Why Smart Businesses View a Data Fabric as an Inevitable Approach to Becoming Data-driven

Judith Hurwitz
President & CEO

Daniel Kirsch
Managing Director

HURWITZ & ASSOCIATES
Insight to Action

Sponsored by IBM

Introduction

Business executives know that their company needs to adopt a data-driven strategy. However, gaining value from data remains a massive challenge, despite the fact that businesses are drowning in data. As businesses begin to adopt a distributed hybrid computing strategy, they must figure out how to achieve a unified view of all of their data. Simply put, you cannot achieve a hybrid computing strategy if you can't abstract data from its underlying location. To be successful, business leaders need to have the ability to gain an understanding of the holistic view of data across any location, any cloud, and any application. The business requirement to analyze and manage your data -- no matter where it resides -- has given rise to an emerging approach and architecture, called a data fabric.

For decades management has demanded the capability to unify data across silos but they have been without success. For example, IT organizations created data warehouses or data lakes to serve as a centralized repository for data, but these attempts required moving large amounts of data, and a supporting organization to continually manage, update and secure the environment. The situation has only gotten more complicated as the amount of data and the number of data sources continues to expand..

The key to business success in an environment of continual change is the ability to turn your data into a strategic and manageable asset. Unfortunately, understanding your data in context and across teams and environments remains a major challenge. In this paper, we will discuss why businesses are starting to adopt a data fabric approach that allows you to connect to your data no matter where it resides. This defining characteristic of a data fabric will allow businesses to more effectively use their information as a strategic business asset. A data fabric is the only practical approach to creating the next generation of data management by delivering a manageable, predictable, and resilient platform with the help of artificial intelligence and machine learning. Automation and built-in AI will allow the data fabric to be flexible enough to support your evolving data landscape without the need for constant stewardship. In addition, we will provide an update on IBM's journey to establishing an intelligent data fabric in the next generation of IBM Cloud Pak for Data.

Data can allow you to anticipate client demand and shifts in product sentiment, operate more efficiently, and spot outstanding employee talent. But how can you unlock those insights?

Data is everywhere in a hybrid and multicloud world

Data can allow you to anticipate client demand and shifts in product sentiment, operate more efficiently, and spot outstanding employee talent. But how can you unlock those insights? It would be convenient if all of your corporate data sat in a single database, was of a single type, or was at least behind your firewall. Instead, nearly every business utilizes multiple database technologies across not only multiple physical sites, but also a variety of cloud services, and now extend out to IoT devices. There are typically hundreds of applications that generate massive amounts of data across the business. With data spread across so many platforms, how can leaders hope to gain insights? Below are four

HURWITZ & ASSOCIATES
Insight to Action

common areas of concern that come up as companies push to become more data driven:

- **Trust** – How can executives confidently make decisions based on insights from AI if they don't trust that their models are utilizing the most complete and accurate information? Are the usage statistics that a model relies on current or was that data backfilled with last month's information? How confident are you in the accuracy of your model? Do you have processes for updating models once they're in production to ensure they remain accurate and don't suffer from drift as business circumstances change?

- **Security, privacy, governance and compliance** – Is your data secure throughout the entire analytics process? Is the sanctity of private customer, employee, or proprietary data exposed as you try to gain insights? How do you ensure security is maintained as data is moved? And how can you keep customers confident that you treat their private data in an appropriate way? Are you confident that your data meets international, federal, state, and industry compliance and governance standards? If you are moving data to a new location, does that new location for analysis have the same compliance standards?

- **Costs & ROI** – To embrace a culture of data and analytics, you need to make data investments. How can you make sure that those investments will pay off? Although the cloud has made massive amount of compute power and storage accessible to all businesses, it doesn't come cheap. If, for example you are transferring all of your data to a centralized location for analysis, costs for data transfers and storage can quickly balloon.

- **Existing Skills & Analytics Efficiency** – Do you have the internal capabilities to become a data-driven company? Your data scientists and other experts might spend nearly all of their time on data prep and blending rather than delivering insights that have a business impact. How can you help data teams become more efficient so that they can spend time on driving insights? Can routine tasks be automated?

These challenges, along with the need to rapidly transform your business to exceed customer challenges, have led to the emerging approach of a data fabric.

Defining a Data Fabric

Being able to make sense of the vast volume and variety of data sources to create a unified way of understanding the business has been a headache for business management for decades. The situation has only gotten worse as the amount and diversity of data has exploded leading to an increasing interest in a data fabric.

