



# Maintaining Sun Master Indexes (Repository)



Sun Microsystems, Inc.  
4150 Network Circle  
Santa Clara, CA 95054  
U.S.A.

Part No: 820-3408-15  
September 2008

Copyright 2008 Sun Microsystems, Inc. 4150 Network Circle, Santa Clara, CA 95054 U.S.A. All rights reserved.

Sun Microsystems, Inc. has intellectual property rights relating to technology embodied in the product that is described in this document. In particular, and without limitation, these intellectual property rights may include one or more U.S. patents or pending patent applications in the U.S. and in other countries.

U.S. Government Rights – Commercial software. Government users are subject to the Sun Microsystems, Inc. standard license agreement and applicable provisions of the FAR and its supplements.

This distribution may include materials developed by third parties.

Parts of the product may be derived from Berkeley BSD systems, licensed from the University of California. UNIX is a registered trademark in the U.S. and other countries, exclusively licensed through X/Open Company, Ltd.

Sun, Sun Microsystems, the Sun logo, the Solaris logo, the Java Coffee Cup logo, docs.sun.com, Java, and Solaris are trademarks or registered trademarks of Sun Microsystems, Inc. or its subsidiaries in the U.S. and other countries. All SPARC trademarks are used under license and are trademarks or registered trademarks of SPARC International, Inc. in the U.S. and other countries. Products bearing SPARC trademarks are based upon an architecture developed by Sun Microsystems, Inc.

The OPEN LOOK and Sun<sup>TM</sup> Graphical User Interface was developed by Sun Microsystems, Inc. for its users and licensees. Sun acknowledges the pioneering efforts of Xerox in researching and developing the concept of visual or graphical user interfaces for the computer industry. Sun holds a non-exclusive license from Xerox to the Xerox Graphical User Interface, which license also covers Sun's licensees who implement OPEN LOOK GUIs and otherwise comply with Sun's written license agreements.

Products covered by and information contained in this publication are controlled by U.S. Export Control laws and may be subject to the export or import laws in other countries. Nuclear, missile, chemical or biological weapons or nuclear maritime end uses or end users, whether direct or indirect, are strictly prohibited. Export or reexport to countries subject to U.S. embargo or to entities identified on U.S. export exclusion lists, including, but not limited to, the denied persons and specially designated nationals lists is strictly prohibited.

DOCUMENTATION IS PROVIDED "AS IS" AND ALL EXPRESS OR IMPLIED CONDITIONS, REPRESENTATIONS AND WARRANTIES, INCLUDING ANY IMPLIED WARRANTY OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE OR NON-INFRINGEMENT, ARE DISCLAIMED, EXCEPT TO THE EXTENT THAT SUCH DISCLAIMERS ARE HELD TO BE LEGALLY INVALID.

# Contents

---

|   |    |
|---|----|
| <b>Maintaining Sun Master Indexes (Repository)</b> .....              | 5  |
| Related Topics .....  | 6  |
| Defining Master Index Security (Repository) .....                     | 6  |
| ▼ To Create a Master Index User Account .....                         | 6  |
| Master Index User Roles (Repository) .....                            | 7  |
| Learning About Master Index Reports (Repository) .....                | 9  |
| Master Index Command Line Reports (Repository) .....                  | 9  |
| Master Index Report Configuration (Repository) .....                  | 10 |
| Creating Custom Master Index Reports (Repository) .....               | 11 |
| Masked Data in Master Index Reports (Repository) .....                | 11 |
| Master Index Production Reports (Repository) .....                    | 11 |
| Master Index Activity Reports (Repository) .....                      | 13 |
| Master Index Database Indexes (Repository) .....                      | 14 |
| Working With Master Index Command Line Reports (Repository) .....     | 14 |
| Configuring the Master Index Report Environment (Repository) .....    | 15 |
| Configuring Master Index Command Line Reports (Repository) .....      | 16 |
| Master Index Command Line Report Properties (Repository) .....        | 18 |
| Running Master Index Command Line Reports (Repository) .....          | 20 |
| Maintaining the Master Index Database (Repository) .....              | 21 |
| Backing up the Master Index Database .....                            | 21 |
| Restoring the Master Index Database .....                             | 22 |
| Archiving Master Index Data .....                                     | 22 |
| Implementing Changes to the Master Index Project (Repository) .....   | 22 |
| Modifying Master Index Configuration Files (Repository) .....         | 23 |
| Modifying Standard Master Index Project Components (Repository) ..... | 24 |
| Modifying the Master Index Database (Repository) .....                | 24 |
| Modifying Master Index Security (Repository) .....                    | 24 |
| Modifying the Local ID Format (Repository) .....                      | 24 |



# Maintaining Sun Master Indexes (Repository)

---

The topics listed here provide procedures, conceptual information, and reference information for maintaining a Sun Master Index application.

