**脏词库**

[gservice@192.168.4.68](mailto:gservice@192.168.4.68) 密码：142857@xcwen

pwd：192.168.4.68:/opt/taomee/gservice/dirty/

外网数据库看程序配置

代码在<http://svn.taomee.com/framework/tserver/dirty_agent/trunk/>

**神奇密码**

代码和文档都在svn上

神奇密码 http://svn.taomee.com/platform/card/magic-card/

时空密码 <http://svn.taomee.com/platform/card/time-card/>

对应在10.1.1.63上的代码在10.1.1.63:/home/ping/card/ （time-card下代码提交到svn上有问题。。。）

数据库在10.1.1.28：

mysql -uroot -pta0mee --auto-rehash --default-character-set='utf8' -h 10.1.1.28

**ip库**

windows下用根据将access的库导出成sql语句，在192.168.71.45上插入库中

每次新建一个表，类似t\_city\_ip\_2015\_Q3

mysql -h192.168.71.45 -uhadoop -pHA#2jsOw%x -P3306 --default-character-set='utf8' db\_ip\_distribution\_12\_Q1

登陆脚本在192.168.11.129:~/bin/db\_ip

将192.168.11.129: /opt/taomee/hadoop/register/calc\_register\_transfer.sh中配置的表改成新的

**Hadoop压缩**

现在集群直接的压缩格式有2种：gz和bz2

gz不支持分块，压缩率高。当一个gz文件大于hadoop配置的块大小时，在hadoop计算时，需要把所有数据传输到一台node上先解压，再计算，会严重影响效率。bz2格式支持分块，压缩率比gz低，可以多台node同时解压。现在的策略是当压缩后的文件比64M小时，使用gz，否则使用bz2。

目录在192.168.11.129: /opt/taomee/hadoop/compression

merge.sh 老统计结果数据

statistic.sh 多区服源数据

stat.sh 老统计源数据

test.sh 数据平台源数据

wireless.sh 无线源数据