denodo

Whitepaper

Data Fabric for the Supply Chain

Introduction

The pandemic, which directly impacted the global supply chain with shortages of even basic necessities, started to show signs of recovery in 2023. In June of that year, the United States White House issued a two-year supply chain report card that cited, among other statistics, that the number of backlogged, anchored U.S. vessels dropped from a high of 155 to about a dozen in May.

However, no organization can afford to go back to business as usual, assuming this were even possible. Many companies quickly pivoted to online-only, or hybrid brick-and-mortar/online models. Similarly, IT departments relied more on cloud-based systems, or quickly learned to adopt hybrid on-premises/cloud systems.

Organizations that work with supply chains felt the brunt of this disruption. They learned that technological agility is not a nice-to-have, it is a necessity that makes the difference between success and failure. They learned that regardless of which technology they leverage, it has to support the unexpected.

Unfortunately, many such organizations are unable to gain this level of agility, as they are constrained by data management infrastructures built around traditional data warehouses. These data warehouses are powerful, reliable, and feature-rich, but compared with recent modern alternatives, they are beginning to exhibit their disadvantages, and they just cannot provide the necessary agility to remain unaffected by the next disruption.

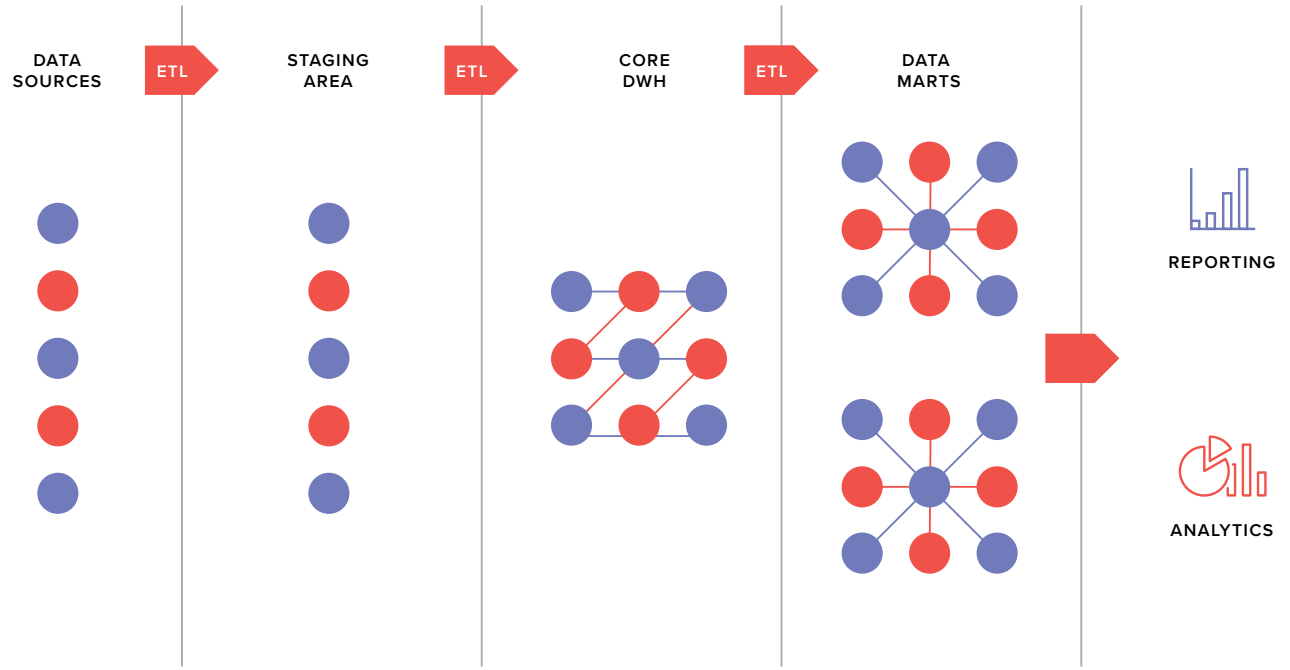
This whitepaper discusses data fabric as an effective alternative. Data fabric is a logical data management solution that can prepare organizations for the slightest change in supply-chain dynamics. The paper begins with a summary of the most disruptive supply-chain challenges, proceeds to a description of data fabric, explains how data fabric can overcome each of the challenges, and ends with a case study of Hastings Deering, which leveraged a logical data access layer, the foundation for a data fabric, to gain numerous supply-chain efficiencies, including accelerated delivery.

