



# 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

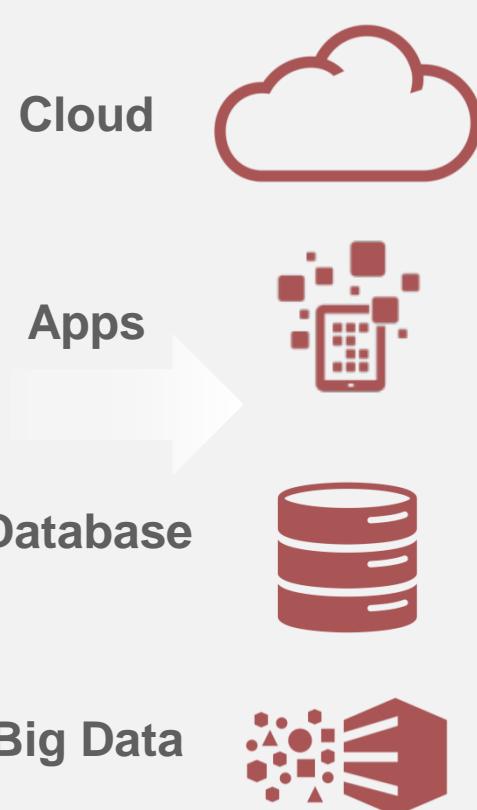


# Обзор интеграционной платформы

**Oracle Data Integration**

# Oracle Data Integrator

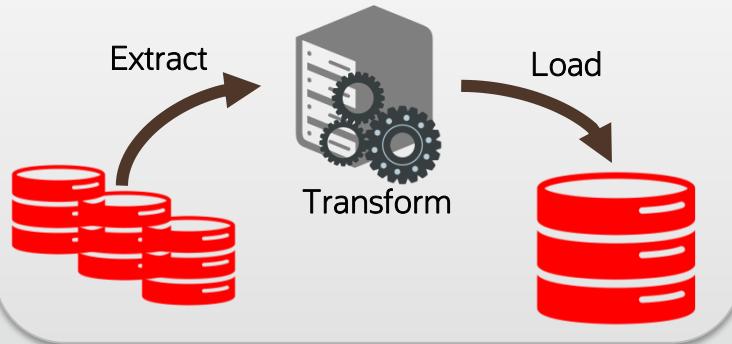
E-LT: Большой объем обработки данных и быстрое преобразование



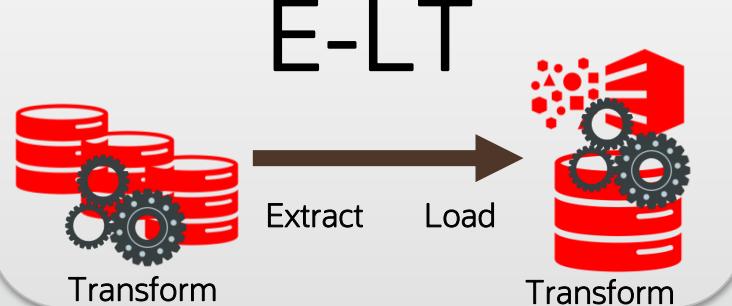
- Сертифицирован для различных технологий: Oracle, Non-Oracle, Big Data и т.д.
- Высокая производительность, низкая стоимость владения E-LT архитектурой
- Легкое развертывание
- Гибкость, легкий доступ к функциональности

