

## BAB IV

### FILE TEKS (LANJUTAN)

#### 4.3 Membaca File Tidak Terstruktur

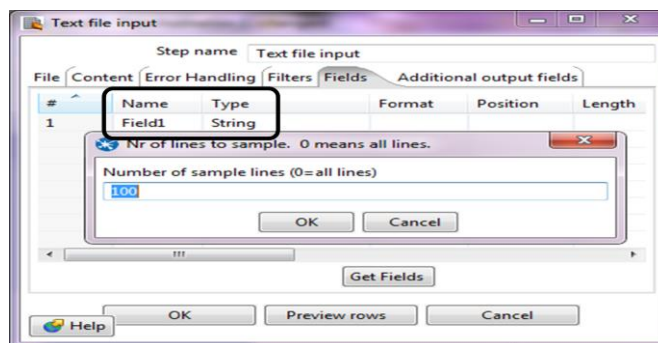
1. Buatlah direktori D:\Latihan Kettle\latihan...
2. Duplikasikan file Propinsi di Pulau Kalimantan.txt
3. Buatlah sebuah transformation baru dengan klik **File > New > Transformation**, atau dengan shortcut **Ctrl+N**. Simpan dengan nama Membaca File Teks Yang Tidak Terstruktur.ktr

Untuk selanjutnya lakukan penyimpanan setiap ada perubahan dengan **File > Save** atau **Ctrl+S**.

4. Bawa step **Text file input** ke ruang kerja.
5. Lakukan seperti pada latihan sebelumnya untuk menyunting file sumber.
6. Pindah ke tab **Content**, tentukan properties berikut :
  - **File type** : CSV.
  - **Header** : Matikan properties ini karena file sumber yang akan diolah tidak memiliki header.
  - **No empty row** : Matikan properties ini agar step **Text file input** juga menganggap baris kosong sebagai data. Kita perlu baris kosong ini untuk digunakan sebagai indikator pemisah data.

Biarkan properties lainnya apa adanya.

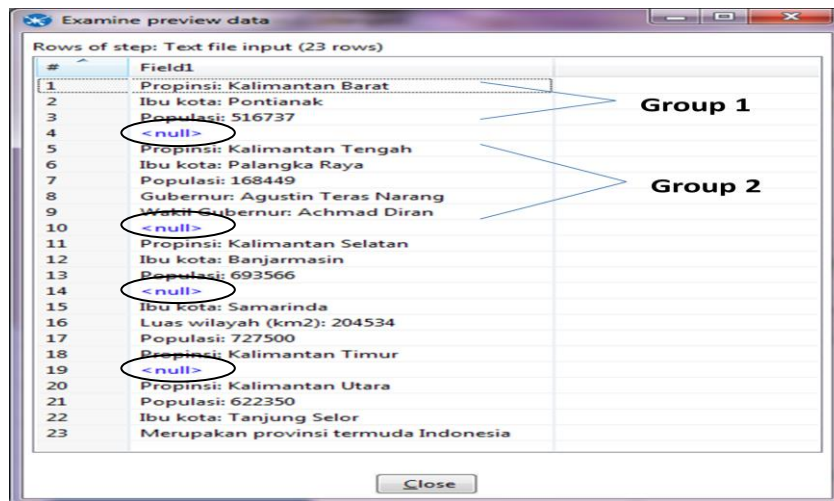
7. Pindah ke tab **Fields**, dan tekan tombol **Get Fields**. Step **Text file input** hanya akan mendefinisikan satu buah field, yaitu **Field1**, dengan tipe String, seperti tampilan berikut :



Tekan tombol **OK** pada window **Nr of lines to sample**.

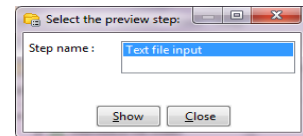
Tekan tombol **Close** pada window **Scan Result**.

8. Lakukan preview data pada step **Text file input** dan hasilnya akan tampil seperti gambar berikut :



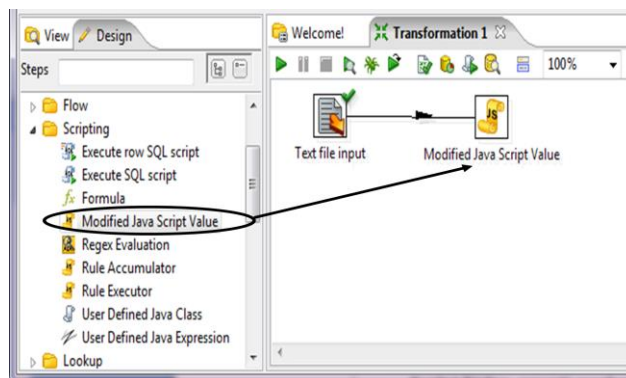
Tekan tombol **Close** untuk menutup window.

Jika muncul tampilan di samping, tekan tombol **Close**.

