

# Data Challenges Past, Present and Future

# BiLET

**Business Innovation  
Leadership & Technology  
Conferences**

**Ben Wright-Jones**

Senior Solution Architect

Industry Solutions Engineering, Data & AI  
Data & AI Solution Area community lead  
Microsoft

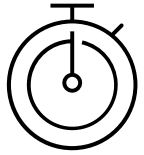
 @datalineage

 /bwrightjones

The perspectives shared are my own and not endorsed by Microsoft



# Session Goals



Where we have been (data past), where are we today (data present) and where we might be going next (data futures)



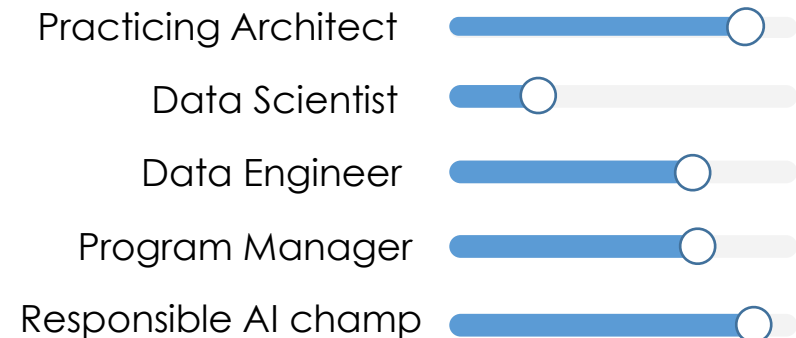
Overview of data architecture patterns and practices influenced by industry and business requirements



Describe data challenges observed working with multiple organisations

# Who am I?

- Senior Solution Architect, Microsoft
  - Data & AI background & focus
- Worldwide Community lead for Microsoft Data & AI Solution Area
  - Facilitating collaboration and knowledge transfer
- Data Governance program lead
- 25 years working across various industries



# Why do we need to talk about data?

- Unprecedented disruptions such as the pandemic, political and economic uncertainty have exposed organisational (digital) fragility
- Need to rapidly innovate in increasing competitive landscape
- Realisation that business needs greater agility

# Data Realities

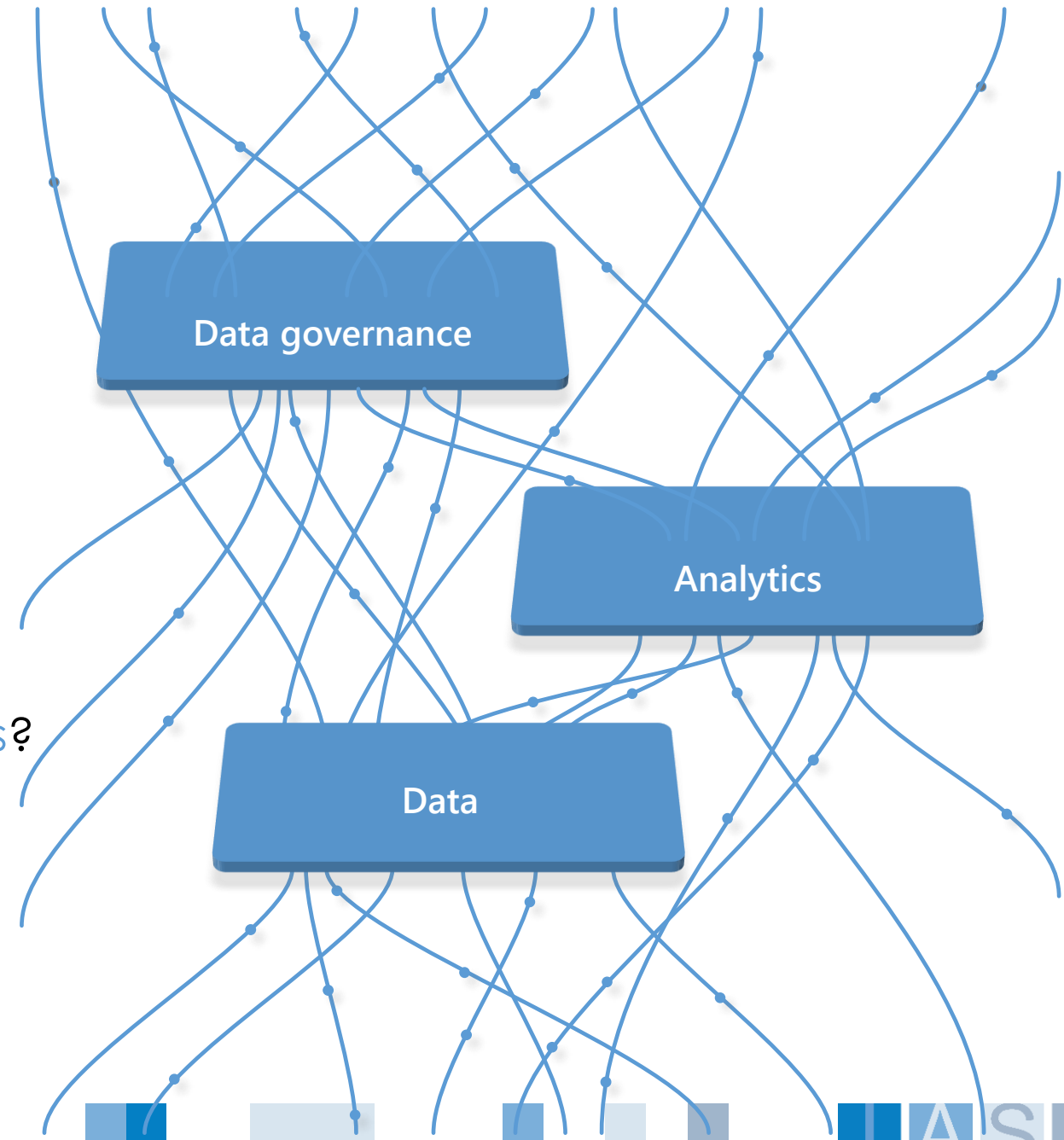
What **data** do I have?

