

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

Практическая работа №2

Тема:

Работа в ETL-системе Talend

Выполнила: Шепелева Е. В., группа: АДЭУ-201

Преподаватель: Т. М. Босенко

Москва

2022

Designer | Code

Работа(Job movies 0.1) Contexts(movies) Компонента Run (Работа movies)

**tFileInputDelimited\_1**

**Basic settings**

Property Type: Repository DELIM:movies

Схема: Repository DELIM:movies - movies\_schema Edit schema

"When the input source is a stream or a zip file, footer and random shouldn't be bigger than 0."

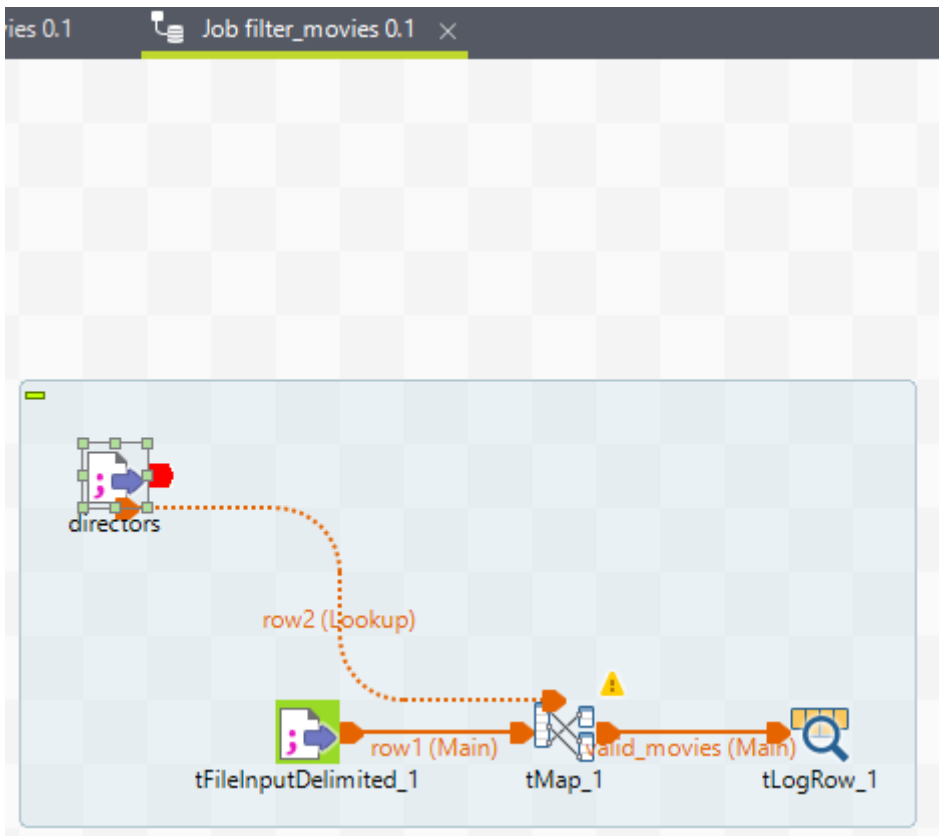
File name/Stream: "C:/Users/10liz/Desktop/input\_data/movies.csv"

Разделитель строк: "\n" Разделитель полей: ";"

☐ CSV options

Header: 1 Footer: 0

☐ Skip empty rows ☐ Uncompress as zip file ☐ Die on error



row1

Column

movieID

title

releaseYear

url

directorID

row2

Property

Lookup Model

Match Model

Join Model

Store temp data

Expr. key

row1.directorID

Value

Load once

Unique match

Left Outer Join

false

Column

directorID

directorName

Find :

Var

valid\_movies

Выражение

row1.movieID

row1.title

row2.directorName

row1.releaseYear

row1.url

Column

movieID

title

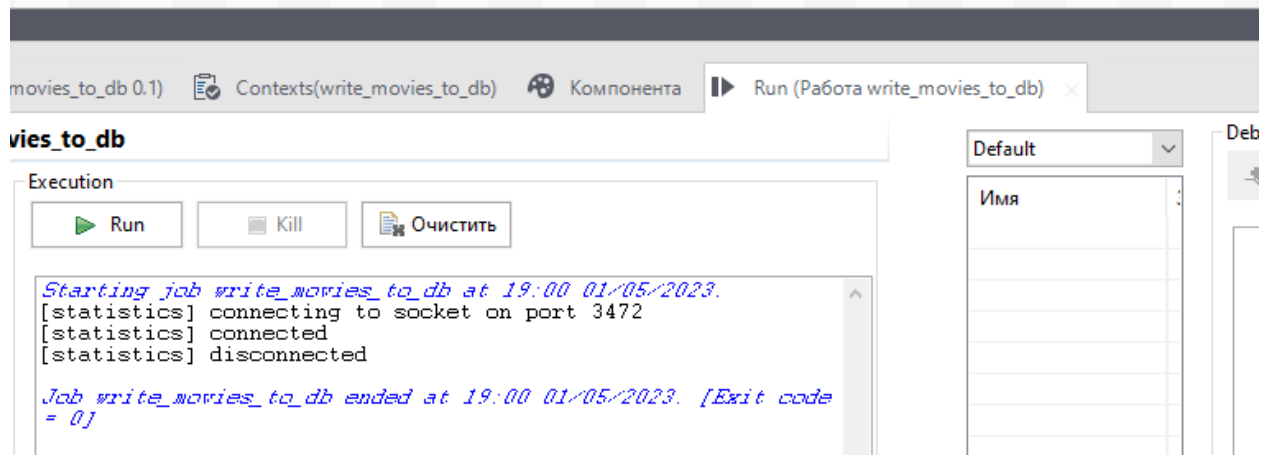
directedBy

releaseYear

url

Schema editor Редактор выражений

row1										valid_movies									
Колонка	Кл...	Тип	<input checked="" type="checkbox"/>	N..	Date Pattern ...	Длина	Precisi...	Defa...	Комм...	Колонка	Кл...	Тип	<input checked="" type="checkbox"/>	N..	Date Pattern ...	Длина	Precisi...	Defa...	Комм...
movieID		Integer	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		4	0			movieID		Integer	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		4	0		
title		String	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		72	0			title		String	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		72	0		
releaseYear		Integer	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		4	0			directedBy		String	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		20			
url		String	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		120	0			releaseYear		Integer	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		4	0		
directorID		String	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		3	0			url		String	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		120	0		



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

Column
directorID
directorName

valid\_movies

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

invalid\_movies

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

Выражение

Column
movieID
title

Schema editor Редактор выражений

row1

Колонка	Кл...	Тип	✓	N...	Date Pattern ...	Длина	Precisi...	Defa...	Комм...
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		

valid\_movies

Колонка	Кл...	Тип	✓	N...	Date Pattern ...	Длина	Precisi...	Defa...	Комм...
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		

phpMyAdmin

Сервер: 127.0.0.1:3306 / База данных: gettingstarted / Таблица: valid\_movies

Обзор Структура SQL Поиск Вставить Экспорт Импорт Привилегии Операции Триггеры

Недавнее Избранное

Создать БД  
gettingstarted  
Новая  
invalid\_movies  
valid\_movies  
information schema  
mysql  
performance schema  
sys

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

✓ Отображение строк 0 - 49 (255 всего, Запрос занял 0,0014 сек.)

SELECT \* FROM `valid\_movies`

☐ Профилирование  
[ Построчное редактирование ] [ Изменить ] [ Анализ SQL запроса ] [ Создать PHP-код ] [ Обновить ]

1 > >> ☐ Показать все Количество строк: 50 Фильтровать строки:

Extra options

movieID	title	directedBy	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?imdb-title-118804
1650	Butcher Boy, The	Neil Jordan	1998	http://us.imdb.com/M/title-exact?imdb-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?imdb-title-118892
691	Dark City	Alex Proyas	1998	http://us.imdb.com/M/title-exact?imdb-title-118929
353	Deep Rising	Stephen Sommers	1998	http://us.imdb.com/M/title-exact?imdb-title-118956

**phpMyAdmin**

Сервер: 127.0.0.1:3306 / База данных: gettingstarted / Таблица: invalid\_movies

Обзор Структура SQL Поиск Вставить Экспорт Импорт Привилегии Операции Триггеры

Недавнее Избранное

Создать БД  
gettingstarted  
Новая  
invalid\_movies  
valid\_movies  
information\_schema  
mysql  
performance\_schema  
sys

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

✓ Отображение строк 0 - 49 (1200 всего, Запрос занял 0,0007 сек.)

`SELECT * FROM `invalid_movies``

☐ Профилирование  
[ Построчное редактирование ] [ Изменить ] [ Анализ SQL запроса ] [ Создать PHP-код ] [ Обновить ]

1 > >> ☐ Показать все Количество строк: 50 Фильтровать строки:

Extra options

movieID	title
1591	Duoluo tianshi
905	Great Expectations
349	Hard Rain
1612	Leading Man, The
913	Love and Death on Long Island
1592	Magic Hour, The
1678	Mat' i syn
1313	Palmetto
1651	Spanish Prisoner, The
1613	Tokyo Fist
1300	'Til There Was You
1643	Angel Baby
1382	Bonheur, Le