



Five Keys to Optimize Your Data Lake with Data Governance



Contents

Introduction	03	Conclusion	11
Begin with the End in Mind	04	Further Reading	12
Key 1. Relevance for Personas All Across the Enterprise	05	About Informatica®	13
Key 2. Streamlined Workflows and Processes	06		
Key 3. Adaptability	07		
Key 4. Data You Can Use	08		
A Success Story	09		
Key 5. AI-Driven Technology Can Streamline Manual Processes	10		

Tip: Click on parts to jump to the particular section you want.



Introduction

This is a story about two data lakes. On the surface, they seem identical. Both hold the same volume of data that's been collected from the same variety of sources.

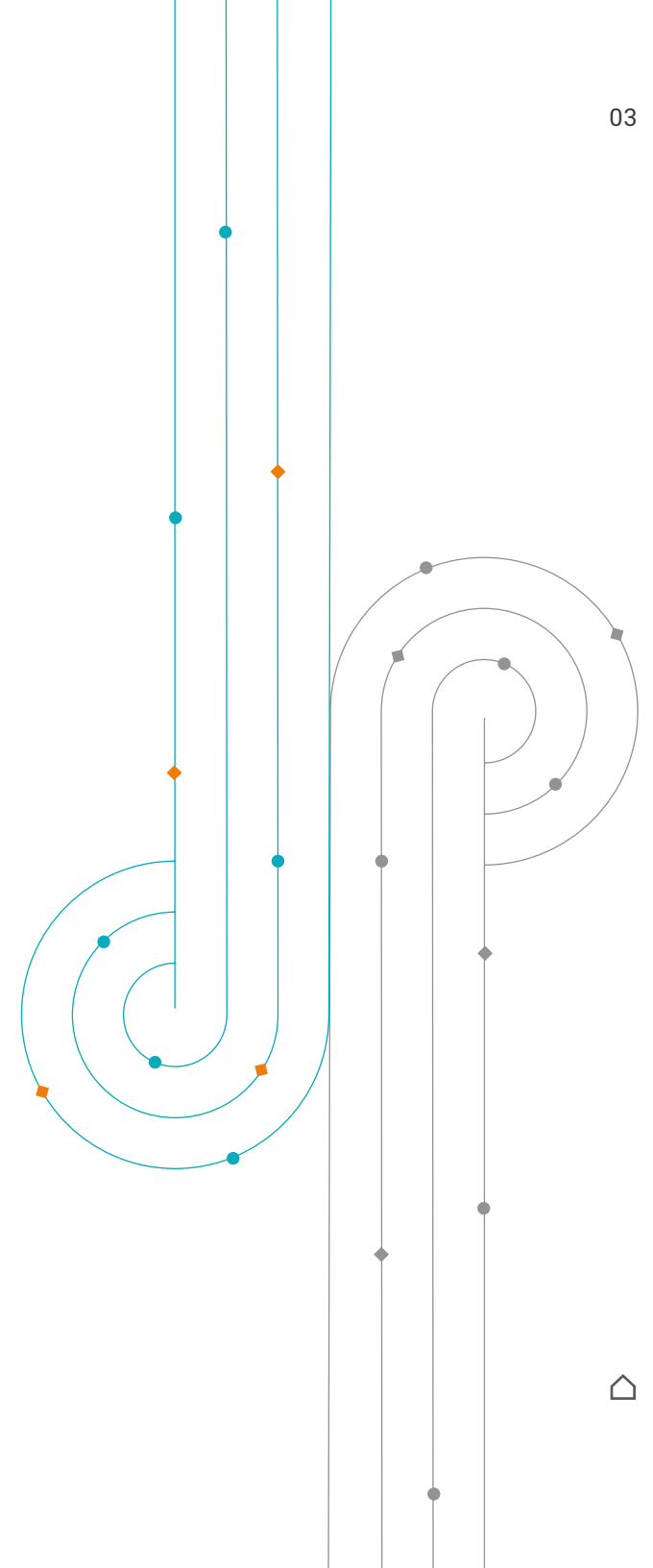
Each data lake is constantly being fed with streams of new data and new data sources. Both data lakes are being used to support agile, exploratory analytics by the same crop of business team members.

