Informatica 9.1 for Agile Data Integration
Best Practices Workshop
LEAN

1. Eliminate waste
2. Automate processes
3. Empower the team
4. Continuously improve
5. Build quality in
6. Plan for change
7. Optimize the whole

AGILE

1. Maximize business value
2. Empirical process control
3. Facilitate collaboration
4. Remove impediments
5. Frequent assessment
6. Adaptive response
7. Balance design & production
Agile Data Integration

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

GMNA Applications Team
(Irina)

Data Warehouse Team

CR Review Committee
Semi-Weekly Review

Development Team
Requirements Review

Marketing Team

Integration Team Manager

Assign Resource

Design Approval
Architectural Review Council

Test Scheduling
Test Team Manager

Design Document
Approved Design

Test Case Development
Test Team

Test Execution
Infrastructure Team

Production Deployment

Value Ratio: Work Time / Lead Time = 0.9%

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

Days/Weeks

Mins/Hrs

Days/Weeks

Mins/Hrs

Days/Weeks

Mins/Hrs

Days/Weeks

Mins/Hrs

1. Business & IT collaborate efficiently to fulfill requirements
2. Rapidly prototype and validate specifications & business rules
3. Leverage business glossary & metadata for search & impact analysis
4. Use data profiling continuously throughout design & development
5. Build data quality into the data integration process
6. Insulate applications from change through data virtualization
7. Minimize delivery time through flexible deployment options
8. Leverage test automation to increase coverage & reduce errors

To-Be Process

Hours

Mins

Hours

Mins

Mins
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
Business Request Scenario

**Fulfill Requests in Days Instead of Weeks/Months**

Agile Data Integration is ideally suited for Data Warehouse & BI projects because of the frequency of change

1. Business requests new information be added to an existing BI report
2. Analyst works with Developer to add a new data source to a data warehouse fact table
**As-Is Process**

1. **Find data sources and targets**
   - Business submits request for new information in report.
   - Analyst receives request from business.
   - Analyst requests clarification from business.

2. **Identify Data Source**
   - App Developer identifies table to meet requirements.
   - Analyst describes requirement to App Developer.

3. **Grant Data Access**
   - DBA grants analyst access to data source (e.g., table).

4. **Define Requirements**
   - Analyst creates requirements definition based on request.

5. **Preview Data**
   - Analyst previews data to confirm data sources.

---

**Complexity vs. # Data Sources**

- Empower analyst to find & preview data on their own.

---

**Establish a data governance framework to ensure confidence, integrity, transparency, & security.**

---

**Which data sources contain the information I need for the report?**

---

**Complexity**

- Days/Weeks
- Mins/Hrs

**Each data source in a different application requires another app developer to get involved.**

---

**Informatica 9**
To-Be Process

Find data sources and targets

Profile data sources for data quality

Which data sources contain the information I need for the report?

Data quality rules

Analyze defines mapping specification

Estimate project scope based on impact analysis

Build, Test, and Deploy

View business term and associated objects

Search for data sources and targets using business terms

Browse and navigate data lineage to find data sources and targets
Govern The Information Value Chain
Build A Solid Data Governance Foundation

Deliver A Single Version Of The Truth Of Information Through Master Data Management

Deliver Context Rich Information Through Metadata Management

Deliver Consistent, Correct, & Complete Information Through Pervasive Data Quality
As-Is Process

1. Find data sources and targets
2. Profile data sources for data quality
3. Identify join conditions and data quality rules
4. Mapping specification
5. Estimate project scope based on impact analysis
6. Build, Test, and Deploy

- **Create Data Dump**
  - **App Developer** creates data dump to spreadsheet or sandbox environment

- **Profile Data**
  - **Analyst** profiles in spreadsheet or creates & runs SQL scripts to profile data

- **Identify Data Quality Rules**
  - **Analyst** works with business to determine data quality rules to fix issues per requirements

Can I trust this data? Are there any data quality issues?

Need to either find another data source or go back to business to identify alternatives

Find data sources and targets

Browser-based tool to profile data without the help of a developer

Days/Weeks
Mins/Hrs

Days/Weeks
Mins/Hrs

Days/Weeks
Mins/Hrs

Days/Weeks
Mins/Hrs

Informatica 9
How do you profile data today?

