

Item 212

git_comments:

git_commits:

1. **summary:** [SPARK-23089][STS] Recreate session log directory if it doesn't exist
message: [SPARK-23089][STS] Recreate session log directory if it doesn't exist ## What changes were proposed in this pull request? When creating a session directory, Thrift should create the parent directory (i.e. /tmp/base_session_log_dir) if it is not present. It is common that many tools delete empty directories, so the directory may be deleted. This can cause the session log to be disabled. This was fixed in HIVE-12262: this PR brings it in Spark too. ## How was this patch tested? manual tests Author: Marco Gaido <marcogaido91@gmail.com> Closes #20281 from mgaido91/SPARK-23089.

github_issues:

github_issues_comments:

github_pulls:

1. **title:** [SPARK-23089][STS] Recreate session log directory if it doesn't exist
body: ## What changes were proposed in this pull request? When creating a session directory, Thrift should create the parent directory (i.e. /tmp/base_session_log_dir) if it is not present. It is common that many tools delete empty directories, so the directory may be deleted. This can cause the session log to be disabled. This was fixed in HIVE-12262: this PR brings it in Spark too. ## How was this patch tested? manual tests

github_pulls_comments:

1. ****[Test build #86195 has finished]**
(<https://amplab.cs.berkeley.edu/jenkins/job/SparkPullRequestBuilder/86195/testReport>)** for PR 20281 at commit [`7fc1e00`]
(<https://github.com/apache/spark/commit/7fc1e007a0d97ee5751117f73195a42759554e4c>). * This patch passes all tests. * This patch merges cleanly. * This patch adds no public classes.
2. LGTM. So looks like the fix is exactly the same as Hive.
3. @cloud-fan would you please take another look on this?
4. LGTM, cc @gatorsmile
5. @gatorsmile , do you have further comment?
6. ****[Test build #86385 has finished]**
(<https://amplab.cs.berkeley.edu/jenkins/job/SparkPullRequestBuilder/86385/testReport>)** for PR 20281 at commit [`8b4eb1c`]
(<https://github.com/apache/spark/commit/8b4eb1c33c525ba3eaab79fe1efa4f61fba7367f>). * This patch passes all tests. * This patch merges cleanly. * This patch adds no public classes.
7. thanks, merging to master/2.3!

github_pulls_reviews:

1. **body:** nit: extra space at the end of line
label: code-design
2. sorry @jiangxb1987 I can't see any extra space at the end of this line...
3. **body:** we need an extra space: `... recreating: " +`
label: code-design
4. oh sorry, now I understand, thanks! Yes, in Hive this space is still missing, I see only now. Thanks.

jira_issues:

1. **summary:** Session log dir cannot be created in some cases
description: We see intermittent error in hiveserver2 log: {code} (HiveSessionImpl.java:setOperationLogSessionDir(211)) - Unable to create operation log session directory:xxxxxx {code} User are not able to retrieve their operation log through Hue.

jira_issues_comments:

1. Attach patch. The patch is not aim to solve the issue, but to: 1. Log warning messages for possible reasons session log dir could not be created 1. If the reason is operation root dir is missing, which might be the most possible one, create the operation root dir
2. +1
3. Here are the results of testing the latest attachment:
<https://issues.apache.org/jira/secure/attachment/12768591/HIVE-12262.1.patch> {color:red}ERROR: {color} -1 due to no test(s) being added or modified. {color:red}ERROR:{color} -1 due to 5 failed/errored test(s), 9694 tests executed *Failed tests:* {noformat}
org.apache.hadoop.hive.cli.TestMiniTezCliDriver.initializationError
org.apache.hadoop.hive ql.io.orc.TestColumnStatistics.testHasNull
org.apache.hadoop.hive ql.io.orc.TestJsonFileDump.testJsonDump
org.apache.hive.hcatalog.api.TestHCatClient.testTableSchemaPropagation
org.apache.hive.jdbc.TestSSL.testSSLVersion {noformat} Test results: <http://ec2-174-129-184-35.compute-1.amazonaws.com/jenkins/job/PreCommit-HIVE-TRUNK-Build/5787/testReport> Console output: <http://ec2-174-129-184-35.compute-1.amazonaws.com/jenkins/job/PreCommit-HIVE-TRUNK-Build/5787/console> Test logs: <http://ec2-174-129-184-35.compute-1.amazonaws.com/logs/PreCommit-HIVE-TRUNK-Build-5787/> 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: 5 tests failed {noformat} This message is automatically generated. ATTACHMENT ID: 12768591 - PreCommit-HIVE-TRUNK-Build
4. Patch committed to master and branch-1. Thanks Thejas for review!
5. Sorry for chiming in late, but wouldn't the patch actually hide the underlying root problem?
6. **body:** [~xuefuz] I haven't been able to reproduce this problem, except by explicitly deleting the root directory of operation logs. However, it is not clear if directory deletion was the root cause of the problem seen in production environments. In case of production environments, users tend to want resolution asap, and they restart HS2 for that, making it very hard to investigate the environment for possible reasons. The HiveServer2 logs didn't have more useful information than what is pasted in description. So the current state wasn't anyway helping in finding the root problem. This change should hopefully keep the production systems up and running without needing a restart. If you have any suggestions for additional debugging that can help in detecting the cause of the failure, please suggest them. We can use a new jira for that.

label: code-design