

# HOME CREDIT



**Virtual Internship Experience**

## Big Data

Data Warehouse OLAP,  
OLAP VS OLTP: What's the difference?,  
ETL VS ELT: What's the difference?

## Data Warehouse - OLAP

Online Analytical Processing Server (OLAP) didasarkan pada model data multidimensi. Hal ini memungkinkan manajer atau data analis untuk mendapatkan informasi dengan akses cepat, konsisten, dan interaktif. Bab ini mencakup jenis OLAP, operasi pada OLAP, perbedaan antara OLAP, dan database statistik dan OLTP.

Terdapat empat jenis server OLAP

- Relational OLAP (ROLAP)
- Multidimensional OLAP (MOLAP)
- Hybrid OLAP (HOLAP)
- Server SQL Khusus Server

### Relational OLAP (ROLAP)

ROLAP server ditempatkan di antara server back-end dan alat front-end klien. Untuk menyimpan dan mengelola data gudang, ROLAP menggunakan DBMS relasional. ROLAP mencakup yang berikut

- Implementasi logika navigasi agregasi.
- Optimalisasi untuk setiap back end DBMS.
- Alat dan layanan tambahan.

### Multidimensional OLAP (MOLAP)

MOLAP bekerja menggunakan mesin penyimpanan multidimensi berbasis array untuk tampilan data multidimensi. Dengan penyimpanan data multidimensi, pemanfaatan penyimpanan mungkin rendah jika ketersediaan data sedikit. Oleh karena itu, banyak server MOLAP menggunakan dua tingkat representasi penyimpanan data untuk menangani kumpulan data yang banyak.

### Hybrid OLAP

Hybrid OLAP adalah kombinasi dari ROLAP dan MOLAP. Ini menawarkan skalabilitas ROLAP yang lebih tinggi dan komputasi MOLAP yang lebih cepat. Server HOLAP memungkinkan penyimpanan volume data yang besar dari informasi rinci. Agregasi disimpan secara terpisah di tempat MOLAP.

## Server SQL Khusus Server

SQL khusus menyediakan bahasa kueri tingkat lanjut dan dukungan pemrosesan kueri untuk kueri SQL melalui star schema dan snowflake schema dalam sistem read-only.

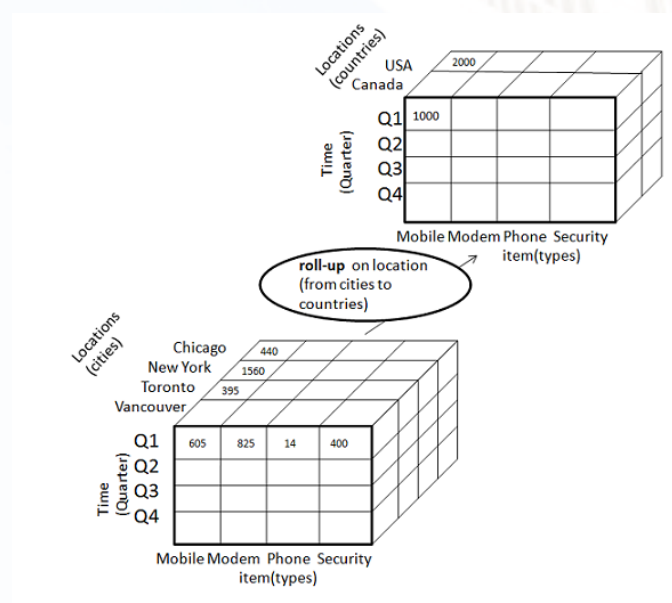
## Operasi OLAP

Karena server OLAP didasarkan pada tampilan data multidimensi, kita akan membahas operasi OLAP dalam data multidimensi. Berikut adalah daftar operasi OLAP :

- Roll-up

Roll-up melakukan agregasi pada kubus data dengan salah cara mengikuti konsep hirarki untuk suatu dimensi dan pengurangan dimensi. Diagram berikut menggambarkan cara kerja roll-up.

