

Informatica Data Quality 9.0.1

Bev Duddridge
Principal Instructor
Global Education Services

bduddridge@informatica.com



Agenda

- Analyst and Developer Tools
- Perform Column, Rule, Join and Mid-Stream Profiling
- Manage reference tables
- Collaborate on projects
- Scorecard data
- Design and develop Mapplets and Rules
- Create standardization, cleansing and parsing routines
- Validate addresses
- Identify duplicate records
- Associate and consolidate matched records
- Migrating from 8.6.2 to 9.0.1
- Logs and troubleshooting 9.0.1

Informatica Analyst 9.0.1

- **Informatica Analyst is a web-based application client that analysts can use to analyze, profile, and score data in an enterprise.**
- **Business analysts and developers use Informatica Analyst for data-driven collaboration.**
- **You can perform column and rule profiling, Scorecarding, bad record and duplicate record management.**
- **You can also manage reference data and provide the data to developers in a data quality solution.**

Informatica Developer 9.0.1

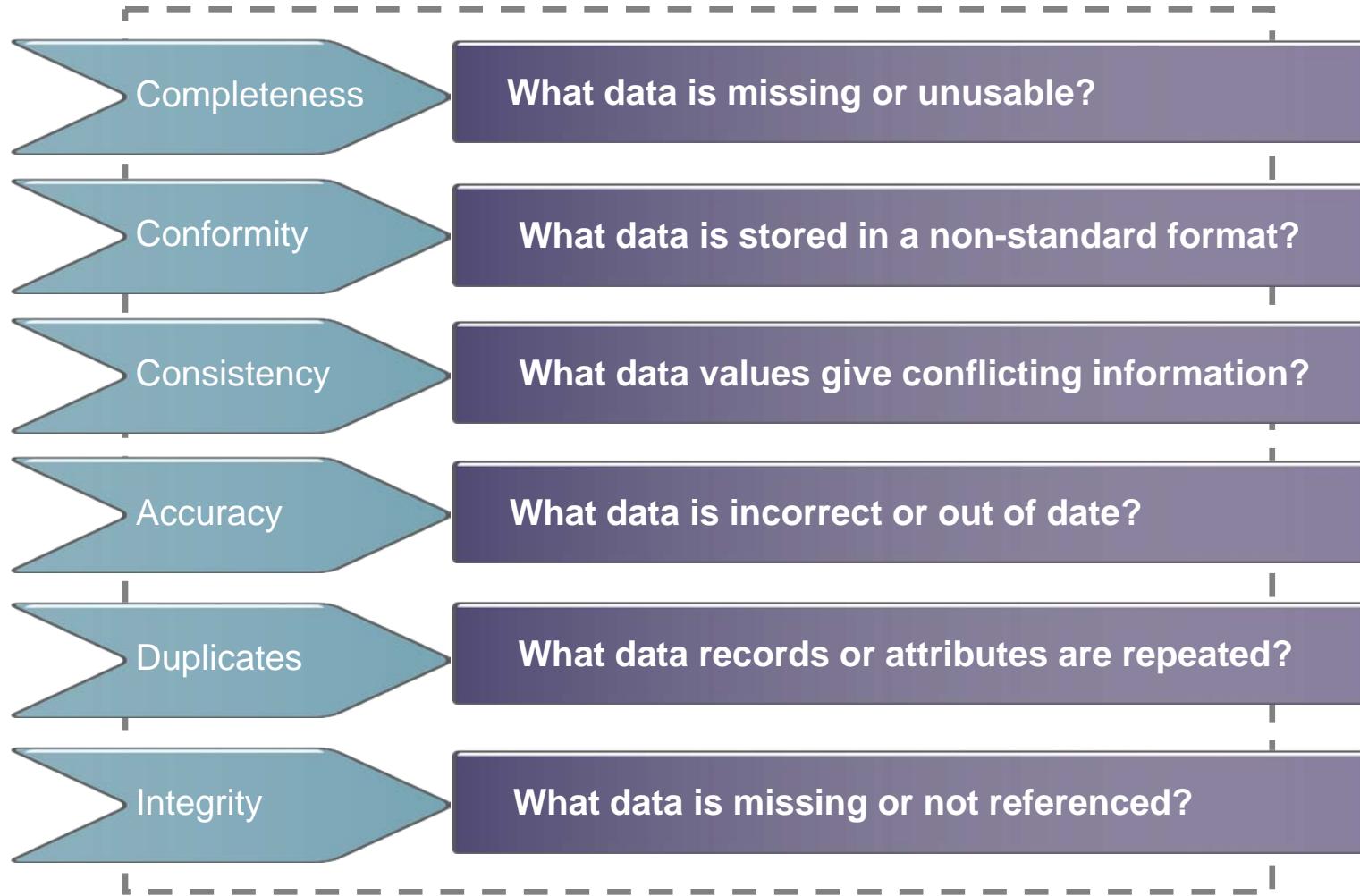
- **Informatica Developer is an application client that developers use to design and implement data quality and data services solutions.**
- **Use the data quality capabilities in the Developer tool to analyze the content and structure of your data and enhance the data in ways that meet your business needs.**
 - **Profile, Standardize and Parse data.**
 - **Validate postal addresses.**
 - **Identify duplicate records.**
 - **Create and run data quality rules.**
 - **Collaborate with Informatica users.**

Introduction to Data Quality Management

What is Data Quality Management?

- A set of processes that measure and improve the quality of important data on an ongoing basis
- Ensures that data dependent business processes and applications deliver expected results

Six dimensions of Data Quality

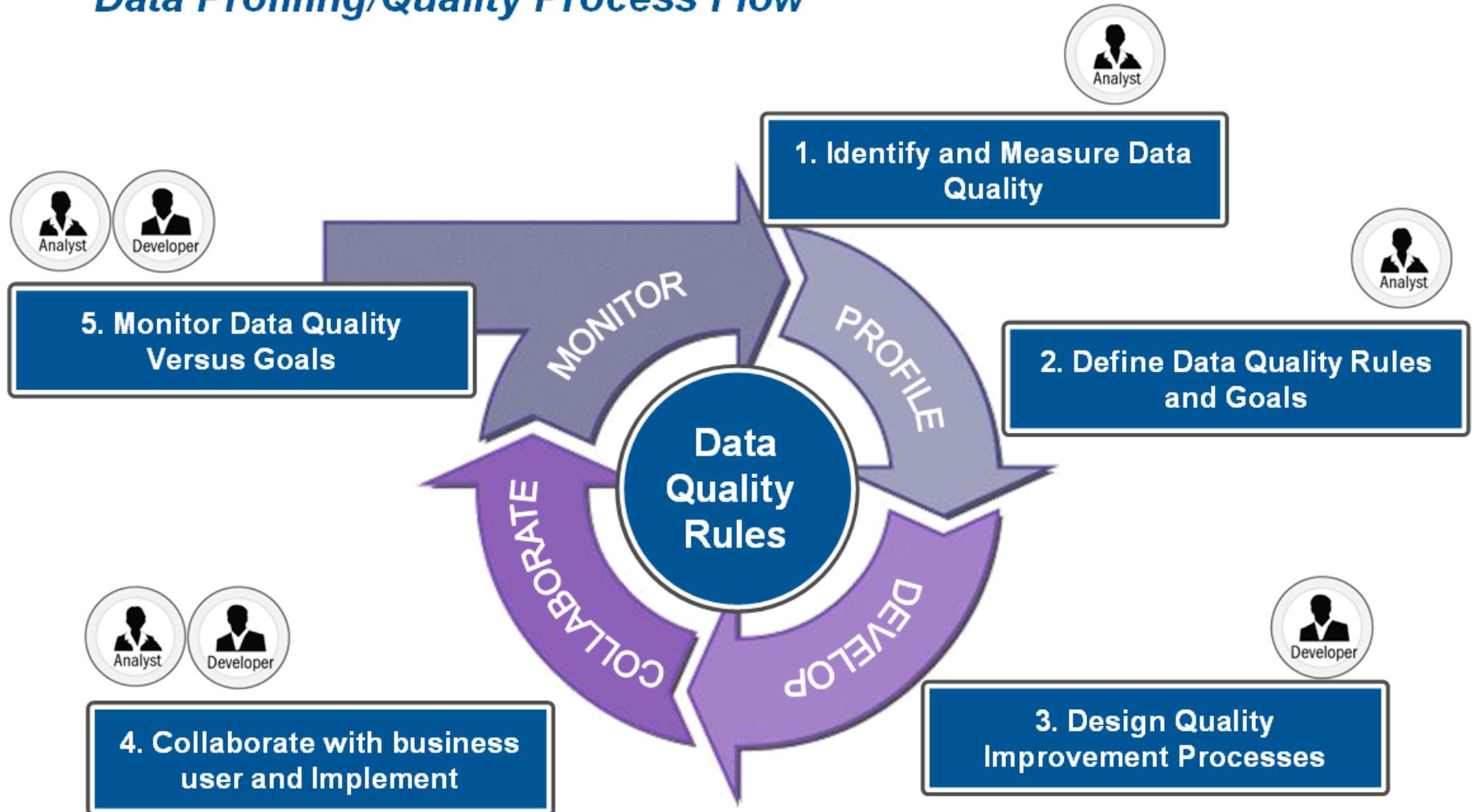


Data Quality Problems

	er	Cust_name	Business Type	Last Invoice Date	Customer Status	Cust. Turnover	sys_address_1	sys_address_2	sys_address_3	sys_address_4
COMPLETENESS	I9479	ALBERT WILSON	PERSON	30/06/2001	LIVE	0	48 HOWARD ROAD	LONDON	NW2 6DR	
	I0186	PETER CLARKE	BUSINESS	10/02/2004	LIVE	87	36 HARBORD STREET	LONDON	SW6 6PJ	
	I0460	JOHN LENAGHAN	PERSON	01/02/2003	TERM	84	KERROMOOR	HIGHAM LANE	TONBRIDGE	KENT
	I0823	JANE DUNNE	PERSON	23/05/2002	LIVE	0	C/O Phillip Jones	14 BUXTON STREET	ACCRINGTON	LANCASHIRE
	I1679	JOSEPH LEDDIN	PERSON	14/09/1999	LIVE	0	7 HEATHFIELD TERRACE	LONDON	W4 4LP	
CONFORMITY	I1353	NIALL MESCALL	PERSON	10/02/2004	TERM	0	25 NIGHTINGALE GROVE	SHEPTON MALLETT	SOMERSET	BA4 5PZ
	I1353	JAMES MAHER	BUSINESS	30/06/2001	LIVE	76	8 YORK STREET	PETERHEAD	ABERDEENSHIRE	AB42 1QA
	I3137	DAVID G LACEY	PERSON	10/02/2004	LIVE	163	43 TREHURST STREET	LONDON	E5 0EB	
	I4346	JOHN O CONNOR	PERSON	01/02/2003	TERM	134	31G	CHAMPION HILL	LONDON	SE5 8BS
	I6593	LORNA MASON	PERSON	23/05/2002	LIVE	108	16 KEMPLAY ROAD	LONDON	NW3 1SY	
CONSISTENCY	I8322	SUSAN KILROY	PERSON	14/09/2004	LIVE	76	23 KELVIN CRESCENT	BEARSDEN	GLASGOW	G61 1BT
	I3990	CHISLAIN DRION	PERSON		TERM	198	235 LONDON ROAD	MORDEN	SURREY	SM4 5PU
	I4622	GEORGE LOUGHRAHAN	PERSON		LIVE	130	3 WORSLEY ROAD	WORSLEY	MANCHESTER	M28 2NN
	I3647	GERARD EGAN	PERSON		LIVE	77	9 FIR GROVE	ELLINGTON	MORPETH	NORTHUMBERLAND
	I3647	MR THOMAS SHEEHAN	BUSINESS	01/02/2003	TERM	509	42 CAMUS AVENUE	EDINBURGH	EH10 6QU	
DUPLICATION	I3647	BK ENGINEERING	PERSON	23/05/2002	LIVE	269	14 DIEPPE CRESCENT	PORTSMOUTH	PO2 9PL	
	I3647	EDWARD LEDESMA	PERSON	14/09/2004	LIVE	65	72 YALE COURT	HONEYBOURNE ROAD	LONDON	NW6 1JQ
	I3647	FLOR CORCORAN	PERSON	01/02/2003	TERM	0	14 SILVERKNOWES PLACE	EDINBURGH	LOTHIAN	EH4 5LS
	I3647	MR MARTIN THATCHER	BUSINESS	23/05/2002	LIVE	82	94 WELLINGTON ROAD	BOSTON	LINCOLNSHIRE	PE21 0LE
	I3647	ROBERT MCGONAGLE	PERSON	14/09/2004	LIVE	291		1 BARROWGATE ROAD	LONDON	W4 4QX
INTEGRITY	I3647	MR PETER KELLEHER	PERSON	10/02/2004	TERM	119	33 COPPIE AVENUE	SALE	WEST MIDLANDS	M33 4ND
	I2564	CHRIS MOLONEY	PERSON	12/10/2001	TERM	1	18 CROYDON ROAD	KENT		
	I3647	PAUL MULLEN	PERSON	10/02/2004	LIVE	669	36 PRIORY GARDENS	LONDON	N6 5QS	
	I3647	PHILIP FAULKNER	BUSINESS	01/02/2003	TERM	69	35 DALVEEN WAY	RUTHERGLEN	GLASGOW	G73 4HJ
	I3647	MR. CHRISTOPHER MOLC	PERSON	30/06/2001	LIVE	1	185 CROYDON ROAD	BECKENHAM	KENT	BR3 3QH
ACCURACY	I3647	GABRIEL MCAREE	PERSON	14/09/2004	LIVE	383	16 SEFTON AVENUE	NEWCASTLE UPON TYNE	NE6 5QR	
	I3647	MR.NOEL KISSANE	PERSON	10/02/2004	TERM	510	166 BANNOCKBURN ROAD	STIRLING	FK7 0EW	
	I3647	FR ANTHONY MULLINS	PERSON	30/06/2001	LIVE	60	6 ARDOCH GARDENS	CAMBUSLANG	GLASGOW	G72 8HB
	I3647	MR NOEL MACAREE DEC	PERSON	10/02/2004	LIVE	0	9 WARRIOR SQUARE	EASTBOURNE	EAST SUSSEX	BN22 7DB
	I3647	EUGENE MAGUIRE	PERSON	01/02/2003	TERM	320	45 MONTFORT RISE	REDHILL	RH1 5DU	
	I3634	JANE POLLARD	PERSON	23/05/2002	LIVE	224	17 OLD BARRACK YARD	LONDON	SW1X 7NP	
	I3647	PAT HEGARTY	PERSON	14/09/2004	LIVE	126	6 INNER LEITHAN WAY	PERTH	PH1 1RN	
	I3647	WILLIAM POLLARD	PERSON	19/03/2002	LIVE	224	17 OLD BARRACK YARD	LONDON	SW1X 7NP	
	I3647	JAMES MURPHY	PERSON	30/06/2001		167	9 CARLISLE PARK	BALLYNAHINCH	COUNTY DOWN	BT24 8HJ

Data Quality Management

Data Profiling/Quality Process Flow



Data Quality Management

1. Profile

- Identify DQ problems through Profiling using either the Analyst or Developer Tools

2. Collaborate

- Developers and Analysts can work together to build the DQ management process

3. Standardize

- Once the problems with the data have been identified, develop your standardization process to cleanse, standardize, enrich and validate your data

Match

- Identify duplicate records in your data using a variety of matching techniques

Consolidate

- Automatically or manually consolidate your matched records

4. Collaborate

- Developers and Analysts can work together to build the DQ management process

Data Quality Management

1. Profile

- Identify DQ problems through Profiling using *either the Analyst or Developer Tools*

2. Collaborate

- **Developers and Analysts can work together to build the DQ management process**

3. Standardize

- Once the problems with the data have been identified, develop your standardization process to cleanse, standardize, enrich and validate your data

Match

- Identify duplicate records in your data using a variety of matching techniques

Consolidate

- Automatically or manually consolidate your matched records

4. Collaborate

- Developers and Analysts can work together to build the DQ management process

Data Quality Management

1. Profile
 - Identify DQ problems through Profiling using either the Analyst or Developer Tools
2. Collaborate
 - Developers and Analysts can work together to build the DQ management process

3. Standardize

- Once the problems with the data have been identified, develop your standardization process to cleanse, standardize, enrich and validate your data

Match

- Identify duplicate records in your data using a variety of matching techniques

Consolidate

- Automatically or manually consolidate your matched records

4. Collaborate

- Developers and Analysts can work together to build the DQ management process

Data Quality Management

1. Profile

- Identify DQ problems through Profiling using *either the Analyst or Developer Tools*

2. Collaborate

- Developers and Analysts can work together to build the DQ management process

3. Standardize

- Once the problems with the data have been identified, develop your standardization process to cleanse, standardize, enrich and validate your data

Match

- Identify duplicate records in your data using a variety of matching techniques

Consolidate

- Automatically or manually consolidate your matched records

4. Collaborate and Monitor

- Developers and Analysts can work together to build the DQ management process

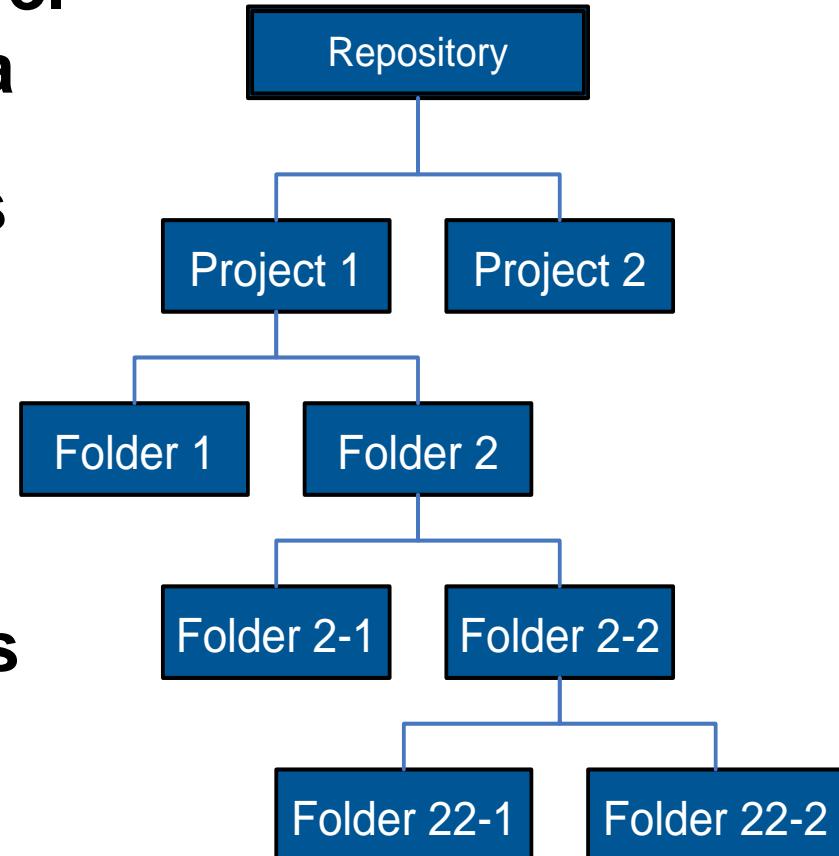
Informatica Analyst

Informatica Analyst Tool

- **Data Objects**
 - Metadata import for Data Sources
 - Data access and preview
- **Profiling**
 - Column Profiling
 - Rule Profiling
 - Expression based Rule creation/editing
 - Project Collaboration
- **Reference Table Manager**
 - Authoring and editing of reference data
 - Auditing of changes
- **Data Quality Scorecarding**
 - Scorecards in the Analyst Tool
- **Data Quality Assistant**
 - Management of Bad Records and Duplicate Records
 - Auditing of changes

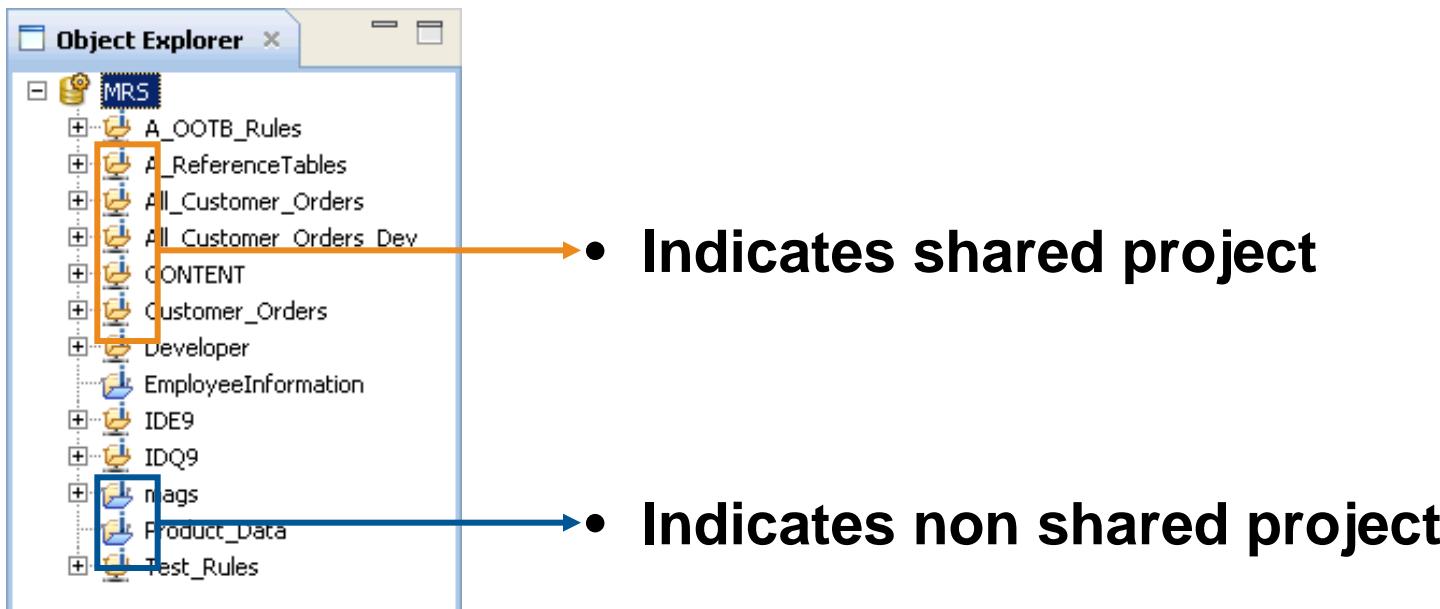
Repository, Projects and Folders

- Projects are the highest level containers for the metadata
- Project can contain objects or Folders
- Folders can be nested
- Organize objects in folders as per your business needs



Projects

- “Shared” option is available at folder creation time only and cannot be changed afterwards
 - Shared Projects
 - Non-shared Projects



The Informatica Analyst GUI

INFORMATICA® Analyst

Administrator Log Out | Manage ▾ | Help ▾ Search All Go

Browse: Projects

Actions

Project Navigator

Project Contents

Profiles

Scorecards

DQA

Data Objects

Reference Tables

Rules

Contents Properties

Navigator Actions

Training_Materials - Folder Contents

Name	Type
Profile (3)	
Profile_All_CustomerOrders	Profile
Profile_EU_Customer_Orders	Profile
Profile_US_Customer_Orders	Profile
Scorecard (1)	
US_Customer_Scorecard	Scorecard
Bad Record (1)	
CUSTOMER	Bad Record
Duplicate Record Table (1)	
CUSTOMERDUP	Duplicate Record Table
Data Object (15)	
Rule (5)	
mplt_MatchCompanyDetail	Rule
rule_Derive_Prefix_from_Firstname	Rule
rule_EUtoUS_DateFormat	Rule
rule_LeadZeroFiveDigitZip	Rule

Properties

No object is selected.

Actions

- Open
- Duplicate...
- Rename...
- Move...
- Delete
- New Flat File...
- New Table...
- New Profile...
- New Reference Table...
- New Data Quality Table...
- Close all Tabs
- Help with Contents...

Physical Data Objects

- **Physical Data Objects**
 - **File**
 - Browse and Upload
 - Network path/shared directory
 - **Table**
- **Data Sources can be**
 - Previewed
 - Profiled
 - Scorecarded

Data Objects

INFORMATICA Analyst

Administrator Sign Off | User Preferences | Help ▾ Search All Go

Browse: Projects Profile_All_Customer... EU_custord

Data Preview Properties Actions ▾

EU_custord - Data Preview (first 100 rows only)

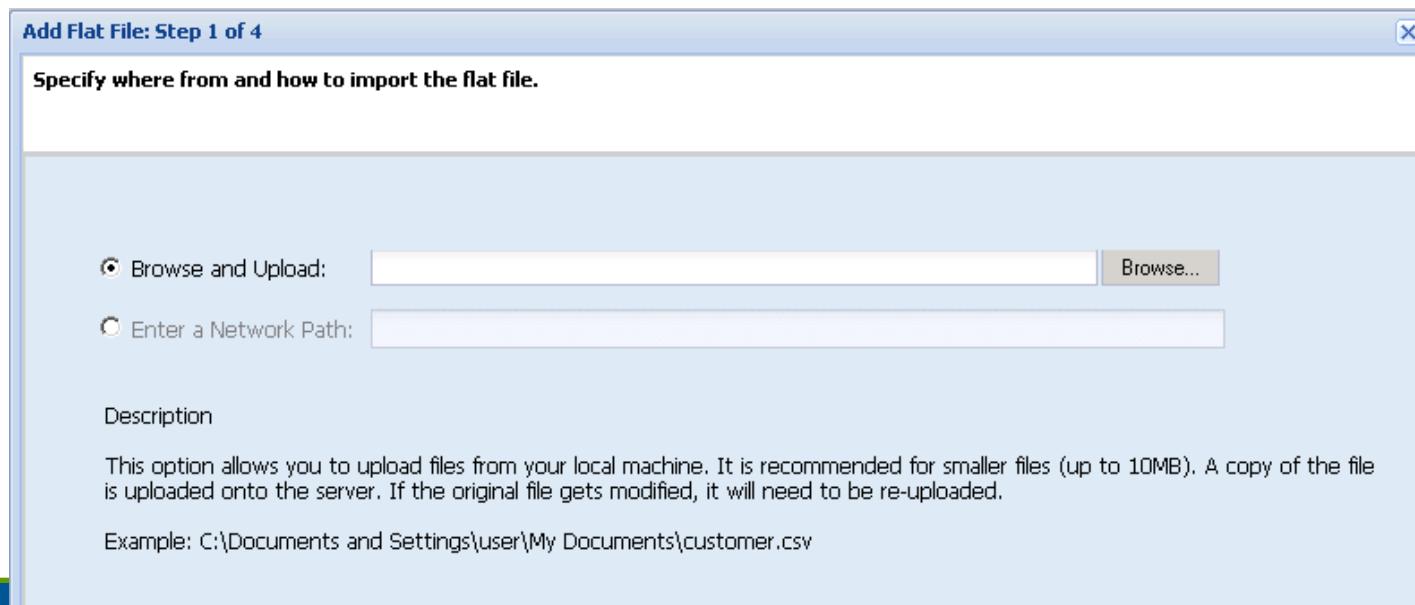
ID	ORDER_NO	PO_NUM	ORDER_DATE	SHIP_DT	COMPANY	ADDR1	ADDR2	ADDR3	
1	5000	2256K	2050	99/99/9999	18/1/2009	Hiperdino	Luis Cernuda ...	OLIVARES ...	SEVILL...
2	5001	23222	2466	11/2/2007	27/2/2009	Alcampo	Santa Coloma 114	BARCELONA ...	BARCEI...
3	5002	2322K	173003	11/2/2007	27/2/2009	Alcampo	Santa Coloma 114	BARCELONA	BARCEI...
4	5003	2288K	181	18/8/2007	25/1/2009	Aldi	11 Almirall Reque...	SOL I VISTA REUS TARRA...	
5	5004	2301K	2414	19/9/2007	14/5/2009	Alcampo	12 Granada	VILLAFRANCA D...	BADAJ...
6	5005	2267K	529	22/12/2008	27/6/2009	Alcampo	14 Virgen De Los...	RIBA-ROJA DE T...	VALEN...
7	5006	23771K	897	22/12/2008	27/6/2009	Alcampo	14 Virgen De Los...	RIBA-ROJA DE T...	VALEN...
8	5007	J4781	8977	27/7/2007	27/1/2009	Lidl	19 Pueblo Garnicia	TORRELAVEGA	CANTA...
9	5008	J2269	2054	27/11/2007	4/6/2009	Alcamplos	2 Julio Nebot	ALBAL	VALEN...
10	5009	J2302	853	27/4/2007	5/2/2009	Al campo	2 Union Y Cultura	VILLA DEL RIO	CORDO...
11	5010	J23406	509	27/4/2007	5/2/2009	Al campo	2 Union Y Cultura	VILLA DEL RIO	CORDO...
12	5011	J2278	107672	23/12/2008	6/10/2009	Carrefours	4 Doctor Fleming	CUDILLERO	ASTURI...
13	5012	23772	1976	23/12/2008	6/10/2009	Carrefours	4 Doctor Fleming	CUDILLERO	ASTURI...
14	5013	J2281	2106	2/12/2008	5/10/2009	Carre four	4 Escuelas	CAMINO MORISCO	CACERI...
15	5014	22921	2056	2/12/2008	5/10/2009	Carre four	4 Escuelas	CAMINO MORISCO	CACERI...
16	5015	23688	878	23/10/2007	25/6/2009	Sabeco	4 Padre Vicent	VALENCIA	VALEN...
17	5016	J2277	56432	23/10/2007	25/6/2009	Sabeco	4 Padre Vicent	VALENCIA	VALEN...
18	5017	J2299	864	8/1/2008	29/4/2009	Alcampo	43 Bernat Guinov...	ALGEMESI GRAND...	VALEN...
19	5018	J2298	13483	9/9/2007	21/4/2009	SPAR	5 QUATRE CAMINS	VILASSAR DE M...	BARCEI...
20	5019	22903	12762	9/9/2007	21/4/2009	SPAR	5 QUATRE CAMINS	VILASSAR DE M...	BARCEI...
21	5020	2293K	574	24/9/2007	3/8/2009	Lidls	6 Doctor Nevado ...	CORDOBA	CORDO...
22	5021	20H09	9073	28/1/2007	5/12/2009	SCOTMID	9 Crookston Road	Glasgow	SCOTL...
23	5022	23290	846	28/1/2007	5/12/2009	SCOTMID	9 Crookston Road	G52 3QE	SCOTL...
24	5023	1887	2369	22/1/2007	26/3/2009	KMK SAVE	Bow Avenue	Ashton Lyne	NULL
25	5024	5944	8230	22/1/2007	26/3/2009	KMK SAVE	4 Bow Street	Ashton Under Ly...	NULL
26	5025	23221	517	11/2/2007	27/2/2009	Alcampo	114 Avenida San...	BARCELONA	NULL

Data
Objects are
listed in
your project

To view,
double click
on the link

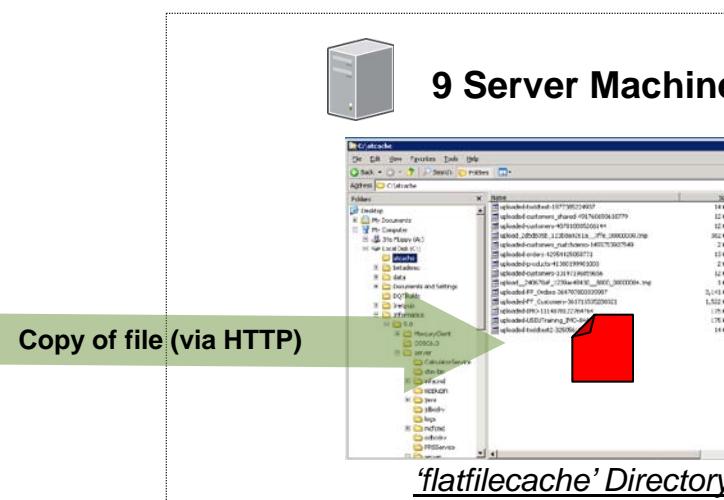
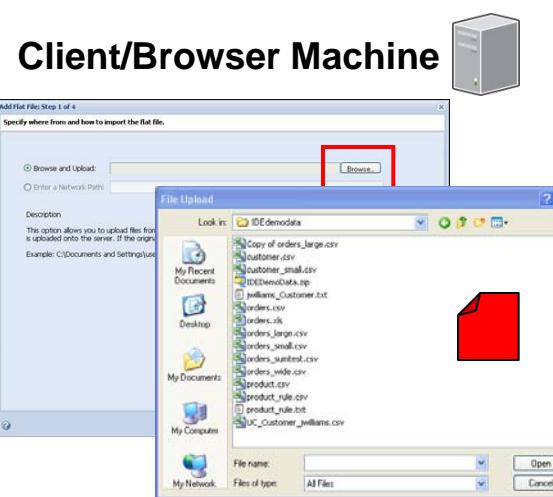
Flat Files

- Analyst enables any browser user to import flat files
- There are 2 import options for flat files:
 - Browse and Upload
 - Network path/shared directory



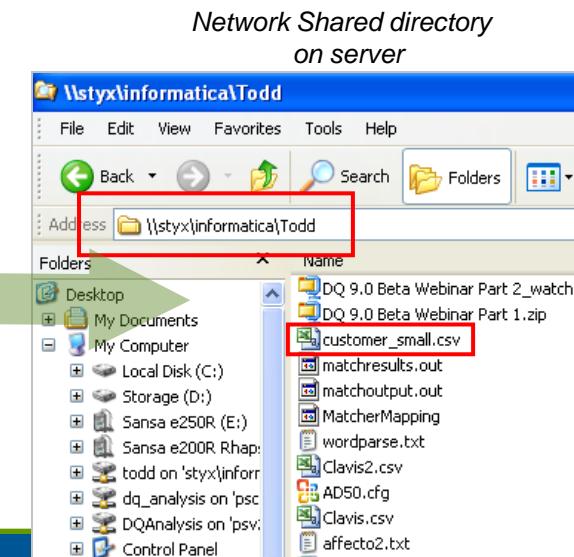
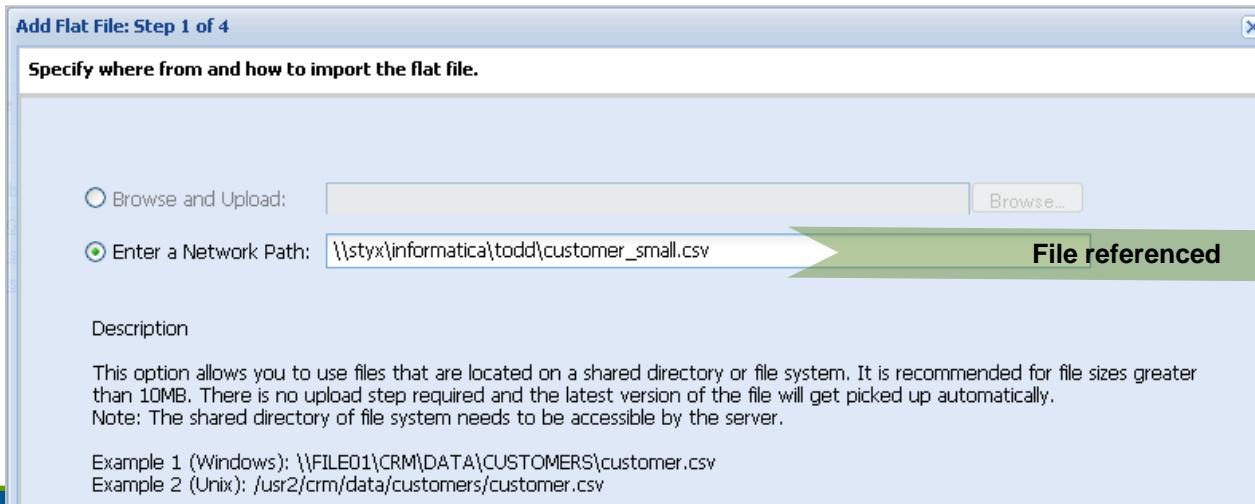
Flat Files - Browse and Upload

- The ‘Browse and upload’ action uploads the file via HTTP to the Server
 - A copy of the file
 - Preview or Profile references the uploaded/copied file not the original
 - Edits made to the local file will not be visible in Preview or Profile
 - Edits to the Uploaded file will be seen
 - Recommended option for files 10MB or smaller



