

# Oracle Data Integrator

*An Oracle Technical White Paper*  
*Febrero 2009*

## **INTRODUCCIÓN**

Los datos son uno de los activos más importantes de las empresas, y la integración de datos constituye la columna vertebral de los sistemas de TI. La integración de datos es el conjunto de tecnologías que realizan este movimiento en una variedad de enfoques de arquitectura, como ETL o E-LT, dependiendo de los requerimientos individuales.

Oracle Data Integrator ofrece una plataforma activa de integración que incorpora todos los estilos de integración de datos: basado en datos, basado en eventos y basado en servicios. Es capaz de transformar grandes volúmenes de datos eficientemente, procesando eventos en tiempo real a través de su capacidad avanzada de Changed Data Capture (CDC), u ofreciendo servicios de datos hacia Oracle SOA Suite, ODI unifica los silos de integración. Además ofrece características de controles de integridad de datos robustas, asegurando la consistencia y calidad de los datos.

Con poderosos diferenciadores principales - E-LT heterogéneo, diseño declarativo y módulos de conocimiento (KM) – Oracle Data Integrator cubre los requerimientos de rendimiento, flexibilidad, productividad, modularidad e interoperabilidad de una plataforma de integración.

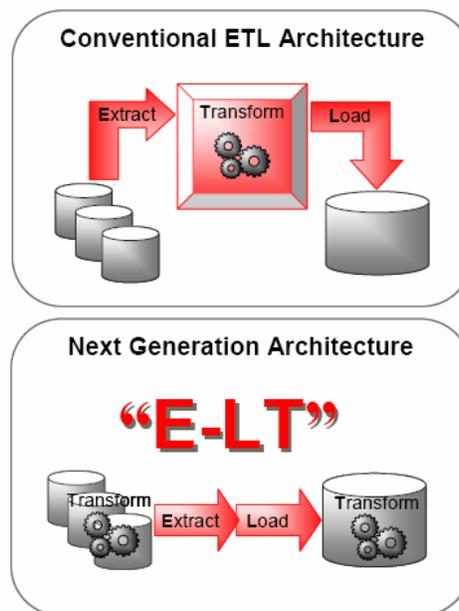
## **E-LT HETEROGENEO**

Las herramientas tradicionales ETL operan primero extrayendo los datos de varias fuentes, transformando los datos en un motor ETL de capa media, propietario, y luego cargando los datos transformados en el datawarehouse o servidor de integración destino. Es por eso que el término ETL representa tanto los nombres como el orden de las operaciones realizadas.

El paso de transformación de datos en el proceso ETL es de lejos el más intensivo y es realizado íntegramente por el motor ETL propietario en un servidor dedicado. El motor ETL realiza transformaciones de datos (y a veces revisiones de calidad de datos) fila por fila, y por lo tanto, puede fácilmente convertirse en el cuello de botella de todo el proceso. Adicionalmente los datos deben moverse sobre la red

dos veces – una entre las fuentes y el servidor ETL, y otra vez entre el servidor ETL y el data warehouse destino.

La arquitectura E-LT incorpora los mejores aspectos tanto de la codificación manual como de las propuestas ETL en la misma solución. El enfoque E-LT cambia el dónde y cómo las transformaciones de datos tienen lugar, y aprovechan las habilidades existentes de los desarrolladores, los motores RDBMS y el hardware del servidor de la mayor forma posible. En esencia, el E-LT reubica el paso de transformación de datos en el RDBMS destino, cambiando el orden de las operaciones a: extraer los datos de las tablas fuentes, cargar las tablas en el servidor destino, y luego transformar los datos en la base de datos objetivo utilizando operadores SQL nativos.



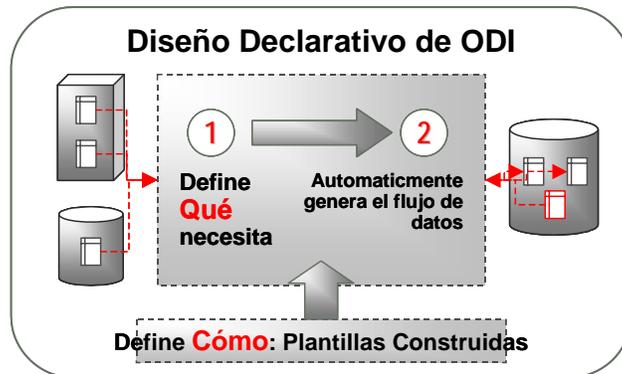
## DISEÑO DECLARATIVO

Oracle Data Integrator acorta los tiempos de implementación con su enfoque de diseño declarativo. Los diseñadores especifican qué es lo que quieren lograr con sus datos, y luego la herramienta genera los detalles de cómo realizar la tarea.

