

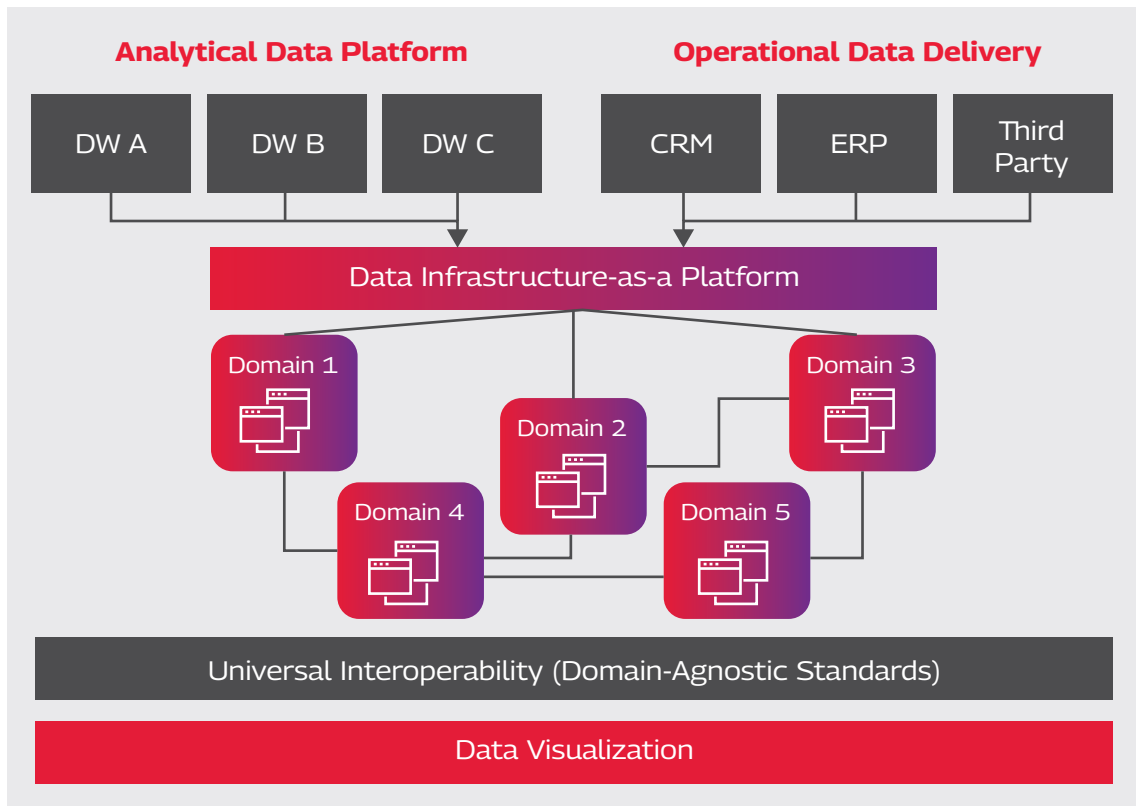
Whitepaper

# Data Mesh - Building a Scalable and Interoperable Data Ecosystem



## Abstract

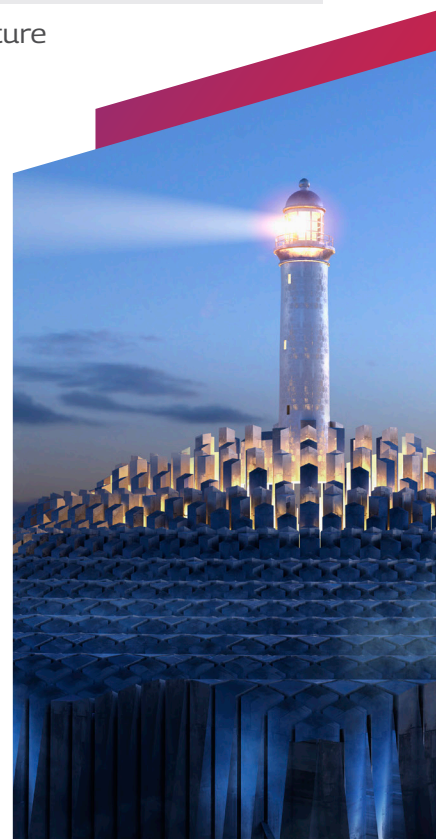
Data mesh essentially refers to the concept of breaking down data lakes and siloes into smaller data domain-specific sets with a self-serve design, to enable data-driven decisions using data products and providing scalable delivery of data with a flexible data governance model. Data mesh is a decentralized approach to sharing, accessing, manage analytical data in a scalable way.



