CLOUDERA

NAVIGATING THE EMEA DATA LANDSCAPE

Compliance concerns and hidden costs of data management

UNLOCKING DATA VALUE

Table of Contents

Key Takeaways:

- Two-thirds (66%) of IT decision makers (ITDMs) believe data is spiraling out of control in their organization. At the same time, 79% cite compliance as their main concern when it comes to managing data.
- Almost two thirds of ITDMs (63%) say siloed data makes it harder to comply with regulations, and 79% believe that point solutions for data analytics and management makes compliance more complex.
- More than three quarters (78%) of ITDMs agreeing that point solutions for data analytics and management is driving up the cost of data, with upfront costs of tools amplified by the need to re-train staff for each new solution.
- 68% of organizations have either deployed a data lakehouse or are planning to, with 62% citing faster access to data as the key driver, and 35% utilizing a lakehouse to underpin their AI/ML strategy. We'd expect to see this figure grow as more AI business use cases are unearthed.

Executive Summary

The compliance landscape is changing rapidly, and the nuances of each new legislation are presenting fresh challenges to organizations almost daily. Data leaders are struggling to keep up, and with regulations such as Schrems III and the EU's Digital Services Act (DSA) already in the pipeline, there's no sign of compliance slowing down or going away.

At the same time, organizations are deploying multiple point solutions to help unlock the full range of capabilities from an ever-expanding pool of data. But each tool requires different skills, specialist training, and has its own complex data architecture. These tools have increased both the cost and complexity of managing data across its lifecycle. As a result, organizations are effectively paying a data integration tax for each solution.

With more organizations looking to embrace AI technology, data will continue to be their most valuable asset. The challenge facing many will be to unlock the value of their data in a compliant and cost-effective way.

Research Methodology

Conducted by Coleman Parkes Research, Cloudera's survey evaluated the opinions of 850 IT decision makers with responsibility for data analytics and tooling in their organization across the UK (200), France (200), Germany (200), Spain (100), Italy (100) and the Middle East (50). Respondents came from organizations with more than 1,000 employees within the following industries: finance, banking, insurance, manufacturing, telecommunications, retail and e-commerce, healthcare and life sciences, government and public sector, technology and software, energy and utilities. *Figures 1-5 are part of the same question, in grid format to ask the extent of agreement to set statements. The research was conducted between March and April 2023.*

01 Compliance Always and Everywhere

Data is Spiraling Out of Control

Data volumes continue to grow. According to IDC, the total volume of global data is expected to reach in excess of 180 Zettabytes by 2025.

This deluge of data is often disparate, coming in different formats — whether structured or unstructured — and sitting across multiple cloud and on-premises environments. In fact, our recent research shows that 68% of organizations currently store data in a hybrid environment, utilizing both cloud and on-premise solutions. This is creating complexity, with two-thirds (66%) of IT decision makers (ITDMs) worrying that data in their organization is spiraling out of control.

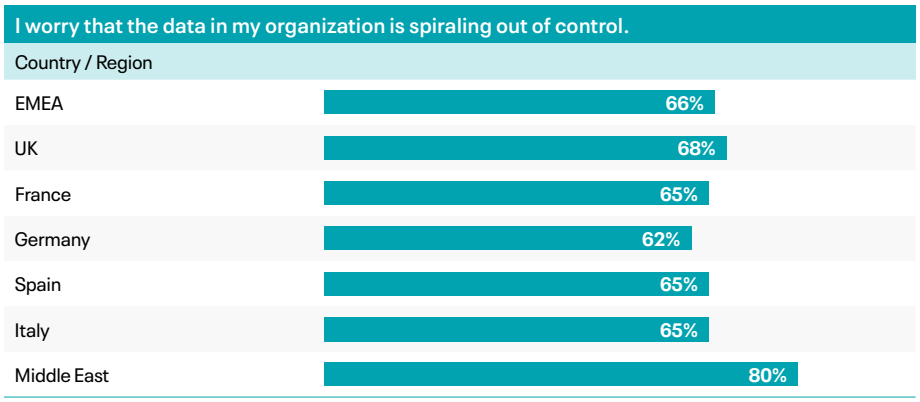

Figure 1. To what extent do you agree with the following statement? I worry that the data in my organization is spiraling out of control.

