Siebel Data Quality Administration Guide

Version 7.7
May 2004

Siebel Systems, Inc., 2207 Bridgepointe Parkway, San Mateo, CA 94404

Contents

Chapter 6: Configuring and Using Data Cleansing

Chapter 7: Working with Data Cleansing and Data Matching in Real-Time and Batch Modes

Chapter 8: Data Quality Configuration Options

1 What's New in This Release

What's New in Siebel Data Quality Administration Guide, Version 7.7

Table 1 lists changes described in this version of the documentation to support release 7.7 of the software.

Table 1. New Product Features in Siebel Data Quality Administration Guide, Version 7.7

Topic	Description
Installation guidelines See "About Installing the SDQ Matching Server" on page 19 and "About Installing the SDQ Universal Connector" on page 25.	Updated guidelines for installing the Siebel Data Quality (SDQ) product modules—the SDQ Matching Server and the SDQ Universal Connector.

This version of the documentation also contains the following change:

- Addition of a new appendix about how to use the Value Match method. See Appendix C, "Using the Value Match Method."

2 Siebel Data Quality Product Overview

This chapter provides an overview of the two Siebel Data Quality product modules—Siebel Data Quality (SDQ) Matching Server and the Siebel Data Quality (SDQ) Universal Connector—as well as a general overview of their functionality and modes of operation. It includes the following topics:

Siebel Data Quality Product Modules

The two product modules available for performing data quality functions within the Siebel enterprise are the Siebel Data Quality (SDQ) Matching Server and the Siebel Data Quality (SDQ) Universal Connector. Table 2 provides an overview and comparison of these two modules.

Table 2. Siebel Data Quality Product Overview

Function	SDQ Matching Server	SDQ Universal Connector[1]
Data matching (deduplication)[2] ■ Real-time mode ■ Batch mode	Yes	Yes
Data cleansing[2]	No	Yes
Identifies duplicate records stored in accounts, contacts, and prospects data table	Yes	Yes
Allows merging of duplicate records	Yes	Yes
Supports multiple languages and platforms	Yes	Yes[3]
Uses prebuilt Siebel business services	Yes	Yes
Requires additional third-party software	No	Yes

1. Driven by capabilities and configuration of third-party software. You must license additional software from third-party vendors to use this data quality product module.
2. For account, contact, and prospect data within the Siebel application.
3. The Siebel Data Quality Universal Connector architecture supports multiple languages and platforms. However, Firstlogic software does not support multiple languages.

CAUTION: You can use both data quality product modules concurrently in certain configurations. However, you cannot simultaneously enable data matching with the SDQ Matching Server and the SDQ Universal Connector for the same Siebel object manager.

In addition to using the SDQ Universal Connector to enable data cleansing, you can also enable data cleansing through the Universal Application Network (UAN) business processes. For more information about how to enable data cleansing through UAN, see the documentation for Siebel Universal Application Network.

About Data Matching and Data Cleansing

The name and address data stored in account, contact, and prospect records in Siebel eBusiness Applications represents your existing and potential customers. Because of the importance of this data, maintaining its integrity is critical.

Siebel Data Quality (SDQ) helps enterprises standardize and consolidate their account, contact, and prospect data in the following ways:

- **Data matching (deduplication).** Data matching identifies possible duplicate-record matches for account, contact, and prospect records, based on administrator-defined parameters. You can merge duplicate records into a single record using the Administration - Data Quality views. This guide generally uses the term data matching, or just matching, for this functionality. The term *deduplication* is also used in some circumstances to describe matching, typically in discussing matches in configuration files, user properties, and other system parameters. You can use the Siebel Data Quality (SDQ) Matching Server or the Siebel Data Quality (SDQ) Universal Connector to perform data matching tasks.
- **Data cleansing**. Data cleansing standardizes the structure of data in the customer records, and is used to standardize name and address information. You use the SDQ Universal Connector to perform data cleansing tasks. Data cleansing typically consists of the following functions:
  - **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, address correction provides ZIP+4 data correction and stores the data in certified U.S. Postal Service format.
  - **Capitalization**. Account, contact, and prospect names are converted to mixed case (uppercase and lowercase letters). Address fields are converted to mixed case, all lowercase, or all uppercase.
  - **Standardization**. Account, contact, and prospect information is stored in a uniform and consistent format.

  You can extend data cleansing functionality to include modifying or enhancing fields from within a customer profile by using the capabilities of external vendors, such as demographic, psychographic, or geocode attributes. Geocode is a standard set of information that many companies sell, including latitude and longitude coordinates, and other location information.

About Siebel Data Quality Modes of Operation

The Siebel Data Quality (SDQ) Matching Server and the Siebel Data Quality (SDQ) Universal Connector operate in real-time and batch mode.

Real-time mode. This mode provides an interactive pop-up window that appears whenever the user attempts to commit account, contact, or prospect data to the database, and detects a possible match with existing data. Whenever a user is entering a new customer record or saving changes to an existing record, the possible duplicates are displayed to that user in real time, allowing the user to select an existing record to continue the task with the correct customer profile. This avoids the creation of duplicate data during data entry.

Batch mode. This mode provides a server administration screen and command-line server manager utility that allows you to submit server component jobs for executing batch tasks. This mode also provides a set of data quality administration views so the administrator can review all the possible duplicates in the database. Within these views, the administrator can review the match scores and other record information for each match candidate and manually merge duplicate records.

3 Siebel Data Quality Matching Server

This chapter provides an overview of the Siebel Data Quality (SDQ) Matching Server functionality and installation guidelines. It includes the following topics:

About the SDQ Matching Server

The Siebel Data Quality (SDQ) Matching Server provides an embedded matching engine that is capable of identifying potential duplicate data within existing accounts, contacts, and prospects in the Siebel transactional database. This matching server is an embedded component that runs within the DeDuplication business service in the Siebel application and does not require third-party software installations to function.

The SDQ Matching Server allows configuration of relevant input fields for data matching (deduplication) using Siebel Tools. The matching server also provides the ability to manually merge duplicate records into a single record through the Administration - Data Quality views. The SDQ Matching Server works across languages and platforms supported by Siebel eBusiness Applications.

For more information about:

- Specifying input fields, see "SDQ Matching Server Rule Modification" on page 18.
- Merging duplicate records, see "Process of Searching for and Merging Duplicate Records" on page 35.
- Languages supported, see "SDQ Matching Server Libraries for Multiple Languages" on page 17.
- Platforms supported, see *System Requirements and Supported Platforms* on Siebel SupportWeb.

NOTE: The SDQ Matching Server uses embedded SSA-NAME3 software from Search Software America. The SSA-NAME3 libraries are embedded in Siebel eBusiness Applications and are included on the *Siebel eBusiness Applications, Base Applications for Platform DVD,* where Platform is either Windows or UNIX-OS (your UNIX operating system, such as HP-UX, AIX, or Solaris). You do not need to install additional third-party software to run the SDQ Matching Server.

For detailed information about SSA-NAME3, see the SSA-NAME3 documentation included on the *Siebel eBusiness Third-Party Bookshelf*.

Typical Data Matching Operations Performed by the SDQ Matching Server

The Siebel Data Quality (SDQ) Matching Server supports data matching on account, contact, and prospect data within the Siebel application. You can enable data matching in real-time mode and run batch jobs to match customer data within the Siebel transactional database.

Typical data matching operations performed by the Siebel Data Quality (SDQ) Matching Server are described in Table 3.

Table 3. Data Matching Operations Performed by the SDQ Matching Server

Type of Operation	Comments
Key generation	The SDQ Matching Server uses multiple keys for each record to detect duplicate records. Keys are generated for each customer record based on a set of deduplication input fields on the customer record for prospects and contacts, or account name for accounts (typically a combination of personal name, company name, address, and identifier information). Later in the search functionality, the matching server detects possible matches by comparing these stored keys against the key for the active record at run time. 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. If no keys are generated for a certain record, that record is ignored as a potential candidate when search and match takes place. **CAUTION:** Typically, you run batch-mode key generation before you run real-time data matching. The SDQ Matching Server requires generated keys in the key tables first before you can run real-time data matching. The Siebel Data Quality (SDQ) Universal Connector also has a similar requirement, but the key generation is done within the deduplication task.
Key refresh	Key refresh is a server component operation type (Key Refresh) specified by the administrator and is a useful administrative tool because it is much faster than key generation. Because key data can become out of sync with the base tables, you need to refresh periodically. A key refresh updates keys only for records that are new or have been modified since your last key generation.
Search and Match	The internal search functionality specifies what ranges of possible keys should be considered for a given record when performing a match. A match score is then computed for each candidate record found within that range of keys. The match score is a number that depicts the similarity of a record with a matched one. It is calculated taking into account a large number of rules along with a number of other factors and weightings.

Data Matching Components for the SDQ Matching Server

The data matching process for the Siebel Data Quality (SDQ) Matching Server consists of two components, Match Key Generation and Search and Match.

Match Key Generation

The matching server must have match keys for the account, contact, and prospect records already in the database. When the matching server performs a matching task, it is not comparing the raw data for each record in the database. Instead, the matching server uses the existing match keys in the database to pick up candidate records for comparison.

These keys are generated by applying an algorithm to the name fields that translates each of the names and words into a set of keys. The keys can be compared for similarity by the matching server. The SDQ Matching Server generates multiple keys for each existing customer record, and the number of keys generated depends on the settings for the key type that you select. If you select the narrower key type (Limited), then the key generation algorithm performs only the most common permutations and generates fewer keys. If you select the wider key type (Standard), the key generation algorithm uses a wider set of permutations to provide the most exhaustive range of keys.

For batch processing, keys are generated for all records that meet the object WHERE clause. For real-time data matching, keys are automatically generated for a record whenever the user saves a new record or modifies and commits an existing record to the database.

Search and Match

From the user's perspective, match key generation is a single task. However, the matching server actually executes two subtasks to complete the match, that is search and match.

After keys are generated for the existing data, the matching server can look for matches, what is referred to as search. In this process, the matching server takes the keys for a selected record (the record entered by a real-time user or the active record in the batch job) and looks for all existing keys that are similar to any keys from the selected record. Based on the values specified (Narrow, Typical, Exhaustive) in the Search Type field in the Data Quality Settings view, the match process scans a smaller range of keys to provide fastest response (Narrow) or scans a wider range of keys to provide the most exhaustive search (Exhaustive). In general, if you are using a wider (more exhaustive) key type, you should also use a wider search type.

When a set of candidate records whose keys fall into the selected search range is found, the matching server computes and assigns a match score to indicate the degree of similarity between the candidate records and the selected record. This match score is based on a combined weighting for all the input fields (personal name, company name, address, and identifiers).

When the match results are returned, the value specified in the Match Threshold field in the Data Quality Settings view determines whether or not the application considers a returned record a match. Match results exceeding the threshold are logged to the S_DEDUP_RESULT match results table and displayed in the user interface (as a real-time pop-up window or as a record in the Administration - Data Quality views). Match results below the threshold are not stored.

Results from running these key tasks vary depending on the values you set in the Data Quality Settings view.

About Data Matching for the SDQ Matching Server

The following is the data matching process used by the Siebel Data Quality (SDQ) Matching Server. This process is organized to present information in a sequence corresponding to the order in which events are likely to occur.

- Keys are generated for the existing customer records in the database.

  Typically, keys are generated and refreshed on a periodic basis by the data administrator. In addition, if real-time deduplication is enabled for end users, keys are also automatically generated for a customer record whenever a user inserts or modifies an existing record.

- When a user enters or modifies a record or the administrator submits a batch deduplication request, the SDQ Matching Server identifies candidate matches for each record by locating existing records whose corresponding keys fall within a range around the master record. 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.

- A match score is computed for each candidate record.

  The match score is a combination of a large number of rules that compensate for how frequently a given name or word appears in the real world. The rules then weigh the similarity of each field on the record according to the real-world frequency of the name. For example, Smith is a common last name, so a match on a last name of Smith would carry less weight than a match on a rare last name. Any existing records in which match scores exceed the threshold specified in the Data Quality Settings are considered as matches.

For information about administrative tasks using the SDQ Matcher Server, see "Process of Using Data Matching (Deduplication)" on page 27.

SDQ Matching Server Libraries for Multiple Languages

The matching rules for the Siebel Data Quality (SDQ) Matching Server are compiled in a set of dynamic-link libraries (DLLs) for various languages. Because the character and name patterns differ substantially between languages, matching rules typically are tuned specifically for each language or language family.

The SDQ Matching Server includes a set of matching libraries that cover a variety of languages and code pages. By default, the installation uses a generalized international library that is built to support a set of Latin1-based languages (languages predominant in the Americas, Western Europe, Australia, and New Zealand). In addition, the Siebel installation DVDs include reference libraries for other regions and code pages. Table 4 provides the matching libraries and languages supported. For information about code pages, see *Global Deployment Guide*.

NOTE: The international library intentionally ignores certain words and abbreviations because those words and abbreviations may have a different meaning in other non-Latin1 languages.

Table 4. Supported SDQ Matching Server Libraries and Languages

International Library	Other Libraries
DAN - Danish	CHS - Simplified Chinese
DEU - German	CHT - Traditional Chinese
ESN - Spanish	CSY - Czech
ENU - English	ELL - Greek
FIN - Finnish	HEB - Hebrew
FRA - French	JPN - Japanese
ITA - Italian	KOR - Korean
NLD - Dutch	PLK - Polish
PTG - Portuguese	
PTB - Brazilian Portuguese	
SVE - Swedish	

NOTE: The matching rules for each language or combination of languages are delivered in the form of DLLs. You can retrieve additional DLLs by installing other language packs on the Siebel Server. For more information about DLLs, see "SDQ Universal Connector Architecture" on page 22.

You can achieve better matching if the region-specific library is used. However, the international library is best if the data is not limited to that region, because a dataset can include a heterogeneous mixture of international names. Installing region-specific libraries for Latin-based languages requires that an administrator replace the library file on the Siebel Server with the language-specific version of the file. For example, for Windows ENU, the library is placed in C:\Siebel\SiebSrvr\bin\enu. For UNIX ENU, the library is placed in /export/home/siebel/siebsrvr/lib/enu.

The library file installed on each Siebel Server should be in sync with the data that is processed from that machine. For example, if the Japanese library is installed, a batch component job for key generation or deduplication should be constrained to Japanese data.

If the Siebel Server is running in Japanese, it loads and references the Japanese version of the matching libraries while the other Siebel Servers (running in a different language) load and reference other matching library files on their own server file systems. The match keys table in the database stores keys generated from different libraries on different Siebel Servers, together with indicators for code page and population (matching library). When a match request is executed, the list of possible match candidates is built based on the match keys from the same code page and population.

NOTE: The SDQ Matching Server does not support the ability to find matches across languages that are not supported by the installed library file. For example, English and French data can be compared using the international library, but Chinese and Spanish data cannot be compared because the matching rules for Chinese require a separate library.

Viewing Settings for SDQ Matching Server Libraries

You can view the settings for the SDQ Matching Server libraries using Siebel Tools.

