Business Intelligence

Modul Praktikum

# File Teks sebagai data Input pada *Kettle Pentaho Data Integration*

## Topik

Membaca file teks terstruktur tipe CSV dan tipe Fixed sebagai data masukan.

## Tujuan

Setelah menyelesaikan praktikum ini, mahasiswa diharapkan mampu:

1. Memahami penggunaan *variable* padaPentaho Data Integration.
2. Mampu menggunakan komponen *step* ***CSV file input*** untuk membaca file teks tipe CSV.
3. Mampu menggunakan komponen *step* ***Fixed file input*** untuk membaca file teks berisi data fixed width.

## Alat dan Bahan

* Komputer dengan **Pentaho Data Integration**  terinstall
* Pentaho Data Integration dapat diperoleh di: <http://community.pentaho.com/>

## Rangkuman/Teori

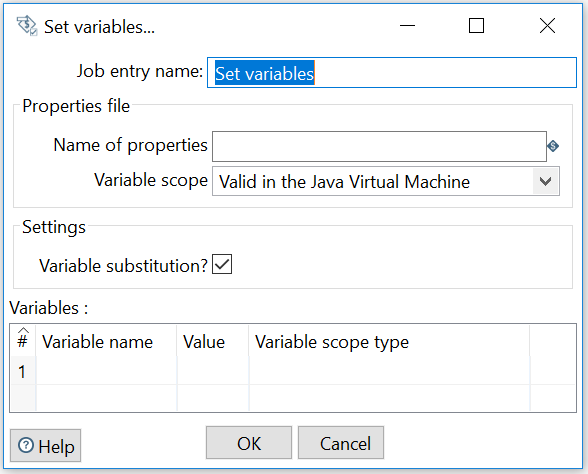
[Menjelaskan *step* (*job entry*) ***Set Variables***, *step* ***Text file input*** untuk membaca file teks **csv** dan file jenis ***fixed width*** (**tab separated**) dan file teks jenis berisi **data tidak terstruktur**, *step* ***Modified Java Script Value***]

### Step Set Variables

Set variables merupakan step (*job entry*) yang berada dalam grup **General**, digunakan untuk membuat variable dalam suatu job.

**Simbol**: 

**Properties Set Variables**:

**Keterangan**:

**Job entry name:** untuk memberi nama step.

**Variable name:** untuk memberi nama variable.

**Value:** untuk mengisi nilai awal variable.

**Variable scope type:** untuk menentukan scope (jangkauan) dari variable.

* Valid in the Java Virtual Machine
* **Valid in the current job**
* Valid in the parent job
* Valid in the root job

Contoh mengisi properties Set Variables:

**Job entry name**: variabel nilai

|  |  |  |
| --- | --- | --- |
| Variable name | Value | Variable scope type |
| ujian | 95 | Valid in the current job |

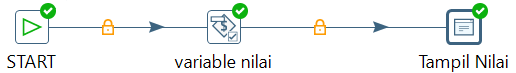
Maka hasilnya adalah:  yang memiliki variabel dengan nama **ujian ,**dengan nilai (*value*) **95** dan variabel hanya berlaku pada job tempat *step Set Variable* tersebut dibuat (*current job*).

**Membaca nilai variabel**:

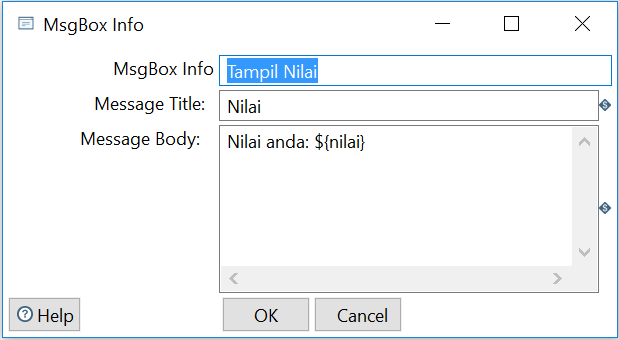
Untuk membaca nilai variabel yang dibuat melalui **Set Variables** gunakan: **${namaVariabel}**

**Contoh**:

Tambahkan step **Start** ke ruang kerja, kemudian tambahkan step **Set Variables** dan step **Display Msgbox Info**. Kemudian susun step dan hop seperti dibawah ini:

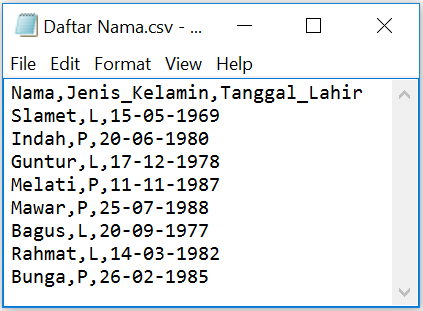


Untuk **Set Variables** isi properties seperti di atas, sedangkan untuk **Display Msgbox Info** isi propertiesnya seperti berikut:



### Step CSV File Input (hanya ada pada Pentaho DI versi 7 ke-atas)

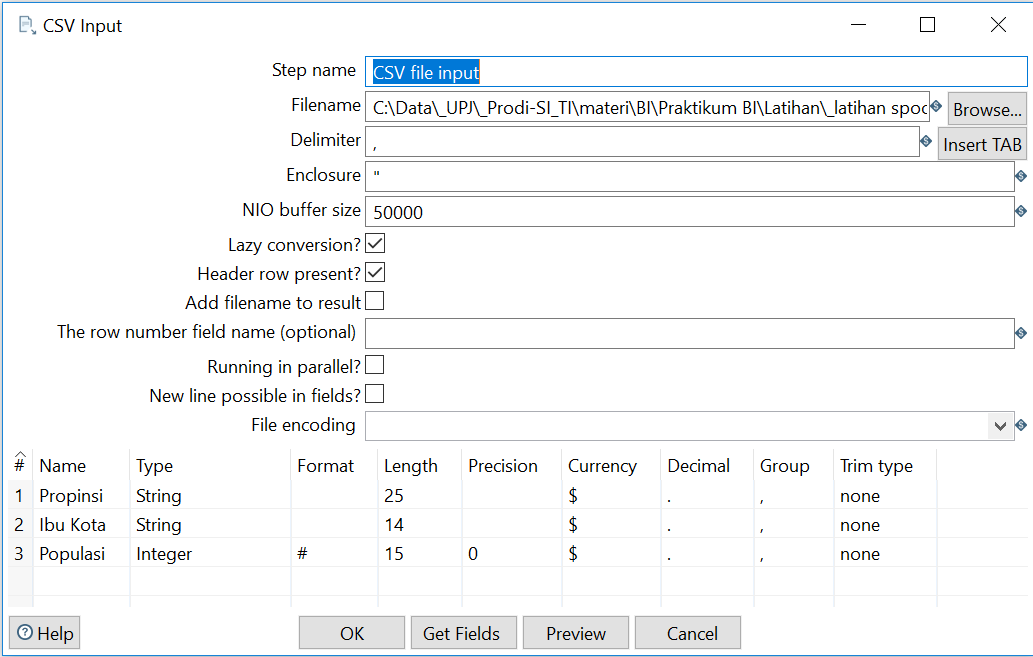
Contoh isi file teks tipe **CSV**:



Step **CSV File Input**  merupakan step yang berada dalam grup **Input** pada suatu **Transformation**, digunakan untuk membaca isi file teks dengan tipe CSV (*comma-separated values*) yang digunakan sebagai input pada suatu transformation. Step **CSV File Input** hanya ada pada lingkungan kerja **Transformation**.

**Simbol: **

**Properties CSV File Input:**



Berikut adalah beberapa item yang sering di-set pada properties **CSV File Input**:

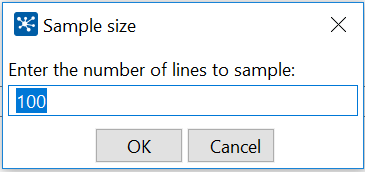
**Step Name :** Untuk memberikan nama pada step.

**Filename :** menentukan nama file dan lokasi file teks CSV yang akan dibuka.

Tekan tombol  untuk mencari file teks CSV yang ingin dibuka.

**Delimiter :** Untuk menentukan simbol pemisah antar field data pada file teks CSV. Biasanya field data pada file teks CSV dipisahkan dengan simbol koma (**,**) atau titik-koma/semicolon (**;**) atau tabulasi/tab.

**Daftar field :** Mendeskripsikan field data yang ada pada file teks CSV, untuk mendapatkan deskripsi field ini cukup tekan tombol  kemudian akan tampil form untuk menentukan jumlah baris data yang akan diambil untuk sample:



Setelah menekan tombol **OK** akan tampil form hasil scan terhadap sampel, kemudian tekan tombol Close untuk memasukkan hasil scan tersebut ke daftar field.

Untuk melakukan preview terhadap data yang akan diambil, tekan tombol  maka akan tampil form berisi data preview.

