

Информация, необходимая для эксплуатации экземпляра программного обеспечения, предоставленного для проведения экспертной проверки программы

«Peraspera DataHunter»

Оглавление

1	СОСТАВ ИНСТАЛЛЯЦИОННОГО ПАКЕТА.....	1
2	ТРЕБОВАНИЯ К КВАЛИФИКАЦИИ СПЕЦИАЛИСТА, ВЫПОЛНЯЮЩЕГО УСТАНОВКУ ПРОГРАММЫ	2
3	ТРЕБОВАНИЯ К ОБОРУДОВАНИЮ	2
4	ТРЕБОВАНИЯ К СИСТЕМНОМУ ПРОГРАММНОМУ ОБЕСПЕЧЕНИЮ	2
5	ПОРЯДОК УСТАНОВКИ	3
6	ИНФОРМАЦИЯ, НЕОБХОДИМАЯ ДЛЯ ЭКСПЛУАТАЦИИ.....	14

1 Состав инсталляционного пакета

В состав инсталляционного пакета программы «Peraspera DataHunter» входит инсталлятор, который устанавливает программу «Peraspera DataHunter», а также указанное в нижеприведенной таблице (Таблица 1) свободно распространяемое ПО, необходимое для функционирования программы «Peraspera DataHunter».

Таблица 1 – Состав свободно распространяемого ПО, необходимого для функционирования программы «Peraspera DataHunter»

№	Свободно распространяемое ПО	Назначение
1	Keycloak	Обеспечение информационной безопасности
2	Postgresql	Хранение данных приложения
3	Scrapyd	Разворачивание и запуск спайдеров
4	Redis	Синхронизация работы спайдеров
5	Browserless chrome	Рендеринг html-контента сайтов
6	HAProxy	Балансировка нагрузки

2 Требования к квалификации специалиста, выполняющего установку Программы

Специалист должен иметь навыки работы с ОС Linux(Ubuntu), Docker, Kubernetes, Apache Kafka, протоколом безопасности OIDC (OpenID Connection).

3 Требования к оборудованию

Минимальные требования к оборудованию для инсталляции и функционирования программы «Peraspera DataHunter»:

Для функционирования узлов программы «Peraspera DataHunter», необходимо оборудование, характеристики которого приведены в таблицах (Таблица 2, Таблица 3).

Таблица 2 – Требования к среде установки

Наименование узла	Выделенное пространство	Объем памяти	Кол-во ядер процессора (CPUs)	Кол-во контроллеров (NICs)
dh-node1	300 GB	16 GB	8	1
dh-node2	300 GB	16 GB	8	1
dh-node3	300 GB	16 GB	8	1
dh-pg-node	500 GB	16 GB	8	1

Таблица 3 – Требования к физическим серверам

Наименование сервера	Выделенное пространство	Объем памяти	Кол-во ядер процессора (CPUs)	Кол-во контроллеров NICs
datahunter-01	300 GB	16 GB	8	2
datahunter-02	300 GB	16 GB	8	2
datahunter-03	300 GB	16 GB	8	2
datahunter-04	500 GB	16 GB	8	2

4 Требования к системному программному обеспечению

Требования, предъявляемые к системному программному обеспечению для установки и эксплуатации Программы «Peraspera DataHunter»:

- ubuntu сервер 18.04 и выше,
- docker 19.03.2 и выше,
- kubernetes 1.18.4 и выше

– Peraspera ESB.

Для функционирования Peraspera ESB в составе «Peraspera DataHunter», необходимо оборудование, характеристики которого приведены в таблицах (Таблица 4, Таблица 5).

Таблица 4 – Требования к среде установки

Наименование узла	Выделенное пространство	Объем памяти	Кол-во ядер процессора (CPUs)	Кол-во контроллеров (NICs)
peraspera-esb-node1	300 GB	16 GB	8	1
peraspera-esb-node2	300 GB	16 GB	8	1
peraspera-esb-node3	300 GB	16 GB	8	1

Таблица 5 – Требования к физическим серверам

Наименование сервера	Выделенное пространство	Объем памяти	Кол-во ядер процессора (CPUs)	Кол-во контроллеров NICs
peraspera-esb-01	300 GB	16 GB	8	2
peraspera-esb-02	300 GB	16 GB	8	2
peraspera-esb-03	300 GB	16 GB	8	2

5 Порядок установки

Для установки программы «Peraspera DataHunter» необходимо:

На оборудовании, соответствующем требованиям, приведенным в пункте 2.1.3 документа “Программа «Программное обеспечение для выполнения задач управляемого автоматизированного поиска и сбора больших объемов целевых данных в интернете и корпоративных источниках. Peraspera DataHunter» Описание функциональных характеристик программного обеспечения и информация, необходимая для установки и эксплуатации программного обеспечения” 72103154.62.01.29.686.001.91:

- 1) Развернуть операционную систему ubuntu сервер версии не ниже 18.04;
- 2) Создать кластер машин;
- 3) Установить Kubernetes версии не ниже 1.18.4;
- 4) Установить Peraspera ESB;
- 5) Загрузить по ссылке <https://disk.yandex.ru/d/ruk2PBnrnSideg> и распаковать архив с дистрибутивом (набором манифестов Kubernetes) в любую удобную директорию;

- 6) Настроить манифесты Kubernetes, скорректировав при необходимости в них значения по умолчанию (см. Таблица 6 вышеназванного документа);
- 7) Установить настроенные манифесты.

Таблица 6 – Значения по умолчанию настраиваемых параметров программы «Peraspera DataHunter»

№	Имя параметра	Значение параметра	Описание группы параметров
values.yaml			
1	global:		
2	ingress:		
3	domain:	data-hunter.company.ru	Домен, на котором располагается сервис.
4	className:	nginx	
5	postgresql:		
9	port:	5432	Порт СУБД
12	scrapyd:		
13	replicaCount:	3	Количество параллельно работающих сервисов.
14	port:	6800	Порт сервиса.
15	contextPath:	scrapyd	Путь расположения сервиса.
16	redis:		
17	port:	6379	Порт Redis.
18	splash:		
19	port:	8050	Порт Splash.
	Headless_chrome:		
	port:	3000	Порт Headless-Chrome.
	minReplicas:	1	Минимальное количество инстансов сервиса.
	maxReplicas:	10	Максимальное количество инстансов сервиса.
20	crawl_dispatcher:		
21	port:	8080	Порт Crawl-Dispatcher.
22	hypervisor:		
23	port:	8080	Порт Hypervisor.
24	config_server:		
25	port:	8888	Порт Config-Server.
Values.peraspera.yaml			
1	global:		

№	Имя параметра	Значение параметра	Описание группы параметров
2	dataHunter:		
3	domain:	data-hunter.ec-leasing.ru	Домен, на котором располагается сервис.
4	useSSL:	true	Флаг, использовать ли SSL сертификат.
5	oidc:		Настройка Keycloak.
6	realm:	data-hunter	Название Realm.
7	clientId:	data-hunter	
8	apiClientId:	data-hunter-api	
	apiClientSecret:	MjEyYTc2OGQtZmU4Ny00NTRILTlkM2EtOTJmNmVkMzllZGFk	
	needAccessForDevelopers:	true	
	users:		
	- name:	admin	Имя учетной записи администратора системы Data-Hunter.
	Password:	MXEwcDJ3OW8=	Пароль учетной записи администратора системы Data-Hunter, закодированный в base64.
	Email:	admin@data-hunter.ec-leasing.ru	Адрес электронной почты администратора Keycloak.
	firstName:	admin	Имя администратора.
	lastName:	admin	Фамилия администратора.
	clientRoles:		
	data-hunter:		
	- admin		
	realm-management:		
	- realm-admin		
	postgresql:		
	databases:		
	keycloak:		
	user:	a2V5Y2xvYWtfdXNlcg==	Имя учетной записи пользователя базы данных для сервиса Keycloak, закодированное в base64.

№	Имя параметра	Значение параметра	Описание группы параметров
	Password:	a2V5Y2xvYWtfdXNlcg==	Пароль учетной записи пользователя базы данных для сервиса Keycloak, закодированный в base64.
	Data_hunter:		
	user:	ZGF0YV9odW50ZXJfdXNlcg==	Имя учетной записи пользователя базы данных для системы Data-Hunter, закодированное в base64.
	Password:	ZGF0YV9odW50ZXJfdXNlcg==	Пароль учетной записи пользователя базы данных для системы Data-Hunter, закодированный в base64.
	Schemas:		
	- dispatcher		
	- hypervisor		
	- config		
	- demonstration		
10	scrapyd:		
11	minAvailable:	1	
12	updateStrategy:	RollingUpdate	
13	image:	devops.ec-leasing.ru/data-hunter/scrapyd	
14	imageTag:	1.3.1-data-hunter	
15	imagePullPolicy:	IfNotPresent	
16	imagePullSecrets:		
17	- name:	nexus-docker-registry	
18	init:		
19	image:	devops.ec-leasing.ru/busybox	
20	imageTag:	1.32.0	
21	configuration:		
22	jobs_to_keep:	5000	
23	max_proc:	32	
24	max_proc_per_cpu:	4	
25	finished_to_keep:	5000	

№	Имя параметра	Значение параметра	Описание группы параметров
26	poll_interval:	0.2	
	pypi:		
	index:	https://devops.ec-leasing.ru/nexus/repository/pypi-public//pypi	
	indexUrl:	https://devops.ec-leasing.ru/nexus/repository/pypi-public/simple	
	trustedHost:	devops.ec-leasing.ru	
	persistence:		
	enabled:	true	
	dataDirSize:	10Gi	
	dataDirStorageClass:	data-hunter	
	redis:		
	image:	devops.ec-leasing.ru/redis	
	imageTag:	6.2.6	
	imagePullSecrets:		
	- name:	nexus-docker-registry	
	init:		
	image:	devops.ec-leasing.ru/busybox	
	imageTag:	1.32.0	
	configuration:		
	numShards:	3	
	headless-chrome:		
	image:	devops.ec-leasing.ru/browserless/chrome	
	imageTag:	1.53-chrome-stable	
	imagePullPolicy:	IfNotPresent	
	imagePullSecrets:		
	- name:	nexus-docker-registry	
	haproxy:		
	image: devops.ec-leasing.ru/haproxy		
	imageTag: 2.6.6		
	imagePullPolicy: IfNotPresent		
	imagePullSecrets:		

№	Имя параметра	Значение параметра	Описание группы параметров
	- name: nexus-docker-registry		
	console:		
	user: YWRtaW4=		
	password: MxEwcDJ3OW8=		
	keycloak:		
	image:	devops.ec-leasing.ru/jboss/keycloak	
	imageTag:	15.0.2	
	imagePullSecrets:		
	- name:	nexus-docker-registry	
	user:	YWRtaW4=	Имя учетной записи администратора Keycloak, закодированный в base64.
	Password:	MxEwcDJ3OW8=	Пароль учетной записи администратора Keycloak, закодированный в base64.
	Persistence:		
	enable:	true	
	claim:		
	name:	keycloak	
	initialization:		
	enable:	true	
	crawl-dispatcher:		
	replicaCount:	1	
	image:	devops.ec-leasing.ru/data-hunter/crawl-dispatcher	
	imageTag:	1.3.0	
	imagePullPolicy:	Always	
	imagePullSecrets:		
	- name:	nexus-docker-registry	
	configuration:		
	tasks:		
	default-process-count:	1	
	kafka:		

№	Имя параметра	Значение параметра	Описание группы параметров
	brokers:		Настройка Kafka.
		“kafka-0.kafka.peraspera-esb-support:9092,\	
		kafka-1.kafka.peraspera-esb-support:9092,\	
		kafka-2.kafka.peraspera-esb-support:9092”	
	outputTopic:	data-hunter-task-responses	
	inputTopic:	data-hunter-tasks	
	cancelTopic:	data-hunter-tasks-cancel	
	springCloudBusInput:	data-hunter-config-input	
	springCloudBusOutput:	data-hunter-config-output	
	10iquibase:		
	defaultSchema:	dispatcher	
	enable:	false	
	timeZone:	Europe/Moscow	Часовой пояс.
	Hypervisor:		
	replicaCount:	1	
	affinity:	{ }	
	image:	devops.ec-leasing.ru/data-hunter/hypervisor	
	imageTag:	1.3.0	
	imagePullPolicy:	Always	
	imagePullSecrets:		
	- name:	nexus-docker-registry	
	configuration:		
	port:	8080	
	10iquibase:		
	defaultSchema:	hypervisor	
	enable:	false	
	kafka:		Настройка Kafka.
	Brokers:		
		“kafka-0.kafka.peraspera-esb-support:9092,\	

№	Имя параметра	Значение параметра	Описание группы параметров
		kafka-1.kafka.peraspera-esb-support:9092,\	
		kafka-2.kafka.peraspera-esb-support:9092”	
	port:	9092	
	searchProducer:		
	outputTopic:	data-hunter-tasks	
	taskResponseMessageProcessor:		
	outputTopic:	data-hunter-task-responses	
	jmx:		
	enabled:	true	
	port:	5555	
	liquibase11:		
	enabled:	true	
	timeZone:	Europe/Moscow	Часовой пояс.
	Config-server:		
	replicaCount:	1	
	affinity:	{ }	
	image:	devops.ec-leasing.ru/data-hunter/config-server	
	imageTag:	1.3.0	
	imagePullPolicy:	Always	
	imagePullSecrets:		
	- name:	nexus-docker-registry	
	timeZone:	Europe/Moscow	Часовой пояс.
	Configuration:		
	kafka:		Настройка Kafka.
	Brokers:		
		“kafka-0.kafka.peraspera-esb-support:9092,\	
		kafka-1.kafka.peraspera-esb-support:9092,\	
		kafka-2.kafka.peraspera-esb-	

№	Имя параметра	Значение параметра	Описание группы параметров
		support:9092”	
	inputTopic:	data-hunter-config-output	
	outputTopic:	data-hunter-config-input	
	l2iquibase:		
	defaultSchema:	config	
	enable:	true	
	data-hunter-web:		
	replicaCount:	1	
	image:	devops.ec-leasing.ru/data-hunter/data-hunter-web	
	imageTag:	prod-1.3.0	
	imagePullPolicy:	Always	
	imagePullSecrets:		
	- name:	nexus-docker-registry	
	configuration:		
	port:	80	
	timeZone:	Europe/Moscow	Часовой пояс.
	Oauth2:		
	clientId:	data-hunter	
	postgresql:		
	replicaCount:	1	
	image:	devops.ec-leasing.ru/postgis/postgis	
	imageTag:	13-3.1	
	imagePullPolicy:	IfNotPresent	
	imagePullSecrets:		
	persistence:		
	enabled:	true	
	dataDirSize:	300Gi	
	dataDirStorageClass:	data-hunter	
	timeZone:	Europe/Moscow	
	admin_user:	cG9zdGdyZXM=	Имя супер-пользователя базы данных.

№	Имя параметра	Значение параметра	Описание группы параметров
	Admin_password:	MXEwcDJ3OW8=	Пароль супер-пользователя базы данных.

6 Информация, необходимая для эксплуатации

6.1 Список терминов и определений

Кластер – группа спайдер-менеджеров, используемых для установки пакетов спайдеров.

Спайдер – программный компонент, выполняющий в зависимости от его типа поиск и сбор либо только сбор целевых данных с использованием поисковых механизмов источников данных.

Пакет спайдеров – набор программных компонент (дистрибутив) в файле формата egg, который содержит в себе один или несколько спайдеров.

Журнал работы кластера – журнал, предоставляющий администратору подробную информацию о работе спайдеров, с возможностью фильтрации по статусу, спайдер-менеджеру, пакету спайдеров, спайдеру и периоду времени.

Версия пакета спайдеров – численно-символьный идентификатор, позволяющий использовать в системе различные модификации одного и того же пакета спайдеров.

Спайдер-менеджер – программный компонент, служащий для управления установкой пакетов спайдеров и запуском спайдеров.

Краулер – спайдер, сконфигурированный с помощью настроек и параметров запуска для дальнейшего использования в конкретных задачах поиска и сбора целевых данных.

Параметры запуска краулера – совокупность конфигурируемых характеристик, используемых в краулере для описания конкретной задачи поиска и сбора целевых данных, реализуемой при помощи спайдера.

Настройки краулера – совокупность конфигурируемых характеристик для управления функционированием спайдера.

Проект – совокупность заданий, расписаний и справочных данных для решения конкретных бизнес-задач.

Задание – перечень объединенных между собой по смыслу задач для определенного краулера, необходимых для поиска и сбора целевых данных.

Задача – набор заполненных пользователем значений параметров запуска краулера для однократного запуска краулера.

Расписание – план запуска одного или нескольких заданий с указанием точного времени запуска или заданной периодичности запуска.

Журнал выполнения заданий – журнал, предоставляющий подробную информацию о выполнении заданий в рамках расписания, с возможностью фильтрации по краулеру, заданию, статусу и временному периоду.

Справочник – набор регулярно используемых списочных данных, необходимых для заполнения значений параметров запуска краулера при формировании заданий.

Журнал регистрации событий – журнал, предоставляющий информацию о действиях пользователя и других значимых событиях, происходящих в процессе работы с Peraspera DataHunter.

Технологический процесс – система взаимосвязанных действий, выполняющихся от момента получения требований к целевым данным до получения результата.

6.2 Продукты для конфигурирования и мониторинга компонент

Эксплуатация программы «Peraspera DataHunter» осуществляется администратором с использованием программных продуктов, представленных в нижеприведенной таблице (Таблица 7).

Таблица 7 – Перечень продуктов для конфигурирования и мониторинга компонент программы «Peraspera DataHunter»

№	Продукт	Назначение
1	СМАК(Cluster Manager for Apache Kafka)	Управление кластерами kafka: - мониторинг состояния кластера; - создание, конфигурирование и удаление тем; - конфигурирование и управление разделами тем
2	Kafka UI	Управление и мониторинг Peraspera DataHunter
3	Kubernetes Dashboard	Конфигурирование и управление Kubernetes: - разворачивание контейнеризированных приложений в кластере Kubernetes; - конфигурирование специализированных ресурсов Kubernetes (Deployments, Pods, Replica Sets и др.) - мониторинг выполняющихся приложений; - мониторинг и управление ресурсами кластера
4	Keycloak	Управление подсистемой информационной безопасности - создание и конфигурирование ролей; - создание и конфигурирование пользователей; - управление доступом пользователей к приложениям.
5	PGAdmin	Конфигурирование и управление Postgresql

6.3 Основные функции Peraspera DataHunter

- управление кластером краулеров;
- управление поиском и сбором данных.

