INTELLIGENCE OF EVERYTHING

Knowledge Mesh
From Data Silos to Data Fabric
2021
What Is Knowledge Mesh?

#KNOWLEDGE_MESH # DOMAIN #DATAFABRIC

1. INNOVATIVE
   An innovative capability for the data-intensive enterprise, provided by BlackSwan Technologies

2. DOMAIN-DRIVEN
   A domain-driven architecture consisting of augmented data cataloging, integration and management tools

3. FIRST-OF-ITS-KIND
   A first-of-its-kind feature with the ability to maximize the effectiveness of data utilization in knowledge graph applications in any domain

4. LIKE NEVER BEFORE
   A combination of advanced data fabric design with composite AI techniques, enabling enterprises to infer insights, incorporate metadata and enhance decision-making abilities
The Challenge

However, the use of traditional centralised data platforms have led to enterprises encountering a number of key issues:

- Numerous business units have different operational systems, meaning centralising this data leaves entity data outdated, incomplete and unnormalised.
- Excessive time, effort and costs required to integrate new data sources.
- Difficulty in incorporating additional unique identifiers about data to completely understand a customer, client or other entity.

+ data discoverability issues heterogeneity of database models heterogeneity of schemas lack of semantic interoperability lack of uniform accessibility rules data granularity issues data quality issues

..in short: **DATA SILOS!**
Up until now, one natural way of handling 360-degree Entity View use-cases has been to use Enterprise Knowledge Graphs:

- Clean, deduplicated (resolved/reconciled), up-to-date view of all entity data
- Entity- and relationship-centric: closer to real-world & whiteboard conceptualizations
- 360-degree view of relevant data coming from multiple sources, under a single, consistent data model
- Powerful data integration framework
- AI & graphs (LPGs, RDF, graph computing, ML on graphs)
- Getting reliable knowledge out of data
Knowledge Graph Construction – ETL Approach

EXTRACT

TRANSFORM

LOAD

Target Schema

Knowledge Reconciliation (incl. Entity Resolution)
Evolution of Traditional, Centralized Data Platforms

**DATA WAREHOUSE**
- Highly Structured | Highly Transformed

**STATIC DATA LAKE**
- Structured | Semi-Structured | Unstructured
- Unified batch and stream processing
- Cloud-based managed services
- Data pipeline execution engines

**REAL TIME DATA LAKE**
- Structured | Semi-Structured | Unstructured | Streaming real-time
- Cloud-based managed services
- Data pipeline execution engines

- Business User
  - External Data
  - Operational Data
- Data Marts
- Data Warehouse
- ETL

- Data Scientist
  - Data preparation & validation
  - Machine Learning
  - Data Marts
  - ETL

- Data Scientist
  - Data preparation & validation
  - Machine Learning
  - Operational Data
  - ETL

- Business User
  - Data Marts
  - ETL

High Volume data | High Variety data | High Velocity data
Today’s Preferred Go-To Data Strategy

#DATA_LAKE #DATA_INTEGRATION

Superior to Data Warehouse

+ Easier to manage and exploit data regardless of format (structured or unstructured), quality or location, and do so at scale

+ The added ability to store raw data, and operate without determining the schema or structure in advance

+ Greater flexibility as any data retrieved or analysed does not affect data stored, so it can be used for other purposes

+ Keep data for longer, improving chances of finding actionable business insights which might get overlooked

+ Less technical debt and achievable at a lower cost
Challenges With Centralization: The Data Swamp

A Data Swamp is:

❌ Created because Data Lakes are hard to manage, and can quickly accumulate vast amounts of uncontrolled data, which is unusable

❌ “Without descriptive metadata and a mechanism to maintain it, the data lake turns into a data swamp. And without metadata, every subsequent use of data means analysts start from scratch” - Gartner

❌ Created due to a lack of process, standards and governance, ultimately making it harder to find data, harder to use data, and therefore result in data consumed out of context

❌ One of the many issues that come as a result of centralization
Challenges With Centralization: ETL Pipelines

#DATA_LAKE #DATA_WAREHOUSE

ETL-based Knowledge Graph construction has a number of same issues as traditional, centralized data integration platforms such as data lakes or data warehouses...

