



An Enterprise Guide to Modern Analytics

Essential Considerations for Managing Data
and Staying Ahead of the Competition

Custom content for Matillion by studioID





The nature of data analytics is changing so fast that many businesses struggle to understand what tools and approaches they should consider.

Companies that fall behind are either failing to act or focusing on the wrong issues. As a result, they are getting outdone by competitors, often digital natives, who understand how to make the most of modern methods. Companies that succeed are doing so based on insights, agility, and the efficiency and the decision-making power that come from working with data they know they can trust. Meanwhile, enterprises that struggle to make data useful and trustworthy — often large, more traditional businesses — are dropping further behind.

Businesses that still rely on and try to manage the three Vs of big data — variety, velocity, and volume — operate from an outdated rubric. What matters most now are the three Ds of data in the cloud: diversity, distribution, and dynamics.

The three Ds concern the different challenges of data, where that data comes from, and how to manage information that is changing moment to moment. Other critical factors for achieving a modern data environment include treating artificial intelligence (AI) and machine learning (ML) as real concepts, not just theoretical ideas; and enabling data to go from anywhere to anywhere — for example, from ML models or cloud data platforms back into operational systems in a more enriched state — rather than following a linear path from source to analysis to dashboard.

In this playbook, we provide an in-depth look at these critical elements and more to help enterprises modernize their analytics strategies.

The Three Ds

The latest data analytics methods reflect a greater degree of data diversity, distribution, and dynamics than previous models. To make the best use of today's predictive tools, companies must work in ways that embrace this fundamental shift.

Diverse

Organizations must grapple with the fact that data now comes in a wider variety of forms, including transactional records, research and development results, inventory reports, financial statements, customer care interactions, marketing efforts, social media, machine logs, and much more. According to an article from the **MIT Sloan School of Management**, external data is particularly important as a business driver. The article states, "Most analytically mature organizations use more data sources, including data from customers, vendors, regulators, and competitors. ... Companies use this external data to augment their decision making, better meet customer needs, predict supply and demand, and more."

There are two keys to managing the diversity of data, according to Mark Balkenende, Senior Director, Product Marketing at Matillion. "Companies need a platform that can ingest all data types, make sense of the data, and use it," he says. "And they must have the right people to run it." But the right data management platform, he adds, "can be managed by a generalist in data analysis, rather than someone who is highly specialized, putting data integration skills in the hands of many rather than a select few."

Best practices for using such platforms include knowing the right level of governance and control for various assets and systems (some need more than others) and giving teams the right tools and platforms to gain access to and work with the data they need. This approach includes converting data into the right formats for reporting, machine learning, or other uses. "A solid platform," states Balkenende, "makes it easier



Distributed

Data is also more distributed than ever, with more data sources in a broader array of environments, including on premises systems, SaaS applications, the Internet of Things, and third-party data sources, among others. To be most effective, companies must find ways to manage the different places where data lives, figure out the best ways to connect to it, and, if needed, move it to or from a data lake or data warehouse.

Balkenende remembers the hassle involved in dealing with distributed data in years past. He recalls, “When we bought data from companies, we had to have DVDs shipped, then load them onto a UNIX machine or run FTPs for hours.” Dealing with a distributed data environment has come a long way since the days of FTP marathons. Today, users can access data directly online, via a link or in the cloud.

“There’s lots of great stuff going on,” says Balkenende, “including enterprises using Snowflake as their central Enterprise data warehouses (EDWs) to collect and aggregate company business data, enabling data analytics to deliver actionable insights. The cloud provides the ability to make data more available to more people within and outside of organizations than ever before. And it’s easier to access third party data sources than most people think.”

Dynamic

Enterprises must also consider how to keep up with data that is constantly fluctuating in real time and subject to API changes, drifting schemas (metadata changes within data sources), and other shifts. Balkenende notes, “Dynamic data is primarily coming from Software as a Service (SaaS) applications, which are especially subject to metadata modifications, and which enterprises are adopting more and more, and have little control over the metadata changes.” To address this issue, he recommends that businesses really understand the key systems and platforms they need to control and recognize how dynamic they are. For example, how often are the endpoints changing? Or are the payloads of the APIs changing, and how often are new fields showing up? Balkenende also suggests that organizations remain flexible in how they manage data, and that they use a robust platform that can automatically adapt to schema drift to avoid data pipeline failure.



