scala\_library(

sources = ["\*.scala"],

platform = "java8",

tags = ["bazel-compatible"],

dependencies = [

":data\_sources",

"3rdparty/src/jvm/com/twitter/scalding:core",

"src/scala/com/twitter/scalding\_internal/dalv2",

"src/scala/com/twitter/scalding\_internal/multiformat/format",

"src/scala/com/twitter/scalding\_internal/source/lzo\_scrooge",

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

"src/thrift/com/twitter/hermit/candidate:hermit-candidate-scala",

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-scala",

"src/thrift/com/twitter/wtf/entity\_real\_graph:entity\_real\_graph-thrift-scala",

],

)

scala\_library(

name = "data\_sources",

sources = [],

description = "DAL datasets we wish to expose externally",

platform = "java8",

tags = ["bazel-compatible"],

dependencies = [

":ads\_fav\_based\_simclusters\_cluster\_to\_tweet\_index-scala",

":ads\_fav\_click\_based\_simclusters\_cluster\_to\_tweet\_index-scala",

":aggregatable\_producer\_simclusters\_embeddings\_by\_fav\_score-scala",

":aggregatable\_producer\_simclusters\_embeddings\_by\_fav\_score\_2020-scala",

":aggregatable\_producer\_simclusters\_embeddings\_by\_fav\_score\_2020\_thrift-scala",

":aggregatable\_producer\_simclusters\_embeddings\_by\_fav\_score\_thrift-scala",

":aggregatable\_producer\_simclusters\_embeddings\_by\_follow\_score\_2020-scala",

":aggregatable\_producer\_simclusters\_embeddings\_by\_follow\_score\_2020\_thrift-scala",

":aggregatable\_producer\_simclusters\_embeddings\_by\_log\_fav\_score-scala",

":aggregatable\_producer\_simclusters\_embeddings\_by\_log\_fav\_score\_2020-scala",

":aggregatable\_producer\_simclusters\_embeddings\_by\_log\_fav\_score\_2020\_thrift-scala",

":aggregatable\_producer\_simclusters\_embeddings\_by\_log\_fav\_score\_relaxed\_fav\_engagement\_threshold\_2020-scala",

":aggregatable\_producer\_simclusters\_embeddings\_by\_log\_fav\_score\_relaxed\_fav\_engagement\_threshold\_2020\_thrift-scala",

":aggregatable\_producer\_simclusters\_embeddings\_by\_log\_fav\_score\_thrift-scala",

":clusters\_members\_connected\_components\_ape\_similarity-scala",

":clusters\_members\_largest\_dim\_ape\_similarity-scala",

":clusters\_members\_largest\_dim\_ape\_similarity\_2\_day\_update-scala",

":clusters\_members\_louvain\_ape\_similarity-scala",

":co\_engagement\_top\_k\_similar\_tweets-scala",

":explore\_mbcg\_user\_embeddings\_kv-scala",

":fav\_based\_evergreen\_content\_simclusters\_cluster\_to\_tweet\_index-scala",

":fav\_based\_simclusters\_cluster\_to\_tweet\_index-scala",

":fav\_based\_video\_simclusters\_cluster\_to\_tweet\_index-scala",

":fav\_inferred\_language\_tfg\_topic\_embeddings-scala",

":fav\_tfg\_topic\_embeddings-scala",

":fav\_tfg\_topic\_embeddings\_2020-scala",

":fav\_tfg\_topic\_embeddings\_2020\_parquet-scala",

":fav\_tfg\_topic\_embeddings\_parquet-scala",

":full\_multi\_type\_graph-scala",

":geopopular\_top\_tweet\_impressed\_topics-scala",

":hashtag\_simclusters\_embeddings\_updated-scala",

":interested\_in\_twice\_by\_largest\_dim-scala",

":interested\_in\_twice\_by\_largest\_dim\_2\_day\_update-scala",

":interested\_in\_twice\_by\_largest\_dim\_fav\_score-scala",

":interested\_in\_twice\_connected\_components-scala",

":interested\_in\_twice\_louvain-scala",

":log\_fav\_reverse\_index\_semantic\_core\_per\_language\_simclusters\_embeddings-scala",

":log\_fav\_semantic\_core\_per\_language\_simclusters\_embeddings-scala",

":log\_fav\_tfg\_topic\_embeddings-scala",

":log\_fav\_tfg\_topic\_embeddings\_parquet-scala",

":multi\_type\_graph\_for\_top\_k\_right\_nodes\_thrift\_50\_m\_scio-scala",

":multi\_type\_graph\_for\_top\_k\_right\_nodes\_thrift\_scio-scala",

":multi\_type\_simclusters\_right\_node\_to\_clusters\_thrift\_50\_m-scala",

":multi\_type\_simclusters\_right\_node\_to\_clusters\_thrift\_fav\_90\_p\_20\_m-scala",

":offline\_cluster\_top\_media\_tweets\_20M\_145K\_2020-scala",

":offline\_tweet\_recommendations\_from\_interested\_in\_20M\_145K\_2020-scala",

":offline\_tweet\_recommendations\_from\_interested\_in\_20M\_145K\_2020\_HL\_0\_EL\_15-scala",

":offline\_tweet\_recommendations\_from\_interested\_in\_20M\_145K\_2020\_HL\_2\_EL\_15-scala",

":offline\_tweet\_recommendations\_from\_interested\_in\_20M\_145K\_2020\_HL\_2\_EL\_50-scala",

":offline\_tweet\_recommendations\_from\_interested\_in\_20M\_145K\_2020\_HL\_8\_EL\_50-scala",

