Integration Patterns for Virtual MDM Implementations

Bhavani Kumar Eshwar
Senior Technical Staff Member
CTO – MDM Engineering, India Software Labs

Coverage – MDM Technology Upgrades, Portfolio Integrations & Cloud/Bluemix/DataWorks Offerings
Please Note …

IBM’s statements regarding its plans, directions, and intent are subject to change or withdrawal without notice at IBM’s sole discretion.

Information regarding potential future products is intended to outline our general product direction and it should not be relied on in making a purchasing decision. The information mentioned regarding potential future products is not a commitment, promise, or legal obligation to deliver any material, code or functionality. Information about potential future products may not be incorporated into any contract. The development, release, and timing of any future features or functionality described for our products remains at our sole discretion.

Performance is based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput or performance that any user will experience will vary depending upon many factors, including considerations such as the amount of multiprogramming in the user’s job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve results similar to those stated here.
Agenda

1. Recap – Virtues of Virtual MDM
2. Integration Patterns Overview
3. Capabilities associated with the Integration Patterns
4. Closing Notes
# Virtues of Virtual MDM – A recap

## Virtual MDM Core facets

- Also called Registry Style MDM
- Source systems author and own the data
- In most cases the Master Data is not modified
- Organizations can match/link duplicate records and produce golden views
- Golden Views but created on the fly i.e. current point in time

## Virtual MDM Benefits

- A virtual MDM style can facilitate a rapid, cost-effective implementation
- Does not require changes to source systems
- Serves best to Organizations in the early stages of a Unified Governance strategy
- Complete, highly accurate, real-time view of master data thoroughly supported by configuration
- Offers quick time-to-value

## Virtual MDM What InfoSphere MDM SE offers

- Virtual master registry to assemble a trusted, tailored view
- Workflow capabilities to implement policies and processes for data stewardship and data governance
- Flexible and extensible data model built for MDM
- Highly accurate probabilistic matching and search
- Policy management to author, monitor and enforce data quality
- Solutions framework to enable and enhance applications and processes with master data
- Data stewardship to inspect and resolve data quality issues
- Relationship and hierarchy management to manage household and B2B relationships
- Last but not the least – Key Integration capabilities
Integration Patterns for Virtual MDM Implementations

INTEGRATION PATTERNS

OVERVIEW
Three key patterns to look at …

Acquire

Patterns for Data Acquisition and Ingestion

Manage

Patterns for managing the Virtual MDM hub

Deliver

Patterns for delivering data from Virtual MDM hub
Integration Patterns for Virtual MDM Implementations

DATA ACQUISITION AND INGESTION
Integration Patterns – Data Acquisition and Ingestion, the big picture

- The first stepping stone to success in Virtual MDM implementations is fluent and extensible methods for Data Onboarding from Sources

![Diagram of data acquisition and ingestion process]

**Acquire**
- **Data Sources**
  - Deal with all kinds of data sources
  - On-Premise Sources
  - SaaS Applications
  - Other Cloud Hosted / Bluemix Applications

**Extract/Transform**
- Employing methods to prepare data
  - Data Analysis
  - Full data extractions
  - New Source addition
  - Transformation
  - Delta Loads
  - Messaging
  - Events

**Curate**
- Early DQ subroutines
  - Optional in Virtual MDM
  - Used for Standardization only for better matching
  - Sometimes used for deterministic de-dupe per source

**Transfer**
- Landing prepared data for MDM onboarding
  - Using ETL Tool
  - Landing Messages for MDM to process
  - Using Client APIs
  - Using Services
  - Using Change Data Capture

**Load**
- Execution of Data onboarding flows
  - ETL Job
  - Message processing
  - Add/Update type Client API calls
  - Add/Update Web Service or Rest API calls
  (CDC can be rolled into ETL Jobs)
Data Acquisition and Ingestion - Patterns & capabilities

- Items to cover for the background
  - Core interfaces
    - Data Analysis (external to Virtual MDM)

