* 简介



图 1

用户特征查询部分是dmp平台和DSP平台的一个接口。它的主要工作是：

1. 存储用户特征
2. 向DSP提供http接口，通过该接口输入用户cookie，可以返回该用户的特征
3. 统计dsp发送特征查询的uv、pv、ips数目
4. 统计dsp发送来的查询用户能够在特征库中查到数据的数目

为了完成这些功能，系统的设计如图1所示：



存储用户数据，目前实现的存储方式为cassandra存储

每个用户的一条特征，加上时间戳，存储在cassandra中。查询时输入用户cookie就能把该用户的所有特征都查询出来。

 从各地dmp平台查询用户特征。该功能通过rpc实现。对应java代码库中cn.clickwise.rpc部分。

统计dsp发送来的查询用户能够在特征库中查到数据的数目，即将dsp发送来的cookie在本地库中查找，统计总的用户数、有查询结果的用户数等

dsp平台通过http接口从系统查询用户数据，格式是：http://183.136.168.79:6579/queryUser/uid=fda077783ebe621112b6254c32d66e4b009

对dsp发送来的用户pv、ips、uv等进行统计。

* 部署

部署的中心是浙江用户特征库，各地dmp平台包括海南、山西、浙江等dmp平台

* 各地dmp平台：

在各地dmp分析平台有公网ip的机器上启动 rpc服务。

部署目录(如112.67.253.101 /home/clickwise/lq)下的文件如下：

log\_rpc.txt rpc.jar runtime.log start\_rpc.sh statistic\_keys user\_click.jar

启动命令是: sh start\_rpc.sh > log\_rpc.txt 2>&1 &

* Dmp提供dsp用户查询服务

该服务要有公网ip，在183.136.168.79机器上，/home/clickwise/liqi文件夹下。该目录下的文件如下所示：



Ardb.conf配置所用ardb的信息，如：

host=192.168.10.29

port=16379

db=10

cassandra.conf配置cassandra的信息，如：

host=192.168.10.29

port=9160

cfName=Users

keySpace=userstore

columnName=title

启动查询服务：

sh start\_feat\_server.sh > log\_server.txt 2>&1 &

在这台机器上还有步骤和步骤的统计，统计结果存储在mysql数据库中。数据库配置文件为mysql.conf，内容如下:

ip=192.168.10.29

port=3306

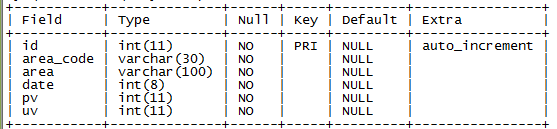
user=root

password=123

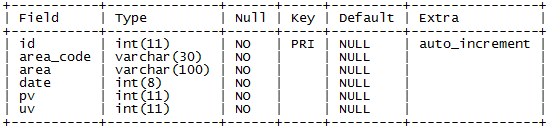
dbname=receipts

共有两张表InquiryReceipts和QueryReceipts，分别对应步骤和步骤的统计结果,

表结构如下：

InquiryReceipts 

QueryReceipts



Area\_code:地区编码

Area:地区名称

Date：日期

Pv：用户请求数

Uv: 不同用户的数目

启动统计服务的命令是：

sh cumulate\_feat\_query.sh > logcumulate.txt 2>&1 &

改程序每半小时统计一次结果，并存入数据库。

* 从各地dmp平台查询用户特征

该功能通过命令：java -cp feathouse.jar cn.clickwise.clickad.feathouse.ScheduleTask > log.txt 2>&1 &

改程序每天从各地dmp平台查询用户数据，并存入本地cassandra库。

各地dmp平台的配置在cn.clickwise.clickad.feathouse.EasyConfigureFactory进行。

修改下面方法：

**public** Dmp[] getDmps() {

// **TODO** Auto-generated method stub

Dmp[] dmps = **new** Dmp[3];

dmps[0] = **new** Dmp();

dmps[0].setName("山西DX");

dmps[0].setArea(**new** Area("山西DX", "023"));

dmps[0].setHost("219.149.148.86");

dmps[0].setUserFeatureTableName("auser\_cates\_keys");

dmps[0].setUidFieldName("uid");

dmps[0].setTmpIdentify("remote\_cookie");

dmps[0].setRpcPort(2733);

