



Project Based Internship

ETL and ELT

Data Warehouse Formation : ETL and ELT

Daftar Isi

| | |
|--|----|
| Exploration Source | 3 |
| A. Introduction | 4 |
| B. ETL vs ELT | 4 |
| C. BAGAIMANA ETL BEKERJA? | 6 |
| 1. Extract | 6 |
| 2. Transform | 7 |
| 3. Load | 7 |
| Studi Kasus : Pengelolaan Data di Bank, manakah yang Sesuai untuk diterapkan? ETL atau ELT ? | 10 |
| References | 13 |

Exploration Source

[Article]

[ETL and Data Warehousing Explained: ETL Tool Basics | Integrate.io](#)

[Apa Perbedaan Antara ETL dan ELT?](#)

[Perbandingan ETL dan ELT: Sejarah Singkat, Definisi, dan Perbedaannya](#)

[Video]

[ETL vs ELT](#)

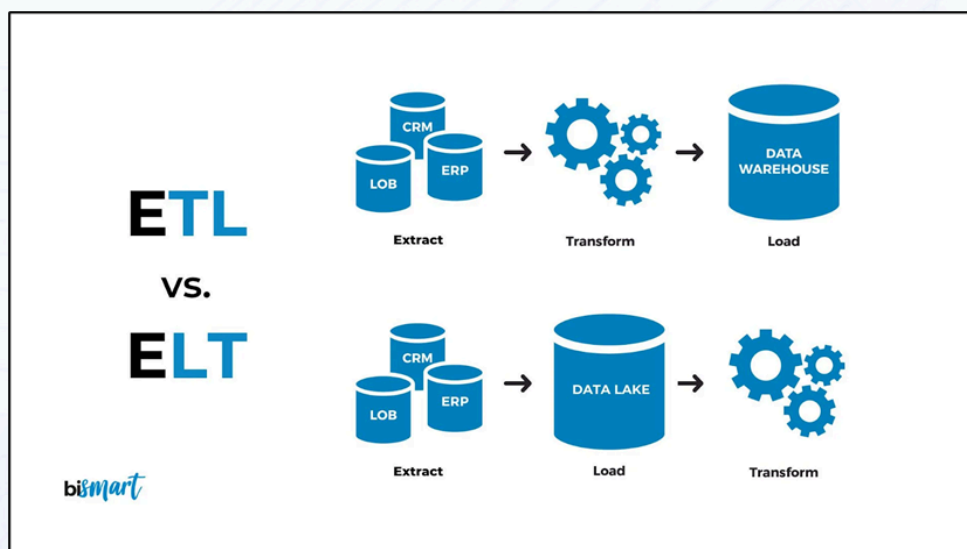
[ETL vs ELT | Modern Data Architectures](#)

A. Introduction

Karena data terus memainkan peran yang semakin penting dalam operasi bisnis modern, kebutuhan untuk mengelola dan mengubah informasi dalam jumlah besar secara efektif menjadi sangat penting. Dua pendekatan untuk tantangan ini adalah ETL (*Extract, Transform, Load*) dan ELT (*Extract, Load, Transform*). Meskipun kedua teknik tersebut bertujuan untuk meningkatkan kualitas dan aksesibilitas data, keduanya berbeda dalam urutan dan pelaksanaannya. Pada artikel ini, kita akan mengeksplorasi perbedaan antara ETL dan ELT dan memahami perbedaan kondisi penerapan ETL dengan ELT, serta contoh implementasinya pada dunia Industri.

B. ETL vs ELT

Perbedaan utama antara ETL dan ELT terletak pada lokasi di mana proses transformasi data dilakukan. Pada ETL, transformasi biasanya dilakukan sebelum data dimuat ke penyimpanan target. Sementara itu, pada ELT, transformasi dilakukan setelah data sudah ada di penyimpanan target.



Perbedaan ETL dengan ELT

Dari gambar di atas dapat dilihat bahwa posisi *Transform* pada ETL berada di tengah-tengah proses, sedangkan pada ELT ada pada akhir proses. Artinya ETL melakukan perubahan pada data di server terpisah, sementara ELT melakukan perubahan pada data di dalam *data warehouse* itu sendiri. Agar lebih mudah dipahami, perhatikan tabel berikut.

