

Data Management Studio

Dave Smith, SAS UK



THE
POWER
TO KNOW®

Agenda

What is Data Management Studio?

A quick tour

- Data Management Methodology
- Possible Applications in Life Sciences

Data Management Studio

Single client interface for all Data Quality needs

- Evolution of DataFlux dfPower Studio

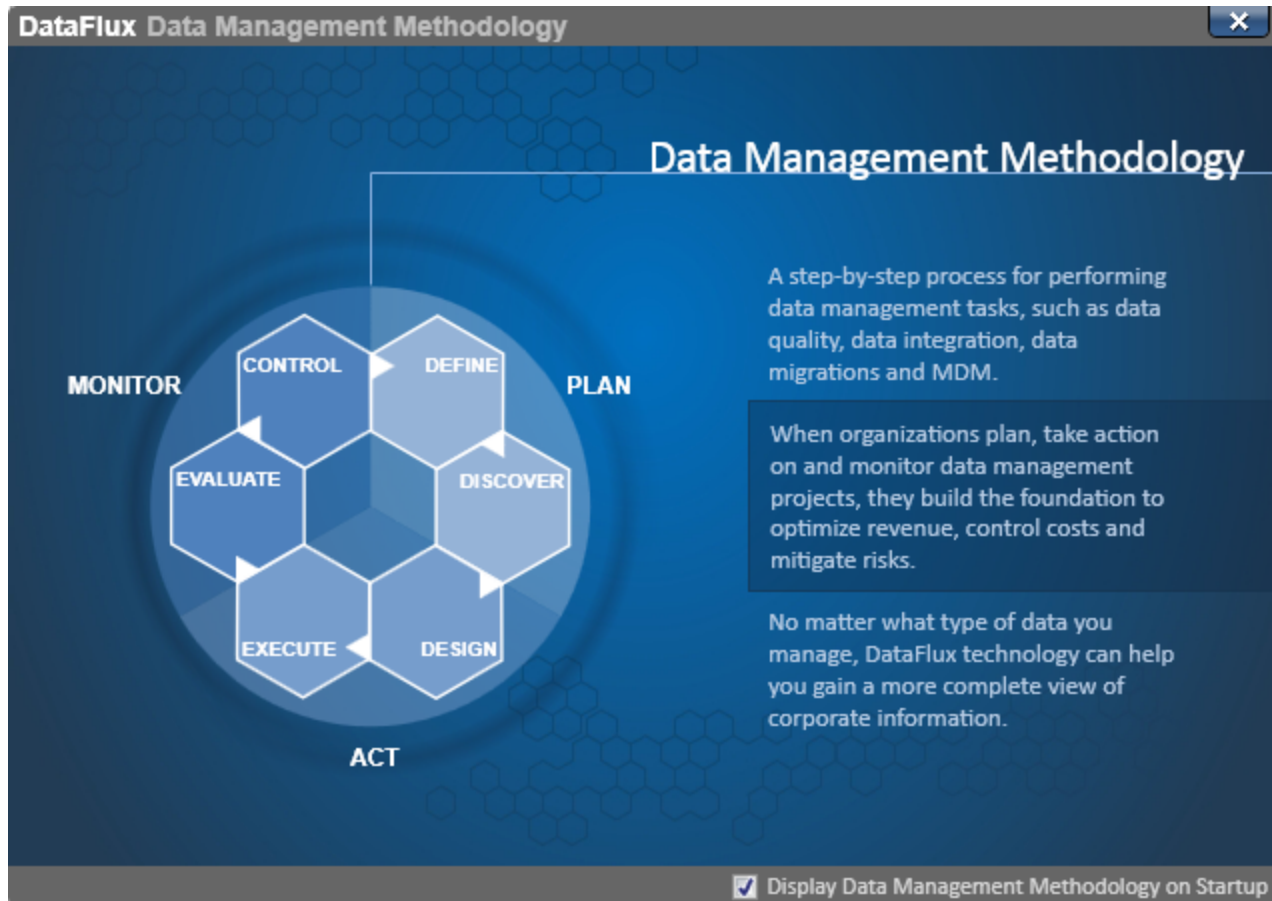
Provides a platform for other solutions

- Eg qMDM

Built around the Data Management Methodology

Integrated with SAS 9.3

Data Management Methodology



Define



DataFlux Data Management Methodology

Back

Connect to Data

Explore Data

Define Business Rules

Build Schemes

Define: Explore Data

Use data explorations to identify data redundancies and extract and organize metadata from multiple sources. Relationships between metadata can be identified and cataloged into types of data by specified business data types and processes. A data exploration is a job that is configured to read data from databases and categorize the fields in the selected tables into categories that have been predefined in the Quality Knowledge Base (QKB).

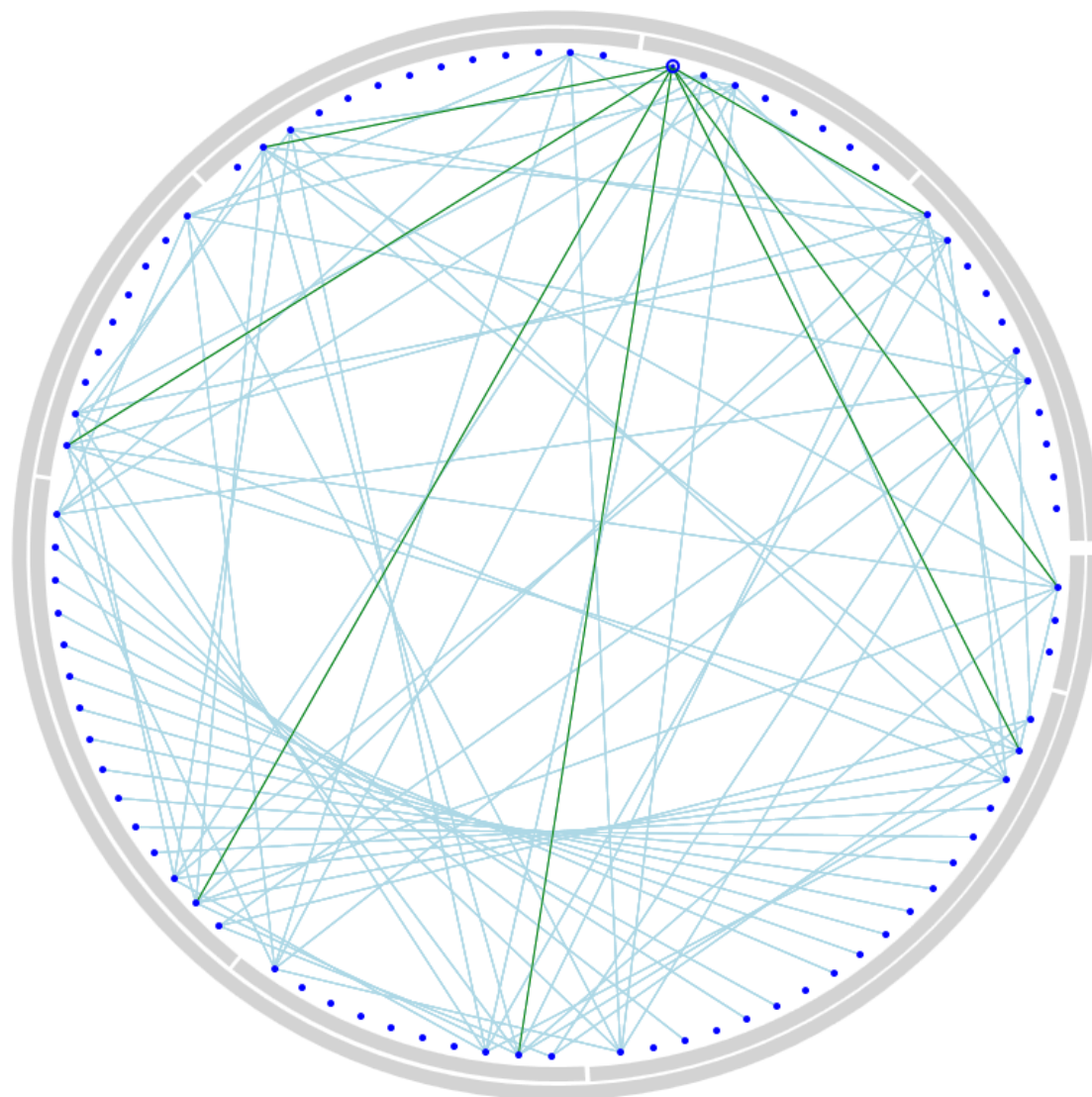
[Click here to learn more...](#)

☒ Display Data Management Methodology on Startup

The screenshot shows a software window titled "DataFlux Data Management Methodology". On the left is a sidebar with four main sections: "Connect to Data" (showing database icons), "Explore Data" (highlighted with a dashed border and a magnifying glass icon), "Define Business Rules" (showing a "Step2: Edit the check details" dialog with rules like "State is null", "State is not uppercase", "Length of State is greater than 2", and "Length of State is less than 2"), and "Build Schemes" (showing a mapping table). The "Explore Data" section on the right contains a descriptive paragraph and a "Click here to learn more..." link. At the bottom right, there is a checkbox labeled "Display Data Management Methodology on Startup" which is checked.

Data	Standard
'S	%
am	AM
c/o	C/O
hq	HQ

Nicardipine Raw



Data Sample

Field: subjId

Table: FUBE

Database: Nicardipine Raw

Sample:

11001
11002
11003
11004
11005
11006
11007
11008
11009
11010
11011
11012
11013
11014
11015
11016
11017
11018
11019
11020
11021
11022
11023
11024
11025
11026
11027
11029
11030
11031
11032
11033
11034
11035
11036
11037
11038
11039
11040
11041
11042
11043
11044
11045
11046
11047
11048
11049
11050
11051
21001

[See more data](#)

Notes

Collections

Profile Task

Discover

DataFlux Data Management Methodology

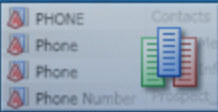
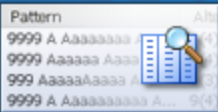
 Back

Explore Data

Profile Data

Organize Fields

Query Data



Discover: Profile Data

Data profiling encompasses discovery and audit activities that help you assess the composition, organization, and quality of databases. Thus, a typical data profiling process helps you to recognize patterns, identify missing information in the data, and calculate frequency and basic statistics. Data profiling can also aid in identifying redundant data across tables and cross-column dependencies.

[Click here to learn more...](#)

☒ Display Data Management Methodology on Startup

Tables

Nicardipine Raw
BASE
SAS
ADVERSE
CHEM
CONMED
DEMO
ECG
FUBE
HEMA
VITAL
CENTER
DIAS
HRTRATE
SYST
TRTGRP
VISIT
chg_dbp
chg_hr
chg_sbp
studyid
subjid

DIAS

Table: VITAL Schema: SAS Catalog: BASE Data Source: Nicardipine Raw

Column Profiling

Frequency Distribution

Pattern Frequency Distribution

Percentiles

Outliers

Primary Key/Foreign Key Analysis

Notes

Metric Name	Metric Value
Ordinal Position	5
Count	3604
Null Count	1099
Percent Null	30.5
Blank Count	(not applicable)
Minimum Value	31
Maximum Value	122
Mode	70
Pattern Count	(not applicable)
Unique Count	79
Uniqueness	3.15
Primary Key Candidate	no
Data Type	DOUBLE
Data Length	15 chars
Actual Type	double
Minimum Length	(not applicable)
Maximum Length	(not applicable)
Mean	76.004391
Median	75
Non-null Count	2505
Nullable	YES
Decimal Places	0
Standard Deviation	12.830845
Standard Error	0.256361

Tables

Collections

Redundant Data Analyses

Tables

Nicardipine Raw
BASE
SAS
ADVERSE
CHEM
CONMED
DEMO
ECG
FUBE
HEMA
VITAL
CENTER
DIAS
HRTRATE
SYST
TRTGRP
VISIT
chg_dbp
chg_hr
chg_sbp
studyid
subjid

VITAL

Catalog: BASE Schema: SAS Data Source: Nicardipine Raw

Standard Metrics Custom Metrics **Business Rules** Alerts Visualizations Notes

Name	Inputs	Number of Violations	Percentage
Systolic BP less than Diastolic BP	DIAS, SYST	0	0.00
Heart Rate between 40 and 180	HRTRATE	2	0.06

Tables

Collections

Redundant Data Analyses

Tables

Nicardipine Raw
BASE
SAS
ADVERSE
CHEM
CONMED
DEMO
ECG
FUBE
HEMA
VITAL
CENTER
DIAS
HRRATE
SYST
TRTGRP
VISIT
chg_dbp
chg_hr
chg_sbp
studyid
subjid

VITAL

Catalog: BASE Schema: SAS Data Source: Nicardipine Raw

Standard Metrics Custom Metrics Business Rules Alerts Visualizations Notes

Name	Inputs	Number of Violations	Percentage
Systolic BP less than Diastolic BP	DIAS, SYST	0	0.00
Heart Rate between 40 and 180	HRRATE	2	0.06

Business Rule Violation Drill Through

CENTER	DIAS	HRRATE	SYST	TRTGRP
04	84	16	164	NIC .15
04	58	28	146	Placebo

Close

Export...

Print...

Tables

Collections

Redundant Data Analyses

Redundant Data Analyses

Subject ID

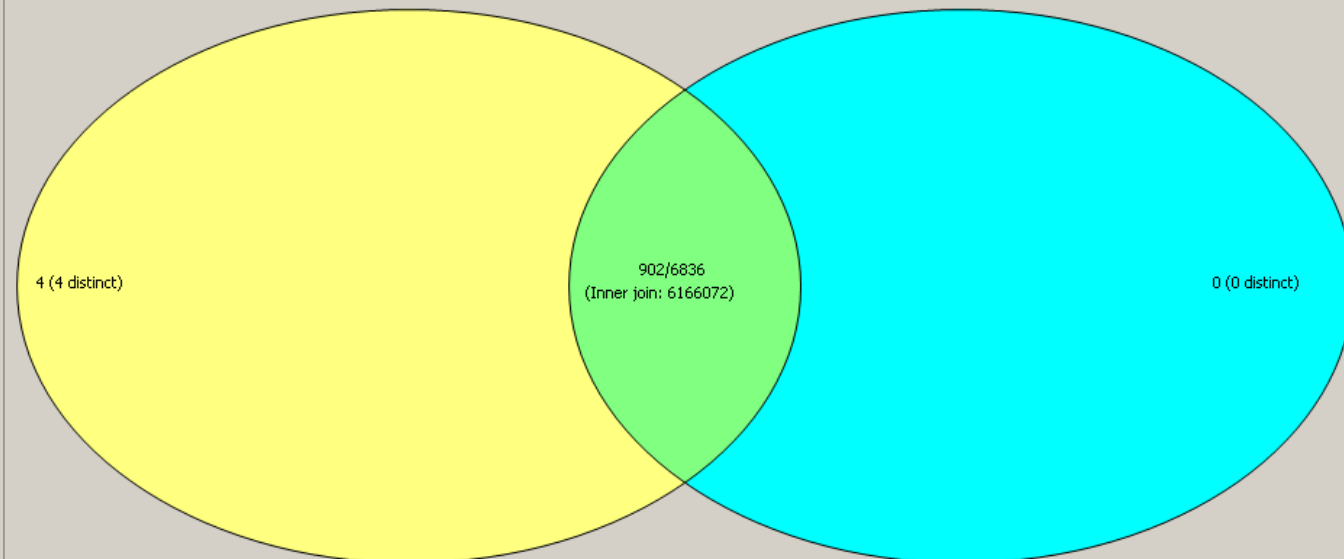
Subject ID

	subjid	subjid	subjid	subjid	subjid	subjid	subjid	subjid
subjid	X	0	0	0	0	0	0	0
subjid		X	52	0	0	3	25	0
subjid			X	2	2	1	0	4
subjid				X	370	159	2978	632
subjid					X	33	540	120
subjid						X	345	68
subjid							X	4
subjid								X

Grid data: ☐ Left outliers ☐ Common ☒ Right outliers

subjid
DEMO
Nicardipine Raw.BASE.SAS

subjid
ADVERSE
Nicardipine Raw.BASE.SAS



Tables

Collections

Redundant Data Analyses

Redundant Data Analyses

Subject ID

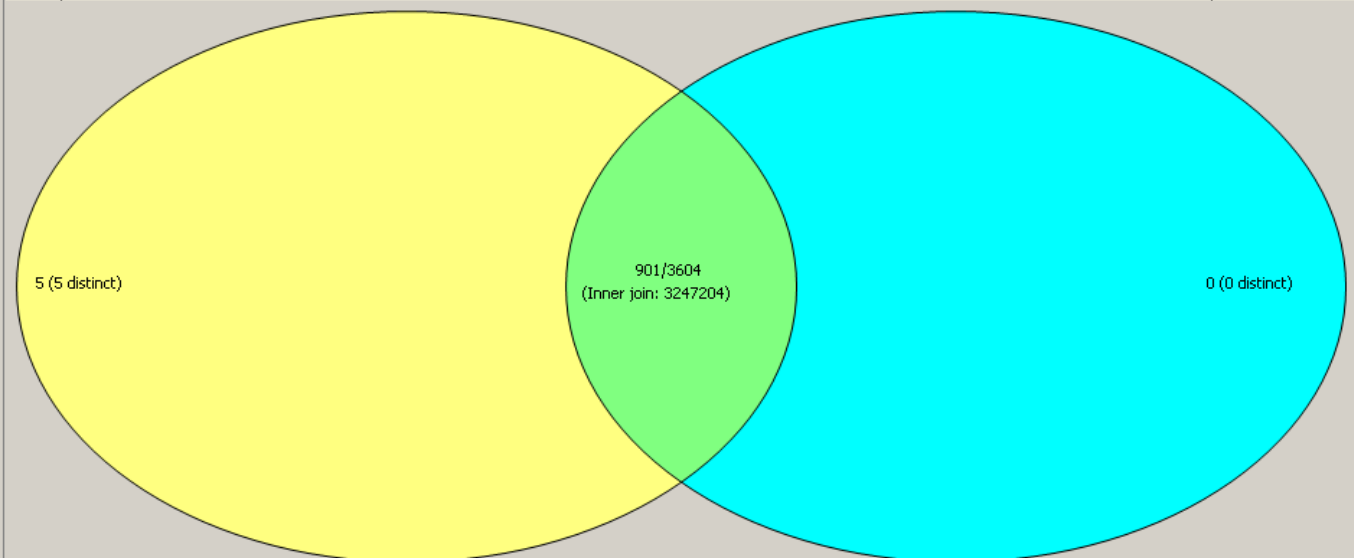
Subject ID

	subjid	subjid	subjid	subjid	subjid	subjid	subjid	subjid
subjid	X	4	1	163	35	19	1	5
subjid		X	7	872	175	154	7	1
subjid			X	7750	1120	1001	0	71
subjid				X	67	59	2	0
subjid					X	47	2	0
subjid						X	1	3
subjid							X	32
subjid								X

Grid data: ☒ Left outliers ☐ Common ☐ Right outliers

subjid
DEMO
Nicardipine Raw.BASE.SAS

subjid
VITAL
Nicardipine Raw.BASE.SAS



Tables

Collections

Redundant Data Analyses

Redundant Data Analyses

Subject ID

Subject ID

	subjid	subjid	subjid	subjid	subjid	subjid	subjid	subjid
subjid	X	4	1	163	35	19	1	5
subjid		X	7	872	175	154	7	1
subjid			X	7750	1120	1001	0	71
subjid				X	67	59	2	0
subjid					X	47	2	0
subjid						X	1	3
subjid							X	32
s								X

Redundant Data Analysis Primary Field Va...

Value	Count
191002	1
51000	1
312001	1
451000	1
392013	1

Export...

Print...

Close

Grid data: ☒ Left outliers ☐ Common ☐ Right outliers

subjid

VITAL

Nicardipine Raw.BASE.SAS

5 (5 distinct)

901/3604
(Inner join: 3247204)

0 (0 distinct)

Tables

Collections

Redundant Data Analyses

Redundant Data Analyses

Subject ID

Subject ID

	subjid	subjid	subjid	subjid	subjid	subjid	subjid	subjid
subjid	X	4	1	163	35	19	1	5
subjid		X	7	872	175	154	7	1
subjid			X	7750	1120	1001	0	71

Redundant Data Analysis Primary Field Value Drill Through

subjid	siteid	invname	WEIGHT	HEIGHT
191002		191A	93	185

Close

Export...

Print...

☐ Outliers
 ☐ Common
 ☐ Right outliers

subjid

VITAL
Nicardipine Raw.BASE.SAS

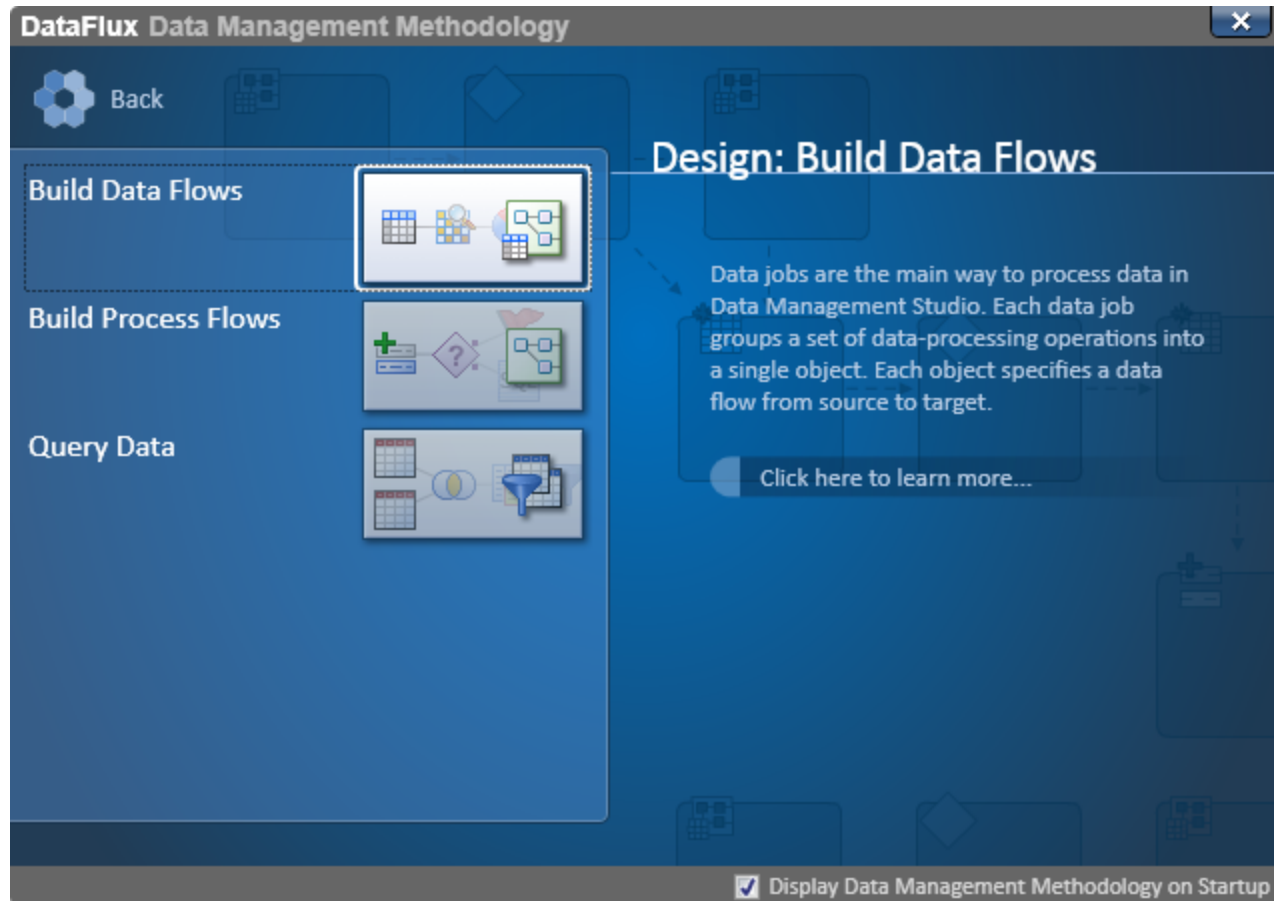
Tables

Collections

Redundant Data Analyses

0 (0 distinct)

Design



[illegible]

Add To Scheme

[illegible]

Add...

Edit...

Delete

Options...