# Лидирующая в отрасли производительность

## Традиционная ETL архитектура



## Архитектура следующего поколения E-LT



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

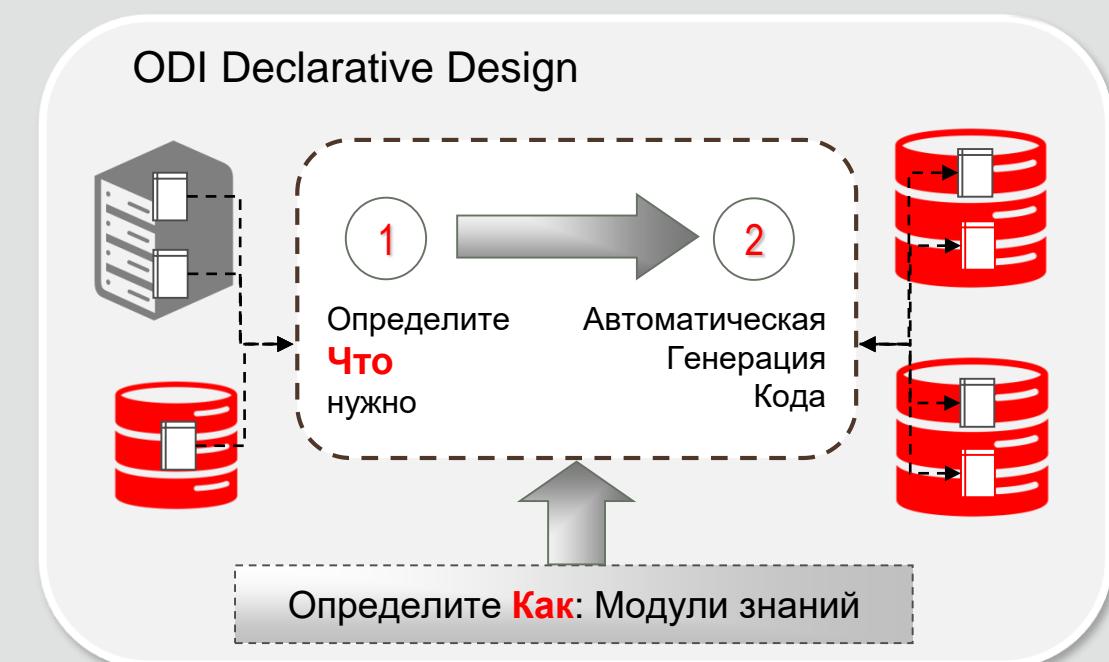
### Преимущества

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

# Создание трансформаций без необходимости программирования

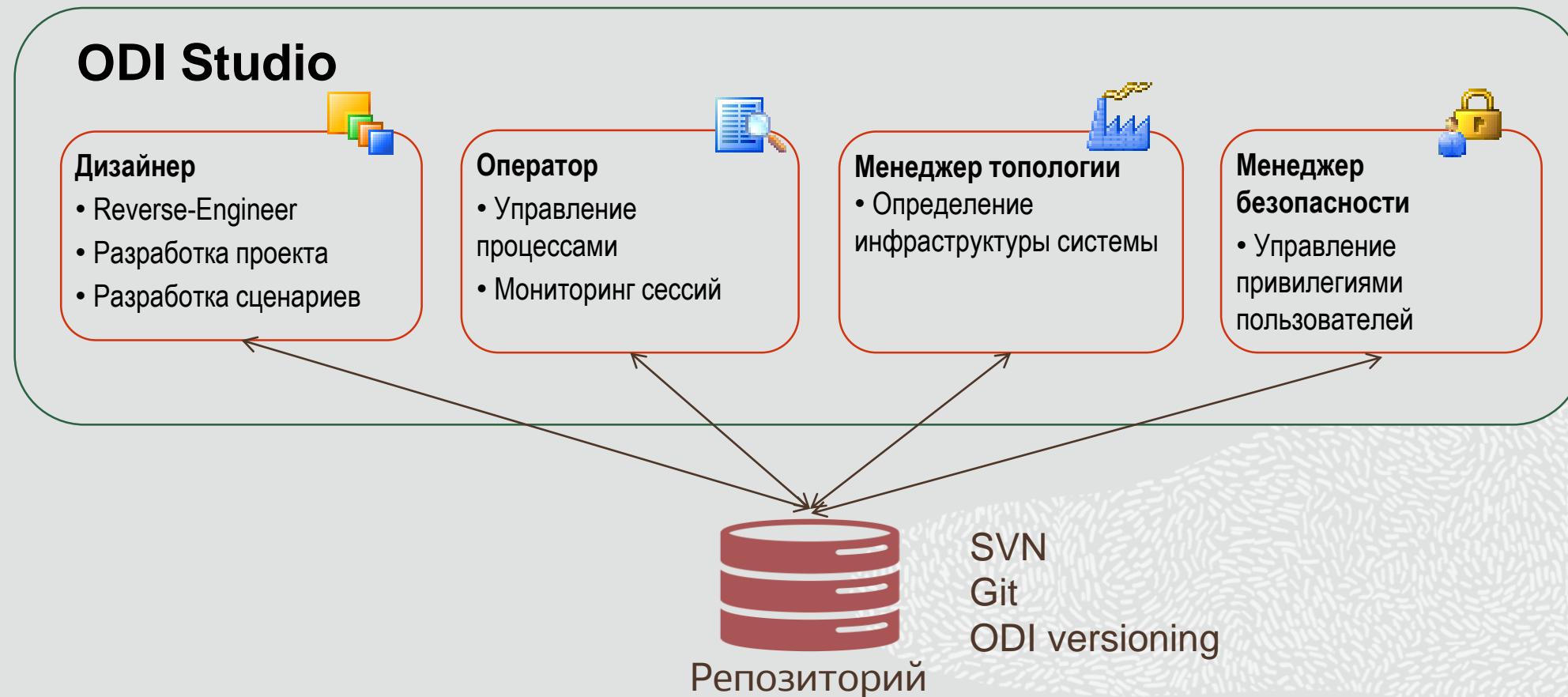
Продуктивность разработчиков и скорость внедрения новых решений

- 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

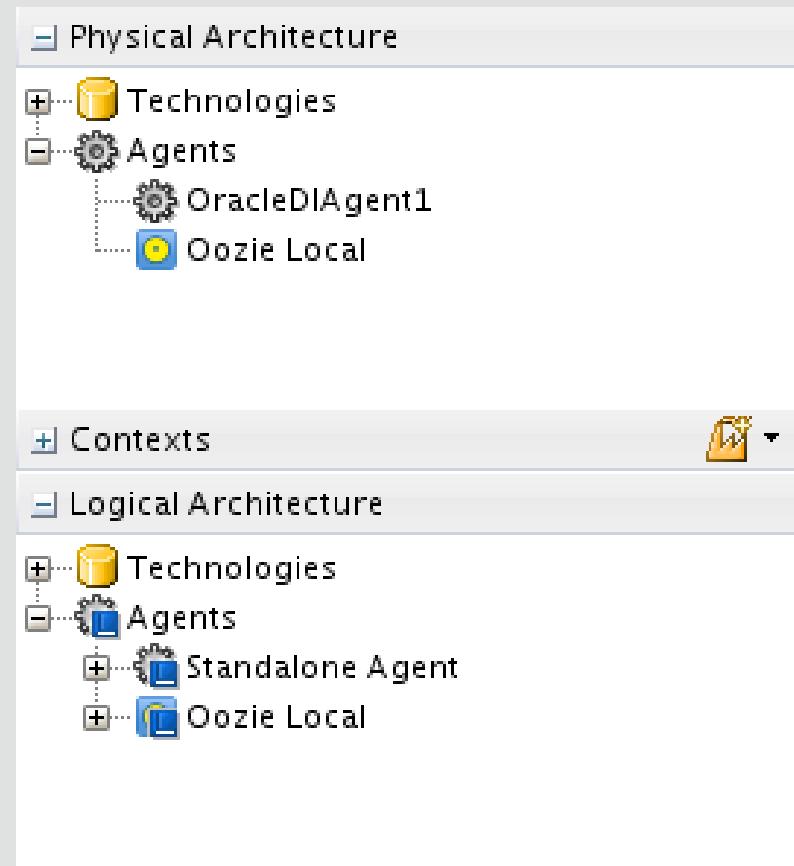
Графическая среда разработки и управления жизненным циклом проектов



# 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.)



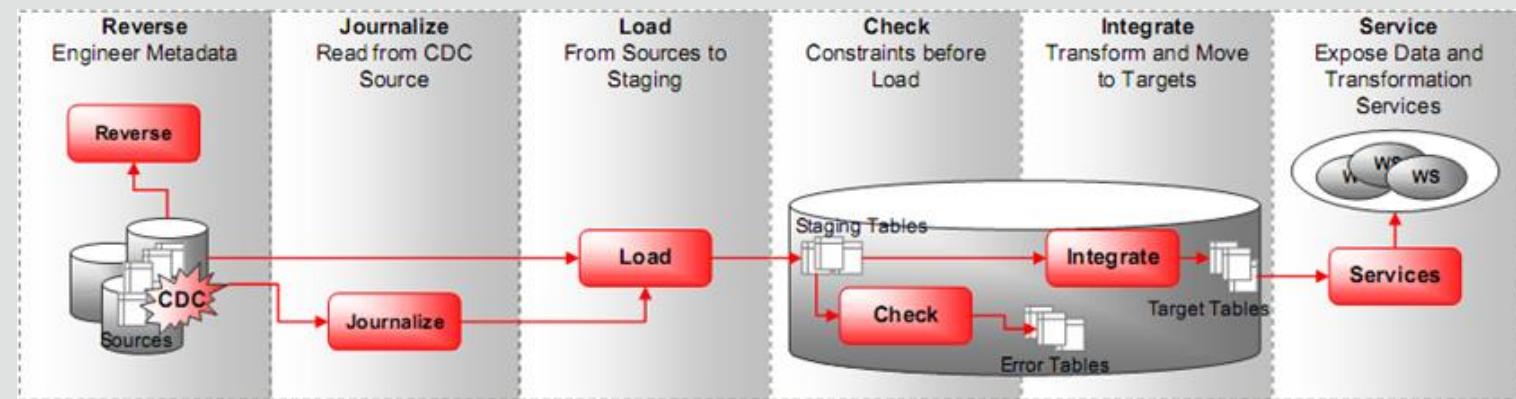
# Модули знаний

Изменяющаяся 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	<?int ODIKM_errorsFound = 0;/*****...>		Jython	<Undefined>	File
Execution Unit Begin			Jython	<Undefined>	File
Execution Unit Main			Oracle	<Undefined>	File
Validate mapping location	<?int ODIKM_errorsFound = 0;/*****...>		Oracle	<Undefined>	File
Initialize environment	import java.sql as sqlimport java.lang as lan...		Oracle	<Undefined>	File
Drop work table	drop table <?=odiRef.getObjectName("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

Copyright © 2019, Oracle and/or its affiliates. All rights reserved.

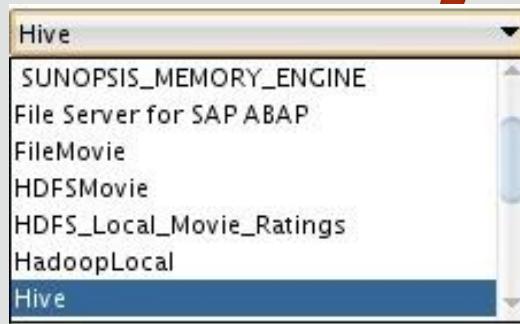
Oracle OpenWorld 2019

ORACLE

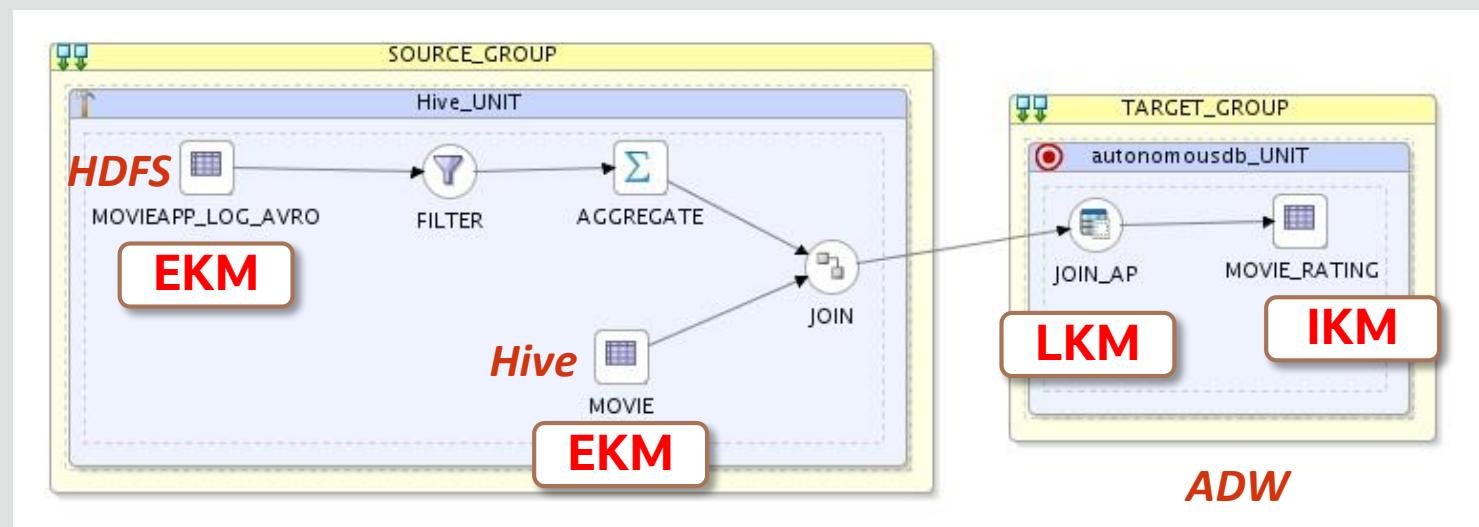
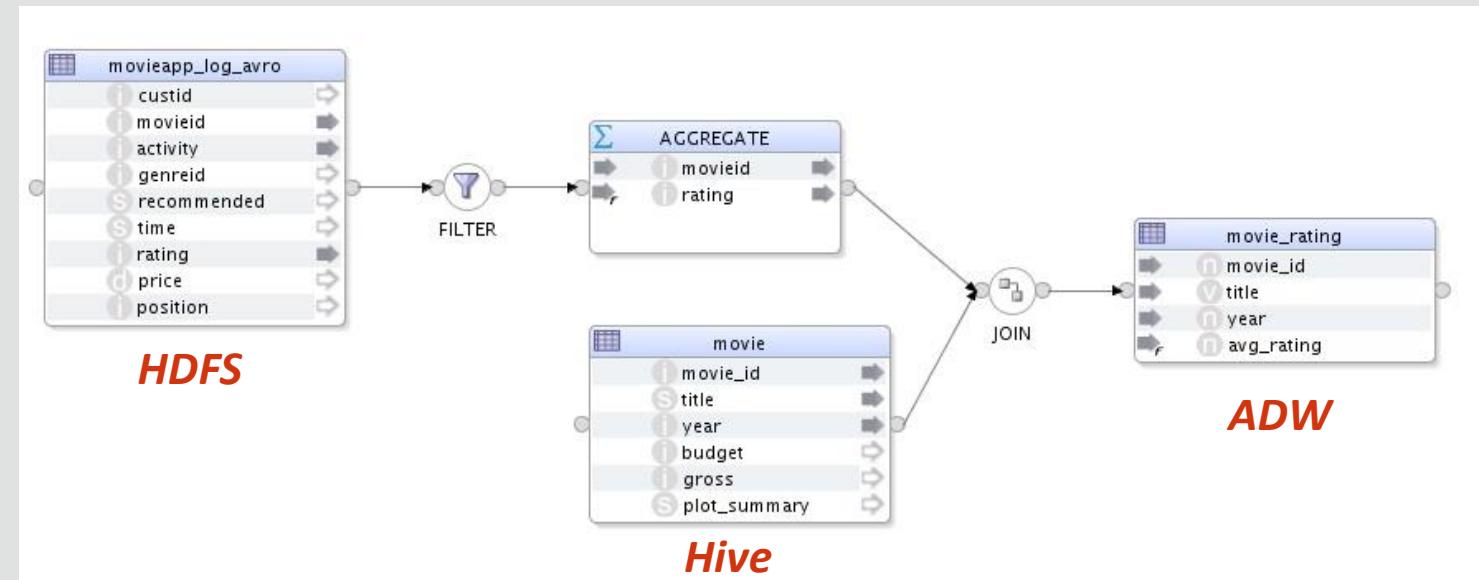
# ODI для работы с Big Data

