# Expert Panel:

## Using a Semantic Layer to Democratize the Enterprise Data Lakehouse

Moderated by James Kobielus

Senior Research Director, Data Management, TDWI

Tuesday January 17, 2023

# SPONSOR

# JAMES KOBIELUS

Senior Research Director,
Data Management
TDWI

# Takeaways

- Semantic layers are key lakehouse infrastructure:
  - Unify access to all enterprise data assets within an open lakehouse architecture
  - Present clear understanding of disparate data across lakehouse subject domains
  - Empower lakehouse users with self-service analytics
  - Accelerate data-driven insights through a high-performance lakehouse architecture abetted by contextual data quality rules and policies
- Data democratization hinges on semantic layers that deliver lakehouse simplicity, clarity, usability, and insight acceleration

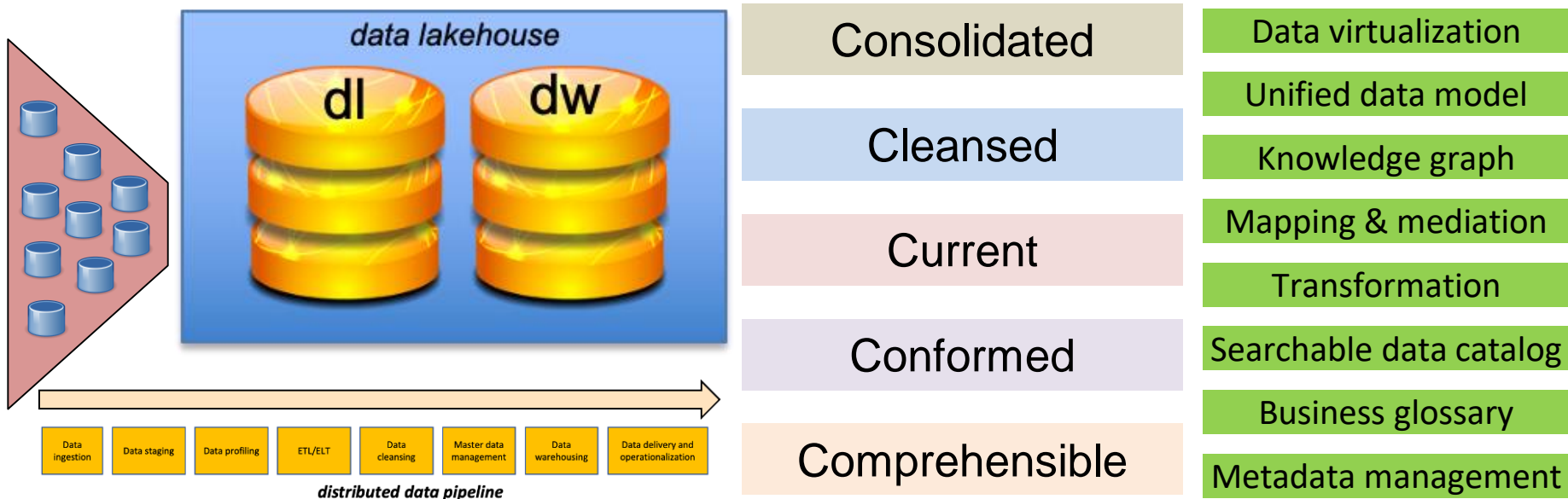# Simplicity: Unifying Access to Enterprise Data

