Navigating Big Data with SAP Data Services

Paul Medaille, Director, Solutions Management, EIM, SAP
Ina Felsheim, Director, Solutions Management, EIM, SAP
Upcoming SAP EIM Webinars:

- **Gain Data Quality Insights with SAP Information Steward**
  - January 29, 2014 1:00 P.M. EST/10:00 A.M. PST
  - [www.sap.com/events](http://www.sap.com/events)

- **Achieve Data Accuracy with SAP Master Data Governance**
  - February 12, 2014 1:00 P.M. EST/10:00 A.M. PST
  - [www.sap.com/events](http://www.sap.com/events)
Agenda

Big Data and Small Data
- How are they different? How are they similar?

Introducing SAP Data Services
- Capabilities
  - Data Integration
  - Data Quality Management
  - Text Data Processing
  - Point-of-Entry Firewall

Value

Why Data Services?
These are the Top Business Issues Where You Need Better Information Management
Solutions to Manage the Complete Information Lifecycle
“You buy new companies, you change your business, you change your systems, you hire new people…but the one thing that remains through it all is your data. “

--customer just starting an EIM initiative
Big Data and Small Data
Big Data Offers New Opportunities
Gain real-time insight from large volumes of a variety of data

- **Volume**: Large volumes (petabyte is normal)
- **Variety**: Multiple data formats
- **Velocity**: Fast collection, processing and consumption
- **Value**: Competitive differentiator for business

**Data Volume**
- 1 Terabyte = 1024 Gigabytes
- 1 Petabyte = 1024 Terabytes
- 1 Exabyte = 1024 Petabytes
- 1 Zettabyte = 1024 Exabytes

7.9 Zettabytes

© 2014 SAP AG or an SAP affiliate company. All rights reserved.
Volume Growth of Big Data
Terabytes to petabytes and more

According to TDWI, by 2015:

Smaller data sets will become less common (2% - 12%)

31% managing 10 – 100 TB

Large volume data sets will become more common (20% – 40%)

Big Data versus Small Data
Discover more patterns & anomalies in large volumes of data

<table>
<thead>
<tr>
<th>TOPCATEGORY_NAME</th>
<th>TITLE</th>
<th>UNIT_PRICE</th>
<th>ASIN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electronics</td>
<td>Rabies 52888146...</td>
<td>479.9</td>
<td>052883469</td>
</tr>
<tr>
<td>Toys &amp; Games</td>
<td>Ball of Whacks</td>
<td>23.37</td>
<td>091112018</td>
</tr>
<tr>
<td>Toys &amp; Games</td>
<td>The Original Wiz...</td>
<td>7.82</td>
<td>0913866687</td>
</tr>
<tr>
<td>Toys &amp; Games</td>
<td>Ticket To Ride</td>
<td>38.0</td>
<td>0975277324</td>
</tr>
<tr>
<td>Electronics</td>
<td>Pecket Cover to ...</td>
<td>399.95</td>
<td>0979526191</td>
</tr>
<tr>
<td>Toys &amp; Games</td>
<td>Malachy Color</td>
<td>17.5</td>
<td>1596344732</td>
</tr>
<tr>
<td>Toys &amp; Games</td>
<td>Wizard Card Gam...</td>
<td>10.29</td>
<td>1578810939</td>
</tr>
<tr>
<td>Toys &amp; Games</td>
<td>Citadels</td>
<td>19.33</td>
<td>1589940329</td>
</tr>
<tr>
<td>Toys &amp; Games</td>
<td>Battlerstar Galactica</td>
<td>38.73</td>
<td>1589946067</td>
</tr>
<tr>
<td>Toys &amp; Games</td>
<td>Tomorrow The M...</td>
<td>45.95</td>
<td>1589014632</td>
</tr>
</tbody>
</table>

Smaller Structured Data Sets
- Run queries to search for insight
- Know what you’re looking for

Large Volumes of Unstructured Data
- Mine data
- Detect nuggets of relevant data while disregarding unimportant data
Use Case: Real Time Analysis