Data Flows

DataFlux Data Management Studio 2.1

Standardise State * x

File Edit View Actions Tools Window Help

Data Job Properties Variables Log

Nodes

- Data Inputs
 - Data Source
 - SQL Query
 - Text File Input
 - Fixed Width File Input
 - External Data Provider
 - Table Metadata
 - COBOL Copybook
 - Job Specific Data
 - XML Input
 - Work Table Reader
- Data Outputs
- Data Integration
- Folders
- Accelerators

Diagram:

```

graph LR
    DS1[Data Source 1  
Data Source  
DSN: DSN=DFSAMPLE...  
✓] --> B1[Branch 1  
Branch  
Completed successfully  
✓]
    B1 --> BS[Bespoke Standardization  
Standardization  
3276 row(s) processed  
✓]
    B1 --> DS2[Default Standardization  
Standardization  
3276 row(s) processed  
✓]
    BS --> FS1[Find State over two characters...  
Expression  
Completed successfully  
✓]
    DS2 --> FS2[Find State over 2 Characters...  
Expression  
Completed successfully  
✓]
  
```

Log

Row	Node Name	Contained In	Instance	Node ID	Node Type	Status	Run Time (minutes)	Start Time	End Time
3	Find State over 2 Characters...			6	Expression	Completed successfully			
4	Bespoke Standardization			8	Standardization	3276 row(s) processed			
5	Branch 1			9	Branch	Completed successfully			
6	Find State over two characters...			10	Expression	Completed successfully			

Job Status: ✓ Completed Successfully Run time: 0 hours 0 minutes 0.473 seconds

Address Verification

DataFlux Data Management Studio 2.1

Standardise State | dfsample_verify_geocode_county

File Edit View Actions Tools Window Help

Data Job Properties Variables Log

Nodes

- Quality
- Enrichment
 - Address Verification (US/Canada)
 - Address Verification (QAS)
 - Address Verification (World)
 - Geocoding
 - Rooftop Geocoding (Tele Atlas)
 - US City/State/Zip Validation
 - County
 - US City/State/Zip Lookup
 - Phone
 - Area Code
 - Canadian Postal Code Lookup
- Folders
- Accelerators

Workflow:

```
graph LR; DS1[Data Source 1] --> AV1[Address Verification 1] --> G1[Geocoding 1] --> C1[County 1] --> DT1[Data Target (Create) 1]
```

Log

Row	Node Name	Contained In	Instance	Node ID	Node Type	Status	Run Time (minutes)	Start Time	End Time
-----	-----------	--------------	----------	---------	-----------	--------	--------------------	------------	----------

Job Status:

Execute

DataFlux Data Management Methodology

Back

Build Data Flows

Build Process Flows

Query Data

Resolve Entities

Run Jobs Remotely

Execute: Resolve Entities

Entity resolution is the process of merging multiple files (or duplicate records within a single file) in such a way that records referring to the same physical object are treated as a single record. Records are matched based on the information that they have in common. The records that you merge appear to be different but can actually refer to the same person or thing.

[Click here to learn more...](#)

☒ Display Data Management Methodology on Startup

Standardise State Entity Resolution on Sample X

File Edit View Actions Tools Window Help

Data Job Properties Variables Log

Nodes

- Data Sorting
- Data Joining
- Data Joining (Non-Key)
- Data Union
- SQL Lookup
- SQL Execute
- Parameterized SQL Query

Quality

Enrichment

Entity Resolution

- Match Codes
- Match Codes (Parsed)
- Clustering
- Cluster Update

Folders

Accelerators

Data Source

Help...

```

graph LR
    DS1[Data Source 1  
Data Source  
DSN: DataFlux S...] --> MC1[Match Codes 1  
Match Codes  
3276 row(s) processed]
    MC1 --> C1[Clustering 1  
Clustering  
Sort setting: actual (specified)]
    C1 --> SRI[Surviving Record Identification  
Surviving Record Identification  
3276 rows processed]
  
```

Basic Properties Node Connections Preview Log

ID	COMPANY	CONTACT	ADDRESS	CITY	STATE
1	First Merit Bank	James E. Briggs	19 East Broad Street	Saint Louis	Missouri
10	DataFlux Corporation	Bob Brauer	6512 Six Forks Road - 404B	Raleigh	North Carolina
100	KAISER HOSPITAL	LUTHER BAKER	3560 E 116TH ST	Cleveland	CA
1000	Transamerica Life Insurance	Irene Greaves	555 W Fifth St	S Northfield	OH
1001	Transamerica Life Ins & Annuity Co	Rob Drain	7718 Elder Way	Clementson	OH
1002	Transamerica Occidental	Tonia Gerstner	1018 N Hayworth Ave	Urb Canas	OH
1003	Transamerica Life Insurance	Joshua Hodgekin	244 Evans	Old Mission	OH
1004	Transamerica Occidental Life	Nancy Weinstock	2134 Estrado Cir	Bartow	(NULL)
1005	Transamerica Financial Services	Mary Little	13845 N 56th Pl	Fort Worth	OH
1006	Transamerica Occidental Life	Kate Lindamood	33032 Lighthouse Ct	Charleston	OH
1007	Transamerica Financial Group	Justin Echavarria	2662 E 2nd St	Crawfordsvile	OH
1008	Transamerica Financial Service	Olivia Bach	PO Box 8179	Lyman	OH
1009	Huntington Beach Union HS Dist	G. Weston	17272 Chapparral Ln	S. Creek	OH
101	KAISER HOSPITAL	KATHY JOHNSON	5798 GROVEWOOD DR	Mentor	CA
1010	Legislative Reference Bureau	V Eberhardt	3112 Quebrada Cir	Okauchee	OH
1011	Gallatin Medical Group Inc.	Jeremy Forero	3405 W Danbury D230	Marion Oaks	OH
1012	Northern Gummman Corp.	X Lauman	10075 Prader Ave	Stony Point	OH

Nodes

- Data Sorting
- Data Joining
- Data Joining (Non-Key)
- Data Union
- SQL Lookup
- SQL Execute
- Parameterized SQL Query

Quality

Enrichment

Entity Resolution

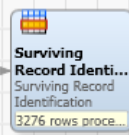
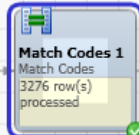
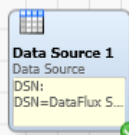
- Match Codes
- Match Codes (Parsed)
- Clustering
- Cluster Update

Folders

Accelerators

Match Codes

Help...



Match Codes 1

ID	COMPANY	CONTACT	ADDRESS	CITY	STATE	COMPANY_MatchCode	CONTACT_MatchCode
1	First Merit Bank	James E. Briggs	19 East Broad Street	Saint Louis	Missouri	GY4~BY~\$	MYF\$C&B_4\$
10	DataFlux Corporation	Bob Brauer	6512 Six Forks Road - 404B	Raleigh	North Carolina	8~GWX\$	MYLY\$M@M\$
100	KAISER HOSPITAL	LUTHER BAKER	3560 E 116TH ST	Cleveland	CA	34Y24N~W\$	M3Y\$W#3_\$
1000	TransAmerica Life Insurance	Irene Greaves	555 W Fifth St	S Northfield	OH	~YP4BY3WGP4YP3\$	FYM\$&YB\$
1001	Transamerica Life Ins & Annuity Co	Rob Drain	7718 Elder Way	Clementson	OH	~YP4BY3WGP4YP3P~\$	8YB\$M@M\$
1002	Transamerica Occidental	Tonia Gerstner	1018 N Hayworth Ave	Urb Canas	OH	~YP4BY3\$	FY4~BY\$~B&\$
1003	Transamerica Life Insurance	Joshua Hodgekin	244 Evans	Old Mission	OH	~YP4BY3WGP4YP3\$	2F3B\$C@42\$
1004	Transamerica Occidental Life	Nancy Weinstock	2134 Estrado Cir	Bartow	(NULL)	~YP4BY3J28P~WG\$	LB4~3\$P&PJ\$
1005	Transamerica Financial Services	Mary Little	13845 N 56th Pl	Fort Worth	OH	~YP4BY3GP3W\$	W~W\$B&YR\$
1006	Transamerica Occidental Life	Kate Lindamood	33032 Lighthouse Ct	Charleston	OH	~YP4BY3J28P~WG\$	WB8B\$J&~2\$
1007	Transamerica Financial Group	Justin Echavarria	2662 E 2nd St	Crawfordsvllle	OH	~YP4BY3GP3W\$	8J2MY&\$C#4~\$
1008	Transamerica Financial Service	Olivia Bach	PO Box 8179	Lyman	OH	~YP4BY3GP3W\$	M3\$W&\$
1009	Huntington Beach Union HS Dist	G. Weston	17272 Chapparel Ln	S. Creek	OH	2P~PF~PM324\$	L4~B\$F\$
101	KAISER HOSPITAL	KATHY JOHNSON	5798 GROVEWOOD DR	Mentor	CA	34Y24N~W\$	CB4B\$J&~2\$
1010	Legislative Reference Bureau	V Eberhardt	3112 Quebrada Cir	Okauchee	OH	WF4W~VYGYP4\$	8MY~\$V\$
1011	Gallatin Medical Group Inc.	Jeremy Forero	3405 W Danbury D230	Marion Oaks	OH	PW~PB83W\$	GY&\$CYB7\$
1012	Northrop Grumman Corp.	Y. Luman	10075 Prader Ave	Stony Point	OH	PY~Y2MYE8D\$	WBR\$

Nodes

- Data Sorting
- Data Joining
- Data Joining (Non-Key)
- Data Union
- SQL Lookup
- SQL Execute
- Parameterized SQL Query

Quality

Enrichment

Entity Resolution

- Match Codes
- Match Codes (Parsed)
- Clustering
- Cluster Update

Folders

Accelerators

Clustering

Help...

