Relational to NoSQL
Northern Trust’s Adoption Journey With MarkLogic

MarkLogic World – May 2018

Maureen Penzenik
Solutions Architect
Northern Trust

Puneet Rawal
Principal Sales Engineer
MarkLogic
@puneetrules
TODAY’S AGENDA

Who We Are
About Northern Trust and our Technology Principles

Our Journey
Data Management Platform Guidance
MarkLogic Applied to Northern Trust’s Business
Mind Shift: Relational to NoSQL

Moving Forward
10 Key Takeaways
What’s Next
NORTHERN TRUST – WHO WE ARE

One of the world’s leading financial services companies

Wealth Management
• Individuals
• Families
• Family offices
• Foundations
• Endowments
• Privately held businesses

As of December 31, 2017

Corporate & Institutional Services
• Pensions
• Sovereign entities
• Fund managers
• Foundations & endowments
• Insurance companies

Integrated Global Operating Platform

$10.7T
ASSETS UNDER CUSTODY/ADMINISTRATION

$1.2T
ASSETS UNDER MANAGEMENT

$139B
BALANCE SHEET

northerntrust.com | © 2018 Northern Trust
DELCIVERING REAL CLIENT BENEFITS

A strategic technology platform that aims to enable the future data and digital needs of all our clients

- **Timeliness**: Access to real or near real time data
- **Data Quality**: Single source of truth as defined by client, overlaid with NT discipline
- **Speed/Agility**: Speed to market, speed of change
- **Data Accessibility**: Data and digital solutions via client/investor portal, APIs
- **Service on Demand**: For clients, distributors and investors
- **Data Insight**: Fund, investor, distribution and analytics data

API: Application program interface
We leverage several key methods for success in service delivery:

- Agile
- Cloud
- Services
- Open Source
- Big Data
RANGE OF DATA PLATFORMS

- SQL
- Hadoop
- MarkLogic
WHICH DATA PLATFORM TO USE?

It isn’t always the shiny new toy
DATA PLATFORM CRITERIA – RELATIVE POSITIONING

- **MARKLOGIC**
- **RDBMS**
- **HADOOP**

<table>
<thead>
<tr>
<th>Criteria</th>
<th>MARKLOGIC</th>
<th>RDBMS</th>
<th>HADOOP</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data Format</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Structured</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unstructured</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Schema Variability</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Advanced Search</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Semantics/Graph Data</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Skill Set Availability</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Infrastructure Availability</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Longer Lead Times</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shorter Lead Times</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Legend: M = Medium, H = High, R = Recommendation
## TRANSACTION HUB OVERVIEW

### Goals

- Build the next generation transaction processing engine on a scalable, rapid delivery, microservices architecture
- Create operational data store of transaction lifecycle and state management
- Leverage ISO 20022 standard

### Requirements

<table>
<thead>
<tr>
<th>Capture…</th>
<th>Validate…</th>
<th>Enrich…</th>
<th>Transform…</th>
</tr>
</thead>
<tbody>
<tr>
<td>…trade messages and archive in original format (SWIFT, XML, fixed length, proprietary)</td>
<td>…trade message is accurate and complete</td>
<td>…trades with additional reference data</td>
<td>…trade messages into a format to route to downstream systems.</td>
</tr>
</tbody>
</table>
## APPLYING MARKLOGIC TO USE CASES

<table>
<thead>
<tr>
<th>Use Case</th>
<th>Business Challenge</th>
<th>Search</th>
<th>ACID Compliance</th>
<th>Flexible Structures</th>
<th>Streamline ETL</th>
<th>Query Metadata</th>
<th>Store JSON</th>
<th>Store XML</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transaction Hub</td>
<td>• Data Format</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Schema Variability</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Scalable infrastructure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wealth Management Goals</td>
<td>• Data Format</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Powered Solutions</td>
<td>• Search</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Versioning</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Application Platform – Event</td>
<td>• Data Format</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Logging*</td>
<td>• Schema Variability</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Search</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Client Account Valuations</td>
<td>• Schema Variability</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Warehouse Modernization</td>
<td>• Schema Variability</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• “Time to Market”</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Governance</td>
<td>• Schema Variability</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Data Format</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(PoC) MDM</td>
<td>• Schema Variability</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Ingesting New Sources</td>
<td>• “Time to Market”</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Request/Event Logging*</td>
<td>• Data Format</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Application Event Logging theme emerging. Store in XML or JSON
MIND SHIFT – FROM RELATIONAL TO NOSQL

MarkLogic is a database with built-in “Google-like” search, built-in triple store, built-in app server, and ACID transactions

- **ODBC/JDBC**
  - Use one call for transactions with MarkLogic since the REST API is stateless

