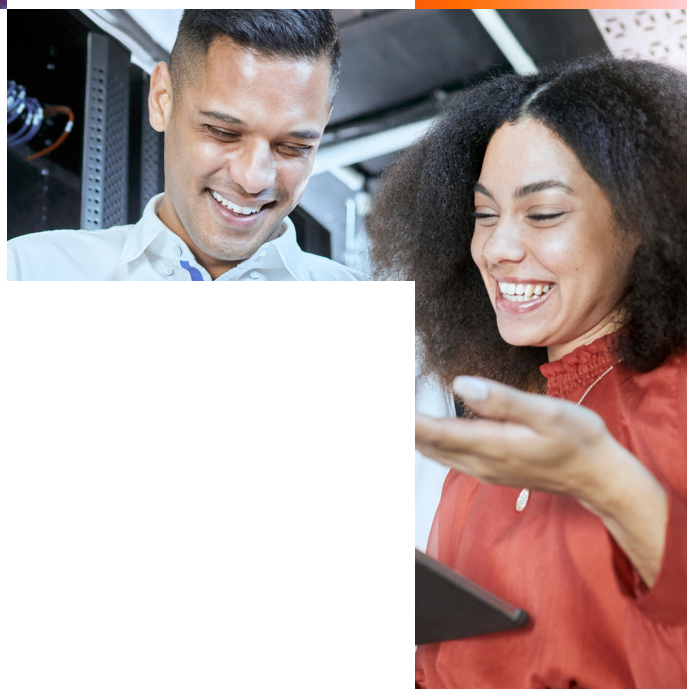

June 27, 2023

Enhancing data mesh: How distributed ledger solutions empower decentralized governance

CGI

Introduction

Data mesh is an innovative approach to data management that offers a solution to the most common challenges organizations face when dealing with large-scale data management. Data mesh concept consists of four key principles for large-scale data management in a multi-tiered, multi-domain organizational structure: domain-oriented ownership, data as product, federated computational governance and a self-service data platform. These all promote a more agile, efficient and scalable data architecture. However, one of the central challenges to implementing data mesh is designing decentralized governance mechanisms that align with its core principles.

This paper examines the potential of distributed ledger technology (DLT) and smart contracts as a mechanism to implement the decentralized governance in data mesh. DLT has proven its value in providing secure, decentralized and tamper-proof data management capabilities. Distributed ledger technology creates a trustless, decentralized governance model that aligns with the principles of data mesh, including domain-centric ownership, secure data sharing and interoperability.

Contents

2. Background and overview of data mesh

2.1 Data mesh approach

In the book "Data Mesh – Delivering Data-Driven Value at Scale," author Zhamak Dehghani, CEO of Nextdata, emphasizes that decentralized data governance and ownership form the cornerstone of effective data management. Dehghani introduced the concept in a 2020 article titled "How to Move Beyond a Monolithic Data Lake to a Distributed Data Mesh." Since then, the adoption of data mesh has gained traction in both government and commercial organizations due to its potential to overcome the challenges of centralized data management.

Data mesh offers several benefits over traditional centralized data management by promoting agility, resilience, scalability and data democratization, which can help organizations to better manage their data and leverage it for insights and decision-making.

Data mesh emphasizes the following:

- *Domain-oriented ownership:* Business domains own the data, rather than central IT. Each domain has the responsibility to manage and govern its own data, which helps to improve data quality and reduce the time to value for data-driven initiatives.
- *Data as a product:* Cross-functional teams build, maintain and deliver data as they would any other product. This approach fosters a culture of accountability and ownership, where teams are responsible for the quality and accuracy of their data products. It promotes the standardization and quality of data by treating it as a product with well-defined characteristics, such as quality, availability and accessibility.
- *Federated computational governance*: Teams keep control over their data while leveraging shared computational resources. This model fosters agility, allowing for faster data processing and analysis. It brings greater flexibility and scalability in data management by allowing domains to collaborate on data management tasks such as quality control, access control and lifecycle management.
- *Self-service data platforms*: This provides teams with the ability to access and use data in a secure and

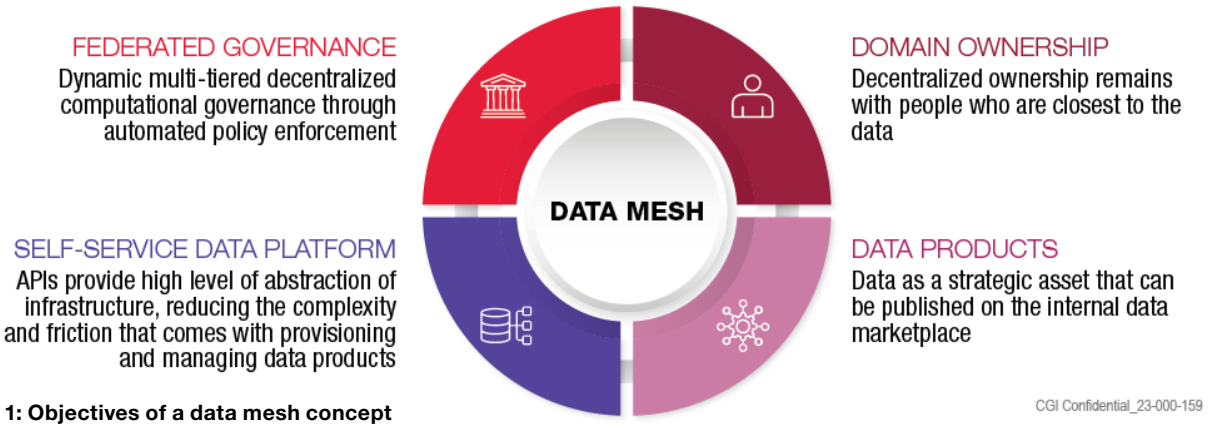

Figure 1: Objectives of a data mesh concept

efficient manner. This principle promotes the democratization of data by providing a platform for self-service data access and analysis.

2.2 Challenges in implementing a data mesh

Data mesh architecture presents five challenges that we describe in further detail below, namely, domain integration, governance modeling, user adoption and training, consensus building and security/privacy

Firstly, there is the technical complexity involved in setting up and integrating technologies within the different domains. Organizations may encounter challenges related to infrastructure setup, protocol selection, security measures and interoperability with existing systems.

Another challenge lies in designing an effective governance model that aligns with the principles of data mesh. This involves defining roles, responsibilities, decision-making processes, and incentive mechanisms that promote collaboration, transparency, and fairness. Additionally, organizations operating in regulated industries may face challenges in adhering to compliance requirements while implementing decentralized governance. Navigating data privacy regulations, data protection, and industry-specific requirements requires careful consideration and compliance measures.

User adoption and education are crucial aspects of decentralized governance implementation. Educating stakeholders about decentralized governance concepts and technologies and addressing their concerns are essential for successful adoption. This may involve providing training, facilitating user understanding of new processes and tools, and managing change within the organization.

Network consensus and governance decision-making present further challenges. Decision-making within a decentralized network can be complex due to diverse stakeholder interests. Designing effective mechanisms for decentralized decision-making, dispute resolution and community governance is vital for the success of the governance model.

Organizations must also address security and privacy considerations. While decentralized governance offers enhanced security through cryptographic mechanisms, organizations need to ensure the integrity of smart contracts, protect private keys, prevent unauthorized access to sensitive data and establish secure communication channels. Privacy concerns may arise when sharing data across different domains within the data mesh architecture, requiring robust data anonymization and access control mechanisms.

Addressing these challenges requires careful planning, collaboration and continuous evaluation. Organizations should approach the implementation of decentralized governance iteratively, learning from experiences and adapting the architecture and governance mechanisms based on real-world insights and evolving best practices.

3. Background and overview of distributed ledger technology

Distributed ledger technology (DLT) enables secure and transparent data sharing among multiple parties without the need for a central authority. There are several types of DLT available, including blockchain and directed acyclic graph (DAG).

Blockchain uses cryptography to ensure the immutability and integrity of data. It creates a chain of blocks, each containing a set of transactions, and uses a consensus mechanism to achieve consensus among network participants. DAG uses a graph structure, allowing the creation of multiple branches and nodes, enabling faster processing of transactions and scalability than blockchain.

3.1 Public vs private blockchain

There are two types of blockchains: public and private. Public blockchains, such as those used for cryptocurrency, allow anyone to participate in the network, validate transactions and create new blocks. These blockchains offer a high level of decentralization and security as they rely on a distributed network of anonymous participants. However, they also come with certain limitations, such as scalability issues and potentially slower transaction speeds.