| Parameter | ETL | ELT |
|------------------------|--|---|
| Proses kirim data | ETL tidak mengirimkan data mentah langsung ke <i>data warehouse</i> | ELT langsung mengirim data mentah ke <i>data warehouse</i> |
| Transformasi ideal | ETL sangat ideal untuk transformasi yang membutuhkan komputasi intensif, sistem dengan arsitektur lama, atau alur kerja data yang memerlukan manipulasi sebelum memasuki sistem target, seperti menghapus informasi identifikasi pribadi (PII) | Transformasi akan memberatkan <i>data warehouse</i> /target karena proses transformasi tidak dilakukan ditempat yang terpisah dengan target |
| Proses penyerapan data | Lebih lambat karena transformasi data dilakukan di server terpisah sebelum proses pemuatan | Lebih cepat karena tidak mengirim data ke server sekunder untuk direstrukturisasi. Bahkan, dengan ELT, data dapat dimuat dan diubah secara bersamaan. |
| Arsip historis | Tidak menghasilkan dataset mentah yang lengkap yang dapat diminta kapan saja | ELT menciptakan arsip historis yang kaya untuk menghasilkan informasi bisnis. Ketika tujuan dan strategi berubah, tim Business Intelligence (BI) dapat meminta data mentah untuk mengembangkan transformasi |

| | | |
|--|--|--|
| | | baru menggunakan dataset yang komprehensif |
|--|--|--|

Tabel Perbedaan ETL dan ELT

Keputusan untuk menggunakan ETL atau ELT dapat dipengaruhi oleh beberapa faktor, termasuk arsitektur data, kebutuhan bisnis, dan teknologi yang digunakan. Beberapa organisasi mungkin lebih memilih ETL jika mereka memiliki kebutuhan untuk membersihkan dan mentransformasi data sebelum menyimpannya ke penyimpanan utama, sementara yang lain mungkin memilih ELT untuk memanfaatkan kekuatan pemrosesan yang ada di dalam penyimpanan target, seperti data warehouse modern yang mendukung analisis di tempat.

C. BAGAIMANA ETL BEKERJA?

Cara termudah untuk memahami cara kerja ETL adalah dengan memahami apa yang terjadi di setiap langkah proses.

1. *Extract*

Selama ekstraksi data, data mentah disalin atau diekspor dari lokasi sumber ke *staging area*. Tim manajemen data dapat mengekstraksi data dari berbagai sumber data, yang dapat terstruktur atau tidak terstruktur. Sumber-sumber tersebut termasuk tetapi tidak terbatas pada :

- SQL Server atau NoSQL
- Sistem CRM dan ERP
- *Flat files*
- E-mail

- Halaman *web*

2. Transform

Di area pementasan, data mentah mengalami pemrosesan data. Di sini, data diubah dan dikonsolidasikan untuk kasus penggunaan analitik yang dimaksud.

Fase ini dapat melibatkan tugas-tugas berikut:

- Memfilter, membersihkan, menghilangkan duplikasi, memvalidasi, dan mengautentikasi data.
- Melakukan perhitungan, penerjemahan, atau peringkasan berdasarkan data mentah. Ini dapat mencakup mengubah tajuk baris dan kolom untuk konsistensi, mengkonversi mata uang atau unit pengukuran lainnya, mengedit string teks, dan lainnya.
- Melakukan audit untuk memastikan kualitas dan kepatuhan data
- Menghapus, mengenkripsi, atau melindungi data yang diatur oleh regulator industri atau pemerintah
- Memformat data ke dalam tabel atau tabel gabungan agar sesuai dengan skema gudang data target.

3. Load

Pada langkah terakhir ini, data yang diubah dipindahkan dari staging area ke gudang data (*data warehouse*) target. Biasanya, ini melibatkan pemuatan awal semua data, diikuti dengan pemuatan berkala perubahan data tambahan dan, lebih jarang, penyegaran penuh untuk menghapus dan mengganti data di gudang. Bagi sebagian besar organisasi yang menggunakan ETL, prosesnya otomatis, terdefinisi dengan baik, berkelanjutan, dan didorong oleh batch.

Biasanya, ETL berlangsung di luar jam kerja saat lalu lintas pada sistem sumber dan gudang data berada pada titik terendah.

D. ETL and *Other Data Integration Methods*

ETL dan ELT hanyalah dua metode integrasi data, dan ada pendekatan lain yang juga digunakan untuk memfasilitasi alur kerja integrasi data. Beberapa di antaranya:

- *Change Data Capture (CDC)* mengidentifikasi dan menangkap hanya sumber data yang telah berubah dan memindahkan data tersebut ke sistem target. CDC dapat digunakan untuk mengurangi sumber daya yang diperlukan selama langkah "ekstrak" ETL; itu juga dapat digunakan secara independen untuk memindahkan data yang telah diubah menjadi data lake atau repositori lainnya secara real time.
- Replikasi data menyalin perubahan dalam sumber data secara *real time* atau dalam *batch* ke *database* pusat. Replikasi data sering terdaftar sebagai metode integrasi data. Faktanya, ini paling sering digunakan untuk membuat cadangan untuk pemulihan bencana.
- Virtualisasi data menggunakan lapisan abstraksi perangkat lunak untuk membuat tampilan data yang terpadu, terintegrasi, dan dapat digunakan sepenuhnya—tanpa menyalin, mengubah, atau memuat data sumber secara fisik ke sistem target. Fungsionalitas virtualisasi data memungkinkan organisasi untuk membuat gudang data *virtual*, *data lake*, dan *data mart* dari sumber data yang sama untuk penyimpanan data tanpa biaya dan kerumitan membangun dan mengelola platform terpisah untuk masing-masingnya.