Traditional Analysis
- Transactional data
- Extract, transform, load from op systems
- Stored in warehouse or data mart
- Request-based query (report/dashboard)
- Analysis after data is consolidated
- Make decisions and take actions based on what has already happened

Real Time Analysis
- Operational, transactional, event data
- Rapid bulk loads or event processing
- Store in-memory or process in real-time
- Ad-hoc or continuous query
- Analysis during processing
- Make decisions and take actions based on what is happening now
SAP Data Services

- Overview
- Capabilities
  - Data Integration
  - Data Quality Management
  - Text Data Processing
  - Point-of-Entry Firewall
SAP Data Services:
One innovative solution with multiple integrated data management capabilities

- One Solution that Provides
  - Data Integration
  - Data Quality
  - Text Data Processing

- One server to execute all capabilities
- One design environment to manage all development
- One administration console to monitor all functions

Lower TCO  Efficient Development  Simplified Landscape
SAP Data Services turns data into real value

Improve efficiency of business process

40% master data quality improvement

Uniting processes, business units, and users

Vision to have data management cycle baked into everyday business

Enable efficient reporting to management

60% reduction in costs by reducing IT involvement

Boost efficiency and eliminate rework

10-50% more allocation of staff to strategic tasks

Ensure compliance with regulations

1st of 14 regional banks to respond to new reporting regulations

Unified view of customers

Ensured reliable reporting by managing multiple sources from M&A activity

Leveraging data to differentiate in call center

5% reduction of overhead while increasing staff productivity

Deeper customer relationships

360 customer view and comprehensive experience

Manage data quality on a continuum

Ensure consistent customer data is maintained across all SAP and Non-SAP systems

Achieving speed and reliability

3 hours loading time for data that once took 3 days
Data Services for ETL
SAP Data Services:
Serving SAP and non-SAP environments

SAP CRM, SAP ERP, SAP NetWeaver, SAP BusinessOne, SuccessFactors

Sybase, HANA, Oracle, SQL, DB2, HADOOP

PeopleSoft, Oracle Apps, Siebel, SFDC, etc

Files, XML, HTML, Mainframe, Excel, text, etc.

SAP ERP, CRM, MDG,...and non-SAP systems

BAPIs / IDocs / Web Services

Data Profiling

Data Services Engine

Data Cleansing, Enrichment

Data Validation

Text Data Processing

Impact Analysis

Data Linéage

Data Migration, Synchronization, Real-Time, ...

Query, Reporting, Analysis, and Dashboards

HANA, BW, Sybase

© 2014 SAP AG or an SAP affiliate company. All rights reserved.
Wide Connectivity to 40+ Systems via Drag-n-Drop Setup

**SAP Data Services** has a long history of providing broad connectivity to many different environments and continues with emerging Big Data sources.

<table>
<thead>
<tr>
<th>Databases</th>
<th>Files/Transport</th>
<th>Applications</th>
<th>Mainframe (with partner)</th>
<th>Unstructured Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>HANA – Tables</td>
<td>Text delimited</td>
<td>SAP Bus Suite – Extractors</td>
<td>ADABAS</td>
<td>Extracts context from any text or binary file types (docs, pdfs, ppts...)</td>
</tr>
<tr>
<td></td>
<td>Text fixed width</td>
<td>– ABAP – BAPI – Idoc</td>
<td>ISAM</td>
<td>Understands context in 31 languages</td>
</tr>
<tr>
<td></td>
<td>EBCDIC</td>
<td>SAP NW BW – SuccessFactors</td>
<td>VSAM</td>
<td></td>
</tr>
<tr>
<td>Oracle</td>
<td>XML</td>
<td>Oracle applications</td>
<td>Enscribe</td>
<td></td>
</tr>
<tr>
<td>DB2</td>
<td>Cobol</td>
<td></td>
<td>IMS/DB</td>
<td></td>
</tr>
<tr>
<td>Sybase ASE</td>
<td>Excel</td>
<td></td>
<td>RMS</td>
<td></td>
</tr>
<tr>
<td>Sybase IQ</td>
<td>HTTP</td>
<td></td>
<td>Both direct and changed data</td>
<td></td>
</tr>
<tr>
<td>SQL Server</td>
<td>JMS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Informix</td>
<td>Web Services</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MySQL</td>
<td>HADOOP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teradata</td>
<td>VCF - Variant Call</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ODBC</td>
<td>Format for storing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Netezza</td>
<td>gene sequence</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>variations</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**New Adapter SDK** simplifies building connectivity to any applications and Big Data sources.
Quick Guide on Creating Data Services Jobs

Normal/Best practice for flows to be Left-Right

drag-n-drop tables, files, transforms from Local Object Library onto Work Canvas

Double-click to drill into object/transform. (All objects/transforms are movable)

click-n-drag from source point to target point to create link

click-n-drop objects from Palette onto Work Canvas
Data Quality Management
How does Data Quality work?

INPUT

Bob oldstead
175 Riviington Ave suite 2
Manhattan, new yourk
10002
How does Data Quality work?

INPUT

Bob oldstead
175 Riviington Ave suite 2
Manhattan, new yourk
10002

1
PARSE
How does Data Quality work?

INPUT

Bob Oldstead
175 Riviington Ave suite 2
Manhattan, New York
10002

PARSE

STANDARIZE
How does Data Quality work?

INPUT

Bob oldstead
175 Rivington Ave suite 2
Manhattan, new york
10002

1. Parse

First Name: Bob
Last Name: Oldstead
AddressL1: 175 Rivington Ave
AddressL2: Suite 2
City: Manhattan
State: New York
Zip Code: 10002

2. Standardize

3. Correct
How does Data Quality work?

INPUT

Bob oldstead
175 Rivington Ave suite 2
Manhattan, new yourk
10002

1. PARSE

First Name: Bob
Last Name: Oldstead
AddressL1: 175 Rivington Ave
AddressL2: Suite 2
City: Manhattan
State: New York
Zip Code: 10002

2. STANDARDIZE

3. CORRECT
How does Data Quality work?

INPUT

Bob Oldstead
175 Rivington Ave suite 2
Manhattan, New York
10002

PARSE

Standardize

First Name: Bob
Last Name: Oldstead
AddressL1: 175 Rivington Ave
AddressL2: Suite 2
City: Manhattan
State: New York
Zip Code: 10002

CORRECT
How does Data Quality work?

1. **PARSE**
   - Input: Bob oldstead
     - 175 Rivington Ave suite 2
     - Manhattan, new york
     - 10002

2. **STANDARDIZE**
   - First Name: Bob
   - Last Name: Oldstead
   - AddressL1: 175 Rivington Ave
   - AddressL2: Suite 2
   - City: Manhattan
   - State: New York
   - Zip Code: 10002

3. **CORRECT**
   - Robert E. Oldstead
     - Manhattan, NY
     - 10002
     - robo@tcabuilders.com
     - 847 442-5555

4. **MATCH**

© 2014 SAP AG or an SAP affiliate company. All rights reserved.
How does Data Quality work?

1. **PARSE**
   - First Name: Bob
   - Last Name: Oldstead
   - AddressL1: 175 Rivington Ave
   - AddressL2: Suite 2
   - City: Manhattan
   - State: New York
   - Zip Code: 10002

2. **STANDARDIZE**
   - First Name: Bob
   - Last Name: Oldstead
   - AddressL1: 175 Rivington Ave
   - AddressL2: Suite 2
   - City: Manhattan
   - State: New York
   - Zip Code: 10002

3. **CORRECT**
   - First Name: Robert
     - Middle Name: E
   - Last Name: Oldstead
   - AddressL1: 175 Rivington Ave
   - AddressL2: Suite 2
   - City: Manhattan
   - State: New York
   - Zip Code: 10002
   - Phone: (847) 442-5555
   - Email: robo@tcabuilders.com

4. **MATCH**

5. **CONSOLIDATE**

Robert E. Oldstead
Manhatten, NY
10002
robo@tcabuilders.com
847 442-5555
How does Data Quality work?

