

Министерство образования Республики Беларусь

Учреждение образования
БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИНФОРМАТИКИ И РАДИОЭЛЕКТРОНИКИ

Инженерно-экономический факультет
Кафедра экономической информатики

Отчёт
по дисциплине
«Системы и технологии интеллектуальной обработки данных»
Лабораторная работа №1
КОНСОЛИДАЦИЯ ДАННЫХ В ETL

Выполнила:

студентка гр.772302
Ковальчук О. Ю.

Проверил:

Кунцевич А. А.

Минск 2020

Задание:

1. Создать CSV файл. Используя Talend, написать сценарий по генерации из этого файла другого CSV файла.
2. Создать Excel файл. Используя Talend, написать сценарий по генерации из этого файла другого Excel файла.
3. Создать Excel файл. Используя элемент tMap Talend(a), написать сценарий по генерации другого Excel файла с заданной структурой полей. Например, некоторые поля одного файла объединены в одно поле другого файла. Таких объединений должно быть 2.

Выполнение:

– Создадим исходный файл in.csv:

	A	B	C	D	E	F	G	H	I
1	id	surname	name	phone	date	town	street	house	passport
2	1	Ковальчук	Ольга	6431212	2000-04-21	Баранови	Платонова	32	AB0329789
3	2	Лицкевич	Алина	4942323	2000-07-05	Минск	Немига	21	AB4532132
4	3	Болбас	Валерия	1338385	1989-10-15	Гродно	Независимости	12	AB3542174
5	4	Тихонович	Наталия	3692581	2000-08-21	Витебск	Красноармейская	45	AB5739292
6	5	Морхат	Глеб	1232323	1999-09-20	Брест	Луговая	12	AB4984773
7	6	Шукайло	Кирилл	4577575	1999-12-23	Минск	Дачная	65	AB4737292
8	7	Щечин	Валера	3787902	1981-04-14	Могилев	Притыцкого	87	AB3484983
9	8	Яницкая	Валерия	2432531	1996-03-12	Гомель	Свердлова	47	AB4888002
10	9	Мороз	Влад	9750540	2002-05-09	Молодеч	Сурганова	34	AB3884739
11	10	Ширко	Никита	7420222	2004-01-24	Солигорск	Ленина	85	AB4636279
12	11	Саванец	Юрий	1234432	1993-10-23	Баорисов	Ромейко	13	AB2363248
13	12	Солтан	Влада	9663366	1988-10-25	Береза	Жукова	115	AB4848322
14	13	Ерш	Павел	2387322	2004-08-11	Пинск	Победителей	76	AB3263131
15	14	Жук	Карина	1233324	2002-06-10	Бобруйск	Павлова	92	AB1233765
16	15	Парханович	Александра	1233455	2000-06-23	Баранови	Гагарина	7	AB3657267

– Создадим сценарий преобразования in.csv в out.csv и выполним работу:

Talend Open Studio for Data Integration (6.1.0.20150902_1604-M2) | lbwrk1 (Соединение: Локально)

Файл Редактировать View Окно Помощь

Learn Ask Upgrade! Exchange

Reposit... LOCAL: lbwrk1

Business Models

Job Designs

job_1 0.1

Контексты

Код

SQL Templates

Metadata

Documentation

Recycle bin

Talend Open Studio *Job job_1 0.1 x

15 rows in 0,04s 15 rows in 0,14s

tFileInputDelimited_1 tLogRow_1 tFileOutputDelimited_1

Designer Code

Работа(Job job_1 0.1) Contexts(job_1) Компонента ▶ Run (Работа job_1)

Работа job_1

Basic Run

Debug Run

Advanced settings

Target Exec

Memory Run

Execution

Run Kill ОЧИСТИ

[statistics] connecting to socket on port 3000
[statistics] connected

tLogRow_1

id	surname	name	phone	date	town	street	house	passport
1	Ковальчук	Ольга	6431212	2000-04-21	Барановичи	Платонова	32	AB0329789
2	Лицкевич	Алина	4942323	2000-07-05	Минск	Немига	21	AB4532132
3	Болбас	Валерия	1338385	1989-10-15	Гродно	Независимости	12	AB3542174
4	Тихонович	Наталия	3692581	2000-08-21	Витебск	Красноармейская	45	AB5739292
5	Морхат	Глеб	1232323	1999-09-20	Брест	Луговая	12	AB4984773
6	Шукайло	Кирилл	4577575	1999-12-23	Минск	Дачная	65	AB4737292
7	Щечин	Валера	3787902	1981-04-14	Могилев	Притыцкого	87	AB3484983
8	Яницкая	Валерия	2432531	1996-03-12	Гомель	Свердлова	47	AB4888002
9	Мороз	Влад	9750540	2002-05-09	Молодечно	Сурганова	34	AB3884739
10	Ширко	Никита	7420222	2004-01-24	Солигорск	Ленина	85	AB4636279
11	Саванец	Юрий	1234432	1993-10-23	Баорисов	Ромейко	13	AB2363248
12	Солтан	Влада	9663366	1988-10-25	Береза	Жукова	115	AB4848322
13	Ерш	Павел	2387322	2004-08-11	Пинск	Победителей	76	AB3263131
14	Жук	Карина	1233324	2002-06-10	Бобруйск	Павлова	92	AB1233765
15	Парханович	Александра	1233455	2000-06-23	Барановичи	Гагарина	7	AB3657267

