![MCKNIGHT CONSULTING GROUP]

Designing a Data Platform for the Future: Principles, Patterns, and Best Practices for Data Fabrics and Data Meshes

Presented by: William McKnight

"#1 2025 Global Thought Leader and Influencer in Cloud and in Data Center" Thinkers360

President, McKnight Consulting Group

3 X   **Inc 5000**

in   **/in/wmcknight**

www.mcknightcg.com
(214) 514-1444

TOP VOICE ★ 2025 ★ thinkers 360

RELTIO

# Designing a Data Platform for the Future – Principles, Patterns, and Best Practices for Data Fabrics and Data Meshes

Gino Fortunato
Solution Consultant, Reltio

Requirements for trusted data available where and when it's needed

Modern data architectures need to:

- Enable a **reliable system of context** for the enterprise AI.

- Deliver **trusted data AND real-time performance**. No room for compromise.

- **Combine centralized, automated** governance with decentralized, agile **domain-oriented management of data products**
  - [Watch our webinar](#) on this best practice

Siloed data is the kryptonite of digital and AI transformation ambitions

Agentic AI	Analytics	Workflows	Integration

Poor recommendations...

Unidentified risk...

Customer defection...

Incorrect diagnoses...

Brand damage...

Compliance problems...

duplicates... inconsistencies... incomplete... stale...

100's of applications...

Adobe SAP Marketo Veeva Microsoft Dynamics 365

ORACLE salesforce Anaplan GUIDEWIRE PEGA workday

servicenow databricks NETSUITE snowflake TERADATA

Dozens of third-party data sources

dun & bradstreet IQVIA LexisNexis zoominfo

experian Loqate MedPro SYSTEMS Linkedin

Reusable trusted data by domains: "data as a product"

Create trusted, unified 360 profiles

Augment domain-based approach with entity relationships

**Maintain security and continuously manage data quality**

Accurate, rich information about customers, products, suppliers, and locations...

Provide real-life context via relationship views of unique entities

Enable self-service with well-governed, reusable assets

A unified data foundation is your competitive advantage in the age of AI

Agentic AI	Analytics	Workflows	Integration

RELTIO
Intelligent Data Graph

Applications

Adobe · SAP · Marketo · Veeva · Microsoft Dynamics 365

ORACLE · salesforce · Anaplan · GUIDEWIRE · PEGA · workday

servicenow · databricks · NETSUITE · snowflake · TERADATA

3rd Party Data Sources

dun & bradstreet · IQVIA · LexisNexis · zoominfo

experian · Loqate · MedPro SYSTEMS · LinkedIn
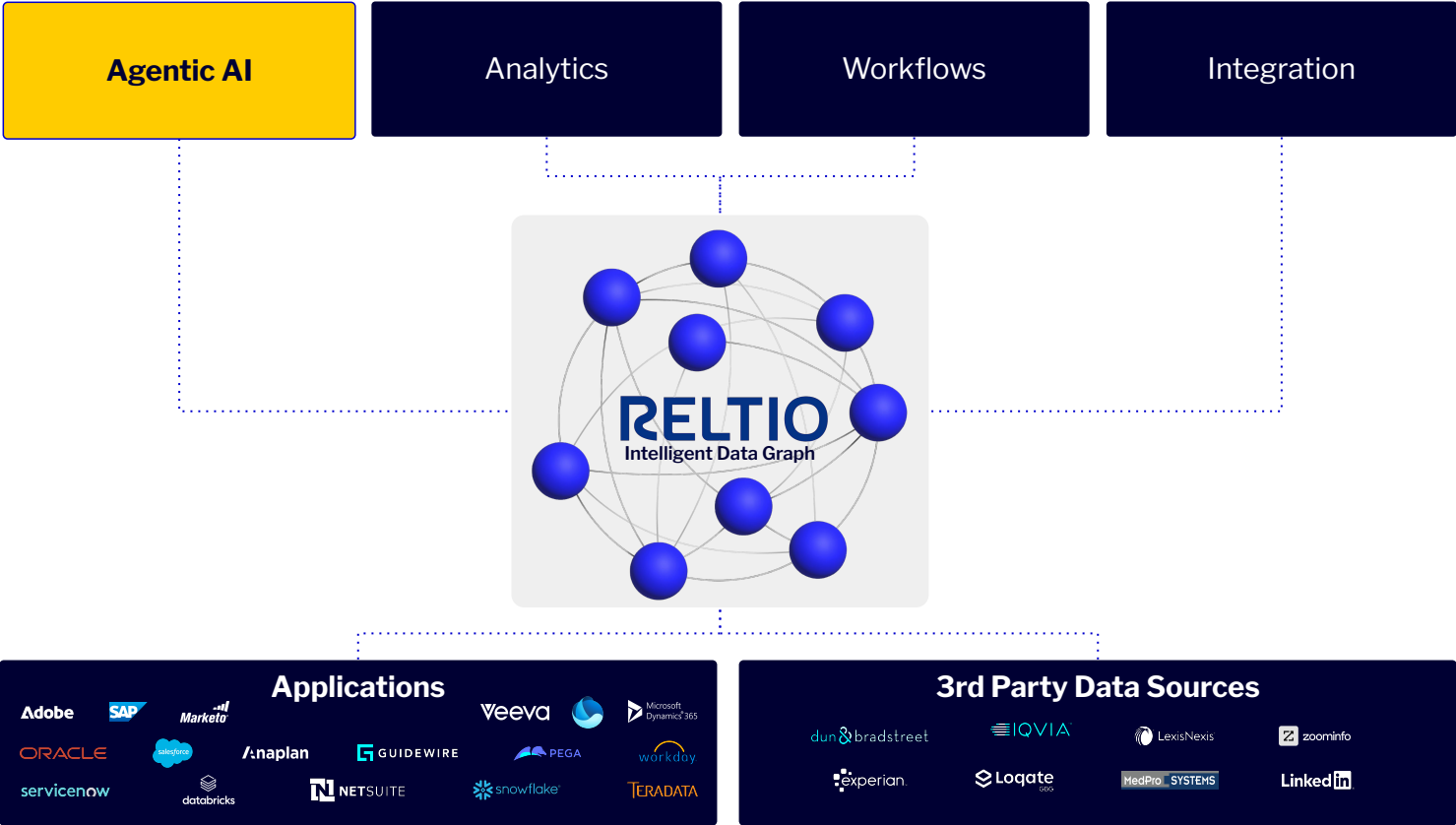
Reltio Intelligent Data Graph - the foundation of the agentic future

