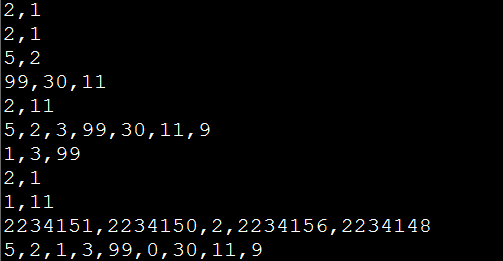
<https://github.com/apache/spark/tree/master/mllib/src/main/scala/org/apache/spark/mllib/fpm>

spark-submit --class "org.apache.spark.mllib.fpm.FPRunner" --driver-memory 10G FPRunner.jar hftp://10.200.91.164:50070/user/liulian/cid\_fp/data fpg/output fpg/res 1000 0.003

参数说明：

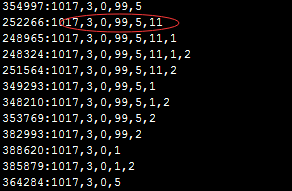
1 hftp://10.200.91.164:50070/user/liulian/cid\_fp/data 输入目录，一行表示一个transaction（比如一个用户看过的所有电影）

数据格式如下



2 fpg/output频繁项集保存的目录

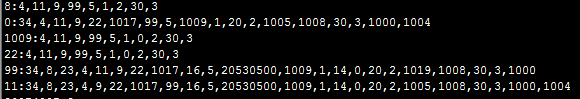
保存的数据格式如下



冒号左边表示支持度，右边表示频繁项集

3 fpg/res 推荐规则结果

数据格式如下



比如第一行表示，喜欢8的，还喜欢4,11等等

注意点：

1 购物篮数据清洗，一个用户当做一个购物篮，如果他对应的物品太少，则剔除。

2 支持度的合理性，该多大才合理呢，有待实践……