Laporan Praktikum Data Warehouse

A. Dimensi Waktu

1. Buatlah sebuah database yang digunakan sebagai OLAP dengan nama dw_LegendVehicle.

CREATE DATABASES

dw LegendVehicle; use

dw_LegendVehicle;

2. Buatlah table untuk menyimpan data master waktu atau yang disebut dengan tabel dimensi. Beri nama table tersebut dengan nama dimDate.

CREATE TABLE dimDate(

id_dimDate int not null AUTO_INCREMENT PRIMARY

KEY date date

year int

month

int day

int

);

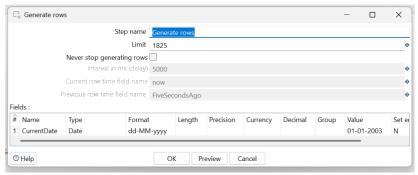
Pada tahapan selanjutnya, untuk membuat tabel dimensi dimDate , maka diperlukan generate data tanggal. Data tanggal yang disiapkan pada tabel dimDate menyesuaikan dengan proses bisnis yang berjalan.

Proses bisnis pada LegendVehicle adalah 5 tahun. Sehingga data pada tabel dimdate yang harus tersedia adalah tanggal selama 5 tahun. Mulai dari 1 Januari 2023

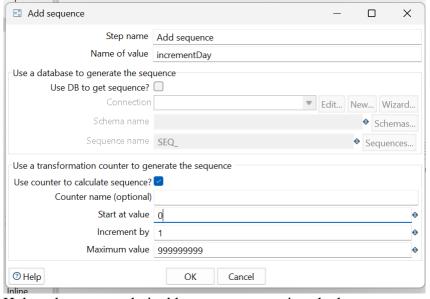
- 1. Buka PDI Spoon. Buat Transformation baru -> File New Transformation.
- 2. Drag and Drop beberapa objek yaitu:
 - Generate Rows: digunakan untuk membuat baris data baru.
 - Add Sequence: digunakan untuk membuat sequence, dalam hal ini membuat data di setiap harinya.
 - Calculator: digunakan untuk menjumlahkan hari dan mengambil data tahun, bulan dan hari.
 - Select Values: digunakan untuk memilih field yang digunakan.
 - Database Lookup: digunakan untuk melihat dan memastikan bahwa data yang akan dimasukkan kedalam tabel dimDate tidak kembar atau sama dengan data yang ada pada tabel dimDate itu sendiri.
 - Filter Rows: digunakan untuk mengambil data yang belum ada pada table dimDate setelah dicek sebelumnya.
 - Table Output: digunakan untuk menyimpan data pada tabel tujuan (dimDate).



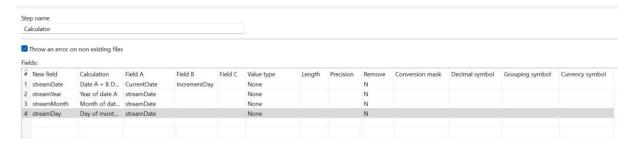
- 3. Konfigurasi pada Generate Rows adalah merubah limit menjadi 1825 dimana memiliki arti bahwa data yang akan dibuat sebanyak 1825 data. 1825 merupakan jumlah hari dalam 5 tahun (365 hari x 5 tahun).
- 4. Membuat fields baru bernama CurrentDate dengan type data Date dan format dd-MM- yyyy serta value awal 01-01-2003.



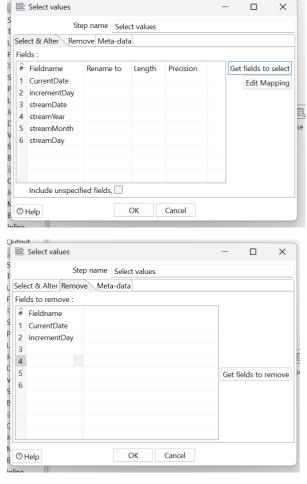
- 5. Hubungkan output dari Generate Rows menuju Add Sequence.
- 6. Konfigurasi pada **Add Sequences** adalah merubah **Name of value** menjadi **incrementDay** dengan **start value** bernilai **0** dan **increment by** bernilai **1**



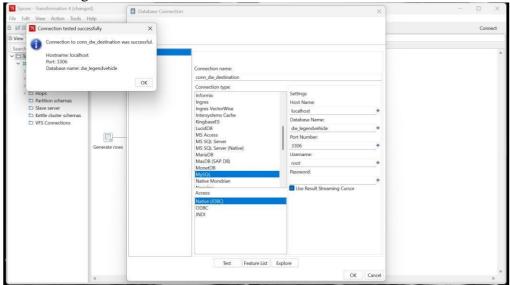
- 7. Hubungkan output dari add sequences menuju calculator.
- 8. Konfigurasi pada calculator dengan membuat fields baru sebagai berikut:
 - streamDate merupakan kalkulasi dari CurrentDate + incrementDay
 - streamYear merupakan Year dari streamDate
 - streamMonth merupakan Month dari streamDate
 - streamDay merupakan Day of month dari streamDate