Concerning Compliance

The increased volume and disparate nature of data impacts organizations in many ways, but none more so than compliance. Four-in-five ITDMs (79%) agree that compliance is their main concern when it comes to data management. Understandably, this is most concerning for ITDMs in highly regulated industries, such as banking and finance (80%) and insurance (82%), which handle a lot of personally identifiable information (PII), such as card details and addresses. Similarly, retail and e-commerce organizations hold a high volume of payment card industry (PCI) data, causing 80% of ITDMs in the industry to express concern over compliance.

Figure 2. To what extent do you agree with the following statement? Compliance is my main concern when it comes to data management

Siloed Data and Point Solutions Adding to Compliance Woes

63% of ITDMs believe siloed data makes it harder for their organization to comply with data compliance regulations. ITDMs in the manufacturing (71%), healthcare and life sciences organizations (71%) and banking and finance (70%) sectors find this the most challenging.

In the banking and finance sector for example, many of the older, established institutions have amassed huge volumes of data over decades, much of which sits on disparate legacy platforms. Comparatively, the digital first challenger banks have been able to build their data pipelines from the ground up.

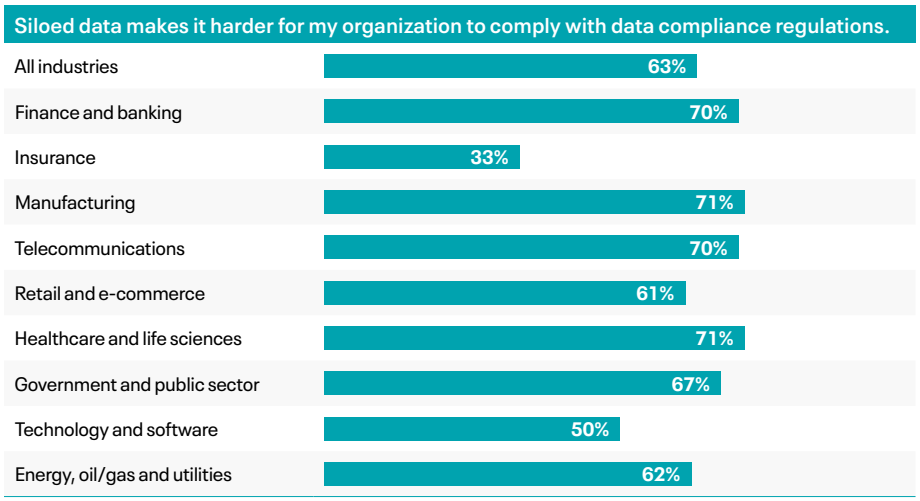

Figure 3. To what extent do you agree with the following statement? Siloed data makes it harder for my organization to comply with data compliance regulations

"As data continues to scale, managing across different environments becomes increasingly complex. The emergence of transformative technologies like AI will only exacerbate this issue if it isn't addressed now. Organizations must start to think about data as a product, utilizing a data mesh to democratize data, and protecting their greatest asset by ensuring that compliance is always on and everywhere."

Chris Royles,
EMEA Field CTO at Cloudera

79%

of organizations agree that integrating point solutions for data analytics and management has made compliance more challenging

In addition, 79% agree that integrating point solutions for data analytics and management has made compliance more challenging.

Every organization should be concerned about governance and ensuring the integrity of their internal data and their customers' PII. Organizations must have a safety net in place to ensure that the integrity of their data isn't compromised. This means understanding all the data across an organization, and having the ability to enact organization-wide governance policies, regardless of whether that data resides in the cloud or on-premises.

02 Integration Tax

The Hidden Costs of Point Solutions

Compliance challenges aren't the only issue that point solutions are creating. While they can be effective in solving specific data processes quickly — giving the illusion of faster time to value, they often create complexity.

Two-thirds of ITDMs (63%) find integrating the different point solutions required for managing data across its lifecycle a challenge, with 34% finding this a significant challenge. This is particularly prevalent across finance and banking (42%) and healthcare and life sciences (43%) organizations.

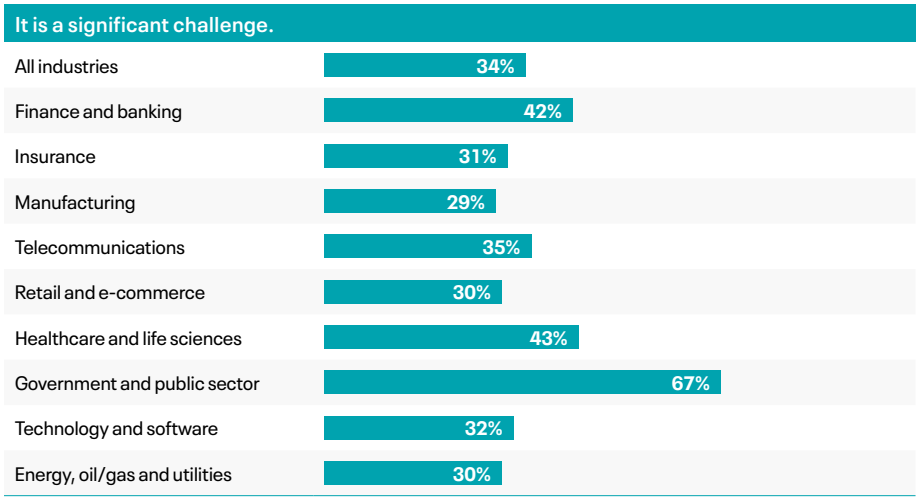

Figure 4. To what extent, if at all, do you find integrating different point solutions for managing data across its lifecycle a challenge? It is a significant challenge

Point solutions are also leaving ITDMs with long-term costs. With the up-front cost of on-boarding tools, the technical cost of integrating them, and staff requiring specialist training to operate them in the long-term, point solutions are drastically cutting the time to value of data.

More than three quarters of ITDMs (78%) believe integrating point solutions for data analytics and management has driven up the cost of data. Across all sectors, ITDMs are effectively paying a data integration tax for each point solution. This is a term used to describe the hidden costs associated with integrating data solutions to process data from disparate sources and for different needs.

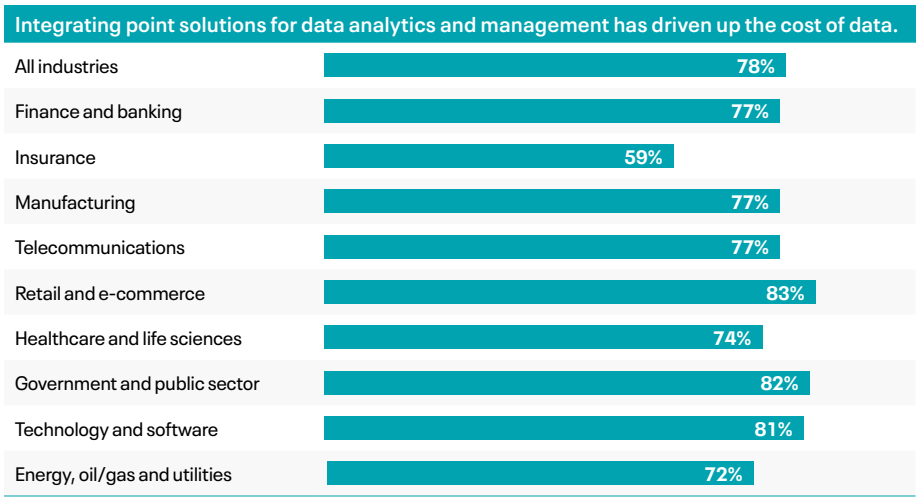

Figure 5. To what extent do you agree with the following statement? Integrating point solutions for data analytics and management has driven up the cost of data

84%

of organizations are collaborating with external partners to develop and execute their data strategy as their data becomes more complex

"To effectively carry out their day-to-day roles, engineers need access to data from across the organization and the ability to self-service. But because of the number of tools organizations deploy for managing data across its lifecycle, this isn't possible as each solution requires its own specialist skillset, which takes time and money to learn."

Chris Royles,
EMEA Field CTO at Cloudera

As a result of the complexity of point solutions, 84% of organizations are collaborating with external partners to develop and execute their data strategy as their data becomes more complex. However, this — combined with the integration tax — means ITDMs are spending more than a quarter (28%) of their annual IT budget on managing data across its lifecycle.

WHEN IT COMES THE MATURITY OF MANAGING DATA ACROSS ITS ENTIRE LIFECYCLE, ORGANIZATIONS ARE THE LEAST WELL-PLACED TO INGEST DATA AND MAKE PREDICTIONS FROM IT (BOTH 34%):

- 34% of organizations are very mature in ingesting data
- 41% of organizations are very mature in preparing data
- 41% of organizations are very mature in analyzing data
- 34% of organizations are very mature in making predictions from data
- 35% of organizations are very mature in publishing data

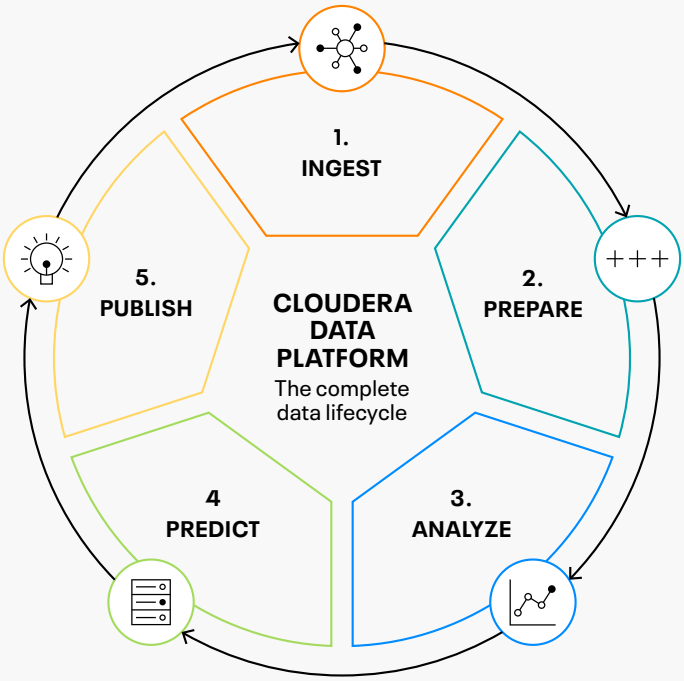

Figure 6. Data lifecycle diagram

If organizations consolidate services and reduce the number of skills required to manage data across its entire lifecycle, they'll cut the time and costs of retraining staff dramatically. This will help to reduce integration tax and deliver a faster time to value for data.

"We're in the midst of an AI revolution, with this transformative technology optimizing a huge number of business processes. But for AI to be truly effective, it needs access to a clean, complete set of proprietary data, which a lakehouse provides. This enables AI to generate accurate insight and negates the possibility of leaking data outside of their organization."

Chris Royles,
EMEA Field CTO at Cloudera

03 Diving into a data lakehouse

Lakehouse Implementation

With data being more important to organizations than ever, driving value from it is becoming vital to operations. A data lakehouse can help to provide a single source of truth for organizations by combining a data lake's ability to store an organization's complete data set at scale, with the quality and richness of data provided by a data warehouse. This essentially enables organizations to reduce duplication and drive value from a complete set of quality data.

WHAT IS A DATA LAKEHOUSE?

Data warehouses offer a single storage repository for structured data and provide a source of truth for organizations. However, organizations must structure and store data inputs in a specific format to be able to extract and efficiently query this data.

On the other hand, data lakes are flexible environments that can store both structured and unstructured data in its native form, enabling organizations to use this data to build artificial intelligence (AI) and machine learning models from large volumes of disparate data sets.

As its name suggests, a data lakehouse combines the structure and accessibility of a data warehouse with the massive storage of a data lake and is built to house both structured and unstructured data.

68% of organizations have already implemented a lakehouse or plan to deploy one. However, this varies between sectors, with industries like telecommunications (75%) and technology and software (74%) ahead of the curve.

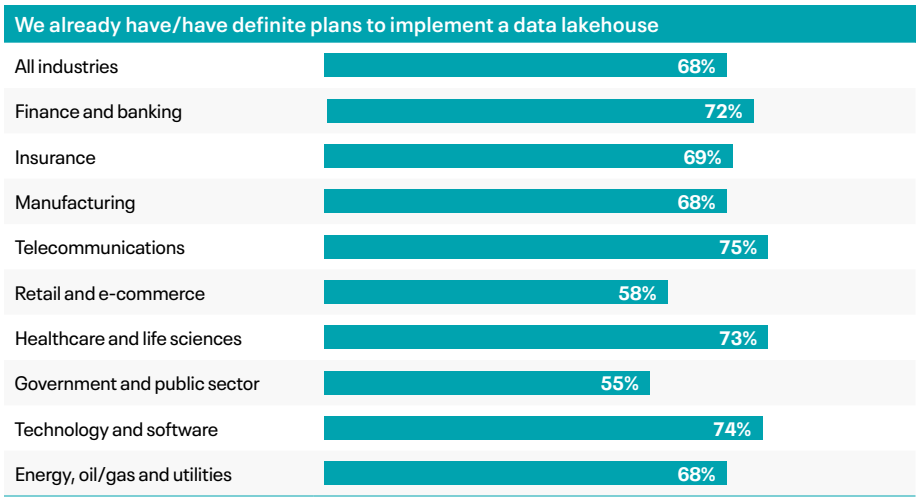

Figure 7. Has your organization already created a data lakehouse or is it considering/planning to create a data lakehouse? We have already implemented a data lakehouse/We have definite plans to implement a data lakehouse.

Enabling faster access to both structured and unstructured data (62%) is the main reason organizations are looking to adopt lakehouses, followed by increasing data collaboration across teams (55%) and supporting a modern data architecture (50%). With a growing number of organizations looking to ride the AI wave, more than a third (35%) of ITDMs say they are implementing a data lakehouse to underpin their AI and machine learning strategy. As more AI use cases emerge, we'd expect data lakehouse adoption to accelerate as organizations look to harness more of their data.

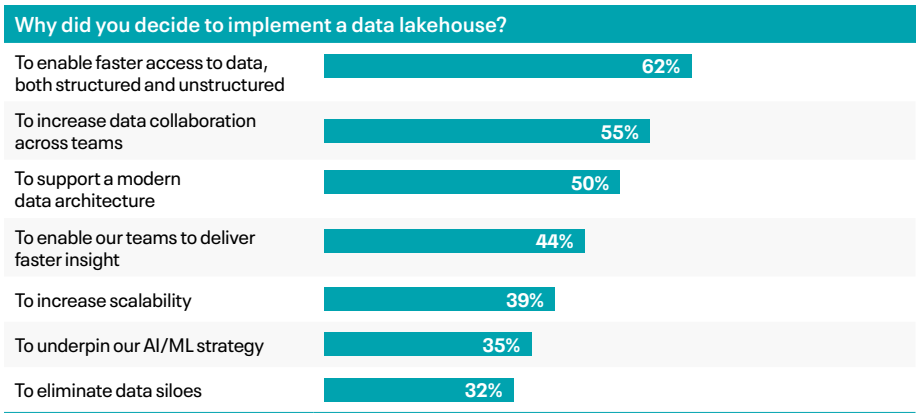

Figure 8. Why did you decide to implement a data lakehouse?

With the volume and velocity of data being generated by an organization today, a lakehouse's single source of truth will enable organizations to drive more value from the full data set. With AI becoming more prevalent, lakehouses will play a crucial role in powering innovation.

Conclusion

The speed at which organizations are finding innovative ways to utilize their data is matched by that of new data regulations, and compliance challenges will continue to grow. Organizations need to be prepared not just for today, but for regulations coming in the next few years. So, taking control of data must be at the heart of governance.

This can only be achieved if governance is always and everywhere. Organizations must develop a set of consistent, global policies that can be applied to data in any environment.

Alongside this, the cost of managing data across its entire lifecycle is soaring as point solutions and the need for re-skilling impose an integration tax on many organizations.

Those organizations that are able to unify their data on a single platform that spans cloud and on-premises environments will be in a position to thrive — cutting costs and enabling continued innovation by utilizing emerging technologies such as AI.

Compliance Control and Abolish Integration Tax

At Cloudera, we manage more than 25 terabytes of data on behalf of customers, helping them to manage and analyze data of all types — machine data, structured data, transactional data, and unstructured data — with data anywhere. In clouds like Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform (GCP). In on-premises data centers. And at the edge where machine data originates.

We enable organizations to manage data across its entire lifecycle, and define and enforce organization-wide policies to ensure compliance on data, no matter where it resides. We deliver streamlined analytics via a lakehouse, without the need for reskilling and multiple tools for data analytics.

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 an enterprise data cloud for any data, anywhere, from the Edge to AI. Powered by the relentless innovation of the open source community, Cloudera advances digital transformation for the world's largest enterprises.

Learn more at cloudera.com