

Микросервисы

Сервис поиска и объединения дублей

Версия 8.0



Эта документация предоставляется с ограничениями на использование и защищена законами об интеллектуальной собственности. За исключением случаев, прямо разрешенных в вашем лицензионном соглашении или разрешенных законом, вы не можете использовать, копировать, воспроизводить, переводить, транслировать, изменять, лицензировать, передавать, распространять, демонстрировать, выполнять, публиковать или отображать любую часть в любой форме или посредством любые значения. Обратный инжиниринг, дизассемблирование или декомпиляция этой документации, если это не требуется по закону для взаимодействия, запрещены.

Информация, содержащаяся в данном документе, может быть изменена без предварительного уведомления и не может гарантировать отсутствие ошибок. Если вы обнаружите какие-либо ошибки, сообщите нам о них в письменной форме.

Содержание

Сервис поиска и объединения дублей	4
Общие принципы работы	4
Схема работы	4
Масштабируемость	5
Совместимость с продуктами Creatio	5
Варианты развертывания	6

Сервис поиска и объединения дублей



Основы

Сервис поиска и объединения дублей (Bulk Duplicate Search) используется для дедупликации (поиска и объединения) дублирующихся записей (дублей), которые могут появиться при добавлении данных в разделах приложения Creatio.

Общие принципы работы

В Creatio реализованы следующие **виды поиска дублей**:

- [Массовый поиск дублей](#) — поиск выполняется по всей базе. Запускается вручную либо автоматически.
- [Локальный поиск дублей](#) — поиск предусматривает проверку существования дублей для конкретной записи. Запускается при создании новой записи в момент ее сохранения в разделе.

Не запуская поиск, можно в ручном режиме выбрать дублирующие записи и выполнить их слияние. Такая опция доступна для всех разделов приложения.

Возможности сервиса поиска дублей:

- Использование преднастроенных правил поиска дублей в разделах [*Контрагенты*] ([*Accounts*]), [*Контакты*] ([*Contacts*]) и [*Лиды*] ([*Leads*]).
- Настройка пользовательских правил поиска дублей контактов, контрагентов и лидов в соответствующих разделах.
- Создание правил поиска дублей для любого раздела системы, в том числе пользовательского.

