October 2005 Volume I Issue 1

ORAtips on CRM

Maintaining "Unique Customer Profile" for Oracle® CRM Implementations

By Anand Kanakagiri

Editor's Note: It is a fairly common business practice for organizations to have customer data in several systems. It is also fairly common practice to have duplicate data, often differentiated by something as simple as an extra comma or capital letter. Wouldn't it be great to be able to organize your customer data and have one definite source of information? While it would seem reasonable to assume that Oracle and other ERP systems would support this business model, Anand Kanakagiri, our CRM Editor, explains that, on the face of it, they do not without some help. In this article, Anand describes how, through networking with third-party vendors, he discovered an approach that not only works, but when implemented with Oracle CRM enhances its benefits. Anand shows you how to leverage existing Oracle functionality and how to use an operational data store architecture to identify, clean, and organize customer data. Anand provides detailed examples and useful screen shots to lead you through the process. Once you have followed this ORAtip, you will find that you have a great foundation for customer data management.

Introduction

Accurate identification and maintenance of unique customer profiles are critical to the success of Oracle CRM implementations. Duplicate customer profiles and inadequate distribution of accurate customer profiles to operational systems frequently result in the inability of organizations to optimally leverage their investments in CRM.

Accurate identification and maintenance of unique customer profiles are critical to the success of Oracle CRM implementations.

A customer data maintenance process should be a part of every CRM implementation. Most CRM packages do not include customer data management functionality. Siebel provides the add-on product Siebel Universal Customer Master that has data management functionality that can be leveraged over many Siebel installations.

This article explores an approach of application integration via customer unique profile management and customer data integration as a catalyst for realizing ROI in large enterprises' multimillion-dollar customer relationship management (CRM) installations. This is specifically applicable to Oracle CRM implementations because Oracle CRM does not provide a data management and customer data integration product. This approach enhances flexibility by enabling integration with Oracle ERP implementations, operational systems, and data warehouses.

The approach is a sample batch architecture for customer data quality and interface to an Oracle CRM system, using third-party data quality products from Trillium, First Logic, etc. All examples of data quality will be from Trillium. The architecture requires an external Operational Data Store (ODS). The data store is used to accumulate, cleanse, enhance, standardize, and remove duplicates from the customer data. This architecture will not support real-time cleansing of customer information.

Definition of the Customer

The definition of the customer depends on the scope of the CRM implementation. Most business analysts instinctively assume that they know who the customer is. However, most large and complex organizations can and do have many types of customers.

For example, customers of a health insurance company may be individual policyholders, the organization that is holding a group policy, and the hospital that is providing the service to the health insurance holder or the insurance broker. Similarly, the customers of a car manufacturing company may be defined as the final users of the car (you and me) or the car dealers. The classes and types of customers that the Oracle CRM implementation will serve should be planned and defined. The business leaders and all stakeholders should sign off on customer definition and scope. Clear definition of the customer and confirmed scope of CRM implementation is essential.

ORAtips on CRM

After the clear definition of the customer and confirmation of scope, Oracle CRM implementations need to plan for and design functionality to manage customer data over a life cycle that will include:

- Capturing customer data
- Cleansing data, correcting spelling and standardizing names, addresses, etc.
- Enhancing customer data through interfaces with other operational systems or external data
- Identifying duplicate customer information.
- Consolidating and creating organization-wide customer identification (ID)
- Processing and architecture to distribute customer ID and authorized information to operational systems and enterprise data warehouses to link customer profiles across multiple applications

Capturing Customer Data

To provide comprehensive and consolidated customer information, all possible sources need to be identified and leveraged. The following are some of the possible sources of customer information.

- The CRM system itself, where new customer and account profiles will be added
- Legacy source systems of customer information
- Operational systems
- External industry sources
- External data service providers like Dun & Bradstreet or IMS for medical data.

New customer information entered in the CRM system directly has not been cleansed, standardized, and checked for duplication. Such information needs to be extracted from the CRM system for scrubbing and re-applied to the CRM with required updates.

Data Cleansing and Standardization

Data received from the above source systems are not in a standardized format. For example, street addresses may be listed as "123 Main St" or "123 Main Street". There may be missing information or errors in the data. Data cleansing processes and standardization are required to correct these anomalies. Some data cleansing and standardization functionality can be custom coded. However, third-party data quality tools provide excellent functionality for data standardization, name and address enhancement, validation, and ability to look up multiple sources for processing.

These tools use the data provided from the United States Postal Service and leverage it against the available address information to provide standardized name and addresses. Use of an appropriate third-party data quality product is highly recommended.

