

Специфика применения ETL в различных предметных сферах.

Урок 8



План курса

1

Модели данных и нормализация таблиц. Схема «звезда».

2

Введение в подготовку данных для аналитиков. Таблицы фактов и таблицы измерений.

3

Получение денормализованных таблиц из нормализованных.

4

Партиционирование данных по дате. Динамическое партиционирование.

5

Обзор возможностей Airflow, установка и настройка.

6

Операторы в Airflow и их применение для ETL.

7

Построение пайплайнов и визуализация потоков данных в Airflow.

8

Специфика применения ETL в различных предметных сферах



Что будет на уроке сегодня



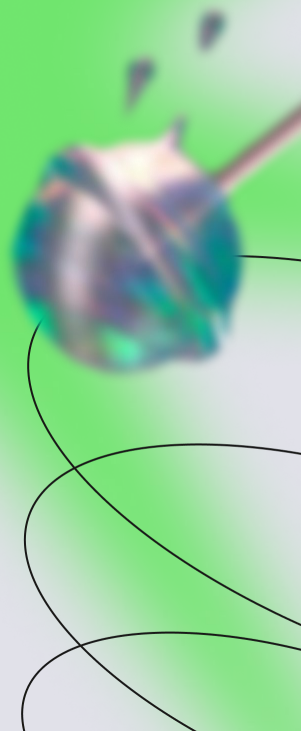
ETL в области науки о клинических данных/здравоохранения



Использование ETL не в полном цикле



ETL в области науки о клинических данных/здравоохранения





Жизненный цикл ETL

- Спецификации
- Проверка данных
- Создание правил и генерация запросов
- Тестирование и отладка
- Отчеты о качестве данных



ETL инструменты

Инструменты ETL — это технологические решения, предназначенные для облегчения процессов ETL.

Enterprise	Open-source	Custom
SAS Data Management	Talend Open Studio	Custom tools using programmming languages
Fivetran	Pentaho Data Integration	
Oracle data integrator	Singer	



Хранилище клинических данных (CDW)

Хранилище данных — это коллекция старых данных, организованная для отчетов и исследований. Это упрощает доступ к данным, объединяя и связывая данные из различных источников, что делает их легко доступными.

Хранилище клинических данных (CDW) или **Хранилище клинических данных (CDR)** — это база данных в режиме реального времени, которая объединяет данные из различных клинических источников для представления единого представления об одном пациенте.

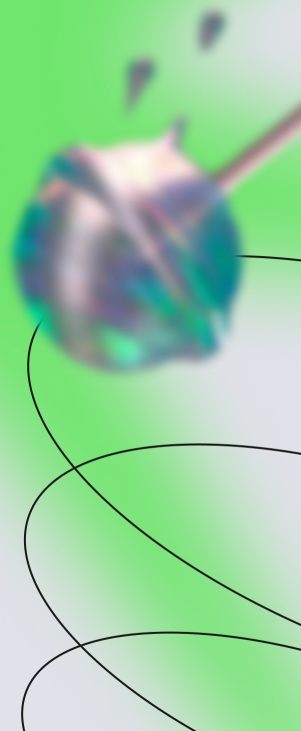


Проблемы с ETL в здравоохранении

- совместимость между исходными и целевыми данными
- качество исходных данных
- масштабируемость процесса ETL
- отсутствию единообразия в данных
- большой объем

