2017.5.3

1.研究druid里面增加字段的问题。

2.hadoop各个组件的含义。

3.spark各个组件的含义。

4.server服务自测、压测。

5.SQL服务理清楚。

6.druid版本迭代，出方案。

7.统计线上服务。

8.druid优化方案。

问题整理：

1.数据结构版本迭代的问题。

2.druid优化的角度。Broker缓存，historical冷热部署，segment聚合时间粒度调整，灌入实时数据分担压力。

3.结构建模，统计字段所有维度，还是统计字段分开维度。

4.日常出问题，一些解决方法。比如查看源码，官方文档等等。如何加深自己对druid的理解。

数据结构上，涉及到版本迭代的问题，每次版本迭代，都需要新增一些字段，而新增的字段，在老数据里面是没有新增的字段的。例如：dim1这个字段是新增的，新数据里面的value取1,2，而同时假设老数据里面dim1的值默认是1，这样要求，dim1=1做筛选的时候，筛选出新旧数据，dim1=2,只筛出新数据。这种情况，有没有固定的解决方案呢？

我们前面提到的，segment可以根据业务进行时间粒度的调整，这算是优化的一方面，我从网上搜资料，historical可以进行冷热部署，以及broker进行外部缓存

需要完成的任务：

1.服务化druid查询API。（自测，提测，研究自动化部署平台部署应用）

2.druid版本迭代，version字段，进而评审和实施。

3.druid优化方案设计，进而评审和实施。

4.SQL解析文档解析描述。