



Modus ETL для управления данными



О КОМПАНИИ

БиАй Про - российский разработчик аналитических решений для корпоративного сектора и государственных органов власти.

Мы специализируемся на разработке и реализации проектов в области BI и Big Data.



Сбор, стандартизация и нормализация данных



Разработка аналитических витрин и визуальных отчетов



Проектирование архитектуры корпоративного хранилища данных

20+

Enterprise проектов
внедрения

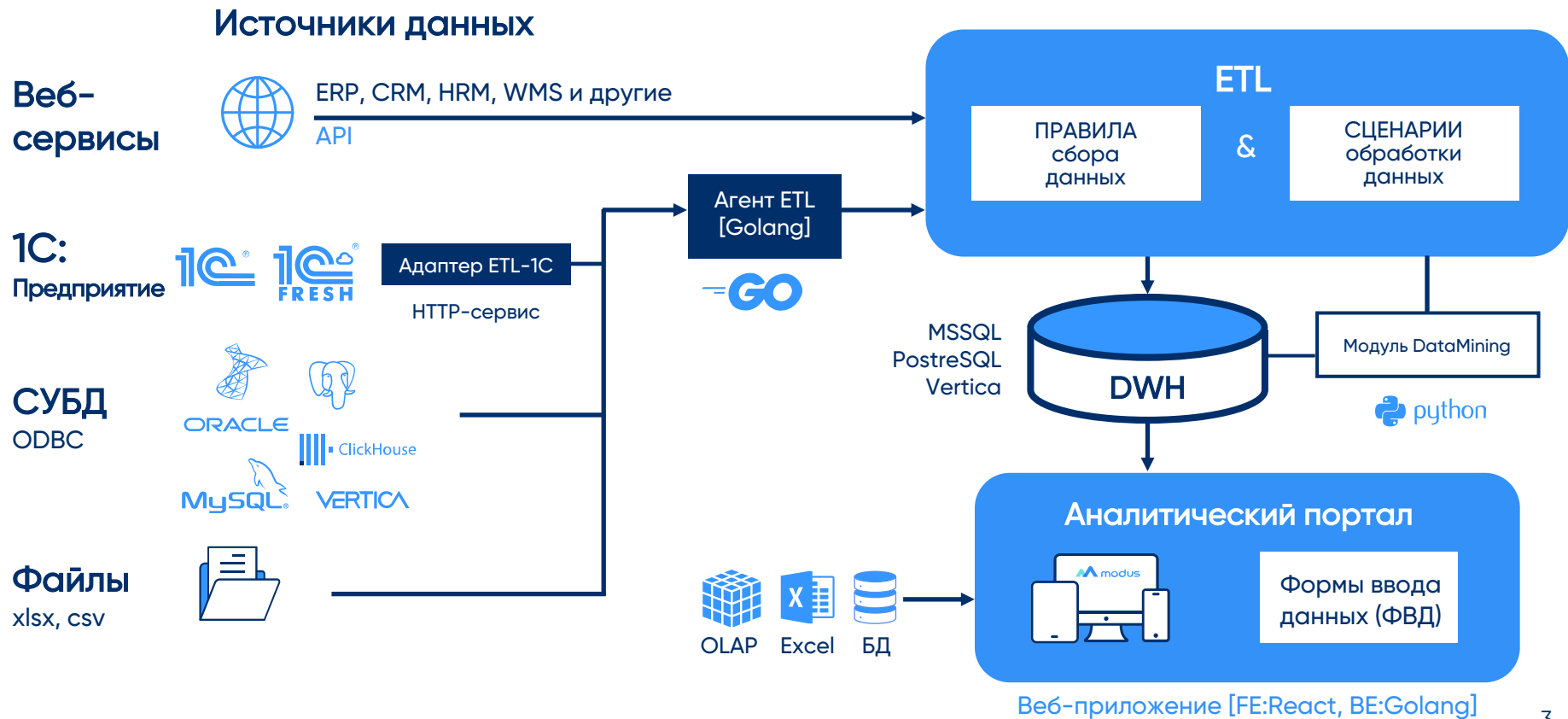
40+

компаний-партнеров

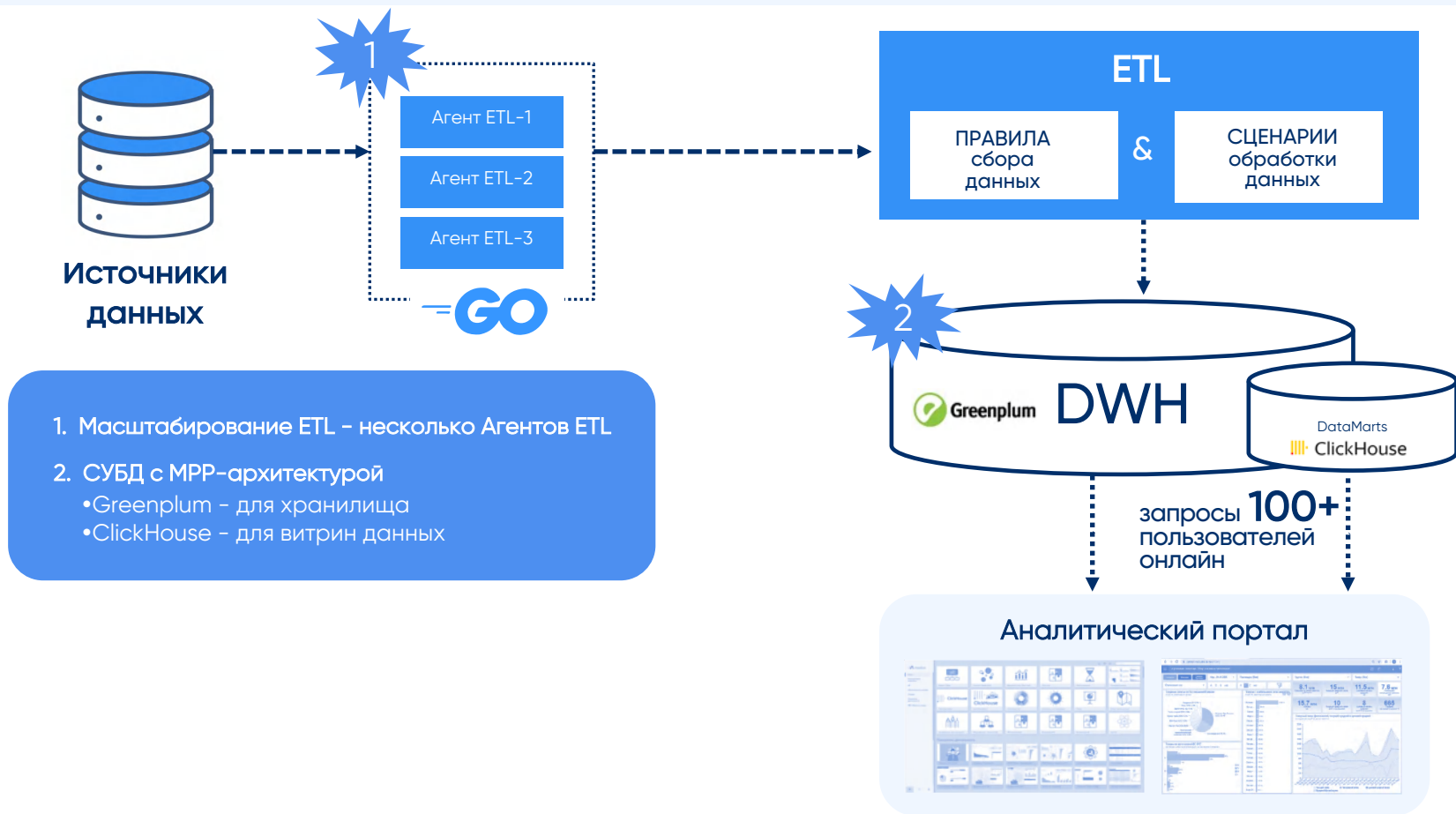
10

регионов внедрения
аналитической системы
для органов власти

ETL и Аналитическая система



Modus ETL и хранилище данных для крупных предприятий



User friendly ETL

Дружественные пользователю
интерфейс и инструменты

- ✓ Мастера и шаблоны ETL-процессов по очистке, обогащению, верификации данных
- ✓ Диаграммы для визуализации ETL-пакетов получения и обработки данных
- ✓ Контроль процессов с простым и понятным интерфейсом
- ✓ ER-диаграммы для настройки и визуализации моделей данных



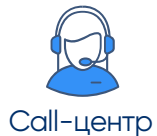


Высокая производительность

Многопоточный сбор данных
с управлением очередями

**Открытый для изменения код
с возможностью самостоятельной
доработки продукта**

Управление хранилищем данных



Хранилище данных

Метаданные



Стандартизация

Эталонный справочник

Первичный справочник
[из учетной системы]



Пример:
Гл.бухгалтер,
Главбух,
главный Бухгалтер
Главный бух-тер



Главный
бухгалтер

Категоризация

Справочник
первичный/эталонный



маппинг



Категория -1

Категория -2

Примеры сущностей:

Начисления ~ Группа начислений

Должности ~ Категория должностей

Учреждения ~ ОИВ

Товары ~ Номенклатурная группа

ETL для управления НСИ

Функции и инструменты

Функции

1

НСИ. Хранение
Универсальная структура в БД

2

НСИ. Управление
справочниками и маппингами

3

НСИ. Использование
для стандартизации и категоризации данных

Инструменты



Результат

Таблицы
в хранилище
данных для НСИ

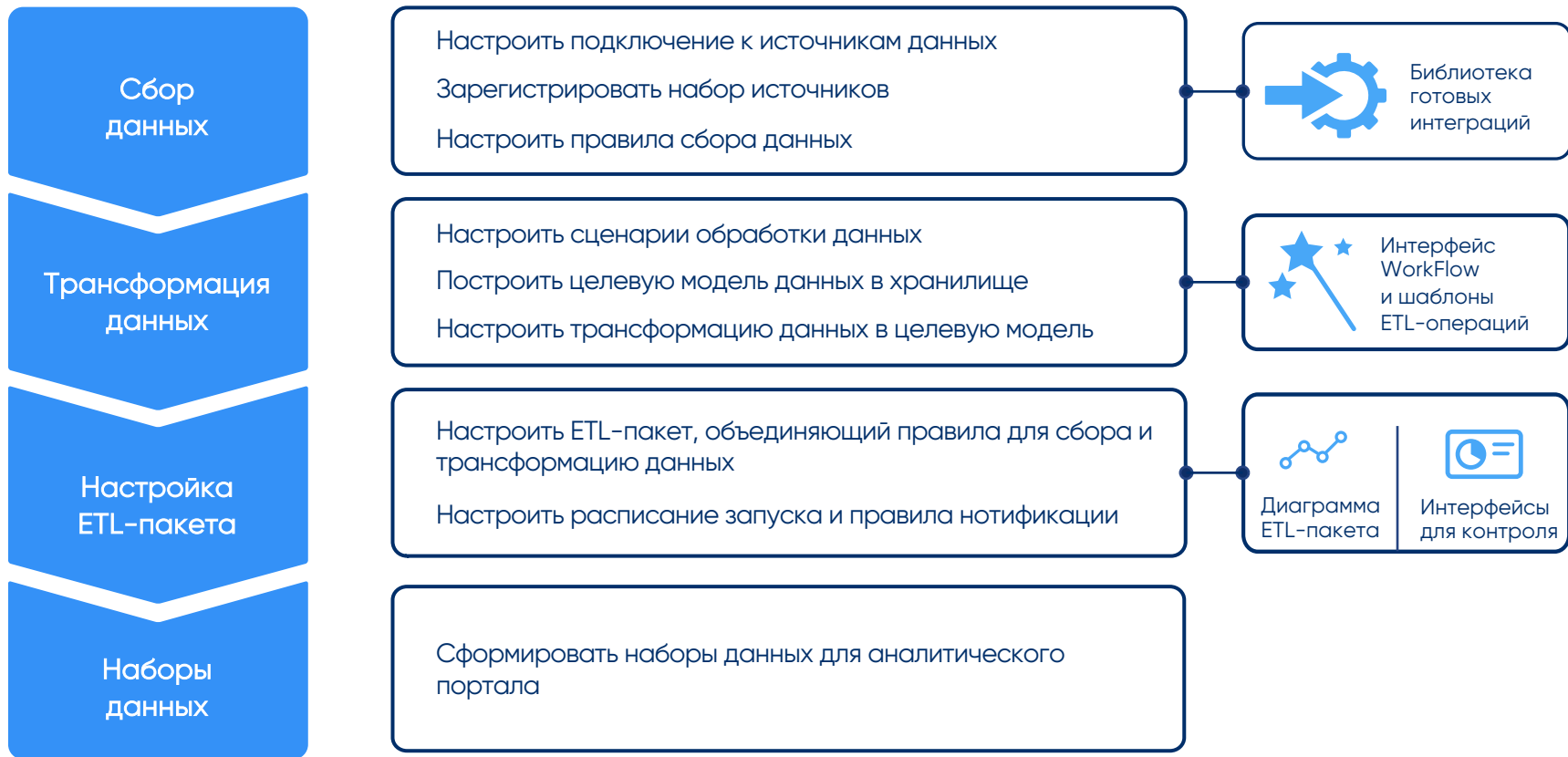
**Справочники и
правила** привязки
сущностей к НСИ

SQL-скрипты для
обработки
данных



РЕЗУЛЬТАТ: стандартизированные и обогащенные данные для аналитики

Последовательность настройки ETL-процессов



1



Сбор данных

Многopоточная загрузка данных с управлением очередями.

Для обеспечения производительной загрузки используется свой "Агент на Go" (веб-сервис).

Для учетных систем на платформе 1С-возможность получать данные, используя подключение к 1С-источнику с помощью "Адаптера ETL-1С" (http-сервис) и нативные способы:

- ✓ запросы на языке 1С
- ✓ подключение и управление схемой компоновки данных СКД-отчетов в источнике

Возможность использовать скрипты для пост-обработки получаемых данных, в т.ч. для обезличивания персональных данных, геокодирования (получение географических координат по адресу) и т.п.

2



Трансформация данных

Трансформация данных предполагает объединение, очистку, дополнение, замену, группировку, форматирование и т.п. данных.

Конструктор WorkFlow для визуального проектирования ETL.

Возможность при трансформации данных использовать многошаговые сценарии, содержащие:

- ✓ шаблоны (мастера) типовых операций трансформации данных (выбора, группировки, маппинга и пр.)
- ✓ продвинутую аналитику (библиотеки Python)
- ✓ SQL-скрипты
- ✓ обработку данных на языке 1С

3 Модуль НСИ – для стандартизации и категоризации данных

- ✓ настройка и хранение справочников (первичных и эталонных)
- ✓ настройка правил сопоставления (маппинга и замены) данных
- ✓ настройка правил проверки и очистки данных

4



Управления хранилищем данных (DWHM)

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
- ✓ Настройка версионирования данных

5

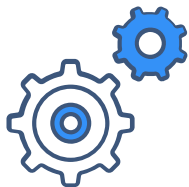


Автоматизация ETL-процессов

Диаграмма для визуализации и настройки ETL-пакета.

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

6

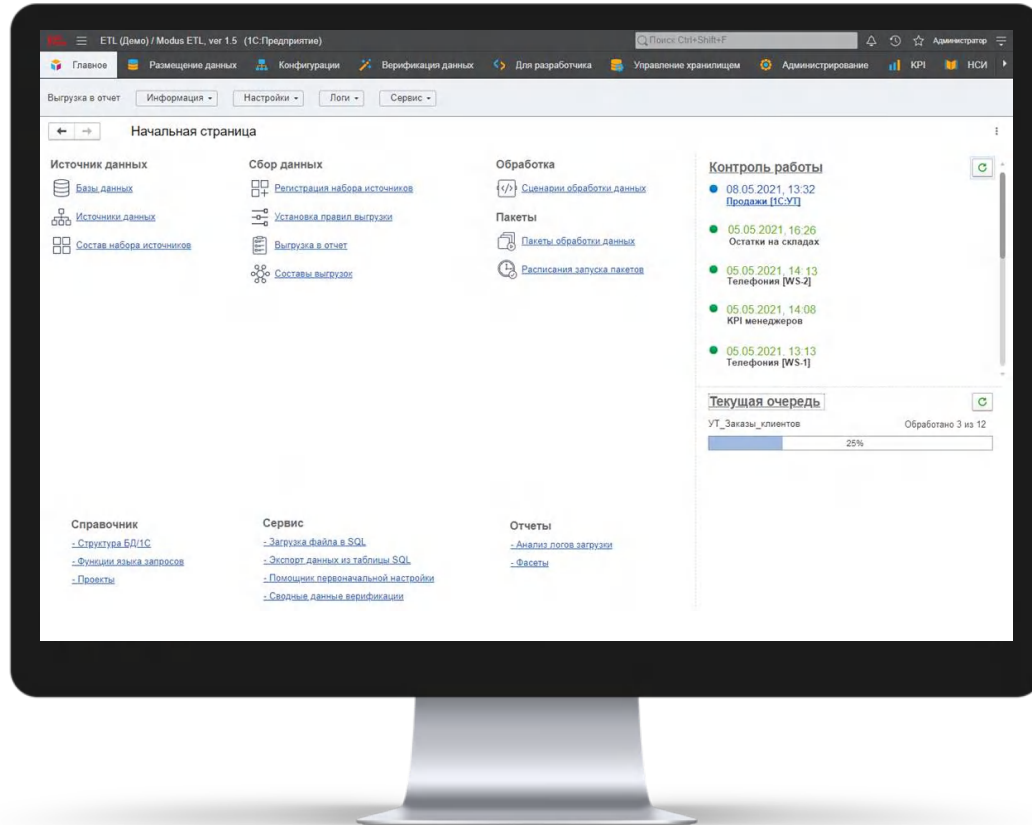


Автоматизация ETL-процессов

- ✓ Загрузка данных из файлов `xlsx`, `csv`
- ✓ Выгрузка данных из БД в файлы `xlsx`, `csv`
- ✓ Интеграция с OLAP-кубами "Полиматика"
- ✓ Загрузка метаданных о структуре данных в источниках (в т.ч. для использования в конструкторе запросов)
- ✓ Конструктор запросов
- ✓ Планировщик запуска ETL-пакетов и т.д.

Примеры интерфейсов

Интерфейс [Начальная страница]



Интерфейсы для контроля работы ETL-пакета

Дата-время старта и Статус последних завершенных процессов

Переход в подробные Логи

Прогресс-бары по очереди обработки источников

Переход в интерфейсы по контролю и управлению очередями

Контроль работы

- 08.05.2021, 13:32
Продажи [1С:УП]
- 05.05.2021, 16:26
Остатки на складах
- 05.05.2021, 14:13
Телефония [WS-2]
- 05.05.2021, 14:08
KPI менеджеров
- 05.05.2021, 13:13
Телефония [WS-1]

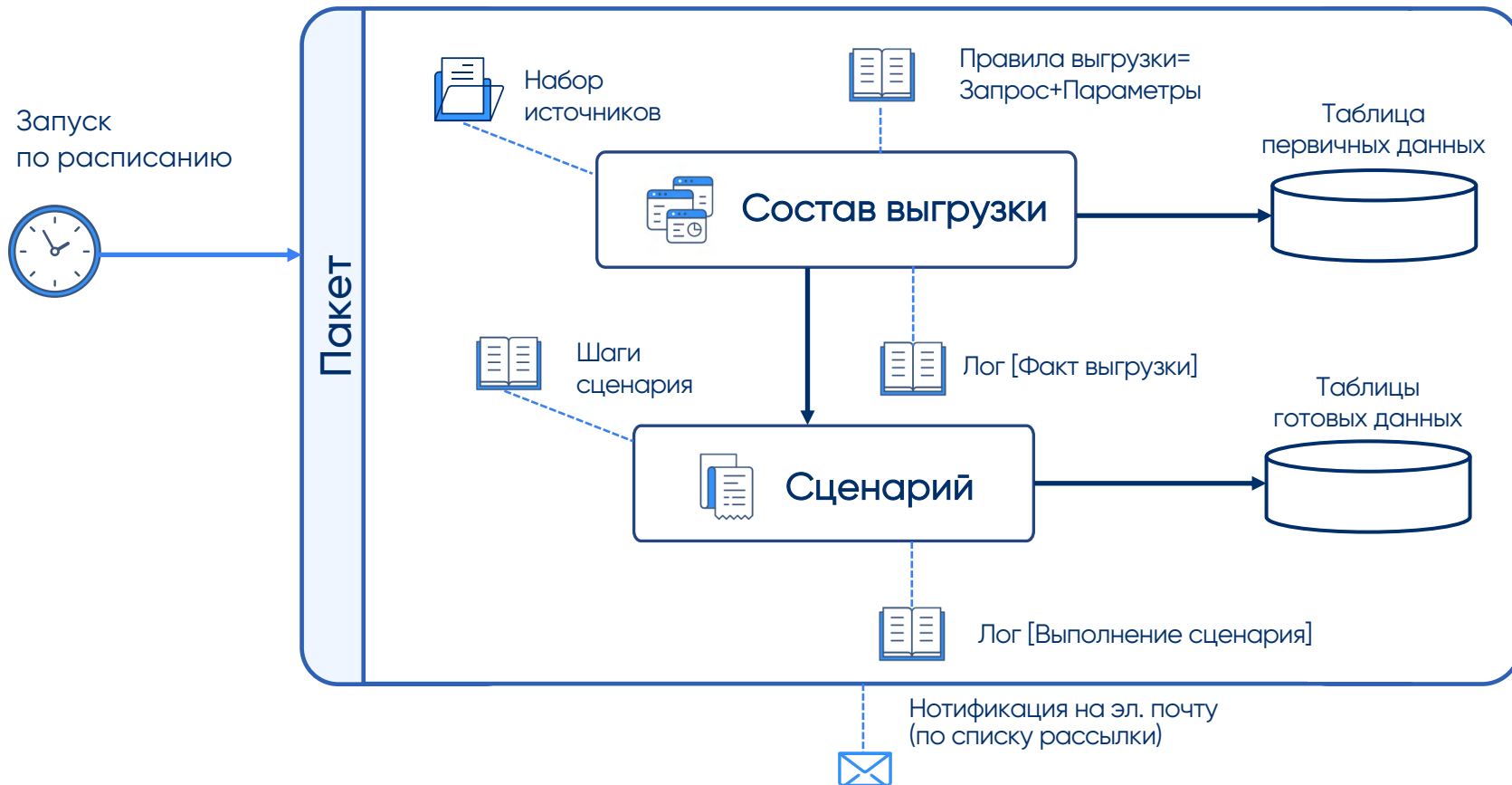
Текущая очередь

УТ_Заказы_клиентов

Обработано 3 из 12



Диаграмма пакета. Интерфейс для визуализации ETL-процессов



WorkFlow: Настройка шагов в мастере

Шаблоны (мастера)

1С:ЕТД (Демо) / Модуль: Управление корпоративным хранилищем данных (ЕТД), редакция 1.4 1С:Предприятие

Главное | Размещение данных | Конфигурации | Верификация данных | Для разработчика | Управление хранилищем | Администрирование | КРП | НСИ | Пресеты

Выгрузка в отчет | Информация | Настройки | Логи | Сервис

ИНФО | ШАБЛОНЫ | СЦЕНАРИЙ

- Произвольный код
 - Произвольный код 1С
 - Произвольный код SQL
- Примитивы обработки данных
 - Произвольная выборка данных
 - Группировка данных
 - Матчинг полей 2х таблиц
 - Оператор Uprivot
 - Временная таблица
 - Создание таблицы БД
- НСИ – дополнение данных
 - НСИ.Дополнение из эталонного справочника
- Рунимп
 - Кластеризация
 - Линейная регрессия
- Сервисные операции
 - Настройка замены пустых значений
 - Выгрузка таблицы в файл
 - Настройка индексов таблицы
 - Настройка удаления пробелов/специмволов из полей таблицы
 - Мастер настройки расчета факетов таблицы
 - Очистка хранилища от бэкапов
 - Кодирование бинарного файла в BASE64
 - Запуск принудительного обновления куба
- Сложный сценарий
 - Вставка результата запроса в таблицу
 - Смена имен таблиц
 - Создание временной таблицы

СОХРАНИТЬ | ЗАПУСТИТЬ | АВТОРАСТАНОВКА

Исходная таблица (шаблон Выборки) → Создание промежуточной таблицы → Заполнение промежуточной таблицы → Транспонирование таблицы

Произвольная выборка данных | Произвольный код | Матчинг полей 2х таблиц | Оператор Uprivot

Шаги сценария настройки обработки данных и связей

Мастер настройки

Мастер настройки выборки данных

Выбор источника | Настройка полей запроса | Связки таблиц | Настройка выбора данных

Источники: Таблица БД DEMO_ЗКСИ_Начисления_VF_инт_2

Имя	Позиция	Поле/Выражение	Таблица
TIMESTAMP	A	TIMESTAMP	DEMO_ЗКСИ_Начисления_VF_инт_2
BADA	A	BADA	DEMO_ЗКСИ_Начисления_VF_инт_2
OBJ	A	OBJ	DEMO_ЗКСИ_Начисления_VF_инт_2
ATTRIBUTE	A	ATTRIBUTE	DEMO_ЗКСИ_Начисления_VF_инт_2
Organization	A	Organization	DEMO_ЗКСИ_Начисления_VF_инт_2
File	A	File	DEMO_ЗКСИ_Начисления_VF_инт_2
Трансформированность	A	Трансформированность	DEMO_ЗКСИ_Начисления_VF_инт_2
Начисленоеудержание	A	Начисленоеудержание	DEMO_ЗКСИ_Начисления_VF_инт_2
y016	A	y016	DEMO_ЗКСИ_Начисления_VF_инт_2
y017	A	y017	DEMO_ЗКСИ_Начисления_VF_инт_2
y018	A	y018	DEMO_ЗКСИ_Начисления_VF_инт_2
контрпос	A	контрпос	DEMO_ЗКСИ_Начисления_VF_инт_2