A data fabric is an architectural approach that abstracts data connections in a dynamic way such that it removes the complexity of requiring organizations to implement integration between individual data silos. In addition, through automation and embedded AI, this approach enables data to be automatically discovered and cataloged. This level of automation means that data scientists can spend their time building solutions to solve business problems rather than

Being able to make sense of the vast volume and variety of data sources to create a unified way of understanding the business has been a headache for business management for decades.

HURWITZ
& ASSOCIATES
Insight to Action

focusing on repetitive tasks that can be handled by AI-based tools. A data fabric is designed to understand the different types of data and can therefore effectively manage how that data needs to be handled.

Why are leading companies adopting a data fabric approach? The reason is straight forward: the data landscape is constantly changing as there are new sources and formats of data, as well as increasingly complex security and compliance requirements. A data fabric ensures that your data analysis, analytics, and AI efforts can be trusted because they are based on a complete view of all your most accurate data across the organization.

A key aspect of a data fabric is that it helps to unlock the value of your data. For example, data that is managed within line of business SaaS applications can be more easily incorporated into your data strategy. Additionally, you can incorporate data that is stored in on-premise databases, multiple cloud storage repositories, and satellite corporate locations.

One of the most important characteristics of a data fabric is that it is designed to link data sources together without moving the data from its source. Unlike many alternative approaches, a data fabric doesn't require bulk transfer of data into a centralized repository. This is quite different from historical data management architectures like data warehouses or data lakes where data needed to be copied or moved from its original location to a separate environment designed for analytics. As you can see in Figure 1 below, a data fabric connects to data wherever it resides. It simplifies the management of your data landscape by easily connecting to new data sources, making your insight and AI project more trusted because the underlying data is always up-to-date and complete. Using a data fabric allows data teams to create data and analytics pipelines that are reusable and sharable within an organization.
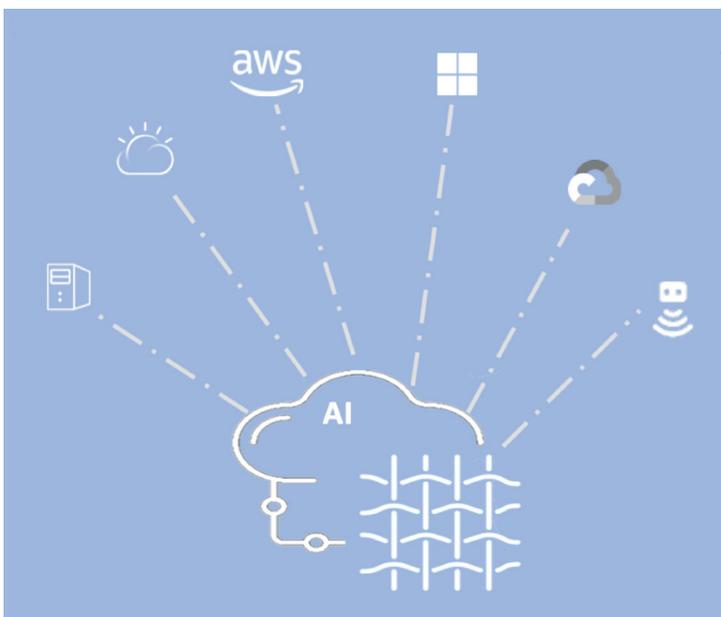
A key aspect of a data fabric is that it helps to unlock the value of your data. For example, data that is managed within line of business SaaS applications can be more easily incorporated into your data strategy.

Figure 1: An AI-enabled data fabric spanning across a variety of on premise and cloud data sources. The AI-driven data fabric has auto discovery and classification to create an intelligent metadata layer.

HURWITZ & ASSOCIATES
Insight to Action

Adopting a data fabric

Businesses that want to become more data driven are beginning to plan for the migration to a data fabric approach. In the past, organizations have attempted to build their own data fabrics by bringing together on premise tools as well as multiple cloud, platform and application tools. This approach creates more problems than it solves. Once a do-it-yourself platform has been developed, you will need a skilled team that can continually update and monitor the health of the data fabric as your data requirements change. In contrast, a data fabric provides the team with the ability to connect with all of their data at a reasonable cost while assuring continuous security and compliance. The data fabric also changes the way data experts can create and use data pipelines -- rather than creating rigid data processes that require large amounts of time and effort to update whenever the data landscape or customer needs change, pipelines can now be flexible and agile. Additionally, the data fabric approach allows data experts to create repeatable pipelines, data ingestion patterns, and a unified metadata layer so that data can be understood across the business. Unlike traditional data management techniques, a data fabric avoids creating a pipeline that is rigid and difficult to manage and take considerable manual efforts to change as the data and business processes change. The data fabric provides an organization with an agile and flexible pipeline that automates processes. This agility enables data scientists to focus on driving insights and creating ML & AI models that can be embedded into applications rather than spending the majority of their time on data preparation.

A data fabric approach helps businesses transition to making more data-driven decisions and embedding AI and machine learning within workflows and applications.

The Elements of a Successful Data Fabric

A data fabric approach helps businesses transition to making more data-driven decisions and embedding AI and machine learning within workflows and applications. To be effective your data fabric must have the following elements:

- Elasticity to grow and evolve as your data landscape changes. You need to be assured that your data fabric can support to data sources and types in the future.

- Open standards to support a variety of pre-built data connectors along with emerging data sources. Adhering to open standards allows partners and internal business units to more easily create connectors to link your data fabric with data across your entire organization.

- Understand data in context by building in automation that can help to auto classify, catalog and discover new data sources.

- Continually govern data so you can ensure compliance and know that your data is accurate and secure.

 It's important to remember, to meet business, customer, and compliance requirements, your data fabric needs to be flexible enough to safely enable your data landscape to evolve as new data types and sources are added. A data fabric that can deliver trusted, compliant, and cost-effective insights must be based on AI and automation. This level of automation will enable organizations to manage all of their data sources as if they were a single data source.

HURWITZ & ASSOCIATES
Insight to Action

IBM Cloud Pak For Data and its Data Fabric

Understanding the platform

An intelligent data fabric that is infused with automation and AI is central to IBM's next generation of IBM Cloud Pak for Data. IBM Cloud Pak for Data is the company's approach to deliver the end-to-end suite of capabilities required to help its clients succeed with data and AI. The platform offers a set of pre-integrated, Kubernetes-based, services that can be flexibly deployed on any cloud. IBM Cloud Pak for Data includes all the components needed to connect, ingest, discover, catalog, govern, and analyze data within unified, collaborative workflows.

One of the most important characteristics of the new intelligent data fabric is the ability to unify data with a universal query engine that provides high performance access to data without the need to move data across locations. The fabric also uses machine learning to discover and understand metadata and automatically map it to business terms. All of the data and analytics services available in IBM Cloud Pak for Data are containerized. The core design point of this offering is unification, collaboration, automation, and governance. IBM Cloud Pak for Data unifies data and AI services across public and private cloud infrastructures. It also provides a consistent way to enable data engineers, data stewards, data scientists, analytics professionals, and business users to collaborate through workflows so they can unlock the value of their data. Furthermore, automation and governance are infused throughout, enabling self-service analytics.

Providing clients with a data fabric

To support the next generation of IBM's Cloud Pak for Data, IBM has added more capabilities to meet the needs of its enterprise clients. IBM is currently building out its intelligent data fabric, an integral part of Cloud Pak for Data. The intelligent data fabric allows clients to create a unified data layer that can connect with and catalog a business' data no matter where that data resides. This approach helps clients more quickly enact analytics-based decision making by embracing a hybrid and multicloud architecture instead of moving data to a single repository. IBM's new data fabric is built upon its existing technologies found within Cloud Pak for Data as well as new offerings. The current capabilities include:

- **DataStage:** AI-Powered data integration for multicloud environments by combining automated data integration with DataOps, governance, and analytics.
- **Knowledge Accelerators:** Pre-created, curated glossaries for key industries that can be imported into IBM Watson Knowledge Catalog to enable governance, regulatory compliance, self-service analytics, and other governance operations.
- **Match360:** Provides users with the ability to quickly build data pipelines for analytics and other data science use cases.

continued

IBM Cloud Pak for Data includes all the components needed to connect, ingest, discover, catalog, govern, and analyze data within unified, collaborative workflows.

HURWITZ
& ASSOCIATES
Insight to Action

- **MettleCI:** A continuous integration and continuous delivery (CI/CD) solution for DataStage that automatically verifies your code, uses best practices, and allows for agile data transformation.

- **Watson Knowledge Catalog:** A cloud-based enterprise metadata repository preparing data for AI, machine learning and deep learning. Allows you to access, curate, categorize, and share data wherever it resides.

