# Department of Computing

**Course Code: CS332**

**Class: BSCS-9ABC**

Lab 08: Getting Started with Apache Solr

**Date: 4th April 2022, 8th April 2022**

**Time: 09:00-12:00; 02:00 – 05:00**

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

|  |
| --- |
| **Submitted By:**  **Fatima Seemab**  **291310**  **Lab 8** |

# Lab 08: Getting Started with Apache Solr

**Introduction**

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://www.tutorialspoint.com/apache\_solr/apache\_solr\_overview.htm**](https://www.tutorialspoint.com/apache_solr/apache_solr_overview.htm)

[**https://solr.apache.org/guide/8\_8/solr-tutorial.html**](https://solr.apache.org/guide/8_8/solr-tutorial.html)

**Description**

Solr is an open-source search platform which is used to build search applications. It was built on top of Lucene (full text search engine). Solr is enterprise-ready, fast and highly scalable. The applications built using Solr are sophisticated and deliver high performance.

It was Yonik Seely who created Solr in 2004 in order to add search capabilities to the company website of CNET Networks. In Jan 2006, it was made an open-source project under Apache Software Foundation. Its latest version, Solr 6.0, was released in 2016 with support for execution of parallel SQL queries.

Solr can be used along with Hadoop. As Hadoop handles a large amount of data, Solr helps us in finding the required information from such a large source. Not only search, Solr can also be used for storage purpose. Like other NoSQL databases, it is a non-relational data storage and processing technology.

In short, Solr is a scalable, ready to deploy, search/storage engine optimized to search large volumes of text-centric data.

**Features of Apache Solr**

Solr is a wrap around Lucene’s Java API. Therefore, using Solr, you can leverage all the features of Lucene. Let us take a look at some of most prominent features of Solr −

Restful APIs − To communicate with Solr, it is not mandatory to have Java programming skills. Instead you can use restful services to communicate with it. We enter documents in Solr in file formats like XML, JSON and .CSV and get results in the same file formats.

Full text search − Solr provides all the capabilities needed for a full text search such as tokens, phrases, spell check, wildcard, and auto-complete.

Enterprise ready − According to the need of the organization, Solr can be deployed in any kind of systems (big or small) such as standalone, distributed, cloud, etc.

Flexible and Extensible − By extending the Java classes and configuring accordingly, we can customize the components of Solr easily.

NoSQL database − Solr can also be used as big data scale NOSQL database where we can distribute the search tasks along a cluster.

Admin Interface − Solr provides an easy-to-use, user friendly, feature powered, user interface, using which we can perform all the possible tasks such as manage logs, add, delete, update and search documents.

Highly Scalable − While using Solr with Hadoop, we can scale its capacity by adding replicas.

Text-Centric and Sorted by Relevance − Solr is mostly used to search text documents and the results are delivered according to the relevance with the user’s query in order.

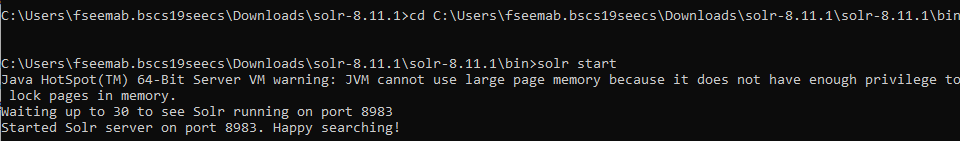
Unlike Lucene, you don’t need to have Java programming skills while working with Apache Solr. It provides a wonderful ready-to-deploy service to build a search box featuring autocomplete, which Lucene doesn’t provide. Using Solr, we can scale, distribute, and manage index, for large scale (Big Data) applications.

