Sep 27th, 2022

C360 SaaS - Automate Match Strategy and Survivorship

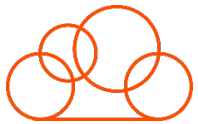
- **Kamal Abrol,** Principal Solutions Architect, CSM

Informatica

Housekeeping Tips

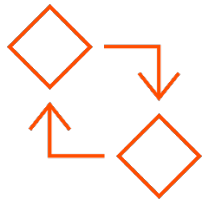
- Today's Webinar is scheduled for 1 hour
- The session will include a webcast and then your questions will be answered live at the end of the presentation
- All dial-in participants will be muted to enable the speakers to present without interruption
- Questions can be submitted to "All Panelists" via 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 INFASupport YouTube channel and 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

Informatica

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

Informatica

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.

Informatica

Agenda

- Why Matching?
- Deduplication Strategy – Usage Recommendation
- Match Services-MDM SaaS
- Survivorship & Conflict Resolution
- Downgrade Trust Score -Validation Importance
- Match Merge job Flow
- Generate Merge Task Workflow
- Live Demo-Ingress + Match + Merge

Why Matching?

John Smith and John Smyth
• Same records – with spelling variations, misspellings, transpositions, omissions, and phonetic variations
• Duplications

• *Configure the match and merge process to identify and resolve duplicate records.*

• *The match and merge process configuration together constitutes the match model.*

• *Based on the outcome of the match process and the merge configuration, the records are either:*

• ***automatically merged***

• ***queued for manual merge by a data steward***

• ***skipped as not a candidate for merge***

SaaS Deduplication Strategy – Usage Recommendation

Match Strategy

Match Strategy	Description	Examples
Exact	A deterministic match that includes records containing identical values. Configure the exact match strategy if the quality of the data is good.	Use the exact match strategy if at least some fields contain unique data, such as IDs, postal codes, and industry codes.
Fuzzy	A probabilistic match that includes variations in data patterns, such as spelling variations, misspellings, transpositions, omissions, and phonetic variations.	Use the fuzzy match strategy to match records based on fields that contain addresses, geographic coordinates, and names of people or organizations.

Merge Strategy

- **AUTO**

Usage Recommendation

• Use the automated merge strategy for declarative rules that are based on unique identifier fields, such as social security number or Passport number.
• Two people cannot have the same Social Security or Passport number, so you do not have to manually review the records.

- **Threshold based**

Available only when you select the fuzzy match strategy. • Use the threshold-based merge strategy for any field. • Based on the match score of the record pairs, the match and merge process determines whether the result is automated, manual, or skipped merge.

- **Manual**

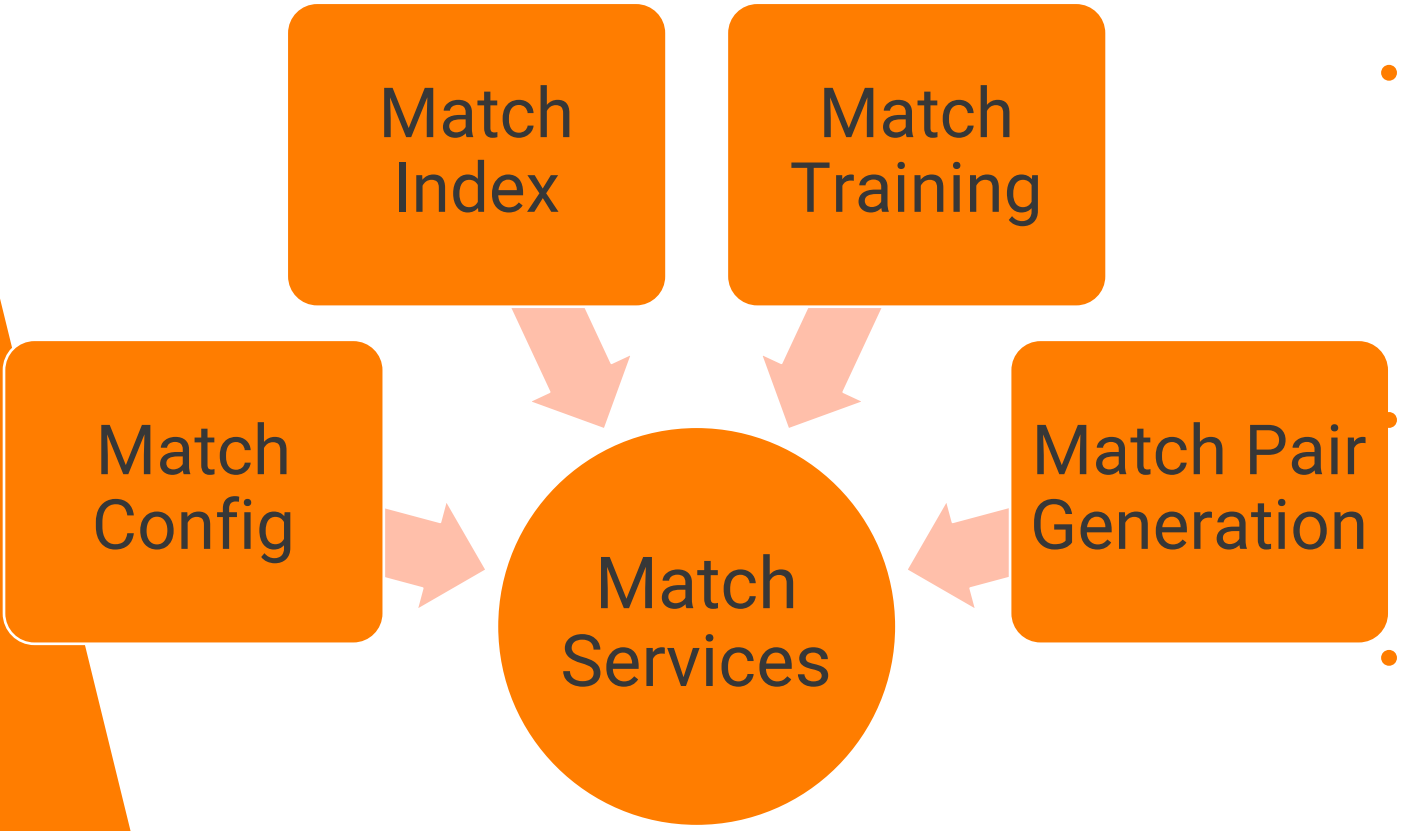
Usage Recommendation

• Use the manual merge strategy for declarative rules that are based on fields, such as the address field.
• Multiple family members can share the same address and the records cannot be definite matches
. • You need to manually review the records to confirm whether these are duplicate records

- **SKIP**

The records in the pairs are not matched against each other if their match scores fall within the range set for the merge threshold for skipping merge

Match Services-MDM SaaS



- **Match Config:**
  - Provides an interface to configure match models i.e candidate selection criteria, DL and AI rules or machine learning model
- **Match Index:**
  - Generates name3 keys/tokens for user data
- **Match Pair Generation:**
  - Deduplicates records by creating match pairs
  - (which in turn are merged for generating golden records)

