

Описание функциональных возможностей POLYHUB

О системе

POLYHUB – это система управления корпоративными данными (далее Система).

Система представляет собой платформу для:

- управления корпоративными хранилищами данных;
- моделирования структуры и потоков данных;
- просмотра и редактирование данных в хранилище;
- обеспечивает сервисы для организации хранения файлов в файловом хранилище;
- работы с различными провайдерами данных и разнородными источниками;
- настройки правил контроля качества данных;
- ведения каталога данных;
- управление словарем терминов и понятий, используемых в контексте работы с данными;
- анализа происхождения данных;
- создания отчетов в различных форматах;
- настройки уведомлений.

Функции

- Корпоративное хранилище с настраиваемым доступом к нему через визуальный интерфейс и из внешних систем по HTTP API
- Сбор структурированных данных из систем-источников из шин данных в хранилище
- Поддержка различных источников данных
 - JSON, PostgreSQL БД, ViQube API metadata/rawdata/query
 - HTML
- Визуальный конструктор для описания объектов системы
 - источники (Файлы: Excel, CSV, HTML, XML, JSON; внешние системы: HTTP API)
 - получателей (таблицы в БД, Visiology ViQube)
 - таблицы хранилища данных (DWH)
 - правила ККД
- Обработка данных из хранилища – декларативное управление данными, редактирование данных, справочников, привязка к ним
- Работа с данными из разных источников в едином документе
- Создание и обновление метаданных о данных, таких как названия таблиц, полей, типы данных
- Организация удобного поиска и навигации по каталогу данных
- Управление словарем терминов и понятий:
 - Создание и редактирование терминологического словаря с определениями терминов и связанными понятиями
 - Установление связей между терминами и их использование в контексте работы с данными
- Анализ происхождения данных:
 - Отслеживание и фиксация источников данных, включая их происхождение, метаданные
 - Построение связей между различными источниками данных для выявления зависимостей
- Генерация отчетов различных типов (xlsx, docx, jrxml) и поддержка HTTP API для внешних систем
- Поддержка форматов различными типами отчетов:
 - JRXML: docx, html, pdf, pptx, rtf, xls, xlsx
 - XLSX: xlsx, с возможностью полноценного редактирования
 - DOCX: docx, с возможностью редактирования
 - HTML: txt, html
- Управление объектами и группами объектов системы

- Отправка уведомлений по событиям, настраиваемым пользователями
- Облачный сервис сбора и хранения файлов с возможностью дальнейшего использования данных в конвейере обработки.
- Управление пользователями и их правами
- Управление настройками пользователей
- Настройка рабочих областей пользователя с удобным расположением объектов в соответствии с его ролью и функциями.

Определения и сокращения POLYHUB

Определения и сокращения POLYHUB представлены в Таблица 1.

Таблица 1. Определения POLYHUB

Термин/Сокращение	Определение
Платформа управления корпоративными данными POLYHUB	Многофункциональное серверное приложение с web-интерфейсом
POLYHUB	Краткое наименование программного обеспечения «Платформа управления корпоративными данными POLYHUB» (далее Система)
Облако	Модуль POLYHUB, реализующий облачный сервис сбора и хранения файлов
JSON-конструктор	Универсальный визуальный редактор POLYHUB с контекстно-зависимой палитрой инструментов для настройки объектов системы (сущностей источников и получателей, документов, отчетов, правила ККД)
Система управления данными Polyflow	Сервис оркестровки сбора и обработки разнородных данных хранилища произвольной архитектуры
Polyflow	Краткое наименование программного обеспечения «Система управления данными Polyflow»
Метаданные Polyflow	Метаданные Polyflow представлены в виде JSON-файлов, которые лежат в файловой системе на сервере Polyflow. Для описания и валидации различных типов сущностей используются JSON-схемы. Метаданные могут содержать сущности разных типов. Метаданные Polyflow могут описывать: <ul style="list-style-type: none">• источники (Файлы: Excel, CSV, HTML, XML, JSON; внешние системы: HTTP API)• получателей (таблицы в БД, Visiology ViQcube)• таблицы хранилища данных (DWH)• правила ККД. Для работы с метаданными в POLYHUB имеется визуальный редактор.
Стенд	Сервер с развернутой Системой
API	«Application Programming Interface», интерфейс программирования приложений, программный интерфейс приложения
ККД	Контроль качества данных. Процесс проверки данных на соответствие определенным стандартам и требованиям. ККД обеспечивает, что данные, используемые в хранилище данных, точны, полны и соответствуют ожиданиям пользователей.
DWH	Англ. Content Repository, Data Warehouse, DWH, хранилище данных
Хранилище данных, ХД	Предметно-ориентированная информационная база данных, сочетающая в себе функции системы управления версиями, поисковой машины и СУБД.
КХД	Корпоративное хранилище данных
Каталог данных	Централизованное хранилище метаданных, содержащее информацию о структуре, связях и характеристиках данных в организации. Он обеспечивает удобный доступ к данным, их