Note that Java CAPS includes two versions of Sun Master Index. Sun Master Index (Repository) is installed in the Java CAPS repository and provides all the functionality of previous versions in the new Java CAPS environment. Sun Master Index is a service-enabled version of the master index that is installed directly into NetBeans. It includes all of the features of Sun Master Index (Repository) plus several new features, like data analysis, data cleansing, data loading, and an improved Data Manager GUI. Both products are components of the Sun Master Data Management (MDM) Suite. This document relates to Sun Master Index (Repository) only.

## What You Need to Know

These topics provide information you should know about maintenance tasks.

- [“Master Index Command Line Reports \(Repository\)”](#) on page 9
- [“Master Index Production Reports \(Repository\)”](#) on page 11
- [“Master Index Activity Reports \(Repository\)”](#) on page 13
- [“Master Index Database Indexes \(Repository\)”](#) on page 14
- [“Backing up the Master Index Database”](#) on page 21
- [“Restoring the Master Index Database”](#) on page 22
- [“Archiving Master Index Data”](#) on page 22
- [“Implementing Changes to the Master Index Project \(Repository\)”](#) on page 22

## What You Need to Do

These topics provide instructions on how to perform maintenance tasks.

- [“Defining Master Index Security \(Repository\)”](#) on page 6
- [“Configuring Master Index Command Line Reports \(Repository\)”](#) on page 16
- [“Running Master Index Command Line Reports \(Repository\)”](#) on page 20

## More Information

These topics provide additional information you should know when maintaining a master index application.

- [“Master Index User Roles \(Repository\)” on page 7](#)
- [“Master Index Command Line Report Properties \(Repository\)” on page 18](#)

## Related Topics

Several topics provide information and instructions for implementing and using a Repository-based master index application. For a complete list of topics related to working with Sun Master Index (Repository), see [“Related Topics” in \*Developing Sun Master Indexes \(Repository\)\*](#).

## Defining Master Index Security (Repository)

Sun Master Index supports security at the user and function level and also supports Secure Sockets Layer (SSL) authentication. A secure user name and password must be defined for each master index application user to connect to the database and to log on to the Enterprise Data Manager (EDM). For each user account you define, you must specify one or more roles in order for that user to be able to perform any functions in the EDM.

In order for security roles to function correctly, authorization security must be enabled in the Enterprise Data Manager file. To enable security, set the *enable-security* element to “true”. By default, this element is set to “false” (the default is “true” for Sun Master Patient Index).

Security for master index applications running on the Sun Java System Application Server is configured using the Admin Console. You can also define security using a Lightweight Directory Access Protocol (LDAP) server, using the roles defined in [“Master Index User Roles \(Repository\)” on page 7](#).

### ▼ To Create a Master Index User Account

- 1 Log on to the Sun Java System Application Server Admin Console.
- 2 In the left portion of the page, expand Configuration, expand Security, expand Realms, and then select File.
- 3 On the Edit Realm page, select Manage Users.
- 4 On the File Users page, select New.
- 5 In the User ID field, enter a name for the user.

- 6 In the **Group List** field, enter one or more of the user roles listed in **“Master Index User Roles (Repository)” on page 7**, separating multiple groups with a comma.
- 7 After you have added all required user roles, enter a password for the user in the **New Password** field.
- 8 In the **Confirm New Password** field, enter the password again.
- 9 Click **OK**.

## Master Index User Roles (Repository)

At a minimum, each user must be assigned to the `eView.Admin` role, or must be assigned to the `eView.User` role and the role that provides access to the initial page as described below (the initial page can be configured in the Enterprise Data Manager file).

- If the Search page is the initial page, then users must be assigned to the `EO.SearchAndViewSBR` role.
- If the Matching Review page is the initial page, then users must be assigned to the `Duplicate.SearchAndView` role.
- If the History page is the initial page, then users must be assigned to the `History.SearchAndView` role.
- If the Create EO page is the initial page, then users must be assigned to the `EO.Create` role.
- If the Reports page is the initial page, then users must be assigned to the `eView.Reports` role.

The user role names listed below are case-sensitive.

**TABLE 1** User Roles and Descriptions

| User Role                          | Description  |
|------------------------------------|--|
| <code>AL.View</code>               | Gives access permission to search for and view audit log entries, and to generate and print the search results report.                           |
| <code>Duplicate.All</code>         | Gives access permission to all potential duplicate functions.  |
| <code>Duplicate.AutoResolve</code> | Gives access permission to permanently resolve potential duplicate records. This permission also requires <code>Duplicate.SearchAndView</code> . |
| <code>Duplicate.Print</code>       | Reserved for future functionality.   |
| <code>Duplicate.Resolve</code>     | Gives access permission to resolve potential duplicate records. This permission also requires <code>Duplicate.SearchAndView</code> .             |

