

Introducing IBM's data fabric: Seller enablement



Content

- Data fabric and related terminology
- Data fabric architecture
- IBM's data fabric solution
- Competitive view
- Why IBM?



Data fabric and related terminology



Data fabric terminology challenges

“Data fabric” is not like “data warehouse”, where there is a consistent understanding of the term

- Analyst definitions are generally consistent, but do have notable differences
- Software vendor definitions reflect their strengths and go-to-market strategies

There are multiple complementary terms:

- DataOps
- Data mesh
- Data virtualization
- Data warehouse
- Data lake
- Data lakehouse

IBM definition of data fabric

*A data fabric is an **architectural approach** to simplify data access in an organization to **facilitate self-service data consumption**.*

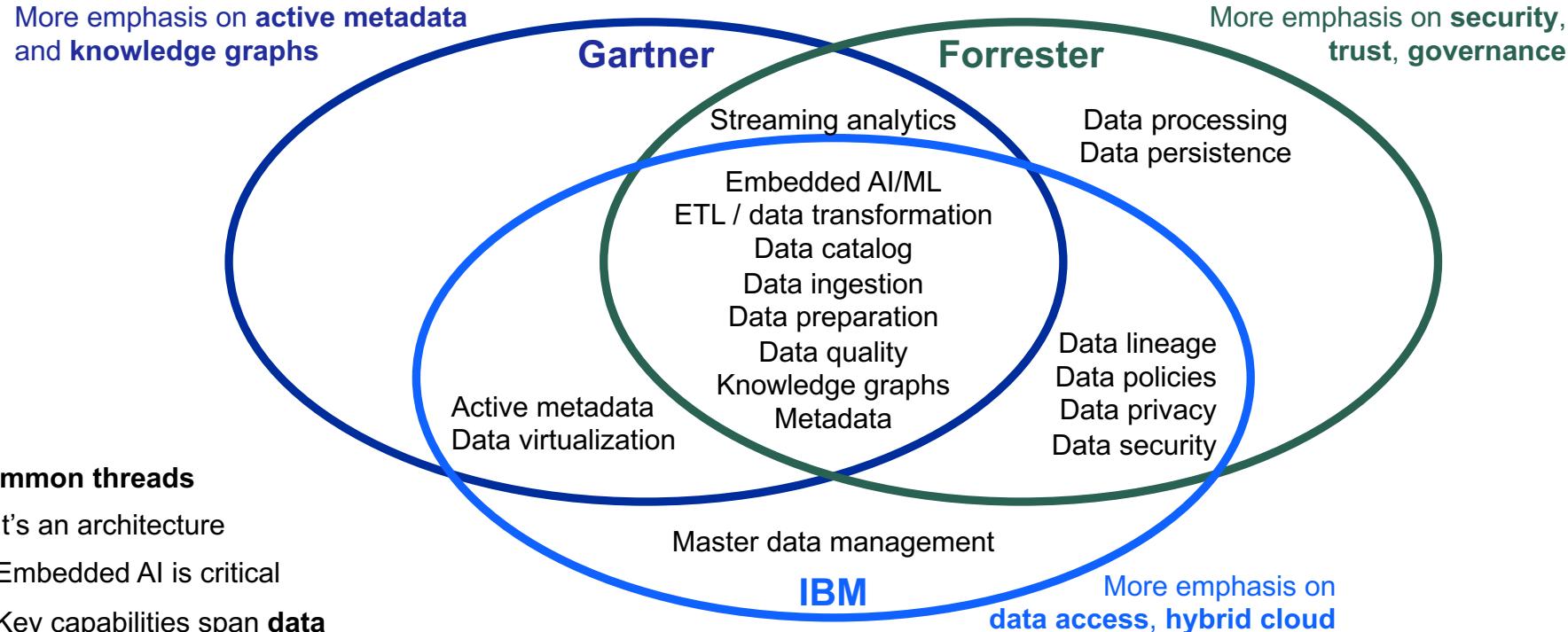
*This architecture is **agnostic to data environments, processes, utility, and geography**, all while **integrating** end-to-end data-management capabilities.*

