package com.twitter.simclusters\_v2.scio

package multi\_type\_graph.multi\_type\_graph\_sims

import com.twitter.dal.client.dataset.KeyValDALDataset

import com.twitter.scalding\_internal.multiformat.format.keyval.KeyVal

import com.twitter.simclusters\_v2.hdfs\_sources.RightNodeCosineSimilarityScioScalaDataset

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

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

import com.twitter.wtf.scalding.jobs.cosine\_similarity.common.ApproximateMatrixSelfTransposeMultiplicationJob

/\*\*

Build:

./bazel bundle src/scala/com/twitter/simclusters\_v2/scio/multi\_type\_graph/multi\_type\_graph\_sims:multi-type-graph-cosine-similarity-scio-adhoc-app

To kick off an adhoc run:

bin/d6w create \

${GCP\_PROJECT\_NAME}/us-central1/multi-type-graph-cosine-similarity-scio-adhoc-app \

src/scala/com/twitter/simclusters\_v2/scio/multi\_type\_graph/multi\_type\_graph\_sims/cosine-similarity-scio-adhoc.d6w \

--jar dist/multi-type-graph-cosine-similarity-scio-adhoc-app.jar \

--bind=profile.project=${GCP\_PROJECT\_NAME} \

--bind=profile.user\_name=${USER} \

--bind=profile.date="2022-01-16" \

--bind=profile.machine="n2d-highmem-16" --ignore-existing

\*/

object RightNodeCosineSimilarityScioAdhocApp extends RightNodeCosineSimilarityScioBaseApp {

override val isAdhoc = true

override val cosineSimKeyValSnapshotDataset: KeyValDALDataset[

KeyVal[RightNode, SimilarRightNodes]

] =

RightNodeCosineSimilarityScioAdhocScalaDataset

override val filterCandidateSimilarityPair: (Double, Double, Double) => Boolean =

ApproximateMatrixSelfTransposeMultiplicationJob.filterCandidateSimilarityPair

}

/\*\*

To deploy the job:

bin/d6w schedule \

${GCP\_PROJECT\_NAME}/us-central1/multi-type-graph-cosine-similarity-scio-batch-app \

src/scala/com/twitter/simclusters\_v2/scio/multi\_type\_graph/multi\_type\_graph\_sims/cosine-similarity-scio-batch.d6w \

--bind=profile.project=${GCP\_PROJECT\_NAME} \

--bind=profile.user\_name=recos-platform \

--bind=profile.date="2021-12-01" \

--bind=profile.machine="n2d-highmem-16"

\*/

object RightNodeCosineSimilarityScioBatchApp extends RightNodeCosineSimilarityScioBaseApp {

override val isAdhoc = false

override val cosineSimKeyValSnapshotDataset: KeyValDALDataset[

KeyVal[RightNode, SimilarRightNodes]

] =

RightNodeCosineSimilarityScioScalaDataset

override val filterCandidateSimilarityPair: (Double, Double, Double) => Boolean =

ApproximateMatrixSelfTransposeMultiplicationJob.filterCandidateSimilarityPair

}