- Loading data into Virtual MDM in a Batch (applicable for Initial Loads and Batch Loads)
  - ETL Tools
  - Change Data Capture

- Processing source systems messages

- Processing source system events
Core Interfaces

- Virtual MDM offerings typically provide certain core interface packages for –
  - Managing data in the Virtual MDM hub
  - Performing certain actions like Searches etc.
  - Extension points to perform additional processing in regular out-of-box flows

- InfoSphere MDM SE has –
  - SDK for Java and Web Services as the Default Client API Interfaces
  - Enterprise Service Oriented Architecture (SOA) Toolkit
  - Message Broker Suite
  - Event Notifications
  - Handlers
Data Analysis - Information Analyzer Integration

**Use Cases**
- For use in inspection of sources
- Data Validation and metrics for source and trusted Data
- Source data Profiling to fix initial Data issues
- Also aids in provide key inputs for Matching configurations

- Sample asset available on IBM Web Management (IWM) site
- The sample includes an InfoSphere Information Analyzer project referred to as the *StartUp IA Project*

- This project includes a set of pre-defined rules that are specific to InfoSphere MDM and the start-up data model

- The IBM InfoSphere Metadata Asset Manager (IMAM) tool is used to import the schema metadata into the InfoSphere Information Server common repository

- InfoSphere Information Analyzer uses the common repository when the data analysis is performed. Data rules and column analysis are run on the InfoSphere Information Server

- The results are analyzed using the reporting capabilities of InfoSphere Information Server
### MDM Analysis Sample Rules

- **Attribute Frequency**
  - What percentage (%) of the Last Names are from source A, source B, and so forth

- **Attribute Completeness**
  - How many Last Name fields are populated, how Last Name fields have no value (NULLs), and so forth

- **Attribute Validity**
  - This includes passing in a set of anonymous (ANON) values and checking which ones are ANONs
  - Validating with an external source
  - Validating reference data

- **Attribute conformance**
  - Data of birth (DOB) falls within a set range, such as from 1910 to 2013
  - Format checking in general, such as Drivers License (DL) always use digits
Loading data into Virtual MDM in a Batch

Source Data Extraction (full or delta)

Virtual MDM bulk loading solution

Change Data Capture from Source Systems

Using Change Data Capture to prepare payload

Virtual MDM bulk loading solution

Virtual MDM Inbound Messaging Processing

If data preparation requires Call backs and complex transformations for multi-row payloads

If data preparation can be covered with simple mappings
Virtual MDM Bulk loading solution

- MDM Connector Stage
- Mpxdata Utility and InfoSphere DataStage
- Legacy Component - Clover ETL
Virtual MDM Bulk loading solution – MDM Connector Stage

- A native Stage in Information Server which can be used in DataStage and QualityStage Jobs
- Currently supports functions to load data into or extract data from InfoSphere MDM SE
- Uses MDM MetaData provided by MDM Workbench’s MetaData Export Capability

MDM Development tasks

MDM Workbench

Exports the project metadata as XMI files for use by MDM Connector

XMI Files

Metadata Integration Tasks

IMAM Asset Manager

Import XMI files, analyze, preview and upload to Metadata Server

MDM Model Bridge

XMeta (MDM, ASCL)

SCM or DevOps Repository

DataStage/Quality Stage Job Development Tasks

DataStage/QualityStage Designer

Configure MDM Stage with MDM Hub Connection and other settings

Query MDM Model in XMeta

Persist Stage configuration in DSX Model

Compile, Deploy & Test Jobs

XMeta (DSX)
Points to note about MDM Connector Stage

- MetaData Export cover InfoSphere MDM SE and InfoSphere MDM AE

- Data Reads and Writes of MDM Connector Stage currently support InfoSphere MDM SE and Healthcare offering which run on InfoSphere MDM SE
  - memget and memput calls are currently supported
  - Implementations need variations and enhancements for the read/write capabilities served by the memget and memput API callouts and this should be available in a future release

- On the Road map
  - memsearch, memscore, memdelete, tskget, tskput, tsksearch etc.
  - MDM SE Engine batch processing functions like Batch Compositor and Bulk Cross Match
  - Support for other Editions
Virtual MDM Bulk loading solution - Mpxdata Utility and InfoSphere DataStage

- Some implementations still leverage mpxdata utility for Data Ingestion into InfoSphere MDM SE
  - MEMCOMPUTE mode in conjunction with madhubload and/or madentload utilities
  - MEMPUT mode

- The challenges around this approach is that there is no standards for Data Acquisition from sources which also elevates the problem of end-to-end automation of the Data Acquisition and Ingestion flows

- A sample InfoSphere DataStage Job is provided to assist in designing this “end-to-end” automation of the Data Acquisition and Ingestion flows
Data Acquisition and Ingestion - Processing source systems’ messages and Events

- InfoSphere MDM SE provides the Message Broker Suite
  - Has various options for inbound and outbound message processing
  - Incoming messages from sources systems either in the form of events or payload messages can be processed by the Message Broker Suite to Ingest data into MDM Standard edition

- On the road map - IBM Integration Bus
  - Good set of our customer are consolidating to have Reliable Messaging component be IBM Integration Bus
  - Multiple requests for Message Broker Suite capabilities to be made available in IBM Integration Bus
  - Current plan -> Technical Preview by November 2015 and OOTB capabilities in 2016
Integration Patterns for Virtual MDM Implementations

MANAGING THE VIRTUAL MDM HUB
Integration Patterns – Managing the Virtual MDM Hub, the big picture

- Managing a Virtual MDM Hub entails many functions some of which are specific to the core Engine and some deal with integrations

Most of the topics around Managing the Virtual MDM Hub relate to Governance
- Integration with IBM’s Business Process Manager, IBM Operational Decision Manager, Cognos etc. will fall under this area
- These is a full session to cover this area
  - Master Data Stewardship & Governance by Jay Limburn - 30 Jun 2015, 10-11AM EST

The key topics we will cover under this section are –
- Federated Query across Virtual MDM Hubs
- Integration with InfoSphere Information Server’s QualityStage
- Specialized Integrations for Free Text Searches – IBM’s Watson Explorer Integration
- Working with Events from InfoSphere MDM SE
Managing the Virtual MDM Hub – QualityStage Integration

- Provides for Address Standardization via a default pre-interaction handler
  - Customer handler too can be written

https://www.youtube.com/watch?v=T5jlMjOrxsQ
Managing the Virtual MDM Hub - Federated Queries

- Provides cross-organization identity governance without a centralized database
- Ideal Information Exchange and cross-jurisdiction public sector use cases

Enhanced Information Consumption
Federated Query for MDM Standard Edition
Managing the Virtual MDM Hub – Flexible Search and Watson Explorer connector

- **Flexible Search**
  - Provides search capabilities for members without relying on bucketing in InfoSphere MDM SE
  - Attributes to be indexed are configurable
  - Free Text and Free Term Searches are enabled (IxnMemTextSearch and IxnMemTermSearch interactions)
  - Java & Rest APIs are provided
  - Member data for the configured attributes is pushed into the Fuzzy Search component used i.e. IBM Watson Explorer

- **IBM Watson Explorer Connector for InfoSphere MDM SE**
  - IBM Watson Explorer also ships with a connector for InfoSphere MDM SE
  - Crawls thru InfoSphere MDM SE repository
    - Member Type Name, Entity Type Name and optionally Composite View Name is used as configuration for crawling over
    - Payload can be calibrated using Segment Code Filters
  - Redbook: Building 360-Degree Information Applications
Enhanced 360° view of the customer

Customer search: MDM draws in all related records: J Robertson, Janet Robertson, Jan Baker

MDM enables a complete purchase history, including Jan Baker’s records from before 2011

Customer info from MDM

Customer’s Products from MDM

Indexed 3rd party information related to customer

Unstructured internal information related to customer
Managing the Virtual MDM Hub - Working with Events from InfoSphere MDM SE

- InfoSphere MDM SE produces events based on certain conditions while processing workloads

