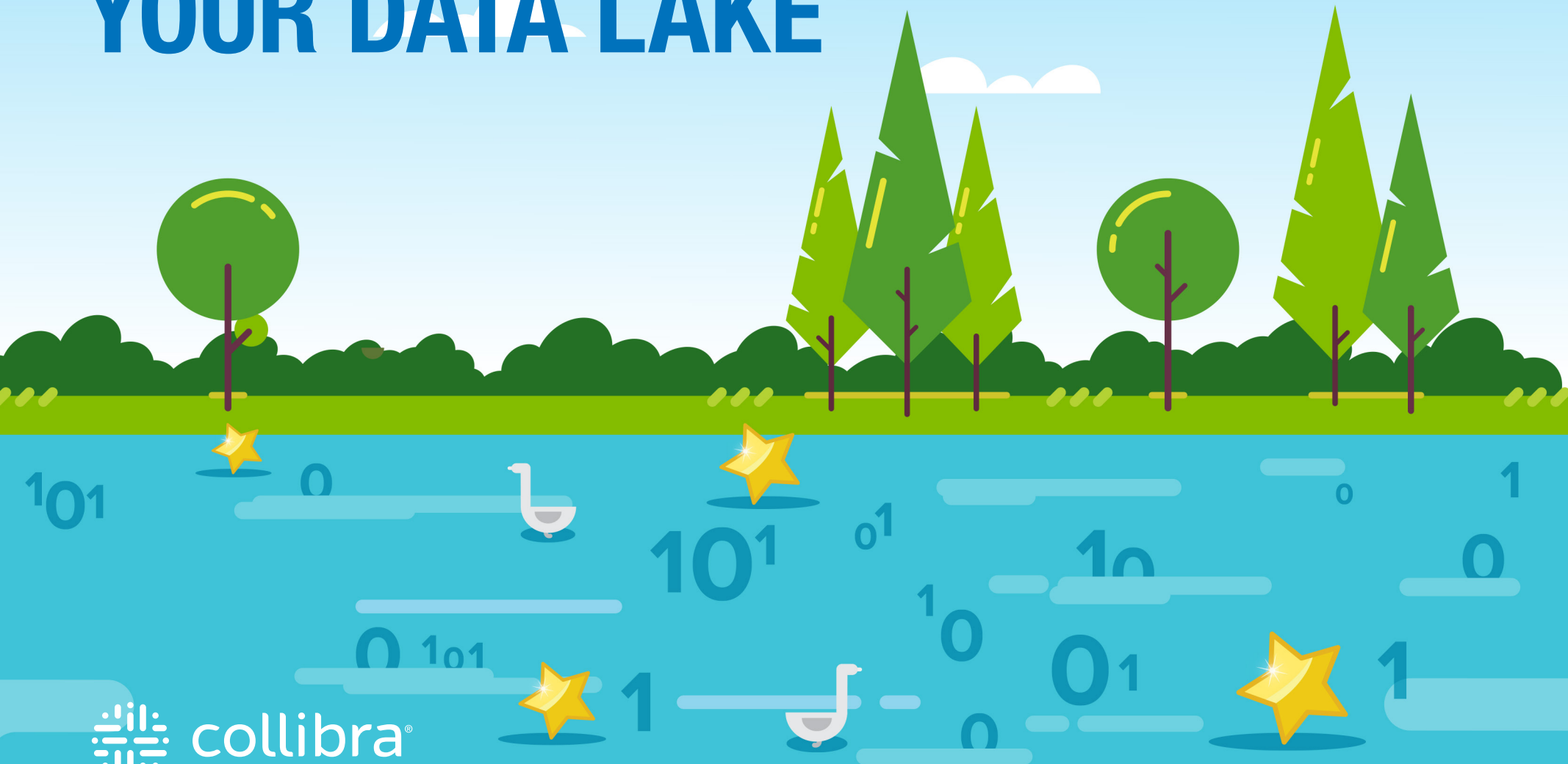


DRIVING VALUE FROM YOUR DATA LAKE



The data lake promise. (It's big.)

Data lakes were supposed to help financial institutions make everything easier. A renewable reservoir of new data and old data. Big data and small data. Fast data and cloud data.

All of it ingested and stored without time-consuming controls. All of it available right when it's needed.

It was the promise of more data, more flexibility, and, ultimately better insights.





