

用StreamSets实现MySQL中变化数据实时写入Kudu

mysqltokudu.json

```
{
  "pipelineConfig" : {
    "schemaVersion" : 5,
    "version" : 9,
    "pipelineId" : "teststream34a0a9ba-af13-4f77-a40d-53fd35e3c132",
    "title" : "mysqltokuducopy",
    "description" : "",
    "uuid" : "015b4cd5-5a02-4a8d-a601-bb24a92fdf8b",
    "configuration" : [ {
      "name" : "executionMode",
      "value" : "STANDALONE"
    }, {
      "name" : "edgeHttpUrl",
      "value" : "http://localhost:18633"
    }, {
      "name" : "deliveryGuarantee",
      "value" : "AT_LEAST_ONCE"
    }, {
      "name" : "startEventStage",
      "value" : "streamsets-datacollector-basic-
lib::com_streamsets_pipeline_stage_destination_devnull_ToErrorNullDTarget::1"
    }, {
      "name" : "stopEventStage",
      "value" : "streamsets-datacollector-basic-
lib::com_streamsets_pipeline_stage_destination_devnull_ToErrorNullDTarget::1"
    }, {
      "name" : "shouldRetry",
      "value" : true
    }, {
      "name" : "retryAttempts",
      "value" : -1
    }, {
      "name" : "memoryLimit",
      "value" : "${jvm:maxMemoryMB() * 0.85}"
    }, {
      "name" : "memoryLimitExceeded",
      "value" : "LOG"
    }, {
```

```

    "name" : "notifyOnStates",
    "value" : [ "RUN_ERROR", "STOPPED", "FINISHED" ]
  }, {
    "name" : "emailIDs",
    "value" : [ ]
  }, {
    "name" : "constants",
    "value" : [ ]
  }, {
    "name" : "badRecordsHandling",
    "value" : "streamsets-datacollector-basic-
lib::com_streamsets_pipeline_stage_destination_recordstolocalfilesystem_ToErrorLocalFSDTarget::1"
  }, {
    "name" : "errorRecordPolicy",
    "value" : "ORIGINAL_RECORD"
  }, {
    "name" : "workerCount",
    "value" : 0
  }, {
    "name" : "clusterSlaveMemory",
    "value" : 2048
  }, {
    "name" : "clusterSlaveJavaOpts",
    "value" : "-XX:+UseConcMarkSweepGC -XX:+UseParNewGC -Dlog4j.debug"
  }, {
    "name" : "clusterLauncherEnv",
    "value" : [ ]
  }, {
    "name" : "mesosDispatcherURL",
    "value" : null
  }, {
    "name" : "hdfsS3ConfDir",
    "value" : null
  }, {
    "name" : "rateLimit",
    "value" : 0
  }, {
    "name" : "maxRunners",
    "value" : 0
  }, {
    "name" : "shouldCreateFailureSnapshot",
    "value" : true
  }, {

