Download Move lens data (1 M): https://grouplens.org/datasets/movielens/1m/

Ratings.dat: userID, movieId, rating,timestamp

RECOMMENDATION CODE

import org.apache.spark.ml.recommendation.ALS

val ratings =spark.read.textFile("/tmp/ratings.dat").map { line => val Array(userID, movieId, rating,timestamp) = line.split("::").map(\_.toInt)(userID, movieId, rating,timestamp) }.toDF("userID", "movieId", "rating","timestamp")

val Array(training, test) = ratings.randomSplit(Array(0.8, 0.2))

val als = new ALS().setMaxIter(5).setRegParam(0.01).setUserCol("userID").setItemCol("movieId").setRatingCol("rating")val alsModel = als.fit(training)

val predictions = alsModel.transform(test)

val perUserPredictions = predictions

.orderBy(col("userID"), col("prediction").desc)

.groupBy("userID")

.agg(expr("collect\_list(movieId) as movies"))

perUserPredictions.collect.take(5)VERIFY RECOMMENDATION:

import org.apache.spark.mllib.evaluation.{

RankingMetrics,

RegressionMetrics}

val regComparison = predictions.select("rating", "prediction").rdd .map(x => (x(0).asInstanceOf[Float].toDouble, x(1).asInstanceOf[Float].toDouble))regComparison.collect.take(5)