

What's New in SAS® Data Integration Studio 4.3

* Supporting information for capability shown is provided in notes



THE
POWER
TO KNOW®

SAS Data Integration Studio 4.3 will be available with SAS 9.3



Data Management



SAS DI Studio 4.3 focus areas

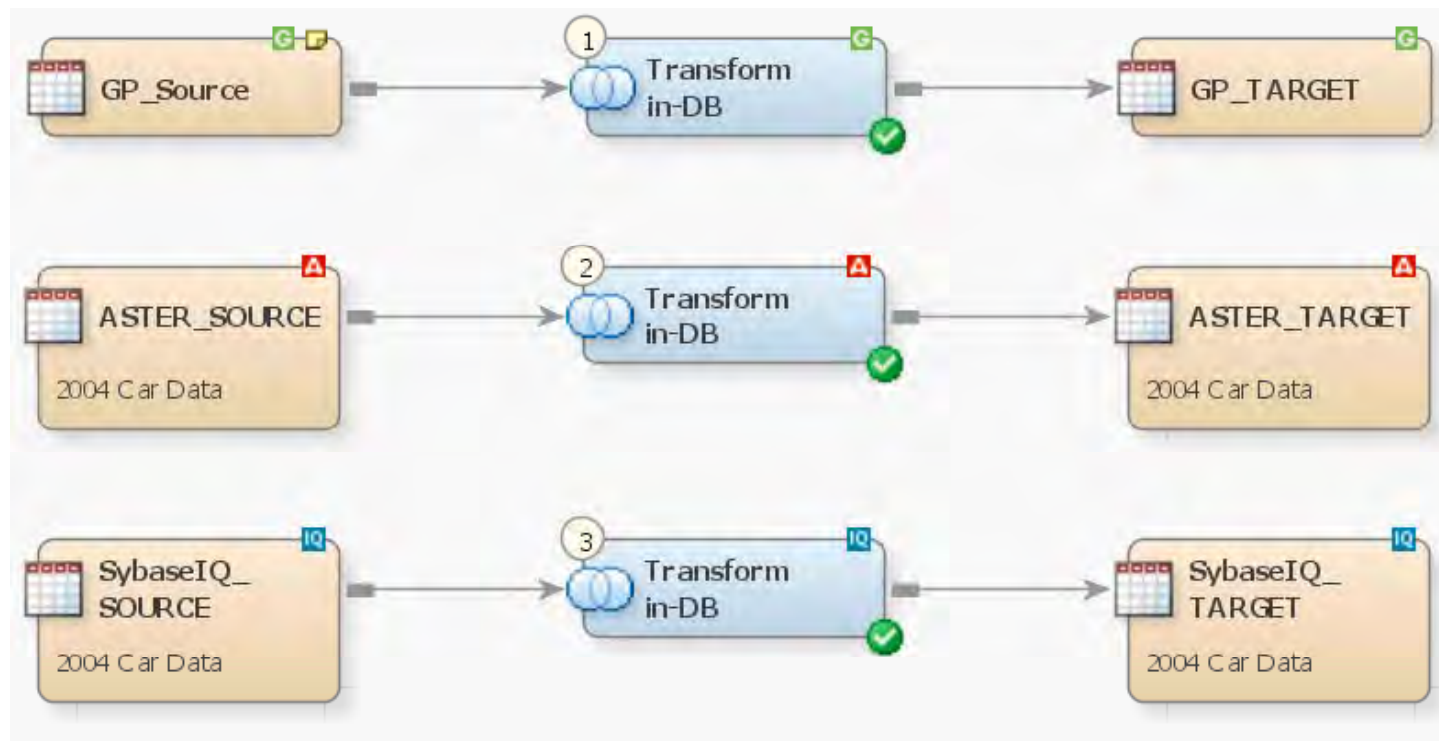
- Governance and Methodology
 - Archiving and Versioning, batch promotion, etc.
- Integration Ubiquity
 - New databases, ...
- Scalability
 - More native ELT, SQL Pushdown, and Grid
- Capability Excellence
 - New transformations
- Unification and Uniformity
 - Innovations in data quality, metadata, and MDM



What's new for users who develop and test SAS Data Management Jobs

Expanded support for MPP databases

- New library templates enable enhanced SQL passthru to support in database SQL processing in:
 - AsterData nCluster
 - Sybase IQ
 - Greenplum

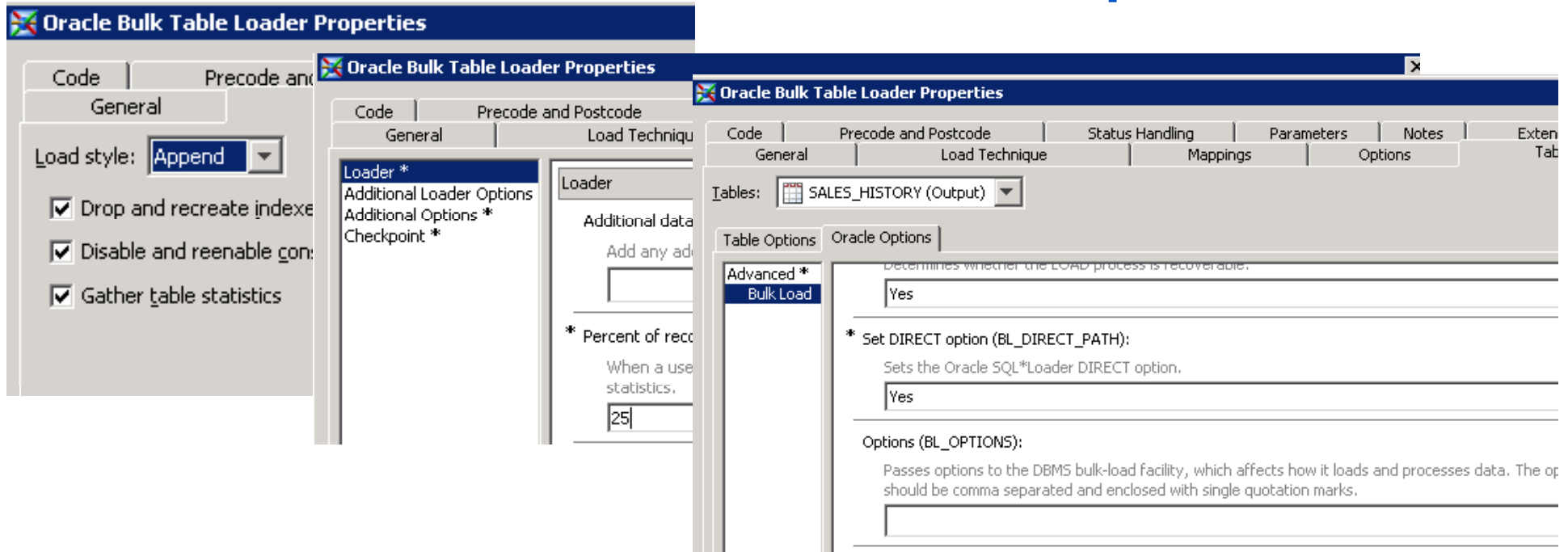


New: Oracle Bulk Table Loader

- Database specific loader that exposes Oracle specific load functions:
 - Direct Path Load
 - Load to Partition
 - Oracle Hints
 - Insert/Update Buffers
- Maintain Indexes, Constraints and Statistics
 - Rebuilds indexes in the proper table space
 - Disable constraints for loading and the re-enable post load
 - Gather statistics based on user defined sample size

Outdated statistics are a leading cause of performance problems in an RDBMS reporting application!

