## Что такое ETL? И чем оно отличается от ELT?

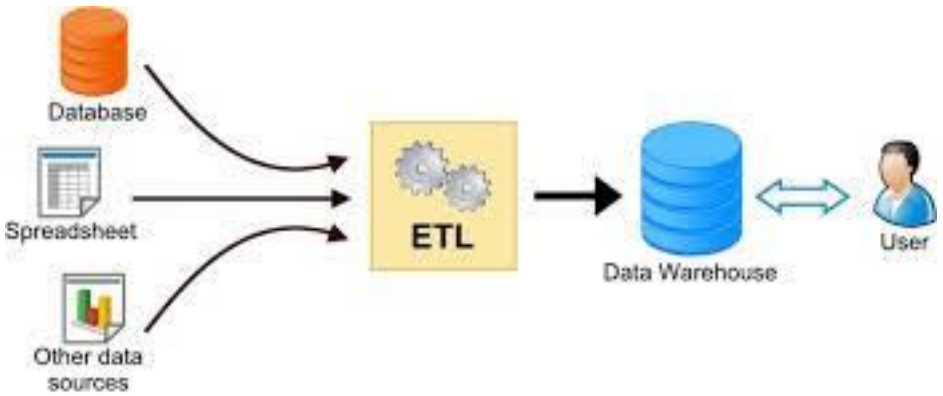
- ETL (*Extract, Transform, Load*) - ничего больше, чем обычный непрерывный процесс с четко определенным рабочим процессом. Идея заключается в том, что сначала вычитывает/извлекает данные из источников (например с Hadoop nodes, elasticsearch). После данные очищаются, преобразуются и сохраняются в хранилище данных.

- ELT (*Extract, Load, Transform*) является подобием ELT с небольшим нюансом. Сначала данные загружаются в хранилище данных, а уже потом обрабатываются, очищаются, и преобразуются.

Зачем такое разнообразие? Как правило, это хорошо работает, данные довольно велики и мощны для обработки. Частные базы данных такие как BigQuery от Google, RedShift от Amazon используются в ELT, потому что они очень эффективны при выполнении преобразований.

*Для чего нужны такие системы?*

Нужно понимать, что твориться в данных (Big Data). Иметь представление о них, вести некую отчетность.
Как то привести данные к одной структуре, уменьшения вероятность возникновения ошибок на уровне считывание/записи/копировании данных. А также детализация структур.



Поэтапный действия в ETL:

- как было сказано раньше, загрузка - является главным, первоначальным этапом в такой системе. Цель - загрузить с источника нужное количество данных. При проверки хешей как в rsync, если не буду совпадать, то является серьезной ошибкой.
- происходит проверка данных на цельность. Логируются ошибки для последующих исправлений.
- так как целью такой системы является привести неоднородные данные в единую систему, то данным этапом является привод к единой таблицы (Mapping)
- Существует две системы OLAP (Online Analitical Processing) & OLTP (Online Transction Processing), где первая является ненормализированой таблицей. Характеризируется тем что имеет довольно низкий обьем транзакций и в последствии - запросы являются очень нетривиальными, которые вмещают в себя агрегации.  OLTP в свою очередь характеризуется тем что имеет великий обьем транзакций. Рассчитан на быстрое записывание, считывание, поддержание интегрирование в мульти-доступных окружениях.
- Итоговым шагом является запись в след систему или целенаправленое хранилище