Данные таблицы DEMO_ЗКСИ_Начисления_VF_инт_2

БАЗА	COL	ATTRIBUTE	TIMESTAMP	Organization	File	Трансформированность	Начисленоеудержание	y016	y017	y018	контрпос
DEMO_ЗКСИ	Начисление	0	2020-05-28T14:49:43Z	Организация	Файл:Путь 0000000016	Дополнительно	6991602704140072830	Группы за час. Показатели	Группы за час. Показатели	Группы за час. Показатели	Группы за час. Показатели
ZKSU_demo	0	ATTRIBUTE	2020-05-28T14:49:43Z	Организация	Файл:Путь 0000000016	Дополнительно	6991602704140072830	Дни неотработанные согласно работам	Дни неотработанные согласно работам	Дни неотработанные согласно работам	Дни неотработанные согласно работам
ZKSU_demo	0	ATTRIBUTE	2020-05-28T14:49:43Z	Организация	Файл:Путь 0000000016	Дополнительно	171464243184829475	Комплексная отгрузка (Отпуск сотрудников)	Комплексная отгрузка (Отпуск сотрудников)	Комплексная отгрузка (Отпуск сотрудников)	Комплексная отгрузка (Отпуск сотрудников)
ZKSU_demo	0	ATTRIBUTE	2020-05-28T14:49:43Z	Организация	Файл:Путь 0000000016	Дополнительно	348976754112928307788	Выплата за стажность и надбавочность	Выплата за стажность и надбавочность	Выплата за стажность и надбавочность	Выплата за стажность и надбавочность
ZKSU_demo	0	ATTRIBUTE	2020-05-28T14:49:43Z	Организация	Файл:Путь 0000000016	Дополнительно	348976754112928307788	Отпуск за отпуск	Отпуск за отпуск	Отпуск за отпуск	Отпуск за отпуск
ZKSU_demo	0	ATTRIBUTE	2020-05-28T14:49:43Z	Организация	Файл:Путь 0000000016	Дополнительно	6991602704140072830	Комплексная отгрузка (Отпуск сотрудников)	Комплексная отгрузка (Отпуск сотрудников)	Комплексная отгрузка (Отпуск сотрудников)	Комплексная отгрузка (Отпуск сотрудников)
ZKSU_demo	0	ATTRIBUTE	2020-05-28T14:49:43Z	Организация	Файл:Путь 0000000016	Дополнительно	6991602704140072830	Отпуск сотрудников	Отпуск сотрудников	Отпуск сотрудников	Отпуск сотрудников
ZKSU_demo	0	ATTRIBUTE	2020-05-28T14:49:43Z	Организация	Файл:Путь 0000000016	Дополнительно	7269722141929166461294	Выплата за стажность и надбавочность	Выплата за стажность и надбавочность	Выплата за стажность и надбавочность	Выплата за стажность и надбавочность

ЛОГ ВЫПОЛНЕНИЯ | ПРЕВЬЮ

Шаг	Наименование шага	%	Начало	Конец	Ошибка	Время, мс
1.1	Исходная таблица (шаблон Выборки)	100	13.05.2021 19:20:47	13.05.2021 19:20:53		4680
1.2	Создание промежуточной таблицы	100	13.05.2021 19:20:53	13.05.2021 19:20:54		749
1.3	Заполнение промежуточной таблицы	100	13.05.2021 19:20:54	13.05.2021 19:20:55		608
1.4	Транспонирование таблицы	100	13.05.2021 19:20:55	13.05.2021 19:20:56		598

Лог выполнения и превью данных

WorkFlow: Статистика и DQM

Статистика по полю (гистограмма)

Измеритель качества данных (DQM)

Статистика по полю (таблица)

Превью данных

The screenshot displays the WorkFlow software interface. At the top, the menu bar includes options like 'Главное', 'Размещение данных', 'Конфигурации', 'Верификация данных', 'Для разработки', 'Управление хранилищем', 'Администрирование', 'KPI', 'НСИ', and 'Пресеты'. Below the menu, there are buttons for 'Выгрузка в отчет', 'Информация', 'Настройки', 'Логин', and 'Сервис'. The main workspace is divided into several panels:

- Left Panel:** A histogram titled 'Статистика по полю experience' showing a distribution of values. Below it, a 'Качество данных' (Data Quality) section shows a progress bar and a legend for 'Не OK', 'NULL', 'Пусто', and 'OK'.
- Center Panel:** A workflow diagram showing 'Исходная таблица (шаблон Выборки)' leading to 'Создание промежуточной таблицы' and then 'Произвольный код'.
- Right Panel:** A configuration window for 'Исходная таблица (шаблон Выборки)'. It contains a table with columns: '№', 'Поле таблицы', 'Тип', 'Тип проверки', and 'Параметр'. The table lists three rows with different fields and their respective check types and parameters.
- Bottom Panel:** A 'ПРЕВЬЮ' (Preview) section showing a table of data execution logs with columns: 'LOG ВЫПОЛНЕНИЯ', 'ФАСЕТ', 'ПРЕВЬЮ', 'BAZA', 'OBL', 'ATTRIBUTE', 'TIMESTAMP', 'Организация', 'ФизЛицо', 'Текущаядолжность', 'Начисленнеудержание', and years from 2016 to 2018.

Настройка правил проверки Качества данных

WorkFlow: Параллельное выполнение шагов

ИНФО ШАБЛОНЫ СЦЕНАРИЙ

- Произвольный код
 - Произвольный код 1С
 - Произвольный код SQL
- Примитивы обработки данных
 - Произвольная выборка данных
 - Группировка данных
 - Маллинг полей 2х таблиц
 - Оператор Unpivot
 - Временная таблица
 - Создание таблицы БД
- НСИ – дополнение данных
 - НСИ Дополнение из эталонного справочника
- Python
 - Кластеризация
 - Линейная регрессия
- Сервисные операции
 - Настройка замены пустых значений
 - Выгрузка таблицы в файл
 - Настройка индексов таблицы
 - Настройка удаления пробелов/ спецсимволов из полей таблицы
 - Мастер настройки расчета фасетов таблицы
 - Очистка хранилища от бэкапов
 - Кодирование бинарного файла в BASE64
 - Запуск принудительного обновления куба
- Сложный сценарий
 - Вставка результата запроса в таблицу
 - Смена имен таблиц
 - Создание временной таблицы

