



Informatica®

White Paper

Lead Your Data Revolution: Unlocking Potential with AI, DataOps and MLOps

Where data
& AI come to **LIFE**™

Contents

The Rise of Data at Scale	3
The Challenges of Becoming a Data-Driven Enterprise	3
Operationalize Your Data: Start with Observability	4
DataOps: Bringing Structure to Your Data	6
The Role of MLOps	7
- Where DataOps and MLOps Come Together	8
How FinOps Ties in	9
- Understanding FinOps	9
- Pillars of FinOps	9
The Systems Thinking Approach to Data Management	10
- Adopting a unified data management platform	10
- Weaving in DataOps to all things data	10
- Overlaying ML to drive insights and business value from the data	11
At the Intersection of DataOps and FinOps	12
Unified Cloud Data Management	13

The Rise of Data at Scale

Data is everywhere¹ in the current digital landscape, and the advent of cloud computing has sparked, and accelerated, an unprecedented era of disruption. When effectively harnessed, this vast ocean of data has the power to redefine customer engagement, crafting unparalleled experiences that set new industry benchmarks. However, the true utility of data extends beyond customer interactions — it's a foundation in delivering value across the board, be it for customers, employees, partners and the greater ecosystem. Accurate interpretation and strategic use of data are pivotal. With that in mind, the pressing imperative for enterprises is to embark on a critical journey toward becoming 'fully aware' and wholly data-driven — a transformation that unlocks potent insights, bolsters decision-making and fuels innovation at scale.

The Challenges of Becoming a Data-Driven Enterprise

While the potential for data to revolutionize enterprise operations is vast, the journey to unlock its full benefits is fraught with challenges that many fusion data teams face. In fact, over 80% of Chief Information Officers (CIOs) have reported falling short in realizing the anticipated business advantages associated with cloud migration.² The hurdles are multifaceted and extend from the arduous task of breaking down data silos — isolated repositories of locked-away data — to grappling with the often-compromised quality of the data itself.

Equally daunting are the challenges related to the integration and utilization of diverse tools that are essential for modern analytics and AI use cases. Beyond mere adoption, there lies the deeper question of how these tools interact within the ecosystem to produce a coherent and efficient workflow. Furthermore, the diverse nature of data across the enterprise requires robust governance mechanisms to maintain integrity, comply with regulations, and ensure data is ethically used.

As organizations try to navigate this complexity, they encounter a pressing need for more nuanced strategies in data orchestration, literacy and culture change management, which are vital for cultivating a truly data-forward enterprise. Understanding data operations and adopting a fluid approach to incorporate the latest practices in DataOps, AI, and MLOps requires a synthesis of discipline (the science) and adaptive strategy (the art) — a combination that must be carefully unraveled and implemented in a methodical progression. Without this, enterprises may find themselves unable to harness the latent power of their data and thus miss out on a transformative opportunity for growth and innovation.

¹ <https://www.cmswire.com/digital-experience/6-critical-success-factors-for-the-data-strategist>

² <https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/cloud-migration-opportunity-business-value-grows-but-missteps-abound>

The challenges of becoming a data-driven enterprise



The quality of data

Siloed and disorganized

More than \$15 million in losses due to the quality of data



Fused data teams

Lacking collaboration and strategic direction

85% of projects fail due to the lack of collaboration



AI/ML tools

A lot still new and unexplored

Less than 10% of the AI/ML initiatives deliver



The spend on data and cloud

Not streamlined

70% of organizations overspending without results

Source: Informatica webcast, "The Power of DataOps: Bring Data Automation to Life."³

Operationalize Your Data: Start with Observability

To truly thrive in this modern economy, enterprises must irrevocably weave a data-driven methodology and mindset into their very fabric — making it an integral part of their organizational DNA. This seismic shift necessitates embedding data-centric thinking and practices into all facets of the business, spanning human resources, operational processes and technological frameworks. It demands a paradigmatic shift in approach, one that is holistic and systematic — a deliberate move away from intuition-based decisions to informed strategies, driven by robust data analysis. Organizations must learn to extract and maximize value from their data through a structured and disciplined practice that is consistently applied at every level, thus transforming data into a strategic asset that propels them ahead of the competition.

