BUSINESS ᵘⁿINTELLIGENCE

CUTTING DATA FABRIC AND DATA MESH TO MEASURE

30 March 2022

Data Warehousing and BI Summit 2022

Dr Barry Devlin
Founder & Principal
9sight Consulting

2106

1

DR. BARRY DEVLIN

Founder and Principal
9sight Consulting, www.9sight.com

Dr. Barry Devlin is a founder of the data warehousing industry and among the foremost authorities worldwide on business intelligence (BI) and beyond. He is a widely respected consultant, lecturer and author of the seminal "Data Warehouse—from Architecture to Implementation". His second book, **"Business unIntelligence—Insight and Innovation Beyond Analytics and Big Data"** *(bit.ly/BunI-TP2) was published in October 2013.*

Barry has 30+ years of experience in the IT industry, previously with IBM, as an architect, consultant, manager and software evangelist.

As founder and principal of 9sight Consulting (www.9sight.com), Barry provides strategic consulting and thought-leadership to buyers and vendors of BI solutions. He is continuously developing new architectural models for all aspects of decision-making and action-taking support, including new thinking on the use and consequences of artificial intelligence.

Based in Cornwall, UK, Barry's knowledge and expertise are in demand both locally and internationally.

Email:
barry@9sight.com

Twitter:
@BarryDevlin

2

2

1

Data Fabric
A distributed Data Management platform whose objective is to combine various types of data storage, access, preparation, analytics, and security tools in a fully compliant manner to support seamless Data Management...

Data Mesh
A domain-driven analytical data architecture where data is treated as a product and owned by teams that most intimately know and consume the data, applying the principles of modern software engineering and the learnings from building object, internet-scale ... true potential of ...

But what exactly do they mean?

3

Digital Systems Information Architecture (DISA)

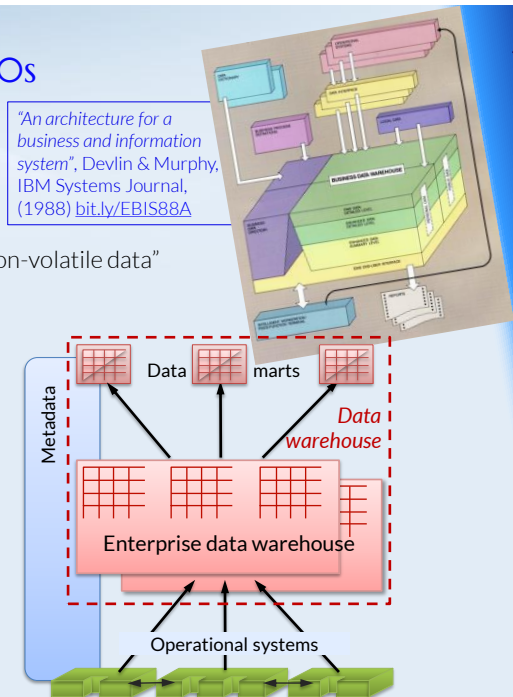"Sometimes it's good to get a different perspective."

A soupçon of architecture as a basis for comparison

4

Data architecture since the mid-'80s

- Design point (1986): internally sourced data
  - Separate operational systems layer to run the business processes
  - Data warehouse for business decision making
    - "Subject-oriented, integrated, time-variant, non-volatile data"
    - *Popularised by Inmon, 1992-'95*

- Two layers within the Data Warehouse
  - EDW – Reconciled data: 3NF → Data Vault
  - Data marts – Users' needs: often star schema

- Characteristics
  - Vertical and horizontal segmentation
  - Hard data only
  - Unidirectional data flow
  - Separate metadata

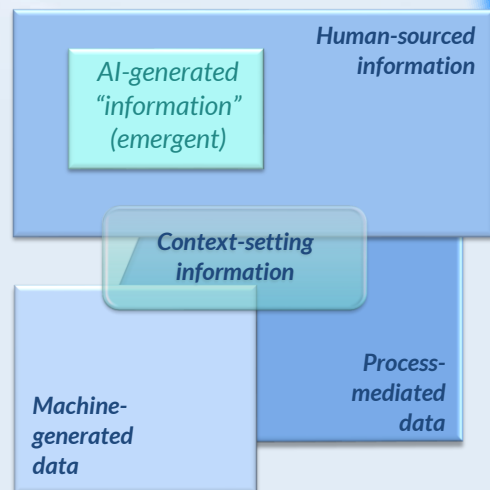- Dive deeper: B. Devlin, "Data Warehouse – from Architecture to Implementation", (1997)