*All users should enjoy unified access to all lakehouse data and analytics assets that they need to be productive.*
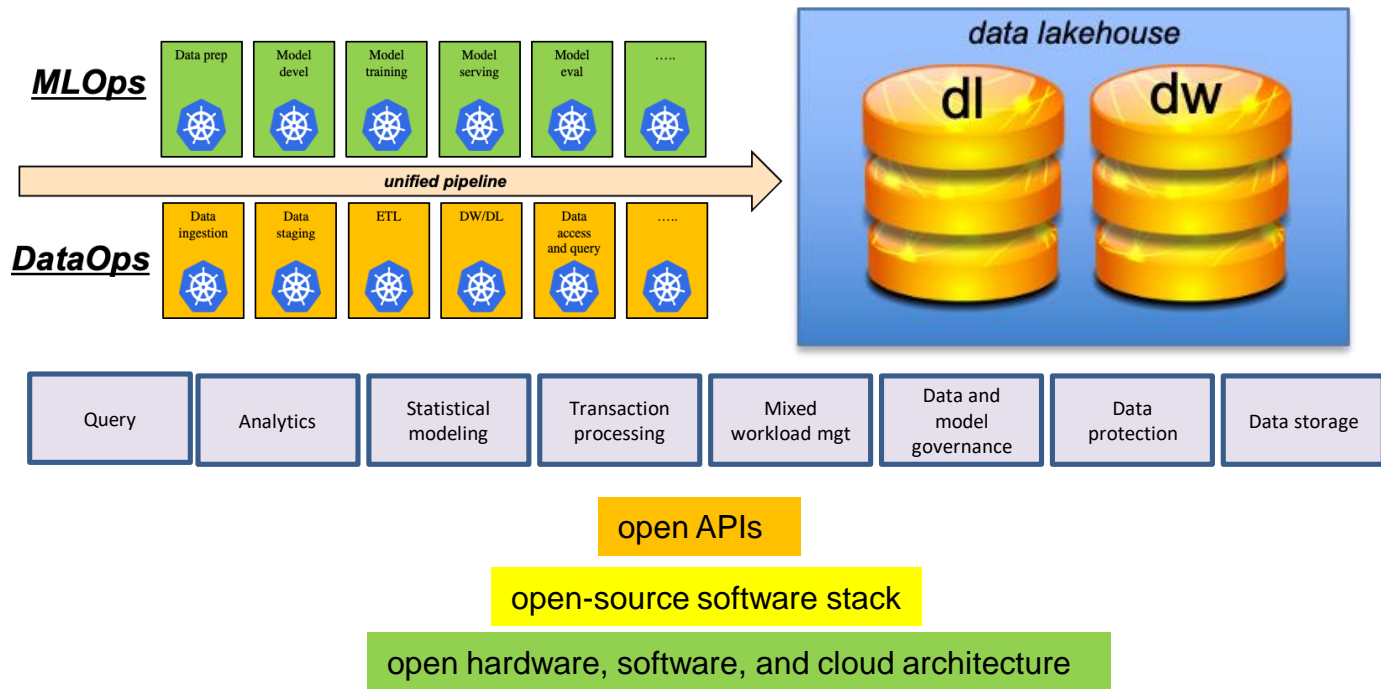


- ✓ Unified view of data and analytic assets

- ✓ Unified DataOps and MLOps pipeline

- ✓ Unified experience for accessing, analyzing, manipulating, and processing it all

- ✓ Unified management, monitoring, and optimization tools

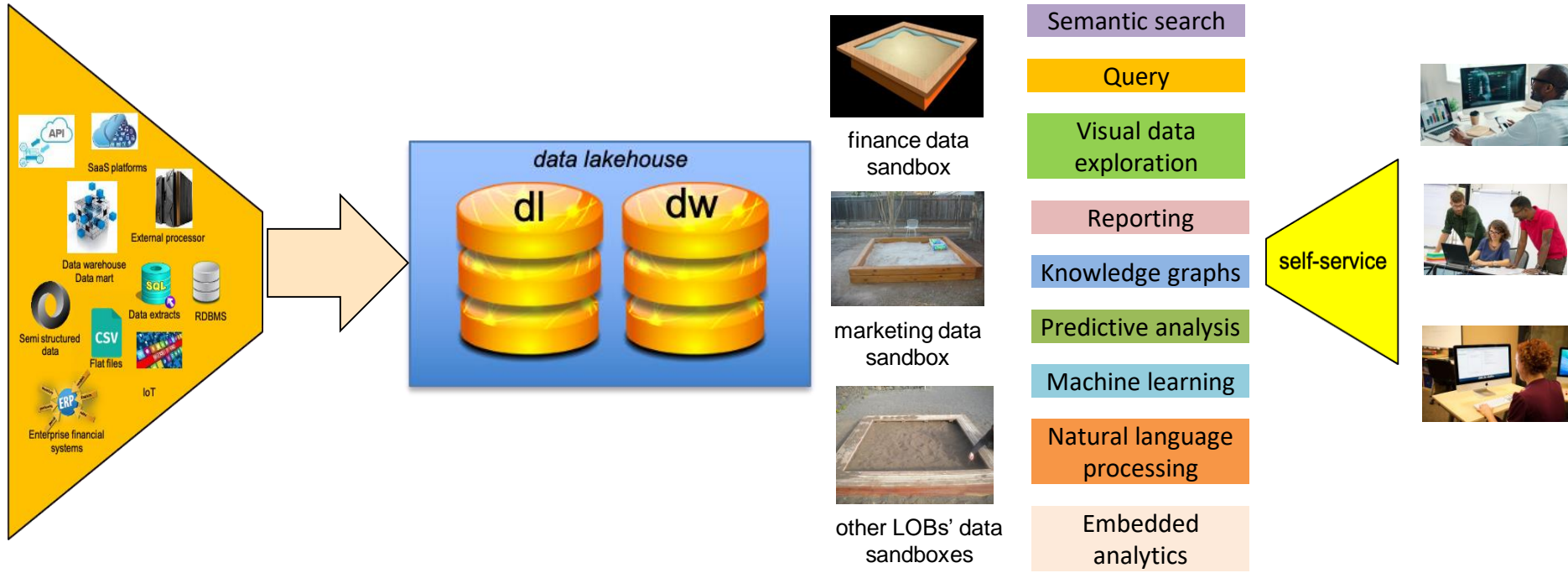- ✓ Unified abstraction for programmability as cloud-native microservices

