SAP BusinessObjects
Data Services
Technical Overview
1. Data Services Technical Architecture
2. Developing in Data Services
3. Data Quality Management
4. Deploying in Data Services
5. Deploying for Performance
6. Managing with Data Services
7. Maintenance with Data Services
Are you spending too much resources learning different tools?

Separate products create the need for redundant investments in training, software management, and life-cycle support.
SAP BusinessObjects Data Services

Data Services is the first single solution combining data integration and data quality.

Data Integrator
- Development User Interface
- Metadata Repository
- Runtime Architecture
- Administration and Connectors

Data Quality
- Development User Interface
- Metadata Repository
- Runtime Architecture
- Administration and Connectors

Data Services
- One Development User Interface
- One Metadata Repository
- One Runtime Architecture
- One Administration Environment
- One Set of Connectors

Access
Transform
Improve
Deliver
Introducing SAP BusinessObjects Data Services

First single, enterprise-class data integration and quality application

- One tool for enterprise-class data integration and data quality
- One intuitive development user interface (UI)
- One administration environment
- One runtime architecture
SAP BusinessObjects Data Services
Gold Award Winner – 2008 Product of the Year

Source: SearchDataManagement.com, April 2009
What is Data Services?

Technology with capabilities to

- Explore (profile), Transform, and Move data anywhere at any frequency
- IMPROVE data (parse, cleanse, enhance, match, consolidate)
- Deliver auditable information
- Maximize developer productivity
- Deliver extreme scalability
  - On single servers
  - On multiple servers (Grid computing)

Benefits:

- Database / Application agnostic
- Total graphical environment maximizes developer productivity with little or no coding
- Investment protected with extreme scalability possibility
Use Case: Data Integrator
Integrate Heterogeneous Data

Disparate sources

- CRM
- RDBMS
- Mainframes
- Excel
- Web
- ERP
- Emails
- Files (Flat, XML)

One or more Targets

- Business Intelligence
- SAP BW
- SAP R/3
- CRM
- Data Mart
- Data Warehouse

Mr. Richard Tan Hock Guan
Future Electornics
2 Shenton Way
SGX Centre 1, #07-01, S068804

Tan, Hok Guan
Future Elect. Co.
7th Floor, SGX Center
2A Shenton Way, #01
Singapore

Hock Guan Tan
Future & Electronics Co.
Shenton House
3 Shenton Way, #01-07
Singapore

Original, Raw data

Data Integrator

Extracts & Transforms data

Loads properly formatted and structured data into Target/s

Richard Tan Hock Guan
Future Electornics
Singapore 068804

Tan Hock Guan
Future Elect. Co.
Singapore

Tan Hock Guan
Future & Electronics Co.
Singapore
Use Case: Data Quality Management
Scheduled Cleansing/Matching

Uncleaned, disparate data

Mr. Richard Tan Hock Guan
Future Electronics
2 Shenton Way
SGX Centre 1, #07-01, S068804

Tan, Hok Guan
Future Elect. Co.
7th Floor, SGX Center
2A Shenton Way, #01
Singapore

Hock Guan Tan
Future & Electronics Co.
Shenton House
3 Shenton Way, #01-07
Singapore

Cleansed data

Richard Tan Hock Guan
Future Electronics
2 Shenton Way
#07-01 SGX Centre 1
Singapore 068804

Tan Hock Guan
Future Elect.
2 Shenton Way
#07-01 SGX Centre 1
Singapore 068804

Tan Hock Guan
Future & Electronics Co.
3 Shenton Way
#07-01 Shenton House
Singapore 068805

De-duped data

Richard Tan Hock Guan
Future Electronics
2 Shenton Way
#07-01 SGX Centre 1
Singapore 068804

Tan Hock Guan
Future & Electronics Co.
3 Shenton Way
#07-01 Shenton House
Singapore 068805

Data Quality Management

Using Dictionaries + Directories Rules

Original, Raw data

Standardized, Corrected data

Matched, De-Duplicated data

Apply “Fuzzy” match techniques
Data Services Technical Architecture
Client-based (for job design) and mostly Web-based (for all other tasks)

Servers run on Windows and UNIX (including Linux), while Client runs only on Windows.
Data Services Physical Architecture

**Client**
- DI Designer (Windows)
- DI Administrator (Web)
- Web Applications

**Repository**
- Central Repository
- Local Repository
- Profiler Repository
- Global Parsing Repository

(Logical – repositories can be separate/different databases on different servers)

**Server**
- (Windows, UNIX, Linux)
- (Windows/Linux 32-bit, UNIX 64-bit)

- Address Directories
- Job Server and Engine
- Real-time Services
- Request-Response Access Server
Developing in Data Services

Development Environment
Data Services Designer

- Project
- Job
- Dataflow
- Repository Window
- Workflow / Dataflow Canvas
- Source
- Target
- Status
  - Job Server
  - Profiler Server
Pre-Built Interfaces

Offers tremendous productivity

- Pre-built *(no coding)* access to common databases, common ERP applications
  - Access to ERP applications is via ERP application layer, maintaining application integrity
  - Exposes the application’s metadata layer

**Databases:**
- Oracle
- DB2
- Sybase & IQ
- SQL Server
- Informix
- Teradata
- ODBC
- MySQL
- Netezza

**Applications:**
- JD Edwards
- Oracle Apps
- PeopleSoft
- Siebel
- SFDC
- SAP BI
- SAP ERP (R/3)
  - ABAP
  - BAPI
  - IDoc

**Files & Transport:**
- Text delimited
- Text fixed width
- EBCDIC
- XML
- Cobol
- Excel
- HTTP
- JMS
- SOAP (Web Services)

**Mainframe** *(with partner)*:
- ADABAS
- ISAM
- VSAM
- Enscribe
- Enscribe
- IMS/DB
- RMS
- Both direct and change data
Connecting to Datastore Source / Target

1. Right-click to create Connection
2. Select Name, Type
3. Enter Connection Details
Connecting to File Source

1. Right-click to create File Schema and Connection

2. Enter Connection Details

[Diagram showing BusinessObjects Data Services Designer interface with steps highlighted]
Data Profiling

Analysis of data beyond viewing
- Frequency distribution
- Distinct values
- Null values
- Minimum/Maximum values
- Data Patterns (e.g. Xxx Xxxx99, 99-Xxx)

Can drill down to view specific records

<table>
<thead>
<tr>
<th>Column</th>
<th>Distinct</th>
<th>Nulls</th>
<th>Nulls %</th>
<th>Zeros</th>
</tr>
</thead>
<tbody>
<tr>
<td>EMPLOYEE_GRP</td>
<td>4</td>
<td>355</td>
<td>21.93%</td>
<td>n/a</td>
</tr>
<tr>
<td>EMPLOYEE_SUB_GRP</td>
<td>20</td>
<td>355</td>
<td>21.93%</td>
<td>n/a</td>
</tr>
<tr>
<td>POSITION_ID</td>
<td>337</td>
<td>355</td>
<td>21.93%</td>
<td>n/a</td>
</tr>
<tr>
<td>ORG_UNIT_ID</td>
<td>117</td>
<td>355</td>
<td>21.93%</td>
<td>n/a</td>
</tr>
<tr>
<td>JOB_CODE</td>
<td>95</td>
<td>355</td>
<td>21.93%</td>
<td>n/a</td>
</tr>
<tr>
<td>PAYROLL_AREA</td>
<td>13</td>
<td>355</td>
<td>21.93%</td>
<td>n/a</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gender</th>
<th>Records</th>
<th>% of total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>1225</td>
<td>68.17%</td>
</tr>
<tr>
<td>Female</td>
<td>540</td>
<td>30.50%</td>
</tr>
<tr>
<td>Unknown</td>
<td>23</td>
<td>1.28%</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0.06%</td>
</tr>
<tr>
<td>Other</td>
<td>0</td>
<td>0.00%</td>
</tr>
</tbody>
</table>

View Data... - EMP_DEMOGRAPHICS (HR_IW.SAP_TARGET)
**Relationship Analysis**

Comparison of values between data sets to determine fit

- Shows % of non-matching values among
  - Table - Table
  - Flat file - Flat file
  - Table - Flat file

Can drill down to view actual records
Interactive Debugger

Very useful for checking the logic of a dataflow

- Can examine and modify data row-by-row
- Can place filters and breakpoints to pause execution of job and returns control to user
- Can break after processing a number of rows
Address Assignment Levels for Countries

Last-Line (or Country, Locality/City, Postcode) (248 countries/territories):

- Singapore, Malaysia, India, etc.