³ <https://gateway.on24.com/wcc/eh/3857765/lp/4147356/the-power-of-dataops-bring-automation-to-life?partnerref=Bambu&cext=infasocial-emp&blaid=4363928>

First, enterprises must take the concrete step of implementing the concept of data observability to actualize the value of their data assets. Data observability goes beyond traditional monitoring, allowing you to fully monitor and manage the quality, reliability, usage and performance of data across all pipelines. This thorough visibility breeds trust and ensures reliability by providing transparency into the real-time aspects of the data pipeline, allowing for real-time diagnostics and optimizations. You can think of it as assigning unique fingerprints to each data element and enriching them with descriptive metadata.

Each use case of metadata, as shown in Figure 1, serves a different purpose in unfolding the data operations, and observability is important to explore.

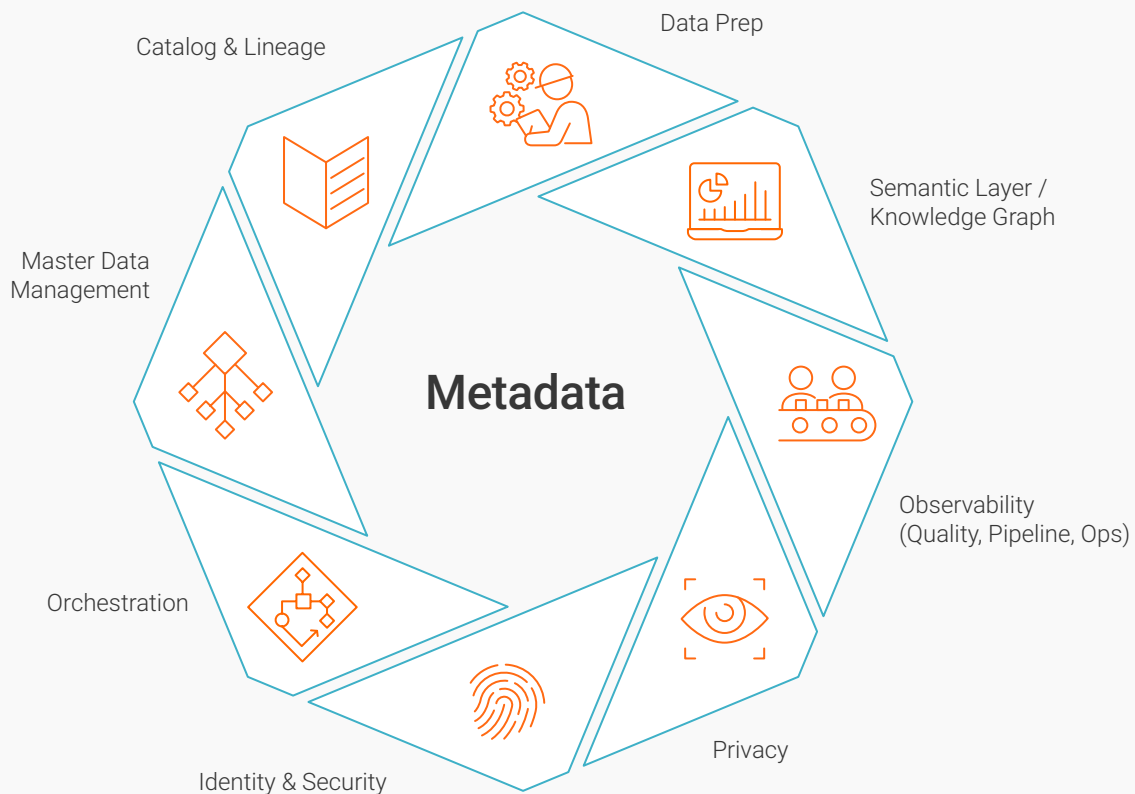


Figure 1. Data observability emerges as a new use case of metadata

DataOps: Bringing Structure to Your Data

DataOps is not a tool or a technology but an approach that stems from observability. It ensures that the quality of the data and the pipelines are intact to drive value and insights and make the right business decisions promptly. To achieve this efficiency, it ties the core pillars of lean, collaboration, and automation into the Ops approach. When AI is fused with DataOps, it augments the system with predictive intelligence, optimizing data processes, forecasting trends and flagging anomalies, enhancing data-driven insights and operational agility.

The big question, though, is how all this ties together to take enterprises from data chaos to nirvana. Through the application's lifecycle and in every phase, as illustrated in Figure 2, an Ops approach ties in to bring the much-needed framework.