# Clarity: Presenting Transparent Cross-Domain Semantics

*All users should be presented with a clear, consistent explanation of lakehouse data's meanings, connections, and contexts.*



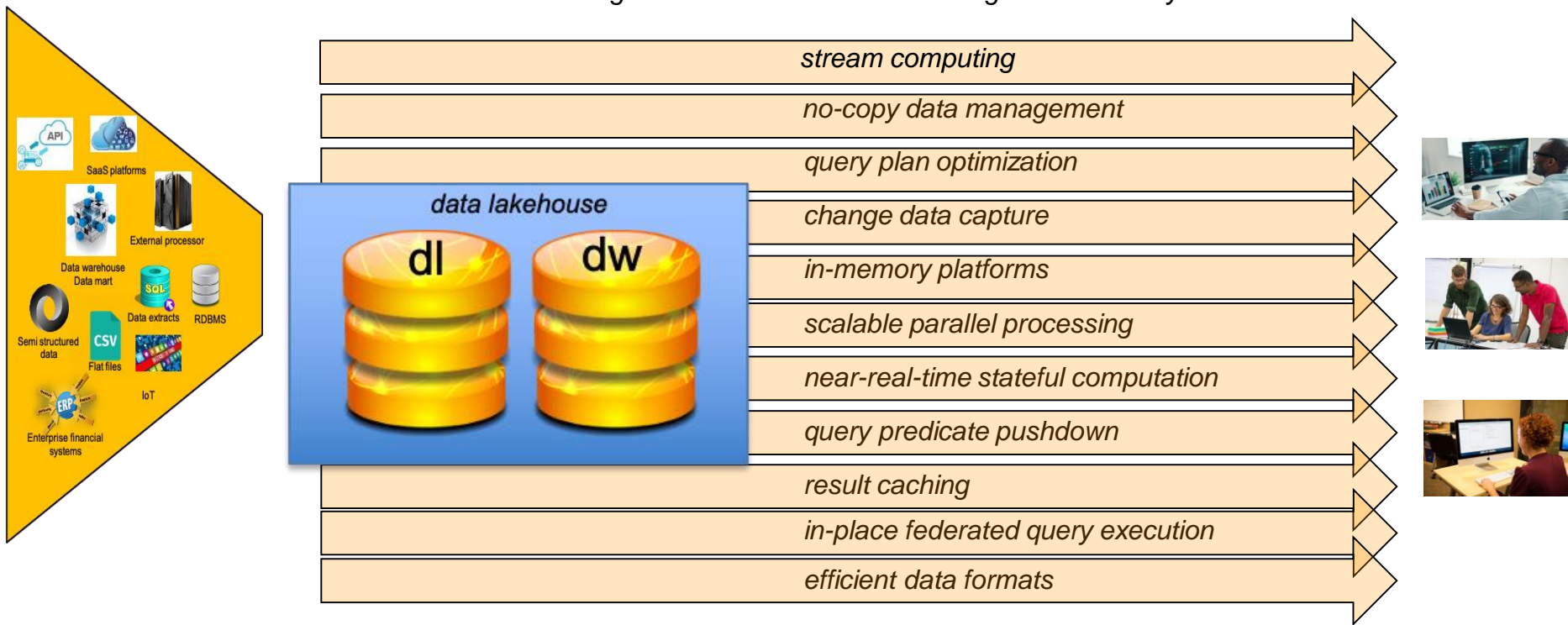| Consolidated | Data virtualization |
| --- | --- |
| Cleansed | Unified data model |
| | Knowledge graph |
| Current | Mapping & mediation |
| | Transformation |
| Conformed | Searchable data catalog |
| | Business glossary |
| Comprehensible | Metadata management |

# Usability: Empowering Users with Self-Service Analytics

*All users should have the right tools they need to explore, analyze, model, and otherwise consume lakehouse data assets.*



data lakehouse

dl    dw

finance data sandbox

marketing data sandbox

other LOBs' data sandboxes

Semantic search

Query

Visual data exploration

Reporting

Knowledge graphs

Predictive analysis

Machine learning

Natural language processing

Embedded analytics

self-service

# Acceleration: Delivering Insights at the Speed of Thought

*All users should consume data-driven insights in real time delivered through a low-latency lakehouse infrastructure.*



- stream computing
- no-copy data management
- query plan optimization
- change data capture
- in-memory platforms
- scalable parallel processing
- near-real-time stateful computation
- query predicate pushdown
- result caching
- in-place federated query execution
- efficient data formats

**tdwi** TRANSFORMING DATA WITH **INTELLIGENCE**™

THANK YOU!!!!!

