怎么把大数据传到大数据集群的 HDFS 文件系统？

什么是HDFS文件系统？

怎么录入，对数据格式的要求？

对网路新闻数据进行清理处理

为什么？是不是爬下来的数据有问题？

怎么进行清理处理？

Dataiku是什么，该怎么用？

根据现有的数据字段属性情况

是不是意味着该建数据库，数据库的属性有哪些？

对互联网新闻数据做主题提取分析，将不同的主题和对应的关键词抽取出来，对新闻正文

，对新闻正文内容做分词创建词库，提取主题，做主题建模

这句话该怎么理解？

可不可以理解为预处理，去停用词，分词，进行TF权重处理，建立分类器，生

成模型，评估模型

使用 Fire.Fox 交互式的展现提取的主题信息，可以选择查看具体的信息情况

这是什么？

完全根据用户的个性信息展现？

完全不懂？

LDA算法是不是更好？