



ORACLE

# Oracle Modern Cloud Day 2019

## Oracle Data Integrator

Интеграция больших данных от локальных систем до облака

**Natalia Kusova**

Solution Engineering

## Safe Harbor

---

The following is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. 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, timing, and pricing of any features or functionality described for Oracle's products may change and remains at the sole discretion of Oracle Corporation.

Statements in this presentation relating to Oracle's future plans, expectations, beliefs, intentions and prospects are "forward-looking statements" and are subject to material risks and uncertainties. A detailed discussion of these factors and other risks that affect our business is contained in Oracle's Securities and Exchange Commission (SEC) filings, including our most recent reports on Form 10-K and Form 10-Q under the heading "Risk Factors." These filings are available on the SEC's website or on Oracle's website at <http://www.oracle.com/investor>. All information in this presentation is current as of September 2019 and Oracle undertakes no duty to update any statement in light of new information or future events.

# Содержание

---

Обзор интеграционной платформы

ODI для работы с Big Data

ODI в облаке

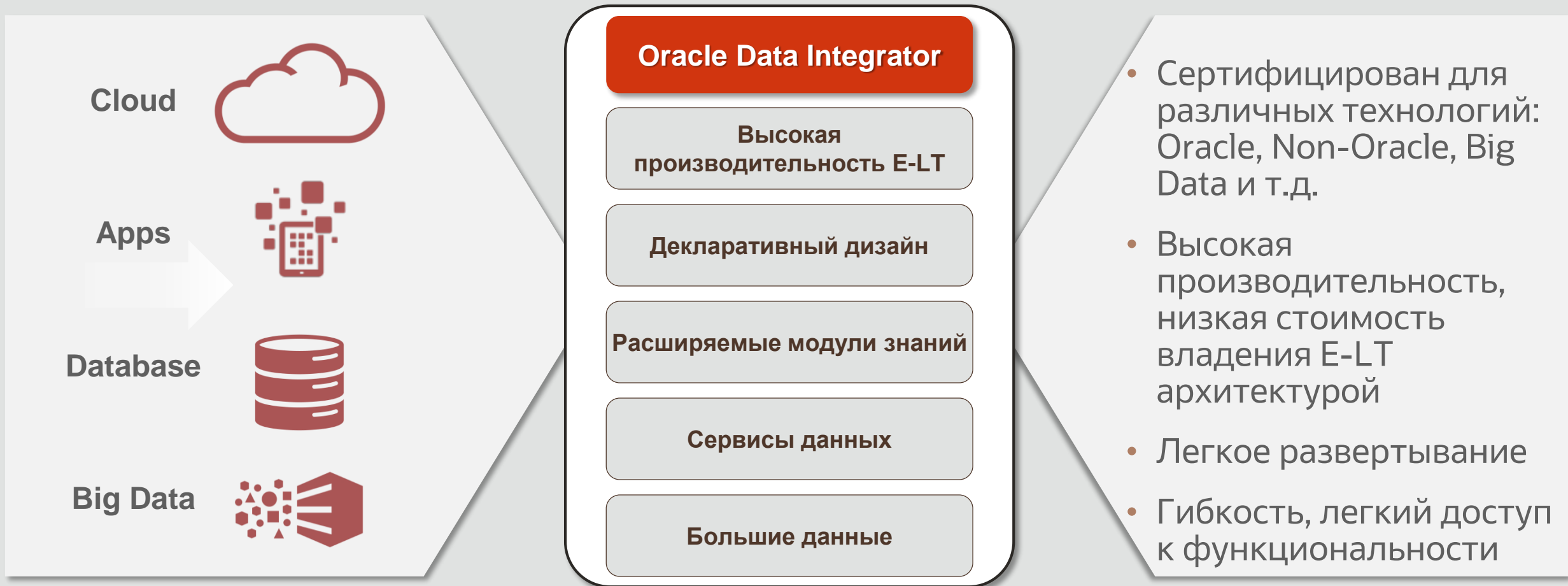
Пример интеграции

# Обзор интеграционной платформы

## Oracle Data Integration

# Oracle Data Integrator

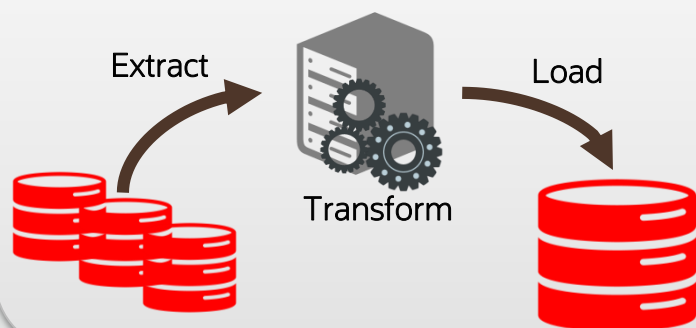
E-LT: Большой объем обработки данных и быстрое преобразование



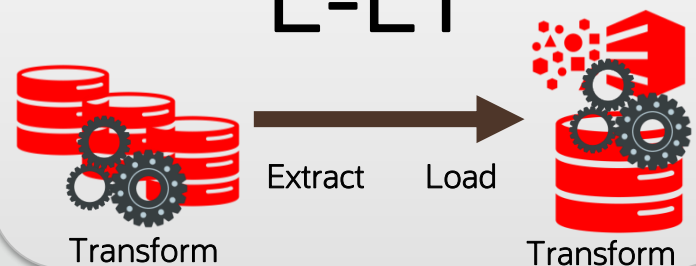


# Лидирующая в отрасли производительность

## Традиционная ETL архитектура



## Архитектура следующего поколения E-LT



E-LT обеспечивает гибкую архитектуру для оптимизации производительности на любой платформе

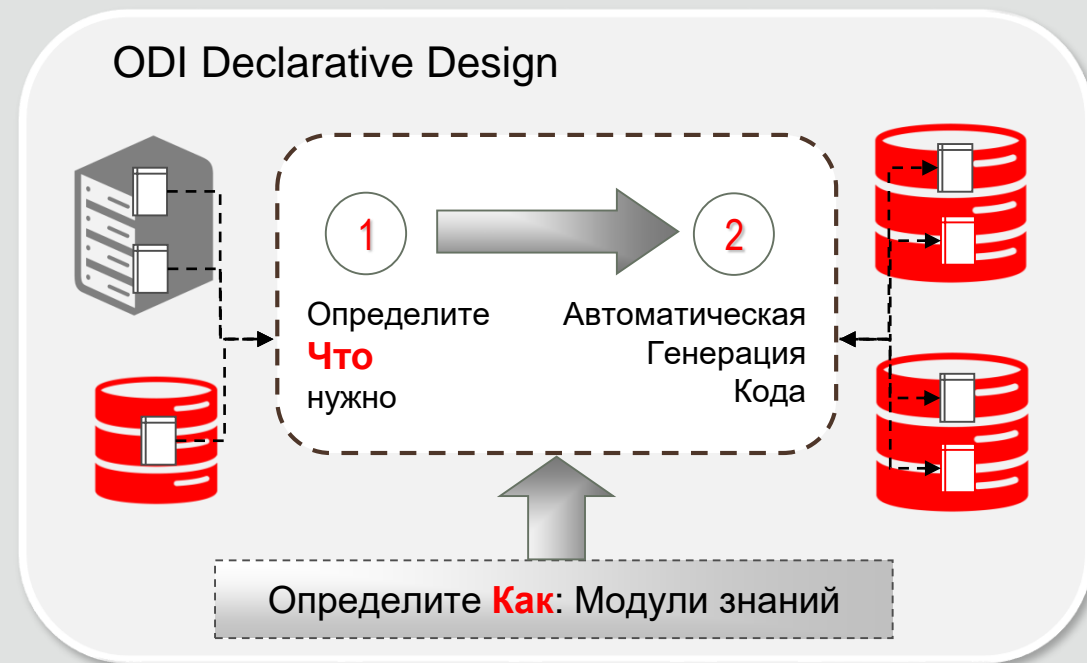
### Преимущества

- Трансформации выполняются в базах
- Улучшает производительность загрузки, нет сетевых прыжков
- Использование преимуществ существующей инфраструктуры: аппаратного и программного обеспечения