INPUT

Bob Oldstead
175 Rivington Ave suite 2
Manhattan, New York
10002

PARSE

First Name: Bob
Last Name: Oldstead
AddressL1: 175 Rivington Ave
AddressL2: Suite 2
City: Manhattan
State: New York
Zip Code: 10002

STANDARDIZE

CORRECT

MATCH

Robert E. Oldstead
Manhattan, NY
10002
robo@tcabuilders.com
847 442-5555

CONSOLIDATE

First Name: Robert
Middle Name: E
Last Name: Oldstead
AddressL1: 175 Rivington Ave
AddressL2: Suite 2
City: Manhattan
State: New York
Zip Code: 10002-2517
Phone: (847) 442-5555
Email: robo@tcabuilders.com

ENHANCE

First Name: Robert
Middle Name: E
Last Name: Oldstead
AddressL1: 175 Rivington Ave
AddressL2: Suite 2
City: Manhattan
State: New York
Zip Code: 10002
Phone: (847) 442-5555
Email: robo@tcabuilders.com
Longitude: 40.7325525
Latitude: -74.004970

How does Data Quality work?
Global Address Cleansing Coverage

Coverage levels

- Basic Address Cleansing: All-World Directory includes Last Line Assignment and Address Line Parsing and Standardization
- Enhanced Address Cleansing: 36+ country-specific Address Directories, Address Line Assignment
## Improve Product and Material Data

### Challenges:
- Unstructured information written in multiple ways prevents high matching rates
- Data does not look professional

### Benefits:
- Splitting the unstructured information into parsed and discrete fields creates improved match results
- Unified and standardized information enables consistent representation of the same information

<table>
<thead>
<tr>
<th>Input Data</th>
<th>Parsed Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glove ultra grip profit 2.3 large black synthetic leather elastic with Velcro Mechanix Wear</td>
<td><strong>Product Category</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Size</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Material</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Trademark</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Cuff Style</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Palm Type</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Color</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Vendor</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Standard Description</strong></td>
</tr>
</tbody>
</table>
Data Quality Firewall
Point-of-Entry Firewall:
How does bad data get into your systems?

Information Governance

- Employee Data Entry
- Purchased or Rented External Data
- Customer Self-Service
- IT Application Updates
- Data Migration Projects

A → B
Point-of-Entry Firewall: How does bad data get into your systems?
What is a Data Quality Firewall?

A data quality firewall cleanses and corrects your data before it gets into your applications systems, reducing the need for data remediation after the fact.

By correcting data at the point of entry, you save the costs associated with poor quality data affecting crucial business processes.
Text Data Processing
Text Data Processing – What is it?

Unstructured Text

1. Extract meaning
2. Transform into structured data for analysis
3. Cleanse and match

Once structured it can be…

- Integrated
- Queried
- Analyzed
- Visualized
- Reported against

...And vital information can be automatically extracted!

- I <STRONGPOSITIVESENTIMENT>
  love </SPS><TOPIC> your product </TOPIC>.
2012 Toyota Highlander, 1G1JF27W8GJ178227. Policy #456-788-99. $500 deductible. Insured are William Thomas Snow and Christine W. Snow. On 1/09/13, received via text message "In accident! Need tow truck." and "Corner of Winter & Cross Rd". Dr. Snow slid through a stop on the snowy road and hit the stop sign. Significant damage to front of the car. He was unhappy at tow truck response time. Contact wtsnow@medicalcenter.com or 651.772.1234. Advance Appraisals Inc. to come out at 8:30am Monday to make further assessment.

Who, What, Where, When, How much, as well as Feelings
Value
So the question is…
Will your business struggle with large amounts of data?

Fact 4: TDWI survey results for Best Practices report on Big Data Analytics
What if?

