



NoSQL数据库课程报告

学院：\_\_\_\_\_\_\_\_\_自动化学院\_\_\_\_\_\_\_\_\_\_\_\_

专业：\_\_\_\_\_\_\_\_\_物联网工程\_\_\_\_\_\_\_\_\_\_\_\_

年级：\_\_\_\_\_\_\_\_\_\_\_\_15级\_\_\_\_\_\_\_\_\_\_\_\_\_\_

班级：\_\_\_\_\_\_\_\_物联网工程4班\_\_\_\_\_\_\_\_

学号：\_\_\_\_\_\_\_\_\_3115001471\_\_\_\_\_\_\_\_\_

学生姓名：\_\_\_\_\_\_\_李子贺\_\_\_\_\_\_\_\_\_\_\_\_\_\_

指导教师：\_\_\_\_\_\_\_李学识\_\_\_\_\_\_\_\_\_\_\_\_\_\_

合作队友：\_\_\_\_\_3115001467\_江国洲\_\_

\_\_\_\_\_3115001330\_陈住辉\_\_\_

2018 年 5 月 17 日

**题目：基于NoSQL的科学数据大数据存储与统计系统**

一、设计目的

1. 数据库可以采用Hadoop、MonogoDB、Redis中的任意一种
2. 对提供的科学测量数据（大数据文件）进行分布式存储。要求对数据库采用分布式集群安装，伪分布式也可以接受（性能较低）
3. 初期可以用较小的科学测量数据文件《sample.txt》
4. 后期程序完善后，可以试验较大的科学测量数据文件《export\_low\_small.txt》
5. 采用Mapreduce进行科学测量数据的分布式读取
6. 在科学测量数据分布式读取之后，采用编程软件（Java、Python或者Matlab）进行数据作图和显示
7. 设计思路

整体思路为：mongodb用于数据的存储，pyecharts用于图形绘制，Flask框架用于web展示，ajax用于数据交互以及图形渲染（Flask + pyecharts + mongodb + ajax）

1、选定数据库，设计数据库的存储结构

2、设计查询方式，即Mapreduce的设计

3、先利用ipython notebook初步测试完成数据存储读取并画图的代码架构

4、利用轻量级Web框架Flask进行数据与图像的交互

三、运行环境

1、操作系统：linux deepin15.5桌面版

2、数据库：mongo v3.6.4  
3、语言环境：python v3.6.4、pymongo v3.6.1、flask v0.12.2、pyecharts

四、文件结构说明

.

