

What's New in SAS® Data Management

Nancy Rausch, SAS Institute Inc., Cary, NC

ABSTRACT

The latest releases of SAS® Data Integration Studio and DataFlux® Data Management Platform provide an integrated environment for managing and transforming your data to meet new and increasingly complex data management challenges. The enhancements help develop efficient processes that can clean, standardize, transform, master, and manage your data. The latest features include capabilities for building complex job processes, new web-based development and job monitoring environments, enhanced ELT transformation capabilities, big data transformation capabilities for Hadoop, integration with the analytic platform provided by SAS® LASR™ Analytic Server, enhanced features for lineage tracing and impact analysis, and new features for master data and metadata management. This paper provides an overview of the latest features of the products and includes use cases and examples for leveraging product capabilities.

INTRODUCTION

The latest releases of SAS® Data Integration Studio and DataFlux® Data Management Studio, and additional SAS® Data Management enhancements provide new and enhanced features to help data warehouse developers, data integration specialists, and data scientists carry out data-oriented tasks more efficiently and with greater control and flexibility. Major focus areas for the latest releases include new features that support data governance and data stewardship, big data, quality, collaboration, and governance, plus new monitoring features. This paper showcases some of the newest features available in the SAS Data Management products.

SAS® DATA LOADER: BIG DATA ETL FOR HADOOP

When traditional data storage or computational technologies struggle to provide either the storage or computational power required to work with large amounts of data, an organization is said to have a big data issue. Big data is frequently defined as the point at which the volume, velocity, or variety (or a combination of these) of data exceeds an organization's storage or computational capacity for accurate and timely decision-making.

The most significant new technology trend that has emerged for working with big data is Apache Hadoop. Hadoop is an open source set of technologies that provide a simple, distributed storage system paired with a fault-tolerant parallel processing approach that is well suited to commodity hardware. Many organizations have incorporated Hadoop into their enterprise, leveraging the ability for Hadoop to process and analyze large volumes of data at low cost.

SAS is integrating with the Hadoop platform to bring the power of SAS to help address big data challenges. SAS, via the SAS/ACCESS® technologies and code accelerator products, has been optimized to push down computation and augment native Hadoop capabilities to bring the power of SAS to the data stored in Hadoop. By reducing data movement, processing times decrease and users are able to more efficiently use computing resources and database systems.

SAS® Data Loader for Hadoop is the SAS offering that provides support for big data. Figure 1 is an overview of the architecture of SAS Data Loader.

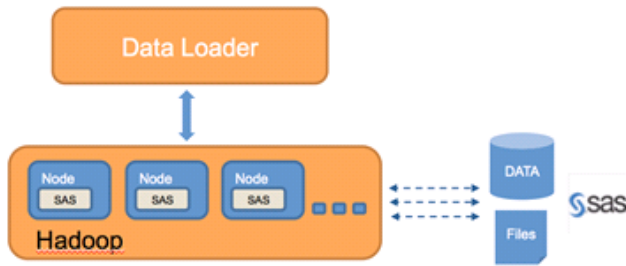


Figure 1: SAS Data Loader for Hadoop Architecture

The SAS Data Loader offering includes a client for building and running jobs in Hadoop that can leverage process capabilities embedded in both Hadoop and SAS. The offering also includes components required to install and distribute SAS on the Hadoop cluster to enable Data Quality, ETL, and Analytic Data Prep features by using Hadoop. Figure 2 shows some of the main transformations and features available to work with data in Hadoop.

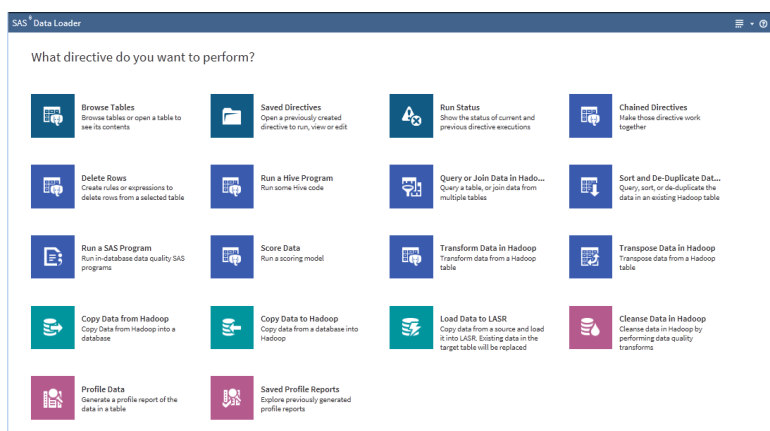


Figure 2: SAS Data Loader Client Main Screen

SAS Data Loader integrates with native Hadoop capabilities such as Oozie, Sqoop, and Hive for reading, transforming, and writing data in parallel by using Hadoop. SAS Data Loader also includes the SAS Embedded Process and generates SAS MultiVendor Architecture and SAS DS2 code to perform various transformations on data in Hadoop. The offering also provides the Data Quality Accelerator for Hadoop and a Quality Knowledge Base offered in a number of different languages to support data quality actions such as standardizing addresses, state codes, and phone numbers, and parsing data into standard or customizable tokens, and identifying data into known types such as names and addresses.

The capabilities of SAS Data Loader include the ability to parallel load data into or out of Hadoop from RDBMS systems, SAS data sets, and delimited files. Data can also be loaded into a SAS® LASR™ Analytic Server. Figure 3 is an example of the Copy Data from Hadoop directive.

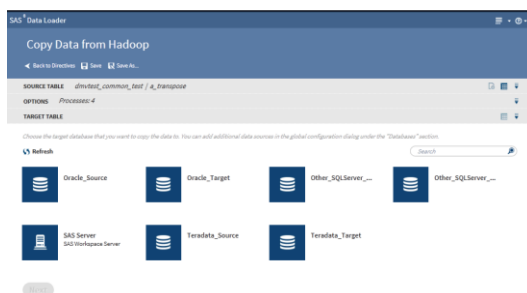


Figure 3: Copy Data from Hadoop Example

Data can be stored in a variety of Hadoop formats, including Parquet, Sequence, and ORC files, and can be delimited using customer delimiters. Figure 4 is an example of the option screen for selecting data persistence format types.

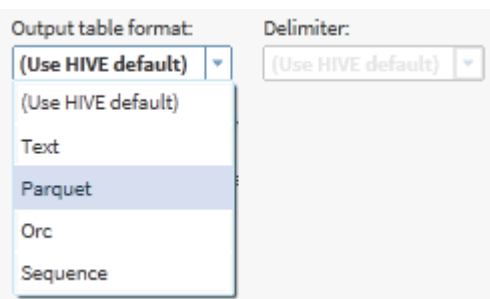


Figure 4: Supported Hadoop Data Persistence Types

You can also build complex joins of 1-N tables or queries by leveraging Hive capabilities. The example in Figure 5 shows a three-table join and the types of joins available.

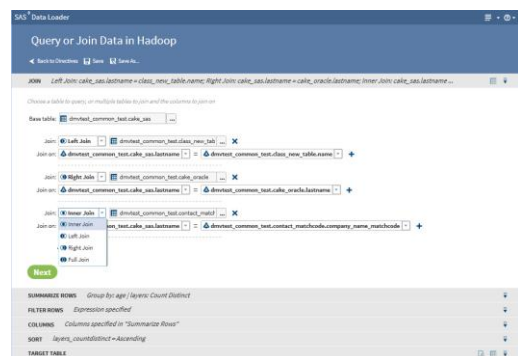


Figure 5: Join Example

Another feature supports the creation of complex mappings from source to target data by using Hive or SAS DS2 syntax. Figure 6 is an example of some of the types of expressions available to the mapping transformation.

