Deduplicate and consolidate your data from multiple data sources using match and merge in Informatica MDM SaaS

Data Sheet | LumenData

Compare your data to identify similar records and combine them into a single, reliable representation of the entity.

PRODUCTS EXPLAINED:

1. Informatica MDM SaaS (IICS account)
2. Business 360 Console and
3. Customer 360

(The user must have the required privileges to access the mentioned services in IICS)

STEP BY STEP GUIDE TO IDENTIFY AND MERGE DUPLICATE OR SIMILAR RECORDS WITHIN THE MASTER DATA

1.CREATING RECORDS

After logging into our IICS account, we navigate to Customer 360.
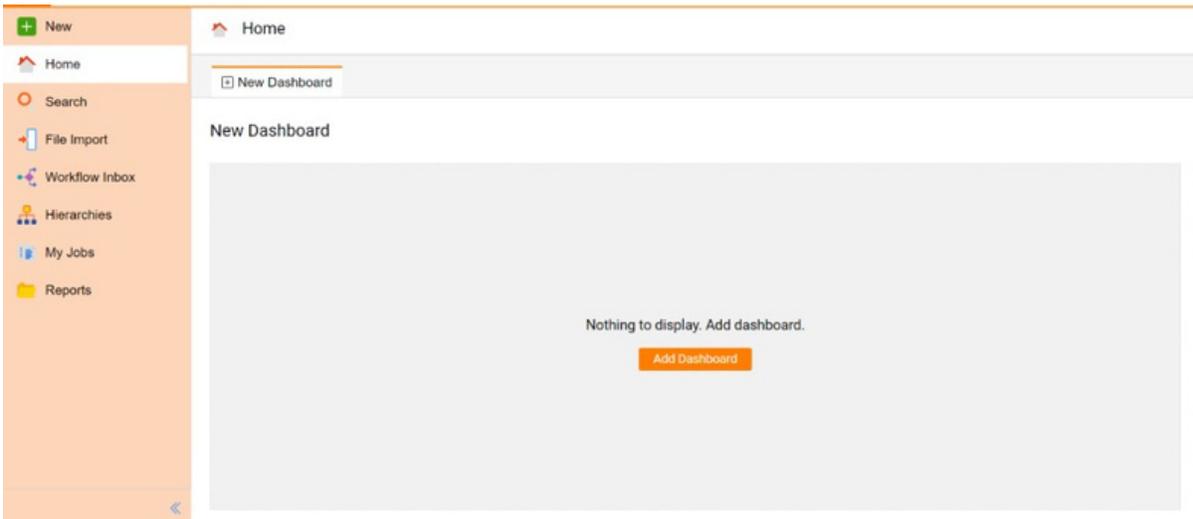
Click on '**new**' and select the record type you want to create for the selected business entity.
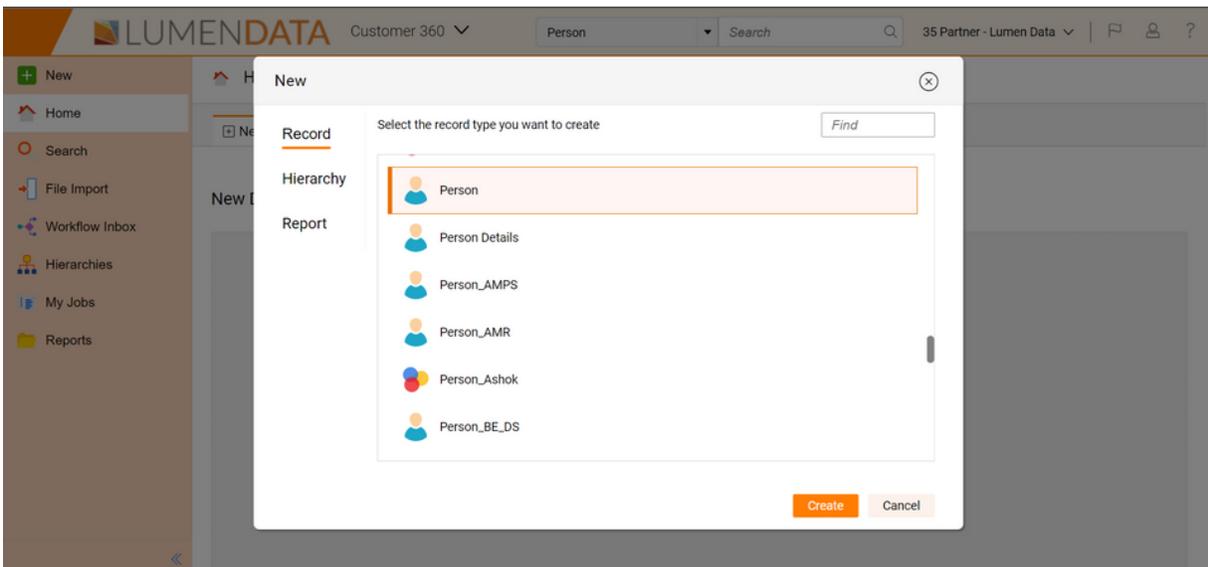
A new window will pop up after you click on create as shown below picture including all the data fields. Fill in all the necessary information and submit the changes which you have made.