СОХРАНИТЬ ▶ ЗАПУСТИТЬ ✎ АВТОРАССТАНОВКА

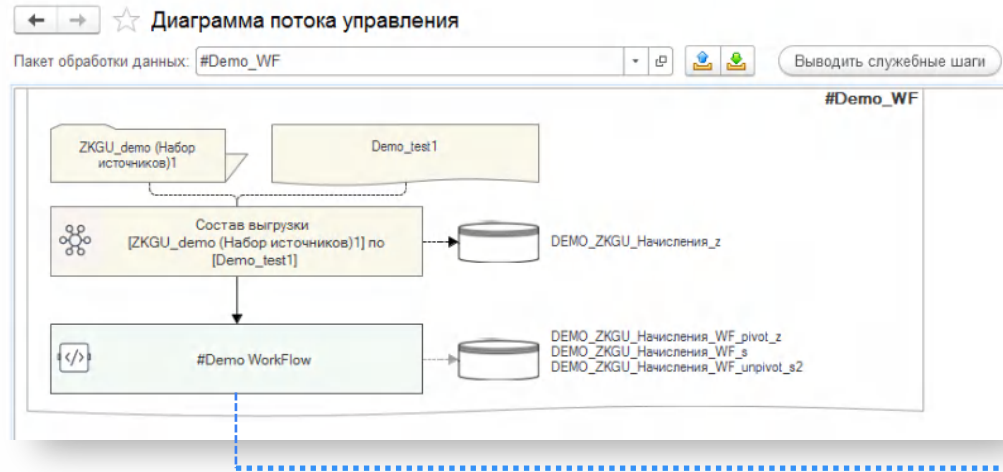
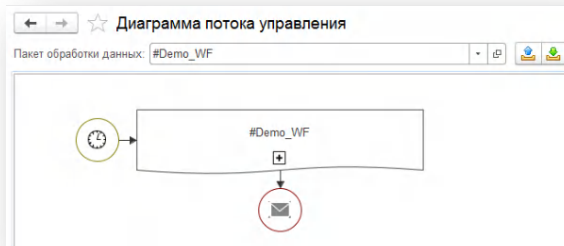
```

graph LR
    Start(( )) --> A[Выборка данных исходной таблицы]
    Start --> B[Произвольная выборка данных]
    A --> C[Промежуточная таблица]
    B --> D[Выборка данных по удержаниям ВЗ]
    C --> E[Запись данных по начислениям]
    D --> F[Произвольная выборка данных]
    E --> G[Маллинг полей 2х таблиц]
    F --> H[Оператор Unpivot]
    G --> I[Транспонирование]
    H --> J[Оператор Unpivot]
    I --> K[Транспонирование из ВЗ]
    J --> L[Сумма начислений по годам ВЗ]
    K --> M[Сумма удержаний по годам ВЗ]
    L --> N[Произвольный код]
    M --> O[Группировка данных]
    N --> P[Выгрузка данных]
    O --> P
    P --> Q[Выгрузка таблицы в файл]
    P --> R[Вывод данных]
    
```

ЛОГ ВЫПОЛНЕНИЯ

Шаг	Наименование шага	%	Начало	Конец	Ошибка	Время, мс
1.1	Выборка данных исходной таблицы	100	25.05.2021 16:23:59	25.05.2021 16:24:02		2678
1.2	Промежуточная таблица	100	25.05.2021 16:24:03	25.05.2021 16:24:03		402
1.3	Запись данных по начислениям	100	25.05.2021 16:24:03	25.05.2021 16:24:04		355
1.4	Выгрузка данных	100	25.05.2021 16:24:04	25.05.2021 16:24:08		3705
2.1	Транспонирование	100	25.05.2021 16:24:05	25.05.2021 16:24:06		805
2.2	Сумма начислений по годам ВЗ	100	25.05.2021 16:24:07	25.05.2021 16:24:07		1