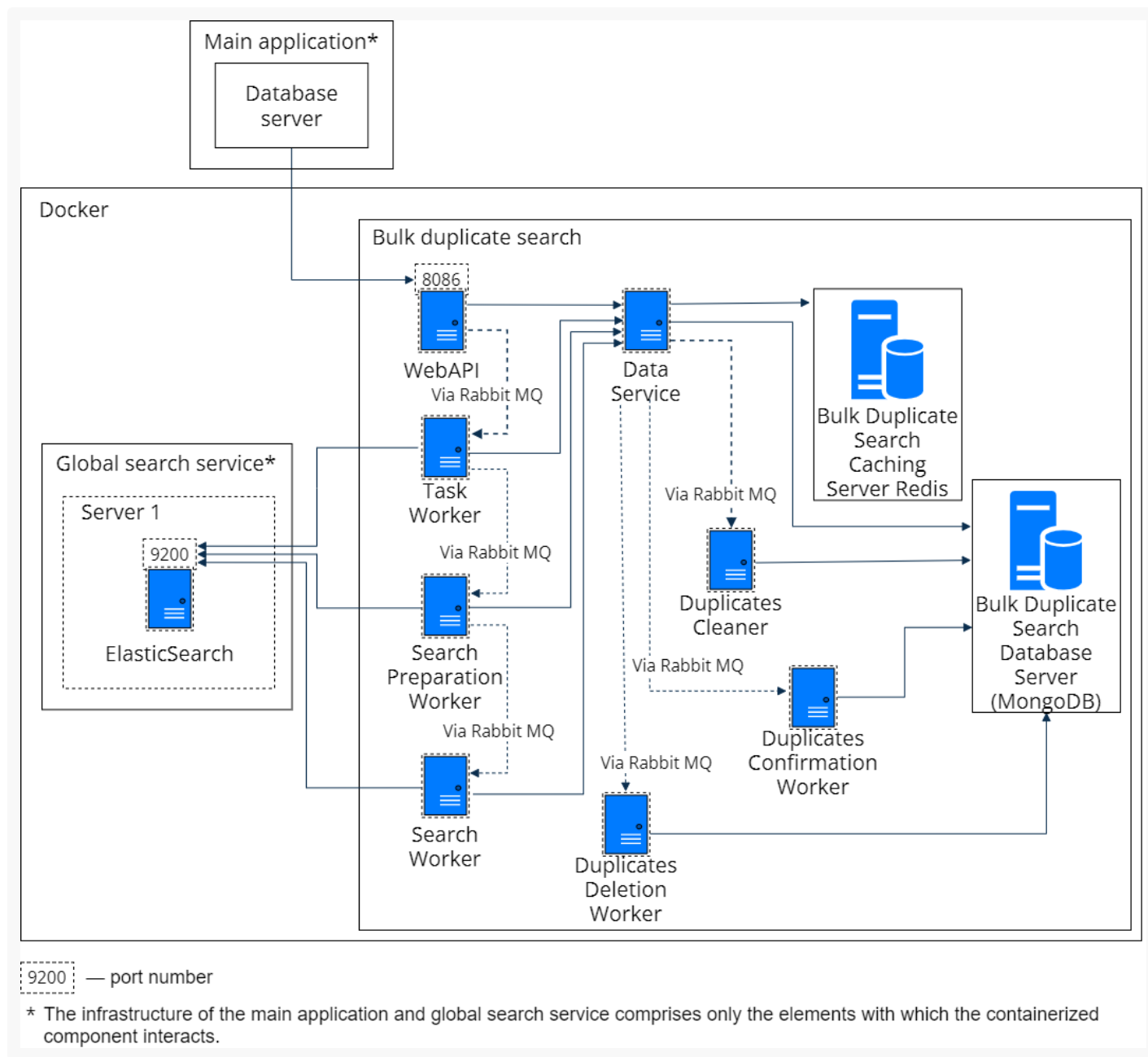
Схема работы

Компоненты сервиса поиска и объединения дублей:

- RabbitMQ — брокер сообщений. Компонент сервиса глобального поиска.
- ElasticSearch — поисковая система. Компонент сервиса глобального поиска.
- Redis — хранилище данных, которое используется для кэширования и быстрого действия.
- MongoDB — документоориентированная система управления базами данных.
- WebAPI — веб-сервис для общения в приложении Creatio.
- Data Service — внутренний сервис коммуникаций с компонентом MongoDB.
- Duplicates Search Worker — компонент поиска дублей.
- Duplicates Deletion Worker — компонент для точечного удаления дублей.
- Duplicates Confirmation Worker — компонент для группировки и фильтрации найденных дублей с учетом уникальности.
- Duplicates Cleaner — компонент очистки дублей.
- Deduplication Task Worker — компонент для постановки задачи дедупликации.

- Deduplication Preparation Worker — компонент для подготовки процесса дедупликации, который формирует запросы для поиска дублей согласно правилам.

Схема работы сервиса поиска и объединения дублей представлена ниже.



Масштабируемость

Использование кластера баз данных в крупных проектах позволяет выполнить масштабирование сервиса поиска и объединения дублей. Описание кластеризации содержится в [документации ElasticSearch](#).

Совместимость с продуктами Creatio

Сервис поиска и объединения дублей имеет версии 1.0-1.5 и 2.0, которые совместимы со **всеми продуктами Creatio** версий 7.14 и выше.

Варианты развертывания

Сервис поиска и объединения дублей можно развернуть on-site и cloud.

Для пользователей приложения Creatio, которое развернуто cloud, функциональность сервиса поиска и объединения дублей включена по умолчанию.

Для пользователей приложения Creatio, которое развернуто on-site, необходимо выполнить предварительную настройку [сервиса глобального поиска](#) в поисковой системе Elasticsearch. Для настройки сервиса поиска и объединения дублей необходим сервер (физическая или виртуальная машина), который должен отвечать [техническим требованиям](#). На сервере должна быть установлена операционная система Linux, которая официально поддерживает программное обеспечение [Docker](#). Перечень операционных систем, которые поддерживает Docker, содержится в [документации Docker](#). Подключение функциональности сервиса поиска и объединения дублей описано в статье [Настроить массовый поиск дублей](#).

Рекомендуем устанавливать актуальную версию сервиса поиска и объединения дублей.