package com.twitter.graph.batch.job.tweepcred

import com.twitter.scalding.\_

import com.twitter.scalding\_internal.job.\_

import com.twitter.scalding\_internal.job.analytics\_batch.\_

/\*\*

\* Register the beginning of the tweepcred job in analytic batch table

\*

\* Options:

\* --weighted: do weighted pagerank

\* --hadoop\_config: /etc/hadoop/hadoop-conf-proc-atla

\*

\*/

class TweepcredBatchJob(args: Args) extends AnalyticsIterativeBatchJob(args) {

def WEIGHTED = args("weighted").toBoolean

override def timeout = Hours(36)

override def hasFlow = false

def descriptionSuffix = " weighted=" + args("weighted")

override def batchIncrement = Hours(24)

override def firstTime = RichDate("2015-10-02")

override def batchDescription = classOf[TweepcredBatchJob].getCanonicalName + descriptionSuffix

override def run = {

val success = super.run

println("Batch Stat: " + messageHeader + " " + jobStat.get.toString)

success

}

def startTime = dateRange.start

def dateString = startTime.toString("yyyy/MM/dd")

override def children = {

val BASEDIR = "/user/cassowary/tweepcred/"

val baseDir = BASEDIR + (if (WEIGHTED) "weighted" else "unweighted") + "/daily/"

val tmpDir = baseDir + "tmp"

val outputDir = baseDir + dateString

val pageRankDir = outputDir + "/finalmass"

val tweepcredDir = outputDir + "/finaltweepcred"

val yesterdayStr = (startTime - Days(1)).toString("yyyy/MM/dd")

val yestPageRankDir = baseDir + yesterdayStr + "/finalmass"

val TWEEPCRED = "/tweepcred"

val curRep = (if (WEIGHTED) baseDir else BASEDIR) + "current"

val todayRep = (if (WEIGHTED) baseDir else BASEDIR) + dateString

val newArgs = args + ("pwd", Some(tmpDir)) +

("output\_pagerank", Some(pageRankDir)) +

("output\_tweepcred", Some(tweepcredDir)) +

("input\_pagerank", Some(yestPageRankDir)) +

("current\_tweepcred", Some(curRep + TWEEPCRED)) +

("today\_tweepcred", Some(todayRep + TWEEPCRED))

val prJob = new PreparePageRankData(newArgs)

List(prJob)

}

private def messageHeader = {

val dateString = dateRange.start.toString("yyyy/MM/dd")

classOf[TweepcredBatchJob].getSimpleName +

(if (WEIGHTED) " weighted " else " unweighted ") + dateString

}

}