问题：一开始集群测试，简单的正则比ASCII匹配要快。单机测试独立功能100万条数据，ASCII是正则性能的两倍。1亿条数据时，差距拉大，ASCII7秒，正则40多秒将近6倍。这还是ASCII复杂度更高。

Producer对象不支持序列化，无法生命成广播变量。需要lazy执行，用到时创建避免了网络传输。跑完一个文件需要时间0.823秒。