

Error coming due to type of delimiter used in PigStorage()

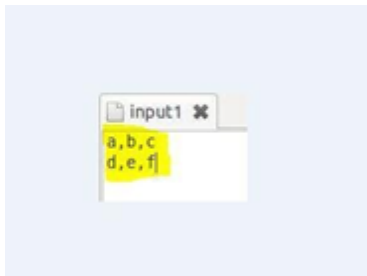
Error scenario: Executing the load command without using built function of pig i.e PigStorage()

When we load a dataset delimited by anything other than tab and we do not specify the built function PigStorage() at the time of loading the relation, we do not get the correct output.

If the dataset is **tab delimited** then we can load the dataset without **PigStorage()** too.

Command: `Y1 = load '/input1' AS(A:chararray, B:chararray, C:chararray);`

The below dataset is comma delimited but we have not specified the built in function PigStorage() in the above command



Note: There is no error at the time of loading but we get the error at the time of the dumping as shown below.

```
Details at logfile: /home/cloudera/pig_1407585651582.log
grunt> Y1 = load '/input1' AS(A:chararray,B:chararray,C:chararray);
grunt>
```

```
2014-08-09 05:05:44,281 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Encountered Warning ACCESSING_NON_EXISTENT_FIELD 4 time(s).
2014-08-09 05:05:44,281 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2014-08-09 05:05:44,302 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2014-08-09 05:05:44,302 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(a,b,c,,)
(d,e,f,,)
grunt>
```

Resolution: We specify the type of delimiter applied in input file using PigStorage function to remove the distortion in the output as shown in the below command.

Correct command: Y1 = load '/input1' using PigStorage(',') AS(A:chararray,B:chararray,C:chararray);

```
cloudera@cloudera-vm:~$ pig
2014-08-09 05:28:30,572 [main] INFO org.apache.pig.Main - Logging error message
s to: /home/cloudera/pig_1407587310570.log
2014-08-09 05:28:30,761 [main] INFO org.apache.pig.backend.hadoop.executionengi
ne.HExecutionEngine - Connecting to hadoop file system at: hdfs://localhost:8020
2014-08-09 05:28:30,942 [main] INFO org.apache.pig.backend.hadoop.executionengi
ne.HExecutionEngine - Connecting to map-reduce job tracker at: localhost:8021
grunt> Y1 = load '/input1' using PigStorage(',') AS(A:chararray,B:chararray,C:ch
ararray);
grunt>
```

On dumping the relation Y1 we get the correct output as shown below:

```
grunt> Y1 = load '/input1' using PigStorage(',') AS(A:chararray,B:chararray,C:ch
ararray);
grunt> dump Y1;
2014-08-09 05:29:43,098 [main] INFO org.apache.pig.tools.pigstats.ScriptState -
Pig features used in the script: UNKNOWN
```

```
2014-08-09 05:29:51,232 [main] INFO org.apache.pig.backend.hadoop.executionengi
ne.mapReduceLayer.MapReduceLauncher - Success!
2014-08-09 05:29:51,242 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileI
nputFormat - Total input paths to process : 1
2014-08-09 05:29:51,242 [main] INFO org.apache.pig.backend.hadoop.executionengi
ne.util.MapRedUtil - Total input paths to process : 1
(a,b,c)
(d,e,f)
grunt>
```