```

```
    "name" : "runnerIdleTime",
    "value" : 60
  }, {
    "name" : "webhookConfigs",
    "value" : [ ]
  }, {
    "name" : "sparkConfigs",
    "value" : [ ]
  }, {
    "name" : "statsAggregatorStage",
    "value" : ""
  } ],
  "uiInfo" : {
    "previewConfig" : {
      "previewSource" : "CONFIGURED_SOURCE",
      "batchSize" : 10,
      "timeout" : 30000,
      "writeToDestinations" : false,
      "executeLifecycleEvents" : false,
      "showHeader" : false,
      "showFieldType" : true,
      "rememberMe" : false
    }
  },
  "fragments" : [ ],
  "stages" : [ {
    "instanceName" : "MySQLBinaryLog_01",
    "library" : "streamsets-datacollector-mysql-binlog-lib",
    "stageName" : "com_streamsets_pipeline_stage_origin_mysql_MysqlDSource",
    "stageVersion" : "2",
    "configuration" : [ {
      "name" : "config.hostname",
      "value" : "192.168.34.148"
    }, {
      "name" : "config.port",
      "value" : "3306"
    }, {
      "name" : "config.username",
      "value" : "maxwell"
    }, {
      "name" : "config.password",
      "value" : "Mysql157@db"
    }, {
```

```
    "name" : "config.serverId",
    "value" : "1"
  }, {
    "name" : "config.maxBatchSize",
    "value" : 1000
  }, {
    "name" : "config.maxWaitTime",
    "value" : 1000
  }, {
    "name" : "config.connectTimeout",
    "value" : 5000
  }, {
    "name" : "config.enableKeepAlive",
    "value" : true
  }, {
    "name" : "config.keepAliveInterval",
    "value" : 60000
  }, {
    "name" : "config.useSsl",
    "value" : false
  }, {
    "name" : "config.startFromBeginning",
    "value" : false
  }, {
    "name" : "config.initialOffset",
    "value" : null
  }, {
    "name" : "config.includeTables",
    "value" : null
  }, {
    "name" : "config.ignoreTables",
    "value" : null
  }, {
    "name" : "stageOnRecordError",
    "value" : "TO_ERROR"
  } ],
  "uiInfo" : {
    "description" : "",
    "label" : "MySQL Binary Log 1",
    "xPos" : -13,
    "yPos" : 78,
    "stageType" : "SOURCE"
  },
}
```

```

    "inputLanes" : [ ],
    "outputLanes" : [ "MySQLBinaryLog_01OutputLane15301760229120" ],
    "eventLanes" : [ ],
    "services" : [ ]
  }, {
    "instanceName" : "StreamSelector_01",
    "library" : "streamsets-datacollector-basic-lib",
    "stageName" : "com_streamsets_pipeline_stage_processor_selector_SelectorDProcessor",
    "stageVersion" : "1",
    "configuration" : [ {
      "name" : "lanePredicates",
      "value" : [ {
        "outputLane" : "StreamSelector_01OutputLane1530179909333",
        "predicate" : "${record:value('/Table')}==cdc_test"
      }, {
        "outputLane" : "StreamSelector_01OutputLane1530179806488",
        "predicate" : "default"
      } ]
    } ],
    "name" : "stageOnRecordError",
    "value" : "TO_ERROR"
  }, {
    "name" : "stageRequiredFields",
    "value" : [ ]
  }, {
    "name" : "stageRecordPreconditions",
    "value" : [ ]
  } ],
  "uiInfo" : {
    "description" : "",
    "label" : "Stream Selector 1",
    "xPos" : 189,
    "yPos" : 76,
    "stageType" : "PROCESSOR"
  },
  "inputLanes" : [ "MySQLBinaryLog_01OutputLane15301760229120" ],
  "outputLanes" : [ "StreamSelector_01OutputLane1530179909333",
"StreamSelector_01OutputLane1530179806488" ],
  "eventLanes" : [ ],
  "services" : [ ]
}, {
  "instanceName" : "StreamSelector_02",
  "library" : "streamsets-datacollector-basic-lib",

```

```

"stageName" : "com_streamsets_pipeline_stage_processor_selector_SelectorDProcessor",
"stageVersion" : "1",
"configuration" : [ {
  "name" : "lanePredicates",
  "value" : [ {
    "outputLane" : "StreamSelector_02OutputLane1530234523718",
    "predicate" : "${record:value('/Type')}='DELETE'"
  }, {
    "outputLane" : "StreamSelector_02OutputLane1530234513027",
    "predicate" : "default"
  } ]
}, {
  "name" : "stageOnRecordError",
  "value" : "TO_ERROR"
}, {
  "name" : "stageRequiredFields",
  "value" : [ ]
}, {
  "name" : "stageRecordPreconditions",
  "value" : [ ]
} ],
"uiInfo" : {
  "description" : "",
  "label" : "Stream Selector 2",
  "xPos" : 396,
  "yPos" : 6,
  "stageType" : "PROCESSOR"
},
"inputLanes" : [ "StreamSelector_01OutputLane1530179909333" ],
"outputLanes" : [ "StreamSelector_02OutputLane1530234523718",
"StreamSelector_02OutputLane1530234513027" ],
"eventLanes" : [ ],
"services" : [ ]
}, {
  "instanceName" : "JavaScriptEvaluator_01",
  "library" : "streamsets-datacollector-basic-lib",
  "stageName" :
"com_streamsets_pipeline_stage_processor_javascript_JavaScriptDProcessor",
  "stageVersion" : "2",
  "configuration" : [ {
    "name" : "processingMode",
    "value" : "RECORD"
  }, {

