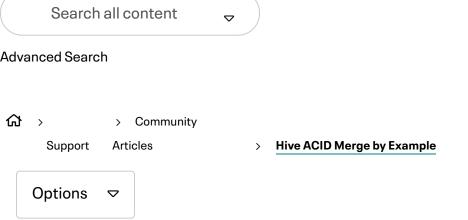
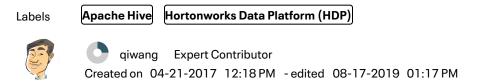


COMMUNITY ARTICLES

Find and share helpful community-sourced technical articles



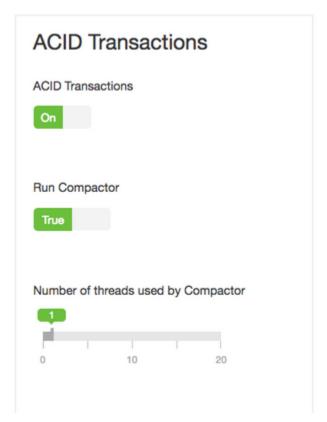
Hive ACID Merge by Example *⊗* ■



With the release of HDP 2.6 comes the Hive ACID Merge feature. This is long sought after feature for people who want to batch update table content in ACID manner. This tutorial will walk you through step-by-step how to use this new feature with some really simple dataset. The example is done on HDP 2.6.0.3 installation. First, ACID in Hive need to be turned on in Ambari UI

Cloudera uses cookies to provide and improve our site's services.

By using this site, you consent to use of cookies as outlined in Cloudera's Privacy and Data Policies.



Then we will create 2 tables, one as the target of merge and one as the source of merge. Please note that the target table must be bucketed, set as transaction enabled and stored in orc format.

```
CREATE TABLE merge_data.transactions(
    ID int,
    TranValue string,
    last_update_user string)
PARTITIONED BY (tran_date string)
CLUSTERED BY (ID) into 5 buckets
STORED AS ORC TBLPROPERTIES ('transactional'='true');

CREATE TABLE merge_data.merge_source(
    ID int,
    TranValue string,
    tran_date string)
STORED AS ORC;
```

Cloudera uses cookies to provide and improve our site's services. Then we will populate the target and source table with some data. By using this site, you consent to use of cookies as outlined in Cloudera's Privacy and Data Policies.

```
INSERT INTO merge_data.transactions PARTITION (tran_date) VALUES
   'value 01', 'creation', '20170410'),
(1,
    'value_02', 'creation', '20170410'),
(2,
(3,
    'value 03', 'creation', '20170410'),
    'value_04', 'creation', '20170410'),
    'value 05', 'creation', '20170413'),
(5,
    'value_06', 'creation', '20170413'),
(6,
    'value_07', 'creation', '20170413'),
(7,
   'value_08', 'creation', '20170413'),
(8,
(9, 'value_09', 'creation', '20170413'),
(10, 'value 10', 'creation', '20170413');
INSERT INTO merge_data.merge_source VALUES
(1, 'value_01', '20170410'),
(4, NULL, '20170410'),
(7, 'value_77777', '20170413'),
(8, NULL, '20170413'),
(8, 'value_08', '20170415'),
(11, 'value_11', '20170415');
```

When we examine the 2 tables, we would expect after the merge, row 1 should be left untouched, row 4 should be deleted (we imply a business rule here: NULL value indicates deletion), row 7 will be update and row 11 will be insertion for new value.

The more interesting use case is with row 8, where it involves move the row from one partition to another. Merge currently does not support change of partition value on the fly. This need to happen as delete in the old partition and insert in the new partition. In real world use case, you need to construct the source table base on this critiria. Then we will create the merge statement as the following. Please note that not all 3 WHEN of the merge statements need to exist, it is fine to have only 2 or even 1 of the WHEN statement. We labeled the data with different last_update_user. For more details on Hive Merge, please refer to Hive document

```
MERGE INTO merge_data.transactions AS T

USING merge_data.merge_source AS S

ON T.ID = S.ID and T.tran_date = S.tran_date

WHEN MATCHED AND (T,TranValue != S.TranValue AND S.TranValue IS NOT NULL) THEN UPDATE SE Cloudera uses cookies to provide and improve our site's services.

WHEN MATCHED AND S.TranValue IS NULL THEN DELETE. By using this site, you consent to use of cookies as outlined in Cloudera's Privacy and Data Policies.

WHEN NOT MATCHED THEN INSERT VALUES (S.ID, S.TranValue, 'merge_insert', S.tran_date);

Accept Cookies
```

As part of the update clause, the set value statement should not contain the target table decorator "T.", otherwise you will get SQL compile error.

Once the merge finished, re-examining the data shows the data is merged just as expected row 1 wasn't changed; row 4 was deleted; row 7 was updated and row 11 was inserted. And row 8, as we see, got moved to a new partition.

	•		
id		transactions.last_update_user	_
1	value_01	creation	20170410
2	value_02	creation	20170410
3	value_03	creation	20170410
5	value_05	creation	20170413
6	value_06	creation	20170413
7	value_77777	merge_update	20170413
8	value_08	merge_insert	20170415
9	value_09	creation	20170413
10	value_10	creation	20170413
11	value 11	merge insert	20170415

This simple example provides instructions on how to use ACID merge with HDP 2.6 or later. Of course the real world use case would be much more complicated comparing to this oversimplified example, but they all follow the same principles. And the beauty behind this is, you don't need to create a ETL process to accomplish this any more.

