

git_comments:**git_commits:**

1. **summary:** [SPARK-13446][SQL] Support reading data from Hive 2.0.1 metastore
message: [SPARK-13446][SQL] Support reading data from Hive 2.0.1 metastore ### What changes were proposed in this pull request? This PR is to make Spark work with Hive 2.0's metastores. Compared with Hive 1.2, Hive 2.0's metastore has an API update due to removal of `HOLD_DDLTIME` in <https://issues.apache.org/jira/browse/HIVE-12224>. Based on the following Hive JIRA description, `HOLD_DDLTIME` should be removed from our internal API too. (<https://github.com/apache/spark/pull/17063> was submitted for it): > This arcane feature was introduced long ago via HIVE-1394 It was broken as soon as it landed, HIVE-1442 and is thus useless. Fact that no one has fixed it since informs that its not really used by anyone. Better is to remove it so no one hits the bug of HIVE-1442 In the next PR, we will support 2.1.0 metastore, whose APIs were changed due to <https://issues.apache.org/jira/browse/HIVE-12730>. However, before that, we need a code cleanup for stats collection and setting. ### How was this patch tested? Added test cases to VersionsSuite.scala Author: Xiao Li <gatorsmile@gmail.com> Closes #17061 from gatorsmile/Hive2.

github_issues:**github_issues_comments:****github_pulls:**

1. **title:** [SPARK-13446] [SQL] Support reading data from Hive 2.0.1 metastore
body: ### What changes were proposed in this pull request? This PR is to make Spark work with Hive 2.0's metastores. Compared with Hive 1.2, Hive 2.0's metastore has an API update due to removal of `HOLD_DDLTIME` in <https://issues.apache.org/jira/browse/HIVE-12224>. Based on the following Hive JIRA description, `HOLD_DDLTIME` should be removed from our internal API too. (<https://github.com/apache/spark/pull/17063> was submitted for it): > This arcane feature was introduced long ago via HIVE-1394 It was broken as soon as it landed, HIVE-1442 and is thus useless. Fact that no one has fixed it since informs that its not really used by anyone. Better is to remove it so no one hits the bug of HIVE-1442 In the next PR, we will support 2.1.0 metastore, whose APIs were changed due to <https://issues.apache.org/jira/browse/HIVE-12730>. However, before that, we need a code cleanup for stats collection and setting. ### How was this patch tested? Added test cases to VersionsSuite.scala

github_pulls_comments:

1. ****[Test build #73458 has finished]**
(<https://amplab.cs.berkeley.edu/jenkins/job/SparkPullRequestBuilder/73458/testReport>)** for PR 17061 at commit [`60af17f`](<https://github.com/apache/spark/commit/60af17f0178ba8ab7d7881c118915334e2c824eb>). * This patch passes all tests. * This patch merges cleanly. * This patch adds no public classes.
2. ****[Test build #73469 has finished]**
(<https://amplab.cs.berkeley.edu/jenkins/job/SparkPullRequestBuilder/73469/testReport>)** for PR 17061 at commit [`febb392`](<https://github.com/apache/spark/commit/febb392116d4540d22309b09e8a70bd62616b0b0>). * This patch passes all tests. * This patch merges cleanly. * This patch adds no public classes.
3. ****[Test build #73470 has finished]**
(<https://amplab.cs.berkeley.edu/jenkins/job/SparkPullRequestBuilder/73470/testReport>)** for PR 17061 at commit [`9ea5850`](<https://github.com/apache/spark/commit/9ea58508224d1de15f2976282f5e8127ea7f8eac>). * This patch passes all tests. * This patch merges cleanly. * This patch adds no public classes.
4. cc @cloud-fan @yhuai @sameeragarwal
5. LGTM.
6. Thank you! @vanzin
7. retest this please
8. we should also update `InsertIntoHiveTable`, it will be great if we can improve it to fail compilation when adding new hive versions.
9. ****[Test build #73860 has finished]**
(<https://amplab.cs.berkeley.edu/jenkins/job/SparkPullRequestBuilder/73860/testReport>)** for PR 17061 at commit [`9ea5850`](<https://github.com/apache/spark/commit/9ea58508224d1de15f2976282f5e8127ea7f8eac>). * This patch passes all tests. * This patch merges cleanly. * This patch adds no public classes.
10. LGTM, pending tests
11. ****[Test build #73872 has finished]**
(<https://amplab.cs.berkeley.edu/jenkins/job/SparkPullRequestBuilder/73872/testReport>)** for PR 17061 at commit [`b713a81`](<https://github.com/apache/spark/commit/b713a81c54efc69e0f9b6e5a2405210e1cb639a7>). * This patch passes all tests. * This patch merges cleanly. * This patch adds no public classes.
12. thanks, merging to master!