- a one-off process instead of continuous evolution
- masses of ETL/ELT code that is only understandable and maintainable by the centrally positioned data engineering team
- the meaning of data lost on the way
- no notion of domain ownership from end-to-end resulting in collaboration bottlenecks and inefficient teams
- heavily centralized architecture and process
- data is moved to new storage spaces giving rise to new data silos
The Future Data Platform

Product: Knowledge Mesh

Architecture: Data Mesh

Concept: Data Fabric
Data Fabric and Data Mesh

- The Data Fabric is a design concept, that requires different data management technologies.
- Aim is to ease complexity in managing different kinds of data and the databases that the data resides in.
- Allows access and sharing in distributed environment.
- Uses a single, secured, controlled data management framework.
- The Data Mesh turns the Data Fabric approach into an architecture, with an added focus on metadata and the use of machine learning.
Knowledge Mesh is BlackSwan's productization of Data Fabric. It goes a step further than any other decentralised architecture.

- Decentralized, cloud-native
- Uniform data access layer
- Data virtualization
- Domain-driven organisation of knowledge assets
- Low-code data integration and analysis
- Data integration and management tools
- Metadata-driven
- Automated knowledge graph application building

**BENEFITS**

- Infer insights
- Incorporate metadata into analysis
- Maximize data utilization in knowledge graph applications in any domain
- Enhance decision-making
Rich Metadata Layer

#METADATA #DATA_CONTEXT #LOW-CODE

POWER OF METADATA

- resource (in particular data) cataloging for increased data discoverability

- low-code (metadata-driven) modeling, tracing and reuse of information assets and flows, e.g.:
  - what sources are mapped to domain X?
  - which business terms are referenced in application Y?

- standard metadata formats and models for increased interoperability between human and machine agents

- declarative specification of interfaces and access control

- teams supported by metadata-driven (no-code) tools, can collaborate around domains
Uniform Data Access Layer & Data Virtualization

SMART DATA FETCHERS
- GraphQL-based universal data serving and querying mechanism
- uniform security and access control
- uniform data description and data query language over any source
- easy deployment on any cloud infrastructure
- semi-automated connectors to 10s of different source types
- natural fit for data virtualization scenarios (app queries are reformulated in real-time into collections of GraphQL queries)
Domain-Oriented Knowledge Organization

## Domain 1

e.g.: compliance

- person
- employee
- organization

## Domain 2

e.g.: cyber

- attack
- asset

### DOMAIN CONTENTS

- A catalog of associated data sources
- Semantic vocabulary(-ies) (data models) covering all key domain terms (entities / relationships / attributes)
- A set of declarative data mappings from the sources to the domain vocabularies
- Dedicated, configurable data science knowledge extraction and reconciliation algorithms pre-trained for that specific domain
Quality Knowledge from Siloed Data Pieces

#DATA_FABRIC #METADATA #KNOWLEDGE_RECONCILIATION

Knowledge (Graph) = Raw data + All we know about it

- semantic metadata
- provenance metadata
- data quality metadata
- source metadata
- reconciliation metadata
- data usage metadata
- ...
Data Fabric: the Story of Two Knowledge Graphs

A graph of relevant domain entities, of direct interest to the user (e.g.: “Jane Doe”), described in terms of the KG schema, with accompanying metadata.

A graph of all resources comprising shared domain knowledge models, data architecture and integration infrastructure (e.g.: “Customer DB”), along with their reach descriptions in terms of metadata models.

