Self-Service in the world of Data Integration

April 2011 San Francisco DAMA Meeting

Diby Malakar
Director Product Management
Agenda

• Introduction
• Business Problem
• Lean and Agile Data Integration
• Self-Service Data Integration Use Cases
• Self-Service and the Role of IT
• Key takeaways
• Q & A
Author Bio

• Mr. Malakar’s primary focus area is building products and solutions around Business-IT Collaboration and Self-Service Data Integration.

• His prior experience includes working as Acting-VP of Engineering and Senior Director of Product Management for Cloud9 Analytics, a SaaS-based startup focused on building sales performance management applications. He has also worked for companies like KPMG, Neoforma and TiVo in various engineering management and product management capacities. He has more than 15 years of experience in the information technology industry and specializes in business intelligence, data warehousing, and data quality management. He holds a Bachelor’s degree in Computer Science and a MBA in Information Systems.

• He is an active member of IAIDQ and is currently serving as the President of the San Francisco chapter of DAMA. He has also worked as a Program Director for the San Francisco chapter of DAMA. He is currently also serving on the DAMA International Board of Directors as VP of Industry Services. He has also been on the Speaker Selection committee for conferences such as the Enterprise Data World Conference. His articles have been published in a variety of places such as TDAN, Wharton’s Leadership Digest, and Oracle Magazine. IT takes too long to deliver data integration projects. Collaboration between the business and IT is not only inefficient but also error-prone in many cases. Business analysts are too dependent on IT to access and understand the data. This presentation will focus on the emerging market trend where in business analysts want to feel empowered to implement business logic end-to-end in a totally self-service way without developer help. They want to be able to provide a great jump-start for business logic that developers can then refine for end-to-end implementation. It will explore the challenges faced and the opportunity to improve productivity for both analysts and developers in the data integration community.
Abstract

IT takes too long to deliver data integration projects. Collaboration between the business and IT is not only inefficient but also error-prone in many cases. Business analysts are too dependent on IT to access and understand the data. This presentation will focus on the emerging market trend where in business analysts want to feel empowered to implement business logic end-to-end in a totally self-service way without developer help. They want to be able to provide a great jump-start for business logic that developers can then refine for end-to-end implementation. It will explore the challenges faced and the opportunity to improve productivity for both analysts and developers in the data integration community.
The Business-IT Challenge

*It Takes Too Long*

**Business Drivers**
- IT takes too long to deliver data integration projects
- Collaboration between the business and IT is inefficient and error-prone
- Business analysts are too dependent on IT to access and understand the data

**BUSINESS**
- Business Process owners
- Application owners
- Data Stewards

**IT**
- Data Analysts
- IT Developers
- Architects
How Long Does a Change Take?

**Reporting Scenario:** On-going requests for data that is **NOT** in the DW

- **Business**
  - Change Request
- **IT**
  - Deploy to Production

**What if?**

Change Request … Approve & Prioritize … Analyze & Design … Build … Test … Deploy

- **66%** of BI requirements change on between a *daily and monthly basis*
- **71%** of the respondents said they have to ask data analysts to **create custom reports** for them
- **36%** of custom report requests require a **custom cube or data mart** to answer the request
- **77%** of respondents cited that it takes **between days and months** to get their BI requests fulfilled

Lean Integration Principles

1. Focus on the customer & eliminate waste
2. Automate processes
3. Continuously improve
4. Empower the team
5. Build Quality In
6. Plan for change
7. Optimize the whole

Agile Data Integration

Taking lessons from lean...

Find data sources and targets
Profile data sources for data quality
Identify join conditions and data quality rules
Analyst defines mapping specification
Estimate project scope based on impact analysis
Build, Test, and Deploy

Cut project throughput times by 90%

Double productivity
What Percentage of Projects Make it Through the First Time?
**Value Stream Map**

DOIT Corporation: Value Stream Map (AS-IS) for Change Request Process
Monday, December 20, 2010

**Notes:**
1. Lead Time includes 5 delay in customer notification
2. Lead Time could be reduced to 24 days with just process changes and using existing tools
3. Lead Time could be reduced to 3 days with a capital investment for automated testing

**Value Ratio:** Work Time / Lead Time = 0.5%

<table>
<thead>
<tr>
<th>Task</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Requirements Review</td>
<td>8.8 Days</td>
</tr>
<tr>
<td>Add CR To List</td>
<td></td>
</tr>
<tr>
<td>Semi-Weekly Review</td>
<td></td>
</tr>
<tr>
<td>Assign Resource</td>
<td></td>
</tr>
<tr>
<td>Integration Team Manager</td>
<td></td>
</tr>
<tr>
<td>Design Approval</td>
<td></td>
</tr>
<tr>
<td>Architecture Review Council</td>
<td></td>
</tr>
<tr>
<td>Test Scheduling</td>
<td></td>
</tr>
<tr>
<td>Test Team Manager</td>
<td></td>
</tr>
<tr>
<td>Test Execution</td>
<td></td>
</tr>
<tr>
<td>Production Deployment</td>
<td></td>
</tr>
<tr>
<td>Production Execution</td>
<td></td>
</tr>
<tr>
<td>Daily ETL Batch Run</td>
<td></td>
</tr>
</tbody>
</table>

