



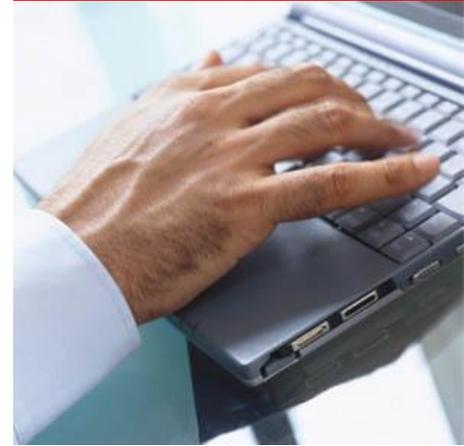
**ORACLE®**

## **Технологии построения хранилищ данных: Oracle Data Integrator**

Родион Константинов

# План презентации

- 4 Ключевых преимущества
- Расширенные возможности
- 6 шагов к внедрению
- Сценарии использования



# Почему интеграция данных?

Информация там и в том виде в каком необходимо

**НУЖНО...**



Business Intelligence



Corporate Performance Management



Business Process Management



Business Activity Monitoring

## Интеграция данных

Migration



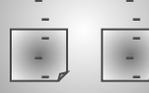
Data Warehousing



Master Data Management



Data Synchronization



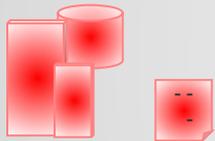
Federation



SOA (Messaging)



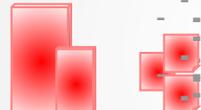
**ИМЕЕМ...** Данные в несогласованных источниках



Legacy



ERP



CRM



Best-of-breed Applications



# Oracle Data Integrator

- Трансформация и перемещение данных из множественных источников в гетерогенной среде

## Преимущества

1. *Скорость:* Архитектура “E-LT”
2. *Гибкость:* Активная интеграция
3. *Продуктивность:* Декларативный дизайн
4. *Расширяемость:* Модули знаний

## Отличительные возможности

# Е-LT архитектура

Высокая производительность

## Специализированный ETL сервер

- Черный ящик
- Слабая производительность
- Высокая цена

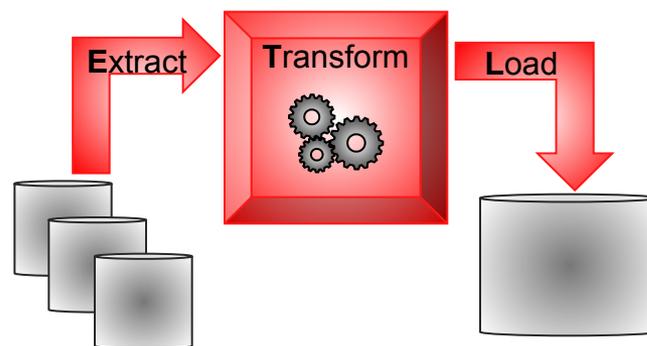
## Преобразования внутри СУБД

- Использование современных возможностей
- Эффективность
- Высокая производительность

## Преимущества

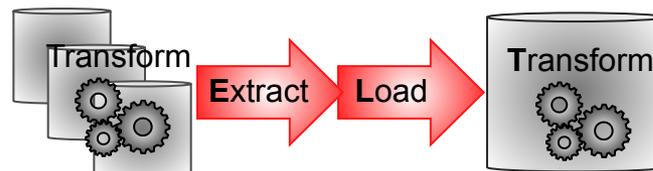
- ✓ Оптимальная производительность и масштабируемость
- ✓ Легкость управления и низкая цена

### «Обычная» ETL Архитектура

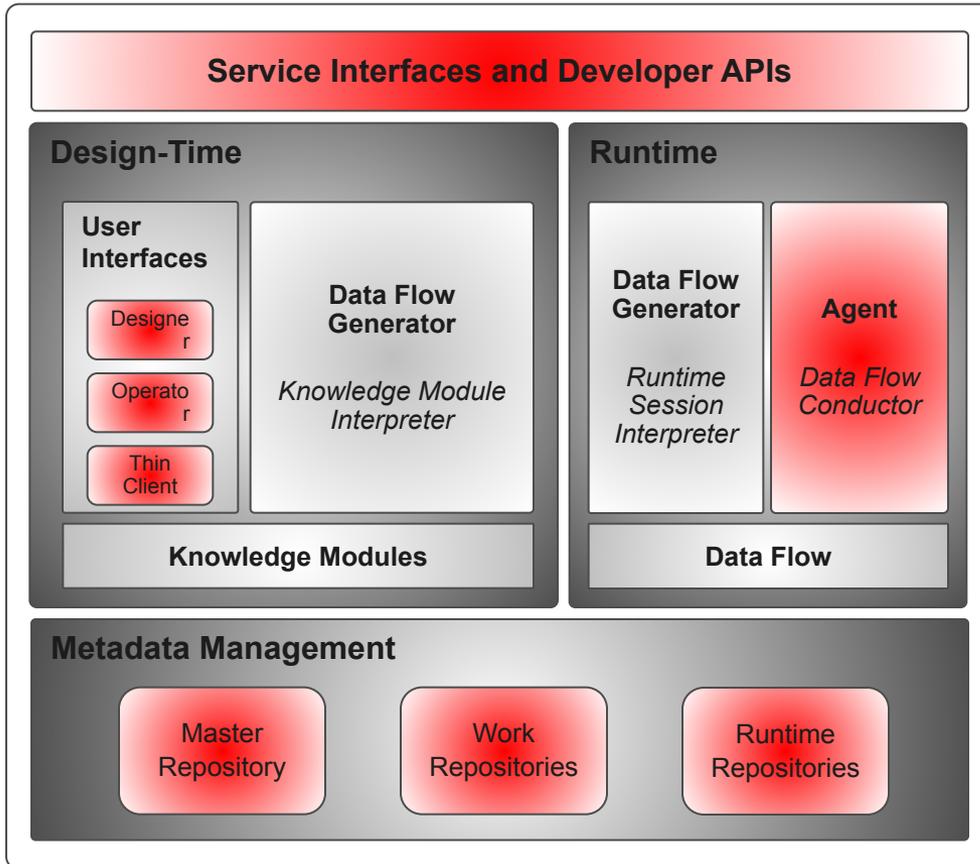


### Архитектура нового поколения

“E-LT”



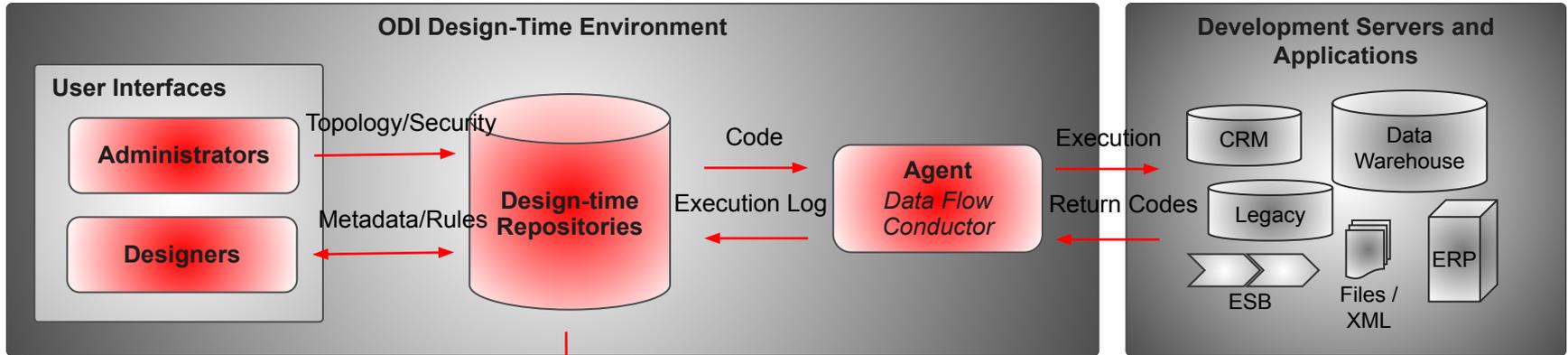
# Архитектура: Концепция



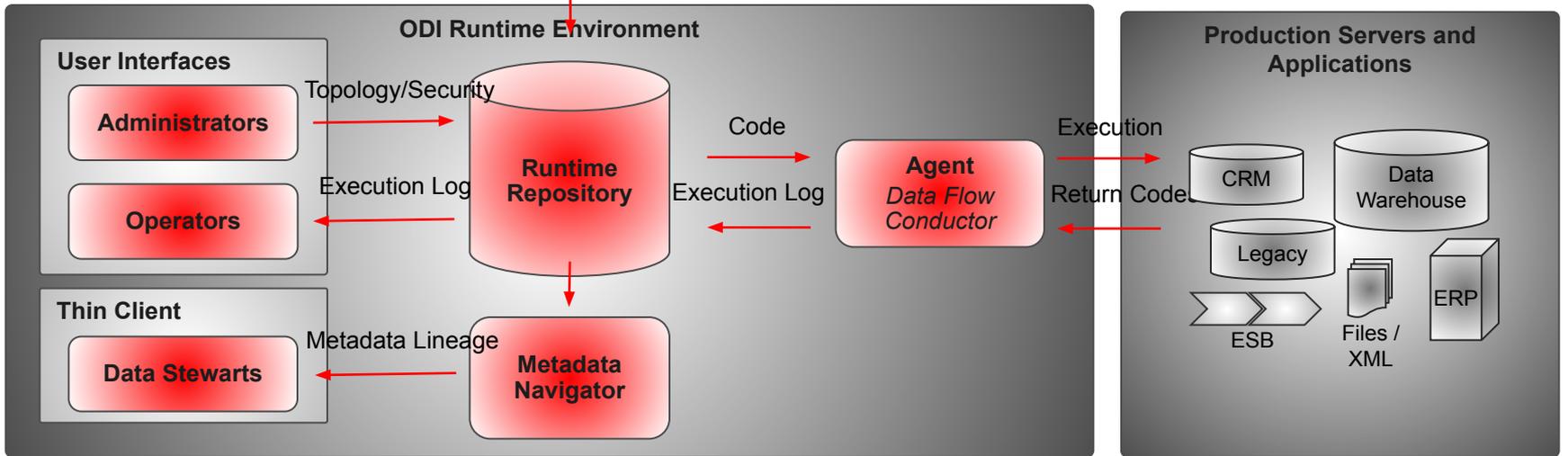
- Java design-time environment
  - Работает на любых платформах
  - Тонкий клиент для просмотра метаданных
- Java runtime environment
  - Работает на любых платформах
  - Управляет выполнением потока задач
- Metadata repository
  - Ставится на любую БД
  - Модульность и расширяемость

