

Oracle Data Integrator: An Executive Overview

*An Oracle White Paper
Updated January 2007*

Oracle Data Integrator: An Executive Overview

EXECUTIVE OVERVIEW

Data is one of your company's most important assets, and data integration constitutes the backbone of your enterprise's IT systems. Choosing the wrong technology for data integration can cause harmful, long-lasting effects that impact not only the IT budget, but also the productivity and responsiveness of critical business divisions within your enterprise.

Oracle Data Integrator, a key component of Oracle Fusion Middleware, provides a strong and reliable integration platform for your IT infrastructure. There are five key reasons why choosing Oracle Data Integrator might end up being one of the smartest decisions you make for your company.

- **Performance**—Market-leading, unique architectural approach
- **Flexibility**—One platform, several data integration styles
- **Productivity**—Declarative design tool that drives reusability
- **Modularity**—Hot-pluggable knowledge modules that support diversity
- **Vision**—Performance-driven, service-oriented architecture—pluggable bulk data integration

"Oracle Data Integrator is helping us turn
our data into gold."

—Jack Garzella,
Vice President, Data Warehousing,
Overstock.com

INTRODUCTION

CIOs and IT leaders are responsible for a number of projects that rely on fast, accurate, and flexible movement of data throughout the IT infrastructure. Data integration is the technology—or, more-accurately, the set of technologies—that perform this movement in a variety of architectural approaches, depending on the individual requirements.

A recent survey of top IT priorities for 2007 cites a host of mission-critical projects, all of which demand a substantial data integration solution component:

- Business process improvement projects
- Business intelligence projects
- Information governance initiatives

- Security and risk management programs
- Service-oriented architecture (SOA) rollouts

Choosing a typical ETL (extract, transform, and load), or even a more-federated EII (Enterprise Information Integration) tool to solve these data integration challenges might lead you down a dead end.

Too often, it is the simple and easy path to keep buying the same old IBM DataStage and Informatica tools, which are the most costly and harmful.

Without evaluating newer options, such as Oracle Data Integrator, you might never know that you can achieve higher performance at a lower cost while also becoming better prepared for an open, standards-driven SOA future.

With so many years invested in the older data integration technology base, it is easy to overlook some of the nagging and persistent problems that it created. These problematic issues include the following:

- The “hub-and-spoke” architecture, still used by many data integration environments, is costly to implement and often becomes a bottleneck that slows down the overall IT systems.
- Many data integration projects are still treated as independent projects, with no overall consistency, resulting in silos of integration. This is exacerbated by most integration platforms’ lack of support for multiple typologies of data latency and volume; for example, few systems support high-volume data batches and trickle-feed real-time data flows within the same platform.
- With top-down procedural design being the norm for many integration platforms, development and maintenance becomes very costly; this lack of flexibility can considerably hinder the time to market for a complete solution.
- Tight coupling of data integration processes and operational systems results in a lack of flexibility; updates are much more costly and cannot be performed with the required agility of today’s fast-paced business environment.

Unlike the legacy data integration solutions from IBM and Informatica, Oracle Data Integrator allows organizations to reduce the cost and complexity of their data integration initiatives, while improving the speed and accuracy of data transformation. Oracle Data Integrator supports all integration needs: for data warehousing, business intelligence, and master data management; for real time; and for SOA data services deployments.

THE THREE “RIGHTS”

Getting the right data in the right place at the right time is the key to a successful data integration strategy.

- **The Right Data:** The data must not only be appropriate for the use that is intended, but must also be accurate and reliable.

“The first thing that struck us was the speed with which we ramped up our ETL developments with Oracle Data Integrator. The implementation and learning curve were extremely quick and, in a few days, we were fully operational with the solution.”

**—Leon Shaigorodsky,
Project Manager, iBasis**

- **The Right Place:** The overall information ecosystem consists of multiple operational and analytical systems, and they all need to benefit from data of the other systems, regardless of their locale.
- **The Right Time:** Data can become stale quickly. A decision support system that does not get the data in time is useless. A shipping application that does not get order information before cutoff time isn't efficient. Getting data in right time—with a latency that is appropriate for the intended use of this data—is one of the most important challenges faced by businesses today.

An Innovative Approach to High-Performance Integration

Oracle Data Integrator provides all the elements of the solution needed to provide these three “rights.”

Its ability to connect to any source or target system—databases, files, packaged applications, and so on—ensures that no island of data will remain isolated from the data integration processes. And its built-in data integrity features guarantee that not only the data processed is accurate, but also it is consistent and compliant.

