DATA INTEGRATION AND GOVERNANCE STRATEGIES FOR THE MODERN ENTERPRISE

Best Practices Series

While the idea of "every business is a software business" has been the catchphrase for organizations for some time now, it's more accurate to say that "every business is a data business." Organizations are being built on data and valued for their data. However, transforming into a data-driven business is hampered by data silos, complexity, dysfunctional corporate cultures, and a lack of vision.

There's much more at stake with well-planned and well-executed data integration and governance than simply keeping all your data in line. It can make or break a business. "Data governance is not just a check box exercise, but a strategic investment that can propel you ahead of the competition," said Jayesh Chaurasia, an analyst with Forrester.

This message resonates in corporate C-suites, a survey of 600 chief data officers (CDOs), released early this year by Informatica, confirms. Data management investments are showing no signs of slowing down this year—despite any economic headwinds. A majority, 68%, predict increased data management investment in 2023. In addition, the survey finds data governance is CDOs' number-one priority.

Data fragmentation and lack of cooperation across businesses are the top issues these data executives are grappling with. A majority, 55%, report having more than 1,000 sources of data in their organization. In addition, 52% cite improved governance over their data and processes as a top data strategy priority for this year. Another 42% plan to invest in data quality and master data management capabilities.

"As data strategies evolve rapidly, data leaders are looking to change and adapt quickly, including exploring new trends such as data mesh, data fabric, data observability, sustainability, and data marketplaces—all of which require data visibility and governance," the report's authors state. "As data leaders bring strategies and tools together to achieve their business priorities, developing metrics and measurements for business outcomes is critical."

Data security needs to remain front and center of all initiatives, preferably built into solutions and processes as they are designed.

Data integration and governance requires engagement from across the enterprise, involving database administrators, data analysts, executives, and line employees.

There is a lot of work that needs to be done, however—especially on the organizational side. In its analysis, Forrester finds that fewer than 10% of enterprises can be considered "advanced" in their insights-driven capabilities. "The key differentiator between beginner, intermediate, and advanced capabilities is not primarily about technology or data; it's about the organizational habits and practices that enable all stakeholders to securely access timely, relevant, clean, complete data for developing insights and applying it to their decisions," said Chaurasia. "This is the realm of data governance."

The following are best practices in developing a forward-looking data integration and governance strategy:

- **Put the business in charge.** Data integration and governance is an extensive effort that touches every part of the enterprise. That means business leaders need to prioritize where and when investments occur in these efforts, as both human and financial resources are involved.
- **Assist the business in planning its data integration and governance strategy.** Data managers and business managers need to work collaboratively in establishing the priorities and requirements for developing a data integration road map and data governance framework. Importantly, this needs to be a continuous process.
- **Promote data ownership, sharing, and collaboration to break down data silos.** Data silos—long the bane of enterprises—continuously interrupt, inhibit, or even prevent the flow of information across business units needed for analytics. "Collaboration, a centralized repository of data knowledge, and the establishment of data-sharing channels are vital to overcome this challenge," Chaurasia urged.
- **Maintain emphasis on data compliance and security.** Data security needs to remain front and center of all initiatives, preferably built into solutions and processes as they are designed. Compliance is a key part of this equation, as data management practices need to comply with local, national, and extra-national laws and regulations. In today's global economy, for example, all companies do business with countries in the European Union, which has stringent privacy and data residency requirements. "Data governance frameworks, policies, and procedures establish data protection and access control mechanisms to ensure that data is only used by authorized people for appropriate purposes," said Chaurasia.
- **Establish a working data governance framework.** Such a framework "defines roles and responsibilities to ensure data ownership that supports data definitions, quality standards, sharing parameters, and access control, enabling users to easily discover, understand, and utilize available data assets," according to Chaurasia.
- **Make data discoverable through a master data management (MDM) plan and data catalog.** With MDM, administrators and users have an awareness of available data assets through a single enterprise repository. Likewise, a data catalog presents available data sources at the click of a button for users.
- **Monitor, measure, and maintain visibility.** Data managers need observability across all data management movement through their enterprises. Data is continuously moving; new sources are constantly coming online; new users are accessing applications and data; and fresh data is required for updating the training of AI models.
- **Automate as much as possible.** Maintaining data operations may be too overwhelming for data staffs with high volumes of data of all types flowing through organizations. With business demands for initiatives around real-time applications and AI on the rise, look to automate as many data integration tasks as possible.
- **Put data accessibility and self-service in place.** Let's face it, data managers simply don't have the time or resources to attend to every user request for data or analytical reporting. Plus, with businesses moving so fast, users often need to assemble analytics reports or apps at a moment's notice. "Well-defined data governance policies provide guidelines for data access, usage, and interpretation, supporting business users with the ability to explore and analyze data on their own," Chaurasia said.
- **Promote data ownership and accountability.** This is an important link in all data integration and governance initiatives. Everyone has a role to play. Data owners are responsible for the quality and timeliness of the data; data engineers and administrators assure the quality and security of data; data stewards help enforce governance policies; data analysts and scientists transform raw data into business value.