Is the data **trustworthy**?

Can people **access the data** needed to make the right decisions?

How can we enable **faster business insights**?

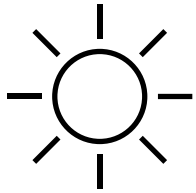
What's our **compliance exposure**?



# Data: Past

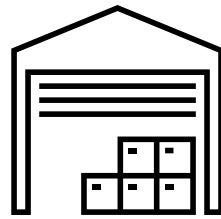


# Data: Past



Dims Facts

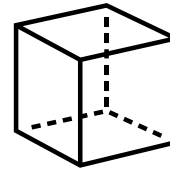
1960



Data Warehouse

1970~

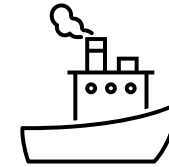
Enterprise  
Schema on write



Data Mart

2002 (1970)

Domain-specific  
Schema on write



Data Lake

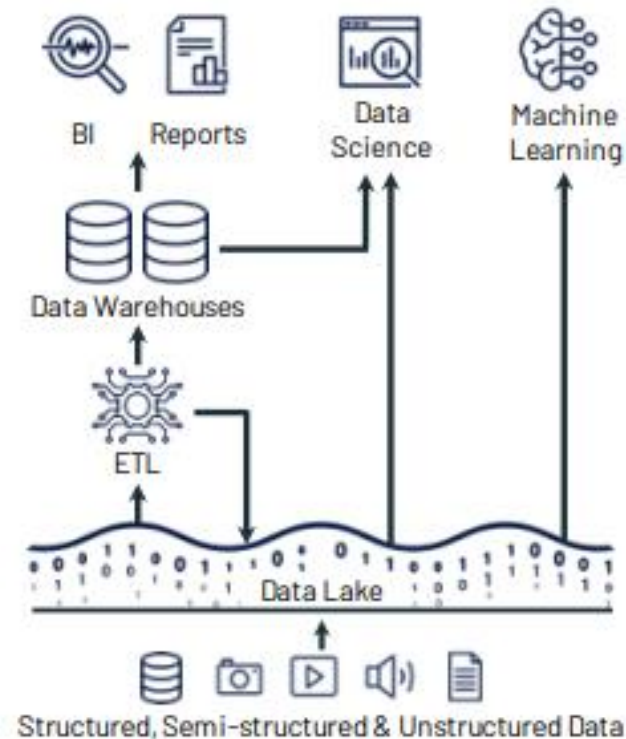
2010

Enterprise  
Schema on read

# Data: Past



Data Warehouse



Data Lake

Source: Armbrust, M., Ghodsi, A., Xin, R. and Zaharia, M., 2021, January. Lakehouse: a new generation of open platforms that unify data warehousing and advanced analytics. In Proceedings of CIDR.

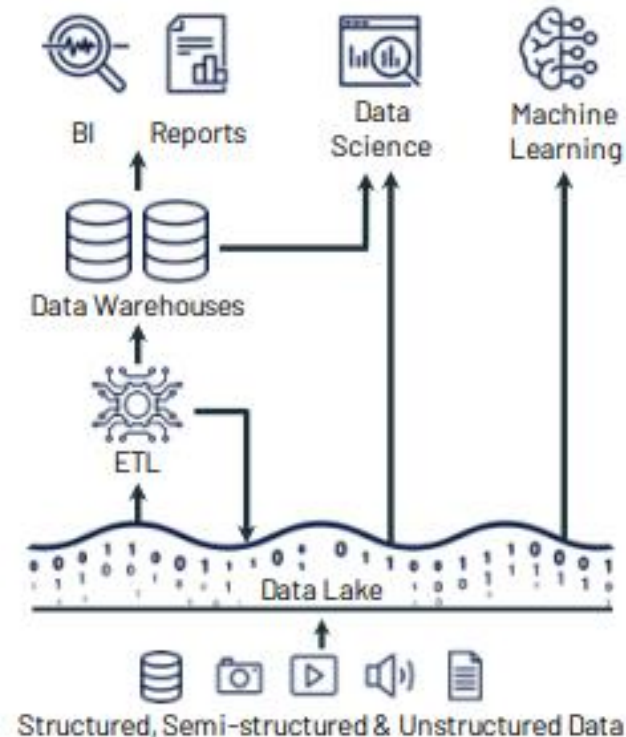


# Data: Past, Challenges

Scalability (compute and storage) constraints, limited flexibility (data formats, data types)



Data Warehouse



Data Lake

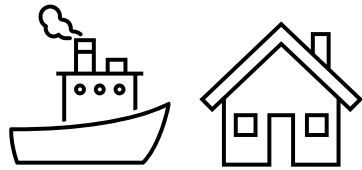
Data quality concerns  
Monolithic solution  
- Data swamp  
ETL/ELT pushed downstream

Source: Armbrust, M., Ghodsi, A., Xin, R. and Zaharia, M., 2021, January. Lakehouse: a new generation of open platforms that unify data warehousing and advanced analytics. In Proceedings of CIDR.

