

Financial Crime Graph Model

Matching Guide

Release 8.0.7.3.0

March 2020

E91246.01

ORACLE®

Financial Services

Table of Contents

| | | |
|----------|--|-----------|
| 1 | Introduction | 4 |
| 2 | Managing Rulesets..... | 6 |
| 2.1 | Accessing the Entity Resolution Component..... | 6 |
| 2.2 | Creating Rulesets | 6 |
| 2.3 | Creating Rules in a Ruleset..... | 7 |
| 2.3.1 | <i>Scoring Method</i> | 8 |
| 3 | Matching Rulesets..... | 10 |
| 3.1 | Example..... | 10 |
| 3.1.1 | <i>Calculation of Score</i> | 14 |

Financial Crime Graph Model

Copyright © 2020 Oracle and/or its affiliates. All rights reserved.

This software and related documentation are provided under a license agreement containing restrictions on use and disclosure and are protected by intellectual property laws. Except as expressly permitted in your license agreement or allowed by law, you may not use, copy, reproduce, translate, broadcast, modify, license, transmit, distribute, exhibit, perform, publish, or display any part, in any form, or by any means. Reverse engineering, disassembly, or decompilation of this software, unless required by law for interoperability, is prohibited.

The information contained herein is subject to change without notice and is not warranted to be error-free. If you find any errors, please report them to us in writing.

If this is software or related documentation that is delivered to the U.S. Government or anyone licensing it on behalf of the U.S. Government, then the following notice is applicable:

U.S. GOVERNMENT END USERS: Oracle programs, including any operating system, integrated software, any programs installed on the hardware, and/or documentation, delivered to U.S. Government end users are “commercial computer software” pursuant to the applicable Federal Acquisition Regulation and agency-specific supplemental regulations. As such, use, duplication, disclosure, modification, and adaptation of the programs, including any operating system, integrated software, any programs installed on the hardware, and/or documentation, shall be subject to license terms and license restrictions applicable to the programs. No other rights are granted to the U.S. Government.

This software or hardware is developed for general use in a variety of information management applications. It is not developed or intended for use in any inherently dangerous applications, including applications that may create a risk of personal injury. If you use this software or hardware in dangerous applications, then you shall be responsible to take all appropriate fail-safe, backup, redundancy, and other measures to ensure its safe use. Oracle Corporation and its affiliates disclaim any liability for any damages caused by use of this software or hardware in dangerous applications.

Oracle and Java are registered trademarks of Oracle and/or its affiliates. Other names may be trademarks of their respective owners.

Intel and Intel Xeon are trademarks or registered trademarks of Intel Corporation. All SPARC trademarks are used under license and are trademarks or registered trademarks of SPARC International, Inc. AMD, Opteron, the AMD logo, and the AMD Opteron logo are trademarks or registered trademarks of Advanced Micro Devices. UNIX is a registered trademark of The Open Group.

This software or hardware and documentation may provide access to or information about content, products, and services from third parties. Oracle Corporation and its affiliates are not responsible for and expressly disclaim all warranties of any kind with respect to third-party content, products, and services unless otherwise set forth in an applicable agreement between you and Oracle. Oracle Corporation and its affiliates will not be responsible for any loss, costs, or damages incurred due to your access to or use of third-party content, products, or services, except as set forth in an applicable agreement between you and Oracle.

For information on third party licenses, click [here](#).

Document Control

| Version Number | Revision Date | Changes Done |
|----------------|---------------------|---|
| 8.0.7.3.0 | Created: March 2020 | Created the first version of the Financial Crime Graph Model Matching Guide for v8.0.7.3.0 Release. |

1 Introduction

In FCC Studio, data is obtained from FCDM (Financial Crime Data Model) to generate Financial Crime Graph Model. The graph model includes nodes for entities such as Customers, Accounts, Events, and Derived Entities, and edges for transactions and relationships.

Entity Resolution compares nodes with the objective to identify pairs or groups of nodes that refer to the same entity. Entity Resolution creates Similarity Edges between nodes by comparing the attributes of the nodes and identifying where the similarity is significant enough to create an edge so the nodes are linked with the graph model and can be analyzed as a single entity.

Entity matching rules are used to compare nodes of different types. For example, deduplicating customers, resolving derived entities, or linking customers or derived entities to external data such as panama papers or sanctions lists with different rules and thresholds.

For example:

A customer holds three different accounts in a bank with three different customer details.