Supply Chain Challenges

Though the supply chain delivers goods, it is controlled by data. The main limitation of traditional data-warehouse-based data infrastructures is that they are not built to support real-time use cases. Instead, they are designed to store and process large volumes of structured data and make it available for reporting via the delivery of large, scheduled batches.

The traditional data warehouse itself is geared for structured data; that is because the expected use case is automated delivery in bulk. Data is delivered to traditional data warehouses via multiple extract, transform, and load (ETL) processes. Also, because traditional data warehouses require structured data, they cannot accommodate many modern data types, such as streaming and cloud data, which is an additional limitation.

Traditional data infrastructures also cannot easily accommodate changes in the data infrastructure, such as the removal, change, or addition of a data source. To support any of these activities, the appropriate ETL script needs to be written or revised, and this usually requires testing in a staging environment before it can be taken live.

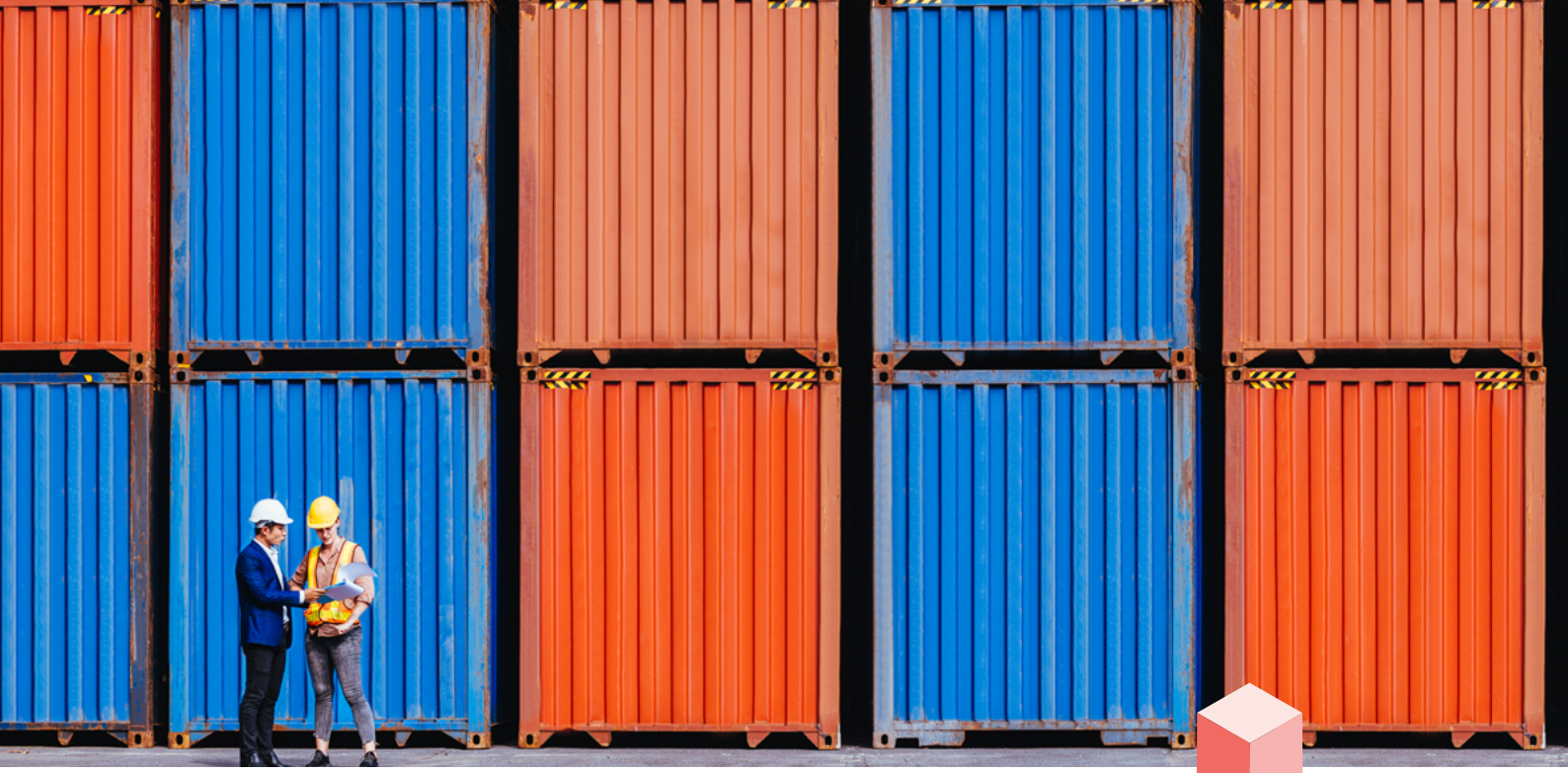


For these reasons, organizations that rely on traditional data warehouse infrastructures often face the following challenges:

POOR VISIBILITY

Traditional data warehouses limit visibility into the supply chain in two ways: First, supply-chain visibility without real-time data is like driving a car based on written reports rather than the view through the windshield. Second, because of the restricted interoperability of traditional data warehouses, and despite all efforts to centralize all available data in the data warehouse, data sources such as cloud-based repositories, data lakes, Internet of Things (IoT) and other streaming sources, remain as data silos. Additionally, newly acquired data assets will also often be siloed for a period of time, due to the development work required to connect them to the data warehouse via an ETL process. This leaves organizations with delayed, fragmented visibility into tier 1 suppliers, tier 2 suppliers, and beyond.

With limited visibility, if a supplier down the line is experiencing a delay, upstream parties might not know until it is too late to take corrective action. As a result, organizations are forced to operate in a reactive mode, rather than a proactive or predictive mode. This limited visibility also results in an inability to accurately forecast demand, which many organizations experienced in a very direct way when the pandemic first appeared.



FREQUENT DATA ERRORS THAT CREATE BOTTLENECKS

In addition to limited visibility, the lack of a unified view over all data results in several other effects that inhibit overall agility. For example, organizations often unknowingly create unnecessary copies of data. This is not only inefficient, but also a recipe for inconsistency, because data copies often get modified individually, or fail, for a variety of reasons, to receive the latest update. Data governance, which would normally prevent such inconsistencies, is inhibited in a siloed data environment, simply because the protocols cannot reach all of the sources simultaneously. For the same reason, there is no way to control the different sources in a straightforward, concerted manner.

THE INABILITY TO EASILY OPTIMIZE FOR ON-TIME DELIVERY

Errors aside, traditional data warehouse infrastructures do not enable the modern supply-chain optimizations provided under Industry 4.0 and smart manufacturing initiatives, with their reliance on real-time intelligence drawn from multiple sources, including streaming IoT data. With the improved automation enabled by this type of data access, organizations would be able to further reduce errors and optimize delivery. In a perfect world, organizations would be able to establish digital twins of all of their shipments, through which they could monitor and manage every moving part in the supply chain. However, such capabilities are not possible without real-time access to all available data.

LIMITED DECISION-MAKING CAPABILITIES

Clearly, the lack of real-time access to unified data would also impact decision making, but this plays out on several different levels, and it is crucial to understand each of them. For example, sometimes the issue is receiving data in a timely fashion, either from the data warehouse, from outside the organization, or from another department, but often, the issue is not necessarily accessing the data but being able to leverage the data for decision making. This condition is often described as being “data rich but insight poor.”

To be used effectively, data often has to be transformed into business-friendly terms, or at least into terms that are friendly to its expected audience. The challenge is that often multiple users will access the same data, from developers, to business users, to customers, to partners. Yet it is difficult to create different versions of the data without creating different copies, which results in a data management quagmire.

Complicating this situation, organizations often have different definitions, standards, and key performance indicators (KPIs) for the same entities. As part of the same data governance issue mentioned above regarding frequent data errors, this condition creates a semantic obstacle that prevents stakeholders from leveraging the data for analysis. This is a huge issue. We often speak of wanting to be “data driven.” Yet having data is not enough, if it can’t be easily leveraged for analytics that move a company forward. So organizations should really think in terms of wanting to be “insights-driven.”

Similarly, organizations should empower business users to access the available data (real-time or not), in a self-service manner, over a simple web browser, ideally via a trusted, real-time dashboard, without having to put in a request to IT. Unfortunately, in a traditional data warehouse environment, someone with specialized technical expertise is necessary to access the different siloed data sources, each with their different requirements, and furnish the necessary results. Traditional business intelligence (BI) is often severely impacted by IT backlogs, queues, and pipelines.

THE TECHNOLOGY IS NOT FUTURE-PROOF

Once again, the pandemic reminded us that we always have to be ready for that next disruption. Technology, by its very nature, underlines this fact. This is why data infrastructures should be as agile and adaptable as possible, not tied to specific technologies or one-off integrations between two specific systems. Traditional data warehouses, the tried-and-true workhorses of the last two decades, and the infrastructures that support them, comprised of large numbers of ETL processes integrating individual systems to the data warehouse, are beginning to be referred to as “legacy” technology, and for good reason.

