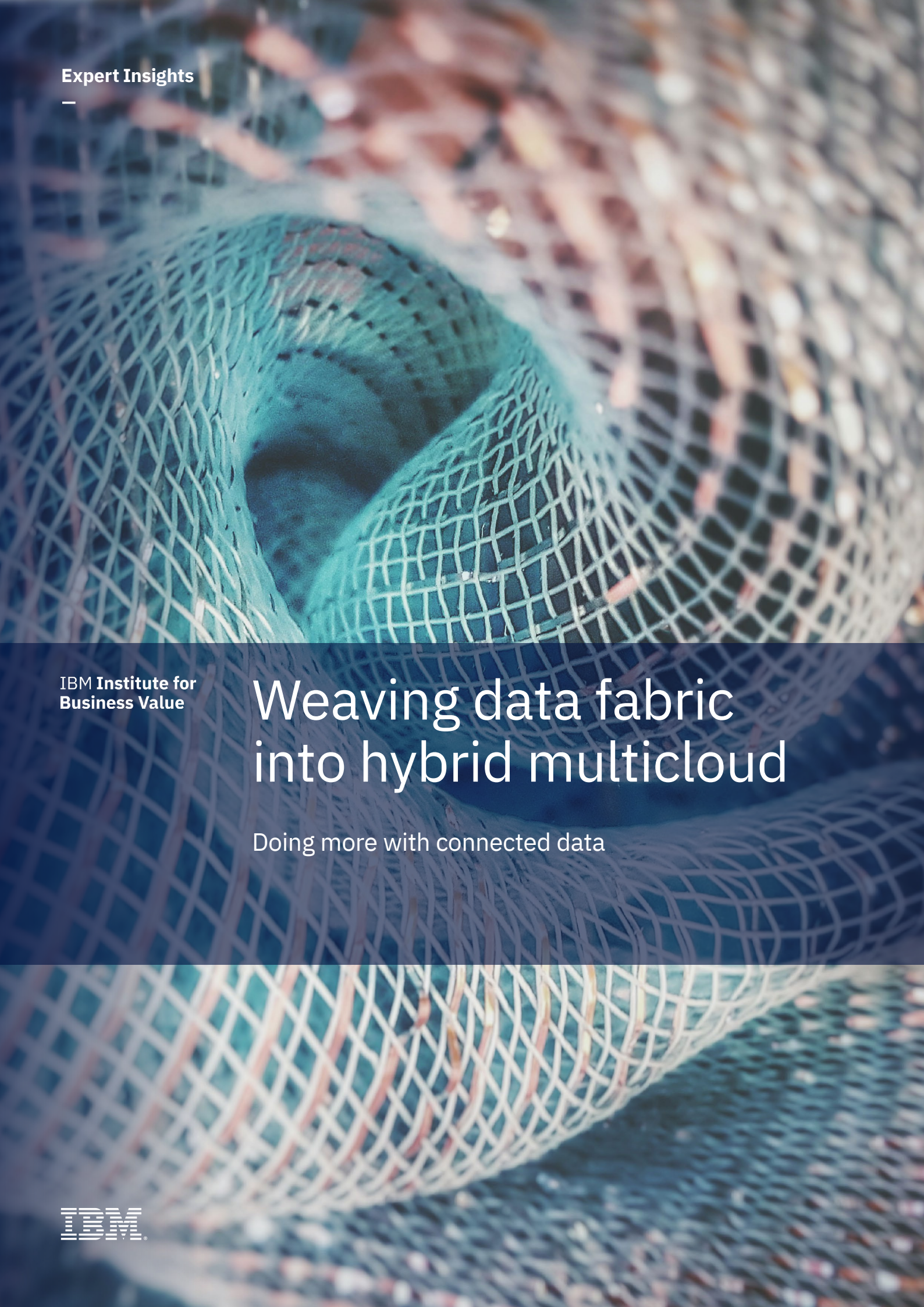


Expert Insights

IBM **Institute for Business Value**

Weaving data fabric into hybrid multicloud

Doing more with connected data

IBM

Experts on this topic

Dr. Sandipan Sarkar

IBM Distinguished Engineer,
Global CTO Data, Hybrid Cloud
Transformation Service Line,
IBM Global Business Services
sandipan.sarkar@in.ibm.com
linkedin.com/in/sandipansarkar