```

```

        "name" : "initScript",
        "value" : ""
    }, {
        "name" : "script",
        "value" : "for(var i = 0; i < records.length; i++) {\n  try { \n      var
newRecord = sdcFunctions.createRecord(true);\n      newRecord.value =
records[i].value['OldData'];\n      newRecord.value.Type = records[i].value['Type'];\n
newRecord.value.Database = records[i].value['Database'];\n      newRecord.value.Table =
records[i].value['Table'];\n      log.info(records[i].value['Type'])\n
output.write(newRecord);\n  } catch (e) {\n      // Send record to error\n
error.write(records[i], e);\n  }\n}"
    }, {
        "name" : "destroyScript",
        "value" : ""
    }, {
        "name" : "stageOnRecordError",
        "value" : "TO_ERROR"
    }, {
        "name" : "stageRequiredFields",
        "value" : [ ]
    }, {
        "name" : "stageRecordPreconditions",
        "value" : [ ]
    } ],
    "uiInfo" : {
        "description" : "",
        "label" : "JavaScript Evaluator-delete",
        "xPos" : 658,
        "yPos" : 1,
        "stageType" : "PROCESSOR"
    },
    "inputLanes" : [ "StreamSelector_02OutputLane1530234523718" ],
    "outputLanes" : [ "JavaScriptEvaluator_01OutputLane15302345785040" ],
    "eventLanes" : [ ],
    "services" : [ ]
}, {
    "instanceName" : "JavaScriptEvaluator_02",
    "library" : "streamsets-datacollector-basic-lib",
    "stageName" :
"com_streamsets_pipeline_stage_processor_javascript_JavaScriptDProcessor",
    "stageVersion" : "2",
    "configuration" : [ {
        "name" : "processingMode",

```

```

        "value" : "RECORD"
    }, {
        "name" : "initScript",
        "value" : ""
    }, {
        "name" : "script",
        "value" : "for(var i = 0; i < records.length; i++) {\n  try { \n      var
newRecord = sdcFunctions.createRecord(true);\n      newRecord.value =
records[i].value['Data'];\n      newRecord.value.Type = records[i].value['Type'];\n
newRecord.value.Database = records[i].value['Database'];\n      newRecord.value.Table =
records[i].value['Table'];\n      log.info(records[i].value['Type'])\n
output.write(newRecord);\n  } catch (e) {\n      // Send record to error\n
error.write(records[i], e);\n  }\n}"
    }, {
        "name" : "destroyScript",
        "value" : ""
    }, {
        "name" : "stageOnRecordError",
        "value" : "TO_ERROR"
    }, {
        "name" : "stageRequiredFields",
        "value" : [ ]
    }, {
        "name" : "stageRecordPreconditions",
        "value" : [ ]
    } ],
    "uiInfo" : {
        "description" : "",
        "label" : "JavaScript Evaluator-upsert",
        "xPos" : 631,
        "yPos" : 147,
        "stageType" : "PROCESSOR"
    },
    "inputLanes" : [ "StreamSelector_02OutputLane1530234513027" ],
    "outputLanes" : [ "JavaScriptEvaluator_02OutputLane15302346847350" ],
    "eventLanes" : [ ],
    "services" : [ ]
}, {
    "instanceName" : "Kudu_01",
    "library" : "streamsets-datacollector-apache-kudu_1_6-lib",
    "stageName" : "com_streamsets_pipeline_stage_destination_kudu_KuduDTarget",
    "stageVersion" : "4",
    "configuration" : [ {

