

The Secret to Harnessing Transformative Al for Business

Cloudera on Amazon Web Services (AWS)

Table of Contents

Introduction	3
What's Stopping You?	
Challenges on the Road to Al	4
Data Integration Barriers	6
Rigid Model Infrastructure	6
Lack of Security, Governance,	
and Transparency	7
Keeping Costs Down	7
Three Pillars for Al Success	8
Data: Enterprise Context	10
Flexibility: Any Model, Anywhere	11
Trust: Security and Governance	12
Cloudera: Making the Promise a Reality	13

Two Paths to AI Transformation	
with Cloudera	16
Cloudera Machine Learning	16
Third-Party Integration	16
Leveraging Amazon Bedrock	18
What Will You Create with	
Generative AI and Cloudera?	19
Get Enterprise-Grade Al Powered by	
Cloudera on AWS	20
Data: Enterprise Context	20
Flexibility: Any Model, Anywhere	20
Trust: Security and Governance	20
Conclusion	21

Introduction

What are some of the ways AI will transform your business? Today, every organization is racing to get the most out of AI for their business.

Most of the promise of Al lies in two areas:

- **Cost reduction:** Cutting costs through optimizing and automating current processes, as well as other opportunities for monetization through data
- **Revenue increase:** Building new revenue streams through new services, as well as adding new Al-infused capabilities to current offerings

Data may be your organization's biggest asset when it comes to getting the most out of AI. Your data is unique, meaning it can create unique value when used to train your own AI.

For example, BloombergGPT, a large language model (LLM) trained on both general data and Bloomberg's proprietary financial data, makes financial data exploration faster and more insightful for Bloomberg customers. It can provide quick answers to natural language questions about financial data (such as, "What stocks are similar to Tesla?") and also powers more complex queries and analyses. What are some ways that Al could benefit your organization?

- Boost efficiency and productivity with automation and streamlined workflows
- Enhance customer and employee experience, engagement, and retention
- Unlock innovation and power data-driven decision-making

But to start achieving these benefits, you must first unlock the business value in your data. Too often, siloed data sources, clunky analytics tools, and skyrocketing costs can hold back Al initiatives.

Cloudera on AWS takes a different approach. Cloudera puts your data at the center, bringing Al to your data, not vice versa. By integrating with Amazon Bedrock, Cloudera empowers your organization to unlock the full potential of Al for your business, no matter where your data resides.

In this ebook, we'll explore some of the main challenges enterprises can encounter along their way to implementing Al and how Cloudera on AWS facilitates the three essential pillars for successful Al transformation. Then, we'll dive into how Cloudera, with the power of Amazon Bedrock, lets you leverage Al to derive maximum value from your data.



What's Stopping You? Challenges on the Road to Al

Al itself isn't just a trend. In fact, 97% of global executives foresee Al revolutionizing data connections in a variety of roles within their organizations, from IT, development, and operations to sales and marketing and customer service and support.

Yet while business leaders are overwhelmingly aware that AI is crucial to their future success, 50% struggle with managing AI-related risk.

Every piece of data you possess can yield valuable strategic business insights, as seen in the exciting ways Bloomberg is using its proprietary financial data to serve customers and drive internal innovation. So why aren't more organizations rushing to make the most of their data? The answer is that while data can boost the power of AI to transform businesses, handling data can also be a big challenge. To ensure trustworthy output, you need to give AI access to all your data while also keeping that data secure and fully governed every step of the way.

There are a number of potential roadblocks between organizations and the innovation they hope to achieve. The roadblocks fall into three general categories: Data integration barriers, Rigid model infrastructure, and Lack of security and transparency.

97%

97% of global executives foresee AI revolutionizing data connections in a variety of roles within their organizations, from IT, development, and operations to sales and marketing and customer service and support.

50% struggle with managing Al-related risk.

50%

Δ

Roadblocks on the way to maximizing Al in your enterprise.

ADOPTION OF ENTERPRISE GENERATIVE AI IS A JOURNEY

Identifying AL challenges in the enterprise

CHALLENGES		WHAT'S MISSING
Data integration barriers		 Streamlined access to fresh enterprise data
Rigid model infrastructure	¢••	• Modularity • Flexibility • Al Ops
Lack of security and transparancy		 Model control End-to-end security Visibility and governance

Figure 1: Roadblocks on the way to maximizing Al in your enterprise.

Data Integration Barriers

The key to leveraging all your data for use with Al is making sure that data is accessible no matter where it resides, across on-premises and cloud environments. This is even more important with generative Al, which needs a rich, accessible data ecosystem.

