IBM Information Server Suite

QualityStage 8 Essentials

Course: DX741

Student Lab Exercises

Copyright, Disclaimer of Warranties and Limitation of Liability

© Copyright IBM Corporation February 2007
IBM Software Group
One Rogers Street
Cambridge, MA 02142

All rights reserved. Printed in the United States.

IBM and the IBM logo are registered trademarks of International Business Machines Corporation.

The following are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both:

AnswersOnLine
AIX
APPN
AS/400
BookMaster
C-ISAM
Client SDK
Cloudscape
Connection Services
Database Architecture
DataBlade
DataJoiner
DataPropagator
DB2
DB2 Connect
UniversalDataWarehouseBlueprint
DB2 Extenders
DB2 Universal Database
Distributed Database
Distributed Relational
DPI
DRDA
DynamicScalableArchitecture
DynamicServer
DynamicServer.2000
DynamicServer with Advanced DecisionSupportOption
DynamicServer with Extended ParallelOption
DynamicServer with UniversalDataOption
DynamicServer with WebIntegrationOption
DynamicServer, WorkgroupEdition
Enterprise Storage Server
FFST/2
Foundation.2000
Illustra
Informix
Informix4GL
InformixExtendedParallelServer
InformixInternet Foundation.2000
Informix RedBrick Decision Server
J/Foundation
MaxConnect
MVS
MVS/ESA
Net.Data
NUMA-Q
ON-Bar
OnLineDynamicServer
OS/2
OS/2 WARP
OS/390
OS/400
PTX
QBIC
QMF
RAMAC
RedBrickDesign
RedBrickDataMine
RedBrick Decision Server
RedBrickMineBuilder
RedBrickDecisionscape
RedBrickReady
RedBrickSystems
RelyonRedBrick
S/390
Sequent
SP
System View
Tivoli
TME
UniData
UniData&Design
UniversalDatabaseComponents
UniversalWebConnect
UniVerse
VirtualTableInterface
Visionary
VisualAge
WebIntegrationSuite
WebSphere

Microsoft, Windows, Window NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

Java, JDBC, and all Java-based trademarks are trademarks or registered trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.

UNIX is a registered trademark of The Open Group in the United States and other countries.

All other product or brand names may be trademarks of their respective companies.

All information contained in this document has not been submitted to any formal IBM test and is distributed on an “as is” basis without any warranty either express or implied. The use of this information or the implementation of any of these techniques is a customer responsibility and depends on the customer’s ability to evaluate and integrate them into the customer’s operational environment. While each item may have been reviewed by IBM for accuracy in a specific situation, there is no guarantee that the same or similar results will result elsewhere. Customers attempting to adapt these techniques to their own environments do so at their own risk. The original repository material for this course has been certified as being Year 2000 compliant.

This document may not be reproduced in whole or in part without the priori written permission of IBM.

Note to U.S. Government Users – Documentation related to restricted rights – Use, duplication, or disclosure is subject to restrictions set forth in GSA ADP Schedule Contract with IBM Corp.

Table of Contents

Lab 1: Case Study

Task: Read WINNCRM case study

Objective: Introduce the customer consolidation/initial load course project
Course Business Case: WINN Insurance CRM Project

Executive Summary
WINN Insurance is a leader in the insurance marketplace, providing their customers with a wide range of homeowner, automobile, and life insurance policies. WINN Corporation has made the strategic decision to change their account-based systems to a customer-based view. A comprehensive understanding of each of their customers is critical to reach WINN Corporation's goals for strategic growth initiatives.

Company Stats:

Name of Business: WINN Insurance

Type: Insurance

Lines of Business: Homeowners, Life, Automobile

The Business Challenge
WINN Corporation has recognized that their current data systems do not lend themselves to providing a comprehensive view of their customer base. While one customer may have purchased several automobile policies and life insurance policies, the systems do not give WINN Corporation visibility to all of the policies purchased by each unique customer. To achieve a customer-centric focus, WINN Corporation is looking to create a single customer repository.

WINN Corporation believes that with a single customer repository they will be able to enhance their organizational efficiencies and profits by getting a true count and accurate view of their unique customers. With this information, they will be able to better market the correct products to their existing customer base and save money by alleviating duplicate mailings.

WINN System and Data Information
WINN Corporation has identified multiple sources as feeds to this repository. The challenge is to identify rules for cleansing the data and providing consolidated views of the data across all sources. They currently maintain three systems, one for each type Insurance Line. They would ultimately like to create a customer information system that represents a consolidated view of their customers with an audit trail showing where the data originates.

WINN has already identified some data quality issues within their existing systems. In addition, they have identified several requirements for establishing their Customer Information (CIF) system. These issues are a serious concern of the management and they would like to see a comprehensive plan for addressing these problems and requirements.

- US records should be split from international address records
- There is customer name and address information spread across free-form text fields. They would like to see this organized into specified fields.

  - They want to match records across all sources
  - They want to remove all duplicate customer records*
- They have inconsistent naming conventions across their systems. They would like to see the name fields separate rather than in freeform text fields.
  - They need to standardize their address formats.
  - They want to establish a unique customer profile
  - They have found blank entries in their account fields
  - They want to maintain a cross-reference file to the legacy systems

Existing Systems

Source System	Description	LOB/Operational Function
AUTOHOME	**Homeowners & Automobile**	**Policy Maintenance**
LIFE	**Life**	**Policy Maintenance**

*Business rules for identifying duplicate customers

1. Same PID number
2. Same First Name, Last Name, Address, City, State, Zip
3. Similar Last Name, Same Address, City, State, Zip
4. Similar Address, Same City, State, Zip, Last Name

For the purposes of this course, the AUTOHOME and LIFE data have already been merged and are in the POLICY.dat file that is supplied on the student CD.

Lab 2: Setup Data Files

Task: Copy data files to disk

1. Place the DX741 student CD in your computer’s CD drive.
2. Copy the WINNCRM folder containing the student files to the C drive. This will create a folder called WINNCRM on the C drive.
3. If you choose another location for the files, you will need to make appropriate modifications to the exercises in this course.
4. Open the C:WINNCRM folder. Three files should be present:
    a. Policy.dat – contains the WINNCRM customer data.
    b. DX741TableDefs.dsx – contains the metadata table description for the Policy.dat data.
    c. DX741Solns.dsx – contains all the QualityStage objects you will build in this course.

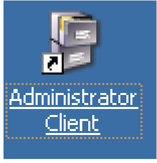
Lab 3: DataStage logon and create WINNCRM project

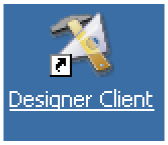
Assumptions:

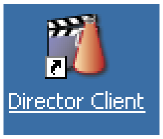
- You have a user id and password assigned by your instructor – exercises used in this course will use admin/admin as the user id and password but this will not necessarily be the one assigned for your system.

Task: DataStage Administrator logon

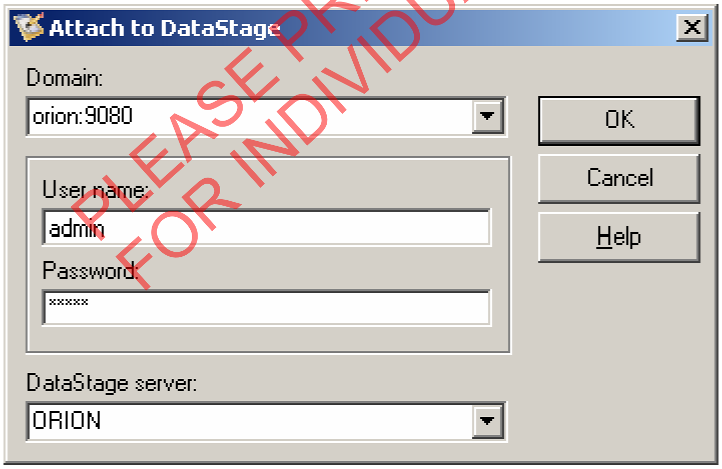
1. Locate the DataStage clients on your desktop.







2. Click the DataStage Administrator icon. The Attach to DataStage screen will appear. Enter the Administrator user id and password assigned by your instructor and then click OK.



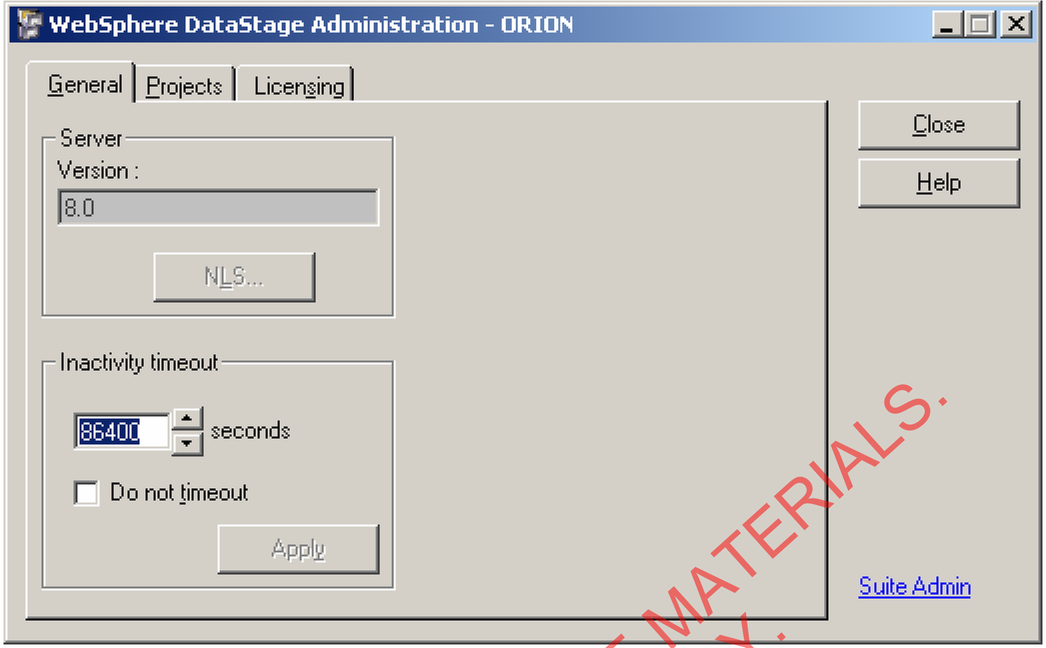
Domain: orion:9080

User name: admin

Password: *****

DataStage server: ORION

3. From the Administrator General tab verify the timeout settings and then click the Projects tab.



WebSphere DataStage Administration - ORION

Server

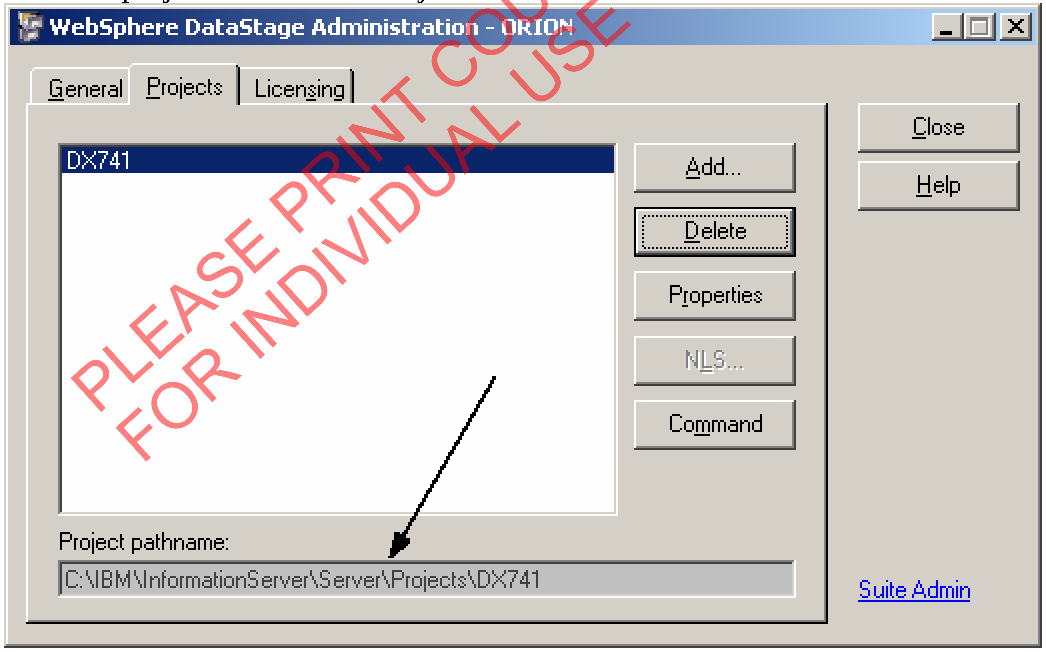
Version : 8.0

Inactivity timeout

86400 seconds

□ Do not timeout

4. Your screen will likely not contain the DX741 project shown in this graphic. However, note the presence of the Project pathname; this is the disk location that will be used to store the project's executable objects.



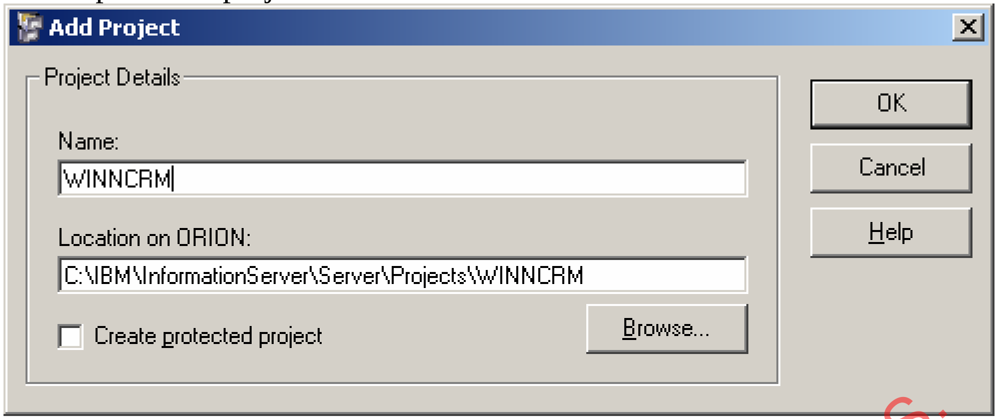
WebSphere DataStage Administration - ORION

DX741

Project pathname: C:\IBM\InformationServer\Server\Projects\DX741

5. Click the Add button and create a new project named WINNCRM. The actual name of your project is really not important; you could name it anything you like but all the screenshots in this course will assume the project name is WINNCRM. Do **not** click the

Create protected project check box.



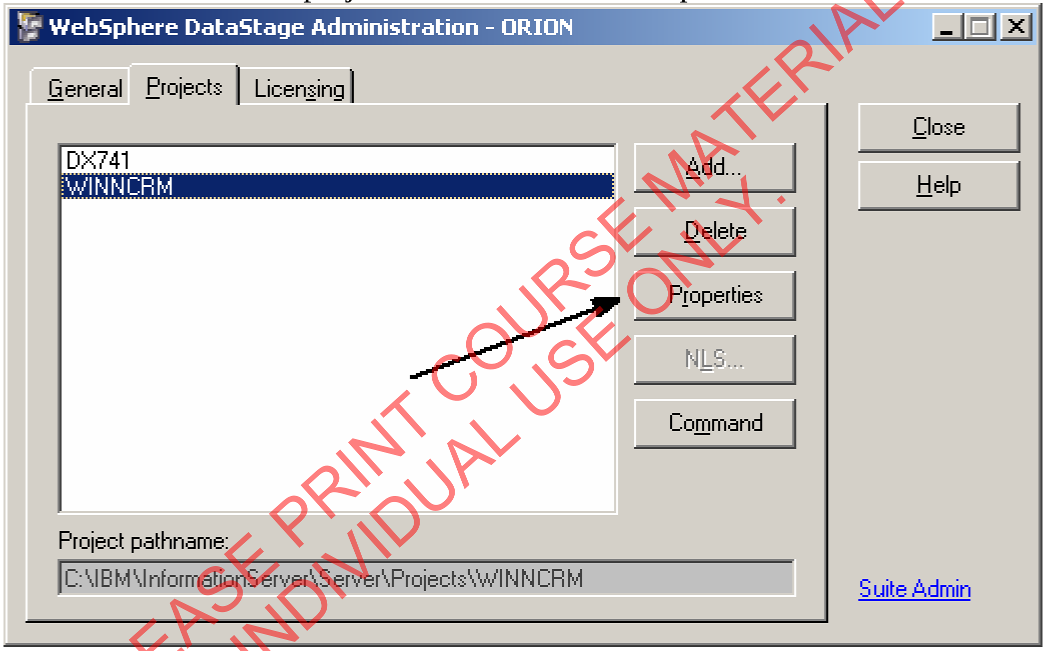
Project Details

Name: WINNCRM

Location on ORION: C:\IBM\InformationServer\Server\Projects\WINNCRM

□ Create protected project

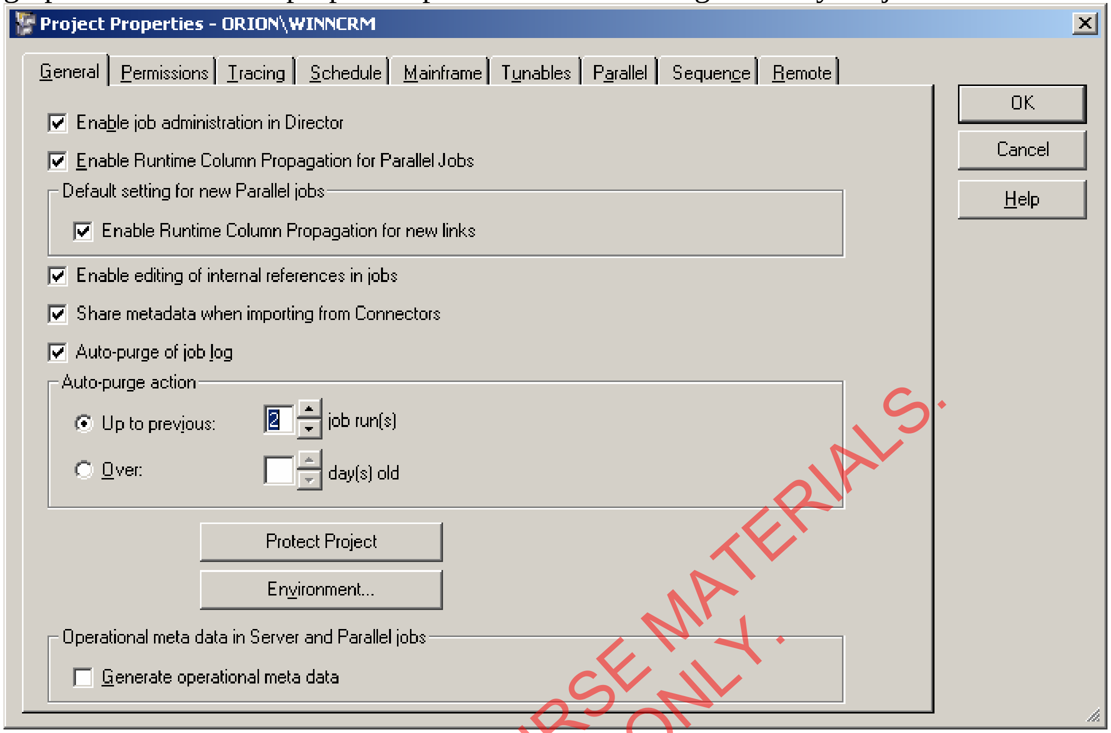
6. Select the WINNCRM project and then click the Properties button.



DX741
WINNCRM

Project pathname: C:\IBM\InformationServer\Server\Projects\WINNCRM

7. You are placed on the project’s General tab by default. Check the properties shown in the graphic below. These properties provide default settings for all your jobs.



☑ Enable job administration in Director

☑ Enable Runtime Column Propagation for Parallel Jobs

Default setting for new Parallel jobs

☑ Enable Runtime Column Propagation for new links

☑ Enable editing of internal references in jobs

☑ Share metadata when importing from Connectors

☑ Auto-purge of job log

Auto-purge action

● Up to previous: 2 job run(s)

○ Over: day(s) old

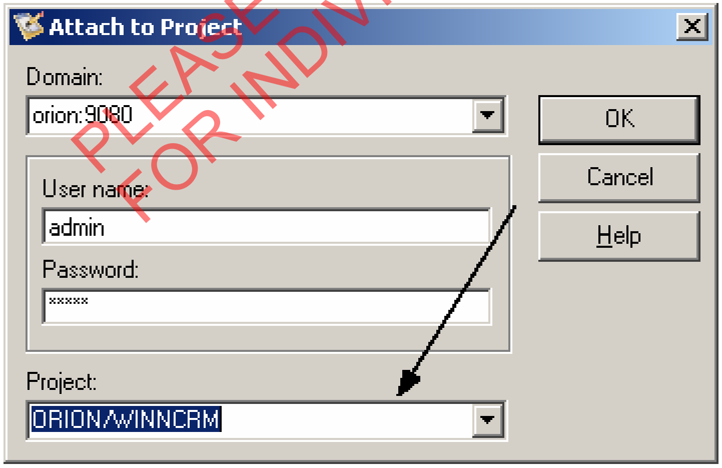
Operational meta data in Server and Parallel jobs

□ Generate operational meta data

8. Click OK to return to the Project listing screen.
9. Click OK to exit DataStage Administrator.

Task: Log onto the WINNCRM project using DataStage Designer

1. Click the Designer Client icon located on your desktop.
2. Complete the entries for user id and password and then click the drop down Project box to select WINNCRM.



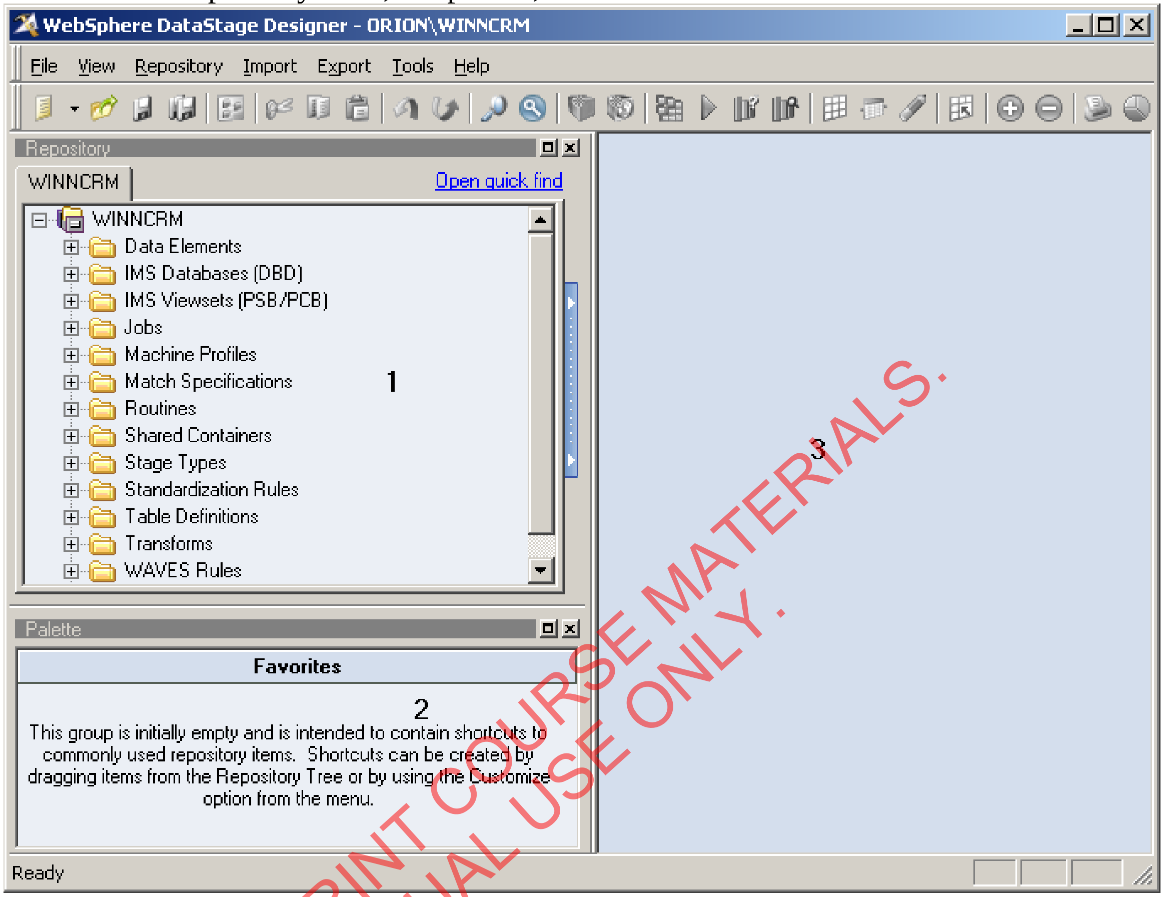
Domain: orion:9080

User name: admin

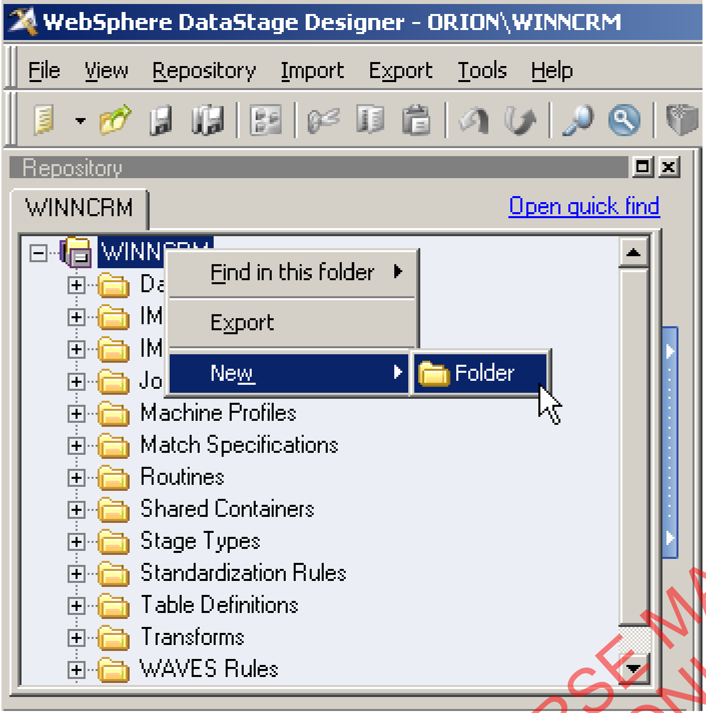
Password: *****

Project: ORION/WINNCRM

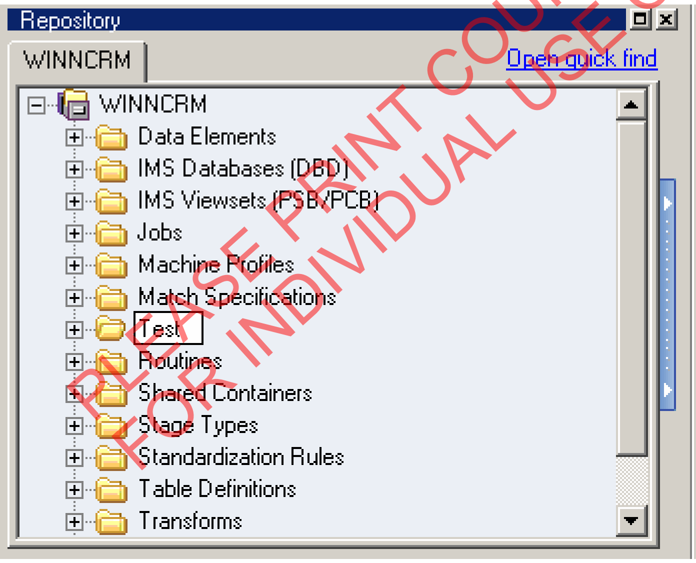
3. The Designer client has three primary areas labeled 1, 2, and 3 in the graphic below. These are 1 – repository view, 2 – palette, 3 – canvas.


WebSphere DataStage Designer - ORION\WINNCRM
File View Repository Import Export Tools Help
Repository
WINNCRM
Open quick find
WINNCRM
Data Elements
IMS Databases (DBD)
IMS Viewsets (PSB/PCB)
Jobs
Machine Profiles
Match Specifications
Routines
Shared Containers
Stage Types
Standardization Rules
Table Definitions
Transforms
WAVES Rules
1
Palette
Favorites
2
This group is initially empty and is intended to contain shortcuts to commonly used repository items. Shortcuts can be created by dragging items from the Repository Tree or by using the Customize option from the menu.
3
Ready

4. You can add folders to the repository hierarchy. Right-click on the WINNCRM root directory and select New ➔ Folder.



5. Rename the New folder to Test.



6. Right-click on the Test folder and then click Delete.

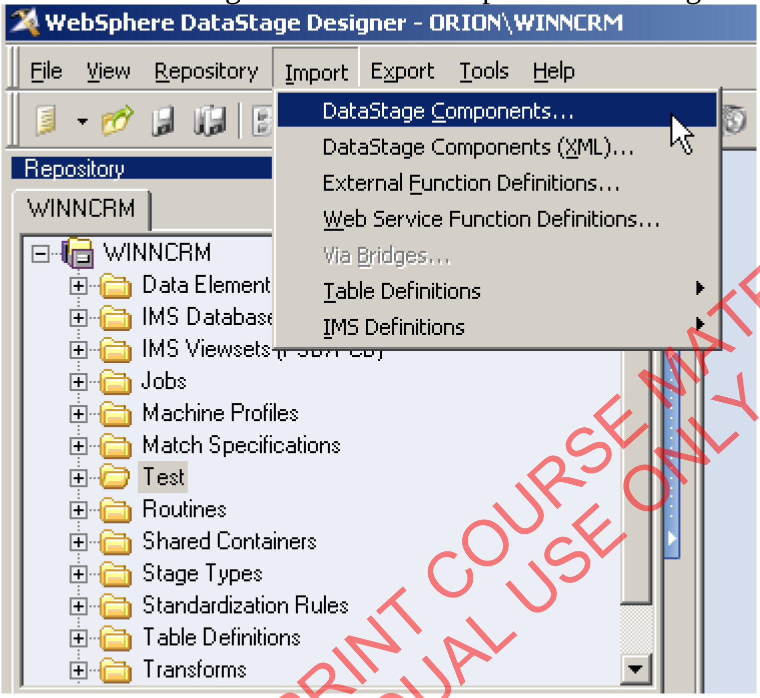
Lab 4: Import table definition meta data

Assumptions:

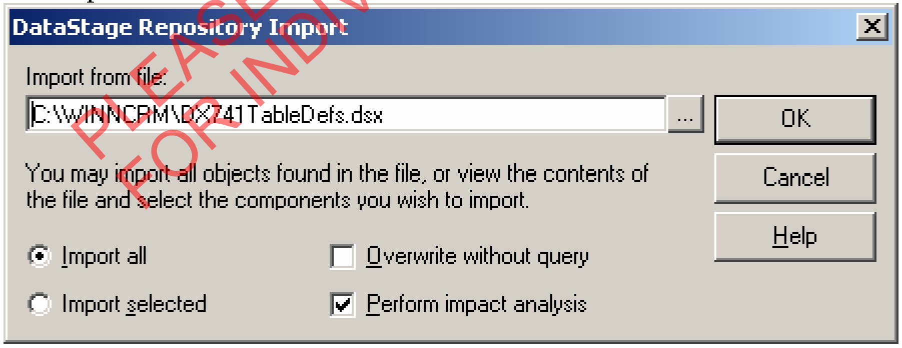
- You have a DataStage project named WINNCRM.

Task: Import table definition

1. While in the Designer Client click Import ➔ DataStage Components.



2. Use the … browse button to locate the C:\WINNCRM\DX741TableDefs.dsx file. Click the Import all radio button and then click OK.



DataStage Repository Import

Import from file: C:\WINNCRM\DX741TableDefs.dsx

You may import all objects found in the file, or view the contents of the file and select the components you wish to import.

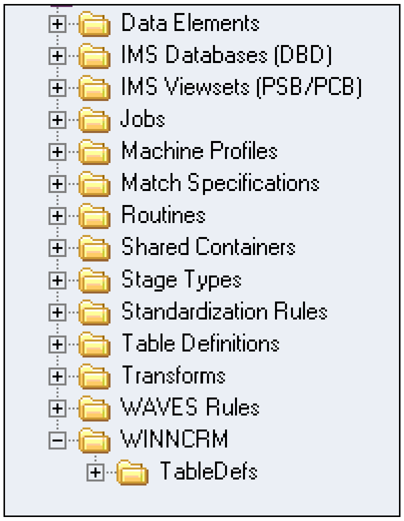
● Import all

○ Import selected

□ Overwrite without query

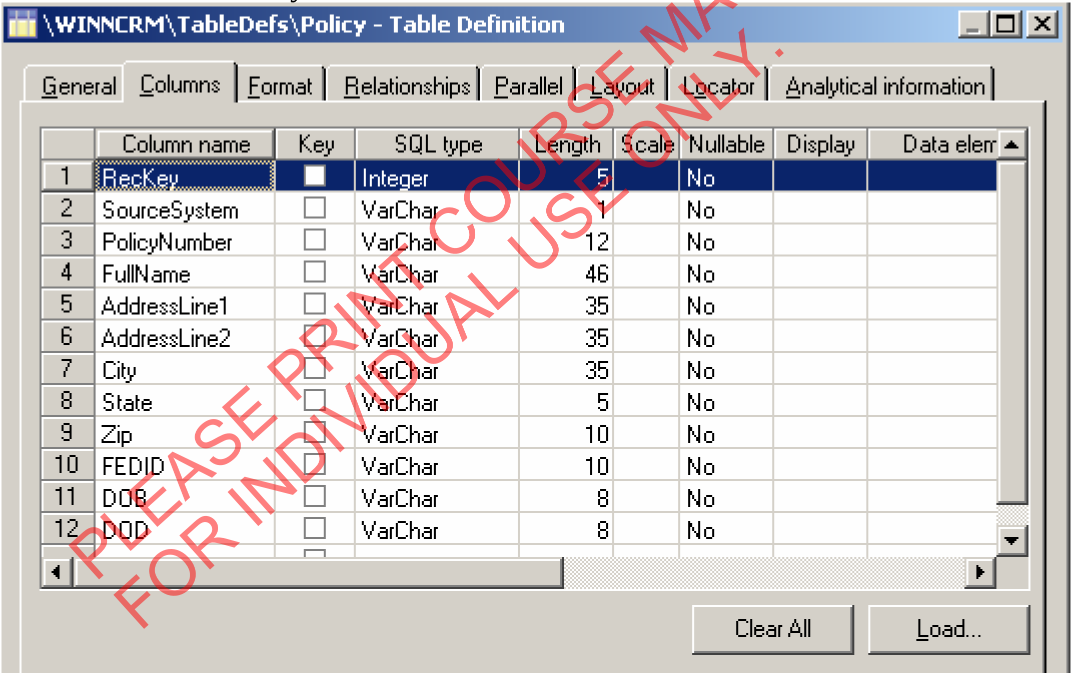
☑ Perform impact analysis

3. A new folder called WINNCRM will appear in the repository view. Open the WINNCRM folder and note the presence of a subfolder named TableDefs.



- Data Elements
- IMS Databases (DBD)
- IMS Viewsets (PSB/PCB)
- Jobs
- Machine Profiles
- Match Specifications
- Routines
- Shared Containers
- Stage Types
- Standardization Rules
- Table Definitions
- Transforms
- WAVES Rules
- WINNCRM
  - TableDefs

4. Open the TableDefs folder to find the Policy table definition.
5. Double-click the Policy table definition and then click the Columns tab.



\WINNCRM\TableDefs\Policy - Table Definition

	Column name	Key	SQL type	Length	Scale	Nullable	Display	Data elem
1	RecKey	☐	Integer	5		No		
2	SourceSystem	☐	VarChar	1		No		
3	PolicyNumber	☐	VarChar	12		No		
4	FullName	☐	VarChar	46		No		
5	AddressLine1	☐	VarChar	35		No		
6	AddressLine2	☐	VarChar	35		No		
7	City	☐	VarChar	35		No		
8	State	☐	VarChar	5		No		
9	Zip	☐	VarChar	10		No		
10	FEDID	☐	VarChar	10		No		
11	DOB	☐	VarChar	8		No		
12	DOD	☐	VarChar	8		No		

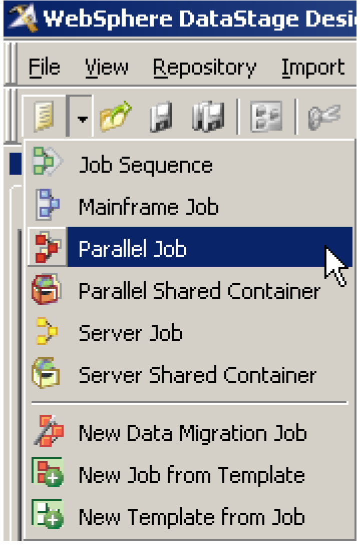
Lab 5: Build and run DataStage read sequential file job

Assumptions:

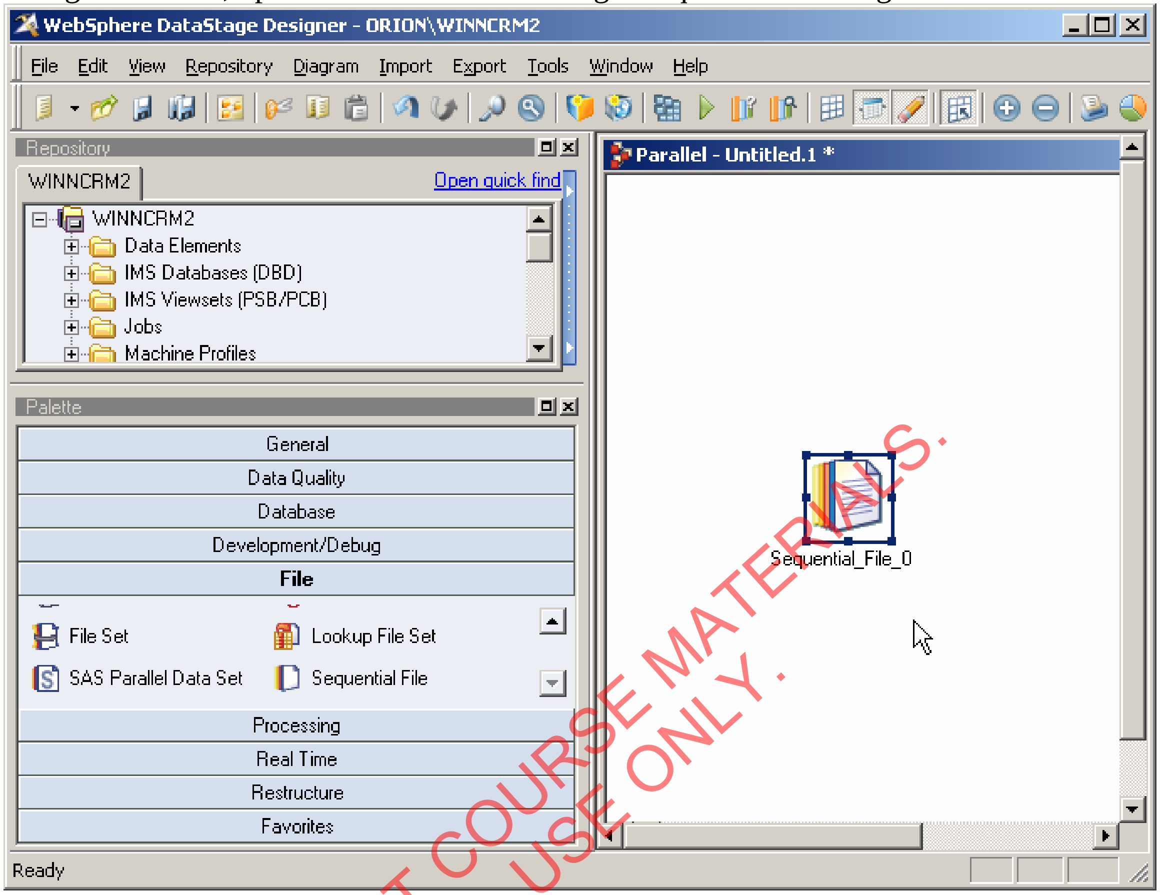
- You have a DataStage project named WINNCRM.

Task: Create DataStage job

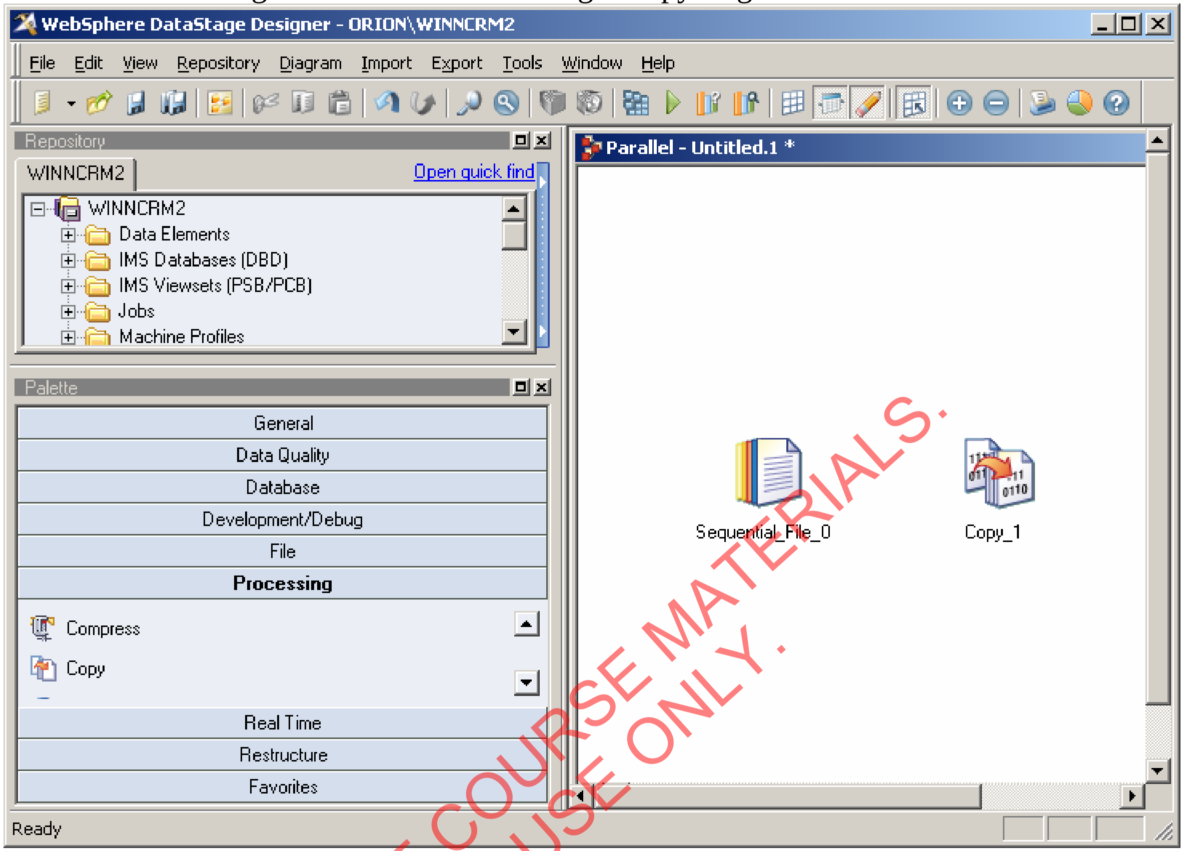
1. In Designer Client click the down arrow immediately to the right of the New icon on the file menu. Select Parallel job.



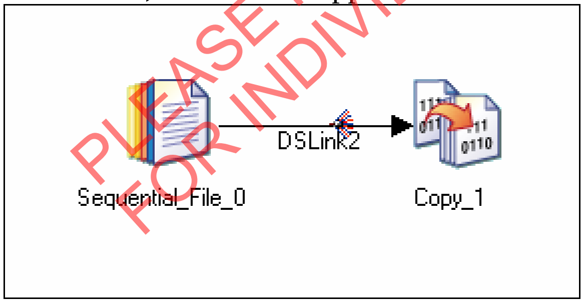
2. Using the Palette, open the File folder and drag a Sequential File stage onto the canvas.



3. From the Processing folder of the Palette drag a Copy stage onto the canvas.

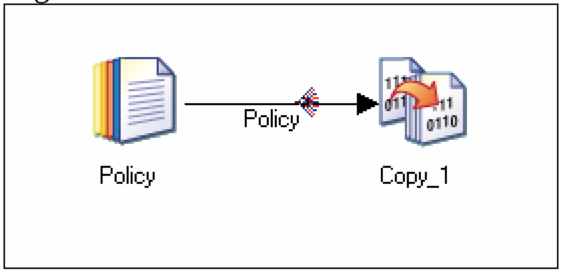


4. Place your cursor over the Sequential File stage and then hold down the right mouse button. Continue holding down the right mouse button, drag the cursor over to the middle of the Copy stage and then release the right mouse button. A link, indicated by a data flow arrow, should now appear between the stages.

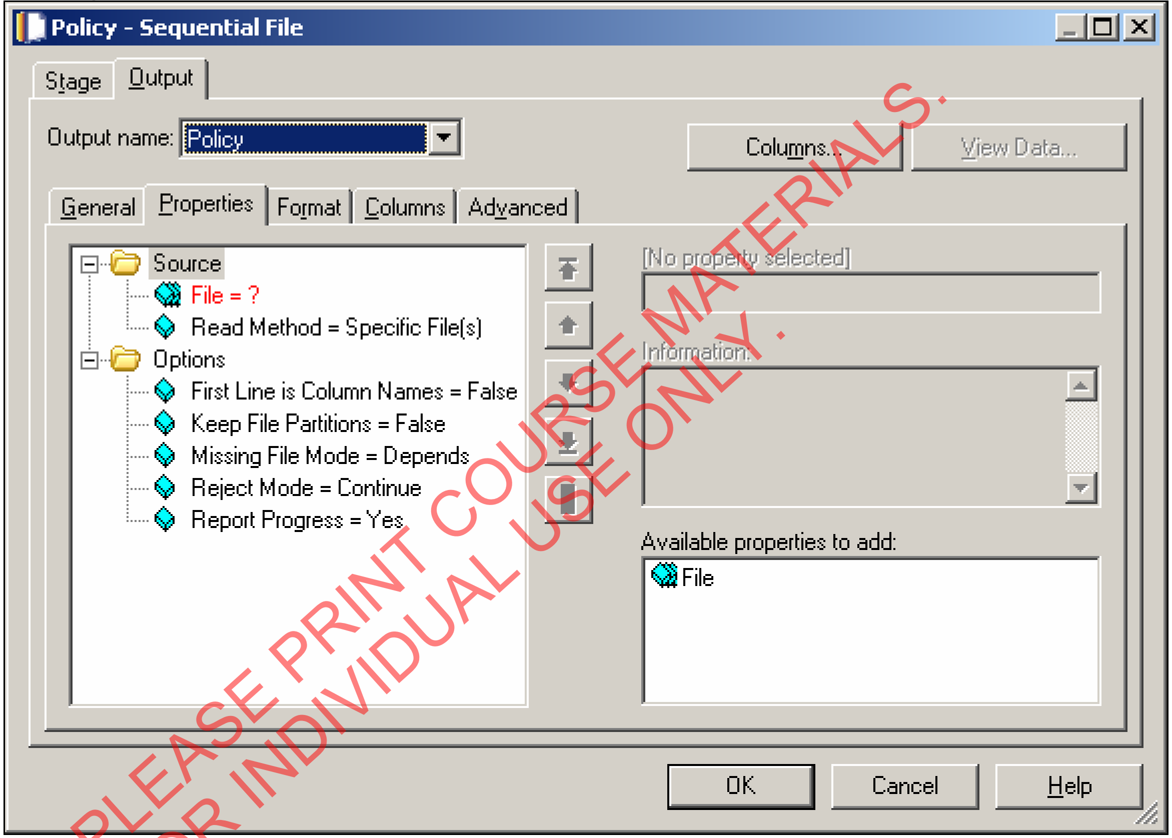


5. Right-click on the source Sequential File stage and rename it to Policy.

6. Right-click on the link and rename it to Policy.



7. Double-click the source Sequential File stage, now named Policy. This action will open the stage properties editor.



Output name: Policy

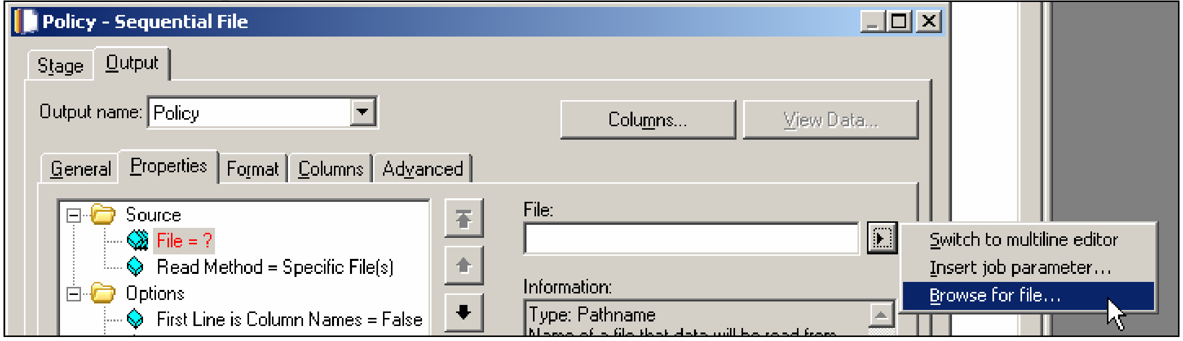
- Source
  - File = ?
  - Read Method = Specific File(s)
- Options
  - First Line is Column Names = False
  - Keep File Partitions = False
  - Missing File Mode = Depends
  - Reject Mode = Continue
  - Report Progress = Yes

Available properties to add:

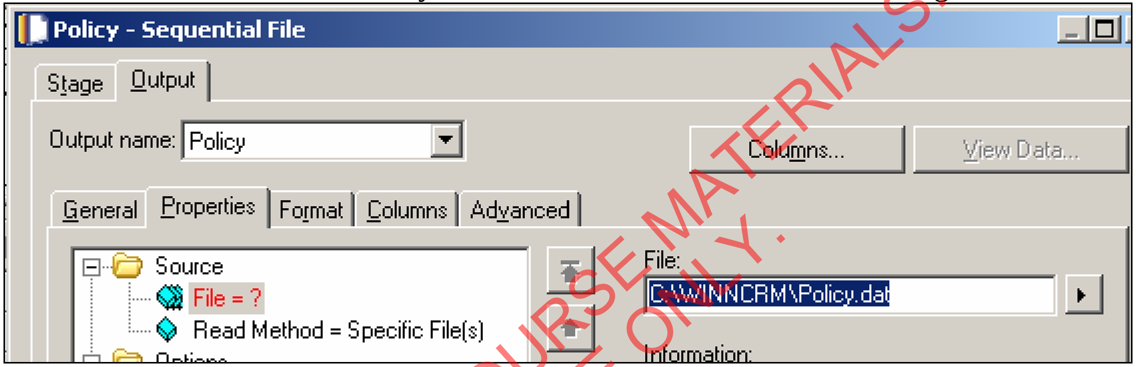
File

8. Required properties that are not yet completed are highlighted in red. This is true for the File property under the Source folder in the stage editor.

9. Select the File property and then click the right arrow in the File text box. Select the Browse for file option.

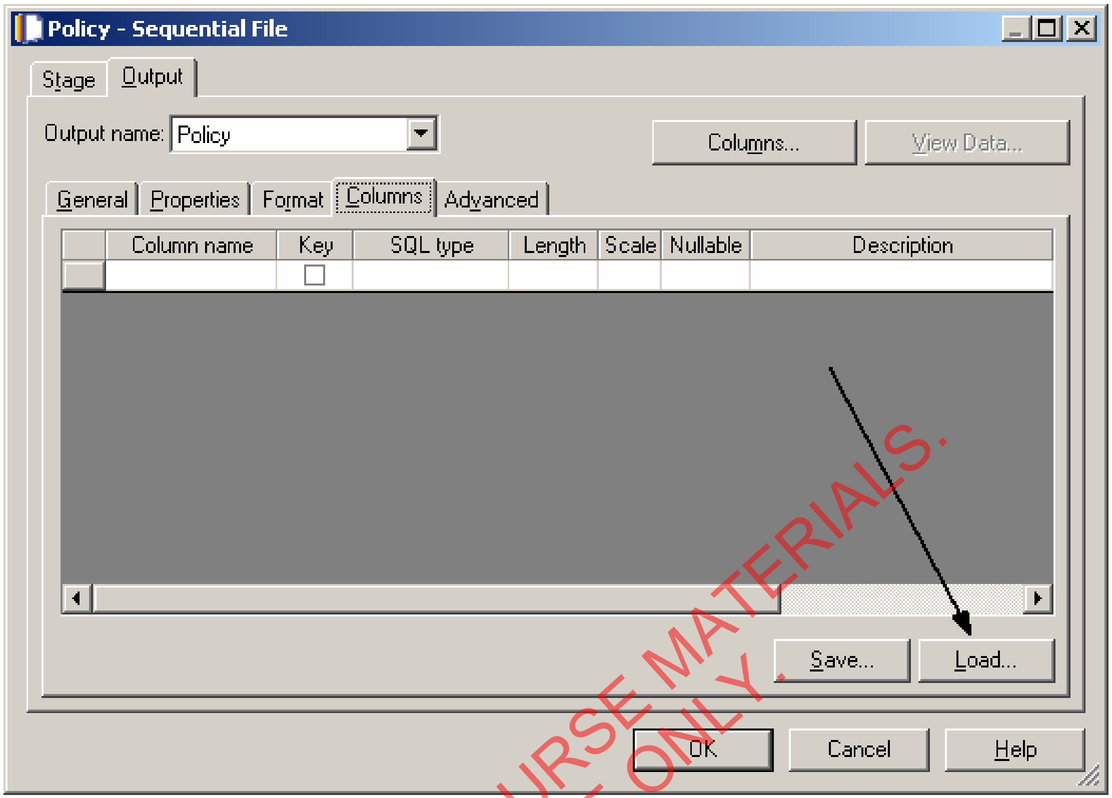


10. Locate the C:\WINNCRM\Policy.dat file and select it to return to the stage editor.



File: C:\WINNCRM\Policy.dat

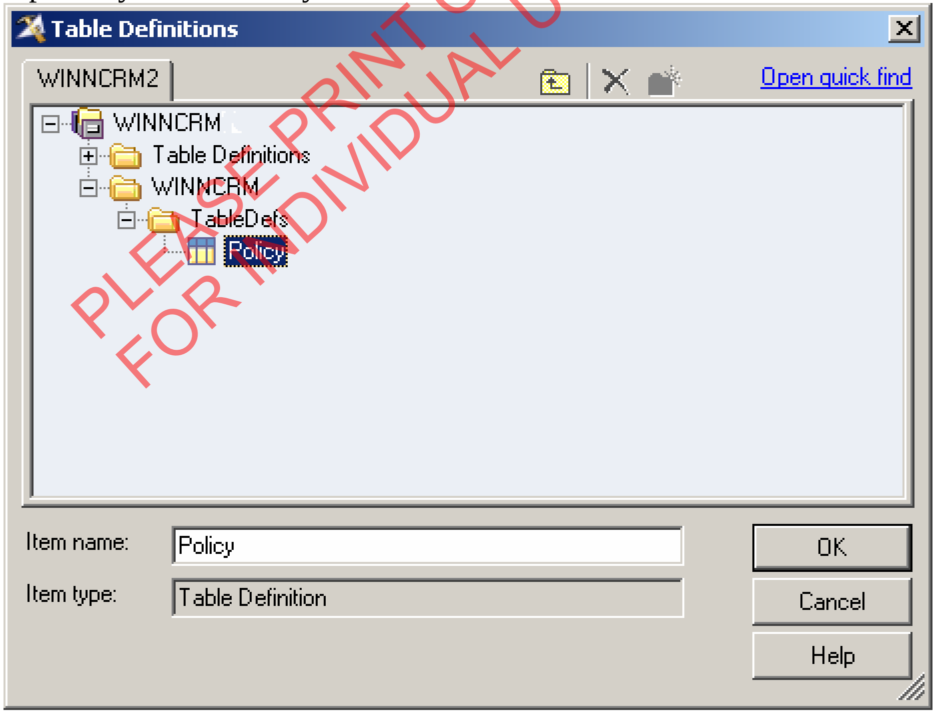
11. Click the Columns tab and then click the Load button.



Output name: Policy

	Column name	Key	SQL type	Length	Scale	Nullable	Description
		□					

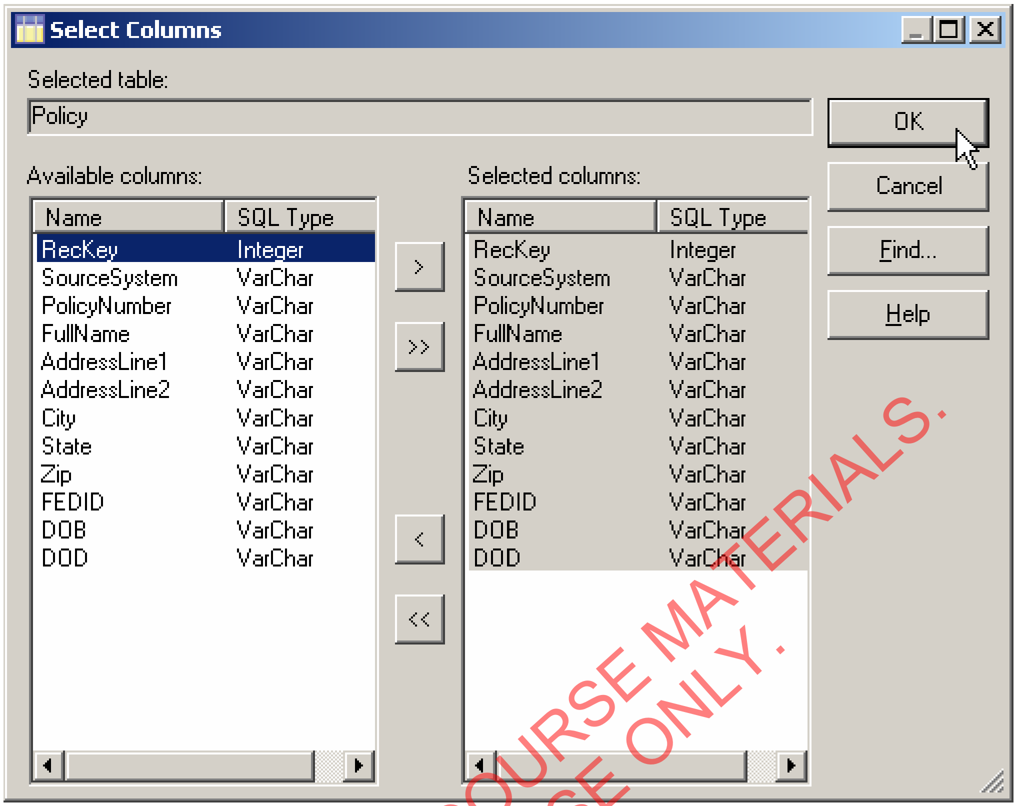
12. Locate the Policy table definition in the WINNCRM ➔ TableDefs folder of the repository. Select Policy and then click the OK button.



Item name: Policy

Item type: Table Definition

13. Click OK to select all columns.



Selected table: Policy

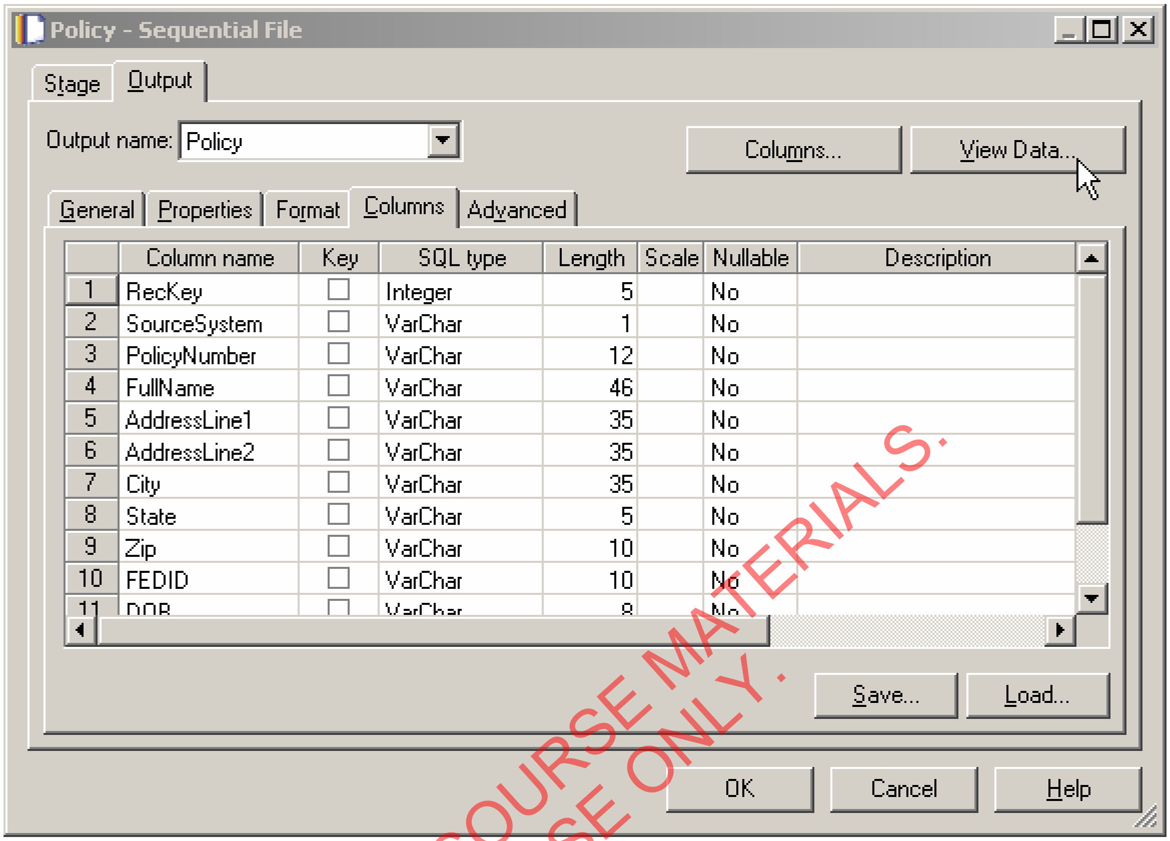
Available columns:

Name	SQL Type
RecKey	Integer
SourceSystem	VarChar
PolicyNumber	VarChar
FullName	VarChar
AddressLine1	VarChar
AddressLine2	VarChar
City	VarChar
State	VarChar
Zip	VarChar
FEDID	VarChar
DOB	VarChar
DOD	VarChar

Selected columns:

Name	SQL Type
RecKey	Integer
SourceSystem	VarChar
PolicyNumber	VarChar
FullName	VarChar
AddressLine1	VarChar
AddressLine2	VarChar
City	VarChar
State	VarChar
Zip	VarChar
FEDID	VarChar
DOB	VarChar
DOD	VarChar

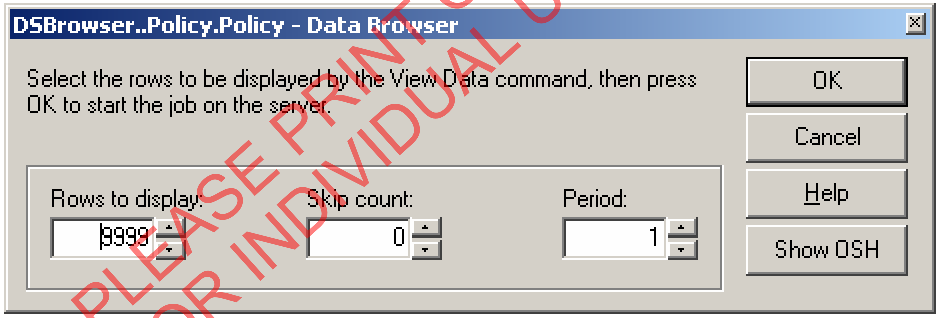
14. From the Columns tab click the View Data button.



Output name: Policy

	Column name	Key	SQL type	Length	Scale	Nullable	Description
1	RecKey	□	Integer	5		No	
2	SourceSystem	□	VarChar	1		No	
3	PolicyNumber	□	VarChar	12		No	
4	FullName	□	VarChar	46		No	
5	AddressLine1	□	VarChar	35		No	
6	AddressLine2	□	VarChar	35		No	
7	City	□	VarChar	35		No	
8	State	□	VarChar	5		No	
9	Zip	□	VarChar	10		No	
10	FEDID	□	VarChar	10		No	
11	DOB	□	VarChar	8		No	

15. Click OK in the data browser window. Note the control windows.



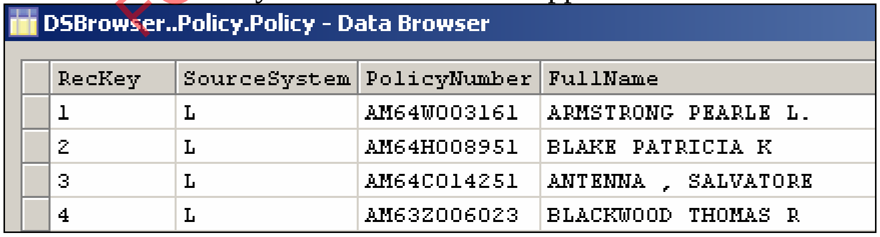
Select the rows to be displayed by the View Data command, then press OK to start the job on the server.

Rows to display: 9999

Skip count: 0

Period: 1

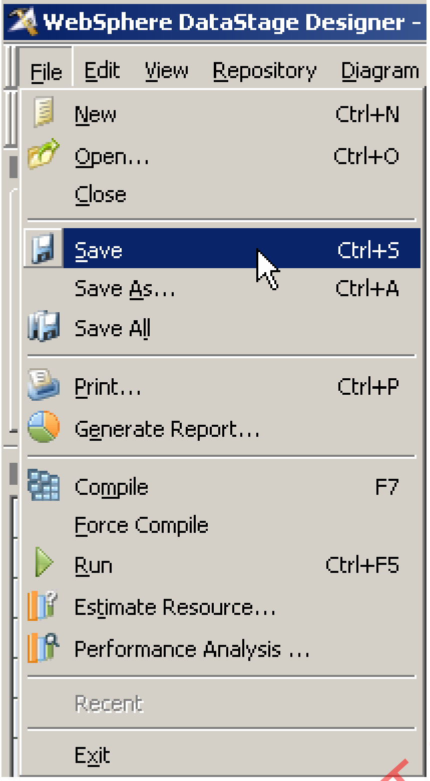
16. Data from the Policy.dat file should now appear.



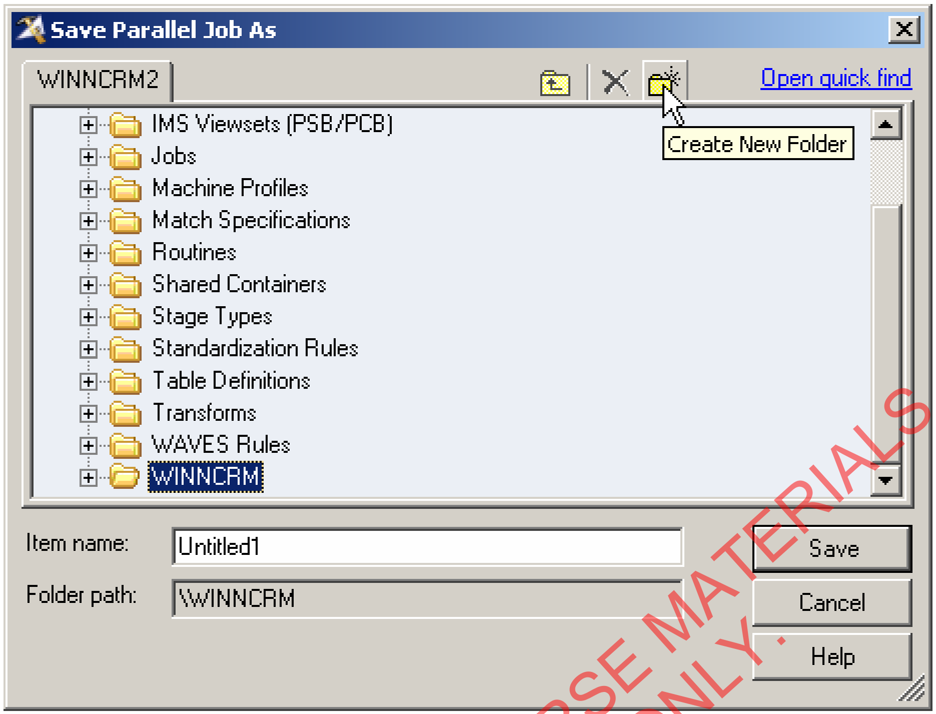
RecKey	SourceSystem	PolicyNumber	FullName
1	L	AM64W003161	ARMSTRONG PEARLE L.
2	L	AM64H008951	BLAKE PATRICIA K
3	L	AM64C014251	ANTENNA , SALVATORE
4	L	AM63Z006023	BLACKWOOD THOMAS R

The View Data button is available on many data stages – both source and target - and should be used to verify connectivity to the data. Note that it will work only on data that is actually present, so if you wish to use it on a target file you will need to first run the job to create the data file.

17. Click the Close button to exit the data browser and then click OK to return to the job canvas.
18. Click File ➔ Save from the file menu.



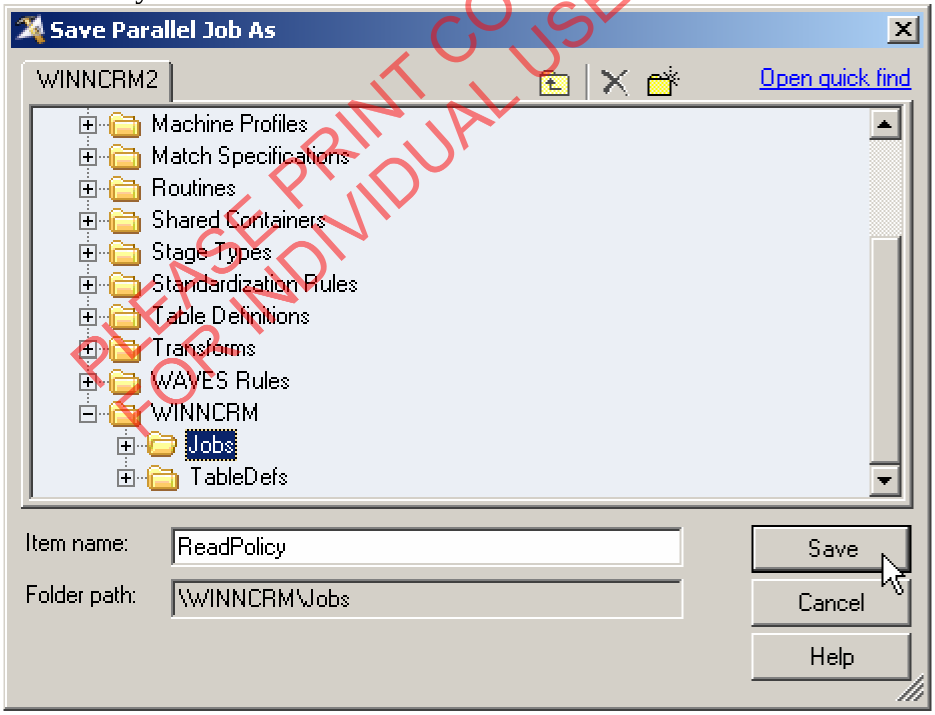
19. Select the WINNCRM folder and then click the Create New Folder icon. Create a folder called Jobs.



Item name: Untitled1

Folder path: \WINNCRM

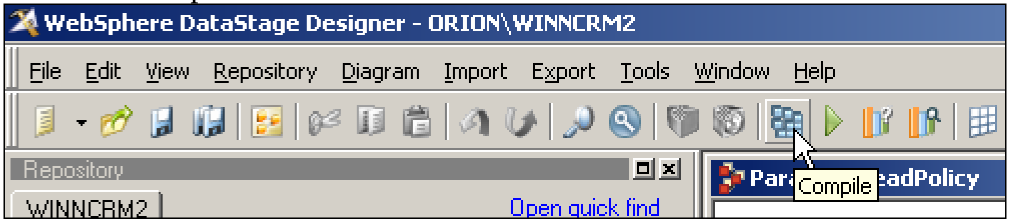
20. Click the Save button to store it in the WINNCRM ➔ Jobs folder and name it ReadPolicy.



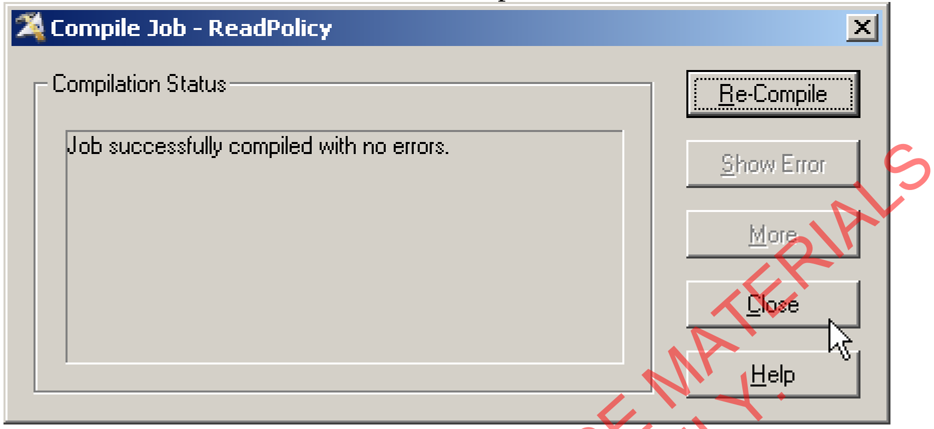
Item name: ReadPolicy

Folder path: \WINNCRM\Jobs

21. Click the Compile icon under the menu bar.


WebSphere DataStage Designer - ORION\WINNCRM2
File Edit View Repository Diagram Import Export Tools Window Help
Repository
WINNCRM2
Open quick find
Para
Compile
adPolicy

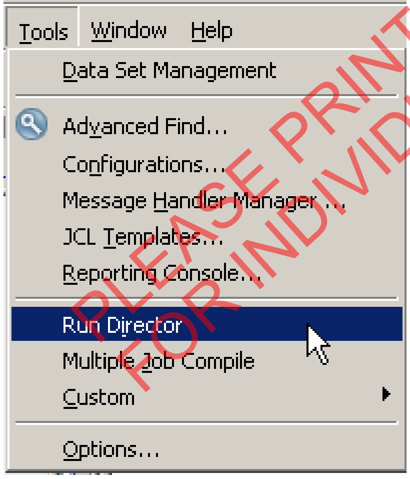
22. Click the Close button to exit the Compile Status screen.



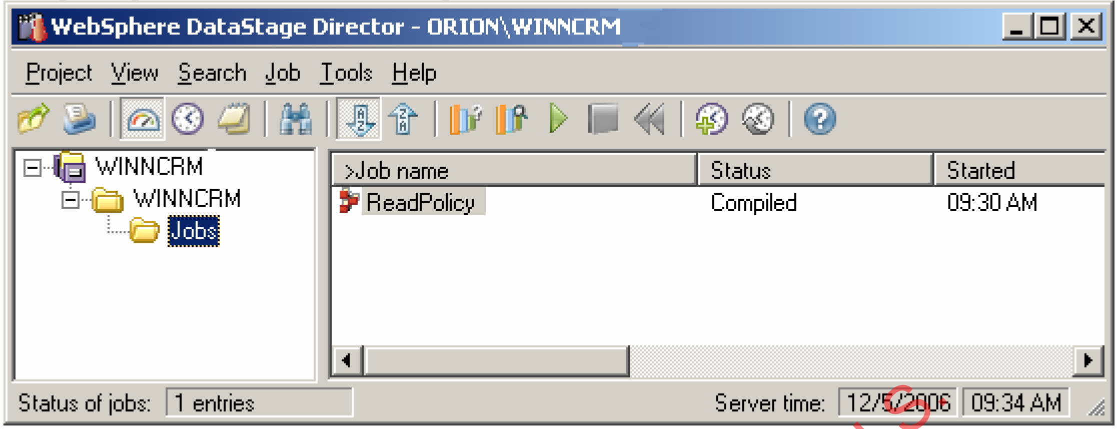
Job successfully compiled with no errors.

Task: Run DataStage job using Director Client

1. From within the Designer Client click Tools ➔ Run Director on the menu bar.



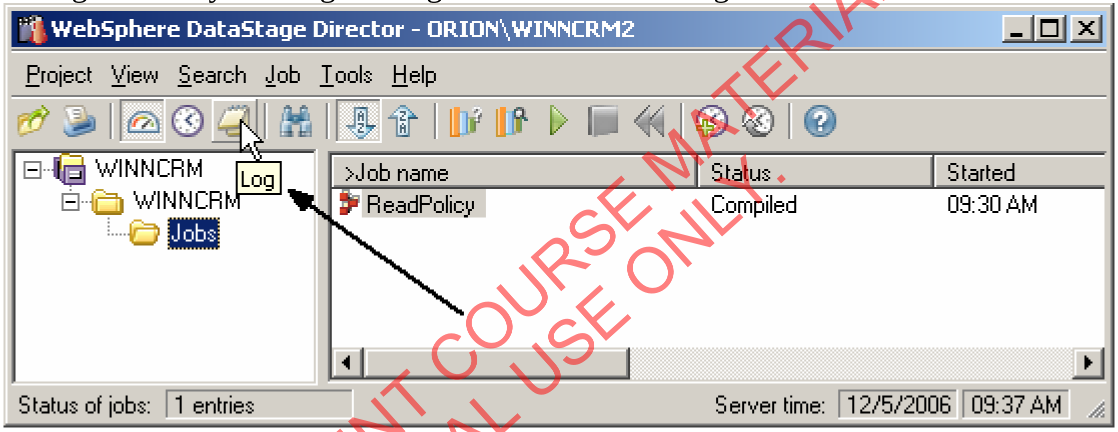
2. The default setting for Director is the Status view. It will show that ReadPolicy is compiled plus the time of the compile.


WebSphere DataStage Director - ORION\WINNCRM
Project View Search Job Tools Help
WINNCRM
WINNCRM
Jobs

>Job name	Status	Started
ReadPolicy	Compiled	09:30 AM

Status of jobs: 1 entries Server time: 12/5/2006 09:34 AM

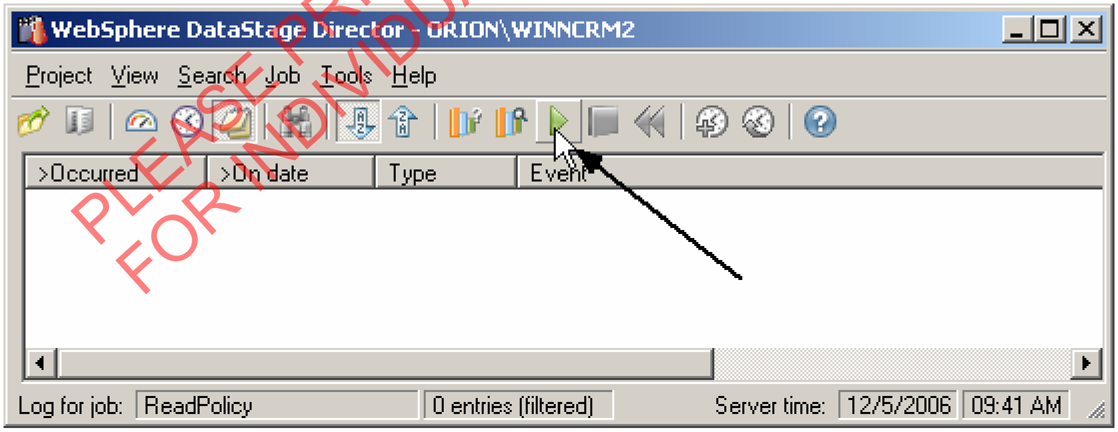
3. Change views by clicking the Log icon located to the right of the status view.


WebSphere DataStage Director - ORION\WINNCRM2
Project View Search Job Tools Help
Log
WINNCRM
WINNCRM
Jobs

>Job name	Status	Started
ReadPolicy	Compiled	09:30 AM

Status of jobs: 1 entries Server time: 12/5/2006 09:37 AM

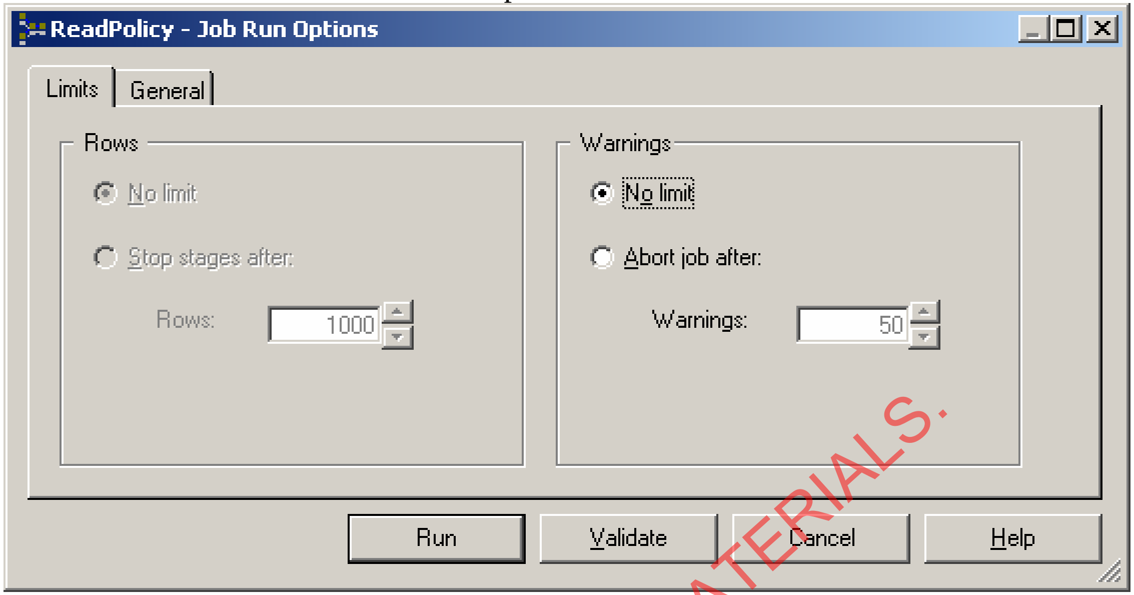
4. Click the Run Now icon.


WebSphere DataStage Director - ORION\WINNCRM2
Project View Search Job Tools Help

>Occurred	>On date	Type	Event

Log for job: ReadPolicy 0 entries (filtered) Server time: 12/5/2006 09:41 AM

5. Click the Run button on the Job Run Options screen.



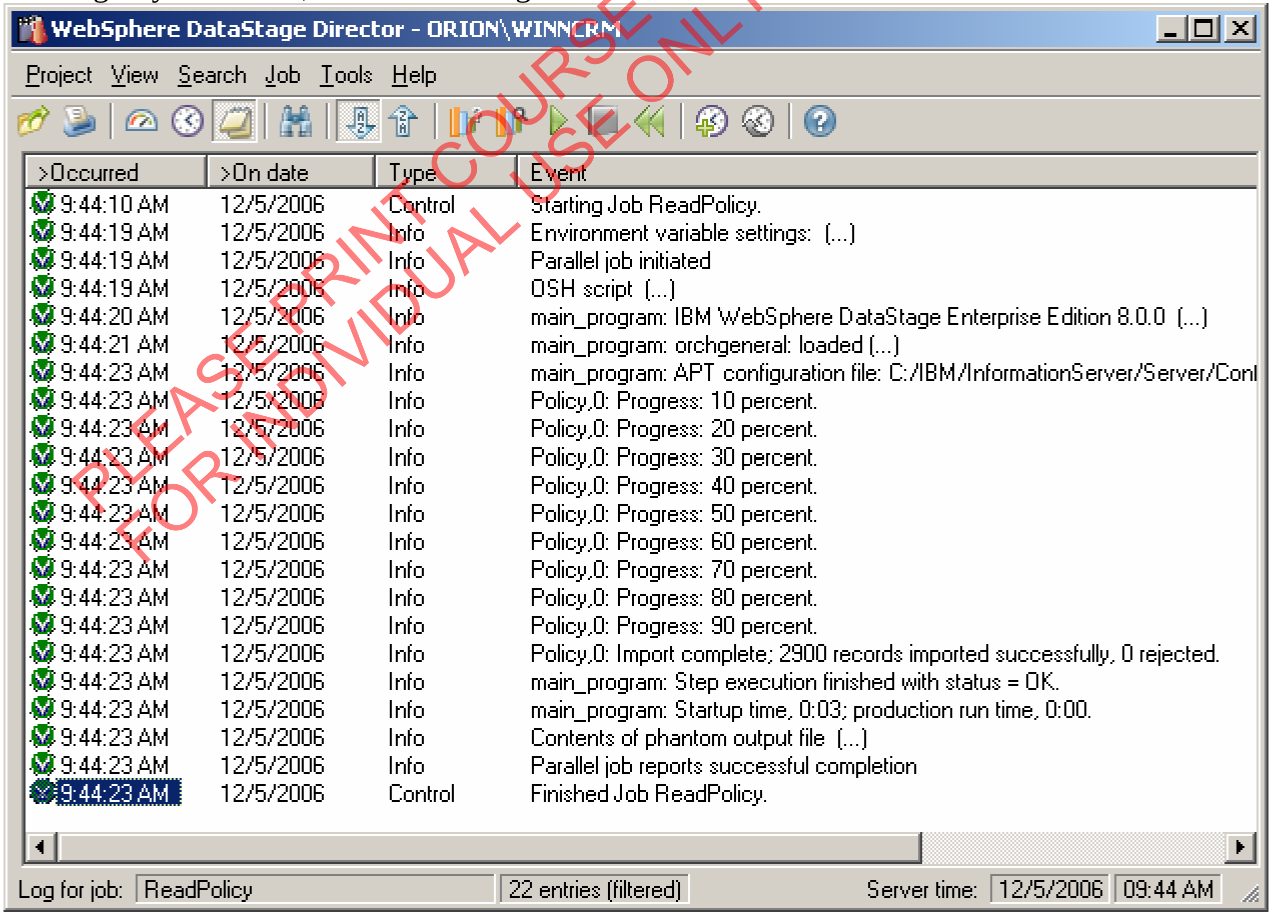
ReadPolicy - Job Run Options

Limits | General

Rows
- ● No limit
- ○ Stop stages after:
- Rows: 1000

Warnings
- ● No limit
- ○ Abort job after:
- Warnings: 50

6. Job log messages will appear. Information messages will have a leading green icon, warnings a yellow icon, and fatal messages will have a red icon.

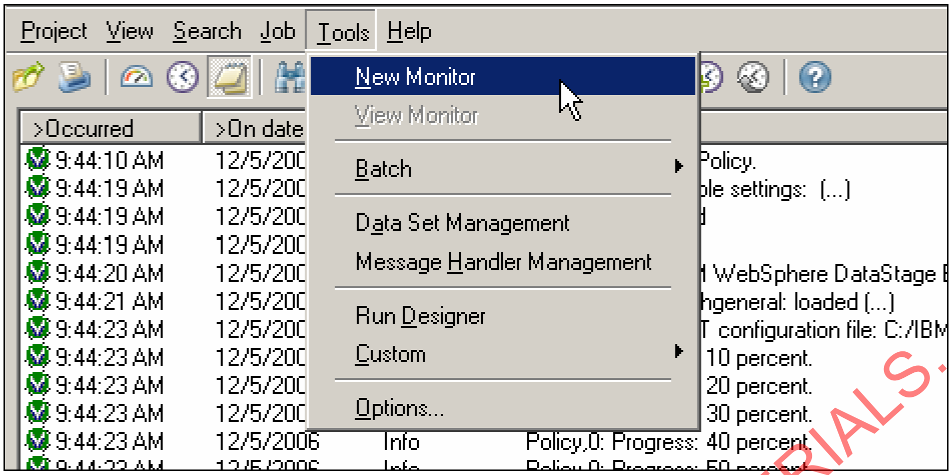


WebSphere DataStage Director - ORION\WINNCRM

```
>Occurred    >On date    Type     Event
9:44:10 AM   12/5/2006   Control  Starting Job ReadPolicy.
9:44:19 AM   12/5/2006   Info     Environment variable settings:  (...)
9:44:19 AM   12/5/2006   Info     Parallel job initiated
9:44:19 AM   12/5/2006   Info     OSH script  (...)
9:44:20 AM   12/5/2006   Info     main_program: IBM WebSphere DataStage Enterprise Edition 8.0.0  (...)
9:44:21 AM   12/5/2006   Info     main_program: orchgeneral: loaded (...)
9:44:23 AM   12/5/2006   Info     main_program: APT configuration file: C:/IBM/InformationServer/Server/Conf
9:44:23 AM   12/5/2006   Info     Policy,0: Progress: 10 percent.
9:44:23 AM   12/5/2006   Info     Policy,0: Progress: 20 percent.
9:44:23 AM   12/5/2006   Info     Policy,0: Progress: 30 percent.
9:44:23 AM   12/5/2006   Info     Policy,0: Progress: 40 percent.
9:44:23 AM   12/5/2006   Info     Policy,0: Progress: 50 percent.
9:44:23 AM   12/5/2006   Info     Policy,0: Progress: 60 percent.
9:44:23 AM   12/5/2006   Info     Policy,0: Progress: 70 percent.
9:44:23 AM   12/5/2006   Info     Policy,0: Progress: 80 percent.
9:44:23 AM   12/5/2006   Info     Policy,0: Progress: 90 percent.
9:44:23 AM   12/5/2006   Info     Policy,0: Import complete; 2900 records imported successfully, 0 rejected.
9:44:23 AM   12/5/2006   Info     main_program: Step execution finished with status = OK.
9:44:23 AM   12/5/2006   Info     main_program: Startup time, 0:03; production run time, 0:00.
9:44:23 AM   12/5/2006   Info     Contents of phantom output file  (...)
9:44:23 AM   12/5/2006   Info     Parallel job reports successful completion
9:44:23 AM   12/5/2006   Control  Finished Job ReadPolicy.
```

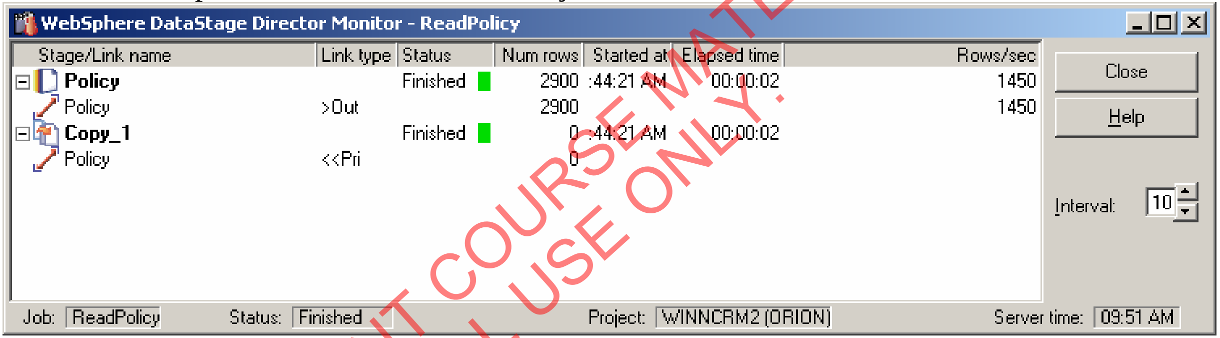
Log for job: ReadPolicy | 22 entries (filtered) | Server time: 12/5/2006 09:44 AM

7. Double-click one of the messages in the log to bring up a message details screen. Click the Close button to exit the message detail screen.
8. Click Tools ➔ New Monitor.



```
>Occurred      >On date
9:44:10 AM     12/5/200                           Policy.
9:44:19 AM     12/5/200                           ble settings:  (...)
9:44:19 AM     12/5/200                           d
9:44:19 AM     12/5/200
9:44:20 AM     12/5/200                           l WebSphere DataStage E
9:44:21 AM     12/5/200                           hgeneral: loaded (...)
9:44:23 AM     12/5/200                           T configuration file: C:/IBM
9:44:23 AM     12/5/200                           10 percent.
9:44:23 AM     12/5/200                           20 percent.
9:44:23 AM     12/5/200                           30 percent.
9:44:23 AM     12/5/2006    Info      Policy,0: Progress: 40 percent.
```

9. The monitor reports record counts for each job link.


WebSphere DataStage Director Monitor - ReadPolicy

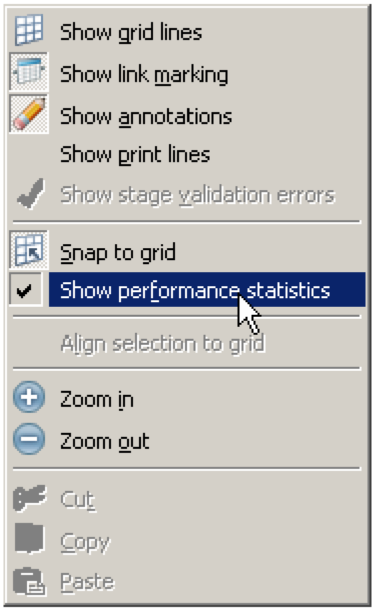
Stage/Link name	Link type	Status	Num rows	Started at	Elapsed time	Rows/sec
Policy		Finished 🟩	2900	:44:21 AM	00:00:02	1450
Policy	>Out		2900			1450
Copy_1		Finished 🟩	0	:44:21 AM	00:00:02	
Policy	<<Pri		0			

Interval: 10

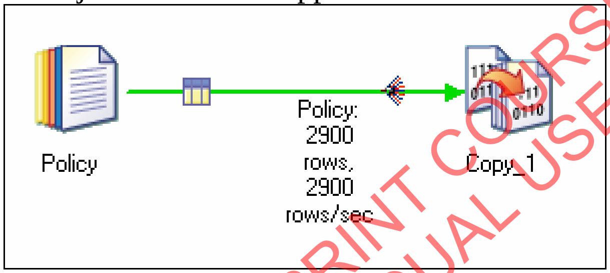
Job: ReadPolicy Status: Finished Project: WINNCRM2 (ORION) Server time: 09:51 AM

10. Click the Close button to return to the Job log view.

11. Return to the Designer Client and view the ReadPolicy job. Right-click on the canvas and select Show Performance Statistics.

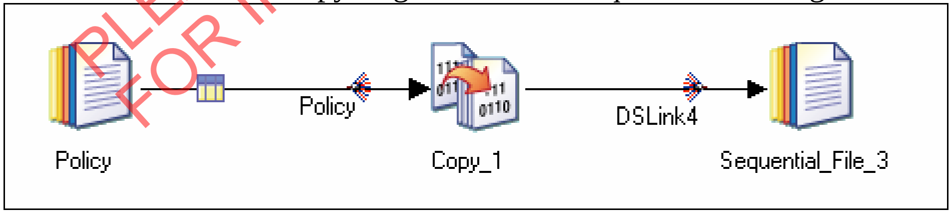


12. Your job should now appear with row count statistics on the link.



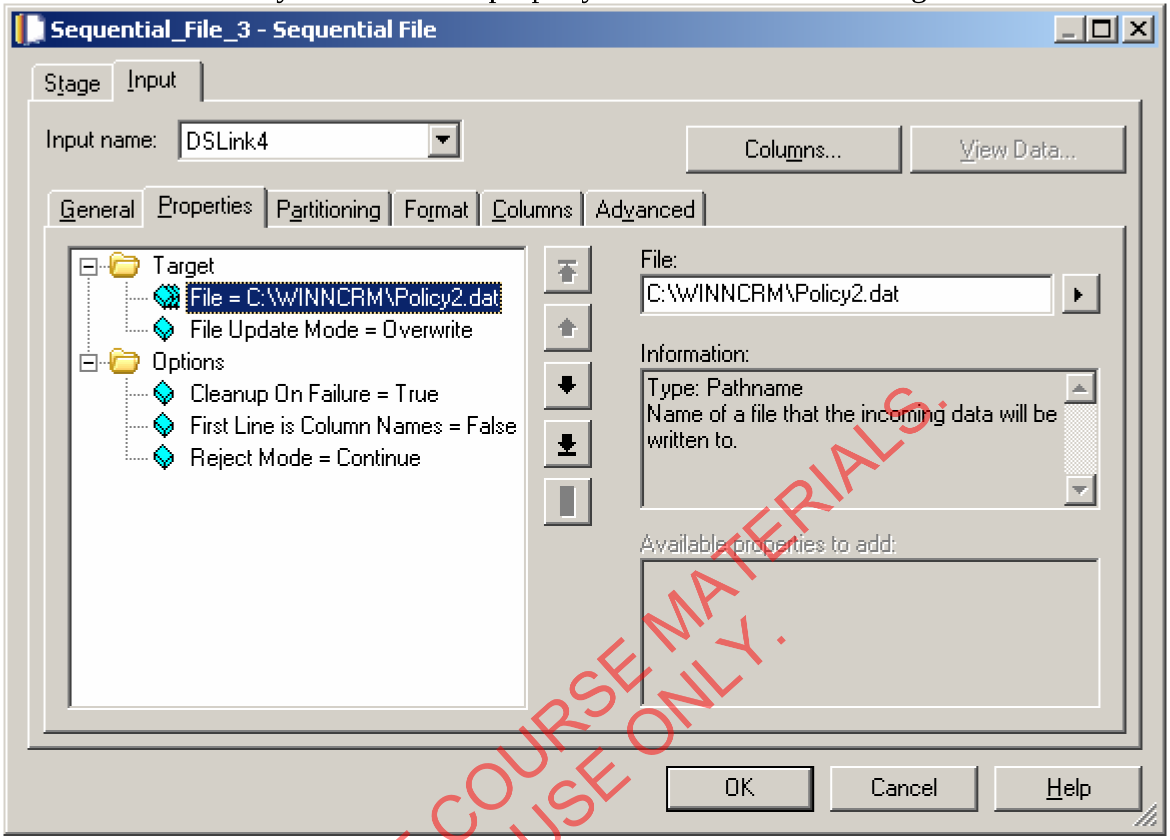
Task: Modify job

1. Using the Palette, drag another Sequential File stage onto the canvas.
2. Draw a link from the Copy stage to the new Sequential File stage.



3. Rename the new link to Policy2 and the new Sequential File stage to Policy2.

4. Open the target Policy2 Sequential File stage and complete the File property to create C:\WINNCRM\Policy2.dat. Note the property folder is now called Target.



Input name: DSLink4

- Target
  - File = C:\WINNCRM\Policy2.dat
  - File Update Mode = Overwrite
- Options
  - Cleanup On Failure = True
  - First Line is Column Names = False
  - Reject Mode = Continue

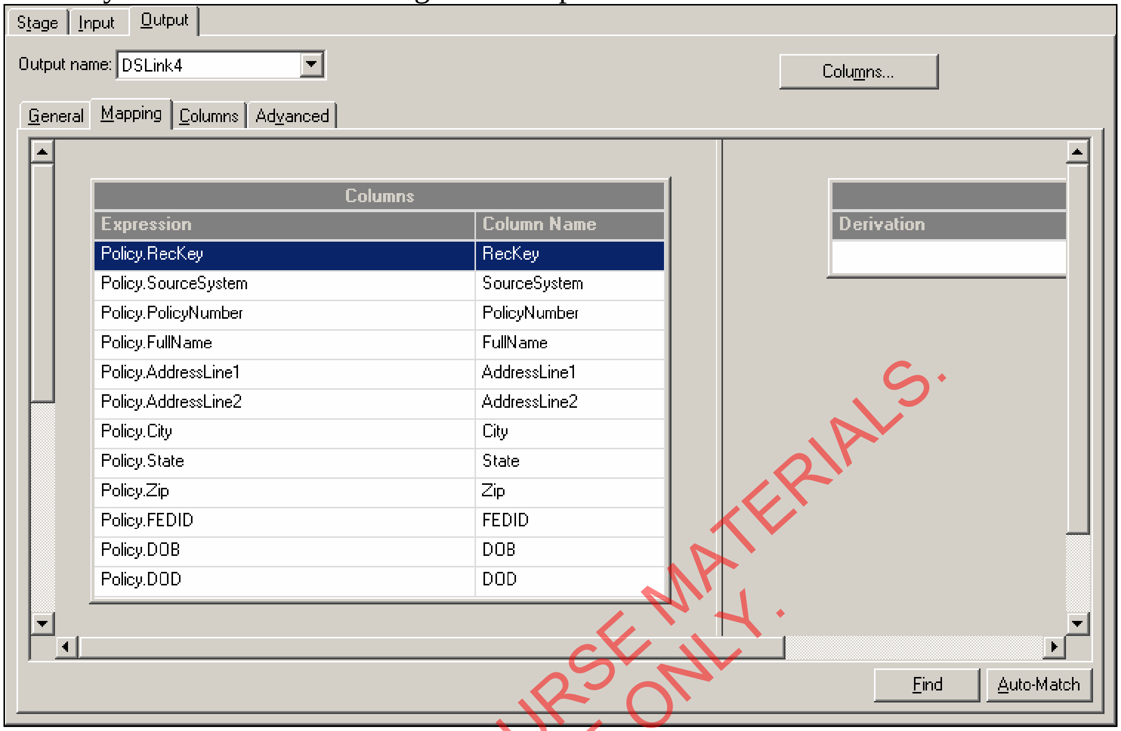
File: C:\WINNCRM\Policy2.dat

Information:
Type: Pathname
Name of a file that the incoming data will be written to.

Available properties to add:

5. Click OK to return to the job canvas.
6. Open the Copy stage.

7. Click the Output ➔ Mappings tabs. This shows how input columns get to the output link; currently no columns are flowing to the output link.

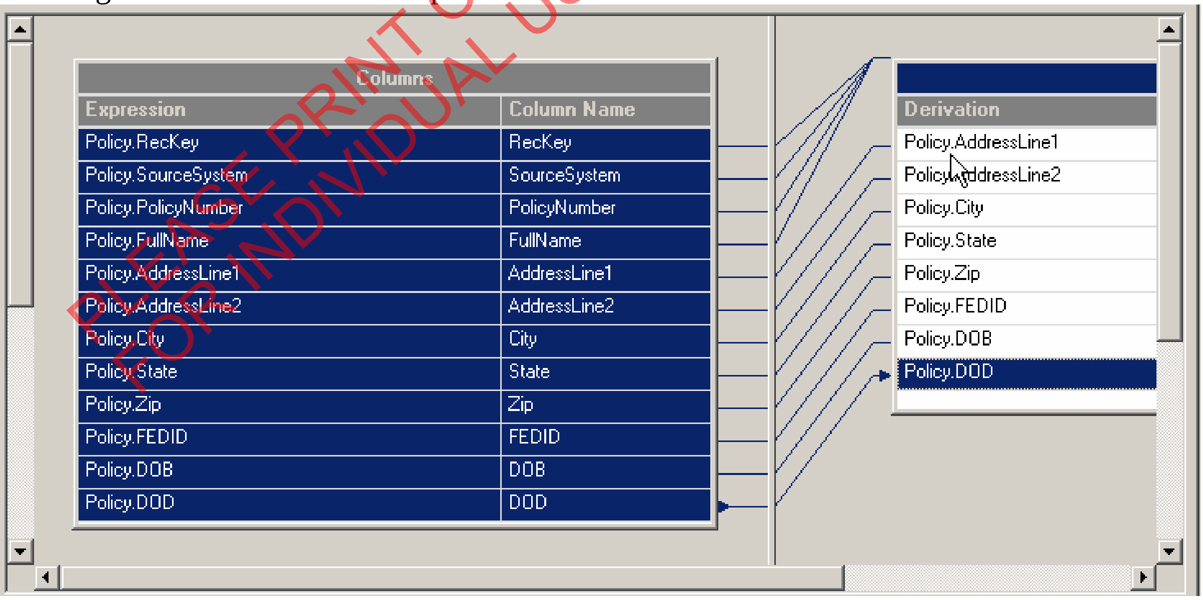


Output name: DSLink4

Columns	
Expression	Column Name
Policy.RecKey	RecKey
Policy.SourceSystem	SourceSystem
Policy.PolicyNumber	PolicyNumber
Policy.FullName	FullName
Policy.AddressLine1	AddressLine1
Policy.AddressLine2	AddressLine2
Policy.City	City
Policy.State	State
Policy.Zip	Zip
Policy.FEDID	FEDID
Policy.DOB	DOB
Policy.DOD	DOD

Derivation

8. Using Windows shift/click selection techniques, select all columns from the input link and drag these columns to the output columns box.



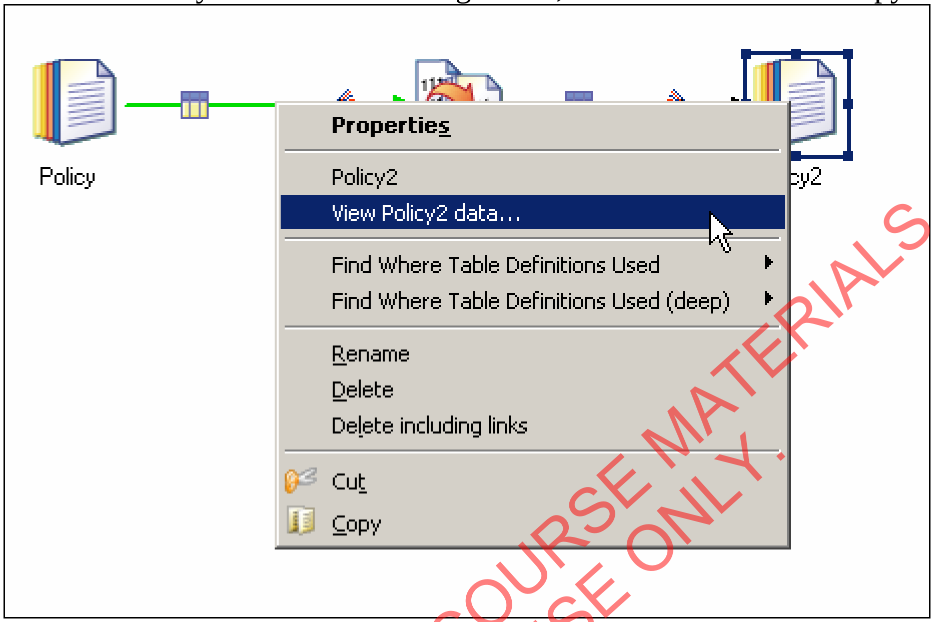
Columns	
Expression	Column Name
Policy.RecKey	RecKey
Policy.SourceSystem	SourceSystem
Policy.PolicyNumber	PolicyNumber
Policy.FullName	FullName
Policy.AddressLine1	AddressLine1
Policy.AddressLine2	AddressLine2
Policy.City	City
Policy.State	State
Policy.Zip	Zip
Policy.FEDID	FEDID
Policy.DOB	DOB
Policy.DOD	DOD

Derivation
Policy.AddressLine1
Policy.AddressLine2
Policy.City
Policy.State
Policy.Zip
Policy.FEDID
Policy.DOB
Policy.DOD

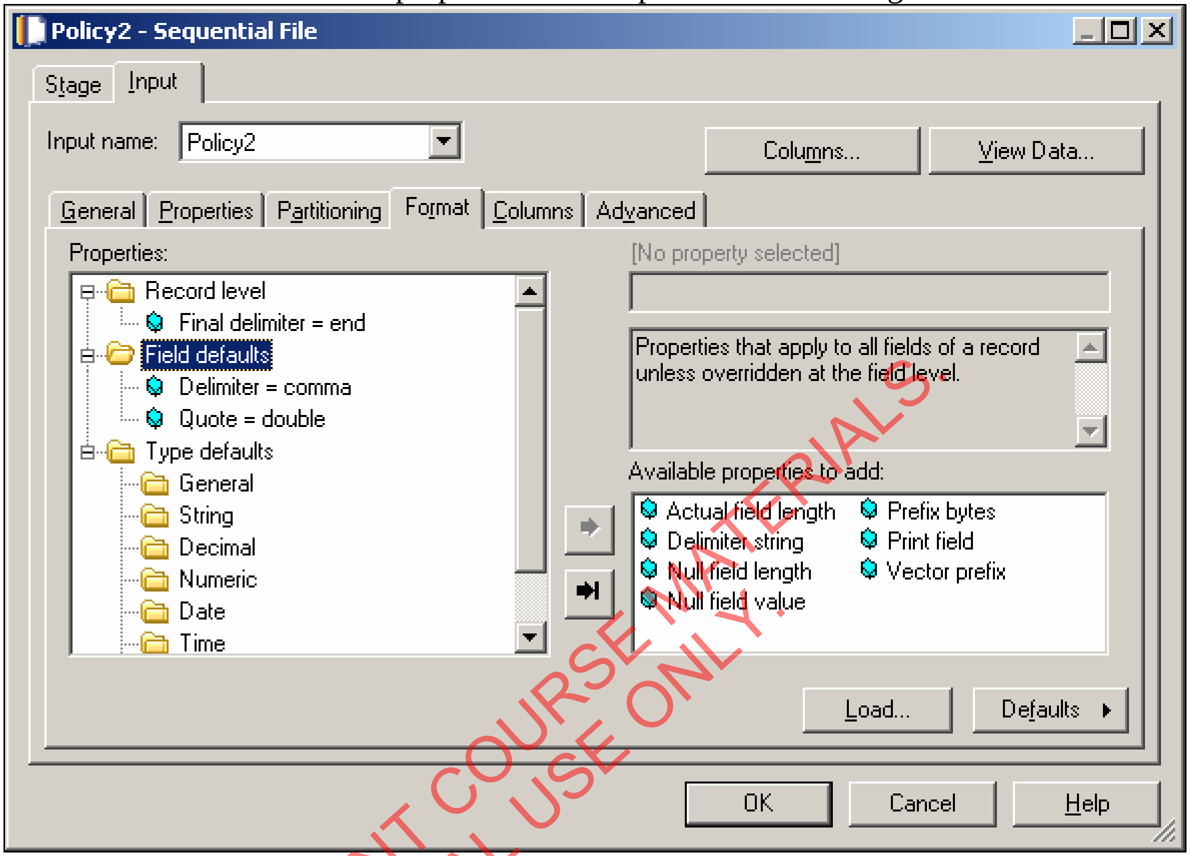
NOTE: Column mappings can also be created by other means. For instance, you could right-click on the Copy stage while on the canvas and choose the propagate columns option.

9. Close the Copy stage and open the target Sequential File stage.

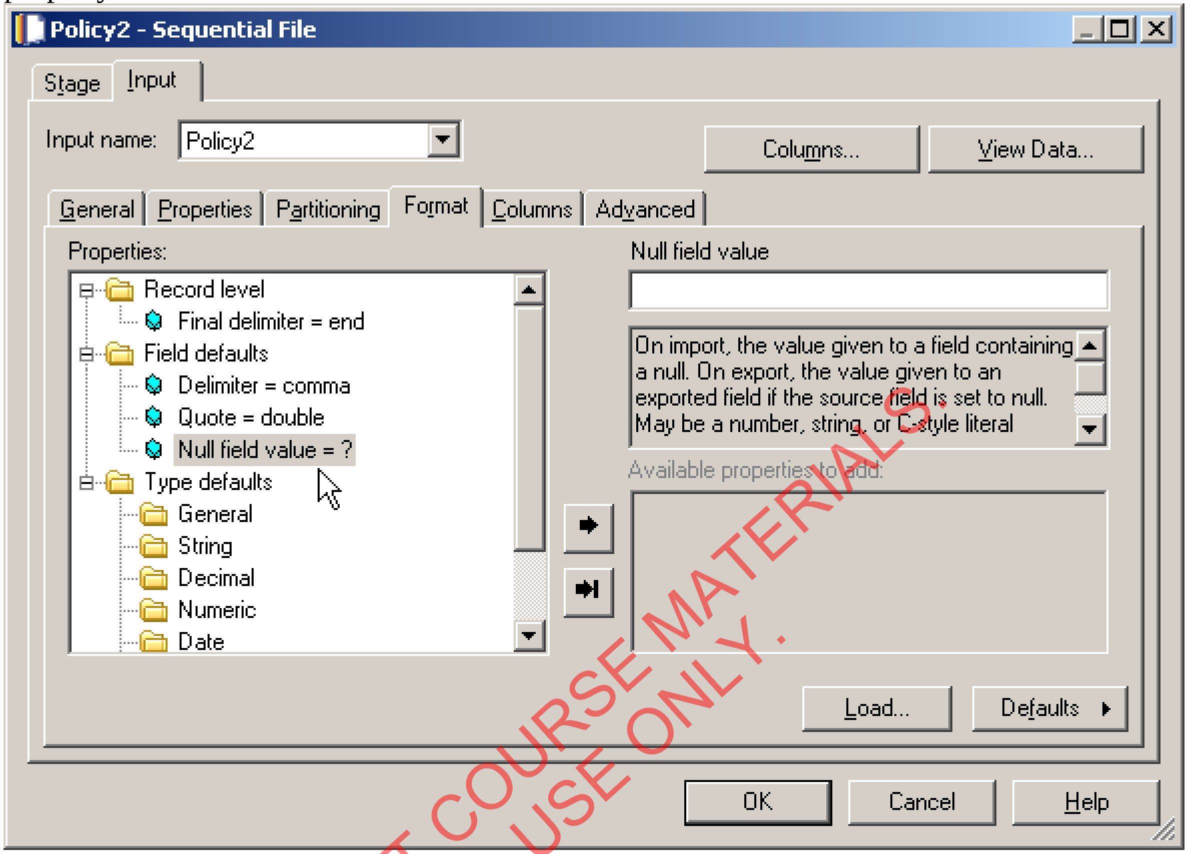
10. Click the Columns tab. Note that the columns are present although you did not explicitly load these from the repository. Rather, the mappings you created in the Copy stage also populated the Columns in the target Sequential File stage.
11. Return to the job canvas.
12. Save, compile, and run your job.
13. From the job canvas, right-click on the target Sequential File stage and click View Data. Continue until you can view the target data; it should be an exact copy of the source data.



14. Open the target Sequential File stage and click the Format tab. Select the Field Defaults folder and note the ‘Available properties to add:’ pane in the lower right.



15. Click the Null field value property from the available properties. This will add the property to the Field defaults folder.



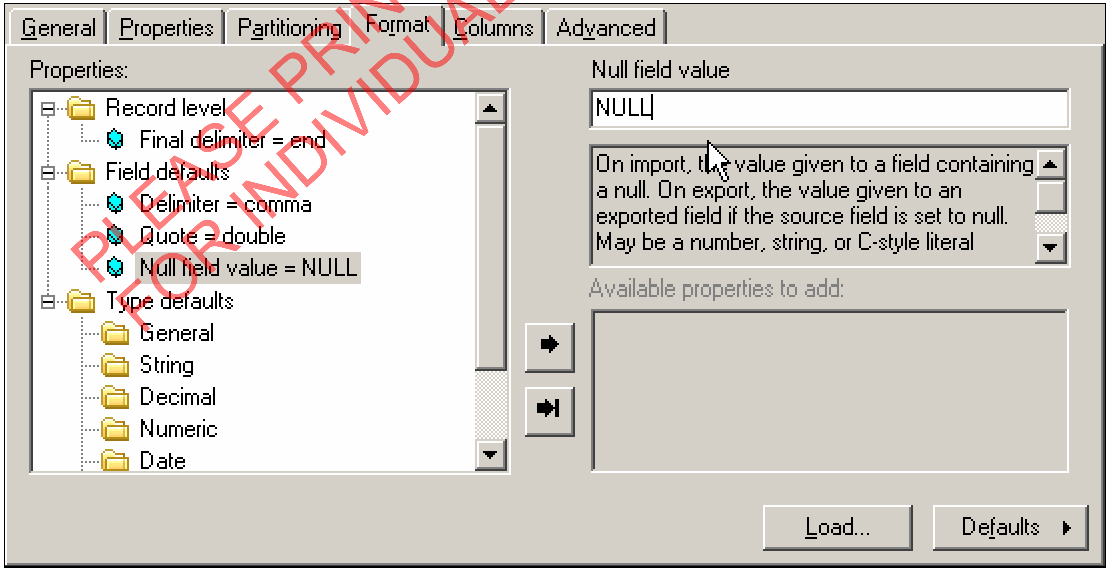
Input name: Policy2

Properties:
- Record level
  - Final delimiter = end
- Field defaults
  - Delimiter = comma
  - Quote = double
  - Null field value = ?
- Type defaults
  - General
  - String
  - Decimal
  - Numeric
  - Date

Null field value

On import, the value given to a field containing a null. On export, the value given to an exported field if the source field is set to null. May be a number, string, or C-style literal

16. Enter NULL into the Null field value text box.



Properties:
- Record level
  - Final delimiter = end
- Field defaults
  - Delimiter = comma
  - Quote = double
  - Null field value = NULL
- Type defaults
  - General
  - String
  - Decimal
  - Numeric
  - Date

Null field value: NULL

On import, the value given to a field containing a null. On export, the value given to an exported field if the source field is set to null. May be a number, string, or C-style literal

DataStage jobs that create target sequential files should always account for the possibility of producing fields with nulls. The technique shown here should always be used; it will replace the null character with the text value NULL (you could make this value any text string as long as you understand that it represents null).

17. Save, compile, and run your job. Did anything change?

Lab 6: Investigate – Character Discrete with C Mask

Assumptions:

- You have a DataStage project named WINNCRM.
- You have the Policy table definition in the repository

Task: Build and run an investigate Data Quality job

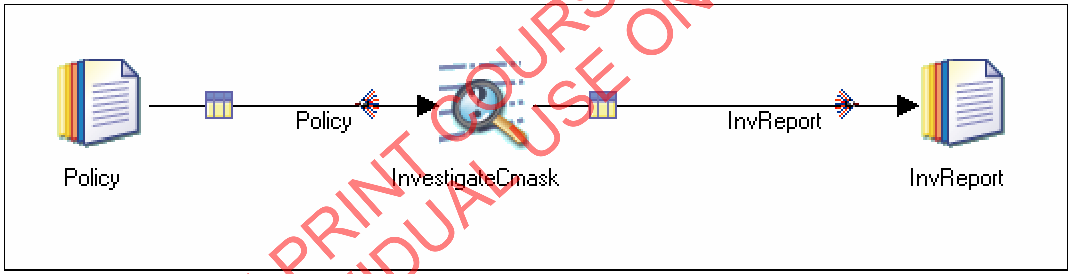
Recall the meaning of "investigate".

Investigate

Parses and analyzes *free-form* and *single-domain* columns by determining the number and frequency of unique values and classifying or assigning a business meaning to each occurrence of a value within a column. If columns contain more than one data element such as part of a name and address, then it is a free-form column. If the column contains only one data element, then it is a single-domain column.

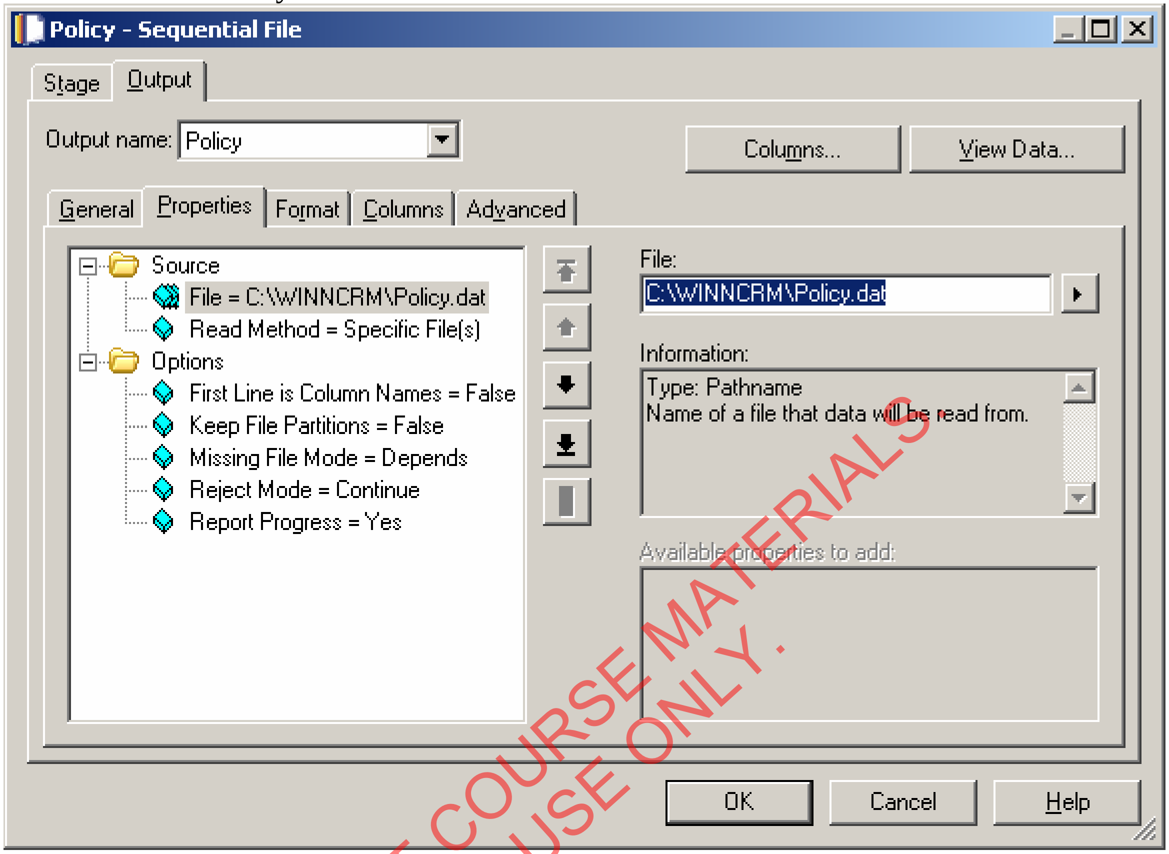
This job will perform a character discrete investigation using a C mask on the SourceSystem, PolicyNumber, FEDID, DOB, and DOD fields of the Policy.dat file.

1. Create a new parallel job named InvCharDisCmask. Name the links and stages as shown below. Note: The Investigate stage can be found in the Data Quality subfolder in the Palette.



2. Open the Policy Sequential File stage and enter the file to be read (or use the browse function to locate it)

C:\WINNCRM\Policy.dat



Output name: Policy

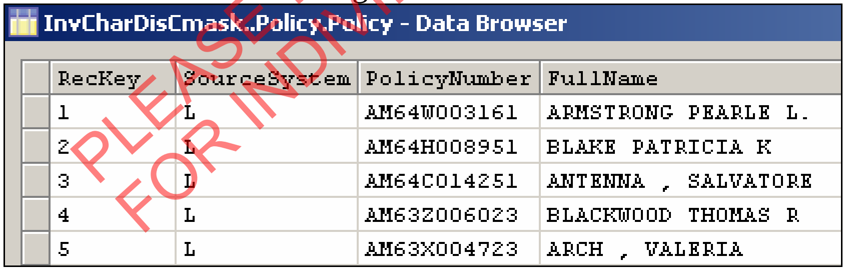
- Source
  - File = C:\WINNCRM\Policy.dat
  - Read Method = Specific File(s)
- Options
  - First Line is Column Names = False
  - Keep File Partitions = False
  - Missing File Mode = Depends
  - Reject Mode = Continue
  - Report Progress = Yes

File: C:\WINNCRM\Policy.dat

Information:
Type: Pathname
Name of a file that data will be read from.

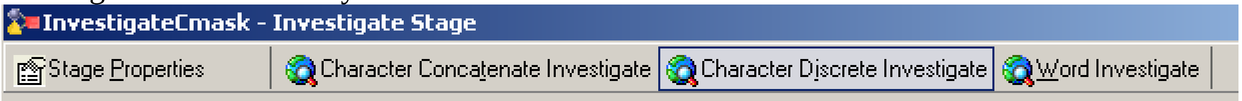
Available properties to add:

3. Click on the Columns tab and click Load.
4. Load all columns from the Policy table located in the WINNCRM ➔ TableDefs folder. Accept all columns.
5. Validate the meta data using the View Data button.

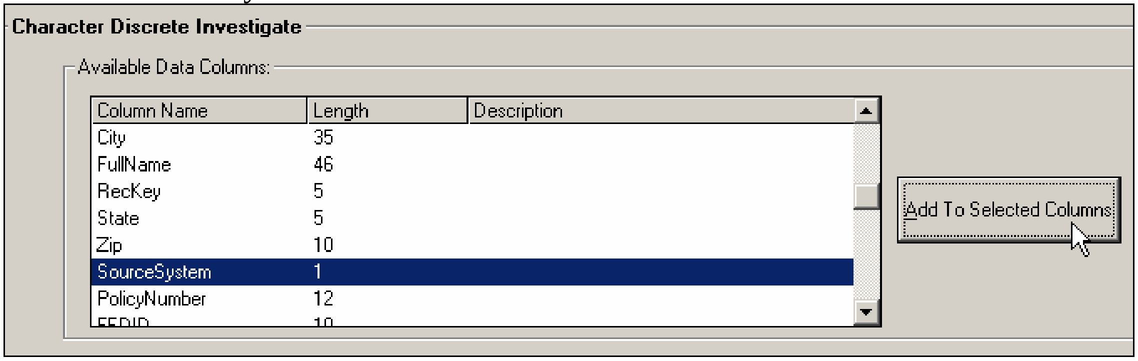


	RecKey	SourceSystem	PolicyNumber	FullName
	1	L	AM64W003161	ARMSTRONG PEARLE L.
	2	L	AM64H008951	BLAKE PATRICIA K
	3	L	AM64C014251	ANTENNA , SALVATORE
	4	L	AM63Z006023	BLACKWOOD THOMAS R
	5	L	AM63X004723	ARCH , VALERIA

6. Close the Sequential File stage.
7. Open the InvestigateCmask Investigate stage. The default tab - Character Discrete Investigate - should already be selected.


InvestigateCmask - Investigate Stage
Stage Properties
Character Concatenate Investigate
Character Discrete Investigate
Word Investigate

8. Select the SourceSystem column and click the Add to Selected Columns button.

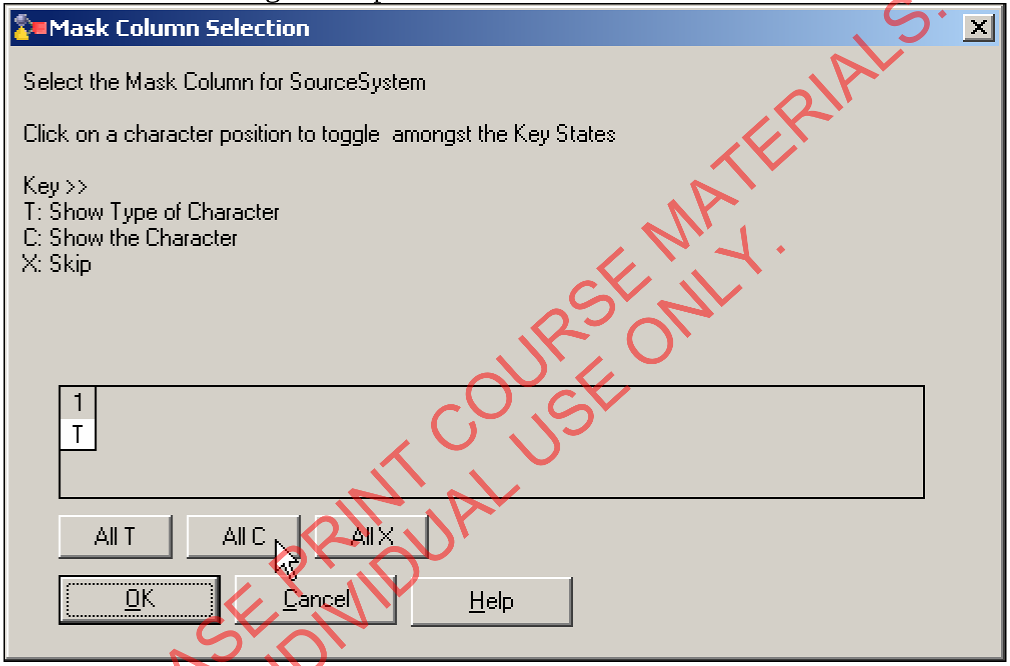


Character Discrete Investigate

Available Data Columns:

Column Name	Length	Description
City	35	
FullName	46	
RecKey	5	
State	5	
Zip	10	
SourceSystem	1	
PolicyNumber	12	
FEDID	10	

9. The Mask Column Selection screen that appears will default to type T mask. Click the All C button to change this option and then click the OK button.



Mask Column Selection

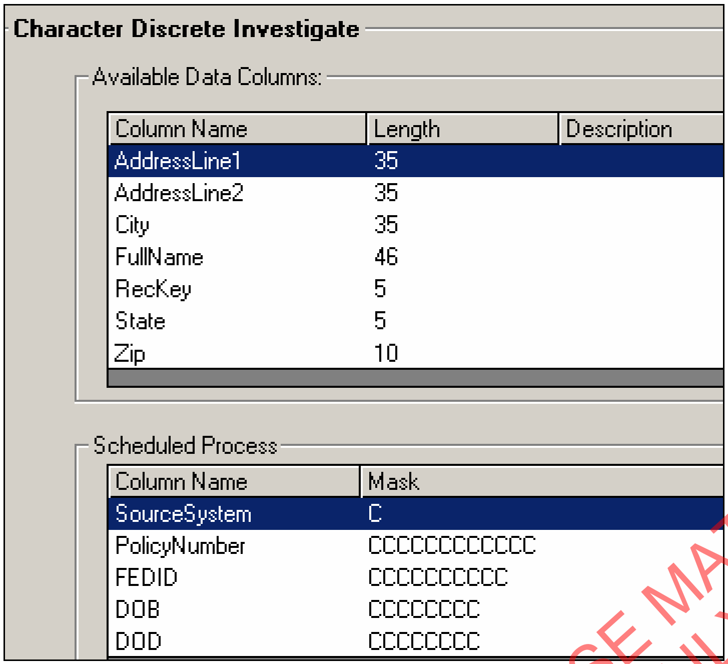
Select the Mask Column for SourceSystem

Click on a character position to toggle amongst the Key States

Key >>
T: Show Type of Character
C: Show the Character
X: Skip

1
T

10. In a similar fashion select columns PolicyNumber, FEDID, DOB, and DOD. Apply the All C mask to each column. Your screen should now resemble:



Character Discrete Investigate

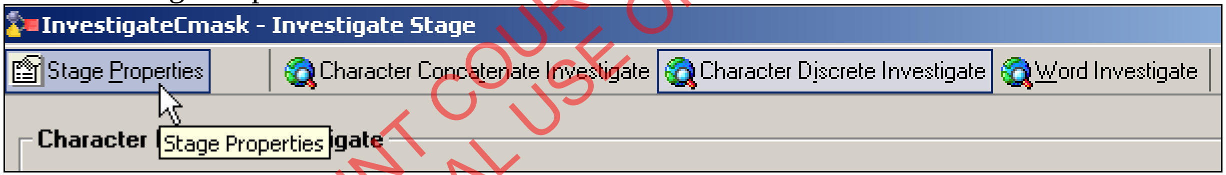
Available Data Columns:

Column Name	Length	Description
AddressLine1	35	
AddressLine2	35	
City	35	
FullName	46	
RecKey	5	
State	5	
Zip	10	

Scheduled Process

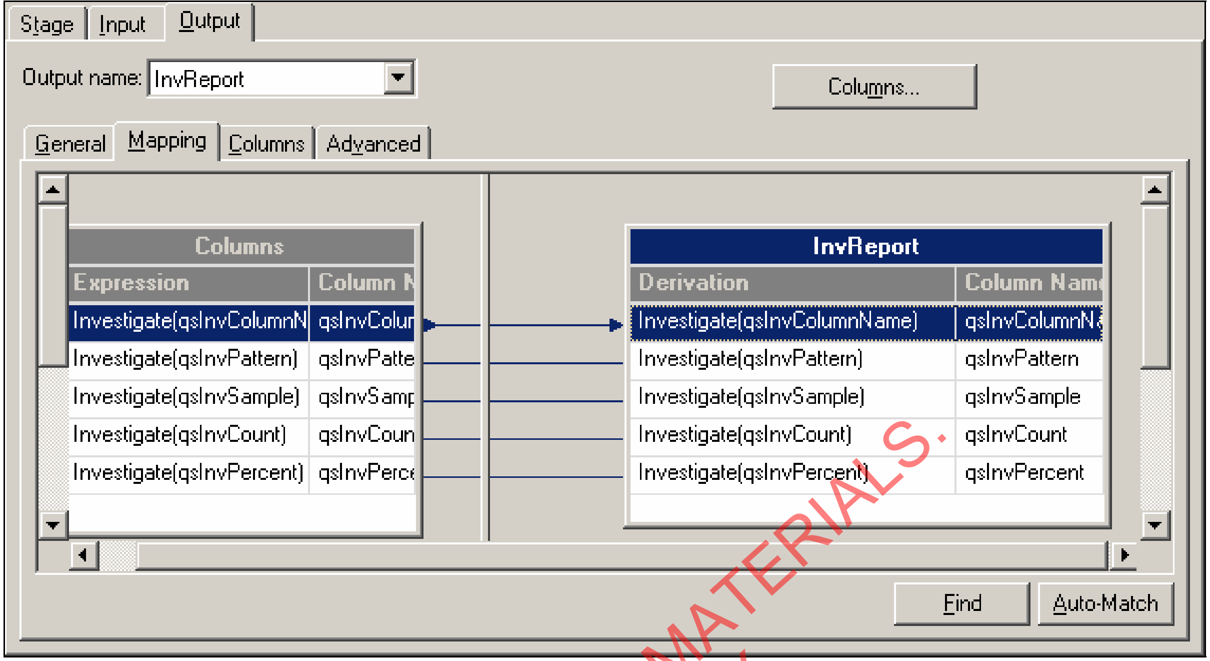
Column Name	Mask
SourceSystem	C
PolicyNumber	CCCCCCCCCCCC
FEDID	CCCCCCCCCC
DOB	CCCCCCCC
DOD	CCCCCCCC

11. Click the Stage Properties tab.


InvestigateCmask - Investigate Stage
Stage Properties
Character Concatenate Investigate
Character Discrete Investigate
Word Investigate
Character
Stage Properties
igate

12. Click the Output ➔ Mapping tabs.

13. Drag all input columns to the output link.

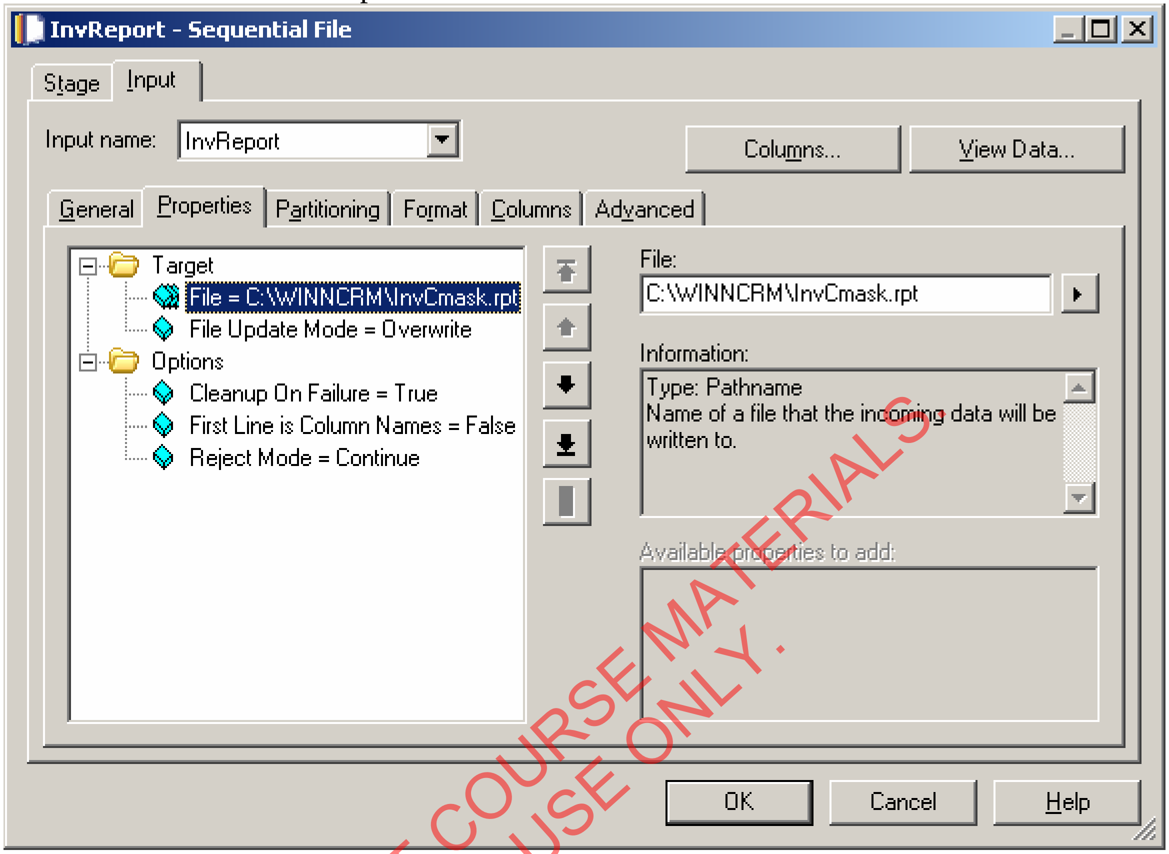


Columns	
Expression	Column N
Investigate(qsInvColumnN	qsInvColur
Investigate(qsInvPattern)	qsInvPatte
Investigate(qsInvSample)	qsInvSamp
Investigate(qsInvCount)	qsInvCoun
Investigate(qsInvPercent)	qsInvPerce

InvReport	
Derivation	Column Nam
Investigate(qsInvColumnName)	qsInvColumnN
Investigate(qsInvPattern)	qsInvPattern
Investigate(qsInvSample)	qsInvSample
Investigate(qsInvCount)	qsInvCount
Investigate(qsInvPercent)	qsInvPercent

14. Click the OK button to return to the main stage screen.
15. Click the OK button to return to the job canvas.
16. Double-click on the target Sequential File stage named InvReport. Complete the File property as

C:\WINNCRM\InvCmask.rpt



Input name: InvReport

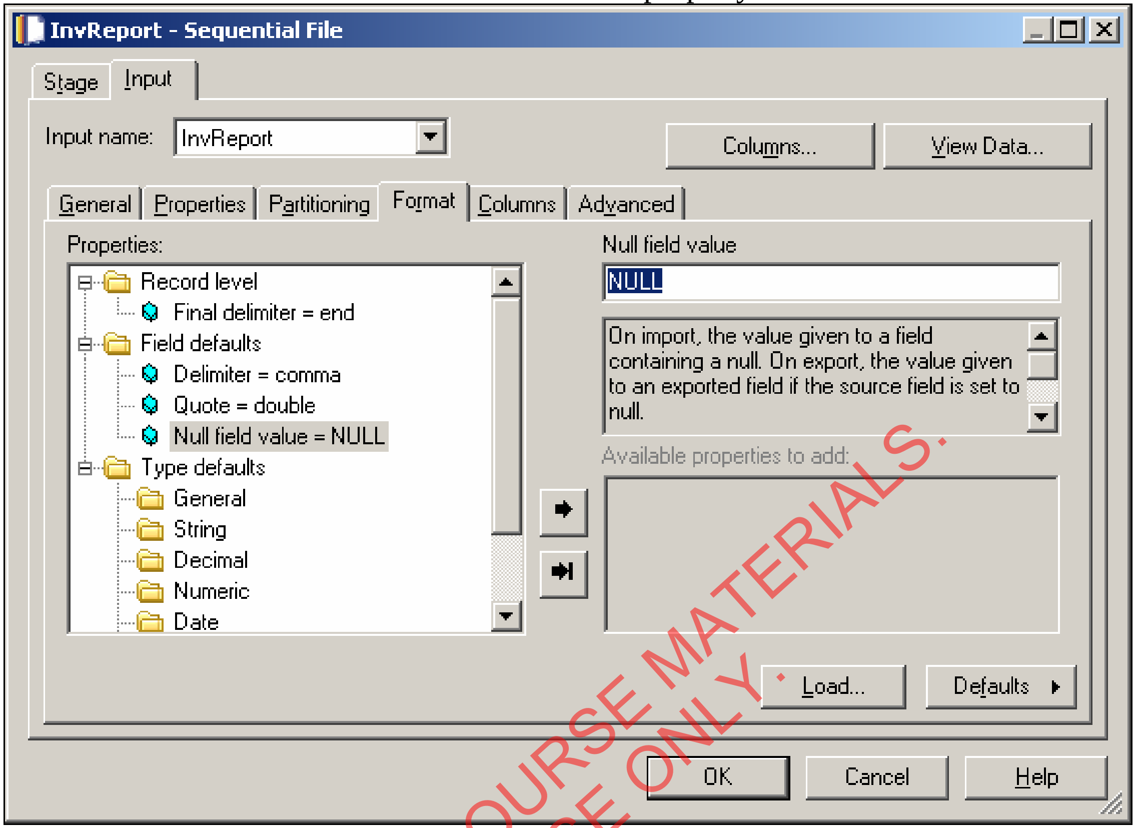
- Target
  - File = C:\WINNCRM\InvCmask.rpt
  - File Update Mode = Overwrite
- Options
  - Cleanup On Failure = True
  - First Line is Column Names = False
  - Reject Mode = Continue

File: C:\WINNCRM\InvCmask.rpt

Information:
Type: Pathname
Name of a file that the incoming data will be written to.

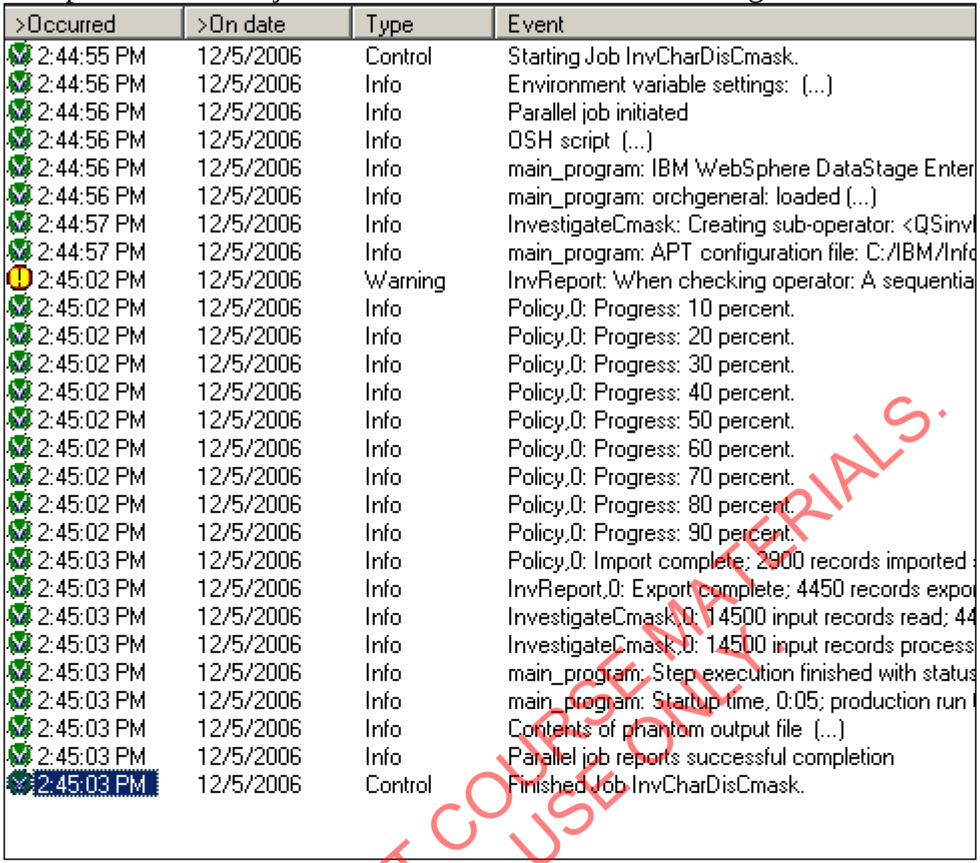
Available properties to add:

17. Click the Format tab and set the Null Field Value property to NULL.



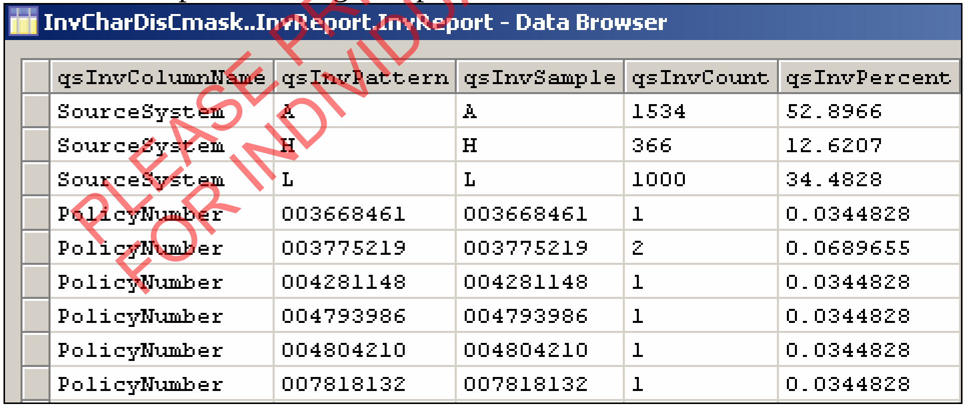
18. Click the OK button to return to the canvas.
19. Save your job as InvCharDisCmask and place it in the WINNCRM➔Jobs folder of the repository.

20. Compile and run the job from the Director Client. Your log should resemble:



```
>Occurred     >On date    Type      Event
2:44:55 PM    12/5/2006   Control   Starting Job InvCharDisCmask.
2:44:56 PM    12/5/2006   Info      Environment variable settings:  (...)
2:44:56 PM    12/5/2006   Info      Parallel job initiated
2:44:56 PM    12/5/2006   Info      OSH script  (...)
2:44:56 PM    12/5/2006   Info      main_program: IBM WebSphere DataStage Enter
2:44:56 PM    12/5/2006   Info      main_program: orchgeneral: loaded (...)
2:44:57 PM    12/5/2006   Info      InvestigateCmask: Creating sub-operator: <QSinvl
2:44:57 PM    12/5/2006   Info      main_program: APT configuration file: C:/IBM/Info
2:45:02 PM    12/5/2006   Warning   InvReport: When checking operator: A sequentia
2:45:02 PM    12/5/2006   Info      Policy,0: Progress: 10 percent.
2:45:02 PM    12/5/2006   Info      Policy,0: Progress: 20 percent.
2:45:02 PM    12/5/2006   Info      Policy,0: Progress: 30 percent.
2:45:02 PM    12/5/2006   Info      Policy,0: Progress: 40 percent.
2:45:02 PM    12/5/2006   Info      Policy,0: Progress: 50 percent.
2:45:02 PM    12/5/2006   Info      Policy,0: Progress: 60 percent.
2:45:02 PM    12/5/2006   Info      Policy,0: Progress: 70 percent.
2:45:02 PM    12/5/2006   Info      Policy,0: Progress: 80 percent.
2:45:02 PM    12/5/2006   Info      Policy,0: Progress: 90 percent.
2:45:03 PM    12/5/2006   Info      Policy,0: Import complete; 2900 records imported :
2:45:03 PM    12/5/2006   Info      InvReport,0: Export complete; 4450 records expo
2:45:03 PM    12/5/2006   Info      InvestigateCmask,0: 14500 input records read; 44
2:45:03 PM    12/5/2006   Info      InvestigateCmask,0: 14500 input records process
2:45:03 PM    12/5/2006   Info      main_program: Step execution finished with status
2:45:03 PM    12/5/2006   Info      main_program: Startup time, 0:05; production run
2:45:03 PM    12/5/2006   Info      Contents of phantom output file  (...)
2:45:03 PM    12/5/2006   Info      Parallel job reports successful completion
2:45:03 PM    12/5/2006   Control   Finished Job InvCharDisCmask.
```

21. Use the View Data button from the target Sequential File stage in Designer to review the results file- a partial listing is depicted below:



InvCharDisCmask..InvReport.InvReport - Data Browser

qsInvColumnName	qsInvPattern	qsInvSample	qsInvCount	qsInvPercent
SourceSystem	A	A	1534	52.8966
SourceSystem	H	H	366	12.6207
SourceSystem	L	L	1000	34.4828
PolicyNumber	003668461	003668461	1	0.0344828
PolicyNumber	003775219	003775219	2	0.0689655
PolicyNumber	004281148	004281148	1	0.0344828
PolicyNumber	004793986	004793986	1	0.0344828
PolicyNumber	004804210	004804210	1	0.0344828
PolicyNumber	007818132	007818132	1	0.0344828

Lab 7: Investigate – Character Discrete with T Mask

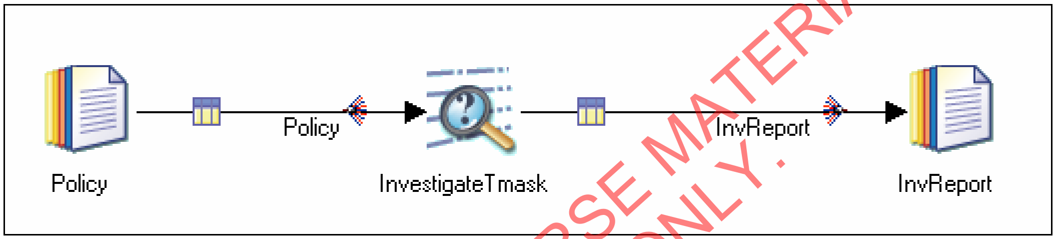
Assumptions:

- You have a DataStage project named WINNCRM.
- You have the Policy table definition in the repository

Task: Build and run an investigate Data Quality job

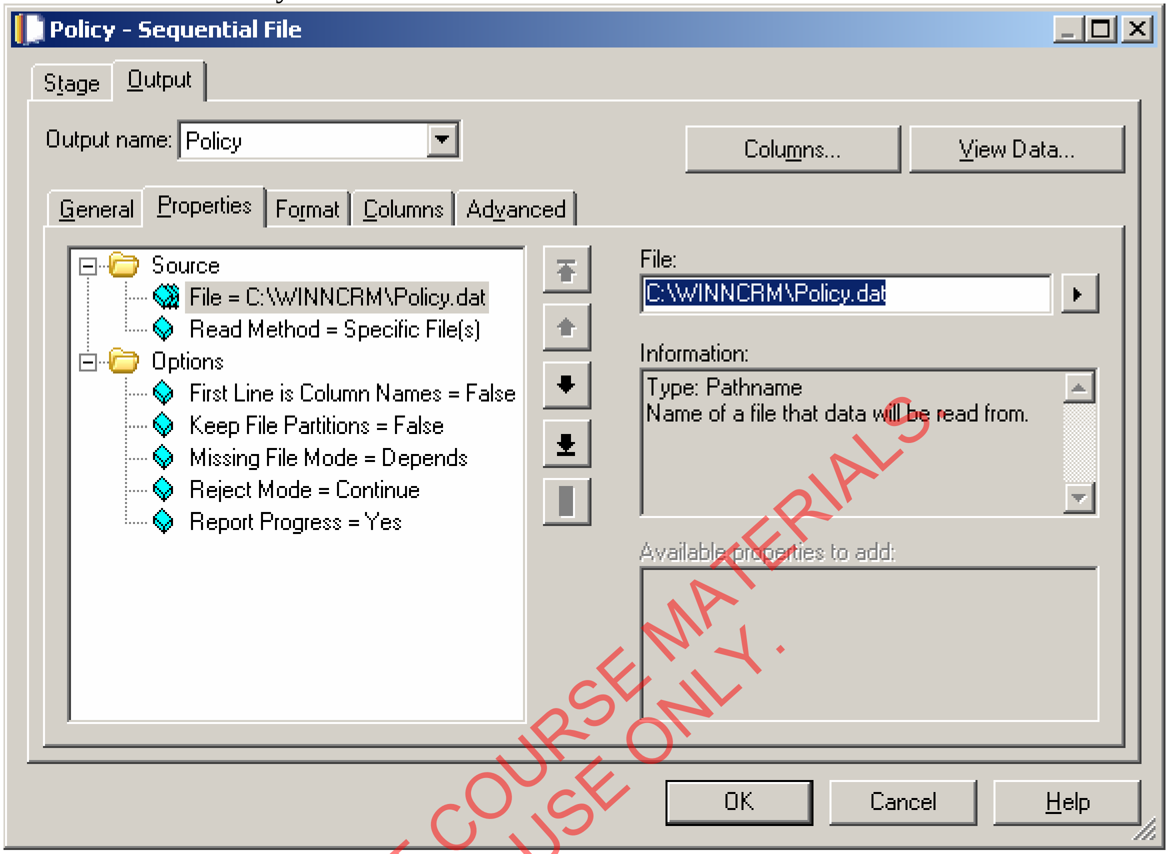
This job will perform a character discrete investigation using a T mask on the SourceSystem, PolicyNumber, FEDID, DOB, and DOD fields of the Policy.dat file. You may build the job from scratch, as described below, or edit your existing InvCharDisCmask job and then save it as InvCharDisTmask into the WINNCRM➔Jobs folder of the repository.

1. Create a new parallel job named InvCharDisTmask. Name the links and stages as shown below. Note: The Investigate stage can be found in the Data Quality subfolder in the Palette.



2. Open the Policy Sequential File stage and enter the file to be read (or use the browse function to locate it)

C:\WINNCRM\Policy.dat



Output name: Policy

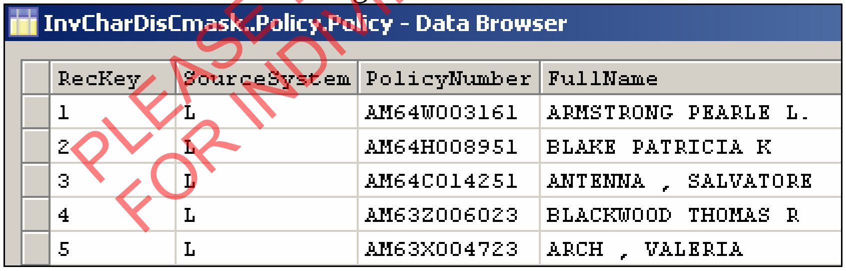
- Source
  - File = C:\WINNCRM\Policy.dat
  - Read Method = Specific File(s)
- Options
  - First Line is Column Names = False
  - Keep File Partitions = False
  - Missing File Mode = Depends
  - Reject Mode = Continue
  - Report Progress = Yes

File: C:\WINNCRM\Policy.dat

Information:
Type: Pathname
Name of a file that data will be read from.

