

Cool Vendors in Data Management

Published 1 June 2021 - ID G00746797 - 14 min read

By Analyst(s): Philip Russom, Ehtisham Zaidi, Jason Medd, Eric Thoo, Robert Thanaraj

Initiatives: [Data Management Solutions](#)

Data and analytics leaders should prepare for a future where many use cases rely on a data fabric that makes heterogeneous, distributed data appear consolidated, easy to access and business-friendly. With this future in mind, “cool vendors” are supplying functionality that can enable data fabrics.

Overview

Key Findings

- Data and analytics leaders are under pressure to squeeze more business value from enterprise data, especially for high-profile use cases in analytics, even when the data is heavily distributed across many systems and geographies, on-premises and on cloud.
- Multiple approaches to managing distributed data are maturing to support intelligent automation for augmented data management. The new functionality will enable a design pattern called the *data fabric*, which Gartner anticipates will eventually be a common way to manage distributed data.

Recommendations

Data and analytics leaders who are looking to modernize their distributed data management environments should:

- Prepare to adopt the data fabric by deploying selected components that are available today. This could involve pilot projects in data catalogs, active metadata, pipelines, orchestration, graph databases and cloud data management.
- Acquire key skills for the virtual and logical data models needed for data fabrics by piloting projects for data virtualization and logical data warehousing.

Strategic Planning Assumption

- By 2023, organizations utilizing active metadata, machine learning and data fabrics to dynamically connect, optimize and automate data management processes will reduce time to integrated data delivery by 30%.

Analysis

This research does not constitute an exhaustive list of vendors in any given technology area, but rather is designed to highlight interesting, new and innovative vendors, products and services. Gartner disclaims all warranties, express or implied, with respect to this research, including any warranties of merchantability or fitness for a particular purpose.

What You Need to Know

Distributed data is an opportunity for greater business value, not just a technology challenge. Distributed data is an enterprise asset; it is a resource that is naturally produced by multiple operational applications, IT use cases and business processes. The opportunity lies in leveraging this resource to create additional organizational advantage, wealth and growth, and greater efficiencies. The problem lies in the advanced techniques required for integrating distributed data, which can be time-consuming, expensive and not well-understood.

The data fabric is a modern approach to integrating distributed data. A data fabric requires a lot of technology, but the real goal of a data fabric is to make the business's use of distributed data more valuable, innovative, agile and automated. These are worthy goals, and trends in the market of data management tools are leading toward credible functionality for achieving them. To take advantage of these emerging capabilities, data and analytics leaders should prepare to design and deploy one or more data fabrics. (For a full definition of data fabric, see [Top Trends in Data and Analytics for 2021: Data Fabric Is the Foundation.](#))

The data fabric concept is defined and guided by a number of characteristics:

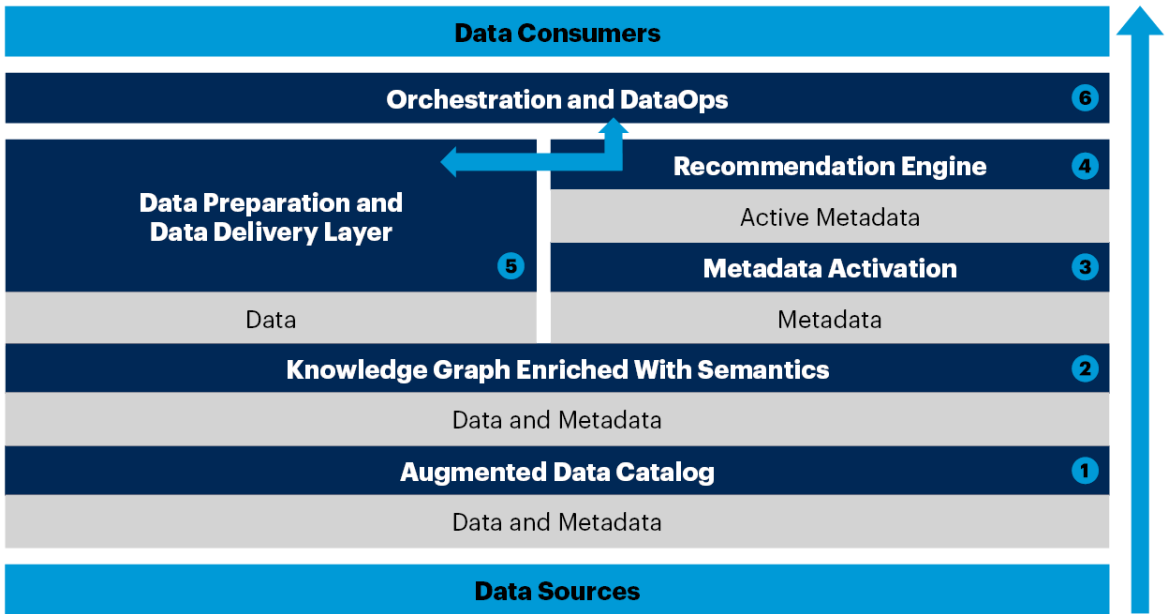
- **A data fabric is a design pattern.** Data developers will work with an organization's unique inventory of sources, targets, datasets and data platforms to create a number of semantics-driven structures that present distributed data to users, tools and applications in innovative ways. (For more data fabric design parameters, see [What Is Data Fabric Design?](#))

- **A data fabric reveals cross-platform relations among distributed data.** For example, virtual models illustrate — at a glance — your enterprise inventory of data, distributed business entities (e.g., the customer), and how data and business processes travel across a data ecosystem.
- **A data fabric is also for data access, not just data organization.** In an organization with a mature data fabric implementation, most data access is pushed through the fabric. This simplifies access to distributed data for better centralization, data standards, governance and security.
- **A mature data fabric has multiple layers.** Many of the layers enable data management functions for data access, integration, quality, cataloging, orchestration, interfaces, DataOps, active metadata and more (as illustrated in Figure 1). Furthermore, multiple virtual models may be layered in a fabric to provide use-case-specific views of distributed data.
- **A data fabric is not a vendor product.** Although some software vendors offer tools that support several data fabric functions, no single vendor addresses the complete data fabric concept in a single, unified toolset. To get the functions and automation you need for fabric maturity, you will need to acquire multiple tools from multiple vendors.

The Cool Vendors selected for this research offer representative examples of tools that enable and automate some of the many functions found in a data fabric. Vendors like these are indispensable for defining and implementing some of the leading-edge functions of a data fabric, and any of them could integrate into your multivendor solution for data fabric.

Figure 1. Key Pillars of a Comprehensive Data Fabric

Key Pillars of a Comprehensive Data Fabric



Source: Gartner
729003_C

Gartner

Nexla

San Mateo, California, U.S. (www.nexla.com)

Analysis by Ehtisham Zaidi

Why Cool: Nexla’s unified data operations platform aims to empower data teams with a tightly integrated and yet loosely coupled set of capabilities. This allows them to connect, transform, prepare, govern and share distributed data from multiple (internal and external) data sources to consumers through automation in data pipeline design and delivery.

Nexla is cool because of its continuous intelligence layer, which is built on data fabric principles and sits atop a universal connector architecture. This layer observes data at a record level, infers metadata from data (active metadata), then combines that with system metadata to generate a deep understanding of an organization's data ecosystem. The result is an automatically generated "Nexset" — which is a logical entity that encapsulates the understanding of data to make raw data usable. Each Nexset represents a logical collection of data that encapsulates schema, samples, error management, audit log, lineage and access control. This is done without making a copy of data, and data is accessed virtually using federated queries, file reads, API calls or stream reads, depending on the data system. By abstracting away the format and schema of the data, Nexsets present users with a consistent interface to connected datasets without need for significant data engineering efforts. As the number of connections increases, more Nexsets are formed that ultimately connect together to represent the metadata of the organization through a knowledge graph instance. The integrated data within Nexsets can be delivered using various popular data formats and structures. Integrated data can also be pushed into SaaS services via APIs and other data-as-a-service mechanisms.

Finally, Nexla can support DataOps design principles. Analysts can set up data flows through a no-code UI, while data scientists can use Nexla's low-code capability to extend the built-in transformation library with custom transforms written as user-defined functions (UDFs) in Python or JavaScript. Sophisticated users like data engineers are given an option to use Nexla APIs and command-line tools to perform heavy customizations, or to connect Nexla to their existing data solutions, such as catalogs, governance tools or continuous integration/continuous delivery (CI/CD) workflows. Moreover, Nexla's option for private installations in addition to its standard cloud SaaS service allows enterprise users to create a unified data fabric across their multicloud and hybrid data environments.

Challenges: Most challenges attributed to Nexla are similar to those that can also be attributed to any relatively small company aiming to disrupt a long-standing traditional data management design. Nexla's main go-to-market is centered around enabling a data fabric architecture by connecting — rather than collecting — data and using continuous intelligence on Nexsets (which are metadata-based knowledge graphs) to inform and automate data access, integration and delivery. Even though this is a potentially disruptive design pattern, this idea is usually met with stiff resistance from traditional IT teams who are more comfortable with established practices centered around physical data consolidation, modeling and delivery.

