

Item 95

git_comments:

git_commits:

1. **summary:** [SPARK-9651] Fix UnsafeExternalSorterSuite.
message: [SPARK-9651] Fix UnsafeExternalSorterSuite. First, it's probably a bad idea to call generated Scala methods from Java. In this case, the method being called wasn't actually "Utils.createTempDir()", but actually the method that returns the first default argument to the actual createTempDir method, which is just the location of java.io.tmpdir; meaning that all tests in the class were using the same temp dir, and thus affecting each other. Second, spillingOccursInResponseToMemoryPressure was not writing enough records to actually cause a spill. Author: Marcelo Vanzin <vanzin@cloudera.com> Closes #7970 from vanzin/SPARK-9651 and squashes the following commits: 74d357f [Marcelo Vanzin] Clean up temp dir on test tear down. a64f36a [Marcelo Vanzin] [SPARK-9651] Fix UnsafeExternalSorterSuite. (cherry picked from commit 4399b7b0903d830313ab7e69731c11d587ae567c) Signed-off-by: Reynold Xin <rxin@databricks.com>

github_issues:

github_issues_comments:

github_pulls:

1. **title:** [SPARK-9651] Fix UnsafeExternalSorterSuite.
body: First, it's probably a bad idea to call generated Scala methods from Java. In this case, the method being called wasn't actually "Utils.createTempDir()", but actually the method that returns the first default argument to the actual createTempDir method, which is just the location of java.io.tmpdir; meaning that all tests in the class were using the same temp dir, and thus affecting each other. Second, spillingOccursInResponseToMemoryPressure was not writing enough records to actually cause a spill.

github_pulls_comments:

1. /cc @rxin @JoshRosen
2. Seems fine to me, pending Jenkins.
3. [Test build #39903 has finished](https://amplab.cs.berkeley.edu/jenkins/job/SparkPullRequestBuilder/39903/console) for PR 7970 at commit [a64f36a] (https://github.com/apache/spark/commit/a64f36a13f5ee6cdeaa09f65e55b8a08bd0936b2). - This patch **fails Spark unit tests**. - This patch merges cleanly. - This patch adds no public classes.
4. Well the test being changed did pass... retest this please.
5. **[Test build #39909 timed out](https://amplab.cs.berkeley.edu/jenkins/job/SparkPullRequestBuilder/39909/console)** for PR 7970 at commit [74d357f] (https://github.com/apache/spark/commit/74d357fa8f60318d795da8a5f6ce601cad33a731) after a configured wait of `175m`.
6. [Test build #239 has finished](https://amplab.cs.berkeley.edu/jenkins/job/SlowSparkPullRequestBuilder/239/console) for PR 7970 at commit [74d357f] (https://github.com/apache/spark/commit/74d357fa8f60318d795da8a5f6ce601cad33a731). - This patch **fails PySpark unit tests**. - This patch merges cleanly. - This patch adds the following public classes _(experimental)_ : `class BlockMatrix(DistributedMatrix);` - `// class DataFrame`
7. I'm going to merge this.
8. [Test build #39930 has finished](https://amplab.cs.berkeley.edu/jenkins/job/SparkPullRequestBuilder/39930/console) for PR 7970 at commit [74d357f] (https://github.com/apache/spark/commit/74d357fa8f60318d795da8a5f6ce601cad33a731). - This patch **fails PySpark unit tests**. - This patch merges cleanly. - This patch adds no public classes.

github_pulls_reviews:

1. This has to be deleted in `tearDown` too I presume in order to fix this? or else it's still reusing a same dir? I checked and there is a similar problem in some other Java tests: `` JavaSaveLoadSuite JavaMetastoreDataSourcesSuite JavaAPISuite ``
2. In this case no; setUp and tearDown are per-test, so each test gets a separate temp dir. I'll take a look at the other tests. Might be worth it to add the temp dir helper method in some common java class.
3. Oh right it makes a subdir every time anyway. Well, it'd be better to clean up the temp dir in `tearDown` if possible.
4. This didn't fail for me locally, which is kind of odd.
5. Fails for me consistently if I use `100000`. `` Running org.apache.spark.util.collection.unsafe.sort.UnsafeExternalSorterSuite Tests run: 2, Failures: 2, Errors: 0, Skipped: 0, Time elapsed: 0.99 sec <<< FAILURE! - in org.apache.spark.util.collection.unsafe.sort.UnsafeExternalSorterSuite
spillingOccursInResponseToMemoryPressure(org.apache.spark.util.collection.unsafe.sort.UnsafeExternalSorterSuite) Time elapsed: 0.426 sec <<< FAILURE!
java.lang.AssertionError: Expected: a value equal to or greater than <1> got: <0> at org.junit.Assert.assertThat(Assert.java:780) at
org.junit.Assert.assertThat(Assert.java:738) at
org.apache.spark.util.collection.unsafe.sort.UnsafeExternalSorterSuite.spillingOccursInResponseToMemoryPressure(UnsafeExternalSorterSuite.java:243)
spillingOccursInResponseToMemoryPressure(org.apache.spark.util.collection.unsafe.sort.UnsafeExternalSorterSuite) Time elapsed: 0.429 sec <<< FAILURE!
java.lang.AssertionError: expected:<0> but was:<69206016> at org.junit.Assert.fail(Assert.java:93) at org.junit.Assert.failNotEquals(Assert.java:647) at
org.junit.Assert.assertEquals(Assert.java:128) at org.junit.Assert.assertEquals(Assert.java:472) at org.junit.Assert.assertEquals(Assert.java:456) at
org.apache.spark.util.collection.unsafe.sort.UnsafeExternalSorterSuite.tearDown(UnsafeExternalSorterSuite.java:150) Results : Failed tests:
org.apache.spark.util.collection.unsafe.sort.UnsafeExternalSorterSuite.spillingOccursInResponseToMemoryPressure(org.apache.spark.util.collection.unsafe.sort.UnsafeExternalSorterSuite) Time elapsed: 0.426 sec <<< FAILURE!
Run 1: UnsafeExternalSorterSuite.spillingOccursInResponseToMemoryPressure:243 Expected: a value equal to or greater than <1> got: <0> Run 2:
UnsafeExternalSorterSuite.tearDown:150 expected:<0> but was:<69206016> ``
6. Ok, I'll do this. I'll fix this test this way, then file a separate bug to have a shared, Java createTempDir() and use it from everywhere. How does that sound? This is the only unit test that consistently fails in our builds so I'd like to at least unblock it.
7. Actually all the other tests you mention do the right thing. This is the only place that was calling the wrong Scala method.

jira_issues:

jira_issues_comments: