ATSCALE

Leveraging governance to improve the value and accuracy of big data

Virtualization enables agile, secure access to consistent data for better decision-making

Performance, security and agility are the elusive trifecta organizations strive to achieve with the adoption of emerging data management platforms. The ability to achieve all three in a single implementation is one of the greatest challenges facing enterprises globally. Anecdotally, data science, one of the most valuable use cases, is plagued by the questionable integrity of the information available for model building activities. According to Paxata[1], data scientists spend only 12% of their time on analysis and as much as 42% profiling/preparing data, and another 16% on data quality evaluation; with these statistics it's difficult to imagine organizations experiencing the necessary return on investment these endeavors require.

As environments become more distributed, a by-product of the emerging cloud data management platforms, enterprises must ensure that data quality and integrity are the foundation for all analysis and data science activities, something easier said than done. The logical response is intelligent data virtualization, which effectively turns large disparate pools of data into a single, logical representation of information, serving as a catalyst for analyses of all kinds—from Business Intelligence (BI) to Artificial Intelligence (AI) and Machine Learning (ML).

> Data governance continues to be the cornerstone of all data management and access activities, ensuring users are within the boundaries of corporate and industry oversight while maximizing the potential value from the underlying asset itself, the data. Ultimately, implementing data governance improves the quality of data and will help organizations to make better business decisions. The goal is to create an insight-driven enterprise where business opportunity is discovered and cultivated with unparalleled performance, security and agility.

Data governance best practices include:

1. Compliance and enforcement with regulatory requirements

2. A single definition of the business relationships within the data

3. Quick, reliable access to data

4. Consistent results across analytical tools/user experiences

1 Paxata, The State of DataQuality in the Enterprise, 2018

1   Engineer compliance with regulatory requirements

Government regulators are implementing laws to protect the privacy and integrity of data related to citizens. Examples of legislation include GDPR and the California Consumer Privacy Act, going into effect in 2020. Such laws will safeguard the data of private citizens by holding accountable organizations who willfully violate consumer privacy by selling or distributing customer data without permission, or who expose consumer information through negligent security practices. In GDPR this requirement is called "Privacy by Design" where data systems must be engineered to secure data from exposure or leakage by default.

As a centralized conduit for orchestrating secure and governed access to the multitude of data sources within the enterprise, intelligent data virtualization is the perfect vehicle for implementing Privacy by Design. Working with virtualized data eliminates the need to create extracts, greatly reducing the chance data will be mishandled.

> **Intelligent data virtualization also covers a variety of data protection scenarios, including data encryption at rest and in flight, further ensuring Privacy by Design.**

2   Create a single logical interface within an increasingly distributed world

Intelligent data virtualization establishes a highly performant, agile and governed view of an enterprise's analytical information regardless of the user persona or query interface. Every user connecting to the intelligent data virtualization fabric has the ability to track the lineage of their data, resulting in greater visibility and certainty that data sources are reliable and consistent with those of their peers. Data analysts are able to create a shared data intellect where recommendations can be made with confidence that they are aligned among fellow analysts and data scientists.

3   Ensure agile, accurate access to data

Organizations can facilitate the creation of the insight-driven enterprise by leveraging autonomous data engineering and acceleration structures. Intelligent data virtualization applies advanced AI and Machine Learning to study query patterns and database statistics to create highly secure and governed data structures that accelerate a broad range of analytical workloads. In many cases this creates an

order of magnitude performance enhancement while simultaneously reducing overall query cost and complexity. Queries return interactively in seconds rather than minutes or hours. The dramatic increases in performance and agility facilitate the creation of insights while simultaneously steering the experimentation of new avenues for research.

4   Provide Consistent Results Across Analytical Tools

Different departments will naturally develop their own preferences in AI/ML and BI tools. Supporting a variety of user interfaces to the data creates the broadest opportunity for team enablement, but also amplifies the challenges of a single, fully compliant data experience that maintains comprehensive security and governance.

The key in any AI/ML or BI user experience is consistency—without a single location for defining the "business context" within the data in a highly scalable, performant and secure way, skepticism rules the day and room for "interpretation" persists. Yet another benefit of intelligent data virtualization is the consistency created by alleviating the different characteristics of each of the individual tools. When leveraging a variety of data management interfaces and heterogenous physical or virtual data platforms, the semantic translation provided by intelligent data virtualization ensures users receive the same responses, irrespective of platform(s) of choice.

USE CASE: Toyota & AtScale

Toyota improved Operational Analytics through Data Governance

- ▲   Created a single, logical data warehouse from dozens of DBs

- ▲   Delivers sub-second queries for Finance & IoT

- ▲   Real-Time user experience for time-to-insight

- ▲   Supports ad hoc queries on Tableau and PowerBI

- ▲   Enable & protect enterprise data consumption

Don't just set-it-and-forget-it. Data governance will require iteration as it changes and grows.

Good data governance requires data to be agile, fast, accurate and consistent for the best insights. These outcomes are often blocked by the complicated and abundant technical hurdles associated with bringing many disparate data sources together. Data governance requires continuous adaptation to new compliance, business, and security challenges. To keep up, the only way to address the size and scale of big data is through intelligent data virtualization which enables an automated orchestration and acceleration—empowering data analysts and business users to quickly derive the insights they need to position the business for competitive, operational and/or revenue insights.

ATSCALE