



**ОПИСАНИЕ ФУНКЦИОНАЛЬНЫХ ХАРАКТЕРИСТИК**  
Модус: Управление корпоративным хранилищем данных (ETL)  
Modus ETL

## Оглавление

Modus ETL.....	3
Список функций Программного продукта .....	3
Общие характеристики.....	3
Функции работы со следующими источниками данных: .....	3
Функции сбора данных:.....	3
Функции трансформации (преобразование) данных.....	4
Функции автоматизации ETL-процессов .....	4
Функции управление хранилищем (DWHM) .....	4
Функции предоставление данных для использования .....	5
Функции разграничения прав доступа.....	5

## Modus ETL

Программа для ЭВМ «Модус: Управление корпоративным хранилищем данных (ETL)», сокращенное название - Modus ETL (далее – «Программный продукт») предназначена для:

- автоматизированного сбора данных из различных источников;
- подготовки данных для аналитики, управления качеством данных;
- управления хранилищем данных.

\*) ETL - от английского Extract, Transform, Load - «извлечение, преобразование, загрузка».

## Список функций Программного продукта

### Общие характеристики

Работа Программного продукта под управлением ОС Windows или Linux.  
Управление хранилищами данных на СУБД MS SQL, PostgreSQL, Vertica.

Возможность установки как внутри закрытой IT-инфраструктуры заказчика так и в облачной инфраструктуре.

Возможность как прямого подключения к базам данных источников так и использования веб-сервисов/API для интеграции.

Открытая архитектура в части доработки новых функций для подключения к источникам, сбора, консолидации, трансформации данных, управления хранилищем и т.п.  
Возможность настраивать ETL-операции без навыков программирования.

### Функции работы со следующими источниками данных

- учетных систем на платформе 1С, 1СFresh;
- баз данных: MS SQL, PostgreSQL, Oracle, Vertica, MySQL и других, используя ODBC;
- веб-сервисов;
- файлов Excel;
- текстовых файлов с разделителями.

### Функции сбора данных

Загрузка данных с запуском по расписанию, в много потоков, с управление очередями, подробным логированием и возможностью автоматической обработки ошибочных ситуаций.

Возможность получение данных с помощью опционального модуля «Агент ETL», - разработанного на языке Golang для производительного многопоточного получение данных. Для горизонтальной масштабируемости ETL-процессов возможно использовать несколько Агентов ETL.

Для получения данных из учетных систем на Платформе 1С используются:

- специально разработанный коннектор «Адаптер ETL-1С» - http-сервис, встраиваемый в источник;
- нативные способы получения данных 1С/СКД-запросами, в том числе подключение к схеме компоновки данных (СКД) существующих отчетов

Для баз данных на СУБД MS SQL, PostgreSQL, Oracle:

- использование запросов на языке СУБД, из которой получают данные.

Для настройки загрузки из файлов:

- использование визуального мастера;
- для “Агента ETL” использование SQL-подобного языка запросов.

Визуальный конструктор запросов для получения данных из 1С-источников и баз данных MS SQL, PostgreSQL, Oracle, Vertica и т.д..

Функция загрузки описания метаданных из структуры данных в источниках с возможностью последующего использования метаданных для составления запросов в конструкторе или вручную.

Функции обработки данных до записи в хранилище:

- Встроенная функция замены символов в данных до записи в хранилище.
- Встроенная функция обезличивания данных до записи в хранилище с применением настраиваемых хэш-функций.
- дополнительная обработка получаемых данных на языке 1С (например, вызов функций/сервисов геокодирования для получения геокоординат по адресу).

### **Функции трансформации (обработки/преобразования) данных**

Обработка данных с использованием многошаговых сценариев с возможностью параллельного выполнения шагов.

Интерфейс “Workflow” для визуального проектирования ETL с возможностью

- настраивать шаги сценария обработки данных;
- связывать шаги обработки друг с другом, настраивая передачу данных/ скриптов для вложенных запросов/ потока управления между шагами;
- запускать сценарий на выполнение;
- логировать и отображать ход, статистику и ошибки выполнения шагов сценария,
- настраивать, сохранять и отображать статистику по данным для входа/выхода шага обработки;

Для настройки шагов сценария возможно:

- использовать готовые мастера (шаблоны) типовых операций трансформации данных: выбора, группировки, объединения, дополнения, категоризации, стандартизации, очистки и пр.; расчета показателей/KPI;
- писать произвольный код на языках SQL или 1С.

Расширенная аналитика реализована в дополнительном модуле DataMining за счёт интеграции с библиотеками Python.

### **Функции автоматизации ETL-процессов**

- Запуск ETL процессов по расписанию;
- Оповещение о статусе выполнения на электронную почту по списку рассылки;
- Логирование ETL операций;
- Интерфейсы для контроля выполнения ETL-процессов.

### **Функции управление хранилищем (DWHM)**

Используемые системы управления базами данных:

- MS SQL

- PostgreSQL
- Vertica

Операции по управлению хранилищем:

- Трансформация первичных данных в схему "звезда" с генерацией структуры объектов;
- Визуализация структуры хранилища: ER-диаграммы моделей данных на основе загруженных метаданных хранилища;
- Создание витрин данных в визуальном интерфейсе;
- Версионирование данных.

#### **Функции управления НСИ для стандартизации и категоризации данных:**

- Настройка структур для хранения нормативно-справочной информации: первичных (полученных из учетных систем) эталонных справочников; правил сопоставления первичных и эталонных справочников (маппингов).
- Мастера (шаблоны) для замены или дополнения данных с учетом справочников и маппингов.

#### **Функции управления качеством данных (DQM):**

- Настройка правил верификации и очистки данных;
- Шаблоны правил проверки и очистки для "доменов данных" (ИНН, телефонов, СНИЛС, адресов и т.д.);

#### **Функции предоставление данных для использования**

- Встроенный коннектор к OLAP-серверу;
- Выгрузка данных из хранилища в форматы файлов Excel;
- Выгрузка данных из хранилища в текстовый файл с разделителями.

#### **Функции разграничения прав доступа**

- Разграничение прав доступа по функционалу – возможность предоставления доступа для настройки ETL-процессов; только запуска ETL-процессов без возможности настройки и изменения.
- Организация ETL-процессов для нескольких проектов с разграничением прав доступа к ним;
- Управление несколькими хранилищами данных с разграничением прав доступа к ним;