But there is one critical difference between the two, and it's just below the surface: one data lake has a data governance program in place. The other data lake does not.

You may know how this story goes: there's the part where the data from the governed lake is useful and trusted and becomes a key business asset that drives positive results for the enterprise.

The data from the ungoverned data lake, however, doesn't fare so well. Data from the ungoverned lake is more likely to generate errors and lead to increased risk, lower chance of reward, and lost opportunity.

To make sure your data lake story ends well, keep reading. We'll show you five ways you can leverage an enterprise data governance program to make sure your data lake becomes a reliable and valuable part of your data ecosystem. Optimizing your data lake with data governance can create the results that matter most to any business: growing revenues and profits, decreasing costs, and reducing risks.



Begin with the End in Mind

Let's start with a quick review of the goals of a well-managed data lake and what happens when you take a thoughtful, well-designed approach to enterprise data governance.

A data lake brings new opportunities to an organization and presents more than just an opportunity to ingest vast amounts of data. Managed properly, a data lake provides an environment to identify and exploit data-driven insights to increase business agility and leverage new opportunities and revenue models:

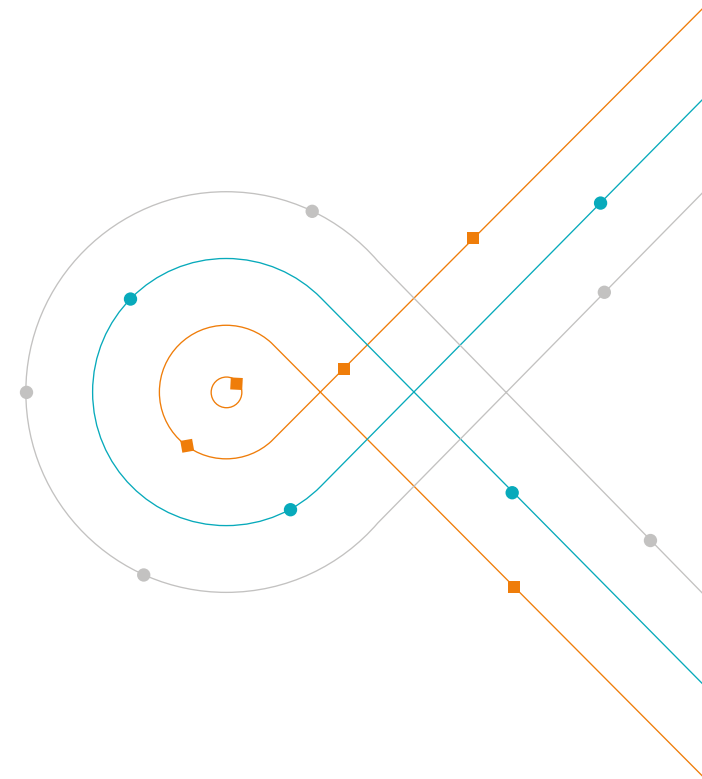
1. Data and business analysts running ad-hoc reports or working with engineers to operationalize reports and analysis
2. Streaming analysis of large volumes of data coming in at high speed, which would become irrelevant if not processed quickly
3. A well-established way of using data from

data warehouses, which comes from data lakes to run business reports and business intelligence in real time

4. A process to prepare and enrich data in the lake so that it's ready for artificial intelligence (AI)/machine learning (ML) modeling
5. Enterprise data prep, preparing and enriching data in the lake so it's ready for advanced analytics

Before you can accomplish any of these end goals, however, you need to be able to govern the data in the lake and keep the data lake from becoming a data swamp.

Data governance is the management of practices and processes that ensure the quality, availability, usability, integrity, and security of enterprise data assets—both on premises and in the cloud.



Key 1. Relevance for Personas All Across the Enterprise

By supporting and facilitating seamless and fluid collaboration between IT and business end-users, data governance should provide role-relevant experiences for all users.

Automated mapping of technical metadata to business terminology supports future scaling, regardless of the ever-expanding volume and diversity of data; mapping that is tailored to different personas accelerates deployment which would otherwise require time-consuming (or completely impractical) manual processes.

This relevance means that data can be used by multiple personas across the enterprise, including:

- Data consumers (such as data scientists, data engineers, and data stewards)

- Information consumers (business functions that use or produce analytics, reporting on compliance, security, privacy, and insights needed by executive personnel)

- Subject matter experts who create or add value to data sets

[Find out if you're ready to reimagine data governance.](#)



Key 2. Streamlined Workflows and Processes

The right data governance solution is just one part of a comprehensive suite of solutions and systems that manage, optimize, and leverage data of all types, from diverse sources, for all types of end users.

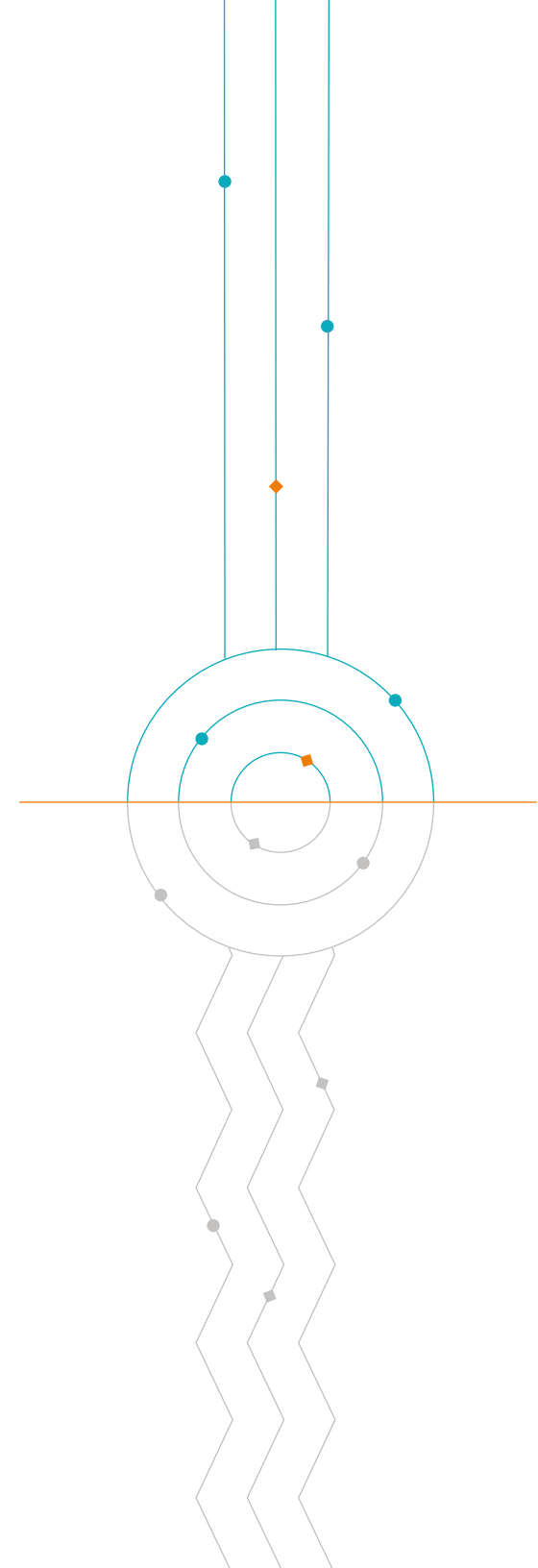
An integrated, modular, and scalable solution from a single vendor will streamline the process of configuring, deploying, and maintaining the data governance solution.

Adopting a “governance by design” approach will not only prevent your data lake from turning into a data swamp, it will enhance your ability to remain compliant with the emerging privacy policies and regulations around the world.

For example, while it’s appropriate for HR employees to have access to unmasked/clear-text data such as salary or date of birth, it would be unacceptable to give non-HR users the same type of access to run enrichment or self-service analytics.

You achieve the greatest value from your processes, workflows, and notifications when you can target the varying needs of different personas and enable more robust and efficient collaboration.

When every team—whether they are in business or IT—is working together, the community as a whole is stronger. Processes, workflows and policies all are agreed upon. The community has documented who is a stakeholder, who owns the data, they know who is involved in key areas, and it’s all agreed upon by the extended community. This streamlines the process of moving data governance away from control and toward collaboration.



Key 3. Adaptability

A key aspect of data governance deals with flexibility. How easy is it for a user to establish the types and extent of data governance?

Some types of data require a high degree of governance—for example, medical records protected by HIPAA, or data that is subject to privacy protections such as the General Data Protection Regulation (GDPR) or the California Consumer Privacy Act (CCPA). In other instances, the use of the data is more casual or is likely to be used in ways that need only to be directionally accurate.

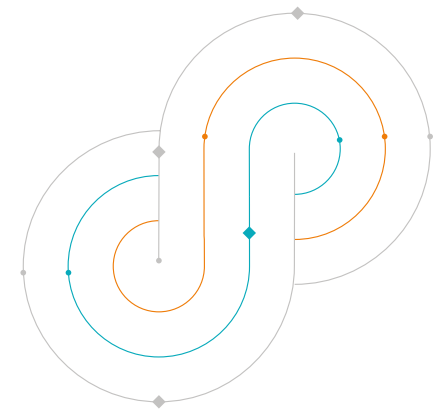
For data that is more sensitive or rigorously protected, it isn't enough to ensure that the data is correct; it is often just as crucial to define the rules regarding access, retention, and deletion. You may not need to meet the same sort of guidelines, or even meet the same levels of precision when you're working with other types of data. In other scenarios, users may have

specific preferences about the types of data they want to factor into their analytics, so the clarity and visibility around the source of the data might be paramount.

Financial data, for example, typically must be precise, come from known and trusted sources, and should be accessible only to people in specific roles in an organization. But other analytics only need data to be “good enough” to help inform a judgement call; knowing that the data comes from a trusted source may be more important than the specific origin of the data.

Consider a car manufacturer deciding on the color options for its new models. Based on what can be gleaned from various trusted sources, the car company may make a data-driven decision to emphasize the “blue” option in its upcoming models but might also choose to discount what the data has to say about the specific shade of blue that customers would prefer.

The right approach to data governance will both minimize risk and liability while still delivering tangible value to the business bottom line.



Key 4. Data You Can Use

For the data in an enterprise data lake to be useful, it must be trustworthy.

Since different end-users have different requirements for creating that foundation of trust, the data should meet all of these requirements:

- Comprehensive and accessible
- Indicate all sources
- Articulate the quality of the data, and communicate the degree of accuracy and completeness
- Clarify the types of use consents obtained
- Document who has used the data, and for what purpose

For optimal reliability, an enterprise data governance program should support data, regardless of location (on-premises, hybrid, and cloud sources). It should offer connectivity with all databases and apps, as well as data from Internet of Things (IoT)-enabled devices. Data should also be updated based on the needs of the end user: for example, data scientists often want to work with raw data, as opposed to business analysts who typically prefer to work with normalized data sets. The solution should also provide visibility into data lineage as well as impact analysis.

Robust connectivity capabilities, data cataloging, and flexible data governance options tailored to specific needs translate into reliable data that can be used for:

- Self-service analytics

- Monitoring and analyzing data streams from IoT devices (such as sensors)
- Understanding unexpected new data sources

Finally, access is key to usability: if the method and process required to access data is too daunting, its utility to end users will be effectively limited. With reliable data available on a unified platform, ramping up new users becomes a far more intuitive process, and yields much more value for the organization.



How does data lake governance help users in different parts of the enterprise leverage data to achieve business objectives, streamline processes, and enable digital transformation?



Worldpay: Supporting digital business transformation with data governance

After some false starts, Worldpay, a leading provider of credit card processing services, was still searching for a data governance solution that would help them manage data challenges and opportunities presented by its digital-driven business transformation and recent corporate acquisitions. They sought to transform their approach to become more context-driven, centralized, forward-looking, and focused on providing consistently high-quality data across the organization. The steps Worldpay took over the course of rolling out the new system—[shared in a recent webinar](#)—provide insights about what an enterprise needs to do to ensure both technical and organizational success.