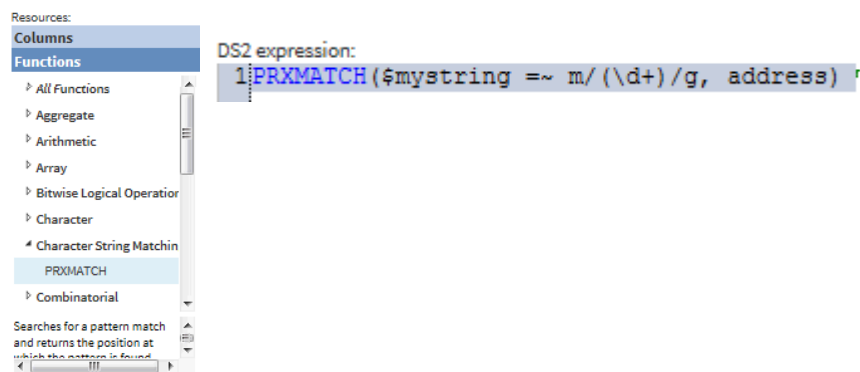


Figure 6: Example of Available Expressions for Mappings

You can also perform data-cleansing operations in-database in multiple languages such as parsing, standardization, filtering, casing, pattern analysis, and others. Figure 7 shows some of the data quality transformations available.

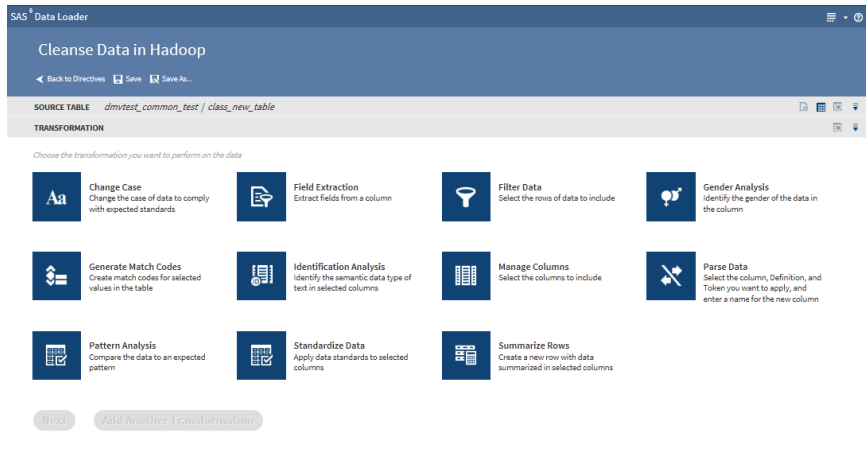


Figure 7: Data Quality Transformations

For example, you can standardize columns to a variety of preshipped standards, or you can add your own. Figure 8 shows some of the available standards that are delivered with the DQ accelerator.

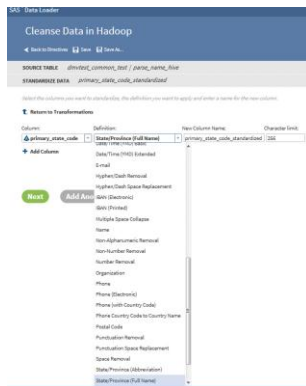


Figure 8: Example of Available Data Quality Standards

Figure 9 shows an example of applying standardization to data in a Hadoop data set. The original data, outlined in blue, has state names that in some records are two-letter abbreviations, and in other records are full names. Using the in-database data quality standardize technique, you can apply a state name standard to your data. The result is placed into a new column in the table. The results column outlined in red shows the new column added to the table with state names standardized to the full state name.

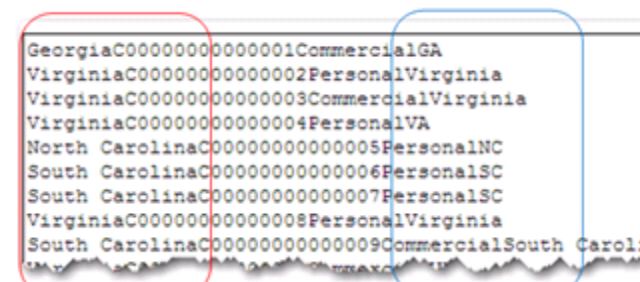


Figure 9: Example of Standardization

Another useful transformation that is available is the ability to transpose data. You can transpose tables by using a variety of techniques. Figure 10 is an example of the transpose feature.

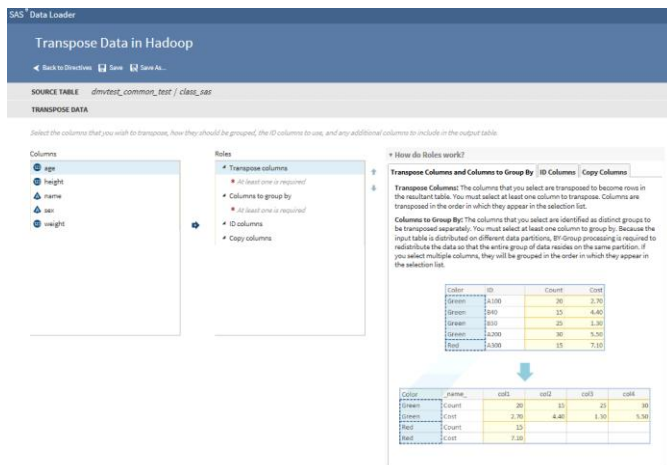


Figure 10: Transpose Example

There are other transforms available that include the ability to delete rows of data by using a WHERE clause and to run in-database scoring models and SAS or Hive programs.

SAS also supports profiling Hadoop data via the new SAS in-database profile engine. Using the in-database features of profile enables you to quickly see the quality of your data stored in Hadoop. The web profile viewer enables you to view a number of data quality metrics associated with your data, such as the count of missing values, the range of values, string patterns, frequency analysis, and other data quality metrics. Figure 11 shows some of the available profile results views of Hadoop data.