ID	COMPANY	CONTACT	ADDRESS	CITY	STATE	COMPANY_MatchCode	CONTACT_MatchCode	Cluster_ID
1	First Merit Bank	James E. Briggs	19 East Broad Street	Saint Louis	Missouri	GY4~BY~\$	MYF\$\$\$\$\$\$\$C&B_4\$\$\$\$	0
4	First Merit	Jim Briggs	19 E. BROAD ST.	St Louis	Mizzou	GY4~BY~\$	MYF\$\$\$\$\$\$\$C&B_4\$\$\$\$	0
3	1st Merit Bank	Mr James Brigs	19 E Broad St	St. Louis	MO.	GY4~BY~\$	MYF\$\$\$\$\$\$\$C&B_4\$\$\$\$	0
10	DataFlux Corporation	Bob Brauer	6512 Six Forks Road - 404B	Raleigh	North Carolina	8~GWX\$	MYLY\$\$\$\$\$\$\$M@M\$\$\$\$	1
8074	DATAFLUX INC.	Mr. Robert J. Brauer	6512 6 Forks Rd Suite 404B	Raleigh	NC	8~GWX\$	MYLY\$\$\$\$\$\$\$M@M\$\$\$\$	1
100	KAISER HOSPITAL	LUTHER BAKER	3560 E 116TH ST	Cleveland	CA	34Y24N~W\$\$\$\$\$\$\$\$\$\$\$\$\$\$\$\$	M3Y\$\$\$\$\$\$\$W#3_\$\$\$\$	2
1000	TransAmerica Life Insurance	Irene Greaves	555 W Fifth St	S Northfield	OH	~YP4BY3WGP4YP3\$\$\$\$\$\$\$\$	FYM\$\$\$\$\$\$\$&YB\$\$\$\$	3
1001	Transamerica Life Ins & Annuity Co	Rob Drain	7718 Elder Way	Clementson	OH	~YP4BY3WGP4YP3P~\$\$\$\$\$\$\$\$	8YB\$\$\$\$\$\$\$M@M\$\$\$\$	4
1002	Transamerica Occidental	Tonia Gerstner	1018 N Hayworth Ave	Urb Canas	OH	~YP4BY3\$\$\$\$\$\$\$\$\$\$\$\$\$\$\$\$	FY4~BY\$\$\$\$\$~B&\$\$\$\$	5
1003	Transamerica Life Insurance	Joshua Hodgekin	244 Evans	Old Mission	OH	~YP4BY3WGP4YP3\$\$\$\$\$\$\$\$	2F3B\$\$\$\$\$\$\$C@42\$\$\$\$	6
1004	Transamerica Occidental Life	Nancy Weinstock	2134 Estrado Cir	Bartow	(NULL)	~YP4BY3J28P~WG\$\$\$\$\$\$\$\$	LB4~3\$\$\$\$\$P&PJ\$\$\$\$	7
1005	Transamerica Financial Services	Mary Little	13845 N 56th Pl	Fort Worth	OH	~YP4BY3GP3W\$\$\$\$\$\$\$\$	W~W\$\$\$\$\$\$\$B&YR\$\$\$\$	8
1006	Transamerica Occidental Life	Kate Lindamood	33032 Lighthouse Ct	Charleston	OH	~YP4BY3J28P~WG\$\$\$\$\$\$\$\$	WB8B\$\$\$\$\$J&~2\$\$\$\$	9
1007	Transamerica Financial Group	Justin Echavarria	2662 E 2nd St	Crawfordsville	OH	~YP4BY3GP3W\$\$\$\$\$\$\$\$	8J2MY&\$\$\$\$\$C#4~\$\$\$\$	10
1008	Transamerica Financial Service	Olivia Bach	PO Box 8179	Lyman	OH	~YP4BY3GP3W\$\$\$\$\$\$\$\$	M3\$\$\$\$\$\$\$#WM&\$\$\$\$	11
1009	Huntington Beach Union HS Dist	G. Weston	17272 Chapparral Ln	S. Creek	OH	2P~PF~PM324\$\$\$\$\$\$\$\$	L4~B\$\$\$\$\$\$\$F\$\$\$\$	12
101	KAISER HOSPITAL	KATHY JOHNSON	5398 GROVEWOOD DR	Mantoloking	CA	34Y24N~W\$\$\$\$\$\$\$\$	CR4R\$\$\$\$\$E18~7\$\$\$\$	13

Nodes

- Data Sorting
- Data Joining
- Data Joining (Non-Key)
- Data Union
- SQL Lookup
- SQL Execute
- Parameterized SQL Query

Quality

Enrichment

Entity Resolution

- Match Codes
- Match Codes (Parsed)
- Clustering
- Cluster Update

Folders

Accelerators

Surviving Record Identification

Help...

ID	COMPANY	CONTACT	ADDRESS	CITY	STATE	COMPANY_MatchCode	CONTACT_MatchCode	Cluster_ID
1	First Merit Bank	James E. Briggs	19 East Broad Street	Saint Louis	Missouri	GY4~BY~\$	MYF\$\$\$\$\$\$\$C&B_4\$\$\$\$	0
10	DataFlux Corporation	Bob Brauer	6512 Six Forks Road - 404B	Raleigh	North Carolina	8~GWX\$	MYLY\$\$\$\$\$\$\$M@M\$\$\$\$\$	1
100	KAISER HOSPITAL	LUTHER BAKER	3560 E 116TH ST	Cleveland	CA	34Y24N~W\$\$\$\$\$\$\$\$\$\$\$\$\$\$\$\$	M3Y\$\$\$\$\$\$\$W#3_\$\$\$\$	2
1000	Transamerica Life Insurance	Irene Greaves	555 W Fifth St	S Northfield	OH	~YP4BY3WGP4YP3P~\$\$\$\$\$\$	FYM\$\$\$\$\$\$\$&YB\$\$\$\$\$	3
1001	Transamerica Life Ins & Annuity Co	Rob Drain	7718 Elder Way	Clementson	OH	~YP4BY3WGP4YP3P~\$\$\$\$\$\$	8YB\$\$\$\$\$\$\$M@M\$\$\$\$\$	4
1002	Transamerica Occidental	Tonia Gerstner	1018 N Hayworth Ave	Urb Canas	OH	~YP4BY3P~\$\$\$\$\$\$	FY4~BY\$\$\$\$~B&\$\$\$\$\$	5
1003	Transamerica Life Insurance	Joshua Hodgakin	244 Evans	Old Mission	OH	~YP4BY3WGP4YP3P~\$\$\$\$\$\$	2F3B\$\$\$\$\$C@42\$\$\$\$\$	6
1004	Transamerica Occidental Life	Nancy Weinstock	2134 Estrado Cir	Bartow	(NULL)	~YP4BY3J28P~WG\$\$\$\$\$	LB4~3\$\$\$\$\$P&PJ\$\$\$\$\$	7
1005	Transamerica Financial Services	Mary Little	13845 N 56th Pl	Fort Worth	OH	~YP4BY3GP3W\$\$\$\$\$	W~W\$\$\$\$\$\$\$B&YR\$\$\$\$\$	8
1006	Transamerica Occidental Life	Kate Lindamood	33032 Lighthouse Ct	Charleston	OH	~YP4BY3J28P~WG\$\$\$\$\$	WB8B8\$\$\$\$\$J&~2\$\$\$\$\$	9
1007	Transamerica Financial Group	Justin Echavarria	2662 E 2nd St	Crawfordsvile	OH	~YP4BY3GP3W\$\$\$\$\$	&J2MY&\$\$\$\$\$C#4~\$\$\$\$\$	10
1008	Transamerica Financial Service	Olivia Bach	PO Box 8179	Lyman	OH	~YP4BY3GP3W\$\$\$\$\$	M3\$\$\$\$\$\$\$#WM&\$\$\$\$\$	11
1009	Huntington Beach Union HS Dist	G. Weston	17272 Chapparral Ln	S. Creek	OH	2P~PF~PM324\$\$\$\$\$	L4~B\$\$\$\$\$\$\$F\$\$\$\$\$	12
101	KAISER HOSPITAL	KATHY JOHNSON	5798 GROVEWOOD DR	Mentor	CA	34Y24N~W\$\$\$\$\$	CB4B\$\$\$\$\$J&~2\$\$\$\$\$	13
1010	Legislative Reference Bureau	V Eberhardt	3112 Quebrada Cir	Okauchee	OH	FW4W~VGY4P4\$\$\$\$\$	&MY~\$\$\$\$\$\$\$V\$\$\$\$\$	14
1011	Gallatin Medical Group Inc.	Jeremy Forero	3405 W Danbury D230	Marion Oaks	OH	FW~PB83W\$\$\$\$\$	GY8&\$\$\$\$\$\$\$CYB7\$\$\$\$\$	15
1012	Northon Gummer Corp	X Lauman	10075 Prader Ave	Stony Point	OH	RY~2YMFYR&\$\$\$\$\$	WBR&\$\$\$\$\$A\$\$\$\$\$	16