After submitting and approving the changes the following window will show up with the new record with all the data fields entered by the user.

NOTE:

Data can also be entered without manual practices using an ingress job that loads the data from source file to the desired business entity using the required connections. This process can be divided into four steps.

 • **Mapping:** Define the source and target connections using the appropriate transformations in data integration.

 • **Mapping Task:** To process the data based on the data flow logic defined in the mapping in data integration.

 • **Task Flow:** Control the execution of the mapping task in data integration.

 • **Ingress Job:** Define a job in the Business 360 Console to run the task flow and load the data into MDM.

2. DEFINING MATCH AND MERGE RULES FOR THE BUSINESS ENTITY IN USE:

In Business 360 Console, go to explore and search for the business entity for which you're creating the record.
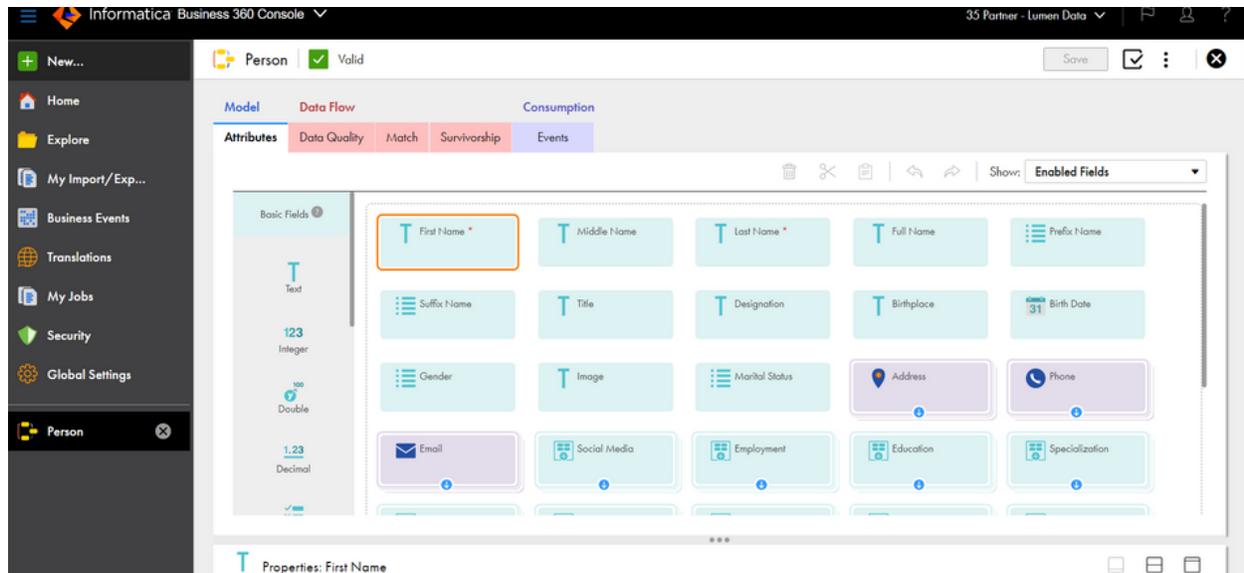
For this case, we are using the business entity as shown above. After creating all the records in Customer 360, define the match rules which we need to work on to match the records. Navigate to the match tab and select the name of the match model which is to be used for the match and merge process.

LUMENDATA
Masters in Data Management

In case there is no suitable match model available to perform the match and merge operation, we will define the model as per our requirement as shown below.

1. Click on the + (plus) symbol to add to a new model.
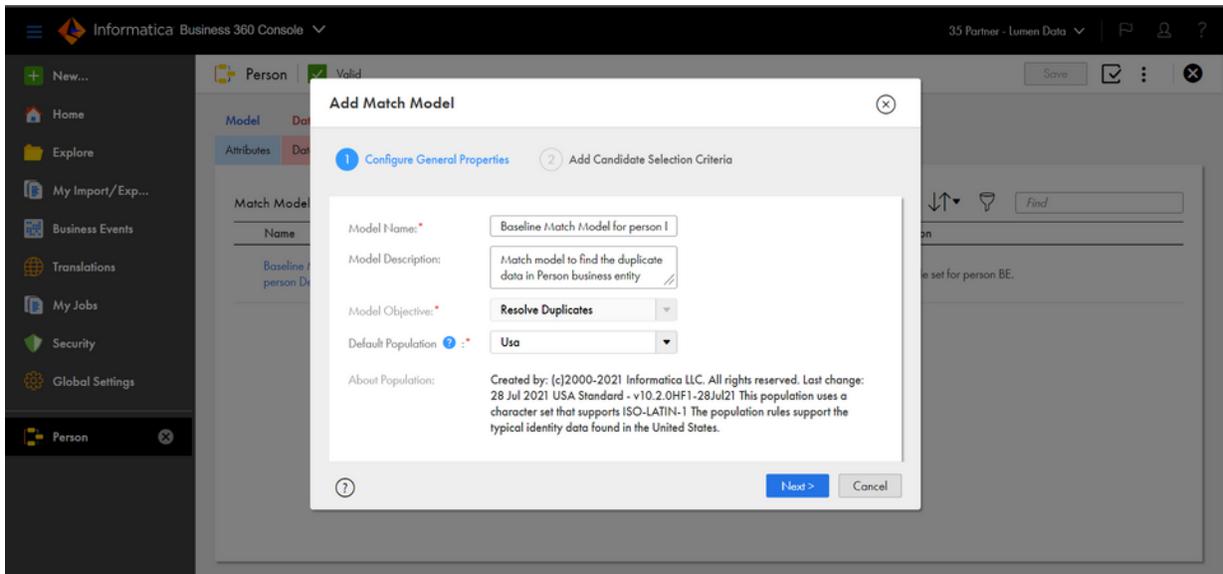
The window shown above will appear where we need to provide the details of the match model. Fill the field as per requirement.

