LTIMindtree

WHITEPAPER

Revolutionize Data Management with GenAI-Powered Data Fabric.

Author: **Shibaji Mitra | Sanjay Biswas | Kallol Dutta**

LTIMindtree

Table of contents

Overview

The volume of data generated by organizations is increasing. Moreover, it is constantly changing and quite dynamic. It is amazing how data has grown and changed throughout time. The amount of data generated has grown enormously during the past decade. Businesses can obtain up-to-date insights from such dynamic data, which facilitates prompt and well-informed decision-making. To manage such massive amounts of dynamically changing data and guarantee accuracy it necessitates robust systems. It is difficult to ensure data consistency and correctness in real-time, and comprehensive analytics tools are needed.

This whitepaper explains the significance of GenAI powered data fabric and its crucial role in offering a unified data view by integrating data from diverse source systems. This whitepaper also discusses the potential building blocks of a data fabric solution.

Intended audience

This paper's intended audience is individuals who facilitate and consume data, i.e., architects, IT department, data analysts and senior management.

This paper is intended for:

01

Anyone who wants to learn more about data fabric and its associated components.

02

Data analysts and senior management–empowering to discover, explore, and generate insights.

03

Data architects and individuals in the IT department who are responsible for designing and implementing an organization's data landscape to ensure that data is integrated, standardized and made consumable.

Key data challenges across industries

In today's world, data is continuously growing. This enormous data growth is fueled by the widespread use of digital devices, the internet, and technological advancements. Everything from smartphones to sensors and appliances keeps generating and streaming data. This data is subsequently analyzed to extract insights that can enhance business efficiency and foster innovation across diverse industries.

Handling these extensive volumes of data presents significant challenges across various capabilities, including data integration, data quality, data governance, master data management, metadata management, and data observability.

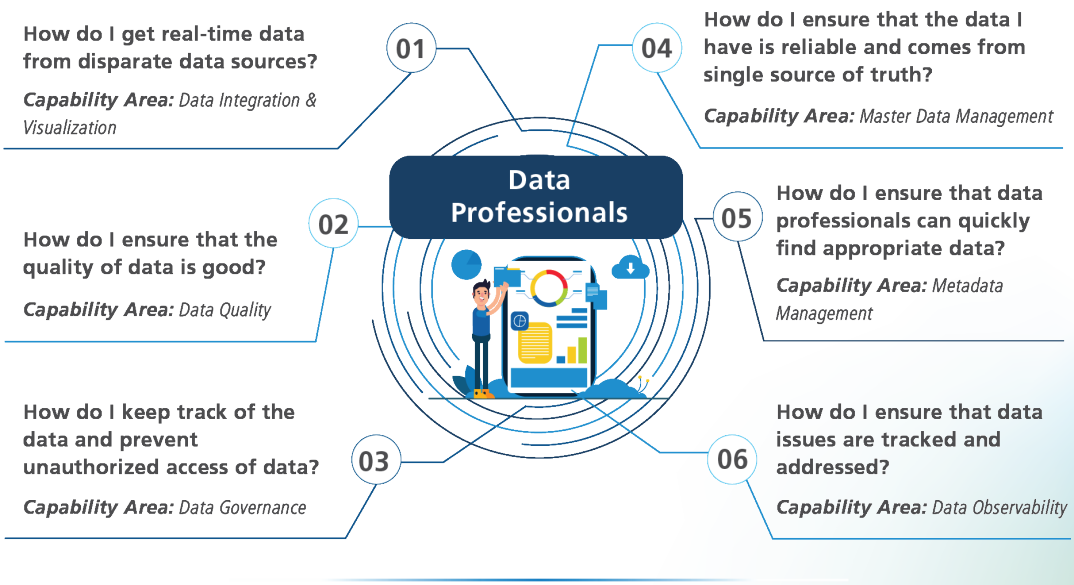


Figure 1: Common data challenges

Based on the Deloitte and McKinsey analysis across industries, data integration, data quality, data consumption and data governance are the biggest data challenge affecting more than 60% of the industries.

As businesses strive to improve their performance and achieve complete digitalization, gaining deeper insights through business and operational data has become more crucial than ever. However, in a rush to become data-driven, many organizations prioritize integrating data from multiple siloed systems over creating a self-service and collaborative workplace. This is where data fabric comes into play as it

accelerates digital transformation by providing a robust foundation for innovation and improved customer experiences. It enhances data accessibility and governance, enabling holistic data analysis and informed decision-making.

Data capabilities needed to address these challenges

01
Data integration

Data integration involves combining data from different sources, whether they are in the cloud or on-premises, into a unified view. This process ensures that data from various systems can be accessed and analyzed together, providing a comprehensive understanding of the information.

02
Data security and privacy

Data security and privacy focuses on protecting data from unauthorized access and ensuring that it is handled in compliance with relevant policies and regulations. This includes implementing measures to monitor data usage, enforce security protocols, and ensure that data privacy is maintained.

03
Data quality

Data quality is about ensuring that the data used in your systems is accurate, complete, and reliable. This involves setting up rules and processes to check for errors, inconsistencies, and other issues that could affect the integrity of the data.

