Вам необходимо сделать отчет и ответить на следующие вопросы. Ответы должны быть развернуты с рисунками или примерами кода.

СУБД Algolia, модель СУБД Search engine

1. История развития СУБД

Algolia была основана в 2012 году Николя Дессанжем и Жюльеном Лемуаном во Франции. Изначально компания специализировалась на оффлайн-поиске в мобильных устройствах. В 2014 году, начав работу с двух центров обработки данных в Европе и США, Algolia открыла третий центр обработки в Сингапуре.

На сегодняшний день Algolia имеет более 70 центров обработки данных в 16 регионах мира, в том числе и в России. Компания обслуживает более 7 500 клиентов, обрабатывая запросы 70 млрд пользователей в месяц.

1. Инструменты для взаимодействия с СУБД

Algolia предлагает широкий выбор продуктовых решений для различных сфер бизнеса и каналов коммуникации: сайты, приложения, голосовые ассистенты и др.

* Поиск и навигация (Site Search) позволяет вашим клиентам получить максимально релевантный результат по своему поисковому запросу. Поисковой движок учитывает контекст и тип запроса, возможные опечатки, синонимы и словоформы, ввод запроса на разных языках и многое другое.
* Голосовой поиск (Voice Search). Все больше людей предпочитают искать информацию с помощью голоса. В режиме реального времени Algolia преобразует такой вид поиска и выводит лучший результат независимо от типа устройства: десктоп, мобильное приложение и даже голосовые ассистенты.
* Гео-поиск (Geo Search) позволяет настроить стратегию ранжирования результатов поиска на основе гео-данных: можно ограничить результаты поиска улицей, городом или континентом.
* Мобильный поиск (Mobile Search). Более 60% всех поисковых запросов производятся на мобильном устройстве. Algolia предлагает фронтальные компоненты для iOS, Android и мобильного Интернета, учитывая специфические концепции, паттерны кодирования и лучшие практики UX/UI своего фреймворка.

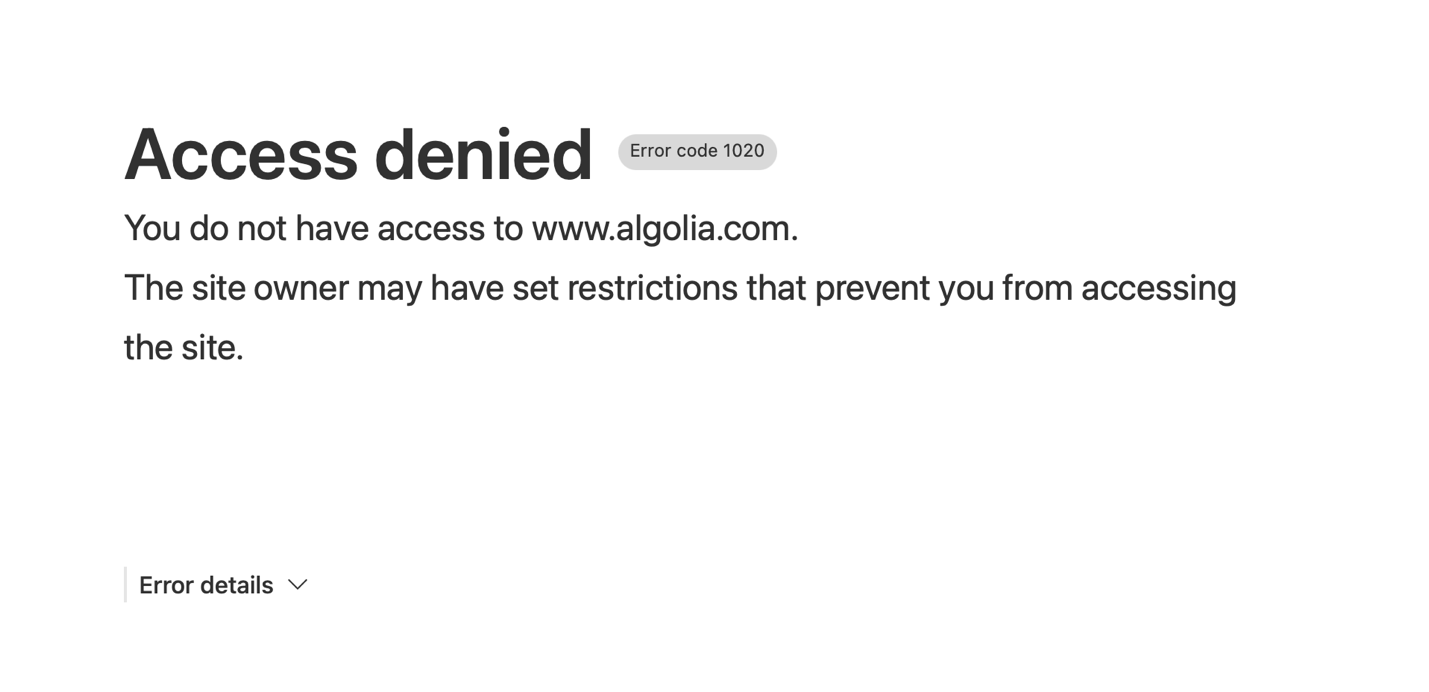
1. Какой database engine используется в вашей СУБД?

