European Journal of Computer Science and Information Technology,13(37),97-,108, 2025

Print ISSN: 2054-0957 (Print)

Online ISSN: 2054-0965 (Online)

Website: https://www.eajournals.org/

Publication of the European Centre for Research Training and Development -UK

AI-Driven Data Mesh with Generative AI for Enterprise Analytics

Lingareddy Alva

IT Spin Inc, USA

doi: https://doi.org/10.37745/ejcsit.2013/vol13n3797108 Published June 07, 2025

Citation: Alva L. (2025) AI-Driven Data Mesh with Generative AI for Enterprise Analytics, *European Journal of Computer Science and Information Technology*,13(37),97-108

Abstract: *This article explores the transformative integration of generative AI capabilities with Data Mesh architecture to revolutionize enterprise analytics. Beginning with examining traditional data architectures' limitations, the discussion highlights how centralized proceeds towards creating bottlenecks that impede innovation and time-to-insight. The Data Mesh paradigm is presented as a fundamental shift that decentralizes data ownership while maintaining federated governance. The integration of generative AI within this framework enables natural language interfaces, synthetic data generation, automated documentation, and intelligent insight creation. Implementation strategies using Databricks platform capabilities demonstrate how organizations can balance domain autonomy with enterprise interoperability. The architecture delivers enhanced analytics through AutoML-powered data quality with generative explanations and event-driven processing that enables real-time, predictive intelligence. Together, these capabilities create a self-improving ecosystem that democratizes data access while ensuring governance, ultimately enabling organizations to move beyond traditional reporting toward autonomous, data-driven operations with cross-domain collaboration.*

Keywords: generative AI, data mesh, domain-driven architecture, real-time analytics, federated governance

INTRODUCTION

The Evolution of Data Architecture in the Generative AI Era

In today's data-driven business landscape, traditional centralized data architectures have become significant bottlenecks, limiting organizational agility, scalability, and innovation. Enterprises are increasingly recognizing the limitations of these centralized approaches as they struggle with growing data volumes and complexity [1]. The centralization of decision-making creates inherent challenges in responsiveness, particularly as organizations attempt to derive value from rapidly expanding and diverse data sources across operational domains.

The convergence of Data Mesh principles and generative AI capabilities represents a paradigm shift in enterprise data management. By embracing decentralized decision-making approaches and domain-oriented ownership models, organizations can address the fundamental limitations of centralized architectures [1]. This decentralization aligns with the growing need for domain-specific expertise in data interpretation and utilization, creating what researchers describe as "democratized data governance" across the enterprise.

AI-driven data mesh addresses enterprise analytics challenges by fundamentally restructuring how data is managed, processed, and consumed. The integration of generative AI into this architecture introduces transformative capabilities across numerous business functions, from marketing and sales to customer operations and software development [2]. This architectural approach enables organizations to deploy AI capabilities close to where data originates, enhancing context-specific analytics while maintaining enterprise-wide governance.

The key components—domain ownership, generative AI, and event-driven analytics—work in concert to create exponential value. Research suggests that generative AI has the potential to add trillions in value to the global economy, with significant impact across industries [2]. When implemented within a domain-oriented data architecture, these technologies enable faster innovation cycles, higher-quality insights, and more responsive decision-making capabilities that are increasingly essential in today's dynamic business environment.

The Limitations of Traditional Data Architectures

Traditional centralized data architectures increasingly struggle to meet the demands of modern enterprise environments, particularly as organizations adopt advanced analytics and generative AI capabilities. Centralized data platforms face significant scaling challenges that impact their performance and viability for modern workloads. As data volumes grow exponentially, the architectural constraints become more pronounced, with performance degradation occurring at critical thresholds where computational demands exceed available resources [3]. This scaling limitation mirrors challenges observed in high-performance computing environments, where interconnect bottlenecks and resource contention become increasingly problematic as system size increases.

Organizational silos and decision-making bottlenecks represent another significant limitation of centralized approaches. In construction and engineering contexts, research has shown that centralized information management leads to coordination challenges and communication barriers between project stakeholders [3]. These barriers are particularly problematic when implementing advanced technologies that require cross-functional collaboration and rapid decision-making. The centralization of data expertise creates

European Journal of Computer Science and Information Technology,13(37),97-,108, 2025
Print ISSN: 2054-0957 (Print)
Online ISSN: 2054-0965 (Online)
Website: https://www.eajournals.org/
Publication of the European Centre for Research Training and Development -UK
dependencies that slow innovation cycles and prevent domain experts from directly leveraging data assets without technical intermediaries.

The disconnect between data producers and consumers creates what researchers refer to as the "translation gap," where business context is lost in transmission between teams. This misalignment becomes particularly problematic for generative AI applications, which require close collaboration between domain experts and technical teams to ensure models are properly trained and validated. The challenges resemble those documented in high-performance computing environments, where users and system administrators often have conflicting priorities and understanding of requirements [4].

Technical limitations further compound these challenges. Traditional architectures face concurrency constraints, high maintenance overhead, and pipeline fragility that limit their effectiveness for modern workloads. As observed in high-performance computing research, system usability and reliability decrease as complexity increases, creating significant operational challenges for maintaining performance at scale [4]. Similar patterns emerge in enterprise data environments, where the complexity of centralized systems requires specialized expertise that creates bottlenecks in development and optimization.

Traditional architectures struggle particularly with generative AI workloads due to their resource-intensive nature. These applications require massive parallel processing capabilities, low-latency data access, and flexible scaling patterns that centralized architectures cannot efficiently provide. The requirements mirror those documented in scientific computing environments, where applications demand both high throughput and low latency across diverse computational patterns [4].

The need for a fundamentally different approach to enterprise data management has never been more apparent. Organizations require architectures that can distribute processing closer to data sources, enable domain-specific optimization, and provide flexible scaling models to support diverse analytical workloads including generative AI applications. This recognition is driving a paradigm shift toward domain-oriented data architectures that better align with both technical requirements and organizational structures.

Table 1: Limitations of Traditional Data Architectures vs. Data Mesh Approach [3, 4]

Aspect	Traditional Centralized Architecture	Data Mesh Approach
Organizational Structure	Centralized data team	Domain-oriented teams
Decision-making	Concentrated among data gatekeepers	Distributed to domain experts
Request Processing	6-8 week delays for analytics requests	Self-service capabilities
Scalability	Performance degradation with growing data volume	Horizontal scaling across domains
Time-to-market	Longer cycles due to dependencies	Faster through domain autonomy
Business Context	Lost in translation between teams	Preserved through domain ownership
Technical Complexity	6-8 processing layers	Simplified domain-specific pipelines
Maintenance	High overhead for existing pipelines	Distributed responsibility

Data Mesh: A Paradigm Shift in Enterprise Data Management

The Data Mesh paradigm, first introduced by Ammara Gafoor et al., represents a fundamental reimagining of enterprise data architecture through domain-driven ownership and data as a product principles. This architectural approach addresses the scalability challenges of centralized data platforms by redistributing ownership to domain teams who understand their data's context and use cases [5]. As research into large language model training demonstrates, decentralized approaches can significantly improve both efficiency and effectiveness when managing complex data ecosystems, particularly when implementing federated learning strategies that preserve domain autonomy while enabling collective improvement.

Data Mesh aligns naturally with microservices architectures and agile methodologies that have transformed application development. Just as microservices decompose monolithic applications into domain-aligned services, Data Mesh decomposes monolithic data platforms into domain-aligned data products. This architectural symmetry creates what researchers describe as "socio-technical congruence," where organizational structures, technical architectures, and data management approaches reinforce rather than contradict each other [5]. The alignment extends to machine learning operations, where domain-specific models can be trained, validated, and deployed using similar patterns to those established for microservices. Decentralized governance with federated computational models forms the backbone of successful Data Mesh implementations. As organizations begin their Data Mesh journey, they establish federated governance teams that define cross-cutting standards while respecting domain autonomy [6]. This balanced approach creates what practitioners call "minimum viable centralization"—just enough standardization to ensure interoperability without stifling innovation or domain-specific optimization. Successful implementations start with a clear data sharing agreement that outlines the responsibilities of both domain teams and central governance functions.

Self-serve data platforms for domain autonomy represent a critical enabler for Data Mesh adoption. Ammara Gafoor et al., implementation experience demonstrates that these platforms must reduce the cognitive load on domain teams while still enabling them to create high-quality data products [6]. The platform capabilities should evolve incrementally, starting with foundational services like discovery, access control, and lineage tracking before expanding to more advanced features such as quality monitoring and automated documentation.

Data Mesh creates an ideal foundation for generative AI integration by addressing fundamental data access, quality, and context challenges. The architectural approach aligns with best practices for distributed data processing required by large language models, which benefit from domain-specific training data with clear provenance and context [5]. By treating data as a product with well-defined interfaces and quality guarantees, domain teams create the essential building blocks for effective generative AI implementation. Cross-domain data collaboration potential represents perhaps the most significant long-term value proposition of Data Mesh architectures. Organizations implementing this approach report improved discovery and reuse of data products across previously siloed domains [6]. The key enabler is what practitioners call "data product thinking"—designing domain datasets with the needs of potential consumers in mind, including comprehensive documentation, clear quality metrics, and standardized access patterns that make cross-domain collaboration both possible and practical.

Generative AI Integration in Data Mesh Architecture

The integration of generative AI capabilities into Data Mesh architectures represents a transformative enhancement to enterprise data management. Natural language interfaces for data discovery and interaction significantly improve data accessibility across organizational domains. Research on AI augmentation in enterprise systems indicates that conversational interfaces reduce the technical barriers that traditionally limit data access, enabling domain experts to interact with data products using natural language rather than specialized query languages [7]. These interfaces transform how business users engage with analytics, shifting from predefined reports to dynamic, contextual interactions that adapt to specific business questions.

Synthetic data generation for testing, development, and augmentation addresses critical challenges in data availability and privacy within domain-oriented architectures. By generating artificial yet statistically representative datasets, domain teams can accelerate development cycles while maintaining compliance with data privacy regulations. This capability proves particularly valuable in regulated industries where data access constraints often create bottlenecks in analysis and AI model development [7]. The synthetic data approach enables parallel development across domains without compromising sensitive information, fundamentally changing how organizations balance innovation with privacy requirements.

Automated documentation and metadata enrichment leverages generative AI to create comprehensive context around domain data products. As research into intelligent systems demonstrates, automated documentation significantly enhances knowledge transfer between different stakeholder groups, addressing what researchers describe as "information asymmetry" between data producers and consumers [8]. This capability proves particularly valuable in Data Mesh environments where domain teams must communicate data product characteristics to potential consumers across organizational boundaries.

AI-powered insight generation and narrative reporting transform raw data into accessible intelligence for business users. By automatically identifying patterns, anomalies, and trends, then expressing them in business-relevant language, these capabilities bridge the gap between technical analyses and business decision-making [8]. The integration with domain-specific data products ensures that generated insights reflect appropriate business context and terminology.

Semantic data discovery through vector embeddings enables organizations to move beyond keyword-based search toward true context and intent understanding. By representing data assets in multi-dimensional semantic spaces, these systems can understand relationships and similarities that traditional metadata approaches cannot capture [7]. This capability enhances cross-domain data discovery, a critical enabler for value creation in Data Mesh architectures.

Cross-domain knowledge graphs enhanced by generative models create unprecedented opportunities for connecting disparate data domains. Research on cybersecurity intelligence systems demonstrates how knowledge graphs can identify complex relationships between entities across domain boundaries [8]. When enhanced with generative capabilities, these knowledge structures automatically expand to incorporate new relationships and entities, creating an evolving map of enterprise data that transcends traditional domain boundaries.