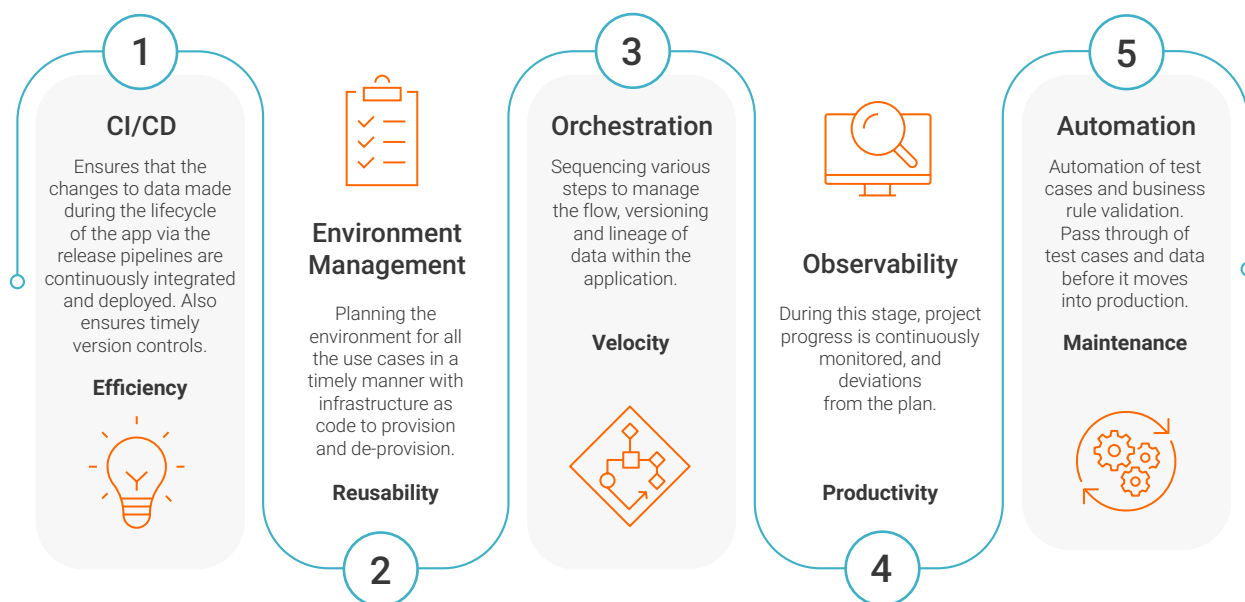


Figure 2. DataOps approach in an App's lifecycle.

The Role of MLOps

MLOps is a new method or approach of collaboration between data scientists and IT teams to automate and productize ML algorithms. MLOps enables the operationalization of the end-to-end pipeline that supports the continuous delivery and integration of ML models in a production environment.

Most organizations engaged in data science have a defined process for building, testing and training ML models. The challenge is what to do after the model is built. Integrating the model with existing systems and data flows, deploying it into a live environment where it can begin to provide value and providing continuous monitoring, are essential to ensure that it performs as expected over time.

This is where software development aligns with ML models, and a new approach or method to manage this fusion comes in (see Figure 3).