Используется база данных NoSQL. Алголия не зависит от базы данных. Можно использовать его с любым источником данных, включая базы данных NoSQL.

1. Как устроен язык запросов в вашей СУБД? Разверните БД с данными и выполните ряд запросов.

Нашла инструкцию по работе с Algolia: <https://www.codementor.io/php/tutorial/how-to-build-a-search-engine-algolia>

Но, к сожалению, при попытке пройти регистрацию на их сайте <https://www.algolia.com> выходит ошибка, что нету прав



1. Возможно ли распределение файлов БД по разным носителям?

Можно ли индексировать файлы PDF, Word и другие типы документов?Да, но не напрямую. Сначала нужно извлечь текстовое содержимое из документов и проиндексировать его в Algolia. Если есть длинные документы, то рекомендуется разделить содержимое на более мелкие фрагменты. Например, разделить текст на абзацы и проиндексировать их независимо друг от друга.

1. На каком языке/ах программирования написана СУБД?

Algolia основана на поисковой системе, созданной с нуля, реализованной на C++ и размещенной на сервере NGINX в виде модуля NGINX.

1. Какие типы индексов поддерживаются в БД? Приведите пример создания индексов.

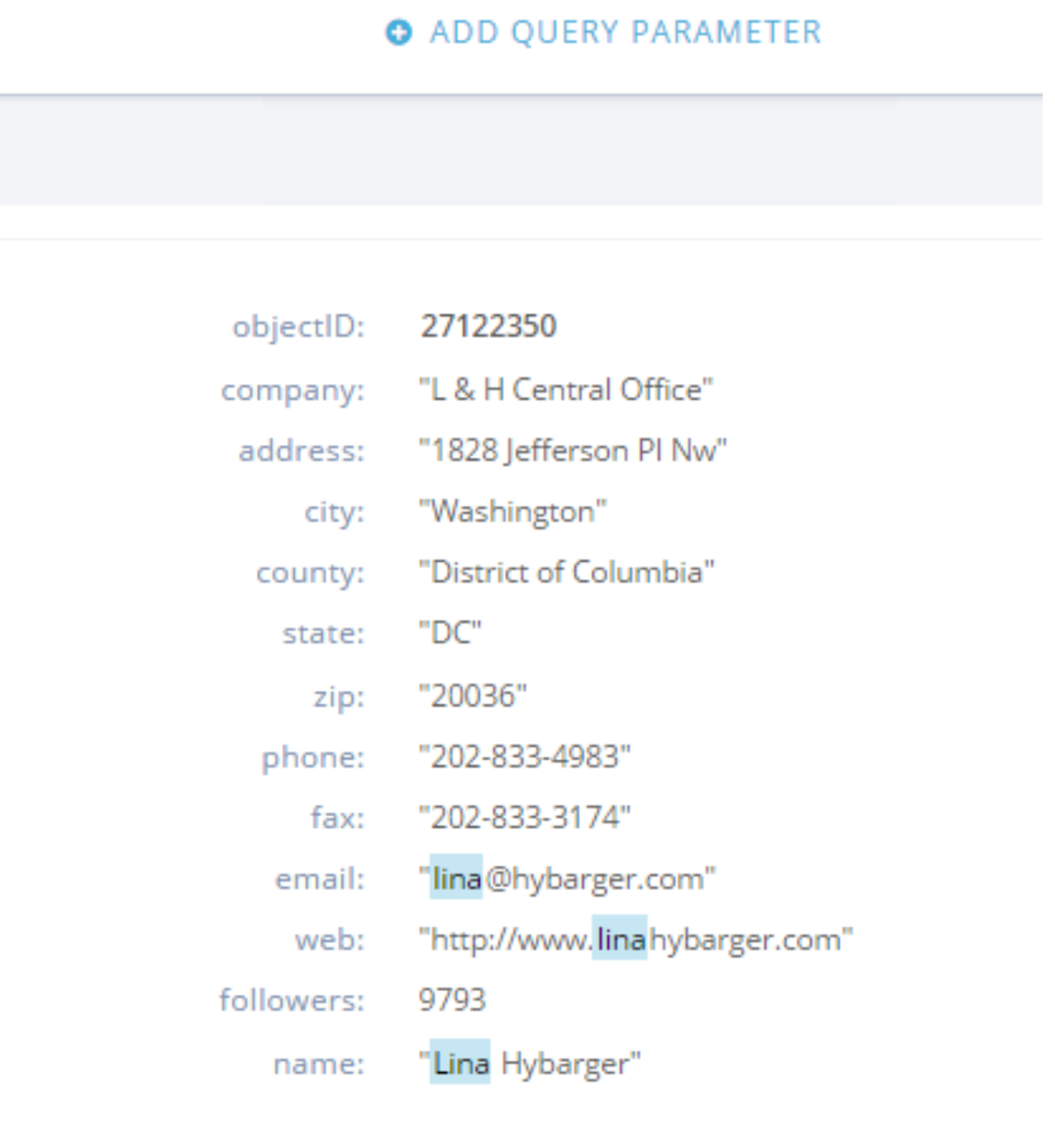
В Algolia встроено ограничение в 1000 общих индексов на приложение, чтобы защитить стабильность и производительность поиска, а также предотвратить потенциально проблемное количество индексов.

Когда выполняется поиск, то мы просматриваем записи индекса Algolia. В реляционных базах данных распределение данных идет по разным таблицам. Мы можем разбивать информацию на небольшие осмысленные блоки, чтобы избежать избыточности. В Algolia эти принципы не имеют значения. При наличии структуры записей и иерархии данных сведенные данные лучше всего подходят для поиска. Это относится к индексному уровню. Хороший подход – создать несколко индексов и сопоставить их с таблицами, где каждый индекс представляет отдельный тип объекта. Например, можно отделить фильмы от актеров и создать индекс для каждого. Но если мы хотим, чтобы наши пользоватли искали и фильмы, и актеров одновременно и чтобы они появлялись в одних и тех же результатах, то в этом случае один индекс работает лучше. Algolia автоматически создает индекс при первом добавлении в него записей.

<https://www.algolia.com/doc/api-client/methods/indexing/>

На сайте (из ссылки на вопрос 4)) есть пример создания индекса:





1. Как строится процесс выполнения запросов в вашей СУБД?

Хорошая статья: <https://www.algolia.com/blog/engineering/inside-the-algolia-engine-part-3-query-processing/>

Чтобы иметь возможность находить результаты, поисковая система должна быть в состоянии глубоко понять, о чем спрашивали. В этом заключается роль обработки запроса: проанализировать запрос и, в конечно итоге, преобразовать его, чтобы упросить его обработку поисковой системой.

