ANT336

Building data mesh architectures on AWS

Ian Meyers

Director, Product Management
AWS Analytics Service

Nivas Shankar

Principal Product Manager
AWS Lake Formation

Travis Muhlestein

Chief Data & Analytics Officer
GoDaddy

ANT336

Building data mesh architectures on AWS

Speakers

- Travis Muhlestein, Chief Data & Analytics Officer, GoDaddy
- Nivas Shankar, Principal Product Manager, AWS
- Ian Meyers, Director of Product Management, Amazon Web Services

Announcements

- None

Takeaways

- Good overview of modern Data Mesh and how it might be realized on AWS
- In 2022 Data Management Hype Cycle Gartner proclaims Data Management Obsolete before plateau.

Azure

- You can build a Data Mesh architecture in Azure too.

< Section Header
< Table of Contents

Modern Data Strategy & Challenges

- See ANT205 – Achieving your modern data architecture

Why Data Mesh

- Acknowledges that data challenges exist and seeks to work within those constraints.
  - Use existing investments in data platforms and treat them as independent domains
  - Improve data governance by pushing policy down to data domains
  - Provides a clear mechanism for centralized data discovery
  - Provides self-service data sharing features to allow owners to grant access to consumers
  - Allows leaders to measure and invest in data products based on usage and business value.

Design Goals & Core Principles - Watch

- Four core principles of a Data Mesh Architecture
  - Data domain ownership
  - Data as a Product
  - Federated computational governance
  - Self-serve Sharing
- Data Mesh Design Goals
  - Enable organizations to get value from data at scale
  - Create business-oriented data products that can support the top strategic goals
  - Allow business domains federated governance through lightweight centralized policy by removing bottlenecks
  - Encourage data-driven agility
  - Support the sharing of data products, with the goal of delighting the experience of data users
- Data Mesh Architecture – How it might look in AWS
  - Decentralized, lightweight federated governance across domain-oriented data systems to drive governed sharing.
  - Lake Formation and Data Zone allow for policies on what the data is and who can access it.

Favorite Quotes

- “Everyone wants to be a data consumer, no one wants to be a producer”

Why GoDaddy built a Data Mesh

- Watch
- Business Outcomes of the Strategy
  - Created hierarchical views of data products at different levels so that business users can analyze information to make quicker business decisions
  - Automated access management framework to enable self-served access to data within and across lines of business
  - Accelerated the data platform adoption to 10+ LOBs and 300+ teams globally
  - Enabled data scientists to find and access data needed to generate ML models across LOBs
  - Achieved 10s of millions of dollars in cost savings from data reuse and better management of purchased datasets.

Why data mesh?

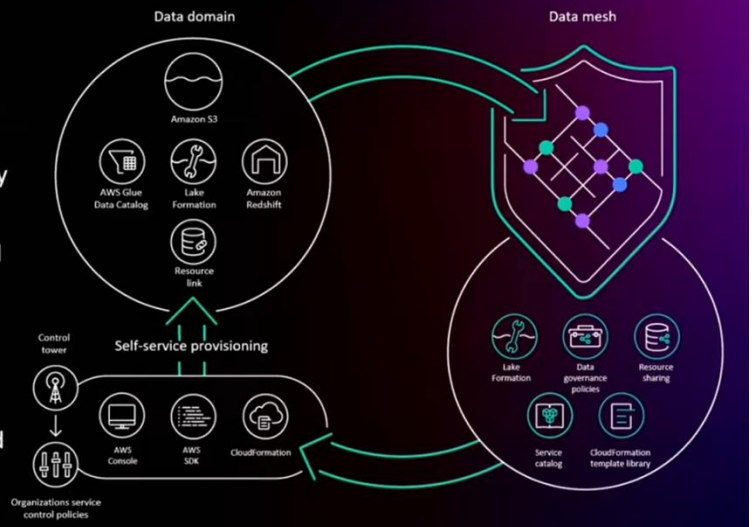
Use existing investments in data platforms and treat them as independent “domains”