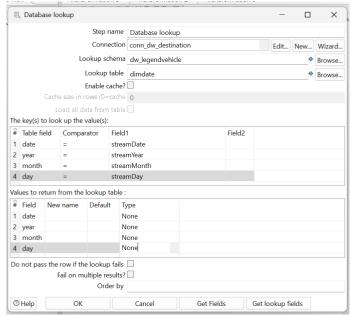
- 9. Hubungkan output dari calculator menuju Select values
- 10. Konfigurasi pada **select values** adalah dengan menekan tombol **Get fields to select** pada tab **Select & Alter**. Secara otomatis semua fields dari data input akan muncul.
- 11. Dikarenakan tidak semua fields digunakan, maka pada tab **Remove** diisikan fields **CurrentDate** dan **incrementDay** dikarenakan kedua fields tersebut tidak digunakan.



- 12. Hubungkan output select values menuju database lookup.
- 13. Sebelum melakukan konfigurasi pada database lookup, buatlah koneksi terlebih dahulu pada database melalui File New Database Connection. Gunakan Connection type MySQL dengan host name , database name, port number, username dan password sesuai konfigurasi MySQL pada device masing-masing. beri nama connection name tersebut dengan nama conn_dw_destination.



- 14. Konfigurasi pada **database lookup** adalah dengan memberikan **connection** dengan koneksi yang sudah dibuat pada step sebelumnya. dengan **schema** nama database yang digunakan dan **tabel dimdate** yang telah dibuat pada langkah pertama.
- 15. Field yang akan dicek untuk melihat kesamaan isi datanya agar tidak kembar adalah:
 - field date pada table dimdate dengan field streamDate
 - field year pada table dimdate dengan field streamYear
 - field month pada table dimdate dengan field streamMonth
 - field day pada table dimdate dengan field streamDay
- 16. Field yang akan di **retrive** adalah field yang ada pada table **dimDate** yaitu **date, year, month,** dan **day**.

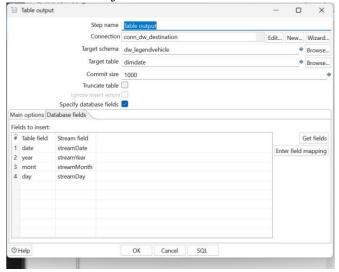


- 17. Hubungkan output dari database lookup dengan filter rows
- 18. Konfigurasi pada **filter rows** adalah dengan melakukan konfigurasi **output true data** pada **table output**. Pada bagian ini data yang tidak memiliki kesamaan pada tahapan sebelumnya akan dicek dimana jika **fields Stream** tidak memiliki kesamaan dengan **field dimDate**, maka **field dimDate** tersebut akan bernilai **null**. Pada pernyataan kondisi tuliskan (**date is null and year is null and month is null and day is nu**ll)



19. Hubungkan output dari filter rows menuju table output.

- 20. Konfigurasi pada table output adalah memberikan koneksi pada conn_dw_destination dengan schema dw_legendvehicle dan table dimdate.
- 21. Aktifkan specify database fields.
- 22. Pada tab Database fields, mapping data input streamDate, streamYear, streamMonth dan streamDay dengan fields yang ada pada dimDate. Pada tahapan ini akan dilakukan insert data menuju tabel dimDate.



23. cek isi table dimdate pada database. Jika sukses maka pada table dimdate akan terisi 1825 data.



TUGAS 1

1. Buka preview tab pada execution result area di setiap proses object. amati input dan output data yang ada. bandingkan di setiap prosesnya. jelaskan perbedaan disetiap prosesnya.

