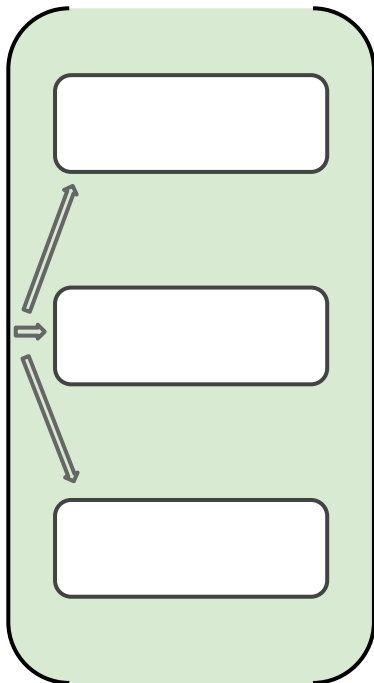


 parallel

Input  
Streaming  
Collection



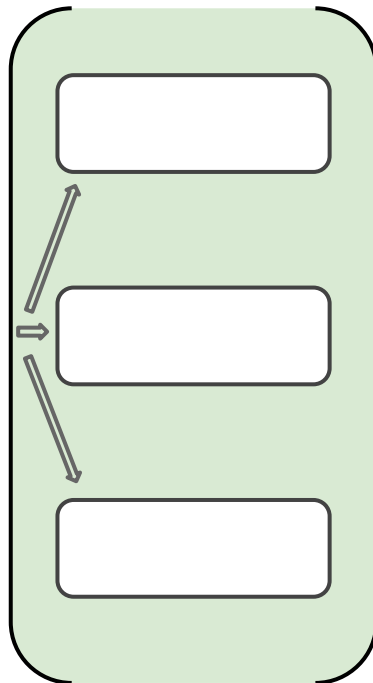
RDD<Tupel3>



Dictionary creation



RDD<LabeledPoints>



Training and Test  
Set creation



MLlib:  
Naive Bayes  
Classifier



**Result**