# Логический и физический дизайн в ODI

Логический  
дизайн



Физический  
дизайн



# Hive: Выполнение преобразований



Run

Context: Production

Physical Mapping Design: Oracle

Logical Agent: Oracle, Hive (selected), Spark

Log Level:

Simulation

Help OK Cancel

Keywords

User

All Executions

- 254 - CalcRatings\_Hive\_SESS - Mar 2, 2015 9:31:52 PM
  - Variables
  - 10 - Hive\_STEP - Mar 2, 2015 9:31:52 PM
    - 10 - SERIAL - MAP\_MAIN
      - 20 - SERIAL - EU - Hive\_UNIT
        - 30 - Prepare Hive Session - HiveBaseKM
        - 40 - Load MOVIE\_RATING - IKM Hive Append

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
17 )
```

Hive SQL

# Spark: Выполнение преобразований



The screenshot shows a user interface for running a Spark job. On the left, a "Run" dialog box is open with the following settings:

- Context: Production
- Physical Mapping Design: Oracle
- Logical Agent: Oracle (selected)
- Log Level: Spark
- Simulation

At the bottom of the dialog are "Help", "OK", and "Cancel" buttons.

To the right of the dialog, a "Session Task Load JOIN\_AP - LKM Spark to Hive" window is displayed. It shows the "Code Type: Executed Code" and the "Target Code" section, which contains the following Scala code:

```
78 AGGREGATE = AGGREGATE.groupBy(lambda data: (data.movieid))
79 AGGREGATE = AGGREGATE.map(lambda (K, G):Row("movieid":K
80 , "rating": AVG(map((lambda row: row.rating), G)))
81 }));
82 movie_APX = movie.keyBy(lambda movie: movie.movie_id);AGGREGATE_APX = AGGREGATE.keyBy(lambda AGGREGAT
83 JOIN = movie_APX.join(AGGREGATE_APX)
84 JOIN = JOIN.map(lambda line: line[1])
85 hiveCtx.hql('CREATE TABLE IF NOT EXISTS movie_rating ( movie_id INT , title STRING , year INT , avg_rating BIGINT )')
86 JOIN = JOIN.map(lambda (movie,AGGREGATE) : {'movie_id': movie.movie_id,'title' : movie.title,'year' : movie.year,'avg_r
87 if 'SchemaRDD' not in type(JOIN).__name__:
88   JOIN = JOIN.map(lambda row: Row(**row) if isinstance(row,dict) else row)
89   JOIN = hiveCtx.inferSchema(JOIN)
90 JOIN.saveAsTable('HIVE_TMP_259')
91 hiveCtx.hql('INSERT OVERWRITE TABLE movie_rating \\\n
```

Below the session task, the "All Executions" tree view shows a completed execution named "259 - CalcRatings\_Spark\_SESS - Mar 2, 2015 10:38:" with several steps expanded, including "Variables", "10 - Spark\_STEP - Mar 2, 2015 10:38:54 PM", "10 - SERIAL - MAP\_MAIN", "20 - SERIAL - EU - Spark\_UNIT", "30 - Load JOIN\_AP - LKM Spark to Hive", and "40 - sparkPyExec - LKM Spark to Hive".



# Oozie

Hadoop Workflow engine

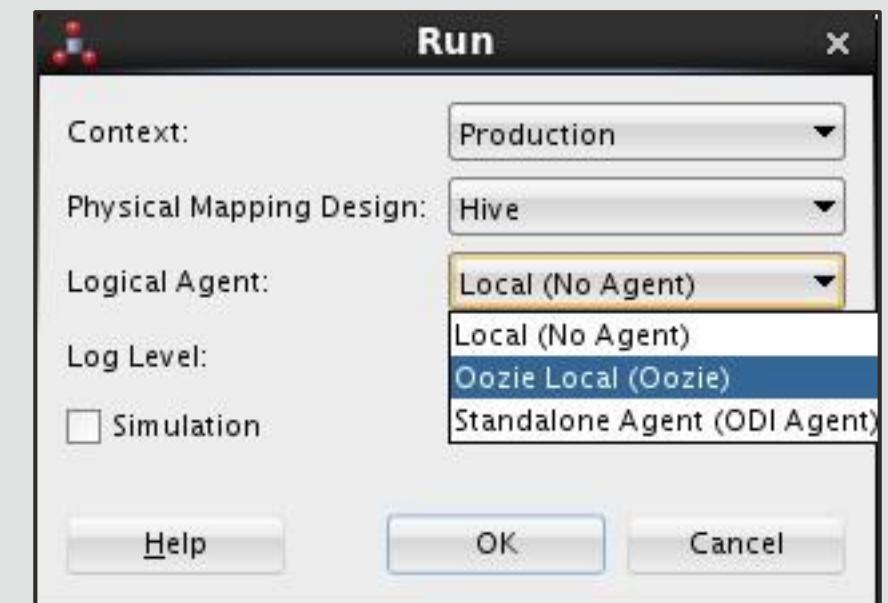
Используется в 90% Hadoop приложений

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. On the left, the 'Session List' pane displays various sessions, including 'Session ODI\_OOZIE\_LEGACY\_SIMPLE\_Scen'. A red arrow points from this session to the 'Execution Details' table on the right. Another red arrow points from the 'Execution Details' table to a detailed view of the MapReduce job 'job\_1416525041653\_0957' on the bottom right.

**Session List:**

- Date
- Agent
  - Internal
    - NEW\_HADOOP\_SVR\_TEST\_OozieEngine
      - Session ODI\_OOZIE\_LEGACY\_SIMPLE\_Scen
        - Variables
        - Session Step Physical\_STEP
          - 10 - SERIAL - MAP\_BEGIN