# Архитектура: Компоненты

## Development



## Production



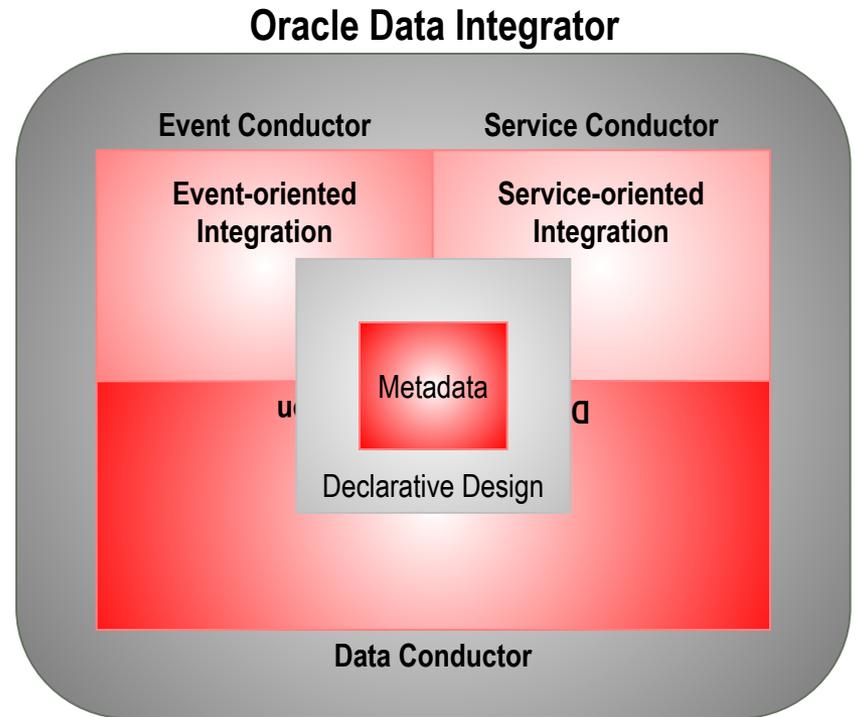
# Активная интеграция

## Событийная и сервис-ориентированная интеграция

- Переход от пакетной загрузки к хранилищам реального времени
- Единообразии инструмента
- Интеграция на лету

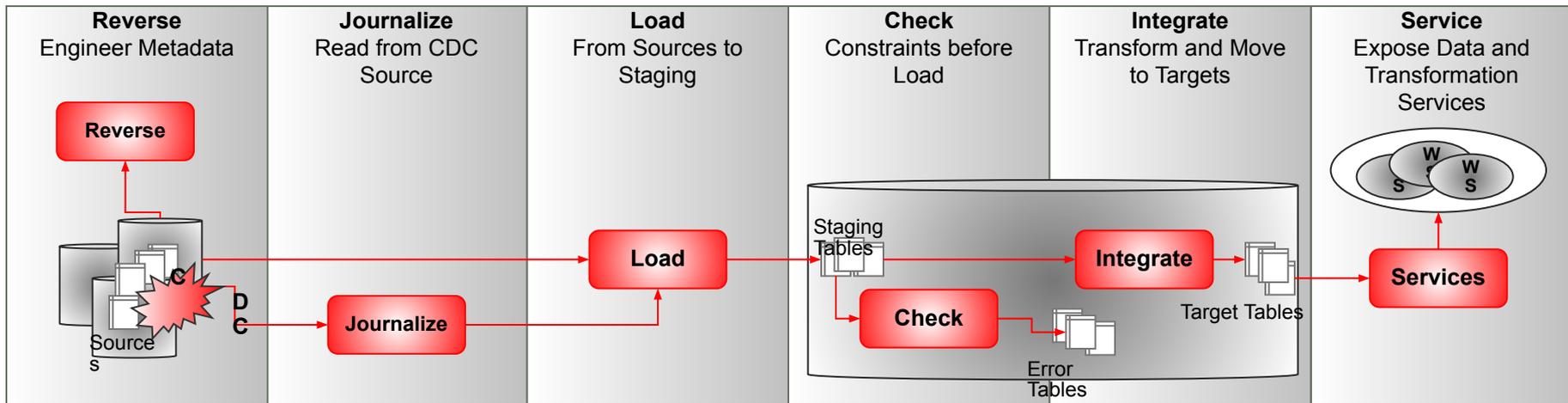
### Преимущества

- ✓ Позволяет построить хранилища реального времени и хабы
- ✓ Можно использовать в SOA

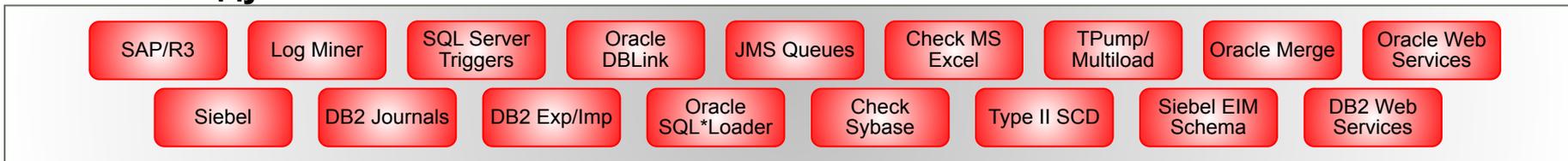


# Модули знаний (Knowledge Modules)

Модульность, гибкость, расширяемость



## Готовые модули знаний

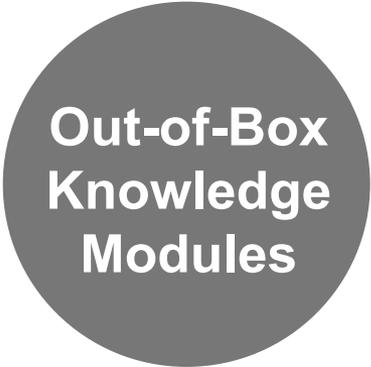


## Преимущества

- ✓ Используется опыт лучших наработок
- ✓ Легко администрировать

# Готовые Knowledge Modules

- Generic SQL DB
- Oracle DB 9i
- Oracle DB 10g
- Oracle DB 10g XE
- IBM DB2/400
- IBM DB2/UDB
- IBM Informix SE
- IBM LDAP Server
- MS SQL Server 2000
- MS SQL Server 2005
- MS SQL Server 2005 SE
- MS Office Access 2000
- MS Office Excel 2000
- MS Active Directory
- Sybase ASA 8.x & 9.x
- Sybase IQ 12.x
- Sonic MQ v7.0
- Teradata V2R5.x
- Teradata V2R6.x
- Netezza Performance Server 2.2.1
- Hyperion Essbase
- PostgreSQL 8.1
- MySQL 4.0
- MySQL 5.0
- Oracle BI Suite 10g
- Oracle BAM 10g
- Oracle Internet Directory 9i
- OpenLDAP 2.3
- Siebel CRM 7.8
- JD Edwards
- PeopleSoft
- SAP R/3
- Oracle EBusiness Suite
- Oracle AQ 10g
- Oracle SOA Suite
- Oracle ESB 10g
- Salesforce.com App Exchange
- Any JMS Standard Implementation



**Out-of-Box  
Knowledge  
Modules**

# Declarative Design

## Продуктивность разработчиков

### Традиционный подход

- Разработчик должен определить каждый шаг сложного ETL процесса
- Требуются специальные ETL навыки
- Существенные усилия по разработке и поддержке

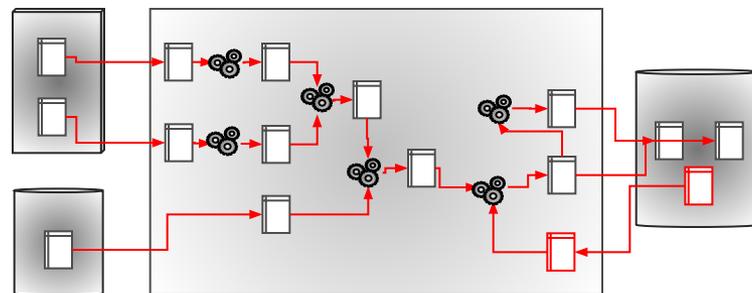
### Declarative Set-based Design

- Уменьшается число шагов
- Автоматическая генерация процедур

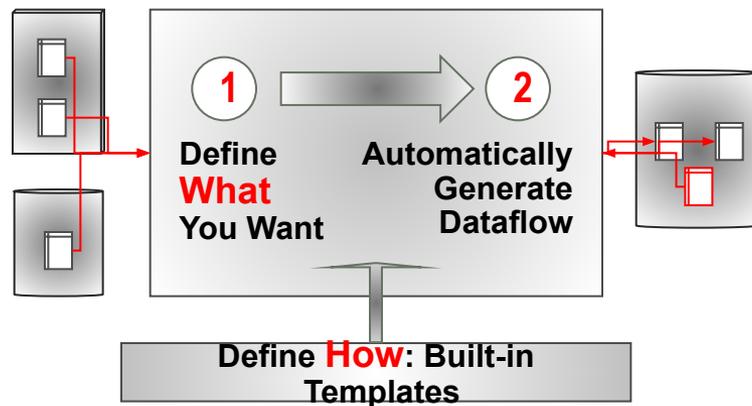
### Преимущества

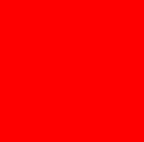
- ✓ Существенное ускорение обучения
- ✓ Уменьшение сроков внедрения
- ✓ Облегчение понимания для не-ИТ персонала