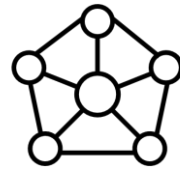
# Data: Present



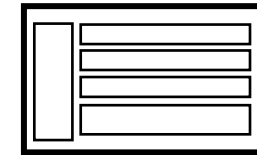
# Data: Present



Data Lakehouse



Data Mesh



Data Fabric

2016

2019

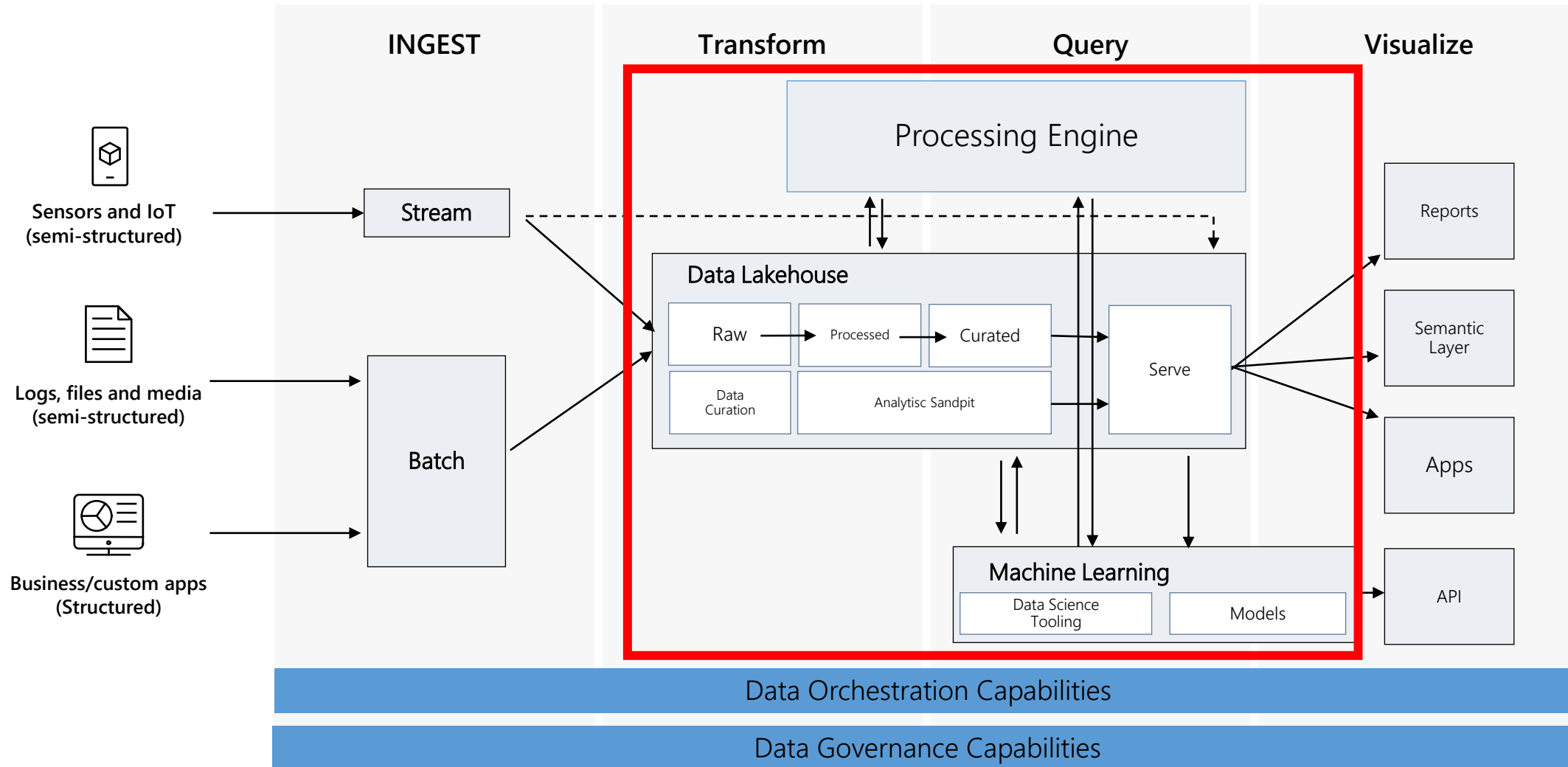
2021

Enterprise  
Schema on write

Decentralized  
Business domain-focused  
Standardized self serve data  
platform  
Product led thinking

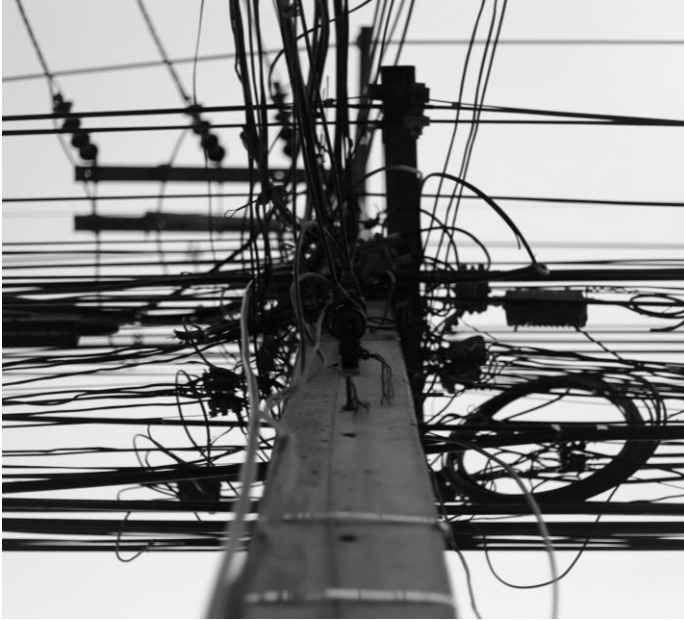
Multi-cloud  
Technology centric  
End to end data  
integration, lifecycle and  
management