### Step Fixed File Input (hanya ada pada Pentaho DI versi 7 ke-atas)

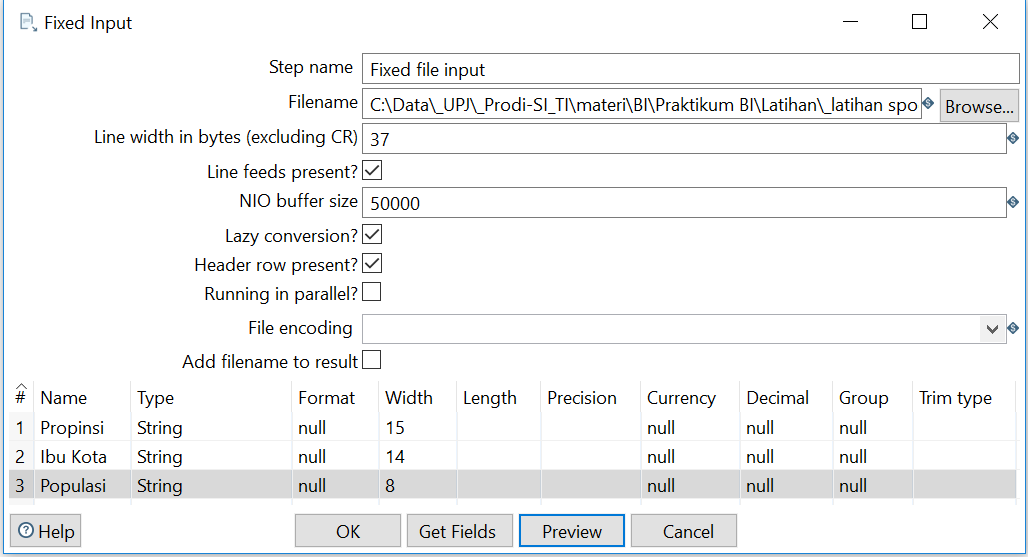
Contoh isi file teks tipe Fixed:



Step **Fixed File Input**  merupakan step yang berada dalam grup **Input** pada suatu **Transformation**, digunakan untuk membaca isi file teks dengan tipe fixed (*lebar tiap kolom data fixed*) yang digunakan sebagai input pada suatu transformation. Step **Fixed File Input** hanya ada pada lingkungan kerja **Transformation**.

**Simbol: **

**Properties Fixed File Input:**



Berikut adalah beberapa item yang sering di-set pada properties **Fixed File Input**:

**Step Name :** Untuk memberikan nama pada step.

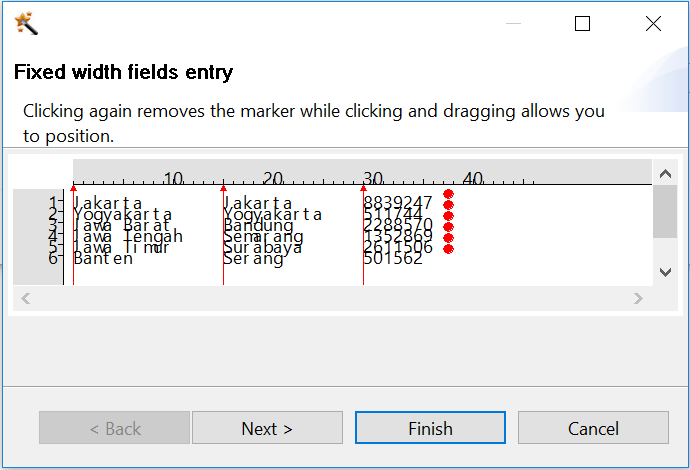
**Filename :** menentukan nama file dan lokasi file teks tipe fixed yang akan dibuka.

Tekan tombol  untuk mencari file teks tipe fixed yang ingin dibuka.

**Line width in bytes :** Untuk menentukan jumlah karakter pada baris data terlebar/terpanjang dalam file teks tipe fixed. (caranya: hitung jumlah krakter (termasuk jumlah spasi) pada baris yang paling lebar dalam file teks).

