

The Business Value of a Data Hub Platform

WHITE PAPER · APRIL 2021

A core element of digital transformation efforts is the ability of organizations to deliver trusted, reliable data to business users exactly when they need it. Businesses have come to recognize that data made available on this basis is the secret sauce of quicker, more accurate decision-making, growth and competitive advantage.

However, the data reality in many organizations today is very different, particularly for those relying heavily on mainframes and other legacy systems of record. Instead of fluid, trusted flows of data to key employees, they are running up against a proliferation of data silos. Business-essential analytics tasks often involve manually compiling data into spreadsheets. IT teams labor setting up data lakes and data warehouses to integrate data from increasingly disparate sources.

And instead of making accurate and speedy business decisions, knowledge workers and managers spend hours finding and often fixing spreadsheet data and questioning how trustworthy the data actually is. With proliferating data silos also come mounting calls to IT for support in getting the data users need.



Time Wasted, Opportunities Lost

For IT, the workarounds put in place to integrate data create serious security and governance problems. It wasn't long ago that a typical enterprise had to deal with a handful of core data services, three to five in many cases.¹

Today, that number has exploded as the result of cloud services, process outsourcing to SaaS, mobile work, automated reporting processes and connected Internet of Things devices, to name a few factors. All this and other data is cascading into warehouses, data lakes, data marts and other data aggregators running throughout a hybrid IT infrastructure.

As McKinsey notes, all these data aggregation additions have drastically *increased* the complexity of data architectures.² They hamper the ability of a business to “deliver new capabilities, maintain existing infrastructure and insure the integrity” of key IT initiatives that rely on durable data, such as artificial intelligence.

Recognizing the data dilemma in which many businesses are mired today, some IT departments have tried various ways of gaining order from data chaos.

The results of these data swamps are just plain bad for business. *MIT Sloan Management Review* estimates that knowledge workers waste up to *half their time* dealing with data issues, much of that involving data trust, quick access and availability.³ Instead of functioning as a data enabler for business decision-making, IT spends far too much time fixing and patching brittle legacy systems and writing numerous, one-off APIs to access data in various places.



Data Swamp Work-arounds

Recognizing the data dilemma in which many businesses are mired today, some IT departments have tried various ways of gaining order from data chaos. But they all have one thing in common: They don't work to deliver what knowledge workers and managers want and what the business needs. Some of the common stopgap efforts include:

- Continue building upon the relational data marts and data warehouses that just increase IT complexity and data chaos, efforts that feed into existing legacy thinking. These constructs can work for data scientists but seldom for knowledge workers and managers.
- Double down on current extract, transform, and load (ETL) strategies in a crazy-quilt effort to integrate data from disparate sources when those sources are rapidly multiplying.
- Copy tidal volumes of data into disparate systems, an often manual and therefore mistake-laden strategy that adds expense and increases time to value.
- All the while, ignore requirements for holistic security and data governance that deliver consistently trusted data.

¹ “Why Today’s Data Explosion Is a Ticking Time Bomb for IT,” *The New Stack*, Sept. 11, 2019

² “How to Build a Data Architecture to Drive Innovation—Today and Tomorrow,” *McKinsey Digital*, June 3, 2020

³ “Seizing Opportunity in Data Quality,” *MIT Sloan Management Review*, Nov. 27, 2017

Data Hubs as a Solution for Creating Durable Data Assets

What businesses need is for IT leaders to deliver a solution that creates a 360-degree view of enterprise data—regardless of its source, volume or data type. That solution is a data hub platform. Think of it as a centralized, one-stop service connecting all current and future IT systems, including platform staples such as ERP and CRM, as well as web applications and SaaS solutions.

Another way to think about a data hub platform is not as a technology but a business approach that provides security and governance as well as visibility into how data will flow seamlessly to knowledge workers and managers, enabling trusted, data-driven business decision-making.

Consider this: If you have, say, 50 systems in your IT environment, which is not uncommon in an enterprise, IT would need to provide as many as 5,000 connections to enable them to talk to one another and share data. It is simply untenable. But sharing data among these systems is also business-critical today. In contrast, with a data hub platform, only one connection per source system is required—50 connections vs. 5,000.