### Традиционный ETL дизайн



### ODI Declarative Design





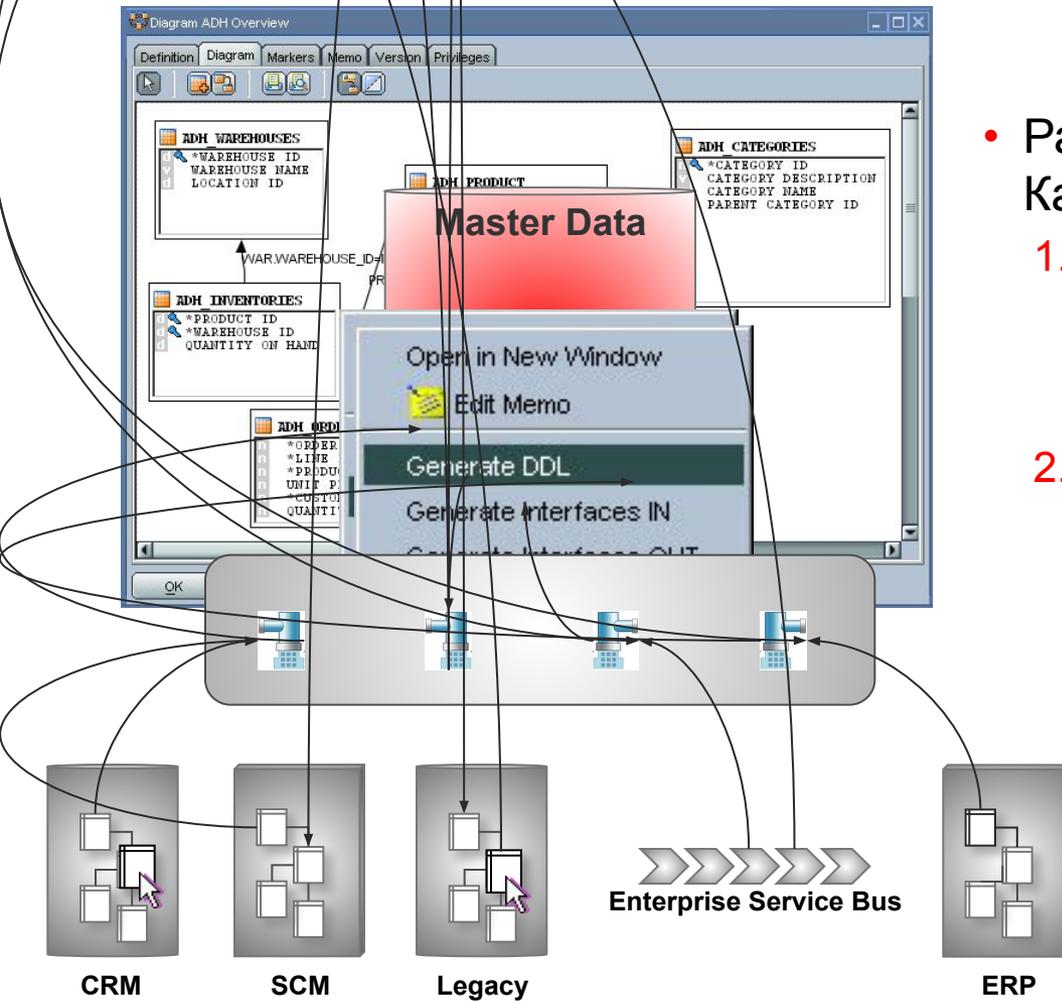
# **Oracle Data Integrator**

## **Расширенные возможности**

# Расширенные возможности

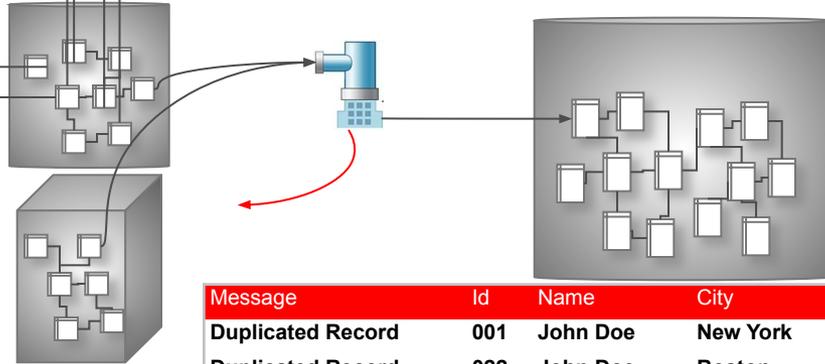
- Master Data Management enabled
  - Создание унифицированных форматов
  - Автоматическое создание канонического формата и трансформаций
  - Встроенная проверка целостности данных
- Real-time enabled
  - Захват изменений
  - Message Oriented Integration (JMS)
- SOA enabled
  - Службы доступа к данным
  - Службы трансформации
- Extensibility
  - Knowledge Modules Framework
  - Языки сценариев
  - Open Tools

# MDM: Canonical Format Design



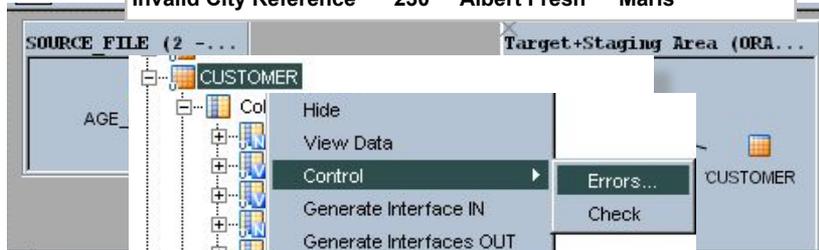
- Разработка и наполнение Канонического формата
  1. Использование имеющихся объектов для построения MDM приложений (entities, fields, relationships)
  2. Создание и поддержка структур эталонных данных

# MDM: Встроенная проверка целостности данных

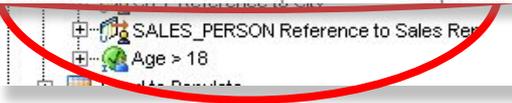


1. Определение ограничений (constraints) на уровне таблиц
2. Дизайн сценариев и проверка целостности
3. Просмотр, очистка или повторная обработка отброшенных записей

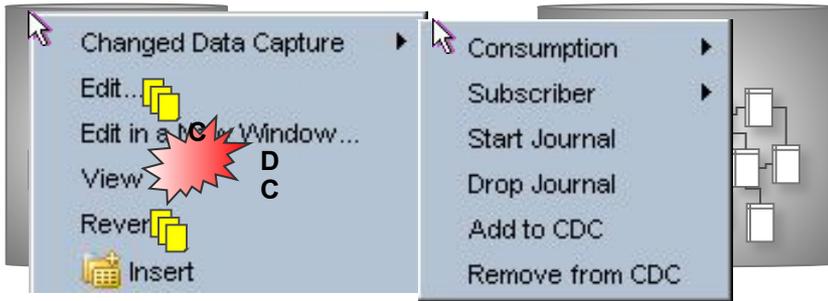
Message	Id	Name	City
Duplicated Record	001	John Doe	New York
Duplicated Record	022	John Doe	Boston
Invalid City Reference	230	Albert Fresh	Maris



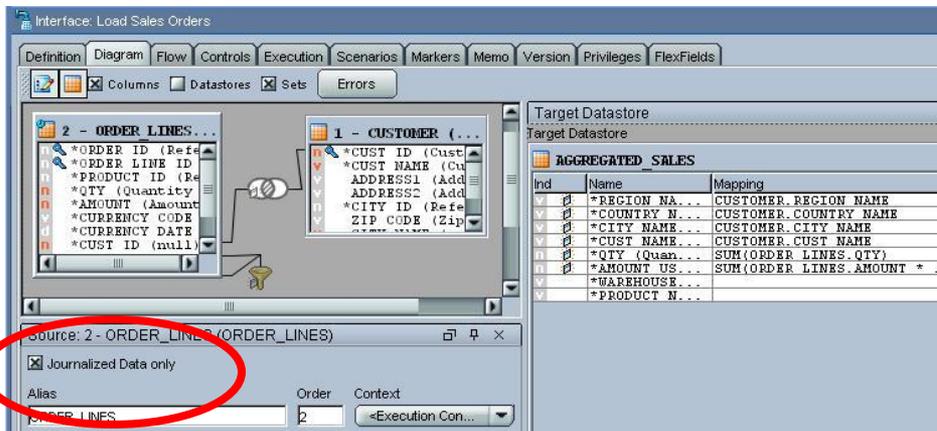
Error Message	Customer Id	First Name	Last Name	Age	City	Sales Rep	Phone	Address	Date Checked
Age should be greater than 18	107	Jacky	Swenson	17	19		(202) 555	64 Imagination Driv	12/23/05
Duplicated Values: "Unique Name and Sales Re	508	Philippe	Michaud	19	23	22	123 456 1	Oakland street 15,	12/23/05
Duplicated Values: "Unique Name and Sales Re	202	Philippe	Michaud	22	23	22	78 21 86 2	197 impasse Renc	12/23/05
Missing Reference (Reference to City) To DEM	103	Peter	Travis	33	342		(510) 555	7835 Hartford Driv	12/23/05
Missing Reference (Reference to Sales Rep) T	105	Tony	Gold	55	19	231	(619) 555	91 Torre drive	12/23/05



# Real-time enabled: Changed Data Capture

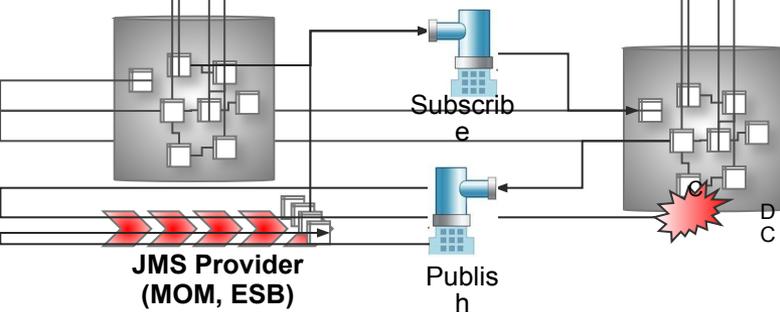


- Публикация и подписка CDC Framework
  - Database logs
  - Triggers
  - Third-tier solutions
- Поддержка целостности транзакций затрагивающих несколько таблиц



1. Создайте или сгенерируйте сценарий
2. Выберите Journalized Data Only
3. Запустите Журналирование

# Real-time enabled: Интеграция основанная на сообщениях



- Подключитесь к провайдеру публикации сообщений и подписки на них (JMS)
- Проверьте, что целостность сообщений не нарушается
- Сложные трансформации
  1. Разрабатывайте сложные сценарии используя очереди, БД и приложения
  2. Используйте JMS Queues как источники и приемники

Interface: Get New Customers from JMS

SS\_0 (-1 - Oracle ESB)

JMS\_CUST

SS\_1 (-1 - ORA MDM)

CITY

Target+Staging Area (ORA...)

CUSTOMER

SS\_0

SS\_1

LKM Selection

LKM: LKM JMS to SQL

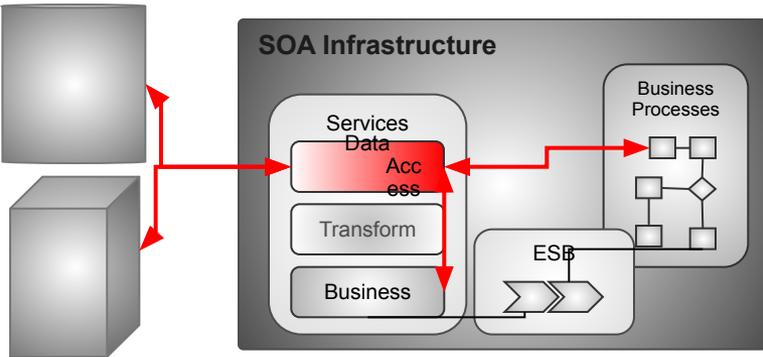
Option	Value
JMS TYPE	<Default>:
CLIENT ID	DWHSubscriber
MESSAGE MAX NUMBER	<Default>: 200
DURABLE	<Default>: D
MESSAGE TIMEOUT	<Default>: 5000

LKM Description

Subscriber identification string.

OK Cancel Apply Help Execute

# SOA Enabled: Data Access Services



- Создавайте и публикуйте службы доступа к данным
  1. Создайте службу доступа к данным
  2. Протестируйте службу
  3. Задействуйте службу доступа к данным в вашей SOA инфраструктуре

The screenshot shows the Oracle SOA Suite configuration and testing environment. On the left is the 'Models' tree with 'ORACLE\_CUSTOMERS' selected. The main window is the 'Model: ORACLE\_CUSTOMERS' configuration dialog, showing the 'Definition' tab. The 'Data Services' section is active, displaying the 'Service KM' tab with a table of data sources. A red circle highlights the 'Generate and deploy...' button. To the right is the 'Test Web Service' window, showing the 'CDCQuery' test results in a table format.

Name	Datastore name	Data Service name	Published entity
<input checked="" type="checkbox"/>	ACCOUNTS	WSAccounts	Accounts
<input checked="" type="checkbox"/>	CITY	WSCity	City
<input checked="" type="checkbox"/>	COUNTRY	WSCountry	Country
<input checked="" type="checkbox"/>	CUSTOMER	WSCustomer	Customer
<input type="checkbox"/>	ORDERS	WSOrder	Order
<input checked="" type="checkbox"/>	REGION	WSRegion	Region

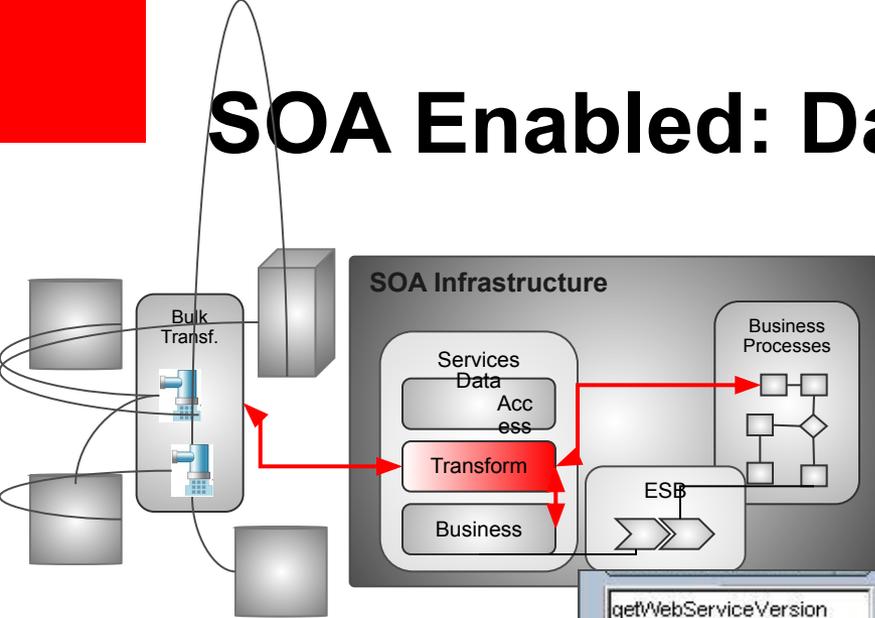
  

Name	Value	Type	Attr
CDCQuery		CDCQuery	
Subscrib...	Data Warehouse	string	

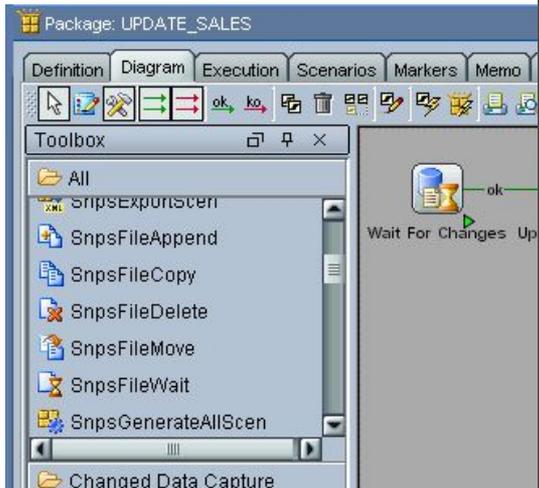
Name	Value	Attribute
CDCCustomer	<JrnDate>2006-11-...	
Customer	<CUST_ID>14567...	
JrnDate	2006-11-23T15:3...	
JrnFlag	1	
CUST_ID	1456778	
FIRST_NAME	Mark	
LAST_NAME	Adams	
AGE	43	
PHONE_NU...	650 765 8954	
ADDRESS	43 King's Court	
CITY_ID	50	
SALES_ID		xsi:nil=true xmlns...
SPONSOR_ID		xsi:nil=true xmlns...
AGE_RANGE	20-50	

# SOA Enabled: Data Flow Services



- Представьте сценарии как Web Services

1. Манипулируйте сценариями
2. Публикуйте сценарии как web-сервисы в вашей SOA инфраструктуре



Name	Value	Type	Attribute																																																																
getWebServiceVersion																																																																			
invokeScenario																																																																			
RepositoryConnection		Repositor...																																																																	
<table border="1"> <thead> <tr> <th>Name</th> <th>Value</th> <th>Type</th> <th>Attribu</th> </tr> </thead> <tbody> <tr> <td>OdiUser</td> <td>SUPERVISOR</td> <td>string</td> <td></td> </tr> <tr> <td>OdiPass...</td> <td>*****</td> <td>string</td> <td></td> </tr> </tbody> </table>				Name	Value	Type	Attribu	OdiUser	SUPERVISOR	string		OdiPass...	*****	string																																																					
Name	Value	Type	Attribu																																																																
OdiUser	SUPERVISOR	string																																																																	
OdiPass...	*****	string																																																																	
Command		Scenario...																																																																	
<table border="1"> <thead> <tr> <th>Name</th> <th>Value</th> <th>Type</th> <th>Attribu</th> </tr> </thead> <tbody> <tr> <td>ScenName</td> <td>UPDATE_SALES</td> <td>string</td> <td></td> </tr> <tr> <td>ScenVer...</td> <td>001</td> <td>string</td> <td></td> </tr> <tr> <td>Context</td> <td>PRODUCTION</td> <td>string</td> <td></td> </tr> <tr> <td>Session...</td> <td>Update DW Sales</td> <td>string</td> <td></td> </tr> <tr> <td>Keywords</td> <td>Prd, PTY1, Sales</td> <td>string</td> <td></td> </tr> <tr> <td colspan="2">Variable...</td> <td colspan="2">Variabl...</td> </tr> <tr> <td colspan="4"> <table border="1"> <thead> <tr> <th>Name</th> <th>Value</th> <th>Type</th> <th>Attrib</th> </tr> </thead> <tbody> <tr> <td>Name</td> <td>REFRESH_DATE</td> <td>string</td> <td></td> </tr> <tr> <td>Value</td> <td>2007-10-18</td> <td>string</td> <td></td> </tr> </tbody> </table> </td> </tr> <tr> <td colspan="2">Agent</td> <td colspan="2">AgentType</td> </tr> <tr> <td colspan="4"> <table border="1"> <thead> <tr> <th>Name</th> <th>Value</th> <th>Type</th> <th>Attribu</th> </tr> </thead> <tbody> <tr> <td>Host</td> <td>coruscant</td> <td>string</td> <td></td> </tr> <tr> <td>Port</td> <td>20040...</td> <td></td> <td></td> </tr> </tbody> </table> </td> </tr> </tbody> </table>				Name	Value	Type	Attribu	ScenName	UPDATE_SALES	string		ScenVer...	001	string		Context	PRODUCTION	string		Session...	Update DW Sales	string		Keywords	Prd, PTY1, Sales	string		Variable...		Variabl...		<table border="1"> <thead> <tr> <th>Name</th> <th>Value</th> <th>Type</th> <th>Attrib</th> </tr> </thead> <tbody> <tr> <td>Name</td> <td>REFRESH_DATE</td> <td>string</td> <td></td> </tr> <tr> <td>Value</td> <td>2007-10-18</td> <td>string</td> <td></td> </tr> </tbody> </table>				Name	Value	Type	Attrib	Name	REFRESH_DATE	string		Value	2007-10-18	string		Agent		AgentType		<table border="1"> <thead> <tr> <th>Name</th> <th>Value</th> <th>Type</th> <th>Attribu</th> </tr> </thead> <tbody> <tr> <td>Host</td> <td>coruscant</td> <td>string</td> <td></td> </tr> <tr> <td>Port</td> <td>20040...</td> <td></td> <td></td> </tr> </tbody> </table>				Name	Value	Type	Attribu	Host	coruscant	string		Port	20040...		
Name	Value	Type	Attribu																																																																
ScenName	UPDATE_SALES	string																																																																	
ScenVer...	001	string																																																																	
Context	PRODUCTION	string																																																																	
Session...	Update DW Sales	string																																																																	
Keywords	Prd, PTY1, Sales	string																																																																	
Variable...		Variabl...																																																																	
<table border="1"> <thead> <tr> <th>Name</th> <th>Value</th> <th>Type</th> <th>Attrib</th> </tr> </thead> <tbody> <tr> <td>Name</td> <td>REFRESH_DATE</td> <td>string</td> <td></td> </tr> <tr> <td>Value</td> <td>2007-10-18</td> <td>string</td> <td></td> </tr> </tbody> </table>				Name	Value	Type	Attrib	Name	REFRESH_DATE	string		Value	2007-10-18	string																																																					
Name	Value	Type	Attrib																																																																
Name	REFRESH_DATE	string																																																																	
Value	2007-10-18	string																																																																	
Agent		AgentType																																																																	
<table border="1"> <thead> <tr> <th>Name</th> <th>Value</th> <th>Type</th> <th>Attribu</th> </tr> </thead> <tbody> <tr> <td>Host</td> <td>coruscant</td> <td>string</td> <td></td> </tr> <tr> <td>Port</td> <td>20040...</td> <td></td> <td></td> </tr> </tbody> </table>				Name	Value	Type	Attribu	Host	coruscant	string		Port	20040...																																																						
Name	Value	Type	Attribu																																																																
Host	coruscant	string																																																																	
Port	20040...																																																																		



# Scripting Framework

Procedure: Sample Scripting

Ord...	Command	Context	Logical ...	Transaction	Commit	Ignore Errors	Log
0	Create Java Class					<input type="checkbox"/>	
10	Run Java Main Clas					<input type="checkbox"/>	
20	Process Socket Inp					<input type="checkbox"/>	
30	Send HTTP Reques					<input type="checkbox"/>	

Command: Process Socket Input in Jython

Name: Process Socket Input in Jython

Log Counter: <Undefined> Log Level: 3 Ignore Errors:

Command on Target: Command on Source

Technology: Jython Transaction Isolation: <Undefined>

Context: <Execution Context> Schema: <Undefined>

Transaction: <Undefined> Commit: <Undefined>

Command:

```
###
# Jython Script: Listen to TCP/IP socket on port 8081
###
import socket
import time

port=8081
s = socket.socket(socket.AF_INET, socket.SOCK_DGRAM)
s.bind(("", port))
print "Waiting for incoming messages..."
time.sleep(10)
while 1:
    data, addr = s.recvfrom(1024)
    print "Msg Received from %s->%s" % (addr, data)
    print "Sending Reply..."
    s.sendto("OK", addr)
```

- Используйте языки сценариев
- Применяйте языки встроенные в базы данных
  - SQL, PL/SQL, Transact SQL, etc.
- Используйте OS shell скрипты
  - Win32 DOS, sh, ksh, csh, OS400 commands, JCL, etc.
- Используйте языки совместимые с Bean Scripting Framework Java, JavaScript, Jython (Java Python), Perl, etc.





# **Oracle Data Integrator**

## **6 шагов к внедрению**

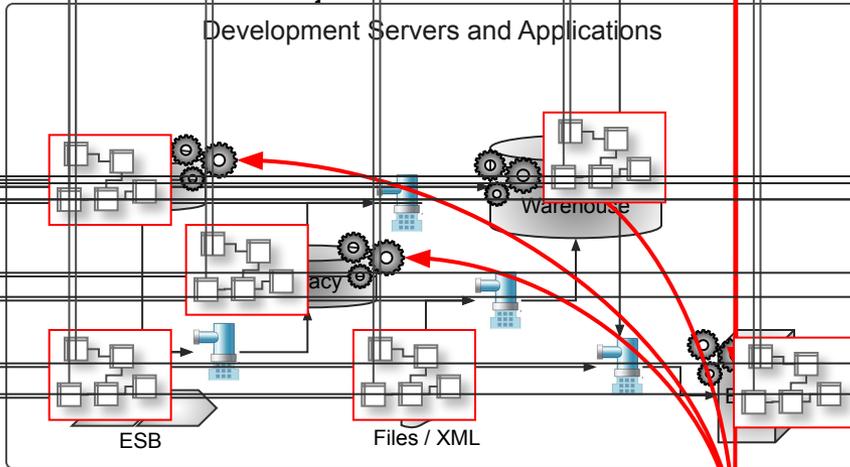
# 6 шагов к внедрению

1. Получите/Создайте метаданные
2. Разработайте трансформации
3. Настройте сценарии выполнения

