

An Oracle White Paper
July 2010

Oracle Data Integrator 11g New Features Overview

Disclaimer

This document is for informational purposes. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. The development, release, and timing of any features or functionality described in this document remains at the sole discretion of Oracle.

This document in any form, software or printed matter, contains proprietary information that is the exclusive property of Oracle. This document and information contained herein may not be disclosed, copied, reproduced, or distributed to anyone outside Oracle without prior written consent of Oracle. This document is not part of your license agreement nor can it be incorporated into any contractual agreement with Oracle or its subsidiaries or affiliates.

Executive Overview

Oracle Data Integrator (ODI) is a best-of-breed data integration platform focused on fast bulk data movement and handling complex data transformations. The 11g version of ODI pushes the state of the art technology even further ahead of the rest of the industry. Oracle has invested in the following key areas for ODI 11g:

- Enterprise Scale Deployment Patterns – availability, failover, and security
- Developer Productivity and Run-Time Performance – new IDE and core features improving integration flow design efficiency, simplicity, and performances
- Component Administration – template-driven deployment, Enterprise Manager integration
- Run-Time Management, Debugging and Diagnosability – code simulation, enhanced error management and session control
- Hot Pluggability and Heterogeneous Connectivity – continuous improvements each release to support new technologies and features for integration

Oracle Data Integrator is fully integrated with the Oracle technology stack, including Database, Exadata, Business Intelligence, and Applications. ODI is the strategic data integration platform for Oracle.

This whitepaper describes in detail the new features and capabilities offered in the ODI 11g data integration platform.

Architecture for Enterprise-Scale Deployment

Oracle Data Integrator 11g dramatically changes the vision of data integration in the information system. In fact, it makes data integration a key service in the information system, providing the backbone infrastructure for Enterprise Information Management.

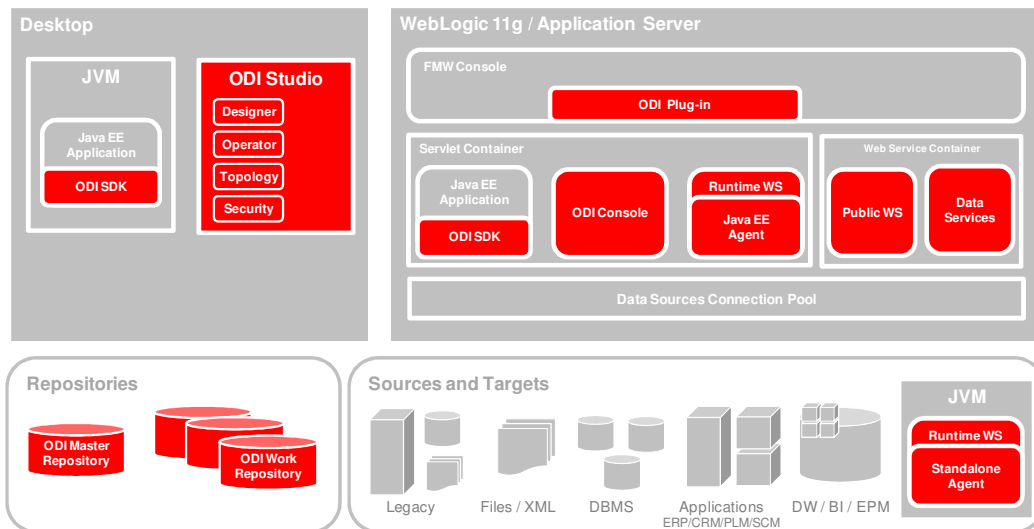


Figure 1 - Oracle Data Integrator 11g Component Architecture

The new Oracle Data Integrator platform integrates in the broader Fusion Middleware platform and becomes a key component of this stack. Oracle Data Integrator 11g provides its run-time components as Java EE applications, enhanced to fully leverage the capabilities of the Oracle WebLogic Application Server. Oracle Data Integrator components include exclusive features for Enterprise-Scale Deployments, high availability, scalability, and hardened security.

High Availability and Scalability

High-Availability (HA) and Scalability is fully supported via clustered deployments for Java EE components. The ODI components deployed in WebLogic Server benefit from the capabilities of the WLS cluster for scalability, including JDBC Connection Pooling and Load Balancing.

In addition to the cluster-inherited HA capabilities, the run-time agent also supports a Connection Retry mechanism to transparently recover sessions running in repositories that are stored in HA-capable database engines such as Oracle RAC.

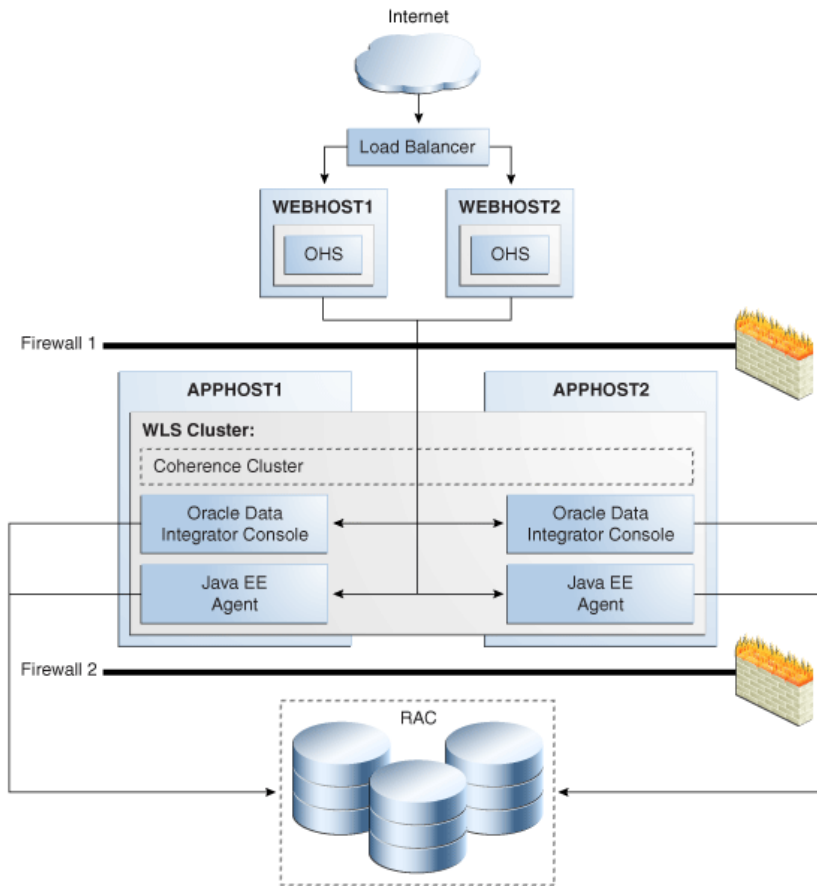


Figure 2 - High Availability Configuration for Oracle Data Integrator JEE Components

Corporate Security Integration

Security in Oracle Data Integrator can be hardened and unified with the corporate security stores.

The source and target data server passwords can be externalized in the enterprise credential store. The externalized passwords are no longer retained in the Oracle Data Integrator Repository.

In addition, the Oracle Data Integrator user can be mapped to users from the corporate identity store (LDAP, Oracle Internet Directory or Active Directory). With this feature, authentication is delegated to the corporate authentication service and Single Sign-On (SSO) can be enabled.

With these external password storage and external authentication features, passwords as well as user identities can be retained in a centralized location meeting any corporate standards for security.

Enterprise-Scale E-LT Architecture

The E-LT architecture mandates the use of lightweight run-agents able to run in strategic locations (for example on a Data Warehouse server) for optimal performances of the integration processes. Oracle Data Integrator 11g continues the support of Standalone Agents. Standalone agents support the same level of service as the Java EE Agent, including corporate security integration and connection retry, yet require a simple Java Machine. Standalone agents also support the ODI built-in load-balancing feature and can be protected using Oracle Process Manager and Notification Server (OPMN).

With this new architecture, every run-time component of Oracle Data Integrator guarantees high scalability, 24/7 availability, and best-in-class security.

Simplified Deployment and Unified Administration

Enterprise-Scale deployments frequently end up being a dilemma between simplicity and efficiency. Indeed, setting up components in highly available and highly secured configurations is a challenge that not every administrator is ready to accept. In addition, monitoring and managing such configurations can become a complex activity.

Simplified Deployment

Oracle Data Integrator simplifies the deployment and administration of the data integration services and components. Java EE components deploy easily and quickly in an Oracle WebLogic Server using preconfigured templates or templates that are generated based on the metadata defined in the topology. It is also possible to create data server definitions for sources and targets in the topology and to deploy these in few clicks as datasources in an Oracle WebLogic Server.

With these deployment capabilities, creating or extending a domain for data integration is a matter of minutes.