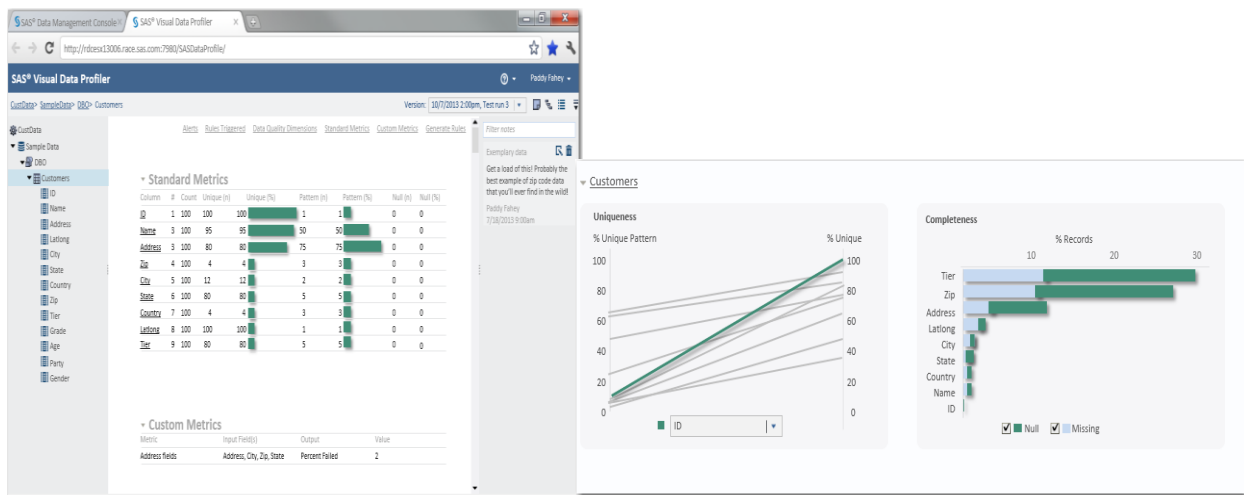


Figure 11: Profile Metrics Examples

You can also view the status of your jobs running in Hadoop. Because jobs running on the cluster are running independently of the client, you can disconnect and come back later to check the progress. Logs are available for jobs that have run, and you can rerun jobs from the job status window. Figure 12 is an example of the Run Status viewer.

SAS® Data Loader

Run Status


← Back to Directives

Show: Last 30 Days Refresh Clear All

Name	Status	Start Time	End Time	Run Time
Cleanse Data in Hadoop	Successful	Feb 28, 2015, 2:12:34 PM	Feb 28, 2015, 2:14:20 PM	00:01:45.865
Cleanse Data in Hadoop	Successful	Feb 28, 2015, 1:55:45 PM	Feb 28, 2015, 1:56:40 PM	00:00:55.645
Cleanse Data in Hadoop	Successful	Feb 28, 2015, 1:42:33 PM	Feb 28, 2015, 1:44:19 PM	00:01:45.640
Cleanse Data in Hadoop	Successful	Feb 28, 2015, 1:23:51 PM	Feb 28, 2015, 1:25:41 PM	00:01:49.892
Cleanse Data in Hadoop	Successful	Feb 28, 2015, 1:19:24 PM	Feb 28, 2015, 1:20:28 PM	00:01:03.343
Profile Data	Successful	Feb 23, 2015, 3:48:29 PM	Feb 23, 2015, 3:55:15 PM	00:06:45.371
hadoop_filter_filter	Successful	Feb 20, 2015, 10:47:02 AM	Feb 20, 2015, 10:49:06 AM	00:02:03.751
GHC_MultiACView	Failed	Feb 20, 2015, 10:47:02 AM	Feb 20, 2015, 10:47:40 AM	00:00:37.518
Run a SAS Program	Successful	Feb 19, 2015, 11:59:06 AM	Feb 19, 2015, 11:59:07 AM	00:00:00.688

Figure 12: Run Status Viewer

SAS Data Loader works with multiple production Hadoop distributions such as Cloudera and Hortonworks. The product is also offered as a 90-day free trial and deploys easily to work with sandbox distributions of Hadoop. See Figure 13 for an example.



SAS Data Loader for Hadoop
Make Hadoop data accessible for everyone.

- Manage Hadoop data -- no coding required.
- Unlock more insights within big data.
- Improve scalability, performance and security.

[Power up your free trial](#)

Figure 13: Trial Download Example

SAS DATA INTEGRATION STUDIO

SAS Data Integration Studio supports traditional ETL and ELT capabilities by using SAS, SQL, and pushdown SQL. The latest release of SAS Data Integration Studio has added a number of new features. One of the key enhancements is a new process fork node that enables you to run multiple flows in parallel in a job. The fork node spawns a new parallel SAS process when it is run inside of a job. All of the nodes between the fork and the fork end transform run in a parallel process. The fork also supports grid processing when Grid is available, and works similarly to the existing Loop transform. Figure 14 is an example of the new fork and fork end transforms, showing two parallel processes.

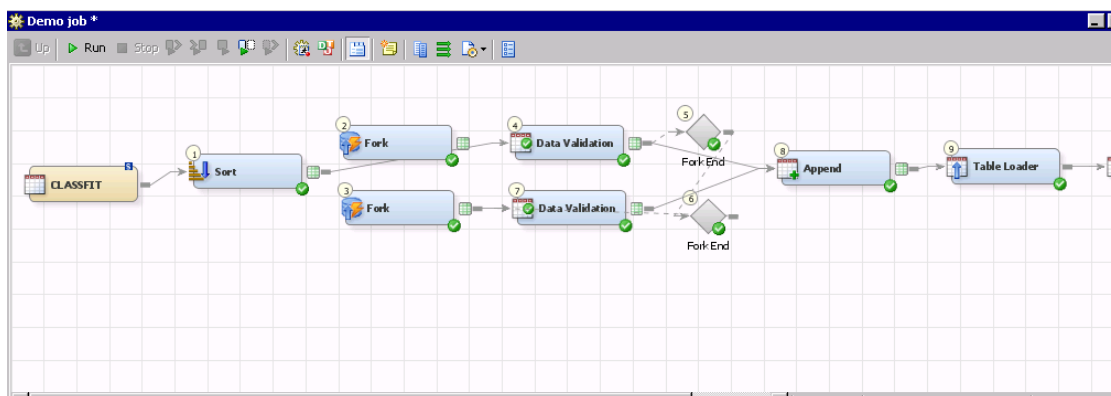


Figure 14: SAS Data Integration Studio Fork Transform

Several new data sources and targets have also been added. Support for Pi data is available and Hawq SQL data sources and targets are now supported via two new SAS/ACCESS engines.

SAS® DATA MANAGEMENT

SAS® Data Management has a number of new features as well. The Data Management products have added support for integration with SAS® Metadata Server. This enables both to support Integrated Windows Authentication and single sign-on, as well as other authentication modes supported by SAS Metadata Server. Shared users and groups are now also supported.

SAS Data Management has added a new public REST API. The REST interfaces support access to most server features including job generation, execution, and status. Figure 15 is an example of the new server REST interface Base URL.

Base URL

```
http://www.example.com/SASDataMgmtRTDataJob/rest
```

```
http://www.example.com/SASDataMgmtRTProcessJob/rest
```

Figure 15: Example of the Data Management Server REST Interface

SAS® FEDERATION SERVER

Data federation is a data integration methodology that allows a collection of data tables to be manipulated as views created from diverse source systems. It differs from traditional ETL and ELT methods because it pulls only the data needed out of the source system. Figure 16 illustrates the differences between data federation and traditional ETL and ELT.

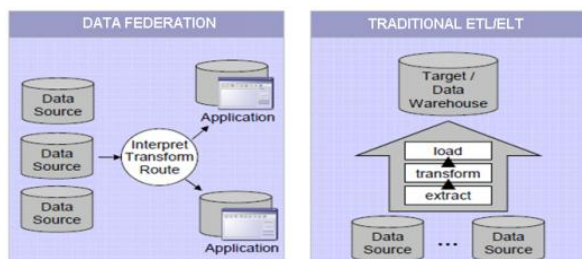


Figure 16: Illustration of the Differences between Data Federation and Traditional ETL and ELT

Typically, a data federation methodology is used when traditional data integration techniques cannot meet the data needs, such as when the data is too large, too proprietary, or too mission-critical to be extracted out of the source systems. Data federation solves this challenge because only the needed information is gathered from the source systems as views that can be delivered to downstream processes. Federation allows the data to be extracted and stored in a persistent cache, which can be updated periodically or scheduled to be refreshed during non-mission-critical times. Federation is also a good choice for systems where data is diversified across source systems. Managing security, user IDs, authorizations, and so on, for all of the various source systems can be a huge burden for a traditional data integration model. Data federation is well suited for this usage scenario because it enables system integrators to have a single point of control for managing diverse system security environments, and for updating views when source systems change.