To view settings for SDQ Matching Server libraries

1. Start Siebel Tools.
2. In the Object Explorer, expand the Business Service object.
3. In the Business Services window, query for DeDuplication.
4. In the Object Explorer, select Business Service User Prop, and then query for SSA Population-Codepage*.

   TIP: If the Business Service User Prop object is not visible in the Object Explorer, you can enable it in the Development Tools Options dialog box. For information about how to view object types in the Object Explorer, see the topic about showing and hiding object types in *Using Siebel Tools*.

For more information about matching libraries, see "SDQ Matching Server Libraries for Multiple Languages" on page 17.

SDQ Matching Server Rule Modification

The administrator uses Siebel Tools to specify the input fields for the embedded matching libraries in the business component user properties for the Account, Contact, and List Mgmt Prospective Contact business components. The input field values are indicated by a lettering nomenclature where different letters indicate standard input types for personal name, company name, address fields, and ID data.

For example, Z indicates postal or ZIP code while I indicates a general unique identifier such as the D-U-N-S number for accounts or social security number for contacts. The field mappings for the business component can be configured to include few fields or modified to map to different fields. For more information about field mappings for business components using the embedded SSA-NAME3 software, see the third-party documentation for SSA-NAME3 on the *Siebel eBusiness Third-Party Bookshelf*. For an example of how to configure the Siebel Data Quality (SDQ) Matching Server using the embedded SSA-NAME3 software for the Account business component, see Appendix B, "Configuring Siebel Data Quality Matching Server Using SSA."

CAUTION: The rules that control the parsing and weighting criteria that contribute to the match score are precompiled and cannot be modified with the standard Data Quality Matching (SDQ) Server module. The custom matching rules must be licensed separately from Search Software America. If your company requires tailored matching rules, contact Siebel Technical Support or Siebel Global Services and ask that they connect you with a representative from Search Software America.

About Installing the SDQ Matching Server

The InstallShield wizard for Siebel Enterprise Server automatically installs the Siebel Data Quality (SDQ) Matching Server files on a Siebel Server. The client install automatically installs the relevant DLLs.

For more information about installing the SDQ Matching Server, see the topic on installing the Siebel Server in the *Siebel Installation Guide* for the operating system you are using. If want language-specific versions of the SDQ Matching Server library files installed, you need to add the required language to your installation.

NOTE: If you are using third-party software for data matching, use the installation instructions provided by that third-party vendor instead.

Table 5 describes the SDQ Matching Server files and folders that are installed.

Table 5. Siebel Data Quality Matching Server Installation Files

Installation Component	Installation Information
SSA library files	For Windows: <siebel_root>\bin\<language>\n3sgsb.dll where <language> is the appropriate language code, such as ENU for American English For Solaris and AIX: <siebel_root>/lib/<language>/n3sgsb.so where <language> is the appropriate language code, such as ENU for American English For HP-UX: <siebel_root>/lib/<language>/n3sgsb.sl where <language> is the appropriate language code, such as ENU for American English For more information about library files, see "SDQ Matching Server Libraries for Multiple Languages" on page 17.
Help files	For Windows: <siebel_root>\bin\ssan3v2.dll <siebel_root>\bin \ssaion3.dll For Solaris and AIX: <siebel_root>/lib/libssan3v2.so <siebel_root>/lib/libssaion3.so For HP-UX: <siebel_root>/lib/ libssan3v2.sl <siebel_root>/lib/libssaion3.sl **NOTE:** These files are generic for all languages.

4 Siebel Data Quality Universal Connector

This chapter provides an overview of the Siebel Data Quality (SDQ) Universal Connector functionality and installation. It includes the following topics:

About the SDQ Universal Connector

The Siebel Data Quality (SDQ) Universal Connector provides a connector to third-party software that allows the Siebel application to use the capabilities of an external application for data matching and data cleansing. The matching, cleansing, and standardization capabilities of the SDQ Universal Connector are driven by the capabilities and configuration options of the third-party software. Using the connector, you can match and cleanse customer data within the Siebel database in real-time or batch mode.

The SDQ Universal Connector works only in conjunction with third-party software from data quality vendors who are certified by Siebel Systems' Technical Alliances program. For information about third-party solutions, see the Alliances section at *http://www.siebel.com*. For detailed information about products that work with this connector, see the Partners section at *http://www.siebel.com*.

Siebel customers who choose to use the SDQ Universal Connector must license additional software separately from third-party companies to make the connector solution functional. For information about which versions of software are validated by Siebel Systems for use with the SDQ Universal Connector, see *System Requirements and Supported Platforms* on Siebel SupportWeb for your Siebel eBusiness Applications. For data cleansing software installation instructions, see the documentation provided by the data cleansing vendor.

Data cleansing functionality is provided through third-party software using the SDQ Universal Connector; it is not provided by the Siebel Data Quality (SDQ) Matching Server. The SDQ Universal Connector can support real-time and batch-mode processing.

SDQ Universal Connector Architecture

The Siebel Data Quality (SDQ) Universal Connector uses prebuilt Siebel business services for data cleansing and deduplication. These business services include a generalized adapter that can communicate to an external data quality application through a set of dynamic-link library (DLL) files. For Windows, the DLL files are stored in the \bin\<language> directory in the Siebel Server root directory, where <language> is the appropriate language code (for instance, ENU for American English). For UNIX, the DLL files are stored in the /lib/<language> directory in the Siebel Server installation directory, where <language> is the appropriate language code (for instance, ENU for American English).

NOTE: If the third-party software does not have multiple-language support, such as is the case with Firstlogic, put the DLL directly under the \bin directory (for Windows) or /lib directory (for UNIX).

The DLL files for Firstlogic are as follows:

- sdqaddress.dll
- sdqname.dll
- sdqmatch.dll

The DLLs for each vendor may be specific to certain platforms or external product versions, so it is important that you confirm with your vendor that you have the correct files installed on your Siebel Server machine.

NOTE: For customers using the SDQ Universal Connector to integrate with Firstlogic applications, the DLLs for the certified link from Firstlogic are specific to each supported middle-tier OS platform for Siebel applications. It is recommended that you copy the correct DLL versions from the installation DVD for the middle-tier platform you are using to the SDQConnector directory in the Siebel Server root directory.

The SDQ Universal Connector requires that you install third-party applications on each Siebel Server machine that has the object managers enabled for data quality functionality.

Data Cleansing and Data Matching with SDQ Universal Connector

The Siebel Data Quality Universal Connector supports both data cleansing and data matching on account, contact, and prospect data within the Siebel application. Through the connector, customer data within the Siebel database is matched and cleansed in real-time or batch mode.

The following topics describe this functionality in detail:

- "Typical Data Cleansing Operations Performed by the SDQ Universal Connector" on page 23
- "About Data Matching (Deduplication) Process for SDQ Universal Connector" on page 24

Typical Data Cleansing Operations Performed by the SDQ Universal Connector

Data cleansing can fix inaccurate and inconsistent data for new or modified account, contact, and prospect records, as described in Table 6.

NOTE: Data cleansing can be done only with the Siebel Data Quality (SDQ) Universal Connector, not with the Siebel Data Quality (SDQ) Matching Server.

Table 6. Data Cleansing Operations

Type of Correction	Comments
Address correction	Address correction updates the fields on an address record with values from a certified external source, typically a directory of addresses from a national postal service or other organization. Typically address correction modifies the following fields: Street Address City State/Province Postal Code/ZIP Code For example, 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

Table 6. Data Cleansing Operations

Type of Correction	Comments
Capitalization	Capitalization standardizes name and address data. For example, a company name and address is converted into title case (Siebel Systems) or all caps (SIEBEL SYSTEMS).
Standardization	Standardization sets the abbreviations and other formatting of a name or address record. For example, Siebel Systems, Incorporated becomes Siebel Systems, Inc. and IBM Corporation becomes IBM Corp. Typically, standardization operates on different sets of fields for account, contact, and prospect records. **Account records**. Typical fields are Account Name and Site fields for account records. **Contact and prospect records**. Typical fields are First Name, Middle Name, Last Name, and Job Title for these records.

About Data Matching (Deduplication) Process for SDQ Universal Connector

The data matching (deduplication) functionality of the Siebel Data Quality (SDQ) Universal Connector uses validated third-party vendor software for the matching rules and algorithms and maintenance of any match keys.

The methodologies and matching capabilities of external applications vary by vendor. Matching rules and weightings are typically configurable within the external application. After running batch deduplication, the SDQ Universal Connector reports the possible matches in the Duplicate views in the Administration - Data Quality screen. A data administrator can then manually merge the records. For information about merging duplicate records, see "Process of Searching for and Merging Duplicate Records" on page 35.

During the batch deduplication process, all records in the database are passed to the third-party software. The software uses an optimized algorithm to separate records into groups to reduce the number of record comparisons. One key difference between the SDQ Universal Connector and the SDQ Matching Server is that the SDQ Universal Connector combines key generation and deduplication into one process. While running batch deduplication using the SDQ Universal Connector, the key values for records are saved in files by the third-party vendor software. During real-time duplication, the third-party software uses the key values stored in the files to find potential duplicates.

TIP: You should run batch deduplication against a business component before running real-time deduplication.

About Installing the SDQ Universal Connector

To use the Siebel Data Quality (SDQ) Universal Connector, you must install the Data Quality Connector feature when running the InstallShield wizard for Siebel Server Enterprise as shown in Figure 1.

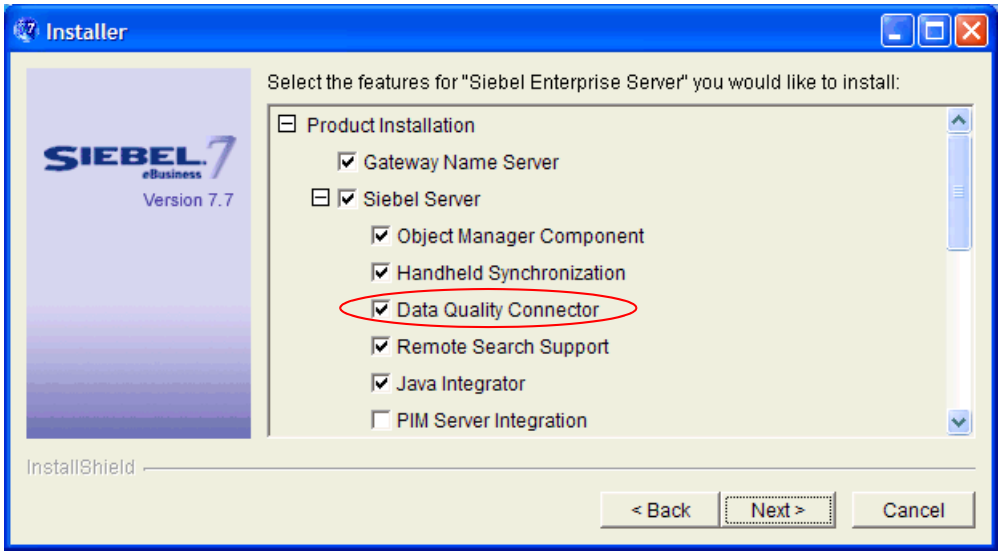

Figure 1. Siebel Enterprise Server Installation Screen Showing Data Quality Connector Feature Selected

For more information about installing Siebel Enterprise Server, see the *Siebel Installation Guide* for the operating system you are using. For information about installing the SDQ Universal Connector on a network, see "About Installing the SDQ Universal Connector Files on the Network" on page 26.

About Installing Third-Party Software for Use with the SDQ Universal Connector

You should install third-party software *after* you install Siebel eBusiness Applications (unlike most other third-party software that is installed before you install Siebel eBusiness Applications). Install the third-party software in the SDQConnector directory where your Siebel eBusiness Applications are installed; the SDQConnector directory is under the root directory of your Siebel Server installation. Include the third-party software configuration files and other software that is not put in the \bin or /lib directories. The third-party software configuration files are automatically loaded for you by the third-party installer. You may need to modify the third-party configuration files. See the documentation provided by the third-party vendor for instructions.

About Installing the SDQ Universal Connector Files on the Network

To perform data cleansing, the third-party vendor software usually needs a set of files for standardization and data cleansing. These files, which Firstlogic calls dictionary files, can be large. If you use Firstlogic and you store the dictionary files in a network directory, you need to explicitly identify the location of the dictionary files in the Firstlogic sdqaddress.cfg configuration file. You specify the location by modifying the value of the US_ENGINE_DIRECTORY_PATH parameter in the sdqaddress.cfg file to point to the network location of the dictionary files. For details, see the FirstLogic documentation.

5 Configuring and Using Data Matching

This chapter explains how to configure and use the data matching functionality. It covers the following topics:

Process of Using Data Matching (Deduplication)

Data matching, sometimes described as deduplication, identifies possible matches for account, contact, and prospect records based on match fields defined in Siebel Tools and administrator-defined matching preferences. After potential duplicates are identified, the administrator can use the Administration - Data Quality views to merge the duplicate profiles together manually.

Data matching functionality is provided by the Siebel Data Quality (SDQ) Matching Server and the Siebel Data Quality (SDQ) Universal Connector. These product modules provide real-time or batch-mode processing, but each process handles duplicates differently:

- In real-time mode, users are prompted when entering a new record that matches another record or when saving changes to an existing record that matches another record.
- In batch mode, the application identifies potentially duplicate records and presents those records to the data administrator for resolution in the Administration- Data Quality views.

To use data matching (deduplication) functionality, you must first perform the following tasks:

- "Setting the Deduplication Data Type for Data Matching" on page 28
- "Enabling Data Matching (Deduplication) for Real-Time Processing" on page 29
- "Applying Siebel Data Quality Settings" on page 31
- (Optional) "Setting User Preference Data Quality Settings" on page 33

NOTE: You can perform data matching without data cleansing enabled. Data matching does not necessarily require enabling data cleansing; however, the quality of the match results can vary with cleansing enabled versus with cleansing disabled, depending on the matching engine used. Some vendors use algorithms that use data cleansing, while others do not.

For more information about the Siebel Data Quality product modules that use data matching, see Chapter 3, "Siebel Data Quality Matching Server," and Chapter 4, "Siebel Data Quality Universal Connector."

Setting the Deduplication Data Type for Data Matching

For you to enable data matching, the Deduplication Data Type parameter must be set for the enterprise. You can choose to use the Siebel Data Quality (SDQ) Matching Server or the Siebel Data Quality (SDQ) Universal Connector as your match solution at the time of your Siebel Server installation. The applicable setting for the DeDuplication Data Type parameter is automatically set during install. However, you can also review or change that setting through the Siebel Server Manager. Use the following procedure to review or change the Deduplication Data Type parameter setting at the enterprise level.

This task is a step in "Process of Using Data Matching (Deduplication)" on page 27.

To review or change the Deduplication Data Type setting for the enterprise

1. Log in to the Siebel application with administrator responsibilities.
2. From the application-level menu, choose Navigate > Site Map > Administration - Server Configuration > Enterprises.
3. Click the Parameters view tab.
4. In the Parameter field in the Enterprise Parameters list, query for the DeDuplication Data Type parameter, and review the settings.
   - CHANGE_ME appears in the Default Value field if you chose None when you installed the Siebel Server.
   - SSA appears in the Default Value field if you chose Siebel Data Quality Matching when you installed the Siebel Server.
   - Vendor1 appears in the Default Value field if you chose Data Quality Connector when you installed the Siebel Server.
5. (Optional) If you want to change the deduplication type, in the Value field, enter another data type.

