

KULIAH CATATAN

7023T Lanjutan Sistem Database

sesi 08

Introducing Extract, Transform,

Beban

HASIL PEMBELAJARAN

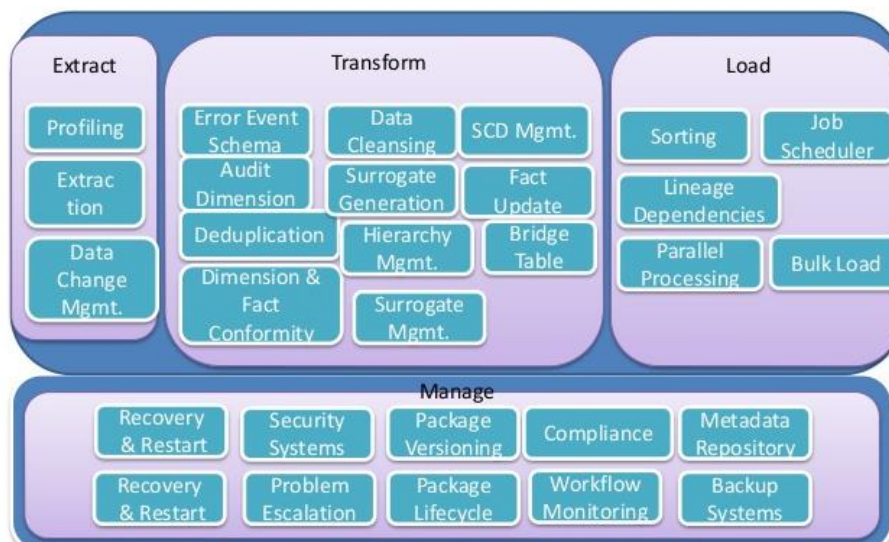
- Peserta diharapkan Mampu Memahami concept dasar dasar Proses ETL (*ekstrak - mengubah - beban*).
- Peserta diharapkan DAPAT menjelaskan persyaratan Dan Kondisi (*Persyaratan Dan paksaan*) Apa Saja Yang Perlu dipenuhi Oleh ETL *alat*.
- Peserta diharapkan Mampu mengidentifikasi kemampuan Dari ETL *alat*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

GARIS MATERI (Sub-Topik):

1. *Perkenalan pada ETL*
2. *ETL Persyaratan dan Kendala*
3. *ETL alat*
4. *ETL Desain*

Pengantar ETL

Seperti Yang Sudah dibahas PADA Bagian sebelumnya, Proses ETL terdiri Dari Tiga Operasi Utama, Yaitu: (E) Data ekstraksi data yang dari Sumber, (T) Data transformasi untuk review melakukan *pembersihan Dan sesuai*, Dan (L) *beban* - ATAU Ke Memuat Data *presentasi* Server. JIKA Operasi Pengelolaan Terhadap Tiga Operasi tersebut (ETL) also dihitung, Maka Terdapat empat dalam Proses Utama. gede Sebagian Perangkat Lunak data warehouse memasukkan ETL *alat* sebagai Bagian Utama Dari Sistem. Satu Hal Yang Penting Adalah diperhatikan untuk review Sekitar 70% Waktu Dan usaha hearts Pengembangan Sistem DW / BI akan PADA difokuskan Proses ETL. Hal lain Yang Perlu diperhatikan Adalah Harus diyakinkan bahwa ETL *alat* Yang digunakan DAPAT BEKERJA DENGAN SEMUA Sumber Data Yang akan digunakan PADA Sistem DW / BI, JIKA TIDAK Maka Proses ETL Harus dilakukan DENGAN menuliskan Cara *naskah* SQL Beroperasi manual. Kimball fungsionalitas membagi ETL Menjadi 34 subsistem seperti diperlihatkan PADA gambar 1. Keseluruhan subsistem tersebut DAPAT dibagi Menjadi 4 Kelompok, yakni *Ekstrak, Transform, beban, Dan Mengelola*. Beroperasi Umum DW / BI *alat* mungkin Saja Checklists Memverifikasi Sedikit Perbedaan fungsionalitas Dari APA Yang Oleh disarankan Kimball.