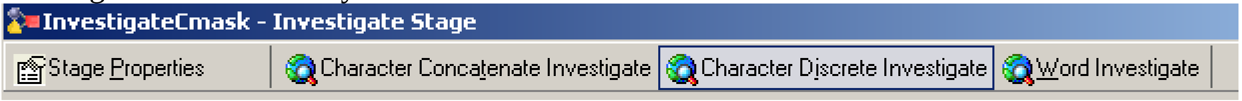
Available properties to add:

3. Click on the Columns tab and click Load.
4. Load all columns from the Policy table located in the WINNCRM ➔ TableDefs folder. Accept all columns.
5. Validate the meta data using the View Data button.

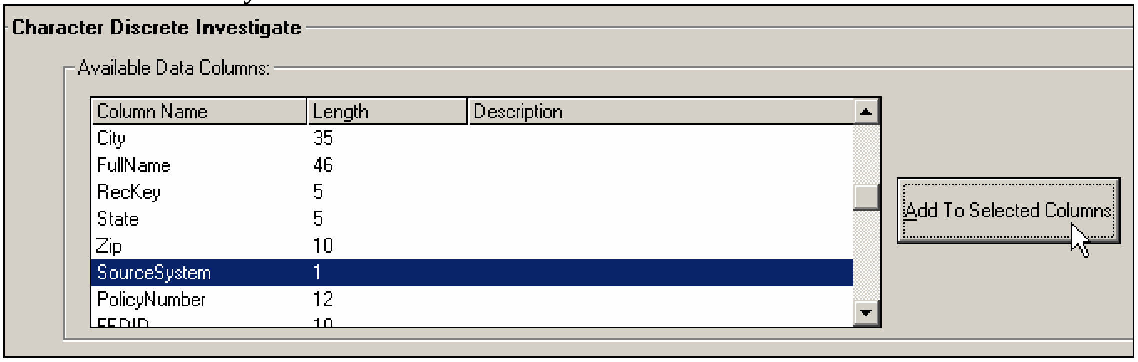


RecKey	SourceSystem	PolicyNumber	FullName
1	L	AM64W003161	ARMSTRONG PEARLE L.
2	L	AM64H008951	BLAKE PATRICIA K
3	L	AM64C014251	ANTENNA , SALVATORE
4	L	AM63Z006023	BLACKWOOD THOMAS R
5	L	AM63X004723	ARCH , VALERIA

6. Close the Sequential File stage.
7. Open the InvestigateTmask Investigate stage. The default tab - Character Discrete Investigate - should already be selected.


InvestigateCmask - Investigate Stage
Stage Properties
Character Concatenate Investigate
Character Discrete Investigate
Word Investigate

8. Select the SourceSystem column and click the Add to Selected Columns button.

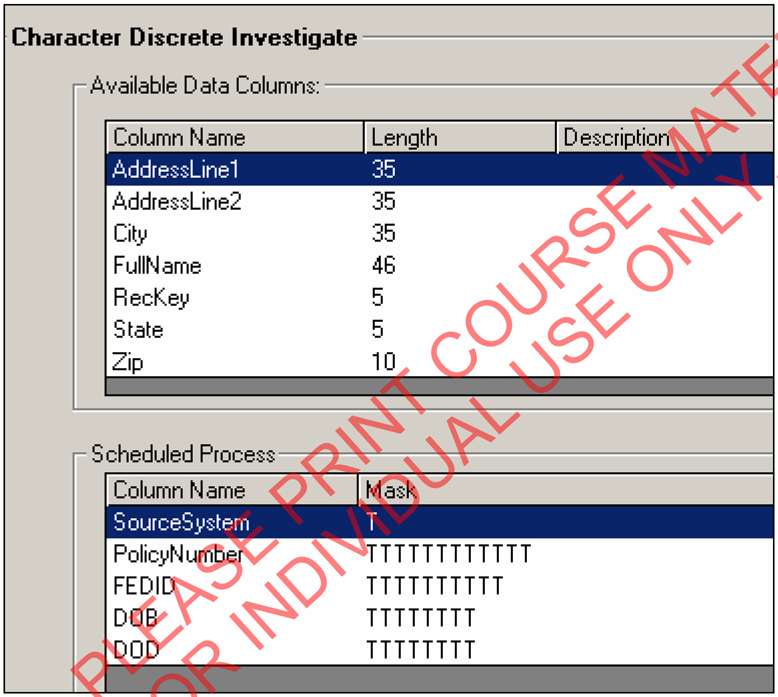


Character Discrete Investigate

Available Data Columns:

Column Name	Length	Description
City	35	
FullName	46	
RecKey	5	
State	5	
Zip	10	
SourceSystem	1	
PolicyNumber	12	
FEDID	10	

9. The Mask Column Selection screen that appears will default to type T mask. Click the OK button to accept this option.
10. In a similar fashion select columns PolicyNumber, FEDID, DOB, and DOD. Apply the All T mask to each column. Your screen should now resemble:



Character Discrete Investigate

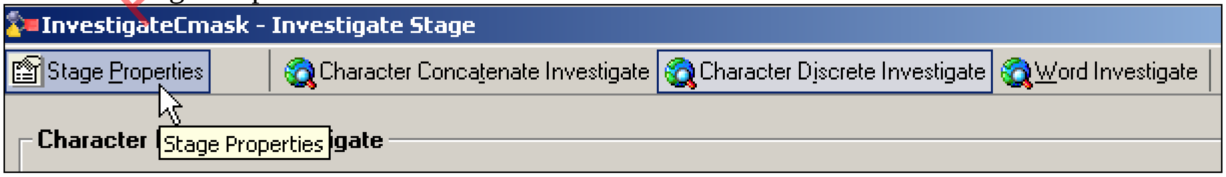
Available Data Columns:

Column Name	Length	Description
AddressLine1	35	
AddressLine2	35	
City	35	
FullName	46	
RecKey	5	
State	5	
Zip	10	

Scheduled Process

Column Name	Mask
SourceSystem	T
PolicyNumber	TTTTTTTTTTTT
FEDID	TTTTTTTTTT
DOB	TTTTTTTT
DOD	TTTTTTTT

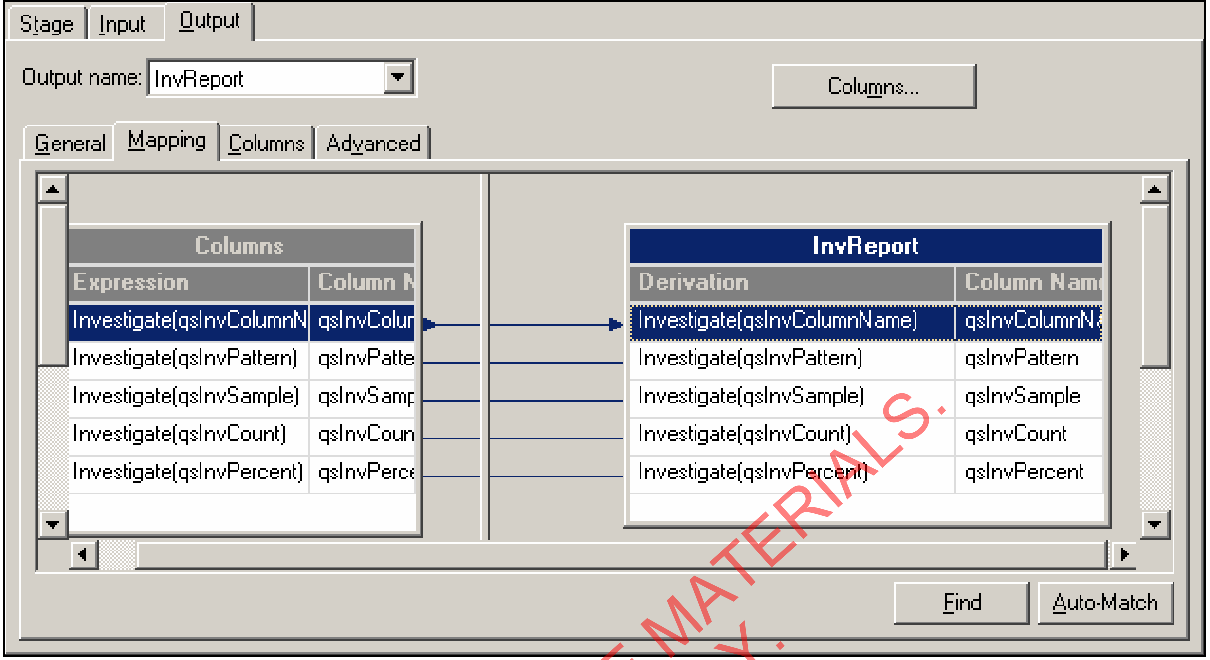
11. Click the Stage Properties tab.



InvestigateCmask - Investigate Stage
Stage Properties
Character Concatenate Investigate
Character Discrete Investigate
Word Investigate
Character Discrete Investigate
Stage Properties

12. Click the Output ➔ Mapping tabs.

13. Drag all input columns to the output link.



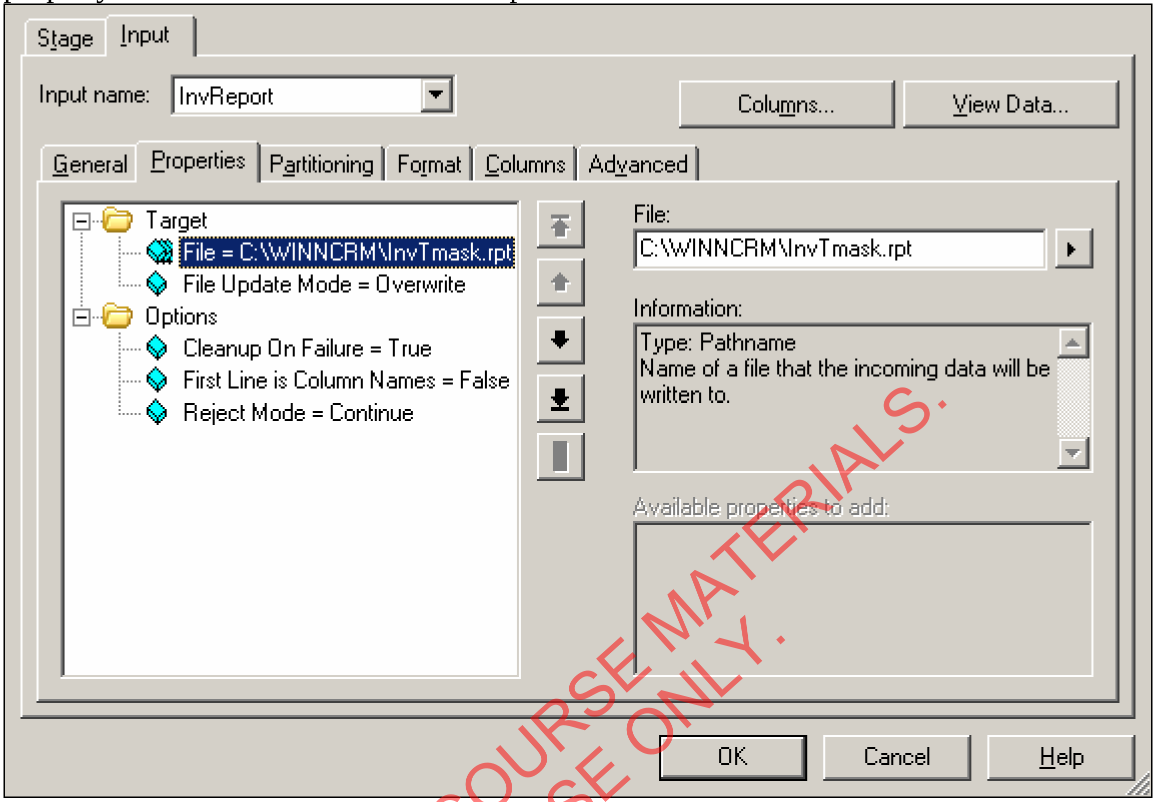
Output name: InvReport

Columns	
Expression	Column N
Investigate(qsInvColumnN	qsInvColur
Investigate(qsInvPattern)	qsInvPatte
Investigate(qsInvSample)	qsInvSamp
Investigate(qsInvCount)	qsInvCoun
Investigate(qsInvPercent)	qsInvPerce

InvReport	
Derivation	Column Name
Investigate(qsInvColumnName)	qsInvColumnN
Investigate(qsInvPattern)	qsInvPattern
Investigate(qsInvSample)	qsInvSample
Investigate(qsInvCount)	qsInvCount
Investigate(qsInvPercent)	qsInvPercent

14. Click the OK button to return to the main stage screen.
15. Click the OK button to return to the job canvas.

16. Double-click on the target Sequential File stage named InvReport. Complete the File property as C:\WINNCRM\InvTmask.rpt



Input name: InvReport

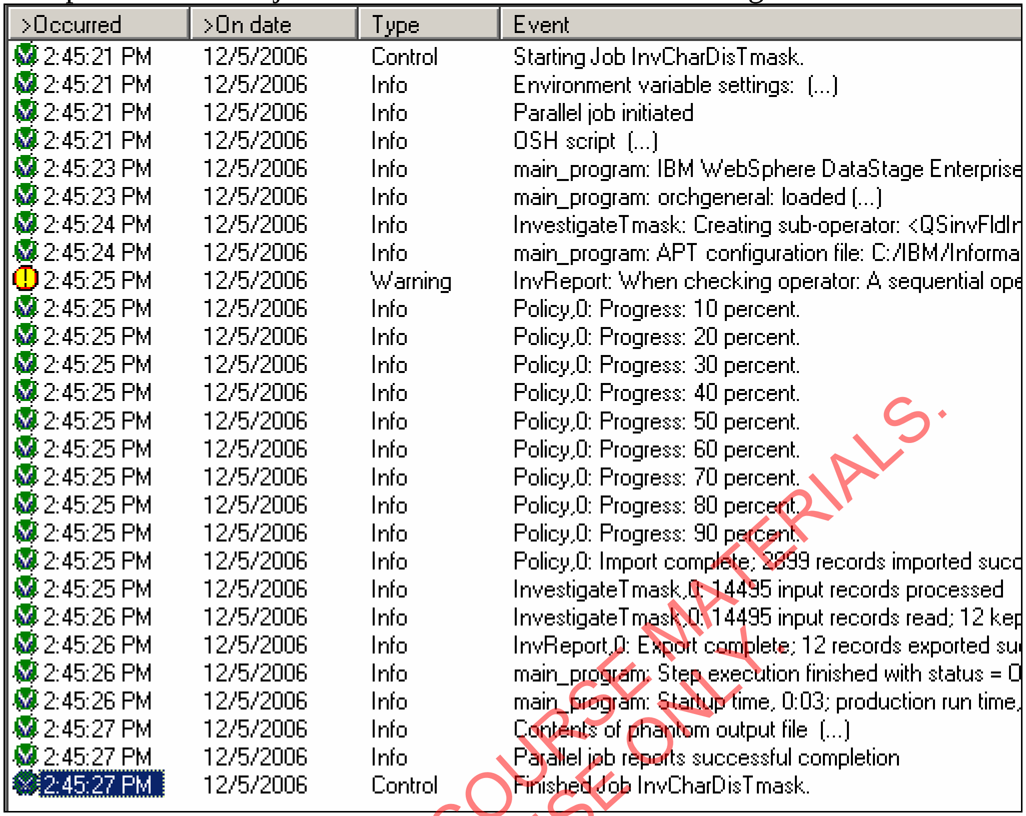
Target
File = C:\WINNCRM\InvTmask.rpt
File Update Mode = Overwrite
Options
Cleanup On Failure = True
First Line is Column Names = False
Reject Mode = Continue

File:
C:\WINNCRM\InvTmask.rpt

Information:
Type: Pathname
Name of a file that the incoming data will be written to.

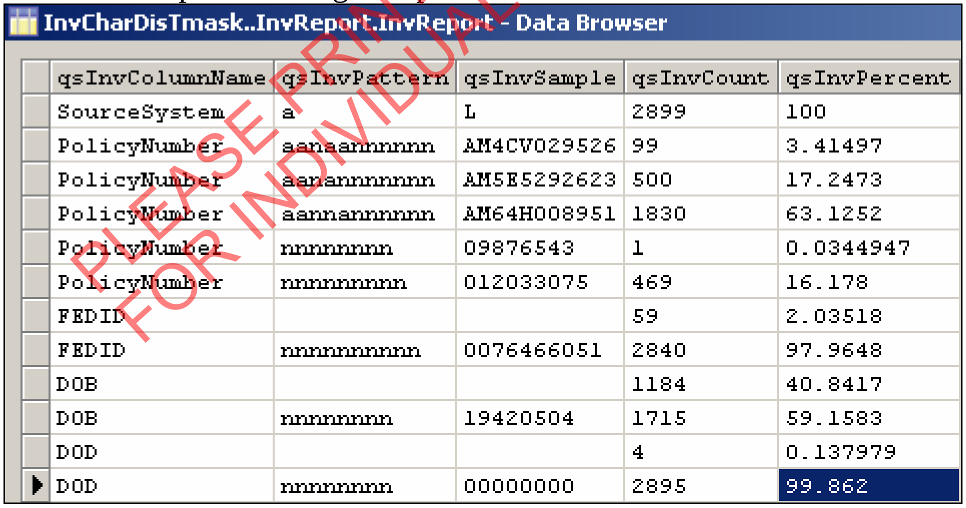
17. Click the Format tab and set the Null Field Value property to NULL.
18. Click the OK button to return to the canvas.
19. Save your job as InvCharDisTmask and place it in the WINNCRM➔Jobs folder of the repository.

20. Compile and run the job from the Director Client. Your log should resemble:



>Occurred	>On date	Type	Event
2:45:21 PM	12/5/2006	Control	Starting Job InvCharDisTmask.
2:45:21 PM	12/5/2006	Info	Environment variable settings: (...)
2:45:21 PM	12/5/2006	Info	Parallel job initiated
2:45:21 PM	12/5/2006	Info	OSH script (...)
2:45:23 PM	12/5/2006	Info	main_program: IBM WebSphere DataStage Enterprise
2:45:23 PM	12/5/2006	Info	main_program: orchgeneral: loaded (...)
2:45:24 PM	12/5/2006	Info	InvestigateTmask: Creating sub-operator: <QSinvFldIn
2:45:24 PM	12/5/2006	Info	main_program: APT configuration file: C:/IBM/Informa
2:45:25 PM	12/5/2006	Warning	InvReport: When checking operator: A sequential ope
2:45:25 PM	12/5/2006	Info	Policy,0: Progress: 10 percent.
2:45:25 PM	12/5/2006	Info	Policy,0: Progress: 20 percent.
2:45:25 PM	12/5/2006	Info	Policy,0: Progress: 30 percent.
2:45:25 PM	12/5/2006	Info	Policy,0: Progress: 40 percent.
2:45:25 PM	12/5/2006	Info	Policy,0: Progress: 50 percent.
2:45:25 PM	12/5/2006	Info	Policy,0: Progress: 60 percent.
2:45:25 PM	12/5/2006	Info	Policy,0: Progress: 70 percent.
2:45:25 PM	12/5/2006	Info	Policy,0: Progress: 80 percent.
2:45:25 PM	12/5/2006	Info	Policy,0: Progress: 90 percent.
2:45:25 PM	12/5/2006	Info	Policy,0: Import complete; 2899 records imported succ
2:45:25 PM	12/5/2006	Info	InvestigateTmask,0: 14495 input records processed
2:45:26 PM	12/5/2006	Info	InvestigateTmask,0: 14495 input records read; 12 kep
2:45:26 PM	12/5/2006	Info	InvReport,0: Export complete; 12 records exported su
2:45:26 PM	12/5/2006	Info	main_program: Step execution finished with status = 0
2:45:26 PM	12/5/2006	Info	main_program: Startup time, 0:03; production run time,
2:45:27 PM	12/5/2006	Info	Contents of phantom output file (...)
2:45:27 PM	12/5/2006	Info	Parallel job reports successful completion
2:45:27 PM	12/5/2006	Control	Finished Job InvCharDisTmask.

21. Use the View Data button from the target Sequential File stage in Designer to review the results file- a partial listing is depicted below:



InvCharDisTmask..InvReport.InvReport - Data Browser

qsInvColumnName	qsInvPattern	qsInvSample	qsInvCount	qsInvPercent
SourceSystem	a	L	2899	100
PolicyNumber	aanaannnnnn	AM4CV029526	99	3.41497
PolicyNumber	aanannnnnnn	AM5E5292623	500	17.2473
PolicyNumber	aannannnnnn	AM64H008951	1830	63.1252
PolicyNumber	nnnnnnnn	09876543	1	0.0344947
PolicyNumber	nnnnnnnnn	012033075	469	16.178
FEDID			59	2.03518
FEDID	nnnnnnnnnn	0076466051	2840	97.9648
DOB			1184	40.8417
DOB	nnnnnnnn	19420504	1715	59.1583
DOD			4	0.137979
DOD	nnnnnnnn	00000000	2895	99.862

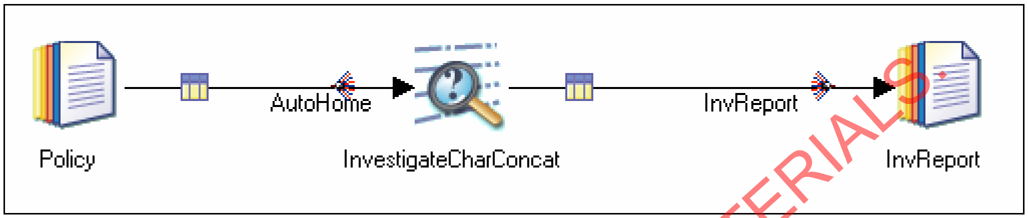
Lab 8: Investigate – Character Concatenate

Assumptions:

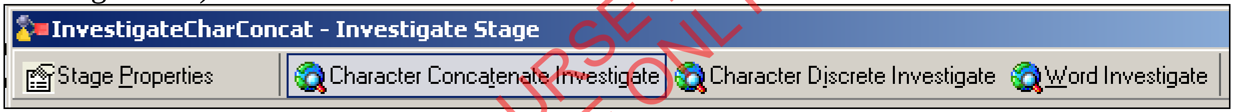
- DataStage project named WINNCRM exists
- The Policy table definition exists in the repository

Task: Build and run a character concatenate Data Quality job

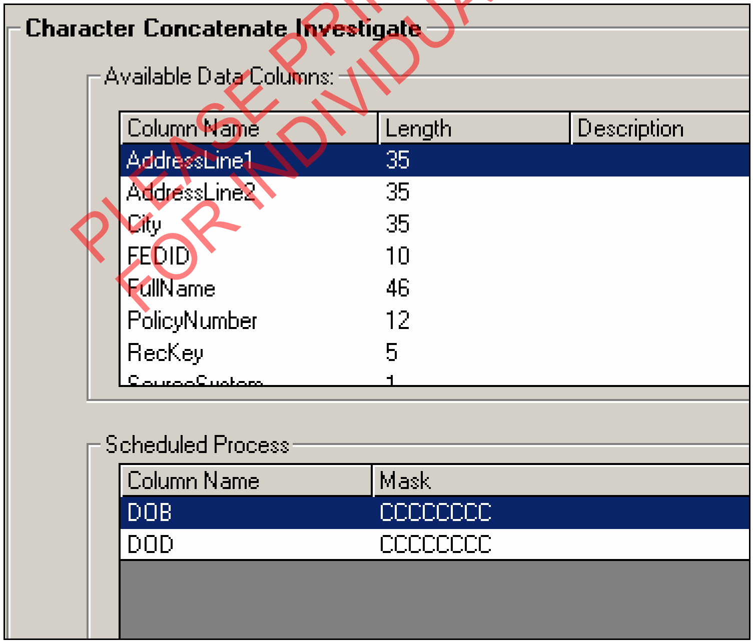
1. Edit the InvCharDisCmask job.
2. Save the job as InvCharConcat. Rename the Investigate stage to create the job depicted below:



3. Open the InvestigateCharConcat stage and select the Character Concatenate Investigate tab. (This will clear the properties you previously assigned in the Character Discrete Investigate tab.)


InvestigateCharConcat - Investigate Stage
Stage Properties
Character Concatenate Investigate
Character Discrete Investigate
Word Investigate

4. Select both the DOB and DOD columns and click the Add to Selected Columns button.
5. The Mask Column Selection screen will appear successively for each field. Click the All C option for each field.
6. Your screen should now look like:



Character Concatenate Investigate

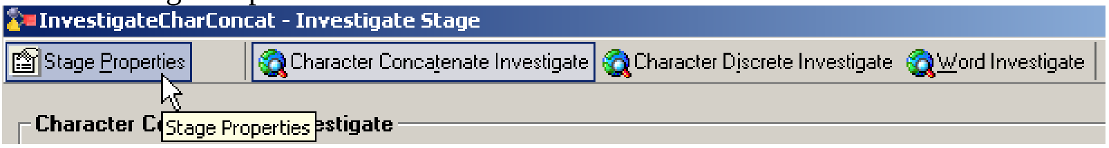
Available Data Columns:

Column Name	Length	Description
AddressLine1	35	
AddressLine2	35	
City	35	
FEDID	10	
FullName	46	
PolicyNumber	12	
RecKey	5	
SourceSystem	1	

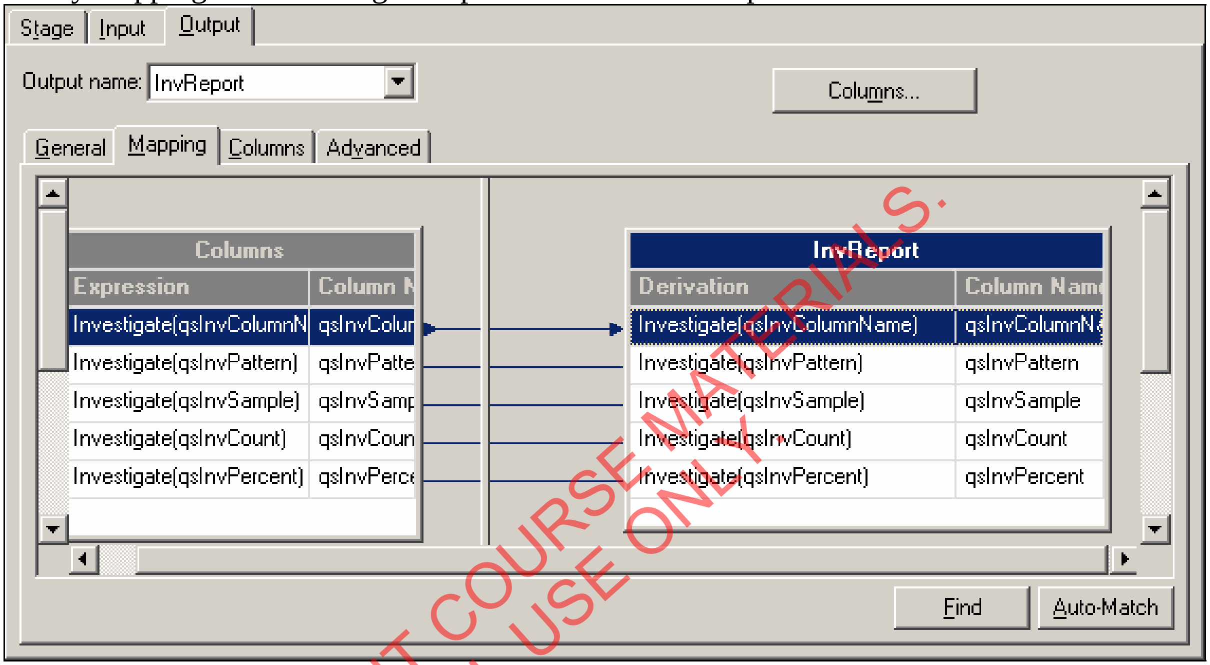
Scheduled Process

Column Name	Mask
DOB	CCCCCCCC
DOD	CCCCCCCC

7. Click the Stage Properties tab.


InvestigateCharConcat - Investigate Stage
Stage Properties
Character Concatenate Investigate
Character Discrete Investigate
Word Investigate
Character Co
Stage Properties
estigate

8. Click the Output ➔ Mapping tabs.
9. Verify mappings exist or drag all input columns to the output link.



Output name: InvReport

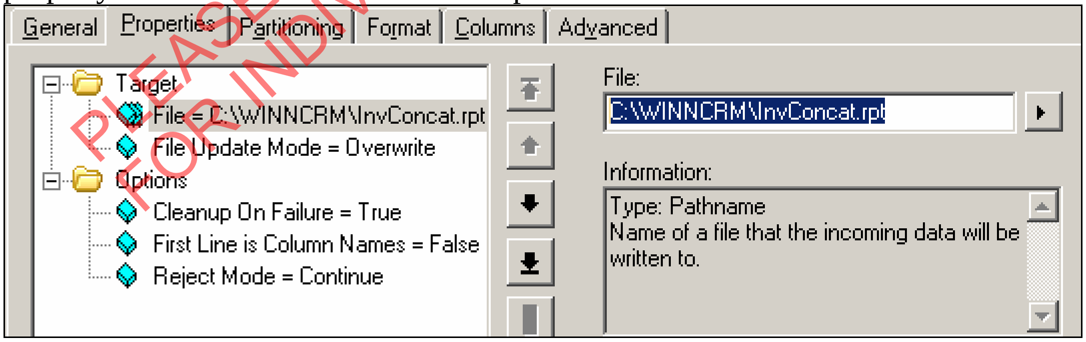
Columns

Expression	Column N
Investigate(qslnvColumnN	qslnvColun
Investigate(qslnvPattern)	qslnvPatte
Investigate(qslnvSample)	qslnvSamp
Investigate(qslnvCount)	qslnvCoun
Investigate(qslnvPercent)	qslnvPerce

InvReport

Derivation	Column Name
Investigate(qslnvColumnName)	qslnvColumnN
Investigate(qslnvPattern)	qslnvPattern
Investigate(qslnvSample)	qslnvSample
Investigate(qslnvCount)	qslnvCount
Investigate(qslnvPercent)	qslnvPercent

10. Click the OK button to return to the main stage screen.
11. Click the OK button to return to the job canvas.
12. Double-click on the target Sequential File stage named InvReport. Complete the File property as C:\WINNCRM\InvConcat.rpt



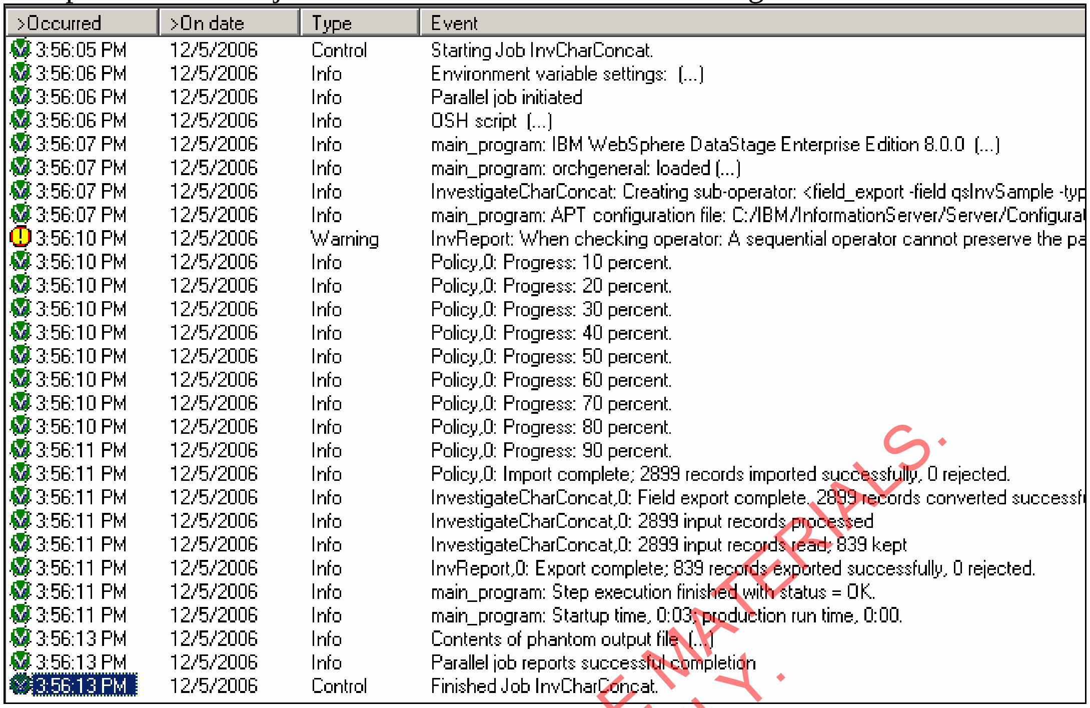
- Target
  - File = C:\WINNCRM\InvConcat.rpt
  - File Update Mode = Overwrite
- Options
  - Cleanup On Failure = True
  - First Line is Column Names = False
  - Reject Mode = Continue

File: C:\WINNCRM\InvConcat.rpt

Information:
Type: Pathname
Name of a file that the incoming data will be written to.

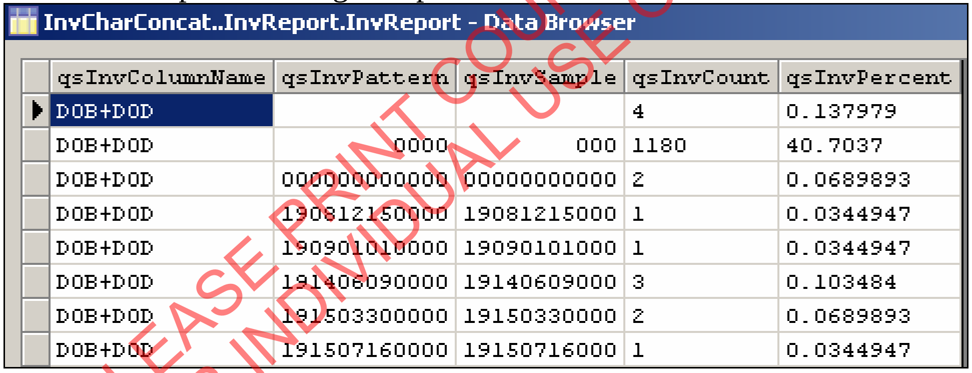
13. Click the Format tab and set the Null Field Value property to NULL.
14. Click the OK button to return to the canvas.

15. Compile and run the job from the Director Client. Your log should resemble:



>Occurred	>On date	Type	Event
3:56:05 PM	12/5/2006	Control	Starting Job InvCharConcat.
3:56:06 PM	12/5/2006	Info	Environment variable settings: (...)
3:56:06 PM	12/5/2006	Info	Parallel job initiated
3:56:06 PM	12/5/2006	Info	OSH script (...)
3:56:07 PM	12/5/2006	Info	main_program: IBM WebSphere DataStage Enterprise Edition 8.0.0 (...)
3:56:07 PM	12/5/2006	Info	main_program: orchgeneral: loaded (...)
3:56:07 PM	12/5/2006	Info	InvestigateCharConcat: Creating sub-operator: <field_export -field qsInvSample -typ
3:56:07 PM	12/5/2006	Info	main_program: APT configuration file: C:/IBM/InformationServer/Server/Configurat
3:56:10 PM	12/5/2006	Warning	InvReport: When checking operator: A sequential operator cannot preserve the pa
3:56:10 PM	12/5/2006	Info	Policy,0: Progress: 10 percent.
3:56:10 PM	12/5/2006	Info	Policy,0: Progress: 20 percent.
3:56:10 PM	12/5/2006	Info	Policy,0: Progress: 30 percent.
3:56:10 PM	12/5/2006	Info	Policy,0: Progress: 40 percent.
3:56:10 PM	12/5/2006	Info	Policy,0: Progress: 50 percent.
3:56:10 PM	12/5/2006	Info	Policy,0: Progress: 60 percent.
3:56:10 PM	12/5/2006	Info	Policy,0: Progress: 70 percent.
3:56:10 PM	12/5/2006	Info	Policy,0: Progress: 80 percent.
3:56:11 PM	12/5/2006	Info	Policy,0: Progress: 90 percent.
3:56:11 PM	12/5/2006	Info	Policy,0: Import complete; 2899 records imported successfully, 0 rejected.
3:56:11 PM	12/5/2006	Info	InvestigateCharConcat,0: Field export complete. 2899 records converted successfu
3:56:11 PM	12/5/2006	Info	InvestigateCharConcat,0: 2899 input records processed
3:56:11 PM	12/5/2006	Info	InvestigateCharConcat,0: 2899 input records read; 839 kept
3:56:11 PM	12/5/2006	Info	InvReport,0: Export complete; 839 records exported successfully, 0 rejected.
3:56:11 PM	12/5/2006	Info	main_program: Step execution finished with status = OK.
3:56:11 PM	12/5/2006	Info	main_program: Startup time, 0:03; production run time, 0:00.
3:56:13 PM	12/5/2006	Info	Contents of phantom output file (...)
3:56:13 PM	12/5/2006	Info	Parallel job reports successful completion
3:56:13 PM	12/5/2006	Control	Finished Job InvCharConcat.

16. Use the View Data button from the target Sequential File stage in Designer to review the results file- a partial listing is depicted below:



InvCharConcat..InvReport.InvReport - Data Browser

qsInvColumnName	qsInvPattern	qsInvSample	qsInvCount	qsInvPercent
DOB+DOD			4	0.137979
DOB+DOD	0000	000	1180	40.7037
DOB+DOD	000000000000	00000000000	2	0.0689893
DOB+DOD	190812150000	19081215000	1	0.0344947
DOB+DOD	190901010000	19090101000	1	0.0344947
DOB+DOD	191406090000	19140609000	3	0.103484
DOB+DOD	191503300000	19150330000	2	0.0689893
DOB+DOD	191507160000	19150716000	1	0.0344947

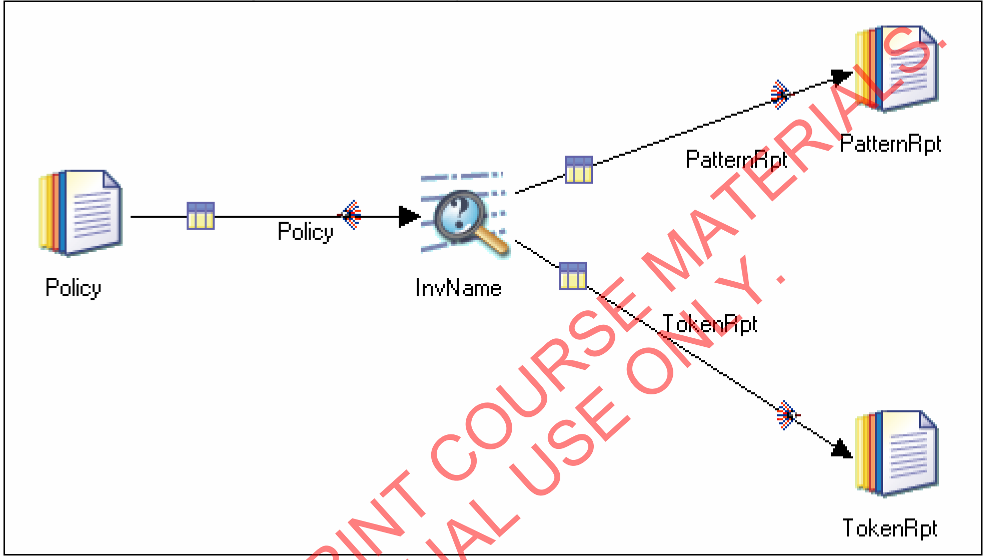
Lab 9: Investigate word

Assumptions:

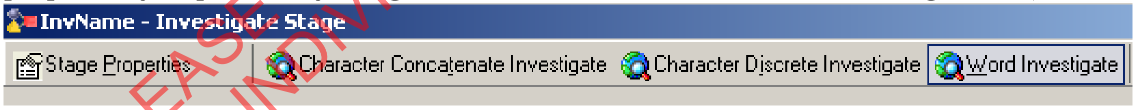
- DataStage project named WINNCRM exists
- Job InvCharDisCmask exists and executed successfully

Task: Build and run a word investigate for Fullname field

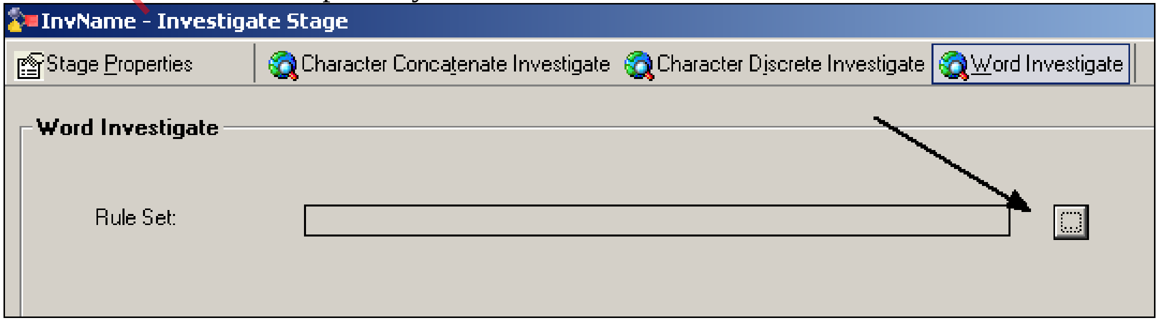
1. Edit your InvCharDisCmask job.
2. Save the job as InvWordName. Add a second target Sequential File stage as depicted below. Rename stages and links to yield:



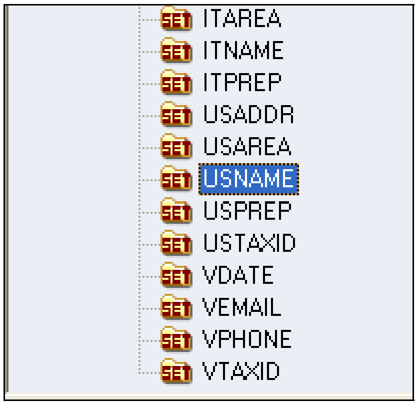
3. Edit the InvName stage and select the Word Investigate tab. (This will clear the properties you previously assigned in the Character Concatenate Investigate tab.)


InvName - Investigate Stage
Stage Properties
Character Concatenate Investigate
Character Discrete Investigate
Word Investigate

4. Click on the browse (….) button to the right of the Rule Set text box. This will allow you to navigate to where the rule sets are stored, typically in the Standardization Rules ➔ Rule Sets folder in the repository.

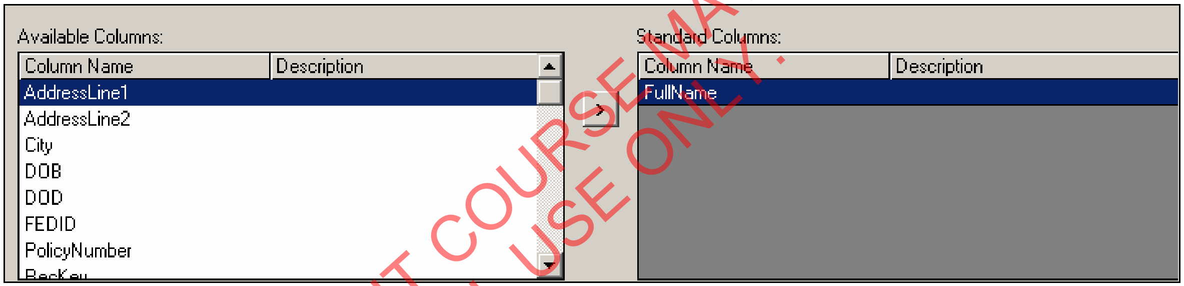

InvName - Investigate Stage
Stage Properties
Character Concatenate Investigate
Character Discrete Investigate
Word Investigate
Word Investigate
Rule Set:

5. Scroll down and select the USNAME rule set.



- ITAREA
- ITNAME
- ITPREP
- USADDR
- USAREA
- USNAME
- USPREP
- USTAXID
- VDATE
- VEMAIL
- VPHONE
- VTAXID

6. Right-click on the USNAME rule set and click the Provision All option. This action copies the rule set to an area needed for job execution.
7. You will be returned to the Word Investigate tab of the Investigation stage.
8. Under available columns, select the Fullname field and click the > button to add FullName to the Standard Columns box.



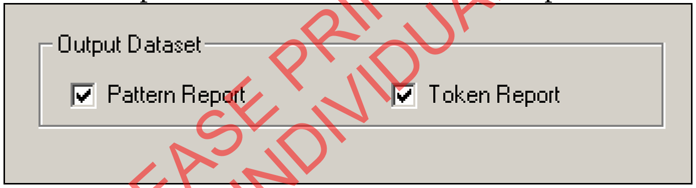
Available Columns:

Column Name	Description
AddressLine1	
AddressLine2	
City	
DOB	
DOD	
FEDID	
PolicyNumber	
RecKey	

Standard Columns:

Column Name	Description
FullName	

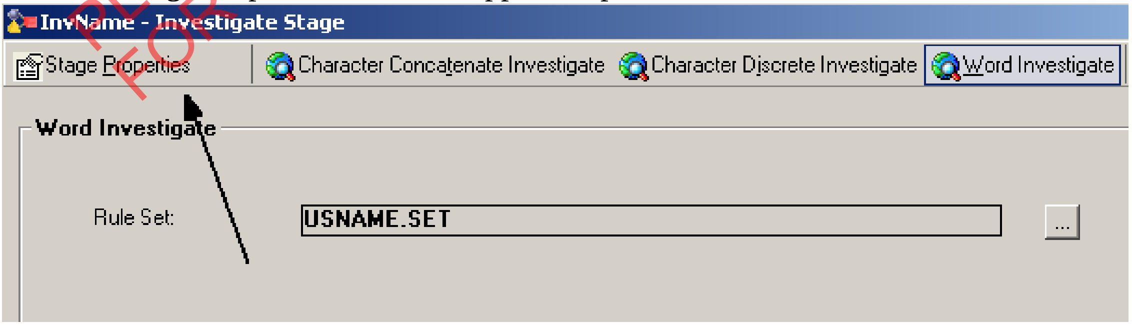
9. Under Output Dataset click the Pattern Report and Token Report checkboxes.



Output Dataset

☑ Pattern Report ☑ Token Report

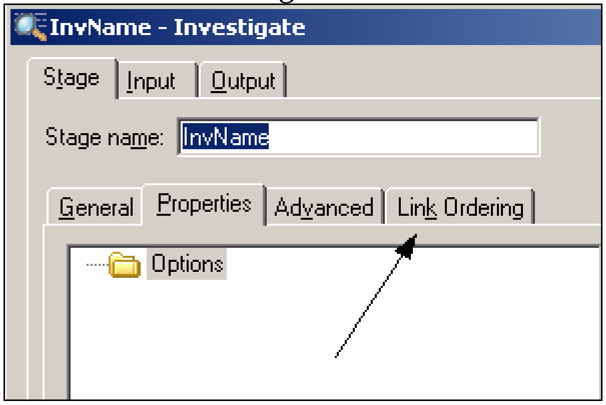
10. Click the Stage Properties tab in the upper left portion of the screen.



InvName - Investigate Stage
Stage Properties
Character Concatenate Investigate
Character Discrete Investigate
Word Investigate
Word Investigate

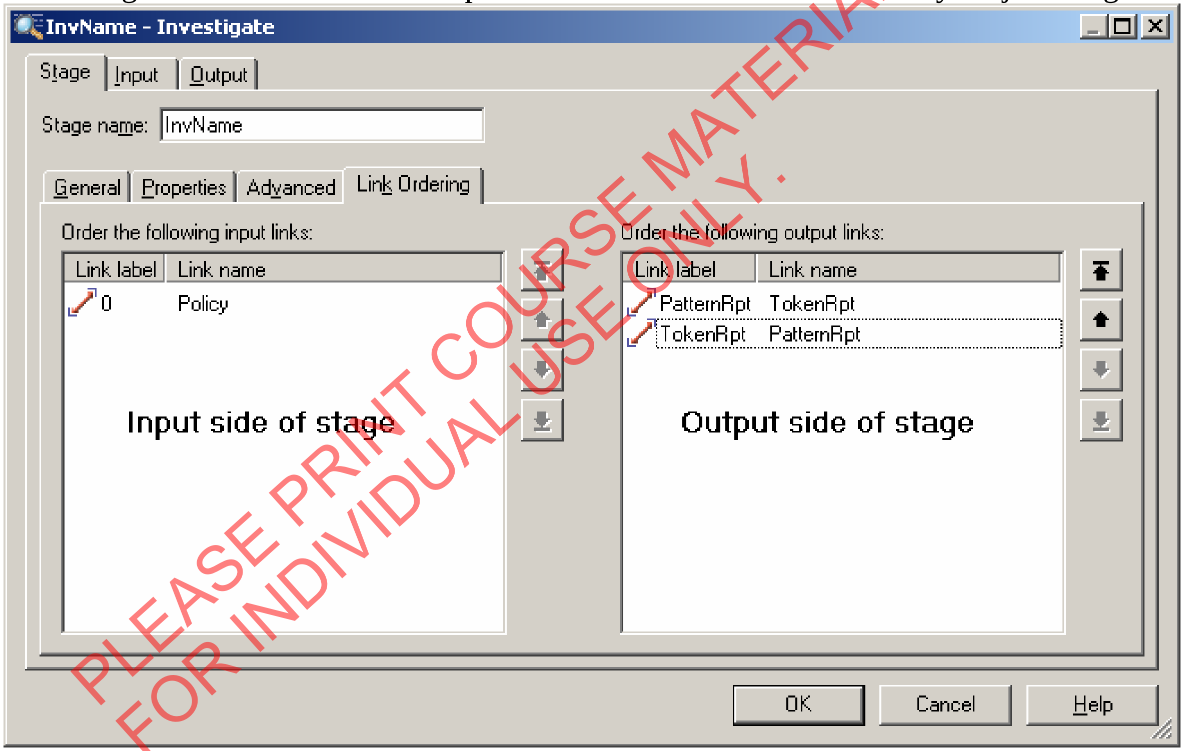
Rule Set: USNAME.SET

11. Click the link ordering tab.



Stage name: InvName

12. The Link Ordering screen has an input side and an output side. Notice that the input side of the stage has one link and the output side has two links. This matches your job design.



Stage name: InvName

Order the following input links:

Link label	Link name
0	Policy

Input side of stage

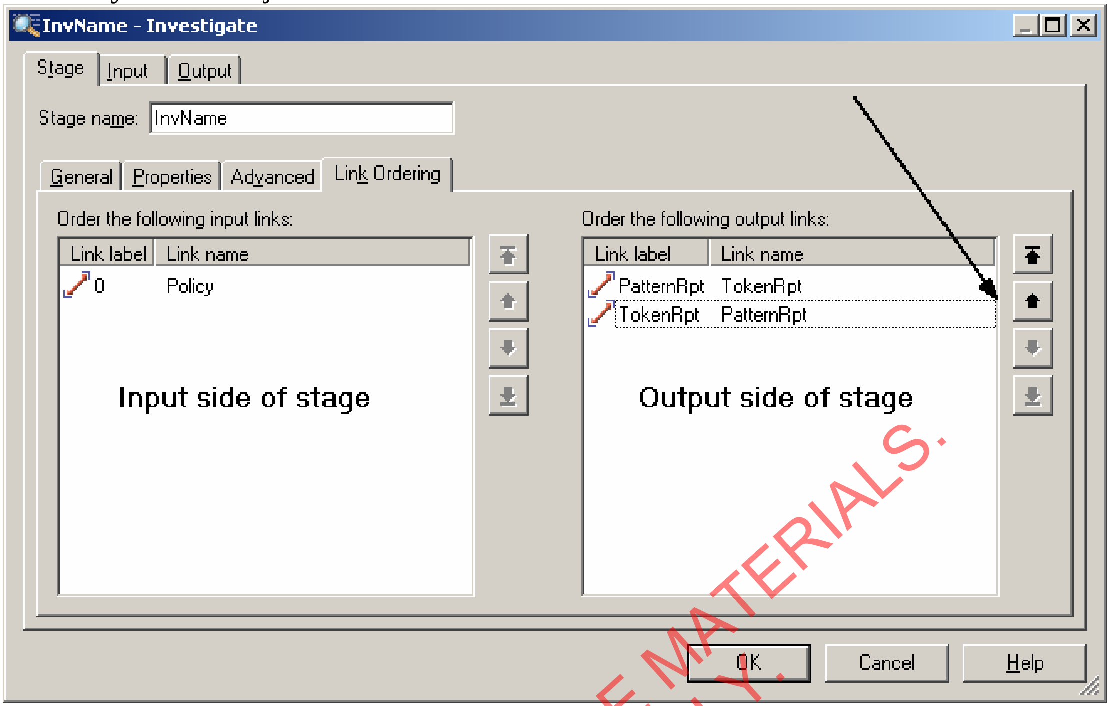
Order the following output links:

Link label	Link name
PatternRpt	TokenRpt
TokenRpt	PatternRpt

Output side of stage

13. You named the two output links PatternRpt and TokenRpt – these are represented under the Link name column. The stage uses the names under Link label for report flow assignments. Note that in this example the link assignments do NOT match the stage assignments. If you find this is true, use the arrow buttons on the right side of the pane to

make any needed adjustments.



Stage name: InvName

Order the following input links:

Link label	Link name
0	Policy

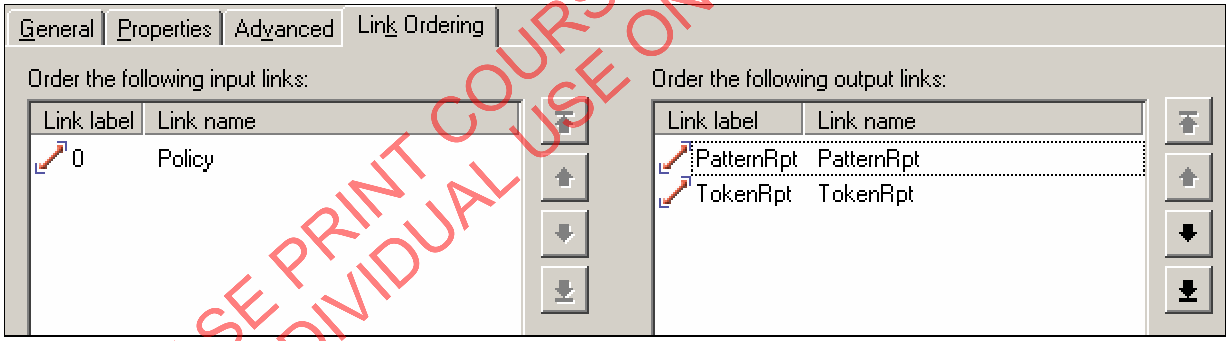
Order the following output links:

Link label	Link name
PatternRpt	TokenRpt
TokenRpt	PatternRpt

Input side of stage

Output side of stage

14. The final result should be:



Order the following input links:

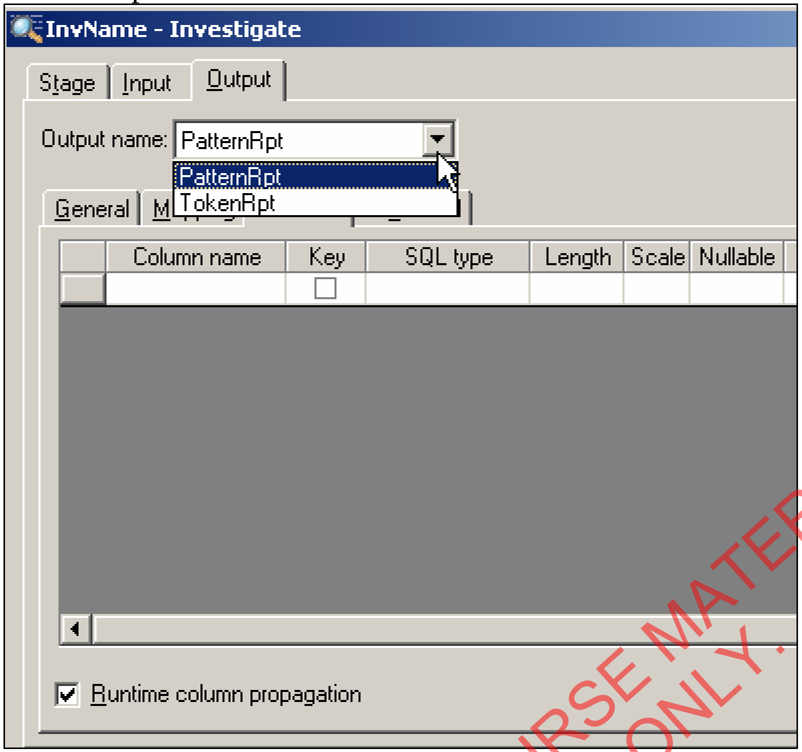
Link label	Link name
0	Policy

Order the following output links:

Link label	Link name
PatternRpt	PatternRpt
TokenRpt	TokenRpt

15. Click the Output ➔ Mapping tab. Since you have two output links, two output mappings are required. You may switch from one link to the other by clicking the drop down arrow

in the Output name box.



InvName - Investigate

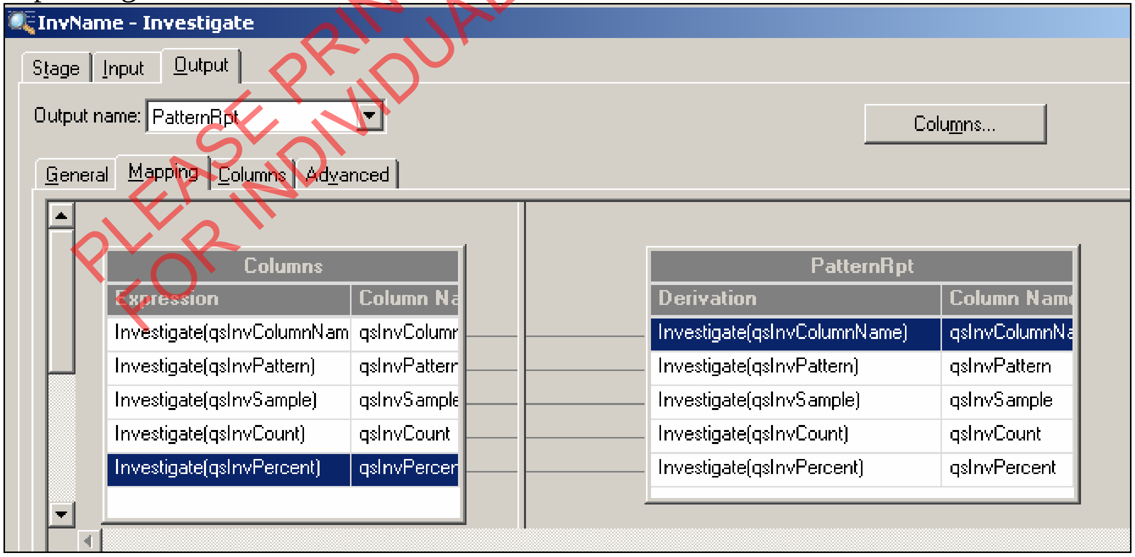
Output name: PatternRpt

PatternRpt
TokenRpt

	Column name	Key	SQL type	Length	Scale	Nullable
		□				

☑ Runtime column propagation

16. First choose one of the links and map all input columns to the output side. Then switch to the other link and map all input columns to the output side. You will now have column mappings for both links: PatternRpt and TokenRpt. Note that the columns differ depending on which link is chosen.

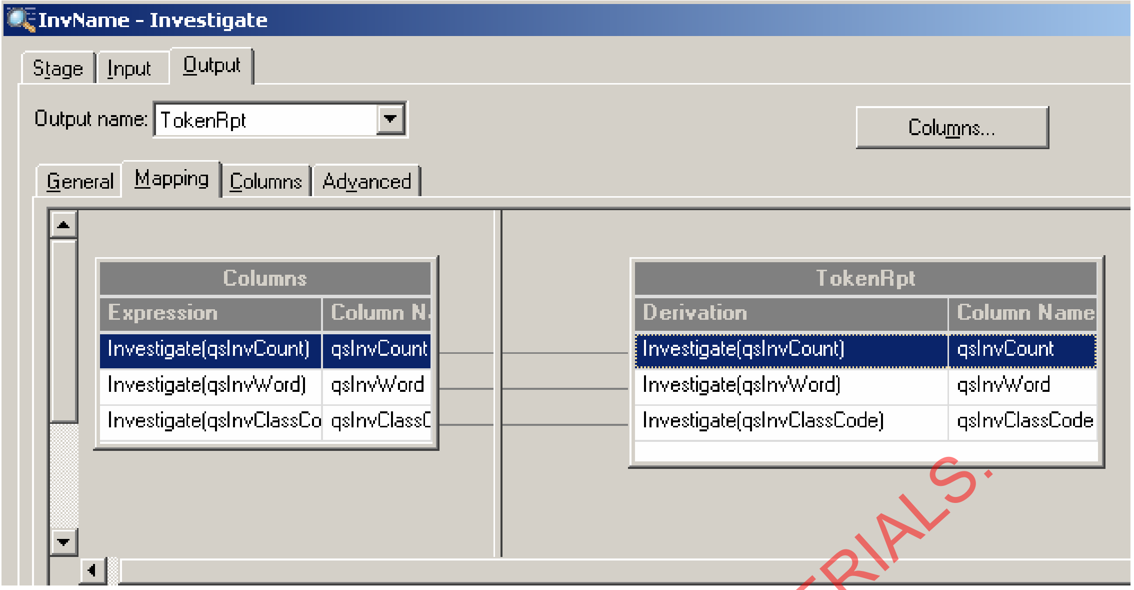


InvName - Investigate

Output name: PatternRpt

Columns	
Expression	Column Na
Investigate(qsInvColumnNam	qsInvColumr
Investigate(qsInvPattern)	qsInvPatterr
Investigate(qsInvSample)	qsInvSample
Investigate(qsInvCount)	qsInvCount
Investigate(qsInvPercent)	qsInvPercer

PatternRpt	
Derivation	Column Nam
Investigate(qsInvColumnName)	qsInvColumnNa
Investigate(qsInvPattern)	qsInvPattern
Investigate(qsInvSample)	qsInvSample
Investigate(qsInvCount)	qsInvCount
Investigate(qsInvPercent)	qsInvPercent

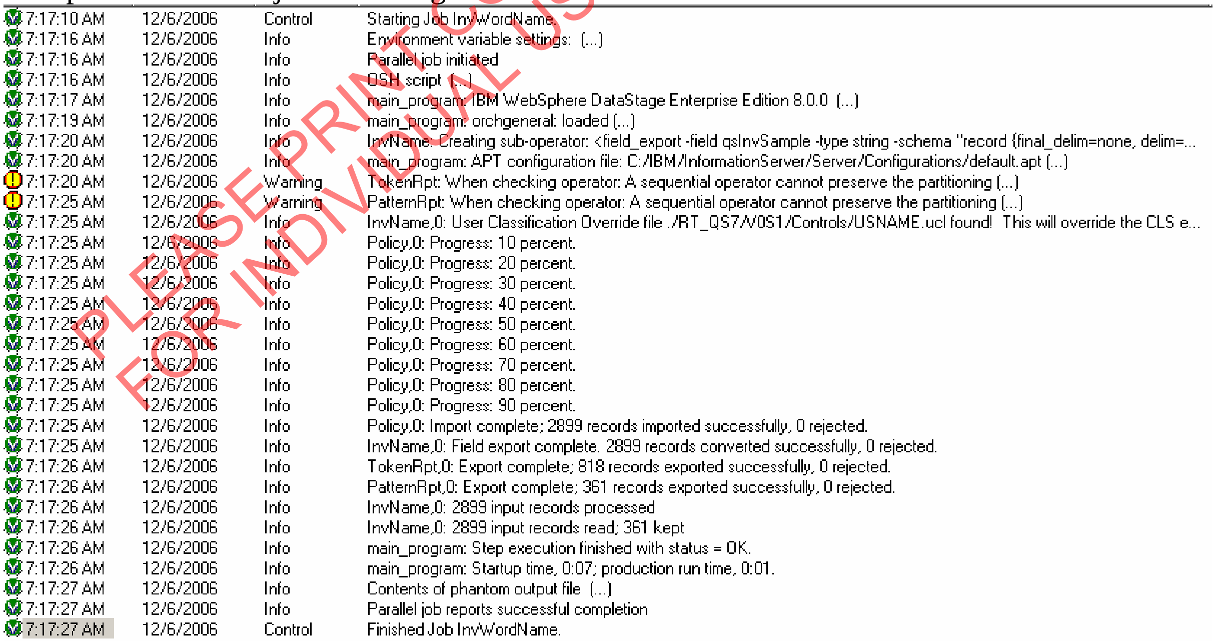


Output name: TokenRpt

Columns	
Expression	Column N.
Investigate(qsInvCount)	qsInvCount
Investigate(qsInvWord)	qsInvWord
Investigate(qsInvClassCo	qsInvClass(

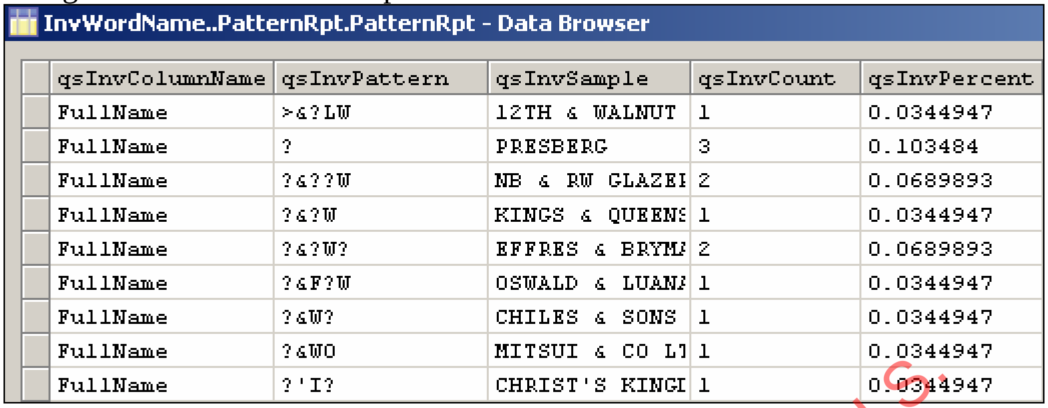
TokenRpt	
Derivation	Column Name
Investigate(qsInvCount)	qsInvCount
Investigate(qsInvWord)	qsInvWord
Investigate(qsInvClassCode)	qsInvClassCode

17. Click the OK button to return to the main Investigate screen.
18. Click OK to return to the job canvas.
19. Open the Sequential File stage named PatternRpt and set the File property to C:\WINNCRM\nvWordNamePattern.rpt.
20. Click OK to return to the job canvas.
21. Open the Sequential File stage named TokenRpt and set the File property to C:\WINNCRM\nvWordNameToken.rpt
22. Compile and run the job. Your log should resemble:



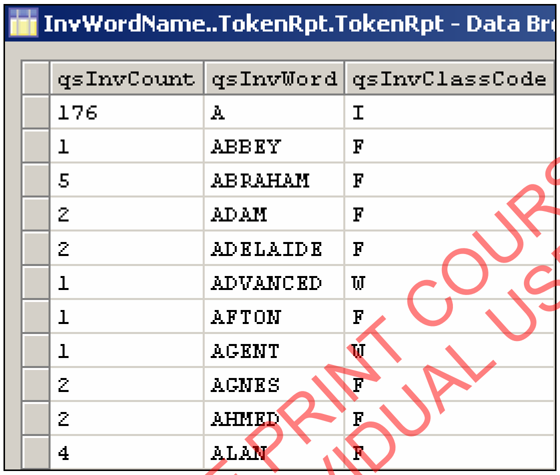
```
7:17:10 AM   12/6/2006   Control   Starting Job InvWordName.
7:17:16 AM   12/6/2006   Info      Environment variable settings:  (...)
7:17:16 AM   12/6/2006   Info      Parallel job initiated
7:17:16 AM   12/6/2006   Info      OSH script  (...)
7:17:17 AM   12/6/2006   Info      main_program: IBM WebSphere DataStage Enterprise Edition 8.0.0  (...)
7:17:19 AM   12/6/2006   Info      main_program: orchgeneral: loaded (...)
7:17:20 AM   12/6/2006   Info      InvName: Creating sub-operator: <field_export -field qsInvSample -type string -schema "record {final_delim=none, delim=...
7:17:20 AM   12/6/2006   Info      main_program: APT configuration file: C:/IBM/InformationServer/Server/Configurations/default.apt (...)
7:17:20 AM   12/6/2006   Warning   TokenRpt: When checking operator: A sequential operator cannot preserve the partitioning (...)
7:17:25 AM   12/6/2006   Warning   PatternRpt: When checking operator: A sequential operator cannot preserve the partitioning (...)
7:17:25 AM   12/6/2006   Info      InvName,0: User Classification Override file ./RT_QS7/V0S1/Controls/USNAME.ucl found!  This will override the CLS e...
7:17:25 AM   12/6/2006   Info      Policy,0: Progress: 10 percent.
7:17:25 AM   12/6/2006   Info      Policy,0: Progress: 20 percent.
7:17:25 AM   12/6/2006   Info      Policy,0: Progress: 30 percent.
7:17:25 AM   12/6/2006   Info      Policy,0: Progress: 40 percent.
7:17:25 AM   12/6/2006   Info      Policy,0: Progress: 50 percent.
7:17:25 AM   12/6/2006   Info      Policy,0: Progress: 60 percent.
7:17:25 AM   12/6/2006   Info      Policy,0: Progress: 70 percent.
7:17:25 AM   12/6/2006   Info      Policy,0: Progress: 80 percent.
7:17:25 AM   12/6/2006   Info      Policy,0: Progress: 90 percent.
7:17:25 AM   12/6/2006   Info      Policy,0: Import complete; 2899 records imported successfully, 0 rejected.
7:17:25 AM   12/6/2006   Info      InvName,0: Field export complete. 2899 records converted successfully, 0 rejected.
7:17:26 AM   12/6/2006   Info      TokenRpt,0: Export complete; 818 records exported successfully, 0 rejected.
7:17:26 AM   12/6/2006   Info      PatternRpt,0: Export complete; 361 records exported successfully, 0 rejected.
7:17:26 AM   12/6/2006   Info      InvName,0: 2899 input records processed
7:17:26 AM   12/6/2006   Info      InvName,0: 2899 input records read; 361 kept
7:17:26 AM   12/6/2006   Info      main_program: Step execution finished with status = OK.
7:17:26 AM   12/6/2006   Info      main_program: Startup time, 0:07; production run time, 0:01.
7:17:27 AM   12/6/2006   Info      Contents of phantom output file  (...)
7:17:27 AM   12/6/2006   Info      Parallel job reports successful completion
7:17:27 AM   12/6/2006   Control   Finished Job InvWordName.
```

23. Use the View Data button for both the PatternRpt and TokenRpt Sequential File stages in Designer to review the two report files.



InvWordName..PatternRpt.PatternRpt - Data Browser

qsInvColumnName	qsInvPattern	qsInvSample	qsInvCount	qsInvPercent
FullName	>&?LW	12TH & WALNUT	1	0.0344947
FullName	?	PRESBERG	3	0.103484
FullName	?&??W	NB & RW GLAZEI	2	0.0689893
FullName	?&?W	KINGS & QUEENS	1	0.0344947
FullName	?&?W?	EFFRES & BRYMA	2	0.0689893
FullName	?&F?W	OSWALD & LUANA	1	0.0344947
FullName	?&W?	CHILES & SONS	1	0.0344947
FullName	?&WO	MITSUI & CO LT	1	0.0344947
FullName	?'I?	CHRIST'S KINGI	1	0.0344947



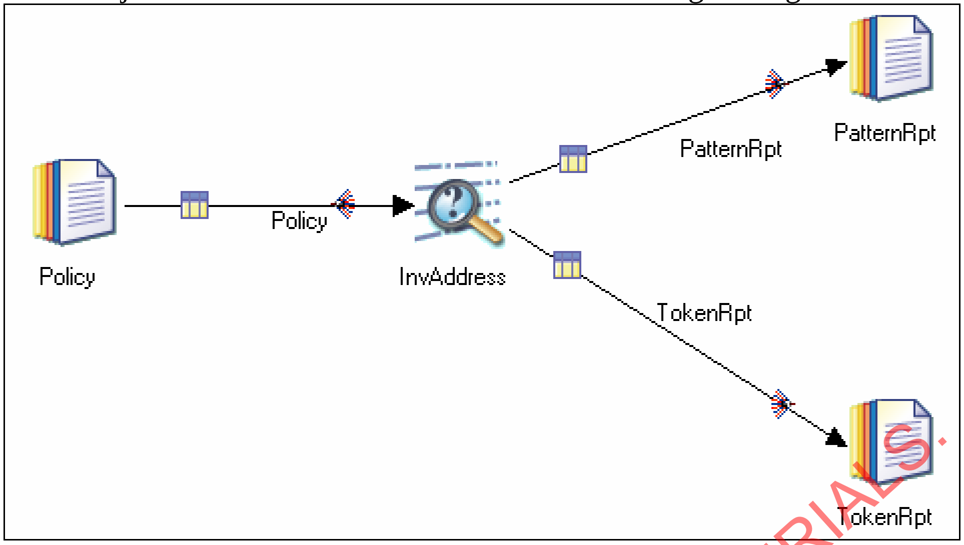
InvWordName..TokenRpt.TokenRpt - Data Br

qsInvCount	qsInvWord	qsInvClassCode
176	A	I
1	ABBEY	F
5	ABRAHAM	F
2	ADAM	F
2	ADELAIDE	F
1	ADVANCED	W
1	AFTON	F
1	AGENT	W
2	AGNES	F
2	AHMED	F
4	ALAN	F

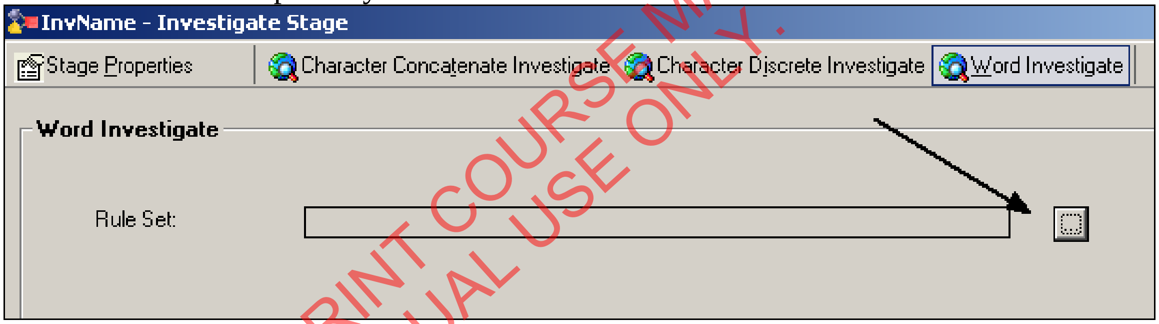
Task: Build and run a word investigate for the Address fields

1. Edit your InvWordName job.

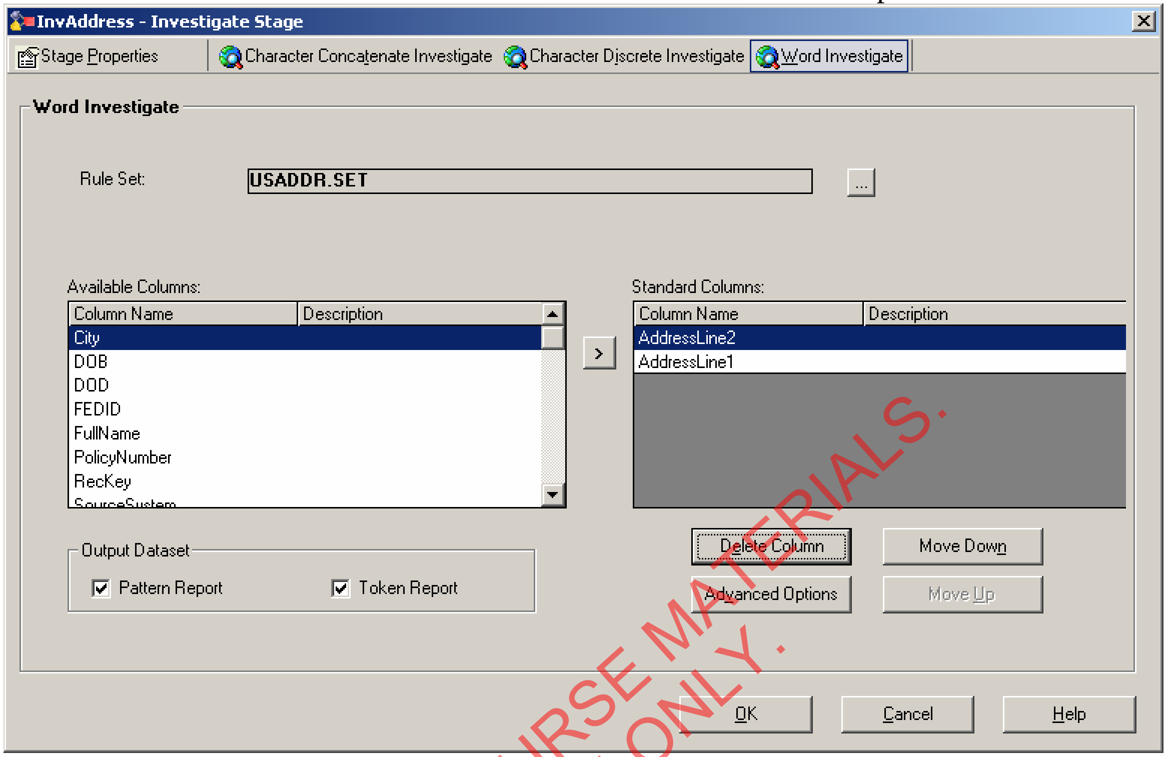
2. Save the job as InvWordAddress. Rename the Investigate stage to InvAddress.



