Java-Success.com

Prepare to fast-track, choose & go places with 800+ Java & Big Data Q&As with lots of code & diagrams.

01b: Convert XML file To Sequence File – writing & reading – Hadoop File System (i.e HDFS)

Posted on December 11, 2016

This extends Convert XML file To Sequence File – writing & reading – Local File System.

Step 1: Upload “report.xml” onto HDFS. E.g using the Cloudera HUE on to path “/user/cloudera/report-data”. You need to create the “report-data” folder.

300+ Java Interview FAQs

300+ Java FAQs 🔥

16+ Java Key Areas Q&As

150+ Java Architect FAQs

80+ Java Code Quality Q&As

150+ Java Coding Q&As

300+ Big Data Interview FAQs

300+ Big Data FAQs 🔥

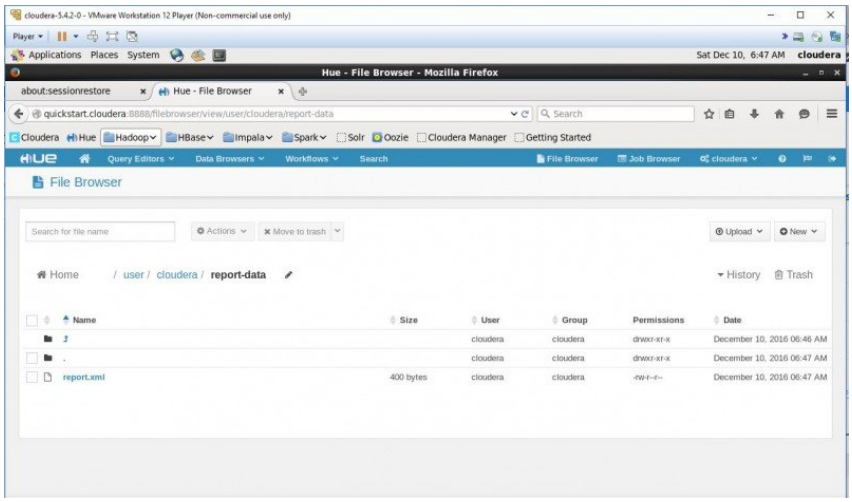
Tutorials - Big Data

TUT - ⏯ Starting Big Data

TUT - Starting Spark & Scala

```
1
2  <?xml version="1.0" encoding="UTF-8"?>
3  <transactionReports xmlns="http://mytutorial.com/
4      <transactionReport>
5          <report>
6              <reportNumber>9999</reportNumber>
7              <createdDatetime>2015-06-15T11:29:52+
8              <processedDatetime>2015-06-15T11:29:5
9              <reportStatusCode>Active</reportStatus
10         </report>
11     </transactionReport>
12 </transactionReports>
13
```

The uploaded file on Hue:



XML file uploaded to HDFS via HUE.

Step 2: Change the code to read from HDFS and write to HDFS. The following code

1) Opens an XML file on HDFS using the HDFS File System

2) Reads the contents into a StringBuilder.

3) Writes as a sequence file to the HDFS using the "SequenceFile.Writer". The "**key**" being "**1**" and the **value** being the "**XML**" file contents.

TUT - Starting with Python
TUT - Kafka
TUT - Pig
TUT - Apache Storm
TUT - Spark Scala on Zeppelin
TUT - Cloudera
TUT - Cloudera on Docker
TUT - File Formats
TUT - Spark on Docker
TUT - Flume
TUT - Hadoop (HDFS)
TUT - HBase (NoSQL)
TUT - Hive (SQL)
TUT - Hadoop & Spark
TUT - MapReduce
TUT - Spark and Scala
TUT - Spark & Java
TUT - PySpark on Databricks
TUT - Zookeeper

800+ Java Interview Q&As

300+ Core Java Q&As
300+ Enterprise Java Q&As
150+ Java Frameworks Q&As
120+ Companion Tech Q&As
Tutorials - Enterprise Java

```
1
2  package com.mytutorial;
3
4  import java.io.BufferedReader;
5  import java.io.IOException;
6  import java.io.InputStreamReader;
7
8  import org.apache.hadoop.conf.Configuration;
9  import org.apache.hadoop.fs.FileSystem;
10 import org.apache.hadoop.fs.Path;
11 import org.apache.hadoop.io.BytesWritable;
12 import org.apache.hadoop.io.IOUtils;
13 import org.apache.hadoop.io.IntWritable;
14 import org.apache.hadoop.io.SequenceFile;
15 import org.apache.hadoop.util.ReflectionUtils;
16
17 public class ConvertXmlToSequence {
18
19     private static final String hdfsUrl = "hdfs:/
20
21     private static final String FILE_IN = "repor
22     private static final String FILE_OUT = "repo
23
24     public static void main(String[] args) throw
25         Configuration conf = new Configuration()
26
27         Path inputFile = new Path(hdfsUrl + "/" +
28         Path outputFile = new Path(hdfsUrl + "/"
29
30         write(conf, inputFile, outputFile); // wr
31         read(conf, outputFile); // read seq file
32     }
33
34     /**
35      * Write a text file to sequence file
36      *
37      * @param conf
38      * @param inputFile
39      * @param outputFile
40      */
41     public static void write(Configuration conf,
42         BufferedReader reader = null;
43         SequenceFile.Writer writer = null;
44
45         try {
46             writer = SequenceFile.createWriter(co
47                     .file(outputFile), SequenceFi
48                     SequenceFile.CompressionType
49                     SequenceFile.Writer.keyClass(
50                     SequenceFile.Writer.valueCla
51
52             IntWritable key = new IntWritable(1)
53
54             FileSystem fs = FileSystem.get(inputF
55
```

```
56              reader = new BufferedReader(new Inpu
57
58              reader = new BufferedReader(new Inpu
59                      fs.open(inputFile)));
60
61              String line;
62              line = reader.readLine();
63              StringBuilder sb = new StringBuilder
64              while (line != null) {
65                  sb.append(line);
66                  line = reader.readLine();
67              }
68
69              BytesWritable value = new BytesWrita
70              writer.append(key, value);
71              writer.close();
72          } catch (IOException e) {
73              System.out.println("Cannot read sequ
74          }
75
76          finally {
77              IOUtils.closeStream(reader);
78              IOUtils.closeStream(writer);
79          }
80      }
81
82      /**
83       * Read a sequence file
84       *
85       * @param conf
86       * @param sequenceFileToRead
87       */
88      public static void read(Configuration conf, 
89          SequenceFile.Reader reader = null;
90          try {
91              reader = new SequenceFile.Reader(con
92                      SequenceFile.Reader.file(seq
93              IntWritable keyRead = (IntWritable) 
94                      reader.getKeyClass(), conf);
95              BytesWritable valueRead = (BytesWrit
96                      .newInstance(reader.getValue
97
98              while (reader.next(keyRead, valueRea
99                  System.out.println("key : " + ke
100                         + new String(valueRead.g
101                 valueRead = (BytesWritable) Refl
102                         reader.getValueClass(), 
103             }
104         } catch (IOException e) {
105             System.out.println("Cannot read sequ
106                     + sequenceFileToRead + e);
107         }
108
109         IOUtils.closeStream(reader);
110     }
```

```
111
112 }
113
114
```

Output:

```
1
2 key : 1 - value : <?xml version="1.0" encoding="UTI
3
```

‹ 03: JAXB with StAX Tutorial step by step for marshalling

06: Avro Schema evolution tutorial ›

Disclaimer