   TIP: The *Value* field in the Enterprise Parameters list appears as the *Current Value* field in the Component Parameters and Server Parameters views.
6. If you changed the data type in Step 5, you must restart the server component for the new settings to take effect.
   a. From the application-level menu, choose Navigate > Site Map > Administration - Server Management > Servers.
   b. Click the Components Groups view tab (if not already active).
   c. In the Servers list (upper applet), select the appropriate Siebel Server (if you have more than one in your enterprise).
   d. In the Components Groups list (middle applet), select the component of your object manager, and use the Startup and Shutdown buttons to restart the component.

   For more information about restarting server components, see *Siebel System Administration Guide*.

Enabling Data Matching (Deduplication) for Real-Time Processing

Real-time data matching (deduplication) is supported only for employee-facing applications. However, it is by default disabled after you install your Siebel Server. To enable data matching for real-time processing for an application, the deduplication parameter must be enabled for the object manager that the application uses. You can enable data matching at the component (application) level using either the graphical user interface (GUI) or the command-line interface of the Server Manager.

NOTE: Because the data deduplication parameters are specified at the object manager level in the Siebel application, data matching (deduplication) and data cleansing can be enabled for one application and disabled for another. However, you cannot enable the Siebel Data Quality (SDQ) Matching Server and the Siebel Data Quality (SDQ) Universal Connector data matching functionality for the same application.

The following tasks are for one step in "Process of Using Data Matching (Deduplication)" on page 27. Choose one of these tasks to enable data matching at the component (application) level.

To enable real-time data matching (deduplication) for an application using the GUI

1. Log in to the Siebel application with administrator responsibilities.
2. From the application-level menu, choose Navigate > Site Map > Administration - Server Configuration > Servers.
3. In the Components list (middle applet), select the object manager where the end users enter and modify customer data, for example, Call Center Object Manager (ENU).
4. Click the Parameters subview tab (bottom applet).
5. In the Parameters field in the Component Parameters list, query for the parameters provided in Table 7 on page 30, and change the settings as indicated.
6. After the component parameters are set, restart the object manager.
   a. From the application-level menu, choose Navigate > Site Map > Administration - Server Management > Servers.
   b. Click the Components Groups view tab (if not already active).
   c. In the Servers list (upper applet), select the appropriate Siebel Server (if you have more than one in your enterprise).
   d. In the Components Groups list (middle applet), select the component of your object manager, and use the Startup and Shutdown buttons to restart the component.

   For information about restarting server components, see *Siebel System Administration Guide*.

Use the parameter settings in Table 7 to enable data matching (deduplication) for an application object manager server component, for example, Call Center Object Manager (ENU).

Table 7. Data Matching (Deduplication) Parameter Settings

Parameter	Value	Description
DeDuplication Data Type	Third-party software name	This parameter is an enterprise-level parameter. The value is set when you install the Siebel application. If you do not select an option during the installation, the value is CHANGE_ME. You can also set this parameter at the component level, for example, Call Center Object Manager, so that the component uses a value other than the enterprise setting value. This setting allows you to run different deduplication types in different object managers. The settings at the component (application) level override the enterprise-level setting. For more information about this parameter, see "Setting the Deduplication Data Type for Data Matching" on page 28.
DeDuplication Enable Flag	False True	False is the default value in the Value field. Set the Value field to True to enable real-time data matching for the application (object manager). **NOTE:** The DeDuplication Enable Flag is only for real-time deduplication. For batch mode, you do not need to set this parameter for Data Quality Manager.

To enable real-time data matching (deduplication) for an application using the command-line interface

1 Log in to the Siebel application with administrator responsibilities.

NOTE: You must have Siebel administrator responsibility to start or run Siebel Server tasks using the Siebel Server Manager command-line interface.

2 Make sure the Siebel Server Manager is running, and then execute the following command to enable data matching:

`change parameter parameter_alias_name1=value1, parameter_alias_name2=value2 for component component_alias_name`

where parameter_alias_name1 and parameter_alias_name2 are the component properties you want changed, value1 and value2 are the values for each component property, and alias_name is the application component name to which you want the change applied.

For example, use the following commands to enable real-time deduplication for the Call Center English application (where SSCObjmgr_enu is the alias name of the English Call Center object manager of the Call Center application.)

- If using Search Software America (SSA) software:

  `change parameter DeDupTypeEnable=True, DeDupTypeType=SSA for component SCCObjMgr_enu`

- If using Universal Connector third-party software:

  `change parameter DeDupTypeEnable=True, DeDupTypeType=Vendor1 for component SCCObjMgr_enu`

For more information on using the command-line interface, see *Siebel System Administration Guide*.

Applying Siebel Data Quality Settings

You can set or change the Siebel Data Quality (SDQ) options for both the Siebel Data Quality (SDQ) Matching Server and the Siebel Data Quality (SDQ) Universal Connector. Use the following procedure to view, set, or change the Data Quality Settings in the Administration - Data Quality screen. You can use the same procedure for both real-time and batch processing.

NOTE: Data Quality Settings can also be specified at the user level in the User Preferences Data Quality view. The settings in user preferences override those set in the Data Quality Settings view in the Administration - Data Quality screen. For more information about applying data quality settings at the user level, see "Setting User Preference Data Quality Settings" on page 33.

This task is a step in "Process of Using Data Matching (Deduplication)" on page 27.

To apply SDQ options for data matching in the Administration - Data Quality screen

1 From the application-level menu, choose Navigate > Site Map > Administration - Data Quality > Data Quality Settings.

2 Add new records for each parameter listed in Table 8 on page 32 and set the value for each parameter appropriately for your installation.

NOTE: By default, the list applet is empty after installation.

3 After making these changes, you must log out of the application and log back in for the changes to take effect.

Table 8 describes the Siebel Data Quality Settings parameters.

Table 8. Siebel Data Quality Settings Parameters

Parameter	Possible Values	Description
Disable DataCleansing	Yes No	Disables data cleansing for real-time mode only. This disables data cleansing for the application the administrator is currently logged in to. Affects business components related to Siebel Data Quality.
Disable DeDuplication	Yes No	Disables data cleansing for real-time mode only. This disables data matching for the application the administrator is currently logged in to. Affects business components related to Siebel Data Quality. **NOTE:** The Disable DeDuplication settings specified in the Administration - Data Quality screen turn off the features even if you enable them using the Administration - Server Configuration screen.
Key Type	Standard Limited	Applicable only for the Siebel Data Quality (SDQ) Matching Server. Match keys are generated by applying an algorithm to the name fields, which translates each of the names and words into a set of keys that can be compared for similarity by the matching server. The SDQ Matching Server generates multiple keys for each existing customer record. The number of keys generated depends on the settings for key type that you select. If you select the narrower key type (Limited), then the key generation algorithm performs only the most common permutations and generates fewer keys. If you select the wider key type (Standard), the key generation algorithm performs a wider set of permutations to provide the most exhaustive range of keys.

Table 8. Siebel Data Quality Settings Parameters

Parameter	Possible Values	Description
Match Threshold	0-100	Applicable only for the SDQ Matching Server. Any records with a match score above this threshold is considered a match. Higher scores indicate closer matches (perfect match = 100). **NOTE:** If you change the Match Threshold value, you must stop and restart the Siebel Server System Service for the change to take effect. For more information about stopping and starting the Siebel Server System Service, see *Siebel System Administration Guide*.
Search Type	Narrow Typical Exhaustive	Applicable only for the SDQ Matching Server. This setting indicates whether the match algorithm should use a narrow set of matching rules or a more exhaustive set of rules. A more exhaustive set of rules looks for additional data permutations, but typically takes more time to process.

For more information about the values specific to Siebel Data Quality Matching Server, see the SSA-Name3 documentation on the *Siebel eBusiness Third-Party Bookshelf*.

Setting User Preference Data Quality Settings

The User Preferences view displays the same options that are set at the application level in the Data Quality Settings view. You can set or change the User Preferences options for both the Siebel Data Quality (SDQ) Matching Server and the Siebel Data Quality (SDQ) Universal Connector. Use the following procedure to view, set, or change the Data Quality Settings in the User Preferences screen. You can use the same procedure for both real-time and batch-mode processing.

NOTE: The values in the user preference settings have a higher priority than the values in the application setting. This means that a user can disable data cleansing for their own login even if data cleansing is enabled for their application. For more information about Data Quality Settings at the application level, see "Applying Siebel Data Quality Settings" on page 31.

This task is a step in "Process of Using Data Matching (Deduplication)" on page 27.

To set SDQ options for data matching in the User Preferences screen

1. Log in as the user who will use the real-time feature.
2. From the application-level menu, choose Navigate > Site Map > User Preferences > Data Quality.
3. In the Data Quality form, set the deduplication and data cleansing parameters for that user.
   a. In the Disable Deduplication field, choose Yes to disable deduplication.

b In the Disable Data Cleansing field, choose Yes to disable data cleansing.

For more information about field values and descriptions, see Table 8 on page 32.

TIP: Setting the Disable Deduplication and Disable Data Cleansing fields to No does not imply that you can enable these features without first enabling them in that object manager. For more information, see "About Disabling Deduplication Without Restarting the Siebel Server" on page 34.

4 If you are using the Siebel Data Quality Matching Server for data matching, set the values for Search Type, Match Threshold, and Key Type by doing the following.

a In the Search Type field, choose a search type (Exhaustive, Narrow, or Typical).

b In the Key Type field, choose a key type (Limited or Standard).

c In the Match Threshold field, choose a value (values are in increments of ten from 50-100).

NOTE: If you change the Match Threshold value, you must stop and restart the Siebel Server System Service for the change to take effect. For more information about stopping and starting the Siebel Server System Service, see the Siebel Server System Service topic in *Siebel System Administration Guide*.

For more information about field values and descriptions, see Table 8 on page 32.

5 Log out of the application and log back in as the user to initialize the new settings.

About Disabling Deduplication Without Restarting the Siebel Server

The Disable DeDuplication settings specified in the Administration - Data Quality screen override those specified in the Administration - Server Configuration screen. If you enabled the deduplication from the Administration - Server Configuration screen, you can disable deduplication from the Data Quality Settings view without restarting the Siebel Server. After you disable deduplication, log out and then log in to the application again for the new settings to take effect. The settings apply to all the object managers in your Siebel Server, whether or not they have been enabled in the Administration - Server Configuration screen. When you are ready to enable deduplication again, reset the Disable DeDuplication field to No, and then log in again.

NOTE: Only specific fields are configured to support data cleansing and deduplication. If you do not enter values in the Disable Data Cleansing and Disable DeDuplication fields when you create a new record, or you do not touch the values in these fields when you modify a record, then data cleansing and deduplication are not triggered. For more information about which fields are preconfigured for different business components, see Appendix A, "Preconfigured SDQ Universal Connector Properties for Firstlogic Applications," and Appendix B, "Configuring Siebel Data Quality Matching Server Using SSA."

For more information, see "Applying Siebel Data Quality Settings" on page 31. In particular, the Disable DeDuplication settings described in that topic are useful for disabling the real-time data quality functionality temporarily or for a certain user, assuming you have enabled it for that object manager.

Process of Searching for and Merging Duplicate Records

When you run a batch process, and depending on the number of duplicates in your system and when you implement the SDQ module, you might find there are potentially hundreds of rows in the Duplicate Resolution views (in the Administration - Data Quality screen). In these instances, it is recommended that you use the following two-step process for searching for and merging duplicate records.

To search for and merge duplicate records, perform the following tasks:

1 "Searching for Duplicate Records" on page 35

   You search for duplicate records by creating a query to find duplicate records and then reviewing the query results. For example, you might want to create a query that includes a subset of all duplicate records where the Name field starts with the letter A.

2 "Merging Duplicate Records" on page 36

   After the query results appear, you merge duplicate records using either the Merge button or the Merge Records option.

Searching for Duplicate Records

Searching for duplicate records involves creating a query to find duplicate records and reviewing the query results. After the query results appear, you click a record to view that record, or click the appropriate Resolution view tab to view the duplicates for that record. For each set of candidates that match the selected record, you specify which record should be retained and request that other records be merged into it.

CAUTION: You should perform batch deduplication first before trying to resolve duplicate records. For more information about batch deduplication, see Chapter 7, "Working with Data Cleansing and Data Matching in Real-Time and Batch Modes," and for information about performance considerations using batch mode, see Chapter 9, "Data Quality Performance Considerations."

This task is a step in "Process of Searching for and Merging Duplicate Records" on page 35.

To search for duplicate records

1 From the application-level menu, choose Navigate >Site Map > Administration - Data Quality.

2 Click one of the following view tabs:

- Duplicate Accounts
- Duplicate Contacts
- Duplicate Prospects

3 Click Query, enter your search criteria, and then click Go.

   The search results appear.

You now decide what you want to do with the duplicate records. For more information, see "Merging Duplicate Records" on page 36.

Merging Duplicate Records

You merge duplicate records using the Merge Records option from the menu list or by clicking the Merge button. There is a difference between these two types of merging operations:

- **Merge Records option**. Performs the standard merge functionality available in Siebel applications for merging records. For more information about the Merge Records menu option, see *Fundamentals* on the *Siebel Bookshelf*.
- **Merge button**. Performs a sequenced merge before performing the standard Merge Records option. This includes populating sparse fields based on the sequence of the records to be merged. It also performs a cleanup in the appropriate Deduplication Results table. Cleanup removes the unnecessary duplicate records that are stored in the results table. This is the preferred method for deduplicating account, contact and prospect records.

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

The following procedure is for one step in "Process of Searching for and Merging Duplicate Records" on page 35. The procedure assumes you have already searched for duplicate records and are ready to determine which records you want to retain and which records you want to merge with the retained record. If you have not already done so, perform the steps provided in "Searching for Duplicate Records" on page 35 before implementing the following procedure.

CAUTION: Merging records is an irreversible operation. You should review all records carefully before initiating this function and using the following procedure.

To merge duplicate records

1. In the Administration - Data Quality screen, click the Duplicate view for the type of record you have selected.

   For example, if you have selected an account record, click the Duplicate Accounts view.

2. If two records appear to be duplicates, enter a sequence number in the Sequence field for each record.

   The last record selected is retained after the merge. Missing fields in the retained record are populated in ascending sequence number order from corresponding fields in the remaining record.

   NOTE: The Siebel Data Quality (SDQ) functionality merges only active fields, which are fields that are visible on the applet. In addition, the functionality extends to merge all the fields designated as Active in the respective business component.

3. Edit the records, if necessary.

4. Click Merge.

   The two records are merged to produce one new record.

For more information about sequenced merges, see "About Sequenced Merges" on page 37.