4. Сгенерируйте сценарии
5. Мониторинг выполнения
6. Проанализируйте влияние

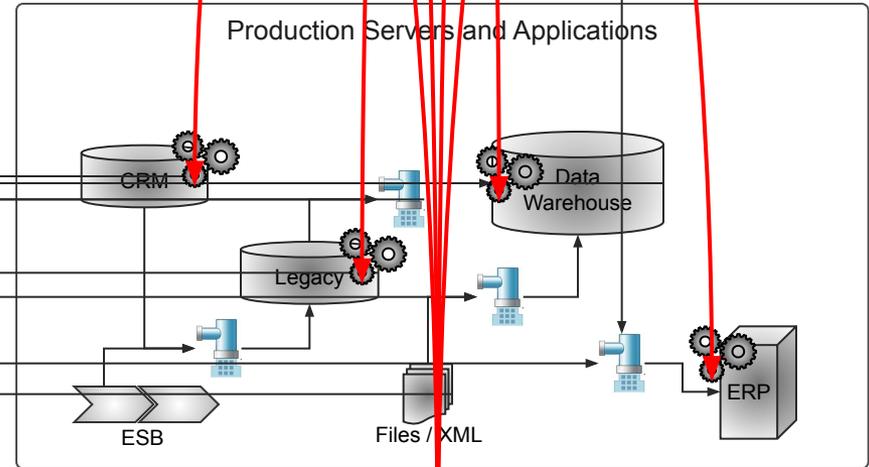
## Development

Development Servers and Applications

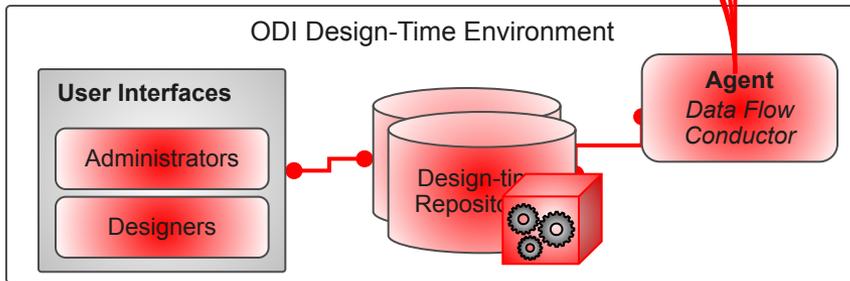


## Productio

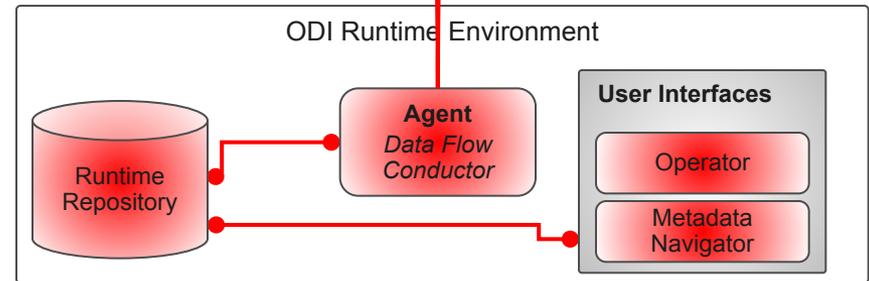
Production Servers and Applications



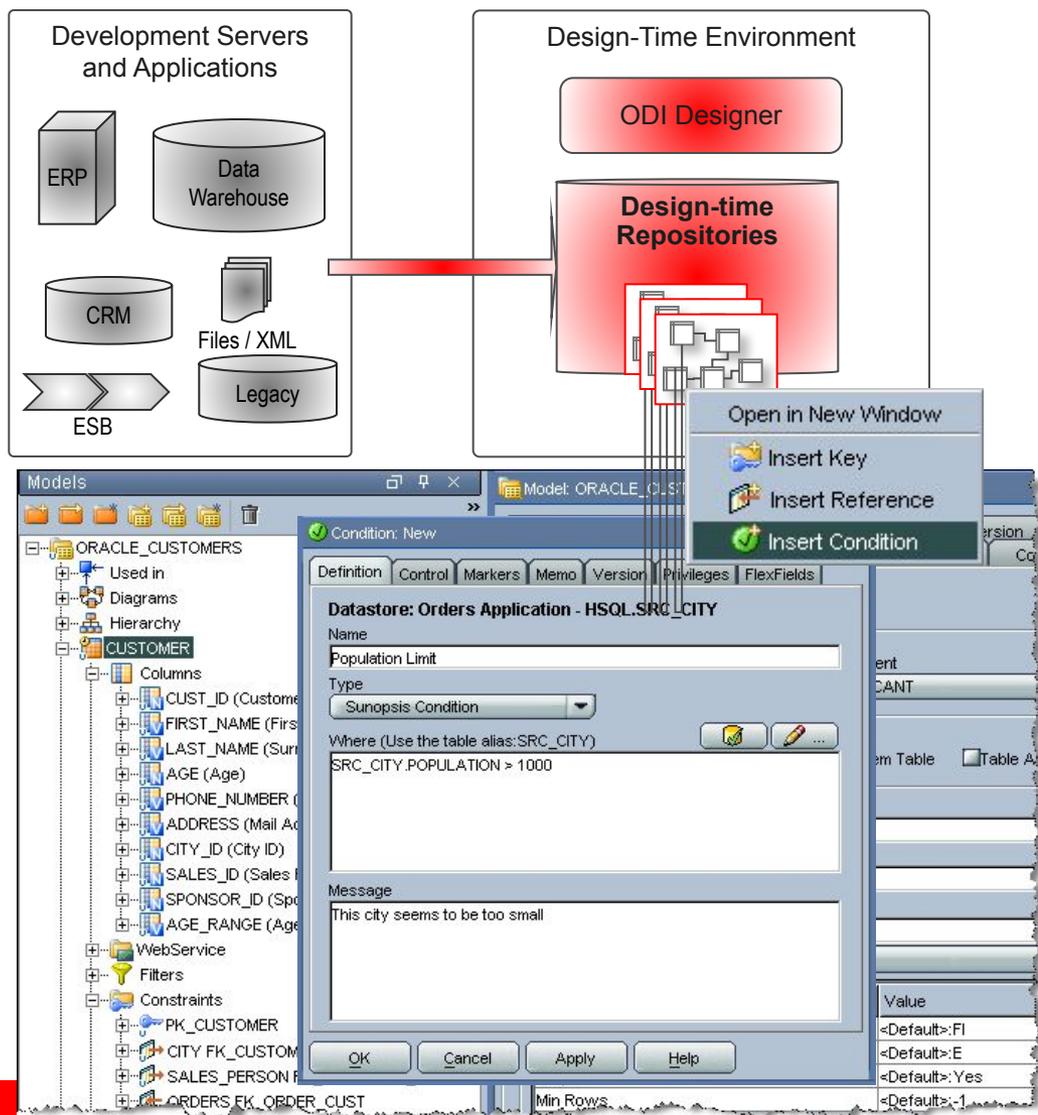
ODI Design-Time Environment



ODI Runtime Environment



# Получите/Создайте метаданные



1. Reverse-engineer
  - Автоматическая
  - Настраиваемая
  - 40+ технологий
2. Создание метаданных
  - Документирование
  - Правила для целостности данных

2

# Разработайте трансформации

1

Определите Что Вы хотите

3

Автоматически сгенерируйте трансформацию

The screenshot displays the Oracle Data Integrator (ODI) interface. The main window shows the 'Interface: Populate Customer' with sources and target datatypes. A mapping rule for 'AGE\_RANGE' is shown in the foreground, detailing a business rule for handling null age values. The 'Projects' window on the right lists various knowledge modules, including CKM and IKM modules for different databases and data integration scenarios.

**Mapping: AGE\_RANGE**

Implementation

```

case
  when AGE_FILE.AGE_RANGE is null then
    UPPER(AGE_FILE.AGE_RANGE)
  else
    'Unknown Age Range for:' || CUSTOMER.AGE
end
  
```

Execute on: -Undefined-

Source Staging Area Target

**Projects**

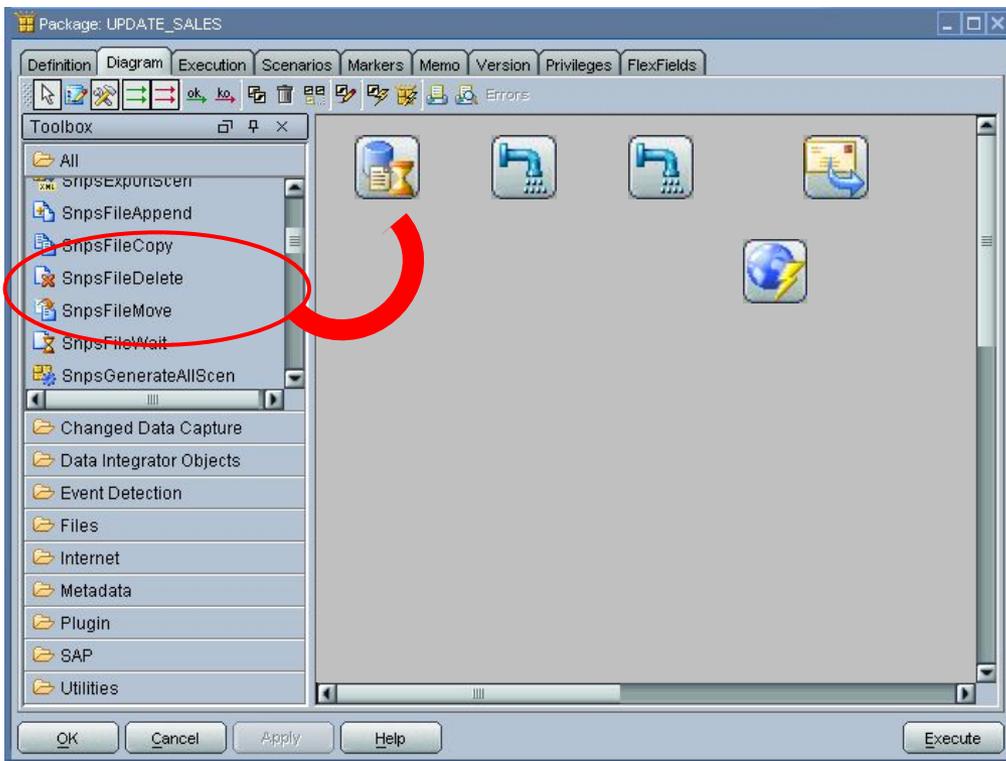
- Knowledge Modules
  - CKM IBM DB2
  - CKM MS SQLServer
  - CKM Oracle
  - CKM SQL
  - CKM Teradata
  - IKM File to Teradata (MultiLoad)
  - IKM Oracle Incremental Update
  - IKM Oracle Incremental Update (MERGE)
  - IKM Oracle Incremental Update (PL SQL)
  - IKM Oracle Slowly Changing Dimension
  - IKM SQL to Teradata (FAST or MULTI LOAD)
  - IKM Teradata Incremental Update
  - IKM Teradata MultiLoad
  - IKM Teradata Replace and Control
  - IKM Teradata Slowly Changing Dimension
  - IKM Teradata to File (FASTEXPORT)
  - JKM DB2 400 Simple
  - JKM DB2/400 - Journal Simple
  - JKM MSSQL Simple

2

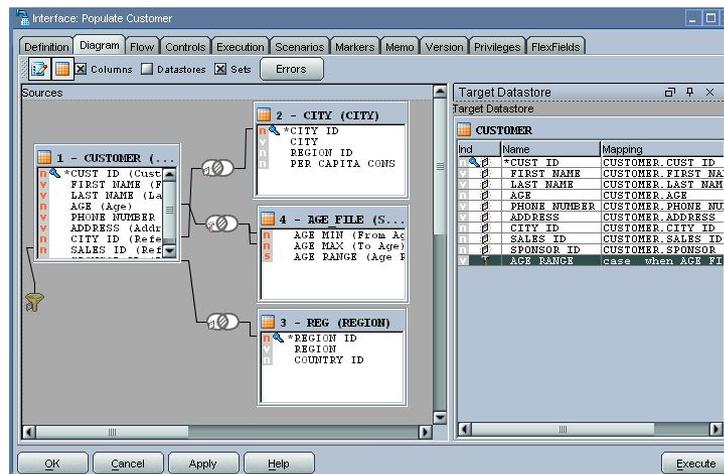
Определите Как делать: Выберите шаблон

Bulk Load • Changed Data Capture • Incremental Update • Slowly Changing Dimension

# Настройте сценарии выполнения



1. Последовательность трансформаций
2. ODI инструменты
  - Data Quality Processes
  - Files/Archives Management
  - Send/Receive Emails
  - Web Services Invokation
  - Event Detection
  - **Создайте собственные инструменты**
3. Используйте
  - Циклы
  - Условия
  - Обработку ошибок



# Сгенерируйте сценарии

1. Скомпилируйте сценарии для выполнения

2. Используйте управление версиями

- Advanced Version Management

3. Перенесите в Production

# Мониторинг выполнения

All Executions

- 2005 - Pop. TRG\_COUNTRY - 2006-02-16 15:01:57.312
  - Variables
  - 1 - Pop. TRG\_COUNTRY - 2006-02-16 15:01:57.406
    - 1 - Loading - SS\_0 - Drop work table
    - 2 - Loading - SS\_0 - Create work table

Session Task: Integration

Definition Description Execution Privileges

Default

insert into I\$\_TRG\_COUNTRY

COUNTRY\_ID,  
COUNTRY,  
IND\_UPDATE

select

Execution

Start	End	Duration
2006-11-17 15:33:12.54	2006-11-17 15:53:13.82	121

Status: Done

Return Code: 0

Message

No. of Rows	127903	No. of Inserts	102843
No. of Updates	25045	No. of Deletes	
No. of Errors	15		

- Просмотр выполнения сессий в реальном времени
- Просмотр кода
- Детальная статистика
- Перезапуск сбойных сессий

261005 - POP TRG\_CITY - 2006-02-15 18:...

260005 - P

259005 - P

258005 - P

257005 - P

256005 - P

255005 - P

254005 - P

Open in New Window

Edit...

