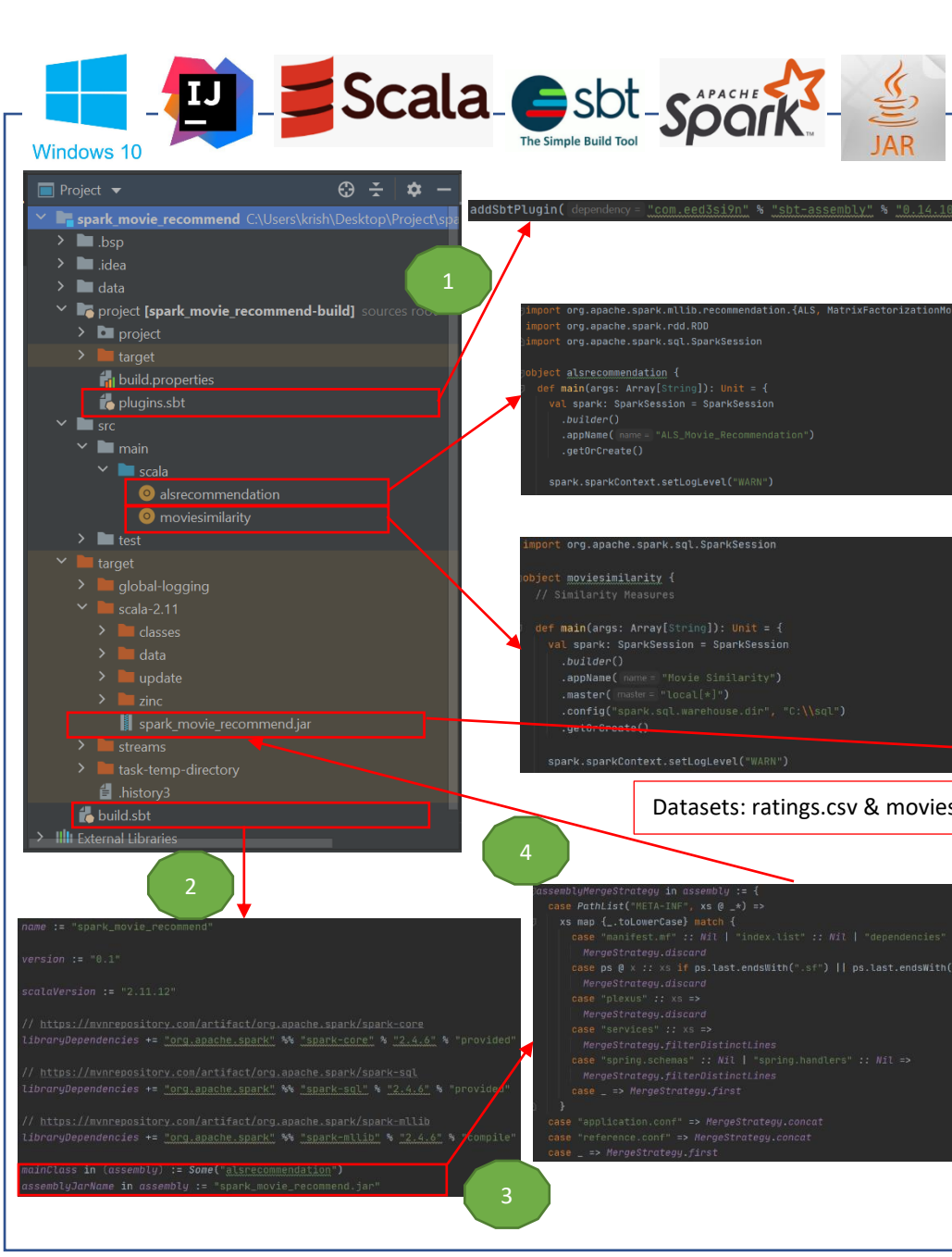
Windows 10
Scala
The Simple Build Tool
APACHE spark
JAR
Project
spark_movie_recommend C:\Users\krish\Desktop\Project\spa
.bsp
.idea
data
project [spark_movie_recommend-build] sources roo
project
target
build.properties
plugins.sbt
src
main
scala
alsrecommendation
moviesimilarity
test
target
global-logging
scala-2.11
classes
data
update
zinc
spark_movie_recommend.jar
streams
task-temp-directory
.history3
build.sbt
External Libraries
1
addSbtPlugin(dependency = "com.eed3si9n" % "sbt-assembly" % "0.14.10")
import org.apache.spark.mllib.recommendation.{ALS, MatrixFactorizationModel
import org.apache.spark.rdd.RDD
import org.apache.spark.sql.SparkSession
object alsrecommendation {
def main(args: Array[String]): Unit = {
val spark: SparkSession = SparkSession
.builder()
.appName(name = "ALS_Movie_Recommendation")
.getOrCreate()
spark.sparkContext.setLogLevel("WARN")
import org.apache.spark.sql.SparkSession
object moviesimilarity {
// Similarity Measures
def main(args: Array[String]): Unit = {
val spark: SparkSession = SparkSession
.builder()
.appName(name = "Movie Similarity")
.master(master = "local[*]")
.config("spark.sql.warehouse.dir", "C:\\sql")
.getOrCreate()
spark.sparkContext.setLogLevel("WARN")
Datasets: ratings.csv & movies.csv
2
name := "spark_movie_recommend"
version := "0.1"
scalaVersion := "2.11.12"
// https://mvnrepository.com/artifact/org.apache.spark/spark-core
libraryDependencies += "org.apache.spark" %% "spark-core" % "2.4.6" % "provided"
// https://mvnrepository.com/artifact/org.apache.spark/spark-sql
libraryDependencies += "org.apache.spark" %% "spark-sql" % "2.4.6" % "provided"
// https://mvnrepository.com/artifact/org.apache.spark/spark-mllib
libraryDependencies += "org.apache.spark" %% "spark-mllib" % "2.4.6" % "compile"
mainClass in (assembly) := Some("alsrecommendation")
assemblyJarName in assembly := "spark_movie_recommend.jar"
3
4
assemblyMergeStrategy in assembly := {
case PathList("META-INF", xs @ _*) =>
xs map {_.toLowerCase} match {
case "manifest.mf" :: Nil | "index.list" :: Nil | "dependencies" :: Nil =>
MergeStrategy.discard
case ps @ x :: xs if ps.last.endsWith(".sf") || ps.last.endsWith(".dsa") =>
MergeStrategy.discard
case "plexus" :: xs =>
MergeStrategy.discard
case "services" :: xs =>
MergeStrategy.filterDistinctLines
case "spring.schemas" :: Nil | "spring.handlers" :: Nil =>
MergeStrategy.filterDistinctLines
case _ => MergeStrategy.first
}
case "application.conf" => MergeStrategy.concat
case "reference.conf" => MergeStrategy.concat
case _ => MergeStrategy.first
aws
Amazon EMR
hadoop
Amazon S3
5
Software configuration
Hardware configuration
Security and access
Cancel
Create cluster
6
Create bucket
Buckets are containers for data stored in S3. Learn more
General configuration
Bucket name
Region
US East (Ohio) us-east-2
Choose bucket
7
8
9
All Applications
Application History
About
Applications
FINISHED
FAILED
KILLED
Tools
Rating:(UserID, MovieID, Rating)
Rating(610,183947,4.472530826402922)
Rating(610,151989,4.455056694885744)
Rating(610,202231,4.388460748498632)
Rating(610,144202,4.382982243183964)
Rating(610,165559,4.315681612768971)