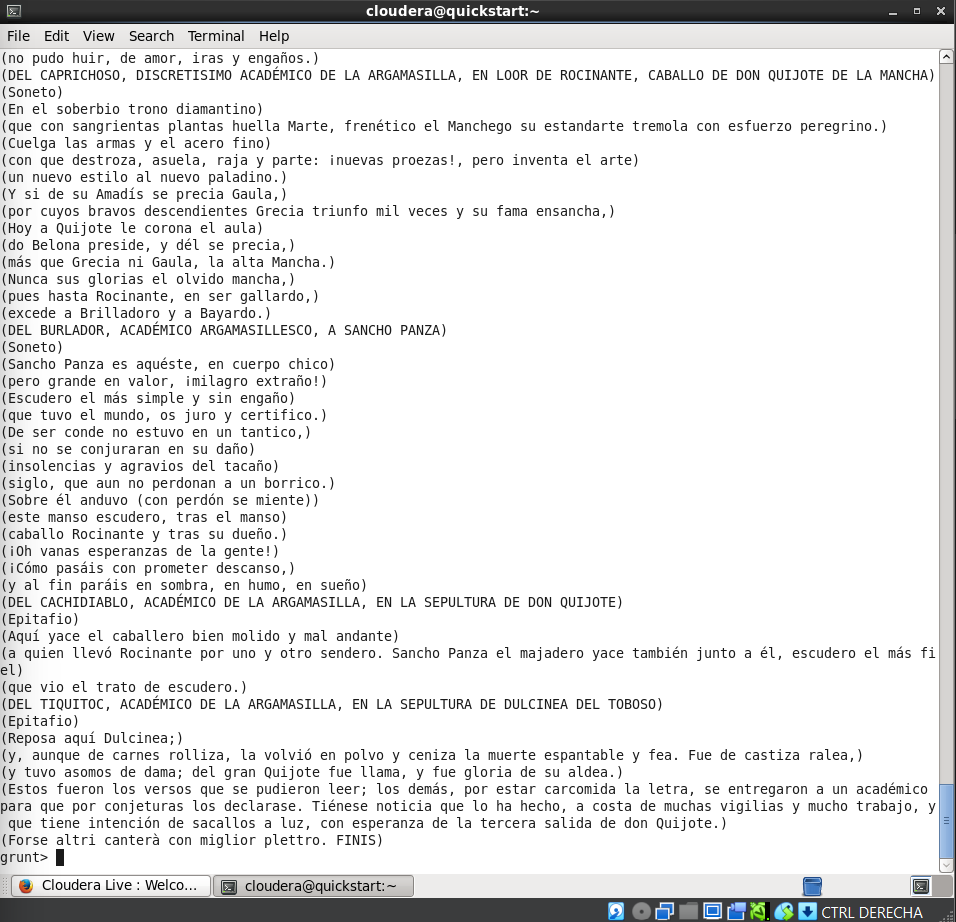
* Cargo el fichero quijote.txt:

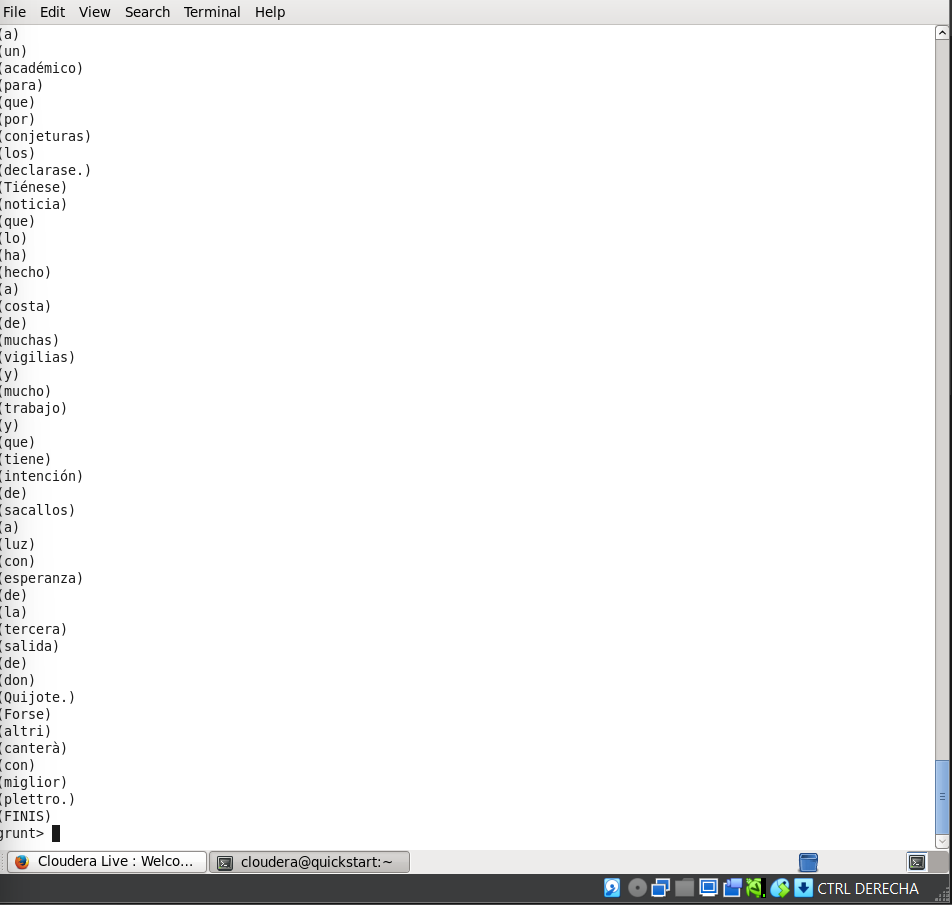
data = LOAD ‘quijote/quijote.txt’ USING PigStorage(‘\n’) AS (line:chararray);

DUMP data;



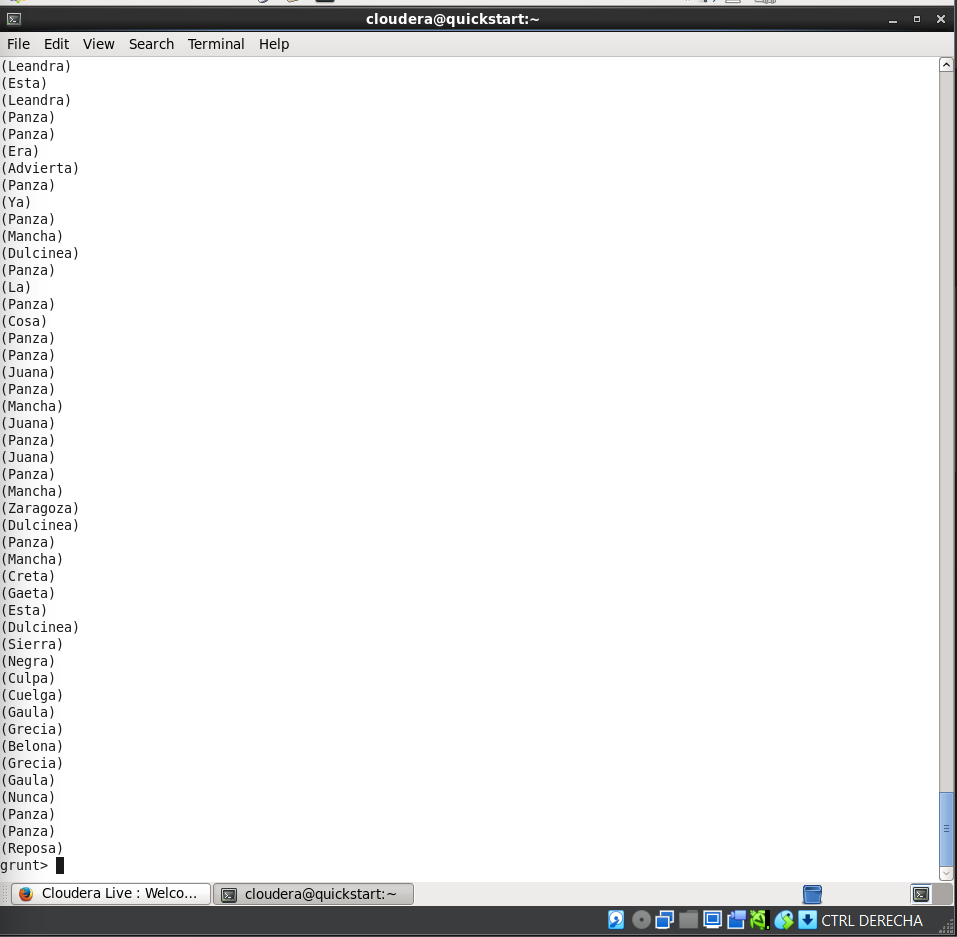
* Separo cada línea en palabras:  
    
  words = FOREACH data GENERATE FLATTEN(TOKENIZE(line)) AS word;

