Bell Labs Consulting

Data mesh and distributed data architecture for telco networks

Enabling autonomous networks

White paper

Abdol Saleh, Astha Sharma and Harold Batteram

Contents

Executive summary

Autonomous networks (AN) are designed to significantly reduce manual intervention. By leveraging advanced automation, AI, and machine learning, AN can self-optimize and manage services within pre-defined reliability thresholds. This reduces operational costs, minimizes human error, and enhances service quality by dynamically adjusting to network conditions in real-time. ANs can also scale services according to demand, ensuring efficient use of network resources. Furthermore, their ability to adapt swiftly to changing business requirements makes them invaluable in addressing new market opportunities, launching innovative services, and responding to evolving customer needs, while maintaining robust security and reliability standards.

Autonomous telecom networks demand a robust, distributed data platform capable of ingesting, processing, transforming, and making transformed data accessible to future network digital twins (NDTs) and a multitude of federated machine learning (ML) algorithms that manage the planning and control functions of the telecom network. Achieving this requires a paradigm shift to a data mesh approach, transforming the organization into a true digital enterprise.

An integrated and distributed data platform — incorporating data mesh principles, data lakehouse design, and data fabric architectures within a distributed hybrid cloud — is essential for enabling autonomous networks in the telecommunications ecosystem. Data mesh principles are particularly suitable for telecom networks that encompass a multitude of domains such as radio access networks (RAN), fixed access networks, core networks, transport networks, and service and application layers. Planning and control in such a complex environment necessitate sophisticated interworking to provide seamless connectivity, high bandwidth, and low-latency services. Utilizing data mesh principles enabled by a distributed data platform ensures efficient data flow across these domains, enabling real-time processing, analytics, and decision-making.

The data mesh concept supports domain-specific data ownership and architectural decentralization, vital for managing the extensive and diverse data generated by various network domains. The data lakehouse architecture complements this by offering a consolidated repository for storing and analyzing structured and unstructured data at scale, optimizing network operations and customer experience. Meanwhile, the data fabric provides a unified data management layer, ensuring data interoperability and accessibility across the telecom network’s disparate domains.

Deployed on a distributed hybrid cloud, this approach addresses the geographical dispersion, scalability demands, data sovereignty, compliance, and operational resilience required by telecom networks. A distributed data platform enables telecom operators to unlock the full potential of autonomous networks, offering advanced functionalities such as predictive maintenance, dynamic resource allocation, and automated service provisioning. These capabilities are crucial for autonomously managing the complex and dynamic nature of modern telecom networks in real time, enhancing the efficiency and reliability of telecom services. Moreover, this approach paves the way for next-generation technologies like 6G and beyond, where the integration of various network domains is critical for delivering innovative services and ensuring customer satisfaction.

Despite the potential, communication service providers (CSPs) have yet to deploy distributed data platforms in real-world scenarios. Historically, CSPs have focused on building large data lakes as a single source of truth. However, these large data lakes present their own challenges. This paper explores how a distributed data platform can address these challenges and provides a conceptual approach for deploying such a platform in a CSP’s environment.

Autonomous networks

With the evolution of new digital service being provided over the network, networks need to provide service delivery assurance in line with the business intent. This, along with additional complexity due to network architecture and increasing demand for hyper connectivity, is resulting in inefficient operation domains in CSP business models. The only way CSPs can scale without compromising business agility, reliability and efficiency is to revise their operations model to one that is anchored on digital automation.

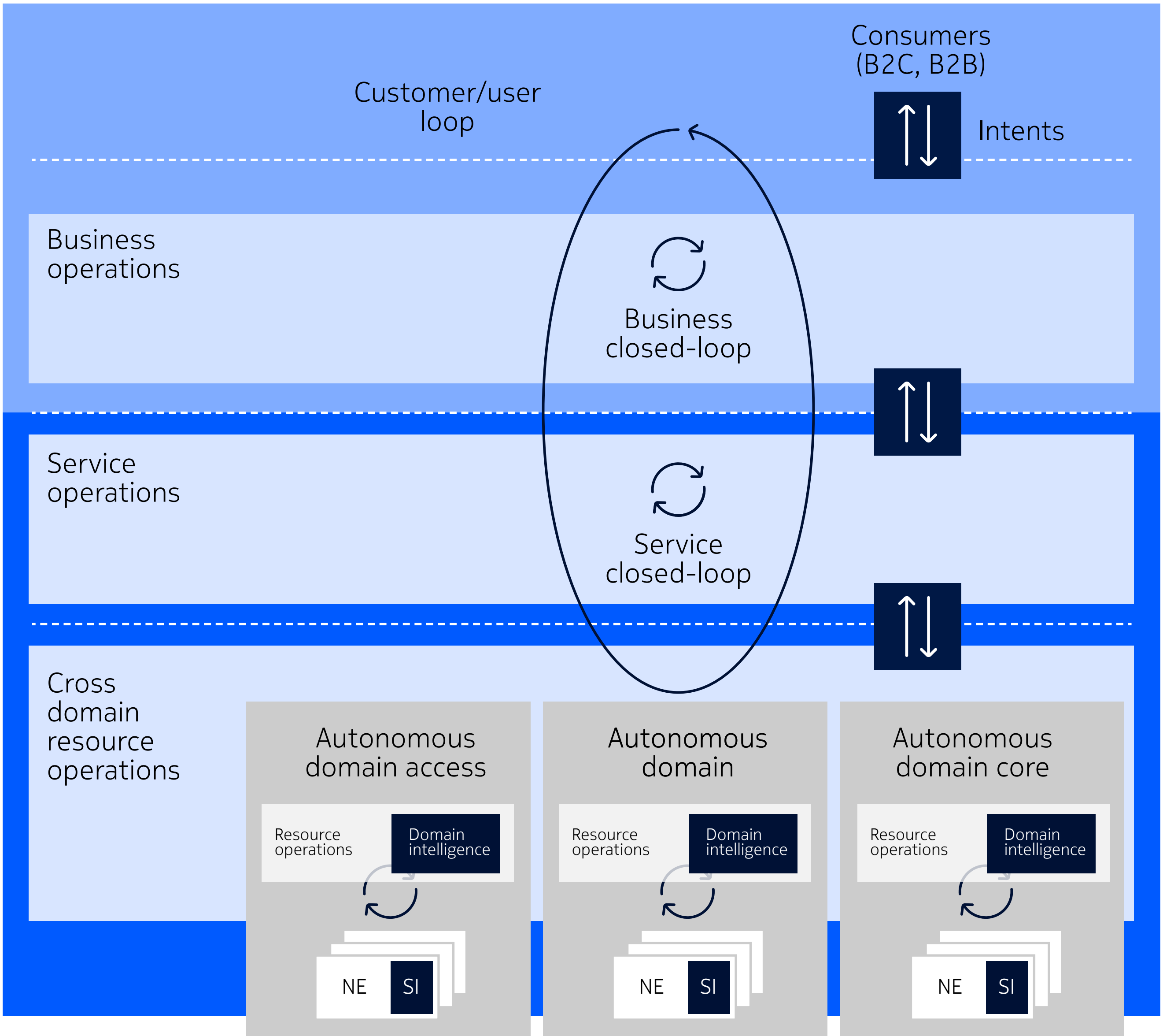
Basics of autonomous networks

ANs are integrated self-optimizing networks that can operate and run with minimal manual intervention and deliver services within reliability thresholds. These networks can deliver and manage services with required scaling and adapt quickly to new business requirements. Autonomous networks are based on three essential principles:

- **Programmable networks:** Dynamic configuration of the network by exposing functionality through open and secure APIs. This ensures automating network and service provisioning, configuration, and lifecycle management of network applications
- **Closed loop automation of network and service operations:** Autonomous management of resources and operations automatically within every domain and across domains.
- **Intent driven:** Understand the business intent by translating it into required policies and configurations as well as starting the course of action to deliver the service that will achieve the expected outcome.

In line with the TM Forum AN blueprint, Bell Labs Consulting recommended a three-layered network architecture for business operations, service operations and resources operation [1]. Each layer is to be managed by a domain-level closed loop and user loop, cutting across all domains (Figure 1).

Figure 1. Bell Labs Consulting blueprint of autonomous networks



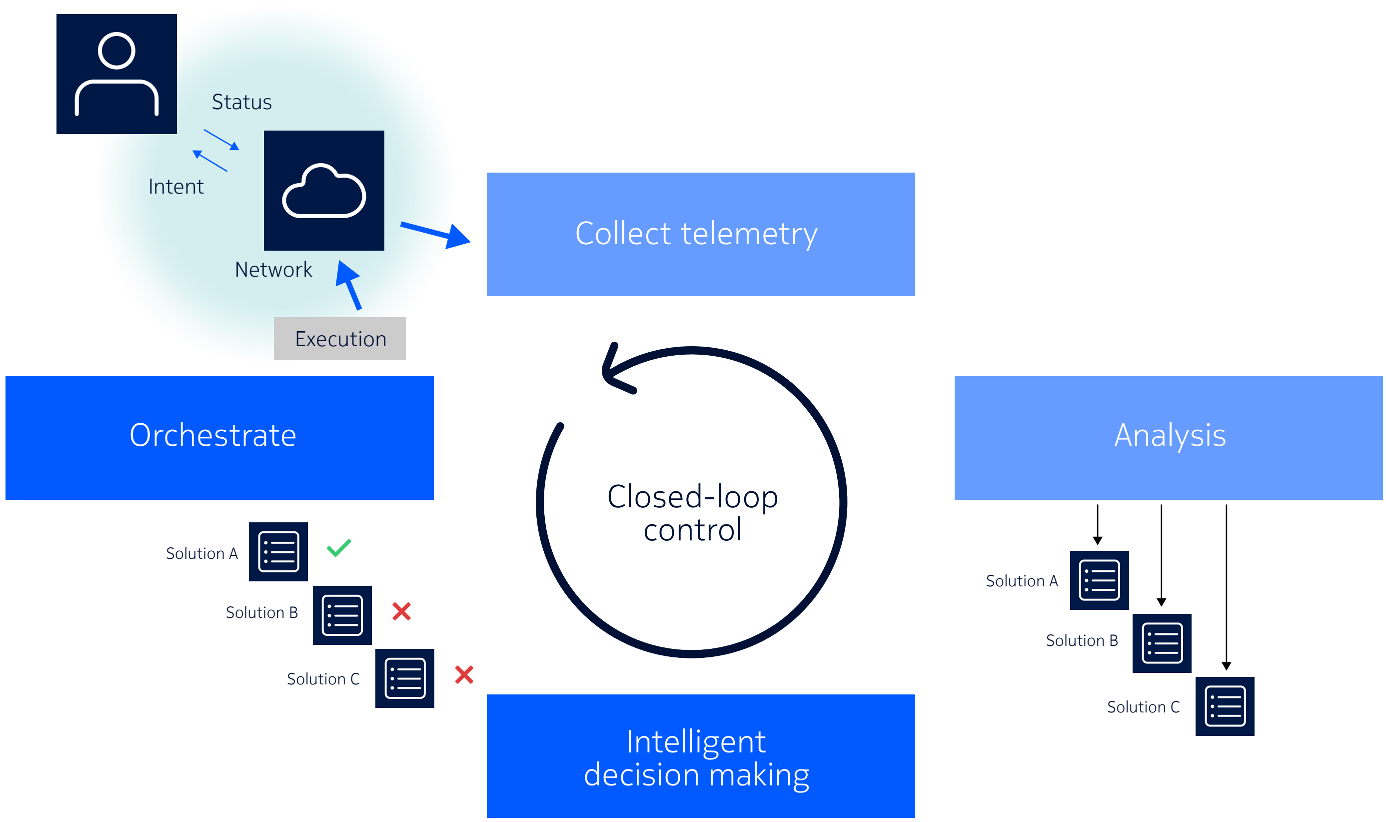
An autonomous domain (AD) represents the lowest granularity in the composition of an autonomous network. It can manage itself and self-heal without human intervention. It interacts with other ADs to deliver the services as per the desired objectives. In line with the AN blueprint, key domains to be considered as part of the framework are access domains, core domains, transport domains, service layer domains and business layer domains. There can be additional ADs depending on the use cases. For example, network slice as a service will require an additional domain defined at the service layer depending on the SLA and policy requirements.

Each of these domains have AI modules working in tandem with the network components and their management systems in real time. Each AI module is responsible for informed decision making on an ongoing basis to predict and prevent unwanted events automated assurance and fulfillment. However, the AI modules are not standalone and require synchronization and coordination. These AI modules, thus, need to be designed and implemented in a federated manner that requires a networked AI/ML architecture across network functions and domains.

AI and data platforms as key enablers

AI plays a central role in the functioning of autonomous networks. These networks utilize vast amounts of data generated by various network domains and use them to train different AI/ML algorithms. Once trained, the algorithms are utilized to make autonomic decisions and optimize network operations. Figure 2 shows the key capabilities required within each domain for closed loop automation.

