

Managing Data Integrity with Oracle Data Integrator

*An Oracle White Paper
Updated December 2006*

Managing Data Integrity with Oracle Data Integrator

Oracle Data Integrator ensures that faulty data is automatically detected and recycled before being inserted into the target application. This is done without programming by following the integrity rules and constraints defined on the target application and in Oracle Data Integrator.

EXECUTIVE OVERVIEW

Many companies underestimate the impact of poor-quality data on their overall business operations. Data integrity concerns become even more critical when new applications are integrated into the information technology infrastructure. Unless a company has implemented a systematic and reliable approach to data integrity, poor-quality data can propagate across the information technology infrastructure and degrades the integrity of data in other applications. This white paper describes how to manage data integrity using Oracle Data Integrator.

INTRODUCTION

An integration process involves the source applications, the processes that transform and integrate data, and the target applications. With Oracle Data Integrator, users can and should manage data integrity in all these subsystems.

By using Oracle Data Integrator Metadata Navigator, the navigational module, developers can understand the information stored by various applications in the information system. This module provides an intuitive, easy-to-use Web interface to reveal the underlying data structures to developers and business users. They can then use this knowledge when they design the declarative rules, data checks, and audit processes essential for maintaining a high level of data integrity.

Data quality in source and target applications is usually managed within the application, at either the graphical user interface (GUI) level or the database level, using database management system constraints. External events or inadequate data-integrity checking, however, can cause these applications to produce inconsistent data. This data can be audited in Oracle Data Integrator using static checks that isolate erroneous information from stored data.

Similarly, controlling data integrity as part of the data integration process ensures that potentially poor-quality data is not propagated across the information system. This kind of data integrity operation is performed through a flow check that isolates erroneous information during an integration process.

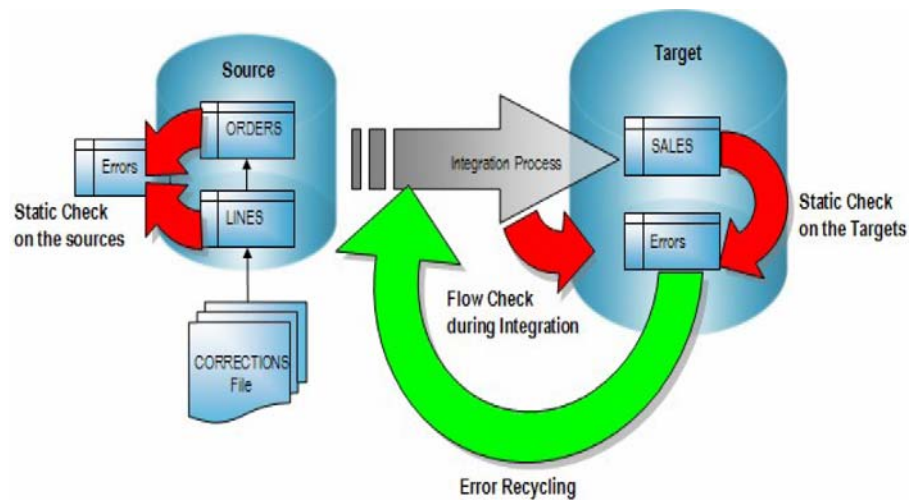


Figure 1: Oracle Data Integrator uses static checks and flow checks to isolate erroneous information.

Ensuring Overall Consistency in Enterprise Information

Oracle Data Integrator allows application designers and business analysts to define declarative rules for data integrity in the centralized Oracle Data Integrator Metadata Repository. These rules are applied to application data to guarantee the overall integrity and consistency of enterprise information.

Defining Business Rules

Oracle Data Integrator can automatically retrieve existing rules that have been defined at the data level (such as database constraints) using a customizable reverse-engineering process. Developers can also create new declarative rules using the Oracle Data Integrator Designer graphical user interface—without the need for coding. These rules can be inferred from data discovery and profiling within Oracle Data Integrator. By using a synchronous check, developers can immediately test the new declarative rules against the data.

