

14 May, 2024

# Managing Data Quality in MDM SaaS

- Anushruti Huilgol, Senior Tech Support Engineer, GCS
- Ramprasad Venugopal, Senior Tech Support Engineer, GCS

Where data & AI come to 

# Housekeeping Tips



- Today's Webinar is scheduled for **1 hour**
- The session will include a webcast and then your questions will be answered live at the end of the presentation
- All dial-in participants will be muted to enable the speakers to present without interruption
- Questions can be submitted to "All Panelists" via the **Q&A option** and we will respond at the end of the presentation
- The webinar is **being recorded** and will be available on our [Success Portal](#) - where you can download the **slide deck** for the presentation. The link to the recording will be emailed as well.
- Please take time to complete the **post-webinar survey** and provide your feedback and suggestions for upcoming topics.

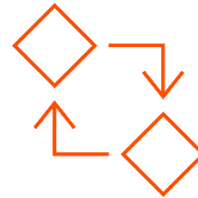
# Feature Rich Success Portal



**Bootstrap trial and  
POC Customers**



**Enriched Customer  
Onboarding  
experience**



**Product  
Learning Paths  
and Weekly  
Expert Sessions**



**Informatica  
Concierge**



**Tailored training  
and content  
recommendations**

# More Information



## Success Portal

<https://success.informatica.com>



## Communities & Support

<https://network.informatica.com>



## Documentation

<https://docs.informatica.com>



## University

<https://www.informatica.com/in/services-and-training/informatica-university.html>

# Safe Harbor

The information being provided today is for informational purposes only. The development, release, and timing of any Informatica product or functionality described today remain at the sole discretion of Informatica and should not be relied upon in making a purchasing decision.

Statements made today are based on currently available information, which is subject to change. Such statements should not be relied upon as a representation, warranty or commitment to deliver specific products or functionality in the future.

Informatica<sup>®</sup>  
**Managing Data Quality in MDM  
SaaS**



**Anushruti Huilgol**

Senior Technical Support Engineer



**Ramprasadh V**

Senior Technical Support Engineer

**Where data  
& AI come to** 

# Agenda

1 Data Quality in MDM SaaS

2 Data Validation

3 Duplicate Record Prevention

4 Data Standardization

5 Error Correction and Handling

6 Best Practices

# Introduction

- Data quality is a crucial aspect of MDM. Poor data quality can lead to incorrect insights, inefficient operations, and poor decision-making.
- To achieve MDM success, organizations must prioritize data quality management.
- Maintaining data quality poses several challenges :
  - Data inconsistency across sources
  - Errors introduced during data entry or processing.
  - Duplication of records
  - Lack of standardization in data formats



# Assess Data Quality Beyond MDM

Data Profiling is a form of statistical analysis used to inspect data sets and determine their inherent quality.

- Count of Null Values
- Max and Min Values
- Data Types and Formats
- Identify Duplicates & missing values
- Identify the contradicting data.

