

CLOUDERA

How SDX accelerates business insight



Table of Contents

Finding Success in the Midst of a Data Storm	3
Take Control of the Data Journey with Consistent Data Context	5
SDX Overview: Use Your Data and Use It Properly	6
Business Use Case #1: An Automatic Mechanism that Scales and Accelerates the Productionization of Data	8
Business Use Case #2: Creating Self-Service Access Without Causing Headaches for IT or Compliance	9
A Case Study in Self-Service Access: How a Financial Services Provider Gained \$76 Million in Annual Deposits	11
Business Use Case #3: Proving Compliance	12
Take Your Next Step—and Control Your Data’s Journey	14

Finding Success in the Midst of a Data Storm

A deluge of data now floods your organization. If you can unlock the insights this data holds, then you can become truly data driven—and undergo digital transformation.

Today's most innovative organizations harness the vast amounts of data at their disposal. They efficiently put data to work for insight and value across every layer and facet of their enterprise—from the bottom to the top, from business users to IT teams. If you hope to do the same and unlock the value of your data, you must first gain control over its journey. This means confidently ensuring governance, security, and control of information no matter where it originates, lives, or moves—so that your organization is free to rapidly derive insights, experiment, and deploy new business use cases at scale.

To become a data-driven organization, you must be able to ingest, enrich, dashboard, and predict (via machine learning models) information at scale. You must also learn that the data journey isn't linear. Instead, it's a fluid, omni-directional path that enables the free movement of data to whichever analytics or users that unlock its value.

Unfortunately, if your organization is like most, your ability to control the data journey remains hindered by the isolation and fragmentation of siloed systems and applications.

A large, stylized graphic of the number 93% in a light blue-grey color, with diagonal hatching lines inside the digits.

of enterprises say they have a multi-cloud strategy.¹

A large, stylized graphic of the number 87% in a light blue-grey color, with diagonal hatching lines inside the digits.

of enterprises say they have a hybrid cloud strategy.¹

¹ Flexera, "Cloud Computing Trends 2020," May 2020

A Hybrid World on a Fragmented Landscape

While 87% of organizations say they have a hybrid cloud strategy,¹ having a plan to leverage the combined compute and storage resources of different environments is far from enough. Multiple analytic workloads should be able to run against the same diverse data sets for a myriad of business applications. But such workloads—like capturing and ingesting data at the edge in real time or using data warehousing for dashboards and analytics—are bad at sharing the same data. Each workload often exists in its own closed ecosystem—and not just in terms of having its own data store and dedicated compute and processing power; systems also have their own framework for managing data. The only way to achieve any consistency across these systems is through manual synchronization, a task which is difficult enough in a homogenous infrastructure, but one that becomes increasingly complex the more an organization expands across hybrid and multi-cloud environments.

Over time, what you're left with is a fragmented IT landscape fraught with inefficiencies and high overhead costs.

The Fallout of Self-Contained Silos

Your data systems operate as self-contained silos. Each one has its own security, governance, and control needs and frameworks. And the unenviable task of keeping these disparate silos in sync falls to members of your IT team; they have to manage data across infrastructures, keep up with compliance, and provide business users secure access to data. Many IT teams are also forced to work within the confines of legacy systems and proprietary lock-in, and they simply can't keep up with business demands.

And if IT teams are too bogged down by the existing landscape now, just imagine how they'll be impacted by the torrent of new data coming in the months and years ahead.

The Business Impact of Inconsistent Governance, Security & Control

These challenges of maintaining security, governance, and control across fragmented and legacy systems ultimately result in delayed business insight, increased operational overheads for IT and business users, and greater security risks.

Meanwhile, departments continue to adopt their own tools and engage in shadow IT. Though done with the best of intentions (i.e., trying to get their job done in the most productive, efficient way possible), these departments are still circumventing efforts to enforce security and governance. After all, as CIO Magazine notes, "Enterprise IT doesn't operate the speed of business."² The resulting hodgepodge of point solutions eventually leads to a massive integration tax across the organization which hampers the ability to gain insight and value from data quickly.

Take Control of the Data Journey with Consistent Data Context

To be among the innovators, you have to empower different users to work on the same data, with potentially different tools—while adhering to regulations and ensuring consistent security.

An Introduction to Cloudera Data Platform (CDP)

CDP is the world's first Enterprise Data Cloud. It's a single platform that runs flexibly across multiple clouds and data centers, and it uses advanced analytics to help you navigate the full data lifecycle, letting you find value and insights from data at scale. CDP also gives you access to both highly customizable analytics services as well as ready-made end user experiences on a 100% open platform. It's built on open source technology and open for integration with your existing systems and applications.

SDX: Built-In Security & Governance That Brings It All Together

The Shared Data Experience (SDX) is the security and governance layer of CDP that gives you control over data throughout all platform deployments. This always-on service is an integral component

of CDP and a fundamental part of its architecture. It takes data access policies defined in one deployment and applies them seamlessly across every other environment and workload. So, multi-tenant data access and governance policies can be set once, and then automatically enforced across hybrid and multi-cloud environments.

In other words, you preserve a single data context—which is everything pertaining to its metadata, security, and governance—as that data is used in different workloads, moved across heterogeneous environments, and accessed by business and IT users.

This ebook will focus specifically on SDX and its role in CDP as a consistent data context layer. Read on to learn how the consistency of SDX helps your organization derive greater business value and insight from your data.

SDX Overview: Use Your Data and Use It Properly

SDX makes it possible for CDP to deliver seamless data security and governance everywhere—across your complete range of platform deployments and analytic workloads. It works as CDP's integrated set of security and governance technologies built on metadata, enabling CDP to maintain the context of your data consistently across deployments.

SDX is also the underlying service of Cloudera's Data Catalog, which provides a single pane of glass that grants comprehensive control and use of complete data context. You not only know what data you have, but you also know that your data is being used properly across the organization.

What Your Business Gains with SDX

- **Greater Agility**
Easily deploy consistently secure data lakes, so your organization can act quicker on insights yielded from more timely and relevant data.
- **Better Insights**
By granting free data access for all, departments across your organization can experiment more, iterate faster, and deploy more use cases to drive greater insights.

- **Improved Governance**
With a built-in security and governance layer that expands across hybrid deployments, your organization can achieve and maintain regulatory compliance with ease.

What Your Business Reduces with SDX

- **Operational Costs**
You have a single pane of glass to manage security and governance across all deployments. This lowers your cost of operational and staffing overhead.
- **Security Risks**
SDX delivers comprehensive controls for multi-tenant data access. Consistent policies ensure users and groups only have access to the data they are entitled to use.

The 5 Key Capabilities of Consistency

To drive consistent data context anywhere, SDX was built to deliver these five features:

- 1 Robust Data Catalog**
SDX includes a Data Catalog, which provides a single-pane view to manage and discover all data assets. The data is profiled and enhanced with rich metadata, including operational, social, and business context. This creates trusted and reusable data assets, and makes them discoverable.
- 2 Comprehensive Security**
Easily create, manage, and maintain multi-tenant data access policies through standardization and seamless enforcement of granular, dynamic, role- and attribute-based security rules. These capabilities eliminate business and security risks, and it ensures compliance by preventing unauthorized access to sensitive data as different groups apply various analytics.
- 3 Enterprise-grade Governance**
Identify and manage sensitive data, and effectively address regulatory requirements with unified, platform-wide operations, including data classification, lineage, modelling, and auditing. Prove compliance throughout the complete data lifecycle and from ingestion to archive/purge with data management across all analytics and deployments.
- 4 Full Encryption**
Resources are deployed and automatically configured to use Kerberos, with Auto-TLS encrypting data as it transits the network. Data at rest, both on-premises and in the cloud, is protected with enterprise-grade cryptography, supporting best practice tried-and-tested configurations.
- 5 Control across hybrid cloud**
SDX enables infrastructure independence with the ability to intelligently move workloads as well as data, together with its context between on-premises data centers and multiple public clouds. Insights from key metrics like workload performance deliver smart recommendations for optimal resource utilization, balancing performance, cost, and resilience.

