Agenda

01   About SDG&E

02   Becoming a Data-Driven Organization

03   Data Mesh – Modernizing Data Management in the Cloud

04   Lessons Learned & Path Forward

About SDG&E

OUR VALUE

Do the right thing | Champion people | Shape the future

OUR MISSION

We improve lives and communities by building the cleanest, safest and most reliable energy infrastructure and company in America.

~4,600	1.3M	1.5M	908K	17.4K
Employees	Customers	Electric Meters	Gas Meters	Power Line Miles

4,100 sq. mi
in San Diego & S. Orange Counties

259,063
Total Solar/Wind Projects Authorized Under NEM (20% of Customer-Base)

Common Data Challenges

Bad data
Which data set should I use
How is this metric calculated

We need time to fix the data since the system is wrong
How has this data been changed from the source

Missing data

Access

The data I am using is old and not maintained
These data points don't match – which one should I believe

Trust

It takes too long to get answers

How do I request access to the data

Who needs to approve my data access I don't trust my data

I don't know where to find the data

Where did this data comes from

Becoming a Data Driven Organization

Accelerate Time to Insights

Re-use trusted data sources to enable rapid development; provide business leaders the data they need when they need it

Streamline Data Accessibility

Modernize data management; Reduce the amount of time needed to find and securely connect to the right data

Increase Trust in Data

Establish single-sources of truth; Measure and improve data quality for critical data elements

Data Initiatives Driving Outcomes

Initiatives supporting our data-driven transformation...

DATA GOVERNANCE

Management function to implement **enterprise-wide** processes and tools to create trusted, **information-based assets** and enable the management of **data as a strategic asset**
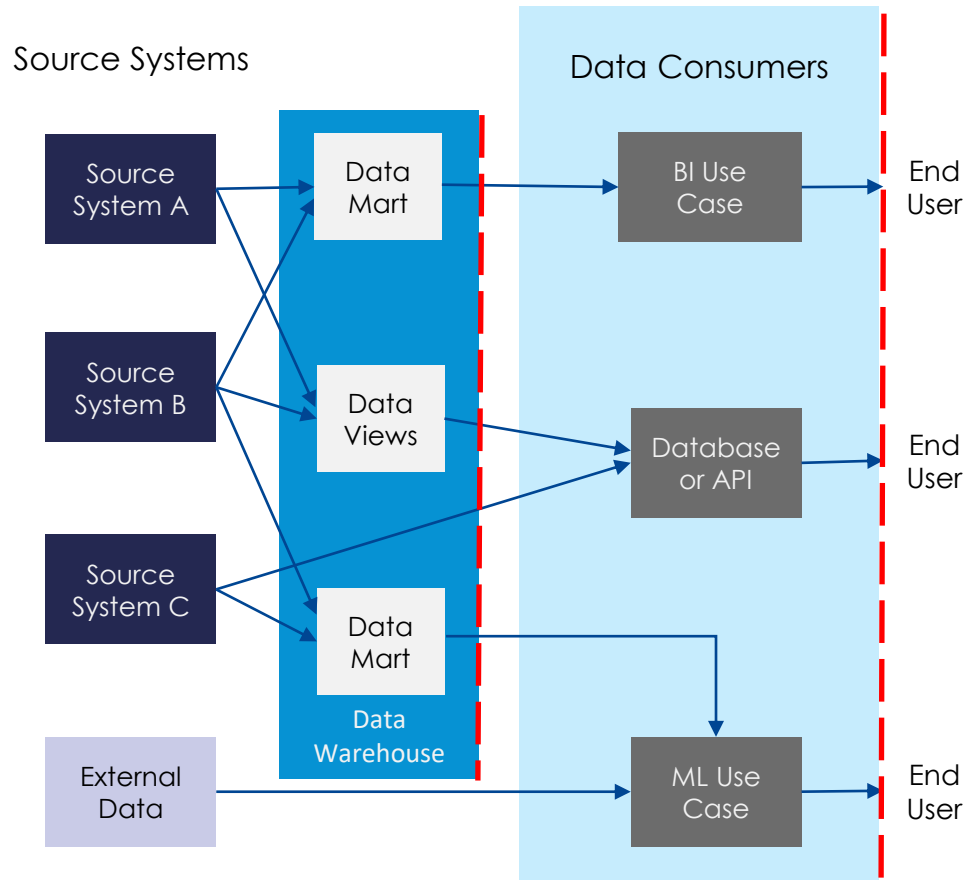
DATA MESH

