

Troubleshooting Guide

- `ClassNotFoundException` for commons-math3
- `OutOfMemoryError` in Hadoop Reduce Phase

ClassNotFoundException for commons-math3

The Apache Commons Math library is utilized by SystemML. The commons-math3 dependency is included with Spark and with newer versions of Hadoop. Running SystemML on an older Hadoop cluster can potentially generate an error such as the following due to the missing commons-math3 dependency:

```
java.lang.ClassNotFoundException: org.apache.commons.math3.linear.RealMatrix
```

This issue can be fixed by changing the commons-math3 scope in the pom.xml file from provided to compile.

```
<dependency>
  <groupId>org.apache.commons</groupId>
  <artifactId>commons-math3</artifactId>
  <version>3.1.1</version>
  <scope>compile</scope>
</dependency>
```

SystemML can then be rebuilt with the commons-math3 dependency using Maven (`mvn clean package -P distribution`).

OutOfMemoryError in Hadoop Reduce Phase

In Hadoop MapReduce, outputs from mapper nodes are copied to reducer nodes and then sorted (known as the *shuffle* phase) before being consumed by reducers. The shuffle phase utilizes several buffers that share memory space with other MapReduce tasks, which will throw an `OutOfMemoryError` if the shuffle buffers take too much space:

```
Error: java.lang.OutOfMemoryError: Java heap space
    at org.apache.hadoop.mapred.IFile$Reader.readNextBlock(IFile.java:357)
    at org.apache.hadoop.mapred.IFile$Reader.next(IFile.java:419)
    at org.apache.hadoop.mapred.Merger$Segment.next(Merger.java:238)
    at org.apache.hadoop.mapred.Merger$MergeQueue.adjustPriorityQueue(Merger.java:348)
    at org.apache.hadoop.mapred.Merger$MergeQueue.next(Merger.java:368)
    at org.apache.hadoop.mapred.Merger.writeFile(Merger.java:156)
    ...
```

One way to fix this issue is lowering the following buffer thresholds.

```
mapred.job.shuffle.input.buffer.percent # default 0.70; try 0.20
mapred.job.shuffle.merge.percent # default 0.66; try 0.20
mapred.job.reduce.input.buffer.percent # default 0.0; keep 0.0
```

These configurations can be modified **globally** by inserting/modifying the following in `mapred-site.xml`.

```
<property>
  <name>mapred.job.shuffle.input.buffer.percent</name>
  <value>0.2</value>
</property>
<property>
  <name>mapred.job.shuffle.merge.percent</name>
  <value>0.2</value>
</property>
<property>
  <name>mapred.job.reduce.input.buffer.percent</name>
  <value>0.0</value>
</property>
```

They can also be configured on a **per SystemML-task basis** by inserting the following in `SystemML-config.xml`.

```
<mapred.job.shuffle.merge.percent>0.2</mapred.job.shuffle.merge.percent>
<mapred.job.shuffle.input.buffer.percent>0.2</mapred.job.shuffle.input.buffer.percent>
<mapred.job.reduce.input.buffer.percent>0</mapred.job.reduce.input.buffer.percent>
```

Note: The default `SystemML-config.xml` is located in `<path to SystemML root>/conf/`. It is passed to SystemML using the `-config` argument:

```
hadoop jar SystemML.jar [-? | -help | -f <filename>] (-config=<config_filename>) ([-args
| -nargs] <args-list>)
```

See [Invoking SystemML in Hadoop Batch Mode \(hadoop-batch-mode.html\)](#) for details of the syntax.