

UTS

DATA WAREHOUSE



Nama : Muhammad Affry

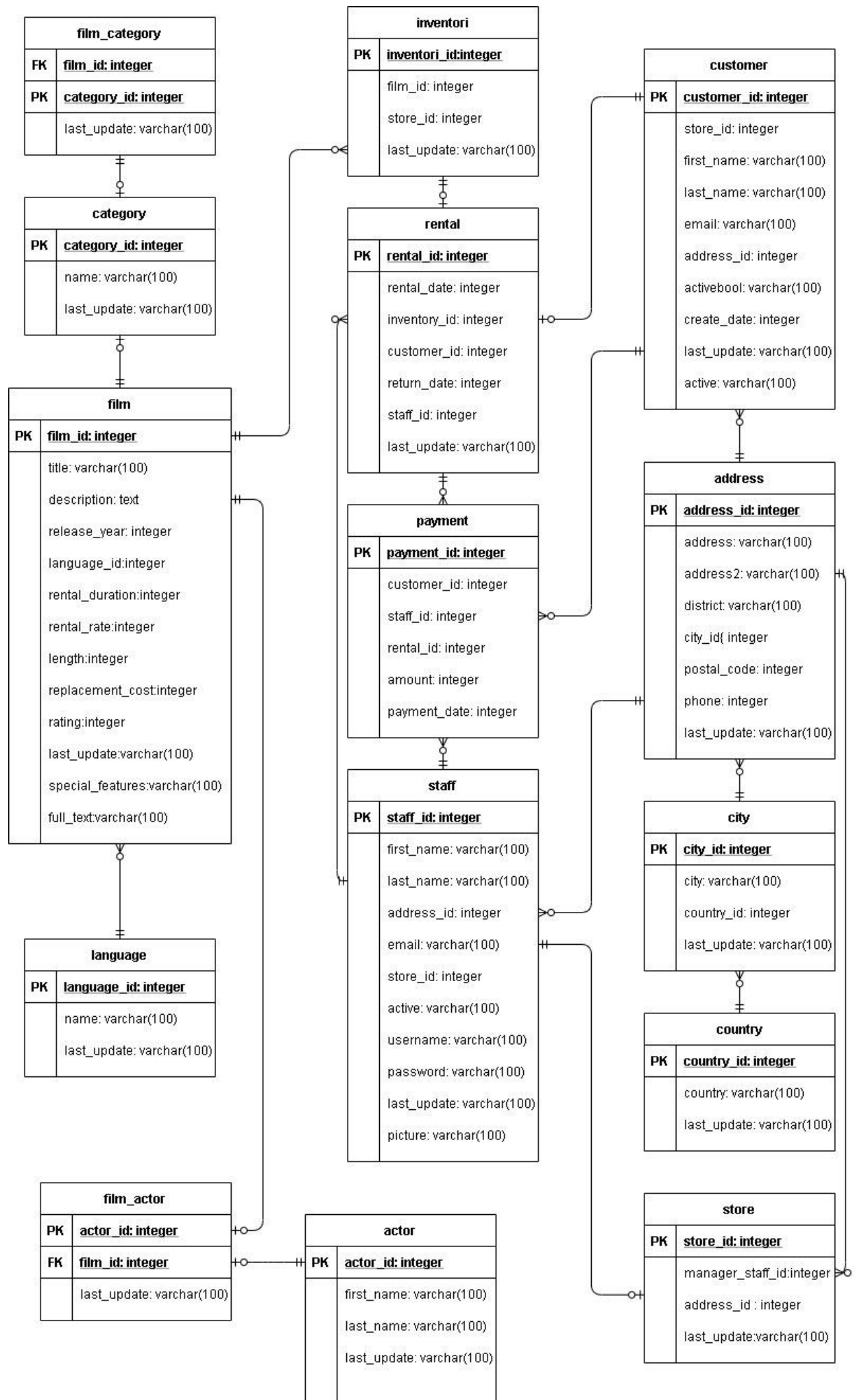
Nim : 18090026

Kelas : 5D

Dosen : Priyanto Tamami, S.Kom.

POLITEKNIK HARAPAN BERSAMA

TEGAL



berkas penjelasan desain, dan berkas proses ETL.

