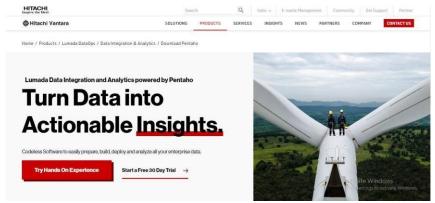
Nama : Arief Lindung NIM : 191402093

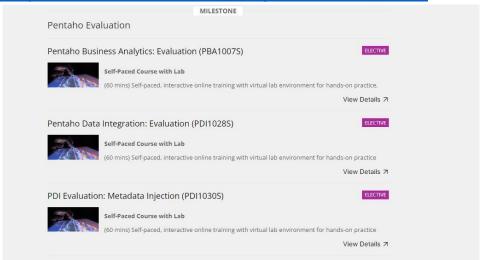
Kom : C

TUGAS DATA WAREHOUSE DAN BISNIS INTELLIGENCE

Halaman utama dari HITACHI Pentaho.

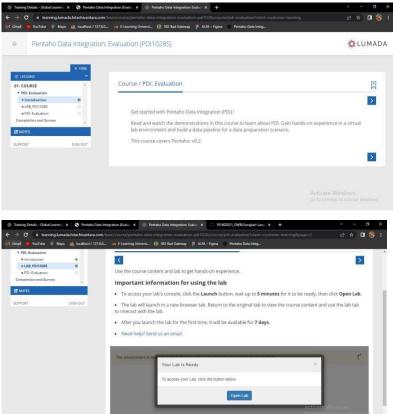


Pada saat kita mencoba Hands On Experience dari Pentaho, kita akan dialihkan ke halaman https://learning.lumada.hitachivantara.com/learning-paths/evaluate-pentaho

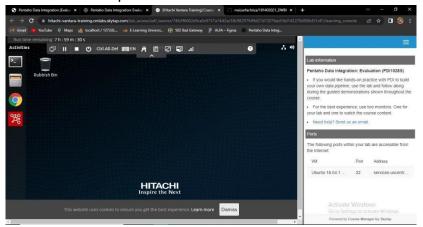


Pentaho Data Integration: Evaluation.

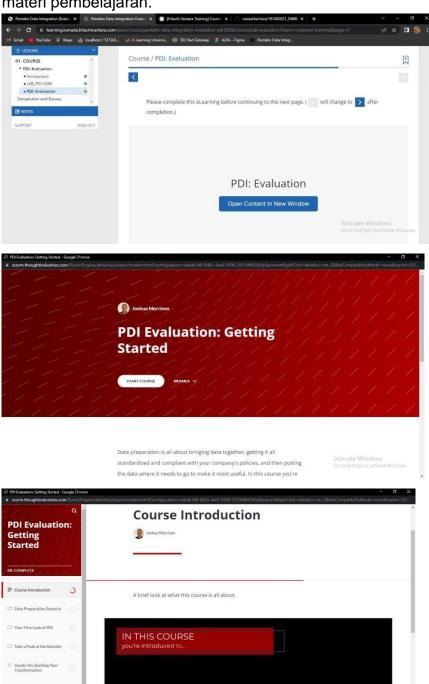
Dengan mendaftarkan diri pada pembelajaran tersebut, kita akan dialihkan ke halaman pembelajaran. Tersedia juga Lab pembelajaran yang dapat diakses via website untuk penerapan pembelajaran yang ada tanpa melakukan instalasi.



 Setelah melakukan permintaan terhadap akses pada lab, kita dapat mengakses virtual machine pada tab browser baru.

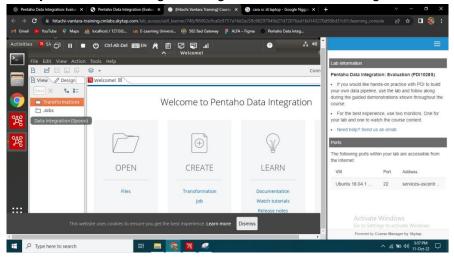


 Selanjutnya pada bagian PDI: Evaluation akan muncul window browser berisikan materi pembelajaran.