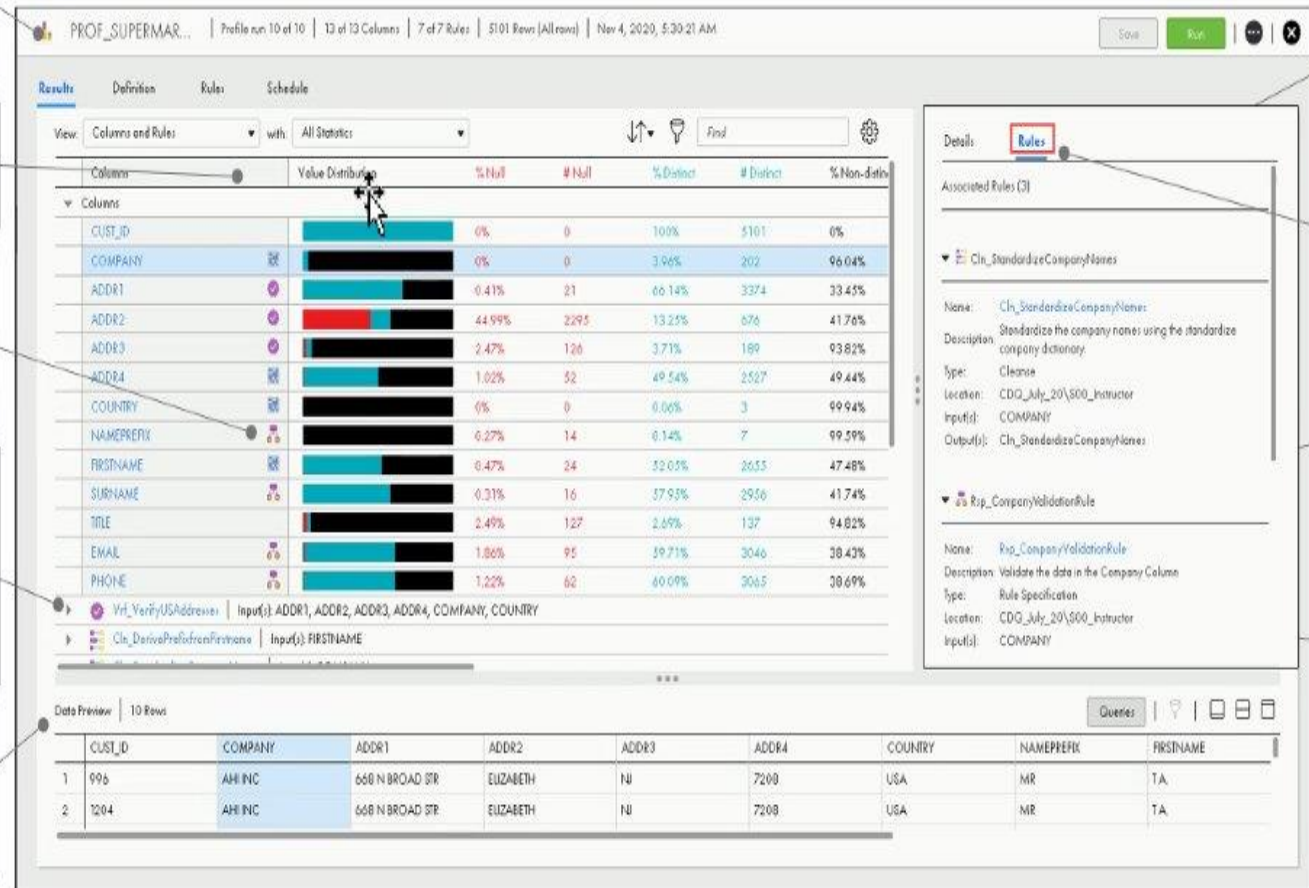
Info about the profiling run

Statistics

Icons indicate a DQ Asset is applied to the column

Rule Spec, Verifier, Cleanse and Parse results

Data Preview & Live Drill Down



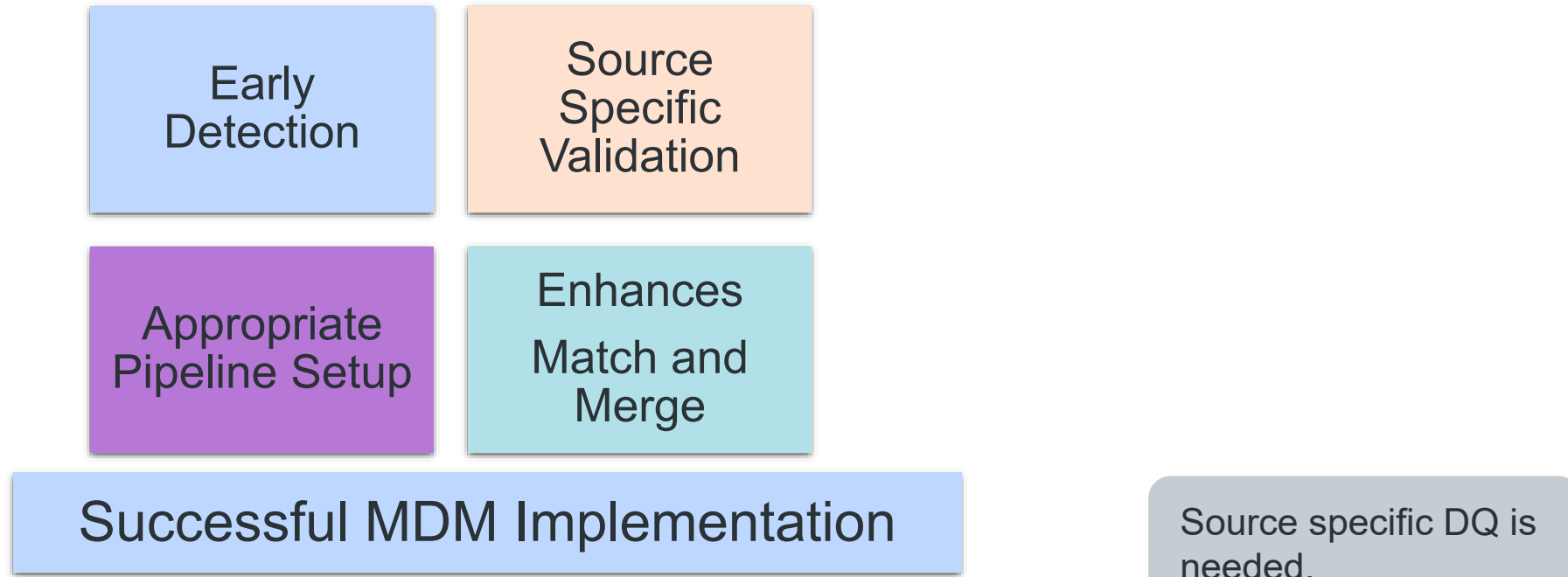
Trends

Rules Tab shows the associated rule details

Documented and Inferred Data Types

Inferred Patterns

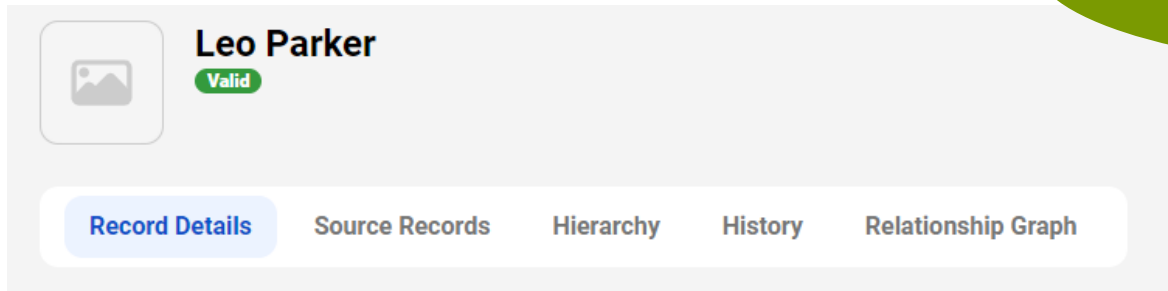
# Benefits



Source System	Purchase Quantity	Price Per Unit	Discount	Total Price
SFA	100	USD 10	2	980
CRM	200	INR 1500	3	300000