Improve data governance by pushing policy down into data domains

Provides a clear mechanism for centralized data discovery

Provides self-service data sharing features to allow domain owners to grant access to consumers

Measure and invest in data products based on usage and business value



Data mesh: Four core principles

Data owner

Data domain ownership

A data mesh features data domains as nodes, which exist in data lake accounts; it is founded in decentralization and distribution of data responsibility to people closest to the data

Data engineer

Data as a product

A data producer contributes one or more data products to a central catalog in a data mesh account; DaaP must be autonomous, discoverable, secure, and correct, and useable

Data steward

Federated computational governance

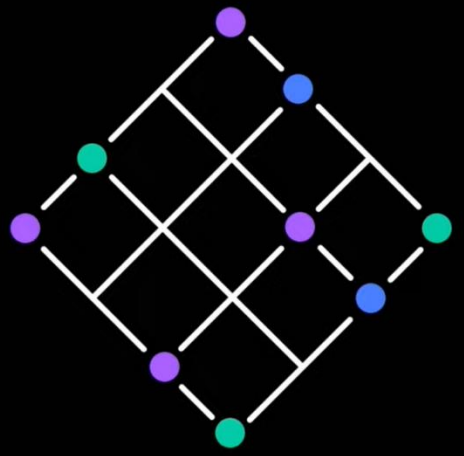
Federated data governance is how data products are shared – delivering discoverable metadata auditability based on federated decision-making and accountability structures

Data consumer

Self-serve sharing

The platform streamlines the experience of data users to discover, access, and use data products; it streamlines the experience of data providers to build, deploy, and maintain data products

Data mesh design goals



Enable organizations to get value from data at scale

Create a business-oriented data products that can support the top strategic goals

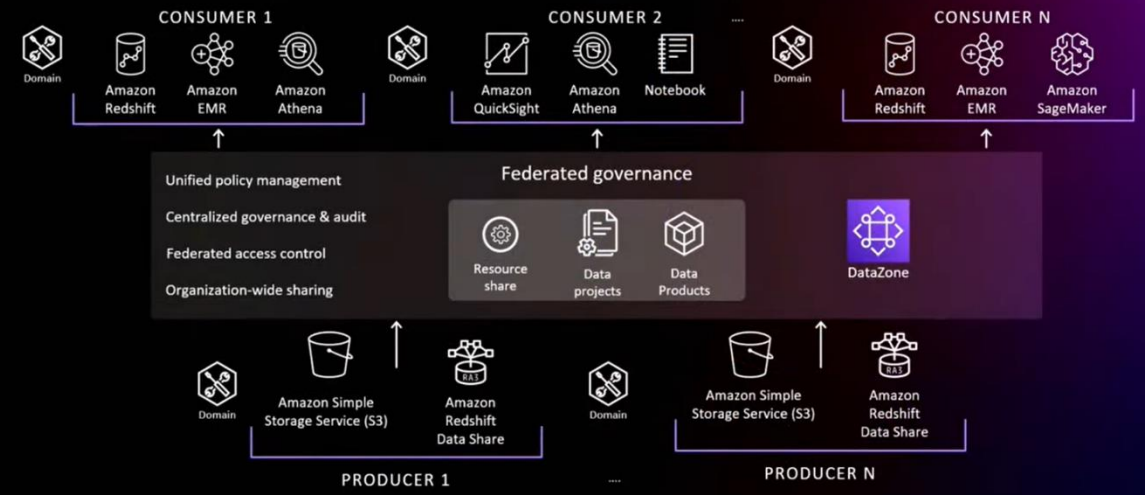
Allow business domains federated governance through lightweight centralized policy by removing bottlenecks

Encourage data-driven agility

Support the sharing of data products, with the goal of delighting the experience of data users

