

Item 253

git_comments:

1. * * Marker interface for streaming output formats. For now, this is used to out out of the * optimizations that do not output empty results to avoid empty files.
2. * * Licensed to the Apache Software Foundation (ASF) under one * or more contributor license agreements. See the NOTICE file * distributed with this work for additional information * regarding copyright ownership. The ASF licenses this file * to you under the Apache License, Version 2.0 (the * "License"); you may not use this file except in compliance * with the License. You may obtain a copy of the License at * * <http://www.apache.org/licenses/LICENSE-2.0> * * Unless required by applicable law or agreed to in writing, software * distributed under the License is distributed on an "AS IS" BASIS, * WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. * See the License for the specific language governing permissions and * limitations under the License.
3. **comment:** TODO# this
label: requirement

git_commits:

1. **summary:** HIVE-14071 : HIVE-14014 breaks non-file outputs (Sergey Shelukhin, reviewed by Pengcheng Xiong)
message: HIVE-14071 : HIVE-14014 breaks non-file outputs (Sergey Shelukhin, reviewed by Pengcheng Xiong)

github_issues:

github_issues_comments:

github_pulls:

github_pulls_comments:

github_pulls_reviews:

jira_issues:

1. **summary:** HIVE-14014 breaks non-file outputs
description: Cannot avoid creating outputs when outputs are e.g. streaming
2. **summary:** HIVE-14014 breaks non-file outputs
description: Cannot avoid creating outputs when outputs are e.g. streaming
3. **summary:** HIVE-14014 breaks non-file outputs
description: Cannot avoid creating outputs when outputs are e.g. streaming
4. **summary:** HIVE-14014 breaks non-file outputs
description: Cannot avoid creating outputs when outputs are e.g. streaming
label: code-design
5. **summary:** HIVE-14014 breaks non-file outputs
description: Cannot avoid creating outputs when outputs are e.g. streaming
6. **summary:** HIVE-14014 breaks non-file outputs
description: Cannot avoid creating outputs when outputs are e.g. streaming
7. **summary:** HIVE-14014 breaks non-file outputs
description: Cannot avoid creating outputs when outputs are e.g. streaming
8. **summary:** HIVE-14014 breaks non-file outputs
description: Cannot avoid creating outputs when outputs are e.g. streaming

jira_issues_comments:

1. [~jdere] [~pxiong] can you take a look? thanks
2. [~sershe], thanks for your fast patch and response. So basically, when the file (to be streamed) is empty, your patch will still create the file (the streaming pipeline)? Is there any other way we can improve that so that we do not create the pipeline at all? If I understand correctly, the root cause of the problem is that streaming client is writing data to streaming server but server may wait infinitely when client does not

have anything to write (file is empty). Is it possible to use "push" model for the client to push data to the server rather than current "pull" model for the server to pull data from client? Thanks.

3. This is writer; this is server pushing data to client. If the server never pushes anything or disconnects, the client just sits around forever waiting for data. At least the close() call should be performed, so that the notification is sent to client and/or the socket is closed. Otherwise the client doesn't know whether server is still doing work, or if it's done. So there's no way to avoid the pipeline.
4. **body:** I had a discussion with [~sershe]. The main draw back of current solution in this patch is that we call "createBucketFiles" and we call "closeWriters" to close it immediately. The reason behind is to keep the client from endless waiting. We explored and discussed different other solutions. As far as we know, it seems that we could not find anything better. I would suggest that we could put some comments about this drawback and some "TODO" comments for future improvement in the code. +1. ccing [~ashutoshc]
label: code-design
5. Here are the results of testing the latest attachment:
<https://issues.apache.org/jira/secure/attachment/12812304/HIVE-14071.patch> {color:red}ERROR: {color} -1 due to no test(s) being added or modified. {color:red}ERROR:{color} -1 due to 3 failed/errored test(s), 10247 tests executed *Failed tests:* {noformat} TestMinimrCliDriver-bucketmapjoin6.q-leftsemijoin_mr.q-bucket5.q-and-1-more - did not produce a TEST-*.xml file
org.apache.hadoop.hive.cli.TestCliDriver.testCliDriver_constantPropagateForSubQuery
org.apache.hadoop.hive.cli.TestMiniSparkOnYarnCliDriver.testCliDriver_index_bitmap3 {noformat}
Test results: <https://builds.apache.org/job/PreCommit-HIVE-MASTER-Build/216/testReport> Console output: <https://builds.apache.org/job/PreCommit-HIVE-MASTER-Build/216/console> Test logs: <http://ec2-50-18-27-0.us-west-1.compute.amazonaws.com/logs/PreCommit-HIVE-MASTER-Build-216/>
Messages: {noformat} Executing org.apache.hive.ptest.execution.TestCheckPhase Executing org.apache.hive.ptest.execution.PrePhase Executing org.apache.hive.ptest.execution.ExecutionPhase Executing org.apache.hive.ptest.execution.ReportingPhase Tests exited with: TestsFailedException: 3 tests failed {noformat} This message is automatically generated. ATTACHMENT ID: 12812304 - PreCommit-HIVE-MASTER-Build
6. Committed to branches; thanks for the review!