While public blockchains have their merits, they might not be the most suitable choice for decentralized governance within a data mesh architecture. The open nature of public blockchains may not align with the need for controlled access and data privacy that domains and teams require. Moreover, the consensus mechanisms used in public blockchains can be resource-intensive and may not be efficient for governance purposes within a data mesh ecosystem.

On the other hand, a single entity or a group of trusted participants control a private blockchain, with restricted access and participation. It offers a more controlled and permissioned environment for decentralized governance in data mesh. For the rest of this whitepaper, we will use the term DLT to refer to a private blockchain or DAG, one that restricts participation to authorized entities, enabling greater control over access rights and data sharing. These DLTs can be tailored to the specific needs of the data mesh ecosystem, ensuring compliance with privacy regulations and providing higher performance and scalability compared to public blockchains.

3.2 Smart contracts, a key feature of distributed ledger technology

Smart contracts are self-executing programs that automatically enforce the terms of a contract between parties. They are a key feature of DLT technology and have gained significant attention

in recent years due to their potential to streamline business processes and reduce the need for intermediaries.

Organizations can use smart contracts to automate a wide range of business processes, including supply chain management, financial transactions and legal agreements. They can be customized to meet the specific needs of different industries and programmed to handle complex logic and decision-making.

Smart contracts are written in code and are stored on a DLT. They contain the terms of the agreement between parties and execute automatically when certain conditions are met. For example, a smart contract could be programmed to release payment to a vendor once a shipment of goods has been received and verified by a third-party logistics provider.

> In March 2021, Walmart Canada rolled out a network known as DL Freight to all of the company's logistics carriers. The system continuously gathers information using DLT technology at every step — from the tender offer from the carrier to the proof of delivery and the approval of payment. This information is automatically captured and synchronized in real time and is visible only to the parties involved in the transaction. Before Walmart Canada implemented DL Freight, 70% of invoices were disputed. Today less than 1% of invoices have discrepancies, and these disputes are easily flagged and quickly resolved.
>
> -Harvard Business Review, 2022

One of the key benefits of smart contracts is that they are transparent and tamper-proof. Once a smart contract is executed, it is recorded on the DLT and cannot be altered or deleted. This makes smart contracts a secure and reliable way to automate business processes and enforce agreements between parties.

Another benefit of smart contracts is that they can reduce the need for intermediaries, such as banks or legal professionals. By automating the enforcement of contracts, smart contracts can help to reduce costs and streamline business processes.

4. Advantages of implementing DLT and smart contracts in data mesh

Distributed ledger technology (DLT) offers a compelling solution to address the challenges of secure data sharing and ownership within the context of a data mesh architecture. By leveraging the inherent properties of DLT, organizations can establish a decentralized and immutable ledger that ensures data integrity, transparency and trust among participating entities.

One of the key features of DLT is its distributed nature, where multiple participants maintain copies of the same ledger. This distributed ledger eliminates the need for a central authority, making it highly resistant to tampering or unauthorized modifications. As a result, data shared and stored on the DLT becomes inherently secure and immune to single points of failure or manipulation.

DLT achieves secure data sharing and ownership through the following mechanisms:

- Immutable data storage: Data stored on the DLT is organized into blocks, with each block containing a cryptographic hash of the previous block. This linkage creates a tamper-evident chain of data, where altering any previous block in the chain would require changing subsequent blocks, making it computationally infeasible. This immutability ensures that once data is recorded on the DLT, it cannot be easily altered or tampered with, providing a strong guarantee of data integrity.
- Decentralized consensus: DLT employs a consensus mechanism, such as proof-of-work or proof-of-stake, to validate and agree on the state of the ledger. Through this decentralized consensus, all participants in the network reach a shared agreement on the validity of transactions and the order in which they are added to the DLT. This consensus process ensures that only authorized and valid transactions are included in the DLT.
- Smart contracts for data governance: In a data mesh, smart contracts can enforce data governance policies and permissions, defining who can access, use and share data within the ecosystem. These smart contracts provide fine-grained control over data ownership, access rights, and usage restrictions, ensuring that data is shared and utilized in accordance with predefined rules and agreements.
- Cryptographic security: DLT utilizes cryptographic algorithms to secure data and transactions. Participants digitally sign transactions using their private keys, ensuring the authenticity and integrity of the data. Public-key cryptography enables secure data sharing without revealing sensitive information, as participants can verify the integrity of data using public keys without the need for disclosing private keys.

