THE DATA FABRIC HANDBOOK

GIGASPACES

Supercharge digital innovation with real-time data services

www.gigaspaces.com

Contents

Introduction

A search on the Internet for the term ‘Data Fabric’ generates hundreds of results. Gartner, for example, defines a data fabric as “an emerging data management design for attaining flexible, reusable and augmented data management and integration” and explains that data fabric architecture “utilizes various data management technologies, such as data catalogs, data integration, data virtualization, orchestration and knowledge graph tools”.

This sounds like heavy stuff. But at its core, data fabric is an architectural design concept that aspires to integrate, manage and enable easy access to enterprise data, wherever it resides.

Why do we need a data fabric?

The need for Data Fabric concepts is driven by the realization that data drives all aspects of modern digital applications. The COVID pandemic accelerated the digital revolution, slashing implementation times for online services. It also speeded up cloud adoption. As a result, most enterprise systems reside both in public and private clouds – as well as on premises. This has led to huge complexity in data environments. **The goal of data fabric is to streamline and simplify the management of diverse data types and data sources and ensure it is accessible to stakeholders and the digital apps that are the foundation of human activity today.**

The key elements of a data fabric

A data fabric is a technology-driven architecture. As such it relies on new and existing technology solutions, and a high degree of automated processes to analyze, classify and integrate data. Below are some of the key elements of a data fabric:

Metadata and Active Metadata

The use of metadata and metadata recommendations is key in data fabric design. Metadata provides contextual information about the data, such as data source, data structure, data lineage and the relationship between types of data as well as its source, structure, lineage, and relationships. Metadata provides descriptive information about the data that improves the understanding and usability of the data. It plays a key role in helping systems analyze and catalog data by facilitating data discovery and search, data interpretation, data lineage and provenance, as well as data integration.

Metadata is a well-established element in data management. Active metadata is an evolving technology that elevates our ability to gain automatic intelligent insights into data.

When to consider a data fabric approach?

- ✔ Mature active metadata analysis tools are **already implemented** in your organization
- ✔ Mature data governance and automation tools are **already in use**

Active Metadata

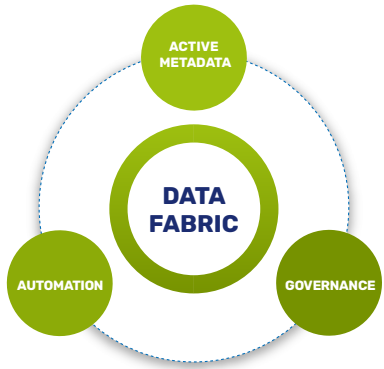
Active metadata takes metadata to a new level. Whereas metadata is static, active metadata is dynamically updated automatically. It is linked to real-time processes and actions, creating a responsive metadata ecosystem that can provide insights into the data. Active metadata discovers data access patterns, usage trends, and data popularity and can provide proactive information on where data should be cataloged and stored, instead of having to rely on human decisions. Active metadata also plays a key role in being able to maintain data governance.

Data Governance

Data governance concerns the management and control of an organization's data. It covers a range of aspects including the policies and processes that ensure data quality, integrity, security, privacy, and compliance. Active metadata supports data governance by monitoring data access, usage, and adherence to governance policies in real-time. It helps to identify compliance violations, ensuring data security, privacy, and regulatory compliance.

Automation

A key differentiator of data fabric architecture is the fact that it strives to be as automated as possible. Automation helps with efficient data processing, integration, management, and governance. Specifically, it streamlines data integration by automatically fetching, transforming, and loading data from various sources into the data fabric. Automation simplifies metadata management by automatically capturing metadata from various data sources and extracting attributes such as data types, schema, lineage, and relationships. In this regard, automation ensures that metadata is up-to-date and accurate. Automation also helps with metadata discovery, enabling the classification and indexing of new data in underlying systems.



Key Elements of a Data Fabric

The technology solutions that make up a data fabric

Data fabric is a relatively new and emerging design concept. It relies on existing technology solutions as well as evolving ones. Since this is a broad design concept, organizations have a large degree of flexibility in utilizing solutions already implemented and augmenting these with new solutions as needed. Below are some key technology components that can contribute to a data fabric.

Governance Solutions

A data governance solution or platform helps establish the structure and guidelines for managing data across an enterprise. It defines roles, responsibilities, and procedures that relate to an organization's data environment. Good governance is important since it enables data stewards to maintain compliance and enforce a company's data usage policies when data resides in multiple systems.

Operational Data Hubs

An Operational Data Hub, also known as a Digital Integration Hub, is a solution designed to efficiently connect data producers and data consumers – primarily for real-time analytics and transactional use cases, such as predictive maintenance in industrial manufacturing, ecommerce (retail) and online banking.

Operational Data Hubs play a key role in data fabric design, by allowing organizations to share data among stakeholders and utilize data that resides in diverse systems for highly performant transactional workloads. Operational Data Hubs comprise CDC and event-driven streaming to ensure data is always current, while low code enables data professionals to expose data easily through APIs or SQL statements. To this end, Operational Data Hubs facilitate the delivery of data products and services in real-time and enable data democratization.

By 2025, 80% of organizations will have deployed multiple data hubs as part of their data fabric to drive mission-critical data and analytics sharing and governance.