Sandipan helps large and global enterprises become data-driven organizations on hybrid multicloud. He is a globally respected expert in exponential technologies, such as artificial intelligence, Internet of Things, and Blockchain. Sandipan holds a Ph.D. in artificial intelligence. He publishes regularly in scientific journals and business magazines and lectures at industry forums and academic institutes.

Varun Bijlani

Global Managing Partner, Hybrid
Cloud Transformation Service,
IBM Global Business Services
varun.bijlani@uk.ibm.com
linkedin.com/in/varunbijlani

Varun leads IBM's global hybrid cloud transformation services business, helping clients design their cloud strategy and architecture, then execute it via migration, modernization and new cloud native capabilities.

He has over 26 years of experience combining domain knowledge and global program management expertise with strategic and operational leadership within both consulting and industry. He holds a PGD in Business Leadership from UWE, UK and Bachelor of Engineering (BE) in Electronics from Mumbai University, India.

Richard Warrick

Global Research Lead,
Cloud Computing
IBM Institute for Business Value
rwarric@us.ibm.com
linkedin.com/in/richardwarrick

Richard leads the IBM institute for Business Value's (IBV) cloud research agenda that explores cloud as part of a new way of working as enterprises evolve to become more fully digital businesses. Richard's consulting career spans over twenty years of helping Global 1000 clients navigate the intersection of business and technology. His work has focused on business + IT strategy, operating model design and implementation, and large-scale transformation programs. He has a Masters of Science degree in Organizational Learning from George Mason University.

Contributors

Subhendu Dey
Executive Architect/Associate
Partner, Cloud advisory for
Data & AI, IBM

Sujay Nandi
Executive IT Architect, Connected
Solutions & IoT, Cloud Advise, IBM

Key takeaways

Clear benefits

While data fabric technology may seem complicated, its benefits are easy to understand: the right data available at the right time to the right application, no matter where that application runs in a hybrid multicloud environment.

Boundaries matter

Data fabric manages three critical boundaries that are important to running intelligent data on distributed, hybrid cloud environments: boundaries shared by data platforms, by clouds and cloud providers, and by an enterprise's transactional and analytical data operations and communities.

Bringing all data capabilities to bear

Enterprises investing in artificial intelligence (AI), machine learning, Internet of Things (IoT), and edge computing can use data fabric to weave together the transactional and analytical sides of their data capabilities.

Clearing the path to value

Getting business value from a data fabric implementation depends a great deal on how well the enterprise has cleared nontechnical obstacles it will face.

—

Data fabric is important—and not just to data engineers

Everyone involved in technology-driven business transformation should understand the concept of data fabric. As large enterprises continue to evolve in response to constant competitive pressures and unforeseen events such as the current pandemic, data fabric will help many of them "pass their genes" along to their organizations' future iterations. Today, enterprise investments in expanded data and data analytics capabilities continue to be strong, as are investments in hybrid multicloud architectures. As software "eats the world," it is doing so on a high-protein diet of *intelligent data* and *hybrid cloud*.

IBM Institute for Business Value research shows that getting more value from data drives sound digital strategies. Outperformers are making enterprise-scale investments in data centers of excellence (CoEs), data scientists, and data analytics tools.[1] Most large enterprises today have strategies for infusing customer-facing applications—digital products—with information that improves the customer's experience, supports customer journeys, and makes new services possible. These strategies work in B2B as well as B2C business models and are central to emerging platform business models as well.

Technology plus scale plus people equals *complexity*. We tend to manage complexity by drawing boundaries around things, allowing us to focus on one part of a complex system at a time. This works up to a point, but we often ignore the crucial business of managing the white space between the *boundaries* we've drawn. Data fabric is important to understand because it provides new ways to manage the boundaries separating applications, data, clouds, and the people who design and create them. It's the management of those boundaries that determines success.

