



CLOUDERA

4 Use Cases for a Hybrid Data Platform

Bring Order to Data Chaos

Table of Contents

Order Out of Data Chaos	3	Data Syndication	13
Hybrid Data Platform: The Question of Why	4	What You Need for Effective Data Syndication Across Cloud Form Factors	14
The Power of “And”	5	A Leading Telecommunications Company’s Hybrid Story	15
Four Hybrid Data Use Cases	6	Virtualize, Abstract, and Leverage Multi-Cloud	16
Develop Once and Run Anywhere	7	What You Need to Virtualize and Commoditize the Cloud	17
What You Need to Enable a Single Code Base	8	A Global Vehicle Company’s Hybrid Story	18
A Leading Electronics Manufacturer’s Hybrid Story	9		
De-risk Cloud Migration	10	Check If You Are Ready	19
What You Need to Migrate Apps Independently to the Public Cloud	11	Why Cloudera?	20
A Financial Services Firm’s Hybrid Story	12	Take the Next Step	21

Order Out of Data Chaos

Organizations that are highly data-driven are three times more likely to report significant improvements in decision-making compared to those who rely less on data.¹

For large enterprises, the effective use of data is only possible when their data analysts, engineers, and scientists have straightforward access to disparate and voluminous data along with the right computational resources to efficiently process it.

Furthermore, to act on and generate revenue from those decisions, platform engineers and application developers must be able to quickly stand-up (and take down) environments by which to develop, test, enrich, and implement the next big idea.

Private cloud works very well for some use cases, public cloud is the answer for others, and together as a whole, the needs of diverse sets of teams can be addressed.

However, to do that, it comes down to the infrastructure/cloud administrators to navigate a chaotic mix of disparate and incompatible cloud form factors to provide it all to their constituents:

- Speed and Control
- Performance and Scale
- Flexibility and Security

To that end, a comprehensive and balanced approach to a hybrid data platform is imperative. This eBook articulates Cloudera's vision of the Hybrid Data Platform through four use cases that we have implemented with leading enterprises that represent diverse industries.

In these stories, infrastructure/cloud administrators will take away a realistic understanding of what they need and the practical steps by which to bring order to what is typically a chaotic experience.

Hybrid Data Platform: The Question of Why

Worldwide spending on big data and business analytics solutions is forecast to increase 10.1% over 2020 to reach \$215.7 billion this year.² That investment goes to waste if the right data and computational resources aren't made available to the teams that can make a difference.

A corporate enterprise is made up of multiple Lines of Businesses (LOBs) and the enterprise's success relies on the ability of each LOB to leverage their expertise to outperform their market. It also helps to increase the bottom-line when they are able to collaborate across the organization and generate a whole that is greater than the sum of its parts.

Standardizing on one data infrastructure often does not address the distinct regulatory and market driven requirements of each LOB. As limitations become apparent, businesses will often strike out on their own to test other options, be it new private cloud infrastructure or the myriad of solutions provided by public Cloud Service Providers (CSPs). In the end, large enterprises often end up with a disjointed and chaotic mix of private, public, and multi-cloud infrastructures.

The decision to 'go private or public cloud' is no longer a binary one. At some point, a realization is made that it depends on the use case and so a comprehensive and balanced approach to a Hybrid Data Platform is imperative.

But before embarking on that journey, it's good to take a step back and ask, "Why?". Some sample questions are listed below.

- Have you repatriated workloads from the public cloud?
- Do you have a need to split workloads across private and public clouds?
- Is your organization using multiple CSPs?
- If you do use multiple CSPs, is one better for one situation over another?

More often than not, from the enterprise perspective, the answer is "Yes".

What Infrastructure/ Cloud Administrators Need

To address the data requirements of the diverse set of teams across their organization, infrastructure/cloud administrators need true hybrid data platform capabilities to help them bring order to what can be a chaotic experience.

This paper touches on some of those capabilities, including:

- Workload isolation and virtualization
- On-demand cloud to cloud bi-directional replication
- Workload repatriation
- Federated shared data experience
- Runtime compatibility
- Cloud native data services
- Common control plane
- Global data catalog search
- Infrastructure abstraction
- Consistent APIs
- Cloud arbitrage

The Power of “And”

Cloudera has worked with more than 1,000 organizations to drive value from their data. We've seen common themes in our customers' challenges as they modernize their platforms and, by helping them find solutions, have learned how to best help companies navigate the journey of joining their private and public cloud experiences.

Critical to cloud success is the ability of infrastructure/cloud administrators to provide the best fit-for-purpose cloud form factor to each of their constituents. Their job is challenging when you consider the inherent complexities of integrating disparate and often incompatible infrastructures. In a binary world of private cloud or public cloud, a lot of time and effort needs to go into weighing the risk and reward of this or that solution

The hybrid data use cases described in this paper show how our customers have been able to reduce complexities by optimizing the data lifecycle across both private and public cloud form factors with consistent data security and governance. The result is a deeply integrated hybrid and multi-cloud solution that provides the power of “And”:

- Speed and Control
- Performance and Scale
- Flexibility and Security

Four Hybrid Data Use Cases

The Cloudera view of a hybrid data platform has evolved one customer success story at a time. We have learned from infrastructure/cloud administrators, platform engineers, application developers, data engineers, and data scientists. The use cases presented below solved challenges within a diverse set of industries, including:

- A **leading electronics manufacturer** that is now able to create applications once and run them anywhere.
- A **financial services firm** that de-risked the migration of real-time workloads to the public cloud without breaking integration with private cloud batch workloads.
- A **leading telecommunications company** in Asia can now apply new data science workloads without changing existing data pipelines or distribution channels.
- A **global vehicle company** that has been able to virtualize, abstract, and leverage multi-cloud in order to support real-time bi-directional replication across regions.

USE CASE

Develop Once and Run Anywhere

Developers who need to write cloud native applications are usually tied to the specific offerings of a CSP. Maybe they are using the CSP's messaging API or possibly one of many shared databases, AI/ML offerings, or even lambda functions. The issue is that once that application has been deployed, it is bound to that specific CSP. Deploying that same application to a different CSP may uncover different performance characteristics or any number of other new issues. There are significant subtle differences between the CSPs that are not usually evident until you start using them. Addressing these nuances often requires CSP specific code that is managed through various approaches depending on the language. The alternative is to develop in multiple languages depending on the CSP, which then requires the maintenance of multiple code bases. None of this is trivial and increases code complexity, development time, and code quality.

What if your developers could write once to a common set of services, not worry about CSP specifics, and have the option to seamlessly move to the best CSP? A leading electronics manufacturer was able to do just that. By deploying a hybrid cloud solution, they are now able to develop assets for private cloud deployment but with options to deploy to many public clouds as needed. All the while, they keep a single code based pattern for development (see the customer story on page 9 for details).

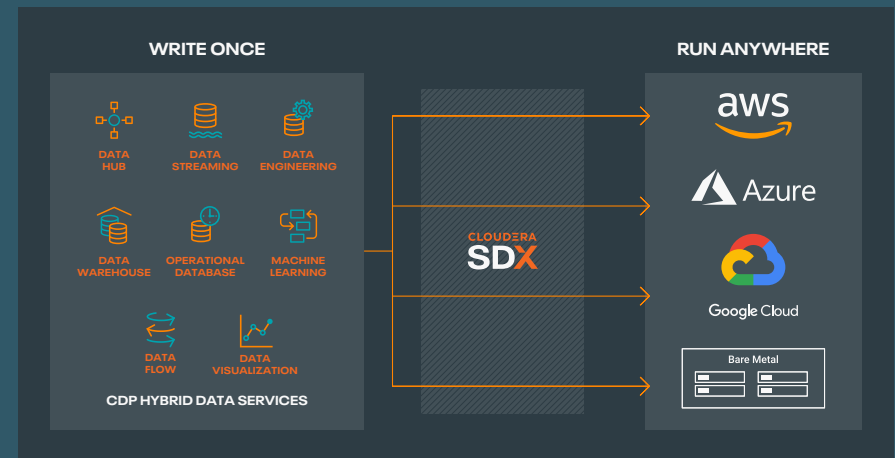


Figure 1: Any workload can be moved between private cloud and multiple public clouds without refactoring.

USE CASE

What You Need to Enable a Single Code Base

Whether it's for your private cloud or public CSPs, when you write an application using cloud specific APIs, you are bound to that environment. If you need the application deployed elsewhere, you are saddled with maintaining multiple code bases.

A common set of services that are available in a multi-cloud environment would facilitate moving an application from one provider to another without modifying the code. For example, having a common messaging API or common database infrastructure would allow the infrastructure/cloud administrator to easily move the application from public to private cloud and vice-versa.

