

ASSIGNMENT

1.

```
grunt> line= load '/hbase/msd.txt' AS (line:chararray);
2019-01-14 20:08:06,991 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
grunt> words =FOREACH lines GENERATE FLATTEN(TOKENIZE(line)) as word;
2019-01-14 20:08:15,389 [main] ERROR org.apache.pig.tools.grunt.Grunt - ERROR 1200: Pig script failed to parse:
<line 6, column 15> Undefined alias: lines
Details at logfile: /home/acadgild/install/pig/pig-0.16.0/bin/pig_1547475496061.log
grunt> grouped= GROUP words BY word;
grunt> wordcount= FOREACH grouped GENERATE group, COUNT(words);
grunt> █
```

- a.load command is used to load database from hdfs to pig .
- b.foreach command is used to apply a condition to each data.
- c.generate flatten command changes the structure of tuples and bags.
- d.group by commands groups the element of the table according to a condition.

2.input text.

```
-----
(Mahendra Singh Dhoni is an Indian international cricketer who captained the Indian national team in limited-overs formats from 2007 to 2016 and in Test cricket from 2008 to 2014.)
grunt> █
```

3.output:word count:

```
ne.util.MapRedUtil - Total input paths to process : 1
(an,1)
(in,2)
(is,1)
(to,2)
(and,1)
(the,1)
(who,1)
(2007,1)
(2008,1)
(2016,1)
(Test,1)
(from,2)
(team,1)
(2014.,1)
(Dhoni,1)
(Singh,1)
(Indian,2)
(cricket,1)
(formats,1)
(Mahendra,1)
(national,1)
(captained,1)
(cricketer,1)
(international,1)
(limited-overs,1)
grunt> █
```

TABLES FOR TASK 2 AND 3 ARE NOT AVAILABLE ON THE GIVEN LINK.