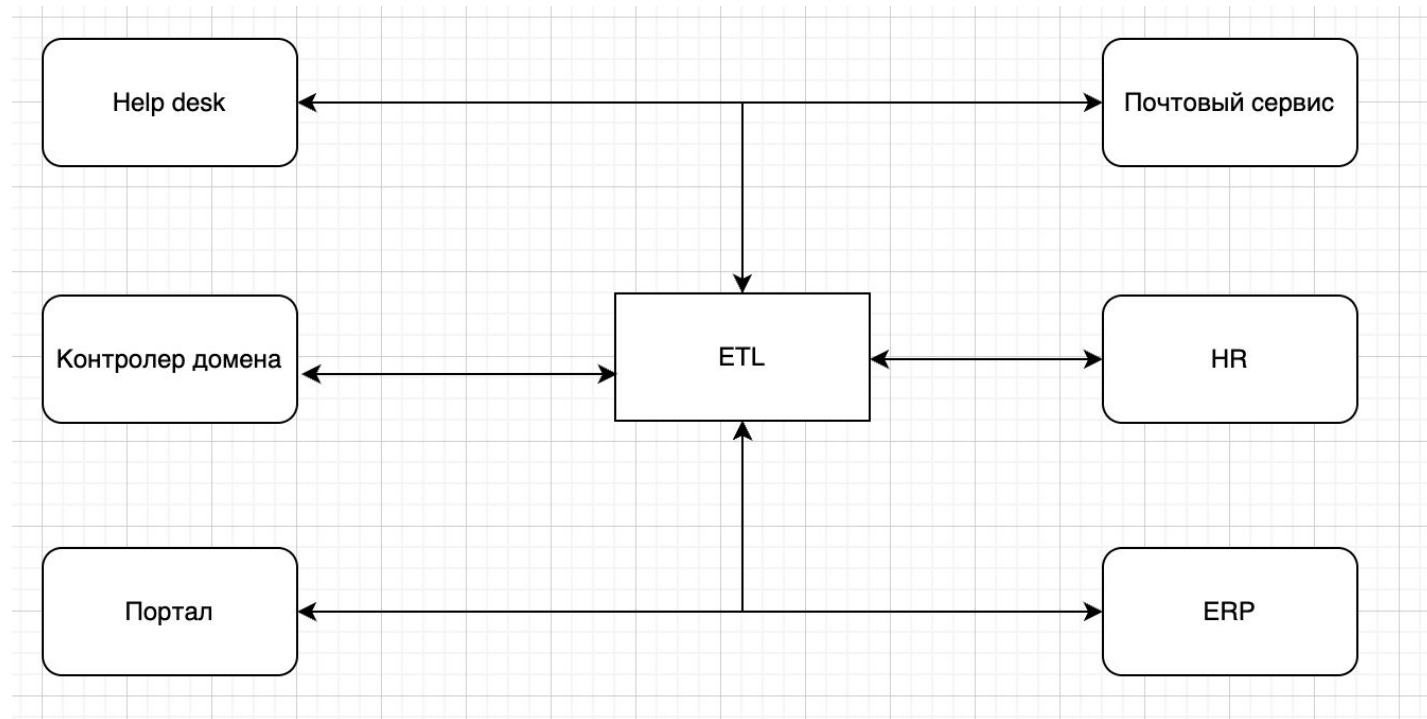


Использование ETL
не в полном цикле



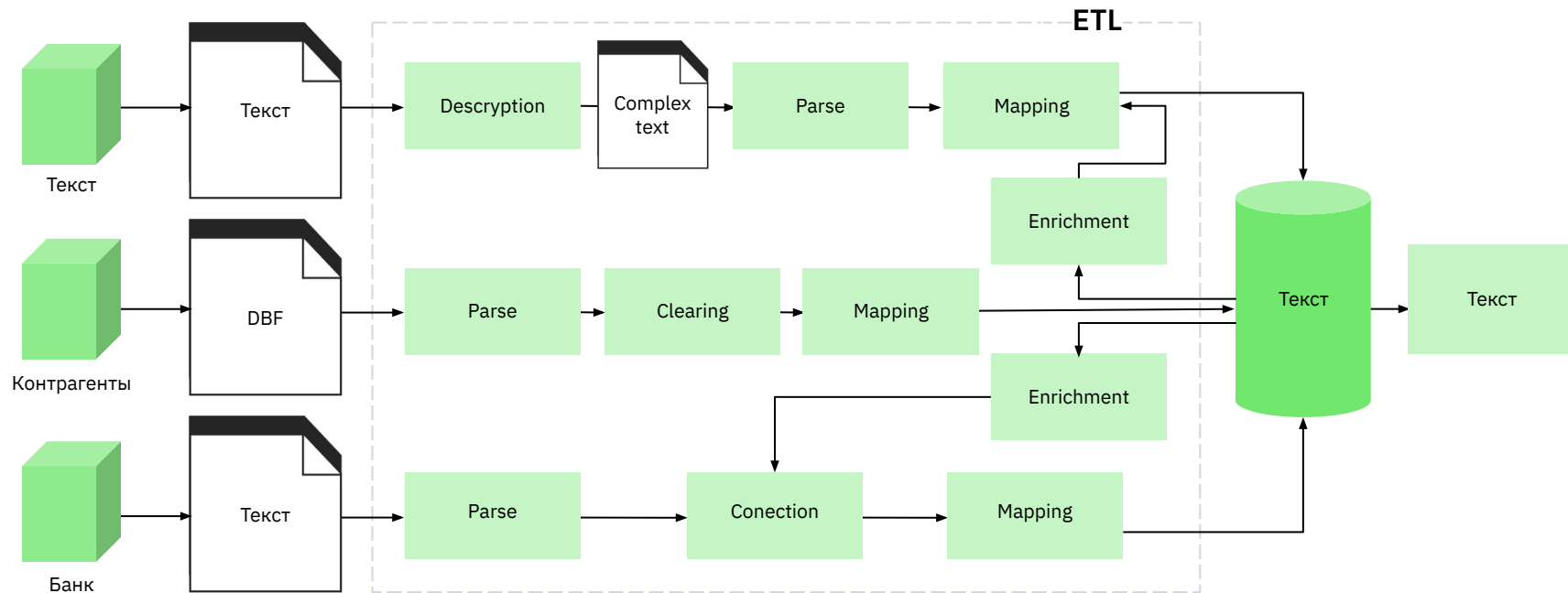


История первая. Приём на работу



