

Table of Contents

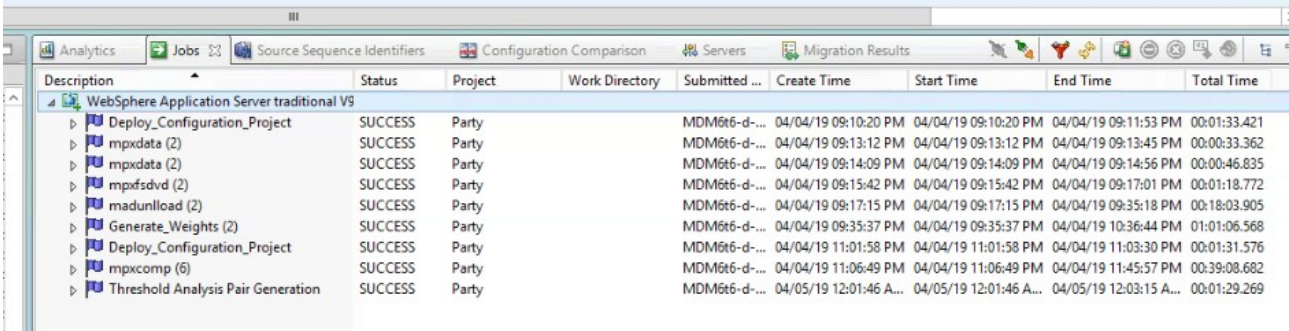
1 Summary of all PME jobs :.....	1
2 Edit algorithm : eg Add Personal Email.....	2
3 data load #1: mpxdata.....	6
4 data load #2 mpxdata.....	8
5 Generate BXM Data mpxfsdvd.....	10
6 load data files to MDM DB : madunlload.....	11
7 Generate Weights.....	12
8 Run the Bulk Cross Match processes.....	14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

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

Ian Dallas

iandallas@uk.ibm.com

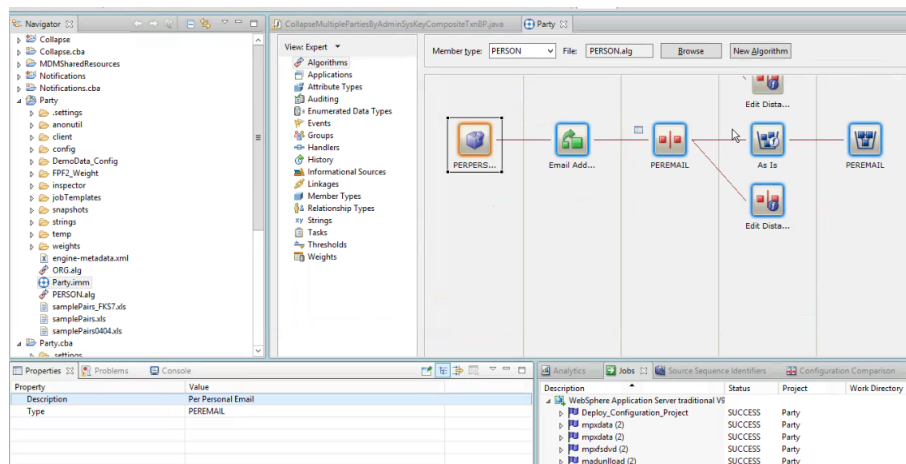
1 Summary of all PME jobs :



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.