04
Metadata management

Metadata management involves organizing and maintaining data about your data. This includes creating a catalog of data assets, ensuring that metadata is accurate and up-to-date, and making it easily accessible and usable for business purposes. This helps in building a solid data foundation that supports efficient data management and usage.

05
Master data management

Master data management (MDM) is the practice of managing the core data that is essential to the operations of a business. This includes data about customers, products, employees, and other key entities. MDM ensures that this data is consistent, accurate, and available across the organization.

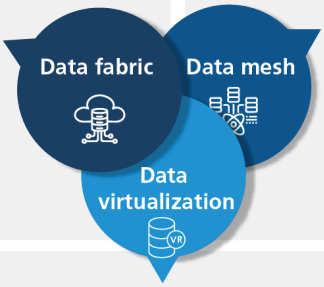
06
Data consumption

Data consumption refers to the ways in which data is accessed and used for analysis and decision-making. This includes enabling rapid analytics on both processed and raw data to derive insights that can inform business strategies and operations.

Various architecture approaches to address data challenges

Data mesh, data virtualization and data fabric are the popular architecture approaches to address the data challenges faced by organizations.

Decentralized data ownership: Data is managed by domain-specific teams, promoting accountability
Data as a product: Each data domain treats its data as a product, ensuring quality and usability
Self-serve data infrastructure: Provides tools and platforms for teams to manage their data independently
Federated governance: Centralized policies with decentralized execution, ensuring compliance and security



Real-time data access: Allows real-time access to data without moving it from its source
Unified data view: Provides a single, unified view of data from multiple sources
Reduced data movement: Minimizes the need for data replication and movement
Enhanced data security: Data remains in its original location, reducing security risks

Unified data architecture: Integrates various data sources into a cohesive architecture
Automation and intelligence: Uses AI and machine learning for data discovery, classification, and enrichment
Enhanced data governance: Ensures data privacy, security, and compliance across the data landscape
Smart integration: Supports various data integration methods like data virtualization, ETL, and streaming

Table 1: Various architectural approaches

Aspect	Data Mesh	Data Virtualization	Data Fabric
Definition	Decentralized data management approach focusing on domain-driven data ownership.	Data integration technique creating a virtual layer to access data from multiple sources.	Unified data management framework creating a seamless, interconnected data ecosystem.
Data Ownership	Domain-oriented, decentralized ownership.	Centralized virtual layer, but data remains in original sources.	Centralized, cohesive architecture with unified data access.
Data Integration	Federated data architecture enabling data sharing across domains.	Virtual abstraction layer connecting disparate data sources.	Seamless integration of various data sources and systems.
Implementation Complexity	High, due to decentralized nature and need for domain expertise.	Moderate, requires setting up virtual layers and connections.	Moderate to high, depending on the existing data infrastructure.
Use Cases	Best for organizations with strong domain boundaries and autonomy.	Ideal for quick access to data across multiple sources without moving data.	Suitable for organizations needing a unified, real-time data access layer.
Cost	High implementation cost due to decentralized data ownership and domain-specific data products.	Less implementation cost as it does not require physical data movement.	High initial cost due to the need for a unified data architecture and integration of various data sources. Cost reduces with time.

Table 2: Key differences between data mesh, data virtualization and data fabric

What is GenAI-powered data fabric?

Gartner defines data fabric as, "A data fabric is an emerging data management design for attaining flexible, reusable and augmented data integration pipelines, services and semantics. A data fabric supports both operational and analytics use cases delivered across multiple deployment and orchestration platforms and processes."

Data fabric, **powered by metadata and knowledge graph**, acts as a framework for **machine-driven data integration**. Data fabric provides a **unified layer on top of heterogeneous data** and enables a consistent view regardless of storage location or user access. Data fabric seamlessly connects data pipelines across on-premises, cloud, and hybrid platforms and integrates capabilities like **data quality, master data management, metadata management, data governance** and **data consumption** under one umbrella.

Artificial Intelligence (AI) and **Generative AI (GenAI)** further enhance the data fabric by automating data integration processes, improving data quality, and enabling advanced analytics. AI algorithms can identify patterns and anomalies in data, while GenAI can generate synthetic data to fill gaps, simulate scenarios, and support decision-making. Together, they empower organizations to derive deeper insights and drive innovation from their data assets.

Why data fabric and benefits it offers?

A data fabric is an architecture and set of data services that provide consistent capabilities across a choice of endpoints spanning hybrid multi-cloud environments. Here's how each point fits into the data fabric framework:

Integrates heterogeneous data

A data fabric automates the integration of disparate data sources, providing a unified view and seamless access to data across various environments, whether on-premises or in the cloud.

Enhances data quality

Within a data fabric, automated data quality checks ensure that data remains consistent and accurate across all sources, enhancing overall data reliability.

Better metadata management

A data fabric automatically captures, catalogs, and maintains metadata, making it easier to manage and utilize data assets effectively. This ensures that metadata is always up-to-date and accessible.

Improves data governance

Data fabrics enforce real-time governance protocols and policies, ensuring compliance with regulations and maintaining data security. This helps in monitoring and governing data usage effectively.