A diferencia del diseño ETL convencional, con ODI, el usuario de negocio o el desarrollador especifica las reglas a aplicar en el proceso de integración. La herramienta genera automáticamente los flujos de datos, administra su complejidad, y administra las instrucciones correctas para los diversos sistemas fuentes y destinos.

Con el diseño declarativo, el número y complejidad de pasos es ampliamente reducido, lo cual a su vez acorta los tiempos de implementación. La generación de

código automática reduce la curva de aprendizaje para los desarrolladores y direcciona el acceso de profesionales no técnicos en la definición de sus procesos de integración y formatos de datos.



## MODULOS DE CONOCIMIENTO (KM)

Los Módulos de Conocimiento (Knowledge Modules) son el núcleo de la arquitectura de Oracle Data Integrator. Ellos hacen a los procesos de ODI modulares, flexibles y extensibles.

Los Módulos de Conocimiento implementan los flujos de datos actuales y definen las plantillas para la generación de código a través de múltiples sistemas involucrados en cada proceso.

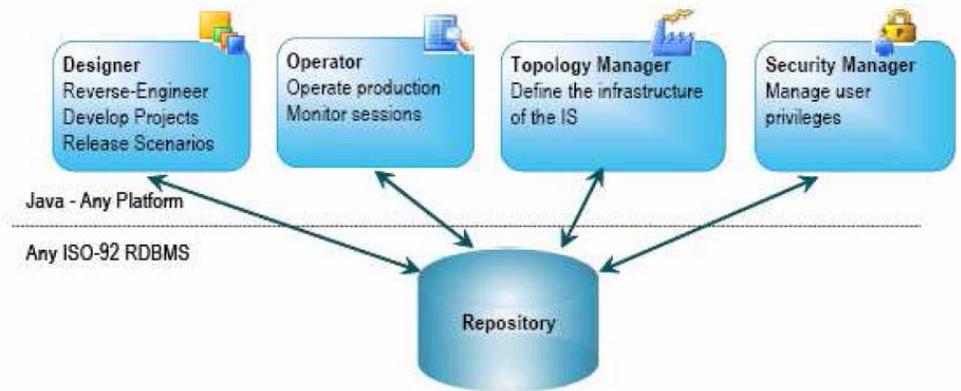
Los Módulos de Conocimiento son genéricos debido a que permiten que los flujos de datos sean generados sin importar las reglas de transformación. Y son altamente específicos, debido a que el código que ellos generan y la estrategia de integración que implementan son finamente afinados para una tecnología dada.

Oracle Data Integrator ofrece una librería extensa de Módulos de Conocimiento, los cuales pueden ser acortados para implementar mejores prácticas existentes (por ejemplo, para el más alto rendimiento, para adherirse a estándares corporativos, o para un conocimiento vertical específico).

Ayudando a capturar y reutilizar el conocimiento y mejores prácticas, el framework de Módulos de Conocimiento reduce el costo de propiedad, y permite la extensibilidad de la funcionalidad de producto para cubrir los retos de integración de datos más demandantes.

## COMPONENTES DE ORACLE DATA INTEGRATOR

La arquitectura de Oracle Data Integrator (ODI) está organizada sobre la base de un repositorio modular, el cual es accedido por agentes y módulos con una interfaz gráfica en arquitectura cliente/servidor, los cuales están escritos completamente en Java. La arquitectura incluye una aplicación web, Metadata Navigator, que permite a los usuarios acceder a las informaciones a través de un navegador.



### MODULOS GRAFICOS

Los cuatro módulos gráficos son: Designer, Operator, Topology Manager y Security Manager. Estos módulos pueden ser instalados en cualquier plataforma que soporte una Java Virtual Machine 1.5 (J2SE), incluyendo Windows, Linux, HP-UX, Solaris, AIX y Mac OS, entre otras.

A continuación presentamos una visión general de los módulos referidos.