However, enterprises often struggle to integrate all their data securely. This is especially true with hybrid storage, which spreads data across the AWS cloud and on-premises environments, increasing complexity and potential security risks for secure integration.

Generative AI needs context

Al success hinges on trusted data. And problems with data, in turn, create unreliable models and missed opportunities. What kinds of problems?

- Siloed data: Models will not have all the data they need for meaningful results because they may be missing vital context.
- **Insecure data:** Models will be unreliable if the underlying data can't be trusted or may have been tampered with.

For generative AI to thrive, you need data accessibility and models that integrate smoothly and securely within the platform and services you are currently using. This will reduce friction and risk and accelerate adoption across your organization.

Rigid Model Infrastructure

Rigid infrastructure can inhibit innovation and slow down AI application development on AWS. This is particularly true when it comes to generative AI, where new state-of-the-art models are being released weekly.

You need a platform that is both flexible and modular, letting your team continuously monitor and re-evaluate which models are available, then swapping them out when a better model is available. This could be as frequently as every one or two weeks.

With a flexible model infrastructure, you'll be able to sustain this pace of change, meaning your applications can quickly graduate from the lab and into production. Plus, by finetuning foundation models, they can be reused for multiple use cases: Q&A, document summarization, natural language to SQL, or code assistance. For generative AI to thrive, you need data accessibility and models that integrate smoothly and securely within the platform and services you are currently using



Lack of Security, Governance, and Transparency

Trust and governance are significant concerns when training models on sensitive data. Security, governance, and transparency need to extend beyond your data lake to include all models and model components using your data. In addition, transparency around the decision-making process of your generative Al solution is essential.

Yet with so many solutions, many of these considerations are added in as an afterthought. For example, narrowly focused solutions, even when integrated with one another, often can't provide a holistic big picture of security across all data and applications.

Few things are as essential to successful Al transformation as securing data in real time, at rest, and in transit, along with ensuring governance across in-house and third-party solutions. By choosing a unified platform that's inherently secure and governed from the ground up, you can be confident in the integrity and security of your Al solutions.

Keeping Costs Down

One final roadblock on the way to Al optimization (not included in the Figure above) is the challenge of achieving a reasonable total cost of ownership (TCO) for compute resources. You need to be able to customize your infrastructure to the workload, including balancing cost-effectiveness with performance for generative Al.

Cost control optimization along with security is essential to help organizations realize the promise of generative AI while keeping services affordable for the customer.

Many of the challenges mentioned in this section can be addressed by using an efficient unified data solution like Cloudera. Cloudera directly enables companies to get past all the roadblocks mentioned above on the way to strategic Al transformation, as we'll explore in the following section. Cost control optimization along with security is essential to help organizations realize the promise of generative Al while keeping services affordable for the customer.



Three Pillars for Al Success

Since 2008, Cloudera has been a trusted data partner for enterprises everywhere, and today, it is also known as a partner for Al transformation.

Cloudera simplifies your AI initiatives with a centralized open data lakehouse that includes integrated AI capabilities, providing a seamless transition to a wide range of AI use cases across your entire organization.

As an expert in data management, Cloudera on AWS puts your data at the center, making Al easier, more economical, and safer. By bringing Al models to your data, and using a single streamlined solution for data, analytics, and Al, you can achieve optimal availability, flexibility, cost-effectiveness, and security.

Let's explore the three pillars that Cloudera on AWS has been built around to help your organization harness AI for business success. Each pillar addresses one of the first three challenges discussed in the previous section.



Cloudera simplifies your Al initiatives with a centralized open data lakehouse that includes integrated Al capabilities.

Data

Core Problem

Without context from your enterprise data, Al solutions are generic and incomplete.

Challenges

Making all data available to provide essential context for Al is a major concern.

Cloudera Solution

An open data lakehouse makes your enterprise context—all of your proprietary, trusted data available to any analytics or AI solution.



Flexibility

Core Problem

Without flexibility, you'll be locked into a limited solution, unable to pivot when you need to.

Challenges

Rigid models are difficult to train and modify, and they don't change with your needs.

Cloudera Solution

Its flexible, cost-effective approach integrates with a wide range of models for power and agility.



Trust

Core Problem

Without trust, your Al transformation could fail to achieve its goals and business benefits will be limited.

Challenges

Lack of governance and trust can undermine the quality of results and confidence in results, degrading business value.

Cloudera Solution

You get a single integrated platform with data as well as model security and governance built right in for maximum visibility and optimal control.

