Swinburne University of Technology Hawthorn Campus
Department of Computing Technologies

COS20028 Big Data Architecture and Application

Assignment 2 - *Semester 2, 2023*

Name: Nguyen Nam Tung
Student ID: 103181157
Submission Date: 26/10/2023

Assignment Task

1. Find a way with tools taught in this unit to check whether the language_code attribute in the source data is unique. **Note that you cannot eyeball the result. The result should be found by combining valid tools in a sequence.**
   1.1. List the sequence of tools you used for finding the answer:
   Ans: Hive can be utilized to find the answers.

   1.2. Give a short explanation of which tool did what.
   Ans: Hive counts the number of all the language_code attribute.
   If the count is greater than 1 then it is not unique.

   1.3. List the code/command/statement and the outcome screenshot of the step in the sequence: (Note that only meaningful screenshot is required. For example, the screenshot of the MapReduce execution result is necessary, but the screenshot of the outcome from the "cd .." command is not meaningful

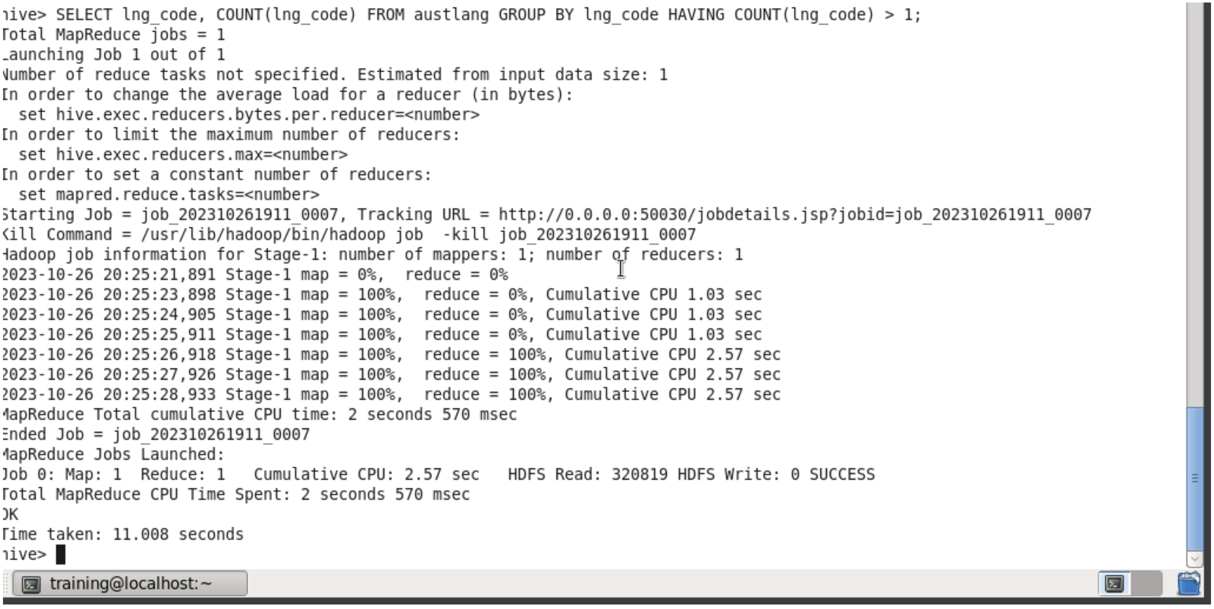
[Phase 1 – Querying to check that the language_code attribute is unique]

Query:

SELECT lng_code, COUNT(lng_code) FROM austlang GROUP BY lng_code HAVING COUNT(lng_code) > 1;

Outcome:

No Data Available for this query so that language code is unique



```
ıive> SELECT lng_code, COUNT(lng_code) FROM austlang GROUP BY lng_code HAVING COUNT(lng_code) > 1;
Total MapReduce jobs = 1
-aunching Job 1 out of 1
\umber of reduce tasks not specified. Estimated from input data size: 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 mapred.reduce.tasks=<number>
5tarting Job = job_202310261911_0007, Tracking URL = http://0.0.0.0:50030/jobdetails.jsp?jobid=job_202310261911_0007
<ill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_202310261911_0007
ladoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-10-26 20:25:21,891 Stage-1 map = 0%,  reduce = 0%
2023-10-26 20:25:23,898 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.03 sec
2023-10-26 20:25:24,905 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.03 sec
2023-10-26 20:25:25,911 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.03 sec
2023-10-26 20:25:26,918 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 2.57 sec
2023-10-26 20:25:27,926 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 2.57 sec
2023-10-26 20:25:28,933 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 2.57 sec
1apReduce Total cumulative CPU time: 2 seconds 570 msec
Ended Job = job_202310261911_0007
1apReduce Jobs Launched:
Job 0: Map: 1  Reduce: 1   Cumulative CPU: 2.57 sec   HDFS Read: 320819 HDFS Write: 0 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 570 msec
OK
Time taken: 11.008 seconds
ıive> 
```

2. Find a way with tools taught in this unit to list the unique values for all entities (tables with the solid boundary) except lng_id.

2.1. What is the best tool of choice to perform this task?

Ans: MapReduce is utilized to perform this task.

2.2. Assume you choose MapReduce to be the tool. Should it be the Map-only design or the complete MapReduce with mapper and reducer involved in the process?

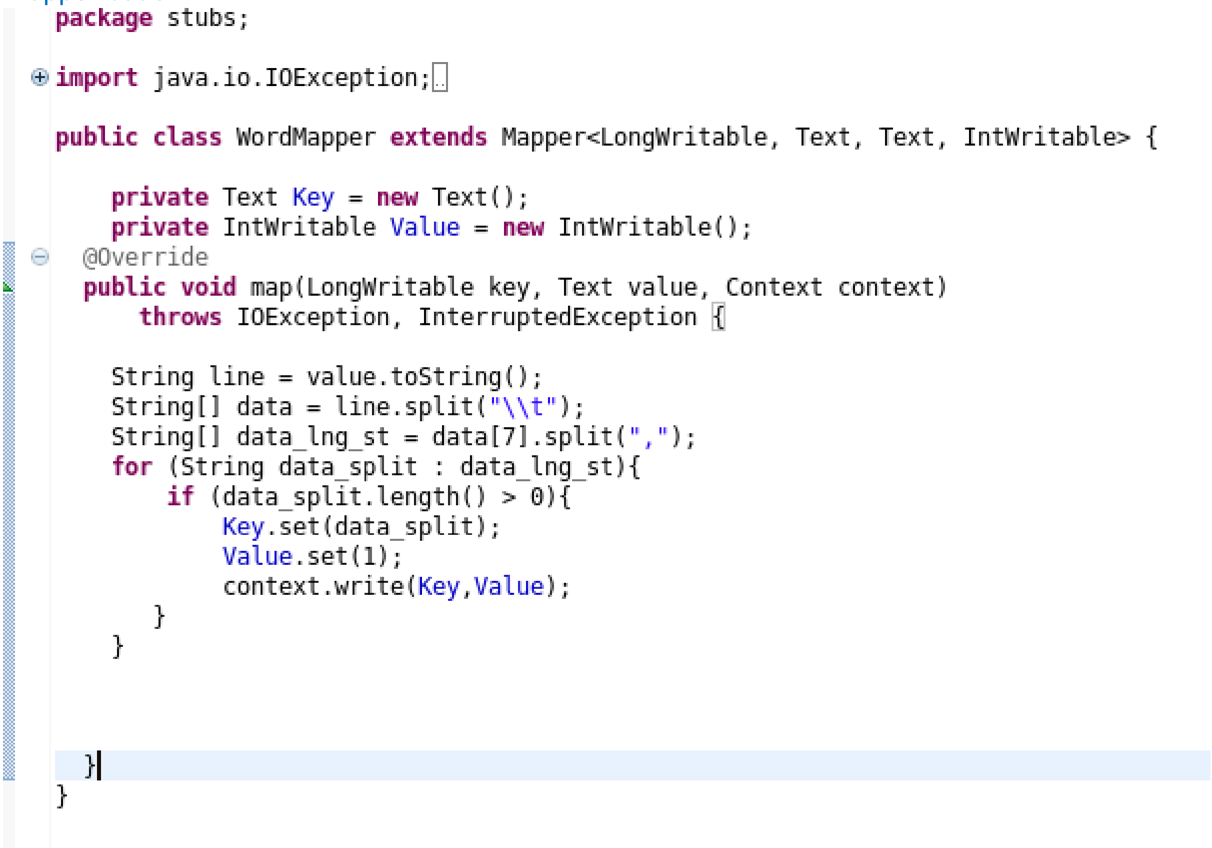
Ans: No, it should not be the Map – only design. It should have both Mapper and Reducer as we need a MapReduce design to aggerate the data to find the unique values.

2.3. Assume you choose MapReduce to be the tool. Does the count in the output matter for preparing the data for creating the entities?

Ans: No because we only want to list the unique values for all the entities. But for example, in question 2.5, the count in the output might be mattered to find the counts of a particular state.

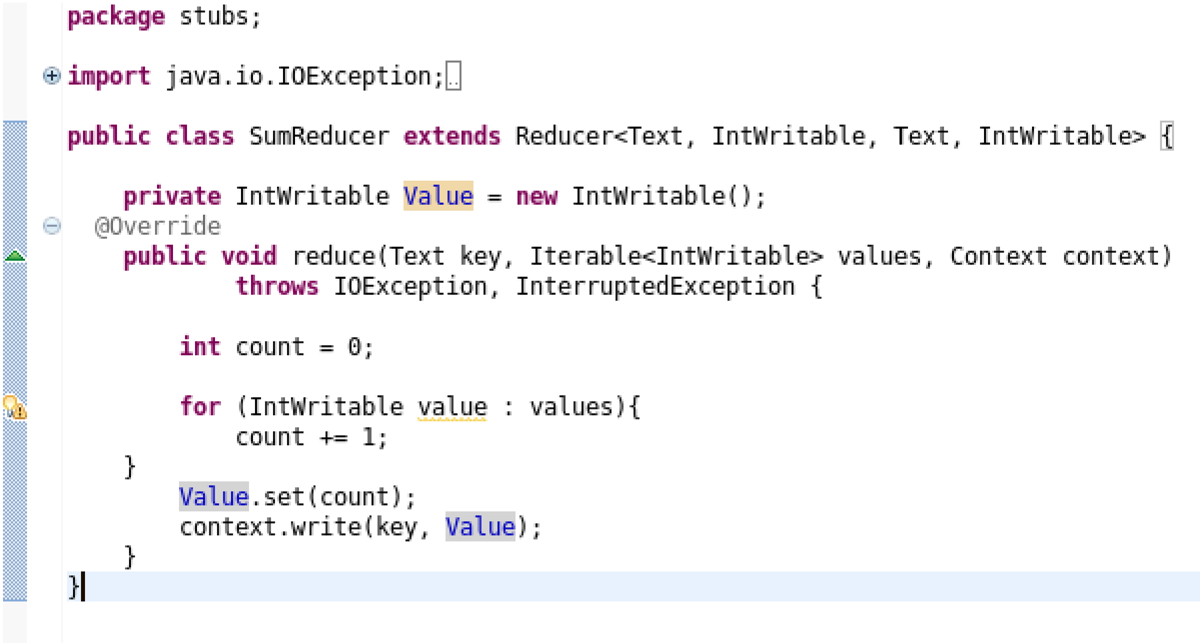
2.4. List the code/command/statement and the outcome screenshot of preparing the data for lng_st.

Mapper Code:



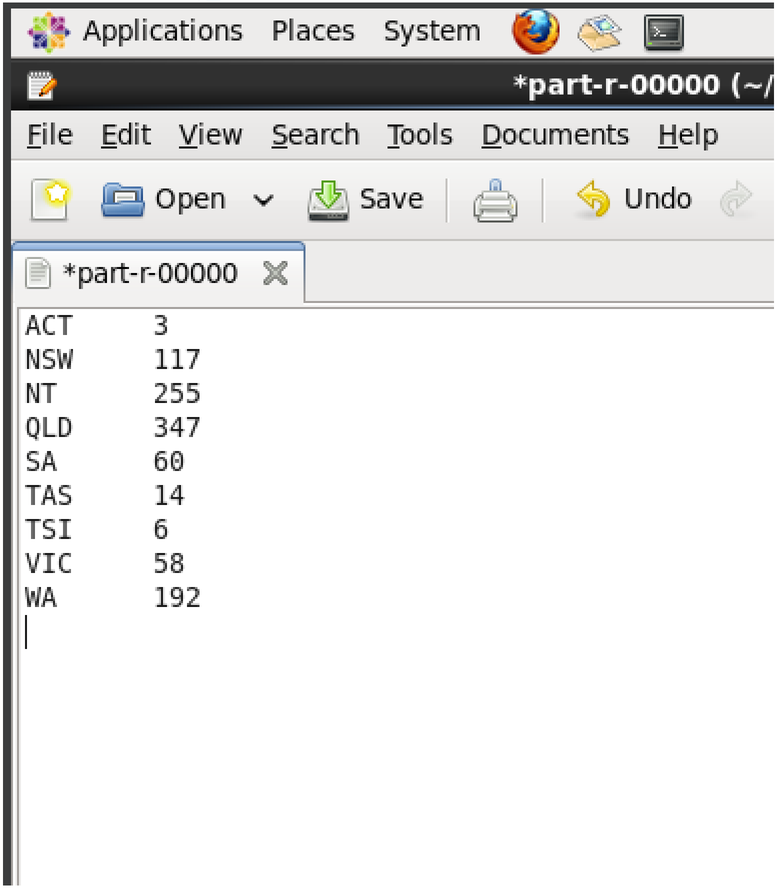
```
package stubs;

import java.io.IOException;

public class WordMapper extends Mapper<LongWritable, Text, Text, IntWritable> {

    private Text Key = new Text();
    private IntWritable Value = new IntWritable();
  @Override
  public void map(LongWritable key, Text value, Context context)
      throws IOException, InterruptedException {

    String line = value.toString();
    String[] data = line.split("\\t");
    String[] data_lng_st = data[7].split(",");
    for (String data_split : data_lng_st){
        if (data_split.length() > 0){
            Key.set(data_split);
            Value.set(1);
            context.write(Key,Value);
        }
    }



  }
}
```

Reducer Code:



```java
package stubs;

import java.io.IOException;

public class SumReducer extends Reducer<Text, IntWritable, Text, IntWritable> {

    private IntWritable Value = new IntWritable();
  @Override
    public void reduce(Text key, Iterable<IntWritable> values, Context context)
            throws IOException, InterruptedException {

        int count = 0;

        for (IntWritable value : values){
            count += 1;
    }
        Value.set(count);
        context.write(key, Value);
    }
}
```

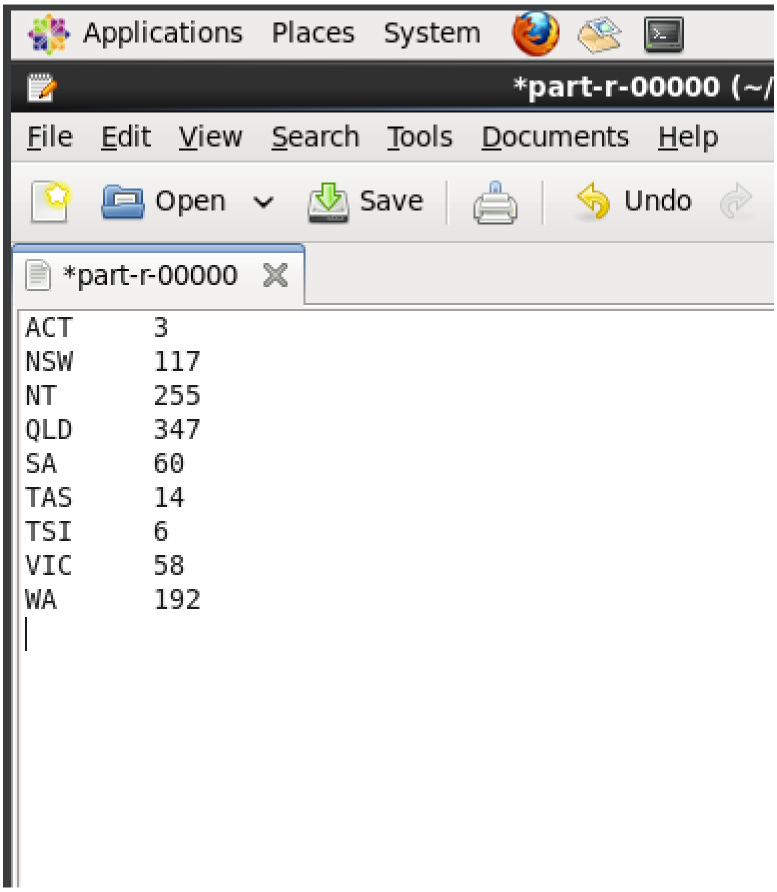
Result Set for lng_st:



```
ACT     3
NSW     117
NT      255
QLD     347
SA      60
TAS     14
TSI     6
VIC     58
WA      192
```

2.5. How many counts does QLD have?

We can check the output file for the desired answer.



```
ACT	3
NSW	117
NT	255
QLD	347
SA	60
TAS	14
TSI	6
VIC	58
WA	192
```

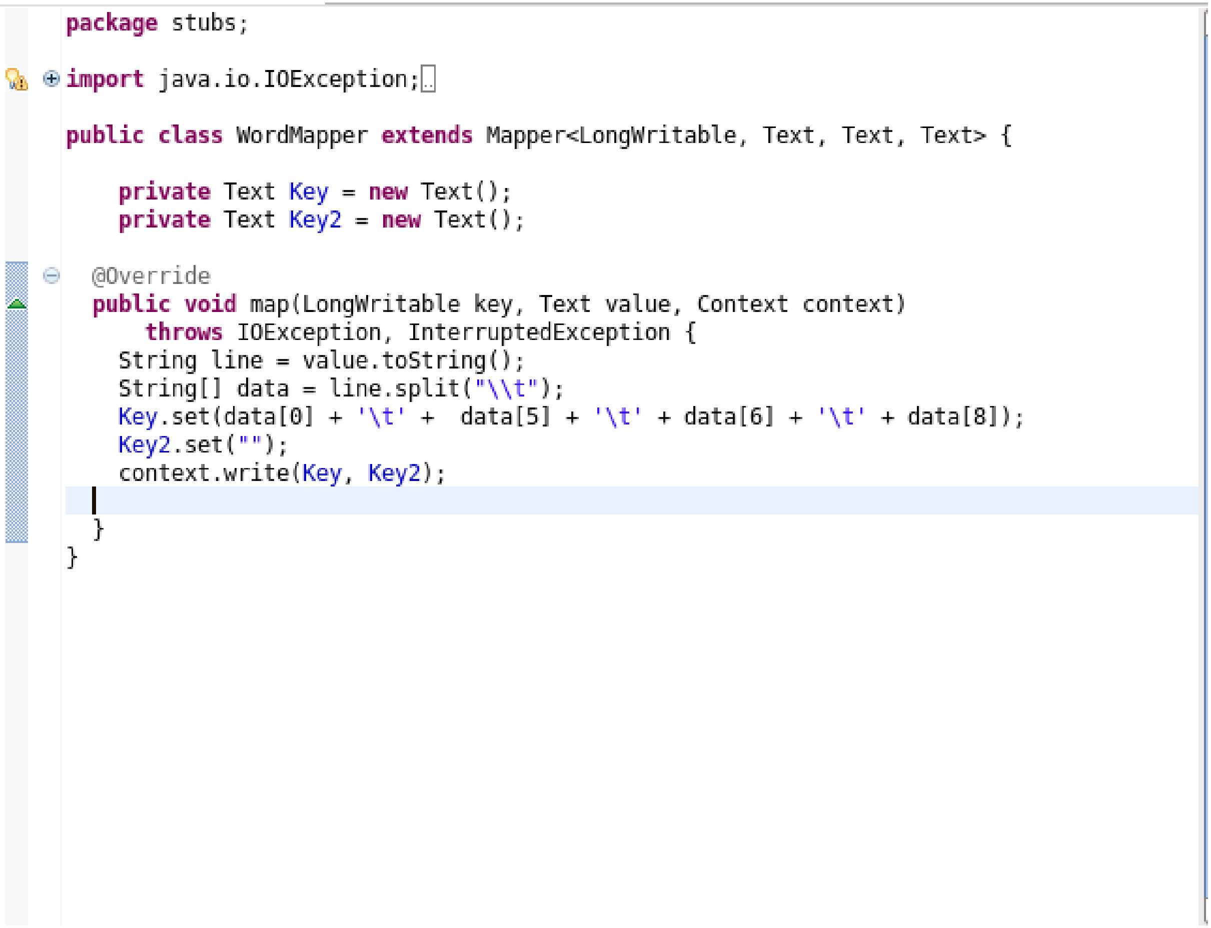
Ans: 347

3. Find a way with tools taught in this unit to prepare the data for all entities (tables with the solid boundary) for MySQL and Hive.
   3.1. Assume your tool of choice is MapReduce for this task. Will this be a Map-only job or a complete MapReduce job for preparing the desired data?

   Ans: In my opinion, it should be a full MapReduce job. This is because, we need to aggregate the data to make sure that a unique key is generated for each record. However, on the other hand, a map-only job may be enough to get the records from particular fields from the original dataset.