**CR Approval**

- Change Management Board
- CR Approval & Schedule
- CR Approval Request

**Data Dictionary**

- To clarify requirements (13 days)
- Data Dictionary
- Approved Designs
- Design Docs & Schedule

**Bypass Council**

- for simple CR’s (26 days)
- Bypass Committee for simple changes (8 days)

**Automated Workflow/Tracking**

- (Cust satisfaction)
- Automated Workflow
- Automated Testing

**CR Request**

- Confirmation Request
- Telephone Tag
- CLAR LEK
- Change Request
- Telephone Tag
- Approval Request
- Change Request

**Test Results**

- Distribution

**Integration Team Manager**

- Approve Changes
- Forward CR Request To Developer

**Test Execution**

- Test Execution
- Data Warehouse Team
- Test Request

**CR Review Committee**

- Semi-Weekly Review
- Requirements Review

**Data Warehouse Team**

- Data Dictionary

**GMNA Applications Team (Irina)**

- CR's P1x12
  P2x5
  P3x14

**Development Team**

- Design Document
- Approved Designs
- Automated Workflow
- Automated Testing

**Production CR Submission**

- Data Warehouse Team
- CR Approval & Schedule
- CR Approval Request

**Testing Handoff**

- Test Execution
- Test Team
- Test Case Development

**Test Case Development**

- Test Team
- Test Execution

**Production Deployment**

- Infrastructure Team
- CR Approval & Schedule
- CR Approval Request
- CR Approval & Schedule

**Value Ratio**

- Work Time: 510 Minutes
- Lead Time: 75.6 Days
- Work Time: 0.3 Days

**NOTES:**
1. Lead Time includes 5 delay in customer notification
2. Lead Time could be reduced to 24 days with just process changes and using existing tools
3. Lead Time could be reduced to 3 days with a capital investment for automated testing
Agile Data Integration

Best Practices

As-Is Process

To-Be Process

Leverage test data management and automation wherever possible to increase test coverage and reduce errors
How many of you use agile development methodologies?

• Use Agile today
• Plan to use Agile within the next 12 months
• No plans to adopt Agile in the near future
Informatica’s self-service data integration doubles productivity by eliminating manual steps and empowering analysts to do more on their own. Analysts can define and validate source-to-target specifications in an intuitive browser-based tool without a data architect or DBA. On top of that, once the analyst creates the source-to-target specification, the mapping logic is automatically generated for a developer to deploy to production.”

*Sean Hickey, Manager Data Integration, T-Mobile*
Business-IT Collaboration

Fulfill Requests in Days Instead of Weeks/Months

1. Business requests new information be added to an existing BI report

2. Analyst works with Developer to first deploy as a SQL view or web service and then to a data warehouse table
Self Service Data Integration

*Empower Business Users to Provision Data for Reports*

1. Business needs to add data to a report from a data warehouse and external data source.

2. Business finds data, defines mapping specification, and automatically generates SQL view or web service for BI report.
Self Service Data Integration

*Empower Business Users to Perform Data Extracts*

1. **Business needs to extract data from a data warehouse**

2. **Business finds data, specifies criteria, and extracts data on their own**
Self-Service Data Integration
Empowering the Business While IT Retains Control

Accelerators
- "Migrate my Apps to SAP"
- "Upload customer data into Registry"
- "Add a Product Dimension to my DW"
- "Backup my App tables for HA/DR"
- "Mask data for Testing"

Business

- Analyst
- Specifications
- Business Terms

IT

- Architect
- Templates
- Mappings

Connectivity

- Cloud Computing
- Application
- Database
- Unstructured Data
- Partner Data
Key takeaways

1. Analysts want to be empowered with tools to accomplish tasks themselves. Analysts do not want to depend on IT for everything

2. Agile methodologies will gain more adoption in the data warehousing and BI space

3. Analysts want to collaborate with developers in the same environment sharing the same artifacts

4. Re-usability of data integration assets and implementation patterns will become very important for saving project costs

5. Integration of metadata management and business glossary into products will be key to improving productivity
Q & A

• For further questions, please contact:
  • Diby Malakar – Director Product Management
  • Email: dmalakar@informatica.com
  • Cell: 408-375-2570