华为实习总结

# 基本情况

姓名:袁浩 工号:81020709 学号:2013302580115 项目:MCS 2.0

岗位:软件开发工程师(我主要进行大数据集群环境的搭建以及系统和数据库性能优化)

# 项目介绍

MCS 2.0是基于老的MCS 1.0项目迁移而来的,主要的任务是.

1. 迁移MCS 1.0的内容进来
2. 看网能力
3. 存量分析
4. LTE业务与4K视频承载业务

项目预期:整个项目的预期是能够建立一个基础的存量信息库,不仅仅是一个平台,还要能够有自定义的挖掘和学习的能力.

# 项目进度及完成情况

截至今日,项目已经进入第四个迭代(每个迭代持续大概一个月).

第一轮迭代:数据库模型的建立(我未参与)

第二轮迭代:网络数据报表分析和生成(我部分参与)

第三轮迭代:流量分析及系统性能优化,MCS 1.0迁移(我全程参与)

第四轮迭代:性能优化,uNetBuilder数据导入(正在参与)

第一轮迭代主要就是建立模型,数据库,准备一些数据,确定数据库结构,以及弄一些数据进去.

第二轮迭代更进一步,需要在第一轮迭代的基础上构建起基本的统计分析和生成报表的能力,要求生成可阅读的Excel表格或者Word文档.不仅如此,还要求能够在前端进行直观的展现.

第三轮迭代加入了流量数据,要求能够对数据进行流量分析,还原网络基本拓扑,生成热力图等等.其维度主要是国家地区代表处和运营商等等.

第四轮迭代才刚刚开始,但是这一轮迭代除了继续进行性能优化之外,还要开始导入uNetBuilder的大量数据.不仅如此,还要在MCS 2.0这个平台上构建起一些应用,用来检验平台的架构的扩展性和模型的正确性.

其中,我参与了迭代二和迭代三的项目演示.

第二轮迭代的演示是很成功的.

第三轮迭代的演示不是很成功,主要是数据量还不够大.

# 我所做的工作

第二轮迭代:

最开始的时候我就看了两天书,是关于HTTP协议的内容,讲得很深入,我强忍着困意看完了一本书之后,感觉收获并不多,因为这本书本身类似于协议标准,很多内容偏于记忆,是用来查阅的,而不适合用来学习.

过了两天,就开始阅读代码,java代码一般都有较高的抽象程度和较为深的接口封装,有时候你都不知道到底实现在哪里,在具体的项目中有很多是多线程调用执行的代码,并不是直接函数调用,所以代码执行路径并不是很清晰.看起来有一些困难,而且缺少文档,所以费了不小的力气.

之后开始了解本部门的业务,说起来我们部门的业务还是很重要的,由于担心涉密,就不介绍业务具体内容了.

抽空温习了一下Web Service,下个周在圈子里面分享.

之后开始了解项目的数据库结构了,比如数据库表的定义,数据库字段的名称和含义以及数据类型.导师最开始也只是交给了我一些比较简单的工作,比如了解一些设备的规格,比如NE40这个路由器,容量多大,有多少单板,每个单板有哪些端口,每块单板有何作用,可以承载什么业务等等问题.

第三轮迭代:

这轮迭代开始了解spark框架和mybatis这个工具.

Mybatis是一个数据库操作的工具,用来帮助我们构建sql语句以及将数据库查询结果转换为对象.在学习Mybatis的过程中,遇到的问题大多是不匹配造成的,比如数据类型不正确,id和函数名不匹配,没有默认构造函数,没有getter,setter方法等等,在查询了一些错误之后,得以熟悉了Mybatis的使用.

Spark是一个高性能的数据分析引擎,是从Hadoop发展过来的,主要进行两类操作,transformation和action,对应于Hadoop的Mapreduce,spark框架也是非常的复杂,他拥有一个类似于Hadoop的Hbase的DataFrame组件,负责将分布式数据映射为数据库表的形式,Hadoop的Hive和Spark的Shark是sql解析引擎,负责将sql语句解析为MapReduce操作.当然Spark和Hadoop还有很多其他的内容,比如HDFS,比如Spark SQL,比如Spark Streaming,非常复杂.另外Spark还集成了机器学习框架MLlib和图算法框架GraphX.

其后开始解决在进行项目移植的过程中Spark出现的各种问题,比如jar包缺失,jar包版本冲突,动态链接库缺失,无法读取本地配置文件等等.