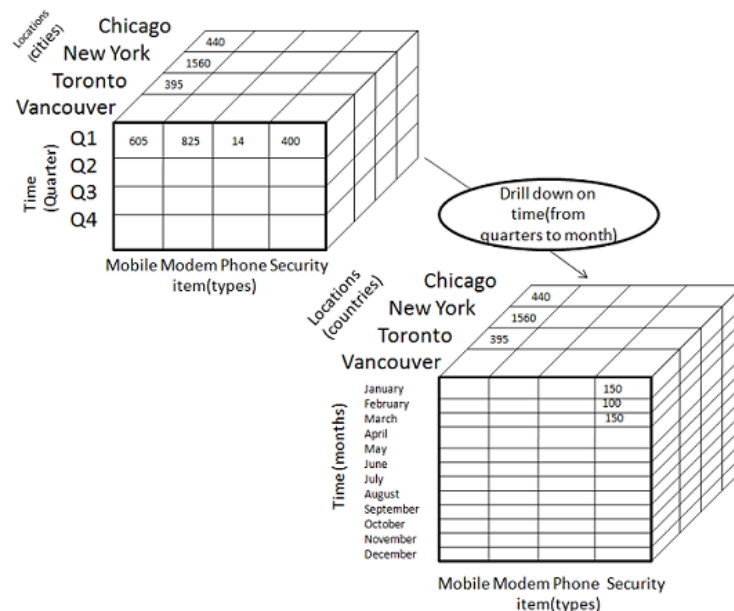
- Roll-up dilakukan dengan mengikuti konsep hierarki untuk lokasi dimensi.
- Awalnya konsep hirarki adalah "jalan < kota < provinsi < negara".
- Saat bergulir, data dikumpulkan dengan naik hierarki lokasi dari tingkat kota ke tingkat negara.
- Data dikelompokkan ke dalam kota, bukan negara.
- Saat roll-up dilakukan, satu atau lebih dimensi dari kubus data akan dihapus.



## ● Drill-down

Drill-down adalah operasi kebalikan dari roll-up. Ini dilakukan dengan salah satu cara dengan menurunkan konsep hierarki untuk suatu dimensi dan dengan memperkenalkan dimensi baru. Diagram berikut mengilustrasikan cara kerja drill-down:

- Drill-down dilakukan dengan menurunkan konsep hierarki untuk dimensi waktu.
- Awalnya konsep hierarki adalah "hari < bulan < kuartal < tahun."
- Saat menelusuri, dimensi waktu diturunkan dari level kuartal ke level bulan.
- Ketika drill-down dilakukan, satu atau lebih dimensi dari kubus data ditambahkan.
- Ini mengarahkan data dari data yang kurang rinci ke data yang sangat rinci.

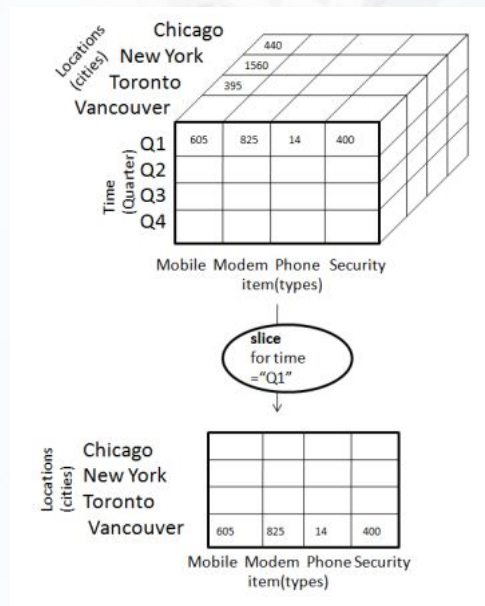


## ● Slice

Operasi Slice memilih satu dimensi tertentu dari kubus yang diberikan dan menyediakan sub-kubus baru. Perhatikan diagram berikut yang menunjukkan cara kerja irisan.

- Slice dilakukan untuk dimensi "waktu" menggunakan kriteria waktu = "Q1".
- Ini akan membentuk sub-kubus baru dengan memilih satu atau lebih dimensi.

