统计平台项目组各系统说明

目录

[统计平台项目组各系统说明 1](#_Toc380069468)

[1、stat.taomee.com 2](#_Toc380069469)

[1.1、系统结构图以及各模块介绍 2](#_Toc380069470)

[1.1.2、stat\_proxy 3](#_Toc380069471)

[1.1.3、stat\_server 4](#_Toc380069472)

[1.1.4、db:mysql 5](#_Toc380069473)

[1.1.5、stat\_task 6](#_Toc380069474)

[1.1.6、misc & new-misc 6](#_Toc380069475)

[1.1.7、stat-daily(统计日报) 6](#_Toc380069476)

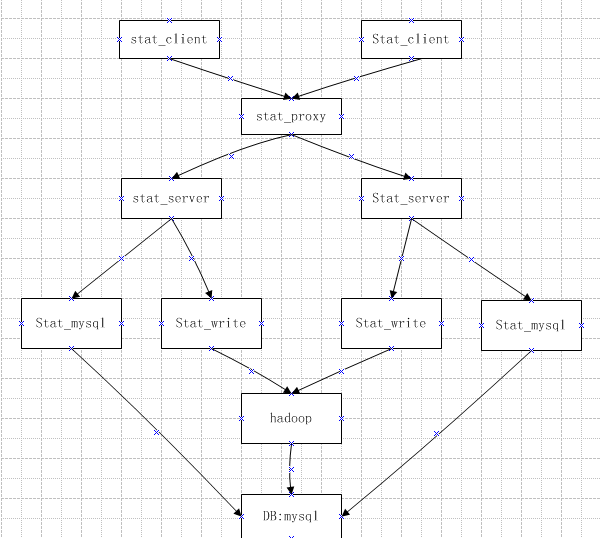
[2、tongji.taomee.com 6](#_Toc380069477)

[3、data.taomee.com 7](#_Toc380069478)

[4、wireless-stat.taomee.com 8](#_Toc380069479)

# 1、stat.taomee.com

## 1.1、系统结构图以及各模块介绍

1.1.1、stat\_client

用于将业务落下来的统计消息发送出去,通常在运维装机时都会默认部署好统计客户端的;如遇运维同事未部署或部署错误的情况下，需要按照如下步骤进行部署

* 获取需要安装统计客户端的机器IP地址以及所在的机房，并确保建立有stat账户，根目录为/opt/taomee/stat/,账户密码跟其他stat帐户密码统一
* 进入数据库192.168.71.68:3307 db\_stat\_config(可通过登录192.168.71.56执行srv\_login.sh快速进入)，检查IP地址是否已经加入到了表t\_host\_info\_tmp以及配置是否正确,其中proxy\_ip和proxy\_port这两个字段指定了该客户端应该链接哪个stat\_proxy服务和端口,另外需要检查客户端到对应的proxy\_ip的路由是否通
* 登入要部署客户端的机器，进入/opt/taomee/stat目录，从192.168.71.56上将/opt/taomee/stat/backup/stat-bak/stat\_client.tar.gz拷贝过来并解压,解压后将spool目录及其子目录的权限改为777
* 进入程序的配置目录/opt/taomee/stat/stat-cli-new/conf/，确保配置文件中的remote\_host和remote\_port这两项的配置是否与数据库中的t\_host\_info\_tmp表记录的一致，数据库中写的是IP，配置文件中是域名;其他配置项通常不需要更改
* 回到程序目录/opt/taomee/stat/stat-cli-new，执行start\_client.sh启动统计客户端

