Item 268
git_comments:

git_commits:

1. **summary:** [MINOR] fix get classname for hive sync (#2008)
   message: [MINOR] fix get classname for hive sync (#2008)

github_issues:

1. **title:** [SUPPORT] hudi hive-sync in master branch (0.6.1) can not run by spark
   body: **Describe the problem you faced** Hudi in master branch (0.6.1) can not use `hive-sync` to sync to hive with error ``` Caused by: java.lang.ClassNotFoundException: parquet.hadoop.ParquetInputFormat ``` **Steps to reproduce the behavior:** 1. run a `HoodieDeltaStreamer` task by master `local[2]` and sync hudi table to hive 2. when sync to hive, it report error: ``` java.lang.NoClassDefFoundError: parquet/hadoop/ParquetInputFormat at org.apache.hadoop.hive.ql.io.parquet.MapredParquetInputFormat.<init>(MapredParquetInputFormat.java:46) ~ [hive-exec-1.2.1.spark2.jar:1.2.1.spark2] at org.apache.hudi.hadoop.HoodieParquetInputFormat.<init> (HoodieParquetInputFormat.java:67) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.hadoop.utils.HoodieInputFormatUtils.getInputFormat(HoodieInputFormatUtils.java:82) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.hadoop.utils.HoodieInputFormatUtils.getInputFormatClassName(HoodieInputFormatUtils.java:92) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.hive.HiveSyncTool.syncSchema(HiveSyncTool.java:159) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.hive.HiveSyncTool.syncHoodieTable(HiveSyncTool.java:130) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.hive.HiveSyncTool.syncHoodieTable(HiveSyncTool.java:98) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.utilities.deltastreamer.DeltaSync.syncMeta(DeltaSync.java:510) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.utilities.deltastreamer.DeltaSync.writeToSink(DeltaSync.java:425) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.utilities.deltastreamer.DeltaSync.syncOnce(DeltaSync.java:244) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.utilities.deltastreamer.HoodieDeltaStreamer.lambda$sync$2(HoodieDeltaStreamer.java:161) ~ [hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.common.util.Option.ifPresent(Option.java:96) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.utilities.deltastreamer.HoodieDeltaStreamer.sync(HoodieDeltaStreamer.java:159) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] *** *** Caused by: java.lang.ClassNotFoundException: parquet.hadoop.ParquetInputFormat at java.net.URLClassLoader.findClass(URLClassLoader.java:382) ~ [na:1.8.0_251] at java.lang.ClassLoader.loadClass(ClassLoader.java:418) ~[na:1.8.0_251] at sun.misc.Launcher$AppClassLoader.loadClass(Launcher.java:355) ~[na:1.8.0_251] at java.lang.ClassLoader.loadClass(ClassLoader.java:351) ~[na:1.8.0_251] ... 20 common frames omitted ``` **Environment Description** * Hudi version : 0.6.1 * Spark version : 2.4.3 * Hive version : 2.3.3 * Hadoop version : 2.8.5 * Storage (HDFS/S3/GCS..) : HDFS * Running on Docker? (yes/no) : no **Additional context** I checked the error code: ```java public static FileInputFormat getInputFormat(HoodieFileFormat baseFileFormat, boolean realtime, Configuration conf) { switch (baseFileFormat) { case PARQUET: if (realtime) { HoodieParquetRealtimeInputFormat inputFormat = new HoodieParquetRealtimeInputFormat(); inputFormat.setConf(conf); return inputFormat; } else { HoodieParquetInputFormat inputFormat = new HoodieParquetInputFormat(); inputFormat.setConf(conf); return inputFormat; } default: throw new HoodieIOException("Hoodie InputFormat not implemented for base file format " + baseFileFormat); } } public static String getInputFormatClassName(HoodieFileFormat baseFileFormat, boolean realtime, Configuration conf) { FileInputFormat inputFormat = getInputFormat(baseFileFormat, realtime, conf); return inputFormat.getClass().getName(); } ``` I think new a `ParquetInputFormat` may not a good idea for hudi in spark. In `hive-sync` package hudi just need a `FileInputFormat` class name, there is no need to new an object and get this class name. Meanwhile, spark also doesn't have total hive jars to do some action like hive.
2. **title:** [SUPPORT] hudi hive-sync in master branch (0.6.1) can not run by spark
   body: **Describe the problem you faced** Hudi in master branch (0.6.1) can not use `hive-sync` to sync to hive with error ``` Caused by: java.lang.ClassNotFoundException: parquet.hadoop.ParquetInputFormat ``` **Steps to reproduce the behavior:** 1. run a `HoodieDeltaStreamer` task by master `local[2]` and sync hudi table to hive 2. when sync to hive, it report error: ``` java.lang.NoClassDefFoundError: parquet/hadoop/ParquetInputFormat at org.apache.hadoop.hive.ql.io.parquet.MapredParquetInputFormat.<init>(MapredParquetInputFormat.java:46) ~

[hive-exec-1.2.1.spark2.jar:1.2.1.spark2] at org.apache.hudi.hadoop.HoodieParquetInputFormat.<init> (HoodieParquetInputFormat.java:67) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.hadoop.utils.HoodieInputFormatUtils.getInputFormat(HoodieInputFormatUtils.java:82) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.hadoop.utils.HoodieInputFormatUtils.getInputFormatClassName(HoodieInputFormatUtils.java:92) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.hive.HiveSyncTool.syncSchema(HiveSyncTool.java:159) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.hive.HiveSyncTool.syncHoodieTable(HiveSyncTool.java:130) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.hive.HiveSyncTool.syncHoodieTable(HiveSyncTool.java:98) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.utilities.deltastreamer.DeltaSync.syncMeta(DeltaSync.java:510) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.utilities.deltastreamer.DeltaSync.writeToSink(DeltaSync.java:425) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.utilities.deltastreamer.DeltaSync.syncOnce(DeltaSync.java:244) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.utilities.deltastreamer.HoodieDeltaStreamer.lambda$sync$2(HoodieDeltaStreamer.java:161) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.common.util.Option.ifPresent(Option.java:96) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.utilities.deltastreamer.HoodieDeltaStreamer.sync(HoodieDeltaStreamer.java:159) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] *** *** Caused by: java.lang.ClassNotFoundException: parquet.hadoop.ParquetInputFormat at java.net.URLClassLoader.findClass(URLClassLoader.java:382) ~[na:1.8.0_251] at java.lang.ClassLoader.loadClass(ClassLoader.java:418) ~[na:1.8.0_251] at sun.misc.Launcher$AppClassLoader.loadClass(Launcher.java:355) ~[na:1.8.0_251] at java.lang.ClassLoader.loadClass(ClassLoader.java:351) ~[na:1.8.0_251] ... 20 common frames omitted ``` **Environment Description** * Hudi version : 0.6.1 * Spark version : 2.4.3 * Hive version : 2.3.3 * Hadoop version : 2.8.5 * Storage (HDFS/S3/GCS..) : HDFS * Running on Docker? (yes/no) : no **Additional context** I checked the error code: ```java public static FileInputFormat getInputFormat(HoodieFileFormat baseFileFormat, boolean realtime, Configuration conf) { switch (baseFileFormat) { case PARQUET: if (realtime) { HoodieParquetRealtimeInputFormat inputFormat = new HoodieParquetRealtimeInputFormat(); inputFormat.setConf(conf); return inputFormat; } else { HoodieParquetInputFormat inputFormat = new HoodieParquetInputFormat(); inputFormat.setConf(conf); return inputFormat; } default: throw new HoodieIOException("Hoodie InputFormat not implemented for base file format " + baseFileFormat); } } public static String getInputFormatClassName(HoodieFileFormat baseFileFormat, boolean realtime, Configuration conf) { FileInputFormat inputFormat = getInputFormat(baseFileFormat, realtime, conf); return inputFormat.getClass().getName(); } ``` I think new a `ParquetInputFormat` may not a good idea for hudi in spark. In `hive-sync` package hudi just need a `FileInputFormat` class name, there is no need to new an object and get this class name. Meanwhile, spark also doesn't have total hive jars to do some action like hive.
label: test

3. **title:** [SUPPORT] hudi hive-sync in master branch (0.6.1) can not run by spark
   body: **Describe the problem you faced** Hudi in master branch (0.6.1) can not use `hive-sync` to sync to hive with error ``` Caused by: java.lang.ClassNotFoundException: parquet.hadoop.ParquetInputFormat ``` **Steps to reproduce the behavior:** 1. run a `HoodieDeltaStreamer` task by master `local[2]` and sync hudi table to hive 2. when sync to hive, it report error: ``` java.lang.NoClassDefFoundError: parquet/hadoop/ParquetInputFormat at org.apache.hadoop.hive.ql.io.parquet.MapredParquetInputFormat.<init>(MapredParquetInputFormat.java:46) ~[hive-exec-1.2.1.spark2.jar:1.2.1.spark2] at org.apache.hudi.hadoop.HoodieParquetInputFormat.<init> (HoodieParquetInputFormat.java:67) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.hadoop.utils.HoodieInputFormatUtils.getInputFormat(HoodieInputFormatUtils.java:82) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.hadoop.utils.HoodieInputFormatUtils.getInputFormatClassName(HoodieInputFormatUtils.java:92) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.hive.HiveSyncTool.syncSchema(HiveSyncTool.java:159) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.hive.HiveSyncTool.syncHoodieTable(HiveSyncTool.java:130) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.hive.HiveSyncTool.syncHoodieTable(HiveSyncTool.java:98) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.utilities.deltastreamer.DeltaSync.syncMeta(DeltaSync.java:510) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.utilities.deltastreamer.DeltaSync.writeToSink(DeltaSync.java:425) ~[hudi-utilities-bundle_2.11-