*A data fabric **automates** data discovery, governance, and consumption, enabling enterprises to use data to maximize their value chain.*

*With a data fabric, enterprises elevate the value of their data by **providing the right data, at the right time**, regardless of where it resides.*

Data Fabric definitions from Gartner and Forrester

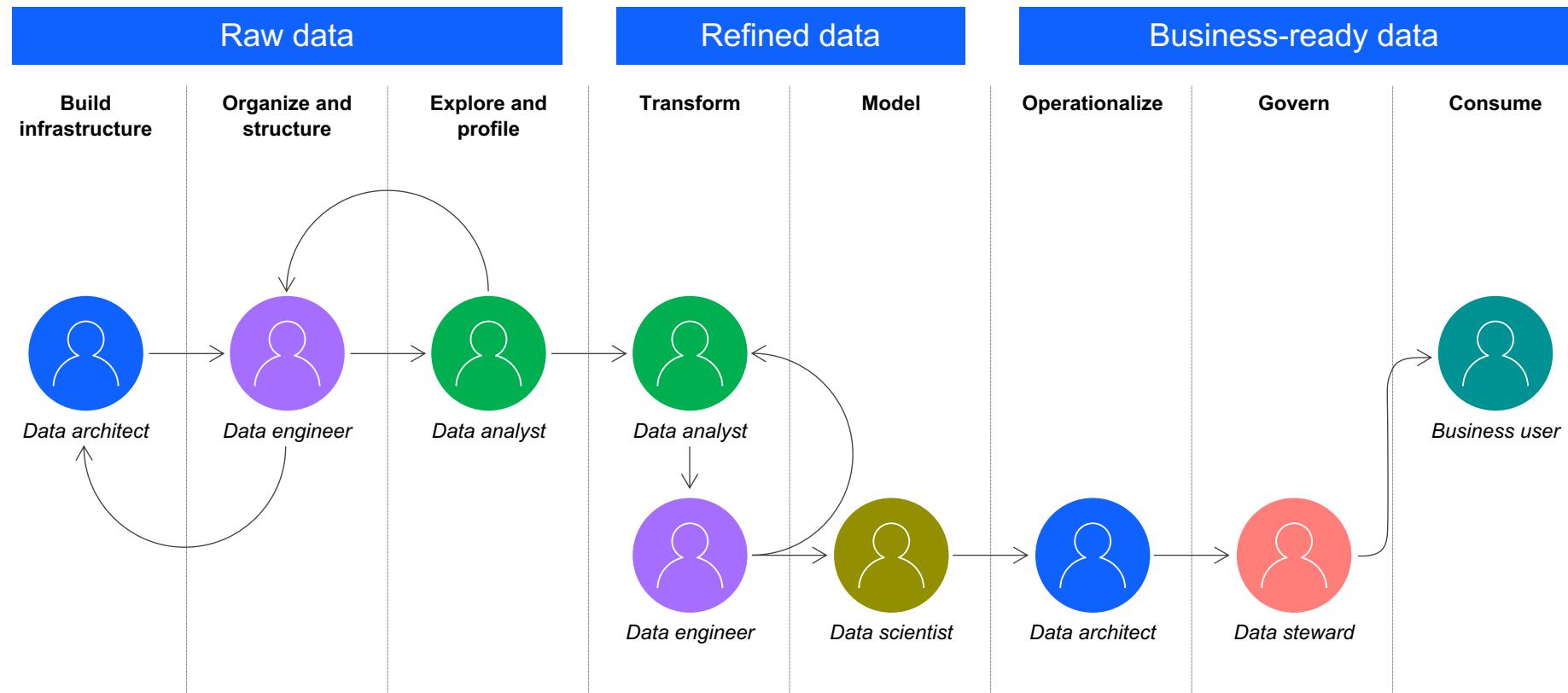
IBM's definition is mostly aligned



Common threads

- It's an architecture
- Embedded AI is critical
- Key capabilities span **data integration** and **governance**
- Connects disparate data

DataOps workflow



DataOps and data fabric

*Data operations (DataOps) is the **orchestration** of people, processes, and technology to **deliver trusted, high-quality data** to data citizens fast.*

Data fabric architectures need an agile and effective DataOps practice

DataOps is the booster for a data fabric solution, ensuring that those components are used to their full potential to facilitate self-service data exploration and experimentation.