SAS® Federation Server provides these capabilities. It includes a data federation engine, multi-threaded I/O, pushdown optimization support, in-database caching of query results, an integrated scheduler for managing cache refresh, a number of data source native engines for database access, full support for SAS data sets, auditing and monitoring capabilities, many security features including table, column, and row security, and a number of other federation features. Access to the federation server is available via a number of interfaces including a REST API, JDBC, ODBC, and a SAS/ACCESS engine.

Figure 17 is a high-level overview of SAS Federation Server.

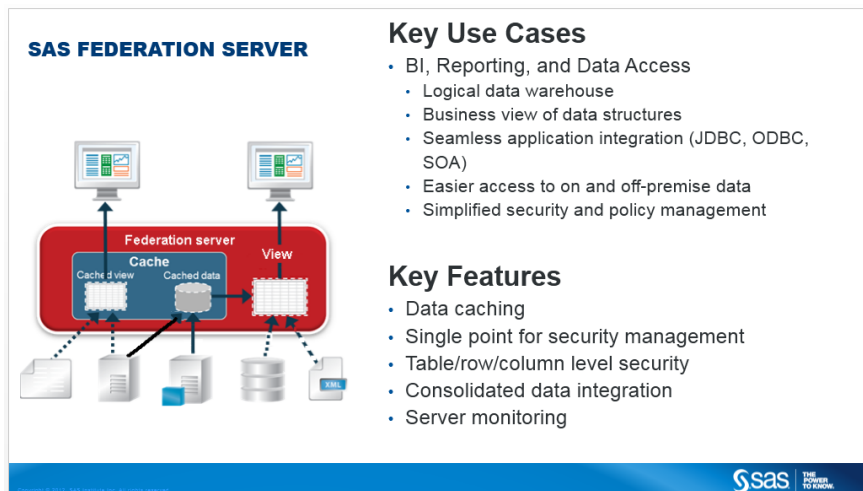


Figure 17: SAS Federation Server Overview

SAS Federation Server has been updated in the latest release to integrate with the SAS Metadata Server and the SAS® Web Infrastructure Platform Database. These features enable SAS Federation to support enterprise-class features such as Integrated Windows Authentication, common authentication, and authorization features including shared user management, consistency in installation, configuration and administrative tasks, and standard SAS web infrastructure capabilities such as custom theming support.

SAS Federation Server also includes support for data masking. Three functions, ENCRYPT, DECRYPT, and HASH, support the ability to mask sensitive information in your data tables. Here is some sample source code that uses data masking:

```
// Create table w/ encrypted NAME column:
create table "EMPLOYEES_ENCR" as
select *,
       syscat.dm.mask('ENCRYPT', "NAME",
                     'alg', 'AES',
                     'deterministic', 'yes',
                     'cta_values', 'yes',
                     'key', 'xyzzzy') as "NAME_ENCRYPTED"
from "EMPLOYEES";
```

Figure 18 is an example of applying this sample source code to some data. The first column in the data set on the left contains the original, unencrypted name. Applying the data masking function to the data results in the data set on the right. The results set has masked the name column.

1	Alfred	M	14	116.94168422...	135.0706559852119
2	Alice	F	13	68.906553332...	85.63003016134127
3	Barbara	F	13	105.26025321...	117.89926300818239
4	Carol	F	14	96.375045159...	107.2893197211607
5	Henry	M	14	98.982069499...	110.14093775704566
6	James	M	12	70.63208847...	68.4074407508043

1	D022CDAB	M	14	116.94168422...	135.07065598...
2	F0282E2E	F	13	68.906553332...	85.630030161...
3	F028CDA5	F	13	105.26025321...	117.89926300...
4	F0280141	F	14	96.375045159...	107.28931972...
5	F0280EE8	M	14	98.982069499...	110.14093775...
6	B01E453D	M	12	70.63208847...	68.40744075...

Figure 18: Before and after Data Example Using a Federation Server Data Masking Function

There are a number of available new source database types, including Apache Hadoop Hive data, Postgres SQL, SAP HANA, and SASHDAT format for writing data files to a SAS analytics server.

Another key feature is the ability to persist content in memory. This can be very useful if there are critical resources that need to be available on demand. You can persist views, tables, and data caches in the in-memory storage facility and cycle it out of memory when it is no longer needed.

Federation servers can also be chained so that one federation server can be used as a source or target

of another federation server. This can be useful if you need to synchronize between different sites such as headquarters and regional offices. You can use data federation to help manage resources between the various locations by using features that the federation server provides such as data caching, security, and fast table access using the in-memory store.

Federation Server also supports the ability to run SAS DS2 programs on the data flowing through the federation server to support a variety of useful functions such as data cleansing, data consolidation, and ETL features such as joins, updates, and queries. Figure 19 is an example of a federation server DS2 program.

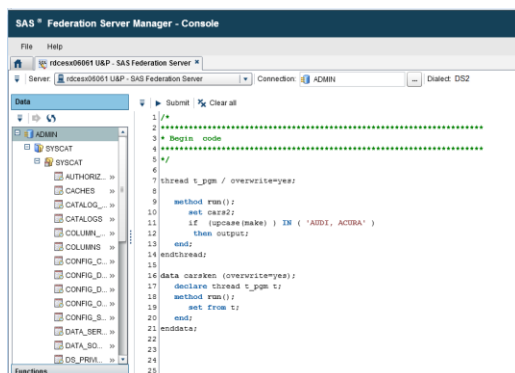


Figure 19: Example Federation Server DS2 Program

SAS® BUSINESS DATA NETWORK

SAS® Business Data Network enables collaboration of domain knowledge between business, technical, and data steward users. Business Data Network can be used as a single entry point for all data consumers to better understand their data. It consists of a web user interface that documents business terms and their associated rules, jobs, applications, data, documentation, and other information. Technical users use the network to document information about tables and columns that implement the business terminology, to relate jobs and other information to terms, to share knowledge about data transformations, and as a data dictionary to describe details of data models and other data-related information. Data stewards can view data from a business standpoint to better visualize problem areas by domain in order to identify and fix data issues more effectively. Figure 20 shows the Business Data Network main view.