          - 40 - SERIAL - MAP\_MAIN
          - 90 - SERIAL - MAP\_CLEANUP

**Execution Details:**

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

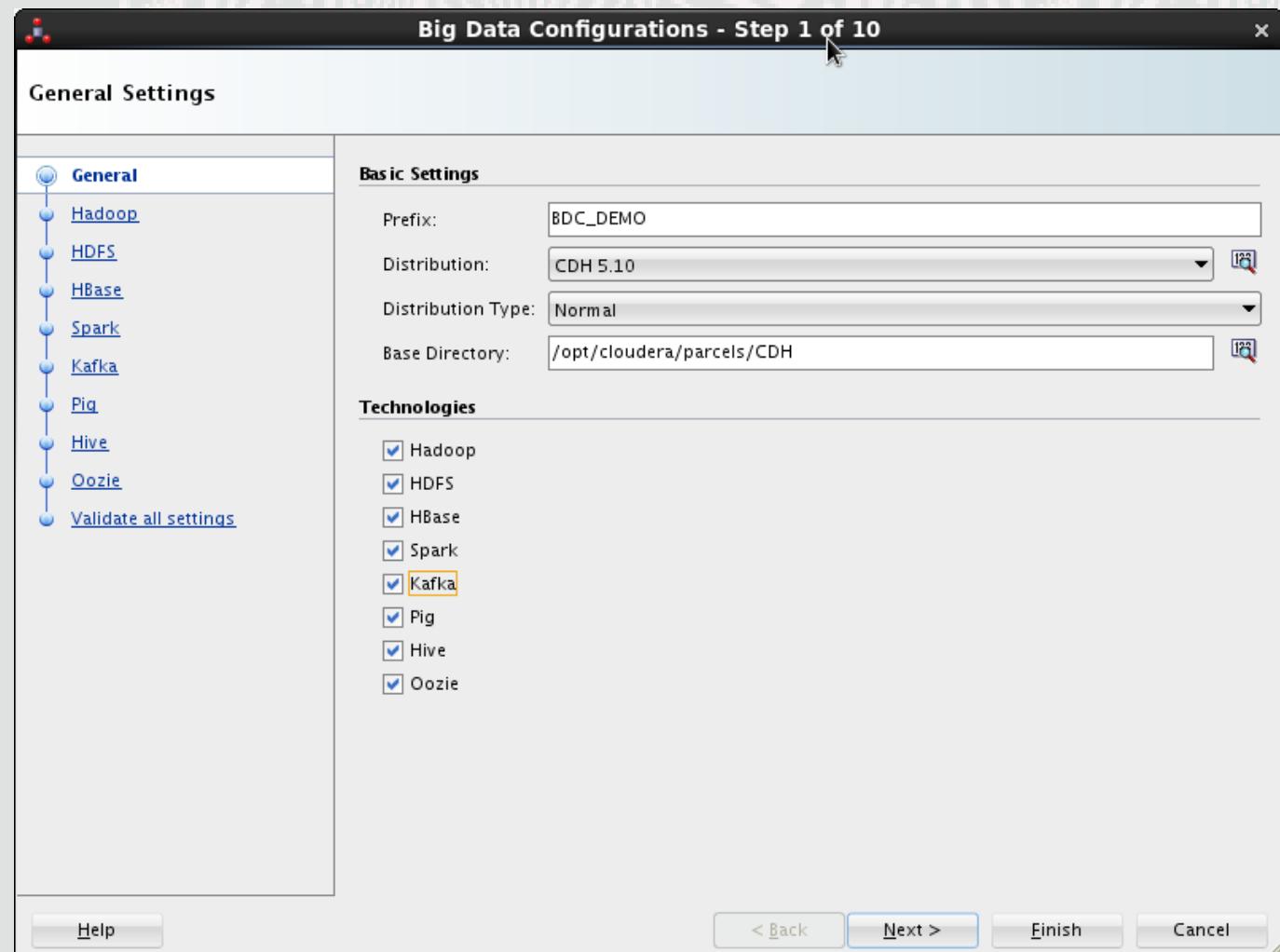
**MapReduce Job job\_1416525041653\_0957**

Job Overview	
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.)



**ORACLE**

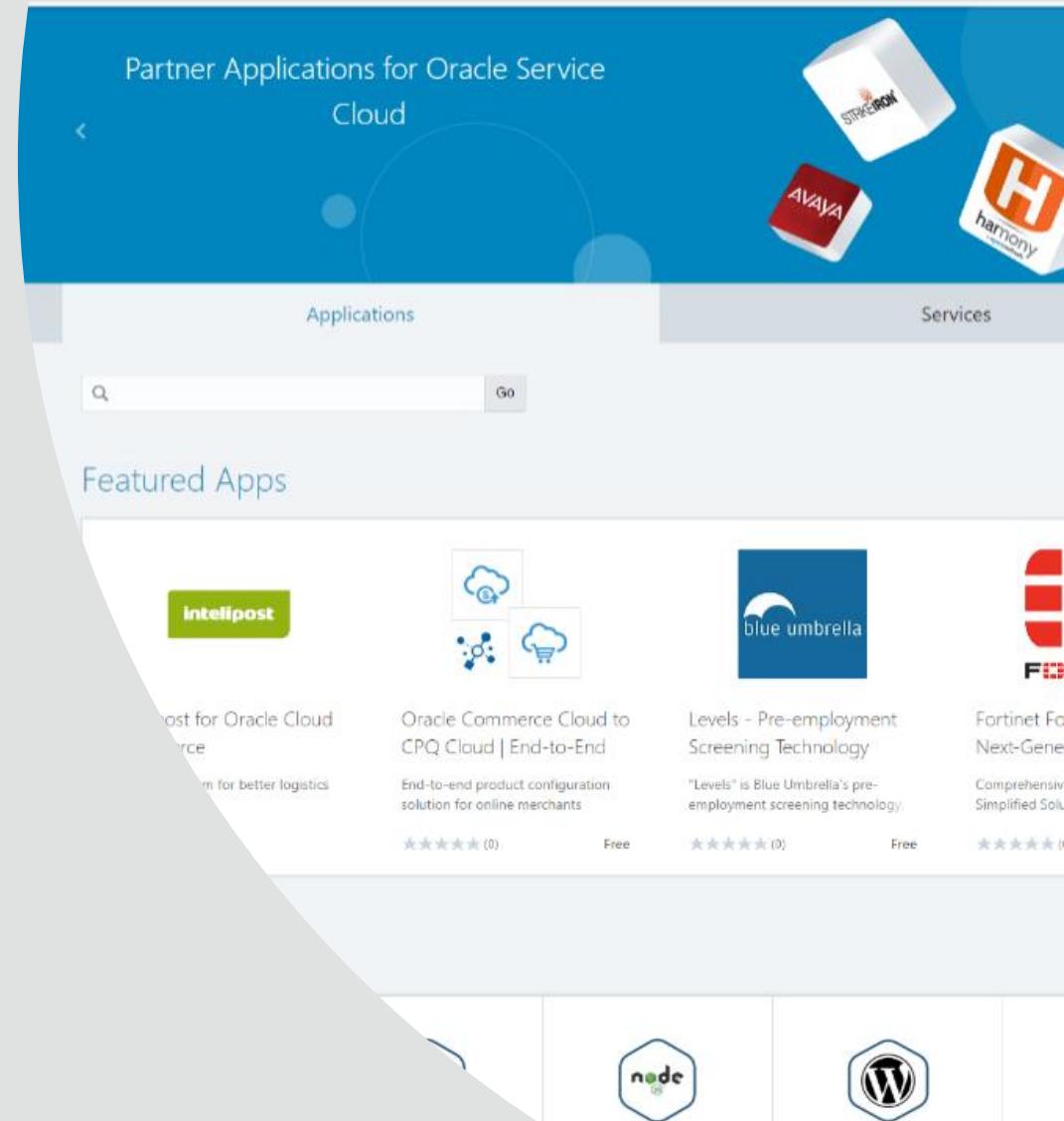
# ODI в облаке

## Oracle Cloud Marketplace



# OCI Маркетплейс

- Новейшие данные 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 аренды OCI и опции VPN для поддержки любой on-premise или не-Oracle облачной интеграции





## Marketplace

## All Applications

## All Applications

Deployed Applications

## Filters

[clear](#)

## TYPE

Any



## PUBLISHER

Oracle



## CATEGORY

Data Integration



## PRICE

Any

ORACLE  
Enterprise Data  
QualityOracle Enterprise Data  
Quality on TomcatA full instance of EDQ, Oracle's  
data quality management...

Type: Image | Price: BYOL

ORACLE  
Enterprise Data  
QualityOracle Enterprise Data  
Quality on WebLogicA full instance of EDQ, Oracle's  
data quality management...

Type: Image | Price: BYOL

ORACLE  
GoldenGateOracle GoldenGate 19c for  
OracleOracle 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

A full instance of ODI, Oracle Data Integrator software

Oracle Data Integrator

Data Integration, Data Integration, Data Integration, Data Integration

[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

[Terms of Use and Privacy](#) [Cookie Preferences](#)

Type

Stack

Version

ODI Marketplace Stack ▾

Compartment

Platform ▾

cissandbox (root)/Platform

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

[Launch Stack](#)

## Support

**Contacts:**

Support hotline ((1) 800.223.1711)

**Links:**<#> <#> <#> <#>

Copyright © 2019, Oracle and/or its affiliates

## Create Stack

- Stack Information
- Configure Variables
- 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 AAAAB3NzaC1yc2EAAAQABQAAQEAj/kmBloC89VzI0qa1ZctvNSbLHct6Hjjv
```