Sementara virtualisasi data dapat digunakan bersama ETL, ini semakin dilihat sebagai alternatif untuk ETL dan metode integrasi data fisik lainnya.

- *Stream Data Integration* (SDI) persis seperti namanya—ia terus mengkonsumsi aliran data secara *real time*, mengubahnya, dan memuatnya ke sistem target untuk dianalisis. Kata kuncinya di sini adalah terus menerus. Alih-alih mengintegrasikan *snapshot* data yang diambil dari sumber pada waktu tertentu, SDI mengintegrasikan data secara konstan saat tersedia. SDI memungkinkan penyimpanan data untuk mendukung analitik, pembelajaran mesin, dan aplikasi waktu nyata untuk meningkatkan pengalaman pelanggan, deteksi penipuan, dan lainnya.

Studi Kasus : Pengelolaan Data di Bank, manakah yang Sesuai untuk diterapkan? ETL atau ELT ?

Latar Belakang :

Sebuah bank besar, yang memiliki jaringan cabang dan layanan digital yang berkembang pesat, ingin meningkatkan pengelolaan dan analisis data mereka. Bank ini memiliki berbagai sumber data, termasuk data pelanggan, transaksi keuangan, dan informasi kepatuhan. Tujuan utama adalah meningkatkan efisiensi operasional, mengoptimalkan keamanan data, dan memfasilitasi analisis data real-time untuk mendukung pengambilan keputusan strategis.

Tantangan :

1. Volume Data yang Besar

Bank menghadapi tantangan dalam menangani volume data yang terus bertumbuh dengan cepat dari berbagai sumber, termasuk transaksi harian, data pelanggan, dan aktivitas keuangan.

2. Keragaman Data

Data berasal dari berbagai sumber dan memiliki format yang beragam, termasuk data terstruktur dan tidak terstruktur.

3. Kepatuhan dan Keamanan

Penting untuk memastikan kepatuhan terhadap regulasi keuangan yang ketat dan mengimplementasikan keamanan data yang kuat untuk melindungi informasi pelanggan.

4. Analisis Real-time

Bank ingin memiliki kemampuan untuk melakukan analisis data real-time untuk mendukung keputusan operasional dan strategis.

SOLUSI :

Pilihan Solusi : ETL atau ELT ?

1. ETL (Extract, Transform, Load) :

- Pendekatan
 - 1.) Ekstraksi data dari berbagai sumber ke dalam *data warehouse*.
 - 2.) Transformasi data di server terpisah sebelum dimuat ke penyimpanan utama.
 - 3.) Pemuatan (*load*) data ke dalam *data warehouse*.
- Keuntungan
 - 1.) Kontrol penuh terhadap transformasi data.
 - 2.) Cocok untuk transformasi data yang kompleks.
 - 3.) Implementasi kebijakan keamanan dan kepatuhan selama transformasi.
- Solusi

Implementasi ETL untuk memproses data pelanggan, menyatukan transaksi, dan menyediakan data yang bersih dan terstruktur ke dalam data warehouse. Ini akan memungkinkan pengelolaan kualitas data dan kepatuhan dengan lebih baik.

2. ELT (Extract, Load, Transform) :

- Pendekatan
 - 1.) Ekstraksi data dari berbagai sumber ke dalam penyimpanan utama.
 - 2.) Pemuatan data mentah ke dalam penyimpanan utama.
 - 3.) Transformasi data dilakukan di dalam penyimpanan utama.

- Keuntungan
 - 1.) Penyerapan data yang lebih cepat.
 - 2.) Mendukung analisis *real-time*.
 - 3.) Fleksibel dan *scalable*, terutama untuk volume data besar.

- Solusi

Implementasi ELT untuk menyediakan analisis real-time dari data transaksi, mendukung inisiatif digital, dan memberikan kecepatan dalam penyerapan data.

Implementasi Hybrid :

Karena kebutuhan yang kompleks, bank dapat memilih pendekatan hybrid yang memanfaatkan keunggulan keduanya. Misalnya, menggunakan ETL untuk transformasi data yang memerlukan pemrosesan yang intensif dan ELT untuk mendukung analisis real-time dari data mentah di penyimpanan utama.

Kesimpulan

Dalam konteks perbankan yang memiliki tantangan data yang kompleks, keputusan untuk memilih ETL atau ELT harus didasarkan pada analisis mendalam terhadap kebutuhan bisnis spesifik, sumber daya yang tersedia, dan tujuan jangka panjang. Pendekatan hibrida dapat memberikan solusi yang terbaik untuk menggabungkan keuntungan dari kedua dunia, meningkatkan efisiensi operasional dan memberikan wawasan yang mendalam.



References

[Data warehouse development life cycle model](#)

[ETL Case Study with Prefect: Building a Data Pipeline for Sales Data](#)

[Apa itu ETL? Penjelasan tentang ETL - AWS](#)