Item 142
**git_comments:**

**git_commits:**

1. **summary:** [SPARK-7672] [CORE] Use int conversion in translating kryoserializer.buffer.mb to kryoserializer.buffer
   **message:** [SPARK-7672] [CORE] Use int conversion in translating kryoserializer.buffer.mb to kryoserializer.buffer In translating spark.kryoserializer.buffer.mb to spark.kryoserializer.buffer, use of toDouble will lead to "Fractional values not supported" error even when spark.kryoserializer.buffer.mb is an integer. ilganeli, andrewor14 Author: Nishkam Ravi <nravi@cloudera.com> Author: nishkamravi2 <nishkamravi@gmail.com> Author: nravi <nravi@c1704.halxg.cloudera.com> Closes #6198 from nishkamravi2/master_nravi and squashes the following commits: 171a53c [nishkamravi2] Update SparkConfSuite.scala 5261bf6 [Nishkam Ravi] Add a test for deprecated config spark.kryoserializer.buffer.mb 5190f79 [Nishkam Ravi] In translating from deprecated spark.kryoserializer.buffer.mb to spark.kryoserializer.buffer use int conversion since fractions are not permissible 059ce82 [Nishkam Ravi] Merge branch 'master' of https://github.com/apache/spark into master_nravi eaa13b5 [nishkamravi2] Update Client.scala 981afd2 [Nishkam Ravi] Check for read permission before initiating copy 1b81383 [Nishkam Ravi] Merge branch 'master' of https://github.com/apache/spark into master_nravi 0f1abd0 [nishkamravi2] Update Utils.scala 474e3bf [nishkamravi2] Update DiskBlockManager.scala 97c383e [nishkamravi2] Update Utils.scala 8691e0c [Nishkam Ravi] Add a try/catch block around Utils.removeShutdownHook 2be1e76 [Nishkam Ravi] Merge branch 'master' of https://github.com/apache/spark into master_nravi 1c13b79 [Nishkam Ravi] Merge branch 'master' of https://github.com/apache/spark into master_nravi bad4349 [nishkamravi2] Update Main.java 36a6f87 [Nishkam Ravi] Minor changes and bug fixes b7f4ae7 [Nishkam Ravi] Merge branch 'master' of https://github.com/apache/spark into master_nravi 4a45d6a [Nishkam Ravi] Merge branch 'master' of https://github.com/apache/spark into master_nravi 458af39 [Nishkam Ravi] Locate the jar using getLocation, obviates the need to pass assembly path as an argument d9658d6 [Nishkam Ravi] Changes for SPARK-6406 ccdc334 [Nishkam Ravi] Merge branch 'master' of https://github.com/apache/spark into master_nravi 3faa7a4 [Nishkam Ravi] Launcher library changes (SPARK-6406) 345206a [Nishkam Ravi] spark-class merge Merge branch 'master_nravi' of https://github.com/nishkamravi2/spark into master_nravi ac58975 [Nishkam Ravi] spark-class changes 06bfeb0 [nishkamravi2] Update spark-class 35af990 [Nishkam Ravi] Merge branch 'master' of https://github.com/apache/spark into master_nravi 32c3ab3 [nishkamravi2] Update AbstractCommandBuilder.java 4bd4489 [nishkamravi2] Update AbstractCommandBuilder.java 746f35b [Nishkam Ravi] "hadoop" string in the assembly name should not be mandatory (everywhere else in spark we mandate spark-assembly*hadoop*.jar) bfe96e0 [Nishkam Ravi] Merge branch 'master' of https://github.com/apache/spark into master_nravi ee902fa [Nishkam Ravi] Merge branch 'master' of https://github.com/apache/spark into master_nravi d453197 [nishkamravi2] Update NewHadoopRDD.scala 6f41a1d [nishkamravi2] Update NewHadoopRDD.scala 0ce2c32 [nishkamravi2] Update HadoopRDD.scala f7e33c2 [Nishkam Ravi] Merge branch 'master_nravi' of https://github.com/nishkamravi2/spark into master_nravi ba1eb8b [Nishkam Ravi] Try-catch block around the two occurrences of removeShutDownHook. Deletion of semi-redundant occurrences of expensive operation inShutDown. 71d0e17 [Nishkam Ravi] Merge branch 'master' of https://github.com/apache/spark into master_nravi 494d8c0 [nishkamravi2] Update DiskBlockManager.scala 3c5ddba [nishkamravi2] Update DiskBlockManager.scala f0d12de [Nishkam Ravi] Workaround for IllegalStateException caused by recent changes to BlockManager.stop 79ea8b4 [Nishkam Ravi] Merge branch 'master' of https://github.com/apache/spark into master_nravi b446edc [Nishkam Ravi] Merge branch 'master' of https://github.com/apache/spark into master_nravi 5c9a4cb [nishkamravi2] Update TaskSetManagerSuite.scala 535295a [nishkamravi2] Update TaskSetManager.scala 3e1b616 [Nishkam Ravi] Modify test for maxResultSize 9f6583e [Nishkam Ravi] Changes to maxResultSize code (improve error message and add condition to check if maxResultSize > 0) 5f8f9ed [Nishkam Ravi] Merge branch 'master' of https://github.com/apache/spark into master_nravi 636a9ff [nishkamravi2] Update YarnAllocator.scala 8f76c8b [Nishkam Ravi] Doc change for yarn memory overhead 35daa64 [Nishkam Ravi] Slight change in the doc for yarn memory overhead 5ac2ec1 [Nishkam Ravi] Remove out dac1047 [Nishkam Ravi] Additional documentation for yarn memory overhead issue 42c2c3d [Nishkam Ravi] Additional changes for yarn memory overhead issue 362da5e [Nishkam Ravi] Additional changes for yarn memory overhead c726bd9 [Nishkam Ravi] Merge branch