Trillium is one such external tool with a module named "Trillium Parser" that provides functionality to:

- Recognize and standardize personal and business names
- Recognize and standardize address and geography (including foreign addresses)
- Process relationship information

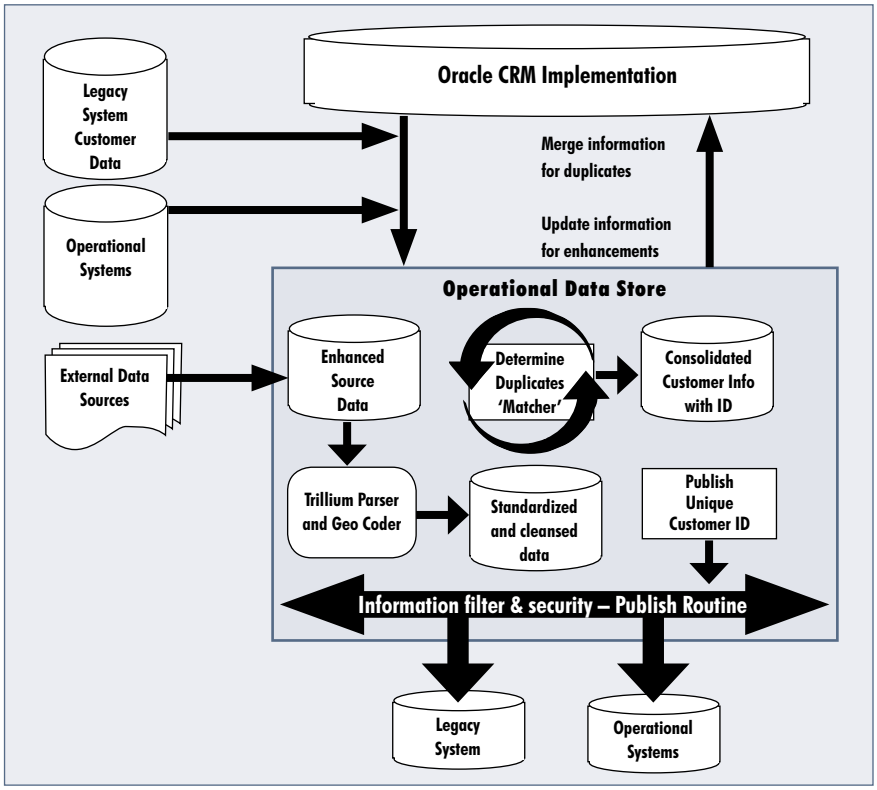

Figure 1 – Operation Data Store

October 2005 Volume I Issue 1

ORAtips Journal

These features can be utilized to clean and standardize critical customer information. Examples of name standardization are shown in the tables below:

Original Business Name	Standardized Name (after Trillium Processing)
ABC Forge Co.	ABC Forge Company
Abc Forge	ABC Forge Company
A B C Forge Company	ABC Forge Company.

Table 1 – Sample Business Names

Original Name	Standardized Name (after Trillium Processing)
Cathrine B St James	Cathrine B St James
St James, Catherine B	Cathrine B St James
CATHERINE B ST JAMES	Cathrine B St James

Table 2 – Sample Personal Names

Enhancing Customer Data

Building interfaces to internal data repositories like Oracle ERP, operational systems, or data warehouses can enhance customer data. Several external data service providers provide industry customer information. For example, IMS Health provides medical information and D&B provides financial information. All available sources should be accessed to enhance the customer data to make it as comprehensive as possible.

Identify Duplicate Customer Information

Most third-party tools provide advanced matching logic architecture that allows complex matching logic to be implemented easily within Oracle. These tools allow implementation of logic that can match based on less than 100% match.

You can easily implement logic such as:

'If the Name matches > 98% and Street Address Matches >91%'

These two records are related to the same customer.

Trillium uses "Matcher Routines" to match fields from two separate customers and assigns a score to the match. The better the match, the higher the score, with a perfect match having a score of 100. Any mismatch results in a reduction of the score.