TABLE 1 User Roles and Descriptions (Continued)

| User Role               | Description  |
|-------------------------|--|
| Duplicate.SearchAndView | Gives access permission to search for and view potential duplicate records, and to view and print the potential duplicate search results report.   |
| Duplicate.Unresolve     | Gives access permission to unresolve potential duplicate records that were previously resolved. This permission also requires Duplicate.SearchAndView.   |
| EO.All                  | Gives access permission to all enterprise object functions described below.  |
| EO.Activate             | Gives access permission to activate enterprise records.  |
| EO.Create               | Gives access permission to create new enterprise records.  |
| EO.Compare              | Gives access permission to compare enterprise records.   |
| EO.Deactivate           | Gives access permission to deactivate enterprise records.  |
| EO.Edit                 | Gives access permission to modify the SBR in enterprise records.   |
| EO.Merge                | Gives access permission to merge enterprise records.   |
| EO.OverwriteSBR         | Gives access permission to modify the SBR and to lock SBR fields for overwrite.  |
| EO.PrintComparison      | Reserved for future functionality.   |
| EO.PrintSBR             | Reserved for future functionality.   |
| EO.SearchAndViewSBR     | Gives access permission to search for and view single best records, and to generate and print the search results report. This group must be assigned to each user except those assigned the eView.Admin group. |
| EO.Unmerge              | Gives access permission to unmerge enterprise records.   |
| EO.ViewMergeTree        | Gives access permission to view a merge history of an enterprise object.   |
| eView.Admin             | Gives access permission to all functions of the Enterprise Data Manager.   |
| eView.Reports           | Gives access permission to generate and view reports. (Note that this is not required to print search results reports, which is granted by the individual search access permissions.)                          |
| eView.User              | Gives access to the EDM. This group must be assigned to each user except those assigned the eView.Admin group.   |
| eView.VIP               | Gives permission to view fields masked by any custom masking logic specified by the Enterprise Data Manager file.  |
| History.All             | Gives access permission to all history functions described below.  |
| History.Print           | Reserved for future functionality.   |
| History.SearchAndView   | Gives access permission to search for and view the transaction history of enterprise records and to generate and print the search results report.  |



TABLE 1 User Roles and Descriptions (Continued)

| User Role  | Description   |
|------------|---|
| SO.All     | Gives access permission to all system record functions described below. |
| SO.Add     | Gives access permission to add system records.                          |
| SO.Edit    | Gives access permission to modify system records.                       |
| SO.Merge   | Gives access permission to merge system records.                        |
| SO.Print   | Reserved for future functionality.                                      |
| SO.Remove  | Gives access permission to delete system records.                       |
| SO.Unmerge | Gives access permission to unmerge system records.                      |
| SO.View    | Gives access permission to view system records.                         |

## Learning About Master Index Reports (Repository)

Several standard reports are provided with master index applications that allow you to monitor and review the state of the information in the master index database. You can either run these reports through the EDM or from a command line. The following topics provide an overview of each report.

- [“Master Index Command Line Reports \(Repository\)”](#) on page 9
- [“Master Index Report Configuration \(Repository\)”](#) on page 10
- [“Creating Custom Master Index Reports \(Repository\)”](#) on page 11
- [“Masked Data in Master Index Reports \(Repository\)”](#) on page 11
- [“Master Index Production Reports \(Repository\)”](#) on page 11
- [“Master Index Activity Reports \(Repository\)”](#) on page 13
- [“Master Index Database Indexes \(Repository\)”](#) on page 14

### Master Index Command Line Reports (Repository)

Sun Master Index provides a set of production and activity reports that can be generated from a command line or from the EDM. You need to download the report client separately using the Java CAPS Uploader. This is described in [“To Download and Expand ZIP Files Using the Java CAPS Uploader”](#) in *Using the Java CAPS 6 Installation GUI*.

The production reports provide information about transactional changes to the data in the master index application and about the current state of that data, helping you monitor stored data and determine how that data needs to be updated. This information also helps verify that the matching logic and weight thresholds are defined correctly. Activity reports provide statistical information for transactions over specific periods of time.

In order to run the command line reports, you must have the Java Runtime Environment (JRE) 1.5.13 or later installed on the machine where the report files reside. For additional reporting needs, the database is accessible using any commercially available ODBC-compliant reporting tool. You can also define reports using Java, PL/SQL, or SQL.

## About Production Reports

Production reports should be run daily and provide information about the transactions that are processed through the master index database. These reports provide lists of potential duplicate records, merge transactions, unmerge transactions, assumed matches, updates, and deactivated records for a specified time period. The information you find in these reports helps you analyze your matching threshold configuration, and provides valuable information about how data is being processed with your current configuration. In addition to running the production reports daily, you should run them against any data that has been loaded from existing systems into the master index database in batch format.

## About Activity Reports

Activity reports should be run weekly, monthly, and yearly to obtain statistical data about the transactions that are processed through the master index database. These reports give the number of each type of transaction performed for the specified week, month, or year. They also provide cumulative information for the week, month, or year to date. The information you find in these reports helps analyze the matching threshold configuration and the condition of your data by giving you the number of potential duplicates created, the number of assumed matches, and so on.