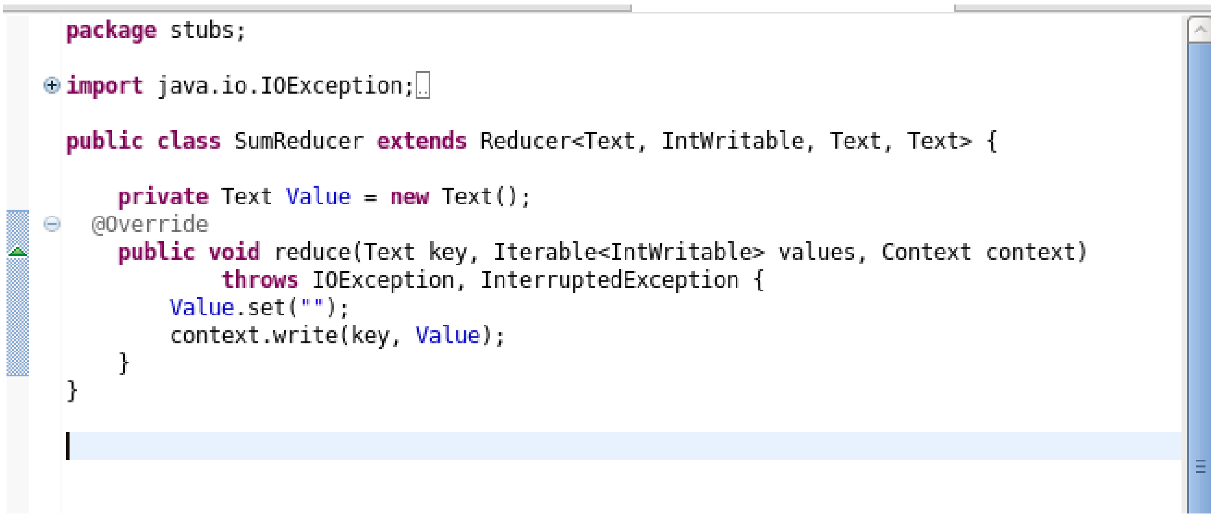
   3.2. List the code/command/statement and the outcome screenshot of preparing the data for lng_id.

Mapper Code:



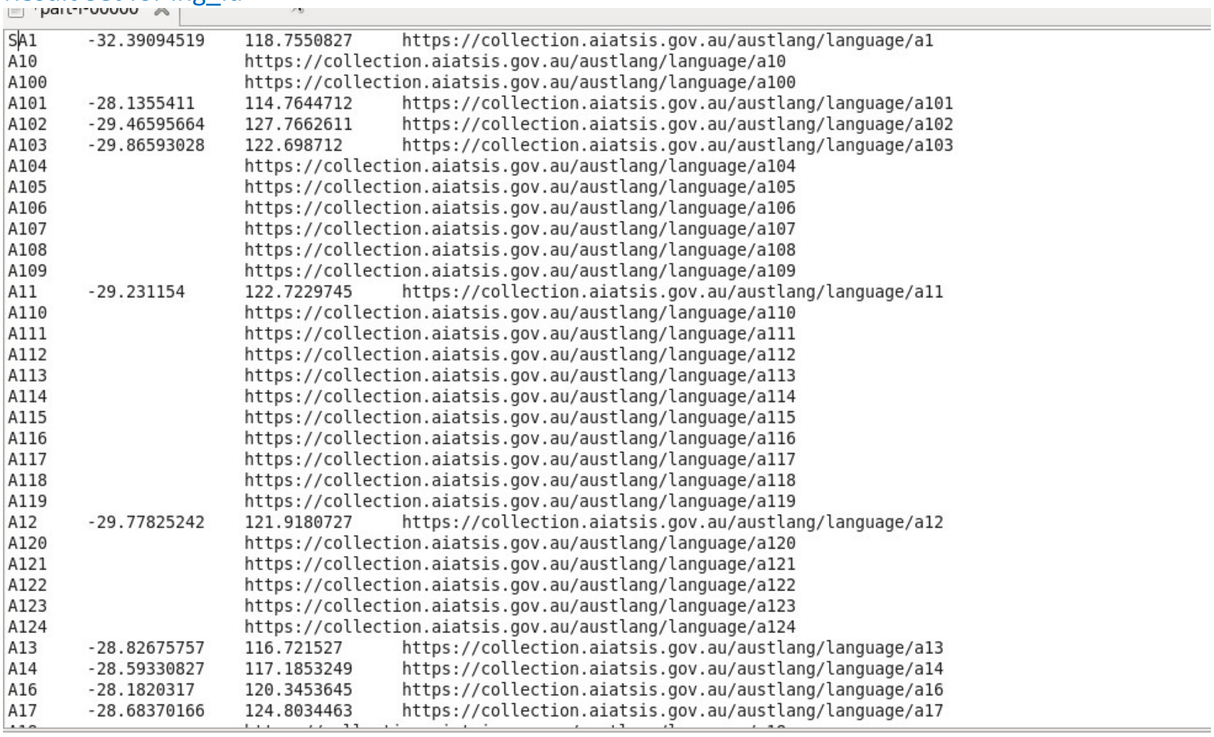
```
package stubs;

import java.io.IOException;

public class WordMapper extends Mapper<LongWritable, Text, Text, Text> {

    private Text Key = new Text();
    private Text Key2 = new Text();

  @Override
  public void map(LongWritable key, Text value, Context context)
      throws IOException, InterruptedException {
    String line = value.toString();
    String[] data = line.split("\\t");
    Key.set(data[0] + '\t' +  data[5] + '\t' + data[6] + '\t' + data[8]);
    Key2.set("");
    context.write(Key, Key2);

  }
}
```

Reducer Code:



```
package stubs;

import java.io.IOException;

public class SumReducer extends Reducer<Text, IntWritable, Text, Text> {

    private Text Value = new Text();
  @Override
    public void reduce(Text key, Iterable<IntWritable> values, Context context)
            throws IOException, InterruptedException {
        Value.set("");
        context.write(key, Value);
    }
}
```

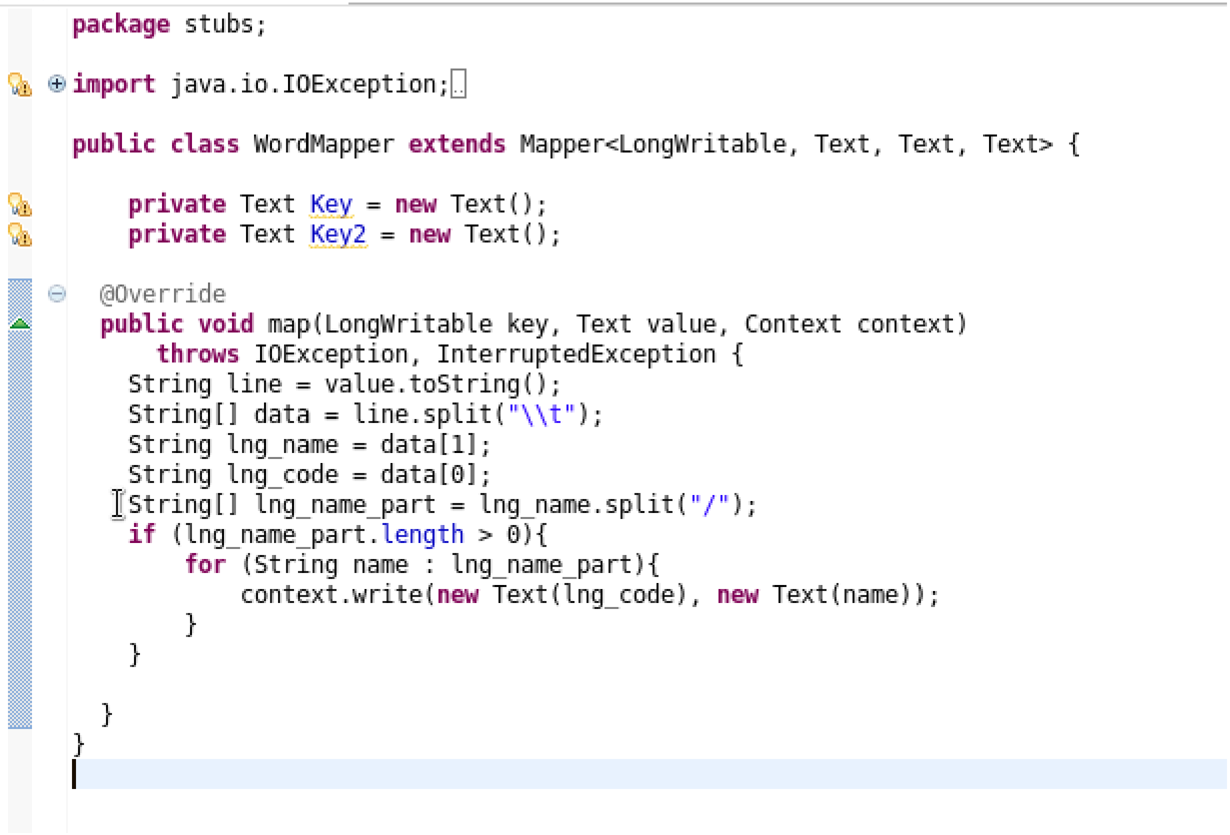
Result Set for lng_id



```
SA1     -32.39094519    118.7550827     https://collection.aiatsis.gov.au/austlang/language/a1
A10                     https://collection.aiatsis.gov.au/austlang/language/a10
A100                    https://collection.aiatsis.gov.au/austlang/language/a100
A101    -28.1355411     114.7644712     https://collection.aiatsis.gov.au/austlang/language/a101
A102    -29.46595664    127.7662611     https://collection.aiatsis.gov.au/austlang/language/a102
A103    -29.86593028    122.698712      https://collection.aiatsis.gov.au/austlang/language/a103
A104                    https://collection.aiatsis.gov.au/austlang/language/a104
A105                    https://collection.aiatsis.gov.au/austlang/language/a105
A106                    https://collection.aiatsis.gov.au/austlang/language/a106
A107                    https://collection.aiatsis.gov.au/austlang/language/a107
A108                    https://collection.aiatsis.gov.au/austlang/language/a108
A109                    https://collection.aiatsis.gov.au/austlang/language/a109
A11     -29.231154      122.7229745     https://collection.aiatsis.gov.au/austlang/language/a11
A110                    https://collection.aiatsis.gov.au/austlang/language/a110
A111                    https://collection.aiatsis.gov.au/austlang/language/a111
A112                    https://collection.aiatsis.gov.au/austlang/language/a112
A113                    https://collection.aiatsis.gov.au/austlang/language/a113
A114                    https://collection.aiatsis.gov.au/austlang/language/a114
A115                    https://collection.aiatsis.gov.au/austlang/language/a115
A116                    https://collection.aiatsis.gov.au/austlang/language/a116
A117                    https://collection.aiatsis.gov.au/austlang/language/a117
A118                    https://collection.aiatsis.gov.au/austlang/language/a118
A119                    https://collection.aiatsis.gov.au/austlang/language/a119
A12     -29.77825242    121.9180727     https://collection.aiatsis.gov.au/austlang/language/a12
A120                    https://collection.aiatsis.gov.au/austlang/language/a120
A121                    https://collection.aiatsis.gov.au/austlang/language/a121
A122                    https://collection.aiatsis.gov.au/austlang/language/a122
A123                    https://collection.aiatsis.gov.au/austlang/language/a123
A124                    https://collection.aiatsis.gov.au/austlang/language/a124
A13     -28.82675757    116.721527      https://collection.aiatsis.gov.au/austlang/language/a13
A14     -28.59330827    117.1853249     https://collection.aiatsis.gov.au/austlang/language/a14
A16     -28.1820317     120.3453645     https://collection.aiatsis.gov.au/austlang/language/a16
A17     -28.68370166    124.8034463     https://collection.aiatsis.gov.au/austlang/language/a17
```

4. Find a way with tools taught in this unit to prepare the data for all weak entities (tables with the dashed boundary) for MySQL and Hive.
    4.1. Which tool would be the best choice for handling this task?
    Ans: MapReduce can be utilized to handle this task.

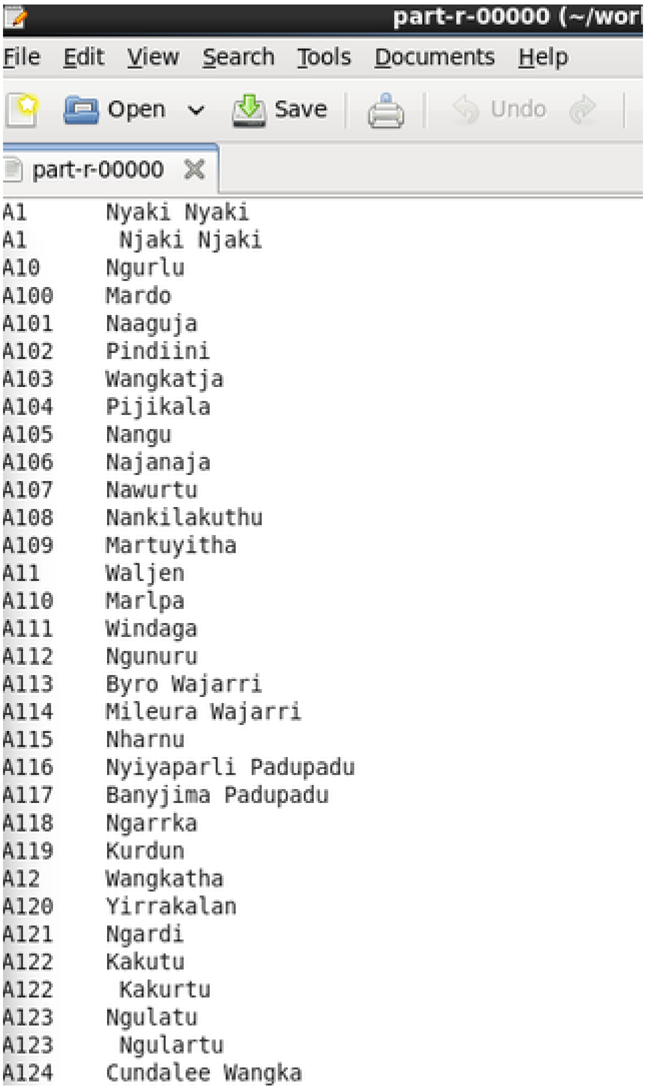
    4.2. List the code/statement of preparing the data for all weak entities.
    Mapper Code:



```java
package stubs;

import java.io.IOException;

public class WordMapper extends Mapper<LongWritable, Text, Text, Text> {

    private Text Key = new Text();
    private Text Key2 = new Text();

  @Override
  public void map(LongWritable key, Text value, Context context)
      throws IOException, InterruptedException {
    String line = value.toString();
    String[] data = line.split("\\t");
    String lng_name = data[1];
    String lng_code = data[0];
    String[] lng_name_part = lng_name.split("/");
    if (lng_name_part.length > 0){
        for (String name : lng_name_part){
            context.write(new Text(lng_code), new Text(name));
        }
    }

  }
}
```

4.3. Post the first-page screenshot of the rel_code_name processed result.

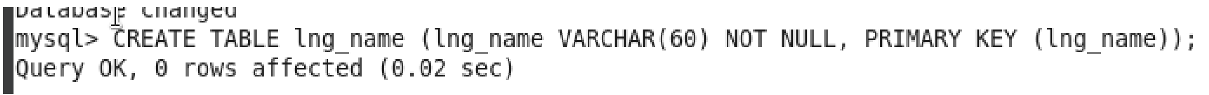
Ans:



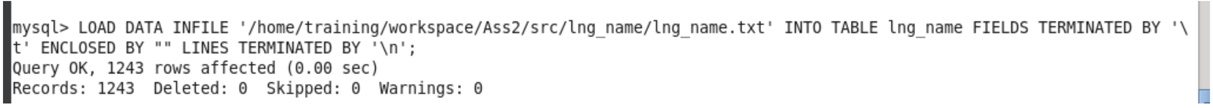
```
A1      Nyaki Nyaki
A1       Njaki Njaki
A10     Ngurlu
A100    Mardo
A101    Naaguja
A102    Pindiini
A103    Wangkatja
A104    Pijikala
A105    Nangu
A106    Najanaja
A107    Nawurtu
A108    Nankilakuthu
A109    Martuyitha
A11     Waljen
A110    Marlpa
A111    Windaga
A112    Ngunuru
A113    Byro Wajarri
A114    Mileura Wajarri
A115    Nharnu
A116    Nyiyaparli Padupadu
A117    Banyjima Padupadu
A118    Ngarrka
A119    Kurdun
A12     Wangkatha
A120    Yirrakalan
A121    Ngardi
A122    Kakutu
A122     Kakurtu
A123    Ngulatu
A123     Ngulartu
A124    Cundalee Wangka
```

5. Follow the given ERDs, create tables and import data into the created tables. All tables should be put in the database entitled “indigenous”.

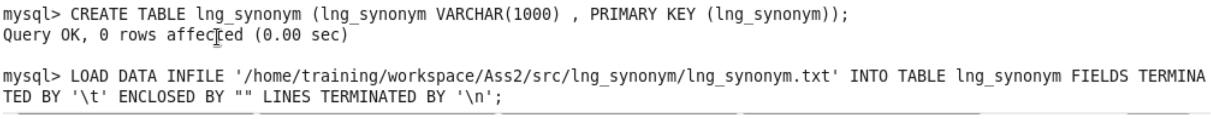
5.1. List all statements of this task for MySQL:



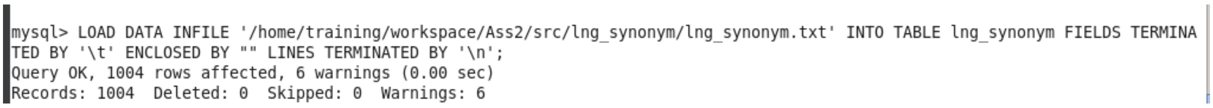
```
Database changed
mysql> CREATE TABLE lng_name (lng_name VARCHAR(60) NOT NULL, PRIMARY KEY (lng_name));
Query OK, 0 rows affected (0.02 sec)
```



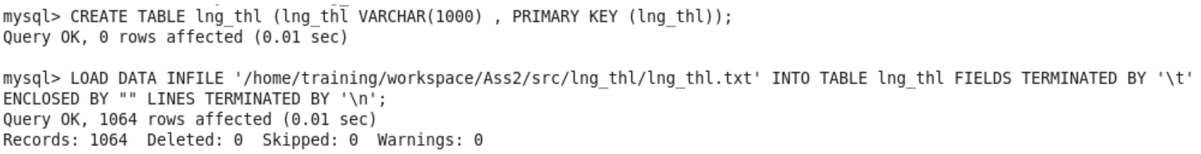
```
mysql> LOAD DATA INFILE '/home/training/workspace/Ass2/src/lng_name/lng_name.txt' INTO TABLE lng_name FIELDS TERMINATED BY '\
t' ENCLOSED BY "" LINES TERMINATED BY '\n';
Query OK, 1243 rows affected (0.00 sec)
Records: 1243  Deleted: 0  Skipped: 0  Warnings: 0
```



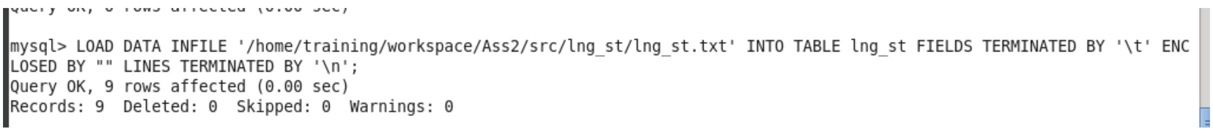
```
mysql> CREATE TABLE lng_synonym (lng_synonym VARCHAR(1000) , PRIMARY KEY (lng_synonym));
Query OK, 0 rows affected (0.00 sec)

mysql> LOAD DATA INFILE '/home/training/workspace/Ass2/src/lng_synonym/lng_synonym.txt' INTO TABLE lng_synonym FIELDS TERMINA
TED BY '\t' ENCLOSED BY "" LINES TERMINATED BY '\n';
```



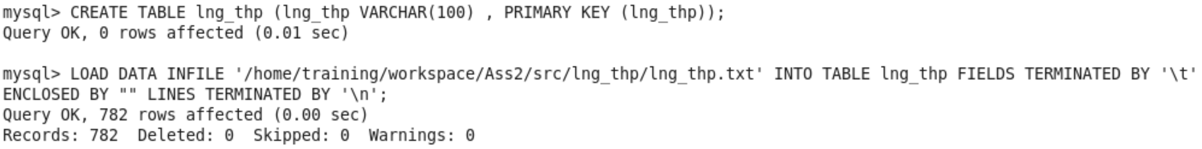
```
mysql> LOAD DATA INFILE '/home/training/workspace/Ass2/src/lng_synonym/lng_synonym.txt' INTO TABLE lng_synonym FIELDS TERMINA
TED BY '\t' ENCLOSED BY "" LINES TERMINATED BY '\n';
Query OK, 1004 rows affected, 6 warnings (0.00 sec)
Records: 1004  Deleted: 0  Skipped: 0  Warnings: 6
```



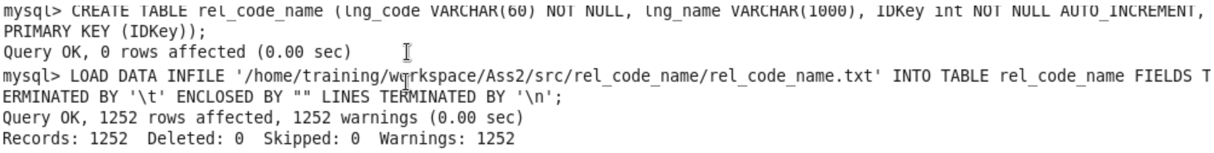
```
mysql> CREATE TABLE lng_thl (lng_thl VARCHAR(1000) , PRIMARY KEY (lng_thl));
Query OK, 0 rows affected (0.01 sec)

mysql> LOAD DATA INFILE '/home/training/workspace/Ass2/src/lng_thl/lng_thl.txt' INTO TABLE lng_thl FIELDS TERMINATED BY '\t'
ENCLOSED BY "" LINES TERMINATED BY '\n';
Query OK, 1064 rows affected (0.01 sec)
Records: 1064  Deleted: 0  Skipped: 0  Warnings: 0
```



