scala\_library(

sources = [

"\*.scala",

"common/\*.scala",

],

compiler\_option\_sets = ["fatal\_warnings"],

platform = "java8",

tags = [

"bazel-compatible",

"bazel-only",

],

dependencies = [

"src/scala/com/twitter/simclusters\_v2/common/clustering",

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources:aggregatable\_producer\_simclusters\_embeddings\_by\_log\_fav\_score-scala",

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources:clusters\_members\_connected\_components\_ape\_similarity-scala",

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources:clusters\_members\_largest\_dim\_ape\_similarity-scala",

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources:clusters\_members\_largest\_dim\_ape\_similarity\_2\_day\_update-scala",

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources:clusters\_members\_louvain\_ape\_similarity-scala",

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources:interested\_in\_twice\_by\_largest\_dim-scala",

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources:interested\_in\_twice\_by\_largest\_dim\_2\_day\_update-scala",

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources:interested\_in\_twice\_by\_largest\_dim\_fav\_score-scala",

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources:interested\_in\_twice\_connected\_components-scala",

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources:interested\_in\_twice\_louvain-scala",

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources:user\_user\_normalized\_graph-scala",

"src/scala/com/twitter/simclusters\_v2/scalding/common",

"src/scala/com/twitter/simclusters\_v2/scalding/embedding",

"src/scala/com/twitter/wtf/scalding/jobs/common:execution\_app",

"src/scala/com/twitter/wtf/scalding/jobs/common:sources",

"src/scala/com/twitter/wtf/scalding/jobs/common:stats\_util",

],

)

# ========================

# ADHOC JOB CONFIGURATIONS

# Note: Please change mapreduce.job.reduces and --num-reducers together.

# ========================

scalding\_job(

name = "interested\_in\_twice\_largest\_dim-adhoc",

main = "com.twitter.simclusters\_v2.scalding.embedding.twice.InterestedInTwiceLargestDimAdhocApp",

args = [

"--date 2021-08-31",

"--num-reducers 4000",

],

config = [

("hadoop.combine-input", "true"),

("hadoop.map.jvm.total-memory", "3072m"),

("hadoop.reduce.jvm.total-memory", "3072m"),

("hadoop.submitter.jvm.total-memory", "5120m"),

("submitter.tier", "preemptible"),

],

hadoop\_cluster = "qus1-bluebird",

hadoop\_properties = [

("mapreduce.job.reduce.slowstart.completedmaps", "1.0"),

("scalding.with.reducers.set.explicitly", "true"),

("mapreduce.job.reduces", "4000"),

("mapreduce.task.timeout", "0"),

],

platform = "java8",

role = "cassowary",

runtime\_platform = "java8",

tags = [

"bazel-compatible:migrated",

"bazel-only",

],

dependencies = [":twice"],

)

scalding\_job(

name = "interested\_in\_twice\_largest\_dim\_fav\_score-adhoc",

main = "com.twitter.simclusters\_v2.scalding.embedding.twice.InterestedInTwiceLargestDimMaxFavScoreAdhocApp",

args = [

"--date 2022-07-01",

"--num-reducers 4000",

],

config = [

("hadoop.combine-input", "true"),

("hadoop.map.jvm.total-memory", "3072m"),

("hadoop.reduce.jvm.total-memory", "3072m"),

("hadoop.submitter.jvm.total-memory", "5120m"),

("submitter.tier", "preemptible"),

],

hadoop\_cluster = "qus1-bluebird",

hadoop\_properties = [

("mapreduce.job.reduce.slowstart.completedmaps", "1.0"),

("scalding.with.reducers.set.explicitly", "true"),

("mapreduce.job.reduces", "4000"),

("mapreduce.task.timeout", "0"),

],

platform = "java8",

role = "cassowary",

runtime\_platform = "java8",

tags = [

"bazel-compatible:migrated",

"bazel-only",

],

dependencies = [":twice"],

)

scalding\_job(

name = "interested\_in\_twice\_louvain-adhoc",

main = "com.twitter.simclusters\_v2.scalding.embedding.twice.InterestedInTwiceLouvainAdhocApp",

args = [

"--date 2021-08-31",

"--num-reducers 4000",

"--cosine\_similarity\_threshold 0.5",

],

config = [

("hadoop.combine-input", "true"),

("hadoop.map.jvm.total-memory", "3072m"),

("hadoop.reduce.jvm.total-memory", "3072m"),

("hadoop.submitter.jvm.total-memory", "5120m"),

("submitter.tier", "preemptible"),

],

hadoop\_cluster = "qus1-bluebird",

hadoop\_properties = [

("mapreduce.job.reduce.slowstart.completedmaps", "1.0"),

("scalding.with.reducers.set.explicitly", "true"),

("mapreduce.job.reduces", "4000"),

("mapreduce.task.timeout", "0"),

],

platform = "java8",

role = "cassowary",

runtime\_platform = "java8",

tags = [

"bazel-compatible:migrated",

"bazel-only",

],

dependencies = [":twice"],

)

scalding\_job(

name = "interested\_in\_twice\_connected\_components-adhoc",

main = "com.twitter.simclusters\_v2.scalding.embedding.twice.InterestedInTwiceConnectedComponentsAdhocApp",

args = [

"--date 2021-08-31",

"--num-reducers 4000",

"--cosine\_similarity\_threshold 0.5",

],

config = [

("hadoop.combine-input", "true"),

("hadoop.map.jvm.total-memory", "3072m"),

("hadoop.reduce.jvm.total-memory", "3072m"),

("hadoop.submitter.jvm.total-memory", "5120m"),

("submitter.tier", "preemptible"),

],

hadoop\_cluster = "qus1-bluebird",

hadoop\_properties = [

("mapreduce.job.reduce.slowstart.completedmaps", "1.0"),

("scalding.with.reducers.set.explicitly", "true"),

("mapreduce.job.reduces", "4000"),

("mapreduce.task.timeout", "0"),

],

platform = "java8",

role = "cassowary",

runtime\_platform = "java8",

tags = [

"bazel-compatible:migrated",

"bazel-only",

],

dependencies = [":twice"],

)

# ============================

# SCHEDULED JOB CONFIGURATIONS

# Twice jobs have been descheduled

# ============================