mentionedTrend.py,

这个文件要最先运行，因为其他文件都可能会用到它产生的结果

可能需要修改的地方：

1. match字段

等东软的数据好了之后，读取WEIBO\_POST\_Keywords 表的match字段（party文件夹下的read\_key\_words.py,这个你很熟悉了），请确认好“春晚”有关的数据对应的match字段是什么，然后在Brand\_echo.py Line 65做相应的修改，目前我暂时写了“春晚”；

1. 按天抓取

这个根据孙总的提醒，之前就把代码改成按天抓取了，修改在line49，逻辑之前说过一次，相信你已经很清楚啦：如果是抓取当日的数据，inter设为0，3天前的数据，inter设为3，以此类推。

1. 日期格式

因为没见过数据，不确定格式，目前我推测数据格式是像这样“2016-04-25 21:47”，但我们自己入库就不需要小时分钟了，只留“2016-04-25”，所以做了split，如果跟我预计的一样，那就不用改了，如果不一样，请根据刚刚的意思意会一下，做相应的修改哈。

mentionedTrendHourly.py

Note:

同上

Brand\_echo.py

在此之前，需先运行mentionedTrend.py,并将附件的gala\_sponsor.sql（目前晓健只提供了一条数据）复制过去。两者是该py文件的输入。

Note:

同上

ciyun\_chunwan.py

Note:

同上

gala\_attention\_degree.py

将附件中的gala\_celebrity.sql复制过去，另外需导入工具包tools，将文件夹至于同目录下就可以

Note：

同上

gala\_discussion.py

同上

region\_interaction.py

同上

customerEvaluation.py

将hbase\_info.sql复制过去

这里有可能需要修改的地方类似上文提到的，match关键字，比如“军情解码”为例，如果WEIBO\_POST\_Keywords 表的match字段有关“军情解码”的关键字就是“军情解码”，那么不用做任何修改，如果不一样的话，根据相关情况在line102修改：目前我的“one\_program”是从sql表hbase\_info中读取的program字段

latest\_news.py

1. 按天抓取

这个根据孙总的提醒，之前就把代码改成按天抓取了，修改在line49，逻辑之前说过一次，相信你已经很清楚啦：如果是抓取当日的数据，inter设为0，3天前的数据，inter设为3，以此类推。

1. 日期格式

因为没见过数据，不确定格式，目前我推测数据格式是像这样“2016-04-25 21:47”，但我们自己入库就不需要小时分钟了，只留“2016-04-25”，所以做了split，如果跟我预计的一样，那就不用改了，如果不一样，请根据刚刚的意思意会一下，做相应的修改哈。

LDA的部署略难一点（参看lda\_new文件夹下的read\_me），实在不行，等我回来再部署吧^^。