# Step-by-step Guide Data Harmonization & Unification

## Introduction to Data Harmonization

Data Harmonization is an essential part of Data Cloud. With Data Harmonization (Data Modeling), we can create a 360 view of our customers by merging data from multiple sources. This gives us the power to analyze a full view of the customer, allowing for the creation of quick & easy insights.

Data Harmonization helps to ensure consistency & identity resolution, in addition to allowing for analytics across data sources. Ultimately, the goal of ***data harmonization*** is to remodel data to align it with our commercial enterprise and to cleanse, standardize, and structure that data to ensure consistency & accuracy.

In this exercise, we will use our example company - Sunshine Trails Hospitality, and use Data Harmonization to consolidate guest data to get a clear understanding of past stays & reservations. Let's tap into the power of Data Harmonization to unify guest data, lay the foundation for future analysis, & provide the best experience possible to Sunshine Trails Hospitality guests!

## Technical Benefits of Data Harmonization & Unification

The technical benefits of data harmonization & unification within Data Cloud include the effective facilitation of data analysis, and the consolidation of discrepancies and inaccuracies in data types, formats, structures, and semantics. Data Harmonization within Data Cloud uses a data model consisting of 80+ pre-defined Data Model Objects (DMOs). As mapping occurs between the Data Lake Object (DLO) & DMO, this high volume of DMOs is designed to accommodate a wide variety of use cases. Data Cloud also uses a Fully Qualified Key (FQK), designed to avoid conflicts when data from multiple sources is harmonized.

**Key Technical Features:**

- **Supports Dynamic Use Cases** like personalized customer experiences, real-time analytics, and AI/ML-driven insights.
- **Uses Machine Learning** to suggest data quality improvements
- **Employs Advanced Data Deduplication**, **Cleansing, and Enrichment Techniques** to ensure high-quality, actionable data.
- **Prepares Data for AI/ML Applications** by ensuring it is clean, consistent, and contextually relevant.
- **Facilitates Advanced Analytics and Predictive Modeling** with harmonized datasets.

## Key Concepts of Data Harmonization

- Harmonization (data modeling) enables all subsequent downstream activity in Data Cloud
- With harmonization, you have different expressions of the exact same data points and you're unifying them.

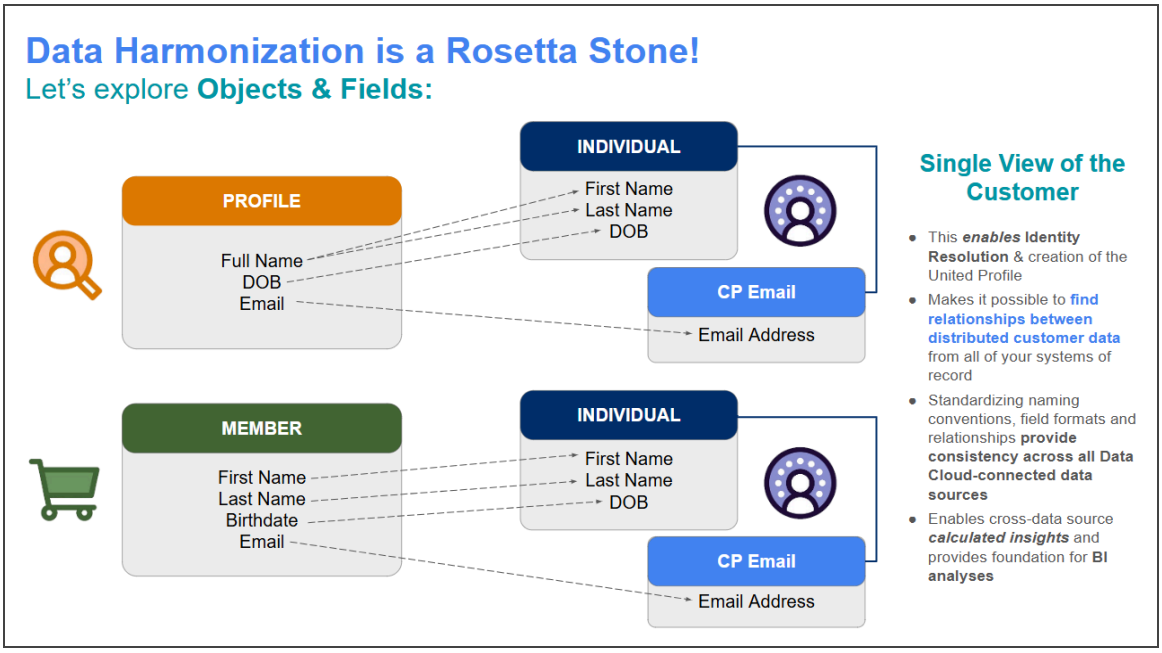
# Step-by-step Guide Data Harmonization & Unification

- Harmonization gives us a standardized view of the data from the different sources which feed into Data Cloud
- We use Data Cloud to translate the source data into one common pool of data which is combined into a common data model
- Harmonization is the standardization of naming conventions, field formats, etc.

| Key Word | Description |
|---|---|
| Data Source | Available information from various data sources used to populate Data Cloud. Ex: Sales Cloud, Service Cloud, Amazon S3, Google Cloud Storage. |
| Data Streams | Extract specific information from diverse data sources & establish connections to data repositories. |
| Data Source Object (DSO) | Temporary storage for staging data; remains in its raw format. |
| Data Lake Object (DLO) | Primary workspace for data inspection & preparation. Allows users to map fields & apply additional transformations. |
| Data Model Object (DMO) | Serves as a virtual gateway to the Data Lake. Generates real-time results from queries based on data stored in the DLOs. |
| Data Space | Allows you to segregate your data, metadata, and processes by categories such as brand, region, or department. |

Data Harmonization acts as our ***Rosetta Stone*** - translating data from multiple sources to Unified Profiles which can serve as a single source of truth for our team!

# Step-by-step Guide Data Harmonization & Unification




**Engaging Data Harmonization**

When discussing Data Harmonization, it's helpful to explore how it affects our data in Salesforce Data Cloud. Data Harmonization allows us to map ingested data into the Customer 360 data model, making it available for unification, segmentation, and activation. In our Sunshine Trails Hospitality example, the use cases shown below help to illuminate some of the ways that Data Harmonization can begin to bring out data together for that 360 view of the customer which helps Data Cloud soar!

| Data Type | Description | Use Case |
|---|---|---|
| Reservation Data | Includes reservation history, room preferences, check-in/check-out times, booking source (OTA, direct, GDS), booking time, and cancellation history. | To understand guest preferences, predict stay behaviors, and provide personalized check-in service. |
| Guest Profile Data | Contact information, communication history, guest preferences, historical interactions, and any marketing engagements | To create and maintain an accurate guest profile for tailored marketing, communication, and personalization of services |
| Survey & Feedback Data | Ratings, review comments, survey responses, Net Promoter Scores (NPS). | To gauge guest satisfaction and improve services based on direct feedback. |
| IoT & Connected Devices Data | In-room IoT data such as room temperature preferences, Wi-Fi usage, and other connected devices. | To personalize the in-room experience (e.g., adjust room conditions based on preferences). |

| | | |
|---|---|---|
| Service Request & Support Data | Service requests, complaint resolutions, chat transcripts, and support call logs. | To enhance guest experience by tracking and resolving issues quickly, and prevent future complaints. |
| Messaging Data | Text messages, email responses, RTP, WhatsApp interactions, and short codes. | To enable communication with guests via their preferred channels and provide personalized messages. |
| Booking Channel Data | Information on the source of booking (direct, OTA, GDS), cancellation history, and booking time. | To understand guest booking behavior and improve marketing strategies for each booking channel. |

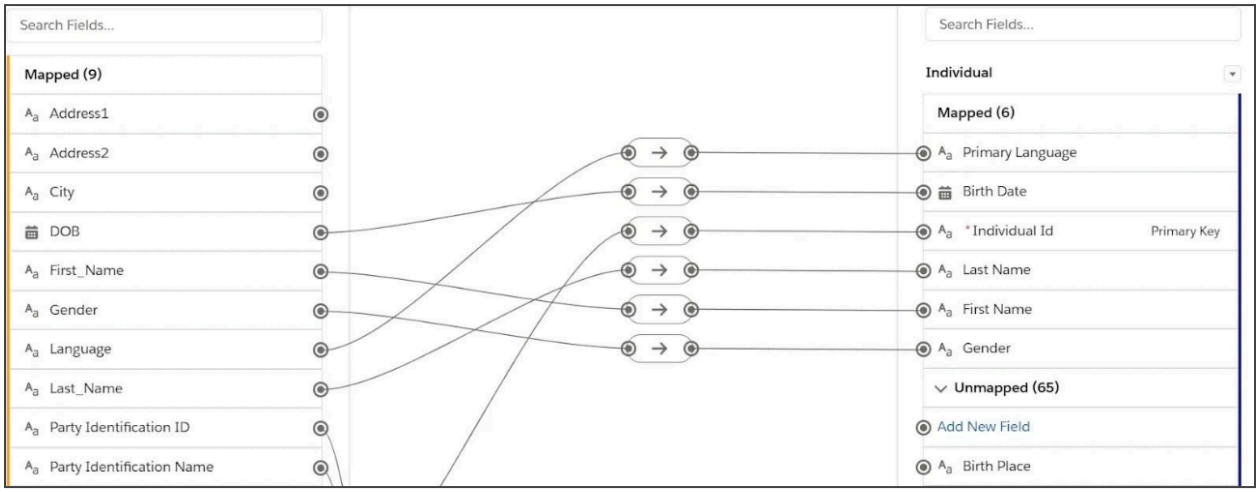
Now that we understand the power of Data Harmonization, let's explore how it can be harnessed & optimized in Salesforce Data Cloud.

**Harmonize Data with Data Cloud**

When harmonizing your data in Data Cloud, keep in mind the data models & concepts we previously discussed. This will help as you dive into the key steps to harmonizing your data.

**Map Data**

1. Navigate to a Data Stream > click the Details section > click  Review in Data Mapping.
2. Select objects and map necessary fields to Data Model Entities.
3. Create custom fields or objects if the standard fields do not meet requirements.
4. Extend the original source schema with formula fields to cleanse & transform data.
5. Save the schema as data lake objects (DLOs) for better data management.



# Step-by-step Guide Data Harmonization & Unification

**Extend Schema with Custom Fields**

1. Add a Formula Field
    a. Go to the Data Streams tab > select the data stream.
    b. Navigate to the data stream Details page >select **New Formula Field** > Enter the formula > click **Save.**
2. Add a Source Field
    a. Go to the Data Streams tab > select the data stream.
    b. Navigate to the data stream Details page > select **Add Source Fields**.
    c. Select **Add Fields Manually** or **Add Discovered Fields** >The connector displays the new fields > select the fields to import > click **Save**.
3. Save the schema as Data Lake Objects (DLOs).

Tip: If a DLO is associated with a data stream, you can't add fields to it from the Data Lake Objects tab. You'll need to make those updates on the Data Streams tab.

**Update a Schema (for a formula)**

1. Navigate to the field list on the data streams details page.
2. Find the specific formula field you wish to modify > select **Edit Formula** from the dropdown.
3. Update the formula syntax > click **Save.** (This overwrites the formula)

Tip: Updates to previously activated formulas are not retroactive; they will not be applied to existing fields. New fields will be updated after the next refresh of the data extension.

# Step-by-step Guide Data Harmonization & Unification

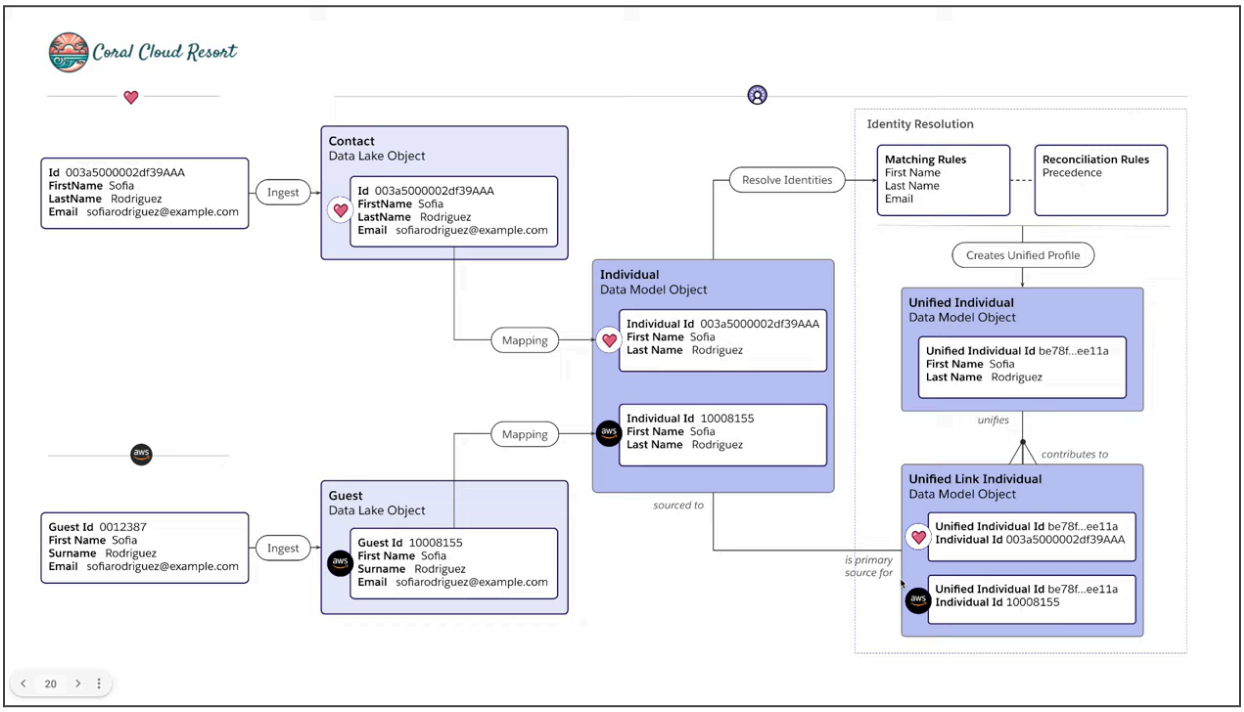
## Introduction to Data Unification

Now that we have our data harmonized, let's explore Data Unification in Data Cloud. The goal of ***data unification*** is to bring together data from disparate sources to create one unified profile. Keep in mind that you can not have data unification without first having data harmonization.

Data Unification is the larger process of taking data from disparate sources, bringing it together, and creating a unified profile. With Data Unification, we're cleansing & normalizing data while also creating unique identifiers for each entity. With Data Unification, our ultimate goal is to create one comprehensive unified view of our data, leading to better decision making & more informed analysis.

**Key Concepts of Data Unification**

- Data unification creates unique identifiers for each entity.
- Data unification aims to create a consistent data set




**Unifying Data**

For our Sunshine Trails Hospitality example, let's explore some of the potential ways that Data Unification supports our work. Unification allows us to configure rules for individual matching across sources of data, establish a preference for unified attributes & reconciliation rules.

**Steps to Unify Data**

Data Unification consists of Data Consolidation, Data Standardization, and Data Identification Rules, all of which lead to Data Integration. Let's explore these steps, below.

# Step-by-step Guide Data Harmonization & Unification

**Step 1: Data Consolidation**

**Create a Data Space**

1. Go to Data Cloud Setup > under Feature Management, and click **Data Spaces**.
2. Click New > give the data space a unique name.
3. Enter a unique data space prefix (after you save the data space, you can't change its prefix because it becomes part of the API name used to differentiate objects that exist in multiple data spaces)
4. Add an optional description about the purpose of the data space >click **Save**.

**Set up a Salesforce connection in Data Cloud**

1. From Data Cloud Setup, select Salesforce CRM under Salesforce Integrations > click **New**.
2. Click **Connect to a Salesforce org** *or* **Connect to a Sandbox org** > click **Next.**
3. Create an alias for your org (you can't change the alias later) > click **Save.**
4. Log in to the org using your Salesforce credentials > click **Allow** in the Allow Access menu.
5. Your org is now available from the Salesforce CRM Connectors page.

Tip: Setting up external integrations (connections) is a process that can be quickly and easily done in your Data Cloud Setup. To establish an External Integration, you'll need Administrative Access to the desired external system. From there, log in with your credentials and follow the steps to create a new connection.

**Create a Salesforce CRM Data Stream**

1. Navigate to the Data Streams tab > click **New** > select the Salesforce CRM data source > click **Next**.
2. To create your data stream, select a Salesforce org. If you have only one Salesforce org connected to Data Cloud, it's selected by default.
3. Select the All Objects tab > click **Next.**
4. Select the Salesforce Object that you want to include >click **Next**.
5. Review the fields to include in your data stream > click Next.
6. In the Header Label, deselect the fields not required for your data stream.
7. If needed, add these formula fields > click **Next**.
    a. Field Label—The display name for a data stream field.
    b. Field API Name—The programmatic reference for a data stream field.
    c. Formula Return Type—The data type corresponding to the new field. The options are number, text, and date.
8. Enter the deployment details.
    a. The data stream name defaults to the object label and the alias that's provided when setting up connection to a Salesforce CRM org. Connections that were created before the alias option was available use the object label and Salesforce org ID for the data stream name. You can change the name of a data stream.

   b. The refresh and frequency are set by default. For more information, see Data Stream Schedule in Data Cloud.
9. Click **Deploy**. Your Salesforce CRM data stream is now created. (To create more data streams, repeat these steps.)

**Set Up a Snowflake Data Federation Connection**

1. Navigate to Data Cloud Setup > under Configuration, select Snowflake > click **New**.
2. Enter a connection name, connection API name, and the authentication details. You can't use an encrypted private key. When you enter the key, don't include BEGIN PRIVATE KEY at the start of the key and END PRIVATE KEY at the end.
3. Click **Nex**t > select a database > click **Save**.

**Create a Snowflake Data Stream**

1. Navigate to the Data Streams tab > click **New**.
2. Select Snowflake as the source of the data stream > click **Next**.
3. Select from the available Snowflake connections > select the desired database > Select the objects that you want to include > click **Nex**t.
4. Select the category to specify the type of data to ingest > for Primary Key, select a unique field to identify a record.
5. *(Optional)* Select a record modified field. If data is received out of order, the record modified field provides a reference point to determine whether to update the record. The record with the most up-to-date timestamp is loaded.
6. *(Optional)* For Organization Unit Identifier, select a business unit to use in a record's data lineage. Data Cloud Implementation BYOL Data Federation
7. Select the source fields > click **Next**.
8. For Data Space, if the default data space isn't selected, assign the data stream to the appropriate data space.
9. If you want to query the data in your Snowflake account with reduced latency, keep the Enable acceleration option checked, and choose the acceleration schedule.
10. Click **Deploy**. You can now map your data lake object to the semantic data model to use the data in segments, calculated insights, and other use cases.

**Set Up an Amazon S3 Connection**

1. Navigate to Data Cloud Setup > in the quick find box, search for "Other Connectors."
2. Click **New** > Click **Amazon S3**.
3. Create a new Amazon S3 source > Enter a Connection Name and Connection API Name, and provide the Authentication and Connection details.
4. Enter the IAM Role Name > enter the External ID (copy & paste).
5. Enter the Connection details (Bucket Name & Parent Directory).
6. To review your connection, click **Test Connection** > click **Save.**

**Create an Amazon S3 Data Stream**

1. Navigate to the Data Streams tab > click **New.**
2. Under Other Sources, select the Amazon S3 connection source > click **Next**.
3. Select from the available Amazon S3 connections > click **Next**.