Data mesh principles

Domain ownership

End-to-end data ownership belongs to the applicable line of business teams and departments

Data as a product

Data needs to be tied to business goals and the value they generate

Self-service data infrastructure

Any data consumer, regardless of role must have easy access to the data they need

Federated computational governance

Enables self-service for easy end user data consumption and collaboration

Data mesh and data fabric

*Data mesh is an **approach** centered on **organizational processes** to enable agile, **domain-specific ownership** and **creation of reusable data products**. It is **technology agnostic** and domain owners are responsible for the entire data lifecycle.*

Data fabric is a technology architecture, while data mesh represents processes and principles – not technology; these are not competing concepts.

As the end-goal of both data fabric and data mesh is to facilitate self-service data consumption, a data fabric architecture can be the implementation layer that supports data mesh processes.

Data virtualization

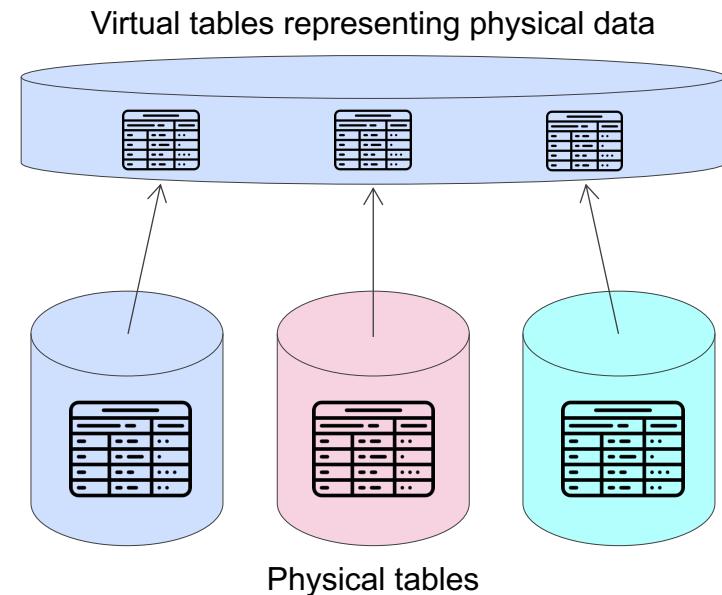
A tool that enables physical data from various sources to be represented in a virtual manner so that the data can be viewed, accessed, manipulated, and analyzed, without the need to know or understand its physical format or location, and without having to move or copy it.

