hive (niit)> create external table customerext( custno string, firstname string, lastname string, age int, profession string)

> row format delimited

> fields terminated by ','

> stored as textfile

> location '/user/hive/warehouse/niit.db/customer';

OK

Time taken: 0.03 seconds

hive (niit)> describe formatted customerext;

OK

col\_name data\_type comment

# col\_name data\_type comment

custno string

firstname string

lastname string

age int

profession string

# Detailed Table Information

Database: niit

Owner: hduser

CreateTime: Sat Dec 23 13:39:50 IST 2017

LastAccessTime: UNKNOWN

Protect Mode: None

Retention: 0

Location: hdfs://localhost:54310/user/hive/warehouse/niit.db/customer

Table Type: EXTERNAL\_TABLE

Table Parameters:

COLUMN\_STATS\_ACCURATE false

EXTERNAL TRUE

numFiles 0

numRows -1

rawDataSize -1

totalSize 0

transient\_lastDdlTime 1514016590

# Storage Information

SerDe Library: org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe

InputFormat: org.apache.hadoop.mapred.TextInputFormat

OutputFormat: org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat

Compressed: No

Num Buckets: -1

Bucket Columns: []

Sort Columns: []

Storage Desc Params:

field.delim ,

serialization.format ,

Time taken: 0.044 seconds, Fetched: 37 row(s)

hive (niit)> select count(\*) from customerext;

Query ID = hduser\_20171223134310\_91474869-5b4e-49de-8844-93e8363b6e7b

Total jobs = 1

Launching Job 1 out of 1

Number of reduce tasks determined at compile time: 1

In order to change the average load for a reducer (in bytes):

set hive.exec.reducers.bytes.per.reducer=<number>

In order to limit the maximum number of reducers:

set hive.exec.reducers.max=<number>

In order to set a constant number of reducers:

set mapreduce.job.reduces=<number>

Starting Job = job\_1514000082488\_0022, Tracking URL = http://rootuser:8088/proxy/application\_1514000082488\_0022/

Kill Command = /usr/local/hadoop/bin/hadoop job -kill job\_1514000082488\_0022

Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1

2017-12-23 13:43:13,773 Stage-1 map = 0%, reduce = 0%

2017-12-23 13:43:17,891 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.08 sec

2017-12-23 13:43:20,970 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 2.29 sec

MapReduce Total cumulative CPU time: 2 seconds 290 msec

Ended Job = job\_1514000082488\_0022

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 2.29 sec HDFS Read: 399159 HDFS Write: 6 SUCCESS

Total MapReduce CPU Time Spent: 2 seconds 290 msec

OK

\_c0

10000

Time taken: 12.695 seconds, Fetched: 1 row(s)

Loading Data From CSV File:

hive (niit)> CREATE TABLE csv\_table (

> student\_name string,

> subject string,

> marks INT)

> ROW FORMAT DELIMITED FIELDS TERMINATED BY ','

> STORED AS TEXTFILE;

OK

Time taken: 0.045 seconds

hive (niit)> LOAD DATA LOCAL INPATH "/home/hduser/student.csv" OVERWRITE INTO TABLE csv\_table;

Loading data to table niit.csv\_table

Table niit.csv\_table stats: [numFiles=1, numRows=0, totalSize=103, rawDataSize=0]

OK

Time taken: 0.133 seconds

hive (niit)> select \* from csv\_table;

OK

csv\_table.student\_name csv\_table.subject csv\_table.marks

Amit Maths 91

Amit Physics 48

Amit Chemistry 66

Sanjay Maths 96

Sanjay Physics 64

Sanjay Chemistry 73

NULL NULL

Time taken: 0.047 seconds, Fetched: 7 row(s)

hive (niit)> INSERT OVERWRITE TABLE avro\_table

> SELECT student\_name, subject, marks FROM csv\_table;

Query ID = hduser\_20171223182515\_1921eb79-8406-4dcf-bfde-c9f2ee0a0520

Total jobs = 3

Launching Job 1 out of 3

Number of reduce tasks is set to 0 since there's no reduce operator

Starting Job = job\_1514000082488\_0025, Tracking URL = http://rootuser:8088/proxy/application\_1514000082488\_0025/

Kill Command = /usr/local/hadoop/bin/hadoop job -kill job\_1514000082488\_0025

Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0

2017-12-23 18:25:18,542 Stage-1 map = 0%, reduce = 0%

2017-12-23 18:25:31,880 Stage-1 map = 100%, reduce = 0%

Ended Job = job\_1514000082488\_0025 with errors

Error during job, obtaining debugging information...