The democratization of data access through conversational AI represents perhaps the most transformative aspect of generative AI integration, enabling all stakeholders to leverage data assets regardless of technical expertise, fundamentally reshaping how organizations derive value from their data investments.

Table 2: Generative AI Integration Points in Data Mesh Architecture [7, 8]

Integration Point	Capability	Business Impact
Data Discovery	Natural language interfaces	Enhanced accessibility for non-technical users
Data Generation	Synthetic data creation	Accelerated development and testing cycles
Documentation	Automated context generation	Improved data product adoption and usage
Insight Generation	Narrative reporting	Better comprehension and faster decisions
Data Discovery	Vector-based semantic search	More relevant and context-aware results
Knowledge Management	Cross-domain knowledge graphs	Identification of previously hidden relationships
User Interaction	Conversational interfaces	Democratized access to analytics capabilities

Implementing an AI-Driven Data Mesh with Databricks

Implementing an AI-driven Data Mesh requires sophisticated architectural components that balance domain autonomy with enterprise interoperability. A well-designed enterprise generative AI architecture consists of multiple layers, including data ingestion, preprocessing, model selection, prompt engineering, and output handling [9]. When implemented within a Data Mesh framework using Databricks, these components work in concert to create a system that maintains domain-specific context while enabling enterprise-wide capabilities. This architectural approach addresses what enterprise architects identify as the "distributed governance challenge"—maintaining coherence across autonomous domains without creating centralized bottlenecks.

Databricks Unity Catalog integration with Large Language Models (LLMs) represents a transformative enhancement for data discovery within the Data Mesh paradigm. By connecting metadata repositories with natural language interfaces, organizations enable semantic search capabilities that significantly improve how users find and interact with data assets. The discovery experience leverages technologies like vector embeddings and retrieval-augmented generation to ensure responses are grounded in enterprise data rather than purely generative [9]. This factual grounding is particularly critical in enterprise environments where accuracy and provenance are essential requirements.

Delta Sharing capabilities provide the foundation for secure cross-domain collaboration, addressing what has historically been a significant challenge in decentralized architectures. IEEE research on data trustworthiness in collaborative smart systems demonstrates the importance of maintaining data quality, provenance, and security across distributed environments [10]. Delta Sharing implements these principles through secure access protocols and data integrity verification, enabling domains to share data products while maintaining appropriate controls and audit capabilities.

Vector search implementation for semantic data discovery fundamentally changes how users interact with domain data products. Unlike traditional metadata-based approaches, vector search enables understanding of contextual relationships and conceptual similarities that keyword approaches cannot capture. This capability aligns with enterprise generative AI architecture patterns that emphasize semantic layers between raw data and consumption interfaces [9]. By representing data assets as vector embeddings, organizations enable "similarity-based discovery" that dramatically improves user experiences, particularly for domain experts without deep technical knowledge.

Foundation model integration for synthetic data generation addresses critical data challenges across the enterprise. These capabilities allow domain teams to generate realistic test data that maintains statistical properties and relationships without exposing sensitive information. The approach aligns with IEEE research on trusted data management, which emphasizes the need for privacy-preserving mechanisms in collaborative environments [10]. By integrating foundation models within domain-specific contexts, organizations enable "privacy-by-design" data practices essential for regulatory compliance.

Implementation patterns for domain-specific generative AI applications leverage fine-tuning and retrieval augmentation to adapt general-purpose models for specialized domains. This approach creates a multi-tiered model architecture where base capabilities are enhanced with domain-specific knowledge and constraints [9]. Successful implementations carefully balance the computational efficiency of centralized models with the contextual specificity of domain-oriented adaptations.

Balancing domain autonomy with enterprise-wide AI governance represents perhaps the most significant implementation challenge. Research on data trustworthiness emphasizes the need for governance frameworks that establish clear responsibilities across organizational boundaries [10]. Effective implementations create what practitioners call "federated AI governance"—where enterprise standards for ethics, security, and compliance complement domain-specific policies for model selection and application-specific thresholds.