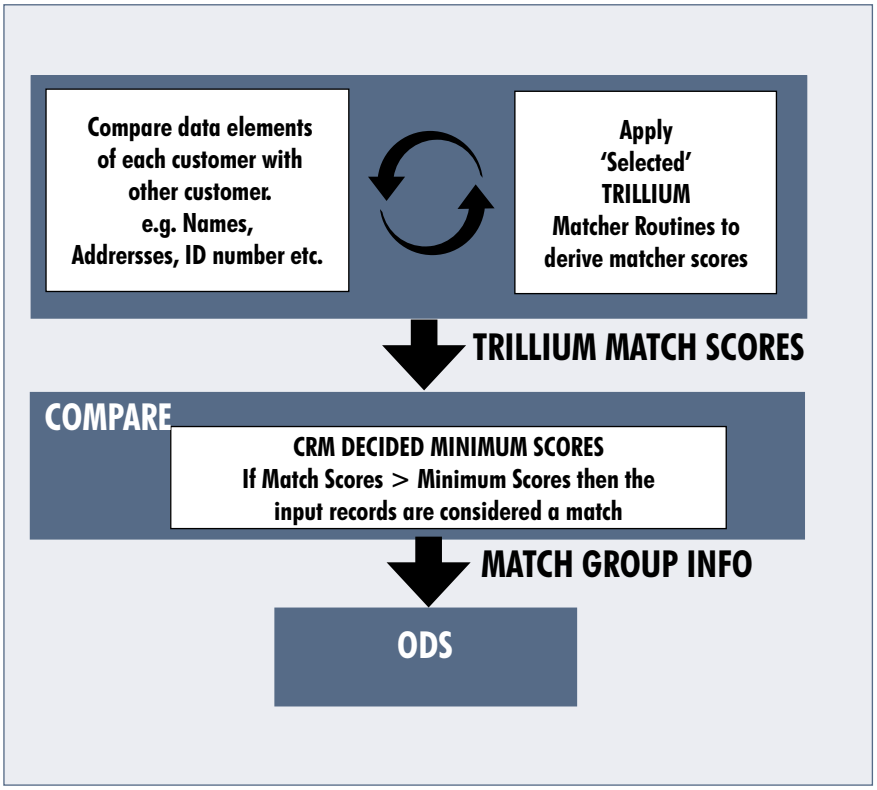

Figure 2 – The Matcher Process

The Oracle Application designers need to decide the scores that are required for two inputs to be considered a match. For example, the implementers can decide that if two addresses have a match score of more than 95 (out of 100) then the application should consider the two input addresses to be the same.

Using an example, let us walk through a case where we are trying to match customer records for retail customers. The customer data contains the following data columns that are used to match the records.

In the next example, we need to determine if record 1 is a match with record 2 and if record 3 is a match with record 4. To achieve this, we

use the pre-defined "Matcher" routines provided within the third-party tool Trillium. These routines apply to specific kinds of fields and a score is assigned based on a specific logic to compared values. Please refer to Table 4 for the sample matcher routines used for specific fields.

Sl#	First Name	Last name	Prefix	Generation	House Number	Street Code	City/State	Zip
1	JoeAnne	Baker	Ms	NULL	506	Bedford Street	Stamford, CT	06901
2	JoeAnn	Baker	Null	NULL	NULL	Bedford Street	Stamford, CT	06901
3	William	McIntyre	Mr.	Jr.	1311A	Main Street	Houston, TX	77010
4	Bill	McIntyre	Mr.	NULL	1311	NULL	Houston, TX	77010

Table 3 – Sample Customer Records

Field Name	Trillium Matcher Routine Used	Description	Scoring Logic (in brief)
First Name, Last Name and Non Blank Middle Name	FRSTNAME	This routine was designed to compare two first names. It assumes the field are left justified and blank filled. If the two fields do not match exactly, the routine will calculate two field lengths by excluding trailing blanks.	Scoring for FRSTNAME: 100: exact match (excluding blank versus blank) 99: blank field value versus blank field value 98: blank field value versus anything 95: shorter field value is one character in length and it equials the first character of the longer field 90: if shorter field value is greater than one character length and it is an exact starting substring of the longer field 80: if there is a mismatch of the first characters of both fields
Prefix	PREFIX	This routine is used for Prefix Column. This routine was designed to compare two name prefix values. Fields are assumed to be 4 characters in length.	Scoring for Prefix: 80: for MRS versus MISS 95: for MISS versus MS 90: for MISS versus blank field value 85: for MISS versus anything except MS/blank field values 90: Or MS versus blank field value 85: for MS versus anything except MRS/blank field values 95: for blank versus anything except MR/MSTR MRS/MISS/MS field values 0: for any other value
Generation	GENER	This routine was designed to compare Generation values. Fields are assumed to be two characters in length.	Scoring for GENER 100: exact match (including blanks versus blanks) 95: blank value versus a "SR" value or "JR" value 90: blank value versus a nonblank non "SR" value 0: any other exception
House Number	HOUSENO	This routine was designed to compare two house numbers. It assumes the fields are right justified. If the two fields do not match exactly, the routine will calculate two field lengths and starting positions by excluding leading blanks.	Scoring for House Number: 100: exact match (excluding blank versus blank) 99: blank field value versus non-blank 99: all zeroes field value versus anything 98: blank field value versus blank field value 98: all zeros field value versus all non-zeros field value.