Different Sources  
Different Units

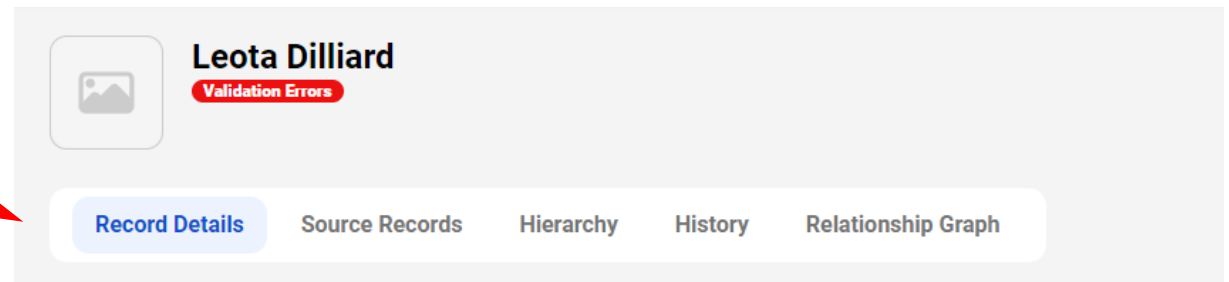
# Record States



Record card for Leo Parker. The name is bolded. Below the name is a green pill with the word "Valid". To the left of the name is a placeholder for a profile picture. Below the name and status is a horizontal menu with five items: "Record Details" (highlighted in blue), "Source Records", "Hierarchy", "History", and "Relationship Graph".

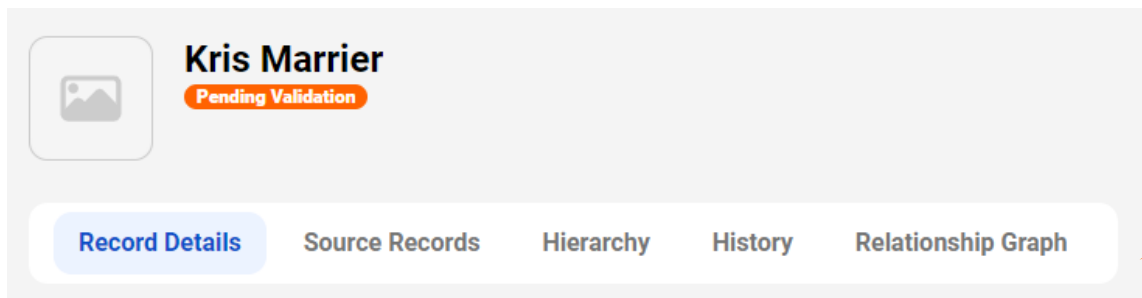
Passed all the DQ rules configured. Valid Record

Failed at least one of the DQ rule. Invalid Record



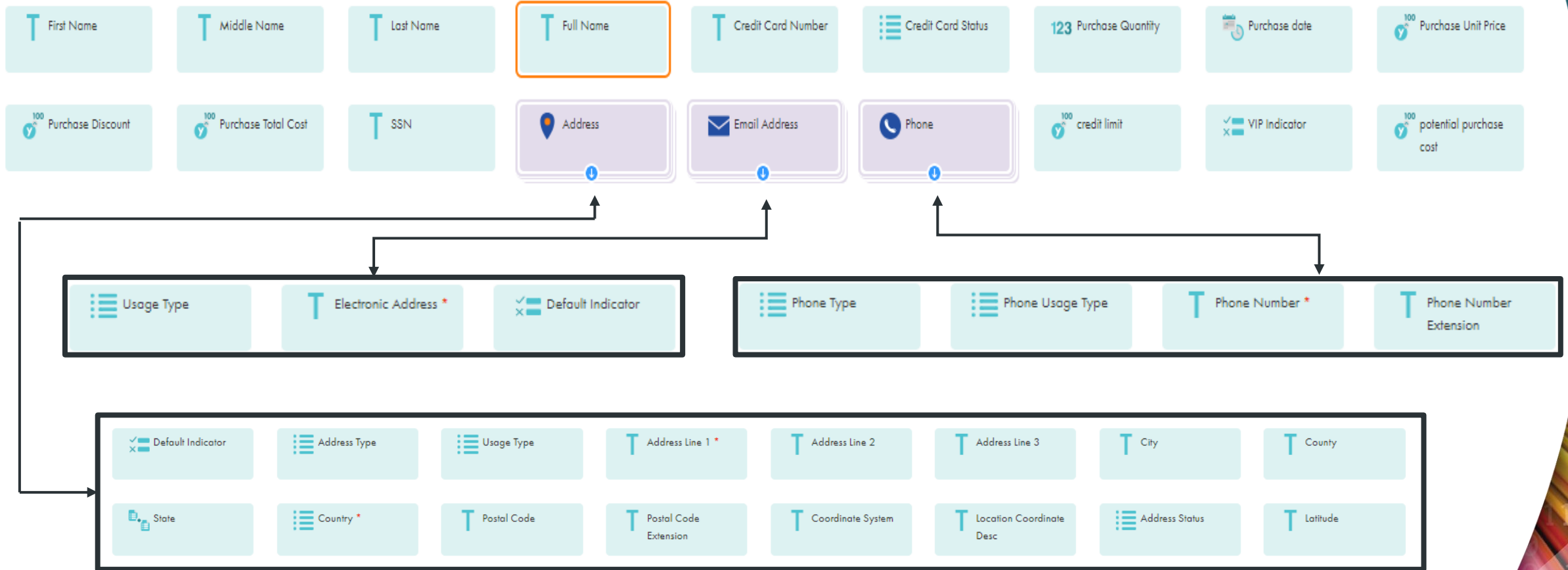
Record card for Leota Dilliard. The name is bolded. Below the name is a red pill with the words "Validation Errors". To the left of the name is a placeholder for a profile picture. Below the name and status is a horizontal menu with five items: "Record Details" (highlighted in blue), "Source Records", "Hierarchy", "History", and "Relationship Graph".

Validation is pending on this record.



Record card for Kris Marrier. The name is bolded. Below the name is an orange pill with the words "Pending Validation". To the left of the name is a placeholder for a profile picture. Below the name and status is a horizontal menu with five items: "Record Details" (highlighted in blue), "Source Records", "Hierarchy", "History", and "Relationship Graph".

# Data Model





# Data Quality **Validation**

# Basic Rules

Basic DQ rules can be used to check length and alphanumeric content validation on First and Last names.

First Name	Last Name
J	Ken
Kian	R0n

Invalid length

Alphanumeric characters

# Configuration

- Configured using the pre-defined functions at the attribute level.
- These are light weight DQ rules.
- Validation severity can be configured to **Information, Warning, Error** and specific custom error messages can also be defined.

X Person | Data Quality > MinLength

**General**

Enable rule association

Applied Field: First Name

Name\* MinLength

Internal ID MinLength\_1715088575198

Description

---

**Condition**      **Validation**

Configure a condition to transform or validate the field values.

Function\* Minimum Length

Description Validates whether the length of a string is greater than or equal to the specified length.

Minimum Length\* Enter a Value 3

**General**

Enable rule association

Applied Field: Last Name

Name\* Last Name Characters Check

Internal ID Last\_Name\_Characters\_Check\_1715088696134

Description

---

**Condition**      **Validation**

Configure a condition to transform or validate the field values.

Function\* Regular Match

Description Validates whether a string matches with the specified regular expression pattern.

Input\* Enter a Value ^[A-Za-z\.\s]+\$

# Advanced Rules

Person	Credit Limit	Potential Purchase Cost	VIP Indicator
Person A	10000	2000	<input checked="" type="checkbox"/>
Person B	5000	1000	<input checked="" type="checkbox"/>

An Advanced DQ rule can be invoked when a data steward tries to enable the Vip indicator for a person.

$$(0.7*(\text{credit\_limit}/10000)*100)+(0.3*(\text{PotentialPurchaseLimit}/2000)*100)$$

Properties: PrimaryRuleSet ⓘ

General

- or if VipIndicator is "" then 'ok'

Inputs

- or if VipIndicator is not NULL then (0.7\*(credit\_limit/10000)\*100)+...

[Rule Logic](#)

- if Result of # 1 is less than 80 then 'not vip'

# External Validation

- Consider the following requirement where we need to verify the validity of a credit card number based on Luhn Algorithm or also called as Modulus 10 Algorithm.
- These cannot be achieved using basic rules with pre-defined functions.

First Name	Last Name	Credit Card Number
John	Ken	223444565578
Kian	Ron	223477654459

# Configuration

- Validation through Application Integration process.
- Configured in the basic rule with function name as external validation.
- This process should be designed on Application Integration service.
- Error severity can be set on this external validation basic rule based on the output of the process.

**Condition**      **Validation**

---

Configure a condition to transform or validate the field values.

**Function\***     

**Description**      Validates the field data based on external validation process.

**Process URL\***           

**Condition**      **Validation**

---

Configure the message to display and the percentage to downgrade the trust score when the validation fails.

**Error Severity ?**     

**Error Message\***     

**Downgrade Trust Score On ?**      Credit Card Number

**Downgrade Trust Score (%) ?**

# DaaS Validation

## Email and Phone

Sending out information on inaccurate contact information can lead to loss of information and other safety concerns.

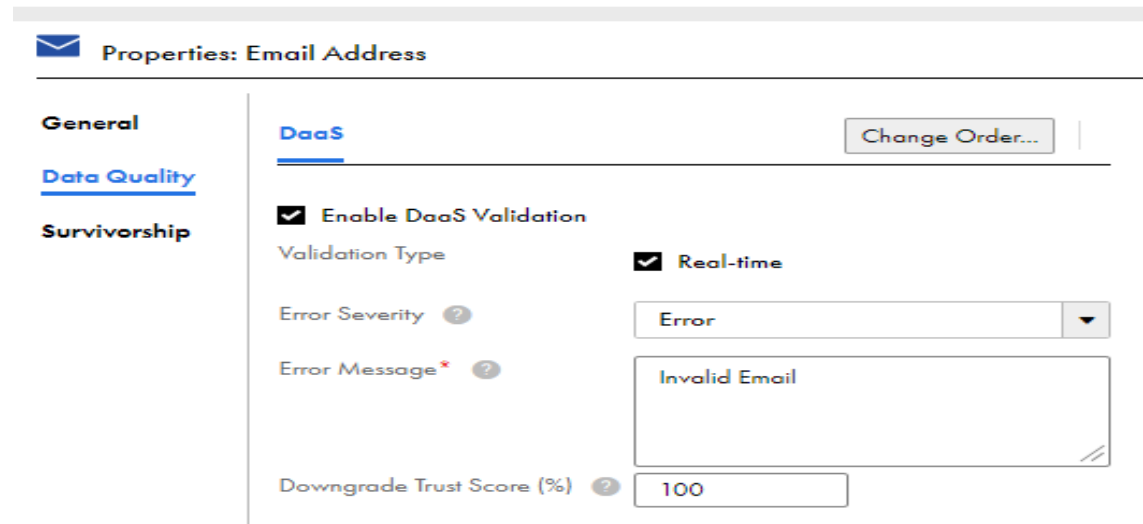
Electronic mail	Phone Number
jken@demotest.com	571-222-88

Incomplete Number

Unsafe Domain

# Configuration

- Real-time Validation.
- Format check for email and phone.
- Domain check for email
- License keys for email and phone verification needs to be set in the global settings in B360.
- We can configure the validation **only for phone and email smart fields**.



Properties: Email Address

**General**

**Data Quality**

**Survivorship**

**DaaS** Change Order...

Enable DaaS Validation

Validation Type  Real-time

Error Severity ?

Error Message\* ?

Downgrade Trust Score (%) ?

# Validation

- After saving the record, we can see the errors on the UI
- On the source records tab, we can see the error returned by the enricher service

Email Address (1)  
 ⚠ Some of the values in an entry are not valid. Enter valid values

Personal  ✎ 🗑

jken@demotest.com

Phone (1)  
 ⚠ Some of the values in an entry are not valid. Enter valid values

Personal

571-222-88

NDC Indicator	:	
Verification St...	:	401
Verification M...	:	Invalid number ... Invalid number - Number is too short by ...
Effective Start ...	:	
Effective End ...	:	
Status	:	

Electronic Add...	:	jken@demotes... jken@demotest.com
Default Indicat...	:	
Hygiene Status	:	Parked Parked
Electronic Add...	:	
Verification St...	:	260 260
Verification M...	:	Domain Confir... Domain Confirmed, Potentially Dangerous

# Duplicate Record Prevention

Where data & AI come to **LIFE**

# UI – Similar Records

- When entering the data in the UI, we would want to avoid adding records that are similar to already existing records.
- While entering record details, Customer 360 displays the matching/similar records as Recommendations.
- We use advanced search to find the similar records in MDM SaaS.
- Click the similar records tab to view. If you find a similar record, you can select it and continue working on it. The record that you were creating is lost.

# Configuration

- Configured through the business application under the search tab.
- Default is Enabled

Customer 360

Model Pages Search

Settings **Similar Records**

Suggestions:  Enabled  Disabled

Search History:  Enabled  Disabled

Similar Records:  Enabled  Disabled

Minimum Characters to Trigg...: 5

Maximum Number of Sugge...: 10

Number of Previous Searches: 5

Results Per Page: 25

Customer 360

New visual experience  material Search ram\_gcs\_org

New Person

General Information

First Name:\* person Middle Name: Last Name:\* Full Name

Prefix Name: Suffix Name: Title: Designation

Birthplace: Birth Date: MM/dd/yyyy Gender: Image:

Marital Status: organization:

CLAIRE™ Similar Records (4)

Some of the existing records are similar to the record that you are trying to create. To use a similar record, select one of the following records:

- ▶ person, 2
- ▶ person, 3
- ▶ person, 1
- ▶ person, 4

Similar records

# API –Search Match

- One more option to check for duplicate records before insert is the search match API.
- We can use Search Match API to search for matching records based on the published match model configuration. The required input parameters for the REST API vary based on the business entity and its match model.
- If the record is ingressed with population name, the search match API should also be fired with the same population name to retrieve the results.

```
POST https://emw1-mdm.dm-em.informaticacloud.com/match-ng-pair-gen/public/api/v1/searchmatch/c360.person?matchmodel=9...

Params Authorization Headers (9) Body Pre-request Script Tests Settings
none form-data x-www-form-urlencoded raw binary GraphQL JSON

1 {
2   "searchControls": {
3     "maxRecordsToReturn": 5,
4     "searchLevel": "Typical",
5     "fileRecordLimit": 1000,
6     "population": "india",
7     "matchPreference": "FIRST_RULE"
8   },
9   "data": {
10    "searchRecord": {
11      "fullName": "Anoop AGGARWAL",
```

```
Body Cookies Headers (19) Test Results
Pretty Raw Preview Visualize JSON

1 {
2   "searchResult": {
3     "totalRecords": 2,
4     "records": [
5       {
6         "_meta": {
7           "score": 100.0,
8           "matchRule": "THD1",
```



# Data Quality Standardization

# Basic Standardization

First Name	Last Name	Full Name
John	Kennedy	John Chris
Mia	Ron	Mia K Ronald

Mismatch or discrepancy

# Configuration

- We can leverage the pre-defined functions provided in the basic rules to build the use-case.
- This allows to have standard full name format ( First Last) across all sources.

The screenshot shows the configuration page for a rule named 'CONCAT\_WITH\_SPACES' under the 'Person | Data Quality' category. The page is divided into two main sections: 'General' and 'Condition'.

**General Section:**

- Enable rule association**
- Applied Field: Full Name
- Name\*: CONCAT\_WITH\_SPACES
- Internal ID: c360.person.LWConfig.rule.fullNameConcatenation
- Description: Concatenate first name , middle name and last name to form full name

**Condition Section:**

Configure a condition to transform or validate the field values.

Function\*: Concatenate With Spaces

Description: Concatenates two or more strings with spaces in between.

Input 1\*: Select a Field (First Name)

Input 2\*: Select a Field (Middle Name)

Input 3\*: Select a Field (Last Name)

# Rule Execution

## General Information

First Name:

Middle Name:

Last Name:

Full Name:



INPUT

## General Information

First Name:

John

Middle Name:

Last Name:

Kennedy

Full Name:

John Kennedy



OUTPUT

# Advanced Data Standardization

- The total price is calculated as  $TP = \text{Purchase Quantity} * \text{PPU} - \text{Discount}$ .
- How can we incorporate mathematical calculations to avoid such errors?

Purchase Quantity	Price Per Unit	Discount	Total Price
100	USD 10	2	980
200	USD 15	3	8967

# Configuration

- Use the advanced DQ rules to define logic that cannot be achieved through basic rules
- We can configure the input and output fields based on rule requirements and set the error severity.

Data Quality ▾

### Purchase Total Cost v2 Advanced

Definition **Configuration**

#### Rule Editor

```
1 IF Quantity=null AND discountRate >=0 AND discountRate <=1
2 THEN TotalCost=Quantity * unitPrice*(1- discountRate)
3 ELSE TotalCost=Quantity*unitPrice
4 ENDIF
```

X Person | Data Quality > Total cost calculation

#### General

Enable rule association  
 Run the rule association even if all input fields are empty

Applied Field: Purchase Total Cost  
Rule Specification: Purchase Total Cost v2

Name: Total cost calculation

Internal ID: Totalcostcalculation\_1714707211064  
Description:

#### Input and Output Fields

Validation

Map the rule specification fields to the business entity fields, or configure values for the business entity fields.

Rule Specification Field	Data Type	Business Entity Field
Quantity	Integer	Purchase Quantity
unitPrice	Float	Purchase Unit Price
discountRate	Float	Purchase Discount

Rule Specification Field	Data Type	Business Entity Field
TotalCost	Float	Purchase Total Cost

# Address Validation and Enrichment

Address Line 1	City	County	State	Country	Postal Code
2369 LONGTOWN RD	Ridgeway		SC	US	

Incomplete Address

- The postal address information received from source could have problems with accuracy, completeness, non-standard formatting, Ambiguity etc.
- Having incomplete or inaccurate address is a potential loss of data and can be costly for the business.
- How can we standardize this data?

# Configuration

- Address verifier and enricher can be enabled of both batch and real-time processing.
- Validates the address against the reference address databases and enriches.
- Setup licenses in global settings in B360.
- We can configure this for **Address smart field as well as any custom field group**.
- Note : When you enable DaaS validation, you can't add basic and advanced rule associations to the fields in the field group.

Properties: Address

General

Data Quality

Survivorship

DaaS Batch    DaaS Real-time    Change Order...

Enable DaaS Validation

Validation Type     Batch

[Configure Address Verification](#)

# Using guides in MDM Business application UI

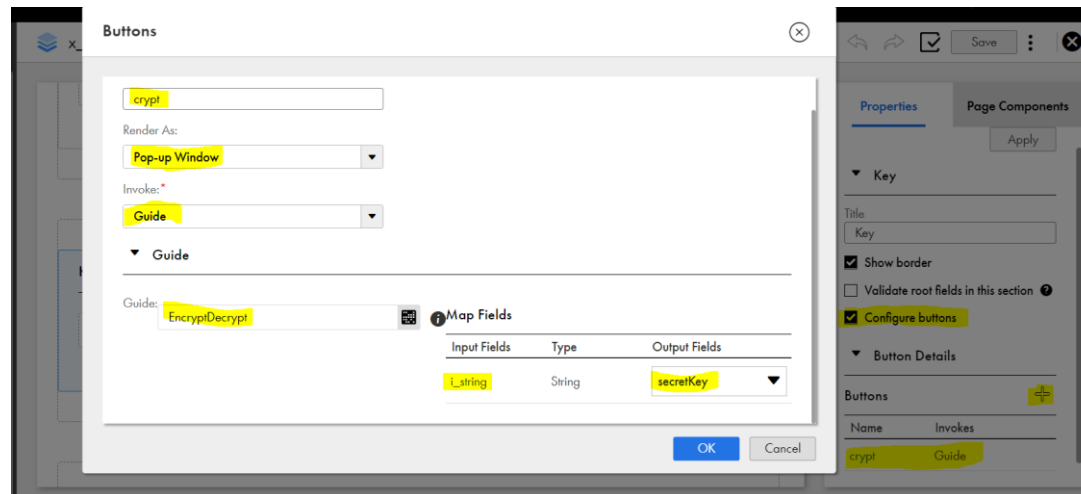
- For the sensitive data coming such as that of SSN , we would want to encrypt the data before it enters the MDM system.
- Assume we have a few secure users who are allowed to view this secure PII information.
- For these specific set of users, we should be able to decrypt the data on the UI.

Sensitive PII

First Name	Last Name	Full Name	Credit Card Number	Social Security Number
John	Kennedy	John K	223456789087	234-45-6678
Mia	Ron	Mia K Ron	542376899987	546-77-9856

# Using guides in MDM Business application UI

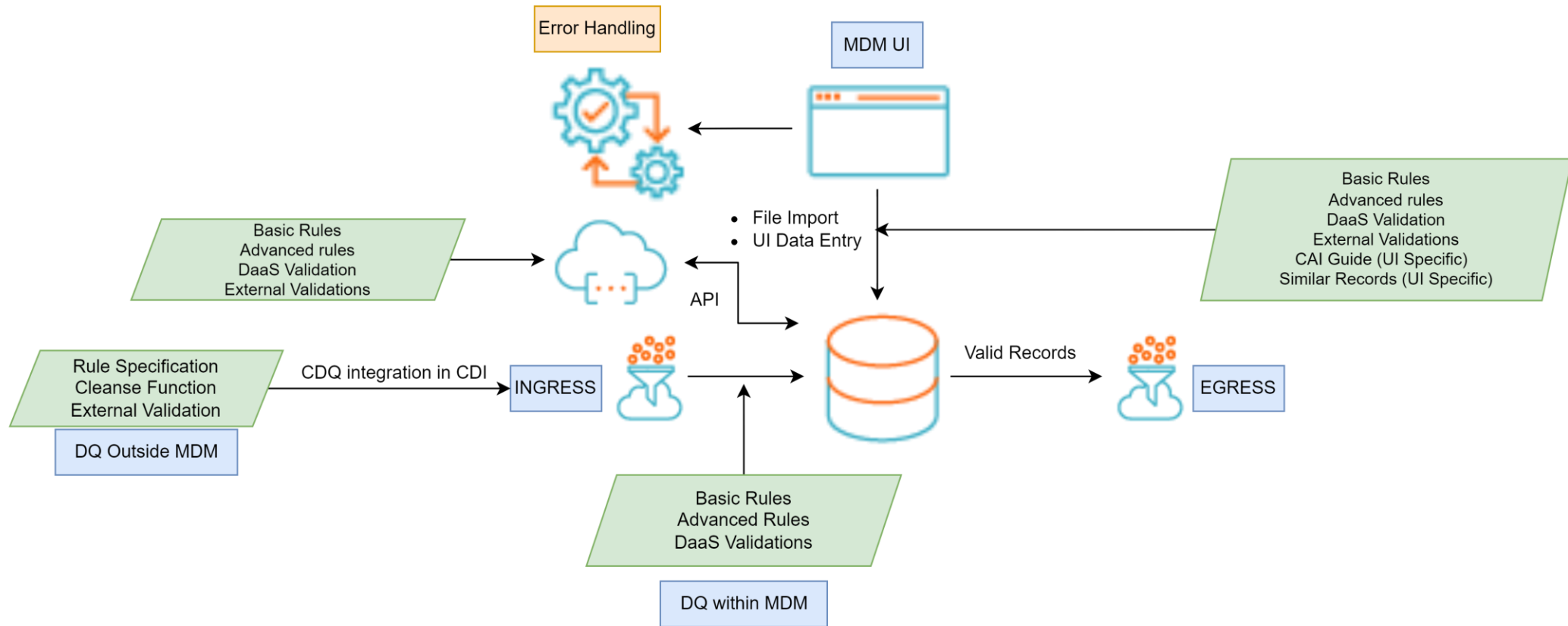
- The guide can be used call any deployed CAI process.
- These guides can be used in scenarios where we need to transform/validate data without the need of saving the values.
- We can add buttons on the create/edit page layouts which, on click, can call a guide.
- Example scenarios where we can use guide in MDM business application UI:
  - Search for a record to check if it exists before creating a new one.
  - Decrypting the encrypted data for a specific set of users.





# Data Quality Recap

# Data Quality in MDM SaaS



# Data Quality

## Error Handling and Correction



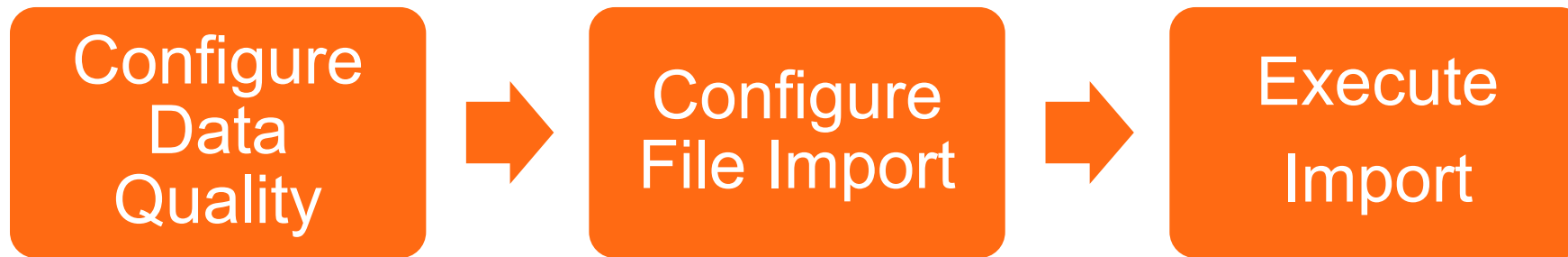
# Data Quality Lifecycle



During ingress/ file import, The records that **fail the data quality** are imported into MDM with a **validation errors status**.

- Identify these records
- Take corrective actions

# Use -Case



**How can we verify the status of the records imported?**

# Job Metrics

- We can verify the number of valid records during a file import using the job metrics.
- The **report provides a detailed overview on data quality failures** for each record.
- This report can be used for further analysis to fix the validation errors and re-ingress the corrected records

Step Metrics Copy JSON

**500** Total Records    **0** Error Records    **500** Success Records

**View by Asset**  
Asset: Person

Starting Records for Asset	500	Applied Data Quality Rules	1000
Error Records	0	Failed Data Quality Rules	803
Successful Records	500	Successful Data Quality Rules	197

**Ingress**

Job Instance: **Transform** | Load | Index for Search | Generate Tokens

**Transform**

Overview | Runtime Parameters | Metrics | **Errors**

Name	Rejected Records	Imported Records with Validation Errors
Person	0	803 <a href="#">Download</a>
Person.Root	0	319 <a href="#">Download</a>
Person.Address	0	484 <a href="#">Download</a>

# Search View

- The search tab offers facets which help us find the records with errors.
- We can filter records on basis of validation status, error severity and error message.
- Validation status, error severity and error type are system fields and searchable by default.
- This search is for master records that failed the validation only.
- We can select multiple records to perform bulk edits or sequential edits
- We can select and export erroneous records for further analysis.
- Note : Existing records (pre-error handling) will not be affected.

The screenshot displays the Informatica Customer 360 Search View interface. The top navigation bar includes the Informatica logo, 'Customer 360', a toggle for 'New visual experience', a search bar with 'Person' entered, and user information 'Org - Anushruti'. The left sidebar contains navigation options: New, Home, Search (selected), File Import, Workflow Inbox, Hierarchies, My Jobs, and Reports. The main content area is titled 'Search / Search For \*' and shows a search for 'Person'. The 'Filters' section on the left includes: 'Validation Status' (All: 500, Validation Errors: 494, Valid: 6), 'Error Type' (All: 801, Error: 484, Warning: 317), and 'Errors' (All: 801, Enter A Valid Value: 482, Person Postal Address DaaS Rule: 2, The Minimum Length For Name Is 7: 317). The search results show 'Search Results: \* (Showing 25 of 494) 3 Selected'. The results list includes: 'Eladia Saulter' (State: New Jersey, City: Ramsey, United States), 'Merlyn Lawler' (State: New Jersey, City: Jersey City, United States), and 'TySmith' (State: New Jersey, City: Hackensack, United States). A context menu is open over the selected records, showing options: Bulk Edit, Edit Selected, Export, Merge Records, Select All, Select Loaded, and Clear All. The bottom left of the interface has a 'Collapse' button.