Connect, access, and govern any data without the need for data movement

Abstract complexity from data consumers

Create virtual views over multiple data sources

Governance integration and security



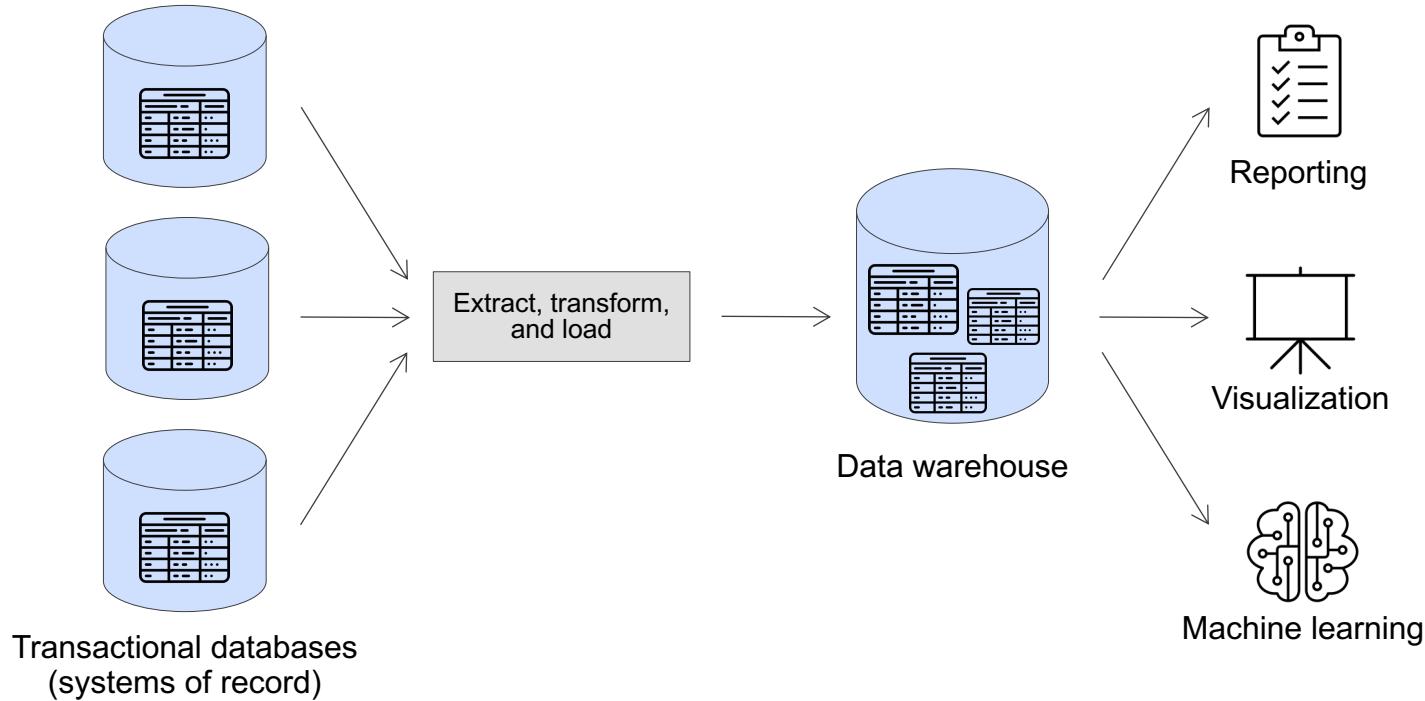
Data virtualization and data fabric

*Data virtualization **connects multiple data sources across locations** and turns all this data into virtual data views. These read-only data views **enable users to view, access, and analyze data without knowing its location.***

Data virtualization is a key technology in data fabric architectures.

Data fabric architectures provide self-service access to data, regardless of location or format. This requires automated, easy-to-use data integration tools including data virtualization.

Data warehousing



Data warehouses and data fabric

*A data warehouse contains data pulled together from different sources into a single data repository, where the **data is structured to answer specific queries from analytics applications** like dashboards, reports, or machine learning training pipelines.*

A data fabric architecture includes data warehouses as data sources.

Data warehouses contain cleansed and modeled data, while created for specific applications still has broad appeal for line of business users who are asking different questions.

Data lake architecture



Data lakes and data fabric

A data lake is a **lower-cost system** (than data warehouses) designed to **store large volumes of data in different formats**.

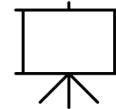
As a data source for data fabric architectures, data lakes can be a sandbox where experimentation work happens.

Data fabric architectures can use data lakes to provide efficient, cost-effective access to data that is otherwise only available from production systems.

Data lakehouse architecture



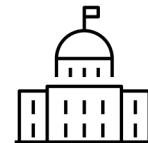
Reporting



Visualization



Machine learning



Governance
framework



Data
processing



Structured data



Document data



Unstructured data



Cloud object
storage

Data lakehouses and data fabric

A *data lakehouse* is a cloud-based data architecture designed to **support the best of both the data lake and data warehouse worlds**: inexpensive, high-volume, and flexible storage, and support for high-performance SQL-based workloads.

