Democratizing Data with Data Fabric and Data Mesh

Powered by Data Virtualization

Chris Walters
Senior Sales Engineer
Data landscape is complex

- Data in multiple silos
- Every business data need is a data integration project
- Inconsistencies

Avg. of 400 sources used for analytics in big organizations

Source: IDG, 2021
What’s Denodo’s recipe?

- **Embrace distributed data landscape**
  - Embrace the fact that data resides in multiple locations or systems – on-prem, hybrid, multi-cloud. All data needs to be managed with consistency.

- **Use a Logical approach to manage it**
  - Consumers access data through a centralized semantic model, decoupled from data location and physical schemas, that can enforce security and governance requirements.
Logical Abstraction of Disparate Datasets

1. Source Abstraction
2. Data Model
   Combine, Transform & Semantics
3. Publish
4. Dev/Ops

Consuming Applications

Sources
Data Fabric Principles

• Definition
  • A unified data architecture that integrates data from various sources and makes it accessible to users across an organization

• Core Principles
  • Data integration - integrates data from multiple sources and makes it available in unified views
  • Data governance - ensures that data is managed in a consistent, compliant, and secure manner
  • Data democratization - enables self service access to easily and quickly discover and analyze data
  • Data intelligence - provides insights into data, relationships, usage context, and data lineage
  • Data interoperability - allows data to be seamlessly exchanged between different systems and applications
Data Fabric Overview

Data Fabric Delivers Integrated Data To All Data Consumers

- Compounds
- Customers
- Products
- Claims

Data Fabric

RDBMS
- Flat Files
- Third Party
- Legacy
- Data Warehouses/ Marts

Traditional Analytics/BI
- Hadoop, File Stores
- Cloud Data Warehouses, Cloud Data Lakes

Data Lakes

Cloud Data Stores

Apps and Doc. Repositories
- XML, JSON, AVRO, PDF, DOC, WEB
Data Mesh Principles

• Definition
  • A set of data management principles for breaking down monolithic data architectures into smaller, more autonomous units and managing data in a distributed, decentralized way.

• Core Principles
  • **Domain-driven design** – systems built around business domains where data is organized and managed by domain experts who have a deep understanding of the data's business context
  • **Decentralization** – data ownership and governance are decentralized to individual domains, rather than centralized in a single team or department
  • **Self service** – Domains are provided with self-service capabilities which enable domain experts to build and manage their own data products without needing to rely on centralized IT teams
  • **Data as a product** – data is treated as a product, with clear ownership which helps to ensure that data is of high quality, reliable, and valuable.
  • **Federated governance** - is governed through a federated model, which enables domains to collaborate and share data products while maintaining control over their own data
Data Mesh Overview
Demo
**Demo Scenario**

**Objectives**
- Launch new digital transformation initiative
- Requires strategic use of data assets
- Integrate data across all business units for a new project
- Provide data to business users to make data driven decisions
- Empower users to find and gain access to data in a secure manner

**Solution**
- Organize data using Data Mesh principles
- Enable data access using Data Fabric principles

**Data Distribution**
- Marketing data is in a Cloud app
- Sales data in a data lake
- Customer data in CRM database
2. Domains connect their data sources. A common domain may be useful to centralized data products common across domains.

1. Each domain is given a separate virtual schema. A common domain may be useful to centralized data products common across domains.

4. Domains SMEs can model their Data Products. Products can be used to define other products.

7. Products can be access via SQL, MDX or exposed as an API. No coding is required.

6. A central team can set guidelines and governance to ensure interoperability.

5. For execution, Products can be served directly from their sources, or replicated to a central location, like a lake.

3. Metadata is mapped to relational views. No data is replicated.

8. Infrastructure can easily scale out in a cluster.

7. Products can be access via SQL, MDX or exposed as an API. No coding is required.

Data Mesh Powered by Data Virtualization

- SQL
- REST
- GraphQL
- OData
- MDX

**Common Domain**
- Customer
- Event
- Location

**Event Management**
- Product
- Event
- Location

**Human Resources**
- Employee

**Operational**
- SaaS APIs

**Data Lakes**
- EDW

**Files**
Resources

Coming Up!
Dataversity Demo Days
Data Catalogs
April 19, 2023

Dataversity Demo Days
Data Fabric
September 20, 2023

Try Denodo with the 30-day free trial in the cloud marketplaces

**CHOICE**
Under your cloud account

**SUPPORT**
Community forum AND remote sales engineer

**OPPORTUNITY**
30 minutes free consultation with Denodo Cloud specialist

denodo.com/free-trials