":offline\_tweet\_recommendations\_from\_mts\_consumer\_embeddings-scala",

":producer\_norms\_and\_counts-scala",

":producer\_top\_k\_simcluster\_embeddings\_by\_fav\_score-scala",

":producer\_top\_k\_simcluster\_embeddings\_by\_fav\_score\_2020-scala",

":producer\_top\_k\_simcluster\_embeddings\_by\_fav\_score\_updated-scala",

":producer\_top\_k\_simcluster\_embeddings\_by\_follow\_score-scala",

":producer\_top\_k\_simcluster\_embeddings\_by\_follow\_score\_2020-scala",

":producer\_top\_k\_simcluster\_embeddings\_by\_follow\_score\_updated-scala",

":push\_open\_based\_simclusters\_cluster\_to\_tweet\_index-scala",

":reply\_based\_simclusters\_cluster\_to\_tweet\_index-scala",

":retweet\_based\_simclusters\_cluster\_to\_tweet\_index-scala",

":reverse\_index\_hashtag\_simclusters\_embeddings\_updated-scala",

":reverse\_index\_semantic\_core\_per\_language\_simclusters\_embeddings-scala",

":reverse\_index\_semantic\_core\_simclusters\_embeddings-scala",

":reverse\_index\_semantic\_core\_simclusters\_embeddings\_2020-scala",

":reverse\_index\_semantic\_core\_simclusters\_embeddings\_updated-scala",

":right\_node\_cosine\_similarity\_scio-scala",

":right\_node\_sim\_hash\_scio-scala",

":rux\_faved\_top\_k\_tweets-scala",

":semantic\_core\_embeddings\_from\_producer-scala",

":semantic\_core\_per\_language\_simclusters\_embeddings-scala",

":semantic\_core\_simclusters\_embeddings-scala",

":semantic\_core\_simclusters\_embeddings\_2020-scala",

":semantic\_core\_simclusters\_embeddings\_updated-scala",

":simcluster\_embedding\_top\_k\_producers\_by\_fav\_score-scala",

":simcluster\_embedding\_top\_k\_producers\_by\_fav\_score\_2020-scala",

":simcluster\_embedding\_top\_k\_producers\_by\_fav\_score\_updated-scala",

":simcluster\_embedding\_top\_k\_producers\_by\_follow\_score-scala",

":simcluster\_embedding\_top\_k\_producers\_by\_follow\_score\_2020-scala",

":simcluster\_embedding\_top\_k\_producers\_by\_follow\_score\_updated-scala",

":simclusters\_inferred\_entities\_from\_interested\_in-scala",

":simclusters\_inferred\_entities\_from\_interested\_in\_keyed\_by\_cluster-scala",

":simclusters\_inferred\_entities\_from\_known\_for-scala",

":simclusters\_offline\_cluster\_top\_k\_tweets-scala",

":simclusters\_offline\_tweet\_cluster\_scores-scala",

":simclusters\_offline\_tweet\_top\_k\_clusters-scala",

":simclusters\_v2\_cluster\_details-scala",

":simclusters\_v2\_cluster\_details\_20m\_145k\_2020-scala",

":simclusters\_v2\_cluster\_details\_20m\_145k\_updated-scala",

":simclusters\_v2\_cluster\_details\_lite-scala",

":simclusters\_v2\_cluster\_details\_lite\_20m\_145k\_2020-scala",

":simclusters\_v2\_cluster\_details\_lite\_20m\_145k\_updated-scala",

":simclusters\_v2\_embeddings\_lite-scala",

":simclusters\_v2\_global\_language\_embedding-scala",

":simclusters\_v2\_global\_language\_embedding\_thrift-scala",

":simclusters\_v2\_interested\_in-scala",

":simclusters\_v2\_interested\_in\_20M\_145K\_2020-scala",

":simclusters\_v2\_interested\_in\_20M\_145K\_updated-scala",

":simclusters\_v2\_interested\_in\_from\_aggregatable\_producer\_embeddings\_20M\_145K\_2020-scala",

":simclusters\_v2\_interested\_in\_from\_producer\_embeddings\_20M\_145K\_updated-scala",

":simclusters\_v2\_interested\_in\_lite\_20M\_145K\_2020-scala",

":simclusters\_v2\_known\_for\_20M\_145K\_2020-scala",

":simclusters\_v2\_known\_for\_20M\_145K\_2020\_thrift-scala",

":simclusters\_v2\_known\_for\_20M\_145K\_dec11-scala",

":simclusters\_v2\_known\_for\_20M\_145K\_updated-scala",

":simclusters\_v2\_known\_for\_20M\_145K\_updated\_thrift-scala",

":simclusters\_v2\_raw\_interested\_in\_20M\_145K\_2020-scala",

":simclusters\_v2\_raw\_interested\_in\_20M\_145K\_dec11-scala",

":simclusters\_v2\_raw\_interested\_in\_20M\_145K\_updated-scala",

":simclusters\_v2\_raw\_interested\_in\_lite\_20M\_145K\_2020-scala",

":simclusters\_v2\_raw\_known\_for\_20M\_145K\_2020-scala",

":simclusters\_v2\_raw\_known\_for\_20M\_145K\_dec11-scala",

":simclusters\_v2\_raw\_known\_for\_20M\_145K\_updated-scala",

":simclusters\_v2\_user\_to\_interested\_in\_20M\_145K\_2020-scala",

":simclusters\_v2\_user\_to\_interested\_in\_20M\_145K\_dec11-scala",

":simclusters\_v2\_user\_to\_interested\_in\_20M\_145K\_updated-scala",

":simclusters\_v2\_user\_to\_interested\_in\_from\_aggregatable\_producer\_embeddings\_20M\_145K\_2020-scala",

":simclusters\_v2\_user\_to\_interested\_in\_lite\_20M\_145K\_2020-scala",

":similar\_topics\_from\_topic\_follow\_graph-scala",

":similar\_users\_by\_fav\_based\_producer\_embedding-scala",

":similar\_users\_by\_follow\_based\_producer\_embedding-scala",

":top\_k\_right\_nouns-scala",

":top\_k\_right\_nouns\_scio-scala",

":top\_locale\_topics\_for\_producer\_from\_em-scala",

":top\_producers\_for\_locale\_topics\_from\_topic\_follow\_graph-scala",

":topic\_top\_producers\_em-scala",

":truncated\_multi\_type\_graph-scala",

":truncated\_multi\_type\_graph\_scio-scala",

":tweet\_evaluation\_timelines\_reference\_set-scala",

":user\_topic\_weighted\_embedding-scala",

":user\_topic\_weighted\_embedding\_parquet-scala",

":user\_user\_fav\_graph-scala",

":user\_user\_graph-scala",

":user\_user\_normalized\_graph-scala",

":video\_view\_based\_simclusters\_cluster\_to\_tweet\_index-scala",

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

],

)

create\_datasets(

base\_name = "user\_user\_fav\_graph",

java\_schema = "com.twitter.simclusters\_v2.thriftjava.EdgeWithDecayedWeights",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.thriftscala.EdgeWithDecayedWeights",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

java\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-java",

],

scala\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-scala",

],

)

create\_datasets(

base\_name = "producer\_norms\_and\_counts",

java\_schema = "com.twitter.simclusters\_v2.thriftjava.NormsAndCounts",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.thriftscala.NormsAndCounts",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

java\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-java",

],

scala\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-scala",

],

)

create\_datasets(

base\_name = "user\_user\_normalized\_graph",

java\_schema = "com.twitter.simclusters\_v2.thriftjava.UserAndNeighbors",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.thriftscala.UserAndNeighbors",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

java\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-java",

],

scala\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-scala",

],

)

create\_datasets(

base\_name = "multi\_type\_simclusters\_right\_node\_to\_clusters\_thrift\_fav\_90\_p\_20\_m",

java\_schema = "com.twitter.simclusters\_v2.thriftjava.RightNodeWithClusters",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.thriftscala.RightNodeWithClusters",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

java\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-java",

],

scala\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-scala",

],

)

create\_datasets(

base\_name = "multi\_type\_simclusters\_right\_node\_to\_clusters\_thrift\_50\_m",

java\_schema = "com.twitter.simclusters\_v2.thriftjava.RightNodeWithClusters",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.thriftscala.RightNodeWithClusters",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

java\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-java",

],

scala\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-scala",

],

)

create\_datasets(

base\_name = "user\_user\_graph",

java\_schema = "com.twitter.simclusters\_v2.thriftjava.UserAndNeighbors",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.thriftscala.UserAndNeighbors",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

java\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-java",

],

scala\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-scala",

],

)