6.4. Роли пользователей Peraspera DataHunter

Peraspera DataHunter предусматривает следующие пользовательские роли:

- Администратор;
- Аналитик данных.

6.4.1. Роль Администратора Peraspera DataHunter, обзор функций

Функционал Администратора:

- Управление пакетами спайдеров: установка, настройка, редактирование, удаление.
- Управление краулерами: создание, настройка, редактирование, удаление.
- Управление проектами: создание, настройка, редактирование, удаление.
- Просмотр статуса кластера и хостов.
- Просмотр статистики задач по хостам.
- Просмотр статистики задач по проектам и заданиям.
- Просмотр графиков динамики сбора элементов.
- Просмотр журнала выполнения заданий.
- Просмотр журнала регистрации событий.
- Просмотр личного профиля.
- Просмотр и поиск справочной информации.

6.4.2. Роль Аналитика данных Peraspera DataHunter, обзор функций

Функционал Аналитика данных:

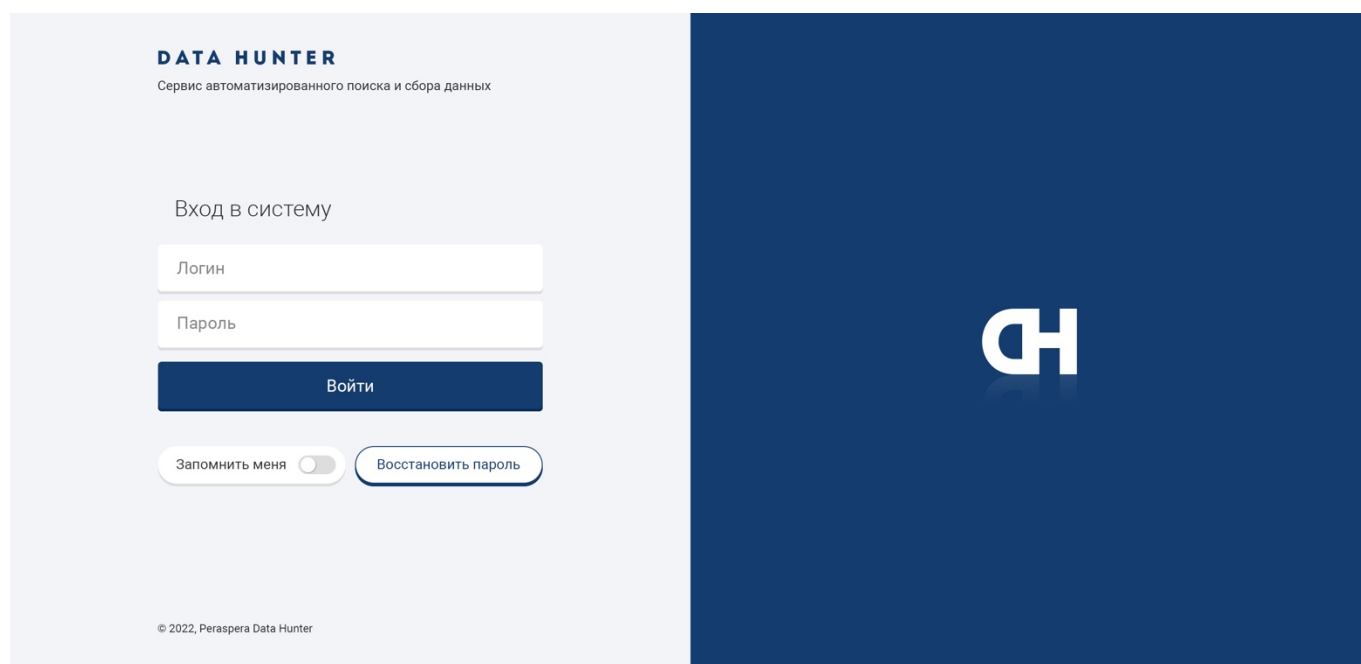
- Просмотр списка проектов.
- Управление заданиями: добавление, настройка, редактирование, удаление.
- Управление расписаниями: добавление, настройка, редактирование, удаление.
- Управление справочниками: добавление, настройка, редактирование, удаление.
- Просмотр журнала выполнения заданий.
- Просмотр журнала регистрации событий.
- Просмотр личного профиля.
- Просмотр и поиск справочной информации.

6.5. Администрирование Peraspera DataHunter

Интерфейс Администратора состоит из следующих логических блоков:

- Авторизация.
- Статус кластера и хостов.
- Статистика задач по хостам.
- Статистика задач по проектам и заданиям.
- Управление пакетами спайдеров.
- Управление краулерами.
- Управление проектами.
- Журнал выполнения заданий.
- Журнал регистрации событий.
- Профиль.
- Справочная информация.

6.5.1. Авторизация



Форма авторизации состоит из компонентов:

- **Поле «Логин»** – поле для введения логина или email, полученного для авторизации у сотрудника ИБ.
- **Поле «Пароль»** – поле для введения пароля, полученного для авторизации у сотрудника ИБ.
- **Кнопка «Войти»** – кнопка для авторизации в системе Peraspera DataHunter после ввода логина и пароля.
- **Опция «Запомнить меня»** – возможность оставаться пользователю в личном кабинете после закрытия сайта и повторного возвращения на него без ввода логина и пароля.

- Кнопка «Восстановить пароль» – функция, вызывающая диалоговое окно восстановления пароля.

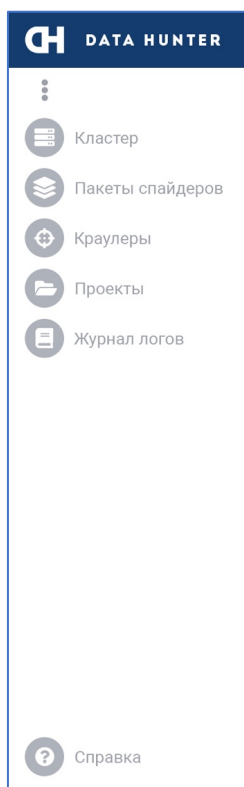
6.5.2. Личный кабинет: верхняя панель



Верхняя панель личного кабинета Администратора состоит из компонентов:

- Логотип – ссылка на главную страницу личного кабинета «Кластер»
- Заголовок – наименование текущего раздела личного кабинета
- Кнопка «Установить пакет спайдеров» – быстрый доступ к функции установки нового пакета спайдеров.
- Кнопка «Создать краулер» – быстрый доступ к функции создания нового краулера.
- Кнопка «Создать проект» – быстрый доступ к функции создания нового проекта.
- Кнопка «Журнал регистрации событий» – ссылка на страницу «Журнал регистрации событий».
- Кнопка «Профиль» – ссылка на страницу «Профиль».
- Кнопка «Выход из системы» – функция выхода из системы.

6.5.3. Личный кабинет: боковое меню



Боковое меню личного кабинета Администратора является системой навигации и состоит из компонентов:

- **Кнопка оптимизации меню** – функция изменения формы меню, скрывающая/открывающая текст меню.
- **Пункт меню «Кластер»** – ссылка на главную страницу: раздел «Кластер».
- **Пункт меню «Пакеты спайдеров»** – ссылка на раздел «Пакеты спайдеров».
- **Пункт меню «Краулеры»** – ссылка на раздел «Краулеры».
- **Пункт меню «Проекты»** – ссылка на раздел «Проекты».
- **Пункт меню «Журнал логов»** – ссылка на раздел «Журнал выполнения заданий».
- **Пункт меню «Справка»** – ссылка на раздел «Справочная информация».

6.5.4. Главная страница – Кластер: Статус кластера и хостов, статистика задач по хостам

The screenshot displays the 'Кластер' (Cluster) section of the DATA HUNTER application. The top navigation bar includes the 'DATA HUNTER' logo and a 'Кластер' tab. Below the navigation bar, there are three buttons: 'Установить пакет спайдеров', 'Создать краулер', and 'Создать проект'. The main content area is divided into two tabs: 'Статус кластера и хостов, статистика задач по хостам' (selected) and 'Статистика задач по проектам и заданиям'. The selected tab contains a table with the following data:

Кластер, хосты	Статус	Запланированных задач	Выполняющихся задач	Завершенных задач	Завершенных задач с предупреждениями	Завершенных задач с ошибками	Элементов	График
Кластер ЕСЛ СОФТ	✓	0	0	0	0	0	0	Смотреть
Наименование хоста # 1	✓	0	0	0	0	0	0	Смотреть
Наименование хоста # 2	✓	0	0	0	0	0	0	Смотреть
Наименование хоста # 3	✓	0	0	0	0	0	0	Смотреть
Наименование хоста # 4	✓	0	0	0	0	0	0	Смотреть

Below the table, there is a pagination control showing 'Строк на странице: 100' and a set of buttons for navigating between pages (1, 2, 3, >, >>, В конец).

The left sidebar contains a vertical menu with the following items: 'Кластер', 'Пакеты спайдеров', 'Краулеры', 'Проекты', 'Журнал логов', and 'Справка'.

Раздел «**Кластер**» личного кабинета Администратора является главной страницей и состоит из двух таблиц: «**Статус кластера и хостов, статистика задач по хостам**» и «**Статистика задач по проектам и заданиям**».

Таблица «**Статус кластера и хостов, статистика задач по хостам**» состоит из полей:

- **Кластер, хосты** – наименование кластера и хостов.
- **Статус** – статус состояния кластера и хостов.
- **Количество запланированных задач** – числовое значение запланированных задач.
- **Количество выполняющихся задач** – числовое значение выполняющихся задач.
- **Количество завершенных задач** – числовое значение завершенных задач.
- **Количество завершенных задач с предупреждениями** – числовое значение завершенных задач с предупреждением о возможных ошибках.
- **Количество завершенных задач с ошибками** – числовое значение задач, завершенных с ошибкой.
- **Количество собранных элементов** – числовое значение собранных на хосте элементов.

- **График динамики сбора элементов** – всплывающее окно с графиком сбора элементов.

В таблице предусмотрена возможность фильтрации полей по каждому из столбцов в алфавитном порядке в обе стороны по нажатию на иконку фильтрации (вверх/вниз) справа от наименования/иконки поля в верхней строке.

Под таблицей расположены:

- Форма настройки отображения количества строк в таблице.
- Постраничная навигация по таблице.

6.5.5. Главная страница – Кластер: Статистика задач по проектам и заданиям

Скриншот интерфейса «DATA HUNTER» в разделе «Кластер». В верхней панели расположены кнопки: «Установить пакет спайдеров», «Создать краулер», «Создать проект». В боковой панели — меню: «Кластер», «Пакеты спайдеров», «Краулеры», «Проекты», «Журнал логов». Основное содержимое — таблица «Статистика задач по проектам и заданиям».

Проекты, задания	Планируемых задач	Выполняющихся задач	Завершённых задач	Завершённых задач с предупреждениями	Элементов	График
Наименование проекта # 1	0	0	0	0	0	Смотреть
– Наименование задания # 1 проекта # 1	0	0	0	0	0	Смотреть
– Наименование задания # 2 проекта # 1	0	0	0	0	0	Смотреть
Наименование проекта # 2	0	0	0	0	0	Смотреть
Наименование проекта # 3	0	0	0	0	0	Смотреть

В нижней части таблицы — элементы управления: «Строк на странице: 100», кнопки навигации (1, 2, 3, >, >>) и «В конец».

Раздел «**Кластер**» личного кабинета Администратора является главной страницей и состоит из двух таблиц: «**Статус кластера и хостов, статистика задач по хостам**» и «**Статистика задач по проектам и заданиям**».

Таблица «**Статистика задач по проектам и заданиям**» состоит из полей:

- **Проекты, задания** – наименование проекта с декомпозицией на задания на втором уровне.
- **Количество запланированных задач** – числовое значение запланированных задач.
- **Количество выполняющихся задач** – числовое значение выполняющихся задач.
- **Количество завершённых задач** – числовое значение завершённых задач.
- **Количество завершённых задач с предупреждениями** – числовое значение завершённых задач с предупреждением о возможных ошибках.

- **Количество завершенных задач с ошибками** – числовое значение задач, завершенных с ошибкой.
- **Количество собранных элементов** – числовое значение собранных на хосте элементов.
- **График динамики сбора элементов** – всплывающее окно с графиком сбора элементов.

В таблице предусмотрена возможность фильтрации полей по каждому из столбцов в алфавитном порядке в обе стороны по нажатию на иконку фильтрации (вверх/вниз) справа от наименования/иконки поля в верхней строке.

Над таблицей расположена форма фильтрации данных таблицы с возможностью выдачи результатов либо за все время, либо за установленный период.

Под таблицей расположены:

- Форма настройки отображения количества строк в таблице.
- Постраничная навигация по таблице.

6.5.6. Главная страница – Кластер: График динамики сбора элементов

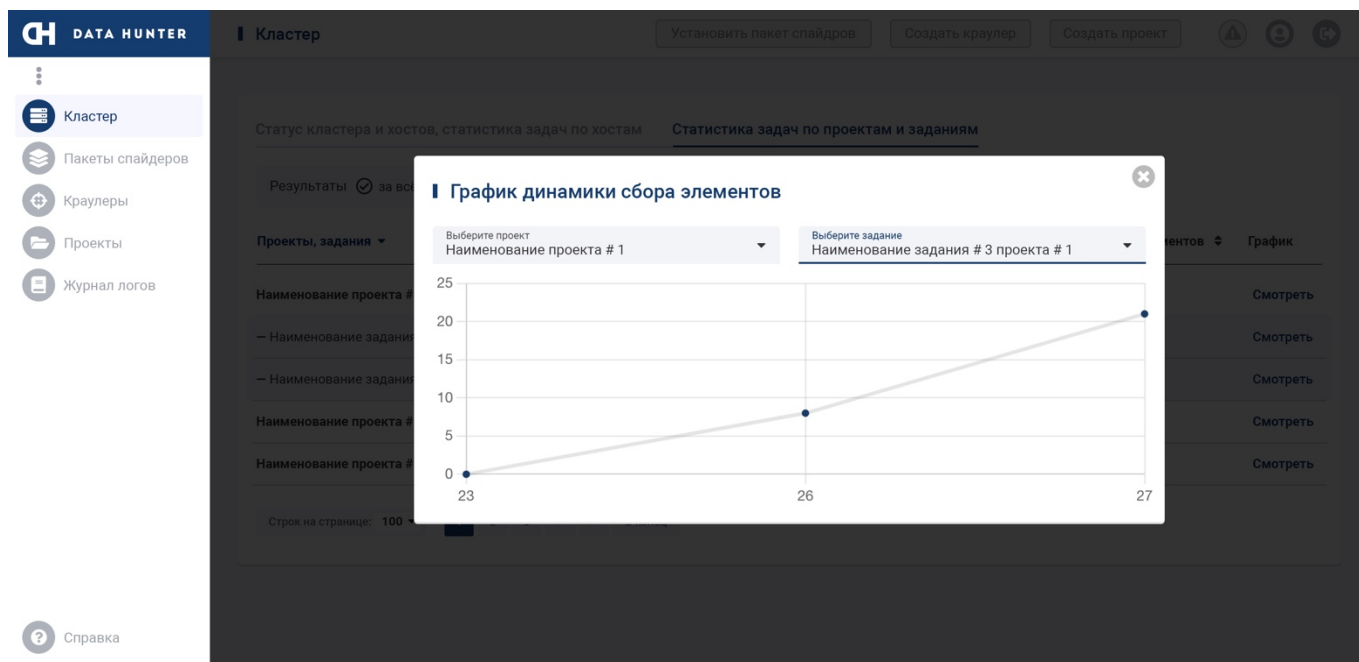


График динамики сбора элементов отображает количество собранных элементов по установленному Администратором проекту и заданию проекта за период выполнения задания.

Для отображения графика Администратором должны быть определены и проект, и задание в рамках выбранного проекта.

6.5.7. Пакеты спайдеров

Установить новый пакет спайдеров

Пакеты спайдеров	Добавлен	Хосты	Описание
Наименование пакета спайдеров # 1	28-05-22 10:20	Scrapy-1, Scrapy-2, Scrapy-3	Спайдеры для поиска в группах Telegram
– Версия SP001 пакета спайдеров # 1	28-05-22 10:20	Scrapy-1, Scrapy-2, Scrapy-3	Описание версии пакета спайдеров # 1
– Спайдер # 1 пакета # 1	28-05-22 10:20	Scrapy-2, Scrapy-3	Спайдер для поиска телефонов в Telegram
– Спайдер # 2 пакета # 1	28-05-22 10:20	Scrapy-1	Спайдер для поиска телефонов в Telegram
– Версия SP002 пакета спайдеров # 1	28-05-22 10:20	Scrapy-1, Scrapy-2, Scrapy-3	Описание версии пакета спайдеров # 2
Наименование пакета спайдеров # 2	28-05-22 10:20	Scrapy-2	Спайдеры для поиска в группах VK
Наименование пакета спайдеров # 3	28-05-22 10:20	Scrapy-1, Scrapy-3	Спайдеры для поиска в группах Одноклассни ...

Строк на странице: 100 1 2 3 > >> В конец

Раздел «Пакет спайдеров» личного кабинета Администратора состоит из компонентов:

- Кнопка «Установить новый пакет спайдеров» – открытие диалогового окна с установкой нового пакета спайдеров.
- Таблица пакетов спайдеров – таблица установленных пакетов спайдеров.

Таблица установленных пакетов спайдеров состоит из полей:

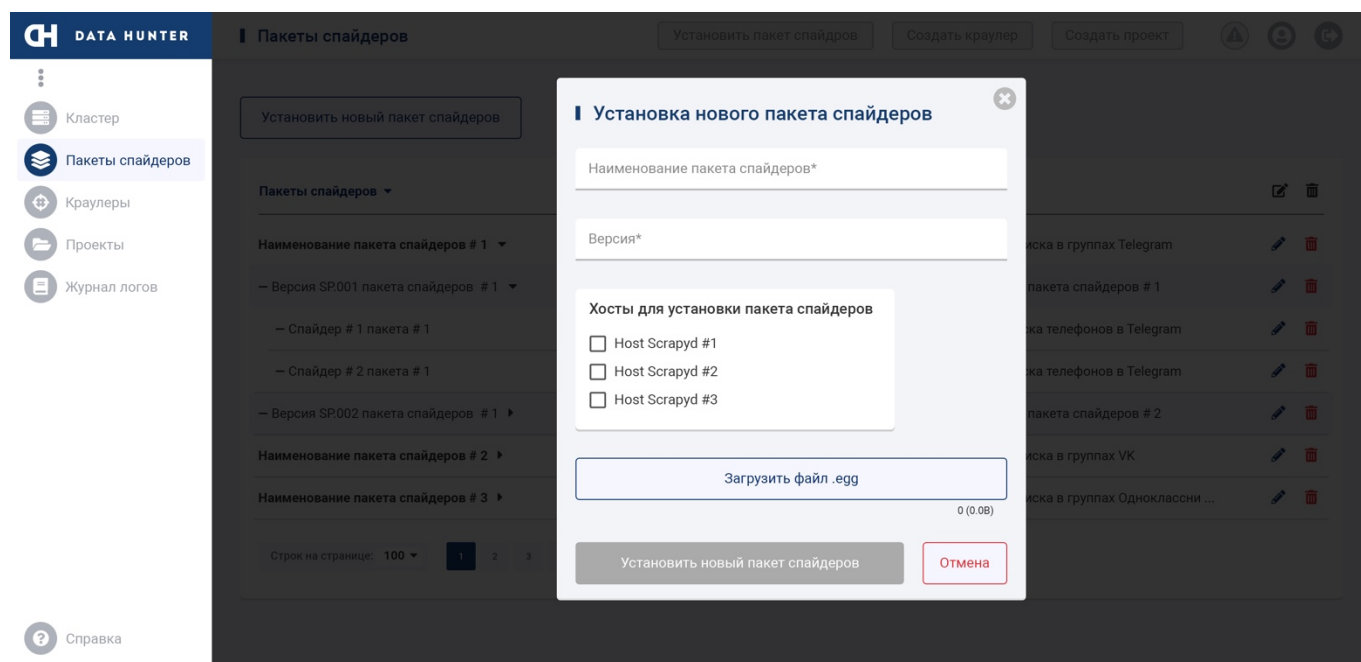
- Пакеты спайдеров – наименование установленного пакета спайдеров с декомпозицией на версии на втором уровне и на спайдеры, из которых эти версии состоят, на третьем уровне.
- Добавлен – дата добавления пакета спайдеров.
- Хосты – хост или список хостов, на которые установлен пакет спайдеров.
- Описание – текстовое описание пакета спайдеров, предусмотренное разработчиком пакета спайдеров.
- Редактирование – диалоговое окно с редактированием пакета спайдеров.
- Удаление – удаление установленного пакета спайдеров из системы.

В таблице предусмотрена возможность фильтрации полей по каждому из столбцов в алфавитном порядке в обе стороны по нажатию на иконку фильтрации (вверх/вниз) справа от наименования/иконки поля в верхней строке.

Под таблицей расположены:

- Форма настройки отображения количества строк в таблице.
- Постраничная навигация по таблице.

6.5.7.1. Установка нового пакета спайдеров



Диалоговое окно «Установка нового пакета спайдеров» доступно:

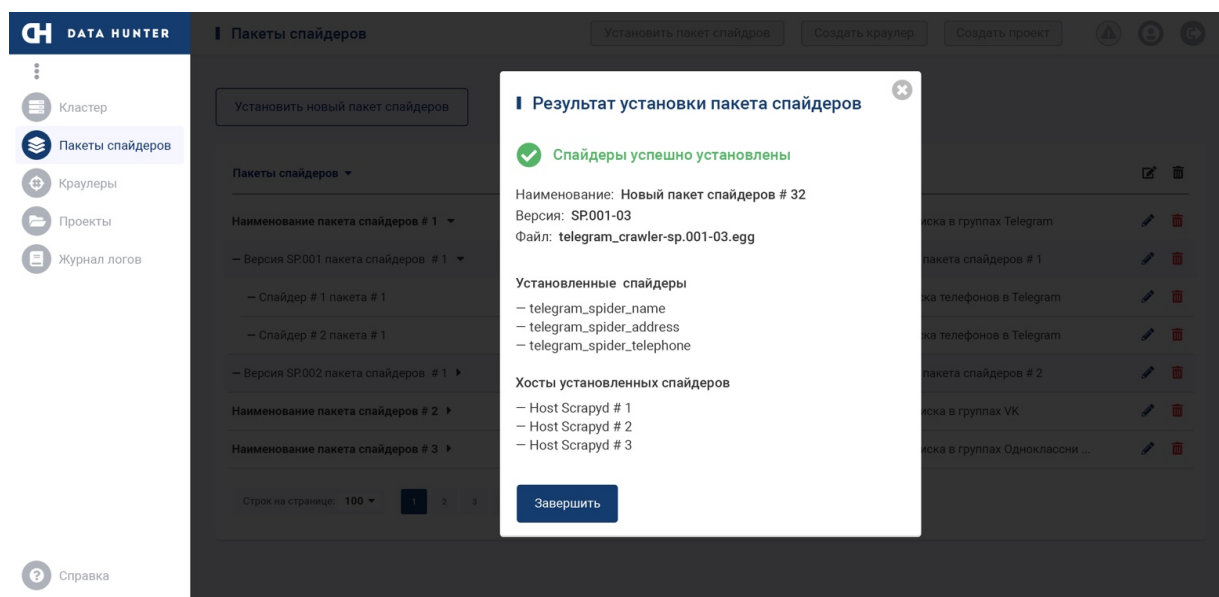
- Из раздела «Пакеты спайдеров» по нажатию на кнопку «Установить новый пакет спайдеров».
- Из верхней панели личного кабинета Администратора по нажатию на кнопку «Установить пакет спайдеров».

Для установки нового пакета спайдеров нужно вызвать диалоговое окно «Установка нового пакета спайдеров» и совершить следующие действия:

- Указать наименование пакета.
- Указать версию пакета спайдеров.
- Указать хосты из списка доступных для размещения пакета спайдеров.
- Загрузить файл формата egg с дистрибутивом пакета спайдеров.
- Подтвердить действия нажатием на кнопку «Установить новый пакет спайдеров».

Все действия в диалоговом окне «Установка нового пакета спайдеров» обязательны для выполнения.

После завершения установки нового пакета спайдеров система выдаст результат установки пакета спайдеров, после чего новый пакет спайдеров появится в таблице установленных пакетов спайдеров в разделе «Пакеты спайдеров».



6.5.7.2. Действия с пакетами спайдеров: редактирование, удаление

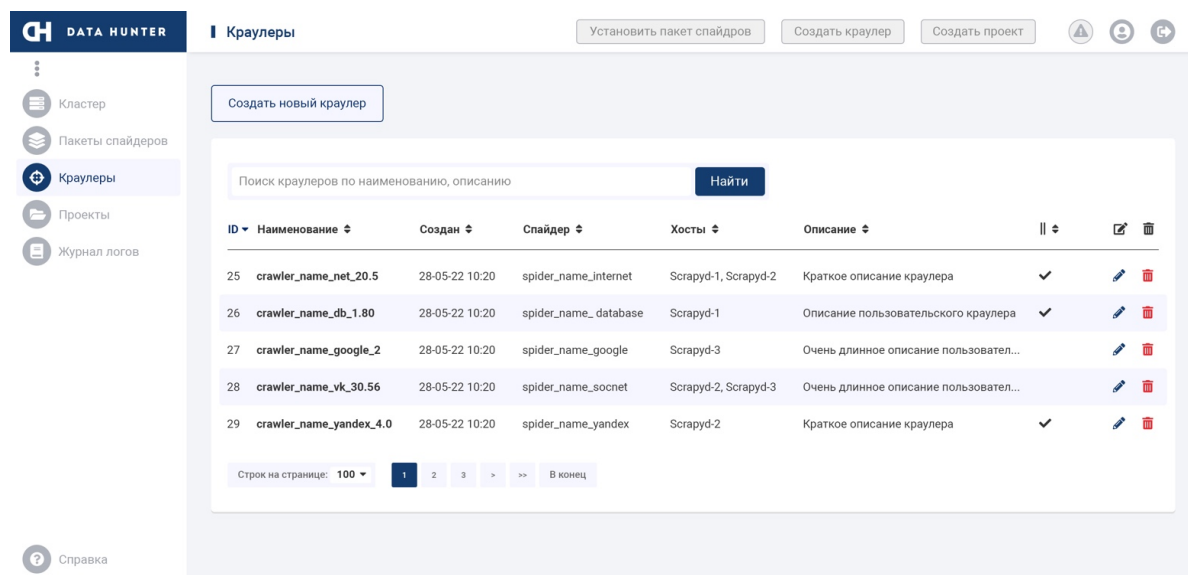
Редактирование пакета спайдеров

Диалоговое окно «Редактирование пакета спайдеров» вызывается из таблицы установленных пакетов спайдеров в разделе «Пакеты спайдеров» по нажатию на иконку карандаша в строке нужного пакета спайдера. Изменения, внесенные в пакет спайдеров через диалоговое окно «Редактирование пакета спайдеров», подтверждаются кнопкой «Сохранить изменения».

Удаление пакета спайдеров

Пакет спайдера может быть удален из таблицы установленных пакетов спайдеров в разделе «Пакеты спайдеров» по нажатию на иконку корзины в строке нужного пакета спайдера.

6.5.8. Краулеры



Раздел «Краулеры» личного кабинета Администратора состоит из компонентов:

- **Кнопка «Создать новый краулер»** – открытие диалогового окна с созданием нового краулера.
- **Таблица краулеров** – таблица созданных краулеров.

Таблица краулеров состоит из полей:

- **ID** – идентификационный номер краулера, автоматически присваиваемый системой при создании.
- **Наименование** – наименование краулера, установленное Администратором при создании.
- **Добавлен** – дата создания краулера.
- **Спайдер** – спайдер, из которого сконфигурирован.
- **Хосты** – хост или список хостов, на которые установлен спайдер, из которого сконфигурирован краулер.
- **Описание** – текстовое описание краулера, установленное Администратором при создании.
- **Распараллеливание** – отметка об активации функции параллельного поиска и сбора данных.
- **Редактирование** – диалоговое окно с редактированием краулера.
- **Удаление** – удаление созданного краулера из системы.

В таблице предусмотрена возможность фильтрации полей по каждому из столбцов в алфавитном порядке в обе стороны по нажатию на иконку фильтрации (вверх/вниз) справа от наименования/иконки поля в верхней строке.

Над таблицей расположено поле поиска краулера по наименованию и описанию.

Под таблицей расположены:

- Форма настройки отображения количества строк в таблице.
- Постраничная навигация по таблице.

6.5.8.1. Создание нового краулера

Диалоговое окно «Новый краулер» доступно:

- Из раздела «Краулеры» по нажатию на кнопку «Создать новый краулер».
- Из верхней панели личного кабинета Администратора по нажатию на кнопку «Новый краулер».

Для создания нового краулера нужно вызвать диалоговое окно «Новый краулер» и пройти три этапа:

- Идентификация краулера и выбор спайдера.
- Конфигурация настроек.
- Параметры запуска.

Создание краулера, шаг 1: Идентификация краулера и выбор спайдера

Скриншот интерфейса «DATA HUNTER» с диалоговым окном «Новый краулер: шаг #1. Идентификация, спайдер».

Диалоговое окно содержит следующие элементы:

- Прогресс-бар с тремя шагами (1, 2, 3), где шаг 1 является активным.
- Поле «Наименование краулера*».
- Поле «Описание краулера*».
- Выпадающий список «Пакет спайдеров*».
- Выпадающий список «Версия*».
- Выпадающий список «Спайдер*».
- Чекбокс «Активировать функцию параллельного выполнения поиска и сбора данных».
- Кнопки «Сохранить и перейти к настройкам» и «Отмена».

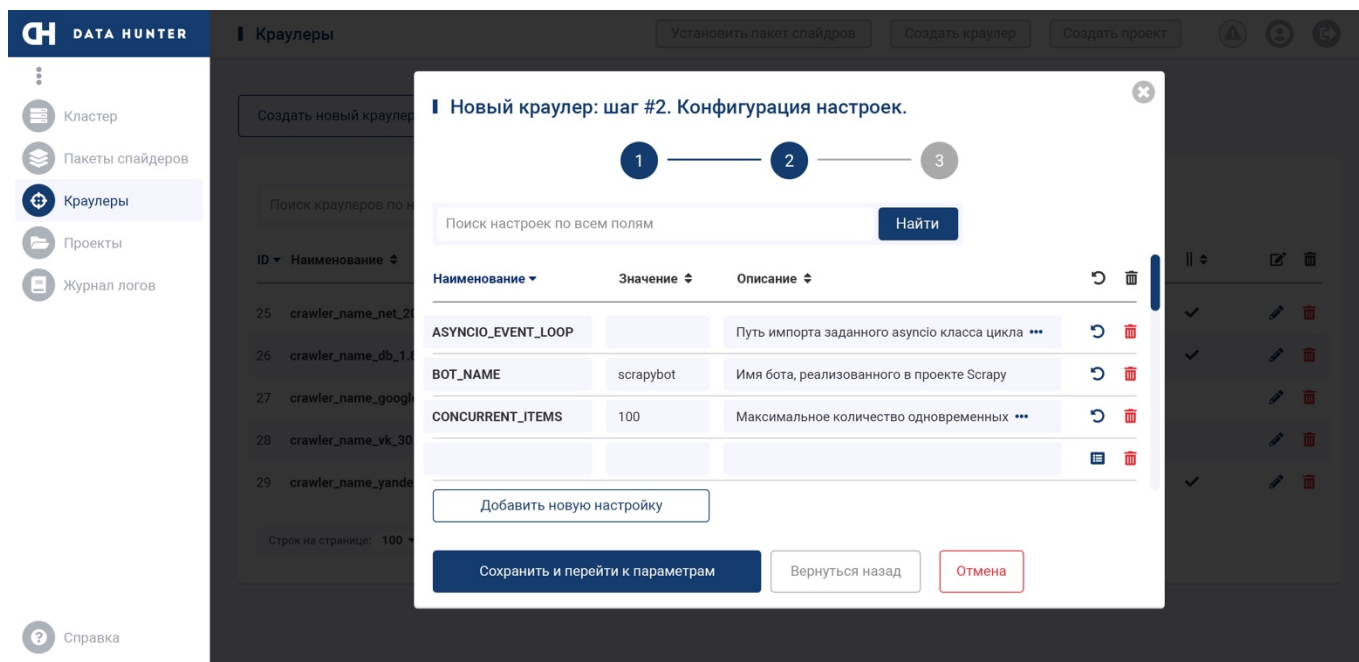
На заднем плане виден интерфейс приложения с меню слева и таблицей существующих краулеров.

На первом шаге создания краулера в диалоговом окне «Новый краулер: шаг #1. Идентификация, спайдер.» нужно выполнить следующие действия:

- Указать наименование краулера.
- Указать описание краулера.
- Выбрать пакет спайдеров, содержащий спайдер для создания краулера.
- Выбрать версию пакета спайдеров.
- Выбрать спайдер для конфигурации краулера.
- Опционально активировать функцию параллельного выполнения поиска и сбора данных.
- Сохранить данные и перейти к конфигурации настроек по нажатию на кнопку «Сохранить и перейти к настройкам».

Все действия в диалоговом окне «Новый краулер: шаг #1. Идентификация, спайдер.» обязательны для выполнения, кроме активации функции параллельного выполнения поиска и сбора данных.

Создание краулера, шаг 2: Конфигурация настроек



На втором шаге создания краулера в диалоговом окне «Новый краулер: шаг #2. Конфигурация настроек.» нужно выполнить следующие действия:

- Сконфигурировать настройки краулера в таблице настроек краулера.
- Сохранить данные и перейти к параметрам запуска по нажатию на кнопку «Сохранить и перейти к параметрам».

Конфигурация настроек краулера при создании краулера

Настройки краулера при его создании конфигурируются в таблице настроек краулера в диалоговом окне «Новый краулер: шаг #2. Конфигурация настроек.» на втором шаге создания краулера в разделе «Краулеры».

Таблица настроек краулеров состоит из полей:

- **Наименование** – наименование настройки краулера на английском языке.
- **Значение** – значение настройки краулера.
- **Описание** – текстовое описание настройки краулера.
- **Сброс** – функция сброса настроек краулера до состояния по умолчанию (к настройкам спайдера).
- **Удаление** – удаление настройки краулера.

В таблице предусмотрена возможность фильтрации полей по каждому из столбцов в алфавитном порядке в обе стороны по нажатию на иконку фильтрации (вверх/вниз) справа от наименования/иконки поля в верхней строке.

Над таблицей расположено поле поиска настройки краулера по всем полям.

Под таблицей расположена форма добавления новой настройки краулера, наименование, значение и описание которой можно:

- Заполнить вручную в соответствующих полях строки новой настройки краулера.
- Выбрать из списка настроек спайдера по нажатию на иконку списка справа от полей настроек краулера.

Новая настройка краулера добавляется по нажатию на кнопку «Добавить новую настройку» под таблицей настроек краулера.

Конфигурация настроек краулера не является обязательной к выполнению для перехода к следующему шагу. По умолчанию краулер наследует настройки спайдера.

Создание краулера, шаг 3: Параметры запуска

Скриншот интерфейса «DATA HUNTER» на шаге 3 создания краулера. В центре экрана открыто диалоговое окно «Новый краулер: шаг #3. Параметры запуска». В окне есть прогресс-бар с тремя шагами (1, 2, 3). Под ним находится строка поиска параметров со значком «Найти». Ниже — таблица параметров:

Наименование	Значение по умолчанию	Описание
query		Поисковый запрос
max_page	5	Количество страниц поисковой выдачи
time	Вчерашний день	Генератор относительного времени

Внизу таблицы — кнопка «Добавить новый параметр». В самом низу — три кнопки: «Сохранить и завершить создание краулера», «Вернуться назад» и «Отмена».