#### Designer

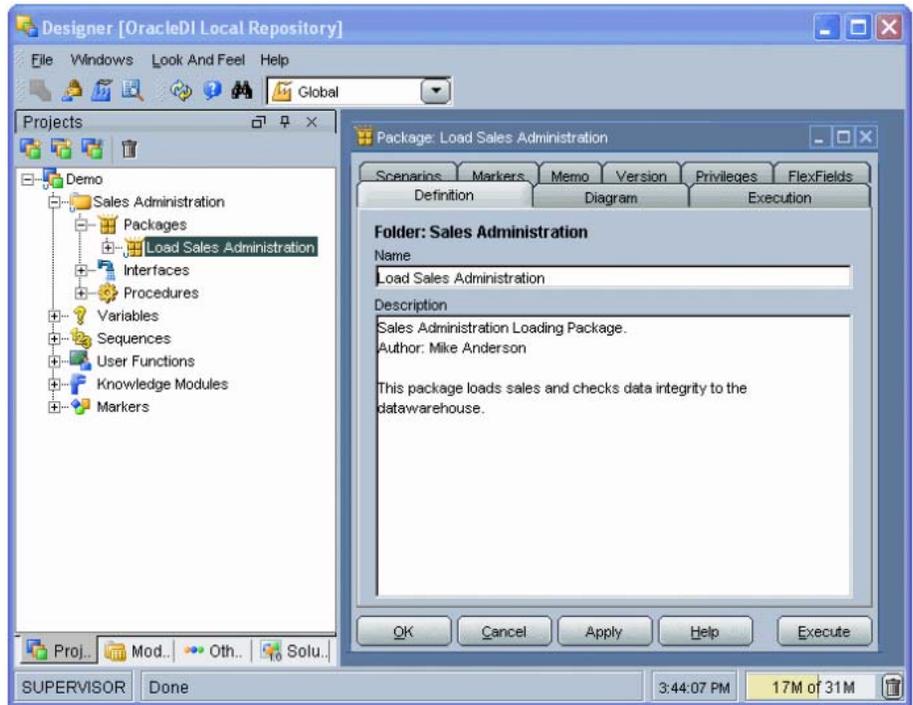
Define las reglas declarativas de integridad y de transformación de datos.

Todo el desarrollo de los proyectos es desarrollado en este módulo, siendo aquí donde los metadatos de las aplicaciones y bases de datos son importados y definidos. El módulo de Designer utiliza metadatos y reglas para generar los escenarios de producción, representando el principal módulo para los desarrolladores y administradores de los metadatos.

Las principales funciones desarrolladas en este módulo son:

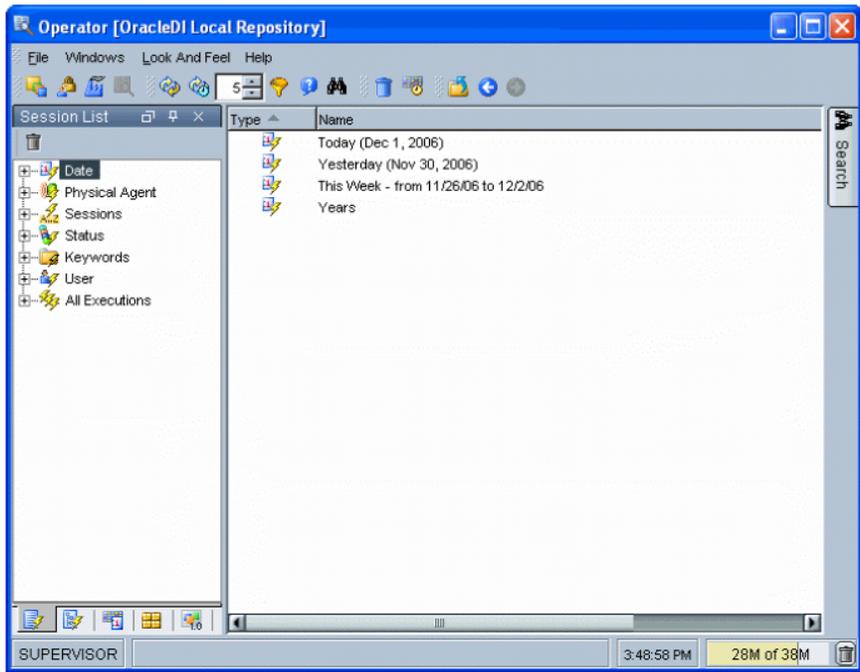
- Ingeniería reversa automática de aplicaciones y bases de datos

- Desarrollo gráfico y mantenimiento de las interfaces de transformación e integración
- Visualización de los flujos de datos de las interfaces
- Administración automática de la documentación
- Customización de código generado