A high-level data mesh logical architecture

Data is collected from analytical and operational platforms and collected in a data lake. Our point of view on data mesh encompasses the following principles:

- Rather than going through a centralized technical team the business domain should own the data end-to-end
- A self-service platform is fundamental to a successful data mesh architectural approach
- Data as a product
- Data products must be built with governance and compliance that is automated and federated



## Challenges and Best Practices



### Data Mesh Challenges

- Effective insights from huge volumes of telemetry and internet of things (IoT) data to improve customer experience and retention
- Managing regional compliance and security standards for multinational organizations specifically in BFSI, retail, and telecommunication sector
- Enabling advanced analytics for each stakeholder group as per individual use cases
- Availability of timely, consistent, and accurate data for robust and reliable forecasting analytics



### Data Mesh Best Practice

- Use case-based data products - create consumption-oriented use case-based data products to optimize data usage and reduce data duplicity with an eye on performance and scalability
- Hybrid and agile data governance - Federated governance across data products with a touch of centralized data governance for key areas such as data security, data quality and data ingestion framework
- Master data management (MDM) and Data quality as a service - Provide MDM and data quality processes as a centralized services to maintain data consistency across enterprise





## Key Use Cases



### Regional Compliance and Regulations

Establish individual data products for regional compliance and operational data (multilingual). For example, multinational retail or finance organizations can address their local compliance using a data product specific to their local region. The 'federated' aspect of data mesh means that governance policies are still defined centrally, but they are implemented by each domain team, which is responsible for making all the data that is produced in its domain and discoverable to other domains.



### Sales and Marketing

Demographic based sales and demand forecasting. For example, usage reports for video streaming organizations customer behavior analysis for retail.



### Federated Enterprise Governance

Federated data governance for key data elements such as customers and products to enable independent data and business processes. For example, customer data platform (CDP) for banking customers, B2B & B2C inventory management for individual regions for global manufacturer.



### Manage Excessive Data Growth

Managing operational analytical data for large organization with a very high data, for example, internet of things (IoT) data from sensors and equipment can be massive. Now instead of putting a global operations team, organizations can create data products and assign owners specific to the logical Data products for better managing the operations. Data mesh architecture helps in essentially decoupling the computational governance from the domain ownership where the domain owners really worry about "shipping" the data product while the infrastructure team can scale for the additional computational capacity as needed. A similar use case could be on traffic data analysis for large cities.

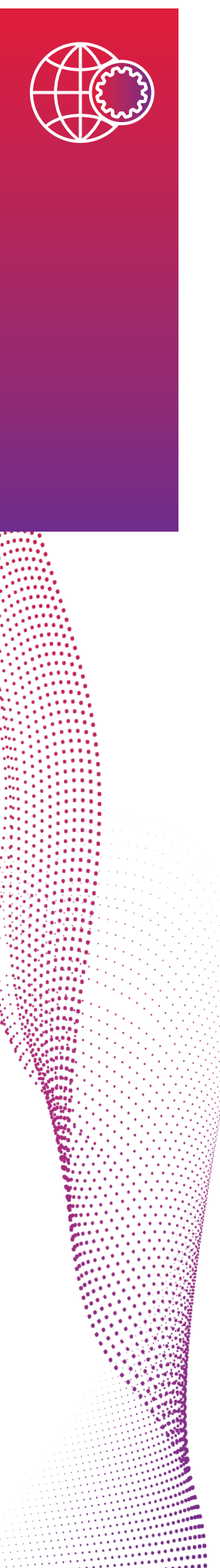




### Data Mesh Empowers Data Ownership for Data Scientists and Analysts


Data mesh enables data scientists and data analysts get access to higher quality data. Data scientists and data analysts suffer when data is incomplete, incompatible, or outdated. Data mesh addresses this problem by keeping data with the groups or teams that know it best. Once they become responsible for the data, and for making it easily and readily available to other groups within the organization, data analysts and data scientists end up working with higher quality data. This, in turn, yields more accurate results.


## Our Solution


Following are the implementation journey, which starts with readiness assessment and could be summarized in a 9-step approach as shown below not necessarily in the exact sequence but should be aligned to the steps.


- 
- 


**01. Readiness Assessment**  
Data maturity level, data sources/domains, and governance framework
  - 


**02. Data Product Assessment**  
Feasibility, cost vs ROI, and KPI
  - 


**03. Architecture Road Map**  
End state definition, transition state definition, adaptive global source/subscriber/third party onboarding
  - 