# Создание трансформаций без необходимости программирования

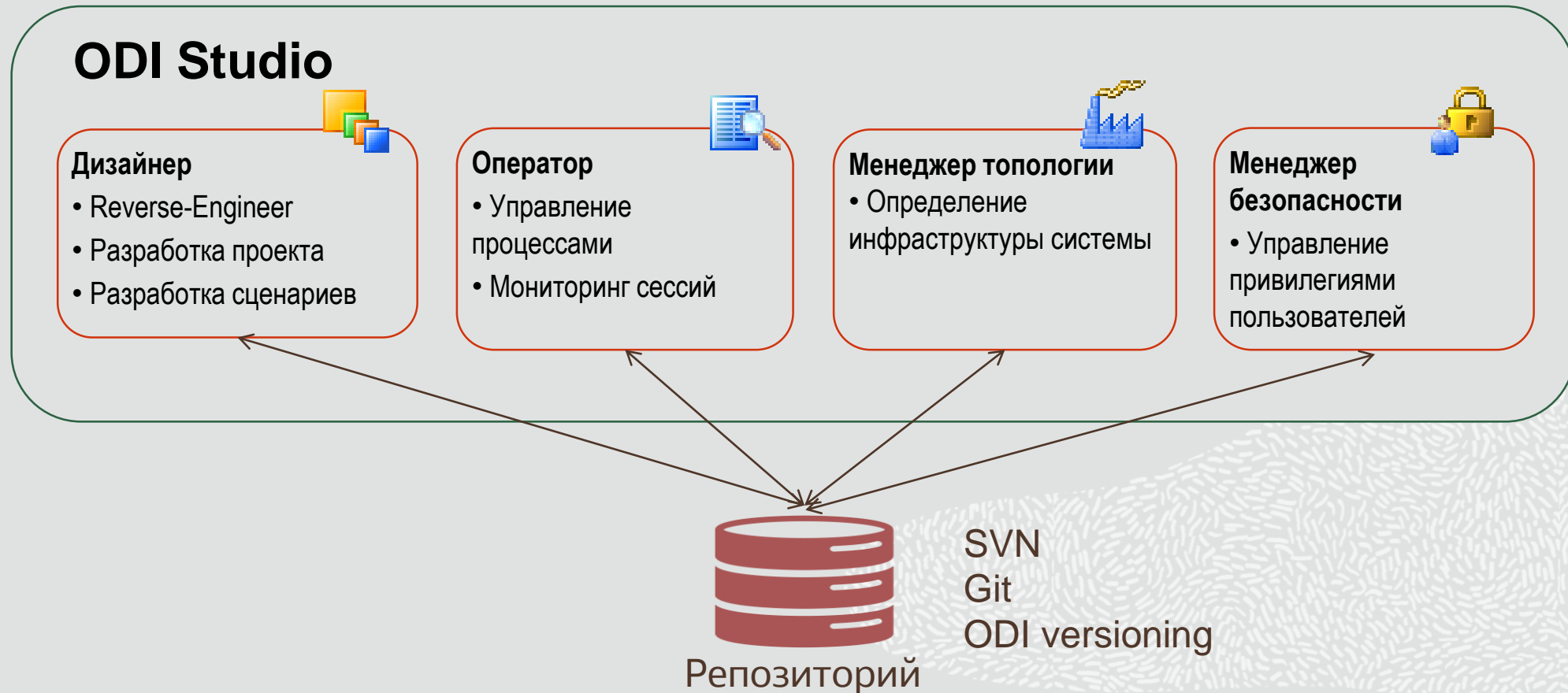
Продуктивность разработчиков и скорость внедрения новых решений

- **Простой и понятный GUI** для создания логики преобразований
- **Используйте с любыми технологиями:** Логика преобразования не зависит от технологий
- **Скорость** внедрения новых цепочек преобразований без необходимости писать сложные процедуры преобразований
- **Прозрачная** система управления жизненным циклом



# ODI Studio

Графическая среда разработки и управления жизненным циклом проектов

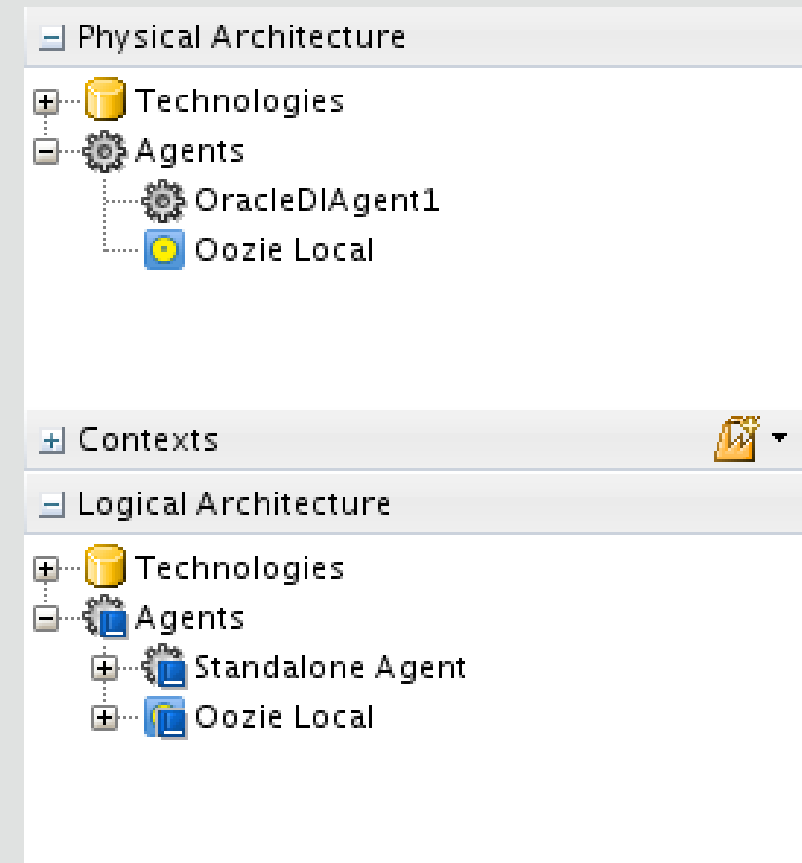




# Oracle Data Integrator

## Агент

Агент – java процесс, управляющий этапами и технологиями исполнения задачи с момента запуска и до завершения (SQL\*Loader, BCP Loader, data pump, etc.)



