Apr 26th, 2022

Cloud Data Integration with Cloud Data Quality

Shiv Patel, Principal Customer Success Technologist,
Customer Success Management

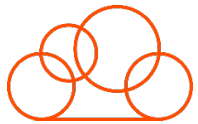
Informatica

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 INFASupport YouTube channel and 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.

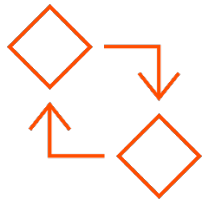
Informatica

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

Informatica

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

Informatica

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

Cloud Data Integration with Cloud Data Quality

Informatica

Agenda

1. Data Quality
2. Profiling
3. Data Quality Assets
4. Demo

Data Quality

Measure

- Review Progress
- Threshold Alerts
- Scorecards

Discover

- Profiling
- Identify Data Issues
- Set Data Quality Goals

Apply

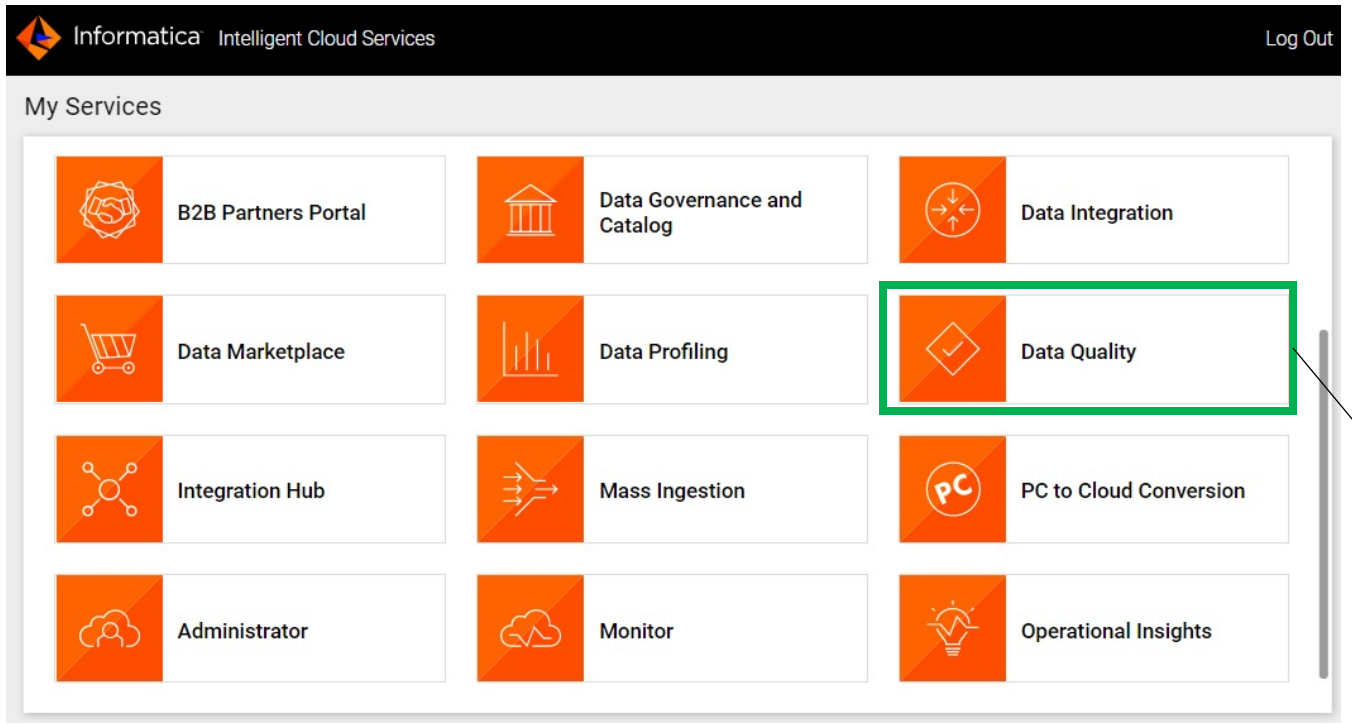
- Mapping Generation
- Standardization / Validation
- Matching / Consolidation