## Master Index Report Configuration (Repository)

The reports are configured by XML files. For the command line reports, the configuration files are located in the report home directory in the `config` subdirectory. The file `eViewCompanyReport.xml` provides an example of how the file might be configured for a company object; the file `eIndexPersonReport.xml` provides an example of how the file might be configured for a person object. You can use either file for your reports. When you create a new master index application, you can specify the fields that appear on reports.

The configuration files allow you to specify which reports to run, the time period of the transactions to include in each report, and the name and location of the report files. You can also define various report details, such as the name of each report, which fields to include, and the names and sizes of the report columns. Most of these changes should only need to be made one time, before you first run the reports.

## Creating Custom Master Index Reports (Repository)

If the standard reports do not provide you with all the information you need, you can create custom reports using PL/SQL, SQL, or Java (using the “lookup” methods in the `MasterController` class). You can also access the database using any ODBC-compliant report writer (such as Crystal Reports), providing you with the flexibility to report on any information contained in the master index database.

## Masked Data in Master Index Reports (Repository)

The EDM can be configured to hide certain fields from users who do not have the appropriate security permissions. However, reports will display hidden data if those fields are configured to appear on the reports. Be sure to only give access to users who should be able to view this information, or do not include hidden fields in the reports.

## Master Index Production Reports (Repository)

The standard production reports help you to monitor and analyze the data in the master index database. You can view information about the transactions processed and about any potential duplicates or assumed matches that result from these transactions.

Each report has certain fields that are always displayed and certain fields that are configured to display. You can customize the configured fields that appear on each report as needed. By default, `eViewCompanyReport.xml` configures all reports to include the company name, type, stock symbol, primary contact, street address, city, and telephone number fields. `eIndexPersonReport.xml` configures all reports to include the first name, last name, date of birth, SSN, and address line 1 and 2 fields. The fields that are always displayed are described for each report in the following sections.

Production reports can be run for the current day, the previous day, or for a date range you specify. If you run your daily reports in the evening, you should run the current day's reports. If you run your daily reports in the morning, you should run the previous day's reports.

### Assumed Match Report

This report displays information about any records that were automatically updated by incoming data during the specified time period. The information in this report, in combination with data from the potential duplicate report, helps you determine whether the matching threshold for assumed matches is accurate. You should review this report daily to ensure that no assumed matches were made in error. The master index application provides the ability to undo an assumed match that was made in error.

The assumed match report always includes the following information about the record that was updated: enterprise-wide unique identifier (EUID), system code, local ID, and matching weight. The report provides the same information for the incoming message that updated the existing record with the exception of the EUID. You can configure the report to include any additional fields from the defined object structure in the Object Definition file in the master index project.

## **Deactivated Record Report**

This report displays a list of all enterprise records that were deactivated during the specified time period. This report does not include system records that were deactivated. Review this report daily to ensure that no records were deactivated in error. The master index application provides the ability to reactivate any deactivated record. The deactivated record report always includes the EUID of the deactivated record, and you can configure the report to include any additional fields from the defined object structure in the Object Definition file in the master index project.

## **Potential Duplicate Report**

This report displays information about records that were marked as potential duplicates of one another during the specified time period. The information provided on this report can help you determine whether the matching (or upper) threshold and the duplicate threshold are configured accurately. The information for each record on the potential duplicate report always includes the EUIDs of both records, the system code, and the matching weight between each potential duplicate pair. You can configure the report to include any additional fields from the defined object structure in the Object Definition file in the master index project.

If same system matching is not enabled and two duplicate records from the same system on this report have a matching weight above the match threshold, it is an indication that the records most likely represent the same person. Review the potential duplicate report daily to determine if two records need to be merged or if they can be resolved. Use this report as a work list when working with potential duplicates.

## **Merge Transaction Report**

This report displays a list of all enterprise records that were merged during the specified time period. Review this report daily to ensure that no records were merged in error. The master index application provides the ability to unmerge any merged records. The merge transaction report always includes the EUID of each record affected by the merge. You can also configure the report to include any additional fields from the defined object structure in the Object Definition file in the master index project.

## **UnMerge Transaction Report**

This report displays a list of all enterprise records that were unmerged during the specified time period. This report always includes the EUIDs of both records involved in the unmerge transaction, and you can configure the report to include any additional fields from the defined object structure in the Object Definition file in the master index project.

## Update Report

This report displays records whose information was updated during the specified time period. Review this report daily to verify the updates made in a given day. This report can help explain why a resolved potential duplicate listing was reinstated to the potential duplicate list. The update report always includes the following information about the record that was updated: EUID, system code, and local ID. You can configure the report to include any additional fields from the defined object structure in the Object Definition file in the master index project. The updated fields might not necessarily appear on this report.

## Master Index Activity Reports (Repository)

The activity reports help you to monitor and analyze the transactions in the master index database by providing statistical data about each transaction type. Unlike the production reports, the information displayed on the activity reports is not configurable. The information displayed on these reports is described for each report in the following sections. Activity reports can be run for any week, month, or year you specify.