**04. Federated Data Governance**  
Governance and standards, automation and computation, domain specific governance
  - 

**05. Design of Data Product**  
Code, infrastructure, data and metadata, modelling (including DP relationships)
  - 

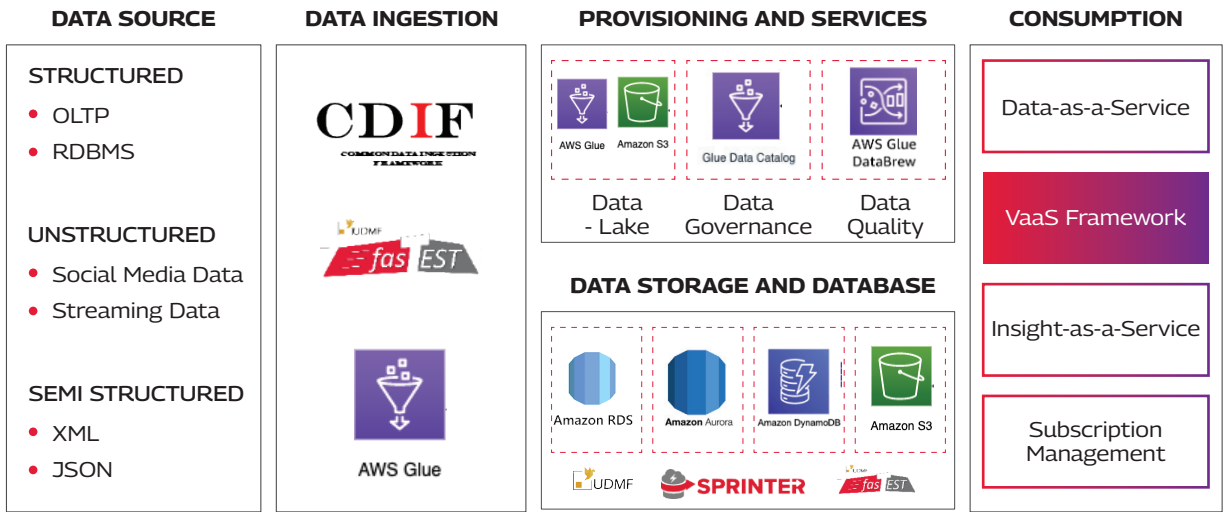
**06. Implementation Methodology**  
Agile/Waterfall methodology, plan, CRP/demo-based checkpoints, Data-Ops
  - 

**07. Data and Integration Infra**  
Data storage, data processing optimization plan, and data growth computation
  - 

**08. Visualization**  
Global/data validation and quality rules, databases and file data extraction, domain specific, and global UI platform
  - 

**09. AI/ML Platform**  
Adaptive and augmented governance and metadata management

# Tech Mahindra and AWS Joint Solution Enabling Technology Mapping



## Key AWS Components

- **Data Lake** - Provide a data storage and analytics solution
- **Amazon RDS** - Commercial and open-source Databases
- **Amazon Aurora** - MySQL and PostgreSQL compatible
- **Amazon DynamoDB** - Non relational database service
- **AWS Glue** - Data quality and integration service

**CDIF**  
COMMON DATA INGESTION FRAMEWORK

Self Service Ingestion Framework for Modern Data Analytics Platforms

- Metadata Driven Automated Ingestion

**UDMF**

Data Migration & Quality Platform

- DQ assessment, consulting & recommendation
- Data cleansing, standardization & reconciliation

**SPRINTER**

Any Source to Any Destination Cloud Migration Accelerator

- Data Search & Discovery
- Schema & SQL Migration

**InfoWise**  
Enterprise metadata 360  
Augmented by AI

Self Service Ingestion Framework for Modern Data Analytics Platforms

- Data migration assessment and strategy
- Report Rationalization



## Benefits



### Automation

Global API and batch based framework with data register for source and subscriber onboarding, adaptive governance, metadata graph and adaptive process automation.



### Flexible Integration Framework

Enable global and data product framework to automatically share data from individual and across data products in both real time (streaming) and batch mode (global ETL framework).



### Model Data as Products

Create data products based on specific domains, consumer groups or business lines (e.g., sales, services, finance, engineering, analytics) that is self-described and connected.



### Data as a Service

Provide reliable, performance intensive data as a service to consumers across platforms, format and medium.



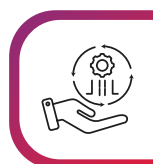
### Self Service

Automated integration for operational and analytical applications using data registry and global API framework.



### Reliable Data and Self-Heal

Data quality (DQ) as a service based on governance policies with robust error framework along with use case based self-healing capability.



### Seamless Performance

Scalable and monitored seamless performance through aligned reporting needs to data product models.



## Authors



### Rajeev Dubey

Rajeev Dubey is the CoE Head for data management in Data and Analytics competency at Tech Mahindra



### Monoj Roy

Monoj Roy is a Global Data Management Leader in Data and Analytics competency at Tech Mahindra



### Sathish Arumugam

Sathish Arumugam is a Sr. Partner Solution Architect at Amazon Web Services.



### Piyush Patra

Piyush Patra is a Partner Solutions Architect at Amazon Web Services.



### Kalyan Kumar Neelampudi

Kalyan Kumar Neelampudi (KK) is a Specialist Partner Solution Architect (Data & Analytics) at AWS.

