**Steps for working with sample.log**

log = LOAD ‘/input/sample.log’; [Note please copy file from windows – linux via WinSCP and then use –copyFromLocal to move to HDFS.]

LEVELS = foreach log generate REGEX\_EXTRACT($0, '(TRACE|DEBUG|INFO|WARN|ERROR|FATAL)', 1) as LOGLEVEL;

dump LEVELS;



FILTEREDLEVELS = FILTER LEVELS by LOGLEVEL is not null;

dump FILTEREDLEVELS

//, which is an alias, that holds these words from each record – TRACE, DEBUG,INFO, WARN, ERROR & FATAL and removes any NULL or EMPTY words.

grunt> GROUPEDLEVELS = GROUP FILTEREDLEVELS by LOGLEVEL;

//Group all of the log levels into their own row (counting is not done yet; it occurs in next step):

grunt> FREQUENCIES = foreach GROUPEDLEVELS generate group as LOGLEVEL, COUNT(FILTEREDLEVELS.LOGLEVEL) as COUNT;

//For each group, count the occurrences of log levels. These is the frequencies of each log level:

grunt> dump FREQUENCIES;

//Dump the FREQUENCIES alias to see the count of the occurrences of each word in a group.

grunt> RESULT = order FREQUENCIES by COUNT desc;

STORE RESULT into ‘/output/pigresults’