1.1.2、stat\_proxy

统计中转服务，负责接收统计客户端的消息，并将消息根据消息ID转发出去，转发规则是先读取数据库配置表t\_route\_info\_tmp中的记录

* 如消息ID不在该配置表中，那么会判断是否在t\_message\_translate表中，如果在，那么就根据服务端的数量，对消息ID取模运算后转发给对应的一个服务端,服务端的数量是读取配置表t\_server\_list\_tmp获取的；如果不在，会直接返回给统计客户端，由统计客户端将消息写入/opt/taomee/stat/spool/unrouted/该目录下
* 如消息ID在该配置表中，且对应的IP和PORT为0，那么不进行转发，直接将消息退回给统计客户端,统计客户端会将这类消息内容写入/opt/taomee/stat/spool/unrouted/该目录下
* 如消息ID在该配置表中，且对应的IP和PORT不为0，那么就将该消息转发给对应的IP和PORT的服务上，当前有3个消息ID是属于这种情况,分别是
  + - 0x1:校巴newsfeed的数据，转发到天津123.150.163.2:9121，由该服务直接转发给校巴项目部提供的一个服务192.168.37.55：7000
    - 0x3:收集B03的无线统计日志，转发到北京211.151.105.182：9203,由该服务转发到192.168.21.135:65533这个服务，由这个服务将日志写到本地
    - 0x05210A00:收集赛尔号的登录IP，转发到上海192.168.1.72：9121，由该服务转发到192.168.1.72: 65533这个服务，由这个服务写入本地文件

因统计中转是部署在各个机房的，因此要求配置数据库开放外网访问权限

因统计中转和服务端是走外网通信，故要求这些中转机器开通外网网卡

公司当前有5个机房，我们的中转服务部署情况如下：

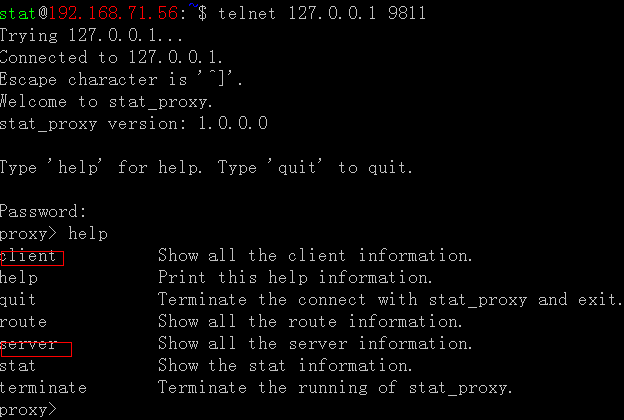
|  |  |  |  |
| --- | --- | --- | --- |
| 机房 | IP：PORT | 绑定的域名 | 部署目录 |
| 苏州 | 192.168.71.56:9900 | sz.proxy.taomee.com | /opt/taomee/stat/stat-proxy |
| 上海 | 192.168.1.72:9900 | sh.proxy.taomee.com | /opt/taomee/stat/stat-proxy |
| 北京 | 192.168.21.182:9900 | bj.proxy.taomee.com | /opt/taomee/stat/stat-proxy |
| 天津 | 192.168.33.2:9900 | tj.proxy.taomee.com | /opt/taomee/stat/stat-proxy |
| 佛山 | 192.168.112.134:9900 | fs.proxy.taomee.com | /opt/taomee/stat/stat-proxy |

在各个中转服务器还部署了stat-monitor-new程序，用于监控连接到中转服务的客户端和服务端的状态，有异常情况会发告警短信以及邮件，该程序是基于PHP实现的，访问中转服务的monitor\_ip和monitor\_port获取信息的

另外中转还提供了查看各个客户端和服务端状态的功能

telnet 127.0.0.1 9811

这里的IP和PORT通过配置文件中的web\_ip和web\_port配置即可，登录密码为123456

用较多的是client和server两个命令

* client可以查看连接到该中转服务的所有客户端IP，以及每个客户端的连接状态，最后一次发送消息的时间以及最后的消息ID，如果客户端有异常情况，会标红显示
* server命令会列出当前系统中存在的所有服务端，以及中转跟服务端之间的连接情况

中转服务对应的程序文件放在192.168.71.56: /opt/taomee/stat/backup/stat-bak/stat\_proxy.tar.gz 如果需要部署新的中转，拷贝过去解压并修改配置文件的IP地址即可

