Apache Solr - Adding Documents (XML)

In the previous chapter, we explained how to add data into Solr which is in JSON and .CSV file formats. In this chapter, we will demonstrate how to add data in Apache Solr index using XML document format.

Sample Data

Suppose we need to add the following data to Solr index using the XML file format.

Student ID	First Name	Last Name	Phone	City
001	Rajiv	Reddy	9848022337	Hyderabad
002	Siddharth	Bhattacharya	9848022338	Kolkata
003	Rajesh	Khanna	9848022339	Delhi
004	Preethi	Agarwal	9848022330	Pune
005	Trupthi	Mohanty	9848022336	Bhubaneshwar
006	Archana	Mishra	9848022335	Chennai

Adding Documents Using XML

To add the above data into Solr index, we need to prepare an XML document, as shown below. Save this document in a file with the name **sample.xml**.

```
<add>
   <doc>
      <field name = "id">001</field>
      <field name = "first name">Rajiv</field>
      <field name = "last name">Reddy</field>
      <field name = "phone">9848022337</field>
      <field name = "city">Hyderabad</field>
   </doc>
   <doc>
      <field name = "id">002</field>
      <field name = "first name">Siddarth</field>
      <field name = "last name">Battacharya</field>
```

```xml
         <field name = "phone">9848022338</field>
         <field name = "city">Kolkata</field>
   </doc>
   <doc>
         <field name = "id">003</field>
         <field name = "first name">Rajesh</field>
         <field name = "last name">Khanna</field>
         <field name = "phone">9848022339</field>
         <field name = "city">Delhi</field>
   </doc>
   <doc>
         <field name = "id">004</field>
         <field name = "first name">Preethi</field>
         <field name = "last name">Agarwal</field>
         <field name = "phone">9848022330</field>
         <field name = "city">Pune</field>
   </doc>
   <doc>
         <field name = "id">005</field>
         <field name = "first name">Trupthi</field>
         <field name = "last name">Mohanthy</field>
         <field name = "phone">9848022336</field>
         <field name = "city">Bhuwaeshwar</field>
   </doc>
   <doc>
         <field name = "id">006</field>
         <field name = "first name">Archana</field>
         <field name = "last name">Mishra</field>
         <field name = "phone">9848022335</field>
         <field name = "city">Chennai</field>
   </doc>
</add>
```

As you can observe, the XML file written to add data to index contains three important tags namely, <add> </add>, <doc></doc>, and < field >< /field >.

- **add** − This is the root tag for adding documents to the index. It contains one or more documents that are to be added.

- **doc** − The documents we add should be wrapped within the <doc></doc> tags. This document contains the data in the form of fields.

- **field** − The field tag holds the name and value of the fields of the document.

After preparing the document, you can add this document to the index using any of the means discussed in the previous chapter.

Suppose the XML file exists in the **bin** directory of Solr and it is to be indexed in the core named **my_core**, then you can add it to Solr index using the **post** tool as follows −

```
[Hadoop@localhost bin]$ post -c my_core sample.xml
```

On executing the above command, you will get the following output.

```
/home/Hadoop/java/bin/java -classpath
/home/Hadoop/Solr/dist/Solr-
core6.2.0.jar -Dauto = yes -Dc = my_core -Ddata = files
org.apache.Solr.util.SimplePostTool sample.xml
SimplePostTool version 5.0.0
Posting files to [base] url
http://localhost:8983/Solr/my_core/update...
Entering auto mode. File endings considered are
xml,json,jsonl,csv,pdf,doc,docx,ppt,pptx,
xls,xlsx,odt,odp,ods,ott,otp,ots,rtf,htm,html,txt,log
POSTing file sample.xml (application/xml) to [base]
1 files indexed.
COMMITting Solr index changes to
http://localhost:8983/Solr/my_core/update...
Time spent: 0:00:00.201
```

Verification

Visit the homepage of Apache Solr web interface and select the core **my_core**. Try to retrieve all the documents by passing the query ":" in the text area **q** and execute the query. On executing, you can observe that the desired data is added to the Solr index.