About Sequenced Merges

When records are merged using a sequence merge, the following events take place:

- All non-NULL fields from the surviving record are kept.
- Any fields that were NULL in the surviving record are populated in the surviving record by information (if any) from the respective fields in the nonsurviving records, based on the sequence number of the nonsurviving record.
- The children and grandchildren (for example, activities, orders, assets, service requests, and so on) of the nonsurviving records are reparented to the retained record.

For example, 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 copies 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 to pick from. It is especially useful if many fields are empty, such as when a contact record with a Sequence of 1 has a value for Email address, but its Work Phone # field is empty, and a contact record with a Sequence number of 2 has a value of Work Phone #. If the field Email address and Work Phone # in the retained record are empty, then the value of Email address is picked from sequence number 1 and the value of Work Phone # is picked from record of sequence number 2.

NOTE: For the merge functionality to operate properly, you must enter a sequence number in the Sequence field. The sequence number is required even if there are only two records (one to be merged and one surviving). However, the retained record is based on the last record you selected using the keyboard or mouse, not the sequence number entered. The sequence number indicates only the sequence in which null fields should be populated from the losing records to the winning record. Do not specify a sequence number for the winning (retained) record.

Field Characteristics for Sequenced Merges

A field must have specific characteristics to be eligible for use in a sequenced merge:

- The field must be active.
- The field must not be a calculated field.

For more information about field characteristics for sequenced merges, perform a search for data quality merge on Siebel SupportWeb (*http://ebusiness.siebel.com/supportweb/*).

6 Configuring and Using Data Cleansing

This chapter explains how to enable and disable, configure, and use data cleansing. It covers the following topics:

About Siebel Data Cleansing

Data cleansing, sometimes described as standardization or data enhancement, modifies account, contact, and prospect records based on cleansing fields defined in Siebel Tools and administrator-defined cleansing preferences.

Data cleansing functionality is provided through third-party software using the Siebel Data Quality (SDQ) Universal Connector; it is not provided by the Siebel Data Quality (SDQ) Matching Server. The SDQ Universal Connector can support real-time and batch-mode processing.

- In real-time mode, data is modified when a user tries to save a new or modified record back to the database. For more information about real-time mode, see Chapter 7, "Working with Data Cleansing and Data Matching in Real-Time and Batch Modes."
- In batch mode, you as the administrator, can select the candidate data set based on specified constraints. You can choose to cleanse the whole database or a set of records by setting a search criteria. For more information about the search functionality, see "SDQ Matching Server Rule Modification" on page 18. For more information about batch mode, see Chapter 7, "Working with Data Cleansing and Data Matching in Real-Time and Batch Modes."

NOTE: You can perform data cleansing without enabling data matching.

For more information about configuring data, see Chapter 8, "Data Quality Configuration Options," and Appendix A, "Preconfigured SDQ Universal Connector Properties for Firstlogic Applications."

Setting the Data Cleansing Type for Data Cleansing

For you to enable data cleansing, the Data Cleansing Type parameter must be set for your enterprise. Because Siebel eBusiness Applications have only one data cleansing solution available, this parameter is set by default after you install your Siebel Server. However, you can also review or change that setting through the Siebel Server Manager. Use the following procedure to review or change the Data Cleansing Type parameter setting at the enterprise level.

To review or change the Data Cleansing Type setting for the enterprise

1. Log in to the Siebel application with administrator responsibilities.
2. From the application-level menu, choose Navigate > Site Map > Administration - Server Configuration > Enterprises.
3. Click the Parameters view tab.
4. In the Parameter field in the Enterprise Parameters list, query for the Data Cleansing Type parameter and review the settings.

   vendor1 appears in the Value field because the Siebel application uses the Data Quality Universal Connector as the data cleansing solution.
5. (Optional) In the Value field, enter another data type, if you want to change the deduplication setting.
6. If you changed the data type in Step 5, you must restart the server component for the new settings to take effect.
   a. From the application-level menu, choose Navigate > Site Map > Administration - Server Management > Servers.
   b. Click the Components Groups view tab (if not already active).
   c. In the Servers list (upper applet), select the appropriate Siebel Server (if you have more than one in your enterprise).
   d. In the Components Groups list (middle applet), select the component of your object manager, and use the Startup and Shutdown buttons to restart the component.

   For more information about restarting server components, see *Siebel System Administration Guide*.

Enabling Data Cleansing for Real-Time Processing

Real-time data cleansing is supported only for employee-facing applications. However, it is by default disabled after you install your Siebel Server. For you to enable data cleansing for real-time processing for an application, the data cleansing parameters must be enabled for the object manager that the application uses. Use the following procedure to enable data cleansing at the component (application) level.

To enable real-time data cleansing for an application

1. Log in to the Siebel application with administrator responsibilities.
2. From the application-level menu, choose Navigate > Site Map > Administration - Server Configuration > Servers.
3. In the Components list (middle applet), select the object manager where the end users enter and modify customer data, for example, Call Center Object Manager (ENU).
4. Click the Parameters subview.
5. In the Parameter field in the Component Parameters list, query for the parameters provided in Table 9 on page 42, and change the settings as indicated.
6. After the component parameters are set, restart the object manager.
   a. From the application-level menu, choose Navigate > Site Map > Administration - Server Management > Servers.
   b. Click the Components Groups view tab (if not already active).
   c. In the Servers list (upper applet), select the appropriate Siebel Server (if you have more than one in your enterprise).
   d. In the Components Groups list (middle applet), select the component of your object manager, and use the Startup and Shutdown buttons to restart the component.

   For information about restarting server components, see *Siebel System Administration Guide*.

Use the parameter settings in Table 9 to enable data cleansing for an application object manager server component, such as Call Center Object Manager (ENU).

Table 9. Data Cleansing Parameters

Parameter	Value	Description
Data Cleansing Enable Flag	False True	False is the default value in the Current Value field. Set the Current Value field to True to enable real-time data cleansing for the application (object manager). **NOTE:** The Data Cleansing Enable Flag is only for real-time data cleansing. For batch mode, you do not need to set this parameter for Data Quality Manager. For more information about batch mode, see Chapter 7, "Working with Data Cleansing and Data Matching in Real-Time and Batch Modes."
Data Cleansing Type	Third-party software name	This parameter is an enterprise-level parameter. The value is set when you install the Siebel application. The default value is Vendor1. You can also set this parameter at the component level, for example, Call Center Object Manager, so that the component uses a value other than the enterprise setting value. This setting allows you to run different data cleansing types in different object managers. The settings at the component (application) level override the enterprise-level setting. For more information about this parameter, see "Setting the Data Cleansing Type for Data Cleansing" on page 40.

NOTE: Because the data cleansing parameters are specified at the object manager level, Siebel data cleansing and deduplication can be enabled for one application and disabled for another.

Disabling Data Cleansing for Specific Records

You can disable data cleansing on a record-by-record basis for both real-time and batch mode.

To disable data cleansing for a record

1. Navigate to and select the record.
2. Check the column on the record labeled Disable Data Cleansing.

   This flag is unchecked (cleansing allowed) by default for new records.

If the Disable Data Cleansing field does not appear, take the following action.

To make the Disable Data Cleansing field appear

- For the list applet, choose Columns Displayed from the View drop-down menu to expose this field.
- For the form applet, click the Show More button to display this field.

For more information about real-time and batch mode, see Chapter 7, "Working with Data Cleansing and Data Matching in Real-Time and Batch Modes."

About Disabling Data Cleansing Without Restarting the Siebel Server

The Disable DataCleansing setting specified in the Administration - Data Quality screen overrides those settings specified in the Administration - Server Configuration screen. If you enabled the data cleansing from the Administration - Server Configuration screen, you can disable data cleansing from the Data Quality Settings view (View > Site Map > Administration - Data Quality > Data Quality Settings) without restarting the Siebel Server. After you disable the settings in Data Quality Settings, log out and then log in to the application again for the new settings to take effect. The settings apply to all the object managers in your Siebel Server, whether or not they have been enabled in the Administration - Server Configuration screen. When you are ready to enable data cleansing again, reset the Disable DataCleansing field to No, and then log in again.

7 Working with Data Cleansing and Data Matching in Real-Time and Batch Modes

This chapter explains how to use data matching and data cleansing in real-time and batch modes and how to customize data quality component jobs. It includes the following topics:

About Running Data Matching and Data Cleansing in Real-Time Mode

In real-time mode, Siebel Data Quality (SDQ) is invoked when you save a new or modified record. If both data cleansing and data matching are enabled in the same object manager, data cleansing executes first.

- If data cleansing is enabled, a set of fields preconfigured to use data cleansing are standardized before committing the record.

  For more information about standardization, see "About Data Matching and Data Cleansing" on page 11.

- 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 or click Ignore All, 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 using the Pick button. The duplicate record is removed from the system in one of the following ways:

  - If you are in the process of creating a new record, that record is not saved.

- If you are in the process of modifying a record, the change is not made to the record.

NOTE: Only certain fields are configured to support data cleansing and deduplication. If you do not enter values in these fields when you create a new record, or you do not touch the values in these fields when you modify a record, then data cleansing and deduplication are not triggered. For more information about which fields are preconfigured for different business components, see Appendix A, "Preconfigured SDQ Universal Connector Properties for Firstlogic Applications," and Appendix B, "Configuring Siebel Data Quality Matching Server Using SSA."

For more information about enabling real-time mode, see "Enabling Data Matching (Deduplication) for Real-Time Processing" on page 29 and "Enabling Data Cleansing for Real-Time Processing" on page 40.

About Running Data Matching and Data Cleansing in Batch Mode

CAUTION: Before enabling batch mode for your implementation, you should familiarize yourself with *Siebel System Administration Guide*. In particular, you should read the chapters about the Siebel Enterprise Server architecture, using the Siebel Server Manager GUI, and using the Siebel Server Manager command-line interface.
When you run a process in batch mode, any visibility limitation against your targeted data set is ignored. It is recommended that you allow only a small group of people to access the Siebel Server Manager to run your data quality tasks and require a login with an administrator ID, such as sadmin.

You can run data matching and data cleansing in batch mode. Batch mode lets you match and cleanse a large number of records at one time. You can run batch mode jobs as stand-alone tasks (see "About Running Data Quality Batch Mode Requests from the Command Line" on page 50) or schedule them on a recurring basis (see "About Customizing Data Quality Server Component Jobs for Batch Mode" on page 51).

- **Batch data cleansing**. Use data cleansing to standardize the structure of data in customer profiles. In batch mode, the application standardizes and corrects a group of accounts, contacts, prospects, or business addresses.
- **Batch data matching (deduplication)**. Use data matching to identify possible duplicate-record matches for account, contact, and prospect records. In batch mode, the application identifies potentially duplicate records and presents them to the data administrator for resolution.

Running data matching and data cleansing in batch mode requires the use of a separate server component called Data Quality Manager. Although you can run real-time data quality features in any interactive Siebel object manager, you should run batch mode data quality features using the Data Quality Manager server component. You use a native management console called Siebel Server Manager to manage, submit, and monitor your data quality tasks. For more information about using the Siebel Server Manager, see *Siebel System Administration Guide*.

In most implementations, you should run batch-mode key generation before you run real-time data matching. The Siebel Data Quality (SDQ) Matching Server requires generated keys in the key tables first before you can run real-time data matching. The Siebel Data Quality (SDQ) Universal Connector also has a similar requirement, but the key generation is done within the deduplication task.

CAUTION: Do not put Visual Basic (VB) logic in business components that are used for batch mode tasks. These tasks execute in the background and may not trigger logic that activates user interface features, such as pop-up windows.

About Enabling the Data Quality Component Group for Batch Mode Requests

For you to run batch mode data quality features, you must first enable the Data Quality component group. Data Quality Manager is the preconfigured component in the Data Quality component group that you use to run your data quality tasks.

When you install Siebel Server, you are asked to specify the list of component groups you want to enable. You can enable the Data Quality component group at that screen. If you do not choose to enable the Data Quality component group during installation, you can at a later time using the Server Manager. For more information about enabling server component groups, see the chapter about using the Server Manager graphical user interface (GUI) in *Siebel System Administration Guide*.

After the Data Quality component group is enabled, you can start your data quality tasks. You use the Components view in the Servers screen (Navigate > Site Map > Administration - Server Configuration > Servers > Components) to run data quality component jobs. For more information about administering component jobs, see the chapter about using the Server Manager user interface in *Siebel System Administration Guide*.

About Running Data Cleansing in Batch Mode Using SDQ Universal Connector

For you to run data cleansing in batch mode using the Siebel Data Quality Universal Connector, you run a server component job. For more information about and procedures for running component jobs, see the chapter about using the Server Manager graphical user interface (GUI) in *Siebel System Administration Guide*.

The business components preconfigured to support data cleansing using the Siebel Data Quality Universal Connector are:

- Account
- Business Address (or CUT Address for users of Siebel Industry Applications)
- Contact
- List Mgmt Prospective Contact

Table 10 provides the component parameter values you should enter for each data cleansing task.

Table 10. Data Cleansing Component Parameter Values

Data Cleansing Task	BusComp Name	Business Object Name	Object WHERE Clause	Operation Type	Reserved Option
Account	Account	Account	`Disable DataCleansing <> 'Y'`	Data Cleansing	0 (zero)
Address	Business Address[1]	Business Address[1]	`Disable DataCleansing <> 'Y'`	Data Cleansing	0 (zero)
Contact	Contact	Contact	`Disable DataCleansing <> 'Y'`	Data Cleansing	0 (zero)
Prospect	List Mgmt Prospective Contact	List Mgmt	`Disable DataCleansing <> 'Y'`	Data Cleansing	0 (zero)

1. For users of Siebel Industry Applications, the CUT Address business component should be used instead.

For information about how to configure data cleansing business components, see "Configuring Business Components to Support Data Cleansing" on page 58.

About Running Key Generation Using SDQ Matching Server in Batch Mode

Before you run data matching using the Siebel Data Quality Matching Server, you must first run key generation. For you to run key generation, you run a server component job. For more information about and procedures for running component jobs, see the chapter about using the Server Manager graphical user interface (GUI) in *Siebel System Administration Guide*. For information about running data matching for the Siebel Data Quality Matching Server and the Siebel Data Quality Universal Connector, see "About Running Data Quality Batch Mode Requests from the Command Line" on page 50.

NOTE: If you are using the Siebel Data Quality Universal Connector to run data matching, you can skip this section.

The business components preconfigured to support data matching using the Siebel Data Quality Matching Server are:

- Account
- Contact
- List Mgmt Prospective Contact

Table 11 provides the component parameter values you should enter for each key generation task.

Table 11. Component Job Parameter Values for Key Generation

Key Generation Task	BusComp Name	Business Object Name	Operation Type for Key Generation	Operation Type for Key Refresh
Account	Account	Account	Key Generate	Key Refresh
Contact	Contact	Contact	Key Generate	Key Refresh
Prospect	List Mgmt Prospective Contact	List Mgmt	Key Generate	Key Refresh

About Running Data Matching in Batch Mode for SDQ Matching Server and Universal Connector

For you to run data matching (deduplication) in batch mode for the Siebel Data Quality Matching Server and the Siebel Data Quality Universal Connector, you run a server component job. For more information about and procedures for running component jobs, see the chapter about using the Server Manager graphical user interface (GUI) in *Siebel System Administration Guide*.