1.1.3、stat\_server

统计服务端负责将中转转发过来的数据解析成单个统计项，然后根据统计项的类型分发给stat\_mysql和stat\_write处理,这一步主要是读取t\_message\_translate、t\_report和t\_message\_protocol三张表完成的.当前系统中在使用的数据类型主要有：

* 实时的算术运算sum、max,转发给stat\_mysql处理,数据结果直接入库，均为分钟数据
* 唯一数去重运算ucount、唯一值运算uvalue、多区服统计日志转发给stat\_write处理,stat\_write会将唯一数和唯一值类型的数据写入/opt/taomee/stat/stat-writer/old\_stat\_data目录下，多区服的数据写入/opt/taomee/stat/stat-writer/data目录下,同时在相应的服务器上部署有hadoop的上传脚本/opt/taomee/hadoop/sync/sync\_data.sh、/opt/taomee/hadoop/sync/sync\_stat.sh负责将数据上传至hadoop以及删除脚本/opt/taomee/stat/stat-writer/del\_old\_files.sh，确保本地磁盘不会被写满，删除5天前的本地文件

当前统计服务端部署(记录在t\_server\_list\_tmp表中)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| IP:PORT | 对应内网IP | client\_seqno\_table\_name | stat\_mysql | stat-write | 用途 |
| 61.155.182.58:9200 | 192.168.71.58 | 24 | 192.168.71.58:8500 | 192.168.71.58:8888 | 正常的统计服务端 |
| 61.155.182.44:9200 | 192.168.71.44 | 26 | 192.168.71.44:8500 | 192.168.71.44:8888 | 正常的统计服务端 |
| 61.155.182.133:9200 | 192.168.71.33 | 25 | 192.168.71.33:8500 | 192.168.71.33:8888 | 正常的统计服务端 |
| 123.150.163.2:9121 | 192.168.33.2 | 21 | 无 | 无 | 转发校巴newsfeed数据 |
| 211.151.105.182：9203 | 192.168.21.182 | 20 | 无 | 无 | 转发无线统计的日志 |
| 192.168.1.72:9121 | 192.168.1.72 | 18 | 无 | 无 | 记录赛尔号IP |

因统计服务端和中转是走公网通信的，故需要绑定在外网地址上，另外上述表格中有个client\_seqno\_table\_name,表示用数据库表t\_client\_seqno\_26记录对应客户端上消息发送的位置信息，每个服务端的这项配置必须区分开。

从以上部分可以看出，**如果需要新增加一个统计服务端，需要按照如下操作**:

* + 确保服务器开通了外网网卡，将服务器信息记录到t\_server\_list\_tmp表中，检查是否有空闲的t\_client\_seqno\_表，如没有就找DBA建新表
  + 服务器程序放在192.168.71.56：/opt/taomee/stat/backup/stat-bak/ stat\_server.tar.gz,将该文件拷贝过来解压，修改相应的配置文件，bench.conf更改绑定地址和端口、 stat\_server.ini中对应的stat\_mysql地址和stat\_write地址，以及client\_seqno\_table\_name，其他项正常情况下无需更改
  + 配置stat账户和hadoop账户的crontab，确保stat-server按时重启以及数据能够上传至hadoop,crontab的配置方法在压缩包中,配置过程中需要检查路径是否正确
  + 按顺序启动stat\_mysql、stat-write、stat\_server，执行相应的启动脚本
  + 重启各个中转服务stat\_proxy

