Harmonization & Unification (Identity Resolution)

In **Salesforce Data Cloud**, **Harmonization** and **Unification** are the processes that combine data from multiple sources into a **single, trusted, and complete profile** for each customer or entity.

- **Harmonization** – Standardizes and aligns data from multiple systems (cleans, formats, and normalizes values).
- **Unification (Identity Resolution)** – Merges duplicate or related records into a single unified record (also called a *golden record*).

These two steps together create the foundation of the **Customer 360 profile**.

---

2. The Identity Resolution Pipeline

The end-to-end process looks like this:

Step	Process	Description
1	Ingest Data	Collect raw data from multiple systems into DLOs
2	Harmonize	Standardize and align field formats
3	Match	Identify duplicate or related records using match rules
4	Reconcile	Merge and prioritize attributes to create the final unified record
5	Unify	Store the final unified record in the unified object (e.g., Unified Individual)

---

3. Example Scenario

Let's imagine a company, **ShopEase**, collects customer data from multiple systems:

- CRM system
- E-commerce website
- Support system

Each system may record the same person differently.

Raw Data Before Harmonization

Source	First Name	Last Name	Email	Phone	City	Last Purchase Date
CRM	John	Smith	john.smith@email.com	9876543210	New York	2025-10-15
Web	Jonathan	Smith	john.s@email.com	(987) 654-3210	NY	2025-10-14
Support	Jon	S.	john.smith@email.com	9876543210	New York City	2025-10-15

Clearly, these represent the **same person**, but they appear **slightly different** across systems.

---

4. Step 1: Harmonization

Definition

Harmonization is the process of cleaning and standardizing data before matching.

Goal: Ensure data from different systems uses consistent formats and conventions.

Field	Harmonization Action	Example Result
Name	Standardize casing and abbreviations	“Jonathan Smith”
Phone	Remove formatting symbols	“9876543210”
City	Normalize regional names	“New York”
Email	Convert to lowercase	“john.smith@email.com”

Harmonized Data

Source	First Name	Last Name	Email	Phone	City	Last Purchase Date
CRM	John	Smith	john.smith@email.com	9876543210	New York	2025-10-15
Web	Jonathan	Smith	john.s@email.com	9876543210	New York	2025-10-14
Support	Jon	Smith	john.smith@email.com	9876543210	New York	2025-10-15

---

5. Step 2: Matching Rules

Definition

Matching rules define **how Salesforce Data Cloud identifies records that represent the same individual or entity**.

Each rule compares fields between records to determine if they should be merged.

There are **three match rule types**:

1. **Exact Match**
2. **Normalized Match**
3. **Fuzzy Match**

5.1 Exact Match

- Compares fields that must match exactly.
- Used for stable, unique identifiers (email, customer ID, loyalty ID).
- Case-insensitive by default.

Example Rule:

```
email == email
```

Result:

Records with the same email (john.smith@email.com) are matched.

5.2 Normalized Match

- Matches fields that may vary in format but represent the same data.
- Normalizes before comparison (removes spaces, dashes, parentheses).
- Common for phone numbers, postal codes, or address fields.

Example Rule:

```
normalize(phone) == normalize(phone)
```

Result:

Matches (987) 654-3210 with 9876543210.

5.3 Fuzzy Match

- Used when data has **variations or typos**.
- Uses similarity algorithms (phonetic, Levenshtein distance, etc.) to compare names or text.
- Useful for names, company names, or addresses.

Example Rule:

```
fuzzy(first_name) AND fuzzy(last_name) AND exact(email_domain)
```

Result:

Matches “Jon Smith” and “John Smith”.

---

6. Choosing the Right Field Combinations

Different combinations of fields increase match accuracy and reduce false positives.

Match Rule Type	Fields Used	Use Case	Boosted Match Rate
Exact	Email	Same email across systems	High precision
Exact + Normalized	Email + Phone	Reduces false duplicates	High precision and recall
Fuzzy + Exact	Name (fuzzy) + Email (exact)	Accounts for misspellings	Balanced accuracy
Fuzzy + Normalized	Name + Phone	Works where email missing	Broader matching coverage

Example:

Match rule = Fuzzy(Name) + Normalized(Phone) can link:

- “Jon Smith”, phone “(987)654-3210”
- “John Smyth”, phone “9876543210”

Both would resolve as the same individual.

---

7. Step 3: Reconciliation Rules

Once matches are found, Salesforce must **decide which data values to keep** in the unified record.

Definition

Reconciliation rules determine **which source wins** when multiple records contain different values for the same field.

There are three main reconciliation strategies:

Rule Type	Description	Example Behavior
Source Priority	Choose the value from a preferred data source	CRM data overrides Web or Support
Most Recent	Choose the latest updated value	Picks data from the most recent record
Most Frequent	Choose the value that appears most often across matched records	Uses the most common city or email

---

Example: Conflicting Data

Source	Email	City	Last Purchase Date	Updated Date
CRM	john.smith@email.com	New York	2025-10-15	2025-10-15
Web	john.s@email.com	NY	2025-10-14	2025-10-16
Support	john.smith@email.com	New York City	2025-10-15	2025-10-13

Applying Reconciliation Rules

Field	Rule	Selected Value	Reason
Email	Source Priority (CRM)	john.smith@email.com	CRM preferred source
City	Most Frequent	New York	Appears most often
Last Purchase Date	Most Recent	2025-10-15	Latest transaction date

Unified Record Output

Unified Field	Final Value	Rule Used
First Name	John	Most Frequent
Last Name	Smith	Most Frequent
Email	john.smith@email.com	Source Priority
Phone	9876543210	Exact
City	New York	Most Frequent
Last Purchase Date	2025-10-15	Most Recent

---

8. Step 4: Warnings and Validation

During unification, Salesforce Data Cloud may generate **warnings** or **validation alerts**, such as:

Warning Type	Description	Action
Missing Primary Key	One or more records lack a unique identifier	Fix DLO/DMO mapping
Mismatched Data Type	Inconsistent field types between sources	Adjust schema or mapping
Low Match Confidence	Fuzzy match below confidence threshold	Tune match rules
Duplicate Retention Conflict	Multiple candidates meet same match rule	Review reconciliation priorities

Use **Data Explorer** or **Identity Resolution Logs** to review and correct issues.

---

9. Example: End-to-End Flow

Step 1: Input (Harmonized Data)

Source	Name	Email	Phone	City
CRM	John Smith	john.smith@email.com	9876543210	New York
Web	Jonathan Smith	john.s@email.com	(987)654-3210	NY
Support	Jon Smith	john.smith@email.com	9876543210	New York City

Step 2: Match Rules Applied

Fuzzy Match (Name) + Normalized (Phone) + Exact (Email Domain)

→ All 3 records matched as one individual.

Step 3: Reconciliation

- **Email:** CRM wins (source priority)
- **City:** "New York" (most frequent)
- **Name:** "John Smith" (most frequent)
- **Phone:** 9876543210 (normalized exact)

Step 4: Unified Profile Output

Unified ID	Name	Email	Phone	City	Last Purchase Date
U1001	John Smith	john.smith@email.com	9876543210	New York	2025-10-15

---

10. Real-World Use Case: Retail Customer 360

Background

A national retailer uses Salesforce Data Cloud to unify data from CRM, website, and loyalty systems.

They struggled with duplicate records — the same customer appeared under multiple variations.

Implementation

Step	Configuration	Description
Match Rules	Fuzzy (Name), Normalized (Phone), Exact (Email)	Captures spelling and format variations
Reconciliation Rules	Source Priority (CRM), Most Recent (purchase)	CRM trusted, but always keeps latest order
Outcome	Unified “Customer 360” profiles	One record per real person

Results

Metric	Before	After
Duplicate records	20%	< 2%
Match rate	72%	95%
Marketing accuracy	Improved by 40%	Higher personalization
Data confidence	Low	High

---

11. Best Practices

Area	Best Practice	Why It Matters
Field Selection	Include multiple identifiers (Email, Phone, Name)	Improves match accuracy
Harmonization	Clean data formats before unification	Reduces false mismatches
Match Rules	Test combinations on sample data	Optimize for accuracy and recall
Reconciliation Rules	Use source priority for trusted systems	Keeps high-quality data
Validation	Review match rates and confidence logs	Detect and fix issues early
Documentation	Maintain mapping and rule configuration sheets	Ensures repeatability and governance

---

12. Summary

Process	Purpose	Example
Harmonization	Standardize and clean raw data	Convert “NYC” to “New York”
Match Rules	Identify duplicates using Exact, Normalized, or Fuzzy logic	Match “Jon” with “John”
Reconciliation Rules	Resolve conflicting field values	Choose latest email or city
Unification	Produce a single, golden customer record	One record per person
Output	Unified Individual (Customer 360)	Ready for activation and analytics

---

Key Takeaway

In Salesforce Data Cloud, **Harmonization and Unification (Identity Resolution)** are the foundation of Customer 360.

Using the right combination of **match rules (Exact, Normalized, Fuzzy)** and **reconciliation logic (Source Priority, Most Recent, Most Frequent)** ensures you get **accurate, deduplicated, and trusted profiles** ready for segmentation and personalization.