2. Add the candidate selection criteria required for our matching conditions. You can provide multiple criteria for a single match model.
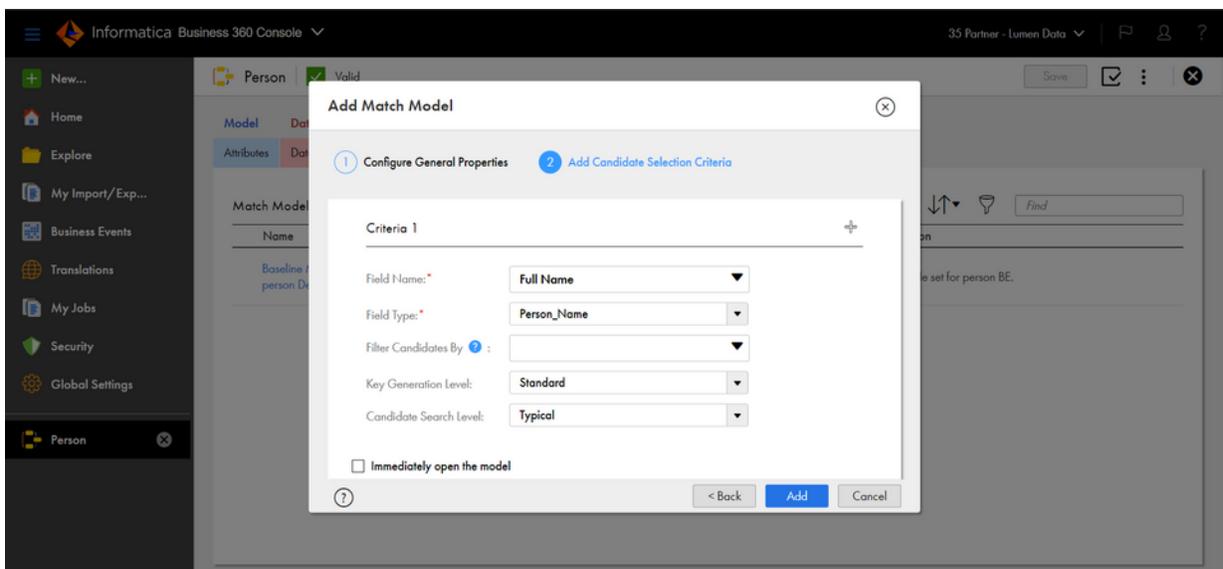
3. Define declarative rules to configure the match properties. Select the match strategy intended to use according to the requirement.
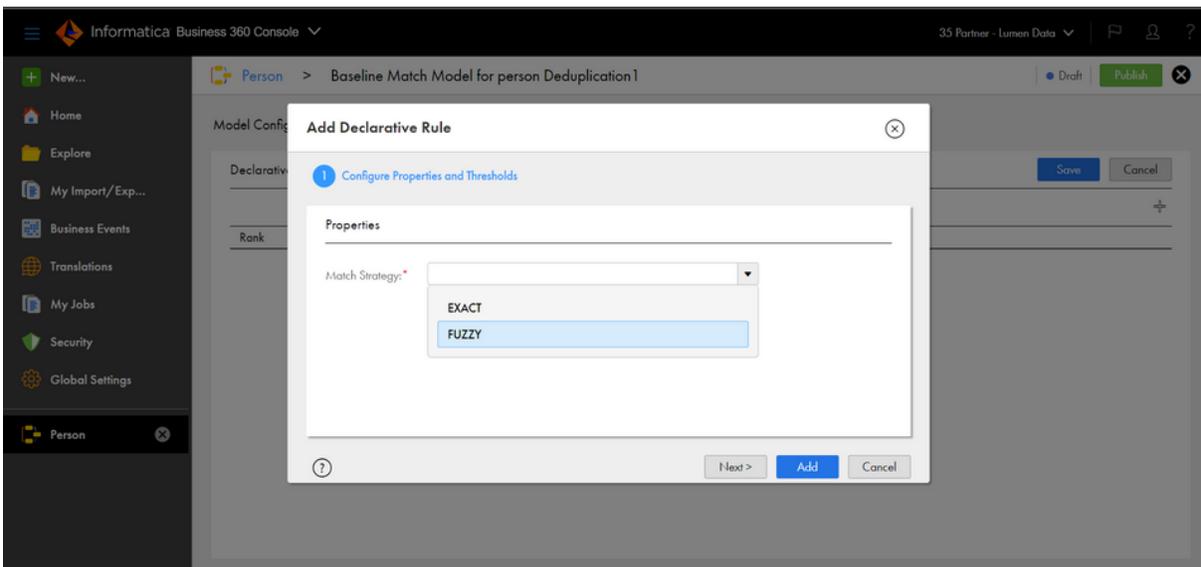
We have two match strategies named **Fuzzy match** and **Exact match.**

Exact Match: A match strategy that matches only records that are identical. You can perform exact matching on any data type.

