Informatica

25 June, 2024

Match and Merge Use-cases

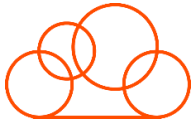
- Sourya Dass, Principal Customer Success Architect, CSA

Where data & AI come to LIFE

Housekeeping Tips

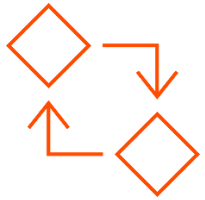
- Today's Webinar is scheduled for 1 hour
- The session will include a webcast and then your questions will be answered live at the end of the presentation
- All dial-in participants will be muted to enable the speakers to present without interruption
- Questions can be submitted to "All Panelists" via the Q&A option and we will respond at the end of the presentation
- The webinar is being recorded and will be available on our Success Portal - where you can download the slide deck for the presentation. The link to the recording will be emailed as well.
- Please take time to complete the post-webinar survey and provide your feedback and suggestions for upcoming topics.

Feature Rich Success Portal



Bootstrap trial and POC Customers

Enriched Customer Onboarding experience



Product Learning Paths and Weekly Expert Sessions

Informatica Concierge

Tailored training and content recommendations

More Information

Success Portal

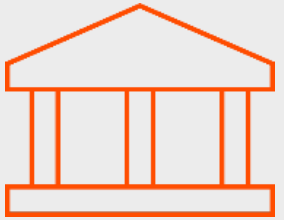
https://success.informatica.com

Communities & Support

https://network.informatica.com

Documentation

https://docs.informatica.com



University

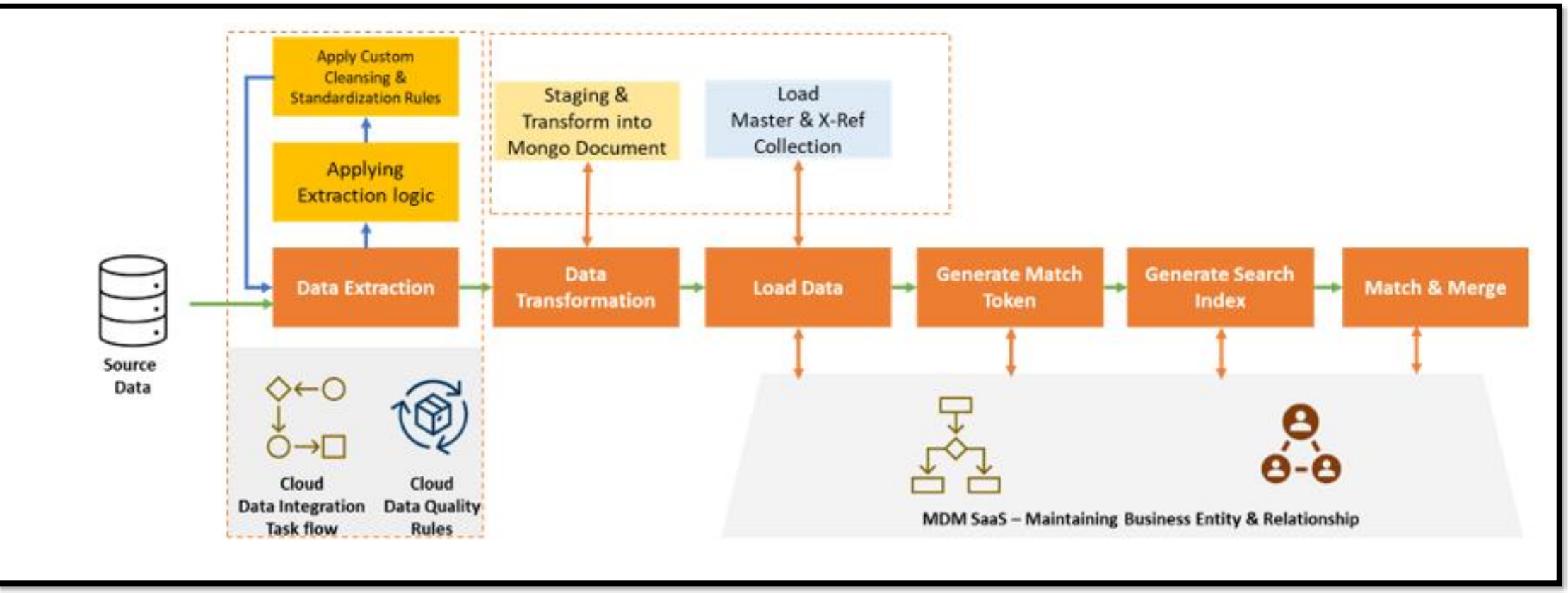
https://www.informatica.com/in/services-and-training/informatica-university.html

