

Департамент образования и науки города Москвы  
Государственное автономное образовательное учреждение  
высшего образования города Москвы  
«Московский городской педагогический университет»  
Институт цифрового образования  
Департамент информатики, управления и технологий

ДИСЦИПЛИНА:

«Инструменты для хранения и обработки больших данных»

**пр\_02**

**Тема:**

**«Работа в ETL-системе Talend»**

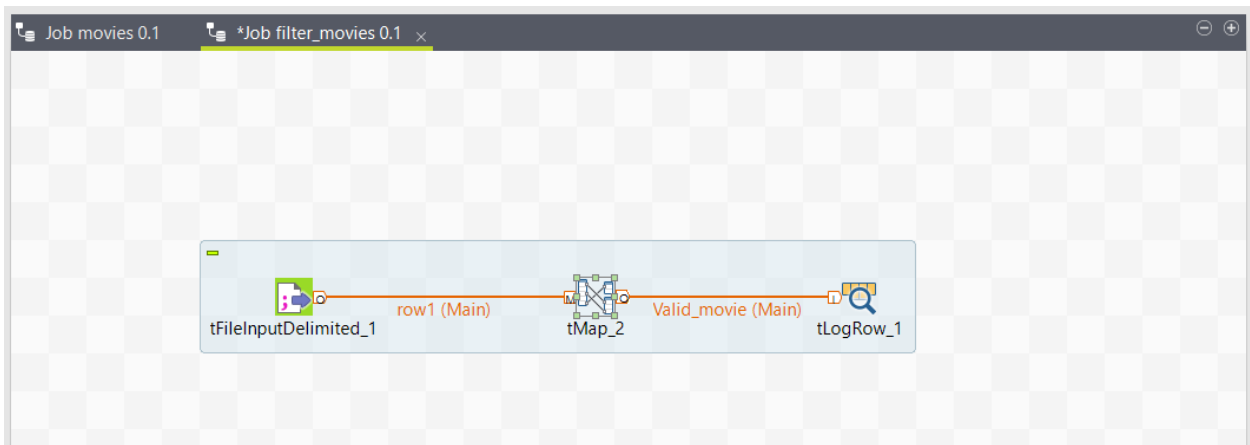
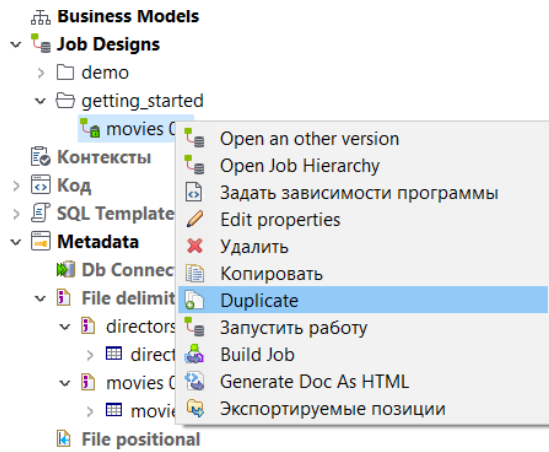
Выполнила: Николаева С.Г., АДЭУ-201

Преподаватель: Босенко Тимур Муртазович

Москва

2023





Talend Open Studio for Data Integration - tMap - tMap\_2

**row1**

Column
movieID
title
releaseYear
url
directorID

**row2**

Property	Value
Lookup Model	Load once
Match Model	Unique match
Join Model	Left Outer Join
Store temp data	false

Expr. key: row1.directorID      Column: directorID, directorName

Find:

Var:

**Valid\_movies**

Выражение	Column
tMap.movieID	movieID
tMap.title	title
tMap.releaseYear	releaseYear
tMap.url	url
tMap.directorID	directorID

**Valid\_movie**

Выражение	Column
	movieID
	title
	releaseYear
	url
	directorID

Options

Inner Join

Left Outer Join

OK Cancel

Schema editor / Редактор выражений

row2

Колонка	K..	Тип	И	N..	Date Pattern (Ctrl+...)	Длина	Precision	Default	Коммента...
directorID		String							
directorName		String				20	0		

Valid\_movies

Колонка	K..	Тип	И	N..	Date Pattern (Ctrl+...)	Длина	Precision	Default	Коммента...
movieID		Integer				4	0		
title		String				72	0		
releaseYear		Integer				4	0		
url		String				120	0		
directorID		String				3	0		

Применить    Ок    Отменить

Find:

Var

row1

Column	
movieID	
title	
releaseYear	
url	
directorID	

row2

Property	Value
Lookup Model	Load once
Match Model	Unique match
Join Model	Inner Join
Store temp data	false

Expr. key

row1.directorID	Column
	directorID
	directorName

Valid\_movies

Выражение	Column
tMap.movieID	movieID
tMap.title	title
tMap.releaseYear	releaseYear
tMap.url	url

Valid\_movie

Выражение	Column
	movieID
	title
	releaseYear
	url
	directorID

Schema editor / Редактор выражений

row2

Колонка	K.	Тип	N.	Date Pattern (Ctrl+...)	Длина	Precision	Default	Коммента...
directorID	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		20	0		
directorName	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>					

Valid\_movies

Колонка	K.	Тип	N.	Date Pattern (Ctrl+...)	Длина	Precision	Default	Коммента...
movieID	<input type="checkbox"/>	Integer	<input checked="" type="checkbox"/>		4	0		
title	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		72	0		
directedBy	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		20			
releaseYear	<input type="checkbox"/>	Integer	<input checked="" type="checkbox"/>		4	0		
url	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		120	0		

Применить Ок Отменить

19:05 15.03.2023

Find:

Var

row1

Column	
movieID	
title	
releaseYear	
url	
directorID	

row2

Property	Value
Lookup Model	Load once
Match Model	Unique match
Join Model	Inner Join
Store temp data	false

Expr. key

row1.directorID	Column
	directorID
	directorName

Valid\_movies

Выражение	Column
row1.movieID	movieID
row1.title	title
row2.directorName	directedBy
row1.releaseYear	releaseYear
row1.url	url

Valid\_movie

Выражение	Column
	movieID
	title
	releaseYear
	url
	directorID

Schema editor / Редактор выражений

row2

Колонка	K.	Тип	N.	Date Pattern (Ctrl+...)	Длина	Precision	Default	Коммента...
directorID	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		20	0		
directorName	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>					

Valid\_movie

Колонка	K.	Тип	N.	Date Pattern (Ctrl+...)	Длина	Precision	Default	Коммента...
movieID	<input type="checkbox"/>	Integer	<input checked="" type="checkbox"/>		4	0		
title	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		72	0		
releaseYear	<input type="checkbox"/>	Integer	<input checked="" type="checkbox"/>		4	0		
url	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		120	0		
directorID	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		3	0		

Применить Ок Отменить

19:13 15.03.2023

Talend Open Studio for Data Integration (7.3.1.20200219\_1130) | practice02 (Соединение: Локально)

Файл Редактировать View Окно Помощь

Repository LOCAL: practice02

- Business Models
  - demo
    - getting\_started
      - filter\_movies 0.1
        - movies 0.1
- Контексты
- Код
- SQL Templates
- Metadata
  - Db Connections
  - File delimited
    - directors 0.1
      - metadata
        - Ссылка(2)
  - movies 0.1
    - File positional
    - File regex
    - File xml
    - File Excel
    - File Idif

- Outline Просмотр кода
- tFileInputDelimited\_1
- tFileInputDelimited\_2 (directors)
- tLogRow\_1
- tMap\_2

Job filter\_movies 0.1

Design: Code

Pa6opa filter\_movies

Execution

Run Kill Очистить

Basic Run

Advanced settings

Target Exec

Memory Run

key value

movieID | 200

title | Shining, The

directorBy | Stanley Kubrick

releaseYear | 1980

url | http://us.imdb.com/M/title-exact?Shining.%20The%20(1980)

[statistics] disconnected

Job filter\_movies ended at 22:20 15/03/2023. [exit code = 0]