Master Data Management

DataFlux® qMDM Solution - Windows Internet Explorer

http://localhost:8000/qmdm/qmdm.html?locale=en_US#qmdm

DataFlux® qMDM Solution

DataFlux® qMDM Solution
Master Repository Manager

Entities Relationships Reports Administration Services Accelerators

Welcome rona Help About Log off

Administration Tasks

Administration

- Administration Tasks
 - Modify/Add Attribute Relationship Types
 - Modify/Add Attributes
 - Modify/Add Cluster Grouping Types
 - Modify/Add Cluster Relationship Types
 - Modify/Add Countries
 - Modify/Add Entity Grouping Types
 - Modify/Add Entity Types
 - Modify/Add Languages
 - Modify/Add Source Systems
 - Modify/Add Attribute Relationships

Modify/Add Attributes

Add Row Refresh Row Single Row View Save Load File Reload Service Run Service Excel Preferences

	mdm_attr_name	mdm_entity_type	mdm_display_group_name	mdm_df_definition	mdm_attr_datatype	mdm_def_attr_label	mdm_attr_desc
1	MDM_ENTITY_ID	PERSON	MDM Identifiers		INTEGER	MDM Entity ID	MDM Entity ID from Hub
2	MDM_ENTITY_CLUSTER_ID	PERSON	MDM Identifiers		INTEGER	MDM Cluster ID	MDM Cluster ID from Hub
3	MDM_ENTITY_GROUPING...	PERSON	MDM Identifiers		STRING	MDM Best Record	MDM Best Record from Hub
4	MDM_SRC_SYS_ID	PERSON	MDM Identifiers		INTEGER	MDM Source System ID	MDM Source System ID fro...
5	MDM_SRC_SYS_REC_ID	PERSON	MDM Identifiers		STRING	MDM Source System Recor...	MDM Source System Recor...
6	MDM_COUNTRY_CD	PERSON	MDM Identifiers		STRING	MDM Country Code	MDM Country Code from Hu
7	MDM_MODIFIED_DTTM	PERSON	MDM Identifiers		DATE	MDM Last Modified	MDM Last Modified from Hut
8	POLICY_ID	PERSON	Policy Identifier		STRING	Policy Number	Policy Number from Src Sys
9	FULL_NAME	PERSON	Personal Information	FULLNAME	STRING	Full Name	Full Name from Src Sys
10	NAME_TITLE	PERSON	Personal Information		STRING	Name Title	Name Title from Src Sys
11	GIVEN_NAME1	PERSON	Personal Information		STRING	Given Name	Given Name from Src Sys
12	GIVEN_NAME2	PERSON	Personal Information		STRING	Middle Name	Middle Name from Src Sys
13	SURNAME1	PERSON	Personal Information	FAMILYNAME	STRING	Surname	Surname from Src Sys
14	SURNAME2	PERSON	Personal Information		STRING	Additional Surname	Additional Surname from Sr
15	NAME_SUFFIX	PERSON	Personal Information		STRING	Name Suffix	Name Suffix from Src Sys
16	NAME_APPENDAGE	PERSON	Personal Information		STRING	Name Appendage	Name Appendage from Src
17	GENDER_CD	PERSON	Personal Information		LIST	Gender Code	Gender Code from Src Sys
18	BIRTH_DATE	PERSON	Personal Information		DATE	Birth Date	Birth Date from Src Sys
19	ADDRESS1	PERSON	Address	ADDRESSLINE	STRING	Address Line 1	Address Line 1 from Src Sy
20	ADDRESS2	PERSON	Address		STRING	Address Line 2	Address Line 2 from Src Sy
21	CITY	PERSON	Address	CITY	STRING	City	City from Src Sys
22	COUNTY	PERSON	Address		STRING	County	County from Src Sys
23	STATE_PROVINCE	PERSON	Address	STATEPROVINCE	STRING	State/Province	State/Province from Src Sys
24	POSTALCODE	PERSON	Address	POSTALCODE	STRING	Postal Code	Postal Code from Src Sys
25	COUNTRY	PERSON	Address	COUNTRY	STRING	Country	Country from Src Sys
26	PHONE1	PERSON	Phone	PHONE	STRING	Primary Phone Number	Primary Phone Number from
27	PHONE2	PERSON	Phone	PHONE	STRING	Secondary Phone Number	Secondary Phone Number f
28	EMAIL1	PERSON	Email		STRING	Primary Email Address	Primary Email Address from
29	EMAIL2	PERSON	Email		STRING	Secondary Email Address	Secondary Email Address f
30	PRIVACY_FLG	PERSON	Flags		BOOLEAN	Privacy Flag	Privacy Flag from Src Sys
31	DF_FULL_NAME_MC	PERSON	Personal Information		STRING	Full Name Match Code	Full Name Match Code from
32	DF_SURNAME1_MC	PERSON	Personal Information		STRING	Surname Match Code	Surname Match Code from

Service completed successfully http://localhost:21036 105 rows Print Export

Evaluate

DataFlux Data Management Methodology

 Back

Profile Data

Define Business Rules

Resolve Entities

Monitor Tasks

Customize QKBs

Pattern

9999 A Aaaaaaaa

9999 Aaaaaa Aaaa

999 AaaaaAaaaa

9999 A Aaaaaaaa A...

Step2: Edit the check details:

State is null

State is not upper

Length of State is greater

Length of State is less than



Identification Analysis

Phone Validation

Table-Based Identical

Preprocessing

Evaluate: Monitor Tasks

The Monitor section on the Information Riser Bar is used to view summary and dashboard information about the results of the monitoring tasks executed in the registered repositories. The Summary tab shows a history of triggers (identified data elements) for all registered repositories as well as provides links to a summary listing of all monitoring task executions in a particular repository. There is also a dashboard tab for each registered repository that provides graphical monitoring information.

[Click here to learn more...](#)

☒ Display Data Management Methodology on Startup

Business Rules - Sandbox

File Edit View Tools Window Help

Tasks

- Control Character Check
- Vital Signs Checks

Rules

- Check for non-numeric character
- Check if Numeric field is zero
- Check Where Percent Null
- Heart Rate between 40 and 180
- Systolic BP less than Diastolic BP

Rule Sets

- Fields
 - DiaBP
 - HRTRATE
 - ID
 - Numeric
 - SysBP
- Field Sets
 - Identifiers
 - Numeric Fields
- Custom Metrics
- Sources

Rule: Systolic BP less than Diastolic BP

Type: Row

Rule number: 4

Edit Rule...

Rule Properties

Name:

User name:

Type: Row

Definition

SysBP

Check

Expression Builder

Step1: Select the check to perform:

- Compare Entire Field**
 - Compare a field to a value or to another field
 - Compare a field to a list of values
 - Compare a field to a field in a data source
 - Determine if a field is inside or outside a range of values
 - Determine how a field differs (amount/percentage) from another field
 - Determine if a field matches a pattern or regular expression
- Compare Portion of Field**
 - Compare a portion of a field to a value

Step2: Edit the check details:

SysBP is less than or equal to DiaBP

Edit...