0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at
org.apache.hudi.utilities.deltastreamer.DeltaSync.syncOnce(DeltaSync.java:244) ~[hudi-utilities-bundle_2.11-
0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at
org.apache.hudi.utilities.deltastreamer.HoodieDeltaStreamer.lambda$sync$2(HoodieDeltaStreamer.java:161) ~
[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at
org.apache.hudi.common.util.Option.ifPresent(Option.java:96) ~[hudi-utilities-bundle_2.11-0.6.1-
SNAPSHOT.jar:0.6.1-SNAPSHOT] at
org.apache.hudi.utilities.deltastreamer.HoodieDeltaStreamer.sync(HoodieDeltaStreamer.java:159) ~[hudi-utilities-
bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] *** *** Caused by: java.lang.ClassNotFoundException:
parquet.hadoop.ParquetInputFormat at java.net.URLClassLoader.findClass(URLClassLoader.java:382) ~
[na:1.8.0_251] at java.lang.ClassLoader.loadClass(ClassLoader.java:418) ~[na:1.8.0_251] at
sun.misc.Launcher$AppClassLoader.loadClass(Launcher.java:355) ~[na:1.8.0_251] at
java.lang.ClassLoader.loadClass(ClassLoader.java:351) ~[na:1.8.0_251] ... 20 common frames omitted ```
Environment Description * Hudi version : 0.6.1 * Spark version : 2.4.3 * Hive version : 2.3.3 * Hadoop
version : 2.8.5 * Storage (HDFS/S3/GCS..) : HDFS * Running on Docker? (yes/no) : no **Additional context** I
checked the error code: ```java public static FileInputFormat getInputFormat(HoodieFileFormat baseFileFormat,
boolean realtime, Configuration conf) { switch (baseFileFormat) { case PARQUET: if (realtime) {
HoodieParquetRealtimeInputFormat inputFormat = new HoodieParquetRealtimeInputFormat();
inputFormat.setConf(conf); return inputFormat; } else { HoodieParquetInputFormat inputFormat = new
HoodieParquetInputFormat(); inputFormat.setConf(conf); return inputFormat; } default: throw new
HoodieIOException("Hoodie InputFormat not implemented for base file format " + baseFileFormat); } } public
static String getInputFormatClassName(HoodieFileFormat baseFileFormat, boolean realtime, Configuration conf)
{ FileInputFormat inputFormat = getInputFormat(baseFileFormat, realtime, conf); return
inputFormat.getClass().getName(); } ``` I think new a `ParquetInputFormat` may not a good idea for hudi in
spark. In `hive-sync` package hudi just need a `FileInputFormat` class name, there is no need to new an object and
get this class name. Meanwhile, spark also doesn't have total hive jars to do some action like hive.

4. **title:** [SUPPORT] hudi hive-sync in master branch (0.6.1) can not run by spark
   body: **Describe the problem you faced** Hudi in master branch (0.6.1) can not use `hive-sync` to sync to hive
   with error ``` Caused by: java.lang.ClassNotFoundException: parquet.hadoop.ParquetInputFormat ``` **Steps to
   reproduce the behavior:** 1. run a `HoodieDeltaStreamer` task by master `local[2]` and sync hudi table to hive 2.
   when sync to hive, it report error: ``` java.lang.NoClassDefFoundError: parquet/hadoop/ParquetInputFormat at
   org.apache.hadoop.hive.ql.io.parquet.MapredParquetInputFormat.<init>(MapredParquetInputFormat.java:46) ~
   [hive-exec-1.2.1.spark2.jar:1.2.1.spark2] at org.apache.hudi.hadoop.HoodieParquetInputFormat.<init>
   (HoodieParquetInputFormat.java:67) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at
   org.apache.hudi.hadoop.utils.HoodieInputFormatUtils.getInputFormat(HoodieInputFormatUtils.java:82) ~[hudi-
   utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at
   org.apache.hudi.hadoop.utils.HoodieInputFormatUtils.getInputFormatClassName(HoodieInputFormatUtils.java:92)
   ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at
   org.apache.hudi.hive.HiveSyncTool.syncSchema(HiveSyncTool.java:159) ~[hudi-utilities-bundle_2.11-0.6.1-
   SNAPSHOT.jar:0.6.1-SNAPSHOT] at
   org.apache.hudi.hive.HiveSyncTool.syncHoodieTable(HiveSyncTool.java:130) ~[hudi-utilities-bundle_2.11-0.6.1-
   SNAPSHOT.jar:0.6.1-SNAPSHOT] at
   org.apache.hudi.hive.HiveSyncTool.syncHoodieTable(HiveSyncTool.java:98) ~[hudi-utilities-bundle_2.11-0.6.1-
   SNAPSHOT.jar:0.6.1-SNAPSHOT] at
   org.apache.hudi.utilities.deltastreamer.DeltaSync.syncMeta(DeltaSync.java:510) ~[hudi-utilities-bundle_2.11-
   0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at
   org.apache.hudi.utilities.deltastreamer.DeltaSync.writeToSink(DeltaSync.java:425) ~[hudi-utilities-bundle_2.11-
   0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at
   org.apache.hudi.utilities.deltastreamer.DeltaSync.syncOnce(DeltaSync.java:244) ~[hudi-utilities-bundle_2.11-
   0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at
   org.apache.hudi.utilities.deltastreamer.HoodieDeltaStreamer.lambda$sync$2(HoodieDeltaStreamer.java:161) ~
   [hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at
   org.apache.hudi.common.util.Option.ifPresent(Option.java:96) ~[hudi-utilities-bundle_2.11-0.6.1-
   SNAPSHOT.jar:0.6.1-SNAPSHOT] at
   org.apache.hudi.utilities.deltastreamer.HoodieDeltaStreamer.sync(HoodieDeltaStreamer.java:159) ~[hudi-utilities-
   bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] *** *** Caused by: java.lang.ClassNotFoundException:
   parquet.hadoop.ParquetInputFormat at java.net.URLClassLoader.findClass(URLClassLoader.java:382) ~
   [na:1.8.0_251] at java.lang.ClassLoader.loadClass(ClassLoader.java:418) ~[na:1.8.0_251] at
   sun.misc.Launcher$AppClassLoader.loadClass(Launcher.java:355) ~[na:1.8.0_251] at
   java.lang.ClassLoader.loadClass(ClassLoader.java:351) ~[na:1.8.0_251] ... 20 common frames omitted ```
   Environment Description * Hudi version : 0.6.1 * Spark version : 2.4.3 * Hive version : 2.3.3 * Hadoop
   version : 2.8.5 * Storage (HDFS/S3/GCS..) : HDFS * Running on Docker? (yes/no) : no **Additional context** I
   checked the error code: ```java public static FileInputFormat getInputFormat(HoodieFileFormat baseFileFormat,

boolean realtime, Configuration conf) { switch (baseFileFormat) { case PARQUET: if (realtime) { HoodieParquetRealtimeInputFormat inputFormat = new HoodieParquetRealtimeInputFormat(); inputFormat.setConf(conf); return inputFormat; } else { HoodieParquetInputFormat inputFormat = new HoodieParquetInputFormat(); inputFormat.setConf(conf); return inputFormat; } default: throw new HoodieIOException("Hoodie InputFormat not implemented for base file format " + baseFileFormat); } } public static String getInputFormatClassName(HoodieFileFormat baseFileFormat, boolean realtime, Configuration conf) { FileInputFormat inputFormat = getInputFormat(baseFileFormat, realtime, conf); return inputFormat.getClass().getName(); } ``` I think new a `ParquetInputFormat` may not a good idea for hudi in spark. In `hive-sync` package hudi just need a `FileInputFormat` class name, there is no need to new an object and get this class name. Meanwhile, spark also doesn't have total hive jars to do some action like hive.

5. **title:** [SUPPORT] hudi hive-sync in master branch (0.6.1) can not run by spark
   body: **Describe the problem you faced** Hudi in master branch (0.6.1) can not use `hive-sync` to sync to hive with error ``` Caused by: java.lang.ClassNotFoundException: parquet.hadoop.ParquetInputFormat ``` **Steps to reproduce the behavior:** 1. run a `HoodieDeltaStreamer` task by master `local[2]` and sync hudi table to hive 2. when sync to hive, it report error: ``` java.lang.NoClassDefFoundError: parquet/hadoop/ParquetInputFormat at org.apache.hadoop.hive.ql.io.parquet.MapredParquetInputFormat.<init>(MapredParquetInputFormat.java:46) ~ [hive-exec-1.2.1.spark2.jar:1.2.1.spark2] at org.apache.hudi.hadoop.HoodieParquetInputFormat.<init> (HoodieParquetInputFormat.java:67) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.hadoop.utils.HoodieInputFormatUtils.getInputFormat(HoodieInputFormatUtils.java:82) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.hadoop.utils.HoodieInputFormatUtils.getInputFormatClassName(HoodieInputFormatUtils.java:92) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.hive.HiveSyncTool.syncSchema(HiveSyncTool.java:159) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.hive.HiveSyncTool.syncHoodieTable(HiveSyncTool.java:130) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.hive.HiveSyncTool.syncHoodieTable(HiveSyncTool.java:98) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.utilities.deltastreamer.DeltaSync.syncMeta(DeltaSync.java:510) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.utilities.deltastreamer.DeltaSync.writeToSink(DeltaSync.java:425) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.utilities.deltastreamer.DeltaSync.syncOnce(DeltaSync.java:244) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.utilities.deltastreamer.HoodieDeltaStreamer.lambda$sync$2(HoodieDeltaStreamer.java:161) ~ [hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.common.util.Option.ifPresent(Option.java:96) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.utilities.deltastreamer.HoodieDeltaStreamer.sync(HoodieDeltaStreamer.java:159) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] *** *** Caused by: java.lang.ClassNotFoundException: parquet.hadoop.ParquetInputFormat at java.net.URLClassLoader.findClass(URLClassLoader.java:382) ~ [na:1.8.0_251] at java.lang.ClassLoader.loadClass(ClassLoader.java:418) ~[na:1.8.0_251] at sun.misc.Launcher$AppClassLoader.loadClass(Launcher.java:355) ~[na:1.8.0_251] at java.lang.ClassLoader.loadClass(ClassLoader.java:351) ~[na:1.8.0_251] ... 20 common frames omitted ``` **Environment Description** * Hudi version : 0.6.1 * Spark version : 2.4.3 * Hive version : 2.3.3 * Hadoop version : 2.8.5 * Storage (HDFS/S3/GCS..) : HDFS * Running on Docker? (yes/no) : no **Additional context** I checked the error code: ```java public static FileInputFormat getInputFormat(HoodieFileFormat baseFileFormat, boolean realtime, Configuration conf) { switch (baseFileFormat) { case PARQUET: if (realtime) { HoodieParquetRealtimeInputFormat inputFormat = new HoodieParquetRealtimeInputFormat(); inputFormat.setConf(conf); return inputFormat; } else { HoodieParquetInputFormat inputFormat = new HoodieParquetInputFormat(); inputFormat.setConf(conf); return inputFormat; } default: throw new HoodieIOException("Hoodie InputFormat not implemented for base file format " + baseFileFormat); } } public static String getInputFormatClassName(HoodieFileFormat baseFileFormat, boolean realtime, Configuration conf) { FileInputFormat inputFormat = getInputFormat(baseFileFormat, realtime, conf); return inputFormat.getClass().getName(); } ``` I think new a `ParquetInputFormat` may not a good idea for hudi in spark. In `hive-sync` package hudi just need a `FileInputFormat` class name, there is no need to new an object and get this class name. Meanwhile, spark also doesn't have total hive jars to do some action like hive.