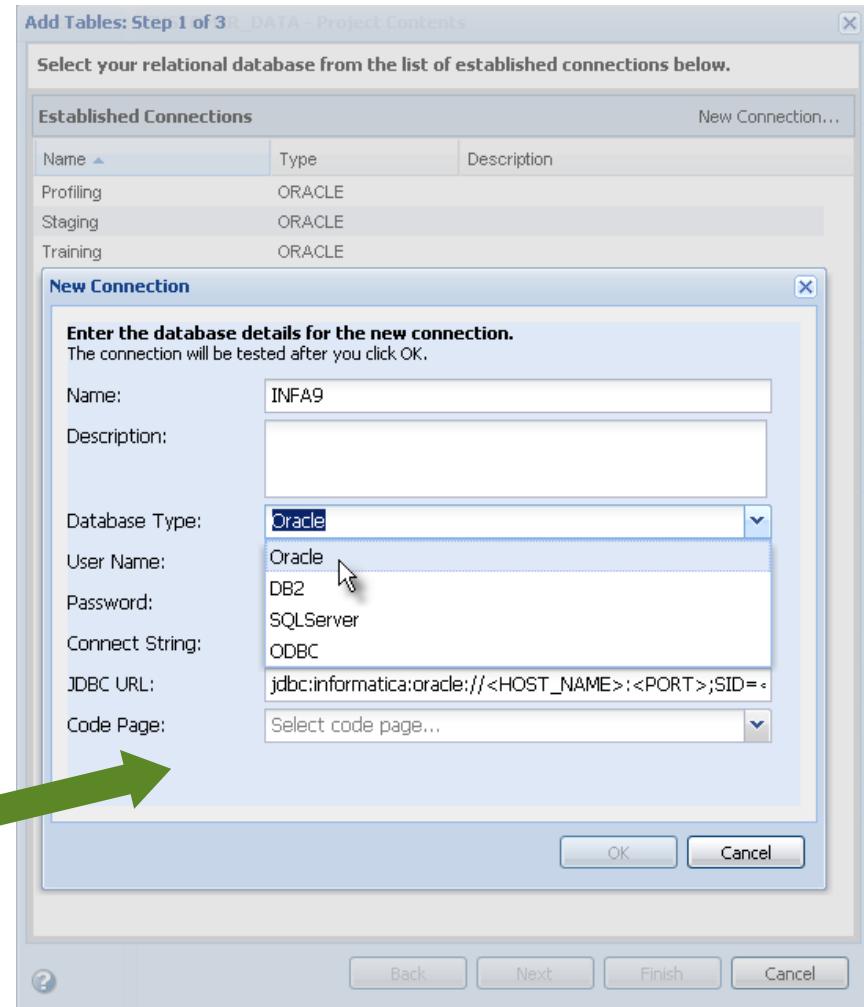
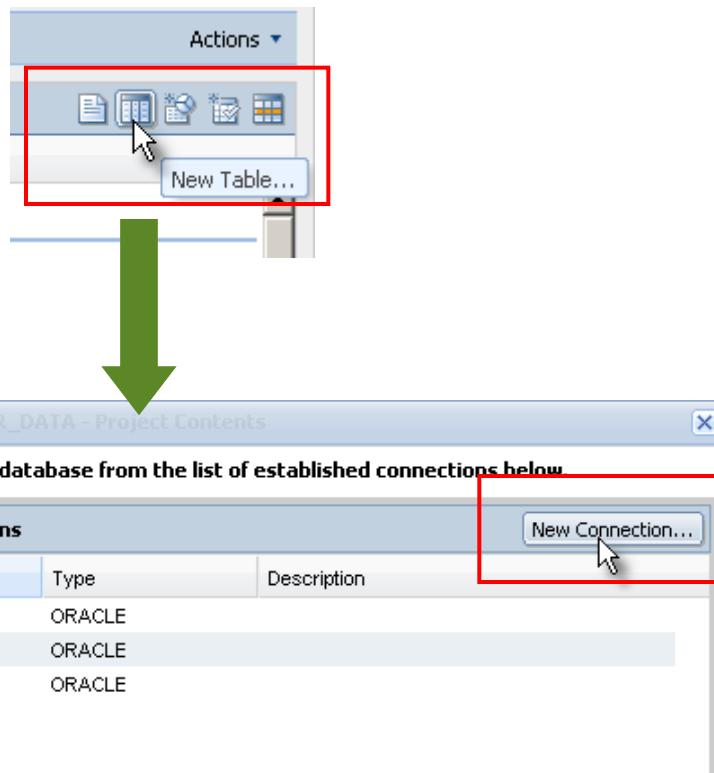
Flat Files - Network Path/Directory

- References files located in a shared directory or file system
 - Share is specific to Server machine - not browser client
 - No browse option for this reason
 - File referenced – no lag in time for ‘Upload’
 - Preview/Profile references the file on network share
 - Edits to the network shared file will be seen
 - Recommended option for files larger than 10MB



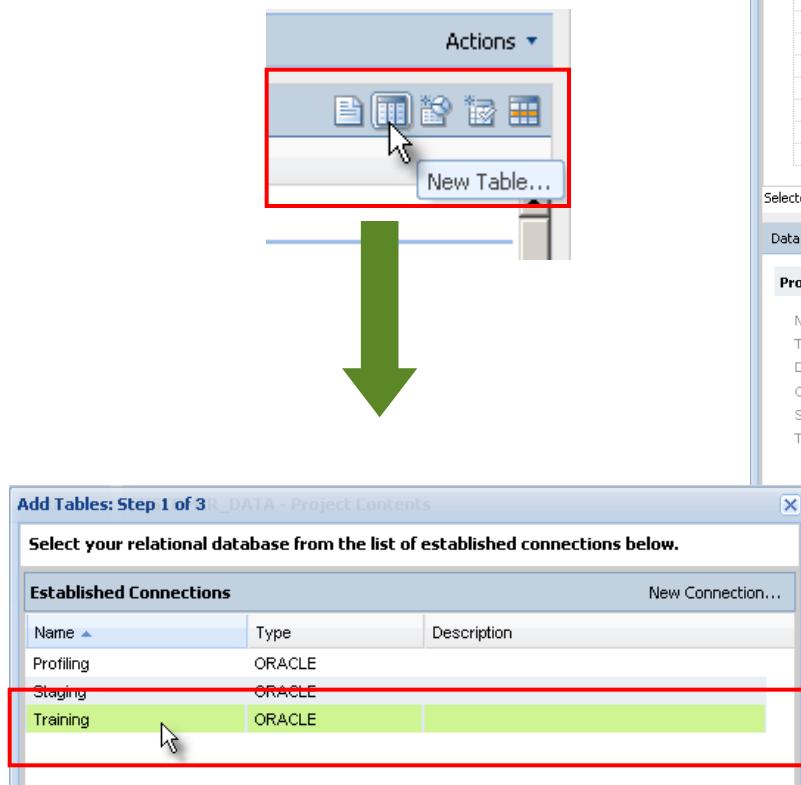
Relational Tables

- Analyst users can
 - Create new DB connections



Relational Tables

- Analyst users can
 - Leverage existing DB connections



Data Profiling

Why profile data?

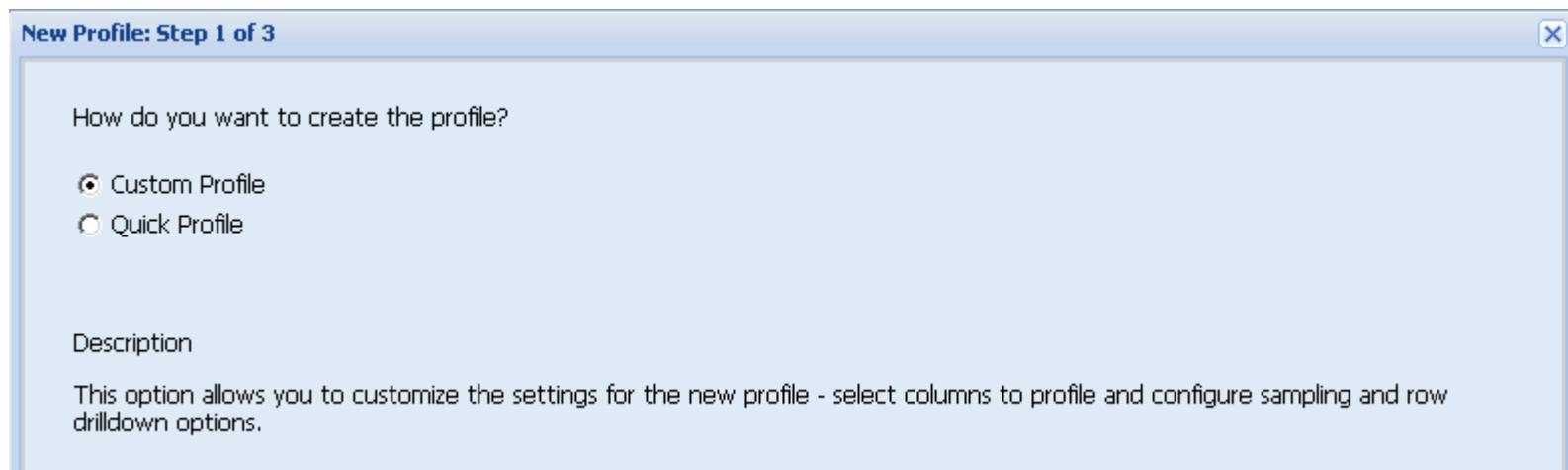
- Data profiling examines data in an existing data source, in order to identify possible data quality problems and issues that may exist.
- It collects statistics and information about the data to:
 - Assess the quality levels of the data, including whether the data conforms to particular standards or patterns.
 - Understand the type of data quality issues that exist.
 - Find out whether existing data can easily be used for other purposes.

Analyst Profiling

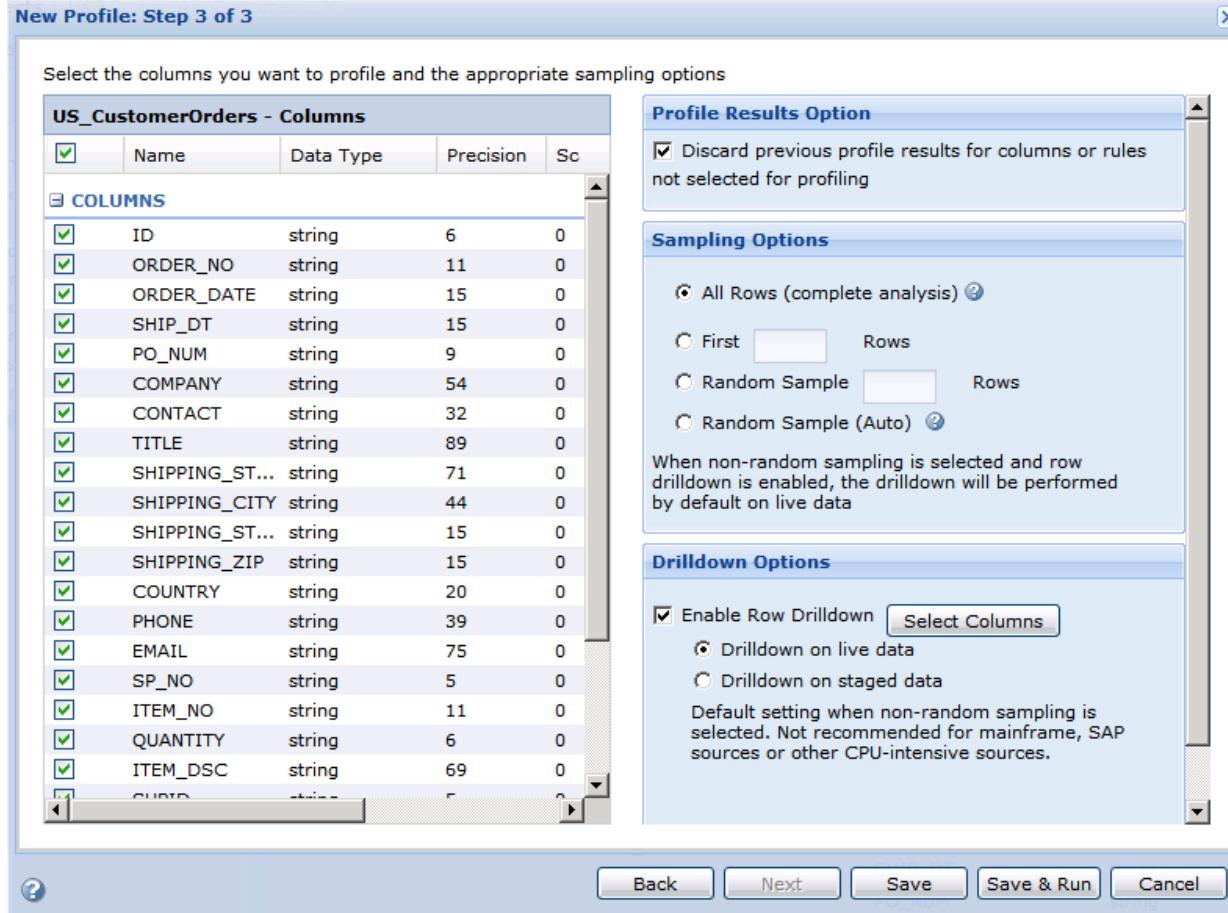
- There are two types of profiling available in the Analyst Tool:
 - Column and Rule Profiling
- Column Profiling:
 - A process of discovering physical characteristics of each column in a file.
 - It is the analysis of data quality based on the content and structure of the data.
 - Review Column profiling results to:
 - Identify possible anomalies in the data
 - Build reference tables
 - Apply or build Rules
 - Develop Scorecards

Column Profiling

- Two methods of creating profiles exist:
 - Quick Profile
 - Default Name Profile_
 - Profiles all columns and rows
 - Drill down on live data
 - Custom Profile
 - User can select settings



Custom Profile



- Specify *name and location*
- Select *columns to profile*
- Discard/keep profile results for columns not selected
- Select number of *rows to profile*
- Drilldown on live or staged data
 - Select Columns to view in drilldowns

Column Profiling

INFORMATICA Analyst

Browse: Projects **Profile_EU_Customer_Orders**

Administrator Log Out | Manage ▾ | Help ▾ Search All Go

Column Profiling Properties Data Preview

Profile_EU_Customer_Orders - Column Profiling

Name	Unique V...	Unique %	NULL	NULL %	Datatype	Inferred %	Documented Dat...	Max Value	Min Va
SOURCE NAME									
RECORD_ID	670	100.00	-	-	Integer(4)	100.00	string(6)	5669	2
ORDER_NO	668	99.70	-	-	String(11)	100.00	string(17)	M2295	111
PO_NUM	642	95.82	1	.15	Integer(6)	100.00	string(9)	99986	
ORDER_DATE	277	41.34	2	30	String(10)	100.00	string(15)	99/99/9...	1/1/20
SHIP_DATE	163								
COMPANY	162								
ADDR1	501	74.70	0	.30	String(52)	100.00	string(78)	Vivienda...	114 Ave
ADDR2	325	48.51	46	6.87	String(25)	100.00	string(38)	york	Kiddern
ADDR3	161	24.03	80	11.94	String(29)	100.00	string(44)	yorkshire	ALACA
ADDR4	390	58.21	20	2.99	String(9)	100.00	string(14)	YO30 4WZ	E7 8
COUNTRY_CODE	5	.75	-	-	String(3)	100.00	string(5)	UK	
CONTACT	577	86.12	-	-	String(36)			ZORAID...	ADALA
TITLE	65	9.70	7	1.04	String(38)	100.00	string(57)	Vice Pre...	Lead Pr
PHONE	366	54.63	12	1.79	String(20)	100.00	string(30)	Tel:+ 34...	230 980
EMAIL	360	53.73	18	2.69	String(46)	100.00	string(69)	zoraida....	@hiperc
SP_NO	42	6.27	5	.75	Integer(3)	100.00	string(5)	99	1
ITEM_NO	187	27.91	90	13.43	String(71)	100.00	string(11)	7T-26	45

Column & Rule Profiling

Drilldown: COUNTRY_CODE = 'ES'(All 29 rows)

	RECORD_ID	ORDER_NO	PO_NUM	ORDER_DATE	SHIP_DATE	COMPANY	ADDR1	ADDR2	ADDR3	ADDR4	COUNTRY_CODE	CONTACT
1	5000	2256K	2050	99/99/9999	18/1/2009	Hiperdino	Luis Cernuda	OLIVARES	SEVILLA	41804	ES	MR JAMES A
2	5001	23222	2466	11/2/2009	27/2/2009	Alcampo	Santa Coloma	BARCELONA		8030	ES	Miss GABI E
3	5002	2322K	173003	11/2/2009		Alcampo	Colombia	ELONA	BARCELONA	8030	ES	Ms GABI CI
4	5003	2288K	181	18/8/2009		Alcampo	Colombia	VISTA REUTARRAGONA		43206	ES	Mr CRISTIA
5	5004	2301K	2414	19/9/2009	14/5/2009	Alcampo	12 Granada	VILLAFRANCA DEBADAJOZ		6220	ES	MR. JOSHU
6	5005	2267K	529	22/12/2009	27/6/2009	Alcampo	Virgen De Los RIBA-ROJA DE TLA	VALENCIA		46190	ES	Mr. WILLIE
7	5006	23771K	897	22/12/2009	27/6/2009	Alcampo	14 Virgen De Los RIBA-ROJA DE TLA	VALENCIA		46190	ES	Mr RALPH P

Drilldown

Value/Patterns/Statistics

Drilldowns

INFORMATICA Analyst

Administrator Log Out | Manage | Help | Search All Go

Browse: Projects Profile_US_Customer_... Column Profiling Properties Data Preview

Profile_US_Customer_Orders - Column Profiling Sampling... Actions

Name	Unique V...	Unique %	NULL	NULL %	Datatype	Inferred %	Docu...
ID	2212	100.00	-	-	Integer(4)	100.00	integer
ORDER_NO	845	38.20	-	-	Integer(5)	98.33	string
ORDER_DATE	607	27.44	25	1.13	String(10)	100.00	string
SHIP_DT	486	21.97	33	1.49	String(10)	100.00	string
PO_NUM	810	36.62	8	.36	Integer(6)	100.00	string
COMPANY	270	12.21	-	-	String(36)	100.00	string
CONTACT	743	33.59	-	-	String(21)	100.00	string
TITLE	128	5.79	115	5.20	String(59)	100.00	string
SHIPPING_STREET	683	30.88	15	.68	String(47)	100.00	string
SHIPPING_CITY	406	18.35	29	1.31	String(21)	100.00	string
SHIPPING_STATE	51	2.31	46	2.08	Fixed Length Stri...	98.92	string
SHIPPING_ZIP	480	21.70	32	1.45	String(10)	100.00	string
COUNTRY	7	.32	479	21.65	Fixed Length Stri...	98.96	string
PHONE	680	30.74	49	2.22	String(26)	100.00	string
EMAIL	470	21.25	77	3.48	String(50)	100.00	string
SP_NO	41	1.85	-	-	Integer(3)	100.00	decimal

Drilldown:COUNTRY = 'null' OR 'U.S.A' OR 'US' OR 'AMERICA' OR 'US ' OR 'UNITED STATES'(First 100 rows only)

HIP_DT	PO_NUM	COMPANY	CONTACT	TITLE	SHIPPING_STREET	SHIPPING_CITY	SHIPPING_STATE	SHIPPING_ZIP	COUNTRY
'3/2010	9066	V.G.'s Food	ALICE CURLING	Business Analyst	7461 N Genesee	iGenesee	MI	48437	
'2/2010	477	A&P ..	Anna DYER	Sr Proj Manager		Little Silver	NJ	7739	
'1/2010	186	Publix SuperMark	Anna Ellis	Lead Sys Analyst		Owens Cross Roa	AL	35763-9204	U.S.A
'1/2010	8962	Schnucks	Anna FORD	Owner/Consultan		East Saint Louis	IL	62205	
'18/2010	2427	Winn Dixie Stores	Anna MORRISON	Info Systems Spe		GOLDSBORO	NC	27530	

Click on the Drilldown arrow in value frequencies to drill down to the associated records.

To drill down on multiple values select the values in the viewer, right click and choose Show Matching Rows

Column Profiling - Values

Value	Frequ...	Percent	Chart	Drilldown
USA	2188	75.92		
UK	476	16.52		
GBR	73	2.53		
ESP	66	2.29		
ES	29	1.01		
GB	26	0.90		
U.S.A	13	0.45		
US	7	0.24		
AMERICA	2	0.07		
US	1	0.03		
UNITED STATES	1	0.03		

Distinct values for the Column, with their frequencies

- *Value*: The column values in order of decreasing frequency.
- *Frequency*: The number of times each value appears
- *Percent*: The percentage that each value represents
- *Chart*: Bar graph representing the percentage of each value found
- *Drilldown*: Click the arrow to see the associated records

Column Profiling - Patterns

Pattern	Frequ...	Percent	Chart	Drilldown
XXX	2327	80.74		
XX	538	18.67		
X.X.X	13	0.45		
X(7)	2	0.07		
xxb	1	0.03		
X(6)bX(6)	1	0.03		

- **Patterns inferred for the Column, with their frequencies and the percentage of values matching each**
 - *Patterns*: The patterns that exist in each column
 - *Frequency*: The number of values in the data profiled which match each pattern
 - *Percent*: The percentage of the values in the data profiled which match each pattern
 - *Chart*: A bar graph representing the percentage of the data which match each pattern
 - Drilldown: Click the arrow to see the associated records

Column Profiling - Statistics

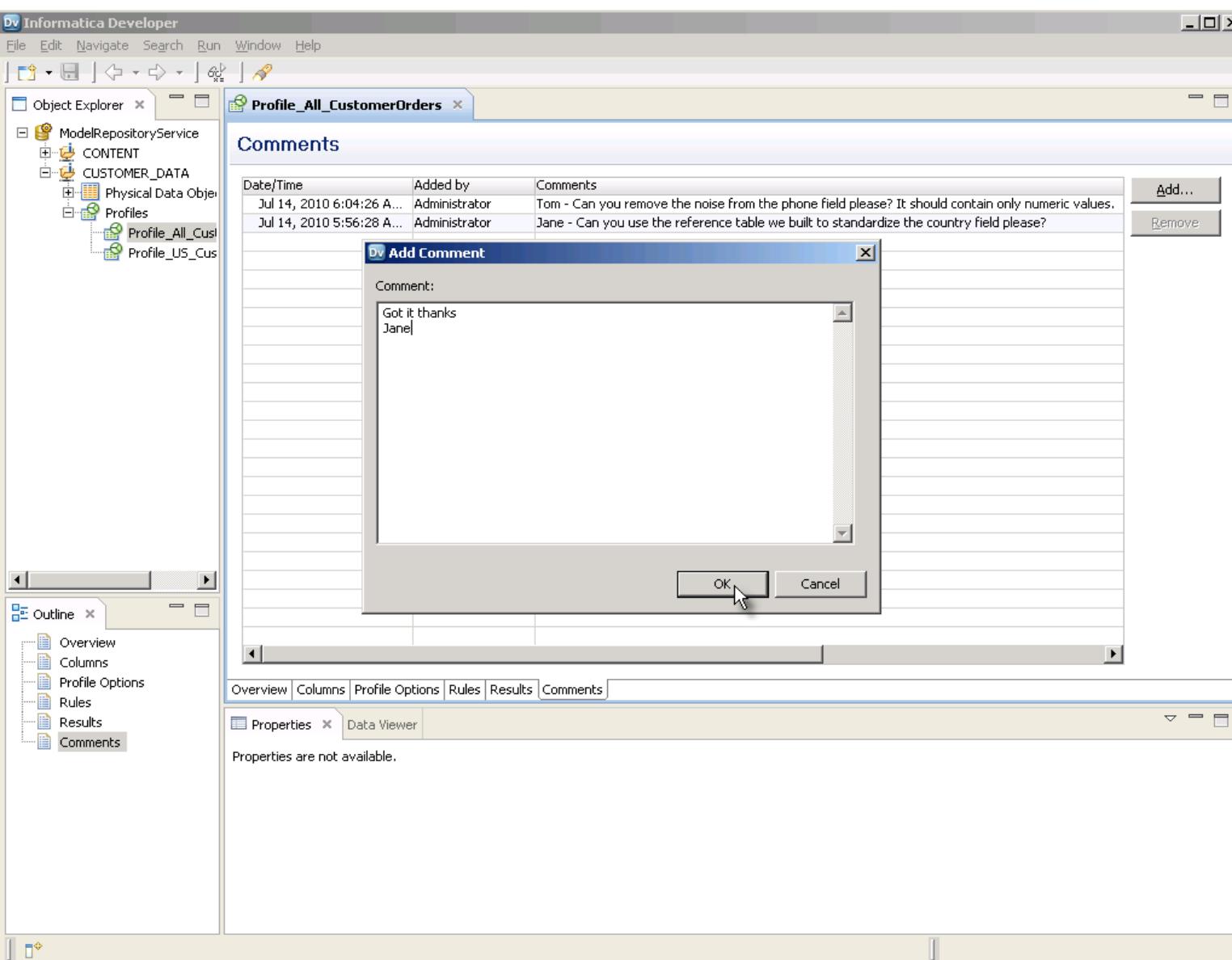
Statistics	
Statistic	Value
Maximum Length	13
Minimum Length	2
Bottom (5)	AMERICA ES ESP GB GBR
Top (5)	USA US us UNITED STATES UK

- The statistics include statistics about the column values, such as average, length, and top and bottom values.
 - Average: Average of the values for the column (integer).
 - Standard Deviation*: The variability between column values (integer).
 - Maximum Length*: Length of the longest value for the column.
 - Minimum Length*: Length of the shortest value for the column.
 - Bottom 5*: lowest values for the column.
 - Top 5*: highest values for the column.

Project Collaboration

- **Seamless collaboration between Analysts and Developers**
 - Projects created in either tool are visible in the other
 - Team members can easily communicate and share work & findings through comments, bookmarks, shared data profiles & data quality scorecards
 - Data can be easily exported from profiles or rules and emailed to the appropriate owner for review or correction

Collaboration - Comments



Analysts and Developers can use comments in profiles to collaborate on projects.

Lossless translation of information.

Collaboration - Exporting data

Browse: Projects Profile_US_Customer_.... x

Column Profiling Properties Data Preview Actions ▾

Profile_US_Customer_Orders - Column Profiling Sampling... Actions ▾

Name Unique V... Unique % NULL NULL % Datatype Inferred % Docu...

SOURCE NAME

ID	2212	100.00	-	-	Integer(4)	100.00	integer
ORDER_NO	845	38.20	-	-	Integer(5)	98.33	string
ORDER_DATE	607	27.44	25	1.13	String(10)	100.00	string
SHIP_DT	486	21.97	33	1.49	String(10)	100.00	string
PO_NUM	810	36.62	8	.36	Integer(6)	100.00	string
COMPANY	270	12.21	-	-	String(36)	100.00	string
CONTACT	743	33.59	-	-	String	100.00	string
TITLE	128	5.79	115	5.20	String	100.00	string
SHIPPING_STREET	683	30.88	15	.68	String	100.00	string
SHIPPING_CITY	406	18.35	29	1.31	String	100.00	string
SHIPPING_STATE	51	2.31	46	2.08	Fixe...	100.00	string
SHIPPING_ZIP	480	21.70	32	1.45	String	100.00	string
COUNTRY	7	.32	479	21.65	Fixe...	100.00	string
PHONE	680	30.74	49	2.22	String	100.00	string
EMAIL	470	21.25	77	3.48	String	100.00	string

Drilldown:COUNTRY = 'null' OR 'U.S.A' OR 'US' OR 'AMERICA' OR 'US' OR 'U.S.A' OR 'AMERICA'

Values Patterns Statistics

Value	Frequ...	Percent	Chart	Drilldown
USA	1709	77.26	Bar	▼
NULL	479	21.65	Bar	▼
U.S.A	13	0.59	Bar	▼
US	7	0.32	Bar	▼
AMERICA	2	0.09	Bar	▼
US	1	0.05	Bar	▼
UNITED S...	1	0.05	Bar	▼

Export data to a file String(31) 100.00 string US

File Name: Profile_US_Customer_Orders_COUNTRY_DRILLDC

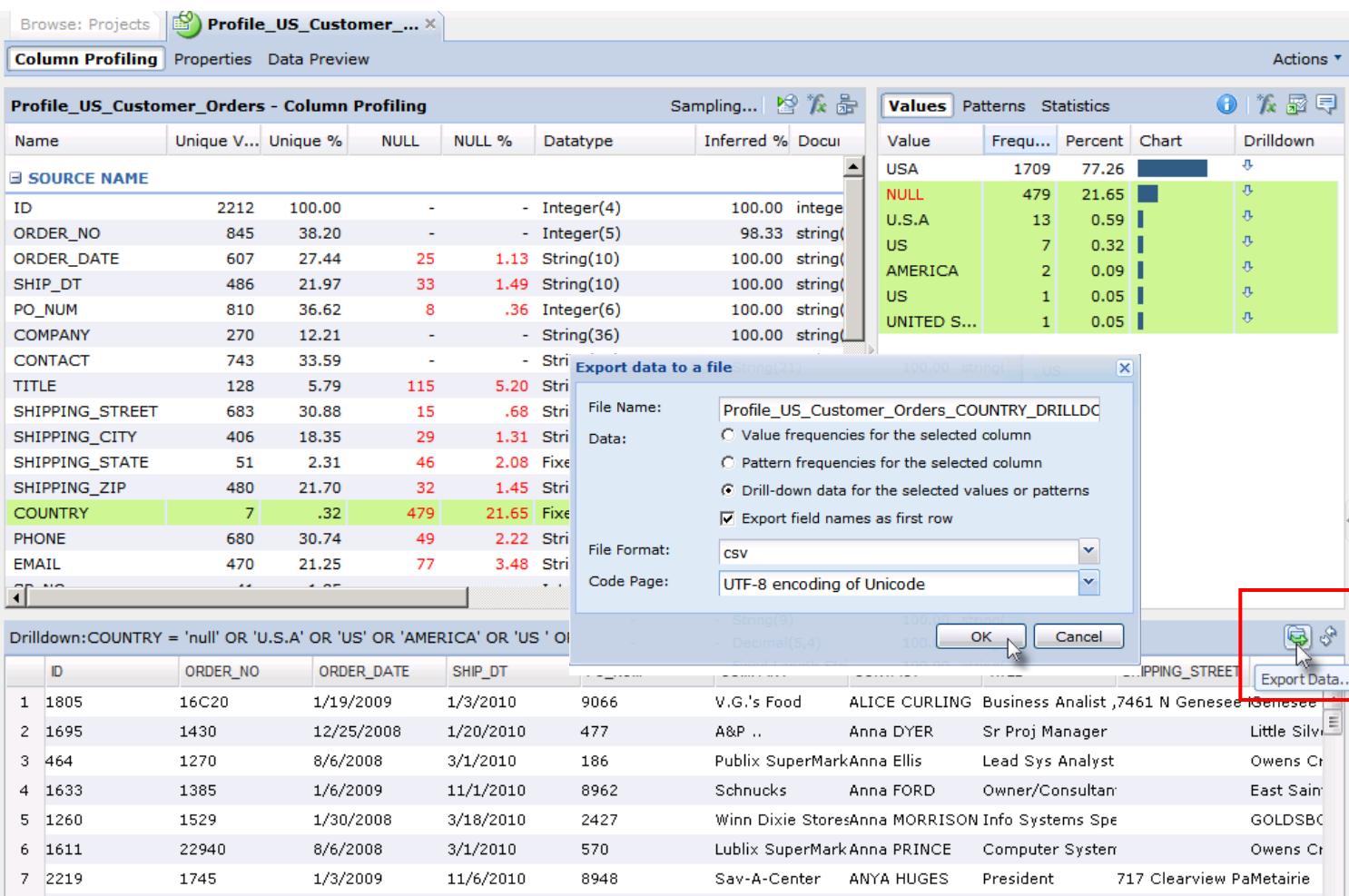
Data: Value frequencies for the selected column
 Pattern frequencies for the selected column
 Drill-down data for the selected values or patterns
 Export field names as first row

File Format: CSV

Code Page: UTF-8 encoding of Unicode

OK Cancel

Export Data...



To export drilldown results click on the Export Data button

Choose what you want to export:

Value frequencies
Pattern frequencies
Drill-down results

The file can sent to the appropriate data owner

Collaboration - Metadata Bookmarks(URLs)

The screenshot shows the Informatica Analyst interface. A browser window titled "Informatica Analyst - Windows Internet Explorer" is open, displaying a URL: http://infa-server:8085/AnalystTool/com.informatica.at.AnalystTool/index.jsp#p=U;glUc85vrEd-hHDerIG_Hjw&i=U;3UbC-8b6I. This URL is highlighted with a red box. Below the browser, the Informatica Analyst application interface is visible, showing a "Profile_US_Customer..." tab and a "Column Profiling" section for "Profile_US_Customer_Orders". A message window titled "Tax Rate and Bill Code columns - Message" is open, containing an email message. The message body reads:

Hey Tom
Can you take a look at the Tax Rate and Bill code columns and let me know if we need to keep them going forward?
Both only contain 1 value and NULL.
I've attached the link to the profile.
Thank you
Jen

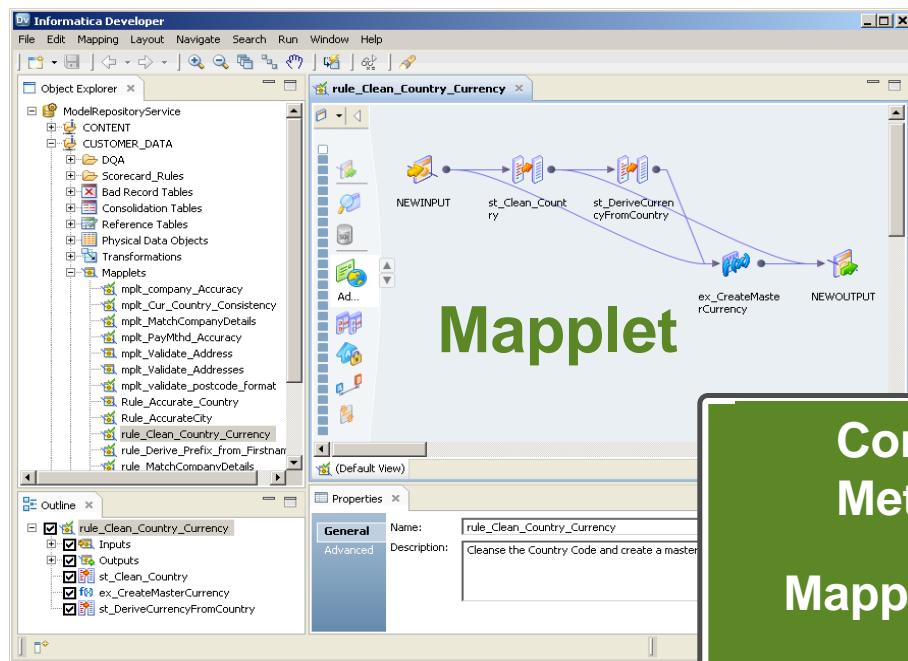
Below the message, a blue link is shown: <http://infa-server:8085/AnalystTool/com.informatica.at.AnalystTool/index.jsp#p=U;ZKW3a0POEd-xGiofrXW2eA&i=U;DW1flIgIEd-XqXSj9sgQWQ&c=com.informatica.profiling.services.model.persist.ProfileDefinition>. This link is also highlighted with a red box.

Collaboration via simple URL in email, portals, links in docs/specs, etc

HTTPS Protocol Supported

Metadata Bookmarks: All objects sharable via common metadata

High-Fidelity Collaboration



INFORMatica® Developer

Common
Metadata

Mapplet=Rule

Administrator Sign Off | User Preferences | Help ▾

rule_Clean_Country_Currency

Cleanse the Country Code and create a master Currency

Inputs

Name	Description	Data Type	Precision	Input Column
In_country		string	20	COUNTRY
In_currency		string	20	SHIP_DT PO_NUM COMPANY CONTACT TITLE SHIPPING_STREET SHIPPING_CITY SHIPPING_STATE SHIPPING_ZIP COUNTRY PHONE EMAIL SP_NO ITEM_NO QUANTITY

Outputs

Name	Description	Data Type	Precision
new_country		string	20
new_currency		string	10

Rule

Drilldown:TAX_RATE

ID

1

2

3

4

5

1489

23407

4/28/2007

3/13/2010

5340

Super Fresh ALAYNE YOSHII... Owner / Consult... 25 Greentree Dr...

272

1275

99/99/9999

3/20/2010

156350

Publix SuperMarket ALCOT PREVOST MIS Manager 4851 Rice Mine R...

< Back

Next >

Cancel

INFORMatica® Analyst

INFORMatica®

Rule Profiling

- A Rule is a constraint written against data that is used to identify possible inconsistencies in the data.
 - Rule creation and editing (Expression based)
 - Leveraging OOTB Rules / Developer created rules
- Join Analysis and mid stream profiling are preformed in the Developer Tool only

Rule Profiling

INFORMATICA Analyst

Browse: Projects Profile_US_Customer_... x

Administrator Log Out | Manage | Help

Column Profiling Properties Data Preview

New Rule : Step 2 of 4

Select the rule you want to use

Rules

- Projects
- CONTENT
 - Dictionaries
 - Rules
 - Address_Data_Cleansing
 - Contact_Data_Cleansing
 - Corporate_Data_Cleansing
 - General_Data_Cleansing
 - rule_Remove_Period
 - rule_Parse_Text_in_Single_Quotes
 - rule_Parse_Number_At_Start_Of_Line
 - rule_Remove_Quotation
 - rule_Add_Space_Around_Ampersand
 - rule_Remove_Apostrophe
 - rule_Date_Parse
 - rule_UpperCase
 - rule_Add_Spaces_Around_Period
 - rule_Parse_Text_Between_Parentheses
 - rule_Completeness
 - rule_Reverse_String_Input
 - rule_Remove_Punctuation
 - rule_IsNumeric
 - rule_Luhn_Algorithm

- **Apply rules within profiles and analyze results in-line with original source data**

Select from one of the prebuilt rules or create your own

Apply Rules to the profile

INFORMATICA Analyst

Administrator Log Out | Manage | Help | Search All Go

Browse: Projects Profile_US_CustomerO... X

Column Profiling Properties Data Preview Actions

Profile_US_Customer_Orders - Column Profiling Sampling... Actions

Name	Unique V...	Unique %	NULL	NULL %	Datatype	Inferred %	Documented C
SUPID	16	.72	-	-	Integer(3)	100.00	decimal(3)
UNIT_COST	80	3.62	-	-	String(9)	100.00	string(9)
TAX_RATE	1	.05	-	-	Decimal(5,4)	100.00	string(6)
BILLCODE	1	.05	-	-	Fixed Length Stri...	100.00	string(10)
PAY_MTHD	18	.81	-	-	String(17)	100.00	string(17)
PAY_TERM	9	.41	107	4.84	Integer(3)	100.00	string(3)

Rule: rule_Date_Validation

Conformant_OrderDate	3	.14	-	-	String(23)	100.00	string(25)
----------------------	---	-----	---	---	------------	--------	------------

Rule: rule_Date_Validation [2]

Conformant_Shipping...	3	.14	-	-	String(23)	100.00	string(25)
------------------------	---	-----	---	---	------------	--------	------------

Rule: rule_Email_Validation

Conformant_Email	2	.09	-	-	Fixed Length Stri...	96.20	string(10)
------------------	---	-----	---	---	----------------------	-------	------------

Rule: rule_IsNumeric

Conformant_OrderNo	2	.09	-	-	Fixed Length Stri...	98.60	string(10)
--------------------	---	-----	---	---	----------------------	-------	------------

Rule: rule_IsNumeric [2]

Conformant_Phone	2	.09	-	-	Fixed Length Stri...	98.92	string(10)
------------------	---	-----	---	---	----------------------	-------	------------

Drilldown:Conformant_Email = 'Invalid'(All 84 rows)

CONTACT	TITLE	SHIPPING_STREET	SHIPPING_CITY	SHIPPING_STATE	SHIPPING_ZIP	COUNTRY	PHONE	EMAIL	SP_NO
BARRY PUGH	Automated System	Gilmore Rd	Fairfield	OH	45014	USA	(513) 881-0100		77
BELA HELBLING	Owner	377 Chestnut St	Needham	MA	2492		781-444-0411		101
Bill Hatch	Programmer	5AN AP6A-2 100	Abbott Park	IL	60064-3500	USA	708-937-6505		80
Bill Hatch	Programmer	103 Abbott Park	Abbot Park	IL	60064-3500		708-937-6501		95
Bill Hatch	Programmer	5AN AP6A-2 100	Abbott Park	IL	60064-3500	USA	708-937-6501		100
Chet Galek	Business Analyst	235 East 42nd St	New York	NY	10017		Main: 212-573-72		76

Apply the rules to the profile

Run the profile to view the results

Value Frequency Rules

INFORMATICA Analyst

Browse: Projects Profile_All_Customer... Column Profiling Properties Data Preview

Profile_All_CustomerOrders - Column Profiling Sampling... Actions Values Patterns Statistics

New Rule : Step 2 of 3

Create the expression you want to use

Name *: rule_ValidCountries Save as a reusable rule in: /CONTENT/Rules

Description:

Functions Columns

- All Functions
- Character
- Conversion
- Data Cleansing
- Date
- Decision
- Encoding
- Financial
- Numerical
- Scientific
- Special
- Test

Expression Validate

```
COUNTRY='USA' OR COUNTRY='GBR' OR  
COUNTRY='ESP'
```

AND OR NOT () = != < <= >= >

Return Type Details Edit...

Return Type N/A
Precision 0
Scale 0

< Back Next > Save Save & Run Cancel

The screenshot shows the INFORMATICA Analyst interface. A 'New Rule : Step 2 of 3' dialog box is open, overlaid on a 'Column Profiling' window for 'Profile_All_CustomerOrders'. The dialog box has a red border and contains fields for 'Name' (set to 'rule_ValidCountries') and 'Description'. It also includes a 'Save as a reusable rule in:' checkbox and a dropdown menu. Below these are sections for 'Functions' (with a tree view of available functions) and 'Columns'. An 'Expression' section contains the SQL-like expression 'COUNTRY='USA' OR COUNTRY='GBR' OR COUNTRY='ESP'. At the bottom of the dialog are 'Return Type Details' and buttons for navigation and saving. The background shows the main interface with various tabs like 'Sampling...', 'Values', 'Patterns', and 'Statistics'.

Select the value frequency results to include in the Rule, right click and choose Add Rule

Choose to create a Value Frequency Rule

The expression is written based on your selection

Can be reusable

Run profile (on all or just the rule column)

Value Frequency Rules

INFORMATICA Analyst

Browse: Projects Profile_All_Customer... ×

Administrator Log Out | Manage ▾ | Help ▾ Search All Go

Column Profiling Properties Data Preview Actions ▾

Profile_All_CustomerOrders - Column Profiling Sampling... Values Patterns Statistics

Name	Unique V...	Unique %	NULL	NULL %	Datatype	Inferred %	Document
COUNTRY	11	.38	-	-	String(13)	100.00	string(13)
CONTACT	1320	45.80	-	-	String(36)	100.00	string(52)
TITLE	133	4.61	122	4.23	String(59)	100.00	string(59)
PHONE	1043	36.19	61	2.12	String(26)	100.00	string(26)
EMAIL	828	28.73	95	3.30	String(50)	100.00	string(50)
SP_NO	42	1.46	5	.17	Integer(3)	100.00	string(3)
ITEM_NO	221	7.67	189	6.56	String(7)	100.00	string(7)
QUANTITY	565	19.60	9	.31	Integer(4)	100.00	decimal(4)
ITEM_DSC	458	15.89	-	-	String(46)	100.00	string(46)
SUPID	18	.62	2	.07	Integer(3)	100.00	string(6)
UNIT_COST	117	4.06	-	-	String(9)	100.00	string(11)
PAY_MTHD	19	.66	1	.03	String(17)	100.00	string(17)
PAY_TERM	9	.31	108	3.75	Integer(3)	100.00	string(3)
CURRENCY	7	.24	1	.03	Fixed Length Str...	98.61	string(9)

Rule: rule_ValidCountries

rule_ValidCountries	2	.07	-	-	Integer(1)	100.00	integer(10)
---------------------	---	-----	---	---	------------	--------	-------------

Drilldown:rule_ValidCountries = '0'(First 100 rows only)

PO_NUM	ORDER_DATE	SHIP_DT	COMPANY	ADDR1	ADDR2	ADDR3	ADDR4	COUNTRY	CON
186	8/6/2008	3/1/2010	Publix SuperMark	Owens Cross RoaAL			35763-9204	U.S.A	Anna
2050		1/18/2009	Hiperdino	Luis Cernuda	OLIVARES	SEVILLA	41804	ES	MR. J
2466	2/11/2009	2/27/2009	Alcampo	Santa Coloma 11-BARCELONA	BARCELONA		8030	ES	Miss
173003	2/11/2009	2/27/2008	Alcampo	Santa Coloma 11-BARCELONA	BARCELONA		8030	ES	Ms G
181	8/18/2009	1/25/2009	Aldi	11 Almirall RequeSOL I VISTA REUTARRAGONA			43206	ES	Mr C
2414	9/19/2009	5/14/2009	Alcampo	12 Granada	VILLAFRANCA DEBADAJOZ		6220	ES	MR.

After running the profile click on the new frequency rule created

1: represents the records that met the criteria

0: represents the records that did not meet the criteria

The rule will be available as a maplet in the Developer tool

Reference Table Management

What are Reference Tables?

- Reference tables enable validation, parsing, enrichment and enhancement of data.
- Reference data can include accurate and standardization values that can be used by analysts and developers in cleansing and validation rules.
- Create, edit, and import data quality dictionary files as reference tables.

Sample Reference Table

INFORMATICA Analyst

Administrator Sign Off | User Preferences | Help ▾

Search All Go

Browse: Projects Standard_Address_... X

Reference Table Audit Trail Properties Actions ▾

Standardize_Address_Suffix - Reference Table

Find [] in All [] Replace with [] Next [] Previous [] Highlight All Replace Replace All (Use Empty Boxes To Look For)

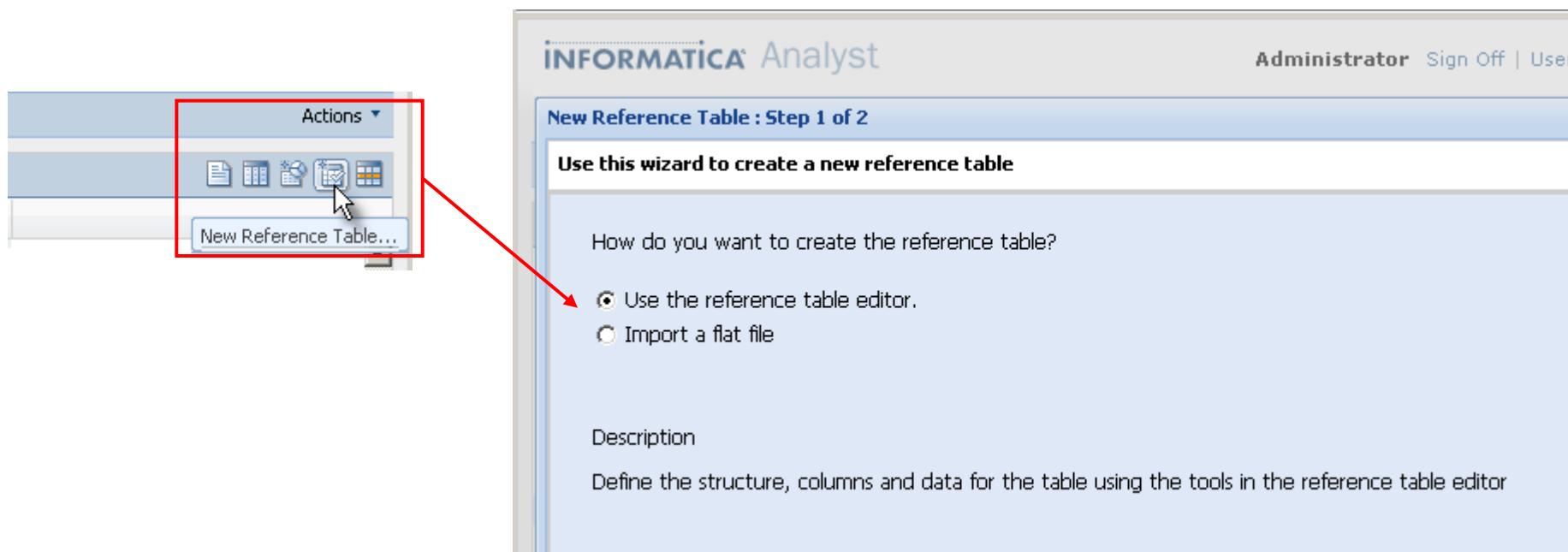
	COLUMN1	COLUMN2	COLUMN3	COLUMN4	COLUMNS5	COLUMN6
< click to add a new row >						
<input type="checkbox"/>	ACRS	ACRS	ACRES	NULL	NULL	NULL
<input type="checkbox"/>	AVE	AVE	AVENUE	AV	AVE.	AV.
<input type="checkbox"/>	BNK	BNK	BANK	NULL	NULL	NULL
<input type="checkbox"/>	BLV	BLV	BOULEVARD	BLVD	NULL	NULL
<input type="checkbox"/>	BLD	BLD	BUILDING	NULL	NULL	NULL
<input type="checkbox"/>	BLDS	BLDS	BUILDINGS	NULL	NULL	NULL
<input type="checkbox"/>	CSTL	CSTL	CASTLE	NULL	NULL	NULL
<input type="checkbox"/>	CNTR	CNTR	CENTRE	CTR	CTR.	NULL
<input type="checkbox"/>	CNTR	CNTR	CHASE	NULL	NULL	NULL
<input type="checkbox"/>	CLNC	CLNC	CLINIC	NULL	NULL	NULL
<input type="checkbox"/>	CL	CL	CLOSE	CL.	NULL	NULL
<input type="checkbox"/>	CTGE	CTGE	COTTAGE	NULL	NULL	NULL
<input type="checkbox"/>	CT	CT	COURT	CRT	CT.	NULL
<input type="checkbox"/>	CRES	CRES	CRESCENT	CRESENT	CRES.	CRS.

- Use the icons to find, edit and modify the data and the reference table

How to create Reference Tables

- **Reference Tables are created in the Analyst Tool and also in the Developer Tool and can be created:**
 - using the reference table editor
 - by importing a flat file
 - from a column profile
- **They can be edited to add columns and rows, or make changes to the data values.**
 - Search and replace values
 - Editing activities tracked in the audit trail log
 - View properties for the reference table in the Properties view

How to create Reference Tables



Reference Table Editor

New Reference Table : Step 2 of 2

Give Reference Table Details

*Table Name: Country_Codes

Description: Valid country codes

Default Value:

Column Attributes

	Valid	Name	Data Type	Precision	Scale	Description	Delete
1	<input checked="" type="radio"/>	column1	string	10	N/A		X
2	<input type="radio"/>	column2	string	10	N/A		X

1. Define the table structure

INFORMATICA Analyst Administrator Sign Off | User Preferences | Help ▾ Search All Go

Browse: Projects Country_Codes

Include a column for row-level de

Audit Note:

Reference Table Audit Trail Properties Actions

Country_Codes - Reference Table

	column1	column2	
	USD		✓ X

2. Add the data values

Import Flat File

New Reference Table : Step 2 of 4

Select the flat file to add.

Locate a file and upload it for preview

*File Name:

Enter name and description for the new flat file source

*Name: →

Description:

Default Value:

Choose the format that best describes your data

Code Page: →

Preview of file C:\INFA_Shared\Dictionary\Training\country_code.dic

12	ASM,ASM,AS,AMERICAN SAMOA
13	ATA,ATA,AQ,ANTARCTICA
14	ATF,ATF,TF,FRENCH SOUTHERN TERRITORIES
15	ATG,ATG,AG,ANTIGUA AND BARBUDA
16	AUS,AUS,AU,AUSTRALIA,COMMONWEALTH OF AUSTRALIA
17	AUT,AUT,AT,AUSTRIA,REPUBLIC OF AUSTRIA
18	AZE,AZE,AZ,azerbaijan,REPUBLIC OF AZERBAIJAN
19	BDI,BDI,BI,BURUNDI,REPUBLIC OF BURUNDI
20	BEL,BEL,BE,BELGIUM,KINGDOM OF BELGIUM

Back

Browse and Upload the file

Enter Name

Define Code Page

Scorecarding

What are Data Quality Scorecards?

- A scorecard is the graphical representation of valid values for a column in a profile.
- Scorecards can be easily shared with Stakeholders via a URL.
- Further DQ rules can be created in the Developer and applied to the profile in the Analyst Tool.
- Use scorecards to measure data quality progress.

Data Quality Scorecards

- Scores based on value frequencies
 - Includes ‘Virtual Columns’ – output of any rule
- Single scorecard supports scores from multiple Data Objects
- Scores added to scorecard via profiles:
 - Are not connected to the profile(s) from which column/virtual column originated from.
 - Delete the profile without impacting the scorecard
 - Deleting the source would invalidate both the profile and the scorecard

Scorecard - Valid Values per column

Browse: Projects Profile_US_Customer_... CLOSE

Column Profiling Properties Data Preview

Profile_US_Customer_Orders - Column Profiling Sampling... Document Values Patterns ...

Name Unique V... Unique % NULL NULL % Datatype Inferred % Document

Add to Scorecard...

Name	Unique V...	Unique %	NULL	NULL %	Datatype	Inferred %	Document
ID	2212	100.00	-	-	Integer(4)	100.00	integer(10)
ORDFR NO	845	38.20	-	-	Integer(5)	98.33	string(6)

Run Profile and select Add to Scorecard

Add to Scorecard: Step 1 of 3 100.00 Jun 28, 2010 9:45:58 AM CDT CLOSE

Confirm the profile columns that you want to add as scores to the scorecard. For a quicker setup, accept the default Score Names (copies from the columns) and skip the Description. You can always set these up later.

Columns

Column Name	Score Name	Description
SOURCE NAME		
BILLCODE	BILLCODE	
COMPANY	COMPANY	
CONTACT	CONTACT	
<input checked="" type="checkbox"/> COUNTRY	Accurate_Country	
EMAIL	EMAIL	
ID	ID	
ITEM_DSC	ITEM_DSC	
ITEM_NO	ITEM_NO	
ORDER_DATE	ORDER_DATE	
ORDER_NO	ORDER_NO	
<input checked="" type="checkbox"/> PAY_MTHD	Accurate_PayMethod	
<input checked="" type="checkbox"/> PAY_TERM	Accurate_PayTerm	
PHONE	PHONE	
PO_NUM	PO_NUM	
QUANTITY	QUANTITY	
SHIPPING_CITY	SHIPPING_CITY	
SHIPPING_STATE	SHIPPING_STATE	
SHIPPING_STREET	SHIPPING_STREET	

Back Next Finish Cancel

Add and rename columns/rules you want to include in the scorecard

Scorecard - Valid Values per column

Add to Scorecard: Step 3 of 3 100.00 Jun 28, 2010 9:45:53 AM CDT

Select each score and adjust its settings. Scores in the scorecard are the percentage of valid values.

Scores

Score Name	Column Name	Remove
Accurate_Country	COUNTRY	
Accurate_PayMethod	PAY_MTHD	
Accurate_PayTerm	PAY_TERM	
Conformant_Email	Conformant_Email	
Conformant_OrderDate	Conformant_OrderDate	
Conformant_OrderNo	Conformant_OrderNo	
Conformant_Phone	Conformant_Phone	
Conformant_ShippingDate	Conformant_ShippingDate	

Score Settings: COUNTRY

Set Custom Thresholds for this Score

95	to	100	%	Good
85	to	94	%	Acceptable
0	to	84	%	Unacceptable

Score using: Values

Value	Frequency	Is valid
USA	1709	<input checked="" type="checkbox"/>
NULL	479	<input type="checkbox"/>
U.S.A	13	<input type="checkbox"/>
US	7	<input type="checkbox"/>
AMERICA	2	<input type="checkbox"/>
UNITED STATES	1	<input type="checkbox"/>
US	1	<input type="checkbox"/>

Back Next Finish Cancel

Create/add to an existing scorecard/group

Select the valid values from the frequency list for the column

Once completed choose Finish

Scorecards can be modified after creation

Scorecard - Out Of The Box Rules

Add to Scorecard: Step 3 of 3

Select each score and adjust its settings. Scores in the scorecard are the percentage of valid values.

Score Name	Column Name
Conformant_Email	Conformant_Email

Score using: Values

Value	Frequency	Is Valid
Valid	2850	<input checked="" type="checkbox"/>
Invalid	30	<input type="checkbox"/>

Score Settings: Conformant_Email

Score Thresholds Set Custom Thresholds for this Score

67	to	100	%	Good
33	to	66	%	Acceptable
0	to	32	%	Unacceptable

Back Next Finish Cancel

67 to 100 % Good
33 to 66 % Acceptable
0 to 32 % Unacceptable

Add the rule to the profile

From the profile add the measure to your scorecard

In the Scorecard, select the valid/true value

Scorecard – Custom Rules

Edit Scorecard

Scores		
Score Name ▾		Remove
Accurate_Company		
Accurate_Country		
Accurate_Currency		
Accurate_PayMethod		
Accurate_PayTerm		
Conformant_Phone		
Conformant_PONum		
Conformant_Postcode		
Consistant_Cur_Country		

Score Settings: Consistant_Cur_Country

Score Thresholds Set Custom Thresholds for this Score

80	to	100	%	Good
51	to	79	%	Acceptable
0	to	50	%	Unacceptable

Save Cancel

Score using: Values		
Value	Frequency	IsValid
Consistent_Cur_Country	2416	<input checked="" type="checkbox"/>
Inconsistent_Cur_Country	512	<input type="checkbox"/>

Build the rule in Developer and “Validate as a Rule”

Add the rule to the profile and from the profile add the measure to your scorecard

In the Scorecard, select the valid values

Edit the Scorecard to move the measures into the Group

Scorecard

INFORMATICA Analyst

Browse: Projects Profile_US_Customer_... x US_Customer_Scorecar... x

Administrator Log Out | Manage | Help | Search All Go

Scorecard

US_Customer_Scorecard - Scores

Last Run On: Jun 28, 2010 10:50:10 AM CDT

Name	Total R...	Invalid...	Score(...)	Score Bar	Data Object	Source	Source Type	Dr
Conformant_Email	2212	84	96	<div style="width: 96%; background-color: green;"></div>	Conformant_Email	Reusable Rule		
Conformant_OrderDate	2212	441	80	<div style="width: 80%; background-color: red;"></div>	Conformant_OrderDate	Reusable Rule		
Conformant_OrderNo	2212	32	98	<div style="width: 98%; background-color: green;"></div>				
Conformant_Phone	2212	2191	0	<div style="width: 0%; background-color: lightgreen;"></div>				
Conformant_ShippingDate	2212	430	80	<div style="width: 80%; background-color: red;"></div>				
Accurate_Country	2212	503	77	<div style="width: 77%; background-color: red;"></div>				
Accurate_PayMethod	2212	426	80	<div style="width: 80%; background-color: red;"></div>				
Accurate_PayTerm	2212	184	91	<div style="width: 91%; background-color: yellow;"></div>				

Conformity

Accuracy

Trend Chart Detail

Accurate_PayTerm 91 Rose 91 points and became Acceptable on Jun 29, 2010 7:46:47 AM

Good
Acceptable
Unacceptable

Valid **Invalid**

Drilldown: Conformant_Phone = 'True' (All 21 rows)

ID	ORDER_NO	ORDER_DATE	SHIP_DT	PO_NUM
1	1846	23332	3/18/2009	8/6/2010
2	467	2312M	9/14/2008	5/14/2010
3	1485	1369	9/14/2008	5/14/2010
4	2380	23152	10/5/2009	4/18/2010
5	220	23361	3/27/2009	11/1/2010
6	1635	1370	3/27/2009	11/1/2010
7	530	2061	5/7/2009	3/28/2010

Close

SHIPPING_STATE SHIP

C T 601
H 324
H 324
H 218
NIA 218

Shaw's. CHAD KNOX Computer Systemized P.O. Box 260 BRAUNFELD, NH
CORILLA STOUGHTON Sys Analyst Pr276 West Main St HILLSBOROUGH, NH
CORILLA STOUGHTON Database Admin:276 West Main St HILLSBOROUGH, NH
The Carpet Store DEMITRI LEE Manager #4 of D102 Mid St South Bedhill, SURRY, NH
CORILLA STOUGHTON Sys Analyst Pr276 West Main St HILLSBOROUGH, NH

Informatica Developer Overview

Informatica Developer GUI

The screenshot shows the Informatica Developer GUI with several windows open:

- Object Explorer**: Located in the bottom-left corner, it displays a tree view of objects in the ModelRepositoryService, including CONTENT, CUSTOMER_DATA, Physical Data Objects (All_CustomerOrders, EU_Customer_Orders, US_Customer_Orders), Mapplets (rule_EUtoUS_DateFormat, rule_LeadZeroFiveDigitZip, rule_PayTerm_Std, rule_Retain_Numbers), Mappings (m_1_Join_CustomerOrders), Profiles, and Scorecards.
- Editor**: The central window shows a mapping diagram with four components: Customer_Detail, kg_Generate_Keys, mt_Match_Data, and MatchData, connected by arrows.
- Properties**: A floating window showing properties for the 'mt_Match_Data' component. It includes sections for General, Ports, Match Type, Strategies, Match Output, and Advanced. Under Strategies, 'Define match' is selected, and under Match Strategy, 'Bigram Distance' is chosen with a value of 0.9. Other options like Edit Distance, Edit Distance, Hamming Distance, and Edit Distance are listed with lower values.
- Connection Explorer**: Located in the top-right, it shows a tree view of a Domain, specifically the Training section, which includes IDQ, Tables (CUSTOMER, CUSTOMERDUP, CUSTOMER_ISSUE), Views, and Synonyms.
- Multiple objects can be opened simultaneously**: A callout bubble points to the Connection Explorer window, indicating that multiple objects can be viewed at once.
- Properties**: Another floating window showing properties for the 'Run Mapping' button. It lists Match Fields and Properties for COMPANY_1, COMPANY_2, ADDRESS_1_1, ADDRESS_1_2, ADDRESS_2_1, ADDRESS_2_2, ADDRESS_3_1, ADDRESS_3_2, ADDRESS_4_1, ADDRESS_4_2, and COUNTRYDV_1, COUNTRYDV_2. The properties for ADDRESS fields show 'Both Null Match Value: 0.5, Nu'.
- Bottom Left Buttons**: A red box highlights a set of small, semi-transparent buttons in the bottom-left corner, likely for quick access to common functions.

Informatica Developer GUI

The screenshot shows the Informatica Developer GUI interface. At the top, there's a menu bar with File, Edit, Mapping, Layout, Navigate, Search, Run, Window, and Help. Below the menu is a toolbar with various icons. The main workspace contains three tabs: m_2_Standardize_Customer_Orders, m_4_MatchCustomer, and Connection Explorer. The Connection Explorer tab is currently active, showing a tree view of domains, training, IDQ, tables, views, and synonyms. In the center, there's a mapping diagram with components like Customer and MatchData. Two green callout boxes point to the left pane: one labeled "View/edit Properties" pointing to the Object Explorer, and another labeled "Preview Data" pointing to the Data Viewer pane at the bottom.

View/edit Properties

Preview Data

Properties Data Viewer

Configuration: (Default Settings) ... Run Show: (All Outputs) Choose...

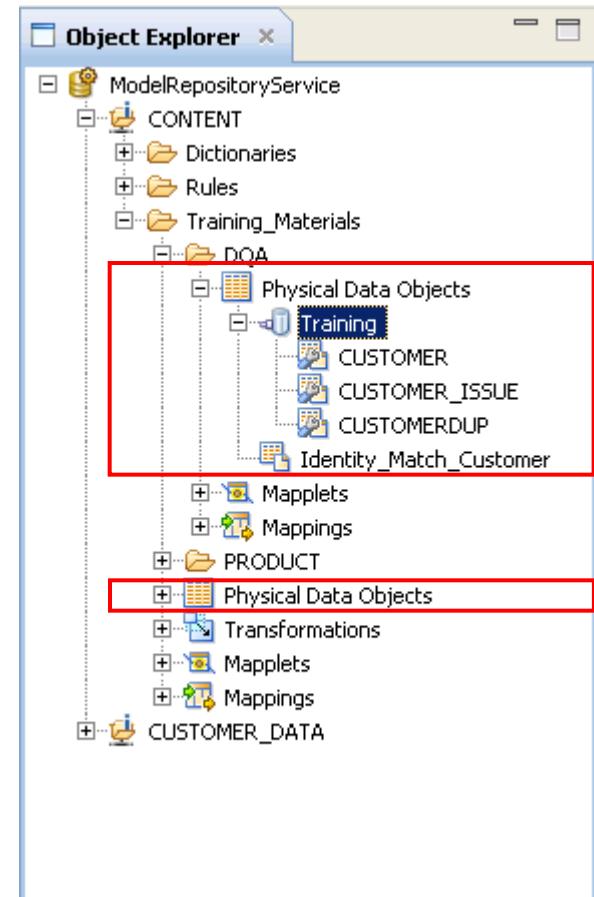
Output

Name: mt_Match_Data.Output

ID	COMPANY	ADDR1	ADDR2	ADDR3	ADDR4	COUNTRY	CONTACT	TITLE	PHONE
1	YEovale STOR...	1 MARGROVE T...	BARNSTAPLE	DEVON	EX32 7AD	GBR	MRS OPPORTU...	DIRECTOR OF IS	01271342156
2	YEovale STOR...	1 MARGROVE T...	BARNSTAPLE	DEVON	EX32 7AD	GBR	MRS OPPORTU...	PROGRAMMER	01271342156
3	WORLDSPAN	N8J1-1 300 GA...	ATLANTA	GA	30339	USA	MR CHIP KAYS	SENIOR ANALY...	7705637447
4	WORLDSPAN	N8J1-1 300 GA...	ATLANTA	GA	30339	USA	MR CHIP KAYS	SENIOR ANALY...	7705637448
5	WORLDSPAN	N8J1-1 300 GA...	ATLANTA	GA	30339	USA	MR CHIP KAYS	SENIOR ANALY...	7705637449
6	WORLDSPAN	N8J1-1 300 GA...	ATLANTA	GA	30339	USA	MR CHIP KAYS	SENIOR ANALY...	7705637450
7	WORLDSPAN	N8J1-1 300 GA...	ATLANTA	GA	30339	USA	MR CHIP KAYS	SENIOR ANALY...	7705637451
8	WORLDSPAN	N8J1-1 300 GA...	ATLANTA	GA	30339	USA	MR CHIP KAYS	SENIOR ANALY...	7705637447

Physical Data Objects

- Represents the native metadata in physical data sources and how it is accessed in the tool
- Physical data objects are used as sources, targets or lookups
- Relational tables are organized by connection names
- Connections are name-based reference



Relational Physical Data Objects

Relational PDO

The screenshot shows the 'INSURED' object in the Relational PDO tab. The 'General' section displays the name 'INSURED' and a description field. To the right, the 'INSURED' table structure is shown with columns: INS_INS_ID, INS_NAME, INS_ADDR_ID, INS_MAIN_PH_N..., INS_BUS_EMAIL_..., INS_CREATED_DT, and INS_MODIFIED_DT. Below this, the 'Columns' section lists the same columns with their native types (number, varchar2), precision, scale, and other metadata. At the bottom, tabs for Overview, Keys, Relationships, and Advanced are visible.

- **PDOs that represent just the native metadata**
- **Reuse the native metadata and customize read/write at the mapping level. For e.g., provide different filter, join conditions, etc**

Customized PDO

The screenshot shows the 'USADDRESS' object in the Customized PDO tab. The 'General' section displays the name 'USADDRESS' and a description field. To the right, the 'USADDRESS' table structure is shown with columns: ADDR_ID, ADDRESS1, ADDRESS2, CITY, REGION, and POSTAL_CD. Below this, the 'Columns' section lists the columns with their native types (number, varchar2), precision, scale, and visibility. A note indicates 'When column metadata changes:' followed by two options: 'Synchronize input and output' (selected) and 'Do not synchronize'. At the bottom, tabs for Overview, Read, Write, Parameters, and Advanced are visible.

- **PDOs that represent both the native metadata and the configuration rules for read/write**
- **Reuse customized PDO in mappings, cannot overwrite further at the mapping level**

Configuring Physical Data Objects - File

The screenshot shows the Informatica PowerCenter interface with the mapping 'US_Customer_Orders' open. On the left, the 'Source' tab displays the 'US_Customer_Orders' source object with columns: ID, ORDER_NO, ORDER_DATE, SHIP_DT, PO_NUM, and COMPANY. On the right, the 'Output' tab displays the target object with the same columns. Arrows connect the source columns to the target columns. Below the tabs, there are buttons for Overview, Read, Write, Parameters, and Advanced. The 'Write' tab is selected. In the bottom right corner of the window, the text 'server based' is visible.

Properties x Data Viewer

Overview Read Write Parameters Advanced

Name Value

Input type File

Source type Direct

File

Source file name US_custord.csv

Source file directory \\student01\\infa_shared\\SrcFiles

Command

Command

Truncate string null

Line sequential buffer length 8000

Input type

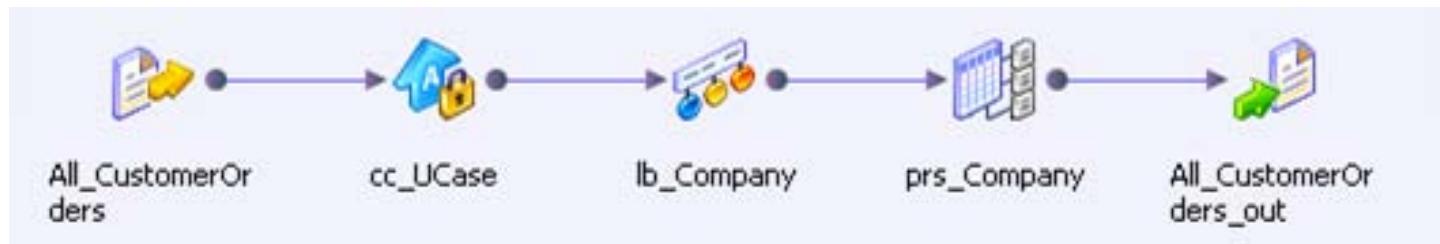
File or command

Configure the Read and Write tabs to indicate where the source file will be read from and written to (server based)

Configured in the Physical Data Objects, not at mapping level

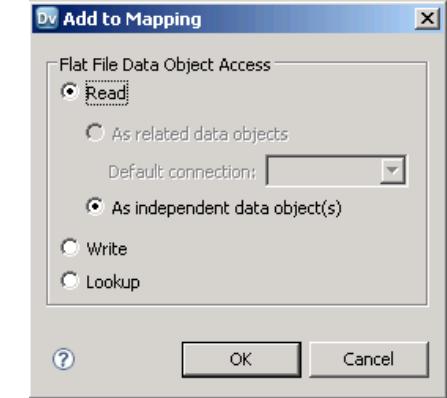
Mappings

- A Mapping reads data from sources, applies transformation logic to data and writes transformed data to targets.
- They can be used in IDQ to logically define the Data Quality/Integration Process.



Mapping elements

- **Physical Data Objects with Read access - Sources**
 - file-based
 - database
- **Operational transformations**
 - tools to cleanse, enhance and match the data
- **Physical Data Objects with Write access - Target**
 - file-based
 - database
- **Reference tables enable validation, parsing, enrichment and enhancement of data**

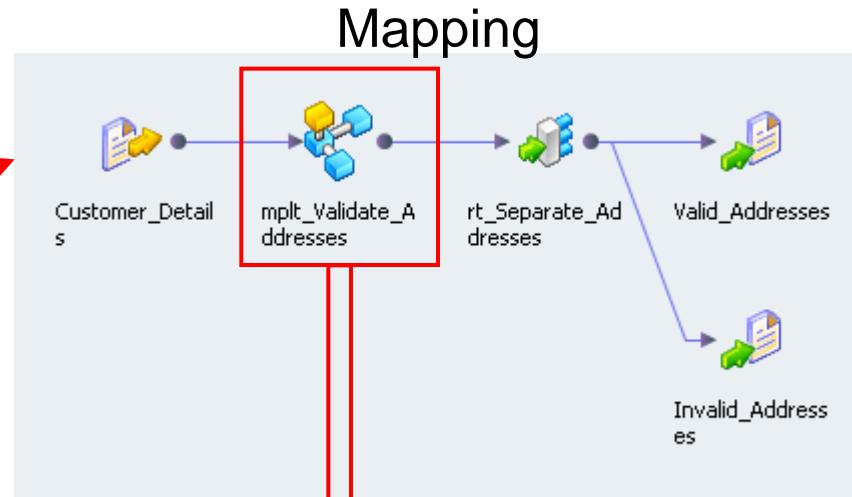


Mapplets and Rules

- A reusable object containing a set of transformations that you can use in multiple mappings.
- Use a mapplet in a mapping or, validate the mapplet as a rule and use in Informatica Analyst.
- When you use a mapplet in a mapping, you use an instance of the mapplet.
 - Changes made are inherited by all instances of the mapplet.

