

Table of Contents

- 1 Summary of all PME jobs :.....1
- 2 Edit algorithm : eg Add Personal Email.....2
- 3 data load #1: mpxdata.....6
- 4 data load #2: mpxdata.....8
- 5 Generate BXM Data mpxfsdvd.....10
- 6 load data files to MDM DB : madunlload.....11
- 7 Generate Weights.....12
- 8 Run the Bulk Cross Match processes.....14
- 9 Perform Analysis Pairs Generation.....15

PME Process For PME Tuning : Specifically Targetted for MDM AE Projects

Ian Dallas

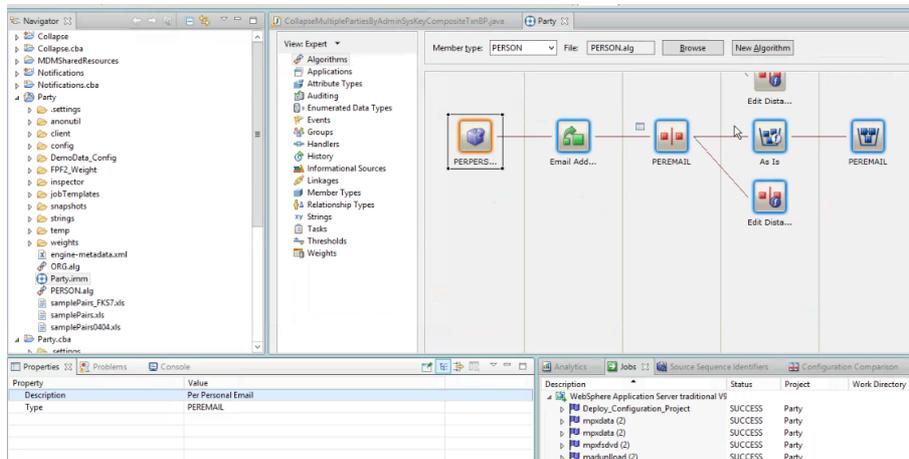
iandallas@uk.ibm.com

1 Summary of all PME jobs :

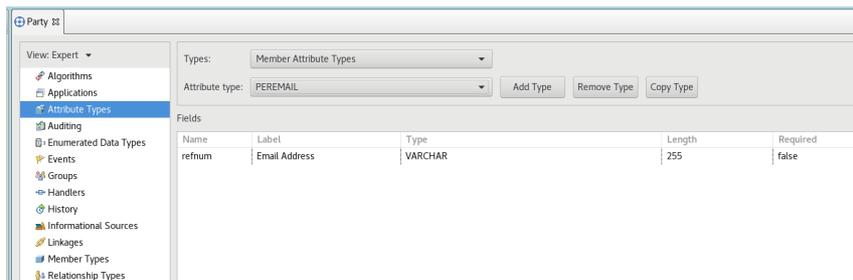
Description	Status	Project	Work Directory	Submitted ...	Create Time	Start Time	End Time	Total Time
WebSphere Application Server traditional V9								
Deploy_Configuration_Project	SUCCESS	Party		MDM6t6-d-...	04/04/19 09:10:20 PM	04/04/19 09:10:20 PM	04/04/19 09:11:53 PM	00:01:33.421
mpxdata (2)	SUCCESS	Party		MDM6t6-d-...	04/04/19 09:13:12 PM	04/04/19 09:13:12 PM	04/04/19 09:13:45 PM	00:00:33.362
mpxdata (2)	SUCCESS	Party		MDM6t6-d-...	04/04/19 09:14:09 PM	04/04/19 09:14:09 PM	04/04/19 09:14:56 PM	00:00:46.835
mpxfsdvd (2)	SUCCESS	Party		MDM6t6-d-...	04/04/19 09:15:42 PM	04/04/19 09:15:42 PM	04/04/19 09:17:01 PM	00:01:18.772
madunlload (2)	SUCCESS	Party		MDM6t6-d-...	04/04/19 09:17:15 PM	04/04/19 09:17:15 PM	04/04/19 09:35:18 PM	00:18:03.905
Generate_Weights (2)	SUCCESS	Party		MDM6t6-d-...	04/04/19 09:35:37 PM	04/04/19 09:35:37 PM	04/04/19 10:36:44 PM	01:01:06.568
Deploy_Configuration_Project	SUCCESS	Party		MDM6t6-d-...	04/04/19 11:01:58 PM	04/04/19 11:01:58 PM	04/04/19 11:03:30 PM	00:01:31.576
mpxcomp (6)	SUCCESS	Party		MDM6t6-d-...	04/04/19 11:06:49 PM	04/04/19 11:06:49 PM	04/04/19 11:45:57 PM	00:39:08.682
Threshold Analysis Pair Generation	SUCCESS	Party		MDM6t6-d-...	04/05/19 12:01:46 A...	04/05/19 12:01:46 A...	04/05/19 12:03:15 A...	00:01:29.269

2 Edit algorithm : eg Add Personal Email

Create a new swim lane containing person email PEREMAIL (and optionally other email types as required). If used for Householding also add a comparison step.

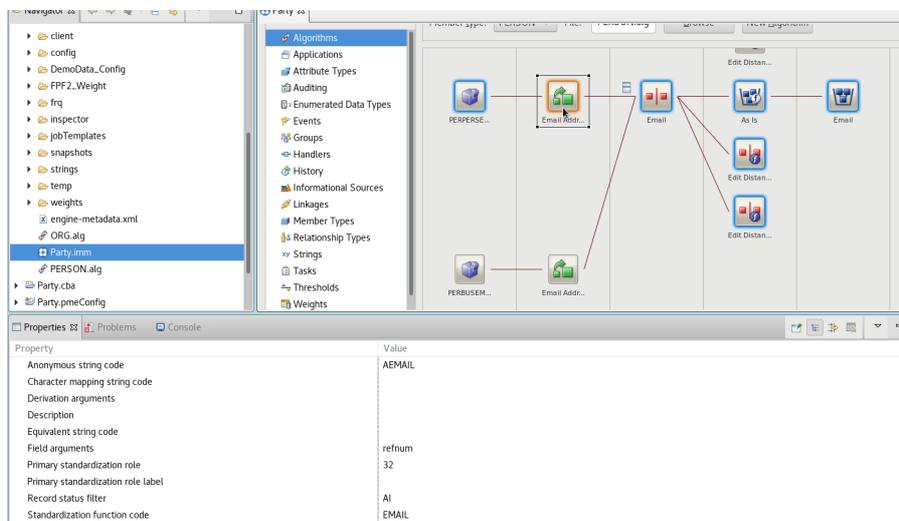


Change the PEREMAIL attribute type to “refnum” : this is required by MDM AE because email address is supplied using PartyContactMethodBOObj “refnumber” field (Usage Type 7)



Now set the properties of the various elements in the new swim lane.

Set field arguments to “refnum” for the email Standardiser



The screenshot shows the Party tool interface. The left sidebar contains a project tree with 'Party.inm' selected. The central panel displays a diagram with several nodes: 'PERPERSE...', 'Email Addr...', 'Email', 'As Is', and another 'Email'. The 'Email' node is highlighted with a red box. The right sidebar lists various tool categories like Algorithms, Attribute Types, etc. The bottom panel shows the properties for the selected 'Email' node.

Property	Value
Comparison role	13
Description	Email
Label	Email

The screenshot shows the Party tool interface. The left sidebar contains a project tree with 'Party.inm' selected. The central panel displays a diagram with several nodes: 'PERPERSE...', 'Email Addr.', 'Email', 'As Is', and another 'Email'. The 'Email' node is highlighted with a red box. The right sidebar lists various tool categories like Algorithms, Attribute Types, etc. The bottom panel shows the properties for the selected 'Email' node.

Property	Value
Anonymous string code	
Bucketing function code	ATTR
Derivation arguments	
Description	
Equivalent string code	
Generation function code	As Is
Maximum tokens	1
Minimum tokens	0

The screenshot shows the Party tool interface. The left sidebar contains a project tree with 'Party.inm' selected. The central panel displays a diagram with several nodes: 'PERPERSE...', 'Email Addr.', 'Email', 'As Is', and another 'Email'. The 'Email' node is highlighted with a red box. The right sidebar lists various tool categories like Algorithms, Attribute Types, etc. The bottom panel shows the properties for the selected 'Email' node.

Property	Value
Comparison arguments	
Comparison function code	DRIDIC
Comparison group	0
Comparison mode	Use for match, link and search
Comparison specification code	EMAIL
Description	
Enable Review Identifier Task	No
Entity type	mdnpr
Equivalent string code	
Kind	1 Role, 1 Dimension
Type	Full
Weight table percentage cut-off	80

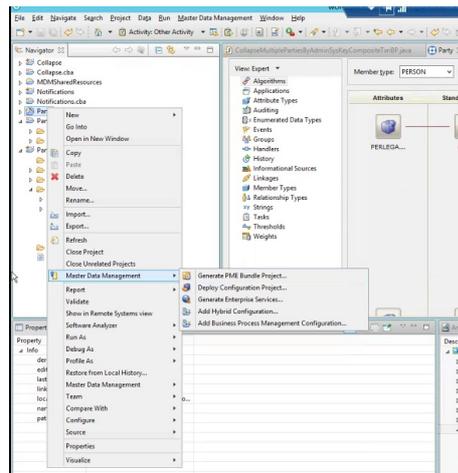
The screenshot shows a software interface with a diagram and a Properties window. The diagram features several nodes: PERPERSE, Email Addr, Email, and As Is. Lines connect these nodes, with a central 'Email' node being a hub. The Properties window is open, displaying a table of properties and their values.