Line limit 100 Wrap

Default

Имя Значение

Найти компонент

Favorites

Recently Used

- tMap
- tFileInputDelimited
- tLogRow

Big Data

Business Intelligence

- DB SCD
- Jasper
- OLAP x6
- Splunk

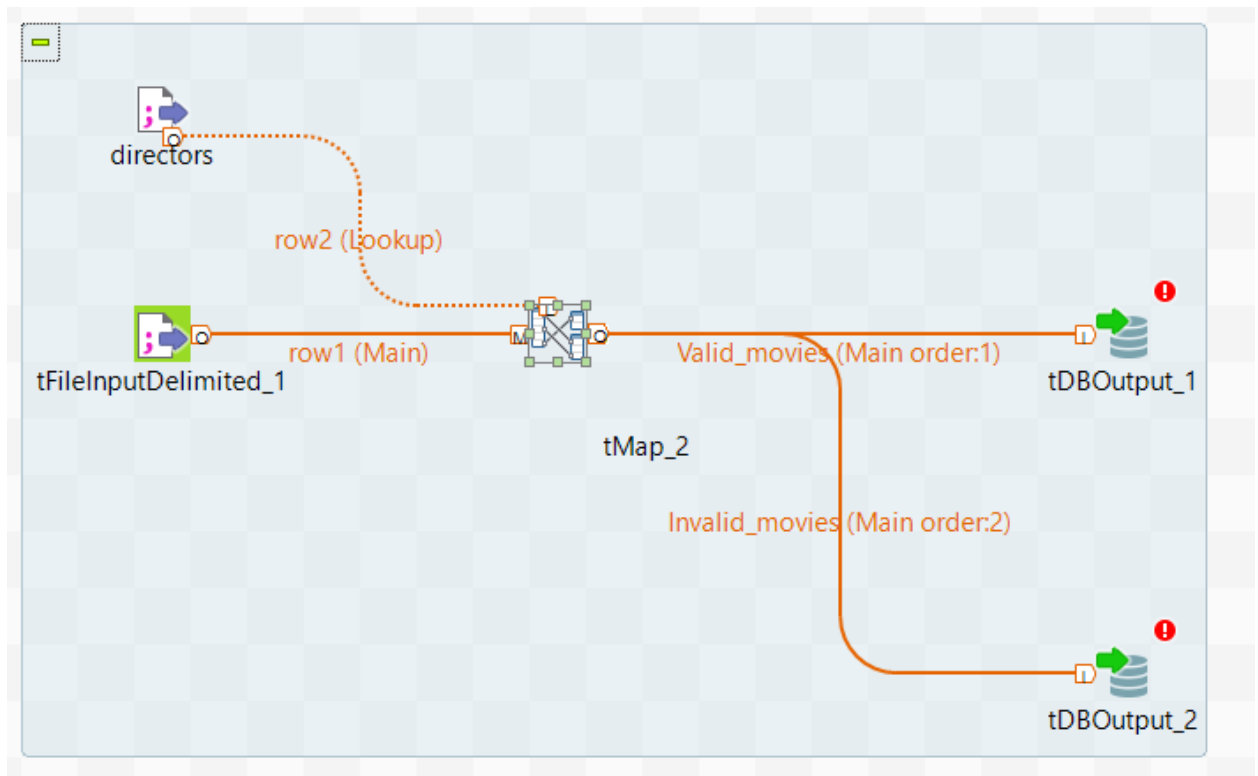
Группы

Business

- JIRA
- Marketo
- NetSuite
- Salesforce

Cloud

- Amazon
- Azure Storage
- Azure
- Box
- Dropbox
- Google Drive
- Google Storage
- Google
- Marketo
- NetSuite
- Salesforce
- ServiceNow
- Snowflake
- iCloudStart



Talend Open Studio for Data Integration - tMap - tMap\_2

Find:

Var:

Auto map!

**row1**

Column
movieID
title
releaseYear
url
directorID

**row2**

Expr. key	Column
row1.directorID	directorID
	directorName

**Valid\_movies**

Выражение	Column
row1.movieID	movieID
row1.title	title
row2.directorName	directorBy
row1.releaseYear	releaseYear
row1.url	url

**Invalid\_movies**

Property	Value
Catch output reject	false
Catch lookup inner join reject	true
Schema Type	Built-In