```



```

    "name" : "kuduConfigBean.kuduMaster",
    "value" : "nn"
  }, {
    "name" : "kuduConfigBean.tableNameTemplate",
    "value" : "impala::default.${record:value('/Table')}"
  }, {
    "name" : "kuduConfigBean.fieldMappingConfigs",
    "value" : [ { } ]
  }, {
    "name" : "kuduConfigBean.defaultOperation",
    "value" : "DELETE"
  }, {
    "name" : "kuduConfigBean.changeLogFormat",
    "value" : "MySQLBinLog"
  }, {
    "name" : "kuduConfigBean.consistencyMode",
    "value" : "CLIENT_PROPAGATED"
  }, {
    "name" : "kuduConfigBean.mutationBufferSize",
    "value" : 100000
  }, {
    "name" : "kuduConfigBean.operationTimeout",
    "value" : 100000
  }, {
    "name" : "kuduConfigBean.unsupportedAction",
    "value" : "DISCARD"
  }, {
    "name" : "stageOnRecordError",
    "value" : "TO_ERROR"
  }, {
    "name" : "stageRequiredFields",
    "value" : [ ]
  }, {
    "name" : "stageRecordPreconditions",
    "value" : [ ]
  } ],
  "uiInfo" : {
    "description" : "",
    "label" : "Kudu-delete",
    "xPos" : 878,
    "yPos" : 1,
    "stageType" : "TARGET"
  },

```

```
"inputLanes" : [ "JavaScriptEvaluator_01OutputLane15302345785040" ],
"outputLanes" : [ ],
"eventLanes" : [ ],
"services" : [ ]
}, {
  "instanceName" : "Kudu_02",
  "library" : "streamsets-datacollector-apache-kudu_1_6-lib",
  "stageName" : "com_streamsets_pipeline_stage_destination_kudu_KuduDTarget",
  "stageVersion" : "4",
  "configuration" : [ {
    "name" : "kuduConfigBean.kuduMaster",
    "value" : "nn"
  }, {
    "name" : "kuduConfigBean.tableNameTemplate",
    "value" : "impala::default.${record:value('/Table')}"
  }, {
    "name" : "kuduConfigBean.fieldMappingConfigs",
    "value" : [ { } ]
  }, {
    "name" : "kuduConfigBean.defaultOperation",
    "value" : "UPSERT"
  }, {
    "name" : "kuduConfigBean.changeLogFormat",
    "value" : "MySQLBinLog"
  }, {
    "name" : "kuduConfigBean.consistencyMode",
    "value" : "CLIENT_PROPAGATED"
  }, {
    "name" : "kuduConfigBean.mutationBufferSize",
    "value" : 100000
  }, {
    "name" : "kuduConfigBean.operationTimeout",
    "value" : 100000
  }, {
    "name" : "kuduConfigBean.unsupportedAction",
    "value" : "DISCARD"
  }, {
    "name" : "stageOnRecordError",
    "value" : "TO_ERROR"
  }, {
    "name" : "stageRequiredFields",
    "value" : [ ]
  }, {
```

```

        "name" : "stageRecordPreconditions",
        "value" : [ ]
    } ],
    "uiInfo" : {
        "description" : "",
        "label" : "Kudu-upsert",
        "xPos" : 873,
        "yPos" : 157,
        "stageType" : "TARGET"
    },
    "inputLanes" : [ "JavaScriptEvaluator_02OutputLane15302346847350" ],
    "outputLanes" : [ ],
    "eventLanes" : [ ],
    "services" : [ ]
}, {
    "instanceName" : "Trash_01",
    "library" : "streamsets-datacollector-basic-lib",
    "stageName" : "com_streamsets_pipeline_stage_destination_devnull_NullIDTarget",
    "stageVersion" : "1",
    "configuration" : [ ],
    "uiInfo" : {
        "description" : "",
        "label" : "Trash 1",
        "xPos" : 387,
        "yPos" : 162,
        "stageType" : "TARGET"
    },
    "inputLanes" : [ "StreamSelector_01OutputLane1530179806488" ],
    "outputLanes" : [ ],
    "eventLanes" : [ ],
    "services" : [ ]
} ],
    "errorStage" : {
        "instanceName" : "WritetoFile_ErrorStage",
        "library" : "streamsets-datacollector-basic-lib",
        "stageName" :
"com_streamsets_pipeline_stage_destination_recordstolocalfilesystem_ToErrorLocalFSDTarget",
        "stageVersion" : "1",
        "configuration" : [ {
            "name" : "directory",
            "value" : "/var/log/streamsets"
        } ], {
            "name" : "uniquePrefix",

