

1. Введение

1.1. Цель документа

Цель данного документа — сформировать исчерпывающую спецификацию требований к ETL Flow Manager — графическому приложению для управления ETL-процессами (Extract, Transform, Load).

1.2. Область применения

Документ предназначен для:

- Команды разработки (проектирование, разработка, тестирование)
- Аналитиков данных и администраторов БД (конечные пользователи)
- Заинтересованных сторон (оценка функциональности)

1.3. Термины и определения

Термин	Определение
ETL	Процесс извлечения, трансформации и загрузки данных
SCD2	Медленно меняющиеся измерения типа 2
GUI	Графический интерфейс пользователя

1.4. Ссылки

- ISO/IEC/IEEE 29148:2018 (требования к инженерии требований)
- Документация SQLAlchemy, PyQt5, pandas

2. Общее описание

2.1. Представление продукта

ETL Flow Manager позволяет:

- Подключаться к PostgreSQL/MySQL
- Загружать данные из CSV и таблиц БД
- Очищать и преобразовывать данные
- Создавать таблицы с ограничениями (PK, FK)
- Выполнять SQL-запросы

2.2. Пользовательские роли

Роль	Функции
Аналитик	Запуск ETL, работа с запросами данных
Администратор БД	Настройка подключений, управление таблицами

2.3. Ограничения

- Поддержка только PostgreSQL и MySQL
- Максимальный размер CSV: 100 МБ

3. Функциональные требования

3.1. Управление подключениями

ID	Требование	Критерии приемки
1	Подключение к БД	Статус "Connected successfully"
2	Обработка неверных учетных данных	Сообщение об ошибке

3.2. Работа с данными

ID	Требование	Критерии приемки
3	Загрузка CSV	Данные в таблице предпросмотра
4	Удаление дубликатов	Количество строк уменьшается

3.3. Управление таблицами

ID	Требование	Критерии приемки
5	Создание таблицы с РК/ФК	Таблица отображается в списке
6	Добавление колонки	Структура таблицы обновляется

Функциональные требования

3.1. Управление соединениями с базами данных

- Система должна поддерживать подключение к различным СУБД (PostgreSQL, MySQL).
- Пользователь должен иметь возможность сохранять и загружать параметры подключения (хост, порт, имя базы данных, логин, пароль).
- Система должна проверять корректность подключения и отображать статус соединения.

3.2. Загрузка и предварительный просмотр данных

- Загрузка данных из CSV
 - Пользователь должен иметь возможность выбрать CSV-файл через интерфейс.
 - Система должна отображать предварительный просмотр данных (первые 10 строк).
- Загрузка данных из таблицы базы данных
 - Пользователь должен иметь возможность выбрать таблицу из подключенной базы данных.
 - Система должна отображать предварительный просмотр данных.

3.4. Очистка данных

- Система должна предоставлять инструменты для удаления дубликатов.
- Пользователь должен видеть изменения в данных после очистки.

3.5. Создание и управление таблицами

- Создание таблиц
 - Пользователь должен иметь возможность задать имя таблицы, её колонки (имя, тип данных, первичный ключ, nullable).
 - Система должна поддерживать добавление ограничений (внешние ключи).

- Управление существующими таблицами
 - Пользователь должен иметь возможность просматривать структуру таблиц (колонки, ключи).
 - Система должна поддерживать операции: добавление колонки, очистка таблицы, удаление таблицы.

3.6. Маппинг полей и ETL-процессы

- Пользователь должен иметь возможность сопоставлять поля источника и целевой таблицы.
- Система должна поддерживать настройку SCD2 (медленно меняющихся измерений).
- Пользователь должен иметь возможность запускать ETL-процесс и видеть его статус.

3.7. Выполнение SQL-запросов

- Пользователь должен иметь возможность вводить и выполнять SQL-запросы.
- Система должна отображать результаты запросов в табличном виде.
- Пользователь должен иметь возможность экспортировать результаты в CSV.

Нефункциональные требования

1. Требования к производительности

- Время отклика:
 - Основные операции (подключение к БД, загрузка данных, выполнение ETL) должны выполняться не более чем за 5 секунд при нормальной загрузке системы.
- Одновременная нагрузка:

- Система должна поддерживать работу 50+ одновременно авторизованных пользователей без существенной деградации производительности.

2. Требования к безопасности

- Шифрование:
 - Передача учётных данных и данных подключения к БД должна осуществляться по HTTPS.
- Разграничение прав:
 - Администраторы должны иметь полный доступ к функционалу.
 - Обычные пользователи должны иметь ограниченный доступ (например, запрет на удаление таблиц).
- Защита от несанкционированного доступа:
 - Механизм блокировки аккаунта после 5 неудачных попыток входа.

3. Требования к интерфейсу

- Интерфейс должен быть интуитивно понятным и адаптивным (подстраиваться под размер окна).
- Система должна поддерживать тёмную и светлую тему (опционально).

4. Требования к совместимости

- Система должна работать на ОС Windows, Linux и macOS.
- Поддерживаемые версии Python: 3.8 и выше.

5. Требования к документации

- Система должна включать руководство пользователя и техническую документацию для разработчиков.