Primary/Street name: (41 countries):

- Brazil, Cyprus, French Guiana, French Polynesia, Greece, Greenland, Guadeloupe, Guam, Martinique, Mayotte, Monaco, New Caledonia, **New Zealand**, Northern Mariana Islands, Palau, Reunion, Saint Pierre and Miquelon, San Marino, Wallis and Futuna, *(including the ones listed under Primary/Street range)*

Primary/Street range (to unit level - 25 countries):

- American Samoa, **Australia***, Austria, Belgium, Canada*, Denmark, Faroe Islands, Finland, France, Germany, Italy, **Japan***, Liechtenstein, Luxembourg, Netherlands, Norway, Poland, Portugal, Puerto Rico, Spain, Sweden, Switzerland, United Kingdom, United States*, U.S. Virgin Islands,

Note:

1. * - Requires Country-specific Engine
2. Address-line (or Unit) assignment level for Singapore before end 2008 through license with SingaporePost
More on Address Cleansing / Verification

Original Addresses

<table>
<thead>
<tr>
<th>Addr1</th>
<th>Addr2</th>
</tr>
</thead>
<tbody>
<tr>
<td>500 E GRAND AVENUE</td>
<td>BELOIT WI  53511</td>
</tr>
<tr>
<td>3024 MATTHEW LN # 2</td>
<td>HOMEWOOD IL 60430</td>
</tr>
<tr>
<td>2005 STEWART DR #317</td>
<td>WACO TX 76708</td>
</tr>
<tr>
<td>768 MAGIE AVE</td>
<td>FAIRFIELD OH 45014</td>
</tr>
<tr>
<td>1338 N LEAVITT ST</td>
<td>CHICAGO IL 60622</td>
</tr>
<tr>
<td>PO BOX 703</td>
<td>OBION TN 38240</td>
</tr>
<tr>
<td>485 W MAPLE ST</td>
<td>NASHVILLE TN 62263</td>
</tr>
<tr>
<td>14977 ARCOLA</td>
<td>LIVONIA MI 48154</td>
</tr>
<tr>
<td>1225 S ROOSEVELT ST</td>
<td>HARRISBURG PA 62946</td>
</tr>
<tr>
<td>15208 PORCHESTER DR</td>
<td>NOBLESVILLE IN 46060</td>
</tr>
<tr>
<td>500 E GRAND AVE</td>
<td>BELOIT WI  53511</td>
</tr>
<tr>
<td>1281 FLUSHING AV</td>
<td>CLEARWATER FL 33764</td>
</tr>
<tr>
<td>924 SIR ECHO DR</td>
<td>KINGSPORT TN 37663</td>
</tr>
<tr>
<td>PO BOX 530</td>
<td>KOHLER WI  53044</td>
</tr>
<tr>
<td>9225 VARDEN PARK #104</td>
<td>KANSAS CITY MO 64114</td>
</tr>
<tr>
<td>1741 CROMO DR</td>
<td>MONTGOMERY AL 36106</td>
</tr>
<tr>
<td>444 HIGHLAND DR</td>
<td>KOHLER WI  53044</td>
</tr>
<tr>
<td>1014 HIGHWAY 6 E</td>
<td>OXFORD MS  38655</td>
</tr>
<tr>
<td>444 HIGHLAND DR</td>
<td>KOHLER WI  53044</td>
</tr>
<tr>
<td>135 GREENCREST DR</td>
<td>SLIDELL LA  70458</td>
</tr>
<tr>
<td>1007 GRESHAW ST</td>
<td>HOUSTON TX 77088</td>
</tr>
<tr>
<td>2615 167TH AVE</td>
<td>DEORAHIA IA 52101</td>
</tr>
<tr>
<td>444 HIGHLAND DR</td>
<td>KOHLER WI  53044</td>
</tr>
<tr>
<td>830 WATTEN LN</td>
<td>WESTERLY RI 02893</td>
</tr>
<tr>
<td>2555 S DURHAM DR</td>
<td>MOBILE AL  36502</td>
</tr>
<tr>
<td>13000 W CUSTER AVE</td>
<td>BUTLER PA  16001</td>
</tr>
</tbody>
</table>

