package com.twitter.simclusters\_v2.scalding.common

import com.twitter.finagle.mtls.authentication.ServiceIdentifier

import com.twitter.scalding.DateRange

import com.twitter.simclusters\_v2.common.Timestamp

import com.twitter.simclusters\_v2.common.TweetId

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

import com.twitter.strato.scalding.StratoManhattanExportSource

import com.twitter.strato.thrift.ScroogeConvImplicits.\_

object PersistentTweetEmbeddingSource {

// hdfs paths

val FavBasedUpdatedHdfsPath: String =

"/atla/proc/user/cassowary/manhattan-exporter/fav\_based\_tweet\_20m\_145k\_updated\_embeddings"

val LogFavBasedUpdatedHdfsPath: String =

"/atla/proc/user/cassowary/manhattan-exporter/log\_fav\_based\_tweet\_20m\_145k\_updated\_embeddings"

val LogFavBased2020HdfsPath: String =

"/atla/proc/user/cassowary/manhattan-exporter/log\_fav\_based\_tweet\_20m\_145k\_2020\_embeddings"

// Strato columns

val FavBasedUpdatedStratoColumn: String =

"recommendations/simclusters\_v2/embeddings/favBasedTweet20M145KUpdated"

val LogFavBasedUpdatedStratoColumn: String =

"recommendations/simclusters\_v2/embeddings/logFavBasedTweet20M145KUpdatedPersistent"

val LogFavBased2020StratoColumn: String =

"recommendations/simclusters\_v2/embeddings/logFavBasedTweet20M145K2020Persistent"

}

/\*\*

\* The source that read the Manhattan export persistent embeddings

\*/

// Defaults to Updated version.

class FavBasedPersistentTweetEmbeddingMhExportSource(

hdfsPath: String = PersistentTweetEmbeddingSource.FavBasedUpdatedHdfsPath,

stratoColumnPath: String = PersistentTweetEmbeddingSource.FavBasedUpdatedStratoColumn,

range: DateRange,

serviceIdentifier: ServiceIdentifier = ServiceIdentifier.empty)

extends StratoManhattanExportSource[(TweetId, Timestamp), PersistentSimClustersEmbedding](

hdfsPath,

range,

stratoColumnPath,

serviceIdentifier = serviceIdentifier

)

// Defaults to 2020 version.

class LogFavBasedPersistentTweetEmbeddingMhExportSource(

hdfsPath: String = PersistentTweetEmbeddingSource.LogFavBased2020HdfsPath,

stratoColumnPath: String = PersistentTweetEmbeddingSource.LogFavBased2020StratoColumn,

range: DateRange,

serviceIdentifier: ServiceIdentifier = ServiceIdentifier.empty)

extends StratoManhattanExportSource[(TweetId, Timestamp), PersistentSimClustersEmbedding](

hdfsPath,

range,

stratoColumnPath,

serviceIdentifier = serviceIdentifier

)