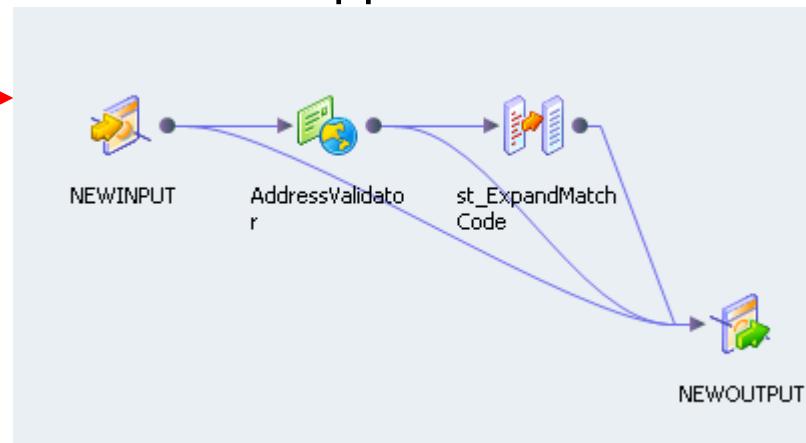
Maplet Example

**Source and target
data defined outside
the Maplet**



Maplet

**Maplet Input
transformation**
Passes data from a
mapping into a
maplet



**Maplet Output
transformation**
Passes data from a
maplet into a
mapping

Transformations

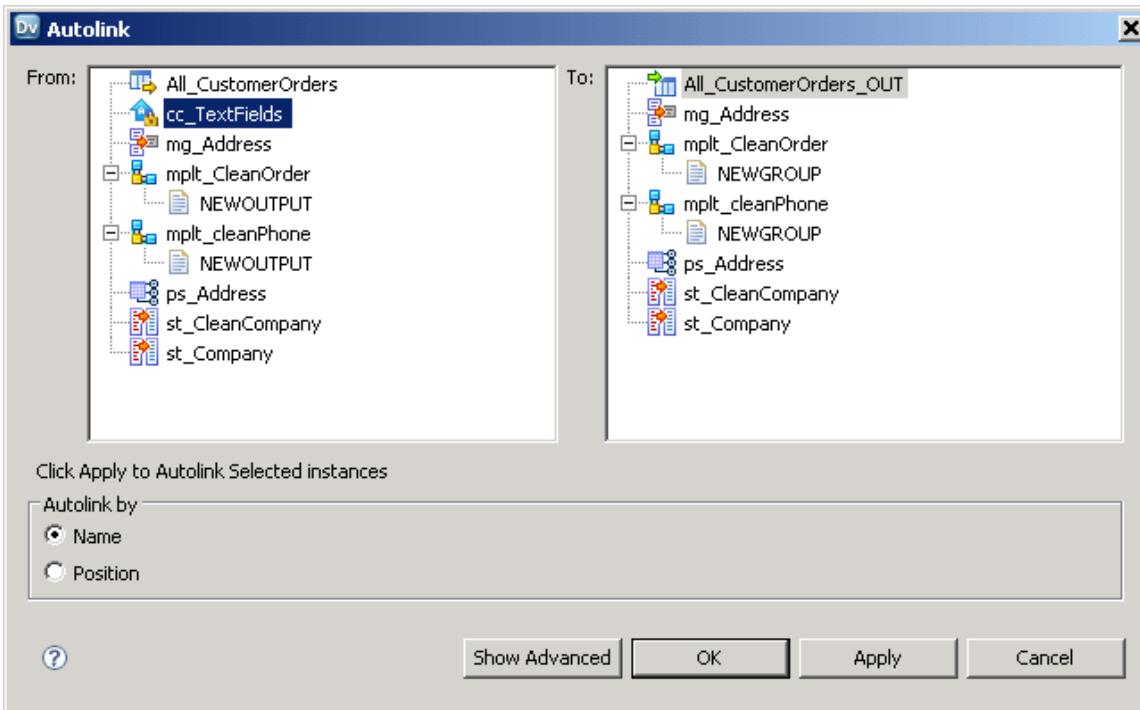
- Data passes through linked ports in a mapping / mapplet.
- An object that generates, modifies, or passes data.
- Reusable transformations:
 - Can be used in multiple mappings or mapplets.
 - All instances inherit changes.

Input Ports
→

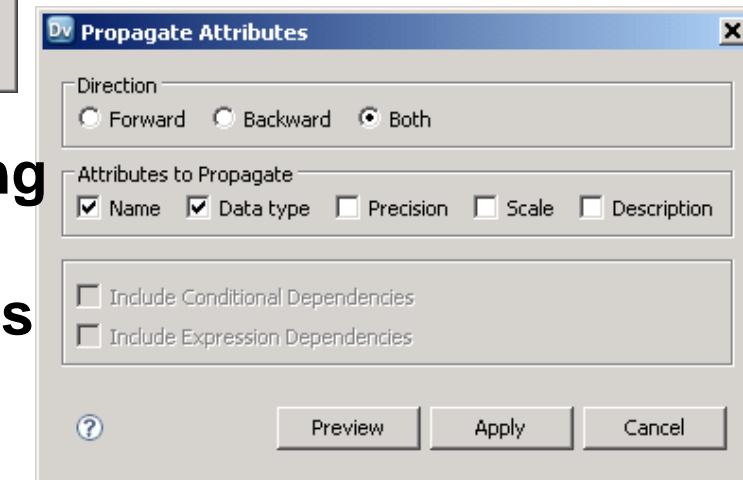
Name	Type	Pre
Input (13)		
COMPANY	string	25
ADDR1	string	25
ADDR2	string	25
ADDR3	string	25
ADDR4	string	25
COUNTRY_CODE	string	25
NAMEPREFIX	string	25
FIRSTNAME	string	25
LASTNAME	string	25
TITLE	string	25
EMAIL	string	25
ITEM_DSC	string	25
CURRENCY	string	25
Output (13)		
COMPANY	string	25
ADDR1	string	25
ADDR2	string	25
ADDR3	string	25
ADDR4	string	25
COUNTRY_CODE	string	25
NAMEPREFIX	string	25
FIRSTNAME	string	25
LASTNAME	string	25
TITLE	string	25
EMAIL	string	25
ITEM_DSC	string	25
CURRENCY	string	25

Output
Ports
←

Autalink & Propagate Attributes

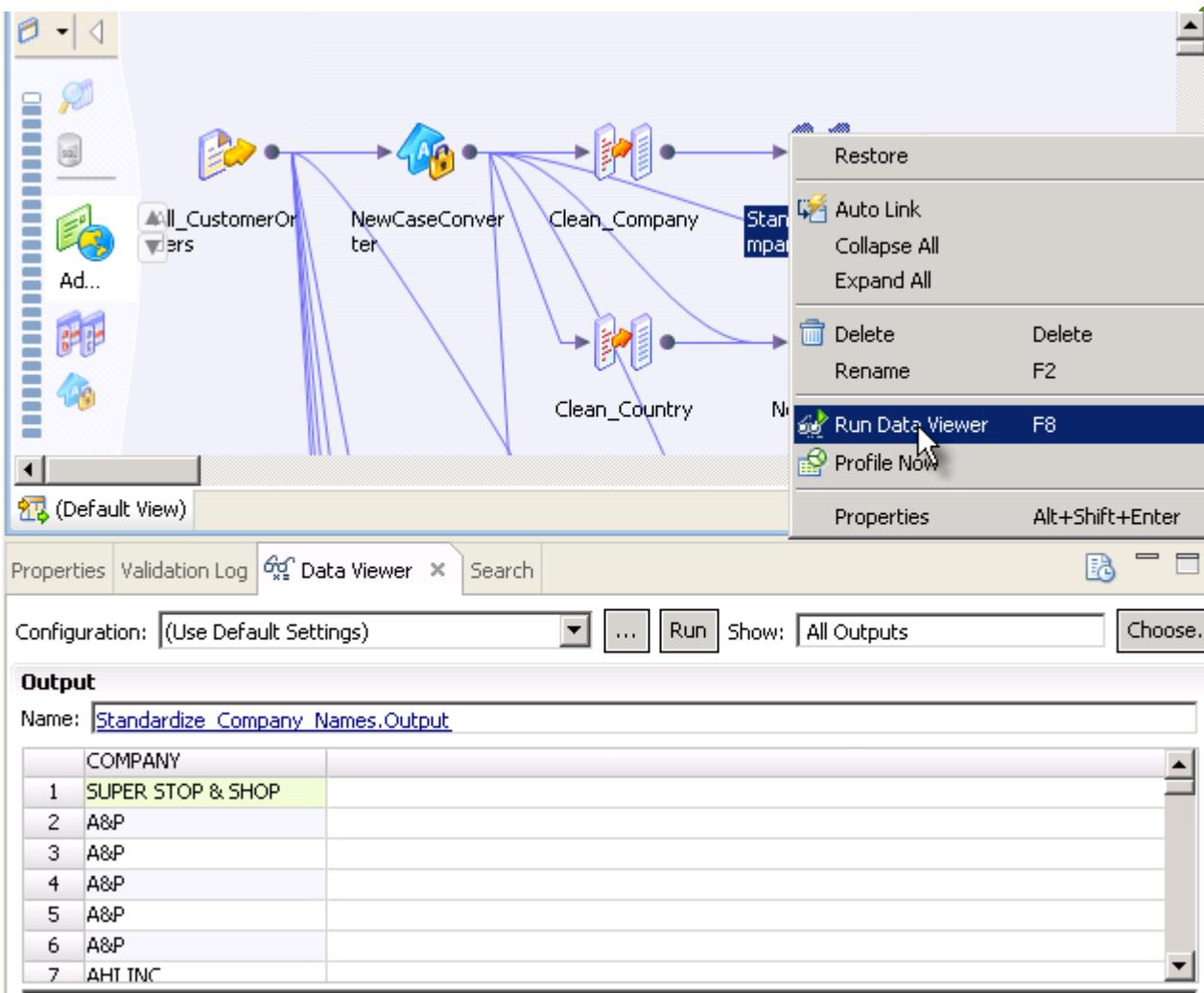


- Autalink ports from one transformation to another
- Autalink by using Prefix / Suffix



- Propagate attribute changes in mapping
- Doesn't affect reusable transformations

Data Preview

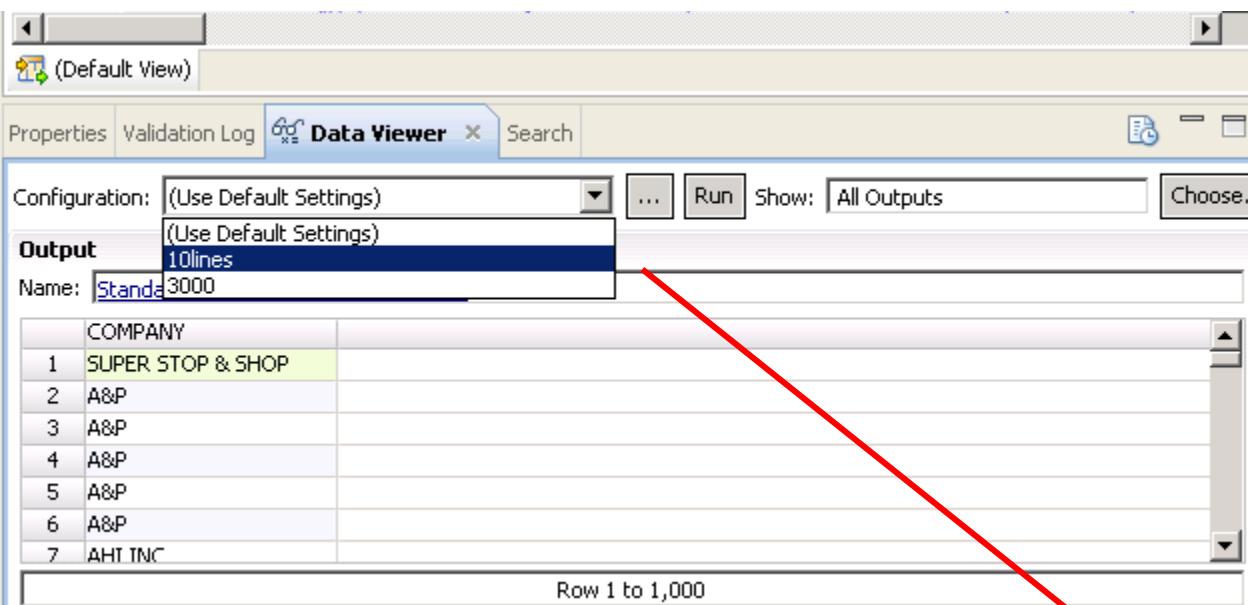


Data can be previewed – even in incomplete partially valid mappings

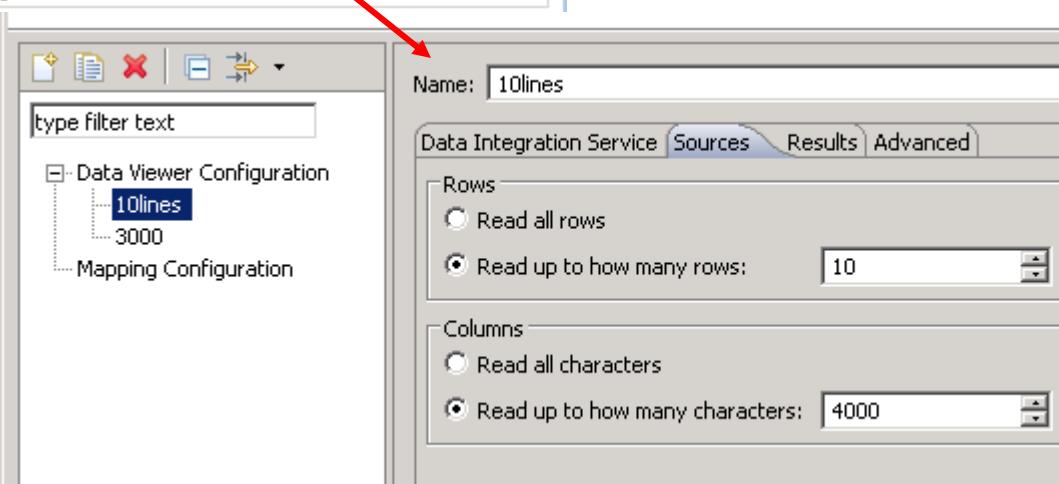
Immediate feedback as you develop, high productivity gains

Shows output ports only

Data Preview



- You can configure how many rows are read and displayed during the preview.



- You can also configure how many rows are processed when running/testing mappings.

Troubleshooting

The screenshot shows the Informatica PowerCenter interface with the 'Validation Log' tab selected. A log file titled 'm_Standardize_Customer_Orders1260480780812.txt' is open in the main pane. The log file contains the following entries:

```
2009-12-10 21:32:36 INFO: [MappingService] Preview job submitted with ID: iQSx7OXTEd6PTHi5
2009-12-10 21:32:41 FINER: APP_DTM-pool-2-thread-1, DTM_1000, Time taken to serialize me
2009-12-10 21:32:43 FINER: APP_9176, TM_6489, Time taken to de-serialize metadata: [
2009-12-10 21:32:44 INFO: APP_9176, VAR_27079, Warning: server variable :[$PMEExtProcDir],
2009-12-10 21:32:44 INFO: APP_9176, DISP_20305, The [Master] DTM with process id [3640] is

2009-12-10 21:32:44 INFO: APP_9176, TM_6964, Date format for the Session is [MM/DD/YYYY]
2009-12-10 21:32:44 INFO: APP_8484, TM_6721, Started [Partition Group Formation].
2009-12-10 21:32:44 INFO: APP_8484, TM_6722, Finished [Partition Group Formation]. It
2009-12-10 21:32:44 INFO: APP_8484, TM_6014, Initializing session [m_Standardize_Custom
2009-12-10 21:32:44 INFO: APP_8484, TM_6101, Mapping name: m_Standardize_Customer_Order
2009-12-10 21:32:44 INFO: APP_8484, TM_6708, Using configuration property [LkupRemoveOr
2009-12-10 21:32:44 INFO: APP_8484, TM_6708, Using configuration property [CreateDirect
2009-12-10 21:32:44 INFO: APP_8484, TM_6708, Using configuration property [lookupOverri
2009-12-10 21:32:44 INFO: APP_8484, TM_6703, Session [m_Standardize_Customer_Orders] is
2009-12-10 21:32:44 INFO: MANAGER, PETL_24058, Running Partition Group [1].
2009-12-10 21:32:44 INFO: MANAGER, PETL_24000, Parallel Pipeline Engine initializing.
2009-12-10 21:32:44 INFO: MANAGER, PETL_24001, Parallel Pipeline Engine running.
2009-12-10 21:32:44 INFO: MANAGER, PETL_24003, Initializing session run.
2009-12-10 21:32:44 INFO: MAPPING, CMN_1569, Server Mode: [UNICODE]
2009-12-10 21:32:44 INFO: MAPPING, CMN_1570, Server Code page: [MS Windows Latin 1 (ANS
2009-12-10 21:32:44 INFO: MAPPING, TM_6151, The session sort order is [Binary].
2009-12-10 21:32:44 INFO: MAPPING, TM_6185, Warning. Code page validation is disabled
```

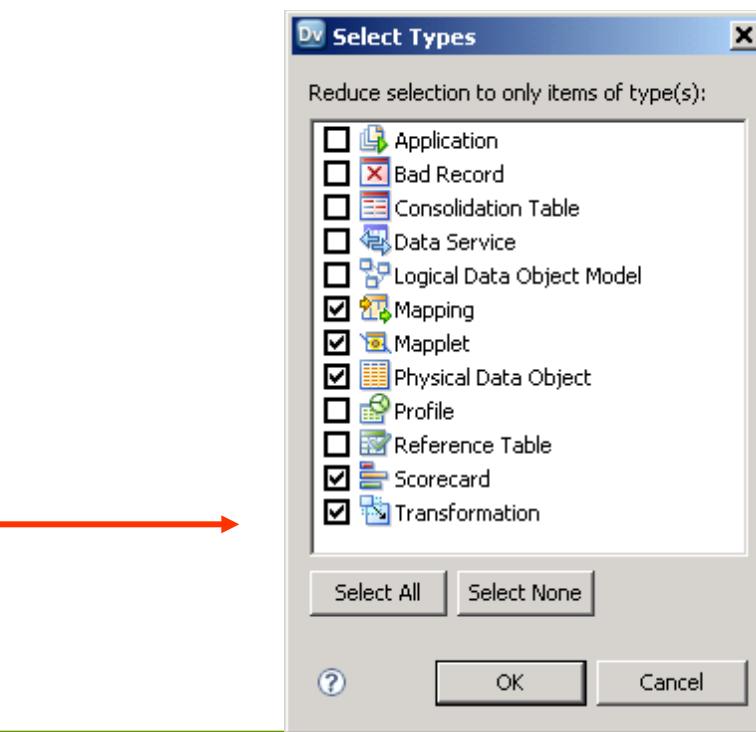
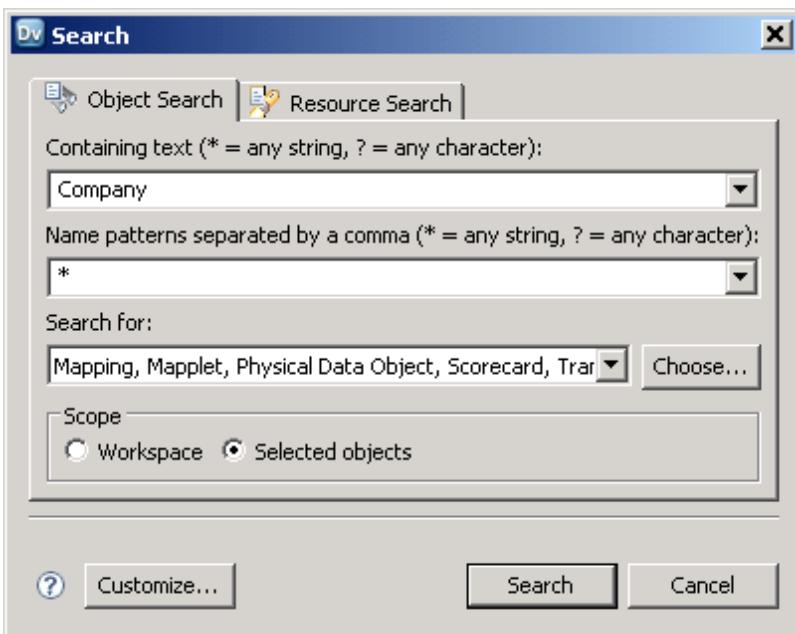
- First error is displayed in the Output view
- View log file to get more detailed information

Search



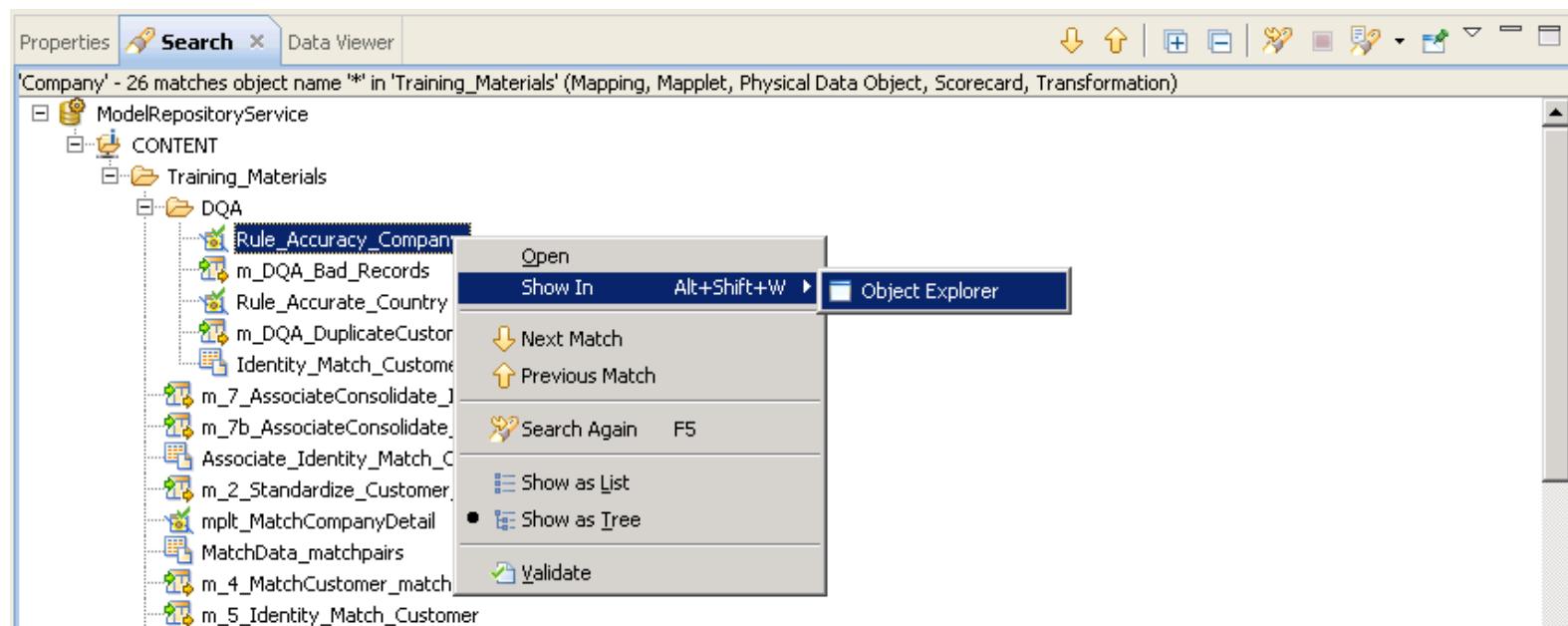
- **Search within a particular context**

- **Search within a particular folder**
- **Search within search**



Search Results

- Double-click or right-click on results to open directly
- Show In Object Explorer (Available elsewhere as well)



Developer Profiling

Column Profiling

Dv Informatica Developer

File Edit Navigate Search Run Window Help

Object Explorer

- ModelRepositoryService
 - CONTENT
 - Dictionaries
 - Rules
 - Training_Materials
 - DQA
 - PRODUCT
 - Bad Record Tables
 - Consolidation Tables
 - Physical Data Objects
 - Transformations
 - Mapplets
 - Mappings
 - Profiles
 - JoinAnalysis_US_EU_Customer
 - Profile_All_CustomerOrders
 - Profile_EU_Customer_Orders
 - Profile_US_Customer_Orders
 - Scorecards
 - CUSTOMER_DATA

Outline

- Overview
- Columns
- Profile Options
- Rules
- Results
- Comments

*Profile_All_CustomerOrders x

Results

Column Profiling

Column Profiling

Drilldown	Column	Unique Values	Unique %	Null Values	Null %	Datatype
	All_CustomerOrders					
<input checked="" type="checkbox"/>	ID	2882	100.00	-	-	- Integer(4) [100.00%]
<input checked="" type="checkbox"/>	ORDER_NO	1508	52.32	-	-	- Integer(5) [97.05%]
<input checked="" type="checkbox"/>	PO_NUM	1391	48.27	9	0.31	- Integer(6) [100.00%]
<input checked="" type="checkbox"/>	ORDER_DATE	855	29.67	51	1.77	String(10) [100.00%]
<input checked="" type="checkbox"/>	SHIP_DT	644	22.35	37	1.28	String(10) [100.00%]
<input checked="" type="checkbox"/>	COMPANY	420	14.57	-	-	- String(36) [100.00%]
<input checked="" type="checkbox"/>	ADDR1	1180	40.94	21	0.73	String(52) [100.00%]
<input checked="" type="checkbox"/>	ADDR2	709	24.60	75	2.6	String(25) [100.00%]
<input checked="" type="checkbox"/>	ADDR3	207	7.18	126	4.37	String(29) [100.00%]
<input checked="" type="checkbox"/>	ADDR4	868	30.12	52	1.8	String(10) [100.00%]
<input checked="" type="checkbox"/>	COUNTRY	11	0.38	-	-	- String(13) [100.00%]
<input checked="" type="checkbox"/>	CONTACT	1320	45.80	-	-	- String(36) [100.00%]
<input checked="" type="checkbox"/>	TITLE	133	4.61	122	4.23	String(59) [100.00%]
<input checked="" type="checkbox"/>	PHONE	1043	36.19	61	2.12	String(26) [100.00%]
<input checked="" type="checkbox"/>	EMAIL	828	28.73	95	3.3	String(50) [100.00%]
<input checked="" type="checkbox"/>	SP_NO	42	1.46	5	0.17	Integer(3) [100.00%]
<input checked="" type="checkbox"/>	ITEM_NO	221	7.67	189	6.56	String(7) [100.00%]
<input checked="" type="checkbox"/>	QUANTITY	565	19.60	9	0.31	Integer(4) [100.00%]
<input checked="" type="checkbox"/>	ITEM_DSC	458	15.89	-	-	- String(46) [100.00%]
<input checked="" type="checkbox"/>	SUPID	18	0.62	2	0.07	Integer(3) [100.00%]
<input checked="" type="checkbox"/>	UNIT_COST	117	4.06	-	-	- String(9) [100.00%]

Details

Show: Values

Value	Frequ...	Percent	Chart
EQUATE INC	70	2.43%	
DISTRICT OF COLUM...	64	2.22%	
Boots	64	2.22%	
A&P	51	1.77%	
SPAR	50	1.73%	
SCOTMID	50	1.73%	
FAA	49	1.70%	
US NAVY	47	1.63%	
DEPARTMENT OF ED...	43	1.49%	
AHI INC	40	1.39%	
PENDLETON GROUP	39	1.35%	
One Stop	30	1.05%	
TECHNAUT	Send to	28%	
EDS	Show Matching Rows	28%	
NATIONAL SECURITY...	36	1.25%	
FAIRFAX COUNTY PL...	35	1.21%	

Value & Pattern Frequencies

GEORGIA TECH RESE...

Properties Data Viewer Validation Log Progress

Show: (All Outputs)

Output

Name: CONTENT\Training_Materials\Profile_All_CustomerOrders

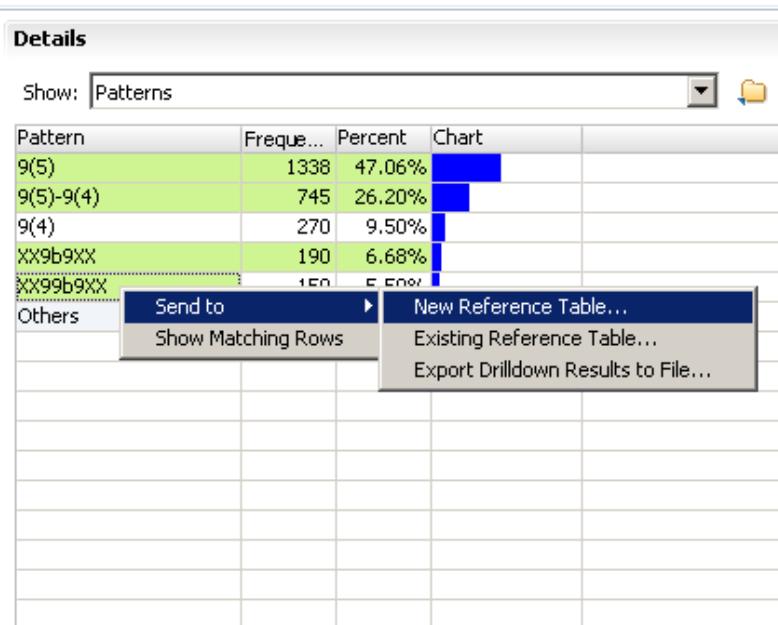
ID	ORDER_NO	PO_NUM	ORDER_DATE	SHIP_DT	COMPANY	ADDR1	ADDR2	ADDR3	ADDR4
1 5090	1983	5338	02/01/2009	06/24/2009	One Stop	96/98 Whitegat...	Blackpool		FY19BZ
2 5091	23594	176	02/01/2009	06/24/2009	One Stop	96/98 Whitegat...	Blackpool		FY19BZ
3 5163	1986	833	07/06/2008	06/05/2009	One Stop	1 3 Grosvenor...	Gloucester		GL1 1SZ
4 5188	1994	1258	09/27/2009	06/20/2009	One Stop	8 Belmont Road	Hereford		HR2 7JE

Drill Down Results

Row 1 to 39

Value Frequencies

- Create or update reference tables using frequency values output from profiling



The screenshot shows the 'New Reference Table' dialog and a preview window. The dialog title is 'New Reference Table' and it contains a 'Reference Table Columns' section. A column named 'ADDR4' is defined with a 'string' data type, precision of 10, and scale of N/A. Below the table are two checkboxes: 'Include a column for row-level descriptions (Note: this will be used as descriptive metadata only)' and 'Include frequency'. At the bottom are 'OK', 'Cancel', and 'Help' buttons. A preview window titled 'ADDR4' shows four rows of data: '99999', '99999-9999', 'XX9 9XX', and 'XX99 9XX'.

Exporting Profiling Results

- 200 value frequencies are displayed. To see more, export to a CSV output
- Drill Down results can also be exported for review

The screenshot shows the Informatica Developer interface with the following details:

- Object Explorer:** Shows a tree structure with nodes like ModelRepositoryService, CONTENT, Dictionaries, Rules, Training_Materials, DQA, PRODUCT, Bad Record Tables, Consolidation Tables, Physical Data Objects, Transformations, Mapplets, Mappings, Profiles, Scorecards, and CUSTOMER_DATA.
- *Profile_All_CustomerOrders Results:** A tab showing "Column Profiling".
 - Drilldown:** A table with columns: Column, Unique Values, Unique %, Null Values, Null %, and Datatype. It lists various columns from the All_CustomerOrders table with their respective statistics.
 - Details:** A table showing value frequencies. The "Show" dropdown is set to "Values". The table includes columns: Value, Freq., Percent, and Chrt. Red boxes highlight the "Chrt" column header and the "Values" dropdown.
- Data Viewer:** A table showing sample data for the All_CustomerOrders table. The columns are ID, ORDER_NO, PO_NUM, ORDER_DATE, SHIP_DT, COMPANY, ADDR1, ADDR2, ADDR3, ADDR4, and COUN. Red boxes highlight the "Run" button and the "Output" section.

Export Value Frequencies

Export Drill down Results

Join Analysis Profiling

Results

Join Profile

Last run on: Jan 25, 2010 12:41:08 PM

Left Table	Right Table	Left Only Rows	Right Only Rows
empinfo:EMPID	empinfo2:EMPID	220	203

Join Condition

Venn Diagram with join results

Details

empinfo
250 rows, 220 (88%) orphan, 0 (0%) null, 30 (12%) right rows

empinfo2
235 rows, 203 (86%) orphan, 0 (0%) null, 32 (14%) left rows

Join
30 (12%) left rows, 32 (14%) right rows

Overview Columns Join Analysis Results Comments

Properties Data Viewer

Show: All Outputs Choose... Run

Output

Name: CUSTOMER_DATA\JoinAnalysis_empinfo

CITY	DATE_JOINED	DEPTID	DEPTNAME	EMPID	FIRSTNAME	GENDER	LASTNAME
1 Newark	6/21/1997	200	MIS	216	Airaj		Masood
2 Minneapolis	12-Jul-92	200	MIS	134	Allison	F	Swenson
3 Minneapolis	5/30/1996	200	MIS	164	Allison	F	Park
4 Houston	1/13/1997	100	Sales	255	Amy	F	Barbles
5 Fairfax	2/26/1992	200	MIS	273	Anna	1	Karneh
6 Bellevue	3/29/1995	200	MIS	163	Anne	F	Brastow
7 New York	10/8/1990	200	MIS	120	Audrey	F	Simpson
8 Washington	4/25/1991	300	Administration	206	Denise	1	Brinker
9 Seattle	6/24/1993	200	MIS	123	Heather	F	Santry

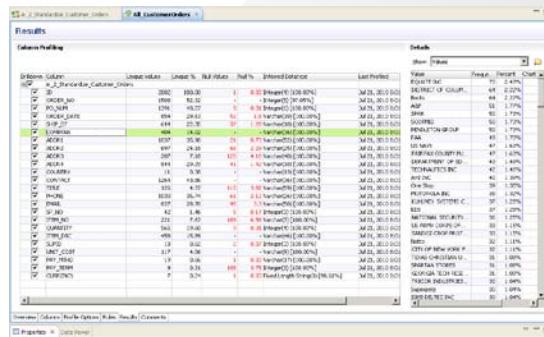
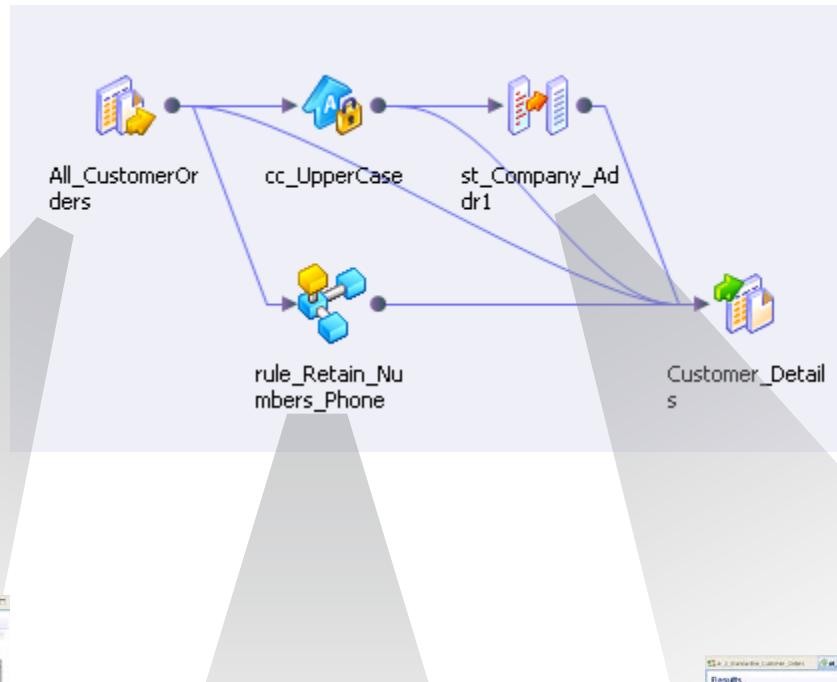
Use Join Analysis to evaluate the degree of overlap between two columns

Click on the Join Condition to view the Venn Diagram

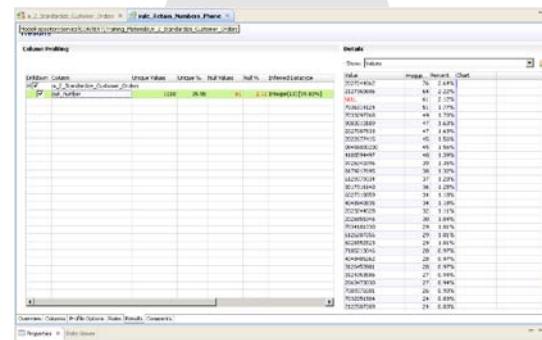
Double click on the area in the Venn Diagram to view the join/orphan records

Mid-Stream Profiling – Profile at any point within a Mapping

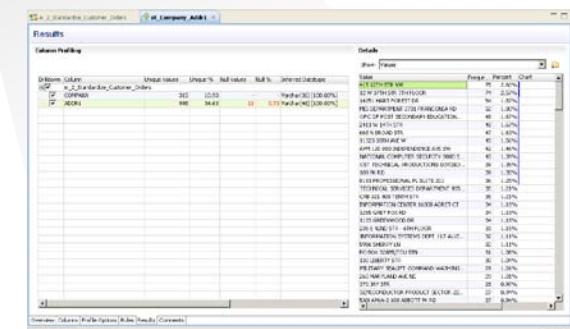
***Targets can not be profiled**



Profile Source



Profile Maplet/Rule



Profile any Transformation

Data Standardization

What is Standardization?