Data lakes were supposed to make getting to your data as easy as pouring a cool glass of water. And quench your thirst for information that would drive decision-making for smarter trading and investment strategies, create more personalized customer experiences, and help transform your business. At their best, data lakes were meant to deliver real business value by helping data users across your enterprise find what they needed to drive new insights. At least that's what your CEO was expecting.

The data lake dilemma. (It's conquerable.)

If you've managed to implement a data lake that pulls data from the edge, integrates with the cloud, breaks down silos, and speeds discovery, kudos. We know it wasn't easy.





But here's the hurdle: expectations for your data lake will only keep rising. Pilot programs for regulatory reporting, fraud detection, and other use cases are turning into enterprise-wide initiatives. New—less expert—data consumers such as portfolio managers, risk officers, and claims administrators are diving into the data lake every day. To them, the data lake is a well-stocked pond, a wealth of resources ready for the catching. But too often, they are coming up empty-handed, unable to find and use the trustworthy data they need to drive decision-making and solve real business problems.

You know the data is there.
But if data users can't find,
understand, and trust that data,
it will never be consumed
and transformed into real
business intelligence.

You might have seen the signs:



1

Data users—from experts to business users—are having a hard time finding the information they need.

2

Confusion about what the data means.



3

Data users who can't trust their data because it's unclear where it came from.

