



Informatica® MDM - Relate 360
10.1

Release Guide

This software and documentation are provided only under a separate license agreement containing restrictions on use and disclosure. No part of this document may be reproduced or transmitted in any form, by any means (electronic, photocopying, recording or otherwise) without prior consent of Informatica LLC.

U.S. GOVERNMENT RIGHTS Programs, software, databases, and related documentation and technical data delivered to U.S. Government customers are "commercial computer software" or "commercial technical data" pursuant to the applicable Federal Acquisition Regulation and agency-specific supplemental regulations. As such, the use, duplication, disclosure, modification, and adaptation is subject to the restrictions and license terms set forth in the applicable Government contract, and, to the extent applicable by the terms of the Government contract, the additional rights set forth in FAR 52.227-19, Commercial Computer Software License.

Informatica and the Informatica logo are trademarks or registered trademarks of Informatica LLC in the United States and many jurisdictions throughout the world. A current list of Informatica trademarks is available on the web at <https://www.informatica.com/trademarks.html>. Other company and product names may be trade names or trademarks of their respective owners.

The information in this documentation is subject to change without notice. If you find any problems in this documentation, report them to us at infa_documentation@informatica.com.

Informatica products are warranted according to the terms and conditions of the agreements under which they are provided. INFORMATICA PROVIDES THE INFORMATION IN THIS DOCUMENT "AS IS" WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND ANY WARRANTY OR CONDITION OF NON-INFRINGEMENT.

Publication Date: 2019-06-27

Table of Contents

Preface	6
Informatica Resources.	6
Informatica Network.	6
Informatica Knowledge Base.	6
Informatica Documentation.	6
Informatica Product Availability Matrices.	7
Informatica Velocity.	7
Informatica Marketplace.	7
Informatica Global Customer Support.	7
 Part I: Version 10.1.....	 8
 Chapter 1: New Features and Changes (10.1).....	 9
New Features (10.1).	9
Probable Matching Records.	9
 Part II: Version 10.0.0.....	 10
 Chapter 2: New Features and Changes (10.0 HotFix 9).....	 11
New Features (10.0 HotFix 9).	11
Retrieving the Preferred Record Count.	11
Changes (10.0 HotFix 9).	11
Version Number of the RESTful Web Services.	11
Searching Multiple Records.	12
Deleting Records in the Repository.	12
Matching Null Values.	12
Optimal Preferred Record Search.	12
Retrieving Records in a Cluster.	12
 Chapter 3: New Features and Changes (10.0 HotFix 8).....	 13
New Features (10.0 HotFix 8).	13
Kerberos Authentication for Streaming Data.	13
Changes (10.0 HotFix 8).	13
Resetting the Offset Position for a Spark Instance.	13
Search Client is Deprecated.	14
 Chapter 4: New Features, Changes, and Release Tasks (10.0 HotFix 7).....	 15
New Features (10.0 HotFix 7).	15
Relationship Graph.	15
Batch Jobs.	15

RESTful Web Services.	16
Changes (10.0 HotFix 7).	16
Version Number of the RESTful Web Services.	16
Chapter 5: New Features, Changes, and Release Tasks (10.0 HotFix 6).	17
New Features (10.0 HotFix 6).	17
Consolidation Process.	17
Consolidation Rules.	17
Retrieving the Preferred Records.	17
Hive Enabler Job.	18
Changes (10.0 HotFix 6).	18
Version Number of the RESTful Web Services.	18
Chapter 6: New Features, Changes, and Release Tasks (10.0 HotFix 5).	19
New Features (10.0 HotFix 5).	19
Pagination Support.	19
Distributed Search.	19
Security for the RESTful Web Services.	19
Processing Data in Motion.	20
Region Splitter Job.	20
Post-Clustering Job.	20
Changes (10.0 HotFix 5).	20
Updating Linked Data in Hive.	20
License File.	21
Version Number of the RESTful Web Services.	21
Deprecated Batch Jobs and Operations.	21
Chapter 7: New Features, Changes, and Release Tasks (10.0.0 HotFix 4).	22
New Features (10.0.0 HotFix 4).	22
Support for the Kerberos Authentication in Cloudera.	22
Debug Mode for the MULTISEARCH Operation.	22
Changes (10.0.0 HotFix 4).	22
Search.	23
Chapter 8: New Features, Changes, and Release Tasks (10.0.0 HotFix 3).	24
New Features (10.0.0 HotFix 3).	24
Tokenization.	24
MapReduce Jobs.	24
Search.	25
Changes (10.0.0 HotFix 3).	26
Matching Rules File.	26