Figure 2. Cognitive loop per domain for an AN



As the figure indicates, the process starts from sensing and collection of the telemetry data from various sources, including network functions, devices, sensors, and user interactions. This data includes information about network status, performance, traffic patterns, device behavior, and security threats. Data further needs to be processed in real-time to take measures and realize the objectives of self-heal, self-configure and/or self-manage for closed loop automation. Therefore, the underlying data platform should facilitate handling of large volumes of data types generated by the network elements. That is, it should be able to handle what is referred to as the five V's, i.e. volume, velocity, variety, veracity, and volatility.

The data platform should be able to ingest, process, transform, and present data in real time in line with the functionalities required by the AI/ML algorithms embedded in the ANs. At the same time, the data platform should also seamlessly integrate with different sources and ensure timely data accessibility.

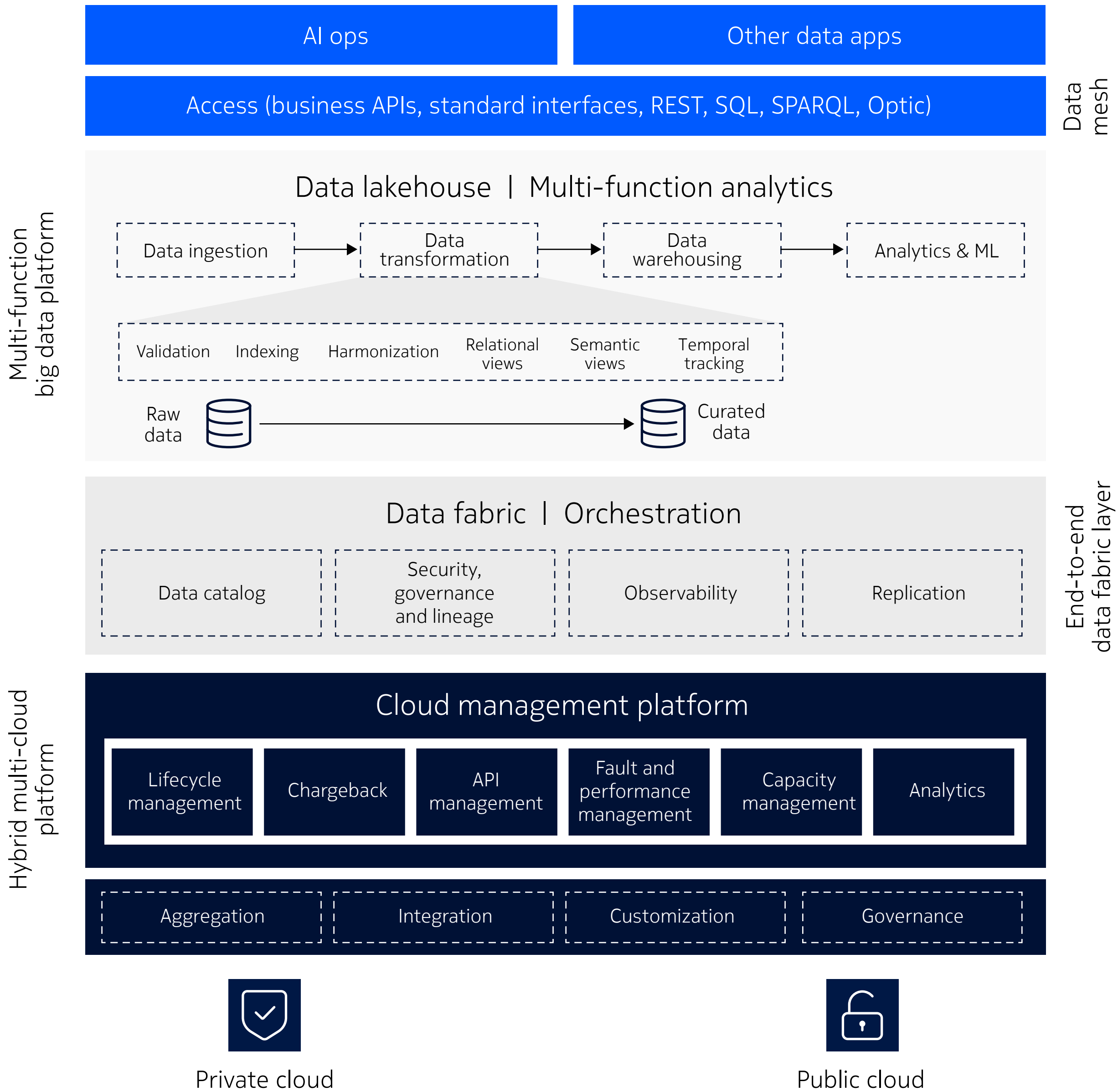
In the absence of a well-defined data platform, the foundation of the AN architecture will be weak, resulting in its inability to meet the expected outcomes of end-to-end automation.

Distributed data architecture for ANs

Traditionally service providers started with centralized data warehouses to enable data-driven decisions following the industry practice of several decades ago. It enabled them to cross-pollinate the data and create a holistic view of customer and business operations to improve various management functions. Faced with the five “V” challenges, data platforms have also evolved from data warehouse to data lake to data lakehouse. While the data platform evolution has improved data reliability and availability, the data operation remains a challenge, as having a central data organization can lead to bottleneck scenarios. The central data organization also has limited proficiency in various domains that otherwise can have the required knowledge to effectively comprehend the underlying context and apply them to the generation of meaningful insights. Isolated domain data management can often result in non-interoperable data silos that prevent identification and generation of interdependent decisions.

The data mesh approach addresses the challenges faced by traditional centralized data warehouse/ lakehouse. Resolution of such issues has paved the path to design and implementation of distributed data architectures. Such a framework allows domains to maintain the decentralized ownership of data and builds a single, virtual management layer atop distributed data to avoid data siloes.

Figure 3. Distributed data platform in a hybrid multi-cloud environment



Following are the key capabilities to implement a distributed data platform in a hybrid multi-cloud environment as shown in Figure 3 [2]:

- **Data mesh approach:** a decentralized, agile approach to enterprise data management based on several core concepts: domain ownership, data as a product, federated computational governance, and self-service design. Data mesh architecture principles were introduced in 2019 by Zhamak Dehghani [3].
- **Efficient, multi-function big data platform:** includes the ability to collect, curate, and store the data in its original format, multi-model, without loss of contextual information; it also includes a multi-model database (MMDB).
- **End-to-end data fabric layer:** a unified, secure platform that connects diverse data repositories to help manage, move, protect, govern, and analyze data residing anywhere in the hybrid cloud.
- **Hybrid multi-cloud enabled data platform:** managed by a centralized provisioning platform that allows deployment of data management services across multiple cloud environments and fully leverages the use of native cloud resources such as compute and storage.

In summary, a distributed data platform enables decentralized data ownership and management within the organization while maintaining the data accessibility across domains through federated governance.

Incorporating data mesh principles in a telco distributed data architecture

Defining data domains

Designing a data mesh-based distributed data architecture for telcos involves implementing a decentralized approach to data ownership and management. Therefore, the first step is to identify the telco data domains.

Data domain definition needs to be aligned with the autonomous domains specified as part of the overall architecture blueprint (as shown in Figure 1) by analyzing the various network components, functionalities and processes involved. The domain definition also requires an understanding of the data requirements of different functional areas and their respective sources.

Figure 4. Data domain for telco ecosystem aligned with AN architecture

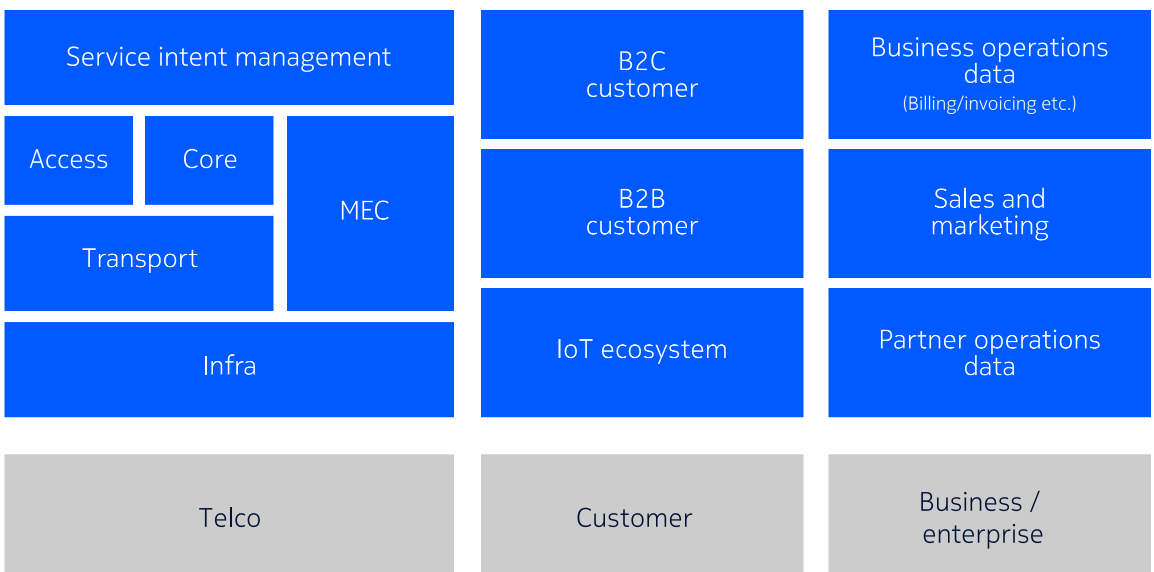


Figure 4 highlights the key categories of the different domains representing the telco ecosystem. The overall AN architecture is layered, hence a few of the data domains will not only be the producers of the data but also the consumers of data from multiple domains. Below are the key categories which are further split into data domains:

- **Telco:** Data related to network configuration, execution and performance, including network availability, latency, bandwidth utilization, and quality of service. In line with the layered network architecture, the domain ownership is defined for access, core, transport, infra, etc. It is also recommended to have a separation between network domain data and service intent data. The service domain collects selective data from different network components to calculate QoS SLA parameters, which are then consumed by the customer operations team.
- **Customer:** Collects, processes and manages customer-related data, including demographics, preferences, usage patterns, customer interactions, experience metrics, revenue per user and subscription data at the customer level. The domain boundaries can be defined based on the customer segment (e.g., B2B vs B2C).

- **Business:** This category handles the data related to business operations including financial data, sales data, product related data and partner data.

With the rise of the internet of things (IoT) and digital enterprise, we recommend defining a separate domain dedicated to IoT data. These are just a few examples of domains that can exist within a telco's data mesh architecture. The specific domains may vary depending on the telco's business model, services offered and strategic priorities.

The domain teams establish clear data scopes and provide data products that meet the specific needs of their respective domains. This empowers teams to make data-driven decisions independently and reduces the reliance on centralized data management. This also allows each team to have control over their own data, ensuring data quality, accessibility and reliability. The data exposure that is based on consumer teams requirements hides the underlying complexity from other teams. As a result, the self-service platform and the common data fabric are critical to the architecture to avoid any potential data siloes and enable seamless data-driven collaboration between different teams and network components.

Data mesh

The concepts of data mesh principles were introduced by Zhamak Dehghani in the paper "How to Move Beyond a Monolithic Data Lake to a Distributed Data Mesh" [3]. The paper brought a key shift in the approach to data ownership. The framework of data mesh based distributed architecture is based on four principles:

- **Domain ownership:** Ownership of the data stays with the domain that generated the data so that the resulting data platform architecture is distributed, and each domain is responsible for creating and maintaining their respective data pipelines
- **Data as a product:** Respective domain teams are also responsible for providing access to the data post-ETL (extract, transform and learn) to other users
- **Self-service design:** The platform is domain agnostic to ensure that data can be consumed as a product between consumers and producers seamlessly
- **Federated computational governance:** A governance framework that includes a set of standards and policies in line with organization policies and regulatory requirements.

By treating data as a product, each team is responsible for defining data products, including data models, APIs and documentation for consumers. The decentralized ownership allows different teams to maintain the necessary context with the data as well as manage the data effectively. Data mesh principles require an underlying platform with unified access to all the data products being exposed by different domain owners. Telcos can deploy a common data fabric layer to enable the principles of data mesh.

Self-service platform: removing siloes through data fabric

A self-service platform in a data mesh architecture is a foundational component that provides the necessary infrastructure, tools and services for domain teams to independently manage, govern and utilize their data products. This platform empowers domain teams by offering capabilities that streamline the processes of data ingestion, processing, storage and consumption without relying heavily on centralized IT or data teams.

The self-service platform in a data mesh approach is designed to support decentralization, enabling domain teams to operate independently by managing their own data ownership and processes. It provides automated tools and services for data operations, such as ingestion, transformation, quality checks and monitoring, and reducing the reliance on centralized IT teams. Ensuring interoperability and standardization across different domain teams, the platform facilitates the sharing and integration of data products.

Additionally, it is built to scale with the growth of data and the number of data products, incorporating robust governance frameworks to maintain data quality, security and compliance without hindering the agility of the domain teams.

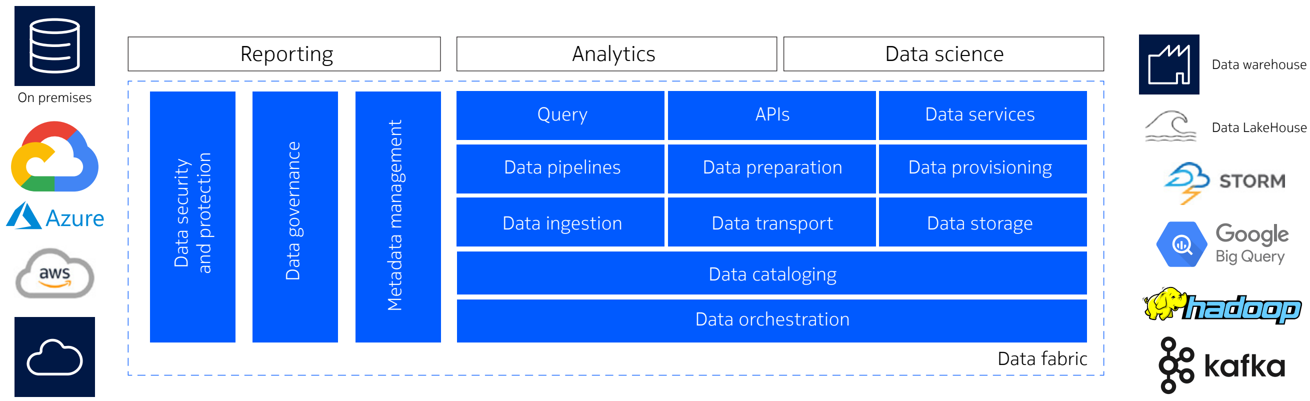
The platform's key functions include tools for data discovery and cataloging, helping teams to document and find their data products efficiently. It provides pipelines and services for data ingestion and transformation from various sources, along with optimized storage solutions for different types of data such as structured and unstructured. Mechanisms are in place to ensure data quality and track data lineage, which facilitates auditing and troubleshooting. APIs and query interfaces enable seamless data access and consumption by various users and applications; while monitoring and observability tools help in overseeing data pipelines, detecting anomalies and maintaining the health of data products.

The benefits of a self-service platform in a data mesh approach are significant. It empowers domain teams to innovate and respond swiftly to changing business needs, enhancing agility and promoting innovation by providing the necessary tools for independent data management. By decentralizing ownership and responsibility, it improves data quality and governance. The platform supports the scaling of data operations across the organization, enabling domain teams to handle their data management tasks effectively. Furthermore, it reduces dependency on centralized IT resources, potentially lowering costs and increasing operational efficiency.

End-to-end data fabric

An end-to-end data fabric supports a self-service platform by providing a comprehensive and integrated architecture that enables seamless access, management and utilization of data across the organization. It is critical in a data mesh architecture for a unified and integrated data architecture that enables seamless data access, sharing and analysis across various systems and applications. It is essentially a metadata-driven way of connecting a disparate collection of data tools that address key pain points in big data projects in a cohesive and self-service manner as shown in Figure 5. It acts as an abstraction layer for data management and provides telcos the capabilities in the areas of data access, discovery, transformation, integration, security, governance, lineage, and orchestration. By providing a way to bring these data management capabilities to bear on data spanning all these silos, a data fabric can help alleviate core data management challenges that are holding companies back from higher-level data use cases, including advanced analytics and AI in the cloud.

Figure 5. Overview of big data fabric



A data fabric supports [4]:

Unified data access and integration: An end-to-end data fabric connects disparate data sources, both on-premises and in the cloud, into a single interface, enabling domain teams to easily discover, access and utilize data without navigating multiple systems. This unified access allows teams to focus on leveraging data rather than managing its complexity.

Automation and orchestration: The data fabric automates data operations like ingestion, transformation, quality checks and pipeline orchestration, thus reducing manual effort and errors. This automation ensures domain teams have reliable and up-to-date data, enhancing their ability to perform analytics and generate insights independently.

Data governance and security: A robust data fabric incorporates governance frameworks that enforce data quality, security and compliance, providing visibility into data lineage and facilitating auditing. This governance ensures domain teams have access to trusted and compliant data, allowing them to manage and use data confidently without compromising on governance requirements.

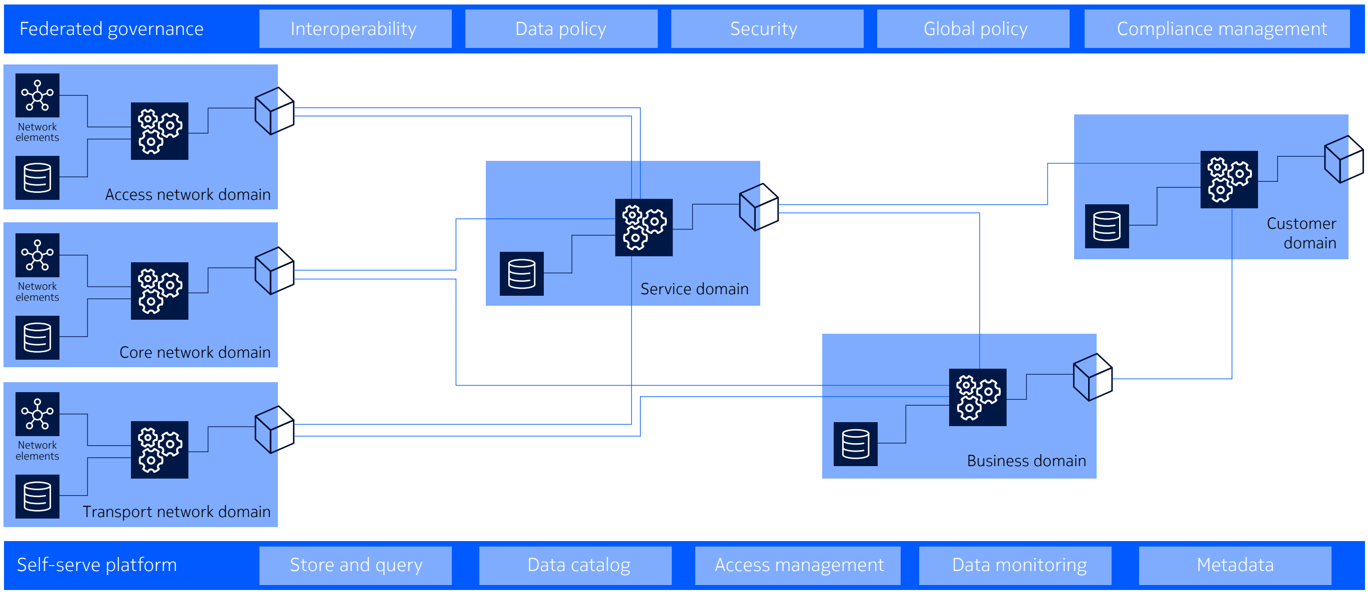
Scalability and performance: Designed to handle large data volumes, the data fabric scales seamlessly as the organization grows, optimizing storage and processing for high performance and low latency. This scalability allows domain teams to manage increasing data and complex workloads efficiently, maintaining productivity and efficiency.

