Spotlight by Bloor

Author **Andy Hayler**

January 2024

Data Fabric Overview

Spotlight

Data Fabric is an approach to gaining a unified view of data in an enterprise without moving copies of the source data around. Instead a semantic layer is created that maps the core source data and enables it to be navigated and queried.

Executive summary

A data fabric is an architecture for data management involving a virtual layer of software that documents the location of data throughout the enterprise, and then serves up that data on demand in response to user queries. It differs from approaches such as data warehouses in that it leaves data in situ rather than gathering it together into a central hub. Instead, queries are executed in a distributed fashion, on demand.

Data fabric is a growing area and dozens of vendors claim to support it, with tools ranging from data catalogs to "knowledge graphs" to artificial intelligence. The industry appears to be in an early adopter phase at this point, with considerable excitement and marketing activity by vendors, and rather fewer documented case studies of non-anonymous companies with customer testimonials and quantified business benefits. As Bloor researches this area further it will investigate the current state of the data fabric industry, getting beneath the PowerPoint layer of the software and into the reality.

> **Data Fabric is an approach to gaining a unified view of data; it differs from master data management, data quality, data mesh and data warehousing but has relationships to these other areas.**
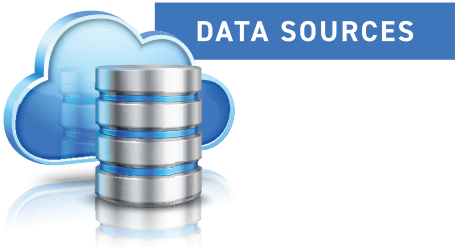
What is a data fabric?

Global organisations have long struggled to manage their data assets. The average company has hundreds of applications (different surveys show different figures, but they are all in the hundreds) scattered across on-premise data centres, public clouds like Azure or AWS, private clouds and beyond. Over the years many attempts have been made to wrangle the data herd. Enterprises have tried data warehouses (copy what you need into a central place for analytics), master data management (manage the key data that crosses those apps, such as customer, product and location) and enterprise resource planning (try and replace all those apps with a giant app from a single vendor). It is fair to say that, with the number of enterprise applications steadily increasing in number despite massive investment in these initiatives, none has proved entirely and universally successful. In 2023 just a third of executives trust their data, roughly the same as the level of trust as some years ago, according to several independent surveys.

A data fabric is a data architecture and infrastructure that attempts to organise and supervise the entire enterprise data resource, providing all systems and applications with timely access to all the data and services they may need. This spans both structured and unstructured data and may utilise existing metadata. Incidentally, it is different from a *"data mesh"*, which is a mostly organisational approach to data management, focusing on key data domains through decentralised data ownership.

BUSINESS USERS

FABRIC LAYERS

DATA ORCHESTRATION

SEMANTIC KNOWLEDGE GRAPH

ACTIVE METADATA

DATA CATALOG

DATA INTEGRATION

FABRIC LAYERS

DATA SOURCES

> **A data fabric is a data architecture and infrastructure that attempts to organise and supervise the entire enterprise data resource, providing all systems and applications with timely access to all the data and services they may need.**

Why should you care?

Assuming that this can all be made to work, then a data fabric promises a kind of nirvana: a way to visualise and inquire about data across the enterprise without the need to pre-integrate and copy swathes of data around into new data sources like data warehouses or data lakes. A central data catalog provides a kind of treasure map of data across the enterprise, hooked into a business glossary that allows users to navigate data using business terms like *"customer"* and *"product"* and *"asset"* without having to know about the physical locations of this data. The data fabric technology should be able to resolve inconsistencies and duplication of such data (the average company has six *"master"* definitions of *"customer"*, and nine of *"product"* according to one survey) in real-time.

It should be noted that data fabric does not necessarily replace all other existing data management structures. For example, a master data management hub would just be another source system in a data fabric architecture. Indeed, this might be a very useful one, since it is unclear how a data fabric virtual layer is going to dynamically resolve the inconsistencies between the various different customer or product hierarchies that exist in most companies. An MDM hub has *"survivorship rules"* that classify the reliability of the various sources, and applies assorted data quality rules and merge/matching algorithms to resolve data inconsistencies. All this considerable amount of work could be utilised by a data fabric virtual layer that plugged into an existing MDM hub. If no such hub exists then the data fabric technology is going to have to deal with the same issues itself.

> **❝** ...a data fabric promises a kind of nirvana: a way to visualise and inquire about data across the enterprise without the need to pre-integrate and copy swathes of data around into new data sources like data warehouses or data lakes. **❞**

What is the benefit?

If you have managed to deploy a working data fabric solution then you can potentially see many benefits. A commercial enterprise could gain a unified view of its customers, for example, across all the various enterprise touchpoints that exist. Hospitals could gain a single, coherent view of patients, encompassing patient records, treatment history and more. Financial institutions would get a better assessment of trading risk, achieving a better knowledge of who the true counterparties are in each trade.

This would allow all kinds of useful business analytics that have proven elusive to deliver in the past, such as knowing who your most profitable products and customers are, a simple notion that is tough to actually deliver. In principle, an all-encompassing data fabric would grow with your business, adapting to new or changed data sources and business models. If it works as promised then data quality, another thorny data problem that has remained stubbornly hard to eradicate, should also improve.

There are significant challenges in making such an architecture work in practice, especially at enterprise scale. Executing complex queries in real-time against a large single database takes considerable resources. To execute such queries against multiple distributed databases, with data scattered amongst them, is a much tougher problem, requiring a highly efficient distributed query capability. Few currently deployed mainstream databases can seriously claim this ability, though some specialist niche databases are designed with distributed query in mind. Even if that problem is ignored, there is the issue of data quality, which is something that continues to plague operational systems despite decades of attempts to address it with assorted data quality software and data governance processes. The level of trust of executives in their data remains stubbornly low even in 2023 surveys, so just being able to execute a query is one thing; being confident that the answer is correct is another. Previous approaches such as data warehouses and master data management put considerable emphasis on data quality (around a third of a typical master data management project is spent resolving data quality issues). A data fabric spurns the movement of data into such hubs, preferring to resolve queries dynamically, but it is unclear how the very real issue with underlying data quality can be wished away. Certainly, any data fabric initiative needs to be able to explain how it will successfully deal with such challenges. If you spend some time, as I have, searching for data fabric case studies, then you will find plenty of vendor claims but very few that have named end user companies with customer testimonials and quantified business benefits. Although data fabric is not a new idea, it would appear to still be in the early adopter phase based on the dearth of properly documented case studies. There are plenty that are anonymous, mostly promoted by data fabric vendors and consultants, who clearly have a vested interest in making such claims.

> **The level of trust of executives in their data remains stubbornly low even in 2023 surveys, so just being able to execute a query is one thing; being confident that the answer is correct is another.**

The solution(s)

There are several underlying components to a data fabric. A virtual software layer is established that maps data assets across the source applications and attempts to transform and process that data as needed to satisfy business needs, all while leaving the source data in place. At its heart is a data catalog with a business glossary, some form of a recommendation engine (perhaps with an AI component) and a way to visualise the data landscape (such as a *"knowledge graph"*, a visual set of linked descriptions of data entities), a data preparation layer to retrieve data as needed, and some data orchestration component to manage the end-to-end workflow of data, including on-premise and cloud.

Dozens of vendors claim to have data fabric solutions today. Some of these are giant vendors with a long history of data management solutions, and some are newer companies that have assorted capabilities, from data catalogues to artificial intelligence to *"active metadata"*. Few have end-to-end solutions covering all components, though some claim to.

> **Dozens of vendors claim to have data fabric solutions today... Few have end-to-end solutions covering all components, though some claim to.**

Summary

Data fabric is an emerging architecture that is generating considerable marketing activity by software vendors, keen to help customers realise the promised gains of a more flexible data architecture that can work without building yet more data silos and instead dynamically access data in situ in operational source systems. There are considerable practical challenges facing such an approach. The underlying issues that caused previous approaches such as data warehouses, data lakes and master data management to struggle have not gone away, and will need to be resolved if data fabric is to be cut from a truly different cloth from these previous approaches.

About the author

ANDY HAYLER
Practice Leader: Data as an Asset

Andy is an established software industry authority, an independent strategy consultant advising corporations, venture capital firms and software companies. He is the founder of Kalido, which under his leadership was the fastest growing business intelligence vendor in the world in 2001. Kalido was recognised as an innovator in data warehousing, and then launched arguably the first true master data management product, a market which at the time did not exist but is now a well recognised and fast growing industry. Andy was the only European named in Red Herring's "Top 10 Innovators of 2002". He was a pioneer in blogging with his award winning "Andy On Enterprise Software" blog.

Andy started his career with Esso, working in a number of technology roles before moving to Shell. He was Technology Planning Manager of Shell UK, then Principal Technology Consultant for Shell International. He later established a global information management consultancy, which under his leadership grew to 300 staff.

Bloor overview

Technology is enabling rapid business evolution. The opportunities are immense but if you do not adapt then you will not survive. So in the age of Mutable business Evolution is Essential to your success.

We'll show you the future and help you deliver it.

Bloor brings fresh technological thinking to help you navigate complex business situations, converting challenges into new opportunities for real growth, profitability and impact.

We provide actionable strategic insight through our innovative independent technology research, advisory and consulting services. We assist companies throughout their transformation journeys to stay relevant, bringing fresh thinking to complex business situations and turning challenges into new opportunities for real growth and profitability.

For over 25 years, Bloor has assisted companies to intelligently evolve: by embracing technology to adjust their strategies and achieve the best possible outcomes. At Bloor, we will help you challenge assumptions to consistently improve and succeed.

Bloor