

KULIAH CATATAN

7023T Lanjutan Sistem Database

sesi 09

Merancang dan Mengembangkan ETL Sistem

HASIL PEMBELAJARAN

- Peserta diharapkan Mampu Memahami concept dasar dasar Proses ETL (*ekstrak - mengubah - beban*).
- Peserta diharapkan DAPAT menjelaskan persyaratan Dan Kondisi (*Persyaratan Dan paksaan*) Apa Saja Yang Perlu dipenuhi Oleh ETL *alat*.
- Peserta diharapkan Mampu mengidentifikasi kemampuan Dari ETL *alat* Yang sebelumnya Saat ini distributes.
- Peserta diharapkan Mampu Memahami Proses perancangan ETL.

GARIS MATERI (Sub-Topik):

1. Pendahuluan
2. STM *dan pementasan Tabel*
3. Pengujian
4. Penanganan Kesalahan

Pendahuluan

Kimball mengusulkan 10 Langkah untuk review Membangun Sebuah Sistem ETL, baik Yang dikerjakan Beroperasi pengguna DENGAN *naskah* maupun ETL using *alat*. gambar 1 memperlihatkan diagram Yang menjelaskan ke-10 Langkah tersebut Dan kaitannya DENGAN 34 subsistem ETL Dari. Kondisi Yang Harus dipenuhi SEBELUM melakukan Proses ETL diantaranya ketersediaan desain logis, Rencana *arsitektur tingkat tinggi*, Dan STM (*sumber ke Target pemetaan*) Data SEMUA Untuk. Desain Fisik Dan Implementasi also Harus dipersiapkan sebelumnya. Hal lain Yang Penting Adalah data yang melakukan *profil* SEBELUM memulai ETL Pengembangan.

ETL PROCESS STEP	ETL SUBSYSTEM			
	EXTRACTING DATA	CLEANING AND CONFORMING	DELIVERING FOR PRESENTATION	MANAGING THE ETL ENVIRONMENT
Plan				
Create a high level, one-page schematic of the source-to-target flow.	1			
Test, choose, and implement an ETL tool (Chapter 5).				
Develop default strategies for dimension management, error handling, and other processes.	3	4, 5, 6	10	
Drill down by target table, graphically sketching any complex data restructuring or transformations, and develop preliminary job sequencing.		4, 5, 6	11	22
Develop One-Time Historic Load Process				
Build and test the historic dimension table loads.	3	4, 7, 8	9, 10, 11, 12, 15	
Build and test the historic fact table loads, including surrogate key lookup and substitution.	3	4, 5, 8	13, 14	
Develop Incremental Load Process				
Build and test the dimension table incremental load processes.	2, 3	4, 7, 8	9, 10, 11, 12, 15, 16, 17	
Build and test the fact table incremental load processes	2, 3	4, 5, 8	13, 14, 16, 18	
Build and test aggregate table loads and/or OLAP processing.			19, 20	
Design, build, and test the ETL system automation.		6	17, 18, 21	22, 23, 24, 30

Gambar 1. 10 Langkah ETL * Menurut Kimball Dan kaitannya DENGAN 34 subsistem Dari ETL

Rencana Kerja hearts Bentuk *Rencana tingkat tinggi* Seperti Yang diilustrasikan PADA gambar 2 Dan

ETL menentukan *alat* Yang DAPAT memenuhi Kebutuhan development. Selanjutnya tim

mengembangkan Pengembangan strategi untuk review melaksanakan aktifitas Utama Dari Proses ETL. Penghasilan kena pajak

Langkah dilakukan tersebut, tim Mulai melakukan transformasi untuk review menghasilkan masing-

Target tabel masing. Penghasilan kena pajak ITU, proses menerjemahkan pemuatan data yang historis Ke hearts tabel-tabel dimensi

Dan tabel Fakta dilakukan. Mengingat ukuran data yang historis Yang PADA umumnya Sangat gede,

Proses pemuatan Data DAPAT berlangsung Cukup lama, Hingga memakan Waktu sehari-hari.

Penghasilan kena pajak SEMUA tabel dimuat DENGAN historis data, Langkah berikutnya Adalah memulai Proses

inkremental, diawali DENGAN tabel-tabel dimensi kemudian dilanjutkan DENGAN tabel Fakta.

