

Описание функциональных возможностей Polyflow

Polyflow

Система управления данными Polyflow (далее — Система) — это решение для управления корпоративными хранилищами данных (КХД).

Модуль представляет собой сервис оркестровки, сбора и обработки разнородных данных в хранилищах произвольной архитектуры. Polyflow позволяет управлять следующей функциональностью:

- сбор данных из разных источников:
 - файлы
 - внешние системы
 - базы данных
- описание сущностей-получателей и источников;
- мониторинг и управление выполняемыми процессами;
- контроль качества данных;
- трансформация данных.

Определения и сокращения Polyflow

Определения и сокращения Polyflow представлены в Таблица 1.

Таблица 1. Определения и сокращения Polyflow

Термин/Сокращение	Определение
Хранилище данных (англ. Content Repository, Data Warehouse, DWH)	Предметно-ориентированная информационная база данных, сочетающая в себе функции системы управления версиями, поисковой машины и СУБД
КХД	Корпоративное хранилище данных
Система управления данными Polyflow	Сервис оркестровки сбора и обработки разнородных данных хранилища произвольной архитектуры
Polyflow	Краткое наименование программного обеспечения «Система управления данными Polyflow»
API	«Application Programming Interface», интерфейс программирования приложений, программный интерфейс приложения
ККД	Контроль качества данных – оценка соответствия данных заданным критериям

Архитектура решения

Состав компонент

Система включает в себя следующие модули:

1. Планировщик/оркестратор
2. Модули расширения функционала
3. Модуль ETL
4. Модуль управления ККД
5. Модуль управления метаданными

Планировщик/оркестратор

Выполняет функции планирования, мониторинга и запуска процессов системы, синхронизации и ведения истории их выполнения.

Модули расширения функционала

Дополнительные модули, упрощающие взаимодействие с планировщиком и оркестратором, настройку процессов, а также взаимодействие компонентов системы.

Модуль ETL

Модуль, выполняющий функции извлечения, трансформации и загрузки данных из различных источников.

Модуль управления ККД

Модуль, решающий задачи контроля качества данных.

Модуль управления метаданными

Модуль, предоставляющий информацию о моделях для других компонентов системы.

Основные взаимодействия

На Рисунке 1 представлена схема решения и основные взаимодействия между его частями и внешними системами:

1. Запрос на скачивание файла из внешнего хранилища файлов;
2. Подготовка файла для скачивания;
3. Скачивание файла;
4. Инициирование извлечения данных из внешнего источника через API;
5. Ответ внешней системы: код ошибки или получение данных;
6. Инициирование извлечения данных из внешней базы данных;
7. Ответ внешней системы: код ошибки или получение данных;
8. Инициирование и загрузка данных в хранилище;
9. Инициирование запуска правил ККД на стороне хранилища;
10. Ответ с результатом выполнения правил ККД;
11. Взаимодействие с метаданными хранилища;
12. Взаимодействие сервисов планировщика и оркестратора с модулями расширения;
13. Взаимодействие модулей расширения с модулем ETL;
14. Взаимодействие модулей расширения с модулем управления ККД;
15. Взаимодействие модулей расширения с модулем управления метаданными;

16. Получение информации о моделях, используемых в ETL (описание сущностей-источников и получателей);

17. Получение информации о моделях, используемых в ККД (описание правил).

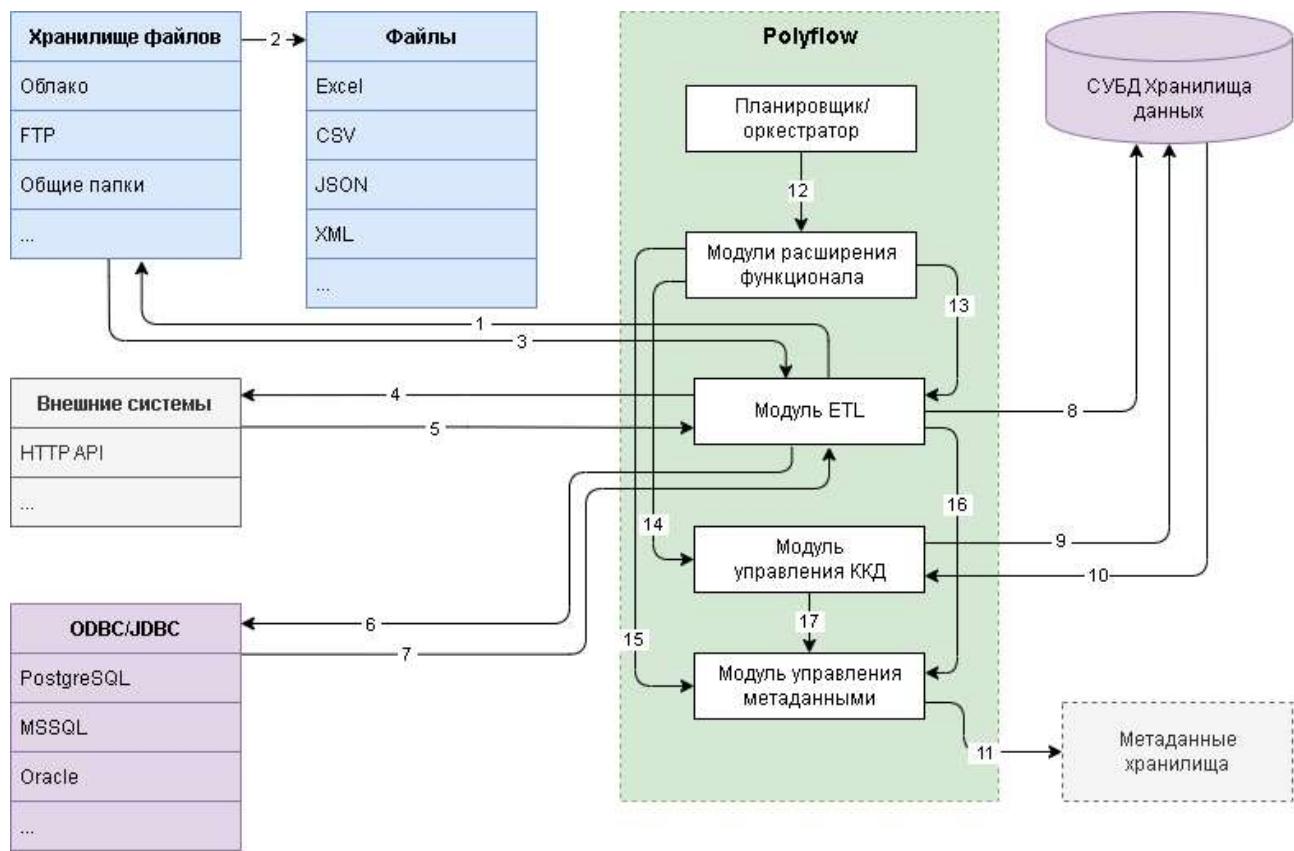


Рисунок 1. Архитектура решения