- **Java Abstraction Layer**
  - Push database operations to the database (i.e., keep low level “data logic” in MarkLogic)

- **ORM (Object Relational Mapping)**
  - Leverage the document model where possible, don’t try to model every data element
## MIND SHIFT – APPLIED TO SPECIFIC PROJECTS

<table>
<thead>
<tr>
<th>Project</th>
<th>Business Challenge</th>
<th>Search</th>
<th>ACID Compliance</th>
<th>Flexible Structures</th>
<th>Streamline ETL</th>
<th>Query Metadata</th>
<th>Store JSON</th>
<th>Store XML</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wealth Management Goals</td>
<td>• Data Format</td>
<td>☢️</td>
<td>☢️</td>
<td>☢️</td>
<td>☢️</td>
<td>☢️</td>
<td>☢️</td>
<td>☢️</td>
</tr>
<tr>
<td>Powered Solutions</td>
<td>• Search</td>
<td>☢️</td>
<td>☢️</td>
<td>☢️</td>
<td>☢️</td>
<td>☢️</td>
<td>☢️</td>
<td>☢️</td>
</tr>
<tr>
<td></td>
<td>• Versioning</td>
<td>☢️</td>
<td>☢️</td>
<td>☢️</td>
<td>☢️</td>
<td>☢️</td>
<td>☢️</td>
<td>☢️</td>
</tr>
<tr>
<td>Application Platform – Event</td>
<td>• Data Format</td>
<td>☢️</td>
<td>☢️</td>
<td>☢️</td>
<td>☢️</td>
<td>☢️</td>
<td>☢️</td>
<td>☢️</td>
</tr>
<tr>
<td>Logging</td>
<td>• Schema Variability</td>
<td>☢️</td>
<td>☢️</td>
<td>☢️</td>
<td>☢️</td>
<td>☢️</td>
<td>☢️</td>
<td>☢️</td>
</tr>
<tr>
<td></td>
<td>• Search</td>
<td>☢️</td>
<td>☢️</td>
<td>☢️</td>
<td>☢️</td>
<td>☢️</td>
<td>☢️</td>
<td>☢️</td>
</tr>
<tr>
<td>Client Account Valuations</td>
<td>• Schema Variability</td>
<td>☢️</td>
<td>☢️</td>
<td>☢️</td>
<td>☢️</td>
<td>☢️</td>
<td>☢️</td>
<td>☢️</td>
</tr>
</tbody>
</table>
TOOLS UTILIZED

Slush Data Explorer
TOOLS UTILIZED

**Slush Data Explorer**

**Docker Sandbox**
TOOLS UTILIZED

- Slush Data Explorer
- Docker Sandbox
- ML Gradle for Process Automation
PILLARS OF DATA GOVERNANCE

Cataloging
- Are there agreed upon business terms and definitions?

Data Quality
- Is the data fit for purpose? Is it accurate, timely, consistent, etc.?

Provenance & Lineage
- Where did the data originate, and how did it change?
IMPLEMENTING DATA GOVERNANCE AT NORTHERN TRUST

Existing Infrastructure
• Rich set of tools (custom and vendor provided)
• Tightly integrated with RDBMS from design, ETL, to production management

Data Governance Goals with MarkLogic

Short Term
• Define critical data elements (CDE) that every project is required to identify
• Data is extracted via ODBC and utilized in existing tools

Long Term
• Real-time integrations
• REST services that can return reports
• Inform MarkLogic’s product roadmap
10 Key Takeaways
XQuery vs JavaScript
<table>
<thead>
<tr>
<th>JSON</th>
<th>XML</th>
<th>RDF</th>
<th>JSON/XML + RDF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Structured data stored as objects</td>
<td>Structured / unstructured data &amp; text</td>
<td>Facts &amp; relationships</td>
<td>Systems of entities &amp; relationships</td>
</tr>
<tr>
<td>- Schema-agnostic</td>
<td>- Schema-agnostic</td>
<td>- Define facts and relationships</td>
<td>- Documents can contain triples</td>
</tr>
<tr>
<td>- Query with JavaScript</td>
<td>- Query with XQuery</td>
<td>- Atomic structure</td>
<td>- Triples can annotate documents</td>
</tr>
<tr>
<td>- Compact, fast to parse</td>
<td>- Can store objects, sets, and many data types such as dates, durations, integers, etc.</td>
<td>- Universal standards (RDF and SPARQL)</td>
<td>- Graphs of triples can contain documents</td>
</tr>
<tr>
<td>- 6 kinds of values: objects, arrays, floats, strings, booleans, nulls</td>
<td>- Uses namespaces, comments, and attributes</td>
<td>- Used for reference data, metadata</td>
<td>- Enhanced querying:</td>
</tr>
<tr>
<td>- Avoids namespaces, comments, attributes</td>
<td>- More maturity than JSON as a data model</td>
<td></td>
<td>- Expand/Restrict document search</td>
</tr>
<tr>
<td>- Common data format for the web</td>
<td></td>
<td></td>
<td>- Enhance graph search</td>
</tr>
</tbody>
</table>
Shared Cluster vs Dedicated Cluster
**SHARED VS DEDICATED RESOURCE? IT DEPENDS**