Safe Harbor

The information being provided today is for informational purposes only. The development, release, and timing of any Informatica product or functionality described today remain at the sole discretion of Informatica and should not be relied upon in making a purchasing decision.

Statements made today are based on currently available information, which is subject to change. Such statements should not be relied upon as a representation, warranty or commitment to deliver specific products or functionality in the future.

MDM SaaS Batch Flow

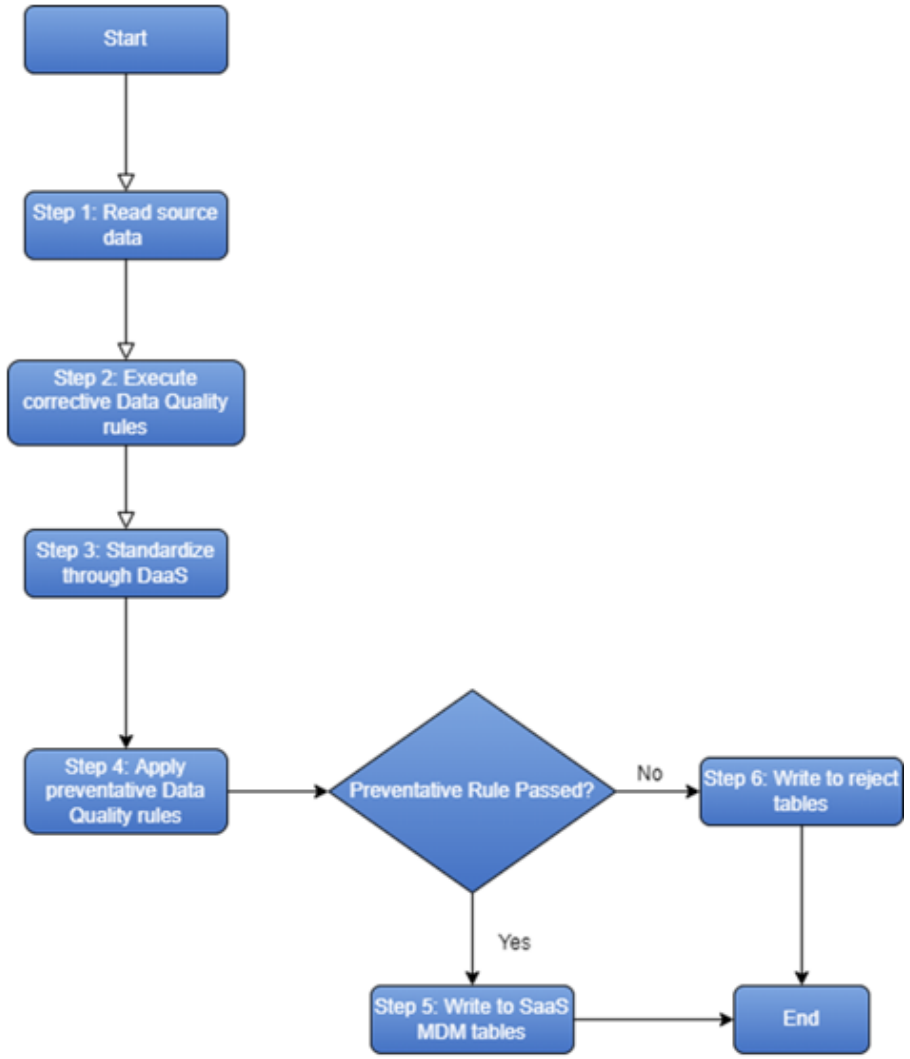


MDM SaaS Batch Flow

- **Data Transformation Phase** – This is the second step of data ingress process. In the step the cleansed data will be transformed and converted into document.
- **Data Load Phase** – In this phase business entity will be created in the MDM database. Generally Master Collection, Cross-Ref Collection will be loaded.
- **Generate Match Token Phase** – In this phase match token will be generated and the master collection will be updated with the match token.
