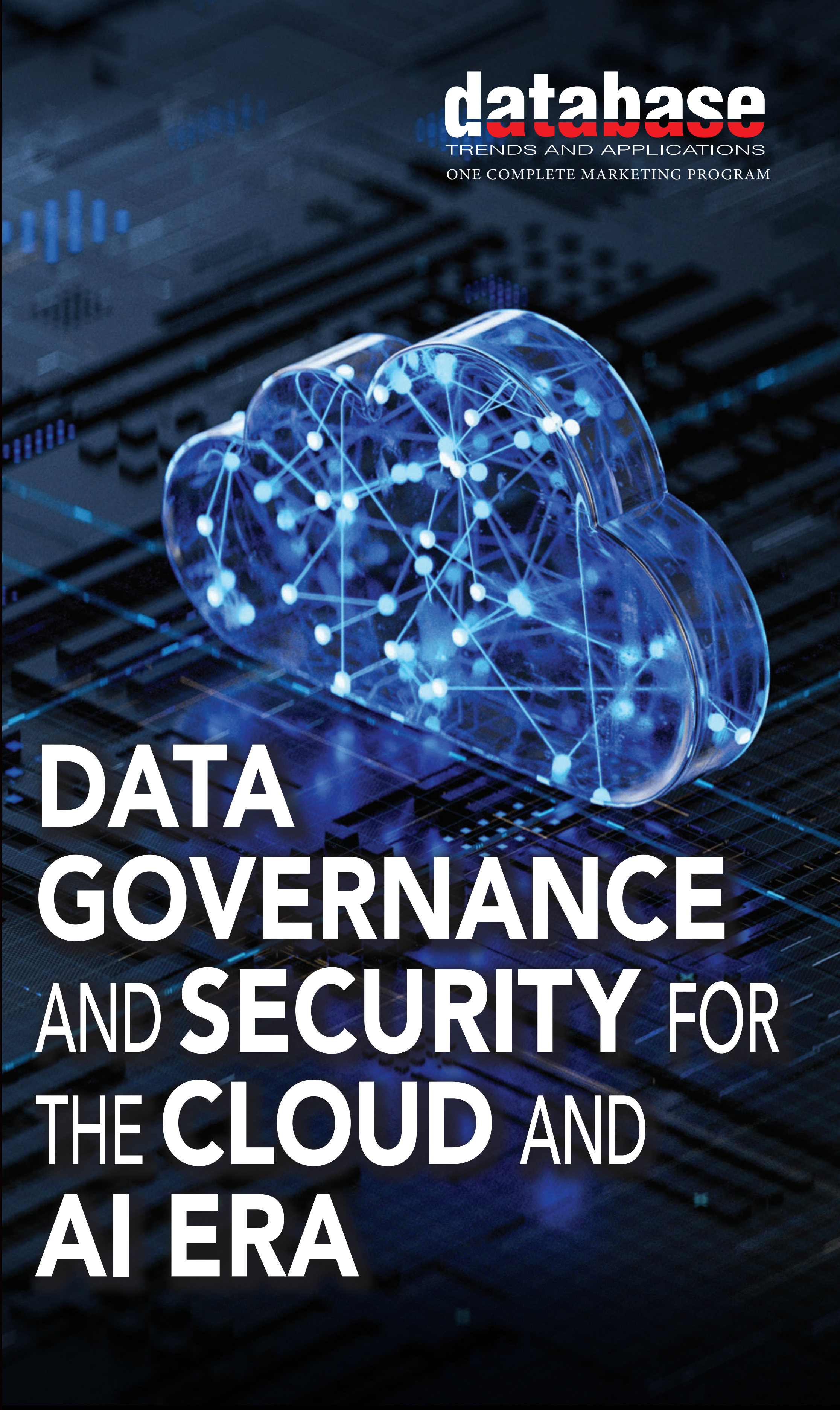
IRI

GOVERNING DATA FOR
CLOUD & AI ENVIRONMENTS

database
TRENDS AND APPLICATIONS
ONE COMPLETE MARKETING PROGRAM



DATA GOVERNANCE AND SECURITY FOR THE CLOUD AND AI ERA

database
TRENDS AND APPLICATIONS

Best Practices Series

DATA GOVERNANCE AND SECURITY FOR THE CLOUD AND AI ERA

Best Practices Series

As AI and machine learning (ML) take hold across enterprises, there's a push to ensure that models and algorithms associated with these emerging technologies are viable for the business and that the data employed is secure. This calls for new standards for both secure and responsible AI-based data. The data needs to be protected against cyberthreats and deliver fair and unbiased results to decision makers and business leaders.

As organizations develop new data analytics and AI projects to compete in today's digital economy, the protection and management of data flowing into AI systems have become challenging aspects for IT and data leaders. This is key, as many companies are betting their businesses on AI-powered applications such as predictive analytics, customer relations, and product development.

