An Oracle White Paper

May 2011

Oracle Data Integrator 11*g* 11.1.1.5 New Features Overview

ORACLE

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.

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.

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

The 11*g* version of ODI has pushed the state of the art technology in data integration further ahead of the rest of the industry. Since the first 11*g* release in July 2010, Oracle has continued its investments on this strategic data integration platform.

For this new ODI 11.1.1.5 patch level, Oracle has consolidated 18 one-off patches and included new fixes for many other bugs. This whitepaper describes in detail some of the new bug fixes, and explains more about the new features and capabilities offered in the 11.1.1.5 version of the Oracle Data Integrator 11*g* platform.

Feature Highlight: SOA and Application Integration

Oracle Data Integrator is the only ETL tool that can run entirely within a Service Oriented Architecture. ODI is the most deeply integrated ETL product for running data integration as a service, accessing web services and writing to web services as a target. ODI 11.1.1.5 has been further enhanced for facilitating application integration and more advanced interaction with a SOA stack.

Complex File Technology

Complex file formats (multiple record files) can now be integrated using the new Complex File technology. This technology leverages a new built-in driver that converts transparently complex file formats into a relational structure using a Native Schema (nXSD) description file.

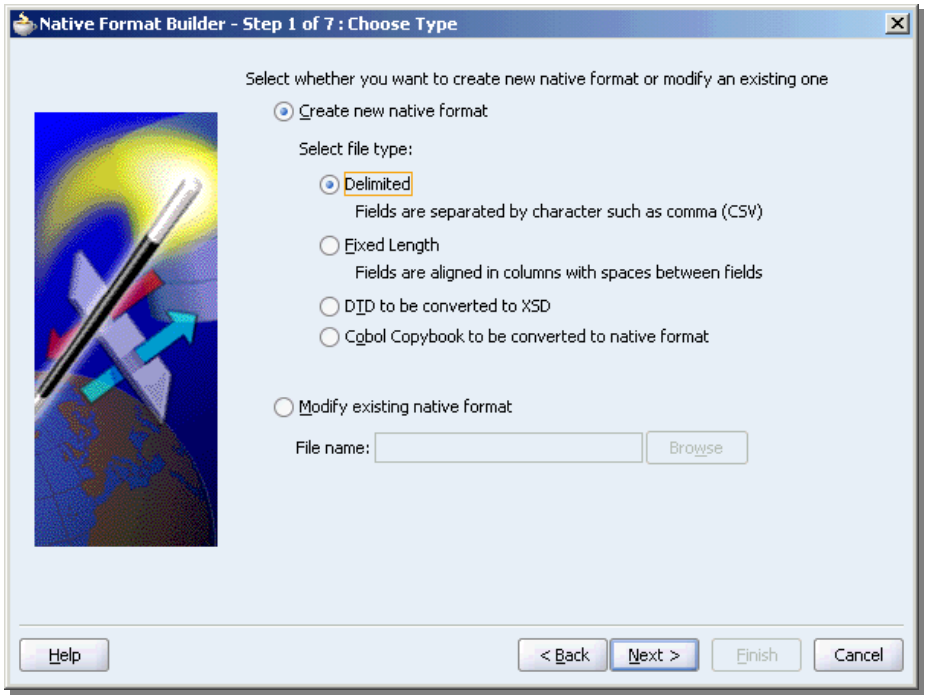


Figure 1 – Complex files described with the Native Format Builder can be integrated in Oracle Data Integrator.

With ODI 11.1.1.5 the nXSD formats may be created manually, or with a graphical tool called the Native Format Builder Wizard (included with BPEL Process Manager). The Native Format Builder Wizard guides you through the creation of a native schema file from the following formats shown in Figure 1. A sample data file format for the selected type must already exist; you cannot create a native schema without one. You can also select to modify an existing native schema previously created with this wizard, except for those generated from a Document Type Definition (DTD) or COBOL copybook.

Web Services Enhancements

Web service support in Oracle Data Integrator has been enhanced with the following capabilities:

Web Service Container Based Authentication

When external authentication and container based authentication with Oracle Platform Security Services (OPSS) are configured, authentication can be passed to the ODI Run-Time Web Services using HTTP basic authentication, WS-Security headers, SAML tokens and so forth.

Asynchronous Web Services and Callbacks

Oracle Data Integrator batch jobs are long running sessions in a web service context. Long running transactions are handled using asynchronous web service calls and callbacks. Starting with this release, a scenario or session can be started using the Run-Time Web Services on a one-way operation. When the session completes, the result of the execution can trigger an operation on a callback address. This pattern can be used for handling any WSAddressing compliant long running sessions, started for example with Oracle BPEL.

Full SOAP Support for Outbound Calls

The OdiInvokeWebService tool now supports full-formed SOAP messages as an input, including the SOAP header and body. Access to the SOAP header improves the ease of interaction with web service security, authentication, and other application-specific information.

Features Highlights: BI and Data Warehousing

The following features have been added in this patch level for customers using Oracle Data Integrator in Business Intelligence and Data Warehousing Projects.