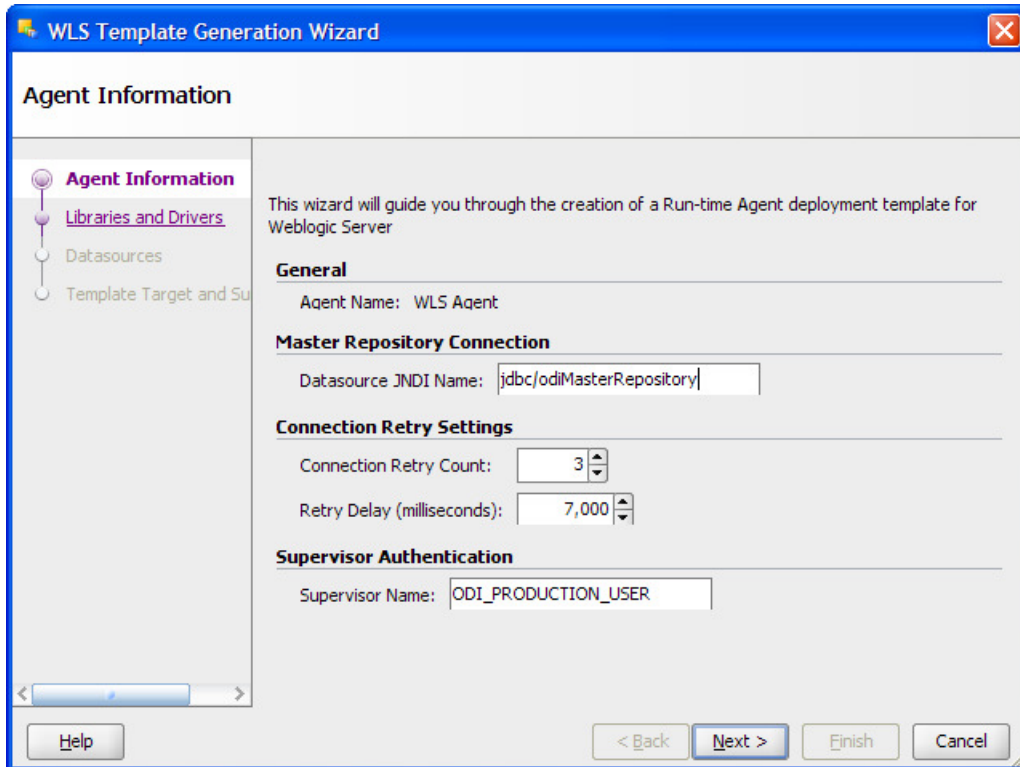


Figure 3 - Wizard for creating an Agent Template for WLS

Unified Administration and Monitoring

The Data Integration Architecture includes a number of components, including the Java EE and Standalone Agents as well as the Oracle Data Integrator Console web application. To manage and monitor these components, Oracle Data Integrator provides a new plug-in that integrates in Enterprise Manager Fusion Middleware Control.

With this plug-in, the Fusion Middleware administrator can manage from a single screen his data integration components along with his WebLogic servers, his deployed web applications, or his SOA composite applications.

This component allows the administrator to monitor the status, metrics, and notifications of Oracle Data Integrator components, including Repositories, Agents and Oracle Data Integrator Console instances.

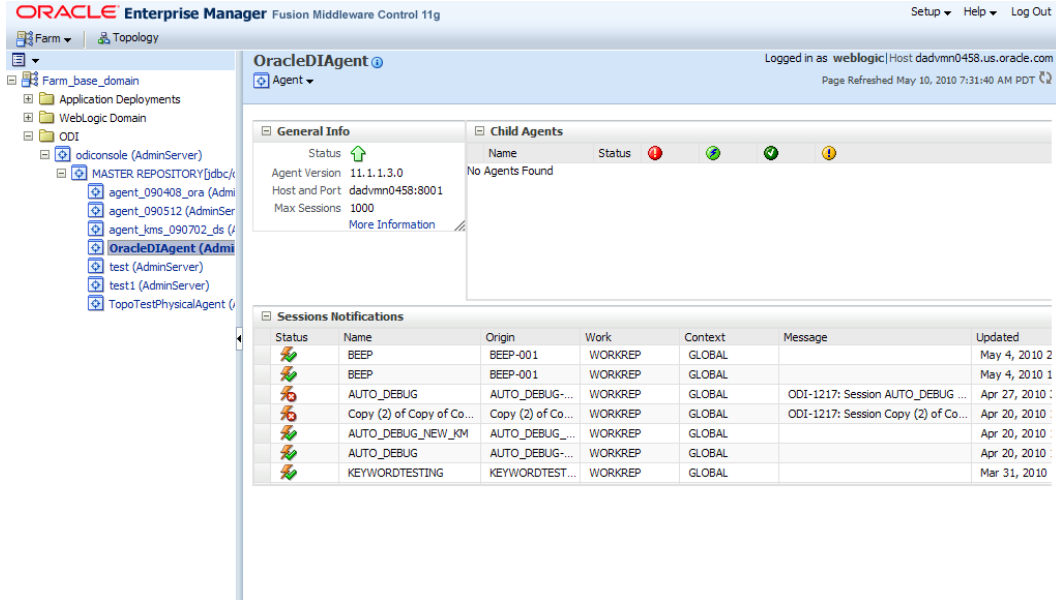


Figure 4 - Oracle Data Integrator in Oracle Enterprise Manager

Better Control over Production

Having a large-scale deployment involves frequently hundreds of integration flows and requires extreme tools for monitoring the production environment.

Oracle Data Integrator enhances the control capabilities available to the production operators by providing unique features.

Oracle Data Integrator Console

The new Oracle Data Integrator Console replaces and unifies the former 10g Metadata Navigator and Lightweight Designer components. This web interface for production operations has been rewritten using the ADF-Faces Ajax Framework for a rich user experience.