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**

# Готовый к работе ODI

IPSecVPN  
Fast Connect

Compute » Instances » Instance Details

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-gs... [Show](#) [Copy](#)  
**Subnet:** [Public Subnet RJwG:EU-FRANKFURT-1-AD-1](#)

After status *Running* + 20 minutes automatic repository creation

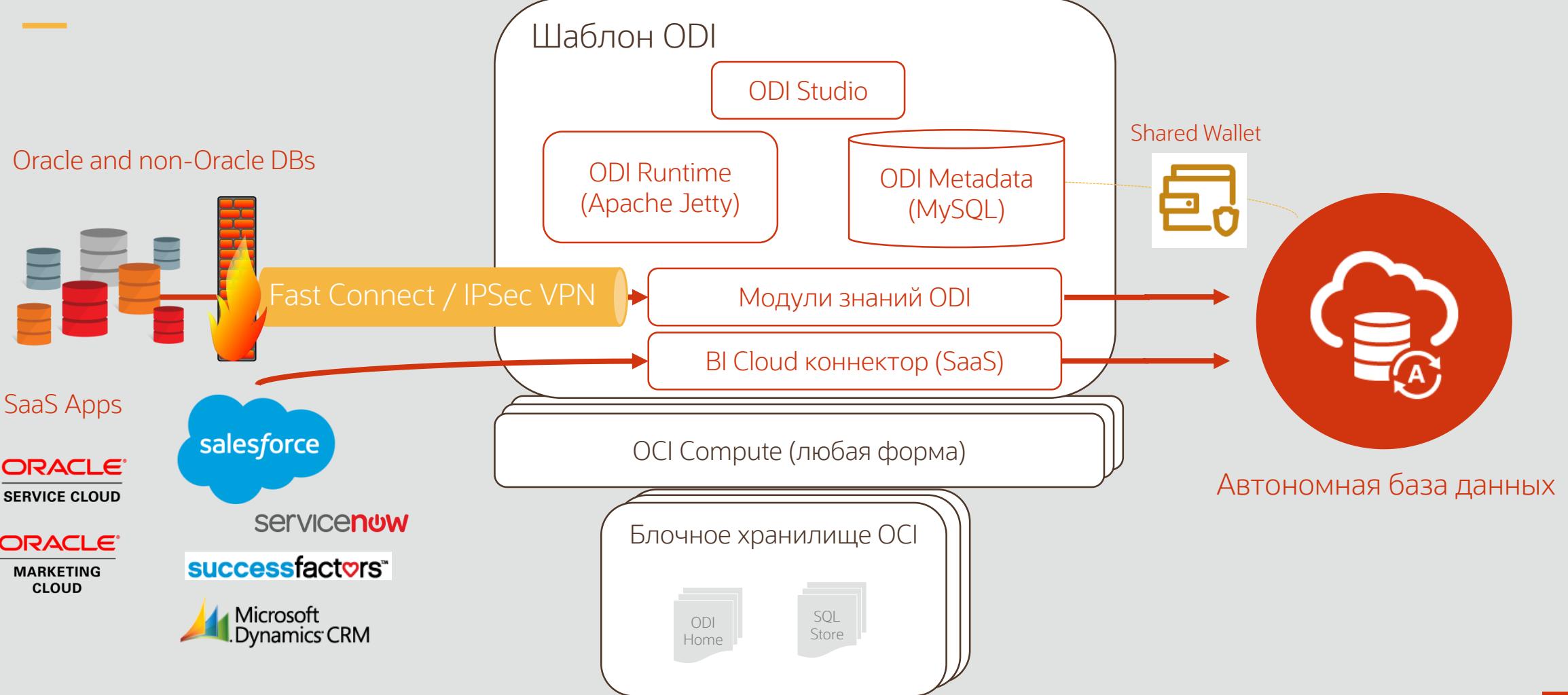


PuTTY  
SSH



Real VNC

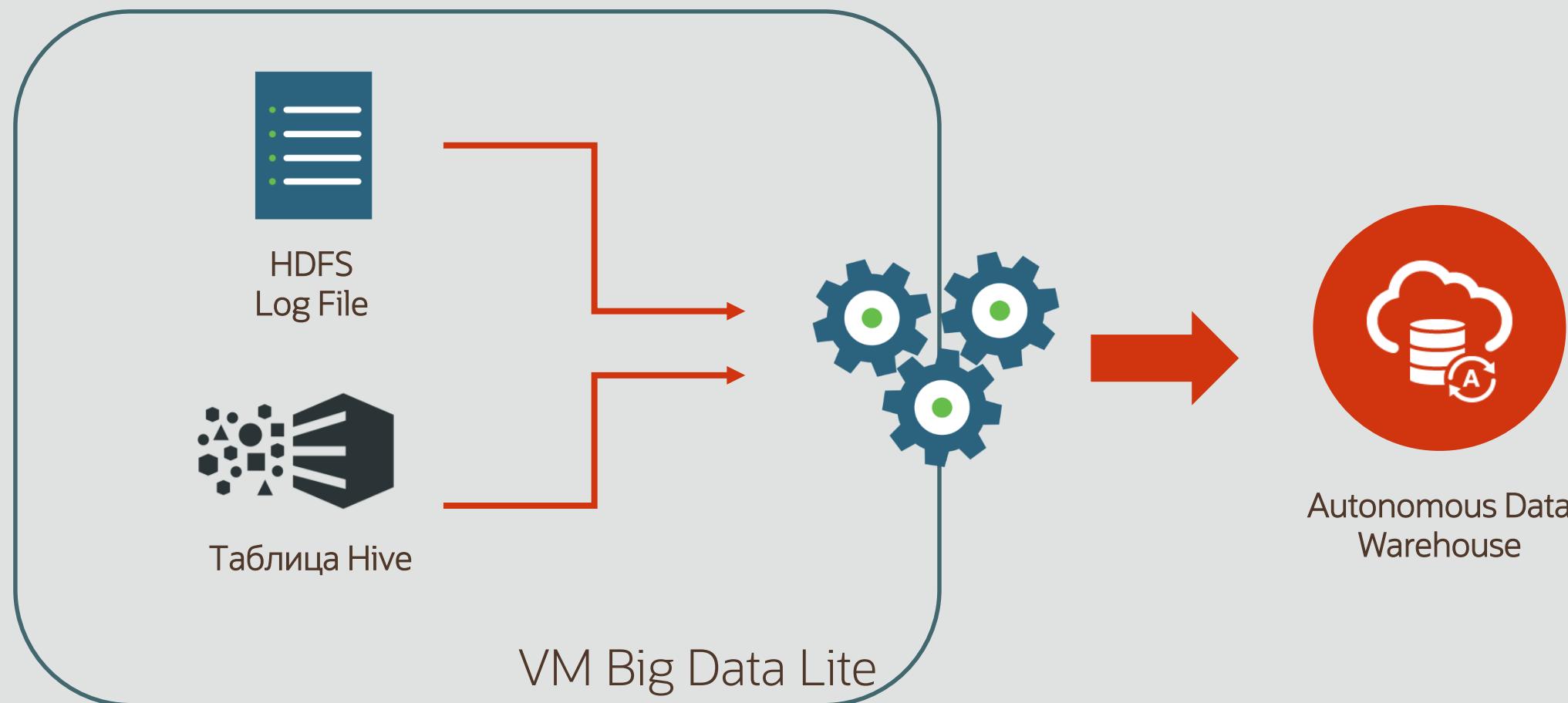
# Архитектура | ODI в маркетплейс



# Пример интеграции

# Пример маппинга

Oracle MoviePlex





oracle

Tue Oct 29, 2:33 PM

File Edit View Search ODI 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

Oracle Data Integrator Studio 12c : ODI Movie Demo



ORACLE

Learn &amp; Explore

Community

What's New

Featured Tutorials

Release Notes

Getting Started Guide

Training Resources

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

A - Load Movies (Sqoop)

DATASTORE: movie\_updates

Issue Details

Severity

Issue

... 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...



Mappings

SUPERVISOR dev\_odi\_repo WORKREP1

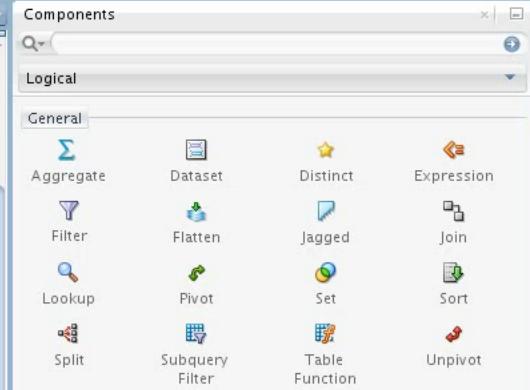
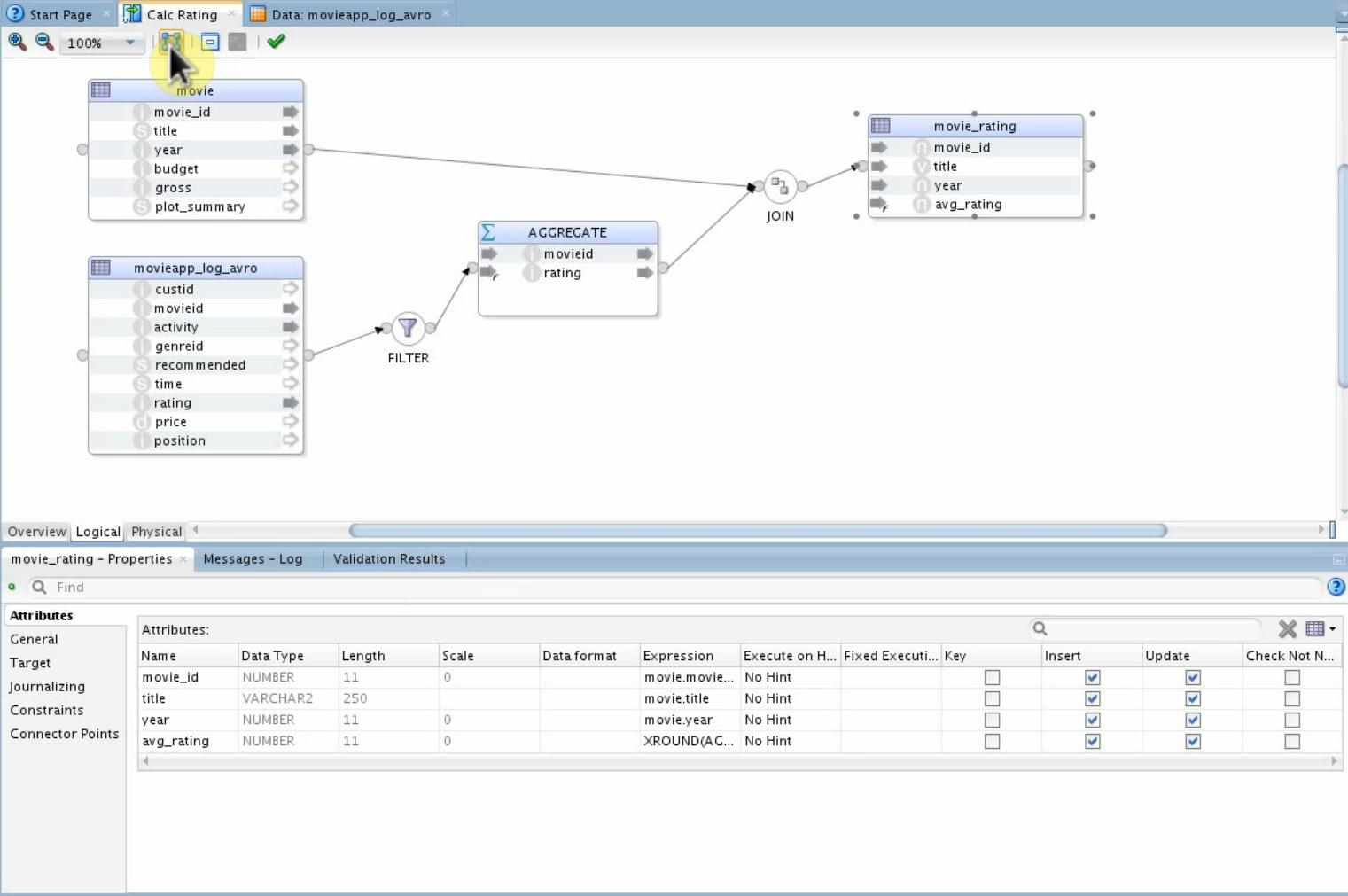


### Physical Architecture

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

- Agents
  - + OracleDIAgent1
  - + Oozie Local

### Oracle Data Integrator Studio 12c : ODI Movie Demo



## 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