TAGS (5) Data Processing hdp-2.6.0 Hive How-ToTutorial merge

Comments

Cloudera uses cookies to provide and improve our site's services.

By using this site, you consent to use of codies as outlined in Poedera's Privacy and Data Policies.



:

Re: Hive ACID Merge by Example



lakshmana_maddi New Contributor
Created on 08-20-2018 12:57 PM

Please let me know why below merge command is throwing the error message.

---- merge command ----

MERGE INTO merge_data.transactions AS T USING merge_data.merge_source AS S ON T.ID = S.ID WHEN MATCHED THEN UPDATE SET TranValue = S.TranValue, last_update_user = 'merge_update' WHEN NOT MATCHED THEN INSERT VALUES (S.ID, S.TranValue, 'merge_insert', S.tran_date);

----- Error ----

Error: Failure while running task: java.lang.RuntimeException: java.lang.RuntimeException: Hive Runtime Error while closing operators: org.apache.hadoop.hive.ql.metadata.HiveException:

org.apache.hadoop.hive.ql.metadata.HiveException: Error evaluating

cardinality_violation(_col0,_col1)

Caused by: java.lang.RuntimeException: Hive Runtime Error while closing operators:

org.apache.hadoop.hive.ql.metadata.HiveException:

org.apache.hadoop.hive.ql.metadata.HiveException: Error evaluating

cardinality_violation(_col0,_col1)

Caused by: java.lang.RuntimeException: Cardinality Violation in Merge statement: [61, 3,

0],20170415

Re: Hive ACID Merge by Example



nishan_unilex Not applicable Created on 08-22-2018 05:55 AM

very nice article. thank you for sharing.

Re: Hive ACID Merge by Example



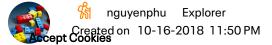
nishan_unilex Not applicable Created on 08-22-2018 05:57 AM

great article

Cloudera uses cookies to provide and improve our site's services.

Re: Hive ACID Merge by Example

By using this site, you consent to use of cookies as outlined in Cloudera's Privacy and Data Policies.



:

Is it possible to batch insert into a table that contains a field of type array<string>. What is the syntax for that?

Explore the Community

 \triangle

Already a User?

Don't have an account?

Sign In

Register

ANNOUNCEMENTS

Product Announcements

[ANNOUNCE] New Cloudera JDBC 2.6.20 Driver for Apache Impala Released

Product Announcements

Transition to private repositories for CDH, HDP and HDF

Product Announcements

[ANNOUNCE] New Applied ML Research from Cloudera Fast Forward: Few-Shot Text Classification

Product Announcements

[ANNOUNCE] New JDBC 2.6.13 Driver for Apache Hive Released

Product Announcements

[ANNOUNCE] Refreshed Research from Cloudera Fast Forward: Semantic Image Search and Federated Learning

View More Announcements

VERSION HISTORY

Revision #: 2

2 of 2

Last update: 08-17-2019

01:17 PM

Updated by:



qiwang

Cloudera uses cookies to provide and improve our site's services.

By using this site, you consent to use of cookies as outlined in Cloudera's Privacy and Data Policies.

Accept Cookies
CONTRIBUTORS



Top Kudoed Authors

THE STATE OF THE S	TimothySpann (e)	₫ 758
	sshimpi	₼ 379
	♦ KuldeepK	 316
	cstanca	₼ 309
	abajwa 🥝	☆ 268

View Full Leaderboard

Cloudera uses cookies to provide and improve our site's services.

By using this site, you consent to use of cookies as outlined in Cloudera's Privacy and Data Policies.

RECOMPRIENDATIONS

	Merging EDW with Hadoop
	JDBC error with Hive ACID Enabled
Q	Are there optimizations for ACID Merge with 100+ m
	Storm Hive Streaming Example
	BMC Spark Hive Example
₽,✓	ACID and CUD transactions
	Hadoop: the Definitive Guide - Example Source Cod
₽,✓	Hive Compaction for ACID Transactions
₽,✓	Hive ACID Table count query launching too many map
₽,✓	unable to view the content of acid table in sp

My Cloudera Home

Community

Documentation

Knowledge Base

Partners

Support

PRODUCTS

Cloudera Data Platform

Data Hub

Data Warehouse

Machine Learning Cloudera uses cookies to provide and improve our site's services.

EXPISITE WAS CONSENT to use of cookies as outlined in Cloudera's Privacy and Data Policies.

Data Flow

Accept Cookies
Data Science Workbench

Enterprise Data Hub			
SOLUTIONS			
Customer Analytics			
oT / Connected Products			
Security, Risk, & Compliance			
ndustry Solutions			
Solutions Gallery			
DOWNLOADS			
Cloudera Data Flow			
CDP Private Cloud			
Cloudera Director			
MORE			
About			
News & Blogs			
nvestor Relations			
Events			
Customers			
Careers			
Security			
Community Guidelines			
Community FAQs			
Contact Us Holted States oblies 8 769 de 488 improve our site's services.			
Noteer වියිසි දිට විශ්වේ සිට වියිසිට ගත්තම improve our site's services. Outsings bliss යියි.ye il කියිවෙනිරිව ගින්සිති cookies as outlined in Cloudera's Privacy and Data Policies.			

Accept Cookies © 2021 Cloudera, Inc. All rights reserved.