# Data: Present (Lakehouse)



# Data: Present

## Triggers for Change



Monolithic  
Complexity



Accelerate  
Innovation  
& Value



Agility &  
Flexibility

# Data Topologies

Cultural Shift

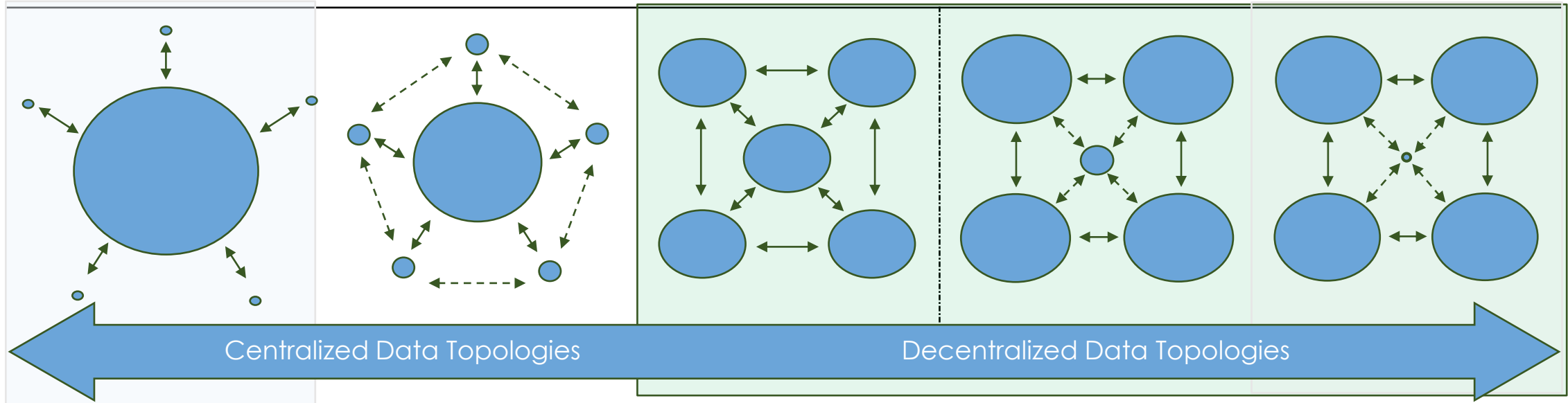
Centralized

Selective

Balanced

Federated

Autonomous



Provides a single point of control and decision making, with limited input from outside IT

Selected decisions and activities are owned by the business

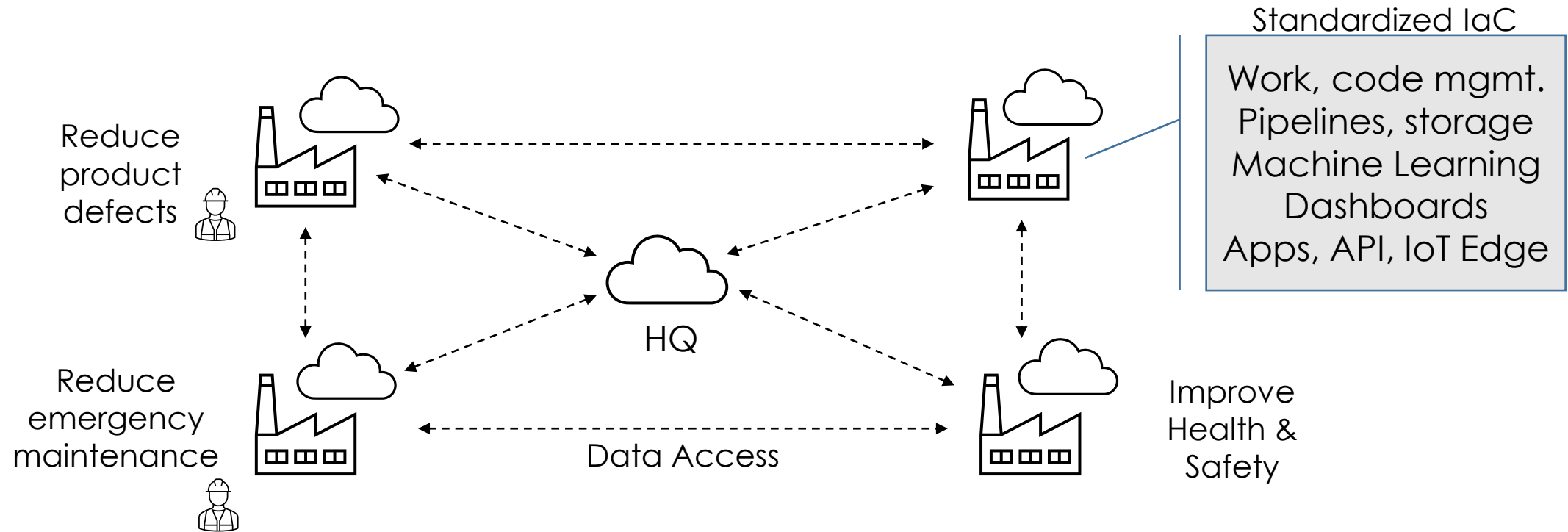
Decisions and activities are jointly shared between IT and the business

