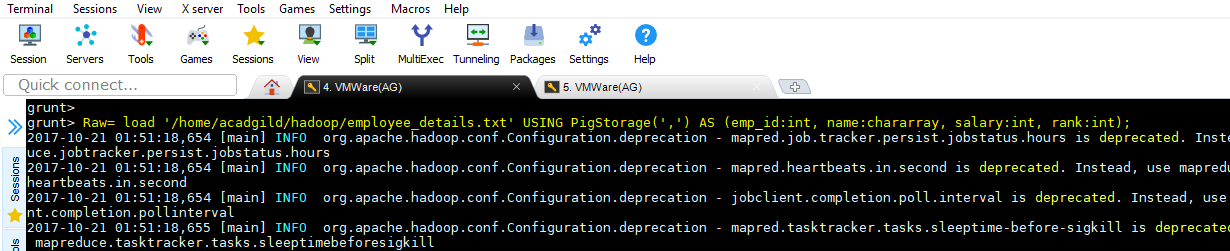
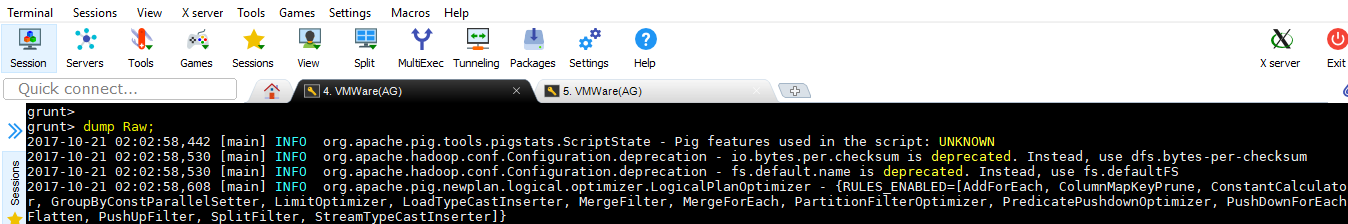
**Tokenize**- Tokenize splits a string(group of words) into a bag of words.

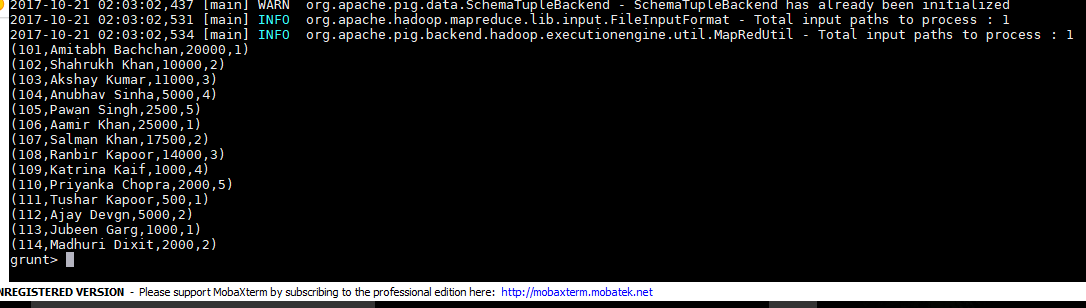
Load the file using below relation-

Raw= load '/home/acadgild/hadoop/employee\_details.txt' USING PigStorage(',') AS (emp\_id:int, name:chararray, salary:int, rank:int)



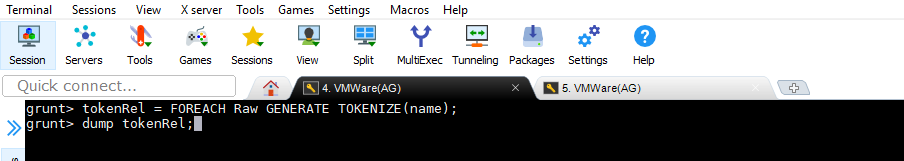
Dumping records-



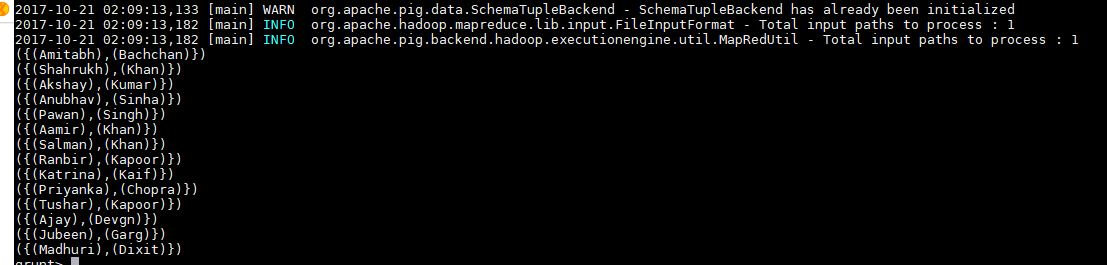


Use below relation to tokenize name field which will generate bag of names-

tokenRel = FOREACH Raw GENERATE TOKENIZE(name);



O/P-



The above output contains name being displayed separately in tuples but in same bag.