- **Generate Search Index Phase –** In this phase search index will be generated for the business entity. The search index will be used to find any document.
- **Match & Merge Phase –** The data deduplication is 2 step process. The first step is match step, in this step system will use the configured match model and identify the duplicate document based on the configured match rule. The second step is merge step. In this step duplicate document will be consolidated and create the golden document. The consolidation process is a complex process, in the process the attribute level value will be consolidated based on the survivorship rule. The process will survive the best value based on the BVT (Best Value of Truth) calculation.

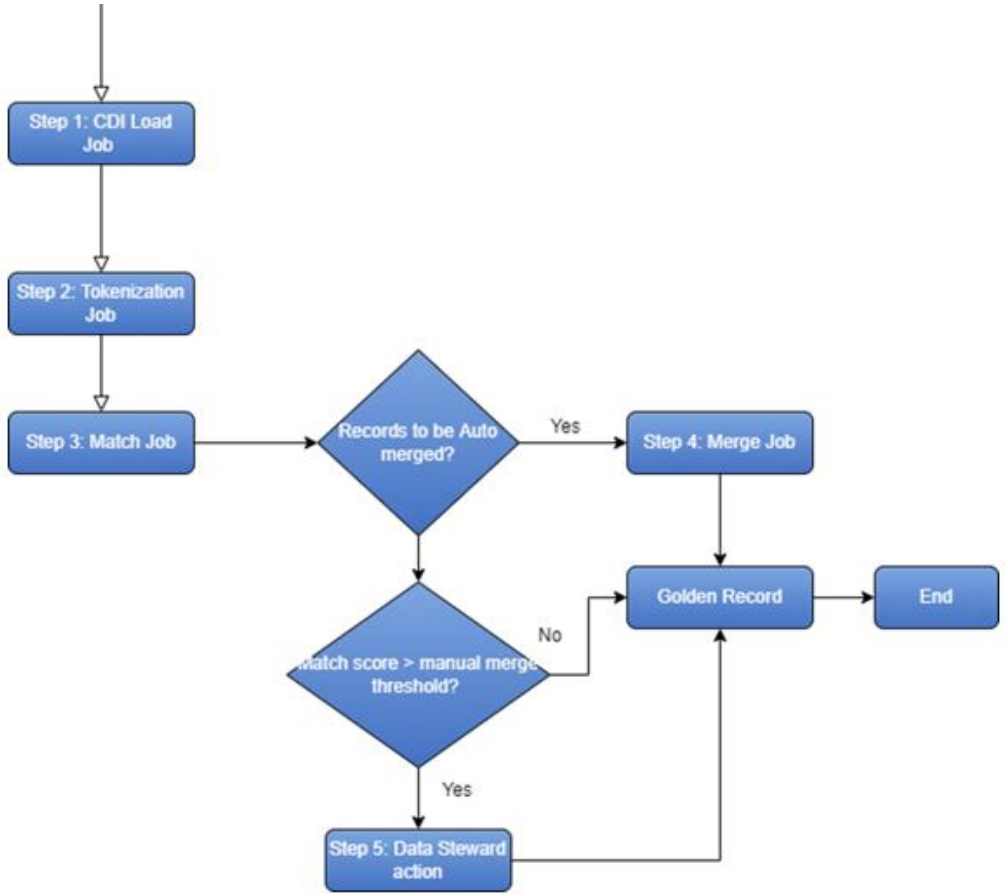
Data Quality Validation & Cleansing

- Step 1: Read data from the source.
- Step 2: Apply corrective data quality rules.
- Step 3: Standardize the name, phone, and address attributes through custom DQ rules.
- Step 4: Apply preventative data quality rules.
- Step 5: If the preventative data quality rules pass, write to SaaS MDM tables.
- Step 6: If the preventative data quality rules don't pass, write to reject tables.



Data Quality Validation & Cleansing

- Step 1: The CDI Load job would load the data into SaaS MDM tables.
- Step 2: Tokenization job would generate match tokens for the records loaded in step 1. The tokens are encrypted data which helps to identify potential duplicates.
- Step 3: Match job generates match score among potential duplicate records. If the match score is less than manual merge threshold, golden record is created.
- Step 4: The merge job automatically merges the records which are identified for auto merge. During the merge, survivorship rules would be applied to generate the golden record.
- Step 5: If the match score is greater than manual merge threshold, records are placed in Data Steward's queue. Data Steward would decide whether the records are duplicate or unique. Based on Data Steward's decision, golden record is created through merge or discarding duplicate records.



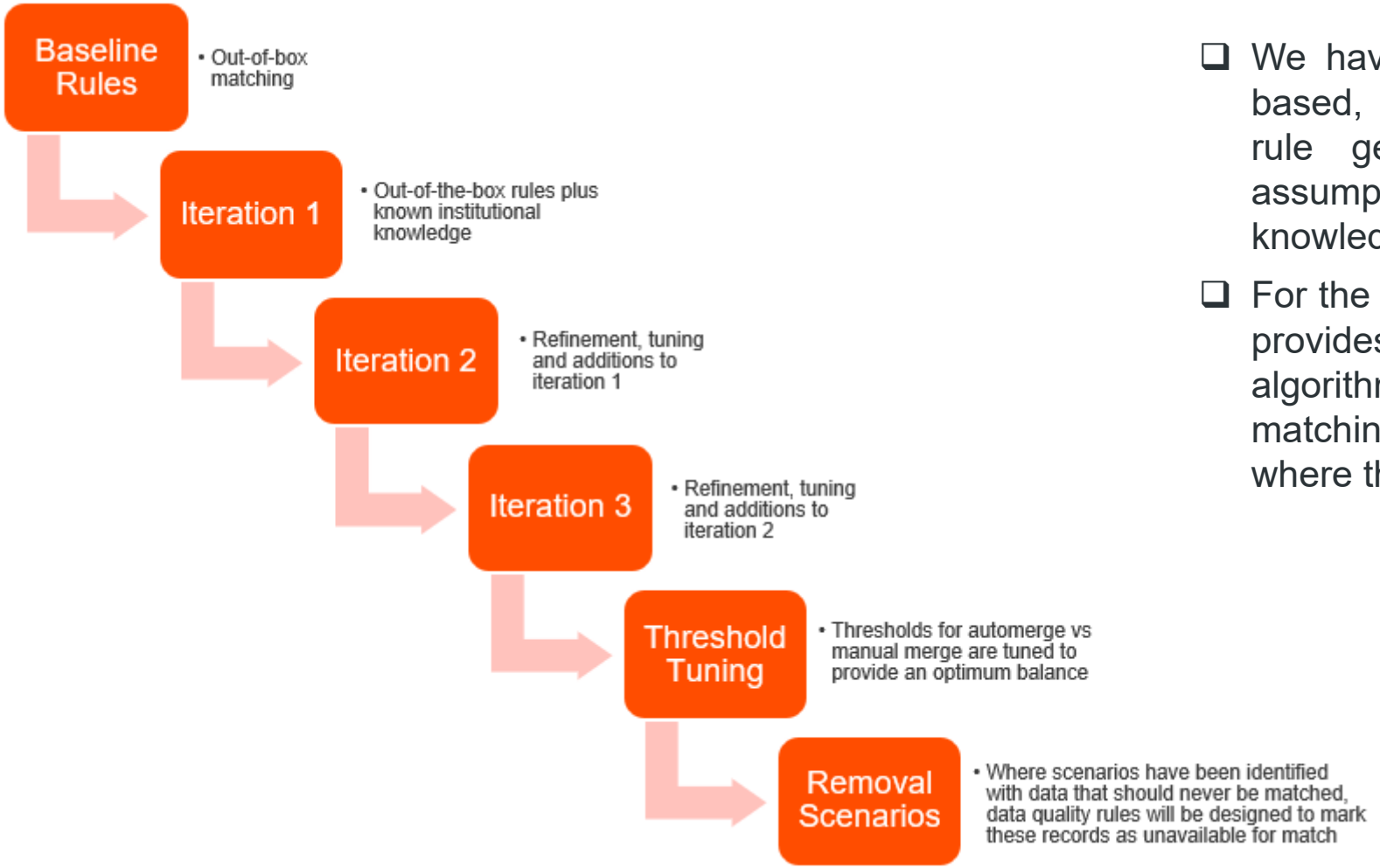
MDM SaaS Matching

- Informatica performs data de-duplication based on agreed match rules. Each match rule is defined as a set of one or more match columns that it needs to examine for points of similarity. The match column determines what constitutes a match during the match process. Match columns determine whether two records are similar enough to consolidate.
- Further, the match column can be configured as Fuzzy or Exact.
  - **Fuzzy** - Probabilistic match that considers spelling variations, possible misspellings, and other differences that can make matching records non-identical.
  - **Exact** - Matches only records that are identical.
- Match rules are configured by setting the conditions for identifying matching records within and across source systems.
- There can be following two types of match rules:
  - **Auto Match Rules:** Match rules that are configured to identify definite matches for auto-merging
  - **Suspect Match Rules:** Match rules that are configured to identify close matches for manual merging

MDM SaaS Consolidation Process

- The goal of merging data in Informatica MDM SaaS is to eliminate all duplicate records by combining them together into a single, consolidated record—while maintaining full traceability, it can determine which systems, and which cells from those systems, contributed to the Golden record. Duplicate records are identified during the match process.
- Informatica MDM SaaS compares records based on precedence (Trust) or recency. The data in the record with the highest trust level/recency are survived to form the Golden Record. The Trust levels indicate, as a percent, the level of confidence in the accuracy of data.