<table>
<thead>
<tr>
<th>SHARED</th>
<th>DEDICATED</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Shared Resource Diagram" /></td>
<td><img src="image" alt="Dedicated Resource Diagram" /></td>
</tr>
</tbody>
</table>
Put “Data Logic” in the Database
PUSHING DATA OPERATIONS TO MARKLOGIC

TRADITIONAL APPROACH

DATABASE

JAVA LAYER

MARKLOGIC APPROACH

MARKLOGIC

JAVA LAYER
Use Profiling and Debugging Tools
Think about URI Naming Standards
**THINK** ABOUT YOUR URI – IT’S THE KEY TO YOUR DOCUMENT!

Example of a standard:

URI: /edw/valuation/cashPosition/CALCULATION/23423423/2017-01-01.json

- **Prefix:** described the domain or project.  *Example: /edw/
- **Document Type:** the name of the document. This can also include the domain such as: /valuations/cashBalance/
- **Unique IDs:** the set of values in the document that define the document as unique.
- **Extension:** either 'json' or 'xml'.

Example Unique ID for document below:

CALCULATION/23423423/2017-01-01

```
{
  "action": "CALCULATION",
  "generationDate": "2017-01-01",
  "accountId": 23423423,
  "value1": 100.00,
  "value2": 101.00,
  "intermediateCalculation": {
    "sum1": 10,
    "sum2": 20
  }
}
```
Java Client API: Say No to the POJO
SAY NO TO THE POJO REPOSITORY

With the Pojo Repository:

- *Object names can become embedded in your document*
- *Default URIs are generated*

Document URI:

```
com.ntrs.thub.tbor.service.file.TransactionFileEntity/25a71c32-a594-40fd-af34-9279ea85c228.json
```
Store Data *As Is*
STORING DATA AS IS

- Store the original data **as is**: in XML, JSON, or text – do **not** wrap it in quotes as text
3

Store/Validate XSD (ISO 20022 schemas)
STORING / VALIDATING XSD IN MARKLOGIC
Use the Envelope Pattern
USE THE ‘ENVELOPE’ PATTERN TO ENRICH DOCUMENTS

Benefits:

• **Canonical**/enterprise data model
• **Metadata**
• **Source Content** *as is*

```json
{
    "envelope": {
        "harmonized": [],
        "metadata": [],
        "source": {
            "your data": "goes here"
        }
    }
}
```
ENVELOPE PATTERN EXAMPLE

2 Sections:
- **Metadata** for lineage
- **Content** or “core” of the document
Choose Your Journey and Become MarkLogic Certified

Administrator Track
Need to manage your environment? Learn to deploy, monitor, and manage MarkLogic.

LEARN MORE

Business User Track
Need a flexible and trusted technology for a project? Learn how MarkLogic brings value to your business.

LEARN MORE

Data Architect Track
Need to design and manage your data architecture? Learn how to use MarkLogic to integrate your data.

LEARN MORE

Developer Track
Want to build that awesome app? Get off the ground quickly with the MarkLogic developer track.

LEARN MORE

marklogic.com/learn/university/
SUMMARY – WHAT’S NEXT?

- Establish MarkLogic Center of Excellence
- Continue MarkLogic Community of Practice
- Continue to build a repository for documentation, knowledge share, best practices
- Leverage ISO 20022 standards and ontologies (FIBO)
- Integrate ML Gradle into CI/CD processes.
- Explore Data Hub Framework
IMPORTANT INFORMATION

This information is not intended to be and should not be treated as legal advice, investment advice or tax advice and is for informational purposes only. Readers, including professionals, should under no circumstances rely upon this information as a substitute for their own research or for obtaining specific legal or tax advice from their own counsel. All information discussed herein is current only as of the date appearing in this material and is subject to change at any time without notice.

Confidentiality Notice: This communication includes confidential material and is meant only for the intended recipients in relation to the presentation accompanying this information. All materials contained in this presentation, including the description of Northern Trust, its systems, processes and methodologies, are proprietary information of Northern Trust. By accepting these materials, the recipient agrees to keep such material strictly confidential and to not distribute such materials or any part thereof to any other person.