6. **title:** [SUPPORT] hudi hive-sync in master branch (0.6.1) can not run by spark
   body: **Describe the problem you faced** Hudi in master branch (0.6.1) can not use `hive-sync` to sync to hive with error ``` Caused by: java.lang.ClassNotFoundException: parquet.hadoop.ParquetInputFormat ``` **Steps to reproduce the behavior:** 1. run a `HoodieDeltaStreamer` task by master `local[2]` and sync hudi table to hive 2. when sync to hive, it report error: ``` java.lang.NoClassDefFoundError: parquet/hadoop/ParquetInputFormat at org.apache.hadoop.hive.ql.io.parquet.MapredParquetInputFormat.<init>(MapredParquetInputFormat.java:46) ~

[hive-exec-1.2.1.spark2.jar:1.2.1.spark2] at org.apache.hudi.hadoop.HoodieParquetInputFormat.<init>(HoodieParquetInputFormat.java:67) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.hadoop.utils.HoodieInputFormatUtils.getInputFormat(HoodieInputFormatUtils.java:82) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.hadoop.utils.HoodieInputFormatUtils.getInputFormatClassName(HoodieInputFormatUtils.java:92) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.hive.HiveSyncTool.syncSchema(HiveSyncTool.java:159) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.hive.HiveSyncTool.syncHoodieTable(HiveSyncTool.java:130) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.hive.HiveSyncTool.syncHoodieTable(HiveSyncTool.java:98) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.utilities.deltastreamer.DeltaSync.syncMeta(DeltaSync.java:510) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.utilities.deltastreamer.DeltaSync.writeToSink(DeltaSync.java:425) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.utilities.deltastreamer.DeltaSync.syncOnce(DeltaSync.java:244) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.utilities.deltastreamer.HoodieDeltaStreamer.lambda$sync$2(HoodieDeltaStreamer.java:161) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.common.util.Option.ifPresent(Option.java:96) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.utilities.deltastreamer.HoodieDeltaStreamer.sync(HoodieDeltaStreamer.java:159) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] *** *** Caused by: java.lang.ClassNotFoundException: parquet.hadoop.ParquetInputFormat at java.net.URLClassLoader.findClass(URLClassLoader.java:382) ~[na:1.8.0_251] at java.lang.ClassLoader.loadClass(ClassLoader.java:418) ~[na:1.8.0_251] at sun.misc.Launcher$AppClassLoader.loadClass(Launcher.java:355) ~[na:1.8.0_251] at java.lang.ClassLoader.loadClass(ClassLoader.java:351) ~[na:1.8.0_251] ... 20 common frames omitted ``` **Environment Description** * Hudi version : 0.6.1 * Spark version : 2.4.3 * Hive version : 2.3.3 * Hadoop version : 2.8.5 * Storage (HDFS/S3/GCS..) : HDFS * Running on Docker? (yes/no) : no **Additional context** I checked the error code: ```java public static FileInputFormat getInputFormat(HoodieFileFormat baseFileFormat, boolean realtime, Configuration conf) { switch (baseFileFormat) { case PARQUET: if (realtime) { HoodieParquetRealtimeInputFormat inputFormat = new HoodieParquetRealtimeInputFormat(); inputFormat.setConf(conf); return inputFormat; } else { HoodieParquetInputFormat inputFormat = new HoodieParquetInputFormat(); inputFormat.setConf(conf); return inputFormat; } default: throw new HoodieIOException("Hoodie InputFormat not implemented for base file format " + baseFileFormat); } } public static String getInputFormatClassName(HoodieFileFormat baseFileFormat, boolean realtime, Configuration conf) { FileInputFormat inputFormat = getInputFormat(baseFileFormat, realtime, conf); return inputFormat.getClass().getName(); } ``` I think new a `ParquetInputFormat` may not a good idea for hudi in spark. In `hive-sync` package hudi just need a `FileInputFormat` class name, there is no need to new an object and get this class name. Meanwhile, spark also doesn't have total hive jars to do some action like hive.

7. **title:** [SUPPORT] hudi hive-sync in master branch (0.6.1) can not run by spark
   body: **Describe the problem you faced** Hudi in master branch (0.6.1) can not use `hive-sync` to sync to hive with error ``` Caused by: java.lang.ClassNotFoundException: parquet.hadoop.ParquetInputFormat ``` **Steps to reproduce the behavior:** 1. run a `HoodieDeltaStreamer` task by master `local[2]` and sync hudi table to hive 2. when sync to hive, it report error: ``` java.lang.NoClassDefFoundError: parquet/hadoop/ParquetInputFormat at org.apache.hadoop.hive.ql.io.parquet.MapredParquetInputFormat.<init>(MapredParquetInputFormat.java:46) ~[hive-exec-1.2.1.spark2.jar:1.2.1.spark2] at org.apache.hudi.hadoop.HoodieParquetInputFormat.<init>(HoodieParquetInputFormat.java:67) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.hadoop.utils.HoodieInputFormatUtils.getInputFormat(HoodieInputFormatUtils.java:82) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.hadoop.utils.HoodieInputFormatUtils.getInputFormatClassName(HoodieInputFormatUtils.java:92) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.hive.HiveSyncTool.syncSchema(HiveSyncTool.java:159) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.hive.HiveSyncTool.syncHoodieTable(HiveSyncTool.java:130) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.hive.HiveSyncTool.syncHoodieTable(HiveSyncTool.java:98) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.utilities.deltastreamer.DeltaSync.syncMeta(DeltaSync.java:510) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.utilities.deltastreamer.DeltaSync.writeToSink(DeltaSync.java:425) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at

org.apache.hudi.utilities.deltastreamer.DeltaSync.syncOnce(DeltaSync.java:244) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.utilities.deltastreamer.HoodieDeltaStreamer.lambda$sync$2(HoodieDeltaStreamer.java:161) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.common.util.Option.ifPresent(Option.java:96) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.utilities.deltastreamer.HoodieDeltaStreamer.sync(HoodieDeltaStreamer.java:159) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] *** *** Caused by: java.lang.ClassNotFoundException: parquet.hadoop.ParquetInputFormat at java.net.URLClassLoader.findClass(URLClassLoader.java:382) ~[na:1.8.0_251] at java.lang.ClassLoader.loadClass(ClassLoader.java:418) ~[na:1.8.0_251] at sun.misc.Launcher$AppClassLoader.loadClass(Launcher.java:355) ~[na:1.8.0_251] at java.lang.ClassLoader.loadClass(ClassLoader.java:351) ~[na:1.8.0_251] ... 20 common frames omitted ``` **Environment Description** * Hudi version : 0.6.1 * Spark version : 2.4.3 * Hive version : 2.3.3 * Hadoop version : 2.8.5 * Storage (HDFS/S3/GCS..) : HDFS * Running on Docker? (yes/no) : no **Additional context** I checked the error code: ```java public static FileInputFormat getInputFormat(HoodieFileFormat baseFileFormat, boolean realtime, Configuration conf) { switch (baseFileFormat) { case PARQUET: if (realtime) { HoodieParquetRealtimeInputFormat inputFormat = new HoodieParquetRealtimeInputFormat(); inputFormat.setConf(conf); return inputFormat; } else { HoodieParquetInputFormat inputFormat = new HoodieParquetInputFormat(); inputFormat.setConf(conf); return inputFormat; } default: throw new HoodieIOException("Hoodie InputFormat not implemented for base file format " + baseFileFormat); } } public static String getInputFormatClassName(HoodieFileFormat baseFileFormat, boolean realtime, Configuration conf) { FileInputFormat inputFormat = getInputFormat(baseFileFormat, realtime, conf); return inputFormat.getClass().getName(); } ``` I think new a `ParquetInputFormat` may not a good idea for hudi in spark. In `hive-sync` package hudi just need a `FileInputFormat` class name, there is no need to new an object and get this class name. Meanwhile, spark also doesn't have total hive jars to do some action like hive.