Manage master data

In a data fabric, master data is managed centrally by integrating and harmonizing data from various sources into a single, consistent view. This ensures that key business entities are accurately represented.

Empowered data consumption

Data fabrics enable business intelligence (BI) reporting and analytics by providing integrated, real-time data views tailored to user roles. This supports efficient decision-making by delivering relevant insights to the right users.

Increases flexibility

A data fabric supports various data consumption patterns, allowing organizations to rapidly adapt to changing business needs and leverage data in multiple ways.

Reduces technical debt

By minimizing redundant data copies and complex pipelines, a data fabric reduces technical debt, lowering maintenance costs and simplifying data management.

Evolution of GenAI-powered data fabric

The concept of 'data fabric' traces its roots back to the traditional data warehouses implementation process started in the 1980s, which primarily dealt with a structured format of data storage. As data volumes expanded, big data platforms and data lakes emerged, allowing the management of vast amounts of unstructured data. Building upon these foundations, data fabric integrates diverse data sources into a cohesive architecture. Its network-based design ensures seamless data access and effective data governance. This evolution facilitates data integration including real-time data processing to perform various advanced analytics that serve to produce more data-driven insights for fostering innovation and efficiency.

Incorporating generative AI into the data fabric architecture further enhances its capabilities. Generative AI can automate the creation of data models, predict data trends, and generate synthetic data to fill gaps in datasets. This integration allows for more sophisticated data analysis and decision-making processes. By leveraging AI, data fabrics can provide deeper insights, automate complex data tasks, and support advanced analytics, ultimately driving greater innovation and operational efficiency.

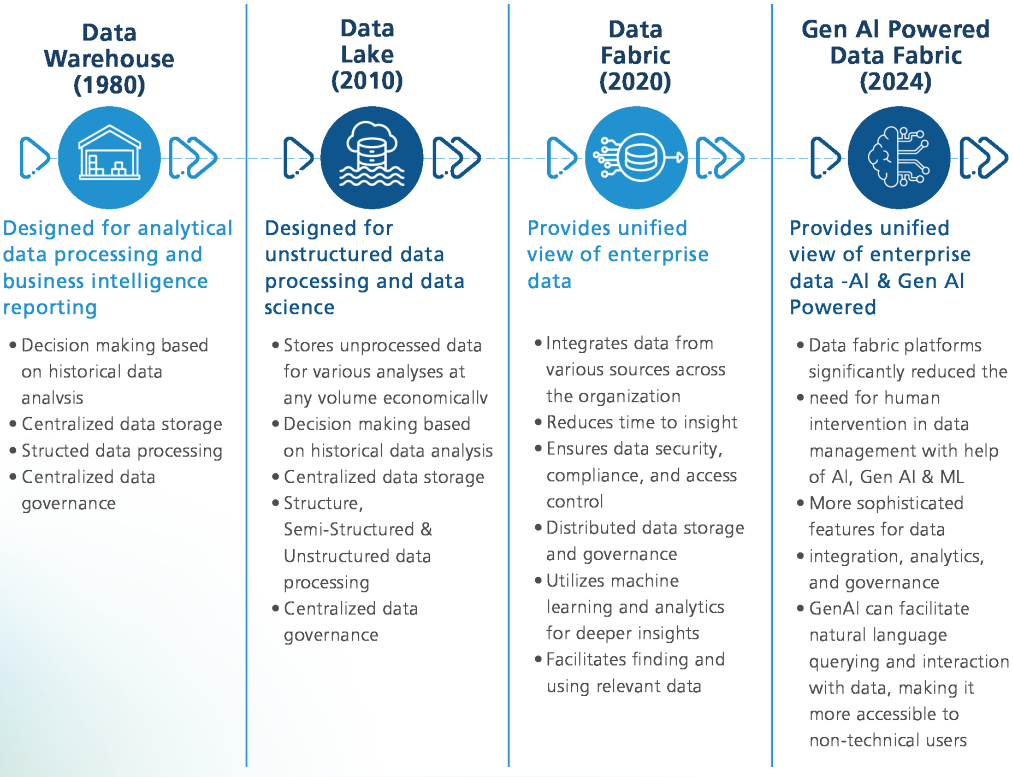


Figure 2: Evolution of GenAI-powered data fabric

GenAI-powered data fabric– logical architecture building blocks

The diagram below shows the functional components of a data fabric architecture.

