



**SIEBEL DATA QUALITY  
ADMINISTRATION GUIDE**

**MIDMARKET EDITION**

***eBUSINESS APPLICATIONS***

*VERSION 7.0, REV. A*

12PA1-DQ01-07000

*MARCH 2002*

Siebel Systems, Inc., 2207 Bridgepointe Parkway, San Mateo, CA 94404  
Copyright © 2002 Siebel Systems, Inc.  
All rights reserved.  
Printed in the United States of America

No part of this publication may be stored in a retrieval system, transmitted, or reproduced in any way, including but not limited to photocopy, photographic, magnetic, or other record, without the prior agreement and written permission of Siebel Systems, Inc.

The full text search capabilities of Siebel eBusiness Applications include technology used under license from Fulcrum Technologies, Inc. and are the copyright of Fulcrum Technologies, Inc. and/or its licensors.

Siebel, the Siebel logo, TrickleSync, TSQ, Universal Agent, and other Siebel product names referenced herein are trademarks of Siebel Systems, Inc., and may be registered in certain jurisdictions.

Other product names, designations, logos, and symbols may be trademarks or registered trademarks of their respective owners.

U.S. GOVERNMENT RESTRICTED RIGHTS. Programs, Ancillary Programs and Documentation, delivered subject to the Department of Defense Federal Acquisition Regulation Supplement, are “commercial computer software” as set forth in DFARS 227.7202, Commercial Computer Software and Commercial Computer Software Documentation, and as such, any use, duplication and disclosure of the Programs, Ancillary Programs and Documentation shall be subject to the restrictions contained in the applicable Siebel license agreement. All other use, duplication and disclosure of the Programs, Ancillary Programs and Documentation by the U.S. Government shall be subject to the applicable Siebel license agreement and the restrictions contained in subsection (c) of FAR 52.227-19, Commercial Computer Software - Restricted Rights (June 1987), or FAR 52.227-14, Rights in Data—General, including Alternate III (June 1987), as applicable. Contractor/licensor is Siebel Systems, Inc., 2207 Bridgepointe Parkway, San Mateo, CA 94404.

**Proprietary Information**

Siebel Systems, Inc. considers information included in this documentation and in Siebel eBusiness Applications Online Help to be Confidential Information. Your access to and use of this Confidential Information are subject to the terms and conditions of: (1) the applicable Siebel Systems software license agreement, which has been executed and with which you agree to comply; and (2) the proprietary and restricted rights notices included in this documentation.

# Contents

## Introduction

|   |         |
|---|---------|
| Who Should Use This Guide . . . . .             | Intro-2 |
| How This Guide Is Organized . . . . .           | Intro-3 |
| Additional Documentation . . . . .              | Intro-4 |
| What's New in This Release . . . . .            | Intro-5 |
| Contacting Siebel Technical Support . . . . .   | Intro-5 |
| Siebel Systems Welcomes Your Comments . . . . . | Intro-7 |

## Chapter 1. Siebel Data Quality Overview

|   |     |
|---|-----|
| About Siebel Data Quality . . . . .                 | 1-2 |
| Siebel Data Quality Matching . . . . .              | 1-4 |
| Siebel Connector for Firstlogic Libraries . . . . . | 1-5 |

## Chapter 2. Siebel Data Quality Matching

|   |     |
|---|-----|
| About Siebel Data Quality Matching . . . . .            | 2-2 |
| About Installing Siebel Data Quality Matching . . . . . | 2-3 |

## Chapter 3. Siebel Connector for Firstlogic Libraries

|  |     |
|--|-----|
| Siebel Connector for Firstlogic Libraries . . . . .  | 3-2 |
| Installing Dictionary Files on the Network . . . . . | 3-2 |
| Siebel Installation Options . . . . .                | 3-3 |
| Configuration Files . . . . .                        | 3-5 |
| ACEAUX.CFG . . . . .                                 | 3-5 |
| TNAUX.CFG . . . . .                                  | 3-6 |

|   |     |
|---|-----|
| ACE.CFG                                 | 3-6 |
| TN.CFG                                  | 3-7 |
| FIRM.CFG                                | 3-7 |
| FIRMINDV.CFG                            | 3-7 |
| About Modifying the Configuration Files | 3-8 |

## **Chapter 4. Configuring and Using Data Matching**

|   |      |
|---|------|
| About Data Matching   | 4-2  |
| Enabling Data Matching  | 4-3  |
| Enabling Server Component Parameters                                | 4-3  |
| Activating Siebel Data Quality Matching on the Dedicated Web Client | 4-5  |
| Setting Siebel Data Quality Options                                 | 4-5  |
| About Enabling and Disabling Without Restarting the Siebel Server   | 4-7  |
| Data Matching in Real-Time Mode                                     | 4-8  |
| Searching for and Resolving Duplicate Records                       | 4-8  |
| Merging Records   | 4-10 |
| Data Matching at the Business Component Level                       | 4-12 |
| Deduplication Tokens for Siebel Connector for Firstlogic Libraries  | 4-12 |
| Deduplication Tokens Using Siebel Data Quality Matching             | 4-13 |
| Business Component User Properties                                  | 4-14 |
| Configuring the DeDuplication Results Applet                        | 4-16 |
| Configuring Business Components and Fields                          | 4-17 |
| User Properties Used in Data Matching                               | 4-20 |
| Operational Requirements  | 4-24 |
| Table Size  | 4-24 |
| DB2 Database Parameter Optimization                                 | 4-25 |
| Server Tasks  | 4-25 |
| Ongoing Deduplication and Key Generation Operations                 | 4-26 |
| Performance Considerations  | 4-27 |
| Performance and the Data Quality Setting Parameter                  | 4-29 |

## Chapter 5. Configuring and Using Data Cleansing

|  |     |
|--|-----|
| Enabling Data Cleansing . . . . .                                      | 5-2 |
| Activating Siebel Data Cleansing on the Dedicated Web Client . . . . . | 5-3 |
| Activating Data Cleansing in Real-Time Mode . . . . .                  | 5-4 |
| Data Cleansing at the Business Component Level . . . . .               | 5-6 |
| Business Component User Properties . . . . .                           | 5-6 |
| DataCleansing Type . . . . .   | 5-7 |
| DataCleansing Field . . . . .  | 5-7 |

## Chapter 6. Batch Mode

|  |      |
|--|------|
| About Batch Mode . . . . .   | 6-2  |
| Customized Request Components . . . . .  | 6-3  |
| Checking the Data Quality Manager Component . . . . .                                | 6-4  |
| Creating New Request Components . . . . .  | 6-5  |
| Activating New Components . . . . .  | 6-7  |
| Submitting a Server Request . . . . .  | 6-8  |
| Starting Siebel Data Quality Manager Server Component . . . . .                      | 6-11 |
| Data Cleansing Parameters for Siebel Connector<br>for Firstlogic Libraries . . . . . | 6-12 |
| Component Parameters for Account Data Cleansing . . . . .                            | 6-12 |
| Component Parameters for Contact Data Cleansing . . . . .                            | 6-13 |
| Component Parameters for Prospect Data Cleansing . . . . .                           | 6-13 |
| Component Parameters for Address Data Cleansing . . . . .                            | 6-14 |
| Data Matching Parameters for Siebel Data Quality Matching . . . . .                  | 6-15 |
| Component Parameters for Account Key Generation . . . . .                            | 6-16 |
| Component Parameters for Contact Key Generation . . . . .                            | 6-16 |
| Component Parameters for Prospect Key Generation . . . . .                           | 6-17 |
| Component Parameters for Account Key Refresh . . . . .                               | 6-17 |
| Component Parameters for Contact Key Refresh . . . . .                               | 6-18 |
| Component Parameters for Prospect Key Refresh . . . . .                              | 6-18 |

Additional Parameters for Data Matching . . . . . 6-19  
    Component Parameters for Account DeDuplication . . . . . 6-19  
    Component Parameters for Contact DeDuplication . . . . . 6-20  
    Component Parameters for Prospect DeDuplication . . . . . 6-20  
Conflict\_IDs . . . . . 6-21

**Index**

# Introduction

|   |         |
|---|---------|
| Who Should Use This Guide . . . . .             | Intro-2 |
| How This Guide Is Organized . . . . .           | Intro-3 |
| Additional Documentation . . . . .              | Intro-4 |
| What's New in This Release . . . . .            | Intro-5 |
| Contacting Siebel Technical Support . . . . .   | Intro-5 |
| Siebel Systems Welcomes Your Comments . . . . . | Intro-7 |

# Who Should Use This Guide

This guide describes the process of finding and removing duplicate records.

This book will be useful primarily to people whose titles or job descriptions match one of the following:

- |  |   |
|--|---|
| <b>Database Administrators</b>           | Persons who administer the database system, including data loading, system monitoring, backup and recovery, space allocation and sizing, and user account management. |
| <b>Siebel Application Administrators</b> | Persons responsible for planning, setting up, and maintaining Siebel applications.  |
| <b>Siebel Application Developers</b>     | Persons who plan, implement, and configure Siebel applications, possibly adding new functionality.  |
| <b>Siebel System Administrators</b>      | Persons responsible for the whole system, including installing, maintaining, and upgrading Siebel applications.   |

## How This Guide Is Organized

This guide provides information necessary to install, configure, and use Siebel Data Quality Matching, MidMarket Edition and Siebel Connector for Firstlogic Libraries, MidMarket Edition. These two products provide data matching and data cleansing for the Siebel eBusiness Applications.

---

**NOTE:** All Siebel MidMarket product names include the phrase MidMarket Edition to distinguish this product from other Siebel eBusiness Applications. However, in the interest of brevity, after the first mention of a MidMarket product in this document, the product name will be given in abbreviated form. For example, after Siebel Call Center, MidMarket Edition, has been mentioned once, it will be referred to simply as Siebel Call Center. Such reference to a product using an abbreviated form should be understood as a specific reference to the associated Siebel MidMarket Edition product, and not any other Siebel Systems offering. When contacting Siebel Systems for technical support, sales, or other issues, note the full name of the product to ensure its proper identification and handling.

---

# Additional Documentation

The product documentation set for Siebel eBusiness Applications is provided on the *Siebel Bookshelf* CD-ROM. For general information about Siebel product documentation, see the *Siebel Bookshelf* home page and *Documentation Roadmap, MidMarket Edition*.

Siebel Systems, Inc., reserves the right to modify the documentation for Siebel eBusiness Applications at any time. For updates to Siebel documentation, go to the SupportWeb site (<http://ebusiness.siebel.com/supportweb/>).

If you want to order additional Siebel documentation and copies of the *Siebel Bookshelf*, go to Books Online at <http://ebusiness.siebel.com/booksonline>.

To access both SupportWeb and Books Online, you will need to provide the user name and password you received from Siebel Support Services (support@siebel.com).

## What's New in This Release

For a list of features new in this release, see the “What’s New” books included on the *Siebel Bookshelf*. Your Siebel implementation may not have all the features described in those guides, depending on which software modules you have purchased.

It is also strongly recommended that you read *Fundamentals, MidMarket Edition* so that you can make optimal use of your Siebel application, especially if you are new to Siebel software.

## Contacting Siebel Technical Support

Do you know how to access Siebel Technical Support? It is crucial that you understand the requirements for getting support before you encounter technical issues that require Siebel Technical Support’s assistance. This will facilitate smooth and timely resolution of your issues. If you have questions, please don’t hesitate to contact us.

To maximize your knowledge of Siebel products and your return on investment:

- You must attend Siebel University training to become a *designated contact*.
- Your Siebel-trained designated contacts provide technical support to your users. Siebel Technical Support provides support directly to your designated contacts only.

To provide efficient, timely support and access to the Technical Support knowledge base:

- Siebel Technical Support is primarily Web-based; it can be accessed through Siebel SupportWeb (<http://ebusiness.siebel.com/supportweb/>). Please submit new service requests to us through SupportWeb, where you can also search the knowledge base for solutions.
- Designated contacts receive read/write access to Siebel SupportWeb. All other project team members at your company receive read-only accounts so that they can access the knowledge base.

To register for Siebel training, access <http://siebeluniversity.siebel.com/edportal/jsp/index.jsp> and choose Implementation Team Training.

Please submit your technical issues and updates to Siebel SupportWeb (<http://ebusiness.siebel.com/supportweb/>). If you do not have a SupportWeb account, or if you have a question, please contact us at [support@siebel.com](mailto:support@siebel.com) or call your local Siebel Support Center:

- **North America:** + 1 800 214 0400 or + 1 650 341 0700
- **Brazil (São Paulo):** + 55 11 3444 0800
- **UK (London):** + 44 (0) 1784 494949
- **Germany (Munich):** + 49 89 957 18 400
- **France (Paris):** + 44 1784 494949
- **Ireland (Galway):** + 44 1784 494949
- **Japan (Tokyo):** 0120 606 750 (toll-free, Japan domestic only),  
+ 81 3 5464 7948 (outside of Japan)
- **Singapore:** + 65 6 212 9266

Outside of local support center hours, Gold and Rollout Support Option customers can call + 1 800 214 0400 or + 1 650 341 0700.

We appreciate your business and look forward to working with you.

## **Siebel Systems Welcomes Your Comments**

To help us improve our products, we want to know about any corrections or clarifications to this guide that you would find useful. Please include in your message:

- The title and version of the guide (very important)
- The name and version number of the Siebel application you are using
- Your name, job title or functional area, company name, phone number, and email address

Contact us through regular mail or email at:

Siebel Systems, Inc.  
Technical Publications Department  
2207 Bridgepointe Parkway  
San Mateo, CA 94404-5009

[doc@siebel.com](mailto:doc@siebel.com)

We appreciate your feedback.

## **Introduction**

*Siebel Systems Welcomes Your Comments*

# Siebel Data Quality Overview

# 1

|   |     |
|---|-----|
| About Siebel Data Quality . . . . .                 | 1-2 |
| Siebel Data Quality Matching . . . . .              | 1-4 |
| Siebel Connector for Firstlogic Libraries . . . . . | 1-5 |

# About Siebel Data Quality

The name and address data stored in Account, Contact, and Prospect records in Siebel eBusiness Applications represents your interaction with existing and potential customers. Because of the importance of this data, maintaining its integrity is critical.

Siebel Data Quality helps enterprises standardize their Account, Contact, and Prospect data in the following ways:

- **Data matching (deduplication).** Identifies possible matches for Account, Contact, and Prospect records, based on user-defined criteria. Duplicate records can be merged into a single record.

This book generally uses the term *data matching* for this functionality. The term *deduplication* is also used. Typically, it is used in configuration files, user properties, and so on.

- **Data cleansing.** Data cleansing resolves the matches identified by data matching. Data cleansing consists of:

- **Address correction.** Street address, city, state, and postal code information is stored in a uniform and consistent format, as mandated by United States postal requirements.

For recognized U.S. addresses it provides ZIP + 4 data correction and stores the data in standard U.S. Postal Service format.

- **Standardization.** Account, Contact, and Prospect information is stored in a uniform and consistent format.
- **Capitalization.** Account, Contact, and Prospect names are converted to mixed case. And address fields are converted to mixed case, all lowercase, or all uppercase.

Data matching and data cleansing functionality are provided by two third-party solutions, NAME3 software from SearchSoftware America, Inc. and Firstlogic software using the Siebel Connector for Firstlogic Libraries. Data cleansing functionality is provided using the Siebel Connector for Firstlogic Libraries. The following sections contain an overview of these products. The rest of the chapters in this book contain more detailed information about installing and using them.

---

**NOTE:** You can enable matching using Siebel Data Quality Matching and cleansing using Siebel Connector for Firstlogic Libraries. You cannot, however, enable matching with Siebel Data Quality Matching and Siebel Connector for Firstlogic Libraries simultaneously.