Informatica

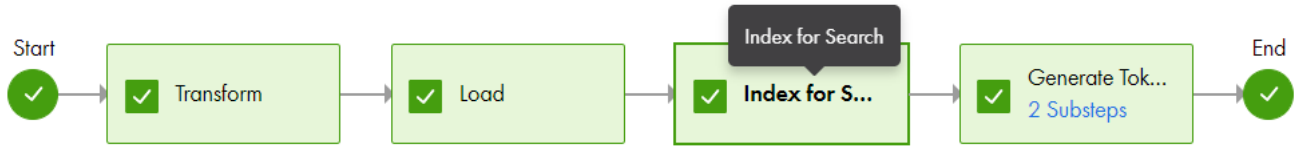
Match Index Service

MATCH_DIRTY-
Date Ingested

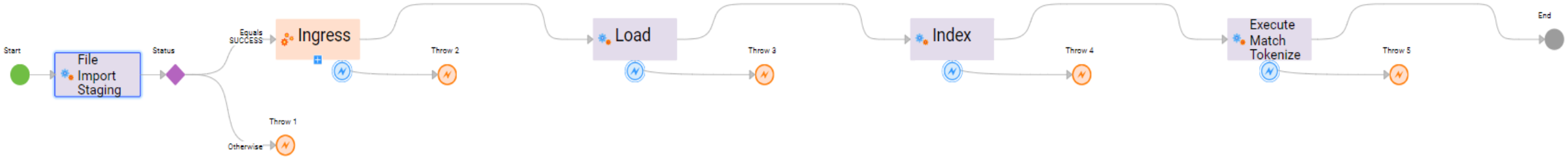
Index Configuration

MATCH_INDEXED-
Post generate Tokenization step

Ingress Job flow- Data getting indexed



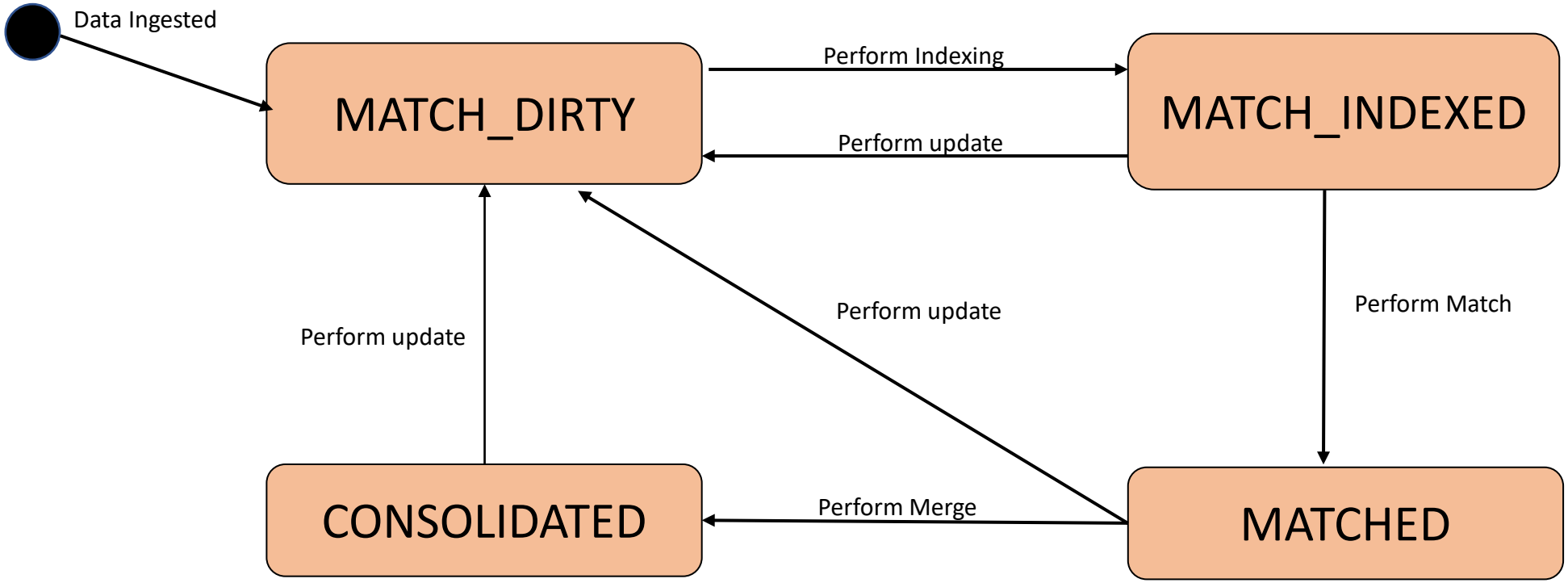
File Import Flow Diagram



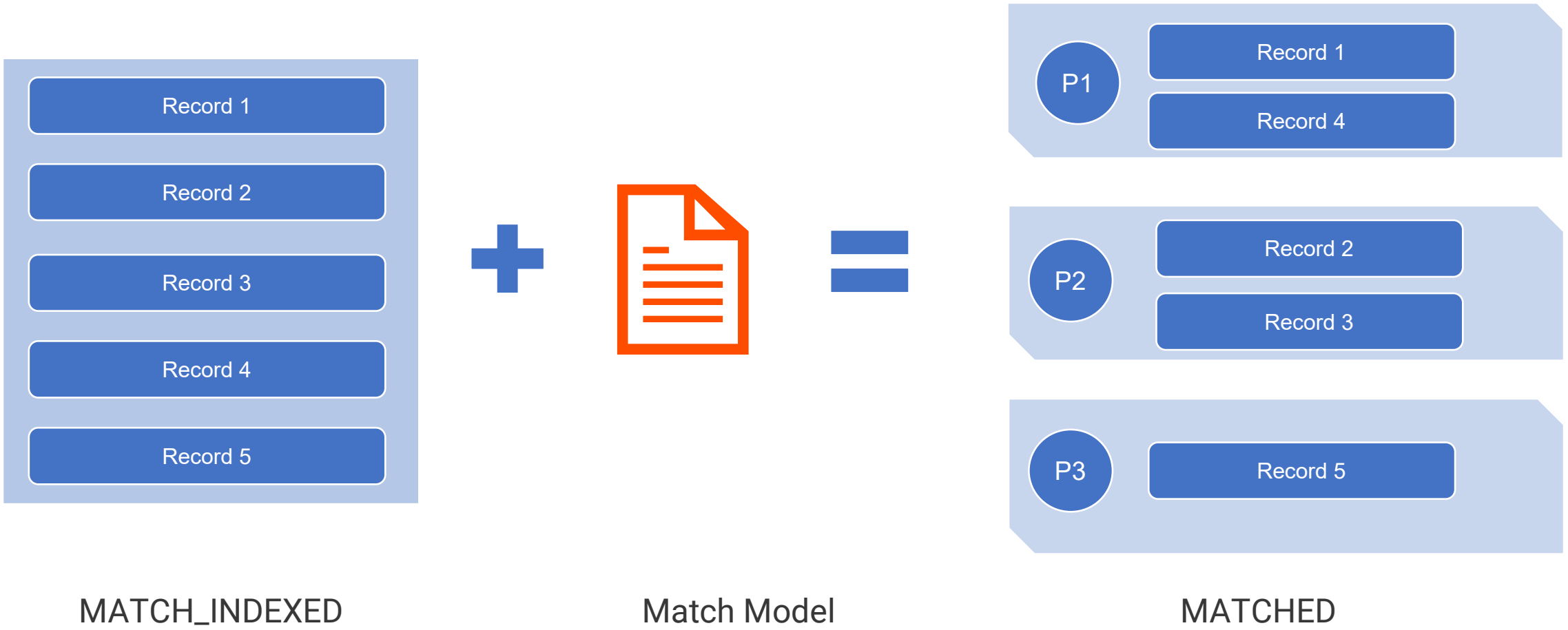
Informatica


Ingress
DEMO

XREF State Management



Match Pair Service



MATCH_INDEXED

Match Model

MATCHED

Survivorship- Master creation from multiple matching pairs

- Which records survive?
- Which source systems are most trusted?
- Though highly trusted, what are the basic validations that need to be done on the attributes coming from those sources?
- If the records are not valid, how do you downgrade the trust scores?

- **Field Level Survivorship**
  - Compete by each entry
- **Block Level Survivorship**
  - Compete the whole underline group
- **Document/root Level survivorship**
  - The whole xref overwrites the master

Field	Xref X1	Xref X2	Master M1 (if X1 name wins) (if X2 age wins)
Name	Tom	Tommy	Tom
Age	20	21	21

Block	Xref X1	Xref X2	Master M1 (if X1 address wins)
Address.city	Toronto	Vancouver	Toronto
Address.province	ON	BC	ON

❖ Survivorship Configuration-

Define attribute/field specific source system ranking and decay configuration

❖ How Survivorship Configuration Works

✓ If survivorship configuration do not exists

- TS will always be 0.

