REAL TIME DATA QUALITY
PRESENTED BY: BRAD HATHAWAY
REAL TIME DATA QUALITY

WHAT ARE THE REASONS?
WHAT ARE THE REASONS?

• Helps to stop or slow down the proliferation of bad data in the company – and create a «good data area»

• Stops bad quality data from being inserted in company systems – at the origin – Data Quality Firewall

• Creates another «data point» for Data Governance processes
REAL TIME DATA QUALITY

EXPOSING DATA QUALITY SERVICES IN REAL-TIME APPLICATIONS

• Ad-hoc checks for real-time error verification
  • Duplicates: insertion of the same people, patients, companies, products, …
• Controlling consistency between fields being inserted
  • Fiscal Codes with personal details
  • Email addresses
  • Dates, types
  • City, province, region
• Material Classification
• Medical prescription search
• Address enrichment
  • Postal codes
  • Latitude and longitude
• Auto-completion

Centralize DQ checks in a hub or Data Quality Firewall
ARCHITECTURE

DATA MANAGEMENT STUDIO & DATA MANAGEMENT SERVER

End User

Enterprise Service Bus

DM Studio

Data Management Server

Authentication Server

LDAP / AD

Authentication Providers

Enterprise Data

End User
REAL TIME DATA QUALITY

TECHNOLOGY

- Uses web services to support interoperability between applications
  - Real-Time Data Services: services to manage and manipulate data
  - Real-Time Process Services: process orchestration services
DATA MANAGEMENT STUDIO

REAL TIME DATA SERVICE – DATA JOB

- External data provider
  - Input fields:
    - Address
    - City
    - State
    - Zip
- Standardization
  - City and State and produce flags
- Lookup
  - Verify address and generate new columns:
    - Address1
    - Locality
    - PostalCode
    - AdministrativeArea
- Standardization
  - Standardize the AdministrativeArea to 2 characters
- Expression
  - Compare true flags for state/city/zip
- Field layout
  - Prepare fields for output
- Job specific data
  - Used to simulate input data to test job
• Multithread and multi-platform server
• Batch Jobs
• Profile Jobs
• Real Time Data Services
• Real Time Process Services
A web service call is activated via an XML request transmitted using the http protocol to a specific port where Data Management Server is active.
Data Management Web Services MDM

A couple simple examples that illustrate calling a Process Jobs on the Data Management Server (required either local or remote address). This example leverages the JSON/XML bi-directional conversion capabilities of the Toth Service Adapter to simplify the JavaScript code in the web page.

<table>
<thead>
<tr>
<th>Company</th>
<th>Bernasconi Elettricità SRL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Person</td>
<td>Angelo Miglio</td>
</tr>
<tr>
<td>Address</td>
<td>Via Fogazzaro, 3</td>
</tr>
<tr>
<td>City</td>
<td>Varese</td>
</tr>
<tr>
<td>State</td>
<td>VA</td>
</tr>
<tr>
<td>Zip</td>
<td>21100</td>
</tr>
</tbody>
</table>

Validate web form data using Data Quality

Please fill out the form and click the "Standardize/Validate" button.

This example calls the "examples/standardization_validation.dof" job on the Data Management Server to first standardize, then validate the form data (i.e., address/city/state/zip). The corrected (if needed) data is then returned and the original form data is replaced.

Standardize/Validate

JavaScript DMP WSDL
## Data Management Web Services MDM

A couple simple examples that illustrate calling a Process Jobs on the Data Management Server (required either local or remote address). This example leverages the JSON/XML bi-directional conversion capabilities of the Toth Service Adapter to simplify the JavaScript code in the web page.

### Standardization/Validation

Validate web form data using Data Quality.

Please fill out the form and click the "Standardize/Validate" button.

This example calls the "examples/standardization_validation.do" job on the Data Management Server to first standardize, then validate the form data (i.e. address/city/state/zip). The corrected (if needed) data is then returned and the original form data is replaced.

<table>
<thead>
<tr>
<th>Company Name</th>
<th>Bemasconi Electricità SRL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Person Name</td>
<td>Angelo Miglio</td>
</tr>
<tr>
<td>Address</td>
<td>Via Fogazzaro, 3</td>
</tr>
<tr>
<td>City</td>
<td>Varese</td>
</tr>
<tr>
<td>State</td>
<td>VA</td>
</tr>
<tr>
<td>Zip</td>
<td>21100</td>
</tr>
</tbody>
</table>

### Clear Form Fields

Click to clear the form fields.

### JavaScript DMP WSDL

- **Determine Gender**
- **Search MDM Hub**
- **RESTful MDM**
A couple simple examples that illustrate calling a Process Jobs on the Data Management Server (required either local or remote address). This example leverages the JSON/XML bi-directional conversion capabilities of the Toth Service Adapter to simplify the JavaScript code in the web page.

This example illustrates how MDM's generic WSDL can be simplified - the request only has 2 variables (strongly typed - e.g. int and string) and the response is also strongly-typed and "flattened" (i.e. not as nested/hierarchical as the original). The server-side "transformation" code demonstrates both REST (JSON response) and simplified XML support for applications that cannot consume complex hierarchical XML requests/responses. This specific example was created to prototype a MDM/Decision Services integration (AKA RDM). The Query Survivor Person is only method/operation currently supported.

Mouse-over the example link below to see more detailed info about it in its tooltip.

Request URL: http://localhost:7070/qmdm/simple/querySurvivor?sourceName=Rossi%20Marcello&sourceMaster=Y

Browser JavaScript Server JavaScript Simplified qMDM WSDL Original DMP WSDL
Exposed services are “consumed” using the following standards:
- SOA – Service Oriented Architecture
- Web Services and WSDL
- HTTP client to make web/REST/web service calls
- JSON
MDM Services can be called via batch or in real-time

- Create new records
- Update existing records
- Retire Records in the MDM Hub
- Build Relationships
- Start a Workflow
- Send records to Data Remediation

SAS Federation Server

RESTful Data Services
SAS Data Quality Accelerator for Teradata
- Over 500% faster than using traditional data quality techniques
- Parsing, Extraction Pattern, Identification, Gender, Standardization, Matching
- Rules for cleansing can be personalized in the Quality Knowledge Base (QKB)

SAS Data Quality Accelerator for Hadoop
- Release in second half 2014
- Parsing, Extraction Pattern, Identification, Gender, Standardization, Matching
- Data Quality as an «embedded process» for big data
REAL TIME DATA QUALITY IN ACTION

Demo
SAS DataFlux Data Management Server

SAS DataFlux Data Management Studio

SAS Data Quality Accelerator for Teradata

SAS Federation Server
THANK YOU!