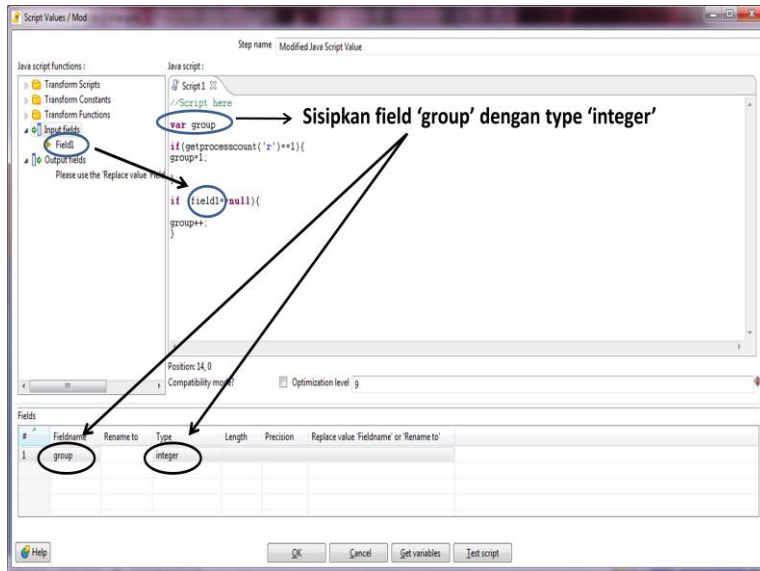


Setiap grup data dipisahkan oleh **null**. Tutup window **Examine preview data**.

9. Bawa step **Modified Java Script Value** yang ada pada grup **Scripting** ke ruang kerja dan buatlah hop antara step **Text file input** dan **Modified Java Script Value**, seperti gambar berikut :



10. Sunting properties dari step **Modified Java Script Value**, seperti pada gambar berikut :



Pada text box **Script1**, masukkan kode JavaScript berikut :

```

var group

if (getprocesscount('r')==j){
group=j;
};

if (fieldj==null){
group++;
}

```

Tekan tombol **OK** untuk menutup window **Modified Java Script Properties**.

- Lakukan preview data pada Modified Java Script Properties untuk melihat baris data yang keluar dari step tersebut. Pastikan **Modified Java Script Properties** sudah dipilih sebelum menekan tombol **Preview this transformation**. Tekan tombol **Quick Launch**.

Tampilan row data keluaran dari **Modified Java Script Value** seperti berikut :

Examine preview data

Rows of step: Modified Java Script Value (23 rows)

#	Field1	group
1	Propinsi: Kalimantan Barat	1
2	Ibu kota: Pontianak	1
3	Populasi: 516737	1
4	<null>	2
5	Propinsi: Kalimantan Tengah	2
6	Ibu kota: Palangka Raya	2
7	Populasi: 168449	2
8	Gubernur: Agustin Teras Narang	2
9	Wakil Gubernur: Achmad Diran	2
10	<null>	3
11	Propinsi: Kalimantan Selatan	3
12	Ibu kota: Banjarmasin	3
13	Populasi: 693566	3
14	<null>	4

Buttons: Close

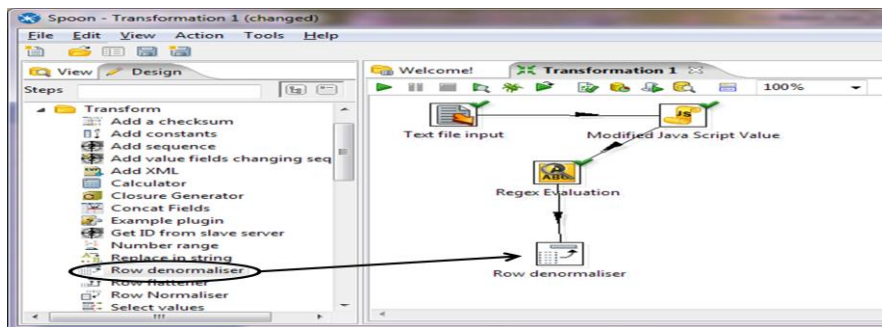


Properties lainnya biarkan apa adanya, tekan tombol **OK** untuk menutup window step **Regex Evaluation**.

14. Untuk melihat hasil baris data keluaran dari step **Regex Evaluation** lakukan preview, hasilnya akan tampil seperti berikut :

#	Field1	group	result	Key	Value
1	Propinsi: Kalimantan Barat	1	Y	Propinsi	Kalimantan Barat
2	Ibu kota: Pontianak	1	Y	Ibu kota	Pontianak
3	Populasi: 516737	1	Y	Populasi	516737
4	<null>	2	N	<null>	<null>
5	Propinsi: Kalimantan Tengah	2	Y	Propinsi	Kalimantan Tengah
6	Ibu kota: Palangka Raya	2	Y	Ibu kota	Palangka Raya
7	Populasi: 168449	2	Y	Populasi	168449
8	Gubernur: Agustin Teras Narang	2	Y	Gubernur	Agustin Teras Narang
9	Wakil Gubernur: Achmad Diran	2	Y	Wakil Gubernur	Achmad Diran
10	<null>	3	N	<null>	<null>
11	Propinsi: Kalimantan Selatan	3	Y	Propinsi	Kalimantan Selatan
12	Ibu kota: Banjarmasin	3	Y	Ibu kota	Banjarmasin
13	Populasi: 693566	3	Y	Populasi	693566
14	<null>	4	N	<null>	<null>