Job Tracking URL: http://rootuser:8088/proxy/application\_1514000082488\_0025/

Examining task ID: task\_1514000082488\_0025\_m\_000000 (and more) from job job\_1514000082488\_0025

Task with the most failures(4):

-----

Task ID:

task\_1514000082488\_0025\_m\_000000

URL:

http://0.0.0.0:8088/taskdetails.jsp?jobid=job\_1514000082488\_0025&tipid=task\_1514000082488\_0025\_m\_000000

-----

Diagnostic Messages for this Task:

Error: java.lang.RuntimeException: org.apache.hadoop.hive.ql.metadata.HiveException: Hive Runtime Error while processing row {"student\_name":"","subject":null,"marks":null}

at org.apache.hadoop.hive.ql.exec.mr.ExecMapper.map(ExecMapper.java:172)

at org.apache.hadoop.mapred.MapRunner.run(MapRunner.java:54)

at org.apache.hadoop.mapred.MapTask.runOldMapper(MapTask.java:450)

at org.apache.hadoop.mapred.MapTask.run(MapTask.java:343)

at org.apache.hadoop.mapred.YarnChild$2.run(YarnChild.java:163)

at java.security.AccessController.doPrivileged(Native Method)

at javax.security.auth.Subject.doAs(Subject.java:422)

at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1628)

at org.apache.hadoop.mapred.YarnChild.main(YarnChild.java:158)

Caused by: org.apache.hadoop.hive.ql.metadata.HiveException: Hive Runtime Error while processing row {"student\_name":"","subject":null,"marks":null}

at org.apache.hadoop.hive.ql.exec.MapOperator.process(MapOperator.java:518)

at org.apache.hadoop.hive.ql.exec.mr.ExecMapper.map(ExecMapper.java:163)

... 8 more

Caused by: org.apache.hadoop.hive.ql.metadata.HiveException: Avro could not validate record against schema (record = {"student\_name": "", "subject": null, "marks": null}) (schema = {"type":"record","name":"student\_marks","namespace":"abc","fields":[{"name":"student\_name","type":"string"},{"name":"subject","type":"string"},{"name":"marks","type":"int"}]})

at org.apache.hadoop.hive.ql.exec.FileSinkOperator.process(FileSinkOperator.java:789)

at org.apache.hadoop.hive.ql.exec.Operator.forward(Operator.java:837)

at org.apache.hadoop.hive.ql.exec.SelectOperator.process(SelectOperator.java:88)

at org.apache.hadoop.hive.ql.exec.Operator.forward(Operator.java:837)

at org.apache.hadoop.hive.ql.exec.TableScanOperator.process(TableScanOperator.java:97)

at org.apache.hadoop.hive.ql.exec.MapOperator$MapOpCtx.forward(MapOperator.java:162)

at org.apache.hadoop.hive.ql.exec.MapOperator.process(MapOperator.java:508)

... 9 more

Caused by: Avro could not validate record against schema (record = {"student\_name": "", "subject": null, "marks": null}) (schema = {"type":"record","name":"student\_marks","namespace":"abc","fields":[{"name":"student\_name","type":"string"},{"name":"subject","type":"string"},{"name":"marks","type":"int"}]})

at org.apache.hadoop.hive.serde2.avro.AvroSerializer.serialize(AvroSerializer.java:102)

at org.apache.hadoop.hive.serde2.avro.AvroSerDe.serialize(AvroSerDe.java:193)

at org.apache.hadoop.hive.ql.exec.FileSinkOperator.process(FileSinkOperator.java:720)

... 15 more

FAILED: Execution Error, return code 2 from org.apache.hadoop.hive.ql.exec.mr.MapRedTask

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 HDFS Read: 0 HDFS Write: 0 FAIL

Total MapReduce CPU Time Spent: 0 msec

hive (niit)> CREATE VIEW afe\_45plus AS select \* from customer where age>45;

OK

custno firstname lastname age profession

Time taken: 0.054 seconds

hive (niit)> show tables;

OK

tab\_name

afe\_45plus

avro\_table

csv\_table

customer

customerext

nyse

txnrecords

Time taken: 0.022 seconds, Fetched: 7 row(s)

It will show afe\_45plus in showtable command but not in HDFS physically.

hive (niit)> describe formatted afe\_45plus;

OK

col\_name data\_type comment

# col\_name data\_type comment

custno int

firstname string

lastname string

age int

profession string

# Detailed Table Information

Database: niit

Owner: hduser

CreateTime: Sat Dec 23 18:30:16 IST 2017

LastAccessTime: UNKNOWN

Protect Mode: None

Retention: 0

Table Type: VIRTUAL\_VIEW

Table Parameters:

transient\_lastDdlTime 1514034016

# Storage Information

SerDe Library: null

InputFormat: org.apache.hadoop.mapred.SequenceFileInputFormat

OutputFormat: org.apache.hadoop.hive.ql.io.HiveSequenceFileOutputFormat

Compressed: No

Num Buckets: -1

Bucket Columns: []

Sort Columns: []

# View Information

View Original Text: select \* from customer where age>45

View Expanded Text: select `customer`.`custno`, `customer`.`firstname`, `customer`.`lastname`, `customer`.`age`, `customer`.`profession` from `niit`.`customer` where `customer`.`age`>45

Time taken: 0.037 seconds, Fetched: 31 row(s)