4. Внешние интерфейсы

4.1. Интерфейсы пользователя

Графический интерфейс (GUI) реализован на базе PyQt5 и включает:

- Вкладки управления:
 - Подключение к БД (параметры: хост, порт, логин, пароль)
 - Загрузка данных (CSV/БД) с предпросмотром
 - Очистка данных (удаление дубликатов)
 - Создание и редактирование таблиц (с неограниченным числом столбцов)
 - Настройка ETL-процессов (маппинг полей, SCD2)
 - SQL-запросы с экспортом в CSV
- Элементы управления:
 - Таблицы с сортировкой и пагинацией
 - Выпадающие списки для выбора типов данных
 - Кнопки действий ("Загрузить", "Очистить", "Выполнить")

4.2. Интерфейсы программного обеспечения

- Интеграция с СУБД:
 - PostgreSQL: через `psycopg2`
 - MySQL: через `mysql-connector-python`
 - SQLAlchemy ORM для абстракции запросов
- Обработка данных:
 - Pandas для трансформации (удаление дубликатов, маппинг)
 - CSV-парсер с поддержкой кодировок (UTF-8, Windows-1251)

4.3. Интерфейсы оборудования

- Серверные требования:
 - Минимум: 4 ГБ ОЗУ, 2 ядра CPU, 50 ГБ HDD
 - Рекомендуется: 8 ГБ ОЗУ, 4 ядра CPU, SSD-диск
 - Поддержка многопоточной обработки данных
- Клиентские требования:
 - Разрешение экрана: 1280×720+
 - Поддержка ОС: Windows 10+, Linux, macOS

4.4. Интерфейсы связи

- Протоколы:
 - HTTPS (TLS 1.2+) для всех соединений с БД
 - SSH для удаленного администрирования
- Защита данных:
 - Шифрование учетных данных (AES-256)
 - Ограничение IP-адресов для доступа к БД (firewall)
- Формат обмена:
 - JSON для конфигураций ETL-процессов
 - CSV/Excel для экспорта результатов

5. Требования к системе

5.1 Аппаратные требования

Для работы сайта требуется сервер с минимальными характеристиками: 4 ГБ оперативной памяти, 2 процессора и достаточное дисковое пространство для хранения контента.

5.2 Программные требования

Система должна быть развернута на сервере с установленным Python 3.8+, а также необходимыми библиотеками и зависимостями.

5.3 Операционные системы и платформы

Программное обеспечение может быть запущено на Windows 10+, Ubuntu 20.04+, macOS 12+.

6. Критерии приемки

6.1 Критерии функциональности

Приемка будет основана на том, что все указанные функции работают корректно и без сбоев в ходе тестирования.

6.2 Критерии производительности

Доступность сайта должна быть проверена в условиях нагрузки.

6.3 Критерии безопасности

Проведение тестирования на уязвимости сайта и проверка на наличие необходимых защитных мер.

6.4 Критерии удобства использования

Полное тестирование интерфейса с участием пользователей для оценки удобства взаимодействия.

7. Составляющие SRS

7.1 Указание на предыдущие и связанные документы

Будут предоставлены ссылки на документы, описывающие технологические решения и предыдущие версии сайта, если таковые имелись.

7.2 Понятия и ссылки

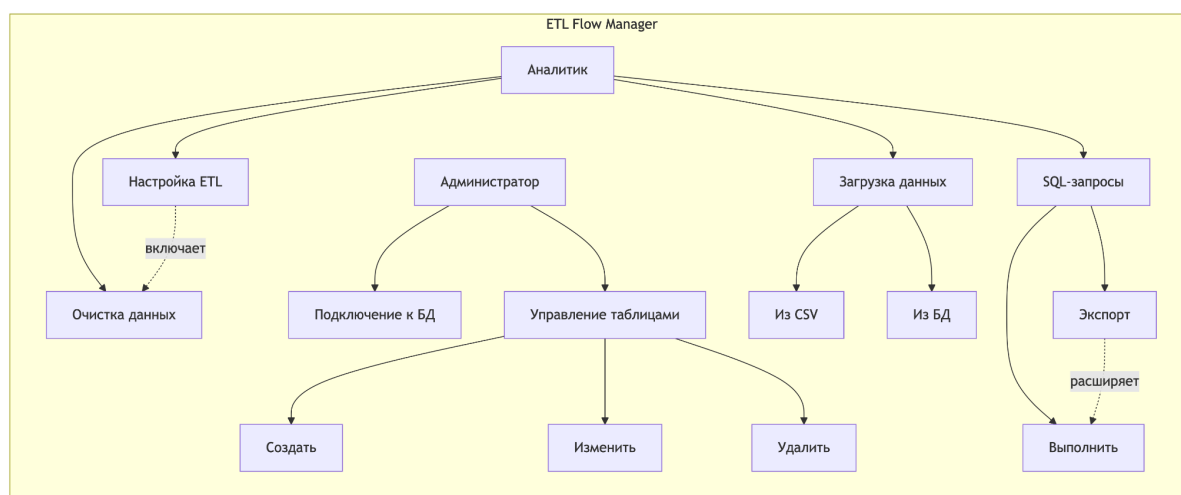
Включение дополнительной информации о используемых технологиях и методах разработки.

7.3 Согласования и утверждения

Установление процесса для проверки и утверждения документа всеми задействованными сторонами.

8. Приложения

Рисунок 1. Диаграмма прецедентов



8.1 Глоссарий

Объяснение специализированных терминов и акронимов, использованных в документе для улучшения понимания.

8.2 .Дополнительные диаграммы

Включение схем и диаграмм для представления архитектуры системы и взаимодействий между компонентами (если требуется).

8.3 Ссылки на стандарты и нормативные документы

Ссылки на документацию, стандарты и лучшие практики, которые были использованы при разработке данного сайта.

Заключение

Документ определяет требования к ETL Flow Manager для корректной реализации. Изменения согласовываются через систему управления версиями.