### Weekly Activity Report

This report displays a summary of transactions that occurred against the database on each day for the specified calendar week (always Sunday through Saturday). The information provided in this summary includes the number of each of the following transactions performed each day.

- Add
- Update
- EUID Deactivate
- EUID Merge
- EUID Unmerge
- LID Merge
- LID Unmerge
- LID Transfer

### Monthly Activity Report

This report displays a summary of transactions that occurred against the database during the specified month. You can run this report for any calendar month. The information provided in this summary includes the number of each of the following transactions that were performed for the month:

- Add
- EUID Deactivate
- EUID Merge
- EUID Unmerge
- LID Merge

- LID Unmerge
- Unresolved Potential Duplicates
- Resolved Potential Duplicates

## Yearly Activity Report

This report displays a summary of transactions that occurred against the database for the specified calendar year. You can run this report for any calendar year. The information provided in this report includes a summary of each transaction listed for the monthly activity report above.

## Master Index Database Indexes (Repository)

Some of the reports you run can grow quite large, impacting the performance of the report client. The following indexes are created in the database to improve performance.

```
CREATE INDEX SBYN_POTENTIALDUPLICATES3 ON SBYN_POTENTIALDUPLICATES
  (TRANSACTIONNUMBER ASC);

CREATE INDEX SBYN_ASSUMEDMATCH2 ON SBYN_ASSUMEDMATCH (TRANSACTIONNUMBER ASC);

CREATE INDEX SBYN_TRANSACTION4 on SBYN_TRANSACTION (EUID2 ASC, TIMESTAMP ASC);

CREATE INDEX SBYN_TRANSACTION3 on SBYN_TRANSACTION (TIMESTAMP ASC,
  TRANSACTIONNUMBER ASC);
```

---

**Note** – These indexes should be removed prior to performing an initial load or batch load of data.

---

## Working With Master Index Command Line Reports (Repository)

The following topics provide information and instructions for configuring and running reports from a command line using a Java command. For information about running the reports from the Enterprise Data Manager, see *Working With the EDM for Sun Master Index*.

- “Configuring the Master Index Report Environment (Repository)” on page 15
- “Configuring Master Index Command Line Reports (Repository)” on page 16
- “Running Master Index Command Line Reports (Repository)” on page 20

Make sure the reports have been installed as described in *Using the Java CAPS 6 Installation GUI*. You must also have the Java 2 Platform, Standard Edition v. 1.5.13 or later installed on the

machine from which the reports are run. Be sure you have configured the database connection for the master index application using the Sun Java System Application Service Admin Console.

## Configuring the Master Index Report Environment (Repository)

Before running the master index reports from a command line, you must configure the report environment. Configuring the environment consists of two steps: copying the generated project files and setting environment variables.

The reports rely on two files, `stc_eindex_client.jar` and `stc_eindex_util.jar`, that are generated in the master index project. You need to export these files to the reports directory. If the project is regenerated at any time, export the files to the reports directory again when the generate process is complete.

### ▼ To Copy the Generated Files

- 1 In the NetBeans Projects window, expand the master index server project.
- 2 Select, and then right click, the *Application\_stc\_eindex\_client.jar* file (where *Application* is the name of the master index application).
- 3 On the context menu that appears, click **Export**.  
The Save dialog box appears.
- 4 In the Save In field, enter, or navigate to, the `lib` subdirectory in the reports home directory.
- 5 Click **Save**.
- 6 Repeat steps 2 through 5 to export the *Application\_stc\_eindex\_util.jar* file (where *Application* is the name of the master index application).
- 7 In Windows Explorer, navigate to the `lib` subdirectory and rename the file *Application\_stc\_eindex\_client.jar* to `stc_eindex_client.jar`, and rename the file *Application\_stc\_eindex\_util.jar* to `stc_eindex_util.jar`.

### ▼ To Set up the Environment

- 1 If you install or move the reports files to a machine other than the application server machine, make sure JRE 1.5.13 or later is installed on the machine where the files reside.
- 2 Set up all Java environment variables as specified in the Java documentation.

- 3 **Create one environment variable, `JAVA_HOME`, and set it to the home directory of the JRE installation.**
- 4 **If you run the reports using the Java command and not the supplied batch file, modify the `CLASSPATH` variable before running the reports for the first time by adding the absolute path and filename of the files in the `lib` subdirectory of the reports home directory to the `CLASSPATH` variable.**

## Configuring Master Index Command Line Reports (Repository)

Before running any reports from the command line, you must customize the XML configuration file. You can use either of the files located in the reports directory in the eView or eIndex subdirectory. A default XML file named `eIndexPersonReport.xml` is defined for a person object and a default XML file named `eViewCompanyReport.xml` is defined for a company object. You can use either of these as a basis for your production configuration file. Report configuration includes two steps: defining the overall report configuration and configuring the individual reports.

### Defining the Command Line Report Configuration

The first section of the report configuration file is indicated by the `DOCTYPE` and the `report` elements and tells the report client how to connect to the application server, which application to run the reports against, and where to output the report files.