The business components preconfigured to support batch mode data matching (deduplication) for both the Siebel Data Quality Matching Server and the Siebel Data Quality Universal Connector are:

- Account
- Contact
- List Mgmt Prospective Contact

Table 12 provides the component parameter values you should enter for each data matching (deduplication) task.

Table 12. Component Job Parameter Values for Data Matching (Deduplication)[1]

Data Matching Task	BusComp Name	Business Object Name	Operation Type
Account	Account	Account	DeDuplication
Contact	Contact	Contact	DeDuplication
Prospect	List Mgmt Prospective Contact	List Mgmt	DeDuplication

1. Supports data matching for the Siebel Data Quality Matching Server and the Siebel Data Quality Universal Connector.

About Running Data Quality Batch Mode Requests from the Command Line

You can start your batch mode component jobs using the srvrmgr program from the command line. For more information about running component jobs from the command line, see the chapter on using the Server Manager command-line interface in *Siebel System Administration Guide*.

Use the following syntax for your batch mode requests:

```
start 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, the following code starts a batch data cleansing request for account data:

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

For more information about running batch mode data cleansing using the Siebel Data Quality Universal Connector, see "About Running Data Cleansing in Batch Mode Using SDQ Universal Connector" on page 47.

About Customizing Data Quality Server Component Jobs for Batch Mode

After you configure the Data Quality Manager server component, you can use the Administration - Server Configuration views to preset the Siebel Data Quality Manager parameters, rather than setting them each time you run the server job. After you customize a component, the component is saved with the correct parameters for the task.

It is recommended that you create your customized component jobs by copying the Data Quality Manager (alias DQMgr) component definition and setting the parameters as provided in Table 13 on page 52 through Table 16 on page 54. For more information about creating custom component definitions, see *Siebel System Administration Guide*.

You may need to enable the Data Quality Manager component before the new settings take effect. For information about enabling the Data Quality Manager component, see "About Enabling the Data Quality Component Group for Batch Mode Requests" on page 47. For sample customization recommendations, see "Sample SDQ Component Customizations for Batch Mode" on page 51.

Sample SDQ Component Customizations for Batch Mode

You can customize Siebel Data Quality (SDQ) components for the Siebel Matching Server and the Universal Connector. This topic provides sample data you can use for customizing these components.

NOTE: It is recommended that you use the same component and alias names shown Table 13 on page 52 through Table 16 on page 54 to provide consistency in locating the relevant log files.

Sample Component Customization for SDQ Matching Server

Table 13 on page 52 through Table 15 on page 53 provide the recommended custom component definitions for Account, Contact, and Prospect objects for the SDQ Matching Server.

For more information about customizing SDQ components, see "About Customizing Data Quality Server Component Jobs for Batch Mode" on page 51.

Table 13. Recommended Custom Component Definitions for SDQ Matching Server for Accounts

Component Alias	Component Name	Description	Component Parameter	Value
DQMgrAcctKGen	DQ Account Key Generation	Data quality key generation for accounts	Buscomp Name	Account
			Business Object Name	Account
			Operation Type	Key Generate
DQMgrAcctKRef	DQ Account Key Refresh	Data quality key refresh for accounts	Buscomp Name	Account
			Business Object Name	Account
			Operation Type	Key Refresh
DQMgrAcctDDup	DQ Account Key DeDuplication	Data quality deduplication for accounts	Buscomp Name	Account
			Business Object Name	Account
			Operation Type	DeDuplication

Table 14. Recommended Custom Component Definitions for SDQ Matching Server for Contacts

Component Alias	Component Name	Description	Component Parameter	Value
DQMgrContKGen	DQ Contact Key Generation	Data quality key generation for contacts	Buscomp Name	Contact
			Business Object Name	Contact
			Operation Type	Key Generate
DQMgrContKRef	DQ Contact Key Refresh	Data quality key refresh for contacts	Buscomp Name	Contact
			Business Object Name	Contact
			Operation Type	Key Refresh
DQMgrContDDup	DQ Contact Key DeDuplication	Data quality deduplication for contacts	Buscomp Name	Contact
			Business Object Name	Contact
			Operation Type	DeDuplication

Table 15. Recommended Custom Component Definitions for SDQ Matching Server for Prospects

Component Alias	Component Name	Description	Component Parameter	Value
DQMgrPrspKGen	DQ Prospect Key Generation	Data quality key generation for prospects	Buscomp Name	List Mgmt Prospective Contact
			Business Object Name	List Mgmt
			Operation Type	Key Generate
DQMgrPrspKRef	DQ Prospect Key Refresh	Data quality key refresh for prospects	Buscomp Name	List Mgmt Prospective Contact
			Business Object Name	List Mgmt
			Operation Type	Key Refresh
DQMgrPrspDDup	DQ Prospect Key DeDuplication	Data quality deduplication for prospects	Buscomp Name	List Mgmt Prospective Contact
			Business Object Name	List Mgmt
			Operation Type	DeDuplication

Sample Component Customization for SDQ Universal Connector

Table 16 provides the recommended custom component definitions for Account, Contact, Prospects, and Address objects for the SDQ Universal Connector.

Table 16. Recommended Custom Component Definitions for SDQ Universal Connector

Component Alias	Component Name	Description	Component Parameter	Value
DQMgrAcctDCIns	DQ Account Data Cleansing	Data quality data cleansing for accounts	Buscomp Name	Account
			Business Object Name	Account
			Operation Type	Data Cleansing
DQMgrContDCIns	DQ Contact Data Cleansing	Data quality data cleansing for contacts	Buscomp Name	Contact
			Business Object Name	Contact
			Operation Type	Data Cleansing
DQMgrPrspDCIns	DQ Prospect Data Cleansing	Data quality data cleansing for prospects	Buscomp Name	List Mgmt Prospective Contact
			Business Object Name	List Mgmt
			Operation Type	Data Cleansing
DQMgrAddrDCIns	DQ Address Data Cleansing	Data quality data cleansing for addresses	Buscomp Name	Business Address
			Business Object Name	Business Address
			Operation Type	Data Cleansing

8 Data Quality Configuration Options

This chapter explains how to configure fields and add field types for data quality implementation. It includes the following topics:

Overview of Data Quality Configuration Options

The Siebel business components for accounts, contacts, and prospects have preconfigured fields that support data cleansing and data matching. You can modify these preconfigured fields and add new fields to include additional field types using Siebel Tools.

TIP: You can also add data quality functionality to more business components. However, these configuration changes may require changes to the Siebel Repository as well as table schema changes in the database. It is recommended you contact Siebel Technical Support for such changes.

CAUTION: You should be familiar with Siebel Tools before performing configuration tasks. For more information about using Siebel Tools, see *Using Siebel Tools*.

About Configuring Connector Mappings to External Vendors

In standard Siebel eBusiness Applications, three business components are enabled for data cleansing and data matching:

- Account
- Contact
- List Mgmt Prospective Contact

In addition, the Business Address (or CUT Address for Siebel Industry Applications) business component is enabled only for data cleansing. The Business Address business component does not support deduplication.

Each of these business components provides a set of user properties that define the correct field names and parameters that communicate with the external data quality vendor.

In the standard Siebel application, these business components have three user properties defined in the Siebel repository, as shown in Table 17. Each user property makes up a connector that can interface to an external data quality application.

Table 17. User Properties for Data Quality Business Components

User Property Name	Comments
DataCleansing Connector - Vendor 1	The value is configurable for accessibility to various vendors. The default value ("namefirm") is set for Firstlogic applications.
DeDuplication Connector - Vendor 1	The value is configurable for accessibility to various vendors. The default value ("match") is set for Firstlogic applications.
DeDuplication Connector - SSA	The connector mapping for the Siebel Data Quality Matching Server. You should not modify any values in this map.

For information about the values for each these user properties for use with the Siebel Data Quality Universal Connector, see Appendix A, "Preconfigured SDQ Universal Connector Properties for Firstlogic Applications." For information about viewing or changing the default values for each these user properties for use with the Siebel Data Quality Matching Server, see Appendix B, "Configuring Siebel Data Quality Matching Server Using SSA."

About Data Quality Field Mappings

The DeDuplication Connector - SSA is the one connector mapping for the Siebel Data Quality Matching Server. The two connector mappings for the Siebel Data Quality Universal Connector are the DataCleansing Connector and the Deduplication Connector.

For each connector there are a set of field mappings that indicate the Siebel business component field name and the field name in the external application that identifies the same information. For example, the Siebel field name for Account Name is Name, whereas the Firstlogic field name drl_ifirmname1.

You can specify one external application for data cleansing and a different application for data matching (deduplication). This is done by setting the correct input values for each external application in the Value column for the user properties for that connector.

Data Quality Deduplication Field Mapping Syntax

Use the following syntax for each field in the deduplication field mapping user properties:

- The user property name is labeled as: `[connector name value] DeDuplication Field [n]` where *n* is a sequential integer
- The user property value consists of a pair of strings enclosed in double quotation marks and separated by a comma and a space, such as:

  ```
  "external vendor field name", "Siebel business component field name"
  ```

  TIP: Make sure there is a comma and a space between the two field values. Deduplication can fail if this syntax is not adhered to.

For example, if the Vendor1 connector is identified as `"match"`, the user properties for the first three fields in the field mappings might be as shown in Table 18.

Table 18. Deduplication Field Mappings

User Property Name	Value
match DeDuplication Field 1	"MATCH_FIELD_FIRM", "Name"
match DeDuplication Field 2	"MATCH_FIELD_FIRMLOC", "Location"
match DeDuplication Field 3	"MATCH_FIELD_UNPADDRLINE", "Street Address"

Data Quality Data Cleansing Field Mapping Syntax

Use the following syntax for each field in the data cleansing field mapping user properties:

- The user property name is labeled as: `[connector name value] DataCleansing Field [n]` where *n* is a sequential integer
- The user property value consists of two or three strings enclosed in double quotation marks and separated by a comma and a space, such as:

  ```
  "Siebel Field", "Vendor Input Field"
  ```

  or

  ```
  "Siebel Field", "Vendor Input Field", "Vendor output field"
  ```

  TIP: Make sure there is a comma and a space between the two field values. Deduplication can fail if this syntax is not adhered to.

For example, for the connector identified as "namefirm", the two user properties preconfigured for the Account business component are shown in Table 19.

Table 19. Data Cleansing Field Mappings

User Property Name	Value
namefirm DataCleansing Field 1	"Name", "drl_ifirmname1", "drl_firm_std1"
namefirm DataCleansing Field 2	"Location", "drl_ifirmloc1", "drl_firm_loc_std1"

TIP: The data cleansing field mappings can use two or three values (the third value is optional). For some types of data, the connector can use the same field for input and output, while for other types of data the connector must use different fields for input and output. You need to consult with the third-party vendor to learn which field or fields to use. If you do not specify the Vendor output field, the Siebel application does not update the cleansed value to the field.

Configuring Business Components to Support Data Cleansing

You use Siebel Tools to define which fields for the Account, Contact, Prospect, or Business Address business components should be enabled for data cleansing. The installed product includes default settings for these business components to support integration to Firstlogic applications, but the fields can be configured according to your requirements or to support integration to other vendors. This section explains how to configure these business components for data cleansing. For more information about the default data quality configuration for these business components, see Appendix A, "Preconfigured SDQ Universal Connector Properties for Firstlogic Applications."

NOTE: For Siebel Industry Applications, the CUT Address business component is enabled for data cleansing rather than the Business Address business component.

In real-time mode, data cleansing is triggered when a record is saved after a field that is defined as an active data cleansing field is updated. An example of an active data cleansing field for the Contact business component is Last Name. For more information about real-time mode, see Chapter 7, "Working with Data Cleansing and Data Matching in Real-Time and Batch Modes."

TIP: Modifying an inactive field does not trigger data cleansing. Only fields that are indicated as data cleansing fields in the business component user properties trigger real-time data cleansing.

The data cleansing functionality is implemented in a Data Cleansing business service. Use the following procedure to use the Data Cleansing business service with a new Siebel business component.

To configure a business component to support data cleansing

1 Base the business component on the CSSBCBase class property to support real-time data cleansing.

NOTE: The CSSBCBase class includes the specific logic to invoke the data cleansing business services.

2 Associate the business component to a connector using user properties for the business component.

For example, add a new user property to the business component called DataCleansing Connector - VendorName. For more information, see "Configuring the SDQ Universal Connector" on page 63.

3 Create the field mappings between the Siebel fields that you want to cleanse and the field names recognized by the external vendor.

For more information, see "Configuring the SDQ Universal Connector" on page 63.

4 (Optional) If you want to prevent data cleansing on a selected record, perform the following:

a Add an extension column to the base table and map it to a business component field called Disable DataCleansing.

For example, the fields used in the Business Address business component are:

Field Name:	Disable DataCleansing
Column:	DISA_CLEANSE_FLG
Predefault value:	N
Text Length:	1
Type:	DTYPE_BOOL

b Map this field to your applet to disable data cleansing for certain records from the user interface.

5 (Optional) Configure a field called Last Clnse Date so that the Data Cleansing business service can mark the current date and time for the records.

After a record is cleansed, the Data Cleansing business service attempts to update the Last Clnse Date business component field to the current date and time. This field is useful for future batch data cleansing, because the administrator can apply an Object WHERE Clause to cleanse only records that have changed since the last cleanse date. For example, the following values appear in the Account business component:

Field Name:	Last Clnse Date:
Join:	S_ORG_EXT
Object Name. Column:	OBJ_NAME
Column:	DEDUP_DATACLNSD_DT
Type:	DTYPE_UTCDATETIME

6 (Optional) Use the DataCleansing Conflict Id Field user property to specify the conflict Id field for a business component.

In most implementations, user keys are defined in the database schema for each table. These user keys make sure that no more than one record has the same set of values in specific fields. For example, the S_ORG_EXT table used by the Account business component uses columns NAME, LOC (Location), and BU_ID (organization id) in the user keys. Before you run data cleansing against your database, you may have similar, but not exactly the same records, in your database.

After these records are cleansed, they may cause user key violations because the cleansed values become exactly the same value. You can use the Conflict Id field to resolve this issue. Add the CONFLICT_ID system column (given this table column exists in the database schema) to the user keys and then configure a user property called DataCleansing Conflict Id Field in that business component. The following example is for the Account business component:

User Property: DataCleansing Conflict Id Field
Property Value: S_ORG_EXT.Conflict Id

If a user key violation occurs when the Siebel application writes the cleansed records to the database, the application tries to update the Conflict Id field to the record's row Id to make the record unique and bypass the user key violation. After the entire database is cleansed, you can perform data matching to catch these records and resolve them.

CAUTION: Before modifying user keys, contact Siebel Technical Support.

Configuring Business Components to Support Data Matching (Deduplication)

The matching functionality is implemented into a DeDuplication business service. For you to use the DeDuplication business service with a new Siebel business component, use the following procedure.

To configure a business component to support data matching (deduplication)

1. Base the business component on the CSSBCBase class property to support real-time data matching.

   NOTE: This class includes the specific logic to invoke the DeDuplication business services.

2. Associate the business component to a connector using user properties for the business component.

   For example, add a new user property to the DeDuplication Connector - VendorName business component. For more information, see "Configuring the SDQ Universal Connector" on page 63.

3. Create the field mappings between the Siebel fields that you want to cleanse and the field names recognized by the external vendor.

   For more information, see "Configuring the SDQ Universal Connector" on page 63.

4. Configure a DeDuplication Results business component based on the S_DEDUP_RESULT table with the following field values:

Dup Object Id. Column:	DUP_OBJ_ID
Object Id. Column:	OBJ_ID
Object Name. Column:	OBJ_NAME
Request Id, Column:	DEDUP_REQ_ID
Total Score. Column:	TOT_SCORE_VAL

   The Siebel DeDuplication business service stores the ROW_ID of the matched pairs in the OBJ_ID and DEDUP_OBJ_ID columns. You can use these columns to join your business component to the primary data table to expose more information of the matched records.

   NOTE: The Siebel matching process uses the S_DEDUP_RESULT table to store the matched pairs with a weighted score. The DeDuplication Results business component is required to insert matched pairs into the S_DEDUP_RESULT table as well as display the duplicate records in a DeDuplication Results list applet to users.

5. Add the new DeDuplication Result business component to the DeDuplication business object.

6 Add the new business component to the business object of the view where you want to enable real-time deduplication.

   In your primary business component, add a user property called DeDuplication Results BusComp and specify the DeDuplication Results business component that you just configured.

7 Configure an applet as your DeDuplication Results List Applet based on the business component you configured in Step 4 on page 61.

   This applet is used to display the duplicate records for real-time mode.

   TIP: It is recommended you make a copy of an existing applet, such as the DeDuplication Results (Account) List Applet, and then make changes to the values (applet title, business component, and list columns). You may want to add join tables and fields to your DeDuplication Results business component and map these fields to your list applet so that you can see the duplicate records rather than their row Ids.

8 For you to trigger real-time matching, perform the following:

   a Modify the applet in which users enter or modify the customer data and base it on the CSSFrameListBase for a list applet or CSSFrameBase for a form applet.

   b Add a user property called DeDuplication Results Applet and specify the DeDuplication Results List Applet you configured in Step 7 in the value column.

9 For Siebel Data Quality Matching Server only. Create a match key table in the database and configure a match key business component.

   For more information. see Appendix B, "Configuring Siebel Data Quality Matching Server Using SSA."

10 Configure duplicate resolution views and add them to the Administration - Data Quality screen.

   It is recommended you use the Account Duplicates View and the Account Duplicates Detail View as your examples to learn how to configure new views.

11 Add a field called Merge Sequence Number to the business component and a user property called Merge Sequence Number Field.

   This configuration is used for sequenced merges. For more information about sequenced merges, see "About Sequenced Merges" on page 37.

   TIP: You do not need to map the Merge Sequence Number field to a database column. Instead, set the Calculated attribute to TRUE.

Configuring the SDQ Universal Connector

This section provides procedures for configuring the Siebel Data Quality (SDQ) Universal Connector.

Three processes are needed to support data quality features using the Siebel Data Quality Universal Connector. These processes are divided into the following three phases:

- "Phase 1: Creating a Connector Configuration" on page 63
- "Phase 2: Associating the Connector to a Business Component" on page 64
- "Phase 3: Mapping Connector Fields to Business Component Fields" on page 65

Each phase contains a set of instructions or steps. Perform the phases and the steps in the order provided.

NOTE: These processes do not cover vendor-specific configuration. You should work with Siebel-certified alliance partners to enhance data quality features for your applications. For more information, contact Siebel Technical Support.

Phase 1: Creating a Connector Configuration

The connector configuration is done in the Business Service user property.

To support:

- The data cleansing connector, you configure the Data Cleansing business service.
- The data matching connector, you configure the DeDuplication business service.

To configure a connector

1 Determine vendor name and vendor's solution names for the connector.

User Property: VendorName Connector *n*

Property Value: SolutionName

where *n* is a number starting from 1

NOTE: A vendor may have more than one solution for either data cleansing or data matching.

2 For each solution, you need to configure the following attributes:

User Property: SolutionName Attribute - Library Base Name
User Property: SolutionName Attribute - Language Dependent
User Property: SolutionName Parameter *n*
where *n* is a number starting from 1.

TIP: A solution may have more than one parameter.

The Library Base Name tells the Siebel application how to load the vendor's dynamic-link library (DLL). The vendors follow Siebel naming convention to build the DLLs. They first pick a base name for their libraries, for instance BASE. Then, the Windows DLL must be called BASE.dll. In AIX and Solaris, the libraries must be called libBASE.so. In HP-UX, the libraries must be called libBASE.sl. The Siebel application loads the DLL from the \bin subdirectory under the Siebel installation directory ($SIEBEL_ROOT/bin) for Windows or the /lib directory for UNIX.

For information about the preconfigured attributes for the Firstlogic connectors, see Appendix A, "Preconfigured SDQ Universal Connector Properties for Firstlogic Applications."

CAUTION: The Business Service user properties in the Siebel application are specifically designed to support multiple vendors in the Universal Connector architecture without the need for additional code. The values of these properties must be provided by Siebel Systems and third-party vendors. Usually these values cannot be changed because specific values are required by each software vendor.

Phase 2: Associating the Connector to a Business Component

In this phase you associate the connector to a Siebel business component.

To associate the connector to a business component

1 Use one of the following business component user properties to associate a connector to a business component:

- DeDuplication Connector - VendorName (for data matching)
- DataCleansing Connector - VendorName (for data cleansing)

The value of the property specifies the solution you want to use from this vendor for this business component.

NOTE: You can configure up to three connectors for a business component. This feature is especially useful for a company that has name cleansing and address cleansing in two different solutions. For example, the List Mgmt Prospective Contact business component uses both "namefirm" and "address" because Firstlogic implements these two functionalities into different connectors.

2 Configure a user property to assign a type name to your business component.

For example, assign SolutionName Type to the user property.

NOTE: This user property is not required, however it does provide a way for the DLL to recognize the type of records the Siebel application passes to it. The vendor can require this type as a predefined keyword. If it is not required as a keyword, developers can instead make the type a unique word so that the DLL identifies the type change when switching from one type to another in Siebel eBusiness Applications.

For information about the preconfigured settings, see Appendix A, "Preconfigured SDQ Universal Connector Properties for Firstlogic Applications."

Phase 3: Mapping Connector Fields to Business Component Fields

In this phase you map the connector fields to Siebel business component fields.

To map connector fields to business component fields

1 Define the data cleansing fields in the business component user properties as follows:

User Property: SolutionName DataCleansing Field *n*
Value: "Siebel Field", "Vendor Input Field", "Vendor output field"
or
Value: "Siebel Field", "Vendor Input Field"
where *n* is a number starting from 1.

TIP: You can have more than one field mapping.

NOTE: If you do not specify the Vendor output field, the Siebel application does not update the cleansed value to the field. In this case. this field is used only for input information. For example, in the Siebel application, the address records may have a predefined LOV (list of values) of State Names and Country Names, so that users can pick a value from the picklist. When you configure these fields as data cleansing fields, you do not want them updated. Instead, you should let the Siebel application pass these fields to the vendor DLL to make the cleansing more accurate.

2 Define the deduplication fields as follows:

User Property: SolutionName DeDuplication Field *n*
Value: "Vendor Input Field", "Siebel Field"
where *n* is the number starting from 1.

TIP: You can have more than one field mapping.

NOTE: You should contact the specific vendor for the list of fields they support for data cleansing and data matching.

About Troubleshooting Data Cleansing

If data cleansing is not working properly in real-time mode, check the following:

- **License key**. Verify that the license key allows Siebel Data Quality functionality.

  NOTE: There are different license keys for the Siebel Data Quality Matching Server and the Siebel Data Quality Universal Connector.

- **Application object manager configuration file**. Verify that data cleansing has been enabled for the application you are logged into. For more information, see "To enable real-time data cleansing for an application" on page 41.
- **User Preferences**. Verify that data cleansing has been enabled for the user. For more information, see "Setting User Preference Data Quality Settings" on page 33.
- **Third-party software**. Verify that the third-party software is installed and services for the third-party vendor are enabled.

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.
- **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.

9 Data Quality Performance Considerations

This chapter explains maintenance of the Siebel Data Quality Universal Connector and Siebel Data Quality Matching Server product modules as well as provides suggestions for optimizing their performance. It includes the following topics:

Optimizing Data Cleansing Performance

The following recommendations for data cleansing should help you achieve good performance when working with large volumes of data:

- You can include only new or recently modified records in the batch data cleansing process.

  If you run data cleansing on a record twice, sometimes the record can change the second time. However, cleansing all records in the Siebel database each time a data cleansing is performed can cause performance issues. It is recommended you include only new or recently modified records in the batch data cleansing process. These records can be identified using the Object WHERE clause when you submit your server component job, as shown in Table 20.

Table 20. Recommended Data Cleansing Object WHERE Clause Solutions

To Cleanse	Use This in Your Object WHERE Clause
Updated records	[Last Clnse Date] < [Updated]
New records	[Last Clnse Date] IS NULL
Updated and new records	[Last Clnse Date] < [Updated] OR [Last Clnse Date] IS NULL

- You can copy address files to your local machine.

  Address data cleansing (for business address and prospect data) needs to access the address data files frequently, so you should copy these files to your local machine.

- You can set the ReservedOption parameter.

  For you to speed up the data cleansing task for large databases, set the ReservedOption component parameter to 0 (ReservedOption=0), and then cleanse a smaller number of records at a time using an Object WHERE clause. For more information, see "About Running Data Cleansing in Batch Mode Using SDQ Universal Connector" on page 47.

- You can split the tasks into smaller tasks and run them concurrently.

Optimizing Data Matching Performance

The following recommendations for data matching should help you achieve good performance when working with large volumes of data:

- You can work with a database administrator to verify that the SIEBEL_4K table space is large enough to hold the records generated during the deduplication process.

  During the batch deduplication process, the information of the deduplication records is stored in the S_DEDUP_RESULT table in the format of a pair of row Ids of the duplicate records and the match scores between them. The number of records in the results table S_DEDUP_RESULT can include up to six times the number of records in your account and contact tables combined. You should consider the following:

  - If the base tables include many deduplicates, more records are inserted in the results table.
  - If different search types are used, a different number of duplicate records may be found and are inserted in the results table.
  - If you use a low match threshold (in the lower range of 100), the matching process generates more records to the results table.

- You can remove obsolete matching results records manually.

  When a duplicate record is detected, it is automatically placed in the S_DEDUP_RESULT table, whether or not the same duplicate record exists in that table. Running multiple batch deduplication tasks results in a large number of duplicate records in the these tables. Therefore, it is recommended that you manually remove the existing records in the S_DEDUP_RESULT tables before running a new batch deduplication task. You can remove the records using any utility that allows you to submit SQL statements. For more information about running batch deduplication, see Chapter 7, "Working with Data Cleansing and Data Matching in Real-Time and Batch Modes."

  NOTE: Removing the records from the S_DEDUP_RESULT tables does not cause a loss of data because these tables are again populated when a new batch deduplication task is run.

Optimizing SDQ Matching Server Performance

The three key tables S_PER_DEDUP_KEY, S_PRSP_DEDUPKEY, and S_ORG_DEDUP_KEY may include six times more records than their corresponding base tables, depending on the key type used in the key generation stage. For limited key type, they may include two to four times more records. For standard key type, they may include at least six times more records. Work with a database administrator and follow recommendations for the Siebel Data Quality Matching Server to optimize the parameters for your database:

- You can 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.
- When you want to run key generation, you can remove all keys in the key tables first.
- When you want to run deduplication, you can remove all records in the result table first.
- You can allocate space for the SIEBEL_4K table space following the sizing recommendations in Table 21.

Table 21. Table Size Recommendations

Table	Sizing Recommendation
S_PER_DEDUP_KEY S_ORG_DEDUP_KEY S_PRSP_DEDUPKEY	These tables may include between two and six times more records than their corresponding base tables, depending on the key type used during the key generation stage, as follows: ■ **Limited key type.** May include between two and four times more records than the corresponding base table. ■ **Standard key type.** May include up to an estimated six times more records than the corresponding base table.
S_DEDUP_RESULT	This table may include between five and six times the number of records in the three base tables combined. Use the following guidelines to help determine table size: ■ If a Typical or Exhaustive search type is used, more records are inserted into the results table. ■ If a low match threshold is used, such as a threshold in the lower 100 range, the matching process generates a larger number of records that are inserted into the results table.

- For the DB2 platform, you can use the following REORG utility commands on the DEDUP_KEY column after key generation on the key tables. The DEDUP_KEY column is based on the M1 clustered index.
  - `reorgchk current statistics on table siebel.S_PER_DEDUP_KEY`

    This command checks F4 for the M1 index to see whether a reorganization is needed.
  - `reorg table siebel.S_PER_DEDUP_KEY`

    This command reorganizes the table and usually takes about 30 minutes to run on 1 to 2 million records.
  - `reorgchk update statistics on table siebel.S_PER_DEDUP_KEY`

    This command updates the statistics.

- For the DB2 platform, if your performance seems degraded, you can run the following command on all tables associated with SDQ. (This includes tables such as: S_PER_DEDUP_KEY, S_ORG_DEDUP_KEY, S_ORG_EXT, S_PRSP_CONTACT, S_CONTACT, S_PRSP_CONTACT, S_PARTY, S_PARTY_PER, and S_DEDUP_RESULT).

