Exploiting Enterprise Data for Transformational Projects

Introducing Anzo: A Data Discovery and Integration Layer for the Data Fabric

The fastest way to deliver valuable blended data products

Sean Martin, CTO
Cambridge Semantics Inc.
sean@cambridgesemantics.com
We apply semantics and graph to a data fabric – so anyone can find, understand, blend, and deliver enterprise data.

At a Glance:
• Based in Boston
• 100+ employees
• Origins in IBM and Netezza
• Anzo 4.0 GA 2017
• Added enterprise-scale OLAP graph data engine in 2015
“Unprecedented levels of data scale and distribution are making it almost impossible for organizations to effectively exploit their data assets”

Source: How To Use Semantics to Drive the Business Value of Your Data, Gartner Group, Guido De Simoni, 27 Nov. 2018
A modern data discovery and integration platform for your enterprise data fabric.

Anzo lets users find, connect, and blend enterprise data into analytic ready data products.
Data Products Speed Time-to-Answer

example data product:

Retail Banking
Customer

Customer Information
CRM, credit score, demographic, household

Home Loan
Agreement, origination, billing, asset

Auto Lease
agreement, payments, service, car

Credit Card
Balance, terms, payments, transactions

Marketing
Segment, profitability, channels, offers

Project
Contracts, Materials, Permits, Financing

Claim
Policy, Incident, Appraisal, Billing, Payment

Shipment
Consignee, Carrier, Documents, Cargo

Product
Supplier, Inventory, Demand, Promotions

Loan
Agreement, Origination, Borrower, Payments, Asset

Drug
R&D, Clinical Trial, Adverse Events, FDA, EMR, Clinical

Build-Your-Own
Build a new data product to provide analytic-ready dataset
Examples of Anzo in Action


- **Credit Suisse**: Accelerate value from enterprise data lake investments across multiple business units.

- **Johnson & Johnson**: Operationalized text analytics for consumer feedback.

- **Merck**: Modernizing clinical data standards management to accelerate drug development.

- **Bristol-Myers Squibb**: Integrated data layer to take cross-trial analytics questions from weeks to hours.

- **Lilly**: Accelerate drug development through an enterprise data fabric.

- **Mayo Clinic**: Building a platform of clinical and treatment data for oncology patients and their doctors to improve diagnosis and care.

- **M2Gen**: Accelerate scale and growth of M2Gen’s ORIEN Program and Pharma partnerships.

- **BNY Mellon**: Modernizing Customer 360 and Data Center Management through semantics and graph.
<table>
<thead>
<tr>
<th>Anzo Difference:</th>
<th>Graph Data Models &amp; Semantics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simplifies</td>
<td>access to complex blended data to address unanticipated questions</td>
</tr>
<tr>
<td></td>
<td>Quickly profiles, connects and harmonizes data from multiple sources, including unstructured textual sources</td>
</tr>
<tr>
<td></td>
<td>Presents tailored views and experiences to different personas with conceptual models that use business terms</td>
</tr>
<tr>
<td></td>
<td>Flexibly accommodates new data sources and use cases on the fly, with minimal impact</td>
</tr>
<tr>
<td></td>
<td>Scales horizontally to accommodate enterprise data fabric scale</td>
</tr>
</tbody>
</table>
Semantics make the impossible practical.

They make it possible for business people to understand enterprise data in business terms, regardless of how or where it is stored and formatted.

<table>
<thead>
<tr>
<th>WITHOUT SEMANTICS</th>
<th>WITH SEMANTICS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data is described in technical terms.</td>
<td>Data is described in business terms.</td>
</tr>
<tr>
<td>No common model of key concepts across the business.</td>
<td>A common model unifies the business, enables consistency and collaboration</td>
</tr>
<tr>
<td>Data is fragmented in separate siloed systems.</td>
<td>Data from across the enterprise is connected and blended.</td>
</tr>
<tr>
<td>A single entity is described many ways in different systems.</td>
<td>Multiple descriptions of an entity are linked and harmonized.</td>
</tr>
</tbody>
</table>
Graph data models flexibly connect and transform new data sources.

### Claims

<table>
<thead>
<tr>
<th>Claim ID</th>
<th>Process Date</th>
<th>Subscriber ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>44223</td>
<td>10/3/2015</td>
<td>C12345</td>
</tr>
<tr>
<td>44224</td>
<td>10/7/2015</td>
<td>C23412</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

### Electronic Health Records

<table>
<thead>
<tr>
<th>Patient ID</th>
<th>Condition</th>
<th>Drug Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>BA213</td>
<td>Sleep Apnea</td>
<td>Narcoleptol</td>
</tr>
<tr>
<td>CS289</td>
<td>Type II Diabetes</td>
<td>Insulin</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

### On Site Doctor Note

On July 3, 2016 Patient BA213 experiencing **headache and nausea** following 500mg dosage of sleep aid therapeutic, **Narcoleptol**.
Simple Healthcare Graph Data Model

What’s happening in my organization in terms of….