A data lakehouse is a supporting technology in a data fabric architecture, where it is both a data source and sandbox for experimentation.

Data fabric architectures can use data lakes to provide efficient, cost-effective access to data that is otherwise only available from production systems.

Elements of a data fabric design for data-driven organizations

Data consumers

Self-service



Find, collaborate, and access to high-quality data assets

Augmented knowledge



Unified view of data and metadata with actionable insights

Smart integration



A range of policy-driven integration styles with intelligent automation

Multi-modal data governance, security, and compliance



Unified definition and enforcement of data governance policies

Unified lifecycle



End-to-end lifecycle to build, test, deploy and manage the data fabric

IBM Cloud

AWS

Microsoft

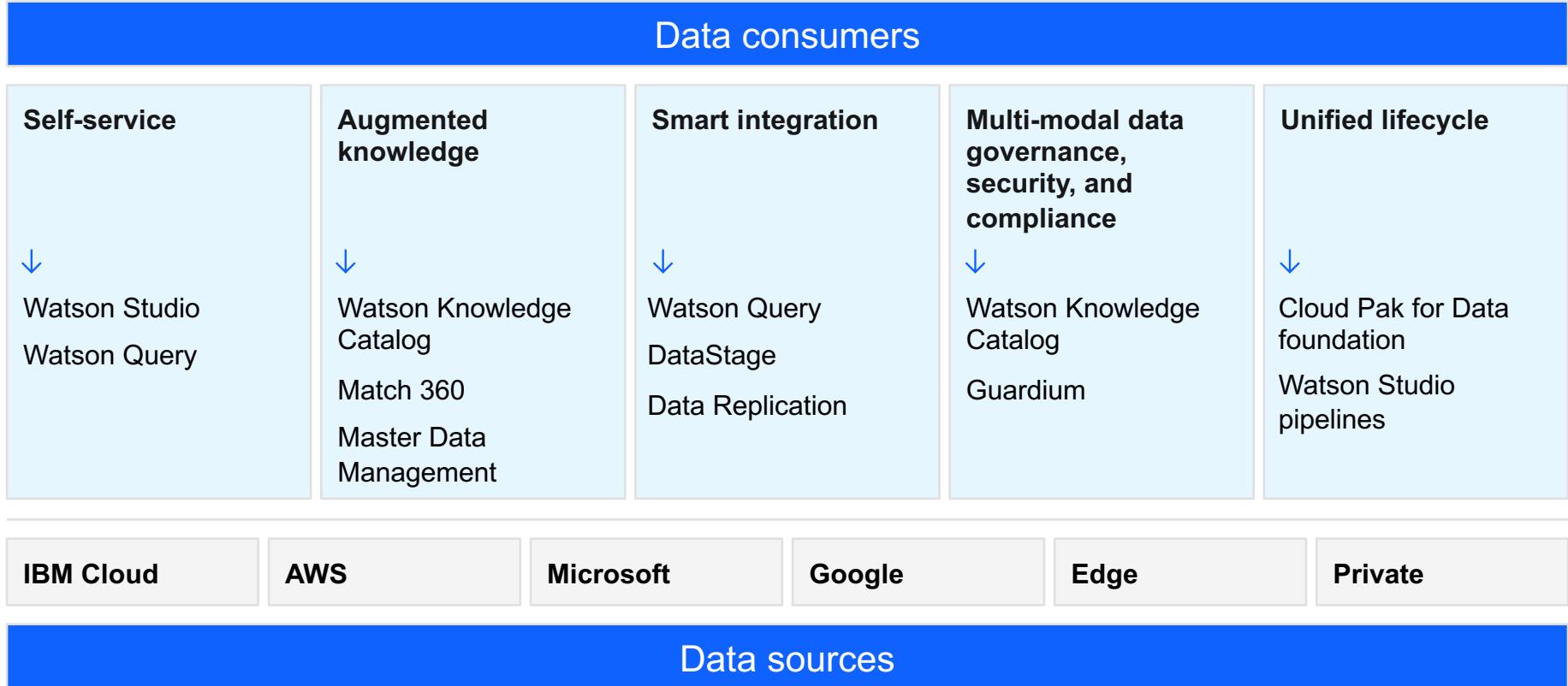
Google

Edge

Private

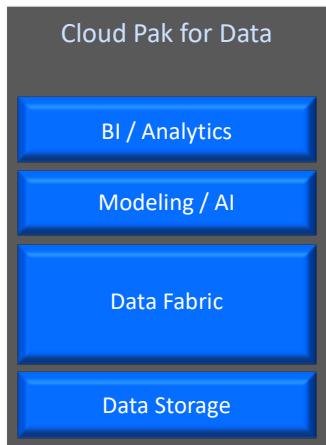
Data sources

IBM's data fabric solution within IBM Cloud Pak for Data



How does IBM's data fabric solution fit into an organization's tool landscape?

Fully integrated stack
based on Cloud
Pak for Data



Use multi-vendor tools
and storage with an
IBM data fabric



Use different tools and
storage with a mixed
data fabric and
federated catalogs



Data fabric competitive landscape



Data fabric vendor landscape

File / storage platforms	<ul style="list-style-type: none">Vendors in this group focus on software (and sometimes hardware) to enable a data lake or lake house environment (rather than data fabric).They may offer some data management technology, but the focus is on data storage rather than data management. From a data fabric perspective, those aren't data fabric solutions.
Data integration and access	<ul style="list-style-type: none">Vendors in this group focus on moving or accessing data. Data access here typically refers to vendors offering data virtualization technology.These vendors may also have some capabilities in metadata management but often do not offer the completeness in data cataloging, privacy and semantics needed for data fabric solutions.
Knowledge accelerators	<ul style="list-style-type: none">Vendors in this group focus on establishing a knowledge hub about data assets.These often utilize graph technology to establish the relationships between data assets or between data assets and a canonical model (a design pattern used to communicate between different data formats).Some vendors call those knowledge hubs data fabrics despite a clear lack of other necessary data fabric capabilities including full governance, data access or integration, cleansing or orchestration of data.
DataOps (integration + quality + catalog)	<ul style="list-style-type: none">From a data fabric perspective, vendors in this group provide a more complete solution than any other group.While these vendors support many capabilities required for a data fabric solution, IBM's data fabric offerings are differentiated by superior support for flexible (hybrid/multi cloud) deployments, scalability, and the ability to integrate other software components into their solution.
Data science / AI	<ul style="list-style-type: none">One dominant use case for data fabric is the acceleration of insights for data science. Therefore, the tight coupling between data fabric and model development & operation can be a key requirement.Vendors in this group may focus more on the model aspect rather than the data aspect and those solutions should not be confused with a data fabric.
Data hubs	<ul style="list-style-type: none">Vendors in this group provide a "physical" data hub.This data hub shouldn't be confused with a file platform as a data hub is a (multi) domain environment which creates a master data environment mapping source data to those master data domains.

High-level competition matrix

Not for customer-facing use

Vendors marked in Orange either call their solution Data Fabric or have been pro-active in marketing their product as a data fabric solution

File / storage platforms	Data integration and access	Knowledge accelerators	DataOps (integration + quality + catalog)	Data science and AI	Data hubs
HPE (MapR) ★★★	Talend ★★★	Cambridge Semantics ★★☆	Informatica ★★★	DataRobot ★★★	Cinchy ★★★
Cloudera ★★★	Denodo ★★☆	Palantir ★★☆	SAP ★★☆	Databricks ★★★	K2View ★★☆
NetApp ★★★	Precisly ★★☆	Stardog ★★☆	TIBCO ★★☆	Google ★★★	Nexla ★★★
Dell (VMWare) ★★★	Oracle ★★☆	Collibra ★★☆	Microsoft ★★☆	Alteryx ★★☆	CluedIn ★★★
Snowflake ★★★	Starburst ★★☆				
IBM ★★★	IBM ★★★	IBM ★★☆	IBM ★★★	IBM ★★★	IBM ★★★

Note: Data integration and access is a subset of DataOps. Additionally, some vendors have additional capabilities beyond their main focal area.

