



## DATAWAREHOUSE

PERTEMUAN-02

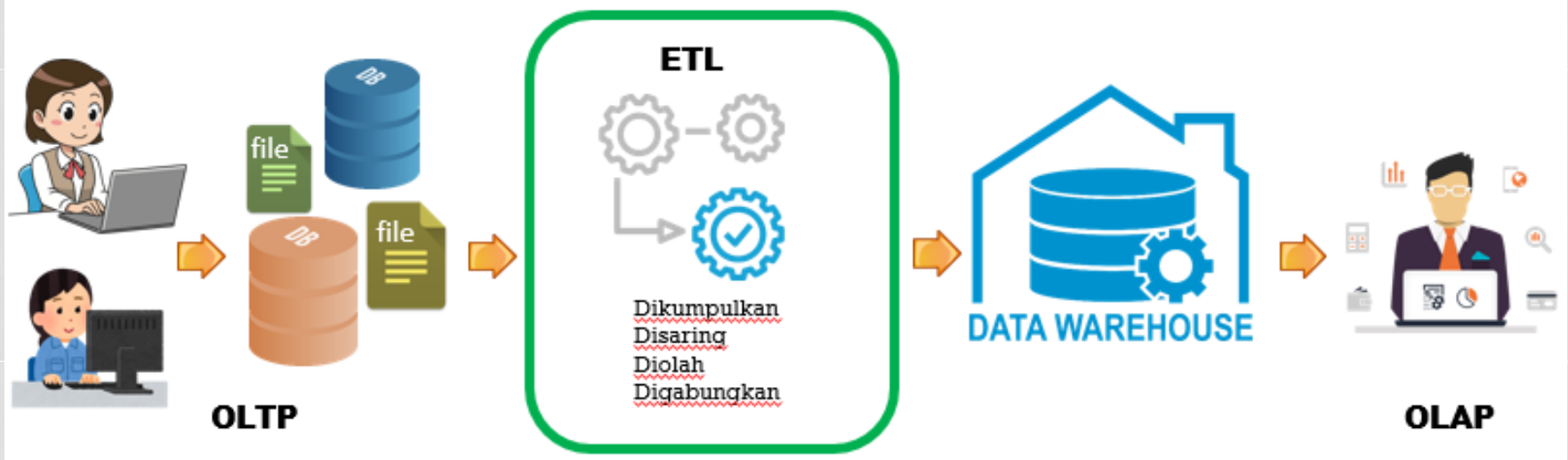
## KONSEP ETL

*Team Datawarehouse 2020*



# Apasih ETL itu ?

- ETL → Extract Transform Load
- Adalah fase pemrosesan data dari sumber data masuk ke dalam data warehouse.
- mengumpulkan, menyaring, mengolah dan menggabungkan data-data yang relevan dari berbagai sumber untuk disimpan ke dalam data warehouse.
- Tujuan Utama adalah membentuk kembali data dari resources menjadi informasi/data yang baik dan relevan pada gudang
- Kelebihan : dapat mempermudah suatu pihak untuk menganalisa dan atau memproses ke tahap lebih lanjut sehingga dapat menemukan kesimpulan/hasil dari analisa data.





- Proses mengambil data dari berbagai macam jenis sumber data dari OLTP
- Karakteristik ekstraksi data :
  - Data yang diolah memiliki volume besar (Big Data)
  - Membutuhkan memory yang cukup besar dalam proses ekstraksi, karena data yang besar dan proses yang cepat.
- Pada Ekstraksi data ada 2 jenisnya:
  - Full Extraction
  - Partial Extraction



- Pengubahan format agar sesuai dengan struktur yang ada di database DW
- Tujuannya adalah melakukan pembersihan data dengan menghilangkan data yang tidak bermakna/ tidak dibutuhkan, kemudian mengubah bentuk data sesuai dengan bentuk yang dibutuhkan.
- Transformasi data ada 3 jenis:
  - Selection
  - Matching
  - Data Transforming

# ETL LOAD



- Memasukkan data yang sudah disesuaikan formatnya ke dalam tabel-tabel yang ada pada Data Warehouse.
- Bagian akhir dalam proses dari ETL
- Data sudah sesuai dengan kebutuhan dan siap diolah pada data warehouse
- Data Loading ada 3 strateginya:
  - Initial Load
  - Incremental Load
  - Full Refresh

## TUGAS



- Buka dan pelajari Jobsheet 00 dan kerjakan tugas sesuai dengan perintah. Kumpulkan ke dosen pengampu !



**TERIMA KASIH**