- If there is no trust defined on any of the sources, always the record which is updated recently survives. However, in event of two records having exactly same update timestamp at source then the record which is inserted last in the system will be considered as the most trustable update.
- The above consolidation process is applicable to all the records which are consolidated into Golden Record.

MDM SaaS Matching



- We have generally found that an evidence-based, data-driven approach to deterministic rule generation, rather than relying on assumptions or even established institutional knowledge, will yield the best results.
- For the MDM SaaS in particular, CLAIRE also provides advanced machine learning algorithms which can help to recognize matching patterns especially in the case where the source data is not well understood.

MDM SaaS Matching

Rule No #	Match Model	Population	Field Group	Match Column	Match Type	Merge	Match Criterion	Overall Rule Match Type Strength (Loose/Typical/ Conservative)	Merge Strategy
#1	ABC Corp Customer Match Model	USA	Organization.Root	Location Type	Exact	Auto Merge	Organization	Typical	Thresold Base Skip Merge (0-50) Manual Merge (51-89) Automerge (90-100)
			Organization.Root	Organization Name (Legal Entity Name)	Fuzzy				
			Organization.Root	Code	Exact				
			Organization.Root	Location Id	Exact				
#2	ABC Corp Customer Match Model	USA	Organization.Root	Location Type	Exact	Manual Merge	Organization	Typical	
			Organization.Root	Organization Name (Legal Entity Name)	Fuzzy				
			Organization.Root	Code	Exact				
			Organization.Address	AddressType	Exact				
			Organization.Address	Address_Part1(Address Line 1, Address Line 2, Address Line 3)	Fuzzy				
			Organization.Address	State/ Province	Exact				
			Organization.Address	Postal/ Zip	Exact				
			Organization.Address	Country code	Exact				
#3	ABC Corp Customer Match Model	USA	Organization.Root	Location Id	Exact	Auto Merge	Organization	Typical	Thresold Base Skip Merge (0-50) Manual Merge (51-89) Automerge (90-100)
			Organization.Root	Organization Name (Legal Entity Name)	Fuzzy				
			Organization.Address	AddressType	Exact				
			Organization.Address	Address_Part1(Address Line 1, Address Line 2, Address Line 3)	Fuzzy				
			Organization.Address	State/ Province	Exact				
			Organization.Address	Postal/ Zip	Exact				
			Organization.Address	Country code	Exact				
#4	ABC Corp Customer Match Model	USA	Organization.Root	Organization Name (Legal Entity Name)	Exact	Auto Merge	Organization	Typical	Thresold Base Skip Merge (0-50) Manual Merge (51-89) Automerge (90-100)
			Organization.Root	Code	Fuzzy				
			Organization.Root	DUNS Number	Exact				
			Organization.Root	DUNS Suffix Number	Exact				
#5	ABC Corp Customer Match Model	USA	Organization.Root	Location Id	Exact	Auto Merge	Organization	Typical	Thresold Base Skip Merge (0-50) Manual Merge (51-89) Automerge (90-100)