---

Data cleansing is also provided by third-party integration solutions validated by Siebel Systems' Technical Alliances program. For information about third-party solutions, see the Alliances page at [www.siebel.com](http://www.siebel.com).

## Siebel Data Quality Matching

Siebel Data Quality Matching provides data matching functionality for data stored in the Account, Contact, and Prospect tables within Siebel eBusiness Applications. It allows administrators to specify matching fields in Siebel Tools and identify the records that match the criteria. It also provides the ability to merge duplicate records into a single record. Siebel Data Quality Matching works across the languages and platforms supported by Siebel eBusiness Applications.

### Siebel Data Quality Matching Functionality

Siebel Data Quality Matching provides the following functionality:

- **Key generation.** Siebel Data Quality Matching uses multiple keys in order to detect duplicate records. Keys are generated based on a person's name (First Name, Middle Name, Last Name) for prospects and contacts, or account name for accounts.
- **Search.** The search functionality generates ranges of possible keys for Person Name, Company Name, or Address fields.
- **Match.** Match scores are based on similarities between the specified match fields. Different combinations of fields representing Person Name, Company Name, Address Line1, Address Line2, and ZIP Code are used for matching comparison.

The match process is twofold:

- First, Siebel Data Quality Matching identifies candidate matches by locating those records whose corresponding keys fall within a range. Like the keys, these ranges are based on a person's name (First Name, Middle Name, Last Name) for prospects and contacts, and account name for accounts.
- Next, a match score is computed for each candidate record. Candidates with scores that exceed the specified threshold are identified as matches.

## **Siebel Connector for Firstlogic Libraries**

Siebel Connector for Firstlogic Libraries is an out-of-the-box connector to the ACE, TrueName, and Match software available from Firstlogic Corporation. Siebel Connector for Firstlogic Libraries provides access to data cleansing and data matching software for addresses in the United States. Customers who choose to use Siebel Connector for Firstlogic Libraries must buy the software separately from Firstlogic Corporation.

---

**NOTE:** The Siebel Connector to Firstlogic Libraries supports Firstlogic's U.S. address product, Firstlogic ACE. Firstlogic Corporation also provides validated integration between its international address correction engine (IACE) and Siebel applications. For more information, please contact Firstlogic at [www.firstlogic.com](http://www.firstlogic.com).

---

This guide describes the functionality provided through Siebel Connector for Firstlogic Libraries. More detailed information about the Firstlogic product is provided in the third-party section of the *Siebel Bookshelf*.

## Data Cleansing Functionality

Data cleansing can fix inaccurate and inconsistent data for new or modified Account, Contact, and Prospect records, as described in [Table 1-1](#).

**Table 1-1. Data Cleansing Operations**

| Type of Correction | Comments  |
|--------------------|---|
| Address correction | <p>The following fields in an account address may be modified:</p> <ul style="list-style-type: none"> <li>■ Street Address</li> <li>■ City</li> <li>■ State</li> <li>■ ZIP Code</li> </ul> <p>Note that the Country field is not subject to data cleansing.</p> <p>For recognized U.S. addresses, the application reconciles the Address fields with their corresponding ZIP + 4 postal codes. It then stores these fields in standard U.S. Postal Service format. For example, 100 South Main Street, San Mateo, CA 94401 becomes 100 S. Main St., San Mateo, CA 94401-3256.</p>   |
| Standardization    | <p>Standardization operates on different fields for Account, Contact, and Prospect records.</p> <ul style="list-style-type: none"> <li>■ <b>Account records.</b> Standardization works on the Account Name and Site fields for Account records.</li> <li>■ <b>Contact and Prospect records.</b> Standardization operates on First Name, Middle Name, Last Name, and Job Title for these records.</li> </ul> <p>Examples of standardization are as follows:</p> <ul style="list-style-type: none"> <li>■ Siebel Systems, Incorporated becomes Siebel Systems, Inc.</li> <li>■ IBM Corporation becomes IBM Corp.</li> </ul> |
| Capitalization     | <p>The application stores and presents all Name and Job Title fields in mixed case for Account, Contact, and Prospect records. As an example, ACCOUNT REPRESENTATIVE will be changed to Account Representative and john smith will be changed to John Smith.</p>  |

## **Matching Functionality**

Unlike Siebel Data Quality Matching, the matching functionality of Siebel Connector for Firstlogic Libraries uses a single-key approach.

- **Key Generation.** One key—also called a *dedup token*—is generated per record. Although configurable, the dedup token is based on the Last Name, Account Name, and Address Postal Code fields for contacts and prospects, or the Account Name, Postal Code, and Street Address for accounts. Because the dedup token is generated during data cleansing, there is no need to run an explicit key-generation process.
- **Match.** Candidate matches are found by searching for records with identical dedup tokens and computing a score for each candidate. The score is based on similarity between match fields. Records with scores higher than the defined threshold are identified as a match.

## **Siebel Data Quality Overview**

---

*About Siebel Data Quality*

# Siebel Data Quality Matching

# 2

|   |     |
|---|-----|
| About Siebel Data Quality Matching . . . . .            | 2-2 |
| About Installing Siebel Data Quality Matching . . . . . | 2-3 |

# About Siebel Data Quality Matching

Siebel Data Quality Matching is provided by NAME3 software from SearchSoftware America, Inc. SSA-NAME3 is included on the Siebel eBusiness Applications CD-ROM and is integrated with Siebel eBusiness Applications. Customers who purchase Siebel Data Quality Matching receive the SSA software and the integration from Siebel eBusiness Applications to SSA-Name3. No additional software needs to be purchased or deployed to run Siebel Data Quality Matching.

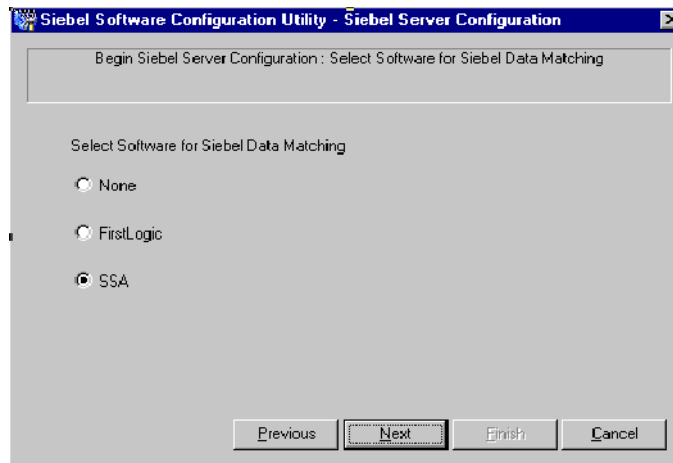
---

**NOTE:** For detailed information on SSA-NAME3, please refer to the SSA-NAME3 documentation that is included in the third-party section on the *Siebel Bookshelf*.

---

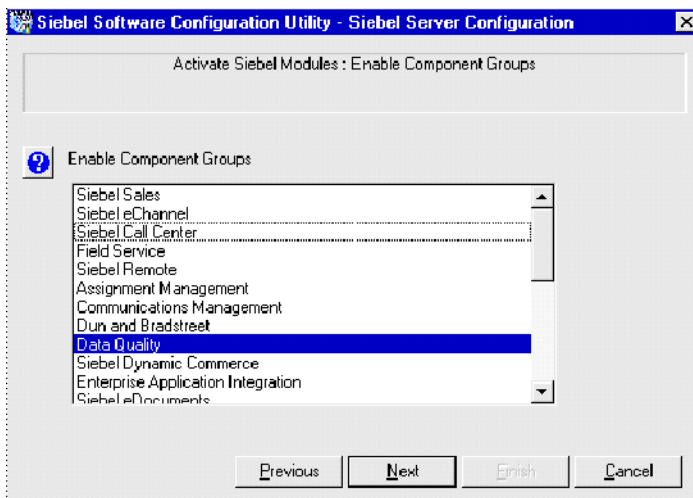
## About Installing Siebel Data Quality Matching

During the installation process for Siebel Server, choose SSA in the Siebel Software Configuration Utility - Siebel Server Configuration screen, as shown in [Figure 2-1](#), to enable Siebel Data Quality Matching. Selecting SSA enables matching for both Siebel Web Client and Siebel Dedicated Web Client.



**Figure 2-1. Selecting SSA to Install Siebel Data Quality Matching**

If you plan to run Siebel Data Quality Matching in batch mode, enable the Data Quality component group in the configuration screen as shown in [Figure 2-2](#). You can also do this later from the Server Administration screen within the application.



**Figure 2-2. Enabling the Data Quality Component**

The Data Quality Matching files and folders shown in [Table 2-1](#) are installed.

**Table 2-1. Siebel Data Quality Matching Files and Folders**

| File or Folder                                    | Description   |
|---|---|
| < siebel_root > \bin\<br>< language > \n3sgsb.dll | NT version.<br>A DLL (dll) or shared object (so) containing the routines linked with the SSA-NAME3 (Siebel Data Quality Matching) interface.<br>Note that you must replace < language > with the appropriate language code, such as ENU for American English. |

# Siebel Connector for Firstlogic Libraries

# 3

- Siebel Connector for Firstlogic Libraries . . . . . 3-2
  - Installing Dictionary Files on the Network . . . . . 3-2
  - Siebel Installation Options . . . . . 3-3
- Configuration Files . . . . . 3-5
  - ACEAUX.CFG . . . . . 3-5
  - TNAUX.CFG . . . . . 3-6
  - ACE.CFG . . . . . 3-6
  - TN.CFG . . . . . 3-7
  - FIRM.CFG . . . . . 3-7
  - FIRMINDV.CFG . . . . . 3-7
  - About Modifying the Configuration Files . . . . . 3-8

# Siebel Connector for Firstlogic Libraries

As discussed in “[Siebel Connector for Firstlogic Libraries](#)” on page 1-5, you need to purchase Siebel Connector for Firstlogic Libraries from Firstlogic Corporation ([www.firstlogic.com](http://www.firstlogic.com)). For information about which version of Siebel Connector for Firstlogic Libraries to install, see the system requirements and supported platforms documentation for your Siebel eBusiness Applications.

It is recommended that you install the Firstlogic software before installing Siebel eBusiness Applications.

- To implement data cleansing in the Siebel eBusiness Applications environment, you need to install the ACE Library and the TrueName Library.
- To implement the Siebel Connector for Firstlogic Libraries data matching, install the Match Library. ACE, TrueName, and Match Libraries are sold by Firstlogic Corporation and must be purchased separately.

The default installation path for Firstlogic programs is C:\PW. Please refer to the Firstlogic documentation provided with the product for instructions on how to install these products.

## Installing Dictionary Files on the Network

You have the option of placing your Firstlogic dictionary files in a network directory instead of placing separate copies of the dictionary files on each user’s computer.

If you place the dictionary files in a network directory, you must explicitly identify the location of the dictionary files in the aceaux.cfg file. For example, if your data dictionaries are installed in the directory \acme\common\Data\_Quality\PW\dirs, then the related entries in aceaux.cfg should be:

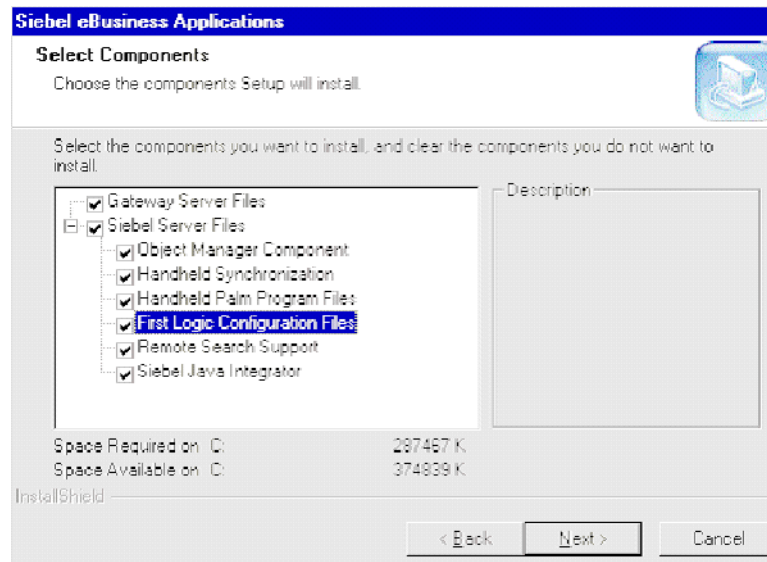
**Table 3-1. Sample ACEAUX.CFG File Entries**

|                |  |
|----------------|--|
| ACE_Dir_City   | \\acme\common\Data_Quality\PW\dirs\city07.dir      |
| ACE_Dir_ZCF    | \\acme\common\Data_Quality\PW\dirs\zcf07.dir       |
| ACE_Dir_ZIP4_1 | \\acme\common\Data_Quality\PW\dirs\dirs\zip4us.dir |

## Siebel Installation Options

You must select very specific options while installing Siebel eBusiness Applications that allow the Siebel installer to make modifications to the Firstlogic configuration files. Make the following choices during the Siebel eBusiness Applications installation process:

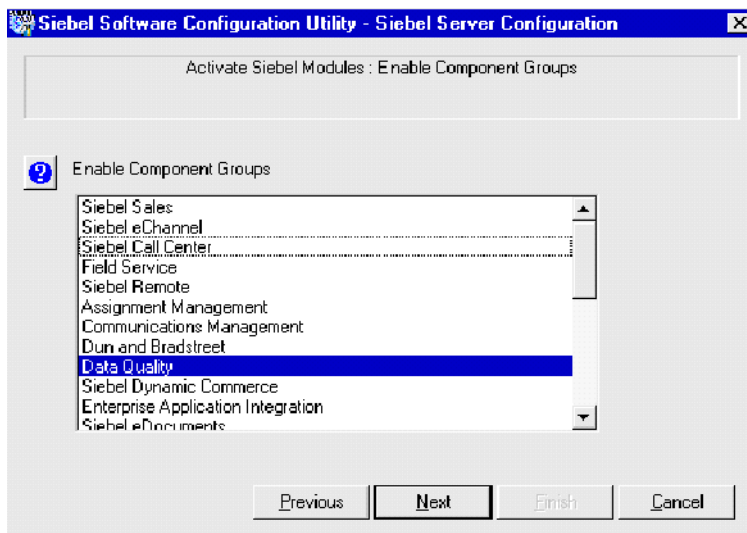
- **Setup Type.** Select the Custom installation option from the Setup Type dialog box.
- **Configuration files.** Select First Logic Configuration Files as one of the Siebel Server Files components, as shown in [Figure 3-1](#).



**Figure 3-1. Selecting Firstlogic Configuration Files**

- **Data-matching software.** Select software for data matching. See [Figure 2-1 on page 2-3](#) for a sample data-matching software selection dialog box.
- **Root directory.** Identify the Firstlogic root directory when you are asked in the Siebel Software Configuration Utility - Siebel Server Configuration dialog box. The suggested path is < drive > :\pw, such as c:\pw on Windows NT.

- **Batch mode.** If you plan to run Siebel Connector for Firstlogic Libraries in batch mode, you need to enable the Data Quality component in the configuration screen, as shown in [Figure 3-2](#), or later in the process from the Server Administration screen from within the application.



**Figure 3-2. Enabling the Data Quality Component**

## Configuration Files

If the Siebel Connector for Firstlogic Libraries programs are installed after installing Siebel eBusiness Applications, or if the Siebel installer does not locate their installation paths, verify that the Firstlogic configuration parameters match those in the tables below.

All the configuration files—including ACEAUX.CFG, TNAUX.CFG, ACE.CFG, TN.CFG, FIRM.CFG, and FIRMINDV.CFG—can be found in  
< siebel\_root > \idcentric\ < language > .