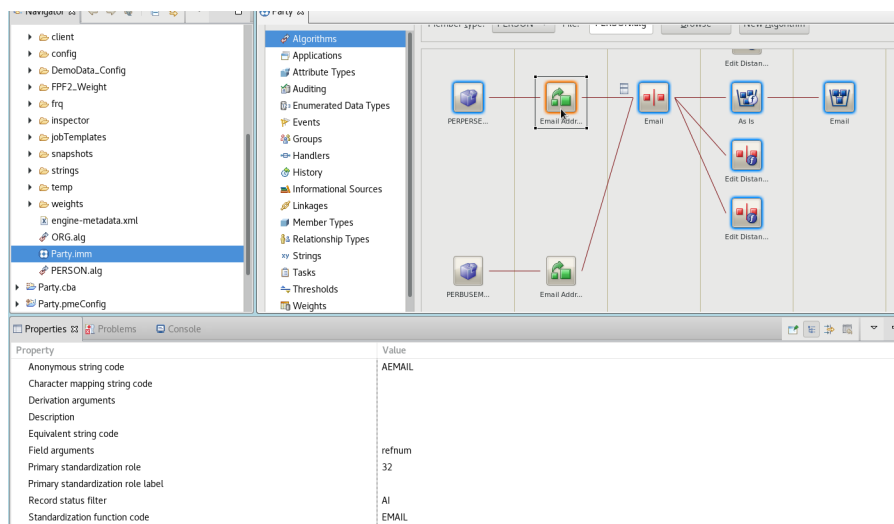


Change the PEREMAIL attribute type to “refnum” : this is required by MDM AE because email address is supplied using PartyContactMethodBOObj “refnumber” field (Usage Type 7)

The screenshot shows the MDM AE interface with the 'Party' swim lane selected. The 'View: Expert' pane displays the 'Attribute Types' section. The 'Types' dropdown is set to 'Member Attribute Types'. The 'Attribute type' is 'PEREMAIL'. The 'Fields' table shows the 'Name' as 'refnum', 'Label' as 'Email Address', 'Type' as 'VARCHAR', 'Length' as '255', and 'Required' as 'false'.

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

Set field arguments to “refnum” for the email Standardiser



Navigator

- client
- config
- DemoData_Config
- FFP2_Weight
- frq
- inspector
- jobTemplates
- snapshots
- strings
- temp
- weights
- engine-metadata.xml
- ORG.alg
- Party.imm**
- PERSON.alg
- Party.cba
- Party.pmeConfig

Party

- Algorithms
- Applications
- Attribute Types
- Auditing
- Enumerated Data Types
- Events
- Groups
- Handlers
- History
- Informational Sources
- Linkages
- Member Types
- Relationship Types
- Strings
- Tasks
- Thresholds
- Weights

PERPERSE... Email Addr... Email As Is Email

PERBUSEM... Email Addr...

Properties

Property	Value
Comparison role	13
Description	Email
Label	Email

Navigator

- client
- config
- DemoData_Config
- FFP2_Weight
- frq
- inspector
- jobTemplates
- snapshots
- strings
- temp
- weights
- engine-metadata.xml
- ORG.alg
- Party.imm**
- PERSON.alg
- Party.cba
- Party.pmeConfig

Party

- Algorithms
- Applications
- Attribute Types
- Auditing
- Enumerated Data Types
- Events
- Groups
- Handlers
- History
- Informational Sources
- Linkages
- Member Types
- Relationship Types
- Strings
- Tasks
- Thresholds
- Weights

PERPERSE... Email Addr... Email As Is Email

PERBUSEM... Email Addr...

Properties

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

Navigator

- client
- config
- DemoData_Config
- FFP2_Weight
- frq
- inspector
- jobTemplates
- snapshots
- strings
- temp
- weights
- engine-metadata.xml
- ORG.alg
- Party.imm**
- PERSON.alg
- Party.cba
- Party.pmeConfig

Party

- Algorithms
- Applications
- Attribute Types
- Auditing
- Enumerated Data Types
- Events
- Groups
- Handlers
- History
- Informational Sources
- Linkages
- Member Types
- Relationship Types
- Strings
- Tasks
- Thresholds
- Weights

PERPERSE... Email Addr... Email As Is Email

PERBUSEM... Email Addr...