Chapter 9: New Features, Changes, and Release Tasks (10.0.0 HotFix 2). . . . 27

New Features (10.0.0 HotFix 2).	27
Advanced Matching.	27
MapReduce Jobs.	27
Search.	28
Changes (10.0.0 HotFix 2).	28
Simple Matching.	29
Updating Indexed and Linked Data.	29

Preface

The *Informatica MDM - Relate 360 Release Guide* lists new features and enhancements, behavior changes between versions, and tasks you might need to perform after you upgrade from a previous version. The *Informatica MDM - Relate 360 Release Guide* is written for users of Informatica MDM - Relate 360. This guide assumes that you have knowledge of the features for which you are responsible.

Informatica Resources

Informatica provides you with a range of product resources through the Informatica Network and other online portals. Use the resources to get the most from your Informatica products and solutions and to learn from other Informatica users and subject matter experts.

Informatica Network

The Informatica Network is the gateway to many resources, including the Informatica Knowledge Base and Informatica Global Customer Support. To enter the Informatica Network, visit <https://network.informatica.com>.

As an Informatica Network member, you have the following options:

- Search the Knowledge Base for product resources.
- View product availability information.
- Create and review your support cases.
- Find your local Informatica User Group Network and collaborate with your peers.

Informatica Knowledge Base

Use the Informatica Knowledge Base to find product resources such as how-to articles, best practices, video tutorials, and answers to frequently asked questions.

To search the Knowledge Base, visit <https://search.informatica.com>. If you have questions, comments, or ideas about the Knowledge Base, contact the Informatica Knowledge Base team at KB_Feedback@informatica.com.

Informatica Documentation

Use the Informatica Documentation Portal to explore an extensive library of documentation for current and recent product releases. To explore the Documentation Portal, visit <https://docs.informatica.com>.

Informatica maintains documentation for many products on the Informatica Knowledge Base in addition to the Documentation Portal. If you cannot find documentation for your product or product version on the Documentation Portal, search the Knowledge Base at <https://search.informatica.com>.

If you have questions, comments, or ideas about the product documentation, contact the Informatica Documentation team at infa_documentation@informatica.com.

Informatica Product Availability Matrices

Product Availability Matrices (PAMs) indicate the versions of the operating systems, databases, and types of data sources and targets that a product release supports. You can browse the Informatica PAMs at <https://network.informatica.com/community/informatica-network/product-availability-matrices>.

Informatica Velocity

Informatica Velocity is a collection of tips and best practices developed by Informatica Professional Services and based on real-world experiences from hundreds of data management projects. Informatica Velocity represents the collective knowledge of Informatica consultants who work with organizations around the world to plan, develop, deploy, and maintain successful data management solutions.

You can find Informatica Velocity resources at <http://velocity.informatica.com>. If you have questions, comments, or ideas about Informatica Velocity, contact Informatica Professional Services at ips@informatica.com.

Informatica Marketplace

The Informatica Marketplace is a forum where you can find solutions that extend and enhance your Informatica implementations. Leverage any of the hundreds of solutions from Informatica developers and partners on the Marketplace to improve your productivity and speed up time to implementation on your projects. You can find the Informatica Marketplace at <https://marketplace.informatica.com>.

Informatica Global Customer Support

You can contact a Global Support Center by telephone or through the Informatica Network.

To find your local Informatica Global Customer Support telephone number, visit the Informatica website at the following link:

<https://www.informatica.com/services-and-training/customer-success-services/contact-us.html>.

To find online support resources on the Informatica Network, visit <https://network.informatica.com> and select the eSupport option.

Part I: Version 10.1

This part contains the following chapter:

- [New Features and Changes \(10.1\), 9](#)

CHAPTER 1

New Features and Changes (10.1)

This chapter includes the following topic:

- [New Features \(10.1\), 9](#)

New Features (10.1)

This section describes new features in version 10.1.