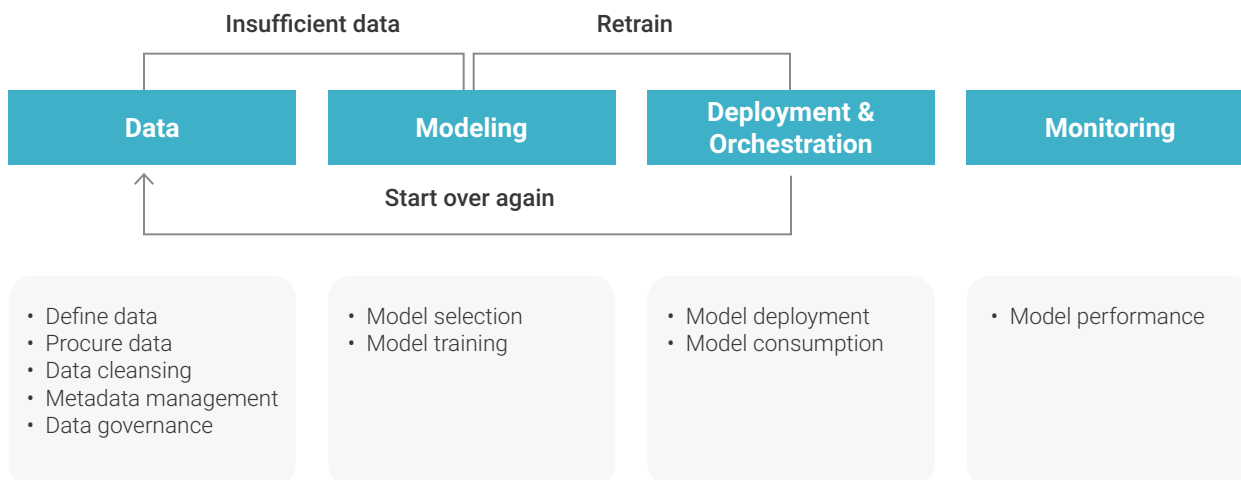


Figure 3. Steps to operationalize AI and ML models.

Where DataOps and MLOps Come Together

The primary area where MLOps and DataOps come together is during the data ingestion and integration phase (or pre-processing) step, as illustrated in Figure 4. This critical phase is where the machine learning algorithms are applied to extract value and insights from vast datasets. Data-centric AI approaches center on achieving better model accuracy by refining and pre-processing the datasets used for training, thereby enhancing the quality of the input data rather than extensively modifying the entire model. In contrast, DataOps establishes the overarching framework and practices that ensure the creation and maintenance of high-quality, reliable data – the foundational elements upon which ML models can leverage to derive meaningful outcomes. These two disciplines are mutually reinforcing: DataOps ensures a pipeline of clean, well-organized data, while MLOps focuses on the application of this data to develop, deploy and monitor machine learning models in production environments.

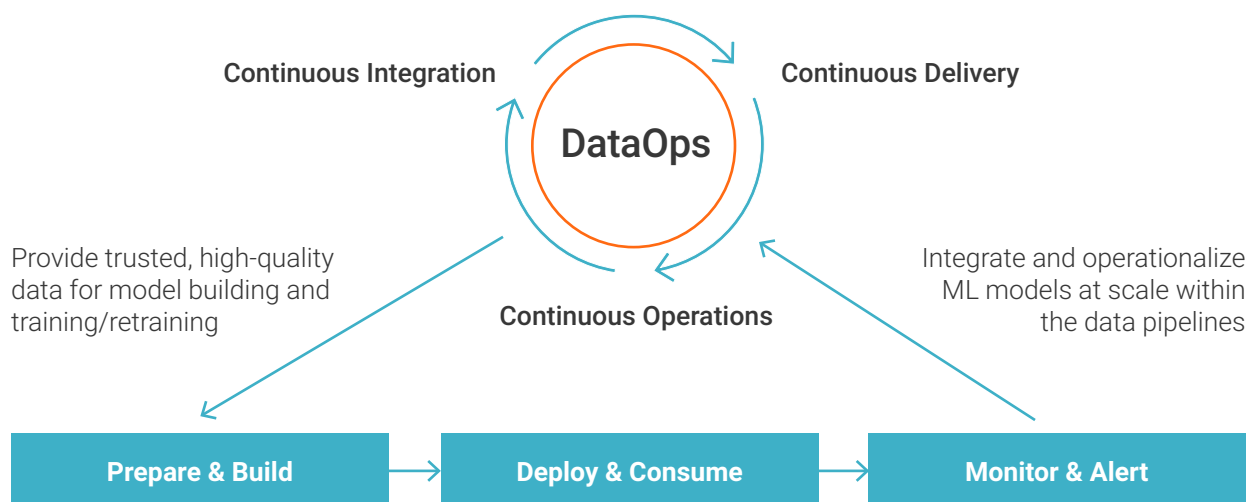


Figure 4. Depicting the seamless flow between DataOps and MLOps.

How FinOps Ties in

Understanding FinOps

While organizations have embarked on the cloud journey, fewer than one-third have fully realized the cloud's potential benefits. A mere 15% have succeeded in effectively mapping their cloud spending to gain business value. This is where the concept of FinOps comes to the fore.

FinOps, like DataOps, is an approach that helps to understand the financial implications at each stage of data management. It brings finance, business and operations together in one framework to drive financial accountability yet accelerate business value from every dollar spent. It's a shared responsibility for the organization's cloud computing infrastructure costs.