4. Configure the file & Source Details
    a. Select the File Type (CSV or Parquet) > provide the Import from Directory path.
    b. *Optionally,* specify a File Name or use wildcards to retrieve files.
5. Create a Data Lake Object (DLO) or use an existing one > click **Next**.
6. Map fields & configure the Data Model > click **Deploy**.

**Set Up an Amazon Kinesis Connection**

1. Navigate to Data Cloud Setup > under Configuration, select **More Connectors** > click **New**.
2. On the Source tab, select **Amazon Kinesis** > click **Next**.
3. Enter a connection name and connection API name, and provide the authentication and connection details.
    a. Enter the AWS access key ID and secret access key to authenticate. You must use the ID and secret access key together.
    b. Enter the name of your Kinesis stream for the AWS account that you want Data Cloud to read data from.
    c. Enter the region where your AWS Kinesis exists.
    d. Enter the Amazon Kinesis data stream service endpoint to connect programmatically to an AWS service.
4. To review your configuration, click Test Connection > click **Save**.

**Create an Amazon Kinesis Data Stream**

1. Navigate to the Data Streams tab > click **New.**
2. Under Other Sources, select the Amazon Kinesis connection source > click **Next**.
3. Select from the available Amazon Kinesis connections.
4. Select the object that you want to import > click **Next**.
5. Under Object Details, for Category, identify the type of data in the data stream.
6. For Primary Key, select a unique field to identify a record. If a primary key isn't listed in the dropdown, you must create one using a formula field.
7. *(Optional)* Select a record modified field. If data is received out of order, the record modified field provides a reference point to determine whether to update the record. The record with the most up-to-date timestamp is loaded.
8. *(Optional)* For Organization Unit Identifier, select a business unit to use in a record's data lineage.
9. Click **Next.**
10. For Data Space, if the default data space isn't selected, assign the data stream to the appropriate data space.
11. Click **Deploy**.

Tip: When the Last Run Status is successful, you can see how many records were processed and the total number of records that were loaded. You can now map your data lake object to the semantic data model to use the data in segments, calculated insights, and other use cases.

**Create a Batch Data Transform**

1. Navigate to the Data Transforms tab > click **New** select **Batch Data Transform** > click **Next**.
2. Select the data source. If you selected data model objects as the source, select a data space from the dropdown menu.
3. Select the initial input data for the batch data transform.
    a. Click Add Input Data > select the objects that you want to add.
    b. Under Selected Columns, choose which of the object's columns to include in your input data, or select Name to select all columns.
        i. If you added multiple source objects, select each object one at a time and then choose its columns. Selected Columns shows only the columns for the selected object.
    c. Click **Next**.
4. Click the Add Node button to add a node to the batch data transform > select the node type.
5. To connect, join, or append nodes in the canvas, drag the Add Node button (period icon) next to a node or Branch button (branch icon) on top of another node in the canvas. Select one of the Connect Nodes options.
6. End the data transform with one or more output nodes.
        i. • If your data source is a data lake object, select the object to write the results to, and then click **Apply**.
        ii. • If your data source is a data model object, select the object to write the results to or create a new object.

**Step 2: Data Standardization**

**Map Data Model Objects**

1. Navigate to the Data Streams tab > click the data stream you want to map > click **Start Data Mapping**.
2. Click **Select Objects** > click **New Custom Object**.
    a. Review Object Label, Object API Name, and Fields. You can't change the field properties after you save the object.
    b. Save your changes.
3. Select the field to map in the DLO and click the related field in the DMO to establish a mapping connection. A confirmation appears with an arrow connecting the fields.
4. If you don't see a field in the DMO that represents the field in your source data, you can add a field to the DMO.
    a. Click the arrow next to unmapped fields > click **Add New Field.**
    b. Enter a field label or display name > choose the field data type > Save.

**Step 3: Data Identification Rules**

**Configure Identity Match Rules**

To specify how identity resolution should identify matching records, you'll need to configure identity resolution match rules.

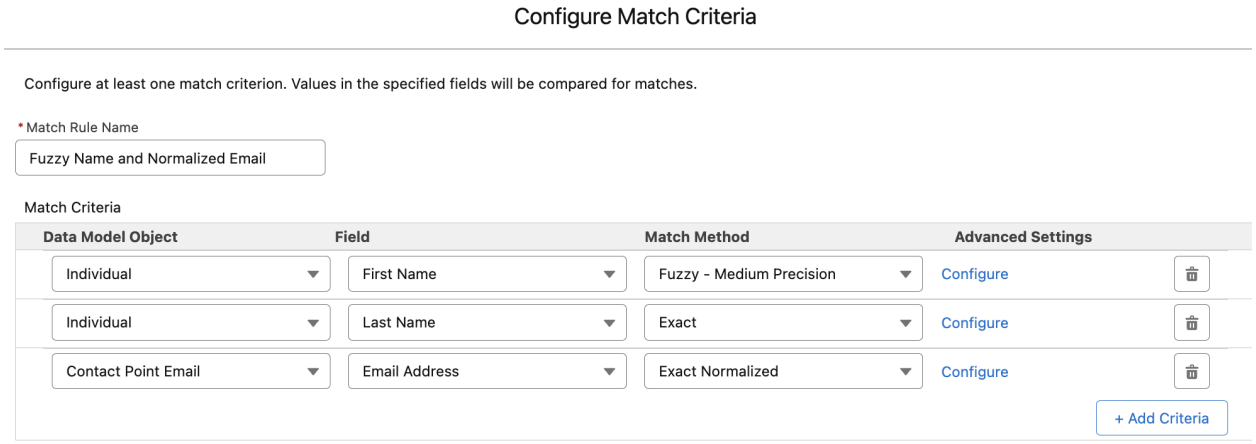
1. Navigate to the record homepage of your identity resolution ruleset.
2. In Match Rules, click **Configure** or **Edit** > click **Add Match Rule**. (Select a default match rule or click **Custom Rule** to create your own.)
3. Click **Next** > Edit, add, or delete criteria for your match rule.
4. Name your match rule (80 characters or less) > click **Save**.

Tip: Each ruleset must have at least one match rule. You can't delete the last match rule from a ruleset. To stop a ruleset from running, delete it completely.

**Create Identity Resolution Rulesets**

Identity Resolution creates your unified customer profile in Data Cloud. It can be broken down into two parts: matching and reconciliation. Matching is the process of grouping profiles together through commonly shared criteria. Reconciliation is the process of summarizing key attributes that have been unified.

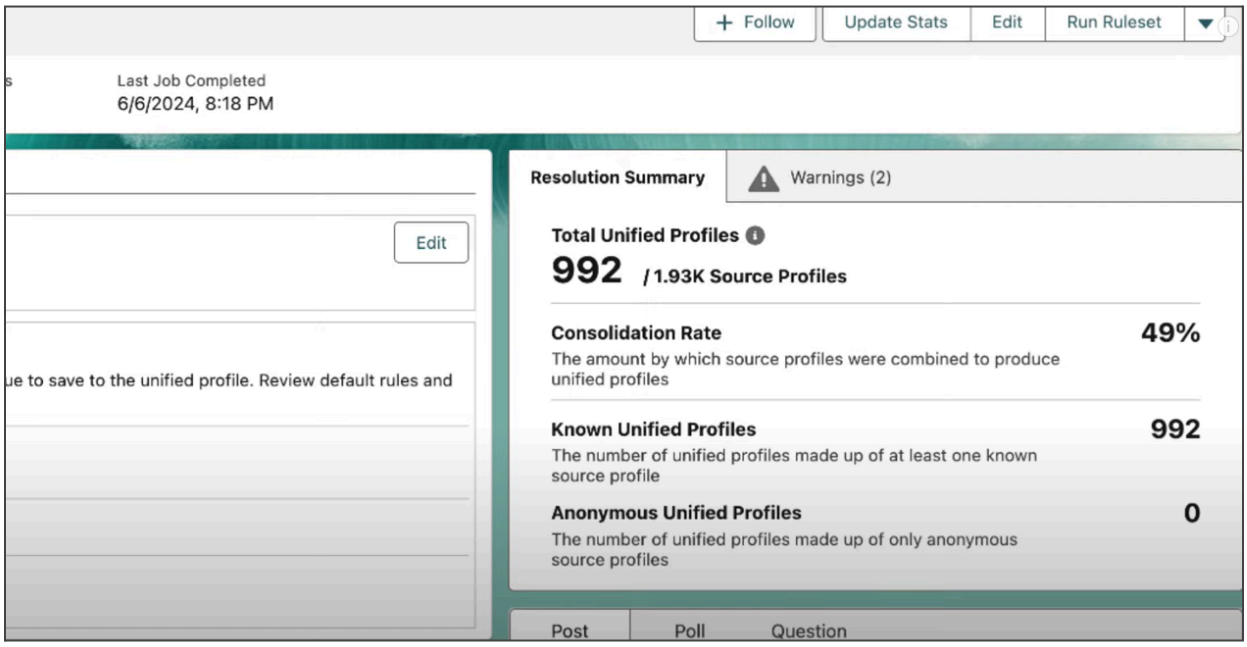
1. Navigate to the Identity Resolutions tab > click **New** > select a data space.
2. Select the primary data model object that your ruleset will be based on. If you don't see any objects available, make sure your data is modeled properly and that you haven't reached the maximum number of rulesets.
3. Add an optional ruleset ID of up to four text characters. Optionally, add a ruleset ID of up to four text characters. The ruleset ID can't be changed and becomes part of the data model object names and API names. If you have two rulesets for an object, at least one must have a ruleset ID.




Tip: Use test as a ruleset ID. After using the test ruleset to test different combinations of match and reconciliation rules, delete the ruleset. If you delete a ruleset and want to use the same ruleset ID again later, you could encounter an

error. Wait for Data Cloud to finish deleting the ruleset with your desired ruleset ID and try again.

4. Click **Next** > name your ruleset > add an optional description.
5. Make a note of the data model object and API names created by the ruleset > click **Save**.




**Set a Default Reconciliation Rule**

1. Navigate to the record home page of your identity resolution ruleset > select an object.
2. Click the pencil icon > select a reconciliation rule from the dropdown menu.
3. *Optionally,* select **Ignore Empty Values**. (Identity resolution ignores the Ignore Empty Values selection for key fields.)
4. If you're using Source Priority to select a value, order the data sources so that the highest priority source is at the top of the list. The value selected for the unified field is the value that appears in the highest priority data source.
5. Click **Save**.

**Best Practices**

| |
|---|
| • Remember to check data fidelity. |
| • Confirm that the data being used for Identity Resolution is accurate & valid. |
| • Keep in mind that Rulesets are set to run automatically. |
| • Remember to reduce the rows being processed whenever possible. |
| • Consider using formula fields in the Data Stream whenever possible. |

**Conclusion**

# Step-by-step Guide Data Harmonization & Unification

Data Harmonization & Unification help to unlock previously siloed data, unlocking a 360 customer view in Data Cloud. By transforming data into a consistent format, we can create a single source of truth for our data, standardizing elements and resolving discrepancies. The steps outlined in this guide will help you to quickly & easily access the power of Data Cloud by maximizing your use of these tools.

**Resources**

Salesforce Help: Harmonize Your Data
Salesforce Help: About Data Cloud
Salesforce Help: Data Cloud & Cross-Cloud Implementation Guides
Salesforce Help: Configure Identity Resolution Match Rules
Salesforce Help: Create Identity Resolution Rulesets
Trailhead: Customer 360 Data Model in Data Cloud
Trailhead: Get Hands on with Data Cloud
Trailhead: Ingestion & Modeling in Data Cloud