Probable Matching Records

You can configure a lower threshold value in the configuration file to identify the probable matching records. If a match score is between the threshold and lower threshold values, the record is considered as a probable matching record.

The initial clustering job writes the probable matching records to an output directory in HDFS. If you use Kafka in your environment, you can write the probable matching records to a Kafka topic.

For more information about configuring a lower threshold value, see the *Informatica MDM - Relate 360 Installation and Configuration Guide*.

Part II: Version 10.0.0

This part contains the following chapters:

- [New Features and Changes \(10.0 HotFix 9\), 11](#)
- [New Features and Changes \(10.0 HotFix 8\), 13](#)
- [New Features, Changes, and Release Tasks \(10.0 HotFix 7\), 15](#)
- [New Features, Changes, and Release Tasks \(10.0 HotFix 6\), 17](#)
- [New Features, Changes, and Release Tasks \(10.0 HotFix 5\), 19](#)
- [New Features, Changes, and Release Tasks \(10.0.0 HotFix 4\), 22](#)
- [New Features, Changes, and Release Tasks \(10.0.0 HotFix 3\), 24](#)
- [New Features, Changes, and Release Tasks \(10.0.0 HotFix 2\), 27](#)

CHAPTER 2

New Features and Changes (10.0 HotFix 9)

This chapter includes the following topics:

- [New Features \(10.0 HotFix 9\), 11](#)
- [Changes \(10.0 HotFix 9\), 11](#)

New Features (10.0 HotFix 9)

This section describes new features in version 10.0 HotFix 9.

Retrieving the Preferred Record Count

Use the Get Preferred Record Count web service or command-line operation to retrieve the total number of preferred records stored in the repository.

For more information about the Get Preferred Record Count web service or command-line operation, see *the Informatica MDM - Relate 360 User Guide*.

Changes (10.0 HotFix 9)

This section describes changes in version 10.0 HotFix 9.

Version Number of the RESTful Web Services

Effective in version 10.0 HotFix 9, all the RESTful web services use v5.0 as the version number in the URLs.

The following sample URL uses v5.0 as the version number:

```
http://BDRMServer:8080/MDMBDRMCustomer360/v5.0/Customer360/MULTISEARCH
```

Previously, all the RESTful web services used v4.0 as the version number in the URLs.

Searching Multiple Records

The following web services or command-line operations can process more than one input record:

- Get Record
- Get Preferred Record

Previously, the Get Record and Get Preferred Record web services and command-line operations could process only one input record.

For more information about the Get Record and Get Preferred Record web services and command-line operations, see the *Informatica MDM - Relate 360 User Guide*.

Deleting Records in the Repository

The repository data deletion job can delete all the records that belong to a source.

Previously, you could delete only the specified records from the repository.

For more information about the repository data deletion job, see the *Informatica MDM - Relate 360 User Guide*.

Matching Null Values

During the linking process, two matched records link to the same cluster. When you set null_ind=2 and if one of the linked records contains a null value, the matching process excludes it from further matching.

Previously, the record with the null value continued to match with other records that contain non-null values and ended up linking irrelevant records to the same cluster.

For more information about configuring the matching rules, see the *Informatica MDM - Relate 360 Installation and Configuration Guide*.

Optimal Preferred Record Search

When you search for preferred records, you can specify whether you want to perform a detailed or optimal search. An optimal search might return fewer matched records but can provide better search performance.

Previously, you could perform only a detailed search.

For more information about the Preferred Record Search web service and command-line operation, see the *Informatica MDM - Relate 360 User Guide*.

Retrieving Records in a Cluster

You can use the Get Cluster web service to retrieve all the records in a cluster based on the cluster identifier.

Previously, you could retrieve all the records based on the primary key column value of a record that is part of the cluster.

For more information about the Get Cluster web service and command-line operation, see the *Informatica MDM - Relate 360 User Guide*.

CHAPTER 3

New Features and Changes (10.0 HotFix 8)

This chapter includes the following topics:

- [New Features \(10.0 HotFix 8\), 13](#)
- [Changes \(10.0 HotFix 8\), 13](#)

New Features (10.0 HotFix 8)

This section describes new features in version 10.0 HotFix 8.

