

Data Analytics in iPinyou

沈学华

x@ipinyou.com

iPinyou is the leading behavior targeting ads platform in China



Incorporated	2008
Offices	Beijing, Shanghai, Guangzhou, Silicon Valley

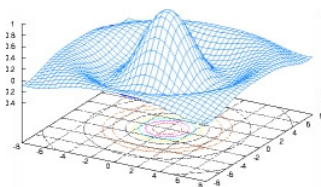
用户



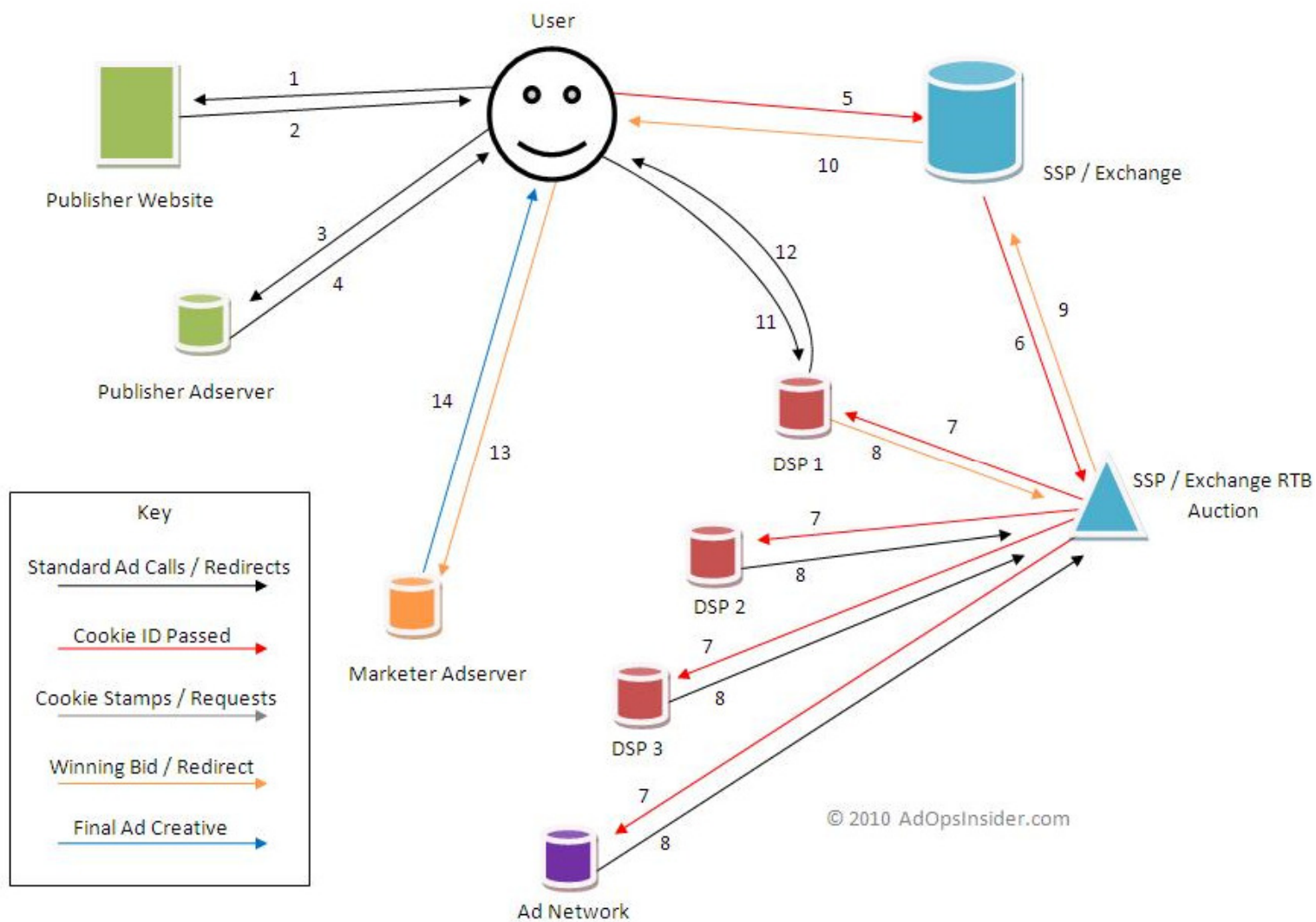
媒体

OPTIMUS优驰™

人群智能广告投放平台
广告主



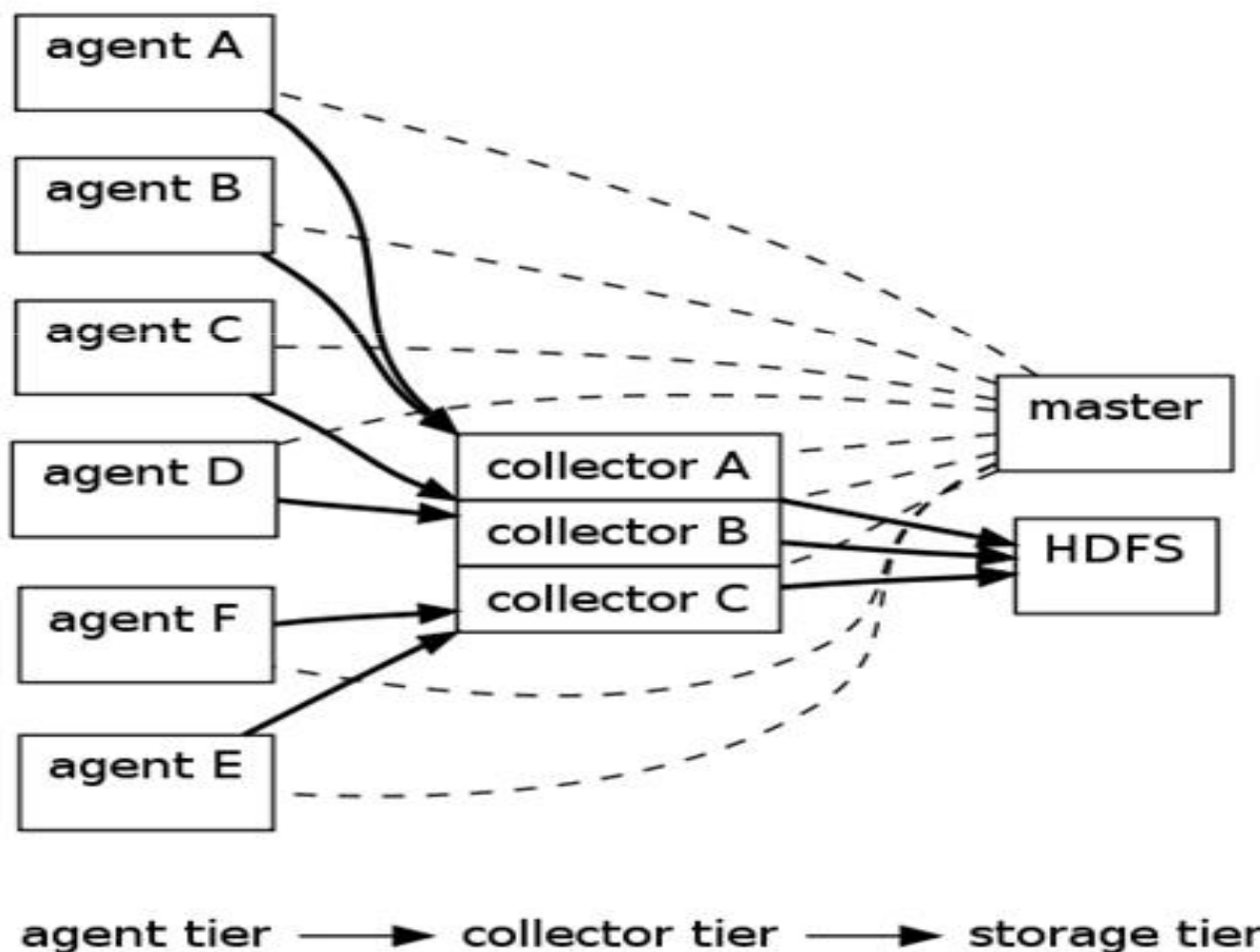




Data



Streaming Log Collection



CPU(%): 237.6 -> 77.7

Memory(%): 16.3 -> 10.2



SQL-to-hadoop

Easy to Use



Import (SQL -> Hadoop)

```
sqoop import \
```

```
--query 'SELECT a.*, b.* FROM a JOIN b on  
(a.id == b.id) WHERE $CONDITIONS'
```

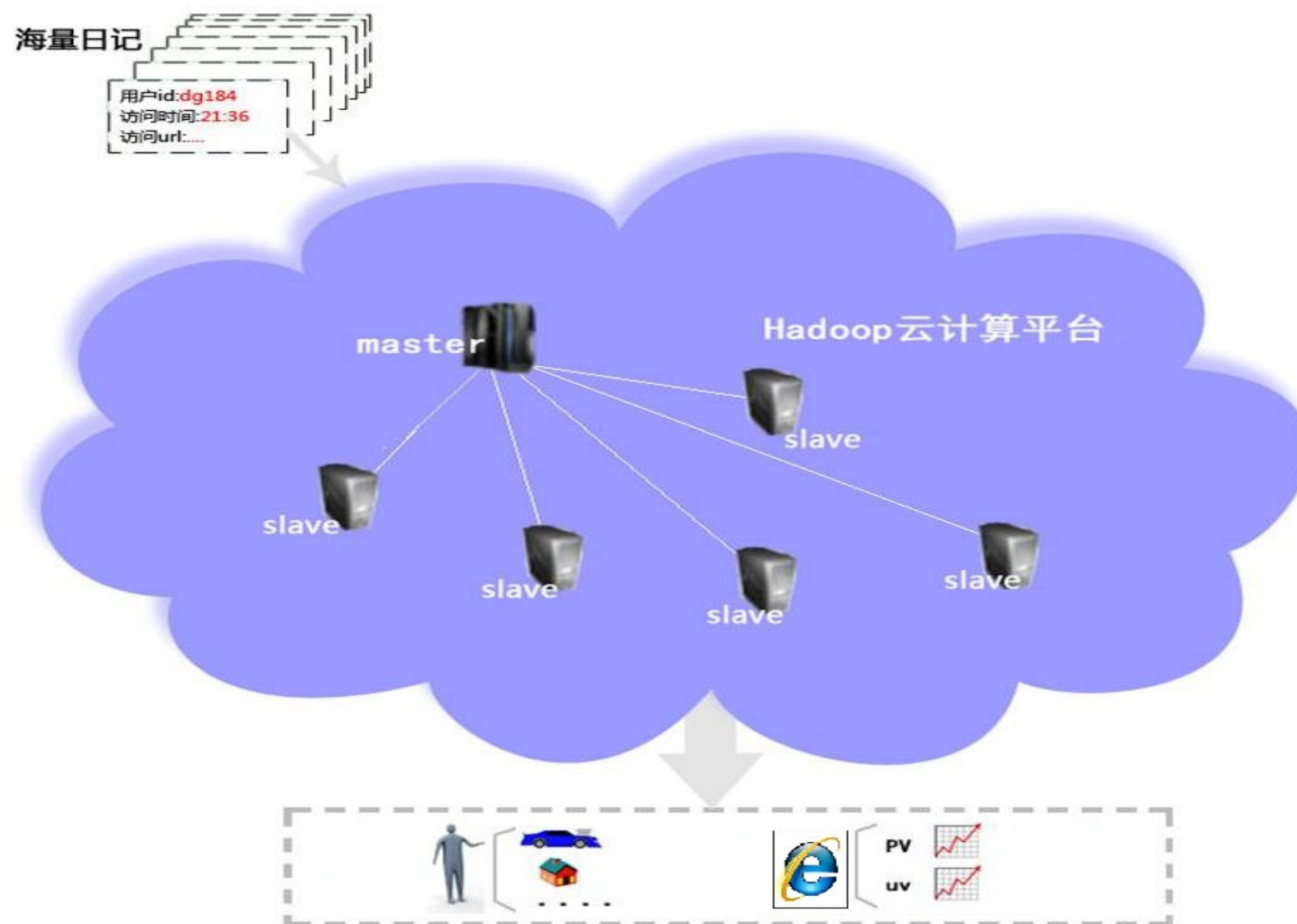
```
-m 1 --target-dir /user/foo/joinresults
```