---

**NOTE:** Substitute a valid language code, such as ENU for American English, for  
< language > .

---

The following tables are based on the assumption that your system is running Windows NT and the Firstlogic files reside in the c:\PW directory. You will need to modify the path in the Change To column if the Firstlogic files are stored in a different location.

### ACEAUX.CFG

Change the following entries in ACEAUX.CFG. Be sure to verify that you are using the most recent version of the Firstlogic files, including the Postal file for ACE\_Dir\_City.

| Entry                  | Change From         | Change To                |
|------------------------|---------------------|--------------------------|
| ACE_Form_3553          | \acelib\ace3553.frm | c:\pw\acelib\ace3553.frm |
| ACE_Form_NDI           | \acelib\acendi.frm  | c:\pw\acelib\acendi.frm  |
| ACE_Dir_City           | \dirs\city06.dir    | c:\pw\dirs\city07.dir    |
| ACE_Dir_REVZIP4        | \dirs\revzip4.dir   | c:\pw\dirs\revzip4.dir   |
| ACE_Dir_ZCF            | \dirs\zcf06.dir     | c:\pw\dirs\zcf07.dir     |
| ACE_Dir_ZIP4_1         | \dirs\zip4us.dir    | c:\pw\dirs\zip4us.dir    |
| ACE_Dct_Address_Line   | \acelib\addrln.dct  | c:\pw\acelib\addrln.dct  |
| ACE_Dct_Last_Line      | \acelib\lastln.dct  | c:\pw\acelib\lastln.dct  |
| ACE_Dct_Capitalization | \acelib\pwcas.dct   | c:\pw\acelib\pwcas.dct   |
| ACE_Dct_Firm_Line      | \acelib\firmln.dct  | c:\pw\acelib\firmln.dct  |

## TNAUX.CFG

Change the following entries in TNAUX.CFG.

| Entry                  | Change From        | Change To               |
|------------------------|--------------------|-------------------------|
| TN_Dct_Parsing         | \nlib\parsing.dct  | c:\pw\tnlib\parsing.dct |
| TN_Dct_Capitalization1 | \tnlib\pwcap.dct   | c:\pw\tnlib\pwcap.dct   |
| TN_Keyfile             | \tnlib\tnlib.key   | c:\pw\tnlib\tnlib.key   |
| TN_Rule_File_Firm      | \tnlib\fprules.gcf | c:\pw\tnlib\fprules.gcf |
| TN_Rule_File_Multiline | \tnlib\mlrules.gcf | c:\pw\tnlib\mlrules.gcf |

Change the values in the following .cfg files (ACE.CFG, TN.CFG, FIRM.CFG, and FIRMINDV.CFG) to point to the location where your Siebel Server files are installed.

The following tables are based on the assumption that the Siebel Server files are in the c:\Siebel directory. You will need to modify the path in the Change To column if the Siebel Server files are in a different location.

---

**NOTE:** Change < language > to a valid language code, such as ENU for American English.

---

## ACE.CFG

Change the following entries in ACE.CFG.

| Entry                        | Change To                                   |
|------------------------------|---|
| ACE_Auxiliary_Config_File    | C:\Siebel\IdCentric\<>language>\aceaux.cfg  |
| ACE_Input_Fields_Config_File | C:\Siebel\IdCentric\<>language>\aceflds.cfg |
| ACE_Options_Config_File      | C:\Siebel\IdCentric\<>language>\aceopts.cfg |
| ACE_Auxiliary_Config_File    | C:\Siebel\IdCentric\<>language>\aceaux.cfg  |
| ACE_Input_Fields_Config_File | C:\Siebel\IdCentric\<>language>\aceflds.cfg |

## TN.CFG

Change the following entries in TN.CFG.

| Entry                       | Change To   |
|-----------------------------|---|
| TN_Auxiliary_Config_File    | C:\Siebel\IdCentric\ <language>\tnaux.cfg</language>  |
| TN_Input_Fields_Config_File | C:\Siebel\IdCentric\ <language>\tnflds.cfg</language> |
| TN_Options_Config_File      | C:\Siebel\IdCentric\ <language>\tnopts.cfg</language> |

## FIRM.CFG

Change the following entries in FIRM.CFG.

| Entry                           | Change To   |
|---------------------------------|---|
| MTC_Key_Config_File             | C:\Siebel\IdCentric\ <language>\firm_key.cfg</language> |
| MTC_General_Config_File         | C:\Siebel\IdCentric\ <language>\firm_gen.cfg</language> |
| MTC_Keyfield_Config_File        | C:\Siebel\IdCentric\ <language>\firm fld.cfg</language> |
| MTC_Auto_Options_Config_File    | C:\Siebel\IdCentric\ <language>\auto_opt.cfg</language> |
| MTC_Rule_Definition_Config_File | C:\Siebel\IdCentric\ <language>\firm_rul.cfg</language> |

## FIRMINDV.CFG

Change the following entries in FIRMINDV.CFG.

| Entry                           | Change To   |
|---------------------------------|---|
| MTC_Key_Config_File             | C:\Siebel\IdCentric\ <language>\fmin_key.cfg</language> |
| MTC_General_Config_File         | C:\Siebel\IdCentric\ <language>\fmin_gen.cfg</language> |
| MTC_Keyfield_Config_File        | C:\Siebel\IdCentric\ <language>\fmin fld.cfg</language> |
| MTC_Auto_Options_Config_File    | C:\Siebel\IdCentric\ <language>\auto_opt.cfg</language> |
| MTC_Rule_Definition_Config_File | C:\Siebel\IdCentric\ <language>\fmin_rul.cfg</language> |

### **About Modifying the Configuration Files**

When you install Siebel eBusiness Applications, the default Firstlogic configuration files are loaded into the \idCentric\ < language > subdirectory under the Siebel installation directory.

After the installation, you can open and modify Firstlogic configuration files in order to alter the cleansing and matching behavior of ACE Library, TrueName Library, and Match Library. For example, modify the files if you want to convert all text to uppercase.

Please refer to the documentation provided with the Siebel Connector for Firstlogic Libraries product for information about modifying configuration files.

# Configuring and Using Data Matching

# 4

|  |      |
|--|------|
| About Data Matching  | 4-2  |
| Enabling Data Matching   | 4-3  |
| Enabling Server Component Parameters                                   | 4-3  |
| Activating Siebel Data Quality Matching on the<br>Dedicated Web Client | 4-5  |
| Setting Siebel Data Quality Options                                    | 4-5  |
| About Enabling and Disabling Without<br>Restarting the Siebel Server   | 4-7  |
| Data Matching in Real-Time Mode  | 4-8  |
| Searching for and Resolving Duplicate Records                          | 4-8  |
| Merging Records  | 4-10 |
| Data Matching at the Business Component Level                          | 4-12 |
| Deduplication Tokens for Siebel Connector for Firstlogic Libraries     | 4-12 |
| Deduplication Tokens Using Siebel Data Quality Matching                | 4-13 |
| Business Component User Properties                                     | 4-14 |
| Configuring the DeDuplication Results Applet                           | 4-16 |
| Configuring Business Components and Fields                             | 4-17 |
| User Properties Used in Data Matching                                  | 4-20 |
| Operational Requirements   | 4-24 |
| Table Size   | 4-24 |
| DB2 Database Parameter Optimization                                    | 4-25 |
| Server Tasks   | 4-25 |
| Ongoing Deduplication and Key Generation Operations                    | 4-26 |
| Performance Considerations   | 4-27 |
| Performance and the Data Quality Setting Parameter                     | 4-29 |

# About Data Matching

Data matching, also known as *deduplication*, identifies possible matches for Account, Contact, and Prospect records based on match fields defined in Siebel Tools and administrator-defined matching preferences. You can perform data matching without data cleansing being enabled.

Siebel Data Quality Matching and Siebel Connector for Firstlogic Libraries can run in real-time or batch mode. Real-time mode and batch mode handle duplicates differently:

- In real-time mode, users are prompted when trying to enter duplicate records. For more information about operating in real-time mode, see [“Data Matching in Real-Time Mode” on page 4-8](#) and [“Activating Data Cleansing in Real-Time Mode” on page 5-4](#).
- In batch mode, the application identifies potentially-conflicting records and presents them to the system administrator for resolution. For more information about operating in batch mode, see [“Batch Mode” on page 6-1](#).

## Enabling Data Matching

Enable individual component parameters such as Call Center or Sales from the Server Administration screen.

The parameters specified in the Server Administration screen or in the configuration (.cfg) files for the dedicated Web client are always checked first. Because these parameters are specified at the object manager level, Siebel Data Quality can be enabled for one application and disabled for another. You cannot enable both the Siebel Data Quality Matching and Siebel Connector for Firstlogic Libraries data matching functionality simultaneously.

### Enabling Server Component Parameters

You can enable data matching from the Server Administration screen, as shown in [Figure 4-1](#) and the following procedure.

The screenshot shows the Siebel Server Administration interface. At the top, there are navigation elements: 'Show: Components', 'History', and 'Favorites'. Below this is the 'Component Parameter' section. The main area is titled 'Server Components' and contains a table with columns: Siebel Server, Name, Component State, Running Tasks, Running MTS Pro, Start Time, and End Time. The table lists several components like Call Center Object Manager, Data Quality Manager, Enterprise Integration Mgr, File System Manager, List Manager, Sales Object Manager, and Server Manager.

Below the 'Server Components' table is the 'Component Parameters' section. It has tabs for 'Component Tasks', 'Component Event Configuration', 'Component State Values', 'Component Statistics', and 'Component Parameters'. The 'Component Parameters' tab is active, showing a 'Query Results' table with columns: Parameter, Type, Effective Inmed?, Current Value, Value on Restart, Subsystem, and Description. The table lists three parameters: DeDuplication Data Type, DeDuplication Enable Flag, and DeDuplication Type Directory.

| Siebel Server | Name                       | Component State | Running Tasks | Running MTS Pro | Start Time          | End Time |
|---------------|----------------------------|-----------------|---------------|-----------------|---------------------|----------|
| smtb50a012    | Call Center Object Manager | Running         | 2             | 1               | 10/08/01 8:47:13 PM |          |
| smtb50a012    | Data Quality Manager       | Online          | 0             | 1               | 10/08/01 8:47:13 PM |          |
| smtb50a012    | Enterprise Integration Mgr | Online          | 0             |                 | 10/08/01 8:47:13 PM |          |
| smtb50a012    | File System Manager        | Online          | 0             | 1               | 10/08/01 8:47:13 PM |          |
| smtb50a012    | List Manager               | Online          | 0             | 1               | 10/08/01 8:47:13 PM |          |
| smtb50a012    | Sales Object Manager       | Running         | 3             | 1               | 10/08/01 8:47:13 PM |          |
| smtb50a012    | Server Manager             | Running         | 2             |                 | 10/08/01 8:47:13 PM |          |

| Parameter                    | Type    | Effective Inmed? | Current Value      | Value on Restart   | Subsystem           | Description                             |
|------------------------------|---------|------------------|--------------------|--------------------|---------------------|---|
| DeDuplication Data Type      | String  | ✓                | FirstLogic         | FirstLogic         | Infrastructure DeDu | Type of software used for DeDuplication |
| DeDuplication Enable Flag    | Boolean | ✓                | False              | False              | Infrastructure DeDu | Enable Flag for DeDuplication           |
| DeDuplication Type Directory | String  | ✓                | /export/home/postw | /export/home/postw | Infrastructure DeDu | Directory used for DeDuplication        |

Figure 4-1. Enabling Data Quality Matching Using the Server Administration Screen

#### To enable component parameters

- 1 Log on to an account with administrator responsibilities. From the application-level menu, choose View > Site Map > Server Administration > Components.
- 2 Select the appropriate object manager from the Server Components list, for example Call Center Object Manager.
- 3 Select the Component Parameters tab.
- 4 Locate the following parameters using a query and change them as indicated in the following table.

| Parameter                    | Value             | Description   |
|------------------------------|-------------------|---|
| Deduplication Enable Flag    | False             | False is the default value.   |
|                              | True              | Set to <code>true</code> to enable data matching for the application (object manager).  |
| Deduplication Data Type      | SSA or Firstlogic | Name of the data-quality matching vendor: <ul style="list-style-type: none"><li>■ For Siebel Data Quality Matching, set to <code>SSA</code>.</li><li>■ For Firstlogic, set to <code>Firstlogic</code>.</li></ul>  |
| Deduplication Type Directory | varies            | Set this parameter according to the vendor you are using: <ul style="list-style-type: none"><li>■ For Siebel Data Quality Matching, leave the field blank. SSA software is included in Siebel eBusiness Applications.</li><li>■ For Firstlogic, set to the path where the Firstlogic files are installed, such as <code>C:\PW</code>.</li></ul> |

## Activating Siebel Data Quality Matching on the Dedicated Web Client

To enable Siebel Data Quality Matching on the dedicated Web client, you modify the appropriate application configuration file, such as siebel.cfg for Siebel Sales or uagent.cfg for Siebel Call Center.

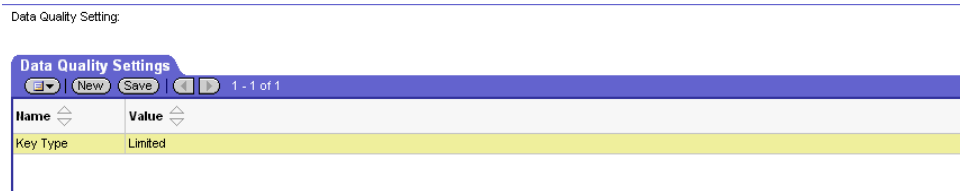
The parameters used for matching are in the [DeDuplication] section of the configuration files:

```
[DeDuplication]
Enable = True
Type = SSA
```

In the example above, matching is set to Siebel Data Quality Matching (Type = SSA). You select the application you want to use for data matching at installation time. By default, data matching is not enabled (Enable = FALSE).

## Setting Siebel Data Quality Options

You can set or change the Siebel Data Quality options from the Data Quality Settings view. The Data Quality Setting list, as seen in [Figure 4-2](#), is visible only if you have a license for the Siebel Data Quality Matching product. If you are using Siebel Connector to Firstlogic Libraries, you will not have access to the list. If that is the case, you set the values from the Server Administration screen, as described in [“Enabling Data Matching” on page 4-3](#).



**Figure 4-2. Data Quality Options Accessed from the Application Administration Screen**

#### To set a Siebel Data Quality option

- 1 From the application-level menu, choose View > Site Map > Application Administration > Data Quality Setting.
- 2 Set the values as shown in the following table.

| Name                  | Possible Values                 | Description   |
|-----------------------|---------------------------------|---|
| Disable DataCleansing | Y (Yes)<br>N (No)               | Affects business components related to Siebel Data Quality. |
| Disable DeDuplication | Y (Yes)<br>N (No)               | Affects business components related to Siebel Data Quality. |
| Key Type              | Standard<br>Limited             | Applicable only for Siebel Data Quality Matching (SSA).     |
| Match Threshold       | 0-100                           | Applicable only for Siebel Data Quality Matching (SSA).     |
| Search Type           | Narrow<br>Typical<br>Exhaustive | Applicable only for Siebel Data Quality Matching (SSA).     |

**NOTE:** For more information about the SSA-specific values, see the SSA-Name3 documentation in the third-party section of the *Siebel Bookshelf*.

## **About Enabling and Disabling Without Restarting the Siebel Server**

The Disable DataCleansing and Disable DeDuplication settings specified in the Application Administration screen override those specified in the Server Administration screen. If you have enabled them from the Server Administration screen, you can disable them from the Application Administration screen without restarting the Siebel server. After you disable them, log off and then log on to the application for the new settings to take effect. The settings will apply to all the object managers in your Siebel server, no matter whether or not they have been enabled in the Server Administration screen. When you are ready to enable data cleansing or deduplication again, reset the two fields to N and log on again.