**Lab Tasks**

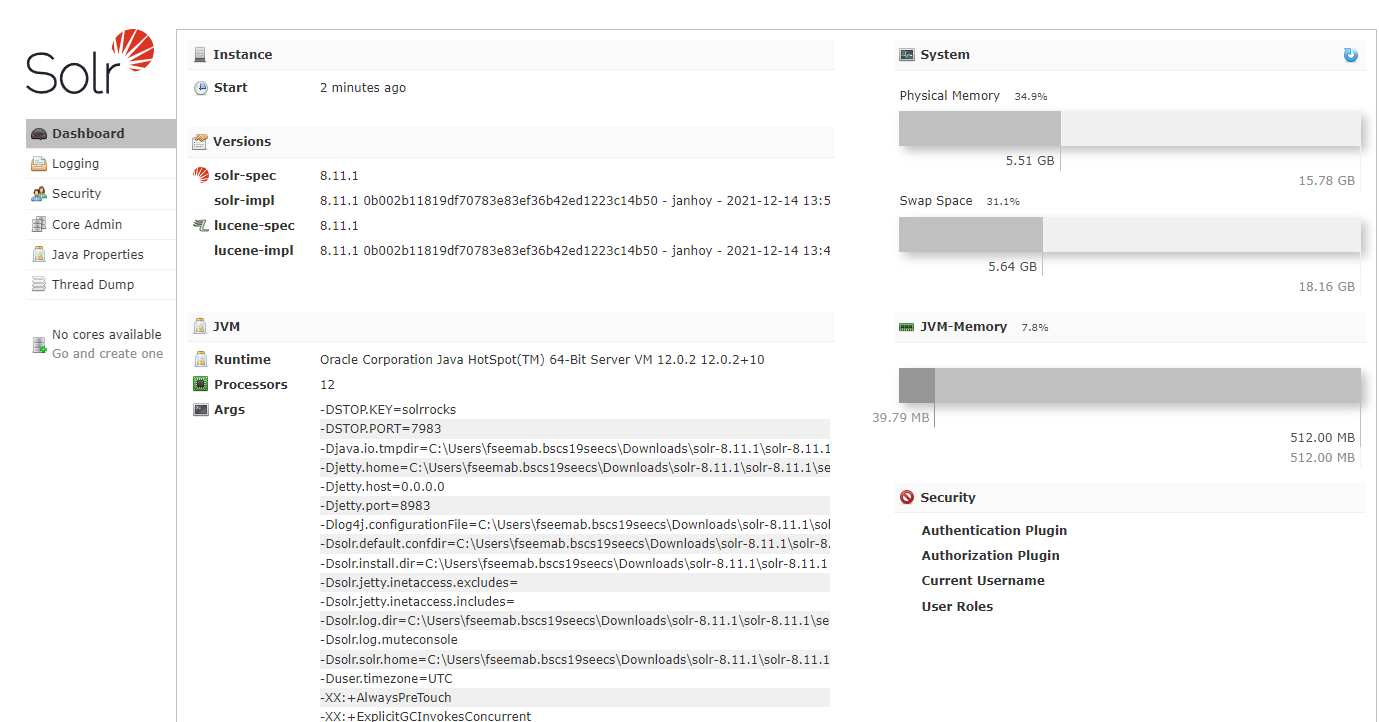
**Task 1**

|  |
| --- |
| Install & configure solr. |

**CMD OUTPUT:**



**Localhost:8983**



Install Apache Solr on windows

Useful link:

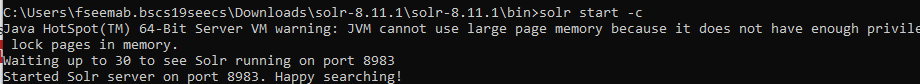
<https://www.youtube.com/watch?v=_fW98B84lXA&t=90s&ab_channel=HowTo>

<https://www.tutorialspoint.com/apache_solr/apache_solr_on_windows_environment.htm>

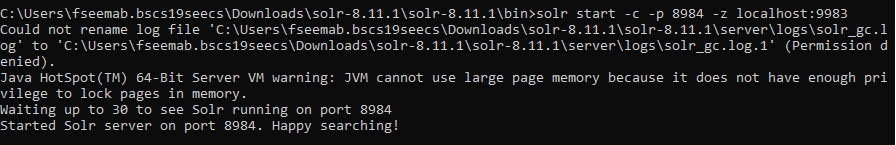
**Task 2:**

|  |
| --- |
| Create a cluster of servers(shards). |

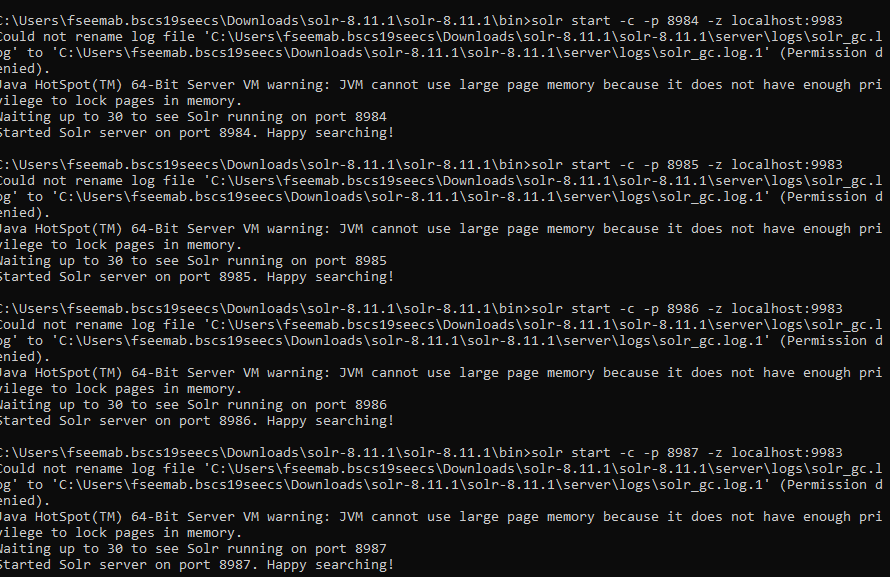
**Running 1st server**



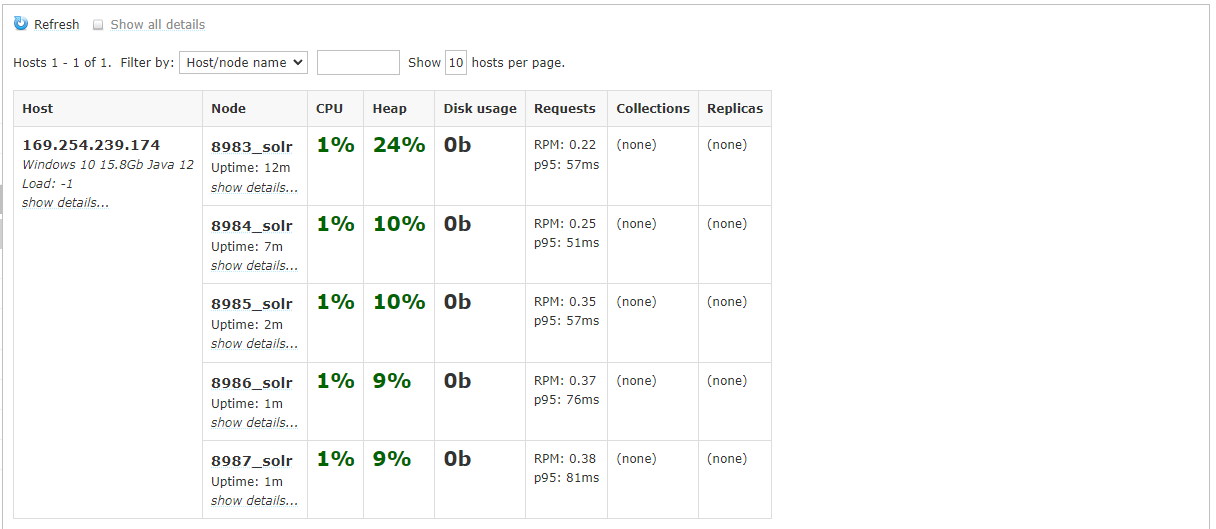
**Running 2nd server**



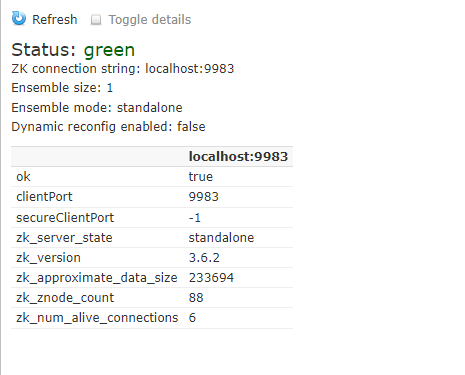
**Running 5 servers**



**5 servers in cloud**



**Zookeper**



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

**Deliverables**

Upload single word file with screenshots of your installation and cluster creation on LMS.