Property	Value
Comparison arguments	DRID1C
Comparison function code	0
Comparison group	
Comparison mode	Use for match, link and search
Comparison specification code	HHEMAIL
Description	
Enable Review Identifier Task	No
Entity type	hh
Equivalent string code	
Kind	1 Role, 1 Dimension
Type	Full
Weight table percentage cut-off	80

This screenshot shows the same software interface as above, but with a different Properties window open. The diagram is visible in the background. The Properties window displays a table of properties and their values.

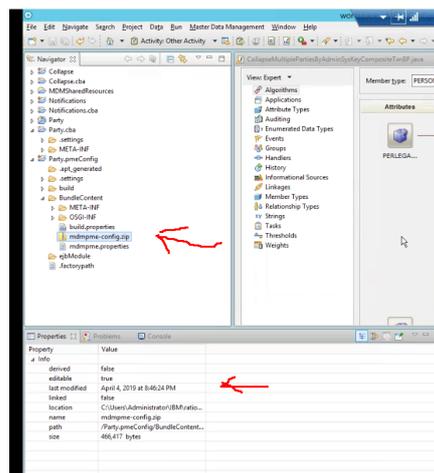
Property	Value
Bucketing role	9
Derivation group	9
Description	
Label	Email
Maximum attribute tokens	1
Maximum bucket size	1000
Minimum attribute tokens	1

Note : if you have already created the and associated Party.cba projects before altering the algorithm, then you will need to delete them and re-create.



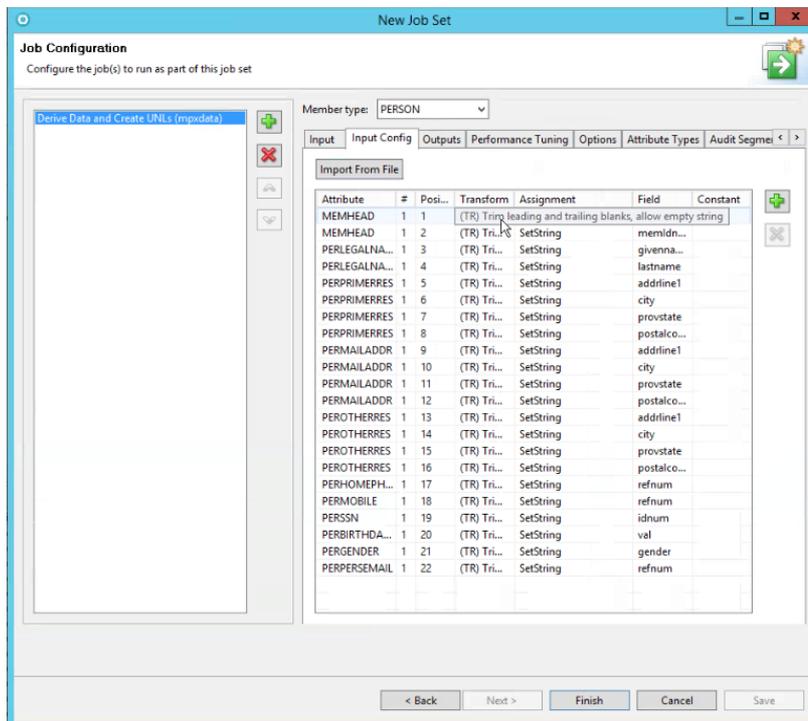
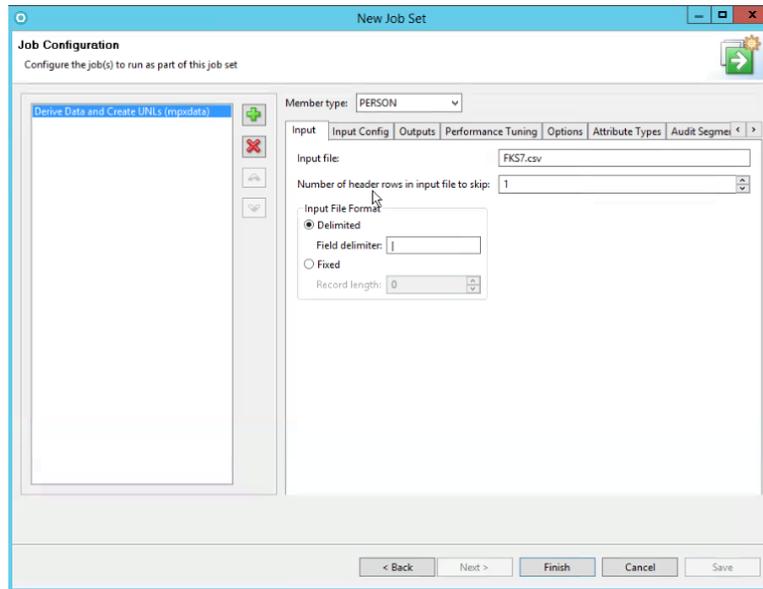
This is because the pmeConfig project with the mdmpme-config.zip file is only created once