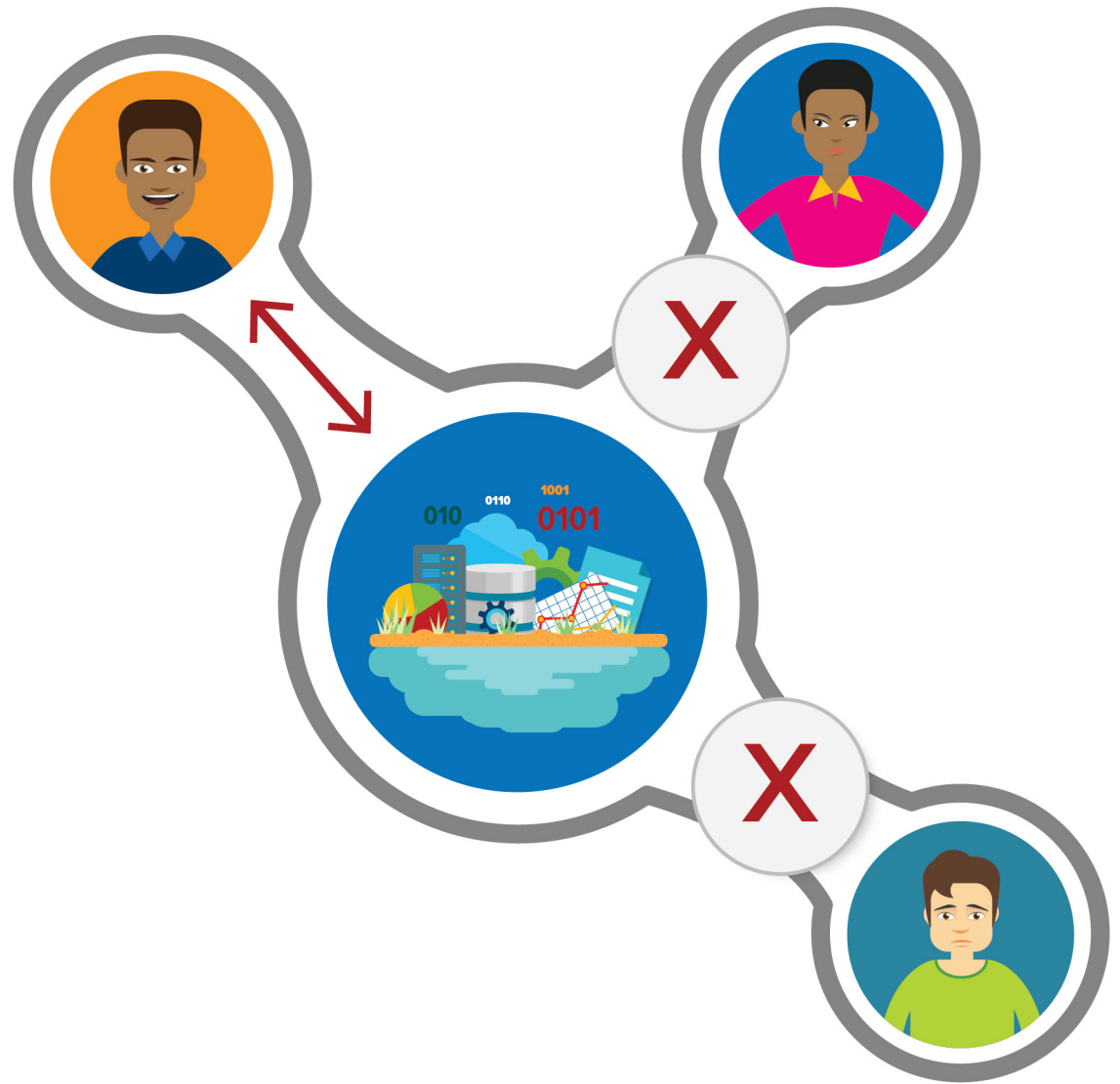
4

Difficulty understanding what data poses a privacy risk.



5

Not enough people across the enterprise taking advantage of data.



Transforming your data lake

Your organization probably had a goal in mind when it implemented its data lake. But if that goal wasn't any more precise than "build us a data lake," you might be having a hard time delivering real value.



The imperative to drive value from your data lake is likely to continue as the need for actionable insight grows.

Data governance adds a layer of intentionality that can ease compliance with regulatory requirements by helping you answer some fundamental business questions:

What data belongs in my data lake?

In the early days of data lake implementations, a lot of organizations answered that question with an enthusiastic, “Everything!” The idea of storing data at scale in a single repository had its allure.





But data collection is not a strategy.

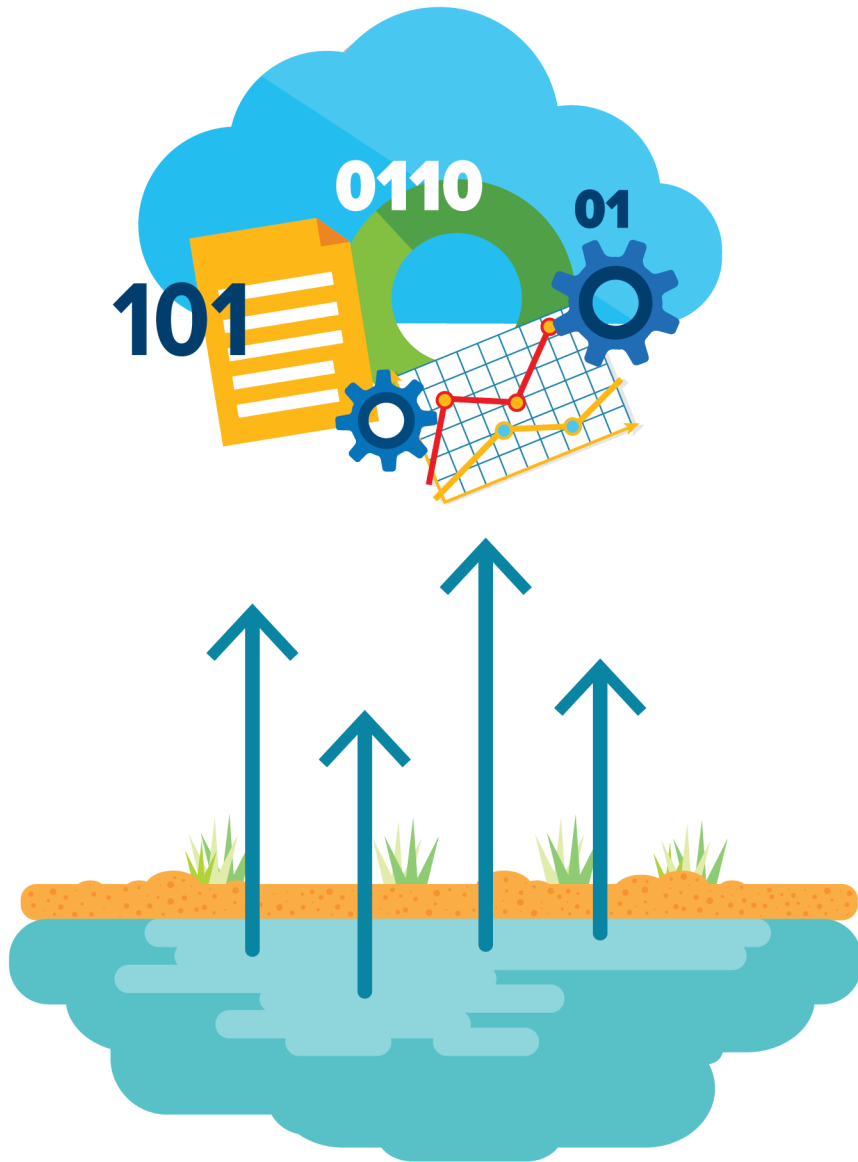
The data lake is a wonderful repository of raw material. But like any natural resource, it needs to be refined to be valuable. Data lakes delivering real value start with agreement about what data actually belongs in the lake. Data stakeholders engaged in a governance project work together with business users to prioritize what's important to the business, and clarify what data is needed to pursue those objectives. And with policies and processes in place for data ingestion, data users will know how to ask for additional data sources to be added to the data lake.