Export (Hadoop ->SQL)

```
sqoop export --connect
```

```
jdbc:mysql://db.example.com/foo --table
```

```
bar --export-dir /results/bar_data
```





Easy to use

Time tradeoff

UDF



Documentation

Oozie Web Console (v1) [/oozie/]

Workflow Jobs									
System Info									
Instrumentation									
Coordinator Jobs									
All Jobs Active Jobs Done Jobs Custom Filter									
Job Id	Name	Status	Run	User	Group	Created	Started	Last Modified	
1 0000022-111109104424065-oozie-...	pig-wf	SUCCEEDED	0	root	users	Fri, 11 Nov 2011 08:06:54 GMT	Fri, 11 Nov 2011 08:06:54 GMT	Fri, 11 Nov 2011 08:07:46 GMT	
2 0000021-111109104424065-oozie-...	pig-wf	SUSPENDED	0	root	users	Fri, 11 Nov 2011 08:06:08 GMT	Fri, 11 Nov 2011 08:06:08 GMT	Fri, 11 Nov 2011 08:06:10 GMT	
3 0000020-111109104424065-oozie-...	pig-wf	SUCCEEDED	0	root	users	Fri, 11 Nov 2011 07:59:04 GMT	Fri, 11 Nov 2011 07:59:04 GMT	Fri, 11 Nov 2011 08:00:12 GMT	
4 0000019-111109104424065-oozie-...	pig-wf	SUCCEEDED	0	root	users	Fri, 11 Nov 2011 07:56:12 GMT	Fri, 11 Nov 2011 07:56:12 GMT	Fri, 11 Nov 2011 07:56:25 GMT	
5 0000018-111109104424065-oozie-...	pig-wf	SUSPENDED	0	root	users	Fri, 11 Nov 2011 07:53:36 GMT	Fri, 11 Nov 2011 07:53:36 GMT	Fri, 11 Nov 2011 07:56:44 GMT	
6 0000017-111109104424065-oozie-...	pig-wf	SUSPENDED	0	root	users	Fri, 11 Nov 2011 07:50:29 GMT	Fri, 11 Nov 2011 07:50:29 GMT	Fri, 11 Nov 2011 07:50:31 GMT	
7 0000016-111109104424065-oozie-...	pig-wf	SUCCEEDED	0	root	users	Fri, 11 Nov 2011 07:47:06 GMT	Fri, 11 Nov 2011 07:47:06 GMT	Fri, 11 Nov 2011 07:47:07 GMT	
8 0000015-111109104424065-oozie-...	pig-wf	SUCCEEDED	0	root	users	Fri, 11 Nov 2011 07:45:02 GMT	Fri, 11 Nov 2011 07:45:02 GMT	Fri, 11 Nov 2011 07:45:04 GMT	
9 0000014-111109104424065-oozie-...	pig-wf	KILLED	0	root	users	Fri, 11 Nov 2011 07:34:57 GMT	Fri, 11 Nov 2011 07:34:57 GMT	Fri, 11 Nov 2011 07:34:59 GMT	
10 0000013-111109104424065-oozie-...	oozie-pig	SUCCEEDED	0	hdfs	users	Wed, 09 Nov 2011 04:09:19 GMT	Wed, 09 Nov 2011 04:09:19 GMT	Wed, 09 Nov 2011 04:10:11 GM	
11 0000012-111109104424065-oozie-...	oozie-pig	KILLED	0	hdfs	users	Wed, 09 Nov 2011 04:06:51 GMT	Wed, 09 Nov 2011 04:06:51 GMT	Wed, 09 Nov 2011 04:10:04 GM	
12 0000011-111109104424065-oozie-...	oozie-pig	SUSPENDED	0	hdfs	users	Wed, 09 Nov 2011 03:49:21 GMT	Wed, 09 Nov 2011 03:49:21 GMT	Wed, 09 Nov 2011 03:52:28 GM	
13 0000010-111109104424065-oozie-...	oozie-pig	SUCCEEDED	0	root	users	Wed, 09 Nov 2011 03:47:48 GMT	Wed, 09 Nov 2011 03:47:48 GMT	Wed, 09 Nov 2011 03:48:40 GM	
14 0000009-111109104424065-oozie-...	oozie-pig	SUCCEEDED	0	root	users	Wed, 09 Nov 2011 03:38:27 GMT	Wed, 09 Nov 2011 03:38:27 GMT	Wed, 09 Nov 2011 03:39:20 GM	
15 0000008-111109104424065-oozie-...	oozie-pig	KILLED	0	root	users	Wed, 09 Nov 2011 03:36:46 GMT	Wed, 09 Nov 2011 03:36:46 GMT	Wed, 09 Nov 2011 03:36:56 GM	
16 0000007-111109104424065-oozie-...	oozie-pig	SUSPENDED	0	root	users	Wed, 09 Nov 2011 03:17:56 GMT	Wed, 09 Nov 2011 03:17:56 GMT	Wed, 09 Nov 2011 03:18:05 GM	
17 0000006-111109104424065-oozie-...	oozie-pig	SUCCEEDED	0	root	users	Wed, 09 Nov 2011 03:00:11 GMT	Wed, 09 Nov 2011 03:00:11 GMT	Wed, 09 Nov 2011 03:00:20 GM	
18 0000005-111109104424065-oozie-...	oozie-pig	SUCCEEDED	0	root	users	Wed, 09 Nov 2011 02:58:00 GMT	Wed, 09 Nov 2011 02:58:00 GMT	Wed, 09 Nov 2011 02:58:09 GM	
19 0000004-111109104424065-oozie-...	oozie-pig	KILLED	0	root	users	Wed, 09 Nov 2011 02:56:22 GMT	Wed, 09 Nov 2011 02:56:22 GMT	Wed, 09 Nov 2011 02:56:32 GM	
Page 1 of 2									
1 - 50 of 53									