Let's take a closer look at how Cloudera uses these three pillars to facilitate your journey to data-powered Al.

Data: Enterprise Context

Enterprise context supercharges AI with your proprietary data.

Data essentially serves as the core of any Al application your business seeks to create. The power of Al lets you get maximum value out of your data—now and into the future when emerging applications will use your data in new ways. Yet to achieve this, you need to be able to access all your data, at any time.

You probably have many different data sources and types, stored in many different ways across hybrid cloud environments. Only a holistic view of all this data will let you achieve truly optimal personalization and performance with Al. And the only way to achieve this is with flexible, scalable, and efficient data infrastructure, such as a data lakehouse, fabric, or mesh.

These modern architectures ensure that your data is stored in a way that's orchestrated (as well as governed and secured, which we'll discuss later) across multi-cloud and on-premises environments. Data lakehouses, in particular, provide an open architecture that lends itself to a variety of applications, giving you the benefits of data lakes, such as flexibility and scalability, combined with the data quality of warehouses. They also give you more flexible, democratic access to data and analytics, with an emphasis on safety, security, and governance.

A connected data lakehouse makes your enterprise context—all of your proprietary, trusted data—available to any analytics or Al solution.

How Cloudera Makes This Happen

Cloudera's data lakehouse and Shared Data Experience (SDX) provide a robust foundation, ensuring that your analytics and enterprise AI solutions have protected access to fresh, secure, and well-managed data, boosting their reliability and accuracy.

Cloudera is also open and adaptable, fostering continuous Al innovation by integrating data from any source, anywhere, so you can future-proof your Al initiatives to meet your needs now and tomorrow. Cloudera's data lakehouse and Shared Data Experience (SDX) provide a robust foundation, ensuring that your analytics and enterprise AI solutions have protected access to fresh, secure, and well-managed data, boosting their reliability and accuracy. **Pillars for Al Success**

Flexibility: Any Model, Anywhere

Flexibility lets your AI solutions adapt and improve to serve you into the future.

Al will deliver more value to your enterprise if you can easily choose from the best-performing models to develop and fine-tune more Al applications.

So, to achieve a successful Al transformation, you need the built-in flexibility to choose from a range of models (open- or closed-source) on cloud or in the data center, regardless of where your data lives. This gives you the freedom to move your data and analytics around to different environments to adjust to business requirements.

How Cloudera Makes This Happen

Cloudera on AWS gives you the most flexible, cost-effective approach to Al development while letting you take advantage of a hybrid architecture that minimizes TCO as your data grows. Additionally, Cloudera's Accelerators for ML Projects (AMPs) expedite experimentation, allowing for rapid model testing and refinement.

Integrations like Amazon Bedrock further expand your options, giving you access to a vast selection of pre-trained models, making it much simpler to leverage cutting-edge AI capabilities with your enterprise data.

Integrations like Amazon Bedrock further expand your options, giving you access to a vast selection of pre-trained models, making it much simpler to leverage cutting-edge Al capabilities with your enterprise data. **Pillars for Al Success**

Trust: Security and Governance

Building AI solutions around trust is the only way to achieve accurate, consistent, and reliable results.

Achieving trustworthy Al within your organization demands that you establish trust across all data, models, and processes. To achieve this, you first need consistent data security and governance frameworks. Security protects your data from breaches, leaks, and manipulation that could lead to inaccurate Al models and biased outputs that can leave you open to financial or reputational risks.

Beyond this, effective governance ensures that all data is known and understood, defines clear access controls, and establishes ownership. When AI models have access to reliable information, this lowers the risk of biased or nonsensical results.

Finally, with strong security and data governance practices in place, users will grow more confident over time in the quality of Al solutions and Al-based insights, leading to even greater adoption and future innovation.

How Cloudera Makes This Happen

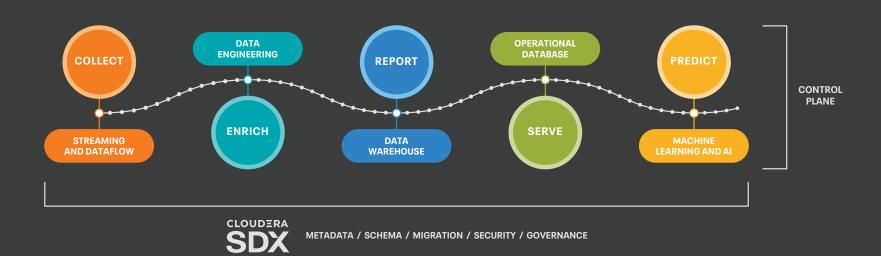
It's normal for security to be a primary concern when it comes to AI. How will your data be used? How will confidentiality be preserved?