A paradigm shift to building Data as a Product, enabling **data democratization, transparency and secure, streamlined access**
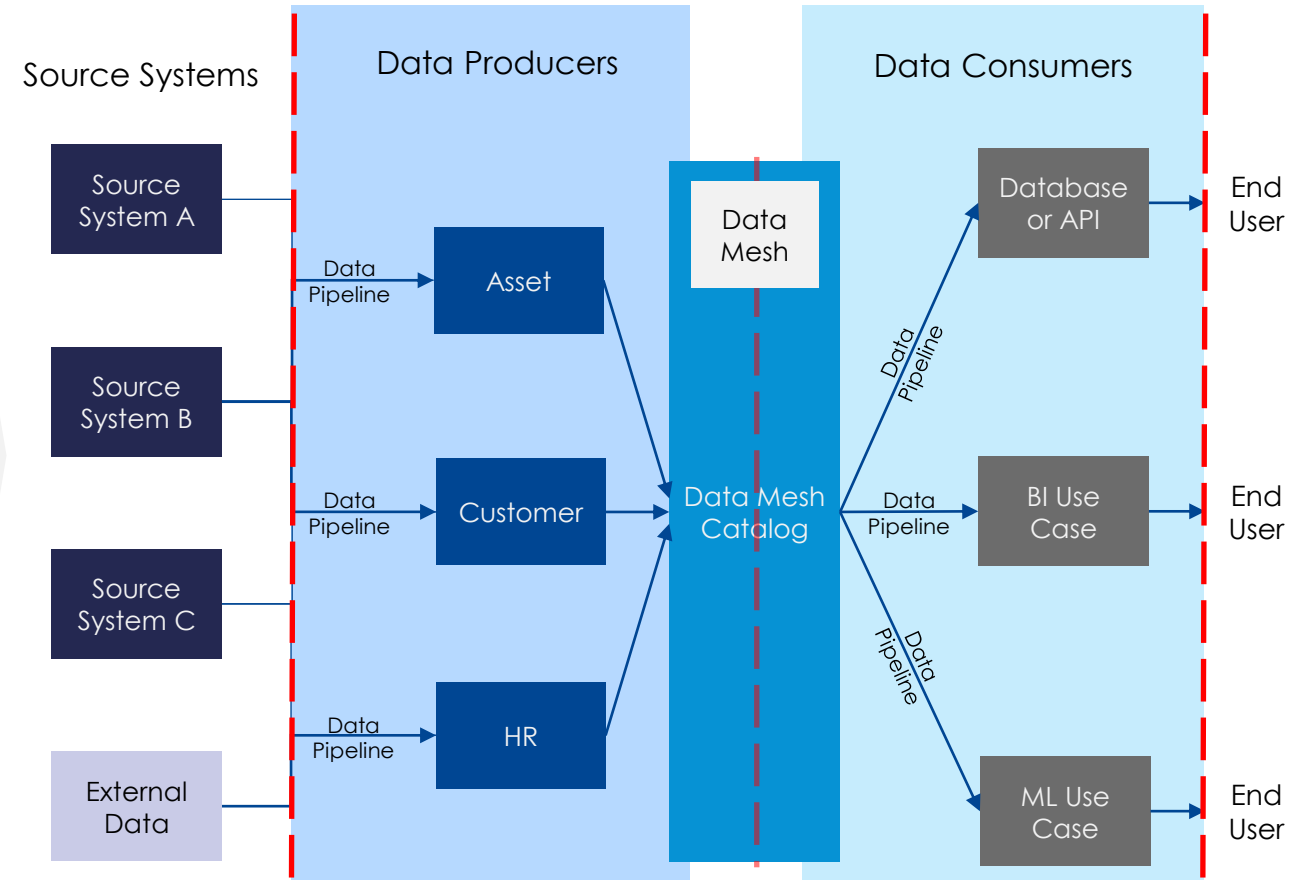
DATA SCIENCE

Looking to establish an Office of Data Science (target 2023)

From Data Warehouse to Data Mesh

Traditional BI/Data Warehouse (Today)

Data Mesh (Future)

Source Systems

Data Consumers

Source System A
Source System B
Source System C
External Data

Data Mart
Data Views
Data Mart
Data Warehouse

BI Use Case → End User
Database or API → End User
ML Use Case → End User

Source Systems

Data Producers

Data Consumers

Source System A
Source System B
Source System C
External Data

Data Pipeline → Asset
Data Pipeline → Customer
Data Pipeline → HR

Data Mesh
Data Mesh Catalog

Data Pipeline → Database or API → End User
Data Pipeline → BI Use Case → End User
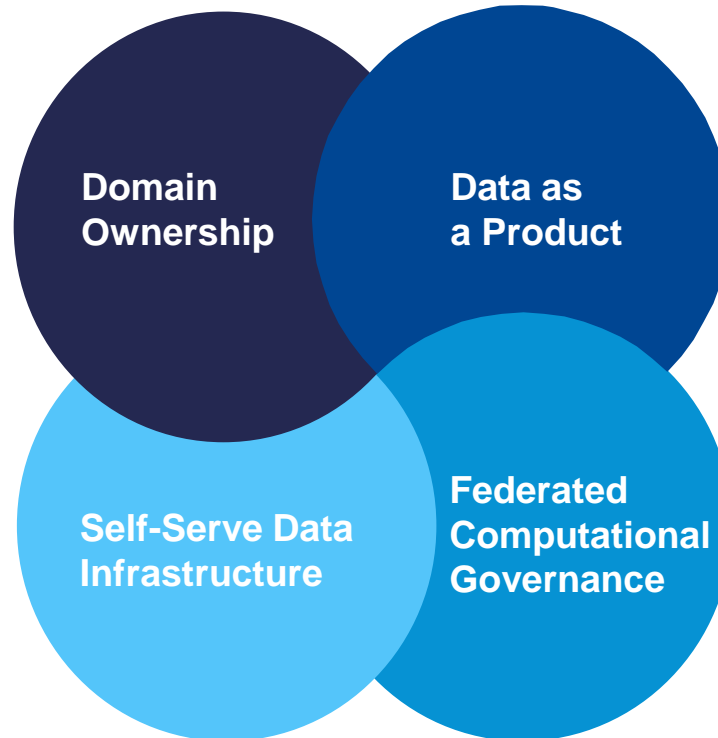Data Pipeline → ML Use Case → End User

Data Mesh Principles

People and Process over Technology

Domain Ownership

Business domains own the development and ongoing management of Data Products within the solution framework. Domains have autonomy, but Data products must adhere to common principles.
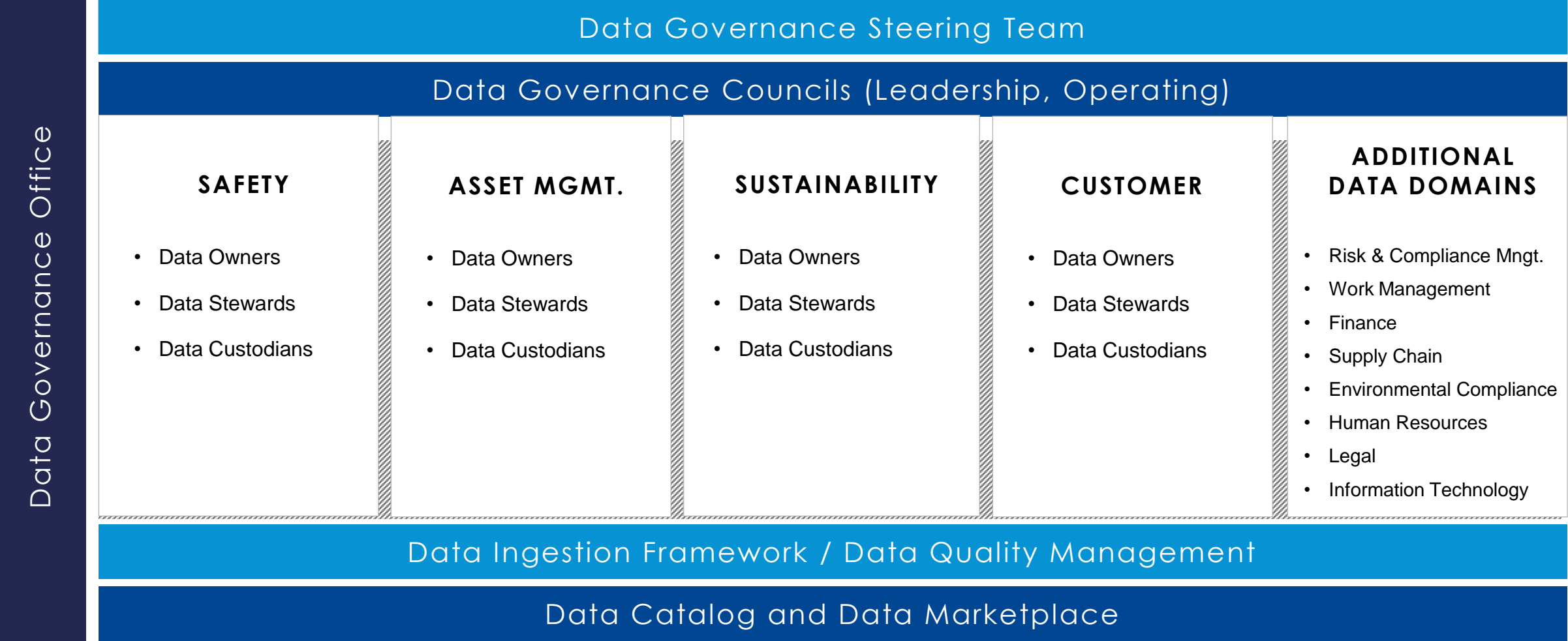
Data as a Product

Data should be forward-engineered to meet existing and emerging business needs instead of built-in arrears on a project-by-project basis. Products must have a clear purpose and value-centric user experience.

Domain Ownership

Data as a Product

Self-Serve Data Infrastructure

Federated Computational Governance

Self-Serve Data Infrastructure

Data products should be easily and flexibly addressable by domain-approved use cases while benefiting from a common backbone of security, data management, auditability, and other architectural goods.

Federated Computational Governance

Domains are responsible for decisions about their data, but we allow for centralization of cross-domain portfolio and master/reference data. Where possible, push decision-making down into the domains.

Data Domain Governance Structure

Data Governance Office

Data Governance Steering Team

Data Governance Councils (Leadership, Operating)

SAFETY	**ASSET MGMT.**	**SUSTAINABILITY**	**CUSTOMER**	**ADDITIONAL DATA DOMAINS**
• Data Owners • Data Stewards • Data Custodians	• Data Owners • Data Stewards • Data Custodians	• Data Owners • Data Stewards • Data Custodians	• Data Owners • Data Stewards • Data Custodians	• Risk & Compliance Mngt. • Work Management • Finance • Supply Chain • Environmental Compliance • Human Resources • Legal • Information Technology

Data Ingestion Framework / Data Quality Management

Data Catalog and Data Marketplace

Data Products

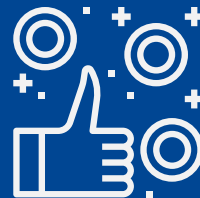**Definition & Expected Attributes of Data Products (Independent of Data Mesh)**

DATA PRODUCTS SHOULD HAVE THE FOLLOWING ATTRIBUTES:

Discoverability

Addressability

Trustworthiness

Descriptiveness

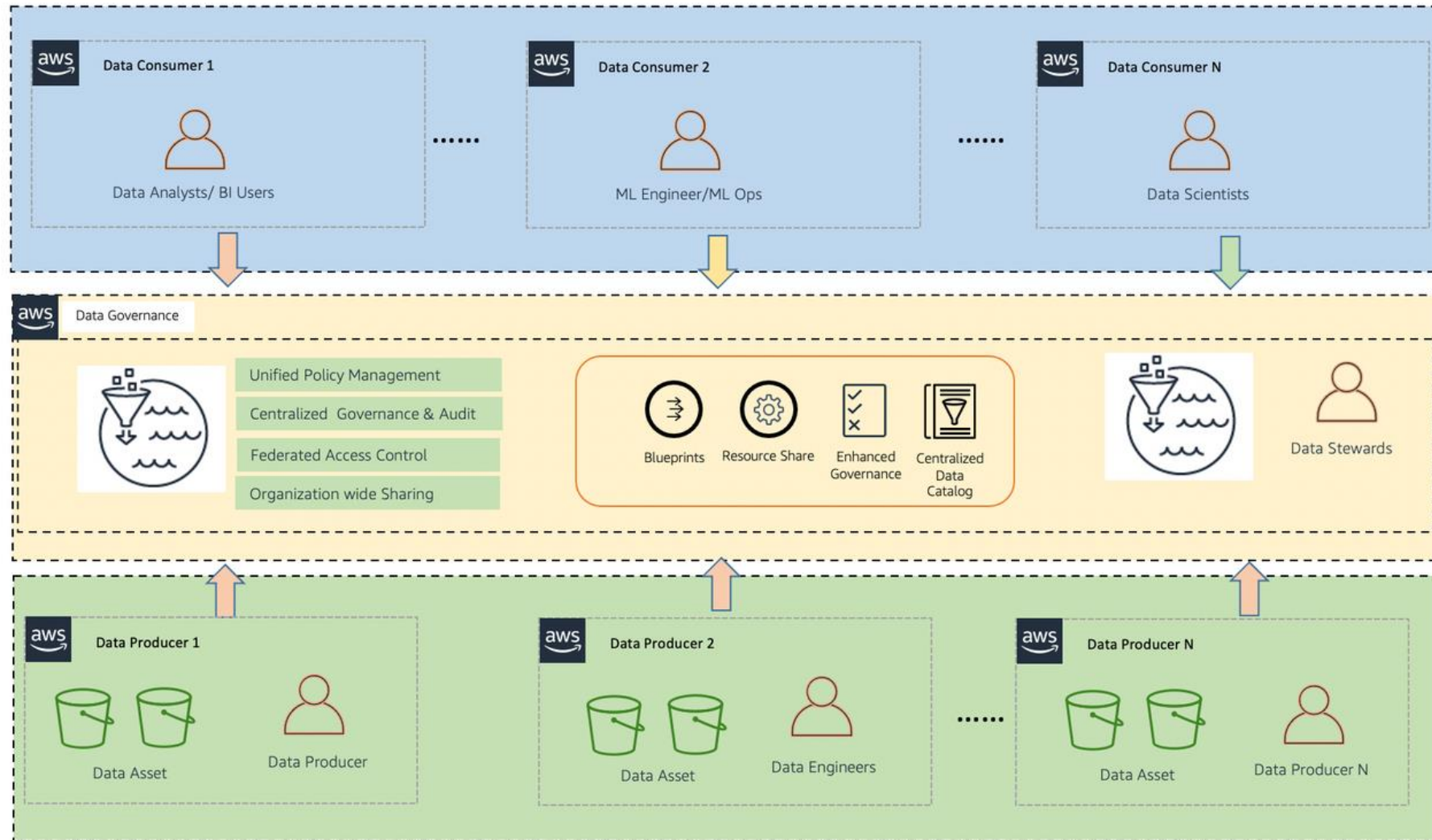**Interoperability**

Security & Accessibility

DATA PRODUCTS:

A Data Product is a reusable data asset, engineered to deliver a trusted dataset for a specific purpose at a guaranteed level of service

Data Products can include:
- *Data Sets*
- *API Endpoints*
- *Data Streams*
- *Machine Learning Models*
- *BI Dashboards*

Data Mesh Conceptual Architecture

Reference: https://aws.amazon.com/blogs/big-data/design-a-data-mesh-architecture-using-aws-lake-formation-and-aws-glue/

Where We are Going – Data Fabric

Exploring, Accessing, and Connecting to Data Products

DATA GOVERNANCE

Experience Layer (UX)

Metadata Plane (Data Mesh)

Data Plane (Data Products)

Metadata Plane (Data Mesh)
- Manages programmatic access to Data Products across all platforms
- Single source for Access Control
- Contains Data Product Metadata (Data Catalog)
- Data Definitions
- Business Data Models
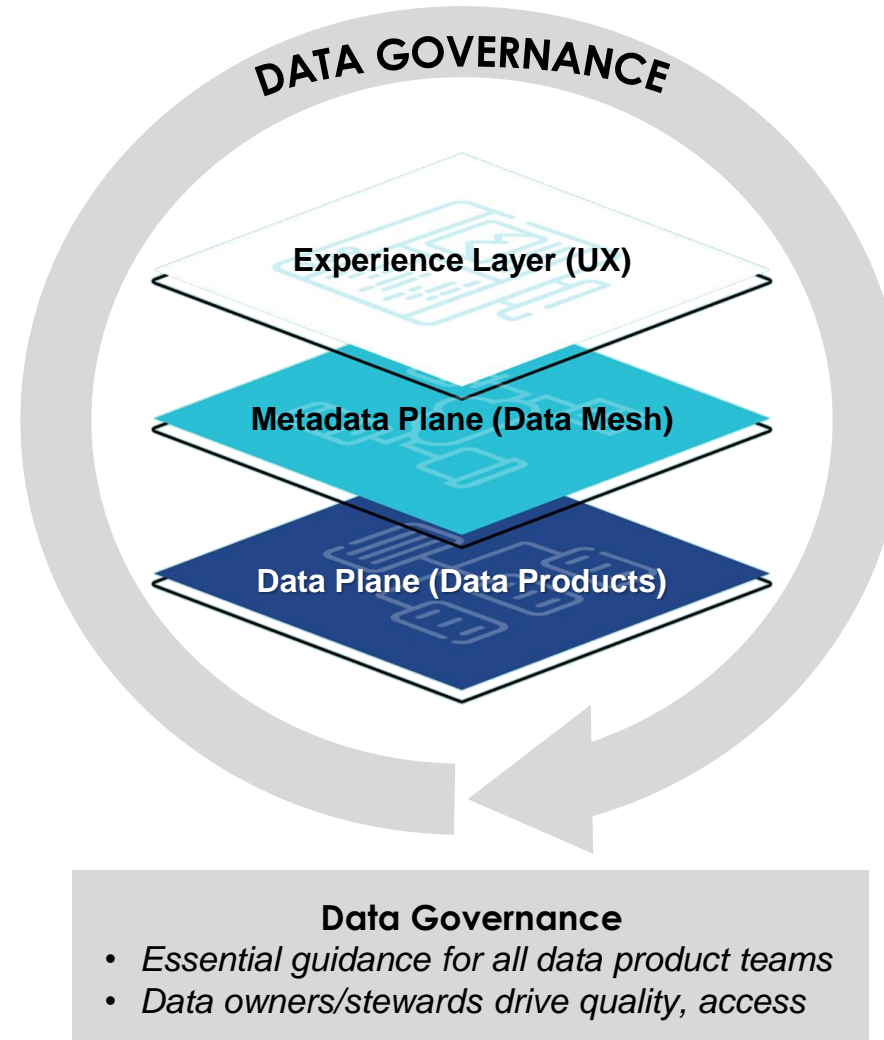
Experience Layer (UX)
- Front-end for Data Discovery
- Displays information on all Data Products
- Search engine queries for Products using information in the Metadata Plane
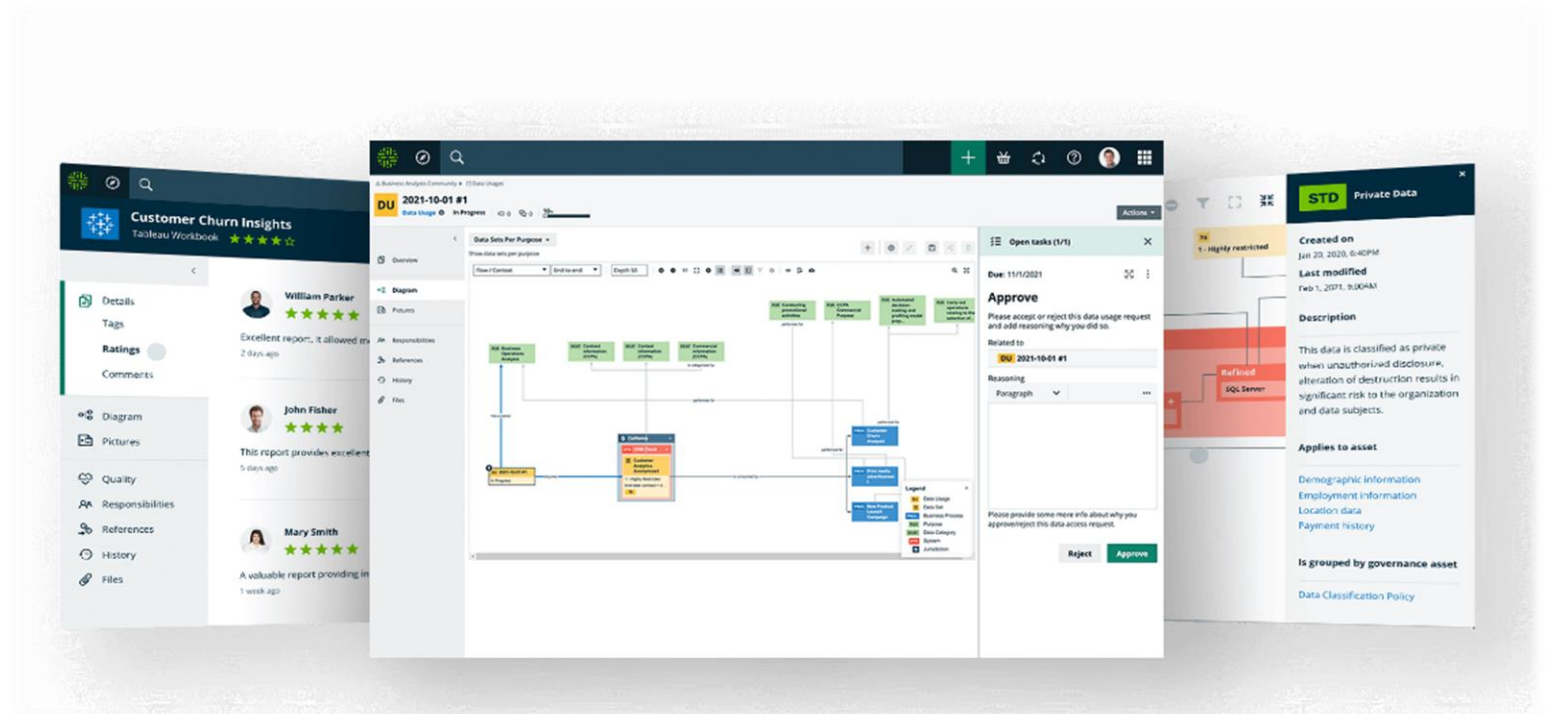- Data Marketplace

Data Plane (Data Products)
- Storage and Compute layer for Data Products
- APIs/data sets/streams/models
- Lives across multiple platforms

Data Governance
- *Essential guidance for all data product teams*
- *Data owners/stewards drive quality, access*
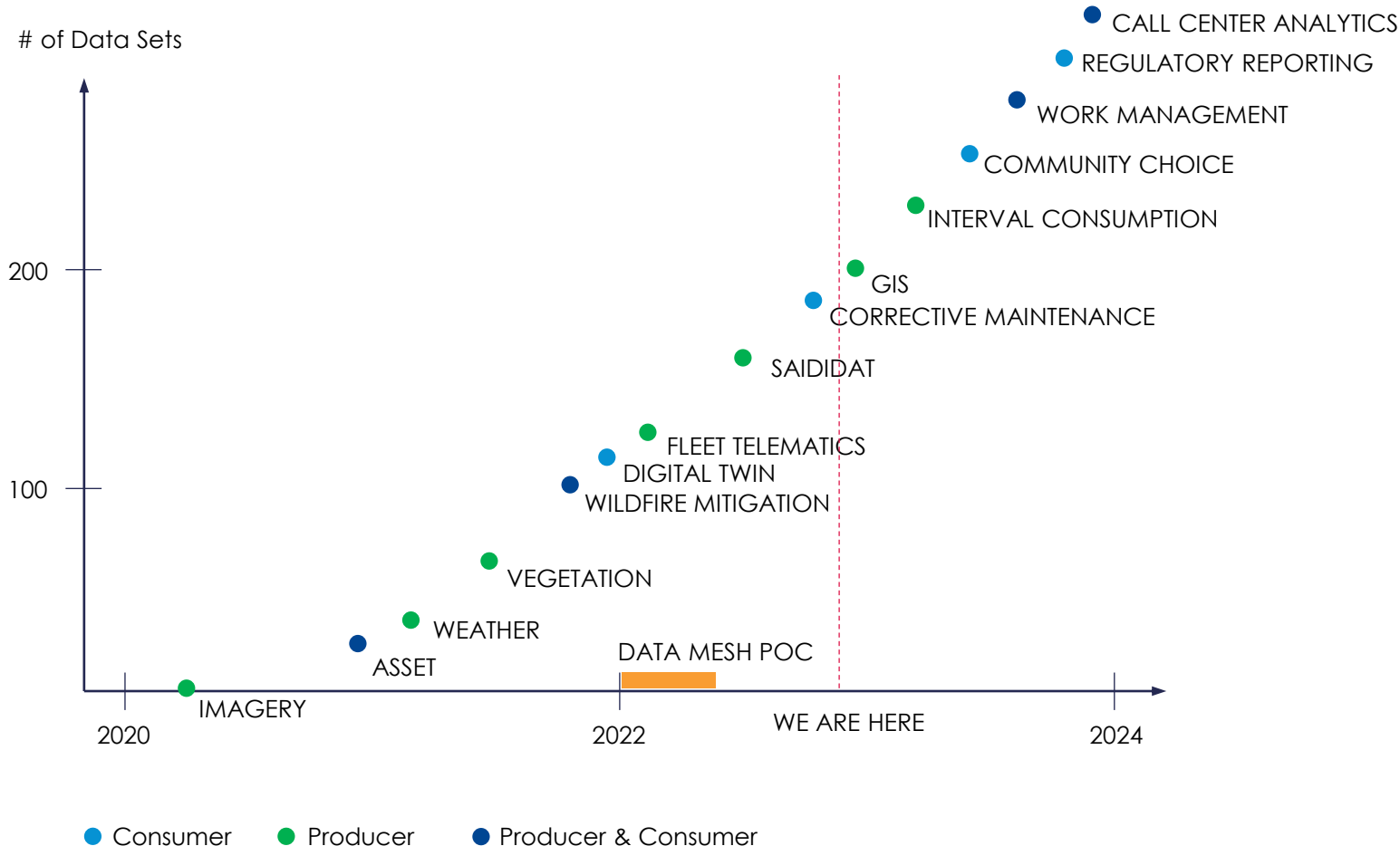
Self-Service Data Marketplace