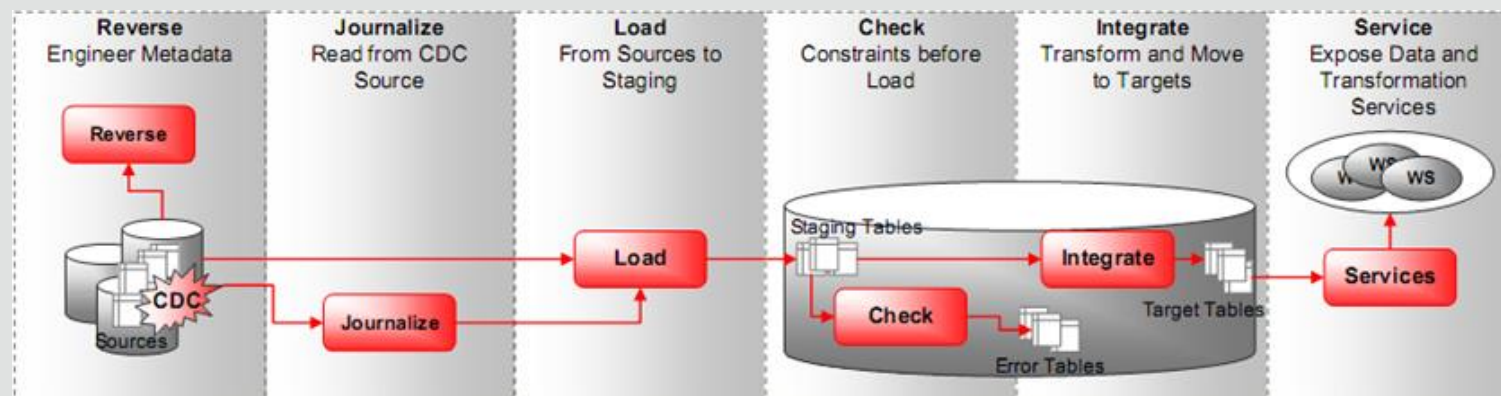
# Модули знаний

## Изменяемая архитектура модулей знаний



### Пример встроенных модулей знаний

Oracle	Sqoop	Hive	Pig
Spark	HBase	Oracle Merge	SAP ERP
SAP BW	Oracle Datapump	Oracle DBLink	JMS
External Tables	Teradata	Oracle Spatial	Siebel
eBusiness Suite	IBM DB2	Netezza	DBaaS



### Ключевые преимущества

- Ускорение разработки и упрощение обслуживания с использованием шаблонов
- Легко расширить и добавить новые лучшие практики
- Обеспечивает предсказуемость и снижает стоимость владения



# Модули знаний. Пример (LKM File to Oracle).

Definition

Tasks

Options

Markers

Memo

Version

Privileges

Flexfields

Task Hierarchy	Target Command	Source Command	Target Technology	Target Commit	Source Technology
Mapping Begin					
Execution Unit Begin					
Execution Unit Main					
Validate mapping location	<?int ODIKM_errorsFound = 0; /*****		Jython	<Undefined>	File
Initialize environment	import java.sql as sqlimport java.lang as lan...		Jython	<Undefined>	File
Drop work table	drop table <?=odiRef.getObjectNme("L", "...		Oracle	<Undefined>	File
Drop work view	drop view <%=odiRef.getTable("L", "COLL_...		Oracle	<Undefined>	File
Create Oracle directory	create or replace directory dat_dir AS '<%=...		Oracle	<Undefined>	File
Grant dir access to work schema	<? // Do we need a view on top of external ...		Oracle	<Undefined>	File
Create external table	createTblCmd = r'''create table <?=odiRef...		Jython	<Undefined>	File
Close connections	myCon.close()		Jython	<Undefined>	File

Пример:

INSERT INTO

<%=odiRef.getOption("INSERT\_HINT")%>

<%=odiRef.getTable("L", "TARG\_NAME", "A")%>

(

<%=odiRef.getColList("", "[COL\_NAME]", ",\n\t", "", "((INS and !TRG) and REW)")%>

<%=odiRef.getColList("", "[COL\_NAME]", ",\n\t", "", "((INS and TRG) and REW)")%>

)

SELECT

<%=odiRef.getOption("SELECT\_HINT")%>

<%=odiRef.getColList("", "[COL\_NAME]", ",\n\t", "", "((INS and !TRG) and REW)")%>

<%=odiRef.getColList("", "[EXPRESSION]", ",\n\t", "", "((INS and TRG) and REW)")%>

FROM

<%=odiRef.getTable("L", "INT\_NAME", "A")%>

INSERT /\*+ APPEND PARALLEL \*/

INTO ODI\_DEMO.TRG\_PRODUCT

(

PRODUCT\_ID,

FAMILY\_ID,

PRICE,

PRODUCT

)