Gambar 1. 34 ETL subsistem

Persyaratan ETL dan Kendala

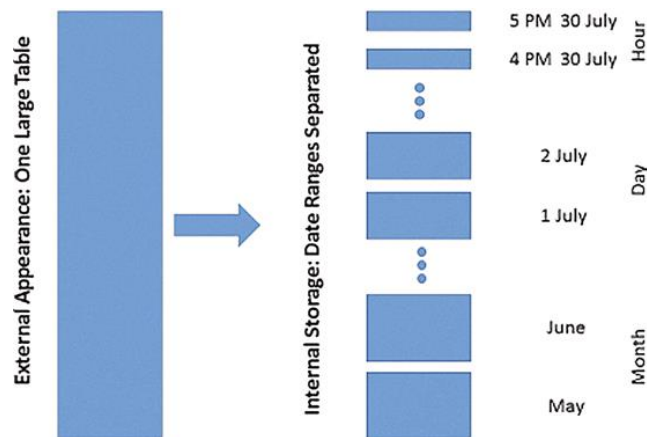
Tujuan Utama Dari Komponen ETL, seperti Komponen lain data warehouse Dari Sebuah Adalah untuk review memenuhi Kebutuhan pengguna Dari Kalangan Bisnis. Kebutuhan Sebagian

tersebut dikumpulkan Saat Tahap *definisi kebutuhan bisnis*, SEMENTARA lainnya dikumpulkan Saat dilakukan Investigasi Terhadap Sumber Data PADA Tahap *pemodelan data*. Tim ETL akan mencocokkan keseluruhan Kebutuhan *level tinggi* Terhadap Sumber Data Yang Jumlah: Tersedia. PADA Saat ditemukan ketidakcocokkan, Perlu dilakukan penyesuaian Kebutuhan agar Sesuai data yang DENGAN ketersediaan.

Sistem ETL also Perlu untuk review mengikuti beberapa persyaratan (*persyaratan dan kendala*), seperti *kepatuhan hukum, keamanan, Dan pengarsipan*. Hal tersebut ditangani Oleh subsistem Ke 32 (*sistem keamanan*) Dan subsistem 33 (*Manajer kepatuhan*). aktifitas Beberapa Yang Perlu dilakukan hearts Rangka memenuhi persyaratan tersebut, seperti MEMBUAT Salinan *backup*, Protokol memenuhi Yang Sudah ditentukan, Serta mendokumentasikan algoritma Dan bersaing. Fakta di Lapangan menunjukkan bahwa pelanggaran Terhadap *keamanan* Banyak LEBIH Berasal Dari hearts organisasi serta dibanding Dari Luar, Oleh KARENA ITU Sangat dianjurkan untuk review Checklists Memverifikasi *keamanan berbasis peran* PADA SEMUA Data Dan metadata ETL. *Pemenuhan hearts* Hal Penyanyi memiliki Makna mendokumentasikan Seluruh Proses Yang Data Berlangganan DENGAN, sehingga SETIAP Data DAPAT diketahui Dari mana asalnya, Proses transformasi APA Yang Sudah Terjadi, keasliannya, Serta salinannya.