Table 4 – Sample Matcher Routines – continued on next page

			If the fields do not match according to the above criteria, then perform the following to determine if there is a match: 98: Step 1) Check the numbers to see if either of the numbers is of the form N-N. If both numbers are of the form N-N, then there is no exact match; continue with Step 2) If one of the numbers is of the form N-N, then check if the other number is in the range N-N. If it is, then return a score of 98. For example, 103 is in the range of 101-105. 100: Remove all leading blanks and characters that are not digits or letters from both strings and left justify them. Return 100 if there is an exact match.
Street	STREET	This routine was designed to compare street names. Common street abbreviated words are then normalized and the spelling algorithm is applied. If the spelling yields below an 80, a modified sound algorithm is applied. If the score is still below 80 for numeral streets, a word comparison routine is applied. If the two fields do not match exactly, the routine will calculate two field lengths by excluding trailing blanks	Scoring for Streets: 100: for an exact match 90-99: varying degrees of acceptable differences 95: for neither field value blank nor one field an exact starting substring (6 characters or more in length) of the other, but the difference in length is not great than the two characters. 88: for an exact match of blank field value versus blank 80: for blank field versus nonblank field
City, State and Zip Code	ABSOLUTE	This routine was designed to look for an exact match between two fields.	Scoring for Absolute: 100: exact match, including blank versus blank 0: any exception

Table 4 – Sample Matcher Routines

A series of business rules can be defined and applied to the scores to determine if the input records match or not. Some of the definitions can be as follows:

In the rules below, perfect match on city, state and Zip code is assumed.

Rule 1 If the score is >94 on first name, >94 on Last Name, 100 on House No and >95 for Match on Street then there is a match.

Rule 2 If the score is 100 on first name, 100 on Last Name, >95 on House No and >95 for Match on Street then there is a match.

Rule 3 If the score is >=90 on first name, 100 on Last Name, >95 on House No and 100 for Match on Street then there is a match.

Consider the following example:

In the above example, these two rows are a match as they pass rule 3. The data present in the two rows are to be considered belonging to the same customer.

Rows	First Name	Last Name	House Number	Street
Row 1 Data	JoeAnne	Baker	506	Bedford Street
Row 2 Data	JoeAnn	Baker	NULL	Bedford Street
Matcher Routine	FRSTNAME	FRSTNAME	HOUSENO	STREET
Trillium Assigned Scores	90	100	99	100
Reason for score	The shorter field value is greater than one character in length and it is an exact starting substring of the longer field	Exact Match	Null versus non null comparison	Exact Match

Table 5 – Sample Customer Data With Scores That Match

Consider our second example:

In this case, even though the data looks similar, it does not satisfy any of our rules. So the data in these two rows are considered to be belonging to two different customers.

Rows	First Name	Last Name	House Number	Street
Row 3 Data	William	McIntyre	1311A	Main Street
Row 4 Data	Bill	McIntyre	1311	NULL
Matcher Routine	FRSTNAME	FRSTNAME	HOUSENO	STREET
Trillium Assigned Scores	0	100	100	80
Reason for score	No Match	Exact Match	Remove all leading blanks and characters that are not digits or letters from both strings and left justify them. Return 100 if there is an exact match.	Exact Match

Table 6 – Sample Customer Data With Scores That Do Not Match

Customer Information Consolidation and Creation of Organization Wide Customer ID

All available data in the duplicate customer needs to be consolidated to extract all required information and create a unique survivor record that contains the "best" available information from all the records. An identification number can be assigned to this customer information.

All efforts should be made to publish this customer ID in all operational systems and data warehouses to facilitate sharing of data. Data warehouses can use this ID to link information from different source systems resulting in a complete view of the customer.

Implementing an external operational data store that maintains customer data through the life cycle can add value to the Oracle CRM implementation. It provides a design to reduce and eliminate duplicate customer data and provide unique customer IDs to subscribing operational systems.

Unique customer IDs allow the organization to link customer information together. A well-designed and integrated operational data store will give a solid foundation for integrated customer information.

Sources

Trillium Documentation and User Manuals.

Anand Kanakagiri – Anand is Assistant Vice President, IT Department, with a major reinsurance corporation. He has over 13 years of experience in Oracle Databases, focusing on data management (Data warehousing, Database administration, Performance tuning, ETL, etc.). He has written several white papers internally within the company and for several fortune 500 clients on Oracle data management. Anand may be contacted at Anand.Kanakagiri@ERPtips.com.

October 2005 Volume I Issue 1

ORAtips Journal