8. **title:** [SUPPORT] hudi hive-sync in master branch (0.6.1) can not run by spark
   body: **Describe the problem you faced** Hudi in master branch (0.6.1) can not use `hive-sync` to sync to hive with error ``` Caused by: java.lang.ClassNotFoundException: parquet.hadoop.ParquetInputFormat ``` **Steps to reproduce the behavior:** 1. run a `HoodieDeltaStreamer` task by master `local[2]` and sync hudi table to hive 2. when sync to hive, it report error: ``` java.lang.NoClassDefFoundError: parquet/hadoop/ParquetInputFormat at org.apache.hadoop.hive.ql.io.parquet.MapredParquetInputFormat.<init>(MapredParquetInputFormat.java:46) ~[hive-exec-1.2.1.spark2.jar:1.2.1.spark2] at org.apache.hudi.hadoop.HoodieParquetInputFormat.<init>(HoodieParquetInputFormat.java:67) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.hadoop.utils.HoodieInputFormatUtils.getInputFormat(HoodieInputFormatUtils.java:82) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.hadoop.utils.HoodieInputFormatUtils.getInputFormatClassName(HoodieInputFormatUtils.java:92) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.hive.HiveSyncTool.syncSchema(HiveSyncTool.java:159) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.hive.HiveSyncTool.syncHoodieTable(HiveSyncTool.java:130) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.hive.HiveSyncTool.syncHoodieTable(HiveSyncTool.java:98) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.utilities.deltastreamer.DeltaSync.syncMeta(DeltaSync.java:510) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.utilities.deltastreamer.DeltaSync.writeToSink(DeltaSync.java:425) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.utilities.deltastreamer.DeltaSync.syncOnce(DeltaSync.java:244) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.utilities.deltastreamer.HoodieDeltaStreamer.lambda$sync$2(HoodieDeltaStreamer.java:161) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.common.util.Option.ifPresent(Option.java:96) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.utilities.deltastreamer.HoodieDeltaStreamer.sync(HoodieDeltaStreamer.java:159) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] *** *** Caused by: java.lang.ClassNotFoundException: parquet.hadoop.ParquetInputFormat at java.net.URLClassLoader.findClass(URLClassLoader.java:382) ~[na:1.8.0_251] at java.lang.ClassLoader.loadClass(ClassLoader.java:418) ~[na:1.8.0_251] at sun.misc.Launcher$AppClassLoader.loadClass(Launcher.java:355) ~[na:1.8.0_251] at java.lang.ClassLoader.loadClass(ClassLoader.java:351) ~[na:1.8.0_251] ... 20 common frames omitted ``` **Environment Description** * Hudi version : 0.6.1 * Spark version : 2.4.3 * Hive version : 2.3.3 * Hadoop version : 2.8.5 * Storage (HDFS/S3/GCS..) : HDFS * Running on Docker? (yes/no) : no **Additional context** I checked the error code: ```java public static FileInputFormat getInputFormat(HoodieFileFormat baseFileFormat, boolean realtime, Configuration conf) { switch (baseFileFormat) { case PARQUET: if (realtime) {

HoodieParquetRealtimeInputFormat inputFormat = new HoodieParquetRealtimeInputFormat();
inputFormat.setConf(conf); return inputFormat; } else { HoodieParquetInputFormat inputFormat = new
HoodieParquetInputFormat(); inputFormat.setConf(conf); return inputFormat; } default: throw new
HoodieIOException("Hoodie InputFormat not implemented for base file format " + baseFileFormat); } } public
static String getInputFormatClassName(HoodieFileFormat baseFileFormat, boolean realtime, Configuration conf)
{ FileInputFormat inputFormat = getInputFormat(baseFileFormat, realtime, conf); return
inputFormat.getClass().getName(); } ``` I think new a `ParquetInputFormat` may not a good idea for hudi in
spark. In `hive-sync` package hudi just need a `FileInputFormat` class name, there is no need to new an object and
get this class name. Meanwhile, spark also doesn't have total hive jars to do some action like hive.

9. **title:** [SUPPORT] hudi hive-sync in master branch (0.6.1) can not run by spark
   body: **Describe the problem you faced** Hudi in master branch (0.6.1) can not use `hive-sync` to sync to hive
   with error ``` Caused by: java.lang.ClassNotFoundException: parquet.hadoop.ParquetInputFormat ``` **Steps to
   reproduce the behavior:** 1. run a `HoodieDeltaStreamer` task by master `local[2]` and sync hudi table to hive 2.
   when sync to hive, it report error: ``` java.lang.NoClassDefFoundError: parquet/hadoop/ParquetInputFormat at
   org.apache.hadoop.hive.ql.io.parquet.MapredParquetInputFormat.<init>(MapredParquetInputFormat.java:46) ~
   [hive-exec-1.2.1.spark2.jar:1.2.1.spark2] at org.apache.hudi.hadoop.HoodieParquetInputFormat.<init>
   (HoodieParquetInputFormat.java:67) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at
   org.apache.hudi.hadoop.utils.HoodieInputFormatUtils.getInputFormat(HoodieInputFormatUtils.java:82) ~[hudi-
   utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at
   org.apache.hudi.hadoop.utils.HoodieInputFormatUtils.getInputFormatClassName(HoodieInputFormatUtils.java:92)
   ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at
   org.apache.hudi.hive.HiveSyncTool.syncSchema(HiveSyncTool.java:159) ~[hudi-utilities-bundle_2.11-0.6.1-
   SNAPSHOT.jar:0.6.1-SNAPSHOT] at
   org.apache.hudi.hive.HiveSyncTool.syncHoodieTable(HiveSyncTool.java:130) ~[hudi-utilities-bundle_2.11-0.6.1-
   SNAPSHOT.jar:0.6.1-SNAPSHOT] at
   org.apache.hudi.hive.HiveSyncTool.syncHoodieTable(HiveSyncTool.java:98) ~[hudi-utilities-bundle_2.11-0.6.1-
   SNAPSHOT.jar:0.6.1-SNAPSHOT] at
   org.apache.hudi.utilities.deltastreamer.DeltaSync.syncMeta(DeltaSync.java:510) ~[hudi-utilities-bundle_2.11-
   0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at
   org.apache.hudi.utilities.deltastreamer.DeltaSync.writeToSink(DeltaSync.java:425) ~[hudi-utilities-bundle_2.11-
   0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at
   org.apache.hudi.utilities.deltastreamer.DeltaSync.syncOnce(DeltaSync.java:244) ~[hudi-utilities-bundle_2.11-
   0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at
   org.apache.hudi.utilities.deltastreamer.HoodieDeltaStreamer.lambda$sync$2(HoodieDeltaStreamer.java:161) ~
   [hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at
   org.apache.hudi.common.util.Option.ifPresent(Option.java:96) ~[hudi-utilities-bundle_2.11-0.6.1-
   SNAPSHOT.jar:0.6.1-SNAPSHOT] at
   org.apache.hudi.utilities.deltastreamer.HoodieDeltaStreamer.sync(HoodieDeltaStreamer.java:159) ~[hudi-utilities-
   bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] *** *** Caused by: java.lang.ClassNotFoundException:
   parquet.hadoop.ParquetInputFormat at java.net.URLClassLoader.findClass(URLClassLoader.java:382) ~
   [na:1.8.0_251] at java.lang.ClassLoader.loadClass(ClassLoader.java:418) ~[na:1.8.0_251] at
   sun.misc.Launcher$AppClassLoader.loadClass(Launcher.java:355) ~[na:1.8.0_251] at
   java.lang.ClassLoader.loadClass(ClassLoader.java:351) ~[na:1.8.0_251] ... 20 common frames omitted ```
   Environment Description * Hudi version : 0.6.1 * Spark version : 2.4.3 * Hive version : 2.3.3 * Hadoop
   version : 2.8.5 * Storage (HDFS/S3/GCS..) : HDFS * Running on Docker? (yes/no) : no **Additional context** I
   checked the error code: ```java public static FileInputFormat getInputFormat(HoodieFileFormat baseFileFormat,
   boolean realtime, Configuration conf) { switch (baseFileFormat) { case PARQUET: if (realtime) {
   HoodieParquetRealtimeInputFormat inputFormat = new HoodieParquetRealtimeInputFormat();
   inputFormat.setConf(conf); return inputFormat; } else { HoodieParquetInputFormat inputFormat = new
   HoodieParquetInputFormat(); inputFormat.setConf(conf); return inputFormat; } default: throw new
   HoodieIOException("Hoodie InputFormat not implemented for base file format " + baseFileFormat); } } public
   static String getInputFormatClassName(HoodieFileFormat baseFileFormat, boolean realtime, Configuration conf)
   { FileInputFormat inputFormat = getInputFormat(baseFileFormat, realtime, conf); return
   inputFormat.getClass().getName(); } ``` I think new a `ParquetInputFormat` may not a good idea for hudi in
   spark. In `hive-sync` package hudi just need a `FileInputFormat` class name, there is no need to new an object and
   get this class name. Meanwhile, spark also doesn't have total hive jars to do some action like hive.