“The key is knowing the subject owner of the data and its lineage. Previously, this was a time-consuming task, and now it's more or less immediate.”

Can my data users find the data they need?

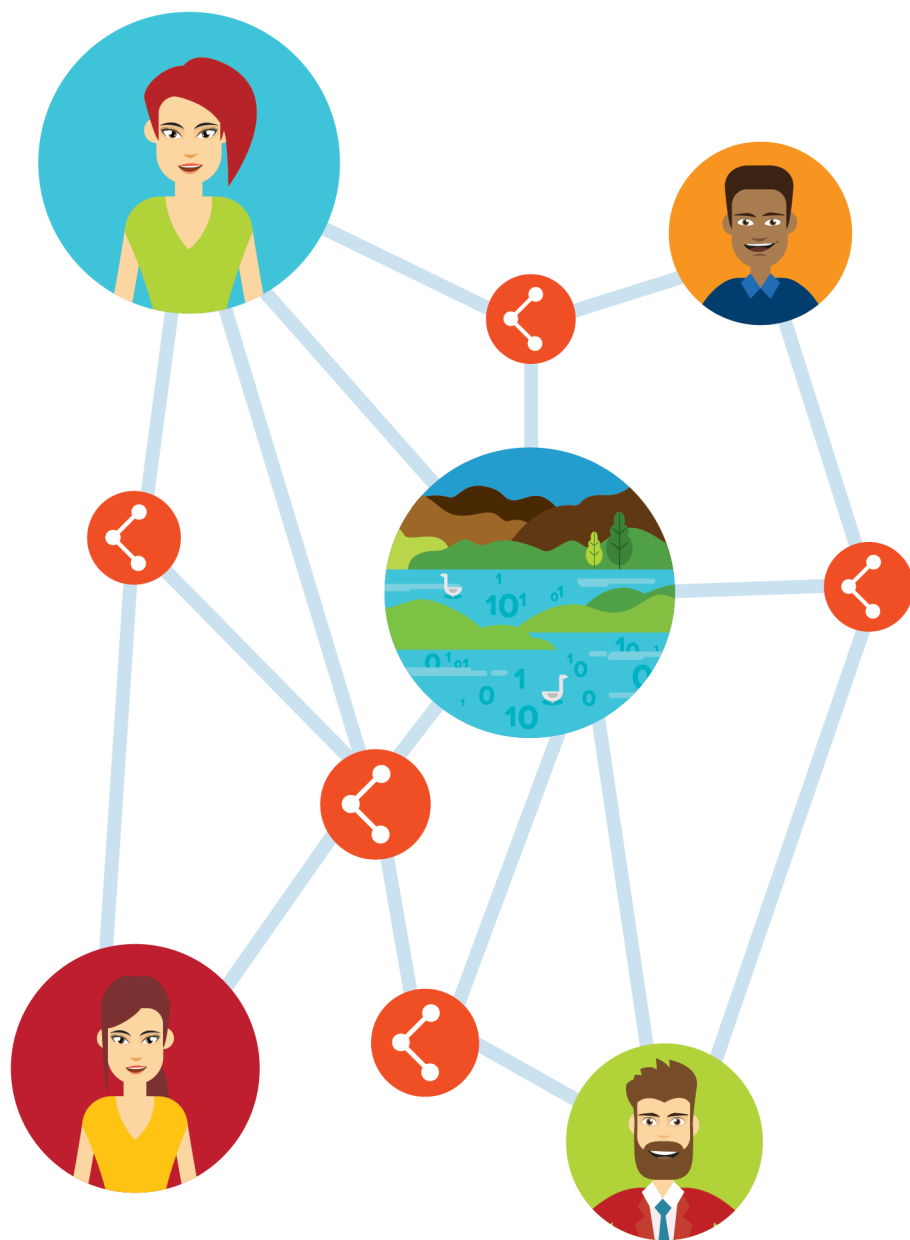
Insight—not data—will move your organization toward its goals. Actionable insight is in demand. But if your data scientists can't navigate your data lake efficiently to create the outputs your business users need, and if your business users can't trust that the information they have is the information they need to answer critical business questions, your data lake is no longer an asset. In fact, if it's slowing down your analytics and preventing new insights from reaching decision-makers, it's a liability.





