package com.twitter.simclusters\_v2.hdfs\_sources

import com.twitter.scalding.DateOps

import com.twitter.scalding.DateRange

import com.twitter.scalding.Days

import com.twitter.scalding.TypedPipe

import com.twitter.scalding\_internal.dalv2.DAL

import com.twitter.scalding\_internal.dalv2.remote\_access.ExplicitLocation

import com.twitter.scalding\_internal.dalv2.remote\_access.ProcAtla

import com.twitter.simclusters\_v2.thriftscala.NormsAndCounts

import com.twitter.simclusters\_v2.thriftscala.UserAndNeighbors

import java.util.TimeZone

object DataSources {

/\*\*

\* Reads production normalized graph data from atla-proc

\*/

def userUserNormalizedGraphSource(implicit dateRange: DateRange): TypedPipe[UserAndNeighbors] = {

DAL

.readMostRecentSnapshotNoOlderThan(UserUserNormalizedGraphScalaDataset, Days(14)(DateOps.UTC))

.withRemoteReadPolicy(ExplicitLocation(ProcAtla))

.toTypedPipe

}

/\*\*

\* Reads production user norms and counts data from atla-proc

\*/

def userNormsAndCounts(

implicit dateRange: DateRange,

timeZone: TimeZone

): TypedPipe[NormsAndCounts] = {

DAL

.readMostRecentSnapshot(ProducerNormsAndCountsScalaDataset, dateRange.prepend(Days(14)))

.withRemoteReadPolicy(ExplicitLocation(ProcAtla))

.toTypedPipe

}

}