- These event’s obviously have implications and the key to those are the ones which may entail Stewardship actions or delivery of the output of the associated views to consuming systems

- Will be covered in the next section, where we look into Patterns for delivering data from Virtual MDM hub

- Some of the key Events produced by InfoSphere MDM SE

  - Entity Created
  - Entity Deleted
  - Entity Updated
  - Member Process
  - Member Created
  - Member Deleted
  - Member Linked
  - Member Merged
  - Member Reprocessed
  - Member Undeleted
  - Member Unlinked
  - Member UnMerged
  - Member Updated
  - Member No-link
  - Task Created
  - Task Deleted
  - Task Resolved
  - Task Updated
Integration Patterns for Virtual MDM Implementations

DELIVERING DATA FROM THE VIRTUAL MDM HUB
Integration Patterns – Delivering data from the Virtual MDM Hub, the big picture

**Deliver**

**Real-time Data Access**
- APIs providing real-time access to golden views

**Events**
- Handling Key events provided by the Virtual MDM Hub
- Processing MDM Events to notify end points or routing
- Routing MDM Events to ETL Tools

**Full & Delta Extracts**
- Getting data out of the Virtual MDM Hub based on requirements
- Using ETL Tool
- Selective Extracts
- Full Extracts

**Data Transformation**
- Managing adherence to Rules and Policies
- Using ETL Tool
- Pre-processing data
- Mapping to End-points

**End-point delivery**
- Dropping off data to various end-points
- On-Premise Endpoints
- Warehouses
- Big Data Environments

**SaaS Applications**
- Salesforce
- Workday

**Other Cloud Hosted / Bluemix Applications**
- Bluemix
- ERP

• Using Client APIs
• Using Services
• Using Flex Search
Delivering Data from the Virtual MDM Hub - Patterns & capabilities

- Real-time Data Access
  - Covered in Core Interfaces
  - APIs to Search/Get Data

- Handling Events
  - OutBound Brokers
  - Custom Message Processing
  - Hooking Queues to ETL for Processing

- ETL Usage – Virtual MDM Bulk Extract Solution
  - Triggers: Event, Full Extracts, IDs for Selective Extract
  - Data Collations and Transformation and delivery to end points
  - MDM Connector Stage
  - Legacy Component - Clover ETL
Virtual MDM Bulk Extract solution – MDM Connector Stage

### InfoSphere Information Server

- **START**
- Implementation Specific Job trigger
  (Event or Full Extract of Delta Extract)
- Stage with Implementation specific MDM Keys sourcing strategy
  - Hash key (data) values, Range, Round robin, Random, Entire, Modulus, Tiered database partitioning, Dynamic
- Information Server Meta Data Repository
- MDM Metadata
  - MDM Connector
    - (memget API)
  - MDM Connector Stage with Implementation specific MDM Keys sourcing strategy
  - Hash key (data) values, Range, Round robin, Random, Entire, Modulus, Tiered database partitioning, Dynamic
- Information Server's MDM Interface components
  - MDM Standard Edition Instance
    - (Person, Patient, Provider, Clinical Data Hub)
  - MDM Database
- Configurations
  - Composite View
    | ID | Last | First | Street |
    |----|------|-------|--------|
    |    |      |       |        |
- Columns to Output Link Map
- Row-sets
  - MDM Connector
  - Information Server's MDM Interface components
- Output Link
  - Any outbound processing Stages ...
  - File
  - PureData for Analytics
  - Salesforce.com
  - Business Applications
  - BigInsights etc.
- Master Data Extract Sample Jobs are provided

---

© 2015 IBM Corporation
Integration Patterns for Virtual MDM Implementations

SPECIALIZED INTEGRATIONS
Unstructured Text Correlations
Analyze Unstructured text sources and manage associations with master data

How It Works
• Discover linkages between text and relevant master data entities
• Link to additional information
• Investigative tool for text analysis
• Complete, accurate and timely views

Client Value
• Enhanced customer service
• Quicker, more accurate issue resolution
• Increased upsell/cross opportunity