Using this console, production users can set up an environment, export and import the repositories, manage run-time operations, monitor the sessions, diagnose the errors, browse design-time artifacts and generate lineage reports.

This web interface integrates seamlessly with Oracle Fusion Middleware Control Console and allows Fusion Middleware administrators to drill down into the details of any Oracle Data Integrator design-time and run-time object.

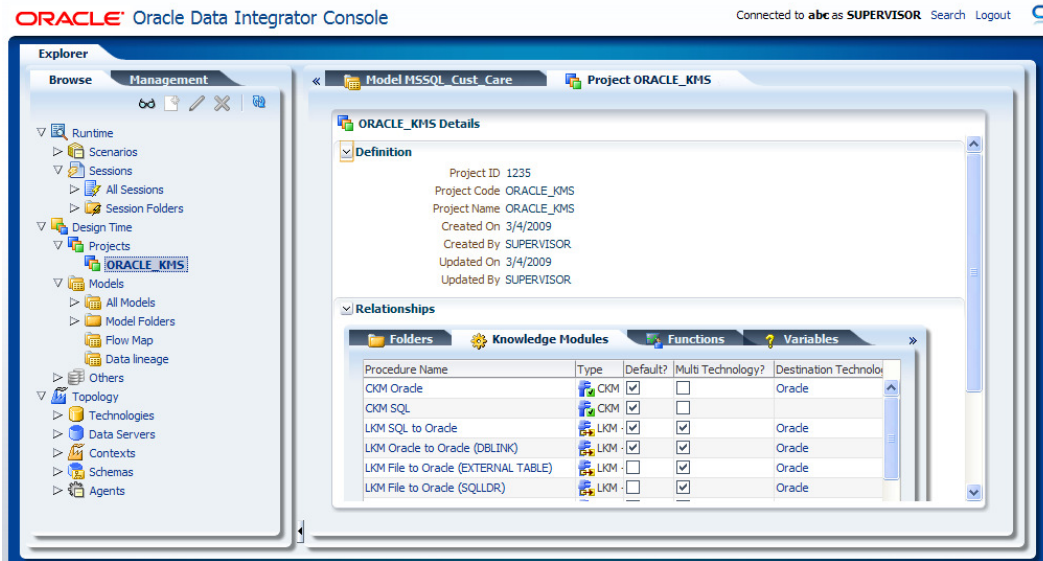


Figure 5 - Oracle Data Integrator Console

Enhanced Session Control

Sessions can now be stopped in an immediate mode. This new mode aborts the current operation (for example, a long-running SQL statement started on a database engine) instead of waiting for its completion before stopping the session.

In addition to this mechanism, sessions that failed due to a repository or the run-time agent failure are automatically identified as stale and moved from the *Running* state to the *Error* state for operator review.

Enhanced Error Management

Error messages raised by Oracle Data Integrator Components and Sessions have been enhanced to provide administrators and production operators with precise information for troubleshooting and fixing the status of the architecture and debugging the sessions.

Enhanced messages cover component lifecycle (agent startup, shutdown, schedule refresh, and so forth) as well as session lifecycle (incorrect scenario version, load balancing issue, agent not available, and so forth). They render clearer errors starting at the lowest level of the session (steps and task). These error messages are enriched with context and infrastructure information to help development or production users quickly find the source and solution for an issue.

Errors raised by components or sessions are automatically notified to the operation administrator in Enterprise Manager Fusion Middleware Control.

Design-Time Productivity

Oracle Data Integrator 11g introduces a new JDeveloper-based integrated development environment (IDE) called the Oracle Data Integrator Studio. This client is entirely redesigned in this release to dramatically increase the developer's productivity and make advanced features more accessible.

Design-Time Experience

New IDE Based on JDev

The new Oracle Data Integrator studio is used as a replacement for all Oracle Data Integrator 10g modules (Designer, Topology, Operator, and Security Manager). All the features of these modules now appear as Navigators within the Oracle Data Integrator Studio.

This new IDE provide state-of-the-art features for developer productivity, including redesigned and reorganized editors in addition to enhanced windows, improved docking, and document navigation and management.