---

**Note** – The `DOCTYPE` element indicates the type of document being generated. Do not change this value.

---

### ▼ To Define the Command Line Report Configuration

- 1 **In the `SYSTEM` element, enter the location of the DTD file for the reports.**

By default, this file is named `report.dtd`, and is located in the `config` directory. You should not need to modify this attribute unless you move `report.dtd`
- 2 **In the `appserver` element, enter the IIOP address for the application server.**

This must be in the format `corbaname:iiop:host:port`, where `host` is the name of the server and `port` is the ORB port number.
- 3 **In the `application` element, enter the name of the primary object used by the master index application.**



- 4 In the *output-folder* element, enter the location in which the generated reports will be placed. If an output directory is specified in the command line, that directory overrides the one specified here. If the output directory already exists, the report client issues a warning that any existing report files will be overwritten and gives you the option of cancelling the reports.

## Configuring Command Line Reports

A configuration section is defined for each of the six report templates. Use these sections to configure each report to display information as you want to view it. You can also specify which reports to run.

### ▼ To Configure Command Line Reports

For each report, make the following modifications before running the reports. Each element or attribute mentioned in the following instructions is defined in . There are six stanzas for you to modify, one for each report.

- 1 In the XML file you will use for your implementation, scroll to the *report* element.
- 2 Name the report in the report *name* attribute.
- 3 Specify whether or not to run the report in the *enable* element.
- 4 Define the name of the output file in the *output-file* element.
- 5 Specify a time period for the report by modifying the *type* element and, optionally, the *from-date* and *to-date* elements.
- 6 Define the fields to include on the report by modifying the elements in the *fields* element.
- 7 When you have finished configuring each report, save and close the file.

A sample report configuration appears below.

```
<report name="Potential Duplicate Today"
  template="Potential Duplicate">
  <enable>true</enable>
  <output-file>pot_dup_t.txt</output-file>
  <max-result-size>0</max-result-size>
  <page-size>100</page-size>
  <criteria>
    <dates type="today" from-date="" to-date=""/>
    <status></status>
  </criteria>
  <fields>
    <field path="Person.FirstName" label="First Name" width="10"/>
    <field path="Person.LastName" label="Last Name" width="10"/>
  </fields>
</report>
```

```

<field path="Person.SSN" label="SSN" width="9"/>
<field path="Person.DOB" label="DOB" width="10"/>
<field path="Person.Address.AddressLine1"
  label="AddressLine1" width="30"/>
<field path="Person.Address.AddressLine2"
  label="AddressLine2" width="30"/>
</fields>
</report>

```

## Master Index Command Line Report Properties (Repository)

The following table lists and describes the elements in the report configuration files that define the configuration of each production and activity report.

| Element/Attribute | Description   |
|-------------------|---|
| report            | Defines each report run by the batch file. Each report is defined by a report element.  |
| report/name       | The descriptive name of the report. This can be any string, and appears as the title in the specified report.   |
| report/template   | The template to use for the type of report being generated. You should not need to modify this element, but you can specify any of the following templates. <ul style="list-style-type: none"> <li>■ Assumed Match</li> <li>■ Potential Duplicate</li> <li>■ Deactivated</li> <li>■ Merged</li> <li>■ Unmerged</li> <li>■ Update</li> <li>■ Weekly Activity</li> <li>■ Monthly Activity</li> <li>■ Yearly Activity</li> </ul> |
| enable            | Specifies whether to run the report for the current run. Specify <b>true</b> to run the report; specify <b>false</b> to disable the report. This option allows you to run one report at a time.   |
| output-file       | The name of the file generated by the report client. This file is created in the output directory defined earlier in the file or in the output directory specified in the command line (the command line output directory overrides the configuration file output directory).   |