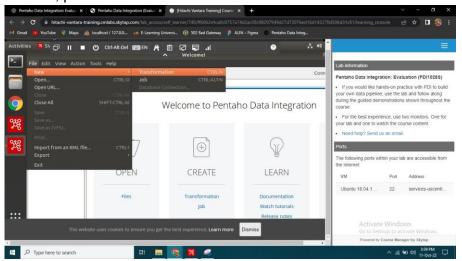


Reading Customer Data (CSV File Input Step)

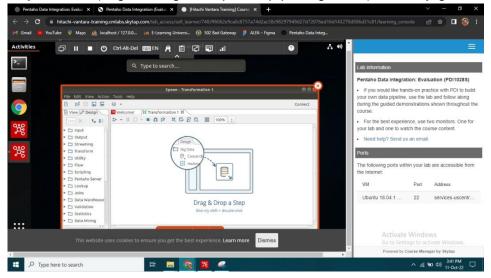
 Langkah pertama, kita akan membaca inputan data terlebih dahulu. Kita akan menggunakan data daftar customer. Sebelumnya kita akses terlebih dahulu Spoon pada virtual machine yang sudah tersedia. Spoon merupakan produk dari Hitachi yang bertujuan untuk menangani intergrasi data, dikenal sebagai Pentaho Data Integration.



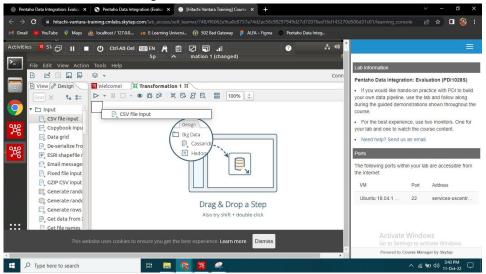
2) Pada penerapan kali ini, kita akan mencoba transformasi data pelanggan dari data yang belum bersih menjadi data yang bersih dan siap dipakai dalam database. Padabagian kiri atas, pilih new kemudian transformation.



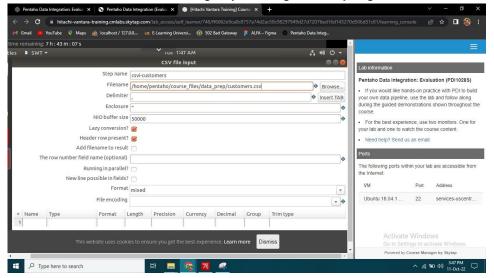
3) Kemudian akan muncul lembar kerja baru untuk melakukan transformation data. Lembar kerja baru ini (bagian Drag & Drop a Step pada gambar) dikenal juga sebagai canvas.



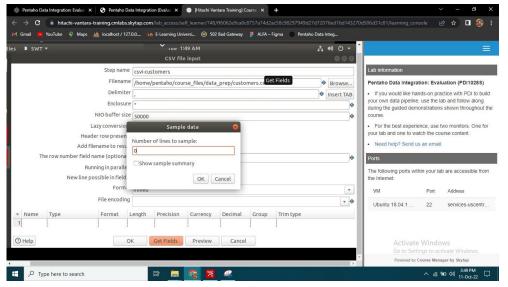
4) Kita mulai dengan membuat step untuk input customer data dengan mendrag CSV file input pada tab Design ke dalam canvas. Diikuti dengan double-click pada step tersebut untuk melakukan konfigurasi.



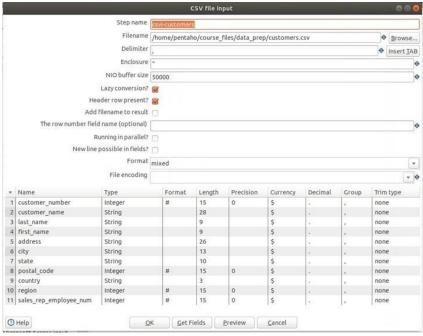
5) Konfigurasi yang dilakukan dapat berupa step name, filename sebagai sumber data, delimiter, enclosure, dan sebagainya sebagaimana yang sesuai untuk menginput data.



6) Kita lanjutkan dengan click pada Get Fields untuk mendapatkan nama kolom yang ada pada input data. Tetapkan 0 pada bagian Number of lines sample untuk mendapatkan nama kolom data secara keseluruhan.

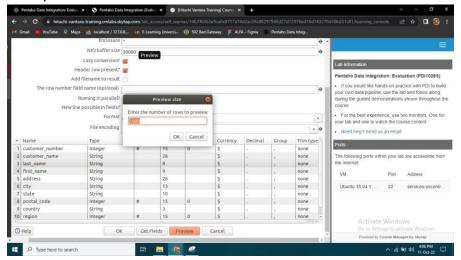


7) Maka nama kolom beserta propertinya akan muncul pada bagian bawah dialog konfigurasi



step.