Properties

Property	Value
Comparison arguments	
Comparison function code	DR1D1C
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

Navigator

- client
- config
- DemoData_Config
- FPF2_Weight
- frq
- Inspector
- jobTemplates
- snapshots
- strings
- temp
- weights
 - engine-metadata.xml
 - ORG.alg
 - Party.inm
 - PERSON.alg

Party

- Applications
- Attribute Types
- Auditing
- Enumerated Data Types
- Events
- Groups
- Handlers
- History
- Informational Sources
- Linkages
- Member Types
- Relationship Types
- Strings
- Tasks
- Thresholds
- Weights

Properties

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

Navigator

- client
- config
- DemoData_Config
- FPF2_Weight
- frq
- Inspector
- jobTemplates
- snapshots
- strings
- temp
- weights
 - engine-metadata.xml
 - ORG.alg
 - Party.inm
 - PERSON.alg

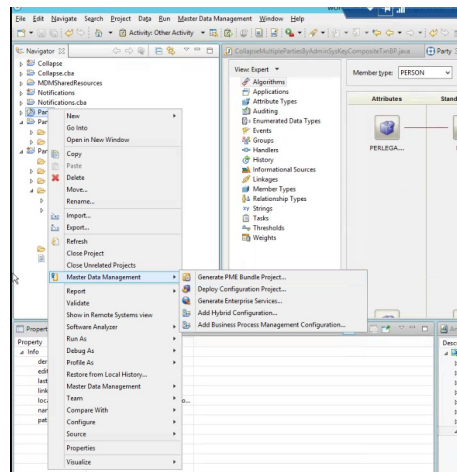
Party

- Applications
- Attribute Types
- Auditing
- Enumerated Data Types
- Events
- Groups
- Handlers
- History
- Informational Sources
- Linkages
- Member Types
- Relationship Types
- Strings
- Tasks
- Thresholds
- Weights

Properties

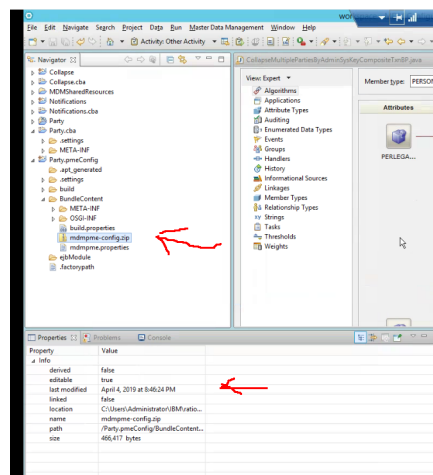
Property	Value
Bucketing role	9
Derivation group	9
Description	
Label	Email
Maximum attribute tokens	1
Maximum bucket size	1000
Minimum attribute tokens	1

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.

New Job Set

Job Configuration
Configure the job(s) to run as part of this job set

Member type: PERSON

Input file: FKS7.csv

Number of header rows in input file to skip: 1

Input File Format

☒ Delimited

Field delimiter: |

☐ Fixed

Record length: 0

< Back Next > Finish Cancel Save

New Job Set

Job Configuration
Configure the job(s) to run as part of this job set

Member type: PERSON

Input Input Config Outputs Performance Tuning Options Attribute Types Audit Segmei < >