These infrastructures are built around the process of IT implementing the physical transfer of data before it can be analyzed, and Gartner, Forrester, and other analysts have been pointing out that not only does this scenario exhibit profound inefficiencies, but the future does not necessarily need to look this way. To be ready for that next disruption, whatever it may look like, organizations need future-proof infrastructures, and traditional data management approaches are just not going to serve this goal.

Introducing Data Fabric

Data fabric is a logical data management architecture. Rather than relying on the physical replication of data from multiple source systems into a central data warehouse, data fabric enables real-time access to data, across myriad different kinds of data sources, while leaving the source data in its original locations. Leveraging real-time data from across the organization, data fabric goes one step further: By leveraging active metadata from across the organization, it enables the intelligent automation and augmentation of many tasks, for a fully flexible, responsive data-access layer worthy of the name "data fabric." For these reasons, data fabric can overcome each of the challenges discussed in the introduction.

Data fabric is an extension of another logical data management architecture, the logical data warehouse, which is, in turn, an extension of the traditional data warehouse. So before we talk more about the benefits of data fabric, we will first give the logical data warehouse a proper introduction.

THE LOGICAL DATA WAREHOUSE

In 2008, Gartner analyst Mark Beyer first used the phrase "logical data warehouse" in a conversation with a client, and he further developed the concept in a blog post in 2011, in which he refers to himself as the "father of the logical data warehouse."[1]

In this post, Beyer describes the logical data warehouse not as a "new and improved" data warehouse, but as a semantics layer that sits above a data warehouse. The word "semantics" here is key, because two systems can only share data if they speak the same language. In Beyer's architecture, the semantics layer acts as an interpreter, able to speak to any underlying data source. To access any data set, data consumers talk to the semantics layer which, in turn, talks to the source.

Since that post in 2011, the logical data warehouse architecture has been extended and developed by other analysts. Philip Russom, when he was senior director of The Data Warehouse Institute (TDWI), defined the logical data warehouse as simply "a logical or virtual layer of the data warehouse architecture that integrates the physical layers of architecture under it." In his definition, the "logical or virtual" layer maps to Gartner's "semantic" layer, and the "physical layers" map to the "traditional data warehouse."

In both definitions, data consumers access the data through a layer that sits above the separate data sources, bringing them together for reporting. Note that the logical data warehouse contains no actual data; it only contains the metadata for accessing each of the sources.

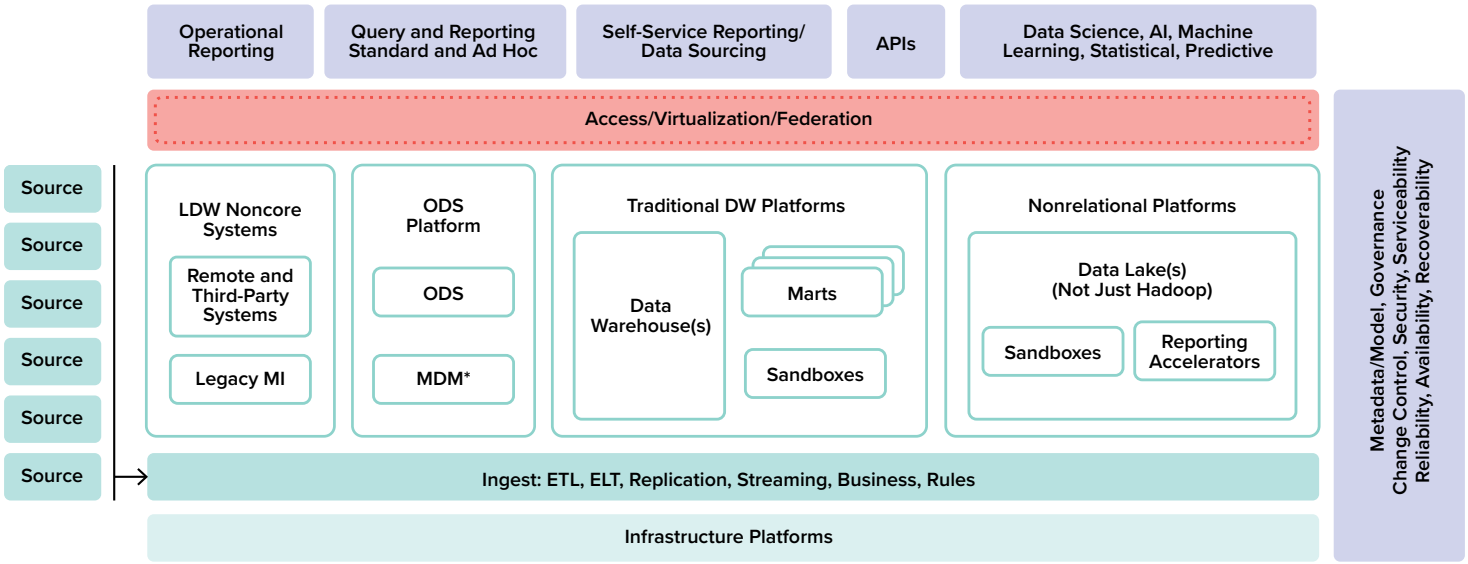
[1] See https://blogs.gartner.com/merv-adrian/2011/11/03/mark-beyer-father-of-the-logical-data-warehouse-guest-post/

