Group 7: Explore the use of text search tools.

Group Members:

1. **Rishabh Jain (rjain35@hawk.iit.edu, A20495530)**
2. **Siddhant Bhatia (sbhatia14@hawk.iit.edu , A20500508)**
3. **Utkarsha (umalegaonkar@hawk.iit.edu, A20493621)**

Project Proposal

Introduction:

After a healthy discussion, we have finalized that we will be using ElasticSearch as the text search tool for our project. ElasticSearch is an Apache Lucene(Java search Library) based real-time text search and analytics tool. It is based on NoSQL and uses JSON to store data and is built on Java. Therefore, we can use ElasticSearch on different platforms very easily.

In ElasticSearch, an index is similar to tables in RDBMS. Every table is a collection of rows, just as every index is a collection of documents in ElasticSearch.

The mechanism of ElasticSearch is something like this:

Beats → LogStash → ElasticSearch → Kibana

A flawless combination of all these, is further called as “Elastic Stack”.

Beats - Beats is a free and open platform for single-purpose data shippers. They send data from hundreds or thousands of machines and systems to Logstash or Elasticsearch.

LogStash - Log Stash provides an input stream of data to the ELasticSearch for storage and search.

Kibana - Kibana is a web-based UI by which we interact with the data and the indexes which we can use, which is prepared by ElasticSearch. Using Kibana we can create dashboards and visualize data in a variety of charts, tables and maps.

Use cases of ElasticSearch:

1. Logging
2. Metrics
3. Security Analytics E.g., Slack
4. Business Analytics E.g., Tinder
5. Today many applications like Instacart, uber use ElasticSearch to find quick and relevant results to the users.

There are many other similar tools available in the market, such as Apache SOLR, and many others.

ElasticSearch VS Apache Solr

1. Elastic search is heavier than Apache Solr in terms of storage consumed, and memory used.
2. Elasticsearch used Yaml formatted configuration files and Apache Solr used XML based configuration files.

Using ElasticSearch, we as a team will be creating a Search Engine based auto-complete System.

Objectives:

1. We will be creating an auto-complete system that will work based on the relevant searches that we type in the search engine.
2. It will show up the matching results related to the data that we type in the search engine.
3. We will be using REST APIs with an endpoint from where we will query the data and will be using Kibana to show and visualize the results.
4. Like Google Search provide recommendations based on the search, our system using the Elastic stack will be able to provide the same functionality based on the dataset that we are using.

References:

https://www.elastic.co/blog/found-uses-of-elasticsearch
https://www.elastic.co/what-is/elk-stack

Following books will be used for reference

1. Amazon Elasticsearch Service: Developer Guide
2. Elasticsearch 7.0 Cookbook: Over 100 recipes for fast, scalable, and reliable search for your enterprise
3. Learning Elasticsearch 7.x: Index, Analyze, Search and Aggregate Your Data Using Elasticsearch

Research Paper:

1. Paper on Searching and Indexing Using Elasticsearch Darshita Kalyani, Dr. Devarshi Mehta Gujarat, India
2. Research Document Search using Elastic Search.