iJRASET
International Journal For Research in Applied Science and Engineering Technology



# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 12 **Issue:** V **Month of publication:** May 2024

**DOI:** https://doi.org/10.22214/ijraset.2024.62241

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538*
*Volume 12 Issue V May 2024- Available at www.ijraset.com*

# Data Fabric

Diwakar Manna
*LTIMindtree, USA*

## I. INTRODUCTION

The concept of a data fabric is relatively new, and its definition is broad. To understand where the concept comes from, it makes sense to step back and analyze the evolution of data management practices over the past decade. The era dominated by the enterprise data warehouse (EDW) is passing. New big data initiatives have fuelled the rise of emerging disciplines like Machine Learning (ML) and data science for business decision making. Citizen integrators and power users were brought onto the business intelligence stage for self service, as IT departments struggled to manage more systems with lower budgets. Cloud and software-as-a-service (SaaS) solutions addressed some of these issues, but this also increased data distribution and siloes, creating new challenges.

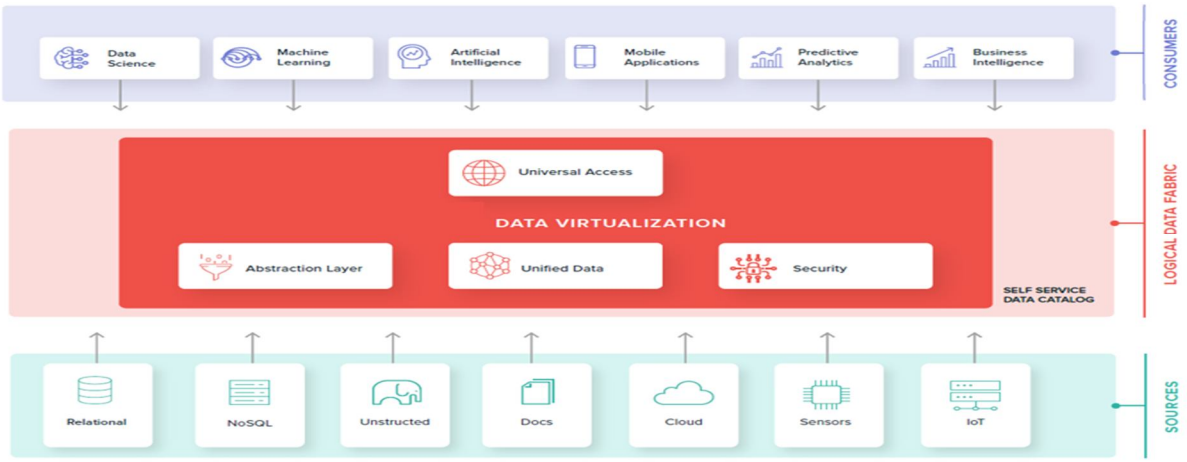
## II. IN SUMMARY

1) New methods in advanced analytics and machine learning practices gave rise to increasingly complex data requirements.
2) The evolution of different specialized tools addressing different data management needs impeded organizations in establishing a “single version of the truth.” These new tools include EDWs, data marts, relational databases (RDBMS), data lakes, no SQL systems, internal and external REST APIs, real-time data feeds including social media feeds, and many more.
3) Multiple personas now require access to the data: BI analysts, citizen integrators, data scientists, data stewards, IT, and data security professionals, each with different skills and requirements.
4) Transitions to the cloud (or to multiple cloud platforms), which creates hybrid ecosystems in which data becomes physically fragmented. IT needs flexibility to adapt to new architectures while supporting the business with minimal interruption.
5) Organizations must demonstrate higher standards in compliance and governance to fulfill specific legal frameworks (GDPR, CCPA) and external threats.
6) Securing and governing a hybrid ecosystem can be complex and error prone.

## III. DATA FABRIC-ARCHITECTURE BASED ON THESE CORE IDEAS

1) A common access layer for all data sources and all consumers, which hides the complexity of the deployment and provides a single logical system for consumption.
2) It should be compatible with various data delivery styles (including, but not limited to, ETL, streaming, replication, messaging, and data virtualization or data microservices).
3) Availability of multiple data integration strategies to be used seamlessly, depending on the use case, for both analytical and operational scenarios.
4) Additional semantics to make data elements (and the relationships and connections between them) easier to consume, operate, and manipulate.
5) Broader governance, documentation, and security features across the board, geared toward providing stronger trust and confidence in the data.
6) Automation, leveraging active metadata and AI, to make it significantly easier to develop, operate, and use such a system.
7) It should support all types of data users – including IT users (for complex integration requirements) and business users (for self-service data preparation).