✓ If survivorship configuration exists

- The first rule executed will be the source ranking rule
- If another rule is configured for the xref element, it will be executed and the calculated score will become the new TS

❖ When trust calculation triggered ?

- ✓ Every time an Xref is Upsert both ways batch job as well as Patch update (From BUI)
- ✓ When a manual override is performed (Survivorship Override)
- ✓ The Data Quality rules can affect the trust score by downgrading with a calculated percentage (DTSP).

Survivorship -Conflict resolution

- The source ranking rule will act as a tie break rule in case of the multiple records from different source systems with the same TS.
- In case of multiple XREFS with the same TS from the same source system the value from the XREF with the newer LUD will survive.
- In the case there are multiple Xrefs with the same TS from the same source system with the same LUD, the record with the highest mongo id (_id) will survive
- There can be a tie in the mongo id. This can happen when picking a winning element in a multivalue field, the highest 'element _id' will be the one that will survive for this case.

Downgrade Trust Score -Validation Importance

Merge Scenario

Case: A

- Informatica Customer 360 is the most trusted source – Rank 1.
- CRM is the next trusted source – Rank 2.
- A record coming from Informatica Customer 360 is NULL, whereas there is a valid record coming from CRM. Which value will be considered for the Match?
- Obviously, CRM – though less trusted.
- To avoid such a situation a DQ Validation rule is a MUST!
- When validation fails, you can downgrade the trust score of the field, and that helps determining the confidence level on the source data.

