

А. О. Куценко
(ГГУ имени Ф. Скорины, Гомель)
Науч. рук. **В. Н. Леванцов**, ст. преподаватель

ELASTICSEARCH КАК NOSQL БАЗА ДАННЫХ

Elasticsearch – это RESTful-система распределенного поиска и аналитики с открытым исходным кодом, построенная на Apache Lucene. Elasticsearch также является базой данных NoSQL, поэтому данные в ней хранятся в неструктурированном виде. Elasticsearch очень эффективен для выполнения сложного поиска по большим объемам данных благодаря Lucene. Elasticsearch использует структуру, основанную на документах, а не таблицах и схемах, и поставляется с обширным REST API интерфейсом для хранения и поиска данных.

Данные в Elasticsearch хранятся в виде документов JSON. Документы являются основной единицей информации, которую можно проиндексировать. Каждый документ имеет уникальный идентификатор, а данные в нем определяются полями, состоящими из ключей и значений. Ключ является именем поля, а значение может быть элементом многих различных типов, например, строкой, числом, логическим выражением, другим объектом или массивом значений. Документы можно сравнить со строкой в таблице реляционной базы данных. Документы со схожими характеристиками объединяются в индексы. Индекс является сущностью самого высокого уровня, которую можно запрашивать в Elasticsearch. Его можно сравнить с базой данных в мире реляционных баз данных. Индекс идентифицируется именем, которое используется для ссылки на него при выполнении операций поиска, обновления и удаления документов.

Elasticsearch является распределенной системой, состоящей из одного и более узлов, которые работают как одно целое, что позволяет масштабировать и распределять нагрузку. Узел в Elasticsearch представлен отдельным сервером, который является частью кластера. Узел хранит часть данных, а также занимается их индексацией. По умолчанию каждому узлу автоматически назначается уникальный идентификатор, который используется в целях управления и становится еще более важным в кластерной среде. Кластер состоит из одного или нескольких узлов, которые вместе содержат полный набор данных. Кластер занимается распределением задач, поиском и индексацией по всем узлам.