10. **title:** [SUPPORT] hudi hive-sync in master branch (0.6.1) can not run by spark
    body: **Describe the problem you faced** Hudi in master branch (0.6.1) can not use `hive-sync` to sync to hive
    with error ``` Caused by: java.lang.ClassNotFoundException: parquet.hadoop.ParquetInputFormat ``` **Steps to
    reproduce the behavior:** 1. run a `HoodieDeltaStreamer` task by master `local[2]` and sync hudi table to hive 2.
    when sync to hive, it report error: ``` java.lang.NoClassDefFoundError: parquet/hadoop/ParquetInputFormat at
    org.apache.hadoop.hive.ql.io.parquet.MapredParquetInputFormat.<init>(MapredParquetInputFormat.java:46) ~
    [hive-exec-1.2.1.spark2.jar:1.2.1.spark2] at org.apache.hudi.hadoop.HoodieParquetInputFormat.<init>

(HoodieParquetInputFormat.java:67) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.hadoop.utils.HoodieInputFormatUtils.getInputFormat(HoodieInputFormatUtils.java:82) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.hadoop.utils.HoodieInputFormatUtils.getInputFormatClassName(HoodieInputFormatUtils.java:92) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.hive.HiveSyncTool.syncSchema(HiveSyncTool.java:159) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.hive.HiveSyncTool.syncHoodieTable(HiveSyncTool.java:130) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.hive.HiveSyncTool.syncHoodieTable(HiveSyncTool.java:98) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.utilities.deltastreamer.DeltaSync.syncMeta(DeltaSync.java:510) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.utilities.deltastreamer.DeltaSync.writeToSink(DeltaSync.java:425) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.utilities.deltastreamer.DeltaSync.syncOnce(DeltaSync.java:244) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.utilities.deltastreamer.HoodieDeltaStreamer.lambda$sync$2(HoodieDeltaStreamer.java:161) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.common.util.Option.ifPresent(Option.java:96) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.utilities.deltastreamer.HoodieDeltaStreamer.sync(HoodieDeltaStreamer.java:159) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] *** *** Caused by: java.lang.ClassNotFoundException: parquet.hadoop.ParquetInputFormat at java.net.URLClassLoader.findClass(URLClassLoader.java:382) ~[na:1.8.0_251] at java.lang.ClassLoader.loadClass(ClassLoader.java:418) ~[na:1.8.0_251] at sun.misc.Launcher$AppClassLoader.loadClass(Launcher.java:355) ~[na:1.8.0_251] at java.lang.ClassLoader.loadClass(ClassLoader.java:351) ~[na:1.8.0_251] ... 20 common frames omitted ``` **Environment Description** * Hudi version : 0.6.1 * Spark version : 2.4.3 * Hive version : 2.3.3 * Hadoop version : 2.8.5 * Storage (HDFS/S3/GCS..) : HDFS * Running on Docker? (yes/no) : no **Additional context** I checked the error code: ```java public static FileInputFormat getInputFormat(HoodieFileFormat baseFileFormat, boolean realtime, Configuration conf) { switch (baseFileFormat) { case PARQUET: if (realtime) { HoodieParquetRealtimeInputFormat inputFormat = new HoodieParquetRealtimeInputFormat(); inputFormat.setConf(conf); return inputFormat; } else { HoodieParquetInputFormat inputFormat = new HoodieParquetInputFormat(); inputFormat.setConf(conf); return inputFormat; } default: throw new HoodieIOException("Hoodie InputFormat not implemented for base file format " + baseFileFormat); } } public static String getInputFormatClassName(HoodieFileFormat baseFileFormat, boolean realtime, Configuration conf) { FileInputFormat inputFormat = getInputFormat(baseFileFormat, realtime, conf); return inputFormat.getClass().getName(); } ``` I think new a `ParquetInputFormat` may not a good idea for hudi in spark. In `hive-sync` package hudi just need a `FileInputFormat` class name, there is no need to new an object and get this class name. Meanwhile, spark also doesn't have total hive jars to do some action like hive.

11. **title:** [SUPPORT] hudi hive-sync in master branch (0.6.1) can not run by spark
    body: **Describe the problem you faced** Hudi in master branch (0.6.1) can not use `hive-sync` to sync to hive with error ``` Caused by: java.lang.ClassNotFoundException: parquet.hadoop.ParquetInputFormat ``` **Steps to reproduce the behavior:** 1. run a `HoodieDeltaStreamer` task by master `local[2]` and sync hudi table to hive 2. when sync to hive, it report error: ``` java.lang.NoClassDefFoundError: parquet/hadoop/ParquetInputFormat at org.apache.hadoop.hive.ql.io.parquet.MapredParquetInputFormat.<init>(MapredParquetInputFormat.java:46) ~[hive-exec-1.2.1.spark2.jar:1.2.1.spark2] at org.apache.hudi.hadoop.HoodieParquetInputFormat.<init>(HoodieParquetInputFormat.java:67) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.hadoop.utils.HoodieInputFormatUtils.getInputFormat(HoodieInputFormatUtils.java:82) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.hadoop.utils.HoodieInputFormatUtils.getInputFormatClassName(HoodieInputFormatUtils.java:92) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.hive.HiveSyncTool.syncSchema(HiveSyncTool.java:159) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.hive.HiveSyncTool.syncHoodieTable(HiveSyncTool.java:130) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.hive.HiveSyncTool.syncHoodieTable(HiveSyncTool.java:98) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.utilities.deltastreamer.DeltaSync.syncMeta(DeltaSync.java:510) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.utilities.deltastreamer.DeltaSync.writeToSink(DeltaSync.java:425) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.utilities.deltastreamer.DeltaSync.syncOnce(DeltaSync.java:244) ~[hudi-utilities-bundle_2.11-

0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at
org.apache.hudi.utilities.deltastreamer.HoodieDeltaStreamer.lambda$sync$2(HoodieDeltaStreamer.java:161) ~
[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at
org.apache.hudi.common.util.Option.ifPresent(Option.java:96) ~[hudi-utilities-bundle_2.11-0.6.1-
SNAPSHOT.jar:0.6.1-SNAPSHOT] at
org.apache.hudi.utilities.deltastreamer.HoodieDeltaStreamer.sync(HoodieDeltaStreamer.java:159) ~[hudi-utilities-
bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] *** *** Caused by: java.lang.ClassNotFoundException:
parquet.hadoop.ParquetInputFormat at java.net.URLClassLoader.findClass(URLClassLoader.java:382) ~
[na:1.8.0_251] at java.lang.ClassLoader.loadClass(ClassLoader.java:418) ~[na:1.8.0_251] at
sun.misc.Launcher$AppClassLoader.loadClass(Launcher.java:355) ~[na:1.8.0_251] at
java.lang.ClassLoader.loadClass(ClassLoader.java:351) ~[na:1.8.0_251] ... 20 common frames omitted ```
Environment Description * Hudi version : 0.6.1 * Spark version : 2.4.3 * Hive version : 2.3.3 * Hadoop
version : 2.8.5 * Storage (HDFS/S3/GCS..) : HDFS * Running on Docker? (yes/no) : no **Additional context** I
checked the error code: ```java public static FileInputFormat getInputFormat(HoodieFileFormat baseFileFormat,
boolean realtime, Configuration conf) { switch (baseFileFormat) { case PARQUET: if (realtime) {
HoodieParquetRealtimeInputFormat inputFormat = new HoodieParquetRealtimeInputFormat();
inputFormat.setConf(conf); return inputFormat; } else { HoodieParquetInputFormat inputFormat = new
HoodieParquetInputFormat(); inputFormat.setConf(conf); return inputFormat; } default: throw new
HoodieIOException("Hoodie InputFormat not implemented for base file format " + baseFileFormat); } } public
static String getInputFormatClassName(HoodieFileFormat baseFileFormat, boolean realtime, Configuration conf)
{ FileInputFormat inputFormat = getInputFormat(baseFileFormat, realtime, conf); return
inputFormat.getClass().getName(); } ``` I think new a `ParquetInputFormat` may not a good idea for hudi in
spark. In `hive-sync` package hudi just need a `FileInputFormat` class name, there is no need to new an object and
get this class name. Meanwhile, spark also doesn't have total hive jars to do some action like hive.

12. **title:** [SUPPORT] hudi hive-sync in master branch (0.6.1) can not run by spark
    body: **Describe the problem you faced** Hudi in master branch (0.6.1) can not use `hive-sync` to sync to hive
    with error ``` Caused by: java.lang.ClassNotFoundException: parquet.hadoop.ParquetInputFormat ``` **Steps to
    reproduce the behavior:** 1. run a `HoodieDeltaStreamer` task by master `local[2]` and sync hudi table to hive 2.
    when sync to hive, it report error: ``` java.lang.NoClassDefFoundError: parquet/hadoop/ParquetInputFormat at
    org.apache.hadoop.hive.ql.io.parquet.MapredParquetInputFormat.<init>(MapredParquetInputFormat.java:46) ~
    [hive-exec-1.2.1.spark2.jar:1.2.1.spark2] at org.apache.hudi.hadoop.HoodieParquetInputFormat.<init>
    (HoodieParquetInputFormat.java:67) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at
    org.apache.hudi.hadoop.utils.HoodieInputFormatUtils.getInputFormat(HoodieInputFormatUtils.java:82) ~[hudi-
    utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at
    org.apache.hudi.hadoop.utils.HoodieInputFormatUtils.getInputFormatClassName(HoodieInputFormatUtils.java:92)
    ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at
    org.apache.hudi.hive.HiveSyncTool.syncSchema(HiveSyncTool.java:159) ~[hudi-utilities-bundle_2.11-0.6.1-
    SNAPSHOT.jar:0.6.1-SNAPSHOT] at
    org.apache.hudi.hive.HiveSyncTool.syncHoodieTable(HiveSyncTool.java:130) ~[hudi-utilities-bundle_2.11-0.6.1-
    SNAPSHOT.jar:0.6.1-SNAPSHOT] at
    org.apache.hudi.hive.HiveSyncTool.syncHoodieTable(HiveSyncTool.java:98) ~[hudi-utilities-bundle_2.11-0.6.1-
    SNAPSHOT.jar:0.6.1-SNAPSHOT] at
    org.apache.hudi.utilities.deltastreamer.DeltaSync.syncMeta(DeltaSync.java:510) ~[hudi-utilities-bundle_2.11-
    0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at
    org.apache.hudi.utilities.deltastreamer.DeltaSync.writeToSink(DeltaSync.java:425) ~[hudi-utilities-bundle_2.11-
    0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at
    org.apache.hudi.utilities.deltastreamer.DeltaSync.syncOnce(DeltaSync.java:244) ~[hudi-utilities-bundle_2.11-
    0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at
    org.apache.hudi.utilities.deltastreamer.HoodieDeltaStreamer.lambda$sync$2(HoodieDeltaStreamer.java:161) ~
    [hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at
    org.apache.hudi.common.util.Option.ifPresent(Option.java:96) ~[hudi-utilities-bundle_2.11-0.6.1-
    SNAPSHOT.jar:0.6.1-SNAPSHOT] at
    org.apache.hudi.utilities.deltastreamer.HoodieDeltaStreamer.sync(HoodieDeltaStreamer.java:159) ~[hudi-utilities-
    bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] *** *** Caused by: java.lang.ClassNotFoundException:
    parquet.hadoop.ParquetInputFormat at java.net.URLClassLoader.findClass(URLClassLoader.java:382) ~
    [na:1.8.0_251] at java.lang.ClassLoader.loadClass(ClassLoader.java:418) ~[na:1.8.0_251] at
    sun.misc.Launcher$AppClassLoader.loadClass(Launcher.java:355) ~[na:1.8.0_251] at
    java.lang.ClassLoader.loadClass(ClassLoader.java:351) ~[na:1.8.0_251] ... 20 common frames omitted ```
    Environment Description * Hudi version : 0.6.1 * Spark version : 2.4.3 * Hive version : 2.3.3 * Hadoop
    version : 2.8.5 * Storage (HDFS/S3/GCS..) : HDFS * Running on Docker? (yes/no) : no **Additional context** I
    checked the error code: ```java public static FileInputFormat getInputFormat(HoodieFileFormat baseFileFormat,
    boolean realtime, Configuration conf) { switch (baseFileFormat) { case PARQUET: if (realtime) {
    HoodieParquetRealtimeInputFormat inputFormat = new HoodieParquetRealtimeInputFormat();