Enterprise Analytics Capabilities with Gen AI and Data Mesh

The integration of generative AI with Data Mesh creates unprecedented enterprise analytics capabilities that transform how organizations derive value from their data assets. AutoML-powered data quality with generative enhancement represents a fundamental shift in how organizations maintain data integrity at scale. Research on adaptive generative AI for streaming analytics indicates that these systems can dynamically adjust to changing data distributions, enabling what researchers describe as "continuous adaptation" essential for maintaining quality in real-time environments [11]. This adaptive approach becomes particularly valuable within domain-oriented architectures where data characteristics may evolve independently across different business contexts.

Anomaly detection with natural language explanations transforms how organizations identify and respond to data quality issues. By combining statistical detection methods with generative language capabilities, these systems not only flag outliers but provide contextual explanations that business users can understand. Missing value imputation through generative models significantly improves analytical accuracy by preserving distributional characteristics and relationships between variables. The generative approach demonstrates particular strength in handling complex data types like images, text, and time-series data where traditional methods often fall short [11]. Data drift detection with automated root cause analysis addresses the challenge of maintaining model relevance as underlying data patterns evolve, a critical capability for organizations implementing real-time analytics at scale.

Explainable quality metrics with narrative context transform technical indicators into business-relevant insights, bridging the gap between data quality measurements and their operational implications. The adaptive systems can learn from user feedback to continuously improve both detection accuracy and explanation quality, creating a self-improving cycle that enhances both technical performance and business utility [11].

Event-driven AI analytics with generative capabilities enables organizations to implement real-time predictive analytics that transform business operations. By processing events as they occur, these systems create what practitioners call "predictive processing" - where insights and actions are generated before business impact occurs [12]. Real-time insight generation with narrative summaries provides business users with immediately actionable intelligence, while predictive modeling with synthetic event generation enhances forecasting accuracy even with limited historical data.

Natural language alerts with context and recommendations transform notifications into actionable guidance, significantly improving response effectiveness. The event-based architecture enables scenario simulation for proactive decision support, allowing organizations to evaluate potential actions before implementation.

European Journal of Computer Science and Information Technology,13(37),97-,108, 2025
Print ISSN: 2054-0957 (Print)
Online ISSN: 2054-0965 (Online)
Website: https://www.eajournals.org/
Publication of the European Centre for Research Training and Development -UK

As noted in research on predictive analytics, these capabilities provide significant competitive advantages by shifting organizations from reactive to proactive operational models [12].

Continuous intelligence loops with feedback mechanisms create self-improving analytics systems by capturing outcome data and automatically incorporating it into future predictions. This approach addresses what industry leaders identify as the "analytics value gap" - where technical capabilities exist but fail to deliver business impact due to adoption or usability challenges. By closing this gap through natural language interfaces and continuous improvement, organizations can achieve what researchers describe as "pervasive analytics" - where data-driven decision-making becomes embedded throughout the enterprise rather than limited to specialized roles or functions [12].

Category	**Capability**	**Operational Benefits**
Data Quality	Anomaly detection with explanations	Faster resolution of quality issues
	Generative missing value imputation	Higher analytical accuracy with incomplete data
	Automated drift detection	Maintained model relevance over time
	Explainable quality metrics	Increased business stakeholder engagement
Real-time Analytics	Event-driven insight generation	Immediate response to changing conditions
	Predictive modeling with synthetic data	Improved forecasting for rare events
	Contextual alerts and recommendations	More effective first responses
	Scenario simulation	Better evaluation of potential actions
	Intelligence loops with feedback	Self-improving analytics over time

Table 3: Enterprise Analytics Capabilities Enabled by Gen AI and Data Mesh [11, 12]