'master' of https://github.com/apache/spark into master_nravi f00fa31 [Nishkam Ravi] Improving logging for AM memoryOverhead 1cf2d1e [nishkamravi2] Update YarnAllocator.scala ebcde10 [Nishkam Ravi] Modify default YARN memory_overhead-- from an additive constant to a multiplier (redone to resolve merge conflicts) 2e69f11 [Nishkam Ravi] Merge branch 'master' of https://github.com/apache/spark into master_nravi efd688a [Nishkam Ravi] Merge branch 'master' of https://github.com/apache/spark 2b630f9 [nravi] Accept memory input as "30g", "512M" instead of an int value, to be consistent with rest of Spark 3bf8fad [nravi] Merge branch 'master' of https://github.com/apache/spark 5423a03 [nravi] Merge branch 'master' of https://github.com/apache/spark eb663ca [nravi] Merge branch 'master' of https://github.com/apache/spark df2aeb1 [nravi] Improved fix for ConcurrentModificationIssue (Spark-1097, Hadoop-10456) 6b840f0 [nravi] Undo the fix for SPARK-1758 (the problem is fixed) 5108700 [nravi] Fix in Spark for the Concurrent thread modification issue (SPARK-1097, HADOOP-10456) 681b36f [nravi] Fix for SPARK-1758: failing test org.apache.spark.JavaAPISuite.wholeTextFiles

**github_issues:**

**github_issues_comments:**

**github_pulls:**

1. **title:** [SPARK-7672][CORE] Use int conversion in translating kryoserializer.buffer.mb to kryoserializer.buffer
   **body:** In translating spark.kryoserializer.buffer.mb to spark.kryoserializer.buffer, use of toDouble will lead to "Fractional values not supported" error even when spark.kryoserializer.buffer.mb is an integer. @ilganeli, @andrewor14

**github_pulls_comments:**

1. Can one of the admins verify this patch?
2. ok to test
3. @nishkamravi2 Good catch. Thanks for fixing it. Can you verify @ilganeli?
4. Merged build triggered.
5. Merged build started.
6. [Test build #32845 has started] (https://amplab.cs.berkeley.edu/jenkins/job/SparkPullRequestBuilder/32845/consoleFull) for PR 6198 at commit [`5190f79`] (https://github.com/apache/spark/commit/5190f79e663d44d07ec4b2081f288f8cfc6d7bec).
7. [Test build #32845 has finished] (https://amplab.cs.berkeley.edu/jenkins/job/SparkPullRequestBuilder/32845/consoleFull) for PR 6198 at commit [`5190f79`] (https://github.com/apache/spark/commit/5190f79e663d44d07ec4b2081f288f8cfc6d7bec). - This patch **fails Spark unit tests**. - This patch merges cleanly. - This patch adds no public classes.
8. Merged build finished. Test FAILed.
9. Test FAILed. Refer to this link for build results (access rights to CI server needed): https://amplab.cs.berkeley.edu/jenkins/job/SparkPullRequestBuilder/32845/ Test FAILed.
10. @andrewor14 @nishkamravi2 Good catch. The fix looks good. Thanks!
11. Thanks for verifying @andrewor14 @ilganeli. Looked at the failed tests and they seem unrelated.
12. Yes, pretty sure they're not related. retest this please
13. LGTM, but could you add a test for this? Something like this, in `SparkConfSuite::test("deprecated configs")`: ``` conf.set("spark.kryoserializer.buffer.max.mb", "1.1") assert(conf.getSizeAsKb("spark.kryoserializer.buffer") === 1100) ```
14. Merged build triggered.
15. Merged build started.
16. [Test build #32853 has started] (https://amplab.cs.berkeley.edu/jenkins/job/SparkPullRequestBuilder/32853/consoleFull) for PR 6198 at commit [`5190f79`] (https://github.com/apache/spark/commit/5190f79e663d44d07ec4b2081f288f8cfc6d7bec).
17. Good suggestion @vanzin.
18. Merged build triggered.
19. Merged build started.

20. [Test build #32855 has started] (https://amplab.cs.berkeley.edu/jenkins/job/SparkPullRequestBuilder/32855/consoleFull) for PR 6198 at commit [`5261bf6`] (https://github.com/apache/spark/commit/5261bf6922579f31b480452ac3e7b0cb36636cb8).
21. Merged build triggered.
22. Merged build started.
23. [Test build #32858 has started] (https://amplab.cs.berkeley.edu/jenkins/job/SparkPullRequestBuilder/32858/consoleFull) for PR 6198 at commit [`171a53c`] (https://github.com/apache/spark/commit/171a53cd443a4283ea632af4e74c5eae674a5e4c).
24. [Test build #32855 has finished] (https://amplab.cs.berkeley.edu/jenkins/job/SparkPullRequestBuilder/32855/consoleFull) for PR 6198 at commit [`5261bf6`] (https://github.com/apache/spark/commit/5261bf6922579f31b480452ac3e7b0cb36636cb8). - This patch **fails Spark unit tests**. - This patch merges cleanly. - This patch adds no public classes.
25. Merged build finished. Test FAILed.
26. Test FAILed. Refer to this link for build results (access rights to CI server needed): https://amplab.cs.berkeley.edu/jenkins/job/SparkPullRequestBuilder/32855/ Test FAILed.
27. [Test build #32853 has finished] (https://amplab.cs.berkeley.edu/jenkins/job/SparkPullRequestBuilder/32853/consoleFull) for PR 6198 at commit [`5190f79`] (https://github.com/apache/spark/commit/5190f79e663d44d07ec4b2081f288f8cfc6d7bec). - This patch **passes all tests**. - This patch merges cleanly. - This patch adds no public classes.
28. Merged build finished. Test PASSed.
29. Test PASSed. Refer to this link for build results (access rights to CI server needed): https://amplab.cs.berkeley.edu/jenkins/job/SparkPullRequestBuilder/32853/ Test PASSed.
30. [Test build #32858 has finished] (https://amplab.cs.berkeley.edu/jenkins/job/SparkPullRequestBuilder/32858/consoleFull) for PR 6198 at commit [`171a53c`] (https://github.com/apache/spark/commit/171a53cd443a4283ea632af4e74c5eae674a5e4c). - This patch **passes all tests**. - This patch merges cleanly. - This patch adds no public classes.
31. Merged build finished. Test PASSed.
32. Test PASSed. Refer to this link for build results (access rights to CI server needed): https://amplab.cs.berkeley.edu/jenkins/job/SparkPullRequestBuilder/32858/ Test PASSed.

**github_pulls_reviews:**

1. I expect the test to fail, but my example code was bad (copy & pasted the wrong property). This should be `spark.kryoserializer.buffer.mb`. Sorry about that.
2. And I thought I had verified your example code, skipped both of us.