Cleansed Addresses

<table>
<thead>
<tr>
<th>BEST_ADDRESS</th>
<th>LOCALITY</th>
<th>REGION</th>
<th>POSTCODE</th>
</tr>
</thead>
<tbody>
<tr>
<td>500 E Grand Ave</td>
<td>Beloit</td>
<td>WI</td>
<td>53511-6332</td>
</tr>
<tr>
<td>13000 W Custer Ave</td>
<td>Butler</td>
<td>WI</td>
<td>53007-1127</td>
</tr>
<tr>
<td>4402 Redwood Dr</td>
<td>Rockford</td>
<td>IL</td>
<td>61109-3541</td>
</tr>
<tr>
<td>300 Clara Dr</td>
<td>Monroe</td>
<td>LA</td>
<td>71203-6706</td>
</tr>
<tr>
<td>1277 Seminary Ave</td>
<td>Saint Paul</td>
<td>MN</td>
<td>55104-1443</td>
</tr>
<tr>
<td>4949 N 453 W</td>
<td>Huntington</td>
<td>IN</td>
<td>46750-8945</td>
</tr>
<tr>
<td>315 Ellis Ave Apt 9</td>
<td>Iowa City</td>
<td>IA</td>
<td>52246-3748</td>
</tr>
<tr>
<td>835 Aspen Cir</td>
<td>Saint Paul</td>
<td>MN</td>
<td>55109-1071</td>
</tr>
<tr>
<td>E14381 Morning Star Rd</td>
<td>La Farge</td>
<td>WI</td>
<td>54639-8589</td>
</tr>
<tr>
<td>500 E Grand Ave</td>
<td>Beloit</td>
<td>WI</td>
<td>53511-6332</td>
</tr>
<tr>
<td>321 Larkin St</td>
<td>Madison</td>
<td>WI</td>
<td>53705-5119</td>
</tr>
<tr>
<td>15587 Hubbs Rd</td>
<td>Pride</td>
<td>LA</td>
<td>70770-9714</td>
</tr>
<tr>
<td>2817 Irving Ave S</td>
<td>Minneapolis</td>
<td>MN</td>
<td>55408-1803</td>
</tr>
<tr>
<td>2132 80th St</td>
<td>Somerset</td>
<td>WI</td>
<td>54025-7416</td>
</tr>
<tr>
<td>1131 Gillian St</td>
<td>Lemont</td>
<td>IL</td>
<td>60439-4523</td>
</tr>
<tr>
<td>1215 S 250 E</td>
<td>Knox</td>
<td>IN</td>
<td>46534-9481</td>
</tr>
<tr>
<td>6760 50th Ave N</td>
<td>Crystal</td>
<td>MN</td>
<td>55428-4310</td>
</tr>
<tr>
<td>PO Box 360</td>
<td>Smithville</td>
<td>MS</td>
<td>38870-0360</td>
</tr>
<tr>
<td>1821 Sunset Blvd</td>
<td>Redington Shores</td>
<td>FL</td>
<td>33708-1049</td>
</tr>
<tr>
<td>27261 Varnado Cutoff Rd</td>
<td>Picayune</td>
<td>MS</td>
<td>39466-8511</td>
</tr>
<tr>
<td>1760 NW 80th Ctr</td>
<td>Clive</td>
<td>IA</td>
<td>50325-5608</td>
</tr>
<tr>
<td>38431 Waldorf Dr</td>
<td>Clinton Twp</td>
<td>MI</td>
<td>48038-2683</td>
</tr>
<tr>
<td>PO Box 184</td>
<td>Council Grove</td>
<td>KS</td>
<td>66846-0184</td>
</tr>
<tr>
<td>109 Beacon Dr</td>
<td>Hobart</td>
<td>IN</td>
<td>46342-5174</td>
</tr>
</tbody>
</table>

For some countries, Address Cleanse can correct or add postal codes.
More on Data (Customer) Cleansing