- Patients
- Encounters
- Providers
- Medications
- Costs
- Care Plans
- Claims
- Etc.
Real World Graphs
Get Big Fast

Vast
Hundreds of sources, representing thousands of entity types

Siloed
Different technologies, schemas, formats

Complex
Sprawling disconnected schemas, wide flat tables, and cryptic names

Unstructured
documents, emails, logs

Valuable
Hidden connections and common business definitions
Virtual Enterprise Knowledge-Graph

Unstructured Data Sources
Structured Data Sets

Blended Data Products

Data Layers
- Load graph subsets
- Refine data
- Link/Connect data
- Infer new data

Load graph subsets
Refine data
Link/Connect data
Infer new data
What does it take to make graph work at scale in a large company?

- High Performance OLAP
- Work with existing landscape
- End-to-end capabilities
- Support diverse users
- Collaboration and reuse
- Security and governance
<table>
<thead>
<tr>
<th>High Performance OLAP</th>
<th>Diverse Users</th>
<th>End-to-end Capabilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Large data volumes</td>
<td>• GUI</td>
<td>• Data pipelines</td>
</tr>
<tr>
<td>• Diverse data types, formats</td>
<td>• No code, code optional</td>
<td>• Catalog, Tech and Biz view</td>
</tr>
<tr>
<td>• Heavy load complex queries</td>
<td>• Guided Experience</td>
<td>• Blend, prepare, transform</td>
</tr>
<tr>
<td>• Lots of users</td>
<td>• Built in functional toolkit</td>
<td>• Visualize, analyze</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Use</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Security and Governance</th>
<th>Collaboration and Reuse</th>
<th>Existing Landscape</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Authorization / Authentication</td>
<td>• Business metadata, tags</td>
<td>• Data sources</td>
</tr>
<tr>
<td>• Data Access Controls</td>
<td>• Shared data sets</td>
<td>• Desktop tools</td>
</tr>
<tr>
<td>• Logging</td>
<td>• Shared prep processes</td>
<td>• Metadata Repository / Catalog</td>
</tr>
<tr>
<td>• Encryption</td>
<td>• Data “zones” – raw to ready</td>
<td>• Schedulers</td>
</tr>
<tr>
<td>• Standards and processes</td>
<td></td>
<td>• LDAP/ AD</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Data pipelines or ETL processes</td>
</tr>
</tbody>
</table>
Graph Is Becoming the Right Solution for an Expanding Range of Mainstream Use Cases

**TRADITIONAL “NICHE” GRAPH USE CASES**
- Integrated data but small volumes
- Complex data structures
- Need flexible modeling
- Graph algorithms
  - R&D Acceleration (pharma)
  - Fraud Detection
  - Recommendation Engines
  - Network and IT operations
  - Search
  - Master data management

**MODERN MAINSTREAM ANALYTIC USE CASES**
- Big / Broad data analytics revolution
- Expanded analytic-ready data
- Speed to answer
- Digital transformation
  - Self-service on-demand data
  - Data discovery / integration in the enterprise data fabric
  - ALL enterprise data, blended and analytics ready

©2018 Cambridge Semantics Inc. All rights reserved.
4 Step User Experience

ON-BOARD
Catalog and map your existing data assets – structured or unstructured.

MODEL
Translate dataset into graph models. Add business definitions, object types, and relationships with semantics.

BLEND
Create blended analytic ready datasets. Connect graph models. Transform data. Harmonize into canonical models.

ACCESS
Analyze data using semantic and graph models. Export data for use with BI, analytics, and machine learning tools.
A modern data discovery and integration platform that lets business users find, connect, and blend enterprise data into analytic ready datasets.

**ON-BOARD**
Catalog your existing data assets. (RDBMS, flat files, JSON, XML, PDF, DOC, JPG, PNG, on-prem or cloud)

**MODEL**
Translate dataset into graph models. Add business definitions, object types, and relationships with semantics

**BLEND**
Connect graph models. Transform data. Harmonize towards canonical model. Create blended analytic ready datasets

**ACCESS**
Analyze data using semantic and graph models. Alternately export data for use with BI, analytic and data visualization tools
Begin your journey.

1. Identify your initial use case
2. Define the IT/business partnership
3. Quick start deployment 4 - 8 weeks
4. Leverage CSI technical expertise /staff
Thank You