├── clean\_data -- 便于画图以及交互从整体数据集里抽离的数据

│   ├── freq.txt -- 所有频率，用于构建web交互的select下拉框

│   ├── phi.txt -- 所有phi，用于构建web交互的select下拉框

│   └── theta.txt -- 构建画图的x轴所需的参数

├── docs -- 课设文件描述以及数据

│   ├── angle\_data\_plot.m

│   ├── export\_low\_small.txt

│   ├── NoSQL数据库课程报告\_模板.docx

│   ├── NoSQL数据库课程考核.docx

│   ├── NoSQL数据库课程考核-数据文件说明.doc

│   └── sample.txt

├── draw -- web展示的Flask代码

│   ├── \_\_pycache\_\_

│   │   └── server.cpython-36.pyc

│   ├── server.py -- 后台代码，用于响应请求

│   └── templates -- 前端代码html模板

│   ├── index.html -- 主界面，用于数据交互以及画图请求

│   └── pyecharts.html -- pyecharts画图的渲染界面

├── img -- 实验结果截图

│   ├── mapreduce查询结果.png

│   ├── mapreduce的python实现.png

│   ├── mapreduce构建.png

│   ├── mongodb存入数据的部分展示.png

│   ├── pandas读取数据部分展示.png

│   ├── web画图结果展示.png

│   ├── 加载绘图所需的参数选择.png

│   ├── 数据处理以及导入.png

│   └── 响应POST请求进行绘图.png

├── insert\_data.py -- 将数据导入至数据库

├── main.ipynb -- 初步思路的代码构建

├── NoSQL数据库课程报告.docx -- 实验报告

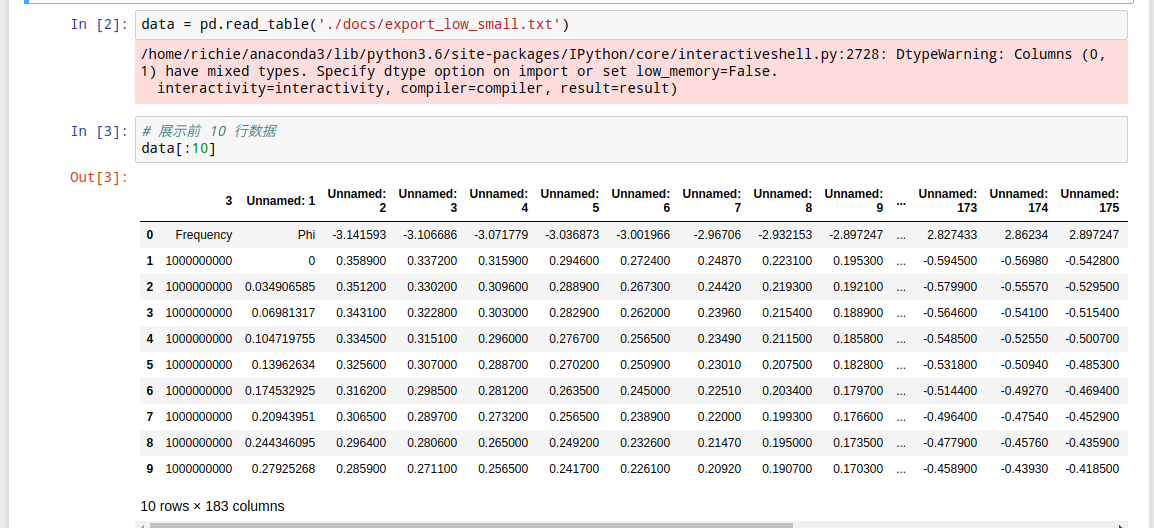
├── readme.md -- 部分文档说明

└── vedio.mp4 -- 程序结果录屏展示

五、设计步骤及结果测试

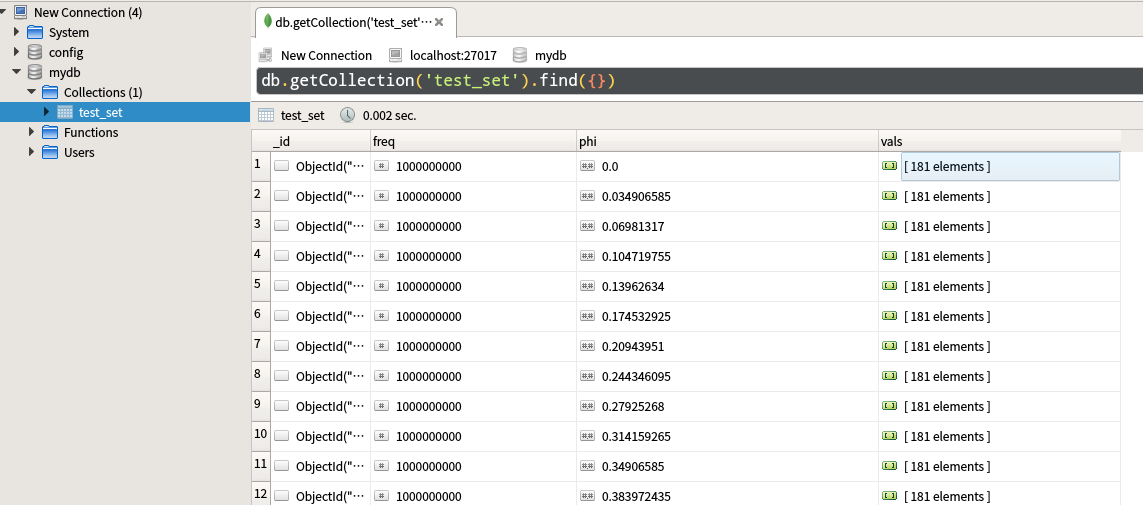
1、选定数据库，设计数据库的存储结构：

选定mongodb用于作为存储的数据库，观察数据发现



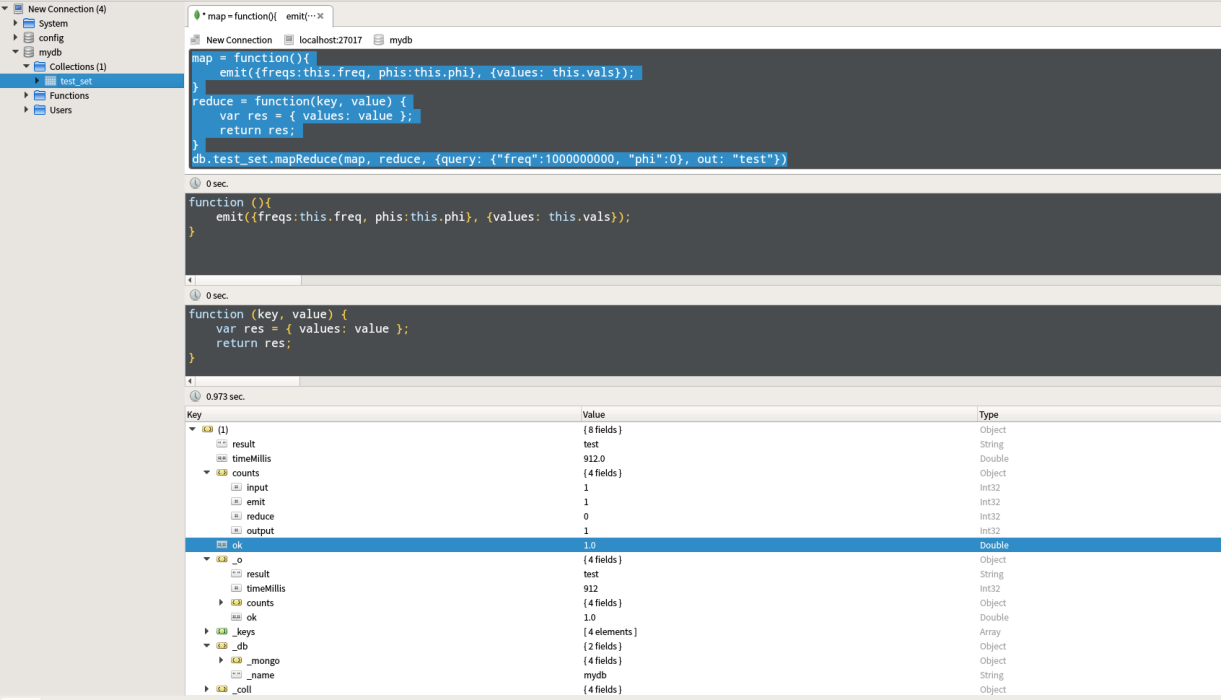
不同的frequency和phi会对应一串电场的值，而因变量theta与所有索引都有关联，故先直接抽离，phi的变化十分固定（范围从0～π），所以选择根据用Frequency+phi来索引一串数值序列（即电场值）

将数据插入mongodb以后效果如下所示



2、设计查询方式，即Mapreduce的设计：

Mapreduce的构建方式，对vals(数值序列)与(freq,phi)建立映射，reduce直接返回value的值没有做什么处理，根据query的限制来进行mapreduce查询（也是数据交互的原理）



Mapreduce查询结果

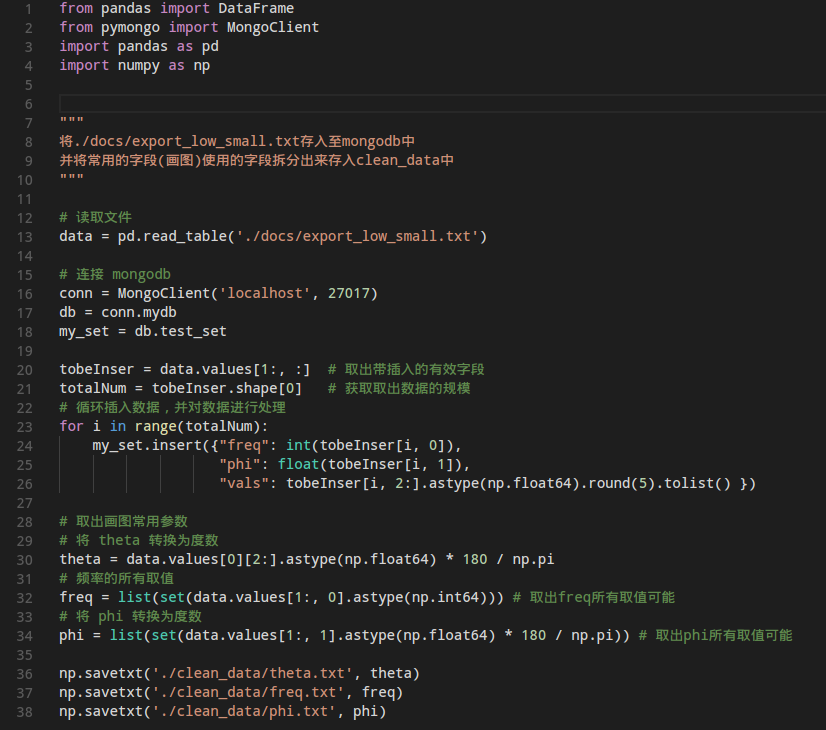


1. 先利用ipython notebook初步测试完成数据存储读取并画图的代码架构

Ipython notebook代码可在浏览器访问：<http://nbviewer.jupyter.org/github/15zhazhahe/nosql-test/blob/master/main.ipynb>

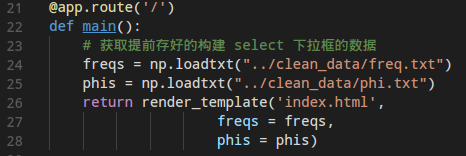
或使用Ipython notebook打开进行浏览

将初步的构想抽离为draw文件夹下的Flask代码（用于数据交互以及作图）和insert\_data.py（用于数据处理以及导入）



1. 利用轻量级Web框架Flask进行数据与图像的交互

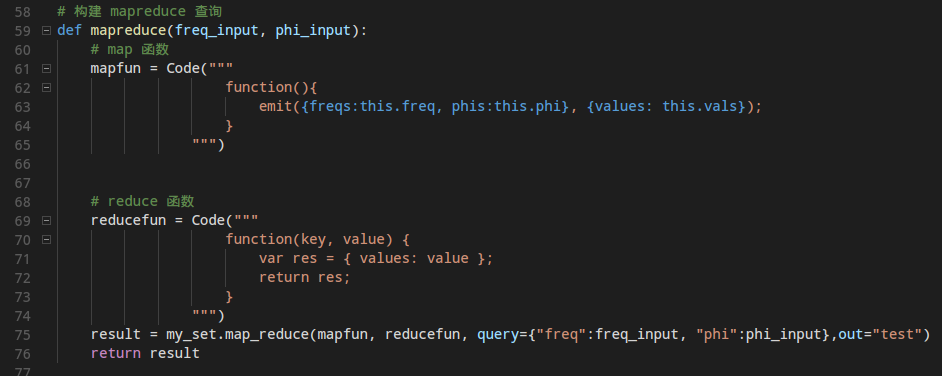
主页渲染，加载图形绘制所需要的参数（即select下拉列表的数据）



对POST请求进行响应，通过POST提交的数据来进行mapreduce查询，用查询的结果来绘制图像



用python构建mongodb的mapreduce函数，用于满足查询的需求



Web展示结果如下，除了基本的数据展示外，还标记了最大值和最小值



六、参考文献

[1] MongoDB教程|菜鸟教程(http://www.runoob.com/mongodb/mongodb-tutorial.html)  
[2] HTML教程|W3school(http://www.w3school.com.cn/html/index.asp)  
[3] AJAX教程|W3school(http://www.w3school.com.cn/ajax/index.asp)  
[4] Flask中文文档(https://yiyibooks.cn/yiyi/flask\_011\_ch/index.html)  
[5] pyecharts 中文文档](http://pyecharts.org/#/zh-cn/)  
[6] 博客:MongoDB:MapReduce基础及实例(https://www.cnblogs.com/Joe-T/p/4264910.html)  
[7] 《NoSQL数据库技术实战》皮雄军 著