Secure, real-time access

APPLICATIONS

LEGACY APPS

DATA LAKES

CATALOGS

DATA ENRICHMENT

API

API

AI & AI Agents

BI & Analytics

Data Lakes

Applications

CLAIM
- Incident
- Adjudication
- Claim amount
- Adjuster
- Service provider

QUOTE

ASSET / PROPERTY
- Asset Type
- Owner
- Replacement value
- Age

LOCATION
- Address
- Theft rate
- Fire risk score
- Flood risk score

AGENT

CLIENT
- Contact info
- History (purchase, lapses, claims)
- Credit Score, Insurance Score
- Engagement signals
- Affiliation, household discount eligible

POLICY
- Type,
- Effective, renewal date
- Coverage limits/, deductibles
- Discount %

HOUSEHOLD

CAMPAIGN
- Segment
- Channel
- Offer
- Metrics/KPI

RELTIO

RELTIO

6

Global insurance broker and risk management advisor accelerates RFI response time by 99%

AI agents help producers populate RFIs in one hour

BUSINESS CHALLENGES

- Fragmented technology stack
- 80 individuals manually mapping and reconciling data
- Missed cross-sell/upsell opportunities
- Declining customer satisfaction
- RFIs take 30–50 hours per report
- High risk of manual errors

SOLUTION

- Created a unified view of each customer
- Data at their fingertips, instantly
- AI Agent for RFI

OUTCOMES

- Reduced information gathering from 4 weeks to 1 hour
- $17M estimated annual savings
- Reduced manual processes and errors

RELTIO

Learn More at Reltio.com

SCAN ME
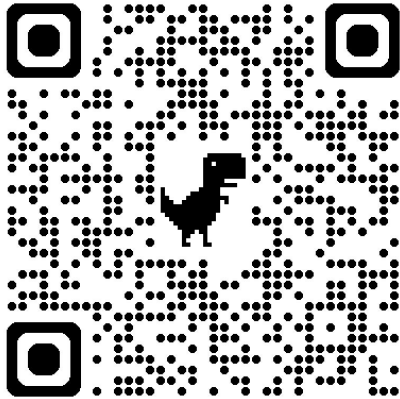
Thank you!

McKnight Consulting Group Partial Technology Implementation Expertise

Big/Analytic/Vector/Mixed Data Management

databricks · Pinecone · Google BigQuery · snowflake · Azure Synapse · MySQL HeatWave · teradata · amazon REDSHIFT · CLOUDERA · SAP Datasphere · Exasol · VERTICA · Starburst · ACTIAN activate your data · Azure HDInsight · SingleStore · Apache Impala · amazon EMR · Yellowbrick · Cloud Dataproc · ORACLE AUTONOMOUS DATABASE · ORACLE

Data Movement and APIs

Airbyte · Informatica · Qlik Replicate® · Fivetran · SYNADIA · CONFLUENT · apigee · talend · MATILLION · Stitch · Kong · Astera

Data Management

MONTE CARLO · Anomalo · SNOWPLOW · SAP Datasphere · alteryx DESIGNER CLOUD · Profisee · splunk> · NGINX Part of F5 · Collibra · IMMUTA · elasticsearch · Azure MLOps

Operational/Transactional Data Management

DIAMANTI · Cockroach · Couchbase · DataStax · SQLite · MongoDB · PostgreSQL · yugabyteDB · neo4j · ORACLE · Google Cloud Spanner · redis · Azure Cosmos DB · Microsoft SQL Server · Amazon Aurora

MCKNIGHT CONSULTING GROUP

The Goal of Modern Enterprises

- **Automating complex tasks:**
  enabling unprecedented levels of efficiency and productivity.

- **Providing smart data environments:**
  offering personalized assistance for data exploration and insights.

- **Driving proactive decision-making:**
  unlocking new levels of data insights and automation.

- **Simplifying business processes:**
  streamlining operations and improving ROI.

- **Facilitating autonomous decision-making:**
  enabling hyper-personalization and data-driven decision-making.

The World is Changing

- **Old Way:**

  Ramp up: Spend months getting familiar with the company

  Data discovery: Take hours to find and understand relevant data

  Waiting on data engineers: Spend hours to days waiting for data engineers to ingest data

  Data optimization: Wait days to weeks for data engineers to optimize data

  Dashboard creation: Spend days to weeks building business dashboards and spaces

- **New Way:**

  Instant insights: AI agents interface with structured and unstructured data

  Smart navigation: Semantic layer guides agents to the data instantly, providing business context

  Unified access: Zero-ETL federation gives unified access to data instantly

  Fast answers: Agents deliver answers instantly

  Optimized queries: Agents optimize queries for everyone instantly

The Answer: The Enterprise Data Platform

Reusable Extensions

Encourages adding modular extensions to the core architecture for enhanced functionality.

Tailored Structures

Involves creating custom data structures to meet specific needs.

Data Reusability

Promotes storing data in a way that allows for multiple uses without duplication.

Diverse Processing Approaches

Emphasizes the importance of integrating various data processing methods.

Enterprise Usability

Focuses on ensuring the architecture is widely accessible and usable across the enterprise.

Redundancy Minimization

Aims to reduce unnecessary data duplication to enhance efficiency.

Most Will Need More than an "AI Band Aid"

- Data Silos
- Lack of Governance
- Data Hoarding
- Unwieldy Legacy Pipelines
- Lack of an Architecture Focus

MCKNIGHT
CONSULTING GROUP

Distributed Data Architecture Patterns

A single data lake architecture has shortcomings that lead to unfulfilled promises at scale

- Monolithic
- Data Management Complexity
- Performance Limitations
- Governance Issues
- Hyper specialized ownership

Pros and Cons of Following Architectural Patterns

Pros
- Theoretically, it's science and has been validated
- Decisions addressed you were unaware of
- Understandable

Cons
- Can lose focus on the business priorities
- May not be right for you
  - Can be overkill
- Limited flexibility

The Key Questions for moving to an Architectural Pattern

- How many stakeholders regularly search for and access data?
- How many stakeholders are independently analyzing data to inform decisions?
- How many data-driven applications are being developed to drive business transformation?
- What is the usage and adoption rate of your data catalog, analytics, and AI tools?
- Are there many types of data frequently used to drive business outcomes?
- Is there an active data community, and how engaged are its members?