15. Yang kita perlukan adalah baris data dengan fields : **Propinsi**, **IbuKota** dan **Populasi**. Sedangkan baris yang keluar dari step **Regex Evaluation** : **Field1**, **group**, **result**, **Key** dan **Value**. Merubah isi kolom **Key** menjadi nama **Field** dengan menggunakan **Row Denormaliser** yang ada dalam grup **Transform** dan buat hop antara **Regex Evaluation** dan **Row Denormaliser** seperti tampilan berikut :



16. Sunting properties dari step **Row Denormaliser**, seperti pada tampilan berikut :

#	Target fieldname	Value fieldname	Key value	Type	Format
1	Propinsi	Value	Propinsi	String	
2	Ibu Kota	Value	Ibu Kota	String	
3	Populasi	Value	Populasi	String	

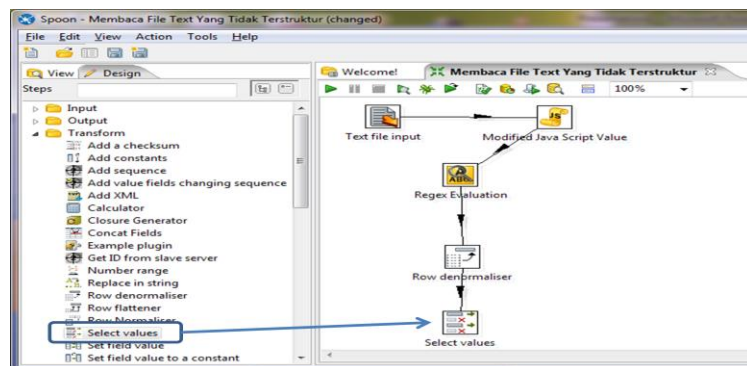
Tekan tombol **OK** untuk menutup window properties step **Row Denormaliser**. Jika muncul window **Warning**, tekan tombol **I Understand**. Warning tersebut muncul karena step **Row Denormaliser** memerlukan data yang sudah terurut berdasarkan field kunci.

17. Lakukan preview pada step **Row Denormaliser**, akan muncul seperti tampilan berikut :

#	Field1	group	result	Propinsi	Ibu kota	Populasi
1	Populasi: 516737	1	Y	Kalimantan Barat	Pontianak	516737
2	Wakil Gubernur: Achmad Diran	2	Y	Kalimantan Tengah	Palangka Raya	168449
3	Populasi: 693566	3	Y	Kalimantan Selatan	Banjarmasin	693566
4	Propinsi: Kalimantan Timur	4	Y	Kalimantan Timur	Samarinda	727500
5	Merupakan provinsi termuda Indonesia	5	N	Kalimantan Utara	Tanjung Selor	622350

Dari tampilan di atas tidak ada field **Key** dan **Value** lagi, sudah menjadi 1 field baru yaitu **Propinsi**, **Ibu Kota** dan **Populasi**.

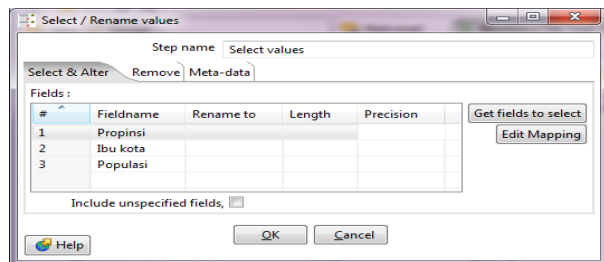
18. Field yang diperlukan sudah diperoleh, namun masih terdapat field lain yang pada dasarnya sudah tidak diperlukan. Untuk menyeleksi field gunakan step **Select Value**. Bawa step **Select Value** yang ada dalam grup **Transform** ke ruang kerja, buat hop dari step **Row Denormaliser** ke step **Select Value**, seperti gambar berikut :



19. Sunting properties dari step **Select Value**. Pada tab **Select & Alter** tekan tombol **Get fields to select**, field dari baris data masukan akan muncul pada grid **Fields**.

#	Fieldname	Rename to	Length	Precision
1	Field1			
2	group			
3	result			
4	Propinsi			
5	Ibu kota			
6	Populasi			

Ada 1 fields yang harus dibuang. Tekan tombol **Ctrl** pada keyboard, lalu tekan **field1**, **group** dan **result** satu per satu, sehingga latar belakang ketiganya berubah menjadi biru muda. Lepaskan tombol **Ctrl**. Tekan tombol **Del** pada keyboard. Ketiga field yang dipilih akan hilang dari grid **Fields**. Seperti pada gambar berikut :



20. Lakukan preview pada step **Select Value**, hasilnya akan seperti gambar berikut :

Lalu tutup window **Examine preview data**.