James Kobielus
Senior Director of Research for Data Management
TDWI (www.tdwi.org)
jkobielus@tdwi.org
@jameskobielus

# Keon Shahab

Solutions Architect
Databricks

# Databricks

## The Lakehouse Company

Inventor and pioneer of the **data lakehouse**

Gartner recognized leader in both
- Database Management Systems
- Data Science and Machine Learning Platforms

Creator of highly successful OSS data projects

Raised over $3B in investment

4500+ employees across the globe

## Global adoption

Over 7000 customers, from F500 to unicorns
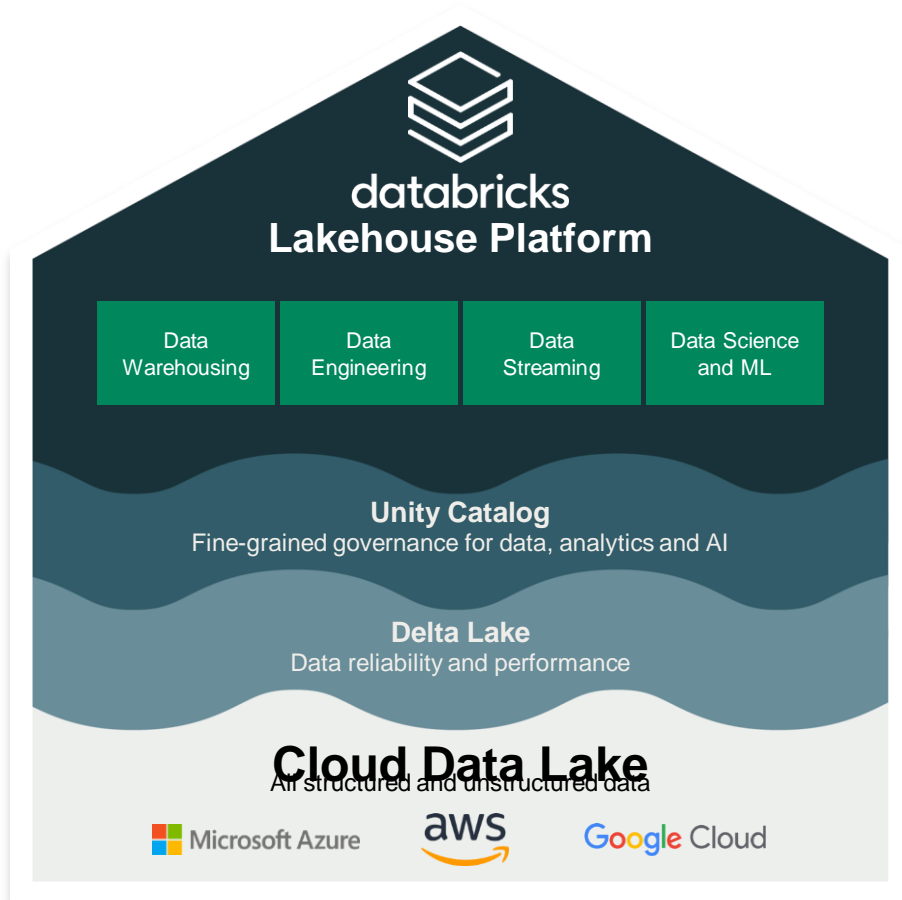
# Databricks Lakehouse Platform