Load Plans

Oracle Data Integrator is often used for populating very large data warehouses. In these use cases, it is common to have thousands of tables being populated using hundreds of scenarios. The execution of these scenarios has to be organized in such a way that the data throughput from the sources to the target is the most efficient within the batch window. Load Plans help the user organizing the execution of scenarios in a hierarchy of sequential and parallel steps for these types of use cases.

Load Plans are new objects introduced in this release to organize at a high level the execution of packages and scenarios. Load Plans provide features for parallel, sequential, and conditional scenario execution. They also support advanced restartability and exception handling capabilities.

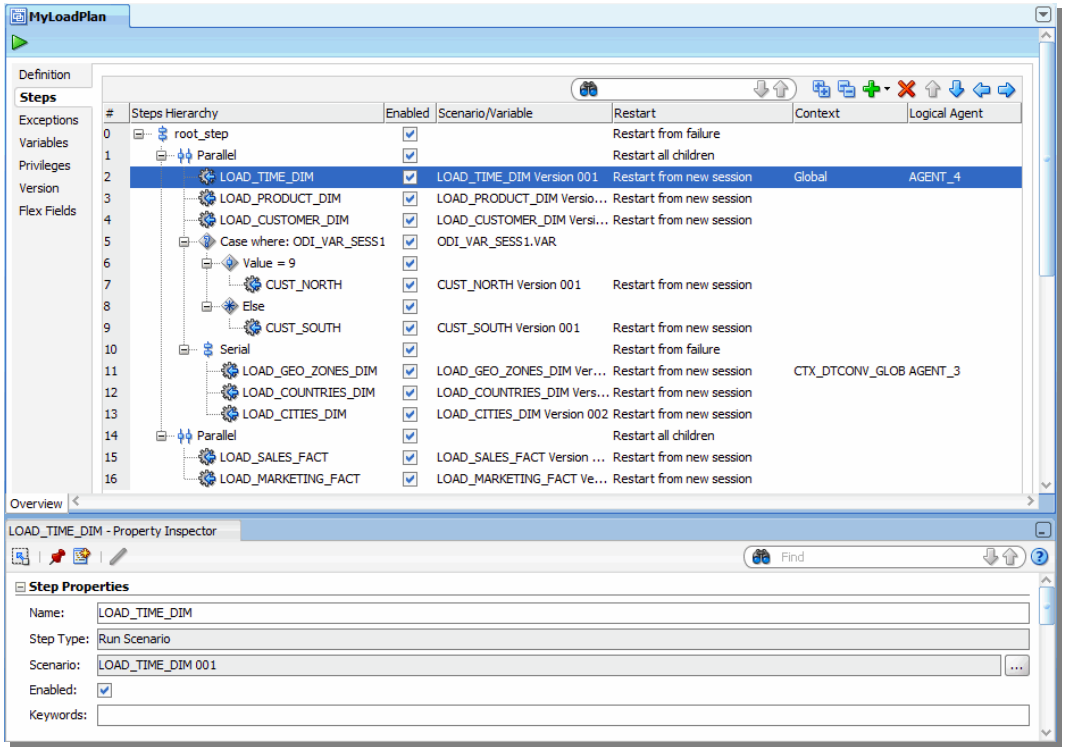


Figure 2 – Load Plans support parallel, sequential and conditional execution

Load Plans can be created and modified in production environments. They can be started, stopped, and restarted from a command line, from Oracle Data Integrator Studio, Oracle Data Integrator Console or a Web Service interface. They can also be scheduled using the run-time agent's built-in scheduler or an external scheduler.

Load plans are not substitutes for packages or scenarios. They are used to organize at a higher level the execution of packages and scenarios. Unlike packages, Load Plans provide native support for parallelism, restartability and exception handling. Load plans are moved to production as is, whereas packages are moved in the form of scenarios. Load Plans can be created in Production environments.

The Load Plan instances and Load Plan runs are similar to Sessions. The difference is that when a session is restarted, the existing session is overwritten by the new execution. The new Load Plan Run does not overwrite the existing Load Plan Run, it is added after the previous Load Plan Runs for this Load Plan Instance. The Load Plan Instance cannot be modified at run-time.

A scenario is an atomic runtime component that is given to the production. When in production, the production administrator can orchestrate the various scenarios using Load Plans in order to optimize the time taken to execute the ETL jobs in the given timeframe and ease the administration of all the ETL jobs.