- **Standardization addresses the data quality issues identified through data profiling**
- **The key objectives in data standardization are:**
 - to transform and parse data from single multi-token fields to multiple fields
 - to correct completeness, conformity, and consistency problems
 - to standardize field formats and extract important data from free text fields

Pattern Based Parsing

Results

Column Profiling

Drilldown	Column	Unique Values	Unique %	Null Values	Null %	Datatype
<input checked="" type="checkbox"/>	m_2_Standardize_Customer_Orders					- String(30) [100.00%]
<input checked="" type="checkbox"/>	Labeled_Output	15	0.52	-	-	- String(30) [100.00%]
<input checked="" type="checkbox"/>	Tokenized_Data	1314	45.59	-	-	- String(35) [100.00%]

New Reference Table

Reference Table

Reference Table Columns

Name	Data Type	Precision	Scale	Description
ict_Pattern_Output	string	52	N/A	

Show: Values

Value	Freque...	Percent	Chart
Firstname Surname	1509	52.36%	Bar
Firstname Firstname	582	20.19%	Bar
Nameprefix Firstname Surname	564	19.57%	Bar
Nameprefix Firstname Firstname	123	4.27%	Bar
	47	1.63%	Bar
	19	0.66%	Bar

Send to > Show Matching Rows

Value	Percent
6	0.21%
5	0.17%
3	0.10%
2	0.07%
2	0.07%
1	0.03%
1	0.03%

Overview | Columns | Profile Options | Rules | Result

Properties Data Viewer

Configuration: (Default Settings)

Include a column for row-level descriptions (Note:this will be used as descriptive metadata only)

Include frequency

Output

Name: pp_Parse_Contact.Output

nameprefix	firstname	parsed_by
1	ADAM	Parsed by Pattern - [3]
2	ADAMSEN	Parsed by Pattern - [1]
3	AIDAN	Parsed by Pattern - [1]
4	ALAYNE	Parsed by Pattern - [1]
5	ALCOT	Parsed by Pattern - [1]
6	ALCOT	Parsed by Pattern - [1]
7	ALEXANDER	Parsed by Pattern - [1]
8	ALEXANDRE	Parsed by Pattern - [3]
9	ALEXANDRE	Parsed by Pattern - [1]
10	ALFREDIA	Parsed by Pattern - [1]
11	ALIA	Parsed by Pattern - [1]
12	ALICE	Parsed by Pattern - [1]
13	ALISHA	Parsed by Pattern - [1]

INFORMATICA

Create a reference table using output from the labeler

Add a Pattern Parser and apply the new reference table

- Parse the patterns

Output fields:
Parsed Data
Parse Status
Overflow

Standardization Transformations



The Case Converter transformation creates data uniformity by standardizing the case of strings in input data.



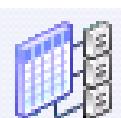
The Merge transformation reads the data values from multiple input fields to create a single output field.



The Standardizer transformation standardizes characters and strings in data. It also can be used to remove noise from a field.



The Decision transformation can be used to build rules



The Parser transformation can parse input data using the following methods:

- Token set.
- Regular expression.
- Reference table.

Address Validation

Address Validation

- **240+ Countries Supported by a single vendor – AddressDoctor**
- **Ability to parse addresses**
- **New input strategies to support different customer storage formats**
- **Additional output transformations to support varying international addressing formats**
- **CASS and SERP reports**
- **Standardized address match codes across all countries**
- **Significant improvements in performance with the ability to multistream**
- **Improved data download processes via Address Doctor fulfilment processes**
- **Single License Key Location**

Output Groups

- **Predefined output groups:**
 - **Geo Coding**
 - Latitude
 - Longitude
 - **Country**
 - Country name
 - ISO country code
 - **Status Info**
 - Information on the quality of each input address
 - **Formatted Address Line**
 - Formats addresses for mailing
 - **Residue**
 - Unrecognized elements in the input address

Address Validation Configuration



Address Validator

The screenshot shows the 'Outputs' section of the configuration interface. It lists several output ports: Postcode1 (string, 15), Province1 (string, 50), CountryName (string, 50), Organization... (string, 50), BuildingNum... (string, 20), and StreetComple... (string, 50). The 'Outputs' section is expanded, and the 'Address Elements' group is also visible.

- Define default/force country
- Define Casing
- Define Mode
- Define input template
- Add input ports
 - Select ports from one input group only

The screenshot shows the 'Properties' tab with the 'Advanced' section selected. The table lists various configuration options:

Name	Value
Tracing Level	Normal
Options	
Matching Alternatives	All
Matching Scope	All
Optimization Level	Standard
Input Format Type	All
Output Format Type	All
Input Format With Country	<input type="checkbox"/>
Output Format With Country	<input type="checkbox"/>
Country Type	EN
Country of Origin	
Preferred Script	Database
Preferred Language	Database
Execution Instances	1

- Add output ports
 - Add ports from multiple output groups

Configure advanced settings *performance improvements X5+

Address Validation level: A+

- **A+: Street or Building coverage for more than 98% of the country. The following countries are available:**

Argentina	Finland	Isle of Man	Paraguay	United Kingdom
Australia	France	Italy	Poland	United States
Austria	French Guiana	Japan	Portugal	Virgin Islands, U.S.
Belgium	Germany	Jersey	Puerto Rico	
Bermuda	Greenland	Liechtenstein	Réunion	
Brazil	Guadeloupe	Luxembourg	Russian Federation	
Canada	Guam	Martinique	Saint Barthélemy	
Cyprus	Guernsey	Mayotte	Singapore	
Denmark	Holy See	Mexico	Slovenia	
Dominican Republic	Hong Kong	Netherlands	Sweden	
Estonia	Indonesia	New Zealand	Switzerland	

Address Validation Level: A

- A: Street, Building or Block coverage for major parts of the country. The following countries are available:

American Samoa
Armenia
Bangladesh
Belarus
Brunei Darussalam
Chile
China
Czech Republic
Greece
Guatemala

Haiti
Hungary
Iceland
India
Korea, Republic of
Latvia
Lithuania
Malaysia
Maldives
Malta
Marshall Islands

Micronesia, Federated States of
Moldova
Monaco
Morocco
Northern Mariana Islands
Norway
Palau
Philippines

Romania
Slovakia
Spain
Turkey
Ukraine
Uruguay
Venezuela

Address Validation Level: B

- B – Locality and Postal Code. Countries include:

Afghanistan
Åland Islands
Albania
Algeria
Andorra
Angola
Anguilla
Antarctica
Antigua and Barbuda
Aruba
Azerbaijan
Bahamas
Bahrain
Barbados
Belize
Benin

Bhutan
Bolivia
Bosnia and Herzegovina
Botswana
British Indian Ocean Territory
Bulgaria
Burkina Faso
Burundi
Cambodia
Cameroon
Cape Verde
Cayman Islands
Central African Republic
Chad
Christmas Island
Cocos Islands

Colombia
Comoros
Congo
Congo, The Democratic Republic of the
Cook Islands
Costa Rica
Côte d'Ivoire
Croatia
Cuba
Djibouti
Dominica
Ecuador
Egypt
El Salvador
Equatorial Guinea
Eritrea

Ethiopia
Falkland Islands
Faroe Islands
Fiji
French Polynesia
Gabon
Gambia
Georgia
Ghana
Gibraltar
Grenada
Guinea
Guinea-Bissau

Guyana
Honduras
Iran, Islamic Republic of
Iraq
Israel

Address Validation Level: B

Jamaica	Mongolia	Pitcairn	South Georgia and the South Sandwich Islands	United Arab Emirates
Jordan	Montenegro	Qatar	Sri Lanka	Uzbekistan
Kazakhstan	Montserrat	Rwanda	Sudan	Vanuatu
Kenya	Mozambique	Saint Helena	Suriname	Viet Nam
Kiribati	Myanmar	Saint Kitts and Nevis	Svalbard and Jan Mayen	Virgin Islands, British
Korea, Democratic People's Republic of	Namibia	Saint Lucia	Swaziland	Wallis and Futuna
Kuwait	Nauru	Saint Pierre and Miquelon	Syrian Arab Republic	Western Sahara
Kyrgyzstan	Nepal	Saint Vincent and the Grenadines	Tajikistan	Yemen
Lao, People's Democratic Republic	Netherlands Antilles	Samoa	Tanzania, United Republic of	Zambia
Lebanon	New Caledonia	San Marino	Thailand	
Lesotho	Nicaragua	Sao Tome and Principe	Togo	
Liberia	Niger	Saudi Arabia	Tokelau	
Libyan Arab Jamahiriya	Nigeria	Senegal	Tonga	
Macedonia, The Former Yugoslav Republic of	Niue	Serbia	Trinidad and Tobago	
Madagascar	Norfolk Island	Seychelles	Tunisia	
Malawi	Oman	Sierra Leone	Turkmenistan	
Mali	Pakistan	Solomon Islands	Turks and Caicos Islands	
Mauritania	Panama	Somalia	Tuvalu	
Mauritius	Papua New Guinea	South Africa	Uganda	
	Peru			

GeoCoding

- **GeoCoding is available for the following countries**
 - Andorra
 - Australia
 - Austria
 - Belgium
 - Canada
 - Croatia
 - Czech Republic
 - Denmark
 - Estonia
 - Finland
 - France
 - Germany
 - Gibraltar
 - Greece
 - Hungary
 - Italy
 - Latvia
 - Liechtenstein
 - Luxembourg
 - Mexico
 - Monaco
 - Netherlands
 - Norway
 - Poland
 - Portugal
 - San Marino
 - Singapore
 - Slovakia
 - Slovenia
 - Spain
 - Sweden
 - Switzerland
 - United Kingdom
 - United States

Address Validation Parameters

INFORMATICA® Administrator

Administrator Log Out | Manage ▾ Help ▾

Domain Logs Monitoring Reports Security

DataIntegrationService : Available

Actions

Domain Navigator Actions ▾

- Domain node01
- ModelRepositoryService
- DataIntegrationService**
- AnalystService
- Education

DataIntegrationService Properties Processes Applications Permissions

Node	Node Status	Process Configuration	Process State
node01	Available	Enabled	Running

Service Process Properties

⋮ Data Integration Service Security Options Edit

HTTP Port 8095

⋮ Logging Options Edit

Log Directory ./dsLogs

⋮ Execution Options Edit

Temporary Directories ./dsTemp

Maximum Execution Pool Size 10

⋮ Address Validation Properties Edit

License Key 1234567890091132132

Reference Data Location C:\Informatica\9.0.1\services\DQContent\INFA_Content\av\default

Full Pre-Load Countries ALL

Partial Pre-Load Countries

No Pre-Load Countries

Full Pre-Load Geocoding Countries ALL

Partial Pre-Load Geocoding Countries

No Pre-Load Geocoding Countries

Memory Usage 1024

Maximum Address Object Count 15

Maximum Thread Count 5

Cache size LARGE

⋮ Advanced Properties Edit

Address Validation Properties

Property	Value
License Key	1234567890091132132
Reference Data Location	C:\Informatica\9.0.1\services\DQContent\INFA_Content\av\default
Full Pre-Load Countries	ALL
Partial Pre-Load Countries	
No Pre-Load Countries	
Full Pre-Load Geocoding Countries	ALL
Partial Pre-Load Geocoding Countries	
No Pre-Load Geocoding Countries	
Memory Usage	1024
Maximum Address Object Count	15
Maximum Thread Count	5
Cache size	LARGE

Define the License key in Informatica Administrator (separate license for Geocoding)

Define the location of the reference data

License expires (not data except CASS data)

Grouping and Matching

Matching Theory

- Consider the following records. How many duplicates are there?
- There are 2 records that could be considered matches. How did you work that out?
- There are 3 logical phases in the matching process:
 - Pair Generation
 - Scoring (matching)
 - Processing

Name	Address
George W Bush	Texas
William J Clinton	New York
Hilary Rodham Clinton	New York
Nancy Pelosi	San Francisco
George H W Bush	Texas

I. Matching Theory - Pair Generation

- In this example, each record in the dataset will be compared with all others. This gives a total of 10 pairs.

Name1	Address1	Name2	Address2
George W Bush	Texas	William J Clinton	New York
George W Bush	Texas	Hilary Rodham Clinton	New York
George W Bush	Texas	Nancy Pelosi	San Francisco
George W Bush	Texas	George H W Bush	Texas
William J Clinton	New York	Hilary Rodham Clinton	New York
William J Clinton	New York	Nancy Pelosi	San Francisco
William J Clinton	New York	George H W Bush	Texas
Hilary Rodham Clinton	New York	Nancy Pelosi	San Francisco
Hilary Rodham Clinton	New York	George H W Bush	Texas
Nancy Pelosi	San Francisco	George H W Bush	Texas

II. Matching Theory - Scoring

- The next phase assigns a score (1 indicates they are identical) to each pair, which indicates how similar they are.

Name1	Address1	Name2	Address2	Score
George W Bush	Texas	William J Clinton	New York	0
George W Bush	Texas	Hilary Rodham Clinton	New York	0
George W Bush	Texas	Nancy Pelosi	San Francisco	0
George W Bush	Texas	George H W Bush	Texas	0.9
William J Clinton	New York	Hilary Rodham Clinton	New York	0.6
William J Clinton	New York	Nancy Pelosi	San Francisco	0
William J Clinton	New York	George H W Bush	Texas	0
Hilary Rodham Clinton	New York	Nancy Pelosi	San Francisco	0
Hilary Rodham Clinton	New York	George H W Bush	Texas	0
Nancy Pelosi	San Francisco	George H W Bush	Texas	0

III. Matching Theory - Processing

- The same number of rows that were originally received are output with an identifier added to each row. Rows that are similar will have the same identifier or ClusterID.
- To determine if two rows are related, we specify a threshold value. Pairs with a score equal to or above the threshold are deemed to match.
- Our threshold is 0.8. Only one pair meets the threshold.

Name	Address	ClusterID
George W Bush	Texas	1
William J Clinton	New York	2
Hilary Rodham Clinton	New York	3
Nancy Pelosi	San Francisco	4
George H W Bush	Texas	1

Transformations

- **Matching Transformations:**



Key Generator – used to group the data



Match - used to match the data

- **Typically the following will be used in Matching Mapplets:**



Comparison



Weighted Average

Grouping

- The number of pairs that a dataset with N records will generate is given by the formula: $\frac{(n^2 - n)}{2}$
 - 5 records will create 10 pairs
 - 50 records will create 1225 pairs
 - 500 will create 124,750
 - 5,000 records will generate nearly 12.5 million pairs.
- We need to consider ways to reduce the number of pairs created, and so reduce the impact on performance.
 - To do this, we should only generate pairs for records that are likely to match only comparing records that share one (or more) particular characteristics.

1. Grouping

- We do this by nominating a Group Key. All records that have the same Group Key are compared against each other.
 - If we nominate Address as the Group Key, we only get two pairs created.

Name	Address	Name	Address
George W Bush	Texas	George H W Bush	Texas
William J Clinton	New York	Hilary Rodham Clinton	New York

- If a data set of 5,000 records is grouped so there are 10 groups of 500 records, it will generate 1.2 million pairs instead of 12 million.

IDQ Grouping and matching

- In matching, the records within each group are compared against each other.
- Matching is not performed across groups, therefore be sure to group on a complete and accurate field.

Configuration: (Use Default Settings)

Output

Name: [kg_Generate_Keys.Output](#)

SequenceId	GroupKey	ID	COMPANY	ADDR1	ADDR2	ADDR3	ADDR4
1	Y1	5336	YEOVALE STOR...	1 MARGROVE T...	BARNSTAPLE	DEVON	EX32
2	Y1	5337	YEOVALE STOR...	1 MARGROVE T...	BARNSTAPLE	DEVON	EX32
3	W6	301	WORLDSPAN	N8J1-1 300 GA...	ATLANTA	GA	3033
4	W6	439	WORLDSPAN	N8J1-1 300 GA...	ATLANTA	GA	3033
5	W6	752	WORLDSPAN	N8J1-1 300 GA...	ATLANTA	GA	3033
6	W6	753	WORLDSPAN	N8J1-1 300 GA...	ATLANTA	GA	3033
7	W6	1103	WORLDSPAN	N8J1-1 300 GA...	ATLANTA	GA	3033
8	W6	1320	WORLDSPAN	N8J1-1 300 GA...	ATLANTA	GA	3033
9	W6	1691	WORLDSPAN	N8J1-1 300 GA...	ATLANTA	GA	3033
10	W6	1755	WORLDSPAN	N8J1-1 300 GA...	ATLANTA	GA	3033

• Group 1

• Group 2

Key Generator Transformation

- The Key Generator transformation has three purposes:
 - Assign a unique identifier to each record in a dataset if one does not exist.
 - Apply an operation to a field so that it is more suitable for grouping
 - Sort the outgoing data so that rows with the same group key value are contiguous.
 - Only required for classic matching

Key creation strategy

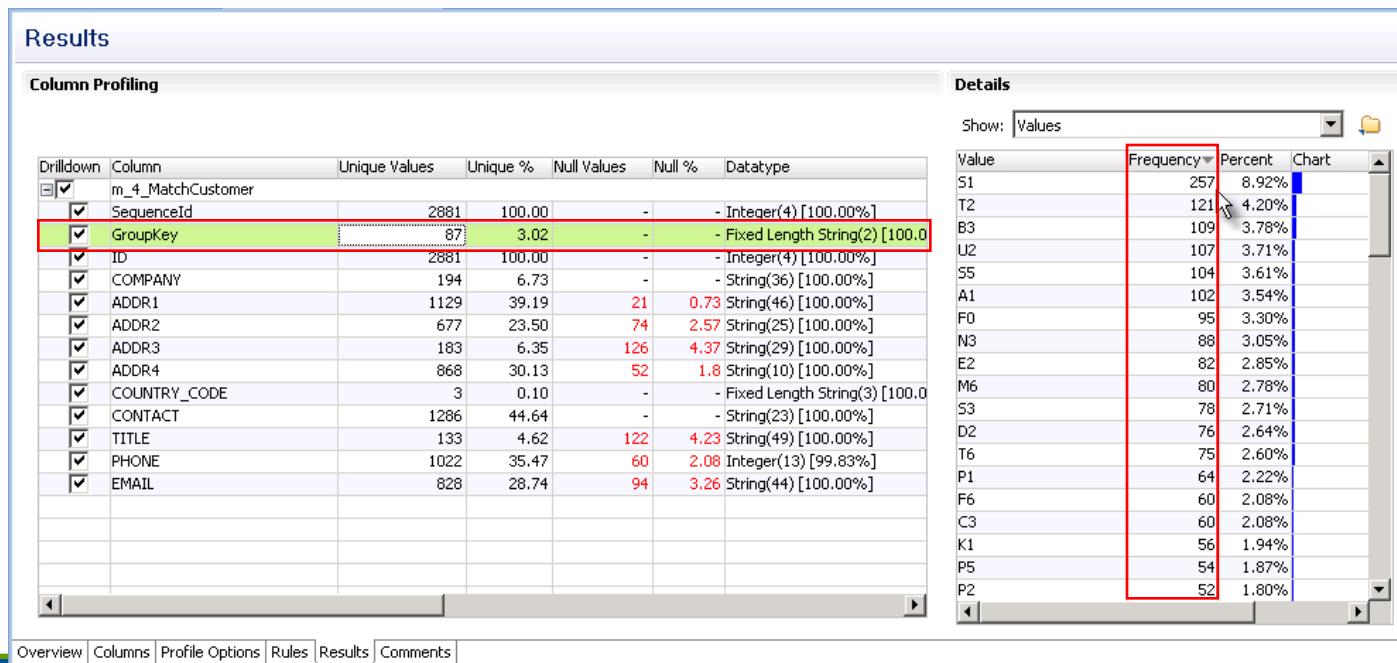
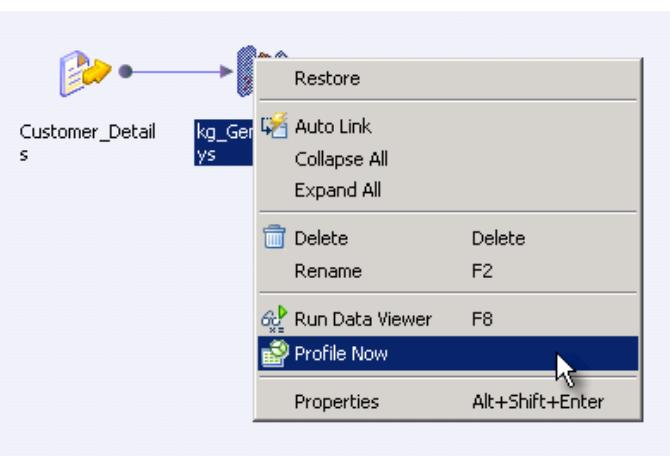
- **String**
 - Builds a group key using the first or last number of characters
- **NYSIIS**
 - The NYSIIS transformation converts a word into its phonetic equivalent.
- **Soundex**
 - The Soundex generates an alphanumeric code that represents the characters at the start of a string. It creates a code based on how the word sounds and takes variations of spelling into account.

Mid-Stream Profiling for Group Analysis

Profile Key Generator Transformation

- **Review:**

- **Number of records per group**
- **NULL keys**
- **Single record groups**



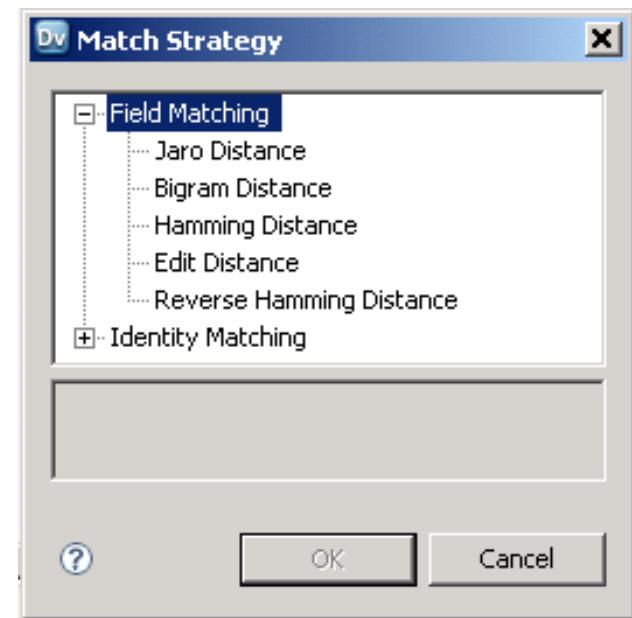
2. Data Matching

- Matching will identify related or duplicate records within a dataset or across two datasets.
- Matching scores records between 0 and 1 on the strength of the match between them, with a score of 1 indicating a perfect match between records.
- Informatica 9 provides a wide range of matching capabilities for each data type.
- Users have the flexibility to decide which algorithms they would like to use as well as configuring null rules, weightings and thresholds.

Matching

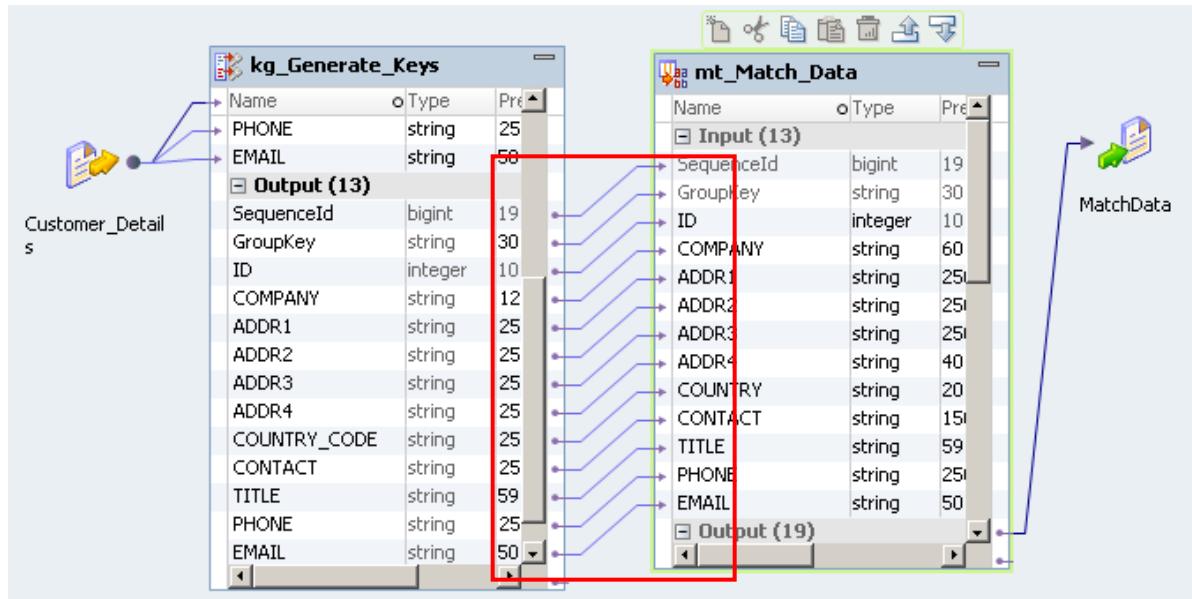
The Match transformation reads values in selected input columns and calculates match scores representing the degrees of similarity between the pairs of values.

- Match Type (Pair Generation)
 - Strategies(Scoring)
 - Match Output (Processing)
-
- Classic Matching strategies:
 - Jaro Distance
 - Bigram Distance
 - Hamming Distance
 - Edit Distance
 - Reverse Hamming Distance



Match Transformation

1 - Pair Generation



The screenshot shows the "Properties" panel for the Match Transformation component, specifically the "Match Type" section:

- General**: General properties.
- Ports**: Input and output ports.
- Match Type**: Selected "Field Match (Single Source)" (checkbox checked). Other options include "Field Match (Dual Source)", "Identity Match (Single Source)", and "Identity Match (Dual Source)".
- Strategies**: Matching strategies.
- Match Output**: Options for handling matched data.
- Advanced**: Advanced settings.

The "Properties" table on the right is currently empty.

Input ports:

- Unique Sequence ID
- Group Key
- Sorted Data
- Match fields

Algorithm Based

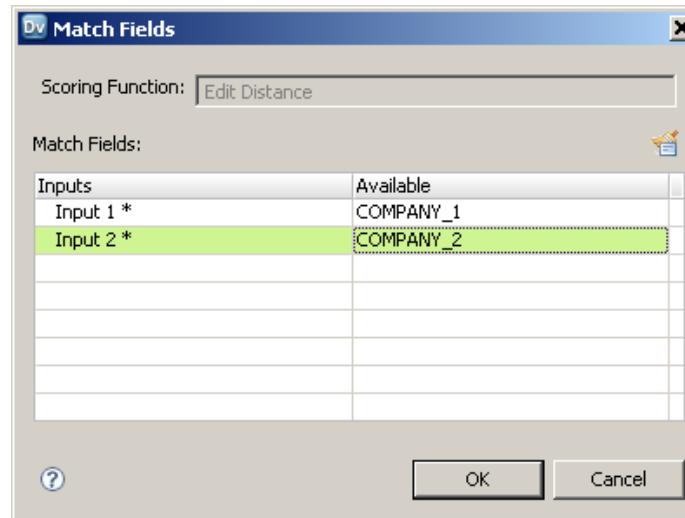
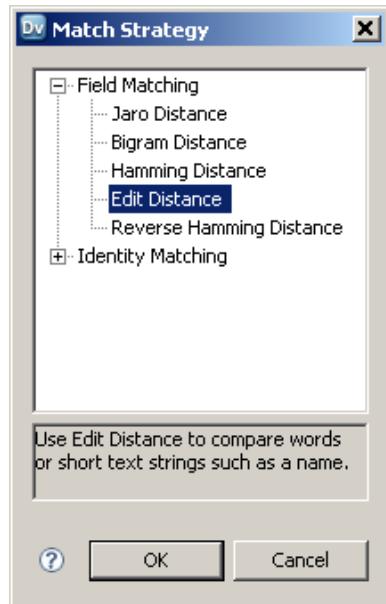
- Single/Dual Source

Identity (covered later)

- Single/Dual Source

Match Transformation

2 - Strategies



A screenshot of the Informatica Properties window. The left sidebar shows tabs for General, Ports, Match Type, Strategies (which is selected and highlighted with a red box), Match Output, and Advanced. Under "Strategies", there are two sections: "Define match strategies" and "Use match rule". The "Define match strategies" section contains a table with columns: Match Strategy, Custom Name, Weight, Match Fields, and Properties. The table data is as follows:

Match Strategy	Custom Name	Weight	Match Fields	Properties
Bigram Distance		0.6	COMPANY_1,COMPANY_2	Both Null Match Value: 0.5, Null Mat...
Bigram Distance		0.9	ADDR1_1,ADDR1_2	Both Null Match Value: 0.5, Null Mat...
Edit Distance		0.4	ADDR2_1,ADDR2_2	Both Null Match Value: 0.5, Null Mat...
Edit Distance		0.5	ADDR3_1,ADDR3_2	Both Null Match Value: 0.5, Null Mat...
Hamming Distance		0.6	ADDR4_1,ADDR4_2	Both Null Match Value: 0.5, Null Mat...
Edit Distance		0.3	COUNTRY_1,COUNTRY_2	Both Null Match Value: 0.5, Null Mat...

Use match rule

Name: **Unspecified**

Match Transformation

3 – Match Output



- Clustered or Matched Pairs
- Select threshold that must be met for records to be identified as a match
- Choose the Scoring method

The screenshot shows the 'Properties' dialog box for a Match Transformation component. The 'Match Output' tab is selected in the left sidebar. The main area displays the 'Match output type:' section, which contains a list box with three options: 'Match Output Type', 'Clusters' (which is checked and highlighted with a red box), and 'Matched Pairs'. To the right of this list is a 'Properties:' table:

Name	Value
Cache Directory	.
Cache Size	400000
Threshold	0.9
Scoring Method	Link Score

Example – product data

Type	Material	Shelf	Weight	Quantity	Color
SP	CHKS IN JY CKN	24M	3KG	X6	Red
SC	CHKS IN JY CKN + BF	12M	1.3KG	X6	Red

EDIT BIGRAM HAMMING HAMMING HAMMING EDIT

0.5 0.83871 0.333 0 1 1

Weights

0.734402

Define the threshold that must be met before records will be output as a possible match

Comparison Transformation

- Evaluates the similarity between pairs of input strings and calculates the degree of similarity for each pair as a numerical score.
- To configure, select a pair of input columns and assign a matching strategy to them.
- Outputs match scores in a range from 0 to 1, where 1 indicates a perfect match.
 - The strategies available are also available in the Match transformation.
 - Used to define match comparison operations in a matching mapplet.
 - Multiple Comparison transformations can be added to the mapplet.



Comparison

Comparison Transformation

The screenshot shows the Informatica PowerCenter interface with the 'Properties' tab selected. On the left, the 'Strategies' section is open, showing various matching algorithms under 'Field Matching'. The 'Bigram Distance' option is highlighted. In the center, the 'Inputs' section displays two input ports: 'Input_1' and 'Input_2', both mapped to 'COMPANY_1' and 'COMPANY_2' respectively. On the right, the 'Parameters' section contains fields for 'Single null match value' (set to 0.5) and 'Both null match value' (set to 0.5). The top part of the interface shows the flow of data from 'NEWINPUT' to the 'bigram' transformation.

Expects pairs
of records to
be passed to
it

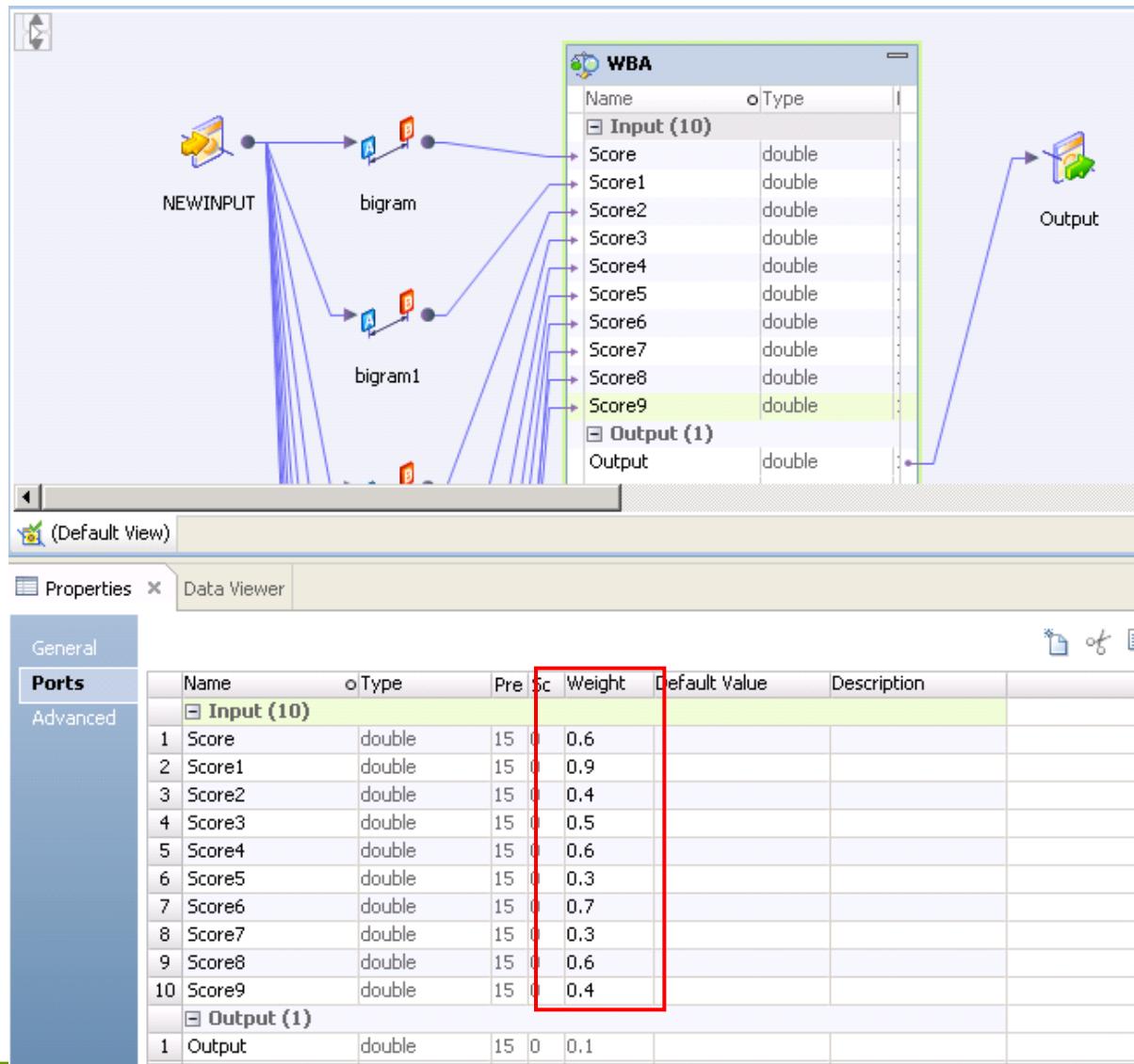
Outputs a
Score

Specify the
Algorithm to
use

Specify the
Input ports

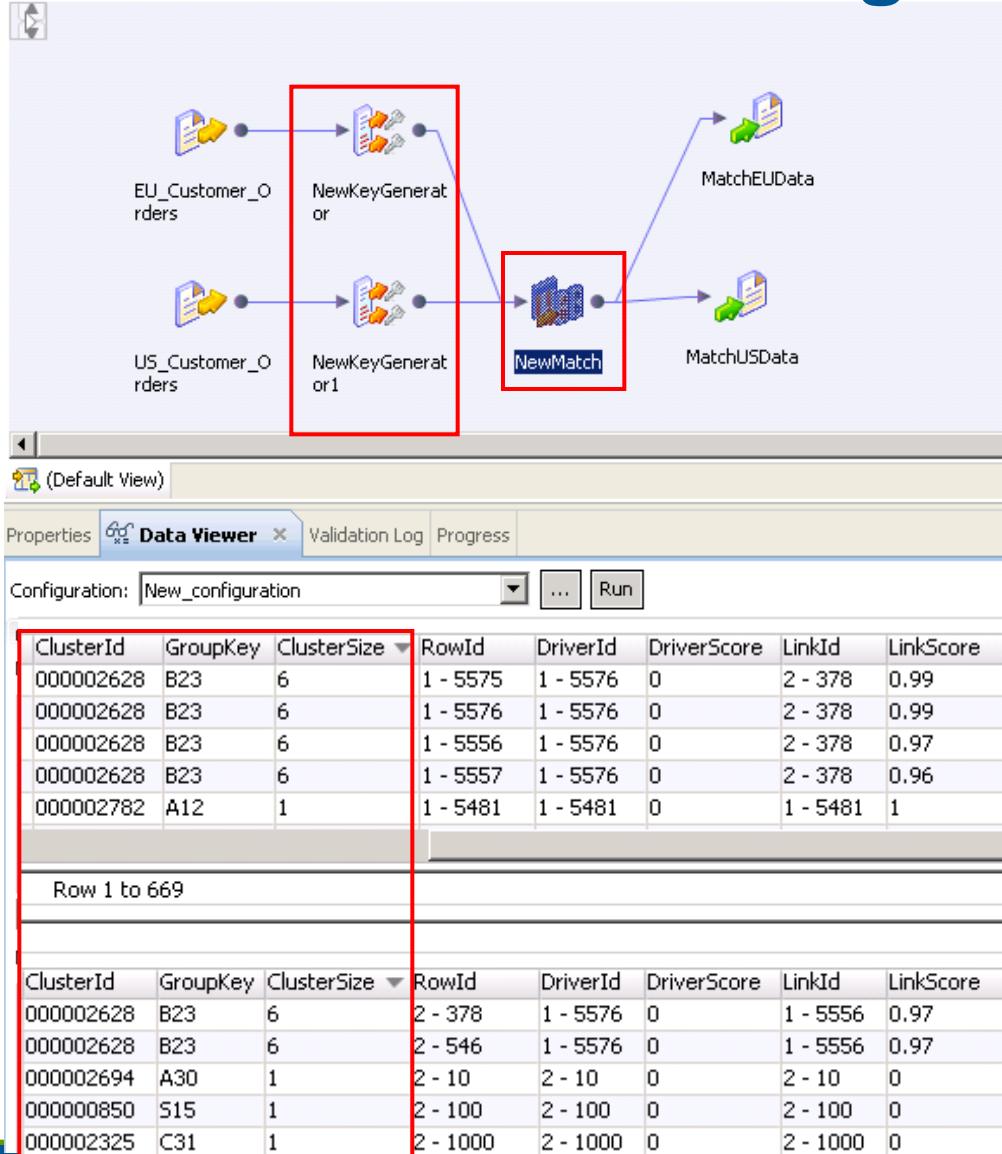
Define Match
Parameters

Weighted Average Transformation



- **Inputs:**
Similarity scores
- **Outputs:**
Weighted Average of Similarity scores

Dual-Source Matching



Select Dual Source Pair Generation Option

Two Key Generators to group data

Single Match Transformation

An output group per source

ClusterID is the same for records in the same group

Identity Matching

What is Identity Matching?