**Simple**

Unify your data warehousing and AI use cases on a single platform

**Multicloud**

One consistent data platform across clouds

**Open**

Built on open source and open standards

# Navin Sharma

VP, Product
Stardog

# The data Set - British Royal Family

| Person ID | Full Name | Gender |
|---|---|---|
| 0 | Queen Elizabeth II | F |
| 1 | King Charles III | M |
| 2 | Diana, Pricess of Wales | F |
| 3 | Anne, Princess Royal | F |
| 4 | Captain Mark Phillips | M |
| 5 | Andrew, Duke of York | M |
| 6 | Edward, Earl of Wessex | M |
| 7 | Sophie, Countess of Wessex | F |
| 8 | William, Prince of Wales | M |
| 9 | Catherine, Princess of Wales | F |
| 10 | Harry, Duke of Sussex | M |
| 11 | Meghan, Duchess of Sussex | F |
| 12 | Peter Phillips | M |
| 13 | Autumn Phillips | F |
| 14 | Zara Tindell | F |
| 15 | Mike Tindall | M |
| 16 | Sarah, Ducess of York | F |
| 17 | Lady Lousie Windsor | F |
| 18 | James, Viscount Severn | M |
| 19 | Princess Beatrice | F |
| 20 | Princess Eugenie | F |
| 21 | Edoardo Mapelli Mozzi | M |
| 22 | Jack Brooksbank | M |
| 23 | Sienna | F |
| 24 | August | F |
| 25 | Archie | M |
| 26 | Lilibet | F |
| 27 | Prince George of Wales | M |
| 28 | Princess Charolette of Wales | F |
| 29 | Prince Louis of Wales | M |
| 30 | Savannah | F |

| Parent ID | Child ID |
|---|---|
| 0 | 1 |
| 0 | 3 |
| 0 | 5 |
| 0 | 6 |
| 1 | 8 |
| 1 | 10 |
| 2 | 8 |
| 2 | 10 |
| 8 | 27 |
| 8 | 28 |
| 8 | 29 |
| 9 | 27 |
| 9 | 28 |
| 9 | 29 |
| 10 | 25 |
| 10 | 26 |
| 11 | 25 |
| 11 | 26 |
| 3 | 12 |
| 3 | 14 |
| 4 | 12 |
| 4 | 14 |
| 12 | 30 |
| 12 | 31 |
| 13 | 30 |
| 13 | 31 |
| 14 | 32 |
| 14 | 33 |
| 14 | 34 |
| 15 | 32 |
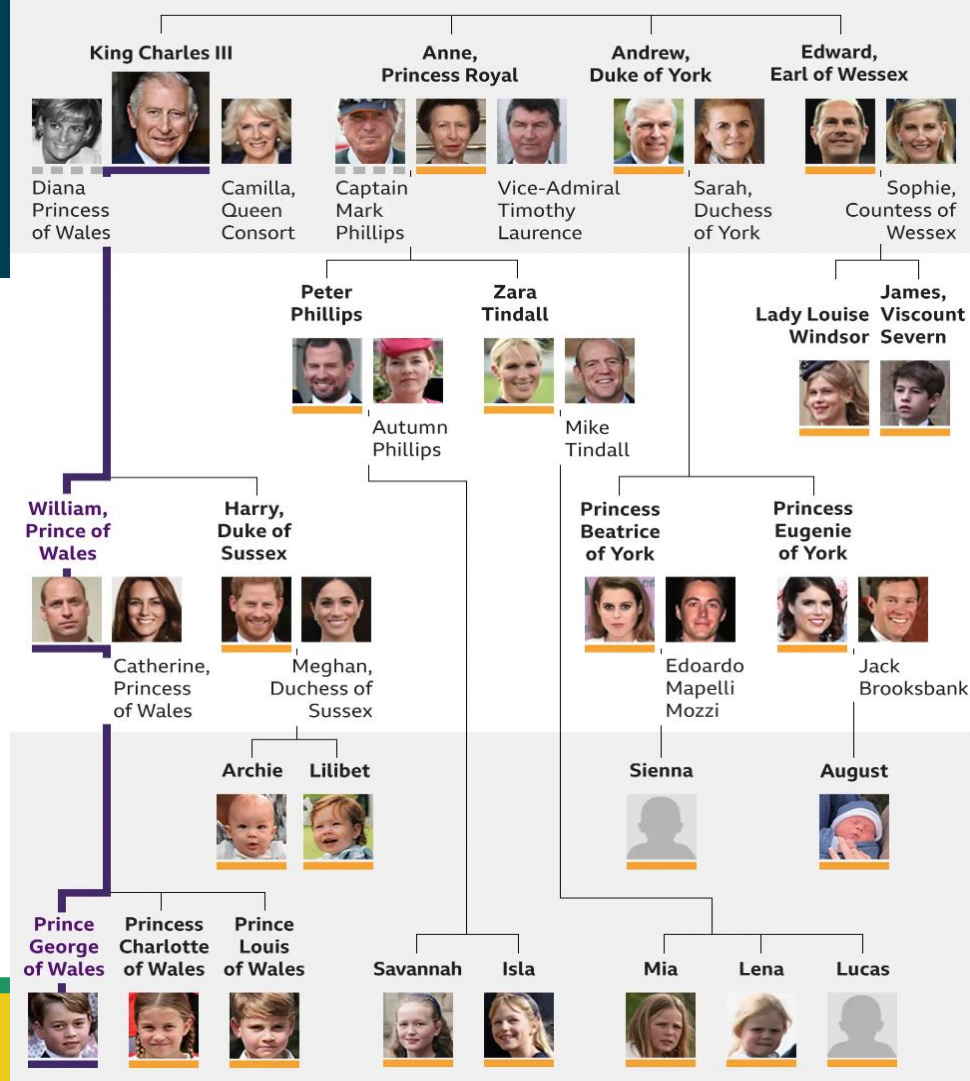
# Not your grandfather's Semantic Layer

**Notice that the data has only members and parent-children information. Semantic Terms like Brothers, Sisters, Siblings, Aunts, Uncles etc do not appear anywhere in the data.**

Examples of **Statements about the data -**
  "Queen Elizabeth *has_Child* Charles"
  "Harry *has_Child* Archie"

Examples of **Semantic model, expressed with annotated business rules annotated without changes to the underlying source data or source data model.**
These are statements about the Domain (irrespective of the data)
  "If person has child then person is parent"
  "A grandparent is the parent of a parent."
  "The male sibling of a parent is the Uncle of the child"

Examples of **Statements that enrich the data with Semantics**
  "Queen Elizabeth is the **Grandmother** of Prince William"
  "Harry is the **Uncle** of Charlotte"

**Stardog takes statements from Subject Matter Experts and expresses them in a way that easily manipulates and expresses data into knowledge ie Semantically Enriched Data.**

# Find the nephews of any aunts

| Members (1) | | hasParent | → | Members (2) |
|---|---|---|---|---|
| ⊕ Members (2) | | hasSibling | → | Members (3) |
| ⊕ Members (3) | Gender | = | F | |
| ⊕ Members (1) | Gender | = | M | |

William, Prince of...

James, Viscount...

Harry, Duke of Sussex
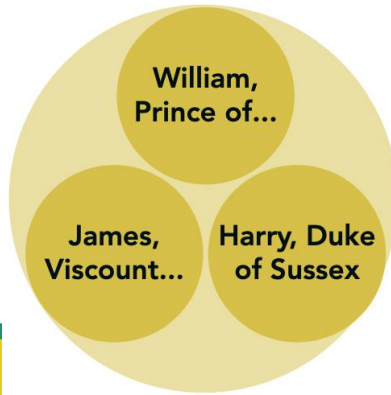
# What is an Enterprise Knowledge Graph?

Enables a flexible, semantic data layer for answering complex queries across a diverse but connected enterprise data landscape by:

- Harmonizing data, metadata & rules based on business meaning

- Limiting data sprawl with federated data access to heterogeneous sources

- Enabling citizen data users to self-serve in context.
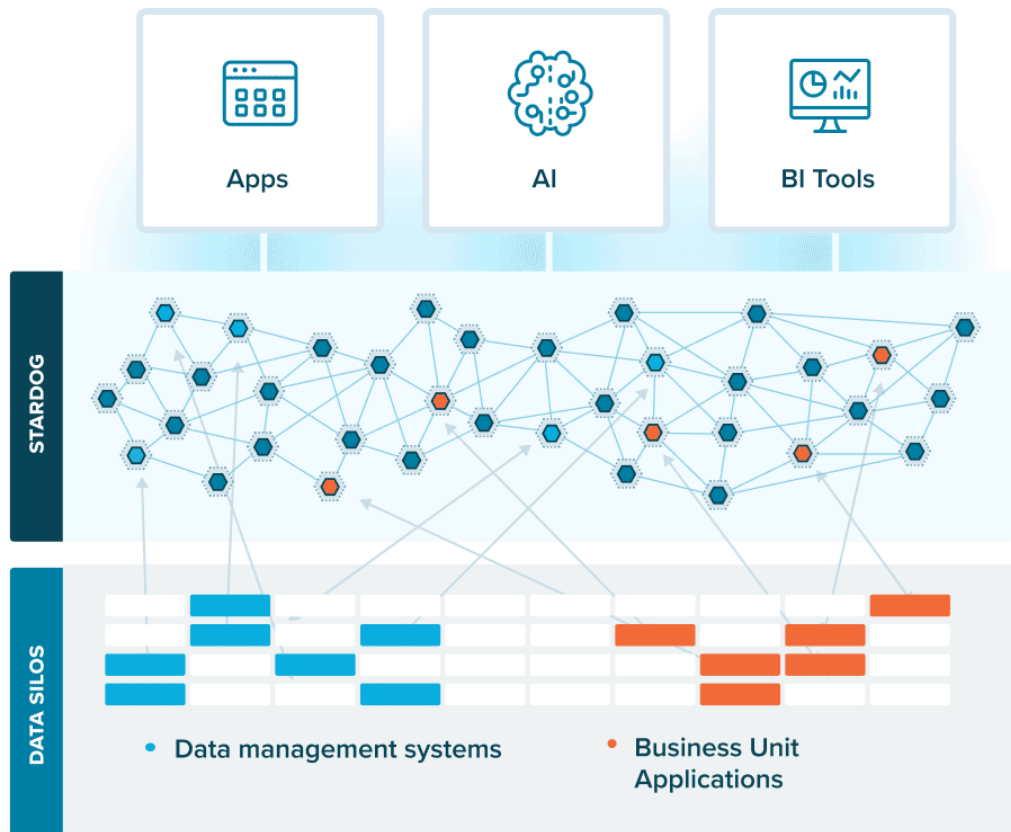
# Semantic Lakehouse Reference Architecture
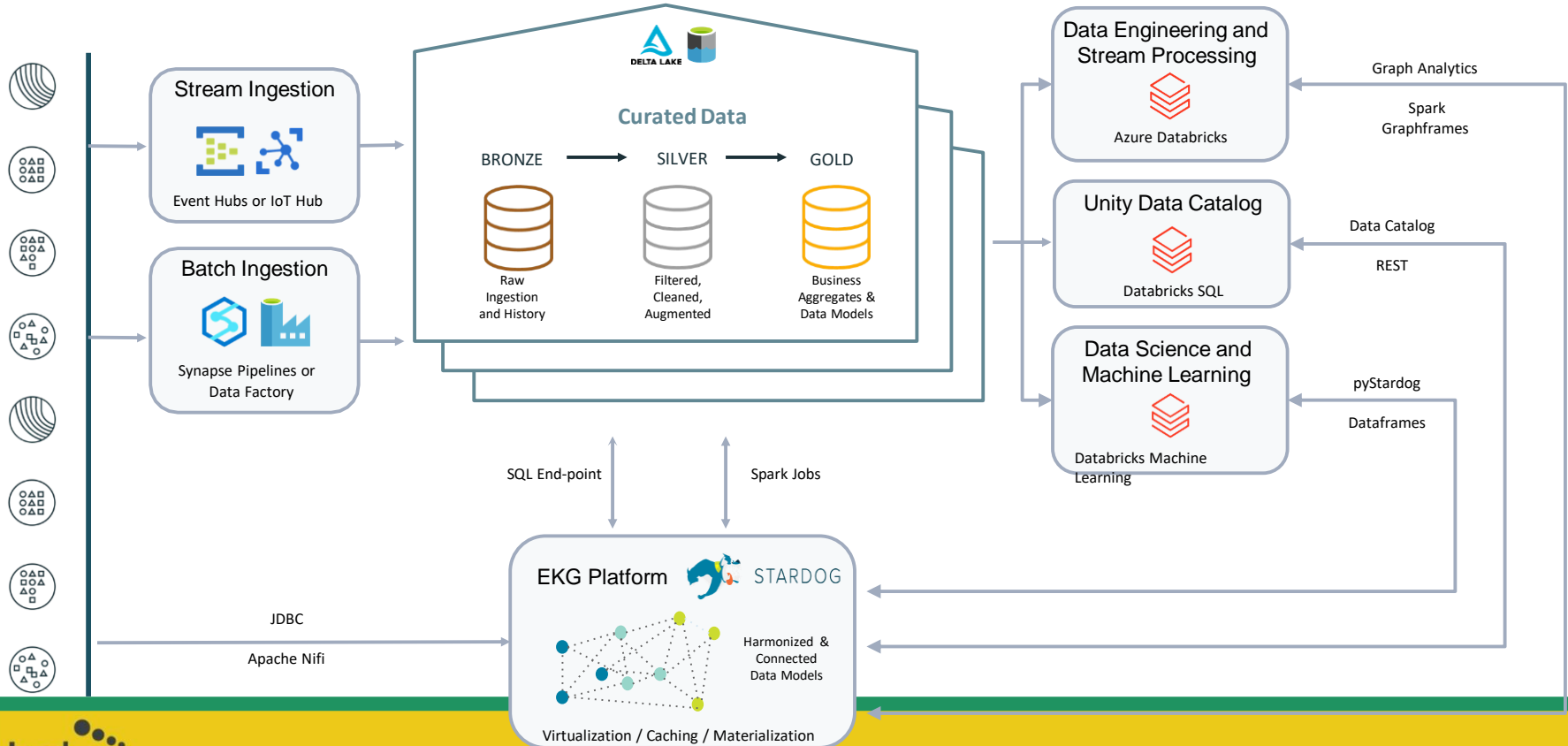
# Enabling use-cases across industries with connected data needs, like, but not limited to