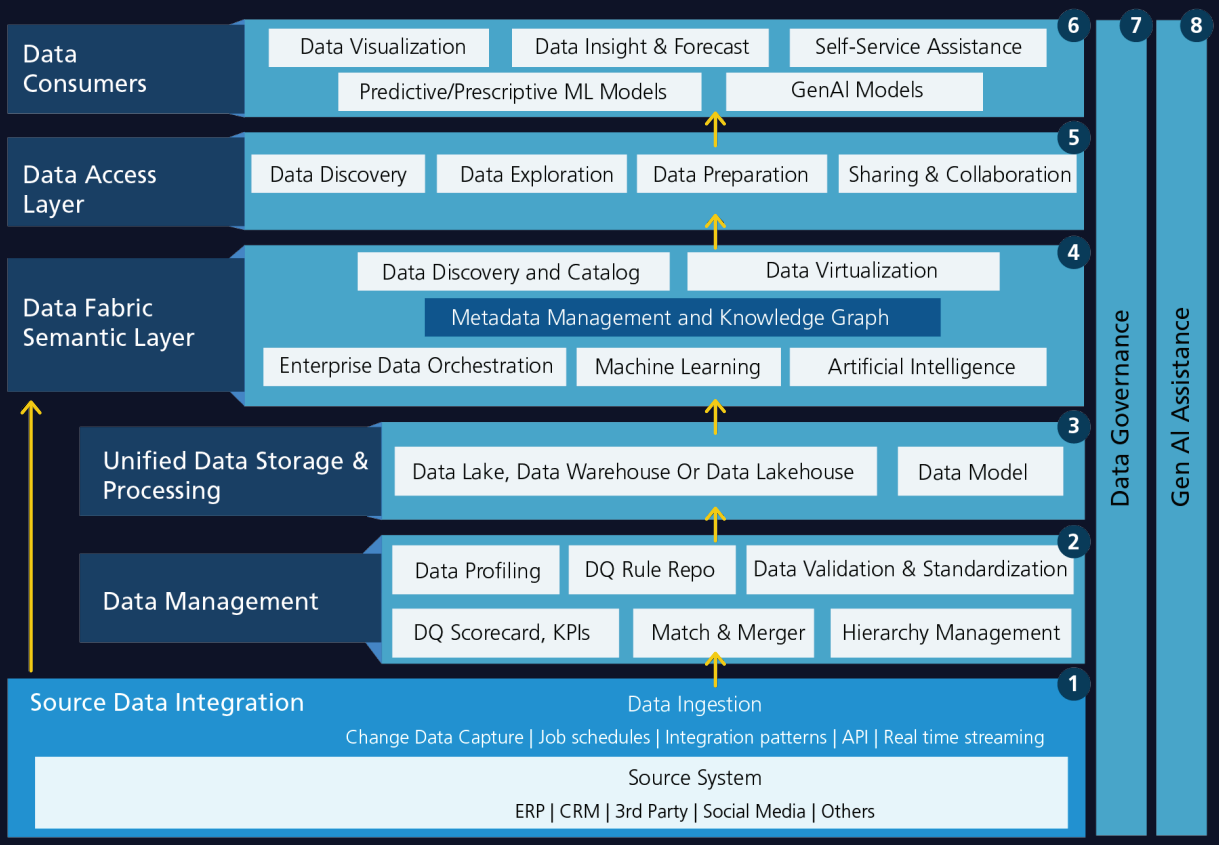


Figure 3: Data fabric solution building blocks

Source data integration

The solution leverages a metadata-driven framework to ensure a high level of automation, enabling seamless integration and assimilation of any kind of data, regardless of its source or rate, from on-site, cloud, or SaaS environments. It supports scalable streaming, batch processing, change data capture, and APIs, all enhanced by extensive and fast networking capabilities. This comprehensive approach ensures efficient and reliable data management across diverse platforms.

GenAI assistance

- Automatically maps data from different sources to target reducing effort and errors.
- Identifies and corrects data inconsistencies, missing values, and duplicates.
- Implements advanced encryption and access control mechanisms.
- Designs and manages complex data workflows.
- Automates repetitive tasks and improves efficiency, lowering costs.

Data management

The engine, automated by AI/ML algorithms supported by metadata, unifies, cleans up, and standardizes enterprise data that is shared but dispersed over different domains. By combining common and duplicate data into a "golden master," it ensures that this consolidated data is accessible for data fabric, enhancing data quality and accessibility across the organization.

GenAI assistance

- GenAI generates summaries of data profiling reports.
- Provides predictive insights based on root cause analysis and data lineage tracking.
- Recommends data validation rules with business context to speed up validation.
- Detects inconsistencies, duplicates, outliers, and missing data to help adopt data standards.
- Allows stewards and business users to interact with data quality tools and explore data assets using natural language.

Unified data storage and processing

Data persistence is primarily required to enhance performance, while also provisioning direct access to source data. Appropriate data management technologies are employed to store both structured and unstructured data. This approach utilizes data lakes, warehouses, and other analytical data storage solutions to ensure efficient and effective data management.

GenAI assistance

- GenAI aids strategic planning and efficient storage use, recommending timely maintenance to prevent system failures.
- Simplifies search and management capabilities.
- Uses large language models (LLMs) to define relationships between entities and attributes in structured and unstructured data.
- Analyzes query patterns and suggests optimizations like indexing, partitioning, and caching to enhance query performance.

Data fabric semantic layer

ML/AI builds a knowledge graph to record connections between data and business processes. A data catalog offers a semantically searchable repository of this metadata, which includes data lineage and data profiling outcomes. Enterprise data orchestration synchronizes data delivery and these semantic support procedures, ensuring efficient and coherent data management across the organization.

GenAI assistance

- Aids strategic planning and efficient storage capacity utilization to recommend timely maintenance, preventing system failures or performance issues.
- Leverages large language models (LLMs) to provide recommendations for defining relationships between entities and attributes from structured and unstructured data.
- Analyzes query patterns and suggests optimizations like indexing, partitioning, and caching to maximize query performance.