Match Merge Flow Diagram

Manual Review Workflow

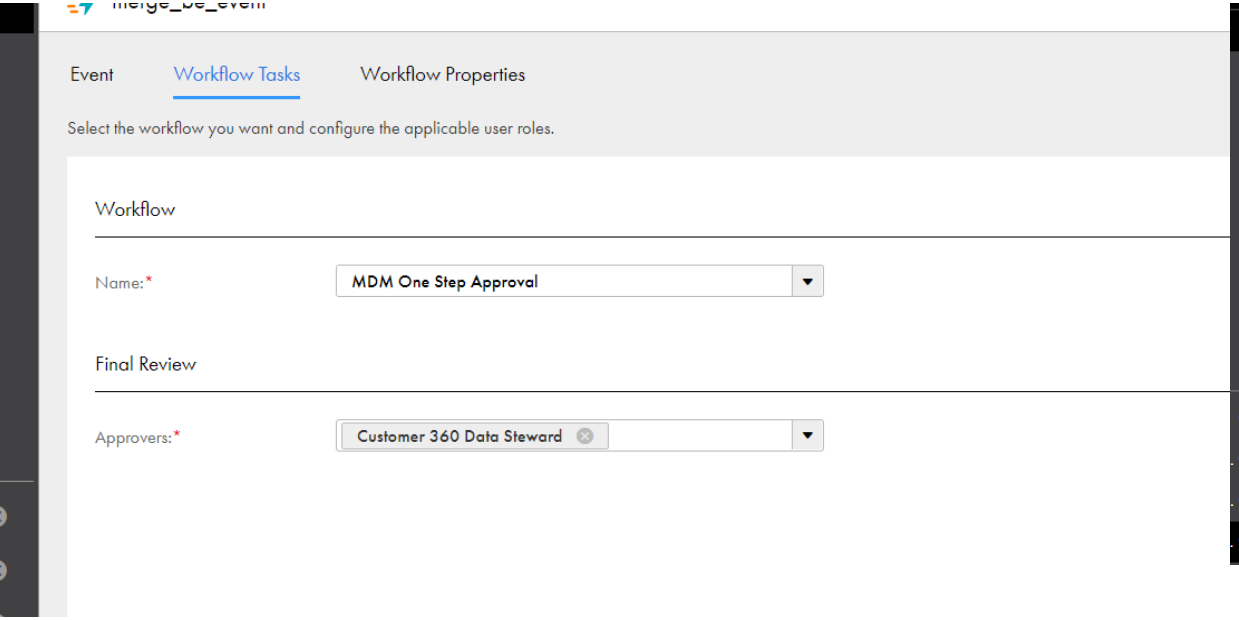
Generate Merge Task Flow-

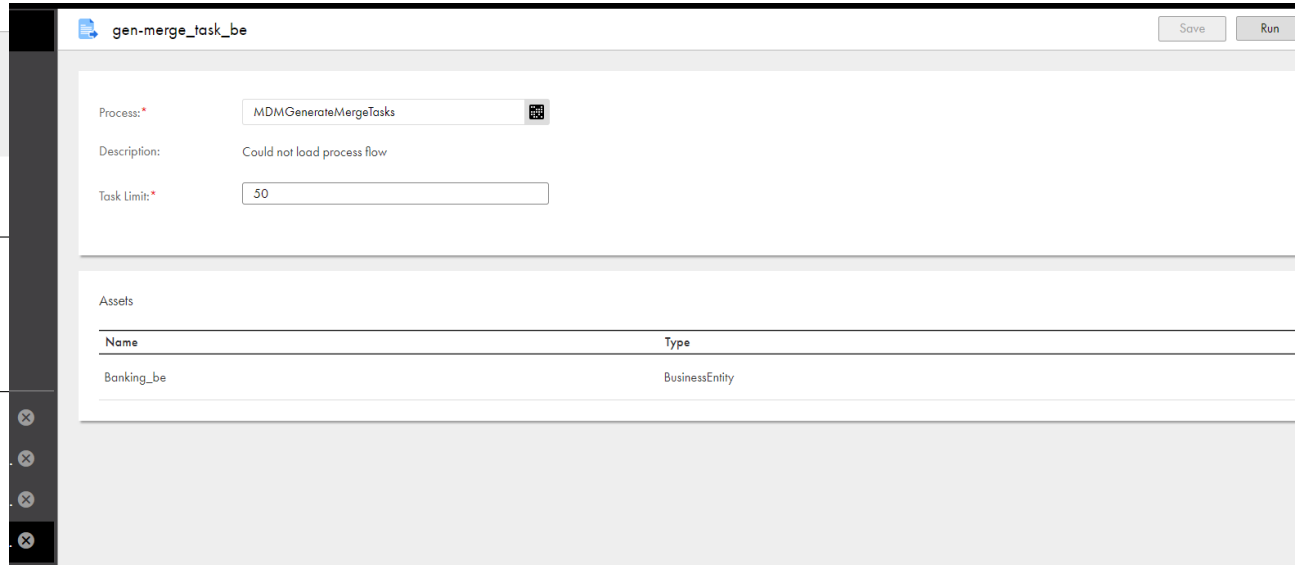
Configure the business event

- For creating review task, we need to configure a system generated business event where we will select the approver here, who will receive the tasks in workflow inbox to review the pairs and can perform appropriate action on them.

