##################TYPES OF LOGS###################

* Web Server Access Logs
* Web Server Error Logs
* Application Server Logs

###########TYPES OF LOG FILES####

A) Common Log File format

E.G: ==== "%hostname %logname %username %time \"%r\" %>statuscode %bytes"

Data == 64.242.88.10 - - [07/Mar/2014:16:47:12 -0800] "GET /robots.txt HTTP/1.1" 200 68

1. DEFINE ApacheCommonLogLoader org.apache.pig.piggybank.storage.apachelog.CommonLogLoader();

2. logs = LOAD 'logfile1' USING ApacheCommonLogLoader AS (addr: chararray, logname: chararray, user: chararray, time: chararray, method: chararray, uri: chararray, proto: chararray, status: int, bytes: int);

3.addrs = GROUP logs BY addr;

4.counts = FOREACH addrs GENERATE flatten($0), COUNT($1) as count;

5. DUMP counts;

2) Combined Log File format

Example:

"%hostname %logname %username %time \"%r\" %>statuscode %bytes \"%{Referer}i\" \"%{User-Agent}i\""

Data:

122.172.200.100 - - [17/Oct/2014:00:04:36 -0400] "GET /tag/hbase-sink/ HTTP/1.1" 200 15997 "https://www.google.co.in/" "Mozilla/5.0 (X11; Linux x86\_64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/33.0.1750.117 Safari/537.36"

1.logs = LOAD 'logfile2' USING org.apache.pig.piggybank.storage.apachelog.CombinedLogLoader() AS (addr: chararray, logname: chararray, user: chararray, time: chararray,method: chararray, uri: chararray, proto: chararray,status: int, bytes: int, referer: chararray, userAgent: chararray);

2. refers = GROUP logs BY referer;

3. refer\_counts = FOREACH refers GENERATE flatten($0), COUNT($1) as count;

4. sorted\_refer\_counts = ORDER refer\_counts BY count DESC;

5. top10\_refers = LIMIT sorted\_refer\_counts 10;

6. STORE top10\_refers INTO '/out/topreferrers/' USING PigStorage('\t');

3) Custom Log File Format

Example:

Data:

1!!!one~i

1. A = LOAD 'logfile3' USING org.apache.pig.piggybank.storage.MyRegExLoader('(\\d+)!+(\\w+)~+(\\w+)');