Fuzzy Match: A match strategy that uses probabilistic matching, which considers spelling variations, possible misspellings, and other differences that can make matching records non-identical. You can perform fuzzy matching on any data type.

After configuring the match strategy, select the merge strategy we need to implement.
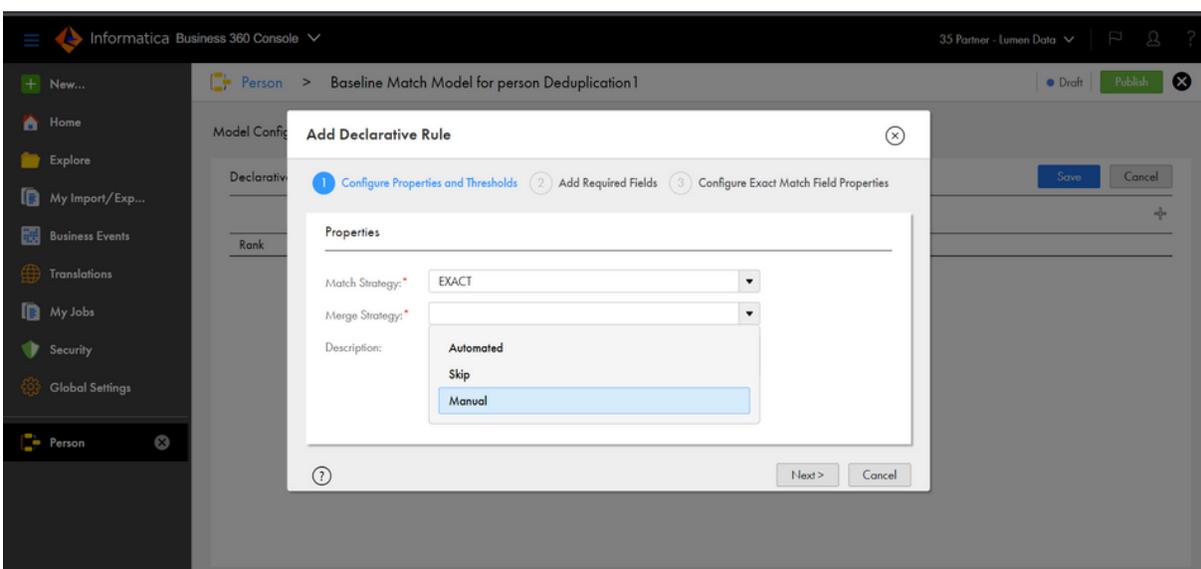
The three merge strategies available are as follows:

- **Automated**: Merge the record automatically after successful matching.

- **Skip**: Skip the merging of record.

- **Manual**: Manually merge the records after applying the match strategy.