MDM SaaS De-Duplication Rules

Match Model	Declarative Rules					
Match Model for Building Deduplication	**Rule No #**	**Match Column**	**Match Type (Match Criterion)**	**Overall Rule Match Type Strength** (Loose, Conservative, Typical)	**Merge Strategy**	**Description**
	#1	Building Name	**Exact**	**Typical**	**Manual**	Exact Building Name
	#2	Building Code	**Exact**	**Typical**	**Auto**	Exact Building Code

Match Model	Declarative Rules					
Match Model for Address Deduplication	**Rule No #**	**Match Column**	**Match Type (Match Criterion)**	**Overall Rule Match Type Strength** (Loose, Conservative, Typical)	**Merge Strategy**	**Description**
	#1	Address_Part1 (Address Line 1, City, Postal Code)	**Fuzzy (Address Part1)**	**Typical**	**Manual Merge**	Strongly similar address and exactly matching state
		State	**Exact**			

MDM SaaS Match Iteration Results

Match Pairs by Rule

Rule ID	rule Desc	Iteration 1		Iteration 2		Iteration 3		POST Iteration 3 (49M)	
		30,348,170		39,541,310		12,317,899		33,584,747	
1674071972721	Rule 1 (name, address,exact phone, exact email)	n/a		15,100,269	38%	5,926,463	48%	13,067,627	39%
1674058587761	Rule 2 (fuzzy name,address, phone, email)	n/a		67,024	0%	2,820,435	23%	6,754,109	20%
1670446557065	Rule 3 (exact phone)	221,976	1%	7,574,199	19%	2,760,250	22%	8,263,814	25%
1670445886996	Rule 4 (exact address, fuzzy zip)	25,022,415	82%	16,264,111	41%	437,924	4%	4,645,062	14%
1670446514208	Rule 5 (exact email)	3,716,112	12%	371,336	1%	195,987	2%	566,496	2%
1670446605045	Rule 6 (exact payment)	1,387,667	5%	164,371	0%	176,840	1%	287,639	1%