That's why the data flowing into AI systems needs to be of the highest possible integrity and quality. As businesses aggressively move into both generative and operational AI, they require massive amounts of accurate and timely data. However, only 23% of data managers surveyed by Unisphere Research, the research arm of *Database Trends and Applications*, expressed full confidence in their data. Close to one-third said that data quality is a constant, ongoing issue. What's more, new data analytics and AI projects are now surfacing many of these issues. The survey, part of a series launched in 2021, reports no letup in the growing confidence gap in the data needed to support next-generation initiatives ("Unfinished Business: Taking on the Data Quality Challenge in the Age of AI," Unisphere Research, a division of Information Today, Inc., November 2023).

The stakes are much higher now. The rise of large language models (LLMs)—both publicly available as well as contained within enterprises—to support business decision making and customer communications means data is being pressed into service in new and highly demanding ways as training data and real-time streaming feeds. The models need to support growing demand for intelligent, customer-facing applications such as chatbots and conversational interfaces, as well as intelligent assistants for internal enterprise operations.

AI data governance, of course, isn't limited to protecting a company's assets and brand—it also helps protect society at large, especially in terms of compliance with laws and regulations issued by governments.

Add to this the ongoing incidents of cyberattacks, hacks, and security breaches that have become almost daily events. If ever there was a solid business case for AI data governance, this is it.

AI data governance, of course, isn't limited to protecting a company's assets and brand—it also helps protect society at large, especially in terms of compliance with laws and regulations issued by governments. For much of the world beyond enterprise walls, AI has become a hot-button issue.

The following are key considerations for building an AI data governance framework:

- **Begin the AI data governance process as a business initiative.** AI data governance is an initiative that needs to extend well beyond the IT department—ownership and responsibility for AI should reside with anyone who works with technology. Technical teams can implement solutions and ensure that systems, models, and data are delivered. However, it's up to the business to establish the mission and purpose of AI initiatives along with the guardrails that ensure data isn't compromised. For many organizations, this may mean a cultural change, including a push to incorporate responsible AI into day-to-day operations, as well as longer-term strategies. Education and training are essential to developing an AI-ready culture.
- **Assess the current AI footprint.** Inventory your current AI systems and assets. This may require working across enterprise lines to better understand what AI tools, platforms, or applications are being used or considered—including the more informal use of outside AI resources, such as ChatGPT.
- **Understand that trust is at the core of governance.** While enterprises and their employees are enthusiastically embracing AI—particularly generative AI (GenAI)—to assist in their work, there is still a lingering mistrust of AI-generated decisions, especially among executives and managers, who continue to rely on gut-level decisions. Effective governance will help assure that the insights produced by AI systems will be based on well-vetted and relevant data. AI data governance should be geared to instill trust in the data that supports AI insights and recommendations.
- **Integrate data governance with AI governance.** These two areas are closely connected. For instance, governing data applies to AI, which is now part of the realm. Data governance, which seeks to oversee the security, quality, and business value of data, lays the groundwork for AI implementations. There are already standards and procedures, developed across several decades, governing data use. The challenge now is to expand upon these governance tenets to extend to AI data. Efforts to bring these two realms together, however, can prove to be challenging, as AI and data teams tend to work separately and need to be brought together.
- **Employ AI to also power governance solutions.** Ensuring the viability of data going into AI initiatives is one side of the coin. The other is the emerging role of AI in supporting governance, through automatic monitoring and providing course corrections for governance and security, based on continuous learning cycles. Solutions are emerging that not only employ GenAI to build SQL queries but also ensure security and compliance.
- **Understand that AI data governance is about security—and much more.** Successful AI data governance addresses potential AI risk factors, as well as assessing business value, tracking data use, providing transparency, assuring privacy, flagging ethical risks, and assuring compliance. In addition, AI presents potential risks with violating intellectual property—even if it is restricted to internal use.
- **Remember that AI data governance can't be "outsourced" to cloud or third-party providers.** The onus is on enterprises to put processes in place to ensure their data is secure and relevant. Cloud providers can deliver broad security from the outside, but it's up to organizations to oversee internal access and adherence to policies and procedures.
- **Keep humans in the loop.** No systems, no matter how current the algorithms and data feeds, are perfect. It's important that the people charged with deploying and using AI are trained in the ethics and risks of these systems. Importantly, people need to be encouraged and ready to hit the "stop" switch and challenge the results delivered by AI systems.

Comprehensive AI data governance will help deliver the new capabilities AI is poised to offer to organizations. Ensuring AI data is both secure and responsible stems from regular and sustained engagement by decision makers and technology teams. This will build trust in AI and the results it delivers, which promise to accelerate business growth in today's economy. ■

—Joe McKendrick

Governing Data for Cloud & AI Environments

IRI Voracity
An Insatiable Appetite for Data

As stewards for the newer sources of data in the cloud and feeding AI models, today's GRC teams and developers must grapple with increasingly complex requirements for data security, compliance, and quality.