SELECT

PRODUCT\_ID,

FAMILY\_ID,

PRICE,

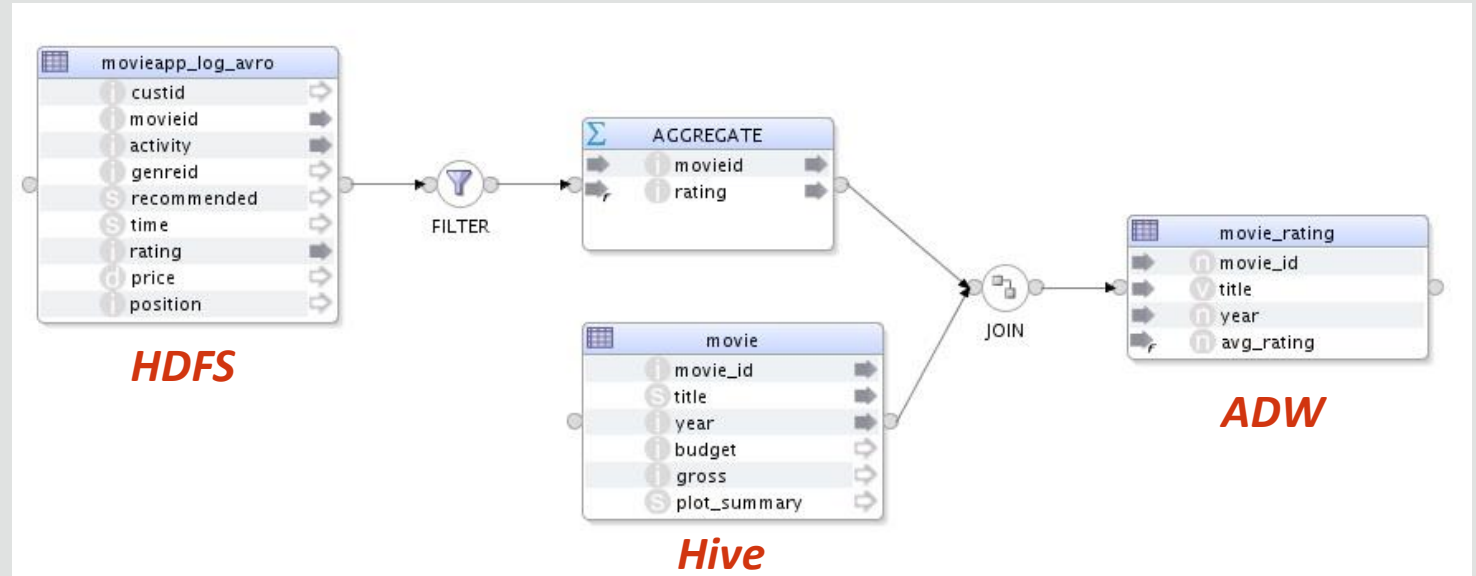
PRODUCT

FROM C\$\_SRC\_PRODUCT

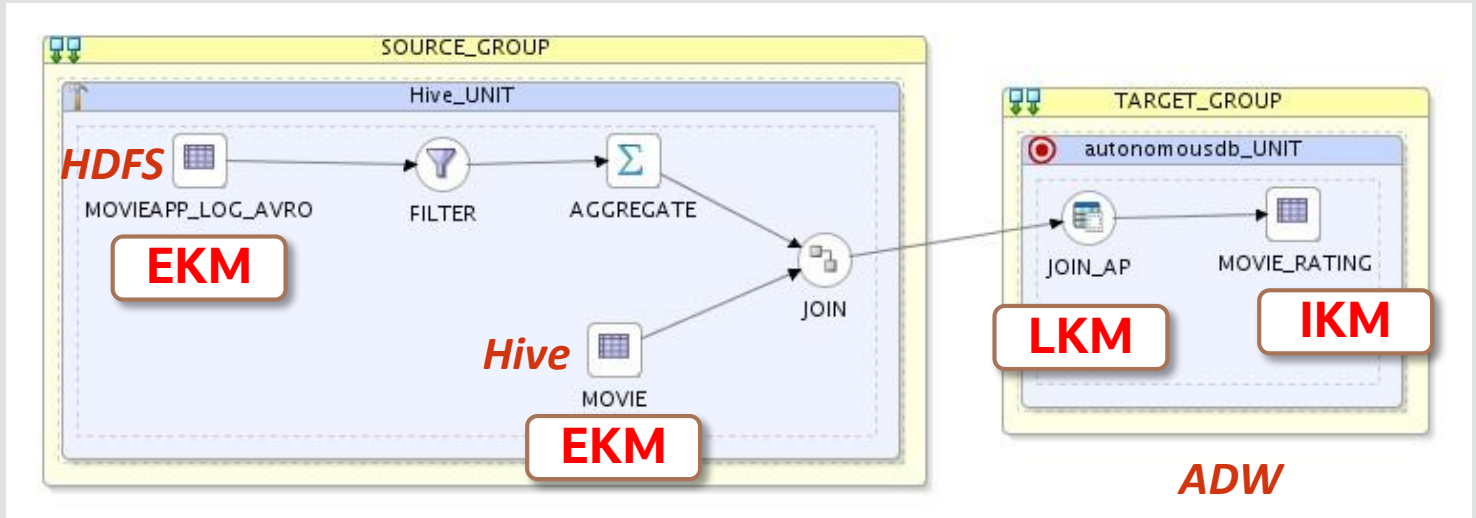
# ODI для работы с Big Data

# Логический и физический дизайн в ODI

Логический  
дизайн



Физический  
дизайн





# Hive: Выполнение преобразований



**Run**

Context: Production

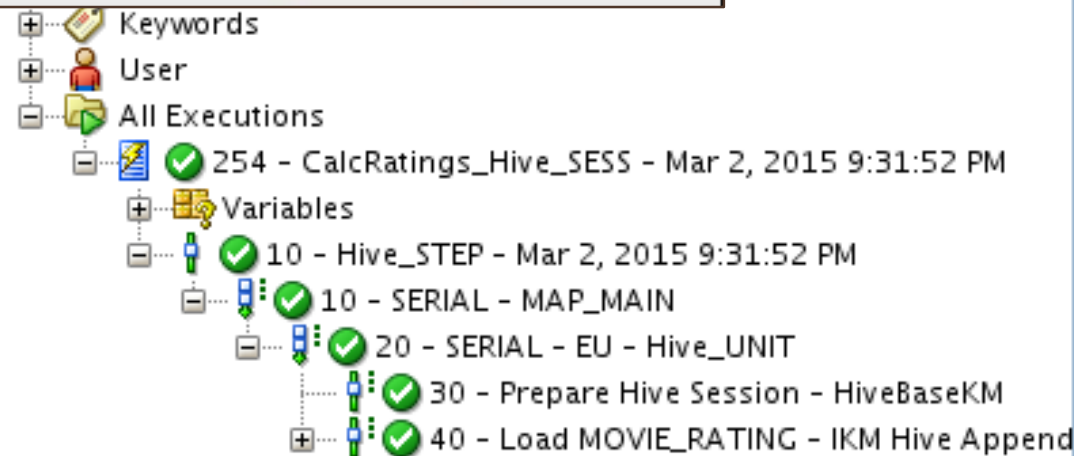
Physical Mapping Design: Oracle

Logical Agent: Oracle  
Hive  
Spark

Log Level:

☐ Simulation

Help OK Cancel



CalcRatings x Session Task Load MOVIE\_RATING - IKM Hive Append x

Definition

**Code**

Connection

Privileges

Code Type: Executed Code

**Target Code**

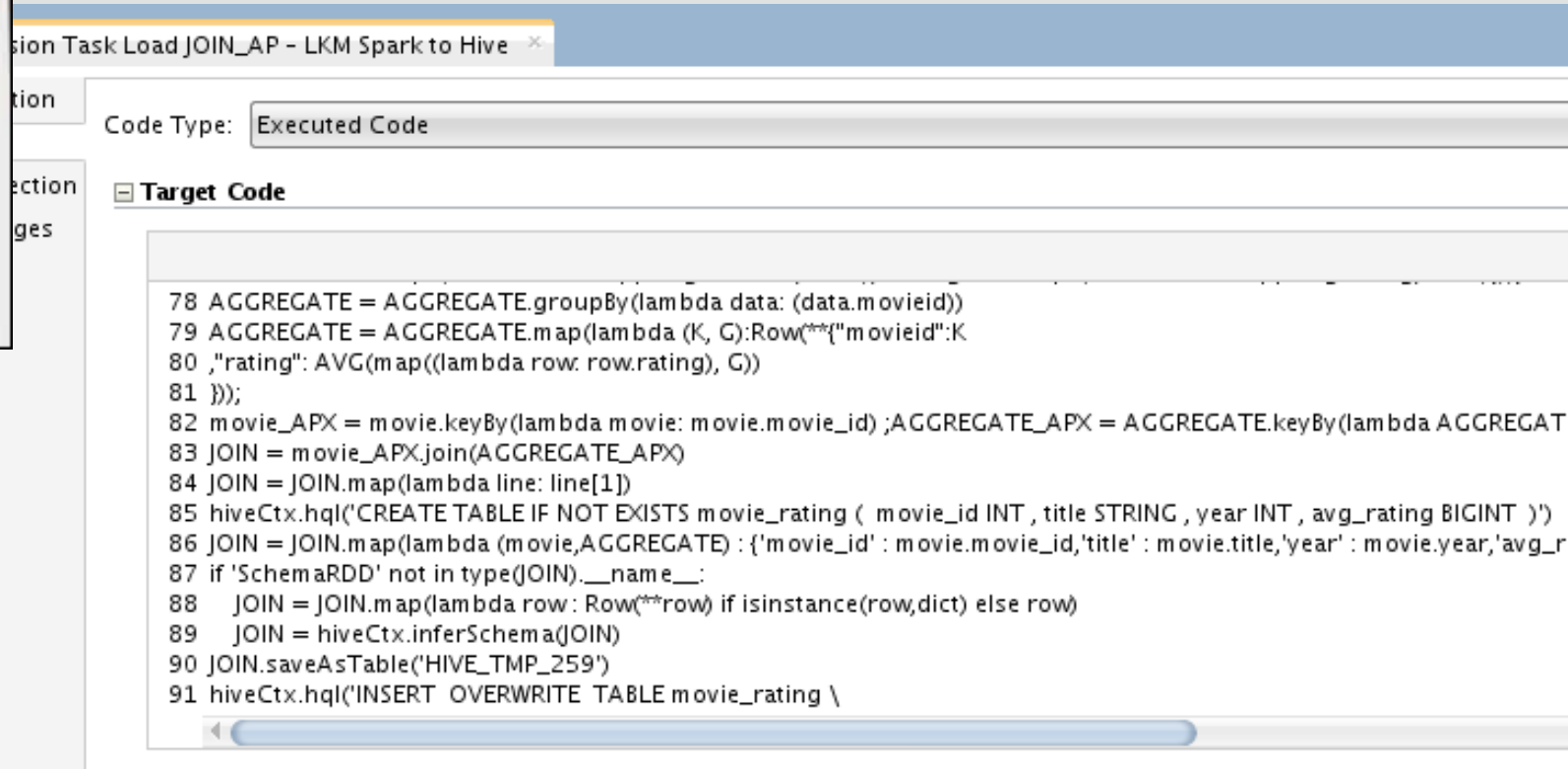
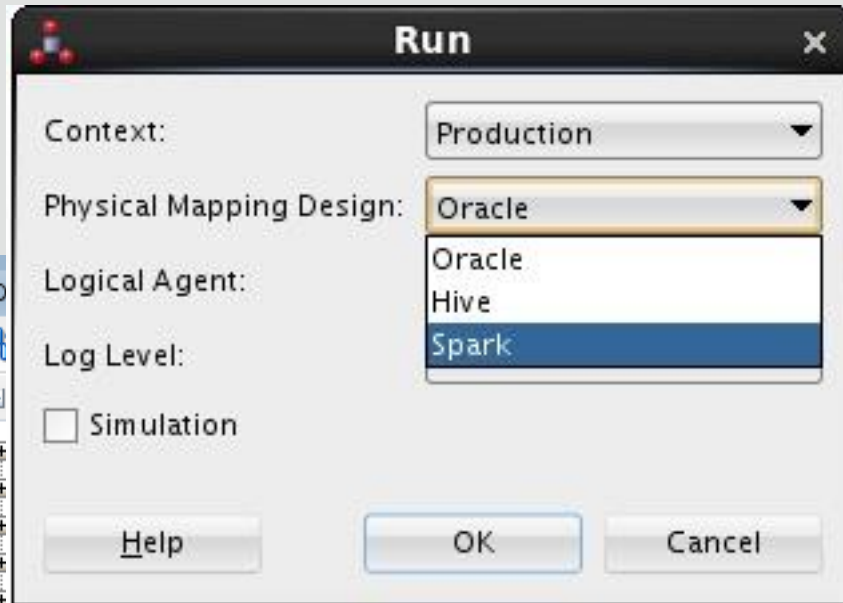
Edit and use as Pre-execution Code Revert to original Pre

```
2 INSERT OVERWRITE TABLE default.movie_rating
3 SELECT
4   MOVIE.movie_id movie_id,
5   MOVIE.title title,
6   MOVIE.year year,
7   ROUND(MOVIEAPP_LOG_1.rating) avg_rating
8 FROM
9   default.movie MOVIE JOIN (
10  SELECT
11    MOVIEAPP_LOG.movieid movieid,
12    AVG(MOVIEAPP_LOG.rating) rating
13 FROM
14   default.movieapp_log MOVIEAPP_LOG
15 WHERE
16   (MOVIEAPP_LOG.activity < 2
```