### HCLS

Drug discovery

Pre-clinical R&D

Molecule-to-market

### Manufacturing

Digital twin & thread

Product 360

Bill of Materials

### Financial Services

Customer 360

Financial crimes

Operational Risk

STARDOG

Supercharge your analytics & AI directly within your applications
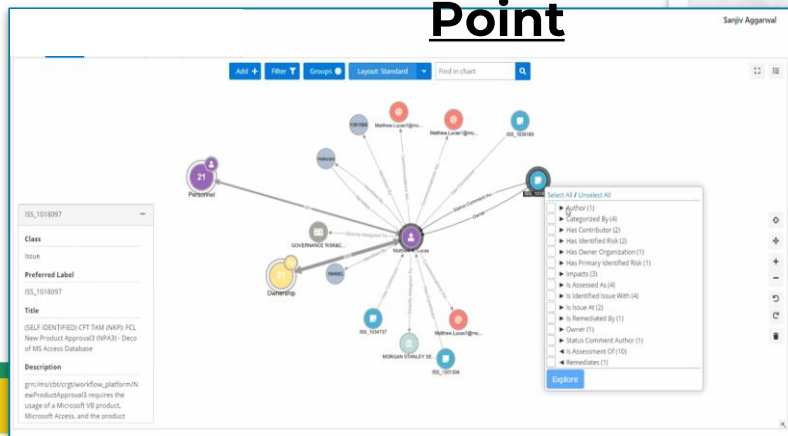
**Python/ Spark Data Frames**

**SQL End- Point**

**GraphQL/RE ST APIs**

# Get Started for Free on your EKG journey

https://www.stardog.com/get-started/
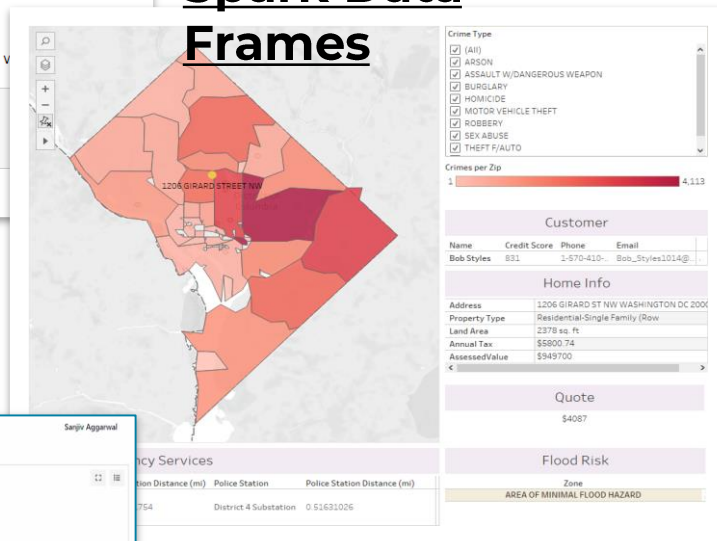
RECOMMENDED

## Stardog Free

Everything you need to learn and build knowledge graphs as a single developer. No credit card required.

**$0/month**
(Free)

Start Free now!

## Stardog Essentials

A perfect fit for jump-starting your proof of concept and getting to your first knowledge graph.

**$99/month**

Start Essentials now!

## Stardog Enterprise

Our mission-critical, large-scale production environment for your enterprise.

**Let's talk**

Contact us!

# Databricks users? Look for Stardog under Partner Connect

# Panel Discussion

tdwi **TRANSFORMING DATA WITH INTELLIGENCE**™

# What is a semantic layer?

How can a semantic layer accelerate the democratization of your enterprise data lakehouse?

TDWI TRANSFORMING DATA WITH INTELLIGENCE™

How can enterprises leverage a semantic layer to harmonize data models with standard, well-understood concepts and vocabularies?

How can knowledge graphs support flexible viewing of semantic relationships among business concepts across connected data domains in your enterprise lakehouse?

**tdwi** TRANSFORMING DATA WITH INTELLIGENCE™

How can a graph database enable fast, efficient, robust, and ACID-compliant semantic processing of connected data domains?

How can business vocabularies be enhanced through application of contextual data quality rules within a semantic layer?

# QUESTIONS?

# CONTACT INFORMATION

If you have further questions or comments:

James Kobielus, TDWI
jkobielus@tdwi.org

Keon Shahab, Databricks
keon.shahab@databricks.com

Navin Sharma, Stardog
navin.Sharma@stardog.com

tdwi **TRANSFORMING DATA WITH INTELLIGENCE**™

# SPONSOR