

Процесс ETL (ETL-process)

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

Разделы: [Бизнес-задачи](#)

Процесс ETL представляет собой комплекс операций, реализующих процесс переноса первичных данных из различных источников в [аналитическое приложение](#) или поддерживающее его [хранилище данных](#). Является составной частью этапа [консолидации](#) в [анализе данных](#).

ETL-приложения извлекают информацию из одного или нескольких источников, преобразуют ее в формат, поддерживаемый системой хранения и обработки, которая является получателем данных, а затем загружают в нее преобразованную информацию.

Изначально ETL-системы использовались для переноса информации из более ранних версий различных информационных систем в более новые. В настоящее время они находят все более широкое применение именно для консолидации данных с целью их дальнейшего анализа.

Очевидно, что поскольку хранилища данных могут строиться на основе различных моделей данных ([многомерных](#), [реляционных](#), [гибридных](#)), то и процесс ETL должен разрабатываться с учетом всех особенностей используемой в хранилище модели. Кроме этого, желательно, чтобы ETL-система была универсальной, т.е. могла извлекать и переносить данные как можно большего числа типов и форматов.

Независимо от особенностей построения и функционирования ETL-система должна обеспечивать выполнение трех основных этапов процесса переноса данных:

- **Извлечение данных** — на этом шаге данные извлекаются из одного или нескольких источников и подготавливаются к этапу [преобразования](#). При этом следует отметить, что для корректного представления данных после их загрузки в хранилище должны быть извлечены не только сами данные, но и информация, описывающая их структуру, из которой будут сформированы [метаданные](#) для хранилища.
- **Преобразование данных** — производится преобразование форматов и кодировки данных, а также их [интеграция](#) и [очистка](#);
- **Загрузка данных** — запись преобразованных, интегрированных и очищенных данных в соответствующую систему хранения.

Все операции над данными в процессе ETL производятся в так называемой **промежуточной области**, где для этого создаются временные таблицы.

Альтернативным вариантом процесса переноса данных из различных источников в хранилище или аналитическую платформу является процесс ELT — extract, load, transform (извлечение, загрузка, преобразование). Как следует из названия, он отличается от классического ETL-процесса тем, что после этапа извлечения сразу следует этап загрузки данных. А процесс преобразования проходит уже в структурах хранилища.