

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

Submitted by: Nanjesh Ramesh
CIN: 307402152

```
INFO : Status: Running (Executing on YARN cluster with App id application_1553054135127_0008)
INFO : Map 1: -/- Reducer 2: 0/1
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
+-----+
| message |
+-----+
| What is so special about red? |
| Why does the red one cost ten times more than the others? |
+-----+
2 rows selected (15.903 seconds)
0 jdbc:hive2://sparksoop-bdcsce-3.compute-6> SELECT *
0 jdbc:hive2://sparksoop-bdcsce-3.compute-6> FROM products
0 jdbc:hive2://sparksoop-bdcsce-3.compute-6> WHERE name LIKE '%16 GB USB Flash Drive%' AND brand="Orion";
+-----+
| products.prod_id | products.brand | products.name | products.price | products.cost | products.shipping_wt |
+-----+
| 1274672 | "Orion" | "16 GB USB Flash Drive (Red)" | 4299 | 4001 | 1 |
| 1274674 | "Orion" | "16 GB USB Flash Drive (Green)" | 4299 | 4001 | 1 |
| 1274675 | "Orion" | "16 GB USB Flash Drive (Blue)" | 4299 | 4001 | 1 |
+-----+
3 rows selected (0.093 seconds)
0 jdbc:hive2://sparksoop-bdcsce-3.compute-6> SELECT context_ngrams(sentences(LOWER(message)), array(null, null), 10) AS onegram FROM ratings;
INFO : Session is already open
INFO : Dag name: SELECT context_ngrams(sentences(LO...ratings(Stage-1)
INFO :
INFO : Status: Running (Executing on YARN cluster with App id application_1553054135127_0008)
INFO : Map 1: -/- Reducer 2: 0/1
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
+-----+
| onegram |
+-----+
| [{"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}] |
+-----+
1 row selected (6.744 seconds)
```

```
-- ssh nmandya@129.150.66.211 ... -- ssh nmandya@129.150.66.211
+-----+
| onegram |
+-----+
| [{"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}] |
+-----+
1 row selected (6.744 seconds)
0 jdbc:hive2://sparksoop-bdcsce-3.compute-6> SELECT EXPLODE(context_ngrams(sentences(LOWER(message)), array(null, null), 10)) AS onegram 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: 1/1
+-----+
| onegram |
+-----+
| [{"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 (1.315 seconds)
0 jdbc:hive2://sparksoop-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
```

```

-- ssh nmandya@129.150.66.211
{"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://sparksgoop-bdcsce-3.compute-6> SELECT EXPLD(context_ngrams(sentences(LOWER(message)), array(null, null), 10)) AS bigram FROM ratings;
INFO : Session is already open
INFO : Dag name: SELECT EXPLD(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://sparksgoop-bdcsce-3.compute-6> SELECT EXPLD(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 EXPLD(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://sparksgoop-bdcsce-3.compute-6>

```