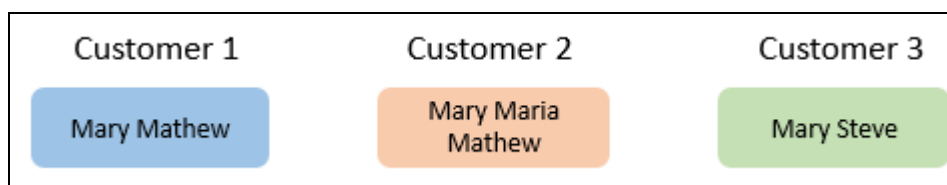


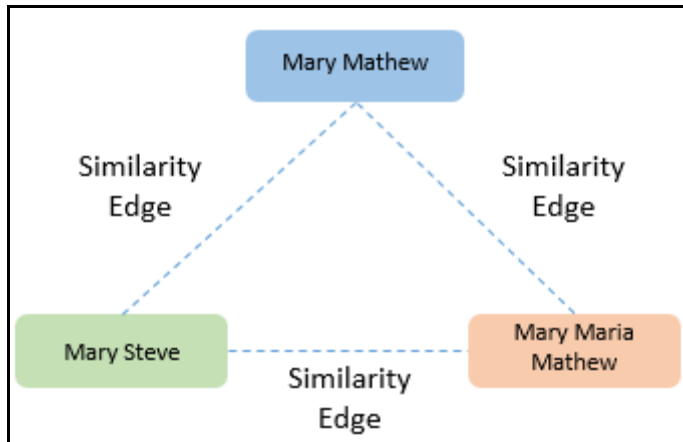
Table 1: Customer Details

| Bank | | | |
|------------------|-----------------------|-----------------------|----------------------|
| Customer Details | Customer 1 | Customer 2 | Customer 3 |
| Source | Source System 1 | Source System 2 | Source System 3 |
| Name | Mary Mathew | Mary Maria Mathew | Mary Steve |
| Email | Mary.Mathew@gmail.com | Mary.Mathew@gmail.com | Mary.Steve@gmail.com |
| Phone | Phone Number 1 | Phone Number 2 | Phone Number 3 |
| Country | United States | United States | United States |
| State | California | California | Washington |
| Address | Redwood City | Redwood City | 15th St NW |
| DOB | 1 Jan 1995 | 1 Jan 1995 | 1 Jan 1995 |
| Tax ID | Tax ID 1 | - | Tax ID 1 |

Using Entity Resolution, you can execute the Customer to Customer Ruleset on the customer data to compare the nodes such as Customer Name, Email, Phone, Country, State, Address, DOB, and Tax ID.

The result obtains an exact match on the DOB and TaxID, and fuzzy match on the Customer Name, Email ID, Address, State, and Country, and no match for the phone number.

This helps to derive to a conclusion to draw Similarity Edges between the three customers in the Bank.



2 Managing Rulesets

A Ruleset is a set of rules that are applied to the defined source and target entities, compares the attributes of the entities to derive a match.

2.1 Accessing the Entity Resolution Component

To access the Entity Resolution component, enter the URL in the following format into the browser:

`http://<Entity Resolution Hostname>:<Entity Resolution Port>/er/Rule.jsp`

The *Entity Resolution* page is displayed with all the out-of-the-box rulesets.

| RULESET NAME | DESCRIPTION | AGGREGATION | SOURCE ENTITY | TARGET ENTITY | ENABLE |
|---------------------------------------|--|-------------|----------------|---------------------------|-------------------------------------|
| Customer To Ext Source- Paradise Addr | Match Customer Address To External Source - Paradise Address | Maximum | customer | external_address_paradise | <input checked="" type="checkbox"/> |
| Customer To Ext Source - Panama | Customer To External Source - Panama | Maximum | customer | external_entity_panama | <input checked="" type="checkbox"/> |
| Customer to Customer Match | Match Customer to Customer based on their attributes | Maximum | customer | customer | <input checked="" type="checkbox"/> |
| Customer To Ext Source- Panama Addr | Match Customer Address To External Source - Paradise Address | Maximum | customer | external_address_panama | <input checked="" type="checkbox"/> |
| Customer To Derived Entity | Customer To Derived Entity | Maximum | customer | derived_entity | <input checked="" type="checkbox"/> |
| Derived Entity To Derived Entity | Derived Entity To derived Entity Match | Maximum | derived_entity | derived_entity | <input checked="" type="checkbox"/> |
| Customer To Ext Source- Offshore | Match Customer To External Source - Offshore | Maximum | customer | external_entity_offshore | <input checked="" type="checkbox"/> |

2.2 Creating Rulesets

To create a ruleset, follow these steps:

1. Navigate to the *Entity Resolution* page.

The Rulesets are displayed.