3. Edit the InvAddress Investigate stage.
4. Use the Rule Set browse button to select the USADDR rule set from the Standardization Rules folder in the repository.


InvName - Investigate Stage
Stage Properties
Character Concatenate Investigate
Character Discrete Investigate
Word Investigate
Word Investigate
Rule Set:

5. Select AddressLine1 and AddressLine2 from the available columns pane.



Rule Set: USADDR.SET

Available Columns:

Column Name	Description
City	
DOB	
DOD	
FEDID	
FullName	
PolicyNumber	
RecKey	
SourceSystem	

Standard Columns:

Column Name	Description
AddressLine2	
AddressLine1	

Output Dataset

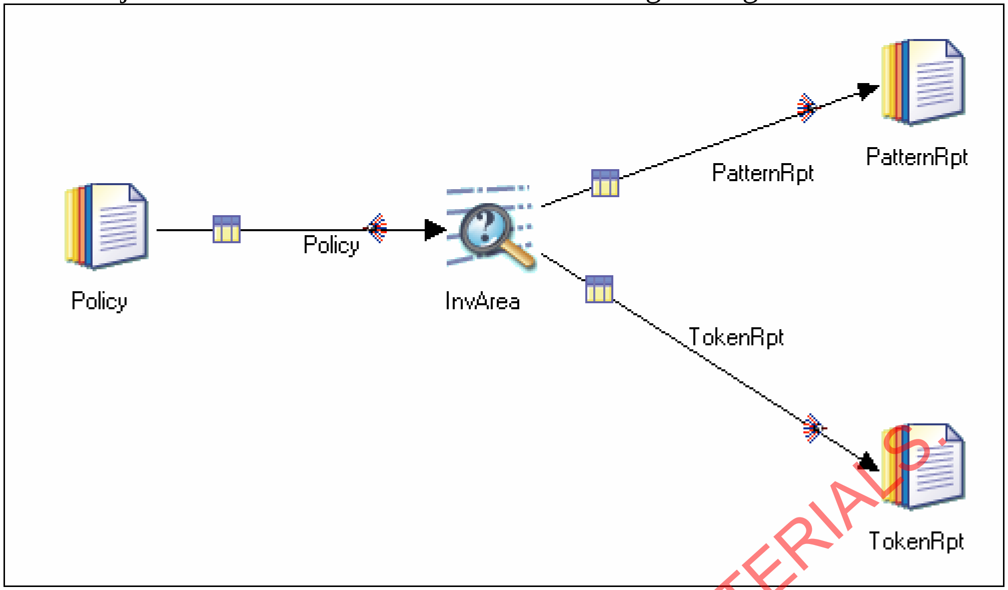
☑ Pattern Report ☑ Token Report

6. Leave the Output Dataset checkboxes checked.
7. The Stage Property settings – link ordering and column mapping – are already correct.
8. Click the OK button to return to the canvas.
9. Edit the PatternRpt Sequential File stage and set the File property to C:\WINNCRM\InvWordAddressPattern.rpt. Return to the job canvas.
10. Edit the TokenRpt Sequential File stage and set the File property to C:\WINNCRM\InvWordAddressToken.rpt. Return to the job canvas.
11. Compile and run your job.
12. Review the pattern and token reports.

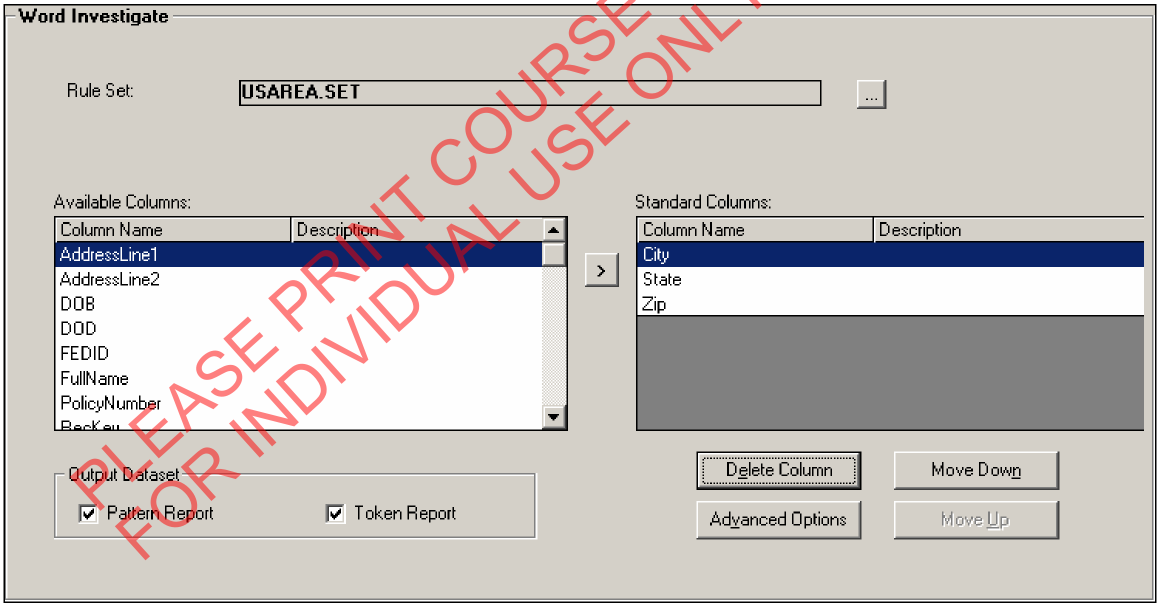
Task: Build and run a word investigate for the Area fields

1. Edit your InvWordName job.

2. Save the job as InvWordArea. Rename the Investigate stage to InvArea.



3. Edit the InvArea Investigate stage.
4. Modify the Investigate stage to use the USAREA rule set on the City, State, and Zip fields. The result should be:



Word Investigate

Rule Set: USAREA.SET

Available Columns:

Column Name	Description
AddressLine1	
AddressLine2	
DOB	
DOD	
FEDID	
FullName	
PolicyNumber	
RecKey	

Standard Columns:

Column Name	Description
City	
State	
Zip	

Output Dataset

☑ Pattern Report ☑ Token Report

5. Close the Investigate stage.
6. Edit the PatternRpt Sequential File stage and change the File property to C:\WINNCRM\InvWordAreaPattern.rpt and then close the stage.
7. Edit the TokenRpt Sequential File stage and change the File property to C:\WINNCRM\InvWordAreaToken.rptand then close the stage.
8. Compile and run the job.
9. View the pattern and token reports.

Lab 10: Standardize Country

Objectives:

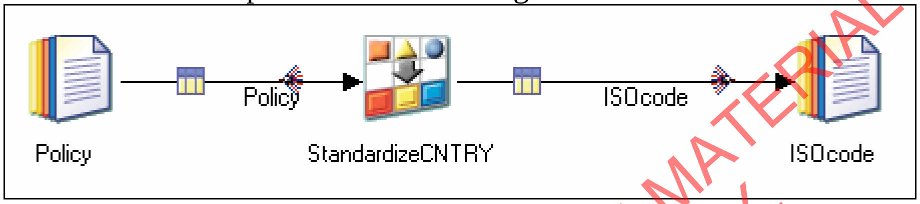
- Add ISO country code based on address and area

Assumptions:

- DataStage project named WINNCRM exists
- Job InvCharDisCmask exists and executed successfully

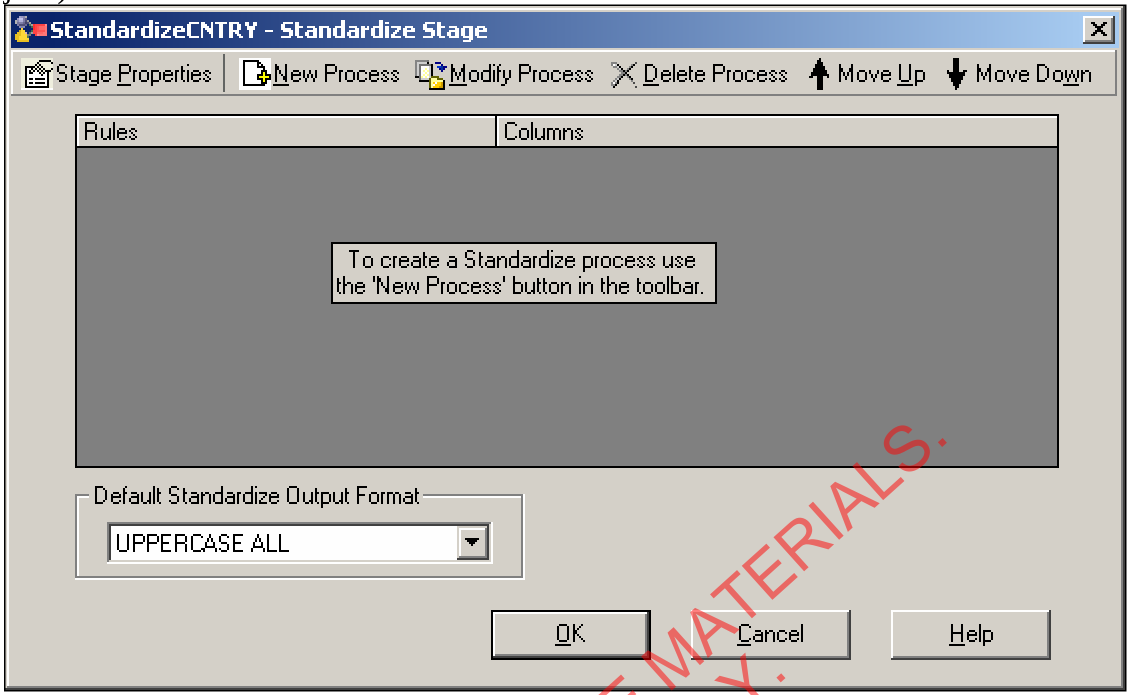
Task: Build and run a Standardize to add ISO country code

1. Create a new parallel job. Add Sequential File and Standardize stages as well as associated links to produce the following:



2. Save your job and name it stanCNTRY.
3. Open the Policy Sequential File stage and complete the File property.
4. Select the Columns tab and import the column definitions from the Policy table. Check your settings by using the View Data button.
5. Open the StandardizeCNTRY stage. Note the tabs near the top and the hint in the middle of the screen. You will use the New Process tab to point to the appropriate standardization rule set (these are the same rule sets you used in the Word Investigation

jobs).



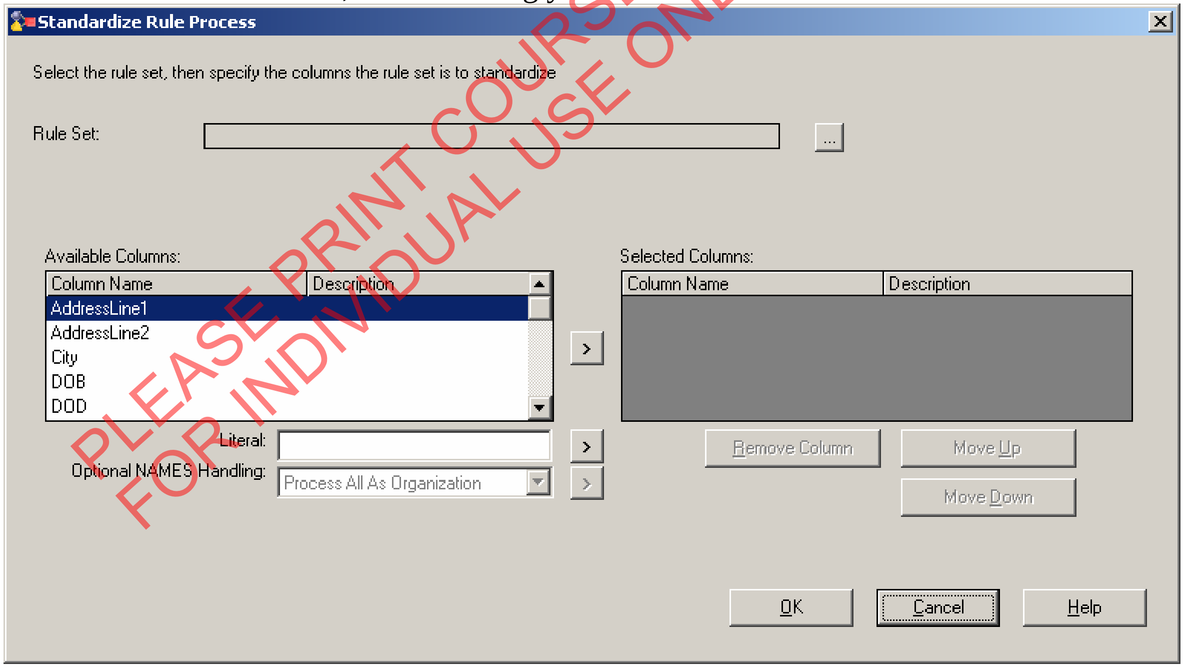
StandardizeCNTRY - Standardize Stage

Rules	Columns

To create a Standardize process use the 'New Process' button in the toolbar.

Default Standardize Output Format: UPPERCASE ALL

6. Click the New Process tab; this will bring you to the rule set selection screen.



Standardize Rule Process

Select the rule set, then specify the columns the rule set is to standardize

Rule Set:

Available Columns:

Column Name	Description
AddressLine1	
AddressLine2	
City	
DOB	
DOD	

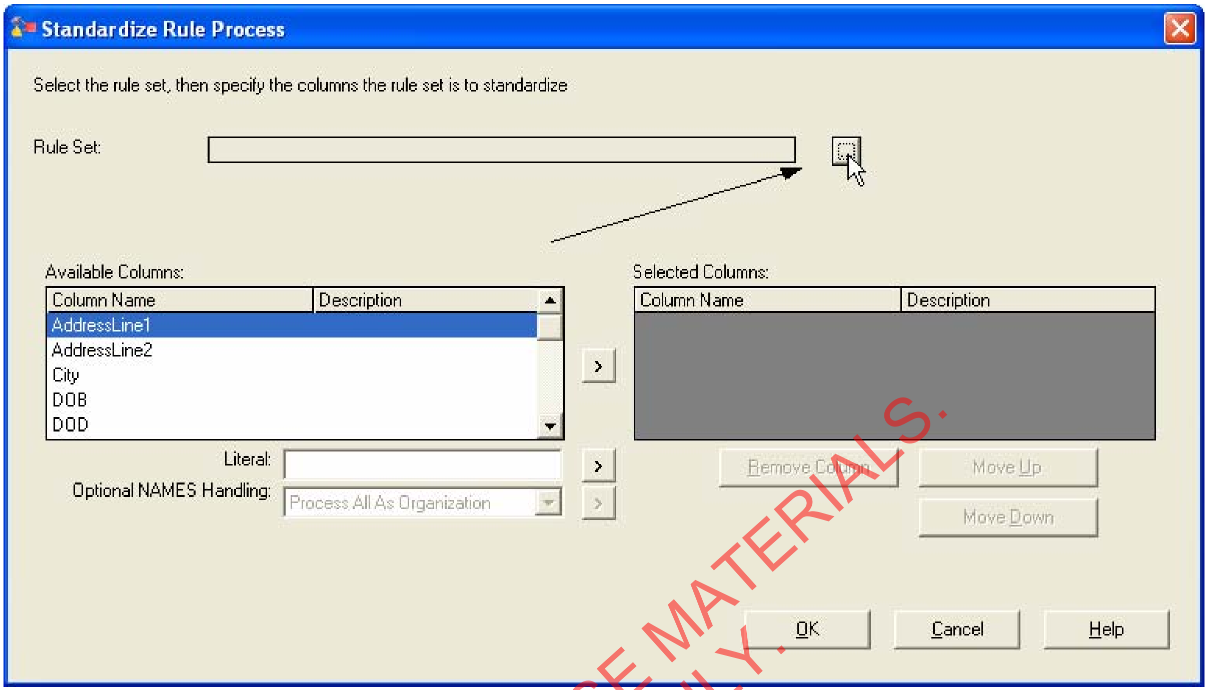
Selected Columns:

Column Name	Description

Literal:

Optional NAMES Handling: Process All As Organization

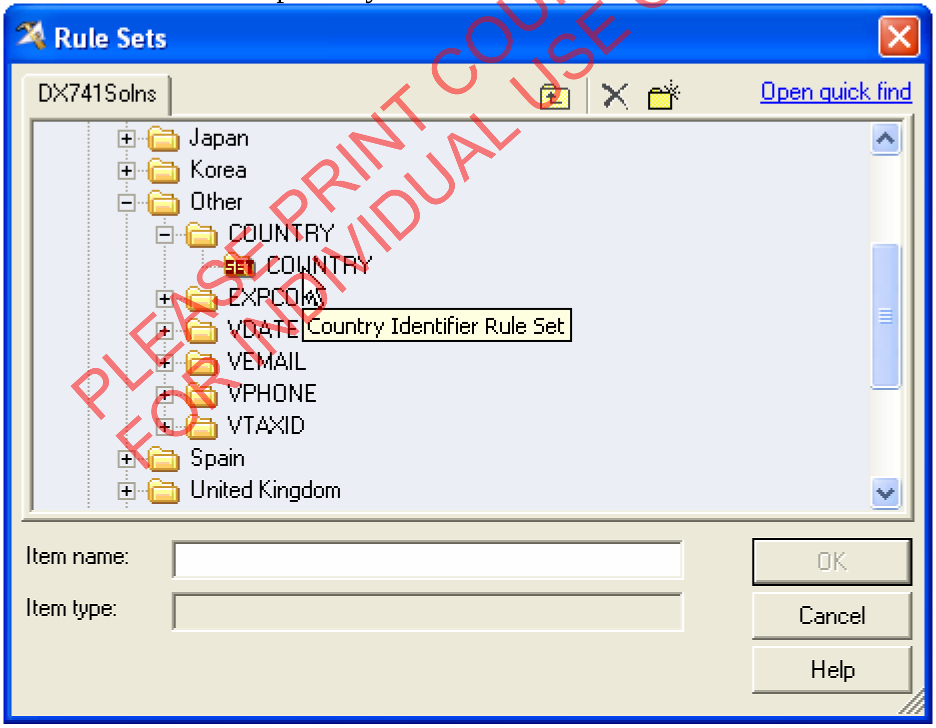
7. Click the …. Browse button.



Standardize Rule Process

Select the rule set, then specify the columns the rule set is to standardize

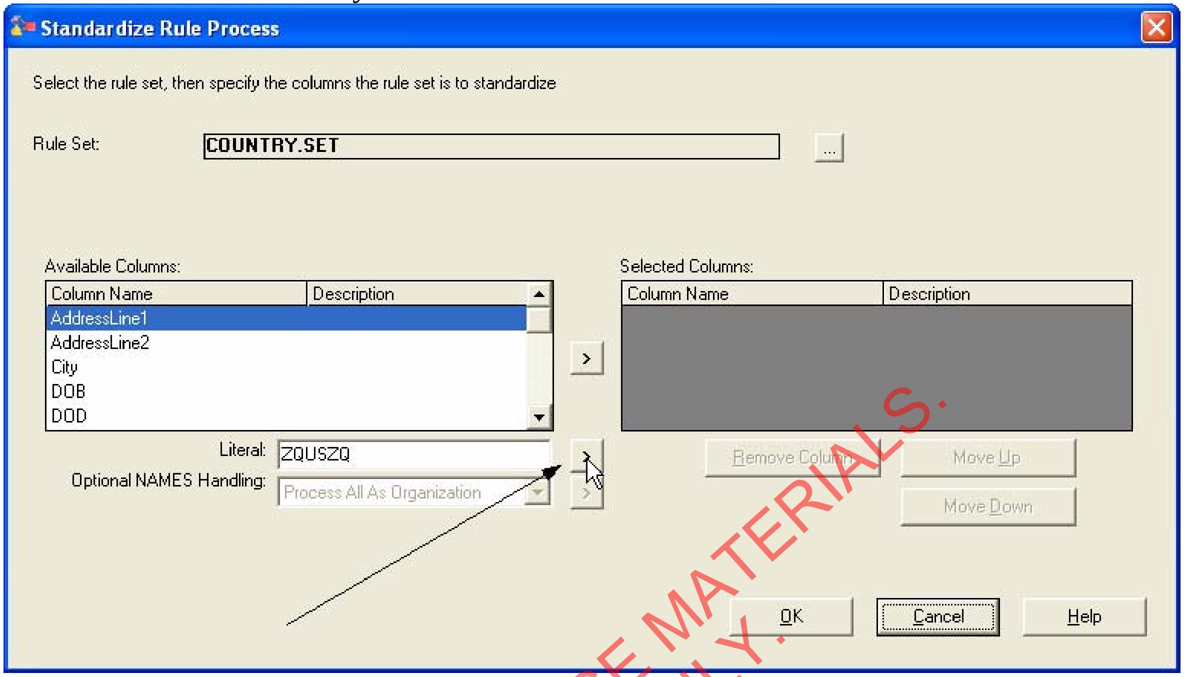
8. Find and select the COUNTRY rule set from the Other folder in the Standardization Rules folder of the repository.



Rule Sets

9. Enter ZQUSZQ in the Literal text box and click the > icon to add to the Selected Columns pane. The string "ZQ" is used to delimit values that you intend to pass as

parameters to a rule set. In this case "US" is being passed to the rule set to provide a default value for the country code.



Standardize Rule Process

Select the rule set, then specify the columns the rule set is to standardize

Rule Set: COUNTRY.SET

Available Columns:

Column Name	Description
AddressLine1	
AddressLine2	
City	
DOB	
DOD	

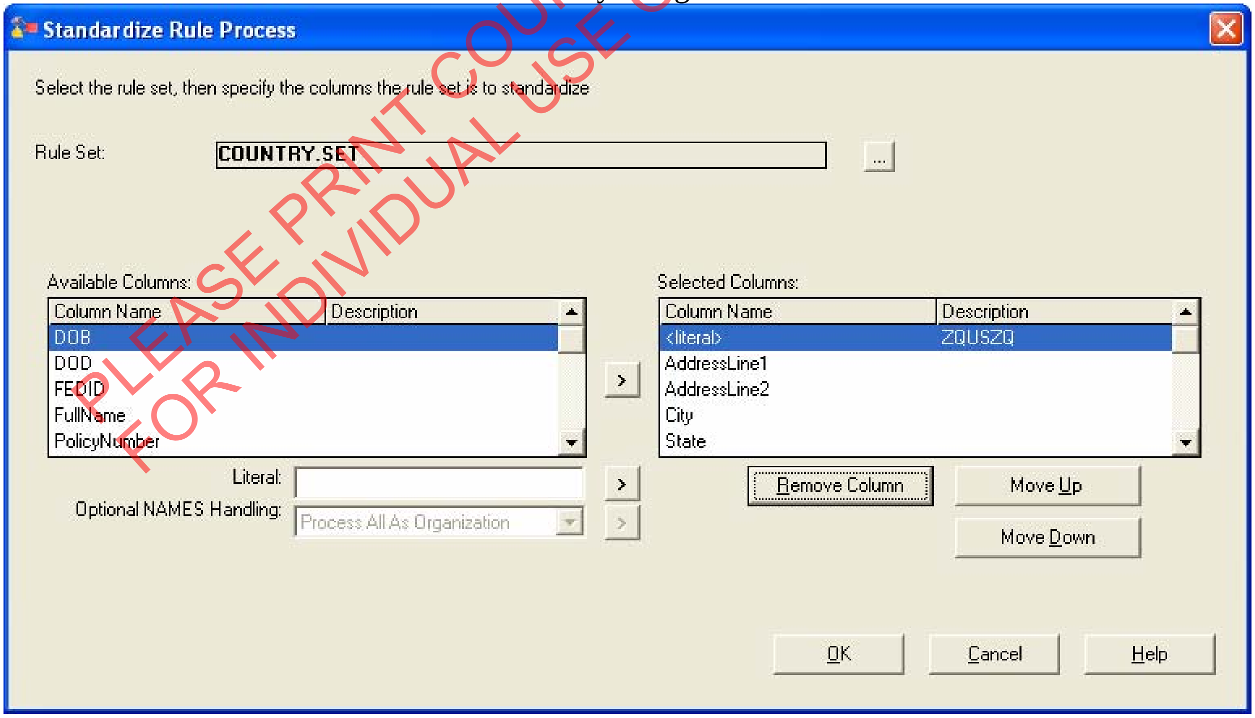
Selected Columns:

Column Name	Description

Literal: ZQUSZQ

Optional NAMES Handling: Process All As Organization

10. Select the columns AddressLine1, AddressLine2, City, State, and Zip from the Available Columns and move to the Selected Columns by using the > icon.



Standardize Rule Process

Select the rule set, then specify the columns the rule set is to standardize

Rule Set: COUNTRY.SET

Available Columns:

Column Name	Description
DOB	
DOD	
FEDID	
FullName	
PolicyNumber	

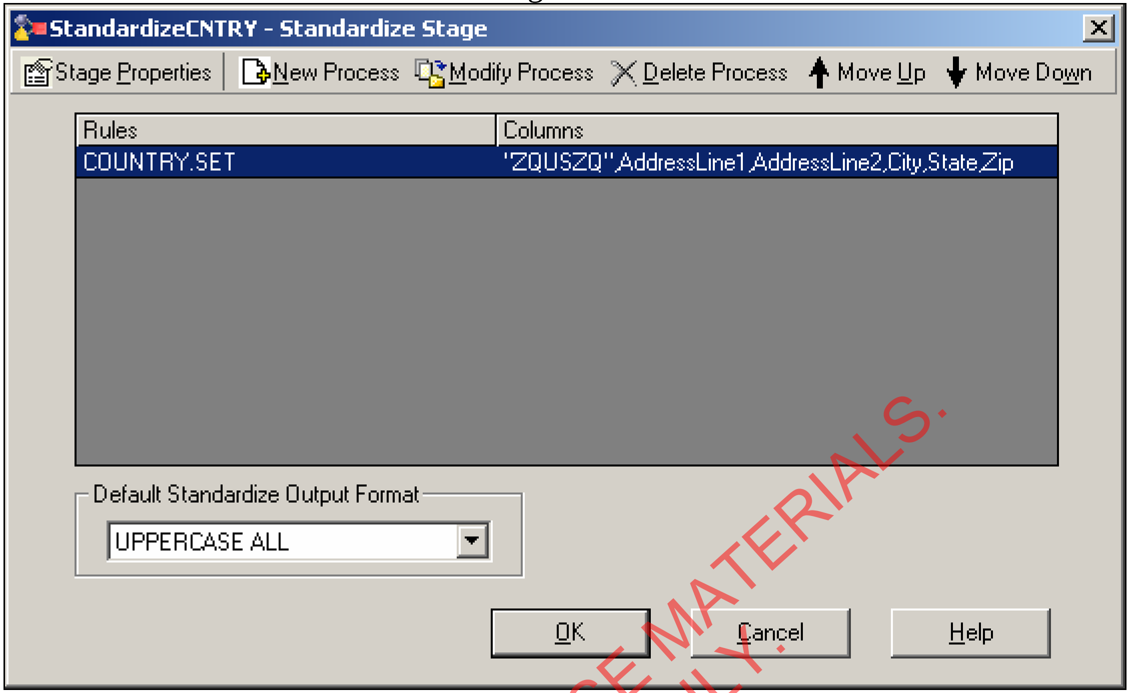
Selected Columns:

Column Name	Description
<literal>	ZQUSZQ
AddressLine1	
AddressLine2	
City	
State	

Literal:

Optional NAMES Handling: Process All As Organization

11. Click OK to return to the Standardize stage main screen.



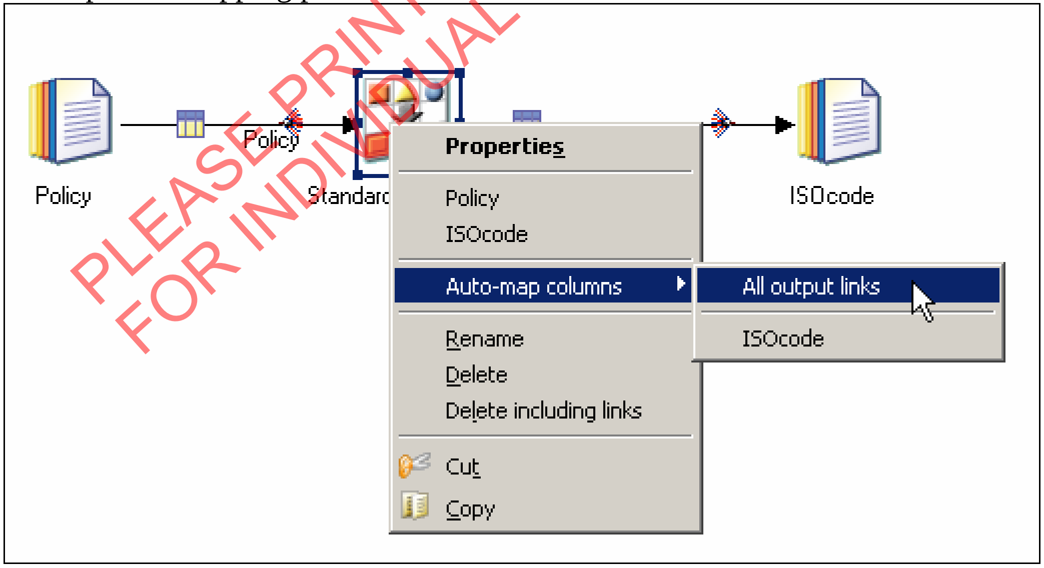
StandardizeCNTRY - Standardize Stage

Rules	Columns
COUNTRY.SET	"ZQUSZQ",AddressLine1,AddressLine2,City,State,Zip

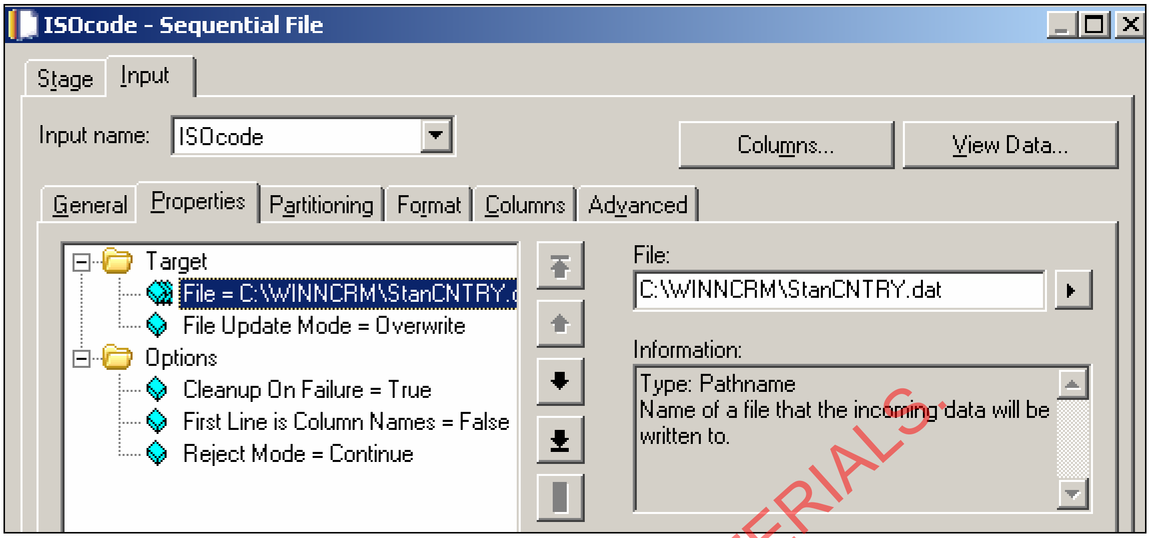
Default Standardize Output Format: UPPERCASE ALL

12. Click OK to return to the job canvas.

13. Right-click on the Standardize stage and select the Auto-map columns > All output links option. This will map all input columns plus columns from the standardization process to the output link. Most stages contain this option as an alternative to the Stage Properties ➔Output ➔ Mapping process used in earlier exercises.



14. Edit the ISOcode Sequential File and set the File property to C:\WINNCRM\StanCNTRY.dat



Input name: ISOcode

- Target
  - File = C:\WINNCRM\StanCNTRY.d
  - File Update Mode = Overwrite
- Options
  - Cleanup On Failure = True
  - First Line is Column Names = False
  - Reject Mode = Continue

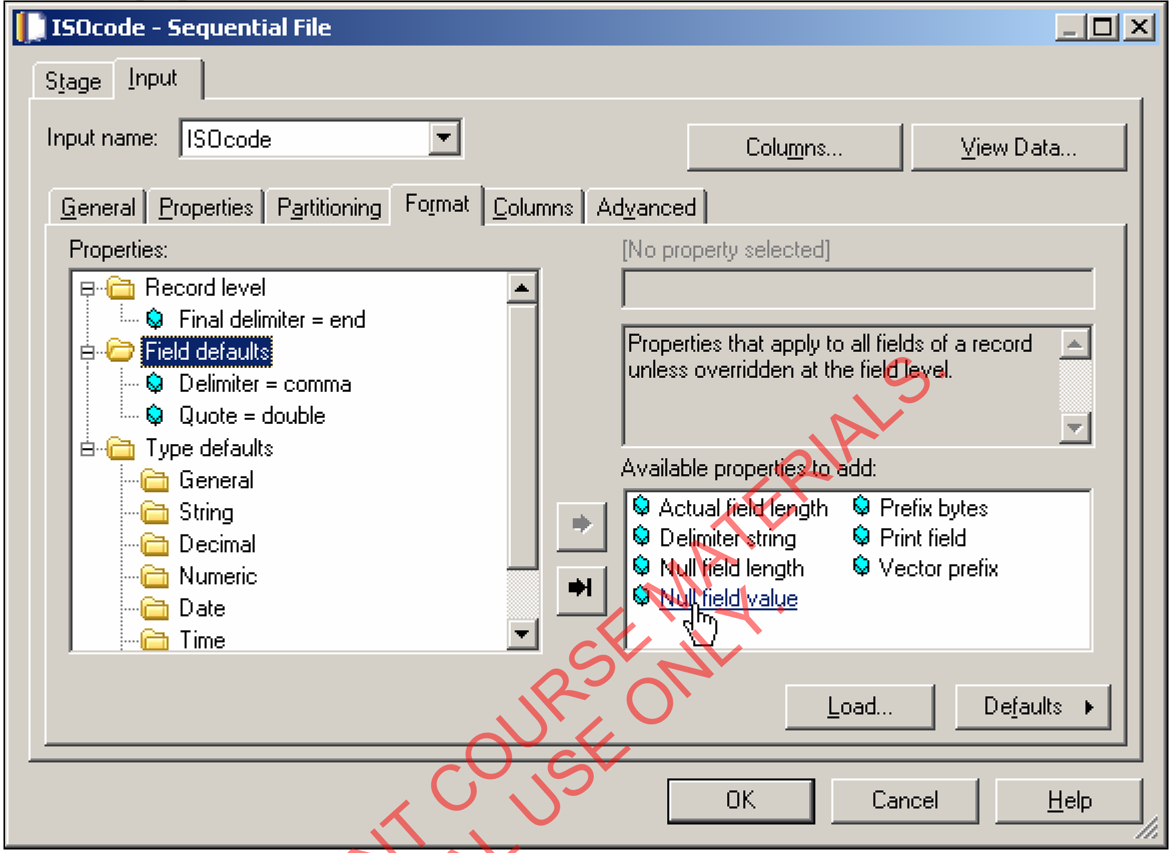
File: C:\WINNCRM\StanCNTRY.dat

Information:
Type: Pathname
Name of a file that the incoming data will be written to.

15. Compile and run your job. Be sure to examine the log file. If your log file contains a number of warnings about null handling, go to the format tab of your target file and verify that the Null field value has been specified. This problem can occur because the sequential file stage will not write records to disk that contain a null value in any of the fields. For each record effected, a log message is generated.
    Two methods to handle this condition are available:
    1. Create a null handling condition in the target sequential file stage. This is the method used earlier in this course by setting the Null field value.
    2. Use a dataset stage in place of the target sequential file stage.

    Both methods are described next.
16. Edit the ISOcode Sequential File stage and select the Format tab.

17. Select the Field Defaults branch and then click on the Null field value property in the Available properties to add window.

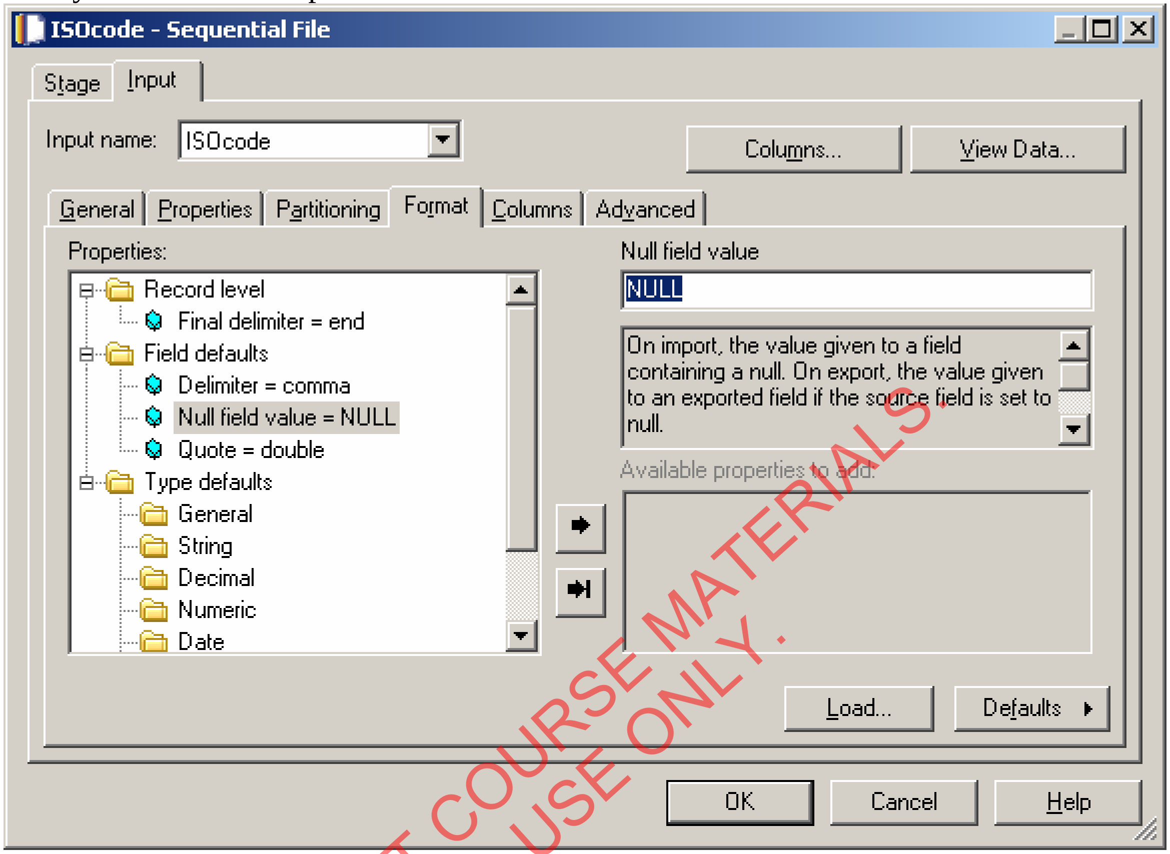


Record level
Final delimiter = end
Field defaults
Delimiter = comma
Quote = double
Type defaults
General
String
Decimal
Numeric
Date
Time

Properties that apply to all fields of a record unless overridden at the field level.

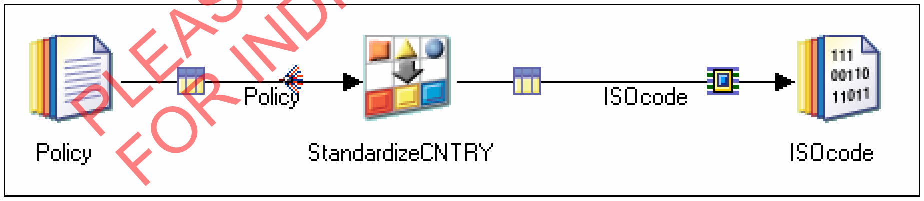
Actual field length
Delimiter string
Null field length
Null field value
Prefix bytes
Print field
Vector prefix

18. Set the Null field value to NULL. There is nothing significant about this string; it could be any value that is not present in the data.



On import, the value given to a field containing a null. On export, the value given to an exported field if the source field is set to null.

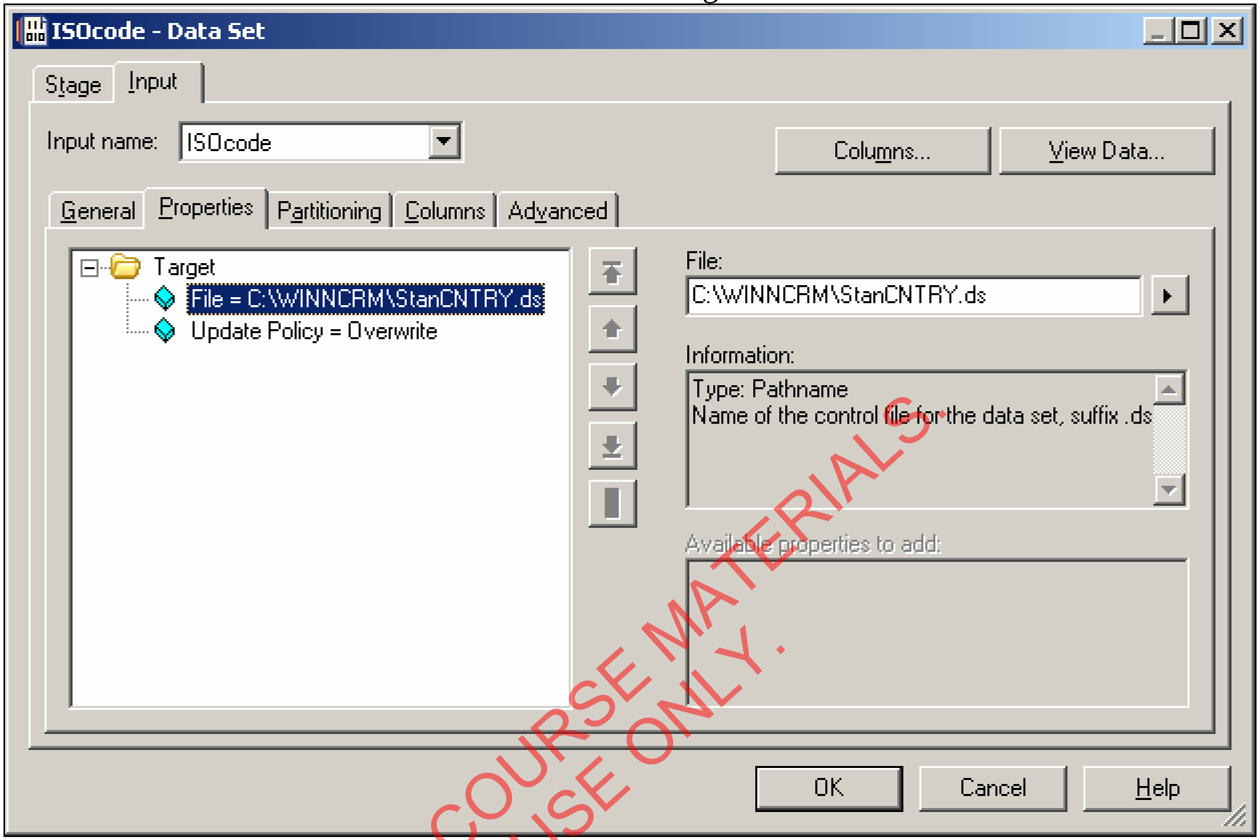
19. Save your job, recompile and run. The warnings should largely disappear.
20. Use the view data button in the target sequential file to examine the results. Note the presence of the literal “NULL” in some of the field values.
21. Method 2 consists of replacing the ISOcode Sequential File stage with a Data Set stage such as:



You can build this configuration by deleting the Sequential File stage, dragging a Data Set stage from the File folder in the Palette and reattaching the link.

22. Open the ISOcode Data Set stage and set the File property to C:\WINNCRM\StanCNTRY.ds. You should always use the .ds suffix with a Data Set

file. Note that no Format tab exists for the Data Set stage.



Input name: ISOcode

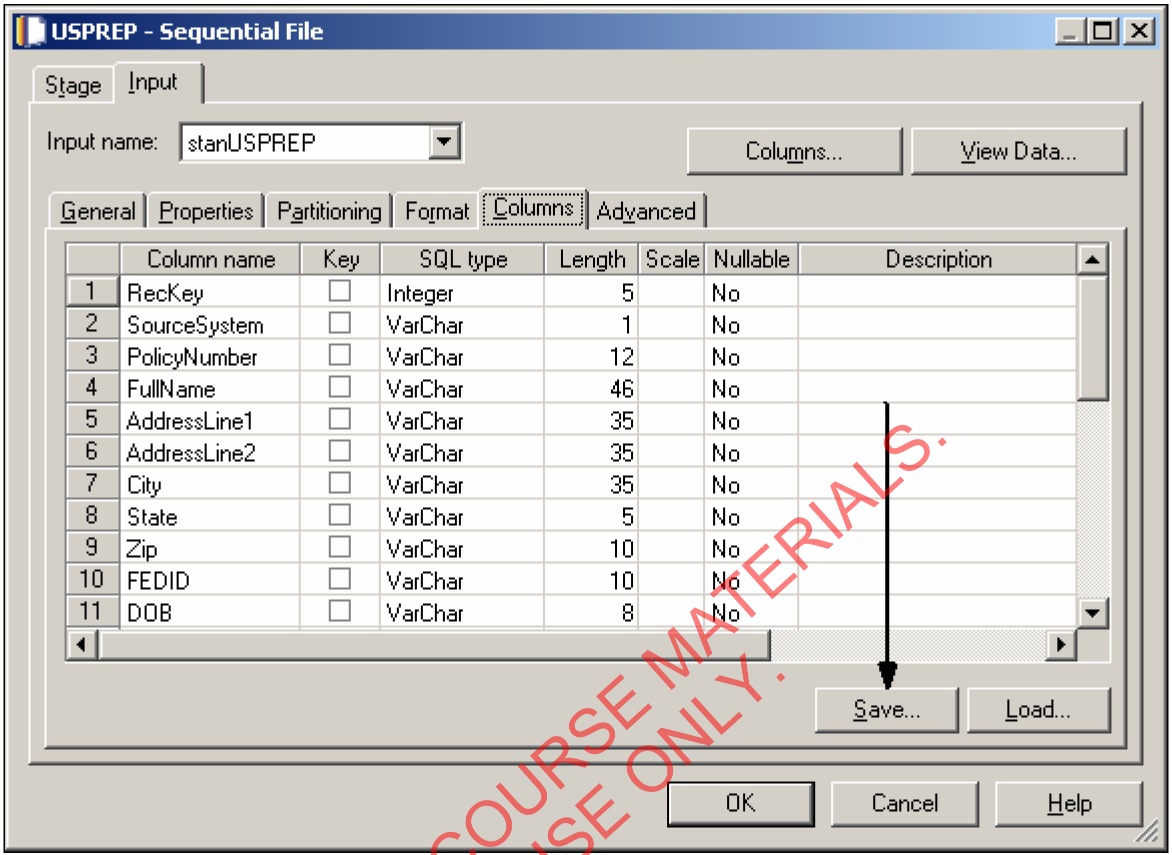
Target
File = C:\WINNCRM\StanCNTRY.ds
Update Policy = Overwrite

File: C:\WINNCRM\StanCNTRY.ds

Information:
Type: Pathname
Name of the control file for the data set, suffix .ds

23. Save your job. Recompile and run.
24. Use the Data Set's View Data button from Designer to review the output data set.
25. The target file (either sequential file or data set) contains new columns that were added by the rule set from the standardization stage. Since this target file will be used in subsequent jobs, it is necessary to save the target file's meta data.
26. Edit the target file and click the Columns tab.

27. Click the Save button.

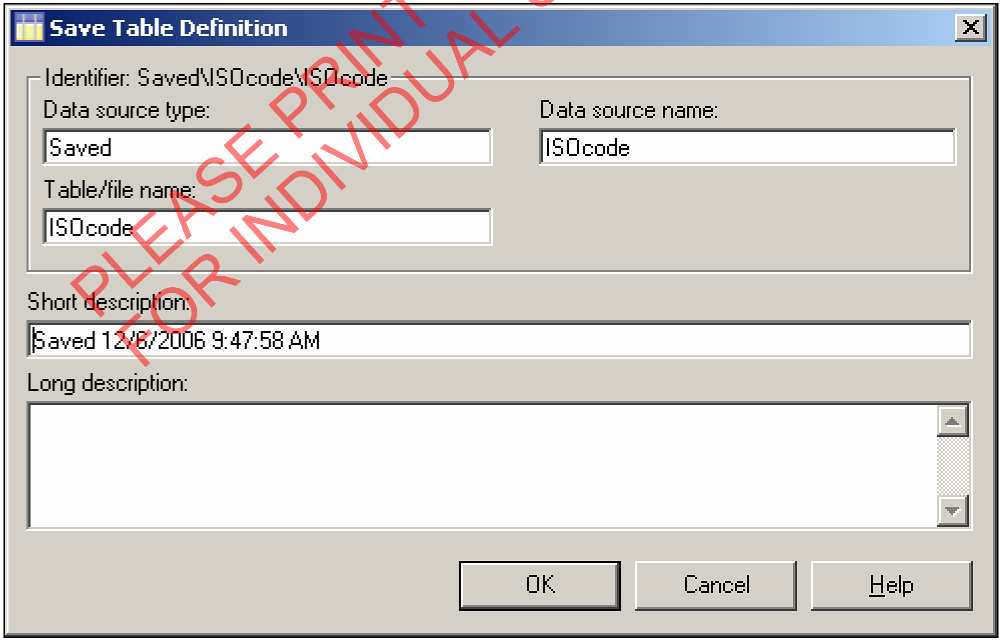


USPREP - Sequential File

Input name: stanUSPREP

	Column name	Key	SQL type	Length	Scale	Nullable	Description
1	RecKey	□	Integer	5		No	
2	SourceSystem	□	VarChar	1		No	
3	PolicyNumber	□	VarChar	12		No	
4	FullName	□	VarChar	46		No	
5	AddressLine1	□	VarChar	35		No	
6	AddressLine2	□	VarChar	35		No	
7	City	□	VarChar	35		No	
8	State	□	VarChar	5		No	
9	Zip	□	VarChar	10		No	
10	FEDID	□	VarChar	10		No	
11	DOB	□	VarChar	8		No	

28. Click OK on the Save Definition screen.



Save Table Definition

Identifier: Saved\ISOcode\ISOcode

Data source type: Saved

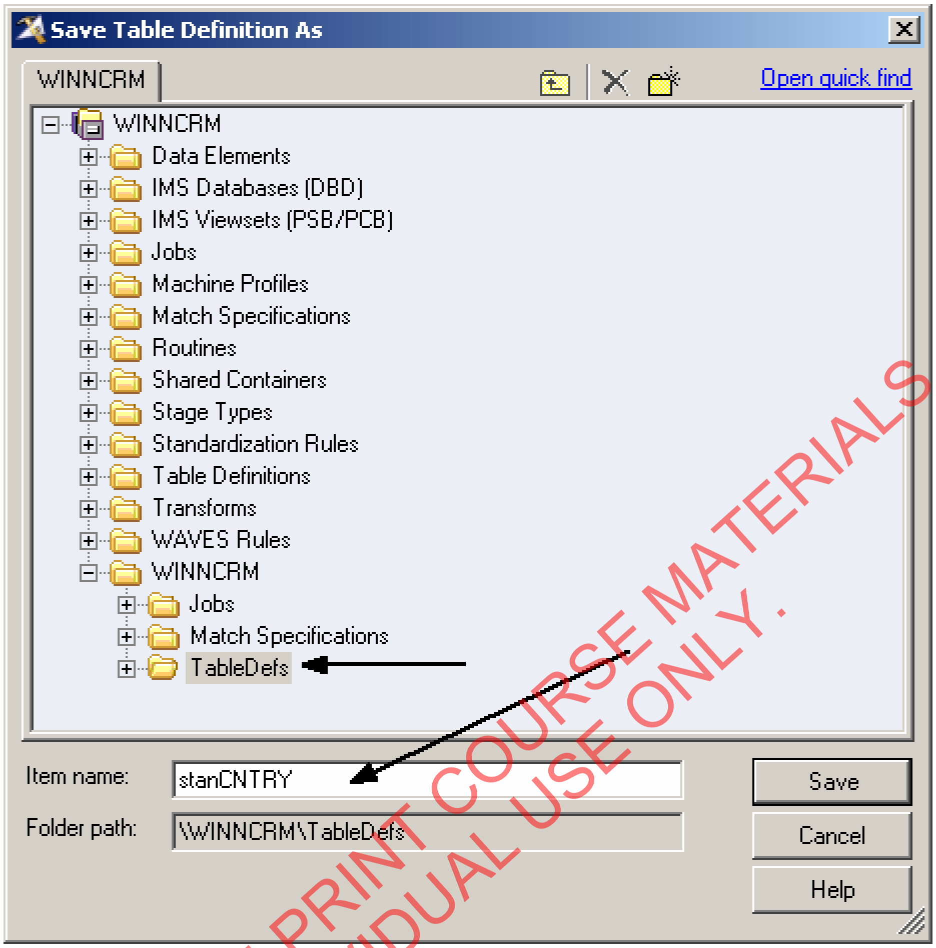
Data source name: ISOcode

Table/file name: ISOcode

Short description: Saved 12/6/2006 9:47:58 AM

Long description:

29. Select the TableDefs folder in the WINNCRM folder of the repository and name your table definition stanCNTRY.



Item name: stanCNTRY

Folder path: \WINNCRM\TableDefs

30. Click the Save button to store the definition. You will be returned to the Columns tab of the stage editor.
31. Click OK to exit the stage editor.

Lab 11: Select US records

Objectives:

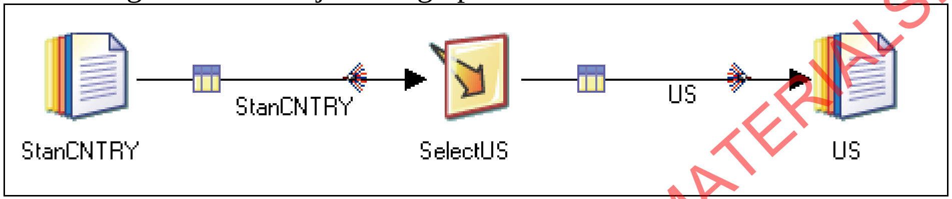
- Use ISO country code to select US records

Assumptions:

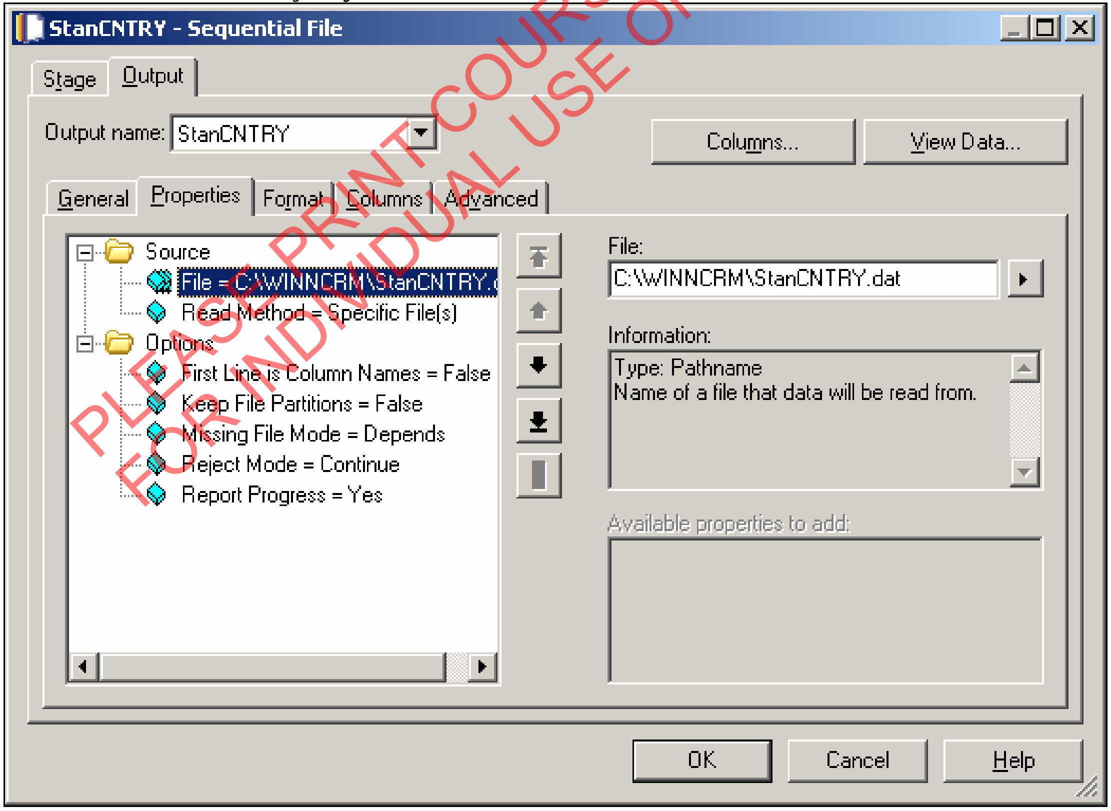
- Job stanCNTRY executed successfully

Task: Build and run a DataStage select job

1. Create a new parallel job. Add Sequential File and Filter stages as well as associated links to produce the following graphic depicted below. The Filter stage is located in the Processing folder of the job design palette.



2. Save your job and name it SelectUS.
3. Edit the StanCNTRY source file stage and set the File property to point to the output file from the StanCNTRY job you created earlier.



Output name: StanCNTRY

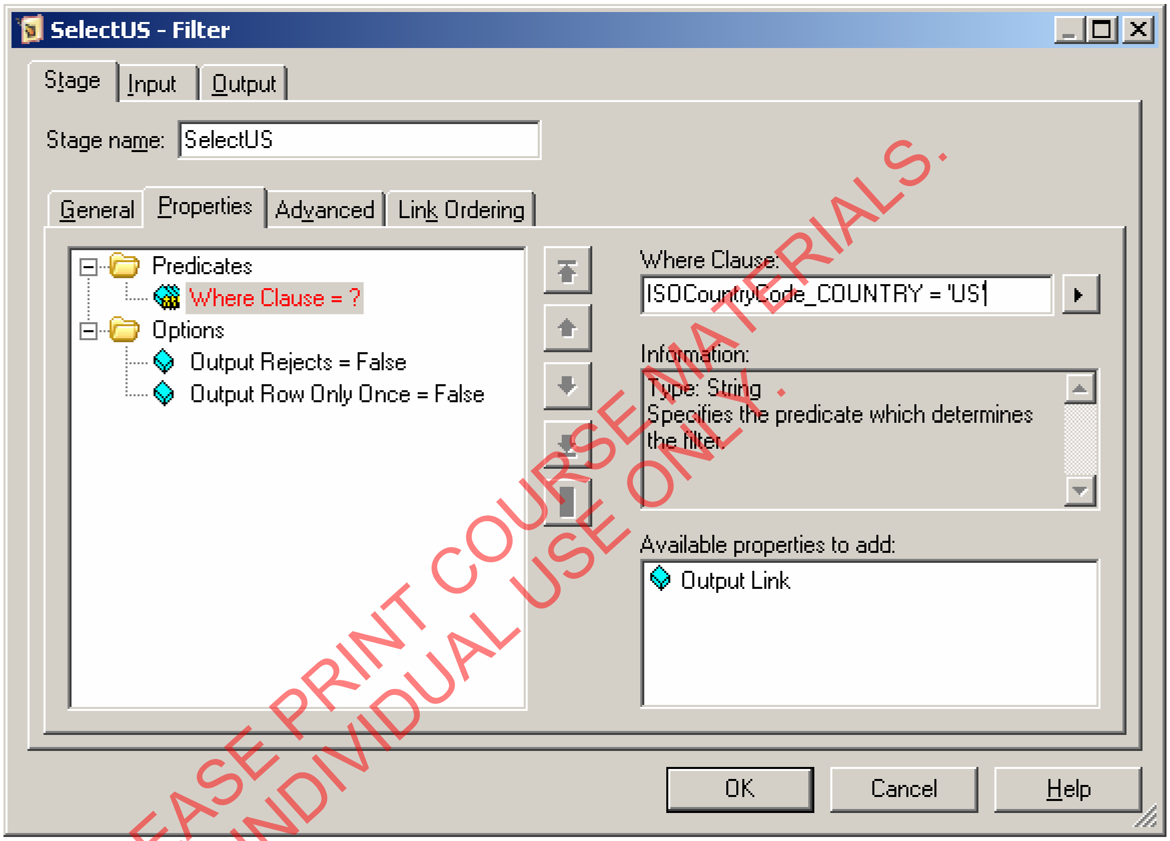
- Source
  - File = C:\WINNCRM\StanCNTRY.
  - Read Method = Specific File(s)
- Options
  - First Line is Column Names = False
  - Keep File Partitions = False
  - Missing File Mode = Depends
  - Reject Mode = Continue
  - Report Progress = Yes

File: C:\WINNCRM\StanCNTRY.dat

Information:
Type: Pathname
Name of a file that data will be read from.

4. Click the Columns tab and then click the Load button.

5. Load the StanCNTRY table definition from the WINNCRM ➔ TableDefs folder in the repository. You will be returned to the stage editor.
6. Use the View Data button to verify connectivity to the source file.
7. Exit the stage editor to return to the job design canvas.
8. Open the SelectUS Filter stage.
   Select the Where Clause = property and enter ISOCountryCode_COUNTRY = 'US'. ISOCountryCode_COUNTRY is a field from the source file.



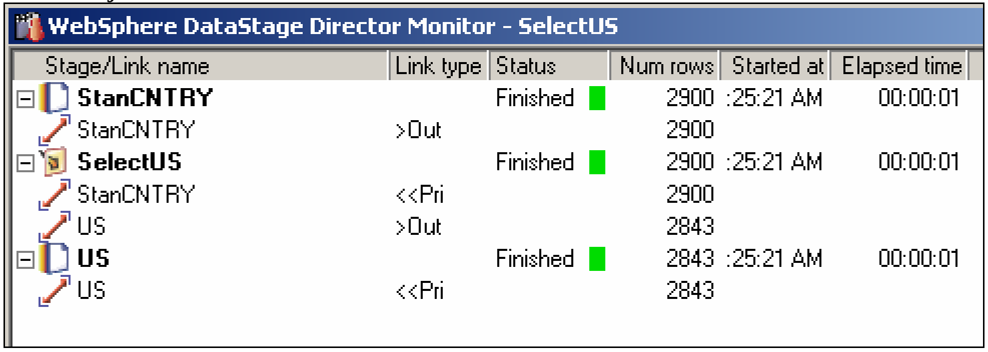
Stage name: SelectUS

Where Clause: ISOCountryCode_COUNTRY = 'US'

Information: Type: String. Specifies the predicate which determines the filter.

9. Click OK to close the stage editor.
10. Right-click on the SelectUS Filter stage and auto-map columns to all output links.
11. Edit the US Sequential File stage and set the File property to C:\WINNCRM\StanCNTRYUS.dat.
12. Click the Format tab and set the Null field value to NULL.
13. Compile and run the job.

14. Use the job monitor to review statistics. In this run 2843 records were selected as US.



WebSphere DataStage Director Monitor - SelectUS

Stage/Link name	Link type	Status	Num rows	Started at	Elapsed time
StanCNTRY		Finished 🟩	2900	:25:21 AM	00:00:01
StanCNTRY	>Out		2900		
SelectUS		Finished 🟩	2900	:25:21 AM	00:00:01
StanCNTRY	<<Pri		2900		
US	>Out		2843		
US		Finished 🟩	2843	:25:21 AM	00:00:01
US	<<Pri		2843		

15. Note: Since no columns were added or deleted by this job, it is not necessary to save the target file's column definitions to the repository. Subsequent jobs can use the StanCNTRY definition against the US-selected records

Lab 12: Standardize USPREP – Domain Pre-processing

Objectives:

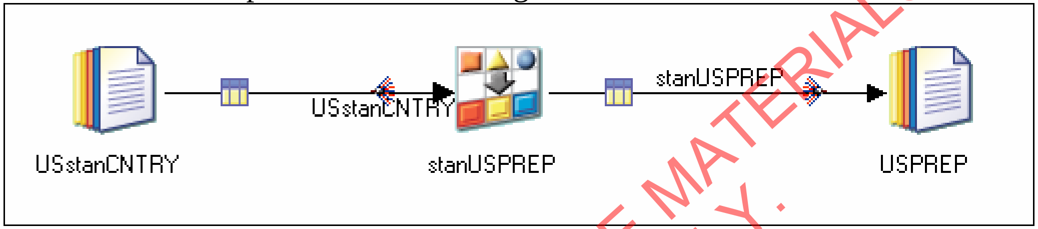
- Apply the USPREP rule set to filter name components from address fields, and area components from address fields

Assumptions:

- SelectUS job ran successfully

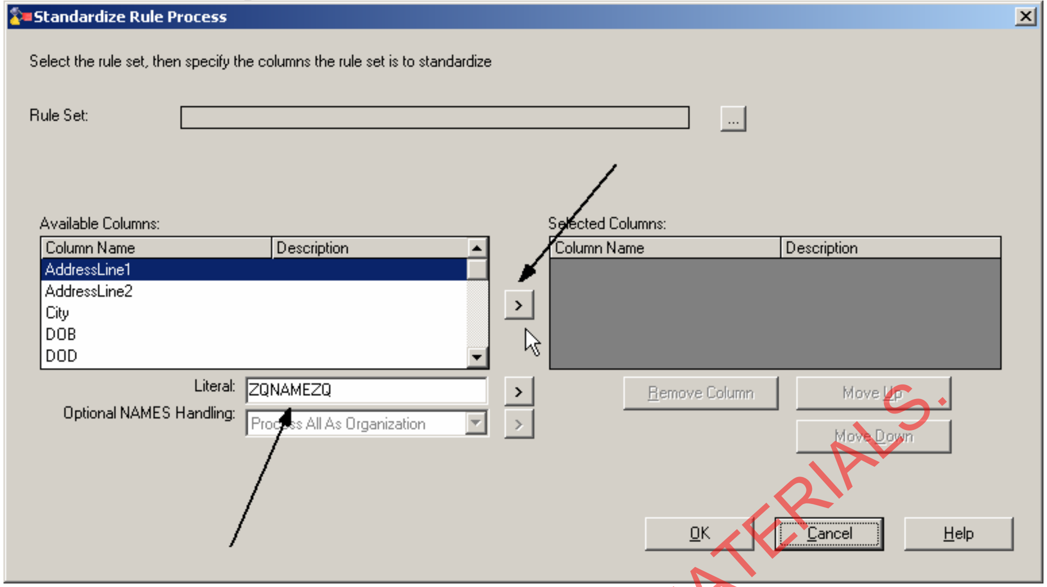
Task: Build and run a Standardize USPREP job

1. Create a new parallel job. Add Sequential File and Standardize stages as well as associated links to produce the following:



2. Save your job and name it StanUSPREP.
3. Edit the USstanCNTRY stage and set the File property to C:\WINNCRM\StanCNTRYUS.dat; this is the file produced by the SelectUS job.
4. Load column definitions from the StanCNTRY meta data.
5. Use the View Data button to verify file connectivity.
6. Close the USstanCNTRY stage editor.
7. Open the stanUSPREP Standardize stage.
8. Click the New Process tab.
9. Using the Rule Set browse button (….) select the USPREP rule set from the Specification Rules folder of the repository.

10. Enter ZQNAMEZQ in the Literal text box. Use the > icon to move this literal to the Selected Columns pane.



Standardize Rule Process

Select the rule set, then specify the columns the rule set is to standardize

Rule Set:

Available Columns:

Column Name	Description
AddressLine1	
AddressLine2	
City	
DOB	
DOD	

Selected Columns:

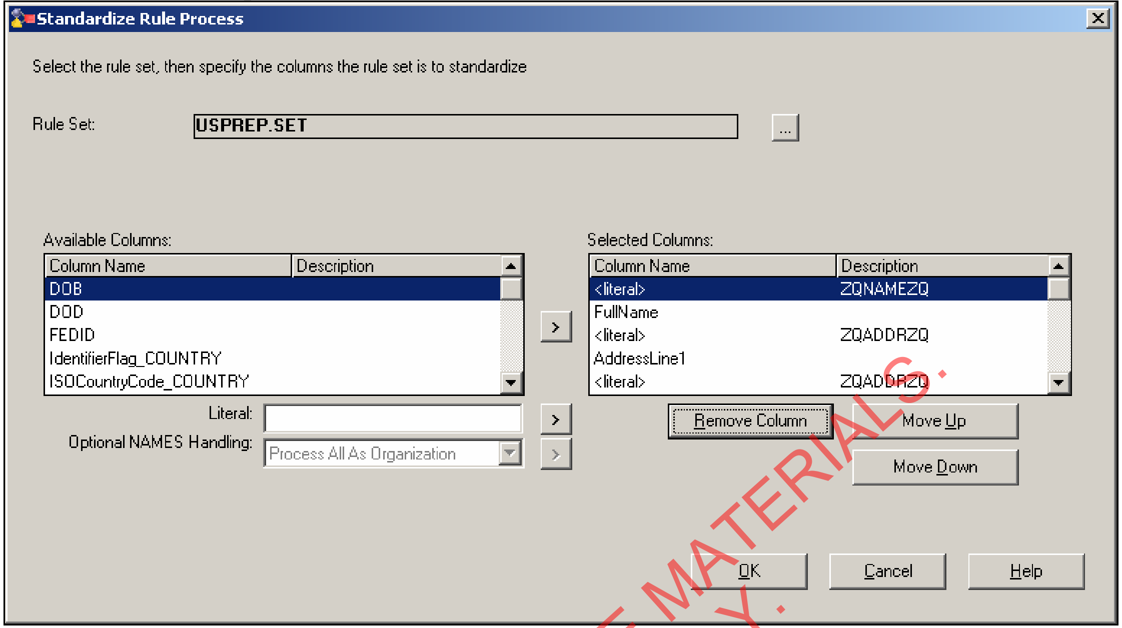
Column Name	Description

Literal: ZQNAMEZQ

Optional NAMES Handling: Process All As Organization

11. Using the appropriate > icon, add the field FullName from the Available Columns pane.
12. Add the Literal ZQADDRZQ.
13. Add the field AddressLine1.
14. Add the Literal ZQADDRZQ.
15. Add the field AddressLine2.
16. Add the Literal ZQAREAZQ.
17. Add the field City.
18. Add the Literal ZQAREAZQ.
19. Add the field State.
20. Add the Literal ZQAREAZQ.
21. Add the field Zip.

22. Your final screen should resemble that shown below (not all columns are visible in the Selected Columns pane).



Select the rule set, then specify the columns the rule set is to standardize

Rule Set: USPREP.SET

Available Columns:

Column Name	Description
DOB	
DOD	
FEDID	
IdentifierFlag_COUNTRY	
ISOCountryCode_COUNTRY	

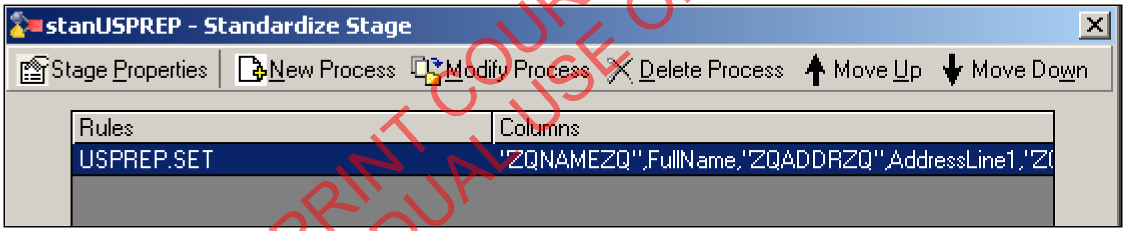
Selected Columns:

Column Name	Description
<literal>	ZQNAMEZQ
FullName	
<literal>	ZQADDRZQ
AddressLine1	
<literal>	ZQADDRZQ

Literal:

Optional NAMES Handling: Process All As Organization

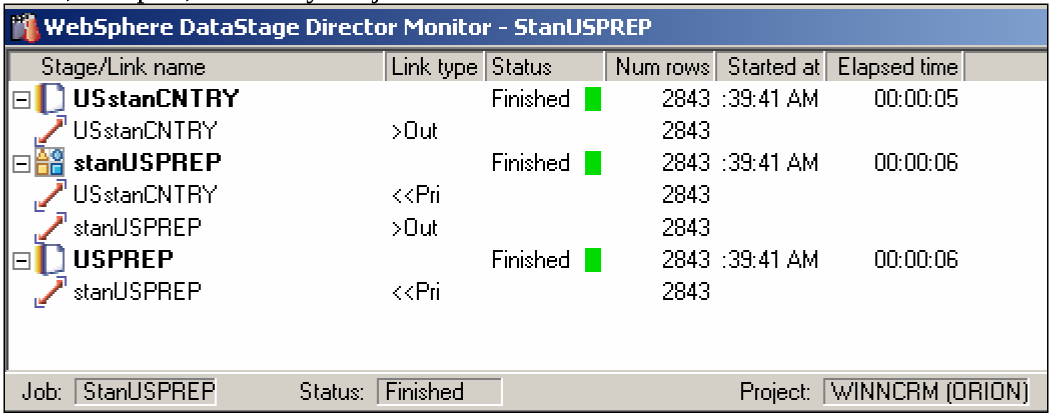
23. Click OK to return to the stage's main screen and review the result. Not all fields are visible.



Rules	Columns
USPREP.SET	'ZQNAMEZQ',FullName,'ZQADDRZQ',AddressLine1,'Z(

24. Click OK to exit the stage editor.
25. Right-click on the Standardize stage and auto-map columns to all output links.
26. Open the target Sequential File stage and set the File property to C:\WINNCRM\USPREP.dat.
27. Click the Format tab and set the Null field value to NULL.
28. The USPREP rule set added columns to the flow and these columns are now part of the target file. Click the Columns tab of the target file stage and save the table definition into the WINNCRM➔TableDefs folder in the repository. Name the table stanUSPREP.
29. Click OK to return to the canvas.