## Data Matching in Real-Time Mode

In real-time mode, Siebel Data Quality is invoked when you save a new or modified record.

- If data cleansing is enabled, application fields are standardized prior to committing the record.
- If data matching is enabled and the new record is a potential duplicate, the Possible Matches dialog box appears. If you do not think the record is a duplicate, close the dialog box. Doing so commits the record to the database. Otherwise, if you think the record is a duplicate, select the best-matching record from the dialog box. The duplicate record will be removed from the system in one of the following ways:
  - If you are creating a new record, that record will not be saved.
  - If you are modifying a record, that record will be deleted from the table.

---

**NOTE:** The Data Administration > Data Quality view is also populated with duplicate records.

---

## Searching for and Resolving Duplicate Records

You manage duplicate records from the administration views.

The searching phase involves creating a query to find duplicate records and reviewing the query results. After the query results appear, you can drill down on a record or click the Resolution tab to view the duplicates for the record. Then, you specify which record should be the retained record and request that the records be merged.

Merging records means that the following events take place:

- All fields from the retained record are kept.
- Any fields that were empty in the retained record are populated in the retained record by information (if any) from the appropriate fields in the records not retained.

- The children and grandchildren of the records not retained are reparented to the retained record.
- Records not retained are flagged and are no longer visible to users or administrators, but they are *not* automatically deleted.

**To search for a duplicate record**

- 1** From the application-level menu, choose View > Site Map > Data Administration > Data Quality.
- 2** Click one of the following tabs:
  - Duplicate Accounts
  - Duplicate Contacts
  - Duplicate Prospects
- 3** Click Query, enter your search criteria, and then click Go.  
The search results appear.

Account Duplicates:

**Duplicate Accounts** | Query | 1 - 5 of 5

| Account        | Site  | Comment |
|----------------|-------|---------|
| Account-1      | SM    |         |
| Pizza Delivery | SM    |         |
| Mascot Jump    |       |         |
| Pizza A GO     |       |         |
| Pizza A GO GO  | Sunny |         |

Query | 2 of 5

**Account:** 
                 
 **Site:**

## Merging Records

Duplicate records can be merged using the Merge button or the Merge Records option from the menu list. There is a difference between these two types of merging operations:

- **Using the Merge button.** Performs a sequenced merge before performing a core merge. This is the preferred method.
- **Using the Merge Records option.** Performs only a core merge.

When multiple records are merged, the child and grandchild records associated with the merged record are associated back to the surviving parent record.

The number in the Sequence field is used when you want to merge multiple records into one record. If one field of the master record is empty, the Merge button will copy the value to the master record from the other records that are going to be merged.

The Sequence number prescribes the order of field values from which to select. It is useful if many fields are empty. For example, Sequence number 1 is the retained record and Sequence number 2 is the next best match and so forth. If the field in Record 1 is empty, then the value from Record 2 will be copied into Record 1.

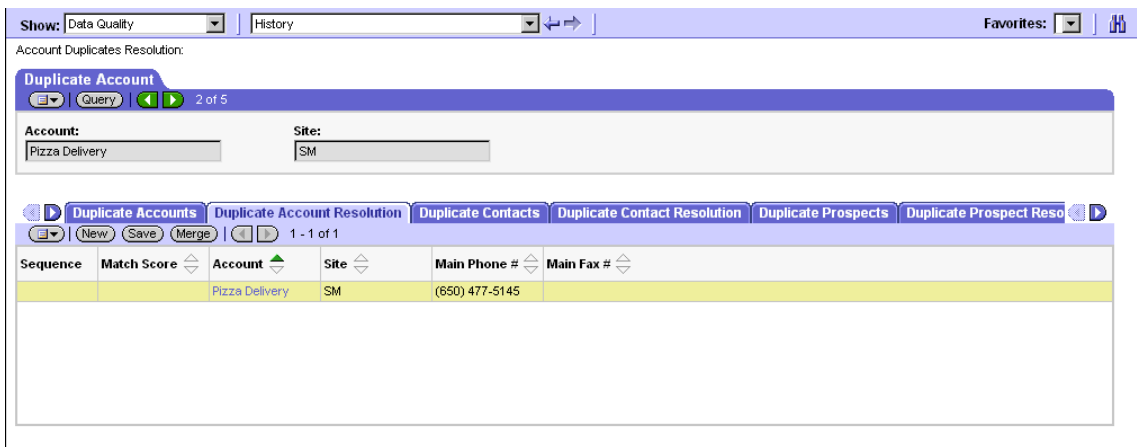
The retained record is based on the last record selected, not the Sequence number entered. The Sequence number only indicates the sequence in which null fields should be populated from the losing record to the winning record.

---

**NOTE:** Merging records is a nonreversible operation. Please review all records carefully before initiating this function.

---

Figure 4-3 shows the Duplicate Account Resolution screen.



**Figure 4-3. Duplicate Account Resolution Screen**

**To resolve duplicate records**

- 1 Click the resolution tab for the type of record you have selected.  
 For example, if you have selected an Account record, click the Duplicate Accounts Resolution tab, as shown in [Figure 4-3](#).
- 2 If two records appear to be duplicates, enter a sequence number in the Sequence field for each record.  
 The record selected last is retained after the merge. Missing fields in the retained record are populated from corresponding fields in the remaining record in ascending Sequence number order.
- 3 Edit the records, if necessary.
- 4 Click Merge.  
 The two records are merged to produce one new record.

# Data Matching at the Business Component Level

The data matching (deduplication) feature is disabled by default. It is enabled for the application through settings in the Siebel Server Administration screen or the corresponding Siebel application .cfg file for the dedicated Web client. After it is turned on at the application level, matching can be turned off for a specific business component by deactivating all of the child user properties (that is, by setting the Inactive property to TRUE). Matching cannot be turned off for individual records.

## Deduplication Tokens for Siebel Connector for Firstlogic Libraries

Data matching using the Siebel Connector for Firstlogic Libraries detects possible matches to records in specified business components during record creation and update. The matching process begins with the *deduplication token*, which is an identifier calculated for each Account, Contact, or Prospect record in the database, and newly-created or modified records. Based on the value of the deduplication token, Siebel eBusiness Applications pass a short list of prequalified, possible matches between existing and new or modified records.

The deduplication token is generated based on the following calculations:

- **Contacts.** The deduplication token consists of a concatenated string of the cleansed five-digit ZIP Code of the account address, the first letter of the cleansed account name, and the first letter of the cleansed last name.

The calculation expression is:

```
IfNull (Left ([Postal Code], 5), "_") + IfNull (Left ([Account], 1), "_") + IfNull (Left ([Last Name], 1), "_")
```

- **Accounts.** The deduplication token consists of a concatenated string of the cleansed five-digit ZIP code of the account address, the first letter of the cleansed account name, and the first nonblank, nonnumeric character of the cleansed street name.

The calculation expression is:

```
IfNull (Left ([Postal Code], 5), "_") + IfNull (Left ([Name], 1), "_") + IfNull (Mid ([Street Address], FindNoneOf ([Street Address], "1234567890 "), 1), "_")
```

- **Prospects.** The deduplication token consists of a concatenated string of the cleansed five-digit ZIP code of the account address, the first letter of the cleansed account name, and the first letter of the cleansed last name.

The calculation expression is:

```
IfNull (Left ([Postal Code], 5), "_") + IfNull (Left ([Primary  
Account Name], 1), "_") + IfNull (Left ([Last Name], 1), "_")
```

Calculation expressions for deduplication tokens follow the same syntax rules as calculated fields.

You can change the calculation expression to adjust the filtering that generates the candidate set. For example, you might want a larger candidate set or one that is more restrictive. You might also want to add or remove a field from the expression.

To change the calculation value, modify the value in the DeDup Token Value user property and run data cleansing in batch mode to regenerate the token values in the database for existing data.

## Deduplication Tokens Using Siebel Data Quality Matching

Siebel Data Quality Matching uses key generation and twofold matching logic to create and use deduplication tokens.

- **Key generation.** Siebel Data Quality Matching uses multiple keys in order to detect duplicate records. Keys are generated based on a person's name (First Name, Middle Name, Last Name) for prospects and contacts, or the account name for accounts.
- **Match.** Match scores are based on similarities between the specified match fields. Different combinations of fields representing Person Name, Company Name, Address Line1, Address Line2, and ZIP Code are used for matching comparison.

The match process is twofold:

- First, Siebel Data Quality Matching identifies candidate matches by locating those records whose corresponding keys fall within a range. Like the keys, these ranges are based on a person's name (First Name, Middle Name, Last Name) for prospects and contacts, and account name for accounts.
- Next, a match score is computed for each candidate record. Candidates with scores that exceed the specified threshold are identified as matches.

### Business Component User Properties

Deduplication has a set of numbered user properties that create correspondences between deduplication fields and fields in business components. These field-mapping properties have names of the form DeDuplication Field  $n$ , where  $n$  is an integer (for example, DeDuplication Field 7). The value consists of a pair of strings enclosed in double quotation marks and separated by a comma and a space for example:

```
DataCleansing Field 1 "Last Name", "Last Name"
```

The first string identifies the deduplication field name and the second string identifies the Siebel name.



**Caution:** If the configuration is not correct, data matching will not take place. For example, if there is *not* a comma and a space between the text strings in the Value user property, the records for the corresponding business component will not be deduplicated.

### Mapped Fields

The set of fields mapped in DeDuplication Field user properties are passed in records of the candidate set to the Firstlogic application. The candidate set consists of records with a deduplication token exactly or partially matching the calculated deduplication token of the record being added or modified, and therefore representing possible duplicates.

Deduplication field names are different for Firstlogic and Siebel Data Quality Matching.

- **Siebel Connector for Firstlogic Libraries.** Uses the DeDuplication Field property in the default configuration.
- **Siebel Data Quality Matching.** Uses the SSA DeDuplication Field. Siebel Data Quality Matching first looks up the SSA DeDuplication Field. If this property is not specified, it looks up the DeDuplication Field.

Other user properties in the business component include:

- **SSA Match Purpose.** This property must be included for Siebel Data Quality Matching.

The choices are Company\_Mandatory, Company\_Optional, Contact\_Mandatory, and Contact\_Optional. By default, Account is set to Company\_Mandatory. Contact and Prospect are set to Contact\_Mandatory.

- If a value is marked mandatory, it implies that the value counts against the total score.
- Values marked Optional do not count toward the total score.

For a more detailed explanation of this property, please refer to the SSA-NAME3 documentation that is included in the third-party section on the *Siebel Bookshelf*.

- **DeDuplication Key BusComp.** This is used with Siebel Data Quality Matching. It is the business component for the key table. You must create it if you want to perform deduplication on objects other than those that come preconfigured with Siebel applications.

---

**NOTE:** Siebel Connector for Firstlogic Libraries has only one key, the DeDup Token. It is stored in the base table.

---

- **DeDuplication Results BusComp.** This is the business component used to populate results.

---

**NOTE:** You should consider the set of fields mapped to the Siebel Data Quality applications in standard Siebel eBusiness Applications a required minimal set. Do not remove these mappings.

---

### About Prospect Data

You need to configure Prospects records differently than Contact and Account records. The reason is that Prospects share name processing capabilities in the Firstlogic application with Contacts. The system assumes that Contact data (rather than Prospect data) is present.

In order to specify that Prospect data is being processed, you need to add the following two user properties:

- **DeDuplication Results BusComp.** Name of the business component that will hold the returned data, typically DeDuplication Results (Prospect). This user property is a child of the business component being processed.
- **DeDuplication Results Applet.** Name of the dialog box that is to be used to prompt the user to resolve duplicates, typically the DeDuplication Results (Prospect) list applet. This user property is a child of the applet from which deduplication is invoked.

### Configuring the DeDuplication Results Applet

Each business component that uses deduplication has a results business component. The business component name is DeDuplication Results (*xxx*), where *xxx* identifies the type of data being validated. For example, DeDuplication Results (Account) is the name of the results business component for account data.

Both real-time and batch results are stored in the Results business component. But, the real-time results are a subset that the real-time dialog box isolates based on a column—called DEDUP\_REQ\_ID—that holds the deduplication request ID.

The dialog box is configurable in Siebel Tools, and has the name DeDuplication Results (*xxx*) list applet, where *xxx* is the Account, Contact, or Prospect. Additional columns from the S\_DEDUP\_RESULT table can be exposed as fields in the DeDuplication Results (*xxx*) business component and as list columns in the DeDuplication Results (*xxx*) list applet. However, you only need to modify the dialog box if you are changing the default configuration of Siebel Data Quality. The DeDuplication Results (*xxxx*) business component must be added to a business object *yyyy* if you want to add duplicate business component *xxxx* in a view based on business object *yyyy*.

The process of configuring business components follows these steps:

- 1** Add extension tables to a Business Component object.
- 2** With Account and Contact, add S\_ORG\_EXT\_X or S\_CONTACT\_X.
- 3** Set the Outer Join Flag to True in order for the Possible Matches dialog box to populate correctly.

---

**NOTE:** For example, go to the DeDuplication Results (Account) business component and open the Join folder. For the S\_ORG\_EXT\_X table, put a check in the Outer Join Flag column.

---

- 4** Recompile.

## **Configuring Business Components and Fields**

In standard Siebel eBusiness Applications, only four business components have data cleansing and deduplication enabled: Account, Business Address (cleansing only), Contact, and List Mgmt Prospective Contact.

If you insert records through noncleansed, nondeduplicated business components, the data is left dirty. Data will be cleansed and deduplicated during batch cleansing and deduplication only if the data can be seen through one of the business components above. You do not need to enable Data Quality on a business component if you do not permit inserting or updating through the business component.

### Specialized Deduplication Fields

Business components have certain fields with specialized deduplication purposes. These fields are:

- **DeDup Token (Firstlogic only).** Text field that represents the DEDUP\_TOKEN column in the base table.

This field stores the calculated dedup token value for each record in the business component. It is required, but it is not exposed in the user interface. If you need to add this field's functionality to another business component, use or create an extension column.

- **Deduplication Match Score.** An integer field whose value is obtained from the TOT\_SCORE\_VAL column in the S\_DEDUP\_RESULT table through a join to that table on the duplicate object ID.

This field is represented in the user interface as the Match Score list column. S\_DEDUP\_RESULT passes the results of evaluation of the candidate set back to Siebel applications from the Siebel Data Quality Matching or Firstlogic application. TOT\_SCORE\_VAL indicates the closeness of the match between the master account record and the detail record in which the value is stored. This field is not required and is used only for the presentation of batch results.

- **Deduplication Object ID.** Row ID field that functions as a foreign key to the master duplicate record in the same business component. It represents the OBJ\_ID column in the S\_DEDUP\_RESULT table and is obtained through the same join as Deduplication Match Score.

This field is not exposed in the user interface and is not required; it is used only for the presentation of batch results.

### Obsolete Records

Using the Merge button does not remove obsolete records in the S\_DEDUP\_RESULT table. Periodically, you need to do this manually using an SQL script that deletes records where the Object ID or the duplicate Object ID are not in the Accounts, Contacts, or Prospects tables or in a configured business component.

## **New Fields**

New fields included in Siebel Data Quality in Siebel 7 include:

- **DeDup Key Modification Date.** Populated when keys are generated. Used by Siebel Data Quality Matching only.
- **DeDup Last Match Date.** Populated when deduplication is run on a record. Used by Siebel Data Matching only
- **Last Cleanse Date.** Populated when cleansing is run on a record. This field is not available for Business Address.
- **Last Update - SDQ.** Used for Siebel Data Quality Matching only.

When running Key Refresh, this field on the record is compared with the DeDup Key Modification Date. If it is not specified (default case) the Update column on the base table is used. For contact and account, the last update of S\_CONTACT and S\_ORG\_EXT respectively are looked at.