Business Use Case #1

An Automatic Mechanism that Scales and Accelerates the Productionization of Data

The world is more virtual than ever. Streams of information and data are hitting organizations at increasing volumes. And with the advent of technologies like 5G, the amount of data that's being generated will increase faster still.

You need to make data available to your end-users as fast as possible, so they can derive insight and value from it. But traditional approaches to data ingestion and preparation can't get it into the hands of those who need it fast enough: more data needs more time. The challenge lies in being able to categorize, classify, and ingest high volumes of data in a way that makes it secure, governed, and usable.

SDX Lets You Productionize Your Data and Make It Rapidly Accessible

Automatic Data Profiling

SDX automatically profiles new data as it is captured. Every single stream of data is profiled at volume in a way that lets you gain immediate insight into a dataset's makeup, characteristics, quality, and more.

Sensitive Data Recognition

Part of SDX's automatic data profiling function is its ability to recognize sensitive data—passport information, bank account numbers, social security

numbers, etc.—and classify that data accordingly. So, if you're ingesting data that must adhere to regulations such as HIPAA or GDPR, then that data is tagged and classified as such.

Attribute-Based Access Control (ABAC)

SDX lets you create and manage data access policies based on data classification. It does this thanks to ABAC, which gives your organization a more dynamic way to use and interact with sensitive data versus traditional role-based access controls. Data classifications are the trigger for how the information is presented to users and groups, preventing unauthorized access to sensitive data. With ABAC, data users across your organization can access the same data needed to run their analytics, but they never see and use the sensitive aspects of it unless they're authorized to do so.

Putting Your Data into Production with SDX

The more your data volumes grow, the greater your need to faster understand data becomes. Fortunately, SDX automatically profiles new data and classifies sensitive information. It also applies the appropriate access policies for compliance, so your users can quickly capitalize on the flood of information coming into your organization.

Business Use Case #2

Creating Self-Service Access Without Causing Headaches for IT or Compliance

How can you make data accessible quickly to the users who need it—while still ensuring consistent security and governance?

With traditional methods, it can take months for IT teams to provide business users with the data they need to perform analytics. This time-consuming process not only throttles innovation, it also leads to increased IT costs and greater security risks. That's because business users aren't afraid to adopt shadow IT methods and solutions in their effort to move quickly and make the best use of the data they have at their disposal while it's still relevant.

To drive insight at scale, you need to give your end-users access to data for experimentation, so they can drive new use cases and derive more insight and value for your organization.

SDX Enables Self-Service Access to Data Across Your Organization

Rapid Data Propagation

SDX lets platform administrators and architects create and manage access policies. They're able to deliver the same data to different users or groups—either as-is, obfuscated, or hidden—based on the interaction of classifications and ABAC. As a result, different users can now see data the way they're entitled to see it—eliminating the need to copy data or set up new clusters.

If data does move, however, SDX makes sure its context (its access policies and governance rules) stays with it. This clears the way for you to rapidly propagate data across the organization in a controlled manner. Data can be sent where it's needed quickly, including all the way to any machine learning models you have in production.



Shadow IT investments can exceed 40% of the formal IT budget.³

³ Gartner, "Integrate 'Shadow IT' and Business-Led IT Into the I&T Operating Model to Enable Enterprise Agility," Simon Mingay, et al, 9 August 2018

Self-Service Access that Doesn't Shortchange Security or Governance

SDX maintains total data lineage across all analytics—from edge to AI—which can be inspected and audited at any time. Data lineage and classification are also part of the context SDX builds around your data.

This data context propagates along the data lineage and so continues to apply to derived data. This ensures the right rules remain applicable as insights are formed across analytics but also across clouds and data centers.

SDX delivers consistent security and governance for multi-tenant data access across all deployments, enabling free access while ensuring safety and compliance. This means your organization can give more business users greater access to data for better, data-driven decisions—while also maintaining accountability and ensuring regulatory compliance.

A Case Study in Self-Service Access

How a Financial Services Provider Gained \$76 Million in Annual Deposits

Banco de Crédito del Perú (BCP), the leading supplier of integrated financial services in Perú, wanted to create a web-based business intelligence tool that would let their clients view KPIs on their business, their customers, and even their competition. To provide this service, the tool would need to analyze a wealth of anonymized financial and transactional data, including 50% of all national credit card transactions.

Faster Business Insights with Self-Service Access

The financial services provider used Cloudera to unlock 90 terabytes of data from its mainframe and data warehouses. Analysts were able to quickly access this wealth of data via its web-based tool in a way they couldn't before, turning previously unused information into a business asset while simplifying security and governance across environments.

As a direct result of making secure, governed data rapidly accessible to analysts, timely insights were derived that enabled to BCP increased its annual deposits by \$76 million.

[Read the full success story.](#)

**"We've turned information into a
business asset that delivers
substantial value to our clients."**

Erick Jaramillo Salazar
Manager, Big Data Center of Excellence, BCP



Business Use Case #3

Proving Compliance

Regulations vary by industry, but they're becoming much stricter across the board. Costs to ensure compliance with regulations such as GDPR and CCPA are also rising—and even then, one wrong move can still result in heavy fines and loss of reputation due to non-compliance.

Proving adherence takes significant time, money, and effort. And if you're a multi-national organization or otherwise subject to several regulations simultaneously, then overhead for your compliance to-do list becomes enormous.

Your organization needs a way to both streamline and strengthen your compliance efforts—without clamping down on data access or increasing costs. Good governance not only ensures compliance, it also adds

business value and opportunity by making data easily accessible for more data-driven insights.

SDX Applies Consistent Context that Moves with Your Data

Compliance should be an effortless side effect of the platform you're using. And when you use CDP, you gain the power to streamline compliance through the SDX layer and remove many of the manual processes typically associated with compliance initiatives.

SDX helps you identify and manage sensitive data without disrupting business processes. You can define classifications that SDX identifies in your data through profiling—and then you can have that contextualized data propagated from the edge to AI.

Annual Compliance Costs Are Starting to Add Up⁴

\$31M

in compliance costs for financial services firms.

\$8M

in compliance costs for media companies.

⁴ Ponemon Institute LLC, "The True Cost of Compliance with Data Protection Regulations," December 2017

You maintain visibility into the complete data lifecycle at all time and can easily prove compliance with reports that show how data is used throughout its lifecycle as well as when data was accessed (or when was access denied) by whom and which data access policy was involved. This saves your team the trouble of having to start up a new compliance project every time a regulator makes a request.

Keep up with Regulation Updates & Policy Changes

Regulatory compliance demands a modern data architecture, one that decreases business and security risks that stem from the ever-expanding data privacy requirements your organization faces.

SDX identifies and manages sensitive data for compliance without disrupting your business processes. It makes it easier to track and stay up to date with new regulations or even minute changes to existing policies—all on a platform that provides transparent, consistent security and governance across all of your workloads.

Take Your Next Step—and Control Your Data's Journey

Use SDX, the fundamental security and governance layer of CDP, to pick your data, pick your analytics, and create insight faster than ever.

Learn more about SDX by visiting our [website](#) or [click here](#) to take a video tour of CDP and see it in action.

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 | US: +1 888 789 1488 | Outside the US: +1 650 362 0488

Sources

- ¹ Flexera, "Cloud Computing Trends 2020," May 2020
- ² CIO Magazine, "How to eliminate enterprise shadow IT," Peter Bendor-Samuel, 11 April 2017
- ³ Gartner, "Integrate 'Shadow IT' and Business-Led IT Into the I&T Operating Model to Enable Enterprise Agility," Simon Mingay, et al, 9 August 2018
- ⁴ Ponemon Institute LLC, "The True Cost of Compliance with Data Protection Regulations," December 2017

© 2020 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. 0000-001 October 9, 2020

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

CLOUDERA