Deloitte. | snowflake®

DATA MESH
A POINT-OF-VIEW FOR IMPLEMENTATION

In today's enterprises, data is spread across multiple regions, business functions, and sometimes different legal entities within the organization. These siloed data systems restrict the organization from being agile in decision-making based on business and market trends. Efforts are being made to consolidate and centralize data for the larger benefit of the organization; however, these attempts face resistance as business functions and entities are reluctant to relinquish the autonomy and governance they enjoy in data management and analysis.

The need of the hour would be to build a data archetype that effectively unites the disparate data sources and links them together through centrally managed data sharing and governance guidelines. The data archetype integrates data silos in a mesh architecture from different business functions within a strong data governance framework, most importantly, without diluting the autonomy of the business functions.

DATA MESH

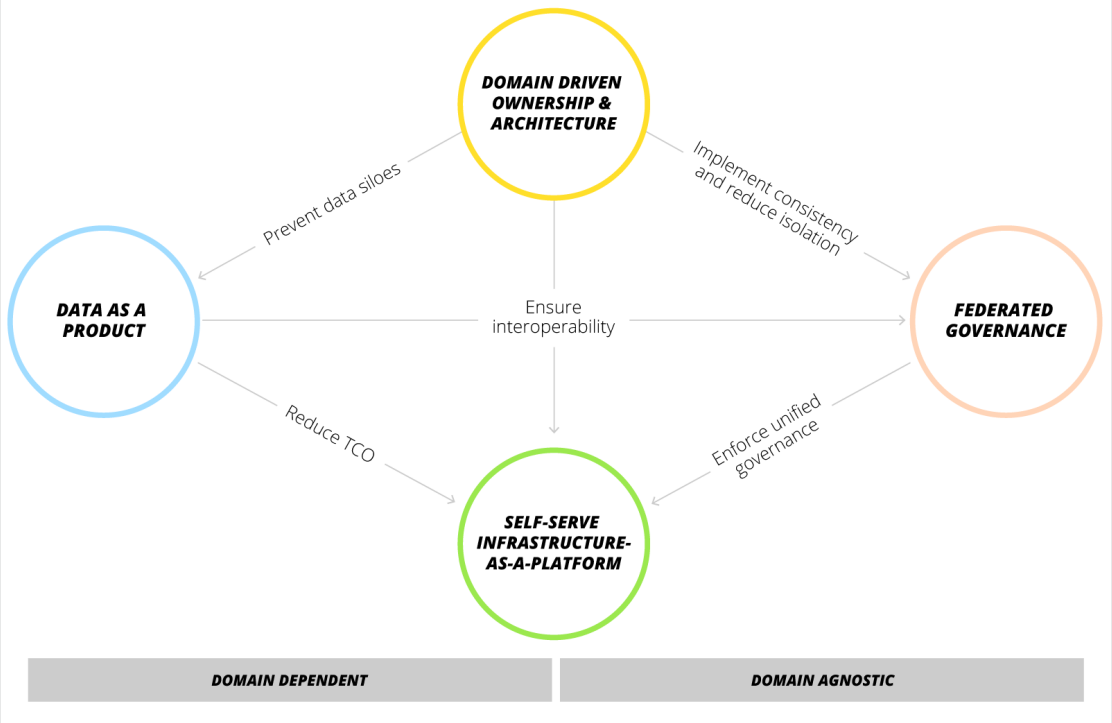
A data mesh is an architectural framework that addresses advanced data governance and security challenges through distributed, decentralized ownership. Business functions can maintain control over how shared data is accessed, who accesses it, and in what formats it is accessed. A data mesh transfers data control to domain and business experts who create suitable data products adhering to a decentralized governance framework. Data consumers can request access to the data products and seek approvals or changes directly from data owners. As a result, everyone in organization gains faster access to relevant data, which can improve business agility.

Potential benefits:

- **Autonomy and accountability:** Empowers business functions and domains to be self-sufficient while respecting overarching governance framework and guidelines.
- **Accelerated (cross-domain) collaboration and productivity:** Simplifies the sharing and access to data across teams, and promotes open standards and interoperability.
- **Scalability and performance:** Data mesh allows an organization to efficiently scale its data management operations by reducing bottlenecks and improving system responsiveness and performance.
- **Adaptability:** In an ever-changing business and market environment, it is important for organizations to stay agile and swiftly adapt to transformations and customer demands.
- **Improved data quality and usability:** Entrust the teams that possess the most knowledge of the data and the domain. Data mesh helps enhance data accuracy, consistency, and reliability, which are crucial for data-driven decision-making and generating valuable insights.
- **Data governance and compliance:** Improved compliance with regulations and security standards, which is crucial for protecting sensitive data and meeting regulatory needs.

PRINCIPLES OF DATA MESH

There are four key principles of data mesh implementation that can help achieve scale, performance, quality and integrity, while adhering to a defined data governance framework.



DOMAIN-DRIVEN OWNERSHIP

By eliminating impediments in architecture and data ownership, domains can gain control over data ownership and architecture components. This can empower the domain or business function to creatively develop data products that address business questions and share examples with the rest of enterprise for easy consumption.

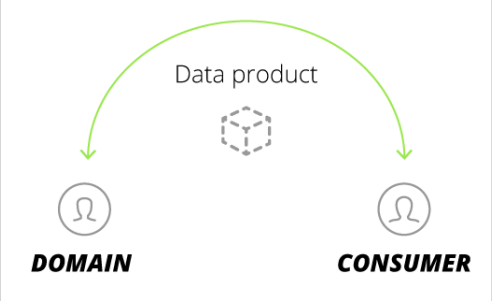
Key guidelines and expectations for domain ownership:

- Use IT organization's data strategy and guidelines to define architecture, tooling, project charters and execution plans.
- Cross-leverage data and tooling competencies with other domains and the IT organization within the enterprise to reduce time-to-market and avoid failures.
- Ensure teams with domain and business knowledge take ownership of data pipelines.
- Develop, manage, and own compressive data cleansing, refinement, pre-aggregation processes based on established guidelines and governance structures.
- Take local responsibility for implementing data governance, privacy, and confidentiality policies, and for frequently updating the business glossary and lineage of the data assets.
- When developing data assets and products, keep consumers in mind and create simple and effective usage protocols.
- Measure and communicate the effectiveness and usage of data products and assets by other business and domain functions.

DATA AS A PRODUCT

To make data products successful, domain owners should prioritize purpose, adaptability, performance, reusability and extensibility to meet the evolving business environments and data requirements.

- Conduct design thinking workshops with key business users to identify business problems, determine how data and intelligent algorithms can measure it and what action is recommended to improve or answer those problems.
- Involving end consumers in understanding how they use the data and what actions they perform to consume it is critical for the success of the data product.
- Defining and creating a backlog of functionalities and features for data products that can be easily understood by consumers.
- Develop an MVP (minimum viable product) of the data product to model key functionalities, identify potential bottlenecks, and address shortcomings.
- Identify the confidentiality and privacy of data attributes, and develop and adapt a compressive security framework within the ecosystem, adhering to governance guidelines.
- Managing the product lifecycle is critical for successful adoption by consumers in an enterprise. Key activities include product versioning and release management, release notes, documentation, and managing end consumer communication and training.



SELF-SERVE DATA INFRASTRUCTURE AS A PLATFORM

To develop analytical data products effectively, it is essential to provide tools and user-friendly interfaces for both analytical end-users and development teams.

Consider implementing a set of organization-approved standard tools across domains or business functions to enable data mesh architecture. This supports data product lifecycle management, self-service capabilities, template creation, and repeatable patterns.

For example, Snowflake enables domain-driven ownership with self-service infrastructure as a platform. It supports:

Self-service infrastructure as a platform | Collaboration on data products | Federated governance

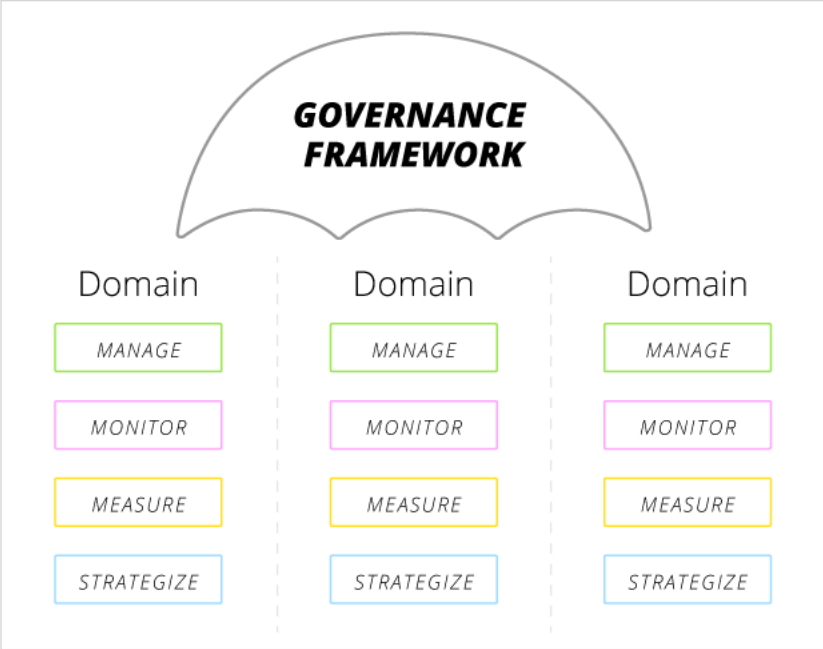
Next, we'll cover Snowflake's data mesh capabilities in detail, including deployment and management strategies.

FEDERATED GOVERNANCE

Balancing centralization and decentralization of governance, along with sharing rules for managing the data product lifecycle and consumption, is critical to the success of the data mesh.

A governance strategy that grants sufficient autonomy to domains enhances efficiency and enables:

- Standardized, repeatable processes that increase efficiency and reduce risk.
- Standardized measurement processes that provide accurate decision making and prioritization.
- A consistent decision and prioritization process that aligns with strategic goals and objectives.
- A unified communication strategy and plan to deliver effective information and involve relevant stakeholders.
- Standardized project methods, patterns, development guidelines, and processes to build consistency and repeatability.
- Clearly defined roles and responsibilities, providing greater visibility and accountability.



DATA MESH VS DATA LAKE

	TRADITIONAL DATA LAKE	DATA MESH
Ownership	Centralized ownership	Decentralized domain ownership
Platform	Centralized, monolithic data platform	Distributed architecture and ecosystem of data products
Infrastructure	Centralized infrastructure management	Decentralized, self-service infrastructure as a platform
Team orientation	Role oriented data teams Segregated data teams	Domain oriented cross-functional teams with embedded data teams
Governance	Top-down governance and team	Federated computational governance with centralized guidelines
Data	Data is by-product of code	Data and code are one unit
Focus	Pipelines as a first-class citizen	Data as a first-class citizen and meant to share and connect

DATA DOMAINS

In the context of a data mesh architecture, a data domain refers to a logical grouping of related data that share common characteristics, such as business context, data type, or subject matter. It represents a high-level categorization of data within an organization. Data products operationalize and deliver data within this architecture, while data domains provide the overarching organizational structure for managing and categorizing data.

DOMAIN OWNERSHIP AND FEATURE ENGINEERING BY DOMAIN TEAMS

Pairing data scientists with data engineers in a domain team facilitates data exploration on clean and transformed data. This collaboration generates features that are stored in a product marketplace. A marketplace is a data repository that serves features for training and inference, and helps track feature versions, metadata, and statistics. This capability enables data scientists in the domain team to work closely with domain experts and keep the features refreshed as data changes within the domain.

DATA PRODUCTS

A data product is a data asset that has been processed, curated, and made available for consumption to various users or applications within an organization. These products are self-contained and autonomous, often encapsulating data pipelines, processing logic and metadata.

Data products are designed to provide well-defined and reusable data assets that can be discovered and accessed by different teams, offering a more decentralized and scalable approach to data management and analytics.

Within a data mesh framework, a data product is a more granular concept compared to a data domain and is associated with a data stewardship model.

Depending on the domain, the data product could be:

	CORE FEATURES
File, Table, Data Set KPI, Metrics ML Model API Message Stream Sub-Graph	• Data product schema registry and documentation • Data product discovery, catalog registration and publishing • Data product lineage • Data product versioning • Data product de-identification • Data product quality metrics (collection and sharing) • Data product monitoring/alerting/logging • Unified data access control and audit logging

INDUSTRY EXAMPLES

Based on our experience of developing the data taxonomy for various sector clients, the examples below represent an initial view of how organizations can start thinking about their own data product taxonomy:

CONSUMER	MANUFACTURING	FINANCIAL SERVICES
Product, Supplier	Bill of Material Item Taxonomy Site	Customer Products
PO, Invoices	Receipts, Product Backlog, Constraints	Risk Compliance
Sales orders, SKU, Shipments, Compliance	Shipments Inventory SKU	Employee
Demand Planning	Invoices, Payments Receivables	Service and Complaints
Promotions, Pricing	PLM Sales Orders	Credit Score
Inventory	Complains Services	KYC

DATA MARKETPLACE

A data marketplace offers data consumers an intuitive, secure, centralized, and standardized data consumption experience. It brings data and associated products closer to data analysts and scientists by utilizing underlying metadata. It also tracks all the data products, which are often stored across a range of data domains.

A data marketplace is typically a thin orchestration layer with an appealing look and feel, offering a unique user experience. Data marketplaces utilize underlying metadata repositories that can be a mixture of homegrown metadata stores and popular catalogues.

Building a data marketplace involves structure, culture, and people. It also requires having trust in users, training personnel, and building awareness.

Key usage functionality includes:

Data preparation:
Providers to curate data products based on features and requirements.

Data services:
Self-service APIs to import data from providers and provide subscription-based access to data consumers.

Subscription:
Allows consumers to access and augment data products by combining with other information and republish the enriched data product

Search and browse:
Allows consumers to search data products based on metadata and tags and provides data previews.

Access workflow:
A workflow-based access provisioning mechanism to grant access to a data product.

DATA MESH WITH SNOWFLAKE

The Snowflake AI Data Cloud platform enables organizations to implement a data mesh architecture by providing a set of enterprise-grade service capabilities. It supports multiple data formats, provides elastic compute, and supports multiple programming languages, allowing domain teams to use their preferred tools and methodologies.

Additionally, Snowflake provides federated governance across distributed products through comprehensive security features and a centralized catalog for managing these products. Its multi-cloud availability and sharing capabilities across regions creates a unified ecosystem, enabling seamless operations across geographic boundaries and cloud providers.

Let's explore the specifics of Snowflake's data mesh implementation and its benefits in more detail in the following sections.

SNOWFLAKE COMPONENTS AT PLAY

Snowflake facilitates decentralized ownership across various business functions, such as sales, marketing, and finance, by enabling each function to operate as an autonomous team. These teams manage their assets through dedicated accounts, databases, or schemas, promoting independence and self-governance. Furthermore, Snowflake's interoperable storage supports multiple formats, allowing teams to work with their preferred formats and tools. The cloud-scale compute warehouses offer serverless capabilities, a pay-per-use model, and scalable resources that can be provisioned instantaneously, providing flexibility and cost-effectiveness.

Snowflake supports multiple programming languages, including Java, SQL, Python, and Scala, providing operational autonomy. The Polaris Catalog empowers teams to choose and operate their preferred processing engines while maintaining interoperability, ensuring seamless integration with existing workflows. Additionally, Snowflake Marketplace allows users to discover, share, and access products, machine learning code, and applications across regions and clouds, fostering collaboration and innovation. Snowgrid, a foundational cross-cloud network, enables independent operations while maintaining centralized control.

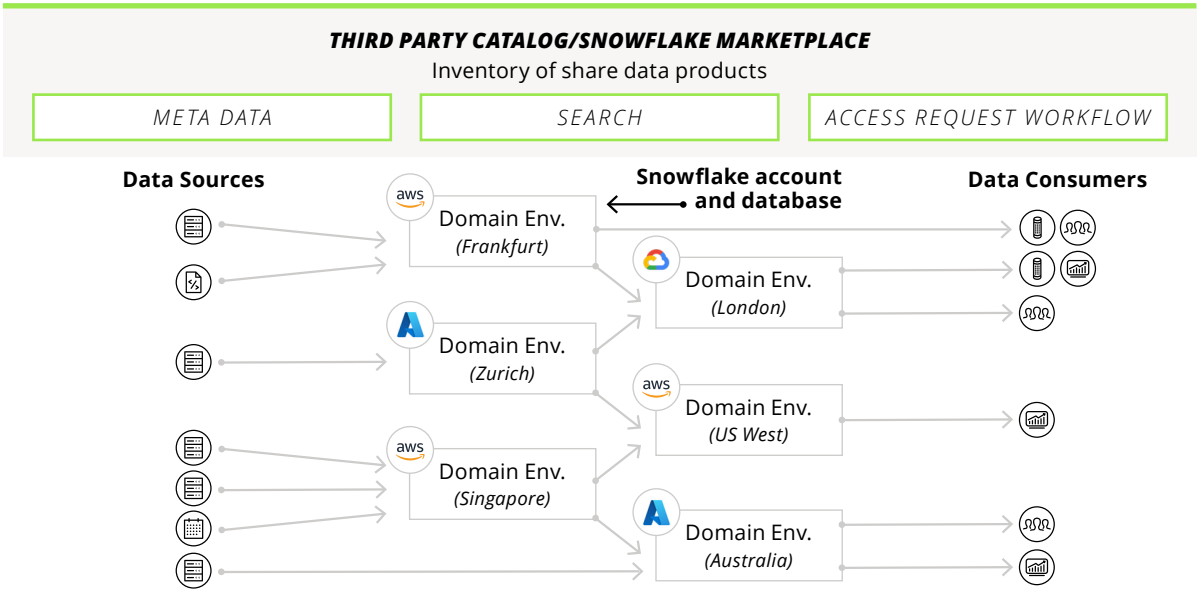
Building upon this infrastructure, the Snowflake Horizon Catalog provides robust governance through integrated compliance, security, and privacy capabilities. Key features include data masking, data quality monitoring, lineage tracking visualization, and role-based granular access controls. The catalog manages assets across organizational accounts, external tables, and third-party connections, supporting end-to-end encryption and various authentication methods, ensuring a secure and compliant data environment.

ARCHITECTURE OVERVIEW

Data mesh architecture requires a flexible and scalable framework for building a unified ecosystem, supporting multiple design patterns that balance autonomy with centralized control and collaboration.

Snowflake's global multi-cloud availability creates a unified ecosystem that spans across AWS, Azure, and Google Cloud platforms, enabling seamless operations across geographic boundaries and cloud providers. This ecosystem can enable organizations to choose from multiple architecture options, striking a balance between autonomy and decentralization. The following topologies are available:

- **Account per environment:** Each environment uses a separate Snowflake account, providing maximum isolation between environments.
- **Database per environment:** Each environment uses one or more separate Snowflake databases, providing simplified management of users, security, and control.
- **Schema per environment:** Each environment uses separate schemas in a single database, providing a lower degree of isolation between environments.
- **Heterogeneous environments:** Environments can use different IT stacks, requiring special consideration and potentially incurring a higher degree of complexity.



Within this architectural framework, organization in Snowflake's data mesh is structured around independent environments, each representing a distinct business context or function. These environments operate as self-contained units responsible for their own products, adhering to the mesh principle of decentralization while maintaining interoperability across the ecosystem.

To facilitate collaboration across these autonomous environments, Snowflake's sharing features enable secure product sharing and cross-environment collaboration. These capabilities provide a centralized platform for interaction while maintaining individual environment control and data lineage, ensuring that autonomy and security remain uncompromised.

Snowflake Horizon complements this architecture by providing a comprehensive governance framework that establishes control in the data mesh architecture. It supports secured environments through features such as federated control, products as a service, and self-serve infrastructure. This framework creates secure cross-environment sharing and collaboration while providing cost control features for accountability and measurement of product value.

SNOWFLAKE DOMAIN-ORIENTED ARCHITECTURE IMPLEMENTATION GUIDE

Organizations adopting a Snowflake-oriented architecture should follow a structured, methodical approach that prioritizes security, governance, and operational efficiency. This section provides an approach for new customers and a transformation strategy for existing customers.

Cross-functional team: Assemble a cross-functional team with comprehensive skills training and change management.

Clear communication: Provide clear communication and executive sponsorship throughout the process.

Technical implementation: Prioritize backward compatibility, seamless migration, and robust rollback mechanisms.

Success metrics: Measure success by reduced complexity, improved security, enhanced governance, increased monetization potential, and lower management costs.

Challenges: Anticipate challenges, technical complexity, skills development, and standardization, and address them through a holistic approach.

RECOMMENDED APPROACH

A structured implementation approach is essential for organizational success, focusing on systematic deployment, robust security, and streamlined operations.

Mapping exercise Conduct a comprehensive exercise to identify core business areas and their specific management requirements.

Topology design Implement an account-per-area topology to provide the highest level of isolation, security, and operational autonomy. Organizations should create dedicated Snowflake accounts for each major business area, ensuring clear boundaries and independent operational capabilities.

Security configuration Mandate security checks such as multi-factor authentication across all accounts, utilizing masking, encryption methodologies of critical data, key-pair as the primary authentication method. Implement network policies that restrict access based on IP ranges, geographical locations, and specific client types. Leverage private connectivity options like PrivateLink to provide maximum network security and isolation.

Policy development Develop a centralized and flexible governance approach, including a standardized tagging strategy for sensitive classification and automated protection mechanisms. Implement granular protection policies as a default configuration, ensuring protection from the outset.

Product standardization Create standardized products to enable efficient sharing and consumption. Develop a centralized exchange mechanism that allows areas to publish and consume products through consistent, governed interfaces. Implement reader accounts to support external sharing while maintaining strict access controls.

Pilot implementation Focus on 3-5 core areas for the initial implementation, allowing for controlled scaling and learning. Phased Expansion: Gradually expand to additional areas, ensuring a smooth transition and minimal disruption.

ESTABLISHING THE SNOWFLAKE DATA MARKETPLACE

Internal to organization

The Snowflake Marketplace functions as a centralized platform where different domain teams within an organization can publish and discover data products through detailed listings. It operates as a "single pane of glass" for data product discovery, procurement, and configuration with single-click functionality.

The internal sharing mechanism enables live, real-time data access for accounts within the Snowflake organization without requiring complex ETL processes, while maintaining robust security through features like row/column-level security and dynamic data masking. By treating data as a product, the Snowflake Marketplace supports the data mesh principle, which promotes clear ownership and quality metrics. This approach enables domain teams to easily search, discover, and request access to relevant data assets, creating a self-service environment that promotes data democratization while maintaining governance standards.

External to organization

For data sharing across external entities, the Snowflake Marketplace offers streamlined mechanisms to share data products with partners, customers, and third-party consumers. Organizations can monetize their data assets or share them freely through public listings, Reader Accounts, and secure data exchanges. External consumers can integrate with technologies and access live data directly from source domains through secure governed data shares.

The Snowflake Marketplace supports cross-cloud and cross-region sharing, providing access to diverse data categories including customer insights, weather data, financial metrics, demographics, and geospatial information. Data providers maintain control over their shared assets with the ability to revoke access at any time, while external consumers benefit from always-current data without the need for copying or complex integration processes.

ESTABLISHING GOVERNANCE WITH SNOWFLAKE HORIZON

Snowflake Horizon provides a comprehensive framework for establishing governance in a data mesh architecture, integrating centralized oversight with decentralized implementation.

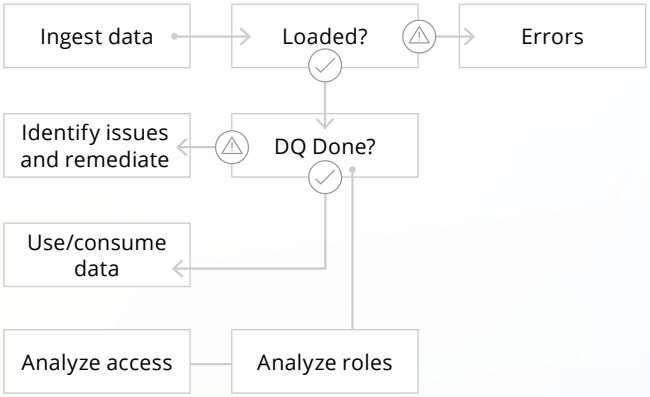
Key features include:

- Federated Computational Governance and technical controls (tagging, classification, lineage)
- Cost governance capabilities for domain accountability and value measurement
- Data masking and clean rooms for secure cross-domain collaboration
- By leveraging these features, organizations can establish a well-governed data ecosystem that balances autonomy with control, while promoting data democratization and maintaining security.

By leveraging these features, organizations can establish a well-governed data ecosystem that balances autonomy with control, while promoting data democratization and maintaining security.

SNOWFLAKE IN ACTION

Data Engineer	• Structured, Semi and Unstructured Data, Iceberg • Batch, RealTime, Ingest Profiling , DQ Rules
Data Governor/Steward	• View/remediate • Data quality issues • Monitor, Sensitive Data
Data User/Data Owner	• View data per role • Share data • Automated Recommendations
Security Admin	• Vulnerabilities • Access Abuse



CONCLUSION

The adoption of a data mesh architecture represents a paradigm shift in how organizations manage and utilize their data. By decentralizing data ownership and promoting a domain-oriented approach, a data mesh addresses many of the limitations inherent in traditional monolithic data architectures. This approach not only enhances scalability and flexibility but also empowers domain teams to take full ownership of their data products, fostering innovation and agility. Looking ahead, the data mesh can offer a robust framework for organizations looking to harness the full potential of their data in a rapidly evolving digital landscape.

For more information, please contact:

Jason Eichenholz
Snowflake Lead Alliance Partner
Deloitte Consulting LLP
jeichenholz@deloitte.com

Srinivas Batny
Executive Director
Deloitte Touche Tohmatsu India LLP
sbatny@deloitte.com

Jayanth Hegde
Partner Development Manager
Snowflake India | Alliances
jayanth.hegde@snowflake.com

Bharath Suresh
Partner Solutions Engineer
Snowflake India | Deloitte Alliance
bharath.suresh@snowflake.com