На третьем шаге создания краулера в диалоговом окне «Новый краулер: шаг #3. Параметры запуска.» нужно выполнить следующие действия:

- Настроить параметры запуска краулера.
- Сохранить данные и завершить создание краулера по нажатию на кнопку «Сохранить и завершить создание краулера».

Настройка параметров запуска краулера при создании краулера

Параметры запуска краулера при его создании настраиваются в таблице параметров запуска краулера в диалоговом окне «Новый краулер: шаг #3. Параметры запуска.» на третьем шаге создания краулера в разделе «Краулеры».

Таблица параметров запуска краулера состоит из полей:

- **Наименование** – наименование параметра запуска краулера на английском языке.
- **Значение по умолчанию** – значение параметра запуска краулера. Поле предусматривает генератор значений, который дает возможность сгенерировать значение параметра краулера в зависимости от дополнительных условий.
- **Описание** – текстовое описание параметра запуска краулера.
- **Удаление** – удаление параметра запуска краулера.

В таблице предусмотрена возможность фильтрации полей по каждому из столбцов в алфавитном порядке в обе стороны по нажатию на иконку фильтрации (вверх/вниз) справа от наименования/иконки поля в верхней строке.

Над таблицей расположено поле поиска параметра запуска краулера по всем полям.

Под таблицей расположена форма добавления нового параметра запуска краулера, наименование, значение по умолчанию и описание которого можно заполнить вручную в соответствующих полях строки нового параметра запуска краулера.

Новый параметр запуска краулера добавляется по нажатию на кнопку «Добавить новый параметр» под таблицей параметров запуска краулера.

Настройка параметров запуска краулера не является обязательной к выполнению для завершения создания краулера. По умолчанию краулер наследует параметра запуска спайдера.

6.5.8.2. Действия с краулерами: редактирование, удаление

Редактирование краулера

Диалоговое окно «Редактирование краулера» вызывается из таблицы созданных краулеров в разделе «Краулеры» по нажатию на иконку карандаша в строке нужного краулера. Изменения, внесенные в краулер через диалоговое окно «Редактирование краулера», подтверждаются кнопкой «Сохранить изменения».

Удаление краулера

Краулер может быть удален из таблицы созданных краулеров в разделе «Краулеры» по нажатию на иконку корзины в строке нужного краулера.

6.5.9. Проекты

Создать новый проект

Поиск проектов по всем полям **Найти**

ID	Наименование	Создан	Описание	Оператор		
25	Сбор данных Банк БКС	28-05-22 10:20	Краткое описание пользовательского проекта	operatornic01		
26	Телефоны поставщиков	28-05-22 10:20	Краткое описание пользовательского проекта	operatornic01, operatornic02		
27	Адреса Вкусно и Точка	28-05-22 10:20	Краткое описание пользовательского проекта	operatornic01, operatornic02, operatornic03		
28	Нелегальные компании	28-05-22 10:20	Краткое описание пользовательского проекта	operatornic03		
29	Нарушения ПДД	28-05-22 10:20	Краткое описание пользовательского проекта	operatornic02		

Строк на странице: 100 1 2 3 > >> В конец

Справка

Раздел «Проекты» личного кабинета Администратора состоит из компонентов:

- Кнопка «Создать новый проект» – открытие диалогового окна с созданием нового проекта.
- Таблица проектов – таблица созданных проектов.

Таблица созданных проектов состоит из полей:

- ID – идентификационный номер проекта, автоматически присваиваемый системой при создании.
- Наименование – наименование проекта, установленное Администратором при создании.
- Создан – дата создания проекта.
- Описание – текстовое описание проекта, установленное Администратором при создании.
- Аналитик данных – логин или список логинов Аналитиков данных, за которыми закреплен проект.
- Редактирование – диалоговое окно с редактированием проекта.
- Удаление – удаление созданного проекта из системы.

В таблице предусмотрена возможность фильтрации полей по каждому из столбцов в алфавитном порядке в обе стороны по нажатию на иконку фильтрации (вверх/вниз) справа от наименования/иконки поля в верхней строке.

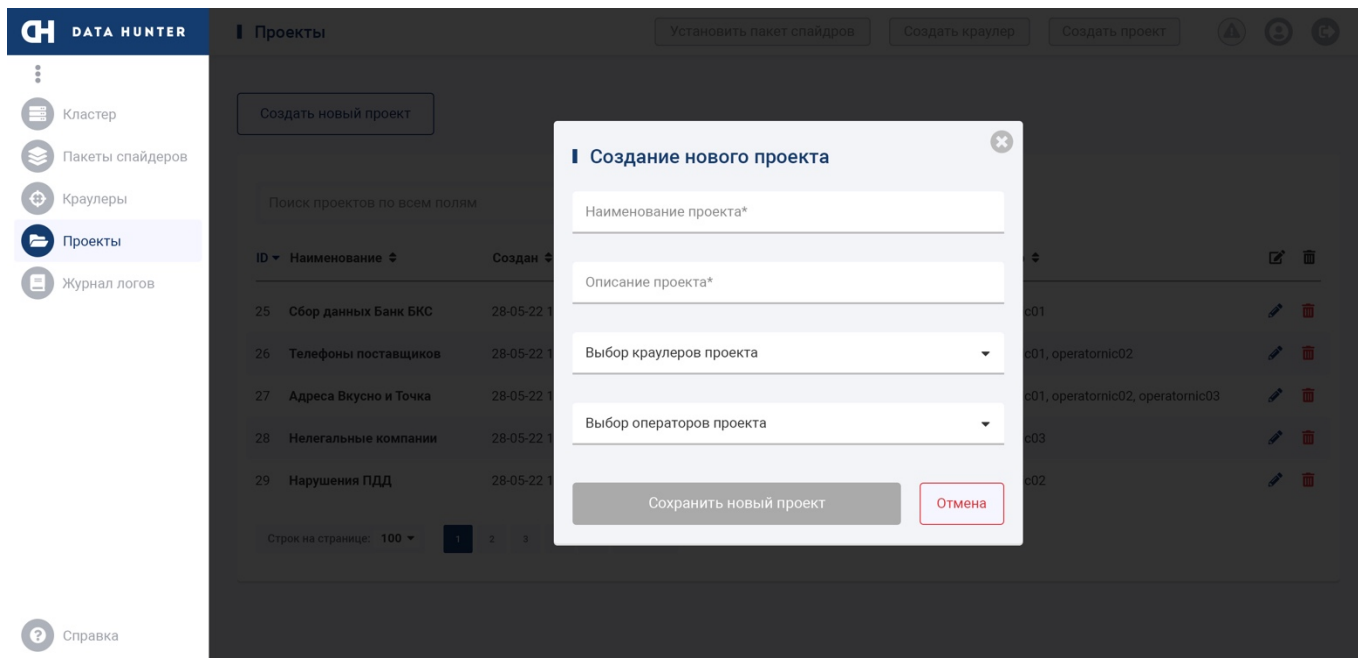
Над таблицей расположено поле поиска проектов по всем полям.

Под таблицей расположены:

- Форма настройки отображения количества строк в таблице.

- Постраничная навигация по таблице.

6.5.9.1. Создание нового проекта



Диалоговое окно «Создание нового проекта» доступно:

- Из раздела «Проекты» по нажатию на кнопку «Создать новый проект».
- Из верхней панели личного кабинета Администратора по нажатию на кнопку «Создать проект».

Для установки нового проекта нужно вызвать диалоговое окно «Создание нового проекта» и совершить следующие действия:

- Указать наименование проекта.
- Указать описание проекта.
- Выбрать краулеры проекта.
- Назначить операторов проекта.
- Подтвердить действия нажатием на кнопку «Сохранить новый проект».

Все действия в диалоговом окне «Создание нового проекта» обязательны для выполнения.

После завершения установки нового проекта система выдаст результат установки проекта, после чего новый пакет спайдеров появится в таблице созданных проектов в разделе «Проекты».

6.5.9.2. Действия с проектами: редактирование, удаление

Редактирование проекта

Диалоговое окно «Редактирование проекта» вызывается из таблицы созданных проектов в разделе «Проекты» по нажатию на иконку карандаша в строке нужного проекта. Изменения, внесенные проект через диалоговое окно «Редактирование проекта», подтверждаются кнопкой «Сохранить изменения».

Удаление проекта

Проект может быть удален из таблицы созданных проектов в разделе «Проекты» по нажатию на иконку корзины в строке нужного проекта.

5.10. Журнал выполнения заданий

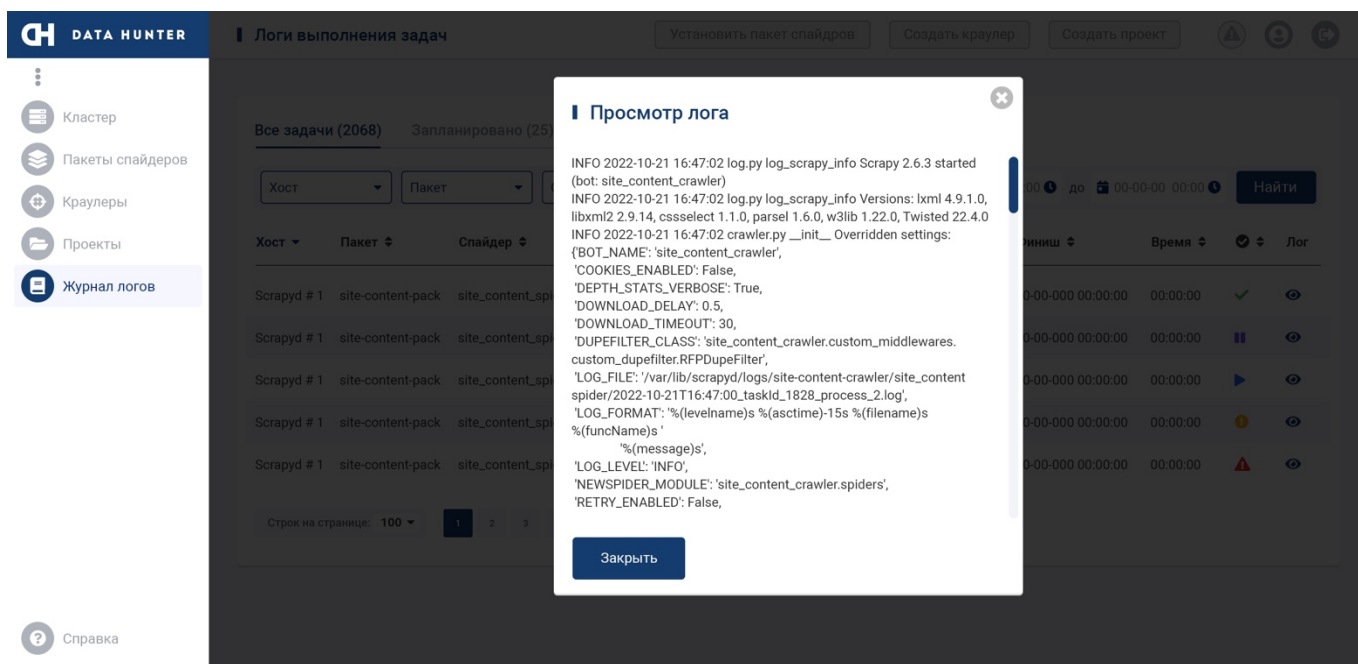
Скриншот интерфейса «Журнал выполнения задач» в Data Hunter. Вверху панели задач отображены кнопки: «Установить пакет спайдеров», «Создать краулер», «Создать проект». Слева находится меню с пунктами: «Кластер», «Пакеты спайдеров», «Краулеры», «Проекты», «Журнал логов». Основное пространство занимает таблица с заголовком «Логи выполнения задач». В таблице есть фильтры: «Хост», «Пакет», «Спаyder», «За всё время», «За период от», «до», «Найти». Таблица имеет следующие столбцы: «Хост», «Пакет», «Спаyder», «Задача», «Старт», «Финиш», «Время», «Статус», «Лог». В таблице представлено 5 записей, все с датой «2022-10-21T16:47:00_taskid_1828_process_2» и временем «00:00:00». Статусы: «За всё время», «За период от», «За период от», «За период от», «За период от». Внизу таблицы есть пагинация: «Строк на странице: 100», «1», «2», «3», «>», «>>», «В конце».

Раздел «Журнал выполнения заданий» личного кабинета Администратора представлен таблицей выполнения задач, состоящей из полей:

- **Хост** – хост или список хостов, на которых выполняется задача.
- **Пакет** – пакет спайдеров, включающий спайдер для выполнения задачи.
- **Спаyder** – спайдер для выполнения задачи.
- **Задача** – идентификатор задачи.
- **Старт** – дата запуска выполнения задачи.
- **Финиш** – дата завершения выполнения задачи.
- **Время** – время выполнения задачи.
- **Статус** – статус задачи с возможными значениями:
 - выполнение задачи завершено;

- выполнение задачи запланировано;
- задача в процессе выполнения;
- выполнение задачи завершено с предупреждением о возможной ошибке;
- выполнение задачи завершено с ошибкой/ошибками.

- **Лог** – иконка глаза для просмотра лога выполнения задачи во всплывающем окне.



В таблице предусмотрена возможность фильтрации полей по каждому из столбцов в алфавитном порядке в обе стороны по нажатию на иконку фильтрации (вверх/вниз) справа от наименования/иконки поля в верхней строке.

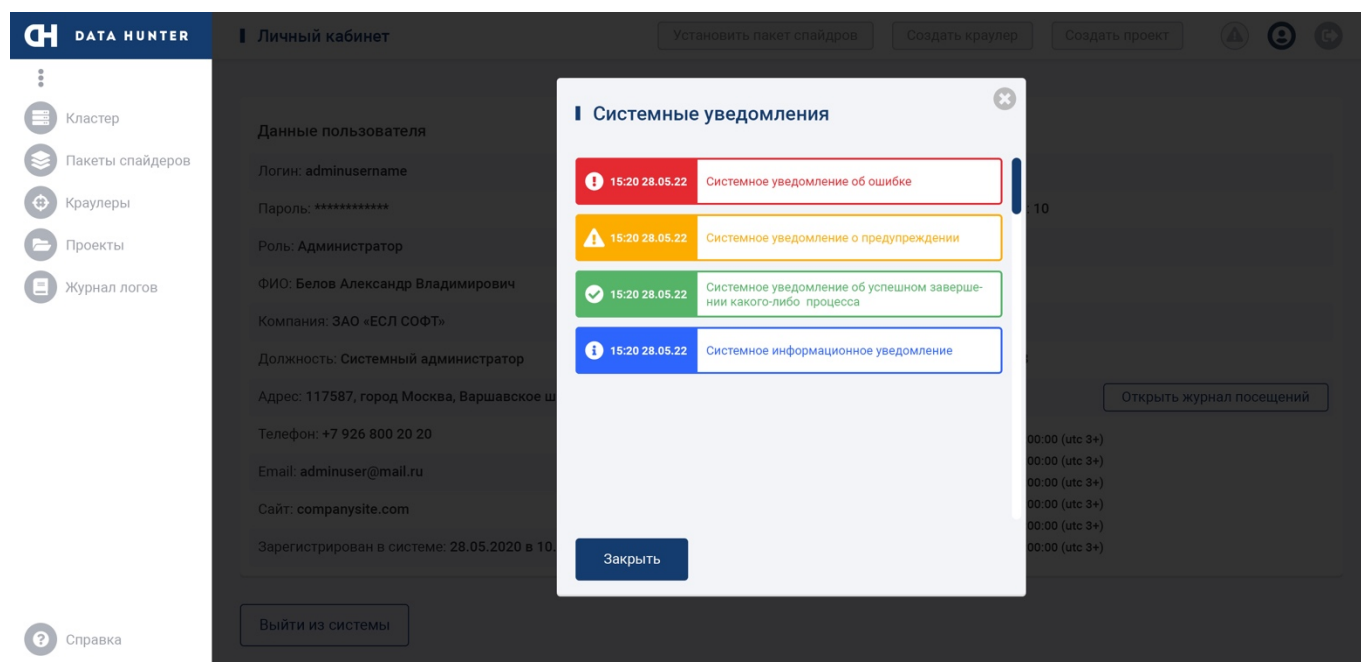
Над таблицей расположена форма фильтрации данных:

- Фильтр по статусу задач: все задачи / запланированные / выполняющиеся / завершённые.
- Фильтр по хосту.
- Фильтр по пакету спайдеров.
- Фильтр по времени выполнения задач с возможностью выдачи результатов либо за все время, либо за установленный период.

Под таблицей расположены:

- Форма настройки отображения количества строк в таблице.
- Постраничная навигация по таблице.

6.5.11. Журнал регистрации событий



Раздел «Журнал регистрации событий» доступен по нажатию на иконку с восклицательным знаком: «Журнал регистрации событий» из верхней панели личного кабинета Администратора.

Статусы записей журнала регистрации событий:

- **Красный** – системное уведомление об ошибке.
- **Желтый** – системное уведомление с предупреждением.
- **Зеленый** – системное уведомление об успешном завершении какого-либо процесса / действия.
- **Синий** – нейтральное системное уведомление.

Системные уведомления появляются в виде всплывающих окон во время сессии Администратора в личном кабинете, информируя Администратора о том или ином событии. Системные уведомления, полученные вне сессии Администратора в личном кабинете доступны в журнале регистрации событий.

6.5.12. Профиль

DATA HUNTER | Личный кабинет

Установить пакет спайдеров | Создать краулер | Создать проект

Кластер
Пакеты спайдеров
Краулеры
Проекты
Журнал логов

Справка

Данные пользователя

Логин: adminusername
Пароль: *****
Роль: Администратор
ФИО: Белов Александр Владимирович
Компания: ЗАО «ЕСЛ СОФТ»
Должность: Системный администратор
Адрес: 117587, город Москва, Варшавское шоссе, д. 125, стр. 1
Телефон: +7 926 800 20 20
Email: adminuser@mail.ru
Сайт: companysite.com
Зарегистрирован в системе: 28.05.2020 в 10.23 (utc +3:00)

Выйти из системы

Статистика профиля

Хостов в кластере: 3
Добавлено пакетов спайдеров: 10
Создано краулеров : 32
Создано проектов: 28
Привязанных операторов: 5
Записей в журнале логов: 2068
Рабочих сессий: 256 [Открыть журнал посещений](#)

29.05.2022 00:00:00 - 29.05.2022 00:00:00 (utc 3+)
28.05.2022 00:00:00 - 28.05.2022 00:00:00 (utc 3+)
26.05.2022 00:00:00 - 26.05.2022 00:00:00 (utc 3+)
25.05.2022 00:00:00 - 25.05.2022 00:00:00 (utc 3+)
23.05.2022 00:00:00 - 23.05.2022 00:00:00 (utc 3+)
22.05.2022 00:00:00 - 22.05.2022 00:00:00 (utc 3+)

Раздел «Профиль» доступен по нажатию на иконку пользователя: «Профиль» из верхней панели личного кабинета Администратора. Раздел «Профиль» личного кабинета Администратора состоит из компонентов:

- Блок «Данные пользователя».
- Блок «Статистика профиля».
- Кнопка выхода из системы «Выйти из системы».

