# Department of Computing

**Course Code: CS332**

**Class: BSCS-9ABC**

Lab 9: Indexing, Importing and Searching data in Apache Solr

**Date: 11th April ,2022, 15th April 2022**

**Time: 09:00 -11:50 am, 2:00 – 5 :00 pm**

**Instructor: Dr. Farzana Jabeen, Shah Khalid**

|  |
| --- |
| **Submitted By:**  **Fatima Seemab**  **291310**  **BSCS 9B** |

Lab 9: Indexing, Importing and Searching data in Apache Solr

**Introduction**

**Indexing Data:**

Indexing enables users to locate information in a document. Indexing collects, parses, and stores documents. Indexing is done to increase the speed and performance of a search query while finding a required document.

**Importing Data**

A Solr index can accept data from many different sources, including XML files, comma-separated value (CSV) files, data extracted from tables in a database, and files in common file formats such as Microsoft Word or PDF.

Here are the three most common ways of loading data into a Solr index:

* Using the [Solr Cell](https://solr.apache.org/guide/6_6/uploading-data-with-solr-cell-using-apache-tika.html#uploading-data-with-solr-cell-using-apache-tika) framework built on Apache Tika for ingesting binary files or structured files such as Office, Word, PDF, and other proprietary formats.
* Uploading XML files by sending HTTP requests to the Solr server from any environment where such requests can be generated.
* Writing a custom Java application to ingest data through Solr’s Java Client API (which is described in more detail in [Client APIs](https://solr.apache.org/guide/6_6/client-apis.html#client-apis)). Using the Java API may be the best choice if you’re working with an application, such as a Content Management System (CMS), that offers a Java API.

**Searching:**

Solr offers a rich, flexible set of features for search. To understand the extent of this flexibility, it’s helpful to begin with an overview of the steps and components involved in a Solr search.

When a user runs a search in Solr, the search query is processed by a request handler. A request handler is a Solr plug-in that defines the logic to be used when Solr processes a request. Solr supports a variety of request handlers. Some are designed for processing search queries, while others manage tasks such as index replication.

Solr is the popular, blazing fast, open source NoSQL search platform from the Apache Lucene project. Its major features include powerful full-text search, hit highlighting, faceted search, dynamic clustering, database integration, rich document handling, and geospatial search. Solr is highly scalable, providing fault tolerant distributed search and indexing, and powers the search and navigation features of many of the world's largest internet sites.

**Objectives**

To install, setup and understanding the basics of Apache solr.

**Tools/Software Requirement**

Apache solr

**Useful tutorial links:**

<https://solr.apache.org/guide>/

<https://blog.kiprosh.com/distributed-indexing-and-searching-with-apache-solrcloud>

<https://solr.apache.org/guide/6_6/distributed-search-with-index-sharding.html>

**Description**

**Lab Prerequisite:**

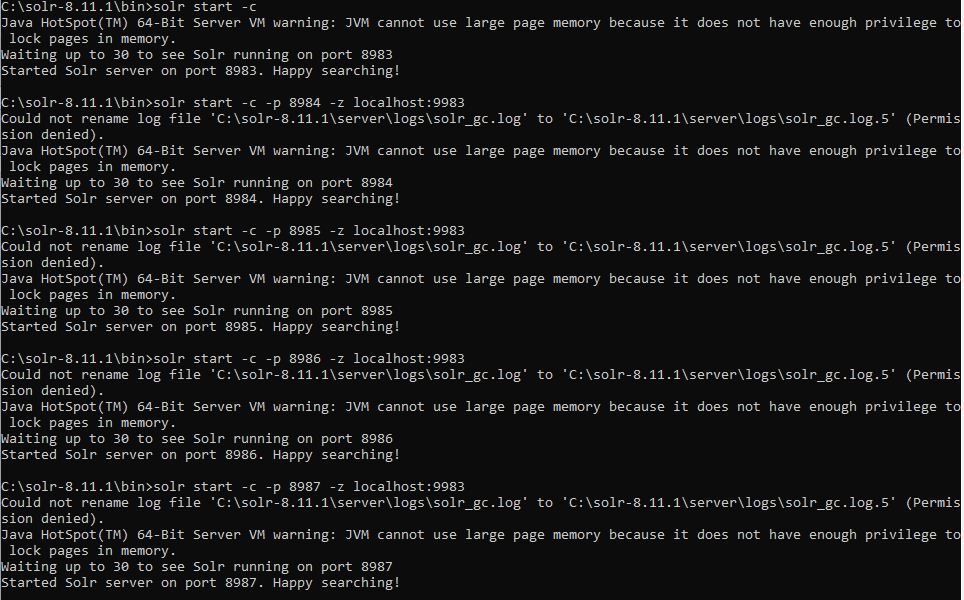
* Install & configure solr.
* Create a cluster of servers(shards).