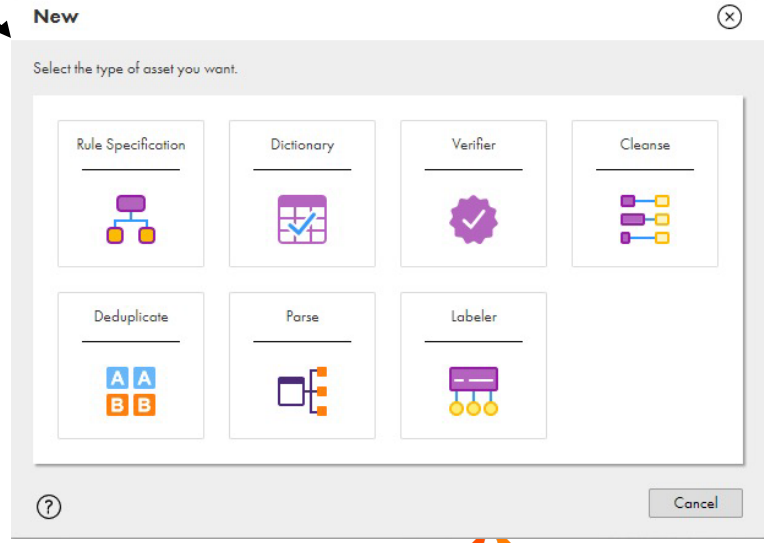
Define

- Dictionaries
- Implement Rules
- Build Cleanse, Parse, Verification, etc. processes

Informatica

Data Quality on IICS





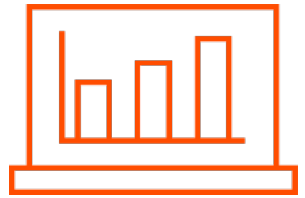
Cloud Data Profiling

- Analyze data schemas
- Determine the quality of data across sources
- Understand the completeness, conformity, and consistency of data in the data sources.