User-friendly interfaces and self-service tools: Data fabrics come with user-friendly interfaces and self-service tools for data cataloging, preparation and analytics. These accessible tools empower domain teams to manage data products, perform analytics and generate insights independently, aligning with the principles of a self-service platform.

Monitoring and observability: Effective data fabrics include monitoring and observability tools that provide real-time insights into data pipelines and processes, helping detect anomalies and ensure smooth operation. These capabilities enable domain teams to proactively manage and troubleshoot data operations, ensuring continuous and reliable data access.

With a singular data fabric virtually overlayed atop the various data repositories across multiple domains, telcos can bring some semblance of unified management to the disparate data sources and downstream consumers, including data stewards, data engineers, data analysts and data scientists (see Figure 6).

Figure 6. Data fabric enabling a self-service data platform



For example, in a telco ecosystem, customer experience is driven by multiple parameters such as network quality and service level performance parameters as well as the business parameters providing billing and usage transparency [5]. Customer care personnel trying to resolve a complaint require data from different domains including network, service and business but not hundreds of network parameters being pushed by different network entities every minute; they only need aggregated data at the customer level. Data fabric streamlines the data analysis and access to the care personnel allowing them to make faster and more informed decisions to resolve customer complaints.

A data fabric provides data owners with capabilities for data asset creation, i.e., cataloging data assets and transforming assets into reusable services that comply with federated governance policies. Therefore, it also needs to have capabilities to publish data assets to the catalog, search and find data assets, and query or visualize data assets by leveraging data virtualization or using published APIs to transform assets into reusable services that comply with federated governance policies.

Data lakehouse

A common analytical approach involves using multiple systems such as a data lake, several data warehouses, and other specialized systems—leading to three main problems:

- **Expensive data movement with dual architecture:** Over 90% of analog/IoT data is stored in data lakes due to their flexibility and low-cost storage. However, to address the data lake's performance and quality issues, enterprises use ETL (Extract/Transform/ Load) to copy a subset of data to a downstream data warehouse for decision support and BI applications. This dual system requires continuous engineering, which is costly and risks data quality issues due to potential ETL failures or bugs.
- **Limited support for machine learning:** Leading ML systems like TensorFlow, PyTorch, and XGBoost do not work well on top of data warehouses. Unlike business intelligence (BI), which extracts small amounts of data, ML systems process large datasets using complex non-SQL code, making data warehouses

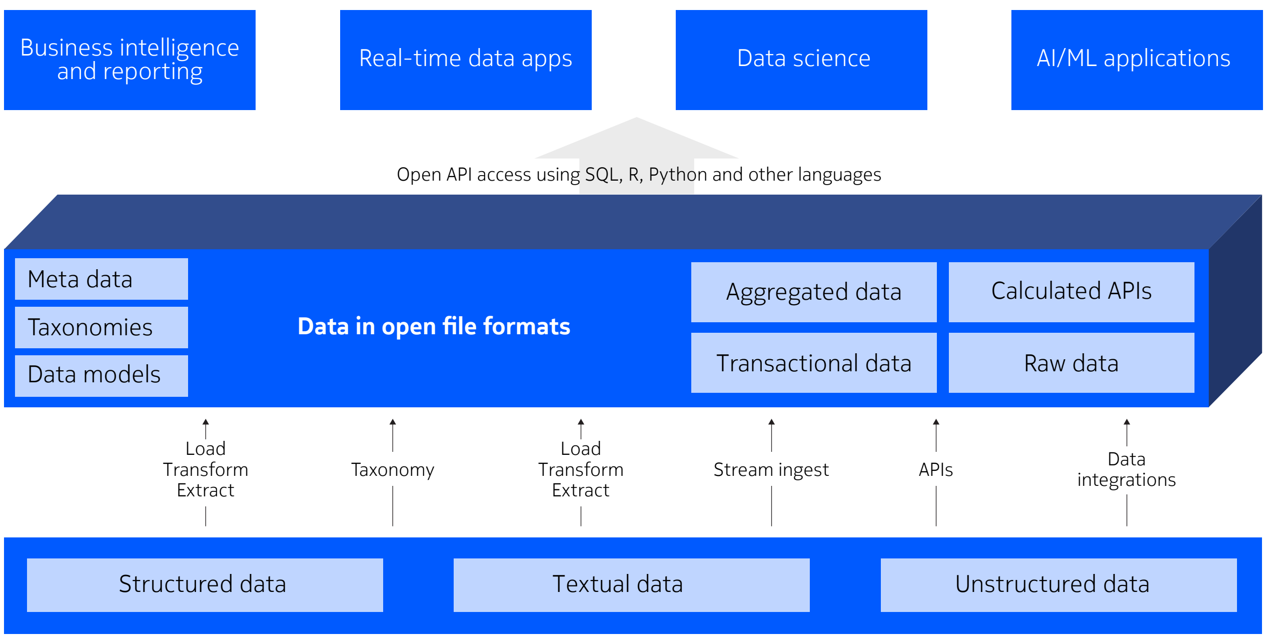
inefficient for ML tasks.

- **Lack of openness:** Data warehouses often lock data into proprietary formats, increasing migration costs and limiting the ability to run other analytics engines, such as machine learning systems, against the data.

From these challenges emerges a new class of data architecture called the data lakehouse [6] as shown in Figure 7, which includes:

- Data from the structured environment
- Data from the textual environment
- Data from the analog/IoT environment
- An analytical Infrastructure.

Figure 7. Data lakehouse architecture overview



This allows data in the lakehouse to be read and understood, implementing similar data structures and management features to those found in data warehouses but operating directly on low-cost storage used for data lakes. The open and standardized system design enables efficient analysis of analog/IoT data while leveraging the low-cost storage and flexibility of data lakes.

The foundational lakehouse logical layer differentiates a data lake from a data lakehouse. While data lakes are suitable for ML, they lack ACID (atomicity, consistency, isolation, and durability) transactions, essential for fast queries during simultaneous ETL. ACID transactions in the lakehouse ensure data integrity during operations, enabling streaming ETL for real-time reporting. For training ML models, data is exported from warehouses to distributed storage in open file formats, ensuring reproducibility and efficient data management.

The data lakehouse architecture presents an opportunity comparable to the one seen during the early years of the data warehouse market. The unique ability of the lakehouse to manage data in an open environment, blend all varieties of data from all parts of the enterprise, and combine the data science focus of the data lake with the end user analytics of the data warehouse will unlock incredible value for organizations.

The data lakehouse integrates various types of data, each with unique physical characteristics. It includes structured data (transaction-based), textual data (conversations and written text), and other unstructured data (analog, IoT, and machine-generated). The volume of data varies across these environments, with structured data being the smallest, textual data larger, and other unstructured data the largest.

Data lakehouse and autonomous networks

Data lakehouse is well suited for supporting the needs of autonomous networking in the telecom industry for several reasons [7]:

- **Unified data management:** The data lakehouse integrates structured, semi-structured, and unstructured data into a single platform. This unified approach is essential for telecom networks, which generate diverse data types, including call records, network logs, customer interactions, and IoT sensor data. The data lakehouse enables seamless data ingestion, storage and processing from various sources.
- **Scalability and performance:** Telecom networks handle massive volumes of data, and a data lakehouse offers the scalability needed to manage this data efficiently. With its support for distributed storage and computing, a lakehouse can scale horizontally to accommodate growing data needs while maintaining high performance for analytics and real-time processing.
- **Advanced analytics and machine learning:** Autonomous networking relies heavily on advanced analytics and ML to optimize network performance, predict failures, and automate decision-making. The data lakehouse supports these needs by providing a robust platform for data scientists and engineers to build, train and deploy machine learning models. The ability to run complex queries and analytics on large datasets in a timely manner is crucial for real-time network management.
- **Real-time data processing:** The telecom industry requires real-time data processing capabilities to monitor and manage network operations effectively. A data lakehouse supports streaming data ingestion and real-time analytics, enabling telecom operators to detect and respond to network anomalies, security threats and performance issues as they occur.
- **Data governance and compliance:** Telecom companies must adhere to stringent data governance and regulatory requirements. A data lakehouse provides comprehensive data governance features, including access controls, data lineage and compliance with industry standards. These features ensure that data is managed securely and transparently, meeting regulatory obligations.
- **Cost efficiency:** By leveraging open file formats and cloud-based storage, a data lakehouse can offer a cost-effective solution for storing and processing large volumes of data. This cost efficiency is particularly important for telecom companies, which must balance the need for extensive data analysis with budget constraints.
- **Interoperability and flexibility:** The open architecture of a data lakehouse ensures interoperability with various data tools and technologies. This flexibility allows telecom operators to integrate their existing systems and tools, ensuring a smooth transition and maximizing the value of their data assets.

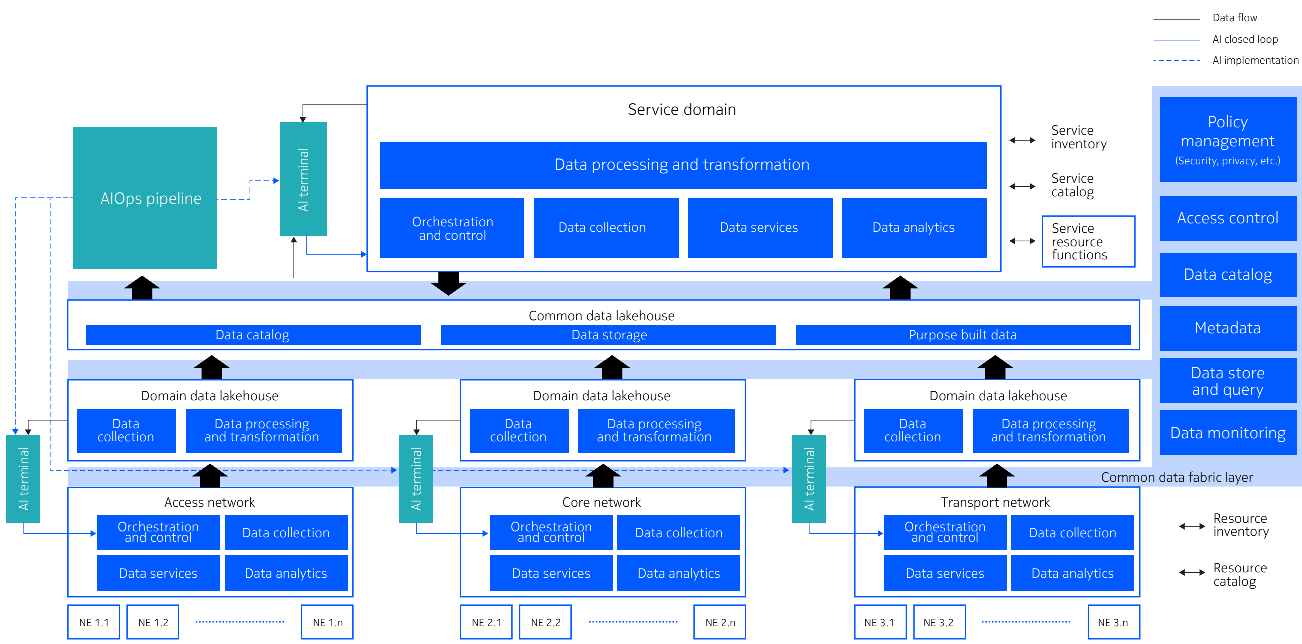
In summary, a data lakehouse provides the scalability, performance, real-time processing, advanced analytics, data governance, cost efficiency, and flexibility needed to support the complex and dynamic requirements of autonomous networking in the telecom industry.

Blueprint of distributed data platform aligned with layered telco architecture

Telcos take a structured approach in designing their network architecture. They organize different domains and services in layers, thus the term layered architecture. Managing a layered architecture requires the data platform to not only manage data ingestion from different sources across the layers but also the scale and complexity of the data.

A distributed data platform is capable of handling requirements of a telecommunication network infrastructure in an efficient way. It ingests data from different data sources across network, service and business domains. Figure 8 provides an overview of a distributed data platform for a layered telco architecture.

Figure 8. Blueprint of distributed data platform for telcos



The data ingestion supports real-time streaming data as well as batch data processing. Each domain manages their distributed data processing and transformation to generate actionable insights. The processed data is also exposed to other domains for consumption as required through the common data fabric layer. The platform also has distributed data storage managed by the domain owners.

Each domain shares their process data to a common data lakehouse not only for storage and cataloging but also to store purpose-built data. The purpose-built data is for the AI modules used by other network functions. A distributed data architecture also improves scalability and resiliency.

Each domain layer also has their AI terminals working with the network objects at runtime leveraging the data collected and processed. Each AI module has MLOps connecting the source of data via the data platform. In an autonomous domain, the AI terminal operates in real time based on data received from the domain and/or outside domain to continuously monitor and improve network performance in a closed loop manner. These AI terminals not only operate on the data received from their own domain but also on other data domains through the common data fabric layer.

The successful evolution towards a distributed data platform and becoming a data-driven organization also requires a cultural transformation in line with a data mesh philosophy. Federated governance plays a significant role in a distributed data platform by ensuring effective management and control of data across multiple entities or organizations.

Federated governance

Federated governance in a data mesh involves distributing the responsibilities of data management and quality assurance to domain teams while ensuring overarching policies and standards are met. The core principles include:

- **Decentralized data ownership:** Each domain owns its data products, maintaining control over data quality, compliance and security within their boundaries
- **Global standards and policies:** Universal standards for data interoperability, security and governance, to which all domains must adhere, ensures consistency across the organization
- **Automated compliance and quality controls:** Implementing automated tools and processes to enforce compliance and data quality standards reduces the manual overhead and potential for human error
- **Collaboration and coordination:** Facilitating collaboration among domain teams to share best practices and coordinate between cross-domain data initiatives.

Application to telecommunication autonomous networks

Telecom autonomous networks are complex systems comprising various technologies from traditional 3G/4G networks to advanced 5G and emerging 6G networks. The autonomous nature of these networks requires a robust and scalable data governance framework to handle the dynamic and distributed data landscape. Federated governance in a data mesh can be particularly beneficial in the following areas:

- **Network data management:** Telecom networks generate massive amounts of data from different sources such as base stations, core networks and user devices. Federated governance enables each network segment (e.g., RAN and core) to manage its data autonomously while adhering to global standards for data sharing and interoperability. This ensures data consistency and reliability across the network.
- **Service automation and optimization:** Autonomous networks leverage AI and machine learning algorithms to optimize network performance and service delivery. Federated governance supports the development and deployment of AI models by ensuring access to high-quality, standardized data across domains. This decentralized approach allows for faster innovation and more effective optimization strategies tailored to specific network segments.
- **Security and compliance:** Telecom networks are subject to stringent regulatory requirements and security standards. Federated governance helps in implementing automated compliance checks and security protocols at the domain level, ensuring that all data products meet regulatory standards without centralizing control. This approach enhances security and compliance while allowing domains to operate independently.
- **Customer experience:** Providing a seamless and high-quality customer experience is critical for telecom operators. Federated governance facilitates the integration of customer data across various touchpoints (e.g., customer service, network usage and billing) to create a unified view of the customer. This holistic perspective enables personalized services and proactive issue resolution, enhancing overall customer satisfaction.

Federated governance in data mesh offers a scalable and flexible framework for managing data in telecom autonomous networks. By decentralizing data ownership and enforcing global standards, telecom operators can achieve a balance between autonomy and coherence, driving innovation and operational efficiency. As telecom networks continue to evolve, adopting federated governance will be key to harnessing the full potential of autonomous networks and delivering superior services to customers.

Conclusion and further discussion

A robust data platform is essential in enabling autonomous networks. The data platform should be able to handle large volumes of diverse datasets without exposing the complex nature of underlying data sources from different network and business applications. A distributed data platform based on data mesh principles can support the required data scalability and accessibility. The domain ownership ensures that the data context understanding is not lost. At the same time, a common data fabric avoids the creation of data siloes.

However, the data platform is just one of the aspects needed to evolve towards autonomous networks. Telcos as organizations also need to shift towards being data-driven digital enterprises. This requires realigning organization structures and operating models with the data mesh principles of domain ownership and data products.

Data platform enables different types of AI modules working as an integrated part of different network and service components to automatically take corrective and preventive decisions to achieve desired results. The distributed data platform achieves the requirement to ingest, curate and store data from different domains to provide inputs to different data modules. This solves one part of the problem, which requires managing large volumes of data flow and different data types from different data sources. The larger question which remains is how to manage the policies and different intent-based services.

The solution lies in operating AI modules in coordination, interacting with different components where the AI governance framework defines the policies to resolve the conflicts arising from actions focused on local versus global optimization. The AI governance framework maintains the larger context under which different AI modules function.

Once AI governance policies are defined and standardized, they should streamline and accelerate telcos' journey towards autonomous networks. Telcos need to start building the foundation by implementing distributed data platforms for a smooth transition.

Abbreviations

ACID	Atomicity, consistency, isolation, and durability
AD	Autonomous domain
AI	Artificial intelligence
AN	Autonomous networks
BI	Business intelligence
CSP	Communications service providers
ETL	Extract, transform and learn
IoT	Internet of things
MMDB	Multi-model database
ML	Machine learning
NDTs	Network digital twins
QoS	Quality of service
RAN	Radio access network
SLA	Service level agreement
SQL	Structured query language

References

[1] TM Forum, “Autonomous Networks Technical Architecture (IG1230), TM Forum web site, Dec. 2022.” Online: https://www.tmforum.org/resources/toolkit/autonomous-networks-technical-architecture

[2] Adapted from Behan, A., “Modern data architecture for telecommunications,” Cloudera blog post, Sep 2022. Online: https://blog.cloudera.com/modern-data-architecture-for-telecommunications/

[3] Dehghani, Z., Data Mesh - Delivering Data-Driven Value at Scale, O Reilly, Mar 2022. Online: https://www.oreilly.com/library/view/data-mesh/9781492092384/

[4] Astorino, S., Simmonds, M. and Sanders, R., Data Fabric - An Intelligent Data Architecture for AI, MC Press Online, LLC, 2022. Online: https://mc-store.com/products/data-fabric-an-intelligent-data-architecture-for-ai

[5] Bhagchandani, N., “Exploring the Synergy of Data Mesh and Data Fabric,” Incedo blog post, Mar 2023. Online: https://www.incedoinc.com/exploring-the-synergy-of-data-mesh-and-data-fabric/

[6] Haelen, B. and Davis, D., Delta Lake: Up and Running: Modern Data Lakehouse Architectures with Delta Lake, O’Reilly Media, Inc., Oct 2023. Online: https://www.oreilly.com/library/view/delta-lake-up/9781098139711/

[7] Inmon, B., Levins, M., and Srivastava, R., Building the Data Lakehouse, Technics Publications, Oct 2021. Online: https://technicspub.com/data-lakehouse-collection/

About Nokia

At Nokia, we create technology that helps the world act together.

As a B2B technology innovation leader, we are pioneering networks that sense, think and act by leveraging our work across mobile, fixed and cloud networks. In addition, we create value with intellectual property and long-term research, led by the award-winning Nokia Bell Labs.

With truly open architectures that seamlessly integrate into any ecosystem, our high-performance networks create new opportunities for monetization and scale. Service providers, enterprises and partners worldwide trust Nokia to deliver secure, reliable and sustainable networks today – and work with us to create the digital services and applications of the future.

Nokia OYJ
Karakaari 7
02610 Espoo
Finland
Tel. +358 (0) 10 44 88 000

Document code: CID214288 (October)