Original Customer Data

<table>
<thead>
<tr>
<th>Company</th>
<th>Name</th>
<th>Misc1</th>
<th>Misc2</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRECISION WOODWORK</td>
<td>E. MATTHEW ALONTE</td>
<td>&lt;Null&gt;</td>
<td>JAN. 27, 2001</td>
</tr>
<tr>
<td>PRECISION WOODWORK</td>
<td>MATTHEW ALONTE</td>
<td>6/1/93</td>
<td>&lt;Null&gt;</td>
</tr>
<tr>
<td>PRECISION WOODWORK</td>
<td>ALAN ALONTA</td>
<td>&lt;Null&gt;</td>
<td>SEP. 24 2000</td>
</tr>
<tr>
<td>STAFF BY MANNING</td>
<td>MARY GRIFFIN</td>
<td><a href="mailto:MARY@MANNING.COM">MARY@MANNING.COM</a></td>
<td>&lt;Null&gt;</td>
</tr>
<tr>
<td>M &amp; I BANK OF BELoit</td>
<td>JUDY ERIKSON</td>
<td>&lt;Null&gt;</td>
<td><a href="mailto:JUDY.ERICKSON@MANDI.COM">JUDY.ERICKSON@MANDI.COM</a></td>
</tr>
<tr>
<td>EFFICIENT OFFICE SUPPLY</td>
<td>WILLIAM J. SIMPSON</td>
<td>&lt;Null&gt;</td>
<td>10/22/2001</td>
</tr>
<tr>
<td>AMHERST BULLETIN</td>
<td>MARJORIE RINAMAN</td>
<td>&lt;Null&gt;</td>
<td></td>
</tr>
<tr>
<td>CALI TRAVEL</td>
<td>DORA HOSTENS</td>
<td>&lt;Null&gt;</td>
<td>&lt;Null&gt;</td>
</tr>
<tr>
<td>GEM SAVINGS</td>
<td>VINESLAVA MALEK</td>
<td>&lt;Null&gt;</td>
<td>JAN 24 1984</td>
</tr>
<tr>
<td>DUNRILE CENTER</td>
<td>BOB CREAMERS</td>
<td>&lt;Null&gt;</td>
<td></td>
</tr>
<tr>
<td>WESTERN FORGE</td>
<td>MARIA LINDSEY</td>
<td>&lt;Null&gt;</td>
<td></td>
</tr>
<tr>
<td>CONCORD OTOARYNGO...</td>
<td>THOMAS MORIARTY</td>
<td>&lt;Null&gt;</td>
<td></td>
</tr>
<tr>
<td>MAURICE HARRINGTON</td>
<td>JANET RECA</td>
<td>&lt;Null&gt;</td>
<td></td>
</tr>
</tbody>
</table>

Cleansed Customer Data

<table>
<thead>
<tr>
<th>DC_FIRM</th>
<th>DC_FNAME</th>
<th>DC_LNAME</th>
<th>DC_TITLE</th>
<th>DC_DATE</th>
<th>DC_EMAIL</th>
</tr>
</thead>
<tbody>
<tr>
<td>M. &amp; I. Bank</td>
<td>Kelly</td>
<td>Stenberg</td>
<td>&lt;Blank&gt;</td>
<td>9/16/1995</td>
<td><a href="mailto:kstenberg@mibank.com">kstenberg@mibank.com</a></td>
</tr>
<tr>
<td>Show Bus. Travel Tours</td>
<td>Hazel</td>
<td>Rutland</td>
<td>&lt;Blank&gt;</td>
<td>9/9/1989</td>
<td>&lt;Blank&gt;</td>
</tr>
<tr>
<td>Precision Woodwork</td>
<td>Matthew</td>
<td>Alonте</td>
<td>&lt;Blank&gt;</td>
<td>6/1/2/1993</td>
<td>&lt;Blank&gt;</td>
</tr>
<tr>
<td>Logic Systems Inc.</td>
<td>Nannette</td>
<td>Smith</td>
<td>Executive Dir. of Mktg.</td>
<td>6/1/1994</td>
<td>&lt;Blank&gt;</td>
</tr>
<tr>
<td>One Hour Cleaner</td>
<td>Keith</td>
<td>Pautler</td>
<td>Owner</td>
<td>6/5/1996</td>
<td>&lt;Blank&gt;</td>
</tr>
<tr>
<td>Electronic Info. Sys.</td>
<td>Antimo</td>
<td>Viglotti</td>
<td>Sales Technician</td>
<td>&lt;Blank&gt;</td>
<td>&lt;Blank&gt;</td>
</tr>
<tr>
<td>Chowchilla News</td>
<td>Charlie</td>
<td>Miranda</td>
<td>&lt;Blank&gt;</td>
<td>&lt;Blank&gt;</td>
<td><a href="mailto:charlie@chowchilla.com">charlie@chowchilla.com</a></td>
</tr>
<tr>
<td>Nat Sherman Inc.</td>
<td>L.</td>
<td>Bradykempf</td>
<td>&lt;Blank&gt;</td>
<td>&lt;Blank&gt;</td>
<td>&lt;Blank&gt;</td>
</tr>
<tr>
<td>Sokol Insurance</td>
<td>Yanie</td>
<td>Dubouchage</td>
<td>&lt;Blank&gt;</td>
<td>&lt;Blank&gt;</td>
<td>&lt;Blank&gt;</td>
</tr>
<tr>
<td>Imperial 500</td>
<td>Dorothy</td>
<td>Keryluk</td>
<td>&lt;Blank&gt;</td>
<td>10/1/1999</td>
<td>&lt;Blank&gt;</td>
</tr>
<tr>
<td>M &amp; I Bank of Beloit</td>
<td>Judy</td>
<td>Erickson</td>
<td>&lt;Blank&gt;</td>
<td>&lt;Blank&gt;</td>
<td><a href="mailto:judy.ericson@mandi.com">judy.ericson@mandi.com</a></td>
</tr>
<tr>
<td>Smith &amp; Harris Mach.</td>
<td>Joseph</td>
<td>Ward</td>
<td>&lt;Blank&gt;</td>
<td>&lt;Blank&gt;</td>
<td>&lt;Blank&gt;</td>
</tr>
<tr>
<td>Palmer Travel Agency</td>
<td>Mark</td>
<td>Deeb</td>
<td>Sales Agent</td>
<td>&lt;Blank&gt;</td>
<td>&lt;Blank&gt;</td>
</tr>
<tr>
<td>Cignaese Construction Inc.</td>
<td>Walter</td>
<td>K measuring</td>
<td>&lt;Blank&gt;</td>
<td>&lt;Blank&gt;</td>
<td>&lt;Blank&gt;</td>
</tr>
</tbody>
</table>