Data integration and governance requires engagement from across the enterprise, involving database administrators, data analysts, executives, and line employees. All participate in a continuous process of discovering, identifying, ingesting, and managing data. With demands for advanced initiatives such as AI and machine learning on the rise, there's never been a more important time for well-designed data integration and governance. ■

—Joe McKendrick

IRI Voracity: Combining Data Integration. Governance and More

IRI Voracity
An Insatiable Appetite for Data

Data integration and data governance work have traditionally been handled as separate disciplines performed in separate silos and software products, managed by people with different agendas and skill sets. Think about how many standalone data profiling, ETL, data cleansing, data masking, master data management, and IAM/auditing technologies exist, and the ones your company has deployed.

Now think about an alternative where you can use simple wizards, modern diagrams, a human-readable 4GL, and a familiar IDE to speed, consolidate, and team-manage these tasks. IRI and Gartner collaborated on that vision in 2015 which culminated in the 2017 release of IRI Voracity.

Voracity is the comprehensive data management platform that incorporates more than 45 years of IRI data manipulation and movement history which began with CoSort. The platform and its component products evolved from mainframe sort and data migration into data warehouse ETL acceleration and reporting, change data capture and slowly changing dimensions, and more recently into data validation and quality, data masking, subsetting, and test data synthesis. All of those jobs use the same back end CoSort SortCL 4GL data definition and manipulation program.

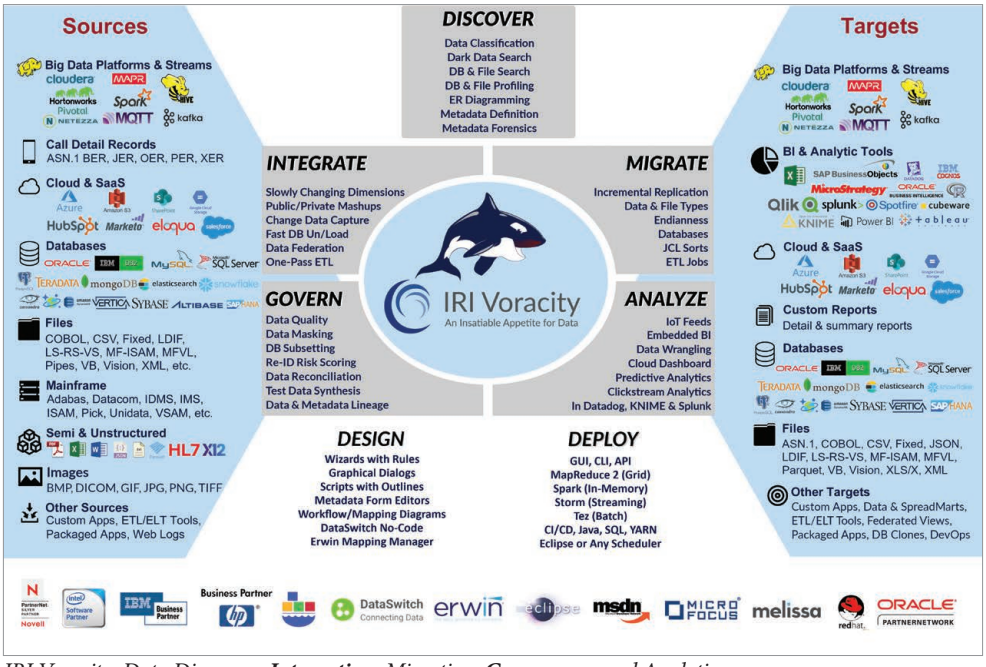
Voracity jobs are built and managed in IRI Workbench, a rich and intuitive IDE built on Eclipse that supports job design and workflow management in multiple graphical and text-based modes, and leverages centralization through EGit. Voracity jobs powered by this long-proven engine can both optimize and combine data transformation, cleansing, masking, and remapping/reporting tasks in one job and I/O pass.

More specifically, you can buy fit-for-purpose Voracity components, or license the whole platform, for:

- Data Discovery—classifying, diagramming, profiling, and searching of structured, semi-structured, and unstructured data sources, on premise or in the cloud
- Data Integration—individually optimized but consolidated same-pass E, T, and L operations, plus CDC, slowly changing dimensions, and ways to speed or leave any legacy ETL platform
- Data Migration—and conversion of data types, file formats, and database platforms, plus incremental or bulk data replication and federation
- Data Governance—PII data masking and re-ID risk scoring, DB subsetting, synthetic test data generation, data validation, cleansing, and enrichment, master and metadata management, etc.
- Analytics—embedded reporting, integrations with DataDog, KNIME, and Splunk, and fast data wrangling for BOBJ, Cognos, Cubeware, Microstrategy, Power BI, Qlik, R, Spotfire, Tableau, etc.

For example, you can combine tasks involving data acquisition, transformation, filtering, cleansing, masking, and reporting. You can also marry those jobs to external tools for IAM, monitoring and workflow, CI/CD and analytics, MDM, data lineage, and so on—and manage everything from the same pane of glass—IRI Workbench—whether you want to run on existing hardware on-premise, or in a single, multi-, or hybrid cloud environment.

Finally, from an affordability standpoint, it's good to know that Voracity subscription or perpetual license models count only engine nodes and not sources, users, or cores. It's a bundle of technical capability designed to deliver a bundle of efficiency, simplicity, and affordability. ■



*IRI Voracity: Data Discovery, **Integration**, Migration, **Governance** and Analytics*

IRI
www.iri.com