"An architecture for a business and information system", Devlin & Murphy, IBM Systems Journal, (1988) bit.ly/EBIS88A

Metadata

Data marts

Data warehouse

Enterprise data warehouse

Operational systems

5

Data and information domains of the Digital Information Systems Architecture (DISA)

- Process-mediated data (PMD)
  - Operational & informational data
  - Via data entry and cleansing

- Machine-generated data (MGD)
  - Machine and sensor output, Internet of Things

- Human-sourced information (HSI)
  - Subjectively interpreted personal experiences
  - From Tweets to Videos

- Context-setting information (CSI)
  - Metadata and more

- AI-generated "information" (AGI)
  - Lacking human understanding, poorly defined, and a Gartner 2022 top strategic technology trend!

Human-sourced information

AI-generated "information" (emergent)

Context-setting information

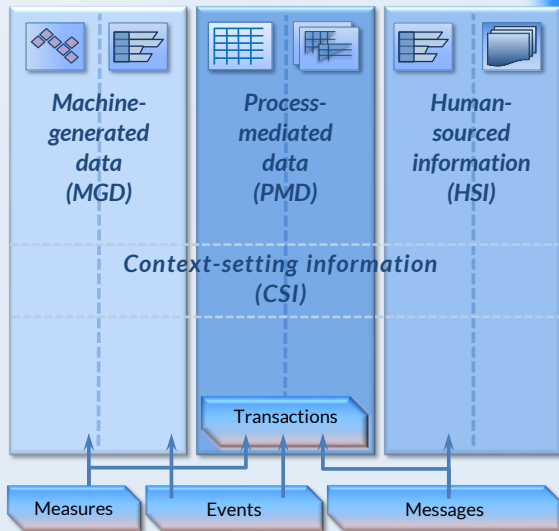*Process-mediated data*
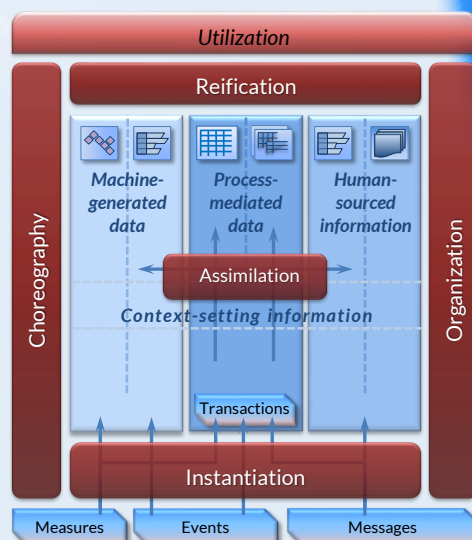
Machine-generated data

6

Introducing Information pillars

- Single architecture for all types of data/information
  - Mix/match technology as needed
    - *Relational, NoSQL, Hadoop, etc.*
  - CSI (metadata+) included

- Data flows as fast as needed and reconciled when necessary
  - Only strictly necessary storage and transformations
  - CSI → pillars rather than silos
  - External data may only be accessed
  - (Contrast layered architecture)

- From sources to stores
  - Events – discrete MGD data points
  - Measures – on-going MGD
  - Messages – HSI
  - Transactions – created from events, measures and messages

Machine-generated data (MGD) — *Process-mediated data (PMD)* — *Human-sourced information (HSI)*

Context-setting information (CSI)

Transactions

Measures | Events | Messages

7

7

The logical (REAL) architecture implements IDEAL concepts

- Processes to create, maintain and use business information
  - **R**ealistic, **E**xtensible, **A**ctionable, **L**abile

  - **Instantiation:** Tools by which raw sources are represented as information instances
    - *File access, ETL, DWA, change capture, streaming...*

  - **Assimilation:** Creation of reconciled & consistent information sets and CSI prior to business use
    - *ETL, ELT and data virtualization*

  - **Reification:** Consistent, cross-pillar information access in real-time via overarching model
    - *Data virtualization*

  - **Utilization:** direct or application use of data/info

  - **Choreography:** manage processes
  - **Organization:** administer the entire environment

- Used mainly by IT

Utilization

Reification

Machine-generated data — *Process-mediated data* — *Human-sourced information*

Choreography

Assimilation

Context-setting information

Organization

Transactions

Instantiation
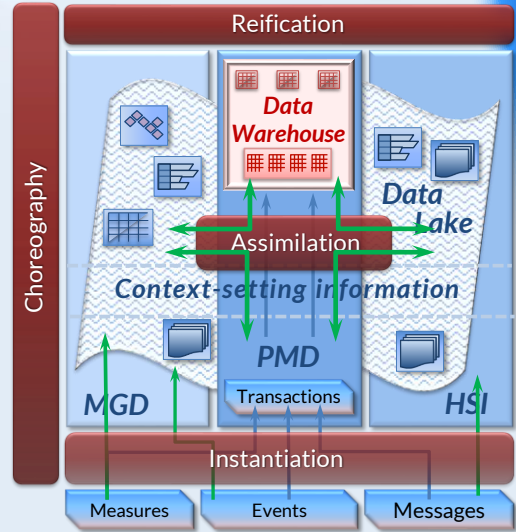
Measures | Events | Messages

8

8

4

DATA WAREHOUSE AND DATA LAKE TODAY

- Data Warehouse – manage ongoing business, based on operations
  - Clean, correct, reconciled, perhaps delayed
  - Governance is vital, architected approach

- Data Lake – predict / influence the future, from digital activities
  - Flexible, innovative, often near real-time
  - Limited data management

- Together: Logical Data Warehouse

Challenges

- Operational / informational overlap
  - Complex data and model flows
  - Data links or extensive copies

- Integration
  - Predictive / prescriptive analytics demands integration of all systems

- Design & maintenance is complex & inflexible

9

9

Data Fabric and Data Mesh

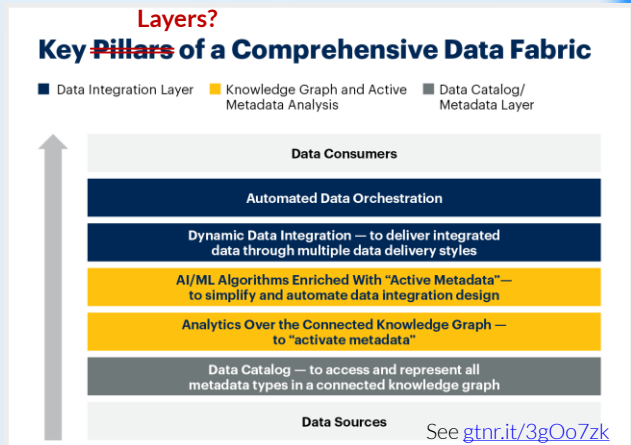# Meaningful definitions & architected pictures

10

10

5

DATA FABRIC – AS DESCRIBED BY PROPONENTS

Gartner – Design concept:

- Integrated layer (fabric) of data and connecting processes
- Continuous analytics over existing, discoverable, and inferenced metadata
- Supporting design, deployment & utilization of integrated & reusable data across all environments, incl. hybrid and multi-cloud platforms
- Leverages both *human* and *machine capabilities* to access data in place or support its consolidation where appropriate
- No. 3 in Gartner Top 10 Data & Analytic Trends 2021

- Not truly "new"
  – Described (in part) since 2013 by Forrester and others

Layers?

Key ~~Pillars~~ of a Comprehensive Data Fabric

■ Data Integration Layer ■ Knowledge Graph and Active Metadata Analysis ■ Data Catalog/ Metadata Layer

Data Consumers

Automated Data Orchestration

Dynamic Data Integration — to deliver integrated data through multiple data delivery styles

AI/ML Algorithms Enriched With "Active Metadata"— to simplify and automate data integration design

Analytics Over the Connected Knowledge Graph — to "activate metadata"

Data Catalog — to access and represent all metadata types in a connected knowledge graph
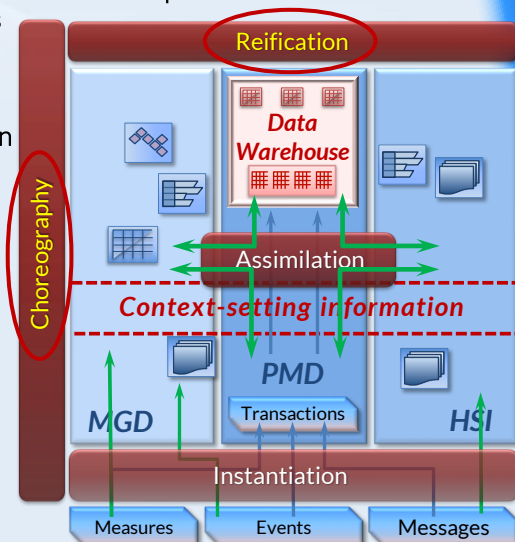
Data Sources

See gtnr.it/3gOo7zk

11

REAL DATA FABRIC PATTERN – (SIMPLY) AN EVOLUTION OF LOGICAL DW?

- Retain and access all current sources and copies of data as required
  – Operational systems and external data sources
  – Data warehouse and data lake

- Data virtualisation to access all data

- Automated, integrated information preparation

- Strong focus on context-setting information
  – Central to data management / governance
  – Key to widespread self-service usage
  – Pervasive – business, technical, operational
  – Actively maintained and used

Challenge: requires significant new / improved technologies, e.g.
  – Choreographed workflows
  – Semantic knowledge graphs
  – Active metadata management
  – Embedded machine learning

Reification

Choreography

Data Warehouse

Assimilation

Context-setting information

PMD

Transactions

MGD   *HSI*

Instantiation

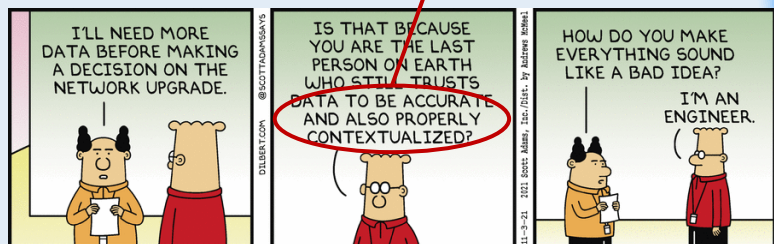Measures   Events   Messages

12

So, do you need a data fabric?

- Number of data warehouses and marts
- Number of data lake instances
- Other data sources (e.g. operational systems) your users access
- Number of information preparation processes
- Rate of change in information preparation processes / marts

- The higher your score the more you need a data fabric, in order to:
  - Limit maintenance costs for data models and information preparation through automation
  - Empower businesspeople to find and easily use the right information

13

Data mesh – as described by proponents (I)

- Drivers
  - Unlock analytical data at scale – autonomy and democratisation
  - Address common failure modes of centralized data lakes & warehouses
  - *"Embrace the reality of [the] ever-present, ubiquitous, and distributed nature of data"*

- Strong focus on organisational concepts and concerns
  - New governance approach
  - Local vs central control

- Principles
  - Learn from DDD, microservices & DevOps from modern software engineering
  - Domain-oriented decentralized data ownership and architecture
  - Data as a product
  - Self-serve data infrastructure as a platform
  - Federated computational governance
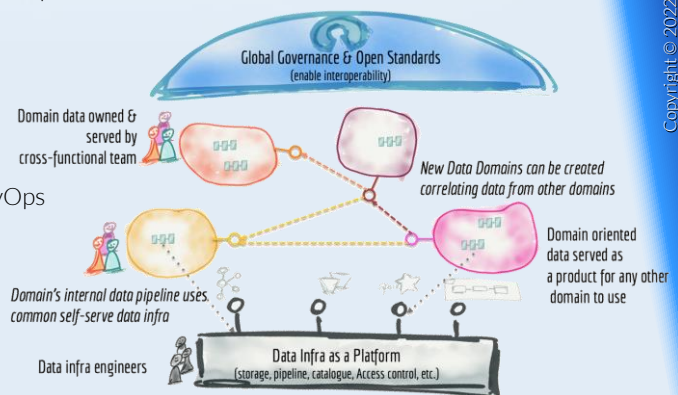 - See: thght.works/3q5QbSR

Image: Zhamak Dehghani, 2019, bit.ly/3dtLrl3

14

Data mesh – as described by proponents (2)

Technical / architectural concepts

- Data product quantum [microservices thinking]:
  Code + Data & Metadata + Infrastructure

- Multi-plane data platform, including:
  – Data infrastructure provisioning plane
  – Data product developer experience plane
  – Data mesh supervision plane

- Computational policies embedded in the mesh
  – Federation / virtualization of data
  – Pull data on demand

- Conceptual complexity
  – Dehghani tweet, 13 August 2021
  – *"There are no warehouses in a mesh. There are autonomous data product quantums that provide multimodal access to domain data for analytical workloads - connected together in a graph - each both transforming and serving/controlling immutable bitemporal data."*
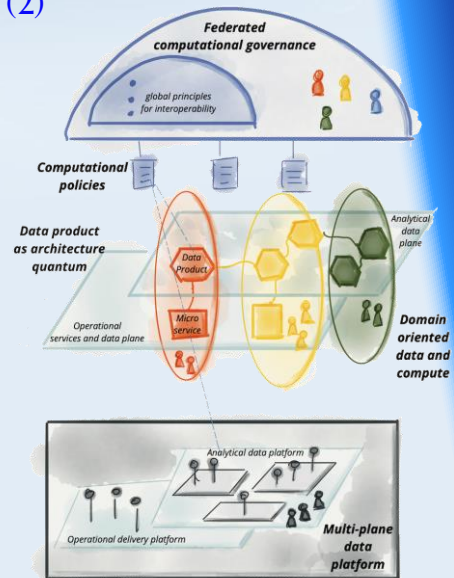

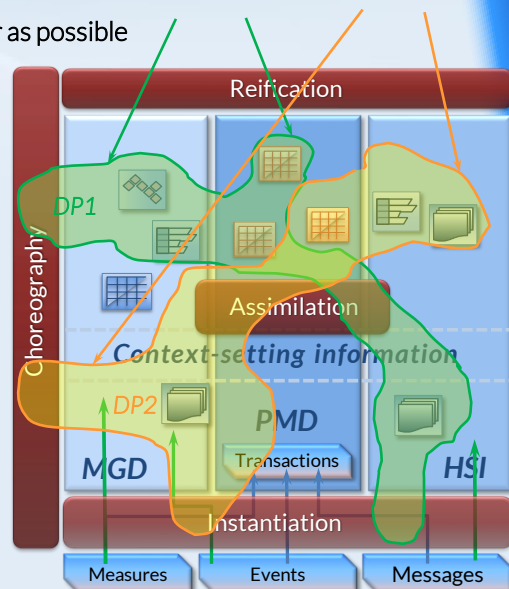*Image: Zhamak Dehghani, 2020, bit.ly/2S7i4xd*

15

Data mesh: Service oriented architecture for data?

- Dispenses with centralised copies of data as far as possible
  – (Although some data mesh approaches keep data warehouse / lake constructs)
  – Instantiation & assimilation in data products – multiple instances across enterprise

- Choreography and reification are key
  – Microservices structure make data products (DPs) available for "pull" usage

- Requires pervasive / active CSI
  – Central to data management / governance
  – Key to microservices

Challenge: requires extensive changes to design thinking and new technologies, e.g.
  – Micro-services operational view
  – Reconciliation across domains (silos?)
  – Gaps in open-source and commercial tooling

16

So, do you need a data mesh?

- Quantity of data sources
- Size of your data team
- Number of data domains
- Data engineering bottlenecks
  - How frequently is the data engineering team a bottleneck to the implementation of new data products on a scale of 1 to 10?
- Data governance
  - How much of a priority is data governance for your organization on a scale of 1 to 10?

- The higher your score the more you need a data mesh!
  - >30 = data mesh sweet spot
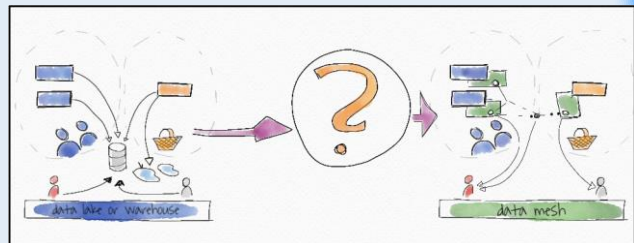 - See *"What is a Data Mesh — and How Not to Mesh it Up"*, Moses, July 2020 , bit.ly/2ZJ4h3B

Image: Sven Balnojan, 2019, bit.ly/3GFIivg

17

17

Are you *really* ready to build a data fabric or mesh?

Data Fabric

- How far toward Logical Data Warehouse (LDW) are you?
- Is your data lake a data swamp or a start toward an LDW?
- A good LDW is already the foundation of Data Fabric

- How good is your metadata / context-setting information (CSI)?
- CSI foundation is vital for Data Fabric

- Data Fabric → integration automation

Data Mesh

- How good are you in microservices, agile dev., and domain-driven design approaches?
- How willing is your data team to adopt and adapt these approaches?
- Data Mesh demands very different organisation and governance

- How important is consolidated / reconciled data to your reporting / analytics?
- Enterprise DW / Vault not formally included

- Data Mesh → significant redesign for agility

18

18

9

CONCLUSIONS

1. **Understanding data fabric and mesh needs common base**
   - ✓ Digital Information Systems Architecture (DISA)
   - ✓ Allows comparison between them
   - ✓ Enables migration planning from warehouse or lake

2. **Data fabric is an evolution of Logical Data Warehouse**
   - ✓ Automation and augmentation via active CSI (metadata)
   - ✓ Knowledge graph for CSI

3. **Data mesh applies domain-driven design to data**
   - ✓ Requires large shift in organisation and methods
   - ✓ Driven by governance thoughts – eliminate centralisation
   - ✓ Microservices approach in technology

19

19

BUSINESS *un* INTELLIGENCE

THANK YOU

QUESTIONS?

Dr Barry Devlin
Founder & Principal
9sight Consulting

2106

20