A European financial institution

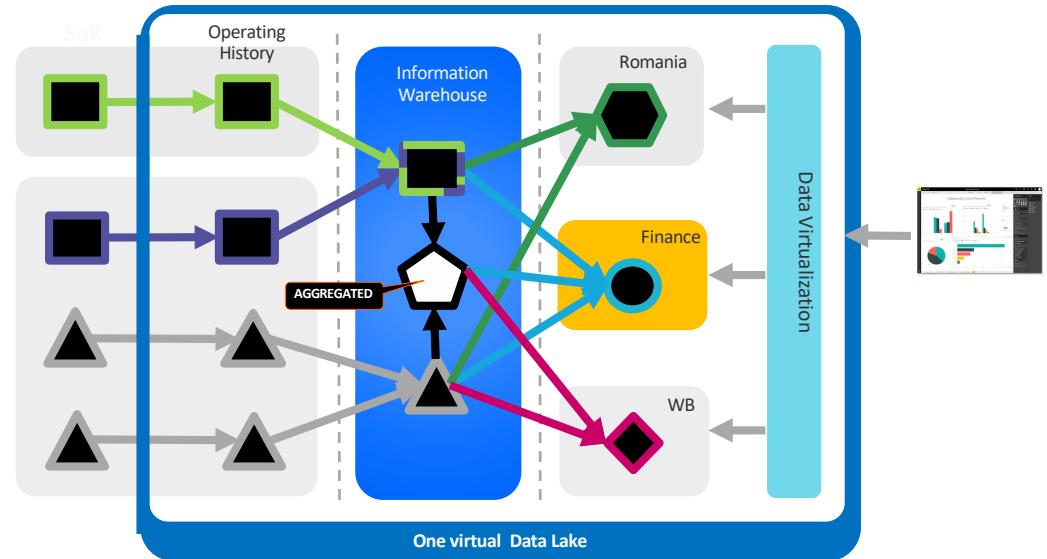
transitions from replicated local data lakes copied to each country to a governed data mesh access and security layer based on metadata

Current Challenges

- Complicated physical architecture and data chain
- Complexity in maintaining multiple data lakes
- Complexity in mapping from country to global

Benefits of the data fabric

- Single governed access pattern across multiple end points and clouds
- Provides access to data independent of physical location, only based on the meaning of data



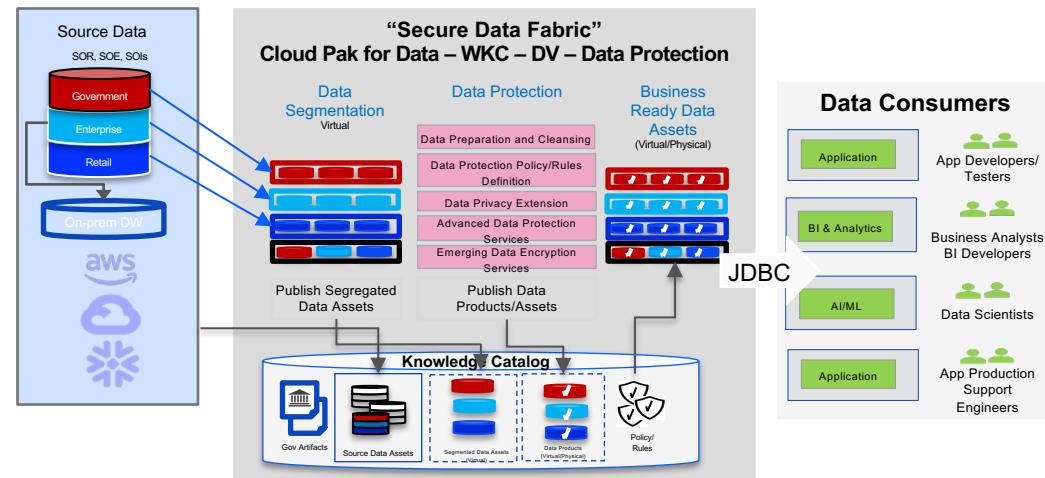
**A US telecommunication firm
with elevated data privacy and security standards enables secure self-service
data access for partners and remote users via a governed, secure data fabric**