# InterestedIn

create\_datasets(

base\_name = "simclusters\_v2\_raw\_interested\_in\_20M\_145K\_dec11",

key\_type = "Long",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.InterestedInInjection.injection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.ClustersUserIsInterestedIn",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "simclusters\_v2\_raw\_interested\_in\_20M\_145K\_updated",

key\_type = "Long",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.InterestedInInjection.injection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.ClustersUserIsInterestedIn",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "simclusters\_v2\_raw\_interested\_in\_20M\_145K\_2020",

key\_type = "Long",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.InterestedInInjection.injection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.ClustersUserIsInterestedIn",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "simclusters\_v2\_raw\_interested\_in\_lite\_20M\_145K\_2020",

key\_type = "Long",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.InterestedInInjection.injection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.ClustersUserIsInterestedIn",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "multi\_type\_graph\_for\_top\_k\_right\_nodes\_thrift\_fav\_90\_p\_20\_m\_scio",

java\_schema = "com.twitter.simclusters\_v2.thriftjava.MultiTypeGraphEdge",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.thriftscala.MultiTypeGraphEdge",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

java\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-java",

],

scala\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-scala",

],

)

create\_datasets(

base\_name = "multi\_type\_graph\_for\_top\_k\_right\_nodes\_thrift\_50\_m\_scio",

java\_schema = "com.twitter.simclusters\_v2.thriftjava.MultiTypeGraphEdge",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.thriftscala.MultiTypeGraphEdge",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

java\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-java",

],

scala\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-scala",

],

)

create\_datasets(

base\_name = "simclusters\_v2\_interested\_in",

key\_type = "Long",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.InterestedInInjection.injection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.ClustersUserIsInterestedIn",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "simclusters\_v2\_interested\_in\_20M\_145K\_updated",

key\_type = "Long",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.InterestedInInjection.injection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.ClustersUserIsInterestedIn",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "simclusters\_v2\_interested\_in\_20M\_145K\_2020",

key\_type = "Long",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.InterestedInInjection.injection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.ClustersUserIsInterestedIn",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "simclusters\_v2\_interested\_in\_lite\_20M\_145K\_2020",

key\_type = "Long",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.InterestedInInjection.injection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.ClustersUserIsInterestedIn",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "simclusters\_v2\_user\_to\_interested\_in\_20M\_145K\_dec11",

java\_schema = "com.twitter.simclusters\_v2.thriftjava.UserToInterestedInClusters",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.thriftscala.UserToInterestedInClusters",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

java\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-java",

],

scala\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-scala",

],

)

create\_datasets(

base\_name = "simclusters\_v2\_user\_to\_interested\_in\_20M\_145K\_updated",

java\_schema = "com.twitter.simclusters\_v2.thriftjava.UserToInterestedInClusters",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.thriftscala.UserToInterestedInClusters",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

java\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-java",

],

scala\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-scala",

],

)

create\_datasets(

base\_name = "simclusters\_v2\_user\_to\_interested\_in\_20M\_145K\_2020",

java\_schema = "com.twitter.simclusters\_v2.thriftjava.UserToInterestedInClusters",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.thriftscala.UserToInterestedInClusters",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

java\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-java",

],

scala\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-scala",

],

)

create\_datasets(

base\_name = "simclusters\_v2\_user\_to\_interested\_in\_lite\_20M\_145K\_2020",

java\_schema = "com.twitter.simclusters\_v2.thriftjava.UserToInterestedInClusters",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.thriftscala.UserToInterestedInClusters",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

java\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-java",

],

scala\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-scala",

],

)

create\_datasets(

base\_name = "simclusters\_v2\_user\_to\_interested\_in\_from\_aggregatable\_producer\_embeddings\_20M\_145K\_2020",

java\_schema = "com.twitter.simclusters\_v2.thriftjava.UserToInterestedInClusters",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.thriftscala.UserToInterestedInClusters",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

java\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-java",

],

scala\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-scala",

],

)

# end of InterestedIn

# KnownFor

create\_datasets(

base\_name = "simclusters\_v2\_raw\_known\_for\_20M\_145K\_dec11",

key\_type = "Long",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.KnownForInjection.injection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.ClustersUserIsKnownFor",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "simclusters\_v2\_raw\_known\_for\_20M\_145K\_updated",

key\_type = "Long",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.KnownForInjection.injection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.ClustersUserIsKnownFor",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "simclusters\_v2\_raw\_known\_for\_20M\_145K\_2020",

key\_type = "Long",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.KnownForInjection.injection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.ClustersUserIsKnownFor",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "simclusters\_v2\_known\_for\_20M\_145K\_dec11",

key\_type = "Long",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.KnownForInjection.injection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.ClustersUserIsKnownFor",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "simclusters\_v2\_known\_for\_20M\_145K\_updated",

key\_type = "Long",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.KnownForInjection.injection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.ClustersUserIsKnownFor",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "simclusters\_v2\_known\_for\_20M\_145K\_updated\_thrift",

java\_schema = "com.twitter.simclusters\_v2.thriftjava.UserToKnownForClusters",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.thriftscala.UserToKnownForClusters",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

java\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-java",

],

scala\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-scala",

],

)

create\_datasets(

base\_name = "simclusters\_v2\_known\_for\_20M\_145K\_2020",

key\_type = "Long",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.KnownForInjection.injection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.ClustersUserIsKnownFor",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "simclusters\_v2\_known\_for\_20M\_145K\_2020\_thrift",

java\_schema = "com.twitter.simclusters\_v2.thriftjava.UserToKnownForClusters",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.thriftscala.UserToKnownForClusters",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

java\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-java",

],

scala\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-scala",

],

)