Import From File

Attribute	#	Posi...	Transform	Assignment	Field	Constant
MEMHEAD	1	1	(TR) Trim leading and trailing blanks, allow empty string			
MEMHEAD	1	2	(TR) Trim...	SetString	memidn...	
PERLEGALNA...	1	3	(TR) Trim...	SetString	givenna...	
PERLEGALNA...	1	4	(TR) Trim...	SetString	lastname	
PERPRIMERRES	1	5	(TR) Trim...	SetString	addline1	
PERPRIMERRES	1	6	(TR) Trim...	SetString	city	
PERPRIMERRES	1	7	(TR) Trim...	SetString	provstate	
PERPRIMERRES	1	8	(TR) Trim...	SetString	postalco...	
PERMAILADDR	1	9	(TR) Trim...	SetString	addline1	
PERMAILADDR	1	10	(TR) Trim...	SetString	city	
PERMAILADDR	1	11	(TR) Trim...	SetString	provstate	
PERMAILADDR	1	12	(TR) Trim...	SetString	postalco...	
PEROTHERRES	1	13	(TR) Trim...	SetString	addline1	
PEROTHERRES	1	14	(TR) Trim...	SetString	city	
PEROTHERRES	1	15	(TR) Trim...	SetString	provstate	
PEROTHERRES	1	16	(TR) Trim...	SetString	postalco...	
PERHOMEPH...	1	17	(TR) Trim...	SetString	refnum	
PERMOBILE	1	18	(TR) Trim...	SetString	refnum	
PERSSN	1	19	(TR) Trim...	SetString	idnum	
PERBIRTHDA...	1	20	(TR) Trim...	SetString	val	
PERGENDER	1	21	(TR) Trim...	SetString	gender	
PERPERSEMAIL	1	22	(TR) Trim...	SetString	refnum	

< Back Next > Finish Cancel Save

New Job Set

Job Configuration
Configure the job(s) to run as part of this job set

Member type: PERSON

Input Input Config Outputs Performance Tuning Options Attribute Types Audit Segme

☒ Generate UNL output
UNL output directory: unl

☒ Generate bucket UNL
☒ Generate comparison UNL
☐ Generate query UNL

☒ Generate BXM output
BXM output directory: bxm

☐ Generate query BXM

Rejects file: rejects.txt

☐ Append to existing data

< Back Next > Finish Cancel Save

New Job Set

Job Configuration
Configure the job(s) to run as part of this job set

Member type: PERSON

Input Input Config Outputs Performance Tuning Options Attribute Types Audit Segme

Mode
☒ MEMCOMPUTE
☐ MEMPUT

Maximum errors before stopping: 100 ☐ Don't stop on errors

Maximum records to process: 0 ☒ Process all records

Processed record report interval: 100000

Starting memreco: 1

Starting audreco: 2

Minimum bucket role: 0

Maximum bucket role: 0

Minimum query role: 10000

Encoding: latin1

Mem mode: complete

Put type: insert_update

Set entity priority: 0 (0 indicates default source priority)

☒ Write audhead records

< Back Next > Finish Cancel Save

New Job Set

Job Configuration
Configure the job(s) to run as part of this job set

Member type: PERSON

Outputs Performance Tuning Options Attribute Types Audit Segments Log Options

☐ Trace logging
☐ Debug logging
☒ Timer logging
☐ SQL logging
☐ Audit logging
☐ Algorithm logging

< Back Next > Finish Cancel Save

4 data load #2 mpxdata

Load data from the 2nd data source.

Job Configuration
Configure the job(s) to run as part of this job set

Member type: PERSON

Input Input Config Outputs Performance Tuning Options Attribute Types Audit Segmei < >

Input file: alka.csv

Number of header rows in input file to skip: 1

Input File Format
☒ Delimited
☐ Fixed

Field delimiter: |

Record length: 0

< Back Next > Finish Cancel Save

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

Job Configuration
Configure the job(s) to run as part of this job set

Member type: PERSON

Input Input Config Outputs Performance Tuning Options Attribute Types Audit Segmei < >

Import From File

Attribute	Posi...	Transform	Assignment	Field	Constant
MEMHEAD	1 1	(TR) Tri...	SetString	srcCode	
MEMHEAD	1 2	(TR) Tri...	SetString	memidn...	
PERLEGALNA...	1 3	(TR) Tri...	SetString	givenna...	
PERLEGALNA...	1 4	(TR) Tri...	SetString	lastname	
PERPRIMERRES	1 5	(TR) Tri...	SetString	addrlne1	
PERPRIMERRES	1 6	(TR) Tri...	SetString	city	
PERPRIMERRES	1 7	(TR) Tri...	SetString	postalco...	
PERMAILADDR	1 8	(TR) Tri...	SetString	addrlne1	
PERMAILADDR	1 9	(TR) Tri...	SetString	city	
PERMAILADDR	1 10	(TR) Tri...	SetString	postalco...	
PERMAILADDR	1 11	(TR) Tri...	SetString	provstate	
PERSN	1 12	(TR) Tri...	SetString	idnum	
PERHOMEPH...	1 13	(TR) Tri...	SetString	refnum	
PERBIRTHDA...	1 14	(TR) Tri...	SetString	val	
PERGENDER	1 15	(TR) Tri...	SetString	gender	
PERPERSEMAIL	1 16	(TR) Tri...	SetString	refnum	