  ```
  runstats on table siebel.S_PER_DEDUP_KEY
  ```

  If the above command returns an error message, use this one instead:

  ```
  runstats on table Siebel.S_CONTACT with distribution indexes all
  ```

- After your initial deduplication or key generation, you can include only new and updated records in deduplication and key generation processes.

  TIP: If you have more than 100,000 records in your base tables, it is time consuming to reprocess all of them.

- When performing deduplication and key generation, you can exclude records that are up to date. You use the DeDup Key Modification Date and DeDup Last Match Date business component fields in your search specifications to exclude records. For example:

  - You can add one of the following to the search specification (Object WHERE Clause) for key generation:

    - For updated records: `([DeDup Key Modification Date] < [Updated])`
    - For new records: `([DeDup Key Modification Date] IS NULL)`
    - For updated and new records (same as key refresh):

      ```
      ([DeDup Key Modification Date] < [Updated]) OR ([DeDup Key Modification Date] IS NULL)
      ```

  - You can add one of the following to the search specification (Object WHERE Clause) for deduplication:

    - For updated records: `([DeDup Last Match Date] < [Updated])dd`
    - For new records: `([DeDup Last Match Date] IS NULL))`
    - For updated and new records:

      ```
      (([DeDup Last Match Date] < [Updated]) OR ([DeDup Last Match Date] IS NULL))
      ```

- You can set the Data Quality Settings values as shown in Table 22.

Table 22. Performance Improvement Suggestions

Improvement for...	Comments
Data Quality Settings	From the application-level menu, choose Navigate > Site Map > Administration - Data Quality > Data Quality Settings to set the following parameters: ■ Key Type (key generation). Set to Limited. ■ Search Type (deduplication). Set to Narrow. ■ Match Threshold (deduplication). Set to a number greater than or equal to 75. The higher the threshold, the faster the deduplication process runs.
Object sort clause	Set this parameter based on the key generation parameters for deduplication. For example, use: ■ Person (contact or prospect). Use Last Name, First Name, Middle Name ■ Company (account). Use Name or Name, Location

- You can set the object sort clause based on the key generation parameters (deduplication) described in Table 22.
- You can execute concurrent Data Quality Manager server tasks to deduplicate data.

  For more information, see the following procedure.

To execute concurrent Data Quality Manager server tasks to deduplicate data

1 Query the base tables to find the search specifications so that each query result set contains between 50,000 and 75,000 records.

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

For example, the following table provides search specification solutions using the Object WHERE Clause to run key generation or deduplication.

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))

2 Start concurrent server tasks using the search specifications provided in Step 1 until the entire table is completely processed.

Optimizing SDQ Universal Connector Using Firstlogic

The following recommendations should help you achieve good performance when working with large volumes of data using Firstlogic software.

- Make sure you have enough disk space for your SDQConnector directory

  Firstlogic creates files under this directory to store the key information of records retrieved from your Siebel database. The size of files varies depending on the number of records in your Account or Contact table. You may need to contact Firstlogic to determine how much disk space is required for your implementation. You can use the Disk Defragment utility to reduce the file fragments so that Firstlogic can perform faster file I/O operations. For more information about using the Disk Defragment utility, contact your company's IT Department.

- Modify the mpbreak.cfg and fmin_rul.cfg files

  The Account, Contact, and Prospect business components each have a different set of configuration files. You can find these configuration files in the Siebel/SDQConnector/match/directory. For mpbreak.cfg, you can specify how you want Firstlogic to break data records into smaller groups. Then, Firstlogic compares records *among* each group but not *across* the groups. The fmin_rule.cfg file specifies the match rules. For more information about how to change the break group and match rules, you should contact Firstlogic technical support.

A Preconfigured SDQ Universal Connector Properties for Firstlogic Applications

This appendix provides preconfigured user property values for the Siebel Data Quality (SDQ) Universal Connector for use with Firstlogic applications and explains how to work with these properties in Siebel Tools. It includes the following topics:

For more information about configuring these properties, see Chapter 8, "Data Quality Configuration Options."

Viewing or Changing SDQ Business Service User Properties

Business service user properties in the Siebel application store configurable global settings for the Siebel Data Quality Universal Connector. Table 23 and Table 24 on page 76 provide the business service user property values for use with the Siebel Data Quality Universal Connector for Firstlogic applications.

This section explains how to view or change business service user properties in general and then provides values for these properties specific to Firstlogic applications. The procedures described are:

To view or change Data Cleansing business service user properties

1. Log in to Siebel Tools.

   For more information about Siebel Tools, see *Using Siebel Tools*.

2. In the Object Explorer, expand the Business Service object.
3. In the Business Services list, query for Data Cleansing.
4. In the Object Explorer, click Business Service User Prop.

   The default Data Cleansing user properties appear in the Business Service User Props list.

Table 23 provides the Data Cleansing business service user properties for Firstlogic applications.

Table 23. Data Cleansing Business Service User Property for Firstlogic Applications

Property Name	Value
Vendor1 Connector 1[1]	namefirm
Vendor1 Connector 2[1]	address
address Attribute - Library Base Name	sdqaddress
namefirm Attribute - Library Base Name	sdqname

1. By default, Vendor1 is mapped to fields for the Firstlogic data quality link for Siebel eBusiness Applications.

To view or change DeDuplication business service user properties

1. Log in to Siebel Tools.

   For more information about Siebel Tools, see *Using Siebel Tools*.
2. In the Object Explorer, expand the Business Service object.
3. In the Business Service list, query for Deduplication.
4. In the Object Explorer, click Business Service User Prop.

   The default Deduplication user properties appear in the Business Service User Props list.

Table 24 provides the DeDuplication business service user properties for Firstlogic applications.

Table 24. DeDuplication Business Service

Property Name	Value[1]
Vendor1 Connector 1[2]	match
match Attribute - Library Base Name	sdqmatch
match Parameter 1	"global", "MATCH_FIELD_LASTNAME", "50"
match Parameter 2	"global", "MATCH_FIELD_FIRSTNAME", "50"
match Parameter 3	"global", "MATCH_FIELD_MIDDLENAME", "50"
match Parameter 4	"global", "MATCH_FIELD_FIRM", "100"
match Parameter 5	"global", "MATCH_FIELD_FIRMLOC", "50"
match Parameter 6	"global", "MATCH_FIELD_UNPADDRLINE", "200"
match Parameter 7	"global", "MATCH_FIELD_CITY", "50"

Table 24. DeDuplication Business Service

Property Name	Value[1]
match Parameter 8	"global", "MATCH_FIELD_STATE", "10"
match Parameter 9	"global", "MATCH_FIELD_ZIP4", "30"
match Parameter 10	"global", "MATCH_FIELD_COUNTRY", "30"

1. The values with three strings show type of parameter, parameter name, and parameter value, in that order. These parameters are provided by third-party vendors and cannot be changed.

2. By default, Vendor1 is mapped to fields for the Firstlogic data quality link for Siebel eBusiness Applications.

Viewing or Changing SDQ Business Component User Properties

Business component user properties store configurable settings specific to a business component, in particular, field mappings from the vendor's field to the Siebel field. The properties are available for both the Siebel Data Quality Universal Connector and the Siebel Data Quality Matching Server.

This section explains how to view or change business component user properties in general and then provides values for these properties specific to Firstlogic applications.

TIP: If you deactivate a field mapping, you may need to shift the remaining fields ahead so that your list of field mappings is sequential. For example, if you deactivate field 2, you may need to change field 3 to field 2, field 4 to field 3, and so on. If you do not shift the remaining fields, the Siebel application recognizes only one field mapping and cannot find field 2 or subsequent fields.

To view or change SDQ business component user properties

1. Log in to Siebel Tools.

   For more information about Siebel Tools, see *Using Siebel Tools*.

2. In the Object Explorer, expand the Business Component object.

3. In the Business Objects list, select the business component.

   For this example, select Account.

4. In the Object Explorer, click Business Component User Prop.

   The default business component user properties appear in the Business Component User Props list.

TIP: You can use a similar procedure for viewing the other data quality business component default user properties.

Table 25 and Table 26 provide the default data cleansing and deduplication user property values for the Account business component for Firstlogic applications.

Table 25. Account Business Component User Properties (Data Cleansing)

Property Name	Value[1]
DataCleansing Connector - Vendor1[2]	"namefirm"
namefirm DataCleansing Field 1	"Name", "drl_ifirmname1", "drl_firm_std1"
namefirm DataCleansing Field 2	"Location", "drl_ifirmloc1", "drl_firm_loc_std1"

1. The values with three strings show type of parameter, parameter name, and parameter value, in that order. These parameters are provided by third-party vendors and cannot be changed.

2. By default, Vendor1 is mapped to fields for the Firstlogic data quality link for Siebel eBusiness Applications.

Table 26. Account Business Component User Properties (Deduplication)

Property Name	Value[1]
DeDuplication Connector - Vendor1[2]	"match"
match Type	Account
match DeDuplication Field 1	"MATCH_FIELD_FIRM", "Name"
match DeDuplication Field 2	"MATCH_FIELD_FIRMLOC", "Location"
match DeDuplication Field 3	"MATCH_FIELD_UNPADDRLINE", "Street Address"
match DeDuplication Field 4	"MATCH_FIELD_CITY", "City"
match DeDuplication Field 5	"MATCH_FIELD_STATE", "State"
match DeDuplication Field 6	"MATCH_FIELD_ZIP4", "Postal Code"
match DeDuplication Field 7	"MATCH_FIELD_COUNTRY", "Country"
match DeDuplication Field 8	"MP_KEYFLD_KEYMISC", ""

1. The values with three strings show type of parameter, parameter name, and parameter value, in that order. These parameters are provided by third-party vendors and cannot be changed.

2. By default, Vendor1 is mapped to fields for the Firstlogic data quality link for Siebel eBusiness Applications.

Table 27 and Table 28 provide the default data cleansing and deduplication user property values for the Contact business component for Firstlogic applications.

Table 27. Contact Business Component User Properties (Data Cleansing)

Property Name	Value[1]
DataCleansing Connector - Vendor1[2]	"namefirm"
namefirm DataCleansing Field 1	"Last Name", "drl_ilname1", "drl_last_name_std1"
namefirm DataCleansing Field 2	"First Name", "drl_ifname1", "drl_first_name_std1"
namefirm DataCleansing Field 3	"Middle Name", "drl_imname1", "drl_mid_name_std1"
namefirm DataCleansing Field 4	"Job Title", "drl_ititle1", "drl_title_std1"

1. The values with three strings show type of parameter, parameter name, and parameter value, in that order. These parameters are provided by third-party vendors and cannot be changed.

2. By default, Vendor1 is mapped to fields for the Firstlogic data quality link for Siebel eBusiness Applications.

Table 28. Contact Business Component User Properties (Deduplication)

Property Name	Value[1]
DeDuplication Connector - Vendor1[2]	"match"
match Type	Contact
match DeDuplication Field 1	"MATCH_FIELD_LASTNAME", "Last Name"
match DeDuplication Field 2	"MATCH_FIELD_FIRSTNAME", "First Name"
match DeDuplication Field 3	"MATCH_FIELD_MIDDLENAME", "Middle Name"
match DeDuplication Field 4	"MATCH_FIELD_FIRM", "Account"
match DeDuplication Field 5	"MATCH_FIELD_FIRMLOC", "Account Location"
match DeDuplication Field 6	"MATCH_FIELD_UNPADDRLINE", "Street Address"
match DeDuplication Field 7	"MATCH_FIELD_CITY", "City"
match DeDuplication Field 8	"MATCH_FIELD_STATE", "State"
match DeDuplication Field 9	"MATCH_FIELD_ZIP4", "Postal Code"
match DeDuplication Field 10	"MATCH_FIELD_COUNTRY", "Country"
match DeDuplication Field 11	"MP_KEYFLD_KEYMISC", ""

1. The values with three strings show type of parameter, parameter name, and parameter value, in that order. These parameters are provided by third-party vendors and cannot be changed.

2. By default, Vendor1 is mapped to fields for the Firstlogic data quality link for Siebel eBusiness Applications.

Table 29 and Table 30 on page 80 provide the default data cleansing and deduplication user property values for the List Mgmt Prospective Contact business component for Firstlogic applications.

Table 29. List Mgmt Prospective Contact Business Component User Properties (Data Cleansing)

Property Name	Value[1]
DataCleansing Connector - Vendor1[2]	"namefirm", "address"
namefirm DataCleansing Field 1	"Last Name", "drl_ilname1", "drl_last_name_std1"
namefirm DataCleansing Field 2	"First Name", "drl_ifname1", "drl_first_name_std1"
namefirm DataCleansing Field 3	"Middle Name", "drl_imname1", "drl_mid_name_std1"
namefirm DataCleansing Field 4	"Job Title", "drl_ititle1", "drl_title_std1"
namefirm DataCleansing Field 5	"Account", "drl_ifirmname1", "drl_firm_std1"
namefirm DataCleansing Field 6	"Primary Account Location", "drl_ifirmloc1", "drl_firm_loc_std1"
address Type	Prospect Address
address DataCleansing Field 1	"Street Address", "nad_line_address1", "nad_line_address1"
address DataCleansing Field 2	"City", "nad_line_locality1", "nad_line_locality1"
address DataCleansing Field 3	"State", "nad_line_region1"
address DataCleansing Field 4	"Postal Code", "nad_line_postcode", "nad_line_postcode"
address DataCleansing Field 5	"Country", "nad_line_country"

1. The values with three strings show type of parameter, parameter name, and parameter value, in that order. These parameters are provided by third-party vendors and cannot be changed.

2. By default, Vendor1 is mapped to fields for the Firstlogic data quality link for Siebel eBusiness Applications.

Table 30. List Mgmt Prospective Contact Business Component User Properties (Deduplication)

Property Name	Value[1]
DeDuplication Connector - Vendor1[2]	"match"
match Type	Prospect
match DeDuplication Field 1	"MATCH_FIELD_LASTNAME", "Last Name"
match DeDuplication Field 2	"MATCH_FIELD_FIRSTNAME", "First Name"
match DeDuplication Field 3	"MATCH_FIELD_MIDDLENAME", "Middle Name"
match DeDuplication Field 4	"MATCH_FIELD_FIRM", "Account"
match DeDuplication Field 5	"MATCH_FIELD_UNPADDRLINE", "Street Address"
match DeDuplication Field 6	"MATCH_FIELD_CITY", "City"

Table 30. List Mgmt Prospective Contact Business Component User Properties (Deduplication)

Property Name	Value[1]
match DeDuplication Field 7	"MATCH_FIELD_STATE", "State"
match DeDuplication Field 8	"MATCH_FIELD_ZIP4", "Postal Code"
match DeDuplication Field 9	"MATCH_FIELD_COUNTRY", "Country"
match DeDuplication Field 10	"MATCH_FIELD_FIRMLOC", "Primary Account Location"
match DeDuplication Field 11	"MP_KEYFLD_KEYMISC", ""

1. The values with three strings show type of parameter, parameter name, and parameter value, in that order. These parameters are provided by third-party vendors and cannot be changed.

2. By default, Vendor1 is mapped to fields for the Firstlogic data quality link for Siebel eBusiness Applications.

Table 31 provides the default data cleansing user property values for the Business Address business component for Firstlogic applications.

Table 31. Business Address Business Component User Properties (Data Cleansing)

Property Name	Value[1]
DataCleansing Connector - Vendor1[2]	n"address"
address Type	Business Address
address DataCleansing Field 1	"Street Address", "nad_line_multiline1", "nad_line_multiline1"
address DataCleansing Field 2	"City", "nad_line_locality1", "nad_line_locality1"
address DataCleansing Field 3	"State", "nad_line_region1"
address DataCleansing Field 4	"Postal Code", "nad_line_postcode", "nad_line_postcode"
address DataCleansing Field 5	"Country", "nad_line_country"
address DataCleansing Field 6	"Street Address 2", "nad_line_multiline2", "nad_line_multiline2"

1. The values with three strings show type of parameter, parameter name, and parameter value, in that order. These parameters are provided by third-party vendors and cannot be changed.

2. By default, Vendor1 is mapped to fields for the Firstlogic data quality link for Siebel eBusiness Applications.

NOTE: You can run data cleansing for the Personal Address business component by configuring the Personal Address business component user properties and the following parameters: Operation Type = Data Cleansing, Business Component = Personal Address, and Business Object = Contact.

B Configuring Siebel Data Quality Matching Server Using SSA

This appendix provides an example of how to configure the Siebel Data Quality (SDQ) Matching Server using the embedded SSA-NAME3 software. It includes the following topics:

Process of Configuring SDQ Matching Server Using SSA

To configure the Siebel Data Quality (SDQ) Matching Server using SSA, perform the following tasks:

1. Associate SSA with a business component.

   For an example using the Account business component, see "Example of Associating SSA to a Business Component" on page 83.

2. Map SSA fields to fields in the Siebel business component you created in Step 1.

   For an example using the Account business component, see "Example of Mapping SSA Fields to Fields in a Siebel Business Component" on page 87.

NOTE: The Account business component is already enabled for DeDuplication, but you can use a similar process to enable deduplication for business components other than the default components (Account, Contact, List Mgmt/List Mgmt Prospective Contact).

Example of Associating SSA to a Business Component

This topic provides a sample configuration using the Account business component as an example. Use the following procedure as an example to associate SSA to a Siebel business component.

NOTE: This procedure is provided only as an example—the Account business component is already enabled for DeDuplication. Use a similar procedure to enable deduplication for business components other than the default components, which are Account, Contact, List Mgmt/List Mgmt Prospective Contact.

To associate SSA with a business component

1. Log in to Siebel Tools.

   For more information about Siebel Tools, see *Using Siebel Tools*.

2 Set the DeDuplication Connector - SSA business component user property value to SSA.

   a In the Object Explorer, expand the Business Component object.

   b In the Business Components list, select Account.

   c In the Object Explorer, click Business Component User Prop.

   d In the Business Component User Props list, query for DeDuplication Connector - SSA.

   e In the Value field, enter "SSA".

3 Assign a type name to your business component.

   a With the Account business component selected, in the Business Component User Properties list, query for SSA Match Purpose.

   b In the value field, enter Company_Mandatory.

   The value choices for the Account business component are: Company_Mandatory and Company_Optional. The value choices for the Contact business component are: Contact_Mandatory and Contact_Optional.

   NOTE: The preceeding four choices are the only Match Purpose values supported by Siebel Data Quality (SDQ).

   By default, the Account business component is set to Company_Mandatory, and the Contact and Prospect business components are set to Contact_Optional. 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.

   NOTE: SSA supports two types of deduplication: company and contact. For more information, see the SSA-NAME3 documentation that is included on the *Siebel eBusiness Third-Party Bookshelf*.

4 Configure the DeDuplication Key business component user property to specify the business component for your key table.

   a Create a key table.

   A key table is a database table that stores the SSA keys used for matching. You can use one of the following existing key tables as a model:

   - S_ORG_DEDUP_KEY
   - S_PER_DEDUP_KEY
   - S_PRSP_DEDUPKEY

   In this example for the Account business component, the S_ORG_DEDUP_KEY key table is used.

   NOTE: The Data Quality Matching Server requires a key table for each business component (whereas the Siebel Data Quality Universal Connector does not). Due to the complexity of creating database tables, it is recommended you contact your database administrator for key table creation.

   b Create a new business component based on the key table you created in Step a.

   You can use one of the following existing key business components as a model:

- DeDuplication - SSA Account Key
- DeDuplication - SSA Contact Key
- DeDuplication - SSA Prospect Key

In this example for the Account business component, the DeDuplication - SSA Account Key business component is used.

c Define the user properties for the business component you created in Step b.

You can use the existing user properties for the following business components as a model:

- DeDuplication - SSA Account Key
- DeDuplication - SSA Contact Key
- DeDuplication - SSA Prospect Key

In this example for the Account business component, the DeDuplication - SSA Account Key business component is used.

For information about how to create business components and define user properties, see *Configuring Siebel eBusiness Applications*.

5 Create a link and the Algorithm Type field for the business component and the key business component you created in Step 4 on page 84.

a Create a link using the following syntax:

<Business Component name>/DeDuplication - SSA <Business Component name> Key

In this example for the Account business component, the link is called: Account/DeDuplication - SSA Account Key.

b Navigate to the Business Component object type and create a multi-value link for the business component you created in Step 4 on page 84 using the following properties:

- Name = DeDuplication - SSA <Business Component name> Key
- Destination Link = <Business Component name>/DeDuplication - SSA <Business Component> Key
- Destination Business Component = DeDuplication - SSA <Business Component name> Key.

In this example for the Account business component, the values are:

- Name = DeDuplication - SSA Account Key
- Destination Link = Account/DeDuplication - SSA Account Key
- Destination Business Component = DeDuplication - SSA Account Key

c Navigate to the Field object type and create a new field for the business component you created in Step 4 on page 84 with the following properties:

- Name = Algorithm Type

- Multivalue Link = DeDuplication - SSA <Business Component name> Key

  In this example for the Account business component, the multi-value link is DeDuplication - SSA Account Key.

For more information about links and multi-value links, see *Configuring Siebel eBusiness Applications*.

6 Configure the DeDup Key Modification Date and DeDup Last Match Date fields for your business component.

a In the Object Explorer, expand the Business Component object.

b In the Business Components list, select Account.

c In the Object Explorer, click Field.

d In the Fields list, query for DeDup Key Modification Date, and enter values.

The following are values for the Account business component:

Field Name:	DeDup Key Modification Date
Join:	S_ORG_EXT
Column:	DEDUP_KEY_UPD_DT
Type:	DTYPE_UTCDATETIME

TIP: After a record is processed during key generation, the DeDuplication business service updates the DeDup Key Modification Date field to the current date and time. This is useful for future batch generations because you can run a key refresh instead of a more time consuming key generation.

e In the Fields list, query for DeDup Last Match Date, and enter values.

The following are values for the Account business component:

Field Name:	DeDup Last Match Date
Join:	S_ORG_EXT
Column:	DEDUP_LAST_MTCH_DT
Type:	DTYPE_UTCDATETIME

TIP: After a record is processed during deduplication, the DeDuplication business service updates the DeDup Last Match Date field to the current date and time. This is useful for future batch deduplication because you can set an Object WHERE Clause to process records that have not changed since the last match date.

Example of Mapping SSA Fields to Fields in a Siebel Business Component

This topic provides a sample configuration using the Account business component as an example. Use the following procedure as an example to map SSA fields to fields in a Siebel business component.

To map SSA fields to fields in a Siebel business component

1. Log in to Siebel Tools.

   For more information about Siebel Tools, see *Using Siebel Tools*.

2. Define the deduplication field mappings user properties for the Account business component.
   a. In the Object Explorer, expand the Business Component object.
   b. In the Business Components list, select Account.
   c. In the Object Explorer, click Business Component User Prop.
   d. In the Business Component User Props list, query for SSA DeDuplication Field*, and enter values using the following syntax.

Property Name	Value
SSA DeDuplication Field *n*, where *n* is a number starting from 1.	"Vendor Input Field", "Siebel Field"

NOTE: You can have more than one field mapping.

TIP: If you deactivate a field mapping, you may need to shift the remaining fields ahead so that your list of field mappings is sequential. For example, if you deactivate field 2, you may need to change field 3 to field 2, field 4 to field 3, and so on. If you do not shift the remaining fields, the Siebel application recognizes only one field mapping and cannot find field 2 or subsequent fields.

Table 32, Table 33 on page 88, and Table 34 on page 89 provide the user property values for the Account, Contact, and List Mgmt Prospective Contact business components, respectively. For information about configuring these properties, see Chapter 8, "Data Quality Configuration Options."

Table 32 provides the user properties for the Account business component.

Table 32. Account Business Component User Properties

User Property Name	Value[1]
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 32. Account Business Component User Properties

User Property Name	Value[1]
SSA DeDuplication Field 5	"Z", "Postal Code"
SSA DeDuplication Field 6	"I", "DUNS Number"
SSA DeDuplication Field 7	"A", "Street Address"

1. The syntax for the values is: "Vendor Input Field", "Siebel Field".

Table 33 provides the user properties for the Contact business component.

Table 33. Contact Business Component User Properties

User Property Name	Value[1]
SSA DeDuplication Field 1	"Z", "Postal Code"
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"
SSA DeDuplication Field 10	"First", "First Name"
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"

1. The syntax for the values is: "Vendor Input Field", "Siebel Field".

Table 34 provides the user properties for the List Mgmt Propitiate Contact business component.

Table 34. List Mgmt Prospective Contact Business Component User Properties

User Property Name	Value[1]
SSA DeDuplication Field 1	"Last", "Last Name"
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"
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 #"

1. The syntax for the values is: "Vendor Input Field", "Siebel Field".

C Using the Value Match Method

This appendix explains the Value Match method and provides a scenario for using the value match feature. It covers the following topics:

About the Value Match Method

Value Match is one of the methods of the DeDuplication business service. Use the Value Match method for tasks that require matching of data in field/value pairs against the data within Siebel business components or when you want to prevent duplicate data from getting into the Siebel application through non-UI data streams. You invoke the Value Match method as you would any other business service method.

For more information about Siebel Workflow, see *Siebel Business Process Designer Administration Guide*. For more information about business services and methods, see *Siebel Developer's Reference*.

Scenario for Data Matching Using the Value Match Method

This topic gives one example of how you can invoke the Value Match business service method using Siebel Workflow. You may use the Value Match method differently, depending on your business model.

Your company needs to import a list of contacts into the Siebel application from other systems in your enterprise. To avoid introducing duplicate contacts in the Siebel application, you might consider invoking Siebel eScript through a workflow process to call the Value Match business service method. In this case, you invoke the Value Match method as a step in your contact import process. This step matches incoming contact information against the contacts within the Siebel business components. To prevent the introduction of duplicate information into the Siebel application, you add processing logic to your script based on the results returned from the value match parameter.

In this scenario, the business component name of the data entity that you want matched must be specified. For example, if you are using SSA to match contacts, then you set the business component name to Contact. Also, make sure the name/value pairs are Siebel field names that are configured for matching (the fields for contact matching appear as values for SSA Deduplication field 1, SSA Deduplication field 2, and so on, of the Contact business component user property). If the input contains fields that are not configured for matching, then those fields do not factor into the match. For example, if Personal Assistant is included in the list of values, but Personal Assistant is not a deduplication field of the matched business component, then this input is ignored.

For information about what the Value Match method does and how to use it, see "Value Match Method" on page 92.

Value Match Method

Value Match is one of the methods of the DeDuplication business service. The Value Match method is used to match data in field/value pairs against the data within Siebel business components or to prevent duplicate data from getting into the Siebel application through non-UI data streams.

Arguments

The Value Match method arguments consist of input and output property sets. Table 35 provides the input property sets, and Table 36 on page 93 provides the output property sets.

CAUTION: The Value Match property sets are specialized. Do not configure these components.

Table 35. Value Match Input Property Sets

Input Property Set	Property	Description	Comments
Adapter Settings[1]	Code Page	Code page.	Optional. Override. Applicable only to SSA. **NOTE:** Properties specified as Override can be used to override the setting information obtained by the service from the administration screens, .cfg file, user properties, and so on.
	Population	Population values.	
	SearchLevel	The search level.	
	Threshold	The threshold score for a duplicate record. A match is considered only if the score exceeds this value.	

Table 35. Value Match Input Property Sets

Input Property Set	Property	Description	Comments
Generic Settings	BC Name	The name of the matched business component.	Required.
	Update Modification Date	If set to N, the match modification date is not updated.	Not required. Default =Y.
	Use Result Table	If set to N, matches are not added to the result table. Instead, matches are determined by the business service.	Not required. Default = Y.
Match Values[1]	<business component field names, value> pairs For example: (Last Name, 'ABC') (First Name, 'XYZ') and so on	The matched business component's field name and the corresponding field value. For example: "Last Name", "Brown"	These name, value pairs are used as the matched value rather than the current row ID of the matched business component. The user properties of the matched business component is used to translate the business component field names to vendor specific field names.

1. Adapter Settings and Match Values are child property sets of the input property set.

Return Value

For each match, a separate child property set called Match Info is returned in the output with properties specific to the match (such as Matchee Row ID and Score), as well as some general output parameters as shown in Table 36.

CAUTION: The Value Match property sets are specialized. Do not configure these components.

Table 36. Value Match Output Property Sets

Output Property Set[1]	Property	Description	Comments
Match Info	Matchee Row Id	The row ID of the matchee.	
	Score	The score of the match.	Matchee row ID.

Table 36. Value Match Output Property Sets

Output Property Set[1]	Property	Description	Comments
Generic Settings	End Time	The run end time.	Applicable only to SSA.
	Num Candidates	The total number of potential matches if scores are not used.	
	Num Results	The number of actual matches.	
	Row Value	The row ID of the match or matches found.	
	Start Time	The run start time.	

1. Match Info is a child property set of the output property set.

Invoked From

Any means by which you can invoke business service methods, such as with Siebel eScript.

Example

The following is an example of using Siebel eScript to invoke the Value Match method by way of a Siebel SmartScript. This script calls the Value Match method to look for duplicates of John Smith from the Contact business component and then returns matches, if any. After the script finishes, determine what you want to do with the duplicate records, that is, either merge or remove them.