dmps[0].setDmpInquiryMethod("/hiveFetchTable");

dmps[0].setDmpStatisticMethod("/hiveStatisticByKeys");

dmps[0].setSourceTableName("astat");

dmps[0].setSourceUidFieldName("user\_id");

dmps[0].setSourceIpFieldName("sip");

dmps[0].setKeyTableName("statistic\_keys");

dmps[1] = **new** Dmp();

dmps[1].setName("海南DX");

dmps[1].setArea(**new** Area("海南DX", "009"));

dmps[1].setHost("112.67.253.101");

dmps[1].setUserFeatureTableName("auser\_cates\_keys");

dmps[1].setUidFieldName("uid");

dmps[1].setTmpIdentify("remote\_cookie");

dmps[1].setRpcPort(2733);

dmps[1].setDmpInquiryMethod("/hiveFetchTable");

dmps[1].setDmpStatisticMethod("/hiveStatisticByKeys");

dmps[1].setSourceTableName("astat");

dmps[1].setSourceUidFieldName("user\_id");

dmps[1].setSourceIpFieldName("sip");

dmps[1].setKeyTableName("statistic\_keys");

dmps[2] = **new** Dmp();

dmps[2].setName("浙江DX");

dmps[2].setArea(**new** Area("浙江DX", "030"));

dmps[2].s*etHost*("192.168.10.138");

dmps[2].setUserFeatureTableName("auser\_cates\_keys");

dmps[2].setUidFieldName("uid");

dmps[2].setTmpIdentify("remote\_cookie");

dmps[2].setRpcPort(2733);

dmps[2].setDmpInquiryMethod("/hiveFetchTable");

dmps[2].setDmpStatisticMethod("/hiveStatisticByKeys");

dmps[2].setSourceTableName("astat");

dmps[2].setSourceUidFieldName("user\_id");

dmps[2].setSourceIpFieldName("sip");

dmps[2].setKeyTableName("statistic\_keys");

**return** dmps;

}

其中setArea设置地区编码

s*etHost设置dmp平台的公网地址*

setUserFeatureTableName设置存储用户数据的hive表名

setUidFieldName 该hive表，用户标识的名称

setTmpIdentify临时文件前缀

setRpcPort rpc服务端口

setDmpInquiryMethod 查询表的方法

setDmpStatisticMethod 通过uid查询表的方法

setSourceTableName 存储用户更多信息的hive表

setSourceUidFieldName该hive表，用户标识的名称

setSourceIpFieldName 该hive表，源ip的名称

setKeyTableName 存储用户id的临时hive表名

* Cassandra平台的部署

共四台机器（浙江平台）：

192.168.10.28

192.168.10.29 seed

192.168.10.30

192.168.10.31 seed

将apache-cassandra-2.1.0-bin.tar.gz拷贝到四台机器的/home/lq文件夹下，解压改

名为cassandra

修改四台机器的cassandra.yaml

28配置：

cluster\_name:’feathouse’

seeds:”192.168.10.29,192.168.10.31”

listen\_address:192.168.10.28

rpc\_address:192.168.10.28

29配置

cluster\_name:’feathouse’

seeds:”192.168.10.29,192.168.10.31”

listen\_address:192.168.10.29

rpc\_address:192.168.10.29

30配置

cluster\_name:’feathouse’

seeds:”192.168.10.29,192.168.10.31”

listen\_address:192.168.10.30

rpc\_address:192.168.10.30

31配置

cluster\_name:’feathouse’

seeds:”192.168.10.29,192.168.10.31”

listen\_address:192.168.10.31

rpc\_address:192.168.10.31

各台机器以root 用户

mkdir –p /var/log/cassandra

chown –R hadoop:hadoop /var/log/cassandra

mkdir –p /var/lib/cassandra

chown –R hadoop:hadoop /var/lib/cassandra

各台机器启动服务

bin/cassandra > log\_cas.txt 2>&1 &

创建表：

bin/cassandra-cli -h 192.168.10.31 -p 9160

show keyspaces;

create keyspace userstore;

use userstore;

create column family Users with comparator=UTF8Type and

default\_validation\_class=UTF8Type and key\_validation\_class=UTF8Type;

set Users[476cb38e3aace0a5d129a147643d8bc3009]['name']='测试';

get Users[476cb38e3aace0a5d129a147643d8bc3009];