How It Works

A data hub platform establishes connections to each IT system, and that connection is shared with any and all other systems that interact with it. This sharply reduces the need for time-consuming and costly data replication to support various business processes.

An operational data hub effectively places a barrier, or “fence,” around legacy systems such as mainframes and aging relational DBMS. The hub permits new applications to go directly to the hub rather than to all existing systems. IT and the business gain an actionable, 360-degree view of enterprise data, as the operational hub [allows the building of new analytical and transactional processes on the data](#).



A multi-model data hub avoids costly ETL by ingesting data just as it is. And the data hub is secure and governed by default, meaning knowledge workers and business managers can fully trust the data on which critical business decisions may rest.

Case Studies: Data Hubs at Work

Data platform modernization at WoodmenLife

This not-for-profit insurer recognized the need to modernize its data infrastructure as it began investing in digital transformation to deliver better customer service. Specifically, IT management at WoodmenLife sought to replace mainframe-based policy management systems and migrate all historical claims data to a hub capable of delivering a complete customer view of its members.

Doing so meant making unified, accurate customer data from all systems readily accessible and easily searchable by users. The company had been [running DB2 on the mainframe and SQL Server elsewhere](#).

Working with [MarkLogic](#), the WoodmenLife team planned a data hub platform, taking only six weeks to complete the proof of concept and launch the first hub-based production application.

MarkLogic built a basic user interface for business users to easily access all policy data and unstructured data, such as PDFs and images, from all systems. WoodmenLife said MarkLogic interoperated smoothly with existing legacy technologies as well.

Next up for WoodmenLife is using MarkLogic to integrate member and operational data across all business lines, enriching it with new external data sets as the company moves ahead with its digital transformation efforts.

Eaton moves from data lakes to a data hub

A venerable global industrial corporation, Eaton has an intense interest in product data, given it makes 7 million different products. Initially, Eaton created a relational product data hub, or data lake. But getting useful data from it in a timely way proved difficult and costly. For example, with XML in the RDBMS, almost no one could get at the XML data they needed without IT having to concoct an expensive, one-off effort.

Eaton's IT management decided to [move all product data into MarkLogic](#) to speed data delivery to business users so they can formulate better and quicker business decisions. Almost immediately, Eaton found the MarkLogic data manipulation tools to be "exceptional," particularly when working with Eaton's voluminous XML documents.

IT also praised MarkLogic's advanced data security capabilities, allowing IT to easily filter product data depending upon who is going to consume it. For example, distributors get certain data sets, suppliers others, and internal users yet others.

The Unique Business Value of MarkLogic

As the visionary and pioneer of operational data hubs, MarkLogic has dedicated its efforts to working with enterprise organizations, like the two above, to create durable data assets. A big part of its unique value proposition is its eponymous multi-model NoSQL database foundation, an elegant solution to the thorny challenge of managing heterogeneous data.

In contrast to polyglot persistence, wherein an application integrates multiple database models, [MarkLogic's multi-model approach offers native support for multiple data models](#) using a single,

integrated back end. Thus, MarkLogic avoids the costly, productivity-sapping silos that result from polyglot persistence requiring complex integration workflows. Instead, MarkLogic provides a single unified interface for data consistency, security and user access.

[As one MarkLogic user remarked](#), "It is a secure, agile and scalable platform providing a multi-model view. The database has great search abilities, and has advanced APIs and tools that help in fast deployment and application."

As the visionary and pioneer of operational data hubs, MarkLogic has dedicated its efforts to working with enterprise organizations.

MarkLogic vs. Relational

As many enterprises today remain attached to their legacy RDBMS, MarkLogic undertook a proof of concept (POC) for a Fortune 500 insurer whose POC team had no previous experience with MarkLogic. Here are the results of that test.

The POC team using MarkLogic delivered across 67 proof points and five extra-credit items, loading both current and historical data and building services atop the database. This took the team 300 hours. The relational SQL team spent 1,700 hours and never even got the data loaded.

Overall, MarkLogic delivered a sixfold improvement in time to value for the company, owing in part to MarkLogic's flexible data model and its ability to load all data "as is" regardless of schema changes. In contrast, relational data models are brittle, requiring extensive ETL and excessive amounts of hardware.

To learn more about how MarkLogic delivers data agility, visit marklogic.com. To see MarkLogic in action, sign up for our live demo.