Pillars of FinOps

- **Inform:** This is a visibility phase on the workloads and analytics running in the cloud to assess their cost and learn about the kinds of teams running them. Transparency into how these resources is consumed is vital to fiscal intelligence.
- **Optimize:** This mainly forecasts and plans cloud use to maximize business value. It includes reducing resource over- and underutilization and modeling the need for new resources based on business objectives.
- **Operate:** This involves focusing on the day-to-day activities of managing and governing cloud infrastructures. Various levers help with this.

To better manage cloud consumption for business value, organizations commonly monitor and measure the metrics in the three FinOps phases, as outlined in Figure 5.

Phase	Metric	Benefit
Inform	The percentage of cloud use and spend allocated to a project, business unit or department.	Driving insights on business value, like customer, product and service profitability, is the foundation of cloud economics.
Optimize	The variance percentage between plan versus actual for cloud resource consumption.	Forecasting accuracy is critical for optimizing cloud value since it impacts costs as well as service levels.
Operate	The percentage of cloud use and spend covered by reservations and committed use discounts.	Reducing consumption outside of discount plans significantly increases return on investment.

Figure 5. Three phases of FinOps and their metrics and benefits.

⁴ https://www.informatica.com/in/lp/how-to-optimize-cloud-consumption-and-costs-with-finops_4544.html

The Systems Thinking Approach to Data Management

Data management is a multi-step process. The full data management lifecycle involves continuous delivery, integration and an operations loop. DataOps, then, is an approach that provides a way to operationalize the data platform by extending the concepts of DevOps to the world of data. It is built on three key constructs: integration, deployment and operations.

Before introducing a cloud-based data management strategy, it is imperative for data leaders to first embrace a systems-thinking perspective for the holistic management of data.

System thinking involves:

1. Adopting a unified data platform
2. Weaving in the DataOps approach to all-things-enterprise-data
3. Overlaying ML to drive insights and more significant business value
4. Bringing governance and security to ensure the proper use and protection of data

Adopting a unified data management platform

A unified cloud data management solution is designed to bring all dispersed data under one roof to drive better business outcomes. Unlike point solutions, this unified approach enables you to share, deliver and democratize data across your business lines and other enterprises with a foundation in governance and privacy. Businesses are now using the power of cloud-native, AI-powered and robust cloud data management to reinvent and bring value via this data.

Weaving in DataOps to all things data

DataOps is the fuel that helps to drive value from all the data. As enterprises develop and deploy data analytics with DataOps, they can attain the same level of productivity that we see in leading software companies. Analytics will be created and deployed rapidly, and statistical process controls will ensure high data quality.

DataOps offers a new approach to creating and operationalizing analytics that can minimize unplanned work, reduce cycle time, and improve the code and data quality. It's a methodology that enables data teams to thrive despite the increasing complexity required to deploy and maintain analytics in the field.

Orchestration	Observability	Environment Management	Deployment Automation	Test Automation
• Orchestrating data pipelines to support any integration patterns at scale • Enabling an open, extensible and embeddable environment with third-party APIs • Handling complex business processes and automating them with built-in intelligence capabilities • Improving process efficiency and productivity of data engineers	• Proactively detecting data quality and anomalies in data pipeline • Automating data delivery, fulfillment and observability with governed workflows • AI-powered capabilities such as self-heal, auto-tune, auto-scale and smart shutdown to process data engineering workloads	• Leveraging fully automated infrastructure management to support all deployment options such as Kubernetes, serverless, etc. • Providing easy-to-manage, wizard-driven compute provisioning in any cloud	• Efficiently comparing, merging and deploying data pipelines • Seamlessly integrating with any cloud ecosystems • Jumpstarting data engineering projects with pre-built templates, demo assets, documentation and hands-on workshops	• Enabling simple and easy to execute cloud-native test validation • Automating testing and validation in a test environment before deploying changes in the production environment

Figure 6. The components of DataOps

Overlaying ML to drive insights and business value from the data

Most organizations engaged in data science have defined a process for building, training and testing ML models. The challenge has been what to do once the model is built. Integration, deployment, and monitoring are essential for providing continuous feedback once the models are in production. This is where the entire process of building ML models aligns more closely with the software development lifecycle than with an analytics project.