PROSES	SS DATA INPUT	SS DATA OUTPUT	KET.
Generate rows	Signature Service Serv	3 Lagging 3 Lancition Hallow 3 Sept Melitics 5 Newton State 5 N	- Step ini digunakan untuk menghasilkan data secara otomatis tanpa input dari sumber lain Pada contoh ini, 1000 baris data dibuat dengan nilai tetap "01-01-2003" pada kolom CurrentDate Hasil output menunjukkan 1000 baris dengan tanggal
Add Sequences	Step action Step action The off of the presence of the step action o	© Legging © Legislation Halloy ⊆ Step Medics C Nerhamania Graph (5 Meins) ◆ Preview dat Onto these C state one C off 1 (101–280) 1 (101–280) 2 (101–280) 3 (101–280) 4 (101–280) 5 (101–280) 6 (101–280) 6 (101–280) 6 (101–280) 7 (101–280) 7 (101–280) 7 (101–280) 1 (101–280)	yang sama. - Step ini digunakan untuk menambahkan nomor urut (sequence) ke setiap baris data yang dihasilkan sebelumnya. - Field baru IncrementDay ditambahkan dengan nilai yang bertambah mulai dari 0, 1, hingga jumlah baris data yang ada. - Hasil output menunjukkan bahwa setiap baris memiliki tanggal CurrentDate yang sama, tetapi memiliki sequence yang bertambah dari 1, 2, 3, hingga baris terakhir.
Calculator	1 1 1 1 1 1 1 1 1 1	□	- Step ini digunakan untuk melakukan operasi matematika pada data. - Pada contoh ini,

	B Morehan — Q x	The integral Streetler Integral Streetler Integral Carl D. Mary 9 Februaries	dilakukan perhitungan streamDate = CurrentDate + incrementDay, sehingga menghasilkan tanggal yang bertambah setiap barisnya Selain itu, data dipecah menjadi beberapa bagian: streamYear, streamMonth, streamDay.
Select values	Supparine Suppar	OT TOTAL (COLUMN) COLUMN	 Step ini digunakan untuk memilih kolom yang akan dipertahankan dalam proses selanjutnya. Kolom yang tidak dibutuhkan akan dibuang agar proses lebih efisien.
Database lookup	Distance Name Processing	Description	- Step ini digunakan untuk mencocokkan data dengan tabel referensi dari database Data yang telah dihitung (misalnya, streamDate) dibandingkan dengan data dari database untuk mendapatkan informasi tambahan Contohnya, hasil output menunjukkan data yang telah diperiksa terhadap tabel referensi dan menambahkan informasi baru seperti date, year, month, day.
Filter rows	Sing the office to the Cale in our American Send the office to the Cale in our American Send the office to the Cale in our American In our American Send the Send of S	United controls of the action of the standard	Step ini digunakan untuk menyaring data berdasarkan kondisi tertentu.



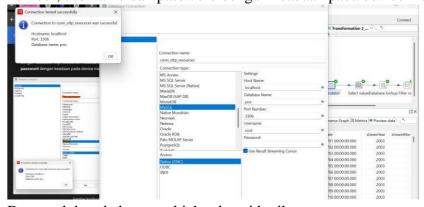
B. Dimensi Pegawai

1. Buatlah tabel dimPegawai pada dw_legendVehicle.

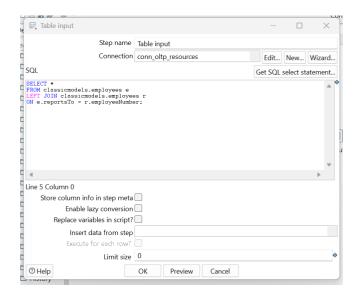
PRIMARY KEY (id_dimEmployees)

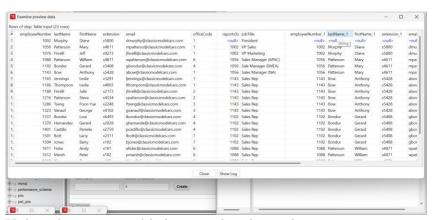
```
CREATE TABLE dimEmployees (
id_dimEmployees INT NOT NULL AUTO_INCREMENT,
employeeNumber INT,
firstName VARCHAR(50),
lastName VARCHAR(50),
jobTitle VARCHAR(50),
boss_firstName VARCHAR(50),
boss_lastName VARCHAR(50),
updated DATETIME DEFAULT CURRENT_TIMESTAMP,
```

2. Pada PDI Spoon buatlah koneksi baru dengan nama conn_oltp_resources yang menghubungkan dengan database oltp. sesuaikan hostname, database name, port number, username dan password dengan keadaan pada device masing-masing.

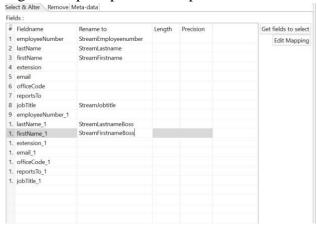


- 3. Drag and drop beberapa objek sebagai berikut:
 - Table input: digunakan mengambil data dari database OLTP.
 - Select values: memeilih field yang digunakan untuk proses Transform dan Load.
 - Database lookup: digunakan untuk melihat data pada tabel dimEmployees untuk memastikan data tidak kembar
 - Filter rows: digunakan untuk memilih data stream yang masih belum ada apada tabel dimEmployees.
 - Table output: Memasukkan data ke dalam tavle dimEmployees
- 4. Konfigurasi pada table input dengan menghubungkan **Connection** pada konesi **conn_oltp_resources**. Untuk mengambil data sumber menggunakan query dibawah ini.