The two KGs are intrinsically interconnected, forming a full context or a data fabric for efficiently utilizing all available and relevant information.
FAIR data via Knowledge Mesh

#KNOWLEDGE_RECONCILIATION

<table>
<thead>
<tr>
<th>Findable</th>
<th>Accessible</th>
<th>Interoperable</th>
<th>Reusable</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Data sources cataloged and searchable</td>
<td>1. All sources accessible via the same API interface</td>
<td>1. All metadata represented using standard formats and models</td>
<td>1. Data virtualization instead of data silos</td>
</tr>
<tr>
<td>2. Contents of each data asset described via shared, semantic, business dictionaries</td>
<td>2. Uniform data description and data query language for exposing sources</td>
<td>2. Deployable on any cloud infrastructure</td>
<td>2. Domain semantic models and data mappings reusable across multiple applications</td>
</tr>
</tbody>
</table>
ELEMENT Knowledge Mesh capabilities address many of the challenges of traditional platforms with the following benefits:

- Complete visibility of data consumption and operational costs
- Enforcement over data usage & compliance with policies
- Seamlessly provide market domain context
- Remove requirement for expensive centralization projects
- Bypass data silos to ensure there are no data discrepancies
- Ensure organisational politics do not cause data asset silos
- Integrate new data sources with ease
BlackSwan Technologies’ ELEMENT of Customer 360™ application, challenges the traditional, centralised approach for creating a Single Customer View of a specific entity.

The application leverages ELEMENT Knowledge Mesh capabilities to achieve a true 360-degree view of a customer, client or profile.

- Ability to distinguish or merge an entity’s records and determine the most accurate descriptive values
- Bypasses data silos, unlike expensive centralisation projects
- Intuitive view and analysis of entity networks
- Minimises data discrepancies between producers and users
- Alignment with other Single Customer View profiles such as KYC
- Can be used as an enabler for investigations, assisting companies with regulatory compliance
Customer 360 View for Financial Institution

**#SUCCESS STORY**

Creating a holistic view of tier 1 bank’s clients and activities using a Data Fabric concept through the ELEMENT of Customer 360 application’s Knowledge Mesh capabilities, providing an up-to-date and accurate Single Customer View.

**Targets**
- Regulatory requirement for Single Customer View
- Overhaul data organisation and data quality to enable investigations such as anti-bribery and corruption
- Ability to determine data is up-to-date and accurate, and resolve entities across a multitude of systems
- Reduction of siloed data assets, harmonization of data sharing between teams
- Ability to identify and match entities and groups, align profiles, create a global network of enterprise relationships and assess related risks

**Addressing The Need**
- Deployed ELEMENT Knowledge Mesh capabilities to consolidate information from a multitude of systems and divisions
- Utilized configurable schema of any number of data points and enriched data from internal, external, structured and unstructured information.
- Accurately identified specific entities across a multitude of sources using market-leading Entity Resolution
- Leveraged ELEMENT Knowledge Discovery’s Graph capabilities to conduct network-based investigations into anti-bribery and corruption and reputational risk for state-owned entities

**Value Proposition**
- Ability to seamlessly meet changing regulatory requirements, minimising monetary losses
- Modernised data infrastructure and governance with no need for an expensive centralisation project
- Surface deep relationships using thorough visual representation of consolidated global network
- Enriching existing assets with fresh sources and insights to realise efficiencies in existing processes such as KYC

**TOP 30** international bank with presence in 50 markets

Millions of global private, corporate and institutional served entities

over 50,000 employees

Inorganic growth meant entities interacting with different business units, with their own operational systems, resulting in data relating to the entity being denormalised
Knowledge Mesh Onboarding Journey

From data silos to data fabric in several steps:
sources → vocabularies → insights → runtime
BlackSwan Technologies is reinventing enterprise software through Agile Intelligence for the Enterprise – a fusion of data, AI, and cloud technologies that generates billions of dollars in economic value for renowned global brands.

Contact us to explore further:

Schedule a briefing  Request a demo  Plan a proof-of-concept

visit blackswantechnologies.ai