Пакет в ETL

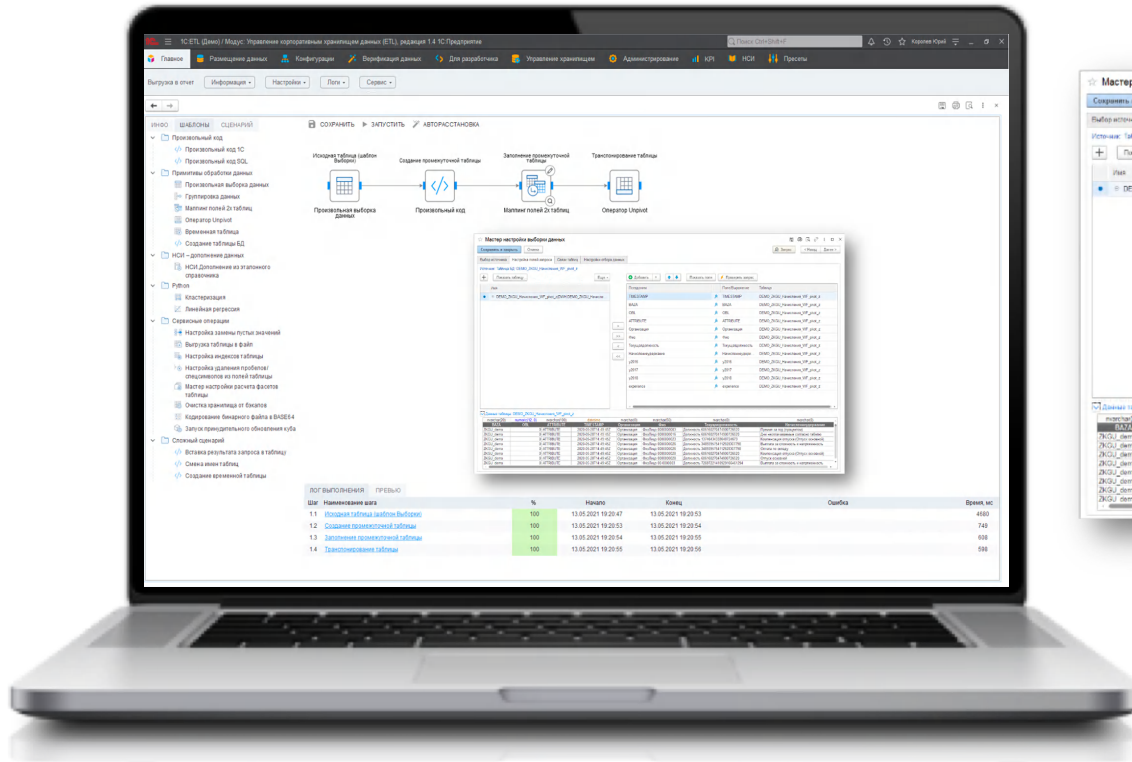


Интерфейс WorkFlow для сценария обработки данных

Screenshot of the WorkFlow interface for a data processing scenario. The interface shows a flow diagram with nodes for "Получение данных", "Создание таблицы", "Преобразование данных", and "Загрузка данных". Below the diagram is a table with columns for "№", "Имя", "Время", "Статус", and "Метрики".

№	Имя	Время	Статус	Метрики
11	Создание таблицы	11.08.2021 10:51:47	Успешно	4000
12	Преобразование данных	11.08.2021 10:52:14	Успешно	174
13	Загрузка данных	11.08.2021 10:52:45	Успешно	400
14	Создание таблицы	11.08.2021 10:53:10	Успешно	400

WorkFlow. Шаги сценария и редактирование шага



Мастер настройки выборки данных

Выбор источника: Таблица [d3: DEMO_ZKQI_Наименование_WF_дпт_2]

Имя: Таблица [d3: DEMO_ZKQI_Наименование_WF_дпт_2]

Имя	Тип	Полное наименование	Таблица
TIME_STAMP	А	TIME_STAMP	DEMO_ZKQI_Наименование_WF_дпт_2
WADA	А	WADA	DEMO_ZKQI_Наименование_WF_дпт_2
CVBL	А	CVBL	DEMO_ZKQI_Наименование_WF_дпт_2
ATTRIBUTE	А	ATTRIBUTE	DEMO_ZKQI_Наименование_WF_дпт_2
Организация	А	Организация	DEMO_ZKQI_Наименование_WF_дпт_2
Фис	А	Фис	DEMO_ZKQI_Наименование_WF_дпт_2
Товародолжность	А	Товародолжность	DEMO_ZKQI_Наименование_WF_дпт_2
Наименованиеуслуги	А	Наименованиеуслуги	DEMO_ZKQI_Наименование_WF_дпт_2
y2016	А	y2016	DEMO_ZKQI_Наименование_WF_дпт_2
y2017	А	y2017	DEMO_ZKQI_Наименование_WF_дпт_2
y2018	А	y2018	DEMO_ZKQI_Наименование_WF_дпт_2
свертываеца	А	свертываеца	DEMO_ZKQI_Наименование_WF_дпт_2

Имя	Тип	Полное наименование	Ссылка	Имя	Тип	Полное наименование	Ссылка			
ZKQI_дпт_2	Д	ATTRIBUTE	2020-05-28T14:49:42	Организация	Физ/Лег	030003003	Должность	605160270414301776203	Финансы	за код. Должность
ZKQI_дпт_2	Д	ATTRIBUTE	2020-05-28T14:49:42	Организация	Физ/Лег	030003014	Должность	605160270414301776202	Доп. информация	Идентификатор таблицы
ZKQI_дпт_2	Д	ATTRIBUTE	2020-05-28T14:49:42	Организация	Физ/Лег	030003023	Должность	13174441333448724878	Клиентские услуги	Отпуск (отпуска)
ZKQI_дпт_2	Д	ATTRIBUTE	2020-05-28T14:49:42	Организация	Физ/Лег	030003026	Должность	248397917413292637796	Выплата	за сложность и неадекватность
ZKQI_дпт_2	Д	ATTRIBUTE	2020-05-28T14:49:42	Организация	Физ/Лег	030003028	Должность	248397917413292637796	Смена	за аренду
ZKQI_дпт_2	Д	ATTRIBUTE	2020-05-28T14:49:42	Организация	Физ/Лег	030003028	Должность	605160270414301776202	Минусовая	отпуска (Отпуска отпусков)
ZKQI_дпт_2	Д	ATTRIBUTE	2020-05-28T14:49:42	Организация	Физ/Лег	030003033	Должность	605160270414301776203	Страна	отпусков
ZKQI_дпт_2	Д	ATTRIBUTE	2020-05-28T14:49:42	Сотрудники	Физ/Лег	034006003	Должность	7269721141821956461294	Бюджет	за сложность и неадекватность

Управляйте бизнесом на основе данных



modusbi.ru



post@modusbi.ru



+7 (495) 118-28-14