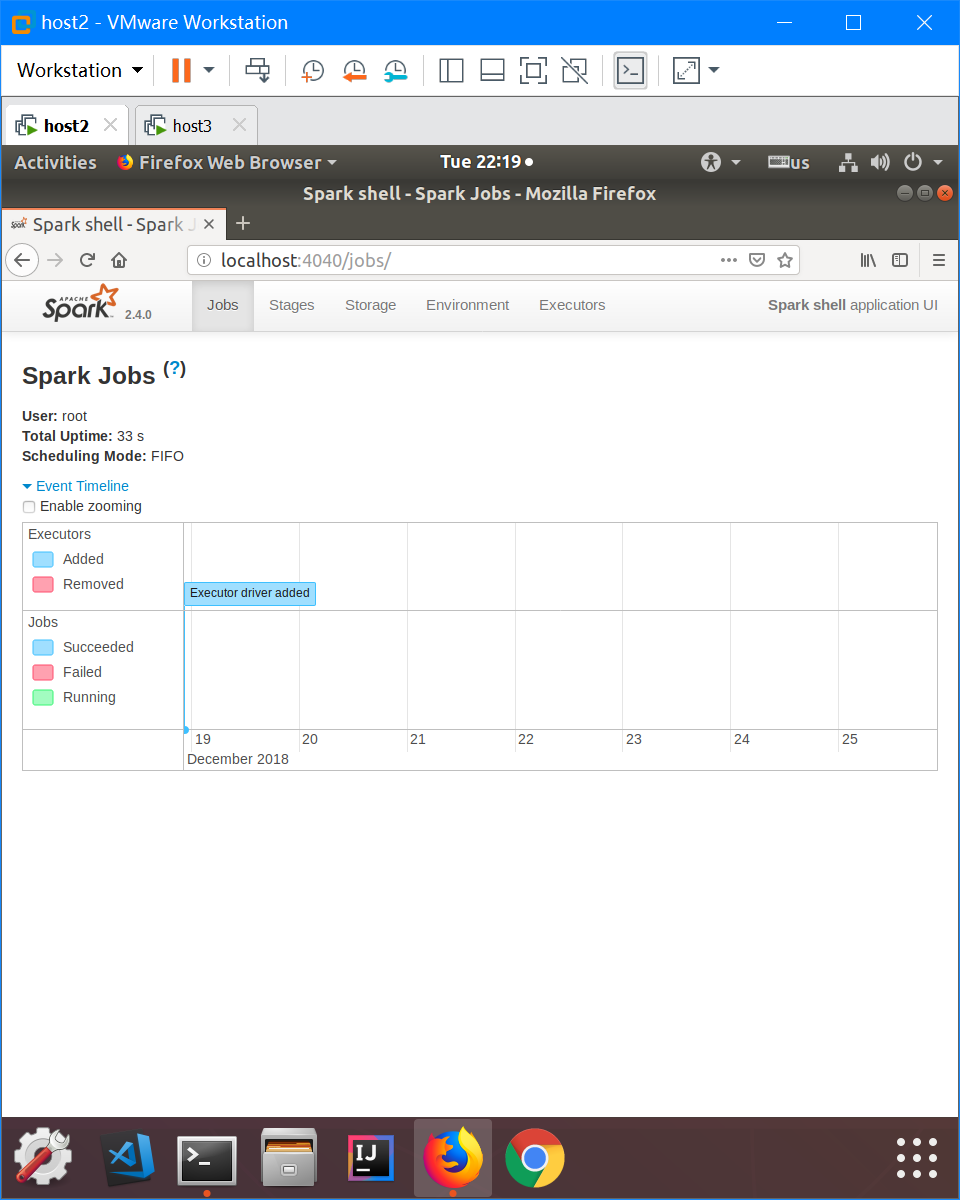
作业9实验报告

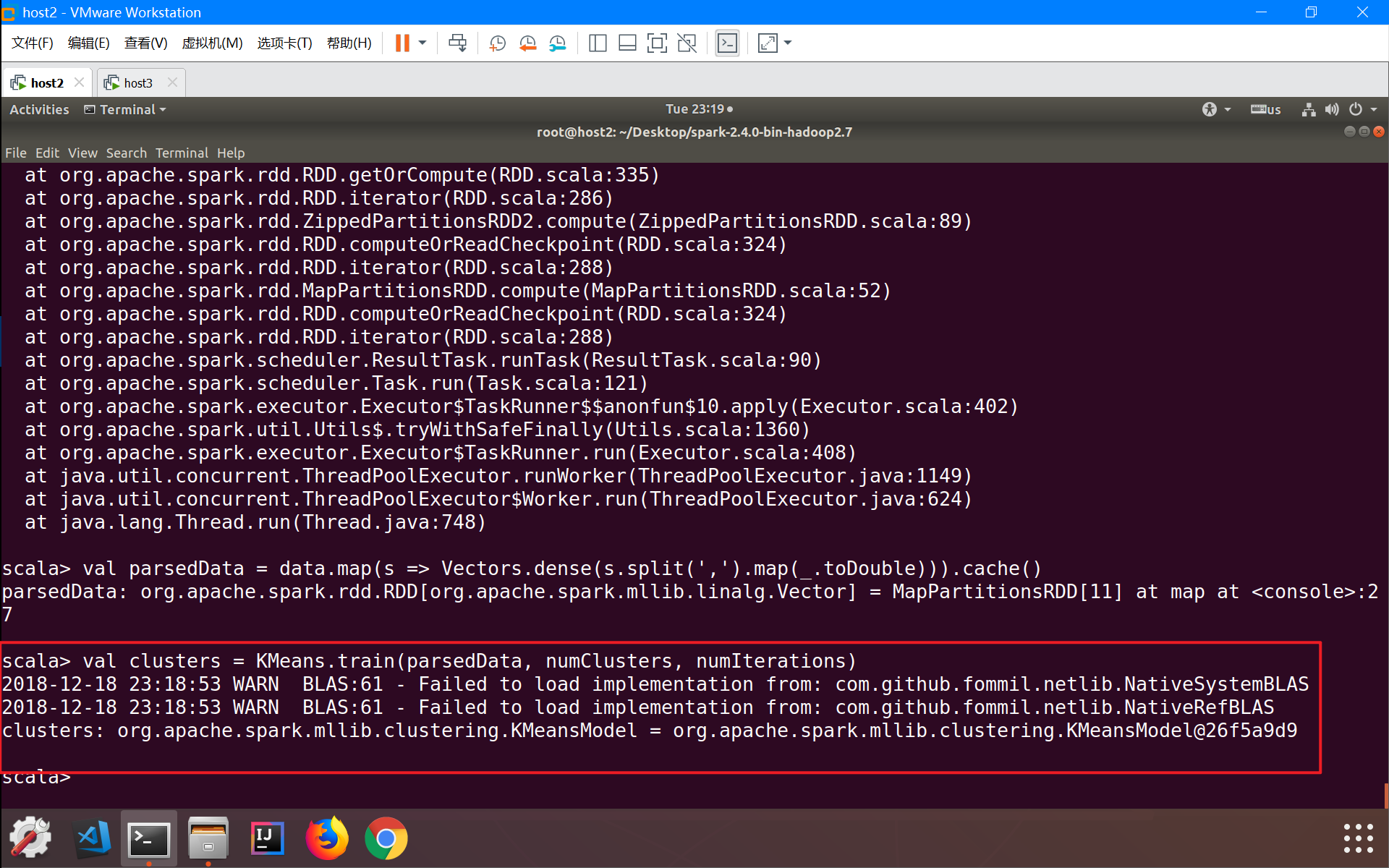
李康 161278015