2. Click the **Add** icon

The *Ruleset Details* page is displayed.

ORACLE Entity Resolution

Ruleset Details Back Reset

| | | | |
|-----------------------------------|--|--------------------------|--|
| Name | Description | Scoring Aggregation Type | Set Threshold |
| <input type="text" value="Name"/> | <input type="text" value="Description"/> | Maximum | <input type="text" value="Threshold"/> |

Rules (0)

Source Target


3. Enter the following details.

| Field | Description |
|-------------|---|
| Name | Indicates the name of the ruleset. |
| Description | Indicates the additional description given for the ruleset. |

| Field | Description |
|--------------------------|---|
| Scoring Aggregation Type | Indicates the scoring aggregation method. Select one of the following options: <ul style="list-style-type: none"> Maximum: Considers the highest score obtained out of all the rules created for a ruleset. Minimum: Considers the lowest score obtained out of all the rules created for a ruleset. |
| Set Threshold | Indicates the threshold value set for a ruleset. A Similarity Edge is generated only when the maximum score obtained for a ruleset is equal to or higher than the threshold value. |
| Source | Indicates the source entity (node). The values are auto-populated from the metadata table that contains the elastic search index names generated as a result of running the Sqoop job. |
| Target | Indicates the target entity (node). The values are auto-populated from the metadata table that contains the elastic search index names generated as a result of running the Sqoop job. |

2.3 Creating Rules in a Ruleset

To create rules in a ruleset, follow these steps:

1. Navigate to a *Ruleset* page.
2. Click the **Add** icon .
A *New Rule* section is displayed.
3. Enter the following details.

| Field | Description |
|----------------|--|
| Name | Indicates the name of the rule. |
| Description | Indicates the description of the rule. |
| Rule Threshold | Indicates the threshold value set for a rule. This rule contributes to the matching, only when the maximum score obtained for a rule is equal to or higher than the threshold value. |

4. Click the **Add** icon  to add new Mappings:

| Field | Description |
|------------------|---------------------------------|
| Source Attribute | Indicates the source attribute. |
| Target Attribute | Indicates the target attribute. |

| Field | Description |
|----------------|---|
| Match Type | <p>Indicates the match type.</p> <p>Select one of the following options:</p> <ul style="list-style-type: none"> Exact: To obtain the matches that are 100% perfect when finding the entities in a database. Fuzzy: To obtain the matches that are less than 100% perfect when finding the entities in a database. |
| Scoring Method | <p>The scoring methods used are as follows:</p> <ul style="list-style-type: none"> Default Jaro Winkler <p>For more information, see Scoring Method.</p> |
| Threshold | <p>Indicates that a score below the mentioned value does not generate a result from the elastic search.</p> |
| Weightage | <p>Indicates the weightage given for the attributes in the rule.</p> |
| Condition | <p>Indicates that this attribute cannot have a null value. This attribute must be populated and must return a value for the matching.</p> |

2.3.1 Scoring Method

The scoring methods used in the entity resolution component are as follows:

- Default Method

The distance is computed by finding the number of edits which transforms one string to another. The transformations allowed are as follows:

- Insertion: Adding a new character
- Deletion: Deleting a character
- Substitution: Replace one character with another

By performing these operations, the algorithm attempts to modify the first string to match the second one. The final result obtained is the edit distance.

For example:

```
a. textdistance.levenshtein('arrow', 'arow')
```

```
1
```

```
b. >> textdistance.levenshtein.normalized_similarity('arrow', 'arow')
```

```
0.8
```

Here, if you insert single 'r' in string 2, that is, 'arow', it becomes same as the string 1. Hence, the edit distance is 1. Similar with Hamming distance, you can generate a bounded similarity score between 0 and 1. The similarity score obtained is 80%.

- Jaro Winkler

This algorithm gives high scores for the following strings:

- The strings that contain same characters, but within a certain distance from one another.

- b. The order of the matching characters is same.

To be precise, the distance of finding similar character is one character less than half of the length of the longest string. So if the longest string has a length of five, a character at the start of the string 1 must be found before or on $((5/2)-1) \sim 2$ nd position in the string 2. This is considered a valid match. Hence, the algorithm is directional and gives high score if matching is from the beginning of the strings.

For example:

- a. `textdistance.jaro_winkler("mes", "messi")`
0.86
- b. `textdistance.jaro_winkler("crate", "crat")`
0.96
- c. `textdistance.jaro_winkler("crate", "atcr")`
0.0

In first case, as the strings are matching from the beginning, high score is given. Similarly, in the second case, only one character was missing and that too at the end of the string 2, hence a very high score is given. In third case, the last two character of string 2 are rearranged by bringing them at front and hence results in 0% similarity.