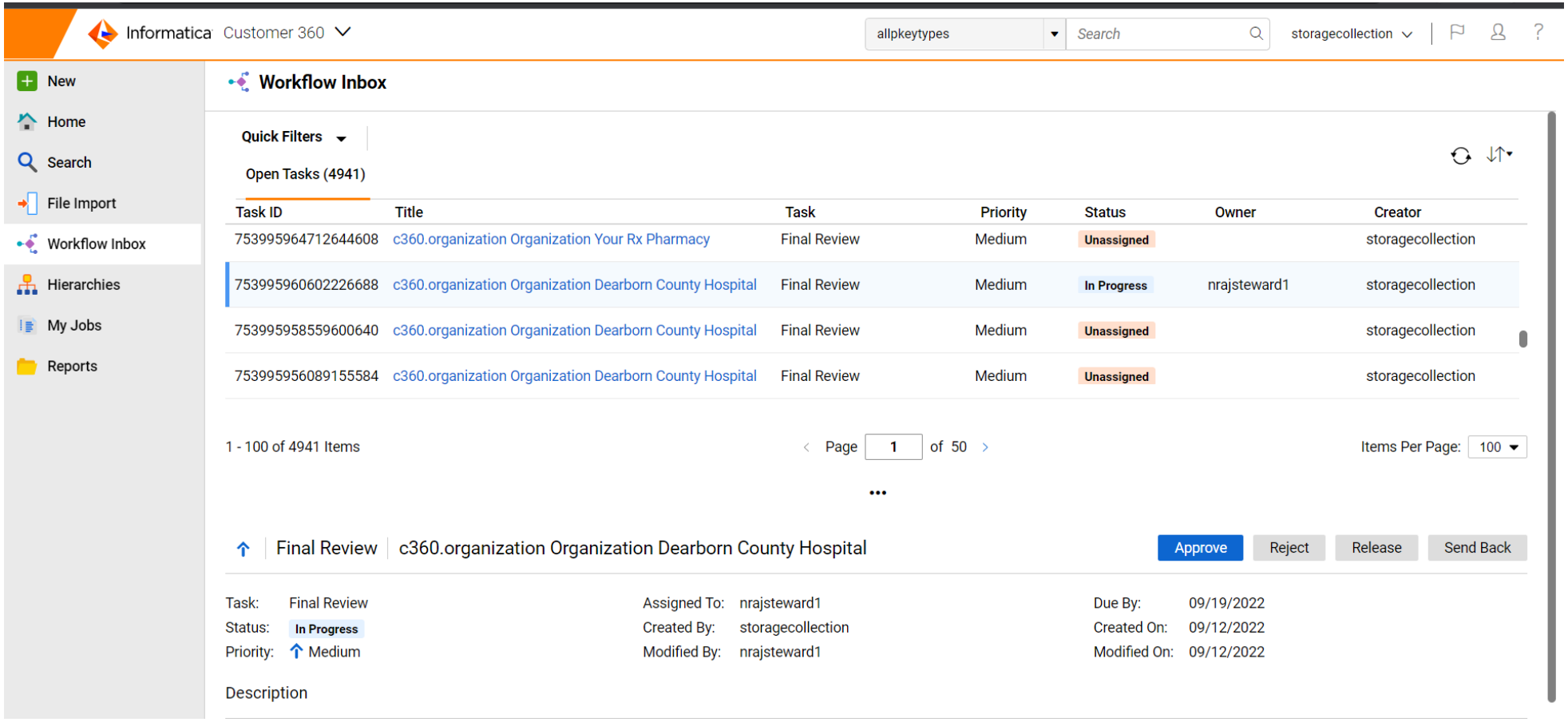
Run the generate merge task job for manual reviews

- We need to configure and run the generate merge task job which creates the task in approver's inbox based on the match outcome of match and merge job.





Login as reviewer and act upon the merge task



High-Level Steps to Configure Match and Merge

1. Configure Source Systems from which you onboard data
2. Configure Survivorship – Set Trust Levels + Decay Rules
3. Create Validation rules to downgrade trust scores
4. Create Declarative Match rules
5. Run Match and Merge jobs
6. For Manual Merge, create Business Events to run workflows


Match Merge
DEMO

Questions?

Kamal Abrol-Speaker
Dilip Yeluguri -Panelist
Customer Success Architect

Thank You

Informatica™

References

- MDM Infa University-https://now.informatica.com/Customer-360-SaaS-for-Business-Users-Instructor-Led.html
- https://infawiki.informatica.com/

Pre steps Best Practices- Matching & Merging

- Analyze the data onboarded and check if matching and merging duplicate data is required.
- Analyze the attributes and quality of the data that you want to consider for the match and merge process.
- Determine the size of the data set, which could impact the performance of the match and merge process.
- Determine if you can opt Automated or a manual merge.
- **Merge Usage Recommendation**
  - ✓ Use the automated merge strategy for declarative rules that are based on unique identifier fields, such as social security number or Passport number.
  - ✓ Use the manual merge strategy for declarative rules that are based on fields, such as the address field.
  - ✓ Multiple family members can share the same address and the records cannot be definite matches.

- Decide the Declarative match rules that best suits the data set.
- Understand the business objectives and requirements for matching and merging duplicate data.
- Determine the match population, which improves match accuracy by accommodating variations and errors that are likely to appear in data for a particular population.