8) Untuk melihat contoh data yang tersedia dari file tersebut, klik tombol preview pada bagian bawah dialog dan masukkan preview size sesuai keinginan banyak data yang ditampilkan.

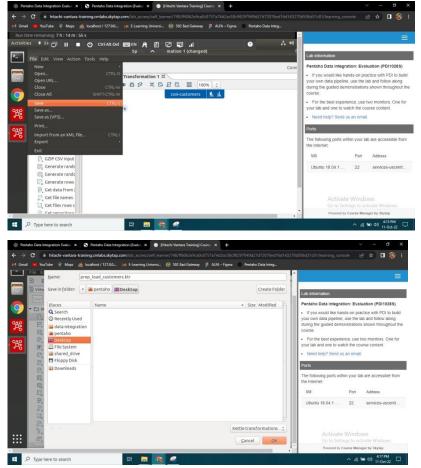


9) Akan muncul kotak dialog baru yang memuat data preview dari file input yang ditentukan. Dari preview ini kita dapat mengamati bagian data yang belum bersih seperti masih terdapatnya null pada kolom sales_rep_employee_num, penulisan state dan address yang tidak konsisten dalam penggunaan tanda titik, hingga tidak konsistennya penggunaan singkatan pada kolom address. Jika kita click pada salah satu data kolom state, ternyata masih terdapat white-space yang tidak diperlukan. Untuk itu kita harus membersihkan data agar dapat lebih efisien untuk



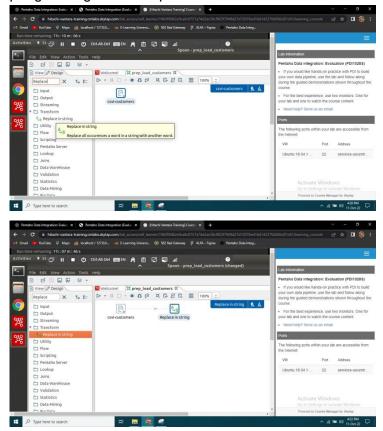
digunakan.

10) Setelah menyelesaikan konfigurasi terhadap file input, click OK. Lanjutkan dengan menyimpan perubahan yang kita lakukan pada canvas transformation terlebih dahulu. Pada tab file pada pojok kiri atas, pilih save dan simpan pada lokasi yang diinginkan.



Standardize Address and State Data (Replace in String Step)

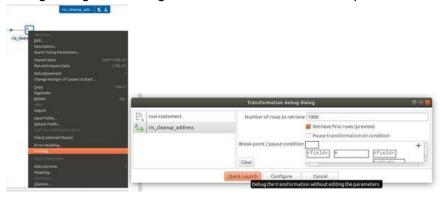
Selanjutnya kita akan melakukan standarisasi format untuk beberapa kolom, dalam contoh penerapan ini kolom address dan state. Sebelumnya, cari terlebih dahulu step Replace in string pada tab design. Untuk menghubungkan kedua step, click terlebih dahulu step csvi-customers dan double-click pada step Replace in string, secara otomatis penghubung kedua step akan terbentuk.



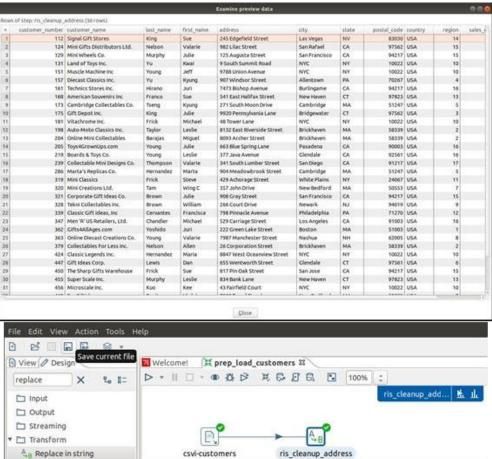
2) Kemudian kita lanjutkan konfigurasi step Replace in string dengan double-click. Standarisasi yang diterapkan berupa menghapus tanda titik pada kolom address dan state, juga mengubah singkatan St dan Ave menjadi Street dan Avenue. Setelah selesai, click OK.



3) Right-click pada step tersebut, pilih preview untuk menampilkan data hasil standarisasi. Sebelum mengakhiri preview, kita dapat melakukan konfigurasi pada transformation debug dialog. Bila konfigurasi selesai dilakukan, click pada Quick Launch.