Installation guide of Apache Solr for windows and Ubuntu are available on LMS if you have not installed it yet, follow the guide and install and configure it first.

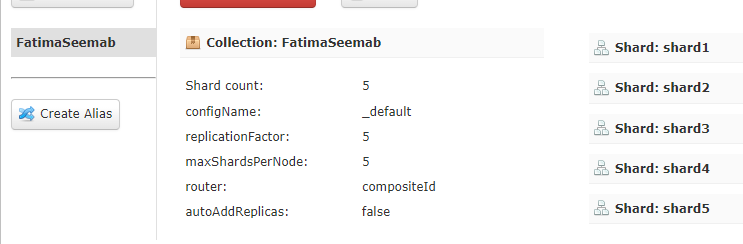
**Task**

You have already created a cluster of servers(shards) now take a sample data Index it, Import and perform searching on it.

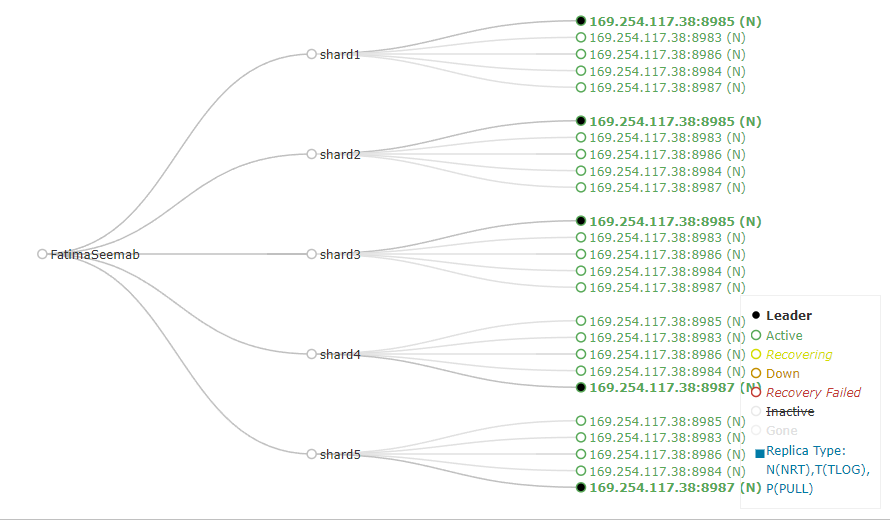
**Creating five shards**



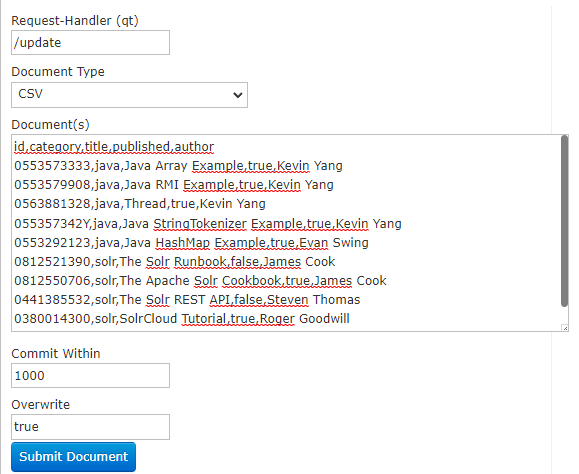
**Creating a Collection**



**Graph**



**Importing Data**

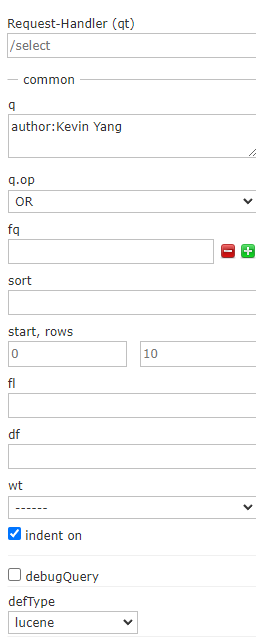




**Query Category : java**

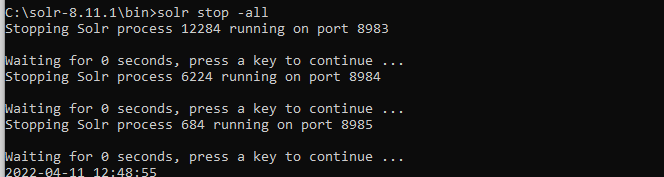


**Query Author: Kevin Yang**



**Response**





Useful Links:   
<https://examples.javacodegeeks.com/apache-solr-clustering-example/>

**Deliverables**

Upload single word file with screenshots of your output and code on LMS.

In case of any confusion or query feel free to contact me at [Aftab.hussain1@seecs.edu.pk](mailto:Aftab.hussain1@seecs.edu.pk)