```

```

        "value" : "sdc-${sdc:id()}"
    }, {
        "name" : "rotationIntervalSecs",
        "value" : "${1 * HOURS}"
    }, {
        "name" : "maxFileSizeMbs",
        "value" : 512
    } ],
    "uiInfo" : {
        "description" : "",
        "label" : "Error Records - Write to File",
        "xPos" : 607,
        "yPos" : 50,
        "stageType" : "TARGET"
    },
    "inputLanes" : [ ],
    "outputLanes" : [ ],
    "eventLanes" : [ ],
    "services" : [ ]
},
"info" : {
    "pipelineId" : "mysqltokuducopyfe7bf9df-2520-4a09-b55f-5419fa6312f0",
    "title" : "mysqltokuducopy",
    "description" : "",
    "created" : 1530317756320,
    "lastModified" : 1530319376374,
    "creator" : "admin",
    "lastModifier" : "admin",
    "lastRev" : "0",
    "uuid" : "015b4cd5-5a02-4a8d-a601-bb24a92fdf8b",
    "valid" : true,
    "metadata" : {
        "labels" : [ ]
    },
    "name" : "mysqltokuducopyfe7bf9df-2520-4a09-b55f-5419fa6312f0",
    "sdcVersion" : "3.3.0",
    "sdcId" : "fa4b0b7a-7965-11e8-8490-d5c4b1792fd8"
},
"metadata" : {
    "labels" : [ ]
},
"statsAggregatorStage" : null,
"startEventStages" : [ {

```

```
    "instanceName" : "Discard_StartEventStage",
    "library" : "streamsets-datacollector-basic-lib",
    "stageName" : "com_streamsets_pipeline_stage_destination_devnull_ToErrorNullDTarget",
    "stageVersion" : "1",
    "configuration" : [ ],
    "uiInfo" : {
      "description" : "",
      "label" : "Start Event - Discard",
      "xPos" : 280,
      "yPos" : 50,
      "stageType" : "TARGET"
    },
    "inputLanes" : [ ],
    "outputLanes" : [ ],
    "eventLanes" : [ ],
    "services" : [ ]
  } ],
  "stopEventStages" : [ {
    "instanceName" : "Discard_StopEventStage",
    "library" : "streamsets-datacollector-basic-lib",
    "stageName" : "com_streamsets_pipeline_stage_destination_devnull_ToErrorNullDTarget",
    "stageVersion" : "1",
    "configuration" : [ ],
    "uiInfo" : {
      "description" : "",
      "label" : "Stop Event - Discard",
      "xPos" : 280,
      "yPos" : 50,
      "stageType" : "TARGET"
    },
    "inputLanes" : [ ],
    "outputLanes" : [ ],
    "eventLanes" : [ ],
    "services" : [ ]
  } ],
  "valid" : true,
  "issues" : {
    "stageIssues" : { },
    "pipelineIssues" : [ ],
    "issueCount" : 0
  },
  "previewable" : true
},
```