Many organizations think data science and machine learning projects are limited to creating models. But once a model is developed and deployed, many other aspects are needed to operationalize it, including:

- Management and monitoring to ensure the model perform optimally within the thresholds defined by the business
- A feedback loop to monitor model drift or degradation
- Tweaking of the model and periodic retraining

Hence, it would help if you had MLOps to establish a continuous delivery cycle of models that form the basis for AI-based systems. MLOps is critical for delivering business value for data science projects.

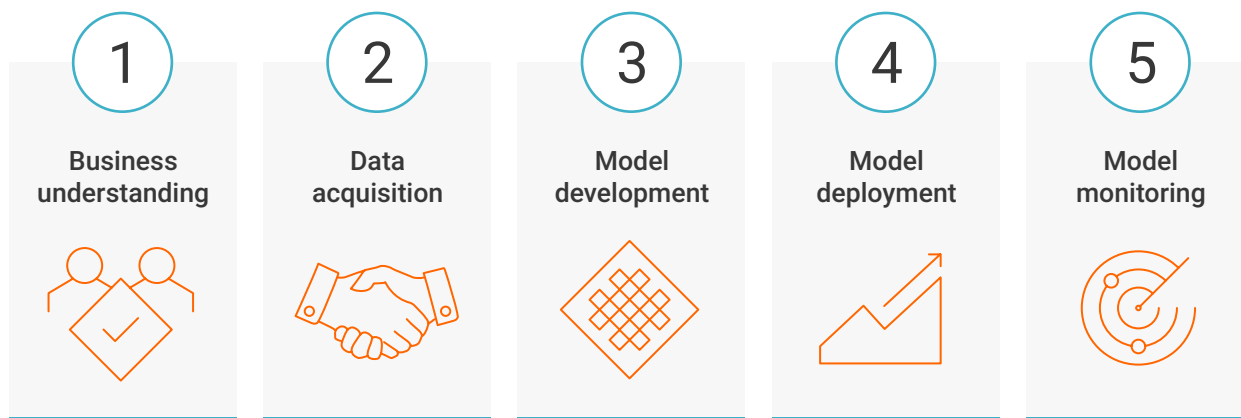


Figure 7. The steps to MLOps

At the Intersection of DataOps and FinOps

Data management capabilities and governance initiatives are key to supporting your FinOps activities and defining the complete unification of the FinOps and DevOps models, as detailed in the following three-step journey:

- 1. Assess your current data management capabilities:** You need a broad array of data management capabilities for FinOps. Assess your current solutions and evaluate the depth of functionality, breadth of capabilities, extent of integration and ability to deploy natively across cloud platforms. This will help you better understand your current circumstances and identify gaps.
- 2. Identify opportunities for architecture optimization:** Your data management architecture will need to support requirements within the same cloud, across different clouds and with on-premises sources. Look for ways to optimize design, like utilizing serverless capabilities for scalability and pushdown optimization for performance. This will help you design cloud infrastructures that ensure effective and efficient operations.
- 3. Establish governance and accountability:** Roles, policies and processes need to be aligned with FinOps objectives. Implement shared metrics and measurement practices to provide consumption and cost visibility to all stakeholders and enable unit economics analysis that guides business value optimization. This will help you drive accountability and collaboration across teams.

Unified Cloud Data Management

Informatica's point of view on DataOps is that managing an end-to-end data pipeline is a multi-step process. Most of these processes are managed manually by more people or through custom-built tools. Vendors could provide point solutions for one-off capabilities, but the point solutions create more complexity and bring more challenges and added cost.

The Informatica Intelligent Data Management Cloud™, the industry's only comprehensive cloud data management solution, provides customers with a proven, AI-powered, scalable cloud data architecture. This enables them to turn data into actionable insights very quickly. It's unified, agile, fully automated and collaborative to address all data challenges to help organizations with their analytics and AI.

To learn more about how Informatica brings your enterprise data to life, visit www.informatica.com.

About Us

Informatica (NYSE: INFA) brings data and AI to life by empowering businesses to realize the transformative power of their most critical assets. When properly unlocked, data becomes a living and trusted resource that is democratized across your organization, turning chaos into clarity. Through the Informatica Intelligent Data Management Cloud™, companies are breathing life into their data to drive bigger ideas, create improved processes, and reduce costs. Powered by CLAIRE®, our AI engine, it's the only cloud dedicated to managing data of any type, pattern, complexity, or workload across any location — all on a single platform.

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
twitter.com/Informatica

Where data & AI come to



[CONTACT US](#)