Блок «Данные пользователя» профиля состоит из полей:

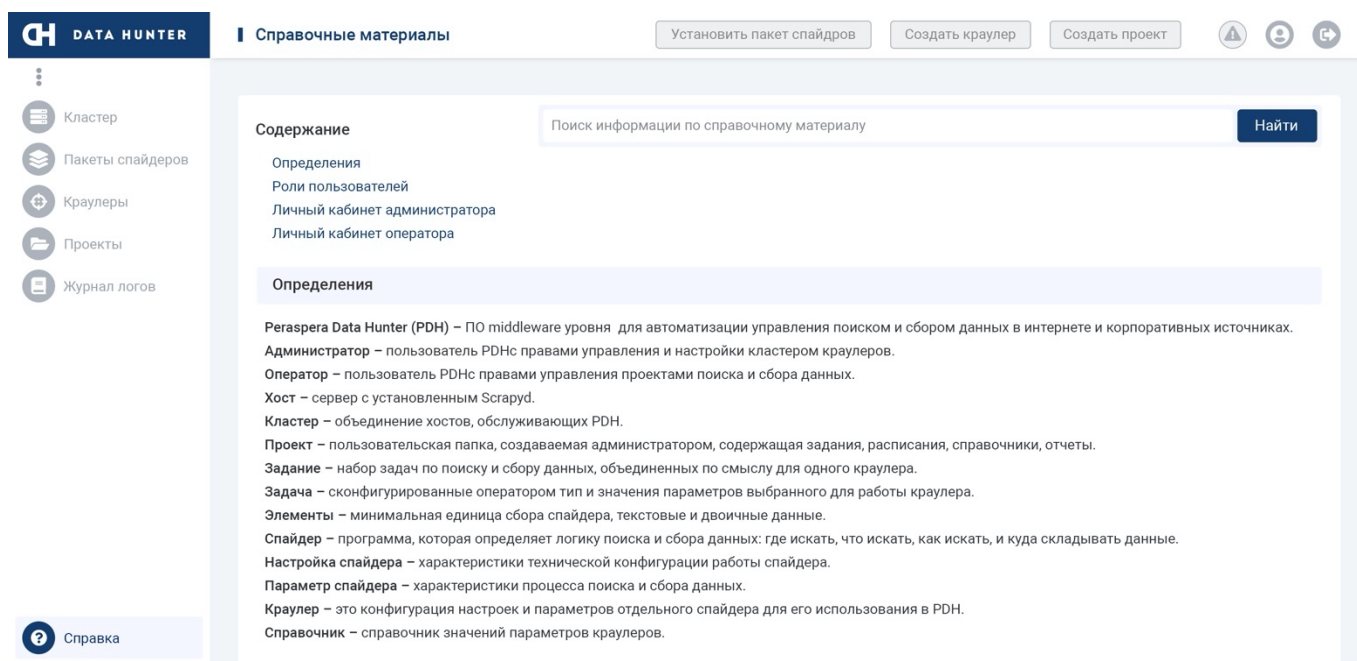
- Логин.
- Пароль.
- Роль.
- ФИО.
- Компания.
- Должность.
- Адрес.
- Телефон.
- Email.
- Сайт.
- Дата регистрации в системе.

Блок «Статистика профиля» профиля состоит из полей:

- Хостов в кластере.
- Добавлено пакетов спайдеров.
- Создано краулеров.

- Создано проектов.
- Привязано операторов.
- Записей в журнале выполнения заданий.
- Рабочих сессий с возможностью просмотра как последних записей, так и всего журнала посещений по нажатию на кнопку «Открыть журнал посещений».

6.5.13. Справочная информация



Раздел «Справочная информация» доступен в боковом меню, пункт меню «Справка». Раздел «Справочная информация» личного кабинета Администратора состоит из компонентов:

- Поле поиска информации по справочным материалам.
- Содержание справочных материалов.
- Руководство пользователя.

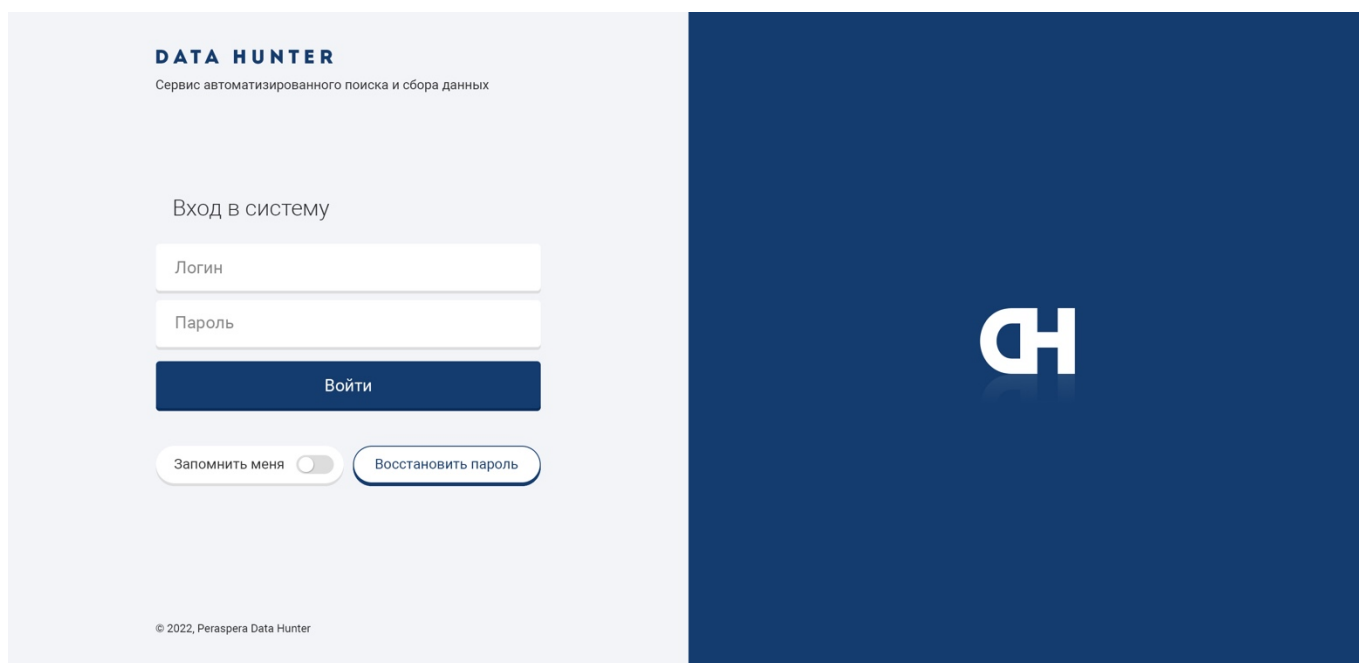
6. Эксплуатация аналитиком данных Peraspera DataHunter

Интерфейс Аналитика данных состоит из следующих логических блоков:

- Авторизация.
- Список проектов.
- Управление заданиями.
- Управление расписаниями.
- Управление справочниками.
- Журнал выполнения заданий.
- Журнал регистрации событий.

- Профиль.
- Справочная информация.

6.6.1. Авторизация



Форма авторизации состоит из компонентов:

- **Поле «Логин»** – поле для введения логина или email, полученного для авторизации у сотрудника ИБ.
- **Поле «Пароль»** – поле для введения пароля, полученного для авторизации у сотрудника ИБ.
- **Кнопка «Войти»** – кнопка для авторизации в системе Peraspera DataHunter после ввода логина и пароля.
- **Опция «Запомнить меня»** – возможность оставаться пользователю в личном кабинете после закрытия сайта и повторного возвращения на него без ввода логина и пароля.
- **Кнопка «Восстановить пароль»** – функция, вызывающая диалоговое окно восстановления пароля.

6.6.2. Личный кабинет: верхняя панель

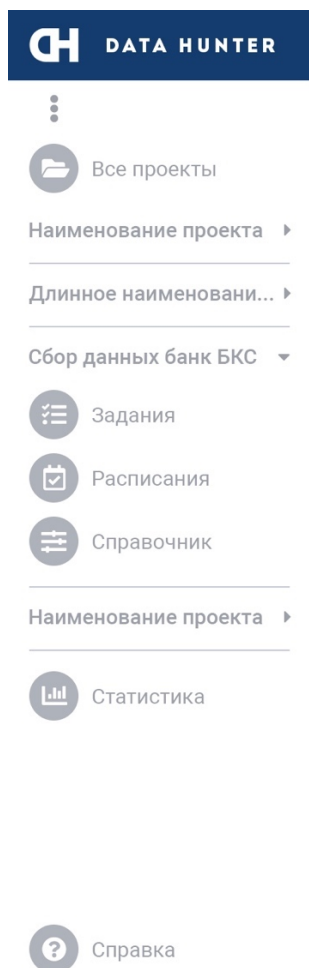


Верхняя панель личного кабинета Аналитика данных состоит из компонентов:

- **Логотип** – ссылка на главную страницу личного кабинета «Проекты».
- **Заголовок** – наименование текущего раздела личного кабинета.
- **Кнопка «Перейти в отчет»** – ссылка на страницу просмотра отчетов.

- Кнопка «Журнал регистрации событий» – ссылка на страницу «Журнал регистрации событий».
- Кнопка «Профиль» – ссылка на страницу «Профиль».
- Кнопка «Выход из системы» – функция выхода из системы.

6.6.3. Личный кабинет: боковое меню



Боковое меню личного кабинета Аналитика данных является системой навигации и состоит из компонентов:

- Кнопка оптимизации меню – функция изменения формы меню, скрывающая/открывающая текст меню.
- Пункт меню «Все проекты» – ссылка на главную страницу: раздел «Проекты».
- Пункт меню «[Наименование проекта]» – раскрывает подпункты меню текущего проекта.
 - Подпункт меню «Задания» – ссылка на раздел «Задания».
 - Подпункт меню «Расписания» – ссылка на раздел «Расписания».
 - Подпункт меню «Справочник» – ссылка на раздел «Справочник».
- Пункт меню «Статистика» – ссылка на раздел «Статистика выполнения заданий».
- Пункт меню «Справка» – ссылка на раздел «Справочная информация».

6.6.4. Главная страница – Проекты

The screenshot shows the 'DATA HUNTER' web application interface. The top bar includes the 'DATA HUNTER' logo and a 'Проекты' (Projects) tab. A 'Перейти в отчет' (Go to report) button is visible. The sidebar on the left contains navigation links: 'Все проекты' (All projects), 'Наименование проекта' (Project name), 'Длинное наименование...' (Long name...), 'Сбор данных банк БКС' (Data collection bank BKS), 'Задания' (Tasks), 'Расписания' (Schedules), 'Справочник' (Reference), 'Наименование проекта' (Project name), 'Статистика' (Statistics), and 'Справка' (Help). The main content area features a 'Запрос на новый проект' (Request for new project) button and a search bar with the text 'Поиск проектов по всем полям' (Search projects by all fields) and a 'Найти' (Find) button. Below the search bar is a table of projects with columns: ID, Наименование (Name), Создан (Created), and Описание (Description). The table contains five rows of project data. At the bottom of the table, there is a pagination control showing 'Строк на странице: 100' (Rows per page: 100) and a set of navigation buttons (1, 2, 3, >, >>, В конце).

ID	Наименование	Создан	Описание
25	Сбор данных Банк БКС	28-05-22 10:20	Краткое текстовое описание пользовательского проекта, заданное админом
26	Телефоны поставщиков	28-05-22 10:20	Краткое текстовое описание пользовательского проекта, заданное админом
27	Адреса Вкусно и Точка	28-05-22 10:20	Краткое текстовое описание пользовательского проекта, заданное админом
28	Нелегальные компании	28-05-22 10:20	Краткое текстовое описание пользовательского проекта, заданное админом
29	Нарушения ПДД	28-05-22 10:20	Краткое текстовое описание пользовательского проекта, заданное админом

Раздел «Проекты» личного кабинета Аналитика данных состоит из компонентов:

- Кнопка «Запрос на новый проект» – сообщение Администратору с запросом на создание нового проекта через форму обратной связи.
- Таблица проектов – таблица привязанных проектов.

Таблица созданных проектов состоит из полей:

- ID – идентификационный номер проекта, автоматически присваиваемый системой при создании.
- Наименование – наименование проекта, установленное Администратором при создании.
- Создан – дата создания проекта.
- Описание – текстовое описание проекта, установленное Администратором при создании.

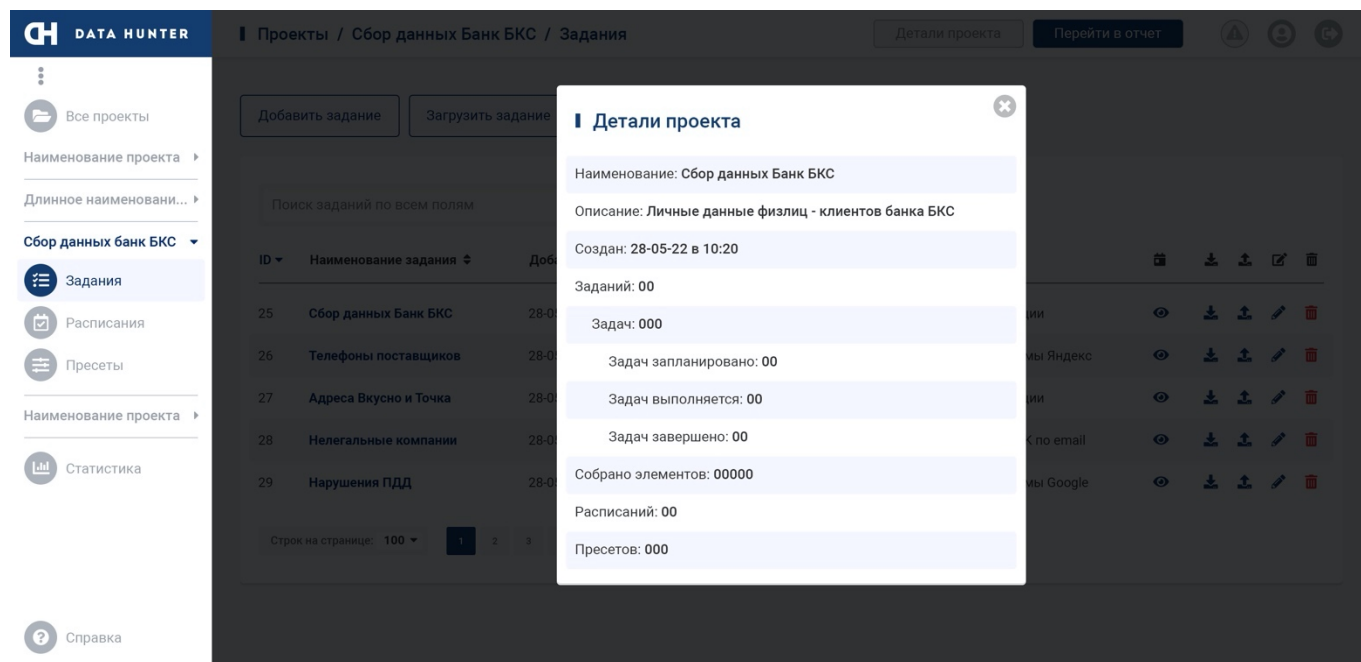
В таблице предусмотрена возможность фильтрации полей по каждому из столбцов в алфавитном порядке в обе стороны по нажатию на иконку фильтрации (вверх/вниз) справа от наименования/иконки поля в верхней строке.

Над таблицей расположено поле поиска проектов по всем полям.

Под таблицей расположены:

- Форма настройки отображения количества строк в таблице.
- Постраничная навигация по таблице.

6.6.5. Детали проекта



Раздел «Детали проекта» доступен по нажатию на кнопку «Детали проекта» из верхней панели личного кабинета Аналитика данных в разделах: Задания / Расписания / Справочники.

Раздел «Детали проекта» личного кабинета Аналитика данных состоит из полей:

- Наименование.
- Описание.
- Дата создания.
- Заданий.
- Задач.
- Задач запланировано.
- Задач выполняется.
- Задач завершено.
- Собрано элементов.
- Расписаний.
- Справочников.

6.6.6. Задания

Добавить задание Загрузить задание

Поиск заданий по всем полям Найти

ID	Наименование задания	Добавлено	Описание задания	Краулер	
25	Сбор данных Банк БКС	28-05-22 10:20	Краткое описание задания	Телеграм для демонстрации	
26	Телефоны поставщиков	28-05-22 10:20	Краткое описание задания	Краулер поисковой системы Яндекс	
27	Адреса Вкусно и Точка	28-05-22 10:20	Краткое описание задания	Телеграм для демонстрации	
28	Нелегальные компании	28-05-22 10:20	Краткое описание задания	Краулер поиска в БД CDEK по email	
29	Нарушения ПДД	28-05-22 10:20	Краткое описание задания	Краулер поисковой системы Google	

Строк на странице: 100 1 2 3 > >> В конец

Раздел «Задания» личного кабинета Аналитика данных состоит из компонентов:

- Кнопка «Добавить задание» – открытие диалогового окна с добавлением нового задания.
- Кнопка «Загрузить задание» – загрузка задания из файла в формате csv.
- Таблица заданий – таблица созданных заданий.