```
function Script_Open ()

{

   TheApplication().TraceOff();
   TheApplication().TraceOn("sdq.log", "Allocation", "All");
   TheApplication().Trace("Start of Trace");

   // Create the Input property set and a placeholder for the Output property set
   var svcs;
   var sInput, sOutput, sAdapter, sMatchValues;
   var buscomp;

   svcs = TheApplication().GetService("DeDuplication");
   sInput = TheApplication().NewPropertySet();
   sOutput = TheApplication().NewPropertySet();
   sAdapter = TheApplication().NewPropertySet();
   sMatchValues = TheApplication().NewPropertySet();

   // Set Generic Settings input property parameters
   sInput.SetProperty("Refresh Keys", "N");
   sInput.SetProperty("BC Name", "Contact");
   sInput.SetProperty("Use Result Table", "N");
   sInput.SetType("Generic Settings");
```

```
    // Set Match Values child input property parameters
    sMatchValues.SetProperty("Last Name", "Smith");
    sMatchValues.SetProperty("First Name", "John");
    sMatchValues.SetType("Match Values");
    sInput.AddChild(sMatchValues);

    // Set Adapter Settings child input property parameters
    sAdapter.SetProperty("Key Level", "Standard");
    sAdapter.SetProperty("Population", "Default");
    sAdapter.SetType("Adapter Settings");
    sInput.AddChild(sAdapter);

    // Invoke the "Value Match" business service
    TheApplication().Trace("Property set created, ready to call Match method");
    svcs.InvokeMethod("Value Match", sInput, sOutput);

    // Get the Output property set and its values
    TheApplication().Trace("Value Match method invoked");
    var propName = "";
    var propVal = "";
    propName = sOutput.GetFirstProperty();
    while (propName != "")
    {

        propVal = sOutput.GetProperty(propName);
        TheApplication().Trace(propName);
        TheApplication().Trace(propVal);
        propName = sOutput.GetNextProperty()

    }

    TheApplication().Trace("End Of Trace");
    TheApplication().TraceOff();

}
```

See Also

For an example of how to invoke the Value Match method using Siebel Workflow, see "Scenario for Data Matching Using the Value Match Method" on page 91. For more information about Siebel Workflow, see *Siebel Business Process Designer Administration Guide*. For more information about business services and methods, see *Siebel Developer's Reference*.

Index

D

F

I

K

L

T

U

V

W