RELTIO

SOLUTION BRIEF

Using machine learning (ML) to speed the journey to faster matches and better data quality

Matching and merging data requires too much time and effort

Creating a single version of your data requires the critical task of aligning your various data sources by deduplicating records that match across systems, and then merging matching records into a single source of truth so that no duplicates exist. Creating match rules for the various source systems can be time consuming, involving close cooperation between business users (who know which records should be considered a match) and IT or other technical resources (who understand how to configure or code business rules in the MDM system to define the match). Match patterns that are not correctly identified or coded would generate incorrect matches or miss valid matches.

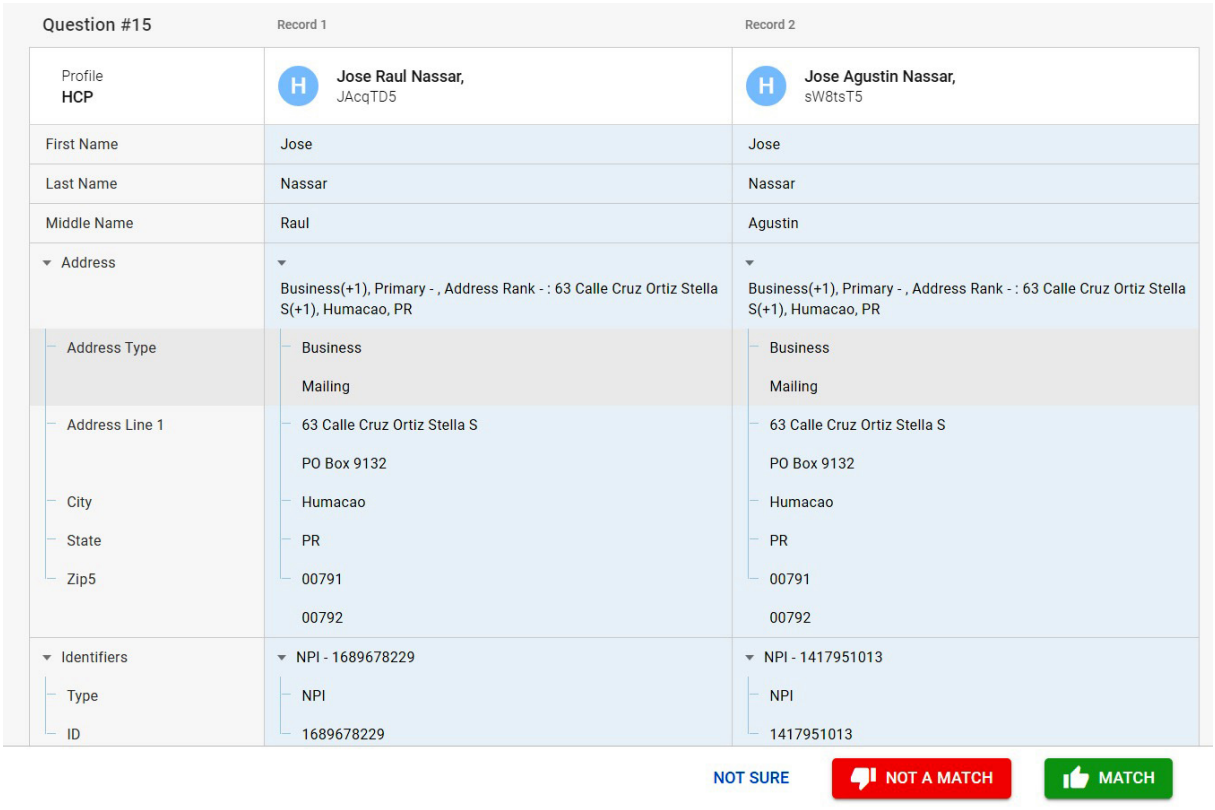
Typically, this is an iterative process. In short, it involves a lot of testing and adjusting matching criteria and configuring or coding—then re-doing it as many times as needed to ensure that all valid matches and only valid matches are identified. The more data sources involved, the longer this can take—it can be weeks or even months to get it right. Of course, any time you add a new data source, this needs to be done again.

Your data management implementation teams, data stewards, and other business users want a streamlined process, using advanced technologies to minimize the manual time and effort involved. At the same time, they want to ensure high data quality and the flexibility to add new source systems quickly.

Match IQ speeds the match-and-merge process

Our ML-based match-and-merge functionality is a core part of our Reltio Connected Data Platform. It uses machine learning technology to simplify and accelerate the matching process so your business users (without IT help) can easily create a model for matching the records, by simply selecting the entity type and related attributes. They can then train the ML model with our active learning process by reviewing pairs of records and indicating which are a match and which are not. As users confirm the matches, machine learning adjusts the matching model and presents additional record pairs to further refine the model. They can also use bulk review
to resolve multiple matches at once.

Potential matches selected used for training ML matching

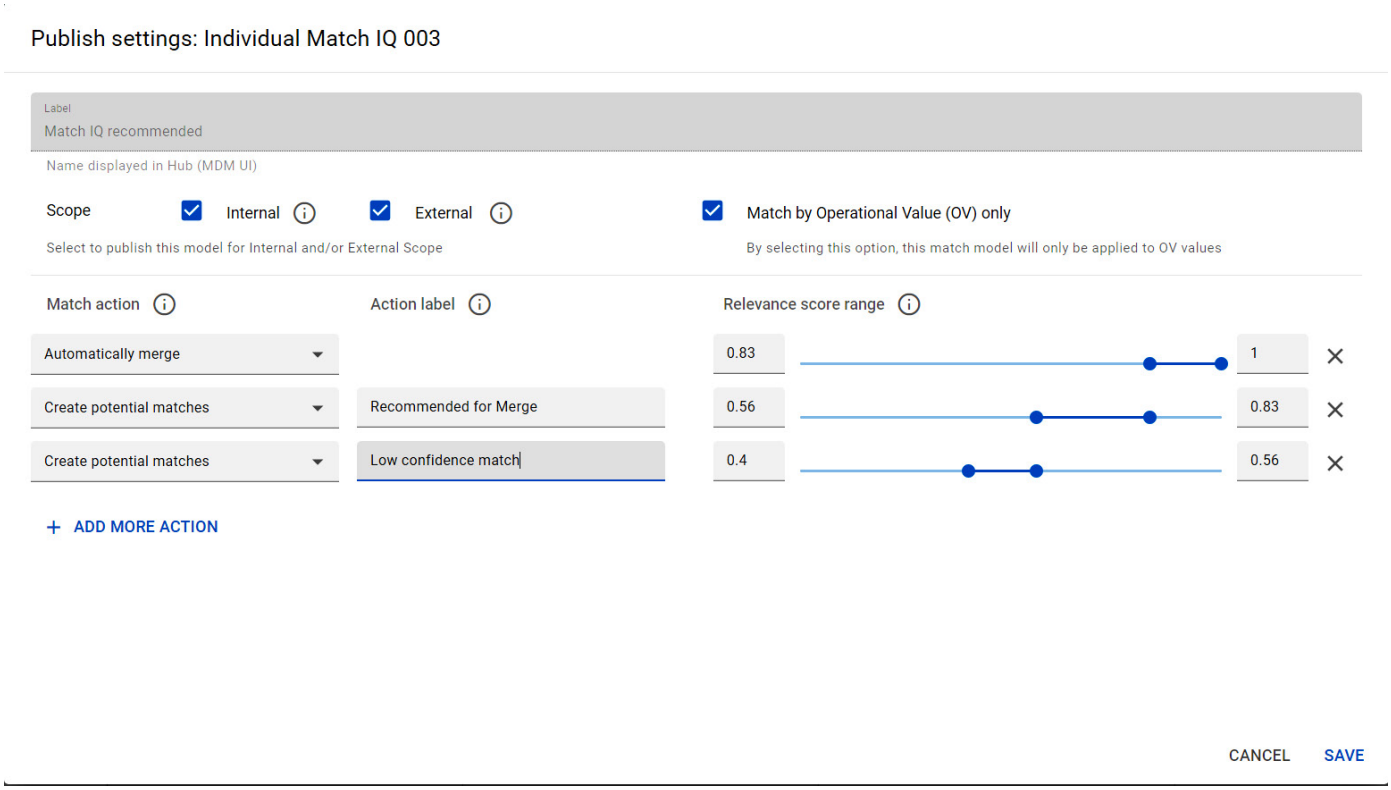


After a significant number of representative record pairs have been matched or not matched, the user can download and review the match results. The downloaded file shows a sample set of match results and a relevance score for each record pair. The higher the relevance score, the more likely the records match. If needed, the user can retrain the model by answering more questions or even creating an alternate model to compare the matching results.

After the results are satisfactory, the data steward or other user with approval authority can approve and publish the model to use with internal and/or external data. The user also provides publishing settings based upon the relevance score range—for example, to define that match pairs with a relevance score of .8 to 1 should be matched and merged.

The end-to-end process, driven and performed by business users, typically takes only a day or two to complete and produces the quality matches customers require.

Settings parameters for matching
and merging based on relevance score range



Data quality for trusted data foundation

Machine learning technology helps ensure unified and reliable data across virtually unlimited data sources. The ML matching model, created with active learning using resolutions of suspected matched pairs, can be effectively applied to future match pairs. This provides a consistent way for your business users and data stewards to match and merge data for increased quality, reliability, and business value.

Unlike many other MDMs, we offer proactive monitoring to identify match rules slowing down your process. Monitoring can automatically bypass poorly performing rules—and present the reason and recommendations for how to correct them.