# Record View

- The record view offers a more detailed overview on the errors.
- Every record **will have a flag which indicates the status**.
- Quick links **provide a detailed insight on fields with errors, warnings, information and other**.
- The address enrichment is denoted by the purple flag indicating address was corrected through the address validation.
- Based on the error messages, corrective measures can be taken to set the record to valid state.

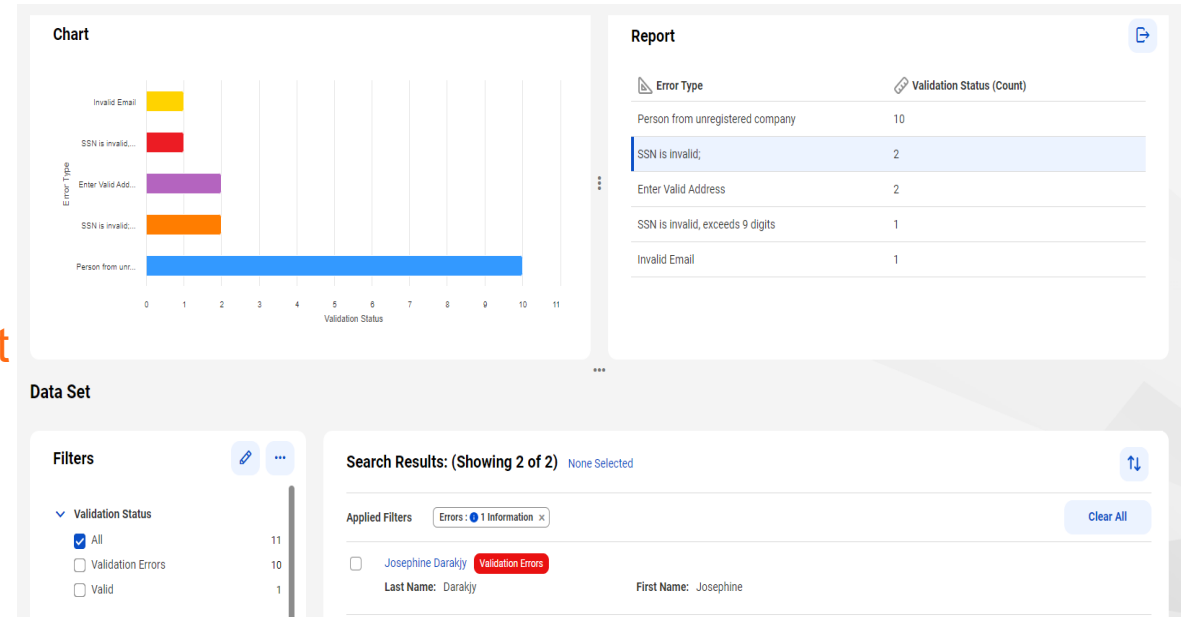
The screenshot shows the record view for 'Eladia Saulter'. At the top, there is a 'Validation Errors' flag. Below it, there are tabs for 'Record Details', 'Source Records', 'Hierarchy', 'History', and 'Relationship Graph'. The 'Record Details' tab is active, showing a 'General Information' section with fields for First Name, Middle Name, Last Name, Prefix Name, Suffix Name, Title, Birthplace, Birth Date, Gender, and Marital Status. A red 'Errors (6)' popup window is open, listing errors for 'Address 1' with messages like 'Enter a valid value.' for Address Line 1, 2, 3, City, and County. At the bottom, there is an 'Address (1)' section with a red error message: 'Some of the values in an entry are not valid. Enter valid values.'

The screenshot shows the record view for 'Lili Paskin'. It features a 'Valid' flag in a green box. Below the name, the address is listed as '9826 N LAKE CREEK PKWY, AUSTIN, Texas, United States'.

The screenshot shows the record view for 'Lili Paskin' with a 'Home' flag in a purple box and a 'Primary' flag in a green box. Below the name, the address is listed as '9826 N LAKE CREEK PKWY, AUSTIN | Texas | United States | 78717-5999'.

# Reports

- We can create **custom reports based on records with validation errors**.
- In the report, we can use different types of charts, dimensions and measure to visualize the validation metrics.
- Each **column is clickable which shows all records that failed that particular rule**.
- For these records, we could either do a bulk edit or do a sequential edit to fix the validation error.
- We can also export this report to perform further analysis.



# Bulk and Sequential edits

- Bulk edit is highly useful when data being edited falls into generalized patterns.
- If we want to make bulk edits to a set of records to correct the validation errors, we could use the bulk edit feature.
- The bulk edit comes with its own workspace and can be invoked through the search tab.
- The bulk edit is similar to search and replace functionality. It has a criteria builder for finding and replacing values.
- Criteria can contain “AND” and “OR” clauses. For example : find city = New York and city = Chicago.
- We can replace any field in the business entity based on this search criteria.
- Note : **The updates are done only to the Master record and not the source record. The updates create a patch XREF**

# Best Practices

Perform Data profiling in initial stages

Have source specific rules.

Basic rules with external validation run only in real-time.

Simple advanced data quality rule associations.

Don't use a rule that contains mapplet in the rule logic

All the advanced rules run during ingress and file import

Advanced rules can run on null values.

# Questions ?

Where data & AI come to **LIFE**



# Thank You

Where data & AI come to **LIFE**



# Where data & AI come to