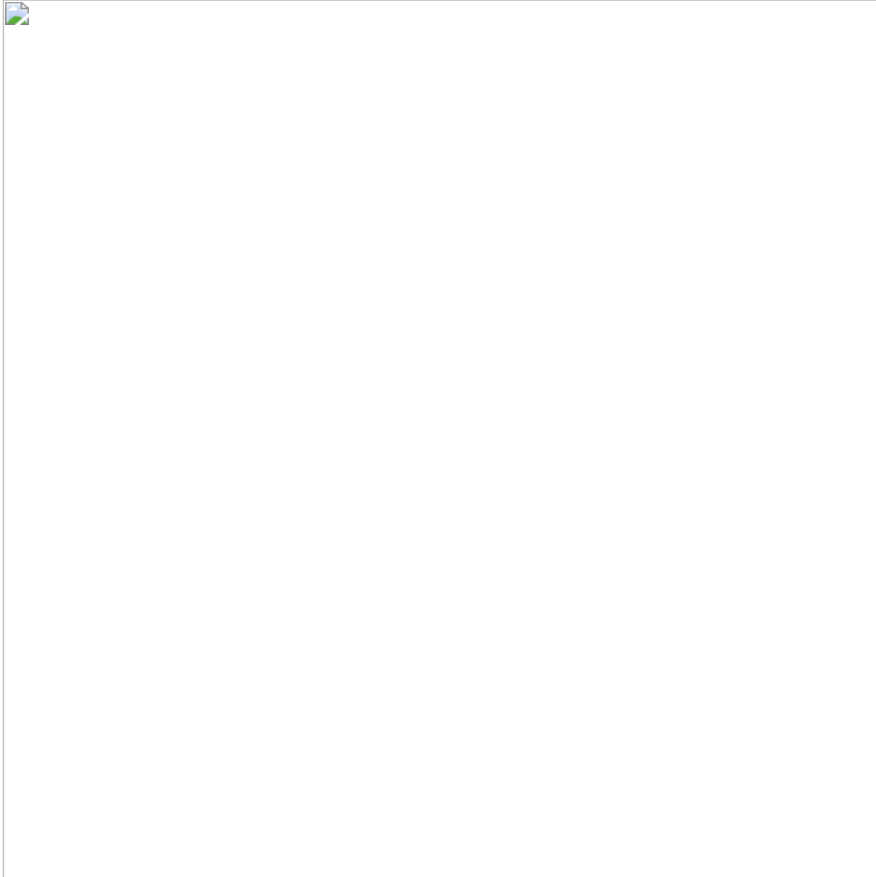
```
"pipelineRules" : {
  "schemaVersion" : 3,
  "version" : 2,
  "metricsRuleDefinitions" : [ {
    "id" : "badRecordsAlertID",
    "alertText" : "High incidence of Error Records",
    "metricId" : "pipeline.batchErrorRecords.counter",
    "metricType" : "COUNTER",
    "metricElement" : "COUNTER_COUNT",
    "condition" : "${value() > 100}",
    "sendEmail" : false,
    "enabled" : false,
    "timestamp" : 1530035416648,
    "valid" : true
  }, {
    "id" : "stageErrorAlertID",
    "alertText" : "High incidence of Stage Errors",
    "metricId" : "pipeline.batchErrorMessages.counter",
    "metricType" : "COUNTER",
    "metricElement" : "COUNTER_COUNT",
    "condition" : "${value() > 100}",
    "sendEmail" : false,
    "enabled" : false,
    "timestamp" : 1530035416648,
    "valid" : true
  }, {
    "id" : "idleGaugeID",
    "alertText" : "Pipeline is Idle",
    "metricId" : "RuntimeStatsGauge.gauge",
    "metricType" : "GAUGE",
    "metricElement" : "TIME_OF_LAST_RECEIVED_RECORD",
    "condition" : "${time:now() - value() > 120000}",
    "sendEmail" : false,
    "enabled" : false,
    "timestamp" : 1530035416648,
    "valid" : true
  }, {
    "id" : "batchTimeAlertID",
    "alertText" : "Batch taking more time to process",
    "metricId" : "RuntimeStatsGauge.gauge",
    "metricType" : "GAUGE",
    "metricElement" : "CURRENT_BATCH_AGE",
    "condition" : "${value() > 200}",
```

```

    "sendEmail" : false,
    "enabled" : false,
    "timestamp" : 1530035416648,
    "valid" : true
  }, {
    "id" : "memoryLimitAlertID",
    "alertText" : "Memory limit for pipeline exceeded",
    "metricId" : "pipeline.memoryConsumed.counter",
    "metricType" : "COUNTER",
    "metricElement" : "COUNTER_COUNT",
    "condition" : "${value() > (jvm:maxMemoryMB() * 0.65)}",
    "sendEmail" : false,
    "enabled" : false,
    "timestamp" : 1530035416648,
    "valid" : true
  } ],
  "dataRuleDefinitions" : [ ],
  "driftRuleDefinitions" : [ ],
  "uuid" : "f7ed1881-7695-479f-8963-4e9fb6335023",
  "configuration" : [ {
    "name" : "emailIDs",
    "value" : [ ]
  }, {
    "name" : "webhookConfigs",
    "value" : [ ]
  } ],
  "ruleIssues" : [ ],
  "configIssues" : [ ]
},
"libraryDefinitions" : null
}

