

СОЗДАНИЕ РАСПРЕДЕЛЕННОГО ВЕБ-ПРИЛОЖЕНИЯ ДЛЯ РАБОТЫ С БОЛЬШИМ КОЛИЧЕСТВОМ ДОКУМЕНТОВ НА ОСНОВЕ ELASTICSEARCH

Вальченко И. Н.

*Белорусский государственный университет, Минск, Беларусь,
e-mail: i.valchenk@hotmail.com*

В настоящее время все больше и больше людей отказываются от использования обычных документов, распечатанных на бумаге, в пользу облачных хранилищ таких как Google Documents, Microsoft SharePoint и им подобные.

Главным достоинством этих систем является доступность, простота в использовании и отказоустойчивость. Обеспечение сохранности и репликации данных ложится на плечи компаний владельцев этих сервисов.

Однако ни один из этих сервисов не решает проблему полнотекстового поиска по документу. Имея в своем хранилище порядка 100 тысяч документов, невозможно досконально знать содержимое каждого из них. Не представляется возможным воспользоваться поиском, предлагаемым этими сервисами, т.к. они предоставляют поиск только по названию файлов или документов.

В силу этого актуальным является разработка универсальной системы с клиент-серверной архитектурой, которая посредством веб-браузера осуществляла бы быструю обработку документов, накопление полученных и сопутствующих данных в базе, расширенный поиск, группировку и сортировку по различным параметрам, обработку накопленных данных, а также предоставление в удобном для пользователя виде итоговых результатов исследований.

Пользователь данной системы должен иметь возможность создавать свои пользовательские свойства (атрибуты) для каждого документа, чтобы в последующем иметь возможность искать, группировать и сортировать по ним.

Планируется, что данная система будет представлять собой веб-сайт с большим количеством посетителей, поэтому ее следует сделать распределенной и масштабируемой. Для этого необходимо иметь несколько сущностей - back-end нашего сайта и балансировщик нагрузки, который будет отвечать за равномерное распределение запросов между всеми сущностями приложения.

Список литературы

1. Рихтер Д. CLR via C#. Программирование на платформе Microsoft .NET Framework 4.5 на языке C# / Д. Рихтер. – 4-е изд. – СПб.: Питер, 2013. – 896 с.
2. Крамп М., Луиджбритс Б. Azure Development Guide / М. Крамп, Б. Луиджбритс. – Microsoft Press, 2017. – 82 с.
3. Кормли К., Тонг З. Elasticsearch: The Definitive Guide / К. Кормли, З. Тонг – O'Reilly Media, 2015. – 684 с.
4. Бёрнс Б. Designing Distributed Systems: Patterns and Paradigms for Scalable, Reliable services / Б. Бёрнс. – O'Reilly Media, 2018. – 166 с.
5. МакКоннелл С. Совершенный код. Мастер-класс / С. МакКоннелл – Русская редакция, 2017. – 914 с.
6. Веббер Д., Парастатидис С., Робинсон И. REST in Practice: Hypermedia and Systems Architecture / Д. Веббер, С. Парастатидис, И. Робинсон – O'Reilly Media, 2010. – 448 с.