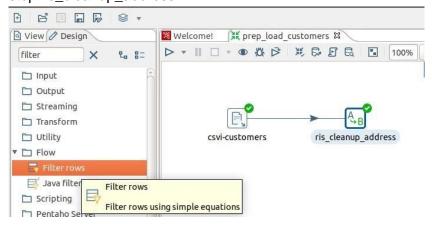


4) Dengan demikian, ktia dapat melihat hasil standarisasi data customer pada kotak dialog baru. Tidak lupa juga untuk melakukan save progress yang kita lakukan sejauh ini dengan click icon save pada bagian kiri atas aplikasi.

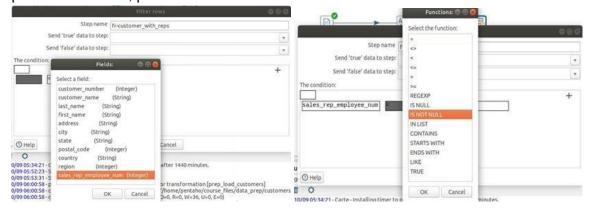


Filter Customers (Filter Rows Step)

Selanjutnya melakukan filtrasi terhadap data terstandarisasi. Pada penerapan kali ini kita akan memfiltrasi kolom sales_rep_employee_num. Sebelumnya, cari step Filter rows pada tab Design terlebih dahulu dan double-click. Pastikan step ini terhubung dengan step ris cleanup address.



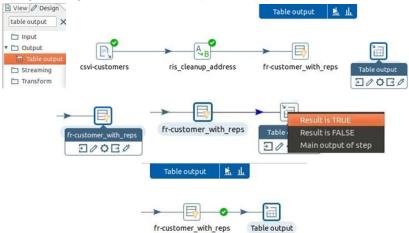
2) Double-click pada step tersebut untuk memulai konfigurasi. Kali ini kita kosongkan saja bagian send 'true' dan 'false' data to step karena kita akan menambahkan penghubung step secara manual nantinya. Beralih ke kotak pada baris kedua bagian condition dan pilih nama kolom yang akan difiltrasi, kolom sales_rep_employee_num. Beralih ke kotak di sebelah kanan, kotak ini bertujuan untuk menentukan function yang akan diterapkan pada nama kolom, pilih IS NOT NULL.



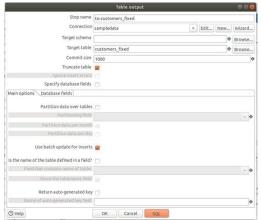
3) Aturan yang ditetapkan pada konfigurasi ini bertujuan untuk filtrasi kolom sales_rep_employee_num yang bernilai null. Melalui step ini, kita dapat mengalihkan data yang memiliki nilai pada kolom tersebut ke step selanjutnya dan data dengan nilai null pada kolom untuk diproses lebih lanjut pada step lainnya.



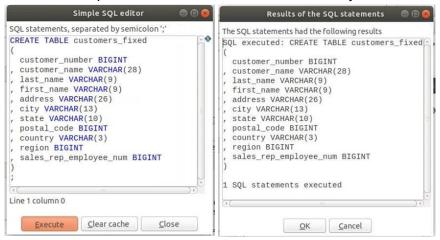
4) Setelah menyelesaikan konfigurasi, click OK. Kita lanjutkan untuk penanganan kondisi/aturan true untuk filtrasi. Pada tab Design, cari step Table output dan drag ke dalam canvas. Hover pada step filter dan pilih icon kedua dari kanan untuk menarik garis penghubung ke step Table output dan pilih Result is TRUE.



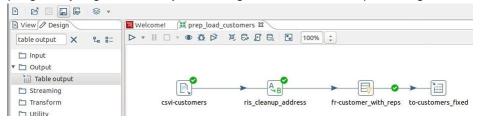
5) Lalu kita double-click pada Table output untuk memulai konfigurasi. Pada dialog konfirgurasi ini, kita juga dapat melihat SQL yang tercipta dengan click pada tombol SQL pada bagian bawah.



Melalui dialog Simple SQL Editor yang muncul setelah click SQL, kita dapat melihat hasil Query berdasarkan konfigurasi sebelumnya. Untuk menjalankan Query, click tombol Execute dan table customers_fixed pada penerapan contoh kali ini akan terbentuk pada database sampledata beserta struktur-struktur tablenya.