30. Save, compile, and run your job. Job monitor counts :



WebSphere DataStage Director Monitor - StanUSPREP

Stage/Link name	Link type	Status	Num rows	Started at	Elapsed time
⊟ **USstanCNTRY**		Finished 🟩	2843	:39:41 AM	00:00:05
USstanCNTRY	>Out		2843		
⊟ **stanUSPREP**		Finished 🟩	2843	:39:41 AM	00:00:06
USstanCNTRY	<<Pri		2843		
stanUSPREP	>Out		2843		
⊟ **USPREP**		Finished 🟩	2843	:39:41 AM	00:00:06
stanUSPREP	<<Pri		2843		

Job: StanUSPREP Status: Finished Project: WINNCRM (ORION)

31. Use the View Data button on the USPREP Sequential File stage to review the results of the job run.

Lab 13: Standardize USNAME USADDR USAREA

Objectives:

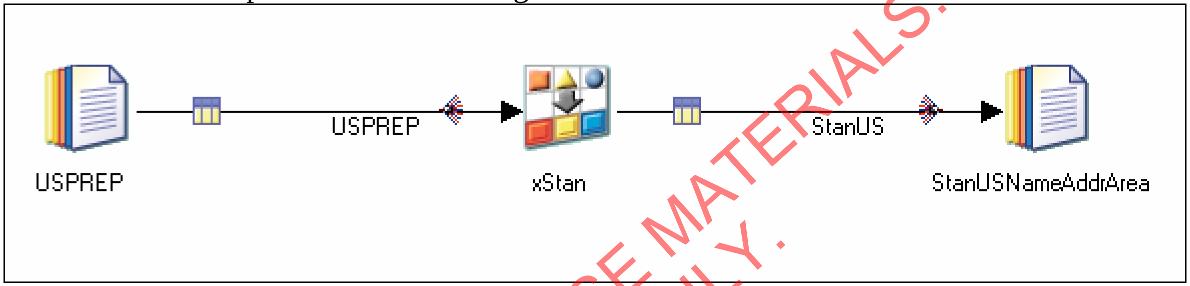
- Standardize the Name, Address, and Area fields

Assumptions:

- Job USPREP ran successfully

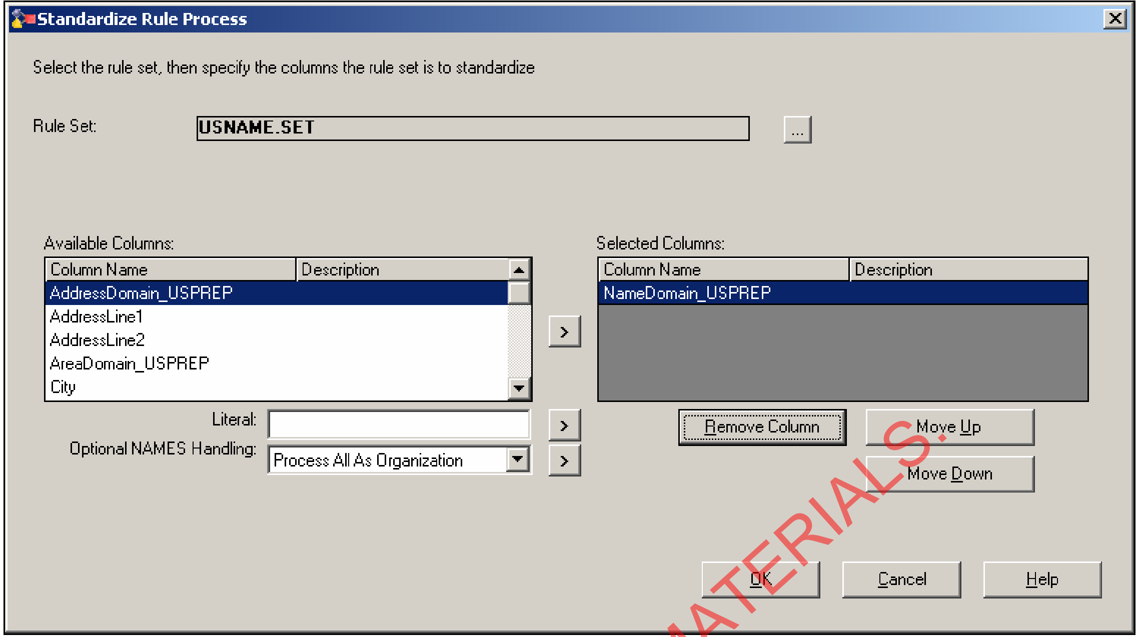
Task: Build and run a job to standardize Name, Address, and Area

1. Create a new parallel job. Add Sequential File and Standardize stages as well as associated links to produce the following:



2. Save your job and name it StanUSNameAddressArea.
3. Open the USPREP stage and set the File property to C:\WINNCRM\USPREP.dat – this file was created by the USPREP job.
4. Click the Columns tab and load the table definition named stanUSPREP.
5. Click the View Data button to validate the meta data and verify connectivity to the file.
6. Open the xStan Standardize stage.
7. Click the New Process tab.

8. Select the USNAME rule set and the NameDomain_USPREP field to yield:



Select the rule set, then specify the columns the rule set is to standardize

Rule Set: USNAME.SET

Available Columns:

Column Name	Description
AddressDomain_USPREP	
AddressLine1	
AddressLine2	
AreaDomain_USPREP	
City	

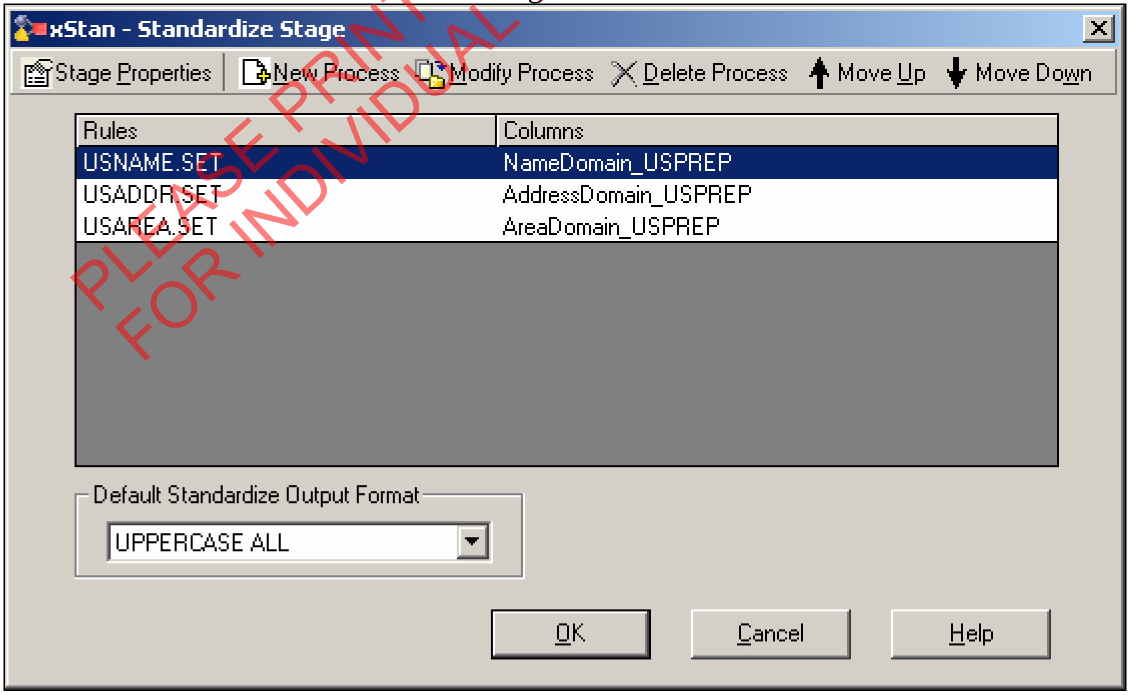
Selected Columns:

Column Name	Description
NameDomain_USPREP	

Literal:

Optional NAMES Handling: Process All As Organization

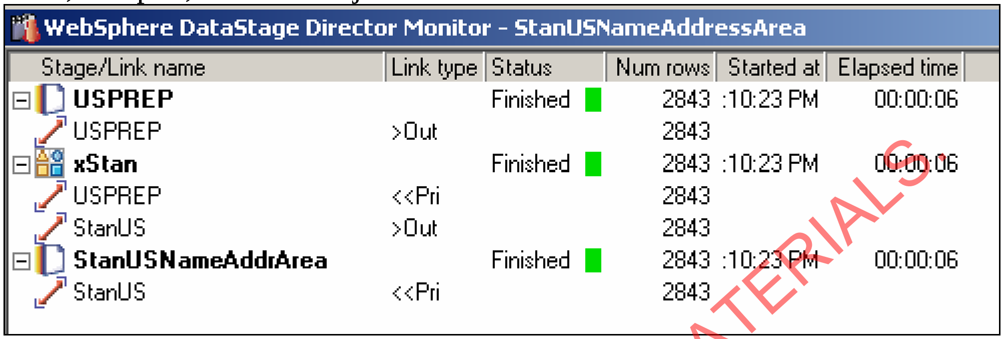
9. Click OK to return to the Standardize stage main screen.
10. Click New Process again.
11. Select the USADDR rule set and the AddressDomain_USPREP field.
12. Click OK to return to the Standardize stage main screen.
13. Click New Process again.
14. Select the USAREA rule set and the AreaDomain_USPREP field.
15. Click OK to return to the Standardize stage main screen.



Rules	Columns
USNAME.SET	NameDomain_USPREP
USADDR.SET	AddressDomain_USPREP
USAREA.SET	AreaDomain_USPREP

Default Standardize Output Format: UPPERCASE ALL

16. Click OK to save settings and return to the canvas.
17. Open the StanUSNameAddrArea target Sequential file stage and set the File property to C:\WINNCRM\StanUSNameAddrArea.dat
18. Click the Format tab and set the Null field value property.
19. Click the Columns tab and then click the Save button.
20. Save the table definition as StanUSNameAddrArea.
21. Click OK as necessary to return to the canvas.
22. Save, compile, and run the job.



WebSphere DataStage Director Monitor - StanUSNameAddressArea

Stage/Link name	Link type	Status	Num rows	Started at	Elapsed time
USPREP		Finished 🟩	2843	:10:23 PM	00:00:06
USPREP	>Out		2843		
xStan		Finished 🟩	2843	:10:23 PM	00:00:06
USPREP	<<Pri		2843		
StanUS	>Out		2843		
StanUSNameAddrArea		Finished 🟩	2843	:10:23 PM	00:00:06
StanUS	<<Pri		2843		

23. Use the View Data button to view the contents of the target file.

Lab 14: Investigate unhandled name patterns

Objectives:

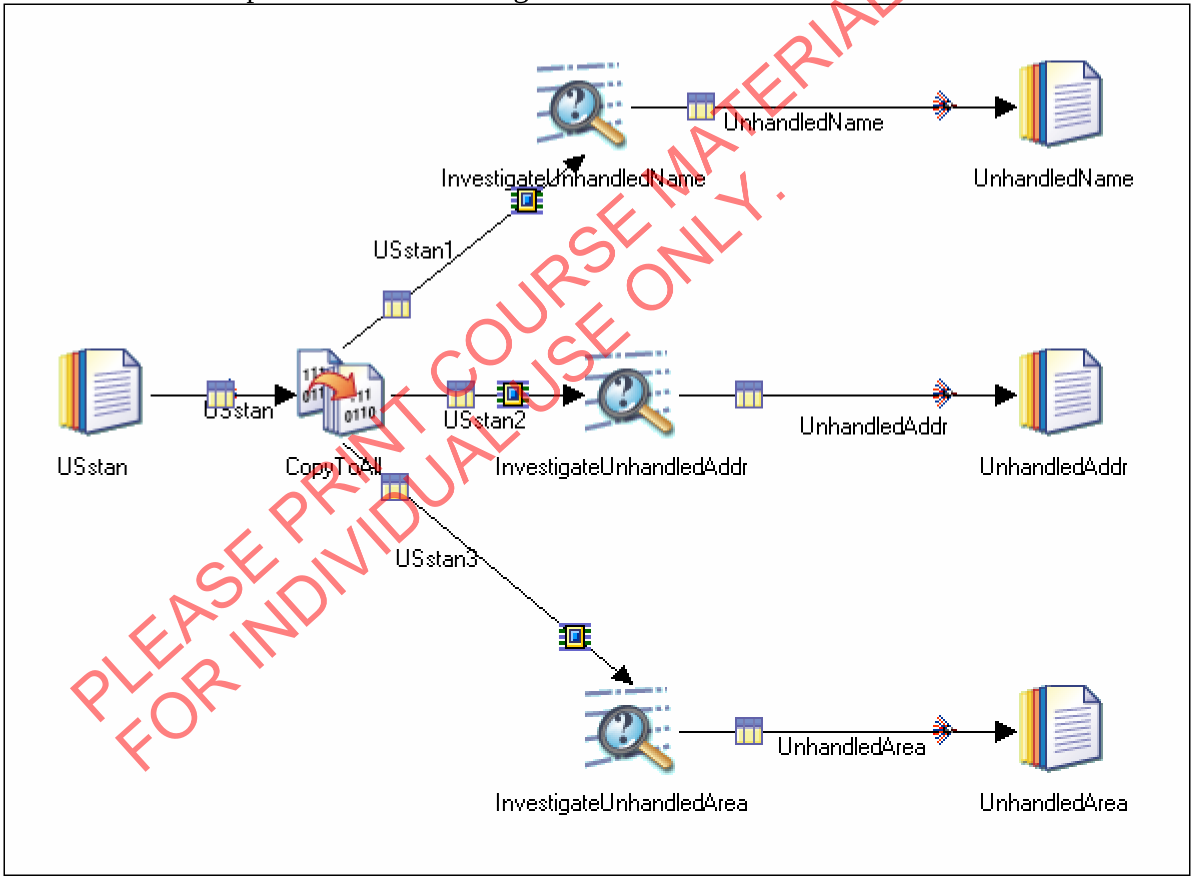
- Use the Investigate Stage to review the results of the USNAME, USADDR, and USAREA standardization.
- Demonstrate use of multiple investigations in one job

Assumptions:

- Job StanUSNameAddressArea exists and executed successfully

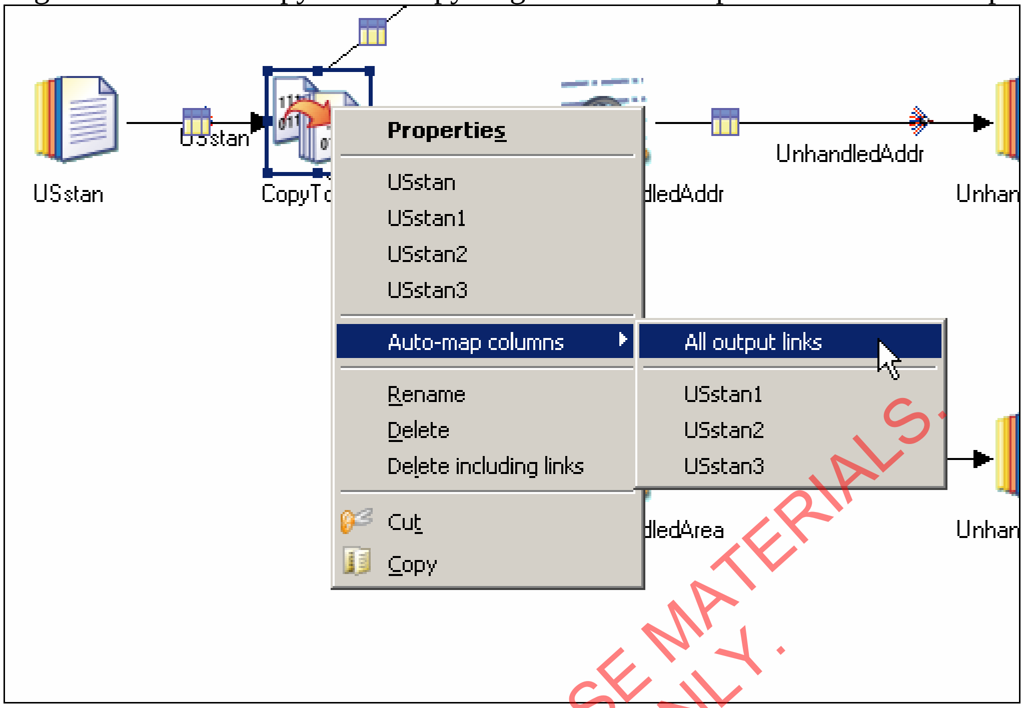
Task: Build and run an Investigation

1. Create a new parallel job. Add Sequential File, Copy, and Investigate stages as well as associated links to produce the following:



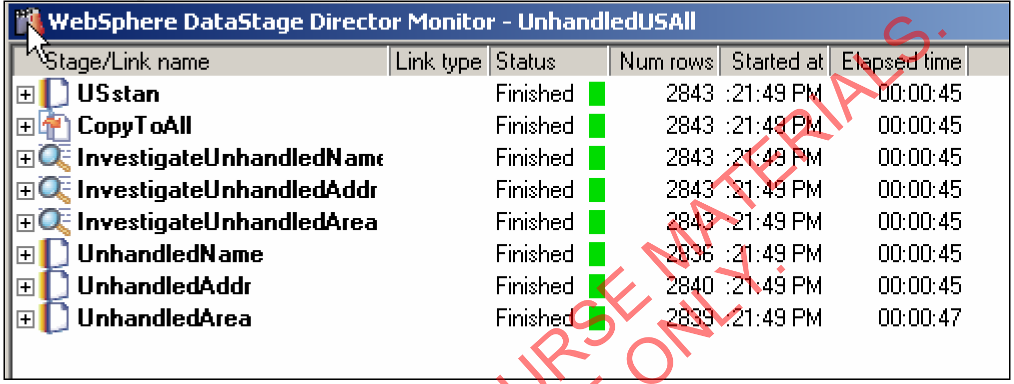
2. Save your job and name it UnhandledUSAll.
3. Open the USstan source stage and set the File property to C:\WINNCRM\StanUSNameAddrArea.dat.
4. Load columns from the StanUSNameAddrArea table you saved in an earlier exercise.
5. Return to the canvas.

6. Right-click on the CopyToAll Copy stage and Auto-map columns ➔ All output links



7. Open the top Investigate stage labeled InvestigateUnhandledName.
8. Click the Character Concatenate Investigate tab.
9. One at a time select columns UnhandledPattern_USNAME, UnhandledData_USNAME, InputPattern_USNAME, and NameDomain_USPREP. Set the mask for UnhandledPattern_USNAME to All C. For all other columns use the All X mask.
10. Close the stage editor.
11. Right-click on the InvestigateUnhandledName Investigate stage and Auto-map columns➔All output links.
12. Open the top target Sequential File stage labeled UnhandledName and set the File property to C:\WINNCRM\USstanUnhandledName.dat.
13. Click the Format tab to set the Null field value property.
14. Close the stage editor.
15. In a similar fashion open the middle Investigate stage labeled InvestigateUnhandledAddr.
16. One at a time select columns UnhandledPattern_USADDR, UnhandledData_USADDR, InputPattern_USADDR, and AddressDomain_USPREP. Set the mask for UnhandledPattern_USADDR to All C. For all other columns use the All X mask.
17. Close the stage editor.
18. Right-click on the InvestigateUnhandledAddr Investigate stage and Auto-map columns➔All output links.
19. Open the middle target Sequential File stage labeled UnhandledAddr and set the File property to C:\WINNCRM\USstanUnhandledAddr.dat.
20. Click the Format tab to set the Null field value property.
21. Close the stage.

22. Open the bottom Investigate stage labeled InvestigateUnhandledArea.
23. One at a time select columns UnhandledPattern_USAREA, UnhandledData_USAREA, InputPattern_USAREA, and AreaDomain_USPREP. Set the mask for UnhandledPattern_USAREA to All C. For all other columns use the All X mask.
24. Close the stage editor.
25. Right-click on the InvestigateUnhandledArea Investigate stage and Auto-map columns➔All output links.
26. Open the bottom target Sequential File stage labeled UnhandledArea and set the File property to C:\WINNCRM\USstanUnhandledArea.dat.
27. Click the Format tab to set the Null field value property.
28. Close the stage.
29. Save, compile and run the job.



WebSphere DataStage Director Monitor - UnhandledUSAll

Stage/Link name	Link type	Status	Num rows	Started at	Elapsed time
USstan		Finished 🟩	2843	:21:49 PM	00:00:45
CopyToAll		Finished 🟩	2843	:21:49 PM	00:00:45
InvestigateUnhandledName		Finished 🟩	2843	:21:49 PM	00:00:45
InvestigateUnhandledAddr		Finished 🟩	2843	:21:49 PM	00:00:45
InvestigateUnhandledArea		Finished 🟩	2843	:21:49 PM	00:00:45
UnhandledName		Finished 🟩	2836	:21:49 PM	00:00:45
UnhandledAddr		Finished 🟩	2840	:21:49 PM	00:00:45
UnhandledArea		Finished 🟩	2839	:21:49 PM	00:00:47

30. Review the sequential file results by using the View Data option on each of the target files.

Lab 15: Apply rule set overrides

Objectives:

- Improve the standardization for USNAME

Assumptions:

- Job UnhandledUSALL ran successfully
- Unhandled results analyzed

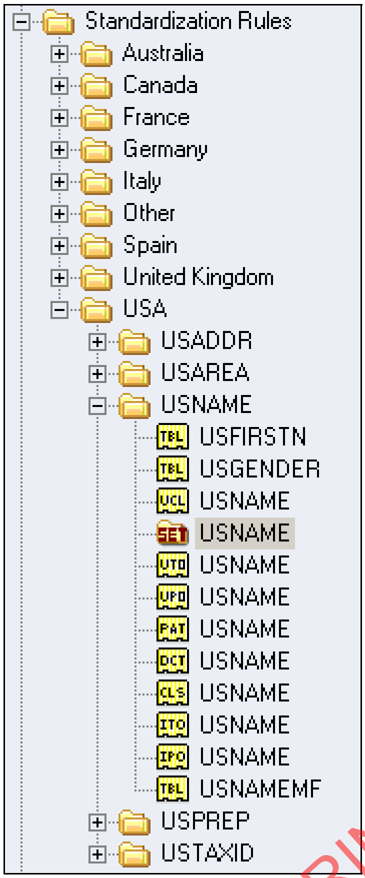
Task: Data Review discussion

Data review for USNAME standardization:
The pattern +FI represents unclassified last name (standard rule set approach), a first name followed by a middle initial.
The data appears to have been classified correctly so a classification override will not process this data correctly.
Notice that the input pattern is identical to the unhandled pattern. This is an indication that the pattern did not match any pattern in the pattern action file and an Input Pattern Override would cause that pattern to be processed.

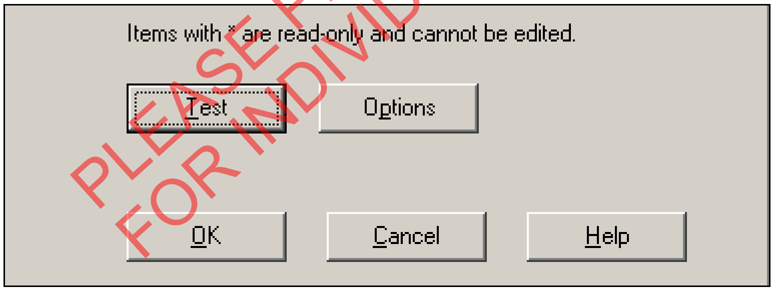
Unhandled Pattern	Unhandled Data	Input Pattern	Input Name Text
+FI	DAMORA WILLIAM H	+FI	DAMORA WILLIAM H
+FI	PEPE NANCY J	+FI	PEPE NANCY J
+FI	KOPPLIN ELDEN E	+FI	KOPPLIN ELDEN E
+FI	KRATOCHWILL TOMAS R	+FI	KRATOCHWILL TOMAS R
+FI	LEININGER SALLY P	+FI	LEININGER SALLY P

Task: Apply input pattern override

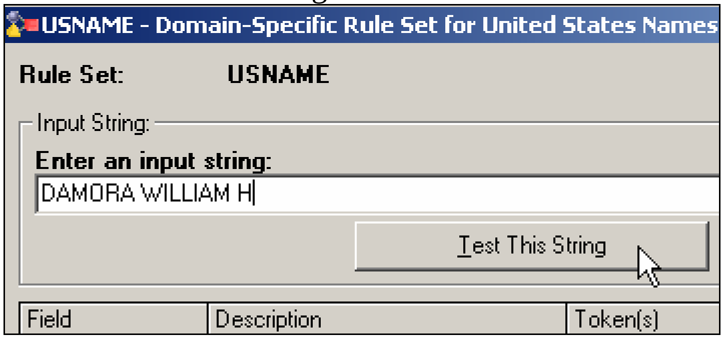
1. Open the Standardization Rules ➔ USA folder of the repository and double-click the USNAME rule set.



2. Click the Test button near the bottom of the screen.


Items with * are read-only and cannot be edited.

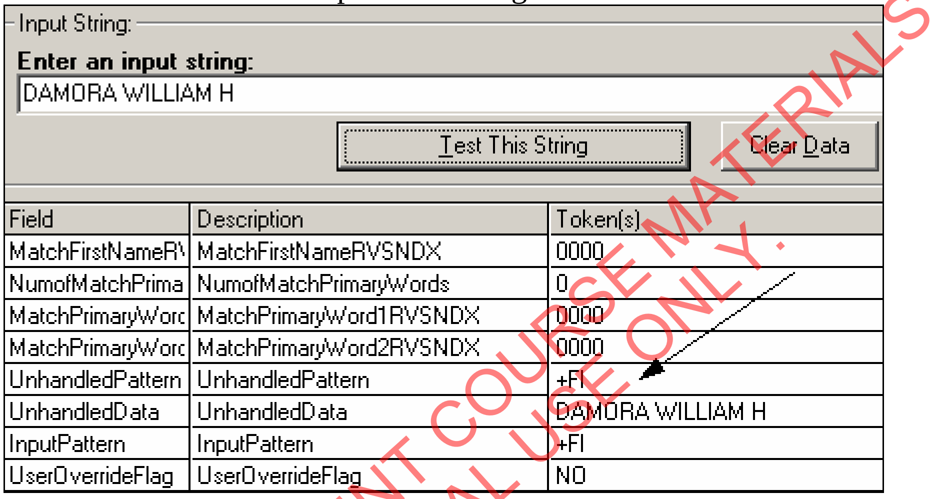
3. Enter the string DAMORA WILLIAM H into the ‘Enter an input string’ text box and click the Test This String button.



Rule Set: USNAME

Enter an input string: DAMORA WILLIAM H

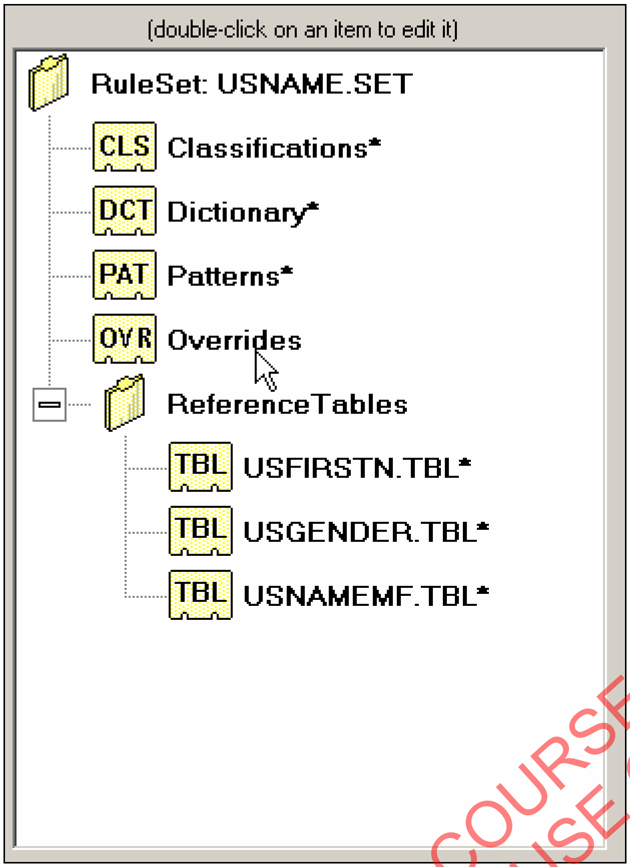
4. Note that the rule set interprets the string as +FI.



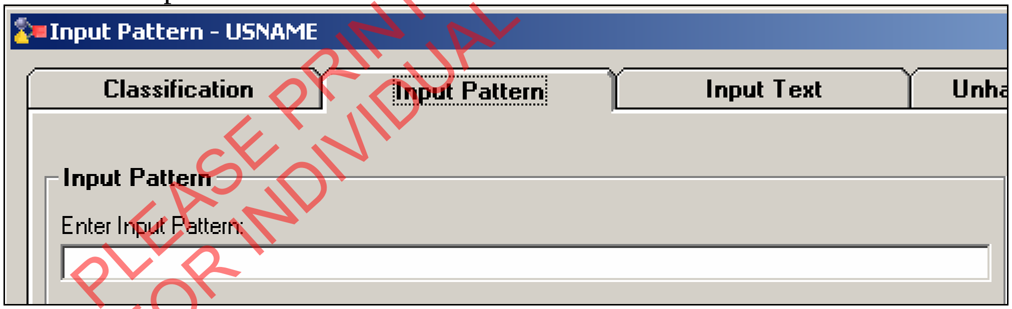
Enter an input string: DAMORA WILLIAM H

Field	Description	Token(s)
MatchFirstNameR\	MatchFirstNameRVSNDX	0000
NumofMatchPrima	NumofMatchPrimaryWords	0
MatchPrimaryWorc	MatchPrimaryWord1RVSNDX	0000
MatchPrimaryWorc	MatchPrimaryWord2RVSNDX	0000
UnhandledPattern	UnhandledPattern	+FI
UnhandledData	UnhandledData	DAMORA WILLIAM H
InputPattern	InputPattern	+FI
UserOverrideFlag	UserOverrideFlag	NO

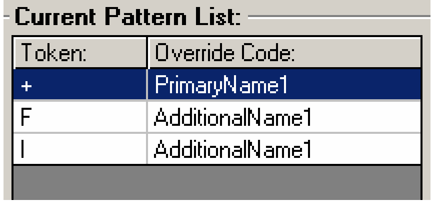
5. Close the test screen and double-click on the Overrides icon.



6. Select the Input Pattern tab.



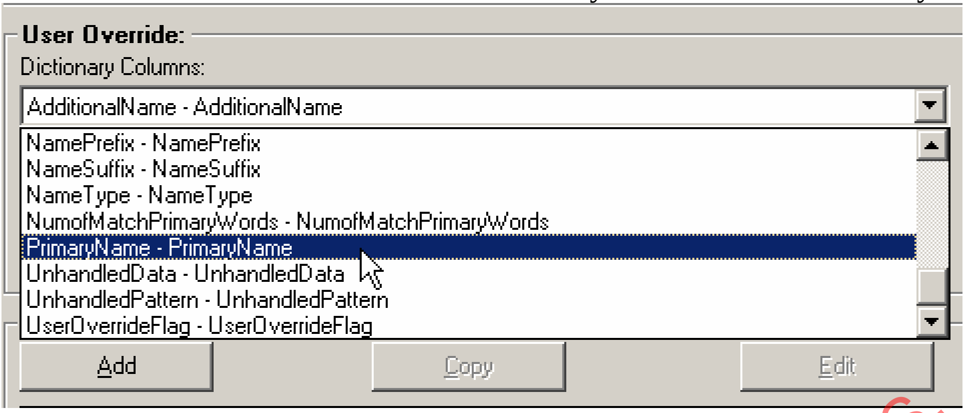
7. Enter the Input Pattern +FI
8. From the Current Pattern List select the first entry, +.



Current Pattern List:

Token:	Override Code:
+	PrimaryName1
F	AdditionalName1
I	AdditionalName1

9. From the User Override down-click Dictionary Columns to find PrimaryName.



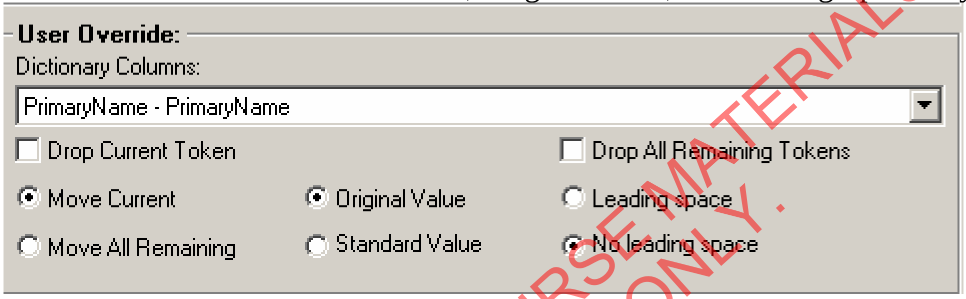
User Override:

Dictionary Columns:

AdditionalName - AdditionalName

- NamePrefix - NamePrefix
- NameSuffix - NameSuffix
- NameType - NameType
- NumofMatchPrimaryWords - NumofMatchPrimaryWords
- PrimaryName - PrimaryName
- UnhandledData - UnhandledData
- UnhandledPattern - UnhandledPattern
- UserOverrideFlag - UserOverrideFlag

10. Click the checkboxes Move Current, Original value, No Leading Space to yield



User Override:

Dictionary Columns:

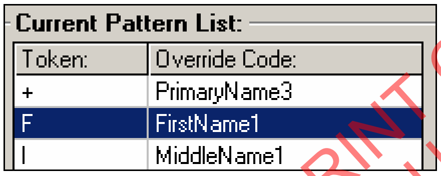
PrimaryName - PrimaryName

□ Drop Current Token □ Drop All Remaining Tokens

● Move Current ● Original Value ○ Leading space

○ Move All Remaining ○ Standard Value ● No leading space

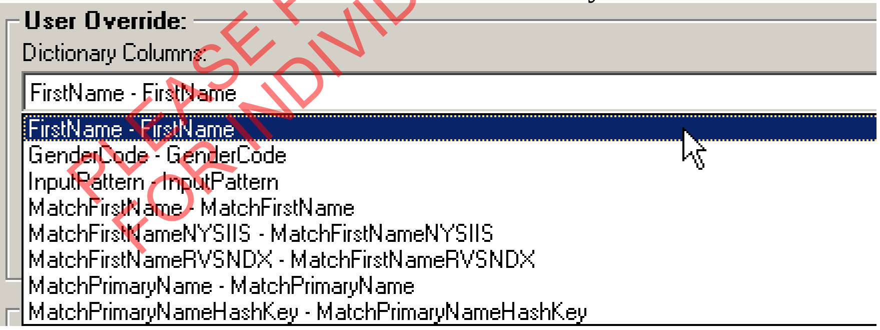
11. From the Current Pattern List select the second entry, F.



Current Pattern List:

Token:	Override Code:
+	PrimaryName3
F	FirstName1
I	MiddleName1

12. From the User Override down-click Dictionary Columns to find FirstName.



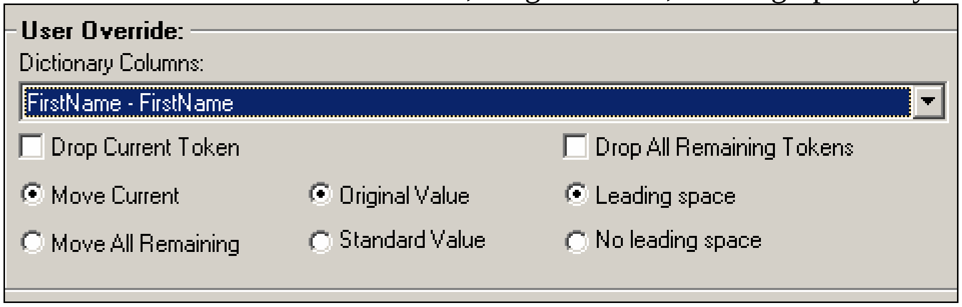
User Override:

Dictionary Columns:

FirstName - FirstName

- FirstName - FirstName
- GenderCode - GenderCode
- InputPattern - InputPattern
- MatchFirstName - MatchFirstName
- MatchFirstNameNYSIIS - MatchFirstNameNYSIIS
- MatchFirstNameRVSNDX - MatchFirstNameRVSNDX
- MatchPrimaryName - MatchPrimaryName
- MatchPrimaryNameHashKey - MatchPrimaryNameHashKey

13. Click the checkboxes Move Current, Original value, Leading Space to yield



User Override:

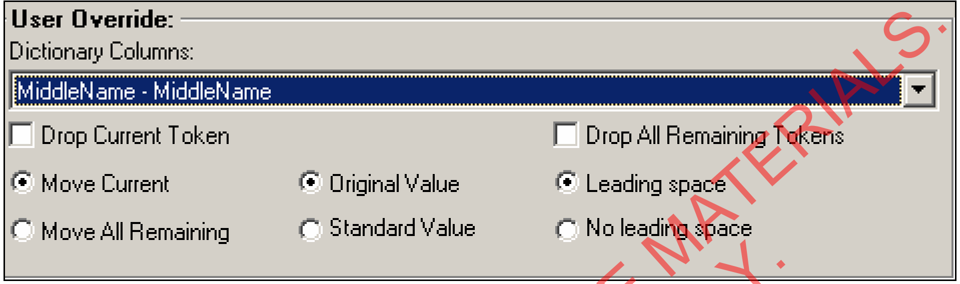
Dictionary Columns: FirstName - FirstName

□ Drop Current Token □ Drop All Remaining Tokens

● Move Current ● Original Value ● Leading space

○ Move All Remaining ○ Standard Value ○ No leading space

14. Repeat this process for the last portion of the pattern – the I portion. Set the Dictionary Column to MiddleName. Set checkboxes Move Current, Original value, and Leading Space to yield:



User Override:

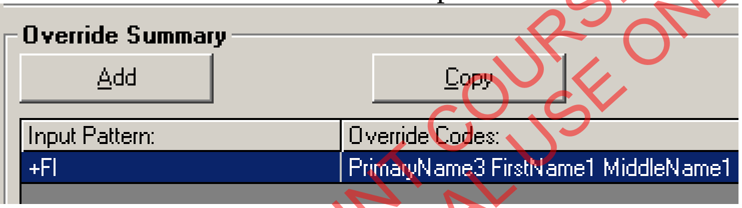
Dictionary Columns: MiddleName - MiddleName

□ Drop Current Token □ Drop All Remaining Tokens

● Move Current ● Original Value ● Leading space

○ Move All Remaining ○ Standard Value ○ No leading space

15. Click the Add button to enter this pattern override.

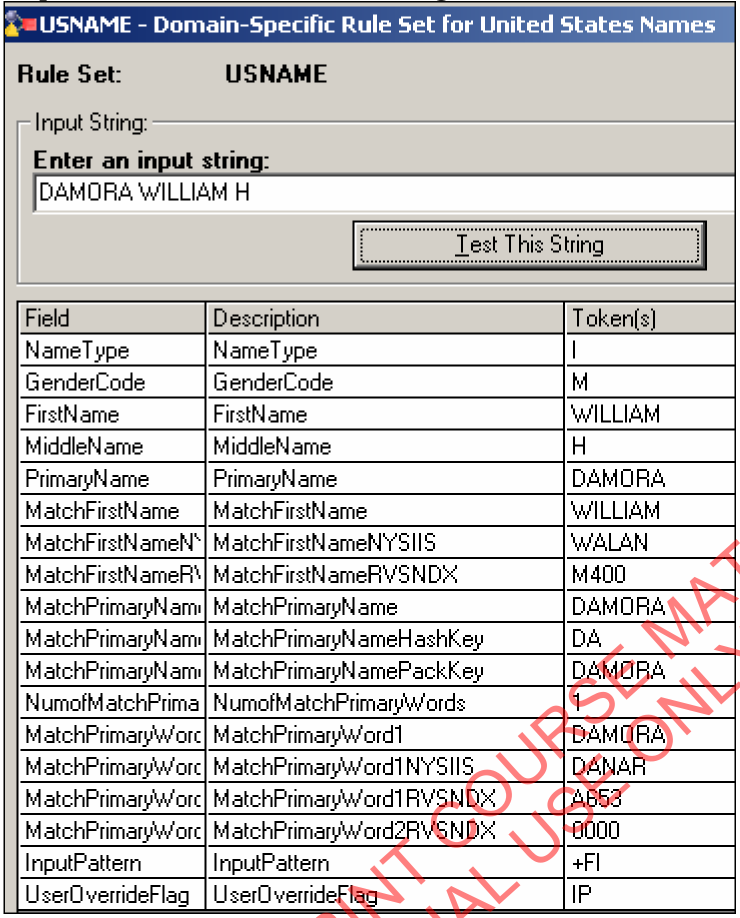


Override Summary

Input Pattern:	Override Codes:
+FI	PrimaryName3 FirstName1 MiddleName1

16. Click the OK button to exit the Override screen.

17. Repeat the earlier test on the string DAMORA WILLIAM H. to validate:



USNAME - Domain-Specific Rule Set for United States Names

Rule Set: USNAME

Input String:

Enter an input string:

DAMORA WILLIAM H

Field	Description	Token(s)
NameType	NameType	I
GenderCode	GenderCode	M
FirstName	FirstName	WILLIAM
MiddleName	MiddleName	H
PrimaryName	PrimaryName	DAMORA
MatchFirstName	MatchFirstName	WILLIAM
MatchFirstNameN	MatchFirstNameNYSIIS	WALAN
MatchFirstNameR	MatchFirstNameRVSNDX	M400
MatchPrimaryNam	MatchPrimaryName	DAMORA
MatchPrimaryNam	MatchPrimaryNameHashKey	DA
MatchPrimaryNam	MatchPrimaryNamePackKey	DAMORA
NumofMatchPrima	NumofMatchPrimaryWords	1
MatchPrimaryWorc	MatchPrimaryWord1	DAMORA
MatchPrimaryWorc	MatchPrimaryWord1NYSIIS	DANAR
MatchPrimaryWorc	MatchPrimaryWord1RVSNDX	A653
MatchPrimaryWorc	MatchPrimaryWord2RVSNDX	0000
InputPattern	InputPattern	+FI
UserOverrideFlag	UserOverrideFlag	IP

Task: Data Review discussion

UNHANDLED PATTERN +,+

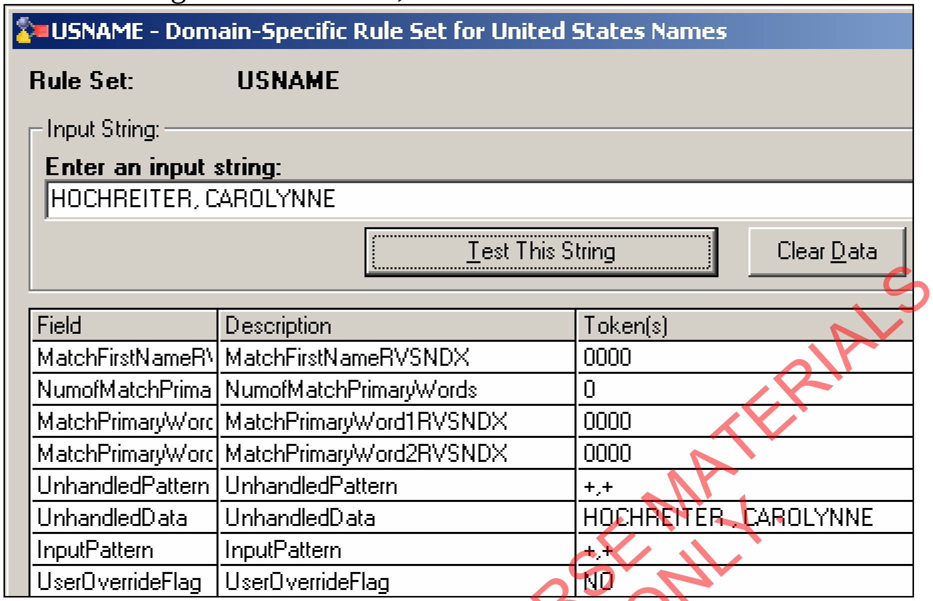
- The pattern +,+ represents unclassified last name (standard rule set approach), a comma, and an unclassified first name.
- The first name values appear to not be classified as first names. One approach would be to review the first names for addition to the classification table.
- You may want to check the frequency of the data value, the more often it occurs the more likely we are to classify that word.

Unhandled Pattern	Unhandled Data	Input Pattern	Input Name Text
+,+	HOCHREITER, CAROLYNNE	+,+	HOCHREITER, CAROLYNNE
+,+	HAYWARD, WINSLOW	+,+	HAYWARD, WINSLOW
+,+	ESHAGHIAN, JOUBIN	+,+	ESHAGHIAN, JOUBIN
+,+	SODIA, MARVYN	+,+	SODIA, MARVYN

+,+ SAKURAZAWA, HARUKO +,+ SAKURAZAWA, HARUKO

Task: Apply classification override

1. Edit the USNAME rule set.
2. Test the string HOCHREITER, CAROLYNNE



USNAME - Domain-Specific Rule Set for United States Names

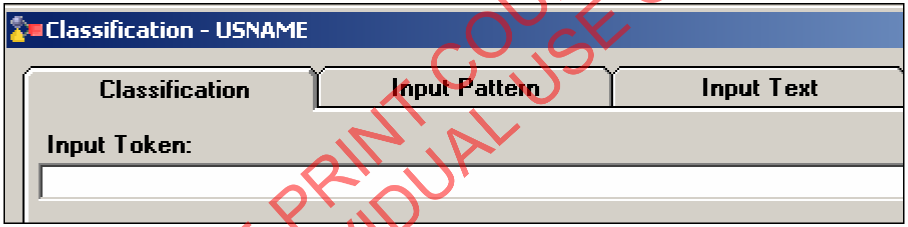
Rule Set: USNAME

Input String:

Enter an input string: HOCHREITER, CAROLYNNE

Field	Description	Token(s)
MatchFirstNameR\	MatchFirstNameRVSNDX	0000
NumofMatchPrima	NumofMatchPrimaryWords	0
MatchPrimaryWorc	MatchPrimaryWord1RVSNDX	0000
MatchPrimaryWorc	MatchPrimaryWord2RVSNDX	0000
UnhandledPattern	UnhandledPattern	+,+
UnhandledData	UnhandledData	HOCHREITER , CAROLYNNE
InputPattern	InputPattern	+,+
UserOverrideFlag	UserOverrideFlag	NO

3. Double-click the Overrides button and select the Classification tab – this is the default.

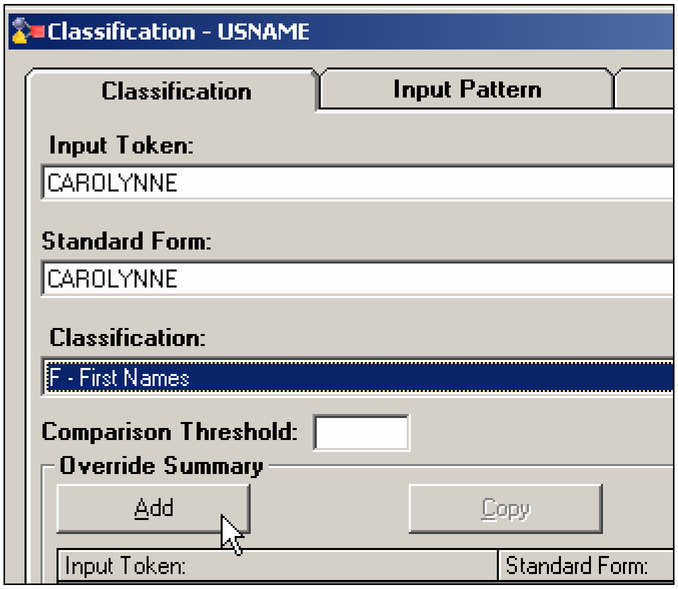


Classification - USNAME

Input Token:

4. Enter the Input Token CAROLYNNE and Standard Form CAROLYNNE.
5. From the classification drop down menu choose, F-First name.

6. Click the Add button.



Classification - USNAME

Input Token: CAROLYNNE

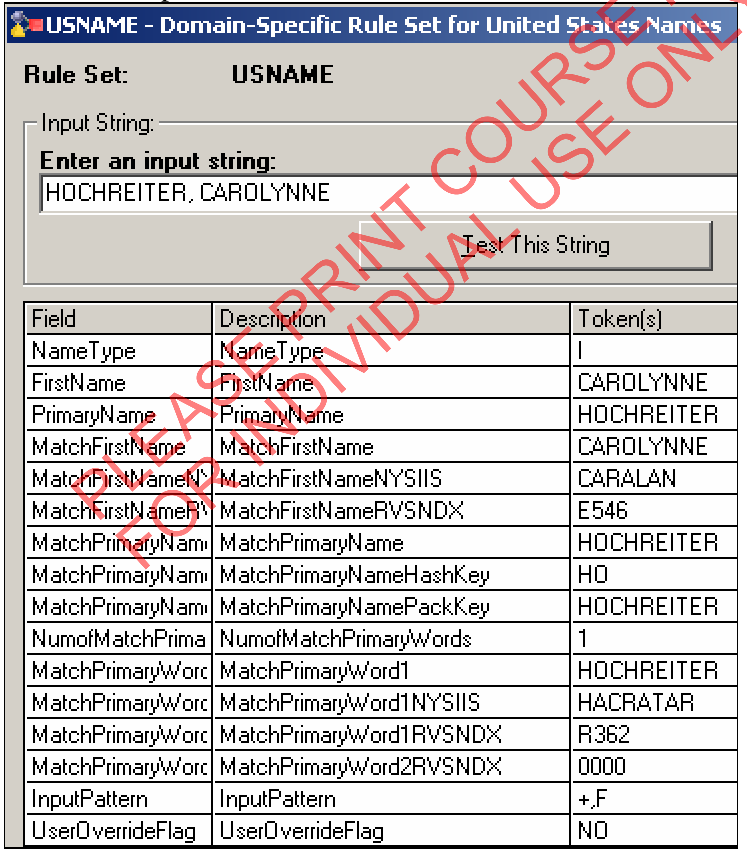
Standard Form: CAROLYNNE

Classification: F - First Names

Comparison Threshold:

Override Summary

7. Repeat this process for WINSLOW, JOUBIN, MARVYN, and HARUKO.
8. Use the test process to validate results.



USNAME - Domain-Specific Rule Set for United States Names

Rule Set: USNAME

Input String:

Enter an input string: HOCHREITER, CAROLYNNE

Field	Description	Token(s)
NameType	NameType	I
FirstName	FirstName	CAROLYNNE
PrimaryName	PrimaryName	HOCHREITER
MatchFirstName	MatchFirstName	CAROLYNNE
MatchFirstNameN'	MatchFirstNameNYSIIS	CARALAN
MatchFirstNameR'	MatchFirstNameRVSNDX	E546
MatchPrimaryNam	MatchPrimaryName	HOCHREITER
MatchPrimaryNam	MatchPrimaryNameHashKey	HO
MatchPrimaryNam	MatchPrimaryNamePackKey	HOCHREITER
NumofMatchPrima	NumofMatchPrimaryWords	1
MatchPrimaryWorc	MatchPrimaryWord1	HOCHREITER
MatchPrimaryWorc	MatchPrimaryWord1NYSIIS	HACRATAR
MatchPrimaryWorc	MatchPrimaryWord1RVSNDX	R362
MatchPrimaryWorc	MatchPrimaryWord2RVSNDX	0000
InputPattern	InputPattern	+,F
UserOverrideFlag	UserOverrideFlag	NO

Task: Data Review discussion

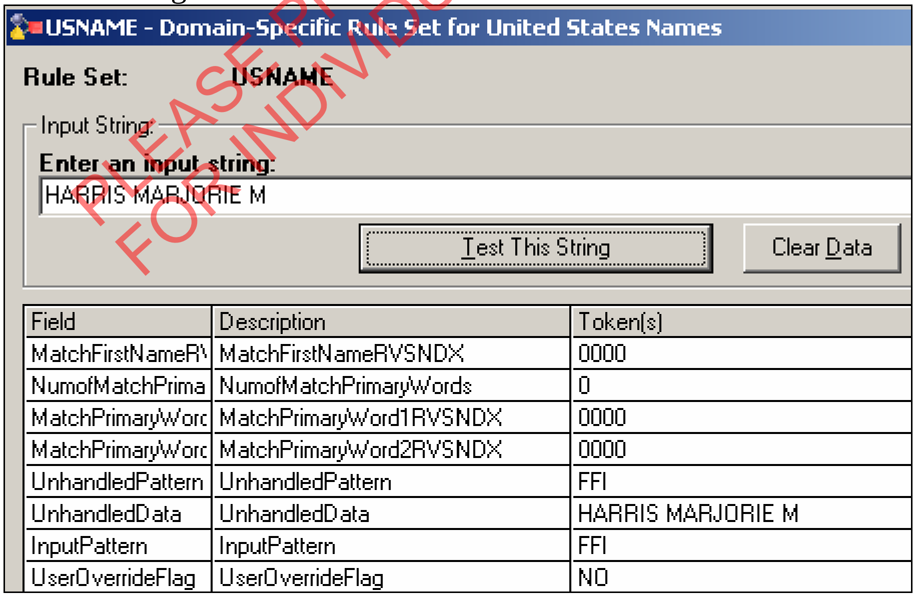
UNHANDLED PATTERN FFI

- The pattern **FFI** represents a last name that was recognized as a first name, a classified first name and an initial. Notice this data does not include a comma providing context to the first and last name tokens.
- In this case the input pattern for all the sample records is not the same as the unhandled pattern. There are 2 distinct input patterns and one distinct unhandled pattern.
- Applying the override to the unhandled pattern will allow us to add one override. If we had chosen the input pattern override then we would need to add an override for each pattern.
- This is an indication that the pattern did not match any patterns in the pattern action file and an **UNHANDLED PATTERN OVERRIDE** would cause this pattern to be processed.

Unhandled Pattern	Unhandled Data	Input Pattern	Input Name Text
FFI	**HARRIS MARJORIE M**	**+FI.**	**HARRIS MARJORIE M.**
FFI	ROSS JOSEPH P	+FI	ROSS JOSEPH P
FFI	**YOUNG THERESA C**	**+FI.**	**YOUNG THERESA C.**
FFI	OLIVA LAWRENCE M	+FI	OLIVA LAWRENCE M
FFI	LANG LEE B	+FI	LANG LEE B

Task: Apply input pattern override

1. Edit the USNAME rule set.
2. Test the string HARRIS MARJORIE M



USNAME - Domain-Specific Rule Set for United States Names

Rule Set: USNAME

Input String:

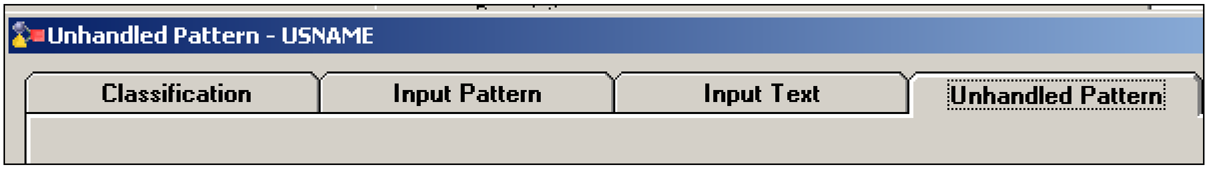
Enter an input string:

HARRIS MARJORIE M

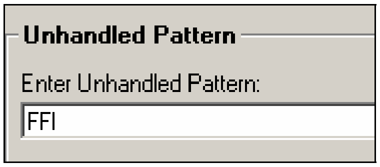
Field	Description	Token(s)
MatchFirstNameR\	MatchFirstNameRVSNDX	0000
NumofMatchPrima	NumofMatchPrimaryWords	0
MatchPrimaryWorc	MatchPrimaryWord1RVSNDX	0000
MatchPrimaryWorc	MatchPrimaryWord2RVSNDX	0000
UnhandledPattern	UnhandledPattern	FFI
UnhandledData	UnhandledData	HARRIS MARJORIE M
InputPattern	InputPattern	FFI
UserOverrideFlag	UserOverrideFlag	NO

3. Exit the test area and open the Overrides.

4. Click the Unhandled Pattern tab.

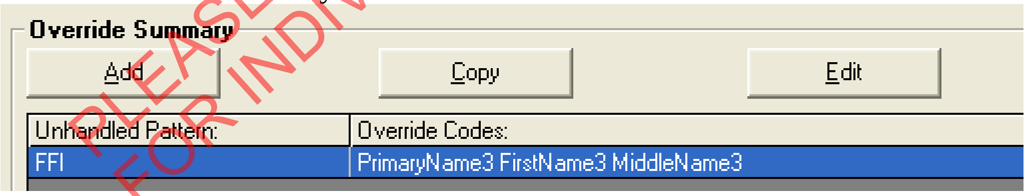


5. Enter Unhandled Pattern



Enter Unhandled Pattern: FFI

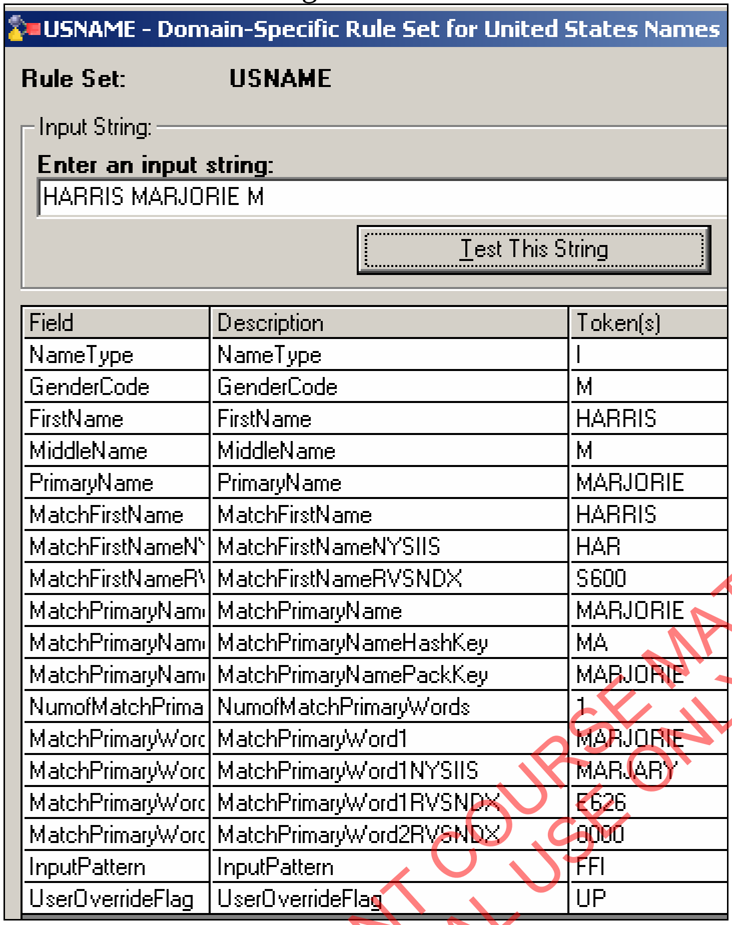
6. From the Current Pattern List select the first entry, F
7. From the User Override Options choose:
   a. Dictionary Fields: PrimaryName
   b. Move Current
   c. Original Value
   d. No Leading Space
8. Repeat the process for the remaining tokens using the following settings
   a. F token
   b. Dictionary Fields: FirstName
   c. Move Current
   d. Original Value
   e. No Leading Space
9. I token
   a. Dictionary Fields: MiddleName
   b. Move Current
   c. Original Value
   d. No Leading Space
10. Under Override Summary click Add.



Unhandled Pattern:	Override Codes:
FFI	PrimaryName3 FirstName3 MiddleName3

11. Click Apply, then OK

12. Test the Overrides using the Test area.



Rule Set: USNAME

Enter an input string: HARRIS MARJORIE M

Field	Description	Token(s)
NameType	NameType	I
GenderCode	GenderCode	M
FirstName	FirstName	HARRIS
MiddleName	MiddleName	M
PrimaryName	PrimaryName	MARJORIE
MatchFirstName	MatchFirstName	HARRIS
MatchFirstNameN	MatchFirstNameNYSIIS	HAR
MatchFirstNameR	MatchFirstNameRVSNDX	S600
MatchPrimaryNam	MatchPrimaryName	MARJORIE
MatchPrimaryNam	MatchPrimaryNameHashKey	MA
MatchPrimaryNam	MatchPrimaryNamePackKey	MARJORIE
NumofMatchPrima	NumofMatchPrimaryWords	1
MatchPrimaryWorc	MatchPrimaryWord1	MARJORIE
MatchPrimaryWorc	MatchPrimaryWord1NYSIIS	MARJARY
MatchPrimaryWorc	MatchPrimaryWord1RVSNDX	E626
MatchPrimaryWorc	MatchPrimaryWord2RVSNDX	0000
InputPattern	InputPattern	FFI
UserOverrideFlag	UserOverrideFlag	UP

Lab 16: Add fields using the Transformer stage

Objectives:

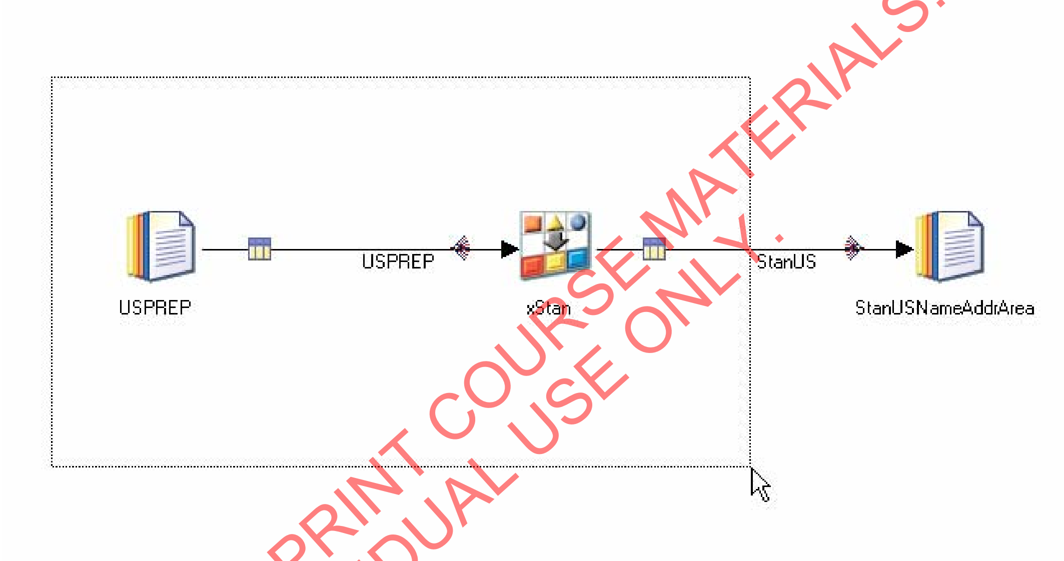
- Add fields needed for match processing

Assumptions:

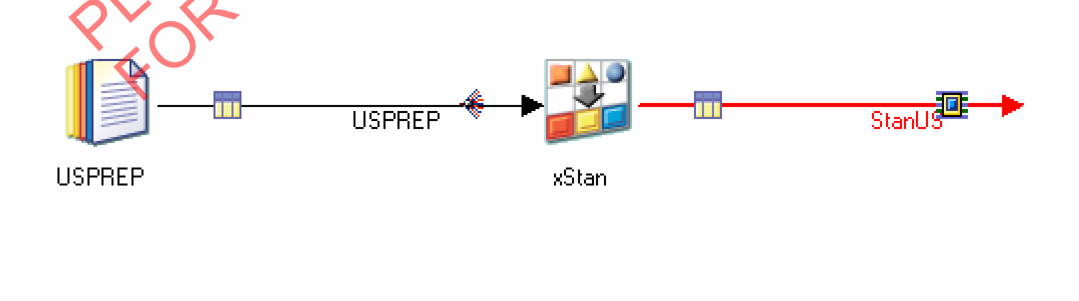
- Job StanUSNameAddressArea exists

Task: Build job to add fields

1. Edit the StanUSNameAddressArea job.
2. Using the Windows mouse drag and select technique, loop the section of your job shown:

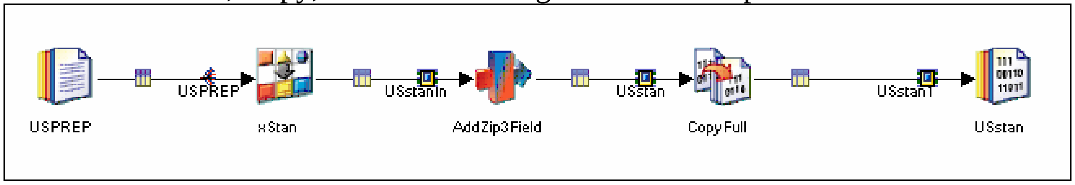


3. Click Edit ➔ Copy from the file menu or cntrl/c. This will place the job components into your Windows clipboard.
4. Create a new parallel job.
5. Paste into the new job.

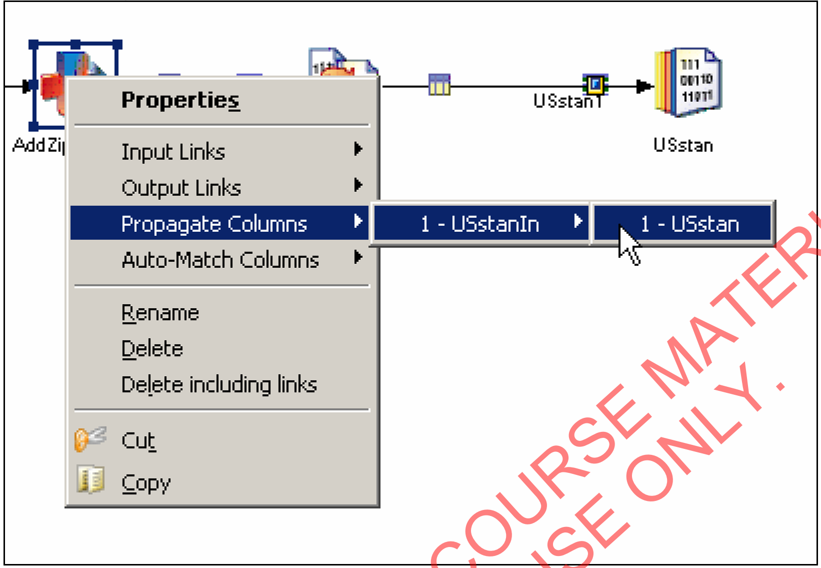


6. Save your job and name it StanAndAddFields.

7. Add Transformer, Copy, and Data Set stages and links to produce:

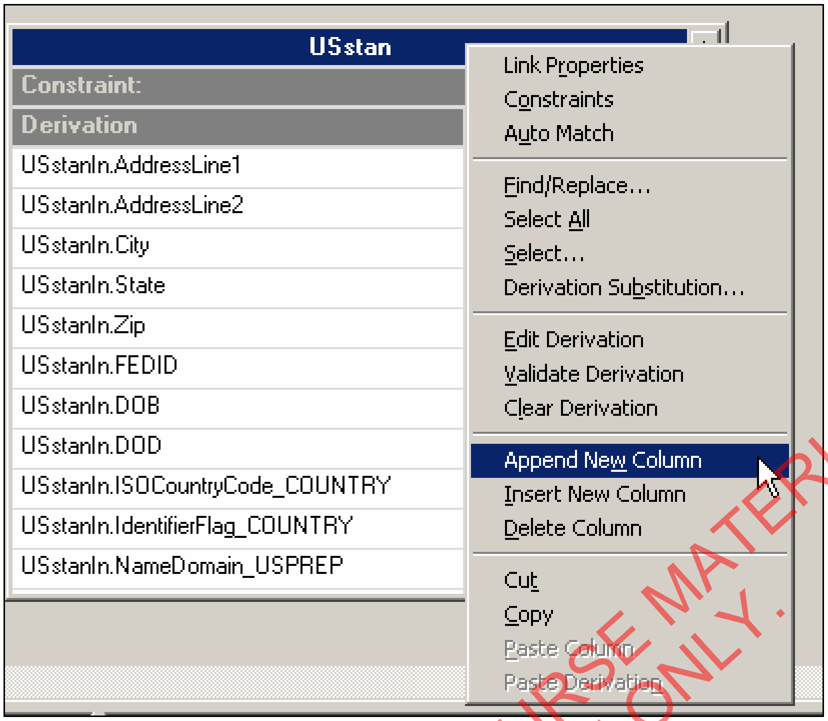


8. Right-click the Transformer stage and click Propagate Columns➔ 1-USstanIn➔ 1-USstan. 1-USstanIn and 1-USstan represent the link names. If you named your links differently then use appropriate values.



9. Double-click the Transformer stage to edit its properties.

10. Right-click on the output link (upper right pane labeled USstan) and click Append New Column.



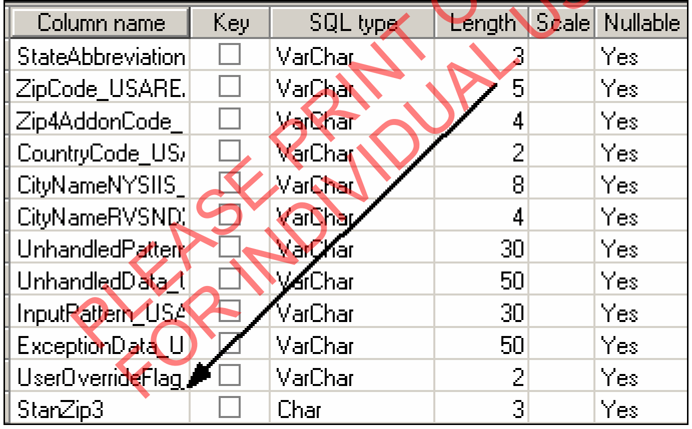
USstan

Constraint:

Derivation

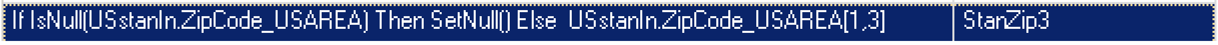
USstanIn.AddressLine1
USstanIn.AddressLine2
USstanIn.City
USstanIn.State
USstanIn.Zip
USstanIn.FEDID
USstanIn.DOB
USstanIn.DOD
USstanIn.ISOCountryCode_COUNTRY
USstanIn.IdentifierFlag_COUNTRY
USstanIn.NameDomain_USPREP

11. In the output link meta data grid, double-click on the new column; it will be the last one.
12. Change the column name to StanZip3, SQL type = char, length = 3.



Column name	Key	SQL type	Length	Scale	Nullable
StateAbbreviation	☐	VarChar	3		Yes
ZipCode_USARE.	☐	VarChar	5		Yes
Zip4AddonCode_	☐	VarChar	4		Yes
CountryCode_US/	☐	VarChar	2		Yes
CityNameNYSIIS_	☐	VarChar	8		Yes
CityNameRVSND	☐	VarChar	4		Yes
UnhandledPatterr	☐	VarChar	30		Yes
UnhandledData_I	☐	VarChar	50		Yes
InputPattern_USA	☐	VarChar	30		Yes
ExceptionData_U	☐	VarChar	50		Yes
UserOverrideFlag_	☐	VarChar	2		Yes
StanZip3	☐	Char	3		Yes

13. Double-click the column derivation for StanZip3 and enter the following code: If IsNull(USstanIn.ZipCode_USAREA) Then SetNull() Else USstanIn.ZipCode_USAREA[1,3] yielding:

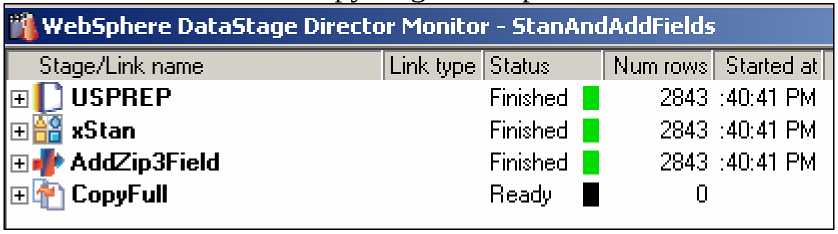


If IsNull(USstanIn.ZipCode_USAREA) Then SetNull() Else USstanIn.ZipCode_USAREA[1,3]	StanZip3

14. Click the OK button to save settings and close the Transformer.
15. Right-click on the Copy stage named CopyFull and click Auto-map columns ➔ All output links. Although the Copy stage is unnecessary for this job, it will be used later to

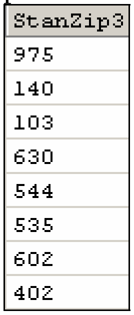
create two output links. For the purposes of the current job, the Copy operator will be optimized out of the flow when the job executes.

16. Open the target Data Set stage labeled USstan and set the File property to C:\WINNCRM\StanUSNameAddrArea.ds – the .ds extension is significant and must be used.
17. Save, compile and execute the job. The monitor may show 0 records for the Copy stage link count because the Copy stage was optimized out of the flow. This is normal.

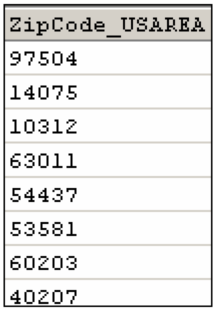

WebSphere DataStage Director Monitor - StanAndAddFields

Stage/Link name	Link type	Status		Num rows	Started at
⊞ USPREP		Finished	🟩	2843	:40:41 PM
⊞ xStan		Finished	🟩	2843	:40:41 PM
⊞ AddZip3Field		Finished	🟩	2843	:40:41 PM
⊞ CopyFull		Ready	■	0	

18. Examine the data in the target Data Set stage by using the View Data option. Verify the presence of the StanZip3 field. You can also validate it against the parent Zip field.



StanZip3
975
140
103
630
544
535
602
402



ZipCode_USAREA
97504
14075
10312
63011
54437
53581
60203
40207

Lab 17: Add Match Frequency

Objectives:

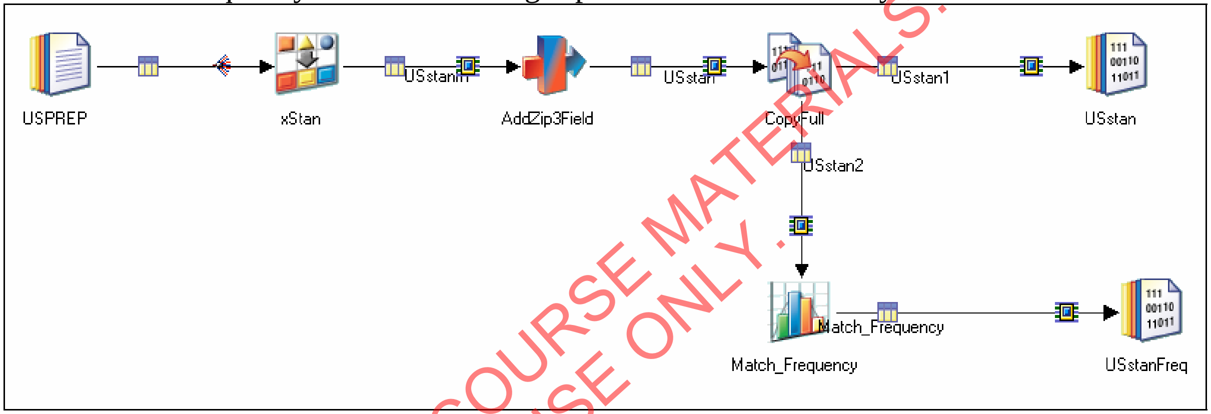
- Produce frequency report data set to be used by match processing

Assumptions:

- Job StanAndAddFields exists

Task: Build and run a Standardize and Match Frequency job

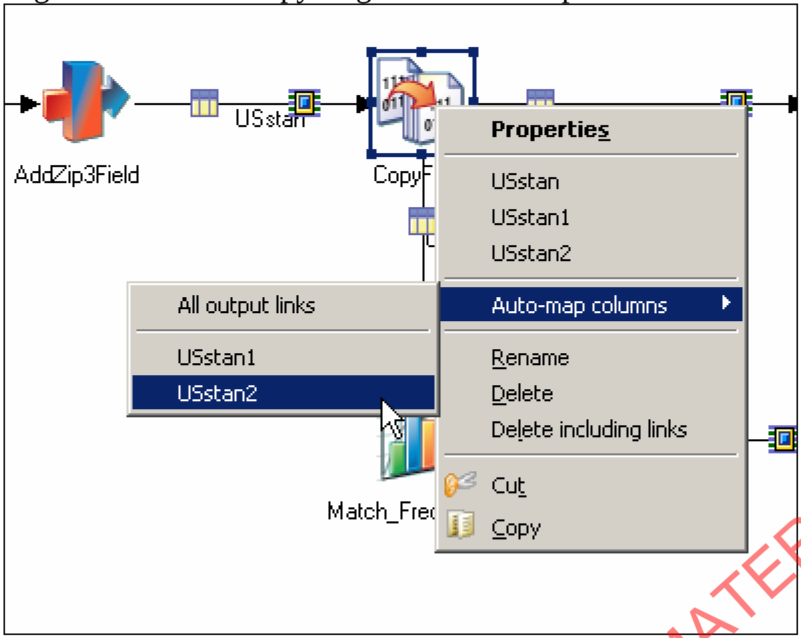
1. Edit the StanAndAddFields job and save as StanAndGenFreq.
2. Insert Match Frequency and Data Set stages plus associated links to yield:



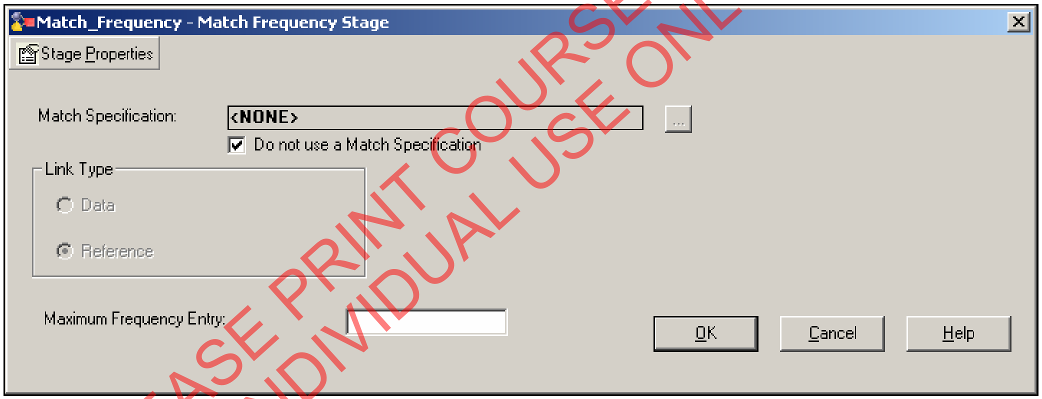
This job standardizes several fields, and then adds a new field via the transformer. The new field (which is an extract of an existing zip code field) will go into the Match Frequency stage to produce frequency counts for data values. The copy stage splits the data flow so that you end up with both the standardized data and a frequency report.

3. Stage properties for the USPREP Sequential File stage, xStan Standardization stage, Transformer stage, and USstan Data Set stage are already assigned.

