



DATA Mesh & DATA Fabric– A Joint Solution TechM & Microsoft

DEC 2022

Key Challenges to moving to Data Mesh & Fabric

Siloed Data and Analytics Solutions

Reduce cost of data engineering

Data and Analytics Operationalization

Poor data quality

Need for Frictionless Data Governance

Too slow moving from data to decision

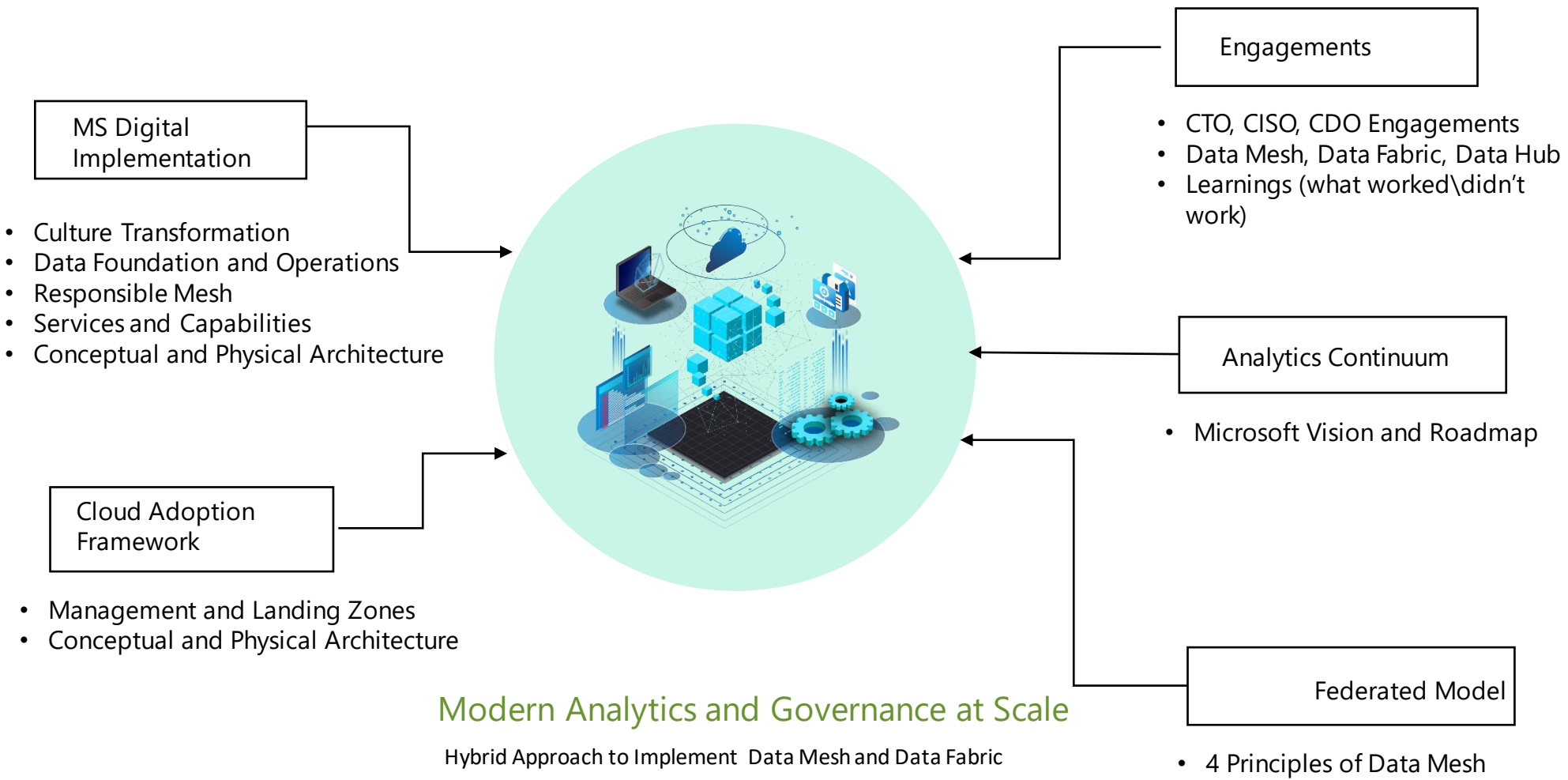
Difficult to balance access and data protection

Enable Lines of Businesses






Unified ecosystem

Project prioritization

Our Hybrid Approach in Alignment to Data Mesh and Data Fabric



Ideally, organizations want to have....

