Migrating SQL to ElasticSearch DB

A mid-semester report submitted to Cluster Innovation Centre
in partial fulfillment of the requirements for

Semester Long Project

(March 2018)

Mentored by:

Mr. Sachin Kumar

Raghav Singh - 11522

Cluster Innovation Centre
University of Delhi
Delhi – 110007

1. Abstract

In recent years the big data has pushed us into a data expansion. So, we must focus on ways to manage our data better. The RDBMS is not good for all causes as it is computationally expensive to store and use. Also, many a times the data cannot be classified to fit in predefined columns and rows as it is irregular and difficult to segregate and process. So, we must focus on NoSQL databases for such projects. The main problem here is that many people who have knowledge of RDBMS (SQL) have to learn a totally different paradigm: NoSQL. Hence if there is a medium which converts their SQL knowledge into NoSQL will help them a lot. Such migration will be very helpful for large projects. Hence we propose to work on this research problem. The project aims to create a javascript web interface with React on the frontend and Node at the backend with effectively (ElasticSearch Express React Node) EERN stack which will take SQL queries as input and convert them back into elasticsearch query DSL. This project will progress as an open source project and will be available on github for collaboration by other developers around the globe.

2. Objectives

- Write automated scripts which will convert a structured sql database table into an unstructured NoSQL based flat documents
- Scripts to create mappings automatically using a sql table info
- Provide a way to users to generate elasticsearch query DSL by writing their commands in SQL

3. Methodology

- Check whether SQL query is valid (syntax)

- Use a library for syntax check and indentation

- Trace SQL query using regex

- Create a parser for softcoding, initially set some rules for parsing (hardcoding) so as to tune it easily

- Support at least CRUD (Create Read Update Delete) operations

- Using the parser convert it into DSL elasticsearch query

4. Timeline

Description	Timeframe
1. Literature Review	January - February
2. Learning web technologies: Javascript (Node and React)	February - March
3. Deployment, Documentation and Debugging	Early March - Early April
4. Early beta and release	Early April
5. Testing and Licensing	Late April

5. References

- Elasticsearch: The definitive guide by Clinton Gormley & Zachary Tong

- Official Documentation: Elasticsearch

- The Complete Node.js Developer Course (2nd Edition) | Udemy by Andrew Mead

- Node with React | Udemy by Stephen Grider