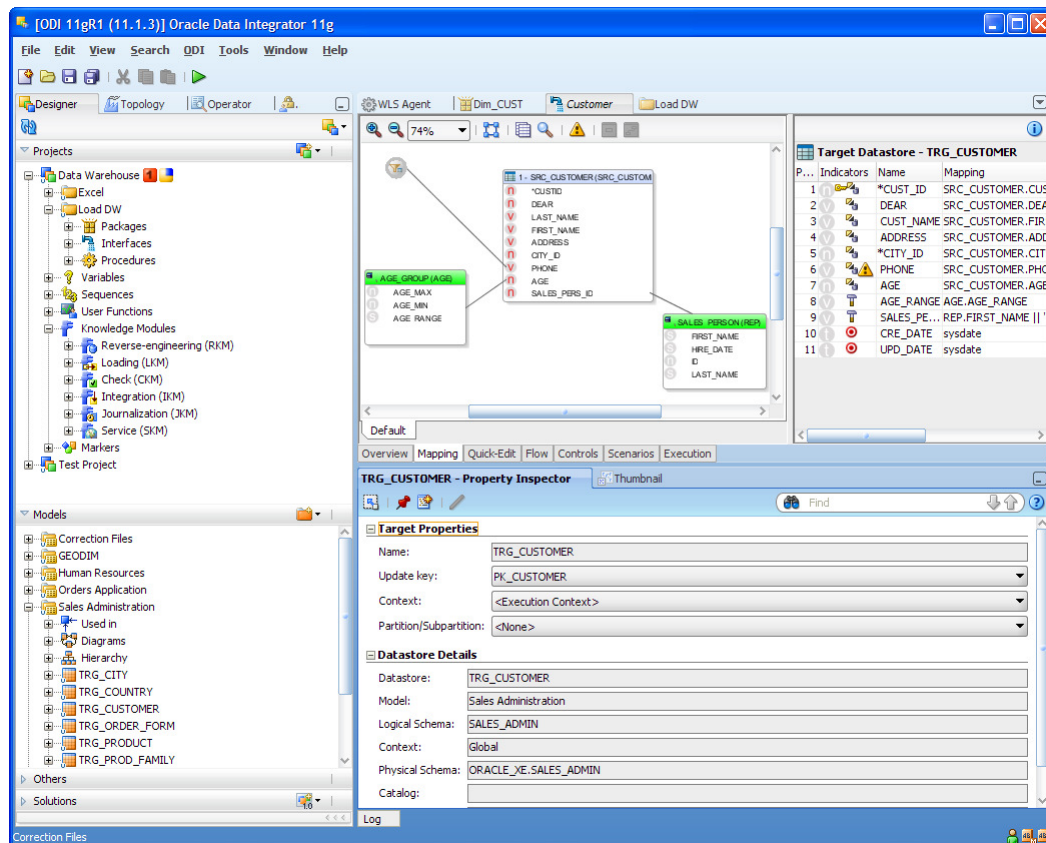


Figure 6 – In Oracle Data Integrator Studio, Editing the “Customer” Interface

Redesigned Interface Editor

The Interface Editor, used to create the integration interfaces, has been entirely redesigned to use the JDeveloper diagramming framework, providing:

- Enhanced diagram look and feel, plus manual or automated diagram layout organization
- Better support of large diagrams via thumbnails, zoom in/out, and bird's eye
- Small but useful features such as source column drag and drop in the target mapping and column sort/selection in the mapping
- Reorganized and searchable Property Inspector

Auto-Fixing

When saving an interface or clicking the Error button in the Interface Editor toolbar, the list of all design errors is displayed with meaningful messages and tips. Automated fixes are suggested and can be applied with a single click.

Quick-Edit

The new Interface Editor includes a new Quick-Edit tab to edit the interface diagram faster. The Quick-Edit Editor displays the interface components in a tabular form, supports mass-updates, and intuitive and accessible keyboard navigation.