Data access layer

A knowledge graph adds depth to data relationships, enabling faster and easier decision-making. Data preparation offers a straightforward interface for gathering, merging, structuring, and organizing data. Additionally, a data marketplace provides a "Netflix of data" experience, making it easier to locate and provide data.

GenAI assistance

- Scans large datasets to identify patterns, trends, and anomalies
- Enhances collaborative platforms by providing real-time suggestions, summarizing discussions, and tracking changes.
- Ensures data sharing complies with relevant regulations and security protocols, protecting sensitive information.
- Assists in transforming data from one format to another, making it easier to integrate different datasets.

Data fabric semantic layer

Leveraging metadata to automate data governance and compliance protection, the system allows for policy enforcements at both local and global levels. Its active data governance (DG) layer facilitates automated data ingestion and self-service, ensuring compliance without any concerns.

GenAI assistance

- Enhances data security and compliance by monitoring access, detecting anomalies, and ensuring adherence to regulations.
- Facilitates metadata management by automatically generating and updating metadata, improving data traceability and understanding.
- Supports policy enforcement by automating the application of data governance policies and procedures across the organization.

Data consumers

In a data fabric architecture, data consumers benefit from seamless access to integrated data, enabling efficient decision-making and advanced analytics. Business analysts, data scientists, executives, and automated systems leverage the unified data environment to perform tasks ranging from generating reports to deploying machine learning models. The data fabric supports self-service capabilities, empowering users across the organization to independently access and analyze data. This democratization

of data fosters a data-driven culture, enhancing operational efficiency and innovation. The architecture ensures compliance and security by leveraging metadata to automate governance protocols, providing a robust framework for data consumption.

GenAI assistance

- Automates insights and visualizations, generating charts, graphs, and dashboards to help visualize data and identify trends.
- Uses historical data to forecast future trends and recommend optimal actions through predictive and prescriptive analytics.
- Allows users to interact with data using natural language queries and receive relevant insights via natural language processing (NLP).

Data fabric logical architecture–technology overlay

Functional Area	Market Leaders
Data integration	Azure Data Factory, aws Glue, Cloud DataFlow, Informatica, snowflake, kafka
Data management and governance	Microsoft Purview, aws Glue, Informatica, Collibra, Alation
Storage and processing	Azure Synapse Analytics, amazon REDSHIFT, Google Big Query, snowflake, databricks
Knowledge graph	neo4j, Amazon Neptune, ArangoDB, OrientDB, TigerGraph
Data visualization	Power BI, +ableau, Looker, amazon QuickSight, Qlik, sisense
Data virtualization	Azure Synapse Analytics, Amazon Redshift Spectrum, Google Big Query, denodo, dremio
Advance analytics	Azure Machine Learning, Amazon SageMaker, vertex.ai, IBM Watson, TensorFlow

Table 3: Data fabric logical architecture technology overlay

Knowledge graph– the core component of GenAI-powered data fabric

Metadata refers to data that provides descriptions about other data. It serves as the bedrock for constructing a data information network. Effective implementation of a data fabric depends on understanding the context of data.

A knowledge graph facilitates joint analysis and interpretation of data and metadata, ensuring clarity and preserving information integrity. By doing so, it establishes a robust framework for uncovering hidden relationships and latent knowledge essential for automating data integration, unification, analytics, and sharing. Leveraging formal semantics, it represents real-world facts and employs automated reasoning to extract additional insights from these facts. Leading providers of knowledge graph are IBM Watson Knowledge Studio, Microsoft Azure Knowledge Graph, AWS Neptune, Neo4j, TigerGraph, etc.

The steps to build knowledge graph are:

01

Data integration

Support a variety of data delivery methods including extract, transform and load (ETL), streaming, messaging and data virtualization, etc.

02

Analyze metadata

Allow data fabric to identify, collect and analyze all forms of metadata.

03

Passive to active metadata

Activate metadata to gather key statistics, business relationships and power AI/ML algorithms.

04

Knowledge graph

Empower data and analytics leaders to extract business value by enriching data with easy-to-understand and intuitive semantics.

Applying graph analytics on the collected metadata helps to train AI/ML models meant for streamlining data integration and management tasks. This empowers data and analytics teams to create knowledge graphs that uncovers the relationships between organization's diverse data assets and their users.

The types of metadata that data fabric deals with are:

Technical metadata

Technical metadata encompasses essential attributes that systems rely on for data processing. These attributes include format, type, length, and storage location.

Business metadata

Business metadata provides context and insights into the data used within an organization. It describes the characteristics, origins, and relationships of the data, enabling users to understand and interpret it accurately.

Operational metadata

Operational metadata offers insights into data usage and the associated processes. It encompasses details such as event dates and times, sourced from various logs, including audit logs and execution logs.

Social metadata

Social metadata enriches data by providing contextual information such as timestamps, geographic locations, tags, and user engagement metrics. This augmentation enhances the sharing and comprehension of content across social networks.