Таблица заданий состоит из полей:

- ID – идентификационный номер задания, автоматически присваиваемый системой при создании.
- Наименование – наименование задания, установленное Аналитиком данных при создании.
- Добавлено – дата добавления задания.
- Описание – текстовое описание задания, установленное Аналитиком данных при создании.
- Краулер – краулер, используемый для выполнения задания.
- Расписание – иконка глаза, по нажатию на которую доступен просмотр расписания выполнения задания.
- Импорт – импорт (загрузка) задания из файла в формате csv.
- Экспорт – экспорт (выгрузка) задания в файле в формате csv.
- Редактирование – диалоговое окно с редактированием задания.
- Удаление – удаление добавленного задания из системы.

В таблице предусмотрена возможность фильтрации полей по каждому из столбцов в алфавитном порядке в обе стороны по нажатию на иконку фильтрации (вверх/вниз) справа от наименования/иконки поля в верхней строке.

Над таблицей расположено поле поиска задания по всем полям.

Под таблицей расположены:

- Форма настройки отображения количества строк в таблице.
- Постраничная навигация по таблице.

6.6.6.1. Добавление задания вручную

Диалоговое окно «Новое задание» доступно из раздела «Задания» по нажатию на кнопку «Добавить задание».

Для добавления нового задания вручную нужно вызвать диалоговое окно «Новое задание» и пройти три этапа:

- Идентификация, выбор краулера.
- Формирование задач.
- Расписание.

Новое задание, шаг 1: Идентификация, выбор краулера

The screenshot shows the 'DATA HUNTER' interface. On the left is a sidebar with navigation items: 'Все проекты', 'Наименование проекта', 'Длинное наименование...', 'Сбор данных банк БКС', 'Задания' (selected), 'Расписания', 'Пресеты', 'Наименование проекта', 'Статистика', and 'Справка'. The main area displays a table of tasks with columns 'ID' and 'Наименование задания'. A modal dialog titled 'Новое задание: шаг #1. Идентификация, краулер.' is open, showing a progress bar with steps 1, 2, and 3. Step 1 is active. The dialog contains the following fields: 'Наименование задания*', 'Описание задания*', and 'Краулер*'. There is also a checkbox labeled 'Автоматическое формирование задач из пресетов'. At the bottom are buttons 'Сохранить и перейти к настройкам задач' and 'Отмена'.

На первом шаге добавления задания вручную в диалоговом окне «Новое задание: шаг #1. Идентификация, выбор краулера.» нужно выполнить следующие действия:

- Указать наименование задания.
- Указать описание задания.
- Выбрать краулер для выполнения задания.

- Сохранить данные и перейти к настройкам задач по нажатию на кнопку «Сохранить и перейти к настройкам задач».

Все действия в диалоговом окне «Новый краулер: шаг #1. Идентификация, спайдер.» обязательны для выполнения.

Новое задание, шаг 2: Формирование задач

The screenshot shows the 'DATA HUNTER' interface. A modal window titled 'Новое задание: шаг #2. Формирование задач.' is open. It features a progress indicator with three steps: 1 (selected), 2 (current), and 3. Below the progress bar is a search bar labeled 'Поиск задач' with a 'Найти' button. The main area contains a table with two columns: 'Наименование параметра' and 'Значение параметра'. The table lists parameters for two tasks: 'Поисковый запрос' and 'Дата поисковой выдачи'. The first task has values '5' and 'Вчерашний день' respectively. The second task has empty fields. At the bottom of the table is a 'Добавить новую задачу' button. Below the table are three buttons: 'Сохранить и перейти к выбору расписания', 'Вернуться назад', and 'Отмена'.

Наименование параметра	Значение параметра
Поисковый запрос	5
Дата поисковой выдачи	Вчерашний день
Поисковый запрос	
Дата поисковой выдачи	

На втором шаге добавления задания вручную в диалоговом окне «Новое задание: шаг #2. Формирование задач.» нужно выполнить следующие действия:

- Сформировать задачи.
- Сохранить данные и перейти к выбору расписания по нажатию на кнопку «Сохранить и перейти к выбору расписания».

Формирование задач при добавлении задания вручную

Формирование задач при добавлении задания вручную – это процесс установки значений параметров запуска краулера, выбранного для выполнения текущего задания, для однократного запуска.

Значения параметров запуска краулера устанавливаются в таблице задач (параметров запуска краулера) в диалоговом окне «Новое задание: шаг #2. Формирование задач.» на втором шаге добавления задания вручную в разделе «Задания».

Таблица задач (параметров запуска краулера) состоит из полей:

- **Наименование параметра** – наименование параметра запуска краулера.

- **Значение параметра** – значение параметра запуска краулера. Поле предусматривает генератор значений, который дает возможность сгенерировать значение параметра краулера в зависимости от дополнительных условий.
- **Удаление** – удаление параметра запуска краулера.

В таблице предусмотрена возможность фильтрации полей по каждому из столбцов в алфавитном порядке в обе стороны по нажатию на иконку фильтрации (вверх/вниз) справа от наименования/иконки поля в верхней строке.

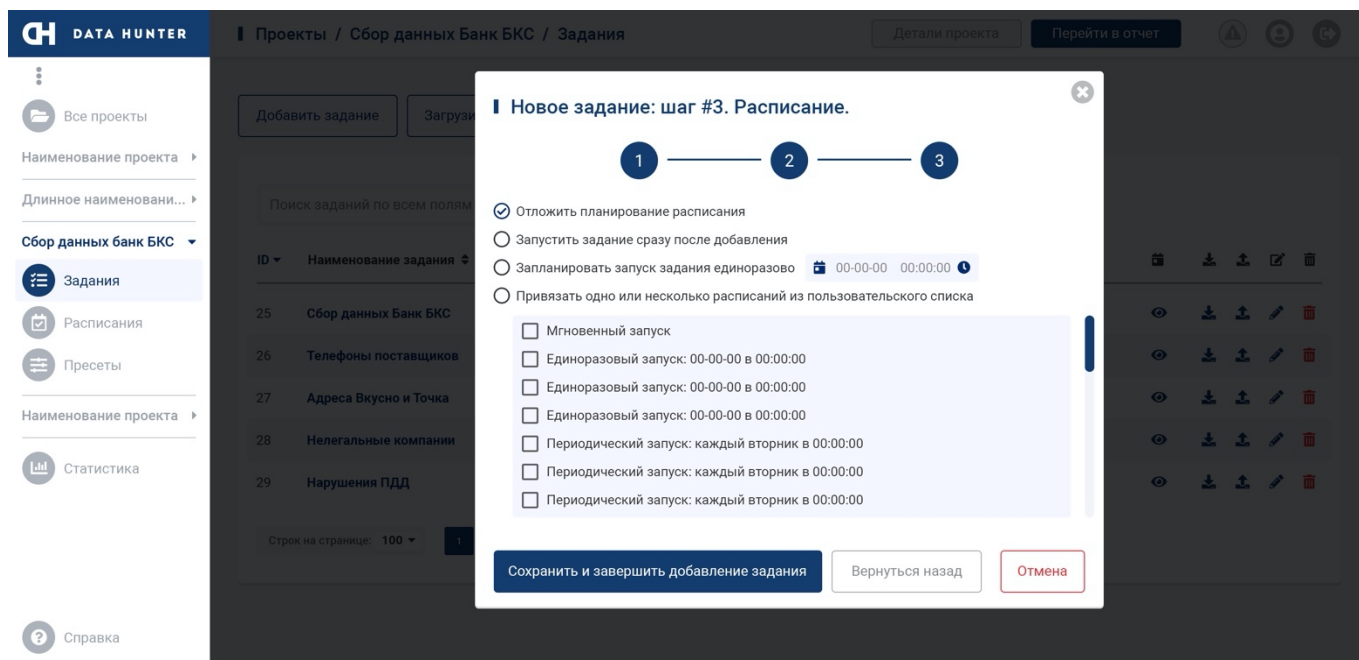
Над таблицей расположено поле поиска задач (параметров запуска краулера и их значений).

Под таблицей расположена форма добавления новой задачи, состоящая из полей параметров запуска краулера и полей их значений, которые нужно заполнить.

Новая задача добавляется по нажатию на кнопку «Добавить новую задачу» под таблицей задач.

Формирование задач не является обязательным для выполнения перехода к следующему шагу. По умолчанию задачи наследуют значения параметров запуска краулера.

Новое задание, шаг 3: Расписание



На третьем шаге добавления задания вручную в диалоговом окне «Новое задание: шаг #3. Расписание.» нужно выполнить следующие действия:

- Определить расписание выполнения задач.

- Сохранить данные и завершить добавление задания по нажатию на кнопку «Сохранить и завершить добавление задания».

Определение расписания выполнения задач

Расписание или расписания выполнения задач устанавливаются в диалоговом окне «Новое задание: шаг #3. Расписание.» на третьем шаге добавления задания вручную в разделе «Задания» выбором одного из методов:

- **Отложить планирование расписание** – (значение по умолчанию) шаг установки расписания для выполнения задач пропускается. Расписание может быть установлено позднее либо из формы редактирования задания, либо из раздела «Расписания».
- **Запуск задания сразу после добавления** – мгновенный запуск выполнения задач сразу после добавления задания.
- **Запланировать запуск задания единоразово** – установка интервала времени для единоразового выполнения задач.
- **Привязать одно или несколько расписаний из пользовательского списка** – установка одного или нескольких пользовательских расписаний для выполнения задач, добавленных ранее Аналитиком данных в разделе «Расписания».

6.6.6.2. Добавление задания с функцией автоматического формирования задач из справочника

Диалоговое окно «Новое задание» доступно из раздела «Задания» по нажатию на кнопку «Добавить задание».

Для добавления нового задания с функцией автоматического формирования задач из справочника нужно вызвать диалоговое окно «Новое задание» и пройти четыре этапа:

- Идентификация, выбор краулера.
- Настройка справочников
- Формирование задач.
- Расписание.

Новое задание, шаг 1: Идентификация, выбор краулера

Скриншот интерфейса системы «DATA HUNTER» с открытым диалогом «Новое задание: шаг #1. Идентификация, краулер.»

Диалог содержит следующие элементы:

- Панель прогресса с 4 шагами, первый из которых активен.
- Поле «Наименование задания*».
- Поле «Описание задания*».
- Выбор «Краулер*».
- Чекбокс «Автоматическое формирование задач из пресетов» (активен).
- Кнопки «Сохранить и перейти к выбору пресетов» и «Отмена».

На заднем плане видна панель навигации и таблица с заданиями:

ID	Наименование задания
25	Сбор данных Банк БКС
26	Телефоны поставщиков
27	Адреса Вкусно и Точка
28	Нелегальные компании
29	Нарушения ПДД

На первом шаге добавления задания с функцией автоматического формирования задач из справочника в диалоговом окне «Новое задание: шаг #1. Идентификация, выбор краулера.» нужно выполнить следующие действия:

- Указать наименование задания.
- Указать описание задания.
- Выбрать краулер для выполнения задания.
- Активировать функцию автоматического формирования задач из справочника.
- Сохранить данные и перейти к настройкам справочников по нажатию на кнопку «Сохранить и перейти к выбору справочников».

Все действия в диалоговом окне «Новый краулер: шаг #1. Идентификация, спайдер.» обязательны для выполнения.

Новое задание, шаг 2: Настройка справочников

Скриншот интерфейса «DATA HUNTER» с диалоговым окном «Новое задание: шаг #2. Настройка пресетов».

В диалоговом окне отображается прогресс-бар с четырьмя шагами (1, 2, 3, 4), где шаг 2 является активным.

В центре окна находится таблица параметров:

Наименование параметра	Метод ввода значений	Пресет / значение
Поисковый запрос	Использовать пресет	Выбрать пресет
Группа в Telegram	Использовать пресет	Мои группы Telegram
Дата поисковой выдачи	Ручной ввод значения	Вчерашний день
Содержит вложения	Ручной ввод значения	true

В нижней части диалогового окна расположены три кнопки: «Сохранить и сформировать задачи», «Вернуться назад» и «Отмена».

На втором шаге добавления задания с функцией автоматического формирования задач из справочника в диалоговом окне «Новое задание: шаг #2. Настройка справочников.» нужно выполнить следующие обязательные действия:

- Настроить справочники.
- Сохранить данные и перейти к формированию задач по нажатию на кнопку «Сохранить и сформировать задачи».

Настройка справочников при добавлении задания с автоматическим формированием задач

Настройка справочников при добавлении задания с функцией автоматического формирования задач из справочника – это процесс выбора справочников для значений параметров перед формированием задач в таблице параметров.

Таблица параметров состоит из полей:

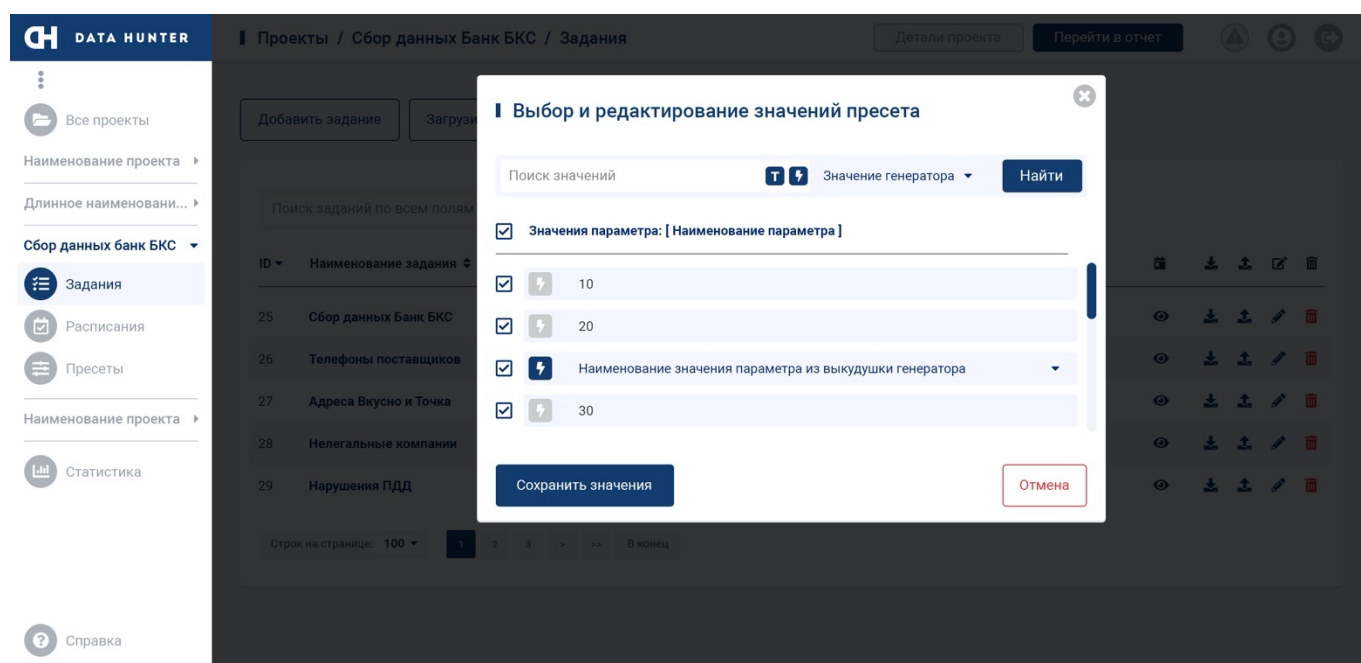
- **Наименование параметра** – наименование параметра выбранного краулера.
- **Метод ввода значений** – ручной / ручной с генератором значений / из справочника / из справочника с генератором значений.
- **Справочник / значение** – значения параметра: введенные вручную / сгенерированные / полученные из справочника.
- **Количество значений** – удаление параметра запуска краулера.

В таблице предусмотрена возможность фильтрации полей по каждому из столбцов в алфавитном порядке в обе стороны по нажатию на иконку фильтрации (вверх/вниз) справа от наименования/иконки поля в верхней строке.

Над таблицей расположено поле поиска параметров.

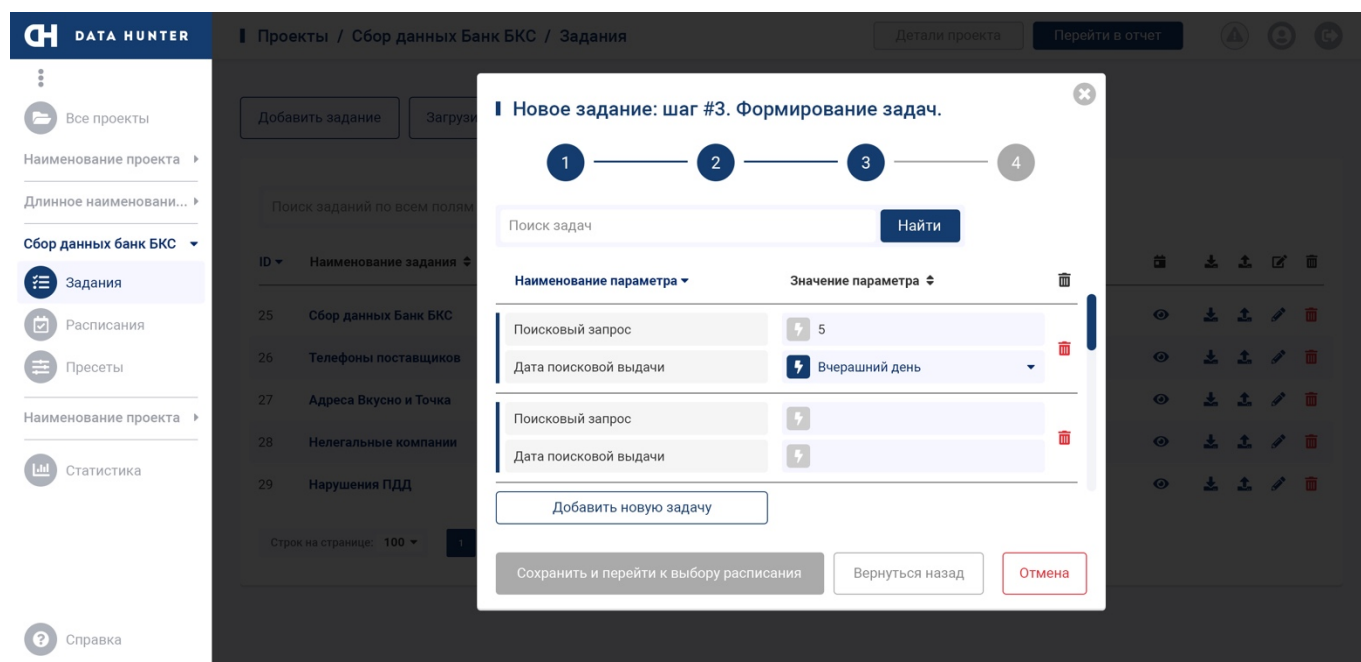
Для добавления значений параметра из справочника в таблице параметров напротив нужного параметра должен быть установлен метод ввода значения «Использовать справочник», после чего в поле «Справочник / значение» активируется возможность выбора справочника. После выбора справочника из списка допустимых значения справочника можно отредактировать по нажатию на иконку карандаша.