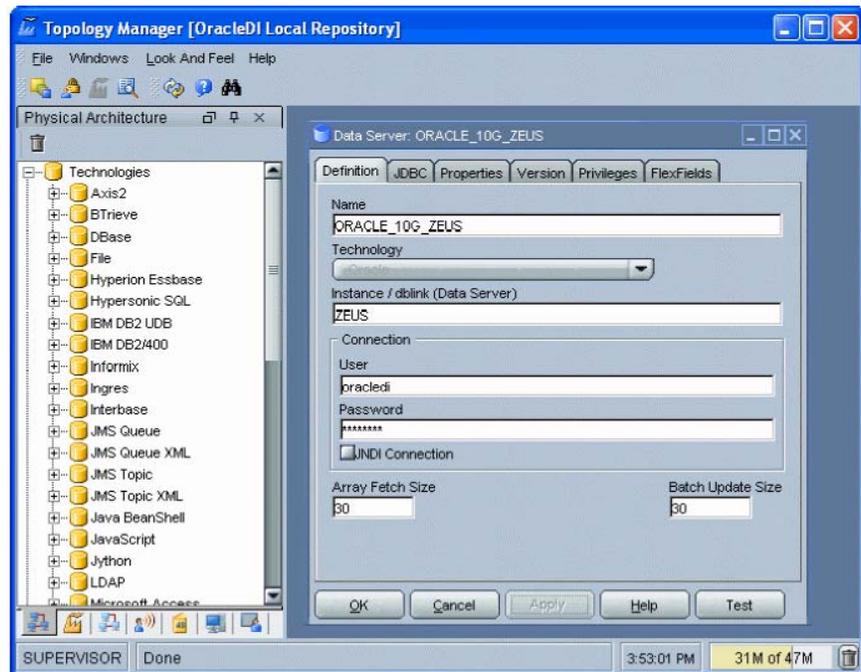
## Operator

Administra y monitorea la producción, siendo destinado para los operadores de producción y presenta los logs de ejecución, con mensajes de error, números de filas procesadas, estadísticas de ejecución, y otros. En tiempo de desarrollo, los desarrolladores pueden también utilizar el Operator con la finalidad de hacer depuración de los procesos.



## Topology Manager

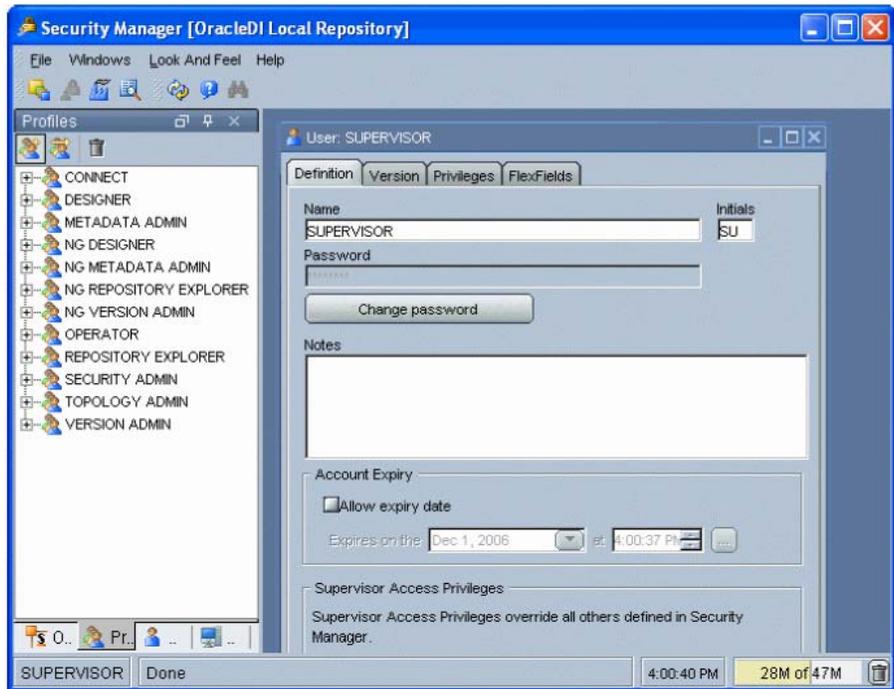
Define la arquitectura física y lógica de la infraestructura. Servidores, esquemas y agentes son registrados en el repositorio maestro a través de este módulo, normalmente por administradores de la infraestructura o del proyecto.



## Security Manager

Administra los usuarios, perfiles y privilegios de acceso. Puede también dar privilegios específicos a objetos y funcionalidades del ambiente ODI, almacenando estas informaciones en el repositorio maestro. Este módulo es normalmente utilizado por administradores de seguridad.

Todos los módulos tienen acceso al repositorio maestro.



## RUNTIME COMPONENTS

En tiempo de ejecución, el agente coordina la ejecución de los escenarios de producción. El agente puede ser instalado en cualquier plataforma que soporte una Java Virtual Machine (J2SE), incluyendo Windows, Linux, HP-UX, Solaris, IBM AIX, iSeries/AS400 y zSeries/OS/390.