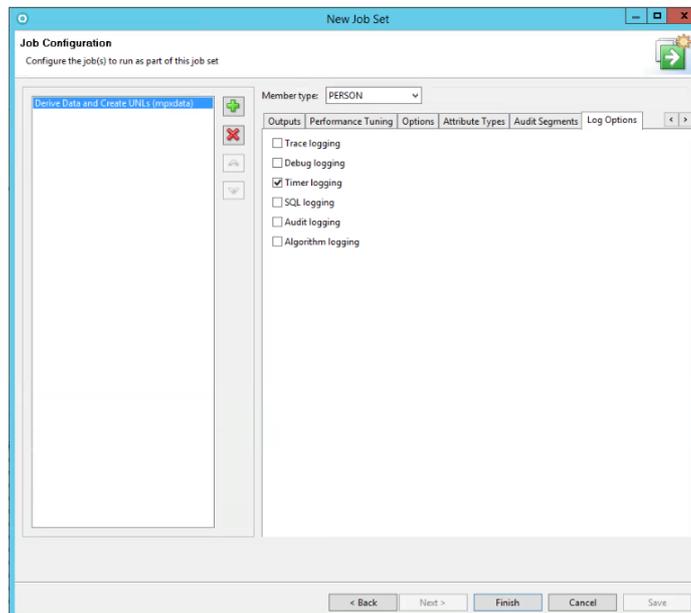
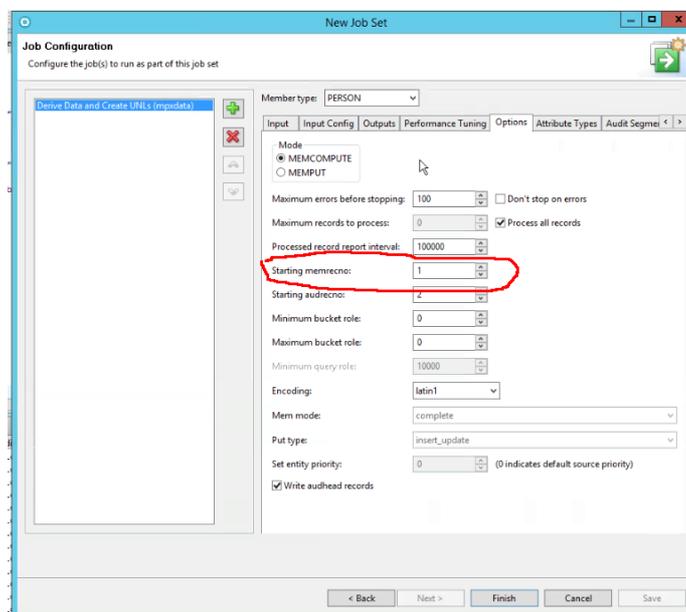
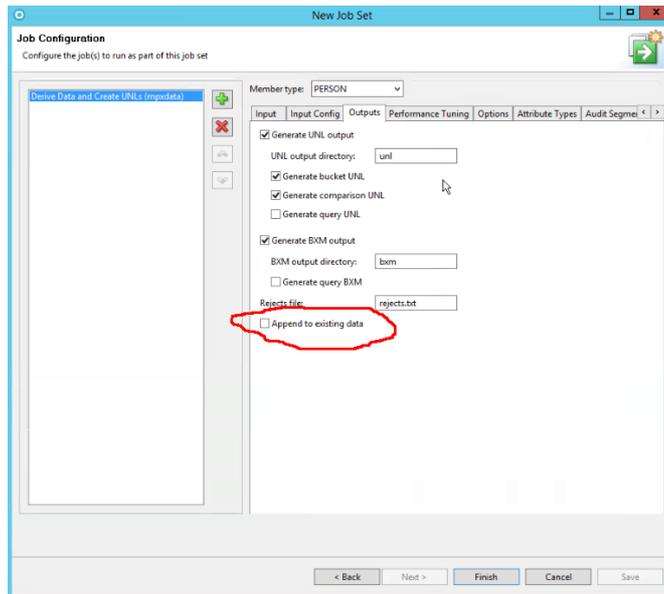
AND IT IS NEVER UPDATED WHEN YOU CHANGE THE ALGORITHM.



3 data load #1: mpxdata

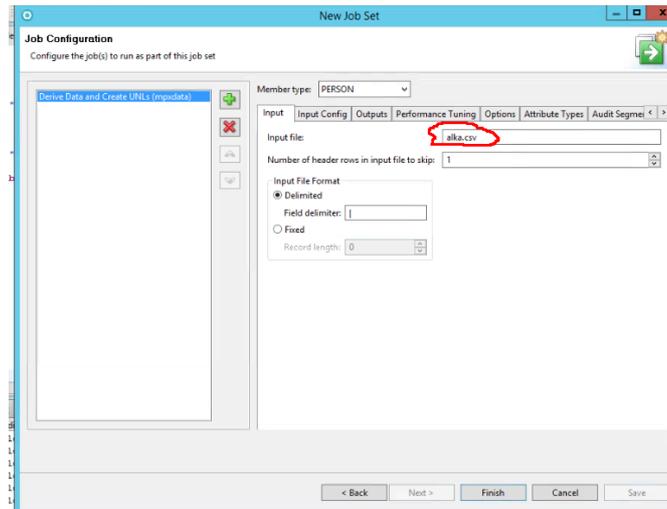
Load data from the 1st data source.



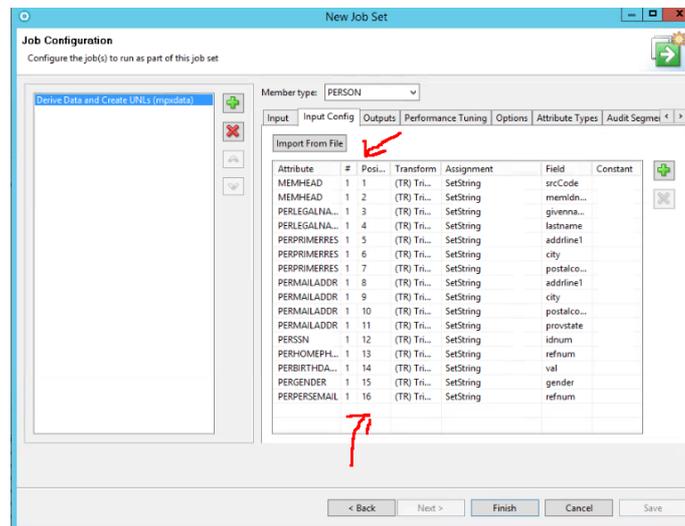


4 data load #2 mpxdata

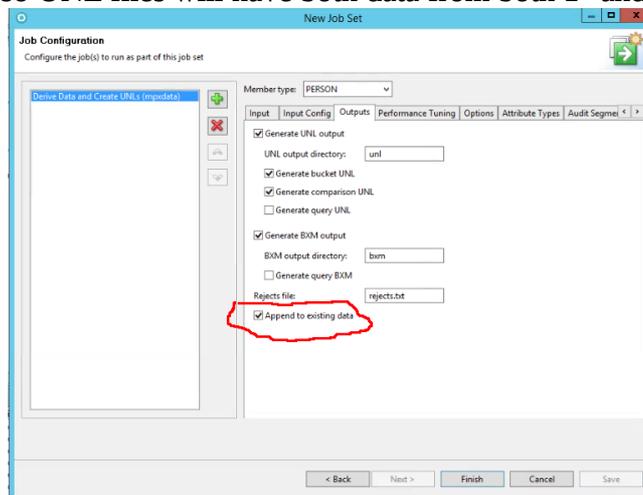
Load data from the 2nd data source.



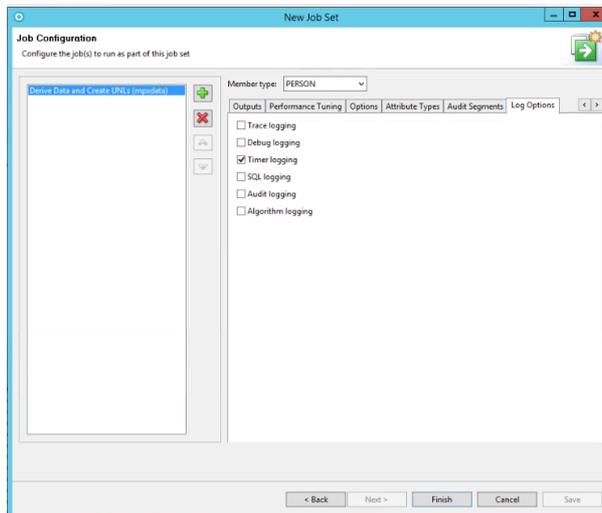
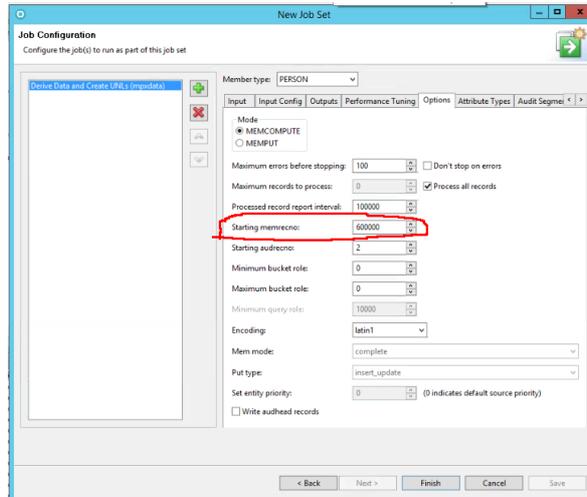
Be careful about the Position numbers : they can get messed up!



Append to existing data so UNL files will have both data from both 1st and 2nd data sources

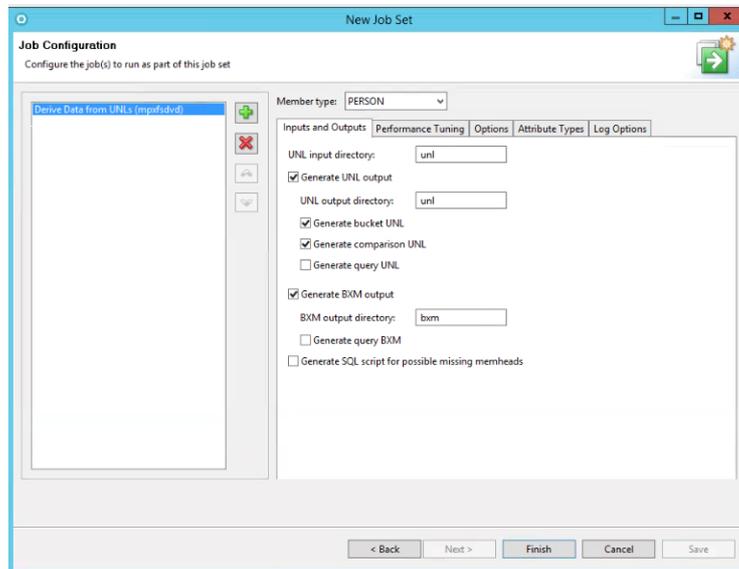


Be careful to set the “Starting memrecnum” so that it is slightly higher figure than the number of records read from the 1st data source. This will avoid error with duplicate rec numbers being loaded to the database in the next steps..



5 Generate BXM Data mpxfsdvd

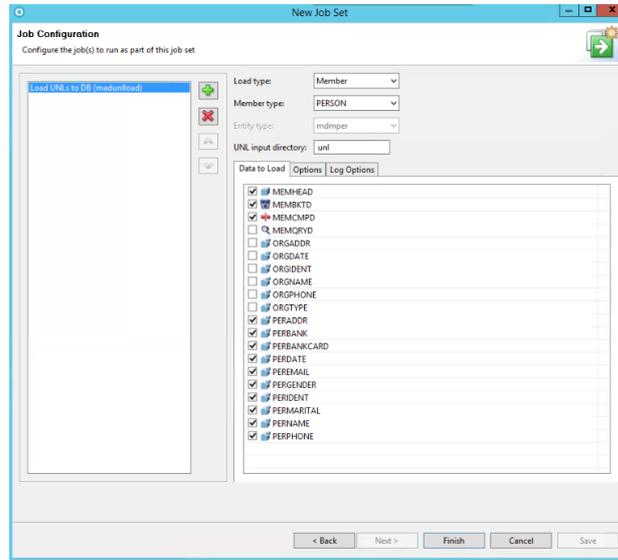
Because each mxpdata load overwrites the BXM data, we need to re-run to populate the BXM load files with data from both data sources.



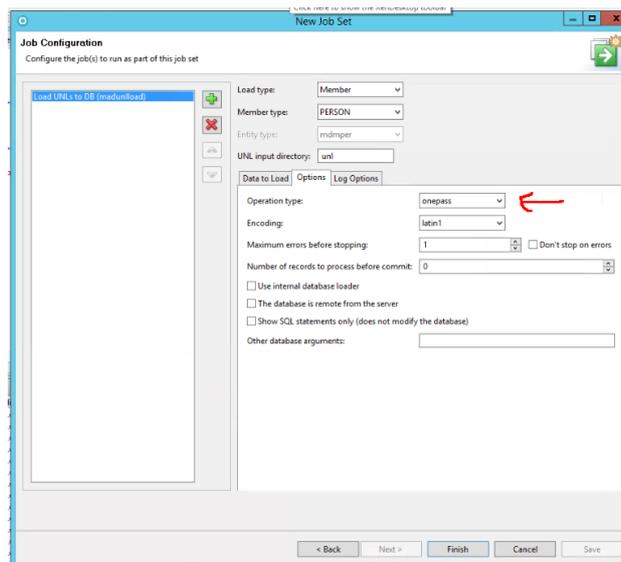
The screenshot shows the 'New Job Set' Job Configuration window. The window title is 'New Job Set' and the subtitle is 'Job Configuration'. Below the subtitle is the instruction 'Configure the job(s) to run as part of this job set'. On the left side, there is a list of data sources with the entry 'Derive Data from UNLs (mpxfsdvd)' selected. On the right side, there are several tabs: 'Inputs and Outputs', 'Performance Tuning', 'Options', 'Attribute Types', and 'Log Options'. The 'Inputs and Outputs' tab is active. The 'Member type' is set to 'PERSON'. The 'UNL input directory' is 'unl'. The 'Generate UNL output' checkbox is checked. The 'UNL output directory' is 'unl'. The 'Generate bucket UNL' checkbox is checked. The 'Generate comparison UNL' checkbox is checked. The 'Generate query UNL' checkbox is unchecked. The 'Generate BXM output' checkbox is checked. The 'BXM output directory' is 'bxm'. The 'Generate query BXM' checkbox is unchecked. The 'Generate SQL script for possible missing memheads' checkbox is unchecked. At the bottom of the window, there are buttons for '< Back', 'Next >', 'Finish', 'Cancel', and 'Save'.

6 load data files to MDM DB : madunload

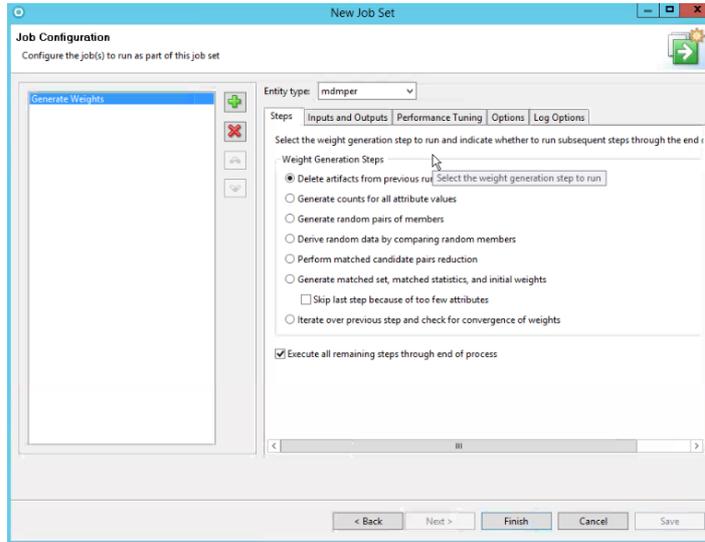
Load the data, buckets etc into the MDM database.



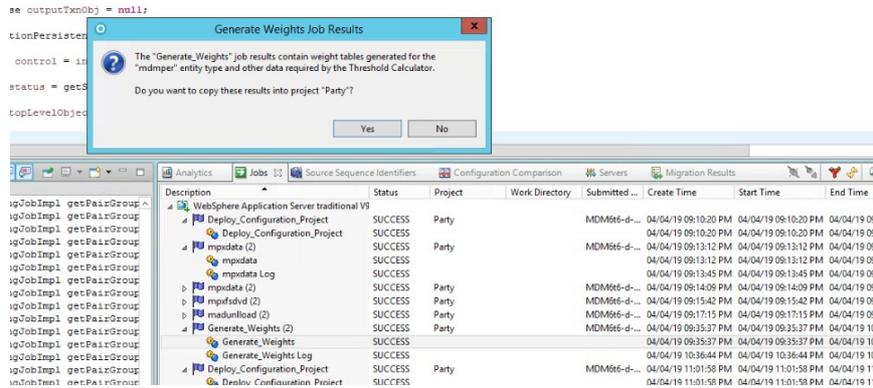
keep default “onepass” option : re-creates the virtual database tables, so don’t worry if previously loaded data is there.. it will be replaced.



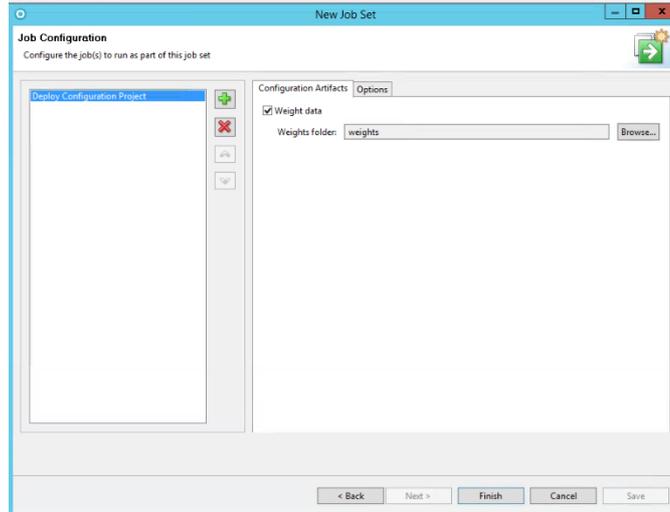
7 Generate Weights



let the process finish, then right click, “get job results” and save the weights in to the Party workspace project in the weights folder



Deploy the configuration again ie new weights get deployed:

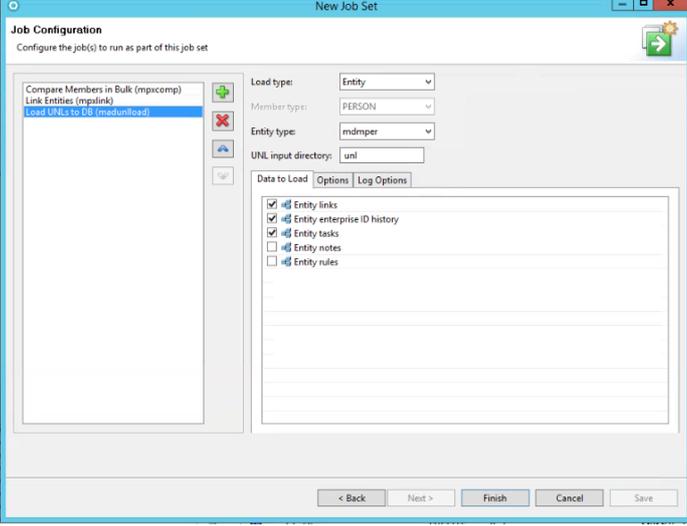
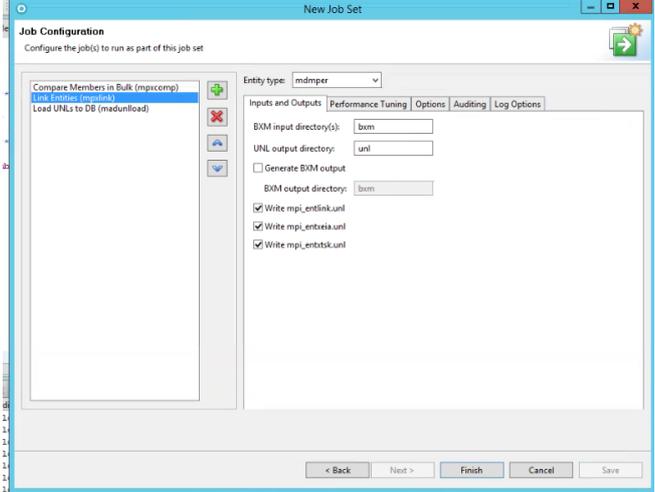
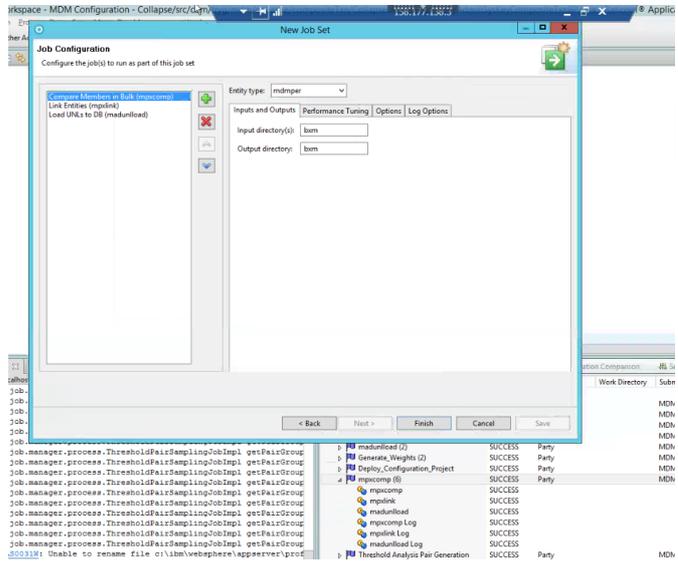


Some of the algorithm “swim lanes”, like business email, if not populated, will error out the weight gen process. So either remove un-needed swim lanes, or change its weight threshold to “1”.

Note : every change to the algorithm, must be followed by a Deploy Configuration and madunload. Re-run Weight Generation as required.

8 Run the Bulk Cross Match processes

(3 of them : mpxcomp mpxlink madunload): combine all 3 jobs and save as a job template (just like recommended for all the other single job processes)



9 Perform Analysis Pairs Generation..

When this process is completed, get the job results, and save as an XLS file which can most easily be analysed/processed using Excel.

The screenshot shows a software window titled "New Job Set" with a sub-header "Job Configuration". Below the sub-header is the instruction "Configure the job(s) to run as part of this job set".

The main configuration area includes the following fields and options:

- Entity type:** A dropdown menu set to "mdmper".
- Input directory:** A text input field containing "bom".
- Number of pairs per score:** A spin box set to "30".
- Include only cross source pairs:** An unchecked checkbox.
- Score Range:** Two spin boxes for "Minimum score" (set to 5.0) and "Maximum score" (set to 40.0).
- Result Order:** Two radio buttons: "Sorted by score" (selected) and "Random".
- Attributes to Return:** A list of attributes with checkboxes. The checked items are "Per DOB" and "Per Call Phone". Other visible items include "Per Abilitac", "Per AKA Name", "Per Alias Name", "Per American Express", "Per Birth Certificate", "Per Business Address", "Per Business Email", "Per Business Name", and "Per Business Phone".

At the bottom of the window are five buttons: "< Back", "Next >", "Finish", "Cancel", and "Save".