OR

OK Cancel Help

[illegible]

Date*	Task	Desc	Rule	# Triggers	Status	# Rows Processed	User ID	Data Source
+ Today								
- Yesterday								
01/06/2011 17:13:44	Vital S							
01/06/2011 17:13:44	Vital S							
01/06/2011 17:10:52	Vital S							
01/06/2011 17:10:52	Vital S							
01/06/2011 17:08:30	Vital S							
01/06/2011 17:08:30	Vital S							
01/06/2011 16:47:46	Vital S							
01/06/2011 16:47:46	Vital S							
01/06/2011 14:24:46	Contr							
01/06/2011 14:24:20	Contr							
+ Last Month								
+ Three months ago								

Record Detail - Vital Signs Checks (Sandbox)

Task Name: Vital Signs Checks

Triggers:

2

Description: none

Rows processed:

3604

Run Date: 01/06/2011 17:13:44

User Id:

sukdhs

Rule: Heart Rate between 40 and 180

Source:

none

HRTRATE

28

16

Page 1 of 1



Rule Code:

```
if (`HRTRATE` < 40 or `HRTRATE` > 180) then  
  return true  
else  
  return false
```

Close

Export...

Print...

Information

Overview

Monitor

Information

Data Connections

Collections

Folders

Data Management Servers

Administration

Monitor

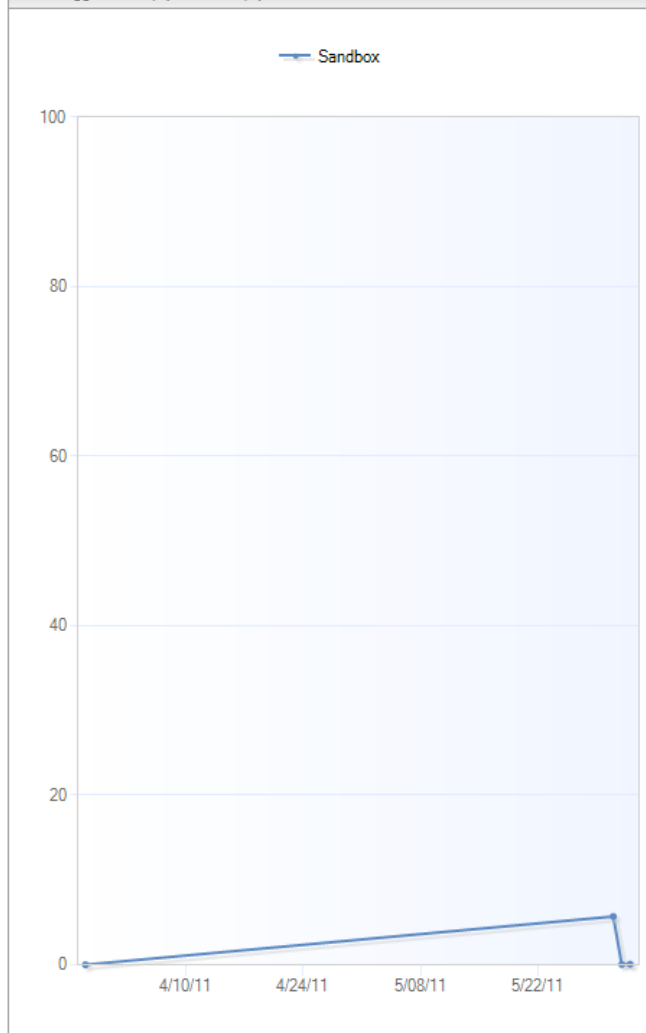
Summary DataFlux Sample - Dashboard Sandbox - Dashboard

Monitor Views

DataFlux Sample

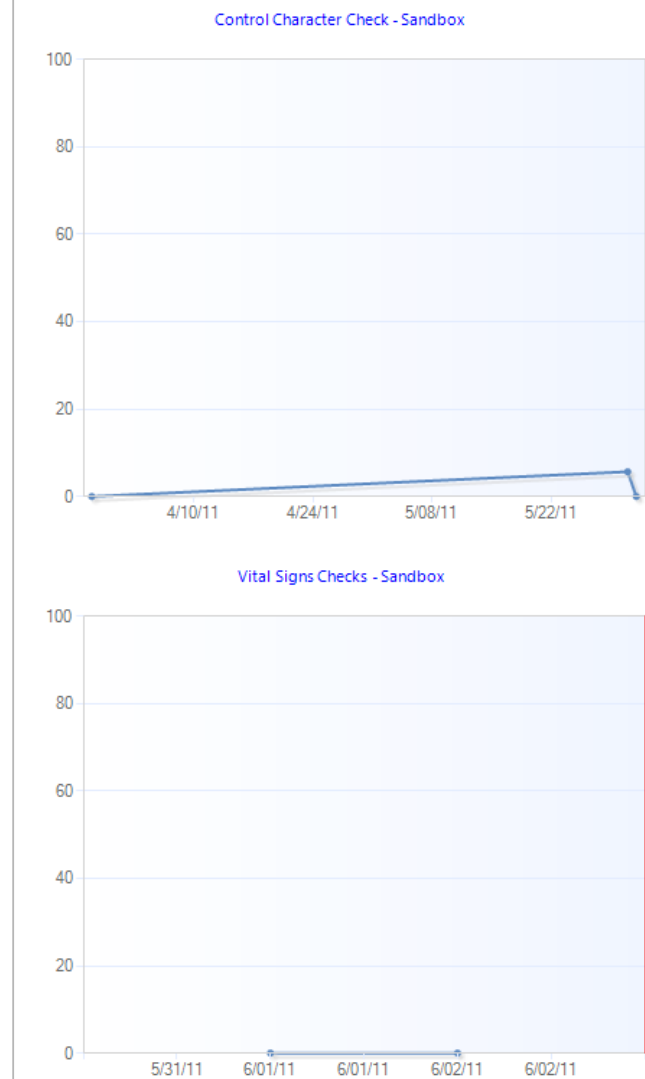
Sandbox

Total Trigger History (Last 90 days)



Task Trigger History (Last 90 days)

Options



Information

Overview

Monitor

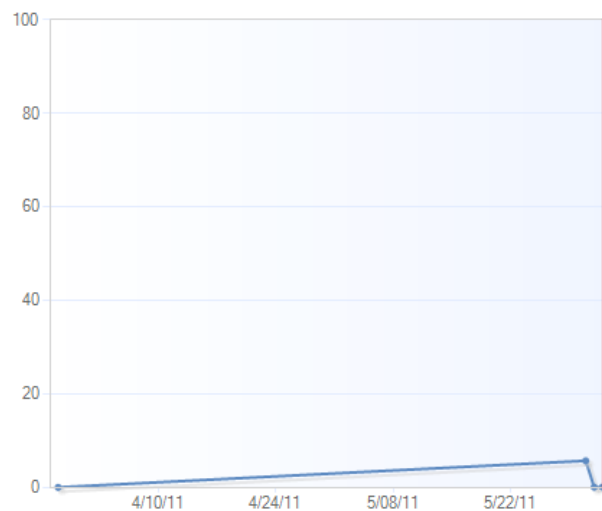
Monitor

Summary

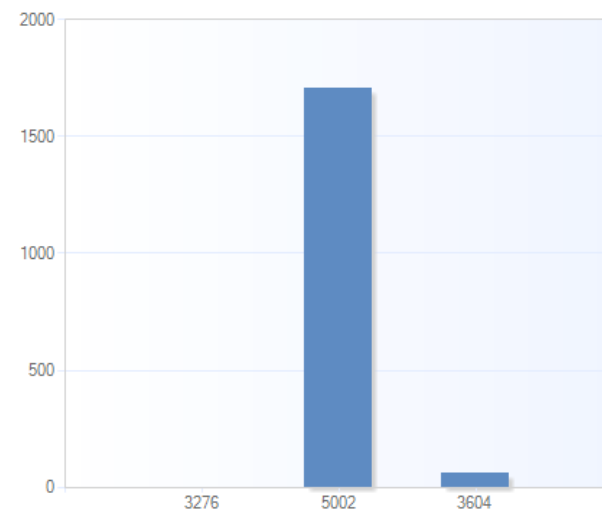
DataFlux Sample - Dashboard

Sandbox - Dashboard

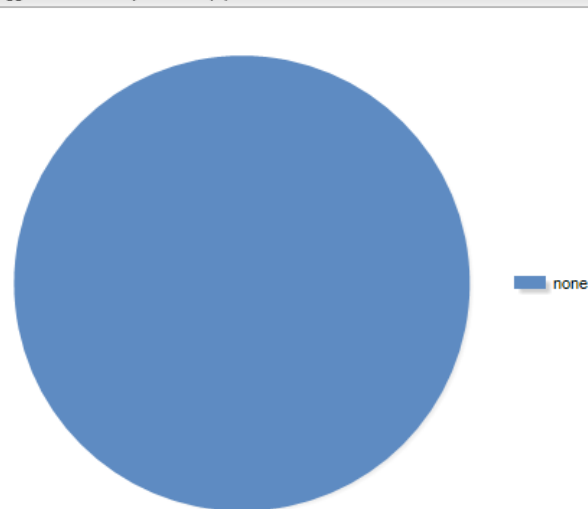
Triggers Per Date (Last 90 days)