4. Right-click on the Copy stage and Auto-map columns to USstan2 (the new output link).



5. Edit the Match Frequency stage and select the “Do not use a match specification” checkbox and then click OK.


Match_Frequency - Match Frequency Stage

Stage Properties

Match Specification: <NONE>

☑ Do not use a Match Specification

Link Type

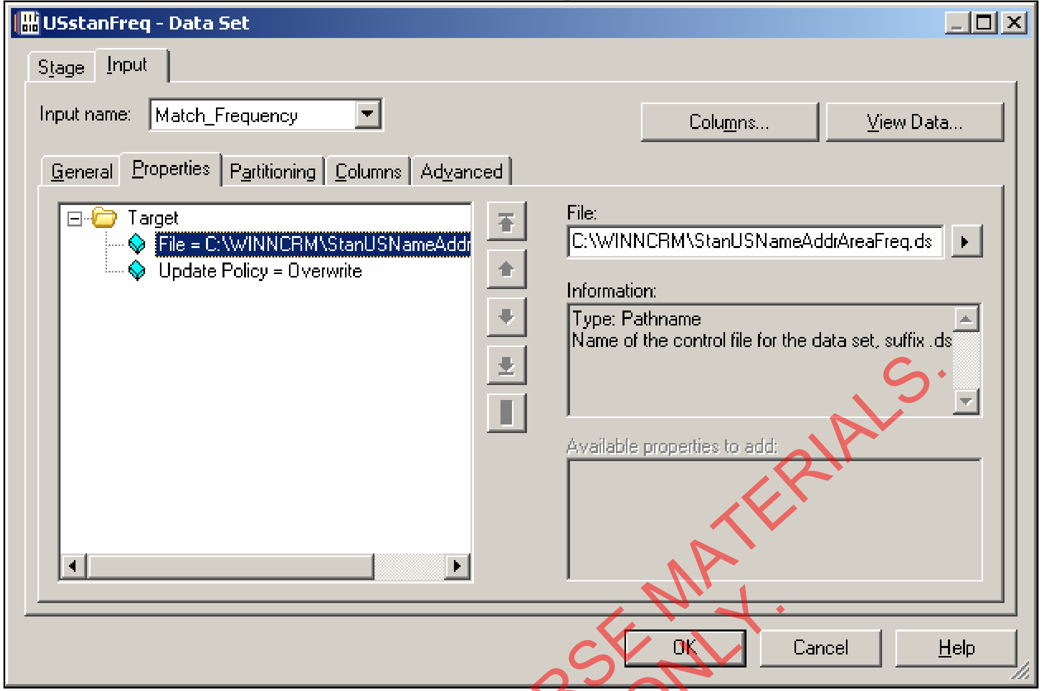
○ Data

● Reference

Maximum Frequency Entry:

6. Auto-map columns for the Match Frequency stage to the output link.

7. Edit the USstanFreq Data Set stage and set the File property to C:\WINNCRM\StanUSNameAddrAreaFreq.ds.



USstanFreq - Data Set

Input name: Match_Frequency

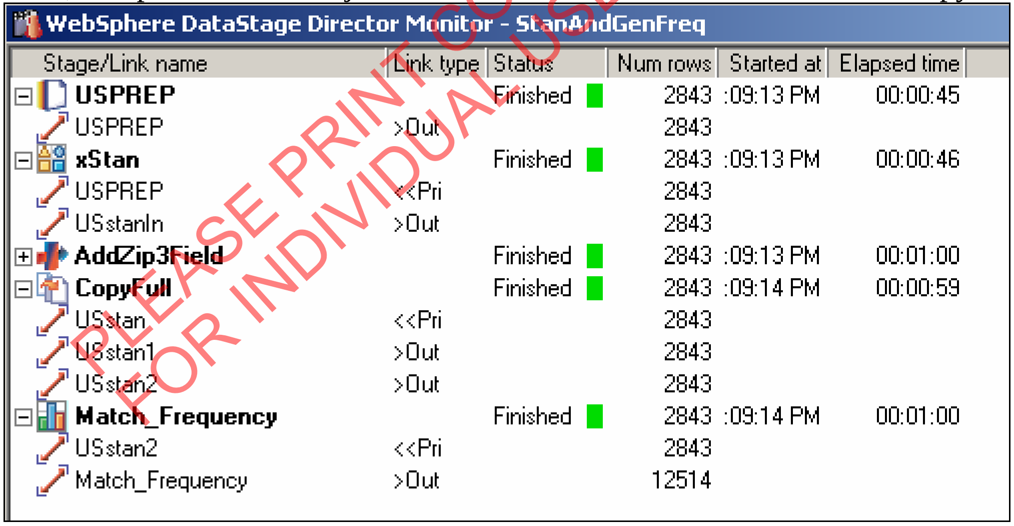
Target
- File = C:\WINNCRM\StanUSNameAddr
- Update Policy = Overwrite

File:
C:\WINNCRM\StanUSNameAddrAreaFreq.ds

Information:
Type: Pathname
Name of the control file for the data set, suffix .ds

Available properties to add:

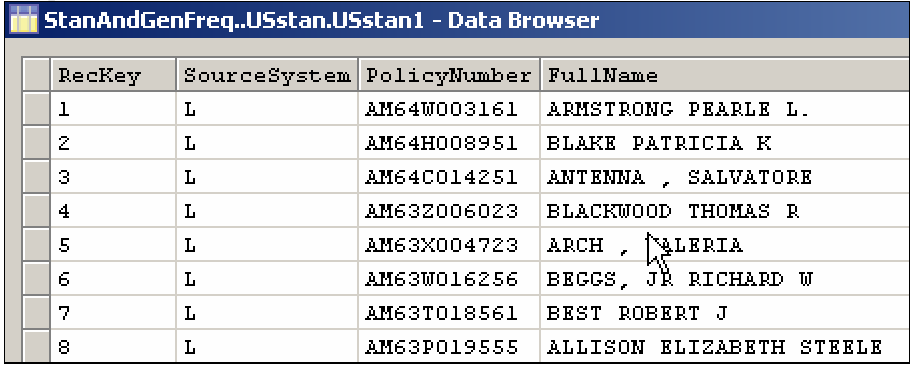
8. Save, compile and run the job. Note the non-zero record count for the CopyFull stage.



WebSphere DataStage Director Monitor - StanAndGenFreq

Stage/Link name	Link type	Status	Num rows	Started at	Elapsed time
USPREP		Finished 🟩	2843	:09:13 PM	00:00:45
USPREP	>Out		2843		
xStan		Finished 🟩	2843	:09:13 PM	00:00:46
USPREP	<<Pri		2843		
USstanIn	>Out		2843		
AddZip3Field		Finished 🟩	2843	:09:13 PM	00:01:00
CopyFull		Finished 🟩	2843	:09:14 PM	00:00:59
USstan	<<Pri		2843		
USstan1	>Out		2843		
USstan2	>Out		2843		
Match_Frequency		Finished 🟩	2843	:09:14 PM	00:01:00
USstan2	<<Pri		2843		
Match_Frequency	>Out		12514		

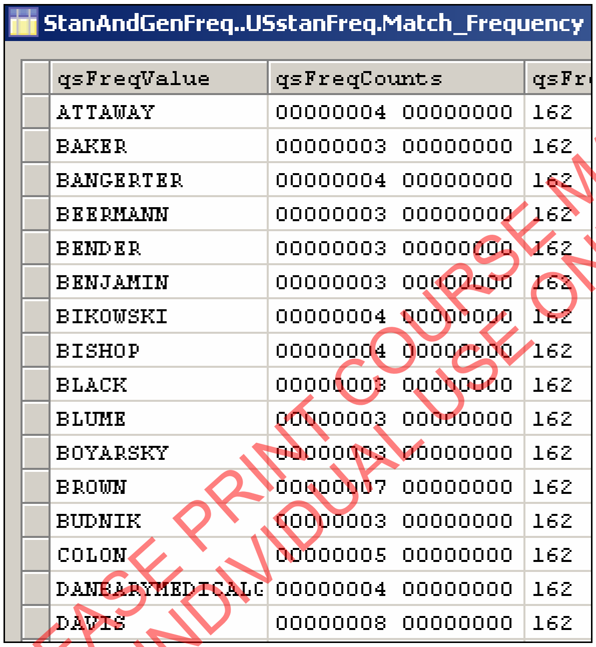
9. Use the View Data button to view the output Data Sets for both standardized data and the frequencies



StanAndGenFreq..USstan.USstan1 - Data Browser

RecKey	SourceSystem	PolicyNumber	FullName
1	L	AM64W003161	ARMSTRONG PEARLE L.
2	L	AM64H008951	BLAKE PATRICIA K
3	L	AM64C014251	ANTENNA , SALVATORE
4	L	AM63Z006023	BLACKWOOD THOMAS R
5	L	AM63X004723	ARCH , VALERIA
6	L	AM63W016256	BEGGS, JR RICHARD W
7	L	AM63T018561	BEST ROBERT J
8	L	AM63P019555	ALLISON ELIZABETH STEELE

Partial listing for standardized data



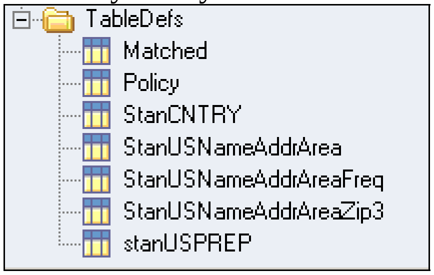
StanAndGenFreq..USstanFreq.Match_Frequency

qsFreqValue	qsFreqCounts	qsFr
ATTAWAY	00000004 00000000	162
BAKER	00000003 00000000	162
BANGERTER	00000004 00000000	162
BEERMANN	00000003 00000000	162
BENDER	00000003 00000000	162
BENJAMIN	00000003 00000000	162
BIKOWSKI	00000004 00000000	162
BISHOP	00000004 00000000	162
BLACK	00000003 00000000	162
BLUME	00000003 00000000	162
BOYARSKY	00000003 00000000	162
BROWN	00000007 00000000	162
BUDNIK	00000003 00000000	162
COLON	00000005 00000000	162
DANBARYMEDICALG	00000004 00000000	162
DAVIS	00000008 00000000	162

Portion of the frequency data set

10. For each of the Data Set stages, go to the Columns tab and click the Save button. Save the column definitions for the USstan stage (save as a table named StanUSNameAddrAreaZip3) and the USstanFreq stage (save as a table named StanUSNameAddrAreaFreq). We will use these definitions in later jobs.

11. Save all table definitions in the WINNCRM ➔ TableDefs folder in the repository. Currently it may look like:



TableDefs
- Matched
- Policy
- StanCNTRY
- StanUSNameAddrArea
- StanUSNameAddrAreaFreq
- StanUSNameAddrAreaZip3
- stanUSPREP

Lab 18: Configure test results database

Objectives:

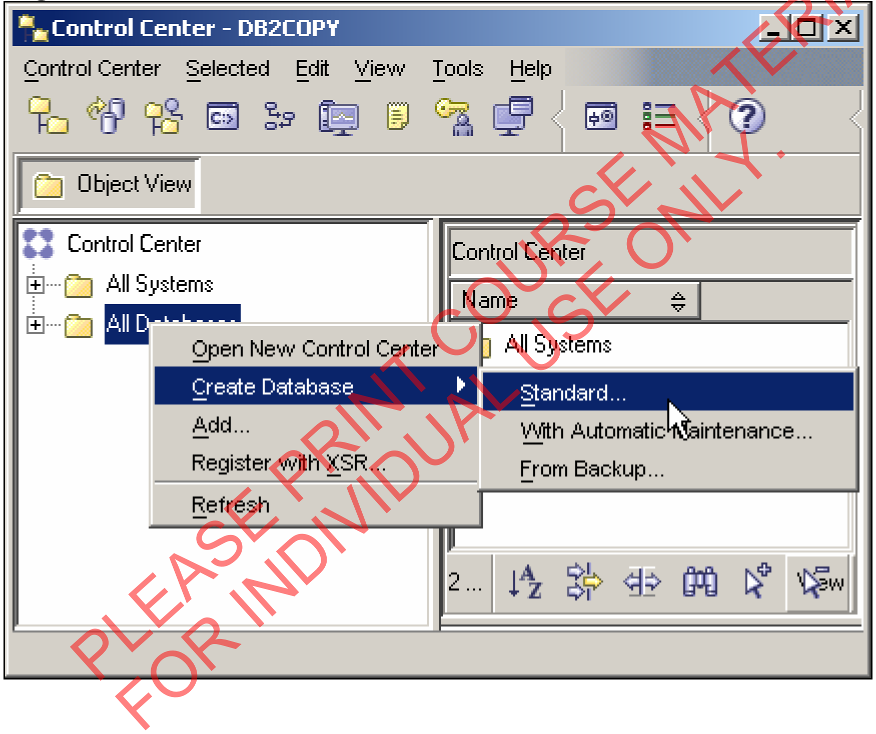
- Prepare for match specification design

Assumptions:

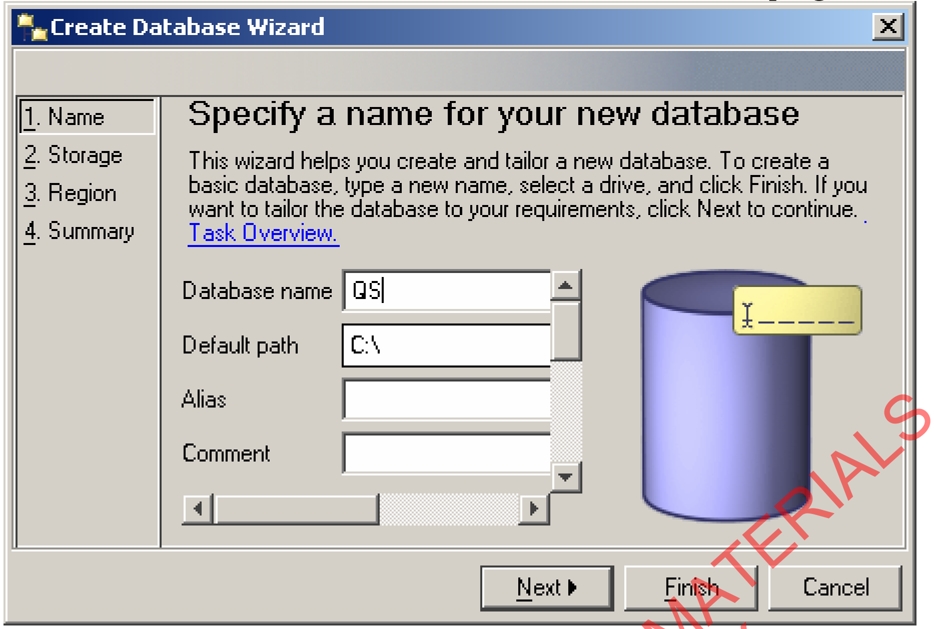
- DB2 installed and available
- Windows platform
- Student knows the DB2 admin id and password

Task: Build DB2 test results database

1. On Windows Start ➔ Programs ➔IBM DB2 ➔DB2Copy (default) ➔General Administration Tools ➔ Control Center
2. Right-click on the All Databases folder. Select Create Database ➔ Standard



3. For Database name enter QS and click the Finish button. A progress screen will appear.



Create Database Wizard

Specify a name for your new database

This wizard helps you create and tailor a new database. To create a basic database, type a new name, select a drive, and click Finish. If you want to tailor the database to your requirements, click Next to continue. Task Overview.

Database name: QS

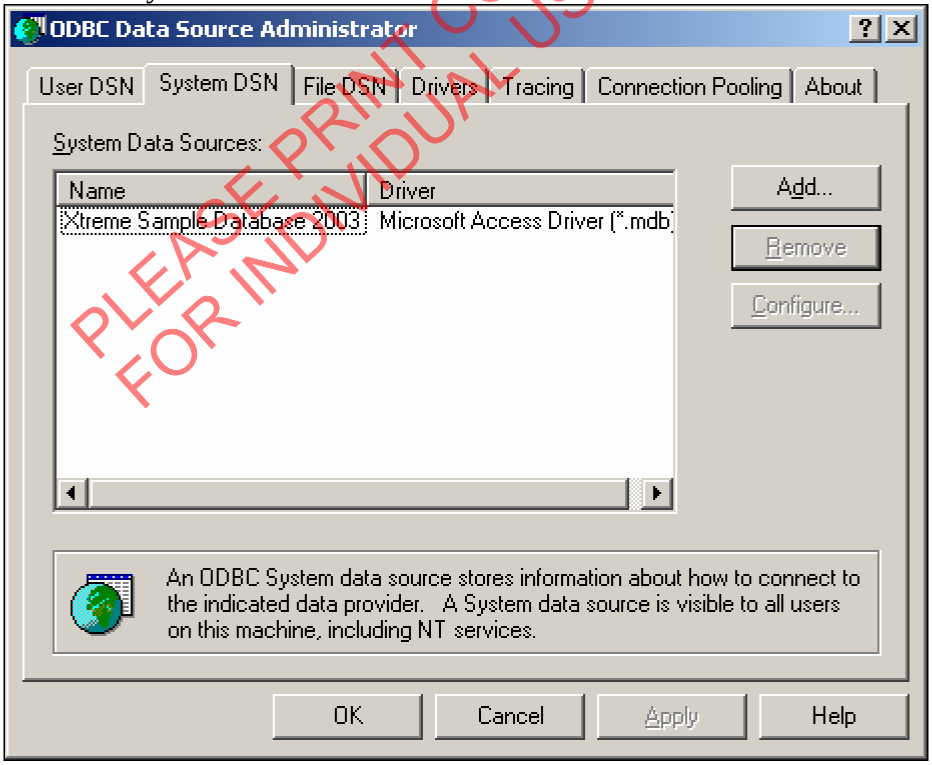
Default path: C:\

Alias:

Comment:

Task: Configure ODBC

1. Go to the Windows ODBC Data Source Administrator. On Windows XP this is done by going to Control Panel ➔ Administrative Tools ➔ Data Sources.
2. Click the System DSN tab.



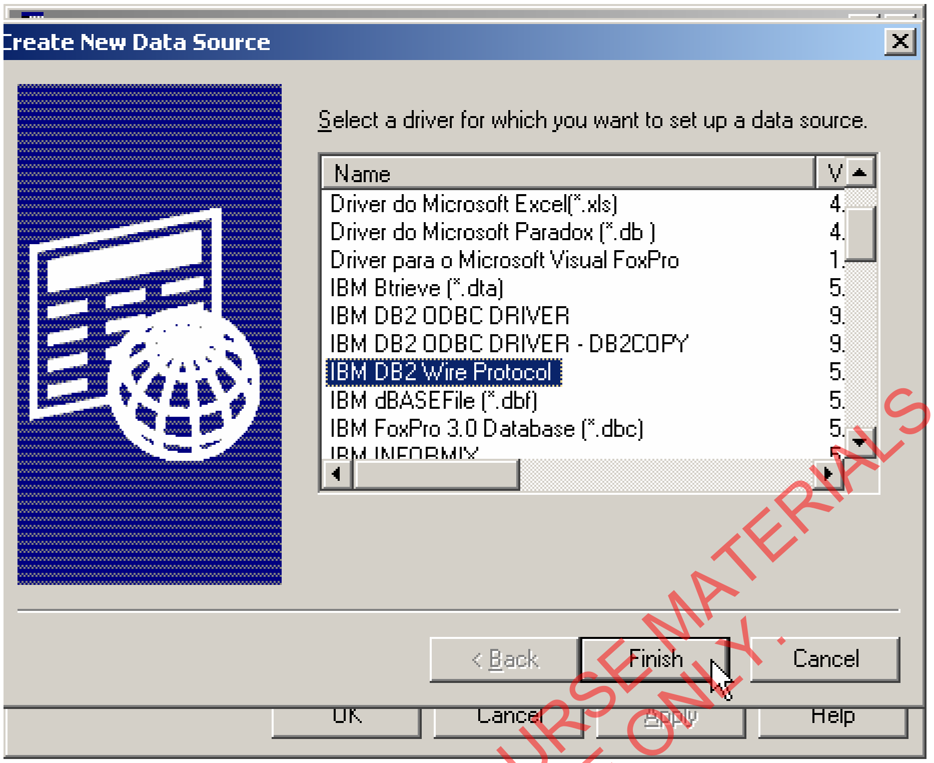
ODBC Data Source Administrator

System Data Sources:

Name	Driver
Xtreme Sample Database 2003	Microsoft Access Driver (*.mdb)

An ODBC System data source stores information about how to connect to the indicated data provider. A System data source is visible to all users on this machine, including NT services.

3. Click the Add button.
4. Scroll down and select the DB2 Wire Protocol driver and then click the Finish button.

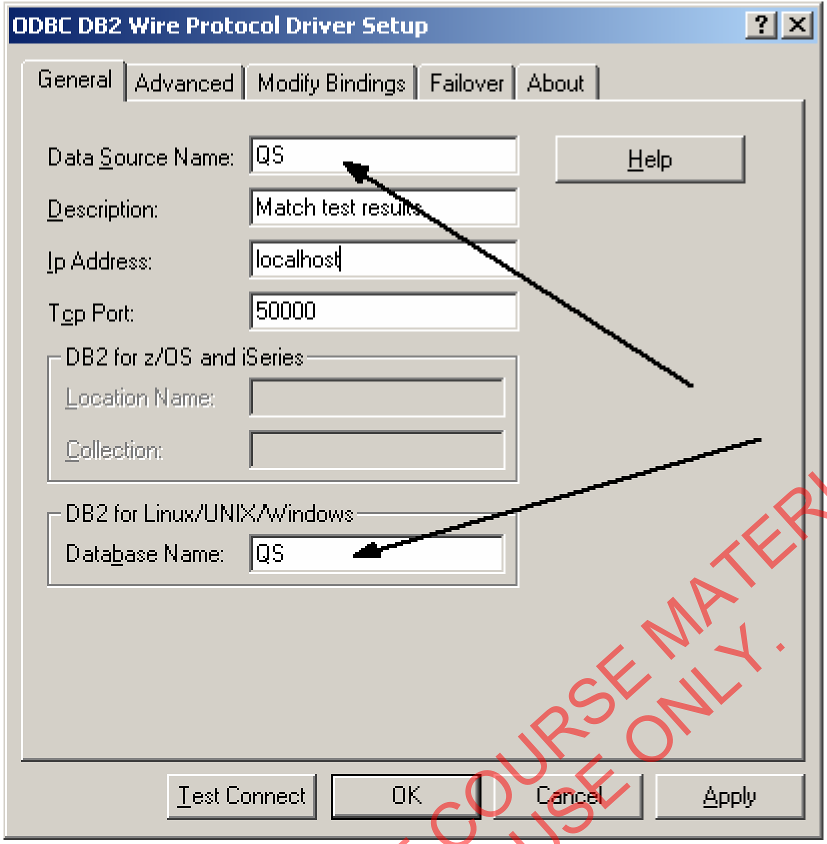


Create New Data Source

Select a driver for which you want to set up a data source.

Name	V
Driver do Microsoft Excel(*.xls)	4.
Driver do Microsoft Paradox (*.db)	4.
Driver para o Microsoft Visual FoxPro	1.
IBM Btrieve (*.dta)	5.
IBM DB2 ODBC DRIVER	9.
IBM DB2 ODBC DRIVER - DB2COPY	9.
IBM DB2 Wire Protocol	5.
IBM dBASEFile (*.dbf)	5.
IBM FoxPro 3.0 Database (*.dbc)	5.
IBM INFORMIX	5.

5. Enter QS for Data Source Name and QS for Database Name.



ODBC DB2 Wire Protocol Driver Setup

Data Source Name: QS

Description: Match test results

Ip Address: localhost

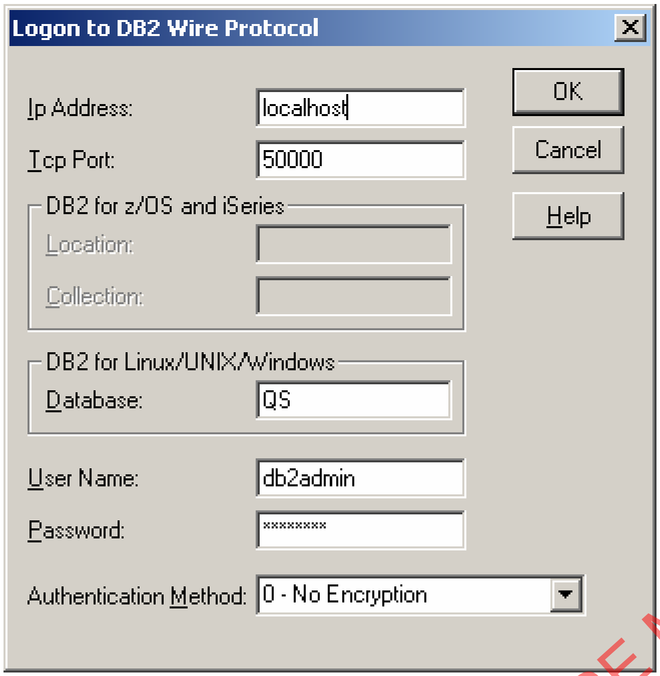
Tcp Port: 50000

DB2 for Linux/UNIX/Windows

Database Name: QS

6. Click the Test Connect button.

7. Enter the correct User Name and Password. In the case of this example these were db2admin/db2admin.



Logon to DB2 Wire Protocol

Ip Address: localhost

Tcp Port: 50000

DB2 for z/OS and iSeries

Location:

Collection:

DB2 for Linux/UNIX/Windows

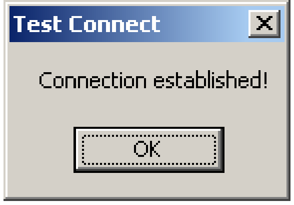
Database: QS

User Name: db2admin

Password: ********

Authentication Method: 0 - No Encryption

8. If your information is correct you should receive a confirmation window.



Test Connect

Connection established!

Lab 19: Match Specification (Unduplicate)

Objectives:

- Use the Match Designer to build a match specification that will be used in a later Match process.

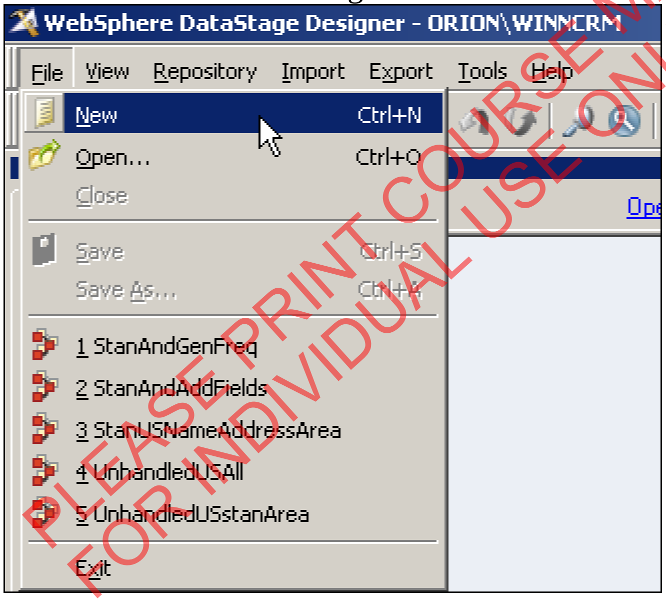
Assumptions:

- StanAndGenFreq job exists and ran successfully
- A database is available via ODBC for a specification test environment

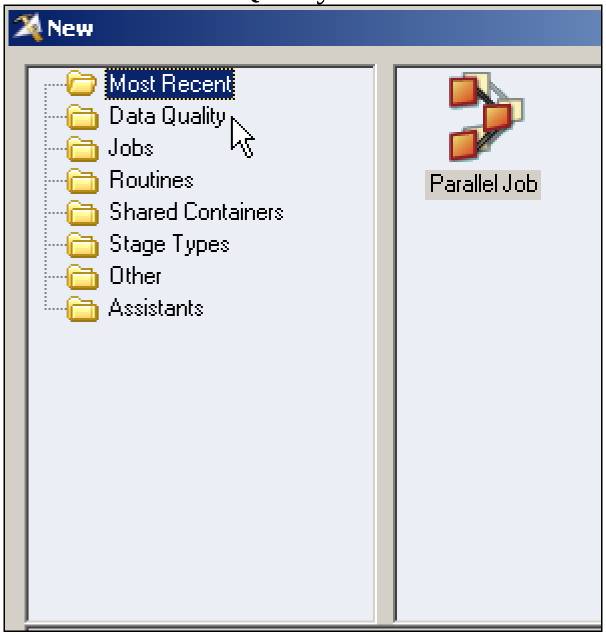
Task: Build a match specification

In this task you build a match specification online and use that specification to test the validity of the match logic. Prior to version 8 match strategy was developed by running jobs in batch. Quality Stage version 8 includes the Match Designer, an online system that lets you create and test match criteria online.

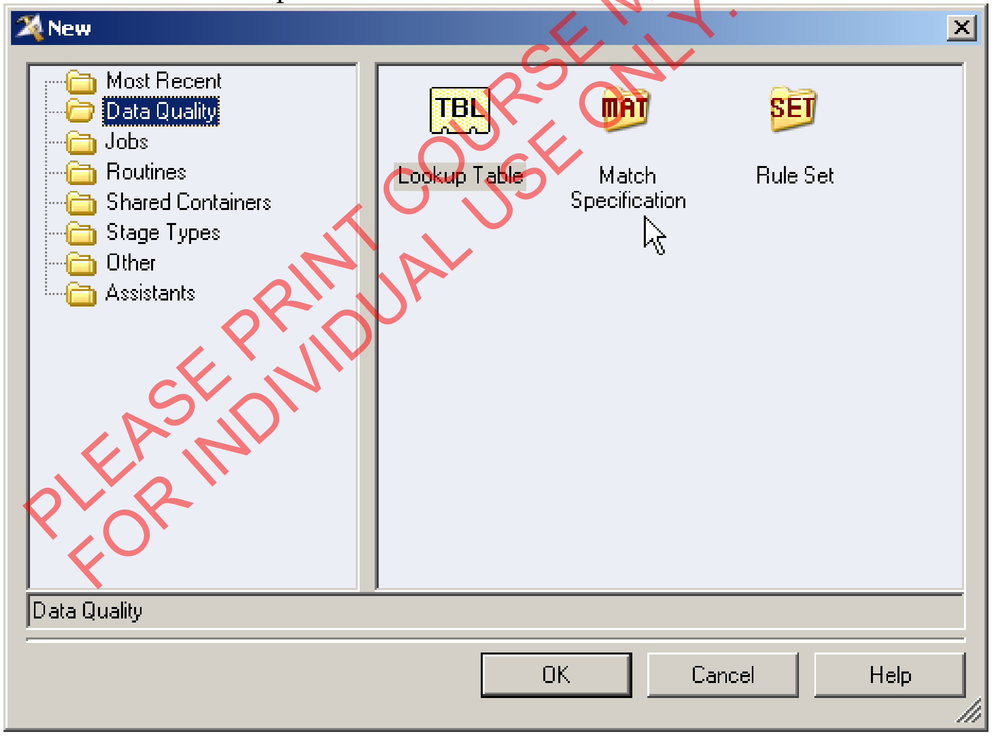
1. From the menu bar in Designer click File ➔ New.



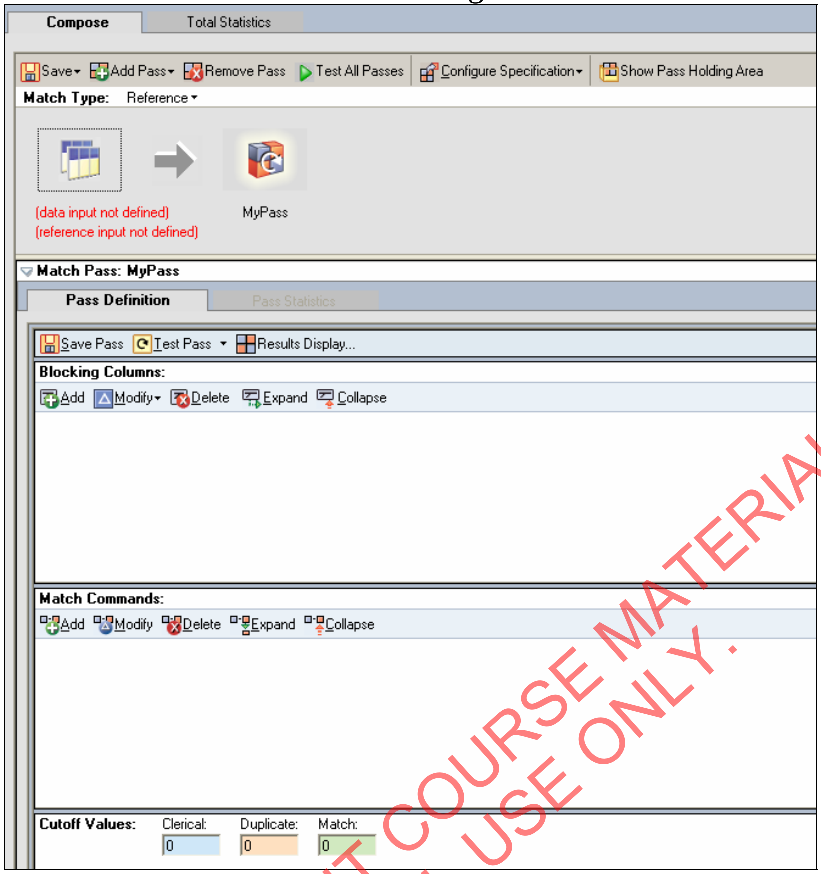
2. Click on the Data Quality folder.



3. Click on the Match Specification icon. And then click the OK button.



4. You should now see the Match Designer.



Match Type: Reference

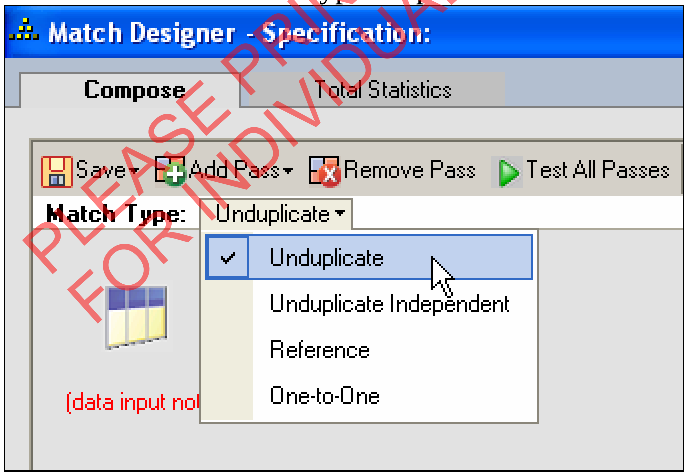
Match Pass: MyPass

Blocking Columns:

Match Commands:

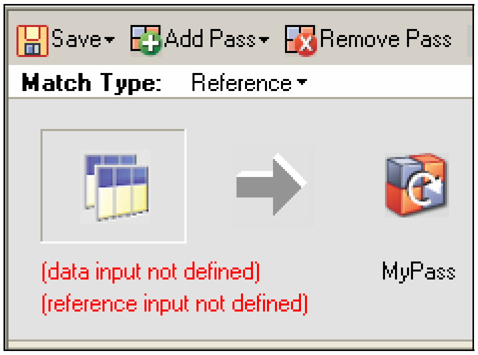
Cutoff Values: Clerical: 0 Duplicate: 0 Match: 0

5. Down-click the Match Type drop down box and select Unduplicate.



Match Type: Unduplicate

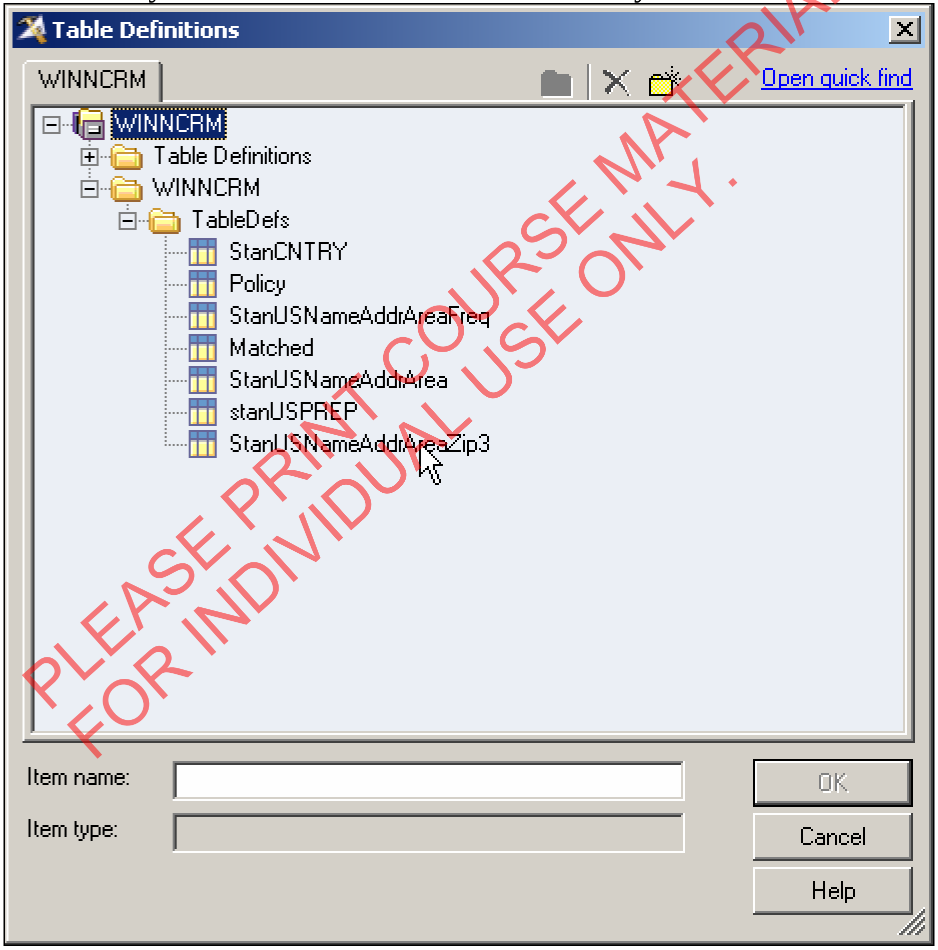
6. Click on the table icon to select the table definition you will use for the match. The table icon can be found to the left of the MyPass icon.



Match Type: Reference

(data input not defined)
(reference input not defined)

7. Use the Load button to retrieve the StanUSNameAddrAreaZip3 column definitions from the WINNCRM➔TableDefs folder in the repository. These are the column definitions you saved from the standardized data you created earlier.

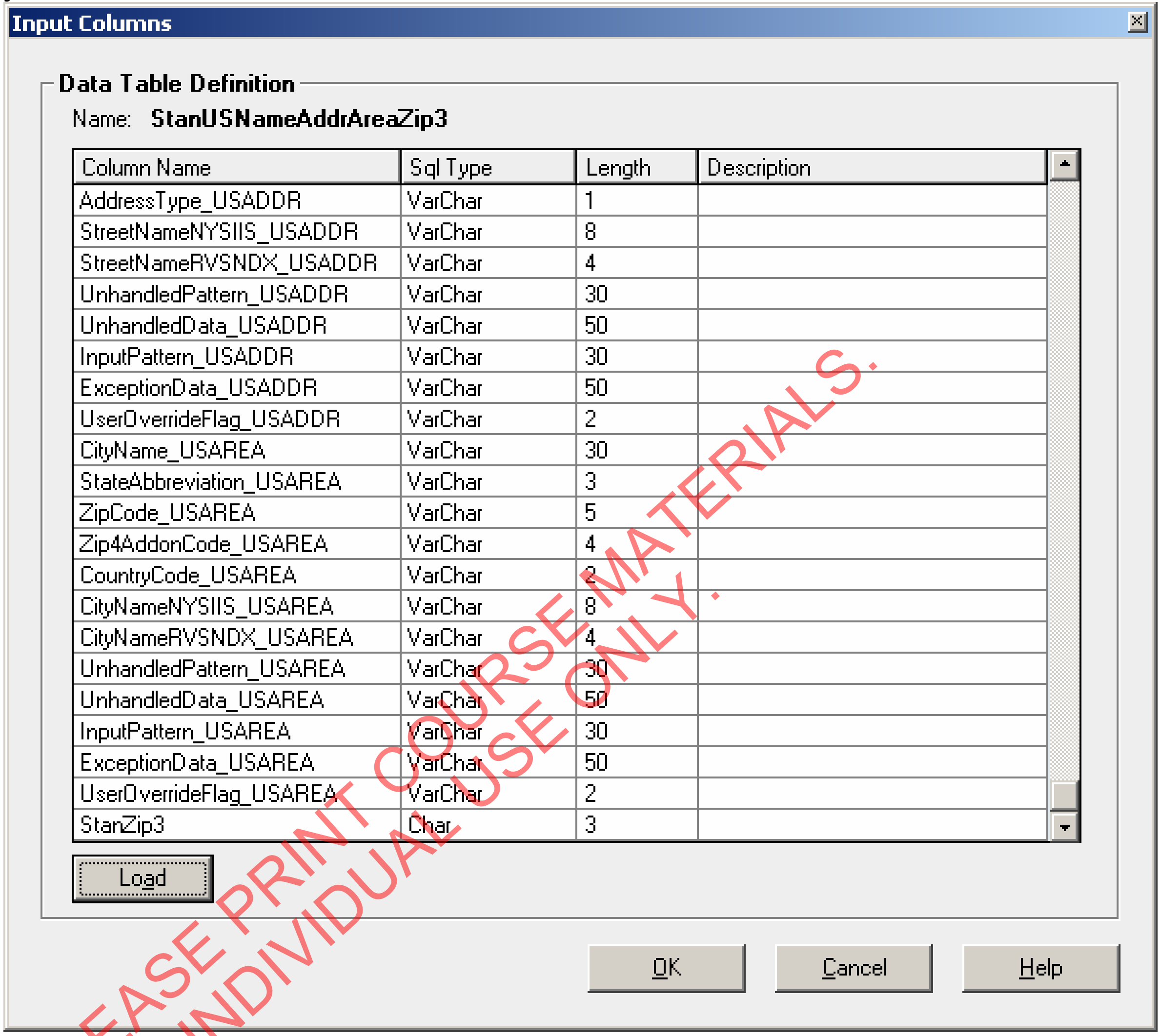


- WINNCRM
  - Table Definitions
  - WINNCRM
    - TableDefs
      - StanCNTRY
      - Policy
      - StanUSNameAddrAreaFreq
      - Matched
      - StanUSNameAddrArea
      - stanUSPREP
      - StanUSNameAddrAreaZip3

Item name:

Item type:

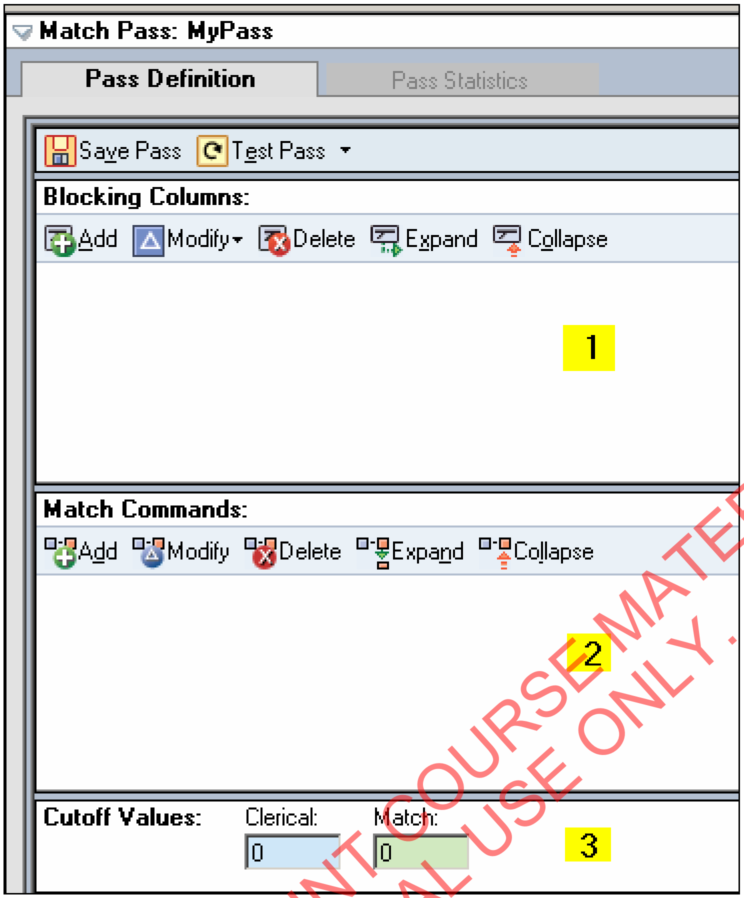
8. To verify you have the correct definition, look for the three character zip code field you added to the end of the table. Click the OK button.



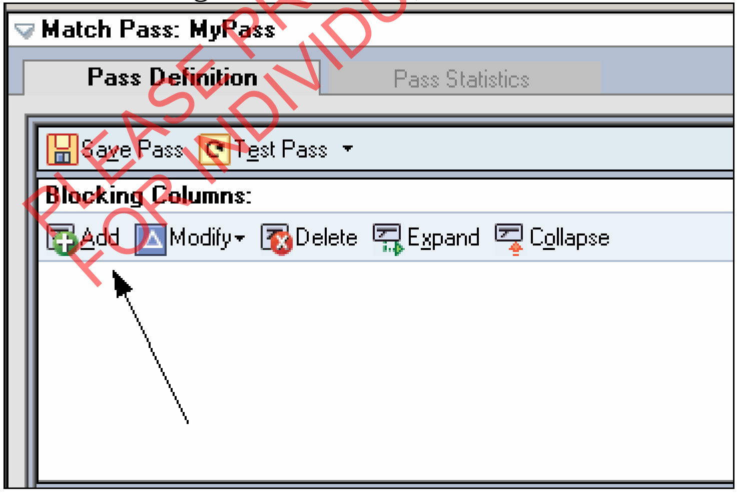
Name: StanUSNameAddrAreaZip3

Column Name	Sql Type	Length	Description
AddressType_USADDR	VarChar	1	
StreetNameNYSIIS_USADDR	VarChar	8	
StreetNameRVSNDX_USADDR	VarChar	4	
UnhandledPattern_USADDR	VarChar	30	
UnhandledData_USADDR	VarChar	50	
InputPattern_USADDR	VarChar	30	
ExceptionData_USADDR	VarChar	50	
UserOverrideFlag_USADDR	VarChar	2	
CityName_USAREA	VarChar	30	
StateAbbreviation_USAREA	VarChar	3	
ZipCode_USAREA	VarChar	5	
Zip4AddonCode_USAREA	VarChar	4	
CountryCode_USAREA	VarChar	2	
CityNameNYSIIS_USAREA	VarChar	8	
CityNameRVSNDX_USAREA	VarChar	4	
UnhandledPattern_USAREA	VarChar	30	
UnhandledData_USAREA	VarChar	50	
InputPattern_USAREA	VarChar	30	
ExceptionData_USAREA	VarChar	50	
UserOverrideFlag_USAREA	VarChar	2	
StanZip3	Char	3	

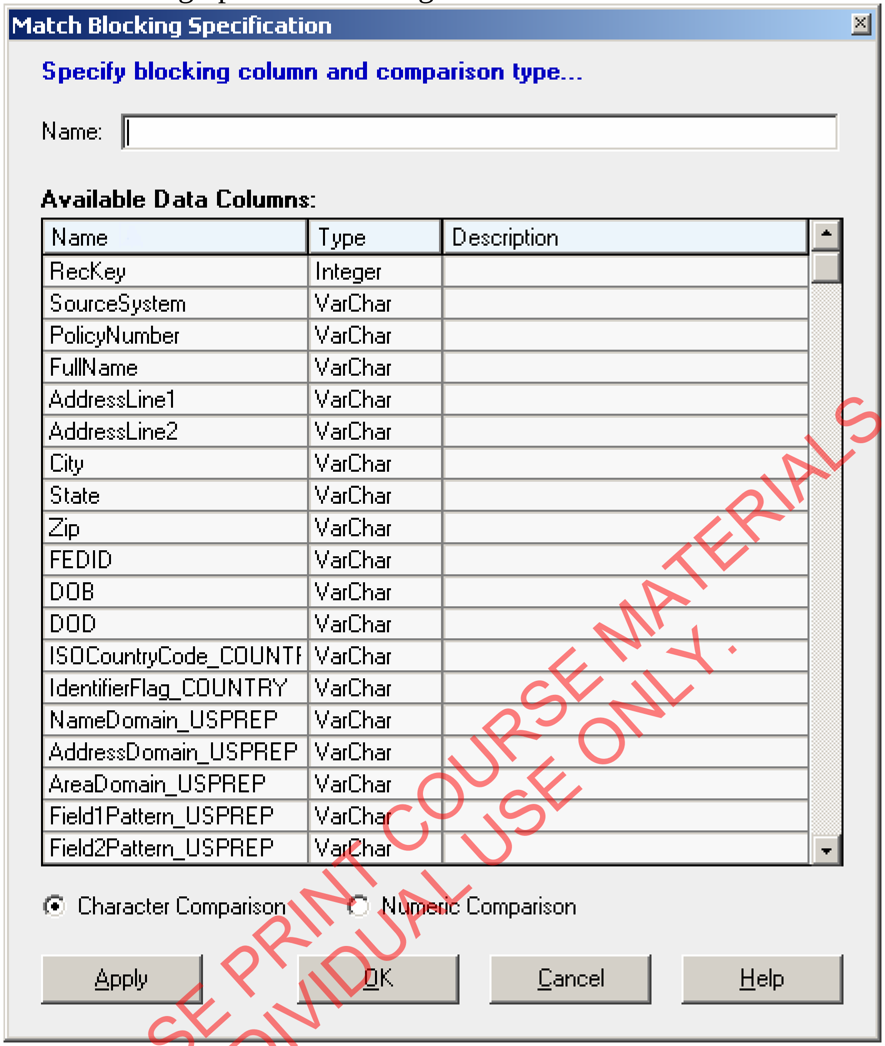
9. Under the Pass Definition are three sections: Blocking Columns, Match Commands, and Cutoff Values.



10. Under Blocking Columns click the Add icon.



11. This will bring up a screen listing all the columns from the table.



Specify blocking column and comparison type...

Name:

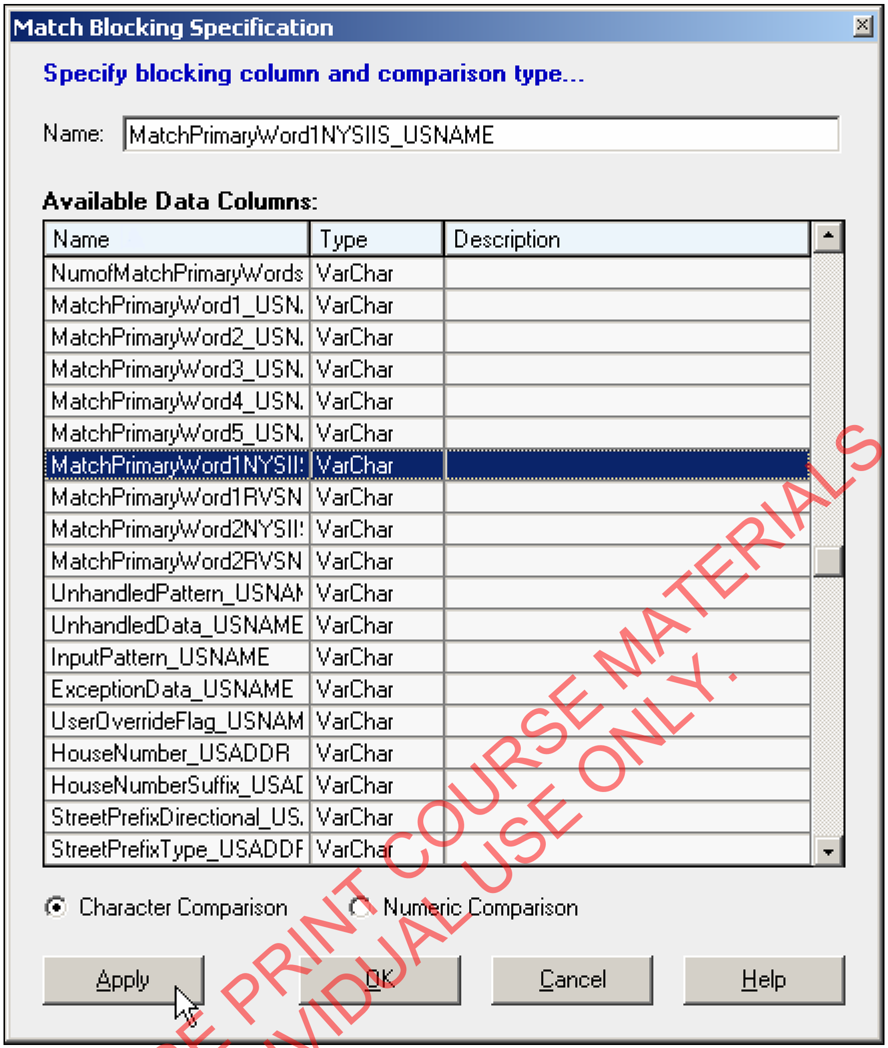
Available Data Columns:

Name	Type	Description
RecKey	Integer	
SourceSystem	VarChar	
PolicyNumber	VarChar	
FullName	VarChar	
AddressLine1	VarChar	
AddressLine2	VarChar	
City	VarChar	
State	VarChar	
Zip	VarChar	
FEDID	VarChar	
DOB	VarChar	
DOD	VarChar	
ISOCountryCode_COUNTI	VarChar	
IdentifierFlag_COUNTRY	VarChar	
NameDomain_USPREP	VarChar	
AddressDomain_USPREP	VarChar	
AreaDomain_USPREP	VarChar	
Field1Pattern_USPREP	VarChar	
Field2Pattern_USPREP	VarChar	

● Character Comparison ○ Numeric Comparison

12. Select the MatchPrimaryWord1NYSIIS_USNAME column. Note the text box near the top of the screen updates to the column name. You could change the text in the box, but take the default instead. The Character Comparison radio button should

remain checked.



Match Blocking Specification

Specify blocking column and comparison type...

Name: MatchPrimaryWord1NYSIIS_USNAME

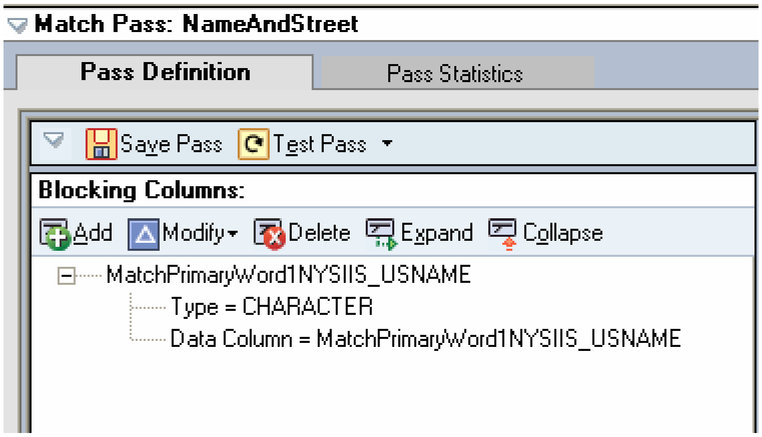
Available Data Columns:

Name	Type	Description
NumofMatchPrimaryWords	VarChar	
MatchPrimaryWord1_USN.	VarChar	
MatchPrimaryWord2_USN.	VarChar	
MatchPrimaryWord3_USN.	VarChar	
MatchPrimaryWord4_USN.	VarChar	
MatchPrimaryWord5_USN.	VarChar	
MatchPrimaryWord1NYSII:	VarChar	
MatchPrimaryWord1RVSN	VarChar	
MatchPrimaryWord2NYSII:	VarChar	
MatchPrimaryWord2RVSN	VarChar	
UnhandledPattern_USNAI	VarChar	
UnhandledData_USNAME	VarChar	
InputPattern_USNAME	VarChar	
ExceptionData_USNAME	VarChar	
UserOverrideFlag_USNAM	VarChar	
HouseNumber_USADDR	VarChar	
HouseNumberSuffix_USAI	VarChar	
StreetPrefixDirectional_US.	VarChar	
StreetPrefixType_USADDI	VarChar	

● Character Comparison ○ Numeric Comparison

13. Click the Apply button. This option allows you to stay on the Match Blocking Specification screen and select other columns for blocking. When you are on your last blocking column you may click the OK button to return to the main Match Designer screen.

14. As you apply blocking columns each data column appears under the Blocking Columns view in the Pass Definition.

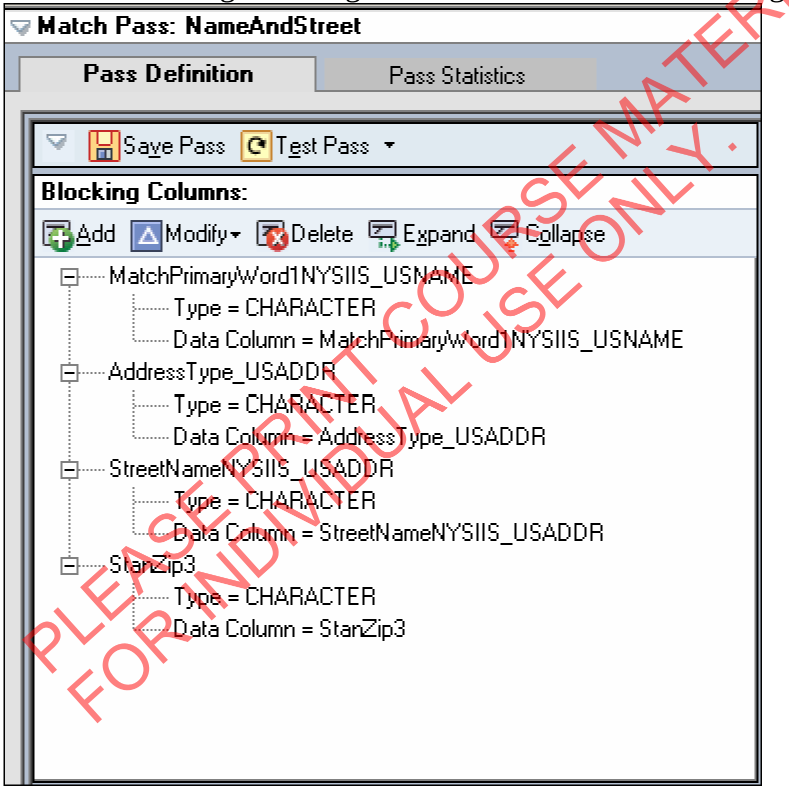


Match Pass: NameAndStreet

Blocking Columns:

```
MatchPrimaryWord1NYSIIS_USNAME
    Type = CHARACTER
    Data Column = MatchPrimaryWord1NYSIIS_USNAME
```

15. Continue building Blocking Columns to obtain the following:



Match Pass: NameAndStreet

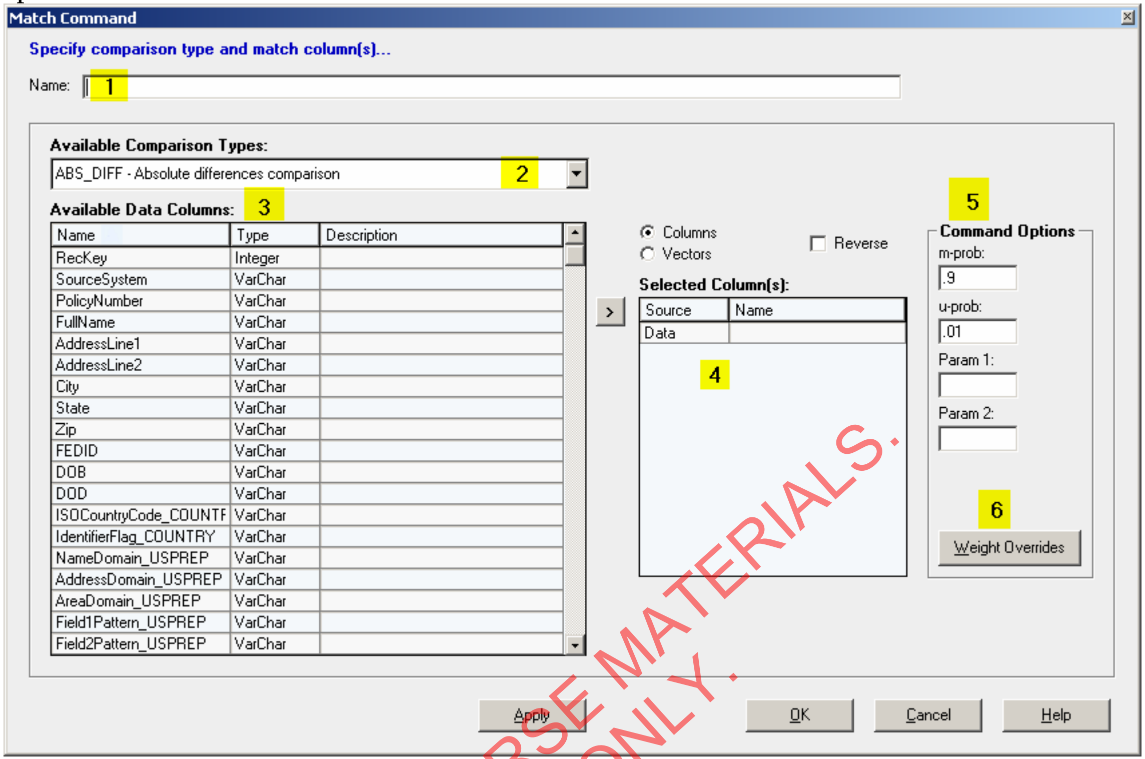
Blocking Columns:

```
MatchPrimaryWord1NYSIIS_USNAME
    Type = CHARACTER
    Data Column = MatchPrimaryWord1NYSIIS_USNAME
AddressType_USADDR
    Type = CHARACTER
    Data Column = AddressType_USADDR
StreetNameNYSIIS_USADDR
    Type = CHARACTER
    Data Column = StreetNameNYSIIS_USADDR
StanZip3
    Type = CHARACTER
    Data Column = StanZip3
```

Recall that you need to click the OK button on your last column added.

16. The second part of the Pass Definition contains the Match Commands and these will be built similar to the Blocking Columns. Click the +Add button under the Match Commands section to get started.
17. Note the position of the 1-Name text box, the 2-Available Comparison Types drop down box, the 3-Available Data Columns grid, the 4-Selected Columns box, the 5-Command Options boxes, and the 6-Override Weights button. You will complete the

specifications for most of these areas.



Specify comparison type and match column(s)...

Name:

Available Comparison Types: ABS_DIFF - Absolute differences comparison

Available Data Columns:

Name	Type	Description
RecKey	Integer	
SourceSystem	VarChar	
PolicyNumber	VarChar	
FullName	VarChar	
AddressLine1	VarChar	
AddressLine2	VarChar	
City	VarChar	
State	VarChar	
Zip	VarChar	
FEDID	VarChar	
DOB	VarChar	
DOD	VarChar	
ISOCountryCode_COUNTF	VarChar	
IdentifierFlag_COUNTRY	VarChar	
NameDomain_USPREP	VarChar	
AddressDomain_USPREP	VarChar	
AreaDomain_USPREP	VarChar	
Field1Pattern_USPREP	VarChar	
Field2Pattern_USPREP	VarChar	

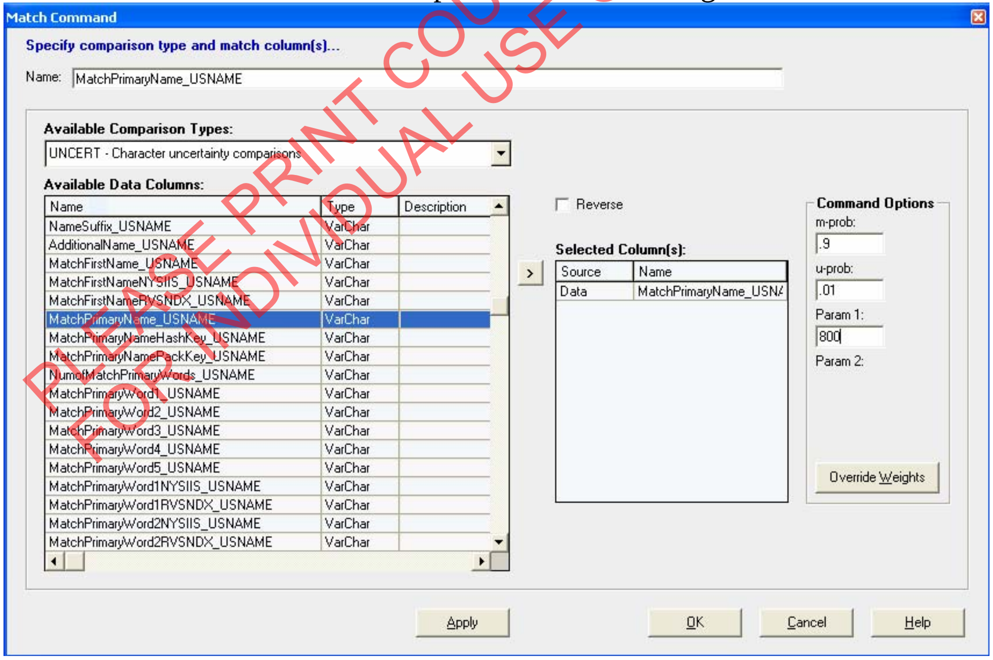
● Columns
○ Vectors
□ Reverse

Selected Column(s):

Source	Name
Data	

Command Options
m-prob: .9
u-prob: .01
Param 1:
Param 2:
Weight Overrides

18. Use this screen to enter values to produce the following:



Specify comparison type and match column(s)...

Name: MatchPrimaryName_USNAME

Available Comparison Types: UNCERT - Character uncertainty comparisons

Available Data Columns:

Name	Type	Description
NameSuffix_USNAME	VarChar	
AdditionalName_USNAME	VarChar	
MatchFirstName_USNAME	VarChar	
MatchFirstNameNYSIIS_USNAME	VarChar	
MatchFirstNameRVSNDX_USNAME	VarChar	
MatchPrimaryName_USNAME	VarChar	
MatchPrimaryNameHashKey_USNAME	VarChar	
MatchPrimaryNamePackKey_USNAME	VarChar	
NumofMatchPrimaryWords_USNAME	VarChar	
MatchPrimaryWord1_USNAME	VarChar	
MatchPrimaryWord2_USNAME	VarChar	
MatchPrimaryWord3_USNAME	VarChar	
MatchPrimaryWord4_USNAME	VarChar	
MatchPrimaryWord5_USNAME	VarChar	
MatchPrimaryWord1NYSIIS_USNAME	VarChar	
MatchPrimaryWord1RVSNDX_USNAME	VarChar	
MatchPrimaryWord2NYSIIS_USNAME	VarChar	
MatchPrimaryWord2RVSNDX_USNAME	VarChar	

□ Reverse

Selected Column(s):

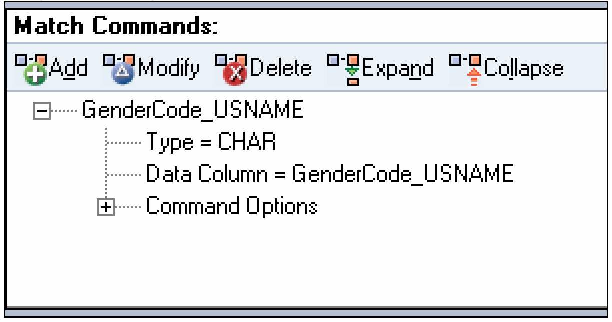
Source	Name
Data	MatchPrimaryName_USN/

Command Options
m-prob: .9
u-prob: .01
Param 1: 800
Param 2:
Override Weights

Name GenderCode_USNAME (built automatically when you move the column to the Select Columns box)
Comparison Type – CHAR - Character comparisons
Selected Column – GenderCode_USNAME

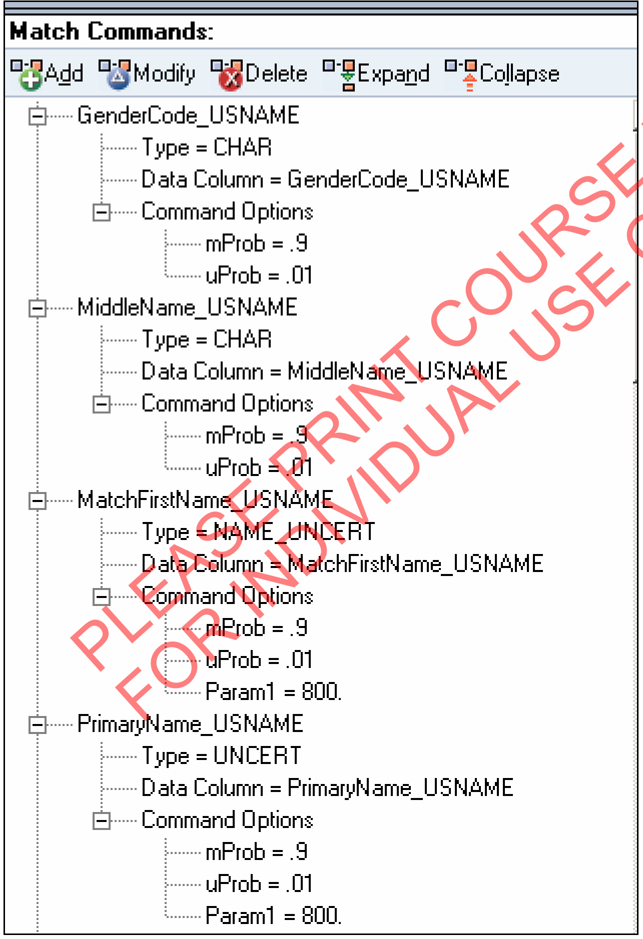
m-prob - .9
u-prob - .01

9. Click Apply to remain on the selection screen. The Match Command window will update as you apply Match Commands. .

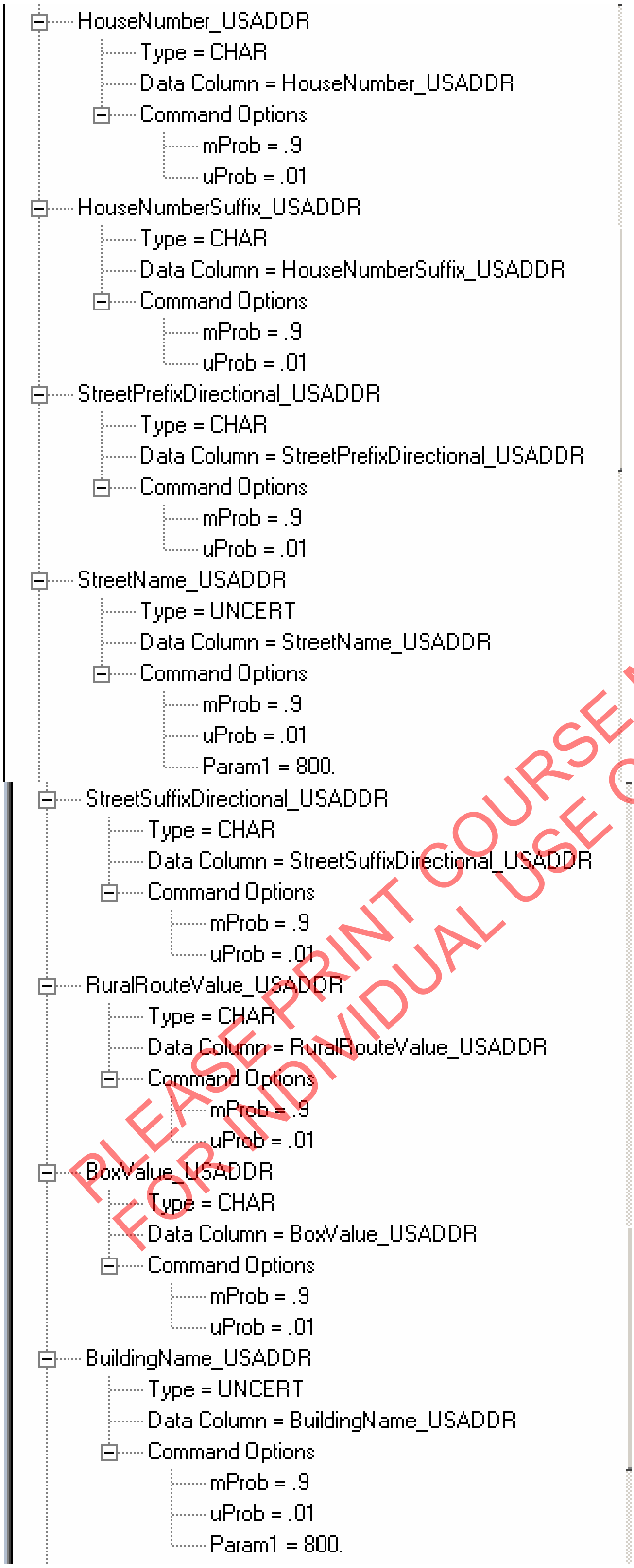


```
GenderCode_USNAME
    Type = CHAR
    Data Column = GenderCode_USNAME
    Command Options
```

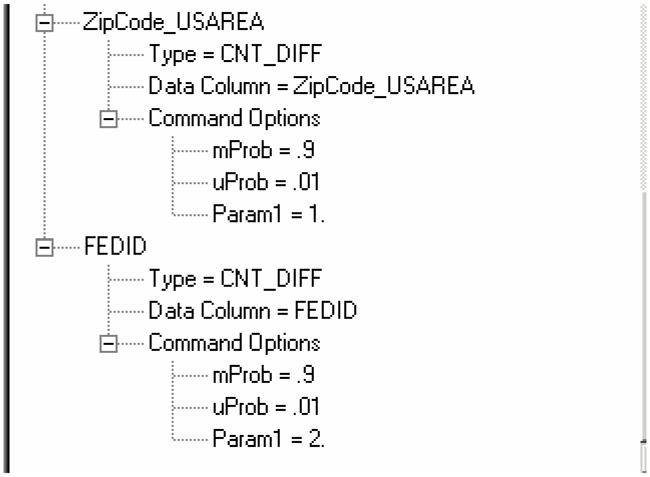
10. Continue building Match Commands to obtain the following:



```
GenderCode_USNAME
    Type = CHAR
    Data Column = GenderCode_USNAME
    Command Options
        mProb = .9
        uProb = .01
MiddleName_USNAME
    Type = CHAR
    Data Column = MiddleName_USNAME
    Command Options
        mProb = .9
        uProb = .01
MatchFirstName_USNAME
    Type = NAME_UNCERT
    Data Column = MatchFirstName_USNAME
    Command Options
        mProb = .9
        uProb = .01
        Param1 = 800.
PrimaryName_USNAME
    Type = UNCERT
    Data Column = PrimaryName_USNAME
    Command Options
        mProb = .9
        uProb = .01
        Param1 = 800.
```

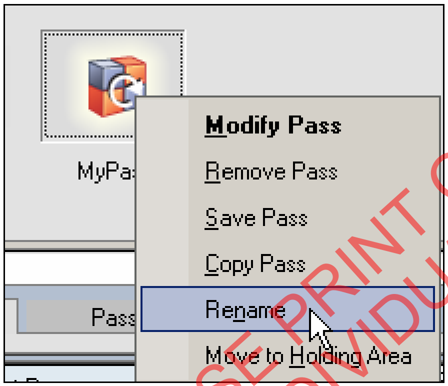


```
HouseNumber_USADDR
    Type = CHAR
    Data Column = HouseNumber_USADDR
    Command Options
        mProb = .9
        uProb = .01
HouseNumberSuffix_USADDR
    Type = CHAR
    Data Column = HouseNumberSuffix_USADDR
    Command Options
        mProb = .9
        uProb = .01
StreetPrefixDirectional_USADDR
    Type = CHAR
    Data Column = StreetPrefixDirectional_USADDR
    Command Options
        mProb = .9
        uProb = .01
StreetName_USADDR
    Type = UNCERT
    Data Column = StreetName_USADDR
    Command Options
        mProb = .9
        uProb = .01
        Param1 = 800.
StreetSuffixDirectional_USADDR
    Type = CHAR
    Data Column = StreetSuffixDirectional_USADDR
    Command Options
        mProb = .9
        uProb = .01
RuralRouteValue_USADDR
    Type = CHAR
    Data Column = RuralRouteValue_USADDR
    Command Options
        mProb = .9
        uProb = .01
BoxValue_USADDR
    Type = CHAR
    Data Column = BoxValue_USADDR
    Command Options
        mProb = .9
        uProb = .01
BuildingName_USADDR
    Type = UNCERT
    Data Column = BuildingName_USADDR
    Command Options
        mProb = .9
        uProb = .01
        Param1 = 800.
```

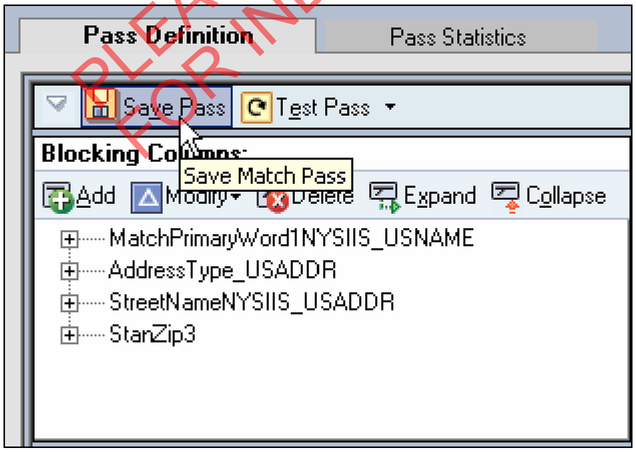


```
ZipCode_USAREA
    Type = CNT_DIFF
    Data Column = ZipCode_USAREA
    Command Options
        mProb = .9
        uProb = .01
        Param1 = 1.
FEDID
    Type = CNT_DIFF
    Data Column = FEDID
    Command Options
        mProb = .9
        uProb = .01
        Param1 = 2.
```

11. Click the OK button when entering the last Match Command to return to the primary Match Designer screen.
12. The third section of the Pass Definition is named Cutoff Values. Set both cutoff values to 25.
13. Near the top of the Match Designer screen, right-click on the MyPass icon and rename it to NameAndStreet.