AI and ML

AI and ML are additional considerations for companies that want to adopt a modern analytics framework. A recent **Forbes Technology Council** post states, “AI will play a big role in making the explanation of answers accessible, consistent and much more engaging. If applied effectively, organizations can spend less time debating results and more time taking action.” These technologies support organizations by:

- **Using powerful computing to make meaningful connections**
- **Transforming those connections into actionable insights**
- **Working and making decisions at a faster speed than human capabilities**
- **Modeling potential decision outcomes and prescribing behaviors based on those insights**

AI and ML can be deployed in two distinct ways. One is to automate data ingestion, which involves figuring

out which data to load from which set of sources and how frequently — relying on metadata and data intelligence — in order to perform data analytics functions. The other is taking the results of AI and ML processing and putting the transformed data back into operational analytics to drive deeper and faster insights.

These technologies are becoming much more prominent for business intelligence (BI) within companies across industries. “It’s one thing to have BI dashboards to look at revenue and other lagging indicators,” comments Balkenende, “but another to use AI- and ML-based prescriptive analytics to know what customers are doing next, whether customers are getting ready to churn, or how the market will change.” For example, insurance companies use these systems to raise or lower rates based on projected weather patterns.

Such uses can be coordinated with SaaS applications like Salesforce to ensure that customers are alerted if the price of their policy changes. “BI systems powered by AI and ML are becoming so essential that teams not using them will likely be unable to compete in the coming years,” says Balkenende.

From Anywhere to Everywhere

Companies who want to manage data effectively must strive to improve the accuracy of analysis by pooling all their data together and drawing from one source of shared, secured, and connected data accessible to everyone who needs it. Cloud Data Platform (CDP) provider **Snowflake** explains, “Data sharing as a technology has become highly relevant for businesses of all sizes, whether they need to disseminate data across a large, global organization or need to augment internal data with broader market data to gain better insights.” This approach increases agility by using data smartly throughout the enterprise and enables the speed needed to quickly answer urgent business questions.

Data lakehouses enable data engineers, scientists, and analysts to share data assets. The lakehouse concept combines the features of a data lake and a data warehouse. Balkenende explains, “The data lake is the primary place to store ingested raw data that hasn’t been transformed. This stage provides the opportunity for AI and ML operations. The data warehouse is where data transformation takes place, making the data conform to business rules to be used for specific purposes like BI and operational analytics.”

“Data sharing as a technology has become highly relevant for businesses of all sizes, whether they need to disseminate data across a large, global organization or need to augment internal data with broader market data to gain better insights.”

— Cloud Data Platform (CDP) provider Snowflake

From Anywhere to Everywhere



These two areas have traditionally existed in different silos and separate hardware, using different platforms. The “lakehouse effect” brings them together on a single platform in the cloud. “You can put it on one single platform,” Balkenende states, “by building a warehouse and using data science practices on one set of data. With this approach, everyone is working off the same starting point of data for their EDW and BI operations.”