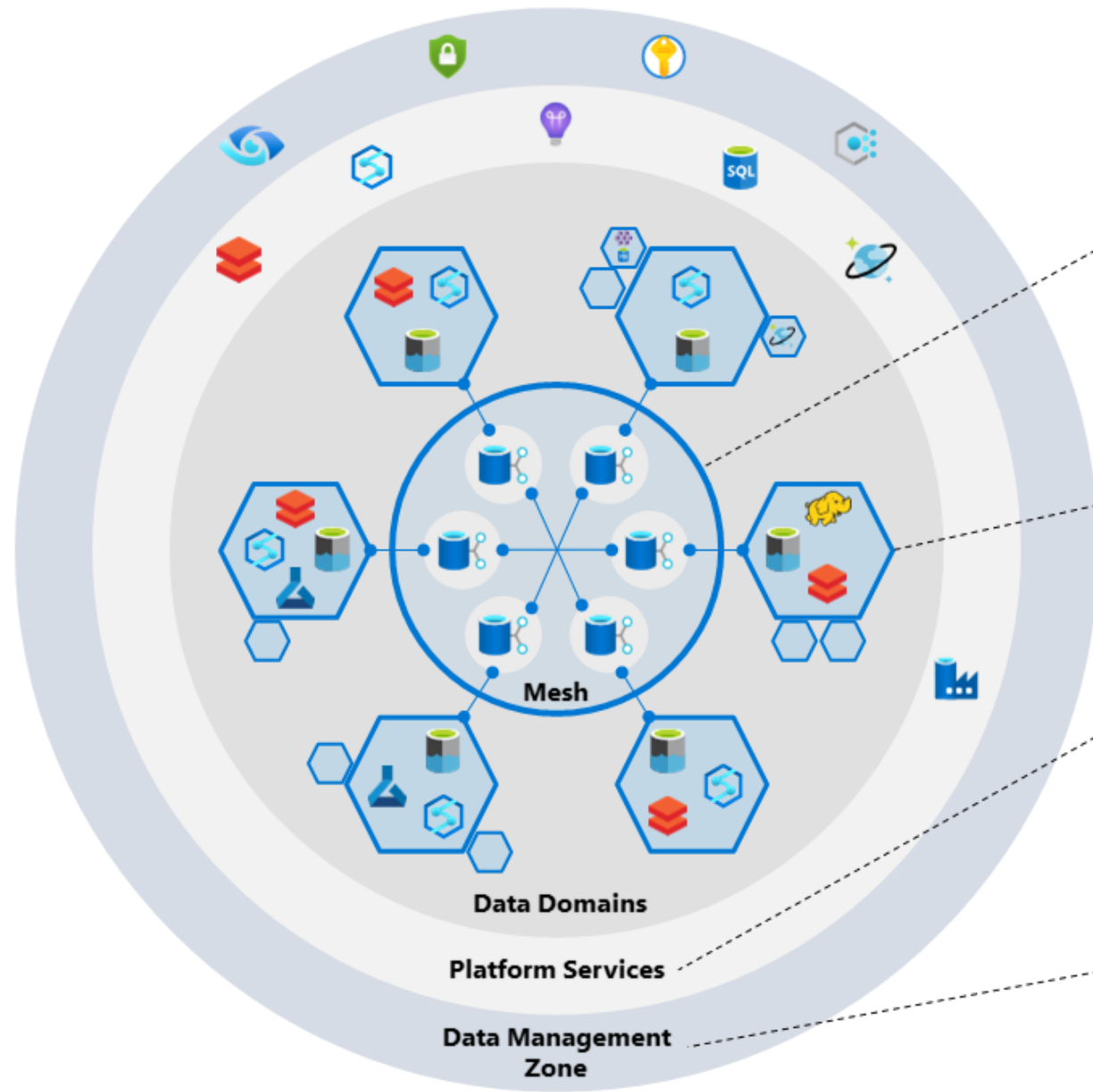
The business exerts control over decision-making and activities, with limited input from IT

Business units operate with autonomy  
– 'data mess' not mesh

# Industry scenario

- Localized domain driven requirements





The **data mesh** intelligently distributes data products between data domains. Read data stores share compute resources. This reduces costs, solves interoperability concerns, and better addresses time-variant and non-volatile concerns of large data consumers.

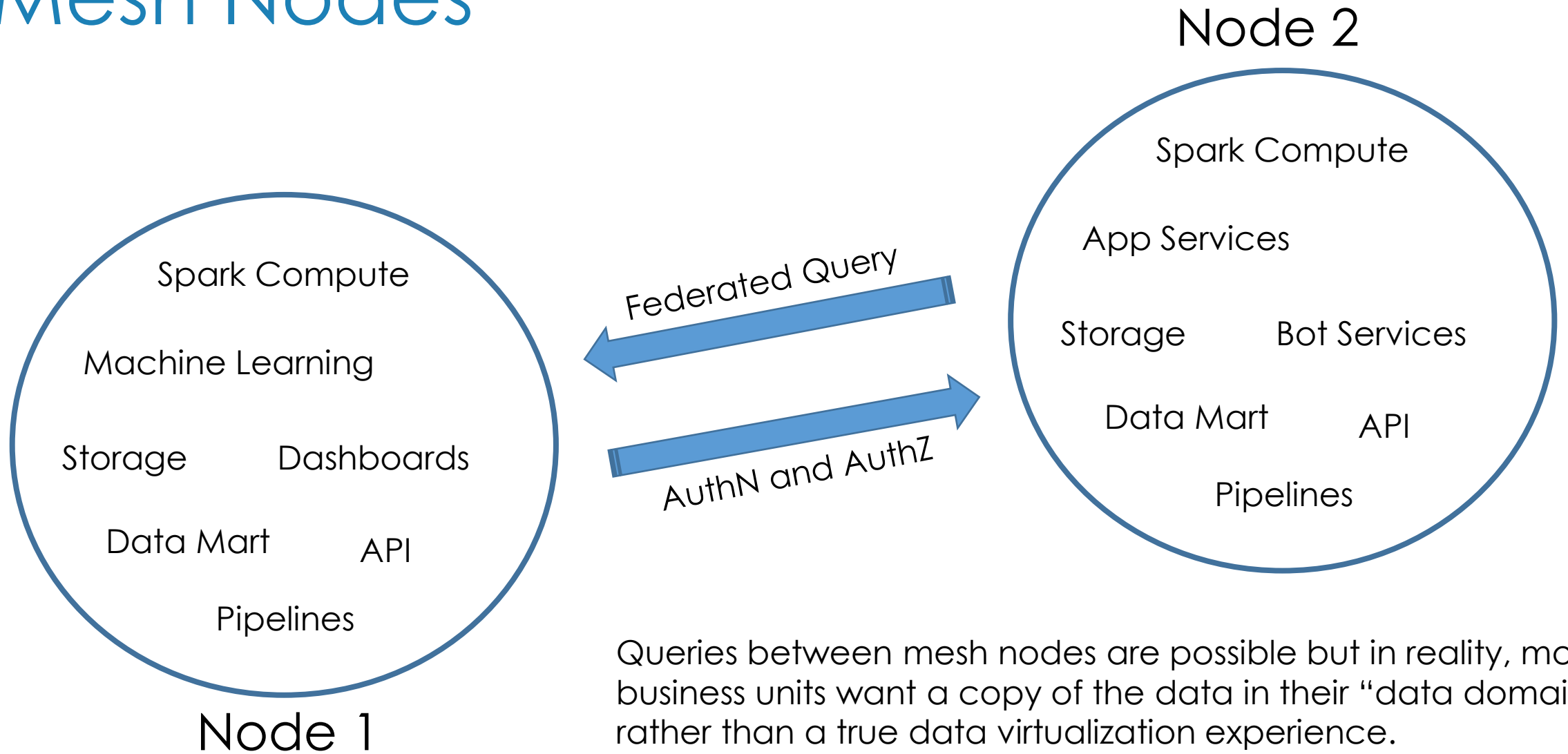
**Data domains** operate their own applications or analytics platforms, whilst adhering to common policies and standards.

