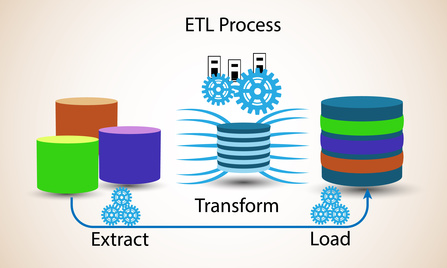
**INTEGRATION MODULE PROCESS**



**APA ITU ETL?**

* ETL adalah singkatan untuk **Extract, Transform, Load**.
* ETL adalah kumpulan proses yang menyiapkan data dari sumber operasi (operational source) untuk data .
* Ia juga merupakan proses – proses yang perlu dilalui dalam pembetukan gudang data (data warehouse).
* Tiga fungsi utama yang perlu dilakukan untuk membuat data tersedia untuk digunakan pada gudang data (data warehouse) ialah ***extration, transformation*** dan ***loading***.
* Proses ETL terdiri daripada tiga fasa iaitu:
* EXTRACT
* Extract ialah proses memilih dan mengambil data dari satu atau beberapa sumber dan menganalisa/mengakses data tersebut.
* Proses ini dapat menggunakan query atau aplikasi ETL.
* Bahagian pertama dalam proses ETL melibatkan pengekstrakan data dari sistem sumber.
* Ia merupakan aspek yang paling penting dalam ETL kerana data yang diekstrak akan menetapkan peringkat untuk kejayaan proses seterusnya.
* Kebanyakan projek gudang data menggabungkan data dari sistem sumber yang berlainan.
* Setiap sistem yang berasingan juga boleh menggunakan organisasi dan / atau format data yang berbeza.
* TRANSFORM
* Dalam fasa ini, data yang telah diekstrak akan dibersihkan dan diubah dari bentuk asli menjadi bentuk yang sesuai mengikut keperluan untuk dimuatkan ke target.
* Sesetengah data tidak memerlukan apa –apa perubahan sama sekali. Data tersebut dikenali sebagai “***direct move***” atau “***pass through***” data.
* Fungsi transformasi yang penting ialah pembersihan data (data cleaning). Ia bertujuan untuk menyampaikan data “tepat” hanya pada sasaran.
* Cabaran ketika berinteraksi dengan sistem yang berbeza ialh interfacing dan komunikasi sistem yang relevan.
* LOAD
* Load adalah fasa terakhir yang berfungsi untuk memasukkan data ke dalam sasaran terakhir yang mungkin menjadi gudang data.
* Cara untuk memasukkan data ialah dengan menjalankan SQL script secara berkala.
* Proses ini berbeza – beza bergantung pada keperluan organisasi.
* Oleh kerana fasa beban berinteraksi dengan pangkalan data, kekangan yang ditakrifkan dalam skema pangkalan data - serta pemicu diaktifkan apabila beban data - digunakan (contohnya, keunikan, integriti rujukan, medan wajib), yang juga menyumbang kepada prestasi keseluruhan data secara keseluruhan daripada proses ETL.
* Oleh kerana pengekstrakan data mengambil masa, adalah perkara biasa untuk melaksanakan tiga fasa (ETL) selari. Walaupun data sedang diekstrak(E), proses transformasi(T) lain dijalankan semasa memproses data yang telah diterima dan mempersiapkannya untuk memuatkan(L) sementara *loading* data bermula tanpa menunggu fasa sebelumnya.
* Sistem ETL biasanya mengintegrasikan data dari pelbagai aplikasi (sistem), yang biasanya dibangunkan dan disokong oleh vendor yang berbeza atau dihoskan pada perkakasan komputer yang berasingan. Sistem yang berasingan yang mengandungi data asal sering diuruskan dan dikendalikan oleh pekerja yang berbeza. Contohnya, sistem perakaunan kos boleh menggabungkan data dari senarai gaji, jualan, dan pembelian.