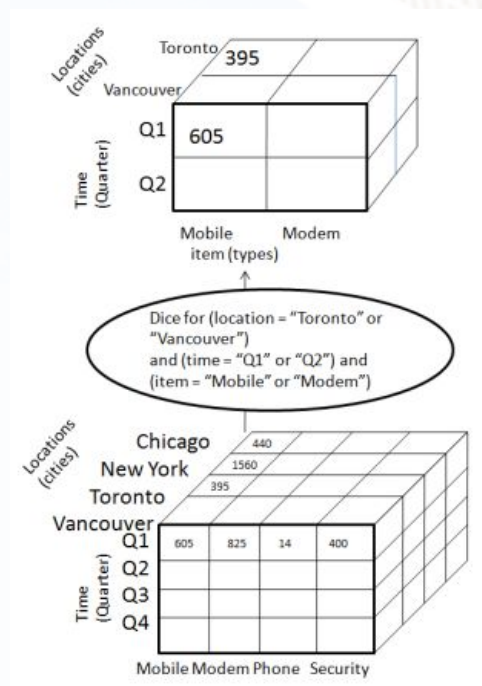




## ● Dice

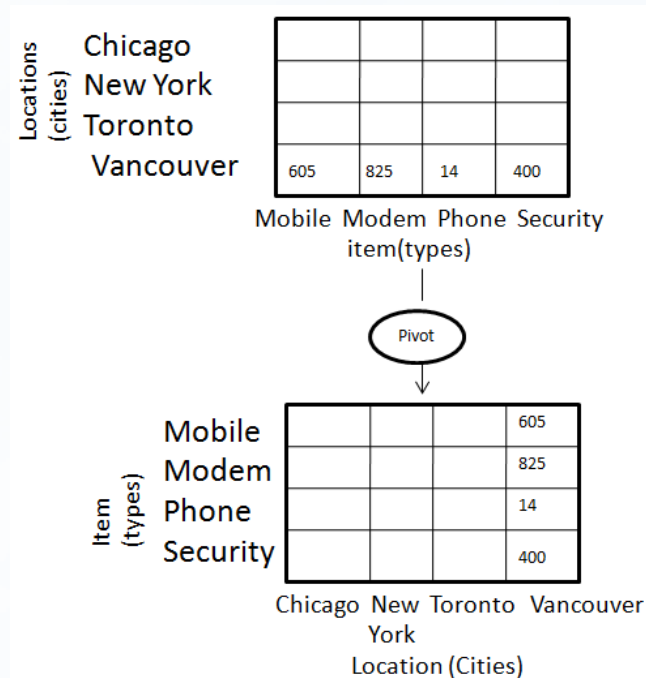
Dice memilih dua atau lebih dimensi dari kubus yang diberikan dan menyediakan sub-kubus baru. Perhatikan diagram berikut yang menunjukkan operasi dadu. Operasi dadu pada kubus berdasarkan kriteria pemilihan berikut melibatkan tiga dimensi.

- (lokasi = "Toronto" atau "Vancouver")
- (waktu = "Q1" atau "Q2")
- (item = "Seluler" atau "Modem")



- Pivot

Operasi pivot juga dikenal sebagai rotasi. Ini memutar sumbu data dalam tampilan untuk memberikan presentasi data alternatif. Perhatikan diagram berikut yang menunjukkan operasi pivot.



## Perbandingan OLAP vs OLTP

Data dari satu atau lebih database OLTP diserap ke dalam sistem OLAP melalui proses yang disebut extract, transform, load (ETL). Dengan alat ETL, pengguna dapat mengumpulkan data dari beberapa sumber dan mengirimkannya ke tujuan, seperti data warehouse OLAP. Berikut ini perbedaan OLAP terhadap OLTP.

No	Data Warehouse (OLAP)	Operational Database (OLTP)
1	Melibatkan pemrosesan informasi secara historis.	Melibatkan pemrosesan sehari-hari.
2	Sistem OLAP digunakan oleh orang berpengalaman seperti eksekutif, manajer dan data analis.	Sistem OLTP digunakan oleh pegawai, DBA, atau profesional database.
3	Berguna dalam menganalisis bisnis.	Berguna dalam menjalankan bisnis.

4	Ini berfokus pada Informasi keluar.	Ini berfokus pada Data masuk
5	Berdasarkan Star Schema, Snowflake schema dan Fact Constellation Schema.	Berdasarkan Model Relasi Entitas.
6	Berisi data historis.	Berisi data terkini.
7	Menyediakan data yang diringkas dan dikonsolidasikan.	Menyediakan data primitif dan sangat detail.
8	Menyediakan tampilan data yang diringkas dan multidimensi.	Memberikan tampilan data relasional yang mendetail.
9	Jumlah atau pengguna mencapai ratusan.	Jumlah penggunanya mencapai ribuan.
10	Jumlah record yang diakses dalam jutaan.	Jumlah record yang diakses mencapai puluhan.
11	Ukuran basis data dari 100 GB hingga 1 TB	Ukuran basis data dari 100 MB hingga 1 GB.
12	Sangat fleksibel.	Memberikan kinerja tinggi.

Secara sederhana, OLTP bersifat operasional, sedangkan OLAP bersifat informasi. Sekilas fitur utama dari kedua jenis menggambarkan perbedaan mendasar mereka, dan bagaimana mereka bekerja bersama. OLTP menyediakan catatan langsung dari aktivitas bisnis saat ini, sementara OLAP menghasilkan dan memvalidasi wawasan dari data tersebut saat dikompilasi dari waktu ke waktu. Perspektif historis tersebut memberdayakan peramalan yang akurat, tetapi seperti halnya insight data, dimana informasi yang dihasilkan dengan OLAP hanya sebaik saluran data dari mana mereka berasal.

## Data Warehousing - Relational OLAP

Relasional Server OLAP ditempatkan di antara server back-end relasional dan alat front-end. Untuk menyimpan dan mengelola data gudang, OLAP relasional menggunakan DBMS relasional atau perluasan relasional. ROLAP mencakup berikut ini:

- Implementasi logika navigasi agregasi

- Optimalisasi untuk setiap back-end DBMS
- Alat dan layanan tambahan.

## Poin yang perlu diingat

- Server ROLAP dapat diskalakan.
- Alat ROLAP menganalisis volume data yang besar di berbagai dimensi.
- Alat ROLAP menyimpan dan menganalisis data yang sangat fluktuatif dan dapat diubah.

## Keuntungan

- Server ROLAP dapat dengan mudah digunakan dengan RDBMS yang ada.
- Data dapat disimpan secara efisien.
- Alat ROLAP tidak menggunakan kubus data yang telah dihitung sebelumnya.
- Server DSS dari strategi mikro mengadopsi pendekatan ROLAP.

## Kekurangan

- Performa kueri yang buruk.
- Beberapa keterbatasan skalabilitas tergantung pada arsitektur teknologi yang digunakan.

## Data Warehousing - Multidimensional OLAP

Multidimensional OLAP (MOLAP) menggunakan mesin penyimpanan multidimensi berbasis array untuk tampilan data multidimensi. Dengan penyimpanan data multidimensi, pemanfaatan penyimpanan mungkin rendah jika kumpulan data sedikit. Oleh karena itu, banyak server MOLAP menggunakan dua bentuk penyimpanan data untuk menangani kumpulan data yang padat dan sedikit.

## Hal yang perlu diketahui

- Alat MOLAP memproses informasi dengan waktu respons yang konsisten terlepas dari tingkat perhitungan yang dipilih.
- Alat MOLAP perlu menghindari banyak kerumitan dalam membuat database relasional untuk menyimpan data untuk analisis.
- Alat MOLAP membutuhkan kinerja secepat mungkin.
- Server MOLAP mengadopsi dua bentuk penyimpanan untuk menangani kumpulan data yang padat dan jarang.
- Sub-kubus yang lebih padat diidentifikasi dan disimpan sebagai struktur array.

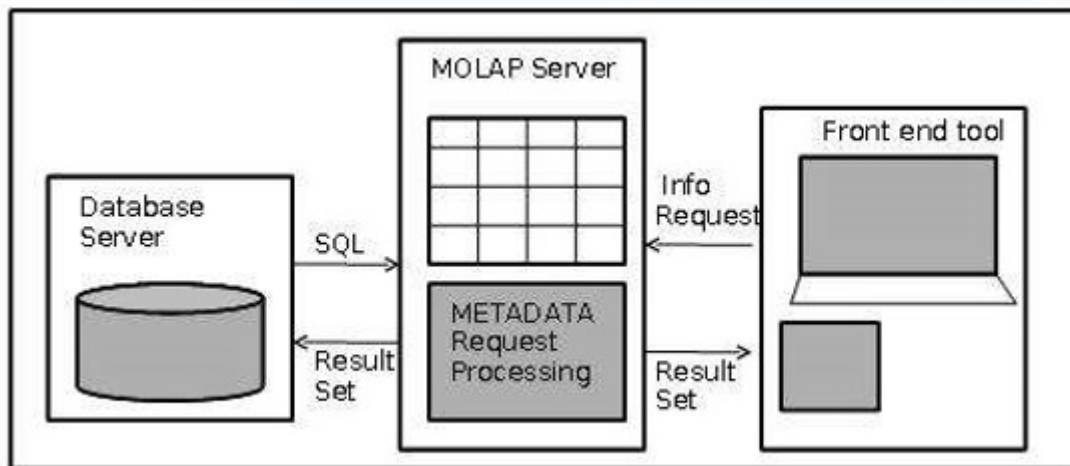


- Sub-kubus yang jarang menggunakan teknologi kompresi.

## MOLAP

MOLAP mencakup komponen-komponen berikut

- Database server.
- server MOLAP.
- Alat ujung depan.



## Keuntungan

- MOLAP memungkinkan pengindeksan tercepat ke data ringkasan yang telah dihitung sebelumnya.
- Membantu pengguna yang terhubung ke jaringan yang perlu menganalisis data yang lebih besar dan kurang jelas.
- Lebih mudah digunakan, oleh karena itu MOLAP cocok untuk pengguna yang belum berpengalaman.

## Kekurangan

- MOLAP tidak mampu memuat data yang detail.
- Pemanfaatan penyimpanan mungkin rendah jika kumpulan data jarang.

## MOLAP vs ROLAP

Sr.No.	MOLAP	ROLAP
1	informasi cepat.	Pencarian informasi relatif lambat.
2	Menggunakan sparse array untuk menyimpan kumpulan data.	Menggunakan tabel relasional.
3	MOLAP paling cocok untuk pengguna yang tidak berpengalaman, karena sangat mudah digunakan.	ROLAP paling cocok untuk pengguna berpengalaman.
4	Memelihara database terpisah untuk kubus data.	Ini mungkin tidak memerlukan ruang selain yang tersedia di gudang Data.
5	Fasilitas DBMS lemah.	Fasilitas DBMS kuat.

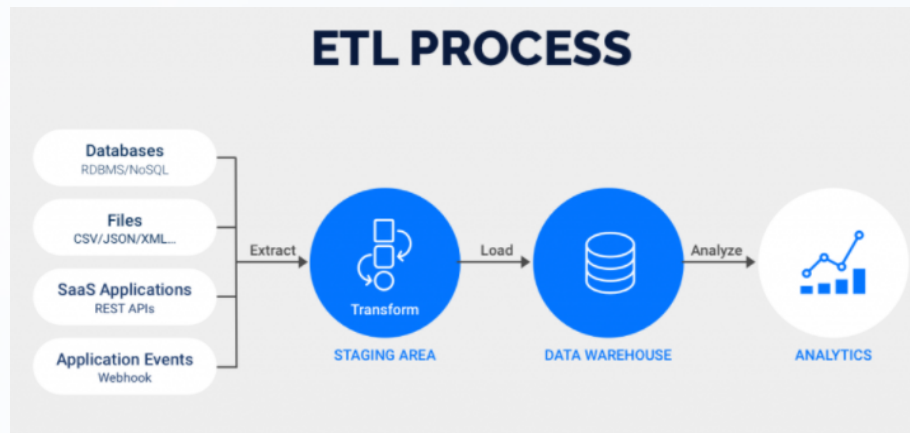
## Perbandingan ETL dan ELT

### Ekstrak, Transform, dan Load (ETL)

Ekstrak, transformasi, dan muat (ETL) adalah integrasi data yang mengekstrak data dari sumber, mengubah data pada server pemrosesan sekunder, dan kemudian memuat data ke database target. ETL digunakan ketika data harus diubah agar sesuai dengan data dari database target. Metode ini muncul pada tahun 1970an, dan database lokal yang memiliki memori dan kekuatan pemrosesan yang terbatas. Online Analytical Processing (OLAP) hanya menerima struktur data berbasis SQL relasional. Dengan jenis data warehouse ini, protokol seperti ETL memastikan kepatuhan dengan mengarahkan data yang diekstraksi ke server

pemrosesan, dan kemudian mengubah data yang tidak sesuai menjadi data berbasis SQL.

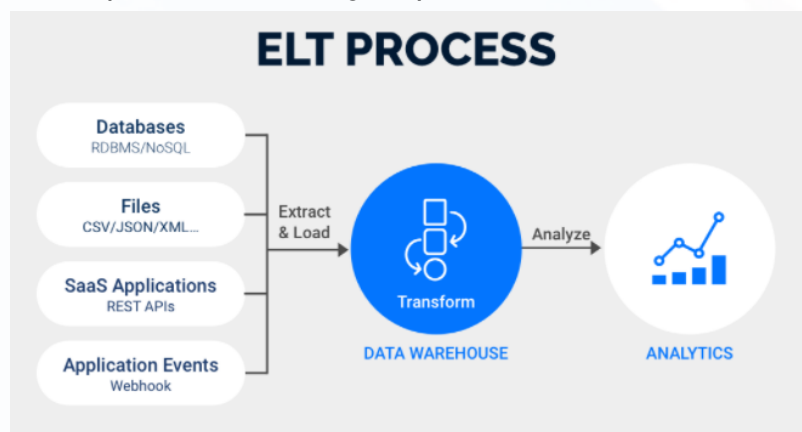
Data yang diekstraksi hanya berpindah dari server pemrosesan ke data warehouse setelah berhasil diubah.



### ETL (Extract, Load, Transform)

Extract, Load, dan Transform (ELT) tidak memerlukan transformasi data sebelum proses loading. ELT memuat data mentah langsung ke data warehouse target dengan memindahkannya ke server pemrosesan untuk transformasi.

Dengan ELT, pembersihan dan transformasi semua terjadi di dalam gudang data itu sendiri. Data mentah disimpan tanpa batas di gudang data, memungkinkan banyak transformasi. ELT adalah perkembangan yang relatif baru, dimungkinkan oleh penemuan gudang data berbasis cloud yang dapat diskalakan. Data warehouse cloud seperti [Snowflake](#), [Amazon Redshift](#), [Google BigQuery](#), dan [Microsoft Azure](#) semuanya memiliki infrastruktur digital, dalam hal penyimpanan dan kekuatan pemrosesan, untuk memfasilitasi penyimpanan data mentah dan transformasi dalam aplikasi. Meskipun ELT tidak digunakan secara universal, metode ini menjadi lebih populer karena perusahaan mengadopsi infrastruktur cloud.



## ETL vs ELT: Apa Bedanya ETL dengan Proses ELT?

ETL dan ELT berbeda dalam dua hal utama. Satu perbedaan adalah di mana data ditransformasikan, selain itu, bagaimana data warehouse tersebut menyimpan data.

- ETL mengubah data pada server pemrosesan terpisah, sementara ELT mengubah data di dalam data warehouse itu sendiri.
- ETL tidak mentransfer data mentah ke data warehouse, sedangkan ELT mengirimkan data mentah langsung ke data warehouse.