Data mesh architecture

DECENTRALIZED, LIGHTWEIGHT FEDERATED GOVERNANCE ACROSS DOMAIN-ORIENTED DATA SYSTEMS TO DRIVE GOVERNED SHARING



Data mesh principle #1: Data domain ownership

DATA OWNERS ARE ACCOUNTABLE FOR THEIR DATA PRODUCTS TO BE RELIABLE, AVAILABLE, AND ACCURATE

Data owner | Accountability for data domain and consumption of data products

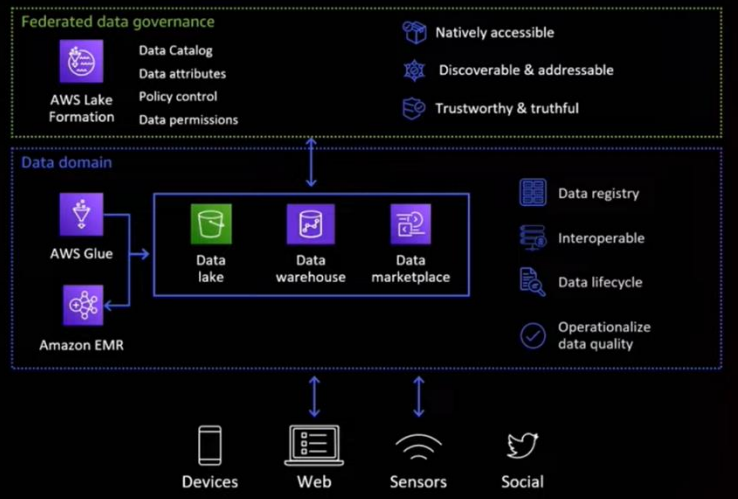
Create
Owns data **created** or **transformed** in my organization

Store
Owns data **stored** in my organization

Share
Owns how my organization **uses** the data

Data mesh principle #2: Data as a product

DOMAIN-DRIVEN DESIGN TECHNIQUES TO FORMULATE AND ESTABLISH BOUNDED CONTEXTS FOR DATA PRODUCTS

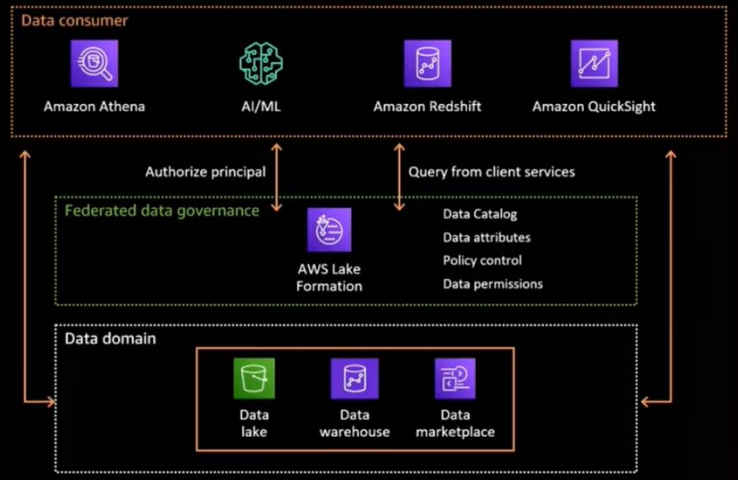


Data engineer

- Each team manages their data and organizes it as data products
- Each product provides an interface(s) to allow others to interact (e.g., APIs, SQL, reports)
- Remove usability frictions, meet the user where they are
- Provide all supporting metadata, lineage
- Data products are valuable on their own

Data mesh principle #3: Self-serve sharing

ECOSYSTEM OF SELF-SERVE DATA INFRASTRUCTURE WITH OPEN PROTOCOLS

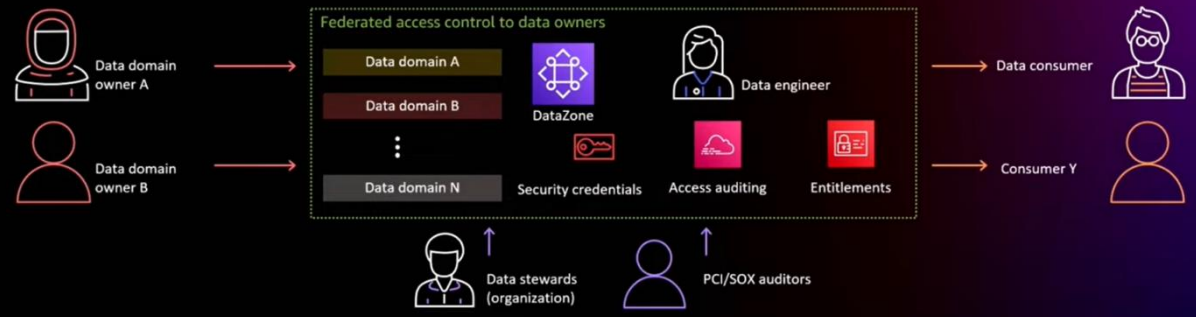


Data consumer

- Design for generalist majority (i.e., make it easy to use and adopt with no specialist skills needed)
- Enable personas to discover, learn, understand, consume, and maintain data products
- Collection of interoperable data products, which enable cross-functional domains to produce and consume data easily and with autonomy and will allow it to scale
- Data products must include data, metadata, code, and policy all as single unit of value
- Abstract complexity through automation

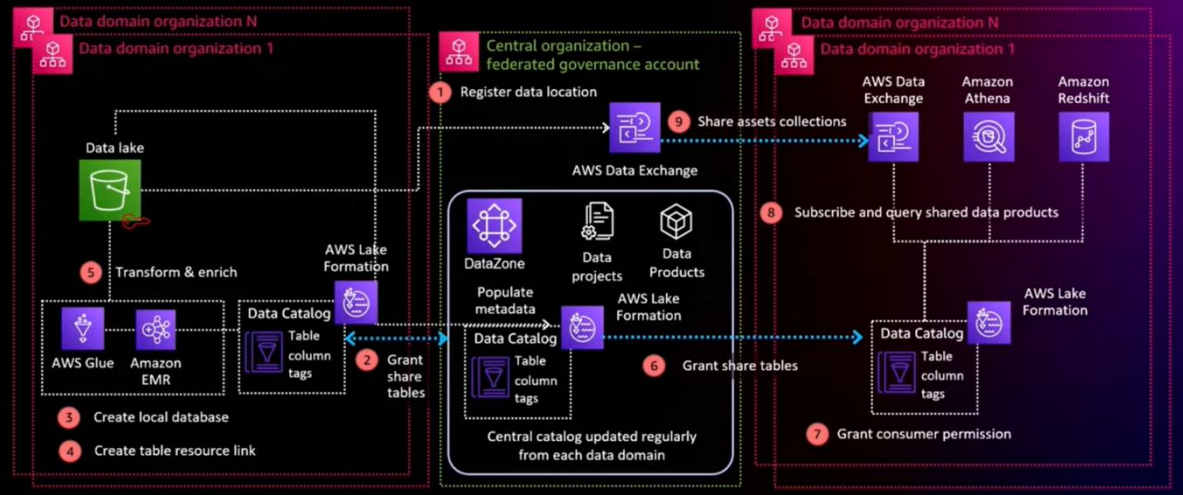
Data mesh principle #4: Federated data governance

GOVERNANCE MODEL THAT EMBRACES DECENTRALIZATION AND DOMAIN SELF-SOVEREIGNTY THROUGH DECISION-MAKING MODEL LED BY FEDERATION OF DATA PRODUCT OWNERS

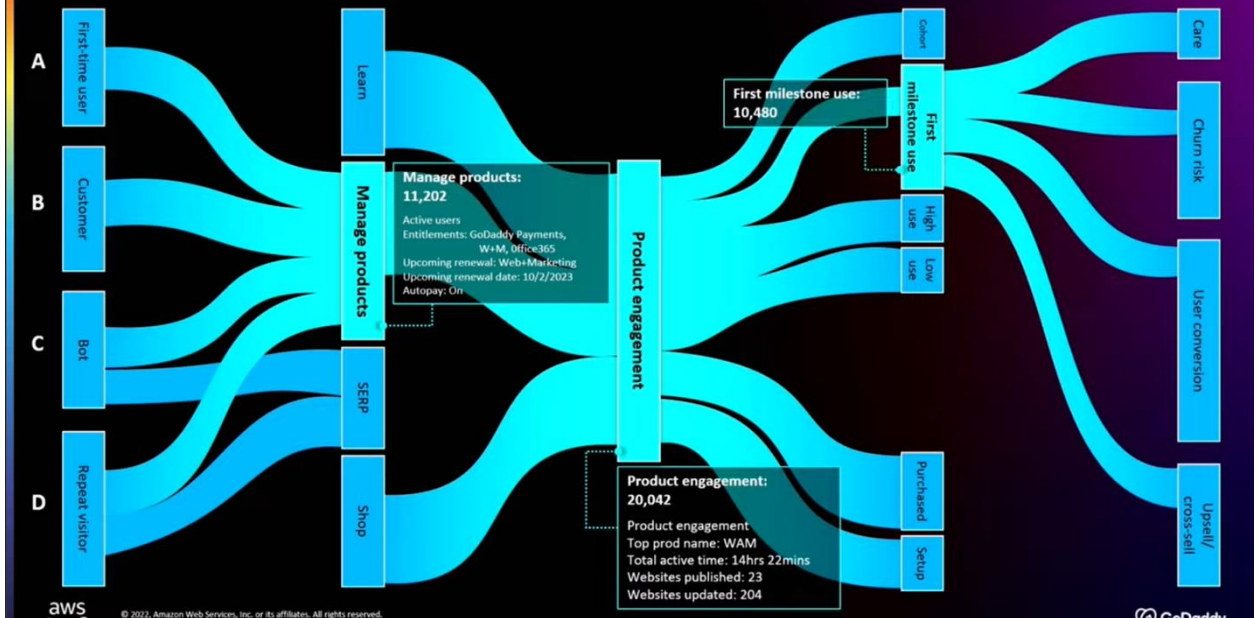


- Decentralization implementation of governance team and standard authorization
- Governance team = a guild consisting of representatives of all teams taking part in the data mesh
- They create global policies and standardization to achieve interoperability
- Automated execution of policies by the data domains (e.g., data classification and privacy, compliance, security, documentation, and interoperability)

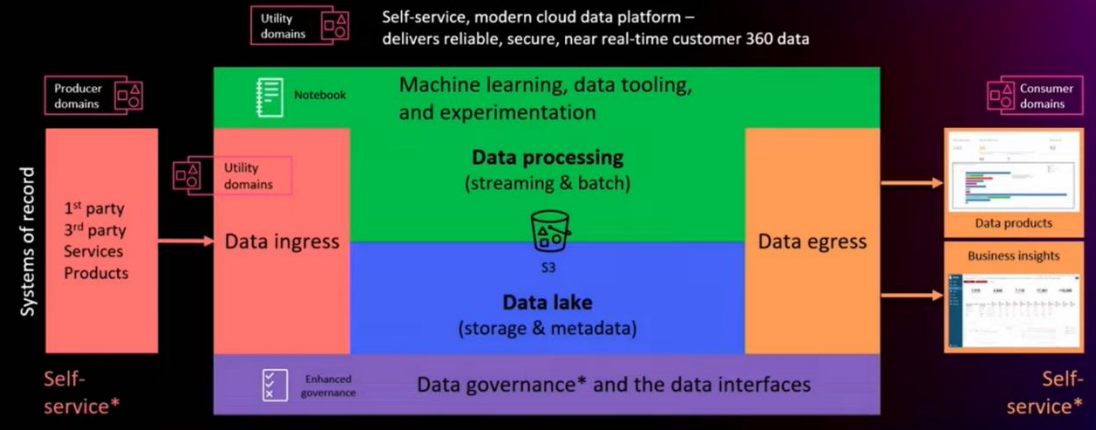
Data mesh architecture pattern: Data lake data products sharing



Multi-domain federated customer actions



Data mesh



GoDaddy conceptual/ domain architecture