With governance, data scientists can find any information asset they need quickly and easily: data, data sets, workbooks, analytics, dashboards, even process flows and rules—without sacrificing the agility that makes the data lake so appealing to you. Because the data has been tagged, indexed, and cataloged, everyday data users can “shop” for the data they need, discover related data sets that they might not know about, and even preview sample data to determine its usefulness.

And if you’re considering moving portions of your data lake to the cloud for better scalability and access to cloud-based tools for new machine learning initiatives, governance (along with catalog search capabilities) can help you make sure your data scientists have a complete view of the data they need now—whether it’s in the cloud or on premises.



Do my data users understand what the data means?

In your data lake, all kinds of data from all kinds of sources live side-by-side. Having a common understanding of what that data means and how it should be used is fundamental to driving value. A governed data lake is a collaborative, crowdsourced asset, where users can add business-driven information about any data set, including technical metadata, business metadata, and data lineage. With a common understanding of what the data means and how it is connected, data users can determine which data is fit-for-purpose—and which should be discarded or ignored because it's incomplete or irrelevant.

“Collibra has helped us realize that there are numerous definitions within our landscape for describing the same thing. We realized the need to standardize.”

Do my data users trust the data they're using?

Because data is explained in business terms, your users have a shared language for understanding data's meaning. And because they can see where the data came from and how it's been used, data users will have a much higher level of trust in the data available to them.

That transparency builds trust. And trust drives value. When your data users have insight into the origins of their data, they will make more informed decisions about what data to include in a standard end-of-day trading report or in an ambitious machine learning project to analyze customer sentiment. If you need to restrict access to sensitive data—including pruning select data from the data lake—you can.

Without that trust, data will never reach the broad base of users it should. Sharing agreements will be impossible to uphold, and data will stagnate and deteriorate.



Governance for everyone

To deliver value, a data lake needs governance. But the quality of that governance matters. Especially in a heavily regulated industry that manages large volumes of highly sensitive data. Today, new concerns about data privacy and new regulations like MiFD in the EU and Consolidated Audit Trail in the U.S. exemplify a move towards increasingly stringent regulations that require more frequent and detailed reporting from financial institutions. And while some business leaders recognize that governance is critical to long-term business success, especially as digitalization continues to transform the competitive landscape, many still consider governance as a problem to be solved, not as an opportunity to be seized.



But when it's done right, governance can be transformative. In its purest sense, governance is about connecting people to useful data. It's about removing barriers so that more people across the organization can use trustworthy data to drive decisions. When governance becomes more collaborative, data assets can be refined and improved by people across your organization, turning the raw materials collected in your data lake into trustworthy, actionable information.