Oracle Bulk Table Loader- Example

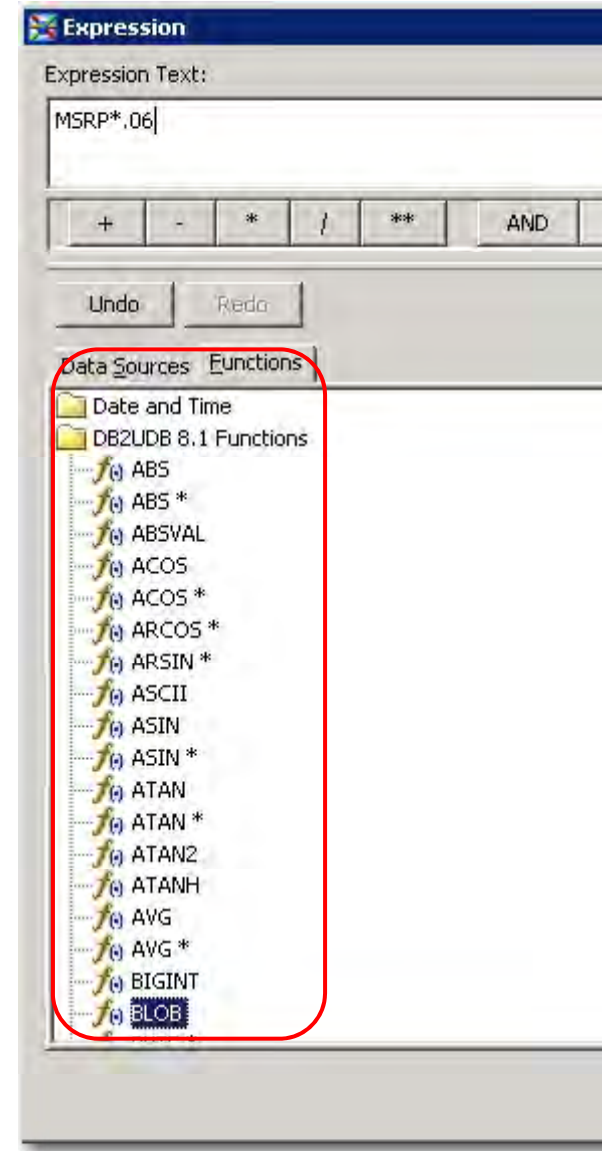


- This example shows bulk load of an Oracle table
 - Append Load Style to incrementally add to the table
 - Index rebuild, constraint management and table statistic gather all enabled
 - » Statistics gather sample set to 25%
 - Direct Path enabled

New: Functions added to Expression builder

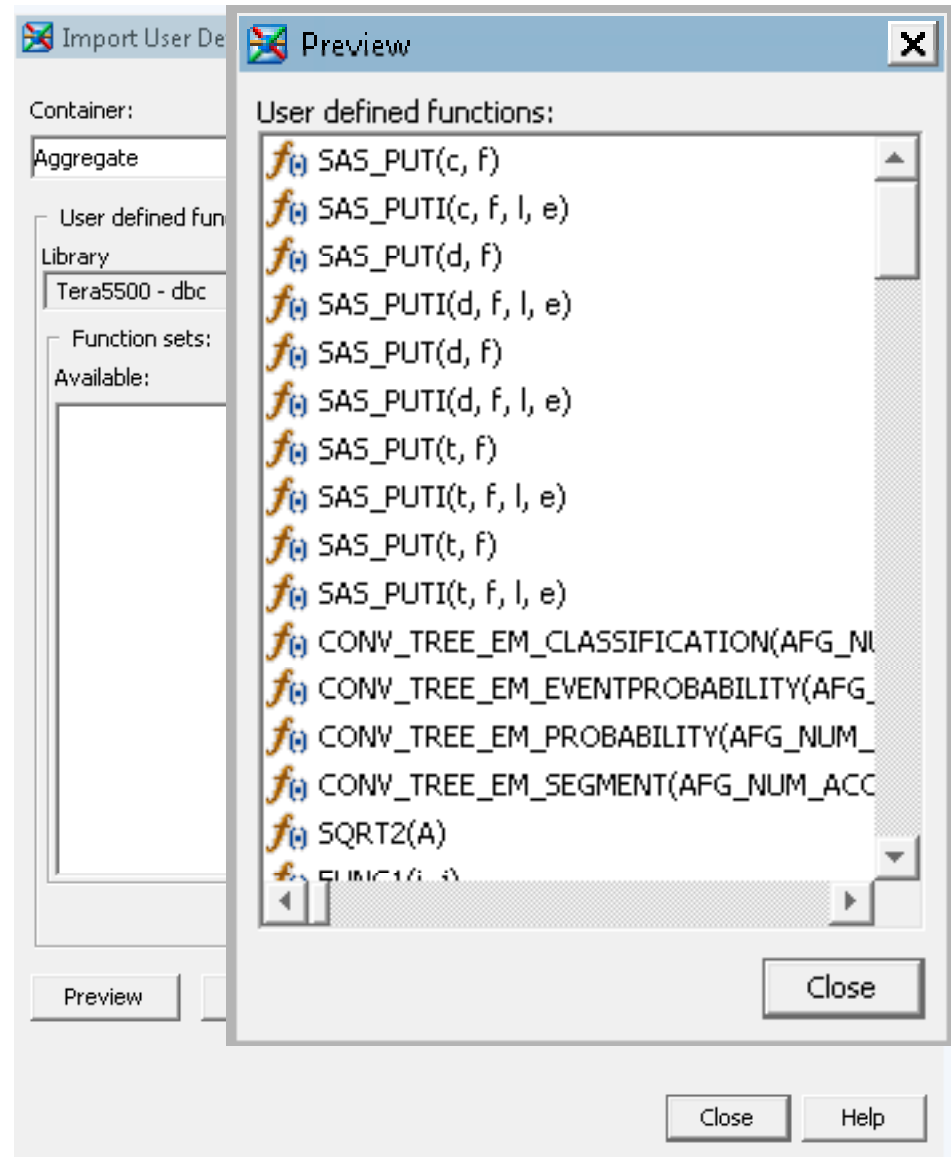
- Teradata 13
- Oracle
- DB2UDB 8.1
- SQL Server
- MySQL
- ODBC
- + User Defined Functions

- '*' Indicates which functions will pushdown with implicit SQL



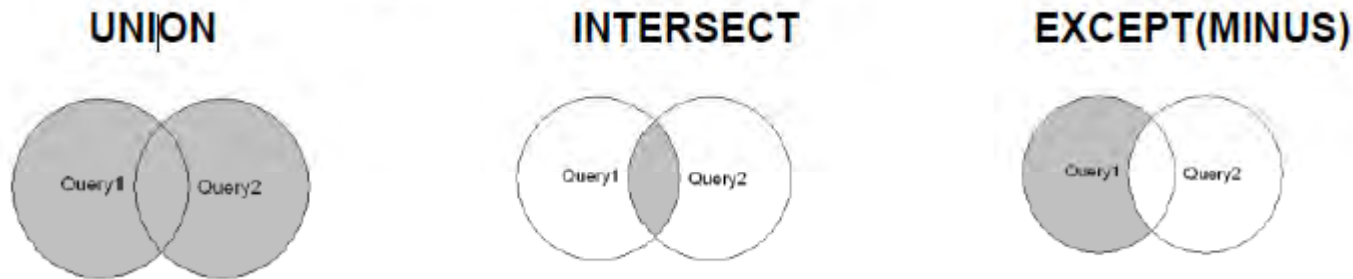
New: Import DBMS User Defined Functions

- Import and use in the Expression builder
- Import any DBMS and Enterprise Miner Analytic UDFs
- Parameter discovery and help
- Expression builder single click mapping for column inserts



New SQL Set Transformation

- DI Studio now supports 3 SQL Set Operators



- Runtime Efficiency
 - SQL Set operators are generally much more efficient than Joins for shared use cases
 - It's possible to accomplish the same thing with complex correlated queries, but such queries are poorly optimized by database query planners (*particularly NOT IN for EXCEPT*)
- Understandability
 - Select the two populations of interest with simple SELECT statements and then specify operators – no complex correlated subqueries

New: SQL Set operator transform

SQL Set Operators Properties

General | Set Operators | Options | Table Options | Code | Precode and Postcode | Status Handling | Parameters | No

Queries:

- Add...
- CONTINENTS_NONAMERICAS (CONTINENTS)
- Intersect**
- CONTINENTS_AMERICAS (CONTINENTS)
- Union
- CONTINENTS2 (CONTINENTS2)
- Union
- CONTINENTS_SPECIAL_CHARS

Set operator:

Set operator type: Intersect

Description: Returns rows that are common from both the two queries

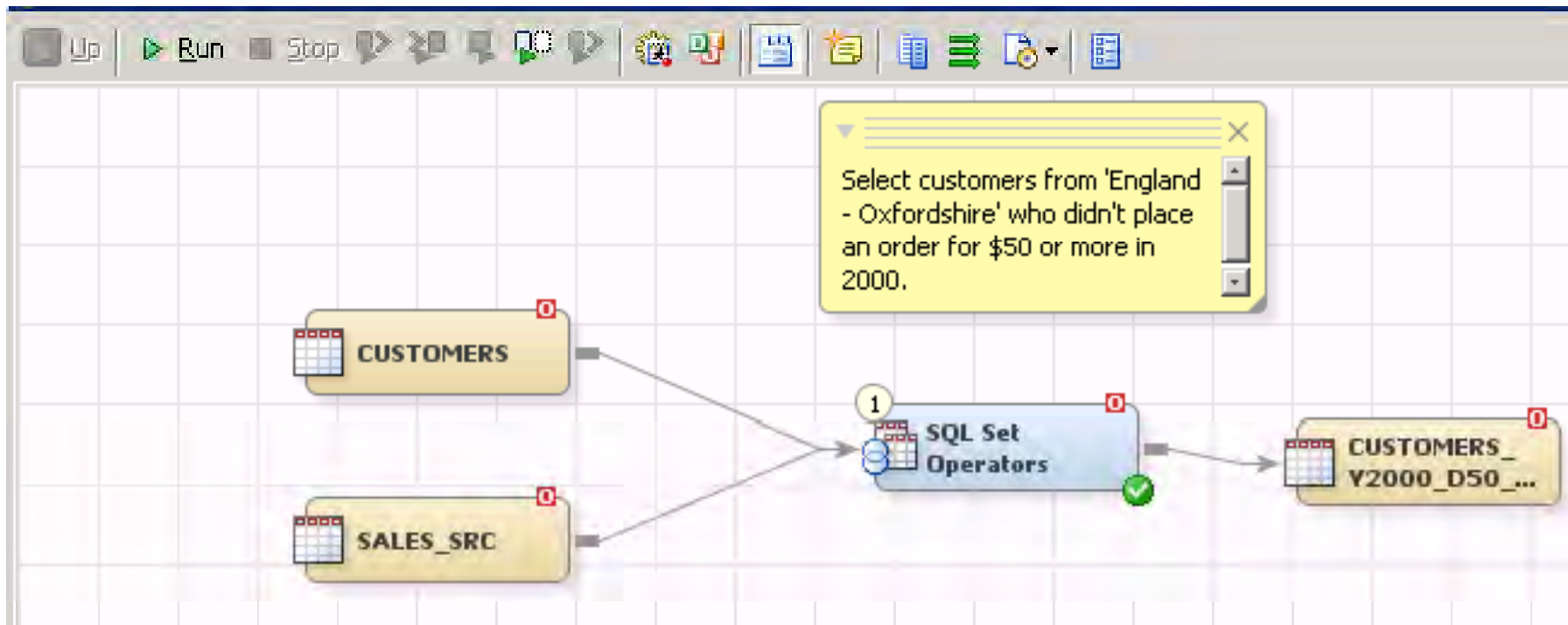
- Include duplicates(ALL)
- Match columns by name(CORRESPONDING)
- Include parenthesis around the clause

The SQL Set Operator transform will generate SQL statements, that will combine the results of two or more queries

New SQL Set Transformation - Example

“Select customers from “England – Oxfordshire who didn’t place an order \geq \$50 in 2000”

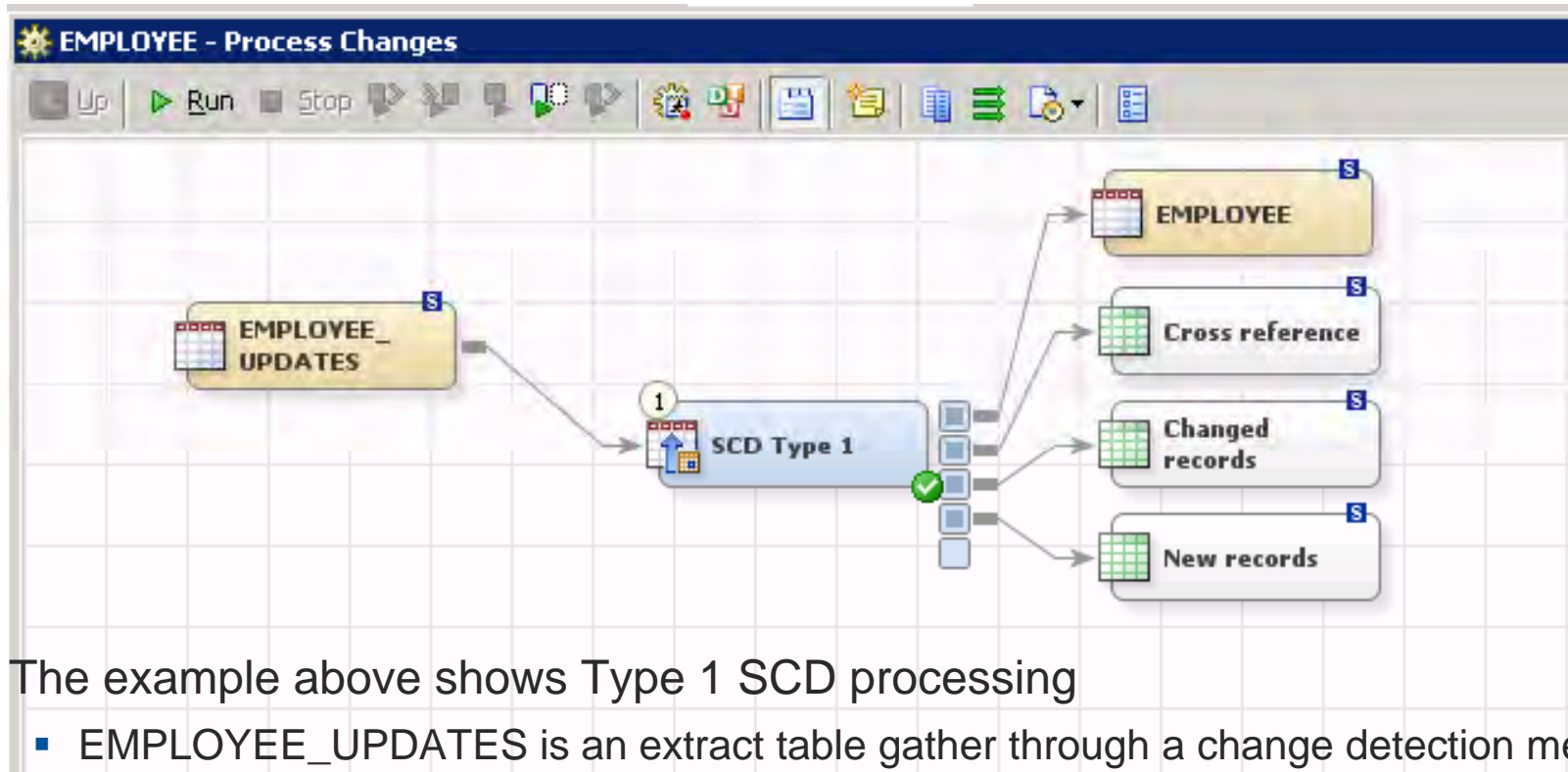
- “England – Oxfordshire” customers selected by first query
- Transactions for 2000 for \geq \$50 selected by second query
- EXCEPT operator used to produce final result



New SCD Type 1 Loader

- Loader for tables where all changes are handled as overwrites (Type 1 SCD)
- Logic:
 - On Incoming, Not on Master:
 - » Action = Insert Rows to Master Table
 - On Incoming, On Master, Changed Columns:
 - » Action = Update Matching Row In Master Table
 - On Incoming, On Master, No Changed Columns
 - » Action = Ignore Row
- Performance
 - Join between master and changes table as Hash lookup or MERGE
- Additional Functionality
 - Generates new surrogate keys
 - Populates original load and last update dates
 - Outputs Cross Reference table that contains Business Key, Surrogate Key and change digest, which is useful for fact table foreign key processing

