An Oracle White Paper
February 2009

# Real-time Data Warehousing with ODI-EE Changed Data Capture

## Executive Overview

Today’s integration project teams face the daunting challenge of deploying integrations that fully meet functional, performance, and quality specifications on time and within budget. Further, these processes must be maintainable over time, and the completed work should be reusable for further, broader integration initiatives.

Traditional “Extract, Transform, Load” (ETL) tools closely intermix data transformation rules with the integration process procedures, requiring the development of both data transformations and data flow. ODI-EE takes a different approach to integration by clearly separating the declarative rules (the “what”) from the actual implementation (the “how”). With ODI-EE, declarative rules describing mappings and transformations are defined graphically, through a drag-and-drop interface, and stored independently from the implementation. ODI-EE automatically generates the data flow, which can be fine-tuned if required. This declarative design approach dramatically reduces the development and maintenance workload—by up to five times—over traditional ETL tools.

This technical brief describes ODI-EE’s use of Changed Data Capture to track changes in source data—changes that are typically performed by other applications. When running integration interfaces, ODI-EE can reduce the volume of source data processed in the flow by extracting only the changed data. Then, only those changes are moved to target systems, tremendously improving the efficiency of data integration processes.

Reducing the volume of source data is useful in many fields, such as data synchronization and replication. And it is essential when setting up an event-driven architecture for integration.

## Introduction

The conventional approach to data integration involves extracting all data from the source system and then integrating the entire set—possibly using an incremental strategy—in the target system. This approach, which is suitable in most cases, can be inefficient when the integration process requires real-time data integration. In such situations, the amount of data involved makes data integration impossible in the given timeframes.

Reducing the volume of the source data to only changed data is a major issue in many fields, such as data synchronization and replication. It is an absolute requirement when implementing event-driven integration processes.

Basic solutions, such as filtering records according to a timestamp column or “changed” flag, are possible, but they might require modifications in the applications. In addition, they usually do not sufficiently ensure that all changes are taken into account.

ODI-EE’s Changed Data Capture identifies and captures data as it is being inserted, updated, or deleted from datastores, and it makes the changed data available for integration processes.

Today Oracle includes Oracle Data Integrator and Oracle Warehouse Builder Enterprise ETL, formerly an add-on option to Oracle Database, as the two components of ODI-EE. Going forward, these products will merge into a single unified data integration technology platform. This strategy fully preserves any existing development investments of all Oracle data integration customers and will provide a seamless, easy upgrade path from the current components to the unified platform. Customers can safely choose either component as the basis for implementations today.  For further advice on the implementation choices, please contact your Oracle sales representative.

## Publish-and-Subscribe Model

Changed Data Capture uses a publish-and-subscribe model. This model works in three steps:

1. An identified subscriber, usually an integration process, subscribes to changes that might occur in a datastore. Multiple subscribers can subscribe to these changes.
2. The Changed Data Capture framework captures changes in the datastore and then publishes them for the subscriber.
3. The subscriber—an integration process—can process the tracked changes at any time and consume these events. Once consumed, events are no longer available for this subscriber.

ODI-EE processes datastore changes in two ways:

- **Regularly in batches (pull mode)**—for example, processes new orders from the Web site every five minutes and loads them into the operational datastore (ODS)

- **In real time (push mode) as the changes occur—**for example, when a product is changed in the enterprise resource planning (ERP) system, immediately updates the on-line catalog

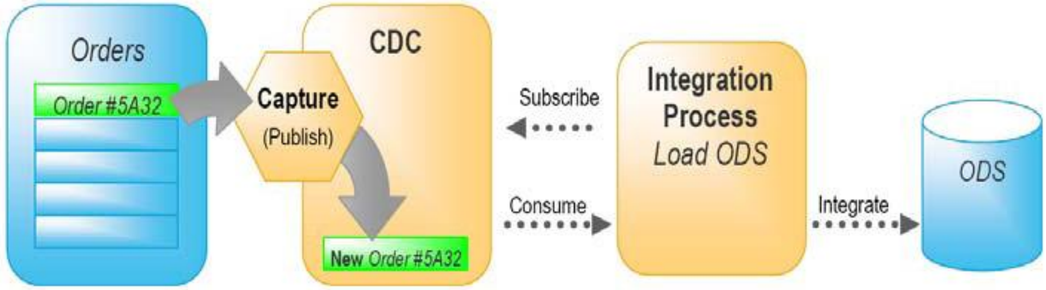

**Figure 1: The Changed Data Capture (CDC) framework uses a publish-and-subscribe architecture.**

# IMPLEMENTING CHANGED DATA CAPTURE

## Methods for Tracking Changes

ODI-EE provides two methods for tracking changes from source datastores to the Changed Data Capture framework: triggers and relational database management system (RDBMS) log mining.

The triggers method creates triggers on the source tables to track changes as data is inserted, updated, or deleted. This method can be implemented on most RDBMS, but it can have an impact on the transactional performance of the source systems.

The second method involves mining the RDBMS logs, which are the internal change history of the database engine. This method has no effect on the system's transactional performance; it is database-specific. This method is supported out-of-the-box for Oracle Database (through the Log Miner feature) and IBM DB2. The Changed Data Capture framework used to manage changes is generic and open. The change tracking method can be customized, and any third-party change provider can be used to load the framework with changes.

## Processing the Changes

ODI-EE employs a powerful declarative design approach, Extract-Load, Transform (E-LT), which separates the rules from the implementation details. Its out-of-the-box integration interfaces use and process the tracked changes.

Developers define the declarative rules for the captured changes within the integration processes in the ODI-EE Designer graphical user interface—without having to code. With the ODI-EE

Designer, customers declaratively specify set-based maps between sources and targets, and then the system automatically generates the data flow from the set-based maps.

The technical processes required for processing the changes captured are implemented in ODI-EE's Knowledge Modules. Knowledge Modules are scripted modules that contain database and application-specific patterns. The runtime then interprets these modules and optimizes the instructions for targets.

## Ensuring Data Consistency

Changes frequently involve several datastores at one time. For example, when an order is created, updated, or deleted, it involves both the orders table and the order lines table. When processing a new order line, the new order to which this line is related must be taken into account.

ODI-EE provides a mode of tracking changes, called Consistent Set Changed Data Capture, for this purpose. This mode allows you to process sets of changes that guarantee data consistency.

# Conclusion

Integrating data and applications throughout the enterprise, and presenting a unified view of them, is a complex proposition. Not only are there broad disparities in data structures and application functionality, but there are also fundamental differences in integration architectures. Some integration needs are data oriented, especially those involving large data volumes. Other integration projects lend themselves to an event-oriented architecture for asynchronous or synchronous integration.

Changes tracked by Changed Data Capture constitute data events. The ability to track these events and process them regularly in batches or in real time is key to the success of an event-driven integration architecture. ODI-EE provides rapid implementation and maintenance for all types of integration projects.

Improve Data Integration with Changed Data Capture
February 2009

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

Oracle is committed to developing practices and products that help protect the environment