Proses inkremental Sangat tergantung Dari Aliran baris data yang Yang Proses merupakan hasil temuan penyisipan, penghapusan, maupun Pembaharuan Data Yang dilakukan sejak Proses pemuatan

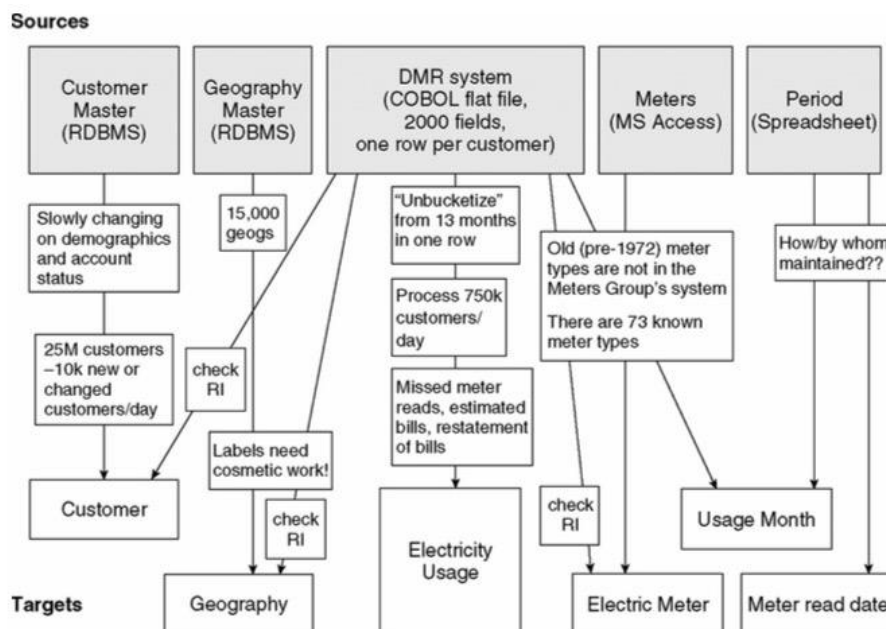
terakhir di dilakukan. Sistem ETL biasanya malakukan Proses inkremental Beroperasi rutin SETIAP

hari. Langkah berikutnya Adalah mempersiapkan tabel agregat Dari tabel Fakta. Beberapa

Langkah terakhir di Yang Perlu dilakukan Adalah mengkonfigurasi Sistem ETL agar DAPAT

beroperasi Beroperasi Otomatis, Hal Penyanyi meliputi penjadwalan, penanganan Kesalahan, Dan Operasi

pemeliharaan Database.



Gambar 2. Contoh Rencana ETL tingkat tinggi

STM dan Tabel Staging

Sistem ETL memanfaatkan tabel *pementasan* Untuk data yang menyimpan Yang Berubah sejak Proses pemuatan terakhir di data. PADA Saat Data tersebut akan dimuat Ke Sistem DW, Proses transformasi Terhadap Data PADA tabel *pementasan* PADA umumnya dilakukan Beroperasi inkremental. Tabel *pementasan* disarankan diletakkan ATAU basis data PADA *skema* terpisah, Hal Suami dimaksudkan agar DAPAT dipindahkan DENGAN Mudah Ke Server lain untuk review mendukung BEBAN Proses ETL Yang Tinggi. tabel Peletakan *pementasan* PADA database yang terpisah also Data Menjaga kerampingan Dan memudahkan Proses tranformasi Beroperasi inkremental. Tabel 1 berikut mengilustrasikan STM dari Sumber Data Ke tabel *pementasan* Dan Dari tabel *pementasan* Ke tabel dimensi DIM_SALES.

Tabel 1a. tabel Ke Ilustrasi STM dari Sumber Data *pementasan*

No	Source system	Source table	Source attribute	Transformation logic	Target table	Target attribute
1	System1	TableA	Manager_name	N/A	STG_TableA	Manager_name

Tabel 1b. tabel Ke Ilustrasi STM dari Sumber Data *pementasan*

No	Source system	Source table	Source attribute	Transformation logic	Target table	Target attribute
1	Staging	STG_TableA	Manager_name	Parse name using a comma separator	DIM_SALES	MGR_FIRST_NAME