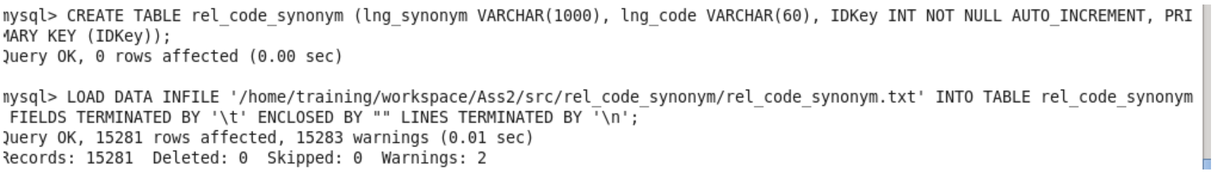
```
mysql> LOAD DATA INFILE '/home/training/workspace/Ass2/src/lng_st/lng_st.txt' INTO TABLE lng_st FIELDS TERMINATED BY '\t' ENC
LOSED BY "" LINES TERMINATED BY '\n';
Query OK, 9 rows affected (0.00 sec)
Records: 9  Deleted: 0  Skipped: 0  Warnings: 0
```



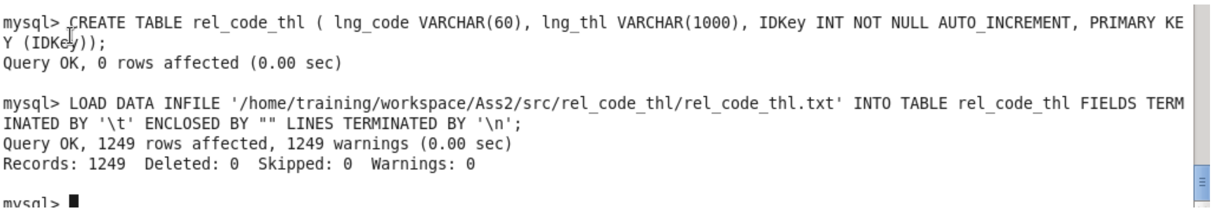
```
mysql> CREATE TABLE lng_thp (lng_thp VARCHAR(100) , PRIMARY KEY (lng_thp));
Query OK, 0 rows affected (0.01 sec)

mysql> LOAD DATA INFILE '/home/training/workspace/Ass2/src/lng_thp/lng_thp.txt' INTO TABLE lng_thp FIELDS TERMINATED BY '\t'
ENCLOSED BY "" LINES TERMINATED BY '\n';
Query OK, 782 rows affected (0.00 sec)
Records: 782  Deleted: 0  Skipped: 0  Warnings: 0
```



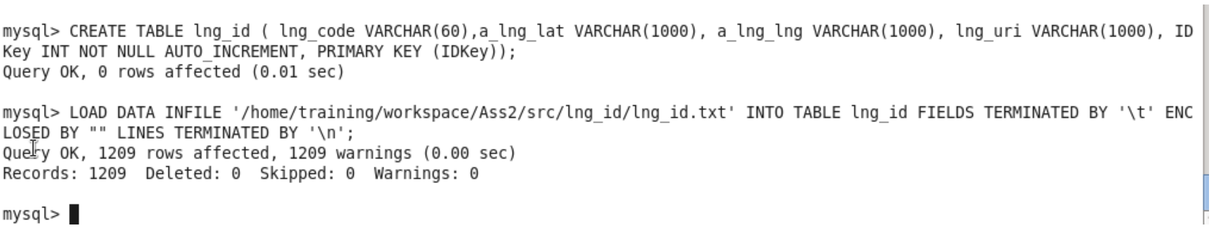
```
mysql> CREATE TABLE rel_code_name (lng_code VARCHAR(60) NOT NULL, lng_name VARCHAR(1000), IDKey int NOT NULL AUTO_INCREMENT,
PRIMARY KEY (IDKey));
Query OK, 0 rows affected (0.00 sec)
mysql> LOAD DATA INFILE '/home/training/workspace/Ass2/src/rel_code_name/rel_code_name.txt' INTO TABLE rel_code_name FIELDS T
ERMINATED BY '\t' ENCLOSED BY "" LINES TERMINATED BY '\n';
Query OK, 1252 rows affected, 1252 warnings (0.00 sec)
Records: 1252  Deleted: 0  Skipped: 0  Warnings: 1252
```



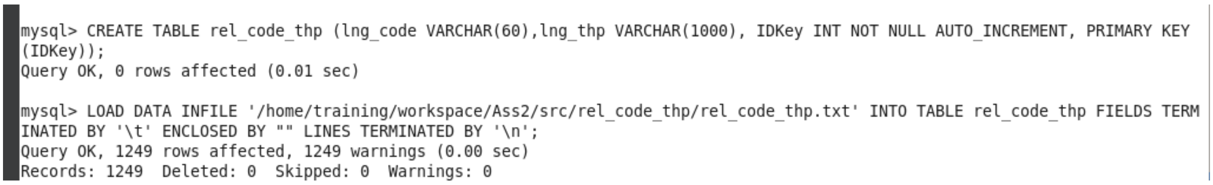
```
mysql> CREATE TABLE rel_code_synonym (lng_synonym VARCHAR(1000), lng_code VARCHAR(60), IDKey INT NOT NULL AUTO_INCREMENT, PRI
MARY KEY (IDKey));
Query OK, 0 rows affected (0.00 sec)

mysql> LOAD DATA INFILE '/home/training/workspace/Ass2/src/rel_code_synonym/rel_code_synonym.txt' INTO TABLE rel_code_synonym
 FIELDS TERMINATED BY '\t' ENCLOSED BY "" LINES TERMINATED BY '\n';
Query OK, 15281 rows affected, 15283 warnings (0.01 sec)
Records: 15281  Deleted: 0  Skipped: 0  Warnings: 2
```



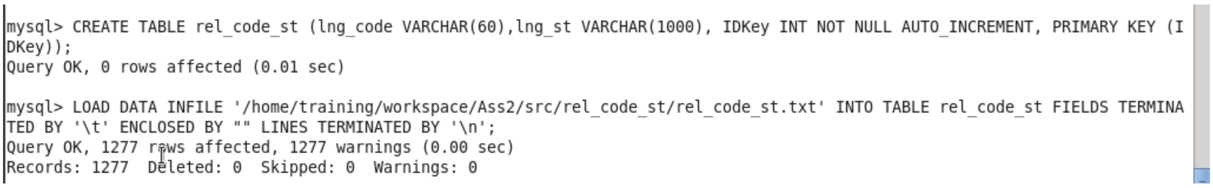
```
mysql> CREATE TABLE rel_code_thl ( lng_code VARCHAR(60), lng_thl VARCHAR(1000), IDKey INT NOT NULL AUTO_INCREMENT, PRIMARY KE
Y (IDKey));
Query OK, 0 rows affected (0.00 sec)

mysql> LOAD DATA INFILE '/home/training/workspace/Ass2/src/rel_code_thl/rel_code_thl.txt' INTO TABLE rel_code_thl FIELDS TERM
INATED BY '\t' ENCLOSED BY "" LINES TERMINATED BY '\n';
Query OK, 1249 rows affected, 1249 warnings (0.00 sec)
Records: 1249  Deleted: 0  Skipped: 0  Warnings: 0

mysql>
```



```
mysql> CREATE TABLE lng_id ( lng_code VARCHAR(60),a_lng_lat VARCHAR(1000), a_lng_lng VARCHAR(1000), lng_uri VARCHAR(1000), ID
Key INT NOT NULL AUTO_INCREMENT, PRIMARY KEY (IDKey));
Query OK, 0 rows affected (0.01 sec)

mysql> LOAD DATA INFILE '/home/training/workspace/Ass2/src/lng_id/lng_id.txt' INTO TABLE lng_id FIELDS TERMINATED BY '\t' ENC
LOSED BY "" LINES TERMINATED BY '\n';
Query OK, 1209 rows affected, 1209 warnings (0.00 sec)
Records: 1209  Deleted: 0  Skipped: 0  Warnings: 0

mysql>
```



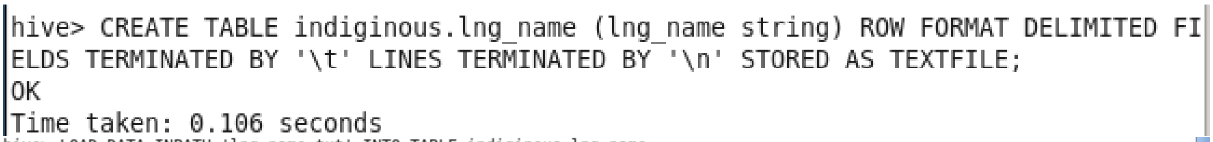
```
mysql> CREATE TABLE rel_code_thp (lng_code VARCHAR(60),lng_thp VARCHAR(1000), IDKey INT NOT NULL AUTO_INCREMENT, PRIMARY KEY
(IDKey));
Query OK, 0 rows affected (0.01 sec)

mysql> LOAD DATA INFILE '/home/training/workspace/Ass2/src/rel_code_thp/rel_code_thp.txt' INTO TABLE rel_code_thp FIELDS TERM
INATED BY '\t' ENCLOSED BY "" LINES TERMINATED BY '\n';
Query OK, 1249 rows affected, 1249 warnings (0.00 sec)
Records: 1249  Deleted: 0  Skipped: 0  Warnings: 0
```



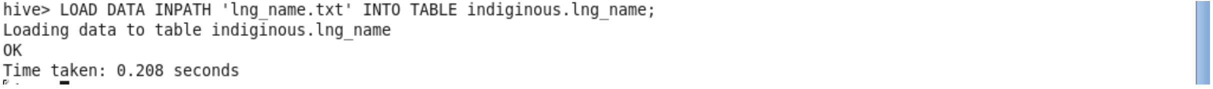
```
mysql> CREATE TABLE rel_code_st (lng_code VARCHAR(60),lng_st VARCHAR(1000), IDKey INT NOT NULL AUTO_INCREMENT, PRIMARY KEY (I
DKey));
Query OK, 0 rows affected (0.01 sec)

mysql> LOAD DATA INFILE '/home/training/workspace/Ass2/src/rel_code_st/rel_code_st.txt' INTO TABLE rel_code_st FIELDS TERMINA
TED BY '\t' ENCLOSED BY "" LINES TERMINATED BY '\n';
Query OK, 1277 rows affected, 1277 warnings (0.00 sec)
Records: 1277  Deleted: 0  Skipped: 0  Warnings: 0
```

5.2. List all statements of this task for Hive:

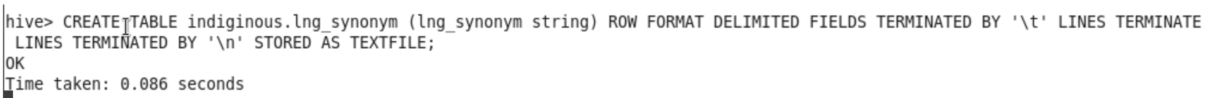
First, we need to put the text all the txt files into HDFS.