-  Platform to actionable Insights to the business
-  Robust data governance
-  Ability to increase the value of hidden data
-  To spend less time preparing data
-  Improve Operational Efficiency

Modern Data Architecture Hybrid Approach – Joint Solution



Focuses on organizations challenges and strategic vision vs landing Framework



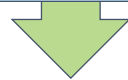
Provide the North Star for Implementing Modern Analytics Ecosystem in alignment to People, Process and Technology



Data Mesh Implementation with a robust Framework

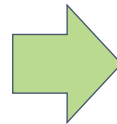
Assessment

- Readiness Assessment** – Data Maturity Level, Data Sources/Domains, Governance Framework, Infrastructure, Cost, Skill availability
- Data Product** – Feasibility, Cost vs ROI, KPI
- Organization Structure** – Federated Data Organization, Domain users and skill availability, Prioritized key business process improvement plan
- Architecture rationalization** – Comparison of requirement to fitment of right methodology – Data Mesh vs Data Fabric vs Data Virtualization



Define Roadmap & Design

- Data Product** – Code, Infrastructure, Data & Metadata, Modelling (including DP relationships)
- Federated Data Governance** – Global governance & standards, Automation & Computation, Domain specific Governance
- Data Sharing and Ingestion (IDP & ODP)** – Multi protocol, medium & formats, CDC/Streaming, Performance, Automated Error detection & resolution framework



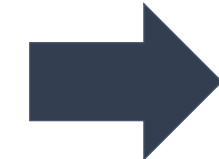
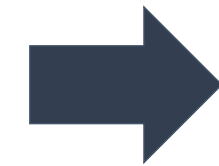
Implementation & Governance

- Business Value Alignment** – KPI centric value realization framework, Iterative KPI centric Rollout
- Implementation Methodology** – Agile/Waterfall methodology, plan, CRP/Demo based checkpoints, Data-Ops
- Architecture Road Map** – End state definition, transition state definition, Adaptive Global source/subscriber/third party onboarding
- Data Organization setup** – Data domain Organization, Global Governance Organization, Stakeholder & Dependency planning