---

**NOTE:** The DeDup Keys field is not used in Siebel 7.

---

## User Properties Used in Data Matching

The user properties described in this section are used for data matching for Account, Contact, and Prospect by both Siebel Data Quality Matching and Siebel Connector for Firstlogic Libraries. Names beginning with SSA (for example, SSA DeDuplication Field 6) are used by Siebel Data Quality Matching. Names not beginning with SSA (for example, DeDuplication Field 1) are used by Siebel Connector for Firstlogic Libraries.

[Table 4-1](#) describes fields for data matching for Account records.

**Table 4-1. Fields Used in Data Matching for Account Records (1 of 2)**

| Name                          | Value  |
|-------------------------------|--|
| DeDup Token Value             | IfNull (Left ([PostalCode], 5), "_") + IfNull (Left ([Name], 1), "_") + IfNull (Mid ([Street Address], FindNoneOf ([Street Address], "1234567890"), 1), "_") |
| DeDuplication CFG File        | firm.cfg   |
| DeDuplication Field 1         | "Firm Name", "Name"  |
| DeDuplication Field 2         | "Firm Location", "Location"  |
| DeDuplication Field 3         | "Street Address", "Street Address"   |
| DeDuplication Field 4         | "City", "City"   |
| DeDuplication Field 5         | "State", "State"   |
| DeDuplication Field 6         | "Zip 10", "Postal Code"  |
| DeDuplication Field 7         | "Country", "Country"   |
| DeDuplication Key BusComp     | DeDuplication - SSA Account Key  |
| DeDuplication Results BusComp | DeDuplication Results (Account)  |
| SSA DeDuplication Field 1     | "C", "Name"  |
| SSA DeDuplication Field 2     | "Country", "Country"   |
| SSA DeDuplication Field 3     | "City", "City"   |
| SSA DeDuplication Field 4     | "State", "State"   |

**Table 4-1. Fields Used in Data Matching for Account Records (2 of 2)**

| <b>Name</b>               | <b>Value</b>          |
|---------------------------|-----------------------|
| SSA DeDuplication Field 5 | "Z", "Postal Code"    |
| SSA DeDuplication Field 6 | "I", "DUNS Number"    |
| SSA DeDuplication Field 7 | "A", "Street Address" |

Table 4-2 lists and describes fields used in data matching for Contact records.

**Table 4-2. Fields Used in Data Matching for Contact Records (1 of 2)**

| <b>Name</b>                     | <b>Value</b>   |
|---------------------------------|--|
| DeDup Token Value               | IfNull (Left ([PostalCode], 5), "_") + IfNull (Left ([Account], 1), "_") + IfNull (Left ([Last Name], 1), "_") |
| DeDuplication Field 1           | "Last Name", "Last Name"   |
| DeDuplication Field 10          | "Country", "Country"   |
| DeDuplication Field 2           | "First Name", "First Name"   |
| DeDuplication Field 3           | "Middle Name", "Middle Name"   |
| DeDuplication Field 4           | "Firm Name", "Account"   |
| DeDuplication Field 5           | "Firm Location", "Firm Location"   |
| DeDuplication Field 6           | "Street Address", "Street Address"   |
| DeDuplication Field 7           | "City", "City"   |
| DeDuplication Field 8           | "State", "State"   |
| DeDuplication Field 9           | "Zip 10", "Postal Code"  |
| DeDuplication Key BusComp       | DeDuplication - SSA Contact Key  |
| DeDuplication Last Update Field | Last Update - SDQ  |
| DeDuplication Results BusComp   | DeDuplication Results (Contact)  |
| SSA DeDuplication Field 1       | "Z", "Postal Code"   |
| SSA DeDuplication Field 10      | "First", "First Name"  |

## Configuring and Using Data Matching

### User Properties Used in Data Matching

**Table 4-2. Fields Used in Data Matching for Contact Records (2 of 2)**

| Name                       | Value                         |
|----------------------------|-------------------------------|
| SSA DeDuplication Field 11 | "Middle", "Middle Name"       |
| SSA DeDuplication Field 12 | "City", "City"                |
| SSA DeDuplication Field 13 | "State", "State"              |
| SSA DeDuplication Field 14 | "Country", "Country"          |
| SSA DeDuplication Field 2  | "T", "Work Phone #"           |
| SSA DeDuplication Field 3  | "T", "Cellular Phone #"       |
| SSA DeDuplication Field 4  | "I", "Social Security Number" |
| SSA DeDuplication Field 5  | "E", "Email Address"          |
| SSA DeDuplication Field 6  | "A", "Street Address"         |
| SSA DeDuplication Field 7  | "T", "Home Phone #"           |
| SSA DeDuplication Field 8  | "C", "Account"                |
| SSA DeDuplication Field 9  | "Last", "Last Name"           |

Table 4-3 lists and describes fields used in data matching for Prospect records.

**Table 4-3. Fields Used in Data Matching for Prospect Records (1 of 2)**

| Name                   | Value  |
|------------------------|--|
| DeDup Token Value      | IfNull (Left ([PostalCode], 5), "_") + IfNull (Left ([Account], 1), "_") + IfNull (Left ([Last Name], 1), "_") |
| DeDuplication Field 1  | "Last Name", "Last Name"   |
| DeDuplication Field 10 | "Address.Country", "Country"   |
| DeDuplication Field 2  | "First Name", "Account"  |
| DeDuplication Field 3  | "Middle Name", "Middle Name"   |
| DeDuplication Field 4  | "Firm Name", "Account"   |
| DeDuplication Field 5  | "Firm Location", "Primary Account Location"  |

**Table 4-3. Fields Used in Data Matching for Prospect Records (2 of 2)**

| <b>Name</b>                   | <b>Value</b>                               |
|-------------------------------|--|
| DeDuplication Field 6         | "Address.Street Address", "Street Address" |
| DeDuplication Field 7         | "Address.City", "City"                     |
| DeDuplication Field 8         | "Address.State", "State"                   |
| DeDuplication Field 9         | "Address.Zip10", "Postal Code"             |
| DeDuplication Key BusComp     | DeDuplication - SSA Prospect Key           |
| DeDuplication Results BusComp | DeDuplication Results (Prospect)           |
| SSA DeDuplication Field 1     | "Last", "Last Name"                        |
| SSA DeDuplication Field 10    | "T", "Work Phone #"                        |
| SSA DeDuplication Field 11    | "T", "Cellular Phone #"                    |
| SSA DeDuplication Field 12    | "I" "Social Security Number"               |
| SSA DeDuplication Field 13    | "E", "Email Address"                       |
| SSA DeDuplication Field 14    | "T", "Home Phone #"                        |
| SSA DeDuplication Field 2     | "First", "First Name"                      |
| SSA DeDuplication Field 3     | "Middle", "Middle Name"                    |
| SSA DeDuplication Field 4     | "C", "Account"                             |
| SSA DeDuplication Field 5     | "A", "Street Address"                      |
| SSA DeDuplication Field 6     | "City", "City"                             |
| SSA DeDuplication Field 7     | "State", "State"                           |
| SSA DeDuplication Field 8     | "Z", "Postal Code"                         |
| SSA DeDuplication Field 9     | "Country", "Country"                       |

## Operational Requirements

This section presents operational recommendations for table size, DB2 database parameter optimization, server tasks, deduplication, and key generation processing.

Batch deduplication is a long process. The recommendations in this section will help you optimize deduplication.

---

**NOTE:** Most of the recommendations in this section are designed for users of SSA.

---

### Table Size

The deduplication process generates many records. Follow the recommendations in [Table 4-4](#) when allocating space for the SIEBEL\_4K table space.

**Table 4-4. Table Size Recommendations**

| Table   | Sizing Recommendation  |
|---|--|
| S_PER_DEDUP_KEY,<br>S_ORG_DEDUP_KEY<br>S_PRSP_DEDUP_KEY | <p>These tables may contain between two and six times more records than their corresponding base tables, depending on the key type used during the key generation stage:</p> <ul style="list-style-type: none"> <li>■ <b>Limited key type.</b> May contain between two and four times more records than the corresponding base table.</li> <li>■ <b>Standard key type.</b> May contain up to six times more records than the corresponding base table. This is a conservative estimate.</li> </ul>   |
| S_DEDUP_RESULT  | <p>This table may contain between five and six times the number of records in the three base tables combined. Use the following guidelines to help determine table size:</p> <ul style="list-style-type: none"> <li>■ If the base tables contain many deduplicate records, more records will be inserted into the results table.</li> <li>■ If a Search Type of Typical or Exhaustive is used, more records will be inserted into the results table.</li> <li>■ If a low match threshold—such as a threshold in the lower 100 range—is used, the matching process generates a larger number of records to be inserted into the results table.</li> </ul> |

## **DB2 Database Parameter Optimization**

After key generation, run the DB2 REORGCHK utility to check if the \*\_DEDUP\_KEY tables need to be reorganized. Check the cluster ratio in the F4 column. If it is less than 94% for the M1 index, use the DB2 REORG utility to reorganize the \*\_DEDUP\_KEY tables based on the M1 index.

If performance seems degraded, use the run\_status utility to update statistics on the following tables:

- S\_PER\_DEDUP\_KEY
- S\_ORG\_DEDUP\_KEY
- S\_ORG\_DEDUP\_KEY
- S\_DEDUP\_RESULT

## **Server Tasks**

Execute concurrent Data Quality Manager server tasks to deduplicate the data. Query the base tables to find the search specifications such that each query result set contains the preferred number of records, between 50,000 and 75,000 per server task. Start concurrent server tasks using these search specifications and continue the deduplication operation until the entire table is completely processed.

## Ongoing Deduplication and Key Generation Operations

The key generation and deduplication timestamp fields, DeDup Key Modification Date and DeDup Last Match Date, respectively, are updated when key generation or deduplication is performed on a record. Use these fields in the search specification to filter records with up-to-date keys and matches.

For users who have over 100,000 records in their base tables, it can be very time consuming to reprocess all the records. For that reason, the recommended record set for subsequent deduplication and key generation runs consists of outdated and new records only. Outdated records are those records that have been modified since they were tagged. In SQL terms, outdated records are those in which [Updated] > (DQ tag).

The examples in [Table 4-5](#) illustrate search specifications—the Object Where clause—to use when running key generation or deduplication using the Contact business component.

**Table 4-5. Sample Search Specifications for Key Generation and Deduplication Runs**

| To Query For...         | Key Generation Example   | Deduplication Example   |
|-------------------------|--|---|
| Updated records         | ( [DeDup Key Modification Date] < [Updated] )  | ( [DeDup Last Match Date] < [Updated] )   |
| New records             | ( [DeDup Key Modification Date] IS NULL )  | ( [DeDup Last Match Date] IS NULL )   |
| Updated and new records | ( [DeDup Key Modification Date] < [Updated] ) OR ( [DeDup Key Modification Date] IS NULL ) | ( ([DeDup Last Match Date] < [Updated] ) OR ([DeDup Last Match Date] IS NULL) ) |

## Performance Considerations

This section presents some general suggestions for improving performance and the use of an option to improve batch performance, as summarized in [Table 4-6](#).

**Table 4-6. Performance Improvement Suggestions (1 of 2)**

| Element   | Comments  |
|---|---|
| General settings on the Application Administration Screen | <p>From the application-level menu, choose View &gt; Site Map &gt; Application Administration &gt; Data Quality Setting to set the following parameters.</p> <ul style="list-style-type: none"> <li>■ <b>Key type (key generation)</b>. Set to Limited.</li> <li>■ <b>Search type (deduplication)</b>. Set to Narrow.</li> <li>■ <b>Match threshold (deduplication)</b>. Set to a number greater than or equal to 75. The higher the threshold, the faster the Deduplication process runs.</li> </ul> |
| Object Sort clause  | <p>Set this parameter based on the key generation parameters for deduplication. For example, use:</p> <ul style="list-style-type: none"> <li>■ <b>Person (contact or prospect)</b>. ‘Use Last Name, First Name, Middle Name’.</li> <li>■ <b>Company (account)</b>. Use “Name” or “Name, Location”.</li> </ul>   |

**Table 4-6. Performance Improvement Suggestions (2 of 2)**

| Element                       | Comments   |
|-------------------------------|--|
| Data quality Delete parameter | <p>Use this parameter to clear the results table, S_DEDUP_RESULT, of any records that correspond to the records that will be processed by the search specification logic. During a deduplication run, the results table must <i>not</i> contain records generated in previous runs. For example, to deduplicate contacts whose last names start with A, the results table cannot contain any records generated during previous runs with last names that start with A.</p> <p>From the application-level menu, choose View &gt; Site Map &gt; Application Administration &gt; Data Quality Setting to set the Delete parameter. For more information, see <a href="#">“Performance and the Data Quality Setting Parameter”</a> on page 4-29.</p> |
| Batch business components     | <p>Use the batch business components for Accounts and Contacts. You may need to configure them if you need to use other fields in the Object Where clause:</p> <ul style="list-style-type: none"><li>■ <b>Contact.</b> Business object is DeDuplication, business component is Contact (SDQ Batch)</li><li>■ <b>Account.</b> Business object is DeDuplication, business component is Account (SDQ Batch)</li></ul>   |

## Performance and the Data Quality Setting Parameter

Setting the Data Quality Setting (DQSetting) parameter to Delete can improve the performance of batch deduplication and key generation processing.

Before populating either the results table (deduplication) or the key table (key generation) with new data, you need to delete the existing keys. Although the application can do this automatically, it is time-consuming. An administrator who is an expert user of SQL may choose to use the Delete option of the DQSetting parameter to delete the data prior to running the task.



**Caution:** Do not attempt to use the Delete option if you are not an expert user of SQL.

## **Configuring and Using Data Matching**

*Performance and the Data Quality Setting Parameter*

# Configuring and Using Data Cleansing

# 5

|  |     |
|--|-----|
| Enabling Data Cleansing . . . . .                                      | 5-2 |
| Activating Siebel Data Cleansing on the Dedicated Web Client . . . . . | 5-3 |
| Activating Data Cleansing in Real-Time Mode . . . . .                  | 5-4 |
| Data Cleansing at the Business Component Level . . . . .               | 5-6 |
| Business Component User Properties . . . . .                           | 5-6 |
| DataCleansing Type . . . . .   | 5-7 |
| DataCleansing Field . . . . .  | 5-7 |

# Enabling Data Cleansing

The Siebel Connector for Firstlogic Libraries is disabled by default. You can enable it from the Server Administration screen or on the dedicated Web client. You can enable data cleansing from the Server Administration screen, as shown in [Figure 5-1](#) and discussed in the following procedure.

The screenshot displays the Siebel Server Administration interface. At the top, there are navigation elements including 'Show: Components', 'History', and 'Favorites'. Below this, the 'Server Components' section is active, showing a table of server components. Below that, the 'Component Parameters' section is active, showing a table of configuration parameters for data cleansing.

| Siebel Server | Name                       | Component State | Running Tasks | Running MTS Pro | Start Time          | End Time |
|---------------|----------------------------|-----------------|---------------|-----------------|---------------------|----------|
| smtb50a012    | Call Center Object Manager | Running         | 2             | 1               | 10/08/01 8:47:13 PM |          |
| smtb50a012    | Data Quality Manager       | Online          | 0             | 1               | 10/08/01 8:47:13 PM |          |
| smtb50a012    | Enterprise Integration Mgr | Online          | 0             |                 | 10/08/01 8:47:13 PM |          |
| smtb50a012    | File System Manager        | Online          | 0             | 1               | 10/08/01 8:47:13 PM |          |
| smtb50a012    | List Manager               | Online          | 0             | 1               | 10/08/01 8:47:13 PM |          |
| smtb50a012    | Sales Object Manager       | Running         | 3             | 1               | 10/08/01 8:47:13 PM |          |
| smtb50a012    | Server Manager             | Running         | 2             |                 | 10/08/01 8:47:13 PM |          |

| Parameter                  | Type    | Effective Inmed? | Current Value      | Value on Restart   | Subsystem             | Description                             |
|----------------------------|---------|------------------|--------------------|--------------------|-----------------------|---|
| Data Cleansing Directory   | String  | ✓                | /export/home/postw | /export/home/postw | Infrastructure Dataac | Directory used for datacleansing        |
| Data Cleansing Enable Flag | Boolean | ✓                | False              | False              | Infrastructure Dataac | Enable Flag for data cleansing          |
| Data Cleansing Type        | String  | ✓                | FirstLogic         | FirstLogic         | Infrastructure Dataac | Type of software used for datacleansing |

**Figure 5-1. Enabling Data Cleansing Using the Server Administration Screen**

### To enable data cleansing

- 1 From the application-level menu, choose View > Site Map > Server Administration > Components.
- 2 Select the appropriate object manager from the Server Component list for example, the Call Center Object Manager.