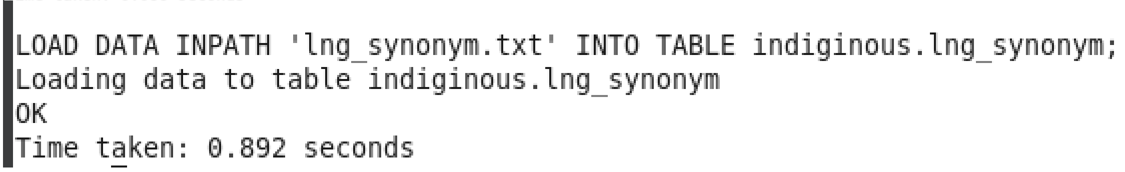
```
hive> CREATE TABLE indiginous.lng_name (lng_name string) ROW FORMAT DELIMITED FI
ELDS TERMINATED BY '\t' LINES TERMINATED BY '\n' STORED AS TEXTFILE;
OK
Time taken: 0.106 seconds
```



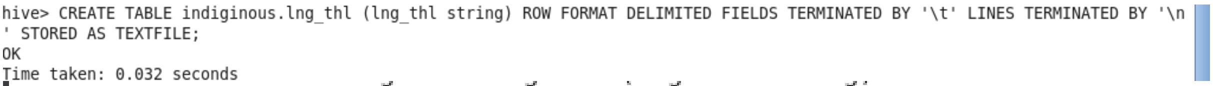
```
hive> LOAD DATA INPATH 'lng_name.txt' INTO TABLE indiginous.lng_name;
Loading data to table indiginous.lng_name
OK
Time taken: 0.208 seconds
```



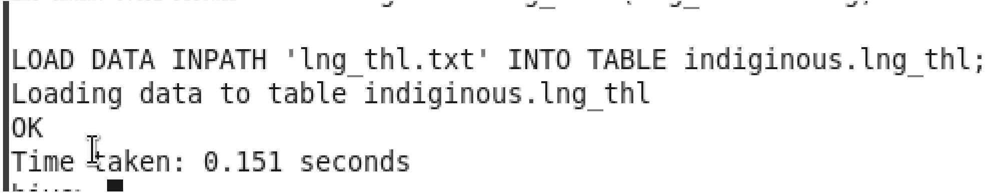
```
hive> CREATE TABLE indiginous.lng_synonym (lng_synonym string) ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' LINES TERMINATE
 LINES TERMINATED BY '\n' STORED AS TEXTFILE;
OK
Time taken: 0.086 seconds
```



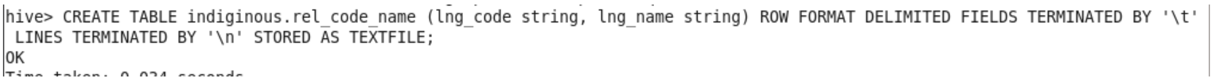
```
LOAD DATA INPATH 'lng_synonym.txt' INTO TABLE indiginous.lng_synonym;
Loading data to table indiginous.lng_synonym
OK
Time taken: 0.892 seconds
```



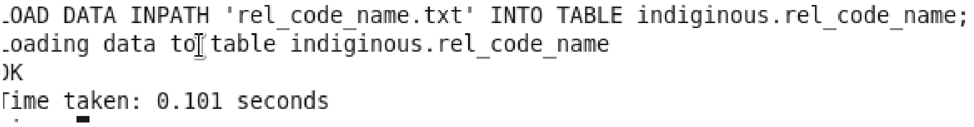
```
hive> CREATE TABLE indiginous.lng_thl (lng_thl string) ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' LINES TERMINATED BY '\n
' STORED AS TEXTFILE;
OK
Time taken: 0.032 seconds
```



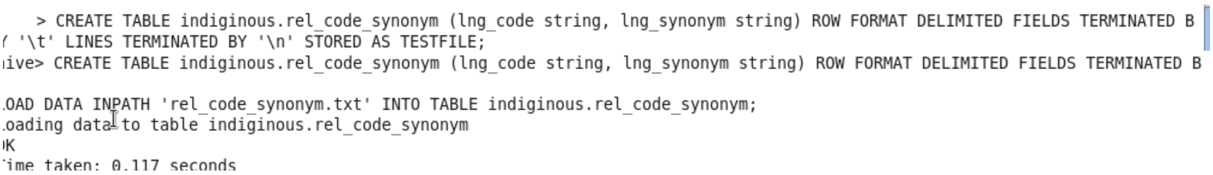
```
LOAD DATA INPATH 'lng_thl.txt' INTO TABLE indiginous.lng_thl;
Loading data to table indiginous.lng_thl
OK
Time taken: 0.151 seconds
```



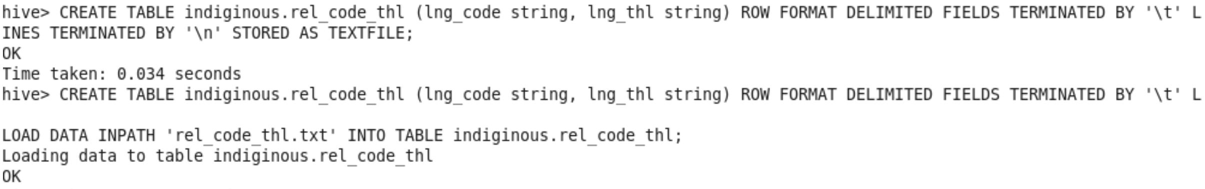
```
hive> CREATE TABLE indiginous.rel_code_name (lng_code string, lng_name string) ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t'
 LINES TERMINATED BY '\n' STORED AS TEXTFILE;
OK
Time taken: 0.034 seconds
```



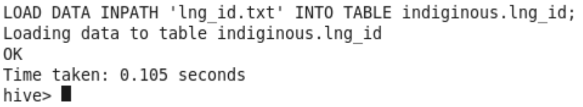
```
LOAD DATA INPATH 'rel_code_name.txt' INTO TABLE indiginous.rel_code_name;
Loading data to table indiginous.rel_code_name
OK
Time taken: 0.101 seconds
```



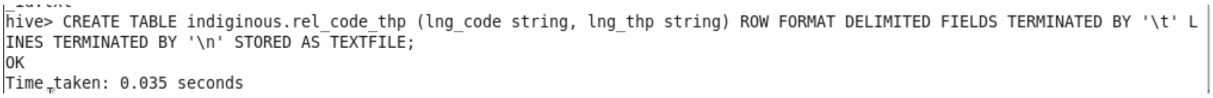
```
   > CREATE TABLE indiginous.rel_code_synonym (lng_code string, lng_synonym string) ROW FORMAT DELIMITED FIELDS TERMINATED B
Y '\t' LINES TERMINATED BY '\n' STORED AS TESTFILE;
hive> CREATE TABLE indiginous.rel_code_synonym (lng_code string, lng_synonym string) ROW FORMAT DELIMITED FIELDS TERMINATED B

LOAD DATA INPATH 'rel_code_synonym.txt' INTO TABLE indiginous.rel_code_synonym;
Loading data to table indiginous.rel_code_synonym
OK
Time taken: 0.117 seconds
```



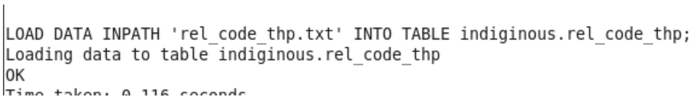
```
hive> CREATE TABLE indiginous.rel_code_thl (lng_code string, lng_thl string) ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' L
INES TERMINATED BY '\n' STORED AS TEXTFILE;
OK
Time taken: 0.034 seconds
hive> CREATE TABLE indiginous.rel_code_thl (lng_code string, lng_thl string) ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' L

LOAD DATA INPATH 'rel_code_thl.txt' INTO TABLE indiginous.rel_code_thl;
Loading data to table indiginous.rel_code_thl
OK
```



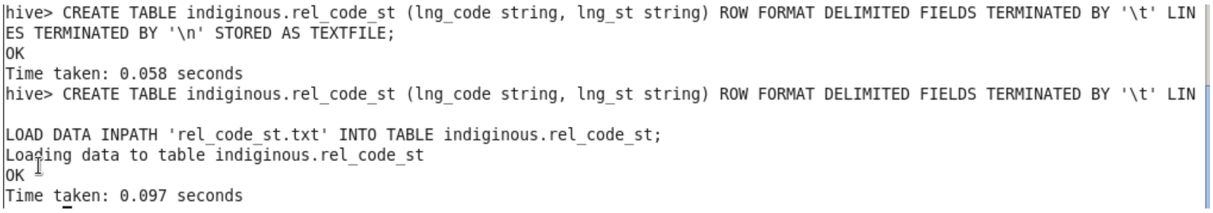
```
LOAD DATA INPATH 'lng_id.txt' INTO TABLE indiginous.lng_id;
Loading data to table indiginous.lng_id
OK
Time taken: 0.105 seconds
hive>
```



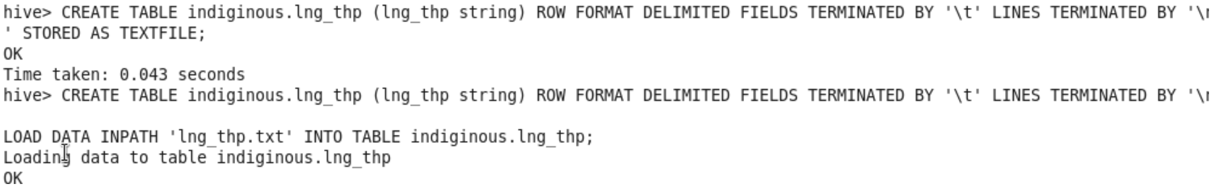
```
hive> CREATE TABLE indiginous.rel_code_thp (lng_code string, lng_thp string) ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' L
INES TERMINATED BY '\n' STORED AS TEXTFILE;
OK
Time taken: 0.035 seconds
```



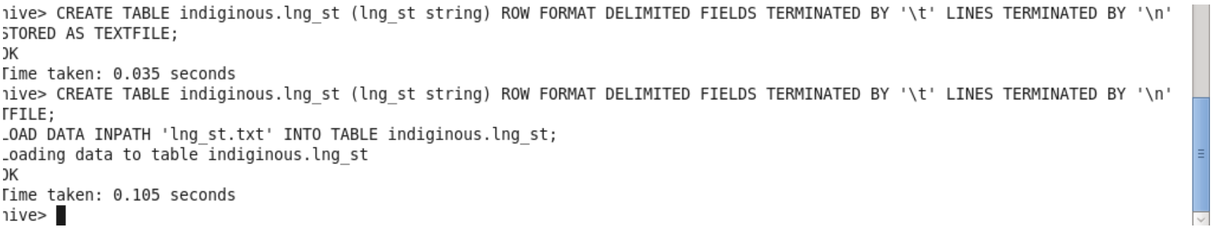
```
LOAD DATA INPATH 'rel_code_thp.txt' INTO TABLE indiginous.rel_code_thp;
Loading data to table indiginous.rel_code_thp
OK
Time taken: 0.116 seconds
```



```
hive> CREATE TABLE indiginous.rel_code_st (lng_code string, lng_st string) ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' LIN
ES TERMINATED BY '\n' STORED AS TEXTFILE;
OK
Time taken: 0.058 seconds
hive> CREATE TABLE indiginous.rel_code_st (lng_code string, lng_st string) ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' LIN

LOAD DATA INPATH 'rel_code_st.txt' INTO TABLE indiginous.rel_code_st;
Loading data to table indiginous.rel_code_st
OK
Time taken: 0.097 seconds
```



```
hive> CREATE TABLE indiginous.lng_thp (lng_thp string) ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' LINES TERMINATED BY '\
' STORED AS TEXTFILE;
OK
Time taken: 0.043 seconds
hive> CREATE TABLE indiginous.lng_thp (lng_thp string) ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' LINES TERMINATED BY '\

LOAD DATA INPATH 'lng_thp.txt' INTO TABLE indiginous.lng_thp;
Loading data to table indiginous.lng_thp
OK
```



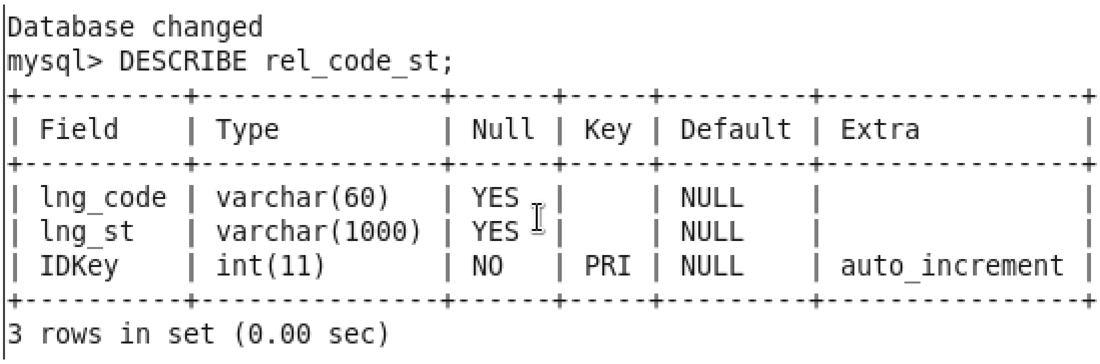
```
hive> CREATE TABLE indiginous.lng_st (lng_st string) ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' LINES TERMINATED BY '\n'
STORED AS TEXTFILE;
OK
Time taken: 0.035 seconds
hive> CREATE TABLE indiginous.lng_st (lng_st string) ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' LINES TERMINATED BY '\n'
TFILE;
LOAD DATA INPATH 'lng_st.txt' INTO TABLE indiginous.lng_st;
Loading data to table indiginous.lng_st
OK
Time taken: 0.105 seconds
hive>
```

6. Show the detailed information of table rel_code_st in MySQL.

6.1. Statement:

Ans: DESCRIBE rel_code_st;

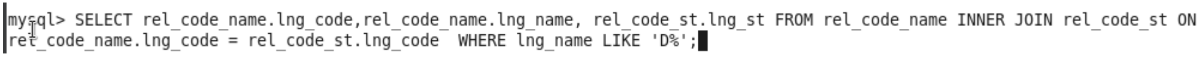
6.2. Result screenshot:



```
Database changed
mysql> DESCRIBE rel_code_st;
+----------+----------------+------+-----+---------+----------------+
| Field    | Type           | Null | Key | Default | Extra          |
+----------+----------------+------+-----+---------+----------------+
| lng_code | varchar(60)    | YES  |     | NULL    |                |
| lng_st   | varchar(1000)  | YES  |     | NULL    |                |
| IDKey    | int(11)        | NO   | PRI | NULL    | auto_increment |
+----------+----------------+------+-----+---------+----------------+
3 rows in set (0.00 sec)
```

7. Collect the results containing "lng_code", "lng_name", and "lng_st" of tuples whose lng_name starts with the upper case "D". Show bot statements and the last page screenshot of your query result in MySQL and Hive.

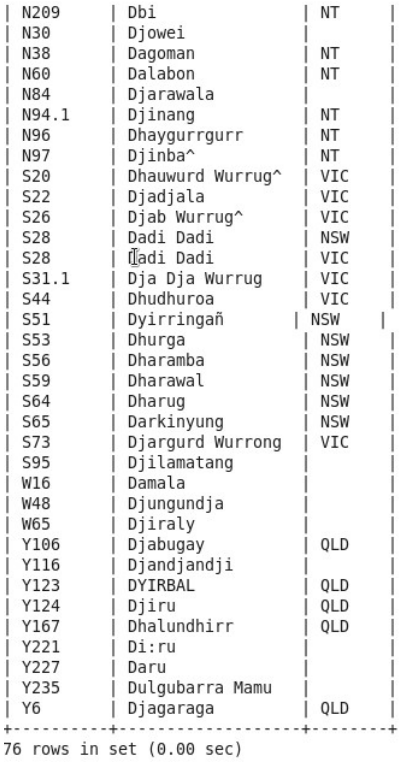
7.1. MySQL query statement:



```
mysql> SELECT rel_code_name.lng_code,rel_code_name.lng_name, rel_code_st.lng_st FROM rel_code_name INNER JOIN rel_code_st ON
rel_code_name.lng_code = rel_code_st.lng_code  WHERE lng_name LIKE 'D%';
```

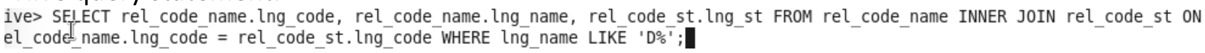
7.2. MySQL result screenshot:

Ans:



```
| N209      | Dbi              | NT     |
| N30       | Djowei           |        |
| N38       | Dagoman          | NT     |
| N60       | Dalabon          | NT     |
| N84       | Djarawala        |        |
| N94.1     | Djinang          | NT     |
| N96       | Dhaygurrgurr     | NT     |
| N97       | Djinba^          | NT     |
| S20       | Dhauwurd Wurrug^ | VIC    |
| S22       | Djadjala         | VIC    |
| S26       | Djab Wurrug^     | VIC    |
| S28       | Dadi Dadi        | NSW    |
| S28       | Dadi Dadi        | VIC    |
| S31.1     | Dja Dja Wurrug   | VIC    |
| S44       | Dhudhuroa        | VIC    |
| S51       | Dyirringañ      | NSW    |
| S53       | Dhurga           | NSW    |
| S56       | Dharamba         | NSW    |
| S59       | Dharawal         | NSW    |
| S64       | Dharug           | NSW    |
| S65       | Darkinyung       | NSW    |
| S73       | Djargurd Wurrong | VIC    |
| S95       | Djilamatang      |        |
| W16       | Damala           |        |
| W48       | Djungundja       |        |
| W65       | Djiraly          |        |
| Y106      | Djabugay         | QLD    |
| Y116      | Djandjandji      |        |
| Y123      | DYIRBAL          | QLD    |
| Y124      | Djiru            | QLD    |
| Y167      | Dhalundhirr      | QLD    |
| Y221      | Di:ru            |        |
| Y227      | Daru             |        |
| Y235      | Dulgubarra Mamu  |        |
| Y6        | Djagaraga        | QLD    |
+-----------+------------------+--------+
76 rows in set (0.00 sec)
```

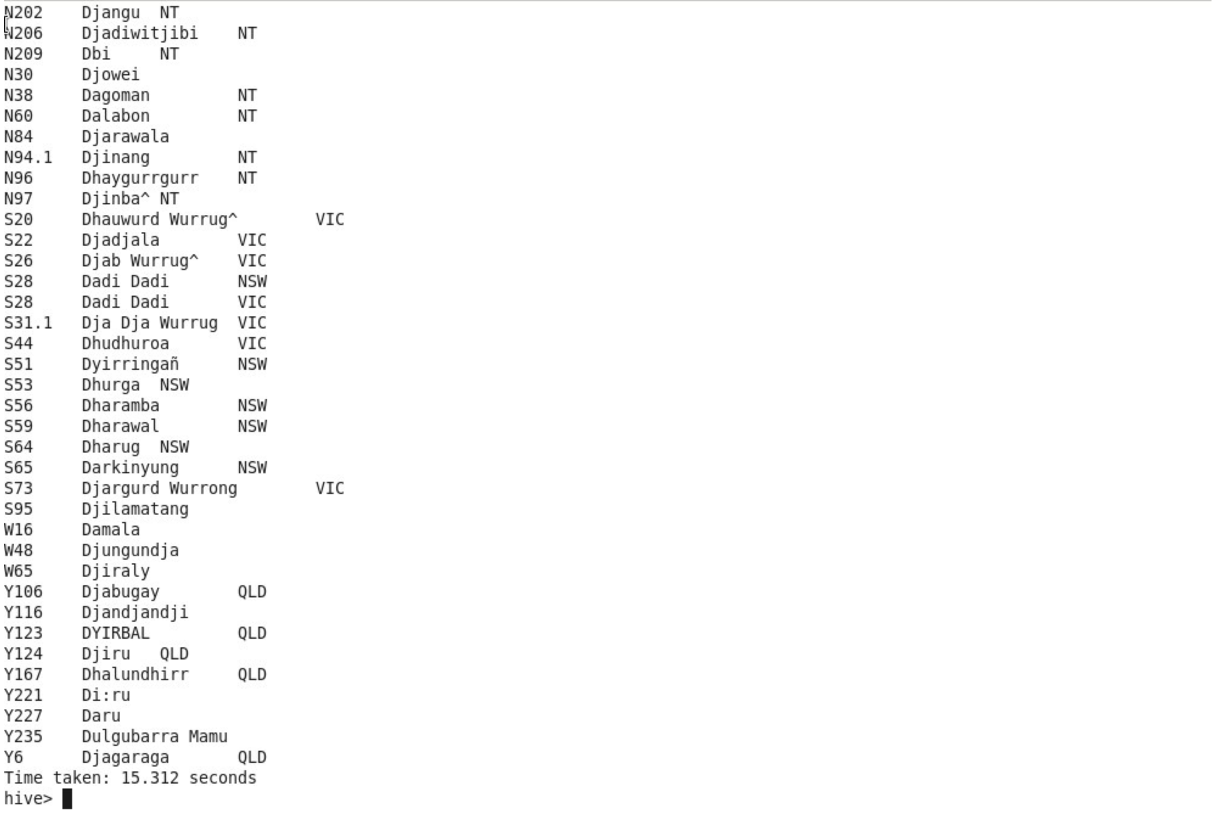
7.3. Hive query statement:



```
ive> SELECT rel_code_name.lng_code, rel_code_name.lng_name, rel_code_st.lng_st FROM rel_code_name INNER JOIN rel_code_st ON
el_code_name.lng_code = rel_code_st.lng_code WHERE lng_name LIKE 'D%';
```