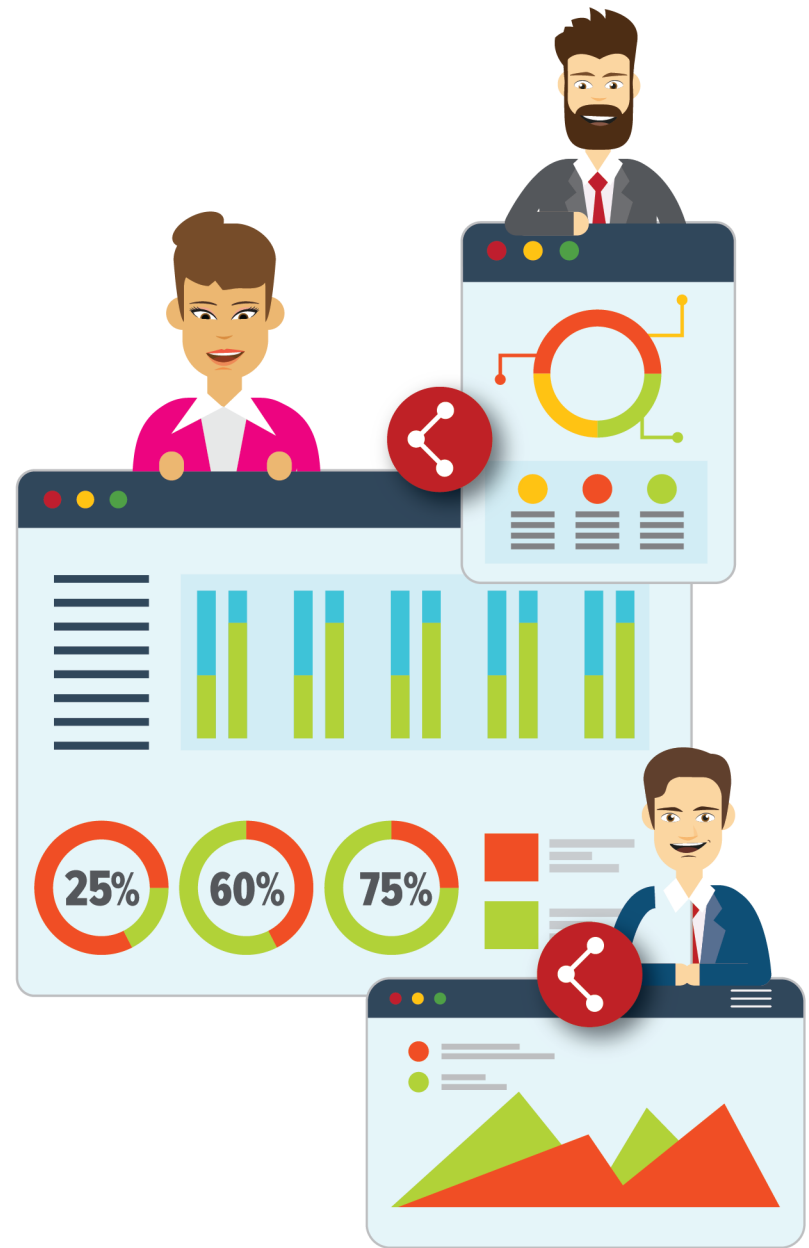


And then something pretty awesome happens.

More people start using the data—the way it was meant to be used. Data sharing agreements extend the power of your data to people outside your organization—with all the right protocols in place. Discovery and sharing tools provide an easy way for people to recommend useful data sets, rate their quality, and tag them for their peers. Transparent workflows help everyone understand who should be using the data and how.

Instead of scrambling to support restrictive top-down policies, your team is orchestrating crowdsourced, collaborative, self-governance that will continue to grow and evolve as the needs of your business change.

“With Collibra, we do a better job of managing and sharing data, both internally and externally. We are finding it easier to share data, and we are more confident in what we have to share.”





Realizing data's potential

Data, big or small, living in a data lake or streaming through the Internet of Things, is simply raw material. Shaping it into a dynamic tool that can drive innovation and transform organizations can only be achieved when that data's meaning and use are clear and when it can be trusted by the people who use that data every day to drive knowledge. With collaborative self-governance, the value of your data lake grows. It becomes a business asset that delivers insights, accelerates decision making, and inspires innovation.

“Our vision with Collibra is to enable our organization to increase the value we derive from our data as a strategic enterprise asset. Our strategy is to create a ‘data aware’ culture where critical data is clearly defined, understood, properly controlled, and accessible as appropriate across the organization.”

Financial institutions are driving value with Collibra on AWS

Collibra is an AWS Advanced Technology Partner and has achieved AWS Financial Services Competency.

Collibra and Amazon Web Services (AWS) enable you to migrate your data, workloads, and regulatory compliance to the cloud. Amazon Simple Storage Service allows easy and secure collection, storage, and analysis of data at massive scale, and Collibra's API integration and platform capabilities provide auto data discovery, profiling, classification, business context, permissioning, and policy management. Amazon Redshift makes data analysis across your data warehouse and data lake simple and cost-effective, while Collibra's integration and catalog capabilities deliver a unified view of data sources. Finally, AWS Glue is a fully managed extract, transform, and load service, and Collibra assists with understanding and trust through lineage and certification, respectively.





Start now.

Connecting the right people to the right data can transform your organization. And your data lake can be the lever that generates value. Help your data users find the assets they need, understand the data sets and reports they find, and use those resources more collaboratively to drive insight. Talk to us today.



collibra®

©2019 Collibra

collibra.com

info@collibra.com

Follow Us