CONCLUSION

The convergence of generative AI with Data Mesh architecture represents a pivotal evolution in enterprise data management, addressing fundamental challenges that have historically limited the business impact of analytics investments. By decentralizing data ownership to domain teams while enriching their capabilities with generative technologies, organizations create an ecosystem where data products become increasingly accessible, trustworthy, and valuable. This architectural paradigm eliminates the traditional trade-offs between governance and agility, centralization and domain autonomy, enabling both simultaneously through federated models and self-service platforms. Natural language interfaces and automated insight generation dramatically expand data utilization beyond technical specialists, creating truly democratized analytics where every business stakeholder can derive value from enterprise data assets. Event-driven processing with generative augmentation transforms decision-making from reactive to proactive, enabling organizations to anticipate and respond to opportunities and challenges in real-time. As enterprises continue navigating increasing data volumes and complexity, this architecture provides a sustainable framework for scaling analytics capabilities while continuously improving data quality, accessibility, and business relevance—ultimately transforming how organizations create value from their data investments.

REFERENCES

1. Leonidas Theodorakopoulos, Alexandra Theodoropoulou, and Constantinos Halkiopoulos, "Enhancing Decentralized Decision-Making with Big Data and Blockchain Technology: A Comprehensive Review," Appl. Sci., 2024. [Online]. Available: https://www.researchgate.net/profile/Alexandra-Theodoropoulou-3/publication/383044846_Enhancing_Decentralized_Decision-Making_with_Big_Data_and_Blockchain_Technology_A_Comprehensive_Review/links/66ba054851aa0775f27aa388/Enhancing-Decentralized-Decision-Making-with-Big-Data-and-Blockchain-Technology-A-Comprehensive-Review.pdf
2. Michael Chui et al., "The economic potential of generative AI: The next productivity frontier," McKinsey Digital, 2023. [Online]. Available: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier
3. Suhyung Jang et al., "Generative AI in architectural design: Application, data, and evaluation methods, "Automation in Construction, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0926580525002146
4. Al Geist and Daniel A Reed, "A survey of high-performance computing scaling challenges," The International Journal of High Performance Computing Applications, 2015. [Online]. Available: https://www.researchgate.net/publication/281605218_A_survey_of_high-performance_computing_scaling_challenges
5. Haoyuan Li and Salman Toor, "Empowering Data Mesh with Federated Learning," arXiv:2403.17878v2 [cs.LG], 2024. [Online]. Available: https://arxiv.org/pdf/2403.17878

6. Ammara Gafoor et al., "Data Mesh in practice: Getting off to the right start (Part I)," ThoughtWorks, 2022. [Online]. Available: https://www.thoughtworks.com/en-in/insights/articles/data-mesh-in-practice-getting-off-to-the-right-start
7. Venkat Sharma Gaddala, "The Future of Enterprise Data is Conversational," Iconic Research And Engineering Journals, 2023. [Online]. Available: https://www.irejournals.com/formatedpaper/17050981.pdf
8. Varun Gupta, "An Empirical Evaluation of a Generative Artificial Intelligence Technology Adoption Model from Entrepreneurs' Perspectives," Systems, 2024. [Online]. Available: https://www.mdpi.com/2079-8954/12/3/103
9. Akash Takyar, "Generative AI architecture for enterprises: Development frameworks, tools, implementation, and future trends," LeewayHertz, 2025. [Online]. Available: https://www.leewayhertz.com/generative-ai-architecture-for-enterprises/#Understanding-the-enterprise-generative-AI-architecture
10. Yizhong Liu et al., "Secure and Scalable Cross-Domain Data Sharing in Zero-Trust Cloud-Edge-End Environment Based on Sharding Blockchain," IEEE Transactions on Dependable and Secure Computing, 2023. [Online]. Available: http://ieeexplore.ieee.org/document/10246351
11. Raghavender Maddali, "Adaptive Generative AI for Real-Time Data Augmentation In Streaming Analytics," ResearchGate, 2023. [Online]. Available: https://www.researchgate.net/publication/390364121_ADAPTIVE_GENERATIVE_AI_FOR_REAL-TIME_DATA_AUGMENTATION_IN_STREAMING_ANALYTICS
12. Confluent, "The Power of Predictive Analytics in Business: Using Generative AI and Confluent," Confluent, Feb. 2024. [Online]. Available: https://www.confluent.io/blog/predictive-analytics-business/