THE NEWER LANDSCAPE

Recent studies by Database Trends and Applications (DBTA) confirm that IT departments are dedicating more time and effort to data governance and security than ever before. Let's delve into the factors driving this trend:

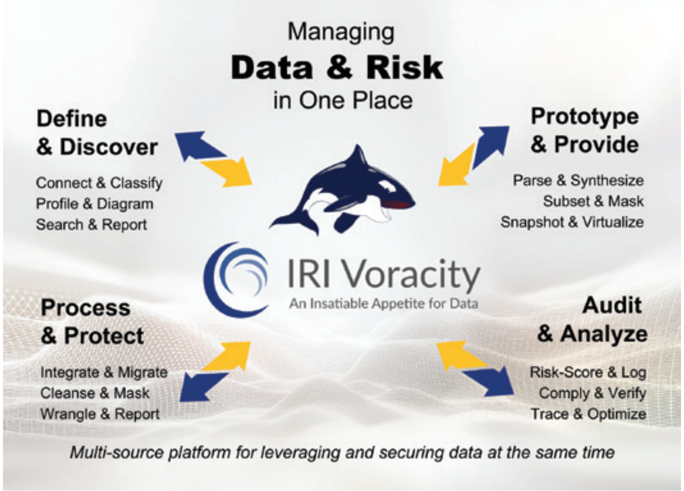
1. **Data Proliferation:** Organizations are facing an explosion of data sources and formats. From structured databases to unstructured logs, the sheer volume of information demands effective governance.
2. **Distribution Challenges:** Data transcends traditional boundaries. It flows rapidly between cloud, on-premise, edge and remote platforms. Ensuring consistent governance across this diverse landscape can be hard.
3. **Insatiable Demand:** Business users crave fast, easy access to data insights. Whether for reporting, analytics, or AI model training, the pressure to deliver timely information is relentless.
4. **Security Imperatives:** Evolving cyber threats pose risks to data integrity and confidentiality. Protecting sensitive information while enabling data access requires a delicate balance, in both production and test environments.
5. **Compliance and Quality:** Stricter regulations demand adherence to data privacy rules. Simultaneously, maintaining data quality standards is essential for accurate decision-making.

MODERNIZING AND IMPROVING DATA GOVERNANCE

IRI Voracity is a comprehensive data management platform designed to address these challenges head-on.

Voracity supports a range of data governance functions—especially for cleaning bad data and masking sensitive data—for a myriad of data sources on-premise and in the cloud. It uniquely offers:

1. **Unified Capabilities:** Voracity integrates data discovery, integration, migration, governance, and analytics within a single, intuitive interface. Built on the Eclipse™ framework, it streamlines the entire data lifecycle.
2. **Integrated Data Quality:** Standalone or during ETL, masking, wrangling, reporting, or migration jobs, Voracity users can validate, filter, scrub, de-duplicate, enrich, and standardize data in structured tables, files, and streams.
3. **Robust Data Masking:** Voracity excels in data masking and test data synthesis—critical data security activities. By classifying, discovering, and consistently obfuscating or synthesizing data, you can comply with data privacy regulations while enabling data utilization and preserving referential integrity.



Managing
Data & Risk
in One Place

Define
& Discover
Connect & Classify
Profile & Diagram
Search & Report

Prototype
& Provide
Parse & Synthesize
Subset & Mask
Snapshot & Virtualize

IRI Voracity
An Insatiable Appetite for Data

Process
& Protect
Integrate & Migrate
Cleanse & Mask
Wrangle & Report

Audit
& Analyze
Risk-Score & Log
Comply & Verify
Trace & Optimize

Multi-source platform for leveraging and securing data at the same time

REDEFINING TRADITION

Modern data governance and security practices must be scalable, agile, and adaptable to accommodate cloud and AI use cases. Voracity embodies this shift from traditional approaches, enabling organizations to harness the full value of their data while simultaneously governing it.

More specifically, Voracity and its included "IRI Data Protector" suite component products like DarkShield in particular, can search, report on, and mask a broad range of data classes pertaining to multiple data privacy laws across an even broader range of structured, semi-structured, or unstructured data sources.

Voracity also supports the ongoing cloud and AI movements through:

1. **Design Efficiency:** Voracity streamlines data processes, optimizing workflows from discovery to analytics. Its unified data management capabilities and consolidated data manipulation reduce project and design complexity, enabling teams to work faster and smarter.
2. **Runtime Speed:** Powered by CoSort and its built-in resource management features, Voracity saves time through smarter algorithms and job configurations.
3. **Team Collaboration:** A familiar GUI and common metadata foster a shared job design environment, bridging typical gaps between IT and business intelligence teams. As collaboration becomes seamless, governance and insight improve. ■

LEARN MORE ABOUT Voracity and the data-centric governance and security capabilities at https://www.iri.com/products/voracity/technical-details#capabilities.