tamr

Tamr Integration With Master Data Management (MDM)

Big Data Has Created Gaps In Master Data Management (MDM)

The concept of Master Data Management (MDM) has been around for decades. MDM is the set of technologies and processes that ensures that organizations have a single, authoritative version of the truth from which they operate. Master Data, or information that refers to those elements core to the operation of a business, is used widely in organizations by both operational and analytical applications. Some examples of master data include customer lists, product lists and supplier lists. Most systems that drive critical business decisions depend on a consistent definition of Master Data, so making sure this data remains consistent is a top imperative for IT organizations.

MDM systems fill this role by providing several pieces of functionality that help create this single view of an entity, and synchronize it with downstream applications. This includes:

Consolidation
Unifying all master data records to create a comprehensive understanding of an entity

Survivorship
Selecting the most appropriate attribute values for each record

Cleansing
Validating the accuracy of the values

Governance
Ensuring compliance with regards to adding or modifying the resulting single 'good' record related to that entity, also called the "golden record".

When MDM is implemented correctly, the result is inevitably a more efficient organization that more widely uses data-driven approaches to decision-making. The specific methods for achieving MDM functionality have worked well in the past, particularly when organizations only had a few critical data sources in their enterprise. Traditional methods are usually very labor-intensive where record matches are typically defined by a set of rules that owners of MDM implementations put in place and rigorously enforce. This requires in-depth knowledge of the data sources themselves.

Unfortunately, in the age of Big Data, companies are now forced to rethink traditional MDM methods. Enterprises across all industries are facing an explosion of data flooding into their environments, from online customer interactions to mobile payments. Business activities such as mergers and acquisitions compound this problem, as organizations are forced to deal with net new IT infrastructure and data that they've never touched. This has exposed limitations in traditional MDM products, including:
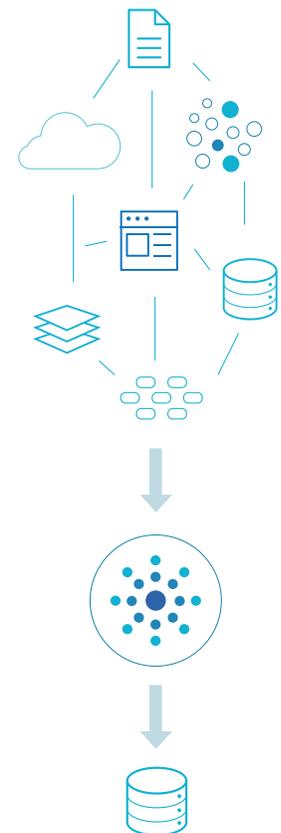
MDM is the set of technologies and processes that ensures that organizations have a single, authoritative version of the truth from which they operate.

+ **Speed and scalability of integrating data sources:** The increasing volume of data sources poses a big threat to speed and scalability. Given the highly manual nature of traditional MDM operations, managing more than a dozen data sources requires an increasingly larger investment in time and money. Consequently, it's often very difficult to economically justify scaling the operation to cover all data sources. Additionally, the speed at which data sources are integrated is often contingent upon how quickly employees can work, which will be at an increasingly unproductive rate.

+ **Problems with agility:** MDM products are very deterministic and up-front in the generation of matching rules. Manual effort is required to understand what constitutes potential matches and then define rules to be implemented around it. For example, in matching addresses, there could be thousands of rules that need to be written. This becomes increasing difficult to manage as data sources become greater in volume, and results in the risk that by the time new rules (or rule changes) have been implemented, business requirements have changed.

Tamr Can Integrate With MDM Products To Improve Speed And Scalability While Maintaining Quality

Tamr was architected to deal with the exact problems that MDM products face in the age of Big Data. It serves as a complementary component to MDM and empowers companies to embrace volume and variety, not shy away from it. Whereas traditional MDM products were built top-down to deeply manage a few data sources, Tamr was built to manage data volume and variety using a unique blend of probabilistic machine learning and expert sourcing. This enables Tamr to effectively and efficiently identify unique entities and attributes within datasets. Specifically, Tamr has the ability to:

+ **Map new sources:** As you bring new sources into Tamr, it will automatically recommend how the schema of the incoming application maps to the unified schema, based on the metadata and the data associated with the new source.