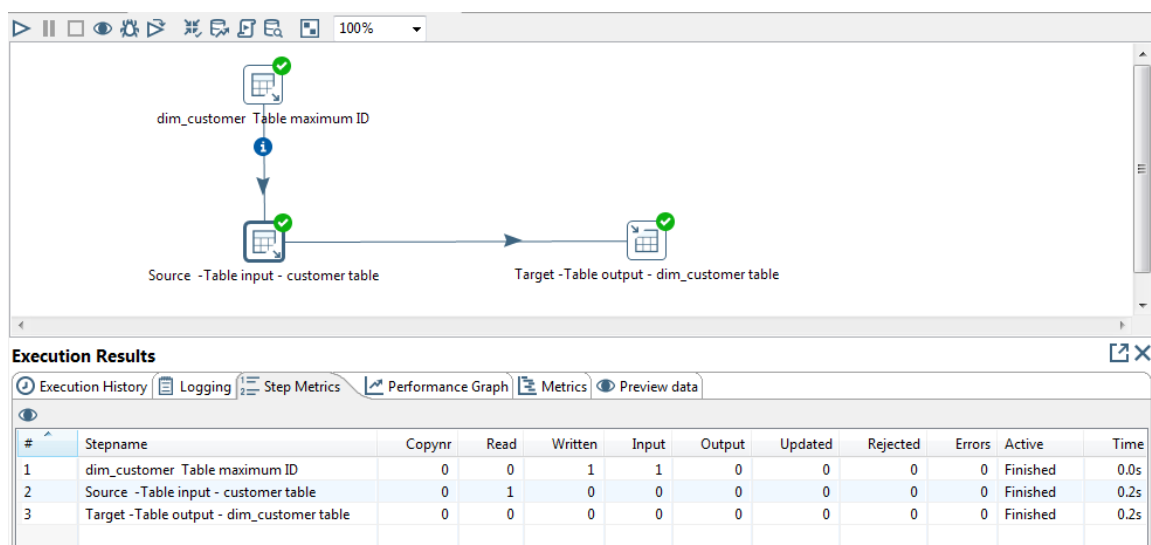
Berdasarkan information PADA tabel 1a, Proses pemuatan dari Sumber Data Ke tabel *pementasan* STG_TableA DAPAT dilakukan Beroperasi Langsung KARENA TIDAK ADA Proses transformasi Yang dilakukan (N / A). Proses Sedangkan Data pemuatan Dari tabel *pementasan* STG_TableA tabel Ke dimensi DIM_SALES Membutuhkan beberapa Proses transformasi seperti pemisahan kata berdasarkan Karakter koma, Teks penggabungan, duplikasi penghapusan, Normalisasi tali, pencocokkan, maupun *formatting*.

Pengujian

Terdapat beberapa pendekatan hearts melakukan Pengujian untuk review melakukan validasi Data Pergerakan PADA SETIAP *lapisan* Sistem DW / BI. Saat ETL Suami *alat* Sudah dilengkapi DENGAN Catatan Berupa log Yang memberikan information Beroperasi rinci Mengenai keseluruhan Proses ETL. Catatan tersebut DAPAT digunakan untuk review Keperluan Pengujian, sebagai contoh DENGAN melakukan PERBANDINGAN Jangka Waktu baris data yang diproses Yang BERHASIL PADA Sebuah *lapisan* Sumber DENGAN *lapisan* tujuan. Pendekatan lainnya, such as inviting participation Pengujian PADA Saat pemuatan Data dilakukan PERBANDINGAN ANTARA information Penjualan PADA masing-masing gerai PADA *lapisan*

Sumber DENGAN Total Penjualan untuk review gerai SEMUA PADA *lapisan* target. Salah Satu praktik Yang dianjurkan Adalah mempersiapkan Pengembangan strategi Pengujian DENGAN cara membuat sekumpulan *naskah* Untuk Pengujian Yang DAPAT digunakan untuk review Menguji Proses pemuatan berkas dan menggunakannya PADA Pengujian pemuatan Data tabel lainnya DENGAN Hanya melakukan perubahan Yang TIDAK signifikan. Pendekatan Suami akan meningkatkan Efisiensi Dan konsistensi Dari Proses Pengujian.

Gambar 3 memperlihatkan contoh log Dari Proses ETL PADA Perangkat Lunak ETL Dari Pentaho Yang disebut alat Pentaho Data Integration. Proses ETL PADA gambar tersebut dilakukan hearts Tiga Langkah, Yaitu membaca masukan doa tabel: pelanggan maximum_ID Dan Dan Memuat hasilnya Ke tabel dimensi dim_customer. Data Jangka Waktu baris Yang diproses (*salinan, membaca, menulis, memperbarui, menolak*) PADA SETIAP Langkah disajikan PADA log, termasuk JIKA ADA Kesalahan. PADA log also disajikan Waktu Yang Dibutuhkan untuk review melakukan masing-masing Langkah tersebut.