For example, the production administrator will:

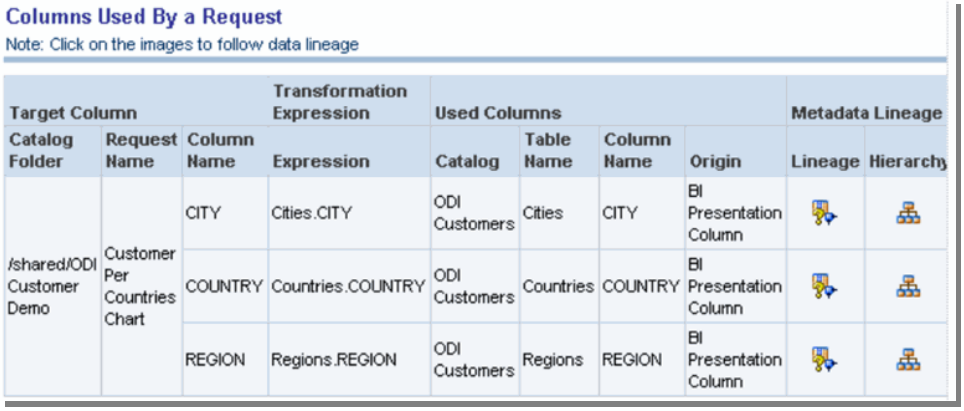
- Rely on Load Plans to orchestrate the execution of the various Scenarios produced by the ETL developers as part of their own projects and will organize the production globally instead of a per-project approach.
- Design the load plans in order to handle the parallelism of the scenarios in order to best use the hardware resources
- Design the load plans in order to handle serial steps to avoid overloading the system with too many scenarios at a given time
- Modify the load plans to take into account the real execution times and tune the load on the systems
- Rely on the load plans to restart only what is needed to be restarted in case of failure. This relies on the fact that the scenarios should be designed in order to be fully restartable (see above)

Oracle Business Intelligence Enterprise Edition Data Lineage

Oracle Business Intelligence Enterprise Edition (OBIEE) users need to know the origin of the data displayed on their reports. When this data is loaded from source systems into the data warehouse using Oracle Data Integrator, it is possible to use the Oracle Data Integrator Lineage for Oracle Business Intelligence feature to consolidate ODI metadata with OBIEE and expose this metadata in a report-to-source data lineage dashboard in OBIEE.

Figure 3 – OBIEE Users can click to view the Lineage of the data from their dashboards.

An important capability that Oracle has provided is the ability to view detailed metadata lineage information from directly within the BI Answers dashboards.



Columns Used By a Request

Note: Click on the images to follow data lineage

Target Column			Transformation Expression	Used Columns				Metadata Lineage	
Catalog Folder	Request Name	Column Name	Expression	Catalog	Table Name	Column Name	Origin	Lineage	Hierarchy
/shared/ODI Customer Demo	Customer Per Countries Chart	CITY	Cities.CITY	ODI Customers	Cities	CITY	BI Presentation Column		
		COUNTRY	Countries.COUNTRY	ODI Customers	Countries	COUNTRY	BI Presentation Column		
		REGION	Regions.REGION	ODI Customers	Regions	REGION	BI Presentation Column		

Figure 4 – OBIEE Users can see the Lineage details directly in their BI dashboards.

Other Features

In addition to the features listed above, the following features have been included in this patch level.

Commands on Connect/Disconnect

It is now possible to define data server commands that will be automatically executed when connections to this data server are created or closed by ODI components or by a session. Such a command can be used for altering and pre-configure sessions (for example, for internationalization of messages or forcing implicit conversions), or in multi-tenant infrastructure where commands must be placed at connection time to access information of a given tenant.

Built-in Technology Additions and Updates

The following technologies used in Oracle Data Integrator have been added and updated to benefit from their new features:

- Embedded HSQL engine is upgraded to version 2.0. This embedded engine is used for the Memory Engine as well as the XML and LDAP Drivers' built-in storage
- Jython BSF engine is updated to version 2.1
- JAX-WS/JRF is now used as a standard stack for web service calls and processing. Axis is no longer used

Groovy Technology

Groovy is added to the list of scripting engines supported by Oracle Data Integrator for use in knowledge modules and procedures.

Ordered and Non-Ordered Join Syntax

Technologies can now support both the ordered or non-ordered (database-specific) syntax for joins. The Oracle DB technology was modified to support both join syntaxes.

New Method for Setting Task Names

A new setTaskName method is available to update at run-time the name of a task. This API can be used in knowledge modules and procedures to modify at run time the task name displayed in the log, making it more readable for production users.

Shared Library for WLS Agent

A new template called Oracle Data Integrator - Agent Libraries includes libraries shared by all the deployed JEE agent in a domain, and must be deployed before the Oracle Data Integrator - Agent default template or a generated template.

This shared library facilitates the deployment, patching and maintenance of ODI domains, as it is deployed once and shared by both the default and user-generated templates.

Performance Optimization

The following optimizations have been made in the design-time and run-time components to improve their performance:

- Long text storage modified to use CLOBs.
- Agent-Repository network communications reduced at run-time
- Agent JDBC to JDBC loading mechanism reviewed and optimized

.

Conclusion

With this ODI 11.1.1.5 patch level Oracle has consolidated numerous one-off patches, and introduced new enhancements for customers using its strategic data integration platform in various contexts - BI, Data warehouse, Application and Service Oriented Integration - while preserving the key product differentiators: Declarative Design, Knowledge Modules, Hot-Pluggability, and E-LT.

Oracle Data Integrator 11*g 11.1.1.5*
New Features Overview
Author: FX Nicolas
May 2011

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

0109