**Hive  
SQL**



# Spark: Выполнение преобразований



# Oozie



## Hadoop Workflow engine

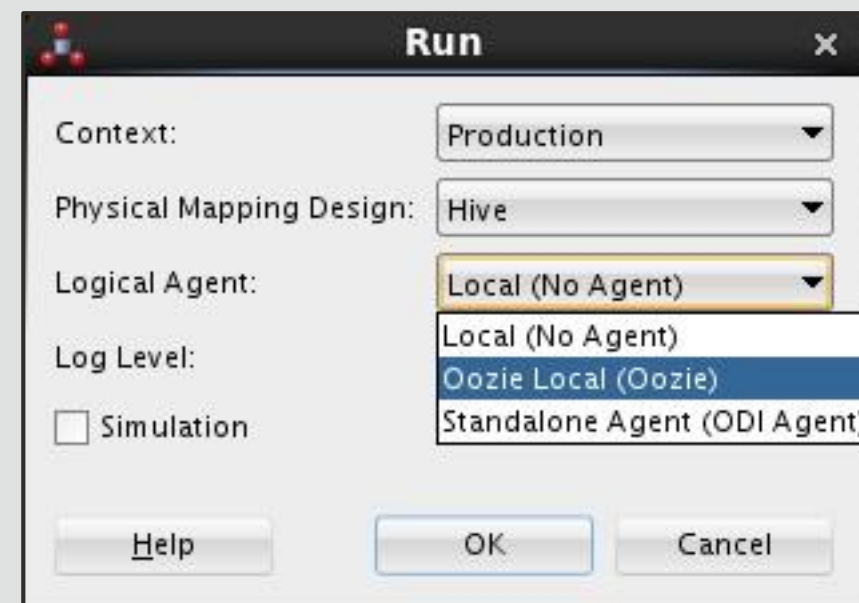
Используется в 90% Hadoop приложений

## Альтернатива ODI агентам в Big Data окружении

Использует нативные технологии Hadoop

Развертывание на Hadoop окружении без следов ODI, не нужна установка

Используются родные Hadoop инструменты для управления процессами и просмотра логов



# Hadoop логи в ODI операторе

The screenshot shows the Oracle Data Integrator Studio interface. The Session List on the left includes a tree structure with the following items:

- Date
- Agent
  - Internal
    - NEW\_HADOOP\_SVR\_TEST\_OozieEngine
      - Session ODI\_OOZIE\_LEGACY\_SIMPLE\_Sc...
      - Variables
        - Session Step Physical\_STEP
        - 10 - SERIAL - MAP\_BEGIN
        - 40 - SERIAL - MAP\_MAIN
        - 90 - SERIAL - MAP\_CLEANUP
      - 1285 - ODI\_STLOG\_WRT\_TEST\_Pkg\_3\_S...
      - 1275 - ODI\_STLOG\_WRT\_TEST\_Pkg\_2\_S...
      - 1255 - ODI\_STLOG\_WRT\_TEST\_Pkg\_1\_S...
      - 1235 - AGENT\_PROCEDURE\_TEST\_PIG\_P...
      - 1225 - ODI\_OOZIE\_PIG\_HIVE\_PIG\_Sc...

The Session Step Physical\_STEP is selected, and its details are shown on the right. The Execution Details table is as follows:

Name	Job Tracker	Status	Started	Finished
Physical_STEP-s10-t30	<a href="#">job_1416525041653_0955</a>	OK	Wed Nov 26 01:21:48 PST 2014	Wed Nov 26 01:23:2
Physical_STEP-s10-t60	<a href="#">job_1416525041653_0956</a>	OK	Wed Nov 26 01:23:21 PST 2014	Wed Nov 26 01:24:5
Physical_STEP-s10-t80	<a href="#">job_1416525041653_0957</a>	OK	Wed Nov 26 01:24:51 PST 2014	Wed Nov 26 01:26:2
Physical_STEP-s10-t110	<a href="#">job_1416525041653_0958</a>	OK	Wed Nov 26 01:26:43 PST 2014	Wed Nov 26 01:27:5

The Hadoop job overview page shows the following details:

**MapReduce Job job\_1416525041653\_0957**

Job Name: oozie:launcher:T=java:W=SCEN\_712:A=Physical\_STEP-s10-t80:ID=0000850-141113105055392-oozie-oozi-W

User Name: gwatters

Queue: root.gwatters

State: SUCCEEDED

Uberized: false

Submitted: Wed Nov 26 01:24:51 PST 2014

Started: Wed Nov 26 01:25:06 PST 2014

Finished: Wed Nov 26 01:26:26 PST 2014

# Легкая настройка Hadoop

Шаблоны для последних версий дистрибутивов Cloudera

Поддержка расширенной безопасности Big Data (Kerberos, Sentry и.т.д.)

Big Data Configurations - Step 1 of 10

General Settings

**General**

- [Hadoop](#)
- [HDFS](#)
- [HBase](#)
- [Spark](#)
- [Kafka](#)
- [Pig](#)
- [Hive](#)
- [Oozie](#)
- [Validate all settings](#)