Tool, Process & Infra Assessment and Selection

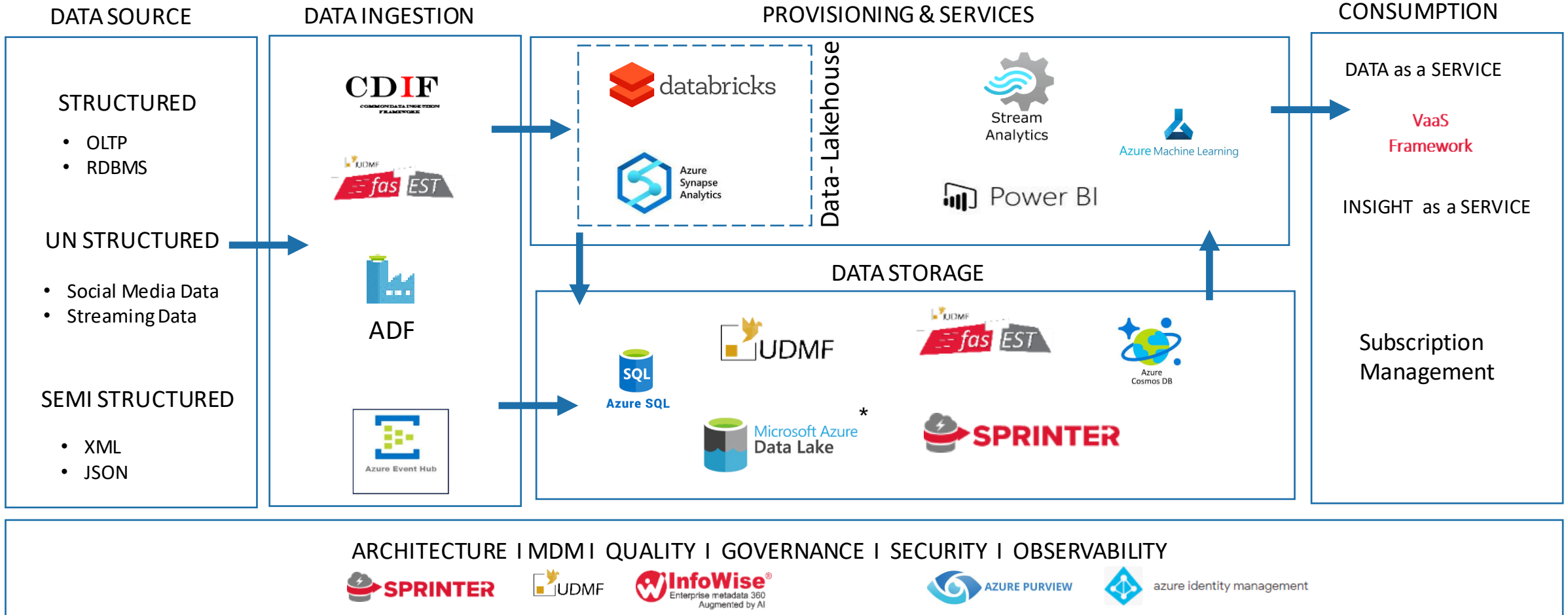
- Data Infra** – Data Code, Data Storage, Data processing optimization plan, Data growth computation
- Integration Infra** – Connectors, Streaming and Batch data integration tools, Transformation tools, error reconciliation
- Visualization** – Global/Data validation and quality rules, Databases and File data extraction, Domain specific and Global UI platform map



Value Realization

- Data as a Product** (Icon: Box with data points)
- Self Serve** (Icon: Network of nodes)
- Federated Governance** (Icon: Circular arrows with a plus sign)
- ML-Enabled Pipeline and Data Quality Services** (Icon: Brain with circuitry)
- DataOps** (Icon: Circular arrows with DATA and OPS labels)

Joint Solution (Technology Mapping)

