Exercise 3:

Learning goal:

- Pig script for a mapreduce program

1. NASA Web hit count

```
data = load 'nasa.log';

b = foreach data generate FLATTEN(STRSPLIT($0,':',2));

d = foreach data generate SUBSTRING($0, LAST_INDEX_OF($0,'[')+1, LAST_INDEX_OF($0,'[')+12) as day;

Groups = GROUP d BY day;

Counts = FOREACH Groups GENERATE  group, COUNT(d);

dump Counts;
```

2. Word Count

```
Lines=LOAD 'input/inputfile.txt' AS (line: chararray);

Words = FOREACH Lines GENERATE  FLATTEN(TOKENIZE(line)) AS word;

Groups = GROUP Words BY word;

Counts = FOREACH Groups GENERATE  group, COUNT(Words);

Top5 = LIMIT Results 5;

STORE Top5 INTO '/output/outputfile.txt';
```