Cloud administrators need a set of common services that allows them to efficiently manage deployment and performance characteristics across many different applications and locations and to do so is only possible if you have the following:

- Workload isolation and virtualization so that you can eliminate noisy neighbor type of problems.
- Cloud to cloud replication using a centralized management tool for replicating and migrating data and metadata between environments.
- Workload repatriation so that you can seamlessly repatriate workloads as needed.
- Federated shared data experience that ensures consistent security and governance regardless of the location of the infrastructure that is replicated across all form factors.

CUSTOMER STORY

A Leading Electronics Manufacturer's Hybrid Story

Challenge

This leading electronics manufacturer enjoys a large global client base across diverse industries with ample opportunities to grow their business.

However, that growth was impeded by a fixed private cloud environment. Additionally resources were inadequate for disaster recovery (DR). Executive leadership did not support a full public cloud migration, assuming that intellectual property (IP) would be at risk.

Why Hybrid?

To address DR challenges, critical data stores (HBase, Hive, and HDFS) were replicated in near real-time to a public CSP. This freed up private cloud resources for new machine learning and operational use cases.

The cloud to cloud replication of common data stores enabled workloads developed in the private cloud to be moved to the public cloud without refactoring.

Results

The DR success begets numerous opportunities to expand global operations:

- DR can follow the sun across regions via multiple public clouds.
- They can choose to run machine learning and operational workloads at any location while managing a single code base.
- The utility of archived data is increased by leveraging unlimited public cloud resources as needed.
- Workload isolation and the physical security of CSPs actually increased protection of IP, thus enabling more apps to be developed once and run anywhere.



4
MONTHS

The time it took
to complete their
hybrid journey

USE CASE

De-risk Cloud Migration

Migrating applications to the public cloud from private cloud can go down several different paths. One is a simple lift and shift, where the application is simply moved to an instance of a Virtual Machine (VM) or database instance managed by the CSP. Another path is full migration, where an application is rearchitected to benefit from the flexible infrastructure capabilities of the public cloud. And, of course, there are many hybrid approaches in-between.

The challenge is that there are subtle differences between private cloud and public clouds, so many distinct migration plans may need to be implemented, even for just one set of applications. This becomes even more complicated when you have applications that depend on each other but you don't want to risk a big bang migration.

Cloudera helped a financial services firm solve this exact challenge (see the customer story on page 12 for details). The organization had, in their private cloud, real-time workloads tightly integrated with batch workloads but market and economic conditions required the real-time workloads to be migrated to a public CSP.

To facilitate de-risked cloud migration, a critical step is to have bi-directional data replication between cloud instances. Having that core piece of infrastructure in place significantly reduces the risk of moving applications from private to public cloud and adds additional comfort of knowing that you can revert back to the original architecture if needed.

The benefit for this financial services firm is that they were able to move the real-time processes to the cloud while maintaining service levels for the batch component in their private cloud. Once they had achieved confidence and understood the implications of using a CSP, they were then able to migrate the batch component at a later date.

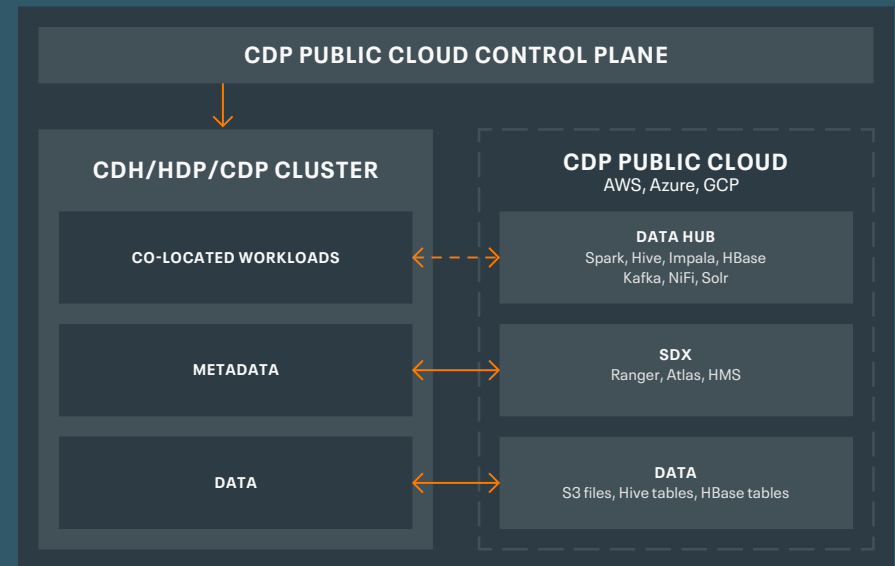


Figure 2: Bi-directional data replication with cloud native data services de-risks cloud migration.