MCKNIGHT
CONSULTING GROUP

These are not Mutually Exclusive

Data Mesh

Data Fabric

Data Lakehouse

Data Warehouse

MCKNIGHT
CONSULTING GROUP

Data Mesh

Low Latency

- Logs (Apps, Web, Devices)
- User tracking
- Operational Metrics
- Sensors
- Applications
- Files

Batch

Txn/ Context Data OLTP/ODS

kafka

Raw Data Topics JSON, AVRO

Processed Data Topics

S3 Data Lake

In-database analytics

Data Warehouse

Reach through

Q

Offload data

Q

MCKNIGHT CONSULTING GROUP

Data Mesh Principles

- Domain Ownership
- Data as a Product
- Self-Service Data
- Distributed Pipelines
- "Serve and Pull"
- Federated Agile Governance
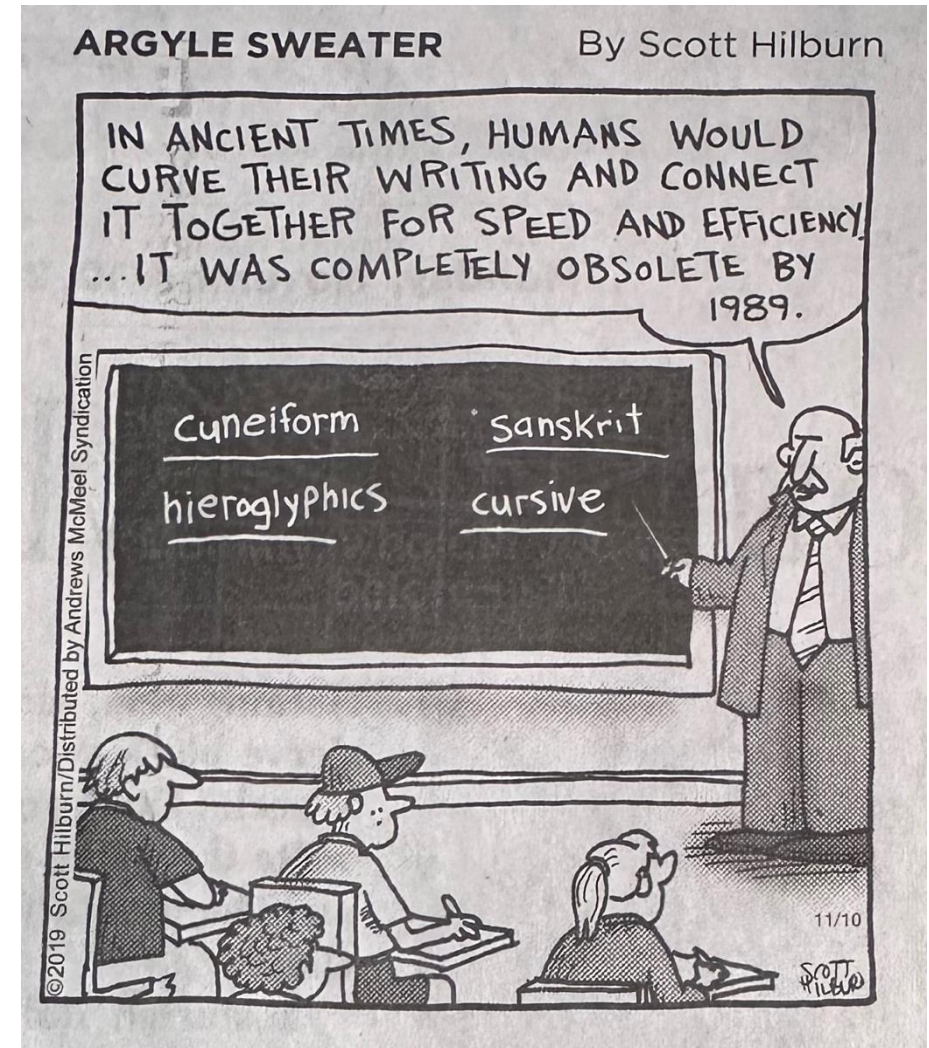
Benefits of a Data Mesh

- Democratization of Data
- Cost Effective
- Reduced Technical Debt
- Collaboration
- Safety
- Adherence
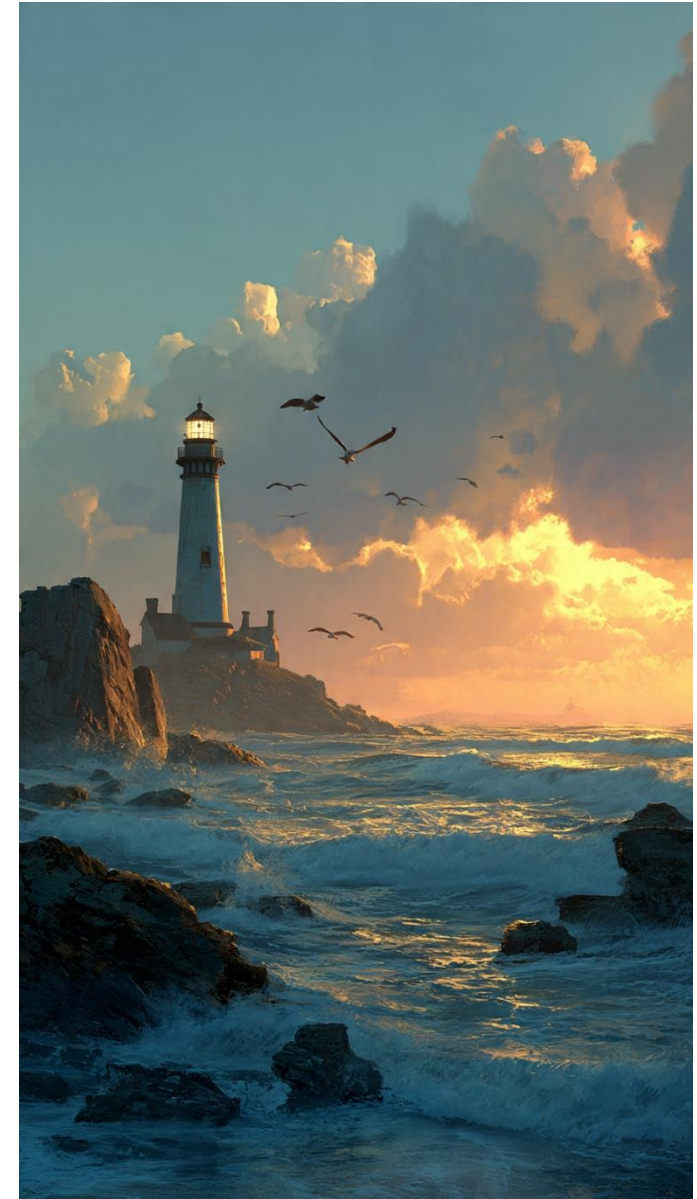
Use Cases for Data Mesh