Диалоговое окно «Выбор и редактирование значений справочника»



Диалоговое окно «Выбор и редактирование значений справочника» позволяет выбрать и установить (отредактировать) значения по выбранному параметру из таблицы значений параметров, предусмотренных справочником.

Новое задание, шаг 4: Формирование задач



На третьем шаге добавления задания с функцией автоматического формирования задач из справочника в диалоговом окне «Новое задание: шаг #3. Формирование задач.» нужно выполнить следующие действия:

- Сформировать задачи.
- Сохранить данные и перейти к выбору расписания по нажатию на кнопку «Сохранить и перейти к выбору расписания».

Формирование задач при добавлении задания с функцией автоматического формирования задач

Формирование задач при добавлении задания с функцией автоматического формирования задач из справочника — это автоматический процесс генерации значений параметров запуска краулера с возможностью ручного редактирования.

Значения параметров запуска краулера генерируются в таблице задач (параметров запуска краулера) в автоматическом режиме после выбора и настройки значений параметров из справочников в диалоговом окне «Новое задание: шаг #3. Формирование задач.» на третьем шаге добавления задания с функцией автоматического формирования задач из справочника в разделе «Задания».

Сгенерированные значения параметров запуска краулера можно отредактировать вручную.

Таблица задач (параметров запуска краулера) состоит из полей:

- **Наименование параметра** — наименование параметра запуска краулера.

- **Значение параметра** – значение параметра запуска краулера. Поле предусматривает генератор значений, который дает возможность сгенерировать значение параметра краулера в зависимости от дополнительных условий.
- **Удаление** – удаление параметра запуска краулера.

В таблице предусмотрена возможность фильтрации полей по каждому из столбцов в алфавитном порядке в обе стороны по нажатию на иконку фильтрации (вверх/вниз) справа от наименования/иконки поля в верхней строке.

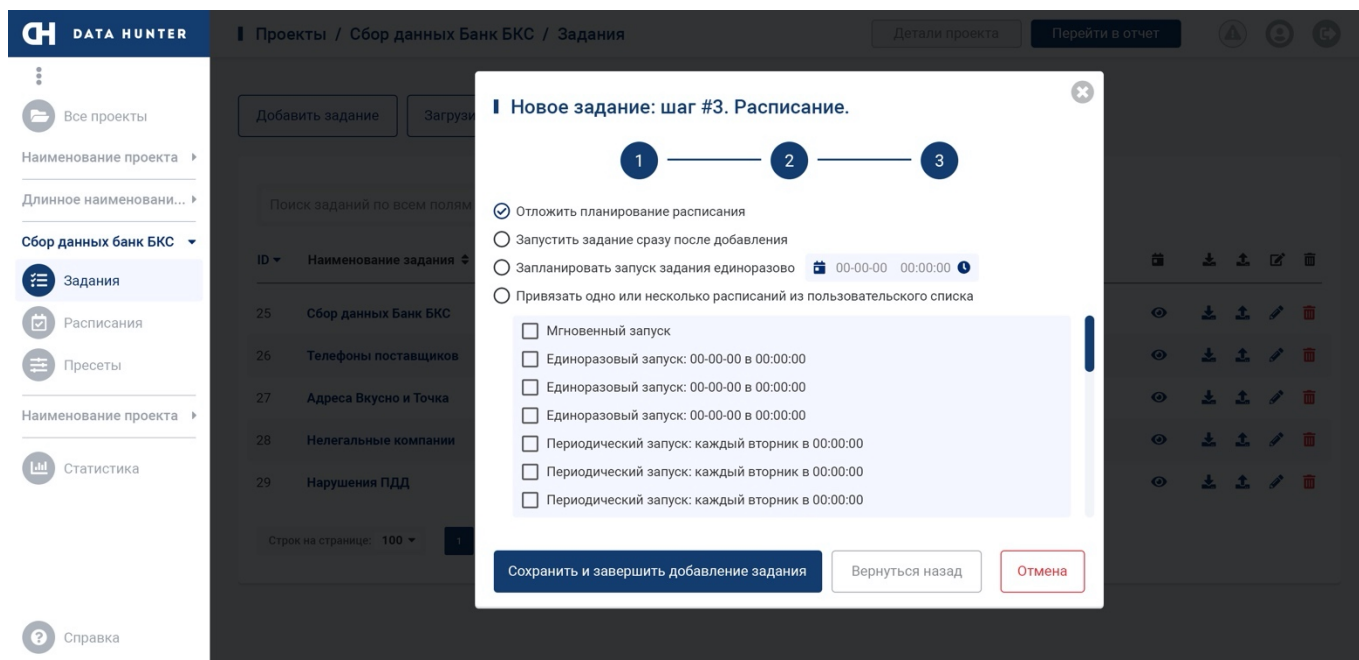
Над таблицей расположено поле поиска задач (параметров запуска краулера и их значений).

Под таблицей расположена форма добавления новой задачи, состоящая из полей параметров запуска краулера и полей их значений, которые нужно заполнить.

Новая задача добавляется по нажатию на кнопку «Добавить новую задачу» под таблицей задач.

Формирование задач не является обязательным для выполнения перехода к следующему шагу. По умолчанию задачи генерируются автоматически.

Новое задание, шаг 4: Расписание



На четвертом шаге добавления задания с функцией автоматического формирования задач из справочника в диалоговом окне «Новое задание: шаг #4. Расписание.» нужно выполнить следующие действия:

- Определить расписание выполнения задач.

- Сохранить данные и завершить добавление задания по нажатию на кнопку «Сохранить и завершить добавление задания».

Определение расписания выполнения задач

Расписание или расписания выполнения задач устанавливаются в диалоговом окне «Новое задание: шаг #4. Расписание.» на четвертом шаге добавления задания с функцией автоматического формирования задач из справочника в разделе «Задания» выбором одного из методов:

- **Отложить планирование расписание** – (значение по умолчанию) шаг установки расписания для выполнения задач пропускается. Расписание может быть установлено позднее либо из формы редактирования задания, либо из раздела «Расписания».
- **Запуск задания сразу после добавления** – мгновенный запуск выполнения задач сразу после добавления задания.
- **Запланировать запуск задания единоразово** – установка интервала времени для единоразового выполнения задач.
- **Привязать одно или несколько расписаний из пользовательского списка** – установка одного или нескольких пользовательских расписаний для выполнения задач, добавленных ранее Аналитиком данных в разделе «Расписания».

6.6.6.3. Действия с заданиями: редактирование, удаление

Редактирование задания

Диалоговое окно «Редактирование задания» вызывается из таблицы заданий в разделе «Задания» по нажатию на иконку карандаша в строке нужного задания. Изменения, внесенные в задание через диалоговое окно «Редактирование задания», подтверждаются кнопкой «Сохранить изменения».

Удаление задания

Задание может быть удален из таблицы заданий в разделе «Задания» по нажатию на иконку корзины в строке нужного задания.

6.6.7. Расписания

Добавить расписание

Поиск расписаний по наименованию и описанию

ID	Наименование расписания	Добавлено	Описание расписания	Статус	Периодичность			
25	Наименование расписания #1	28-05-22 10:20	Краткое описание расписания	Готово к выполнению	28-05-22 10:20:00			
26	Наименование расписания #2	28-05-22 10:20	Краткое описание расписания	Готово к выполнению	в 20.30			
27	Наименование расписания #3	28-05-22 10:20	Краткое описание расписания	Готово к выполнению	28-05-22 10:20:00			
28	Наименование расписания #4	28-05-22 10:20	Краткое описание расписания	Новое	28-05-22 10:20:00			
29	Наименование расписания #5	28-05-22 10:20	Краткое описание расписания	Готово к выполнению	28-05-22 10:20:00			

Строк на странице: 100 1 2 3 > >> В конец

Раздел «**Расписания**» личного кабинета Аналитика данных состоит из компонентов:

- Кнопка «**Добавить расписание**» – открытие диалогового окна с добавлением нового расписания.
- **Таблица расписаний** – таблица созданных заданий.

Таблица расписаний состоит из полей:

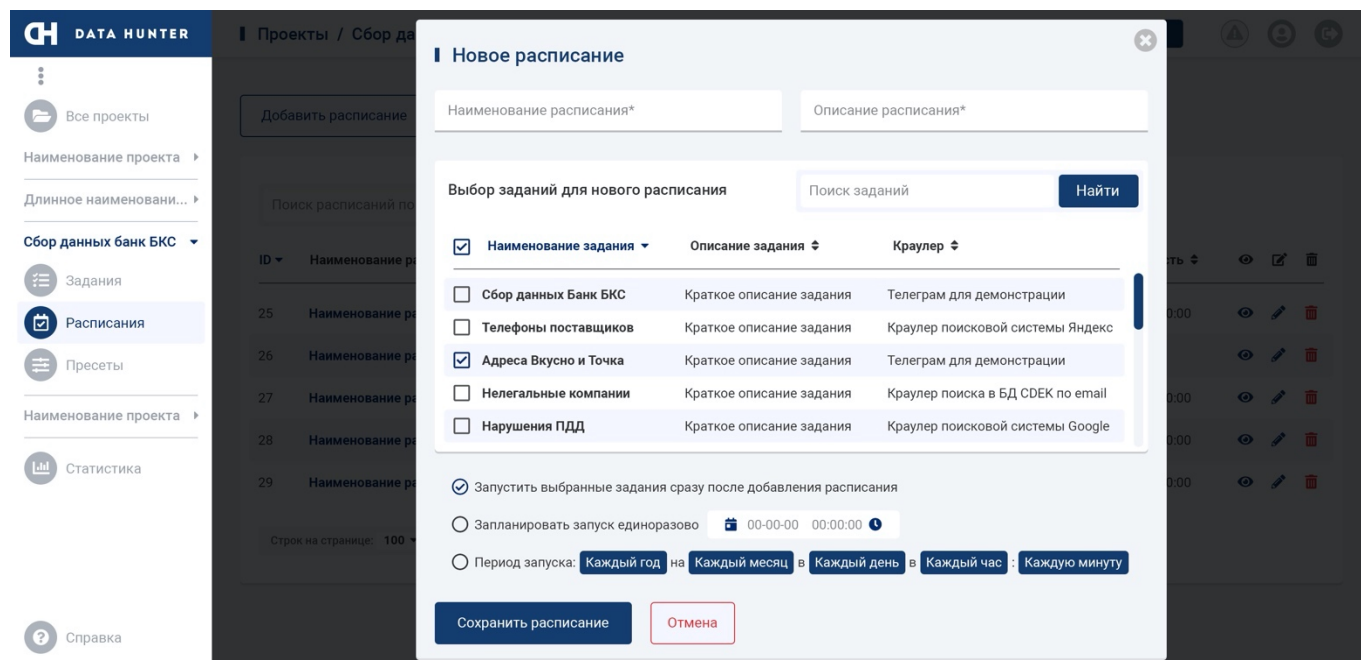
- **ID** – идентификационный номер расписания, автоматически присваиваемый системой при создании.
- **Наименование** – наименование расписания, установленное Аналитиком данных при создании.
- **Добавлено** – дата добавления расписания.
- **Описание** – текстовое описание расписания, установленное Аналитиком данных при создании.
- **Статус** – статус расписания со значениями: готово к выполнению / новое.
- **Периодичность** – периоды времени выполнения задач.
- **Просмотр** – просмотр таблицы выполнения задач.
- **Редактирование** – диалоговое окно с редактированием расписания.
- **Удаление** – удаление добавленного расписания из системы.

Над таблицей расположено поле поиска расписаний по наименованию и расписанию.

Под таблицей расположены:

- Форма настройки отображения количества строк в таблице.
- Постраничная навигация по таблице.

6.6.7.1. Добавление расписания



Диалоговое окно «Новое расписание» доступно из раздела «Расписания» по нажатию на кнопку «Добавить расписание».

Для добавления нового справочника нужно вызвать диалоговое окно «Новое расписание» и выполнить следующие обязательные действия:

- Указать наименование расписания.
- Указать описание расписания.
- Выбрать задания из таблицы «Выбор задания для нового расписания». Доступен множественный выбор.
- Определить метод запуска выполнения задач:
 - Запустить выбранные задания сразу после добавления расписания;
 - Запланировать запуск единоразово (с указанием периода выполнения задач);
 - Установить период запуска выполнения задач.
- Сохранить расписание по нажатию на кнопку «Сохранить расписание».

6.6.7.2. Действия с расписаниями: редактирование, удаление

Редактирование расписания

Диалоговое окно «Редактирование расписания» вызывается из таблицы расписаний в разделе «Расписания» по нажатию на иконку карандаша в строке нужного расписания. Изменения, внесенные в расписание через диалоговое окно «Редактирование расписания», подтверждаются кнопкой «Сохранить изменения».

Удаление расписания

Расписание может быть удален из таблицы расписаний в разделе «Расписания» по нажатию на иконку корзины в строке нужного расписания.

6.6.8. Справочники

Детали проекта | Перейти в отчет

Добавить пресет | Загрузить пресет

Поиск пресетов по наименованию и описанию | Найти

ID	Наименование пресета	Создан	Описание пресета	Краулер	Параметр краулера		
25	Наименование пресета #1	28-05-22 10:20	Краткое описание пресета	Краулер Telegram	Ссылка на канал для поиска		
26	Наименование пресета #2	28-05-22 10:20	Краткое описание пресета	Краулер Telegram	Поиск с первого или последнего соо...		
27	Наименование пресета #3	28-05-22 10:20	Краткое описание пресета	Краулер Telegram	Категория поиска		
28	Наименование пресета #4	28-05-22 10:20	Краткое описание пресета	Краулер Telegram	Категория поискового запроса		
29	Наименование пресета #5	28-05-22 10:20	Краткое описание пресета	Краулер Telegram	Поисковый запрос		

Строк на странице: 100 | 1 | 2 | 3 | > | >> | В конец

Раздел «Справочники» личного кабинета Аналитика данных состоит из компонентов:

- Кнопка «Добавить справочник» – открытие диалогового окна с добавлением нового справочника.
- Кнопка «Загрузить справочник» – загрузка справочника из файла в формате csv.
- Таблица справочников – таблица созданных справочников.

Таблица справочников состоит из полей:

- ID – идентификационный номер справочника, автоматически присваиваемый системой при создании.
- Наименование – наименование справочника, установленное Аналитиком данных при создании.
- Добавлено – дата добавления справочника.
- Описание – текстовое описание справочника, установленное Аналитиком данных при создании.
- Краулер – краулер, параметр запуска которого используется в справочнике для формирования значений.
- Параметр – параметр запуска краулера, из значений которого формируется справочник.
- Редактирование – диалоговое окно с редактированием справочника.
- Удаление – удаление добавленного справочника из системы.

Над таблицей расположено поле поиска справочников по наименованию и расписанию.

Под таблицей расположены:

- Форма настройки отображения количества строк в таблице.
- Постраничная навигация по таблице.

6.6.8.1. Добавление справочника

Диалоговое окно «Новый справочник» доступно из раздела «Справочники» по нажатию на кнопку «Добавить справочник».

Для добавления нового справочника нужно вызвать диалоговое окно «Новый справочник» и пройти два этапа:

- Идентификация, выбор краулера, выбор параметра.
- Значения параметра.

Новый справочник, шаг 1: Идентификация, выбор краулера, выбор параметра

The screenshot shows the 'DATA HUNTER' interface. On the left is a sidebar with navigation links: 'Все проекты', 'Наименование проекта', 'Длинное наименовани...', 'Сбор данных банк БКС', 'Задания', 'Расписания', 'Пресеты' (highlighted), 'Наименование проекта', 'Статистика', and 'Справка'. The main area displays a table of presets with columns 'ID' and 'Наименование пресета'. A modal dialog is open in the center, titled 'Новый пресет: шаг #1. Идентификация, краулер, параметр.'. The dialog has a progress indicator with steps 1 and 2. It contains the following fields: 'Наименование пресета*' (text input), 'Описание пресета*' (text input), 'Краулер*' (dropdown menu), and 'Параметр*' (dropdown menu). Below these are 'Значений:' with a value of '1' and a checkbox for 'Тип значений: Генератор'. At the bottom are two buttons: 'Сохранить и перейти к значениям' and 'Отмена'.

На первом шаге добавления справочника в диалоговом окне «Новый справочник: шаг #1. Идентификация, краулер, параметр.» нужно выполнить следующие действия:

- Указать наименование справочника.
- Указать описание справочника.
- Выбрать краулер, из значений параметра которого будет формироваться справочник.

- Выбрать параметр выбранного краулера, из значений которого будет формироваться справочник.
- Установить количество значений выбранного параметра.
- (Опционально) Задать тип значений выбранного параметра «Генератор».
- Сохранить данные и перейти к значениям параметра по нажатию на кнопку «Сохранить и перейти к значениям».

Все действия в диалоговом окне «Новый справочник: шаг #1. Идентификация, краулер, параметр.» обязательны для выполнения, кроме выбора типа значения «Генератор».

Новый справочник, шаг 2: Значения параметра

Скриншот интерфейса приложения DATA HUNTER. В центре экрана открыто диалоговое окно «Новый пресет: шаг #2. Значения параметра.». В окне видна панель прогресса с шагами 1 и 2, где шаг 2 активен. Вверху диалогового окна находится поле «Поиск значений» и кнопка «Найти». Ниже расположена таблица «Значения параметра: [Наименование параметра]». В таблице три строки: первая и вторая содержат значение «5», третья — выпадающий список с текстом «Наименование значения параметра из выкудушки генератора». Каждая строка имеет иконку молнии слева и кнопку удаления справа. Внизу диалогового окна расположены кнопки: «Добавить новое значение», «Сохранить и завершить добавление пресета», «Вернуться назад» и «Отмена».