	управление, защиту и использование в различных приложениях и системах.
--	--

Архитектура решения

Система состоит из 8 модулей, закрывающих несколько областей работы с данными (Рисунок 1)

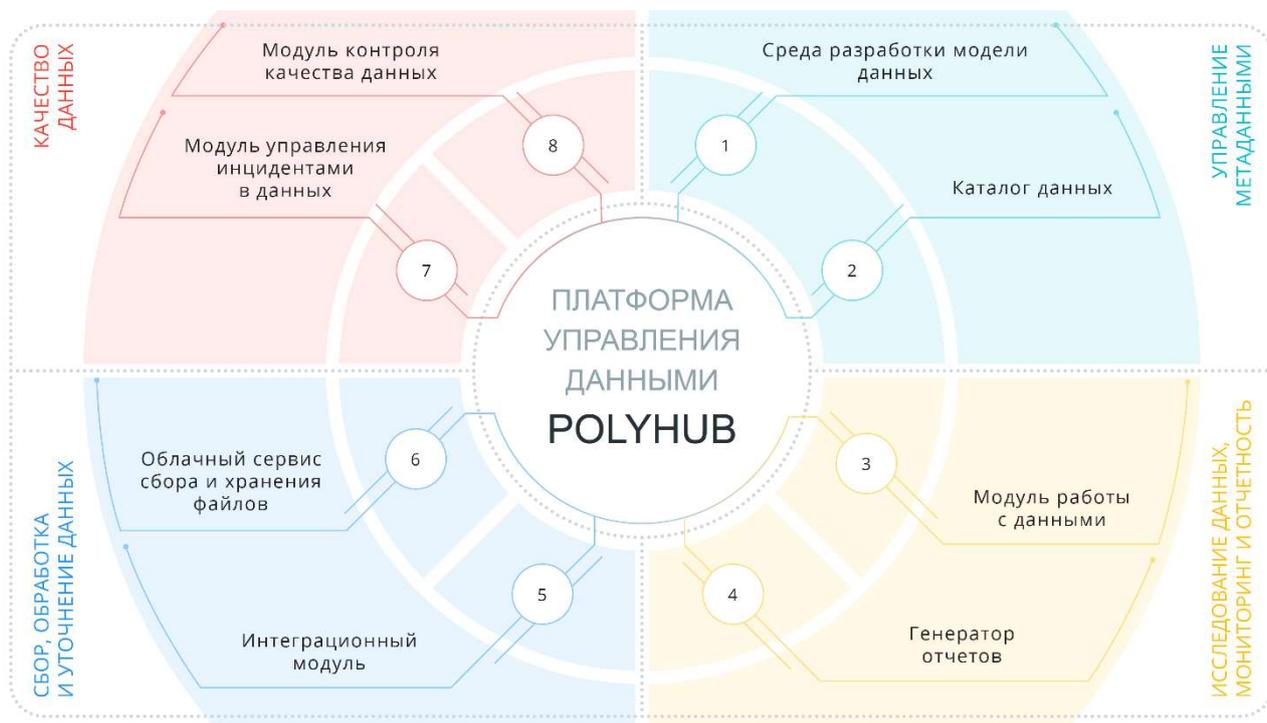


Рисунок 1. Платформа POLYHUB

Состав компонентов

Система включает в себя следующие модули:

1. Модуль разработки модели данных
2. Каталог данных
3. Модуль работы с данными
4. Генератор отчетов
5. Интеграционный модуль
6. Облачный сервис сбора и хранения файлов
7. Модуль управления инцидентами в данных
8. Модуль контроля качества данных
9. Системные модули (для упрощения схемы не показаны на Рисунок 1)

Модуль разработки модели данных

Включает подмодули:

1. Модели DWH
предоставляет возможность настройки и редактирования моделей объектов системы (источников и приемников);
2. JSON-схемы
предоставляет возможность создания описаний моделей различных типов объектов системы (источников и приемников, ККД, документов).