...You could improve
✓ operations
✓ customer satisfaction
✓ sales and services

...By integrating
✓ ALL data
✓ ANY scale
✓ ANY velocity
...And combine it with unstructured content

- Web logs
- Call centers
- Social networks
- Data streams
- Bar-codes
- And sensors
Because the reality is..
To compete, you need to manage **all types of data**

Fact 4: Gartner, Information Governance: 12 Things to Do in 2012, Published: 27 January 2012
...And by extracting useful intelligence you get unprecedented contextual insight from big data
…Which requires that the quality of the data be fit for use

Is the Data...
- Consistent
- Accurate
- Complete
Because if the quality of the data is bad…

Poor data quality is a primary reason for 40% of all business initiatives failing to achieve their targeted benefits.

Is your company investing enough for information governance success?

Through 2016, spending on governing information must increase to five times the current level to be successful.
Does your company have the right technologies for proper information governance?

Governance, though focused on people and process, requires specific technology to support and enforce its practices — that is, to ensure stewardship.
Why Data Services?
What Are My Options?

- **Hand-code each instance of data movement, quality, and point of entry firewall.**
  - Pros: Reuse utilities
  - Cons: Manageability, speed, time, and knowledge transfer

- **Use an open-source tool for EIM**
  - Pros: Cost
  - Cons: Hidden costs, enterprise features and connectivity, integrated holistic suite, tuned for IT developers

- **Use a different enterprise EIM tool**
  - **Gasp**
  - Pros: You may have some existing development; Political bias
  - Cons: Deep connectivity to SAP Business Suite and SAP Analytics; Ease of use

- **Use SAP Data Services**
  - Pros: Deep SAP integrations; Ease of use; Integrated platform from Architect to Archive
  - Cons: Internal education at your company
SAP Data Services Solutions: Proven and Trusted
Powering the world’s leading global organizations

- 7,500+ customers worldwide
- 800+ new customers in 2012
- 90% customer satisfaction rating

#2 market share for data quality & data integration
Leader
Data Integration & Data Quality
Gartner Magic Quads 2013
Leading the EIM Market Revolution

First and Only Data Services Platform to Include Text Analytics

- Unlock hidden insights from unstructured content customer sentiment for competitive advantage and improved customer responsiveness
- Together with Information Steward, enables seamless Data Integration, Data Quality, Text Analytics, Data Profiling, and Metadata Management

Innovative Information Steward Offering for Information Governance

- Data profiling, data quality score/monitor, metadata management, business glossary, and business-user driven UI for creating cleansing rules

First Big Data Innovation with In-Memory Appliance (HANA)

- Data Services, Replication Server are the primary data integration solutions to load HANA
Why SAP?
The best choice for mastering big data

- **Flexible, open real-time data platform** that enables to transact and analyze data simultaneously to enable real-time business

- **Exceptional operational and analytical processing performance** with SAP HANA and SAP Sybase IQ

- **Efficient planning** of your Big Data environment with SAP Sybase PowerDesigner

- **Complete and accurate information** with SAP Data Services and SAP Information Steward

- **Powerful predictive analysis** with support for R language & SAP BusinessObjects Predictive Analysis

- **Seamless user experience of Big Data analysis** with SAP BusinessObjects Business Intelligence tools
Enterprise Information Management with SAP

• Understand the big picture of SAP’s enterprise information management offerings
• Explore step-by-step instructions for working with SAP Data Services
• Learn how to perform the most important tasks in SAP Information Steward, SAP NetWeaver Information Lifecycle Management, SAP Master Data Governance, and more
• All royalties donated to Doctors Without Borders

Reviews are in!

READ THIS ITS EXCELLENT ★★★★★
Posted by Anonymous on Dec 16th, 2012
From start to end this book gives you complete detail information about EIM

Very good book about EIM and Data Services ★★★★★
Posted by Anonymous on Nov 26th, 2012
This book has all information about EIM, manage data quality, migration process and a good overview about SAP HANA.
SAP Solutions for Information Management e-Book
Available on iTunes and as a PDF file

Turn Information into a Strategic Asset
SAP® Solutions for Information Management

Master Data Management
Consolidate, cleanse, and synchronize your data

Leverage complete and trustworthy data to support your operational, analytical, and data compliance initiatives.