Informatica

Cloud Data Profiling Statistics

Metrics Analyzed on the Profiled Data

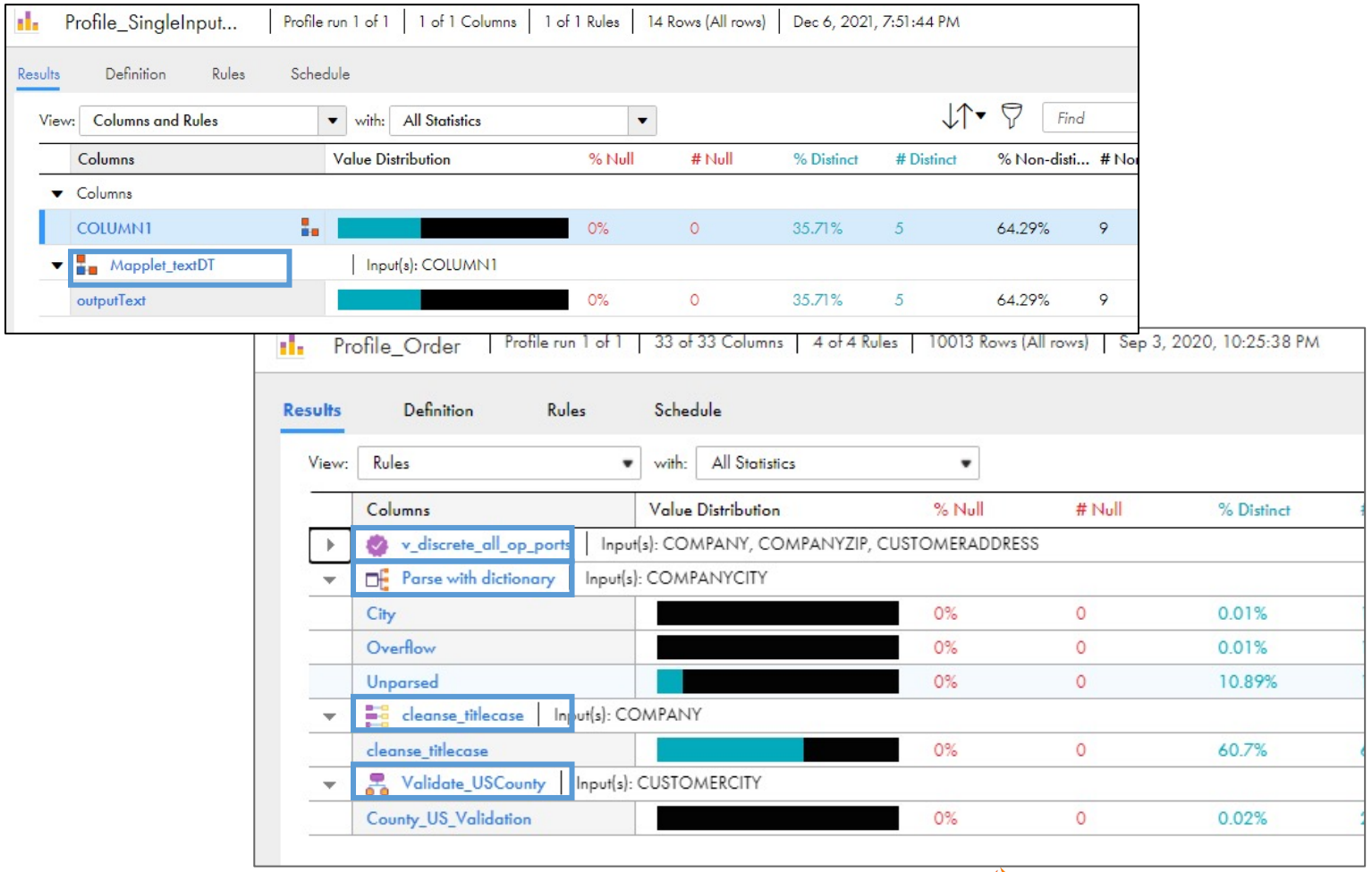


Null (%/#)	Min Value	Patterns
Distinct (%/#)	Max Value	Documented Data Type
Non-Distinct (%/#)	Min Length	Inferred Data Type
Blanks (%/#)	Max Length	Value Frequency
Average, Sum, Std Deviation, **Zeros (%/#)** NEW (for numeric data types)		Pattern and Value Frequency Outlier

Informatica

Cloud Data Profiling

- Perform What-If scenarios by profiling outputs of Data Quality Assets and Mapplets
  - Mapplets
  - Parser
  - Verifier
  - Cleansing
  - Rule Specification



Informatica

Data Quality Assets

Data Quality Assets

- Verifier
- Labeler
- Deduplicate
- Cleanse
- Dictionary
- Rule Specification
- Parse

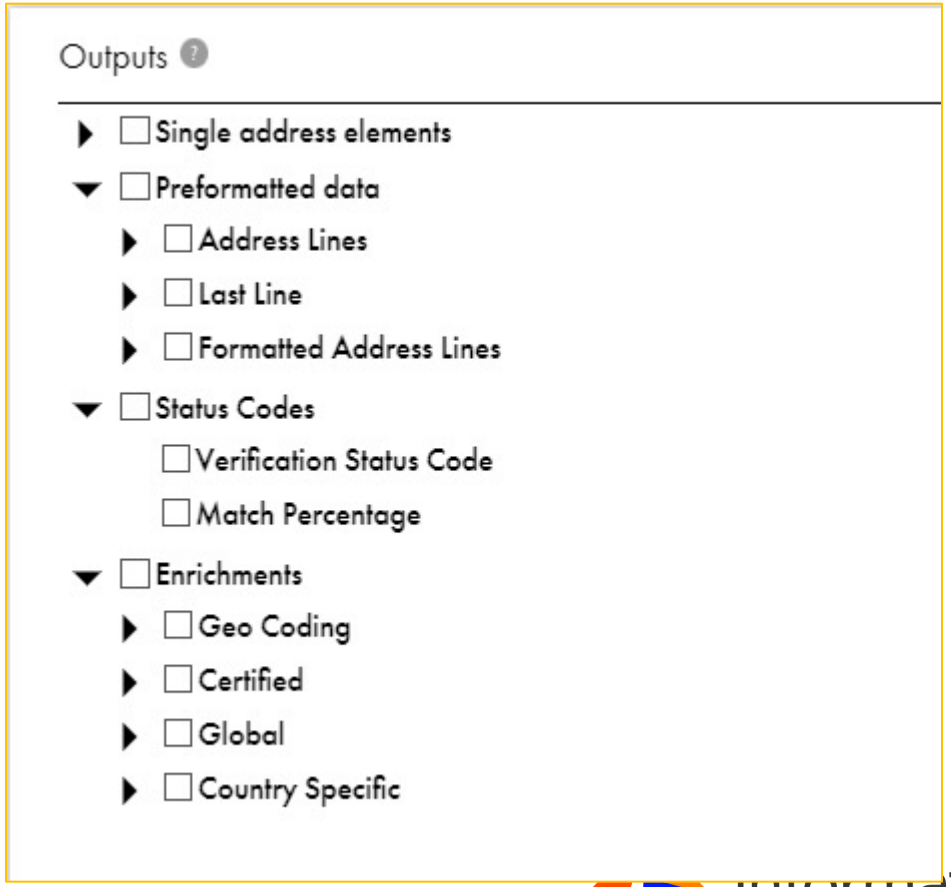
Verifier Assets

- Evaluates the accuracy and deliverability of address records
- Determine the accuracy of the input addresses
- Fix errors in the addresses
- Enhance the addresses where possible with additional information.
- Measure and report on the quality of each address

Informatica

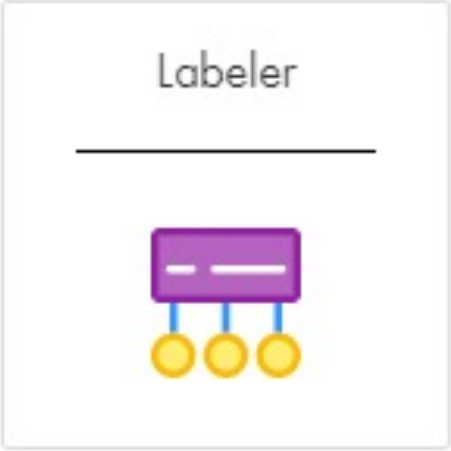
Verifier Assets

- The verifier can perform the following operations
  - Verify
  - Format
  - Suggest
  - Measure
  - Enrich
  - Certify



Labeler Assets

- Derives information about the content and structure of data.
- Perform token labeling or character labeling.
- Token labeling analyzes one or more tokens, or delimited values, in an input field
- Character labeling analyzes the individual characters in the input field.

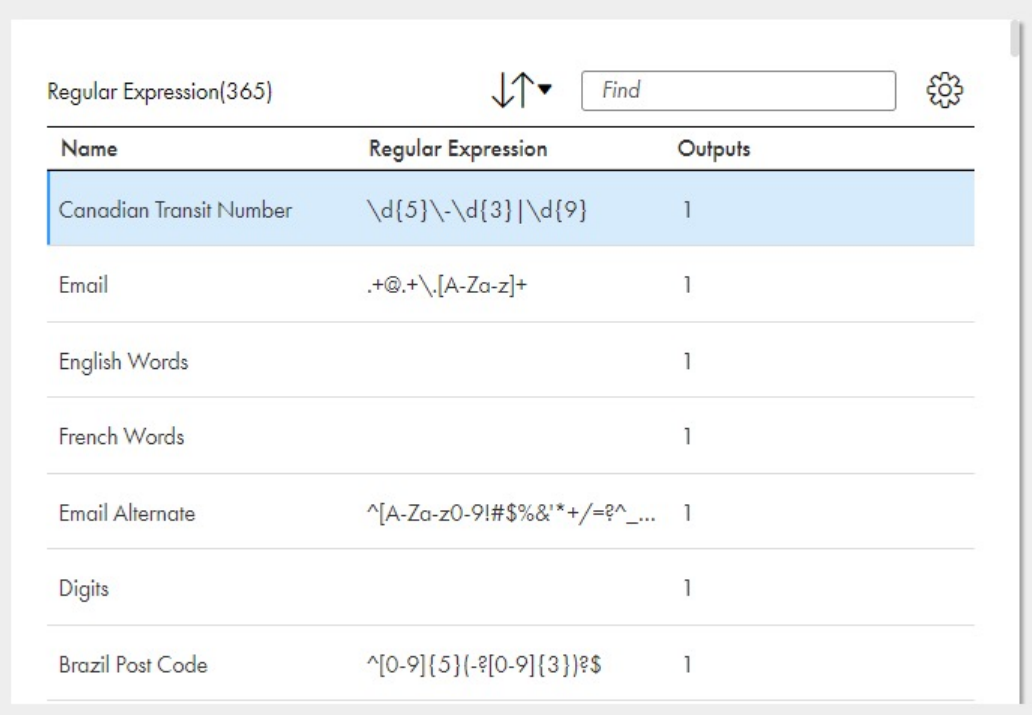


Informatica

Labeler Assets

When to use a labeler asset

- Verify business information with dictionaries
- Identify data by character format
- Review the structure of your input data