**Basic Settings**

Prefix: BDC\_DEMO

Distribution: CDH 5.10

Distribution Type: Normal

Base Directory: /opt/cloudera/parcels/CDH

**Technologies**

- ☒ Hadoop
- ☒ HDFS
- ☒ HBase
- ☒ Spark
- ☒ **Kafka**
- ☒ Pig
- ☒ Hive
- ☒ Oozie

Help < Back Next > Finish Cancel

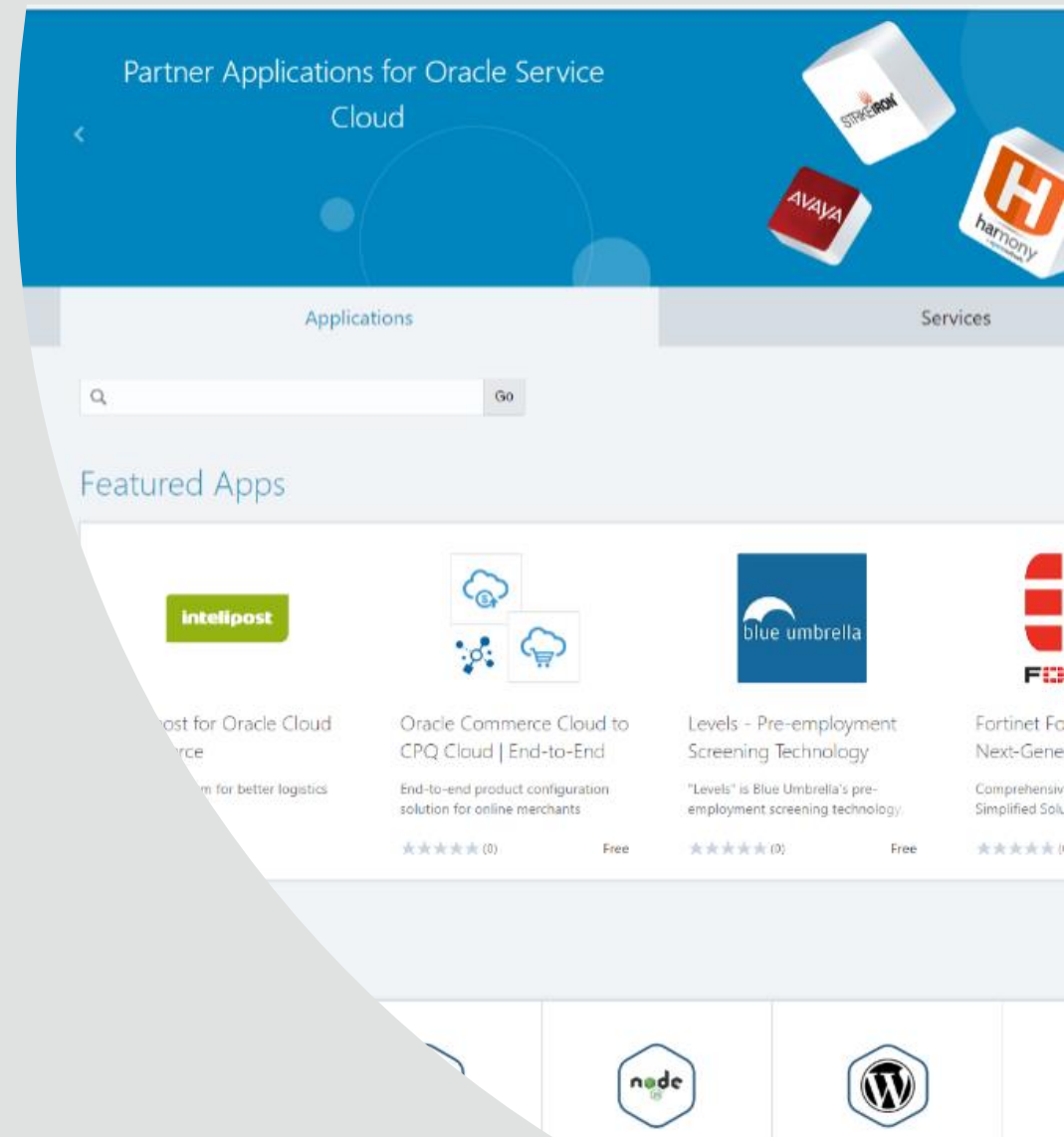
# ODI в облаке

## Oracle Cloud Marketplace



# OCI Маркетплейс

- Новейшие дата центры, агрессивный график сборки
- Высокоскоростные вычисления, сеть и хранилища
- Сборка автоматизирована через скрипты Oracle Terraform
- Будет работать с Oracle Database Cloud, включая Autonomous Database и любые приложения, размещенные на OCI Compute
- Надежная безопасность, стандартные аренда OCI и опции VPN для поддержки любой on-premise или не-Oracle облачной интеграции





## Marketplace

All Applications

Deployed Applications

## Filters

[clear](#)

### TYPE

Any

### PUBLISHER

Oracle

### CATEGORY

Data Integration

### PRICE

Any

## All Applications

**ORACLE**  
Enterprise Data  
Quality

**Oracle Enterprise Data  
Quality on Tomcat**

A full instance of EDQ, Oracle's  
data quality management...

Type: Image | Price: BYOL

**ORACLE**  
Enterprise Data  
Quality

**Oracle Enterprise Data  
Quality on WebLogic**

A full instance of EDQ, Oracle's  
data quality management...

Type: Image | Price: BYOL

**ORACLE**  
GoldenGate

**Oracle GoldenGate 19c for  
Oracle**

Oracle GoldenGate 19c for  
Oracle

Type: Stack | Price: BYOL

**ORACLE**  
Data Integrator

**Oracle Data Integrator**

A full instance of ODI, Oracle  
Data Integrator software

Type: Stack | Price: Free

Marketplace » Oracle Data Integrator

**ORACLE**  
Data Integrator

# Oracle Data Integrator

A full instance of ODI, Oracle Data Integrator software

Oracle Data Integrator

Data Integration, Data Integration, Data Integration, Data Integration

Type  
Stack

Version

ODI Marketplace Stack ▾

Compartment

Platform ▾

cissandbox (root)/Platform

Software Price per OCPU

**\$0.00/hr**

Your actual fees may vary based on your billing model. ⓘ

☒ I have reviewed and accept the [Oracle Standard Terms and Restrictions](#).

**Launch Stack**

**Overview**

Provider

More Apps

Usage Instructions

## App by Oracle

This image provides a full instance of Oracle Data Integrator (ODI) based on 12.2.1.4

The image includes.:

- The option to install the ODI repository on an Autonomous Data Warehouse or use a preconfigured ODI

## Support

**Contacts:**

Support hotline ((1) 800.223.1711)

**Links:**

ORACLE Cloud

Create Stack

✓

Stack Information

2

Configure Variables

3

Review

Configure the variables for the infrastructure resources that this stack will create when you run the apply job for this execution plan.

General Settings

NETWORKING OPTIONS

Existing networking components will be used

The ODI Instance can be created inside a new VCN and Subnet, that will be provisioned and configured for you, or it can be created inside an existing network.

ODI REPOSITORY LOCATION

Create a new ODI Repository in an Autonomous Database

The ODI repository can either be located in an Autonomous Database in this Tenancy, or in an embedded database in the ODI Instance.

RESOURCE DISPLAY NAME PREFIX OPTIONAL

Display name prefix for all generated compute and network resources. If not specified, this will be automatically generated.

TARGET COMPARTMENT

Platform

The target compartment for all of the provisioned resources

Network Configuration

Back

Next

Cancel

Terms of Use and Privacy

Cookie Preferences

Network Configuration

VCN COMPARTMENT

kusova

VCN

odi

SUBNET COMPARTMENT

kusova

SUBNET OPTIONAL

Public Subnet RJwG:EU-FRANKFURT-1-AD-1

☒

ASSIGN PUBLIC IP ADDRESS  
Assign a public ip address

ODI Instance Settings

ODI NODE SHAPE

VM.Standard2.4

The shape for all ODI compute instances

SSH PUBLIC KEY

ssh-rsa AAAAB3NzaC1yc2EAAAABJQAAAQEAg/kmBloC89VziOqa1ZctvNSbLHct6Hjiv

SSH keys are needed to ssh to the ODI vm instance. It can be generated using puttygen. Use the corresponding private key to access the ODI compute instances

ODI AVAILABILITY DOMAIN

RJwG:EU-FRANKFURT-1-AD-1

The name of the availability domain in which to create the ODI compute instances

ODI VNC PASSWORD

\*\*\*\*\*

The VNC password for the ODI instance


Create






23

Copyright © 2019, Oracle and/or its affiliates.