4. Add the required fields for comparison: **Note: At least one exact field match field is required.** Define all the field names for the exact match fields and select the comparison outcome as identical for exact match as shown below.
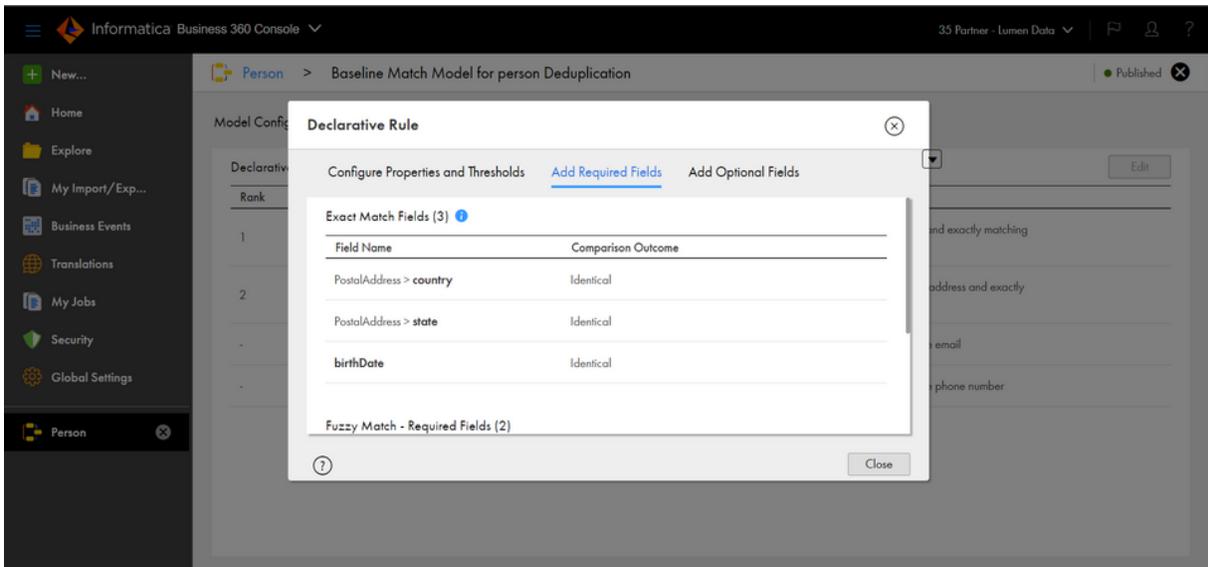
Configure the properties and thresholds to be used in case of a fuzzy match. Define the match strategy, merge strategy, match criteria, match level, and description if required. Define the merge threshold for the selected merge strategy, the