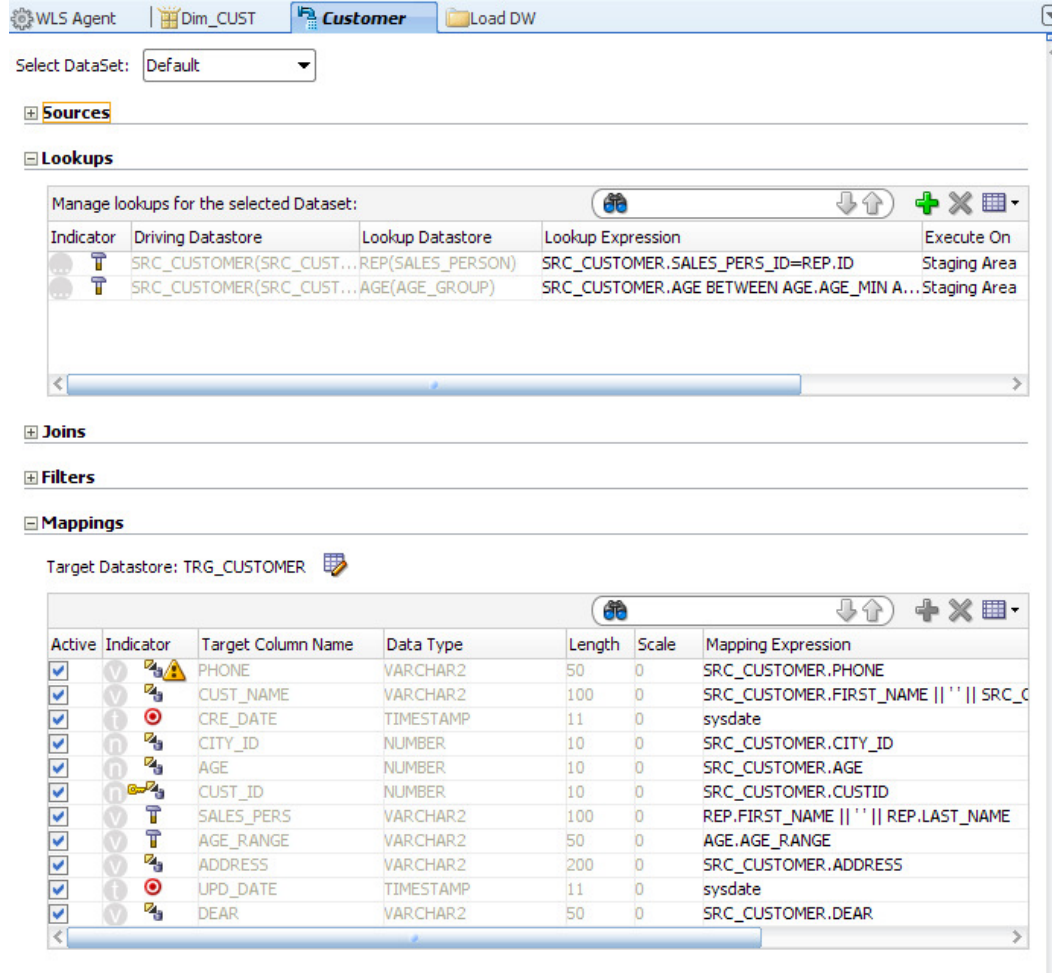


Figure 7 - Quick Edit view for the “Customer” Interface

Small (yet Important) Things

In addition to these major changes, some small improvements have been made in ODI to make the developer’s life easier. For example:

- **Code Simulation** - When running a session at design-time, it is possible to make a code simulation instead of running a full execution. This simulation mode generates a session simulation report displaying the generated code for review.
- **Reverse-Engineering** - When a model is created, the reverse-engineering context is automatically set to the default context. In addition, when performing a selective reverse-engineering, the system tables are now hidden from the display.

- **Scenario Naming Convention** – A user can define a pattern (using for example the object name, folder path, or project name) for the naming convention that is automatically applied to new scenarios created by this user.
- **Long Object Names** - Object names have been extended to support long database object names (128 characters) and repository object labels (400 characters).

Java API for Task Automation

Oracle Data Integrator 11g introduces a new Java API for creating and managing both design-time and run-time operations.

Using this API, it is for example possible to perform the following operations:

- Create or modify the topology contents
- Set up projects and models
- Create and browse interfaces, procedures, and packages
- Start and manage sessions and monitor their status

This API allows developers to programmatically perform operations without using the user interfaces and to create their own applications to access the repository content.

E-LT Features for Better Performances

New core features have been built in Oracle Data Integrator Studio around the E-LT architecture and declarative design approach to support a more productive development process and better performances at run-time.

Datasets and Set-Based Operators

This major enhancement introduces the notion of datasets in interfaces. A dataset represents the data flow coming from a group of joined and filtered source datastores. Each dataset includes the target mappings for this group of sources. Several datasets can be merged into the interface target datastore using set-based operators such as Union and Intersect.

Figure 9 illustrates the use of datasets to merge data from a source file and from an HR system. Each of these datasets (From File, From HR) is a flow with its own set of source tables, joins, filters and mappings.

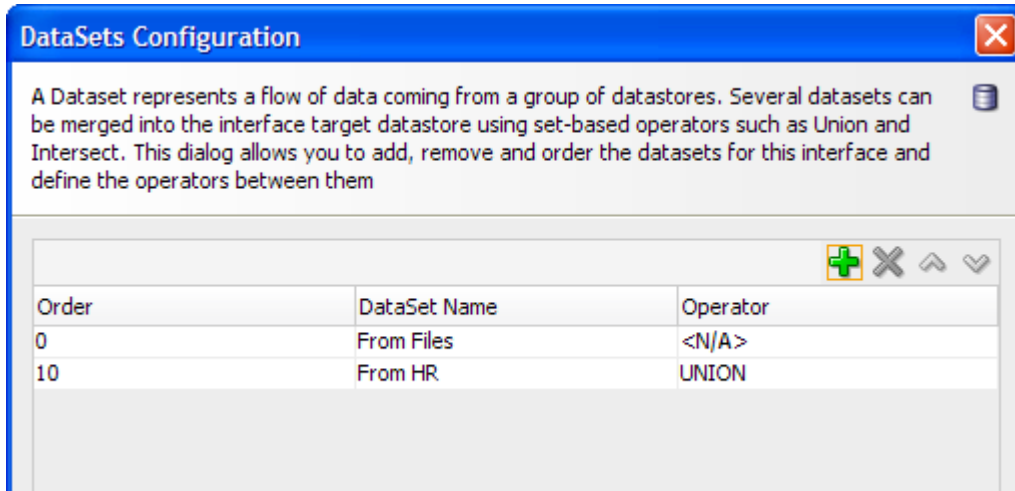


Figure 8 – One interface merges with a UNION operator flows from two diagrams

These two flows appear in different diagram tabs in the Interface Editor. They are merged using the UNION operator as shown in Figure 9.