By integrating DLT into a data mesh architecture, organizations can establish a secure and decentralized data sharing environment. The immutability, distributed consensus, smart contracts, and cryptographic security offered by DLT provide robust mechanisms for ensuring data integrity,

ownership, and controlled access within the data mesh ecosystem.

DLT allows organizations to confidently embrace the principles of data mesh, empowering domain-oriented teams with secure data ownership, autonomous decision-making and transparent data governance.

4.1 Smart contracts facilitate decentralized governance

Smart contracts can play a crucial role in achieving decentralized governance within a data mesh architecture. They allow organizations to automate and enforce governance rules, policies and decision-making processes in a transparent and decentralized manner.

One way smart contracts facilitate decentralized governance is through the automated validation and execution of data sharing agreements. For example, when two domains within a data mesh agree to share data, they can establish a smart contract to define the terms and conditions of the data exchange. The contract can specify the data attributes, access permissions, usage restrictions and any associated compensation or royalties. Once the conditions outlined in the contract are met, such as successful data validation or meeting certain predefined criteria, the smart contract automatically triggers the data exchange process.

Furthermore, smart contracts enable the creation of decentralized data marketplaces within a data mesh. Domains or data providers can list their datasets on the marketplace or catalog, along with the respective terms and conditions of use. Interested consumers or data consumers can interact with the smart contracts to initiate the purchase or access of the desired data. The smart contract handles the transaction, verifies payment if required, and grants access to the data based on the agreed terms.

In terms of governance decision-making, smart contracts can facilitate decentralized voting mechanisms. For example, when a significant change or decision affecting the data mesh ecosystem arises, the parties involved can use smart contracts to enable voting among relevant stakeholders. They can encode each stakeholder's voting rights and influence into the contract, ensuring a fair and transparent decision-making process. The system can then automatically tally the votes and execute the resulting actions based on the predefined rules set within the smart contract.

4.2 How DLT can enable interoperability

Interoperability, the ability of different systems to seamlessly communicate and exchange data, is a critical aspect of modern data ecosystems. In the context of a data mesh architecture, where

data is distributed across multiple domains and owned by different teams, achieving interoperability becomes even more crucial. DLT, specifically designed to support decentralized and transparent data exchange, offers promising capabilities to enable interoperability and foster seamless data sharing among diverse entities within the data mesh ecosystem.

DLT's decentralized and distributed nature plays a fundamental role in enabling interoperability. By leveraging a shared ledger, DLT eliminates the need for a central authority or intermediary, enabling direct peer-to-peer interactions. This decentralized architecture empowers different domains and teams within the data mesh to connect and exchange data without relying on a single point of control or coordination. DLT acts as a trust layer that establishes a common ground for data exchange, ensuring transparency, immutability, and traceability.

DLT can enable interoperability in the following ways:

- Common data formats and standards: DLT networks often encourage consensus on data formats and transaction structures, fostering the adoption of common standards. By defining and adhering to these standards, different entities within the data mesh ecosystem can communicate and exchange data in a standardized manner. This promotes compatibility and reduces complexity when integrating heterogeneous systems.
- Smart contract-based orchestration: Through smart contracts, different domains can define and enforce rules for data exchange, access and usage. These contracts act as programmable protocols, ensuring that all parties involved adhere to predefined agreements. By leveraging smart contracts, interoperability can be achieved through automated and trustless interactions, reducing the need for manual coordination and mitigating the risk of miscommunication or errors.
- Consensus-driven data validation: DLT's consensus mechanisms contribute to ensuring the validity and integrity of shared data. Before adding data to the DLT, participants must reach a consensus on the validity of the information. This consensus-driven validation ensures that only verified and trusted data is included in the shared ledger. By relying on a decentralized consensus process, interoperability is enhanced, as data integrity and trustworthiness are established through a distributed network rather than relying on a centralized authority.
- Interoperable protocols and APIs: DLT networks often provide well-defined protocols and APIs that enable interoperability between different systems and applications. These standardized interfaces allow for seamless integration and data exchange among disparate entities. By adhering to these protocols, domains within the data mesh can interact with each other, leveraging the functionalities and capabilities offered by the DLT network. This promotes compatibility and simplifies the process of integrating and exchanging data between different domains.

DLT-enabled interoperability allows for the efficient utilization of data assets, promotes innovation through cross-domain collaboration, and eliminates data silos. As organizations embrace DLT-based interoperability, they can unleash the full potential of their data mesh architecture, unlocking new insights, opportunities and value from their interconnected data landscape.

5. CGI's proposed data mesh architecture using digital ledger technologies

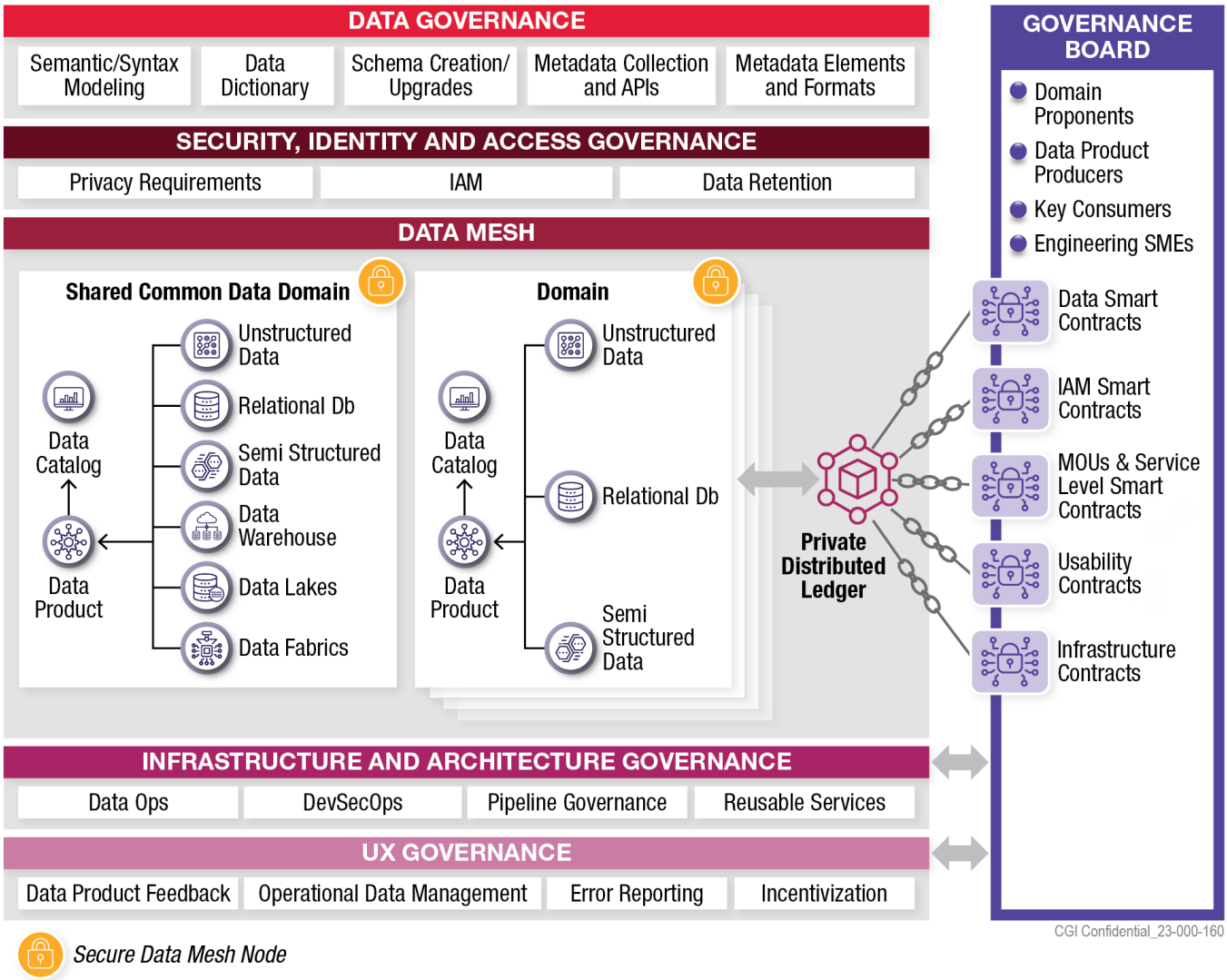

Figure 2: Notional data mesh architecture using distributed ledger technology

CGI proposes the notional architecture of a data mesh implemented using distributed ledger technology. In this architecture a private distributed ledger is employed to act as the interoperability medium between the various domains as explained in section 4.2. The DLT smart contracts provide the governance mechanisms required to implement data governance, security and identity governance, infrastructure/architecture governance and UX governance as explained in section 4.1.

5.1 Alignment with core data mesh principles

CGI's proposed architecture for decentralized governance in data mesh aligns closely with the fundamental principles laid out by Zhamak Dehghani, including:

Domain-oriented ownership:

Data mesh emphasizes domain-oriented ownership, where each domain has full ownership and autonomy over its data assets. CGI's proposed architecture aligns with this principle by enabling domains and teams to create and deploy their smart contracts. These smart contracts define data ownership, access policies, and governance rules specific to each domain. This decentralized approach empowers domains to take ownership of their data and govern it according to their specific needs and objectives.

Data as product:

In data mesh, data is treated as a product, and the architecture supports this principle by providing mechanisms to define data ownership, access rights and usage permissions through smart contracts. The smart contracts act as self-executing agreements, automating the enforcement of data governance policies. This ensures that data is managed as a valuable product, with clear rules and conditions for its usage, sharing and monetization.

Federated computational governance:

The architecture embraces federated computational governance, allowing domains and teams to participate in decision-making processes through token-based voting. The governance tokens, are enabled by using the smart contract features and represent ownership and influence over governance matters, ensuring that entities with a stake in the data have a voice in the decision-making process. This decentralized approach promotes collaboration, transparency and collective intelligence.

Self-service data platform:

The architecture supports the self-service nature of data mesh by providing user-friendly interfaces and tools for participants to interact with the decentralized governance framework. These interfaces enable domains and teams to submit proposals, cast votes, monitor governance activities, and explore the governance history. The self-service nature empowers participants to actively engage in governance processes, promoting agility and autonomy.

By aligning with these principles, the architecture ensures that the decentralized governance framework within data mesh is in harmony with the core values and objectives of the data mesh approach. It enables domains and teams to exercise ownership, define governance rules and participate in decision-making processes, fostering a culture of collaboration, accountability and innovation.

5.2 Additional benefits and enablers

CGI's proposed architecture offers additional benefits when realized.

Resilient network infrastructure:

The architecture starts with a distributed network infrastructure, where participating domains and teams form a peer-to-peer network. Built on a DLT platform and supported by the execution of smart contracts, this network ensures transparency, immutability, and traceability of governance-related transactions. The decentralized nature of the distributed network offer resilience as there is no single point of failure.

Automation of data sharing memoranda of understanding (MOU):

Because each domain or team creates and deploys a smart contract that defines the rules and conditions for data ownership, access rights and usage permissions, these smart contracts act as self-executing agreements and MOUs, automatically enforcing the defined governance policies.

Zero trust based on self-sovereign identities:

Decentralized identifiers (DIDs) provide unique and self-sovereign identities to domains, teams and individual participants within the data mesh ecosystem. DIDs ensure that each entity has a globally resolvable and persistent identifier, which they can use to establish trust and enable secure interactions within the decentralized governance framework. DIDs enable participants to own and control their identities, facilitating secure data sharing and collaboration.

Echeloning of governance:

Organizations have various ownership and influence over the governance processes within the data mesh. CGI's proposed architecture enables such asymmetrical needs by implementing governance tokens. Each domain or team is assigned a specific number of tokens, which they can use to participate in voting on proposals and decision-making. The distribution of tokens can be based on predefined criteria, such as data ownership, contribution to the ecosystem or other agreed-upon factors. Governance tokens provide a mechanism for fair and proportional participation, ensuring that entities with a stake in the data have a say in governance matters.

Decentralized decision-making:

The decentralized decision-making process is facilitated through the use of smart contracts and token-based voting mechanisms. When a governance decision needs to be made, domains can submit proposals in the form of smart contracts. These proposals can range from updates to data schemas, changes to access policies, or modifications to governance rules. Token holders can then vote on these proposals using their governance tokens, with voting power proportional to their token holdings.