merge strategy value can be anything from 0 to 100. However, keep the merge threshold between 90 to 100 for better performance.
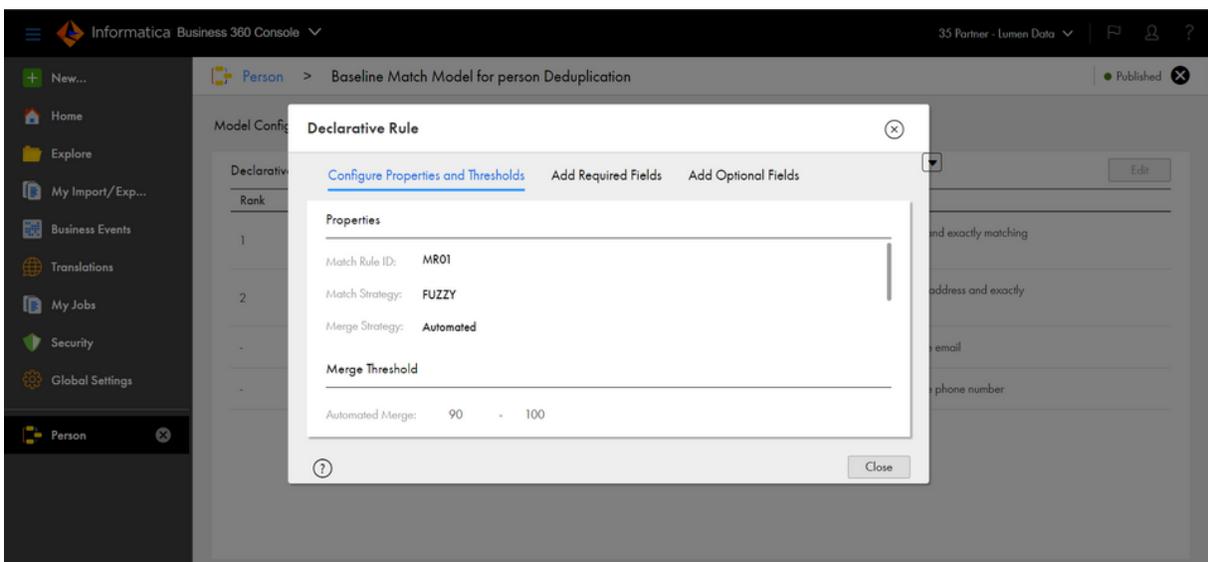
Add the required fields for exact and fuzzy matches as shown below,
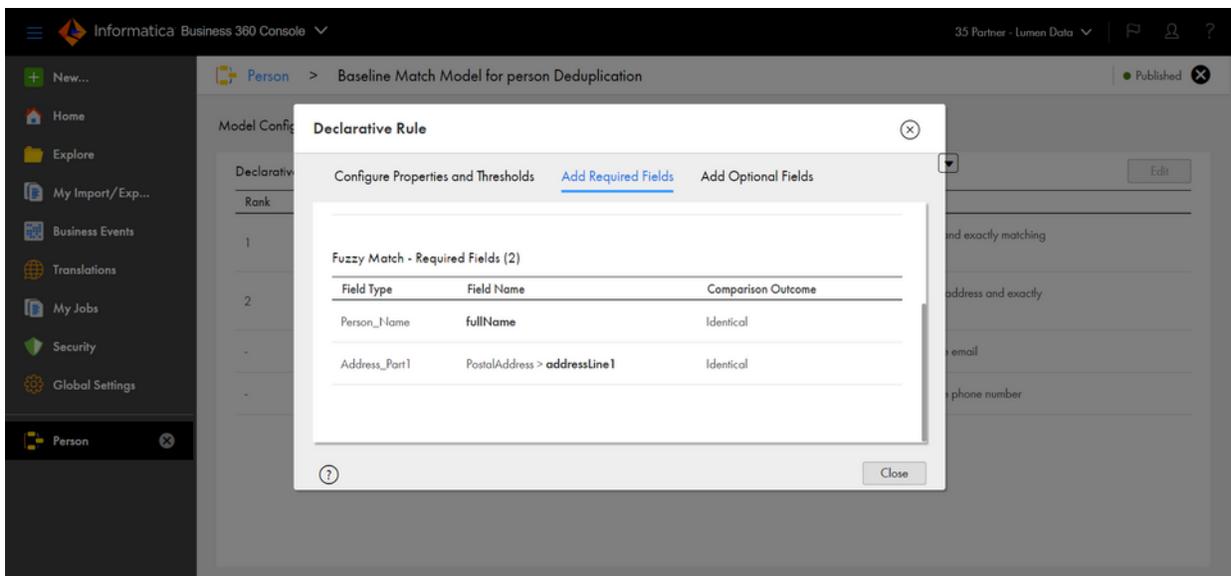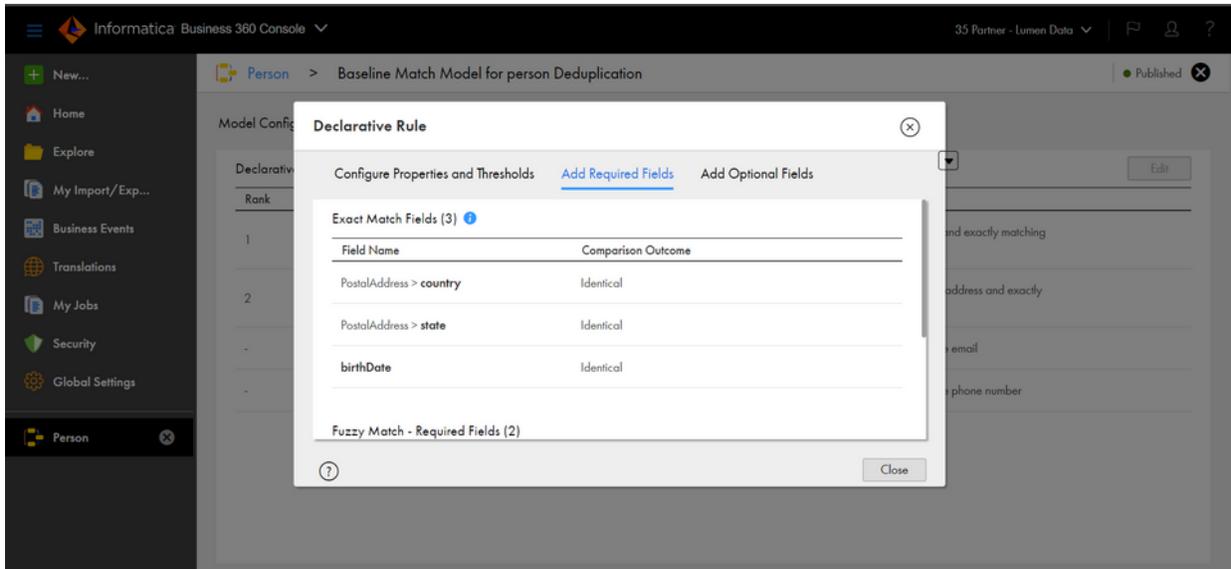
You can add optional field for fuzzy match as per requirements. After configuring all the match and merge strategies, click on save and publish your business entity to use the match model.