DUMP words;



* Filtro las palabras que empiezan por mayúscula y acaban por ‘a’:  
    
  filter\_words = FILTER words BY word MATCHES ‘^[A-Z].\*a$’;

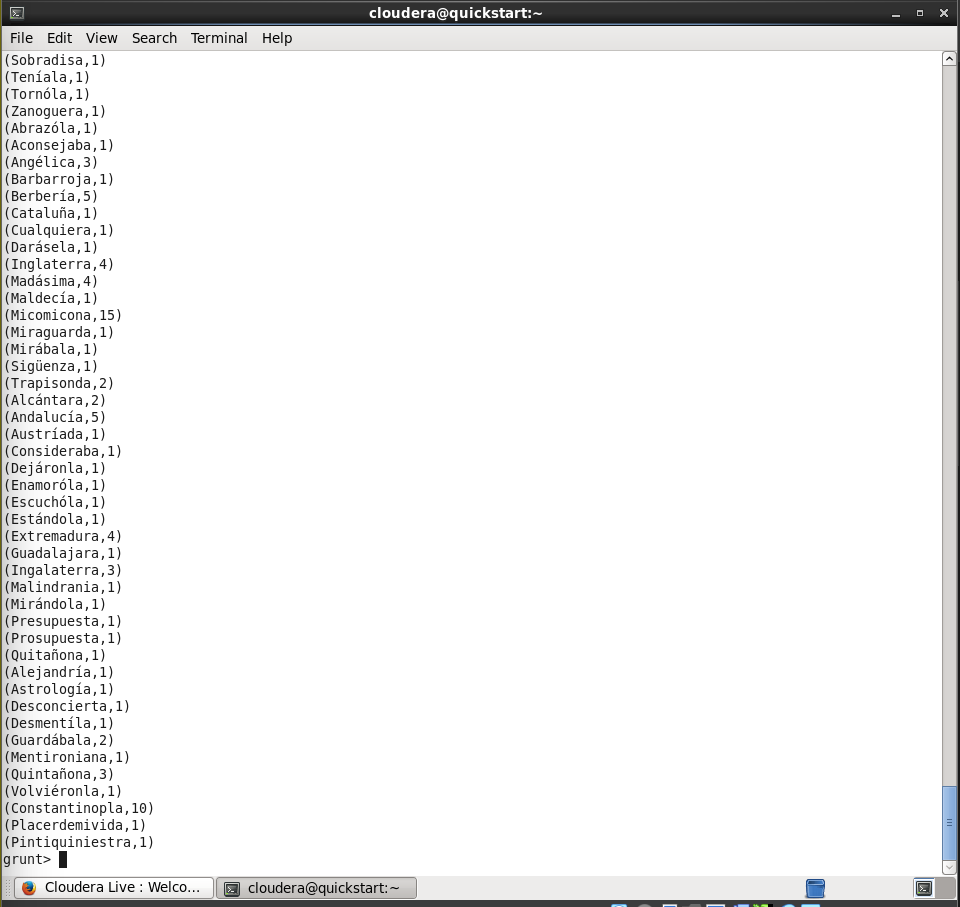
DUMP filter\_words;



* Contar las ocurrencias de cada palabra:

word\_counts = FOREACH (GROUP filter\_words BY word) GENERATE group AS word, COUNT(filter\_words) AS count;

DUMP word\_counts;



* Ordeno el resultado por mayor numero de ocurrencias y por orden alfabético por si hubiera alguna repetida:

sorted\_words = ORDER word\_counts BY count DESC, word ASC;

DUMP sorted\_words

(Las capturas de pantallas de la ejecución las adjuntaré a parte en el directorio)