# end of KnownFor

create\_datasets(

base\_name = "simclusters\_v2\_cluster\_details",

key\_type = "scala.Tuple2[String, Int]",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.ClusterDetailsInjection.injection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.ClusterDetails",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "simclusters\_v2\_cluster\_details\_lite",

java\_schema = "com.twitter.simclusters\_v2.thriftjava.ClusterDetailsLite",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.thriftscala.ClusterDetailsLite",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

java\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-java",

],

scala\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-scala",

],

)

create\_datasets(

base\_name = "simclusters\_v2\_embeddings\_lite",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.thriftscala.EmbeddingsLite",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "simclusters\_v2\_cluster\_details\_20m\_145k\_updated",

key\_type = "scala.Tuple2[String, Int]",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.ClusterDetailsInjection.injection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.ClusterDetails",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "simclusters\_v2\_cluster\_details\_lite\_20m\_145k\_updated",

java\_schema = "com.twitter.simclusters\_v2.thriftjava.ClusterDetailsLite",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.thriftscala.ClusterDetailsLite",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

java\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-java",

],

scala\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-scala",

],

)

create\_datasets(

base\_name = "simclusters\_v2\_cluster\_details\_20m\_145k\_2020",

key\_type = "scala.Tuple2[String, Int]",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.ClusterDetailsInjection.injection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.ClusterDetails",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "simclusters\_v2\_cluster\_details\_lite\_20m\_145k\_2020",

java\_schema = "com.twitter.simclusters\_v2.thriftjava.ClusterDetailsLite",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.thriftscala.ClusterDetailsLite",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

java\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-java",

],

scala\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-scala",

],

)

create\_datasets(

base\_name = "tweet\_evaluation\_timelines\_reference\_set",

description = "A Tweet dataset that contains impressed tweets with engagement labels, parsed from Timelines",

java\_schema = "com.twitter.simclusters\_v2.thriftjava.ReferenceTweets",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.thriftscala.ReferenceTweets",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

java\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-java",

],

scala\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-scala",

],

)

create\_datasets(

base\_name = "semantic\_core\_simclusters\_embeddings",

key\_type = "com.twitter.simclusters\_v2.thriftscala.SimClustersEmbeddingId",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.EntityEmbeddingsInjections.EntitySimClustersEmbeddingInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.SimClustersEmbedding",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "semantic\_core\_simclusters\_embeddings\_updated",

key\_type = "com.twitter.simclusters\_v2.thriftscala.SimClustersEmbeddingId",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.EntityEmbeddingsInjections.EntitySimClustersEmbeddingInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.SimClustersEmbedding",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "semantic\_core\_simclusters\_embeddings\_2020",

key\_type = "com.twitter.simclusters\_v2.thriftscala.SimClustersEmbeddingId",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.EntityEmbeddingsInjections.EntitySimClustersEmbeddingInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.SimClustersEmbedding",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "semantic\_core\_per\_language\_simclusters\_embeddings",

key\_type = "com.twitter.simclusters\_v2.thriftscala.SimClustersEmbeddingId",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.EntityEmbeddingsInjections.EntitySimClustersEmbeddingInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.SimClustersEmbedding",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "log\_fav\_semantic\_core\_per\_language\_simclusters\_embeddings",

key\_type = "com.twitter.simclusters\_v2.thriftscala.SimClustersEmbeddingId",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.EntityEmbeddingsInjections.EntitySimClustersEmbeddingInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.SimClustersEmbedding",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "reverse\_index\_semantic\_core\_simclusters\_embeddings",

key\_type = "com.twitter.simclusters\_v2.thriftscala.SimClustersEmbeddingId",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.EntityEmbeddingsInjections.InternalIdEmbeddingInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.InternalIdEmbedding",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "reverse\_index\_semantic\_core\_simclusters\_embeddings\_updated",

key\_type = "com.twitter.simclusters\_v2.thriftscala.SimClustersEmbeddingId",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.EntityEmbeddingsInjections.InternalIdEmbeddingInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.InternalIdEmbedding",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "reverse\_index\_semantic\_core\_simclusters\_embeddings\_2020",

key\_type = "com.twitter.simclusters\_v2.thriftscala.SimClustersEmbeddingId",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.EntityEmbeddingsInjections.InternalIdEmbeddingInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.InternalIdEmbedding",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "reverse\_index\_semantic\_core\_per\_language\_simclusters\_embeddings",

key\_type = "com.twitter.simclusters\_v2.thriftscala.SimClustersEmbeddingId",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.EntityEmbeddingsInjections.InternalIdEmbeddingInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.InternalIdEmbedding",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "log\_fav\_reverse\_index\_semantic\_core\_per\_language\_simclusters\_embeddings",

key\_type = "com.twitter.simclusters\_v2.thriftscala.SimClustersEmbeddingId",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.EntityEmbeddingsInjections.InternalIdEmbeddingInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.InternalIdEmbedding",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "hashtag\_simclusters\_embeddings\_updated",

key\_type = "com.twitter.simclusters\_v2.thriftscala.SimClustersEmbeddingId",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.EntityEmbeddingsInjections.EntitySimClustersEmbeddingInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.SimClustersEmbedding",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "fav\_tfg\_topic\_embeddings",

key\_type = "com.twitter.simclusters\_v2.thriftscala.SimClustersEmbeddingId",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.EntityEmbeddingsInjections.EntitySimClustersEmbeddingInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.SimClustersEmbedding",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "fav\_tfg\_topic\_embeddings\_2020",

key\_type = "com.twitter.simclusters\_v2.thriftscala.SimClustersEmbeddingId",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.EntityEmbeddingsInjections.EntitySimClustersEmbeddingInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.SimClustersEmbedding",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "fav\_tfg\_topic\_embeddings\_parquet",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.thriftscala.TfgTopicEmbeddings",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

scala\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-scala",

],

)

create\_datasets(

base\_name = "fav\_tfg\_topic\_embeddings\_2020\_parquet",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.thriftscala.TfgTopicEmbeddings",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

scala\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-scala",

],

)

create\_datasets(

base\_name = "fav\_inferred\_language\_tfg\_topic\_embeddings",

key\_type = "com.twitter.simclusters\_v2.thriftscala.SimClustersEmbeddingId",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.EntityEmbeddingsInjections.EntitySimClustersEmbeddingInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.SimClustersEmbedding",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "log\_fav\_tfg\_topic\_embeddings",