New SCD Type 1 Loader- Example

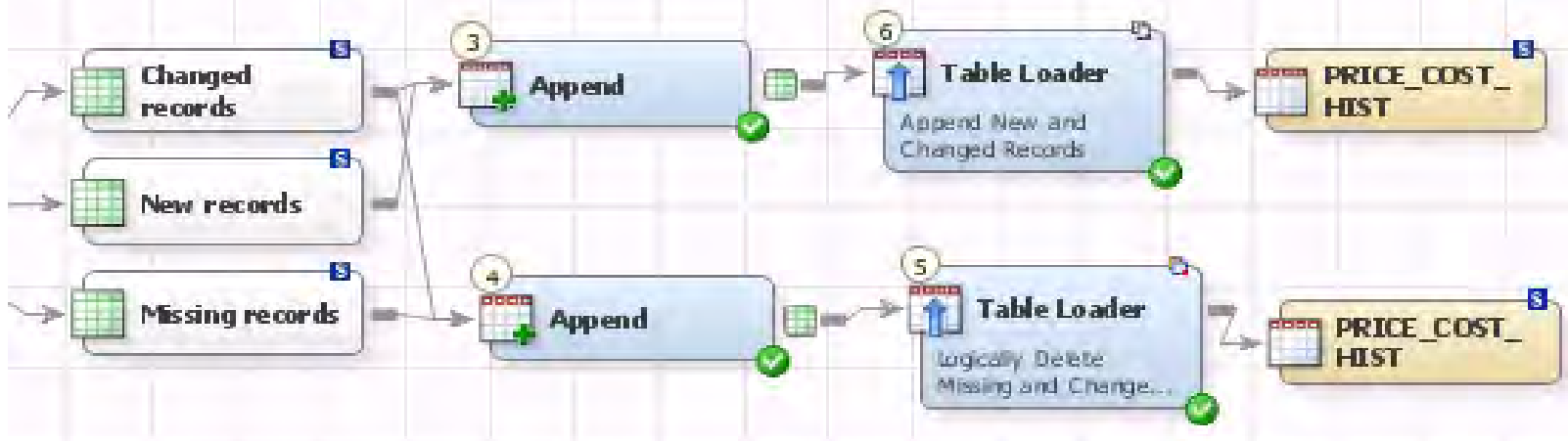
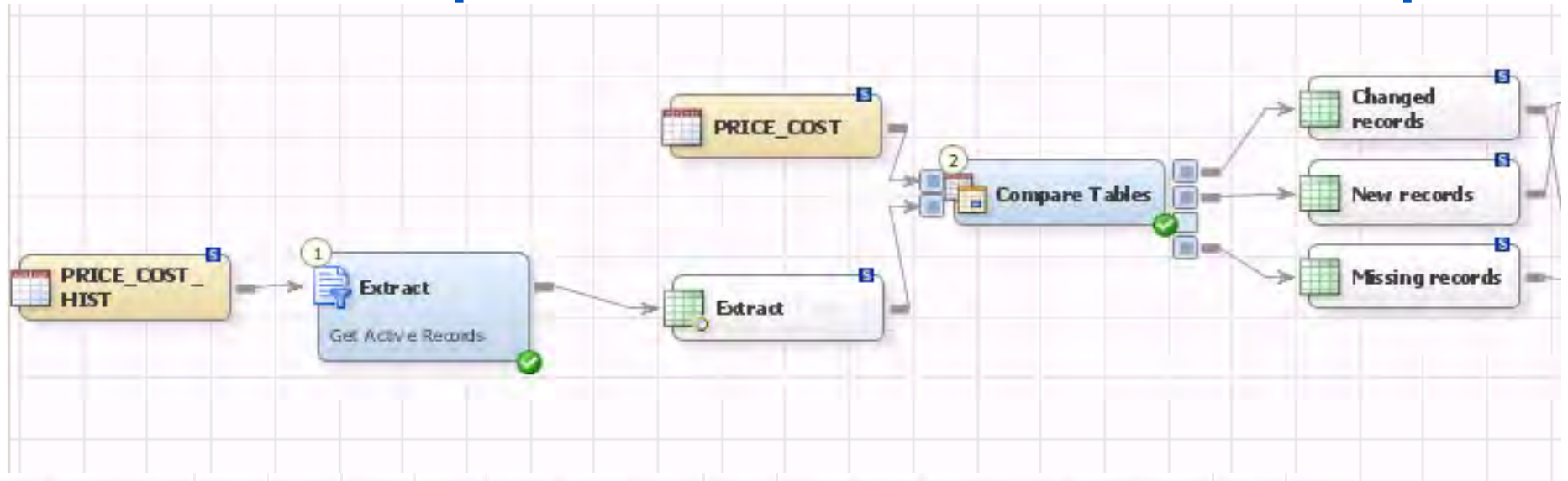


- The example above shows Type 1 SCD processing
 - EMPLOYEE_UPDATES is an extract table gather through a change detection method (source system extract, CDC, etc)
 - Compared against EMPLOYEE for changes
 - New records appended to EMPLOYEE
 - Changed records updated in place on EMPLOYEE and “Update Date” field set

New: Compare Tables Transform

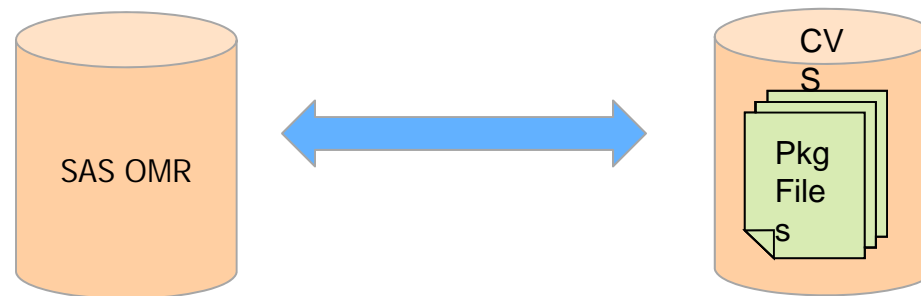
- Compare two tables (by key and change digest), outputs the following four tables:
 1. Records in Table B, NOT in Table A (new)
 2. Records in Table A, NOT in Table B (missing)
 3. Records in both tables, but with differences (updates)
 4. Records in both tables but no differences (unchanged)
- Any output table may be optionally retained/deleted based on your needs
- Supports
 - Generating a digest key in the transformation
 - Two methods of join (DATA Step MERGE vs. Hash Join)