- 3 Select the Component Parameters tab.
- 4 Locate the following parameters using a query and change them as indicated.

| Parameter                  | Value              | Description   |
|----------------------------|--------------------|---|
| Data Cleansing Enable Flag | True<br>False      | Set to <code>true</code> to enable data cleansing for the application (object manager). |
| Data Cleansing Type        | FirstLogic         | Name of the data cleansing vendor.  |
| Data Cleansing Directory   | < directory name > | Set this parameter to the location of the Firstlogic installation.                      |

## Activating Siebel Data Cleansing on the Dedicated Web Client

To enable data cleansing using the Siebel Connector for Firstlogic Libraries on the dedicated Web client, you modify the appropriate application configuration file. The configuration files for two of the key Siebel eBusiness Applications are:

- `siebel.cfg` (Siebel Sales)
- `uagent.cfg` (Siebel Call Center)

The parameters used for matching are in the `[DataCleansing]` section of the configuration files:

```
[DataCleansing]
Enable = True
Type = Firstlogic
Dir = C:\PW
```

---

**NOTE:** The directory specification in the preceding example is for Windows NT or Windows 2000.

---

### Activating Data Cleansing in Real-Time Mode

In real-time mode, data cleansing is triggered when you step off the record after you have updated a field that has been defined as an active DataCleansing Field. An example of an active DataCleansing Field for the Contact business component is Last Name. Modifying an inactive field does not trigger data cleansing. For more information about business components, see [“Data Cleansing at the Business Component Level” on page 5-6](#).

### Troubleshooting

If data cleansing does not appear to be working in real-time mode, check the following areas.

- **License key.** Verify that the license key includes data quality.
- **Configuration file.** Verify that data cleansing has been designated. For more information, see [“To enable data cleansing” on page 5-2](#).

If you have configured a business component for data cleansing, also check the following:

- **Business component Class property.** Verify that the business component Class property is CSSBCBase.
- **DataCleansing Type.** Verify that the DataCleansing Type has been set correctly.
- **User Properties.** Verify that the DataCleansing Field has the correct User Property value and that the value is formatted correctly. For example, there must be a space after a comma in user properties that have a compound value. For more information, see [Table 5-1 on page 5-6](#).

## **Undoing and Disabling Data Cleansing**

You can undo and disable cleansing on a record-by-record basis. You can turn data cleansing off for specific records by checking the column labeled Disable Data Cleansing in various list and multivalue group applets. You can add a Disable Cleansing column to other list and multivalue group applets.

### ***To undo changes made during cleansing in real-time***

- Before the record is saved, choose Edit > Undo.

### ***To prevent a record from being cleansed in real-time or batch mode***

- 1** Navigate to and select the record.
- 2** Check the Disable Cleansing check box.

# Data Cleansing at the Business Component Level

You can configure data cleansing for any business component whose purpose is to hold name, address, or company (account) information. The following data is configured for data cleansing:

- Name and Location for accounts
- First Name, Middle Name, Last Name, and Job Title for contacts and prospects
- Street Address, City, State, and Postal Code for business address records

## Business Component User Properties

Data cleansing is configured in two kinds of business component user properties:

- DataCleansing Field
- DataCleansing Type

One DataCleansing Type user property and one or more DataCleansing Field user properties are required.

Business Component user properties are child object definitions of the business component. Each has a Name property, which specifies the user property to be set, and a Value property, which specifies the value for the user property. [Table 5-1](#) shows the set of child Contact Business Component User Property object definitions that are used for data cleansing configuration.

**Table 5-1. Contact User Properties for Data Cleansing**

| User Property Name    | User Property Value               |
|-----------------------|-----------------------------------|
| DataCleansing Field 1 | "Last Name", "Last Name"          |
| DataCleansing Field 2 | "First Name", "First Name"        |
| DataCleansing Field 3 | "Middle Name", "Middle Name"      |
| DataCleansing Field 4 | "Professional Title", "Job Title" |
| DataCleansing Type    | Contact                           |

## DataCleansing Type

The DataCleansing Type user property is required. It specifies to Siebel Connector for Firstlogic Libraries what kind of data is being validated. It has one of the following values:

- **Contact.** Indicates that the data consists of person name records.
- **Account.** Indicates that it consists of business or office name records.
- **Address.** Indicates that it consists of postal addresses.

Data cleansing operates differently on each of these types. For example, business components with Address cleansing have reconciliation performed between address fields and the ZIP (Postal Code) field. All values have capitalization validated.

## DataCleansing Field

Each DataCleansing Field *n* property specifies a correspondence between a field name in the Firstlogic application and a field name in a Siebel application. The Value property consists of a pair of text strings. They are separated by a comma and a space. Each string is enclosed in double quotation marks.



**Caution:** If the configuration is not correct, data cleansing will not take place. For example, if there is *not* a comma and a space between the text strings in the Value user property, or if DataCleansing Type was not defined, the records for the corresponding business component will not be cleansed.

You can add or remove data cleansing functionality for new and existing fields by setting DataCleansing Field user properties.



|  |      |
|--|------|
| About Batch Mode .....   | 6-2  |
| Customized Request Components .....  | 6-3  |
| Checking the Data Quality Manager Component .....                                | 6-4  |
| Creating New Request Components .....  | 6-5  |
| Activating New Components .....  | 6-7  |
| Submitting a Server Request .....  | 6-8  |
| Starting Siebel Data Quality Manager Server Component .....                      | 6-11 |
| Data Cleansing Parameters for Siebel<br>Connector for Firstlogic Libraries ..... | 6-12 |
| Component Parameters for Account Data Cleansing .....                            | 6-12 |
| Component Parameters for Contact Data Cleansing .....                            | 6-13 |
| Component Parameters for Prospect Data Cleansing .....                           | 6-13 |
| Component Parameters for Address Data Cleansing .....                            | 6-14 |
| Data Matching Parameters for Siebel Data Quality Matching .....                  | 6-15 |
| Component Parameters for Account Key Generation .....                            | 6-16 |
| Component Parameters for Contact Key Generation .....                            | 6-16 |
| Component Parameters for Prospect Key Generation .....                           | 6-17 |
| Component Parameters for Account Key Refresh .....                               | 6-17 |
| Component Parameters for Contact Key Refresh .....                               | 6-18 |
| Component Parameters for Prospect Key Refresh .....                              | 6-18 |
| Additional Parameters for Data Matching .....                                    | 6-19 |
| Component Parameters for Account DeDuplication .....                             | 6-19 |
| Component Parameters for Contact DeDuplication .....                             | 6-20 |
| Component Parameters for Prospect DeDuplication .....                            | 6-20 |
| Conflict_IDs .....   | 6-21 |

# About Batch Mode

Batch mode lets you process a large number of records at one time. You can run batch mode jobs as standalone tasks or schedule them on a recurring basis. You can run both data cleansing and data matching in batch mode.

- **Batch data cleansing.** Use batch data cleansing to standardize and correct a group of accounts, contacts, prospects, or business addresses.
- **Batch data matching.** In batch mode, the application identifies potentially-conflicting records and presents them to the system administrator for resolution.

If you run data cleansing on a record twice, sometimes it will change the second time. However, cleansing the entire database each time can cause performance issues. For that reason, it is recommended that only new or recently-modified records be included in the batch data cleansing process. These records can be identified using SQL Object Where clauses like the sample ones in [Table 4-5 on page 4-26](#). The same rule applies to data matching. For more information about general performance considerations, see [“Ongoing Deduplication and Key Generation Operations” on page 4-26](#).

Running data cleansing and data matching in batch mode requires the use of a separate server process called Siebel Data Quality Manager. You initiate batch data cleansing or data matching by starting a new server task, as described in [“Submitting a Server Request” on page 6-9](#).



**Caution:** Do not put VB logic in any business components that are being used for batch mode tasks. These tasks execute in the background and may not trigger any logic that activates user interface features, such as pop-up message boxes.

## Customized Request Components

Data quality components allow you to preset Siebel Data Quality Manager parameters, rather than setting them each time you run the task. [Table 6-1](#) lists the components and their parameters. After you customize a component, it is saved with the correct parameters for the task.

**Table 6-1. Siebel Data Quality Manager Components**

| Component Name                      | Business Component Name       | Business Object Name | Operation Type           |
|-------------------------------------|-------------------------------|----------------------|--------------------------|
| Account Data Cleansing              | Account                       | Account              | Data cleansing           |
| Account DeDuplication               | Account                       | Account              | Matching (deduplication) |
| Contact Data Cleansing              | Contact                       | Contact              | Data cleansing           |
| Contact DeDuplication               | Contact                       | Contact              | Matching (deduplication) |
| Prospect Data Cleansing             | List Mgmt Prospective Contact | List Mgmt            | Data cleansing           |
| Prospect DeDuplication              | List Mgmt Prospective Contact | List Mgmt            | Matching (deduplication) |
| <sup>1</sup> Address Data Cleansing | Business Address              | Business Address     | Data cleansing           |

1. Data matching is not available for addresses.

The process of creating a customized component is as follows:

- Verify that the Data Quality Manager component is enabled.
- Create the new components.
- Activate the new components.

## Checking the Data Quality Manager Component

Before you can customize a component, you must verify that the Data Quality Manager component has been enabled.

### **To verify that Data Quality Manager is enabled**

- 1 From the application-level menu, choose View > Site Map > Server Administration > Enterprise Operations.
- 2 Click the Component Requests tab.
- 3 In the form, click the Component/Job field, create a query to locate data Quality Manager, and verify that it has been enabled.

If you need to enable the server component, see [“Starting Siebel Data Quality Manager Server Component” on page 6-12](#) for instructions.

The screenshot shows two overlapping windows from the Siebel application. The background window is titled 'Component Requests' and displays a table of request records. The foreground window is titled 'Components/Jobs - Microsoft Internet Explorer' and shows a search result for the 'Data Quality Manager Component'.

| ID        | Component/Job       | Mode         | Status  | Completion Code | Completion Infor | Server | Request Key |
|-----------|---------------------|--------------|---------|-----------------|------------------|--------|-------------|
| 10-2PR1-6 | Optimization Engine | Asynchronous | Queued  |                 |                  |        | 10-2PCV50   |
| 10-2PR1-5 | Optimization Engine | Asynchronous | Success | 0               |                  | HPANI  | 10-2PCV50   |
| 10-2PR1-4 | Optimization Engine | Asynchronous | Success | 0               |                  | HPANI  | 10-2PCV50   |
| 10-2PR1-3 | Optimization Engine | Asynchronous | Success | 0               |                  | HPANI  | 10-2PCV50   |
| 10-2PR1-2 | Optimization Engine | Asynchronous | Success | 0               |                  | HPANI  | 10-2PCV50   |
| 10-2PR1-1 | Optimization Engine | Asynchronous | Queued  |                 |                  |        | 10-2PCV50   |
| 10-2EVI-5 | Optimization Engine | Asynchronous |         |                 |                  |        |             |

| Name                          | Type | Component | Enabled? | Description                             |
|-------------------------------|------|-----------|----------|---|
| Data Quality Manage Component |      |           | ✓        | Cleanses data and de-duplicates records |

## Creating New Request Components

To create a customized component, you create a component record and modify the parameters associated with it, as noted in [Table 6-2](#). Note that you are not creating new parameters, but modifying the values of existing ones.

**Table 6-2. Component Request Parameter Values (1 of 2)**

| Parameter Name        | Description   |
|-----------------------|---|
| Buscomp Name          | <p>Required.</p> <p>Name of the business component.</p> <p>Values are:</p> <ul style="list-style-type: none"> <li>■ Account</li> <li>■ Contact</li> <li>■ List Mgmt Prospective Contact</li> <li>■ Business Address</li> </ul>                          |
| Business Object Name  | <p>Required.</p> <p>Name of the business object with which the business component is associated.</p> <p>Values are:</p> <ul style="list-style-type: none"> <li>■ Account</li> <li>■ Contact</li> <li>■ List Mgmt</li> <li>■ Business Address</li> </ul> |
| Object Sorting Clause | <p>Optional.</p> <p>Used during matching to identify a specific deduplication token.</p>  |

**Table 6-2. Component Request Parameter Values (2 of 2)**

| <b>Parameter Name</b> | <b>Description</b>  |
|-----------------------|---|
| Object Where Clause   | Optional.<br>Searched by fields that exist in the specified business component.<br>Example: Disable DataCleansing = N or Disable DataCleansing IS NULL  |
| Operation Type        | Required.<br>Type of operation to be performed.<br>Values are: <ul style="list-style-type: none"> <li>■ Data Cleansing</li> <li>■ DeDuplication</li> <li>■ Key Generate</li> <li>■ Key Refresh</li> </ul> |

**To create a customized request component**

- 1** From the application-level menu, choose View > Site Map > Server Administration > Enterprise Configuration.
- 2** Click the Component Definitions tab, add a new component, and enter the following values in the component definition fields:

| <b>Component Definition</b> | <b>Value</b>   |
|-----------------------------|--|
| Request Component Name      | Select a name for the component, such as Account Data Cleansing. |
| Component Type              | Dqmgr  |
| Component Group             | Data Quality   |
| Alias                       | Alias that corresponds with the business component name.         |

- 3** Save the record.

- 4 In the lower Component Definitions tab, query to locate the Buscomp Name and Business Object Name parameters, and change their values as described in [Table 6-2 on page 6-6](#).

Enterprise Component Definitions:

| Name                                     | Component Type     | Definition State | Component Group    | Description           | Alias              | Run Mode   |
|--|--------------------|------------------|--------------------|-----------------------|--------------------|------------|
| Account Data Clean Dqmgr                 |                    | Active           | Data Quality       |                       | DqMgr AcctDCIns    | Batch      |
| D&B Update Mgr (Mkt DNB Update           |                    | Active           | Dun and Bradstreet | Updates D&B or Sie    | DNBUprMgnMultiTask | Batch      |
| Data Dictionary Man Data Dictionary Serv |                    | Active           | Marketing          | Connects to External  | DataDictMgr        | Batch      |
| Data Quality Manage Dqmgr                |                    | Active           | Data Quality       | Cleanses data and c   | DQMgr              | Batch      |
| Database Extract                         | DbXtract           | Active           | Siebel Remote      | Extracts visible data | DbXtract           | Batch      |
| DCommerce Alerts                         | DCommerce Alerts   | Active           | DCommerce          | Background proces     | DCommerceAlerts    | Background |
| DCommerce Automate                       | DCommerce Automate | Active           | DCommerce          | Background proces     | DCommerceAutoClo   | Background |

| Parameter            | Fixed | Value   | Data Type | Parameter Type | Description                |
|----------------------|-------|---------|-----------|----------------|----------------------------|
| Buscomp Name         |       | Account | String    | Component      | Name of buscomp to be DQed |
| Business Object Name |       | Account | String    | Component      | Name of business object    |

- 5 In the lower Component Definitions tab, query to locate the Operation Type and change its value as described in [Table 6-2 on page 6-6](#).
- 6 Click the Component Definitions menu in the upper form and select Enable Component Definition.