Definitions: Data fabric

Data fabric is a conceptual approach, not just one specific technology. It can be thought of as an environment that includes an architecture and set of unified data services. Together, these services support consistent data capabilities across an organization's own network—on premise—and on multiple cloud environments.

The term "fabric" can be both misleading and instructive: while data fabric is not a thing, per se, its component parts can be thought of as resembling the woven nature of a fabric, both connecting and holding together the whole.

This paper explores three of those boundaries. First are the boundaries between data platforms. Second are the boundaries between clouds and cloud providers. Third are the boundaries between the enterprise's *transactional* and *analytical* data operations and communities. This third type of boundary may be less familiar to practitioners than the first two, but it is at least as critical, if not more so.

Data fabric—more than just another complicated technology

Like many elements of digital business transformation, data fabric technology and architecture are important but quite complicated. Google "data fabric architecture," and you'll get the idea (see "Definitions: Data fabric"). This complicatedness—and the speed of change driven by hybrid cloud adoption, AI, edge computing, and IoT—can combine to make data fabric feel beyond the grasp of people who are not data scientists, data engineers, AI modelers, or the like. Which means most of us may find it baffling.

Fear not. For this paper, it's enough to understand that data fabric is a new solution to a perennial problem. Although data is valuable and large enterprises have tons of it, it is imperfectly managed. Data is spread all over enterprises in containers large and small: data centers, data warehouses, and data lakes. It's stored and retrieved in different formats using different methods. It moves from place to place, sometimes slowly and with difficulty. It's hard to manage, expensive to store, not always trustworthy, and accessible only through intermediaries under restrictive conditions.

Now, add cloud computing. Cloud computing takes this mess, makes data less expensive to store by moving it to the cloud, and thus creates new places where data can proliferate in more silos, endpoints, and the "walled gardens" created by diverse cloud service providers. Cloud encourages innovation in the form of new on-the-cloud applications, all of which can generate more data silos.

If data is the new oil, most large enterprises are oil wells having a blowout. Data is gushing all over the place, and none of it is worth much in its current crude condition. Data fabric wants to solve this problem.

If data is the new oil, most large enterprises are oil wells having a blowout. Data fabric wants to solve this problem.

Data fabric: move beyond the database

Data fabric's primary concept is to stop thinking of databases, data lakes, data warehouses, or data marts as fixed stores of data. Instead, view data as flowing through a broad network "on tap" where the right data is available to the right location and the right application at the right time. From the user's point of view, it just works.

Moving beyond the database is a way to deal with "data gravity"—as more data winds up in one place, the harder it is to move. Data fabric can make that data easier to move, or it can make data that can't be moved—data that has to stay on premise, for instance—easier to access.

Data fabric and Indra's Net

Millenia-old Eastern teachings provide an apt metaphor for data fabric in the form of *Indra's Net*.[2] In these teachings, Indra is the king of the gods, and his net is a vast, cosmic lattice with a precious jewel everywhere the threads of the lattice cross. The net strings together an infinite number of jewels, and each facet of every jewel reflects each facet of every other jewel. Everything is connected to everything else, with causes and effects rippling across the infinite net (see Figure 1, "Indra's Net: A metaphor for data fabric").

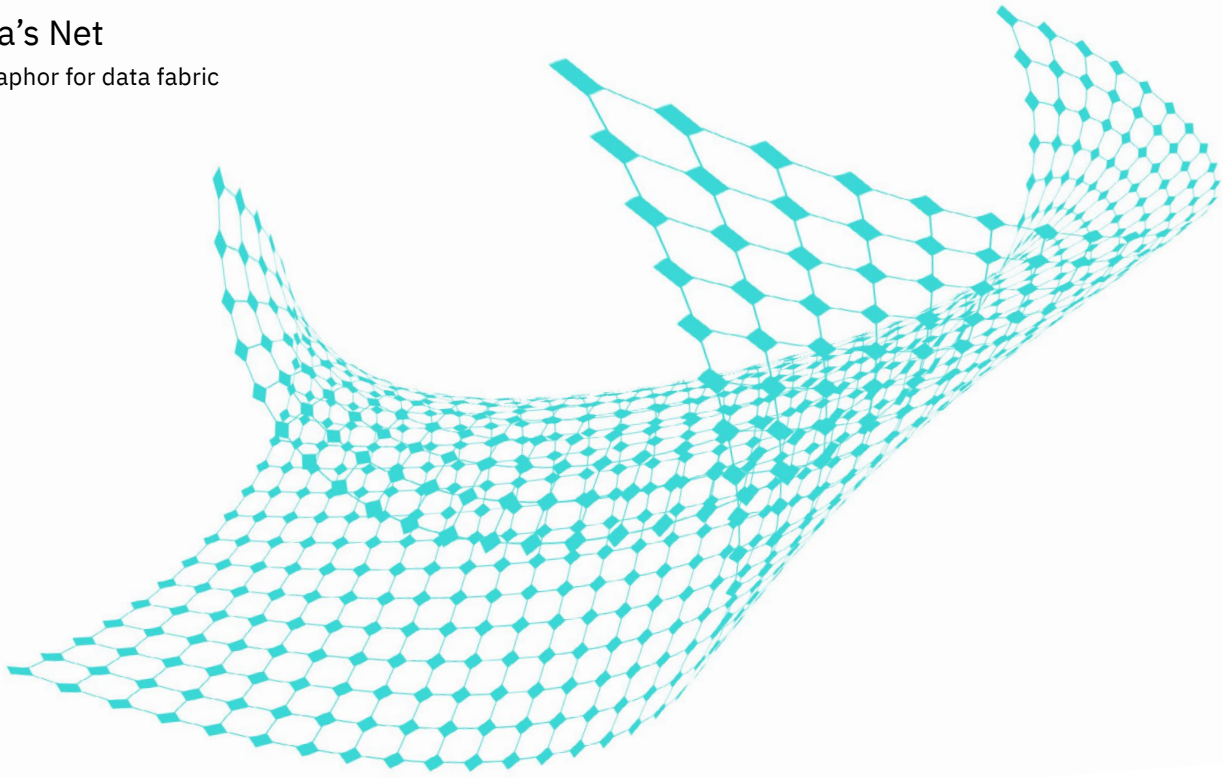
In our metaphor, the jewels in Indra's Net represent each of an enterprise's thousands of applications, as well as applications run by customers, business partners, industry ecosystem participants, and others. Each of those jewels is connected to every other jewel. While each jewel is distinct, it is also connected by the net: the *fabric*. The fabric provides everything each jewel needs: compute power, storage, connectivity, and data. Wherever each jewel appears in the net, it receives the data it needs from whatever source of *data* is best and closest to the jewel.

With that introduction to data fabric and the problems it solves, let's explore three of the boundaries that enterprises encounter across their stores of data and how an approach akin to Indra's Net can help.

Figure 1

Indra's Net

A metaphor for data fabric



The data needed to build an end-to-end view of the customer is unlikely to be located in any single data platform.

Boundaries between data platforms

A saying common among data practitioners is, “once you add a second database, you have an integration problem.” In the years since big data was the king of IT buzzwords, the opportunities to profit from better use of data have compounded, but at a rate fully matched by the challenges of making the right data available to the right applications at *scale*. Even before the cloud, enterprises were building *data platforms*: technology solutions that integrate data located in diverse databases. Data platforms are designed to act as a service. Within guardrails, people who need data can get access to it or have it delivered to users, applications, or other technologies.

The boundary around each data platform is usually defined by the type of data it stores, or by the way that the data is used. Large enterprises might need an HR data platform, or a supply chain data platform, or a customer data platform for a specific business unit.

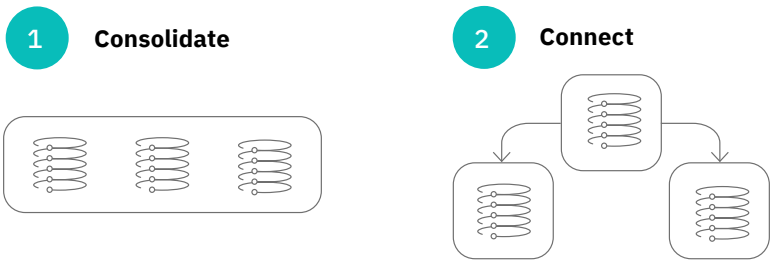
Those boundaries made sense and were an expedient way to get more value from the data available. But today, several changes have made a difference. First, as enterprises deploy new business models and build single views of lifetime customer interactions, business unit silos of customer data are less defensible. Second, as supply chains become increasingly digitized, the data needed to build and operate an end-to-end view of the customer value chain is unlikely to be located in any single data platform. Third, better data analytics mean that there may be connections and insights available across data platforms that would not have occurred to their designers.

There are three approaches to manage the boundaries between data platforms: consolidate, connect, and control (see Figure 2, “Boundary strategies: Three approaches for managing data platform boundaries”). Any of these methods could be part of a data fabric, but the one most suited to a distributed, hybrid cloud environment is the third option. For every method, however, the critical thing is to loosely couple the data in the platforms with the applications that need it. This concept of loose coupling is important throughout data fabric architectures (see “Value propositions for cross-selling customers”).

Figure 2

Boundary strategies

Three approaches for managing data platform boundaries



The data platforms are moved to same infrastructure.

One data platform gets connected to another in an upstream-downstream combination.

Have unified governance and management plane for multiple data platforms, which can be distributed and can evolve at their own pace and strength.

Boundaries between clouds and cloud providers

Large enterprises have adopted hybrid cloud architectures that typically employ up to a dozen clouds and a variety of different cloud service providers. This heterogeneity can have advantages, but it does not necessarily make it easier to share data. The separate clouds and cloud providers can be another form of vertical silos that we saw with data platforms above.

Data fabric plays a key role in managing these boundaries, as do some architectural design decisions. In both cases, the key is to apply a horizontal management layer on top of the clouds that need to be woven together (see "Transcending the data ingress and egress problem"). In a hybrid multicloud environment, managing these boundaries with data fabric means that one cloud provider needs to become "the first among equals," so that provider must employ an open architecture. Applying a data fabric may also reveal that some of the clouds under management may not have been as open as advertised.

Boundaries between transactional and analytical data operations communities

First, some basics. Large enterprises generate a lot of *transactional data*: data that gets created or exchanged during interactions with a customer. The challenges with using and managing transactional data effectively are well understood today, even if they have not been completely solved. For instance, most enterprises struggle to get a single view of a customer because they use and exchange transactional data from many different channels, applications, and products and services.

If not actively managed, transactional data tends to wind up in silos across the enterprise, preventing the kind of data sharing that is important to better business performance. How important? Jeff Bezos of Amazon famously threatened to fire anyone who kept transactional data to themselves.[3]

A second type of data—*analytical data*—is different. This type of data comprises very large and diverse data sets used to reveal hidden patterns and to make predictions and decisions. The big data wave was about analytical data, and today analytical data is the foundation of so-called exponential technologies, such as AI, machine learning, IoT, edge computing, and 5G.

Insight: Value propositions for cross-selling customers

The "control" pattern for negotiating boundaries between data platforms is common among organizations that have grown data platforms around distinct customer offerings. A large bank, for instance, might have a data platform for its mortgage products as well as a data platform for credit card customers. The business case for cross-selling mortgages to credit card customers and vice-versa is good—it's a classic case of monetizing the data a large enterprise already has on hand.

In this example, there's no real need for consolidating the platforms—and the lines of business would not agree to such a thing. Merely connecting the platforms is not very scalable over the long term. The control pattern keeps each platform in the hands of the respective business units but also enables the larger enterprise to play its role in governing and funding how the integrated platforms evolve.

Insight: Transcending the data ingress and egress problem

Boundaries among clouds show up in a very tangible way: higher cloud service provider fees. Cloud service providers base their fees on the premise that once a client's data is in their cloud, it stays there. Adding data to a cloud can increase fees, but taking data out of the provider's cloud and moving it to another cloud can also be very expensive.

Using the previous example of the cross-selling bank, let's imagine that the bank has hosted its mortgage applications with cloud service provider A and its credit card applications with cloud service provider B. In daily operations, moving data directly from one cloud to the other in anything like real time would be difficult and expensive. With a data fabric, the data would go from cloud A to a data platform where it would become a data service accessible not just to cloud B, but to any requesting application on any cloud.

Data fabric can use APIs to deliver insights-as-a-service directly implanted within transactional applications or business workflows.

Two characteristics of analytical data are important to data fabric. First, many uses of analytical data require it to be ingested in large volumes and at high speeds. Second, it may include *unstructured* data: data that can't be fit into a data cell. Video, images, audio, text, social media, sensor data—it's all data, but not the kind that could be captured in a spreadsheet.

Modern analytics need to run on *real-time* or *close-to-real-time* data. Analytics can't rely on conventional data warehouses or data lakes to provide this kind of speed because they require batch processing and expensive movement of data to make the analytics happen. By the time any insights are generated, they may already be outdated.

Data fabric can solve this challenge by providing event streaming, application integration, and real-time analytics that run even when data is on the move. In addition, data fabric offers data virtualization capabilities that allow analytics to be built over distributed data sources. This is especially important as sources of data become distributed across multiple clouds, where moving data in and out of clouds and cloud providers becomes very expensive.

Data fabric can employ APIs to deliver insights-as-a-service directly implanted within transactional applications or business workflows. This means that users of the data do not need to wait on expensive manual interpretations of data dashboards to provide actionable insights. In some cases, those insights can trigger automated actions directly within the system without requiring human intervention.

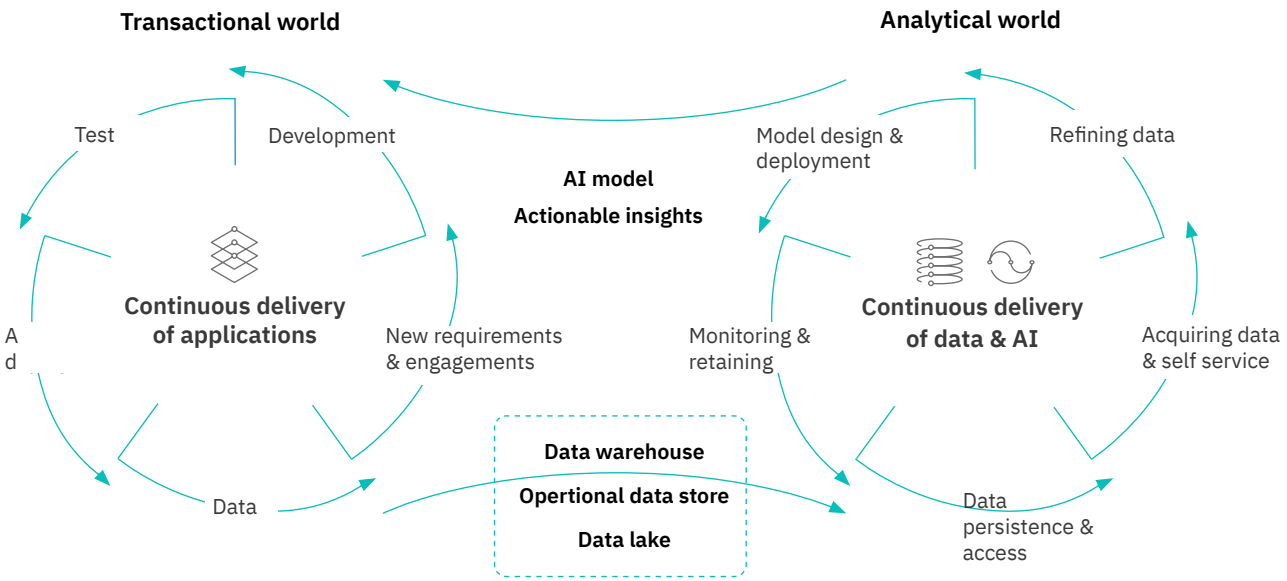
Managing community boundaries creates operating model opportunities

Bringing transactional and analytical worlds together can have profound effects on the operating model of the enterprise.[4] By bringing transactional and analytical worlds closer technologically, data fabric opens up new opportunities to bring application development teams and data and analytics teams closer. The cycle of digital product development and management can now mesh with the cycle of generating data insights and applying models for AI, creating new, value-adding workflows. The teams in both communities can work with the common goal of delivering data-rich value to application users and customers (see Figure 3, "A winning coalition: Managing the boundaries between transactional and analytical data communities").

Figure 3

A winning coalition

Managing the boundaries between transactional and analytical data communities



Fostering customer-provider relationships between *application delivery* and *data analytics* communities

This is where the magic happens—data fabric connects data from across a hybrid multicloud IT environment and makes it consumable by digital products. It also makes the data generated by a digital product consumable by the data CoE. Since data fabric does not require centralized architecture, and because it works well with distributed heterogenous hybrid-multicloud landscapes, it makes innovation possible that otherwise would have been blocked. When data fabric breaks down the technological and organizational barriers typical in large enterprises, good things can happen.

The same advantages can enable a unified governance of data and analytics across all platforms, thereby breaking down more kinds of silos (see "Developing provider and customer relationships"). Unified governance allows business units to move forward with individual platforms, while helping ensure that diverse analytics initiatives can be tied together through a common data governance process. For instance, with data fabric, an enterprise-level data catalogue and model registry can make trusted data and analytics assets in the enterprise visible and available.

Data fabric may be a tough technology to understand, but its usefulness is clear: it can connect disparate data throughout an enterprise and multiple cloud environments the enterprise uses. Perhaps most importantly, it can connect the enterprise's application delivery teams and its data analytics CoEs. We encourage your organization to look very hard at the area where data fabric, hybrid cloud, application development, and data analytics come together. It's an area full of potential business value.

Insight: Developing provider and customer relationships

Using our banking example for a third time, the bank's digital mortgage products and digital credit card products both generate transactional data. The business case for cross-selling between customer sets is based on the idea that some—but not all—of the data generated by mortgage applications will be useful to the credit card business, and vice-versa. The product owners for both types of applications need some analysis of the other's data and perhaps even some higher-order capabilities, such as predictive analytics or workflow automation. For that, they need the data analytics community.

Conversely, the data analytics community needs a steady, high-volume stream of data upon which it can test and build advanced data models and train machine learning and AI products. From this perspective, the two communities are natural providers and customers to each other. But in practice, it's common for them to have little interaction, and the performance of both communities suffers as a result.

Breaking down the barriers between the transactional community and the analytical community is an operating model problem. It requires a shift in our thinking: from "running transactional applications" and "doing data science" as distinct domains, to seeing them instead as contributors to the same horizontal flows of work required to deliver value to customers.

Action guide

Clearing the path to capture business value

Discussions about cloud adoption, data fabric, and intelligent data are increasingly interrelated, as they should be. What's often missing is a similarly integrated conversation about the changes in *management practices* and *mindsets* that make the difference between just implementing the data fabric architecture technology and implementing it in a way that captures business value.

Here are three ways to move beyond implementation to value.

1. Fund data fabric initiatives as products, not projects

Business technology initiatives tend to be funded as projects that stand up, deliver against preset requirements, and then stand down, with ownership shifting to the part of IT that does maintenance and sustainment work. For something like data fabric, this approach won't work.

It's far better to fund and manage data fabric initiatives as products, where funding mirrors a product management lifecycle. Instead of implementing the technology and declaring victory before any business outcomes have been realized, the product approach recognizes that fielding a working version of data fabric is just the beginning.[5] Funding and staffing will need to be persistent over the course of years.