**Schema editor Редактор выражений**

**row1**

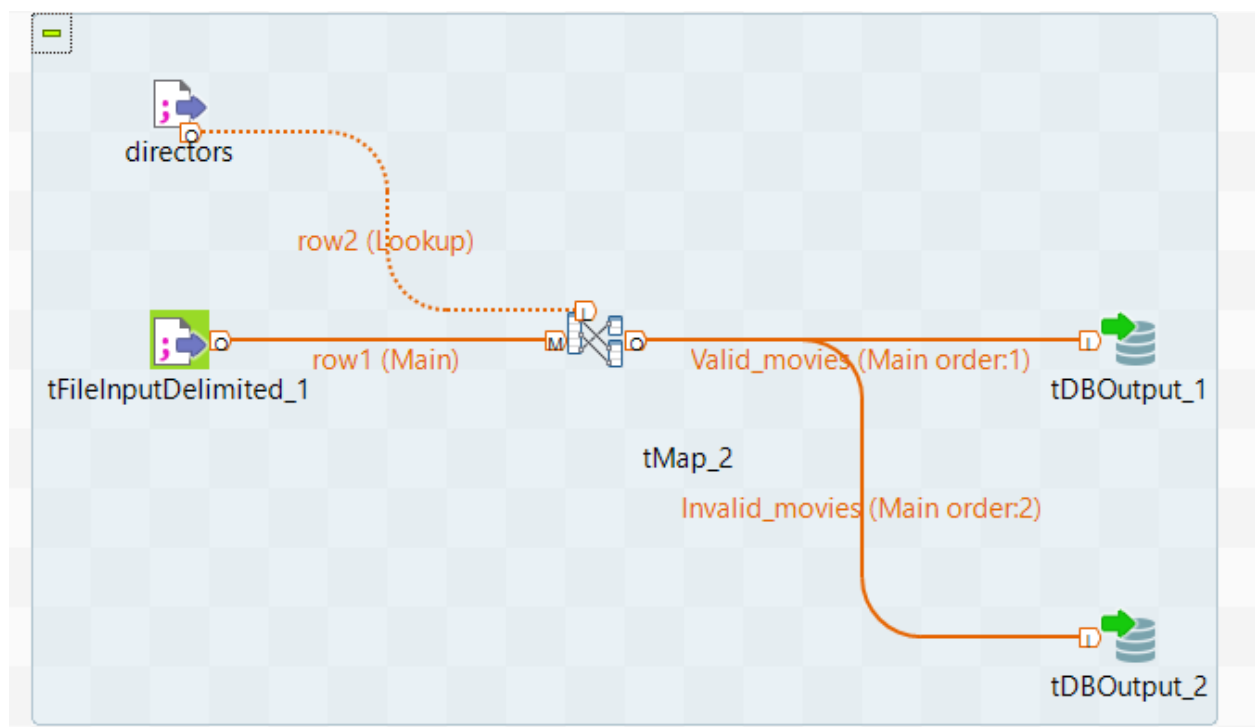
Колонка	K..	Тип	✓ N.	Date Pattern (Ctrl+...)	Длина	Precision	Default	Коммента...
movieID	<input type="checkbox"/>	Integer	<input checked="" type="checkbox"/>		4	0		
title	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		72	0		
releaseYear	<input type="checkbox"/>	Integer	<input checked="" type="checkbox"/>		4	0		
url	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		120	0		
directorID	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		3	0		

**Invalid\_movies**

Колонка	K..	Тип	✓ N.	Date Pattern (Ctrl+...)	Длина	Precision	Default	Коммента...
movieID	<input type="checkbox"/>	Integer	<input checked="" type="checkbox"/>		4	0		
title	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		72	0		

Применить Ок Отменить

22:38 15.03.2023



Talend Open Studio for Data Integration (7.3.1.20200219\_1130) | practice02 (Соединение: Локально)

Файл Редактировать View Проект Окно Помощь Вил

Debug Project Explorer

Job filter\_movies 0.1 Job write\_movies\_to\_db 0.1

Diagram showing data flow from **directors** to **tFileInputDelimited\_1**, then through **row1 (Main)** and **row2 (lookup)** to **tMap\_2**. **tMap\_2** outputs to **Valid\_movies (Main order:1)** and **Invalid\_movies (Main order:2)**, which connect to **tDBOutput\_1** and **tDBOutput\_2** respectively.

**Designer Code**

**tDBOutput\_1(MySQL)**

Basic settings Database: MySQL Apply

Advanced settings Property Type: Built-In

Dynamic settings Версия БД: Mysql 8

View ☐ Использовать существующее соединение

Документация Хост: "localhost" Порт: "3306"

Database: "nikolaevadb"

Имя пользователя: "root" Пароль:  Переменная, прикрепленная к этому параметру: \_USER

Таблица: "valid\_movies"

Action on table: Drop table if exists and create Действие над данными: Обновить

Схема: Built-In Edit schema Sync columns

Talend Open Studio for Data Integration (7.3.1.20200219\_1130) | practice02 (Соединение: Локально)

Файл Редактировать View Проект Окно Помощь Вил

Debug Project Explorer

Job filter\_movies 0.1 Job write\_movies\_to\_db 0.1

Diagram showing data flow from **directors** to **tFileInputDelimited\_1**, then through **row1 (Main)** and **row2 (lookup)** to **tMap\_2**. **tMap\_2** outputs to **Valid\_movies (Main order:1)** and **Invalid\_movies (Main order:2)**, which connect to **tDBOutput\_1** and **tDBOutput\_2** respectively.

