

Deduplication and Data Cloud: You Have an Architecture Problem

By Mark Cane

July 22, 2026

Branded content with [clearMDM](#)

Data Cloud, now Data 360, can unify and activate customer data, but it cannot replace the governance, survivorship, and stewardship needed to create trusted Customer 360 outcomes.

Most Salesforce data quality programmes start in the same place: duplicate Accounts, Contacts, or Leads.

This is understandable. Duplicates reduce user trust, distort reporting, create confusion for sales and service teams, and make automation difficult to rely upon. For many Salesforce teams, duplicate rules, matching rules, or a deduplication tool are the first practical steps towards improving CRM data quality.

However, achieving a true Customer 360 requires more than duplicate reduction. In a modern Salesforce architecture, high-quality, governed customer data directly underpins CRM processes, Flow logic and dashboards, while driving Data 360 insights, personalized journeys, generative AI and Agentforce outputs. Duplicates are only part of the problem. Incomplete, inaccurate, poorly governed, or relationship-blind data can be just as damaging.

Customer data quality is no longer just an administrative clean-up task. It is an architectural concern.

Deduplication Is Necessary, But Not Sufficient

Salesforce provides useful native capabilities for duplicate management. [Duplicate Rules](#) and Matching Rules can help prevent duplicate records at the point of entry, alert users to potential matches, and support basic data hygiene.

For many organizations, this is a sensible foundation, particularly where data volumes are manageable, and the main requirement is to prevent obvious duplicates from being created in the first place.

As environments become more complex, native controls are often extended with third-party deduplication tools from the [AgentExchange](#). These tools can improve match accuracy, support advanced fuzzy logic, automate bulk processing, introduce scoring, and provide more scalable review and merge capabilities.

This is an important progression. Third-party deduplication tools solve real problems that native duplicate management was not designed to address. But there is still a distinction that matters.

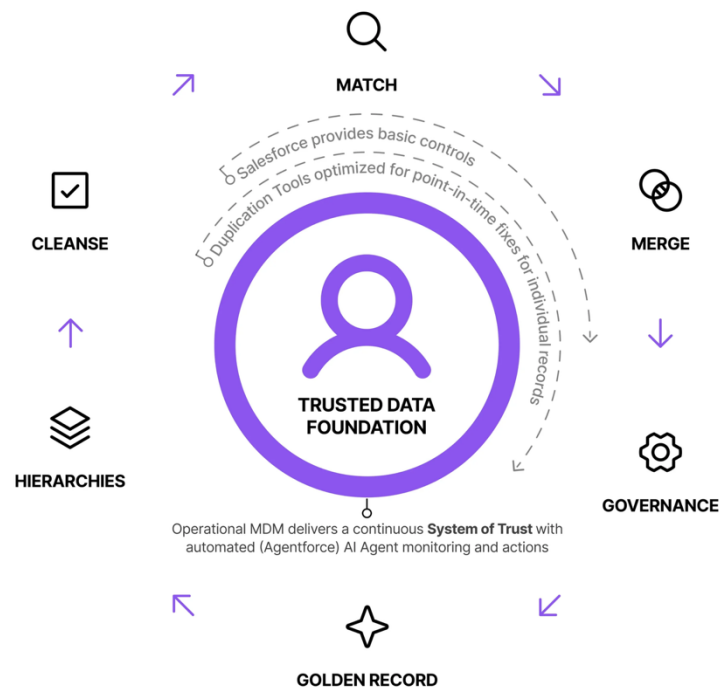
Deduplication tools improve how duplicates are identified and resolved. Master Data Management changes how trusted customer data is governed and maintained over time.

Master Data Management (MDM) provides the governance discipline for trusted customer data. In Salesforce CRM, this is increasingly delivered through **Operational MDM**: an embedded approach that governs Golden Records, survivorship, stewardship, lineage, and relationships where operational work happens.

A Maturity Model for Salesforce Customer Data

Native Salesforce duplicate management focuses on duplicate identification and [manual merging](#). Third-party deduplication tools extend that model with stronger matching, automation, scoring, bulk processing, review queues, and light governance controls around duplicate resolution.

Operational MDM extends further. It introduces Golden Records, survivorship rules, stewardship workflows, data lineage, source ownership, and relationship hierarchies. It is less concerned with a one-off merge event and more concerned with the ongoing question of which version of the customer the business should trust.



Salesforce customer data management maturity progresses from native duplicate management, through deduplication tooling to Operational MDM with Golden Records and real-world relationship hierarchies.

This is not a case of one layer invalidating the others. Native duplicate management, third-party deduplication tools, and Operational MDM all have a place. The question for architects is whether the organization only needs duplicate resolution or whether it needs a governed operating model for trusted customer data.

In many Customer 360 programmes, the latter becomes increasingly important.

Customer 360 Needs More Than a Clean Record

Deduplication answers an important question: *Do these records represent the same customer?*

However, achieving a true Customer 360 requires answering broader, operational questions:

- **Ownership:** Which system owns this customer attribute?
- **Trust:** Which email address should be trusted?
- **Survivorship:** Which phone number should survive?
- **Context:** Which household relationship is current?
- **Compliance:** Which consent indicator takes precedence?
- **Action:** Should records be merged, mastered, suppressed, or retained?
- **Oversight:** Who reviews exceptions?
- **Downstream:** What should target platforms and AI agents rely on?

These are not simply deduplication questions – they are governance questions.

In practice, this is where many Customer 360 programmes discover that data quality is more difficult than expected and not just a tooling decision. Creating a joined-up view of the customer is not just about matching records together. It is about deciding which data is authoritative, how that decision is made, and how trust is maintained as the data changes.

Customer 360 is not just a data model. It is an operating model.

Where Data 360 Fits

[Data 360 \(formerly Data Cloud\)](#) plays an important role in this architecture. It connects data sources, harmonizes customer data, resolves identities, supports segmentation, generates insights, and activates customer data across channels and Salesforce experiences.

That makes it highly valuable. For many organizations, Data 360 is the layer that turns customer data into action. However, Data 360 should not be treated as a substitute for governed source data or as a solution to operational data quality issues.

For example, Data 360 may identify that “Jane Smith” in Sales Cloud, “J. Smith” in a support platform, and “jane.smith@email.com” in Marketing Cloud relate to the same individual. That is valuable identity resolution.

But the organization still needs rules to determine which email address, phone number, account relationship, support preference, or consent value should be treated as trusted. It also needs to know where that decision is governed, how exceptions are handled, and what happens when the underlying data changes.

This is the core architectural distinction: Data 360 unifies and activates customer context, while Operational MDM governs the trusted operational version of the customer at source.

MDM creates trust. Data 360 creates activation. Customer 360 needs both.

Golden Records and Unified Profiles Are Not the Same Thing

One common source of confusion is the difference between a Golden Record and a unified profile.

A Data 360 unified profile is designed for **harmonization, insight, segmentation and activation**. It brings customer data together from multiple sources so the organization can understand, analyze, and act on customer context.

A Golden Record is different: it is the **governed operational version of the customer**. It defines the trusted version of the record that users, processes, reporting, and automation can rely on inside Salesforce CRM.

Both are valuable, but they are not interchangeable. A unified profile helps the organization understand and activate customer data, whereas a Golden Record helps the organization operate on the trusted version of that data.

For Customer 360, the strongest architecture is rarely an either-or choice between Data 360 and MDM. It is Data 360 plus MDM. Operational MDM improves the quality, consistency, and governance of CRM data, while Data 360 leverages that higher-quality foundation for unification, segmentation, insight, and activation.

Agentforce Raises the Bar

The same distinction becomes even more important with Agentforce.

It is true that AI relies on clean data, but the architectural point is more specific: AI agents inherit the trust model of the data beneath them.

If Salesforce contains duplicate, fragmented, or conflicting customer records, an AI agent may use that data confidently but incorrectly. It may recommend an action based on an outdated email address, the wrong or absent household relationship, a conflicting consent value, or an incomplete service history.

This is also where a new operational pattern is emerging: AI-powered data stewardship. Human data stewards are often tasked with reviewing potential duplicate matches, assessing confidence scores, investigating conflicting records, and deciding whether matched records should be accepted, rejected, or investigated further. That process is essential, but it can quickly become a bottleneck at scale.

When AI stewardship is combined with governed Operational MDM, Agentforce can support data stewards with match reasoning, contextual guidance, suggested next actions, and conversational support. In more mature scenarios, AI-assisted stewardship can automate lower-risk decisions while escalating ambiguous cases for human review.

Trusted AI does not depend only on better data. It also depends on better governance processes for maintaining that data over time.

Architect Checklist Before Customer 360 Activation

Before customer data is activated across Data 360, Agentforce, or wider Customer 360 initiatives, architects should be clear on:

- **Data ownership:** Which systems own which customer attributes?
- **Matching logic:** What is the definition of a duplicate for each domain?
- **Survivorship policy:** Which records should be merged, mastered, suppressed, or retained?
- **Field-level lineage:** Which source wins for each field?
- **Mastering strategy:** Where is the Golden Record maintained?
- **Data stewardship:** Which changes require stewardship review?
- **Audit trail:** How is data lineage preserved?
- **Relationship management:** Which customer relationships and hierarchies need to be modelled?
- **AI safety:** Which data is safe for AI Agents to use when recommending or taking action?

These questions move the conversation beyond deduplication and into operational data governance.

Moving Beyond Deduplication With clearMDM

Deduplication remains important. No Salesforce data strategy should ignore duplicate records, weak matching, or inconsistent customer data. But treating deduplication as a one-off clean-up exercise is rarely enough. Customer 360 requires trusted customer data, governed at source and maintained over time.

This is where [clearMDM](#) fits into the Salesforce architecture.

clearMDM is a Salesforce-native Operational MDM platform that helps organizations move beyond point-in-time deduplication towards continuously governed Golden Records, survivorship, and relationship-aware customer data inside Salesforce CRM.

Rather than competing with Data 360, clearMDM complements it by improving the quality, consistency, and trustworthiness of operational CRM data before it is harmonized, segmented, or activated. In this model, clearMDM provides the governed trust layer in Salesforce CRM, while Data 360 provides the broader unification and activation layer across the Salesforce data ecosystem.

clearMDM also extends this model through the [clearMDM Data Steward Agent for Agentforce](#), which supports AI-powered data stewardship use cases such as AI-guided duplicate resolution, conversational stewardship support, AI-generated match reasoning and automated stewardship decisions. The agent works alongside human data stewards to reduce manual review effort, improve consistency, and help maintain trusted Salesforce customer data at scale.

By improving trust at source, clearMDM helps provide the customer data foundation required for Customer 360, Agentforce, and AI-driven activation.

To explore this architecture further, join the [Deduplication and Data Cloud: You Have an Architecture Problem webinar](#) or download the [white paper](#) and learn how Salesforce teams can build trusted customer data foundations for Customer 360, Data 360, and Agentforce.

Prefer to discuss your Salesforce data architecture directly? [Contact clearMDM](#).

The Author



Mark Cane

Mark is the Founder of clearMDM and a Salesforce CTA, certified in 2012.