| Element/Attribute | Description  |
|-------------------|--|
| max-result-size   | The number of records to display on the report. If no value is entered, or if the value is zero (0), the size defaults to 1000 records. To retrieve all records for a report, enter a very large value for this element.   |
| page-size         | The number of records returned to the report generator at one time for each report.  |
| criteria          | Defines the date range for the report.   |
| dates/type        | <p>Indicates the type of date range to use for the report. Specify <b>today</b> to report on transactions with today's date; specify <b>yesterday</b> to report on transactions with yesterday's date; or specify <b>range</b> to enter a specific range of dates. If you specify <b>range</b>, you must enter the date range in the <i>from-date</i> and <i>to-date</i> attributes.</p> <p><b>Note</b> – If you enter a type of today or yesterday and you enter a date range, only the type will be used. For the activity reports, entering today runs the report for the current week, month, or year. Entering yesterday only runs the previous week's report if yesterday was a Saturday.</p>                                    |
| dates/from-date   | <p>The starting date when using a date range for the report. Enter the starting date for the report transactions in <b>YYYYMMDD</b> or <b>YYYYMMDDHHmmss</b> format. If you enter a date in this element, you must enter a later date in the <i>to-date</i> element and specify <b>range</b> in the <i>type</i> element.</p> <p><b>Note</b> – For the activity reports, you can enter the range for the week, month, or year (depending on the type of activity report) on which you want to report. If the dates you specify do not fall within one calendar week, month, or year, the report client creates a report for the calendar week, month, or year containing the <i>from-date</i> and ignores the <i>to-date</i> value.</p> |
| dates/to-date     | The ending date when using a date range for the report. Enter the ending date for the report in <b>YYYYMMDD</b> or <b>YYYYMMDDHHmmss</b> format.   |
| status            | <p>This element is valid for the potential duplicate report only, and indicates the status of the potential duplicate pairs to display on the report. Specify any of the following values:</p> <ul style="list-style-type: none"> <li>■ <b>U</b> - Only unresolved potential duplicates appear on the report.</li> <li>■ <b>A</b> - Only potential duplicates that are permanently resolved (auto-resolved) appear on the report.</li> <li>■ <b>R</b> - Only resolved potential duplicates appear on the report.</li> </ul> <p>Leaving the status blank results in potential duplicates of all statuses appearing on the report.</p>   |
| fields            | A list of fields to display on the report in addition to those that are displayed automatically. This element should be empty for the activity reports. If a list of fields is supplied for these reports, it is ignored.  |

| Element/Attribute | Description   |
|-------------------|---|
| field/path        | The ePath to a field you want to include in the report. For more information about ePaths, see “ <a href="#">Master Index Field Notations</a> ” in <i>Understanding Sun Master Index Configuration Options (Repository)</i> .<br><b>Note</b> – You cannot use the asterisk option in the ePaths you specify here. |
| field/label       | The column label for the specified field in the report.   |
| field/width       | The width of the column for the specified field in the report. If a field value is larger than the width specified, that value will be truncated in the report.   |

## Running Master Index Command Line Reports (Repository)

Once you have configured the reports, you can run them by either running the batch file provided with the reports or using the Java command.



**Caution** – The application server must be running with the master index project deployed and enabled in order to generate command line reports.

### ▼ To Run the Reports Using the Batch File

1 From a command prompt, navigate to the location of the report files.

2 Type the following all on one line:

```
ReportClient.bat -f config_file -d output_directory
```

where *config\_file* is the name of the report configuration file to use, and *output\_directory* is the location to which the reports will be written. This value overwrites the value specified in the configuration file. If this option is not specified, the configuration file value is used.

**Note** – The `ReportClient.bat` file must reside in the reports home directory at the same level as the `lib` and `config` subdirectories in order for the environment variables to be set up correctly.

3 To view the reports, navigate to the location you specified as your output path and open the files in any text editor.

### ▼ To Run the Reports Using a Java Command

#### Before You Begin

Before running the reports for the first time, set up the environment variables as described in “[To Set up the Environment](#)” on page 15.

**1 At the command prompt, type the following all on one line:**

```
java com.stc.eindex.report.ReportClient -f config_file -d output_directory
```

where *config\_file* is the name of the report configuration file to use and *output\_directory* is the location to which the reports will be written. This value overwrites the value specified in the configuration file. If this option is not specified, the configuration file value is used.

---

**Note** – An additional option, -h, can be used to obtain help information for the report client.

---

**2 To view the reports, navigate to the location you specified as your output path and open the files in any text editor.**

## Maintaining the Master Index Database (Repository)

The database requires periodic maintenance tasks, such as backing up information or archiving certain tables. Perform backups regularly, and use the standards and policies of your organization to determine the best methods for backing up data. The following topics provide information about tasks you should perform for standard database maintenance.

- [“Backing up the Master Index Database” on page 21](#)
- [“Restoring the Master Index Database” on page 22](#)
- [“Archiving Master Index Data” on page 22](#)

## Backing up the Master Index Database

The master index database must be backed up on a regular basis. Typically, the database should be backed up once a month or once a quarter, depending on the size of the database and the volume of data being processed. The frequency of your database backups depends on your organization’s internal policies and practices. Use your normal procedures for backing up a high availability database (this procedure should be determined by a database administrator).

### Online Backups

The best practice for backing up the master index database is an online backup during which the database is not shut down. (Note that this does require an offline backup as a starting point to which any online changes can be applied in the event the database must be restored). An online backup will always take a consistent snapshot, though it might not backup all transactions in progress.

Each transaction in the master index application is saved under one commit command, so the state of the database is always consistent when a backup is performed. The history tables always match the transactions in the current tables and no partial transactions are committed. Even if a transaction is underway at the time of the backup, the database is consistent.

For the most reliable backups for Oracle databases, Oracle recommends running the Oracle database in ARCHIVELOG mode. ARCHIVE mode ensures that your database is protected from both instance and media failure and, because all changes made to the database are saved in a redo log, all database updates are available for recovery rather than just the most recent changes. For SQL Server, Microsoft recommends running the SQL Server database using the full recovery model, which allows a database to be recovered to the point of failure. Online backups are available for Oracle and SQL Server databases running in these modes.