SELF-SERVICE VIA CENTRALIZED DATA MARKETPLACE

- Empower data analysts, engineers, and scientists to source data quickly and easily
- Automated workflow for access controls
- Increase trust in data quality and accuracy via business glossary & data lineage
- Accelerate analytics time-to-value

Analytics Progress in Cloud

Reusable Data Products are key to speed-to-value

# of Data Sets

CALL CENTER ANALYTICS
REGULATORY REPORTING
WORK MANAGEMENT
COMMUNITY CHOICE
INTERVAL CONSUMPTION
GIS
CORRECTIVE MAINTENANCE
SAIDIDAT
FLEET TELEMATICS
DIGITAL TWIN
WILDFIRE MITIGATION
VEGETATION
WEATHER
ASSET
DATA MESH POC
IMAGERY

200

100

2020   2022   WE ARE HERE   2024

- Consumer
- Producer
- Producer & Consumer

CURRENT STATE:

Number of Data sets = 220

Number of ingestion jobs = 184

Volume of data = 7 TB

Average Cost over past 6 months = $1900 per month

TARGETED DATA SETS:

- Outage
- SCADA
- Safety
- Materials

- Gas Leaks
- Emissions
- Reference Data (Finance, HR, etc.)

Lessons Learned

But it's only been a few months ….

WHAT WORKED WELL

- Started with Proof of Value (PoV)
- Collaboration with Cloud Office & cybersecurity on account strategy, et al
- Focused on advancing one domain
- Grassroots support for building & sharing data products
- Initial efforts on OpModel and organizational change management
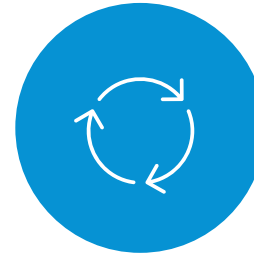
WHAT WE'D DO DIFFERENTLY

- Start Data Governance initiative earlier
- Bring business domain and technical teams together earlier
- Data onboarding backlog/roadmap
- Dedicate resources from business and IT to OpModel and organizational change management

Additional Considerations

- Role of Data Product Owner relative to Data Owner

- Data Tagging Schema

- AWS Account Strategy

- Chargeback Model

- Access Controls / Workflows

What's next in our journey?

SHORT-TERM

- Split out central Data Mesh team
- Complete Glue Data Catalog integration with Collibra
- Migrate sub-domain tagging and access control policies
- Implement monitoring/audit reporting (access, performance, etc.)
- Implement security controls for sensitive data (PI, NERC/CIP, etc.) into central catalog
- Balance use-case driven backlog with strategic data sets

LONG-TERM

- Make transition to Data Product management
  - Business domain leadership
  - Product SLAs, data profiling, ratings
  - Technology team enablement
- Implement Data Marketplace and workflow automation
- Complete onboarding of key domains
- Automate use case onboarding

THANKS FOR ATTENDING

Please fill out an evaluation form and drop it in the collection basket located at the back of the room.

**To learn about membership in the Utility Analytics Institute,
please visit www.utilityanalytics.com**

October 18-20, 2022
San Diego, CA