Logical data warehouses are particularly suited to operate over core enterprise repositories such as traditional data warehouses, operational data stores, data marts, and data lakes.

Going back to Beyer, in his 2011 post he says that logical data warehouses focus on "the logic of information and not the mechanics that are used." This is accomplished through data virtualization.

Data virtualization is a data integration technology, but unlike ETL processes, it not only integrates data without replication, that is, leaving the data in its original location or locations, but it does so in real time.

High Level LDW Architecture



*MDM, likely in a separate system but can be regarded as part of LDW

It is important to note that although logical data warehouses can be implemented as a semantic layer above traditional data warehouses, they do not require traditional data warehouses in order to function. All that a logical data warehouse needs is one or more data sources, and these can easily be comprised of a traditional data warehouse; to the logical data warehouse, a traditional data warehouse is "just another data source."

Logical data warehousing has evolved and matured over the years, and many companies have leveraged it. The next step in modernizing a data infrastructure is to turn a logical data warehouse into an enterprise-wide data fabric.

DATA FABRIC: EXTENDING THE CAPABILITIES OF THE LOGICAL DATA WAREHOUSE

We mentioned that logical data warehouses, supported by data virtualization, leverage metadata to enable real-time, logical connections to traditional data warehouses, operational data stores, data marts, and data lakes.

Data fabric, also supported by data virtualization, performs all of the functions of a logical data warehouse, but it leverages metadata to extend these capabilities in three key ways: First, it leverages metadata to enable connections to an essentially unlimited variety of data sources, including streaming data, such as machine feeds in the Internet of Things (IoT), which is critical for supply-chain management. Second, it leverages not only "passive" metadata such as descriptions of fields and access credentials, it leverages "active" metadata of user activity, data movement, and machines, to enable powerful automation, especially combined with artificial intelligence (AI) and machine learning (ML) capabilities. Third, it can leverage both active metadata and AI/ML capabilities to enable a recommendations engine that suggest different datasets for individual users to access, based on their profile and real-time activity. Data fabric can also augment and automate key data catalog tasks, such as metadata discovery and ingestion.

How Data Fabric Can Overcome Supply-Chain Challenges

In this section, we revisit each of the five challenges covered in the introduction and explain how they can be overcome by data fabric.

IMPROVED VISIBILITY

With data fabric, organizations can gain real-time visibility across all of the different moving parts in the supply chain and across the different tiers in the supplier network. Organizations can leverage data fabric to streamline reporting for internal users and easily expose rich, integrated data sets to external partners via APIs. Any partners that also share supply chain information via APIs can be added as a data source as if it were just another internal system, making it easy for internal consumers to have a complete view of the supply chain. Tier 2 suppliers might still be reluctant to share intelligence with original equipment manufacturers (OEMs), however technology would no longer be an impediment, and organizations would be equipped to see all of the available information at their disposal.

Organizations can adopt a proactive mode rather than a purely reactive mode, and by leveraging data science capabilities, they can also adopt a prescriptive mode. Hastings Deering (see the case study below) leveraged a logical data access layer to enable the prescriptive maintenance of heavy equipment. Cause-and-effect analysis and forecasting are significantly improved with improved visibility.

FEWER DATA ERRORS

Data fabric reduces data errors and improves data quality in four ways. First, by reducing replication and data silos, they reduce the likelihood of data copies becoming out-of-date. Second, by seamlessly interoperating with IoT data sources, they facilitate automation, reducing the number of human errors.

Third, they enable seamless data governance. By being positioned in a layer above an organization's myriad data sources, they enable not only simplified access across the various sources, but also simplified management. Data fabric enables organizations to implement data governance controls, including permission-based data masking, across all of the organization's data sources, simultaneously, from a single point of control.