Kerberos Authentication for Streaming Data

If you use streaming data, you can use the Kerberos authentication to secure data.

For more information about configuring the Kerberos authentication for streaming data, see the *Informatica MDM - Relate 360 Installation and Configuration Guide*.

Changes (10.0 HotFix 8)

This section describes changes in version 10.0 HotFix 8.

Resetting the Offset Position for a Spark Instance

Effective in version 10.0 HotFix 8, before you redeploy an existing Spark instance after a failure, run the `kafka-consumer-groups.sh` script to adjust the offset position. The `kafka-consumer-groups.sh` script resets the offset position to the beginning of the topic or to the latest position of the topic for the Spark instance. After you reset the offset position, redeploy the Spark instance.

Previously, before you redeploy an existing Spark instance, you did not require to run the `kafka-consumer-groups.sh` script to adjust the offset position.

Search Client is Deprecated

Effective in version 10.0 HotFix 8, the Relate 360 search client is deprecated. You can use the RESTful web services or the command-line commands to search for records.

CHAPTER 4

New Features, Changes, and Release Tasks (10.0 HotFix 7)

This chapter includes the following topics:

- [New Features \(10.0 HotFix 7\), 15](#)
- [Changes \(10.0 HotFix 7\), 16](#)

New Features (10.0 HotFix 7)

This section describes new features in version 10.0 HotFix 7.

Relationship Graph

You can use Relate 360 to process different types of data, such as customer data, transaction data, and product data. Each type of data can be a business entity type. You can create relationships between the business entity types. The relationships result in a relationship graph that shows how a record is related to other records.

For more information about creating and viewing the relationship graph, see the *Informatica MDM - Relate 360 User Guide*.

Batch Jobs

To create a relationship graph, run the following batch jobs based on the type of data you process:

- Load pairs job
- Create relationship job

For more information about these batch jobs and their prerequisites, see the *Informatica MDM - Relate 360 User Guide*.

RESTful Web Services

After you create relationships between the business entity types, you can retrieve details about the business entity types, their relationships, and the relationship graph. You can also create or remove the relationship between two records.

To retrieve the relationship details and manage the relationship between two records, run the following RESTful web services:

- Get Graph Metadata
- Get Entity Metadata
- Get Entity Relationship
- Get All Relationships
- Get Entity Details
- Create Relationship
- Remove Relationship

For more information about these RESTful web services, see the *Informatica MDM - Relate 360 User Guide*.

Changes (10.0 HotFix 7)

This section describes changes in version 10.0 HotFix 7.

Version Number of the RESTful Web Services

Effective in version 10.0 HotFix 7, all the RESTful web services use v4.0 as the version number in the URLs.

The following sample URL uses v4.0 as the version number:

```
http://BDRMServer:8080/MDMBDRMCustomer360/v4.0/Customer360/MULTISEARCH
```

Previously, all the RESTful web services used v3.0 as the version number in the URLs.

CHAPTER 5

New Features, Changes, and Release Tasks (10.0 HotFix 6)

This chapter includes the following topics:

- [New Features \(10.0 HotFix 6\), 17](#)
- [Changes \(10.0 HotFix 6\), 18](#)

New Features (10.0 HotFix 6)

This section describes new features in version 10.0 HotFix 6.

Consolidation Process

Use the consolidation process to merge data from different records in a cluster to create a preferred or golden record based on the consolidation rules. The input data can be batch or streaming data.

For more information about the consolidation process, see the *Informatica MDM Big Data Relationship Management User Guide*.

Consolidation Rules

The consolidation process uses the consolidation rules to create a preferred record for each cluster. You must define the consolidation rules in the consolidation rules file. You can use the GETSTRATEGIES web service or command-line operation to get a list of supported consolidation rules.

For more information about the consolidation rules, see the *Informatica MDM Big Data Relationship Management Installation and Configuration Guide*.

Retrieving the Preferred Records

After you consolidate the input data, use the GETPREFERREDRECORD or PREFERREDRECORDSEARCH web service or command-line operation to retrieve the preferred records of the clusters.

For more information about the GETPREFERREDRECORD and PREFERREDRECORDSEARCH web services or command-line operations, see the *Informatica MDM Big Data Relationship Management User Guide*.