# Готовый к работе ODI

IPSecVPN  
Fast Connect



Germany Central (Frankfurt)     

Compute » Instances » Instance Details



RUNNING

## odi-node-1

Start Stop Reboot Move Resource Apply Tag(s) Actions ▾

Instance Information

Tags

### Instance Information

**Availability Domain:** RJwG:EU-FRANKFURT-1-AD-1

**Fault Domain:** FAULT-DOMAIN-1

**Region:** eu-frankfurt-1

**Shape:** VM.Standard2.4

**Virtual Cloud Network:** [odi](#)

**Maintenance Reboot:** -

### Primary VNIC Information

**Private IP Address:** 10.0.0.19

**Public IP Address:** 13\*\*\*\*\*

**Network Security Groups:** None [Edit](#)

This instance's traffic is controlled by its firewall rules in addition to the associated [Subnet's](#) security lists and the VNIC's network security groups.

### Launch Options

**NIC Attachment Type:** VFIO

**Remote Data Volume:** PARAVIRTUALIZED

**Image:** [Published Image: ODI Marketplace Image V2.0.1](#)

**OCID:** ...bvmssa [Show](#) [Copy](#)

**Launched:** Tue, 01 Oct 2019 14:44:16 UTC

**Compartment:** cissandbox (root)/Platform/kusova

**Launch Mode:** NATIVE

**Internal FQDN:** oracle-odi-inst-gsgp... [Show](#) [Copy](#)

**Subnet:** [Public Subnet RJwG:EU-FRANKFURT-1-AD-1](#)

**Firmware:** UEFI\_64

**Boot Volume Type:** PARAVIRTUALIZED

После статуса *Running* еще + 20 минут на автоматическое создание репозитория



PuTTY  
SSH

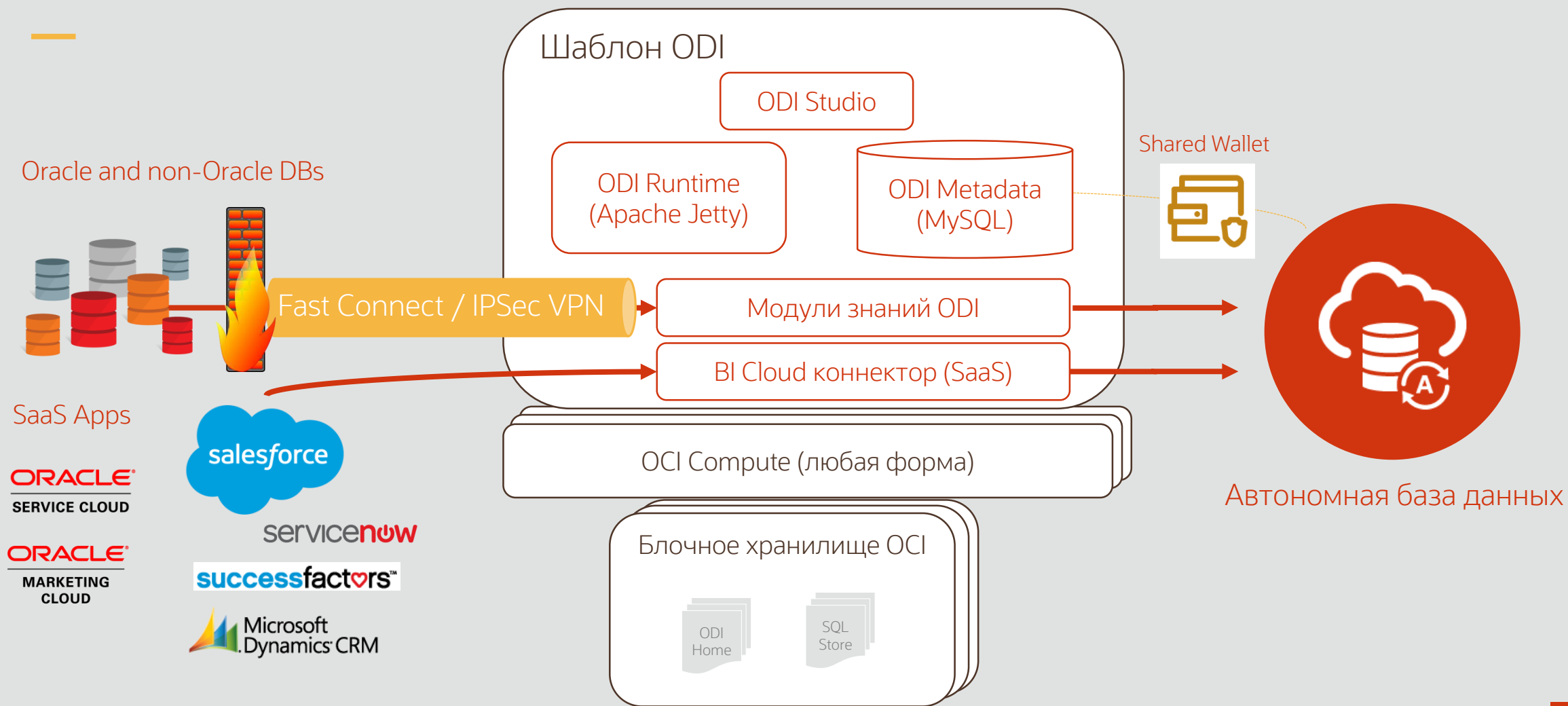


Real VNC





# Архитектура | ODI в маркетплейс

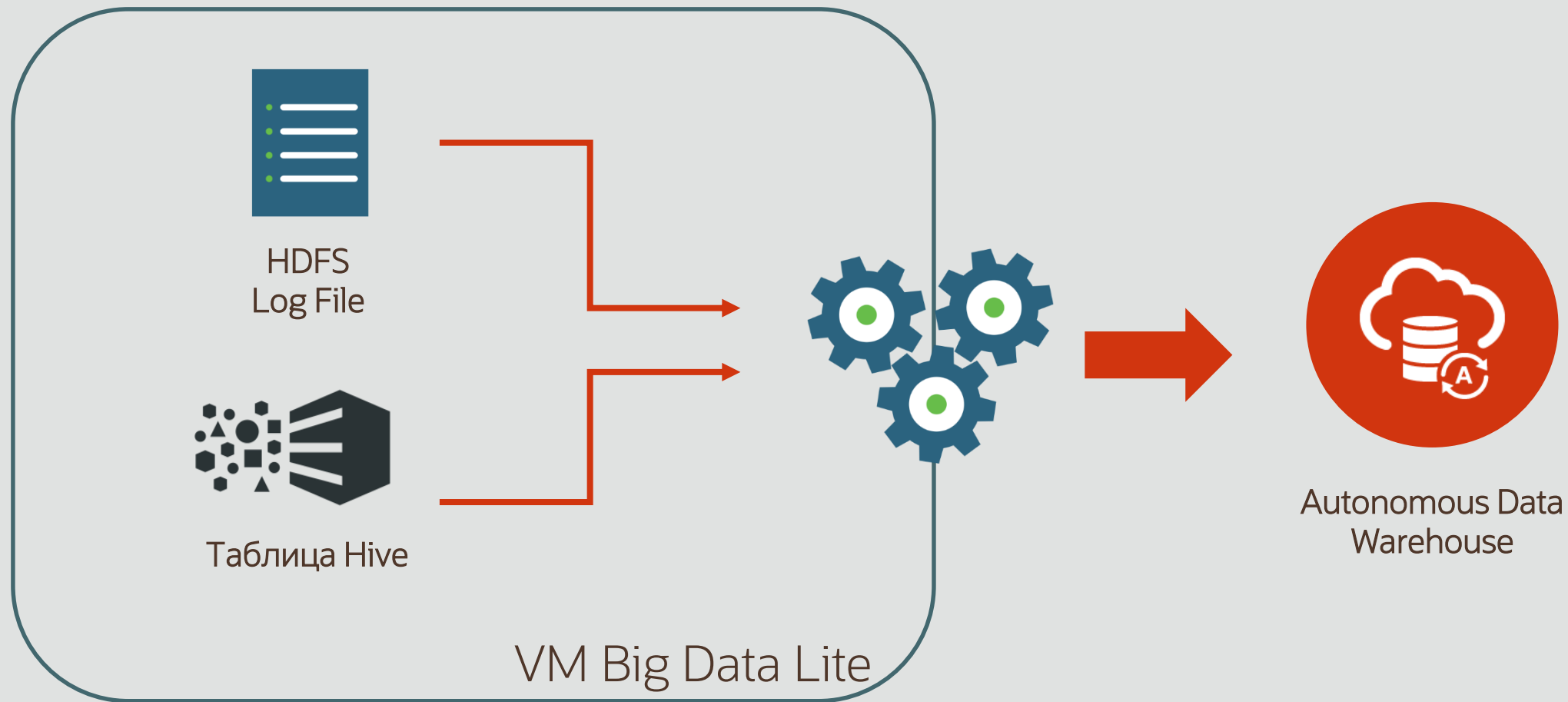


# Пример интеграции



# Пример маппинга

Oracle MoviePlex



Oracle Data Integrator Studio 12c : ODI Movie Demo

File Edit View Search QDI Tools Window Team Run Help

Designer Operator Topology Security Start Page

Projects

- Big Data Demo
  - Demo
    - Packages
      - Mappings
        - A - Load Movies (Sqoop)
        - B - Merge Movies (Hive)
        - C - Calc Ratings (Hive - Pig - Spark)
        - D - Calc Ratings (JSON Flatten)
        - E - Load Oracle (OLH)
        - F - Calc Sales (Big Data SQL)
        - G - Sessionize Data (Pig)
      - Reusable Mappings
      - Procedures
      - Setup
      - Variables
      - Sequences
      - User Functions
      - Knowledge Modules
      - Markers
- Models
  - AutonomousDB
    - Used by
    - Diagrams
    - Hierarchy
    - ODI\_MOVIE\_RATING
    - movie\_rating
    - Hidden DataStores
  - HDFSMovie
  - HiveMovie
    - Used by
    - Diagrams
    - Hierarchy
    - MOVIE2
    - cust
- Dimensions and Cubes
- Load Plans and Scenarios
- Global Objects
- Labels

DATA INTEGRATOR

Learn & Explore

What's New

Release Notes

Training Resources

Community

Featured Tutorials

- Getting Started Guide
- Creating and Connecting to ODI Master and Work Repositories
- Creating an ODI Project and Mapping: Flat File to a Table