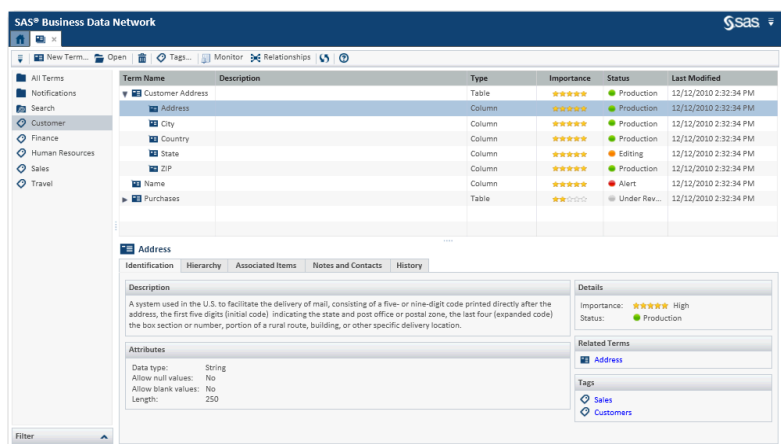


Figure 20: Business Data Network Main View

Typically, users who understand their data model or business terminology would provide the initial information in the network. These users might also attach documents or rules that describe each term. There are also import and export features that support the ability to quickly populate and exchange information. Other users add additional information related to the term such as jobs that are used to

modify the term or physical tables that might implement the term. The network is fully integrated with impact analysis to help you understand how your objects interrelate. Figure 21 shows a typical diagram of relationships stored in the network.

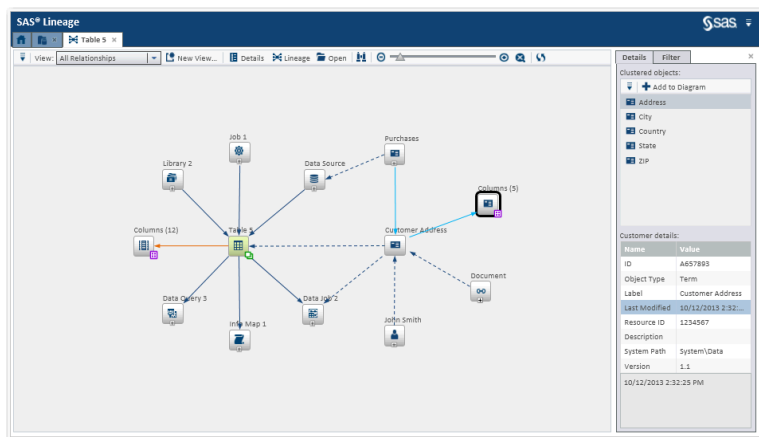


Figure 21: Example of the Relationships View Showing Business and Technical Metadata

There are a number of new features in the latest release of SAS Business Data Network. The user interface now supports roles, capabilities and security for terms, and term attributes. The roles and capabilities are fully customizable to match your site requirements. Figure 22 displays the many roles, capabilities, and security settings available.

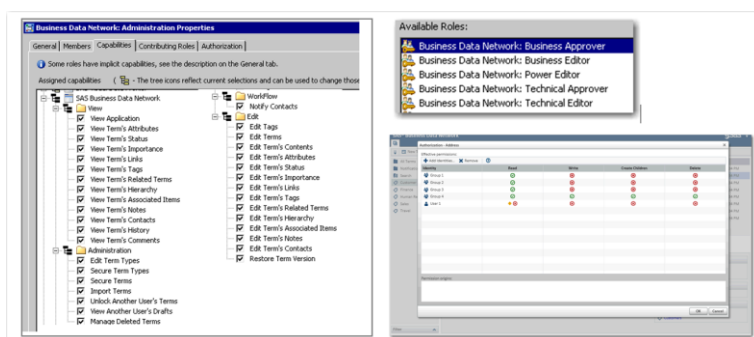


Figure 22: Examples of Security and Role Features Available in Business Data Network

Integration with the SAS workflow is also now available. Users can send terms into the workflow for review and approval before publishing. There are several default workflows available that you can customize, or you can create your own workflows to match your business needs. Figure 23 is an example of a workflow process applied to Business Data Network. The network reads the workflow state and customizes the UI to display buttons to help you interact with each workflow state. The status is also shown at each step in the workflow.

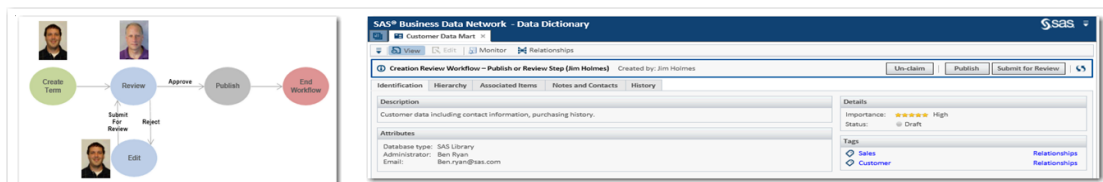


Figure 23: An Example of Using Workflow in Business Data Network

Users can quickly see workflow tasks that are waiting on their input in the task manager view in the data management console and in views in the network. There are also a number of quick actions available for users such as being able to update the workflow for multiple terms together.

Different workflows can be used for different actions in the network. For example, you can have one

workflow that you want to use when creating terms and another workflow when deleting terms. You can also tie different workflows to different term groups; for example, you can have one workflow when working with supplier information, and a different workflow for working with your data dictionary tables and columns.

There are number of collaboration features integrated into the network. As a term progresses through the workflow, each user can add status or comments. A notes section also allows authorized users to collaborate on any term. From a compliance standpoint, every action that is taken on any term is logged in the SAS audit service, which can be retrieved for reporting or compliance purposes. In addition, history is retained on all published content. Authorized users can view the changes that have occurred over time to their terms or term content, as well as restore or retire content in the network. Figure 24 illustrates collaboration between interested parties of terms and some of the available versioning features.

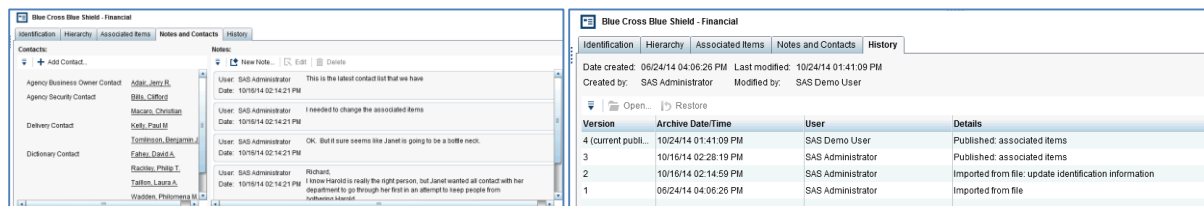


Figure 24: User Collaboration and History Example

Another important new feature is support for multiple, customized term templates. Administrators are now able to create templates with custom attributes for terms and term hierarchies. For example, you might have a set of terms that represent the tables and columns in your data dictionary. You can create a table template with the information that you want to use to describe tables in your system, and a different column template with the information that you want to capture about columns. You can have any number of custom templates that match the information that you want to capture in your terms. You also have options to specify whether the template should be inherited in a hierarchy of terms, whether attributes are required, which can be useful if you want to enforce the collection of standard information for every term that is built from the template, and default values for attributes of a term. Most of the attributes of a term are now fully customizable via the term template. Figure 25 shows some of the features available for term templates.