< Back Next > Finish Cancel Save

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

Job Configuration
Configure the job(s) to run as part of this job set

Member type: PERSON

Input Input Config Outputs Performance Tuning Options Attribute Types Audit Segmei < >

☒ Generate UNL output

UNL output directory: unl

☒ Generate bucket UNL

☒ Generate comparison UNL

☐ Generate query UNL

☒ Generate BXM output

BXM output directory: bom

☐ Generate query BXM

Rejects file: rejects.txt

☒ Append to existing data

< Back Next > Finish Cancel Save

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..

The screenshot shows the 'New Job Set' dialog box with the 'Job Configuration' tab selected. The 'Member type' is set to 'PERSON'. The 'Mode' is set to 'MEMCOMPUTE'. The 'Maximum errors before stopping' is 100. The 'Maximum records to process' is 0. The 'Processed record report interval' is 100000. The 'Starting memrecnum' is 600000, which is highlighted with a red circle. The 'Starting audrecnum' is 2. The 'Minimum bucket role' is 0. The 'Maximum bucket role' is 0. The 'Minimum query role' is 10000. The 'Encoding' is 'latin1'. The 'Mem mode' is 'complete'. The 'Put type' is 'insert_update'. The 'Set entity priority' is 0. The 'Write audthead records' checkbox is unchecked. The 'Finish' button is highlighted.

The screenshot shows the 'New Job Set' dialog box with the 'Job Configuration' tab selected. The 'Log Options' sub-tab is selected. The 'Member type' is set to 'PERSON'. The 'Trace logging' checkbox is unchecked. The 'Debug logging' checkbox is unchecked. The 'Timer logging' checkbox is checked. The 'SQL logging' checkbox is unchecked. The 'Audit logging' checkbox is unchecked. The 'Algorithm logging' checkbox is unchecked. The 'Finish' button is highlighted.

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.

New Job Set

Job Configuration
Configure the job(s) to run as part of this job set

Member type: PERSON

Inputs and Outputs | Performance Tuning | Options | Attribute Types | Log Options

UNL input directory: unl

☒ Generate UNL output

UNL output directory: unl

☒ Generate bucket UNL

☒ Generate comparison UNL

☐ Generate query UNL

☒ Generate BXM output

BXM output directory: bxm

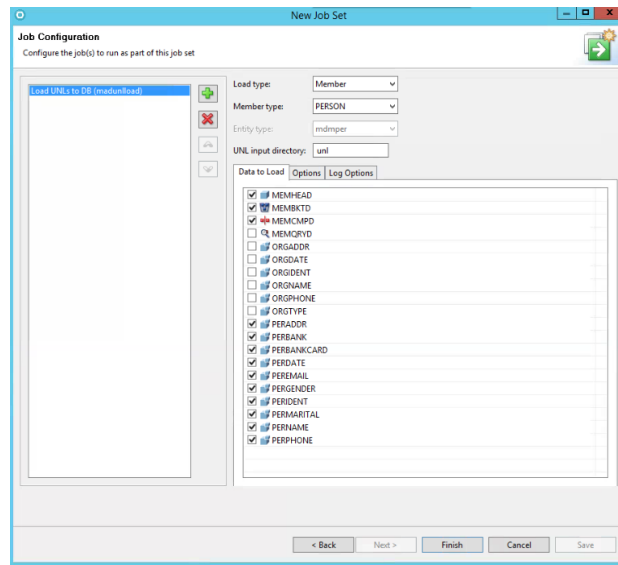
☐ Generate query BXM

☐ Generate SQL script for possible missing memheads

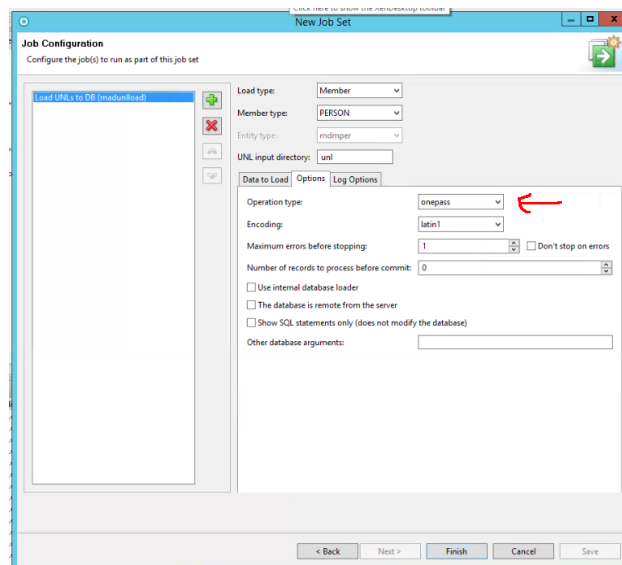
< Back Next > Finish Cancel Save

6 load data files to MDM DB : madunload

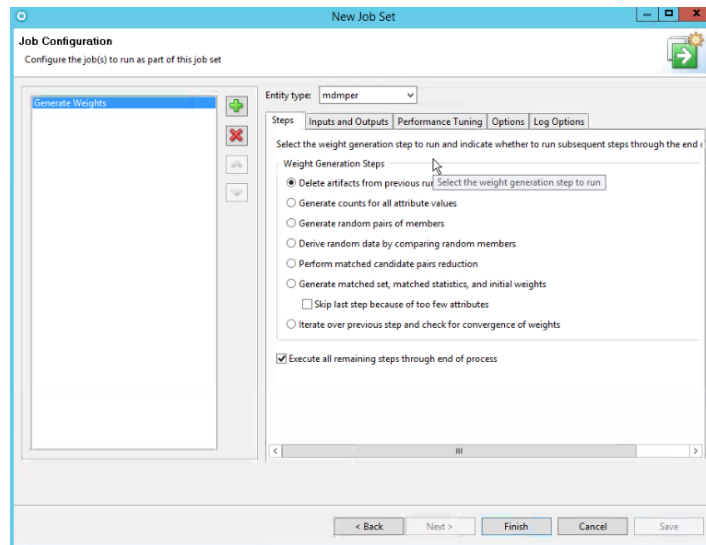
Load the data, buckets etc into the MDM database.



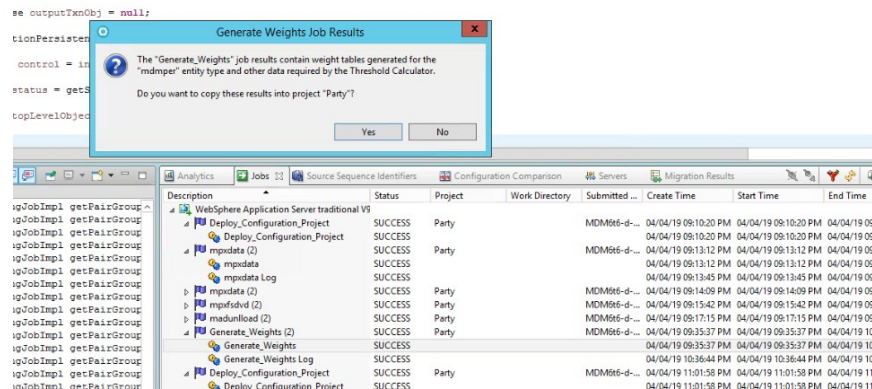
keep default “onepass” option : re-creates the virtual database tables, so don’t worry if previously loaded data is there.. it will be replaced.



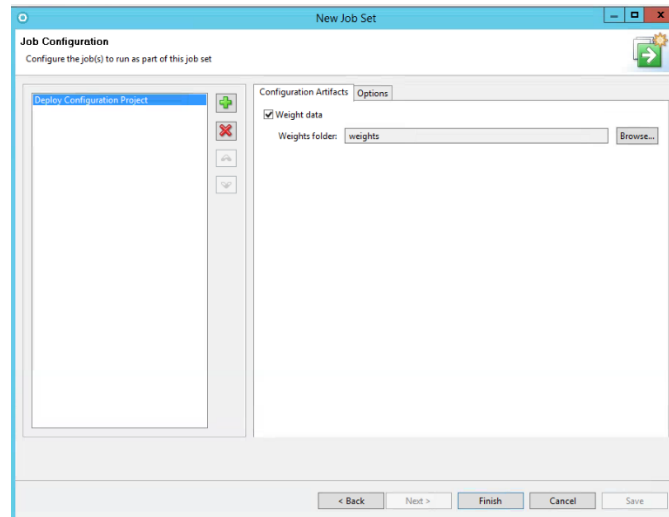
7 Generate Weights



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:

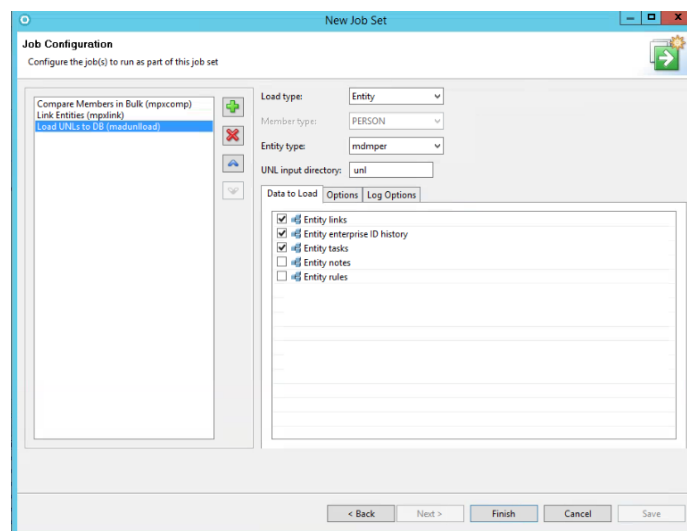
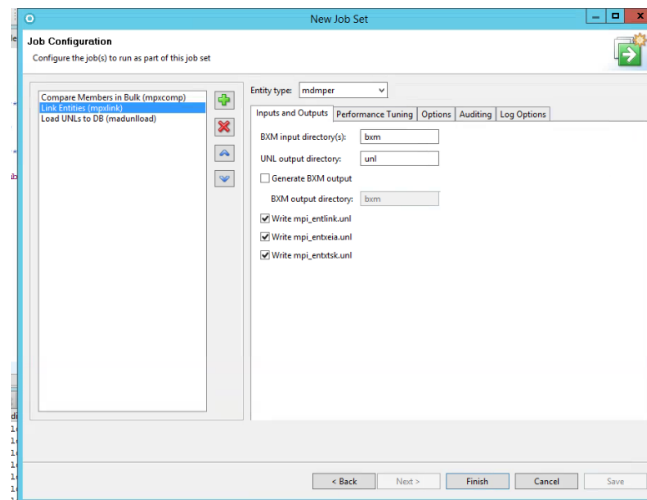
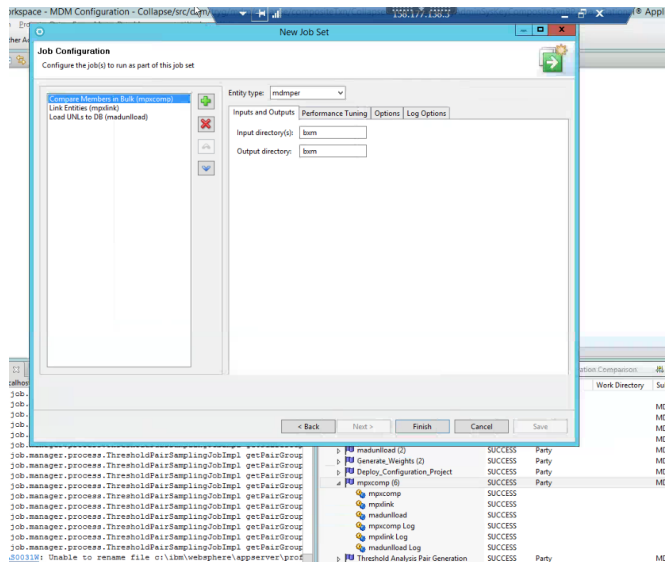


Some of the algorithm “swim lanes”, like business email, if not populated, will error out the weight gen process. So either remove un-needed swim lanes, or change its weight threshold to “1”.

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

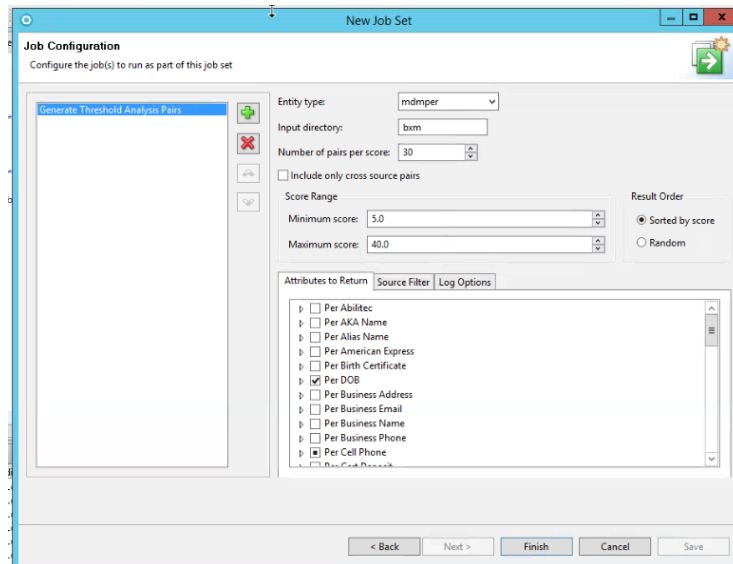
8 Run the Bulk Cross Match processes

(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 the 'New Job Set' window with the 'Job Configuration' tab selected. The window title is 'New Job Set'. Below the title bar, there is a 'Job Configuration' section with the instruction 'Configure the job(s) to run as part of this job set'. On the left, there is a 'Generate Threshold Analysis Pairs' button with a green plus icon. The main configuration area includes the following fields and options:

- Entity type:** A dropdown menu set to 'mdmper'.
- Input directory:** A text field containing 'bim'.
- Number of pairs per score:** A text field containing '30'.
- Include only cross source pairs:** An unchecked checkbox.
- Score Range:** Two text fields for 'Minimum score' (5.0) and 'Maximum score' (40.0).
- Result Order:** Two radio buttons: 'Sorted by score' (selected) and 'Random'.
- Attributes to Return:** A list of attributes with checkboxes:
 - ☐ Per Abiltec
 - ☐ Per AKA Name
 - ☐ Per Alias Name
 - ☐ Per American Express
 - ☐ Per Birth Certificate
 - ☒ Per DOB
 - ☐ Per Business Address
 - ☐ Per Business Email
 - ☐ Per Business Name
 - ☐ Per Business Phone
 - ☒ Per Cell Phone

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