

Virtual Internship Experience

id/x partners

ETL Implementation

Disclaimer

“Dokumen ini memiliki hak cipta. Barang siapa yang menyebarluaskan atau menduplikasi tanpa izin dari instansi terkait dapat diproses sesuai dengan ketentuan hukum yang berlaku.”

Outline

- 1 ETL Implementation**
- 2 ETL Tools**



1 ETL Implementation

Pada Implementasi proses ETL ini, kita akan melakukan ekstraksi data yang berasal dari file excel, kemudian melakukan transformasi data dengan menerapkan filter ke data dan memuat data baru tersebut ke dalam database

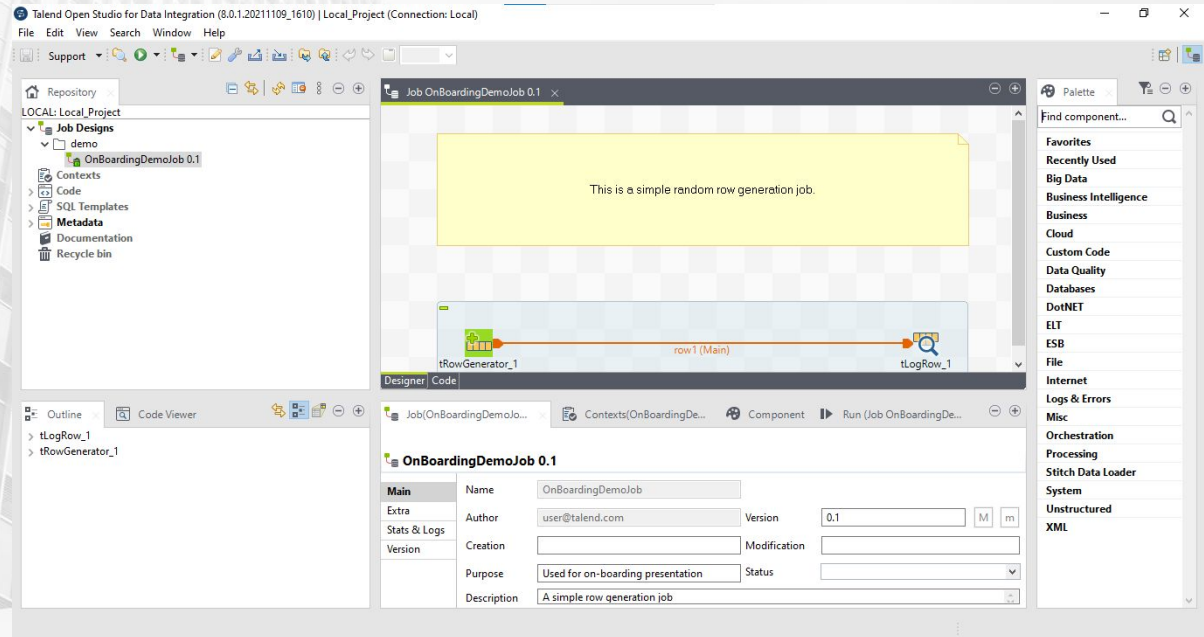
Cust_ID	Cust_Fname	Cust_Lname	Cust_Email	Cust_City	Cust_Type
101	James	Butt	jbutt@gmail.com	New Orleans	VIP
102	Josephine	Darakjy	josephine_darakjy@darakjy.org	Brighton	New
103	Art	Venere	art@venere.org	Bridgeport	Loyal
104	Lenna	Paprocki	lpaprocki@hotmail.com	Anchorage	VVIP
105	Donette	Foller	donette.foller@cox.net	Hamilton	New
106	Simona	Morasca	simona@morasca.com	Ashland	Loyal
107	Mitsue	Tollner	mitsue_tollner@yahoo.com	Chicago	New
108	Leota	Dilliard	leota@hotmail.com	San Jose	Loyal
109	Sage	Wieser	sage_wieser@cox.net	Sioux Falls	VVIP
110	Kris	Marrier	kris@gmail.com	Baltimore	Loyal
111	Minna	Amigon	minna_amigon@yahoo.com	Kulpsville	Loyal
112	Abel	Maclead	amaclead@gmail.com	Middle Island	New
113	Kiley	Caldarera	kiley.caldarera@aol.com	Los Angeles	Loyal

Gambar 4 Sample Data untuk ETL

1 ETL Implementation

Tahap 1: Buat *job* baru dan dari *palette*, seret dan lepas komponen berikut:

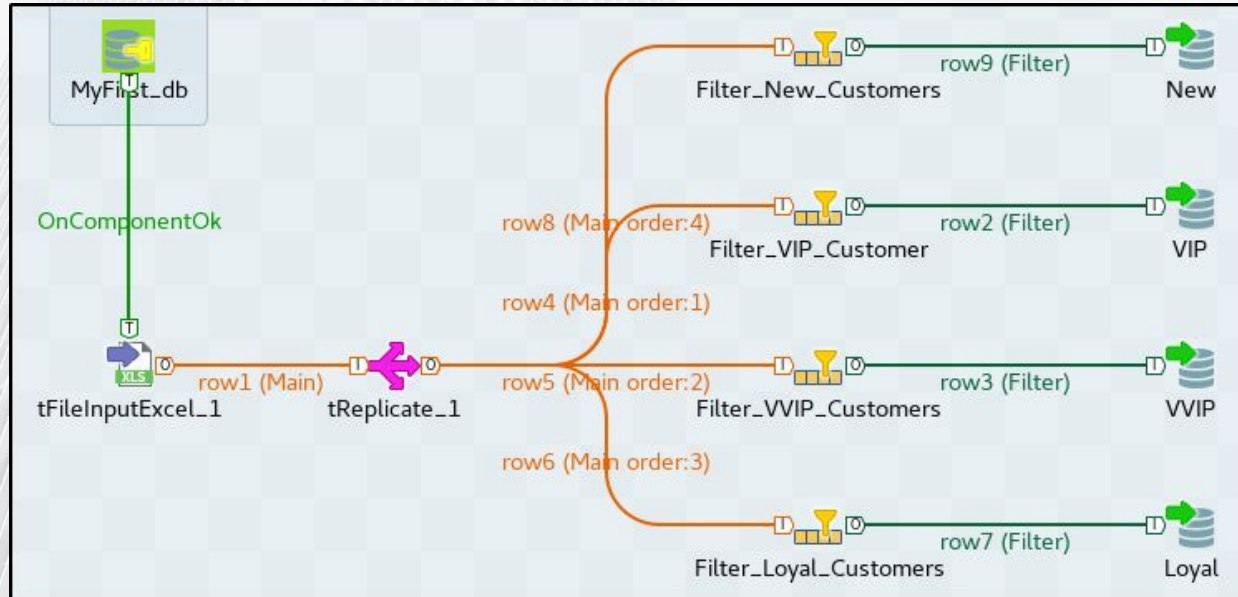
1. tMysqlConnection
2. tFileExcelInput
3. tReplicate
4. (tFilterRow)X4
5. (tMysqlOutput)X4



Gambar 5 Tampilan Awal Talend

1 ETL Implementation

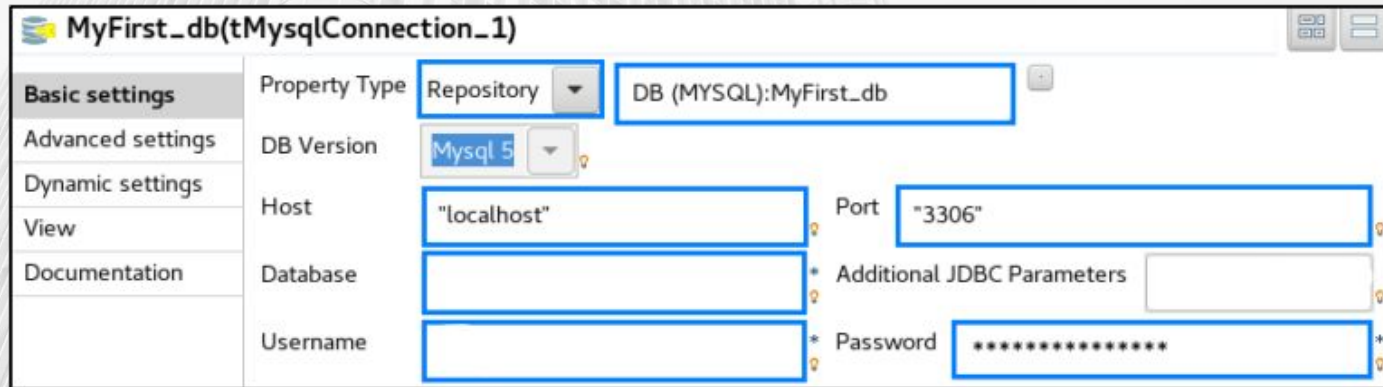
Tahap 2: Hubungkan komponen yang ada sesuai dengan gambar di bawah ini



