

Процесс ELT (ELT-process)

Синонимы: ELT, Extract-Load-Transform, Извлечение-загрузка-преобразование

Разделы: [Бизнес-задачи](#)

ELT — это процесс переноса данных из разнородных источников в [хранилище данных](#) (ХД) с целью их дальнейшего анализа. В целом, процесс ELT выполняет те же функции, что и [ETL](#) с той только разницей, что этапы загрузки и преобразования меняются местами.

ELT извлекает данные из исходных местоположений, но вместо перемещения их в промежуточную область для преобразования, загружает необработанные данные непосредственно в приемник, где их можно преобразовать по мере необходимости («на лету», «по вызову») в соответствии с конкретными целями и задачами анализа.

Однако порядок шагов — не единственное отличие ELT от ETL. В ELT целевым местом [консолидации](#) данных может быть не только ХД, но и [озеро данных \(Data Lake\)](#), которое представляет собой большое централизованное хранилище, где данные содержатся в их исходном, необработанном формате.

ETL преобразует данные перед их перемещением в центральный репозиторий, что упрощает администрирование, поскольку обработка реализуется в соответствии с предварительно заданным регламентом, однако аналитикам может оказаться более интересен процесс ELT, который позволяет им «играть в песочнице» сырых, необработанных данных и выполнять собственные преобразования, адаптированные к конкретным задачам анализа.

Для наглядности поясним отличие процессов ETL и ELT в следующей таблице:

Характеристики	ETL	ELT
Внедрение технологии, доступность инструментов и специалистов	ETL используется десятки лет, является отлаженной технологией с развитым инструментарием и большим количеством специалистов	ELT сравнительно новая технология, поэтому могут возникнуть сложности с поиском специалистов и проектированием процесса

Характеристики	ETL	ELT
Доступность данных	В ХД загружаются только данные, которые нужны для анализа, в соответствии с текущими настройками процесса ETL	Позволяет загрузить в хранилище все данные, а уже затем аналитик сам определит какие из них следует преобразовать и анализировать.
Поддерживаемые хранилища данных	Работают с локальными и облачными ХД, требуют <u>структурированных данных</u> .	Ориентированы на работу с облачными хранилищами и озерами данных, могут работать с неструктурированной, слабоструктурированной и структурированной информацией.
Порядок работы	Преобразование данных происходит сразу после извлечения из источников, в области временного хранения, затем выполняется загрузка	Данные загружаются в хранилище или озеро данных сразу после извлечения. Их преобразование производится по мере необходимости
Время ожидания загрузки данных	Длительное, из-за предшествующего процесса преобразования	Малое, поскольку преобразование производится после загрузки

Применение ELT дает следующие преимущества:

- 1. Быстрота.** Поскольку затратный в вычислительном и временном плане этап преобразования не предшествует загрузке, пользователи получают интересующие данные быстрее. Это особенно важно, в случаях, когда требуемая скорость принятия решений на основе данных высока, а риск ухудшения качества решений из-за того,

что данные не обработаны, невелик. Такое имеет место, когда источники хорошо управляются, и из них поступают качественные данные. Примером может являться фондовый рынок, который генерирует большие объемы данных, используемых для принятия решений (покупать/продавать) в режиме реального времени.

2. **Разделение задач.** Поскольку данные преобразуются после загрузки, у пользователя появляется больше возможностей для управления и манипулирования ими. При этом гарантируется, что ошибки или ограничения этапа преобразования не повлияют на загрузку данных.
3. **Масштабирование.** Затратные в вычислительном плане процессы преобразования данных реализуются распределенными структурами ХД, которые могут по мере необходимости наращивать число узлов обработки. Особенно это выражено в облачных структурах, где доступны вычислительные кластеры, обеспечивающие масштабируемость.
4. **Непрерывность работы.** Для любой среды, требующей быстрого доступа к данным, необходимо обеспечить непрерывную работу. ELT хорошо подходит для данных, используемых в облачных средах, которые часто включают приложения, доступ к которым осуществляется по запросу, но в непрерывном режиме.

Тем не менее, при переходе от архитектуры ETL к архитектуре ELT может возникнуть ряд проблем:

- работа с «сырыми данными» требует от пользователей более высокой квалификации по сравнению с получением отчетов из заранее спроектированного хранилища;
- поскольку информация загружается в ХД в сыром, необработанном виде, может произойти снижение качества полученных аналитических решений;
- сложно обеспечить единый согласованный взгляд на данные, т.к. любой аналитик имеет возможность применить специфичные алгоритмы предобработки;
- отсутствие анонимизации и шифрования данных, которые обычно производятся на этапе преобразования, могут создать проблемы с безопасностью;
- различная логика процессов может потребовать полной реконфигурации информационной инфраструктуры компании.