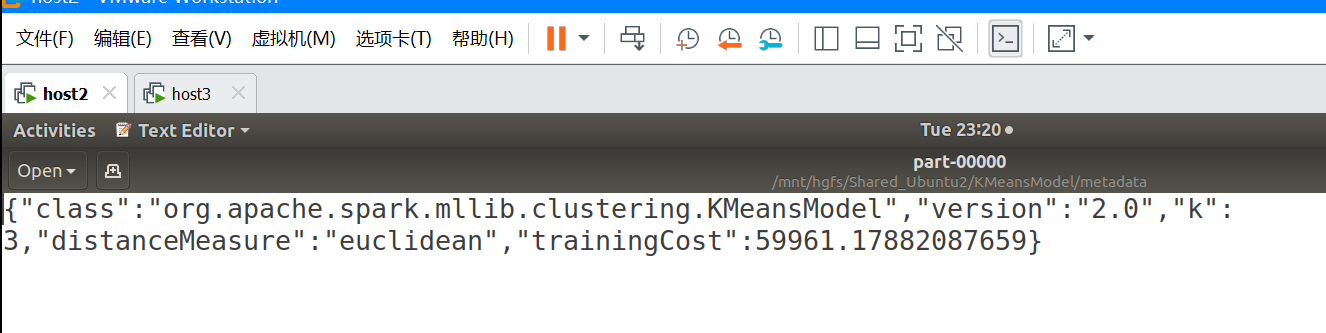
* 可以使用jps去查看有没有开启master进程
  + **JPS 名称:**jps - Java Virtual Machine Process Status Tool