Untuk ETL, proses penyerapan data diperlambat dengan mengubah data di server terpisah sebelum proses pemuatan.

ELT memberikan penyerapan data yang lebih cepat, karena data tidak dikirim ke server sekunder untuk direstrukturisasi. Bahkan, dengan ELT, data dapat dimuat dan diubah secara bersamaan. Retensi data mentah ELT menciptakan arsip historis yang kaya untuk menghasilkan insight bisnis. Saat tujuan dan strategi berubah, tim bisnis dapat meminta data mentah untuk mengembangkan transformasi baru menggunakan kumpulan data yang komprehensif. Selain itu, ETL tidak menghasilkan kumpulan data mentah lengkap yang dapat diminta tanpa henti.

Faktor-faktor ini membuat ELT lebih fleksibel dan efisien, terutama untuk menyerap data dalam jumlah besar, memproses kumpulan data yang berisi data terstruktur dan tidak terstruktur, dan mengembangkan kecerdasan bisnis yang beragam. Di sisi lain, ETL sangat ideal untuk transformasi komputasi intensif dengan arsitektur lama, atau alur kerja data yang memerlukan manipulasi sebelum memasuki sistem target, seperti menghapus informasi pengenalan pribadi (PII).

Kategori	ETL	ELT
Definisi	Data diekstraksi dari sistem sumber, diubah pada server pemrosesan sekunder, dan dimuat ke sistem tujuan.	Data diekstraksi dari sistem sumber, dimuat ke sistem tujuan, dan diubah di dalam sistem tujuan.
Ekstrak	Data mentah diekstraksi menggunakan konektor API.	Data mentah diekstraksi menggunakan konektor API.
Transform	Data mentah ditransformasikan pada server pemrosesan.	Data mentah diubah di dalam sistem target.



Load	Transformasi data dimuat ke sistem tujuan.	Data mentah dimuat langsung ke sistem target.
Kecepatan	ETL adalah proses yang memakan waktu; data ditransformasikan sebelum dimuat ke sistem tujuan.	ELT lebih cepat dibandingkan; data dimuat langsung ke sistem tujuan, dan ditransformasikan secara paralel.
Transformasi Berbasis Kode	Dilakukan pada server sekunder. Terbaik untuk transformasi intensif komputasi & pra-pembersihan.	Transformasi dilakukan dalam database; memuat & mengubah secara bersamaan; kecepatan & efisiensi.
Maturity	Modern ETL telah ada selama 20+ tahun; praktik & protokolnya terkenal dan didokumentasikan.	ELT adalah bentuk integrasi data yang lebih baru; kurang dokumentasi & pengalaman.
Privasi	Transformasi pra-muat dapat menghilangkan PII (membantu untuk HIPPA).	Pemuatan data secara langsung membutuhkan lebih banyak perlindungan privasi.
Pemeliharaan	Server pemrosesan sekunder menambah beban pemeliharaan.	Dengan sistem yang lebih sedikit, beban pemeliharaan berkurang.
Biaya	Server terpisah dapat menimbulkan masalah biaya.	Tumpukan data yang disederhanakan lebih murah.
Permintaan	Data ditransformasikan sebelum memasuki sistem tujuan; oleh karena itu data mentah tidak dapat diminta kembali.	Data mentah dimuat langsung ke sistem tujuan dan dapat diminta kembali tanpa henti.
Kompatibilitas Data Lake	Tidak, ETL tidak memiliki kompatibilitas data lake.	Ya, ELT memang memiliki kompatibilitas data lake.

Data Output	Terstruktur (biasanya).	Terstruktur, semi terstruktur, tidak terstruktur.
Volume Data	Ideal untuk kumpulan data kecil dengan persyaratan transformasi yang rumit.	Ideal untuk kumpulan data besar yang membutuhkan kecepatan & efisiensi.