- Creating an Agent
- Creating Procedures and Scenarios

Featured Documentation

- Get Started
- Develop and Deploy
- Administer

Messages - Log Validation Results

Object	Issue Details	Severity	Issue
A - Load Movies (Sqoop)	DATASTORE: movie_updates	Warning	The size of map attribute MOVIE.OUTPUT1.PLOT_SUMMARY (4,000) is greater than the size of attribute movie_updates.OUTPUT1.plot_summary (250), t...



# Oracle Data Integrator Studio 12c : ODI Movie Demo

File Edit View Search Diagram ODI Tools Window Team Run Help

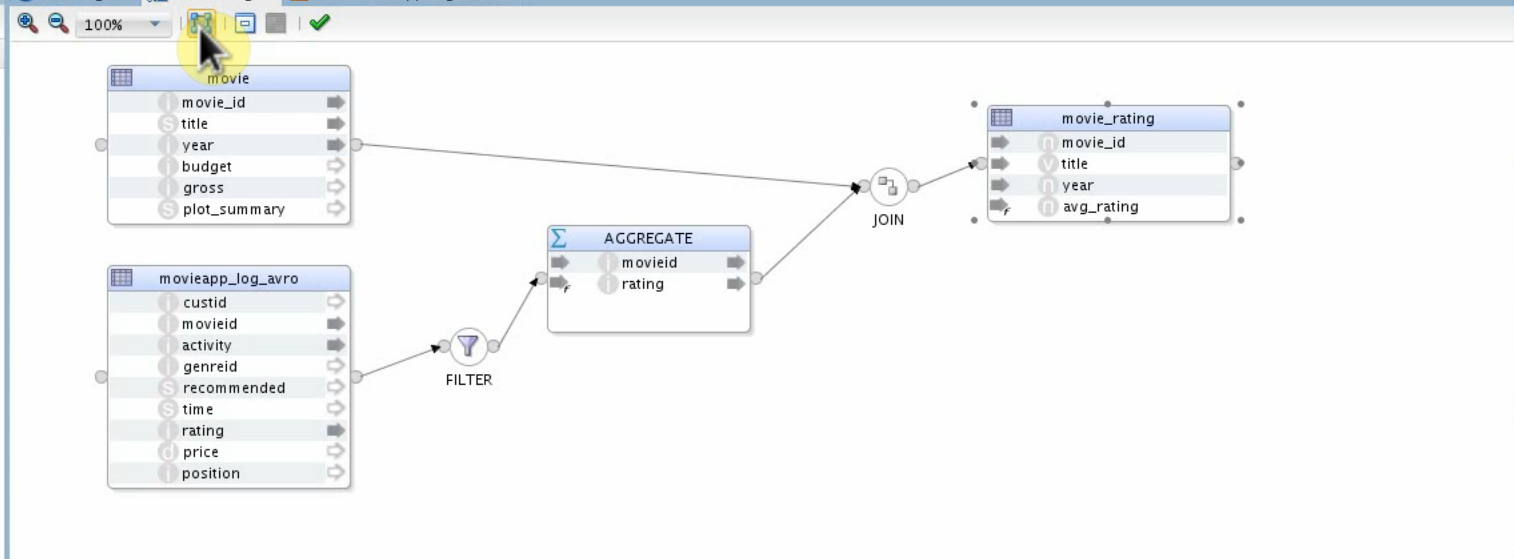
Designer Operator Topology Security

Physical Architecture

- Technologies
  - File
  - HBase
  - HDFS
  - Hadoop
  - Hive
  - In-Memory Engine
  - Oracle
    - AutonomousDB
    - OracleMovie
    - Data Types
    - Actions
    - Index Type
  - Pig
  - Spark Python
  - Spark
    - Data Types
    - Actions
    - Index Type
- Agents
  - OracleDIAgent1
  - Oozie Local

- Contexts
- Logical Architecture
- Languages
- Repositories
- Generic Action

Start Page Calc Rating Data: movieapp\_log\_avro



Overview Logical Physical

movie\_rating - Properties Messages - Log Validation Results

Find

Attributes												
General												
Target												
Journalizing												
Constraints												
Connector Points												
Attributes:												
Name	Data Type	Length	Scale	Data format	Expression	Execute on H...	Fixed Executi...	Key	Insert	Update	Check Not N...	
movie_id	NUMBER	11	0		movie.movie...	No Hint		<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
title	VARCHAR2	250			movie.title	No Hint		<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
year	NUMBER	11	0		movie.year	No Hint		<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
avg_rating	NUMBER	11	0		XROUND(AG...	No Hint		<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	

Components

Logical

General

- Aggregate
- Dataset
- Distinct
- Expression
- Filter
- Flatten
- Jagged
- Join
- Lookup
- Pivot
- Set
- Sort
- Split
- Subquery Filter
- Table Function
- Unpivot

Saved: Spark

Oracle SQL Developer ... Oracle Data Integrator...

SUPERVISOR dev\_odi\_repo WORKREPI

Oracle Data Integrator...

Oracle Data Integrator...

Oracle Data Integrator...

Oracle Data Integrator...

Oracle Data Integrator...

## BigData Lite

---

Пример на компании Oracle MoviePlex, предоставляющей пользователям потоковое видео (смотреть фильмы онлайн).

Преобразование данных, используя Hive|Spark|Oracle, загрузка данных в автономную базу данных Oracle в облаке используя Oracle Loader для Hadoop

Виртуальная машина и материалы для дальнейшего изучения доступны для скачивания:

<http://www.oracle.com/technetwork/database/bigdata-appliance/oracle-bigdatalite-2104726.html>.



# Thank You

---

**Natalia Kusova**

Solution Engineering