- Identity Matching delivers next generation linguistic and statistical matching algorithms to ensure highly accurate matching out of the box for over 60 countries
- Identity matching enables a business user to deliver accurate matches across multiple languages
- Emulates a human expert's ability to determine a match based on numerous fields & attributes
- Despite data having errors, variation, and duplication, Identity delivers the highest possible reliability when, matching data based on names, addresses, descriptions, and other identification data
- Identity Matching works through the use of the prebuilt *population* and matching *strategies (purpose)*

20 Common Data Errors & Variation

Variation or Error	Example
Sequence errors	<ul style="list-style-type: none"> • Mark Douglas or Douglas Mark
Involuntary corrections	<ul style="list-style-type: none"> • Browne – Brown
Concatenated names	<ul style="list-style-type: none"> • Mary Anne, Maryanne
Nicknames and aliases	<ul style="list-style-type: none"> • Chris – Christine, Christopher, Tina
Noise	<ul style="list-style-type: none"> • Full stops, dashes, slashes, titles, apostrophes
Abbreviations	<ul style="list-style-type: none"> • Wlm/William, Mfg/Manufacturing
Truncations	<ul style="list-style-type: none"> • Credit Suisse First Bost
Prefix/suffix errors	<ul style="list-style-type: none"> • MacDonald/McDonald/Donald
Spelling errors	<ul style="list-style-type: none"> • P0rter
Typing errors	<ul style="list-style-type: none"> • Beht

Variation or Error	Example
Transcription mistakes	<ul style="list-style-type: none"> • Hannah, Hamah
Missing tokens	<ul style="list-style-type: none"> • George W Smith
Extra tokens	<ul style="list-style-type: none"> • George Smith, Smith
Foreign sourced data	<ul style="list-style-type: none"> • Khader AL Ghamdi, Khadir A. AlGamdey
Unpredictable use of initials	<ul style="list-style-type: none"> • John Alan Smith, J A Smith
Transposed characters	<ul style="list-style-type: none"> • Johnson, Jhonson
Localization	<ul style="list-style-type: none"> • Stanislav Milosovich – Stan Milo
Inaccurate dates	<ul style="list-style-type: none"> • 12/10/1915, 21/10/1951, 10121951, 00001951
Transliteration differences	<ul style="list-style-type: none"> • Gang, Kang, Kwang
Phonetic errors	<ul style="list-style-type: none"> • Graeme – Graham

Populations

- **Populations contain key building algorithms that have been developed for specific countries and languages.**
- **Rules differ depending on the country/language**
 - **E.G. when building keys using the UK population:**
 - Name field: it assumes the surname is on the right of the field
 - Organization names: assumes the major part of the name is on the left
 - Address: St, Rd, Ave are all markers. The word before is typically the street name
 - Rules differ for each field – for example with the name field Bob = Robert but for address Bob <> Robert

Identity Populations – sample rules

USA

Category Name	Rule Type	Examples
Noise Word	Word is Deleted	e.g. THE, AND
Company Word Delete	Word is Deleted	e.g. INC, LTD, CO
Company Word Skip	Word is marked Skip	e.g. DEPARTMENT, ASSOCIATION
Personal Title Delete	Word is Deleted	e.g. MR, MRS, DR, JR
Nickname Replace Diminutives	Word and its Diminutives are Replaced	e.g. CATH(E,IE,Y) => CATHERINE
Nickname Replace	Word is Replaced	e.g. MIKE => MICHAEL
Word Replace	Word is Replaced	e.g. SVCS => SERVICES
Secondary Lookup	Word generates additional search ranges	e.g. AL => ALBERT, ALFRED

Germany

Category Name	Rule Type	Examples
Noise Word	Word is Deleted	e.g. DAS, UND
Company Word Delete	Word is Deleted	e.g. AG, GMBH, KG
Company Word Skip	Word is marked Skip	e.g. ABTEIL, VEREIN
Personal Title Delete	Word is Deleted	e.g. HR., FR, FRL, DR.,
Nickname Replace Diminutives	Word and its Diminutives are Replaced	e.g. KATHY => CATHERINE
Nickname Replace	Word is Replaced	e.g. HANS => JOHANNES
Word Replace	Word is Replaced	e.g. DIENSTE => DIENST
Secondary Lookup	Word generates additional search ranges	e.g. AL => ALBERT, ALFRED, ALFONS

Match Type – Pair Generation

The screenshot shows the Informatica PowerCenter interface for a Match Type - Pair Generation job named 'mt_IdentityMatch'. The job has 13 inputs. On the left, a source icon 'All_CustomerOrders' is connected to the 'Input (13)' table. The table lists fields: SequenceId (bigint), GroupKey (string), ID (integer), COMPANY (string), ADDR1 (string), ADDR2 (string), ADDR3 (string), and ADDR4 (string). A target icon 'Identity_Match_Customer' is connected to the output of the table. The Properties panel on the left is set to 'Match Type' and shows 'Identity Match (Single Source)' selected. The Properties grid on the right contains the following configuration:

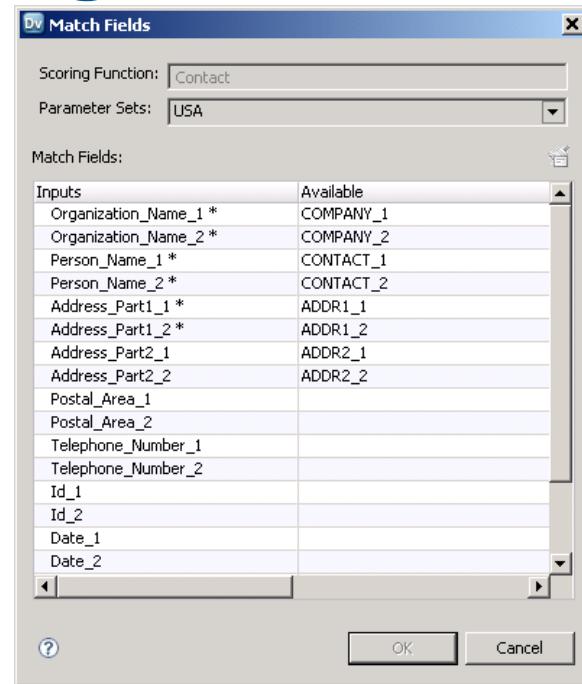
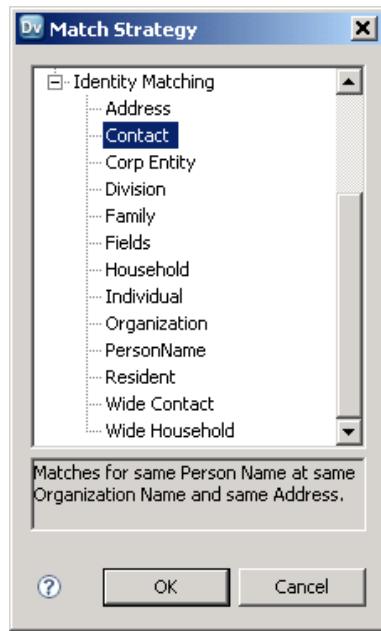
Name	Value
Population	USA
Key Level	Standard
Key Type	Organization_Name
Search Level	Typical
Key Field	COMPANY
Index Folder	C:\Identity\index
Cache Directory	C:\Cache

On the right side of the interface, there is a vertical stack of labels: Population, Key Level, Key Type, Search Level, Key Field, Index Folder, and Cache Directory, each aligned with its corresponding configuration in the Properties grid.

Match Type

- **Key Level, Search Level** – specifies how hard Identity will work to find a candidate.
- **Key Field, Key Type** – specifies which input should be used for keying, and also what type of field it is (*Organization Name, Contact or Address*). Identity logic will change depending on the type selected.
- **Index folder:** The key index folder where the index and data will be written.

Identity Matching



The Properties grid displays match strategies. Under 'Strategies', there is one entry:

Match Strategy	Custom Name	Weight	Match Fields	Properties
Contact		0.5	COMPANY_1,COMPANY_2,CONTACT...	Population: USA, Match Level: Typical

Below this, the 'Use match rule' section is shown with 'Name: Unspecified' and a 'Match Fields...' button.

Identity Match Strategy

- For each Identity Match Strategy, three Match Levels are available:
 - Typical
 - Accepts reasonable matches
 - Default if no Match_Level specified
 - Conservative
 - Accepts close matches
 - Loose
 - Accepts matches with a higher degree of variation

Match Output - Processing

- Identity clustering can only be used with Identity Pair Generation.
 - It is possible to group using the key generator (instead of Identity) and match using Identity matching. In this case check “Field Match” on the Match Type tab

Configuration: (Use Default Settings) Show: All Outputs

Output
Name: mt_IdentityMatch.Output

	ADDR4	COUNTRY	CONTACT	TITLE	PHONE	EMAIL	ClusterId	GroupKey	ClusterSize	RowId	DriverId	DriverScore	LinkId	LinkScore
1	41804	ES	MR JAMES ABRIL	Mgr Coord	Tel: + 34 9644 ...	luiscernuda@hi...	000000001	<null>	1	1 - 5000	1 - 5000	1	1 - 5000	1
2	ONA 8030	ES	Miss GABI HER...	Sr Sys Analyst...	Tel: + 34 9657 ...	santacoloma@a...	000000002	<null>	2	1 - 5001	1 - 5001	1	1 - 5001	1
3	ONA 8030	ES	Ms GABI Cleary	Programmer	Tel: + 34 9657 ...	coloma@alcam...	000000002	<null>	2	1 - 5002	1 - 5001	1	1 - 5001	0.8400000
4	ONA 43206	ES	Mr CRISTIAN P...	Business Analyst	Tel: + 34 9628 ...	almirall@reque...	000000003	<null>	1	1 - 5003	1 - 5003	1	1 - 5003	1
5	IZ 6220	ES	MR. JOSHUA M...	SR Computer S...	Tel: + 34 9645 ...	granada@alca...	000000004	<null>	1	1 - 5004	1 - 5004	1	1 - 5004	1
6	IA 46190	ES	Mr. WILLIE PER...	Computer Speci...	Tel: + 34 9658 ...	virgendetlosdesa...	000000005	<null>	2	1 - 5005	1 - 5005	1	1 - 5005	1
7	IA 46190	ES	Mr RALPH PEREZ	Computer Speci...	Tel: + 34 9658 ...	virgendetlosdesa...	000000005	<null>	2	1 - 5006	1 - 5005	1	1 - 5005	0.8400000
8	IRIA 39300	ES	MR FEDERICO...	Manager Data...	Tel: + 34 9664 ...	pablogarnica@li...	000000006	<null>	1	1 - 5007	1 - 5007	1	1 - 5007	1
9	IA 46470	ES	HELENA SANCH...	Vice President	Tel: + 34 9668 ...	julionebot@alca...	000000007	<null>	1	1 - 5008	1 - 5008	1	1 - 5008	1
10	BA 14640	ES	Mr ROBERTO F...	Business Manag...	Tel: + 34 9641 ...	unionycultura@...	000000008	<null>	2	1 - 5009	1 - 5009	1	1 - 5009	1
11	BA 14640	ES	Mr. ROBERTO HO	President	Tel: + 34 9641 ...	unionycultura@...	000000008	<null>	2	1 - 5010	1 - 5009	1	1 - 5009	0.8400000
12	ES 33150	ES	MS CLARE CAS...	SR Computer S...	Tel: + 34 9643 ...	doctorfleming@...	000000009	<null>	2	1 - 5011	1 - 5011	1	1 - 5011	1
13	ES 33150	ES	Miss CLARE CA...	President	Tel: + 34 9643 ...	doctorfleming@...	000000009	<null>	2	1 - 5012	1 - 5011	1	1 - 5011	1
14	ES 10620	ES	MR DAVID STAR	Analyst	Tel: + 34 9617	000000100	...	2	1 - 5013	1 - 5012	1	1 - 5012	1

List of Identity Populations

- Americas
 - Argentina
 - Brazil
 - Canada
 - Chile
 - Mexico
 - Peru
 - USA
- Industry Solutions
 - AML
 - OFAC
- APAC
 - Australia
 - China (5)
 - India
 - Indonesia
 - Japan (3)
 - Korea (2)
 - Malaysia
 - New Zealand
 - Philippines
 - Singapore
 - Taiwan
 - Thailand (2)
 - Hong Kong
 - Vietnam
- EMEA
 - Arabic (3)
 - Belgium
 - Czech Republic
 - Denmark
 - Finland
 - France
 - Germany
 - Greece (2)
 - Hungary
 - Ireland
 - Italy
 - Luxembourg
 - Netherlands
 - Norway
 - Poland
 - Portugal
 - Spain
 - Sweden
 - Switzerland
 - Turkey
 - United Kingdom

50 countries
65 populations e.g. China
has 5 populations

Automatic Data Consolidation

Association Example

- If we match on all of the columns below, the three records would not be identified as matching.

ID	Name	Address	City	State	Zip	SSN
1	David Jones	100 All Saints Ave	New York	NY	10547	987-65-4320
2	Dennis Jones	1000 Alberta Rd	New Jersey	NY		987-65-4320
3	D. Jones	All Saints Ave	New York	NY	10547-1521	

- In order to identify all three of these records as matching, you need to match on two different criteria:
 - 1) Name and Address
 - 2) Name and SSN

Association Transformation

ID	Name	Address	City	State	Zip	SSN	Name and Address Cluster ID	Name and SSN Cluster ID	Assoc Cluster ID
1	David Jones	100 All Saints Ave	New York	NY	10547	987-65-4320	1	1	1
2	Dennis Jones	1000 Alberta Rd	New Jersey	NY		987-65-4320	2	1	1
3	D. Jones	All Saints Ave	New York	NY	10547 -1521		1	2	1

- After matching on name and address, record 1 and 3 are in the same cluster, however record 2 is in a different cluster
- After matching on name and SSN, record 1 and 2 are in the same cluster and record 3 is in a different cluster
- The Association transformation creates links between records that share duplicate characteristics across more than one data field so they are treated as members of a single set in data consolidation

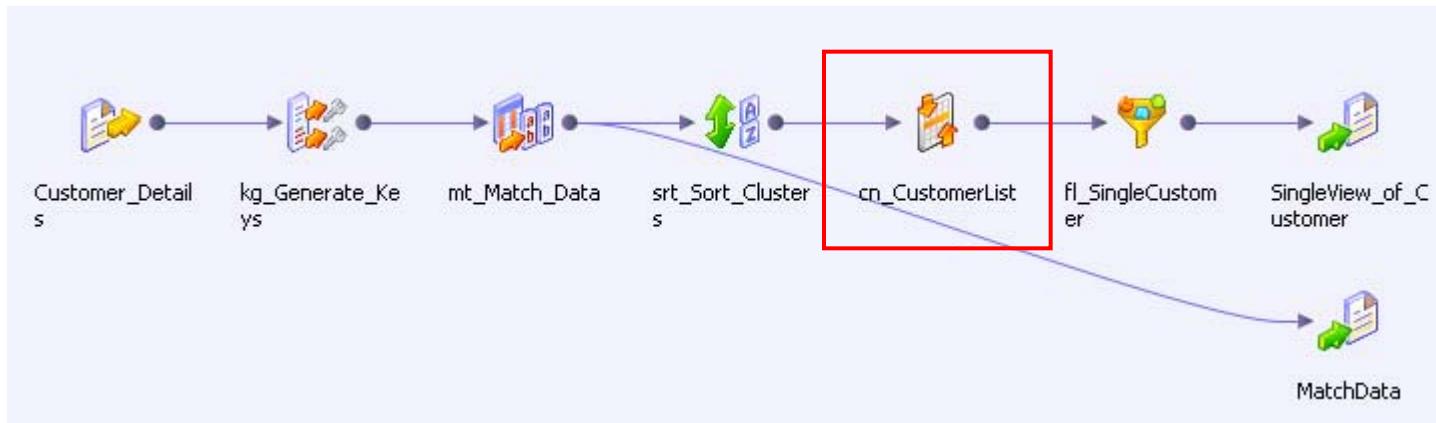
Consolidation Transformation

- Create single version of the truth
- Merges related records, eliminating duplicates (de-duping)
- Append data from additional data set
- Take best data based on rule and/or replacing inaccurate data
- Example: *Consolidation rule = longest string of matched records for each field*

Nick Jones	755 Tramway Av	Onalaska, WI 54650	(555) 555-5555	
Nicholas Jones	755 Tramway Av	Onalaska, WI 54650		njones@informatica.com

Nicholas Jones	755 Tramway Av	Onalaska, WI 54650	(555) 555-5555	njones@informatica.com
----------------	----------------	-----------------------	----------------	--

Consolidation Transformation - Create Survivor Record



Input data from Association or Match Transformation

Properties Data Viewer

General Ports Consolidation Advanced

Group By	Strategy	Output
	Use default	ID
	Average	COMPANY
	Longest	ADDR1
	Maximum	ADDR2
	Minimum	ADDR3
	Most frequent non-blank	ADDR4
	Most frequent non-blank	COUNTRY
	Longest	CONTACT
	Longest	TITLE
	Longest	PHONE
	Most frequent non-blank	EMAIL
	Use default	ClusterId

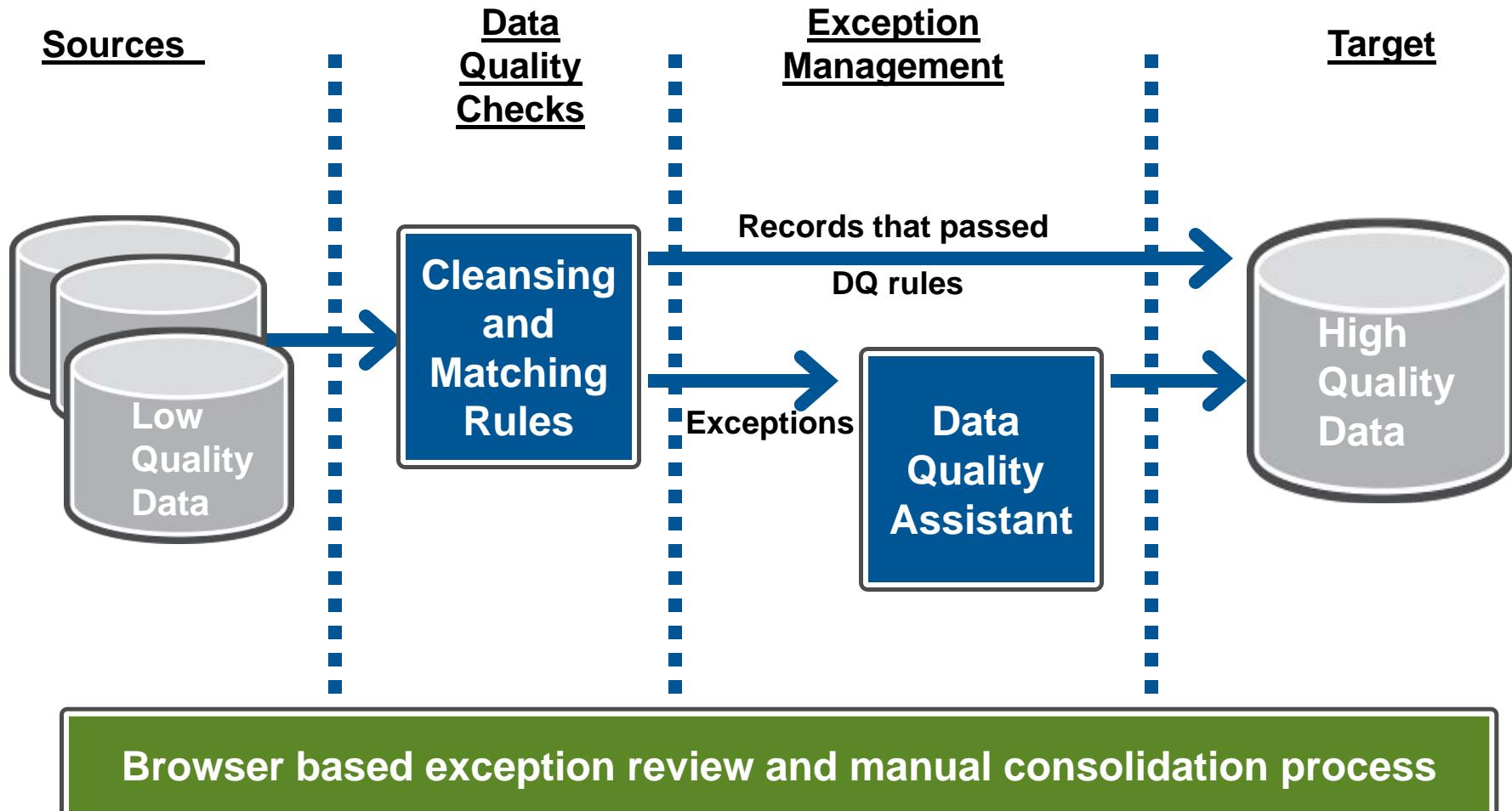
- **Consolidation functions:**
 - Most frequent
 - Most frequent nonblank
 - Longest
 - Shortest
 - Minimum (integer)
 - Maximum (integer)
- **Select Group By Field**

Consolidation Functions

- **MostFrequent**
 - Returns the most frequently occurring value for the port, including blank and null values
- **MostFrequentNonBlank**
 - Returns the most frequently occurring value for the port, ignoring blank and null values
- **Longest**
 - Returns the longest value
- **Shortest**
 - Returns the shortest value
- **Minimum (integer)**
 - Returns the minimum value
- **Maximum (integer)**
 - Returns the maximum value

Data Quality Assistant

Exception Management process



The Data Quality Assistant

- The DQA is a web based (Analyst) application for record management. It works in conjunction with data quality Mappings to sort and filter data records by data quality issue. It can be used to:
 - Manage bad records
 - Users can easily view and update bad data in a table through an easy to use GUI
 - Consolidate duplicate records into a Master Record
 - Users can create a Master record from multiple duplicate records
 - View the audit trail
 - View the audit trail on changes made to the data

Required Tables

- The DQA uses 3 staging tables:
- Bad Record Management
 - The main data table. This table will store your data as well as the matching information after matching is performed. E.g. dqacustomer
 - The issue table. This must have the name of the main data table suffixed with _issue e.g. dqacustomer_Issue. This table stores the issue identified per field
- Consolidation
 - Duplicate record table. This will be used to hold the duplicate record clusters for consolidation.
 - Within each table there are certain columns that must exist and are reserved for use internally by the DQA

Bad Record Table

- Data Quality Assistant (DQA) allows users to review and correct exceptions
- Audit trail of manual changes

The screenshot shows the INFORMATICA Analyst interface. The top navigation bar includes 'INFORMATICA Analyst', 'Administrator', 'Sign Off', 'User Preferences', 'Help', 'Search All', and a 'Go' button. Below the navigation is a toolbar with 'Browse: Projects' and a 'CUSTOMER' tab. The main area has tabs for 'Bad Records' (which is selected), 'Audit Trail', and 'Properties'. A sub-header 'CUSTOMER - Bad Records' is displayed above a table. The table has columns: Quality Issue, ADDR2, ADDR3, ADDR4, COUNTRY, CONTACT, TITLE, PHONE, and EMAIL. The 'Quality Issue' column dropdown is set to 'All' and shows several invalid entries like 'invalid_city', 'invalid_company', etc. The table displays 994 records. The bottom right corner of the table area shows a total of 994 records.

Quality Issue:	ADDR2	ADDR3	ADDR4	COUNTRY	CONTACT	TITLE	PHONE	EMAIL
All	MOUNT VERN...	NY	10550	USA	MR ISACC	SCIENTIFIC S...	9146646671	IWALI
A&P	.. MOUNT VERN...	NY	10550	USA	MR ISACC	SCIENTIFIC S...	9146646671	IWALI
GAP SO	.. MOUNT VERN...	NY	10550	USA	MR ISACC	SCIENTIFIC S...	9146646671	IWALI
NATION	.. FT MEADE	MD	20755	USA	MR JOHN CAM...	SENIOR COMP...	4108594497	JOHN
NATIONAL SE...	.. FT MEADE	MD	20755	USA	MR JOHN CAM...	GRAPHIC DES...	4108594497	JOHN
PENDLETON G...	14251 HART F...	CENTERVILLE	VA	20121	MS YOUNG	OWNER/CONS...	7036314124	YOUN
PENDLETON G...	14251 HART F...	CENTERVILLE	VA	20121	MS KIM J YOUNG	OWNER/CONS...	6314124	YOUN
PENDLETON G...	14251 HART F...	CENTERVILLE	VA	20121	MS KIM YOUNG	OWNER/CONS...	6314124	YOUN
PR DELTEC INC	1265 SAN JUA...	SAN JUAN	PR	009-08877	MR KENNETH Y...	CONTROLLER		KEN@
SUPERPETZ	1275 W PATRI...	FREDERICK	MD	217	MR RUSTICUS ...	DATABASE AD...	3016319010	SUPER
ABBOTT LABO...	5AN AP6A-2 1...	ABBOTT PARK		60064-3500	MR WILLIAM H...	PROGRAMMER	7089376501	BILL@
ABBOTT LABO...	5AN AP6A-2 1...	ABBOTT PARK	IL	60064-3500	MR BILL HATCH	PROGRAMMER	7089376501	BILL@
ABBOTT LABO...	5AN AP6A-2 1...	ABBOTT PARK	IL	60064-3500	MR BILL HATCH	PROGRAMMER	7089376501	BILL@
AHI INC	668 N BROAD ...	ELIZABETH	NJ	7208	T A. FREEDMAN	MIS MANAGER	9083513189	T_A.C
NATIONAL SE...	NATIONAL CO...	FT MEADE	MD	20755	MR JOHNNY C...	SENIOR COMP...	4108594497	JOHN
PENDLETON G...	14251 HART F...	CENTERVILLE	VA	20121	MS KIM YOUNG	OWNER/CONS...	6314124	YOUN
PUBLIX SUPER...	7060 PALMETT...	BOCA RATON	FL	33433 3411	MS BESSY RO...	COMPUTER SP...	5613930022	BESS
TRICOR INDU...	8181 PROFESS...	LANDOVER	MD	20785	MR MICK GOR...	VICE PRESIDE...	3017316140	MICH
TRICOR INDU...	8181 PROFESS...	LANDOVER	MD	20785	MR MICHAEL G...	VICE PRESIDE...	3017316140	MICH

Duplicate Records

- DQA allows for manual record consolidation after duplicates are detected
- Audit trail of manual changes

INFORMATICA Analyst

Administrator Sign Off | User Preferences | Help ▾ Search All Go

Browse: Projects CUSTOMERDUP

Duplicate Records Audit Trail Properties Actions ▾

CUSTOMERDUP - Duplicate Records

443 Clusters: <<11 12 13 14 15 16 17 18 19 20 >>

Cluster	Master	COMPANY	ADDR1	ADDR2	ADDR3	ADDR4	COUNTRY	CONTACT
<input checked="" type="checkbox"/>	<input checked="" type="radio"/>	ABBOTT LABORATORIES	5AN AP6A-2 100 ABBOTT PK	ABBOTT PARK		60064-3500	USA	MR WILLIAM H.
<input checked="" type="checkbox"/>	<input type="radio"/>	ABBOTT LABORATORIES	5AN AP6A-2 100 ABBOTT PK	ABBOTT PARK	IL	60064-3500	USA	MR BILL HATCH
<input checked="" type="checkbox"/>	<input type="radio"/>	ABBOTT LABORATORIES	5AN AP6A-2 100 ABBOTT PK	ABBOTT PARK	IL	60064-3500	USA	MR BILL HATCH
<input checked="" type="checkbox"/>	<input type="radio"/>	ABBOTT LABORATORIES	5AN AP6A-2 100 ABBOTT PK	ABBOTT PARK	IL	60064-3500	USA	MR BILL HATCH
<input checked="" type="checkbox"/>	<input type="radio"/>	ABBOTT LABORATORIES	5AN AP6A-2 100 ABBOTT PK	ABBOTT PARK	IL	60064-3500	USA	MR BILLY HATCH
<input checked="" type="checkbox"/>	<input type="radio"/>	ABBOTT LABORATORIES	5AN AP6A-2 100 ABBOTT PK	ABBOTT PARK	IL	60064-3500	USA	MR BILL HATCH
<input checked="" type="checkbox"/>	<input type="radio"/>	ABBOTT LABORATORIES	5AN AP6A-2 100 ABBOTT PK	ABBOTT PARK	IL	60064-3500	USA	MR BILL HATCH
<input checked="" type="checkbox"/>	<input type="radio"/>	ABBOTT LABORATORIES	5AN AP6A-2 100 ABBOTT PK	ABBOTT PARK	IL	60064-3500	USA	MR WILLIAM H.
<input checked="" type="checkbox"/>	<input type="radio"/>	ABBOTT LABORATORIES	5AN AP6A-2 100 ABBOTT PK	ABBOTT PARK	IL	60064-3500	USA	MR WILLY HAT
<input checked="" type="checkbox"/>	<input type="radio"/>	ABBOTT LABORATORIES	5AN AP6A-2 100 ABBOTT PK	ABBOTT PARK	IL	60064-3500	USA	MR BILL HATCH
<input checked="" type="checkbox"/>	<input type="radio"/>	ABBOTT LABORATORIES	5AN AP6A-2 100 ABBOTT PK	ABBOTT PARK	IL	60064-3500	USA	MR WILLY HAT
<input checked="" type="checkbox"/>	<input type="radio"/>	ABBOTT LABORATORIES	5AN AP6A-2 100 ABBOTT PK	ABBOTT PARK	IL	60064-3500	USA	MR BILL HATCH
<input checked="" type="checkbox"/>	<input type="radio"/>	ABBOTT LABORATORIES	5AN AP6A-2 100 ABBOTT PK	ABBOTT PARK	IL	60064-3500	USA	MR BILL HATCH
<input checked="" type="checkbox"/>	<input type="radio"/>	ABBOTT LABORATORIES	5AN AP6A-2 100 ABBOTT PK	ABBOTT PARK	IL	60064-3500	USA	MR HATCH

Business User - Manage Bad Records

CUSTOMER - Audit Trail												
Actions		Audit Trail										
Status		User	Last Updated	Audit Notes	COMPANY	ADDR1	ADDR2	ADDR3	ADDR4	COUNTRY	CONTACT	TITLE
	Accepted	Administrator	Feb 27, 2010 8:51:23 PM G...		QUALCOMM	125 PEARL ...	BRAINTREE	MA	2,184	USA	MS IDALIA ...	AUTOMATE...
	Accepted	Administrator	Feb 27, 2010 8:51:23 PM G...		AHI INC	668 N BRO...	ELIZABETH	NJ	7,208	USA	T A. FREED...	MIS MANAG...
	Accepted	Administrator	Feb 27, 2010 8:51:23 PM G...		AHI INC	668 N BRO...	ELIZABETH	NJ	7,208	USA	T A. FREED...	MIS MANAG...
	Updated	Administrator	Feb 27, 2010 8:50:56 PM G...		QUALCOMM	125 PEARL ...	BRAINTREE	MA	2,184	USA	MS IDALIA ...	AUTOMATE...
	Updated	Administrator	Feb 27, 2010 8:50:45 PM G...		AHI INC	668 N BRO...	ELIZABETH	NJ	7,208	USA	T A. FREED...	MIS MANAG...
	Updated	Administrator	Feb 27, 2010 8:50:27 PM G...		AHI INC	668 N BRO...	ELIZABETH	NJ	7,208	USA	T A. FREED...	MIS MANAG...
	Rejected	Administrator	Feb 12, 2010 8:53:12 PM G...		A&P	45 DEMERC...	ALLENDALE	NJ	7,401	USA	MR JACOB ...	PRESIDENT
	Reproc...	Administrator	Feb 12, 2010 8:52:55 PM G...		PATHMARK	211 ELMOR...	ELIZABETH	NJ	7,202	USA	MR NEILS F...	COMPUTER ...
	Updated	Administrator	Feb 12, 2010 8:52:49 PM G...		PATHMARK	211 ELMOR...	ELIZABETH	NJ	7,202	USA	MR NEILS F...	COMPUTER ...
	Accepted	Administrator	Feb 12, 2010 8:52:33 PM G...		A&P	null	CLOSTER	NJ	7,624	USA	MS ANNA P...	PRESIDENT
	Accepted	Administrator	Feb 12, 2010 8:52:33 PM G...		A&P	120 N MAYNE	MANVILLE	NJ	8,838	USA	MR BOBBY ...	PARTNER
	Accepted	Administrator	Feb 12, 2010 8:52:33 PM G...		P&C	LINCOLN C...	LINCOLN	NH	3,251	USA	MS MALKA...	DATABASE...
	Accepted	Administrator	Feb 12, 2010 8:52:33 PM G...		AHI INC	668 N BRO...	ELIZABETH	NJ	7,208	USA	T A. FREED...	MIS MANAG...
	Updated	Administrator	Feb 12, 2010 8:45:51 PM G...		A&P	120 N MAYNE	MANVILLE	NJ	8,838	USA	MR BOBBY ...	PARTNER
Status Counts:												
	Updated: 11		Accepted: 8		Rejected: 1		Reprocess: 1					
		Page	1	of 1							Displaying 1 - 19 of 19	

Business User - Record Consolidation

INFORMATICA® Analyst

Administrator Sign Off | User Preferences | Help ▾ Search All Go

Browse: Projects CUSTOMERDUP X

Duplicate Records Audit Trail Properties Actions ▾

CUSTOMERDUP - Audit Trail

From:	02/27/2010	To:	02/27/2010	User:	All	Status:	All	Show		
							All			
Status	User	Last Updated	Audit Notes	COMPANY	ADDR1	ADDR2	Updated	COUNTRY	CONTACT	PHONE
Consolidated	Administrator	Feb 27, 2010 9:46:16 PM G...		JOHNSON &... GLODWICK ... OLDHAM			Consolidated	GBR	MR JUDAS ...	1,616,264,8... JL
Updated	Administrator	Feb 27, 2010 9:46:16 PM G...		JOHNSON &... GLODWICK ... OLDHAM			Rematch	GBR	MR JUDAS ...	1,616,264,8... JL
Consolidated	Administrator	Feb 27, 2010 9:46:13 PM G...		BOOTS	52A LONG ... NOTTINGHAM		Extracted	GBR	MRS CONC... 1,159,587,3...	CM
Updated	Administrator	Feb 27, 2010 9:46:13 PM G...		BOOTS	52 LONG R... NOTTINGHAM	null	null	GBR	MRS CONC... 1,159,587,3...	CM
Consolidated	Administrator	Feb 27, 2010 9:46:09 PM G...		SPAR	FORE STR	ST AUSTELL	BUGLE, CO...	GBR	MS PEARL ... 1,726,850,5...	CH
Consolidated	Administrator	Feb 27, 2010 9:46:09 PM G...		SPAR	FORE STR, ...	BUNGLE	CORNWALL	GBR	MR EARL R... 1,726,850,5...	CH
Consolidated	Administrator	Feb 27, 2010 9:46:09 PM G...		SPAR	FORE STR, ...	BUNGLE	CORNWALL	GBR	MR EARL W... 1,726,850,5...	CH
Updated	Administrator	Feb 27, 2010 9:46:09 PM G...		SPAR	FORE STR	ST AUSTELL	BUGLE, CO...	GBR	MR EARL R... 1,726,850,5...	CH