- Dump data to spreadsheet and profile in spreadsheet
- Dump data to sandbox environment and use SQL scripts
- Use Informatica Profiling
- Other
To-Be Process

1. Find data sources and targets
2. Profile data sources for data quality
3. Identify joint conditions and data quality rules
4. Estimate project scope based on impact analysis
5. Build, Test, and Deploy

Can I trust this data? Are there any data quality issues?

Why is the Customer number not unique?

Identify duplicate records and uniqueness violations

Identify null violation issues (e.g. zip code)

Check column formats (e.g. date)

Drill down to see details and change in state

Informatica 9
How do we join these tables together and fix these data quality issues?

**Identify Join Conditions**
- **Developer** identifies & validates PK-FK relationships using ad-hoc approach

**Identify Join Transformations**
- Developer determines which lookups, filters and transformations to apply in order to meet join conditions

**Perform Join Analysis**
- Developer searches documentation and creates SQL scripts to profile tables

**Build E-R Model**
- Developer builds E-R model based on table relationships

**Request DQ Rules**
- Analyst requests data quality rules from developer to fix data quality issues per requirements

**Fix DQ Issues**
- Developer writes script to resolve data quality issue (which has probably been written before)

**Request DQ Rules**
- Analyst verifies data quality rules fix issues

**Profile data sources for data quality**

**Search common metadata repository for data quality rules already created**

**Complex DQ issues that are not easy to resolve may need to go back to the business**

**As-Is Process**

**Estimate project scope based on impact analysis**

**Build, Test, and Deploy**

<table>
<thead>
<tr>
<th>Days/Weeks</th>
<th>Mins/Hrs</th>
<th>Days/Weeks</th>
<th>Mins/Hrs</th>
<th>Days/Weeks</th>
<th>Mins/Hrs</th>
<th>Days/Weeks</th>
<th>Mins/Hrs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Informatica 9**
How do we join these tables together and fix these data quality issues?

Identify join conditions and data quality rules

Analyst defines mapping specification

Estimate project scope based on impact analysis

Build, Test, and Deploy

Infer primary and foreign keys

View PK-FK relationships

Confirm join condition will work, identify orphan rows, and find redundant values in other tables
As-Is Process

Find data sources and targets

Analyst adds sources and targets to mapping document

Add Sources & Targets

Analyst defines mapping specification

Identify Field Mappings

Analyst describes transformations between source fields and target fields based on requirements

Identify Field Mappings

Developer requests clarification from Analyst

Verify Field Mappings

Analyst verifies source-to-target field mappings meet requirements

Leverage common metadata repository and business glossary to understand objects and relationships

Create Mapping Document

Analyst uses spreadsheet to create mapping document

Build, Test, and Deploy

Estimate project scope based on impact analysis

Days/Weeks

Mins/Hrs

Days/Weeks

Mins/Hrs

Days/Weeks

Mins/Hrs

Days/Weeks

Mins/Hrs

Find data sources and targets

Which source attributes map to target attributes? What transformations and rules are required to map the sources to targets?
How many iterations between the Business and IT does it take to get specifications right?

- Less than 10
- 10 to 24
- 25 to 49
- 50 or more
To-Be Process

Find data sources and targets

Which source attributes map to target attributes? What transformations and rules are required to map the sources to targets?

Analyst defines mapping specification

Estimate project scope based on impact analysis

Build, Test, and Deploy

- Easily map data sources to both physical and virtual data targets
- Specify transformation logic using reusable expressions
- Include pre-built ETL and data quality rules as transformations in mapping specification
- Include transformation descriptions to instruct developer
- Define Specification DEMO
- Mapping Validation & Generation DEMO
Developer reviews code to determine what reports and objects are impacted.

Leverage metadata management with data lineage to perform impact analysis.

Search SCCS

Developer searches source-code control for reports or other objects impacted.

Code Review

Developer reviews code to determine what reports and objects are impacted.

List Impacted Artifacts

Developer compiles list of reports and artifacts that need to be retested as a result of the change.

Estimate QA Time

QA provides estimate on how long to retest reports and other affected objects.

Estimate Dev Time

Developer estimates level of effort to build and unit test objects.