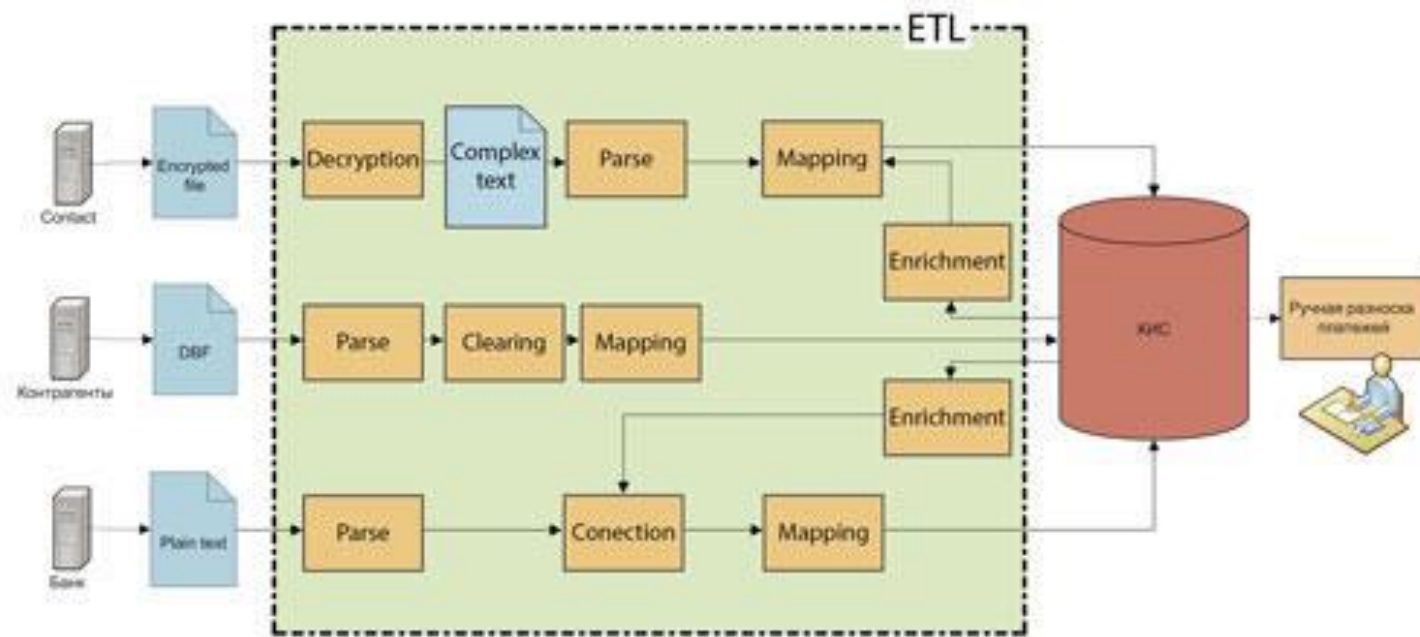


История вторая. Разноска платежей





История вторая. Разноска платежей





Спасибо за внимание