- **Watson Studio:** Operationalizes AI and optimizes decisions anywhere on IBM Cloud Pak for Data. This environment allows teams to collaborate, automate AI lifecycles, and speeds up time-to-value on an open multicloud architecture using open source frameworks.

Figure 2 highlights the new capabilities in the next generation of IBM Cloud Pak for Data and how they are helping to create the intelligent data fabric.
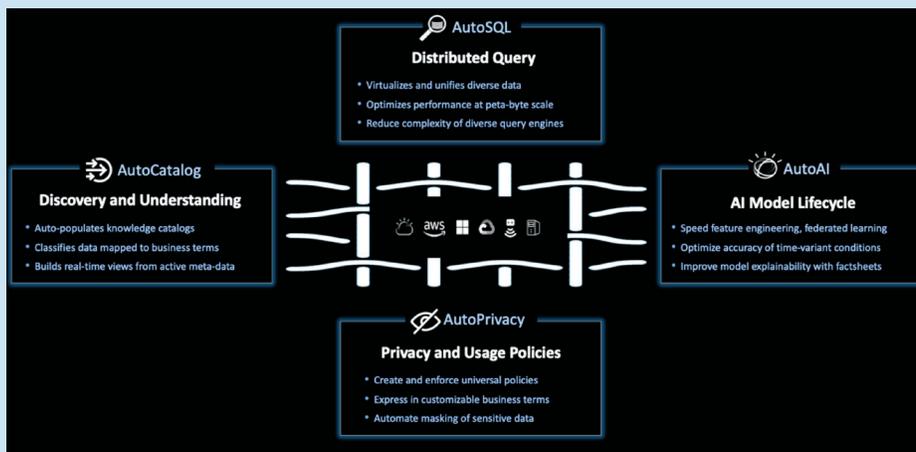
...the automation means that data scientists can focus on insights rather than continually monitoring a data platform.

Figure 2: IBM's Intelligent Data Fabric

These new capabilities heavily rely on automation to help make the data and AI process easier and less error-prone. In addition, the automation means that data scientists can focus on insights rather than continually monitoring a data platform. IBM has announced the following four new capabilities that enable a data fabric:

- **AutoSQL:** A universal query engine that automates how you access, update, and unify data across all data sources (multiple clouds, warehouses, etc.).

- **AutoAI:** Capabilities to automate and simplify the creation of AI models. Federated learning simplifies training across complex and siloed datasets; time series optimizations improve predictive accuracy across complex, time-variant conditions; and dynamic fact sheets improve model quality and explainability, while reducing the risk of policy violations.

HURWITZ & ASSOCIATES
Insight to Action

- **AutoPrivacy:** Provides automated universal data privacy capabilities to ensure that customer, employee and other sensitive data is protected throughout the analytics lifecycle.

- **AutoCatalog:** Allows clients to automatically discover, classify, and catalog data so that it can be easily found by teams across the business.

IBM's intelligent data fabric is designed to allow an organization to become more data-driven as they build trust in their analytical insights. The goals of the data fabric are to:

- Unlock data and insight that reside on multiple clouds and formats.

- Spread trusted AI and business insights throughout the entire business.

- Increase productivity throughout the data and AI lifecycle.

- Reduce the costs and risks associated with siloed data and centralized data repositories.

The data fabric is the only viable mechanism to support the needs of businesses to effectively manage their variety and complexity of data.

Conclusion

The data fabric is the only viable mechanism to support the needs of businesses to effectively manage their variety and complexity of data. With a data fabric: an organization can make effective use of the data already within their grasp; IT and business leaders have the opportunity to collaborate and support the ever-changing business landscape; and organizations can proactively manage all of their data as it exists and it expands.

HURWITZ & ASSOCIATES
Insight to Action

About Hurwitz & Associates

Hurwitz & Associates is a strategy consulting, market research and analyst firm that focuses on how technology solutions solve real world customer problems. Hurwitz research concentrates on disruptive technologies, such as Cloud and Multicloud Management, Big Data and Analytics, Machine Learning and Artificial Intelligence, Information Security, DevOps, and emerging infrastructure platforms that are defining the future of business agility. Their experienced team merges deep technical and business expertise to deliver the actionable, strategic advice clients demand. Additional information on Hurwitz & Associates can be found at www.hurwitz.com.

One Mifflin Place, Suite 400, Cambridge, MA 02138 • Tel: 617-597-1724
www.hurwitz.com