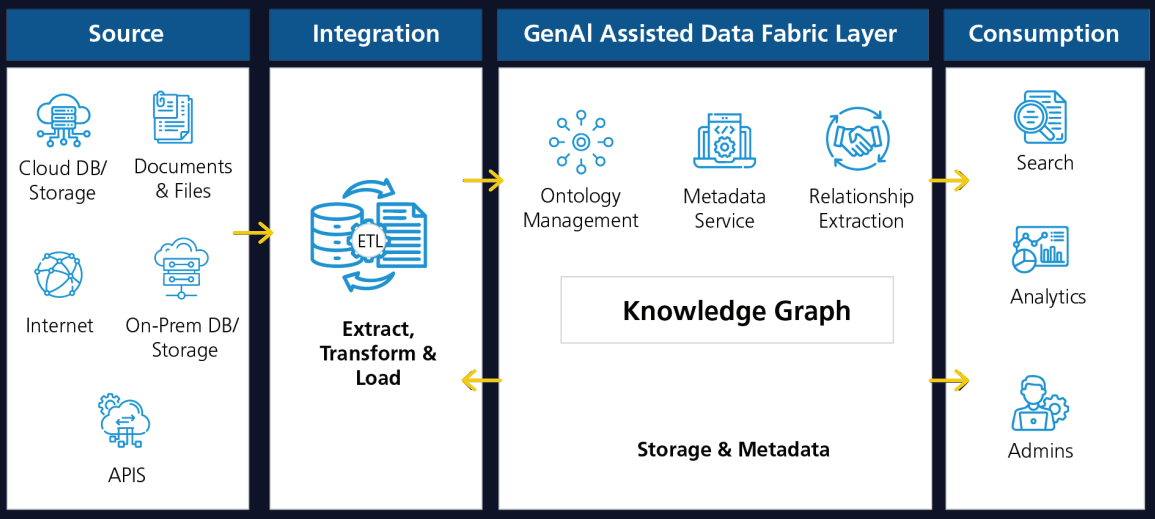


Figure 4: Knowledge graph implementation in data fabric

Features of knowledge graph

- Leverages formal semantics to represent real-world facts and relationships
- Facilitates joint analysis and interpretation of data and metadata
- Ontology management in a knowledge graph within a data fabric enhances data integration, consistency, access, governance, and analysis
- Employs automated reasoning to extract additional insights
- Establishes a robust framework for:
    1. Uncovering hidden relationships
    2. Revealing latent knowledge
    3. Automating data integration, unification, analytics, and sharing

Sample use cases leveraging data fabric architecture

Data fabric caters to varirety of use cases across industries, some of which are listed below.

Across industries	• Real-time analytics • Automated data management • Instant combination of broad spectrum of data • Deeper and quicker data insights • Enhanced product quality and operational efficiency • Implementation of GenAI use cases to identify patterns, instant recommendation and actionable insights
Hospitality industry	Personalized Marketing
Healthcare industry	Enhanced integration of healthcare data for better patient outcomes
Energy industry	Energy Optimization and Distribution
Banking industry	Fraud Detection and Risk Prevention
Manufacturing industry	Supply Chain Efficiency

There are many more use cases that can be addressed by data fabric and not covered in this table. This table highlights key use cases only across different industries.

Data fabric– benefits and considerations

Data fabric aims to resolve the complexities of data management by providing a unified architecture that integrates and manages data across various environments, ensuring seamless access, governance, and security.

The key benefits of data fabric are:

01

Automated data integration

Data fabric automates data management tasks through the application of artificial intelligence and machine learning.

02

Effective data discovery

Data catalog facilitates efficient search and rapid access to essential enterprise data for consumers.

03

Simplifies data access

Data fabric abstracts the intricacies of enterprise data, allowing users to interact with it without needing to comprehend its specific type or location.

However, there are certain constraints of data fabric, such as the complexity of implementation, the need for significant upfront investment, and potential challenges in integrating with legacy systems.

01

Complex and costly

Implementation of data fabric requires multiple data management techniques, contributing to its complexity and cost.

02

Multiple tools and technology

The implementation necessitates integrating multiple disparate tools and technologies.

03

Data security

To ensure secure data access, robust privacy and security protocols must be implemented to prevent breaches.

Data fabric is ideal for scenarios where one needs to enable real-time analytics, automated data management, instantly combine a broad spectrum of data, with deeper and quicker data insights.

Studies by Gartner, IBM and MicroStrategy reveal that data fabric can cut down data access & preparation time by 40-60% and reduce operational cost by 20-30%.

Data fabric may not be best suited for the organizations with strict regulations, as complexity and compliance requirements can cause significant challenges in development. For organizations with minimal data integration requirements or who need immediate benefits from smaller data sets, data fabric can be an over-engineered solution to start with. Lack of skilled resources can further hinder the effective implementation and management of the data fabric.