Types of Rules

Rules for data integrity can include

- Uniqueness rules, such as:
 - “Different customers must not have the same e-mail address”
 - “Different products must have different product and family codes”
- Simple and complex reference rules, such as:
 - “All customers must have a sales representative”
 - “Orders must not be linked to customers marked as “Invalid””

- Validation rules that enforce consistency at the record level, such as:
 - “Customers must not have an empty zip code”
 - “Web contacts must have a valid e-mail address”

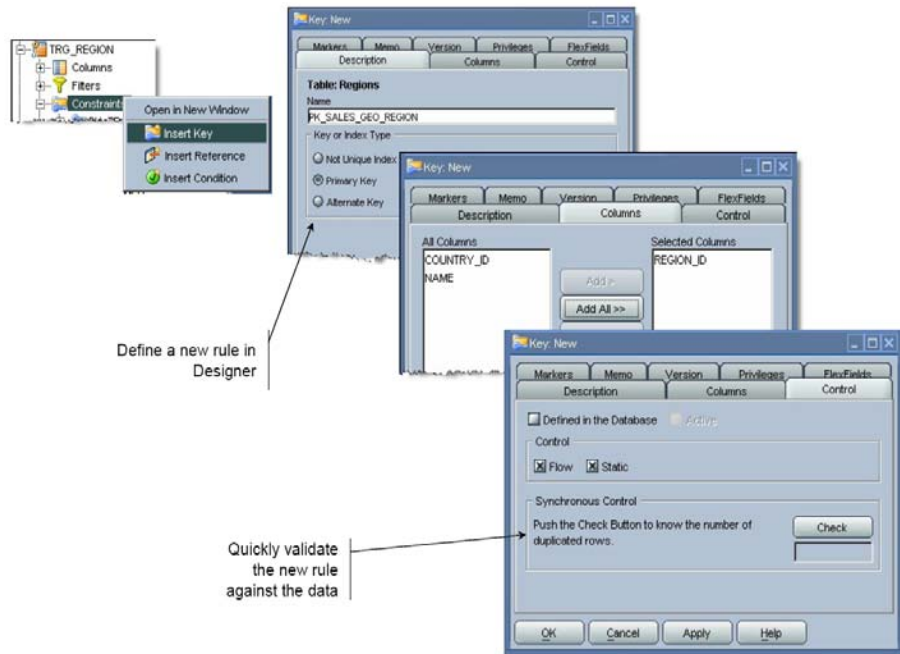


Figure 2: Oracle Data Integrator Designer can immediately check a newly defined declarative rule against the data.

Enforcing Business Rules

Oracle Data Integrator’s customizable Check Knowledge Modules help developers automatically create data integrity audits of their applications based on declarative rules that have been defined in the Oracle Data Integrator Metadata Repository. These Check Knowledge Modules generate the code necessary for the static or flow data checks and the error recycling performed as part of the integration process.

Audits provide statistics on the integrity of application data. They also isolate data that is detected as erroneous by applying the business rules. Once erroneous records have been identified and isolated in error tables, they can be accessed from Oracle Data Integrator Designer or any other front-end application.

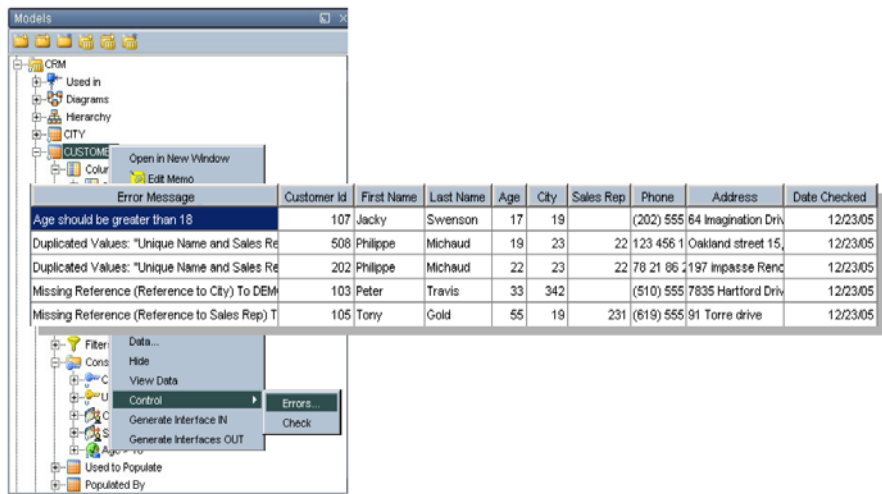


Figure 3: With Oracle Data Integrator Designer's graphical user interface, users can easily view erroneous data.

This extensive audit information on data integrity makes it possible to perform a detailed analysis, so that erroneous data can be handled according to information technology strategies and best practices. For example, erroneous data can be handled in any of the following ways.