## Offline Backups

If needed, you can perform offline backups of the master index database. In this case, you must queue any incoming messages using the JMS IQ Manager and undeploy the master index application before beginning the backup. Once the backup is complete, restart the database, redeploy the master index application, and then process the messages queued by the JMS IQ Manager.

## Restoring the Master Index Database

In the unlikely event that you need to restore the master index database to a previously archived version, you must undeploy the master index application prior to performing the restoration to ensure that the application retrieves the correct sequence numbers from the database once it is restored. Any new transactions that occurred after the archived version was created will be lost, but they can be resent if the JMS IQ Manager is configured to journal all messages.

## Archiving Master Index Data

In addition to regular database backups, some of the master index database tables can grow very large. For performance reasons, you might want to archive the information in the `sbyn_assumedmatch` and the `sbyn_audit` tables.

# Implementing Changes to the Master Index Project (Repository)

After a master index application has been in production, you might need to make changes to your project. For example, if you add a new external system, you need to add that system to the master index database and you might need to modify the object structure and OTDs as well as update the application files. Changes occur as the needs of your end users evolve and as additional external systems are added. Do not make changes to the system hastily. Handle changes using the same change management process that was originally used to deploy your project. Applying this same process of planning, configuration, testing, migration, monitoring, and reevaluation will help ensure successful updates.

- “Modifying Master Index Configuration Files (Repository)” on page 23
- “Modifying Standard Master Index Project Components (Repository)” on page 24
- “Modifying the Master Index Database (Repository)” on page 24
- “Modifying Master Index Security (Repository)” on page 24
- “Modifying the Local ID Format (Repository)” on page 24

## Modifying Master Index Configuration Files (Repository)

Over time, you might need to make changes to your configuration files, such as adding fields or objects to the object structure, changing queries, or fine-tuning the matching process. Whenever you make a change to a master index configuration file, you must disable the master index server project, regenerate the application, and then redeploy the project. If any Java Collaborations in client projects reference the master index application, you must re-import the regenerated .jar files from the master index server project into the Java Collaboration.

This section provides tips for updating components of the configuration files. In order for any of these changes to take affect, you must regenerate the application and rebuild and redeploy the project.

### Updating the Object Structure

If you make any changes to the object structure, keep the following in mind.

- If you want the new fields or objects to appear on the EDM, make sure to add them to the first section of the Enterprise Data Manager file and to any of the page definitions later in the file (this includes search pages).
- If the new fields require normalization, parsing, or phonetic encoding, define the new structures in the Match Field file.
- If a new field will be used for matching, add it to the blocking query used for match processing as well as to the match string in the Match Field file.
- If the new fields or objects will be included in incoming messages, add them to the inbound OTD structure (the outbound OTD will be updated when you regenerate the application).

### Updating Normalization and Standardization Structures

If you define normalization, standardization, or phonetic encoding for fields that are not currently defined in the Match Field file, or if you change existing standardization structures, make sure to do the following.

- Use the appropriate standardization type, domain selector, and field IDs.
- Add the new fields that will store the standardized versions of the original field value to the appropriate objects in the Object Definition file.

- Add new columns to the database to store the standardized field values.

## Updating the Match String

If you make changes to the match string, update the database indexes and the blocking query in the Candidate Select file accordingly. For example, if you remove a field from the match string, you might also want to remove that field from the blocking query and database indexes. If you add a field to the match string, add the field to the blocking query and to the appropriate database index to maintain performance.

## Modifying Standard Master Index Project Components (Repository)

Whenever changes are made to components of a master index project outside of the master index application, such as modifying the Connectivity Map, eWay properties, Collaborations, or OTDs, the master index application does not need to be regenerated; however the project containing the changed components must be redeployed in order for the changes to take effect.

## Modifying the Master Index Database (Repository)

There might be times when you need to modify the master index database. For example, you might need to add or modify a stored procedure or index, or you might need to add new external systems. You must modify the database if you add fields or objects to the object structure in order to reflect the new structure in the database tables. If you make changes to the database, rebuild and redeploy the master index server project to ensure the changes are picked up by the application.

## Modifying Master Index Security (Repository)

You can define new users for the database at any time using standard SQL statements to create the type of user you want to define. You can also add new users for the the Enterprise Data Manager file through the Sun Java System Application Server. Neither of these procedures require any stoppage of the database or of the master index application and redeployment is not required.

## Modifying the Local ID Format (Repository)

If you need to modify the local ID format for an external system, regenerate the application after you make the changes and then redeploy the project. Any Collaborations that reference the master index application must also be recompiled and those projects redeployed. If you extend



the length of a local ID past 20 characters, make sure to increase the length of any database columns containing local IDs. Local ID columns are found in the following tables: *sbyn\_parent\_object*, *sbyn\_assumedmatch*, *sbyn\_enterprise*, *sbyn\_systemobject*, and *sbyn\_transaction*.