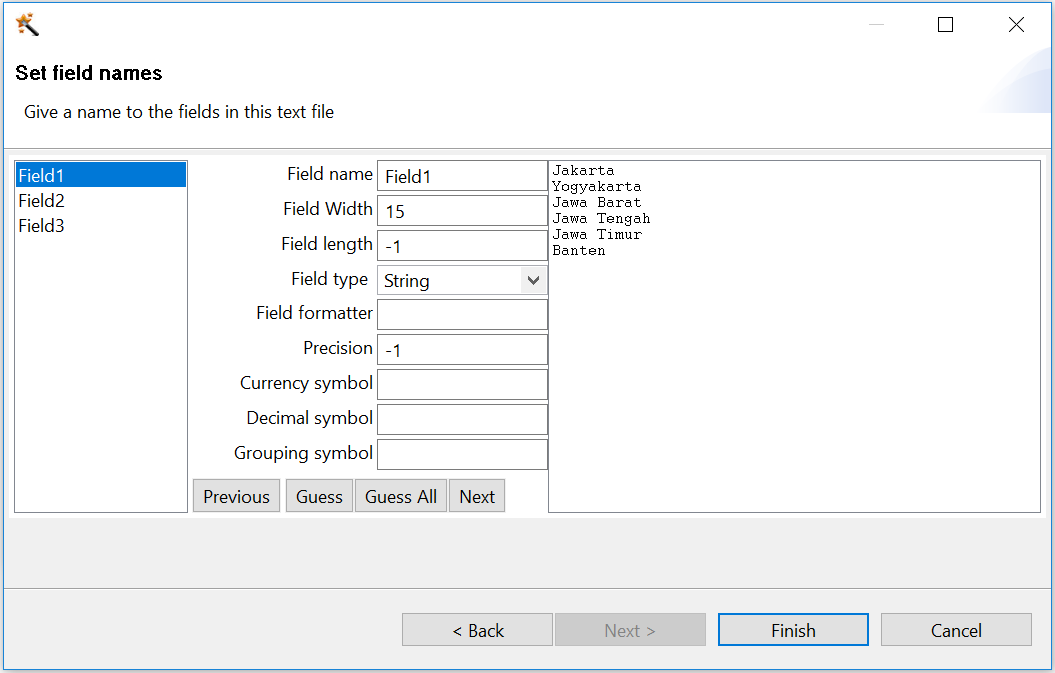
**Daftar field :** Mendeskripsikan field data yang ada pada file teks tipe fixed.

Untuk mendapatkan deskripsi field ini cukup tekan tombol  kemudian akan tampil form untuk menentukan batas awal tiap-tiap kolom pada teks:



Click pada *ruler* (penggaris) untuk membuat garis batas awal field (garis warna merah). Untuk menghapus garis pembatas yang salah, *click* pada garis yang ingin dihapus.

Tekan tombol **Next** sehingga akan tampil form:



*Click* **Field1** , kemudian tentukan **field name** (nama field) dan **field type** (tipe field) untuk field 1, demikian juga untuk **field2**, dan seterusnya.

*Click* tombol **Finish**, untuk memasukkan field ke dalam daftar field.

Untuk melakukan preview terhadap data yang akan diambil, tekan tombol  maka akan tampil form berisi data preview.

## Praktek

Pada bagian ini, akan dibahas mengenai praktek yang akan dilakukan.

## Praktek 1: Mengakses data dari file teks tipe CSV

Ikuti langkah-langkah berikut ini untuk membuat transformation yang akan membaca data nama dari file teks tipe CSV **daftar nama.csv**!

1. Jalankan spoon.bat
2. Buat transformation baru, file 🡪 new 🡪 transformation.
3. Click & Drag (atau *shift+double click*) **CSV file input** dari entries **Input** ke ruang kerja
4. *Double Click* pada step CSV file input, untuk membuka properties
5. Pada properties step name, isikan nama step (=daftar nama).
6. *Click* tombol **browse** untuk mencari file **daftar nama.csv** .
7. Pastikan delimiter sesuai dengan pemisah yang digunakan pada file **daftar nama.csv**.
8. *Click* tombol **Get Fields** untuk mengambil field yang ada pada file.
9. *Click* tombol **Preview** untuk melihat previewdata dari file teks.
10. Simpan file transformation dengan nama **ambil data nama.ktr** yang nantinya bisa digunakan di dalam job.

## Praktek 2: Mengakses data dari file teks tipe Fixed

Ikuti langkah-langkah berikut ini untuk membuat transformation yang akan membaca data nama dari file teks tipe Fixed **daftar nama.txt**!

1. Jalankan spoon.bat
2. Buat transformation baru, file 🡪 new 🡪 transformation.
3. Click & Drag (atau *shift+double click*) **Fixed file input** dari entries **Input** ke ruang kerja
4. *Double Click* pada step Fixed file input, untuk membuka properties
5. Pada properties step name, isikan nama step (=daftar nama).
6. *Click* tombol **browse** untuk mencari file **daftar nama.txt** .
7. Pada line width in bytes, isi dengan jumlah karakter pada baris teks terpanjang dalam file teks.
8. *Click* tombol **Get Fields** untuk field-field yang ada pada file, tentukan titik awal (garis batas awal) tiap field.
9. *Click* tombol **Next**, kemudian ubah nama field dan tipe field sesuai kebutuhan.
10. *Click* tombol **Finish**.
11. *Click* tombol **Preview** untuk melihat previewdata dari file teks.
12. Simpan file trenaformation dengan nama **ambil data nama 2.ktr** yang nantinya bisa digunakan di dalam job.

# Analisa Hasil Praktek

Berdasarkan praktek yang telah Anda lakukan, jawablah beberapa pertanyaan berikut ini:

1. …?

# Tugas

1. ….!