## Sejarah Singkat: Proses ETL & ELT

Seluruh Bisnis pasti memerlukan integrasi data untuk memusatkan, mengakses, dan mengaktifkan data di seluruh organisasi mereka. Bisnis harus memanfaatkan puluhan atau ratusan sumber data yang berbeda, lintas negara, benua untuk mendorong hasil dalam ekonomi berbasis data. Dalam menggabungkan beberapa sumber data ke dalam tampilan terpadu tidak pernah lebih penting. Namun tantangan ini bukanlah hal baru. Integrasi data telah mengganggu organisasi sejak awal era digital.

Penemuan pertama dari era komputer modern meletakkan dasar bagi fungsi unik dari integrasi data. Pada akhir 1960-an, penyimpanan disk menggantikan kartu punch, memungkinkan akses langsung ke data. Tidak lama kemudian, IBM dan perusahaan lain merintis Database Management System (DBMS) yang pertama. Kemajuan ini segera menyebabkan berbagi data antar komputer. Hampir seketika, proses rumit dalam mengintegrasikan data dan sumber data dengan mesin eksternal menjadi tantangan.

ETL, metode standar pertama untuk memfasilitasi integrasi data, muncul pada 1970-an. ETL menjadi terkenal karena bisnis perusahaan mengadopsi sistem komputer multi-cabang dan sumber data yang heterogen. Bisnis ini membutuhkan cara untuk mengumpulkan dan memusatkan data dari transaksi, sistem penggajian, log inventaris, dan data perencanaan sumber daya perusahaan (ERP) lainnya.

Dengan munculnya data warehouse di tahun 1980-an, ETL menjadi semakin penting. Data warehouse dapat mengintegrasikan data dari berbagai sumber, tetapi biasanya memerlukan ETL khusus untuk setiap sumber data. Sehingga hal ini membutuhkan banyak proses ETL. Pada akhir 1990-an, banyak dari solusi ini akhirnya menjadi terjangkau dan terukur untuk bisnis pasar menengah, bukan hanya perusahaan besar.

Dengan munculnya komputasi awan pada tahun 2000-an, cloud data lake dan data warehouse menyebabkan evolusi baru: ELT. Dengan ELT, bisnis dapat memuat tanpa batas langsung ke cloud DWH. Tim engineer dan analis dapat mengeksekusi query SQL dalam jumlah tak terbatas di atas data mentah ini, langsung di dalam cloud data warehouse itu sendiri. Untuk pertama kalinya, bisnis dapat membuka kunci daya

analitik dan efisiensi yang selalu dijanjikan oleh data besar. Dikombinasikan dengan alat visualisasi dan DWH cloud, ELT mengantarkan era baru analitik dan pengambilan keputusan berbasis data.

## Mana yang Lebih Baik: ETL atau ELT?

Data Warehouse Cloud telah memunculkan batas baru dalam integrasi data, tetapi memilih antara ETL dan ELT bergantung pada kebutuhan tim. Meskipun ELT menawarkan keuntungan baru yang menarik, beberapa tim akan tetap menggunakan ETL karena metode ini masuk akal untuk penerapan khusus mereka, infrastruktur lama atau tidak. Apapun pilihannya, tim data di seluruh spektrum menemukan keberhasilan dalam menjalankan strategi integrasi mereka dengan memanfaatkan platform integrasi data.

### Referensi :

- TutorialsPoint , Data Warehousing - OLTP, from [https://www.tutorialspoint.com/dwh/dwh\\_olap.htm](https://www.tutorialspoint.com/dwh/dwh_olap.htm) , [acces:24/03/2022].
- Oracle , What is OLTP, from <https://www.oracle.com/database/what-is-oltp/> , [acces : 24/03/2022]
- Stitch, OLTP and OLAP: A Practical comparison, from <https://www.stitchdata.com/resources/oltp-vs-olap/> [Acces: 26/03/2022].
- Kevin Bartley, ETL VS ELT : What's the difference?, from <https://rivery.io/blog/etl-vs-elt/#:~:text=ETL%20transforms%20data%20on%20a,directly%20to%20the%20data%20warehouse.> [acces: 25/03/22].