**jira_issues:**

1. **summary:** Bug in Configuration.java exposed by Spark (ConcurrentModificationException)
   **description:** The following exception occurs non-deterministically:
   java.util.ConcurrentModificationException at
   java.util.HashMap$HashIterator.nextEntry(HashMap.java:926) at
   java.util.HashMap$KeyIterator.next(HashMap.java:960) at
   java.util.AbstractCollection.addAll(AbstractCollection.java:341) at java.util.HashSet.<init>
   (HashSet.java:117) at org.apache.hadoop.conf.Configuration.<init>(Configuration.java:671) at
   org.apache.hadoop.mapred.JobConf.<init>(JobConf.java:439) at
   org.apache.spark.rdd.HadoopRDD.getJobConf(HadoopRDD.scala:110) at
   org.apache.spark.rdd.HadoopRDD$$anon$1.<init>(HadoopRDD.scala:154) at
   org.apache.spark.rdd.HadoopRDD.compute(HadoopRDD.scala:149) at
   org.apache.spark.rdd.HadoopRDD.compute(HadoopRDD.scala:64) at
   org.apache.spark.rdd.RDD.computeOrReadCheckpoint(RDD.scala:241) at
   org.apache.spark.rdd.RDD.iterator(RDD.scala:232) at
   org.apache.spark.rdd.MappedRDD.compute(MappedRDD.scala:31) at
   org.apache.spark.rdd.RDD.computeOrReadCheckpoint(RDD.scala:241) at

```
org.apache.spark.rdd.RDD.iterator(RDD.scala:232) at
org.apache.spark.rdd.MappedRDD.compute(MappedRDD.scala:31) at
org.apache.spark.rdd.RDD.computeOrReadCheckpoint(RDD.scala:241) at
org.apache.spark.rdd.RDD.iterator(RDD.scala:232) at
org.apache.spark.rdd.MappedRDD.compute(MappedRDD.scala:31) at
org.apache.spark.rdd.RDD.computeOrReadCheckpoint(RDD.scala:241) at
org.apache.spark.rdd.RDD.iterator(RDD.scala:232) at
org.apache.spark.rdd.MapPartitionsRDD.compute(MapPartitionsRDD.scala:34) at
org.apache.spark.rdd.RDD.computeOrReadCheckpoint(RDD.scala:241) at
org.apache.spark.rdd.RDD.iterator(RDD.scala:232) at
org.apache.spark.scheduler.ShuffleMapTask.runTask(ShuffleMapTask.scala:161) at
org.apache.spark.scheduler.ShuffleMapTask.runTask(ShuffleMapTask.scala:102) at
org.apache.spark.scheduler.Task.run(Task.scala:53) at
org.apache.spark.executor.Executor$TaskRunner$$anonfun$run$1.apply$mcV$sp(Executor.scala:213) at
org.apache.spark.deploy.SparkHadoopUtil$$anon$1.run(SparkHadoopUtil.scala:42) at
org.apache.spark.deploy.SparkHadoopUtil$$anon$1.run(SparkHadoopUtil.scala:41) at
java.security.AccessController.doPrivileged(Native Method) at
javax.security.auth.Subject.doAs(Subject.java:415) at
org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1548) at
org.apache.spark.deploy.SparkHadoopUtil.runAsUser(SparkHadoopUtil.scala:41) at
org.apache.spark.executor.Executor$TaskRunner.run(Executor.scala:178) at
java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.java:1145) at
java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.java:615) at
java.lang.Thread.run(Thread.java:744)
```

**jira_issues_comments:**

1. Patch attached.
2. Fix is to move the HashSet initialization to the synchronized block right above it. This patch can potentially by applied to all Hadoop branches.
3. {color:red}-1 overall{color}. Here are the results of testing the latest attachment http://issues.apache.org/jira/secure/attachment/12638356/HADOOP-10456_nravi.patch against trunk revision . {color:green}+1 @author{color}. The patch does not contain any @author tags. {color:red}-1 tests included{color}. The patch doesn't appear to include any new or modified tests. Please justify why no new tests are needed for this patch. Also please list what manual steps were performed to verify this patch. {color:green}+1 javac{color}. The applied patch does not increase the total number of javac compiler warnings. {color:green}+1 javadoc{color}. There were no new javadoc warning messages. {color:green}+1 eclipse:eclipse{color}. The patch built with eclipse:eclipse. {color:green}+1 findbugs{color}. The patch does not introduce any new Findbugs (version 1.3.9) warnings. {color:green}+1 release audit{color}. The applied patch does not increase the total number of release audit warnings. {color:green}+1 core tests{color}. The patch passed unit tests in hadoop-common-project/hadoop-common. {color:green}+1 contrib tests{color}. The patch passed contrib unit tests. Test results: https://builds.apache.org/job/PreCommit-HADOOP-Build/3738//testReport/ Console output: https://builds.apache.org/job/PreCommit-HADOOP-Build/3738//console This message is automatically generated.
4. This is a fix for a non-deterministic bug (a thread synchronization issue). A unit test case for this fix would be flaky. The verification was done by (i) running mvn compile/install/test, (ii) running Spark benchmarks and making sure there are no performance regressions
5. Thank you for taking this JIRA, [~nravi]. +1(non-binding), reviewed and confirmed it's a only critical section about finalParameters without lock within Configuration instance. [~cnauroth], can you take a look, please?
6. +1 for the patch. Thank you, Nishkam and Tsuyoshi. I can't commit this right now. I'll aim to commit tomorrow. If I don't commit it in the next few days, please ping me again in case I forget. :-)
7. Thanks Tsuyoshi and Chris.
8. SUCCESS: Integrated in Hadoop-trunk-Commit #5456 (See [https://builds.apache.org/job/Hadoop-trunk-Commit/5456/]) HADOOP-10456. Bug in Configuration.java exposed by Spark (ConcurrentModificationException). Contributed by Nishkam Ravi. (cnauroth: http://svn.apache.org/viewcvs.cgi/?root=Apache-SVN&view=rev&rev=1584575) * /hadoop/common/trunk/hadoop-common-project/hadoop-common/CHANGES.txt *

/hadoop/common/trunk/hadoop-common-project/hadoop-common/src/main/java/org/apache/hadoop/conf/Configuration.java

9. I committed this to trunk, branch-2 and branch-2.4. Nishkam, thank you for contributing this bug fix. Tsuyoshi, thank you for the code review.

10. SUCCESS: Integrated in Hadoop-Yarn-trunk #529 (See [https://builds.apache.org/job/Hadoop-Yarn-trunk/529/]) HADOOP-10456. Bug in Configuration.java exposed by Spark (ConcurrentModificationException). Contributed by Nishkam Ravi. (cnauroth: http://svn.apache.org/viewcvs.cgi/?root=Apache-SVN&view=rev&rev=1584575) * /hadoop/common/trunk/hadoop-common-project/hadoop-common/CHANGES.txt * /hadoop/common/trunk/hadoop-common-project/hadoop-common/src/main/java/org/apache/hadoop/conf/Configuration.java

11. FAILURE: Integrated in Hadoop-Mapreduce-trunk #1747 (See [https://builds.apache.org/job/Hadoop-Mapreduce-trunk/1747/]) HADOOP-10456. Bug in Configuration.java exposed by Spark (ConcurrentModificationException). Contributed by Nishkam Ravi. (cnauroth: http://svn.apache.org/viewcvs.cgi/?root=Apache-SVN&view=rev&rev=1584575) * /hadoop/common/trunk/hadoop-common-project/hadoop-common/CHANGES.txt * /hadoop/common/trunk/hadoop-common-project/hadoop-common/src/main/java/org/apache/hadoop/conf/Configuration.java

12. FAILURE: Integrated in Hadoop-Hdfs-trunk #1721 (See [https://builds.apache.org/job/Hadoop-Hdfs-trunk/1721/]) HADOOP-10456. Bug in Configuration.java exposed by Spark (ConcurrentModificationException). Contributed by Nishkam Ravi. (cnauroth: http://svn.apache.org/viewcvs.cgi/?root=Apache-SVN&view=rev&rev=1584575) * /hadoop/common/trunk/hadoop-common-project/hadoop-common/CHANGES.txt * /hadoop/common/trunk/hadoop-common-project/hadoop-common/src/main/java/org/apache/hadoop/conf/Configuration.java