A Model-Driven Approach to Data Integration with Entity Services

Justin Makeig, Product Manager, MarkLogic
Charles Greer, Lead Engineer, MarkLogic
LOAD
Load From Any Source in Its Most Natural Form

- No pre-defined schema required
- Entities, messages, hierarchical data naturally as documents
- Out-of-the-box data movement tools
- Parallel ingest with elastic provisioning
Discover and Transform Data Iteratively

- Ask-anything universal indexing: Text, scalars, structure, geo, semantics
- Built-in scripting environment runs real code close to the data, indexes
- Always consistent and secure
Integrate With Anything via Data Services

- Built-in HTTP server simplifies infrastructure and integration
- Transform on read or write
- Real-time query, analytics, and updates are interactive, always consistent
- Deploy anywhere, scale up (or down) on-demand
“An implementation should be conservative in its sending behavior, and liberal in its receiving behavior.”

Postel’s Principle
Jon Postel
RFC 760 (January 1980)
“A data hub should be liberal in its receiving behavior and conservative in its sending behavior”

Makeig’s Principle
Where Do Meaning and Context Live Today?

Application code
Relational schemas, domain models, ERDs
ETL scripts, stored procedures
Payment terms on orders of in-stock products over $10,000 to current customers must be NET30 unless using discount code PROMO1…

Huh?

Domain Expert

Developer
Payment terms on orders of in-stock products over $10,000 to current customers must be NET30 unless using discount code PROMO1.

- What is a “payment term”?
- What is an “order”?
- How is an “order” related to a “product”?
- How do I determine if a product is “in-stock”?
- How do I calculate the total value of an “order”?
- What’s a “customer”?
- How are “customers” related to discount codes?
Catalog the Context and Meaning of Data

- Centrally document entities, properties, and relationships in a queryable model
- Only model what you need, when you need it, *not* before
- Automatically derive services, transformations, configuration from the model
- Extend the model to capture other context
Expect That Over Time, Everything Can Change, at Any Time
1. **MODELING VOCABULARY**
   A structured way to describe entities, properties, and relationships

2. **PERSISTENCE CONVENTION**
   A standard way to represent harmonized entities, metadata, and raw data as documents

3. **APPLICATION SCAFFOLDING**
   Code and configuration generated from the model and designed for extension
1. **MODELING VOCABULARY**
   A structured way to describe entities, properties, and relationships

2. **PERSISTENCE CONVENTION**
   A standard way to represent harmonized entities, metadata, and raw data as documents

3. **APPLICATION SCAFFOLDING**
   Code and configuration generated from the model and designed for extension
Entity Type Model

- **Entities** (noun)
  A *Customer* is something that exists as part of my business/mission

- **Properties** (adjective)
  Customer entities have a *Name* that is of type string and is required

- **Relationships** (verb)
  Customers *place* Orders

Start small

- Entities
  - Properties
  - Relationships

MODELING VOCABULARY
Extending the Entity Type Model

- Governance
- Provenance
- …anything else

Finite core

Infinite extensibility with Semantics
1. MODELING VOCABULARY
   A structured way to describe entities, properties, and relationships

2. PERSISTENCE CONVENTION
   A standard way to represent harmonized entities, metadata, and raw data as documents

3. APPLICATION SCAFFOLDING
   Code and configuration generated from the model and designed for extension
Wrapping in Envelopes

Source systems “as-is”

- Transactional updates
- Granular security
- Indexes
- Bitemporal history
1. **MODELING VOCABULARY**
   A structured way to describe entities, properties, and relationships

2. **PERSISTENCE CONVENTION**
   A standard way to represent harmonized entities, metadata, and raw data as documents

3. **APPLICATION SCAFFOLDING**
   Code and configuration generated from the model and designed for extension
Model-Driven

Data Steward
- Upstream data models
- Business rules
- Domain knowledge

Developer
- Transformation stubs
- TDE
- Index configuration
- Search options
- XML Schemas
- Type version translation
- REST CRUD, query, and discovery services
- Much more…
Entity Services

Data Steward

Model Descriptor

Semantic Model

Developer

Code

Configuration
DEMO
Customer 360º

A
ERP
Customer

B
SUPPORT
Contact

C
MARKETING
Party

“Customer”
Customer 360°

“Customer”

1

2

ERP

Customer
Customer 360°

“Customer”

1. [Diagram of code]

2. [Plug icon]

3. [Person icon]

ERP

Customer

- Harmonized
- Raw source
- Provenance
Customer 360º
Customer 360°

“Customer”

ERP

Customer

Harmonized

Raw source

Provenance
DEMO

Customer 360°
Customer 360°

- Identify “Customer” entities from raw sources
- Model and harmonize Customer properties
- Relate Customer to Order and Product entities
- Extend the model to capture governance rules
Demo

For real, this time
“You’ve always had the power, my dear. You just had to learn it for yourself.”

Glinda, Good Witch of the East
Model-Driven Data Integration

- Centralized definition of entities, properties, and relationships
- Governed and queryable
- Tools to materialize as documents and triples
- Generate mappings, schemas, indexes, config...
- Extend model with custom Semantics
- Iterative, evolutionary, and as-you-need-it
MIFID version 2 requires us to retain Customer records about domestic swap trades for 7 years after execution or any corrections.

- What is a “customer”?
- What is a “swap trade”?
- How is a “swap trade” related to other trade types? Other financial instruments?
- How is “domestic” designated on a trade?
- How are customers associated with trades?
MIFID version 2 requires us to retain Customer records about domestic swap trades for 7 years after execution or any corrections.

- What is a “customer”?
- What is a “swap trade”?
- How is a “swap trade” related to other trade types? Other financial instruments?
- How is “domestic” designated on a trade?
- How are customers associated with trades?
Person is governed by Legislation
Person has property SSN
Person has PII
<<Legislation>>
SSN has property dicted by PII