key\_type = "com.twitter.simclusters\_v2.thriftscala.SimClustersEmbeddingId",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.EntityEmbeddingsInjections.EntitySimClustersEmbeddingInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.SimClustersEmbedding",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "log\_fav\_tfg\_topic\_embeddings\_parquet",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.thriftscala.TfgTopicEmbeddings",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

scala\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-scala",

],

)

create\_datasets(

base\_name = "reverse\_index\_hashtag\_simclusters\_embeddings\_updated",

key\_type = "com.twitter.simclusters\_v2.thriftscala.SimClustersEmbeddingId",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.EntityEmbeddingsInjections.InternalIdEmbeddingInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.InternalIdEmbedding",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "simcluster\_embedding\_top\_k\_producers\_by\_fav\_score",

key\_type = "com.twitter.simclusters\_v2.thriftscala.PersistedFullClusterId",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.ProducerEmbeddingsInjections.SimClusterEmbeddingTopKProducersInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.TopProducersWithScore",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "simcluster\_embedding\_top\_k\_producers\_by\_fav\_score\_updated",

key\_type = "com.twitter.simclusters\_v2.thriftscala.PersistedFullClusterId",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.ProducerEmbeddingsInjections.SimClusterEmbeddingTopKProducersInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.TopProducersWithScore",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "simcluster\_embedding\_top\_k\_producers\_by\_fav\_score\_2020",

key\_type = "com.twitter.simclusters\_v2.thriftscala.PersistedFullClusterId",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.ProducerEmbeddingsInjections.SimClusterEmbeddingTopKProducersInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.TopProducersWithScore",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "producer\_top\_k\_simcluster\_embeddings\_by\_fav\_score",

key\_type = "Long",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.ProducerEmbeddingsInjections.ProducerTopKSimClusterEmbeddingsInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.TopSimClustersWithScore",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "producer\_top\_k\_simcluster\_embeddings\_by\_fav\_score\_updated",

key\_type = "Long",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.ProducerEmbeddingsInjections.ProducerTopKSimClusterEmbeddingsInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.TopSimClustersWithScore",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "producer\_top\_k\_simcluster\_embeddings\_by\_fav\_score\_2020",

key\_type = "Long",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.ProducerEmbeddingsInjections.ProducerTopKSimClusterEmbeddingsInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.TopSimClustersWithScore",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "simcluster\_embedding\_top\_k\_producers\_by\_follow\_score",

key\_type = "com.twitter.simclusters\_v2.thriftscala.PersistedFullClusterId",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.ProducerEmbeddingsInjections.SimClusterEmbeddingTopKProducersInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.TopProducersWithScore",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "simcluster\_embedding\_top\_k\_producers\_by\_follow\_score\_updated",

key\_type = "com.twitter.simclusters\_v2.thriftscala.PersistedFullClusterId",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.ProducerEmbeddingsInjections.SimClusterEmbeddingTopKProducersInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.TopProducersWithScore",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "simcluster\_embedding\_top\_k\_producers\_by\_follow\_score\_2020",

key\_type = "com.twitter.simclusters\_v2.thriftscala.PersistedFullClusterId",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.ProducerEmbeddingsInjections.SimClusterEmbeddingTopKProducersInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.TopProducersWithScore",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "producer\_top\_k\_simcluster\_embeddings\_by\_follow\_score",

key\_type = "Long",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.ProducerEmbeddingsInjections.ProducerTopKSimClusterEmbeddingsInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.TopSimClustersWithScore",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "producer\_top\_k\_simcluster\_embeddings\_by\_follow\_score\_updated",

key\_type = "Long",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.ProducerEmbeddingsInjections.ProducerTopKSimClusterEmbeddingsInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.TopSimClustersWithScore",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "producer\_top\_k\_simcluster\_embeddings\_by\_follow\_score\_2020",

key\_type = "Long",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.ProducerEmbeddingsInjections.ProducerTopKSimClusterEmbeddingsInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.TopSimClustersWithScore",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "similar\_users\_by\_fav\_based\_producer\_embedding",

key\_type = "Long",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.ProducerEmbeddingsInjections.SimilarUsersInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.hermit.candidate.thriftscala.Candidates",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "similar\_users\_by\_follow\_based\_producer\_embedding",

key\_type = "Long",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.ProducerEmbeddingsInjections.SimilarUsersInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.hermit.candidate.thriftscala.Candidates",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "aggregatable\_producer\_simclusters\_embeddings\_by\_log\_fav\_score",

key\_type = "com.twitter.simclusters\_v2.thriftscala.SimClustersEmbeddingId",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.ProducerEmbeddingsInjections.ProducerSimClustersEmbeddingInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.SimClustersEmbedding",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "aggregatable\_producer\_simclusters\_embeddings\_by\_log\_fav\_score\_2020",

key\_type = "com.twitter.simclusters\_v2.thriftscala.SimClustersEmbeddingId",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.ProducerEmbeddingsInjections.ProducerSimClustersEmbeddingInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.SimClustersEmbedding",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "aggregatable\_producer\_simclusters\_embeddings\_by\_follow\_score\_2020",

key\_type = "com.twitter.simclusters\_v2.thriftscala.SimClustersEmbeddingId",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.ProducerEmbeddingsInjections.ProducerSimClustersEmbeddingInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.SimClustersEmbedding",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "aggregatable\_producer\_simclusters\_embeddings\_by\_log\_fav\_score\_relaxed\_fav\_engagement\_threshold\_2020",

key\_type = "com.twitter.simclusters\_v2.thriftscala.SimClustersEmbeddingId",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.ProducerEmbeddingsInjections.ProducerSimClustersEmbeddingInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.SimClustersEmbedding",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "aggregatable\_producer\_simclusters\_embeddings\_by\_fav\_score",

key\_type = "com.twitter.simclusters\_v2.thriftscala.SimClustersEmbeddingId",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.ProducerEmbeddingsInjections.ProducerSimClustersEmbeddingInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.SimClustersEmbedding",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "aggregatable\_producer\_simclusters\_embeddings\_by\_fav\_score\_2020",

key\_type = "com.twitter.simclusters\_v2.thriftscala.SimClustersEmbeddingId",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.ProducerEmbeddingsInjections.ProducerSimClustersEmbeddingInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.SimClustersEmbedding",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "aggregatable\_producer\_simclusters\_embeddings\_by\_log\_fav\_score\_thrift",

java\_schema = "com.twitter.simclusters\_v2.thriftjava.SimClustersEmbeddingWithId",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.thriftscala.SimClustersEmbeddingWithId",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

java\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-java",

],

scala\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-scala",

],

)

create\_datasets(

base\_name = "aggregatable\_producer\_simclusters\_embeddings\_by\_log\_fav\_score\_2020\_thrift",