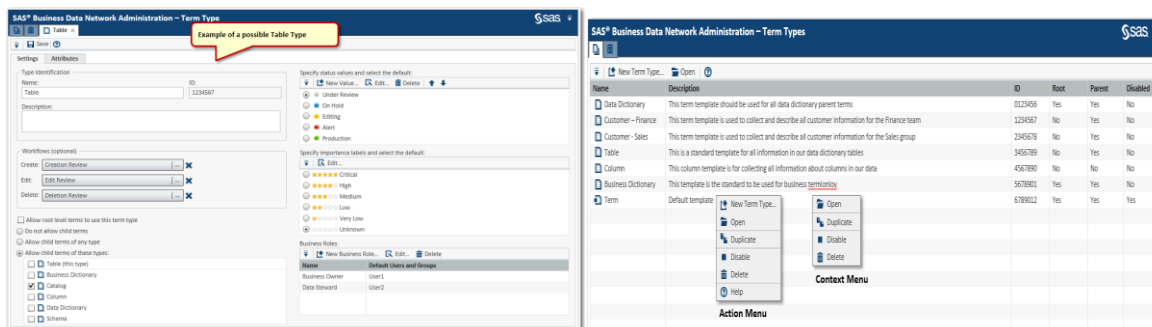


Figure 25: Customizing Term Template Examples in Business Data Network

LINEAGE AND IMPACT ANALYSIS

A number of important new features have been added to support lineage and impact analysis. SAS has created a shared store for all relationship information called the SAS relationship service. Most SAS products and object types are now integrated into the SAS relationship service. The relationship's web viewer supports different views for displaying information stored in the service. Figure 26 is an example of the Impact Data Flow view. There are also views for all Relationships and for Data Governance.

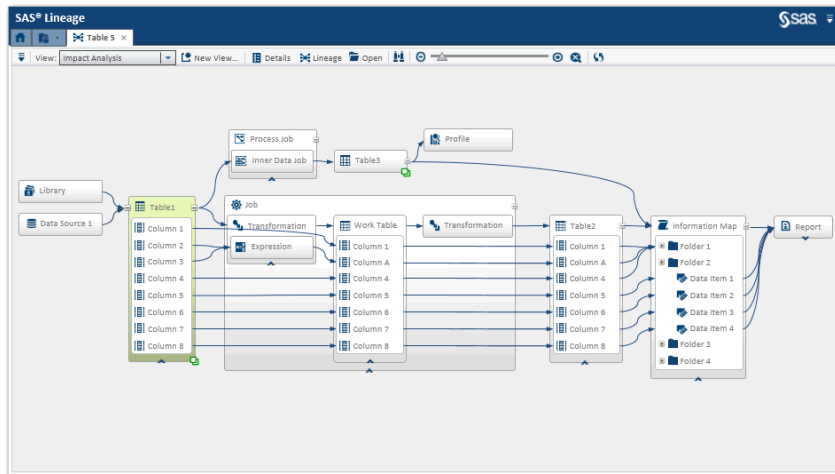


Figure 26: Lineage Viewer Showing Table, Job, and Column Relationships

You can create your own views by using the filtering capabilities of the viewer. This can help you subset the information to only the objects and relationships that you want to see. In addition, there are helpful features such as grouping node sets, which enable you to expand on demand, and an overview window with details of objects. Figure 27 illustrates some of these new features.

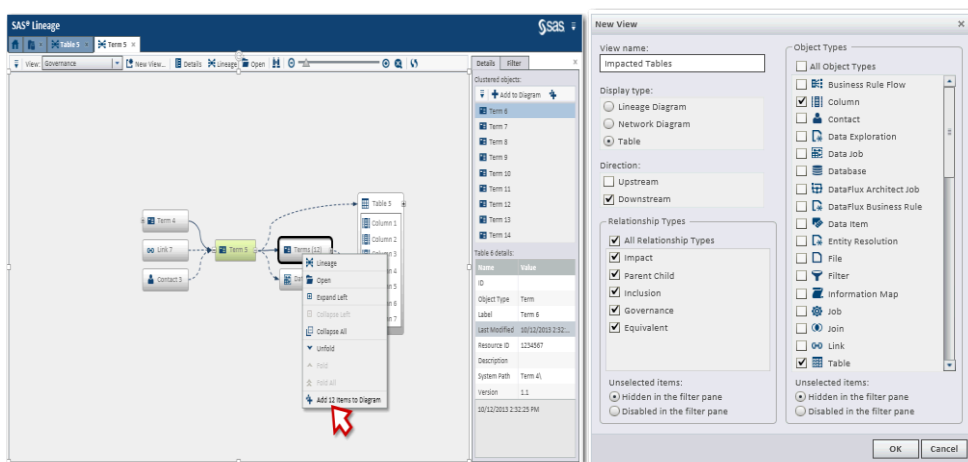


Figure 27: Custom Views, Collapsing and Expanding Multiple Nodes, and Node Details in Lineage Viewer

There are a number of enhancements to lineage and the underlying relationship service that supports the lineage content. A key enhancement is the ability to import content from third-party metadata sources by using the Meta Integration Bridge technology. The import exchange is available for hundreds of third-party sources, including vendors such as SAP BusinessObjects, ERwin Data Modeler, and many other tools. In past releases, import was limited to relational types of metadata, but this restriction has now been lifted. The content types are unlimited, in that all object types from all models can be imported.

Metadata exchange with third-party sources is available via a new command line utility that comes with a SAS installation. Here is an example of the launcher program and its installation location:

```
Launcher name: sas-metabridge-relationship-loader
Install location: !SASHOME\SASMetadataBridges\4.1\tools
```

The user supplies logon information to the relationship service and an administrative user ID and password to perform the import. Other options available during the import include the ability to mark 1-N objects as equivalent to each other so that the viewer shows the object as a single group object instead of separate objects; the ability to specify vendor options when using a particular vendor bridge; and the

ability to schedule the import to occur on a predetermined schedule for better support of synchronizing content. Here is a partial list of the options available in the utility:

```
usage: sas-metabridge-relationship-loader [options...]
```

Example options:

-?, --help	Print help information.
-bridgeDirectory	The location of the SAS Metadata Bridges if different than default location.
-bridgeList	Request the list of available licensed bridges.
-bridgeOptions	Customize the import
-clean	Clean relationships from the third party source.
-loadRelationships	Load relationships from a third party source

...and others

Figure 28 and Figure 29 show imported content from external metadata sources using the bridges.

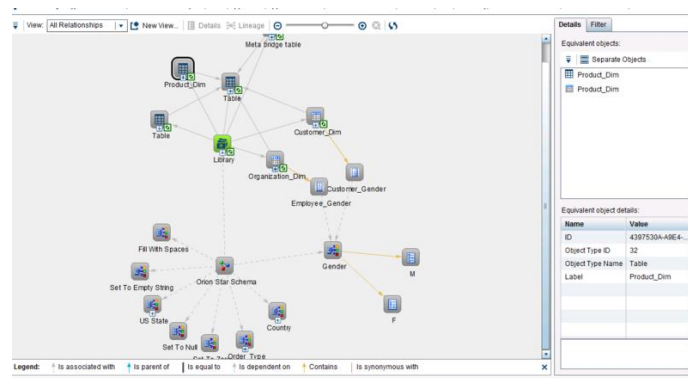


Figure 28: Example of Content Imported from an External Metadata Source

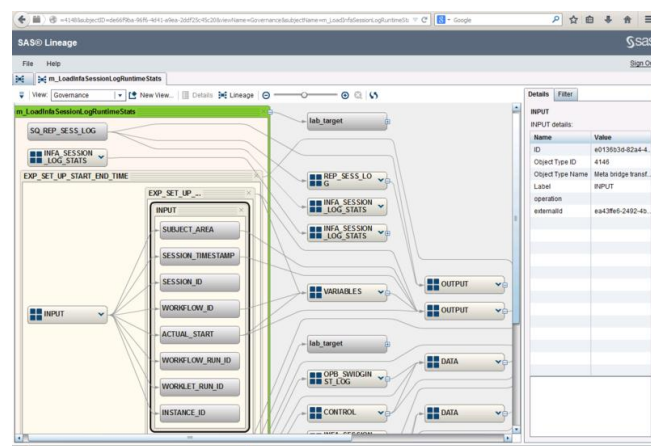


Figure 29: Example of the Governance View from an Import

There are also a number of reporting interfaces that can be used to extract content from the relationship service. The relationship reporter command line utility enables you to extract content stored in the relationship service and output the results to a CSV or text file. For example, you can use the reporter utility to write lineage, impact, equivalency, filters, and other relationship information into a file. Some useful options with the reporter include the ability to view relationships before or after one or more objects; filter on relationship types or object types; filter based on date modified, or a range of dates; and various search options. Here is some sample usage syntax of the utility:

```
usage: sas-relationship-loader [options...] object-paths...
```

Example options:

-before <date> Select only the objects that have been modified before this date.

-excludeSubTypes Excludes any subtypes of the content types

-folder <folder-path> Only objects in this folder will be processed.

-nameMatchType <operator> Search operator used when filtering objects

-types <types> Filter to only these types

...and others

Figure 30 shows sample output from the reporter utility showing the CARS table subject and its associated downstream dependencies.

```
"/BITesting/DataSources/BIDdata/CARS" (Table)
<-- Depends: "/BITesting/DataSources/Maps/cars" (Information map (relational))
<-- Depends: "/BITesting/DataSources/Maps/carswithFilterChevy" (Information map (relational))

"/BITesting/DataSources/SASCubeSources/CARS" (Table)
<-- Depends: "/BITesting/DataSources/Cubes/BID_MDDBCARS" (Cube)
<-- Depends: "/BITesting/DataSources/Maps/OLAPMaps/MDDBCARS" (Information map (OLAP))

"/STPSanity/CARS" (Table)
<-- Depends: "Parameters" (Prompt group)
<-- Includes: "/STPSanity/casCadEBI-stream" (Stored process)
<-- Depends: "Parameters" (Prompt group)
<-- Includes: "/STPSanity/casCadEBI_pak" (Stored process)
<-- Depends: "/User Folders/sasdemo/My Folder/1_test_report.srx" (SAS report)
```

Figure 30: Example Output from the Reporter Utility

MASTER DATA MANAGEMENT

When consolidating data from diverse source systems, there is often a need to select the best record out of all possible records that represents the same data so that you can pass one version of the data to downstream jobs and reports. For example, you might have multiple diverse customer records coming in from various source systems and you need to be able to consolidate on a single, best record with standardized values for fields such as address and phone number.

Figure 31 is an example of a customer best record selected from three different source records.

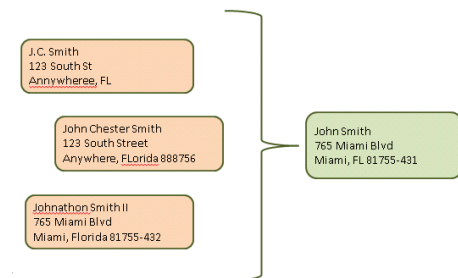


Figure 31: Customer Best Record Selection

SAS® Master Data Management automates the process of selecting the best records from source data. Master Data Management works through a technique called “clustering,” which is difficult to do with traditional SQL transformation logic. The technology supports sophisticated techniques such as probabilistic matching that are able to pull out the best record based on analytical processes. For example, if two records are similar to each other, the technology can create a score based on rules as to how likely or how probable the records match. The best record is automatically selected for you into a single, cleansed, and de-duplicated data record. Figure 32 is an example of a view in Master Data Management showing a set of incoming records and the selected best record.

The screenshot shows the 'MASTER DATA MANAGEMENT ENTITY' window. It has tabs for 'Records', 'Hierarchies', 'Relationship Diagram', and 'History'. The 'Records' tab is active, displaying a table with columns: MDM Entity ID, Source System, Created, Retired, and Full Name. Below the table, there is a detailed view for the selected record (ID: 343).

MDM Entity ID	Source System	Created	Retired	Full Name
10860	Best Record	11/30/2012		Bill Evans
343	ERP	11/30/2012		Will Evans
34123	CRP	10/30/2012		Bill Evans
7656	MD Manager	9/15/2012		Bill Evans
5464	MD Manager	10/30/2012		Bill Evans
744	AARP	11/30/2012		Bill Evans
56	FARC	11/30/2012		Will Evans

Below the table, the detailed view for ID: 343 is shown:

- ID: 343
- System: ERP
- Name: Will Evans
- Email: Bill.Evans@comddot.com

Figure 32: Example Best Record View

SAS Master Data Management has added a number of new features. Data stewards can now view and control how and under what circumstances master data records should be consolidated with contributing source systems for consistency purposes. You can also execute jobs when new surviving records are created, such as pushing data to a reporting environment or synchronizing data in the other systems. Figure 33 is an example of the Source Systems activity viewer that shows source activity and allows you to call various processes when changes occur.

The screenshot shows the 'SAS MDM - Source Systems' window. It has a 'File' menu and a 'Help' button. On the left, there is a tree view of source systems: All Source Systems, Best Record, MD Manager, ERP, Marketing, Orion, Call Center UK, Call Center US, and CRUI. The main area displays a table titled 'Add/Update History: All Source Systems' with columns: Start Date, Elapsed Time, Source System, Entity, User, Records, Added to Hub, and Send for. The table shows a list of activities with their respective dates, times, and record counts.

Start Date	Elapsed Time	Source System	Entity	User	Records	Added to Hub	Send for
02/27/15 07:47:56 AM	00:00:09	Best Record	Individual	Ron Agresta	1	1	0
02/25/15 01:47:12 PM	00:04:10	Marketing	Company	Ron Agresta	14341	14341	0
02/25/15 01:45:56 PM	00:00:49	Marketing	Part	Ron Agresta	2481	2481	0
02/25/15 01:37:04 PM	00:03:31	ERP	Company	Ron Agresta	13629	13629	1
02/25/15 01:36:13 PM	00:00:34	ERP	Part	Ron Agresta	2364	2362	2
02/25/15 01:19:36 PM	00:06:32	ERP	Individual	Ron Agresta	13779	13769	1
02/25/15 12:50:45 PM	00:05:59	Marketing	Individual	Ron Agresta	15244	15244	0
02/25/15 12:43:43 PM	00:04:53	Orion	Individual	Ron Agresta	15258	15258	0
02/25/15 12:41:35 PM	00:01:32	Call Center UK	Individual	Ron Agresta	2556	2556	0
02/19/15 04:19:28 PM	00:03:22	Orion	Individual	Ron Agresta	5000	5000	0
02/19/15 04:15:59 PM	00:03:32	Call Center UK	Individual	Ron Agresta	5000	5000	0
02/19/15 04:14:45 PM	00:03:38	Orion	Individual	Ron Agresta	0	0	0
02/19/15 04:14:36 PM	00:04:05	Marketing	Individual	Ron Agresta	5000	5000	0
02/19/15 04:14:23 PM	00:04:15	ERP	Individual	Ron Agresta	5000	5000	0
02/04/15 02:25:24 PM	00:00:26	ERP	Individual	Ron Agresta	500	500	0
02/04/15 02:15:13 PM	00:00:08	Best Record	Individual	Ron Agresta	1	1	0

Figure 33: Source Systems Activity Viewer

Several other new features include cross-field matching and custom relationship attributes. Cross-field matching enables you to build matching rules that can include other related columns. For example, you can design a rule that matches on name and then across two different phone number fields, the rule looks for similar values. Relationships between master record fields have been expanded to support custom attributes. For example, you might choose to add "Start Date" to an "Employed by" relationship type. Relationship attributes can be viewed and edited in the entity editor or in the relationship diagram panel.

SAS Master Data Management now ships some useful SAS® Visual Analytics reports to gain more insight into their master data. You can see useful information such as batch load statistics, record counts by entity type and source system, and many other aspects of the data in the Master Data Management database. Figure 34 is an example report.

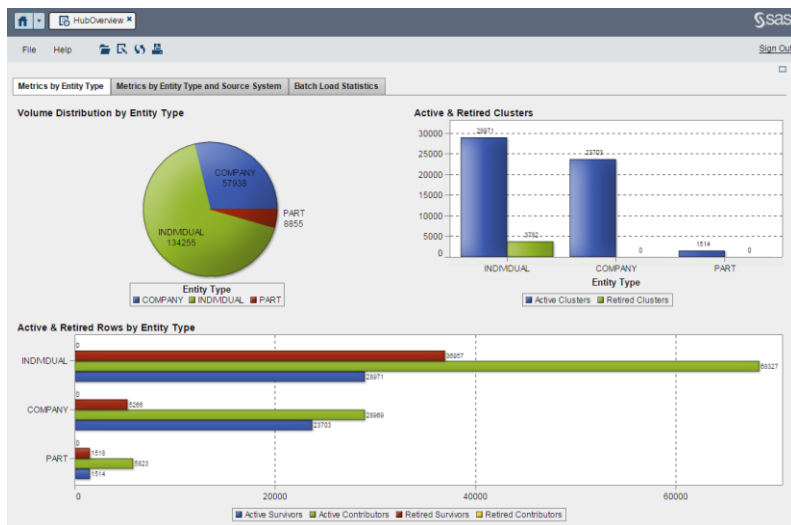


Figure 34: Visual Analytics Sample Report of Master Data Management Data

A number of new features have been added to support workflow and remediation of data when working with data in the Master Data Management hub. With the latest release, you can now send data records into a workflow for users to review, approve, or reject changes, and receive notifications when you or others in your workgroup are required to take some action on the data. For example, you might want to review the best record that was selected or adjust the data first and have it reviewed by your data quality experts before adding it to the hub. Figure 35 is an example of the Data Remediation view.

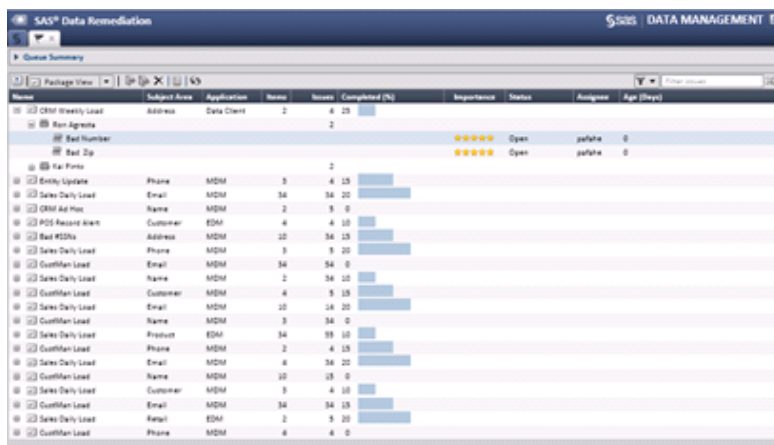


Figure 35: Example of the Data Remediation View

You can group remediation issues by various categories such as Issue type, Importance, and other criteria. Hierarchical views of the data are also available. You can also see that a data quality issue has been logged against a Master Data Management record and you can drill into that record for details. SAS® Data Management Console, as shown in Figure 36, displays many of these details at a glance.

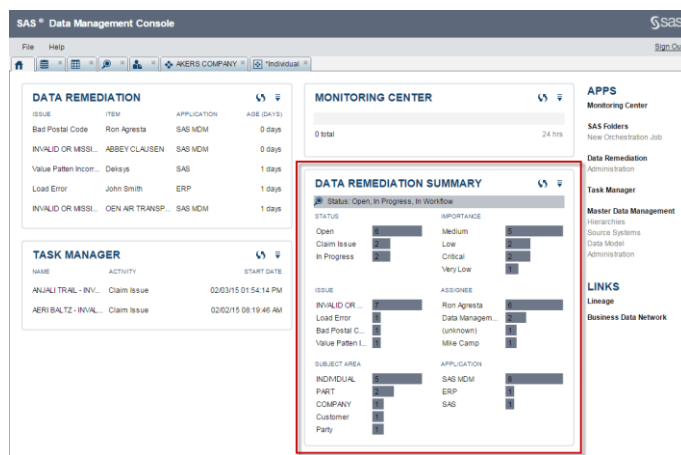


Figure 36: Remediation Details in SAS Data Management Console

CONCLUSION

The latest releases of SAS Data Integration Studio and DataFlux Data Management Platform provide enhancements to help both data warehouse developers and data integration specialists carry out data-oriented processes more efficiently and with greater control and flexibility. Major focus areas for the releases include features for job performance and manageability, enhanced metadata management capabilities, and new features in support of big data. Customers will find many reasons to upgrade to the latest version of SAS Data Management.

RECOMMENDED READING

- SAS® Enterprise Data Management and Integration Discussion Forum. Available at http://communities.sas.com/community/sas_enterprise_data_management_integration.
- SAS® Data Loader for Hadoop Discussion Forum. Available at <https://communities.sas.com/groups/sas-data-loader>.
- Rineer, B. 2015. "Garbage In, Gourmet Out: How to Leverage the Power of the SAS® Quality Knowledge Base." *Proceedings of the SAS Global Forum 2015 Conference*. Cary, NC: SAS Institute Inc. Available at <http://support.sas.com/resources/papers/proceedings15/SAS1852-2015.pdf>.
- Agresta, R. 2015. "Master Data and Command Results: Combine MDM with SAS Analytics for Improved Insights." *Proceedings of the SAS Global Forum 2015 Conference*. Cary, NC: SAS Institute Inc. Available at <http://support.sas.com/resources/papers/proceedings15/SAS1822-2015.pdf>.
- McIntosh, Liz, et al. 2014. "Understanding Change in the Enterprise." *Proceedings of the SAS Global Forum 2014 Conference*. Cary, NC: SAS Institute Inc. Available at <http://support.sas.com/resources/papers/proceedings14/SAS396-2014.pdf>.
- Rausch, Nancy, et al. 2014. "What's New in SAS Data Management." *Proceedings of the SAS Global Forum 2014 Conference*. Cary, NC: SAS Institute Inc. Available at <http://support.sas.com/resources/papers/proceedings14/SAS034-2014.pdf>.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author:

Nancy Rausch
100 SAS Campus Drive
Cary, NC 27513
SAS Institute Inc.
Nancy.Rausch@sas.com

<http://www.sas.com>

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.