©IJRASET: All Rights are Reserved 

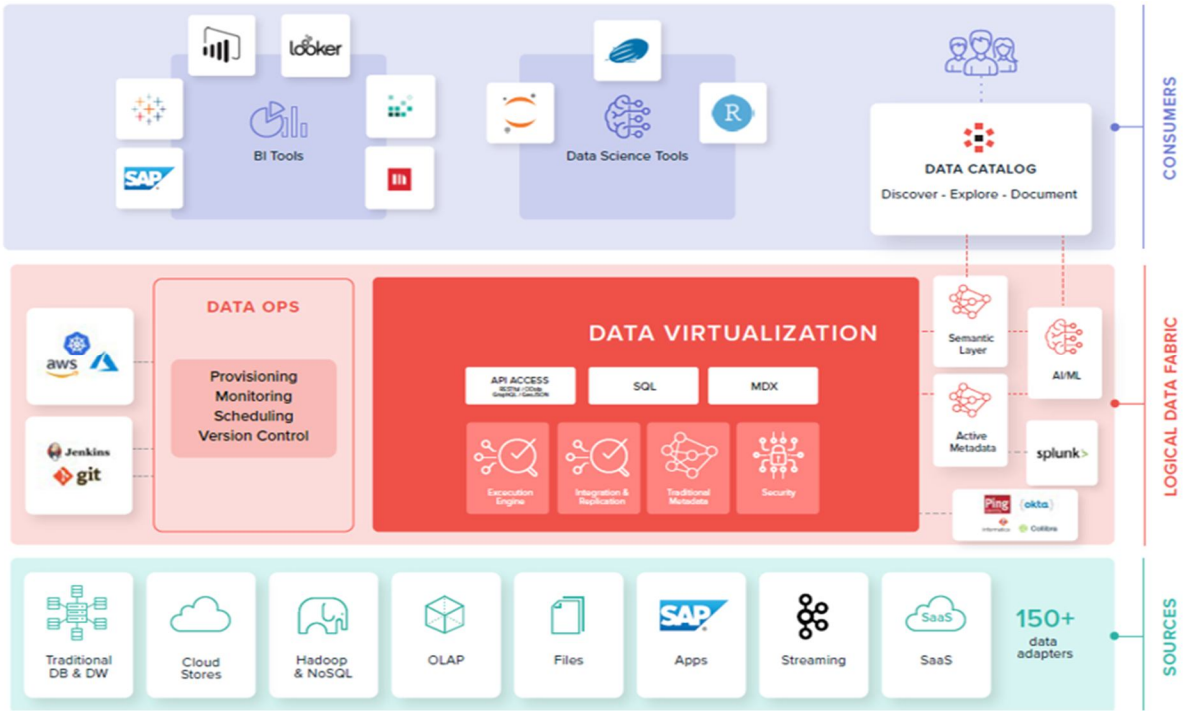



High Level Diagram of a Logical Data Fabric

Leading industry analyst Gartner defines data fabric as "an architecture pattern that informs and automates the design, integration, and deployment of data objects regardless of deployment platforms and architectural approaches. It utilizes continuous analytics and AI/ML over all metadata assets to provide actionable insights and recommendations on data management and integration design and deployment patterns. This results in faster, informed and, in some cases, completely automated data access and sharing."
The final goal of data fabric is, therefore, to enable a more agile, seamless, and, in many cases, automated approach to data access and data integration. It should provide enough complexity to enable advanced analytics while at the same time offering a friendly facade with which business users can interact. A mature data fabric should be able to support both analytical and operational scenarios.

## IV. CORE COMPONENTS OF A DATA FABRIC

The diagram below illustrates the main components of a logical data fabric. However, architects need to understand the specific requirements of each individual implementation and how the components of the logical data fabric work together to meet those requirements. In the next section, we will cover those components in detail.




Functional diagram of a Logical Data Fabric

1) A Data Virtualization Engine: abstracts data and allows for the decoupling of applications/data usage from data sources to provide a common access layer.
2) A key component of this engine is an intelligent query optimizer that helps reduce processing costs and optimizes speed.
3) An Augmented Data Catalog: facilitates data exploration and discovery and improves collaboration and data governance.
4) Active Metadata: enables auditing and historical analysis and serves as the foundation for AI processes.
5) A Semantic Layer with Extended Metadata: enriches traditional technical information with business terms, tags, status, or documentation, to fuel improvements in self-service, security, and governance across all data assets.
6) AI-Based Recommendations: useful throughout the platform to learn from usage and simplify the entire lifecycle of the data management practice, including development, operations, performance tuning, etc.
7) DataOps and Multi-Cloud Provisioning: reduces management and operational cost and enables the system to be cloud-vendor agnostic.

## V. KEY BENEFITS OF A LOGICAL DATA FABRIC

A logical data fabric approach to data management offers numerous benefits, but some of the most salient ones are:

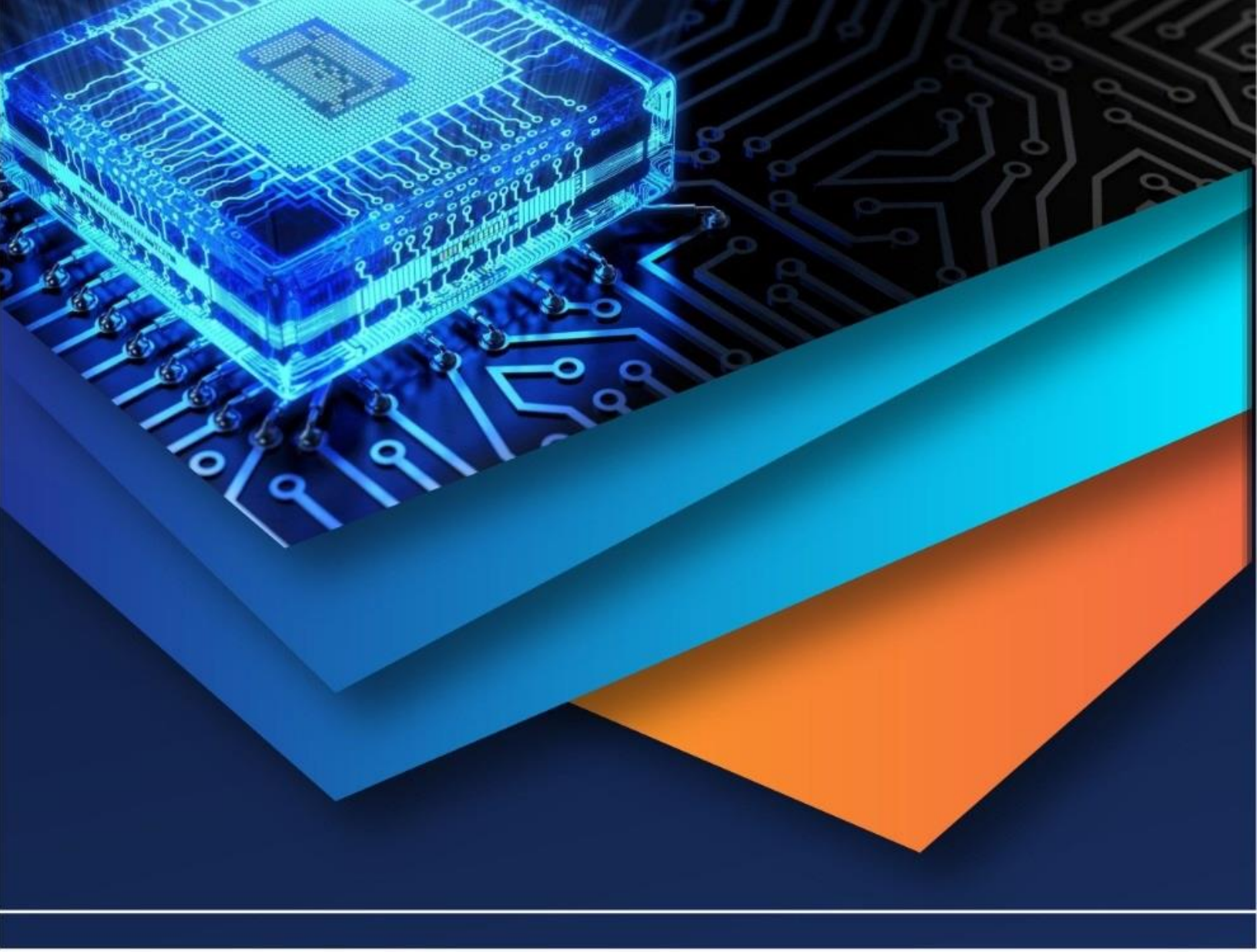
1) Improved business decisions: By combining an integrated data catalog, which offers a simple-yet-powerful tool to explore data across systems, with the advanced use of semantics, which brings meaning and context to data, a logical data fabric makes self-service a more streamlined, trusted activity allowing users to find, access and share data for better decisions.
2) Near instant answers to your queries: Advanced techniques like Smart Query Acceleration and advanced caching options, with an architecture designed to scale, ensuring data queries are returned immediately.
3) Respond to business requests more quickly: IT has a variety of methods and systems at its fingertips to manage data. They can incorporate new datasets in a few clicks and secure them. In addition, they can choose a variety of integration techniques (virtualized federation, full replication to another system, ELT, etc.) from a single, integrated Design Studio.
4) Reduced data management overhead: Automate or augment many data management tasks with activity metadata combined with AI-based recommendations to simplify the usage and operation of the platform, freeing the data team to focus on more strategic initiatives.
5) Enhanced security and governance: Logical data fabric provides a global data access layer with which to enforce security and governance across the organization, regardless of the capabilities of each data source, enabling consistent enforcement of security policies.

## VI. CONCLUSION

Data fabric is not merely a combination of traditional and contemporary technologies but a design concept that changes the focus of human and machine workloads. It is a forward-thinking approach to data management, combining the power of multiple components to create a powerful distributed and logical architecture that aims to address the big challenges of the enterprise data landscape today.

A virtual layer, enabled by a platform like AWS Glue, Denodo, Data Virtuality, SAP HANA Cloud, and Amazon Redshift, offers an enterprise-ready foundation for logical data fabric. In addition to traditional data virtualization capabilities like the federation engine, the Platform provides a thorough set of components that enable multiple integration strategies, advanced performance techniques, and AI-based recommendations. At the same time, platform should offer multiple UIs tailored for different personas (a design studio, a data catalog, a solution manager, etc.) and mature API-based integration points to successfully integrate data across any organization.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)