3. CREATE A JOB TO PERFORM THE MATCH AND MERGE TO OPERATION:

Navigate to Business 360 Console, click on new, select match and merge from the jobs section as shown below,
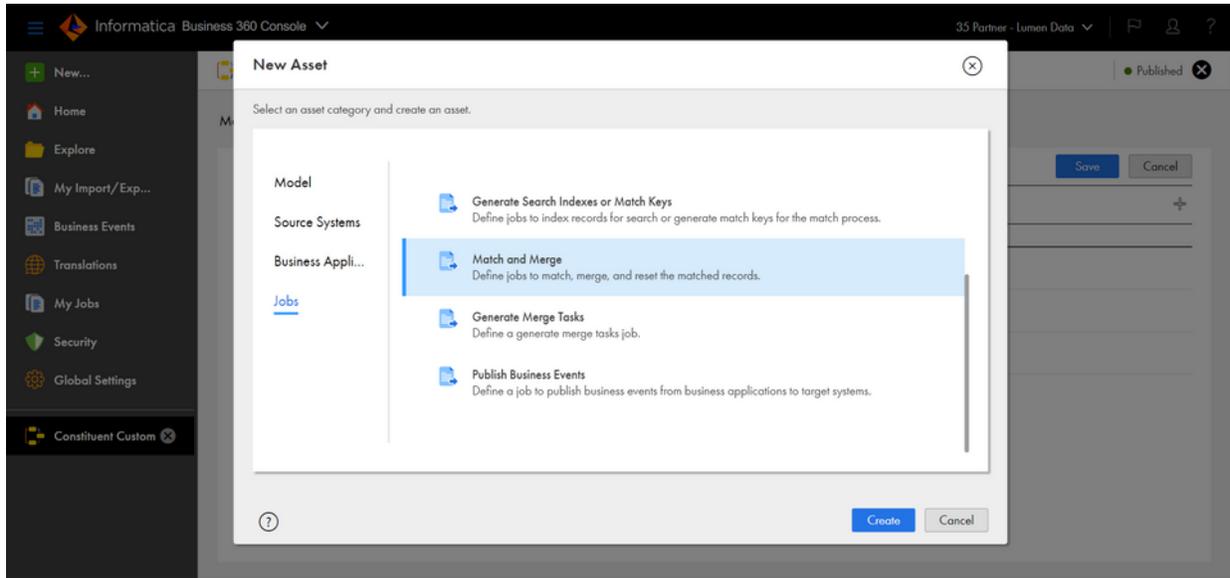
Provide the display name, location and description of the match and merge job and click on OK as shown below,
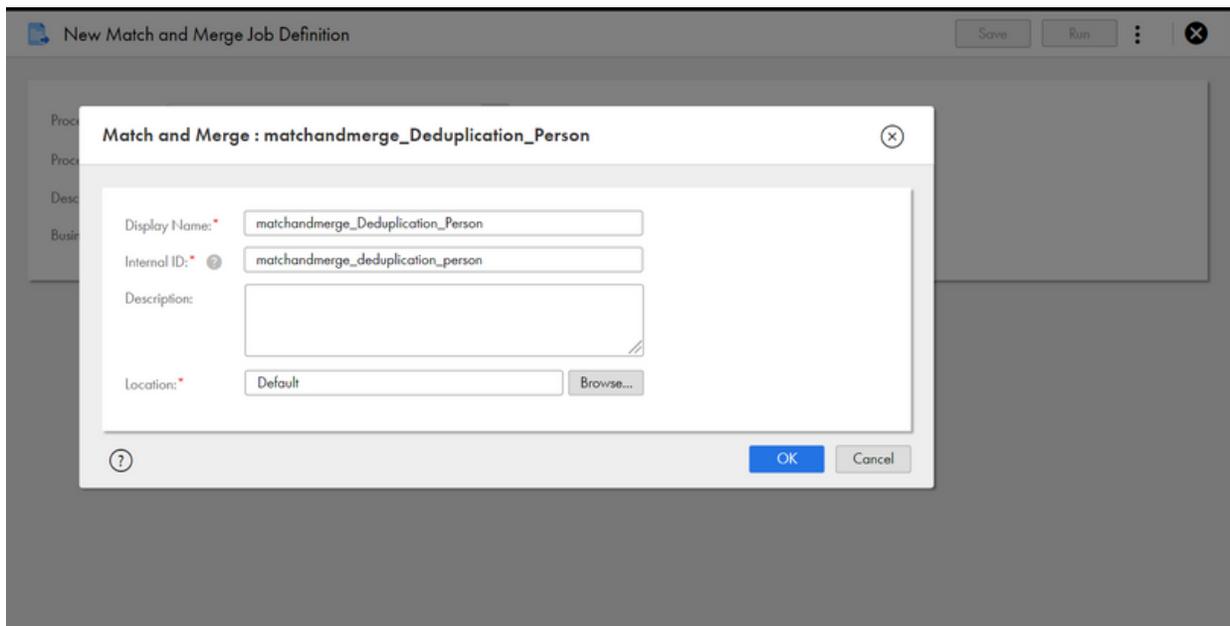
Enter all the required input for the fields as shown in the below picture which specify the business entity upon which we are applying the match and merge process.

Make sure the business entity, match model name, match model version are selected correctly as per the previous configuration. Save the job and run it. Navigate to my jobs in business 360 console to check the status of the running job. Once, the job has been completed successfully download the output file from the output tab from the match process.
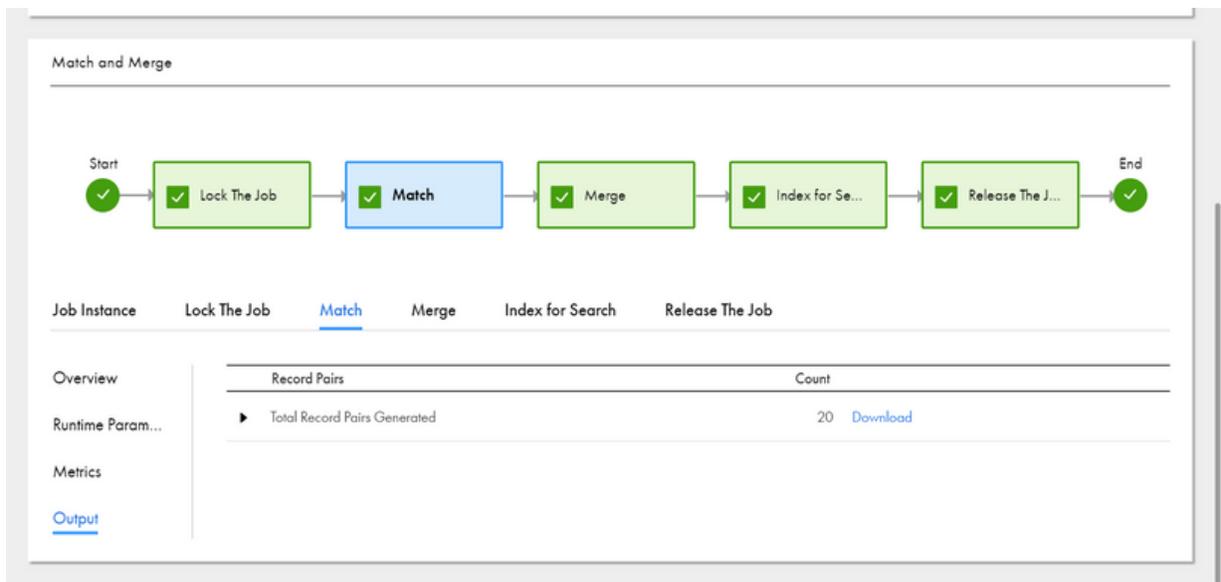
The output file would be a CSV file which will give us all the details about the operation and the records undergoing that operation as below and the match records the match records thus obtained,

Record ID	Record So	Matched F	Matched F	Match Sco	Match Ou	Matched E	Match Mo	Match Mo	Match Mo	Match Rul	Population	Descriptio	Match Crit	Record fir:	Matched F	Record ful	Matched F	R
6450c033	c360.defa	6450c12d	c360.defa	100	AUTO_ME	Declarativ	65984184	Baseline N		1	1	usa			ABHINAV	ABHINAV	ABHINAV	ABHINAV	IK
6450c12d	c360.defa	6450c033	c360.defa	100	AUTO_ME	Declarativ	65984184	Baseline N		1	1	usa			ABHINAV	ABHINAV	ABHINAV	ABHINAV	IK
6450c337	c360.defa	6450c718	c360.defa	100	AUTO_ME	Declarativ	65984184	Baseline N		1	2	usa		Person_N:	VIGNESH	VIGNESH	VIGNESH	VIGNESH	IK
6450c337	c360.defa	6450c7be	c360.defa	100	AUTO_ME	Declarativ	65984184	Baseline N		1	2	usa		Person_N:	VIGNESH	VIGNESH	VIGNESH	VIGNESH	IK
6450c718	c360.defa	6450c337	c360.defa	100	AUTO_ME	Declarativ	65984184	Baseline N		1	2	usa		Person_N:	VIGNESH	VIGNESH	VIGNESH	VIGNESH	IK
6450c718	c360.defa	6450c7be	c360.defa	100	AUTO_ME	Declarativ	65984184	Baseline N		1	2	usa		Person_N:	VIGNESH	VIGNESH	VIGNESH	VIGNESH	IK
6450c7be	c360.defa	6450c337	c360.defa	100	AUTO_ME	Declarativ	65984184	Baseline N		1	2	usa		Person_N:	VIGNESH	VIGNESH	VIGNESH	VIGNESH	IK
6450c7be	c360.defa	6450c718	c360.defa	100	AUTO_ME	Declarativ	65984184	Baseline N		1	2	usa		Person_N:	VIGNESH	VIGNESH	VIGNESH	VIGNESH	IK
6450c97e	c360.defa	6450ca9e	c360.defa	100	AUTO_ME	Declarativ	65984184	Baseline N		1	2	usa		Person_N:	NIHARIKA	NIHARIKA	NIHARIKA	NIHARIKA	S
6450ca9e	c360.defa	6450c97e	c360.defa	100	AUTO_ME	Declarativ	65984184	Baseline N		1	2	usa		Person_N:	NIHARIKA	NIHARIKA	NIHARIKA	NIHARIKA	S

Analysis: After running the match and merge job we navigate to customer 360 to check for the merged record in the MDM. Select the previously used business entity and search for the required record. Once the record is found, check for the source records.

In this case, two different records with different IDs as shown above have been merged into one single surviving record with the details of both the merged records.

Upon searching for the above record in Customer 360, we only found the surviving record in the MDM with the two merged records under the source record section as previously discussed. These records were merged as the ID and Full Name were same for both records and this was our declarative rule defined in the match strategy with the highest priority. The details of the match and merge operation can be found in the output file.

	Vaibhav Singh Negi Vaibhav Si...	Vaibhav Singh Negi Vaibhav Sing...	Vaibhav Singh Negi Vaibhav Sing...
		Source System: Informatica Customer 360	Source System: Informatica Customer 360
System Fields			
Business ID	: MDM000000B...		
ID	: 64660731d3be...	ex62873c19f03d912957697764	ex62873c19f03d912957697761
Updated On	: 05/18/2023	05/18/2023	05/18/2023
Updated By	: Adarsh.ld	Adarsh.ld	Adarsh.ld
Record State	: ACTIVE	ACTIVE	ACTIVE
Population	:		
▶ Match Pair Details			
▼ General Information			
First Name	: Vaibhav	Vaibhav	Vaibhav
Middle Name	: Singh	Singh	Singh
Last Name	: Negi	Negi	Negi
Full Name	: Vaibhav Singh ...	Vaibhav Singh Negi	Vaibhav Singh Negi
Prefix Name	: Mr.	Mr.	Mr.

ABOUT LUMENDATA:

LumenData is a leading provider of Enterprise Data Management, Cloud & Analytics solutions provider. We enable enterprises to modernize their legacy data infrastructure and derive greater insights into their business. Founded in 2008, with locations in multiple countries, LumenData is privileged to serve over 100 of the world's leading companies, including KwikTrip, Versant Health, US Food & Drug Administration, US Department of Labor, Cummins Engine, BCG, and others. LumenData is SOC2 certified and has instituted extensive controls to protect client data including adherence to GDPR and CCPA regulations.

To discuss how we can facilitate data-driven transformation for your organization.

MEET OUR AUTHORS

Adarsh Sinha
Informatica/Data ETL Engineer

Kousalya S
Associate Consultant

Shadwal Srivastava
Technical Lead

LUMENDATA
Masters in Data Management