CIS-5200: Hive N-gram Sentiment Data Analysis in Oracle

Submitted by: Nanjesh Ramesh

CIN: 307402152

```
| {"ngram":["the","product"],"estfrequency":1704.0}  |
| {"ngram":["was","average"],"estfrequency":1375.0}  |
| {"ngram":["item","was"],"estfrequency":1275.0}     |
| {"ngram":["this","is"],"estfrequency":1057.0}      |
| {"ngram":["of","product"],"estfrequency":1011.0}   |
| {"ngram":["value","of"],"estfrequency":1011.0}     |
| {"ngram":["quality","was"],"estfrequency":1009.0}  |
+----------------------------------------------------+--+
10 rows selected (1.315 seconds)
0: jdbc:hive2://sparksqoop-bdcsce-3.compute-6> SELECT EXPLODE(context_ngrams(sentences(LOWER(message)), array(null, null), 10)) AS bigram FROM ratings;
INFO  : Session is already open
INFO  : Dag name: SELECT EXPLODE(context_ngrams(sent...ratings(Stage-1)
INFO  :

INFO  : Status: Running (Executing on YARN cluster with App id application_1553054135127_0008)

INFO  : Map 1: 0/1      Reducer 2: 0/1
INFO  : Map 1: 0(+1)/1  Reducer 2: 0/1
INFO  : Map 1: 1/1      Reducer 2: 0(+1)/1
INFO  : Map 1: 1/1      Reducer 2: 1/1
+----------------------------------------------------+--+
|                      bigram                         |
+----------------------------------------------------+--+
| {"ngram":["product","was"],"estfrequency":2772.0}  |
| {"ngram":["was","ok"],"estfrequency":2641.0}       |
| {"ngram":["it","was"],"estfrequency":2127.0}       |
| {"ngram":["the","product"],"estfrequency":1704.0}  |
| {"ngram":["was","average"],"estfrequency":1375.0}  |
| {"ngram":["item","was"],"estfrequency":1275.0}     |
| {"ngram":["this","is"],"estfrequency":1057.0}      |
| {"ngram":["of","product"],"estfrequency":1011.0}   |
| {"ngram":["value","of"],"estfrequency":1011.0}     |
| {"ngram":["quality","was"],"estfrequency":1009.0}  |
+----------------------------------------------------+--+
10 rows selected (1.177 seconds)
0: jdbc:hive2://sparksqoop-bdcsce-3.compute-6> SELECT EXPLODE(context_ngrams(sentences(lower(message)), array("red", "one", null, null), 4, 10)) AS snippet FROM ratings;
INFO  : Session is already open
INFO  : Dag name: SELECT EXPLODE(context_ngrams(sent...ratings(Stage-1)
INFO  :

INFO  : Status: Running (Executing on YARN cluster with App id application_1553054135127_0008)

INFO  : Map 1: 0/1      Reducer 2: 0/1
INFO  : Map 1: 0(+1)/1  Reducer 2: 0/1
INFO  : Map 1: 1/1      Reducer 2: 0/1
INFO  : Map 1: 1/1      Reducer 2: 0(+1)/1
INFO  : Map 1: 1/1      Reducer 2: 1/1
+------------------------------------------------+--+
|                    snippet                      |
+------------------------------------------------+--+
| {"ngram":["cost","ten"],"estfrequency":71.0}   |
+------------------------------------------------+--+
1 row selected (6.691 seconds)
0: jdbc:hive2://sparksqoop-bdcsce-3.compute-6>
```