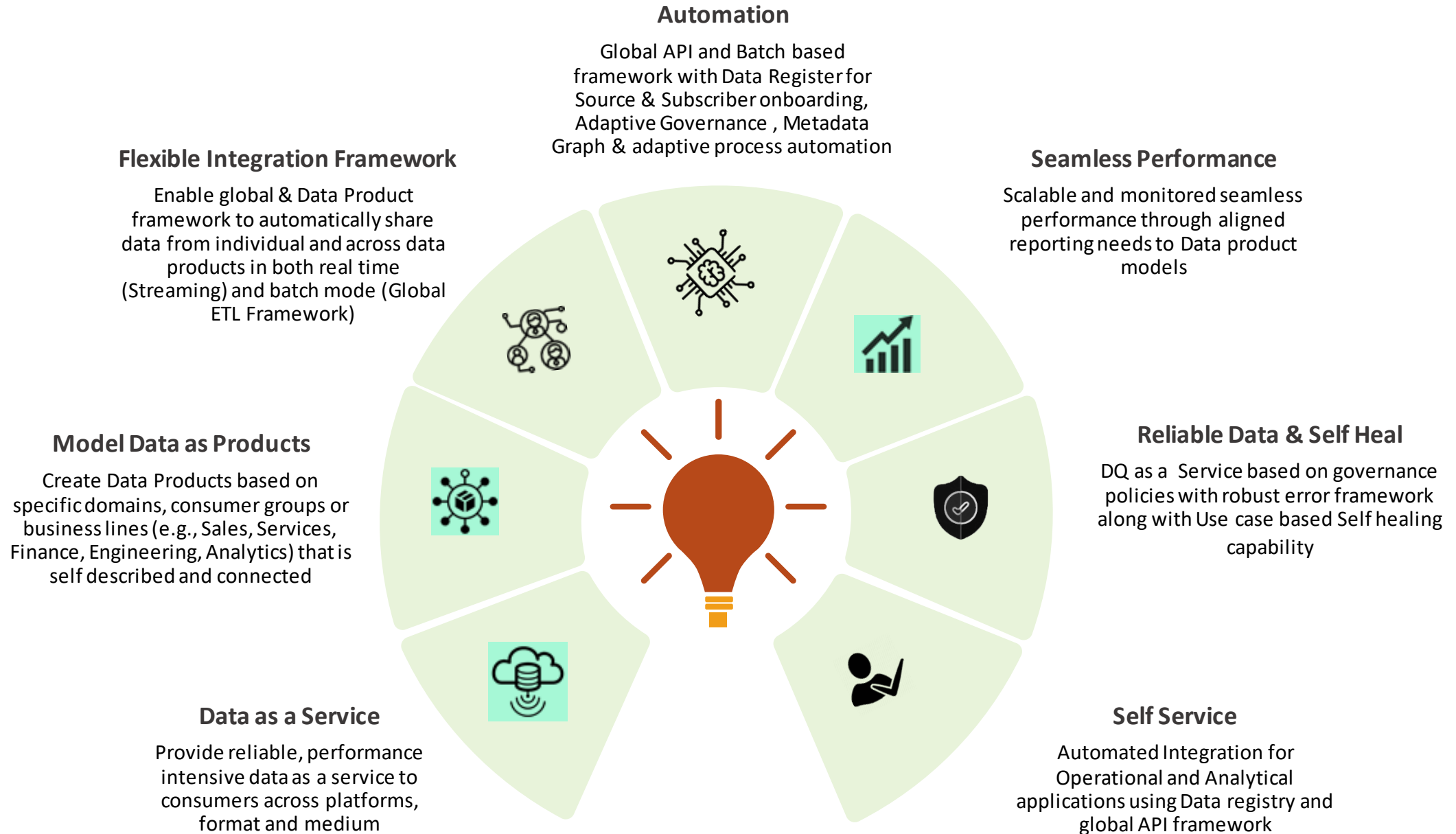


Key Azure Components

- Azure Data Factory (ADF) for orchestrating data integration pipelines to integrate data from across hybrid, multi-cloud, SaaS, and legacy enterprise data source systems
- Azure Databricks for building an open standard data lakehouse using the Delta format
- Power BI for Business Intelligence
- Microsoft Purview and Databricks Unity Catalog Federation for Unified Data Governance
- Azure Machine Learning

Key TechM IPs

- UDMF – Data Migration & Quality platform
- InfoWise - Metadata Governance Platform
- SPRINTER – Any Source to Any Destination Cloud Migration Accelerator
- FASTEST – ETL Test Automation Tool
- CDIF- Self Service Ingestion Framework for Modern Data Analytics Platforms





Data Fabric Best Practice

- **Meta Data based Common Ingestion Framework –**
Establish common ingestion framework based on centralized data model (consolidation based) using meta data graph.
- **Modular Governance Approach –**
Identify data policies based on critical data elements and business processes.
- **Scalable Infra planning & Monitoring -**
Scalable Infra planning based on data and processing or compute growth rate enabled by dashboard reporting through comprehensive monitoring.



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 & Agile Data Governance –**
Federated governance across data products with a touch of centralized data governance for key areas such as data security ,Data Quality & Data ingestion framework.
- **MDM & DQ as a Service –**
Provide MDM & Data Quality processes as a centralized services to maintain data consistency across enterprise

VISION

- HealthCare major GSK embarked on a journey to modernize their Data Ecosystem, migrate their on-prem systems and application to Azure Cloud.

CHALLENGES

- Data separation by Dept and Data Domain
- Setting up Data Governance & Policy (Local Vs Global)
- User Separation based on if the user is an enterprise or domain user
- Infrastructure provisioning

SOLUTION

- Architecture Solution** - *Architecture blueprinting and tech stack selection*
- Use Case Based Access Framework**- *enablement for users via Global access Framework.*
- Pipeline Modernization** - *Enhance and modernize around 10000+ data pipelines*

Value Delivered

Compliance & Governance

Implementation of HIPPA ,GDPR with local and Global policy implementation.