+ **De-duplicate records:** Tamr builds a model that helps recognize when two records belong to the same cluster. The model is able to confidently match 90%+ of records. Only a small percentage remains un-matched, and is then referred to experts.

Tamr produces a reference map (i.e. schema mappings and record keys linked to data sources) which can then be used in Master Data Management systems to produce golden records. Given the capabilities above, Tamr can help an MDM customer:

+ **More quickly add a datasource** by reducing the amount of human effort needed to integrate it. This is primarily due to Tamr's utilization of machine learning and expert sourcing.

+ **Scale to cover more data sources**. Tamr continuously learns from your datasets and input so the incremental cost and time required to integrate new sources is continuously declining. Consequently, organizations end up integrating more of their data, enabling the creation of new golden records (e.g. enhance coverage) or enrichment of existing ones with additional attributes (e.g. enhance completeness).
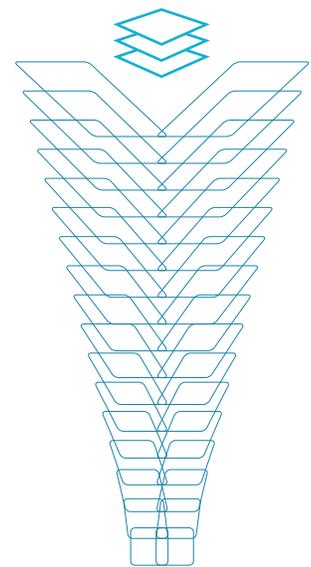
Speed and scalability of integrating data sources: The increasing volume of data sources poses a big threat to speed and scalability. Given the highly manual nature of traditional MDM, managing more than a dozen data sources requires an increasingly larger investment in time and money.

$

Cost to Integrate

Traditional Methods

tamr

Data Sources
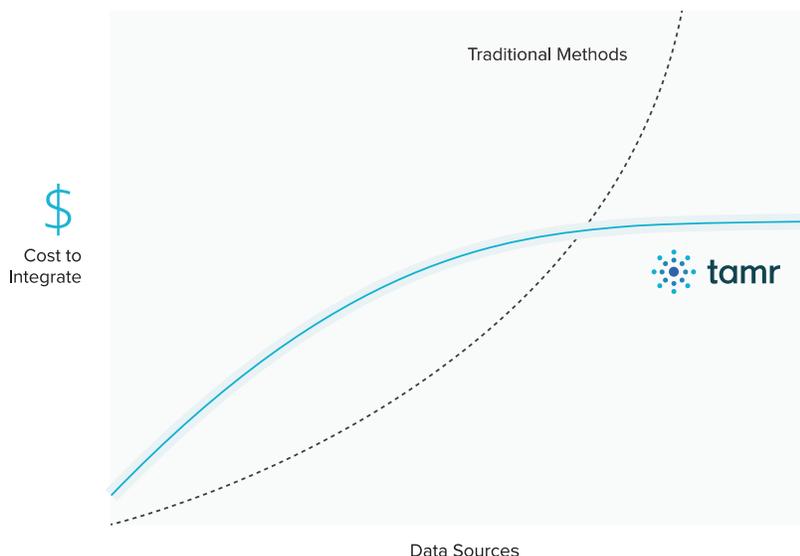
+ **Maintain a high level of data quality.** Any matches that have uncertainty associated with them are reviewed and validated by the people most qualified to make judgements - the business experts. This validation, in turn, helps Tamr improve its own model, further increasing Tamr's accuracy and reducing human effort required in subsequent cycles.

+ **Improve business agility.** Tamr enables organizations to deliver quickly on MDM projects and changes, reducing the risk that business requirements would have already changed by the time the project was delivered.

Specifically, Tamr integrates with Master Data Management products in the following way:
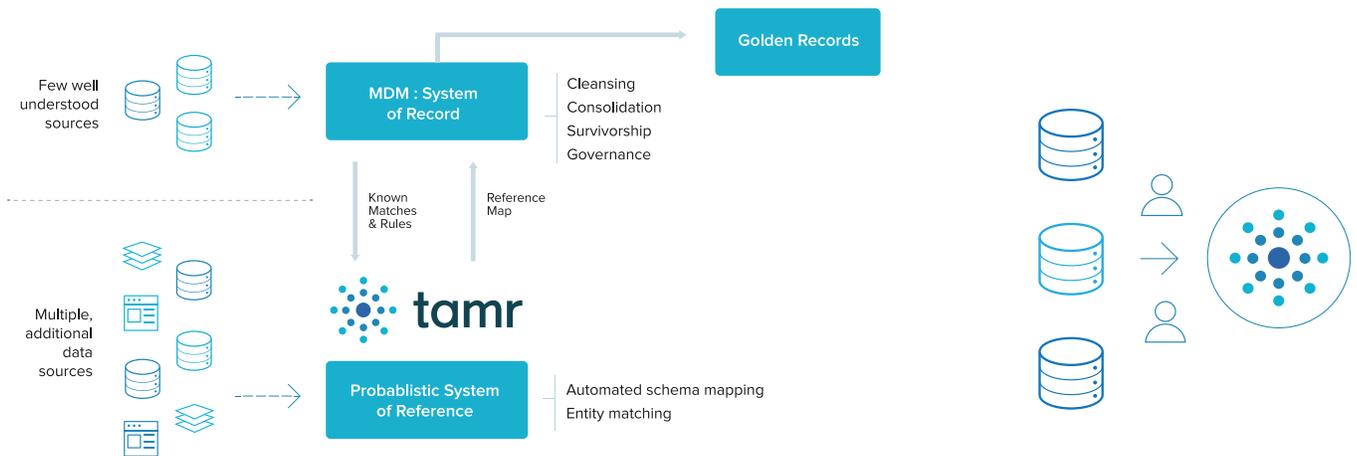
+ A customer's existing MDM implementation will connect to and ingest the existing few, well understood data sources in their organization then generate entity matches and, subsequently, IDs from the data sources.

+ Tamr will act as the system of refererence: the product will connect to and ingest information from the multiple, additional data sources that would not be efficiently and cost-effectively reachable by existing MDM products.

+ The MDM product will pass Tamr the matches it identified as well as the methodologies for creating those matches (e.g. rules).

+ Tamr will utilize MDM's matching methodology to influence it's matching capabilities, which are primarily derived through a combination of machine learning and expert sourcing, and develop IDs for entities that exist in the multiple, additional data sources that MDM does not manage.

+ Once complete, Tamr has essentially developed a "reference map" that identifies all unique entities in the entire enterprise dataset.

Tamr will utilize MDM's matching methodology to influence it's matching capabilities, which are primarily derived through a combination of machine learning and expert sourcing, and develop IDs for entities that exist in the multiple, additional data sources that MDM does not manage.

+ Tamr will then pass the completed reference map to the existing MDM product, which will continue to act as the system of record. Customers who have already invested in MDM and built survivorship rules, etc. within it will continue to take advantage of those features.

+ Finally, the MDM product will leverage Tamr's reference map to either create new golden records (e.g. enhance coverage) or enrich existing ones with additional attributes (e.g. enhance completeness).
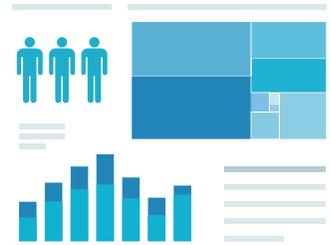
This architecture leverages the capabilities of existing MDM investments while enhancing the process with the exceptional speed and scalability of Tamr.

Proven Success

Tamr's customers have had significant success with this model of master data management. They recognize that although they have worked to build out robust MDM practices, speed and scalability in the new world isn't possible given traditional methods. Tamr was needed to fill that gap. In one particular example, a large media company utilized Tamr to identify unique entities within a multitude of data sources not managed by their MDM installation. The results of this model were very impressive. In the past, it took the customer over six months to integrate a new dataset into their environment, due to the fact they took a traditional approach to MDM and manually worked to map schemas and match records. Tamr helped reduce this time to merely two weeks, resulting in a significantly improved time to value.

About Tamr

Tamr, Inc., provides a data unification platform that dramatically reduces the time and effort of connecting and enriching multiple data sources to achieve a unified view of siloed enterprise data. Using Tamr, organizations are able to complete data unification projects in days or weeks versus months or quarters. For your own personalized Tamr demo, visit www.tamr.com.

This architecture leverages the capabilities of existing MDM investments while enhancing the process with the exceptional speed and scalability of Tamr.