3 Matching Rulesets

Each ruleset comprises of multiple rules. The ruleset compares the attributes that are defined in the rules for the source entity with the target entity.

A list of rulesets that are packaged with the FCC Studio application are as follows.

Table 1: List of Rulesets

| Ruleset Name | Source Node Type | Target Node Type |
|--|------------------|---------------------------|
| Customer To Customer Match | customer | customer |
| Customer To Derived Entity | customer | derived_entity |
| Derived Entity To Derived Entity | derived_entity | derived_entity |
| Customer To Ext Source - Offshore | customer | external_entity_offshore |
| Customer To Ext Source - Bahamas | customer | external_entity_bahamas |
| Customer To Ext Source - Paradise | customer | external_address_paradise |
| Customer To Ext Source - Panama | customer | external_entity_panama |
| Customer To Ext Source - Offshore Addr | customer | external_address_offshore |
| Customer To Ext Source - Bahamas Addr | customer | external_address_bahamas |
| Customer To Ext Source - Paradise Addr | customer | external_address_panama |
| Customer To Ext Source - Panama Addr | customer | external_address_paradise |

3.1 Example

A customer holds three different accounts in a bank with three different customer details.

Table 2: Customer Details

| Bank | | | |
|------------------|-----------------------|-----------------------|----------------------|
| Customer Details | Customer 1 | Customer 2 | Customer 3 |
| Source | Source System 1 | Source System 2 | Source System 3 |
| Name | Mary Mathew | Mary Maria Mathew | Mary Steve |
| Email | Mary.Mathew@gmail.com | Mary.Mathew@gmail.com | Mary.Steve@gmail.com |
| Phone | Phone Number 1 | Phone Number 2 | Phone Number 3 |
| Country | United States | United States | United States |
| State | California | California | Washington |
| Address | Redwood City | Redwood City | 15th St NW |
| DOB | 1 Jan 1995 | 1 Jan 1995 | 1 Jan 1995 |
| Tax ID | Tax ID 1 | - | Tax ID 1 |

The Customer to Customer Match ruleset compares the attributes defined for the source (customer) and target (customer) entities of each rule. If the score of the combination of the result obtained for all the rules in a ruleset is equal to or greater than the threshold set for the ruleset, a Similarity Edge is formed between the source and the target entity.

The Customer to Customer Match ruleset is given as follows:

| Ruleset Details | | | |
|----------------------------|--|--------------------------|---------------|
| Name | Description | Scoring Aggregation Type | Set Threshold |
| Customer to Customer Match | Match Customer to Customer based on their attributes | Maximum | 0.75 |

The rules and the attributes in the Customer to Customer Match ruleset are as follows.

Table 3: Customer to Customer Ruleset

| Ruleset | Rules | Parameters/Attributes |
|------------------------------|------------------------|--------------------------------|
| Customer to Customer Ruleset | Name and Address Match | Name, Alias, Concatenated Name |
| | | Email |
| | | Phone |
| | | Country |
| | | State |
| | | Address |
| | | DOB |
| | Tax ID | Tax ID |

