

# HADOOP & NATURAL LANGUAGE PROCESSING (DATA SCIENCE FLAVOR) ALL IN ONE COURSE



GITHUB · <https://github.com/nitinkaushik01/Hadoop-Natural-Language-Processing>

---

**LOGIN TO AMBARI USING BELOW MENTIONED URL AND CREDENTIALS:**

Ambari : <http://sandbox.hortonworks.com:8080/>

Credentials:      Username: [raj\\_ops](#)      Password: [raj\\_ops](#)

**OPEN THE TERMINAL AND TYPE BELOW COMMAND ON COMMAND LINE**

```
su solr
```

**EDIT THE BELOW MENTIONED PATH SO THAT SOLR CAN RECOGNIZE TWEET'S TIMESTAMP FORMAT BY RUNNING THE COMMANDS**

```
cp -r /opt/lucidworks-hdpsearch/solr/server/solr/configsets/data_driven_schema_configs  
/opt/lucidworks-hdpsearch/solr/server/solr/configsets/tweet_configs
```

```
vi /opt/lucidworks-hdpsearch/solr/server/solr/configsets/tweet_configs/conf/solrconfig.xml
```

**TYPE BELOW MENTIONED LINE. PLEASE NOTE THAT YOU DON'T HAVE TO BE IN INSERT MODE TO DO THIS**

```
/solr.ParseDateFieldUpdateProcessorFactory
```

**INSERT BELOW MENTIONED LINE ABOVE ALL OTHER <STR> TAGS AND BELOW <ARR NAME="FORMAT">**

```
<str>EEE MMM d HH:mm:ss Z yyyy</str>
```

**NOW PRESS ESCAPE KEY ON KEYBOARD AND TYPE :wq TO SAVE THE SOLRCONFIG.XML FILE**

**REPLACE JSON FILE BY USING BELOW MENTIONED COMMANDS TO MOVE THE ORIGINAL FILE AND DOWNLOAD THE REPLACEMENT FILE**

```
cd /opt/lucidworks-hdpsearch/solr/server/solr-webapp/webapp/banana/app/dashboards/
```

```
mv default.json default.json.orig
```

```
wget https://raw.githubusercontent.com/hortonworks/data-tutorials/master/tutorials/hdp/analyzing-social-media-and-customer-sentiment-with-apache-nifi-and-hdp-search/assets/default.json
```

**NOW ADD COLLECTION NAMED AS "tweets" BY USING BELOW MENTIONED COMMAND**

```
/opt/lucidworks-hdpsearch/solr/bin/solr create -c tweets -d tweet_configs -s 1 -rf 1 -p 8983
```

Note: here -c indicates the name  
-d is the config directory  
-s is the number of shards  
-rf is the replication factor  
-p is the port at which solr is running

**LOG OUT OF THE SOLR USER AND GO BACK TO ROOT USER BY TYPING BELOW COMMAND**

[exit](#)

**NAVIGATE TO SOLR UI BY TYPING BELOW MENTIONED COMMAND**

<http://sandbox.hortonworks.com:8983/solr/>