- Business Intelligence Dashboards
- Customer Experiences
- Machine Learning
- Risk Assessment
- Manufacturing Production Optimization
- Supply Chain Optimization
- Regulatory Compliance

MCKNIGHT
CONSULTING GROUP
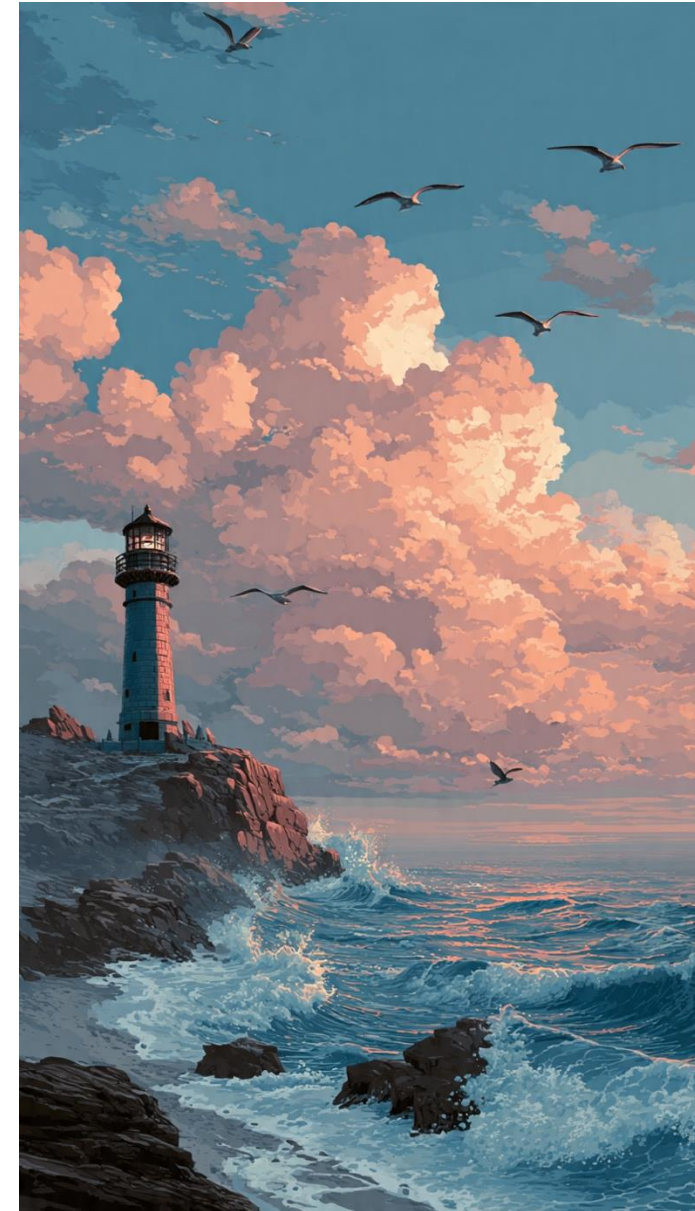
The Data Mesh Way

Centralized Platform

Ingesting data into a central platform

Extracting and *loading* data

Flowing data through centralized pipelines

A centralized *data platform*

Data Mesh

Serving data from domains

Preparing and *sharing* data products

Publishing events as streams from domains

An *ecosystem of data products*

MCKNIGHT
CONSULTING GROUP

Data Lakehouse

In-database analytics

Data Warehouse

Q

Reach through

Q

S3 Data Lake

MCKNIGHT
CONSULTING GROUP
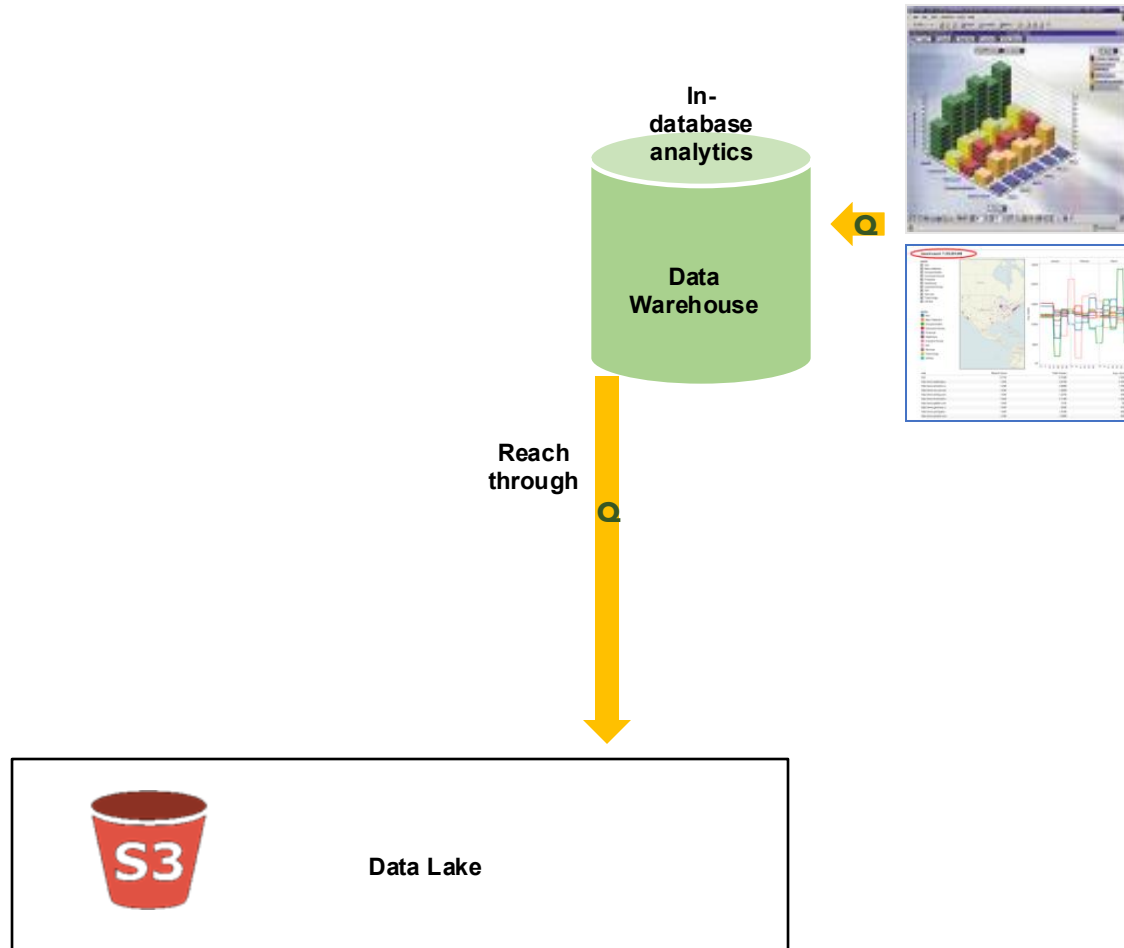
Data Lakehouse Principles