На втором шаге добавления справочника в диалоговом окне «Новый справочник: шаг #2. Значения параметра.» нужно выполнить следующие действия:

- Установить значения параметра.
- Сохранить данные и завершить добавление справочника по нажатию на кнопку «Сохранить и завершить добавление справочника».

Установка значений параметра

Значения параметра устанавливаются в таблице значений параметра в диалоговом окне «Новый справочник: шаг #2. Значения параметра.» на втором шаге добавления справочника в разделе «Справочники».

Таблица значений параметра состоит из полей:

- **Тип значения** — иконка молнии, активная, если тип значения параметра «Генератор».

- **Значение параметра** – значение параметра запуска краулера.
- **Удаление** – удаление параметра запуска краулера.

В таблице предусмотрена возможность фильтрации полей по каждому из столбцов в алфавитном порядке в обе стороны по нажатию на иконку фильтрации (вверх/вниз) справа от наименования/иконки поля в верхней строке.

Над таблицей расположено поле поиска значений параметра.

Под таблицей расположена форма добавления нового значения, состоящая из поля значения, которое нужно заполнить.

Новое значение добавляется по нажатию на кнопку «Добавить новое значение» под таблицей задач.

Установка значений параметра является обязательным для завершения добавления справочника.

6.6.8.2. Действия со справочниками: редактирование, удаление

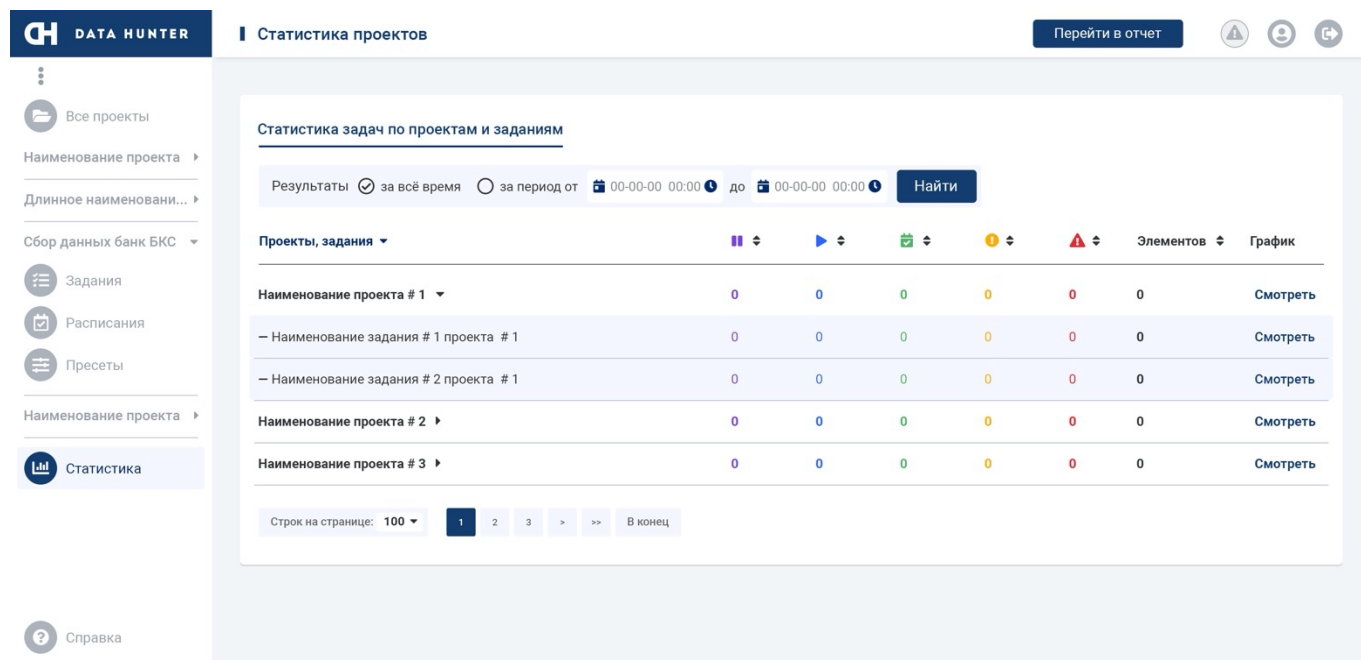
Редактирование справочника

Диалоговое окно «Редактирование справочника» вызывается из таблицы справочников в разделе «Справочники» по нажатию на иконку карандаша в строке нужного справочника. Изменения, внесенные в справочник через диалоговое окно «Редактирование справочника», подтверждаются кнопкой «Сохранить изменения».

Удаление справочника

Справочник может быть удален из таблицы справочников в разделе «Справочники» по нажатию на иконку корзины в строке нужного справочника.

6.6.9. Статистика



Раздел «Статистика проектов» личного кабинета Аналитика данных состоит из таблицы: «Статистика задач по проектам и заданиям».

Таблица «Статистика задач по проектам и заданиям» состоит из полей:

- **Проекты, задания** – наименование проекта с декомпозицией на задания на втором уровне.
- **Количество запланированных задач** – числовое значение запланированных задач.
- **Количество выполняющихся задач** – числовое значение выполняющихся задач.
- **Количество завершенных задач** – числовое значение завершенных задач.
- **Количество завершенных задач с предупреждениями** – числовое значение завершенных задач с предупреждением о возможных ошибках.
- **Количество завершенных задач с ошибками** – числовое значение задач, завершенных с ошибкой.
- **Количество собранных элементов** – числовое значение собранных на хосте элементов.
- **График динамики сбора элементов** – всплывающее окно с графиком сбора элементов.

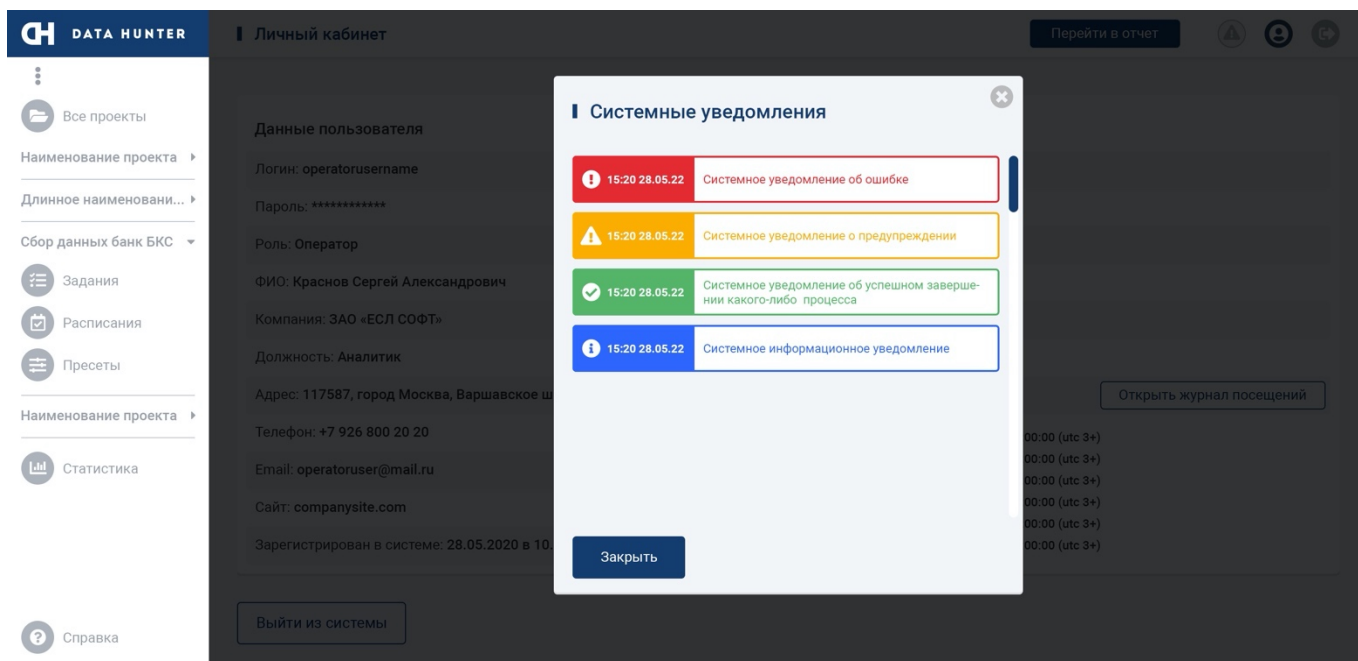
В таблице предусмотрена возможность фильтрации полей по каждому из столбцов в алфавитном порядке в обе стороны по нажатию на иконку фильтрации (вверх/вниз) справа от наименования/иконки поля в верхней строке.

Над таблицей расположена форма фильтрации данных таблицы с возможностью выдачи результатов либо за все время, либо за установленный период.

Под таблицей расположены:

- Форма настройки отображения количества строк в таблице.
- Постраничная навигация по таблице.

6.6.10. Журнал регистрации событий



Раздел «Журнал регистрации событий» доступен по нажатию на иконку с восклицательным знаком: «Журнал регистрации событий» из верхней панели личного кабинета Аналитика данных.

Статусы записей журнала регистрации событий:

- **Красный** – системное уведомление об ошибке.
- **Желтый** – системное уведомление с предупреждением.
- **Зеленый** – системное уведомление об успешном завершении какого-либо процесса / действия.
- **Синий** – нейтральное системное уведомление.

Системные уведомления появляются в виде всплывающих окон во время сессии Администратора в личном кабинете, информируя Аналитика данных о том или ином событии. Системные уведомления, полученные вне сессии Аналитика данных в личном кабинете доступны в журнале регистрации событий.

6.6.11. Отчеты

Результаты поиска

Скачать в .csv

Скачать выбранные поля

Все краулеры: 2068

Yandex: 680

Google: 920

Telegram: 385

Database: 83

Строк на странице: 100

Поиск результатов

☒ По всем полям
 ☒ По запросам
 ☒ По результатам

От 28-05-22 10:20

до 28-05-22 10:20

Найти

☐ ID	Дата добавления	Краулер	Запросы	Результаты	Доп. информация	Скачать	Просмотр
☐ 00022309	28-05-22 10:20:05.328	Yandex	Пассивный доход ...	https://plusworld.ru/pro ...	Как самозанятый может стать ...	Обновить	Просмотреть
☐ 00022308	28-05-22 10:20:05.326	Yandex	Пассивный доход ...	https://kursk.com/finance ...	Финансовая система страны мо ...	Скачать	Просмотреть
☐ 00022307	28-05-22 10:20:05.323	Google	Пассивный доход ...	https://smartmoney.one/n ...	Куда вложить деньги, если вы ...	Скачать	Просмотреть
☐ 00022306	28-05-22 10:20:05.322	Telegram	—	Новинка на блоге Royal ...	https://t.me/profvest/30924	—	—
☐ 00022305	28-05-22 10:20:05.320	Database	{cdek_email: 'sado ...	{_cdek_id: '244594', 'phone ...	Как самозанятый может стать ...	—	—

1

2

3

>

>>

В конец

Раздел «**Результаты поиска**» (Отчеты) личного кабинета Аналитика данных представлен таблицей результатов выполнения задач с возможностью экспорта таблицы как целиком по нажатию на кнопку «Скачать в .csv», так и по выбранным полям по нажатию на кнопку «Скачать выбранные поля».

Таблица результатов выполнения задач состоит из полей:

- **Чекбокс** – выбор полей для экспорта.
- **ID** – идентификационный номер результата, автоматически присваиваемый системой при получении.
- **Дата добавления** – дата получения результата.
- **Краулер** – краулер, участвовавший в выполнении задачи.
- **Запросы** – поисковые запросы.
- **Результаты** – текстовые данные в формате, предусмотренном краулером: текст, ссылка, код, прочее.
- **Доп. информация** – дополнительная текстовая информация к данным, в качестве которой может выступать текстовый пассаж, ссылка и иной формат, предусмотренный краулером.
- **Скачать** – возможность скачивания контента, если результат сбора данных – гиперссылка.
- **Просмотреть** – просмотр web-страницы, если результат сбора данных – гиперссылка.

В таблице предусмотрена возможность фильтрации полей по каждому из столбцов в алфавитном порядке в обе стороны по нажатию на иконку фильтрации (вверх/вниз) справа от наименования/иконки поля в верхней строке.

Над таблицей расположена форма поиска по всем полям / по запросам / по результатам с возможностью фильтрации данных:

- По краулерам.
- По времени выполнения задач с возможностью выдачи результатов либо за все время, либо за установленный период.

Под таблицей расположены:

- Форма настройки отображения количества строк в таблице.
- Постраничная навигация по таблице.

6.6.12. Профиль

Раздел «Профиль» доступен по нажатию на иконку пользователя: «Профиль» из верхней панели личного кабинета Аналитика данных. Раздел «Профиль» личного кабинета Аналитика данных состоит из компонентов:

- Блок «Данные пользователя».
- Блок «Статистика профайла».
- Кнопка выхода из системы «Выйти из системы».

Блок «Данные пользователя» профиля состоит из полей:

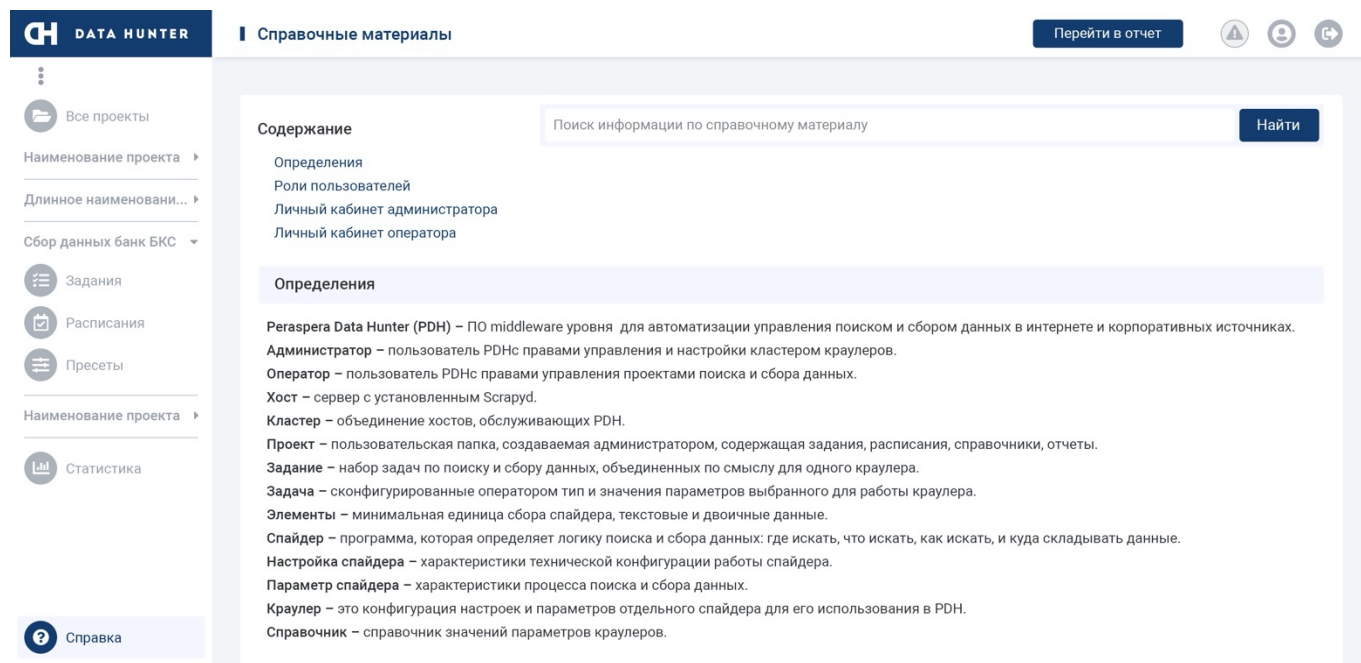
- Login.
- Пароль.
- Роль.

- ФИО.
- Компания.
- Должность.
- Адрес.
- Телефон.
- Email.
- Сайт.
- Дата регистрации в системе.

Блок «Статистика профайла» профиля состоит из полей:

- Проектов.
- Доступных краулеров.
- Добавлено заданий.
- Собрано элементов.
- Добавлено расписаний.
- Добавлено справочников.
- Рабочих сессий с возможность просмотра как последних записей, так и всего журнала посещений по нажатию на кнопку «Открыть журнал посещений».

6.6.13. Справочная информация



Раздел «Справочная информация» доступен в боковом меню, пункт меню «Справка». Раздел «Справочная информация» личного кабинета Аналитика данных состоит из компонентов:

- Поле поиска информации по справочным материалам.
- Содержание справочных материалов.

- Руководство пользователя.

6.7. Предустановленные краулеры

Первый релиз Peraspera DataHunter в базовой сборке на 01.01.2023 года предлагает 9 предустановленных краулеров.

Таблица предустановленных краулеров

#	Краулер	Функции	Элементы контента
1.	Яндекс-краулер	Поиск/сбор данных из результатов поиска Яндекса*	● Ссылка из поисковой выдачи
2.	Google-краулер	Поиск/сбор данных из результатов поиска Google*	● Ссылка из поисковой выдачи
3.	Telegram-краулер	Поиск/сбор данных из открытых каналов/групп Telegram*	● Сообщение из Telegram-канала ● Telegram-канал ● Внешняя ссылка в сообщении канала
4.	Вконтакте-краулер	Поиск/сбор данных из открытых профилей/групп VK*	● Сообщение из группы ● Ссылка в сообщении ● Сообщение со стены пользователя ● Комментарий к сообщению ● Группа ● Реакция на комментарий/сообщение ● Профиль пользователя
5.	Одноклассники-краулер	Сбор данных из открытых профилей/групп ОК*	● Группа ● Топики (сообщения) в группе
6.	Универсальный краулер сайтов	Сбор данных из открытых сайтов*	—
7.	Краулер для баз данных	Поиск/сбор данных из различных баз данных	● Строка таблицы базы данных
8.	Краулер для веб-страниц	Сбор данных из различных веб-таблиц	● Строка из веб-таблицы
9.	Краулер реестров	Сбор файлов с реестрами	● Файл с реестром

* По текстовым запросам