Persyaratan ETL Yang Berlangganan DENGAN penyampaian data yang Ke Komponen-Komponen **Data termasuk lainnya *latency* Dan *format*. Data *latency* Adalah *Dari throughput yang Sistem*** ETL sebagai Bagian Dari keseluruhan Sistem DW / BI. Pemilihan Arsitektur Perangkat keras Yang Tepat Dan Lingkungan Perangkat Lunak termasuk activities Yang dilakukan hearts Rangka **memenuhi persyaratan *Data latency*. Persyaratan lain, seperti *pemrosesan real-time* Adalah** shalat Satu persyaratan Yang Umum daerah hearts inisial. *pemrosesan real-time* bermakna Kebutuhan pengguna DW agar data yang Selalu Diperbaharui Beroperasi kontinu Sepanjang hari. Pembaharuan tersebut DAPAT Saja Beroperasi *instan, sering, ATAU harian*. Pembaharuan Beroperasi Data *saat* Berarti information Yang ditampilkan PADA Layar pengguna DW / BI Diperbaharui Segera Penghasilan kena pajak Terjadi perubahan PADA Sumber data. Pembaharuan *saat* biasa ditemui hearts Sistem *Informasi Integrasi Enterprise (EII)*. Salah Satu pendekatan Yang DAPAT digunakan untuk review Proses mengoptimalkan Pembaharuan Beroperasi Data *waktu sebenarnya* Adalah DENGAN Checklists Memverifikasi teknik *partisi real-time* Yang membagi tabel Fakta Beroperasi Fisik berdasarkan Rentang Waktu Tertentu (Misal: jam SETIAP, SETIAP hari, SETIAP minggu, SETIAP bulan, Dan seterusnya). gambar 2 memperlihatkan ilustrasi penerapan teknik *real-time partisi*. Pembaharuan Beroperasi *sering* Berarti information Yang ditampilkan PADA Layar pengguna DW / BI Diperbaharui

beberapa kali hearts Sehari, such as inviting participation 15 menit mandi Sekali. Pembaharuan biasanya Suami diimplementasikan sebagai *mikro-batch* hearts Arsitektur ETL.



Gambar 2. Ilustrasi teknik *partisi real-time*

ETL *alat*

SEBELUM TERSEDIA Perangkat ETL *alat*, ETL *pembangun* Perlu menghabiskan Waktu Banyak hearts menyiapkan *naskah* Untuk melakukan Proses ekstraksi, transformasi, Dan Memuat Data kedalam data warehouse. ETL *alat* Yang sebelumnya Saat ini Jumlah: Tersedia di pasaran Sangat membantu ETL *pembangun* hearts melaksanakan tanggung-jawabnya. ETL *alat* MENYEDIAKAN Antarmuka Grafis DENGAN kemampuan *drag-and-drop* Untuk merancang Proses ETL Yang akan menghasilkan *naskah* Untuk Proses ekstraksi, transformasi, maupun pemuatan data yang Ke hearts gudang data. Script tersebut Nantinya DAPAT dieksekusi Beroperasi pengguna maupun Otomatis Sesuai DENGAN Jadwal Yang Sudah ditentukan. ETL *alat* DAPAT Mengurangi Waktu Pengembangan Beroperasi signifikan. Beberapa ETL *alat* also dilengkapi DENGAN OPSI *Template* Yang memungkinkan Sebuah Program ETL Disimpan Dan digunakan untuk review Kebutuhan Lain di hari kemudian. Gambar 3 memperlihatkan Antarmuka Grafis Dari Pentaho Data Integration. Generasi terbarunya ETL Dari *alat* dilengkapi DENGAN Dukungan Yang LEBIH Luas Terhadap Sistem Database (Oracle, SQL Server, Teradata, DB2, MySQL, PostgreSQL, dll.) tinggi, sistem ERP (SAP, PeopleSoft, dll.), Perangkat Lunak Keuangan (Netsuite, Quickbook, MYOB, dll.), Maupun *alat rekonsiliasi* seperti Hiperion, Dan Lain-lain.

Beberapa ETL *alat* MENYEDIAKAN kemampuan *kontrol versi* dan Pengembangan hearts Lingkungan *multi-pengguna*. *alat* tersebut also MENYEDIAKAN Catatan Berupa *mencatat* Yang memberikan Laporan Dari Proses ETL Yang Telah dijalankan. Catatan Penyanyi bermanfaat Sangat Untuk Keperluan *penyelesaian masalah* Dan memungkinkan analisis tren membandingkan untuk review

Waktu Yang Dibutuhkan hearts mengolah data yang PADA periode Waktu BERBEDA. Selain ITU. Catatan Berupa *mencatat* DAPAT dimanfaatkan untuk review Keperluan Pelaporan.

ETL *alat* also MENYEDIAKAN PILIHAN apakah Proses pemuatan data yang akan dilakukan

Beroperasi paralel ATAU sekuensial. Pemrosesan Beroperasi paralel memungkinkan pemuatan beberapa tabel DAPAT dilakulan Beroperasi simultan sehingga Mengurangi Waktu Proses Beroperasi keseluruhan.