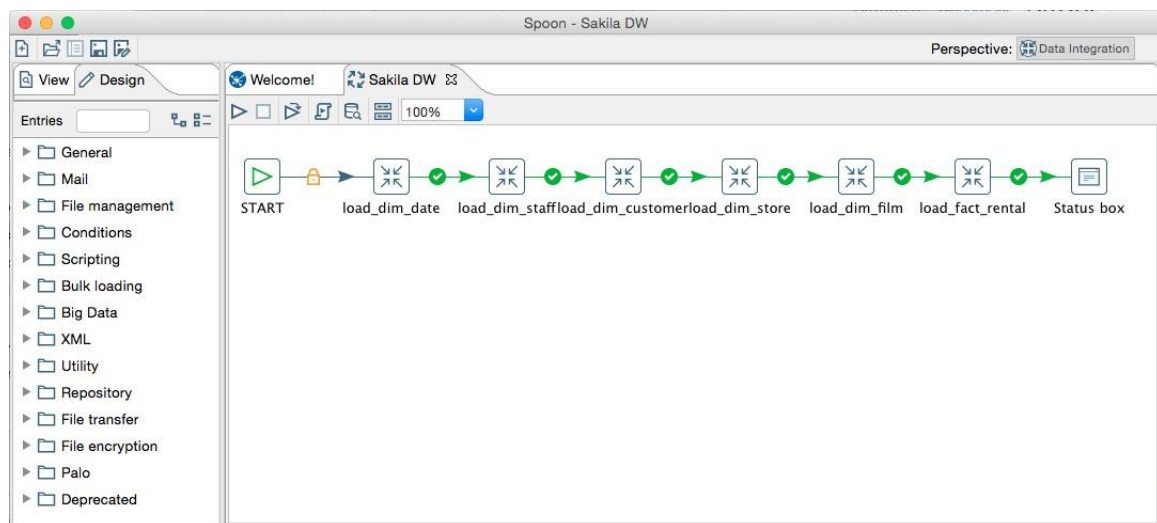


Gambar 3. Contoh log Dari Proses ETL PADA alat Pentaho Data Integration

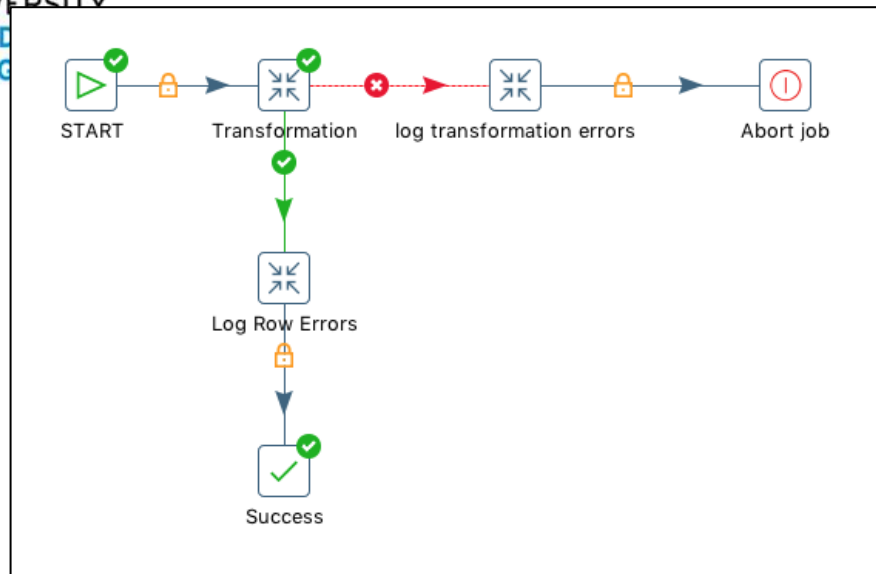
Penanganan Kesalahan

Penghasilan kena pajak SEMUA Proses ETL untuk review Memuat SEMUA tabel Selesai dirancang, SEMUA Proses tersebut Perlu diorganisasikan hearts Suatu *alur kerja* ATAU Urutan Tertentu untuk review melaksanakan Proses pemuatan tabel-tabel dimensi Dan tabel Fakta Sesuai DENGAN Urutan Tertentu. Gambar 4 memperlihatkan Sebuah *alur kerja* ETL PADA Integrasi Pentaho Data alat, PADA *alur kerja* tersebut DAPAT Dilihat bahwa pemuatan tabel-tabel dimensi (dim_date,