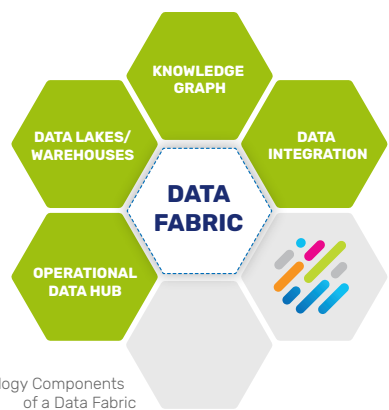
Gartner

Data Lakes/ Data Warehouses

Data lakes and data warehouses are central data repositories that store large amounts of raw and unprocessed data in its native format. It can store diverse data types, such as structured, semi-structured, and unstructured data, without the need for pre-defined schema or data transformations. Data lakes store data as-is, maintaining its original format. Many organizations already collect vast amounts of data in data lakes or warehouses but struggle utilizing this data efficiently. One of the aims of data fabric design is to augment data lakes and warehouses with additional components to enable better access to – and usage of – this stored data.

Active Metadata Capabilities

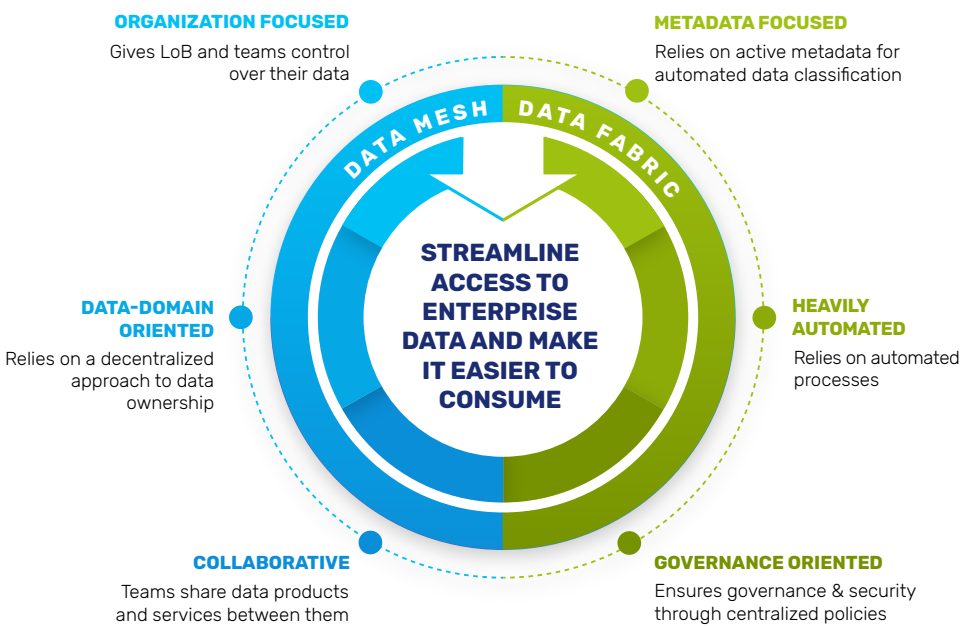
The ability to leverage metadata cuts across most data solutions. The innovative and novel angle in data fabric is the ability to utilize active metadata. This is still an emerging capability and there are not many mature solutions which enable the high degree of automated active metadata that data fabric design envisions. There is likely to be continuing evolution in this solution area to enable better visibility and classification of data.



Select Technology Components of a Data Fabric

The difference between a data fabric and a data mesh

Data mesh is a data design concept that has also been receiving a lot of buzz. At their core data fabric and data mesh aim to achieve the same outcome: make it easier and simpler to access enterprise data and make data easier to consume by various stakeholders in the organization. They differ, however, in how this is achieved. Unlike data fabric, the data mesh approach relies on the human element. In a data mesh, lines of business and functional teams manage and own their own organization in the form of data services or data products. This approach too, places a premium on 'data democratization' but relies on data experts to implement it, rather than the automated approach advocated by a data fabric.



Conclusion

As noted, the goal of a data fabric architecture is to enable easy access to data residing in multiple and diverse enterprise systems. Key components of data fabric design include active metadata, data governance and automation. Organizations can implement data fabric design by using already implemented solutions such as data lakes and augment these with innovative solutions such as those that offer active metadata capabilities and unified platforms that offer real-time highly performant data delivery such as Operational Data Hubs. Ultimately, data-driven organizations need to be able to utilize data in the most efficient way possible, enabling easy access and delivery to drive digital services and real-time analytics. Data fabric design offers a valuable framework for achieving this goal.

About GigaSpaces

GigaSpaces is building on its in-memory computing and operational data store technologies to offer one of the market's first Operational Data Hubs, an out-of-the-box unified real-time data platform that empowers organizations' digital transformation, by unlocking data that resides in legacy and cloud systems. Whether you need to accelerate one application with cache, or modernize your architecture, GigaSpaces can future-proof your investment. Never before has it been this straightforward to accelerate API-powered digital applications and democratize access to data across your enterprise.

Smart DIH is part of the GigaSpaces Smart suite of products, alongside the award-winning Smart Cache solution. GigaSpaces offices are located in the US, Europe and Israel, with partners such as Capgemini and Cognizant around the globe; serving customers such as Morgan Stanley, Bank of America, CSX, Goldman Sachs, Societe Generale, Credit Agricole, American Airlines, Avanza Bank, Avaya, CLSA, and UBS.

For more information visit www.gigaspaces.com.