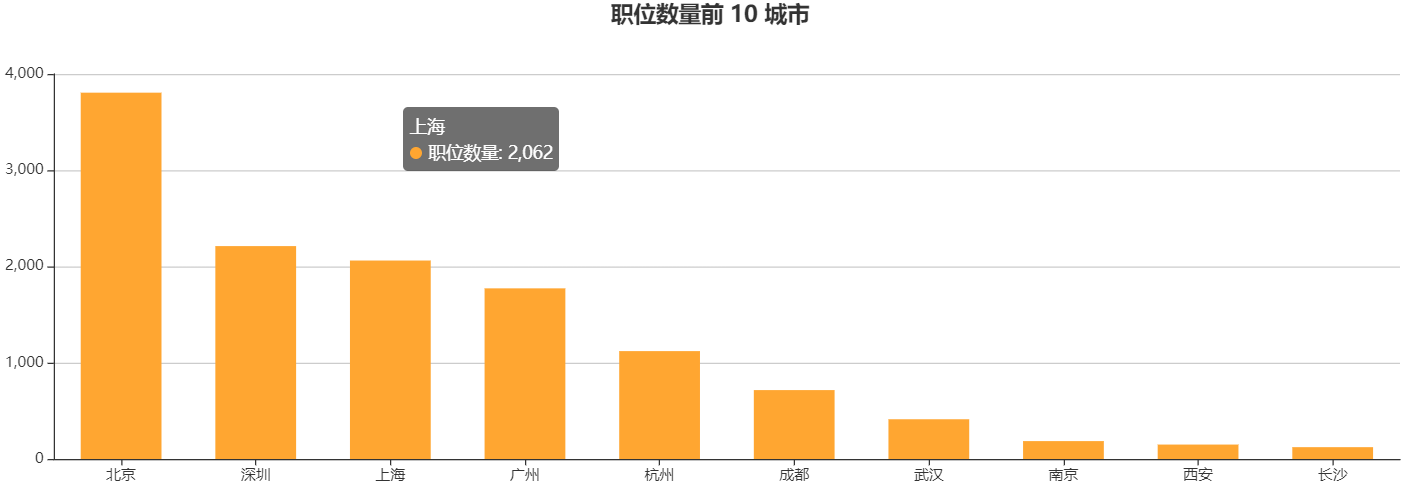
NoSQL数据库课程考核

题目：基于NoSQL的2020年招聘就业数据分析

要求：以下内容均需完成，并且用文字/图片反映在课程报告中。

1. 数据库推荐使用MongoDB，因为提供的数据文件（在“keshe”文件夹内）是BSON格式，可以直接用MongoDB读取到本地数据库。读取方法可以参考文末附录，并且自己再上网查找相关操作信息。如果用其他数据库，则需要自己进行格式转换和数据记录存储。该就业招聘数据是2020年6月份从“拉勾网”进行部分爬取并且存入MongoDB。这里有13604条招聘数据，但是并不完整。数据仅供教学使用，请勿外传。
2. 对提供的2020年就业数据文件进行读取，并且以记录形式存入NoSQL数据库。就业数据文件发送到课程QQ群的群文件。
3. 分布式/伪分布式存储。要求对数据库采用伪分布式或者分布式安装
4. 采用查询语句大致查询就业数据情况。采用Mapreduce编程进行就业数据的分布式读取和统计。
5. 在就业数据读取之后，采用编程软件（Java、Python、Matlab、甚至一些图形开发包如E-chart等）进行数据作图和显示。

如下图显示的是就业职位数量前十的城市。



其中横轴为城市名称，纵轴为职位数量。

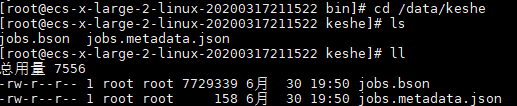
可以从多个角度进行数据分析。比如显示各种职位数量前十的排序和图示，或者求职者学历水平的排序和图示，或者薪酬前十的排序和图示，等等。

验收:

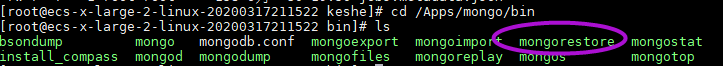
1. 提交电子版课程报告。按照模板格式完善封面信息。
2. **每位同学**将电子版课程设计内容打包发送到老师QQ邮箱20367252@qq.com。在7月15日前发送到老师邮箱：20367252@qq.com。**邮件名和附件名，都要命名为**：学号-姓名-课程报告-日期。（比如：3117001387-蔡楠-课程报告-20200712）。包括：课程报告文本+源代码（工程文件）+数据库运行过程及效果展示视频(可选)。
3. 独立完成设计。课程报告内容请勿雷同，以避免误认为抄袭。
4. 课程结束2周内（7月15日前）务必提交课程报告纸质版以及电子版。可以提前交，迟交则不接收。

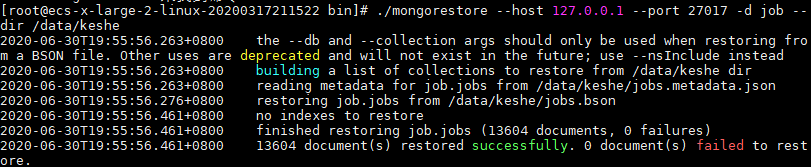
附录:

1. 将“keshe”文件夹拷贝到服务器的本地目录。



1. 用MongoDB的小程序“mongorestore”，将数据文件恢复到本地服务器的mongoDB数据库里面。





如果对mongorestore用法要继续了解，请查看如下链接： https://www.cnblogs.com/ytxiao/p/12919200.html

1. 查看就业数据记录



老师联系方式——

微信：stranger0072000

电话：13427583825