dim_staff, dim_customer, dim_store, Dan dim_film) dilakukan SEBELUM pemuatan tabel Fakta (Fact_rental). ETL *alat* dilengkapi DENGAN kemampuan untuk review merancang Dan Mengimplementasikan Kondisi bersyarat Saat melakukan pemuatan Data Dari different Sumber. PADA Saat Terjadi Kesalahan, Proses ETL DAPAT dikonfigurasi untuk review membatalkan Proses Data pemuatan, eksekusi melanjutkan, ATAU melewati Proses Terhadap baris data yang Yang menyebabkan Kesalahan Dan melanjutkan eksekusi Ke baris data yang berikutnya. gambar 5 memperlihatkan Sebuah contoh *alur kerja* ETL DENGAN penanganan Kesalahan. PADA gambar tersebut apabila Terjadi Kesalahan PADA Proses transformasi Maka Langkah Yang akan dilakukan Adalah MEMBUAT Catatan log Dari Kesalahan tersebut kemudian membatalkan bersaing. Sedangkan JIKA Proses tranformasi BERHASIL dilakukan, Langkah berikutnya Adalah MEMBUAT Catatan log Kegagalan Proses PADA baris Tertentu. ETL *alat* also dilengkapi DENGAN kemampuan untuk review menentukan Nilai Ambang (*ambang*) Untuk meneruskan eksekusi Hingga ditemukan Jangka Waktu Kesalahan Yang LEBIH gede daripada Ambang Yang Sudah ditentukan sebelumnya.



Gambar 4. Contoh *alur kerja* ETL PADA alat Integrasi Pentaho Data



Gambar 5. Contoh alur kerja ETL DENGAN penanganan Kesalahan

Beberapa proses menerjemahkan pemuatan data yang DAPAT diulangi Dari POSISI Dimana Mulai Terjadi Kesalahan, SEMENTARA untuk review Proses lainnya Harus diulangi Mulai dari Awal. Sebagai contoh, Data diasumsikan Penjualan Dari SETIAP stopkontak Harus dimuat SETIAP hari PADA Malam hari agar pengguna Dari Kalangan Bisnis DAPAT melakukan analisis Penjualan SETIAP hari. misalkan Proses pemuatan Data untuk review tabel-tabel dimensi Dan tabel Fakta Membutuhkan Waktu Sekitar 6 selai. Apabila Terjadi Kesalahan PADA Sebuah baris data yang Yang diakibatkan Oleh Kesalahan Format Tanggal, Maka Proses pemuatan Data untuk review baris data yang tersebut DAPAT dilewatkan Dan eksekusi dilanjutkan Ke baris data yang berikutnya. Catatan Kesalahan PADA log DAPAT diinvestigasi PADA kesempatan lain. Penghasilan kena pajak hasil temuan Investigasi menemukan Sumber kesalahannya, Maka Proses pemuatan KHUSUS untuk review Data tersebut DAPAT diulangi Lagi. * Semua Skenario terjadinya Kesalahan Dan penanganan Yang diperlukan Perlu dianalisis DENGAN baik, termasuk diantaranya menemukan Dimana Saja Proses pemuatan data yang mungkin mengalami Kesalahan Dan menentukan Prosedur penanganan Kesalahan Yang Perlu dilakukan untuk review Mengatasi SETIAP Kesalahan Yang DAPAT diidentifikasi.

SIMPULAN

- SEBELUM melakukan Proses ETL, tim ETL Perlu mendokumentasikan Rencana Kerja hearts Bentuk *Rencana tingkat tinggi* agar Proses ETL DAPAT Berjalan Sesuai Rencana DENGAN
- Tabel *pementasan* hearts berperan memproses data yang pemuatan Yang Telah Berubah sejak Proses pemuatan terakhir di dilakukan.
- Proses Pengujian Terhadap Skenario ETL Penting untuk review dilakukan untuk review mensimulasikan Kondisi Yang akan dihadapi Saat Proses ETL Sesungguhnya dieksekusi.
- Skenario Kesalahan Dan penanganannya Perlu dipersiapkan SEBELUM melakukan Proses ETL, sehingga SETIAP Kejadian Kesalahan DAPAT diantisipasi sebelumnya.

DAFTAR PUSTAKA

1. Kimball, R. (2008). *Data Warehouse Lifecycle Toolkit*. John Wiley & Sons.
2. Kimball, R., & Ross, M. (2011). *Data Warehouse Toolkit: Panduan Lengkap untuk Modeling Dimensi*. John Wiley & Sons.
3. Inmon, WH (2005). *Membangun Data Warehouse*. John Wiley & anak.