Fourth, by looping all data sources into a centralized data-access layer, they provide data consumers with a single view of the data, simplifying the process of eliminating errors.

INCREASED ON-TIME DELIVERY

With data fabric, organizations can turn supply-chain optimization into an art. With a fully modern extension to their existing data warehouse, or simply with fully modern, interoperable capabilities, they will have the foundation for Industry 4.0 and smart manufacturing initiatives. Data fabric is perfectly suited to the enablement of digital twins, and as mentioned above, automation.

Regardless of the growing complexity of the data infrastructure, data fabric equips organizations to make the timely moves necessary to optimize delivery times and increase on-time delivery.

UNIFIED DATA FOR DECISION MAKING

The real-time data enabled by data fabric is powerful, but what is perhaps even more powerful is how data fabric enables organizations to gain much more value from their available data than what might otherwise be possible.

Recall how logical data warehouses are established as a semantics layer above an organization's disparate data sources (including traditional data warehouses), acting as an interpreter. Into the semantic layer of a data fabric, organizations are capable of building any number of additional, secondary semantic layers, without affecting the underlying data. This means that organizations can implement "Data Store A," using its own "front end" language, labels, rules, and functionality, to appeal to a certain department or group of users, and implement "Data Store B," to appeal to a completely different groups of people, without affecting the underlying data and without creating unnecessary data copies. Users can have the data "their way," in terms that make sense to them. Also, with enhanced data literacy, data consumers will have one less obstacle when embracing new developments in data sharing.

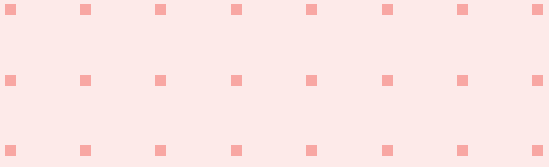
Organizations can also use a semantics layer to implement standard definitions for entities such as KPIs. This empowers data scientists and other data consumers to grab data at will, and know they are always leveraging the company's approved, trusted data.

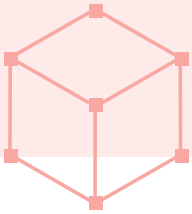
Data fabric also enables powerful data catalogs, and when the data delivery layer is tightly integrated with the data catalog, users not only get a list of all available data assets in a clear, user-friendly interface, it also enables data consumers to access the data directly from the catalog itself.

Going one step further, data fabric provides the necessary foundations for true self-service data access by a variety of users. Data fabric provides more than JDBC or ODBC support for SQL queries by business users – developers have a variety of other delivery options, including REST and GraphQL APIs, Kafka, and JMS message queues. Data can be integrated and brought into virtually any application. This, in turn, frees IT to focus on provisioning data and delivering datacentric functionality, rather than performing ad-hoc integrations and creating reports.

FUTURE-PROOF DATA INFRASTRUCTURES

Finally, due to the unique architecture of data fabric, how it decouples data consumption from the complexities of accessing the underlying data sources, data fabric enables organizations to swap out, maintain, replace, or fully modernize any number of the underlying sources without impacting business operations. With data fabric in place, users will have continuous access to data, even if it is in the process of being migrated. In this way, data fabric enables future-proof infrastructures that can adapt to any changes or innovations in the technology landscape.