New: Compare Tables Transform - Example



New: Versioning and Rollback

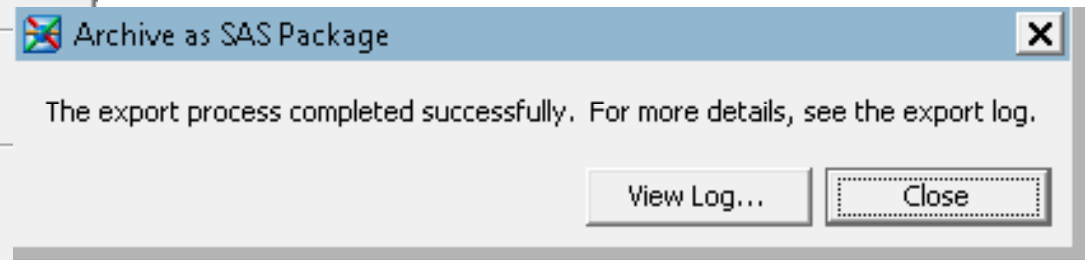
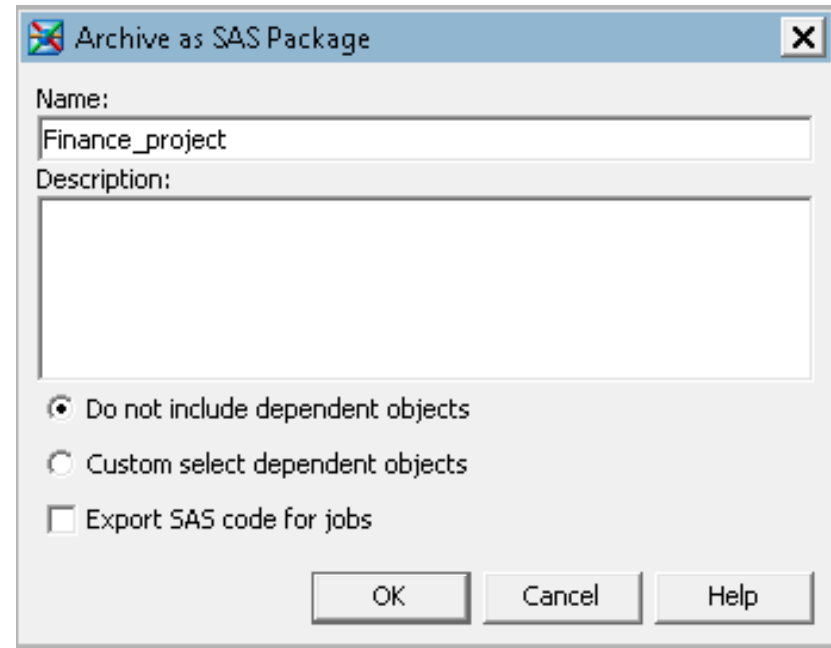
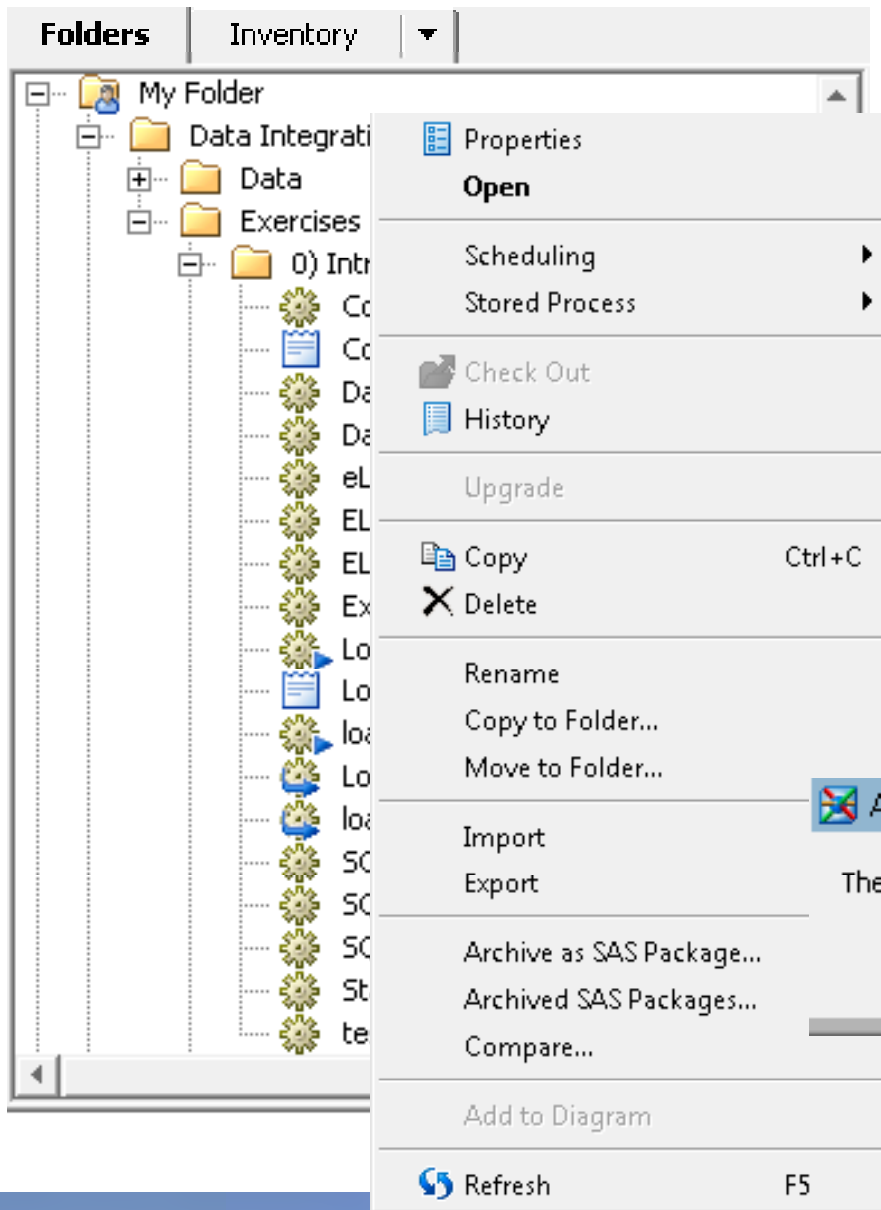
- Integration with two different open source 3rd party source management systems: CVS and SubVersion; you can choose which one you want to use
- Published API for additional source systems as a Consulting or end user option
- Example source code for plug-ins supplied



New: Versioning and Rollback

- Assumption is that there is one versioning system per metadata server
- Version package files (similar to import/export)
 - Better support sets of objects that all work together
 - Includes discovery and optional inclusion of dependent objects
 - Can include content such as SAS Source Code
- A little bit of “extra” metadata to make it easy to find objects in packages
 - Can ask any object: What packages are you in?
- Simple object differencing

New: Versioning and archiving



Import SAS Package Wizard

Select Objects to Import

The following items have been selected for import. Deselect any items that you do not want to import.

Expand All
 Collapse All
 Select All
 Clear All
 Details

- Users
 - sasdemo
 - My Folder
 - Data Integration Examples
 - Exercises
 - 0) Introduction
 - Data
 - Oracle
 - CARS
 - SAS
 - CARS
 - CARS2
 - CLASS
 - NAMES

Options | Properties

Select an item to view its options.

Package contents:

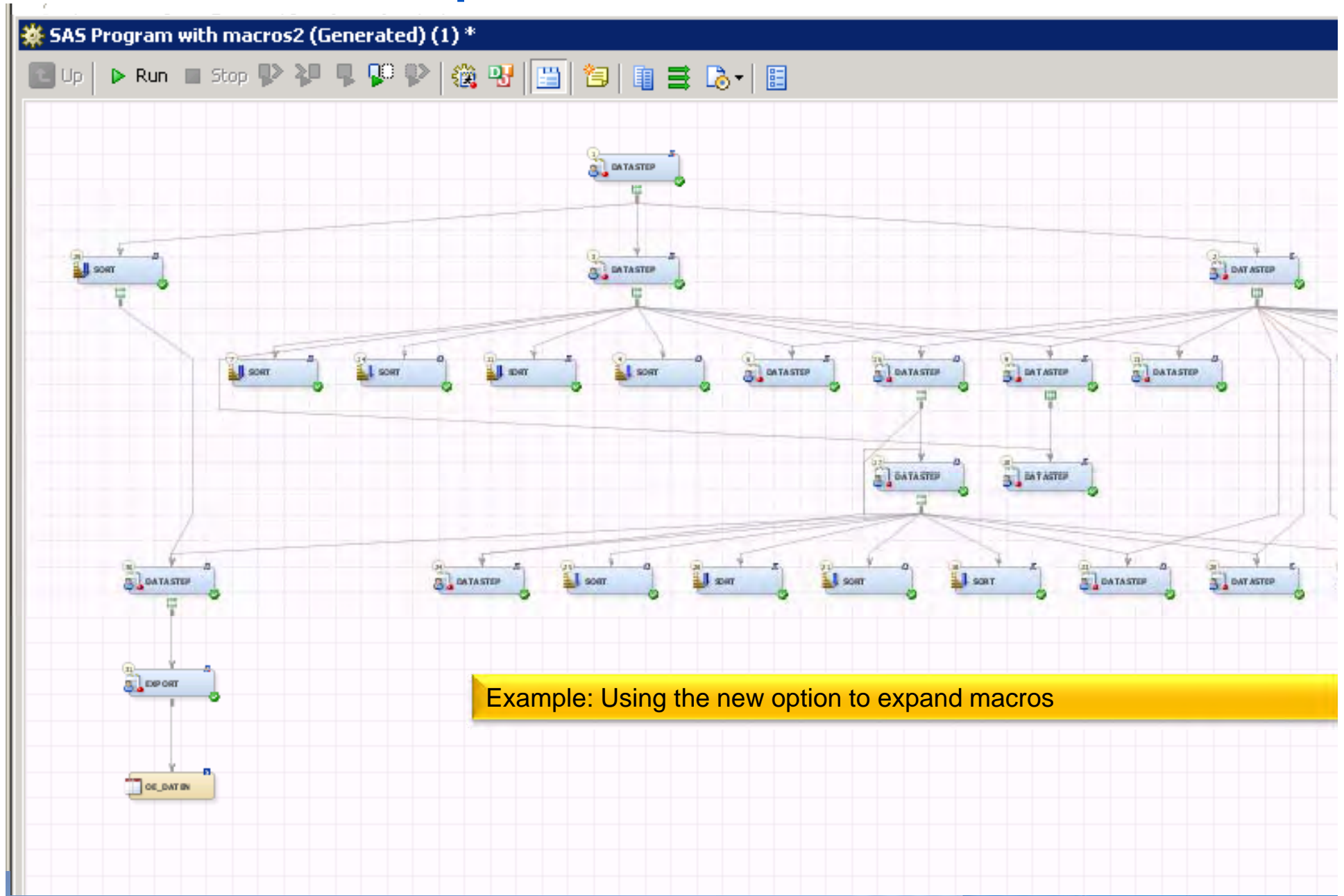
Compare To...