Каталог данных

Обеспечивает централизованное хранение метаданных о данных, их структуре, связях, качестве и использовании.

Модуль работы с данными (Документы)

Предоставляет возможность настройки документов для просмотра и редактирования данных из разных источников, взаимодействие с внешними системами по HTTP API.

Генератор отчетов (Отчеты)

Настройка, генерация и печать отчетов, в том числе по HTTP API.

Интеграционный модуль

Предоставляет публичное API для работы с облаком, документами.

Облачный сервис сбора и хранения файлов (Облако)

Предоставляет возможность загрузки, хранения и скачивания файлов через веб-интерфейс и по HTTP API. С корзинами облака может работать Polyflow, а также другие системы.

Модуль управления инцидентами в данных (Уведомления)

Предоставляет возможность отправки, настраиваемых на события системы, email-оповещений, в том числе по HTTP API.

Модуль контроля качества данных (Правила ККД)

Предоставляет возможность настройки и редактирования правил ККД для использования в Polyflow.

Системные модули

1. Администрирование
работа с пользователями системы, настройками, разрешениями, группами объектов;
2. Провайдеры
предоставляет данные по моделям, а также описывают источники данных (БД, внешние сервисы) и хранения;
3. Рабочие столы
обеспечивают гибкую настройку веб-интерфейса для объектов системы;
4. Подстановки
обеспечивают создание преднастроенных шаблонов для использования при работе с отчетами, уведомлениями.

Решаемые задачи

Далее приведены различные типы задач, которые решают модули Системы (Таблица 2).

Таблица 2. Задачи, решаемые модулями Системы

Тип задач	Модуль	Задачи
Управление метаданными	Модуль разработки модели данных Каталог данных	<ul style="list-style-type: none">– Описание различных источников данных для экстракта– Создание и редактирование терминологического словаря с определениями терминов и связанными понятиями– Установление связей между терминами и их использование в контексте работы с данными

		<ul style="list-style-type: none"> – Предоставление возможности поиска и быстрого доступа к информации из словаря – Отслеживание и фиксация источников данных, включая их происхождение, метаданные
Исследование данных, мониторинг и отчетность	Модуль работы с данными Генератор отчетов	<ul style="list-style-type: none"> – Редактирование данных в хранилище (например, в настроечных таблицах или в справочниках) – Получение регламентных отчетов
Сбор, обработка и уточнение данных	Облачный сервис сбора и хранения файлов (далее Облако) Интеграционный модуль	<ul style="list-style-type: none"> – Корпоративное хранилище – Настройки документов для просмотра и редактирования данных из разных источников, взаимодействие с внешними системами по HTTP API
Качество данных	Модуль контроля качества данных Модуль управления инцидентами в данных	<ul style="list-style-type: none"> – Описание правил ККД – Просмотр результатов правил ККД – Информирование о различных событиях системы, связанных с данными
Системные	Модуль администрирования; Модуль провайдеры Рабочие столы Подстановки	<ul style="list-style-type: none"> – Управление пользователями и разграничение доступа к объектам системы – Предоставление подключений к различным источникам данных и внешним системам – Настройка рабочей области пользователя в соответствии с его ролью и функциями

Основные взаимодействия

На Рисунок 2 представлена схема решения и основные взаимодействия между его частями и внешними системами:

1. Инициация запроса пользователем через интерфейс Системы к сервису front'а
2. Запрос сервиса front'а к сервису back'а
3. Запрос сервиса back'а к сервису работы с отчетами report
4. Запрос сервиса back'а к сервису базы данных db
5. Запрос сервиса back'а к сервису управления уведомлениями alert
6. Запрос сервиса back'а к сервису базы данных (структур данных) redis
7. Запрос сервиса работы с отчетами report к сервису базы данных db
8. Запрос сервиса управления уведомлениями alert к сервису базы данных db

9. Запрос внешними системами к сервису back'a
10. Запрос сервиса back'a к сервису каталога catalog'a
11. Запрос сервиса back'a к сервису каталога origin'a
12. Запрос сервиса back'a к сервису каталога skipper'a
13. Запрос сервиса back'a к сервису каталога orch-dagster-a'a
14. Запрос сервиса catalog'a к сервису каталога opensearch'a.