ETL *alat* also dilengkapi DENGAN kemampuan penjadwalan Yang memungkinkan Proses Data pemuatan dilakukan PADA Waktu Tertentu Sesuai DENGAN Jadwal Yang ditentukan.

Komponen lain Yang Penting Dari Perangkat Lunak ETL Adalah penanganan Kesalahan Dan

Notifikasi. Mencari Google Artikel adanya Komponen Penyanyi, Proses ETL DAPAT dikonfigurasi agar membatalkan

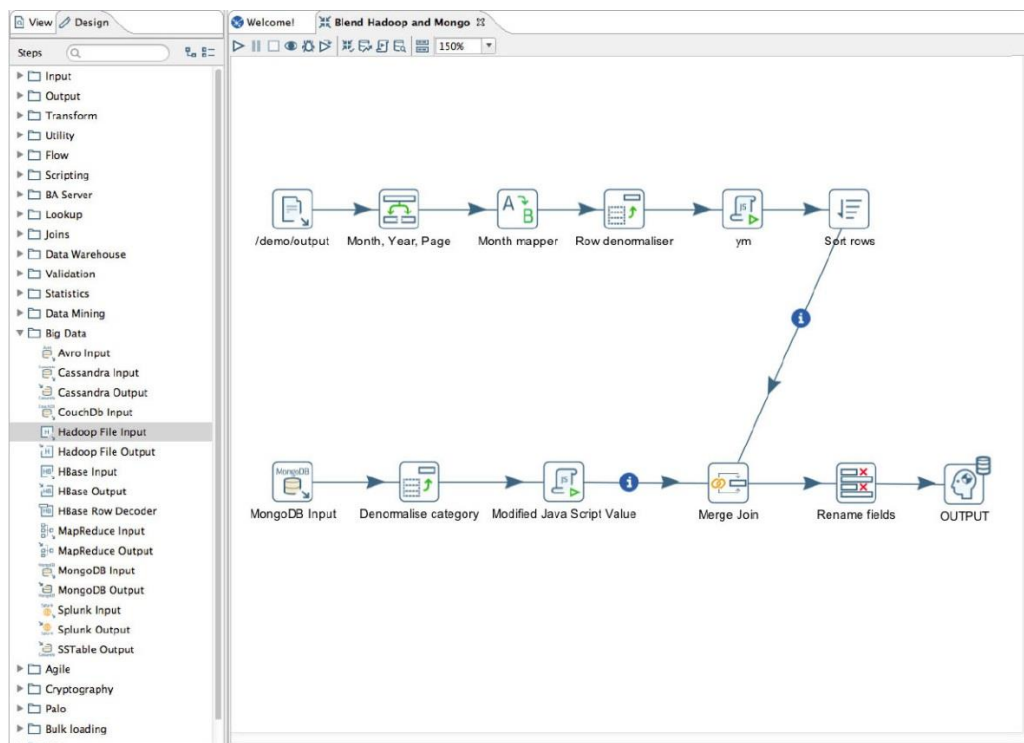
Data pemuatan Segera Penghasilan kena pajak ditemukan Kesalahan Pertama kali, ATAU meneruskan Proses

Ke pemuatan *merekam* berikutnya Penghasilan kena pajak mengalami Kesalahan PADA Saat mengolah *merekam*

Tertentu. Komponen Notifikasi DAPAT diatur agar memberikan information Berlangganan Proses ETL

Yang Sudah dijalankan through email, baik Proses ETL Yang BERHASIL dilakukan maupun PADA