inputFormat.setConf(conf); return inputFormat; } else { HoodieParquetInputFormat inputFormat = new HoodieParquetInputFormat(); inputFormat.setConf(conf); return inputFormat; } default: throw new HoodieIOException("Hoodie InputFormat not implemented for base file format " + baseFileFormat); } } public static String getInputFormatClassName(HoodieFileFormat baseFileFormat, boolean realtime, Configuration conf) { FileInputFormat inputFormat = getInputFormat(baseFileFormat, realtime, conf); return inputFormat.getClass().getName(); } ``` I think new a `ParquetInputFormat` may not a good idea for hudi in spark. In `hive-sync` package hudi just need a `FileInputFormat` class name, there is no need to new an object and get this class name. Meanwhile, spark also doesn't have total hive jars to do some action like hive.

13. **title:** [SUPPORT] hudi hive-sync in master branch (0.6.1) can not run by spark
    body: **Describe the problem you faced** Hudi in master branch (0.6.1) can not use `hive-sync` to sync to hive with error ``` Caused by: java.lang.ClassNotFoundException: parquet.hadoop.ParquetInputFormat ``` **Steps to reproduce the behavior:** 1. run a `HoodieDeltaStreamer` task by master `local[2]` and sync hudi table to hive 2. when sync to hive, it report error: ``` java.lang.NoClassDefFoundError: parquet/hadoop/ParquetInputFormat at org.apache.hadoop.hive.ql.io.parquet.MapredParquetInputFormat.<init>(MapredParquetInputFormat.java:46) ~ [hive-exec-1.2.1.spark2.jar:1.2.1.spark2] at org.apache.hudi.hadoop.HoodieParquetInputFormat.<init> (HoodieParquetInputFormat.java:67) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.hadoop.utils.HoodieInputFormatUtils.getInputFormat(HoodieInputFormatUtils.java:82) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.hadoop.utils.HoodieInputFormatUtils.getInputFormatClassName(HoodieInputFormatUtils.java:92) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.hive.HiveSyncTool.syncSchema(HiveSyncTool.java:159) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.hive.HiveSyncTool.syncHoodieTable(HiveSyncTool.java:130) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.hive.HiveSyncTool.syncHoodieTable(HiveSyncTool.java:98) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.utilities.deltastreamer.DeltaSync.syncMeta(DeltaSync.java:510) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.utilities.deltastreamer.DeltaSync.writeToSink(DeltaSync.java:425) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.utilities.deltastreamer.DeltaSync.syncOnce(DeltaSync.java:244) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.utilities.deltastreamer.HoodieDeltaStreamer.lambda$sync$2(HoodieDeltaStreamer.java:161) ~ [hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.common.util.Option.ifPresent(Option.java:96) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.utilities.deltastreamer.HoodieDeltaStreamer.sync(HoodieDeltaStreamer.java:159) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] *** *** Caused by: java.lang.ClassNotFoundException: parquet.hadoop.ParquetInputFormat at java.net.URLClassLoader.findClass(URLClassLoader.java:382) ~ [na:1.8.0_251] at java.lang.ClassLoader.loadClass(ClassLoader.java:418) ~[na:1.8.0_251] at sun.misc.Launcher$AppClassLoader.loadClass(Launcher.java:355) ~[na:1.8.0_251] at java.lang.ClassLoader.loadClass(ClassLoader.java:351) ~[na:1.8.0_251] ... 20 common frames omitted ``` **Environment Description** * Hudi version : 0.6.1 * Spark version : 2.4.3 * Hive version : 2.3.3 * Hadoop version : 2.8.5 * Storage (HDFS/S3/GCS..) : HDFS * Running on Docker? (yes/no) : no **Additional context** I checked the error code: ```java public static FileInputFormat getInputFormat(HoodieFileFormat baseFileFormat, boolean realtime, Configuration conf) { switch (baseFileFormat) { case PARQUET: if (realtime) { HoodieParquetRealtimeInputFormat inputFormat = new HoodieParquetRealtimeInputFormat(); inputFormat.setConf(conf); return inputFormat; } else { HoodieParquetInputFormat inputFormat = new HoodieParquetInputFormat(); inputFormat.setConf(conf); return inputFormat; } default: throw new HoodieIOException("Hoodie InputFormat not implemented for base file format " + baseFileFormat); } } public static String getInputFormatClassName(HoodieFileFormat baseFileFormat, boolean realtime, Configuration conf) { FileInputFormat inputFormat = getInputFormat(baseFileFormat, realtime, conf); return inputFormat.getClass().getName(); } ``` I think new a `ParquetInputFormat` may not a good idea for hudi in spark. In `hive-sync` package hudi just need a `FileInputFormat` class name, there is no need to new an object and get this class name. Meanwhile, spark also doesn't have total hive jars to do some action like hive.

14. **title:** [SUPPORT] hudi hive-sync in master branch (0.6.1) can not run by spark
    body: **Describe the problem you faced** Hudi in master branch (0.6.1) can not use `hive-sync` to sync to hive with error ``` Caused by: java.lang.ClassNotFoundException: parquet.hadoop.ParquetInputFormat ``` **Steps to reproduce the behavior:** 1. run a `HoodieDeltaStreamer` task by master `local[2]` and sync hudi table to hive 2. when sync to hive, it report error: ``` java.lang.NoClassDefFoundError: parquet/hadoop/ParquetInputFormat at org.apache.hadoop.hive.ql.io.parquet.MapredParquetInputFormat.<init>(MapredParquetInputFormat.java:46) ~ [hive-exec-1.2.1.spark2.jar:1.2.1.spark2] at org.apache.hudi.hadoop.HoodieParquetInputFormat.<init> (HoodieParquetInputFormat.java:67) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at

org.apache.hudi.hadoop.utils.HoodieInputFormatUtils.getInputFormat(HoodieInputFormatUtils.java:82) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.hadoop.utils.HoodieInputFormatUtils.getInputFormatClassName(HoodieInputFormatUtils.java:92) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.hive.HiveSyncTool.syncSchema(HiveSyncTool.java:159) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.hive.HiveSyncTool.syncHoodieTable(HiveSyncTool.java:130) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.hive.HiveSyncTool.syncHoodieTable(HiveSyncTool.java:98) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.utilities.deltastreamer.DeltaSync.syncMeta(DeltaSync.java:510) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.utilities.deltastreamer.DeltaSync.writeToSink(DeltaSync.java:425) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.utilities.deltastreamer.DeltaSync.syncOnce(DeltaSync.java:244) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.utilities.deltastreamer.HoodieDeltaStreamer.lambda$sync$2(HoodieDeltaStreamer.java:161) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.common.util.Option.ifPresent(Option.java:96) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.utilities.deltastreamer.HoodieDeltaStreamer.sync(HoodieDeltaStreamer.java:159) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] *** *** Caused by: java.lang.ClassNotFoundException: parquet.hadoop.ParquetInputFormat at java.net.URLClassLoader.findClass(URLClassLoader.java:382) ~[na:1.8.0_251] at java.lang.ClassLoader.loadClass(ClassLoader.java:418) ~[na:1.8.0_251] at sun.misc.Launcher$AppClassLoader.loadClass(Launcher.java:355) ~[na:1.8.0_251] at java.lang.ClassLoader.loadClass(ClassLoader.java:351) ~[na:1.8.0_251] ... 20 common frames omitted ``` **Environment Description** * Hudi version : 0.6.1 * Spark version : 2.4.3 * Hive version : 2.3.3 * Hadoop version : 2.8.5 * Storage (HDFS/S3/GCS..) : HDFS * Running on Docker? (yes/no) : no **Additional context** I checked the error code: ```java public static FileInputFormat getInputFormat(HoodieFileFormat baseFileFormat, boolean realtime, Configuration conf) { switch (baseFileFormat) { case PARQUET: if (realtime) { HoodieParquetRealtimeInputFormat inputFormat = new HoodieParquetRealtimeInputFormat(); inputFormat.setConf(conf); return inputFormat; } else { HoodieParquetInputFormat inputFormat = new HoodieParquetInputFormat(); inputFormat.setConf(conf); return inputFormat; } default: throw new HoodieIOException("Hoodie InputFormat not implemented for base file format " + baseFileFormat); } } public static String getInputFormatClassName(HoodieFileFormat baseFileFormat, boolean realtime, Configuration conf) { FileInputFormat inputFormat = getInputFormat(baseFileFormat, realtime, conf); return inputFormat.getClass().getName(); } ``` I think new a `ParquetInputFormat` may not a good idea for hudi in spark. In `hive-sync` package hudi just need a `FileInputFormat` class name, there is no need to new an object and get this class name. Meanwhile, spark also doesn't have total hive jars to do some action like hive.