Informatica

Deduplicate Assets

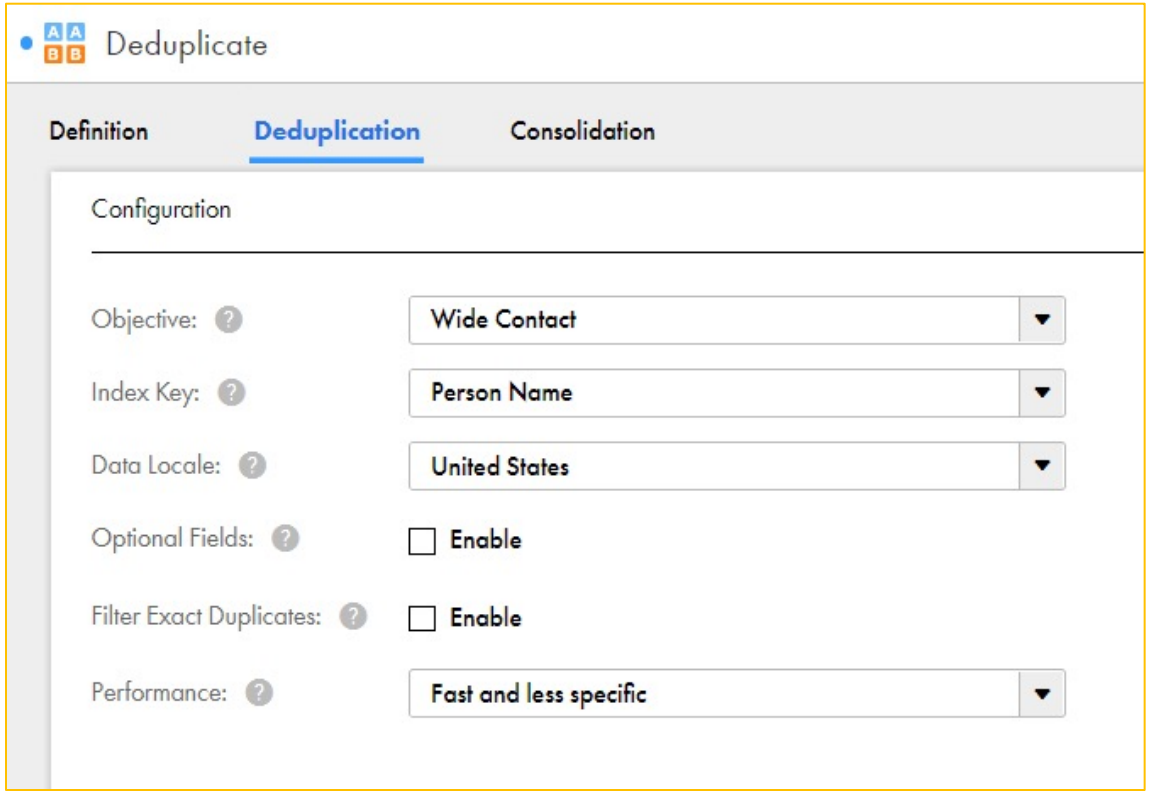
- Calculates degree of similarity between records and generates output data for these calculations
- Based on the Match Transformation
  - Identity matching
  - Single source / single data set
- An identity is a set of data values that collectively identify a person or organization

Informatica

Deduplicate Assets

Duplicate analysis methodology

- It's a set of instructions for a Deduplicate transformation.
- Select the type of identity to search for at run time
- Define the search criteria to the input data.