LibSVM



Key-value storage

Real-time read/write



iRedis developed in iPinyou

Client-side sharding

Multiple sharding strategies



Cluster Report for Fri, 18 Nov 2011 14:51:17 +0800

Get Fresh Data

Metric Last Sorted

[Physical View](#)

Grid > [hadoop](#) > --Choose a Node

Overview of hadoop

CPUs Total:

Hosts up:

Hosts down:

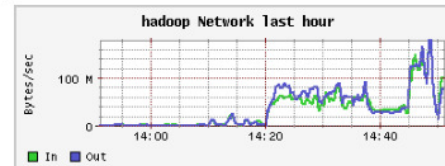
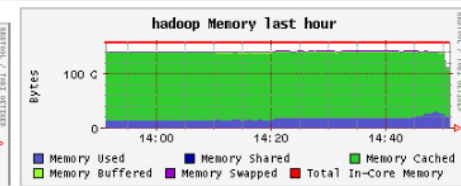
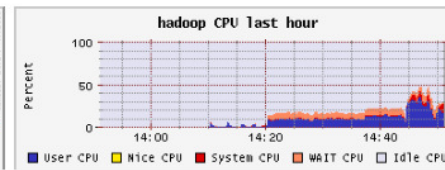
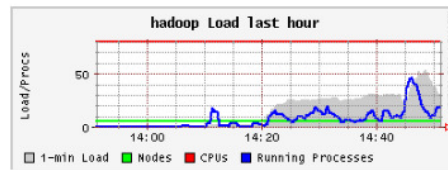
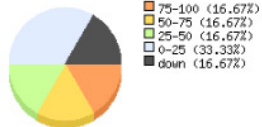
Avg Load (15, 5, 1m):