Data fabrics and DataOps are currently less mature in terms of adoption, established use cases and core capabilities. The relatively small product marketing team at Nexla will therefore have to provide a better connection of data fabrics with overall business outcomes of companies.

Nexla's technology requires bidirectional metadata access, sharing and registry with other data management technology solutions, so it will need to partner with more cloud service providers and other data management vendors (including data catalogs, data quality and data governance tools).

Who Should Care: Data and analytics leaders looking for a CSP-neutral and database-agnostic platform to enable data fabric design and collaborative ways for data teams to create, manage and deliver data pipelines through DataOps principles should investigate and adopt Nexla's Nexsets platform.

Soda

Brussels, Belgium (www.soda.io)

Analysis by Jason Medd

Why Cool: Soda is cool because it offers a highly intuitive environment where business and technical users can collaborate to tackle their data quality problems and address some of the main impediments of sustaining data quality programs. Data teams face the continuous challenge of providing high-quality data to data consumers for analytics and machine learning purposes. Once a data pipeline is up and running, it can fail due to changes to data schemas, new values being introduced or anomalies suddenly appearing in data values. These actions result in bad data being delivered to downstream processes. Data management teams are often caught off guard and only find out there is a problem when someone asks a question — such as “Why are the dashboards not working?” — which sets off a scramble to find and fix the problem. This occurs even in environments where data governance and data quality processes have been set up using best-of-breed metadata management and data quality tools.

Soda's cloud-based deployment model allows for easy spin-up and insertion into data processes with minimal overhead. Soda provides a number of out-of-the-box monitoring rules that can be instantly applied to datasets, accelerating users through their first iteration of the data quality process. The user interface is highly intuitive, allowing users to easily get started with monitoring their data. This design allows both business and technical users to monitor datasets, find anomalies, and collaborate to discuss any problems that are detected and initiate remediation steps. As a result, governance processes for data quality can form organically around specific data assets without first putting in place top-down or infrastructure-heavy governance programs to encourage and sustain user engagement.

Challenges: As more established data quality vendors move their data quality products to the cloud, their current data monitoring capabilities will be revisited, bringing more automation to improve user experiences. Soda will experience challenges differentiating its solution from other data quality solutions provided by established data quality vendors. Data quality monitoring is currently considered a subset of an overall data quality solution. Soda will need to demonstrate unique value to consumers that full-service data quality vendors cannot provide.

Soda recently secured Series A funding, but it is still a small vendor with fewer than 50 employees. Soda needs to invest in sales and product marketing, develop a partner ecosystem, and innovate its product offering to differentiate itself in the market.

Who Should Care: Data and analytics leaders should consider Soda if they have made investments in analytics and machine learning (ML) and want to be more proactive in managing data quality issues in their data integration pipelines. Data and analytics leaders should also consider Soda when they are struggling with ensuring high-quality data passes through their data pipelines, and are also experiencing challenges engaging business and technical users to work on data quality problems. This vendor's product could assist data engineering projects of any size where project teams must constantly react to changes in their data, including missing values, unexpected schema changes or new business conditions that impact data.

Tada Cognitive Solutions

Peoria, Illinois, U.S. (www.tadacognitive.com)

Analysis by Robert Thanaraj

Why Cool: Tada Cognitive Solutions (Tada) is cool because of its patented technology, Digital Duplicate, which can provide full visibility of data across an enterprise in the form of a semantic network. Tada enables business users to easily create the digital fabric of their enterprise's distributed data systems. The digital fabric gives a real-time view of various critical metrics across business processes, such as a supply chain digital network covering supplier networks, inbound freight, manufacturing, warehousing, outbound freight, retailers and customers. Once this digital fabric is established, business leaders can use Tada on their desktop or mobile devices, and be fully aware of any developing risks and mitigate them early.

Tada offers a three-tier data fabric model: connect data, clean data and consume data. It provides various API services, such as data integration services, domain services, model services, dictionary services, visualization services, business design services and advanced analytics services. Internally, Tada uses various open-source software to deliver its core API services for storage, search, messaging, stream processing and system monitoring. Its out-of-the-box application suite enables the creation of interactive applications with workflows, notifications and sharing content. Users can leverage the artificial intelligence or machine learning-based decision engines coupled with exploratory workspace and self-service data preparation for scenario analysis and simulations.

Challenges: Tada claims to be cloud-agnostic with its containerized solutions (IaaS). However, it is not yet available as a self-service offering (PaaS or SaaS) on any cloud service provider.

Tada has benefited from its long-standing partnership with CGN Global, a management consulting company focused on the supply chain business. However, Tada has limited presence beyond the manufacturing vertical and limited reach beyond North America. Finally, development and maintenance processes are mostly manual in Tada, whereas automation is one of the more desirable characteristics of a data fabric. Tada needs to create more and better automation to achieve data fabric standards.

