates across on-premises, public cloud, and private cloud environments. This capability empowers enterprises to build and manage modern data architectures.
- **Unified governance:** With a single centralized control plane, Cloudera customers control and manage data policies, lineage, data catalog, and everything needed for achieving a true end-to-end data sovereignty. This is highly critical for customers migrating workloads to AWS but who still need to keep highly sensitive data on premises or in more than one cloud—and therefore need federated end-to-end governance of their data set.
- **Open data lakehouse:** Cloudera’s open data lakehouse design bridges the gap between data lakes and data warehouses and provides scalable and secure data storage for both structured and unstructured data. It enables businesses to easily collect and visualize data in hybrid environments without compromising scalability and vendor lock-in.
- **Self-service analytics:** Cloudera offers self-service analytics, enabling business users and data scientists to view, visualize, and interpret data in their own ways. This data democratization drives faster insights and gives teams the power to make informed decisions without heavily investing in IT. Cloudera’s data platform is natively PaaS, so all the data services are offered with a “no code” approach, and this self-service cloud experience can be uniquely provided not just in the public cloud but also on premises.
- **Cloudera AI:** When you have trusted data running on a high-performance platform, an organization can accelerate its path to using enterprise data to create unique AI-powered experiences. Cloudera provides a fast route to achieving trusted and secure generative and predictive AI. Facilitating effortless integration with any language or foundation model alongside the latest AI innovations ensures investments remain agile against the rapidly evolving technology landscape.
- **Data transformation and pipelines:** In addition to powerful data transformation and pipeline tools, Cloudera automates and simplifies data preparation for organizations’ analysis. From ingest to analysis, the platform facilitates many different data flow scenarios, delivering high-quality data fit for big data and AI use cases.
- **Scale and performance:** Cloudera aims to scale naturally, meeting the needs of even the largest businesses with vast amounts of data. Its hyper-performance design enables organizations to analyze and consume real-time data to aid mission-critical processes and rapid decision-making.

---

<sup>4</sup> Ibid.

Figure 3. Cloudera True Hybrid Platform

## The **Only** True Hybrid Platform for Data, Analytics, and AI



Source: Cloudera

As shown in Figure 3, Cloudera is built on four key pillars. First, it supports any modern data architecture without requiring additional integration. Second, it enables seamless data management and analytics across all infrastructures, whether multi-cloud or on-premises. Third, it ensures effortless movement of data and workloads between environments, letting applications and use cases be written once and run anywhere. Finally, Cloudera provides unified security and governance. This infrastructure is critical to long-term success in AI and digital transformation. Apache Iceberg, the popular modern data storage technology, also facilitates Cloudera's open data lake.

## Benefits of Migrating to Cloudera on AWS

There are many reasons why organizations look to migrate to Cloudera on AWS. Some of the main use cases and benefits include:

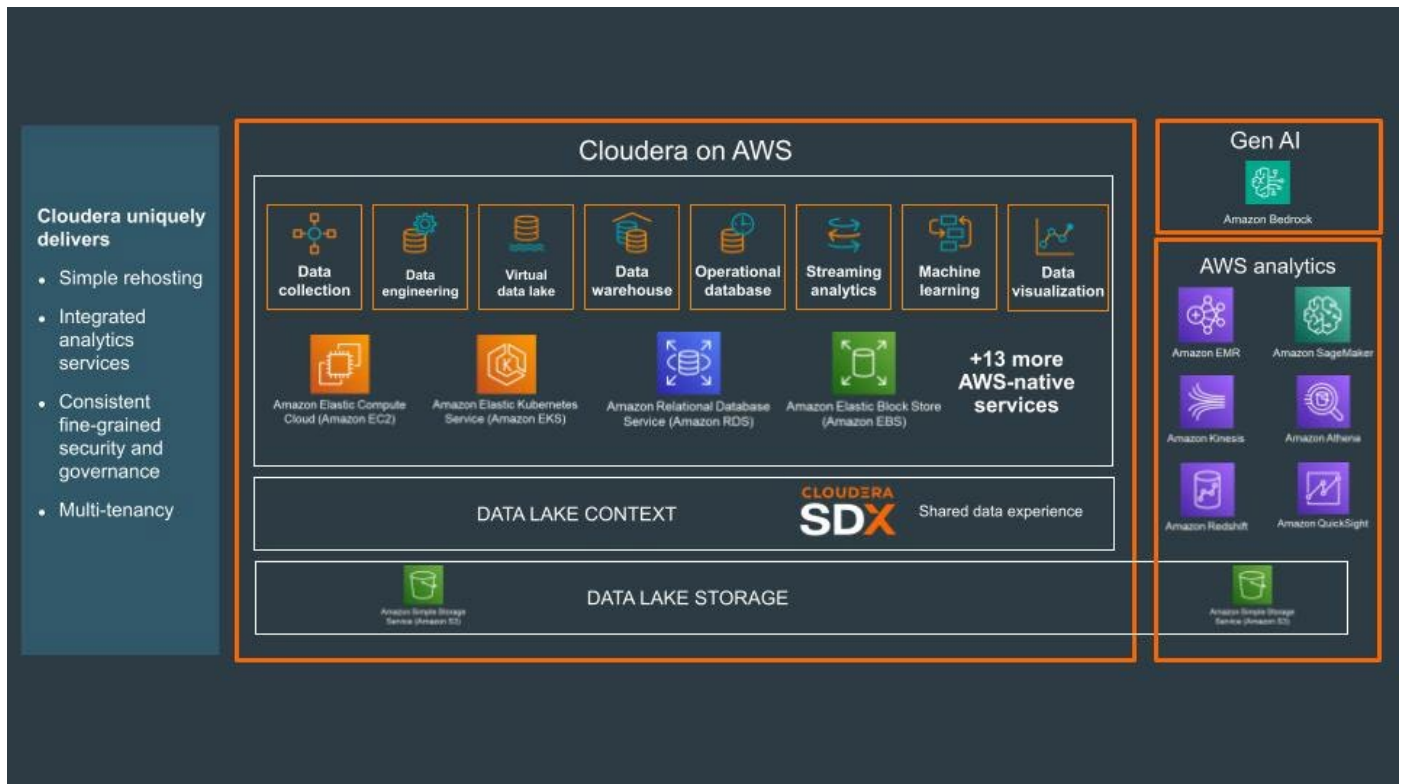
- **Unified data management:** Cloudera on AWS provides a unified platform for managing both structured and unstructured data, enabling seamless data integration and management across hybrid environments.
- **Cost efficiency and scalability:** Cloudera Data Lakehouse on AWS enables businesses to scale their data operations dynamically, optimizing costs using AWS's pay-as-you-go model and reducing the need for overprovisioning.



- **Security and compliance:** Cloudera Data Lakehouse features robust security and compliance, which are enhanced by AWS’s global security standards, ensuring data protection and regulatory compliance.
- **Enhanced analytics and AI:** Organizations can leverage Cloudera to perform advanced analytics and deploy AI models efficiently, utilizing AWS’s scalable infrastructure to handle large data volumes and complex computations.

Organizations can accelerate AI with Cloudera and AWS to offer customers cutting-edge analytics and AI tools (see Figure 4). This integration provides high performance, savings, and the flexibility to use the entire suite of AWS products and Cloudera’s hybrid data infrastructure. As businesses scale, traditional data infrastructure often struggles to meet the demand for seamless data integration, advanced analytics, and real-time insights. For organizations considering migrating to Cloudera on AWS, the opportunity to modernize their data strategy becomes clear. Organizations can tap into a unified, flexible, and future-ready platform by uniting Cloudera’s robust data, analytics, and AI capabilities with AWS’s scalable, secure, and innovative cloud-native services.

Figure 4. Cloudera and AWS



Source: Cloudera



Some of the core benefits of the Cloudera and AWS relationship include:

- **Seamless integration with AWS services:** With 18 AWS native services interoperable with Cloudera, organizations can design a single, scalable data ecosystem. With solutions such as Amazon S3 for safe, scalable data lake storage and Amazon Elastic MapReduce (EMR), Amazon Kinesis, Amazon Athena, Amazon Redshift, and Amazon SageMaker for advanced analytics, real-time streaming, querying, and machine learning (ML) workloads, Cloudera offers one integrated data center for managing and analyzing data. In addition, integration with industry-leading platforms such as Amazon Bedrock enables enterprises to use generative AI tools for novel use cases. It's a tight integration, meaning enterprises can use existing AWS tools and then add the agility and functionality of Cloudera's platform to their workflows to drive value immediately without disrupting existing cloud practices.
- **Comprehensive data lifecycle management:** Cloudera offers true end-to-end data management for the whole data lifecycle. From data acquisition and engineering (ETL and ingestion support) to setting up a centralized virtual data lake on Amazon Simple Storage (Amazon S3), Cloudera is as flexible and scalable as possible. There are also data warehousing and operational databases for structured analytics and transactional processing with low latency, streaming analytics for real-time data, and ML and data visualization applications for complex analytics and AI-enabled decision-making. AWS solutions, such as Amazon SageMaker for ML and Amazon QuickSight for visualization, juggle with Cloudera's products to deliver an ecosystem that converts data from noise to value at each stage of the data flow.
- **Security, governance, and multi-tenancy:** One of Cloudera's key differentiations is its security and governance approach to data with its Shared Data Experience (SDX). SDX provides fine-grained control over data access, migration, and security within hybrid and multi-cloud environments so that organizations meet regulatory and security expectations no matter the location of their data in the cloud. This is further enhanced by AWS's globally scalable platform and secure multi-tenancy, which makes AWS the best choice for enterprises running sensitive or mission-critical workloads. Cloudera on AWS delivers a safe place to build confidence and operate across an increasingly complex data landscape.
- **Cost and operational efficiency:** There are cost and productivity savings when migrating to Cloudera on AWS. By simply rehosting, enterprises can remove the burden and downtime of migrating and save on the use of resources. With AWS's elastic scalability, companies can grow and contract their infrastructure as needed, without overprovisioning, which will ultimately drive costs down and save money. Moreover, Cloudera's analytics services integrated with AWS speed up time-to-insight to make it easier and faster. Together, these enable companies to make their total cost of ownership as low as possible while increasing the productivity and visibility of their data.
- **Innovation and business agility:** Combining Cloudera on AWS opens up an entirely new world of innovation and agility. Coexistence with services such as Amazon SageMaker and Amazon Bedrock helps organizations take AI and ML to the next level by enabling enterprises to leverage them on a scale that can speed up the production of AI solutions. Cloudera empowers organizations by providing access to AWS's entire library of cloud-based tools that help ensure companies can be at the forefront of technological change while remaining consistent and in control of their data operations. This synergy enables organizations react easily to evolving market needs, drive innovation, and stay ahead of the competition.



## Your Data Is Your AI Differentiator

With generative AI empowering organizations to use their data in ways that were not possible a short time ago, data isn't just a strategic asset; it's the cornerstone of building unique and differentiated generative AI solutions—and your secret weapon in the competitive landscape. When you use your internal domain-specific current and historical data in generative AI solutions, you have added unique knowledge and insights that improve accuracy and deliver differentiated experiences with endless possibilities. Internal data reflects your organization's operations, processes, and customer interactions. This distinct data is the vital ingredient for generative AI applications that provide personalized insights, spur innovation, and create solutions tailored precisely to your business challenges and opportunities.

But underpinning any successful generative AI initiative is successfully managed data as a trusted source, access AI models and tools with the ability to test and deploy generative AI solutions leveraging your unique enterprise data. Cloudera on AWS delivers what organizations need to make generative AI solutions a reality as a streamlined process to manage data, AI, and analytics at scale.

## Unique Benefits of Cloudera on Amazon

Cloudera on Amazon empowers businesses with full control over their data. Some of the key benefits include:

- **Seamless integration with Apache Iceberg:** Cloudera is the only data platform that works seamlessly with the popular Apache Iceberg, a modern data storage technology, across both cloud and on-premises systems at the same time. This integration also allows Cloudera to work smoothly with other tools like Snowflake, Amazon EMR, and Amazon Athena, thanks to its advanced Iceberg REST Catalog.
- **Full control for customers:** When Cloudera on Amazon is used, everything runs directly in the organization's own Amazon account, not in one controlled by Cloudera. This gives you complete ownership and transparency. Here are some of the key benefits:
  - **Amazon infrastructure costs stay on your account:** All the services used by Cloudera—like Amazon S3 (storage), EC2 (computing), and others—are billed to the organizations existing Amazon account. This means organizations can take advantage of any special deals or discounts they already have with Amazon, such as reserved storage or computing capacity.
  - **Your data, your control:** Cloudera runs entirely within the organizations private Amazon network (VPC) and your chosen subnets. Unlike many other platforms that require organizations to use their network or a shared system, this setup keeps data completely under your control. It also eliminates concerns about security or unauthorized access to data, whether it's being moved or stored.
  - **Uses Amazon S3 storage:** Cloudera stores your data in Amazon S3 buckets, a widely used and trusted cloud storage service. Unlike other platforms that lock you into proprietary storage systems, Cloudera uses open file formats like Iceberg tables and Parquet files. This means data remains compatible with other Amazon services.
  - **Works with Amazon's data tools:** Cloudera makes it easy to connect data with Amazon's built-in tools, like Amazon EMR for big data processing, Athena for analyzing data, and SageMaker for building ML models. The organizations data in Cloudera looks and feels like it belongs to these tools, making it simple to work with. Cloudera also integrates smoothly with services like Amazon Kinesis (for streaming data) and Amazon Bedrock (for AI-based applications).



By providing a secure, flexible, and fully open environment, Cloudera on Amazon empowers businesses to modernize their data operations while keeping control firmly in their hands. It's designed to work seamlessly with the tools you already use, making it a great choice for businesses of all sizes.

## Conclusion

Transitioning to Cloudera on AWS is an exciting opportunity for organizations wanting to get the most out of their data and a trusted, secure, and forward-looking platform. Migration lets organizations take advantage of Cloudera's hybrid and multi-cloud features, along with the strong cloud, AI, and infrastructure and services of AWS. Together, they provide an ecosystem that can accelerate AI and analytics innovation, simplify workloads, and improve overall productivity at scale.

The 18 integrations with Cloudera and AWS also enable organizations to take advantage of the full power of advanced analytics and AI, enabling deeper insights and faster time-to-value by cutting the complexity out of re-platforming and offering ready-to-use data and AI use cases. Cloudera on AWS positions enterprises to keep up with today's hypercompetitive, data-based world. In addition, Cloudera and AWS bring you one-stop security, governance, and compliance, which are essential building blocks of data trust. This collaboration not only speeds digital transformation but also ensures companies are able to scale quickly and reliably while remaining agile enough to adapt to new market requirements.

In short, Cloudera on AWS is not just a migration, but an investment in an intelligent, scalable, and secure data future. Enterprise Strategy Group highly recommends organizations to engage with the Cloudera team to explore true hybrid and its unique value proposition.

©TechTarget, Inc. or its subsidiaries. All rights reserved. TechTarget, and the TechTarget logo, are trademarks or registered trademarks of TechTarget, Inc. and are registered in jurisdictions worldwide. Other product and service names and logos, including for BrightTALK, Xtelligent, and the Enterprise Strategy Group might be trademarks of TechTarget or its subsidiaries. All other trademarks, logos and brand names are the property of their respective owners.

Information contained in this publication has been obtained by sources TechTarget considers to be reliable but is not warranted by TechTarget. This publication may contain opinions of TechTarget, which are subject to change. This publication may include forecasts, projections, and other predictive statements that represent TechTarget's assumptions and expectations in light of currently available information. These forecasts are based on industry trends and involve variables and uncertainties. Consequently, TechTarget makes no warranty as to the accuracy of specific forecasts, projections or predictive statements contained herein.

Any reproduction or redistribution of this publication, in whole or in part, whether in hard-copy format, electronically, or otherwise to persons not authorized to receive it, without the express consent of TechTarget, is in violation of U.S. copyright law and will be subject to an action for civil damages and, if applicable, criminal prosecution. Should you have any questions, please contact Client Relations at [cr@esg-global.com](mailto:cr@esg-global.com).

---

### About Enterprise Strategy Group

TechTarget's Enterprise Strategy Group provides focused and actionable market intelligence, demand-side research, analyst advisory services, GTM strategy guidance, solution validations, and custom content supporting enterprise technology buying and selling.

 [contact@esg-global.com](mailto:contact@esg-global.com)

 [www.esg-global.com](http://www.esg-global.com)