- Unified Data Management

- Open Formats

- Adaptable Storage

- Curated Data with Zones

- Facilitating the Continuous Flow of Streaming Data

- Handling Varied Tasks

MCKNIGHT
CONSULTING GROUP

Open Data Formats

Table Format

File Format

A method of organizing and structuring a dataset's files within a file system or object store, enabling them to be presented and queried as a unified, logical table.
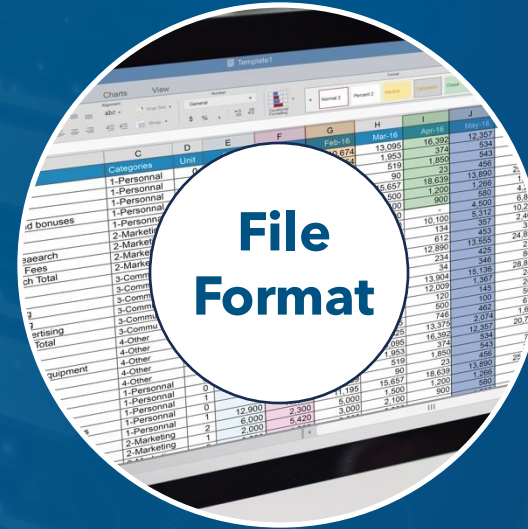
The physical format in which data is stored in files, such as:

- Text formats (e.g., CSV, JSON)
- Binary formats (e.g., Parquet, Avro, ORC, Arrow)

Open Table Format and Open File Format

Data Movement Across the Platform

- Data Movement vs Data Repository
- Data pipelines are becoming commoditized
- Kafka streaming services and CSP-native tools meeting basic needs
- Growing data volume and complexity will drive demand for streaming data services
- Connectors > Data/Event Hub connections

Why Data Products?

Data Sharing   **Data Services**   **Data Interoperability**   **Data Mesh**

MCKNIGHT
CONSULTING GROUP

Data Product Development

Data Ownership

Data Definitions

Data Lineage

Data Contracts

Data Consumption Pattern

Data Use

Data Contracts

- What data constraints, definitions, and validation rules apply?

- What are the service level agreements (SLAs) and objectives (SLOs)?

- What are the data sharing rules, authorized uses, and governance policies?

- What is the purpose or intended use case?

- How does it perform and scale under load?

- What are the quality, reliability, and maintainability characteristics?

- What are the security controls, access rules, and visibility settings?

Data Fabric

Low Latency

Logs (Apps, Web, Devices)