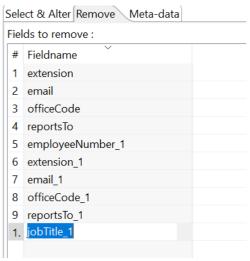




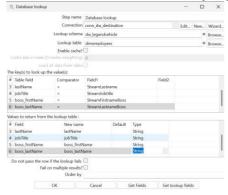
- 5. Hubungkan output table input pada select values.
- 6. Konfigurasi pada Select values yaitu mengambil data dari field employeenumber, lastname, firstname, jobtitle, lastname_1 dan firstname_1 sebagai data stream yang digunakan pada proses ETL pada tab select & alter.



7. Hilangkan field lain yang tidak digunakan pada tab remove.



- 8. Hubungkan output select values pada database lookup.
- 9. Konfigurasi pada database lookup adalah dengan menghubungkan koneksi pada conn_dw_destination dengan table lookup dimEmployees yang telah dibuat pada tahap pertama.
- 10. Field yang di lookup adalah field pada tabel dimEmployees dengan field stream input dari OLTP. sedangkan field yang di retrieve adalah field dari dimEmployees itu sendiri. Jika tidak ada data yang sama maka akan muncul null.

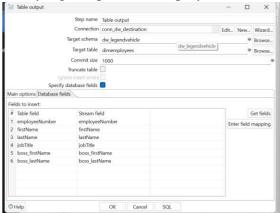


- 11. Hubungkan output databse lookup dengan filter rows.
- 12. Pada filter rows berikan kondisi field yang null pada field dimemployees untuk dimasukkan pada proses selanjutnya. Hal itu menandakan bahwa data stream belum memiliki kesamaan pada data di dimemployees.



13. Hubungkan output dari filter rows dengan table output.

- 14. Pada table output, gunakan connection conn_dw_destination untuk memasukkan data pada tabel dimemployees.
- 15. Aktifkan specify databse fields, dan mapping data stream input dari oltp terhadap field yang ada pada dimemployees.



16. jika proses keseluruhan berhasil maka tabel **dimemployees** akan terisi data pegawai dari database OLTP.



TUGAS 2

1. Buka preview tab pada execution result area di setiap proses object. amati input dan output data yang ada. bandingkan di setiap prosesnya. jelaskan perbedaan disetiap prosesnya.

PROSES OBJEK	SS DATA INPUT	SS DATA OUTPUT	KET.
Table Input	Company Comp	Description	Hasil output adalah seluruh dataset awal sebelum transformasi.
Select values	Company Comp	Company Comp	Input berupa semua kolom dari Table Input, sedangkan output hanya kolom yang dipilih
Database lookup	### Company of the co		Menambahkan data dari tabel lain berdasarkan key,

			output akan memiliki kolom tambahan dari tabel lain
Filter Rows			Input memiliki
	1	1 15 mg 10 10 10 10 10 15 mg 10 10 10 10 10 10 10 10 10 10 10 10 10	semua data sebelum
		17 No. 16 No. 16 No. 16 No. 17 No. 16 No. 16 No. 17 No. 16 No. 17 No. 17 No. 18 No. 1	filtering, sedangkan
			output hanya
			menyisakan data
			yang memenuhi
			kondisi.
Table output	Second S	Societies Togg Transition To the Editional of State Marcoll Once Transit State Institute Annie State Institute Annie State Institute Annie	Data yang masuk ke
	1		sini adalah hasil
		1 10 for Along Markey 10 former for 15 Along long long in Markey 10 former for 15 Along long long long in Markey 17 10 former long long long in Markey 17 10 former long long long long long long long long	akhir setelah semua
	100 0 00 00 00 00 00 00 00 00 00 00 00 0	27 beam 26	proses sebelumnya.

- 2. Jika proses itu di ulangi (di run kembali) apakah data akan redudant? Jawaban: Jika proses ini diulang (di-run kembali), data tidak akan redundant. Hal ini karena ada mekanisme yang mencegah duplikasi
 - Database Lookup: Mengecek apakah data sudah ada sebelum dimasukkan.
 - Filter Rows: Menyaring data yang akan diproses agar tidak terjadi duplikasi.
- 3. Tambahkan nama anda pada table employee di OLTP. jalankan kembali transformasi ini. Amati hasilnya, apa yang terjadi?

Jawaban: Data karyawan baru dengan nama saya berhasil ditambahkan pada table employees di OLTP dengan memastikan employeeNumber unik untuk menghindari duplikasi.