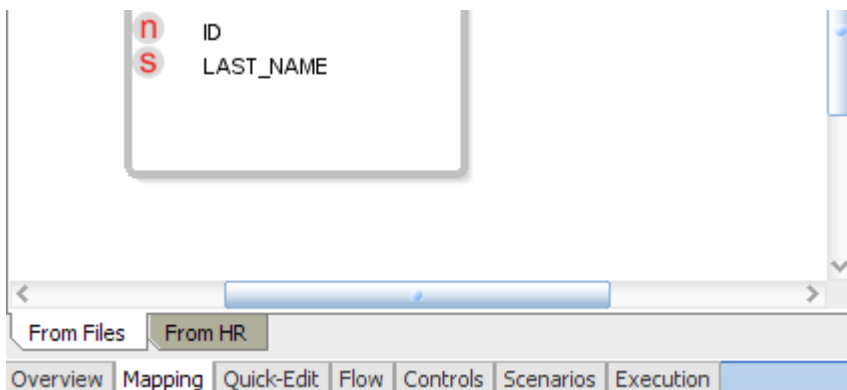


Figure 9 – Two diagram tabs for the datasets in the Interface Editor.

This feature accelerates the interface design and reduces the number of interfaces needed to merge several data flows into the same target datastore.

Partitioning

Oracle Data Integrator now supports partitioning features of the data servers. Partitions can be reverse-engineered using RKM or manually created in models. When designing an interface, it is possible to define the partition to address on the sources and target datastores. Oracle Data

Integrator code generation handles the partition usage syntax for each technology that supports this feature.

Lookups

A wizard is available in the Interface Editor to design a lookup from a driving source datastore to a lookup datastore or interface. These lookups appear as a compact graphical object in the Sources diagram of the interface. The user can choose how the lookup is executed: as a Left Outer Join in the FROM clause or as an expression in the SELECT clause (in-memory lookup with nested loop). This second syntax is frequently more efficient on small lookup tables.

This feature simplifies the design and readability of interfaces using lookups, and allows optimized execution of these lookups.

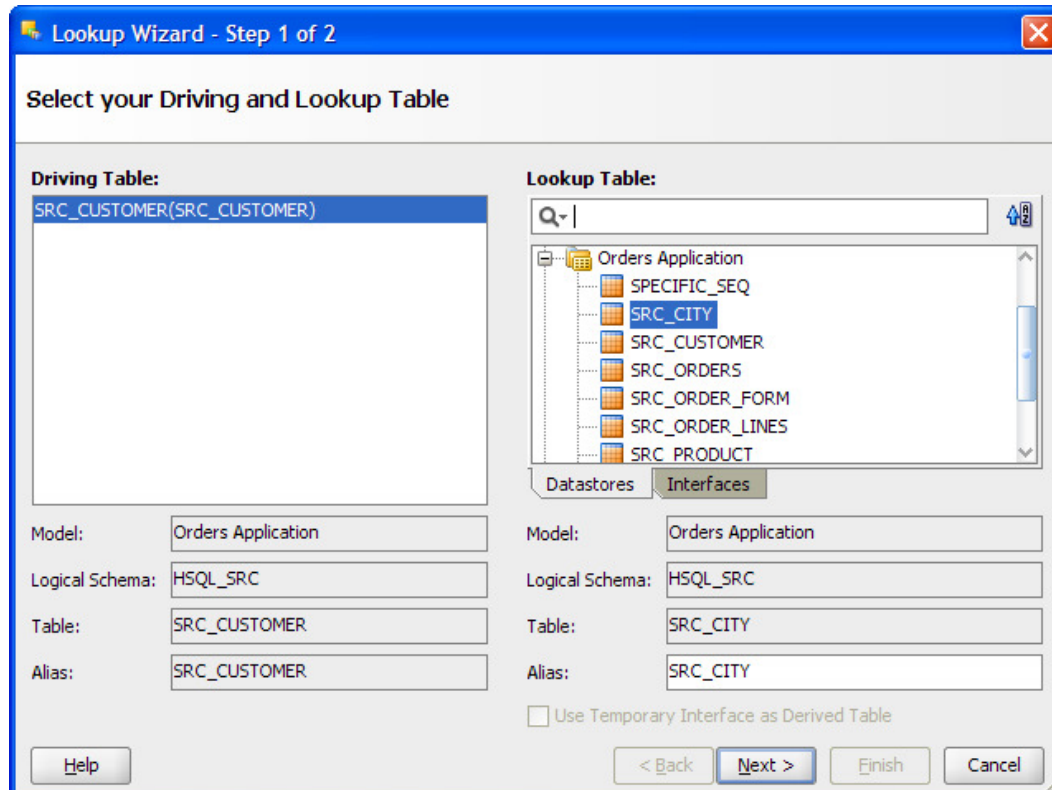


Figure 10 - Lookup Creation Wizard

Derived Select for Temporary Interfaces

When using a temporary interface as a source in another interface, it is possible to avoid persisting the target datastore of the temporary interface by generating instead a Derived Select (sub-select) statement corresponding to the execution of this temporary interface. Consequently, the temporary interface no longer needs to be executed to load the temporary datastore. The code generated for the sub-select can be optionally customized in an IKM.