FABMAT– LTIMindtree data fabric maturity framework

A phased approach is ideal for implementation of data fabric. Consider the big picture but start small focusing on the following steps:

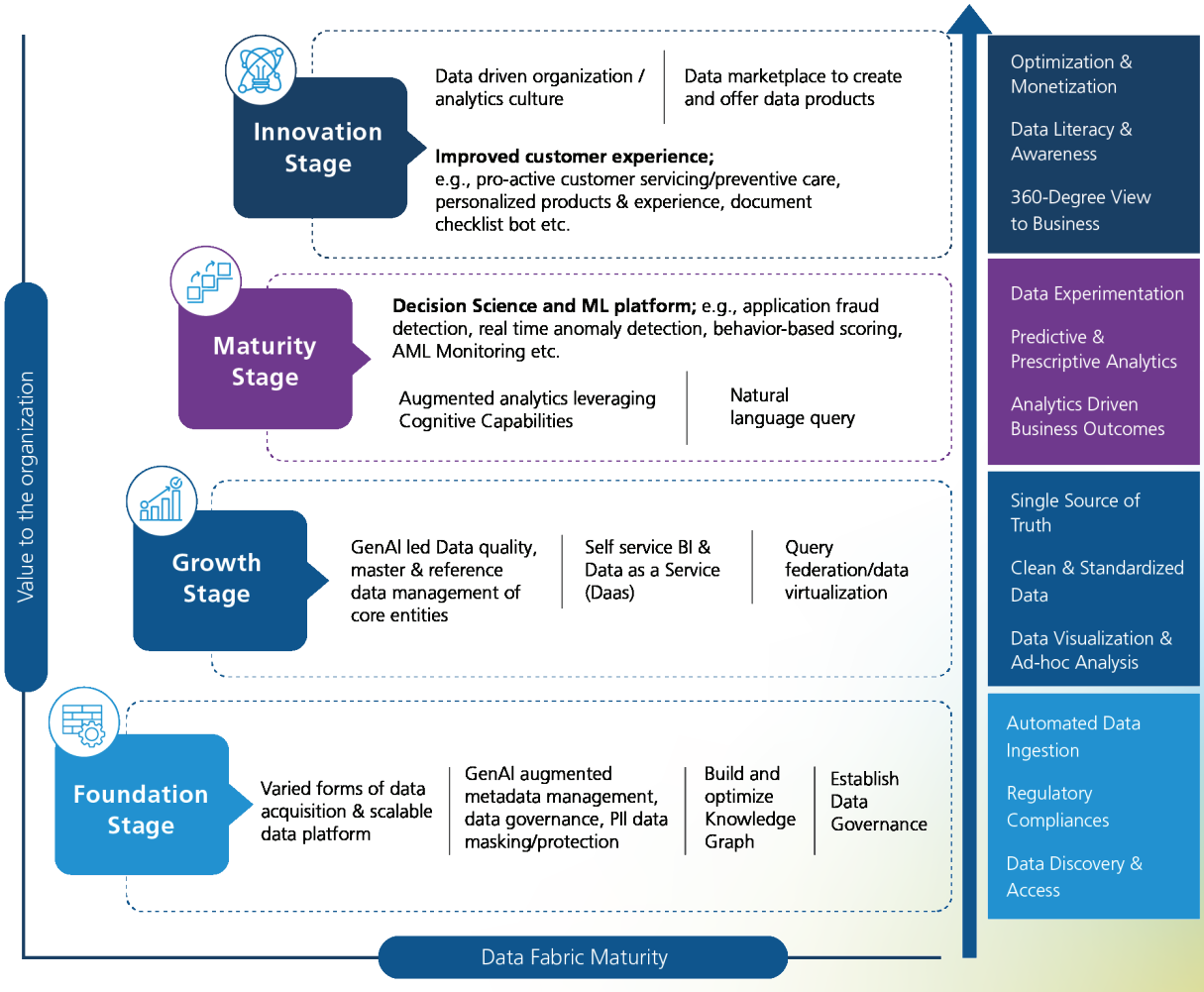
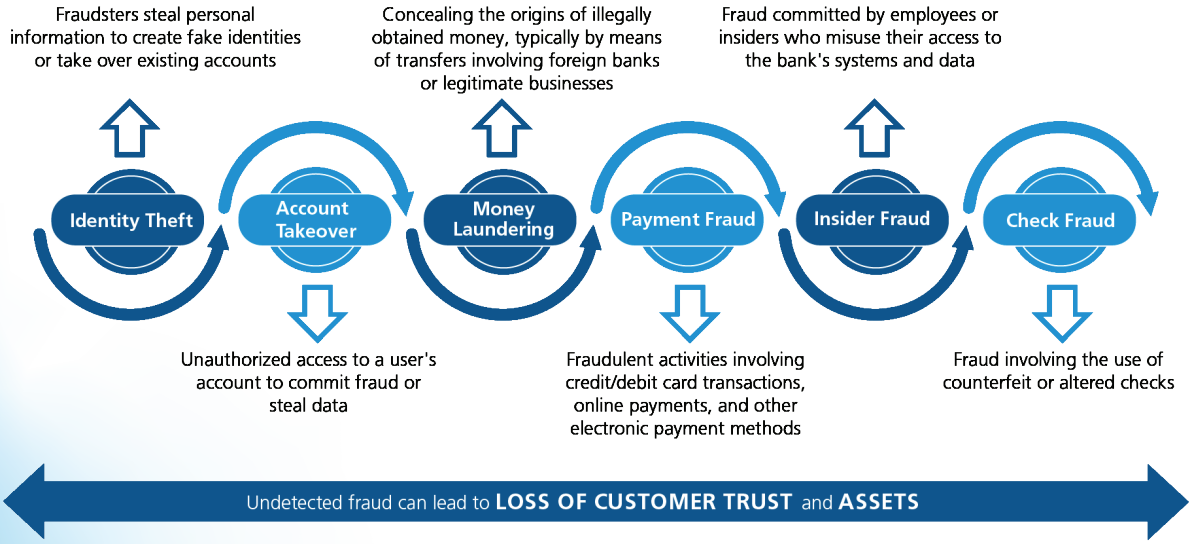