Transparency and auditability:

DLT provides transparency and auditability by recording all governance-related transactions on the DLT. The ledger permanently records every interaction, proposal and vote, ensuring a transparent and immutable history of governance activities. This transparency enhances accountability and enables thorough audits of governance processes.

Dispute resolution:

The architecture incorporates a dispute resolution mechanism. Smart contracts can include predefined dispute resolution processes, such as mediation or arbitration, which can be triggered when conflicts arise. These mechanisms aim to resolve conflicts in a fair and transparent manner, ensuring the smooth functioning of the decentralized governance framework.

Integration with existing tools and systems:

CGI's proposed architecture can integrate with existing tools and systems within the data mesh ecosystem. For example, data discovery and cataloging platforms can interface with the DLT network, providing visibility into data ownership, access rights and governance policies. This integration ensures seamless compatibility and interoperability between the decentralized governance framework and other components of the data mesh architecture.

6. Conclusion

By adopting this proposed architecture, organizations can establish a decentralized governance framework within their data mesh architecture. The use of DLT, smart contracts, decentralized identifiers and governance tokens empowers domains and teams to take ownership of their data and participate in decision-making processes.

As organizations navigate the challenges of managing complex and diverse data ecosystems, CGI's proposed architecture offers a novel and powerful solution for decentralized governance. It promotes trust, transparency and accountability, empowering domains and teams to take ownership of their data and participate actively in decision-making processes. By adopting this architecture, organizations can unlock the full potential of their data assets while fostering a culture of collaboration, innovation and data-driven insights.

7. References

1. Zhamak Dehghani. How to Move Beyond a Monolithic Data Lake to a Distributed Data Mesh. 05/20/2019 : https://martinfowler.com/articles/data-monolith-to-mesh.html

2. Zhamak Dehghani. Data Mesh Paradigm Shift in Data Platform Architecture. San Francisco, USA: 2020 : https://www.youtube.com/watch?v=52MCFe4v0UU

3. Lance Johnson. What is a Data Mesh? 2020: https://trustgrid.io/what-is-a-data-mesh/

4. Zhamak Dehghani. Data Mesh Principles and Logical Architecture 2020: https://martinfowler.com/articles/data-mesh-principles.html

5. Barr Moses. What is a Data Mesh — and How Not to Mesh it Up 07/08/2020: https://towardsdatascience.com/what-is-a-data-mesh-and-how-not-to-mesh-it-up-210710bb41e0

6. Sven Balnojan. Data Mesh Applied 2/17/2019: https://towardsdatascience.com/data-mesh-applied-21bed87876f2

7. Max Schultze & Arif Wider. Data Mesh in Practice: How Europe's Leading Online Platform for Fashion Goes Beyond the Data Lake 2020: https://www.youtube.com/watch?v=eiUhV56uVUc

8. Justin Cunningham. Netflix Data Mesh: Composable Data Processing - Justin Cunningham 2020: https://www.youtube.com/watch?v=TO_IiN06jJ4

9. James Reid & Arup Nanda : Data Lake Strategy via Data Mesh Architecture at JPMorgan Chase; Data Mesh Learning Meetup #005 2022: https://www.youtube.com/watch?v=7iazNKG8XQo

10. Kate Vitasek, John Bayliss, Loudon Owen, and Neeraj Srivastava: How Walmart Canada Uses Blockchain to Solve Supply-Chain Challenges 2022: https://hbr.org/2022/01/how-walmart-canada-uses-blockchain-to-solve-supply-chain-challenges

11. Natasja van der Keijl, Hans Moonen : The what, how and why of data sharing 2023: https://smartport.nl/wp-content/uploads/2023/03/SmartPort_White-paper_Data-Sharing_EN_final.pdf

About CGI

Founded in 1976, CGI is among the largest IT and business consulting services firms in the world.

We are insights-driven and outcomes-based to help accelerate returns on your investments. Across 17 industries in 400 locations worldwide, our 76,000 professionals provide comprehensive, scalable and sustainable IT and business consulting services that are informed globally and delivered locally.

Our commitment: Insights you can act on.

cgi.com

CGI