github_pulls_reviews:

1. Previously, we did not add the configuration generated by `HiveUtils.hiveClientConfigurations`. Here, we added it to see whether it could cause any test failure.
2. is the `exclusions` still useful?
3. This was originally added because `VersionsSuite` would fail without it, since it uses maven/ivy to download the different Hive libraries (because they used to depend on snapshot releases of libraries, or libraries that don't exist anymore on maven central). Perhaps Hive has updated its dependencies enough that some of this can be cleaned up...
4. why this?
5. Hive changed the default from `true` to `false` since 2.0 `` datanucleus.autoCreateSchema Default Value: true Added In: Hive 0.7.0 Removed In: Hive 2.0.0 with HIVE-6113, replaced by datanucleus.schema.autoCreateAll Creates necessary schema on a startup if one does not exist. Set this to false, after creating it once. In Hive 0.12.0 and later releases, datanucleus.autoCreateSchema is disabled if hive.metastore.schema.verification is true. `` `` datanucleus.schema.autoCreateAll Default Value: false Added In: Hive 2.0.0 with HIVE-6113, replaces datanucleus.autoCreateSchema (with different default value) Creates necessary schema on a startup if one does not exist. Reset this to false, after creating it once. datanucleus.schema.autoCreateAll is disabled if hive.metastore.schema.verification is true. `` Without changing the flag, we will get the following error `` 14:59:39.253 WARN DataNucleus.Query: Query for candidates of org.apache.hadoop.hive.metastore.model.MDatabase and subclasses resulted in no possible candidates Required table missing : "DBS" in Catalog "" Schema "". DataNucleus requires this table to perform its persistence operations. Either your MetaData is incorrect, or you need to enable "datanucleus.schema.autoCreateTables" org.datanucleus.store.rdbms.exceptions.MissingTableException: Required table missing : "DBS" in Catalog "" Schema "". DataNucleus requires this table to perform its persistence operations. Either your MetaData is incorrect, or you need to enable "datanucleus.schema.autoCreateTables" at org.datanucleus.store.rdbms.table.AbstractTable.exists(AbstractTable.java:606) at org.datanucleus.store.rdbms.RDBMSStoreManager\$ClassAdder.performTablesValidation(RDBMSStoreManager.java:3365) at org.datanucleus.store.rdbms.RDBMSStoreManager\$ClassAdder.run(RDBMSStoreManager.java:2877) ``
6. See the JIRA: <https://issues.apache.org/jira/browse/HIVE-6113>
7. let's add a comment so that future readers can know this
8. Sure. Will do.

jira_issues:

1. **summary:** Remove HOLD_DDLTIME
description: This arcane feature was introduced long ago via HIVE-1394 It was broken as soon as it landed, HIVE-1442 and is thus useless. Fact that no one has fixed it since informs that its not really used by anyone. Better is to remove it so no one hits the bug of HIVE-1442

jira_issues_comments:

1. Here are the results of testing the latest attachment: <https://issues.apache.org/jira/secure/attachment/12767879/HIVE-12224.patch> {color:green}SUCCESS:{color} +1 due to 2 test(s) being added or modified. {color:red}ERROR:{color} -1 due to 2 failed/errored test(s), 9696 tests executed *Failed tests:* {noformat} 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/5732/testReport> Console output: <http://ec2-174-129-184-35.compute-1.amazonaws.com/jenkins/job/PreCommit-HIVE-TRUNK-Build/5732/console> Test logs: <http://ec2-174-129-184-35.compute-1.amazonaws.com/logs/PreCommit-HIVE-TRUNK-Build-5732/> 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: 2 tests failed {noformat} This message is automatically generated. ATTACHMENT ID: 12767879 - PreCommit-HIVE-TRUNK-Build
2. <https://reviews.apache.org/r/39555/>
3. LGTM +1. The only issue is to remove redundant `./itests/src/test/resources/testconfiguration.properties:472: ddltime.q`
Thanks!
4. Pushed to master.