worldpay

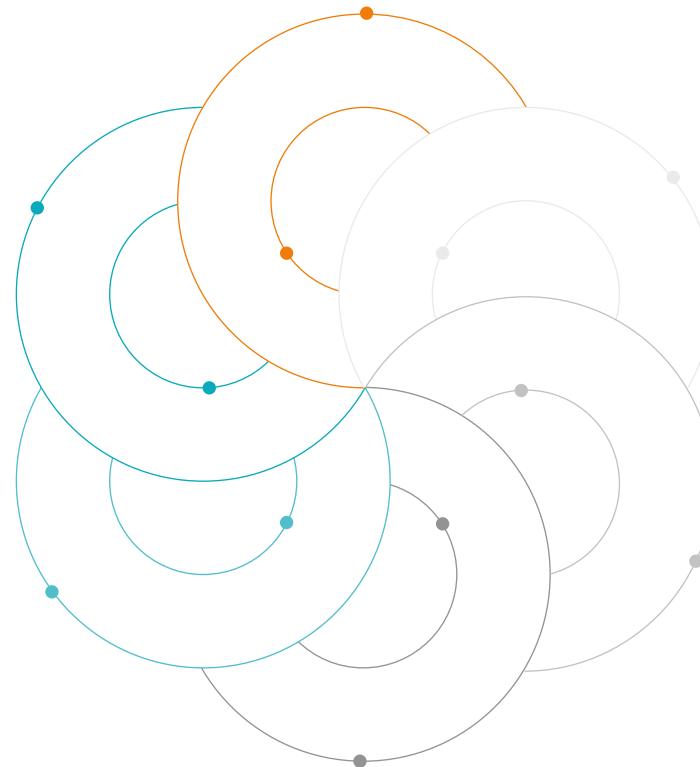


Key 5. AI-Driven Technology Can Streamline Manual Processes

In any enterprise, it's a given that data will continue to scale and evolve. To manage these vast volumes of rapidly changing data, AI-driven automation is a must.

Being able to take an adaptive approach to data governance will also serve the organization well when it comes to emerging technologies and future generations of data governance. For example, AI can employ certain algorithms and prompt data stewards to apply particular governance rules when analyzing different aspects of the data, which can greatly streamline current and future data lake management.

Even when we don't know everything about our businesses—or we don't know where all of our data lives—technology can help us find what we're looking for. For example, an AI-powered solution can help an organization deploy data governance in a modular fashion, with a modest pilot that can quickly and easily ramp up to accommodate robust functionality and build up to unlimited data volume as necessary.

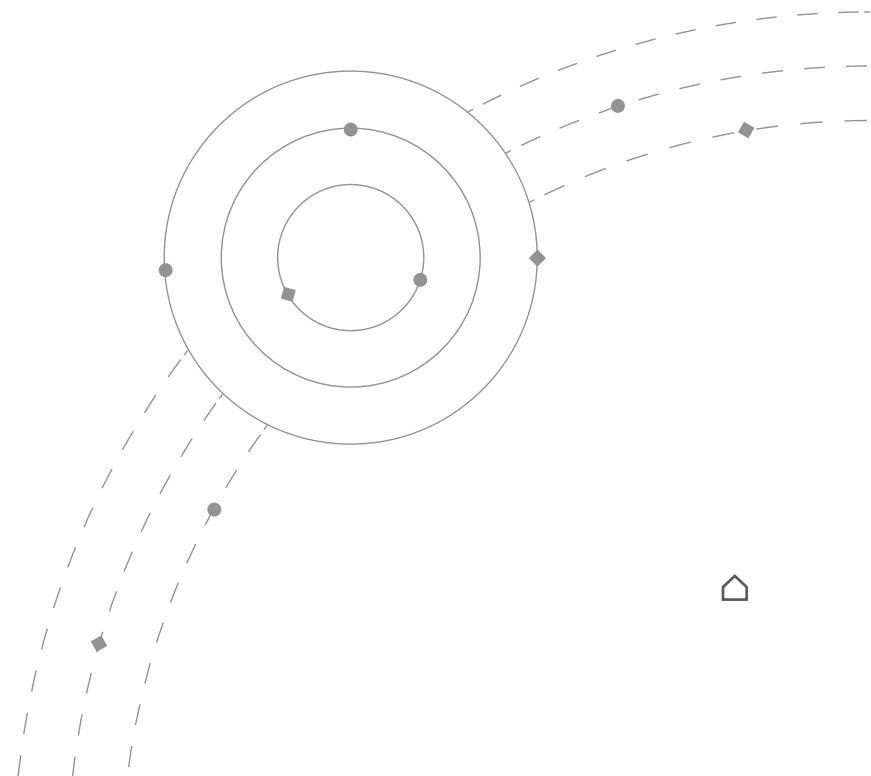


Conclusion

The right data governance program will ensure that your data lake meets the needs of every stakeholder in your organization, from end users to data scientists.

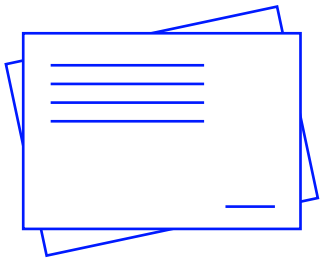
A great data governance program can boost efficiencies and speed up workflows because it enhances the trustworthiness and reliability of the data, and because it provides a shared understanding and a common vocabulary for all users.

When you have the best data governance program for your organization, you'll know that there aren't any inefficiencies, weaknesses, or hidden obstacles lurking beneath the surface of your data lake—and you'll improve your organization's ability to leverage data that drives business transformations that not only reduce risks and cost but create new opportunities.

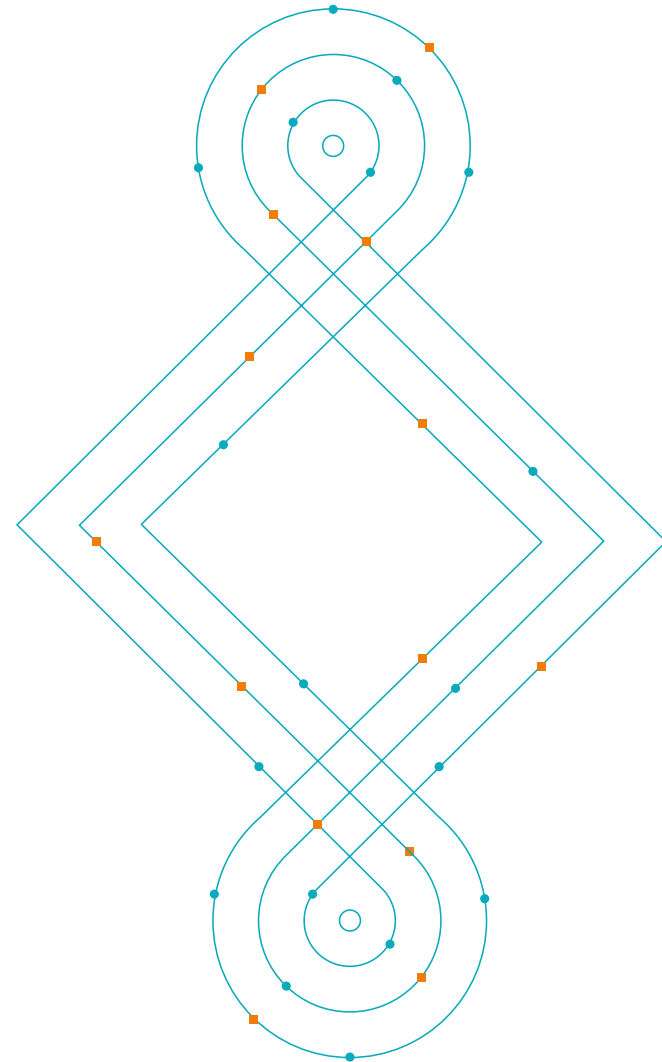


Further Reading

This [handy checklist](#) shows why big data needs a new kind of data governance and can help you get started on improving the quality of the data in your data lake.



[READ MORE](#)



About Informatica

Digital transformation changes expectations: better service, faster delivery, with less cost. Businesses must transform to stay relevant and data holds the answers.

As the world's leader in Enterprise Cloud Data Management, we're prepared to help you intelligently lead—in any sector, category or niche. Informatica provides you with the foresight to become more agile, realize new growth opportunities or create new inventions. With 100 percent focus on everything data, we offer the versatility needed to succeed.

We invite you to explore all that Informatica has to offer—and unleash the power of data to drive your next intelligent disruption.

Worldwide Headquarters

2100 Seaport Blvd, Redwood City, CA 94063, USA

Phone: 650.385.5000

Fax: 650.385.5500

Toll-free in the US: 1.800.653.3871

informatica.com

[linkedin.com/company/informatica](https://www.linkedin.com/company/informatica)

twitter.com/Informatica

[CONTACT US](#)