Cloudera on AWS gives you a single integrated platform, with built-in security and governance. This increases transparency and maximizes your ability to access and leverage your proprietary data while at the same time minimizing risk and facilitating compliance with all applicable privacy standards. All of which adds up to highly trustworthy, reliable results.

Cloudera on AWS gives you a single integrated platform, with built-in security and governance. This increases transparency and maximizes your ability to access and leverage your proprietary data while at the same time minimizing risk and facilitating compliance with all applicable privacy standards.

Cloudera: Making the Promise a Reality

As a fully integrated solution, Cloudera puts your data front and center, reimagining the data lifecycle so that you can unlock the full value contained in information you collect from a wide range of sources.



DATA LIFECYCLE

Figure 2: The Cloudera data lifecycle lets you make the most of all your data with minimal friction.

Collect

Cloudera gives your Al models the latest, up-to-date, complete context with unified data collection from any source, on-premises or cloud, structured or unstructured, for real-time insights and Al-ready analysis, all in a secure and scalable platform.

Example

A manufacturing company gathers real-time sensor data from production lines to monitor equipment health and predict maintenance needs.

Enrich

Cloudera refines raw data into a clean, accessible format for impactful analysis, empowering all stakeholders and business decision-makers.

Example

A retailer cleans and organizes customer purchase data to understand buying patterns and personalize marketing campaigns.

Report

With centralized data, Cloudera enables unified reporting and real-time analytics so you can uncover hidden patterns and optimize your business.

Example

A healthcare provider centralizes patient records, treatment plans, and billing data to generate comprehensive reports and improve care outcomes.

Serve

Cloudera transforms data into actionable insights, delivering ROI by optimizing operations and fueling growth across your business.

Example

A logistics company analyzes traffic patterns and delivery routes to improve delivery schedules and reduce costs.

Predict

Cloudera leverages AI to predict trends, mitigate risks, and generate new content. Secure, automated workflows, powered by Amazon Bedrock, deliver data-driven predictions without complex coding, accelerating AI success.

Example

A finance company leverages machine learning to detect fraudulent transactions and predict customer credit risk.

Based on this holistic data lifecycle, Cloudera on AWS empowers flexible AI transformation by:

- Making all your data accessible—real-time, streaming, stored, and more
- Supporting flexible and fast application development through integrations and AMPs
- Letting your developers use their preferred tech stack
- Ensuring trust through built-in security and governance every step of the way

At Cloudera, Al transformation begins with your data.

Cloudera brings your data together with generative AI to produce the insights you need.

DATA = ENTERPRISE CONTEXT

Your data enables contextually accurate responses from LLMs

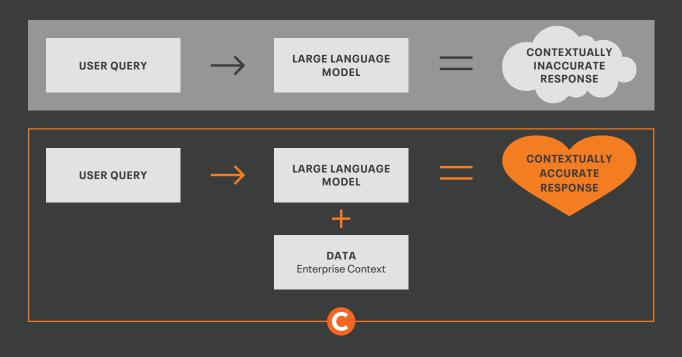


Figure 3: Cloudera brings your data together with generative AI to produce the insights you need.

Without your data, you'll receive generic responses that could be inaccurate and don't add value. Look at these two examples from an ecommerce chatbot. Which would you rather encounter?

User: I'm looking for headphones. **Chatbot:** Here is a selection of our best-selling headphones.

User: I'm looking for headphones. **Chatbot:** Based on your past purchases, here are three models you might enjoy. And since this is your third visit this year, I can offer you a loyal-customer discount today of 10%.

With context from your organization's data, Cloudera delivers AI that offers more relevant, reliable, and helpful results, setting you apart from the competition.

Two Paths to Al Transformation with Cloudera

Cloudera on AWS gives you two paths to harness the power and potential of AI: Cloudera Machine Learning and third-party integration.