И для этого поисковая система будет обрабатывать запрос в два больших этапа, оба из которых могут быть очень сложными.

Обнаружение слов в запросе: этот процесс называется токенизацией и определяет, что является словом, отвечая на такие вопросы, как: «Футболка – это одно слово или два?» Поиск альтернатив для этого набора слов: цель этого шага состоит в том, чтобы быть менее жесткими, добавляя альтернативные исправления или синонимы, чтобы избежать пропуска результатов, которые не содержат точно то, что искали, но что-то эквивалентное. Это уменьшает необходимость переформулировать запрос, чтобы найти то, что вы хотите.

1. Есть ли для вашей СУБД понятие «план запросов»? Если да, объясните, как работает данный этап.

Нету, так как Algolia – это не база данных, а поисковая система.

1. Поддерживаются ли транзакции в вашей СУБД? Если да, то расскажите о нем. Если нет, то существует ли альтернатива?

Есть транзакция для обновления индекса. Транзакция собирает все операции, которые будут выполнены, когда транзакция будет зафиксирована. Нужно избегать перестроения индекса для каждой отдельной операции, избегать хранения всех необходимых данных в памяти, например, путем сброса добавленных объектов во временные файлы на диске. Транзакция может быть создана вызовом OfflineIndex.newTransaction().

1. Какие методы восстановления поддерживаются в вашей СУБД. Расскажите о них.

Сохраняются временные резервные копии для оперативных целей, но восстанавливаются только обновленные или удаленные индексы на уровне приложения для планов надстройкой Enterprise, поэтому обязательно нужно сохранять свои резервные копии.

1. Расскажите про шардинг в вашей конкретной СУБД. Какие типы используются? Принцип работы.

Разделение позволяет масштабировать данные по нескольким индексам. Перегруженные shards демонстрируют снижение производительности.

Algolia обрабатывает все необходимые сегменты за «кулисами», они невидимы для пользователей. Shards можно динамически перебалансировать.

1. Возможно ли применить термины Data Mining, Data Warehousing и OLAP в вашей СУБД?

<https://6sense.com/tech/data-management/algolia-vs-olapreportingtool>

Можно применить OLAP

1. Какие методы защиты поддерживаются вашей СУБД? Шифрование трафика, модели авторизации и т.п.

Серверы Algolia поддерживают шифрование HTTPS, эфемерный обмен криптографическими ключами Диффи-Хеллмана на эллиптических кривых, подписанный с помощью RSA и ECDSA, а для поддерживаемых клиентов также методы полной прямой секретности (PFS), помогающие защитить трафик от скомпрометированного ключа или криптографического взлома.

1. Какие сообщества развивают данную СУБД? Кто в проекте имеет права на коммит и создание дистрибутива версий? Расскажите об этих людей и/или компаниях.

<https://community.algolia.com>

<https://www.algolia.com/ecommerce-merchandising-playbook/set-up-permissions/>

<https://github.com/algolia>

1. Создайте свои собственные данные для демонстрации работы СУБД.

См 4)

1. Как продолжить самостоятельное изучение языка запросов с помощью демобазы. Если демобазы нет, то создайте ее.

<https://video.algolia.com/watch/VqirgnCMcx2wVX5hHV9v18#request-meeting-vidyard>

https://www.algolia.com/interface-demos/568c94ea-ca9b-40a5-9eb9-979c61af3b18

1. Где найти документацию и пройти обучение

<https://resources.algolia.com>

<https://academy.algolia.com/trainings>

<https://www.algolia.com/doc/guides/getting-started/how-algolia-works/>

https://www.algolia.com/doc/

1. Как быть в курсе происходящего

Постоянно читать новые статьи/новости, документацию. Интересоваться данным СУБД.

Пару ссылок, где можно более подробно узнать про Algolia:

<https://www.youtube.com/watch?v=VYjecPkcIcE&list=PLuHdbqhRgWHJDATsVq_Mrj3NjIRjREiM1&index=1>

<https://www.algolia.com/about/news/>