2015/08/13

第一家校招 第二家 盛世全景 第三家 华胜天成

北京XXX宇有限公司

1. Hashmap hashset treemap 区别

HashSet 实现了Set 接口 ，值不能重复，将对象存储挨HashSet 前，需要重写

Equals 和hashCode ,这样才能比较对象的值是否相等，确保set 中没有存储相等的对象，

1. 什么是hashMap

Hashmap 实现了map 接口，map接口对键值对进行映射，Map 不允许重复的键位，HashMap和TreeMap。TreeMap保存了对象的排列次序，而HashMap则不能。HashMap允许键和值为null。HashMap是非synchronized的，但collection框架提供方法能保证HashMap synchronized，这样多个线程同时访问HashMap时，能保证只有一个线程更改Map。



1. 常用标签html

<li> <lo> <font><table><td><tr><br><form> <textarea> <imgsrc=”图片地址”>

<i> <u> <frame>

3 lunce 和solr

**。** 有完整的API族以及对应的实现。你可以利用这些在自己的应用里实现高级查询（基于倒排索引技术的），Lucene对单机或者桌面应用很实用很方便。**但是**Lucene，需要开发者自己维护索引文件，在多机环境中备份同步索引文件很是麻烦。于是，就有了Solr。

1. storm 流处理数据丢失

我猜测是bolt 和bolt 之间通信出了问题 ，zmq

盛世全景面试官(主要是我讲 )

1 .Mapreduce 过程

2. 你的集群多大 每天流量多少

3.你集群中定时任务是怎么做的

Liunx 执行任务的方法

1. 在LINUX中你应该先输入crontab -e，然后就会有个vi编辑界面，再输入0 3 \* \* 1 /clearigame2内容到里面 :wq 保存退出。
2. 在LINUX中，周期执行的任务一般由cron这个守护进程来处理[ps -ef|grep cron]。cron读取一个或多个配置文件，这些配置文件中包含了命令行及其调用时间。

cron的配置文件称为“crontab”，是“cron table”的简写。

1. /var/spool/cron/ 这个目录下存放的是每个用户包括root的crontab任务，每个任务以创建者的名字命名，比如tom建的crontab任务对应的文件就是/var/spool/cron/tom。  
   一般一个用户最多只有一个crontab文件。
2. /etc/crontab 这个文件负责安排由系统管理员制定的维护系统以及其他任务的crontab，/etc/cron.d/ 这个目录用来存放任何要执行的crontab文件或脚本。
3. 6. 在crontab文件中如何输入需要执行的命令和时间。该文件中每行都包括六个域，其中前五个域是指定命令被执行的时间，最后一个域是要被执行的命令。  
    每个域之间使用空格或者制表符分隔。格式如下：   
   minute hour day-of-month month-of-year day-of-week commands   
       合法值 00-59 00-23 01-31 01-12 0-6 (0 is sunday) commands（代表要执行的脚本）  
       除了数字还有几个个特殊的符号就是"\*"、"/"和"-"、","，\*代表所有的取值范围内的数字，"/"代表每的意思,"/5"表示每5个单位，"-"代表从某个数字到某个数字,","分开几个离散的数字。

每天早上6点   
0 6 \* \* \* echo "Good morning." >> /tmp/test.txt //注意单纯echo，从屏幕上看不到任何输出，因为cron把任何输出都email到root的信箱了。

每两个小时   
0 \*/2 \* \* \* echo "Have a break now." >> /tmp/test.txt    
  
晚上11点到早上8点之间每两个小时和早上八点   
0 23-7/2，8 \* \* \* echo "Have a good dream" >> /tmp/test.txt  
  
每个月的4号和每个礼拜的礼拜一到礼拜三的早上11点   
0 11 4 \* 1-3 command line

1. 监控系统做过没

Ganglia 监控系统 或者集群

第三家 华胜天成

问了mapreduce 过程

Mapreduce 优化

Hbase 执行过程

Hbase 优化

Spring 控制反转 依赖注入 ，AOP 及其用途

关系型数据库是怎么把数据导出到Hbase 里的

尚学堂出来的都知道

Hive 的优化

Hive 左右连接

Hive 使用的三种形式

Spark RDD 是什么

Spark 算子

2015/8/14

联龙博通

Hive 分区

1 .单分区建表语句：create table day\_table (id int, content string) partitioned by (dt string);单分区表，按天分区，在表结构中存在id，content，dt三列。

2 .双分区建表语句：create table day\_hour\_table (id int, content string) partitioned by (dt string, hour string);双分区表，按天和小时分区，在表结构中新增加了dt和hour两列。

**添加分区表语法**（表已创建，在此基础上添加分区）：ALTER TABLE table\_name ADD  
partition\_spec [ LOCATION 'location1' ]  
partition\_spec [ LOCATION 'location2' ] ...

3.ALTER TABLE day\_table ADD  
PARTITION (dt='2008-08-08', hour='08')  
location '/path/pv1.txt'

4. **删除分区语法**：ALTER TABLE table\_name DROP  
partition\_spec, partition\_spec,...

用户可以用 ALTER TABLE DROP PARTITION 来删除分区。分区的元数据和数据将被一并删除。例：

ALTER TABLE day\_hour\_table DROP PARTITION (dt='2008-08-08', hour='09');

**数据加载进分区表中语法**：

LOAD DATA [LOCAL] INPATH 'filepath' [OVERWRITE] INTO TABLE tablename [PARTITION (partcol1=val1, partcol2=val2 ...)]

例：

LOAD DATA INPATH '/user/pv.txt' INTO TABLE day\_hour\_table PARTITION(dt='2008-08- 08', hour='08'); LOAD DATA local INPATH '/user/hua/\*' INTO TABLE day\_hour partition(dt='2010-07- 07');当数据被加载至表中时，不会对数据进行任何转换。Load操作只是将数据复制至Hive表对应的位置。数据加载时在表下自动创建一个目录  
基于分区的查询的语句：SELECT day\_table.\* FROM day\_table WHERE day\_table.dt>= '2008-08-08';

**查看分区语句**：

hive> show partitions day\_hour\_table; OK dt=2008-08-08/hour=08 dt=2008-08-08/hour=09 dt=2008-08-09/hour=09

其他的老师给的资料都有都是重复的