

Building One Version of the Truth – the Oracle Warehouse Builder's way  
Senthil Nathan Natarajan  
Technical Consultant, Asparona Limited

---

## Introduction

Data Warehouse provides a wealth of data for analytical reporting. Data quality is a vital factor to gain business value from data warehouse. Diversified business operations, multiple divisions and mergers of companies can make room for many versions of the truth. This paper outlines Oracle's 'truth' using Oracle Warehouse Builder to perform data cleansing by matching and merging data. We can leverage the utilities of Oracle Warehouse Builder and also add our customization to define the business rules for matching and merging. A business case with a merger of two companies is outlined to showcase the business benefits of data consolidation, considering examples of customer and debtor data consolidation.

## Data Quality Importance

Surveys conducted by The Data Warehousing Institute (TWDI) have shown that poor data quality costs U.S. businesses an estimated \$600 billion a year [1].

Data stewardship programs must exist at an enterprise level. Individual divisions or companies within a corporate having different data quality standards result in multiple versions of the truth. Customer interacts with the organization perceiving it as one unit, but gets different stories from the different divisions of the company when enterprise data quality is not in place. As a consequence customer loyalty is at stake, shaking the very purpose of the business.

Data profiling is required to characterize the content, quality and structure of the data. Customer profiling for example can result in better recognition of valued customers and identify bad customers. Savings in costs such as mailing and better debt recovery are clearly possible.

## Case Study Scenario

In today's world merger of companies is frequent. Once the merger is officially announced, blending the operations of the businesses is the challenge.

We illustrate here the merger of two fictitious companies Rogers Limited and Taylors Limited to one company called RT Limited. Rogers Limited has 6,500 customers and Taylors Limited has 4,500 customers. There are common customers as well; the exact number not sure at the time of merger. Both these companies have outstanding payments to receive.

Customers are to be treated as customers of RT Limited after the merger. This means that in terms of loyalty, debt recovery, mailing and other communications we wish to integrate the data and avoid any redundancies. The win-win is better service for the customer and cost savings for the business.

## Oracle Warehouse Builder

Oracle Warehouse Builder (OWB) is a CASE tool to design, develop and implement an ETL. Data can be sourced from various sources (Oracle / non-Oracle / flat files / XML) and mapped into a data warehouse. The 'mapping editor' within the OWB is the visual editor to design the data flow. Various mapping operators from the toolbox assist in joining, splitting, aggregating, expression building, key lookup and similar functions of transforming the data. Some special data cleansing operations to highlight are de-duplication, name & address cleansing and math-merge functionalities.

## Match Merge Functionality

The match-merge operator is a data quality operator that can first match the record and then merge them. The input to the match-merge operator is one which has records which are potentially business duplicates and needs to be consolidated. The output from the match-merge operator is a data set which is consolidated.

### Matching

The input data set can be divided into matching bins. Matching records are searched for within the matching bin rather than the entire data set. Based on the business data we can choose attribute(s) to define the data in the same matching bin.

Matching rules can be defined in various ways. Single or multiple match rules can be defined. When multiple rules are defined, if any one rule being satisfied then the records are considered to be a match. Each match rule is defined against attribute(s). Some of the match rule types are:

- Address
- Conditional
- Firm
- Custom

Conditional matching has pre-defined algorithms such as exact and standardized exact (case insensitive).

### Merging

In the event of records being a match then have to merged and given as one output. Merge rules are defined against single or multiple attributes. Some of the merge rule types are:

- Any
- Any Record
- Custom
- Min Max

Oracle Warehouse Builder assigns a position number and creates a default name for each merge rule. The merge rules are executed in the order of their position number.

More details of the matching from the Oracle Warehouse Builder User Guide / Transformation Guide.

## Savings from Data Quality

The companies are merging to widen the customer base and cross sell the products. We shall illustrate here how the match merge is done and results in savings.

Criteria for matching are: 'Name', 'Street Address' and 'Postcode'. The matching bins are formed by 'Postcode'. The entire data set of 110,000 records is scanned and 15,000 matches are identified. As a consequence there is now an actual customer base of 95,000 customers.

The merging criteria are taken as the first occurrence of the attribute value for all fields, except the 'credit limit', 'amount overdue' and 'loyalty points' amount attributes. The pre-defined merging rule 'Any Record' can be set for the multiple attributes.

For the 'credit limit', the merging criterion is to set the higher credit limit value by coding a PL/SQL and assigning it as a custom rule for this attribute.

For the attributes 'amount overdue' and 'loyalty points' the merge criteria is set to a custom rule to accumulate the values from the merging records.

The business advantage here is that the unified customer record can show the real value of the customer by accumulating the 'loyalty points'. Targeted marketing becomes more effective. Customer is happy to obtain more privileges due to merger of the companies.

The bigger business advantage is the recovery of overdue amount. The merger decides that any overdue amount less than \$800 is written off. Without consolidation many customers may be well below the limit in the individual companies of Rogers Limited and Taylors Limited and written off on both sides. The common customers where the merge identified is 15,000. The details of how the savings are achieved are shown below:

	Rogers Limited	Taylors Limited
Customers	65,000	45,000
Overdue customers	3,000	3,200
Overdue customers below \$800	1,800	950
Consolidated overdue amount, where the individual overdue amount is below \$800 (average \$600 overdue)	1,080,000	570,000

Scenario for the merged RT Limited – after match merge of customer information	
Common customers having overdue less than \$800	700
Potential recovery due to merging of customer information for 700 customers	840,000
Total amount potentially written off if customer information was not merged	1,650,000

## Conclusion

Data Quality prevents losses of dollars and reputation for the business. In our paper we have seen how a business can potentially save \$840,000 due to customer data consolidation. It was one version of the truth in the overdue amount which brought about this savings.

Customers appreciate and stay loyal when the business is accurate about their engagement with the business. Happy customers improve employees satisfaction is improved, consequently reducing staff turnover.

Needless to say quality data results in higher productivity and quality decisions.

## References

[1] Data Quality and Bottom Line: Achieving Business Success through Commitment and High Quality Data by Meighan Berberich, The Data Warehousing Institute