User tracking

Operational Metrics

Sensors

Applications

Files

Batch

Txn/ Context Data OLTP/ODS

kafka

Raw Data Topics JSON, AVRO

Processed Data Topics

In-database analytics

Data Warehouse

S3 and / or **Parquet**

Data Lake

MCKNIGHT CONSULTING GROUP
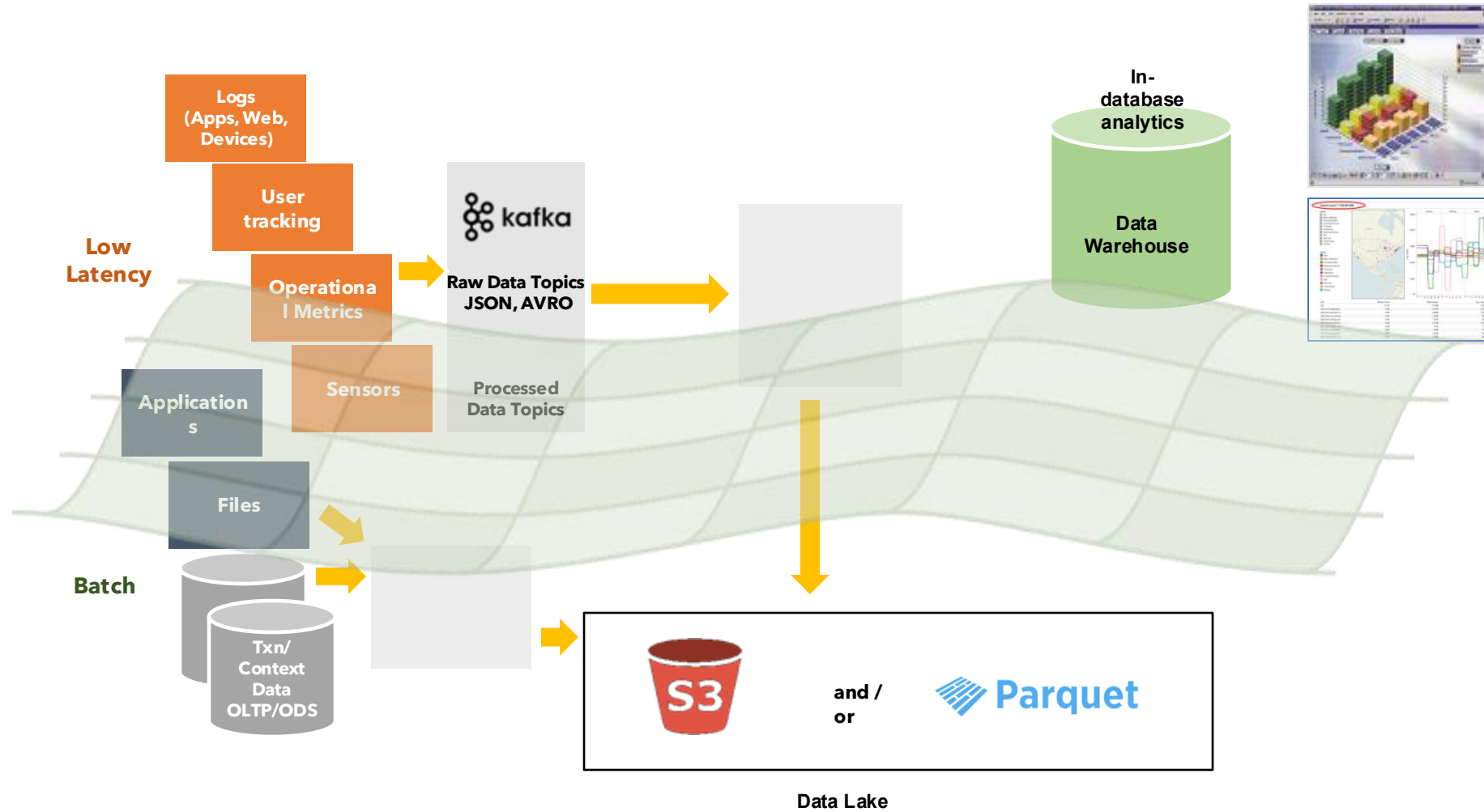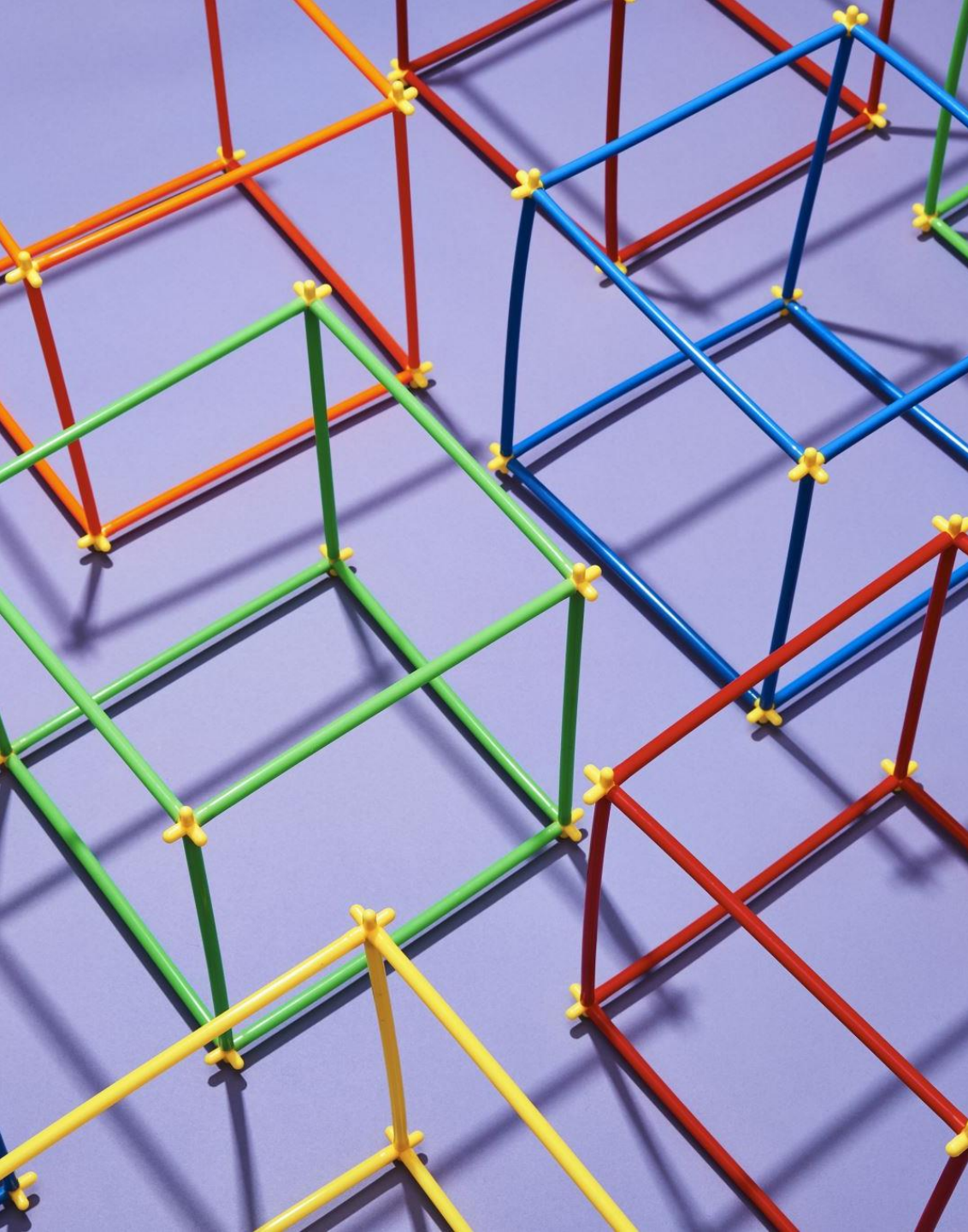
Data Fabric Principles

- Intelligent and Automated
- Unification of Disparate Data Systems
- Access to Integrated Enterprise Data
- Scale Efficiently
- Multi-Cloud Awareness
- Data Democratization
- Integrated Security

Benefits of a Data Fabric

- Boosted Productivity
- Better Collaboration
- Advanced Analytics
- Diverse Data Delivery
- Dynamic Pipeline Scaling
- Cost Optimization

Metadata Examples

- Using data arrival timestamps to identify patterns in data delays.

- Inferring data schema from records to automatically detect schema drift and errors.

- Identifying data processing bottlenecks and dynamically allocating resources (containers, network capacity, etc.) to underperforming areas.

- Recommending new data sources to integrate, enabling self-healing data integration jobs, and optimizing data workloads by choosing the most appropriate execution engine or pipeline structure (e.g., ELT vs. ETL).

MCKNIGHT
CONSULTING GROUP

Use Cases for Data Fabric

- Fraud Detection
- Preventative Maintenance
- Data Discovery
- Customer Profiling
- Risk Modeling
- Automated Analysis

