

Hi Vapukov,

My main issue is to read the huge xml. Even if I want to split it, Talend will have to read it first. This step is the bottleneck. I have tried to change the .ini file to increase the java arguments with -Xms1024m and -Xmx9208m. As well, I have tried to increase the jvm settings of the job runner using specific JVM arguments (-Xms1024m and -Xmx9208m). I have tried with Talend Open Studio 5.6.2 MDM edition and 6.3.0 BigData edition. The computer I use has an SSD hard disk and a total RAM of 16Gb. After 6 hours of running the job, it is still in the status "Starting". The CPU usage is 100%. The memory usage is 14.6Gb.

It is important to mention that I use the generation mode "fast, with low memory consumption SAX".

This is the xml structure that I have use to create the structure in the metadata:

```
<?xml version='1.0' encoding='UTF-8'?>
<m:GenericData xmlns:footer="http://www.sdmx.org/resources/sdmxml/schemas/v2_1/message/footer" xmlns:g="http://www.sdmx.org/resources/sdmxml/schemas/v2_1/data/generic"
xmlns:c="http://www.sdmx.org/resources/sdmxml/schemas/v2_1/common" xmlns:m="http://www.sdmx.org/resources/sdmxml/schemas/v2_1/message"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:xml="http://www.w3.org/XML/1998/namespace">
  <m:Header>
    <m:ID>DATASET_ID_480609846268</m:ID>
    <m:Test>>false</m:Test>
    <m:repared>2016-12-01T16:30:46</m:repared>
    <m:ender id="TEST">
      <c:Name xml:lang="de">TITLE ONE</c:Name>
      <c:Name xml:lang="en">TITLE TWO</c:Name>
      <c:Name xml:lang="fr">TITLE THREE</c:Name>
    </m:ender>
    <m:structure structureID="STR_ID_1_0" dimensionAtObservation="TIME_PERIOD">
      <c:structureUsage>
        <Ref agencyID="TEST" id="AG_ID" version="1.0"/>
      </c:structureUsage>
    </m:structure>
    <m:ataSetID>DATASET_ID</m:ataSetID>
    <m:Extracted>2016-12-01T16:30:46</m:Extracted>
    <m:EmbargoDate>2016-12-02T10:00:00</m:EmbargoDate>
  </m:Header>
  <m:ataSet structureRef="STR_ID_1_0">
    <g:eries>
      <g:eriesKey>
        <g:Value id="FREQ" value="A"/>
        <g:Value id="CURRENCY" value="MIO_EUR"/>
        <g:Value id="BOP_ITEM" value="CA"/>
        <g:Value id="SECTOR10" value="S1"/>
        <g:Value id="SECTPART" value="S1"/>
        <g:Value id="STK_FLOW" value="BAL"/>
        <g:Value id="PARTNER" value="AT"/>
        <g:Value id="GEO" value="MT"/>
      </g:eriesKey>
      <g:bs>
        <g:bsDimension value="2016"/>
        <g:bsValue/>
        <g:Attributes>
          <g:Value id="OBS_FLAG" value="c"/>
        </g:Attributes>
      </g:bs>
    </g:eries>
  </m:ataSet>
</m:GenericData>
```

And it works perfectly with a 6500 records file, but when I change to the 4Gb one, nothing. I have tried 2 types of jobs, all of them failed in reading (jobs keep in starting)

tFileinputXML-> tMAP -> tfileOutputDelimited

tFileinputXML-> tlogRow

If I use lower memory arguments (1Gb or 2Gb as maximum), I get an out of memory exception, but the message talks about Xerces and SAX:

Starting job test_001 at 00:31 11/12/2016.

connecting to socket on port 3947

connected

Exception in thread "main" java.lang.OutOfMemoryError: Java heap space

at java.util.Arrays.copyOf(Unknown Source)

at java.util.Arrays.copyOf(Unknown Source)

at java.util.ArrayList.grow(Unknown Source)

at java.util.ArrayList.ensureExplicitCapacity(Unknown Source)

at java.util.ArrayList.ensureCapacityInternal(Unknown Source)

at java.util.ArrayList.add(Unknown Source)

at org.talend.xml.sax.SAXLoopHandler.endElement(SAXLoopHandler.java:308)

at org.talend.xml.sax.SAXLoopCompositeHandler.endElement(SAXLoopCompositeHandler.java:86)

at org.apache.xerces.parsers.AbstractSAXParser.endElement(Unknown Source)

at org.apache.xerces.impl.XMLDocumentFragmentScannerImpl.scanEndElement(Unknown Source)

at org.apache.xerces.impl.XMLDocumentFragmentScannerImpl\$FragmentContentDispatcher.dispatch(Unknown Source)

at org.apache.xerces.impl.XMLDocumentFragmentScannerImpl.scanDocument(Unknown Source)

at org.apache.xerces.parsers.XML11Configuration.parse(Unknown Source)

at org.apache.xerces.parsers.XML11Configuration.parse(Unknown Source)

disconnected

at org.apache.xerces.parsers.XMLParser.parse(Unknown Source)

at org.apache.xerces.parsers.AbstractSAXParser.parse(Unknown Source)

at org.apache.xerces.jaxp.SAXParserImpl\$JAXPSAXParser.parse(Unknown Source)

at org.apache.xerces.jaxp.SAXParserImpl.parse(Unknown Source)

at org.talend.xml.sax.ComplexSAXLooper.parse(ComplexSAXLooper.java:159)

at org.talend.xml.sax.SAXLooper.parse(SAXLooper.java:160)

at ref2diss.test_001_0_1.test_001.tFileinputXML_1Process(test_001.java:850)

at ref2diss.test_001_0_1.test_001.runJobInTOS(test_001.java:1690)

at ref2diss.test_001_0_1.test_001.main(test_001.java:1547)

Job test_001 ended at 00:48 11/12/2016.

I hope this can provide you enough insight. Please, let me know if you need more information.

Thanks for your help and support