```

效果:



mysql:

```
create database test;
create table cdc_test (
    id varchar(50),
    name varchar(200),
    PRIMARY KEY ( id )
);
insert into cdc_test select uuid(), 'aaa' from cdc_test;
```

impala:

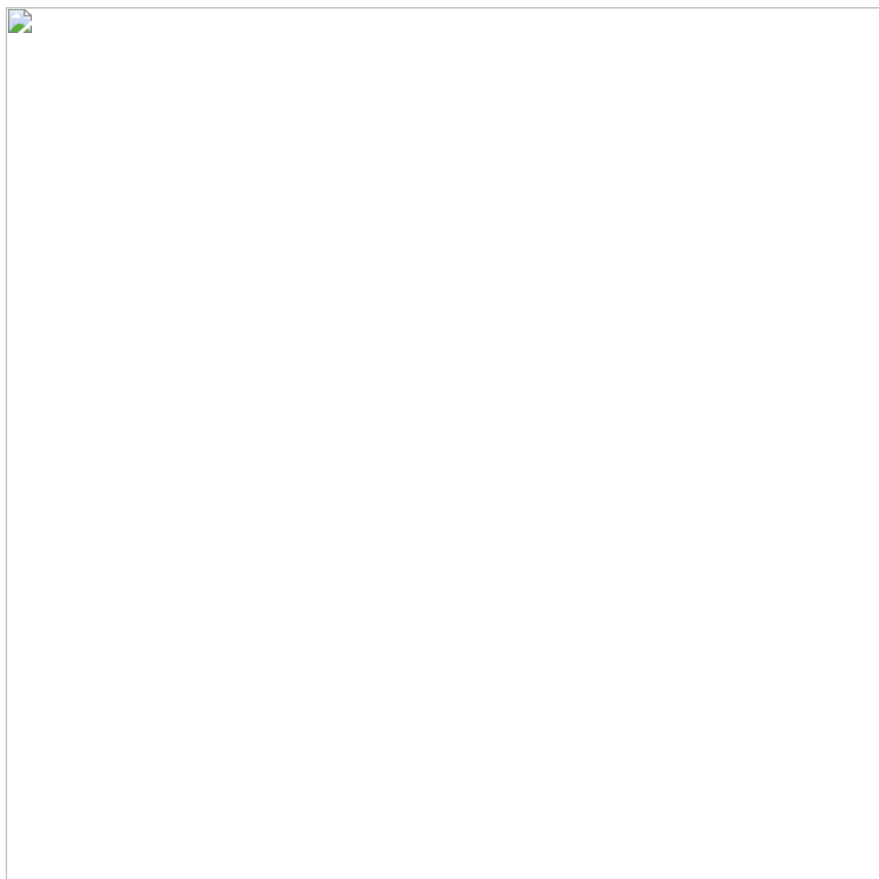
```
create table cdc_test (
    id string,
    name String,
    primary key(id)
)
PARTITION BY HASH PARTITIONS 16
STORED AS KUDU ;
```

性能测试:

插入100万


```
mysql> insert into cdc_test select uuid(), 'aaa' from cdc_test;
Query OK, 1048576 rows affected (14.73 sec)
Records: 1048576 Duplicates: 0 Warnings: 0
```

streamset 用的时间很长，而且大量报错



苦读参数设大：
插入几千万都没有报错了



参考:

<https://www.cloud.tencent.com/developer/article/1100860>