15. **title:** [SUPPORT] hudi hive-sync in master branch (0.6.1) can not run by spark
    body: **Describe the problem you faced** Hudi in master branch (0.6.1) can not use `hive-sync` to sync to hive with error ``` Caused by: java.lang.ClassNotFoundException: parquet.hadoop.ParquetInputFormat ``` **Steps to reproduce the behavior:** 1. run a `HoodieDeltaStreamer` task by master `local[2]` and sync hudi table to hive 2. when sync to hive, it report error: ``` java.lang.NoClassDefFoundError: parquet/hadoop/ParquetInputFormat at org.apache.hadoop.hive.ql.io.parquet.MapredParquetInputFormat.<init>(MapredParquetInputFormat.java:46) ~[hive-exec-1.2.1.spark2.jar:1.2.1.spark2] at org.apache.hudi.hadoop.HoodieParquetInputFormat.<init>(HoodieParquetInputFormat.java:67) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.hadoop.utils.HoodieInputFormatUtils.getInputFormat(HoodieInputFormatUtils.java:82) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.hadoop.utils.HoodieInputFormatUtils.getInputFormatClassName(HoodieInputFormatUtils.java:92) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.hive.HiveSyncTool.syncSchema(HiveSyncTool.java:159) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.hive.HiveSyncTool.syncHoodieTable(HiveSyncTool.java:130) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.hive.HiveSyncTool.syncHoodieTable(HiveSyncTool.java:98) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.utilities.deltastreamer.DeltaSync.syncMeta(DeltaSync.java:510) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.utilities.deltastreamer.DeltaSync.writeToSink(DeltaSync.java:425) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.utilities.deltastreamer.DeltaSync.syncOnce(DeltaSync.java:244) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at

org.apache.hudi.utilities.deltastreamer.HoodieDeltaStreamer.lambda$sync$2(HoodieDeltaStreamer.java:161) ~ [hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.common.util.Option.ifPresent(Option.java:96) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.utilities.deltastreamer.HoodieDeltaStreamer.sync(HoodieDeltaStreamer.java:159) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] *** *** Caused by: java.lang.ClassNotFoundException: parquet.hadoop.ParquetInputFormat at java.net.URLClassLoader.findClass(URLClassLoader.java:382) ~ [na:1.8.0_251] at java.lang.ClassLoader.loadClass(ClassLoader.java:418) ~[na:1.8.0_251] at sun.misc.Launcher$AppClassLoader.loadClass(Launcher.java:355) ~[na:1.8.0_251] at java.lang.ClassLoader.loadClass(ClassLoader.java:351) ~[na:1.8.0_251] ... 20 common frames omitted ``` **Environment Description** * Hudi version : 0.6.1 * Spark version : 2.4.3 * Hive version : 2.3.3 * Hadoop version : 2.8.5 * Storage (HDFS/S3/GCS..) : HDFS * Running on Docker? (yes/no) : no **Additional context** I checked the error code: ```java public static FileInputFormat getInputFormat(HoodieFileFormat baseFileFormat, boolean realtime, Configuration conf) { switch (baseFileFormat) { case PARQUET: if (realtime) { HoodieParquetRealtimeInputFormat inputFormat = new HoodieParquetRealtimeInputFormat(); inputFormat.setConf(conf); return inputFormat; } else { HoodieParquetInputFormat inputFormat = new HoodieParquetInputFormat(); inputFormat.setConf(conf); return inputFormat; } default: throw new HoodieIOException("Hoodie InputFormat not implemented for base file format " + baseFileFormat); } } public static String getInputFormatClassName(HoodieFileFormat baseFileFormat, boolean realtime, Configuration conf) { FileInputFormat inputFormat = getInputFormat(baseFileFormat, realtime, conf); return inputFormat.getClass().getName(); } ``` I think new a `ParquetInputFormat` may not a good idea for hudi in spark. In `hive-sync` package hudi just need a `FileInputFormat` class name, there is no need to new an object and get this class name. Meanwhile, spark also doesn't have total hive jars to do some action like hive.

16. **title:** [SUPPORT] hudi hive-sync in master branch (0.6.1) can not run by spark
    body: **Describe the problem you faced** Hudi in master branch (0.6.1) can not use `hive-sync` to sync to hive with error ``` Caused by: java.lang.ClassNotFoundException: parquet.hadoop.ParquetInputFormat ``` **Steps to reproduce the behavior:** 1. run a `HoodieDeltaStreamer` task by master `local[2]` and sync hudi table to hive 2. when sync to hive, it report error: ``` java.lang.NoClassDefFoundError: parquet/hadoop/ParquetInputFormat at org.apache.hadoop.hive.ql.io.parquet.MapredParquetInputFormat.<init>(MapredParquetInputFormat.java:46) ~ [hive-exec-1.2.1.spark2.jar:1.2.1.spark2] at org.apache.hudi.hadoop.HoodieParquetInputFormat.<init> (HoodieParquetInputFormat.java:67) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.hadoop.utils.HoodieInputFormatUtils.getInputFormat(HoodieInputFormatUtils.java:82) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.hadoop.utils.HoodieInputFormatUtils.getInputFormatClassName(HoodieInputFormatUtils.java:92) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.hive.HiveSyncTool.syncSchema(HiveSyncTool.java:159) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.hive.HiveSyncTool.syncHoodieTable(HiveSyncTool.java:130) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.hive.HiveSyncTool.syncHoodieTable(HiveSyncTool.java:98) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.utilities.deltastreamer.DeltaSync.syncMeta(DeltaSync.java:510) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.utilities.deltastreamer.DeltaSync.writeToSink(DeltaSync.java:425) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.utilities.deltastreamer.DeltaSync.syncOnce(DeltaSync.java:244) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.utilities.deltastreamer.HoodieDeltaStreamer.lambda$sync$2(HoodieDeltaStreamer.java:161) ~ [hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.common.util.Option.ifPresent(Option.java:96) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] at org.apache.hudi.utilities.deltastreamer.HoodieDeltaStreamer.sync(HoodieDeltaStreamer.java:159) ~[hudi-utilities-bundle_2.11-0.6.1-SNAPSHOT.jar:0.6.1-SNAPSHOT] *** *** Caused by: java.lang.ClassNotFoundException: parquet.hadoop.ParquetInputFormat at java.net.URLClassLoader.findClass(URLClassLoader.java:382) ~ [na:1.8.0_251] at java.lang.ClassLoader.loadClass(ClassLoader.java:418) ~[na:1.8.0_251] at sun.misc.Launcher$AppClassLoader.loadClass(Launcher.java:355) ~[na:1.8.0_251] at java.lang.ClassLoader.loadClass(ClassLoader.java:351) ~[na:1.8.0_251] ... 20 common frames omitted ``` **Environment Description** * Hudi version : 0.6.1 * Spark version : 2.4.3 * Hive version : 2.3.3 * Hadoop version : 2.8.5 * Storage (HDFS/S3/GCS..) : HDFS * Running on Docker? (yes/no) : no **Additional context** I checked the error code: ```java public static FileInputFormat getInputFormat(HoodieFileFormat baseFileFormat, boolean realtime, Configuration conf) { switch (baseFileFormat) { case PARQUET: if (realtime) { HoodieParquetRealtimeInputFormat inputFormat = new HoodieParquetRealtimeInputFormat(); inputFormat.setConf(conf); return inputFormat; } else { HoodieParquetInputFormat inputFormat = new

HoodieParquetInputFormat(); inputFormat.setConf(conf); return inputFormat; } default: throw new HoodieIOException("Hoodie InputFormat not implemented for base file format " + baseFileFormat); } } public static String getInputFormatClassName(HoodieFileFormat baseFileFormat, boolean realtime, Configuration conf) { FileInputFormat inputFormat = getInputFormat(baseFileFormat, realtime, conf); return inputFormat.getClass().getName(); } ``` I think new a `ParquetInputFormat` may not a good idea for hudi in spark. In `hive-sync` package hudi just need a `FileInputFormat` class name, there is no need to new an object and get this class name. Meanwhile, spark also doesn't have total hive jars to do some action like hive.

github_issues_comments:

1. @cdmikechen : Curious, Are you not seeing this with older version of Hudi (0.5.x) ?
2. **body:** @cdmikechen : Also, if you look at integration tests ITTestHoodieDemo, we cover the tests with hive syncing and this test has been passing for us. Can you take a look at the tests to see what the difference is ?
   label: test
3. @bvaradar Hudi 0.5.X or 0.6.0 when `HoodieInputFormatUtils` not exists don't happened this. The old code looks like the following which is to get the classname directly instead of the need to new class. ```java String inputFormatClassName = cfg.usePreApacheInputFormat ? com.uber.hoodie.hadoop.HoodieInputFormat.class.getName() : HoodieParquetInputFormat.class.getName(); ```
4. THanks @cdmikechen . Agree this is a new code change in 0.6.x but can you check why the integration tests which enables hive sync for delta-streamer passes.
5. cc @garyli1019 have you encountered this before?
6. > cc @garyli1019 have you encountered this before? @vinothchandar No, I don't use hive sync anywhere. This issue seems like a version mismatch for me. I do agree that we don't need to create a new class here to get the name. People using different versions of Hive might run into a similar issue.
7. @garyli1019 can you please submit a PR fixing this?
8. > @garyli1019 can you please submit a PR fixing this? yes sure EDIT: Before I do this, @cdmikechen are you interested to fix this? If so, please let me know. Happy to have you to drive this fix :)
9. I verified by running deltastreamer using docker demo : The below command ran fine without any issue and I was able to check table registered successfully. But, it is good to fix it. THis is not a blocker for release though. root@adhoc-2:/opt# spark-submit --class org.apache.hudi.utilities.deltastreamer.HoodieDeltaStreamer $HUDI_UTILITIES_BUNDLE --table-type COPY_ON_WRITE --source-class org.apache.hudi.utilities.sources.JsonKafkaSource --source-ordering-field ts --target-base-path /user/hive/warehouse/stock_ticks_cow2 --target-table stock_ticks_cow --props /var/demo/config/kafka-source.properties --schemaprovider-class org.apache.hudi.utilities.schema.FilebasedSchemaProvider --enable-sync --hoodie-conf hoodie.datasource.hive_sync.table=dumm2 --hoodie-conf hoodie.datasource.hive_sync.use_jdbc=true --hoodie-conf hoodie.datasource.hive_sync.jdbcurl=jdbc:hive2://hiveserver:10000 --hoodie-conf hoodie.datasource.hive_sync.partition_extractor_class=org.apache.hudi.hive.MultiPartKeysValueExtractor --hoodie-conf hoodie.datasource.hive_sync.partition_fields=yr,month,day
10. Thanks for confirming balaji!
11. > @cdmikechen : Also, if you look at integration tests ITTestHoodieDemo, we cover the tests with hive syncing and this test has been passing for us. Can you take a look at the tests to see what the difference is ? @bvaradar I checked `hudi-integ-test` package and found the reason: In `hudi-integ-test` pom.xml where contains `ITTestHoodieDemo`, hudi contains `hudi-exec-2.3.1` In pom dependencies. So that if we new a `MapredParquetInputFormat` class, hudi will use this class by `hudi-exec-2.3.1`. ```java package org.apache.hadoop.hive.ql.io.parquet; import java.io.IOException; import org.apache.hadoop.hive.ql.exec.Utilities; import org.apache.hadoop.hive.ql.exec.vector.VectorizedInputFormatInterface; import org.apache.hadoop.hive.ql.io.parquet.read.DataWritableReadSupport; import org.apache.hadoop.hive.ql.io.parquet.read.ParquetRecordReaderWrapper; import org.apache.hadoop.io.ArrayWritable; import org.apache.hadoop.io.NullWritable; import org.apache.hadoop.mapred.FileInputFormat; import org.apache.hadoop.mapred.InputSplit; import org.apache.hadoop.mapred.JobConf; import org.apache.hadoop.mapred.RecordReader; import org.apache.hadoop.mapred.Reporter; import org.slf4j.Logger; import org.slf4j.LoggerFactory; import org.apache.parquet.hadoop.ParquetInputFormat; public class MapredParquetInputFormat extends FileInputFormat<NullWritable, ArrayWritable> implements VectorizedInputFormatInterface { ``` But if we just use a standalone spark environmental without hive-2.3.1 dependencies (like starting a new project and only depend spark lib), hudi will use `hive-exec-1.2.1-spark2`. ```java package org.apache.hadoop.hive.ql.io.parquet; import java.io.IOException; import org.apache.commons.logging.Log; import org.apache.commons.logging.LogFactory; import org.apache.hadoop.hive.ql.exec.Utilities; import org.apache.hadoop.hive.ql.exec.vector.VectorizedInputFormatInterface; import org.apache.hadoop.hive.ql.io.parquet.read.DataWritableReadSupport; import org.apache.hadoop.hive.ql.io.parquet.read.ParquetRecordReaderWrapper; import

org.apache.hadoop.io.ArrayWritable; import org.apache.hadoop.mapred.FileInputFormat; import org.apache.hadoop.mapred.RecordReader; import parquet.hadoop.ParquetInputFormat; public class MapredParquetInputFormat extends FileInputFormat<Void, ArrayWritable> { ```

12. @cdmikechen : The integration test actually brings up a dockerized environment and runs spark-submit command. So, the dependencies specified in hudi-integ-test/pom.xml should not be part of the deltastreamer run. As you can look from my spark-submit command in my previous comment, only HUDI_UTILITIES_BUNDLE is passed. otherwise, it is only spark runtime environment. The ParquetInputFormatClass is part of parquer-hadoop-bundle. root@adhoc-2:/opt# grep -r 'parquet.hadoop.ParquetInputFormat' $SPARK_HOME/jars/* Binary file /opt/spark/jars/parquet-hadoop-1.10.1.jar matches Binary file /opt/spark/jars/parquet-hadoop-bundle-1.6.0.jar matches root@adhoc-2:/opt# root@adhoc-2:/opt# jar tf /opt/spark/jars/parquet-hadoop-bundle-1.6.0.jar | grep ParquetInputFormat.class parquet/hadoop/mapred/DeprecatedParquetInputFormat.class parquet/hadoop/ParquetInputFormat.class Are you using the spark distribution prebuilt with hadoop ?

13. @bvaradar Thanks for your reminder, I finally found my mistake: I use hudi in a maven project with spark dependencies. I noticed that hudi remove `com.twitter:parquet-hadoop-bundle`, so that I also removed this dependency in my project. ``` <exclusions> <exclusion> <groupId>com.twitter</groupId> <artifactId>parquet-hadoop-bundle</artifactId> </exclusion> </exclusions> ``` Therefore, when starting a spark task in this maven project, hudi can not find `parquet-hadoop-bundle-1.6.0.jar` and `parquet.hadoop.ParquetInputFormat` class. If I add this dependency, it should not report this error . Meanwhile, I think my another suggestion which we should avoid new `FileInputFormat` to just get class name should be fixed.

14. Thanks @cdmikechen for clarifying. Agree on not having to instantiate the input format. @garyli1019 has a PR for this : https://github.com/apache/hudi/pull/2008 Closing this ticket !!

github_pulls:

1. **title:** [MINOR] fix get classname for hive sync
   body: ## What is the purpose of the pull request Fix #2005 ## Brief change log - *avoid creating a new class when getting the class name in HoodieInputFormatUtils* ## Verify this pull request This pull request is already covered by existing tests, such as *(please describe tests)*. ## Committer checklist - [ ] Has a corresponding JIRA in PR title & commit - [ ] Commit message is descriptive of the change - [ ] CI is green - [ ] Necessary doc changes done or have another open PR - [ ] For large changes, please consider breaking it into sub-tasks under an umbrella JIRA.

2. **title:** [MINOR] fix get classname for hive sync
   body: ## What is the purpose of the pull request Fix #2005 ## Brief change log - *avoid creating a new class when getting the class name in HoodieInputFormatUtils* ## Verify this pull request This pull request is already covered by existing tests, such as *(please describe tests)*. ## Committer checklist - [ ] Has a corresponding JIRA in PR title & commit - [ ] Commit message is descriptive of the change - [ ] CI is green - [ ] Necessary doc changes done or have another open PR - [ ] For large changes, please consider breaking it into sub-tasks under an umbrella JIRA.

3. **title:** [MINOR] fix get classname for hive sync
   body: ## What is the purpose of the pull request Fix #2005 ## Brief change log - *avoid creating a new class when getting the class name in HoodieInputFormatUtils* ## Verify this pull request This pull request is already covered by existing tests, such as *(please describe tests)*. ## Committer checklist - [ ] Has a corresponding JIRA in PR title & commit - [ ] Commit message is descriptive of the change - [ ] CI is green - [ ] Necessary doc changes done or have another open PR - [ ] For large changes, please consider breaking it into sub-tasks under an umbrella JIRA.

4. **title:** [MINOR] fix get classname for hive sync
   body: ## What is the purpose of the pull request Fix #2005 ## Brief change log - *avoid creating a new class when getting the class name in HoodieInputFormatUtils* ## Verify this pull request This pull request is already covered by existing tests, such as *(please describe tests)*. ## Committer checklist - [ ] Has a corresponding JIRA in PR title & commit - [ ] Commit message is descriptive of the change - [ ] CI is green - [ ] Necessary doc changes done or have another open PR - [ ] For large changes, please consider breaking it into sub-tasks under an umbrella JIRA.

5. **title:** [MINOR] fix get classname for hive sync
   body: ## What is the purpose of the pull request Fix #2005 ## Brief change log - *avoid creating a new class when getting the class name in HoodieInputFormatUtils* ## Verify this pull request This pull request is already covered by existing tests, such as *(please describe tests)*. ## Committer checklist - [ ] Has a corresponding JIRA in PR title & commit - [ ] Commit message is descriptive of the change - [ ] CI is green - [ ] Necessary doc changes done or have another open PR - [ ] For large changes, please consider breaking it into sub-tasks under an umbrella JIRA.

github_pulls_comments:

1. Hi @cdmikechen , would you please review this PR? If this looks good then we can merge.

github_pulls_reviews:

1. this is going to fail now, probably given we have a HFile base now. @garyli1019 can you please rebase and fix.
2. thanks for letting me know

jira_issues:

jira_issues_comments: