# THREE KEY FACTORS FOR GAINING CONTROL OVER YOUR DATA LIFECYCLE

## Discover How Consistent Data Context Transforms How You Use Your Data

METADATA

GOVERNANCE

SECURITY

## Table of Contents

## The Challenges of Becoming Truly Data-Driven

Your digital transformation begins and ends with the data lifecycle. If you still don't have full control and visibility over every aspect of your data, you won't have the capabilities you need to drive innovation. It's that simple.

Your organization has already transformed itself in myriad ways. You may have married your on-premises systems with public and multi-cloud environments. Perhaps you've even employed machine learning and AI technology to help analyze data at the edge. But these efforts will suffer if you don't have the capabilities to locate, access, and harness that data.

This is a common problem. Although 93 percent[1] of enterprises have a multi-cloud strategy, many of them still struggle to gain a complete view of their data. There are many reasons for this. Now that data lives across multiple environments, it can be subject to an array of different management policies, making comprehensive analysis challenging. There are also the constraints of data privacy laws across different regions, the security needs of different teams and divisions, and the confines of legacy systems. The result is a fragmented, siloed data landscape that impedes insights, slows business, and raises costs.

Turning yourself into a true data-driven organization instead means establishing consistent data security and governance policies — context for short — across every part of your data lifecycle. With that in place, you can empower your users on all deployments, allowing them to work with any data, in any form — wherever it lives.
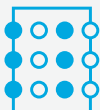
Although 93 percent[1] of enterprises have a multi-cloud strategy, many of them still struggle to gain a complete view of their data.

## Three Steps Toward Consistent Data Context

While overcoming the challenges of making data available, at scale, to users across multiple environments in a safe and compliant manner may seem daunting, it comes down to three fundamental steps. Put these in place, and will be able to give users the data they need to create the insights your business demands.

### 1. Create Visibility

You first need to know what data you have. This is essential to delivering data-driven insights. Yet factors such as siloed platforms and the absence of centralized data stewardship all regularly contribute to a lack of data visibility. This is a serious problem. Poor data visibility, according to 95 percent[2] of businesses, is a primary reason for wasted resources, ineffective business initiatives, and bad customer experiences.

Today, organizations face a data storm. More data is being generated than ever before, making it even more challenging to discover, catalog, and keep track of all this information. This means it's all the more important to take action now. Future business success will predominantly depend on data visibility.

### 2. Enable Access

You should give your users access to the data they need. Doing so will ensure you are maximizing your organization's ability to leverage your data. However, accomplishing this means taking into account the complexities of today's IT infrastructures. In other words, you must be able to manage data accessibility across multiple platforms, environments, and geographies, each of which can come with their own data compliance and privacy regulations.

This requires a proactive governance strategy. It's no longer enough to react to regulatory and compliance requirements as they arise. You must instead drive business value through consistent and comprehensive access policies.

### 3. Gain Control

Finally, you need to gain more granular control over the security of your data. Traditional approaches to data security can often vary by system, environment, or types of data. Disconnected security controls can restrict some users from accessing the data they need while inadvertently giving other users access to data that should remain private.

This imposes a burden on IT teams trying to keep security between disparate systems in sync. Accessing data securely then becomes time-consuming for users and costly for businesses.

Instead, you need a specific and focused approach to data security that you can apply across all of your environments and types of data. This will help ensure your data remains both accessible as well as secure, wherever it resides.

## Cloudera Data Platform (CDP) and its Shared Data Experience (SDX)

Cloudera's mission is to empower people to get more value out of their data. Our Cloudera Data Platform (CDP) is how we turn this mission into your reality.

As the world's first Enterprise Data Cloud, CDP is designed from the ground up to work seamlessly across multiple infrastructures, from legacy on-premises systems to hybrid cloud environments. It allows you to easily collect data, automatically categorize it, then use this comprehensive view to manage that data and deploy highly customizable advanced analytics. With its flexible and scalable approach, CDP solves the challenges of managing and maintaining multiple environments, giving users a radically simple way to pull valuable insights out of their data.

Integral to CDP is the Shared Data Experience (SDX) layer. Through SDX, organizations can use CDP to make data available to their end users while also enabling consistent governance and security across their entire ecosystem. To put this another way, you no longer have to choose between balancing security and usability. It's how we help you take control of your data and digitally transform.

This whitepaper will take a closer look at the capabilities of SDX and how you can use it to gain consistent data context across your entire data lifecycle.

**METADATA**

Between 2020 and 2021, worldwide data production will grow by an estimated 25 percent[3].

SDX goes beyond the typical schematic metadata and profiles information based off of three broad categories:

- **Operational Metadata:** Where data is coming from and where it is going
- **Social Metadata:** What data people are saying about the data and its applications
- **Business Metadata:** The business application and meaning of data

## Create Visibility Across Your Ecosystem With Metadata

Optimizing your data lifecycle, powering data-driven decisions, and delivering consistent data context across your environments all starts with visibility. Simply put, you cannot digitally transform if you do not know what data you have.

While this may sound obvious, it is a primary challenge for modern organizations. We are in the midst of a data storm, collecting and consuming more information than ever before. Between 2020 and 2021, worldwide data production will grow by an estimated 25 percent[3]. The sheer volume of this information poses significant challenges to visibility — but there are also the complexities of knowing which data is most relevant and useful. You need the ability to discover what's in your data in order to build trust in it.

### Metadata Is the Key to Visibility

Metadata has long been associated with data's technical and structural attributes, such as tables and fields. While this is useful, metadata encompasses so much more. You can capture rich profiles of your data using attributes that describe where that information originated, the pathways it took, which other users have touched it, what business applications that data has, and more.

In addition to technical and structural metadata, SDX enhances data with insights across operational, social, and business metadata.

Using this rich metadata, you have complete visibility into the data across your deployments. This forms the basis that lets you intelligently manage access to your data through governance policies, as well as build sophisticated security controls.

### Build Trust, Produce Insights, and Create Value

Data visibility is more than an organizational benefit. The ability to quickly become familiar with multiple aspects of your data can also create tangible, long-term benefits throughout your organization. Let's use the three categories of metadata employed by SDX as an example.

- **Operational metadata builds trust.** If you know where your data originated, you may be able to determine if it is a good source of information for your use case. Likewise, knowing when data was created or refreshed will allow you to select the most up-to-date information for your needs.
- **Social metadata lets you build on this trust with community knowledge.** Discover what data sets other people found useful, what they had questions about, and much more. Just like a good product review, you can quickly determine what will likely lead to the best insights.
- **Business metadata helps you use the right data for the right analytics.** You won't have to figure out what an abbreviated field name means, guess the purpose from a data sample or translate another language. You'll know exactly what the business relevance of a data set is, saving you time and creating additional value.

### Your Data Should Find You

Metadata — and the visibility it creates — is ultimately just a building block for an entirely different way of looking at your data. Rather than spending time searching through reams of information, it should all be so accessible that the right data comes directly to you.

SDX uses metadata in a myriad of ways to do just this. As your platforms ingest information, sensitive data such as credit card numbers, is automatically detected. Assigning the appropriate access policies to this information can be done with a single click, preventing potential privacy issues. You can also tag and classify data assets so that they're easier to find. And all of this data can be viewed and organized within SDX's comprehensive Data Catalog. This gives you a single pane of glass to discover the data most useful to you.

## GOVERNANCE

**Proactive Governance Means More Value**

Turning your governance approach proactive can do more than just create immediate business value. It can also open your organization up to more exploratory and creative ways to create long-term value. Let's take a look at three possible scenarios.[5]

- **Focused Value Discovery:** This is when you enable data access in order to solve a specific business challenge or use case, such as why an individual customer left. It uses data to refine and optimize your insights.

- **Insight and Value Prospecting:** This scenario is slightly broader. It refers to when you allow users to look at data in order to solve for an entire domain, such as customer churn.

- **Free Data Experimentation:** The most unstructured approach, this scenario involves giving users (such as data scientists) access to a wide set of data simply to experiment with. The goal is for them to stumble across an entirely new use or insight.

Although free data experimentation can potentially create tremendous value, most organizations will actually benefit from smaller scale, lower value data initiatives that can be carried out more frequently. However, free yet secure data access enables organizations to explore each scenario for added value. This means that, in many ways, becoming a data-driven organization depends on good governance.

## Enable Access on Your Terms With Consistent Governance

Users need to have access to the right data to produce the insights your organization needs. But you must also comply with data regulation and privacy rules. As IT ecosystems have spread from on-premises platforms to hybrid and multi-cloud environments, keeping up with these compliance rules has become increasingly difficult. Compounding the challenges, many organizations have made data governance an afterthought: In one survey, only 12 percent[4] said they had fully implemented a governance program.

This can be counterproductive. Without a consistent and proactive approach to governance, organizations end up treating each regulation as its own project. These efforts are often redundant and time-consuming, requiring tens of millions of dollars and coming at the cost of impeding innovation.

### Governance Should Be Simple and Accessible

Regulations vary according to multiple factors: industry, geography, even type of data. Recently, as seen with the EU's GDPR and California's CCPA, they have also become more encompassing. But your data governance strategy doesn't have to be complex. Using metadata, you can create simple, dynamic, and consistent governance policies across multiple environments.

For starters and because it profiles all incoming data, SDX already gives you the built-in capabilities to recognize sensitive data and classify it accordingly. SDX also keeps track of where data is going so that you can rest assured compliance rules remain enforced. With this level of visibility, you can create a comprehensive governance structure for all of your data and automatically enforce it across your entire deployment. SDX's data catalog then gives you a single pane of glass to track your data and monitor who has access, helping you prove compliance.

### Turn Governance Into an Opportunity

As your data visibility increases, applying governance policies gets much easier. For example, you can locate where a certain piece of data is, demonstrate that the right people have access to it in the forms that they should, and quickly change your access policies in case they do not. This turns governance from a reactive strategy into a proactive one. More importantly, this allows you to look for and create new opportunities for insights.

What does this look like? Instead of limiting access to data to comply with regulations or privacy laws, you can begin thinking about that data in terms of who can get the most benefit from it. This allows you to eliminate data silos by both giving users access to the specific data they need and offering them open access to data, in a safe and compliant manner, so that they can explore on their own. Eventually, the net effect of this will be to present your users with vital context that allows them to better understand and trust their data, leading to better insights for your business.

SECURITY

## Gain Control Over Your Data With Intelligent Security

Traditional approaches to data security can be cumbersome and ineffective. They typically rely on role-based access controls (RBAC), which give certain individuals or groups access to different types of data. If you need to view data outside of your group, you will have to rely on your IT team to grant permissions. This can take months, slowing innovation, raising costs, and increasing risk of shadow IT. You need a better way to control access to your data across teams, locations, and environments.

### Turn Security On its Head

RBAC alone is often an inefficient approach to data security. Instead, by using metadata to classify your data and enable access, you can take a completely different approach within SDX. Called attribute-based access control (ABAC) or fine-grained access control (FGAC), this security strategy lets you control who sees data based on the attributes of that data, rather than the predefined role of a particular user. This allows you to determine not only whether data is seen by different users or groups, but what aspects of that data they can see — all without ever altering the original data source. It's a more progressive, dynamic approach to security.

### Get Granular Control of Your Data

Approaching security in this way opens up entirely new possibilities for controlling and managing access to your data. Because access has moved from the user to the data, it becomes possible to layer different access controls on top of one another. This can be as straightforward a rule as:

• User X can access this data only when they are in the office.

But this can quickly become more granular as additional layers are added on:

• User X can access this data in its natural form when they are in the office between the hours of 9 a.m. and 5 p.m. Outside of those times (but still in the office), they can see a masked version of the data.

Access to both individual pieces of data and entire groups can continue to be altered as security needs change. SDX will ensure these permissions are standardized and seamlessly enforced across all deployments.

### Intelligent Security Saves Resources

It pays to take a smarter, more focused approach to security. Without ABAC, for example, organizations may resort to spinning up new clusters in order to implement new data access policies. Access to those clusters is then controlled as a means to prevent sensitive data from being seen by some users. And all these clusters take up valuable resources, making it entirely possible for an inefficient security policy to become a significant part of the budget.

But by using metadata to gain more control over your data through ABAC security policies, your IT department can significantly reduce the number of policies they need to manage. Access to data across your deployment suddenly becomes much easier to control, while overhead decreases. It goes to show that good security is not just about making sure the right people have access to the right data. It's also about being more efficient.

## Conclusion

Digital transformation is a journey. It takes time and there will likely be a few wrong turns, but you should never lose sight of your data along the way.

As an integral layer of CDP, SDX allows you to optimize your data lifecycle from end to end while ensuring you always have a handle on your data. With deep metadata insight, strong security, and end-to-end governance, all working as one to deliver a consistent data context, you can streamline control of data across multiple environments. This lets you achieve advanced analytics that lead to better, more actionable insights.

SDX brings you closer to your data, no matter where it is. Learn more about how CDP and SDX can help you transform into a true data-driven organization.

**Sources**

[1] Flexera, "2020 State of the Cloud Survey," 2020

[2] Experian. "2019 Global data management research report," 2019

[3] Statistica, "Total Data Volume Worldwide 2010-2024," Holst, Arne, 23 November 2020

[4] Erwin, "2020 State of Data Governance and Automation," 2020

[5] 451 Pathfinder, "Accelerating Outcomes with Data Governance," Bartley, Paige, November 2020

**CLOUDERA**

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