安装spak成功：

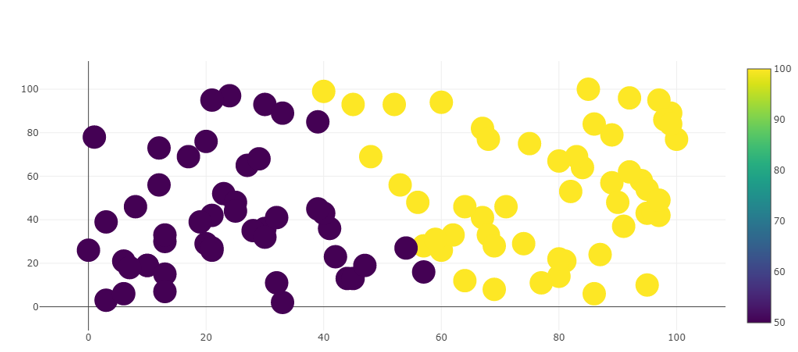
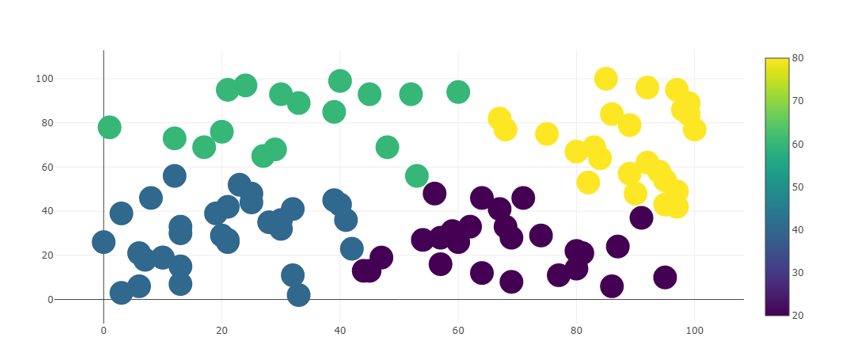
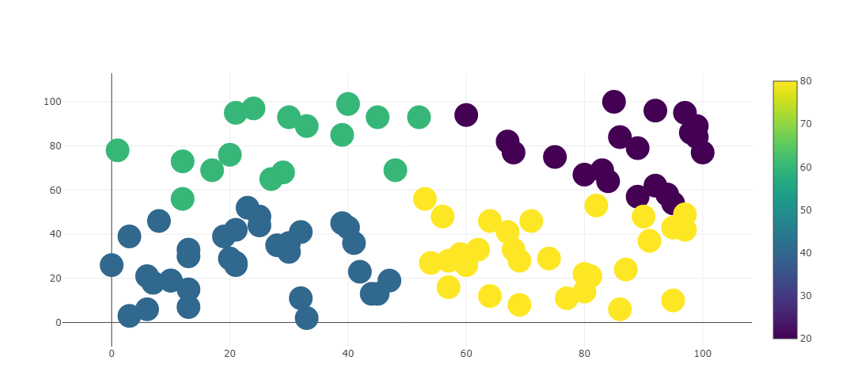


运行成功：





实验结果比较：

* K=4,Loop=20 (hadoop上的运行结果)
  + 
* Spark上的运行结果：
  + 
* 可以看出spark的分类结果与hadoop相差不大。
* 运行代码
  + import org.apache.spark.mllib.clustering.{KMeans, KMeansModel}
  + import org.apache.spark.mllib.linalg.Vectors
  + // Load and parse the data
  + val data = sc.textFile("Instance.txt")
  + val parsedData = data.map(s => Vectors.dense(s.split(' ').map(\_.toDouble))).cache()
  + // Cluster the data into two classes using KMeans
  + val numClusters = 2
  + val numIterations = 20
  + val clusters = KMeans.train(parsedData, numClusters, numIterations)
  + // Evaluate clustering by computing Within Set Sum of Squared Errors
  + val WSSSE = clusters.computeCost(parsedData)
  + println(s"Within Set Sum of Squared Errors = $WSSSE")
  + // Save and load model
  + clusters.save(sc, "KMeansModel")
  + val sameModel = KMeansModel.load(sc, "KMeansModel")