USE CASE

What You Need to Migrate Apps Independently to the Public Cloud

Critical factors by which to mitigate application rewrite includes:

- Runtime compatibility so that platform specific executables are not required.
- The ability to leverage common database types between private and public cloud form factors and seamlessly move data across locations.
- Cloud native data services that reduce/eliminate infrastructure management.
- Bi-directional data replication that keeps data across cloud form factors in sync.

CUSTOMER STORY

A Financial Services Firm's Hybrid Story

Challenge

Mission critical applications and batch workloads were running on a single cluster resulting in batch workloads negatively impacting the SLA for their customer facing applications. A big bang approach to migrate both sets of workloads to the public cloud was deemed as too risky.

Why Hybrid?

Cloudera's hybrid approach enabled the financial services firm to segregate the "batch" workloads from the mission critical apps and those apps were moved to the public cloud. This was possible because the data store used in the private cloud (HBase) was made available in the public cloud with bidirectional replication in real-time.

Results

Executing a practical migration strategy that moved only a subset of workloads to the public cloud without breaking private cloud processes had two beneficial results. First, the migration was de-risked, and second, this customer enjoyed 20-30% savings on infrastructure cost compared to moving everything to the public cloud on a 24x7x365 basis.



20-
30%

20-30% savings on
infrastructure cost

USE CASE

Data Syndication

Data syndication is a process designed to safely provision ready-to-use data assets from disparate data pipelines to data consumers (i.e., data scientists and application developers). This enables the data consumers to experiment and create new products within one area, and by having all of that data in one place, are provided with context across the broader market.

But what happens if you want to leverage the public cloud to apply new machine learning models onto pre-existing data pipelines that are managed in a private-cloud environment? The last thing an infrastructure/cloud administrator wants to do is duplicate data pipelines per cloud form factor. Cloudera helped a leading telecommunications company in Asia leverage a hybrid cloud solution to simplify the process (see the customer story on page 15 for details).

The challenge the telecommunications company had was that the data pipelines needed for the new data science models were pre-existing with a rich data structure that has evolved over years. Furthermore, the new data insights provided needed to be distributed across the enterprise via existing business intelligence and reporting platforms. A hybrid approach was implemented that inserted horizontally scalable compute capabilities of the public cloud into pre-existing processes of data pipelines and distribution (see Figure 3, at right).

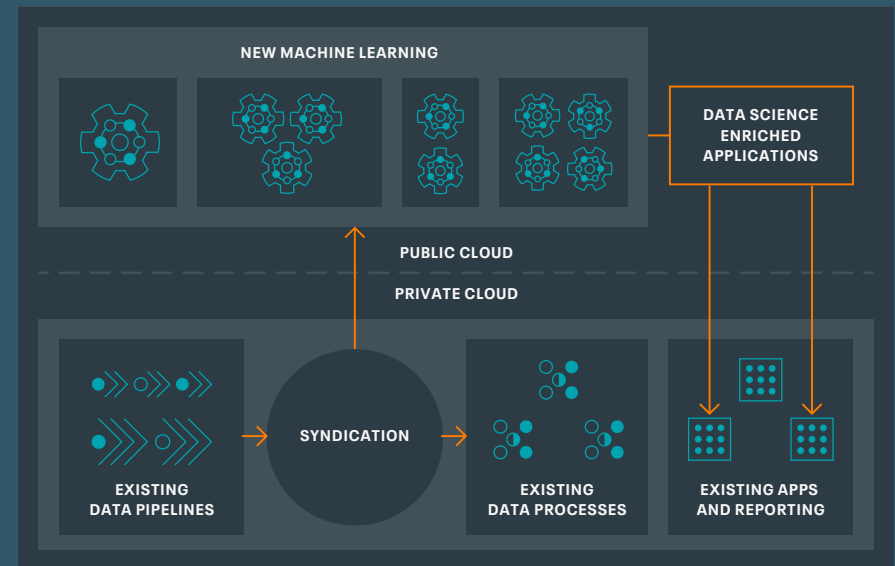


Figure 3: A data syndication model that spans across private and public cloud form factors.