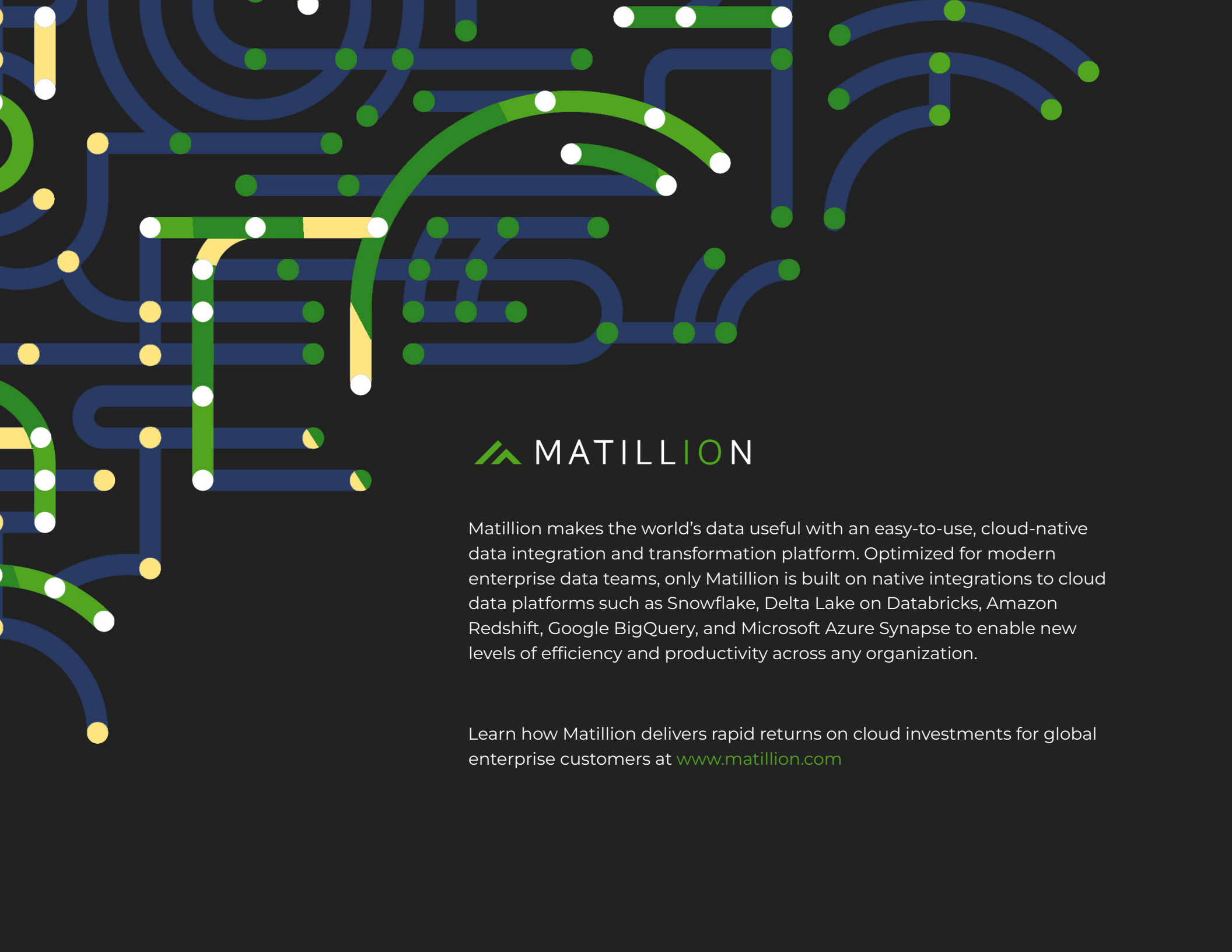
Companies that use the lakehouse approach enable users to perform self-service data access and analytics faster and with better data. By placing data in the cloud, users get access to data sets easier and faster, providing a foundation for increased data literacy, but data literacy is a must. Balkenende notes, “Company programs are teaching employees where the right data is and the importance of the right people using it at the right time. These organizations are then enabling self-governing and self-access, which are also critical for the greatest effectiveness.”



To stay ahead of the competition, modern enterprises must use modern analytics methods and tools. Matillion offers a data integration and transformation platform and a range of services and support to help businesses get started and make the most of their investments.

- **Matillion ETL** is a data management platform that is built for the cloud. Only Matillion's ELT software natively integrates with Snowflake, Delta Lake on Databricks, Amazon Redshift, Google BigQuery, and Microsoft Azure Synapse, to help enterprises take advantage of the power, flexibility, and economics of the cloud.
- **Matillion Data Loader** helps businesses deal with dynamic data and handle schema drift by enabling them to effortlessly load source system data into their cloud data platform. Matillion Data Loader, our SaaS-based integration tool, provides quick access to data, helping enterprises accelerate innovation and make faster, better business decisions. It provides both batch replication and streaming Change Data Capture (CDC) loading functionality.

Companies that want to take full advantage of cloud technology must choose data integration and management tools built for the cloud — not built 20 to 30 years ago for legacy systems that have been retrofitted to the cloud. These modern tools perform actions in ways that take full advantage of cloud functionality and handle diverse, distributed, and dynamic data in ways that are most beneficial. Companies that understand these underpinnings of modern methodology and invest in cloud-native solutions like Matillion are positioned well for success now and in the years ahead.



MATILLION

Matillion makes the world's data useful with an easy-to-use, cloud-native data integration and transformation platform. Optimized for modern enterprise data teams, only Matillion is built on native integrations to cloud data platforms such as Snowflake, Delta Lake on Databricks, Amazon Redshift, Google BigQuery, and Microsoft Azure Synapse to enable new levels of efficiency and productivity across any organization.

Learn how Matillion delivers rapid returns on cloud investments for global enterprise customers at www.matillion.com

studio / **ID**

BY INDUSTRY DIVE

studioID is Industry Dive's global content studio offering brands an ROI rich tool kit: Deep industry expertise, first-party audience insights, an editorial approach to brand storytelling, and targeted distribution capabilities. Our trusted in-house content marketers help brands power insights-fueled content programs that nurture prospects and customers from discovery through to purchase, connecting brand to demand.

[**LEARN MORE**](#)