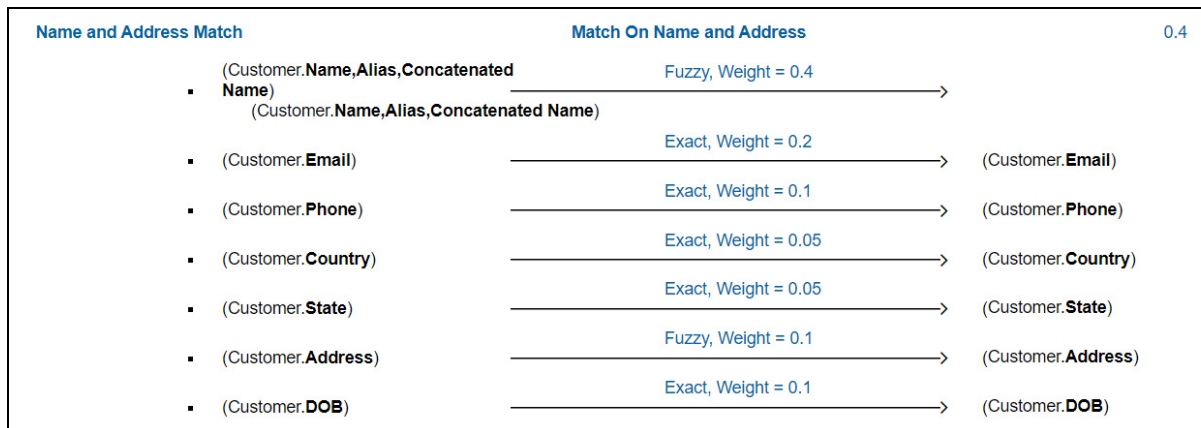


Table 4: Name and Address Match Rule

| Parameters/Attributes | Source Entity | Target Entity | Match Type | Weightage | Threshold | Description |
|--------------------------------|---------------|---------------|------------|-----------|-----------|--|
| Name, Alias, Concatenated Name | Customer | Customer | Fuzzy | 0.4 | 0.5 | <ul style="list-style-type: none"> The Name, Alias, and Concatenated Name attributes of the source entity (Customer) is compared with the target entity (Customer) to obtain a fuzzy match. If the fuzzy match generates a result that is equal to or greater than the threshold value (0.5), a weightage of 0.4 is contributed to this match. |
| Email | Customer | Customer | Exact | 0.2 | 1 | <ul style="list-style-type: none"> The Email address of the source entity (Customer) is compared with the target entity (Customer) to obtain an exact match. If an exact match is obtained, a weightage of 0.2 is contributed to this match. |
| Phone | Customer | Customer | Exact | 0.1 | 1 | <ul style="list-style-type: none"> The phone number of the source entity (Customer) is compared with the target entity (Customer) to obtain an exact match. If an exact match is obtained, a weightage of 0.1 is contributed to this match. |

Table 4: Name and Address Match Rule

| Parameters/Attributes | Source Entity | Target Entity | Match Type | Weightage | Threshold | Description |
|-----------------------|---------------|---------------|------------|-----------|-----------|--|
| Country | Customer | Customer | Exact | 0.05 | 1 | <ul style="list-style-type: none"> The country of the source entity (Customer) is compared with the target entity (Customer) to obtain an exact match. If an exact match is obtained, a weightage of 0.05 is contributed to this match. |
| State | Customer | Customer | Exact | 0.05 | 1 | <ul style="list-style-type: none"> The state of the source entity (Customer) is compared with target entity (Customer) to obtain an exact match. If an exact match is obtained, a weightage of 0.05 is contributed to this match. |
| Addresses | Customer | Customer | Fuzzy | 0.1 | 0.6 | <ul style="list-style-type: none"> The address of the source entity (Customer) is compared with the target entity (Customer) to obtain a fuzzy match. If the fuzzy match generates a result that is equal to or greater than the threshold value (0.6), a weightage of 0.1 is contributed to this match. |
| DOB | Customer | Customer | Exact | 0.1 | 1 | <ul style="list-style-type: none"> The date of birth of the source entity (Customer) is compared with the target entity (Customer) to obtain an exact match. If an exact match is obtained, a weightage of 0.1 is contributed to this match. |



Table 5: Tax ID Rule

| Parameters/ Attributes | Source Entity | Target Entity | Match Type | Weight age | Description |
|------------------------|---------------|---------------|------------|------------|---|
| Tax ID | Customer | Customer | Exact | 1 | <ul style="list-style-type: none"> The Tax ID of the source entity (Customer) is compared with the target entity (customer) to obtain an exact match. If an exact match is obtained, a weightage of 1 is contributed to this match. |

3.1.1 Calculation of Score

Table 6: Calculation of Score

| Customer Details | Customer 1 | Customer 2 | Score | Weight (From Rule) | Weighted Score |
|------------------|-----------------------|-----------------------|-------|--------------------|----------------|
| Name | Mary Mathew | Mary Maria Mathew | 93.07 | 0.4 | 37.22 |
| Email | Mary.Mathew@gmail.com | Mary.Mathew@gmail.com | 100 | 0.2 | 20 |
| Phone | Phone Number 1 | Phone Number 2 | 100 | 0.1 | 10 |
| Country | United States | United States | 100 | 0.05 | 5 |
| State | California | California | 100 | 0.05 | 5 |
| Address | Redwood City | Redwood City | 100 | 0.1 | 10 |
| DOB | 1 Jan 1995 | 1 Jan 1995 | 100 | 0.1 | 10 |
| | | | | Total=1 | Total=97.22 |

The total score obtained is greater than the rule threshold of 40%, a Similarity Edge is created between Customer1 and Customer 2.

Similar calculation is performed for all possible combination of customers like Customer 2 and Customer 3, Customer 1 and Customer 3.

