

git_comments:

git_commits:

1. **summary:** [SPARK-7970] Skip closure cleaning for SQL operations
message: [SPARK-7970] Skip closure cleaning for SQL operations Also introduces new spark private API in RDD.scala with name 'mapPartitionsInternal' which doesn't closure cleans the RDD elements. Author: nitin goyal <nitin.goyal@guavus.com> Author: nitin.goyal <nitin.goyal@guavus.com> Closes #9253 from nitin2goyal/master.

github_issues:

github_issues_comments:

github_pulls:

1. **title:** [SPARK-7970] Skip closure cleaning for SQL operations
body: Also introduces new spark private API in RDD.scala with name 'mapPartitionsInternal' which doesn't closure cleans the RDD elements.

github_pulls_comments:

1. cc @andrewor14
2. ok to test
3. ****[Test build #44238 has finished]**
(https://amplab.cs.berkeley.edu/jenkins/job/SparkPullRequestBuilder/44238/consoleFull)** for PR 9253 at commit [`ca487cb`](https://github.com/apache/spark/commit/ca487cbae6ba4eb2d14d7b007eb54ccc4dd3ee3a). - This patch ****fails to build****. - This patch merges cleanly. - This patch adds no public classes.
4. ****[Test build #44284 has finished]**
(https://amplab.cs.berkeley.edu/jenkins/job/SparkPullRequestBuilder/44284/consoleFull)** for PR 9253 at commit [`6a9f738`](https://github.com/apache/spark/commit/6a9f738bb3008cad7ce855fd33115fbb29d1c0a). - This patch ****fails Spark unit tests****. - This patch merges cleanly. - This patch adds no public classes.
5. @andrewor14 - not sure which test has failed. can we retest this please?
6. retest this please
7. Looks great! I look forward to getting this merged. Once you address the comments I will do so.
8. ****[Test build #44592 has finished]**
(https://amplab.cs.berkeley.edu/jenkins/job/SparkPullRequestBuilder/44592/consoleFull)** for PR 9253 at commit [`6a9f738`](https://github.com/apache/spark/commit/6a9f738bb3008cad7ce855fd33115fbb29d1c0a). - This patch passes all tests. - This patch merges cleanly. - This patch adds no public classes.
9. ****[Test build #44609 has finished]**
(https://amplab.cs.berkeley.edu/jenkins/job/SparkPullRequestBuilder/44609/consoleFull)** for PR 9253 at commit [`36db8a1`](https://github.com/apache/spark/commit/36db8a1f690cc30d027a97f4a307e0aa5e23b5b4). - This patch passes all tests. - This patch merges cleanly. - This patch adds no public classes.
10. Thanks fore reviewing Andrew (@andrewor14). Have addressed your comments. Let me know if it looks good.
11. @nitin2goyal Sorry for the delay. This LGTM. I will merge it once you rebase to master again.
12. ****[Test build #45870 has finished]**
(https://amplab.cs.berkeley.edu/jenkins/job/SparkPullRequestBuilder/45870/consoleFull)** for PR 9253 at commit [`aa4a7ce`](https://github.com/apache/spark/commit/aa4a7ce8f188886d25183ccb2694854fced6ebef). - This patch ****fails Spark unit tests****. - This patch merges cleanly. - This patch adds the following public classes
(experimental):\n * `public class JavaGradientBoostingClassificationExample`\n * `public class JavaGradientBoostingRegressionExample`\n * `public class JavaRandomForestClassificationExample`\n * `public class JavaRandomForestRegressionExample`\n
13. retest this please
14. ****[Test build #45895 has finished]**
(https://amplab.cs.berkeley.edu/jenkins/job/SparkPullRequestBuilder/45895/consoleFull)** for PR 9253 at commit [`aa4a7ce`](https://github.com/apache/spark/commit/aa4a7ce8f188886d25183ccb2694854fced6ebef). - This patch passes all tests. - This patch merges cleanly. - This patch adds no public classes.
15. Should mapPartitions() be replaced with mapPartitionsInternal() in the following classes ? ``
child.execute().mapPartitions { iter =>
/Users/tyu/spark/sql/core/src/main/scala/org/apache/spark/sql/execution/aggregate/TungstenAggregate.scala val
rootType = schemaData.mapPartitions { iter =>
/Users/tyu/spark/sql/core/src/main/scala/org/apache/spark/sql/execution/datasources/json/InferSchema.scala

```

json.mapPartitions { iter =>
/Users/tyu/spark/sql/core/src/main/scala/org/apache/spark/sql/execution/datasources/json/JacksonParser.scala
rows.mapPartitions { iterator =>
/Users/tyu/spark/sql/core/src/main/scala/org/apache/spark/sql/execution/datasources/json/JSONRelation.scala
.mapPartitions { iterator =>
/Users/tyu/spark/sql/core/src/main/scala/org/apache/spark/sql/execution/datasources/parquet/ParquetRelation.scala
.mapPartitions { iter =>
/Users/tyu/spark/sql/core/src/main/scala/org/apache/spark/sql/execution/datasources/text/DefaultSource.scala
child.execute().mapPartitions { iter =>
/Users/tyu/spark/sql/core/src/main/scala/org/apache/spark/sql/execution/debug/package.scala data.mapPartitions
{ iterator => data.mapPartitions { iterator =>
/Users/tyu/spark/sql/core/src/main/scala/org/apache/spark/sql/execution/ExistingRDD.scala
child.execute().mapPartitions { iter =>
/Users/tyu/spark/sql/core/src/main/scala/org/apache/spark/sql/execution/Expand.scala
streamedPlan.execute().mapPartitions { streamedIter =>
/Users/tyu/spark/sql/core/src/main/scala/org/apache/spark/sql/execution/joins/BroadcastHashJoin.scala
streamedPlan.execute().mapPartitions { streamedIter =>
/Users/tyu/spark/sql/core/src/main/scala/org/apache/spark/sql/execution/joins/BroadcastHashOuterJoin.scala val
matchesOrStreamedRowsWithNulls = streamed.execute().mapPartitions { streamedIter =>
/Users/tyu/spark/sql/core/src/main/scala/org/apache/spark/sql/execution/joins/BroadcastNestedLoopJoin.scala
streamed.execute().mapPartitions { streamedIter =>
/Users/tyu/spark/sql/core/src/main/scala/org/apache/spark/sql/execution/joins/LeftSemiJoinBNL.scala
rdd.mapPartitions { iter => inputRDD.mapPartitions { iter =>
/Users/tyu/spark/sql/core/src/main/scala/org/apache/spark/sql/execution/python.scala
child.execute().mapPartitions { iter => child.execute().mapPartitions { iter =>
/Users/tyu/spark/sql/core/src/main/scala/org/apache/spark/sql/execution/rowFormatConverters.scala
child.execute().mapPartitions { iter =>
/Users/tyu/spark/sql/core/src/main/scala/org/apache/spark/sql/execution/sort.scala child.execute().mapPartitions
{ stream => /Users/tyu/spark/sql/core/src/main/scala/org/apache/spark/sql/execution/Window.scala ``` If so,
allow me to open a PR

```

github_pulls_reviews:

1. just use `@param` here
2. can you add that this is mainly for performance improvements? Also you're missing a period at the end.

jira_issues:

jira_issues_comments: