WORKSHOP

DATA ENGINEERING

Jay Sen

Director | Data Engineering
PayPal

Applied Data Mesh Workshop for Scalable Data Platforms

ODSC EAST 2025
THE LEADING AI BUILDERS CONFERENCE

MAY 13-15
BOSTON

Intro & Abstract

As data volumes and business demands grows, traditional centralized data architectures are unable to scale for agility and speed. Data Mesh is a modern architectural paradigm that redefines how organizations produce, share, and consume data—by aligning data ownership with domain teams, treating data as a product, and enabling scalable self-service platforms under federated governance.

This 1-hour workshop is designed for data engineers looking to move beyond monolithic data platforms and explore the principles and practices of Data Mesh in real-world scenarios. Participants will gain a deep understanding of the four core principles—domain-oriented ownership, data as a product, self-serve infrastructure, and federated computational governance—and how they apply to designing and operating modern data platforms.

Through a mix of technical presentations and hands-on exercises, attendees will:

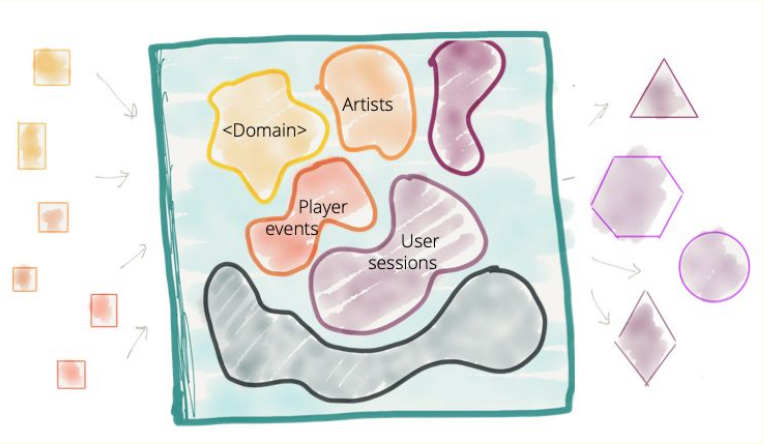
- Model domain-specific data model with ownership boundaries.
- Design and document and implement data products with real consumer needs using domain business events.
- Map out requirements for a self-serve data platform.
- Explore governance trade-offs and automation strategies.

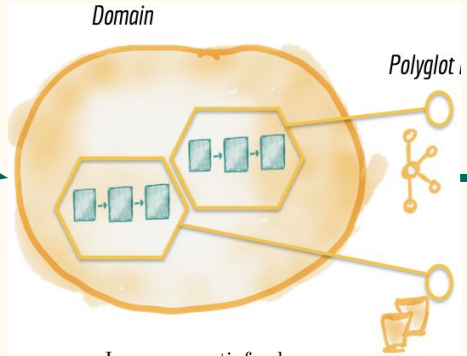
Whether you're building the next-gen data platform or modernizing existing pipelines, this workshop will help you apply Data Mesh patterns to increase autonomy, scalability, and data quality across your organization.

Why Data mesh?

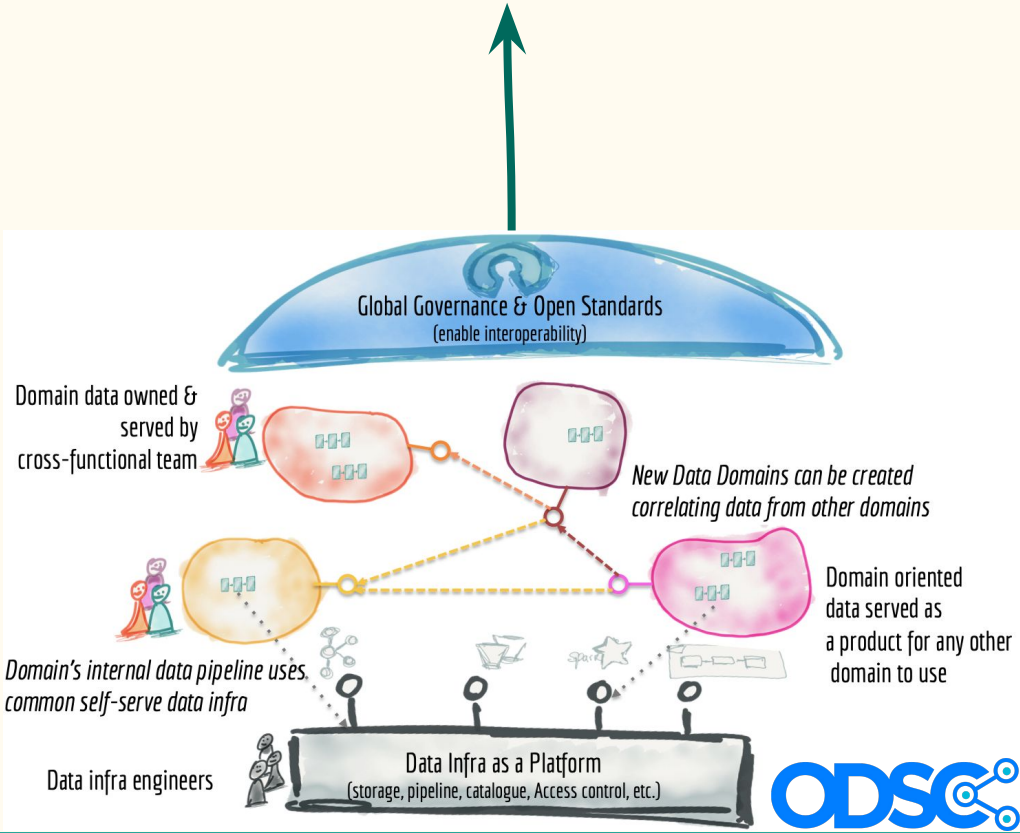
Business Domain:

- ✓ Agility & Scalability, & Speed of data delivery
- ✓ Overall Velocity

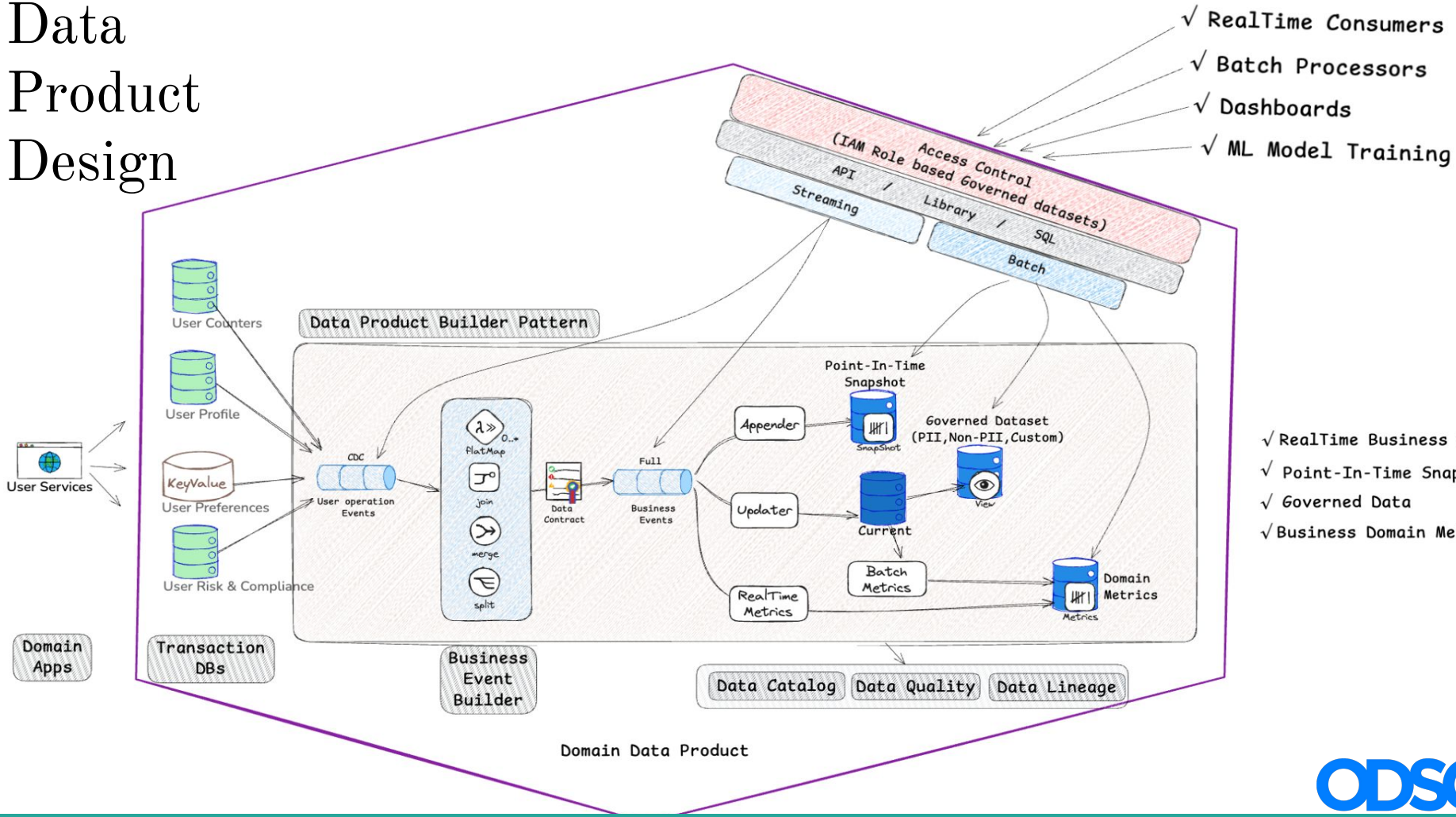




Images: martinfowler.com



Data Product Design



Workshop Prerequisites

https://github.com/jhsenjaliya/data-product-demo/tree/main/src/main/resources

Workshop Goals

- Learn practically the concept of Domain, Business Events & Data Product.
- Understand Pain points of the data consumers.
- Understand Data Mesh principle and how it helps with problems of large enterprises.
    - Domain-oriented decentralized data ownership
    - Data as a product:
    - Self-serve data platform
    - Federated governance
- Learn how to increase velocity for large organizations !

Federated Governance

- **Objective:** Identify areas of governance and create guidelines and policy for all domains to follow.
- **Task:** Simulate how federated governance is applied across domains by:
    - Define global standards (naming, privacy classification, schema types).
    - Decide on policies to automate via platform enforcement. (ex: should not contain teenage user accounts)
    - Creating a mock “data contract” to simulate policy-as-code.
- **Time:** 10 minutes.

Domain Data Modeling

- **Objective:** Identify and define data domains and their ownership boundaries in a sample business scenario.
- **Scenario:** *Imagine "Acme Retail Co."* – it has multiple business domains (User, Transactions, Sales & Marketing etc…). Currently all analytics go through a central data team.
- **Task:** In small groups (or individually), **map out 2 to 4 domain data areas** for Acme Retail. For each domain, list what key data it would own as a product (e.g., Payment domain owns transactions data, Marketing owns campaign data, User domain owns all user data, etc…).
- **Considerations:** Think about how the organization is structured – domains usually align with business functions. Ensure every important data source is owned by one of the domains (no important data left unowned).
- **Time:** 5 minutes to brainstorm and sketch the domain model.

Data as a Product

- **Objective:** Apply the “data as a product” mindset by designing a data product for one of the domains identified earlier.
- **Setup:** Pick one domain from the domain modeling exercise (e.g., User or Payment domain in our retail scenario). Assume your team owns that domain’s data.
- **Task: Define a new data product** for that domain:
    - What valuable dataset or analytics could your domain provide to others? (e.g., Sales domain might offer “Daily Sales Trends” data product for Finance and Marketing to consume in addition to core transaction data).
    - **Consumers:** Who are the intended customers (data consumers) of this data product, and what do they need from it?
    - **Data & Features:** What data will it contain? Outline its key data fields or metrics. Ensure it’s high-quality and documented.
    - **Delivery:** How will consumers get it (batch file, API, query, etc.) and how often is it updated?
    - **Success Metrics:** How will you measure if this data product is successful? (e.g., adoption rate by other teams, data quality metrics, user feedback).
- **Time:** 10 minutes to design.

Self-Serve Platforms

Objective: Identify what platform components and tools would be needed to support the data product your group designed.

Task: For your data product, consider the **infrastructure and tooling** you'd want from a self-serve data platform. Make a short list:

- Data storage/processing: do you need a streaming platform (Kafka) or batch data store (Iceberg, parquet) or both ?
- Pipeline orchestration: how will you build and schedule the data pipeline (perhaps using a tool like Airflow, dagster, prefect)?
- Metadata & discovery: how will others find and understand your data (e.g., a dataHub, OpenMetadata)?
- Monitoring & quality: what tools to monitor pipeline runs or data quality checks?
- Access management: how to manage who can access the data product (Views, Permissions, Roles)?

Tools: You won't be using the actual tools in this exercise – just brainstorming. Use sticky notes or a doc to record the platform features or specific technologies you'd choose for each need.

Time: 5 minutes to outline the platform support for your product, then we'll discuss a few examples.

Q & A

Thank You !