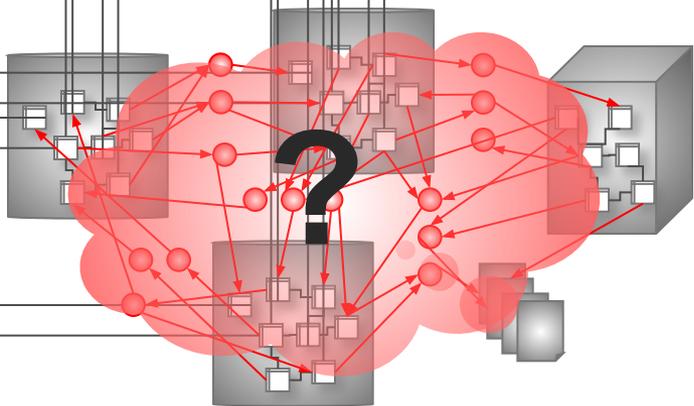
View

Stop

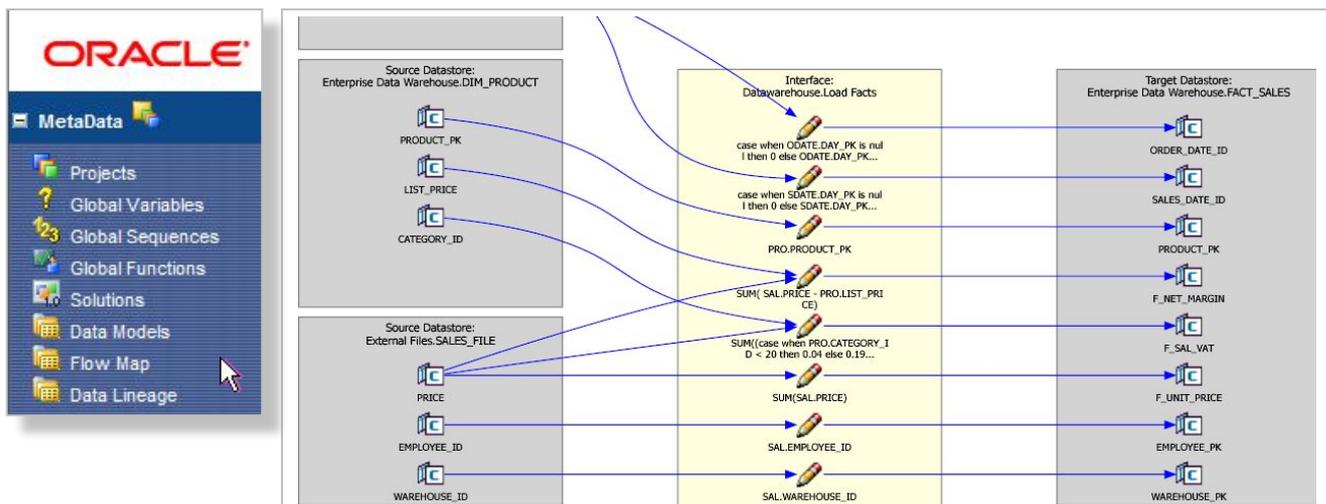
Restart

Delete

# Анализируйте влияния/зависимости



- Поддержка большого числа сценариев в сложных проектах
- Основанный на Web инструмент просмотра
  1. Отследите ваши потоки данных
  2. Следуйте по пути данных
  3. Drill-down к трансформациям



# Сценарии использования

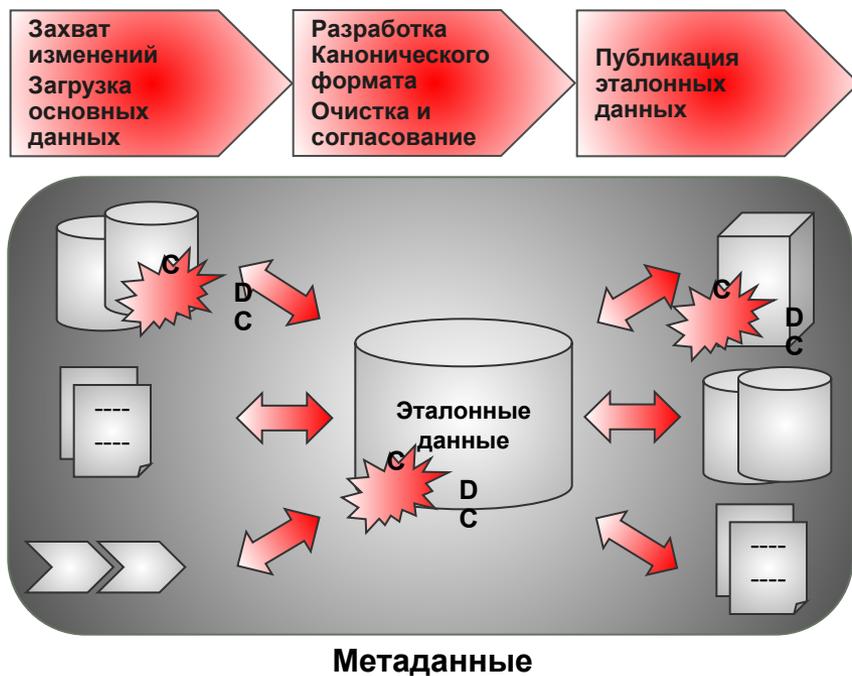
# E-LT для хранилищ данных



- ✓ Гетерогенные источники и цели
- ✓ Инкрементальная загрузка
- ✓ Медленно меняющиеся размерности
- ✓ Целостность данных
- ✓ Захват изменений (CDC)
- ✓ Анализ влияния

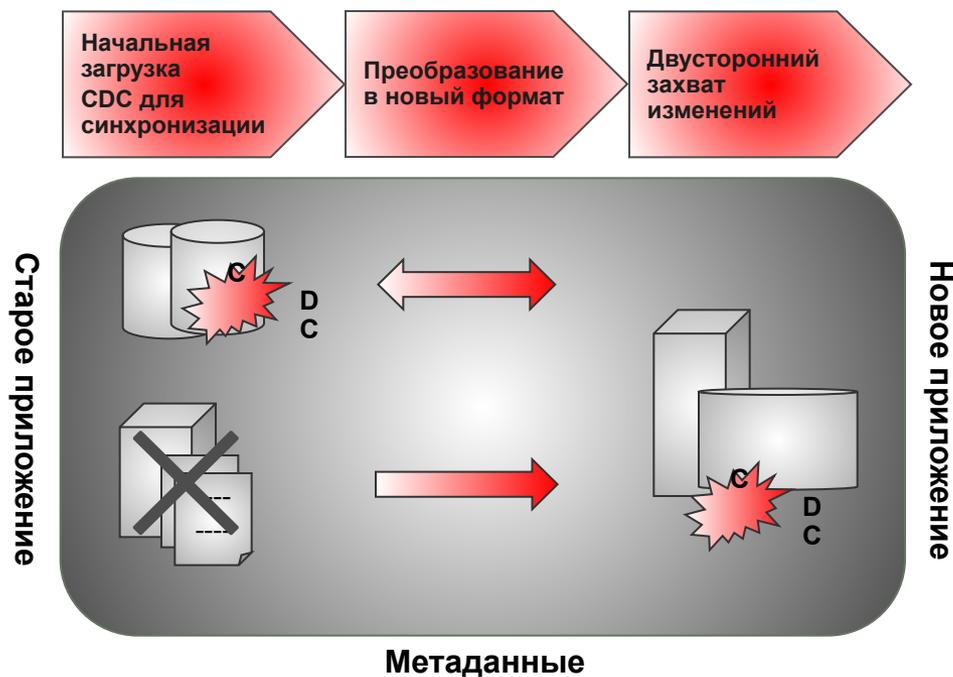
# Master Data Management

(Управление основными данными)



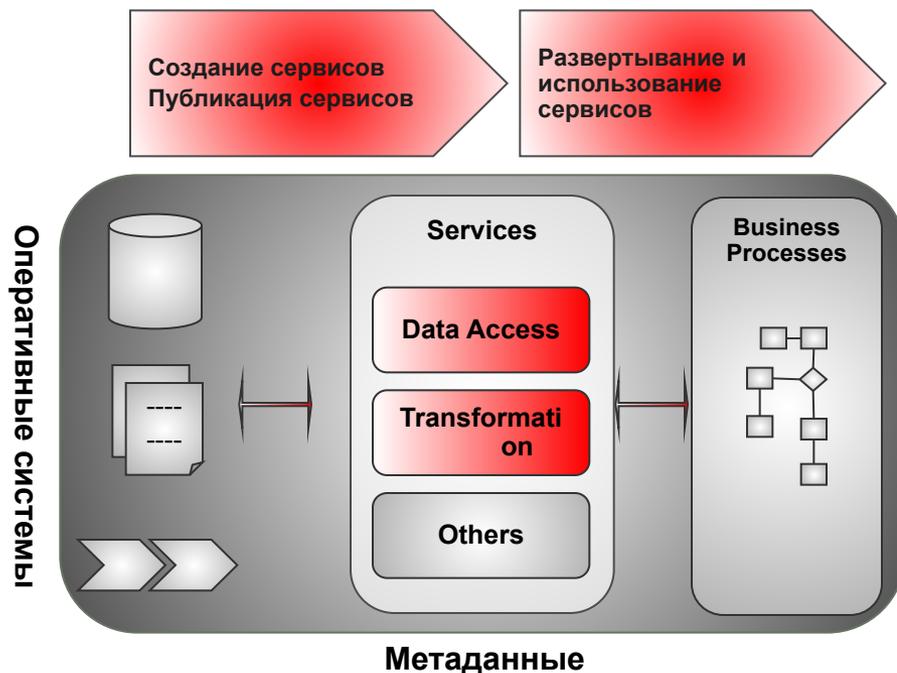
- ✓ Используется как инфраструктура для создания собственного хаба
- ✓ Описания потоков данных
- ✓ Захват изменений (CDC)
- ✓ Очистка и согласование данных
- ✓ Публикация эталонных данных
- ✓ Расширение метаданных

# Миграция



- ✓ Массовая перегрузка данных в новое приложение
- ✓ Преобразование форматов
- ✓ Синхронизация старого и нового приложения в период совместного использования
- ✓ Двусторонний захват изменений (CDC)

# Использование в SOA



- ✓ Вызов внешних сервисов для интеграции
- ✓ Создание сервисов
- ✓ Публикация сервисов