Parsing (Identification and Isolation of specific parts of mixed data) can be extended to non-Customer information
Universal Data Cleanse

Uncleaned Product data

<table>
<thead>
<tr>
<th>ORIG_PRODUCT</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>10-year, $100,000 CD with a 7.10% interest rate</td>
<td>Two-year CD 6.55% rate $50K minimum deposit Imperial Cap. Bnk.</td>
</tr>
<tr>
<td>CD at 5.25% and six months $50K @ Washington Mutual</td>
<td>WAMU 4.25% on five-year CDs $100,000 minimum deposit</td>
</tr>
<tr>
<td>Alliance Bank 4 yr jumbo CD $100,000 minimum deposit 5.55%</td>
<td>$10K min. dep. 5.55% for 1 month CD</td>
</tr>
<tr>
<td>Security Savings Bank 5.51% APY for 2.5 yr jumbo CD $100,000 minimum dep.</td>
<td>3 MONTH $25K CD 5.36% APY DANE COUNTY C.U.</td>
</tr>
<tr>
<td>5.40% 3 MONTH $50K CD RIVER CITY COMMUNITY CREDIT UNION</td>
<td>$10,000 min. deposit 2.5 yr jumbo CD 5.48% Amtrustdirect.com</td>
</tr>
<tr>
<td>6.55% for 3 month CD $10,000 min. deposit</td>
<td>Countrywide Bank $50000 min. deposit 3-month CD 5.40% APY</td>
</tr>
<tr>
<td>10-month CDs $10,000 min. deposit 5.51% APY</td>
<td>5.36% 4 yr jumbo CD $100,000 minimum deposit</td>
</tr>
</tbody>
</table>