1.1.4、db:mysql

数据库当前的部署

|  |  |  |  |
| --- | --- | --- | --- |
| IP | port | 用途 | 是否监听外网 |
| 192.168.71.68 | 3307 | 统计配置库db\_stat\_config | 是 |
| 192.168.71.68 | 3306 | 统计结果库db\_stat\_result\_[1001-1010] | 否 |
| 192.168.71.73 | 3308 | 统计源库db\_stat\_report\_[1-24] | 否 |
| 192.168.71.72 | 3306 | 统计源库db\_stat\_report\_[51-60] | 否 |
| 192.168.71.35 | 3310 | 统计源库db\_stat\_report\_[25-30] | 否 |
| 192.168.71.35 | 3312 | 统计源库统计源库db\_stat\_report\_[31-50] | 否 |
| 192.168.71.35 | 3307 | 统计源库统计源库db\_stat\_report\_[61-80] | 否 |
| 192.168.71.35 | 3309 | 统计源库统计源库db\_stat\_report\_[81-100] | 否 |

1.1.5、stat\_task

1.1.6、misc & new-misc

misc服务于接收通过http协议发送过来的统计消息，由misc服务落统计项；目前部署有2套

* misc.taomee.com(192.168.1.177、192.168.33.132) 只接受老的统计项
* newmisc.taomee.com(192.168.1.136、192.168.33.81)，新添加的统计项走new-misc，格式为：

统计人数/人次： http://newmisc.taomee.com/misc.txt?type=0x13A0124c&mm=424790537&count=1

统计人数：http://newmisc.taomee.com/misc.txt?type=0x13A0124c&mm=424790537

统计人次：http://newmisc.taomee.com/misc.txt?type=0x13A0124c&count=1

type值即为消息id，需要重新分配

1.1.7、stat-daily(统计日报)

统计日报程序放在192.168.71.57: /opt/taomee/stat/stat-daily-new/v2/这个目录下

通过执行./stat\_daily\_v2.sh test发送测试日报

执行./stat\_daily\_v2.sh release发送正式日报

# 2、tongji.taomee.com

# 3、data.taomee.com

主要提供广告效果的检测数据，以及各个游戏的月报数据和KPI数据

数据源来自boss&account上传的日志以及通过web服务器nginx收集的日志信息(pv、uv信息)

PV、UV的日志信息收集：

通过在192.168.112.148和192.168.33.146这两台服务器上部署nginx服务器，然后定时(每小时一次)将nginx的日志上传至hadoop系统，在由hadoop的计算脚本进行计算

用户登录数据以及付费的收集:

由boss&account系统将用户登录和付费数据每小时上传一次至hadoop系统，在由hadoop的计算脚本进行计算

Hadoop计算脚本部署在192.168.11.128的/opt/taomee/hadoop/ads/mapred/work/这个目录下,主要的一个脚本 runPrevDay.sh，根据传入参数的不同，进行不同业务数据的计算

|  |  |
| --- | --- |
| 参数 | 说明 |
| pvuv | 调用runPvUv.sh计算每个广告带来的PV和UV数据 |
| newad | 调用runNewAds.sh计算每个广告给各游戏带来的新注册米米号数，然后根据米米号的注册来源计算对应的新增、活跃、付费数据；以及计算各个游戏的新增留存、活跃留存数据 |
| tp | 调用runTopics.sh,计算各个游戏月度数据,包括月报的数据 |
| actad | 调用runActAds.sh，目前这项数据已经基本无人查看了 |
| today | 同newad ，只不过这里计算的是当天的数据，会根据集群的状况决定是否需要进行计算 |

以上这些脚本均是通过crontab来触发的



# 4、wireless-stat.taomee.com

无线统计平台目前主要提供imole和iseer这两个游戏的数据，其他新开发的无线产品都会要求接入tongji.taomee.com平台

数据源的收集主要分为两类：

* imole通过http协议收集,通过部署在192.168.21.135上的wlstat.61.com来收集，写本地文件，然后上传至hadoop
* iseer通过游戏后台落stat.taomee.com里面的0x3日志，由统计系统的收集服务将数据收集到192.168.21.135，然后上传至hadoop

无线统计的计算脚本放在192.168.11.128: /opt/taomee/hadoop/wireless/run.sh这里;目前维护工作较少，主要是一个无线的日报程序，部署在10.1.1.60上，可参见日报发送的文档