Cleanse Assets

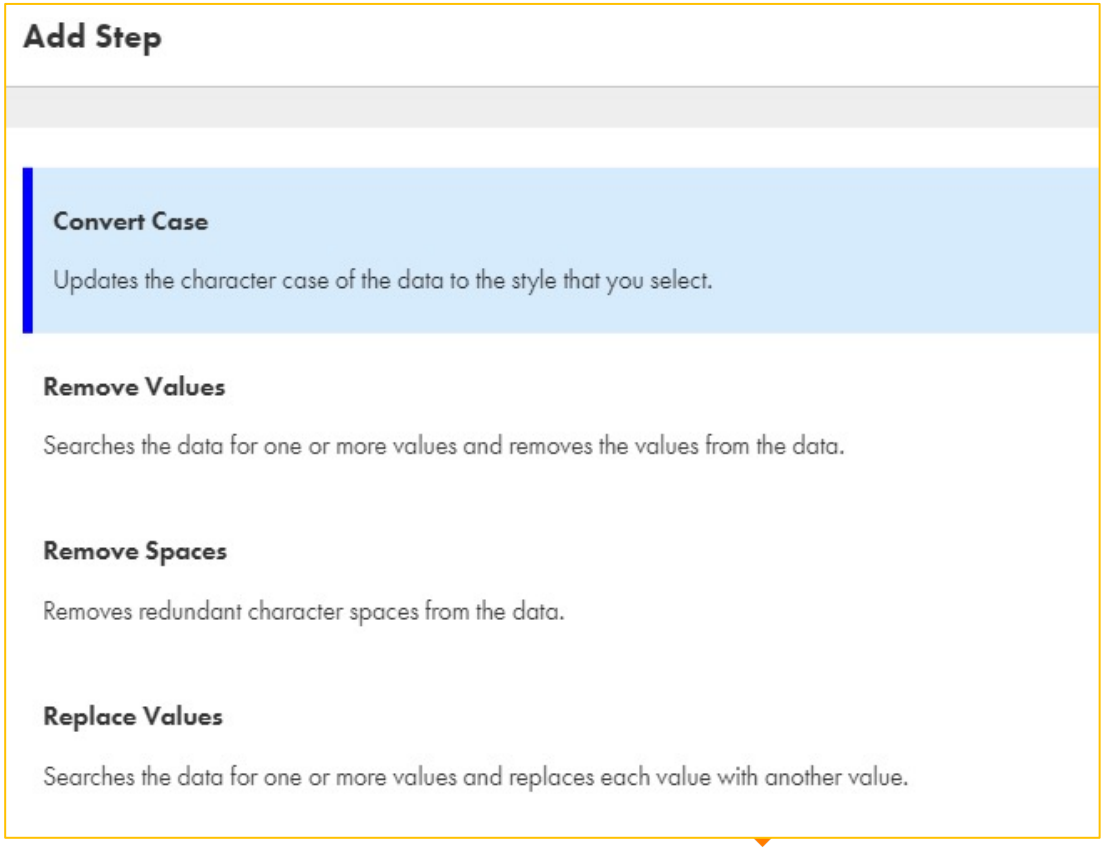
- It's a set of one or more data transformation steps that can standardize the form and content of your data
- Perform cleanse and merge operations on the input fields that a cleanse instance identifies in the asset

Informatica

Cleanse Assets

Standardize the data

- Improve data consistency in a data set
- Fix errors in data
- Comply with regulatory standards
- Prepare for downstream data quality initiatives
- Merge data fields to effectively manage your data set



Dictionary Assets

- Reference data set that you can use to evaluate data in a mapping
- Use dictionaries to verify that the data values on a data source or another object in a mapping are accurate and correctly formatted

Informatica

Dictionary Assets

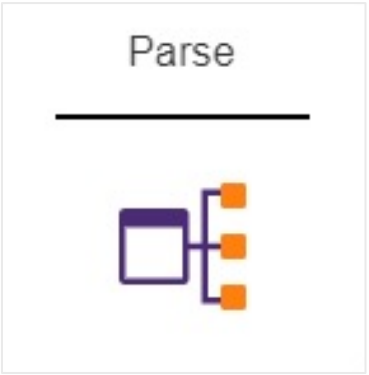
Dictionary Operations

- Compares the input field data to the data in the dictionary
- If match, the transformation performs an action based on definitions

Column 1	Column 2	Column 3	Column 4	Column 5
Mississippi	228	601	662	769
Missouri	314	417	573	636
Montana	406			
Nebraska	308	402	531	
Nevada	702	725	775	
New Hampshire	603			

Parse Asset

- Use a parse asset to improve the structure of your data
- A parse asset defines a set of operations that can identify discrete values in an input field and write the values to appropriate output fields

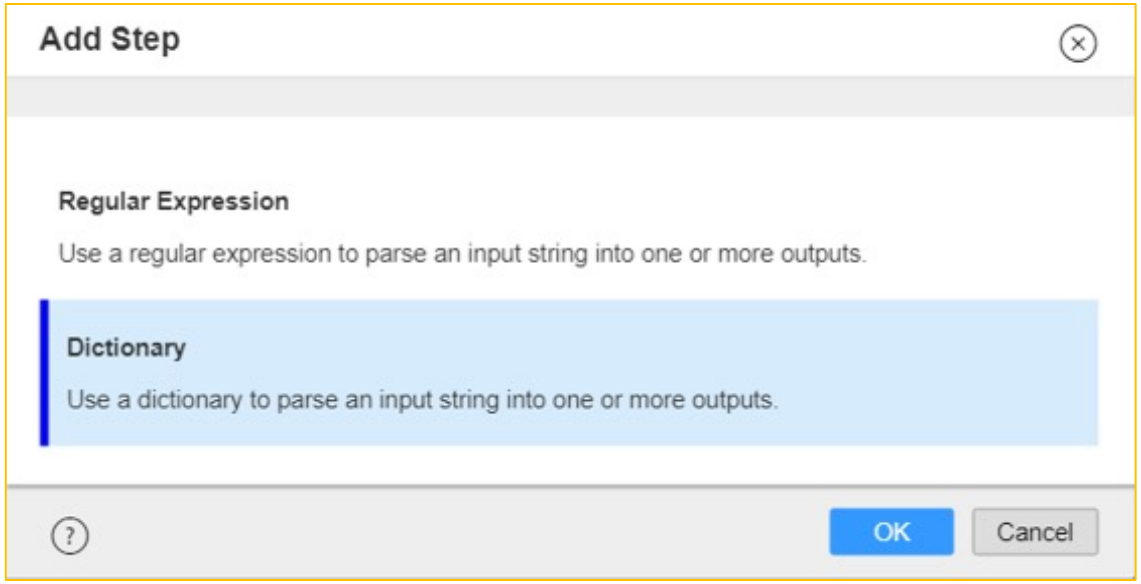


Informatica

Parse Assets

Parse operation to identify values in the following ways:

- Use a dictionary to identify values
- Use a regular expression to identify values
- Use a pre-built pattern to identify values



Rule Specification Asset

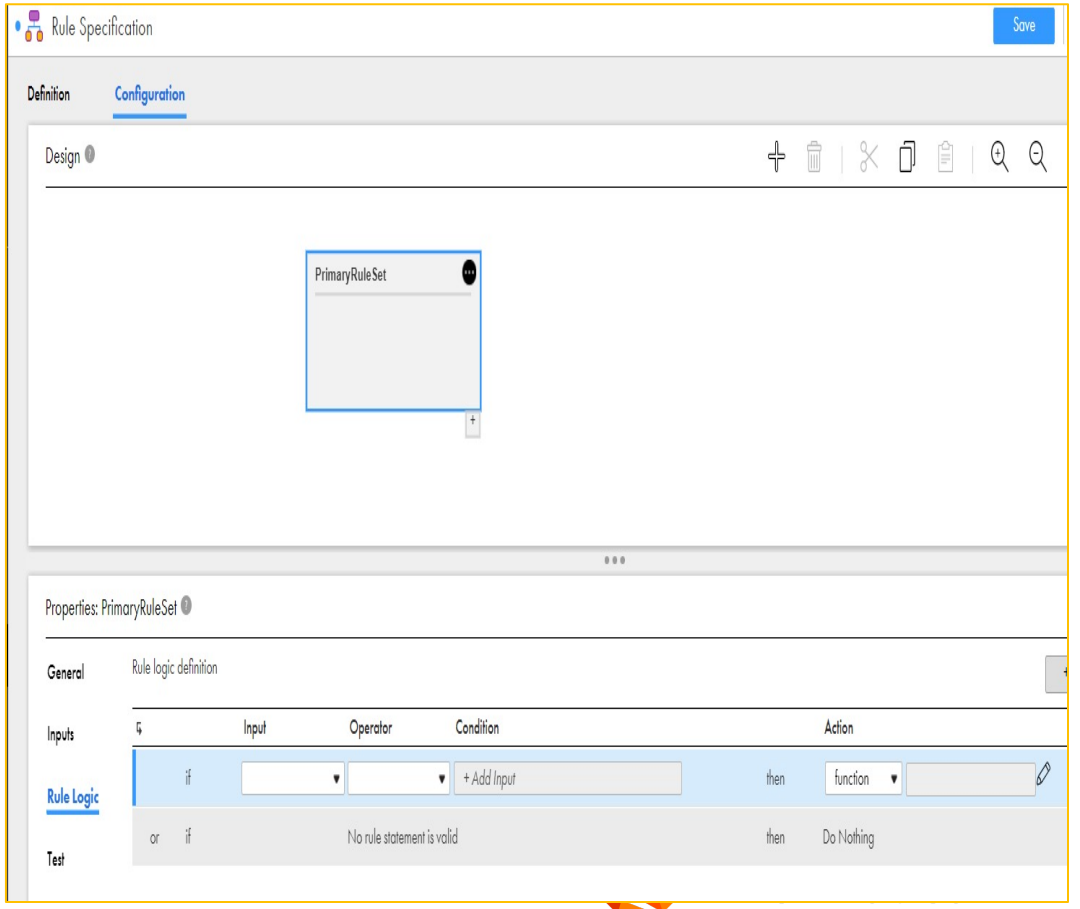
- A rule specification is an asset that represents the data requirements of a business rule in logical form