USE CASE

What You Need for Effective Data Syndication Across Cloud Form Factors

Data syndication, as described above, requires navigating varied security, governance, and API protocols across private and public CSPs. To provide the data scientists rapid and straightforward access to data while making the data engineer's job more efficient, your platform will require the following:

- A common control plane that provides a universal set of tools for management, workload analysis, and data movement across multiple environments.
- Cloud to cloud data replication that not only keeps data in sync but generates a replication plan based on deep analysis of each workload.
- Shared data experience across cloud form factors, so that data security and governance consistently addresses the compliance and regulatory requirements of each use case.
- A global data catalog search to find, curate, and audit your data across all infrastructures.

CUSTOMER STORY

A Leading Telecommunications Company's Hybrid Story

Challenge

This telecommunications company made it a priority to reduce churn by improving customer experiences. They already had a robust and pervasive mobile network in place but needed to advance their data management and analytics capabilities to better tailor how their customers discover, buy, and use their services.

The challenge was to leverage new data science capabilities in the public cloud without changing existing private cloud data pipelines and distribution channels.

Why Hybrid?

Through cloud to cloud replication and a common control plane that provides a universal set of tools for management, workload analysis, and data movement across environments, the data scientists are provided large amounts of horizontally scalable compute that is needed to perform their job without touching pre-existing data pipelines.

Results

This telecommunications company is leveraging analytics to perform customer segmentation that would have been previously impossible, resulting in a better mobile experience and more effective ads.

The new analytical environment ingests massive volumes of real-time, granular network signal information and combines that with batch loads from billing systems, payments, and more.

Organizations across the company benefit from direct access to the platform.

Better mobile experience
and more effective ads
based on previously
impossible customer
segmentation.

USE CASE

Virtualize, Abstract, and Leverage Multi-Cloud

For application developers, data engineers, and data scientists, a single entry point for all cloud resources simplifies development, data science workflows, and enables architectures that take advantage of the unique benefits of each CSP. Data and compute virtualization can make this happen, by bringing multiple clouds together under the umbrella of a single control plane and a single interface. The result is a centralized location for all cloud resources regardless of where they reside.

The unique challenges of a global vehicle company required such a solution because they manage large fleets of vehicles across geographic regions and monitor each one in real-time. Since vehicle activity varies between weekdays and weekends with seasonal spikes, making the best use of resources was impossible to do manually. Cloudera helped them to virtualize, abstract, and leverage a multi-cloud infrastructure to support auto-scaling of real-time bi-directional replication across regions (see the customer story on page 18 for details).

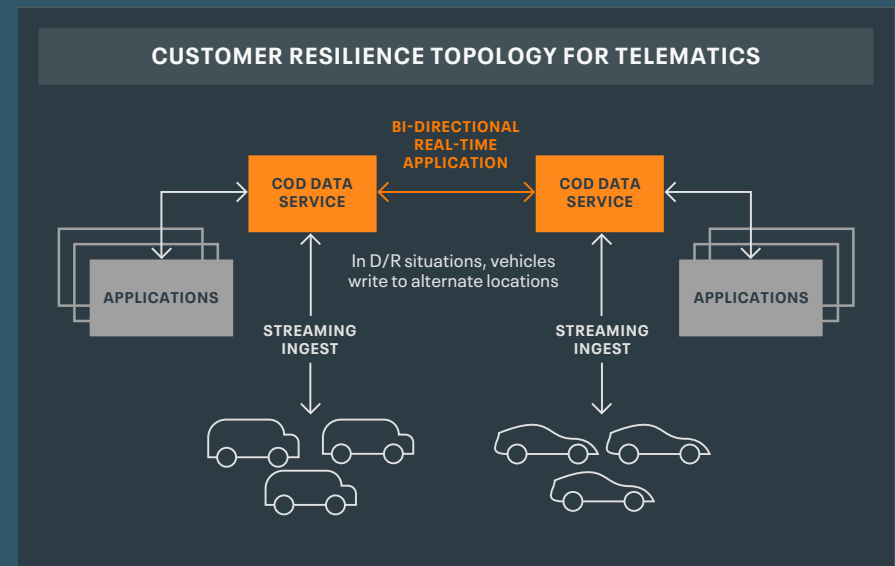


Figure 4: Cloudera enables active-active workload deployments for resiliency and to optimize TCO.

USE CASE

What You Need to Virtualize and Commoditize the Cloud

What if you didn't have to make numerous and manual infrastructure decisions every time a new application or feature is deployed?

What if you could optimize the use of resources based on workload demand and historical usage? The result would be reduced cost and size of the estate that is required to maintain high availability and performance levels. It also mitigates over allocating hardware to ensure applications SLAs can be met during periods of peak utilization, while most of the time remaining idle.