# Oracle Data Integrator: Roadmap

- Focus Areas for Next Major Release
  - Deep Integration with Fusion Middleware
    - Runtime, Design time, Security, Administration, Events
  - Functional Integration with Oracle Warehouse Builder
    - Runtime Integration, Metadata Sharing, Knowledge Module Sharing
  - Deployment of ODI for Embedded Data Integration
    - OracleBI Enterprise Edition, Data Hubs, Application Migrations
  - Enhanced Usability and Debuggability
    - Wizards, New Views, User-definable Debugging
  - Improved Support for Native Oracle Database Features
    - Oracle OLAP

# Заказчики





ORA



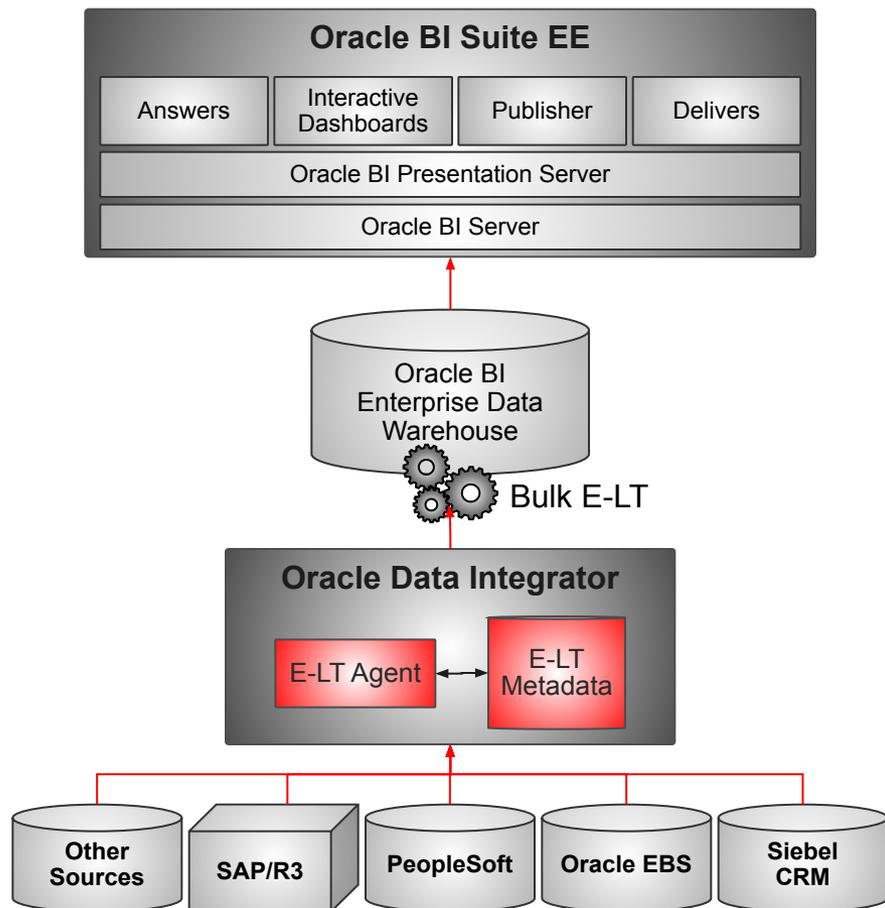
**ORACLE IS THE INFORMATION COMPANY**

**ORACLE®**

# Преимущества Oracle Data Integrator

- ✓ «E-LT» архитектура
- ✓ Активная интеграция (Active Integration)
- ✓ Модули знаний (Knowledge Modules)
- ✓ Подход к проектированию основанный на бизнес правилах (Declarative Design)

# ODI расширяет возможности Oracle BI



## Oracle Business Intelligence Suite EE:

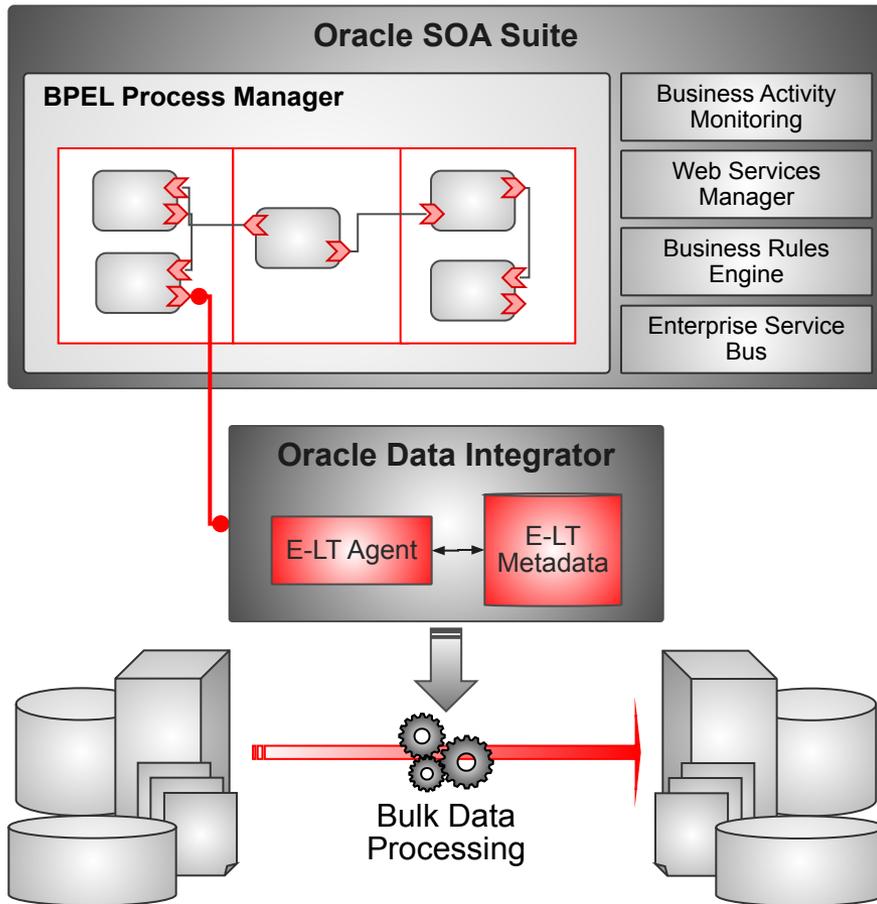
- ✓ Бизнес модель
- ✓ Обработка сложных вычислений
- ✓ Интеллектуальная генерация запросов
- ✓ Оптимизация доступа к данным

## Oracle Data Integrator:

- ✓ Наполнение корпоративного хранилища данных
- ✓ Оптимизация производительности ETL процессов

# ODI Enhances Oracle SOA Suite

## Add Bulk Data Transformation to BPEL Process



### Oracle SOA Suite:

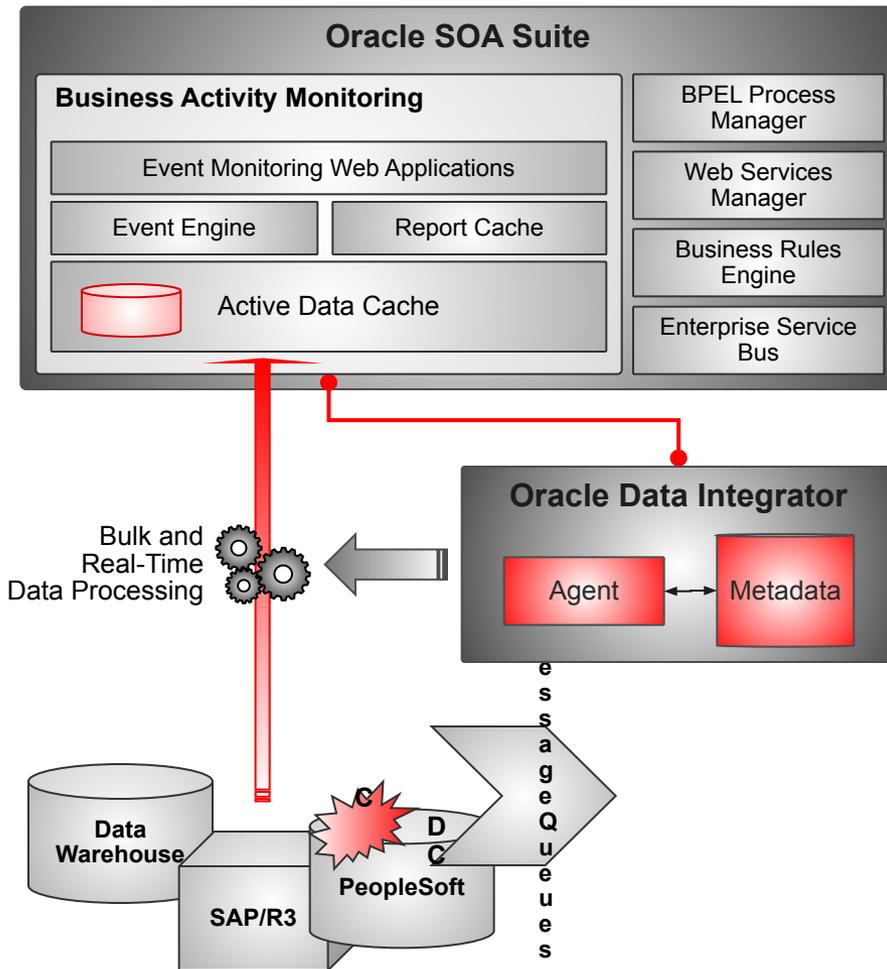
- ✓ *BPEL Process Manager* for Business Process Orchestration

### Oracle Data Integrator:

- ✓ Efficient Bulk Data Processing as Part of Business Process
- ✓ Interact via Data Services and Transformation Services

# ODI Enhances Oracle SOA Suite

## Populate BAM Active Data Cache Efficiently



Oracle SOA Suite:

- ✓ *Business Activity Monitoring* for Real-time Insight

Oracle Data Integrator:

- ✓ High Performance Loading of BAM's Active Data Cache
- ✓ Pre-built and Integrated