The central **platform services** defines blueprints that encompass baseline security, policies, capabilities, and standards.

A key concept for every enterprise-scale analytics and AI implementation is having one **data management zone**. This subscription, which is required for data management, contains resources that'll be shared across all landing zones.



# Mesh Nodes



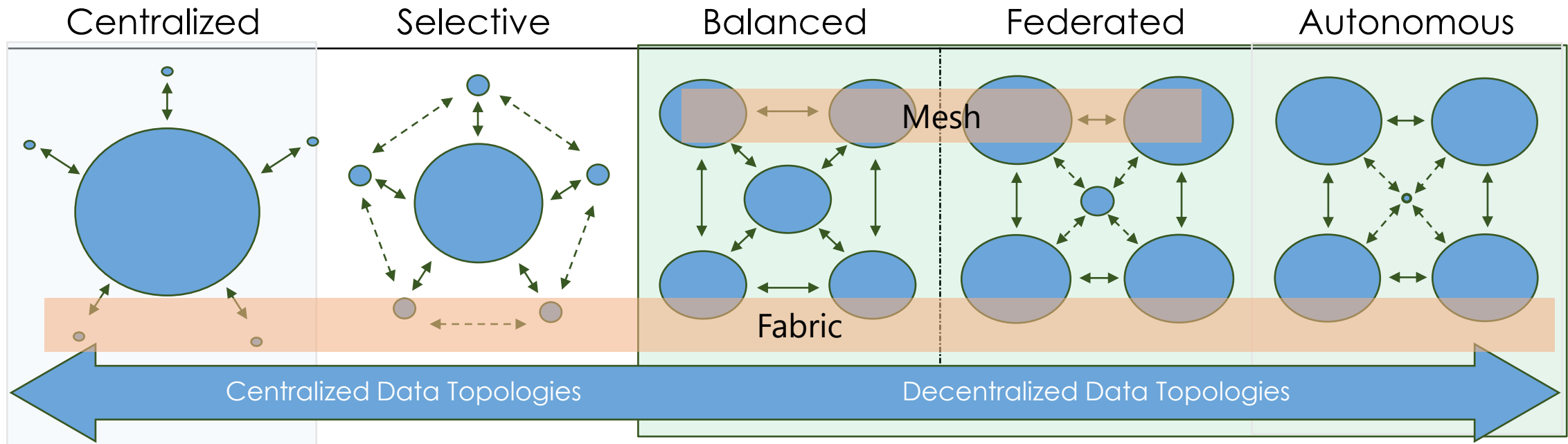
# Fabric and Mesh Compared

	Data Fabric	Data Mesh
Source	Gartner, 2021	Deghani, 2019
Emphasis	Technology oriented (Integration, Metadata, Virtualisation, Lifecycle)	Business oriented
Intent	Multi-cloud data management.	Domain oriented self-serve platform, data as a product
Affiliated Examples	IBM	Microsoft, AWS, GCP
Enabling capabilities	Many	Many
Technical Complexity	High	High
Topology	Open to interpretation	Decentralised
Ownership	Centralised	Shared (EIT & Domains)

Evolving viewpoint

# Data Topologies

Cultural Shift



Provides a single point of control and decision making, with limited input from outside IT

Selected decisions and activities are owned by the business

Decisions and activities are jointly shared between IT and the business

The business exerts control over decision-making and activities, with limited input from IT

Business units operate with autonomy  
– 'data mess' not mesh

# When do I need Mesh or Fabric?



[Episode 32: Is Your Data Fabric a Mesh? | data.world](#)

[Data Fabric or Data Mesh: How to Decide Your Future Data Management Architecture \(gartner.com\)](#)

# Data: Challenges



# What are your data challenges?

Go to [www.menti.com](http://www.menti.com) and use the code 3984 5238



# Challenges

Where do we start?

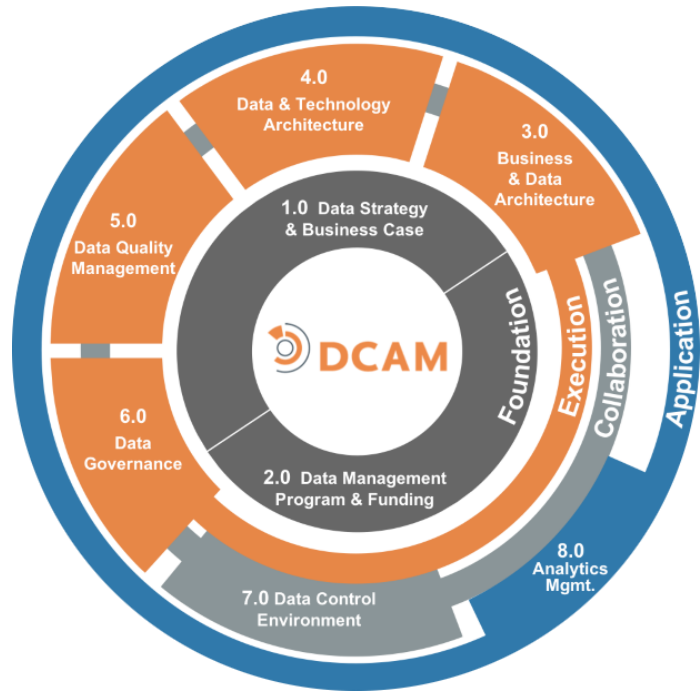


- Technology is the easy part
- Ethical considerations
- How is my data being used?
- Is it a sensitive use case?
- Consider AI impact assessment

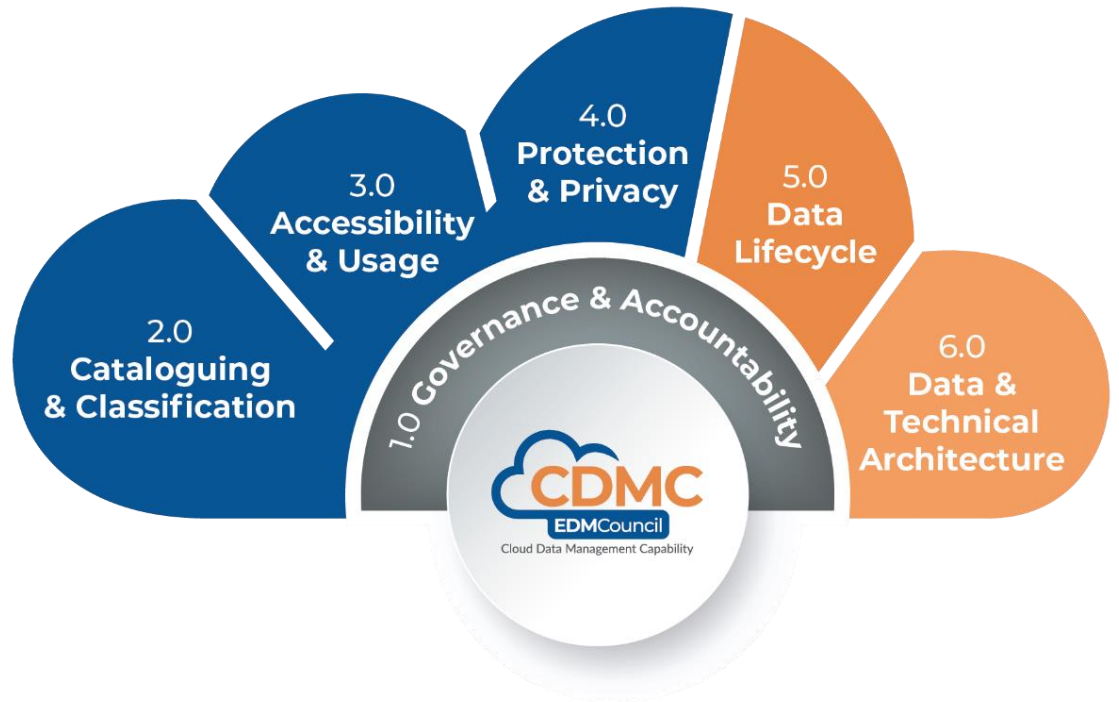


# Challenges

Where do we start?



Use DCAM to assess if you have the established foundations of a data and analytics management initiative