Status Counts: Updated: 21 Consolidated: 41 Rematch: 3 Extracted: 24

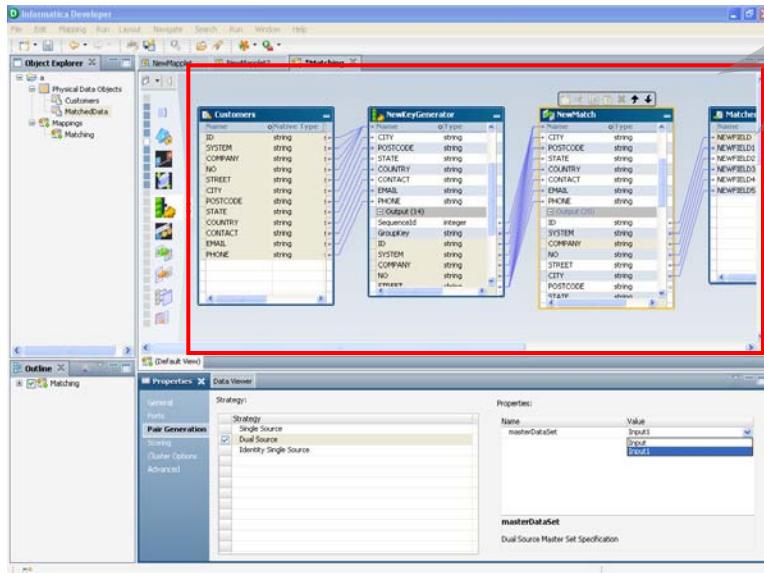
Page 1 of 1 Displaying 1 - 8 of 8

PowerCenter Integration

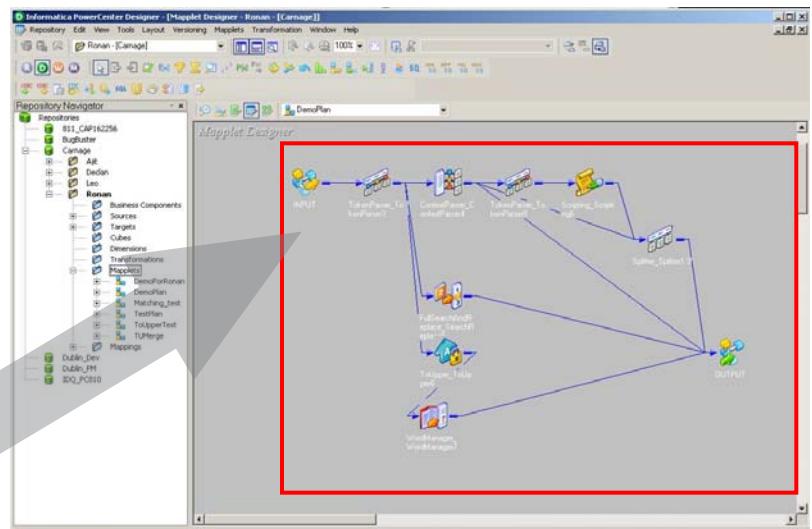
Integration with PowerCenter

- Deployment to PC for
 - Performance
 - Scalability
 - Connectivity
 - Batch access
 - Web Services
 - DQ as part of ETL process

Informatica Developer 9.0.1

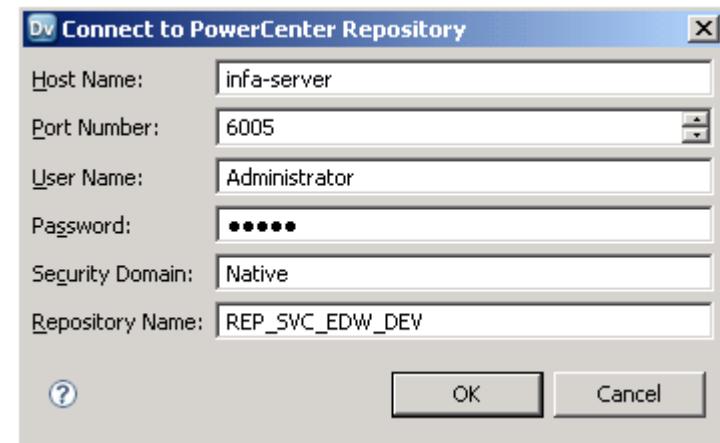
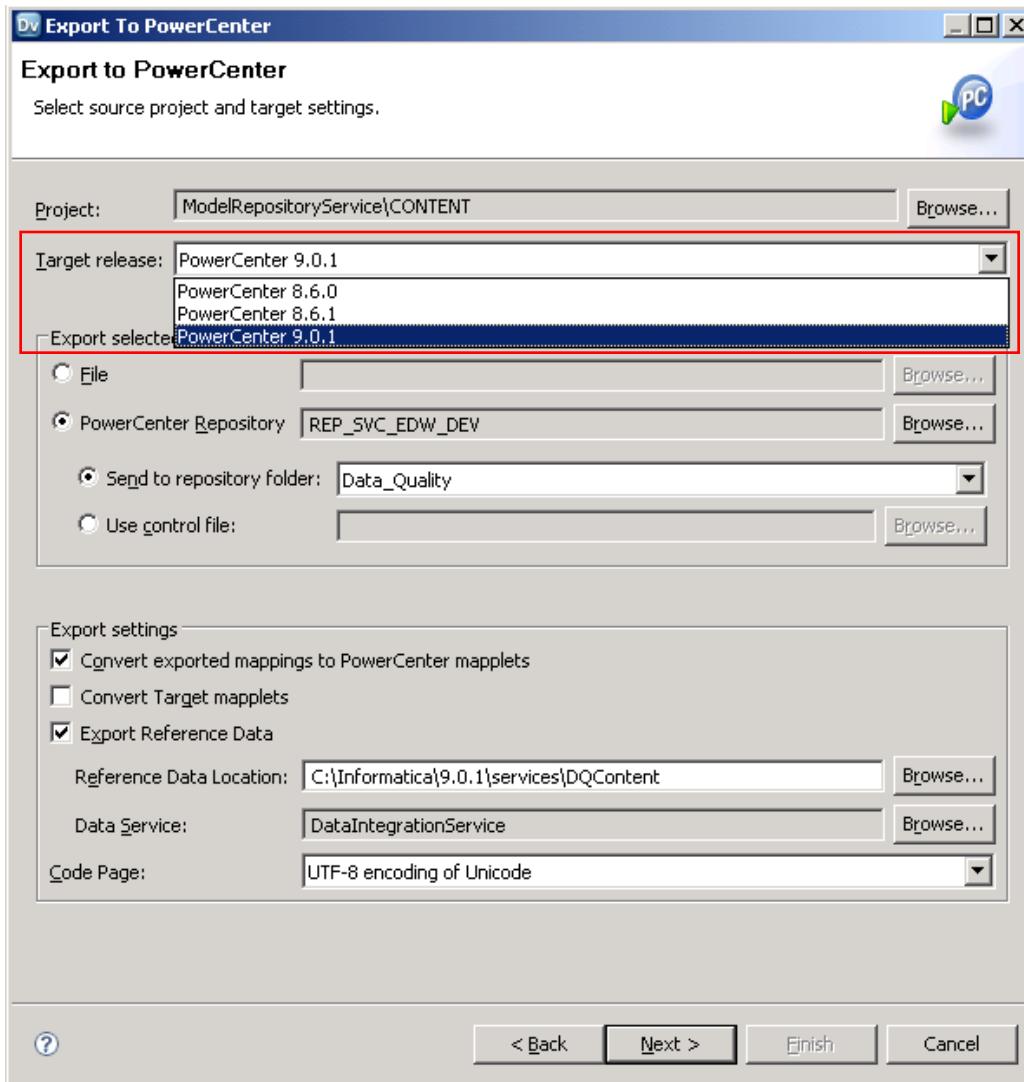


PowerCenter 8.6/8.6.1 or 9.0.1



- **Informatica Developer objects exportable to PC repository**
 - Mappings
 - Mapplets
 - Data Object Read Maps
- **Executed natively within PC**
- **No requirement to install PC Designer on the same machine as 9.0.1 Developer**

Export Options



Choose PC domain and repository

Export:

- **To file**

OR

- **To PC repository**

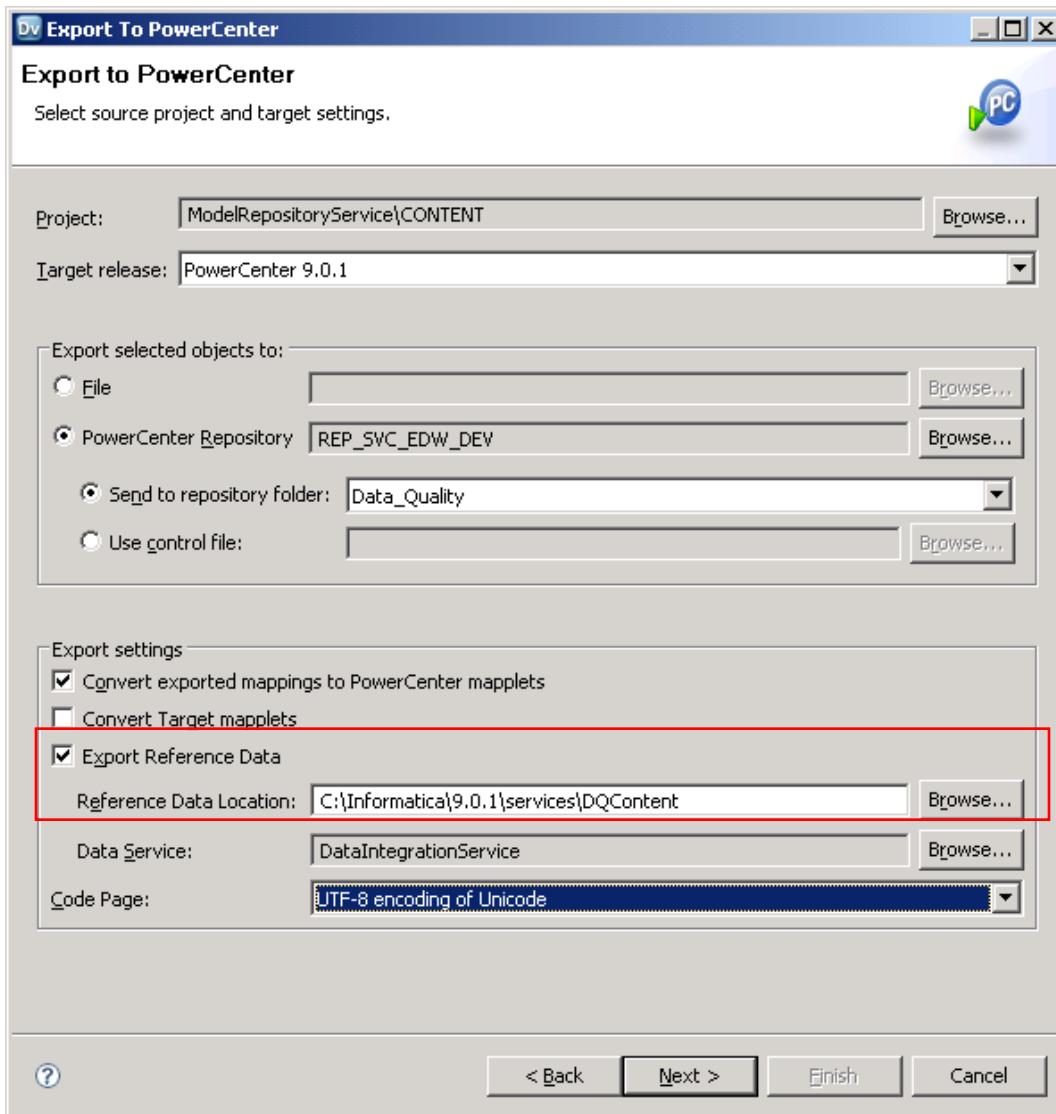
Export mappings:

- **As mappings**

OR

- **As mapplets**

Export Reference Data



Defined content dependencies are identified at Export

RTM tables converted to flat files

DQ/PC Integration Installation

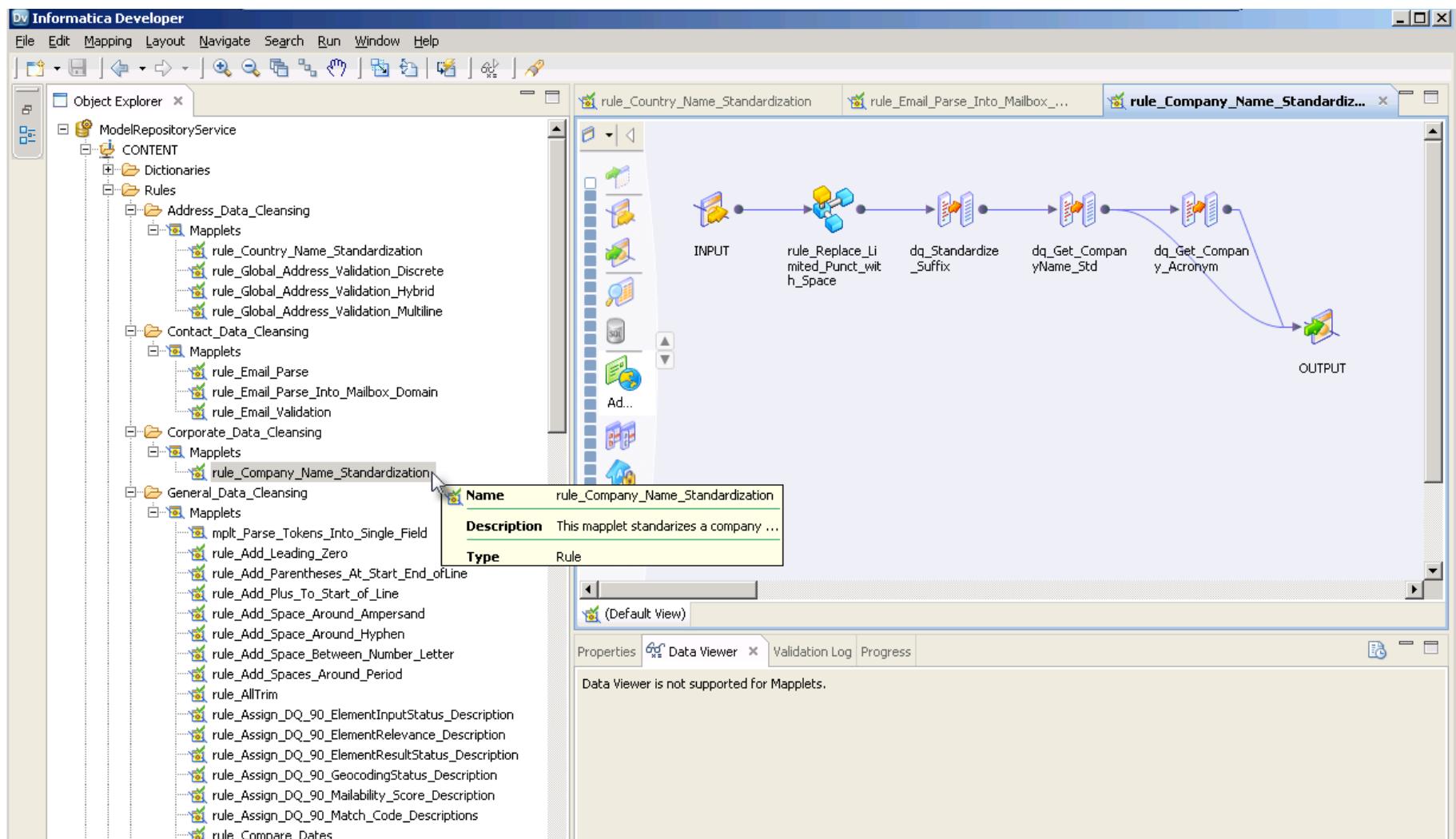
	PowerCenter 9.0.1	PowerCenter 8.6/8.6.1
IDQ 9.0.1	No separate integration installer. All required files placed by the 9.0.1 Universal installer.	DQ/PC Integration installers on both the Client and the Server side

Content

What makes up OOTB Content?

- **Maplets** – Snippets of DQ functionality used by the Developer
- **Rules** – Maplets that have been Validated as Rules for the Analyst to consume
- **Reference Tables** – Reference data used in maplets, rules, and mappings
- **Address Validation data** – Subscription data used with the Address Validator transformation
- **Identity Populations** - Contains metadata on types of personal, household, and corporate identity including algorithms that apply the metadata to input data

Pre-Built Mapplets and Rules

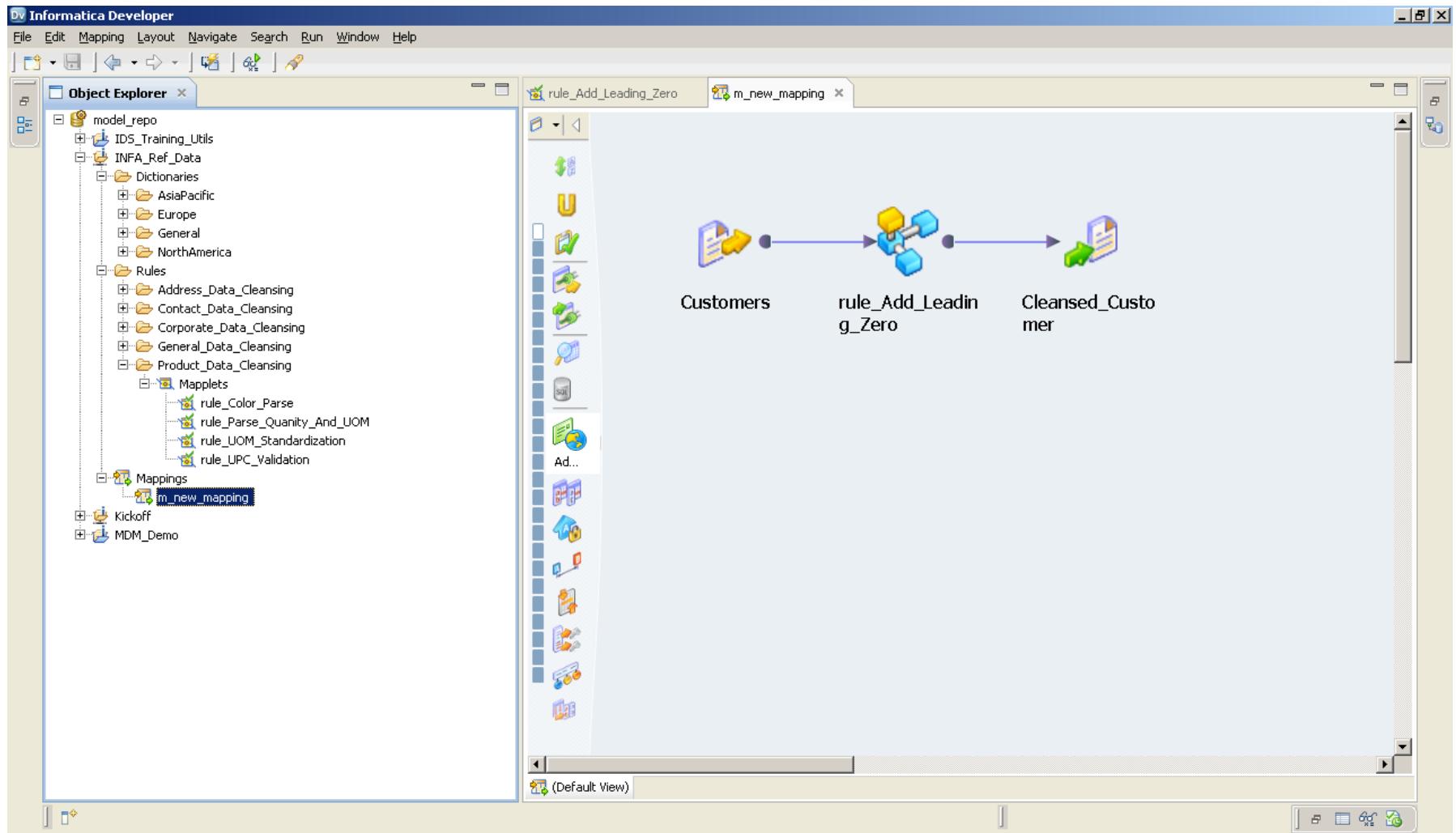


Pre-Built Reference Tables

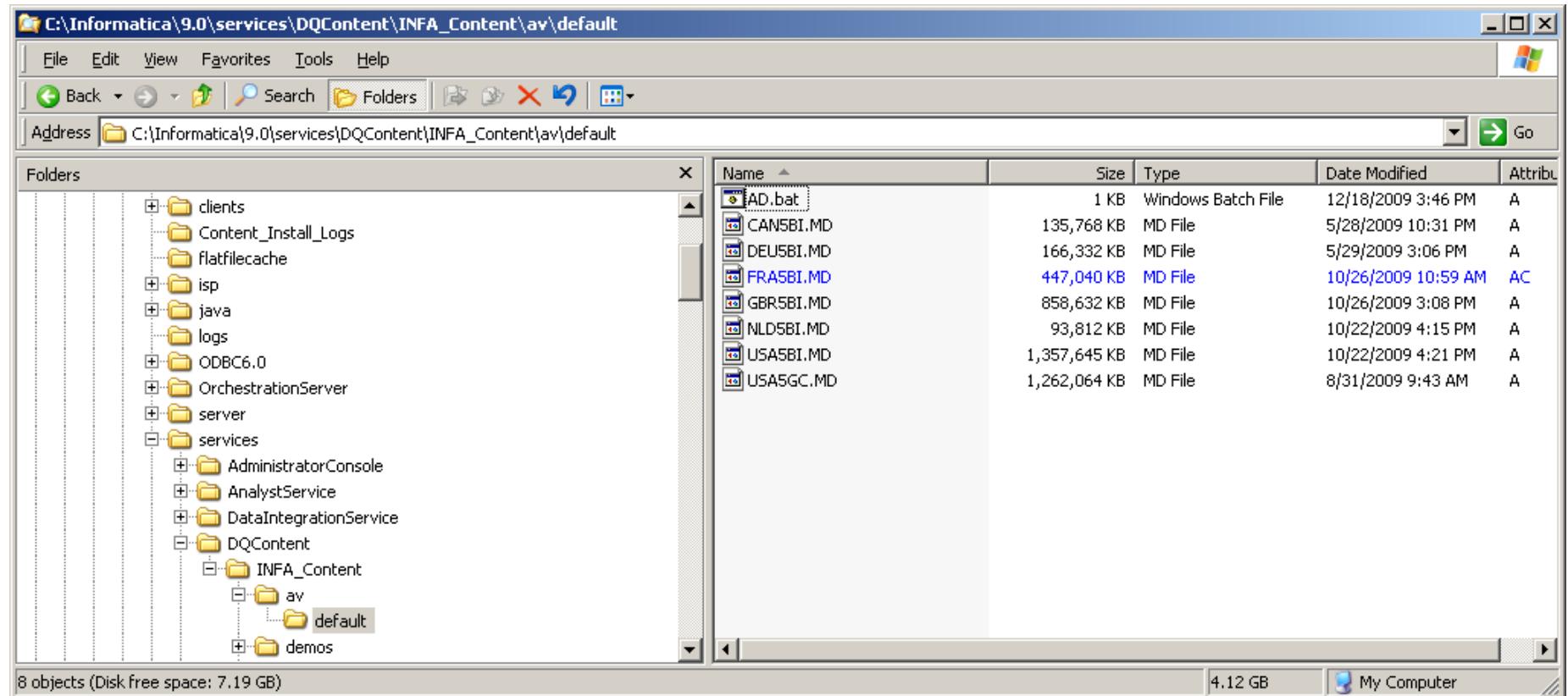
The screenshot shows the Informatica Developer application interface. On the left, the Object Explorer tree view displays a hierarchical structure of reference data. Under the 'INFA_Ref_Data' node, there are nodes for 'Dictionaries' and 'Reference Tables'. The 'Dictionaries' node has sub-nodes for 'AsiaPacific', 'Europe', 'General', and 'NorthAmerica'. The 'Reference Tables' node under 'INFA_Ref_Data' has sub-nodes for 'Australia', 'India', 'Canada', and 'USA'. Each of these sub-nodes contains several specific reference table entries. On the right, a data preview window titled 'aus_company_name_std_infa' is open, showing a table with six columns: column1, column2, column3, column4, column5, and column6. The table contains numerous rows of company names, many of which are truncated or redacted. A button labeled 'Open with Informatica Analyst...' is visible in the top right corner of the preview window.

column1	column2	column3	column4	column5	column6
A AND H TRADIN...	A AND H TRADIN...	A AND H TRADING	A & H TDG	NULL	NULL
A ARCHEN HOWE...	A ARCHEN HOWE...	A ARCHEN HOWE...	NULL	NULL	NULL
A AUSTRALIAN C...	A AUSTRALIAN C...	A AUSTRALIAN C...	AUST CALIBRA...	NULL	NULL
A BASIC CONCE...	A BASIC CONCE...	A BASIC CONCE...	NULL	NULL	NULL
A CLASS METAL ...	A CLASS METAL ...	A CLASS METAL ...	NULL	NULL	NULL
A CLEANER WOR...	A CLEANER WOR...	A CLEANER WOR...	A CLEANER WRL...	NULL	NULL
A CONSOLIDATE...	A CONSOLIDATE...	A CONSOLIDATE...	A CONSOL EMPL...	NULL	NULL
A CURTAIN FACT...	A CURTAIN FACT...	A CURTAIN FCTRY	NULL	NULL	NULL
A DART & CO	A DART & CO	A DART	NULL	NULL	NULL
A FINE KETTLE O'...	A FINE KETTLE O'...	A FINE KETTLE O ...	NULL	NULL	NULL
A GRADE CLEANI...	A GRADE CLEANI...	A GRADE CLNNG ...	NULL	NULL	NULL
A GROUP OF CO...	A GROUP OF CO...	A GROUP OF CO...	A GRP OF COS 2...	NULL	NULL
A HAUSLER PTY L...	A HAUSLER PTY L...	A HAUSLER	NULL	NULL	NULL
A MAN'S TOYSHO...	A MAN'S TOYSHO...	A MANS TOYSHO...	A MANS TOYSHO...	NULL	NULL
A MURRAY & SON...	A MURRAY & SON...	A MURRAY & SON...	A MURRAY AND S...	NULL	NULL
A NOBLE & SON L...	A NOBLE & SON L...	A NOBLE & SON	A NOBLE AND SON	NULL	NULL
A PLUS I T SUPPL...	A PLUS I T SUPPL...	A PLUS I T SUPPL...	A PLUS I T SPPLS	NULL	NULL
A RAPTIS & SON...	A RAPTIS & SON...	A RAPTIS & SONS	A RAPTIS AND S...	NULL	NULL
A ROLLEY & SONS	A ROLLEY & SONS	A ROLLEY AND S...	NULL	NULL	NULL
A SAMIOS PTY LTD	A SAMIOS PTY LTD	A SAMIOS	NULL	NULL	NULL
A SANTALUCIA &...	A SANTALUCIA &...	A SANTALUCIA &...	A SANTALUCIA A...	NULL	NULL
A SWIFT BINS	A SWIFT BINS	NULL	NULL	NULL	NULL
A&B MINTEC AUS...	A&B MINTEC AUS...	A&B MINTEC AUS...	A AND B MINTEC...	NULL	NULL
A&R WHITCOULL...	A&R WHITCOULL...	A&R WHITCOULL...	A AND R WHITC...	NULL	NULL
A1 ACE WASTE	A1 ACE WASTE	A ONE ACE WASTE	NULL	NULL	NULL
A1 ALUMINIUM D...	A1 ALUMINIUM D...	AONE ALUMINIU...	NULL	NULL	NULL
A1 ASPHALTING ...	A1 ASPHALTING ...	A1 ASPHALTING	AONE ASPHALTING	NULL	NULL

Add OOTB rules to Mappings

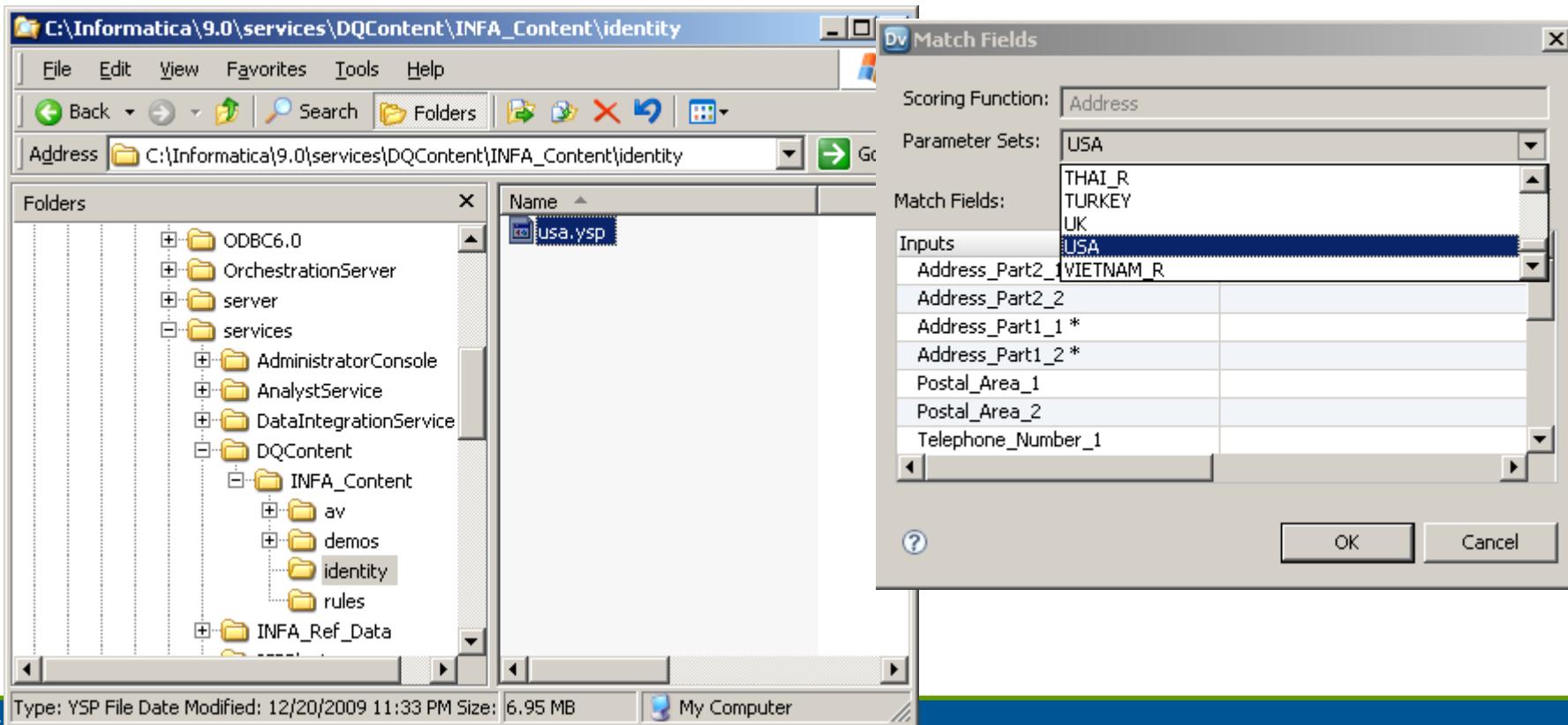


Address Validation Data



Identity Populations

- Populations need to be installed
- Parameter Sets are pre-populated in the Match transformation



Installation Tips and Tricks

- **Client and Server Install**
 - Client install has to be done first
 - Imports the mapplets
 - Server install has to be done second
 - Installs the content
 - Content is Database Specific
- **IN_901_Content_InstallationGuide.pdf**

IDQ 9.0.1 – Migration 8.6.2 to 9.0.1

Why is it Called Migration?

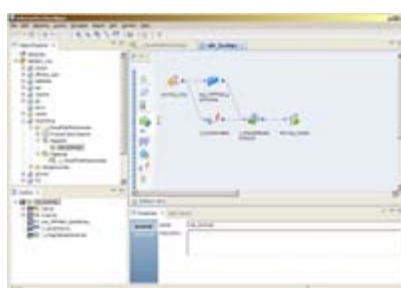
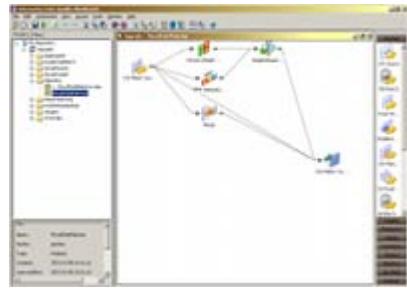
- “*Migrate and convert all user content to implement DQ logic designed in an 8.6.2 environment to an 9.0.1 environment.*”

Why is it Called “Migration?”

- Why isn't it called “Upgrade?”
 - Significant changes to components
 - Significant change from Dictionaries to Reference Tables
 - Significant change in moving Plans from one architecture to another

Overview

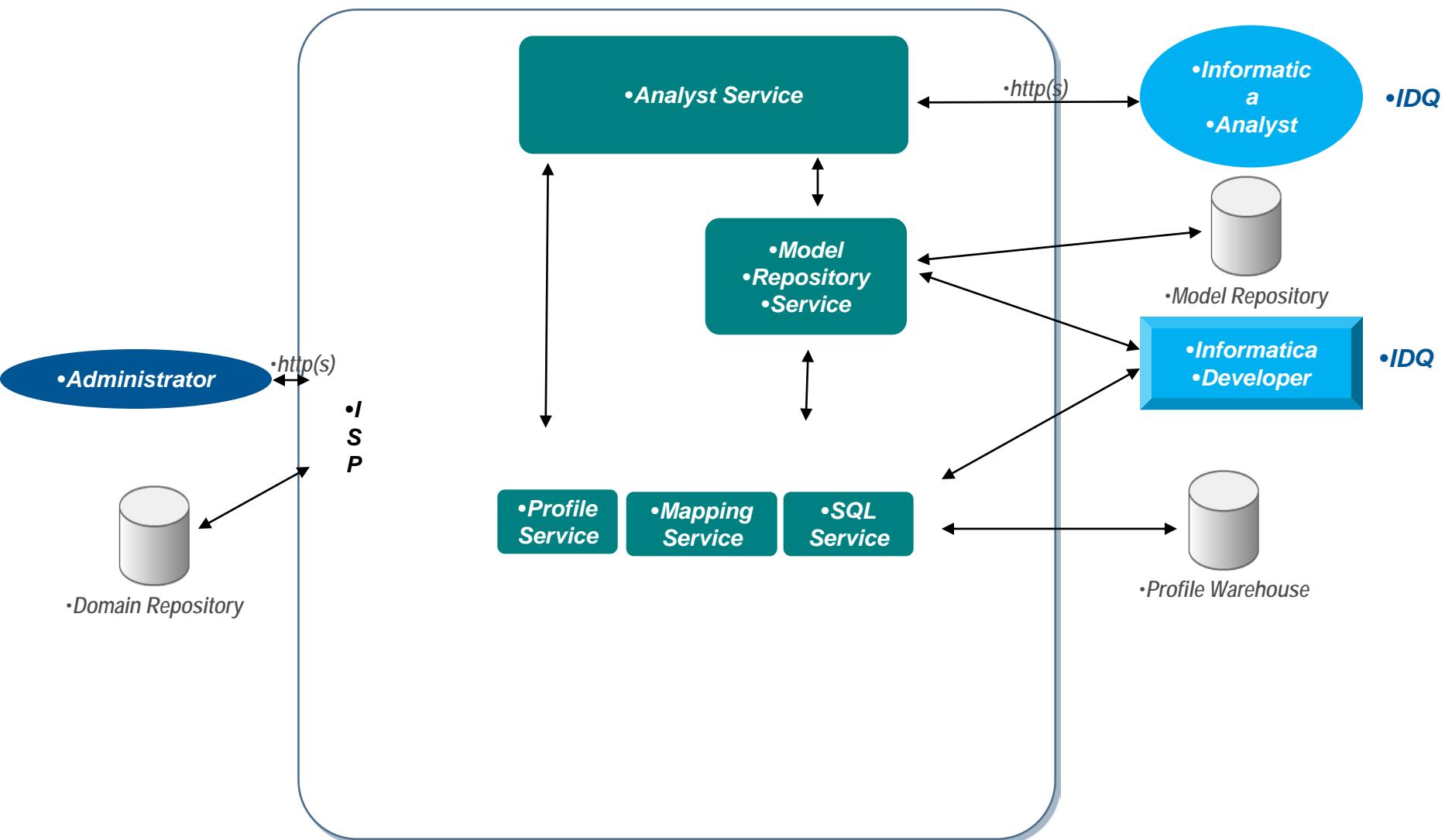
Version Differences



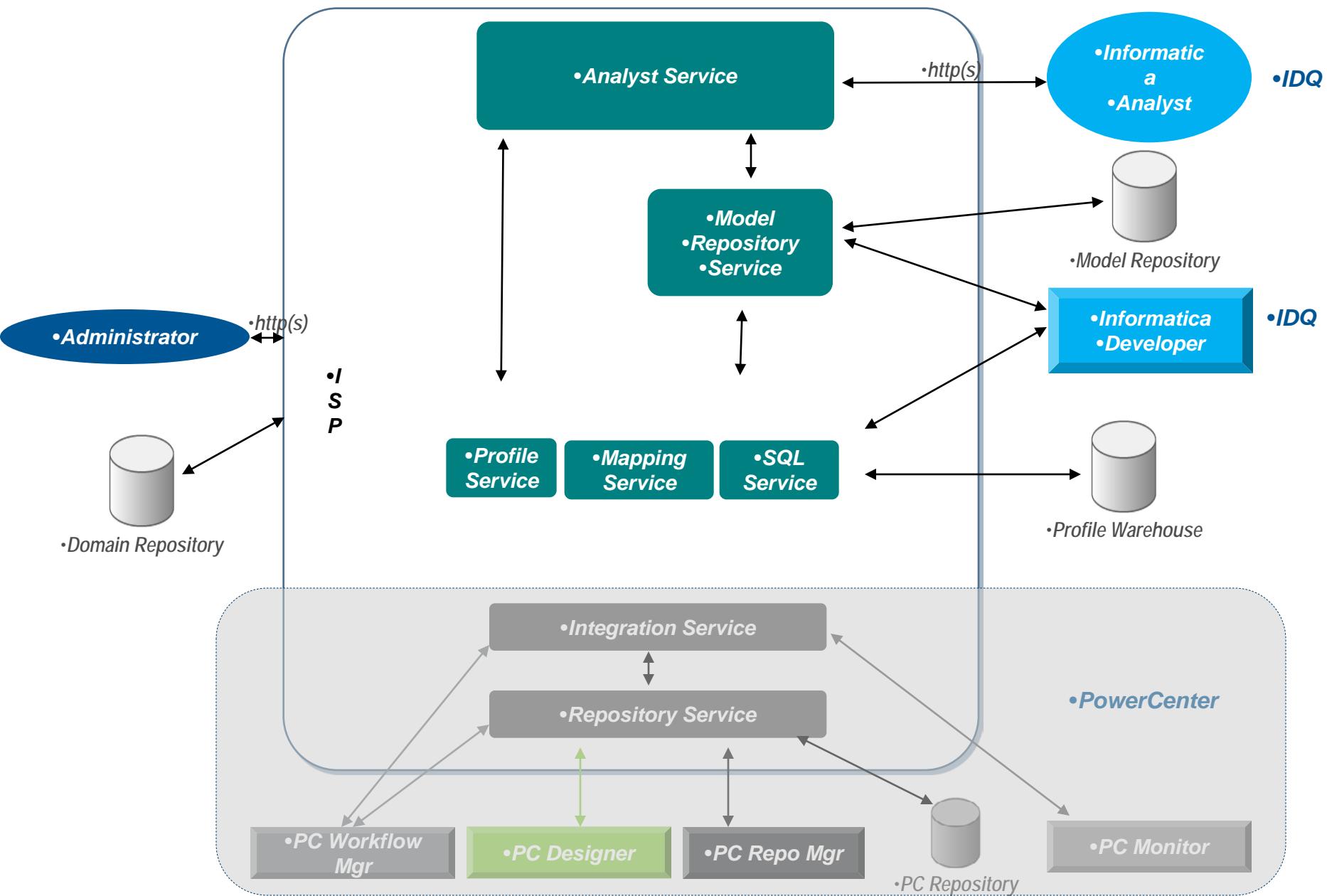
- **8.6.2**
 - One repository per user
 - Reference data on the local file system
 - Data quality metadata contained in IDQ Plan
 - Connection details embedded within IDQ Plan
- **9.0.1**
 - Central repository shared by all users
 - Reference data in the Reference Table Manager
 - Data Quality metadata in 9.0.1 models
 - Connection details stored centrally

Domain

Informatica 9 Architecture for IDQ

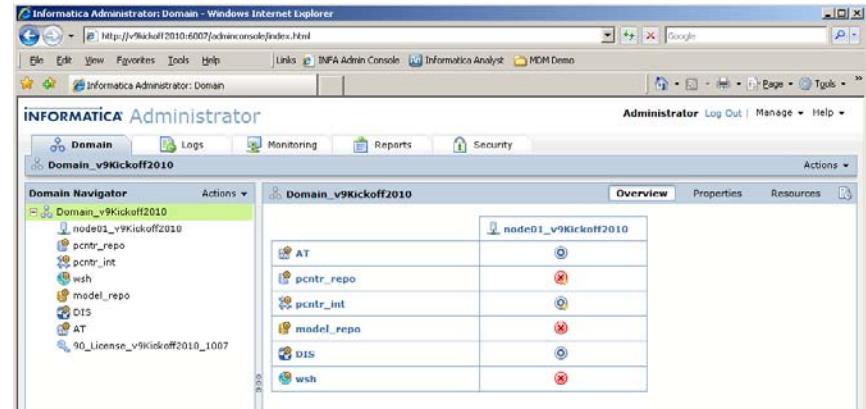


Informatica 9 Architecture for IDQ & PC



Informatica Domain

- The Informatica domain include objects and services for the Informatica platform.
- The Admin console is now known as Administrator
- The Informatica domain includes services for PowerExchange, Informatica Analyst, and Informatica Developer.