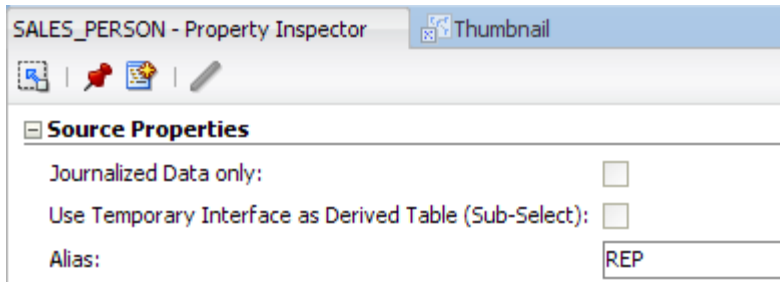


Figure 11 - Use Temporary Interface as Derived Table option can be checked when an interface is used as a source.

This feature eliminates the need for complex packages to handle the execution order of temporary interfaces. It dramatically simplifies the execution of cascades of temporary interfaces.

Support for Native Sequences

Oracle Data Integrator now provides support for a new type of sequences that directly maps to database-defined sequences. Such a sequence is selected from a list retrieved from the database when the sequence is created. Native Sequences are used as regular Oracle Data Integrator sequences, and the code generation automatically handles technology-specific syntax for sequences.

This feature simplifies the use of native sequences in all expressions and enables cross-referencing on native sequence usage.

Support for Natural Joins

Oracle Data Integrator now provides support for Natural joins, defined at technology level. A natural join does not require any join expression to be specified, and is handled by the engine that processes it. This engine matches automatically columns with the same name.

Automatic Temporary Index Management

When creating joins or filters on source tables, it is possible to have Oracle Data Integrator automatically generate temporary indexes for optimizing the execution of these joins or filters. The user simply selects the type of index from a list for the given technology. Knowledge modules automatically generate the code for creating and dropping the temporary index as needed.

This feature accelerates join and filter execution and enables better performances for integration interfaces.

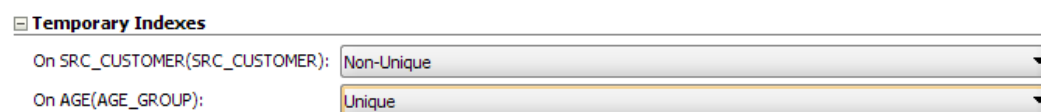


Figure 12 – Two temporary Indexes (Non-Unique and Unique) will be created for optimizing the execution of the join between SRC_CUSTOMER and AGE_GROUP.

Broaden Connectivity

Data Integration is about moving and transforming data from sources to targets in a heterogeneous world. In this context, hot-pluggability, that is being able to access and integrate any source and any target in the most performing way, is a constant challenge that Oracle Data Integrator addresses using the knowledge modules architecture.

The Oracle Data Integrator highly flexible knowledge module architecture allowed the release of significant connectivity improvements in the previous version lifecycles. These improvements included:

- Application Adapters for Oracle E-Business Suite, PeopleSoft, Siebel, JDE EnterpriseOne, SAP ERP, and SAP BW,
- Performance optimizations for Teradata
- Adapters for Oracle OLAP and Oracle Changed Data Capture.
- Adapters for Hyperion Planning, Financial Management, and Essbase
- Changed Data Capture using Oracle GoldenGate.

Oracle Data Integrator 11g adds to this long list the following improvements:

- Adapter for sourcing from ADF-View Objects exposed in an Oracle Business Intelligence Enterprise Edition Server
- Support for Oracle Database Multi-Table Insert
- Support for Teradata Multi-Statements

Of course, the Knowledge Modules have also been enhanced in various manners to support the new productivity and performance enhancements such as datasets and temporary indexes.

Conclusion

In this release, Oracle Data Integrator introduces a new architecture, as well as features for simplifying the deployment and management of enterprise-scale and highly-available configurations. New interfaces based on Fusion Middleware standards (ADF, JDeveloper) accelerate design-time as well as run-time operations for data integration flows. Core features, such as datasets, lookups, derived select increase again development productivity and performance while preserving the key product differentiators: Declarative Design, Knowledge Modules, Hot-Pluggability, and E-LT.



Oracle Data Integrator 11g New Features
Overview
July 2010

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

Copyright © 2010, Oracle and/or its affiliates. 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 is a registered trademark of Oracle Corporation and/or its affiliates. Other names may be trademarks of their respective owners.