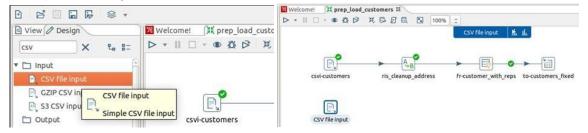


7) Table dan struktur yang terbentuk akan kita lihat pada penerapan selanjutnya. Setelah selesai konfiguras, tidak lupa untuk kita OK dan close. Tidak lupa juga untuk menyimpan progress yang telah dikerjakan dengan click icon save pada bagian kiri atas aplikasi ini.

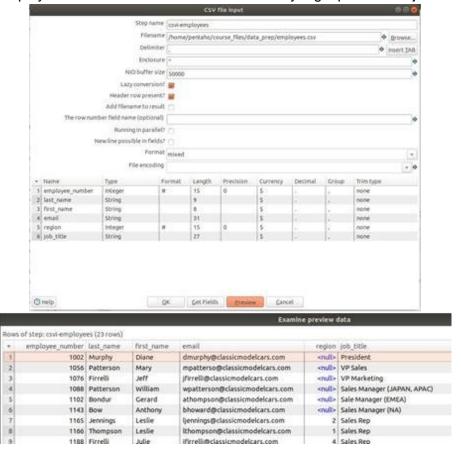


Reading Employee Data (CSV File Input Step)

 Selanjutnya kita akan menginput file employee data sebagai dasar penanganan untuk bagian false pada filtrasi data bernilai null. Sama seperti cara sebelumnya, cari step CSV file input dan drag ke dalam canvas.



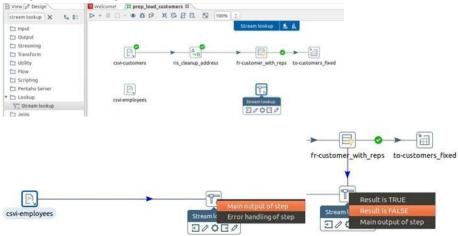
2) Lakukan konfigurasi terhadap file input dengan double-click pada step ini dan jangan lupa untuk Get Fields dan Preview pada data employee ini. Pada penerapan ini, file input data employee sudah bersih untuk kolom tertentu yang diperlukan saja.



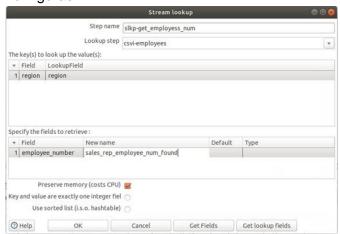
Lookup Employee Number

(Stream Lookup Step)

Selanjutnya, cari step Stream lookup pada tab Design dan drag ke dalam canvas. Jangan lupa untuk menghubungkan csvi-employees dengan step ini, dapat dilakukan dengan menahan Shift dan menarik garis dari step csvi menuju step ini, atur sebagai Main output of step. Hubungkan juga step filtrasi sebelumnya (fr-customer_with_reps) dengan step ini dengan cara yang sama dan pilihan Result is FALSE.

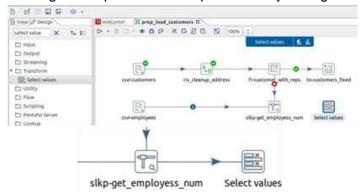


Double-click pada step dan lakukan beberapa konfigurasi. Pada step ini, data akan mengacu pada kolom region sebagai lookup value untuk data employee dan customer (yang tersaring filtrasi sebagai false karena nilai sales_rep_employee_num null). Juga menetapkan employee_number dari data employee sebagai sales_rep_employee_num_found yang nantinya akan menggantikan nilai null agar data customer terintegrasi. Jangan lupa untuk click OK dan save progress setelah selesai konfigurasi.

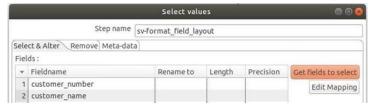


Format Field Layout (Select Values Step)

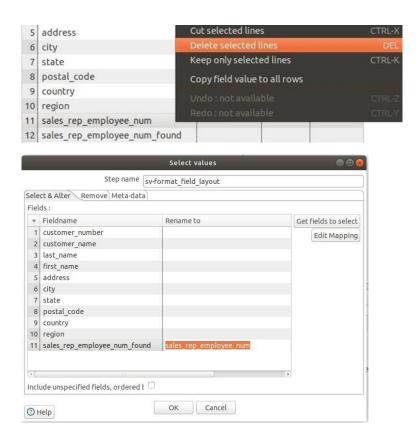
 Selanjutnya, cari step Select Values pada tab Design dan drag pada canvas, serta hubungkan step stream lookup sebelumnya dengan step ini



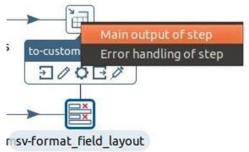
2) Double-click pada step untuk memulai konfigurasi. Silahkan click tombol Get fields to select bila tidak tertera daftar nama kolom (fields) pada tampilan konfigurasi.