## Activating New Components

After you have created all the request components, you need to activate them.

### To activate new request components

- 1 Restart the server.
- 2 From the application-level menu, choose View > Site Map > Server Administration > Enterprise Configuration.
- 3 On the Batch Component tab, click Synchronize.

## Submitting a Server Request

You can start a batch process from the Server Administration screen or the command line.

### To start a server request from the Server Administration screen

- 1 From the application-level menu, choose View > Site Map > Server Administration > Enterprise Operations.
- 2 Click the Component Request tab and then select Submit Request from the Comp Reqs menu, as shown in the following figure.

The screenshot displays the Siebel Server Administration interface. At the top, the 'Show:' dropdown is set to 'Enterprise Operations' and the 'History' dropdown is selected. A 'Favorites:' dropdown is also visible. Below this, a 'Comp Reqs:' menu is open, showing options like 'New Record', 'Edit Record', 'Save Record', etc. The main area shows a table of component requests with columns: Job, Mode, Status, Completion Code, Completion Info, Server, and Request Key. The table contains several rows, with the first row highlighted in yellow. Below the table, there are tabs for 'Enterprise Component Tasks', 'Enterprise Tasks', 'Component Group Assignment', 'Component Requests', and 'Repeating Component R'. The 'Component Requests' tab is active, showing a detailed view of a request with fields for ID, Status, Component/Job, Component, Scheduled Start, Expiration, Delete Interval, Delete Unit, Server, Request Key, Mode, Type, Submit Date, Actual Start, and End Date.

| Job | Mode         | Status  | Completion Code | Completion Info   | Server     | Request Key |
|-----|--------------|---------|-----------------|-------------------|------------|-------------|
| 1   | Asynchronous | Success | 0               |                   | smtb50a012 |             |
| 1   | Asynchronous | Error   | 5700061         | SRB-00061: proces | smtb50a012 |             |
| 1   | Asynchronous | Success | 0               |                   | smtb50a012 |             |
| 1   | Asynchronous | Success | 0               |                   | smtb50a012 |             |
| 1   | Asynchronous | Success | 0               |                   | smtb50a012 |             |
| 1   | Asynchronous | Success | 0               |                   | smtb50a012 |             |

**\*ID:** 1-4RC-7D  
**Status:** Success  
**Component/Job:** List Manager  
**Component:**

**Scheduled Start:** 10/08/01 1:10:14 PM  
**Expiration:**  
**Delete Interval:** 0  
**Delete Unit:**

**Server:** smtb50a012  
**Request Key:**  
**Mode:** Asynchronous  
**Type:** Component

**\*Submit Date:** 10/08/01 1:10:14 PM  
**Actual Start:** 10/08/01 1:52:14 PM  
**End Date:** 10/08/01 1:52:39 PM

**To run Siebel Data Quality Manager from the command line**

- 1 From the /bin directory on the Siebel Server machine, type the following:

```
srvrmgr /g <gateway server name> /e <enterprise server name> /  
u <username> /p <password> /s <Siebel server name> /l  
<language, ENU for English>
```

- 2 Type the following for each batch mode operation:

```
run task for comp DQMgr with BCName = <Buscomp Name>, BobjName  
= <Business Object Name>, OpType = <Operation Type>,  
ObjWhereClause = <Object Where Clause>, ObjSortClause = <Object  
Sorting Clause>
```

For example:

```
run task for comp DQMgr with  
BCName="Account",BObjName="Account",OpType="Data Cleansing"
```

## Batch Mode

### Customized Request Components

#### To check the status of the request

- 1 From the application-level menu, choose View > Site Map > Server Administration > Server.
- 2 Select Siebel Server from the Show drop-down list and click the Server Tasks tab.

The status of the server tasks displays in the Status column, as seen in the following figure.

The screenshot displays the Siebel Server Administration interface. At the top, there is a navigation bar with 'Show: Servers' and a search field containing 'History'. Below this, the 'Siebel Servers' section is active, showing a table with one entry for 'smtb50a012' in a 'Running' state. The 'Server Tasks' tab is selected, displaying a table of tasks with columns for Task, Component, Task State, Status, PID, Start Time, and End Time. The tasks listed include 'Sales Object Manag Running' (Status: Waiting for comman), 'Server Manager Running' (Status: Processing \*List Ta), and 'Call Center Object N Running' (Status: Handling Request).

| Siebel Server | Server State | Host Name  | SiebSrvr PID | Start Time          | End Time |
|---------------|--------------|------------|--------------|---------------------|----------|
| smtb50a012    | Running      | smtb50a012 | 15848        | 10/09/01 11:27:27 # |          |

| Task  | Component            | Task State | Status              | PID    | Start Time          | End Time |
|-------|----------------------|------------|---------------------|--------|---------------------|----------|
| 10263 | Sales Object Manag   | Running    | Waiting for comman  | 19,536 | 10/09/01 11:34:19 # |          |
| 10262 | Sales Object Manag   | Running    | Waiting for comman  | 19,536 | 10/09/01 11:34:15 # |          |
| 10261 | Sales Object Manag   | Running    | Waiting for comman  | 19,536 | 10/09/01 11:34:07 # |          |
| 10260 | Server Manager       | Running    | Processing *List Ta | 12,956 | 10/09/01 11:30:43 # |          |
| 10259 | Call Center Object N | Running    | Handling Request    | 17,410 | 10/09/01 11:29:42 # |          |
| 10258 | Call Center Object N | Running    | Waiting for comman  | 17,410 | 10/09/01 11:29:18 # |          |
| 10257 | Server Request Pro   | Running    |                     | 7,400  | 10/09/01 11:28:14 # |          |

Figure 6-1. Checking the Status of Server Tasks

## Starting Siebel Data Quality Manager Server Component

After the components are created, you need to start the Siebel Data Quality Manager server component.



**Caution:** You must log in with an administrator ID—such as SADMIN—to enable the data-cleansing server task.

### Starting the server component

- 1 From the application-level menu, choose View > Site Map > Server Administration > Servers.
- 2 In the Siebel Servers applet, verify that the Siebel Server is running.
- 3 In the Server Component Groups applet, query for the Data Quality Component Group.
- 4 If the Run state of the group is Shutdown, click Startup to start the Data Quality Component Group.

Server Component Groups:

Server Component Groups

Siebel Servers

Query Enterprise Server: smtb50a012 1 - 1 of 1

| Siebel Server | Server State | Host Name  | SiebSrvr PID | Start Time          | End Time |
|---------------|--------------|------------|--------------|---------------------|----------|
| smtb50a012    | Running      | smtb50a012 | 15848        | 10/09/01 11:27:27 # |          |

Server Component Groups

Server Components Server Tasks Server Parameters Server Event Configuration Server Statistics Server Info Log

Offline Startup Shutdown 1 - 6 of 6

| Component Group      | Run State | # of Components | Enable State | Description                                   |
|----------------------|-----------|-----------------|--------------|---|
| Data Quality         | Online    | 1               | Enabled      | Data Quality Components                       |
| Enterprise Applicati | Online    | 1               | Enabled      | Enterprise Application Integration Components |
| Marketing            | Online    | 1               | Enabled      | Marketing Components                          |
| Siebel Call Center   | Running   | 1               | Enabled      | Siebel Center Components                      |
| Siebel Sales         | Online    | 1               | Enabled      | Siebel Sales Components                       |
| System Management    | Running   | 6               | Enabled      | System Management Components                  |

## Data Cleansing Parameters for Siebel Connector for Firstlogic Libraries

To cleanse data with Siebel Connector for Firstlogic Libraries, add the following four parameters:

- Account data cleansing (alias DqMgrAcctDCIns)
- Contact data cleansing (alias DqMgrContDCIns)
- Prospect data cleansing (alias (DqMgrPrspDCIns)
- Address data cleansing (alias DqMgrAddrDCIns)

### Component Parameters for Account Data Cleansing

The component parameters for account data cleansing and their values are listed in [Table 6-3](#).

**Table 6-3. Component Parameters for Account Data Cleansing**

| Component Parameter   | Value  |
|-----------------------|--|
| Buscomp Name          | Account  |
| Business Object Name  | Account  |
| Object Sorting Clause |  |
| Object Where Clause   | Disable DataCleansing = N<br>or<br>Disable DataCleansing IS NULL |
| Operation Type        | Data Cleansing   |

## Component Parameters for Contact Data Cleansing

The component parameters for contact data cleansing and their values are listed in [Table 6-4](#).

**Table 6-4. Component Parameters for Contact Data Cleansing**

| Component Parameter   | Value  |
|-----------------------|--|
| Buscomp Name          | Contact  |
| Business Object Name  | Contact  |
| Object Sorting Clause |  |
| Object Where Clause   | Disable DataCleansing = N<br>or<br>Disable DataCleansing IS NULL |
| Operation Type        | Data Cleansing   |

## Component Parameters for Prospect Data Cleansing

The component parameters for prospect data cleansing and their values are listed in [Table 6-5](#).

**Table 6-5. Component Parameters for Prospect Data Cleansing**

| Component Parameter   | Value  |
|-----------------------|--|
| Buscomp Name          | List Mgmt Prospective Contact                                    |
| Business Object Name  | List Mgmt  |
| Object Sorting Clause |  |
| Object Where Clause   | Disable DataCleansing = N<br>or<br>Disable DataCleansing IS NULL |
| Operation Type        | Data Cleansing   |

## Component Parameters for Address Data Cleansing

The component parameters for address data cleansing and their values are listed in [Table 6-6](#).

**Table 6-6. Component Parameters for Address Data Cleansing**

| <b>Component Parameter</b> | <b>Value</b>   |
|----------------------------|--|
| Buscomp Name               | Business Address   |
| Business Object Name       | Business Address   |
| Object Sorting Clause      |  |
| Object Where Clause        | Disable DataCleansing = N<br>or<br>Disable DataCleansing IS NULL |
| Operation Type             | Data Cleansing   |

## Data Matching Parameters for Siebel Data Quality Matching

If you are using Siebel Data Quality Matching, you need to add these components:

- Key generation:
  - Account key generation (alias DqMgrAcctKGen)
  - Contact key generation (alias DqMgrContKGen)
  - Prospect key generation (alias DqMgrPrspKGen)
- Key refresh:
  - Account key refresh (alias DqMgrAcctKRef)
  - Contact key refresh (alias DqMgrContKRef)
  - Prospect key refresh (alias DqMgrPrspKRef)

These components are not used with Siebel Connector for Firstlogic Libraries.

Use the information in the following tables to specify these components, depending on the business object against which you want to run Siebel Data Quality Matching. Notice that the value of the Object Where Clause is blank. The reason is that it is a search specification whose value will vary. The following sample Object Where Clause value processes only those records with a value in the Last Name field that begins with an A, B, or C:

```
[Last Name] >= 'A' and [Last Name] <= 'C'.
```

## Component Parameters for Account Key Generation

The component parameters for account key generation and their values are listed in [Table 6-7](#).

**Table 6-7. Component Parameters for Account Key Generation**

| Component Parameter   | Value        |
|-----------------------|--------------|
| Buscomp Name          | Account      |
| Business Object Name  | Account      |
| Object Sorting Clause |              |
| Object Where Clause   |              |
| Operation Type        | Key Generate |

## Component Parameters for Contact Key Generation

The component parameters for contact key generation and their values are listed in [Table 6-8](#).

**Table 6-8. Component Parameters for Contact Key Generation**

| Component Parameter   | Value        |
|-----------------------|--------------|
| Buscomp Name          | Contact      |
| Business Object Name  | Contact      |
| Object Sorting Clause |              |
| Object Where Clause   |              |
| Operation Type        | Key Generate |

## Component Parameters for Prospect Key Generation

The component parameters for prospect key generation and their values are listed in [Table 6-9](#).

**Table 6-9. Component Parameters for Prospect Key Generation**

| Component Parameter   | Value                         |
|-----------------------|-------------------------------|
| Buscomp Name          | List Mgmt Prospective Contact |
| Business Object Name  | List Mgmt                     |
| Object Sorting Clause |                               |
| Object Where Clause   |                               |
| Operation Type        | Key Generate                  |

## Component Parameters for Account Key Refresh

The component parameters for account key refresh and their values are listed in [Table 6-10](#).

**Table 6-10. Component Parameters for Account Key Refresh**

| Component Parameter   | Value       |
|-----------------------|-------------|
| Buscomp Name          | Account     |
| Business Object Name  | Account     |
| Object Sorting Clause |             |
| Object Where Clause   |             |
| Operation Type        | Key Refresh |

## Component Parameters for Contact Key Refresh

The component parameters for contact key refresh and their values are listed in [Table 6-11](#).

**Table 6-11. Component Parameters for Contact Key Refresh**

| Component Parameter   | Value       |
|-----------------------|-------------|
| Buscomp Name          | Contact     |
| Business Object Name  | Contact     |
| Object Sorting Clause |             |
| Object Where Clause   |             |
| Operation Type        | Key Refresh |

## Component Parameters for Prospect Key Refresh

The component parameters for prospect key refresh and their values are listed in [Table 6-12](#).

**Table 6-12. Component Parameters for Prospect Key Refresh**

| Component Parameter   | Value                         |
|-----------------------|-------------------------------|
| Buscomp Name          | List Mgmt Prospective Contact |
| Business Object Name  | List Mgmt                     |
| Object Sorting Clause |                               |
| Object Where Clause   |                               |
| Operation Type        | Key Refresh                   |

## Additional Parameters for Data Matching

The following parameters for matching apply to both Siebel Connector for Firstlogic Libraries and Siebel Data Quality Matching:

- Account DeDuplication (alias DqMgrAcctDDup)
- Contact DeDuplication (alias DqMgrContDDup)
- Prospect DeDuplication (alias DqMgrPrspDDup)

### Component Parameters for Account DeDuplication

The component parameters for account deduplication and their values are listed in [Table 6-13](#).

**Table 6-13. Component Parameters for Account DeDuplication**

| Component Parameter   | Value  |
|-----------------------|--|
| Buscomp Name          | Account  |
| Business Object Name  | Account  |
| Object Sorting Clause | DeDup Token (for Siebel Connector for Firstlogic Libraries)  |
| Object Where Clause   |  |
| Operation Type        | DeDuplication or Cleansing<br>(This parameter is required. If it is missing, the server task will exit with an error.) |

## Component Parameters for Contact DeDuplication

The component parameters for contact deduplication and their values are listed in [Table 6-14](#).

**Table 6-14. Component Parameters for Contact DeDuplication**

| Component Parameter   | Value  |
|-----------------------|--|
| Buscomp Name          | Contact  |
| Business Object Name  | Contact  |
| Object Sorting Clause | DeDup Token (for Siebel Connector for Firstlogic Libraries)  |
| Object Where Clause   |  |
| Operation Type        | DeDuplication or Cleansing<br>(This parameter is required. If it is missing, the server task will exit with an error.) |

## Component Parameters for Prospect DeDuplication

The component parameters for prospect deduplication and their values are listed in [Table 6-15](#).

**Table 6-15. Component Parameters for Prospect DeDuplication**

| Component Parameter   | Value  |
|-----------------------|--|
| Buscomp Name          | Contact  |
| Business Object Name  | Account  |
| Object Sorting Clause | DeDup Token (for Siebel Connector for Firstlogic Libraries)  |
| Object Where Clause   |  |
| Operation Type        | DeDuplication or Cleansing<br>(This parameter is required. If it is missing, the server task will exit with an error.) |

## **Conflict\_IDs**

Conflict\_ID is a system field. It is part of the user key fields, fields that uniquely identify a record. For example, user key fields for an account record are Account Name, Account Location, and Account Organization. If two account records have the same values for these fields, it generates a User Key violation. Before running batch data cleansing, all the records in the database have different keys, although the difference between two records may be very slight, such as two records having ABC Corporation and ABC Corp respectively for Account Name.