Hive Enabler Job

You can use the Hive enabler job to load the consolidated data from HDFS or repository to a Hive table.

For more information about the Hive enabler job, see the *Informatica MDM Big Data Relationship Management User Guide*.

Changes (10.0 HotFix 6)

This section describes changes in version 10.0 HotFix 6.

Version Number of the RESTful Web Services

Effective in version 10.0 HotFix 6, all the RESTful web services use v3.0 as the version number in the URLs.

The following sample URLs use v3.0 as the version number:

```
[
  "http://BDRMServer:8080/MDMBDRMSYS/v3.0/SYS/SEARCH",
  "http://BDRMServer:8080/MDMBDRMSYS/v3.0/SYS/GETCLUSTER",
  "http://BDRMServer:8080/MDMBDRMSYS/v3.0/SYS/GETSEARCHLAYOUT",
  "http://BDRMServer:8080/MDMBDRMSYS/v3.0/SYS/GETCLUSTERLAYOUT",
  "http://BDRMServer:8080/MDMBDRMSYS/v3.0/SYS/GETMULTISEARCHLAYOUT",
  "http://BDRMServer:8080/MDMBDRMSYS/v3.0/SYS/MULTISEARCH",
  "http://BDRMServer:8080/MDMBDRMSYS/v3.0/SYS/RULESEARCH",
  "http://BDRMServer:8080/MDMBDRMSYS/v3.0/SYS/INGEST",
  "http://BDRMServer:8080/MDMBDRMSYS/v3.0/SYS/GETINGESTLAYOUT",
  "http://BDRMServer:8080/MDMBDRMSYS/v3.0/SYS/GETRECORDLAYOUT",
  "http://BDRMServer:8080/MDMBDRMSYS/v3.0/SYS/GETRECORD",
  "http://BDRMServer:8080/MDMBDRMSYS/v3.0/SYS/GETMANAGECLUSTERLAYOUT",
  "http://BDRMServer:8080/MDMBDRMSYS/v3.0/SYS/MANAGECLUSTER",
  "http://BDRMServer:8080/MDMBDRMSYS/v3.0/SYS/DELETERECORD",
  "http://BDRMServer:8080/MDMBDRMSYS/v3.0/SYS/AUTHENTICATE",
  "http://BDRMServer:8080/MDMBDRMSYS/v3.0/SYS/GETPREFERREDRECORDLAYOUT",
  "http://BDRMServer:8080/MDMBDRMSYS/v3.0/SYS/GETPREFERREDRECORD",
  "http://BDRMServer:8080/MDMBDRMSYS/v3.0/SYS/GETSTRATEGIES",
  "http://BDRMServer:8080/MDMBDRMSYS/v3.0/SYS/PREFERREDRECORDSEARCH"
]
```

Previously, all the RESTful web services used v2.0 as the version number in the URLs.

CHAPTER 6

New Features, Changes, and Release Tasks (10.0 HotFix 5)

This chapter includes the following topics:

- [New Features \(10.0 HotFix 5\), 19](#)
- [Changes \(10.0 HotFix 5\), 20](#)

New Features (10.0 HotFix 5)

This section describes new features in version 10.0 HotFix 5.

Pagination Support

When you perform a search, you can enable pagination for the search results and specify the maximum number of results to return. If you enable pagination, a search request returns a token with the search results. You can use the token in the subsequent requests to get the search results from cache to avoid performing the search again.

For more information about pagination, see the *Informatica MDM Big Data Relationship Management User Guide*.

Distributed Search

Use distributed search to distribute a search request to multiple region servers of the repository and perform the search on the region servers. Distributed search optimally uses the available resources and results in improved search performance.

For more information about distributed search, see the *Informatica MDM Big Data Relationship Management Installation and Configuration Guide*.

Security for the RESTful Web Services

You can configure security for the RESTful web services to prevent any unauthorized access to the data. MDM Big Data Relationship Management uses the OpenAM-based authentication system to secure the web services.

For more information about securing the RESTful web services, see the *Informatica MDM Big Data Relationship Management Installation and Configuration Guide*.

Processing Data in Motion

You can link or tokenize the input data in motion. The input data in motion indicates streaming data. Use the JSON format to stream the input data. MDM Big Data Relationship Management processes the input data and persists the linked or tokenized data in the repository.

For more information about processing data in motion, see the *Informatica MDM Big Data Relationship Management Installation and Configuration Guide* and *Informatica MDM Big Data Relationship Management User Guide*.

Region Splitter Job

Use the region splitter job to uniformly distribute the linked or tokenized data across all the regions in the repository. The uniform distribution of the linked data optimally uses the resources and improves the search performance. Run the region splitter job before you run the load clustering job for the first time. A load clustering job can use the output files of a region splitter job to distribute the linked data.

For more information about the region splitter job, see the *Informatica MDM Big Data Relationship Management User Guide*.

Post-Clustering Job

Use the post-clustering job to read the output files of an initial clustering job in HDFS and process the input data based on the mode that you configure.

You can run the post-clustering job in the following modes:

- Skip
- Recluster
- Longtail
- Export

For more information about the post-clustering job, see the *Informatica MDM Big Data Relationship Management User Guide*.

Changes (10.0 HotFix 5)

This section describes changes in version 10.0 HotFix 5.

Updating Linked Data in Hive

Effective in version 10.0 HotFix 5, if you persist the linked data in HDFS, you can update the linked data in Hive with incremental data. If you persist the linked data in a repository, you can link a Hive table to the repository table so that the Hive table accesses the repository table for the linked data.

Previously, you could not update the linked data in Hive.

License File

Effective in version 10.0 HotFix 5, you do not require a license file when you install MDM Big Data Relationship Management.

Previously, when you installed MDM Big Data Relationship Management, you had to copy the license file to the installation directory.

Version Number of the RESTful Web Services

Effective in version 10.0 HotFix 5, all the RESTful web services use v2.0 as the version number in the URLs.

The following sample URLs use v2.0 as the version number:

```
"http://BDRMServer:8080/MDMBDRMSYS/v2.0/SYS/SEARCH",
"http://BDRMServer:8080/MDMBDRMSYS/v2.0/SYS/GETCLUSTER",
"http://BDRMServer:8080/MDMBDRMSYS/v2.0/SYS/GETSEARCHLAYOUT",
"http://BDRMServer:8080/MDMBDRMSYS/v2.0/SYS/GETCLUSTERLAYOUT",
"http://BDRMServer:8080/MDMBDRMSYS/v2.0/SYS/GETMULTISEARCHLAYOUT",
"http://BDRMServer:8080/MDMBDRMSYS/v2.0/SYS/MULTISEARCH",
"http://BDRMServer:8080/MDMBDRMSYS/v2.0/SYS/RULESEARCH",
"http://BDRMServer:8080/MDMBDRMSYS/v2.0/SYS/INGEST",
"http://BDRMServer:8080/MDMBDRMSYS/v2.0/SYS/GETINGESTLAYOUT",
"http://BDRMServer:8080/MDMBDRMSYS/v2.0/SYS/GETRECORDLAYOUT",
"http://BDRMServer:8080/MDMBDRMSYS/v2.0/SYS/GETRECORD",
"http://BDRMServer:8080/MDMBDRMSYS/v2.0/SYS/GETMANAGECLUSTERLAYOUT",
"http://BDRMServer:8080/MDMBDRMSYS/v2.0/SYS/MANAGECLUSTER",
"http://BDRMServer:8080/MDMBDRMSYS/v2.0/SYS/DELETERECORD",
"http://BDRMServer:8080/MDMBDRMSYS/v2.0/SYS/AUTHENTICATE"
```

Previously, all the RESTful web services used v1.0 as the version number in the URLs.

Deprecated Batch Jobs and Operations

Effective in version 10.0 HotFix 5, the following batch jobs and search operations that use simple matching are deprecated, and Informatica will drop support for them in the next major release:

- Initial linking job
- Initial loading job
- Incremental linking job
- Update linking job
- SEARCH operation
- RULESEARCH operation

Informatica recommends that you use initial clustering job, load clustering jobs, and MULTISEARCH operation instead.

CHAPTER 7

New Features, Changes, and Release Tasks (10.0.0 HotFix 4)

This chapter includes the following topics:

- [New Features \(10.0.0 HotFix 4\), 22](#)
- [Changes \(10.0.0 HotFix 4\), 22](#)

New Features (10.0.0 HotFix 4)

This section describes new features in version 10.0.0 HotFix 4.

Support for the Kerberos Authentication in Cloudera

In Cloudera, you can use the Kerberos authentication for the MapReduce jobs to access data in HBase and Hive. To use the Kerberos authentication, you must configure the parameters related to Kerberos in the configuration file.

For more information about the parameters related to the Kerberos authentication, see the *Informatica MDM Big Data Relationship Management User Guide*.

Debug Mode for the MULTISEARCH Operation

You can perform the MULTISEARCH operation in the debug mode. The debug mode returns performance metrics that you can use for troubleshooting purposes.

For more information about configuring the debug mode, see the *Informatica MDM Big Data Relationship Management User Guide*.

Changes (10.0.0 HotFix 4)

This section describes changes in version 10.0.0 HotFix 4.

Search

Effective in version 10.0.0 HotFix 4, the MULTISEARCH and RULESEARCH operations use multiple threads to perform parallel processing for matching. The parallel processing improves the search performance.

Previously, the MULTISEARCH and RULESEARCH operations used a single thread to perform matching.

CHAPTER 8

New Features, Changes, and Release Tasks (10.0.0 HotFix 3)

This chapter includes the following topics:

- [New Features \(10.0.0 HotFix 3\), 24](#)
- [Changes \(10.0.0 HotFix 3\), 26](#)

New Features (10.0.0 HotFix 3)

This section describes new features in version 10.0.0 HotFix 3.

Tokenization

Use the tokenization process to create match tokens for the input data based on the columns that you configure as index fields in the matching rules file. You can persist the tokenized data in HDFS or in a repository, and you can search the tokenized data for matching records.

For more information about the tokenization process, see the *Informatica MDM Big Data Relationship Management User Guide*.

MapReduce Jobs

This section describes new features and enhancements to the MapReduce jobs.

HDFS Tokenization Job

The HDFS tokenization job creates match tokens for the input data based on the columns that you configure as index fields in the matching rules file. You can also run the HDFS tokenization job in the incremental mode to update the tokenized data that a HDFS tokenization job creates.

For more information about the HDFS tokenization job, see the *Informatica MDM Big Data Relationship Management User Guide*.

Repository Tokenization Job

The repository tokenization job creates match tokens for the input data in HDFS and loads the tokenized data into the repository. The repository tokenization job uses the columns that you configure as index fields to

generate the match tokens. The repository tokenization job performs the tasks of a HDFS tokenization job and a load clustering job.

For more information about the repository tokenization job, see the *Informatica MDM Big Data Relationship Management User Guide*.

HDFS Batch Search Job

The HDFS batch search job identifies the matching records for the input data in the output files of a HDFS tokenization job based on the match tokens.

For more information about the HDFS batch search job, see the *Informatica MDM Big Data Relationship Management User Guide*.

Repository Batch Search Job

The repository batch search job identifies the matching records for the input data in the repository based on the match tokens.

For more information about the repository batch search job, see the *Informatica MDM Big Data Relationship Management User Guide*.

Repository Update Job

The repository update job updates the tokenized data in the repository with the input data and creates match tokens for the input data in the repository.

For more information about the repository update job, see the *Informatica MDM Big Data Relationship Management User Guide*.

Search

This section describes new features and enhancements to the search process.

RULESEARCH Operation

The RULESEARCH operation generates key ranges, compares the key ranges with the key ranges in the repository data, and gets the matching records based on the rules that you configure in the matching rules file.

You can use the RULESEARCH operation to search the repository data if you define a single index in the matching rules file and use simple matching to link the records. Use the REST API or the `run_client.sh` script in the command line to perform the RULESEARCH operation.

For more information about the RULESEARCH operation, see the *Informatica MDM Big Data Relationship Management User Guide*.

Sorting Search Results

When you perform the MULTISEARCH or RULESEARCH operation in the command line, you can sort the search results based on a column that you define in the `PZMAP` section of the configuration file. You can also sort the search results based on the score of the search results.

For more information about sorting the search results, see the *Informatica MDM Big Data Relationship Management User Guide*.

Changes (10.0.0 HotFix 3)

This section describes changes in version 10.0.0 HotFix 3.

Matching Rules File

This section describes changes to the matching rules file in version 10.0.0 HotFix 3.

Multiple Populations

Effective in version 10.0.0 HotFix 3, you can define matching rules for multiple populations in the matching rules file.

Previously, you could define matching rules only for a single population in the matching rules file.

For more information about configuring matching rules, see the *Informatica MDM Big Data Relationship Management User Guide*.

Index Fields

Effective in version 10.0.0 HotFix 3, you can define multiple columns as index fields for an index.

Previously, you could define only one column for an index. To use multiple columns as index fields, you had to define multiple indexes.

For more information about configuring indexes in the matching rules file, see the *Informatica MDM Big Data Relationship Management User Guide*.

CHAPTER 9

New Features, Changes, and Release Tasks (10.0.0 HotFix 2)

This chapter includes the following topics:

- [New Features \(10.0.0 HotFix 2\), 27](#)
- [Changes \(10.0.0 HotFix 2\), 28](#)

New Features (10.0.0 HotFix 2)

This section describes new features in version 10.0.0 HotFix 2.

Advanced Matching

Use advanced matching to index multiple columns and configure multiple matching rules for each index. You can define indexes and matching rules in the matching rules file. You can use the matching rules file when you run the MapReduce jobs.

For more information about advanced matching, see the *Informatica MDM Big Data Relationship Management User Guide*.

MapReduce Jobs

This section describes new features and enhancements to the MapReduce jobs.

Initial Clustering Job

An initial clustering job indexes and links the input data based on the rules in the matching rules file and the parameters in the configuration file. The initial clustering job reads the input data in HDFS and persists the indexed and linked data to HDFS. You can also run the initial clustering job in the incremental mode to incrementally update the indexed and linked data in HDFS.

For more information about the initial clustering job, see the *Informatica MDM Big Data Relationship Management User Guide*.

Load Clustering Job

A load clustering job loads the integrated and linked data from HDFS into a repository. Before you run the load clustering job, you must run the initial clustering job in the initial or incremental mode.

For more information about the load clustering job, see the *Informatica MDM Big Data Relationship Management User Guide*.

Delete Clustering Job

A delete clustering job matches the input data in HDFS with the indexed and linked data in HDFS that the initial clustering job creates. The job deletes the matching records from the indexed and linked data in HDFS.

For more information about the delete clustering job, see the *Informatica MDM Big Data Relationship Management User Guide*.

Hive Enabler

If you persist the indexed and linked data in HDFS, you can use the Hive enabler job to load the indexed and linked data into the Hive database.

For more information about the Hive enabler job, see the *Informatica MDM Big Data Relationship Management User Guide*.

Search

This section describes new features and enhancements to the search process.

GETMULTISEARCHLAYOUT Operation

The GETMULTISEARCHLAYOUT operation gets the search layout for the matching rules file. The layout contains the required fields that you can specify as input parameters in the JSON format to perform the MULTISEARCH operation. Use the REST API or the `run_client.sh` script in the command line to perform the GETMULTISEARCHLAYOUT operation.

For more information about the GETMULTISEARCHLAYOUT operation, see the *Informatica MDM Big Data Relationship Management User Guide*.

MULTISEARCH Operation

The MULTISEARCH operation generates keys and keys ranges, compares the key ranges with the key ranges in the repository data, and gets the matching records based on the searching and matching criteria that you specify in the matching rules file. Use the REST API or the `run_client.sh` script in the command line to perform the MULTISEARCH operation.

For more information about the MULTISEARCH operation, see the *Informatica MDM Big Data Relationship Management User Guide*.

Changes (10.0.0 HotFix 2)

This section describes changes in version 10.0.0 HotFix 2.

Simple Matching

Effective in version 10.0.0 HotFix 2, you can define multiple matching rules for an index in the matching rules file to perform simple matching. You can use the matching rules file when you run the MapReduce jobs.

Previously, you could define only a single matching rule for an index in the configuration file.

For more information about defining multiple matching rules, see the *Informatica MDM Big Data Relationship Management User Guide*.

Updating Indexed and Linked Data

Effective in version 10.0.0 HotFix 2, you can incrementally update the indexed and linked data in HDFS.

Previously, you could incrementally update the indexed and linked data in a repository but not in HDFS.