SAP® master data management (MDM) solutions address both enterprise-wide and application-centric processing — and can help you improve processes, decision-making, and governance with a single, accurate version of the truth. This means you can:
- Streamline procurement processes with a centralized and accurate product catalog
- Enhance marketing and sales with complete and accurate customer information
- Improve quarterly reports with reliable financial master data from creation to delivery
- Ensure global data synchronization by standardizing supply chain data with trading partners
- Optimize supplier spend reporting with a consolidated view of material and supplier master data

Learn more:
- Report: SAP master data governance: The MDM Institute
- Report: Next generation master data management (TDW)
- Customer story: Data drives agility at Colgate-Palmolive

Product Demo: SAP master data governance: Centralize and consolidate supplier master data

Product Demo: SAP master data governance: Automate and control financial master data changes while supporting regulatory compliance

Product Demo: SAP master data governance: Empowering business users to make material data changes while upholding data standards

Solution Brief: SAP master data governance for:
- Customer
- Material
- Supplier

Customer Testimonials
Real business issues. Real business benefits.

See how SAP® partners with leading companies across the world for Information Management. Here are some examples of why these companies run SAP® Reap the benefits, understand their objectives, and learn about the measurable benefits they have gained with SAP.

Learn more:
- Video: Almandor N.V.: Improving data quality and reducing IT costs by 60%
- Video: Suncor Energy: Monitoring and expanding competitive advantage
- Video: Reckitt Benckiser: Enabling sustainable growth
- Video: Vodafone: Reducing complexity and improving customer experience
- Video: The Food and Drug Administration: Enhancing data quality and compliance with SAP<br>

Harness the Big Data Opportunity
Integrate data from a variety of sources, gain insight from new types of data, and have confidence in the quality of information

With the volume, variety, and velocity of structured and unstructured data growing exponentially, how can you maximize the business value of this big data flow across your organization?

SAP® solutions for information management (BMI) can help you gain a complete and accurate view of information by integrating and improving data at any scale from any data source. Benefits include:
- Improve decision-making and streamline business processes with access to all relevant information
- Increase productivity and enable more confident decision making with accurate and reliable information
- Enable business transparency and efficiency by understanding enterprise data models to provide a 360-degree view of the enterprise and customers

Learn more:
- Video: What is Big Data? How can it be defined by Tino Brand, SAP Executive and Adam Lippincott, SAP UK & Ireland
- Video: How can you benefit from Big Data? Find tips on how to improve management with Big Data to benefit significantly from Big Data technologies
- Video: Marketing Made Easier with Information Management
- Video: Unlock Big Data with SAP HANA and SAP Data Services
- Video: Customer story: University of Kentucky: Connecting mobile students with a new revenue platform from SAP
- Video: Customer story: Minera Marketing Solutions: Achieving a bigger picture with SAP by enhancing marketing automation with SAP Sybase IQ, SAP HANA, PowerDesigner, and SAP Data Services
- White paper: The Value of Big Data: Using big data to acquire and discover the value in data for accurate analytics for third parties
- White paper: Big Data Comes of Age: Shifting to a New Era of Customer Experience Management (CEM)
- Cost Calculator: How much does it cost to store data costing your business

© 2014 SAP AG or an SAP affiliate company. All rights reserved.
Important Links
Information for EIM, data migration, and information governance

- Where's the Good Stuff: A Comprehensive Walk-through of EIM Resources:
  http://scn.sap.com/docs/DOC-48657
Thanks!

Ina Felsheim, SAP
EIM Solution Management
Ina.Felsheim@sap.com
Twitter: @InaSAP

Paul Medaille, SAP
EIM Solution Management
Paul.Medaille@sap.com