Line limit 100 Wrap

– Создадим исходный файл in.xls:

	A	B	C	D	E	F	G	H	I
1	id	surname	name	phone	date	town	street	house	passport
2	1	Ковальчук	Ольга	6431212	2000-04-21	Баранови	Платонова	32	AB0329789
3	2	Лицкевич	Алина	4942323	2000-07-05	Минск	Немига	21	AB4532132
4	3	Болбас	Валерия	1338385	1989-10-15	Гродно	Независимости	12	AB3542174
5	4	Тихонович	Наталия	3692581	2000-08-21	Витебск	Красноармейская	45	AB5739292
6	5	Морхат	Глеб	1232323	1999-09-20	Брест	Луговая	12	AB4984773
7	6	Шукайло	Кирилл	4577575	1999-12-23	Минск	Дачная	65	AB4737292
8	7	Щечин	Валера	3787902	1981-04-14	Могилев	Притыцкого	87	AB3484983
9	8	Яницкая	Валерия	2432531	1996-03-12	Гомель	Свердлова	47	AB4888002
10	9	Мороз	Влад	9750540	2002-05-09	Молодеч	Сурганова	34	AB3884739
11	10	Ширко	Никита	7420222	2004-01-24	Солигорск	Ленина	85	AB4636279
12	11	Саванец	Юрий	1234432	1993-10-23	Баорисов	Ромейко	13	AB2363248
13	12	Солтан	Влада	9663366	1988-10-25	Береза	Жукова	115	AB4848322
14	13	Ерш	Павел	2387322	2004-08-11	Пинск	Победителей	76	AB3263131
15	14	Жук	Карина	1233324	2002-06-10	Бобруйск	Павлова	92	AB1233765
16	15	Парханович	Александра	1233455	2000-06-23	Баранови	Гагарина	7	AB3657267

– Создадим сценарий преобразования in.xls в out.xls и настроим свойства элемента map:

Talend Open Studio for Data Integration - tMap - tMap_1

Find:

Var

Auto map!

row3

Column

- id
- surname
- name
- phone
- date
- town
- street
- house
- passport

out

Выражение

- row3.id
- row3.surname + " " + row3.name
- row3.phone
- row3.date
- row3.town + " " + row3.street + " " + row3.house
- row3.passport

Column

- id
- name
- phone
- date
- address
- passport

Schema editor Редактор выражений

row3

Колонка	K...	Тип	N...	Date Pattern (Ctrl+...	Длина	Precision	Default	Коммент...
id	<input checked="" type="checkbox"/>	Integer	<input checked="" type="checkbox"/>					
surname	<input checked="" type="checkbox"/>	String	<input checked="" type="checkbox"/>					
name	<input checked="" type="checkbox"/>	String	<input checked="" type="checkbox"/>					
phone	<input checked="" type="checkbox"/>	Integer	<input checked="" type="checkbox"/>					
date	<input checked="" type="checkbox"/>	Date	<input checked="" type="checkbox"/>	"yyyy-MM-dd"				
town	<input checked="" type="checkbox"/>	String	<input checked="" type="checkbox"/>					

out

Колонка	K...	Тип	N...	Date Pattern (Ctrl+...	Длина	Precision	Default	Коммент...
id	<input checked="" type="checkbox"/>	Integer	<input checked="" type="checkbox"/>					
name	<input checked="" type="checkbox"/>	String	<input checked="" type="checkbox"/>					
phone	<input checked="" type="checkbox"/>	Integer	<input checked="" type="checkbox"/>					
date	<input checked="" type="checkbox"/>	Date	<input checked="" type="checkbox"/>	"yyyy-MM-dd"				
address	<input checked="" type="checkbox"/>	String	<input checked="" type="checkbox"/>					
passport	<input checked="" type="checkbox"/>	String	<input checked="" type="checkbox"/>					

Применить Ок Отменить