Cleansed data

<table>
<thead>
<tr>
<th>INSTITUTION</th>
<th>TERM</th>
<th>PRODUCT</th>
<th>YIELD</th>
<th>MIN_DEPOSIT</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;Blank&gt;</td>
<td>10 Yr</td>
<td>CD</td>
<td>7.10%</td>
<td>$100,000</td>
</tr>
<tr>
<td>Imperial Capital Bank</td>
<td>1 Yr</td>
<td>CD</td>
<td>6.55%</td>
<td>$50,000</td>
</tr>
<tr>
<td>Washington Mutual</td>
<td>8 Months</td>
<td>CD</td>
<td>5.25%</td>
<td>$50,000</td>
</tr>
<tr>
<td>Washington Mutual</td>
<td>5 Yr</td>
<td>CD</td>
<td>4.25%</td>
<td>$100,000</td>
</tr>
<tr>
<td>Alliance Bank</td>
<td>4 Yr</td>
<td>Jumbo CD</td>
<td>5.55%</td>
<td>$100,000</td>
</tr>
<tr>
<td>&lt;Blank&gt;</td>
<td>1 Month</td>
<td>CD</td>
<td>5.55%</td>
<td>$10,000</td>
</tr>
<tr>
<td>Security Savings Bank</td>
<td>2.5 Yr</td>
<td>Jumbo CD</td>
<td>5.51%</td>
<td>$100,000</td>
</tr>
<tr>
<td>Dane County Credit Union</td>
<td>3 Month</td>
<td>CD</td>
<td>5.36%</td>
<td>$25,000</td>
</tr>
<tr>
<td>River City Community Credit Union</td>
<td>3 Month</td>
<td>CD</td>
<td>5.40%</td>
<td>$50,000</td>
</tr>
<tr>
<td>AmTrustDirect.com</td>
<td>2.5 Yr</td>
<td>Jumbo CD</td>
<td>5.48%</td>
<td>$10,000</td>
</tr>
<tr>
<td>&lt;Blank&gt;</td>
<td>3 Month</td>
<td>CD</td>
<td>5.36%</td>
<td>$10,000</td>
</tr>
<tr>
<td>Countrywide Bank</td>
<td>10 Month</td>
<td>CD</td>
<td>5.51%</td>
<td>$10,000</td>
</tr>
<tr>
<td>&lt;Blank&gt;</td>
<td>4 Yr</td>
<td>Jumbo CD</td>
<td>5.36%</td>
<td>$100,000</td>
</tr>
<tr>
<td>1st National</td>
<td>1 Month</td>
<td>CD</td>
<td>5.87%</td>
<td>$10,000</td>
</tr>
<tr>
<td>1st National</td>
<td>2.5 Yr</td>
<td>Jumbo CD</td>
<td>5.48%</td>
<td>$100,000</td>
</tr>
<tr>
<td>Independence Bank</td>
<td>3 Month</td>
<td>CD</td>
<td>5.67%</td>
<td>$25,000</td>
</tr>
<tr>
<td>Stearns Bank</td>
<td>3 Month</td>
<td>CD</td>
<td>5.40%</td>
<td>$50,000</td>
</tr>
<tr>
<td>Northwestern Mutual Credit Union</td>
<td>2.5 Yr</td>
<td>CD</td>
<td>5.55%</td>
<td>$50,000</td>
</tr>
<tr>
<td>Tennessee Commerce Bank</td>
<td>3 Month</td>
<td>CD</td>
<td>5.48%</td>
<td>$100,000</td>
</tr>
<tr>
<td>MetLife Bank</td>
<td>5 Yr</td>
<td>CD</td>
<td>5.87%</td>
<td>$100,000</td>
</tr>
<tr>
<td>FirstFedDirect</td>
<td>6 Month</td>
<td>&lt;Blank&gt;</td>
<td>5.35%</td>
<td>$10,000</td>
</tr>
</tbody>
</table>

Data Quality Management

Custom Dictionary / Rules

Original, Raw data

Standardized and properly formatted data
Universal Data Cleanse
Custom Parser

Word breaking
- Break the input down into smaller pieces

Tokenization
- Assign meaning to the pieces

Rule matching (pattern)
- Match the pieces to rules

Actions & Action item assignment
- Create output from rule matches

Word Break ➔ Tokenize ➔ Rule Match ➔ Action

Large mushrooms sausage pepperoni stuffed crust

Rule1 = Size + Topping + Topping + Topping + Crust + Crust;

Action = Pizza;
Pizza = 1 : Pizza_Crust : 5;
Pizza = 1 : Pizza_Crust : 6;
Pizza = 1 : Pizza_Topping : 2;
Pizza = 1 : Pizza_Topping : 3;
Pizza = 1 : Pizza_Topping : 4;
Pizza = 1 : Pizza_Size : 1;
End_Action;

Pizza
  ├── Pizza_Crust
  │       └── Pizza_Size
  └── Pizza_Topping
Universal Data Cleanse
Custom Dictionaries

Dictionary entries can be added or modified

Rule: Replace this word 白い with “white” when it is classified as COLOR

Entire new dictionary can be created
Deploying in Data Services

Deployment Environment
Multi-User Development Environment

DEV
Job Server

Central Repo

TEST
Job Server

Central Repo

PROD
Job Server

Local Repo

Local Repo

Local Repo

Local Repo

Check in

Get

Check in

Get

Server Group

SAP BusinessObjects
Check-In/Check-Out Management

Sophisticated method for sharing/moving DI objects

- Involves check-in/out of objects with versioning
Real-Time vs Batch (Scheduled) Processing Support

Real-Time capabilities built into the engine in the form of Request-Response message processing