之后开始学习提交代码,第一次提交代码的心情还是有些激动的,第一次提交代码之前调试了很久,确认不会出现异常才提交.

这个星期主要写的代码就是对数据库中的四个表定义访问数据库实现增删查的接口,接口是用Mybatis定义的配置文件,对象通过spring进行初始化,数据库连接也配置在XML文件中.

第一次编写实际的代码还是有些不太一样的,尤其是定义接口和多层封装,和以前写的代码都不太一样.

除了完成这些接口的编写之外,我还要负责在数据库对象和业务对象之间进行转换.由于对象比较复杂,转换的过程也是比较繁琐,尤其是组装对象的时候,必须要通过某些字段来进行关联,其间很容易出现错误,导致异常的发生.

这星期的最后一天开始部署spark到公司的集群里面,公司的集群有4台机器,一台master节点3台slave节点,4台机器都是Linux系统的,只有命令行界面,配置集群的过程并不是非常复杂,主要遇到的问题就是集群启动之后slave节点的worker不能和master节点连接,最后经过输出地错误日志进行问题定位,发现是由于网络配置不正确,在进行ip配置的时候没有使用严格的主机名称,导致节点之间无法连接.

集群部署之后我运行了一个小的spark程序,这是一个计算PI的程序,成功的启动了起来.

第四轮迭代:

当我在编写这份文档的时候第四轮迭代还没有结束,所以只完成了一部分工作.

我开始将Hadoop和Yarn部署在集群上.

其间由于SSH配置的问题,导致几台机器之间不能连接启动,最后重新生成了RSA密钥对.

后来在提交的过程中,由于命令行参数写错了,出错了很久才解决.

Spark的问题基本解决了,以后每周都要做一下性能测试,可能以后还要做代码移植啥的,希望我的工作可以帮助后来的人吧.

关于Redis和数据库之间的性能比较,大致的结果也得出了,数据库肯定是比Redis要慢的,但是在不查询idList的情况下,数据库肯定是要快很多的,所以最终决定全面迁移到数据库上面去.

# 总结与感想

在深圳的生活异常艰苦.

受限于资金,我们只能多人合租,而且我选择了一间没有床的房间,就在地上铺了一张凉席,席地而眠.

深圳的物价也很高,平时消费都要精打细算,工资几乎入不敷出.可以说,对于我们来说,在深圳的每一天,都是一种锻炼.

公司8:00就要求我们上班,我每天6:30起床,6:45出门,7:20到公司,7:45吃完早饭,7:50到工位,接个水就7:55.晚上一般12点入睡.

但是,这些生活上的不便丝毫都没有影响到我工作的热情.

很快,我就开始参与到实际的项目中了,也完成了几个任务,我的任务主要都是围绕性能优化来进行的,每天都要分析大量的日志,进行各种优化.

有时候,最困难的时候,一连几天都没有进展,进度几乎停滞,我眉头紧锁,思索可能的方案.

当然也有意气风发的时候,一下子性能提升了接近一倍,顿时觉得欢欣鼓舞.

金鳞岂是池中物,一遇风云便化龙.我一直不知道,我的风云在哪里.直到初来的第二个星期,台风来临,我站在沿海的礁石上,向东而望,只见风起云涌.

这个时候,我终于知道,我的风云,就在这里,就在华为.

实习了一个月了,就要开始答辩了,这个答辩,理论上决定了我这一个月的努力是否能够得到认可,也决定了我是否能够留在华为,所以,我还是很认真的准备了一番.做了一个高大上的PPT,准备好了一些内容,也把自己解决问题的能力展现了出来.

在我的主管问我我的奋斗目标是什么的时候,我忽然想起了我最近看到的一句话:为天地立心,为生民立命,为往圣继绝学,为万世开太平.脱口而出,我不知我的主管是否被我的样子吓到了,但我知道,那就是那一刻我最真实的想法.

来华为这么久了,最大的变化,就是眼光不一样了,在这里,你会用更高的视野来看待问题,而不必拘泥于得失.

终究,我还是要证明自己的.

很久很久之前,我就知道,我的心底蕴藏着磅礴的能量,只要与我眼神对视,便可以看到.一直以来,我都不敢去触碰那股力量,因为它太强大,而我并没有十全的把握掌控它,我担心一旦我开动用了这股力量,就会因为控制不住它,最后伤人伤己.

但终究,这是我的宿命,躲不过,逃不掉,我生来就是要掌控它的.

是时候,释放它了.