Master Data Management
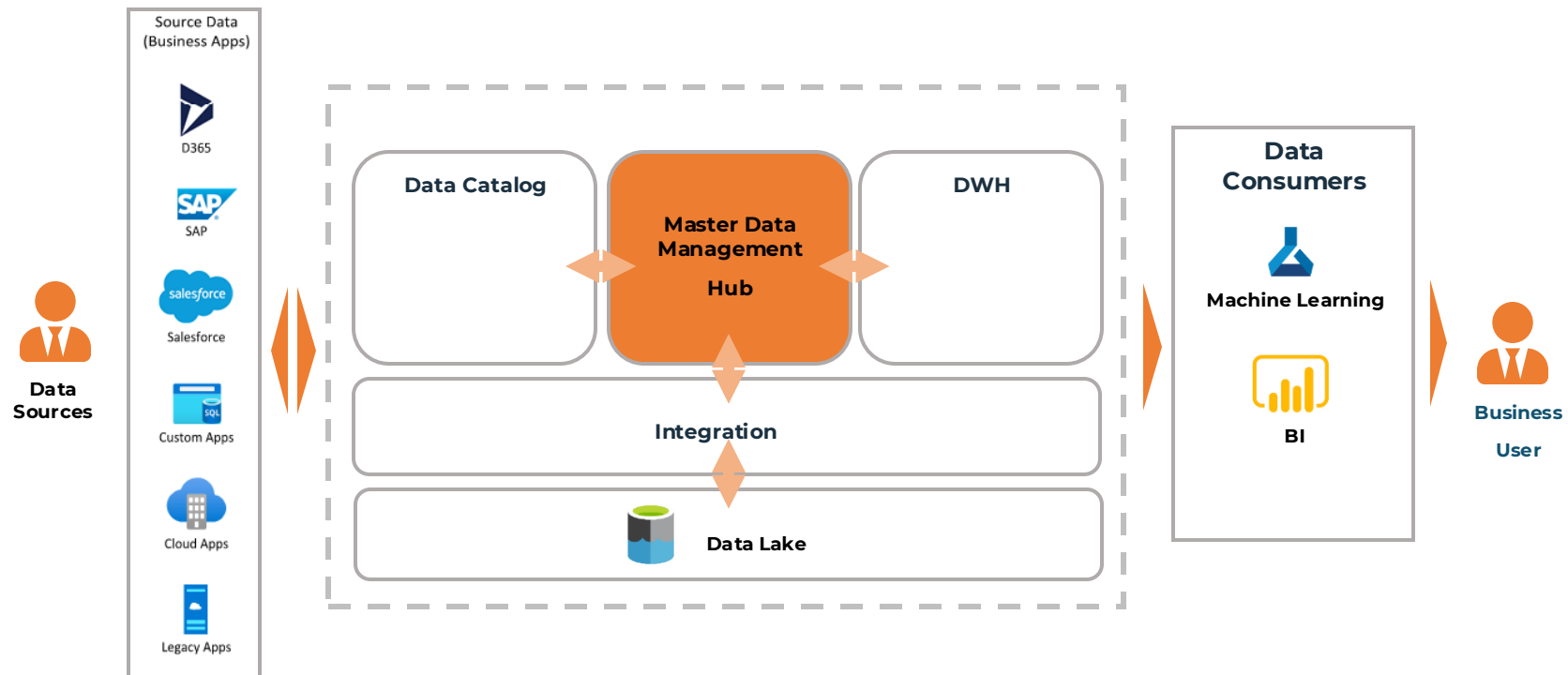
Data Architecture Needs Translytical Workloads

- Fast Streaming Ingest (millions of events/second)
- Low Latency
- High Concurrency (thousands of concurrent users)
- Unlimited Storage
- Pipelines
- Transactional Consistency

A multi-model database is a single, integrated database that can store, manage and query data in multiple models

- Open Linked Data
- JSON
- XML
- Documents
- Binary
- Graph
- Log Files
- Vector
- Images and Multimedia
- Time Series
- RDF
- Spatial

MCKNIGHT
CONSULTING GROUP
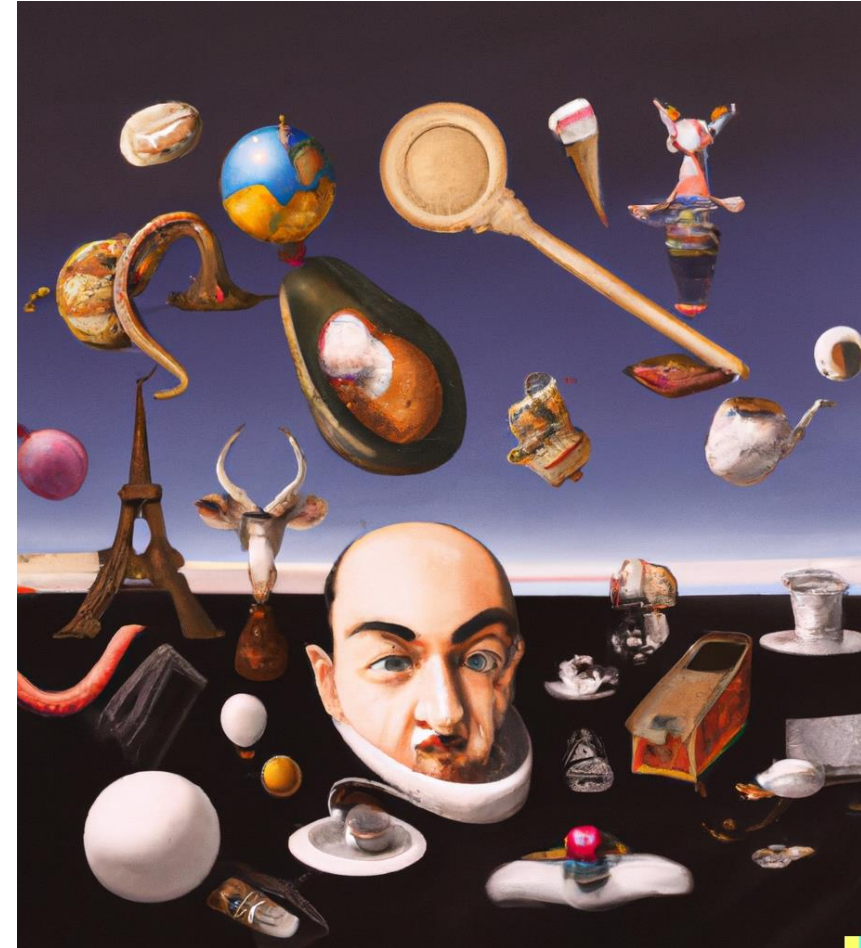
Data at the Edge

- Oil & gas in remote location
- Agriculture & mapping
- Black box equipment – i.e., airplanes
- Self-Driving Cars
- Trading Algorithms
- Smart Homes
- Transportation Networks
- Law Enforcement