The right time is about the performance of the integration processes. Oracle Data Integrator is the only data integration platform on the market to natively implement the high-performance ELT architecture. Instead of moving all the data through an intermediate ETL transformation server, the ELT approach leverages the power of the target database engines to perform the transformations, dramatically improving the performance at a much-lower total cost of ownership. Data goes straight from sources to targets, and is transformed many times faster than other tools.

Unlike the “pushdown optimization” feature introduced by Informatica and IBM WebSphere, Oracle’s native ELT architecture is not a mere afterthought. Most of these so-called pushdown optimization transformations still occur inside the ETL engines, and require the physical data to transit over the network and through their engines anyway. Although pushdown optimization may indeed improve the performance of Informatica or IBM ETL processes in some cases, the Oracle architecture is in a performance class of its own.

Other key features that address the right time business need include Oracle Data Integrator’s advanced capabilities, such as Changed Data Capture that transfers only updated records, and the ability to seamlessly change the latency of all integration processes—from batch to near-real time to real time.

ONE PLATFORM THAT EVOLVES ON DEMAND

Data integration is not just batch ETL. It isn't only SOA data services either. Nor is it only event-driven real-time data flow. Data integration encompasses all of the above—and many more combinations of use cases. Oftentimes, data integration projects start with one specific need; for example, populate the data warehouse, or exchange data in real time between two operational systems. But then integration needs grow in new directions. Regardless of the scope of an initial project, choosing

“We needed a data integration tool that would reduce our dependency on manual coding of ETL scripts and leverage the power of our teradata warehouse for data transformation. With Oracle Data Integrator, we have been able to deploy a standardized process for integrating large volumes of disparate data into the data warehouse with less development effort.”

—Dan Hazel,
Vice President, Sabre Holdings

an integration platform that will grow with your integration needs is crucial to sustaining your operations and getting the most out of your IT investments.

Keys to Flexibility

Oracle Data Integrator supports several integration styles: data, events, and services. It can move data directly from databases to different databases to or from packaged applications such as enterprise resource planning or customer relationship management, and to or from flat files. It can also provide data or events to an Enterprise Service Bus or a message-oriented middleware; it also becomes part of an SOA by providing data services and transformation services.

Oracle Data Integrator also addresses all latencies, from monthly, weekly, or daily batches; to hourly or more-frequent batches; to near-real time and real time. Integration processes can be triggered by a predefined schedule or by events from new data.

Unlike many “nonintegrated” integration platforms available on the market, which were in fact built through successive patchworks of disparate technologies, Oracle Data Integrator is a single platform that addresses all these styles and latencies of integration within a common framework. The development approach and methodology is integrated across the platform, and reusability of shared components within Oracle Data Integrator and across Oracle Fusion Middleware is high.

Oracle Data Integrator addresses all data integration needs without compromise. When used in conjunction with other parts of the Oracle Fusion Middleware it becomes a central component of an end-to-end IT architecture and provides top-level shared data services across the enterprise.

MORE PRODUCTIVE, FEWER INTEGRATION RESOURCES

The design phase of the integration processes is probably the most critical one in the entire IT infrastructure. Poorly designed integration will hinder dramatically the performance of the overall systems. Yet, it is also often a very time-consuming and costly process.

Declarative design, as introduced by Oracle Data Integrator, alleviates a number of the concerns traditionally associated with integration processes design. It allows business users to participate in the design, decreases the workload of the developers, and significantly reduces development and maintenance times.

Enablers for Design and Maintenance Productivity

Conventional approaches to ETL design, such as the ones from Informatica and IBM, require you to design the entire flow of data, positioning transformation and mapping code at the nodes of this flow. Not only is this a time-consuming process, it is also prone to serious performance problems—one misplaced transformation can dramatically impact the overall system performance. As a result, for this legacy

Oracle Data Integrator also addresses all latencies, from monthly, weekly, or daily batches, to hourly or more-frequent batches, to near-real time and real time.

approach to actually produce acceptable performance, it requires highly skilled integration developers who are very proficient in operating system and database tuning. But even worse, these approaches make your developers do repetitive work that needn't be done at all.

Suppose you want to populate a target table from three different source tables. With the conventional ETL approach, you need to design every step of the data flow: (1) connect to the three source tables, (2) sort the first source, (3) filter and look up data from the second source, (4) transform and stage the result of the lookup, (5) read and join the third source, (6) transform and stage the result of the join, and, then finally, (7) write to the target. These time-consuming steps need to be repeated every time you want to populate a different target from different sources.