- **Automatic correction**—Oracle Data Integrator offers a set of tools to simplify the creation of data cleansing interfaces that can be scheduled to run at predetermined intervals.
- **Accept erroneous data (for the current project)**—In this case, interface developers will need precise rules for filtering out erroneous data later, using Oracle Data Integrator filters.
- **Correct the invalid records**—In this case, the invalid data is sent to application end users through various text formats or distribution modes, such as e-mail, HTML, XML, flat text files, and so on, using Oracle Data Integrator packages.
- **Recycle data**—Erroneous data from an audit can be recycled into the integration process.

All these strategies can be automated by using Oracle Data Integrator interfaces and packages to put data integrity at the heart of integration processes.

Performing Complex Data Cleansing Operations

Oracle Data Integrator's open data-integrity architecture is based on customizable Check Knowledge Modules. As a result, the system can be seamlessly integrated with third-party products to perform complex data-quality and cleansing operations.

For example, the Oracle Data Integrator might use Trillium TS Quality features to run name and address cleansing processes such as the following.

Unless a systematic, reliable approach to data integrity is implemented, poor-quality data can propagate across the information technology infrastructure and can degrade the integrity of data in other applications.

- **Standardization and cleansing**—Contact and address information are cleansed and standardized based on the country of origin or according to corporate standards.
- **Address validation and enrichment**—Addresses are validated against national postal authorities' files and possibly corrected. Additional geographic or third-party information (latitude and longitude, or market targeting information, for example) can also be added to the address information.
- **De-duplication**—Duplicate records are identified and merged into a unique "best" record, and then all duplicates are linked to it. Consolidation can then be performed with Oracle Data Integrator to create a unified view with no loss of original data.

Oracle Data Integrator can use TS Quality's data standardization, enrichment, and de-duplicating features in real time (row by row) or in batch mode.

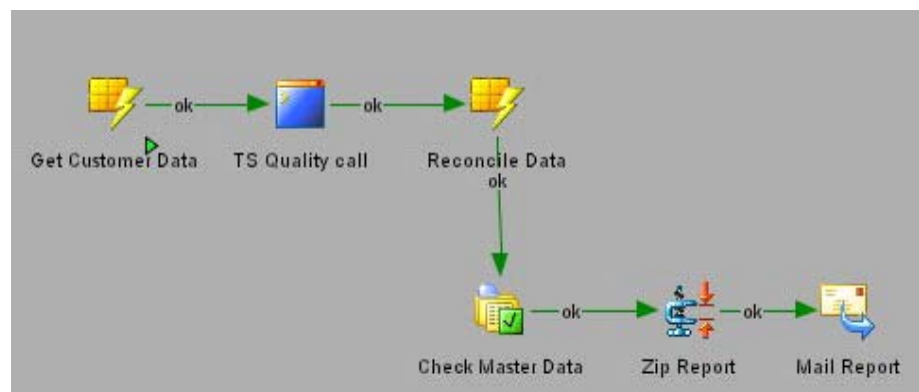


Figure 4: With Oracle Data Integrator, complex name and address cleansing procedures can be automated using Trillium TS Quality.

CONCLUSION

Data integrity is an essential part of the integration process. Although common approaches tend to separate data movement from data integrity control, Oracle Data Integrator's integrated approach ensures that data is adequately verified, validated, and cleansed at every stage of the integration process.



Managing Data Integrity with Oracle Data Integrator
Updated December 2006

Oracle Corporation
World Headquarters
500 Oracle Parkway
Redwood Shores, CA 94065
U.S.A.

Worldwide Inquiries:
Phone: +1.650.506.7000
Fax: +1.650.506.7200
oracle.com

Copyright © 2007, Oracle. All rights reserved.

This document is provided for information purposes only and the contents hereof are subject to change without notice.

This document is not warranted to be error-free, nor subject to any other warranties or conditions, whether expressed orally or implied in law, including implied warranties and conditions of merchantability or fitness for a particular purpose. We specifically disclaim any liability with respect to this document and no contractual obligations are formed either directly or indirectly by this document. This document may not be reproduced or transmitted in any form or by any means, electronic or mechanical, for any purpose, without our prior written permission. Oracle, JD Edwards, PeopleSoft, and Siebel are registered trademarks of Oracle Corporation and/or its affiliates. Other names may be trademarks of their respective owners.