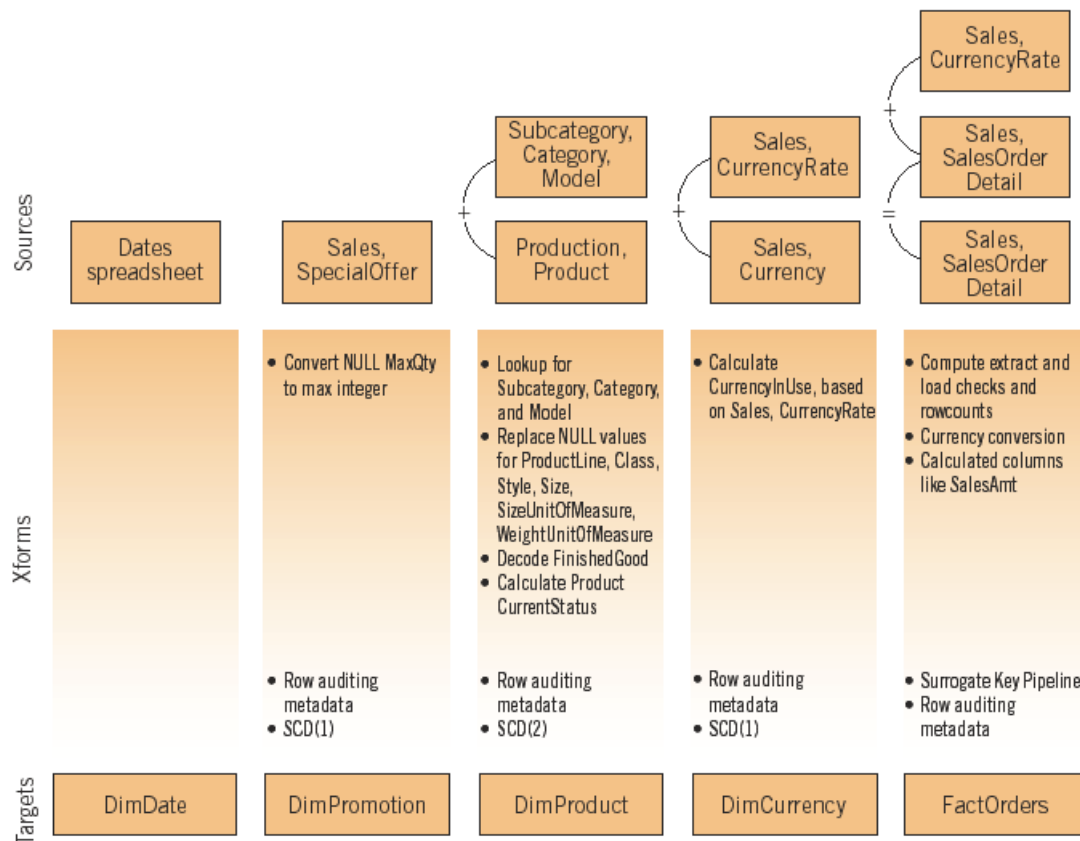
Saat menemui Kesalahan Yang mengakibatkan Proses Kegagalan.



Gambar 3. Antarmuka Grafis Integrasi Dari Pentaho Data

ETL *Desain*

Proses Untuk ETL desain, arsitek ETL Dan *pembangun* Perlu Memahami Dan tinjauan melakukan Model Terhadap Data Data Dan Arsitektur. membantu dimensi Model Proses tinjauan Terhadap tabel-tabel dimensi Dan tabel Fakta. JIKA Jumlah: Tersedia *kamus data*, Maka Hal tersebut DAPAT digunakan sebagai Langkah Awal Untuk mendokumentasikan information Mengenai Sumber Data untuk review tabel-tabel dimensi Dan tabel Fakta, Beserta atribut-atributnya. JIKA TIDAK Jumlah: Tersedia *kamus data*, Maka PT Sumber Data DAPAT Dibuat Model berdasarkan dimensi. Data Model also bermanfaat hearts mempersiapkan Dokumen Pemetaan dari Sumber Ke tujuan (*sumber untuk menargetkan pemetaan - STM*) Yang merupakan Hal Penting hearts Proses ETL. STM membantu untuk review mendokumentasikan Aturan transformasi ATAU Aturan Pemetaan Yang akan diterapkan PADA Dari Saat memindahkan data yang Satu lapisan Ke lapisan lainnya PADA DW / BI (data Sumber • *pementasan database pementasan Database* • *sumber data operasional, sumber data operasional* • DW, DW • *data mining*). gambar 4 diagram memperlihatkan *sumber tingkat tinggi untuk menargetkan pemetaan* Untuk beberapa memetakan data / tabel Sumber (Tanggal spreadsheet, Penjualan, SpecialOffer, subkategori, Cagegory, Model, Produksi, Produk, Penjualan, currencyRate, Mata, SalesOrder Detil) Menjadi empat dalam tabel dimensi (DimDate, DimPromotion, DimProduct, DimCurrency, Dan FactOrder), Proses Serta tranformasi Yang Perlu dilakukan PADA masing-masing data / tabel Sumber untuk review Menjadi tabel tujuan (dimensi Dan Fakta).