Batch data cleansing corrects records and commits them to the database. However, if there is a potential of duplicate records, such as those in the ABC Corp example, the batch data cleansing process updates a hidden Conflict\_ID field with the Row\_ID of the record. Because this field is part of the user key fields, the record now has a unique key and can be committed to the database. Later, you can run batch data matching to catch and resolve these duplicate records. Using the Conflict\_ID as part of the user key prevents a User Key violation.

**Batch Mode**

*Conflict\_IDs*

# Index

## A

- accounts
  - data cleansing component
    - parameters 6-12
  - data matching fields 4-20 to 4-21
  - deduplication calculation 4-12
  - deduplication component
    - parameters 6-19
  - key generation component
    - parameters 6-16
  - key refresh component parameters 6-17
- ACE Library 3-2, 3-8
- ace.cfg, changes to 3-6
- aceaux.cfg
  - changes to 3-5
  - dictionary file entries 3-2
- address correction
  - data cleansing component
    - parameters 6-14
  - Siebel Connector for Firstlogic Libraries 1-6
  - Siebel Data Quality 1-2
- audience for guide Intro-2

## B

- batch mode
  - about 6-2
  - component parameters, about 6-5
  - customized components,
    - creating 6-6 to 6-7
  - data cleansing, disabling 5-5
  - data cleansing, multiple passes of 6-2

- enabling Siebel Data Quality Matching
  - for 2-4
- performance improvement 4-28
- server request, starting 6-8
- server requests, checking status of 6-10
- Siebel Data Quality Manager,
  - enabling 6-4
- VB logic, use of 6-2
- batch mode, Siebel Connector for Firstlogic Libraries
  - account data cleansing parameters 6-12
  - address data cleansing parameters 6-14
  - contact data cleansing parameters 6-13
  - data cleansing parameters and
    - aliases 6-12
  - installation considerations 3-4
  - prospect data cleansing parameters 6-13
- batch mode, Siebel Data Quality
  - server request, starting 6-9
  - starting server manager component 6-11
- business components
  - data cleansing permitted 4-17
  - data cleansing requirements for 5-6
  - data cleansing user properties 5-6 to 5-7
  - data matching, deactivating 4-12
  - data matching, deactivating by business component 4-12
  - deduplication fields 4-18
  - deduplication fields, about 4-14
  - DeDuplication Key BusComp 4-15
  - deduplication permitted 4-17
  - DeDuplication Results BusComp 4-15
  - performance considerations, batch components 4-28

**C**

- candidate set 4-14
- capitalization standards
  - Siebel Connector for Firstlogic Libraries 1-6
  - Siebel Data Quality 1-2
- child records 4-10
- child user properties, deactivating 4-12
- command line operation 6-9
- component parameters
  - account deduplication 6-19
  - contact deduplication 6-20
  - prospect deduplication 6-20
- component parameters, Siebel Data Quality
  - account key generation 6-16
  - account key refresh 6-17
  - contact key generation 6-16
  - Contact Key Refresh 6-18
  - prospect key generation 6-17
  - prospect key refresh 6-18
- concurrent server tasks 4-25
- configuration files
  - ace.cfg 3-6
  - aceaux.cfg 3-5
  - firm.cfg 3-7
  - firmindv.cfg 3-7
  - location 3-5
  - tn.cfg 3-7
  - tnaux.cfg 3-6
- Conflict\_ID 6-21
- contacts
  - data cleansing component parameters 6-13
  - data cleansing user properties 5-6
  - data matching fields 4-21 to 4-22
  - deduplication calculation 4-12
  - deduplication component
    - parameters 6-20
  - key generation component
    - parameters 6-16
  - key refresh component parameters 6-18
- customized components,
  - creating 6-6 to 6-7

**D**

- data cleansing
  - about enabling 5-2
  - about modes 4-2
  - address correction 1-2
  - business component restrictions 4-17
  - capitalization standards 1-2
  - components of 1-2
  - configuration file parameters 5-3
  - configuration, user properties 5-6 to 5-7
  - Conflict\_ID, role of 6-21
  - data requirements 5-6
  - DataCleansing Type 5-7
  - dirty data 4-17
  - disabling at record-level 5-5
  - enabling and disabling 5-5
  - field correspondence 5-7
  - format standardization 1-2
  - multiple passes 6-2
  - real-time mode, troubleshooting 5-4
  - triggering in real-time mode 5-4
  - undoing at record level 5-5
  - user property, about 5-7
- data cleansing, Siebel Connector for Firstlogic Libraries
  - account parameters 6-12
  - address parameters 6-14
  - altering cleansing behavior 3-8
  - contact parameters 6-13
  - functionality overview 1-6
  - parameters and aliases 6-12
  - prospect parameters 6-13
- data matching
  - See also* deduplication
  - Account fields 4-20 to 4-21
  - application and business component levels 4-12
  - batch mode, about starting 6-2
  - Conflict\_ID, role of 6-21
  - Contact fields 4-21 to 4-22
  - duplicate records, locating 4-9
  - duplicate records, resolving 4-11
  - enabling from configuration file 4-5

- enabling from Server Administration
  - screen 4-3
- enabling, overview 4-3
- field mapping 4-14
- key generation, role of 4-13
- match scores, role of 4-13
- merge operation, reversing 4-10
- products for data matching 1-3
- Prospect fields 4-22 to 4-23
- real-time mode, about 4-8
- real-time search and merge, process
  - overview 4-8
- Siebel Connector for Firstlogic Libraries,
  - altering cleansing behavior 3-8
- Siebel Connector for Firstlogic Libraries,
  - functionality overview 1-7
- Siebel Data Quality Matching,
  - functionality 1-4
- Siebel Data Quality Matching, process
  - overview 1-4
- data quality options, setting 4-5 to 4-6
- DataCleansing Field user property 5-6, 5-7
- DataCleansing Type user property 5-6, 5-7
- DB2 parameter optimization 4-25
- dedicated Web client
  - data cleansing, activating 5-3
  - data matching, enabling 4-5
- DeDup Key Modification Date field 4-19
- DeDup Keys field 4-19
- DeDup Last Match Date field 4-19
- DeDup Token field 4-18
- DeDup Token Value user property 4-13
- dedup token, about 1-7
- dedup tokens 4-12 to 4-13
- DEDUP\_REQ\_ID column 4-16
- DEDUP\_TOKEN column 4-18
- deduplication
  - See also* data matching
  - about modes 4-2
  - business component fields 4-18
  - business component restrictions 4-17
  - calculation values, changing 4-13
  - child records, position of 4-10
  - component parameters for
    - accounts 6-19
  - component parameters for contacts 6-20
  - component parameters for
    - prospects 6-20
  - DeDup Token field, about 4-18
  - dedup token, role in
    - process 4-12 to 4-13
  - Deduplication Match Score field 4-18
  - defined 1-2
  - dirty data 4-17
  - grandchild records, position of 4-10
  - ongoing operations 4-26
  - Possible Matches dialog box,
    - configuring 4-16
  - process overview, Siebel Connector for
    - Firstlogic Libraries 4-12 to 4-13
  - prospect data, special
    - requirements 4-16
  - real-time mode, about 4-8
  - recommended record set 4-26
  - record filtering 4-26
  - results business component,
    - sample 4-16
  - DeDuplication Field user property 4-14
  - DeDuplication Key BusComp user
    - property 4-15
  - Deduplication Match Score field 4-18
  - Deduplication Object Id field 4-18
  - DeDuplication Results Applet user
    - property 4-16
  - DeDuplication Results BusComp user
    - property 4-15, 4-16
  - DeDuplication Results list applet 4-16
  - Delete parameter 4-28, 4-29
  - dictionary installation 3-2
  - dirty data 4-17
  - DqMgrAcctDCIns 6-12
  - DqMgrAcctDDup 6-19
  - DqMgrAcctKGen 6-15
  - DqMgrAcctKRef 6-15
  - DqMgrAddrDCIns 6-12
  - DqMgrContDCIns 6-12

- DqMgrContDDup 6-19
- DqMgrContKGen 6-15
- DqMgrContKRef 6-15
- DqMgrPrspDClns 6-12
- DqMgrPrspDDup 6-19
- DqMgrPrspKGen 6-15
- DqMgrPrspKRef 6-15
- DQSetting parameter, performance implications 4-29
- duplicate records
  - deleting from S\_DEDUP\_RESULT table 4-18
  - locating 4-9
  - locating and merging, process overview 4-8
  - merging methods, compared 4-10
  - resolving 4-11

## **F**

- features, new features in this release Intro-5
- field mappings
  - about creating 4-14
  - deleting 4-15
- firm.cfg, changes to 3-7
- firmindv.cfg, changes to 3-7
- format standardization 1-2, 1-6

## **G**

- grandchild records 4-10

## **K**

- key generation
  - component parameters for
    - accounts 6-16
    - component parameters for contacts 6-16
    - component parameters for prospects 6-17
  - Contact component parameters 6-16
  - ongoing operations 4-26
  - prospect component parameters 6-17
  - recommended record set 4-26

- required components, Siebel Data Quality Matching 6-15
- role in data matching 4-13
- Siebel Connector for Firstlogic Libraries 1-7
- Siebel Data Quality Matching 1-4
- key refresh 4-19
  - account key refresh component parameters 6-17
  - component parameters for
    - accounts 6-17
    - component parameters for contacts 6-18
    - component parameters for prospects 6-18
  - contact key refresh component parameters 6-18
  - prospect key refresh component parameters 6-18
  - required components, Siebel Data Quality Matching 6-15
- key table, deleting keys 4-29
- key type
  - performance improvement 4-27
  - relation to table size 4-24

## **L**

- Last Cleanse Date field 4-19
- Last Update - SDQ field 4-19

## **M**

- Match Library 3-2, 3-8
- match scores, role in data matching 4-13
- merged records
  - about 4-8
  - merge operation 4-10
  - obsolete records 4-18
  - Sequence field 4-10
  - undoing a merge 4-10

## **N**

- n3sgsb.dll 2-4
- Name property 5-6

**O**

OBJ\_ID column 4-18  
 object manager 4-3  
 Object Sort clause 4-27  
 Object Where clause, samples 4-26  
 outdated records 4-26

**P**

performance  
   batch mode, tuning 4-28  
   concurrent server tasks 4-25  
   degradation, investigating 4-25  
   DQSetting parameter, implications  
     of 4-29  
   general settings, tuning 4-27  
   Object Sort clause 4-27  
   Object Where clause samples 4-26  
   run\_stats utility 4-25  
 Possible Matches dialog box,  
   configuring 4-16  
 prospects  
   data cleansing component  
     parameters 6-13  
   data matching fields 4-22 to 4-23  
   deduplication calculation 4-13  
   deduplication component  
     parameters 6-20  
   key generation component  
     parameters 6-17  
   key refresh component parameters 6-18  
   special configuration requirements 4-16

**R**

real-time mode  
   about 4-8  
   data cleansing, disabling 5-5  
   data cleansing, triggering 5-4  
   data cleansing, troubleshooting 5-4  
   data matching, process overview 4-8  
   duplicate records, locating 4-9  
   duplicate records, resolving 4-11  
 release, new features Intro-5

request components  
   about customizing 6-3  
   activating 6-7  
 Results applet, configuring for real-time  
   mode 4-16  
 Results business component, naming 4-16  
 Results table  
   Delete parameter performance  
     improvement 4-28  
   deleting keys 4-29  
 retained record, about 4-8  
 run\_status utility 4-25

**S**

S\_DEDUP\_RESULT table  
   deduplication match score 4-18  
   deleting duplicate records 4-18  
   performance improvement 4-28  
   recommended size 4-24  
   Results applet dialog box 4-16  
 S\_ORG\_DEDUP\_KEY, recommended  
   size 4-24  
 S\_PER\_DEDUP\_KEY, recommended  
   size 4-24  
 S\_PRSP\_DEDUP\_KEY, recommended  
   size 4-24  
 Sequence field 4-10  
 server requests  
   checking status of 6-10  
   starting 6-8  
 server tasks, concurrent 4-25  
 Siebel Connector for Firstlogic Libraries  
   account data cleansing parameters 6-12  
   account deduplication calculation 4-12  
   account deduplication component  
     parameters 6-19  
   ace.cfg 3-6  
   aceaux.cfg 3-5  
   address correction, about 1-6  
   address data cleansing parameters 6-14  
   batch mode installation  
     considerations 3-4

- cleansing and matching behavior,
    - altering 3-8
  - configuration file location 3-5
  - contact data cleansing parameters 6-13
  - contact deduplication component
    - parameters 6-20
  - contacts deduplication calculation 4-12
  - data cleansing parameters and
    - aliases 6-12
  - deduplication process overview 4-12
  - dictionary installation 3-2
  - firm.cfg 3-7
  - firmindv.cfg 3-7
  - format standardization 1-6
  - functionality 1-5
  - installation recommendations 3-2
  - modes of operation 4-2
  - prospect data cleansing parameters 6-13
  - prospect deduplication component
    - parameters 6-20
  - prospects deduplication
    - calculation 4-13
  - root directory location 3-3
  - Siebel eBusiness Application installation
    - options 3-3 to 3-4
  - subaccount deduplication
    - calculation 4-12
  - tn.cfg 3-7
  - tnaux.cfg 3-6
  - Siebel Data Quality Manager
    - command line 6-9
    - component parameters 6-5
    - enabling 6-4
    - parameters, preset 6-3
    - starting server component 6-11
  - Siebel Data Quality Matching
    - account deduplication component
      - parameters 6-19
    - account key generation component
      - parameters 6-16
    - account key refresh component
      - parameters 6-17
    - contact deduplication component
      - parameters 6-20
    - contact key generation component
      - parameters 6-16
    - contact key refresh component
      - parameters 6-18
    - deduplication tokens 4-13
    - enabling for batch mode 2-4
    - files installed 2-4
    - functionality 1-4
    - installation requirements 2-3 to 2-4
    - modes of operation 4-2
    - prospect deduplication component
      - parameters 6-20
    - prospect key generation component
      - parameters 6-17
    - prospect key refresh component
      - parameters 6-18
    - required components 6-15
    - software requirements 2-2
  - Siebel Data Quality, options,
    - setting 4-5 to 4-6
  - Siebel Dedicated Web Client, enabling
    - Siebel Data Quality Matching 2-3
  - Siebel technical support Intro-5 to Intro-6
  - Siebel Web Client, enabling Siebel Data Quality Matching 2-3
  - SIEBEL\_4K table space 4-24
  - SSA Deduplication Field property 4-14
  - SSA Match Purpose user property 4-15
  - subaccounts, deduplication
    - calculation 4-12
- T**
- table size recommendations 4-24
  - technical support Intro-5 to Intro-6
  - tn.cfg, changes to 3-7
  - tnaux.cfg, changes to 3-6
  - TOT\_SCORE\_VAL column 4-18
  - TrueName Library 3-2, 3-8

**U**

- User Key violation 6-21
- user properties 4-15
  - contact data cleansing 5-6
  - DataCleansing Field 5-6
  - DataCleansing Type 5-7
  - DataCleansingField 5-7
  - DeDup Token Value 4-13
  - DeDuplication Field 4-14
  - DeDuplication Key BusComp 4-15
  - DeDuplication Results Applet 4-16
  - DeDuplication Results BusComp 4-15,  
4-16
  - SSA Match Purpose 4-15

**V**

- Value property 5-6

**Z**

- ZIP + 4 postal codes 1-6