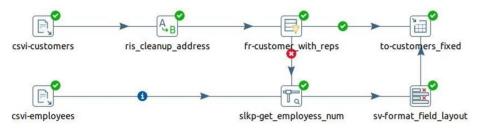
3) Cari nama kolom sales_rep_employee_num dan right-click padanya, kemudian Delete selected lines. Kolom tersebut dihapus karena memuat nilai null saja teruntuk data yang tidak lulus step filtrasi. Dan juga cari kolom sales_rep_employee_num_found dan isi pada kolom rename di sebelah kanannya dengan sales_rep_employee_num. Hal ini bertujuan untuk menetapkan employee_number berdasarkan lookup sebelumnya sebagai nilai baru untuk kolom sales_rep_employee_number dan menyesuaikan dengan kolom-kolom yang ada pada table customer_fixed.



4) Setelah konfigurasi selesai, click OK. Hubungkan juga step ini dengan step table output sebelumnya (to-customer_fixed) dan pilih Main output of step. Jangan lupa untuk melakukan save progress yang ada dengan click icon save pada kiri atas aplikasi ini.

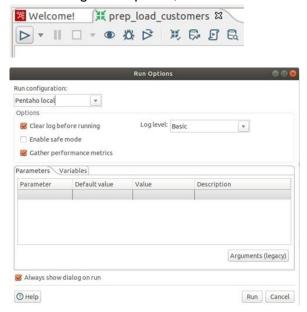


5) Susunan akhir canvas transformation:

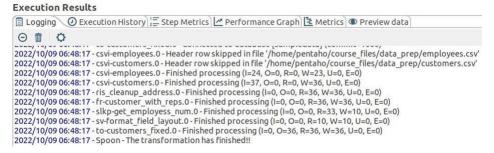


Running Your Transformation

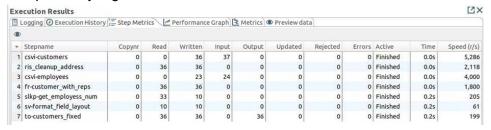
 Setelah melalui penerapan sebelumnya, transformation yang dibentuk untuk membersihkan dan mengintgrasi data customer telah selesai. Untuk menjalankan hasil transformasi, click icon play pada bagian kiri atas canvas transformation. Akan muncul kotak dialog Run Options, click Run setelah selesai konfigurasi.



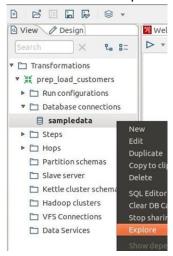
2) Proses dan hasil run transformation akan muncul pada bagian bawah aplikasi pada bagian Execution Results tab Logging. Silahkan cek lagi bila terdapat error pada logging ini dan belum mendapatkan indikasi Spoon – The transformation has finished.



3) Tab Step Metrics bertujuan untuk menunjukkan pemrosesan data pada setiap step, sebagai contoh pada step to-customers_fixed memiliki property output bernilai 36, yang artinya terdapat 36 data yang masuk ke dalam table customer_fixed pada database sampledata yang sudah dibentuk.

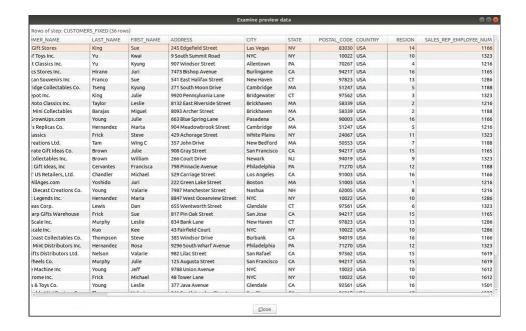


4) Untuk melihat hasil transformation pada table, pilih tab View dan expand folder Transformations, prep_load_customers (nama transformation), Database connections. Pilih database sampledata dan right-click serta pilih explore.



5) Setelah muncul kotak dialog Database Explorer, right-click pada CUSTOMER_FIXED pada bagian Tables dan pilih Preview first 100. Kemudian akan tampil data hasil transformation yang sudah masuk ke dalam table tersebut.





6) Dengan demikian, Hands On Experience untuk Pentaho Data Integration: Evaluation bagian Getting Started telah selesai