java\_schema = "com.twitter.simclusters\_v2.thriftjava.SimClustersEmbeddingWithId",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.thriftscala.SimClustersEmbeddingWithId",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

java\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-java",

],

scala\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-scala",

],

)

create\_datasets(

base\_name = "aggregatable\_producer\_simclusters\_embeddings\_by\_follow\_score\_2020\_thrift",

java\_schema = "com.twitter.simclusters\_v2.thriftjava.SimClustersEmbeddingWithId",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.thriftscala.SimClustersEmbeddingWithId",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

java\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-java",

],

scala\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-scala",

],

)

create\_datasets(

base\_name = "aggregatable\_producer\_simclusters\_embeddings\_by\_fav\_score\_thrift",

java\_schema = "com.twitter.simclusters\_v2.thriftjava.SimClustersEmbeddingWithId",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.thriftscala.SimClustersEmbeddingWithId",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

java\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-java",

],

scala\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-scala",

],

)

create\_datasets(

base\_name = "aggregatable\_producer\_simclusters\_embeddings\_by\_fav\_score\_2020\_thrift",

java\_schema = "com.twitter.simclusters\_v2.thriftjava.SimClustersEmbeddingWithId",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.thriftscala.SimClustersEmbeddingWithId",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

java\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-java",

],

scala\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-scala",

],

)

create\_datasets(

base\_name = "aggregatable\_producer\_simclusters\_embeddings\_by\_log\_fav\_score\_relaxed\_fav\_engagement\_threshold\_2020\_thrift",

java\_schema = "com.twitter.simclusters\_v2.thriftjava.SimClustersEmbeddingWithId",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.thriftscala.SimClustersEmbeddingWithId",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

java\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-java",

],

scala\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-scala",

],

)

# TWICE & Clustering datasets

create\_datasets(

base\_name = "interested\_in\_twice\_by\_largest\_dim",

key\_type = "com.twitter.simclusters\_v2.thriftscala.SimClustersMultiEmbeddingId",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.EntityEmbeddingsInjections.EntitySimClustersMultiEmbeddingInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.SimClustersMultiEmbedding",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "interested\_in\_twice\_by\_largest\_dim\_fav\_score",

key\_type = "com.twitter.simclusters\_v2.thriftscala.SimClustersMultiEmbeddingId",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.EntityEmbeddingsInjections.EntitySimClustersMultiEmbeddingInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.SimClustersMultiEmbedding",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "interested\_in\_twice\_by\_largest\_dim\_2\_day\_update",

key\_type = "com.twitter.simclusters\_v2.thriftscala.SimClustersMultiEmbeddingId",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.EntityEmbeddingsInjections.EntitySimClustersMultiEmbeddingInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.SimClustersMultiEmbedding",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "interested\_in\_twice\_louvain",

key\_type = "com.twitter.simclusters\_v2.thriftscala.SimClustersMultiEmbeddingId",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.EntityEmbeddingsInjections.EntitySimClustersMultiEmbeddingInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.SimClustersMultiEmbedding",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "interested\_in\_twice\_connected\_components",

key\_type = "com.twitter.simclusters\_v2.thriftscala.SimClustersMultiEmbeddingId",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.EntityEmbeddingsInjections.EntitySimClustersMultiEmbeddingInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.SimClustersMultiEmbedding",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "clusters\_members\_largest\_dim\_ape\_similarity",

key\_type = "com.twitter.simclusters\_v2.common.UserId",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.ClusteringInjections.OrderedClustersAndMembersInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.OrderedClustersAndMembers",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "clusters\_members\_largest\_dim\_ape\_similarity\_2\_day\_update",

key\_type = "com.twitter.simclusters\_v2.common.UserId",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.ClusteringInjections.OrderedClustersAndMembersInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.OrderedClustersAndMembers",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "clusters\_members\_louvain\_ape\_similarity",

key\_type = "com.twitter.simclusters\_v2.common.UserId",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.ClusteringInjections.OrderedClustersAndMembersInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.OrderedClustersAndMembers",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "clusters\_members\_connected\_components\_ape\_similarity",

key\_type = "com.twitter.simclusters\_v2.common.UserId",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.ClusteringInjections.OrderedClustersAndMembersInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.OrderedClustersAndMembers",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

# End of TWICE & Clustering datasets

create\_datasets(

base\_name = "simclusters\_offline\_tweet\_cluster\_scores",

description = "A dataset that contains the scores for tweet and cluster pairs",

java\_schema = "com.twitter.simclusters\_v2.thriftjava.TweetAndClusterScores",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.thriftscala.TweetAndClusterScores",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

java\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-java",

],

scala\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-scala",

],

)

create\_datasets(

base\_name = "simclusters\_offline\_tweet\_top\_k\_clusters",

description = "A dataset that contains the top clusters for each tweet",

java\_schema = "com.twitter.simclusters\_v2.thriftjava.TweetTopKClustersWithScores",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.thriftscala.TweetTopKClustersWithScores",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

java\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-java",

],

scala\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-scala",

],

)

create\_datasets(

base\_name = "simclusters\_offline\_cluster\_top\_k\_tweets",

description = "A dataset that contains the top tweets for each cluster",

java\_schema = "com.twitter.simclusters\_v2.thriftjava.ClusterTopKTweetsWithScores",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.thriftscala.ClusterTopKTweetsWithScores",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

java\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-java",

],

scala\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-scala",

],

)

create\_datasets(

base\_name = "simclusters\_inferred\_entities\_from\_known\_for",

key\_type = "Long",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.InferredEntitiesInjections.InferredEntityInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.SimClustersInferredEntities",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "simclusters\_inferred\_entities\_from\_interested\_in",

key\_type = "Long",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.InferredEntitiesInjections.InferredEntityInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.SimClustersInferredEntities",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "simclusters\_inferred\_entities\_from\_interested\_in\_keyed\_by\_cluster",

key\_type = "Int",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.InferredEntitiesInjections.InferredEntityKeyedByClusterInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.SimClustersInferredEntities",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "co\_engagement\_top\_k\_similar\_tweets",

description = "A dataset that contains the top similar tweets based on co-engagement",

java\_schema = "com.twitter.simclusters\_v2.thriftjava.TweetTopKTweetsWithScore",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.thriftscala.TweetTopKTweetsWithScore",

segment\_type = "partitioned",

tags = ["bazel-compatible"],

java\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-java",

],

scala\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-scala",

],

)

create\_datasets(

base\_name = "rux\_faved\_top\_k\_tweets",

description = "A dataset that contains the top similar tweets based on rux fav-to-impression ratio",