Prior to the data fabric:

- Point to point data source connections creating a scale bottleneck
- Limited or delayed access for remote users or partners due to security reasons

With the data fabric:

- Governed data catalog and policy definition
- Policy aware, secure data bus (fabric) for data access
- Data asset/product management
- Secure self-service data consumption
- Data-centric privacy protection



Why IBM?

The only vendor able to deliver an end-to-end, seamlessly integrated data fabric solution for Hybrid Cloud

IBM Cloud Pak for Data is an ideal architectural foundation for data fabric

Completely modular, composable architecture supporting any pattern or entry points for data-driven digital transformation:

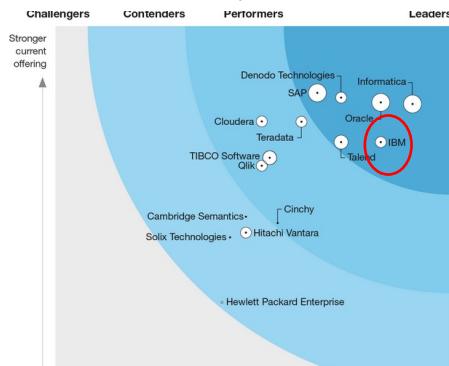
- Any deployment (SaaS, PaaS, IaaS, On-premises)
- Distributed or single hub
- Best of breed or “All in one”
- Any size with elastic scale

Recognized as technology and thought leader in AI – one of the core differentiators between traditional data management and data fabric

IBM has one of the most complete data fabric solutions in the market

Not for customer-facing use

Forrester Wave Enterprise Data Fabric 2022



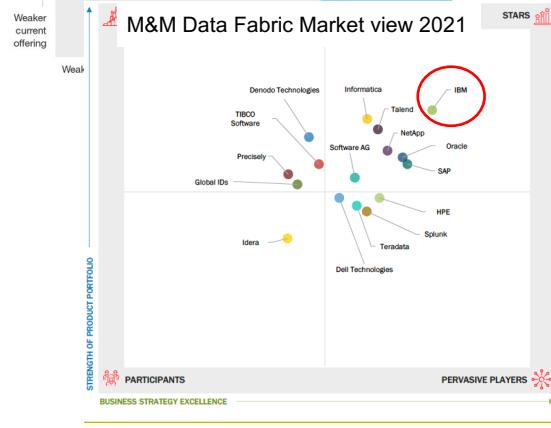
Gartner MQ for Metadata Management 2020



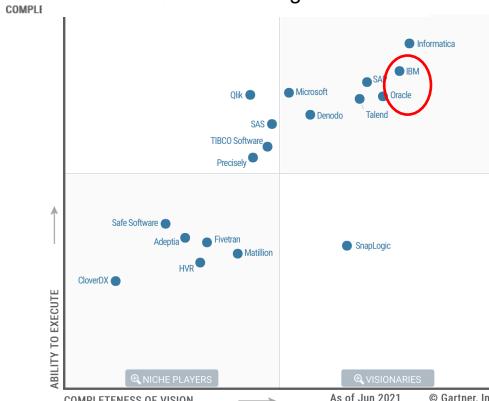
Gartner MQ for Data Quality Solutions 2020



M&M Data Fabric Market view 2021



Gartner MQ for Data Integration Tools 2021



Gartner MQ for Data Science Platforms 2021