- Tabel Actor berisikan first_name, last_name dengan type data varchar, actor_id merelasi film_id yang ada pada table film_actor
- Tabel film_actor berisikan actor_id dengan type data integer
- Tabel film berisi tentang title, description dan lain-lain, table film merelasi table category
- Tabel language berisikan nama dengan type varchar dan language_id dengan data tipe integer
- Tabel category berisikan name, last_pdate, table category merelasikan table film_category
- Tabel inventory berisi id dari film_id, store_id
- Tabel rental yaitu table relasi dari inventory, didalam table rental yaitu rental_id, inventory_id, customer_id, dan lain-lain, dan table rental itu merelasi table paymart
- Table paymart berisikan customer_id, staff_id, rental_id, amount, payment_date. Tabel payment juga merelasi tabel yang lainnya seperti tabel customer dan tabel staff
- Tabel staff adalah tabel yang berisi data staff, didalam tabel tersebut ada first_name, last_name, dan lain-lain
- Tabel customer berisi data customer yang didalamnya berisi data store_id, first_name, last_name, dan lain-lain
- Tabel address adalah tabel yang berisikan data address yaitu address, district, city_id dan lain-lain
- Tabel city berisi data kota seperti city, country_id, last_update
- Table store yang berisi data store yaitu manager_staff_id, address_id dan last_update

ETL atau Extract Transform Load adalah proses integrasi data yang menggabungkan data dari berbagai sumber ke dalam satu penyimpanan yang konsisten dan dimuat ke dalam gudang data atau sistem lainnya.

Langkah 1: ekstraksi data

Sebagian besar perusahaan mengelola data dari berbagai sumber dan menggunakan beberapa alat analisis untuk membuat business intelligence. Namun, ada juga bisnis yang hanya bergantung pada satu jenis data atau sistem.

Jika ingin membuat strategi data yang kompleks berfungsi, data yang digunakan harus bisa bergerak bebas di antara sistem dan aplikasi.

Data harus diekstrak terlebih dahulu dari sumbernya sebelum dipindahkan ke tempat yang lain. Pada langkah pertama proses ETL ini, data terstruktur dan tidak terstruktur diimpor dan dikonsolidasikan ke dalam satu wadah penyimpanan.

Data mentah dapat diekstraksi dari berbagai sumber berikut ini:

Database yang ada dan legacy system.

Cloud, hybrid, dan on-premises environments.

Aplikasi penjualan dan pemasaran.

Mobile devices dan apps.

CRM systems.

Data storage platforms.

Data warehouses.

Analytics tools.

Langkah 2: transformasi

Setelah tahap ekstraksi selesai, data perlu dipindahkan ke sistem target atau ke sistem perantara untuk diproses lebih lanjut. Langkah selanjutnya adalah proses transformasi ETL yang akan membantumu membuat gudang data terstruktur.

Transformasi ETL merupakan pembersihan dan mempersiapkan agregasi untuk analisis. Langkah ini sangat penting dalam proses ETL karena membantu memastikan data yang akan diolah sepenuhnya siap dan kompatibel.

Proses transformasi ETL terbagi menjadi beberapa proses sebagai berikut:

Pembersihan: data yang tidak konsisten dihilangkan.

Standardisasi: memasang aturan pemformatan ke kumpulan data.

Deduplikasi: data yang sama dibuang atau dikecualikan.

Verifikasi: data yang tidak dapat digunakan dihapus dan anomali ditandai.

Pengurutan: data diatur menurut jenisnya.

Tugas lainnya - aturan tambahan yang dapat meningkatkan kualitas data.

Langkah 3: loading/memuat data.

Loading adalah proses terakhir dalam ETL, yaitu memuat data yang sudah diubah ke tujuan baru. Data tersebut dapat dimuat sekaligus (full load) atau interval terjadwal (incremental load).

Full loading

Untuk full loading ETL, semua yang berasal dari transformasi menjadi catatan baru dan unik di gudang data. Full load berguna untuk menghasilkan kumpulan data yang tumbuh secara eksponensial dan sulit untuk diatur.

Incremental loading

Metode yang ini kurang komprehensif, tetapi lebih mudah dikelola. Incremental loading membandingkan data yang masuk dengan data yang sudah ada. Dan hanya akan menghasilkan data tambahan jika ditemukan data yang unik dan baru.