java\_schema = "com.twitter.simclusters\_v2.thriftjava.TweetTopKTweetsWithScore",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.thriftscala.TweetTopKTweetsWithScore",

segment\_type = "partitioned",

tags = ["bazel-compatible"],

java\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-java",

],

scala\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-scala",

],

)

create\_datasets(

base\_name = "semantic\_core\_embeddings\_from\_producer",

key\_type = "com.twitter.simclusters\_v2.thriftscala.SimClustersEmbeddingId",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.EntityEmbeddingsInjections.EntitySimClustersEmbeddingInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.SimClustersEmbedding",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "simclusters\_v2\_interested\_in\_from\_producer\_embeddings\_20M\_145K\_updated",

key\_type = "Long",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.InterestedInInjection.injection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.ClustersUserIsInterestedIn",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "simclusters\_v2\_interested\_in\_from\_aggregatable\_producer\_embeddings\_20M\_145K\_2020",

key\_type = "Long",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.InterestedInInjection.injection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.ClustersUserIsInterestedIn",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "geopopular\_top\_tweet\_impressed\_topics",

key\_type = "String",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.SemanticCoreEntitiesInjections.StringToSemanticCoreEntityScoreListInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.recos.entities.thriftscala.SemanticCoreEntityScoreList",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "similar\_topics\_from\_topic\_follow\_graph",

key\_type = "Long",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.SemanticCoreEntitiesInjections.LongToSemanticCoreEntityScoreListInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.recos.entities.thriftscala.SemanticCoreEntityScoreList",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "top\_locale\_topics\_for\_producer\_from\_em",

key\_type = "com.twitter.recos.entities.thriftscala.UserIdWithLocale",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.SemanticCoreEntitiesInjections.UserWithLocaleToSemanticCoreEntityScoreListInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.recos.entities.thriftscala.SemanticCoreEntityScoreList",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "top\_producers\_for\_locale\_topics\_from\_topic\_follow\_graph",

key\_type = "com.twitter.recos.entities.thriftscala.SemanticCoreEntityWithLocale",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.SemanticCoreEntitiesInjections.SemanticCoreEntityWithLocaleToUsersScoreListInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.recos.entities.thriftscala.UserScoreList",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "topic\_top\_producers\_em",

key\_type = "com.twitter.recos.entities.thriftscala.SemanticCoreEntityWithLocale",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.SemanticCoreEntitiesInjections.SemanticCoreEntityWithLocaleToUsersScoreListInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.recos.entities.thriftscala.UserScoreList",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "adhoc\_abuse\_simcluster\_features",

java\_schema = "com.twitter.simclusters\_v2.thriftjava.AdhocSingleSideClusterScores",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.thriftscala.AdhocSingleSideClusterScores",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

java\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-java",

],

scala\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-scala",

],

)

create\_datasets(

base\_name = "search\_abuse\_simcluster\_features\_manhattan",

key\_type = "Long",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.SingleSideUserScoresInjection.injection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.SingleSideUserScores",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "adhoc\_cross\_simcluster\_block\_interaction\_features",

java\_schema = "com.twitter.simclusters\_v2.thriftjava.AdhocCrossSimClusterInteractionScores",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.thriftscala.AdhocCrossSimClusterInteractionScores",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

java\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-java",

],

scala\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-scala",

],

)

create\_datasets(

base\_name = "adhoc\_cross\_simcluster\_fav\_interaction\_features",

java\_schema = "com.twitter.simclusters\_v2.thriftjava.AdhocCrossSimClusterInteractionScores",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.thriftscala.AdhocCrossSimClusterInteractionScores",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

java\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-java",

],

scala\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-scala",

],

)

create\_datasets(

base\_name = "top\_k\_right\_nouns",

key\_type = "com.twitter.simclusters\_v2.thriftscala.RightNodeTypeStruct",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.MultiTypeGraphInjections.topKRightNounListInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.NounWithFrequencyList",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "top\_k\_right\_nouns\_scio",

key\_type = "com.twitter.simclusters\_v2.thriftscala.RightNodeTypeStruct",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.MultiTypeGraphInjections.topKRightNounListInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.NounWithFrequencyList",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "offline\_cluster\_top\_media\_tweets\_20M\_145K\_2020",

key\_type = "com.twitter.simclusters\_v2.thriftscala.DayPartitionedClusterId",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.ClusterTopMediaTweetsInjection.injection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.TweetsWithScore",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "truncated\_multi\_type\_graph",

key\_type = "com.twitter.simclusters\_v2.thriftscala.LeftNode",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.MultiTypeGraphInjections.truncatedMultiTypeGraphInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.RightNodeWithEdgeWeightList",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "truncated\_multi\_type\_graph\_scio",

key\_type = "com.twitter.simclusters\_v2.thriftscala.LeftNode",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.MultiTypeGraphInjections.truncatedMultiTypeGraphInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.RightNodeWithEdgeWeightList",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "multi\_type\_graph\_for\_top\_k\_right\_nodes\_thrift\_scio",

java\_schema = "com.twitter.simclusters\_v2.thriftjava.MultiTypeGraphEdge",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.thriftscala.MultiTypeGraphEdge",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

java\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-java",

],

scala\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-scala",

],

)

create\_datasets(

base\_name = "full\_multi\_type\_graph",

java\_schema = "com.twitter.simclusters\_v2.thriftjava.MultiTypeGraphEdge",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.thriftscala.MultiTypeGraphEdge",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

java\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-java",

],

scala\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-scala",

],

)

create\_datasets(

base\_name = "right\_node\_sim\_hash\_scio",

java\_schema = "com.twitter.simclusters\_v2.thriftjava.RightNodeSimHashSketch",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.thriftscala.RightNodeSimHashSketch",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

java\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-java",

],

scala\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-scala",

],

)

create\_datasets(

base\_name = "right\_node\_cosine\_similarity\_scio",

key\_type = "com.twitter.simclusters\_v2.thriftscala.RightNode",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.MultiTypeGraphInjections.similarRightNodesInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.SimilarRightNodes",

scala\_dependencies = [

"src/scala/com/twitter/scalding\_internal/multiformat/format",

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "user\_topic\_weighted\_embedding",

key\_type = "Long",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.InterestedInInjection.injection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.ClustersUserIsInterestedIn",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "user\_topic\_weighted\_embedding\_parquet",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.thriftscala.UserTopicWeightedEmbedding",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

scala\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-scala",

],

)

create\_datasets(

base\_name = "explore\_mbcg\_user\_embeddings\_kv",

key\_type = "Long",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.EntityEmbeddingsInjections.UserMbcgEmbeddingInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.ml.api.thriftscala.Embedding",