Self Service – Enablement of Self Service

Centralized Data Quality

Process Efficiency and quality improvement of data

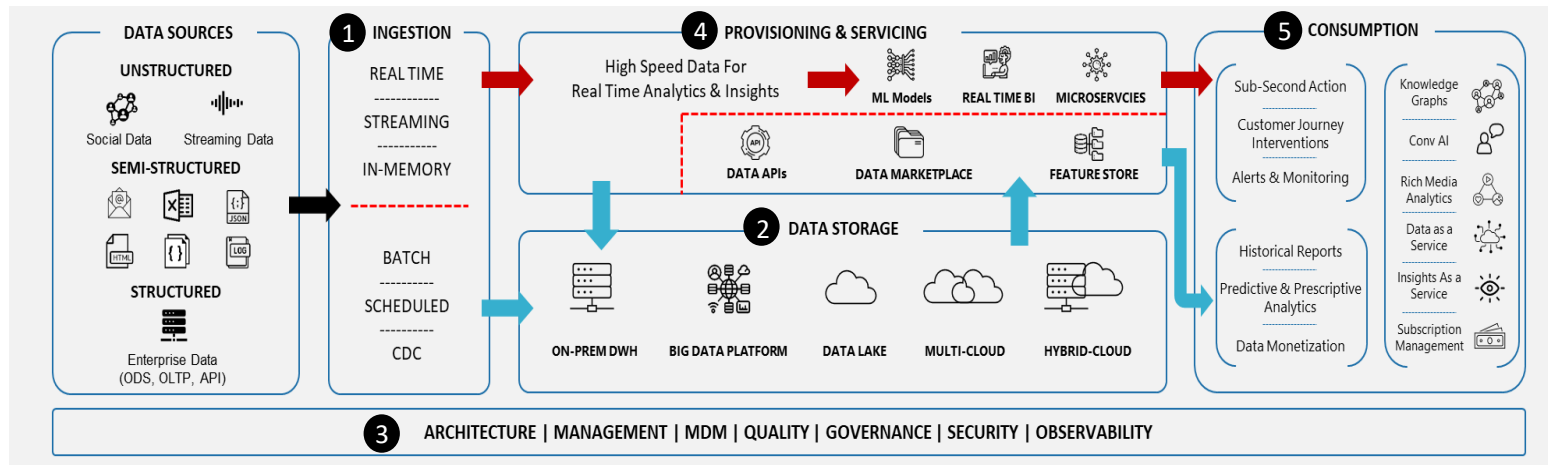
AI/ML Efficiency

Achieving high accuracy in prediction of drug sales



Tools: Azure Data Lake ,Databricks,Synapse,Collibra

Data Platforms Map Across the Data Value Chain Aiming to Reduce Time & Effort at Transforming into a Modern Data Ecosystem



1

CDIF
COMMON DATA INGESTION FRAMEWORK

60%

Self Service Ingestion Framework for Modern Data Analytics Platforms

- Metadata Driven Automated Ingestion
- Azure, AWS and GCP
- Dynamic parameters for ADF & equivalent pipelines from SQL Metadata repositories
- Metadata Configurable in SQL

2 3

UDMF

30%

Data Migration & Quality Platform

- DQ assessment, consulting & recommendation
- Data cleansing, standardization & reconciliation
- Business data movement to cloud
- Automate application assets movement

1 2

fas EST

25-30%

ETL Test Automation Tool

- Auto Test Cases/Scenario Creation
- Data Comparison across large datasets
- Query Building Capability for Multiple Platforms
- Referential Integrity Testing

2 3

SPRINTER

>40%

Any Source to Any Destination Cloud Migration Accelerator

- Data Search & Discovery
- Schema & SQL Migration
- Data Migration & Validation
- Monitoring & Tracking

3

InfoWise
Enterprise metadata 360
Augmented by AI

40-65%

GraphDB based Metadata Governance Platform

- Data migration assessment & strategy
- Report Rationalization
- Live Monitoring/ Data Observability
- Data Lineage & Business Glossary



Tech Mahindra

Connected World. Connected Experiences.



www.youtube.com/user/techmahindra09

www.facebook.com/techmahindra

www.twitter.com/tech_mahindra

www.linkedin.com/company/techmahindra

www.techmahindra.com

top.marketing@techmahindra.com