Cloudera Machine Learning

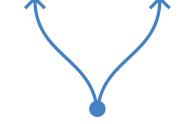
To quickly get your organization's AI journey on track, Cloudera Machine Learning streamlines the path from raw data to actionable insights through its generative AI capabilities built on secure, trustworthy data. Using Cloudera's Accelerators for Machine Learning Projects, the platform accelerates time to value by providing high-quality AI reference examples.

Third-Party Integration

Cloudera's capabilities can extend almost infinitely through integration with third-party AI. This empowers flexible AI development by letting you embrace both open and closed-source models, as needed. One example of this is Cloudera's ability to integrate with Amazon Bedrock, which we'll explore specifically in the next section.

What are some typical reasons for choosing closedvs. open-source AI models?

TYPE OF MODEL	CLOSED-SOURCE	OPEN-SOURCE
Cost	Typically more expensive (subscriptions, licensing fees)	Free to use, but may require resources for customization
Control	Limited control over model internals and functionality	Full control over the model; can be modified and customized
Security & Privacy	Less transparent; security concerns if not from reputable vendors	More transparent; security depends on implementation



Many businesses begin their Al journey with closed-source models. These give you a familiar path with pre-built functionality, making initial testing and exploration straightforward. However, closed-source models can be expensive and lack the transparency and control your organization may want to claim as your use of Al grows.

Open-source models are emerging as a powerful alternative. They are typically smaller, faster, and more cost-effective to run. Perhaps most importantly, opensource models offer greater control and customization. This lets you fine-tune models for your specific needs—and, very critically, work within the context of your data—giving you better performance over the long term than closed-source solutions.

Only Cloudera on AWS gives you the agility and control to pivot as models evolve, continually monitoring and replacing existing models with improvements to keep your Al initiatives at the forefront.

Cloudera's flexibility fosters greater customization, safeguarding your sensitive data and intellectual property, preventing vendor lock-in, and letting you seamlessly integrate with today's best solutions, such as Amazon Bedrock.

Any Model

Use the right model for the right job: closed or open-source. Cloudera gives you the agility and control to choose the best model to meet your needs.

CLOSED-SOURCE	OPEN-SOURCE
Most advanced Al models	Rapidly improving Al models
Great for a wide range of tasks	Excel at more specialized tasks
Usage can easily scale but so can your costs	Better cost planning
Compliance, privacy, and security risks	More control over where & how models are deployed

Figure 4: Cloudera gives you the agility and control to choose the best model to meet your needs.

While Cloudera gives you the flexibility to use any tech stack you prefer, our unique ecosystem facilitates best-in-breed stack integrations to help you build the apps you need with minimal friction.

Cloudera's Friction-free Ecosystem Integrations:

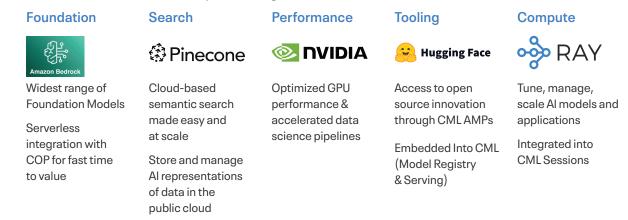


Figure 5: Cloudera's friction-free ecosystem integrations

Leveraging Amazon Bedrock

Integrating with Amazon Bedrock, a fully managed serverless service, significantly extends Cloudera's generative AI capabilities by providing you with a wide range of industry-leading foundation models.

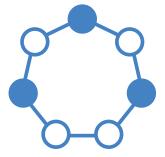
These pre-built and pre-trained AI models, designed for tasks like image analysis, language processing, and coding, give you a major advantage. When fine-tuned using your unique and proprietary data, they can be adapted to meet a wide range of needs, delivering customized AI applications far faster and at a lower cost than building a new model from scratch.

When integrated with Cloudera's data capabilities, Amazon Bedrock lets you quickly and easily build generative AI applications using your own data. Further leveraging familiar AWS tools and capabilities, you can deploy reliable, secure, and scalable generative AI applications.

Amazon Bedrock also makes generative Al accessible, allowing you to embed these foundation models into applications. As a fully managed service, you won't need to set up or manage Amazon Bedrock infrastructure, meaning you can get started building reliable, secure, and scalable generative Al applications immediately, while Cloudera takes care of all your data needs. And because Amazon Bedrock is integrated with familiar AWS tools and security capabilities you may already use, there's no need for the risk analysis required when adding new cloud infrastructure, saving time and effort.

Plus, Amazon Bedrock gives you access to cuttingedge AWS Trainium and AWS Inferentia chipsets, specialized processors designed and developed to perform AI tasks far more efficiently than generalpurpose processors.

Together, Cloudera on AWS and Amazon Bedrock give you the industry's most efficient and lowest-risk path to harnessing generative AI. You'll future-proof your business's AI investment with an open, flexible platform that can easily expand and integrate with other solutions.



What Will You Create with Generative AI and Cloudera?

Looking beyond the hype, generative AI is making a real difference to businesses of all kinds today. From cutting costs thanks to automation and revenue growth to AIpowered services or features to practical benefits like improved efficiency, customer experience, and data-driven decision-making.

The transformation to a holistic approach to AI usually involves two stages: First, extending your current AI capabilities, and later, unlocking hidden patterns and trends fueling game-changing AI applications—from personalized customer experiences to predictive maintenance.



Just a Few of the Nearly Infinite Possibilities When You Combine Cloudera With Al...

Developer productivity gains:

What if your developers had a coding assistant already familiar with your context? The assistant could help developers auto-generate, debug, and improve code, speeding up deployment cycles so your dev teams could focus on innovation. Cloudera makes this process fully secure, with no data exposed to third-party Al services.

Call center improvements:

Cut the time that agents need to handle calls through automated call handling and a self-service online knowledge base that can walk customers through common issues. An internal-use chatbot can leverage the same knowledge base to speed up support team inquiries. Al can also use speech-to-text to transcribe and summarize contact center calls, helping you train new personnel and improve call-handling processes overall. Al can even help detect customer mood and intent to improve net promoter score (NPS) and other satisfaction metrics.

Business-process boosts:

Al can enhance collaboration and save team members' time across your entire organization. For example, using Al to create meeting summaries or review documents to assist in classification and risk management, or to manage internal knowledge within departments, such as helping to codify and streamline HR processes.

Get Enterprise-Grade Al Powered by Cloudera on AWS

Your business needs the power of AI, and you need the smoothest possible path to get there. Cloudera on AWS simplifies your entire data journey, bringing together all your data from a wide range of sources, giving you actionable insights and eliminating complexity at every step.

Cloudera's three pillars all work together to support your successful Al transformation.

Data: Enterprise Context

Cloudera's open data lakehouse provides a central hub for all your data, capable of handling any format, whether on-premises or in the cloud. It cleans, organizes, and prepares your data for clear analysis and reporting. Automation throughout the data lifecycle unlocks its full potential. This open and adaptable approach allows you to integrate data from any source, futureproofing your Al initiatives for long-term success.

Flexibility: Any Model, Anywhere

Cloudera provides maximum flexibility to leverage cuttingedge AI capabilities centered on your data while minimizing costs. With a range of available machine learning and AI frameworks, Cloudera lets you choose the approach that best suits your needs. Additionally, Cloudera provides access to pre-built models, both managed and open source, that can be further customized to meet your industry's requirements. Integrations like Amazon Bedrock further accelerate development by providing a starting point for testing and refinement.

Trust: Security and Governance

Security and governance are an integral part of Cloudera, which gives you robust controls throughout your data's lifecycle, regardless of format or source. With its integrated approach, Cloudera minimizes risks and eases compliance with privacy regulations. By prioritizing security and governance, Cloudera empowers you to confidently access and leverage your data for reliable AI results.



Conclusion

With Cloudera on AWS, you'll be able to focus on creating the personalized generative Al solutions you need centered around your data. And that means you'll have what you need to start deriving value faster and easier than ever before.

Stop grappling with fragmented tools and manual processes. Streamline your Al journey with Cloudera, an integrated platform built for collaboration and efficiency.

Get in touch today to begin a 5-day cloud trial and kick off your organization's AI transformation journey.

LEARN MORE

About Cloudera

At Cloudera, we believe data can make what is impossible today, possible tomorrow. We empower people to transform data anywhere into trusted enterprise AI so they can reduce costs and risks, increase productivity, and accelerate business performance. Our open data lakehouse enables secure data management and portable cloud-native data analytics helping organizations manage and analyze data of all types, on any cloud, public or private. With as much data under management as the hyperscalers, we're the preferred data partner for the top companies in almost every industry. Cloudera has guided the world on the value and future of data, and continues to lead a vibrant ecosystem powered by the relentless innovation of the open source community.

Learn more at cloudera.com | US: +1 888 789 1488 | Outside the US: +1 650 362 0488

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

© 2024 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. 6250-001 July 25, 2024

Privacy Policy | Terms of Service