What is the scope of this change? What other reports and artifacts are impacted?
To-Be Process

Find data sources and targets

To Be Processed

Profile data sources for data quality

Find data sources and targets

Identify join conditions and data quality rules

Profile data sources for data quality

Analyst defines mapping specification

Estimate project scope based on impact analysis

Build, Test, and Deploy

View upstream lineage

View downstream lineage

Lineage object

View downstream impact

View upstream impact

Lineage & Impact Analysis DEMO

What is the scope of this change? What other reports and artifacts are impacted?

View upstream lineage

View downstream lineage

Lineage object

View downstream impact

View upstream impact

Lineage & Impact Analysis DEMO
As-Is Process

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

Clarity Specification
Analyst clarifies source and target mapping

Specify Results
Analyst specifies expected results

System Test
Tests all affected objects and reports

Acceptance Test
Business user performs user acceptance testing

Deploy early and test often to speed up iterations

Use comparative profiling
Automatically compare actual results with expected results

Reuse DI & DQ rules in unified developer environment
Automatically generate mapping logic from specification

Analyst verifies that target data meets business requirements

Verify Requirements
Analyst verifies that target data meets business requirements

Unit Test
Developer performs unit testing

Test results against business requirements; go back to specification step

Days/Weeks Mins/Hrs Days/Weeks Mins/Hrs Days/Weeks Mins/Hrs Days/Weeks Mins/Hrs
How long does it take to deploy a change to the data warehouse affecting > 1 BI report?

- Less than 2 weeks
- 2 to 4 weeks
- 5 to 8 weeks
- More than 8 weeks
Data Architecture for Agile DI

Data Virtualization & Adaptive Data Services

- Support All Projects
- Flexible Deployment
- Implement All Policies
- Business-Friendly Data Abstraction
- Access All Data

Access
Quality
Retention
Privacy
Freshness

CUSTOMER
SUPPORT
PRODUCT
INVOICE

Mainframe
Databases
Unstructured and Semi-structured
Applications
Cloud
Social Media
NoSQL
To-Be Process

1. Find data sources and targets
2. Profile data sources for data quality
3. Identify join conditions and data quality rules
4. Analyst defines mapping specification
5. Estimate project scope based on impact analysis
6. Build, Test, and Deploy

- Mapping logic automatically generated from analyst defined specification
- Extend mapping to dynamically mask sensitive data
- Extend mapping to consume web services

Informatica Developer

DEMO
Consume WS
Dynamic Masking
DEMO
To-Be Process

1. Find data sources and targets
2. Profile data sources for data quality
3. Set data quality rules
4. Analyst defines mapping specification
5. Estimate project scope based on impact analysis
6. Build, Test, and Deploy

- Deploy as a SQL view
- Deploy as a web service
- Deploy to PowerCenter
To-Be Process

Find data sources and targets → Profile data sources for quality → Create data quality rules → Datastax defines mapping specification → Estimate project scope based on impact analysis → Build, Test, and Deploy

What is the fastest way to deploy this change to production?

Compare profiling statistics during development to ensure data quality.

Validate actual target data against expected data.

Actual Results → Expected Results

Data Validation

DEMO

Comparative Profiling

DEMO

INFORMATICA 9
Agile Data Integration

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**

**Project Request** → **Define Requirements** → **Analyze & Design** → **Build** → **Test** → **Deploy** → **Support & Maintain**
Agile Data Integration

1. Analyst quickly finds data sources by searching business glossary and data lineage
2. Analyst profiles data sources to identify data quality issues sharing results with developer through common metadata repository
3. Developer rapidly infers PK-FK relationships, validates join conditions, and finds pre-built data quality rules
4. Analyst defines and validates mapping specification that leverage pre-built ETL and data quality rules
5. Developer extends automatically generated mapping logic to include other transformations and web services
6. Developer rapidly deploys SQL view, web service, and batch ETL process
7. Developer and QA test using comparative profiling and data validation automation
How much faster could you deliver projects using the Agile Data Integration methods described today?

- 2X
- 3X
- 5X
- 10X
Agile Data Integration

Other Project Types

- MDM
- Application Integration
- Application Modernization
- Data Migration
Agile Data Integration
How to Get Started

- Identify a Sponsor and Change Agent
- Identify Early Win Project
- Know Your Customer
- Create As-Is Value Stream
  - Look at one of your last projects
- Create To-Be Value Stream
  - Leverage Informatica Sales and Professional Services
Q & A