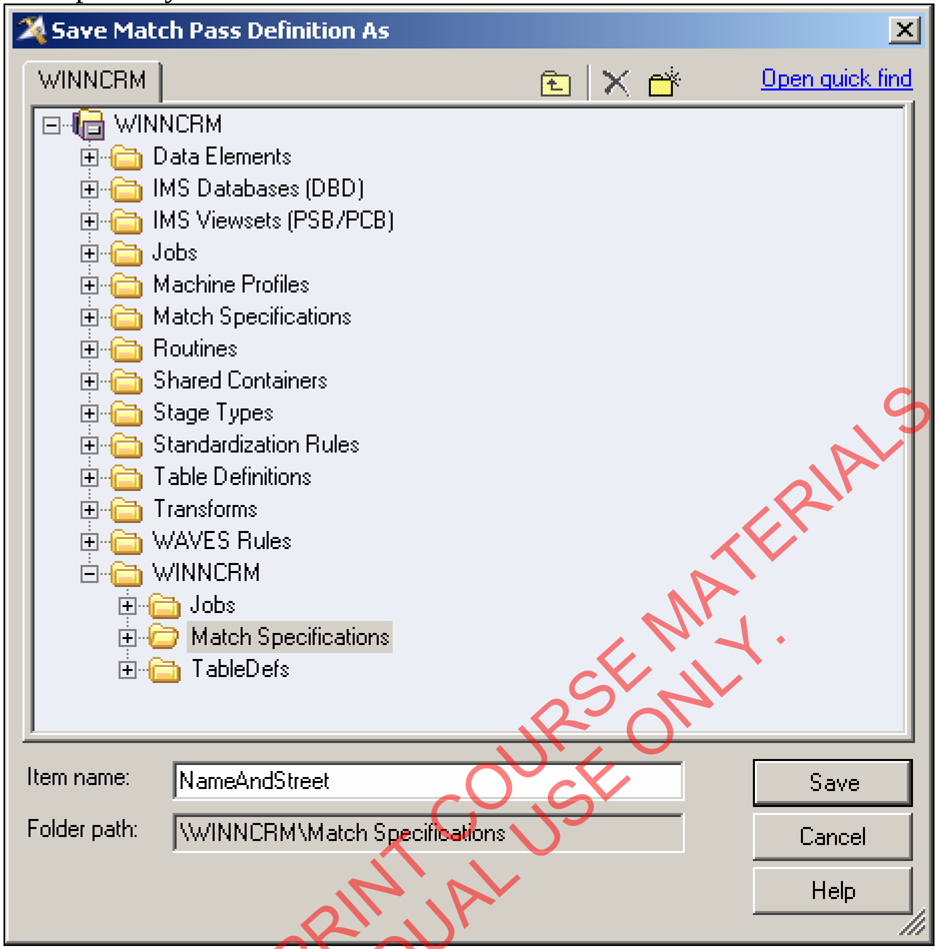


14. Click the Save Pass button under the Pass Definition portion of the Match Designer.



```
MatchPrimaryWord1NYSIIS_USNAME
AddressType_USADDR
StreetNameNYSIIS_USADDR
StanZip3
```

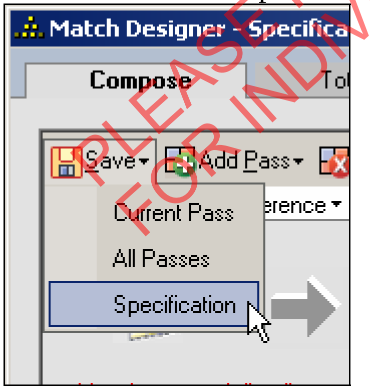
15. Save your NameAndStreet pass into the WINNCRM ➔ Match Specifications folder of the repository.



Item name: NameAndStreet

Folder path: \WINNCRM\Match Specifications

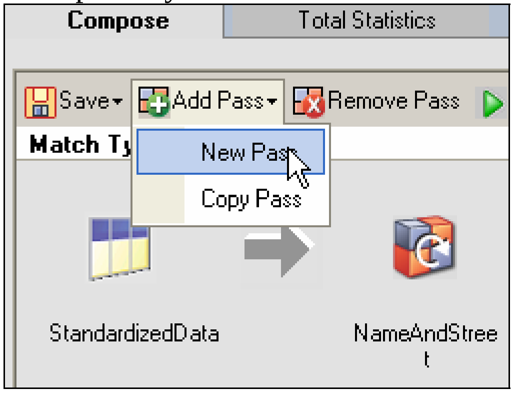
16. Click the Save ➔ Specification button in the top left portion of the Match Designer.



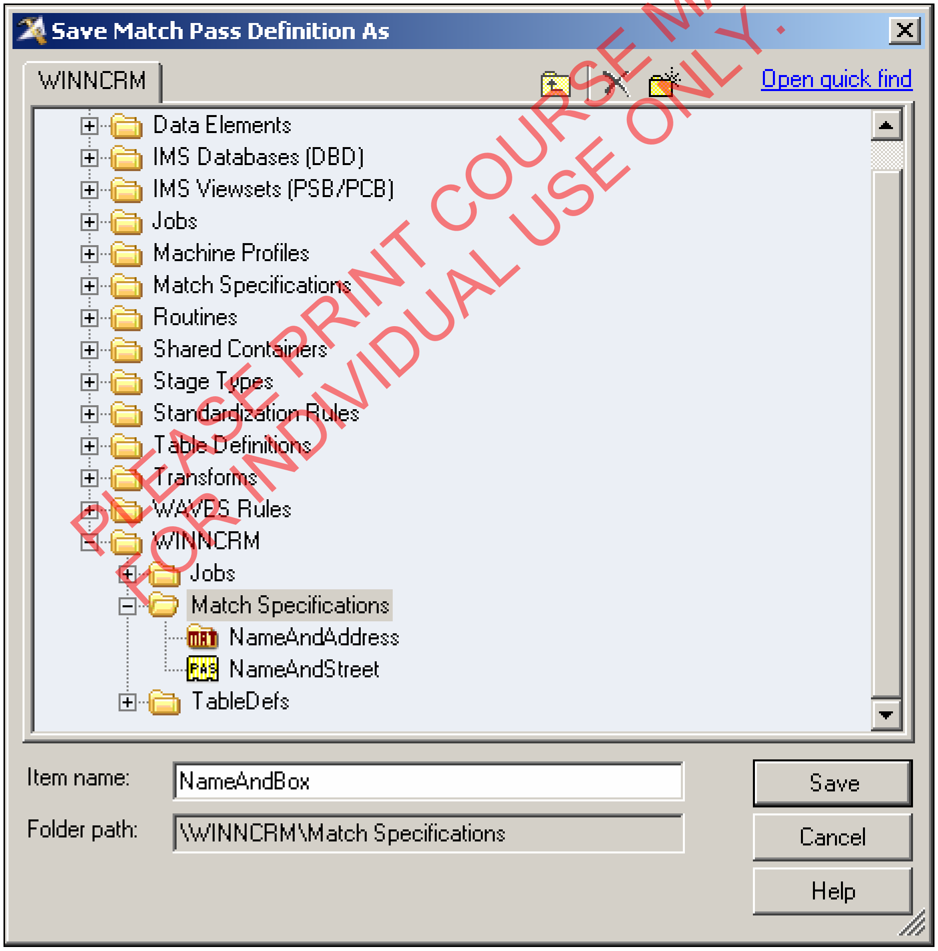
17. Save your Match Specification into the WINNCRM➔Match Specification folder and name it NameAndAddress

Task: Add a second pass

1. Add a second pass by clicking the +Add Pass button immediately under the Compose tab. Select the New Pass option. You will immediately be asked to name this pass – call it NameAndBox and place it in the Passes folder under the Match Specifications folder of the repository.



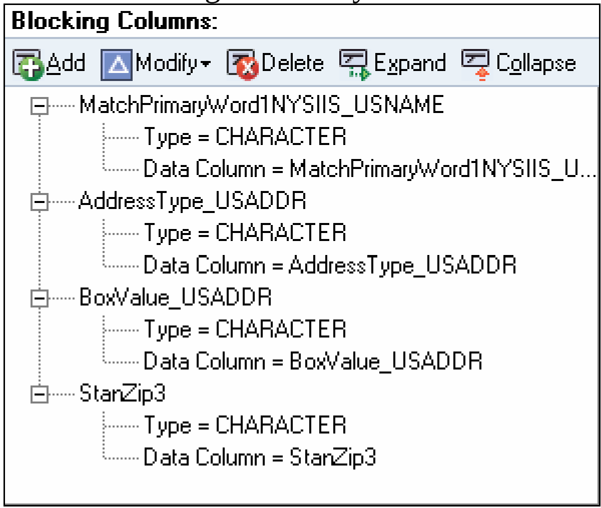
2. Name the new pass NameAndBox and save it to the WINNCRM➔Match Specifications folder.



Item name: NameAndBox

Folder path: \WINNCRM\Match Specifications

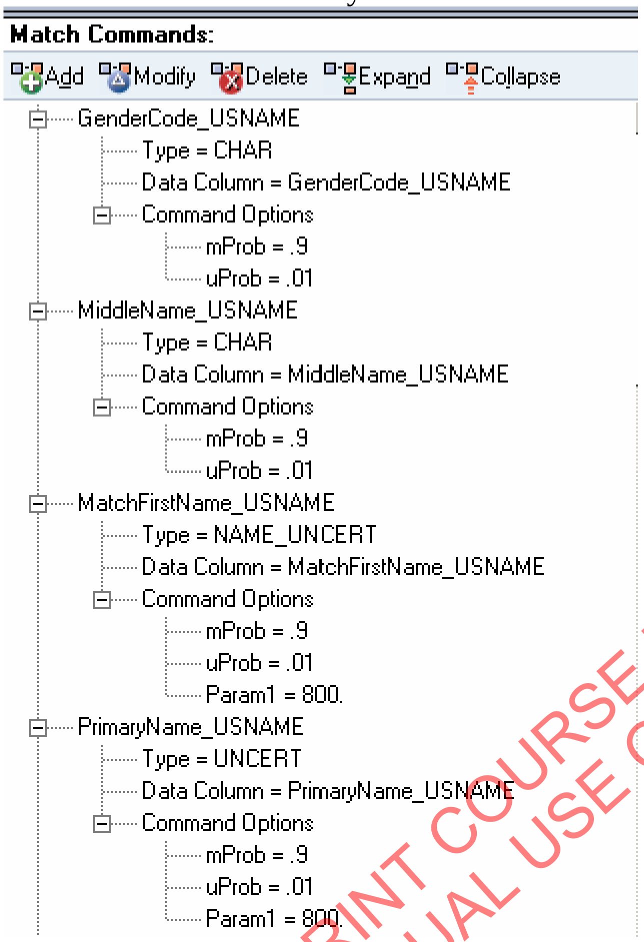
3. Use the same techniques for the second pass utilized in the first; that is:
   a. Add Blocking Columns
   b. Add Match Commands
   c. Provide Cutoff Values
4. The following instructions will provide you with the values for each of these processes.
5. For the Blocking Columns your result should be:



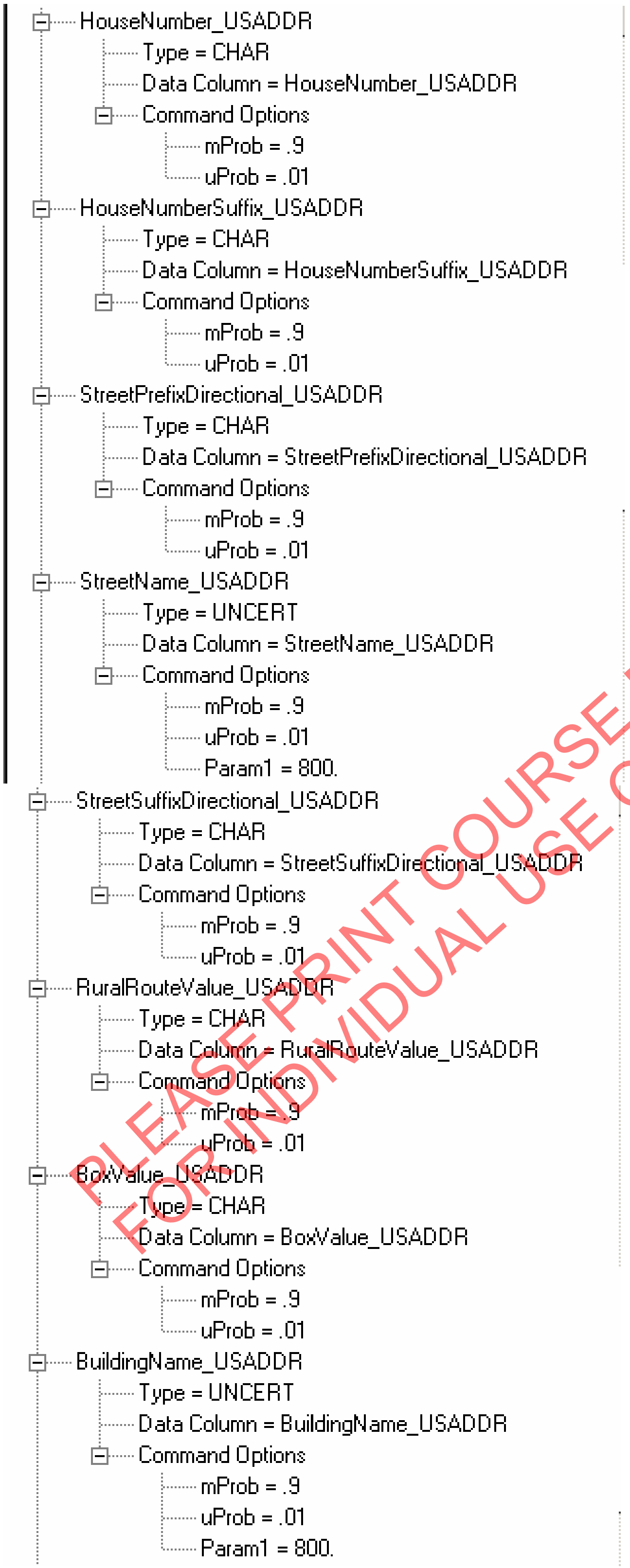
Blocking Columns:

```
⊟ MatchPrimaryWord1NYSIIS_USNAME
    Type = CHARACTER
    Data Column = MatchPrimaryWord1NYSIIS_U...
⊟ AddressType_USADDR
    Type = CHARACTER
    Data Column = AddressType_USADDR
⊟ BoxValue_USADDR
    Type = CHARACTER
    Data Column = BoxValue_USADDR
⊟ StanZip3
    Type = CHARACTER
    Data Column = StanZip3
```

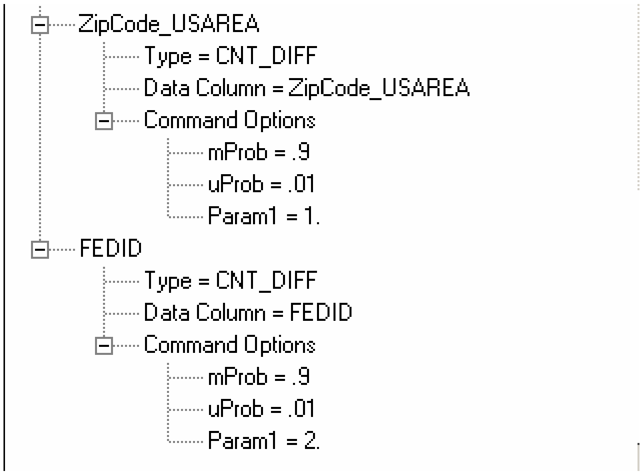
6. For the Match Commands your result should be:



```
GenderCode_USNAME
    Type = CHAR
    Data Column = GenderCode_USNAME
    Command Options
        mProb = .9
        uProb = .01
MiddleName_USNAME
    Type = CHAR
    Data Column = MiddleName_USNAME
    Command Options
        mProb = .9
        uProb = .01
MatchFirstName_USNAME
    Type = NAME_UNCERT
    Data Column = MatchFirstName_USNAME
    Command Options
        mProb = .9
        uProb = .01
        Param1 = 800.
PrimaryName_USNAME
    Type = UNCERT
    Data Column = PrimaryName_USNAME
    Command Options
        mProb = .9
        uProb = .01
        Param1 = 800.
```



```
HouseNumber_USADDR
    Type = CHAR
    Data Column = HouseNumber_USADDR
    Command Options
        mProb = .9
        uProb = .01
HouseNumberSuffix_USADDR
    Type = CHAR
    Data Column = HouseNumberSuffix_USADDR
    Command Options
        mProb = .9
        uProb = .01
StreetPrefixDirectional_USADDR
    Type = CHAR
    Data Column = StreetPrefixDirectional_USADDR
    Command Options
        mProb = .9
        uProb = .01
StreetName_USADDR
    Type = UNCERT
    Data Column = StreetName_USADDR
    Command Options
        mProb = .9
        uProb = .01
        Param1 = 800.
StreetSuffixDirectional_USADDR
    Type = CHAR
    Data Column = StreetSuffixDirectional_USADDR
    Command Options
        mProb = .9
        uProb = .01
RuralRouteValue_USADDR
    Type = CHAR
    Data Column = RuralRouteValue_USADDR
    Command Options
        mProb = .9
        uProb = .01
BoxValue_USADDR
    Type = CHAR
    Data Column = BoxValue_USADDR
    Command Options
        mProb = .9
        uProb = .01
BuildingName_USADDR
    Type = UNCERT
    Data Column = BuildingName_USADDR
    Command Options
        mProb = .9
        uProb = .01
        Param1 = 800.
```

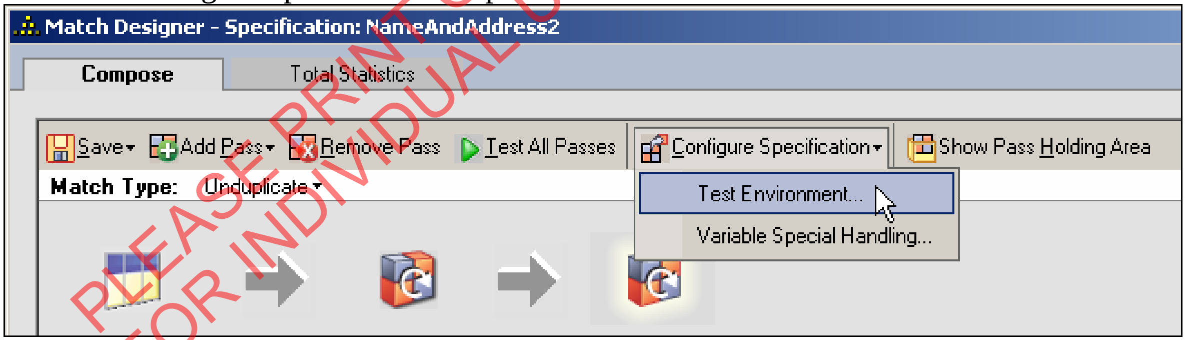


```
ZipCode_USAREA
    Type = CNT_DIFF
    Data Column = ZipCode_USAREA
    Command Options
        mProb = .9
        uProb = .01
        Param1 = 1.
FEDID
    Type = CNT_DIFF
    Data Column = FEDID
    Command Options
        mProb = .9
        uProb = .01
        Param1 = 2.
```

7. Cutoff Values should be: Clerical =0 and Match = 0.
8. Save your second pass as NameAndBox.
9. Click on the Save button immediately below the Compose tab at the top of the Match Designer screen. Select Specification and save your updated specification

Task: Configure a match specification test environment

1. To complete this portion of the exercise you must have a database and ODBC connection available to the Match Designer. The following instructions assume an ODBC connection named QS that connects to an online database.
2. Open the NameAndAddress specification if you are not already in the Match Designer.
3. Click the Configure Specification drop-down menu and select Test Environment.

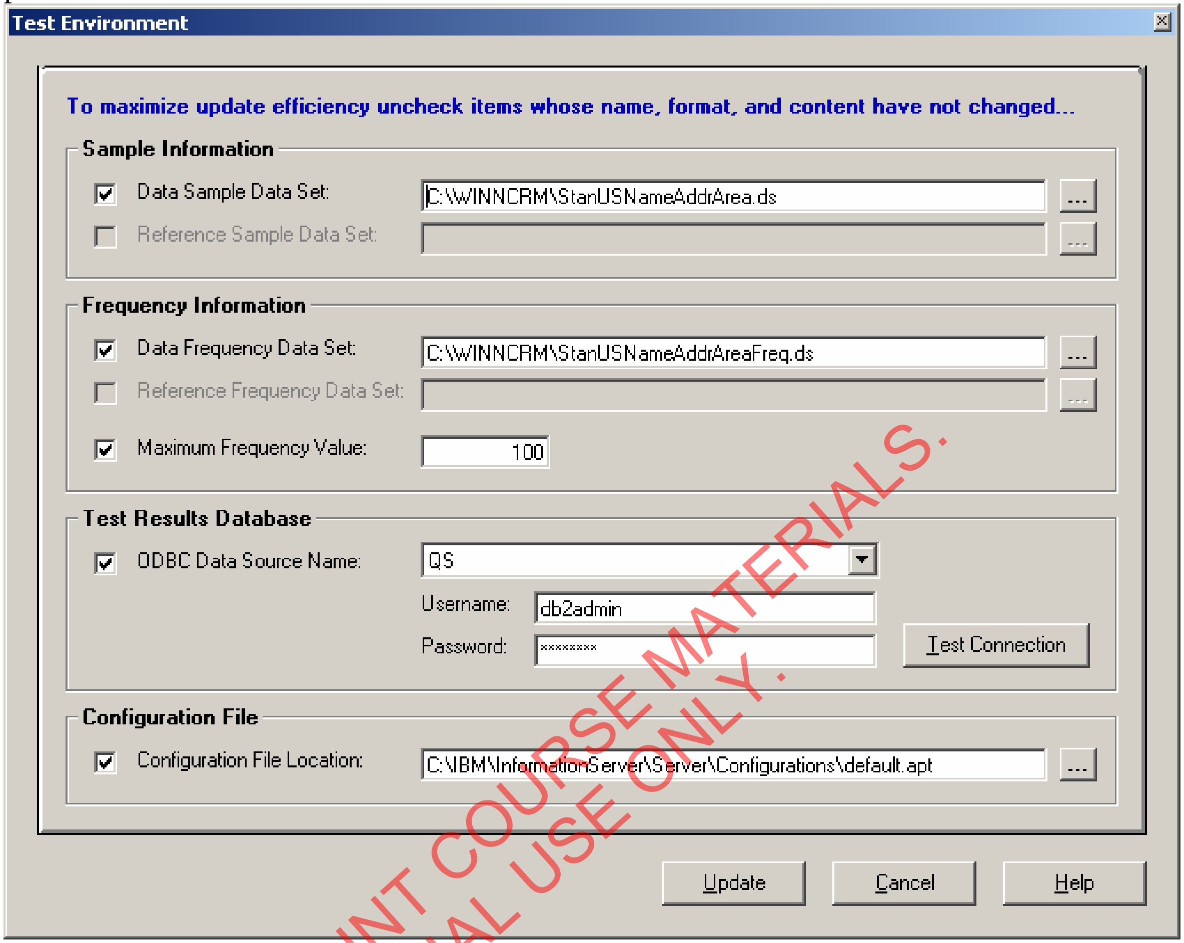


Match Designer - Specification: NameAndAddress2

Match Type: Unduplicate

4. Use the appropriate browse buttons to find your standardized data file (default location is C:\WINNCRM\StanUSNameAddrArea.ds) and its associated frequency report (C:\WINNCRM\StanUSNameAddrAreaFreq.ds).
5. In the Test Results Database section, select an ODBC source (must be predefined) and complete the Username and Password fields. Click the Update button to complete this

process.



Test Environment

To maximize update efficiency uncheck items whose name, format, and content have not changed...

Sample Information

☑ Data Sample Data Set: C:\WINNCRM\StanUSNameAddrArea.ds

☐ Reference Sample Data Set:

Frequency Information

☑ Data Frequency Data Set: C:\WINNCRM\StanUSNameAddrAreaFreq.ds

☐ Reference Frequency Data Set:

☑ Maximum Frequency Value: 100

Test Results Database

☑ ODBC Data Source Name: QS

Username: db2admin

Password: ********

Test Connection

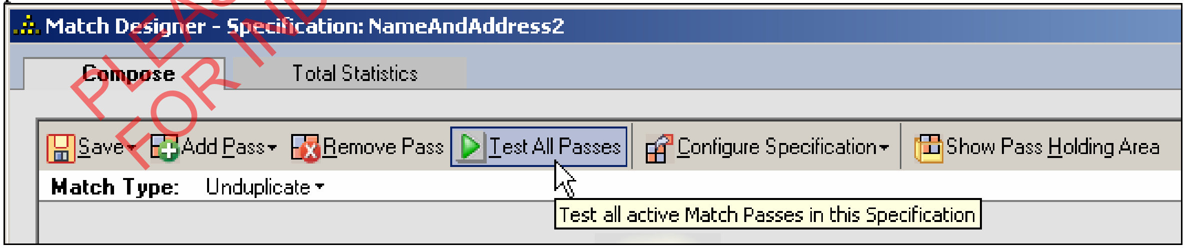
Configuration File

☑ Configuration File Location: C:\IBM\InformationServer\Server\Configurations\default.apt

Update | Cancel | Help

Update will create (but not populate) tables in the test results database and return you to the main screen.

6. Click the Test All Passes button. An hourglass will appear while the records are processed.



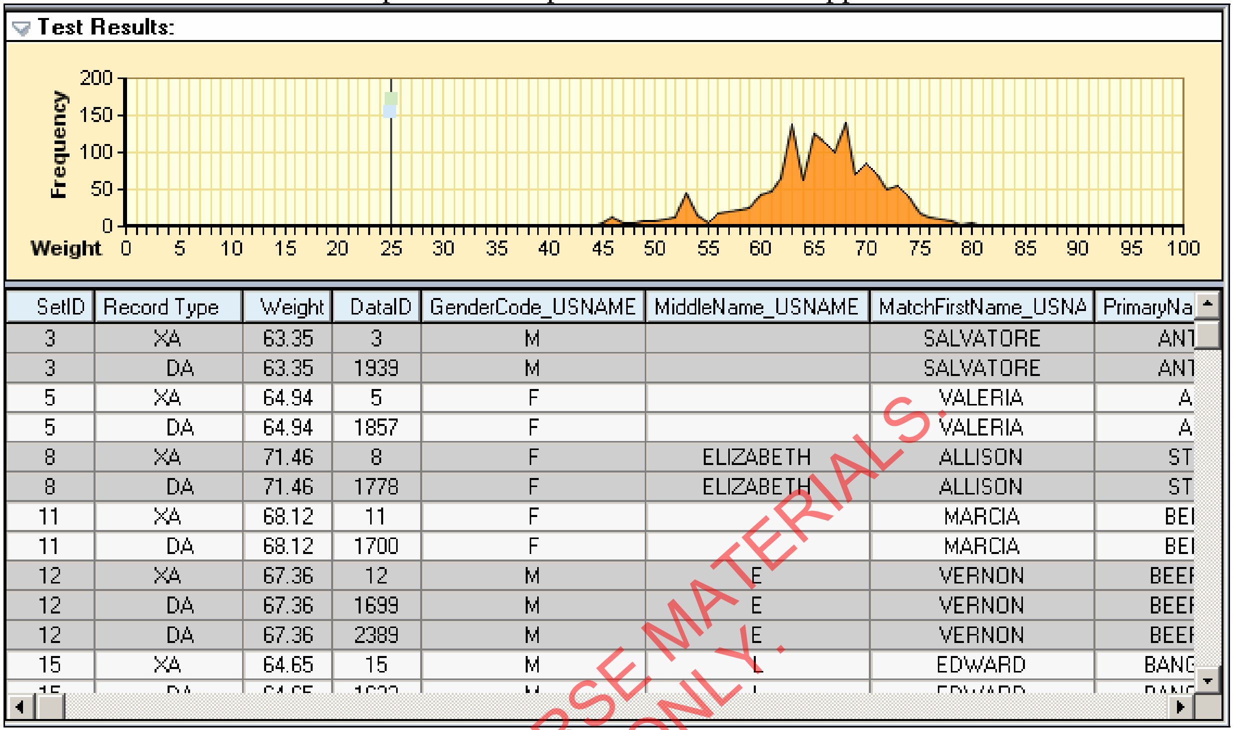
Match Designer - Specification: NameAndAddress2

Match Type: Unduplicate

Test all active Match Passes in this Specification

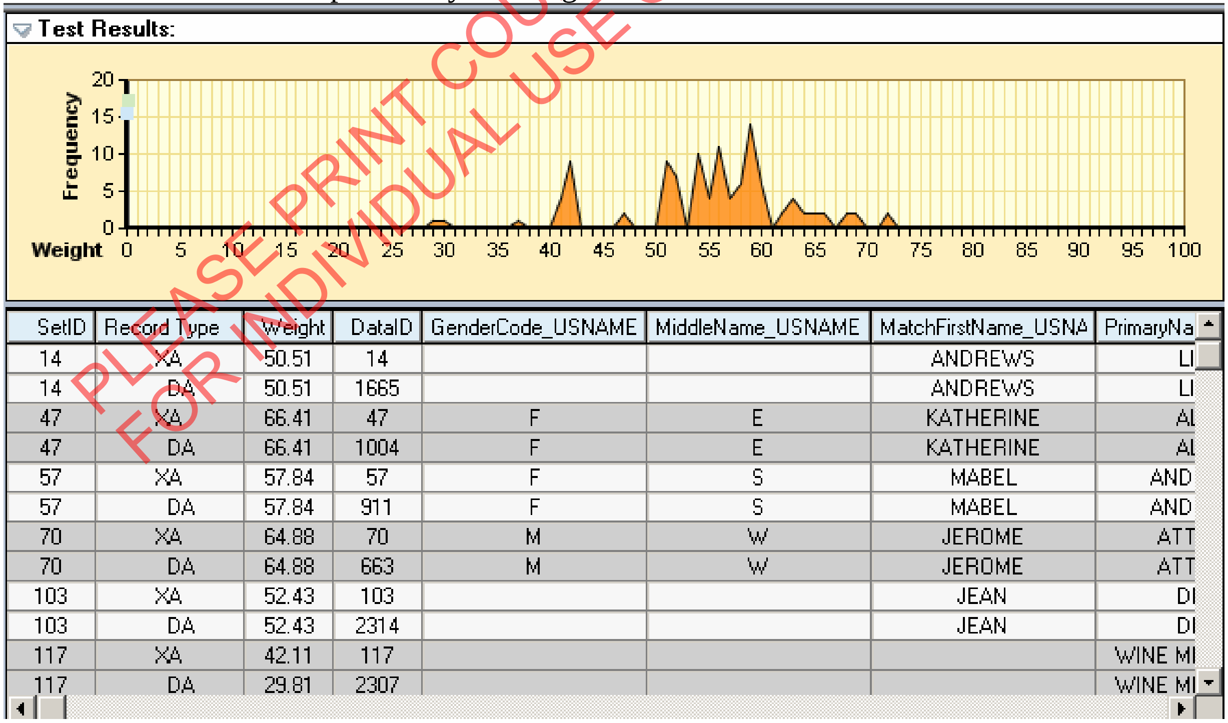
This will populate the statistics tables in your database.

7. Click on the NameAndStreet pass and the pass test results will appear.



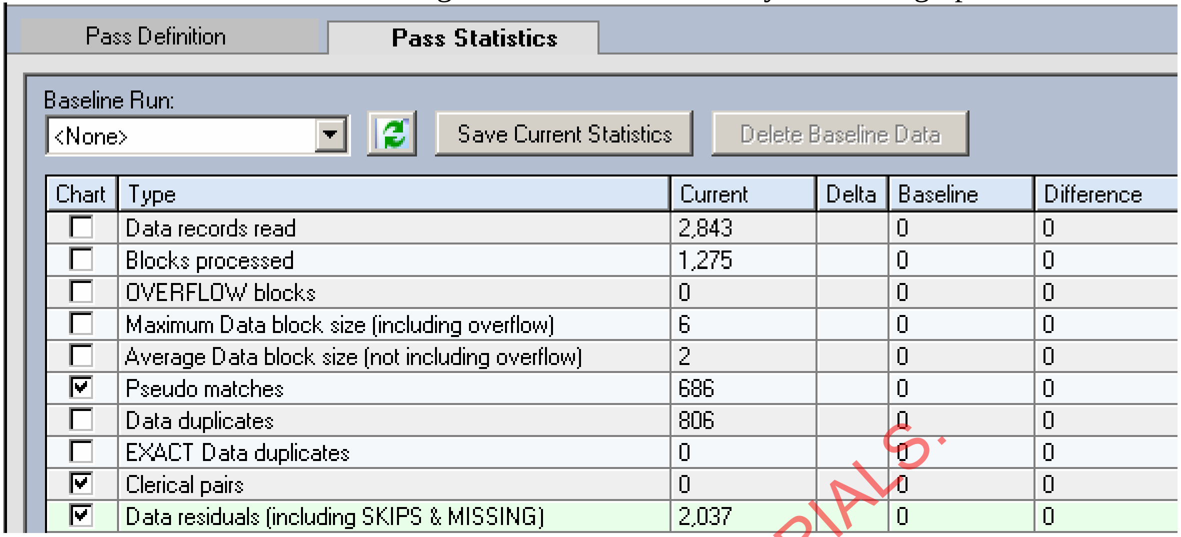
SetID	Record Type	Weight	DataID	GenderCode_USNAME	MiddleName_USNAME	MatchFirstName_USNA	PrimaryNa
3	XA	63.35	3	M		SALVATORE	ANT
3	DA	63.35	1939	M		SALVATORE	ANT
5	XA	64.94	5	F		VALERIA	A
5	DA	64.94	1857	F		VALERIA	A
8	XA	71.46	8	F	ELIZABETH	ALLISON	ST
8	DA	71.46	1778	F	ELIZABETH	ALLISON	ST
11	XA	68.12	11	F		MARCIA	BEI
11	DA	68.12	1700	F		MARCIA	BEI
12	XA	67.36	12	M	E	VERNON	BEEI
12	DA	67.36	1699	M	E	VERNON	BEEI
12	DA	67.36	2389	M	E	VERNON	BEEI
15	XA	64.65	15	M	L	EDWARD	BANG
15	DA	64.65	[illegible]	M	L	EDWARD	BANG

8. Select the NameAndBox pass and you will get a different set of results.



SetID	Record Type	Weight	DataID	GenderCode_USNAME	MiddleName_USNAME	MatchFirstName_USNA	PrimaryNa
14	XA	50.51	14			ANDREWS	LI
14	DA	50.51	1665			ANDREWS	LI
47	XA	66.41	47	F	E	KATHERINE	AI
47	DA	66.41	1004	F	E	KATHERINE	AI
57	XA	57.84	57	F	S	MABEL	AND
57	DA	57.84	911	F	S	MABEL	AND
70	XA	64.88	70	M	W	JEROME	ATT
70	DA	64.88	663	M	W	JEROME	ATT
103	XA	52.43	103			JEAN	DI
103	DA	52.43	2314			JEAN	DI
117	XA	42.11	117				WINE MI
117	DA	29.81	2307				WINE MI

9. Click on the Pass Statistics tab to get both a record summary and chart graphic:



Baseline Run: <None>

Chart	Type	Current	Delta	Baseline	Difference
□	Data records read	2,843		0	0
□	Blocks processed	1,275		0	0
□	OVERFLOW blocks	0		0	0
□	Maximum Data block size (including overflow)	6		0	0
□	Average Data block size (not including overflow)	2		0	0
☑	Pseudo matches	686		0	0
□	Data duplicates	806		0	0
□	EXACT Data duplicates	0		0	0
☑	Clerical pairs	0		0	0
☑	Data residuals (including SKIPS & MISSING)	2,037		0	0

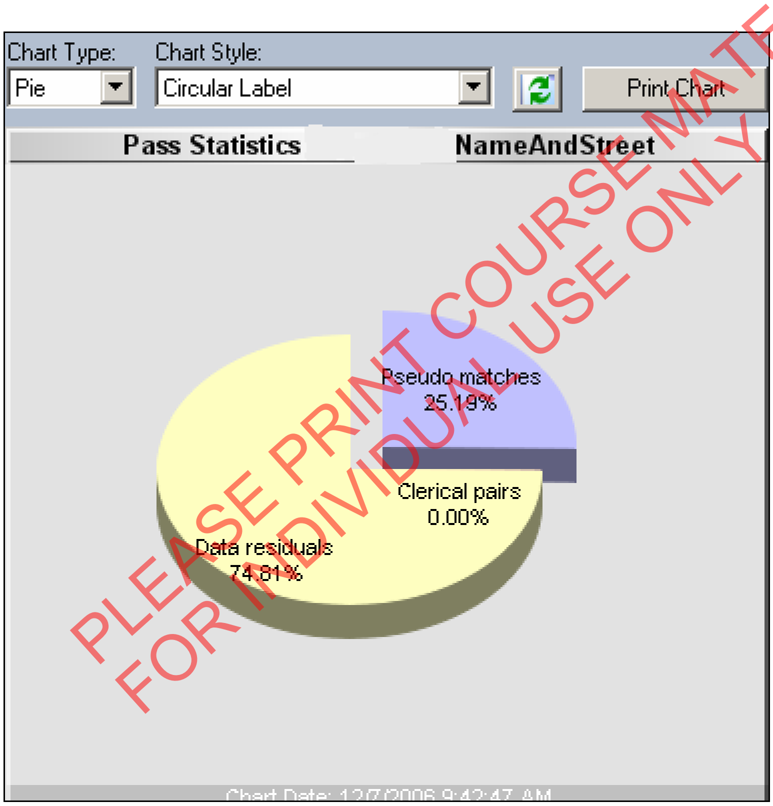
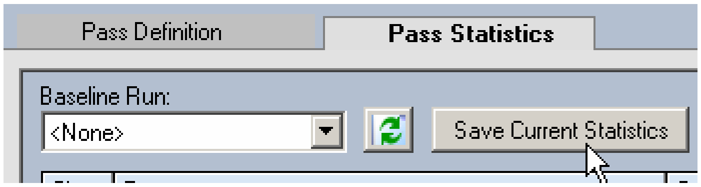


Chart Type: Pie

Chart Style: Circular Label

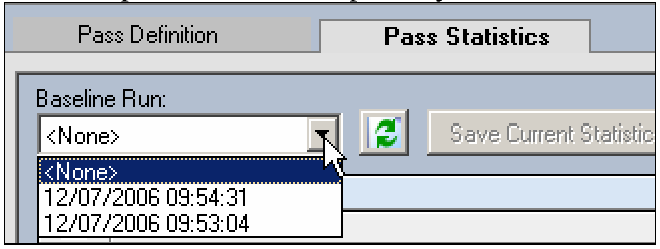
10. Note that you can control the Chart type and Chart style to produce a variety of visual representations of the match pass simulations. When you change options use the refresh chart button to update the image.

11. Click the Save Current Statistics button to create an historical record of this pass.



Baseline Run: <None>

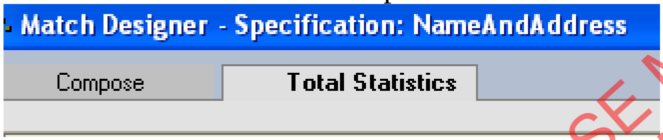
12. Click the drop-down arrow under Baseline Run to see your historical statistics records. As more pass runs are completed you will see further historical records.



Baseline Run: <None>

<None>
12/07/2006 09:54:31
12/07/2006 09:53:04

13. You can also view pass results for the entire run – including both passes. Click on the Total Statistics tab new the top of the screen.



14. Expand the pass results on the left portion of the screen. Note that both passes are represented in the chart; you can also change Chart Type and Chart Style.

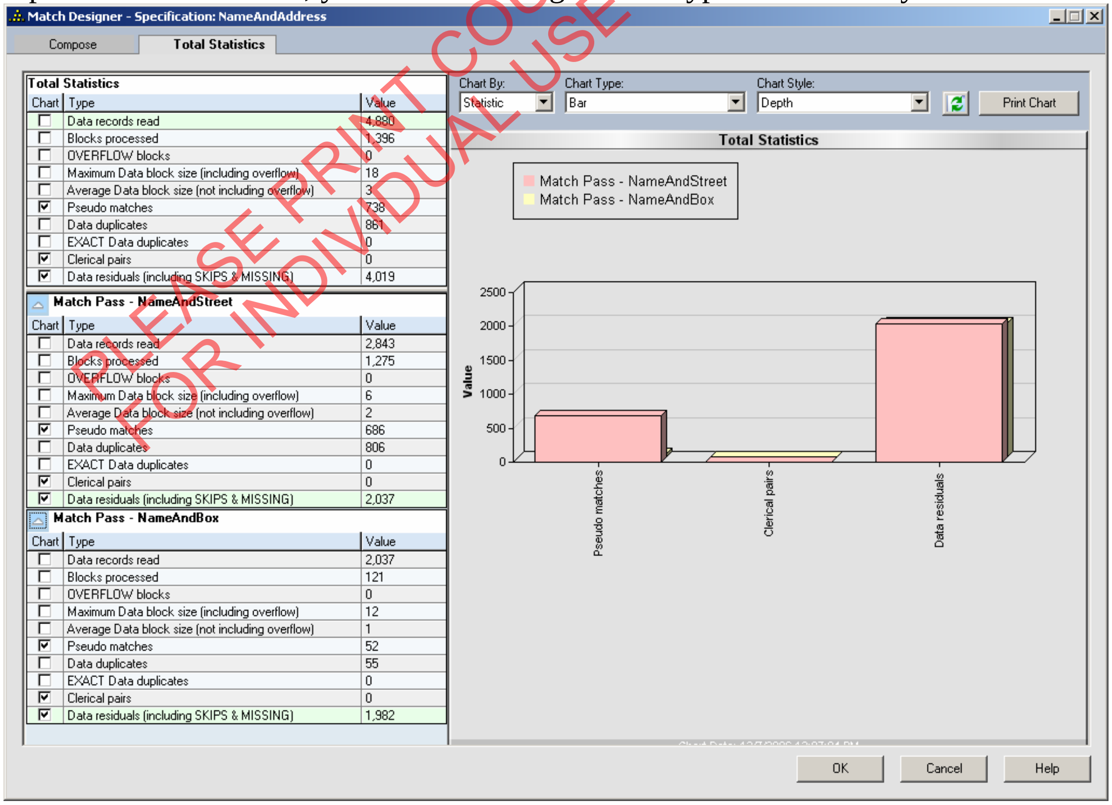


Chart By: Statistic
Chart Type: Bar
Chart Style: Depth

Total Statistics

Chart	Type	Value
□	Data records read	4,880
□	Blocks processed	1,396
□	OVERFLOW blocks	0
□	Maximum Data block size (including overflow)	18
□	Average Data block size (not including overflow)	3
☑	Pseudo matches	738
□	Data duplicates	861
□	EXACT Data duplicates	0
☑	Clerical pairs	0
☑	Data residuals (including SKIPS & MISSING)	4,019

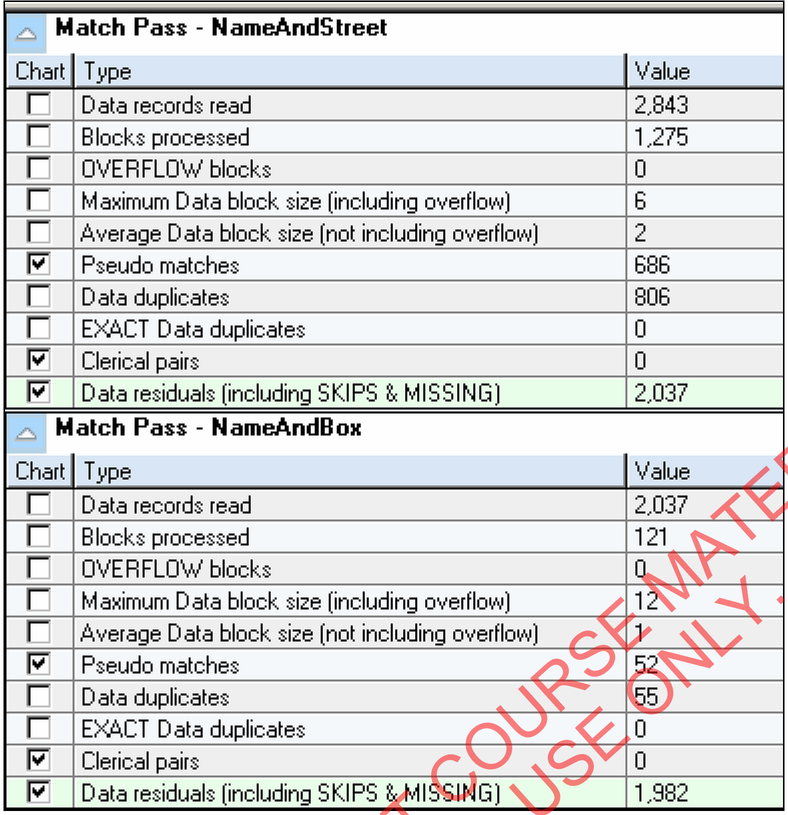
Match Pass - NameAndStreet

Chart	Type	Value
□	Data records read	2,843
□	Blocks processed	1,275
□	OVERFLOW blocks	0
□	Maximum Data block size (including overflow)	6
□	Average Data block size (not including overflow)	2
☑	Pseudo matches	686
□	Data duplicates	806
□	EXACT Data duplicates	0
☑	Clerical pairs	0
☑	Data residuals (including SKIPS & MISSING)	2,037

Match Pass - NameAndBox

Chart	Type	Value
□	Data records read	2,037
□	Blocks processed	121
□	OVERFLOW blocks	0
□	Maximum Data block size (including overflow)	12
□	Average Data block size (not including overflow)	1
☑	Pseudo matches	52
□	Data duplicates	55
□	EXACT Data duplicates	0
☑	Clerical pairs	0
☑	Data residuals (including SKIPS & MISSING)	1,982

15. Examine the counts in the two match passes. This is an Unduplicate dependent pass, meaning that the second pass receives the residuals from the first pass. An Unduplicate independent pass, on the other hand, would run both passes against the full set of records.



Match Pass - NameAndStreet

Chart	Type	Value
☐	Data records read	2,843
☐	Blocks processed	1,275
☐	OVERFLOW blocks	0
☐	Maximum Data block size (including overflow)	6
☐	Average Data block size (not including overflow)	2
☑	Pseudo matches	686
☐	Data duplicates	806
☐	EXACT Data duplicates	0
☑	Clerical pairs	0
☑	Data residuals (including SKIPS & MISSING)	2,037

Match Pass - NameAndBox

Chart	Type	Value
☐	Data records read	2,037
☐	Blocks processed	121
☐	OVERFLOW blocks	0
☐	Maximum Data block size (including overflow)	12
☐	Average Data block size (not including overflow)	1
☑	Pseudo matches	52
☐	Data duplicates	55
☐	EXACT Data duplicates	0
☑	Clerical pairs	0
☑	Data residuals (including SKIPS & MISSING)	1,982

Lab 20: Unduplicate

Objectives:

- Identify duplicates in a single file

Assumptions:

- A two pass Match Specification named NameAndAddress exists
- Job StanAndGenFreq ran successfully

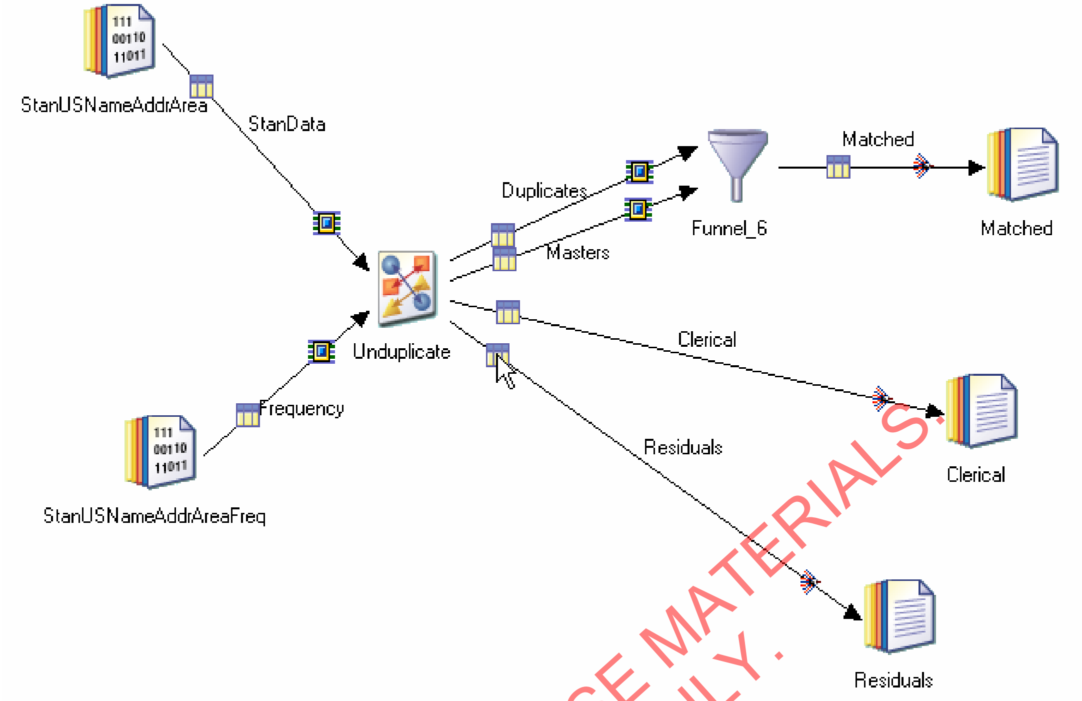
Task: Build and run an Unduplicate job

Using Data Sets
The job created in this exercise will use data sets, one containing standardized data and one containing a match frequency report; these data sets were created in a previous job (StanAndGenFreq). Data files used as intermediate staging points between jobs are usually stored as data sets instead of sequential files; this technique increases parallel job performance.

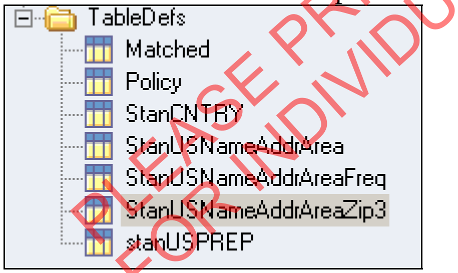
Match Specification
In addition to the stages shown in the job below, your design will use a match specification; this specification gives your job information about blocking and cutoff criteria.

1. Create a new parallel job using Data Set, Unduplicate Match, Funnel, and Sequential File stages. Rename stages and links as shown – the link names are especially important in

this job as you will see when you edit the Unduplicate Match stage.

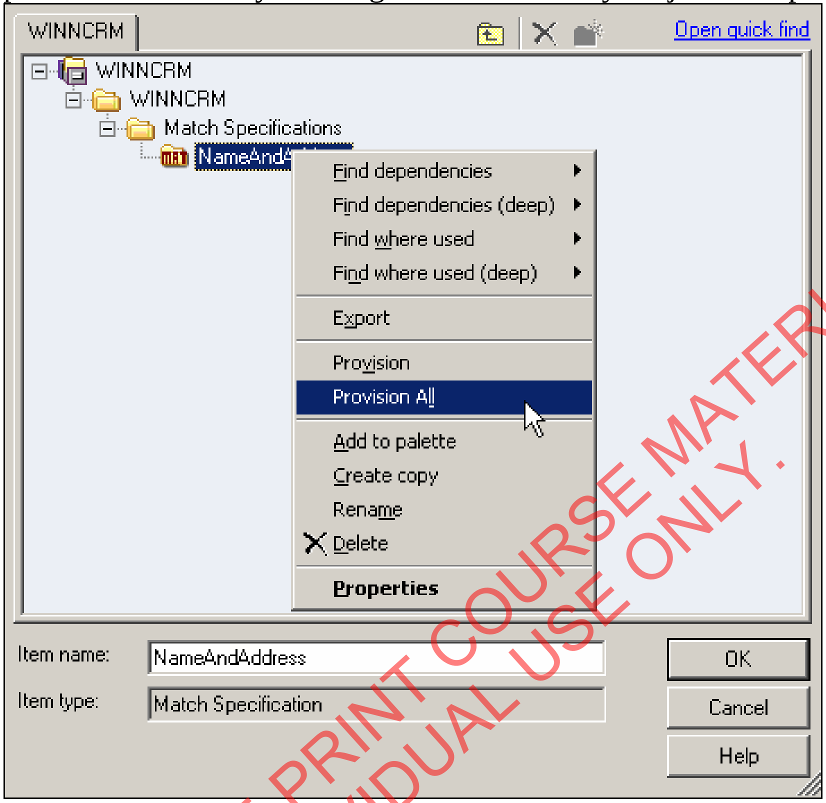


2. The input data sets in this job should already exist and were created by the StanAndGenFreq job.
3. Open the StanUSNameAddrArea Data Set stage. Set the File property to C:\WINNCRM\StanUSNameAddrArea.ds.
4. Click the Column tab and import the column definitions from the StanUSNameAddrAreaZip3 table.



5. Use the View Data button to validate your settings and then click the Ok button to return to the job design canvas.
6. Open the StanUSNameAddrAreaFreq Data Set stage. Set the File property to C:\WINNCRM\StanUSNameAddrAreaFreq.ds.
7. Import column definitions from the StanUSNameAddrAreaFreq table.
8. Use the view data button to validate your settings and then click the Ok button to return to the job design canvas.
9. Open the Unduplicate stage.
10. Select all radio buttons under the Match Outputs section.

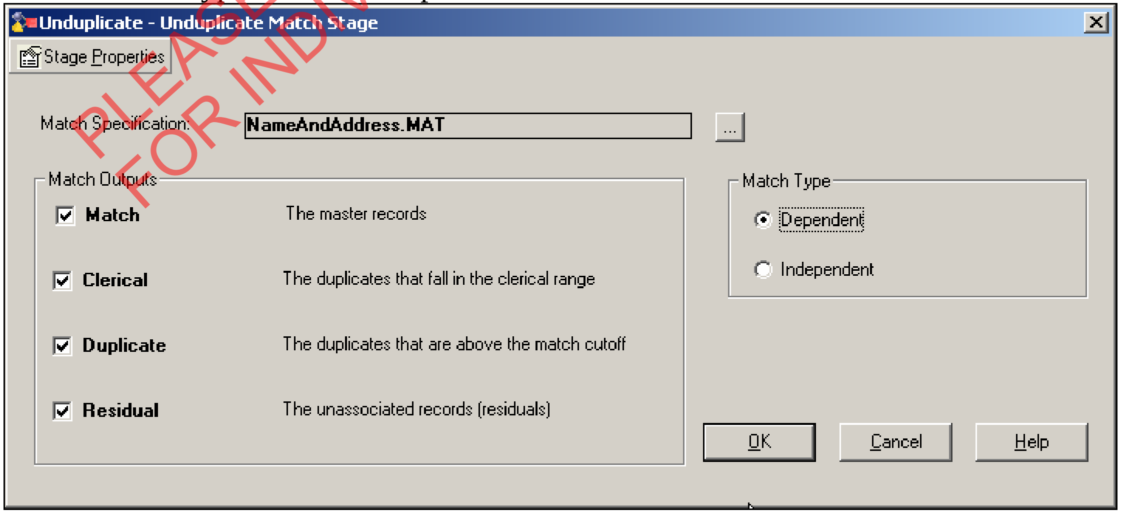
11. Use the Match Specification browse (….) button to locate the NameAndAddress match specification. You placed it earlier in the WINNCRM➔Match Specification folder in the repository.
12. When you find the match specification, right-click on it and select the Provision All option. This will copy the match specification to the job execution area. If you do not perform this action you will get an error when your job attempts to execute.



Item name: NameAndAddress

Item type: Match Specification

13. Under Match Type select the Dependent radio button.



Match Specification: NameAndAddress.MAT

Match Outputs

☑ Match — The master records

☑ Clerical — The duplicates that fall in the clerical range

☑ Duplicate — The duplicates that are above the match cutoff

☑ Residual — The unassociated records (residuals)

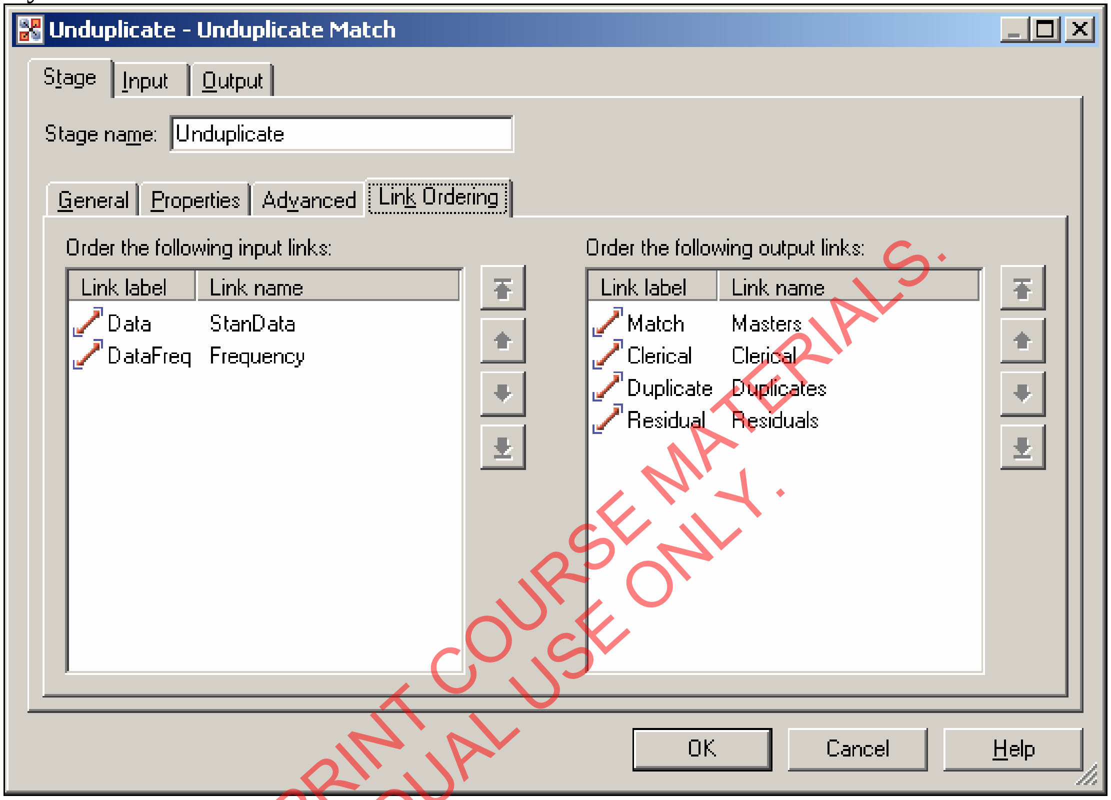
Match Type

● Dependent

○ Independent

The Match Outputs checkboxes direct this stage to produce four output links. Recall that you created four output links and gave them meaningful names. The next step in this task will associate each of the Match Outputs to the corresponding link.

14. Click on the Stage Properties tab and then the Link Ordering tab. Your screen may look different from the one below depending on the order you built the links but the general layout should be the same.



Stage name: Unduplicate

Order the following input links:

Link label	Link name
Data	StanData
DataFreq	Frequency

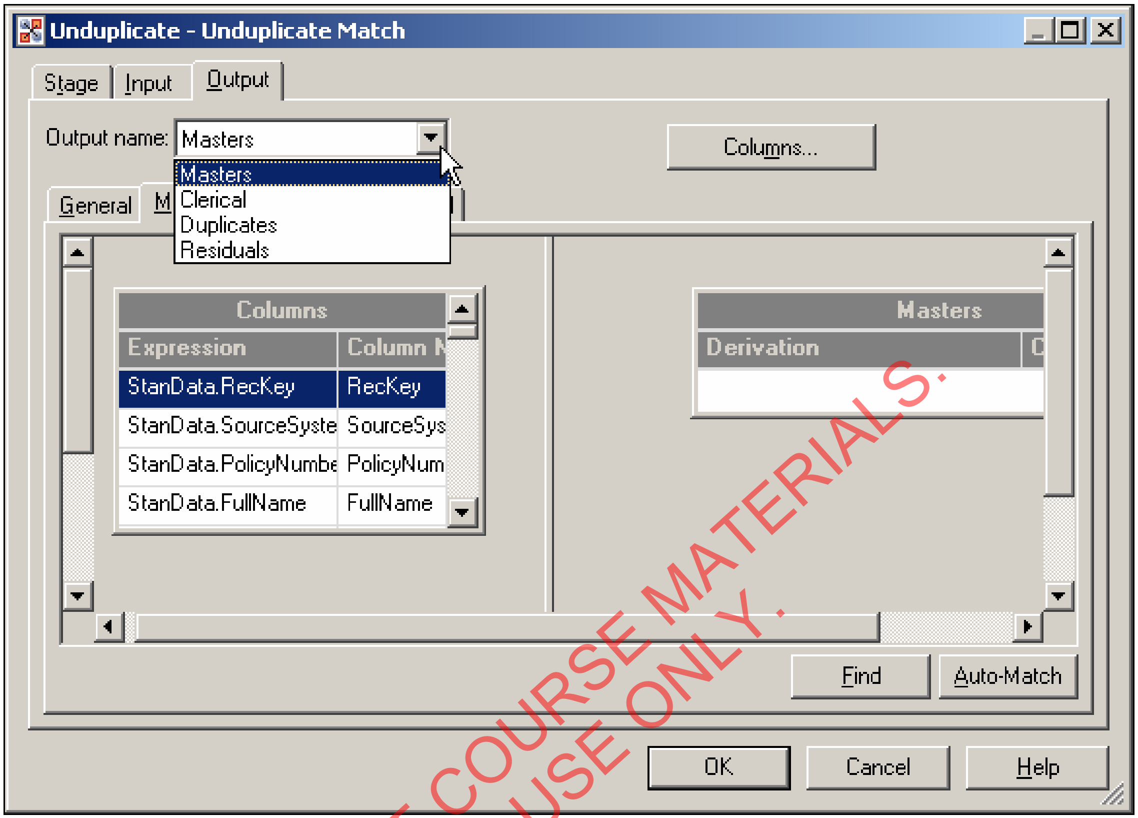
Order the following output links:

Link label	Link name
Match	Masters
Clerical	Clerical
Duplicate	Duplicates
Residual	Residuals

Links coming into the stage are on the left half of the screen, outgoing links are represented on the right half. Note that each half is divided into two columns – Link label and Link name. Recall that the Investigate stage works in a similar fashion.

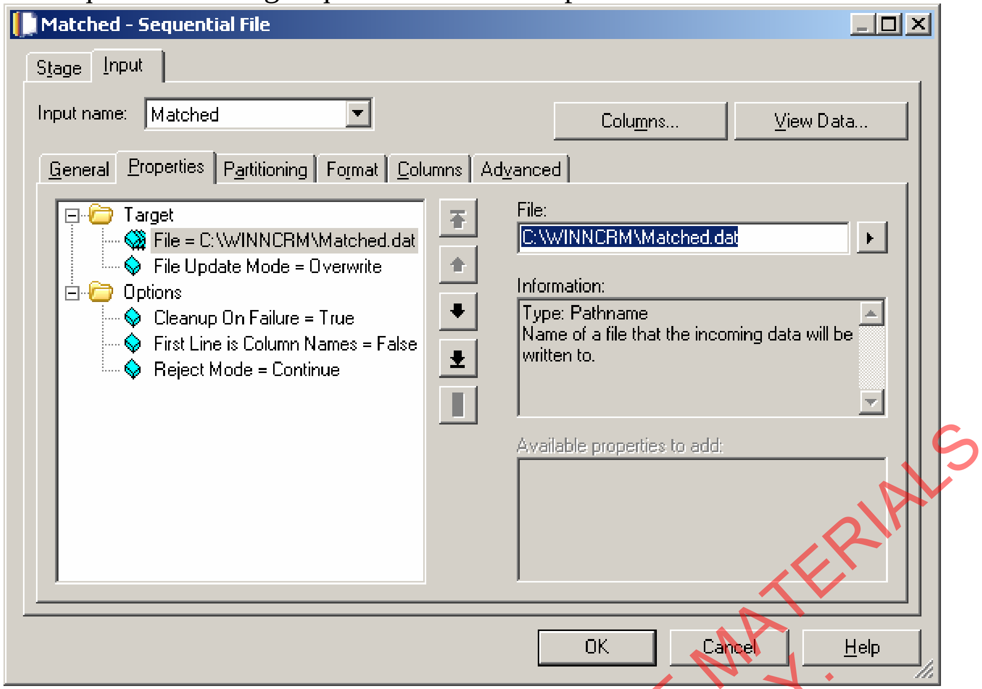
15. Use the appropriate arrows to match the proper Link label with the Link name you assigned when the job was initially created.

16. Click the Output➔Mapping tabs and then click the drop down box to view the output links.



17. For each link, map all input columns to the output.
18. When finished with mappings, click the Ok buttons until you return to the job design canvas.
19. Right-click on the Funnel stage and Auto-map input columns ➔ All output links.
20. Edit the Matched Sequential File stage and set the File property C:\WINNCRM\Matched.dat.

You also need to go to the format tab for each sequential file and set the Null field value to NULL since many of the records will contain a null in some of the fields. Recall this is a consequence of using sequential files as output.



Input name: Matched

- Target
  - File = C:\WINNCRM\Matched.dat
  - File Update Mode = Overwrite
- Options
  - Cleanup On Failure = True
  - First Line is Column Names = False
  - Reject Mode = Continue

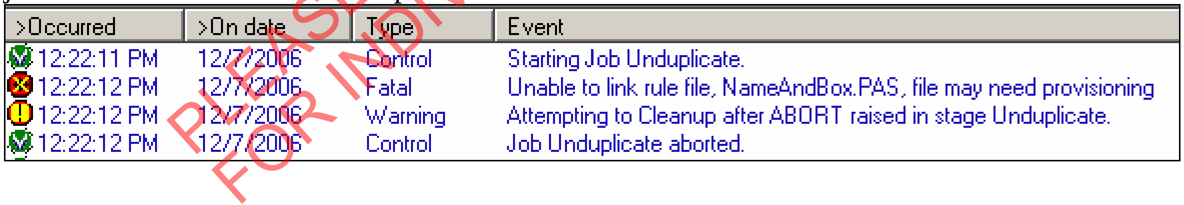
File: C:\WINNCRM\Matched.dat

Information:
Type: Pathname
Name of a file that the incoming data will be written to.

Available properties to add:

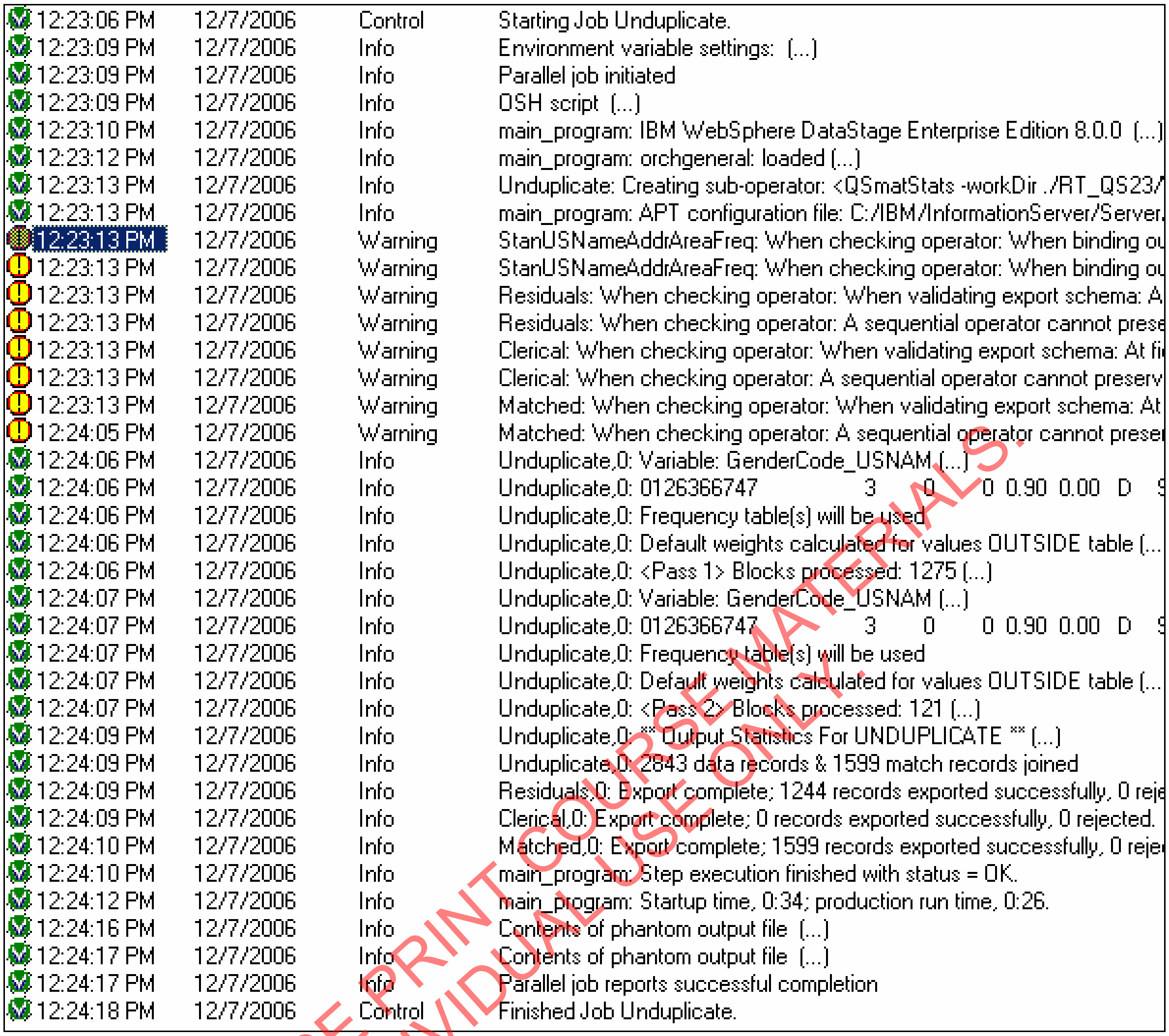
21. Open the Clerical Sequential File stage and set the File property to C:\WINNCRM\Clerical.dat.
22. Open the Residuals Sequential File stage and set the File property to C:\WINNCRM\Residuals.dat.
23. Save your job as Unduplicate.
24. Compile and run.

Job Log follows:
Did you get this error? It means that you need to provision the match specification and reset the job. You do not need to recompile.



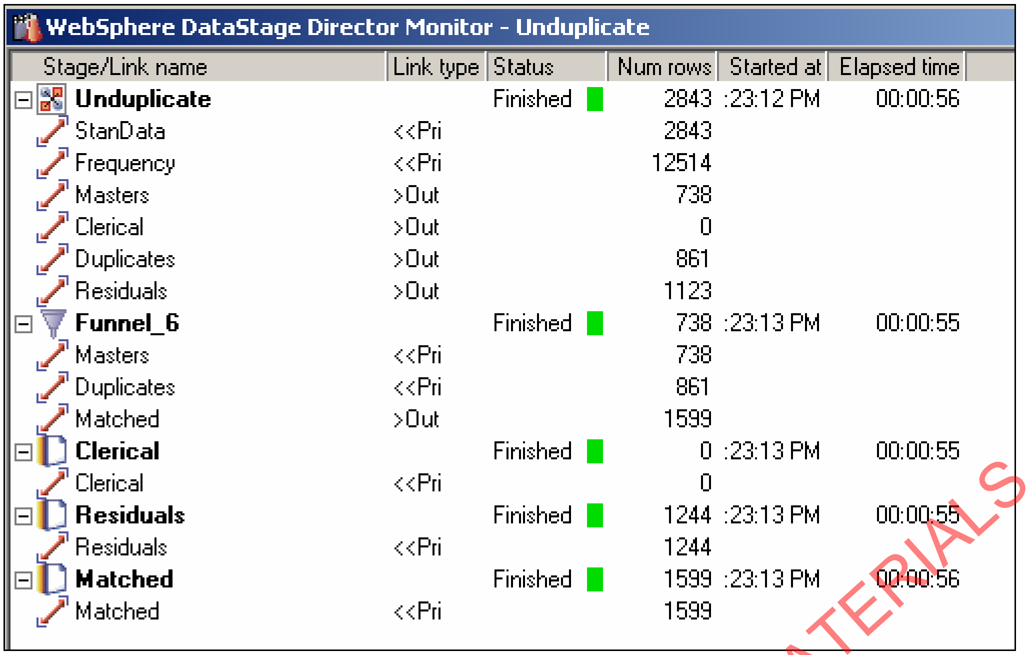
>Occurred	>On date	Type	Event
12:22:11 PM	12/7/2006	Control	Starting Job Unduplicate.
12:22:12 PM	12/7/2006	Fatal	Unable to link rule file, NameAndBox.PAS, file may need provisioning
12:22:12 PM	12/7/2006	Warning	Attempting to Cleanup after ABORT raised in stage Unduplicate.
12:22:12 PM	12/7/2006	Control	Job Unduplicate aborted.

A successful run looks like the following. Note the messages for each pass.



```
12:23:06 PM   12/7/2006   Control   Starting Job Unduplicate.
12:23:09 PM   12/7/2006   Info      Environment variable settings:  (...)
12:23:09 PM   12/7/2006   Info      Parallel job initiated
12:23:09 PM   12/7/2006   Info      OSH script  (...)
12:23:10 PM   12/7/2006   Info      main_program: IBM WebSphere DataStage Enterprise Edition 8.0.0  (...)
12:23:12 PM   12/7/2006   Info      main_program: orchgeneral: loaded (...)
12:23:13 PM   12/7/2006   Info      Unduplicate: Creating sub-operator: <QSmatStats -workDir ./RT_QS23/
12:23:13 PM   12/7/2006   Info      main_program: APT configuration file: C:/IBM/InformationServer/Server/
12:23:13 PM   12/7/2006   Warning   StanUSNameAddrAreaFreq: When checking operator: When binding ou
12:23:13 PM   12/7/2006   Warning   StanUSNameAddrAreaFreq: When checking operator: When binding ou
12:23:13 PM   12/7/2006   Warning   Residuals: When checking operator: When validating export schema: A
12:23:13 PM   12/7/2006   Warning   Residuals: When checking operator: A sequential operator cannot prese
12:23:13 PM   12/7/2006   Warning   Clerical: When checking operator: When validating export schema: At fi
12:23:13 PM   12/7/2006   Warning   Clerical: When checking operator: A sequential operator cannot preserv
12:23:13 PM   12/7/2006   Warning   Matched: When checking operator: When validating export schema: At
12:24:05 PM   12/7/2006   Warning   Matched: When checking operator: A sequential operator cannot prese
12:24:06 PM   12/7/2006   Info      Unduplicate,0: Variable: GenderCode_USNAM (...)
12:24:06 PM   12/7/2006   Info      Unduplicate,0: 0126366747            3      0       0  0.90  0.00   D    9
12:24:06 PM   12/7/2006   Info      Unduplicate,0: Frequency table(s) will be used
12:24:06 PM   12/7/2006   Info      Unduplicate,0: Default weights calculated for values OUTSIDE table (...
12:24:06 PM   12/7/2006   Info      Unduplicate,0: <Pass 1> Blocks processed: 1275 (...)
12:24:07 PM   12/7/2006   Info      Unduplicate,0: Variable: GenderCode_USNAM (...)
12:24:07 PM   12/7/2006   Info      Unduplicate,0: 0126366747            3      0       0  0.90  0.00   D    9
12:24:07 PM   12/7/2006   Info      Unduplicate,0: Frequency table(s) will be used
12:24:07 PM   12/7/2006   Info      Unduplicate,0: Default weights calculated for values OUTSIDE table (...
12:24:07 PM   12/7/2006   Info      Unduplicate,0: <Pass 2> Blocks processed: 121 (...)
12:24:09 PM   12/7/2006   Info      Unduplicate,0: ** Output Statistics For UNDUPLICATE ** (...)
12:24:09 PM   12/7/2006   Info      Unduplicate,0: 2843 data records & 1599 match records joined
12:24:09 PM   12/7/2006   Info      Residuals,0: Export complete; 1244 records exported successfully, 0 reje
12:24:09 PM   12/7/2006   Info      Clerical,0: Export complete; 0 records exported successfully, 0 rejected.
12:24:10 PM   12/7/2006   Info      Matched,0: Export complete; 1599 records exported successfully, 0 reje
12:24:10 PM   12/7/2006   Info      main_program: Step execution finished with status = OK.
12:24:12 PM   12/7/2006   Info      main_program: Startup time, 0:34; production run time, 0:26.
12:24:16 PM   12/7/2006   Info      Contents of phantom output file  (...)
12:24:17 PM   12/7/2006   Info      Contents of phantom output file  (...)
12:24:17 PM   12/7/2006   Info      Parallel job reports successful completion
12:24:18 PM   12/7/2006   Control   Finished Job Unduplicate.
```

Record count from job monitor:



WebSphere DataStage Director Monitor - Unduplicate

Stage/Link name	Link type	Status	Num rows	Started at	Elapsed time
⊟ **Unduplicate**		Finished 🟩	2843	:23:12 PM	00:00:56
StanData	<<Pri		2843		
Frequency	<<Pri		12514		
Masters	>Out		738		
Clerical	>Out		0		
Duplicates	>Out		861		
Residuals	>Out		1123		
⊟ **Funnel_6**		Finished 🟩	738	:23:13 PM	00:00:55
Masters	<<Pri		738		
Duplicates	<<Pri		861		
Matched	>Out		1599		
⊟ **Clerical**		Finished 🟩	0	:23:13 PM	00:00:55
Clerical	<<Pri		0		
⊟ **Residuals**		Finished 🟩	1244	:23:13 PM	00:00:55
Residuals	<<Pri		1244		
⊟ **Matched**		Finished 🟩	1599	:23:13 PM	00:00:56
Matched	<<Pri		1599		

25. Use the appropriate View Data button to review the data in each of the output data files. Note these are sequential instead of data sets because they may be used in other non-DataStage processes.
26. Edit the Matched Sequential File stage and click on the columns tab.
27. Save the column definitions to the repository in the saved folder – name the table definition Matched. These columns will be used in a subsequent job.