“...an incredibly powerful development environment. [...] our development teams increased their productivity by at least 50 percent.”

**—Leon Schurgers,
Director of IT, ABN AMRO Interfinance**

Declarative design, the Oracle Data Integrator approach, allows you to accomplish the same result in only three steps: (1) link the three sources together, (2) declare the mappings between source fields and target fields, and (3) choose the built-in template that will generate the data flow. You don't need to worry about all the intermediate steps required in the conventional approach. They are generated for you by the Oracle Data Integrator Knowledge Modules (KMs).

Therefore, you concentrate on *what* your transformation is—declarative rules—rather than on *how* to do it—technical steps of the process. To populate different targets from different sources, you will always have only these three steps to complete, regardless of what the number of sources and targets are.

Further, suppose we want to update this previous example to add an extra source table. The conventional approach requires additional steps: (1) connect the new source table, (2) figure out the best place for it to get in the process, (3) read and sort it, (4) join it to the appropriate staging result, (5) transform and stage the new result, and (6) reconcile this new result within the existing process.

But with declarative design, you simply need to (1) link this additional source and (2) declare the new mappings between source fields and target ones. The Oracle Data Integrator KM will figure out automatically the best place for the new table to get in the process, and generate the appropriate internal flow for you.

Oracle Data Integrator will instantly provide your development team with

- A shorter learning curve; concentrate on mappings only
- Faster design; fewer steps to implement transformations
- Drastically reduced maintenance costs

Since behavioral changes are localized within the Oracle Data Integrator KMs, the incremental effects of change can be automatically inherited and propagated to all transformations within the Oracle Data Integrator system.

FAST TIME TO LAUNCH, MORE OPTIONS IN THE FUTURE

As previously discussed, the unique productivity benefits achieved with Oracle Data Integrator derive from an industry-first approach to data integration modularity called Knowledge Modules (KM). KMs provide Oracle Data Integrator with unmatched flexibility, productivity, and modularity that directly translate to an improved total cost of ownership.

Layers of Reconfigurable Knowledge Modules

Oracle Data Integrator's KMs are at the core of the product's integration architecture. They make integration processes modular, flexible, and extensible. KMs are master templates that define both the data flow implementation and the code generation.

KMs are generic, because they allow data flows to be generated regardless of the transformation rules. At the same time, they are highly specific, because the code they generate and the integration strategy they implement are finely tuned for a given technology or application platform.

Oracle Data Integrator provides a comprehensive library of KMs, which can be tailored to implement existing best practices. Examples of custom tailored KMs might include performance tuning to a specific database instance, enforcing adherence to corporate standards and policies on the data, or including specific vertical know-how within a given execution flow.

Because each KM is optimized for a given technology, porting a set of processes from one environment to another is as simple as picking the appropriate KM. Such KMs are the foundation of Oracle Data Integrator's hot-pluggability—enabling the Oracle Data Integrator to support the widest possible number of platforms and applications without performance overhead or the requirement of a separate ETL server.

KMs are the underlying components that enable declarative design and accelerate the overall design, development, implementation, and maintenance of the end-to-end data integration processes.

ORACLE'S VISION FOR DATA INTEGRATION

In summary, it should be clear that Oracle Data Integrator provides a unique data integration platform, architected for performance and productivity, which provides a high degree of flexibility and modularity. Compared to conventional ETL approaches from our competition, the Oracle Data Integrator excels in the following areas:

- **Performance:** It has high performance thanks to a market-leading, unique approach—the high-performance ELT architecture.
- **Productivity:** It has declarative design that drives reusability and accelerates development and maintenance.

Knowledge Modules are the underlying components that enable declarative design and accelerate the overall design, development, implementation, and maintenance of the data integration process.

- **Flexibility:** It has a single platform that natively supports several data integration styles and latencies.
- **Modularity:** It has hot-pluggable knowledge modules that support application and database platform diversity.

Oracle's commitment to its customers has always been to provide the highest quality enterprise software with maximum value. Oracle Data Integrator is a cornerstone technology for providing comprehensive data integration solutions and highlights the Oracle Fusion Middleware commitment to hot-pluggability and the broadest possible support for diverse IT environments.

Regardless of the database or applications within your IT ecosystem, the Oracle Data Integrator solution can be optimized to drive the highest-performance bulk or real-time transformations. Oracle's vision is to combine and enable these capabilities from within a next-generation, unbreakable service-oriented architecture that will continue to drive business value within your enterprise for many years to come.



Oracle Data Integrator: An Executive Overview
Updated January 2007
Yves de Montcheuil

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.