Close Help

Enhanced: SAS Code Importer

- Support optional Macro expansion
- New “Analyze” only feature
 - Detailed summary information
- Collision detection with warnings
 - Ex: Library already exists but different path
- Better logging and error checking
- Work tables can be registered as physical tables
- Many more fixes and user requested enhancements

SAS Code Importer Enhancements



Example: Using the new option to expand macros



For Administrators and users of Data Management jobs

New: Metadata Search

- Searches by type, name, date, other options
- Subsetting capabilities by folder, other options
- Searches can be saved
- “Find in” support

The screenshot shows the SAS Data Integration Studio 4.3 - Cypress interface. The 'Search' window is open, displaying a list of metadata items. The search panel on the left allows filtering by folder location, name, types, and date. The main window shows a table of search results.

Name	Type	Folder Location	Description	Date Modified
SAS Management ...	Folder	/System/Applications		Jul 30, 2010 4:54:42 PM
SAS Target Tables...	Folder	/heike/SIM_BWTEST01_ZO...		Aug 6, 2010 10:33:56 AM
SAS Target Tables...	Folder	/heike/SIM_BWTEST01_ZOS...		Aug 6, 2010 3:42:44 PM
SAS Web Analytics	Folder	/System/Applications		Sep 2, 2010 8:55:18 AM
sasadm	Folder	/User Folders		Jul 30, 2010 5:03:38 PM
sasbin	Folder	/User Folders		Aug 2, 2010 10:56:27 AM
sasbcm	Folder	/User Folders		Aug 2, 2010 8:46:44 AM
sasclw	Folder	/User Folders		Jul 30, 2010 5:39:04 PM
sasfjc	Folder	/User Folders		Aug 20, 2010 11:40:54 AM
saspgs	Folder	/User Folders		Aug 2, 2010 4:47:57 PM
sasrfr	Folder	/User Folders		Aug 9, 2010 11:10:57 AM
saszsh	Folder	/User Folders		Aug 30, 2010 2:32:17 PM
SCA	Folder	/User Folders/jhohm/My Folder		Jul 30, 2010 6:02:28 PM
SCA	Folder	/User Folders/klewi/My Folder		Aug 2, 2010 10:43:51 AM
SCD Type1 Testing	Folder	/dozeri		Sep 1, 2010 2:49:59 PM
SCD_Sources	Folder	/dozeri		Sep 9, 2010 1:23:41 PM
SCD Targets	Folder	/dozeri		Sep 9, 2010 1:23:50 PM
scott	Folder	/De Capite/source		Aug 4, 2010 9:48:47 AM
Services	Folder	/System	Folder used to create meta...	Jul 30, 2010 4:51:11 PM
sh	Folder	/De Capite/source		Aug 3, 2010 2:52:39 PM
sh	Folder	/De Capite/target		Aug 3, 2010 2:52:44 PM
Shared Data	Folder		Folder for shared libraries, t...	Jul 30, 2010 4:51:11 PM
SIM	Folder	/heike		Aug 2, 2010 3:58:31 PM
source	Folder	/De Capite		Aug 2, 2010 10:42:29 AM
Subfolder 1	Folder	/Mike		Sep 20, 2010 4:47:45 PM
Subfolder 2	Folder	/Mike		Sep 20, 2010 4:48:46 PM
Subscribers	Folder	/System/Publishing	Folder used to organize sub...	Jul 30, 2010 4:51:11 PM
SybaseIQ Tables	Folder	/Data		Aug 2, 2010 3:09:18 PM
System	Folder		Reserved for SAS system use...	Jul 30, 2010 4:51:11 PM
System	Folder		Reserved for SAS system use...	Sep 2, 2010 8:55:18 AM
S-ARM	User group		SCS	Jul 30, 2010 5:36:51 PM
s0652838	Table	/Data/SybaseIQ Tables		Aug 2, 2010 3:09:59 PM
s0652838	Table	/Data/Greenplum Tables		Aug 2, 2010 3:13:32 PM
s0654036	Job	/User Folders/Mike Writers/...		Sep 13, 2010 2:36:57 PM
SZD_IND_APP_AS...	Job	/User Folders/klewi/My Fold...		Aug 2, 2010 3:39:59 PM
s390deva MVS Ser...	Application ...			Aug 10, 2010 1:36:51 PM
s390deva MVS Ser...	Logical wor...			Aug 10, 2010 1:36:58 PM
S_ACCNT_ATT	Table	/User Folders/klewi/My Fold...	S_ACCNT_ATT	Sep 17, 2010 1:35:35 PM
S_DNB_ORG_EXT...	Table	/User Folders/klewi/My Fold...	S_DNB_ORG EXTRACT	Sep 17, 2010 1:35:36 PM
S_DNB_ORG_JOIN...	Table	/User Folders/klewi/My Fold...	S_DNB_ORG JOINED	Sep 17, 2010 1:35:37 PM
S_DNB_ORG_SORT	Table	/User Folders/klewi/My Fold...	S_DNB_ORG SORT	Sep 17, 2010 1:35:39 PM
saacos	User			Jul 30, 2010 5:36:27 PM
saadub	User			Jul 30, 2010 5:36:40 PM
saaran	User			Jul 30, 2010 5:36:49 PM
sababi	User			Jul 30, 2010 5:36:25 PM

New: Column Metadata Standardization

- Set column attributes for all tables to the same values
- Look across all tables in a library
- Impact Analysis sees impact of changes

Column Standardization Tool

Non-standard Columns

Standardize the non-standard columns Step 2 of 4

Selected Column: **Street** [Standardize] [Rollback] [Impact Analysis]

Column Names

Sort By: By Name

- LName
- Position
- SS
- STATE
- Salary
- Sex
- Street**
- Weight
- Zip
- acctno
- code
- decimal
- desc1
- desc2
- empcode
- hex
- postal

Column Details

Data Type	Length	Format	Informat	Description	Is Nullable
Char	12	\$12.	\$12.		yes

Double click any row in the table below to populate the values in the standard row above.

Include		Table Name	Column Name	Library Name	Data Type
<input checked="" type="checkbox"/>	!	LOOKUPSOURCE	Street	Lookup Source	Char
<input checked="" type="checkbox"/>	!	LOOKUPSOURCE_17164	Street	Lookup Source	Char
<input checked="" type="checkbox"/>	!	LOOKUPSOURCE_PL	Street	Lookup Source	Char

NOTE: The values are not updated in the metadata when you click only standardize. Values will only be updated when you click Finish at the end of this wizard.

< Back [Next >] [Finish] [Cancel] [Help]

New: Automation of job deployment process

- Command line driven, DI Studio client not required
- Deploy supports single and multiple jobs in a folder
- Same features as “deploy for scheduling” driven from a command line
- Useful for
 - Scheduling deployments
 - Scripted job deployments
 - Fast redeployment

Example:

```
C:\>deploySASJobs -metaserver CONF11140 -deploymentdir c:\jobs_to_deploy_
```

New: Reporting on batch job execution

- SAS DI provides a couple of sas jobs that extract information from the SAS logs and create tables containing information such as:
 - Jobs
 - Step Details
 - Warnings
 - Errors
 - Job Statistics
- Tables can be viewed and analyzed and shared with others
- Reports are provided in a number of accessible formats:
 - SAS Stored Processes
 - SAS Web Report Studio Reports
 - SAS Information Maps, and SAS OLAP Cubes

New: Reporting on batch job execution

SAS Web Report Studio : View Report - Windows Internet Explorer

http://brdvm0194.chn.sas.com:8080/SASWebReportStudio/openRVUrl.do?rsRID=5B1P%3A%2F%2FMETASERVER%2FDIStudioJobStats%2FDI+Job- Live Search

File Edit View Favorites Tools Help Links Customize Links

SAS Data Integration Studio ... SAS Web Report Studio : ...