Usage
• Enhance detection of relationships between entities
• Enhance entity resolution from the evidence hidden within the documents.
• Enrich knowledge base by adding additional information to MDM records

Using MDM in fraud investigations
**Example: Public Safety**

1. **Potential suspect identified**

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td>Miram Mada</td>
</tr>
<tr>
<td>Address</td>
<td>Via Riasc 21 Compione Italy</td>
</tr>
<tr>
<td>Citizenship</td>
<td>United States</td>
</tr>
<tr>
<td>Place of birth</td>
<td>Portland, Oregon</td>
</tr>
<tr>
<td>Date of birth</td>
<td>1938-09-15</td>
</tr>
</tbody>
</table>

2. **Search master records in InfoSphere MDM**

3. **Uncover links between entities from documents, emails, blogs etc.**

4. **Matched entities identified; no automatic resolution of master record**

   - Miram Mada
     - EID: 1574 (Person)
     - Matches Entity

   - Maaranda Group
     - EID: 3652 (Person)
     - Matches Entity

   - Miram Mada
     - EID: 5010998 (Person)
     - Document

   - Maaranda Group
     - EID: 5011001 (Person)
     - Document

   - Demo.pdf
     - EID: 5010999 (Document)

   - Learned something interesting yesterday: Miram Mada seems to be in contact with Maaranda Group of Companies to help him execute on ETT. BTW, M.M. born in Portland, Oregon recently spent considerable time in Vienna, Austria where Maaranda is located.

   - BTW, I will be out on a personal trip to Bangalore, India starting next week and won't be back until March 25th.
On-premise Virtual MDM to Cloud Integration

InfoSphere MDM Supercharges Salesforce.com serving LOB

- Enrich SFDC ‘Account’ information by leveraging the broader enterprise master information in InfoSphere MDM
- Improved SFDC ‘Account’ search and remediation & support for both virtual and physical hubs
- Leverages WebSphere Cast Iron Cloud Integration on Information Server pack for SFDC
- MDM Services TIP (Template Integration Pattern) in Cast Iron tooling

Cloud Integration for Bluemix:
Provides the on-ramp for on-prem MDM Services for Data Refinery APIs

- MDM’s Cast Iron Live Orchestration APIs in Bluemix
- Gateway to Data Refinery for on-premise MDM APIs
- Developer appeal

1. Real-time synchronization
2. Near-real-time synchronization with reliable queuing

* v1.5.0.2 needed
Integration Patterns for Virtual MDM Implementations

CLOSING NOTES
Feedback and connect

- Provide Feedback to organizers on your preferences for deep dive topics in one or more integration areas you are interesting in.

- Contact me if you are dealing with issues or specific opportunities where you need help:
  - bheshwar@in.ibm.com
In Summary …

- Finance
- Government
- Retail
- Telco
- Insurance
- Travel & Transportation
- More …

- BPM component based Stewardship and remediation
- Master Data Policy Management
- Extend Master Data Governance to Business Applications
- Support for Mobile Apps

- Billion Scale Matching
- Social Data Integration
- Enterprise Free Text Search
- Data Integration on-ramps for Big Data Analytics

- Cloud Integration
- IaaS
- PaaS / DataWorks
- SaaS Solutions from BPs

- Virtual
- Physical
- Hybrid

- Line of Business oriented

- Industries

- Big Data Ready

- Cloud Enabled

The most Cohesive, Comprehensive, Complete MDM Portfolio

Acquire, Manage, Deliver Critical Information Assets
Thank you
Questions & Answers
Integration Patterns for Virtual MDM Implementations

BACKUP SLIDES
For details on Clover ETL, review the following links

- InfoSphere MDM SE Documentation

- We are continuing our partnerships with Javelin - Clover ETL’s external statements
  http://www.cloveretl.com/ibm-mdm
  http://www.cloveretl.com/content/cloveretl-guarantees-support-ibm-infosphere-mdm-users-and-enhances-their-data-processing-pos
Redbook to know more about Integrating SaaS Applications and MDM