7.4. Hive result screenshot:

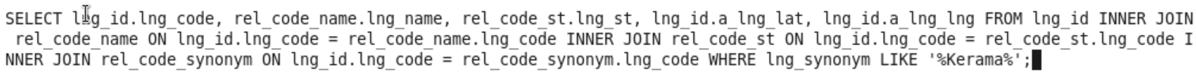
Ans:



```
N202    Djangu  NT
N206    Djadiwitjibi    NT
N209    Dbi     NT
N30     Djowei
N38     Dagoman         NT
N60     Dalabon         NT
N84     Djarawala
N94.1   Djinang         NT
N96     Dhaygurrgurr    NT
N97     Djinba^ NT
S20     Dhauwurd Wurrug^        VIC
S22     Djadjala        VIC
S26     Djab Wurrug^    VIC
S28     Dadi Dadi       NSW
S28     Dadi Dadi       VIC
S31.1   Dja Dja Wurrug  VIC
S44     Dhudhuroa       VIC
S51     Dyirringañ      NSW
S53     Dhurga  NSW
S56     Dharamba        NSW
S59     Dharawal        NSW
S64     Dharug  NSW
S65     Darkinyung      NSW
S73     Djargurd Wurrong        VIC
S95     Djilamatang
W16     Damala
W48     Djungundja
W65     Djiraly
Y106    Djabugay        QLD
Y116    Djandjandji
Y123    DYIRBAL         QLD
Y124    Djiru   QLD
Y167    Dhalundhirr     QLD
Y221    Di:ru
Y227    Daru
Y235    Dulgubarra Mamu
Y6      Djagaraga       QLD
Time taken: 15.312 seconds
hive>
```

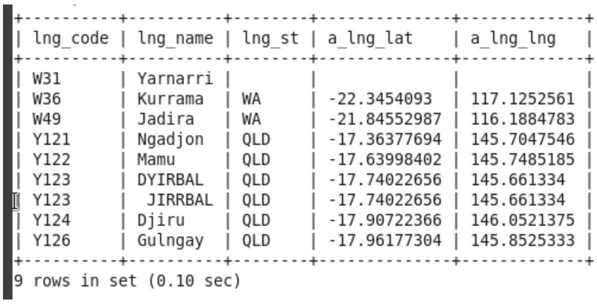
8. Collect the results containing "lng_code", "lng_name", "lng_st", "a_lng_lat", and "a_lng_lng" of tuples whose lng_synonym contains "Kerama". Show both statements and the last page screenshot of your query result in MySQL and Hive.
   8.1. How many tuples are retrieved at the end?
   Ans: 9 tuples are retrieved at the end.

   8.2. MySQL query statement:



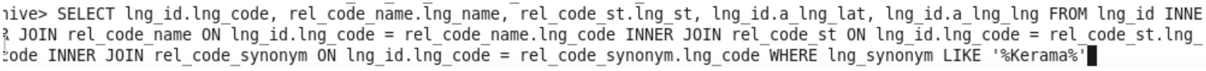
```
SELECT lng_id.lng_code, rel_code_name.lng_name, rel_code_st.lng_st, lng_id.a_lng_lat, lng_id.a_lng_lng FROM lng_id INNER JOIN
 rel_code_name ON lng_id.lng_code = rel_code_name.lng_code INNER JOIN rel_code_st ON lng_id.lng_code = rel_code_st.lng_code I
NNER JOIN rel_code_synonym ON lng_id.lng_code = rel_code_synonym.lng_code WHERE lng_synonym LIKE '%Kerama%';
```

   8.3. MySQL result screenshot:
   Ans:



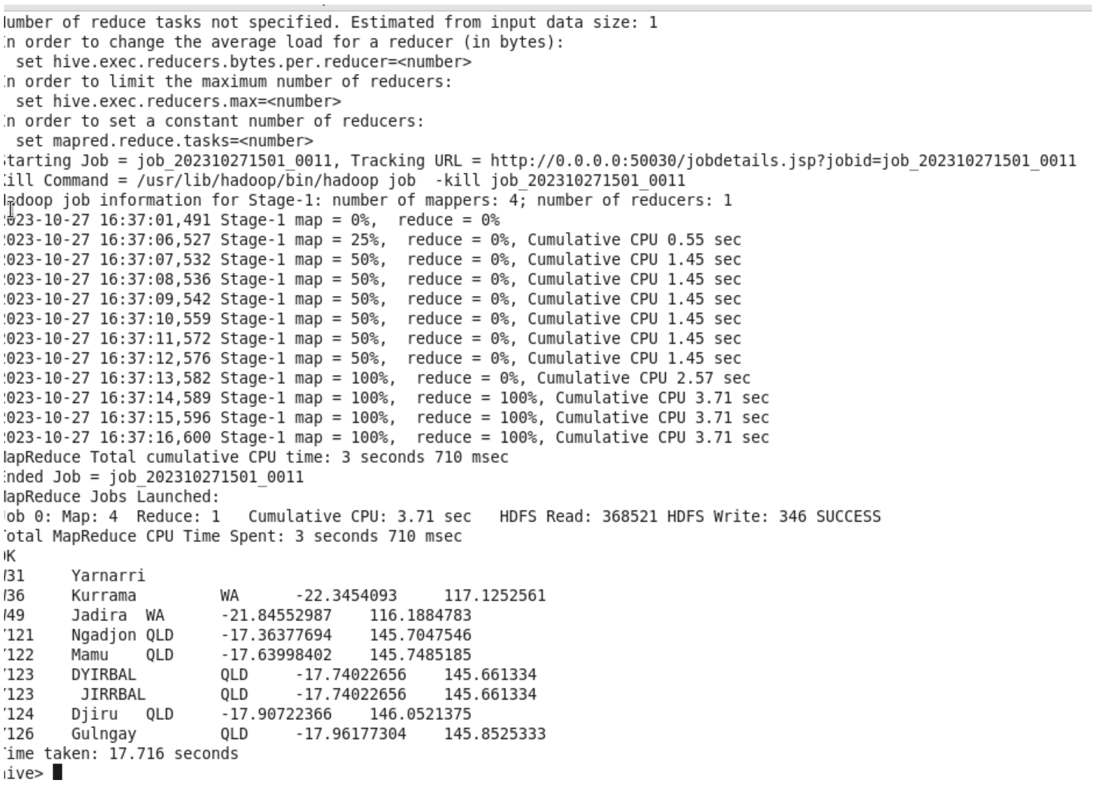
```
+----------+----------+--------+--------------+-------------+
| lng_code | lng_name | lng_st | a_lng_lat    | a_lng_lng   |
+----------+----------+--------+--------------+-------------+
| W31      | Yarnarri |        |              |             |
| W36      | Kurrama  | WA     | -22.3454093  | 117.1252561 |
| W49      | Jadira   | WA     | -21.84552987 | 116.1884783 |
| Y121     | Ngadjon  | QLD    | -17.36377694 | 145.7047546 |
| Y122     | Mamu     | QLD    | -17.63998402 | 145.7485185 |
| Y123     | DYIRBAL  | QLD    | -17.74022656 | 145.661334  |
| Y123     |  JIRRBAL | QLD    | -17.74022656 | 145.661334  |
| Y124     | Djiru    | QLD    | -17.90722366 | 146.0521375 |
| Y126     | Gulngay  | QLD    | -17.96177304 | 145.8525333 |
+----------+----------+--------+--------------+-------------+
9 rows in set (0.10 sec)
```

   8.4. Hive query statement:



```
ive> SELECT lng_id.lng_code, rel_code_name.lng_name, rel_code_st.lng_st, lng_id.a_lng_lat, lng_id.a_lng_lng FROM lng_id INNE
R JOIN rel_code_name ON lng_id.lng_code = rel_code_name.lng_code INNER JOIN rel_code_st ON lng_id.lng_code = rel_code_st.lng_
ode INNER JOIN rel_code_synonym ON lng_id.lng_code = rel_code_synonym.lng_code WHERE lng_synonym LIKE '%Kerama%'
```

   8.5. Hive result screenshot:
   Ans:



```
lumber of reduce tasks not specified. Estimated from input data size: 1
:n order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
:n order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
:n order to set a constant number of reducers:
  set mapred.reduce.tasks=<number>
;tarting Job = job_202310271501_0011, Tracking URL = http://0.0.0.0:50030/jobdetails.jsp?jobid=job_202310271501_0011
:ill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_202310271501_0011
ladoop job information for Stage-1: number of mappers: 4; number of reducers: 1
:023-10-27 16:37:01,491 Stage-1 map = 0%,  reduce = 0%
:023-10-27 16:37:06,527 Stage-1 map = 25%,  reduce = 0%, Cumulative CPU 0.55 sec
:023-10-27 16:37:07,532 Stage-1 map = 50%,  reduce = 0%, Cumulative CPU 1.45 sec
:023-10-27 16:37:08,536 Stage-1 map = 50%,  reduce = 0%, Cumulative CPU 1.45 sec
:023-10-27 16:37:09,542 Stage-1 map = 50%,  reduce = 0%, Cumulative CPU 1.45 sec
:023-10-27 16:37:10,559 Stage-1 map = 50%,  reduce = 0%, Cumulative CPU 1.45 sec
:023-10-27 16:37:11,572 Stage-1 map = 50%,  reduce = 0%, Cumulative CPU 1.45 sec
:023-10-27 16:37:12,576 Stage-1 map = 50%,  reduce = 0%, Cumulative CPU 1.45 sec
:023-10-27 16:37:13,582 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.57 sec
:023-10-27 16:37:14,589 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 3.71 sec
:023-10-27 16:37:15,596 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 3.71 sec
:023-10-27 16:37:16,600 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 3.71 sec
lapReduce Total cumulative CPU time: 3 seconds 710 msec
:nded Job = job_202310271501_0011
lapReduce Jobs Launched:
lob 0: Map: 4  Reduce: 1   Cumulative CPU: 3.71 sec   HDFS Read: 368521 HDFS Write: 346 SUCCESS
'otal MapReduce CPU Time Spent: 3 seconds 710 msec
)K
/31     Yarnarri
/36     Kurrama         WA      -22.3454093     117.1252561
/49     Jadira  WA      -21.84552987    116.1884783
'121    Ngadjon QLD     -17.36377694    145.7047546
'122    Mamu    QLD     -17.63998402    145.7485185
'123    DYIRBAL         QLD     -17.74022656    145.661334
'123     JIRRBAL        QLD     -17.74022656    145.661334
'124    Djiru   QLD     -17.90722366    146.0521375
'126    Gulngay         QLD     -17.96177304    145.8525333
'ime taken: 17.716 seconds
ive>
```