Gambar 6 Konfigurasi Antar Elemen

1 ETL Implementation

Tahap 3: Buka tab komponen tMySQLConnection dan dari 'Property Type' pilih jenis koneksi yang digunakan; *Built-in* atau *Repository*. Jika Anda menggunakan koneksi bawaan maka Anda harus menentukan detail berikut:

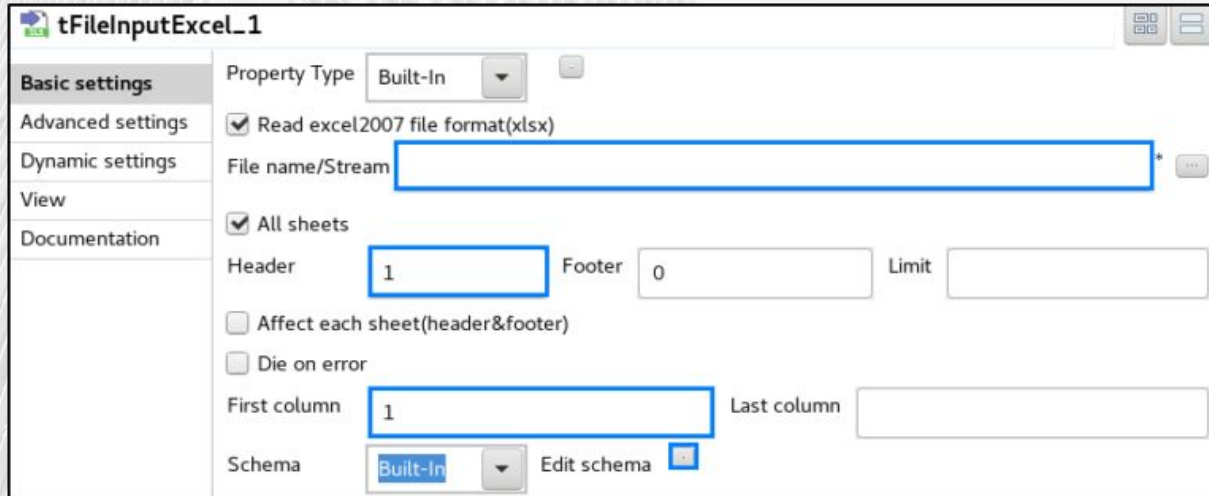


Property Type	Repository	DB (MYSQL):MyFirst_db
DB Version	Mysql 5	
Host	localhost	Port 3306
Database		* Additional JDBC Parameters
Username		* Password ***** *

Gambar 7 Konfigurasi Koneksi tMySQLConnection

1 ETL Implementation

Tahap 4: Buka tab komponen tFileExcelInput dan konfigurasi file Excel yang anda gunakan mulai dari file name hingga nomor kolom yang akan Talend baca sebagai kolom pertama



tFileExcelInput_1

Property Type: Built-In

☒ Read excel2007 file format(xlsx)

File name/Stream:

☒ All sheets

Header: Footer: Limit:

☐ Affect each sheet(header&footer)

☐ Die on error

First column: Last column:

Schema: Built-In

Gambar 8 Konfigurasi tFileExcelInput

1 ETL Implementation

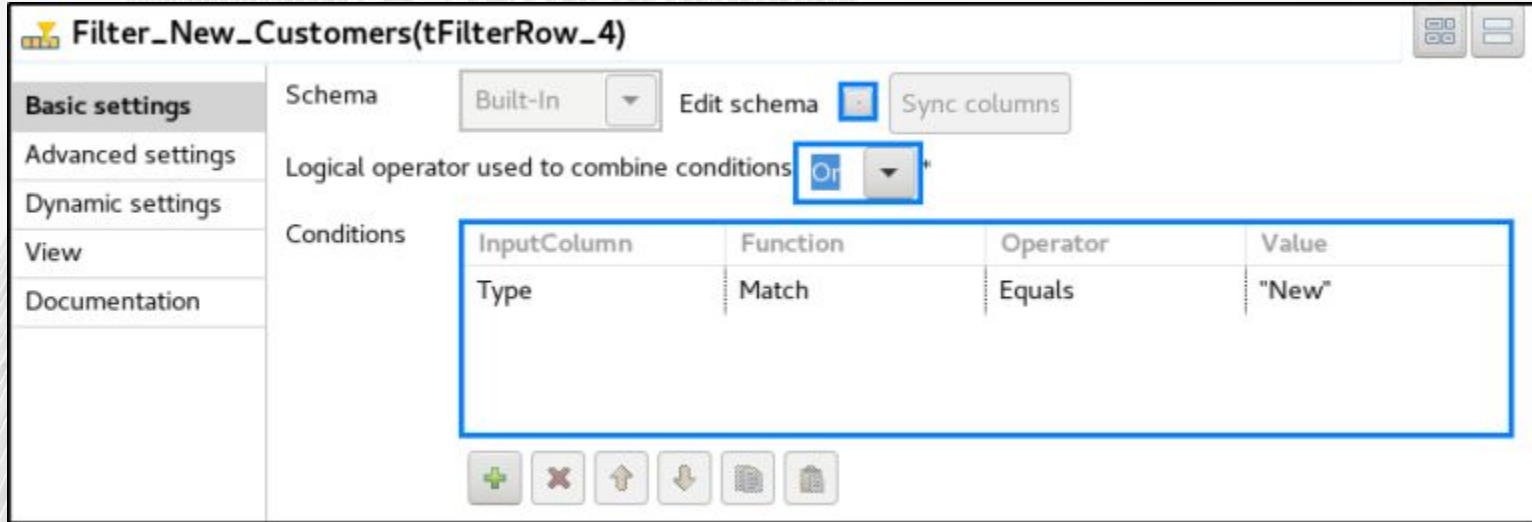
Tahap 5: Di tab komponen tReplicate, klik 'Sync columns'.



Gambar 9 Konfigurasi tReplicate

1 ETL Implementation

Tahap 6: Buka tab komponen tFilterRow pertama dan periksa skemanya. Lalu, isi *Condition* sesuai dengan kondisi yang akan diterapkan pada *filter*. Lakukan pada setiap komponen tFilterRow berikutnya.



Filter_New_Customers(tFilterRow_4)

Schema: Built-In Edit schema Sync columns

Logical operator used to combine conditions: Or

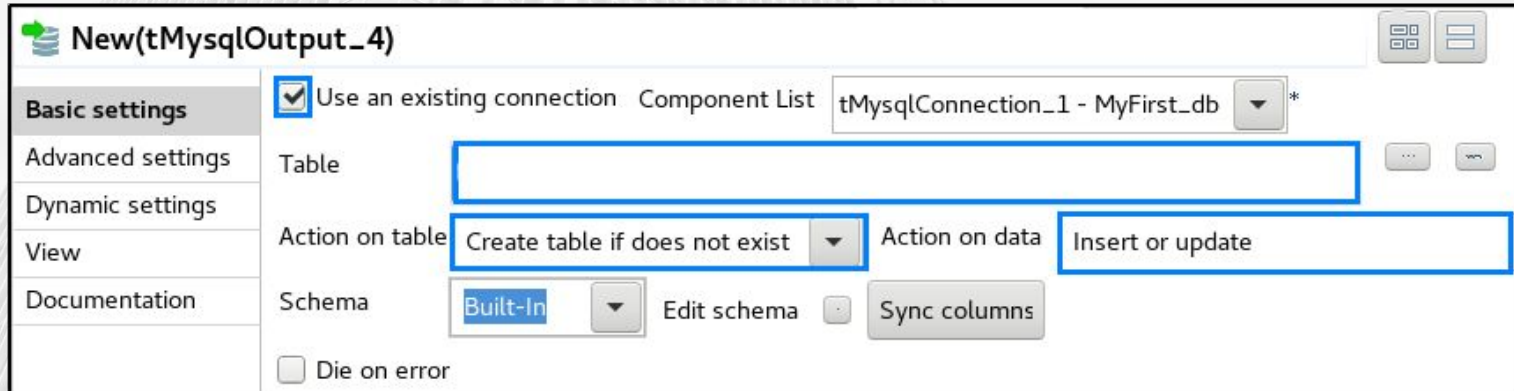
InputColumn	Function	Operator	Value
Type	Match	Equals	"New"

Buttons: +, -, Up, Down, Copy, Paste

Gambar 10 Konfigurasi tFilterRow

1 ETL Implementation

Tahap 7: Dalam tab komponen tMySQLOutput, centang tanda *'Use an existing connection'*. Kemudian tentukan nama tabel di bidang *'Table'* dan pilih *'Action on table'* dan *'Action on data'* sesuai kebutuhan. Lakukan pada setiap komponen tMySQLOutput berikutnya.



New(tMySQLOutput_4)

Basic settings

☒ Use an existing connection Component List tMySQLConnection_1 - MyFirst_db *

Table

Action on table Create table if does not exist Action on data Insert or update

Schema Built-In Edit schema Sync columns

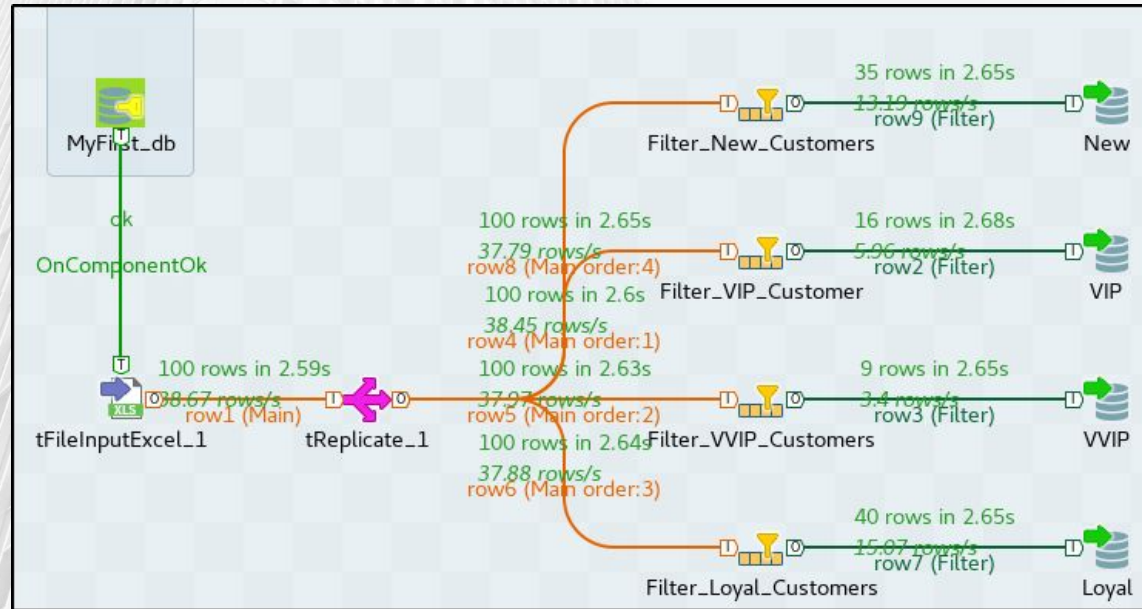
☐ Die on error

Gambar 11 Konfigurasi tMySQLOutput



1 ETL Implementation

Tahap 8: buka tab 'Run' dan jalankan *Job* yang telah dibuat.



Gambar 12 Executed Job

2 ETL Tools

Talend



IBM DataStage



SAS Data Management



Pentaho



Google Data Flow



Azure Data Factory





Thank You!



id/x partners