**ASSIGNMENT -2**

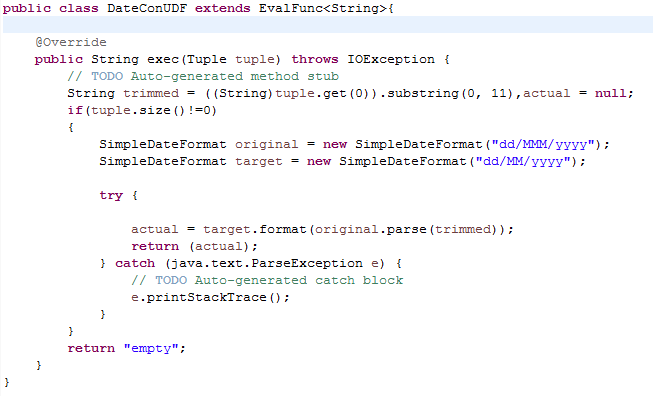
1.Combine Log Loader using piggybank

REGISTER '/usr/lib/pig/piggybank.jar'

combineLogLoaderData = LOAD '/user/logs/log' 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. Custom to ISO and ISO to Custom

DataCon.jar



REGISTER '/home/ec2-user/DataCon.jar'

combineLogDateData = FOREACH combineLogLoaderData GENERATE addr,logname,user,DateConUDF(time) AS date,method,uri,proto,status,bytes,referer,userAgent;

DEFINE CustomFormatToISO org.apache.pig.piggybank.evaluation.datetime.convert.CustomFormatToISO();

convertToIso = FOREACH combineLogDateData GENERATE org.apache.pig.piggybank.evaluation.datetime.convert.CustomFormatToISO(date,'dd/MM/yyyy') as dateISO;

DEFINE ISOToUnix org.apache.pig.piggybank.evaluation.datetime.convert.ISOToUnix();

ISOUnix = FOREACH convertToIso GENERATE ISOToUnix(dateISO) as dateUnix:long;

3. FILTER BY status code 404

combineLogLoader404 = FILTER combineLogLoaderData BY status == 404;

combineLogLoader404ED = FOREACH combineLogLoader404 GENERATE addr,status;

4. Group By IP

combineLogLoaderGroup = GROUP combineLogLoaderData BY addr;

combineLogLoaderGroupCount = FOREACH combineLogLoaderGroup GENERATE group, COUNT(combineLogLoaderData);

5. Group IP and status

combineLogLoaderGroupIS = GROUP combineLogLoaderData BY (addr,status);

combineLogLoaderGroupISCount = FOREACH combineLogLoaderGroupIS GENERATE group, COUNT(combineLogLoaderData);