Lab 21: Survivorship

Objectives:

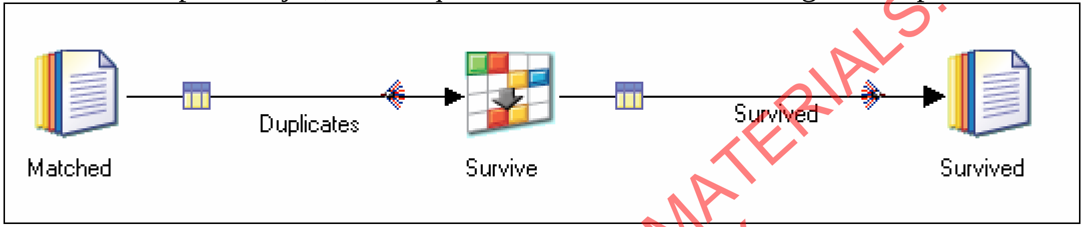
- Survive a Policy record that contains “best of breed” data values

Assumptions:

- Unduplicate job exists and ran successfully

Task: Build and run a Survive Data Quality job

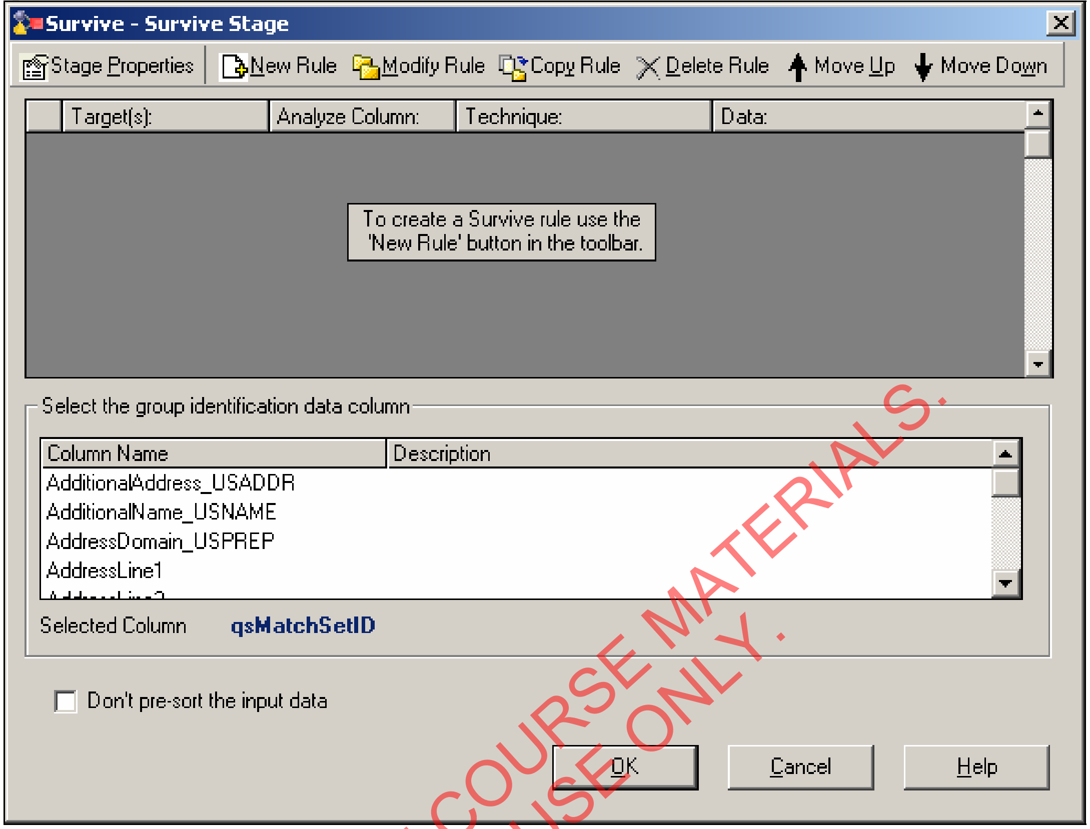
1. Create a new parallel job; add Sequential File and Survive stages as depicted:



The source data for this job is contained in the C:\WINNCRM\Matched.dat file created in the Unduplicate DataStage job from an earlier exercise.

2. Edit the Matched Sequential File stage and set the File property to C:\WINNCRM\Matched.dat.
3. In the Columns tab load the column definitions from the Matched table in the WINNCRM➔TableDefs folder of the repository. Test the validity by using the View Data button.
4. Close the stage.
5. Open the Survive stage.

6. In the section labeled "Select the group identification data column" locate and select the qsMatchSetID column. You should now see:



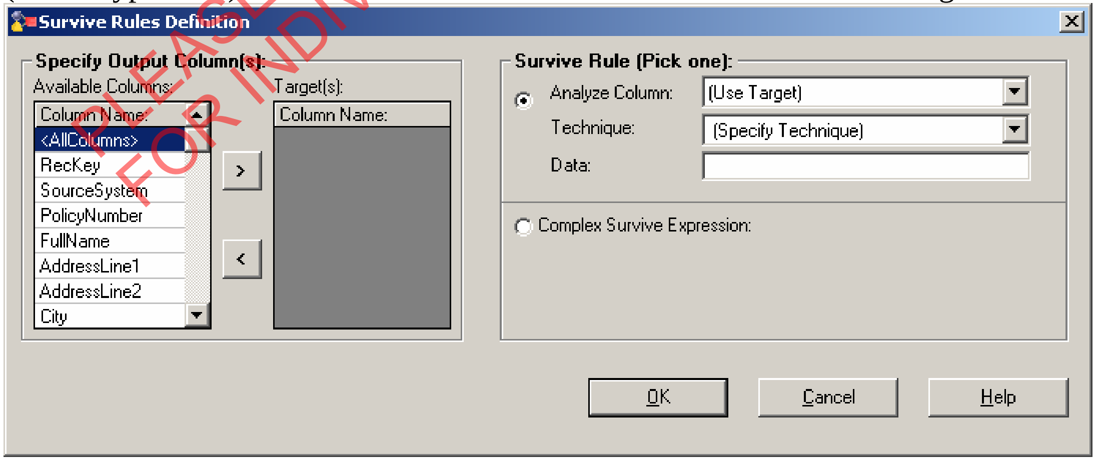
To create a Survive rule use the 'New Rule' button in the toolbar.

Column Name	Description
AdditionalAddress_USADDR	
AdditionalName_USNAME	
AddressDomain_USPREP	
AddressLine1	

Selected Column qsMatchSetID

□ Don't pre-sort the input data

7. Click the New Rule tab to create criteria for survivorship.
8. General procedure: Select a column from the list on the left and move it to the Target(s) using the > button, click the Analyze Column and Technique drop down boxes to achieve the desired result. The Data window is used to enter text selection criteria, such as "MP" (do not type the "). Click the OK button to add the rule to the Survive stage.



Available Columns:

Column Name:
<AllColumns>
RecKey
SourceSystem
PolicyNumber
FullName
AddressLine1
AddressLine2
City

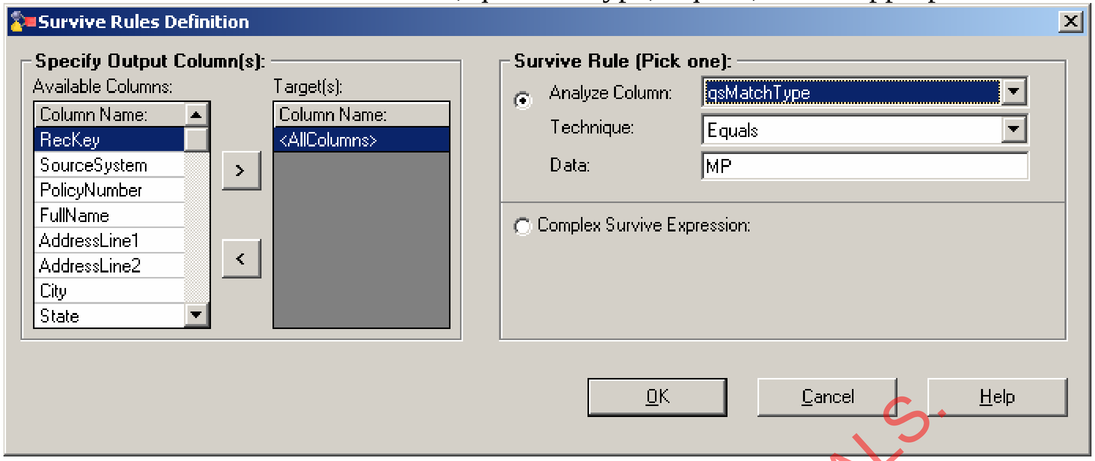
● Analyze Column: (Use Target)

Technique: (Specify Technique)

Data:

○ Complex Survive Expression:

9. For instance: Select <AllColumns>, qsMatchType, Equals, MP in appropriate boxes



Survive Rule (Pick one):

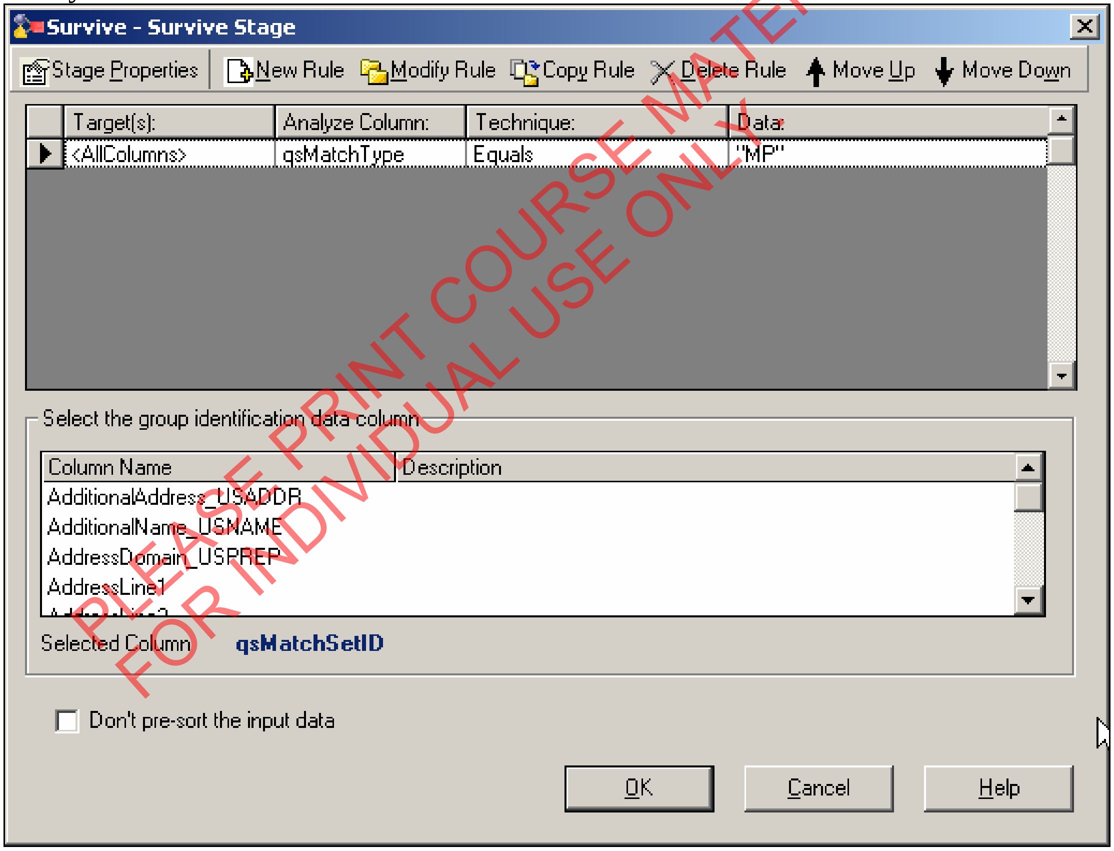
● Analyze Column: qsMatchType
Technique: Equals
Data: MP

○ Complex Survive Expression:

Column Name:
RecKey
SourceSystem
PolicyNumber
FullName
AddressLine1
AddressLine2
City
State

Column Name:
<AllColumns>

Will yield:



Target(s):	Analyze Column:	Technique:	Data:
<AllColumns>	qsMatchType	Equals	"MP"

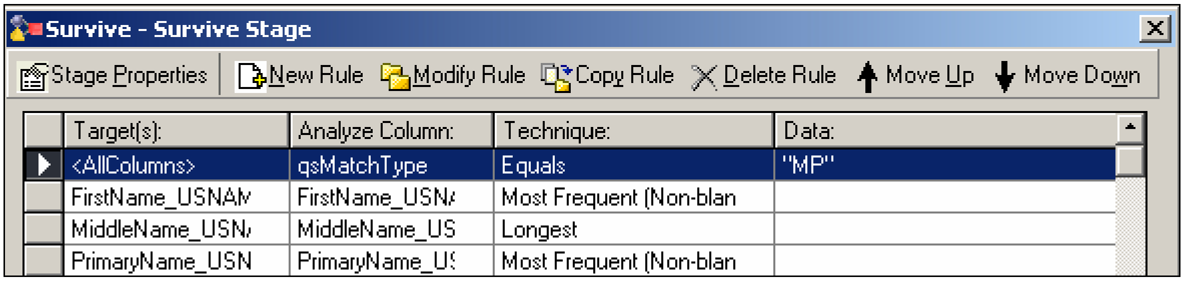
Select the group identification data column

Column Name	Description
AdditionalAddress_USADDR	
AdditionalName_USNAME	
AddressDomain_USPREP	
AddressLine1	

Selected Column qsMatchSetID

□ Don't pre-sort the input data

10. Use the New Rule tab to add criteria for survivorship until you are finished. Final result should be:

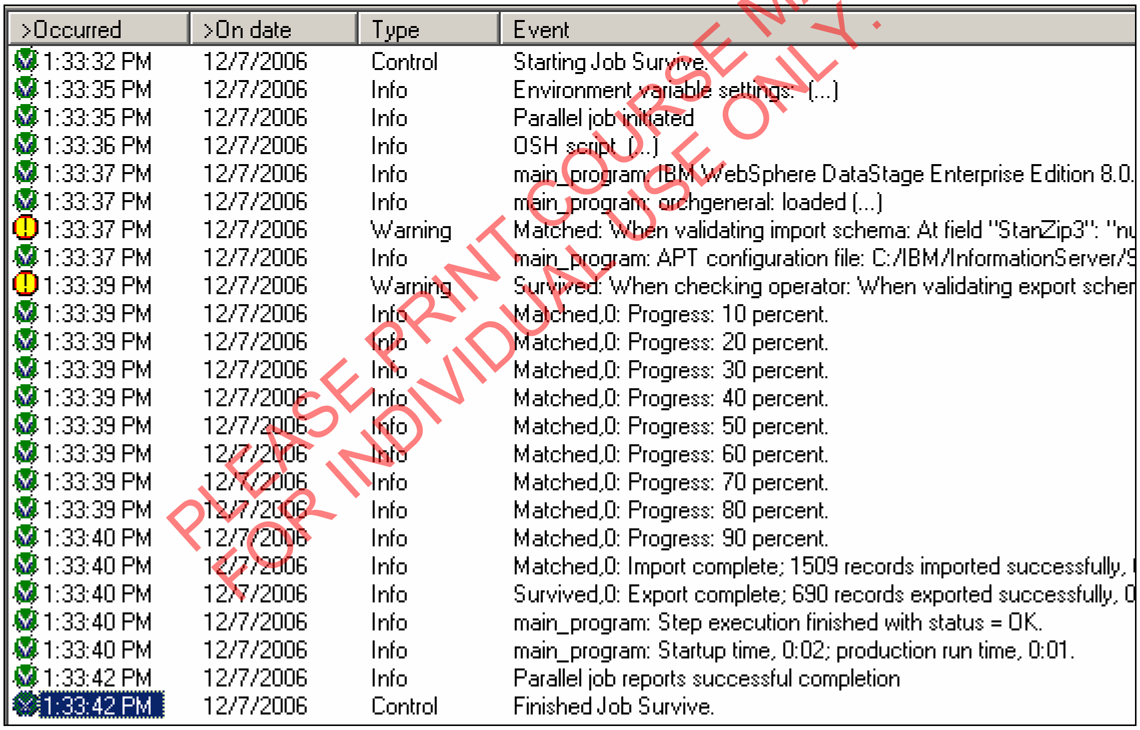


Survive - Survive Stage

Target(s):	Analyze Column:	Technique:	Data:
<AllColumns>	qsMatchType	Equals	"MP"
FirstName_USNAM	FirstName_USNA	Most Frequent (Non-blan	
MiddleName_USN	MiddleName_US	Longest	
PrimaryName_USN	PrimaryName_US	Most Frequent (Non-blan	

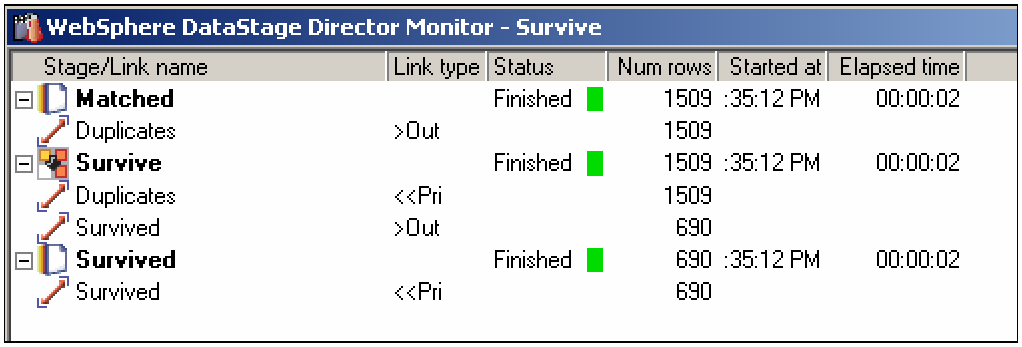
11. Close the Survive stage.
12. Right-click on the Survive stage and select Auto-map columns ➔ All output links.
13. Open the Survived Sequential File stage and set the File property to C:\WINNCRM\Survived.dat.
14. Click the format tab and set the NULL field value to NULL.
15. Save your job as Survive. Compile and run. Use the job monitor to get link counts.

Job Log



>Occurred	>On date	Type	Event
1:33:32 PM	12/7/2006	Control	Starting Job Survive.
1:33:35 PM	12/7/2006	Info	Environment variable settings: (...)
1:33:35 PM	12/7/2006	Info	Parallel job initiated
1:33:36 PM	12/7/2006	Info	OSH script (...)
1:33:37 PM	12/7/2006	Info	main_program: IBM WebSphere DataStage Enterprise Edition 8.0.
1:33:37 PM	12/7/2006	Info	main_program: orchgeneral: loaded (...)
1:33:37 PM	12/7/2006	Warning	Matched: When validating import schema: At field "StanZip3": "nu
1:33:37 PM	12/7/2006	Info	main_program: APT configuration file: C:/IBM/InformationServer/S
1:33:39 PM	12/7/2006	Warning	Survived: When checking operator: When validating export scher
1:33:39 PM	12/7/2006	Info	Matched,0: Progress: 10 percent.
1:33:39 PM	12/7/2006	Info	Matched,0: Progress: 20 percent.
1:33:39 PM	12/7/2006	Info	Matched,0: Progress: 30 percent.
1:33:39 PM	12/7/2006	Info	Matched,0: Progress: 40 percent.
1:33:39 PM	12/7/2006	Info	Matched,0: Progress: 50 percent.
1:33:39 PM	12/7/2006	Info	Matched,0: Progress: 60 percent.
1:33:39 PM	12/7/2006	Info	Matched,0: Progress: 70 percent.
1:33:39 PM	12/7/2006	Info	Matched,0: Progress: 80 percent.
1:33:40 PM	12/7/2006	Info	Matched,0: Progress: 90 percent.
1:33:40 PM	12/7/2006	Info	Matched,0: Import complete; 1509 records imported successfully, (
1:33:40 PM	12/7/2006	Info	Survived,0: Export complete; 690 records exported successfully, 0
1:33:40 PM	12/7/2006	Info	main_program: Step execution finished with status = OK.
1:33:40 PM	12/7/2006	Info	main_program: Startup time, 0:02; production run time, 0:01.
1:33:42 PM	12/7/2006	Info	Parallel job reports successful completion
1:33:42 PM	12/7/2006	Control	Finished Job Survive.

Job Monitor



WebSphere DataStage Director Monitor - Survive

Stage/Link name	Link type	Status	Num rows	Started at	Elapsed time
Matched		Finished 🟩	1509	:35:12 PM	00:00:02
Duplicates	>Out		1509		
Survive		Finished 🟩	1509	:35:12 PM	00:00:02
Duplicates	<<Pri		1509		
Survived	>Out		690		
Survived		Finished 🟩	690	:35:12 PM	00:00:02
Survived	<<Pri		690		

IBM Information Server Business Glossary Lab Exercises

Lab 1: Accessing the IBM Information Server Business Glossary

Objective: Access the Glossary

Although performing tasks within the Business Glossary is very intuitive, the following exercises have been created to aid in your familiarization with the Glossaries basic activities and navigation.

The glossary is accessed by using Microsoft® Internet Explorer version 6.

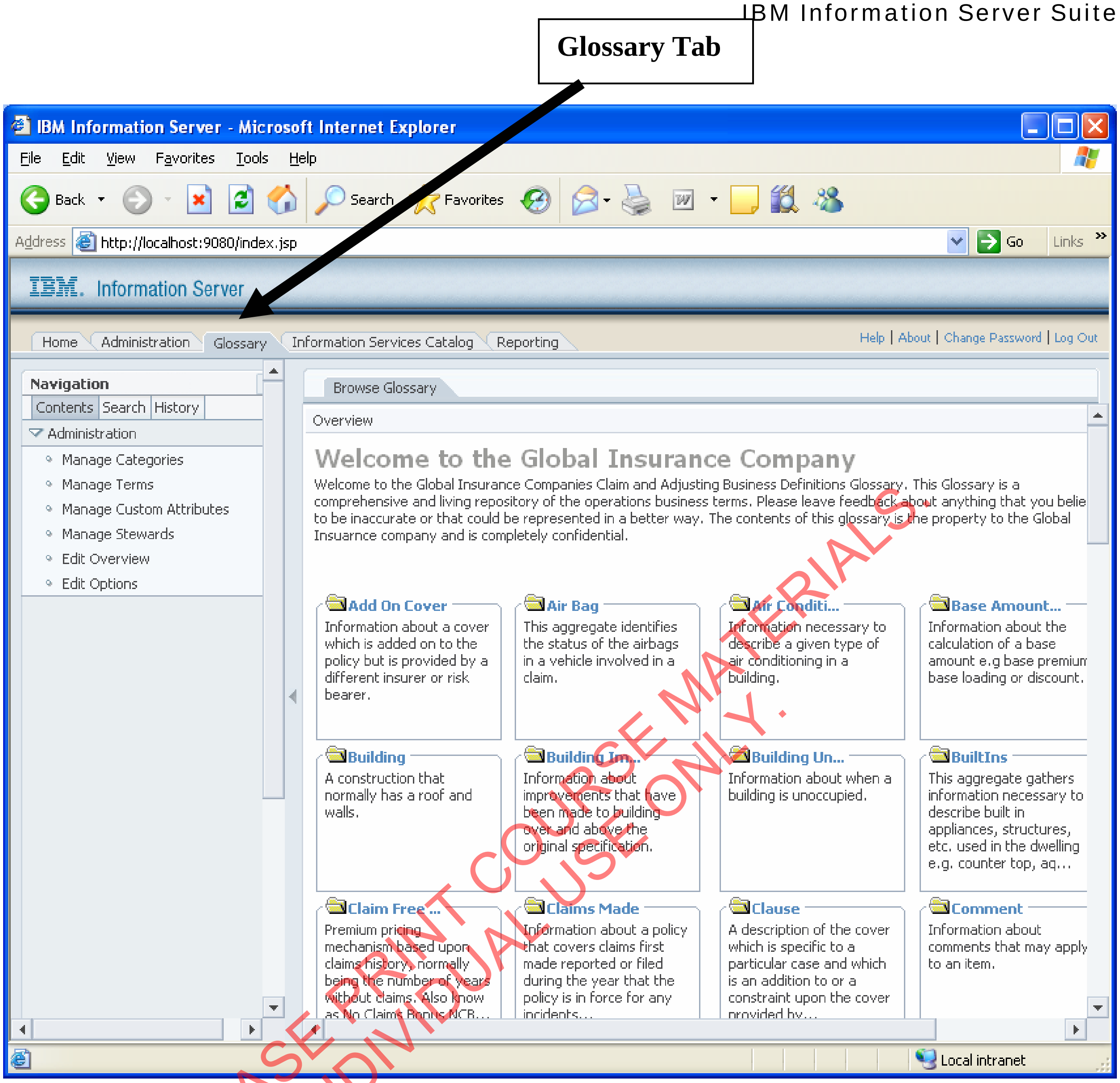
To access IBM Information Server Business Glossary:

1. Open Internet Explorer and connect to the following URL: http://host_server:port, where *host_server* is the name or IP address of the IBM Information Server Web console for the WebSphere Metadata Server that you want to connect to.
2. Enable popups for this site in your Web browser.
3. Type the user name and password for the IBM Information Server Web console, and click **Enter**.
4. Select the **Glossary** tab.

For example:

- To connect to a server on your network named Andros, type **http://Andros:9080**.
- To connect to a server whose IP address is 666.555.44.333, type **http://666.555.44.33:9080**.
- Very possibly for the purpose of this class the server address will be **http://localhost:9080.**

The following screenshot is the IBM Information Server Web console with the Glossary tab showing. This is only visible is you have be granted Business Glossary user, author or administrator role. You will initially have different content than that in the screenshot. If the Glossary tab is not visible then you will have to add the appropriate suite user role. For the purpose of these exercises, grant an existing or new user the Business Glossary Administrator Role. This must be done by a suite administrator.


Glossary Tab
IBM Information Server - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites
Address http://localhost:9080/index.jsp Go Links
IBM. Information Server
Home Administration Glossary Information Services Catalog Reporting
Help | About | Change Password | Log Out
Navigation
Contents Search History
Administration
Manage Categories
Manage Terms
Manage Custom Attributes
Manage Stewards
Edit Overview
Edit Options
Browse Glossary
Overview
Local intranet

Welcome to the Global Insurance Company

Welcome to the Global Insurance Companies Claim and Adjusting Business Definitions Glossary. This Glossary is a
comprehensive and living repository of the operations business terms. Please leave feedback about anything that you belie
to be inaccurate or that could be represented in a better way. The contents of this glossary is the property to the Global
Insuarnce company and is completely confidential.

Add On Cover Information about a cover which is added on to the policy but is provided by a different insurer or risk bearer.

Air Bag This aggregate identifies the status of the airbags in a vehicle involved in a claim.

Air Conditi... Information necessary to describe a given type of air conditioning in a building.

Base Amount... Information about the calculation of a base amount e.g base premiur base loading or discount.

Building A construction that normally has a roof and walls.

Building Im... Information about improvements that have been made to building over and above the original specification.

Building Un... Information about when a building is unoccupied.

BuiltIns This aggregate gathers information necessary to describe built in appliances, structures, etc. used in the dwelling e.g. counter top, aq...

Claim Free ... Premium pricing mechanism based upon claims history, normally being the number of years without claims. Also know as No Claims Bonus NCB...

Claims Made Information about a policy that covers claims first made reported or filed during the year that the policy is in force for any incidents...

Clause A description of the cover which is specific to a particular case and which is an addition to or a constraint upon the cover provided by...

Comment Information about comments that may apply to an item.

Lab 2: Administration Console

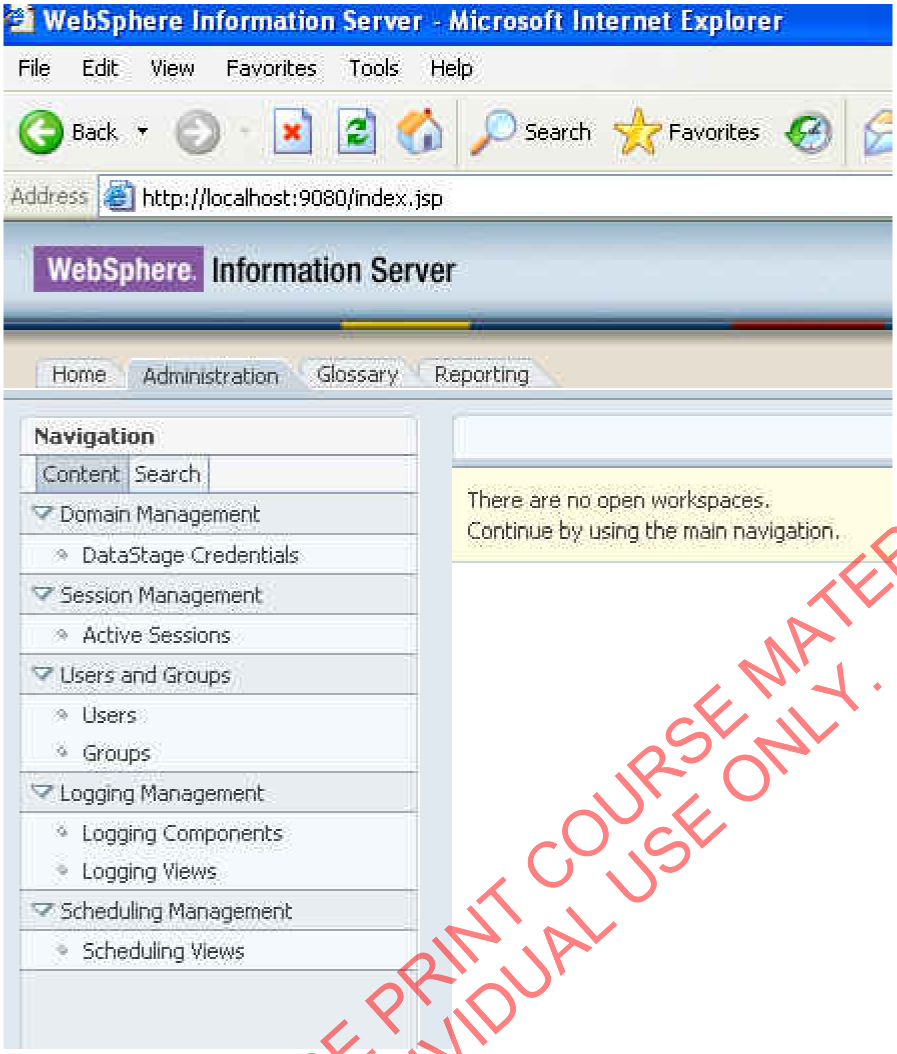
Assumptions

- A Suite user named demohawk has been defined. In this exercise will grant the demohawk user the Business Glossary Administrator role. It is possible that this role has already been granted to the demohawk user. But going through these steps will help in understanding the administration console and the types of privileges and roles utilized in the suite.

Task: Open the Administration Console

22. Open your web browser.
23. Enter the address to your Administration Console, e.g., http://localhost:9080. Here, localhost is an alias for your local machine. If the Administration console is running on a different machine, use the name or IP address of this machine instead of localhost.

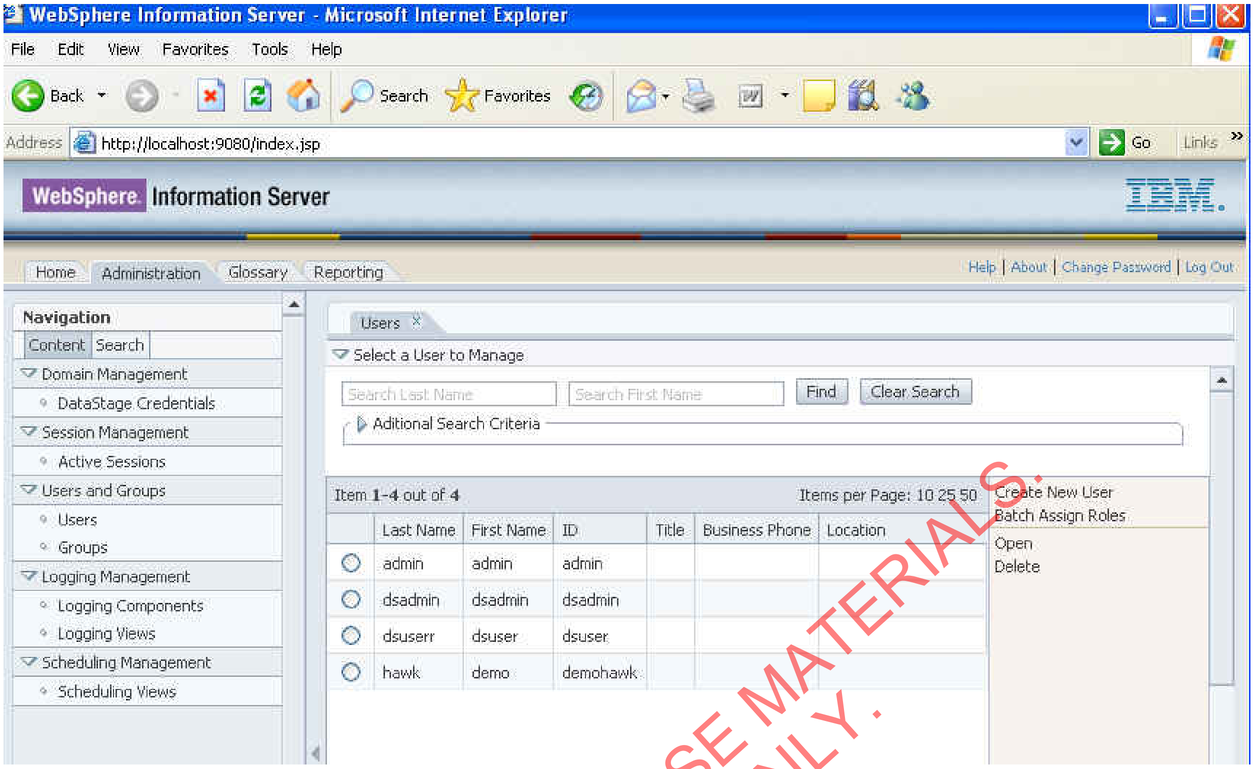
24. Click the Administration tab.



There are no open workspaces.
Continue by using the main navigation.

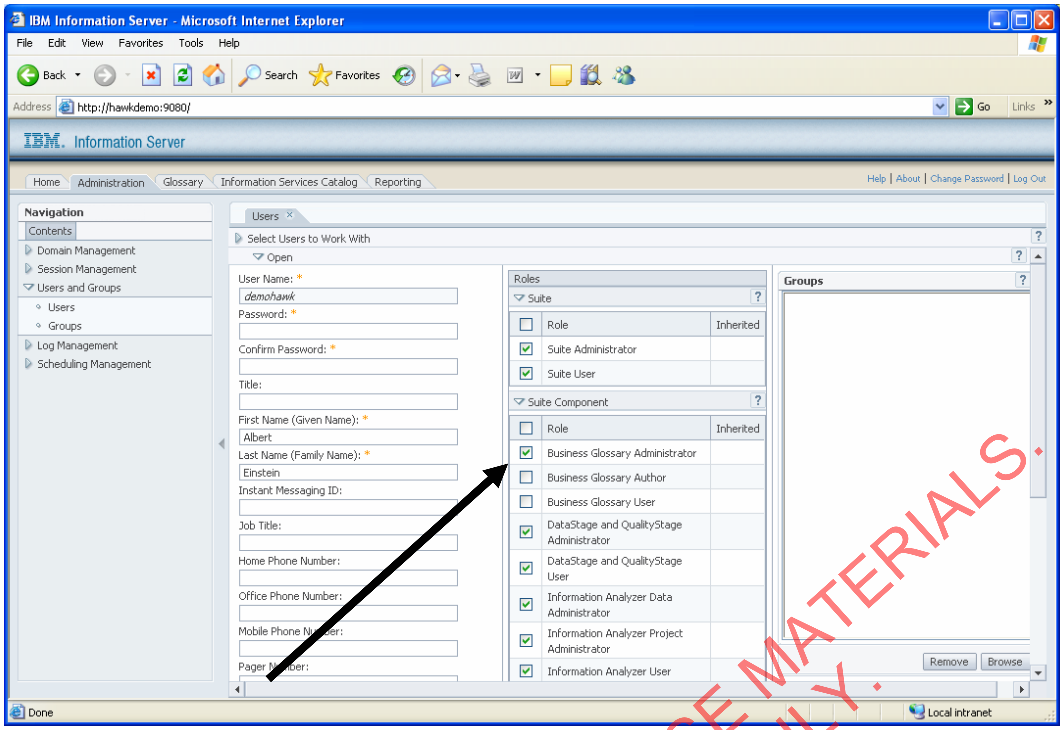
Task: Create add Business Glossary Admin Role to user demohawk

1. Expand the Users and Groups folder and then click Users.



	Last Name	First Name	ID	Title	Business Phone	Location
○	admin	admin	admin			
○	dsadmin	dsadmin	dsadmin			
○	dsuserr	dsuser	dsuser			
○	hawk	demo	demohawk			

2. Select the demohawk user and then click Open.
3. Note the first and last names of this user. Note what Suite Roles and Product Roles that have been assigned to this user.
4. Assign demohawk the Business Glossary Administrator product role.


IBM Information Server - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites
Address http://hawkdemo:9080/
Go Links
IBM. Information Server
Home Administration Glossary Information Services Catalog Reporting
Help | About | Change Password | Log Out
Navigation
Contents
Domain Management
Session Management
Users and Groups
Users
Groups
Log Management
Scheduling Management
Users
Select Users to Work With
Open
Groups
Remove Browse
Done
Local intranet

User Name: * demohawk

Password: *

Confirm Password: *

Title:

First Name (Given Name): * Albert

Last Name (Family Name): * Einstein

Instant Messaging ID:

Job Title:

Home Phone Number:

Office Phone Number:

Mobile Phone Number:

Pager Number:

Roles

Suite

□	Role	Inherited
☑	Suite Administrator	
☑	Suite User	

Suite Component

□	Role	Inherited
☑	Business Glossary Administrator	
□	Business Glossary Author	
□	Business Glossary User	
☑	DataStage and QualityStage Administrator	
☑	DataStage and QualityStage User	
☑	Information Analyzer Data Administrator	
☑	Information Analyzer Project Administrator	
☑	Information Analyzer User	

Lab 3: Upload Categories and Terms

Objectives:

Upload the Business Categories and Terms into the Global Insurance Glossary of business categories and terms.

Assumptions:

- You must have the Business Glossary Administrator role or Business Glossary Author role to perform this task.
- You must prepare an XML file that complies with the IBM® WebSphere® Business Glossary category and term data file schema. Links to the schema and to a sample XML file that complies with the schema appear in the Upload Categories and Terms page.
- Names of categories and terms must start and end with a character that is not a space. Names cannot contain any of the following characters:
  - . (period)
  - , (comma)
  - ; (semicolon)
  - % (percentage sign)
  - " (quotation marks)

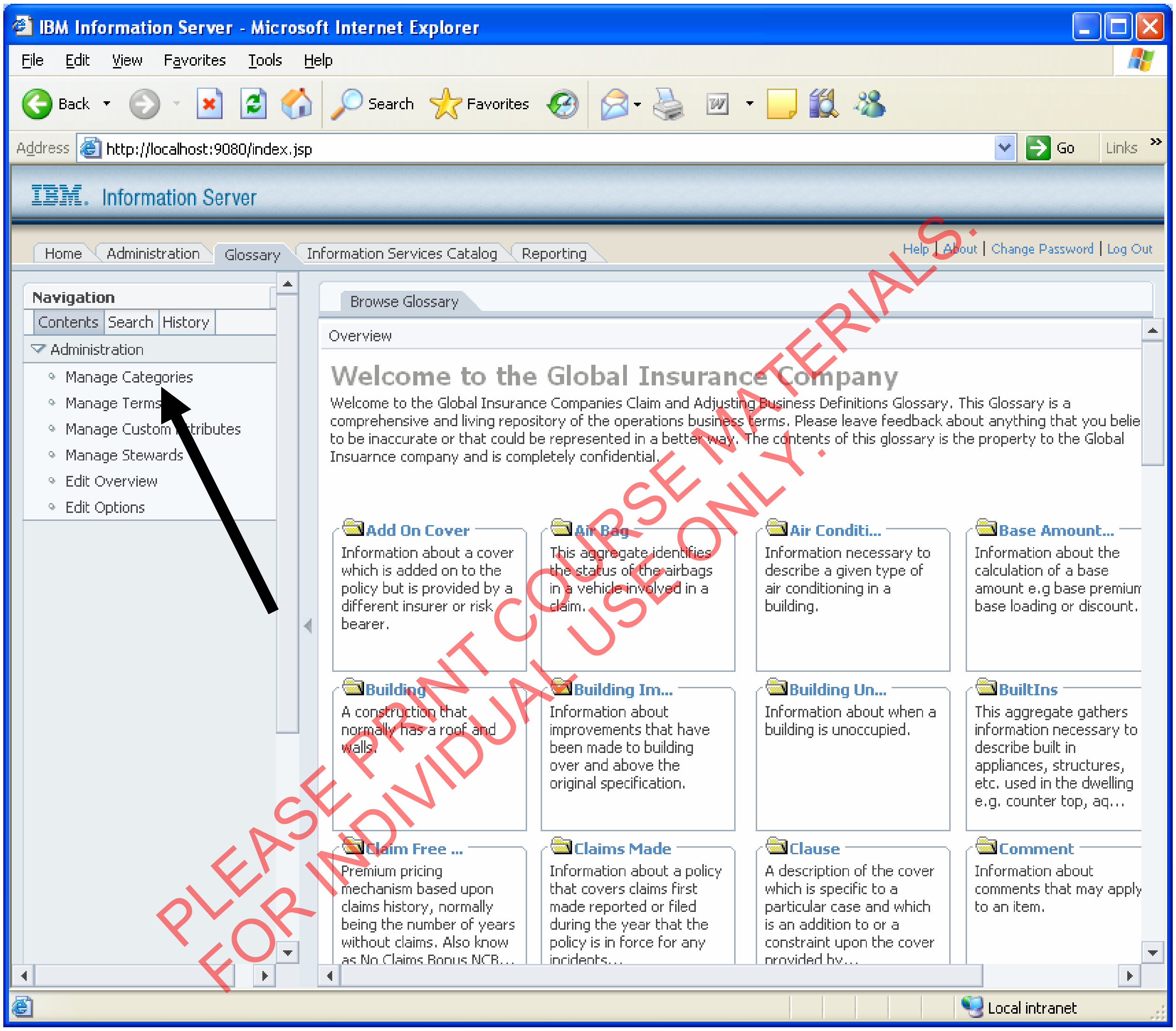
- You must use the appropriate entity references for reserved characters in XML:

For this reserved character	Use this entity reference
< (left angle bracket)	&lt;
> (right angle bracket)	&gt;
" (quotation marks)	&quot;
' (apostrophe)	&#39;
& (ampersand)	&amp;

Task: Upload the Categories and Terms

Several categories and terms have been provided in the appropriate xml format. The filename is **Business Glossary Upload Sample (Insurance version 1).xml**. This file is contained in a folder on your student CD called Business Glossary.

1. To upload categories, terms, and custom attributes:
   In the Navigation pane on the Glossary tab, select **Contents** > **Administration** > **Manage Categories**.



Welcome to the Global Insurance Company

Welcome to the Global Insurance Companies Claim and Adjusting Business Definitions Glossary. This Glossary is a comprehensive and living repository of the operations business terms. Please leave feedback about anything that you belie to be inaccurate or that could be represented in a better way. The contents of this glossary is the property to the Global Insuarnce company and is completely confidential.

Add On Cover: Information about a cover which is added on to the policy but is provided by a different insurer or risk bearer.

Air Bag: This aggregate identifies the status of the airbags in a vehicle involved in a claim.

Air Conditi...: Information necessary to describe a given type of air conditioning in a building.

Base Amount...: Information about the calculation of a base amount e.g base premiur base loading or discount.

Building: A construction that normally has a roof and walls.

Building Im...: Information about improvements that have been made to building over and above the original specification.

Building Un...: Information about when a building is unoccupied.

BuiltIns: This aggregate gathers information necessary to describe built in appliances, structures, etc. used in the dwelling e.g. counter top, aq...

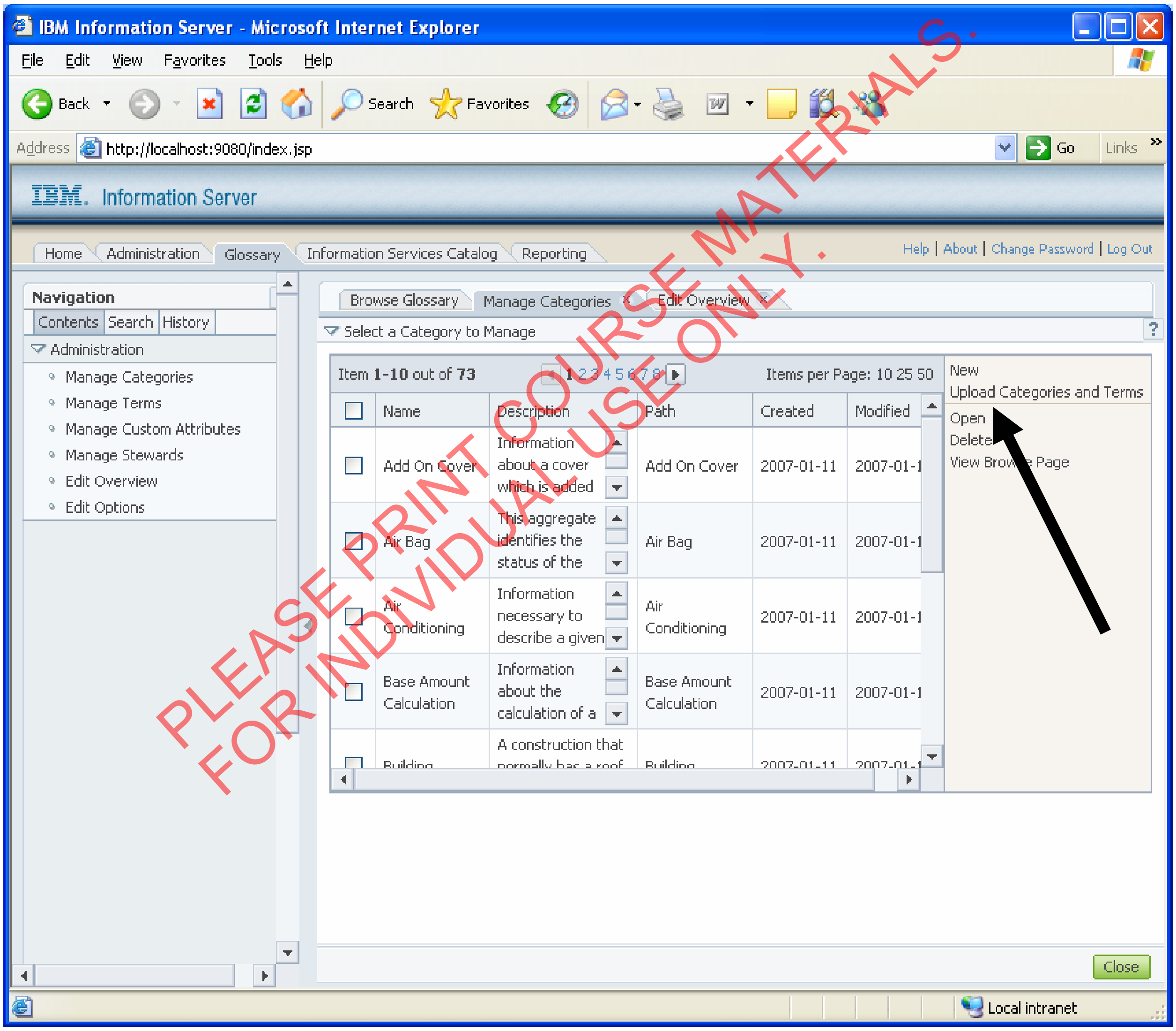
Claim Free ...: Premium pricing mechanism based upon claims history, normally being the number of years without claims. Also know as No Claims Bonus NCB...

Claims Made: Information about a policy that covers claims first made reported or filed during the year that the policy is in force for any incidents...

Clause: A description of the cover which is specific to a particular case and which is an addition to or a constraint upon the cover provided by...

Comment: Information about comments that may apply to an item.

2. Click **Upload Categories and Terms**.


IBM Information Server - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites
Address http://localhost:9080/index.jsp
Go Links
IBM. Information Server
Home Administration Glossary Information Services Catalog Reporting
Help | About | Change Password | Log Out
Navigation
Contents Search History
Administration
Manage Categories
Manage Terms
Manage Custom Attributes
Manage Stewards
Edit Overview
Edit Options
Browse Glossary Manage Categories Edit Overview
Select a Category to Manage
Item 1-10 out of 73
1 2 3 4 5 6 7 8
Items per Page: 10 25 50
New
Upload Categories and Terms
Open
Delete
View Browse Page
Close
Local intranet

□	Name	Description	Path	Created	Modified
□	Add On Cover	Information about a cover which is added	Add On Cover	2007-01-11	2007-01-1
□	Air Bag	This aggregate identifies the status of the	Air Bag	2007-01-11	2007-01-1
□	Air Conditioning	Information necessary to describe a given	Air Conditioning	2007-01-11	2007-01-1
□	Base Amount Calculation	Information about the calculation of a	Base Amount Calculation	2007-01-11	2007-01-1
□	Building	A construction that normally has a roof	Building	2007-01-11	2007-01-1

3. Click **Browse**, navigate to an XML file, select it and click **Open**.

4. Click **Upload**. The Upload Categories and Terms page displays the number of categories and terms that were uploaded.
5. Click **OK**.

At this point you should now have a variety of categories and terms to browse and search in the glossary.

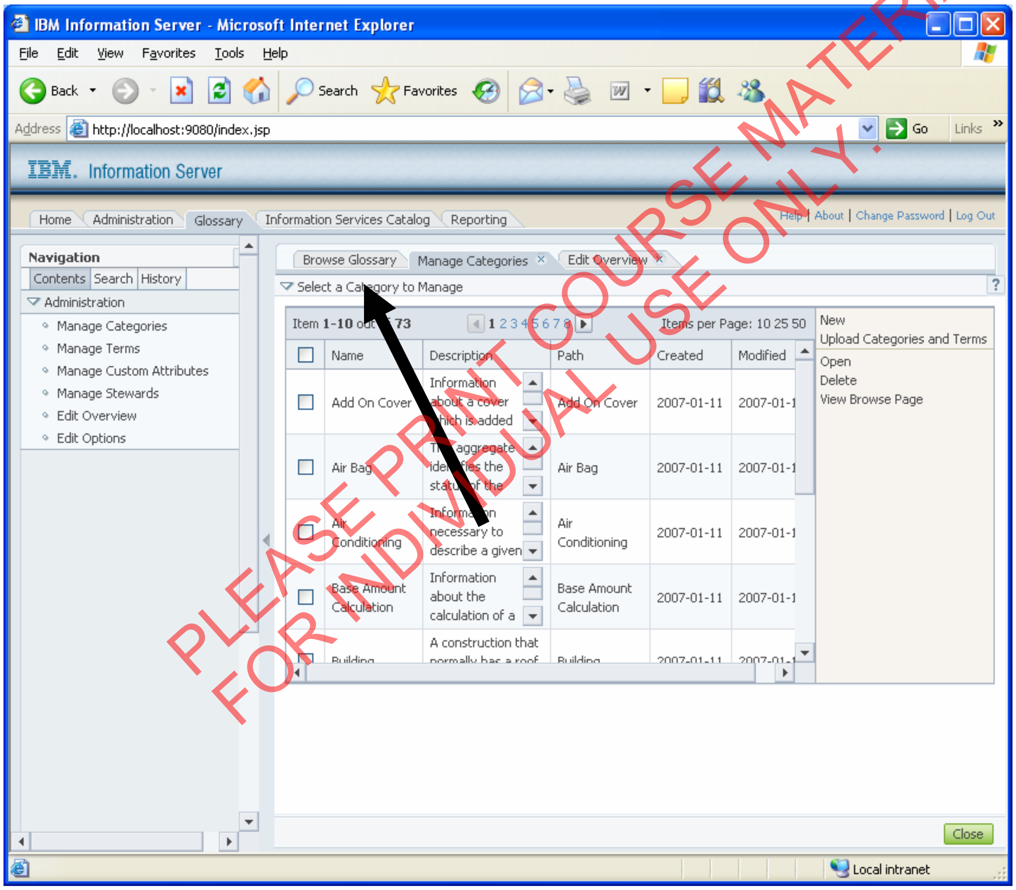
Lab 4: Browsing the glossary

Overview:

You can browse the glossary structure to explore categories, terms, and objects in the repository of your IBM Information Server.

You can start browsing the glossary from the Overview page, which displays the top-level categories that the glossary administrator has designated as most important for navigation in the metadata repository. You can also search for objects and select an object from the search results. When you select an object, the browse page of the object is displayed on the Browse Glossary tab, which lists the name, class, steward and other important properties of the object. You can inspect the attributes of the object, browse its relationships to other objects, and send feedback to the administrator. Administrators and authors can add and edit notes about the object.

1. Navigate back to the “Home” page. The “Browse Glossary tab will bring you home.


IBM Information Server - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites
Address http://localhost:9080/index.jsp
Go Links
IBM. Information Server
Home Administration Glossary Information Services Catalog Reporting
Help | About | Change Password | Log Out
Navigation
Contents Search History
Administration
Manage Categories
Manage Terms
Manage Custom Attributes
Manage Stewards
Edit Overview
Edit Options
Browse Glossary Manage Categories Edit Overview
Select a Category to Manage
Item 1-10 out of 73
1 2 3 4 5 6 7 8
Items per Page: 10 25 50
New
Upload Categories and Terms
Open
Delete
View Browse Page
Close
Local intranet

□	Name	Description	Path	Created	Modified
□	Add On Cover	Information about a cover which is added	Add On Cover	2007-01-11	2007-01-1
□	Air Bag	The aggregate identifies the status of the	Air Bag	2007-01-11	2007-01-1
□	Air Conditioning	Information necessary to describe a given	Air Conditioning	2007-01-11	2007-01-1
□	Base Amount Calculation	Information about the calculation of a	Base Amount Calculation	2007-01-11	2007-01-1
□	Building	A construction that normally has a roof	Building	2007-01-11	2007-01-1

2. Select a category such as **Coverage Option** and browse its contents and available tasks. Remember that the availability of the tasks is governed by the user role.

Task: Searching the Glossary

Overview:

Using the search tool is often the quickest way to find an object in the repository of WebSphere® Metadata Server.
You can perform simple and advanced searches to find repository objects of all classes, including, but not limited to, categories, terms, tables, columns, job definitions, users, and groups. The more information that you can specify about the object that you are searching for, the faster the search results are returned.
The containment path of the object is displayed in the Path column in the search results so you can distinguish between objects with similar or identical names. For example, the containment path for a column might display the names of the containing table, schema, database, and host computer the column was imported from.
When you locate the object in the search results, you can click its name to display the browse page for the object. You can then inspect its attributes, browse its relationships to other objects, and send feedback to the administrator. Administrators and authors can add and edit notes about the object.

Finding Objects with the Simple Search

1. In the Navigation pane on the Glossary tab, select **Search** > **Simple Search**.

2. In the **Simple Search** field, type the search criteria. You can use all or part of a name or short description, or you can use multiple keywords from names or descriptions, separated by spaces or commas. In this case search for **Claims Made.**

3. Click **Search**. The Search Results page displays a list of objects in the metadata repository whose names or short descriptions match the search string. If you typed multiple keywords, the list includes the objects whose names or short descriptions include all of the keywords. As you will notice, many items will be returned. We will learn how to perform a more exacting search in the next task.

4. In the list, click an object name to view the object, its relationships, and its attributes.

Finding objects with an advanced search

You can use multiple criteria when you search for objects that are stored in the metadata repository. In this case, we will once again search for the term **Claims Made**, but for **Claims Made** only.

To find objects with an advanced search:

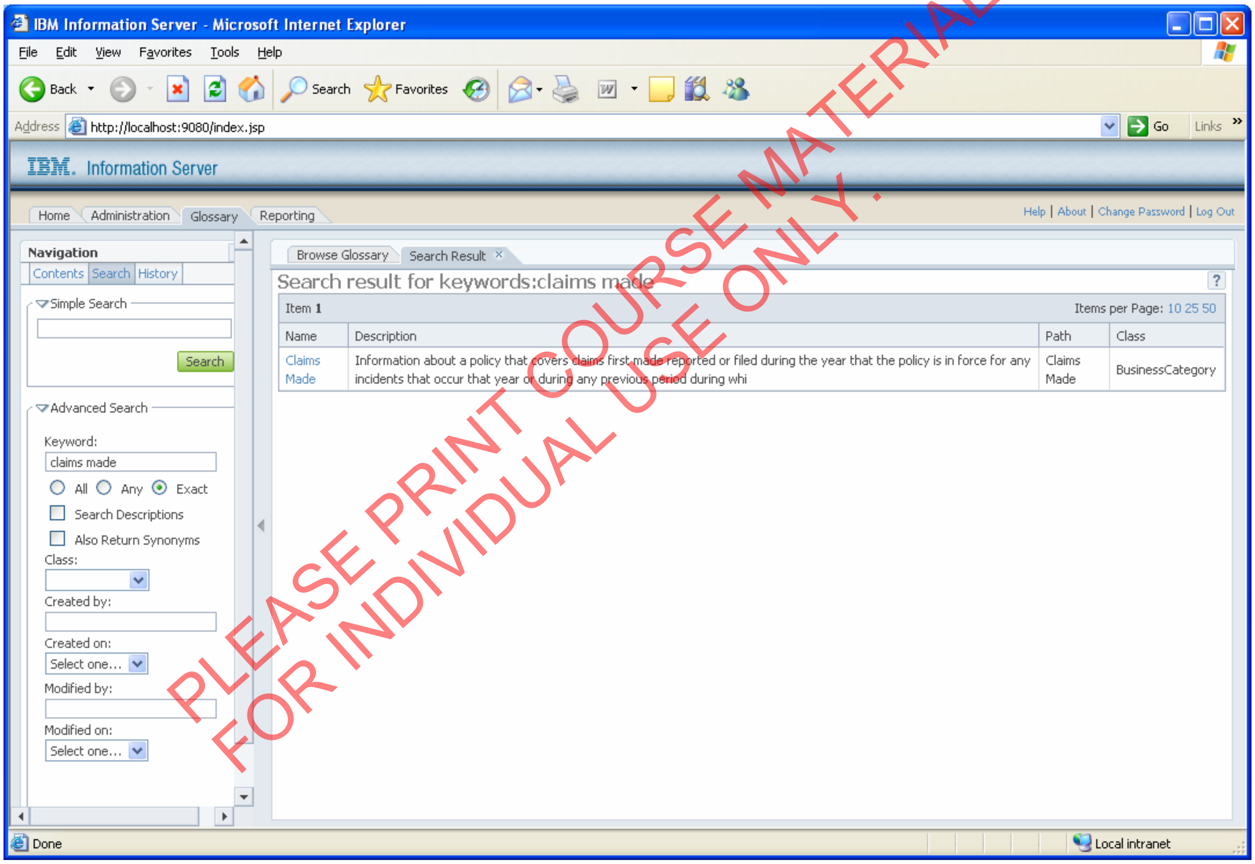
1. In the Navigation pane on the Glossary tab, select **Search** > **Advanced Search**.

   Specify the criteria for the search:

   - The keyword as before will be **Claims Made**.
   - Filter for only exact matches
   - Uncheck the Search Descriptions check box.

.

2. Click **Search**. A list of search results is displayed on the Search Results page. In this case, only the **Claims Made** term should be returned.


IBM Information Server - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites
Address http://localhost:9080/index.jsp
Go Links
IBM. Information Server
Home Administration Glossary Reporting
Help | About | Change Password | Log Out
Navigation
Contents Search History
Simple Search
Search
Advanced Search
Browse Glossary Search Result
Done
Local intranet

Search result for keywords:claims made

Item 1 — Items per Page: 10 25 50

Name	Description	Path	Class
Claims Made	Information about a policy that covers claims first made reported or filed during the year that the policy is in force for any incidents that occur that year or during any previous period during whi	Claims Made	BusinessCategory

Keyword: claims made

○ All ○ Any ● Exact

□ Search Descriptions

□ Also Return Synonyms

Class:

Created by:

Created on: Select one...

Modified by:

Modified on: Select one...

Lab 5: Business Category and Term Creation and Editing

Overview:

Administrators and authors can create and edit categories that contain or reference terms and that serve as subcategories or parent categories to other categories.

Assumptions:

In this task you will create a category named **MyCategory** or any other name you wish. If possible consider something that categorizes a facet of your business that could become a parent of a subcategory or business term. For example: the construction category contains terms such as material, material code, material percent etc.

To create or edit a category:

1. In the Navigation pane on the Glossary tab, select **Contents** > **Administration** > **Manage Business Categories**.

2. On the Manage Categories page, specify whether to create or edit a category:

   - To create a category, click **New**.
   - To edit a category, select a category from the list of all categories, and click **Open**.

3. If you are creating a category, in the **Name** field, type a name for the category. If you are editing an existing category, you can change the name.

4. **Optional:** Specify or change information about the category. You can add or edit descriptions, specify a steward, define relationships to other categories and terms, specify values for custom attributes, and specify whether the category is displayed on the Overview page.

5. Click **Save and Close** to save your changes and close the category.

Task: Business Term Creation and Editing

Overview:

Administrators and authors can create and edit terms to categorize one or more metadata objects in the metadata repository.

Assumptions:

In this task you will create a business term named **MyBusinessTerm** or any other name you wish. Think about how well the category you just created is named in relation to the term you are about to create.

To create or edit a category:

Creating and editing terms

Administrators and authors can create and edit terms to categorize one or more metadata objects in the metadata repository.

Prerequisites: You must have the Business Glossary Administrator role or Business Glossary Author role to perform this task. A category must exist to contain the term. As in the previous To create or edit a term:

1. In the Navigation pane on the Glossary tab, select **Contents** > **Administration** > **Manage Business Terms**.

2. On the Manage Terms page, specify whether to create a term or edit a term:

   - To create a term, click **New**.
   - To edit a term, select a term from the list of all terms, and click **Open**.

3. If you are creating a new term, specify its name and parent category. If you are editing the term, you can edit the name and parent category.

   - In the **Name** field, type a name for the term.
   - Next to the **Parent Category** field, click **Select** to select a parent category to contain the term.

4. **Optional:** Specify a Data Steward when the term is created. This can be done at a later time by editing the term. Data Stewardship is a role that must be granted to a user or group.

5. **Optional:** Specify or change information about the term. You can add or edit descriptions, specify a steward, define relationships to other terms, classify objects, specify values for custom attributes, give an example of the term, specify the status of the term, specify how the term is used, and specify abbreviations for the term.

6. Click **Save and Close** to save your changes and close the term.

Lab 6: Working with Annotations. (Notes)

Overview:

Administrators and authors can add, edit and delete notes on the browse page of any object.

Assumptions:

In this task you will create a note on the business term of your choice. Then you can edit or delete it. The note might be an observation you have made about the data that you feel should be shared with others. This is an example of where it can be very helpful to have a Data Steward assigned to an object.

Task: Work with a note:

1. Display the browse page of an object by any of these methods:

- Browsing from the Overview page
- Finding objects by using a simple search
- Finding objects with an advanced search
- Browsing the properties and relationships of objects

2. Add, edit, or delete the note:

To add a note:	a. In the **Tasks** list, click **Add Note**. b. In the New Note window, type a label and comment for the note and click **OK**. The note is added to the Notes tab.
To edit a note:	a. On the Notes tab, in the row that describes the note that you want to edit, click ✎ (edit note). The icon is not displayed if you do not have authority to edit the note. b. In the Edit Note window, type a label and comment for the note and click **OK**.
To delete a note:	a. On the Notes tab, in the row that describes the note that you want to delete, click 🗑 (delete note). The icon is not displayed if you do not have authority to delete the note. b. Click **Yes** to confirm deletion.

Lab 7: Working with Custom Attributes

Overview:

Administrators can create custom attributes to store information about terms and categories, when that information does not fit into the standard attributes and relationships of the glossary model. You can use custom attributes to apply governance standards, enable architecture frameworks, or provide other metadata that is standard for your organization.

When you create a custom attribute, you specify that it applies to either terms or categories, or to both terms and categories. If you apply the custom attribute to both terms and categories, two separate custom attributes are created, one that applies to terms, and one that applies to categories.

Each custom attribute has a name, a description, and a valid value type. The valid value type can be any string or an enumerated list of string values.

You can change the valid value type for a custom attribute at any time. When you change the type, the change does not affect any values that are currently assigned for the attribute. The change determines what will happen the next time a user edits the value for a custom attribute. If you change the type of a custom attribute to String, when users subsequently edit the attribute for any object, they can enter any string value. If you change the type of a custom attribute to Enumerated, when users subsequently edit the attribute for any object, they must select values from the enumerated list of values.

The value of the custom attribute for any particular term or category is initially null. After you create the custom attribute, you can specify its value separately for each term or category that it applies to.

For example, you might create a custom attribute named Data Sensitivity with the following description

A number from 1 to 5, which indicates the sensitivity of the data. Sensitivity is a subjective measure of the impact of the data being released to unauthorized consumers.

You can specify that Data Sensitivity attribute applies only to terms. You choose the enumerated valid value type and enter the numbers 1 through 5 as valid values. After you create the custom attribute, you choose one of those valid values for each particular term that you want to specify a value for.

Assumptions:

In this task you will create a custom attribute named **Data Sensitivity.** This attribute will be available to all of your business terms. It does not make sense to associate this attribute with any categories, although this is possible within the glossary to do so. This custom attribute will be of the enumerated type. A custom attribute can be created first and then associated with a term or category, or it can be done as part of the term or category editing process.

Task: Create and work with a custom attribute:

Prerequisite: You must have the Business Glossary Administrator role to perform this task.

To create a custom attribute:

1. In the Navigation pane on the Glossary tab, select **Contents** > **Administration** > **Manage Custom Attributes**.

2. Click **New**.

3. Type a name and description for the custom attribute. The name will be **Data Sensitivity**. Valid values are numbers 1-5.

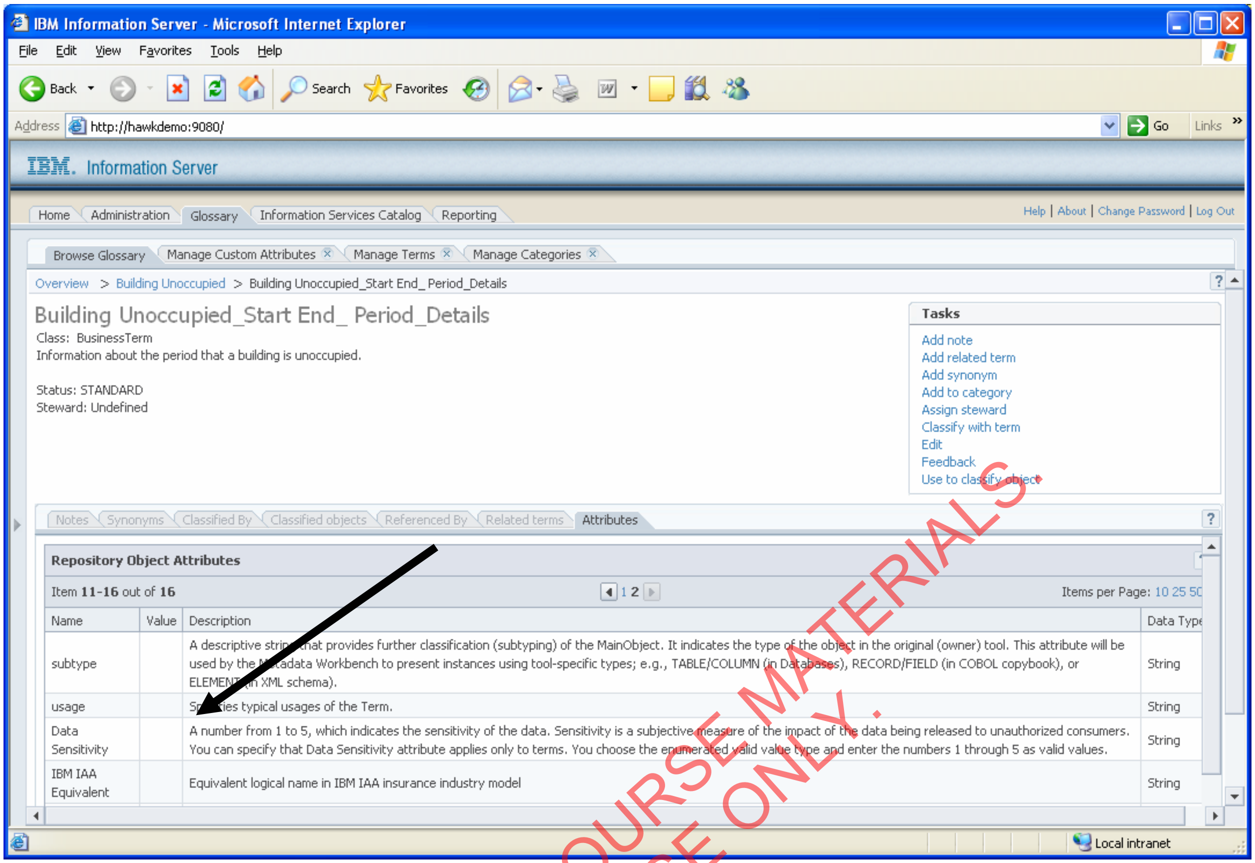
4. Select the class of object that the attribute applies to, either categories, terms, or both. If you select both, two custom attributes are created with the same name and properties. One custom attribute applies to terms and the other applies to categories. In this case, it only makes sense to select **terms.**

5. From the **Attribute Type** drop-down list, select the type of valid value:

Option	**Description**
To specify that any string is a valid value:	Select **String**.
To specify a list of valid string values:	a. Select **Enumeration**. b. Type a single valid value, and click **Add**. c. Repeat step b to add valid values.

6. Click **Save and Close** to save your changes

7. Now browse any of your business terms and find the custom attribute you just created. This will be on the second page of any term you choose. You should notice the custom attribute and that it has no value.


IBM Information Server - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites
Address http://hawkdemo:9080/
Go Links
IBM. Information Server
Home Administration Glossary Information Services Catalog Reporting
Help | About | Change Password | Log Out
Browse Glossary Manage Custom Attributes Manage Terms Manage Categories
Notes Synonyms Classified By Classified objects Referenced By Related terms Attributes
Local intranet

Overview > Building Unoccupied > Building Unoccupied_Start End_ Period_Details

Building Unoccupied_Start End_ Period_Details

Class: BusinessTerm

Information about the period that a building is unoccupied.

Status: STANDARD

Steward: Undefined

Tasks

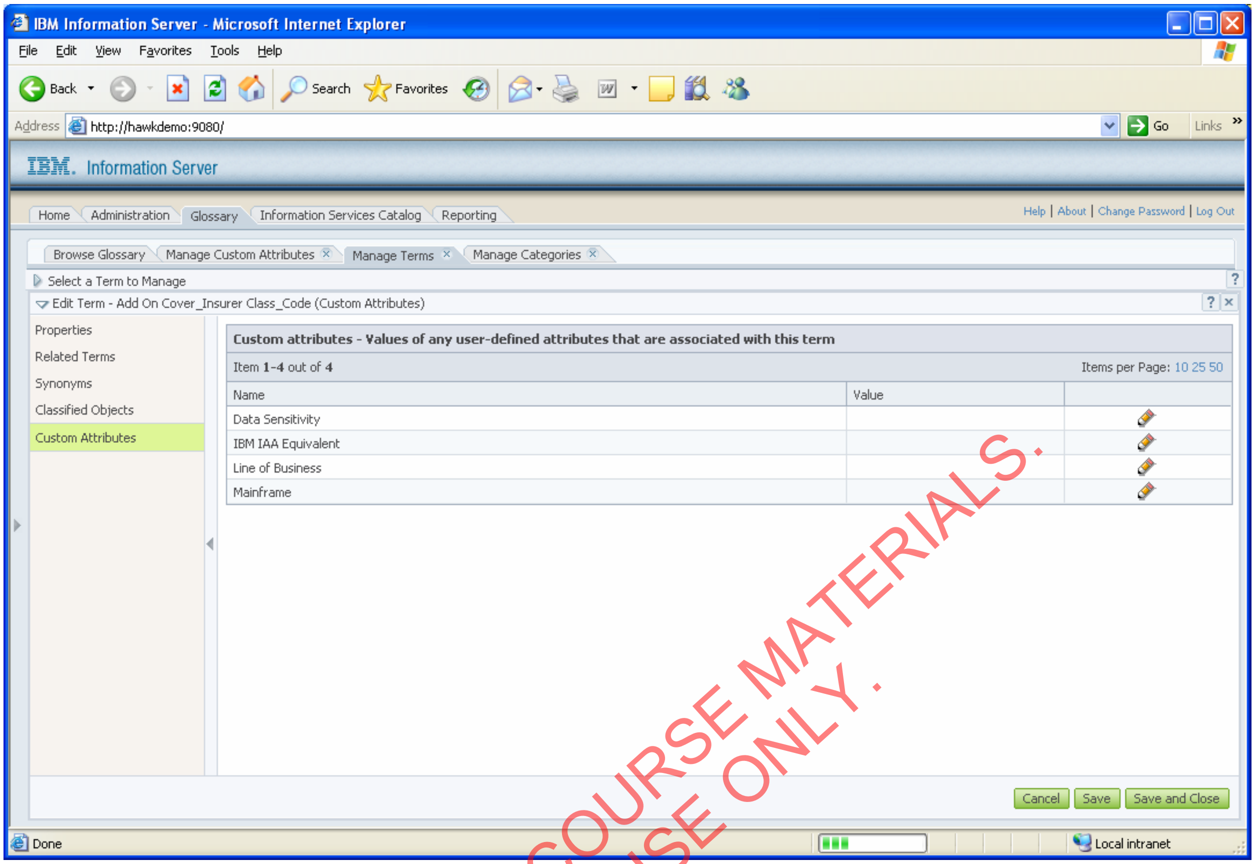
- Add note
- Add related term
- Add synonym
- Add to category
- Assign steward
- Classify with term
- Edit
- Feedback
- Use to classify object

Repository Object Attributes

Item 11-16 out of 16 1 2 Items per Page: 10 25 50

Name	Value	Description	Data Type
subtype		A descriptive string that provides further classification (subtyping) of the MainObject. It indicates the type of the object in the original (owner) tool. This attribute will be used by the Metadata Workbench to present instances using tool-specific types; e.g., TABLE/COLUMN (in Databases), RECORD/FIELD (in COBOL copybook), or ELEMENT (in XML schema).	String
usage		Specifies typical usages of the Term.	String
Data Sensitivity		A number from 1 to 5, which indicates the sensitivity of the data. Sensitivity is a subjective measure of the impact of the data being released to unauthorized consumers. You can specify that Data Sensitivity attribute applies only to terms. You choose the enumerated valid value type and enter the numbers 1 through 5 as valid values.	String
IBM IAA Equivalent		Equivalent logical name in IBM IAA insurance industry model	String

8. Next enter a value for your new custom attribute **Data Sensitivity.** You will have to manage the term to do this.



IBM Information Server - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites
Address http://hawkdemo:9080/
Go Links
IBM. Information Server
Home Administration Glossary Information Services Catalog Reporting
Help | About | Change Password | Log Out
Browse Glossary Manage Custom Attributes Manage Terms Manage Categories
Select a Term to Manage
Properties
Related Terms
Synonyms
Classified Objects
Custom Attributes
Cancel Save Save and Close
Done
Local intranet

Edit Term - Add On Cover_Insurer Class_Code (Custom Attributes)

Custom attributes - Values of any user-defined attributes that are associated with this term

Item **1-4** out of **4** Items per Page: 10 25 50

Name	Value
Data Sensitivity	
IBM IAA Equivalent	
Line of Business	
Mainframe	