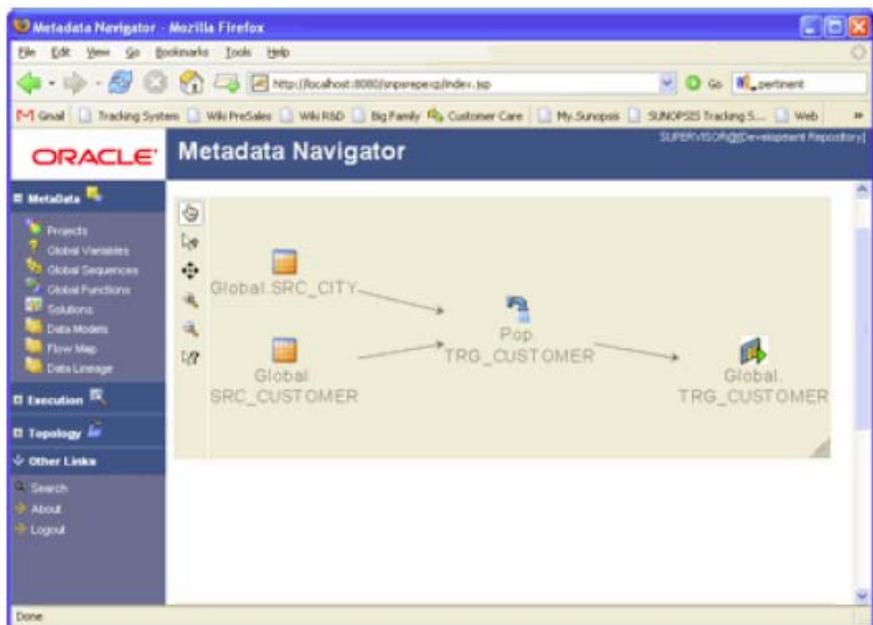
La ejecución de procesos puede ser disparada a partir de uno de los módulos gráficos o también por el agente o desde herramientas de terceros (schedulers).

## El Repositorio

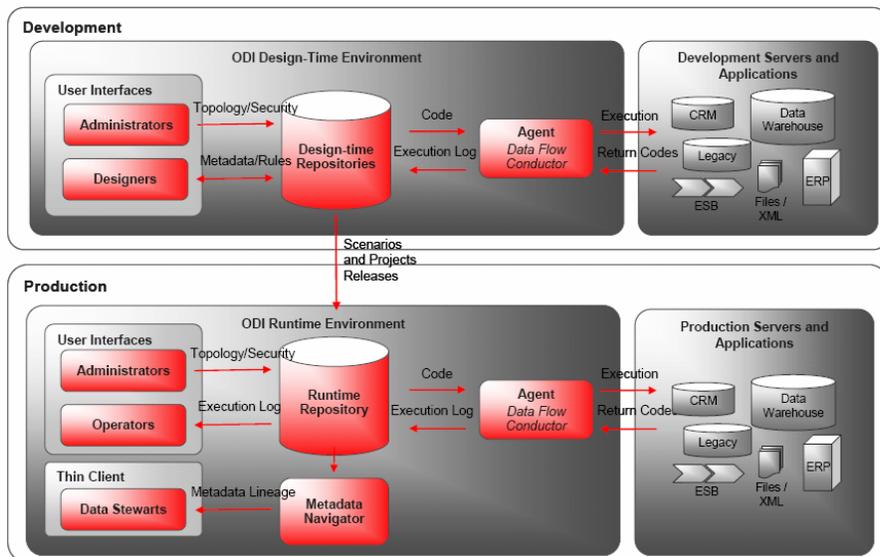
Es un repositorio de metadatos integrado, abierto compartido, que almacena metadatos de las aplicaciones y datos involucrados en las transformaciones y procesos de integración, las versiones de los proyectos desarrollados y todas las informaciones requeridas para el uso de estos (planeamiento, programación y reportes de ejecución).

## Metadata Navigator

Es un cliente web para el acceso al Repositorio de Oracle data Integrator. É un cliente Web, browser based, para acceso ao Oracle Data Integrator Repository. Lista, investiga y analiza todo el contenido del repositorio, además de ofrecer “data lineage tracing”, referencias cruzadas, análisis de impacto, etc. También se puede utilizar esta herramienta para la administración de procesos en ejecución.



Con este escenario de módulos y componentes, la arquitectura operacional de Oracle Data Integrator, en contextos de desarrollo y producción puede ser descrita de acuerdo a la figura de abajo, representando las herramientas, los repositorios con sus integraciones y agentes.





Oracle Data Integrator  
Febrero 2009  
Author: José Villalón

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](http://oracle.com)

Copyright © 2008, Oracle Corporation 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.