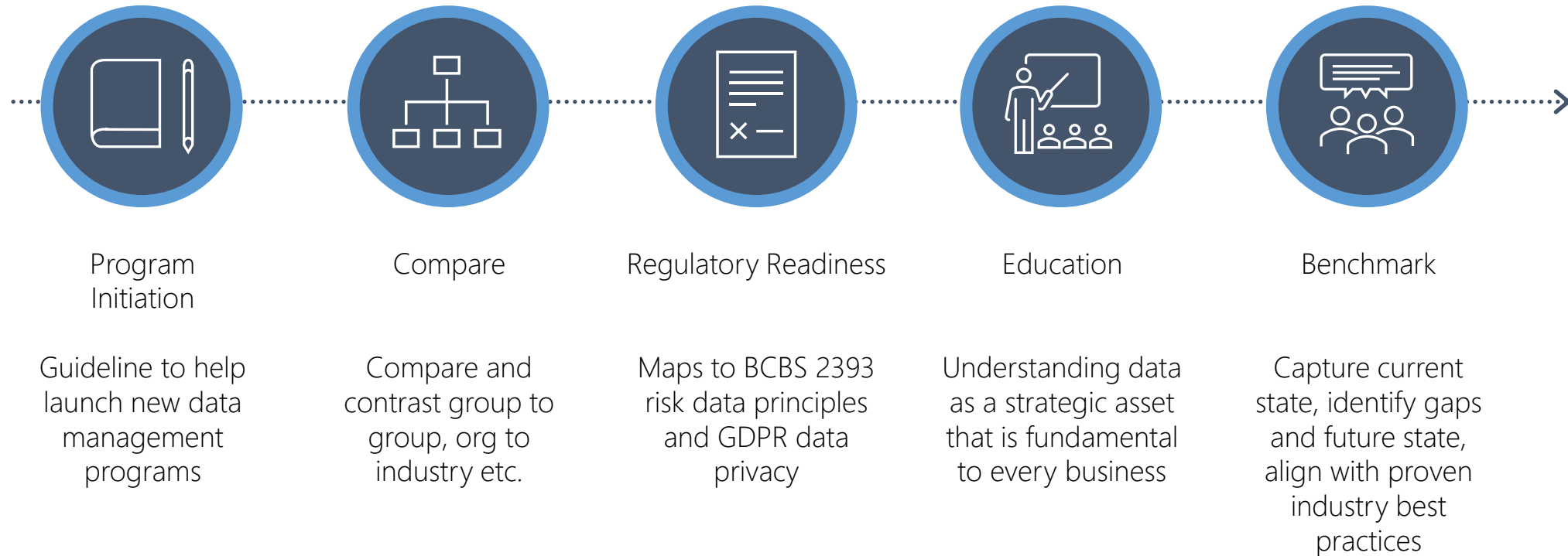


Use CDMC to assess if you have the capabilities to underpin a controlled adoption and migration to the cloud



# DCAM

Understand current state and how to get to target state



# Challenges

## Data Governance

### DISCOVERY

What data do I have?  
Where did the data originate?  
Can I trust it?



Chief Data Officer

### COMPLIANCE

What's my exposure to risk?  
Is my usage compliant?  
What is required by regulation X?



### CLOUD

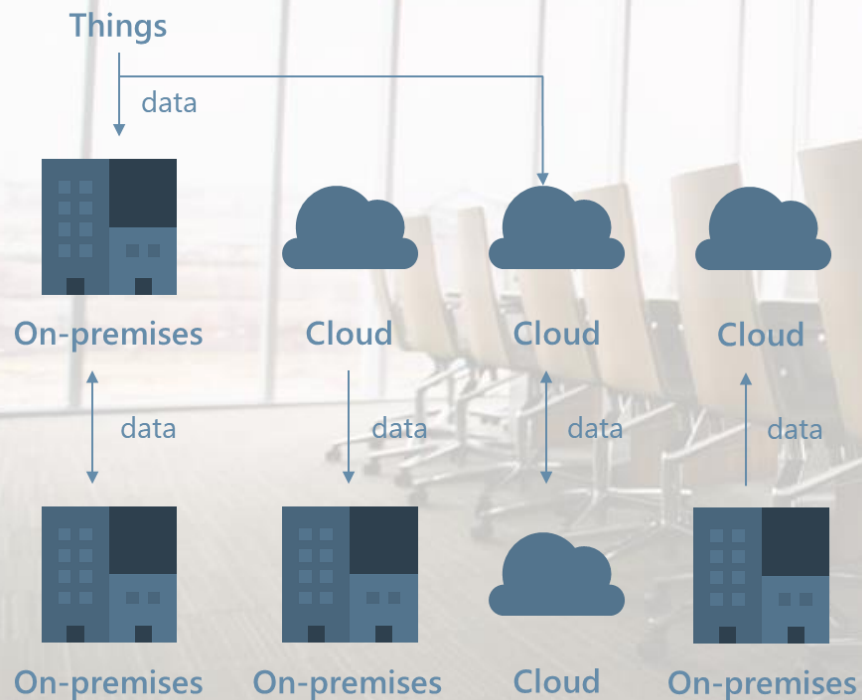
How do I manage data at cloud scale?  
My data estate is becoming more complex,  
how can I gain more control and  
drive more value?



# Challenges

## Data Governance

### Solving data issues in an increasingly complex data landscape



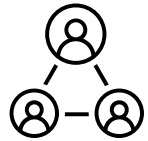
- 92% of organisations have a multi and hybrid cloud strategy
- The number of solution choices is increasing with over 15,000 SaaS products available
- Unstructured content is even more fractured, 80-90% of organizations' data is estimated to be unstructured
- By 2025, there will be over 100 zettabytes of data stored in the cloud

# Data: Futures

(or just around the corner..)

# Data: Futures

Looking round the corner....



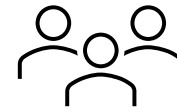
Adaptive  
Data Governance

Need for greater flexibility, select the right style for the business scenario (control, outcomes, agility, autonomous)



Data as a  
Product

Data assets are organized and supported as products, regardless of whether they're used by internal teams or external customers



Data  
Democratisation

Data accessibility and sharing is a key digital transformation capability, improve transparency, open data culture

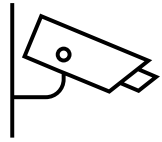


AI TRISM

AI Trust, Risk and Security Management

# Data: Futures

Looking round the corner....



Data Observability

Capture all data, providing more context to inform decision making



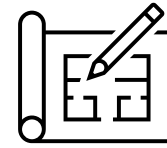
Real-Time

Increasing focus on real-time data for competitive advantage



Data-centric AI

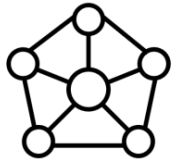
Focus on enhancing and enriching the data for the models, using AI to improve data quality



MV(D)A

“Just enough architecture” to support product release, small increments, flexible, tied to actual requirements

# Key Takeaways



Data mesh and data fabric are complementary architecture patterns, but not always the answer



Let the business strategy, data strategy and use cases influence architectural decisions



Establish and understand data estate maturity as a pre-requisite for successful advanced analytics and AI initiatives



Thanks!

Benjamin Wright-Jones

Senior Solution Architect  
Microsoft, Industry Solutions Engineering  
benjones@microsoft.com



@datalineage



# Questions



# Unlock the possibilities ... Join Us Today

Iasa Global is a non-profit association dedicated to the advancement of all business, enterprise, solution and technology architecture professionals.

We exist to support the development of the architecture profession as a whole.

<http://iasaglobal.org>