Informatica Domain

- **IDQ Migration**
 - Direct migration from 8.6.2 to 9.0.1
 - Direct upgrade from 9.0 to 9.0.1
 - To migrate Pre-8.6.2 installations you must first upgrade to IDQ8.6.2, then migrate to 9.0.1
- **Security**
 - Informatica 9 platform provides full READ, WRITE, EXECUTE and GRANT permissions for domain connection objects.
 - Support for MS SQL Server Trusted connection for hosting the domain repository (MRS)
 - Ability to set and enforce permissions for all services and folders in the domain.

New Services

- **Analyst Service**
 - Application service that runs **Informatica Analyst** in the **Informatica domain**. Create and enable an **Analyst Service** on the **Domain tab** of **Informatica Administrator**. When you enable the **Analyst Service**, the **Service Manager** starts **Informatica Analyst**. You can open **Informatica Analyst** from **Informatica Administrator**.
- **Model Repository Service**
 - Application service that manages the **Model repository**. The **Model repository** is a relational database that stores the metadata for projects created in **Informatica Analyst** and **Informatica Designer**. The **Model repository** also stores run-time and configuration information for applications deployed to a **Data Integration Service**. Create and enable a **Model Repository Service** on the **Domain tab** of **Informatica Administrator**.

Migrating the Repository and Dictionaries

Steps for Migration

- 1. *ClientPackage* - On the IDQ 8.6.2 client – single step process to:**
 - Export IDQ plans from IDQ repository
 - Identify connection details
 - Gather local dictionaries
 - Package data for the next step
- 2. *ServerImport* - On the 9.0.1 Server – single process to:**
 - Unpack data from ClientPackage
 - Create connections
 - Import dictionary data into Reference Table Manager
 - Convert Plans to 9.0.1 mapping XML
- 3. *XML Import* - On 9.0.1 Client**
 - Import mapping XML from ServerImport into 9.0.1 repository via Developer

ClientPackage – Overview

- **Export IDQ plans from IDQ repository**
- **Identify connection details**
- **Gather local dictionaries**
- **Package data for the next step - ServerImport**

ClientPackage - Report

- Default Location:
 - <MigrationPackageLocation>/Package/PackageReport.html

The screenshot shows a report titled "Migration Report" created on 2010-05-10 17:10:21. The report includes sections for "Migration Summary", "Migration Details", "Used dictionaries", and "Unused dictionaries". Red callout boxes highlight specific information: "Identify Dictionaries used by plans and dictionaries that exist but are not used by any plan" points to the "Unused dictionaries" section, and "Database Connections used by plans. One entry for every DSN/Username/Password combination" points to the "Migration Details" section.

Migration Report
Created on 2010-05-10 17:10:21

INFORMATICA®

Migration Summary

Dictionary information:

- [Used dictionaries \(10 dictionaries\)](#)
- [Unused dictionaries \(29 dictionaries\)](#)

[Plan Information \(2 plans\)](#)

[Required Database Connection List \(3 connections\)](#)

Migration Details

Used dictionaries

The following dictionaries are used in the migrated plans and have been packaged.

C:/Program Files/Informatica Data Quality/Dictionary/Vertical_Solutions/Dimensions/DIM_BusinessUnit.dic
C:/Program Files/Informatica Data Quality/Dictionary/Vertical_Solutions/Dimensions/DIM_DQ_Dimension.dic
C:/Program Files/Informatica Data Quality/Dictionary/Vertical_Solutions/Dimensions/DIM_Field_Class.dic
C:/Program Files/Informatica Data Quality/Dictionary/Vertical_Solutions/Dimensions/DIM_JobTracking.dic
C:/Program Files/Informatica Data Quality/Dictionary/Vertical_Solutions/Dimensions/DIM_JobTracking_Date.dic
C:/Program Files/Informatica Data Quality/Dictionary/Vertical_Solutions/General/Date/day_month.dic
C:/Program Files/Informatica Data Quality/Dictionary/Vertical_Solutions/General/Date/db_default_dates.dic
C:/Program Files/Informatica Data Quality/Dictionary/Vertical_Solutions/General/Date/month_numbers.dic
C:/Program Files/Informatica Data Quality/Dictionary/Vertical_Solutions/General/Date/year_range.dic
C:/Program Files/Informatica Data Quality/Dictionary/Vertical_Solutions/General/Rule/pass_fail.dic

[^top](#)

Unused dictionaries

ServerImport – Overview

- **Unpack data from ClientPackage**
- **Create connections**
- **Import dictionary data into Reference Table Manager**
- **Convert 8.6.2 Plans to 9.0.1 Mapping XML**

Steps to perform before ServerImport

- Create new blank project for mappings to be imported to
- Create new folder for imported reference tables
- Install Informatica Content packages in shared project

ServerImport – Summary / Overview Report

- Overall status of conversion
- Links to detail / individual reports
- Default location
 - <MigrationPackageLocation>/migration_reports

The screenshot shows a software interface titled "Migration Report" created on 2010-05-13 09:15:50. The top right corner features the INFORMATICA logo. The interface is divided into several sections:

- Migration Summary**:
 - Plan Information (2 plans)
 - Errors encountered: 8
 - Warnings encountered: 7
- Migration Details**:
 - Plans added to this package**:

Path	Comment
Check_ValuationDate.xml Report File	Error: Rule Based Analyser ('Rule Based Analyzer'): The component has no inputs. The generated mapping will not be valid as transformations without inputs are not allowed. Please use Informatica Designer to review the mapping and add an appropriate input to the decision transformation

ServerImport – Detail Reports

- One Detail report per 8.6.2 plan/9.0.1 mapping
- Component / Port level detail
- Includes warnings / errors
- Default location
 - <MigrationPackageLocation>/migration_reports

**Migration report for IDQ plan
'Check_ValuationDate.xml'**

Created on 2010-05-13 09:26:07

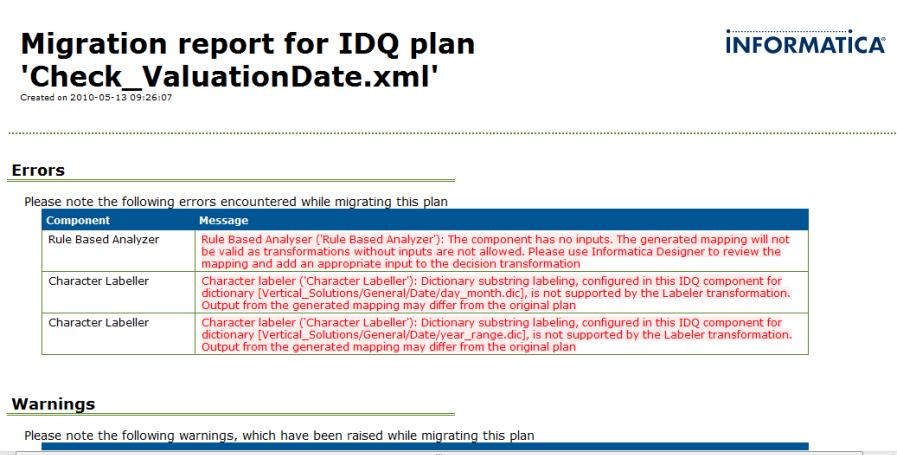
Errors

Please note the following errors encountered while migrating this plan

Component	Message
Rule Based Analyzer	Rule Based Analyzer ('Rule Based Analyzer'): The component has no inputs. The generated mapping will not be valid as transformations without inputs are not allowed. Please use Informatica Designer to review the mapping and add an appropriate input to the decision transformation.
Character Labeller	Character Labeller ('Character Labeller'): Dictionary substring labeling, configured in this IDQ component for dictionary [Vertical_Solutions/General/Date/day_month_dc], is not supported by the Labeler transformation. Output from the generated mapping may differ from the original plan.
Character Labeller	Character Labeller ('Character Labeller'): Dictionary substring labeling, configured in this IDQ component for dictionary [Vertical_Solutions/General/Date/year_range_dc], is not supported by the Labeler transformation. Output from the generated mapping may differ from the original plan.

Warnings

Please note the following warnings, which have been raised while migrating this plan



Client XML Import – Overview

- Import mapping XML generated through **ServerImport** into 9.0.1 repository
 - Through Informatica Designer
 - Through infacmd
- Default location for XML file:
 - <MigrationPackageLocation>/Output/MigratedMappings.xml

XML Import via Developer

The image displays two overlapping windows from the Informatica Developer interface, illustrating the XML Import process.

Left Window: Import File

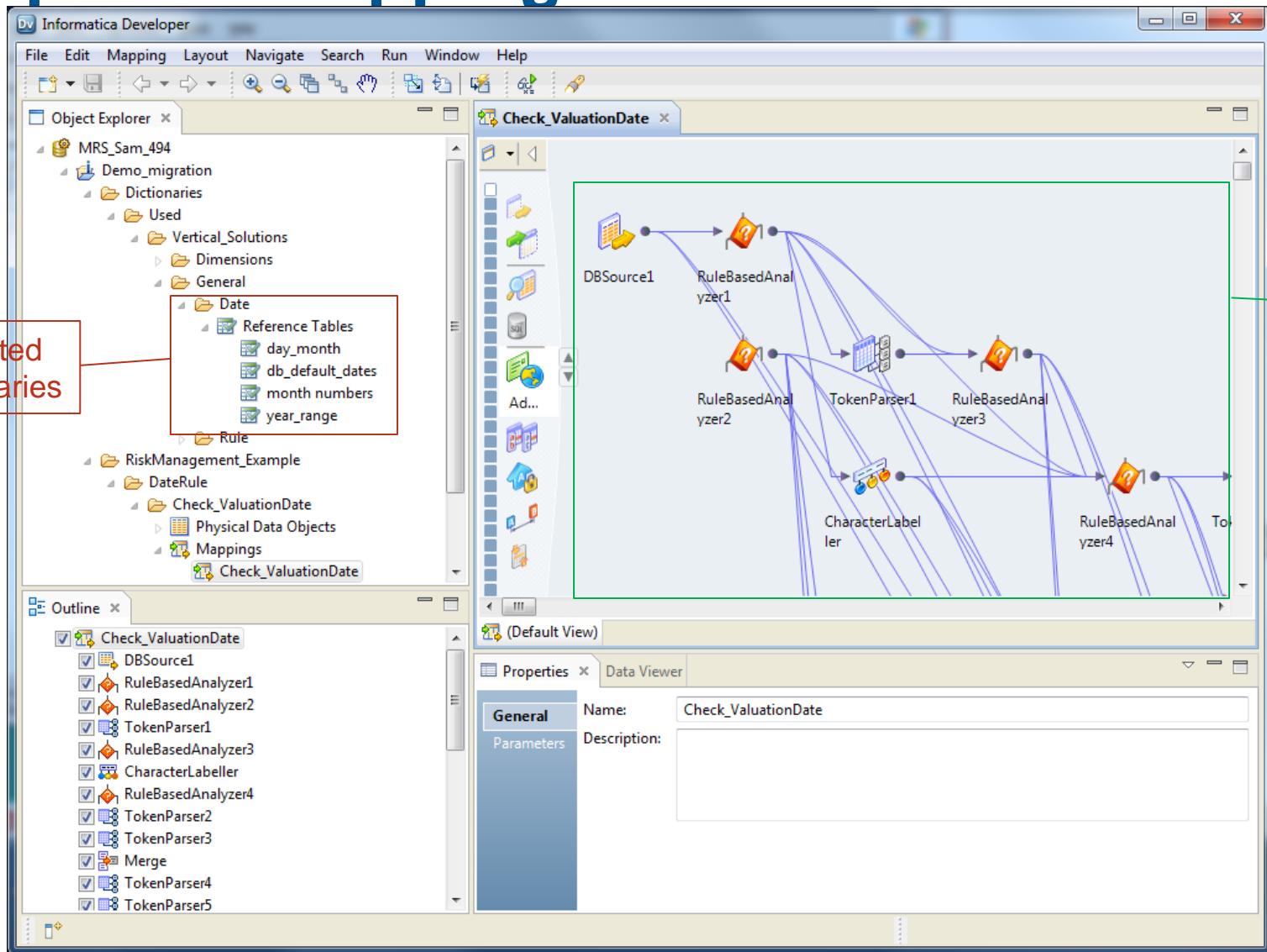
- File name:** T:\4Chris\migration_reports\MigratedMappings.xml
- Project:** Demo_migration
- Location:** MRS_Sam_494\Demo_migration
- Import settings**
 - If duplicates are found:
 - Replace existing object with imported object
 - Rename object before importing
- Import Reference Data settings**
 - Import Reference Data
 - Reference data location: [empty field]
 - Code Page: UTF-8 encoding of Unicode
 - If Reference Data duplicate Found:
 - Replace existing Reference Data
 - Rename Reference Data Table
 - Skip importing Reference Data if it already exists

Right Window: Required Objects

The following objects from the import file are required for this operation.

- Migration
 - Demo_migration
 - Dictionaries
 - Used
 - RiskManagement_Example
 - DateRule
 - Check_ValuationDate
 - Check_ValuationDate
 - RDS_DBSource1
 - RDS_DBTarget1
 - RDS_DBTarget2
 - RatingReviewDate
 - R083_084_122_Obligor_Rating_Review_Date
 - R083_084_122_Obligor_Rating_Review_Date
 - RDS_DBSink1
 - RDS_DBSink2
 - RDS_DBSink3
 - RDS_DBSource1

Imported Mappings



Tips and Tricks - General

- **Migration packages required Java 1.6 or later to be installed**
 - e.g. C:\Informatica\9.0.1\Java\bin
- **Zip files generated by ClientPackage are not editable in WinZip (or similar)**
- **On a 64-bit client, manual export is required due to Java version incompatibility with IDQ 8.6.2 32-bit libraries**
- **Dictionaries from previous All World package are not automatically recognized as Informatica dictionaries.**

Post-Migration Notes

- **Incompatible Components may require editing the Plan in 9.0.1**
- **Address Validation components will require editing in 9.0.1**
 - e.g. QAS and Melissa have been replaced with Address Doctor
- **IDQ8.6.2 Connections that source or target MySQL will have to be edited by hand**

Logging and Logs in IDQ v9

Logs

- The purpose is to identify the logs populated by Informatica 9 IDQ (Informatica Data Quality). What logs exist, where they are located and what are their main purpose. Armed with this information, the user will be able to quickly identify issues during the installation process and with day to day operation. Also, the user will also be able to identify areas requiring periodic maintenance (i.e. Log removal).

Installation Logs

- **Server, Client and Content installation logs are located mostly in the root installation directory.**
- **On windows, the default is C:\informatica\9.0.1. For the rest of the document, it will be referred to by <Informatica install dir>.**
- **There are two logs for each installation. One shows the commands executed and the other shows the output of the installation.**
- **For debugging purposes, you will need to look at the InstallLog files.**

Installation Logs: Client, Server and Content

All these look the same – look for the Summary Information

Summary

Installation: Successful.

18 Successes

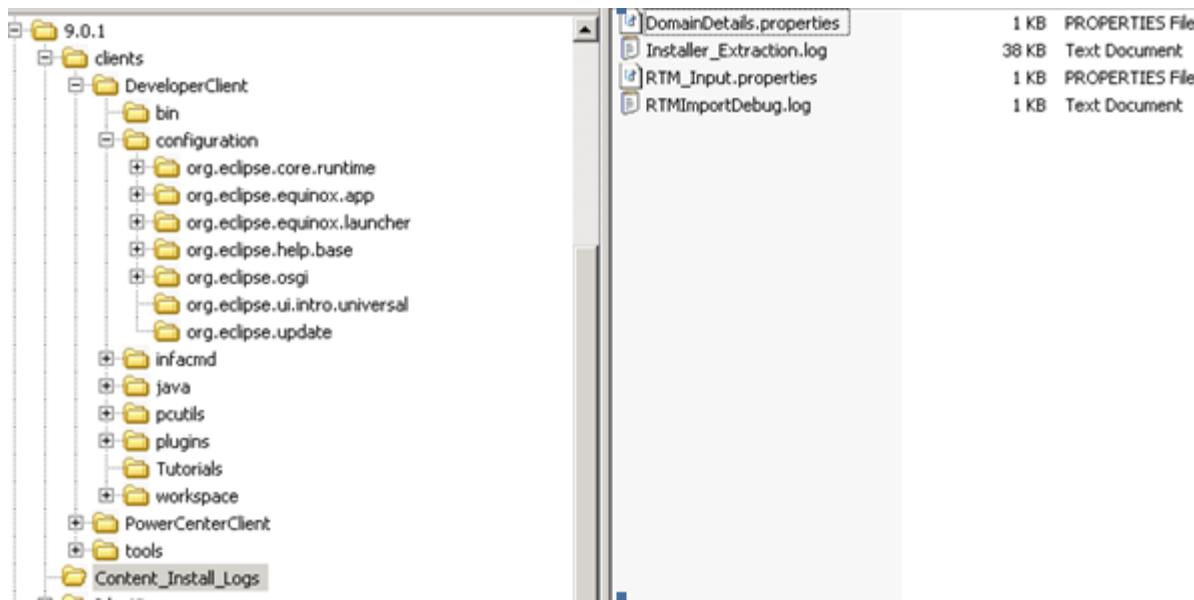
0 Warnings

0 NonFatalErrors

0 FatalErrors

Additional Content Installation Logs

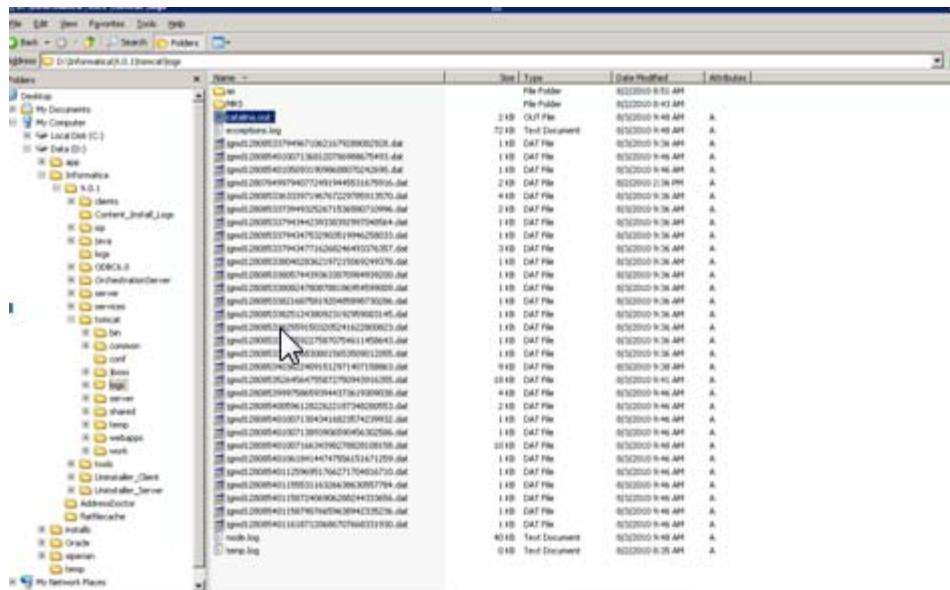
There are also content installation log files located at <Informatica install dir>\Content_Install_Logs



Day to Day Operations

Initial errors when starting up

- When initially starting up the services and they don't start, look here: <Informatica install dir>\tomcat\logs
 - There are two logs of interest. The exceptions.log and catalina.out.



Day to Day Operations

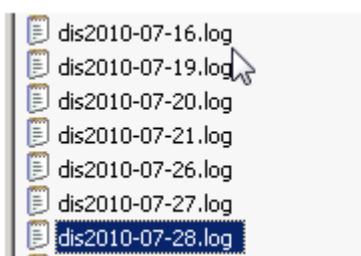
Catalina.out and Exceptions.log

- While the services are up and running, these file are locked.
- Catalina.out has messages about the errors found when the domain starts
- Exceptions.log has messages referring to what happens after the domain has come up such as the status of gateway elections and it is found at

`<Informatica install dir>\tomcat\logs`

Day to Day Operations - Analyst

- When creating a physical object, the Analyst tool uses the data integration service. As it performs the task, it adds entries to the Data Integration Service (DIS) Logs located at:
<Informatica install dir>\tomcat\temp\DIS\logs
- The logs are dated



dis2010-07-16.log	17 KB	Text Document	7/16/2010 1:39 PM	A
dis2010-07-19.log	772 KB	Text Document	7/19/2010 8:46 AM	A
dis2010-07-20.log	17 KB	Text Document	7/20/2010 9:27 AM	A
dis2010-07-21.log	33 KB	Text Document	7/21/2010 11:13 AM	A
dis2010-07-26.log	18 KB	Text Document	7/26/2010 2:38 PM	A
dis2010-07-27.log	17 KB	Text Document	7/28/2010 9:40 AM	A
dis2010-07-28.log	22 KB	Text Document	7/28/2010 10:29 AM	A

Day to Day Operations - Analyst

- Keep this area in mind, because this is one of the areas that will eventually need to be cleaned up.
- The Analyst Tool log (analyst.log) can be found at <Informatica install dir>\tomcat\logs\as

Day to Day Operations – Profiling Logs

- There are two logs created for each profiling job in <Informatica install dir>\tomcat\bin\disLogs\profiling.
- There is a summary log, which just tells you the mappings were complete and the details such as what tables were updated in the profiling warehouse but not a lot of details about the profile itself.
- Live drill down and Export of profiling results will also create log files here.

Day to Day Operations – Profiling Logs

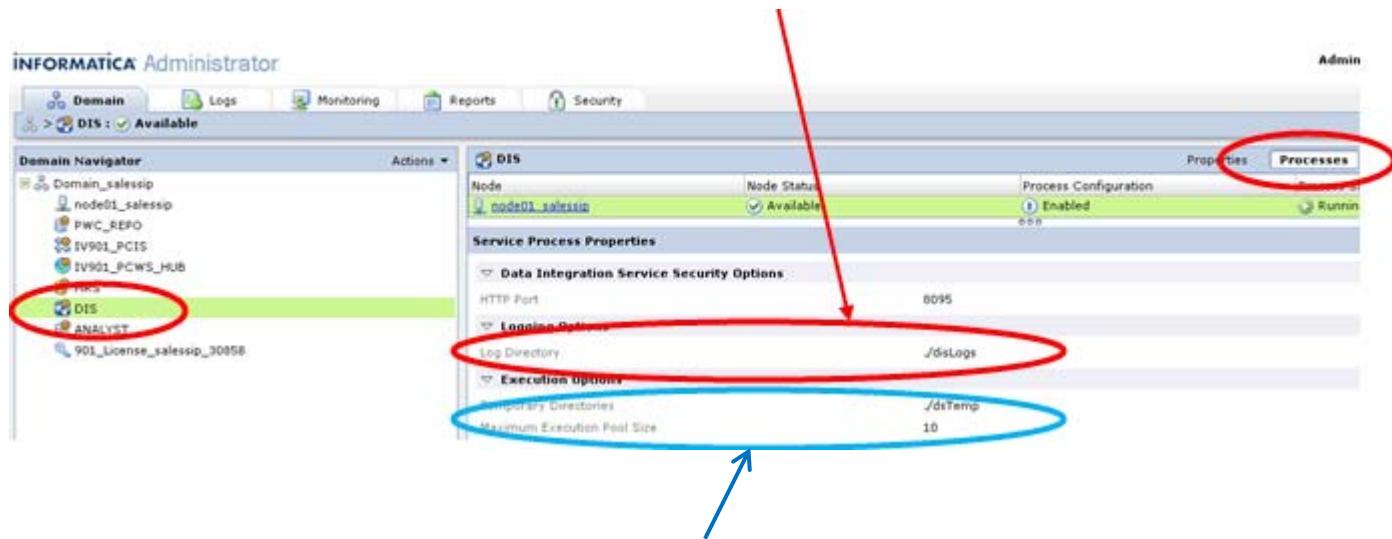
- These logs can and should be moved to a location that is more accessible by the general community.
- Usually, a directory that has software install is usually inaccessible by the general user community. A more logical place than

`<Informatica install dir>\tomcat\bin\disLogs`

would help people find them

Day to Day Operations – Profiling Logs

- The location can be configured in the admin console:



- The temp logs can also be configured somewhere else.

Day to Day Operations – Profiling Logs

- When you do mid-stream profiling, it creates a log in this directory but they are not accessible from the client tool. This is true for any profiling operation (from Dxt -Designer/AT Analyst Tool).

```
rule_USA_Phone_Number_Standardization_6640643544230.log C:\Informatica\9.0.1\tomcat\bin\disLogs\profiling  
rule_USA_Phone_Number_Standardization_Summary_6640646... C:\Informatica\9.0.1\tomcat\bin\disLogs\profiling
```

Day to Day Operations – MRS Logs

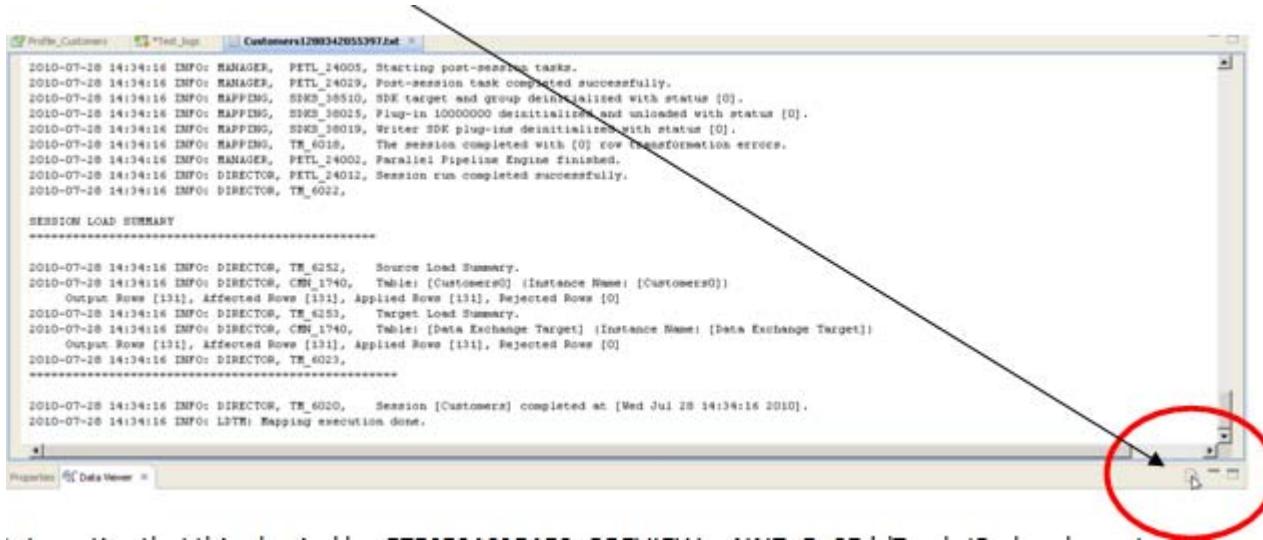
- When the service is initially brought up, an MRS log is started at <Informatica install dir>\tomcat\logs\MRS.
- Also, when you connect to the MRS with the client, its attempt and success is recorded here.
- While the services are up, this file is locked.

Day to Day Operations – Mapping Service Logs

- The mapping service logs are a little more helpful when looking for errors in a mapping. (remember profiling is done by a mapping).
- Among other things, It can confirm that the file was read without errors.
- They can be found at <Informatica install dir>\tomcat\bin\disLogs\ms.
- This is another area that will need occasional maintenance.

Day to Day Operations – Mapping Service Logs

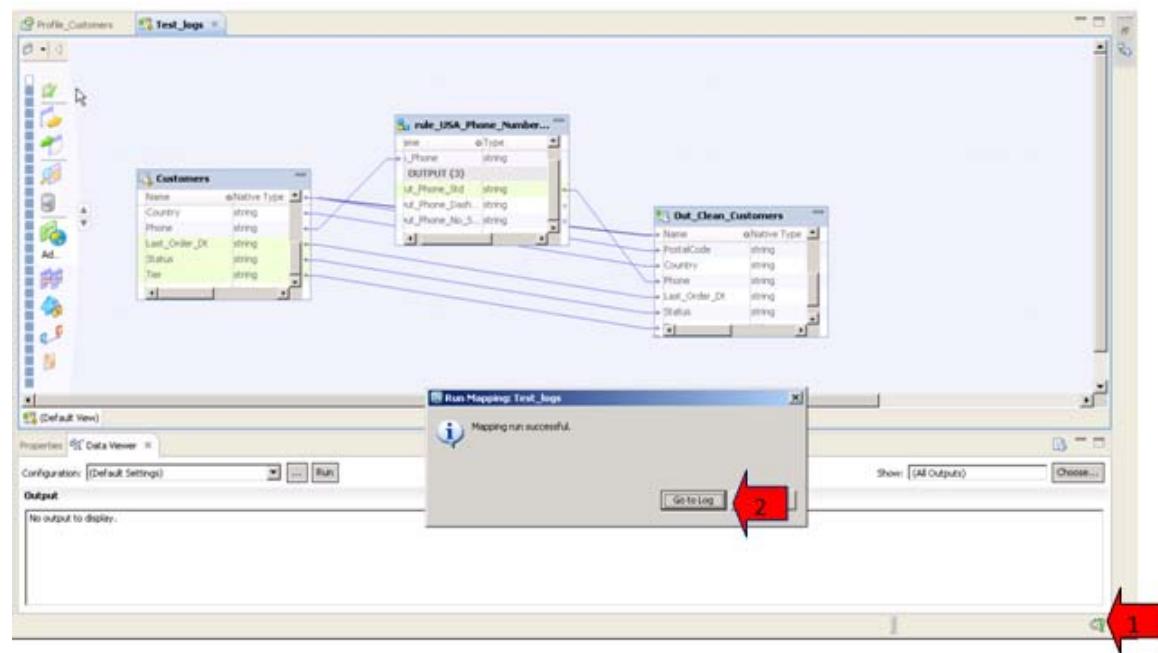
- Anything you do in the client with regards to a mapping will update these logs. They are also assessable from the client. A simple run data viewer produced this log and was accessed via the client by double clicking on the show logs icon.



```
Profile_Customers *Test_Logs Customers12003420553971.edt
2010-07-28 14:34:16 INFO: MANAGER, PETL_24005, Starting post-session tasks.
2010-07-28 14:34:16 INFO: MANAGER, PETL_24019, Post-session task completed successfully.
2010-07-28 14:34:16 INFO: MAPPING, SMNC_38510, SNE target and group deinitialized with status [0].
2010-07-28 14:34:16 INFO: MAPPING, SMNC_38025, Plug-in 10000000 deinitialized and unloaded with status [0].
2010-07-28 14:34:16 INFO: MAPPING, SMNC_38019, Writer SDK plug-in deinitialized with status [0].
2010-07-28 14:34:16 INFO: MAPPING, TM_6018, The session completed with [0] row transformation errors.
2010-07-28 14:34:16 INFO: DIRECTOR, PETL_24002, Parallel Pipeline Engine finished.
2010-07-28 14:34:16 INFO: DIRECTOR, PETL_24012, Session run completed successfully.
2010-07-28 14:34:16 INFO: DIRECTOR, TM_6022,
SESSION LOAD SUMMARY
*****
2010-07-28 14:34:16 INFO: DIRECTOR, TM_6252, Source Load Summary.
2010-07-28 14:34:16 INFO: DIRECTOR, CMH_1740, Table: [Customers] (Instance Name: [Customer0])
Output Rows [131], Affected Rows [131], Applied Rows [131], Rejected Rows [0]
2010-07-28 14:34:16 INFO: DIRECTOR, TM_6253, Target Load Summary.
2010-07-28 14:34:16 INFO: DIRECTOR, CMH_1740, Table: [Data Exchange Target] (Instance Name: [Data Exchange Target])
Output Rows [131], Affected Rows [131], Applied Rows [131], Rejected Rows [0]
2010-07-28 14:34:16 INFO: DIRECTOR, TM_6023,
*****
2010-07-28 14:34:16 INFO: DIRECTOR, TM_6020, Session [Customers] completed at [Wed Jul 28 14:34:16 2010].
2010-07-28 14:34:16 INFO: LPTM: Mapping execution done.
```

Day to Day Operations – Mapping Service Logs

- When you run a mapping, you can view the logs by clicking here



Once you view the log and close it, it is no longer accessible via the client. You would need to go to the <Informatica install dir>\tomcat\bin\disLogs\ms directory and view it there.

Day to Day Operations – Other Logs

- **Reference Table Manager Command Line Interface (CLI) logs:** The Reference Table Manger CLI logs can be found at <Informatica install dir>\server\bin\rtm_cli_logs. They are generated when the reference tables are imported.
- **Import / Export logs:** You can find some import/export logs at the same location: <Informatica install dir>\clients\DeveloperClient\infacmd\rtm_cli_logs

ESG – Additional Training

- Our Classes are available:
 - On-Site – at your company location
 - Virtual Academy – on-line including conference calling
 - Public classes – at our training site throughout the world
- IDQ 9.0.1 – 4 days
- IDQ Migration – 1 day
- List of classes and dates available are at:
 - www.informatica.com
 - Products & Services tab