MDM SaaS Workflow

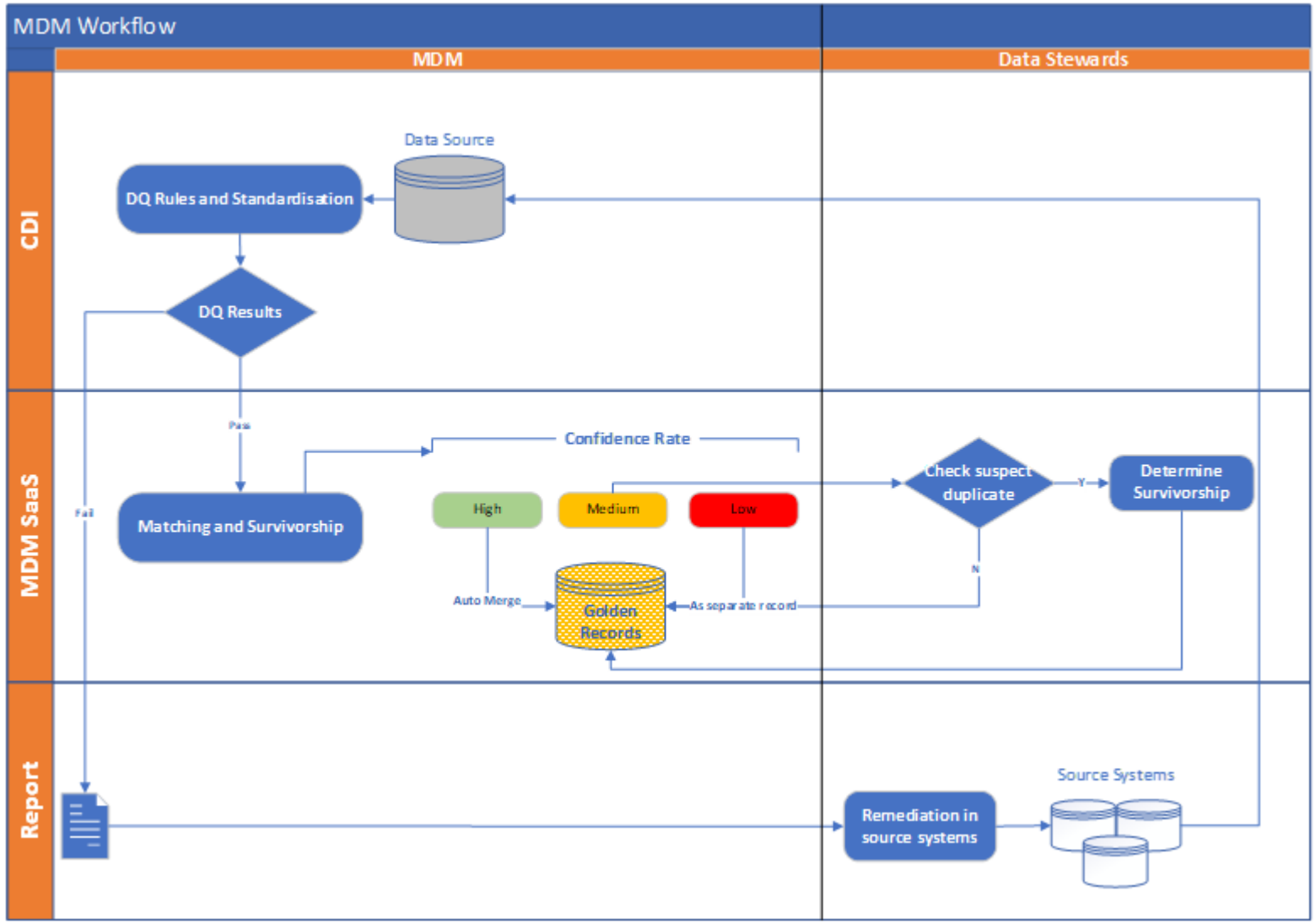


Diagram explains the MDM workflow approval process. Whenever the match score for two records is more than manual merge threshold and less than auto merge threshold, a new task would be created for the Data Steward to review the match records and decide whether the records are duplicate records or are unique records.

MDM SaaS Roles

Role/Access	Create	Read	Update	Delete	Merge/Unmerge
ABC – Administrator (Full Access)	Yes	Yes	Yes	Yes	Yes
ABC – Administrator (Read Only)	No	Yes	No	No	No
ABC – Governance (Full Access)	Yes	Yes	Yes	Yes	Yes
ABC – Governance (Read Only)	No	Yes	No	No	No
ABC – Owner (Full Access)	Yes	Yes	Yes	Yes	Yes
ABC – Owner (Read Only)	No	Yes	No	No	No
ABC – Steward (Full Access)	Yes	Yes	Yes	Yes	Yes
ABC – Steward (Read Only)	No	Yes	No	No	No
ABC – Analyst (Full Access)	Yes	Yes	Yes	Yes	Yes
ABC – Analyst (Read Only)	No	Yes	No	No	No

Logical Steps to MDM SaaS Implementation

- Data Governance Councils
- Finalization of source systems
- Finalization of Lookup Code values
- Finalization of Business attributes
- Define Business & Technical concepts
- Finalize MDM Data Model
- Profile Data Quality

Logical Steps to MDM SaaS Implementation

- Finalize MDM use cases based on requirements and business feedback & inputs
- Design conceptual Match Rules in MDM
- Design Ingress & Egress mechanism
- Finalize Data Quality rules based on use cases
- Design MDM UI
- Finalize User Roles & Privileges
- Start test data load in Development environment

Drilling down to the technical details

CDQ use cases	CDI use cases	MDM use cases	Downstream data syndication	Reporting & Analytics requirements
• Business rules • Technical validations • Exception report • Data corrections	• Source to Target mappings • Data transformations • DnB, Address Doctor integration	• Data Load Strategy • Match rules • Survivorship • Hierarchy • UI	• API (real time) • Message Queue (near real time) • Batch driven • Publish format	• Target system • Schedule • Giving access to teams