MCKNIGHT
CONSULTING GROUP

Hub and Spoke Data Mesh with Data Products
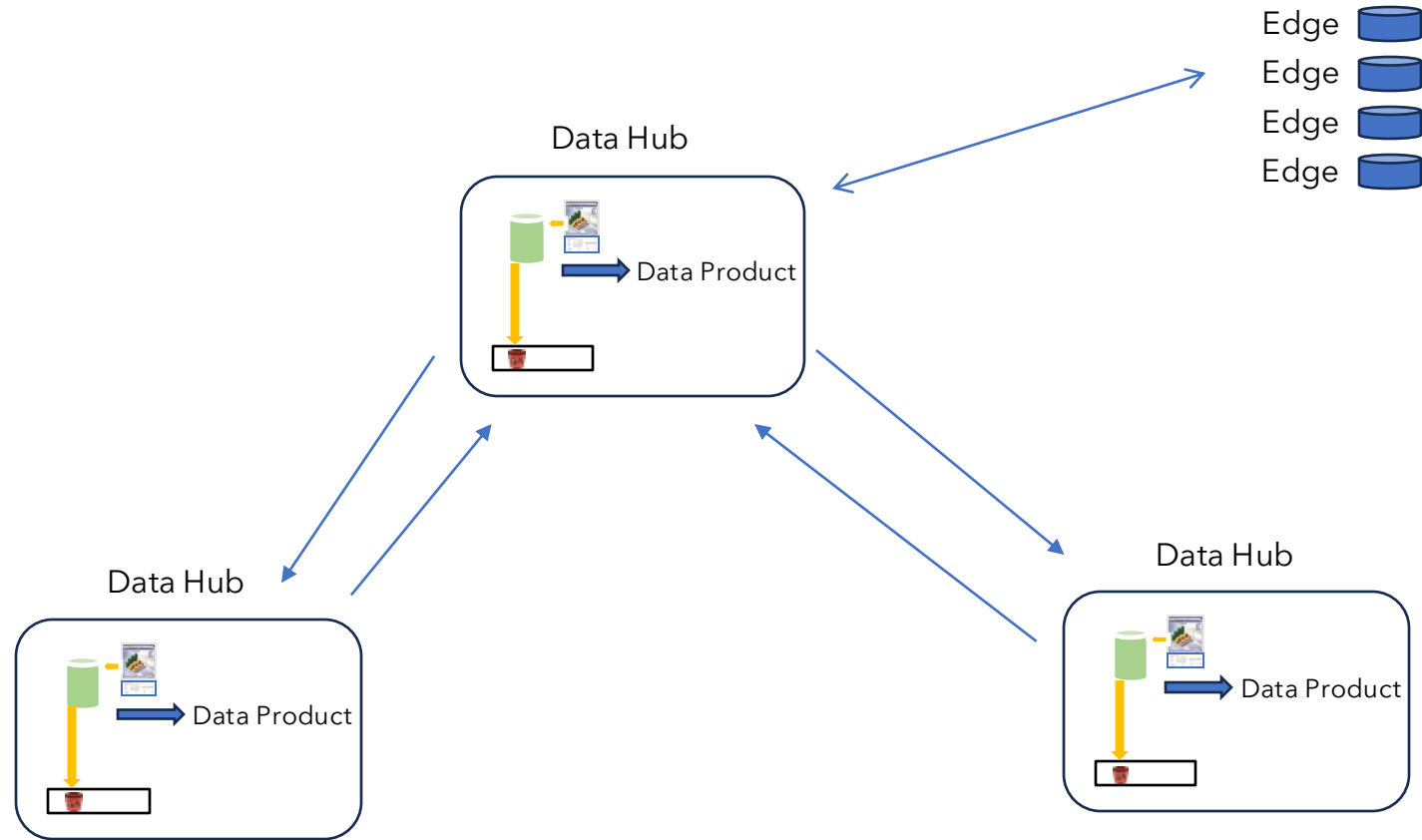
Challenges to Building Data Platforms

It's a moving target with many components.

Technical Challenges

- **Data Quality and Availability**: Models require high-quality, diverse, and relevant data to learn and generate accurate results
- **Explainability**: Models can be difficult to interpret, making it challenging to understand the decision-making process
- **Scalability and Performance**: Models can be **computationally intensive**, requiring significant resources to deploy and maintain

Process Challenges

- Aligning with business goals
- Knowing what you're building – program or project
- Need DevOps/MLOps
- Internal Resistance
- Skills Gap
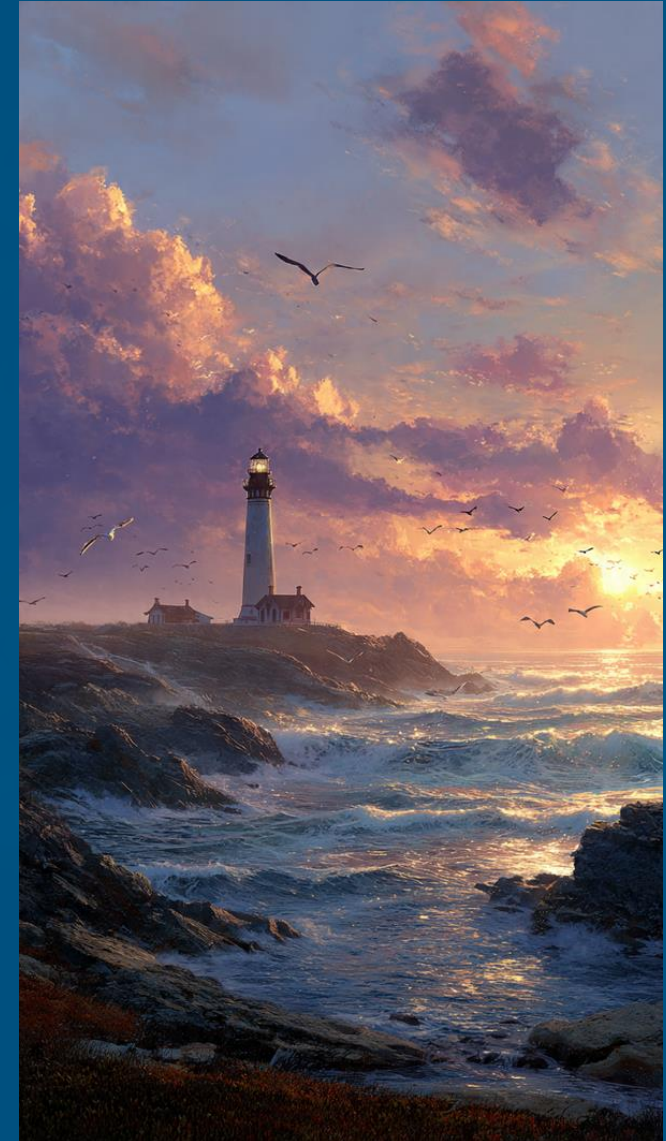
Summary

Build the Enterprise

Data Mesh

Data Lakehouses

Supports AI

- Build the Enterprise Information Architecture for Diverse Interests Data

- Data mesh ensures domain teams own and manage data quality, providing accurate and reliable data for AI models

- Data lakehouses accommodate diverse data types, enabling AI models to leverage structured, unstructured, and semi-structured data:
  - Multi-modal
  - Translytical
  - Edge
  - Data Products
  - Open Data Formats

- This architecture supports AI by:
  - Enabling robust data pipelines for AI model training
  - Providing real-time data for AI model inference
  - Facilitating scalable AI model deployment
  - Supporting diverse AI use cases (e.g., NLP, computer vision)
  - Ensuring data governance and compliance for AI applications