Informatica

Rule Specification Assets

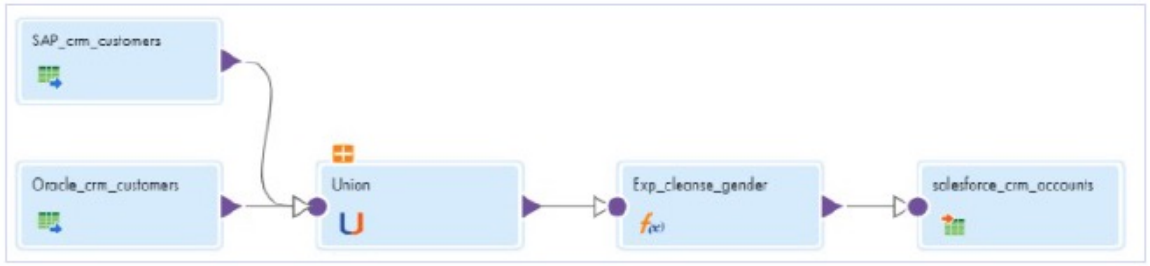
Rule Specification operation:

- Define the types of data that a business data set contains
- Define a set of conditions that the business data must satisfy
- Define the actions to take when the data satisfies the conditions of the business rule
- Define the actions to take when the data fails to satisfy the conditions of the business rule

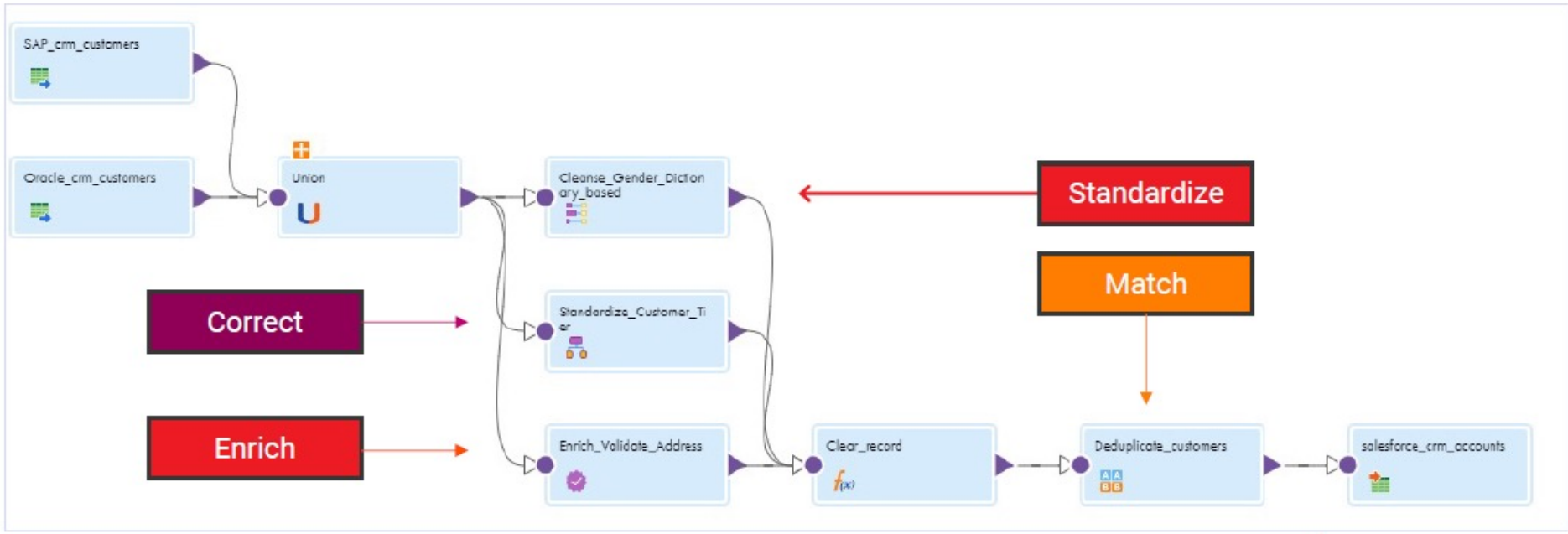


Data Quality in Data Integration Pipeline

Data Pipeline: Integration only



Data Pipeline: Integration and Data Quality



Informatica

Demo

Q&A