To enable above, you need the following:

- Infrastructure abstraction to enable cloud native application design and eliminate administrative overhead.
- Consistent APIs so that you are no longer required to move applications between private and public clouds without having to worry about versioning related issues.
- Cloud arbitrage to guarantee that the resources required are located in the least expensive CSP. This would also enable you to automate financial governance.
- Real-time bi-directional data replication enabling active-active architectures and reducing overall infrastructure requirements by leveraging deployed infrastructure in multiple locations

CUSTOMER STORY

A Global Vehicle Company's Hybrid Story

Challenge

This global vehicle company needed a platform that could perform predictive maintenance with efficient routing and fueling recommendations for their fleet of vehicles.

The challenge was to provide real-time analytics across wide-spread geographic regions, keeping all of those regions in sync and minimizing the operational costs of doing so.

Why Hybrid?

Cloudera's Hybrid Data Platform solution enabled near-real time bidirectional data replication across two regions, enabling a robust active-active solution.

Additionally, infrastructure abstraction capabilities automated how best to compute and autoscale each appropriately.

Results

The company is currently ingesting data from nearly 200,000 vehicles, with plans to increase five times to 1 million vehicles on the road.

Along with added scale, this company is also saving about \$1M/year in public cloud infrastructure costs due to auto-scaling.



\$1M

\$1M per year savings
in public cloud
infrastructure costs
due to auto-scaling

Check If You Are Ready

To enable these use cases for a Hybrid Data Platform, you need to make sure you are ready. Readiness includes your data, your applications, and staff. While each company will have circumstances that are unique to them, below are some guidelines to help you prepare.

- The same data architecture in private cloud and multiple public cloud, including ingest, ETL, analytics, transactions, and machine learning
- Categorizing your data sets so it's clear whether they are suitable for each environment, including InfoSec approval and data privacy considerations
- Capability to move the data, metadata, and any associated workloads/applications that are dependent on the data
- Plan to update ingest and ETL pipelines for the new location of the data
- Understand the ecosystem that is dependent upon any application or workload that is moved
- All 3rd party tooling in the environment where the data will land, especially important when data and workloads are moved

This list will help you understand the operational issues you need to address to ensure you get the full benefits of a hybrid data platform, without causing disruptions to your company's mission-critical operations.

Why Cloudera?

Cloudera has worked with more than 1,000 organizations to drive value from their data and have seen common themes in our customers' challenges as they modernize their platforms. Our experts help them find solutions, and in doing so, have learned how to best help companies navigate the joining of their private and public cloud experiences.

A result of that is how Cloudera Data Platform (CDP) reduces the complexities described in this eBook by optimizing the data lifecycle across both private and public cloud form factors with consistent data security and governance.

What CDP Hybrid Cloud provides is a deeply integrated hybrid and multi-cloud solution that is unified, flexible, secure and fast. A common control plane, workload optimization, scalable resources and an array of other services that make it straightforward for an infrastructure/cloud administrator to deliver all the benefits of a hybrid/multi-cloud environment without all the complexity of working in a multi-cloud environment.

Cloudera de-risks the process with a customer-driven step by step approach that capitalizes on the collective expertise acquired across many industries. Best practices are enhanced with CDP automated tools so that your valuable resources are focused on the intricacies of your organization.

Take the Next Step

CDP has been built to meet the needs of companies on their hybrid cloud journeys. With extensive experience implementing data-driven use cases, Cloudera has the expertise to help you simplify your move to a hybrid data platform and go from pilot to production quickly, painlessly, at lower cost, and with peak performance that is underpinned with consistent data security, governance, and management. Watch [these videos](#) to learn how CDP powers data-driven decisions by easily, quickly, and safely connecting and securing the entire data lifecycle.

About Cloudera

At Cloudera, we believe that data can make what is impossible today, possible tomorrow. We empower people to transform complex data into clear and actionable insights. Cloudera delivers the only hybrid data platform for modern data architectures with data anywhere. 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 | US: +1 888 789 1488 | Outside the US: +1 650 362 0488

Sources

- ¹ Tim Stobierski, "The Advantages of Data-Driven Decision-Making", Harvard Business School. August 2019.
- ² "Global Spending on Big Data and Analytics Solutions Will Reach \$215.7 Billion in 2021, According to a New IDC Spending Guide", businesswire.com. August 2021.

© 2022 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. 5272-001 June 14, 2022

[Privacy Policy](#) | [Terms of Service](#)

CLUDERA