30%, 39%, 33%

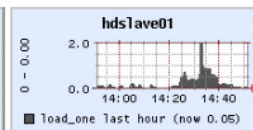
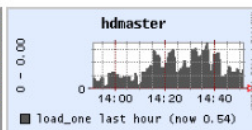
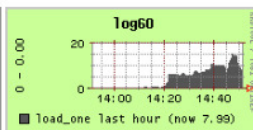
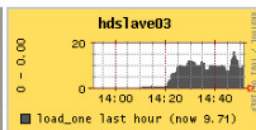
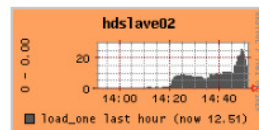
Localtime:

2011-11-18 14:51

Cluster Load Percentages



Show Hosts: ☒ yes ☐ no | [hadoop load_one last hour sorted descending](#) | Columns



Pentaho BI

Pentaho Dashboard

Pentaho Report Designer

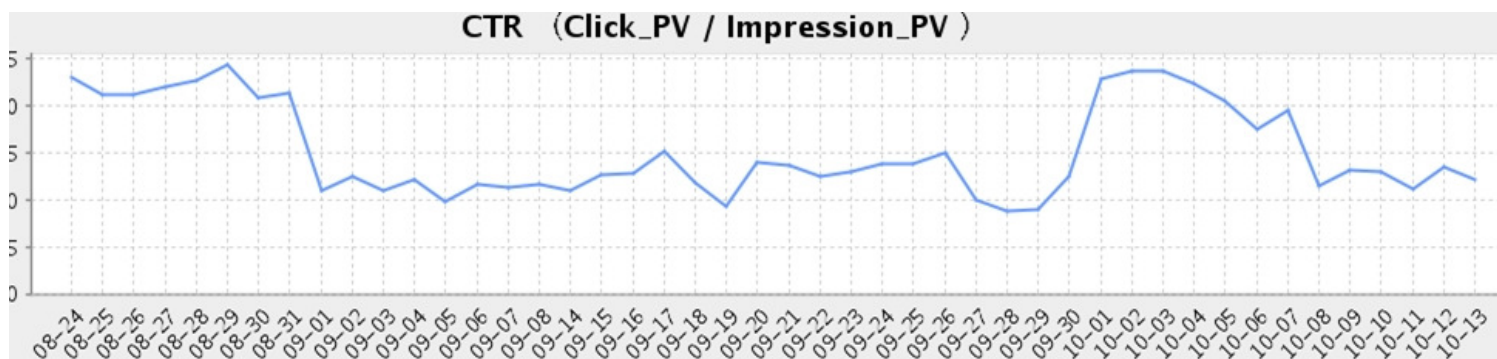
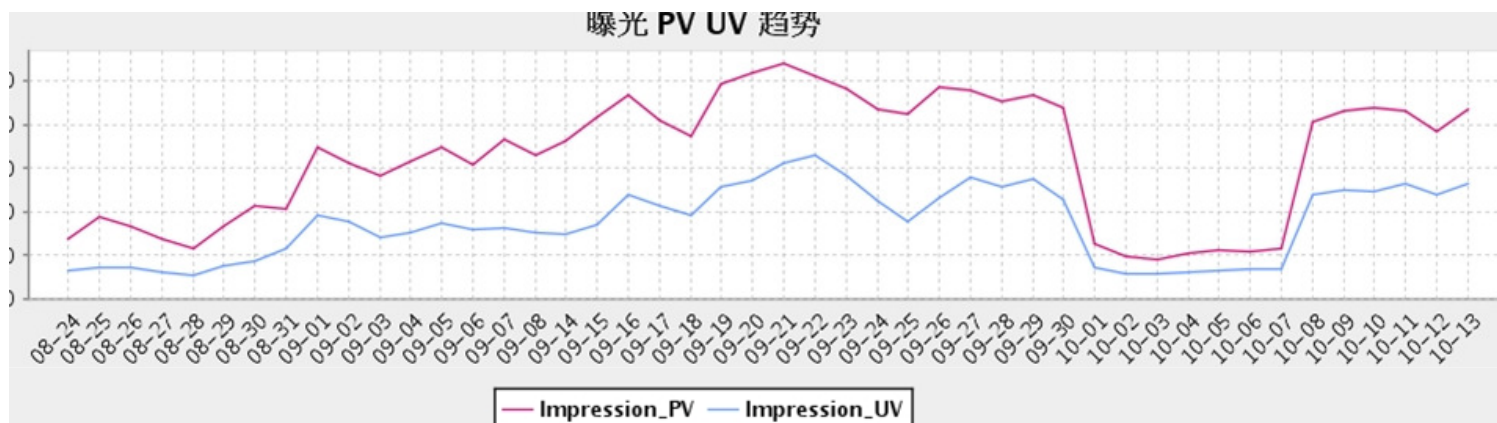
Weka (数据挖掘)