How Hastings Deering Improved Analytics and Delivery Times

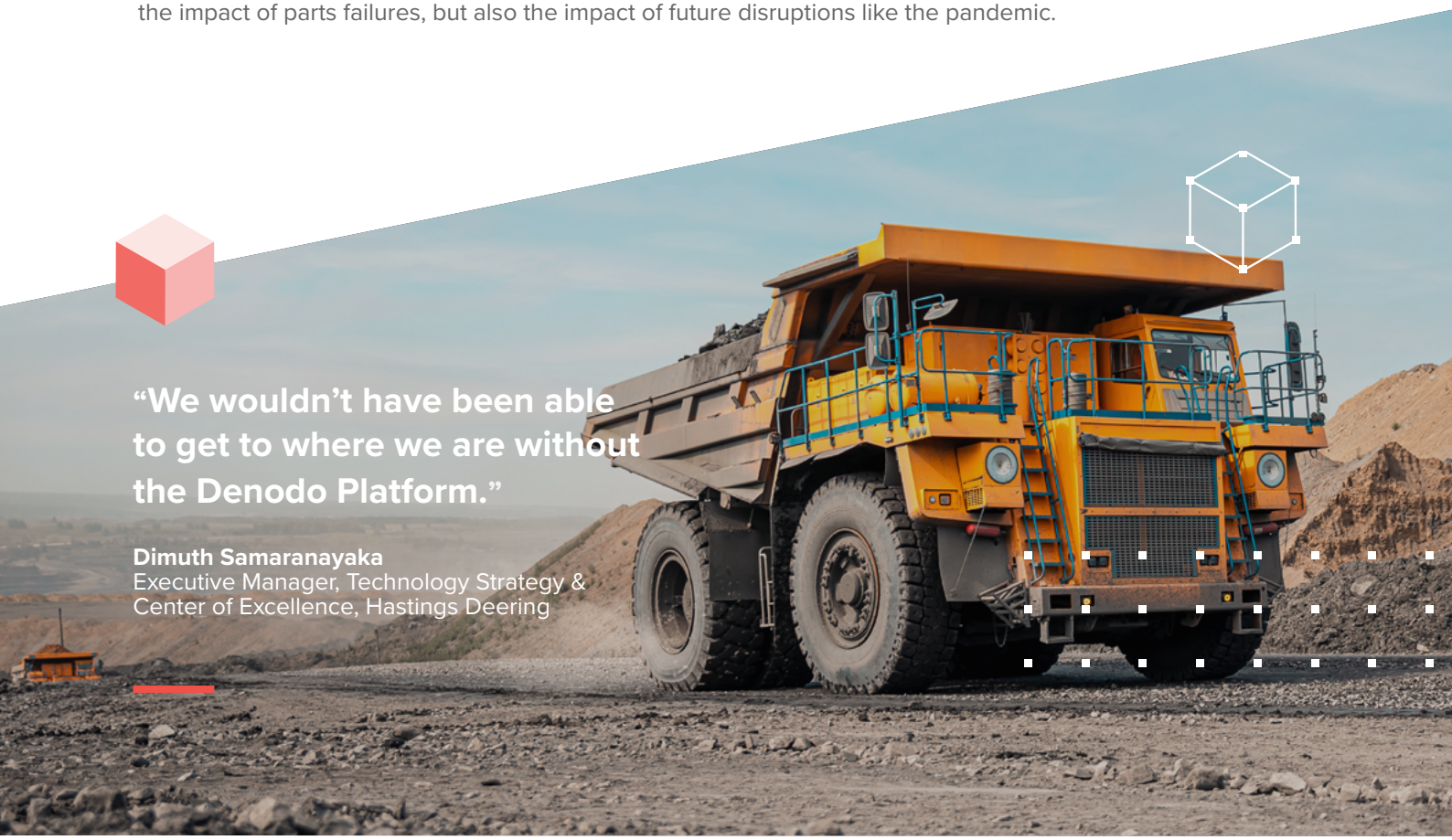
Hastings Deering Pty Ltd (a Sime Darby Industrial company), is one of the largest Cat® dealers globally servicing mining, construction, power systems, and marine industries.

After analyzing its supply-chain visibility and optimization capabilities, Hastings Deering determined that data silos were causing key problems. The company was generating a significant amount of data through multiple enterprise resource planning (ERP) systems, and this data needed to be integrated with several Caterpillar systems. In addition, the company was managing a variety of on-premises and cloud systems, as well as SaaS applications, all of which required time-consuming, highly manual processes in order to integrate the data between them for reporting. The poorly integrated silos also became out-of-synch, causing inconsistencies and outright errors, which delayed both time-to-data and time-to-value, impacting the supply chain and overall trust in the data.

Hastings Deering also began to see a shift in business requests, away from traditional data requests to data sharing requests to assist with quick insights, so the company knew that its data infrastructure needed to be more agile and flexible.

Hastings Deering implemented a logical data access layer, enabled by data virtualization, above the company's data sources.

With this logical data access layer, which enabled Hastings Deering to establish a real-time data marketplace with access across the different data sources, including streaming data from machines, Hastings Deering was able to implement predictive analytics in its Parts division, so that the company could better manage not only the impact of parts failures, but also the impact of future disruptions like the pandemic.



"We wouldn't have been able to get to where we are without the Denodo Platform."

Dimuth Samaranayaka
Executive Manager, Technology Strategy & Center of Excellence, Hastings Deering

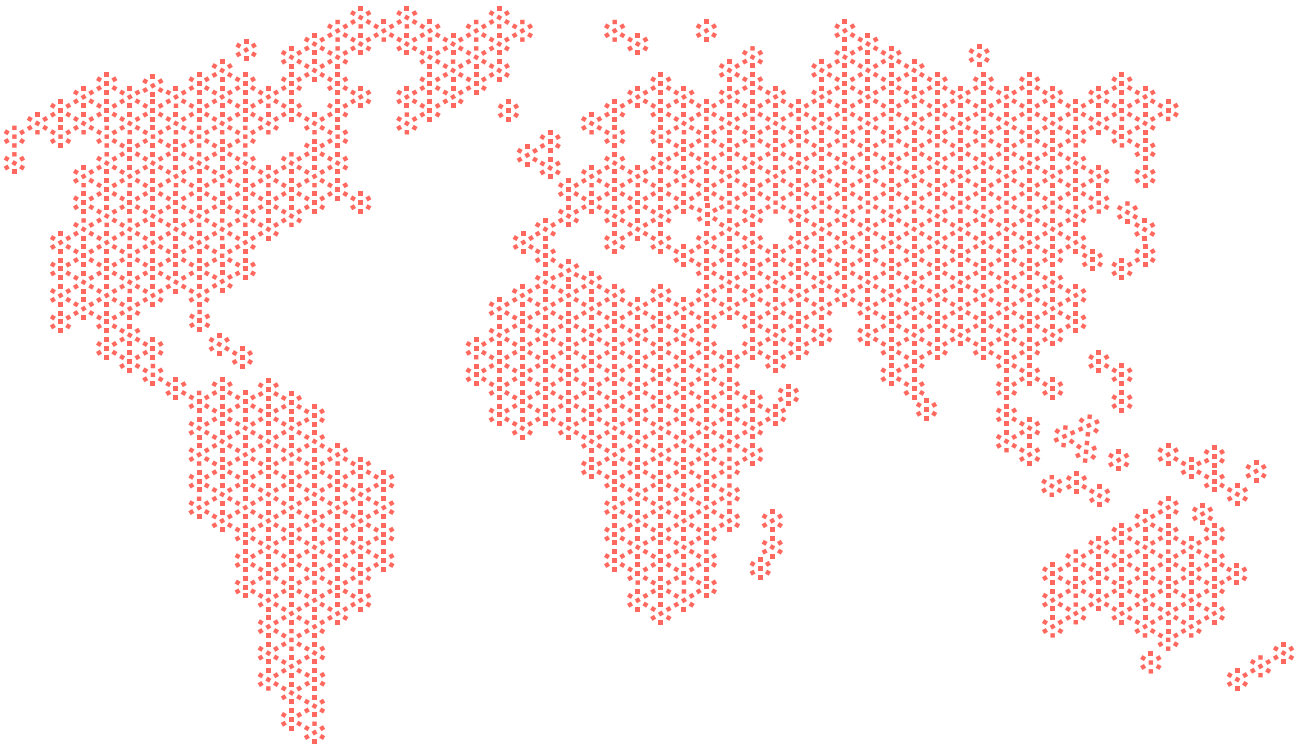
The new data marketplace became the central Hastings Deering data platform for reporting, intelligence, data sharing, and new digital programs. It also helped the company to establish data literacy across the organization as well as to develop comprehensive data governance programs. The new logical data access layer enabled Hastings Deering to develop data catalogs and business glossaries, together with guided self-service capabilities, which has greatly improved general trust in the data, as well as overall data literacy. Hastings Deering offered self-service analytics through a single access point to different personas across the business, from novice users to citizen developers.

Through the company's new technology and methodology initiatives, Hastings Deering has been able to accelerate the acquisition of important data sets, create more value from its data assets, and further enhance supply-chain forecasting capabilities to meet business needs.

Perhaps most importantly, by implementing the logical data access layer, Hastings Deering has been able to deliver a lot more in the same amount of time.

Embracing the Modern

Organizations can no longer accept the limitations of traditional data management approaches. Fortunately, they no longer have to. Data fabric, working seamlessly with existing, legacy infrastructure, or on its own, enables all of the benefits of modern data management, for maximum control over the supply chain.



denodo

Denodo is a leader in data management. The award-winning Denodo Platform is the leading data integration, management, and delivery platform using a logical approach to enable self-service BI, data science, hybrid/multi-cloud data integration, and enterprise data services. Realizing more than 400% ROI and millions of dollars in benefits, Denodo's customers across large enterprises and mid-market companies in 30+ industries have received payback in less than 6 months.

Visit www.denodo.com | Email info@denodo.com | Discover community.denodo.com