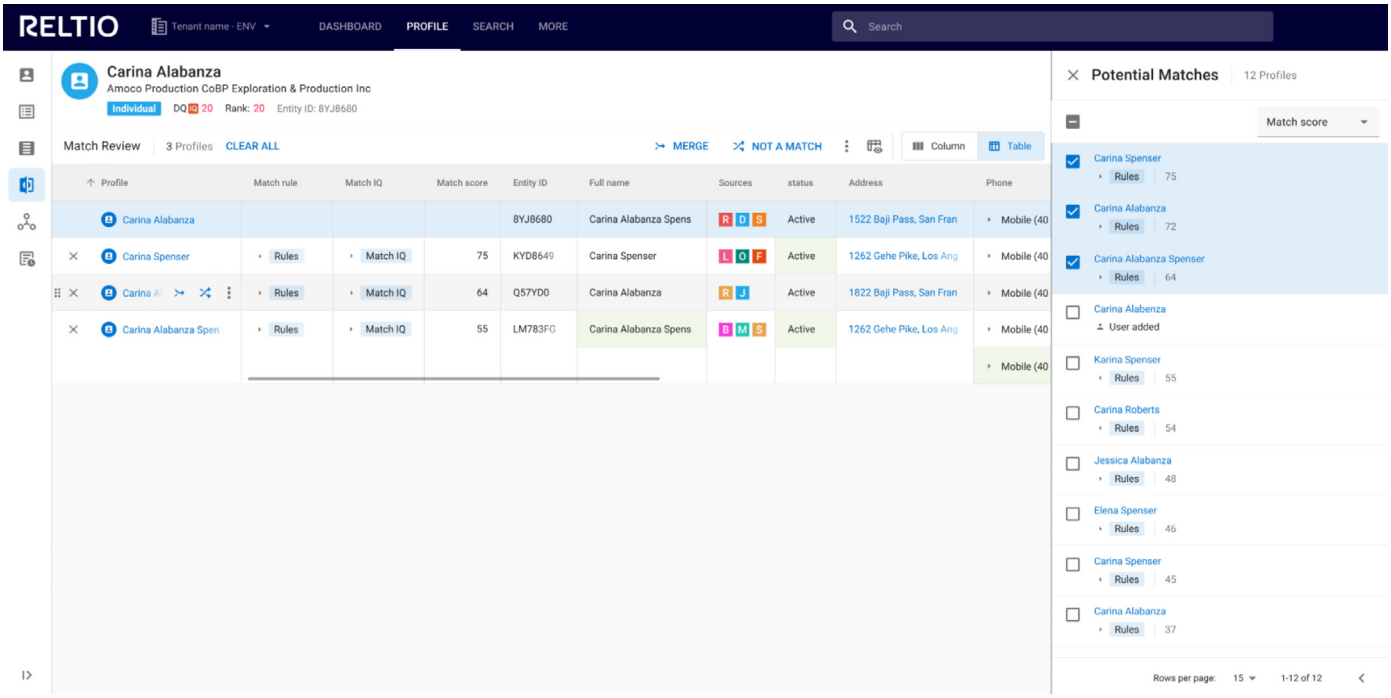
Accelerated time to value

The streamlined matching process—which does not require IT specialists or coding—enables you to get up and running faster and with less effort. Typically customers can progress from initial subscription to completing their match-and-merge operations in a matter of days. Compare this to the weeks or months required by more traditional approaches. This same process is used to perform matching for new data sources as they are added, providing additional time savings and increased productivity.

Match at the scale and speed of business

Once a matching model is trained, no user interaction is required, but the model can be retrained if needed. Because match-and-merge operations are performed using these models and calculated relevance scores, the process is rapid, consistent, and reliable. As your business grows or changes, the models can easily be adjusted to accommodate additional data sources. This enables matching and merging at the scale and speed of business.

Bulk review enables users to resolve multiple matches at once



Reduced errors in a repeatable process

No definition of matching requirements is needed; instead, users select matched pairs and machine learning creates the models. This greatly reduces the possibility of matching requirements not being correctly identified that might generate incorrect matches or miss valid matches. And because machine learning creates and adjusts the matching model without coding by IT specialists, coding errors are a thing of the past. This not only reduces errors in the match-and-merge process, but it also saves significant time as it creates a repeatable process.

Free resources for more meaningful work

With all the time saved by using machine learning for matching, those involved—your data owners, data stewards, IT, and other business users—will find they have more time available for work that adds value to the business. They can use their time to focus on creating better user experiences, data improvement initiatives, or streamlining other processes.

Features

- Simple user interface designed with the business user in mind.
- Machine-learning-based process that "understands" matching patterns and creates models based upon the active learning performed by business users—no coding required
- Creation and testing of multiple models before publishing the model that produces the most accurate results
- Setting match actions such as auto merge, potential match, and custom match actions based upon relevance scoring
- Matching done in real-time or batch mode—and upon internal or external data
- Matching API to move approved models from development or test tenants to target tenant (e.g., production) without retraining

US +1 (855) 360-3282
UK +44 (800) 368-7643

@Reltio
facebook.com/ReltioHub
linkedin.com/company/reltio-inc