Log Off SAS operator | Preferences | Hp

SAS Web Report Studio - DI Job Performance Reports

File View Data Edit View

Applied filters: None
Date > 2010

Columns 1 - 12 of 24

Week	W23				W27				W28			
	CPU Time	Duration	System I/O	System Memory	CPU Time	Duration	System I/O	System Memory	CPU Time	Duration	System I/O	System Memory
Job Name												
Parent_job	0:00:00.72	0:00:07.83	51,631,697	33,894	0:00:01.94	0:00:27.58	170,360,755	162,918	0:00:02.44	0:00:37.81	238,870,713	228,086
loop_job	0:00:00.23	0:00:02.56	6,004,180	12,509	0:00:00.81	0:00:23.55	30,138,660	59,924	0:00:01.11	0:00:28.06	42,226,892	83,894
simpleJob	0:00:01.33	0:00:10.98	59,842,288	12,247	0:00:18.19	0:00:43.10	1926792880	58,614	0:00:27.44	0:01:14.43	2697510032	82,059
test	0:00:00.09	0:00:04.42	3,529,742	12,247	0:00:00.56	0:00:21.08	17,807,430	58,614	0:00:00.55	0:00:29.17	24,930,402	82,059
Total	0:00:02.38	0:00:25.80	121,007,907	70,898	0:00:21.50	0:01:55.30	2145099725	340,070	0:00:31.53	0:02:49.48	3003538039	476,099

CPU Time

Week: W23, W27, W28, W29, W30

CPU Time

Week: W23, W27, W28, W29, W30

Table of Contents: Section1 (No group breaks are defined.)

Section Data Options: DI Job Step Stats Cube Map

- Date Hier
- Job Hier
- CPU Time
- Duration
- User CPU Time
- System I/O
- System Memory

Enhanced: Data Quality Capabilities

- For SAS Enterprise Data Integration Server users
 - DataFlux Data Management Studio
 - » Data Profiling
 - » Metadata Exploration
 - » Quality, Integration, and Entity Resolution
 - DataFlux Data Management Server for SAS
 - » Profile, Quality, Integration, and Entity Resolution
- For SAS Data Integration Server users
 - DataFlux Data Management Studio
 - » Data Profiling

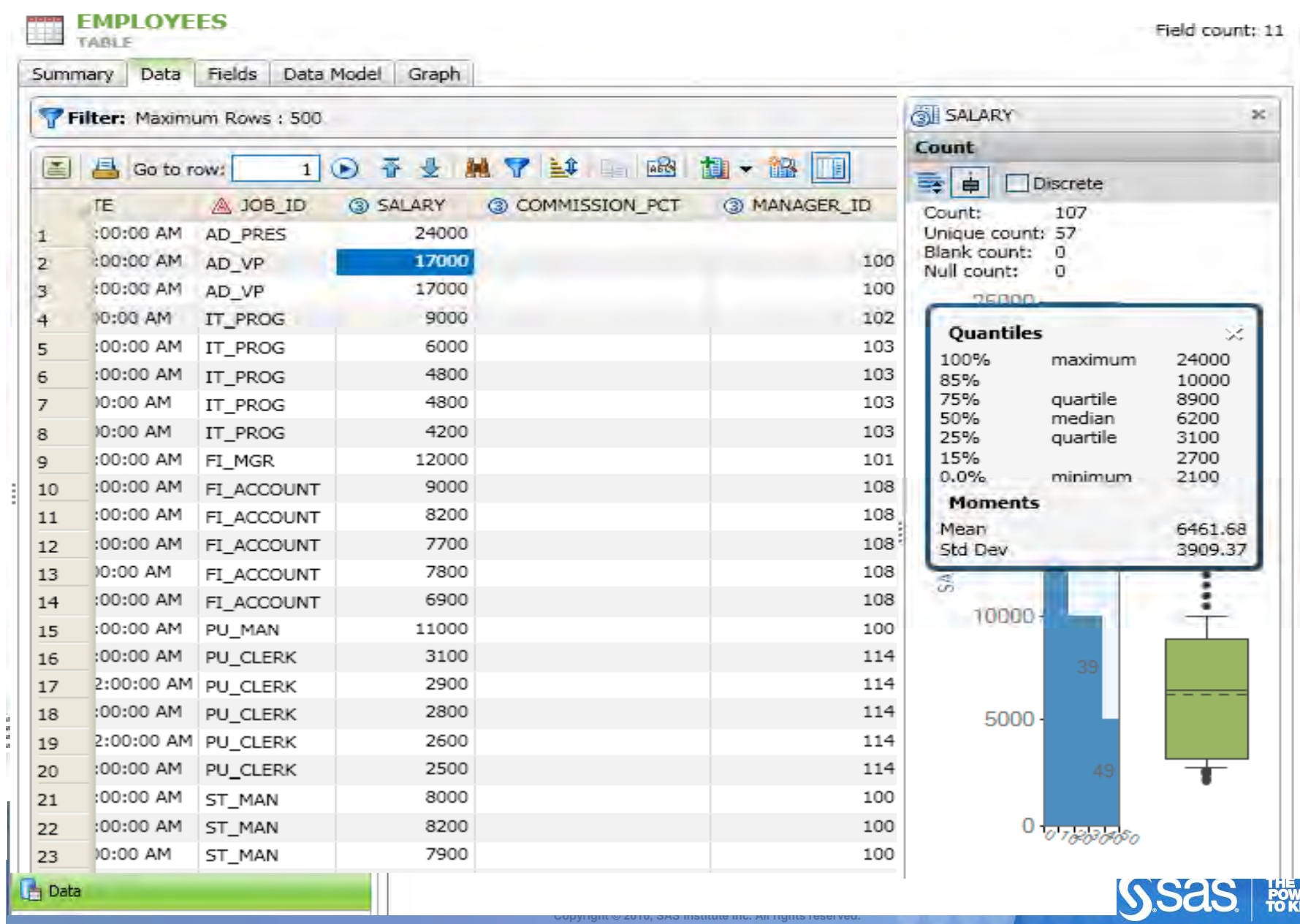
New: Unified Data Quality interface

The screenshot displays the DataFlux Data Management Studio 2.2 interface. The main window is titled "DataFlux Data Management Studio 2.2" and features a menu bar with "File", "View", "Tools", "Window", "Help", and "New". The interface is divided into several panels:

- Information Panel (Left):** Contains a tree view with "Overview" (selected) and "Monitor".
- Overview Panel (Top Center):** Shows "Recent Files" with entries for "Product" and "Organization". Below this is "The Data Roundtable" section, which lists several blog posts with dates and titles, such as "Data Governance and the Occasional Vow of Celibacy" and "Data Quality Leaders Need Power to Sustain a Sense of Urgency". At the bottom of this panel is the "Default Settings" section, listing configuration details like Host ID, Active QKB, License Primary Method, License Primary Location, User Directory, and Application Root.
- Data Management Methodology Panel (Top Right):** Displays a circular diagram representing the Data Management Methodology. The cycle consists of six stages: CONTROL, OFFLINE, PLAN, DISCOVER, DESIGN, and EXECUTE, with arrows indicating a clockwise flow. The word "MONITOR" is positioned to the left of the cycle, and "ACT" is positioned below it.
- Documentation Panel (Bottom Right):** Lists various help and reference guides, including "Data Management Studio Online Help", "Expression Language Reference Guide", "ODBC Driver Online Help", "ODBC Driver Reference Guide", "Federation Server TSSQL Reference Guide", "Data Management Server User's Guide", "Federation Server User's Guide", "Authentication Server User's Guide", "Data Management Server Administrator's Guide", "Federation Server Administrator's Guide", "Authentication Server Administrator's Guide", "Data Management Studio Installation Guide", and "Data Management Studio Release Notes".
- Quick Links Panel (Bottom Right):** Provides links to "DataFlux.com", "News and Events", and "DataFlux Portal".

The bottom of the interface features a navigation bar with icons for "Information", "Data", "Folders", "Glossary", "Data Management Servers", and "Administration".

New: Explore Data Model and Tables



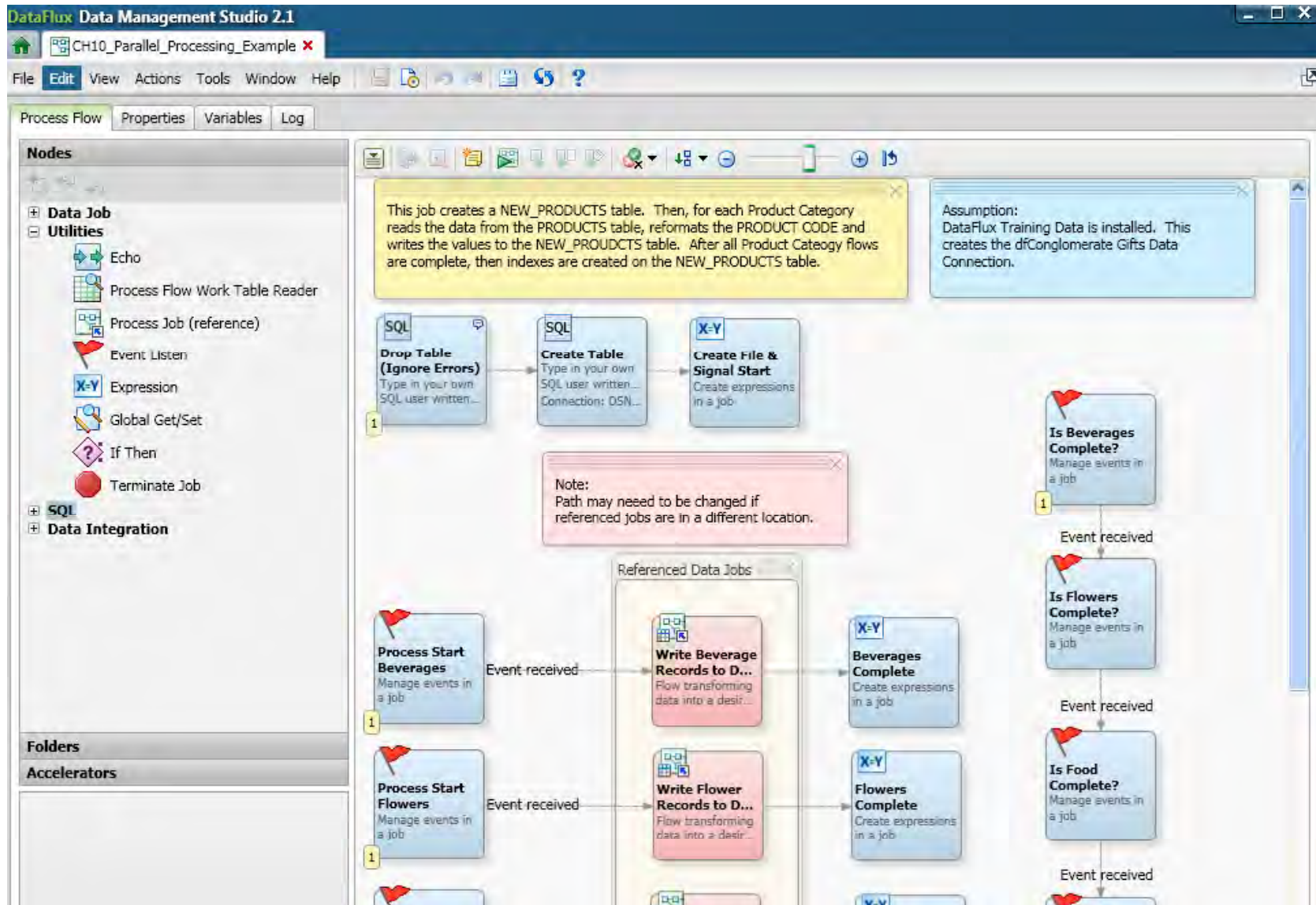
Enhanced: Data and Metadata Discovery

- Sample data to discover content
- Quality Knowledge Base definitions help locate address info, phones, city, zip, other content in your data sources

The screenshot displays the DataFlux Data Management Studio 2.1 interface. The main window shows a 'Field relationship map' with a circular network of nodes and edges. A context menu is open over a node, with 'Matching Fields' selected, showing a list of fields: ADDRESS, Address, Address, STREET_ADDR, and ADDRESS2. The 'Data Sample' panel on the right shows a list of addresses, with 'NULL' highlighted. The 'Identification Analysis' panel on the left shows a tree view of 'All Identification Definitions' with 'Metadata' expanded to show 'CITY'. A table below the 'CITY' definition shows 'Identity matches'.

Field Name	Table
City	Client_Info
City	Prospect_Data
CITY	Contacts
CITY	NC_Customer
CITY	Purchase
res_City	Precinct_01

New: Process orchestration



New: Integrated Master Data Management

- Coming later this year in DMP 2.2
- Chooses best record from all sources
- Wizard for mapping source to target
- Job generator
- Use output in downstream processes
- Easily upgrade to full qMDM for additional features

The screenshot shows a software window titled "New Master Data Project" with a close button (X) in the top right corner. The window is divided into two steps, with "Step 1 of 2" indicated in the top right. The current step is "Identification".

The form contains the following fields and controls:

- * Name:** A text input field containing the word "Organization".
- * Database location:** A text input field that is currently empty, with a small square button to its right.
- * Repository:** A dropdown menu with a folder icon on the left and a downward arrow on the right. The selected option is "MDM". Below this dropdown, a note reads: "Reports, documentations, and jobs will be stored in the repository."
- * Entity definition:** A dropdown menu with a folder icon on the left and a downward arrow on the right. The selected option is "Organization".

At the bottom of the window, there are three buttons: "Create", "Cancel", and "Help".

- Data
- Collections
- Data Connections
- Master Data Projects
 - Organization

Organization
MASTER DATA PROJECT

Field count: 21

Summary Data Clusters Fields

Filter: Minimum cluster size = 2

Page: 1 of 1

	MDM_SRC_SYS_NAME	MDM_SRC_SYS_REC_ID	POLICY_ID	ORG_NAME	ADDRESS1
1	MDM	91	10	DataFlux Corporation	6512 Six Forks Road - 404B
2	MDM	173	1	First Merit Bank	19 East Broad Street
3	MDM	650	1928	Tulare County	4002 Westminister Ave.
4	MDM	803	1182	Farmers Insurance Group Inc.	3145 E Flamingo Rd Ste 1007
5	MDM	810	11267	SAS	6512 Six Forks #404B
6	MDM	929	3045	Queen of Angels Medical Center	3200 E Carson St
7	MDM	1520	102	Jackson Data Co.	25670 W HEDGEWOOD DR.
8	MDM	1551	134	Kaiser Permanente	75 Columbus Avenue
9	MDM	1563	129	Kasier Permanente	75 Columbus Avenue
10	MDM	2395	130	Kaiser Permanente	75 Columbus Ave
11	MDM	3958	133	Kaiser Permanente	75 Columbus Avenue
12	MDM	4433	2981	Desert Samaritan Health Systems	1 Park Plaza, 11th Flr
13	MDM	4534	2380	St. Mary's Hospital	2525 Dupont Dr
14	MDM	4693	2003	Science Applications Int'l Corp.	2001 W Chandler Blvd

Cluster records:

Surviving record:

MDM	173	1	First Merit Bank	19 East Broad Street (null)	Saint Louis (null)
-----	-----	---	------------------	-----------------------------	--------------------

Clustered records:

	MDM_SRC_SYS_NAME	MDM_SRC_SYS_REC_ID	POLICY_ID	ORG_NAME	ADDRESS1	ADDRESS2	CITY	COUNT
1	DataFlux Sample.Contacts	3	3	1st Merit Bank	19 E Broad St	(null)	St. Louis	(null)
2	DataFlux Sample.Contacts	1	1	First Merit Bank	19 East Broad Street	(null)	Saint Louis	(null)
3	DataFlux Sample.Contacts	5	5	First Marret Bank	19 E. Broad	(null)	Saint Louis	(null)

- Information
- Data
- Folders
- Glossary
- Data Management Servers
- Administration

Other enhancements

- Many customer requests and issues addressed (>150)
- Auto mapping of derived columns
- Enhanced Lineage reporting
- Enhanced Grid submits
- Java JRE 1.6
- Tools options...Code Generation....new options for lines less than 80 characters
- New Stored Process (SP Version 2) support (optional)
- Migration support
 - 9.1.3 -> 9.3
 - 9.2 -> 9.3
 - 9.3 -> 9.3 (same OS, optionally can be on different machines)

Summary

SAS Data Integration Studio 4.3

- Significant update for SAS Data Integration users
 - Easier to Manage and Support
 - Delivers Better Performance
 - Innovation in Data Quality and MDM Integration

Continued innovation in data management

- Stay tuned...there is a lot more to look forwards to.
- Keep in touch with SAS...let us know how we can help!



Get started with SAS Data Integration Studio 4.3

- Contact your account team for consultation
- Not sure who to contact? Check www.sas.com/contact for list of global phone and Email resources