Gambar 4. Contoh sumber tingkat tinggi untuk menargetkan pemetaan

Aktivitas hearts desain ETL also termasuk mengidentifikasi Waktu Yang Tepat Untuk melakukan ekstraksi masing-masing Sumber data, memuatnya Ke beberapa lapisan DW / BI, Urutan pemuatan Dari SETIAP tabel, Dan Lain Lain. Ketersediaan Sumber Data Dari Berasal Yang different Sistem Yang BERBEDA mungkin Saja PADA WAKTU Yang BERBEDA. Sebagai contoh, information Penjualan Dari Kantor Cabang di gatra amerika latin mungkin Saja Jumlah: Tersedia Saat Tengah Malam, SEMENTARA information Yang sama Dari Kantor Cabang di gatra asia Jumlah: Tersedia Saat Tengah hari. Oleh KARENA ITU, Proses ETL Terhadap Kantor Cabang Yang terletak di benua Yang BERBEDA tersebut Perlu dilakukan PADA Saat Yang BERBEDA. Jadwal Pelaksanaan ekstraksi data yang also Perlu mempertimbangkan BEBAN Yang memucat minimum Ke Sistem operasional agar TIDAK TIDAK Terlalu mengganggu activities Bisnis Sehari-hari. Apabila Hanya Terdapat Sedikit ATAU TIDAK ADA tranformasi Yang terlibat PADA Saat pemindahan data yang Dari Ke Sumber Data *pementasan daerah*, Maka costs kos Dari Sistem operasional TIDAK akan terganggu Beroperasi signifikan.

PADA umumnya, proses menerjemahkan data yang pemuatan Harus dilakukan DENGAN Urutan Tertentu KARENA adanya dependensi ANTARA tabel-tabel Sumber. Sebagai contoh, tabel-tabel dimensi biasanya dimuat terlebih PT KARYA CIPTA PUTRA SEBELUM tabel Fakta. Hal Suami akan menghindari pelanggaran Terhadap *referensial kendala integritas*. ETL *alat* kini Telah dilengkapi DENGAN OPSI untuk review



mengkonfigurasi *alur kerja* ATAU *urutan* Untuk melakukan pemuatan tabel-tabel dimensi

Dan tabel Fakta Sesuai DENGAN Urutan Yang diinginkan.

SIMPULAN

- Sekitar 70% Waktu Dan usaha hearts Pengembangan Sistem DW / BI difokuskan akan PADA Proses ETL.
- Kimball membagi fungsionalitas ETL Menjadi 34 subsistem Yang DAPAT dikelompokkan Menjadi 4 kategori yakni *Ekstrak, Transform, beban, Dan Mengelola*.
- ETL *alat* sebelumnya Saat ini Sudah dilengkapi DENGAN Antarmuka Grafis Yang Sangat membantu ETL *pembangun* hearts melakukan Proses ETL, sehingga mereka TIDAK Perlu Lagi menghabiskan Waktu untuk review MEMBUAT *naskah* ETL.
- Dokumen STM (*sumber untuk menargetkan pemetaan*) itub hearts desain ETL, Dokumen Penyanyi membantu mendokumentasikan Aturan transformasi ATAU Aturan Pemetaan Yang akan diterapkan PADA lapisan Saat memindahkan data yang Dari Satu lapisan Ke lainnya PADA DW / BI.

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.