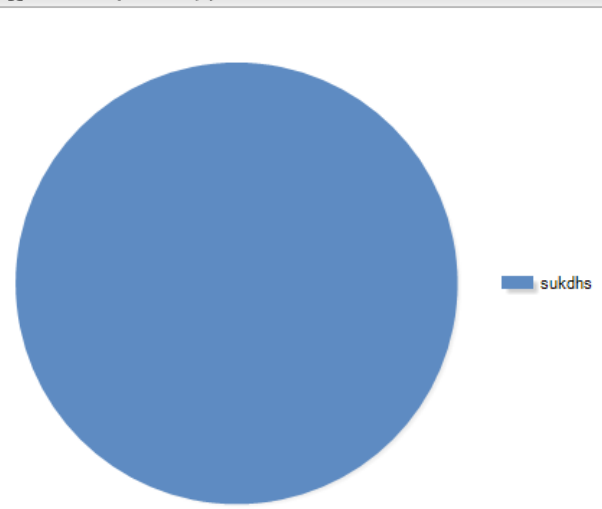
Triggers Per Records Processed (Last 90 days)



Triggers Per Source (Last 90 days)



Triggers Per User (Last 90 days)



Information

Data Connections

Collections

Folders

Data Management Servers

Administration

Control

DataFlux Data Management Methodology

Back

- Define Business Rules
 - Step2: Edit the check details:
 - State is null
 - State is not uppercase
 - Length of State is greater than 2
 - Length of State is less than 2
- Organize Fields
 - PHONE
 - Phone
 - Phone Number
 - Prospect
- Monitor Tasks
 - Bar chart and bell icon
- Run Jobs Remotely
 - Import... Export...
 - Name
 - Finance Data Extr...
 - Finance Data Quality
- Customize QKBs
 - Identification Analysis
 - Phone Translation
 - Table-Based Identical
 - Preprocessing

Control: Run Jobs Remotely

Data Management Server is used to remotely execute Data Management Studio jobs. The Data Management Server Manager, a Microsoft® Windows® client application, enables the management and execution of real-time services and batch jobs on a server from a remote client. Use the Data Management Servers Riser Bar to register and manage the DataFlux Data Management Servers.

[Click here to learn more...](#)

☒ Display Data Management Methodology on Startup

Nodes

Data Job

- Data Job
- Data Job (reference)

Utilities

- Echo
- Process Flow Work Table Reader
- Process Job (reference)
- Event Listen
- Expression
- Global Get/Set
- If Then
- Terminate Job

SQL

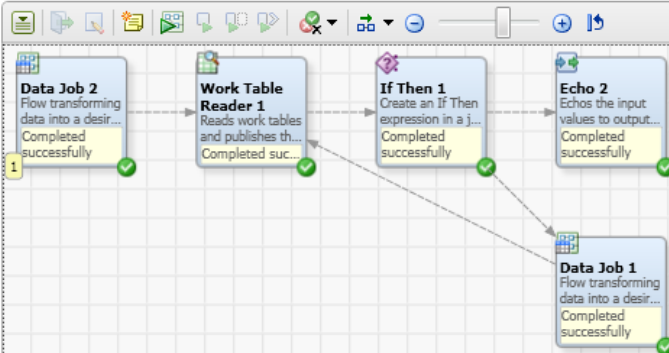
Data Integration

Folders

Accelerators

Echos the input values to output values

Help...

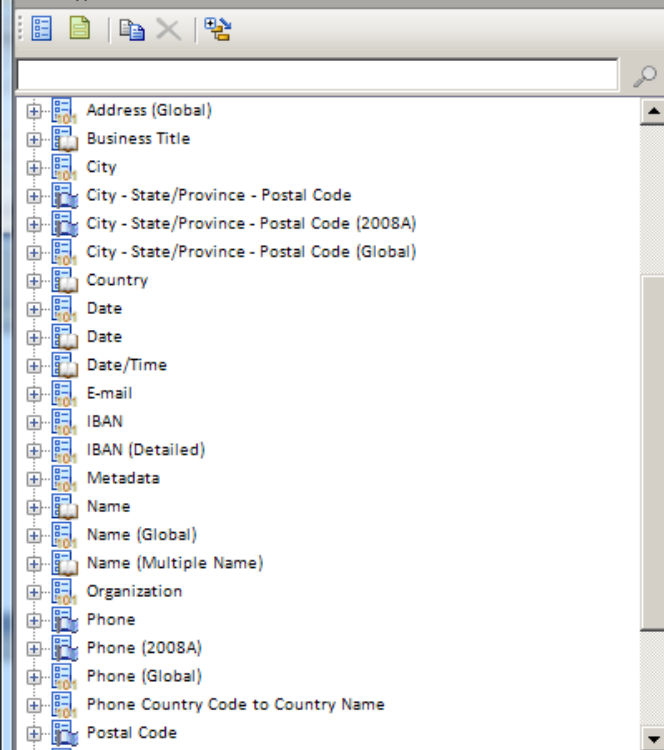


Log



Row	Node Name	Contained In	Instance	Node ID	Node Type	Status	Run Time (min)
0	dmsample_process					Completed successfully	00:05.944
1	Data Job 2	dmsample_process	0	DATAFLOW_5	Data Job	Completed successfully	00:00.195
2	Data Source 1	Data Job 2	0	1	Data Source	DSN: DSN=DataFlux Sample;DFXTYPE=ODBC SQL: SELECT "STATE" FROM "Contacts" 3276 rows read	
3	Frequency Distribution 1	Data Job 2	0	2	Frequency Distribution	Input 3276 rows. Output 61 rows	

Data Types and Definitions



Data Types and Definitions QKB Files Properties Notes Error log

Normalization Regex Li... < >

Vocabularies < >

Vocabulary

Variable Labels

Categorization Regex L... < >

Number Check

Default Categories

1 default category

Testing - Token Mappings

85

Test Values

systolic blood pressure mmhg

Results:

Token	Output
Event	
Measurement	blood pressure
Qualifier	systolic
Group	
Date	
Flag	
Timing	
Observation	
Treatment	
Unit	mmhg
Identifier	
Descriptor	

Data Connections

- Data Connections
 - Appalachian DDS
 - DataFlux Sample
 - dBASE Files
 - dfConglomerate Gifts
 - dfConglomerate Grocery
 - DFSsample
 - Excel Files
 - FoxPro Files
 - MS Access 97 Database
 - MS Access Database
 - Nicardipine Raw
 - BASE
 - SAS
 - ADVERSE
 - CHEM
 - CONMED
 - DEMO
 - ECG
 - FUBE
 - HEMA
 - VITAL
 - Nicardipine SDTM
 - PRISM
 - SalesforceDSN
 - Study 357
 - TestRepository
 - Text Files
 - Visio Database Samples

Information

Data Connections

Collections

Folders

Data Management Servers

Administration

VITAL
TABLE

Summary Data Fields Data Model Graph

Identification

Name: VITAL
Type: Table
Connection: Nicardipine Raw
Database: SAS
ID: dfd://Nicardipine Raw.TKTS/BASE/SAS/VITAL/

Details

Field count: 11

Lineage

VITAL



Possible Applications to Pharma

Exploration before analysis or warehousing

Profiling before analysis – after receipt of database

Standardisation of terms

AE Reconciliation – Redundant Data Analysis

Business Rules monitoring before database lock

Conclusion

Complete

- SAS have a complete solution which includes all the data quality needs of the enterprise

Easy

- The solution is easy for business users to use

Connected

- The solution can easily integrate with source systems using standard methods

Guided

- The data management methodology guides users through the process of improving data quality