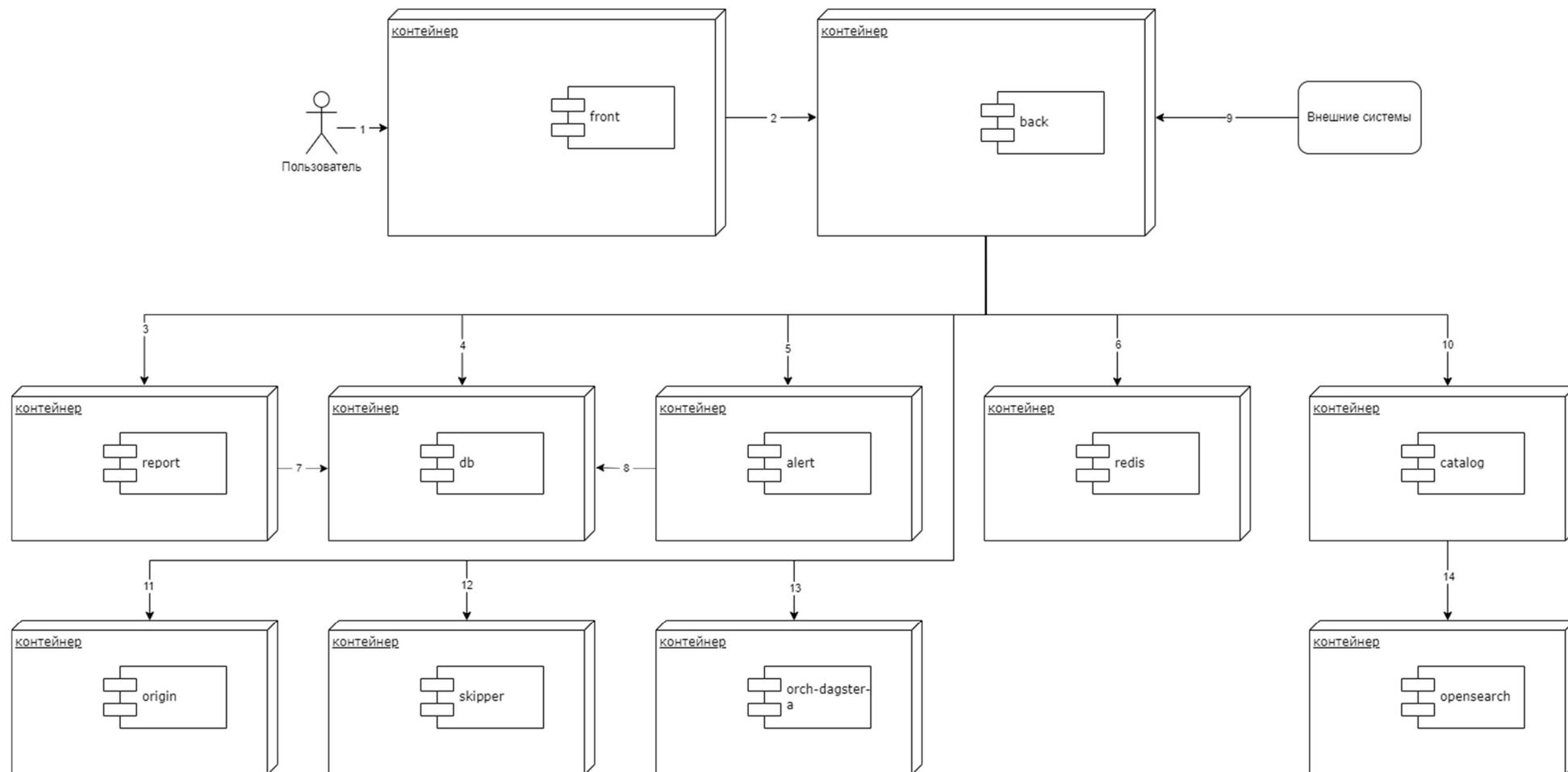


Рисунок 2. Архитектура решения