scala\_dependencies = [

"src/scala/com/twitter/scalding\_internal/multiformat/format",

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "offline\_tweet\_recommendations\_from\_interested\_in\_20M\_145K\_2020",

key\_type = "Long",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.MultiTypeGraphInjections.tweetRecommendationsInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.CandidateTweetsList",

scala\_dependencies = [

"src/scala/com/twitter/scalding\_internal/multiformat/format",

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "offline\_tweet\_recommendations\_from\_interested\_in\_20M\_145K\_2020\_HL\_0\_EL\_15",

key\_type = "Long",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.MultiTypeGraphInjections.tweetRecommendationsInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.CandidateTweetsList",

scala\_dependencies = [

"src/scala/com/twitter/scalding\_internal/multiformat/format",

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "offline\_tweet\_recommendations\_from\_interested\_in\_20M\_145K\_2020\_HL\_2\_EL\_15",

key\_type = "Long",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.MultiTypeGraphInjections.tweetRecommendationsInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.CandidateTweetsList",

scala\_dependencies = [

"src/scala/com/twitter/scalding\_internal/multiformat/format",

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "offline\_tweet\_recommendations\_from\_interested\_in\_20M\_145K\_2020\_HL\_2\_EL\_50",

key\_type = "Long",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.MultiTypeGraphInjections.tweetRecommendationsInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.CandidateTweetsList",

scala\_dependencies = [

"src/scala/com/twitter/scalding\_internal/multiformat/format",

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "offline\_tweet\_recommendations\_from\_interested\_in\_20M\_145K\_2020\_HL\_8\_EL\_50",

key\_type = "Long",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.MultiTypeGraphInjections.tweetRecommendationsInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.CandidateTweetsList",

scala\_dependencies = [

"src/scala/com/twitter/scalding\_internal/multiformat/format",

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "offline\_tweet\_recommendations\_from\_mts\_consumer\_embeddings",

key\_type = "Long",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.MultiTypeGraphInjections.tweetRecommendationsInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.CandidateTweetsList",

scala\_dependencies = [

"src/scala/com/twitter/scalding\_internal/multiformat/format",

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "fav\_based\_simclusters\_cluster\_to\_tweet\_index",

key\_type = "com.twitter.simclusters\_v2.thriftscala.FullClusterId",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.ClusterTopTweetsInjection.clusterIdToTopKTweetsInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.TopKTweetsWithScores",

scala\_dependencies = [

"src/scala/com/twitter/scalding\_internal/multiformat/format",

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "video\_view\_based\_simclusters\_cluster\_to\_tweet\_index",

key\_type = "com.twitter.simclusters\_v2.thriftscala.FullClusterId",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.ClusterTopTweetsInjection.clusterIdToTopKTweetsInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.TopKTweetsWithScores",

scala\_dependencies = [

"src/scala/com/twitter/scalding\_internal/multiformat/format",

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "retweet\_based\_simclusters\_cluster\_to\_tweet\_index",

key\_type = "com.twitter.simclusters\_v2.thriftscala.FullClusterId",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.ClusterTopTweetsInjection.clusterIdToTopKTweetsInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.TopKTweetsWithScores",

scala\_dependencies = [

"src/scala/com/twitter/scalding\_internal/multiformat/format",

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "reply\_based\_simclusters\_cluster\_to\_tweet\_index",

key\_type = "com.twitter.simclusters\_v2.thriftscala.FullClusterId",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.ClusterTopTweetsInjection.clusterIdToTopKTweetsInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.TopKTweetsWithScores",

scala\_dependencies = [

"src/scala/com/twitter/scalding\_internal/multiformat/format",

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "push\_open\_based\_simclusters\_cluster\_to\_tweet\_index",

key\_type = "com.twitter.simclusters\_v2.thriftscala.FullClusterId",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.ClusterTopTweetsInjection.clusterIdToTopKTweetsInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.TopKTweetsWithScores",

scala\_dependencies = [

"src/scala/com/twitter/scalding\_internal/multiformat/format",

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "ads\_fav\_based\_simclusters\_cluster\_to\_tweet\_index",

key\_type = "com.twitter.simclusters\_v2.thriftscala.FullClusterId",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.ClusterTopTweetsInjection.clusterIdToTopKTweetsInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.TopKTweetsWithScores",

scala\_dependencies = [

"src/scala/com/twitter/scalding\_internal/multiformat/format",

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "ads\_fav\_click\_based\_simclusters\_cluster\_to\_tweet\_index",

key\_type = "com.twitter.simclusters\_v2.thriftscala.FullClusterId",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.ClusterTopTweetsInjection.clusterIdToTopKTweetsInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.TopKTweetsWithScores",

scala\_dependencies = [

"src/scala/com/twitter/scalding\_internal/multiformat/format",

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "fav\_based\_evergreen\_content\_simclusters\_cluster\_to\_tweet\_index",

key\_type = "com.twitter.simclusters\_v2.thriftscala.FullClusterId",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.ClusterTopTweetsInjection.clusterIdToTopKTweetsInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.TopKTweetsWithScores",

scala\_dependencies = [

"src/scala/com/twitter/scalding\_internal/multiformat/format",

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "fav\_based\_video\_simclusters\_cluster\_to\_tweet\_index",

key\_type = "com.twitter.simclusters\_v2.thriftscala.FullClusterId",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.ClusterTopTweetsInjection.clusterIdToTopKTweetsInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.TopKTweetsWithScores",

scala\_dependencies = [

"src/scala/com/twitter/scalding\_internal/multiformat/format",

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "simclusters\_v2\_global\_language\_embedding",

key\_type = "String",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.hdfs\_sources.injections.InterestedInInjection.languageInjection",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

val\_type = "com.twitter.simclusters\_v2.thriftscala.ClustersUserIsInterestedIn",

scala\_dependencies = [

"src/scala/com/twitter/simclusters\_v2/hdfs\_sources/injections",

],

)

create\_datasets(

base\_name = "simclusters\_v2\_global\_language\_embedding\_thrift",

java\_schema = "com.twitter.simclusters\_v2.thriftjava.LanguageToClusters",

platform = "java8",

role = "cassowary",

scala\_schema = "com.twitter.simclusters\_v2.thriftscala.LanguageToClusters",

segment\_type = "snapshot",

tags = ["bazel-compatible"],

java\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-java",

],

scala\_dependencies = [

"src/thrift/com/twitter/simclusters\_v2:simclusters\_v2-thrift-scala",

],

)