Who Should Care: Data and analytics leaders seeking to manage external data ingestion and blending with internal data for operational intelligence should consider Tada. Tada can provide end-to-end visibility for the data and processes of an enterprise, and it can highlight critical issues.

Also, supply chain leaders should investigate Tada if they are seeking supplier collaboration for live visibility on supplier risks, material availability and demand management. Reference customers (large enterprises) have found Tada's supplier onboarding experience quick and effective at an affordable price point.

timbr.ai

Tel Aviv, Israel (www.timbr.ai)

Analysis by Eric Thoo

Why Cool: The graph data exploration technology of timbr.ai allows users to visualize the underlying data as a graph, to explore and discover relationships and dependencies in the data. The timbr-based SQL Knowledge Graph provides a graph-style structure for integrated data, based on ontology modeling and semantic reasoning in SQL over diverse data sources, including data lakes and data warehouses.

Capitalizing on demands for new paradigms for knowledge representation and machine learning support of data consumption, timbr.ai sets out to alleviate difficulties in integration and data management to bridge the gap between conventional SQL-fluent databases and relationship-rich smart knowledge graphs.

This enables querying data from abstracted views, using SQL code that is significantly less complex than the SQL required to query the same data via a relational model. The combination of knowledge graph and data virtualization capabilities equips data consumers to explore an extensive range of organizational data represented as a single connected graph.

The timbr SQL Knowledge Graph technology is integrated into a collaborative, AI-enabled platform. This gives researchers, analysts, data scientists and other data-focused roles in organizations the semantic access they need to better understand the data, discover hidden value and expose the data without extracting tables or views before running a query. Graph algorithms are applied along with customizable SQL Knowledge Graph for supporting data pipeline implementation. A library of commonly used graph algorithms aims at easing the process of selecting, configuring and deploying algorithms.

To accelerate time-to-value of data science initiatives and empower marketing and sales teams with access to granular-level insights of visitors' behavior, timbr.ai supports Google Analytics 360 SQL knowledge graph for data delivery to BigQuery.

Challenges: The use of graph enablement is still a new notion among data management practitioners. Increasing innovations of graph-style technology offerings and vendors will put pressure on timbr.ai to gain mind share and positioning in markets related to data management solutions.

While graph capabilities are gaining popularity, it is one among many moving parts of data management capabilities that businesses need as they pursue the multifaceted functionality of data fabric. Hence, timbr.ai's technology will need to interoperate seamlessly with the many tool types and diverse data sources that large organizations currently support spanning emergent, incumbent and legacy data environments.

Who Should Care: The timbr-based SQL Knowledge Graph should be of interest to any data and analytics leaders focused on data management. Architects and modelers seeking to optimize a data fabric design using knowledge graphs to result in significantly easier ways for amassing data-of-interest and accelerating diverse integration and sharing of data should also explore timbr.ai's offering.

Leaders of data-driven solutions should develop an understanding of how integrated data describing the business can be defined independently from disparate or siloed data sources. Understandable formal logic and common terms become key to describing the business so that both humans and machine learning algorithms can uncover the value that data can create.

Recommended by the Authors

Some documents may not be available as part of your current Gartner subscription.

[What Is Data Fabric Design?](#)

[How to Activate Metadata to Enable a Composable Data Fabric](#)

[Top Trends in Data and Analytics for 2021: Data Fabric Is the Foundation](#)

[Top Trends in Data and Analytics for 2021: Graph Relates Everything](#)

[Augmented Data Catalogs: Now an Enterprise Must-Have for Data and Analytics Leaders](#)

[How Augmented Data Management Capabilities Are Impacting MDM and Data Governance](#)

[Augmented Data Quality Represents a New Option for Upscaling Data Quality Capabilities](#)

© 2021 Gartner, Inc. and/or its affiliates. All rights reserved. Gartner is a registered trademark of Gartner, Inc. and its affiliates. This publication may not be reproduced or distributed in any form without Gartner's prior written permission. It consists of the opinions of Gartner's research organization, which should not be construed as statements of fact. While the information contained in this publication has been obtained from sources believed to be reliable, Gartner disclaims all warranties as to the accuracy, completeness or adequacy of such information. Although Gartner research may address legal and financial issues, Gartner does not provide legal or investment advice and its research should not be construed or used as such. Your access and use of this publication are governed by [Gartner's Usage Policy](#). Gartner prides itself on its reputation for independence and objectivity. Its research is produced independently by its research organization without input or influence from any third party. For further information, see "[Guiding Principles on Independence and Objectivity](#)."