Typical Batch Job

1. Message
2. Listener
3. Immediate action upon receiving Message
4. Response

Converted to Real-Time Job

1. Message
2. Real-time Jobs
3. Immediate action upon receiving Response

ERP or Web applications

Access Server

Data Services

Job Server and Engine

© SAP 2008 / Page 31
Invoking/Consuming Data Services-Produced Web Services

1. External Web Service client
2. Data Services Web Server
3. Web Services server
4a. Request-Response
4b. Real-time jobs
4c. Access Server
5a. Invoke Real-time jobs
5b. Invoke Batch jobs

Job Server and Engine
Repository
Consuming 3rd Party Web Services

1. Define Web Service source

2. Use imported Web Service operation as standard function
Deploying for Performance

Deployment Environment
Data Sources/Targets can be switched between different development environments.
# Data Services

## System Configurations

<table>
<thead>
<tr>
<th>Substitution Parameter</th>
<th>DEV</th>
<th>PROD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$$FILE_PATH</td>
<td>C:\TEMP</td>
<td>C:\PROD</td>
</tr>
<tr>
<td>$$LOG_PATH</td>
<td>C:\LOG</td>
<td>C:\LOG</td>
</tr>
</tbody>
</table>

### Datastore Source

- **Configuration 1**
  - **Database type**: Oracle
  - **Server name**: SJ-PROD
  - **Database name**: Source_Data
  - **User**: sd_5263
  - **Password**: *****

### Datastore Target

- **LocalServer**
  - **Database type**: mySQL
  - **Server name**: Localhost
  - **Database name**: Source_Data
  - **User**: sa
  - **Password**: *****

- **ProdServer**
  - **Database type**: Oracle
  - **Server name**: SJ-PROD
  - **Database name**: Source_Data
  - **User**: sd_5263
  - **Password**: *****
Parallel Execution
- via explicit Parallel Data flows

The Dimensions data flows run before the Facts data flow

The Dimension data flows run in parallel
Parallel Execution

- via Partitioned source and target

Non-partitioned data flow at Design Time

Partitioned at Execution time

If Source is partitioned 2 ways

If Target is partitioned 2 ways

Specifying Partitioning scheme
Parallel Execution

- via specifying Degree of Parallelism (DOP) > 1

At Design Time

At Execution Time
Distributing Data Flow Execution - 1

Run as a separate process for the following resource-intensive operations
- Hierarchy_Flattening
- Join
- GROUP_BY
- ORDER_BY
- DISTINCT
- Table_Comparison
- Lookup_ext function
- Count_distinct function
- Associate
- CountryID
- Global Address Cleanse
- Global Suggestion List
- Match
- User-Defined
Distributing Data Flow Execution - 2

Various ways available to improve performance

- Flexible ETL (traditional transform)
- Hybrid “ELT” (transform at target, transform at source)
  - Concept is “Push-Down” processing
“Push-Down” in detail

- “Push-Down” can be staged onto database or file directory

To push down operations to the source server

To push down the GroupBy operation to the target server

Data Flow with Push-Down’s

Push-Down settings

- Enable transfer
- Transfer types: File
- Join ranks: Automatic
- File options:
  - Root directory: c:\temp\emp
  - File name: DataTransfer

Original Data Flow

Distributing Data Flow Execution - 3
Grid deployment (across multiple servers) is easy

- No changes required in Job flows
- Simply specify at execution time what to distribute across servers – Job, Data flow, or sub-Data flow
Distribution Level 101

- Job level distribution
  - All processes belonging to job execute on the same computer

All sub-data flows run on the same computer
Distribution Level 101

- Data flow level distribution
  - All processes of each data flow can execute on a different computer
Distribution Level 101

- Sub data flow level distribution
  - Each sub data flow can execute on a different computer
Managing with Data Services

Management Environment
Web-based Management with graphical dash-board info
Recovery From Unsuccessful Job Execution

Automated Recovery

- With this set, DI records the result of each successful step in a job
- During recovery, DI re-runs uncompleted or failed steps under the same conditions as the original job

There is Manual Recovery, too
Auto-Reporting

Via Web-based Management Console

- Reports automatically generated
- Report details can be interactively chosen and drilled into
- Reports can be printed

Interactive Dataflow graphics
Lineage And Impact

Provides analyzing dependencies for
- Impact (source to which target/s)
- Lineage (target back to which source/s)
Thank you!