**Designer Code**

**tDBOutput\_2(MySQL)**

Basic settings Database: MySQL Apply

Advanced settings Property Type: Built-In

Dynamic settings Версия БД: Mysql 8

View ☐ Использовать существующее соединение

Документация Хост: "localhost" Порт: "3306"

Database: "nikolaevadb"

Имя пользователя: "root" Пароль:  Переменная, прикрепленная к этому параметру: \_USER

Таблица: "invalid\_movies"

Action on table: Drop table if exists and create Действие над данными: Вставить

Схема: Built-In Edit schema Sync columns

127.0.0.1:81 / 127.0.0.1 / nil

127.0.0.1:81/openserver/phpmyadmin/index.php

phpMyAdmin

Сервер: 127.0.0.1:3306 База данных: nikolaevadb

Структура SQL Поиск Запрос по шаблону Экспорт Импорт Операции Ещё

Недавнее Избранное

- Создать БД
- information\_schema
- joomla\_db
- mysql
- nikolaevadb
- performance\_schema
- sys
- wordpress

Таблиц в базе данных не обнаружено.

Создать таблицу

Имя:  Количество столбцов:

Вперёд





127.0.0.181 / 127.0.0.1 / ni

he server time zone value

java - Проблемы с TimeZ

set global timezone open

timezone - How do I set ti

+

127.0.0.181/opensever/phpmyadmin/index.php

phpMyAdmin

НадавниеИзбранное

Создать БД

information\_schema

joomla\_db

mysql

nikolaevadb

Новая

invalid\_movies

valid\_movies

performance\_schema

sys

wordpress

Сервер: 127.0.0.1.3306База данных: nikolaevadbТаблица: valid\_movies

ОбзорСтруктураSQLПоискВставитьЭкспортИмпортПривилегииОперацииТриггеры

⚠ Данное выделение не содержит уникального столбца. Изменение сетки, выставление галочки, редактирование, копирование и удаление невозможно.

✔ Отображение строк 0 - 49 (255 всего, Запрос занял 0,0009 сек.)

SELECT \* FROM `valid\_movies`

ПрофилированиеПострочное редактированиеИзменитьАнализ SQL запросаСоздать PHP-кодОбновить

Показать все

Количество строк: 50

Фильтровать строки

Поиск в таблице

Параметры

movieID	title	directorBy	releaseYear	url
315	Apt Pupil	Bryan Singer	1998	http://us.imdb.com/Title?Apt+Pupil+(1998)
1294	Ayn Rand: A Sense of Life	Michael Paxton	1998	http://us.imdb.com/Title?Ayn+Rand%3A+A+Sense+of+Li...
1679	B. Monkey	Michael Radford	1998	http://us.imdb.com/M/title-exact?B%2E+Monkey+(1998...
1649	Big One, The	Michael Moore	1998	http://us.imdb.com/Title?Big+One,+The+(1997)
362	Blues Brothers 2000	John Landis	1998	http://us.imdb.com/M/title-exact?Blues+Brothers+20...
1645	Butcher Boy, The	Neil Jordan	1998	http://us.imdb.com/M/title-exact?mdb-title-118804
1650	Butcher Boy, The	Neil Jordan	1998	http://us.imdb.com/M/title-exact?mdb-title-118804
1234	Chairman of the Board	Alex Zamm	1998	http://us.imdb.com/Title?Chairman+of+the+Board+(19...
1654	Chairman of the Board	Alex Zamm	1998	http://us.imdb.com/Title?Chairman+of+the+Board+(19...
918	City of Angels	Brad Silberling	1998	http://us.imdb.com/Title?City+of+Angels+(1998)
909	Dangerous Beauty	Marshall Herskovitz	1998	http://us.imdb.com/M/title-exact?mdb-title-118892
691	Dark City	Alex Proyas	1998	http://us.imdb.com/M/title-exact?mdb-title-118929
353	Deep Rsing	Stephen Sommers	1998	http://us.imdb.com/M/title-exact?mdb-title-118956
329	Desperate Measures	Barbet Schroeder	1998	http://us.imdb.com/Title?Desperate+Measures+(1998)
344	Desperate Measures	Barbet Schroeder	1998	http://us.imdb.com/Title?Desperate+Measures+(1998)

Консоль

2303

15.03.2023

ENG

4