This approach can be difficult because pre-digital, pre-cloud funding processes and policies are still in place at most large enterprises. They are often very resilient and well-defended. Rather than seeking direct modernization of legacy funding schemes, some organizations do better by requesting one-off waivers of the current policies.

2. Employ agile implementation principles

There is no data-fabric-by-numbers kit. Each enterprise will need to learn what it takes to make data fabric work for it. This is especially true in cases where work on data fabric is running at the same time as cloud adoption work or work on scaled AI implementation. Changes to one part of the system will create changes to other parts, often in ways that can't be anticipated.

Employing agile principles, such as build a little, test a little, learn a little, helps to cope with the complexity of execution that is inevitable when building a data fabric. The best outcomes use small, cross-disciplinary teams that stay together to build, integrate, and operate successively more complete and scaled versions of data fabric. In cases where cloud and AI initiatives are running at the same time, it's best to support the collective teams with a cross-initiative product office that helps integrate the teams' efforts.

Employing agile implementation principles can be difficult because many large enterprises are accustomed to top-down, command and control management principles. Developing fixed plans, schedules, and milestones at the beginning of the initiative when everyone knows the least about how to make data fabric work is a bad idea. Even with the right talent, the right technology, and useful artifacts such as reference architectures, the fact remains that this initiative has never been done before—not with these people in this organization in this business environment.

3. Run data fabric as a service

Ideally, data fabric runs in the background without its users—development teams, for instance—knowing *how* it is doing *what* it is doing. So development teams should not have to employ full-time data fabric experts. Data fabric is best delivered as a service that users consume in the process of responding to customer requests, sometimes without knowing that they are consuming it. In a service-driven business technology model, data fabric fits perfectly as an ongoing, foundational set of services available to a variety of "customers" elsewhere in the operating model.

One advantage of making data fabric a service consumed by customers is that it automatically keeps the scope and scale of data fabric on target. Data fabric services expand as demand for them expands; data fabric services shrink—or change—as customer-driven demand shrinks or changes.

Running data fabric as a service can be difficult because conventional IT organizations are structured around vertical functional silos, and a service-driven operating model ultimately requires changing to a more horizontal structure that aligns with customer-centered flows of digital product delivery.

About Expert Insights

Expert Insights represent the opinions of thought leaders on newsworthy business and related technology topics. They are based upon conversations with leading subject matter experts from around the globe. For more information, contact the IBM Institute for Business Value at iibv@us.ibm.com.

Notes and sources

1 "From Data Science to Data Diplomacy." IBM, 2019, www.ibm.com/thought-leadership/institute-business-value/c-suite-study/cio?mhsrc=ibmsearch_a&mhq=data+diplomacy. https://www.ibm.com/thought-leadership/institute-business-value/c-suite-study/cio?mhsrc=ibmsearch_a&mhq=data%20diplomacy; Giordano, Tony. "A Blueprint for Data in a Multicloud World." IBM, 2019, www.ibm.com/thought-leadership/institute-business-value/report/multicloud-data-strategy

2 O'Brien, Barbara. "Indra's Net: Buddhist Metaphor for the Interpenetration of All Things." Learn Religions, www.learnreligions.com/indras-jewel-net-449827

3 Settle, Mark. "CIOs Can Weaponize Their API Architectures Through Better Governance." Forbes, Forbes Magazine, 1 June 2020, www.forbes.com/sites/marksettle/2020/05/04/cios-weaponize-apis-through-better-governance/?sh=27972aee6076

4 "How to Move Beyond a Monolithic Data Lake to a Distributed Data Mesh." Martinfowler.com, martinfowler.com/articles/data-monolith-to-mesh.html; "Data Mesh Principles and Logical Architecture." Martinfowler.com, martinfowler.com/articles/data-mesh-principles.html

5 Kersten, Mik. "Project to Product: How Value Stream Networks Will Transform It and Business." IT Revolution PR, 2018. https://itrevolution.com/book/project-to-product

IBM Corporation
New Orchard Road
Armonk, NY 10504
Produced in the United States of America
April 2021