Figure 5: LTIMindtree's FABMAT maturity model framework

Use case illustration– fraud detection and risk prevention

This sample use case highlights current challenges of an organization that lacks matured fraud detection and analytics process

Current challenge: Complaints from customers suggesting a likelihood of fraudulent transactions.



The steps involved in addressing the challenge through a data fabric solution is given below. Each step leads to incremental build of data fabric components starting from automated data ingestion to data marketplace for seamless data exchange.

Assessment and strategy

- Identify source systems for transactions, customer profiles, social media, government databases, historical fraud cases, etc.
- Plan for capabilities, execution roadmap and estimates for fraud detection analytics.

Data management

- Establish policies and procedures for master data management.
- Define & enforce policies for data quality, privacy, and security to ensure compliance and protect sensitive information.

Source data ingestion and knowledge graph

- Identify relationships between varied data assets like transaction amount, time, duration, sender/receiver details, IP addresses, etc.
- Model knowledge graph to establish links between these assets.
- Output to be stored in a graph database.

Date marketplace

- Create a centralized platform for data access and sharing.
- Maintain a catalogue of available data sets, including descriptions, sources, and usage guideline.
- Provide training for effective use.

Insight generation

- Establish graph analytics and GenAI models to identify patterns and anomalies indicative of fraud.
- Implement visualization tools for effective analysis of entities and linkages.

How to implement data fabric

A phased approach is ideal for implementation of data fabric. Consider the big picture but start small focusing on the following steps:

01

Understand

Deconstruct the transformation landscape, including objectives, key affected business units, stakeholders, processes, and technologies/data.

02

Ideate

Map and define potential solutions and priorities based on identified capability needs. Highlight key sequencing, dependencies, and KPIs.

03

Accelerate

Recommend a roadmap to navigate a successful rollout with technology recommendations and budgetary estimates.

04

Foundation

Set up infrastructure, create a sandbox for development and set up an access policy.

05

Implement

Phased implementation of data fabric capabilities.

Glossary

Abbreviation	Abbreviation
AI	Artificial Intelligence
ML	Machine Learning
IoT	Internet of Things
GenAI	Generative Artificial Intelligent

References

How Data Fabric Can Optimize Data Delivery, Insights by Gartner,

https://www.gartner.com/en/data-analytics/topics/data-fabric

Introduction to data fabric, Vivek Pratap, July 11, 2023,

https://developer.ibm.com/articles/introduction-to-data-fabric/

A Guide to Data Fabric, Insights by Teradata,

https://www.teradata.com/insights/data-platform/data-fabric-guide

Data Fabric Smart Data Engineering, Operations, and Orchestration, Dave Wells, September 30, 2019,

https://www.eckerson.com/register?content=data-fabric

Data: A double-edged sword, Paul H. Silverglate, and David Jarvis, Deloitte Insights, September 15, 2022,

https://www2.deloitte.com/us/en/insights/industry/technology/challenges-in-data-management.html

Master data management: The key to getting more from your data, Aziz Shaikh, Jorge Machado, Kayvaun Rowshankish, Rachit Saxena, Rajat Jain, Holger Harreis, Insights by McKinsey Digital,

https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/master-data-management-the-key-to-getting-more-from-your-data

What is Data Fabric and Why are Businesses Using It?, MicroStrategy, September 11, 2024,

https://www.microstrategy.com/blog/what-is-data-fabric-and-why-are-businesses-using-it

Cost Optimization Is Crucial for Modern Data Management Programs, Gartner Research, June 22, 2020,

https://www.gartner.com/en/documents/3986583

How a data fabric overcomes data sprawls to reduce time to insights, Shashank Sabhlok, IBM, April 28, 2023,

https://www.ibm.com/blog/how-a-data-fabric-overcomes-data-sprawls-to-reduce-time-to-insights/

About authors

Shibaji Mitra

Data Architect, Data Architecture and Strategy, LTIMindtree

Shibaji is a member of the Data Architecture and Strategy team. He has more than 20 years of experience in DW BI and has worked on technology consulting and end-to-end implementation of data warehousing and modernization projects across multiple industries.

Sanjay Biswas

Data Architect, Data Architecture and Strategy, LTIMindtree

Sanjay has over 14 years of experience in data analytics strategy, data management consulting, end-to-end implementation, and defining data visualization strategies.

Kallol Dutta

Data Strategy and Data Integration CoE Head in LTIMindtree

Kallol has over 25 years of industry experience and has led multiple strategy consulting projects and implementation engagements. He has authored and published point of views and whitepapers across different data concepts and technologies.