Mondrain (OLAP)

Kettle (ETL工具)

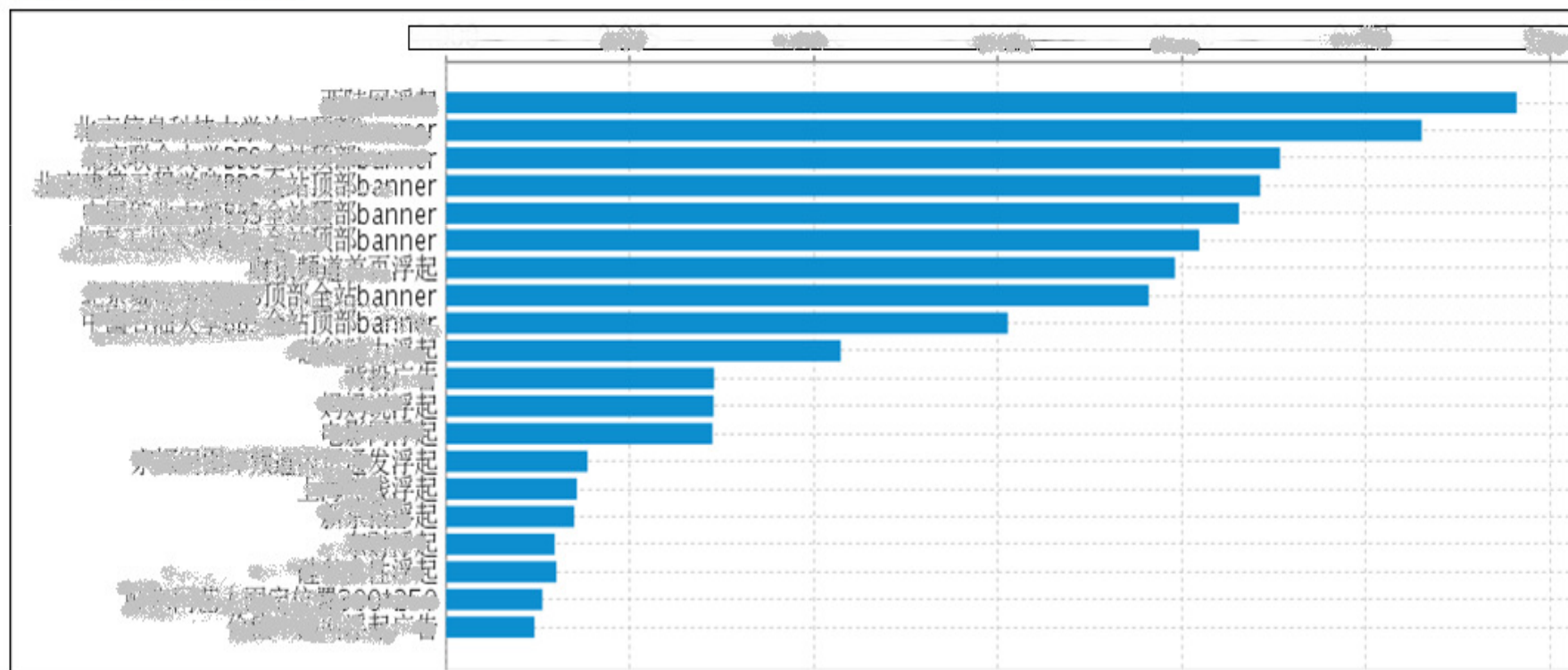
Pentaho
Platform

PV, UV趋势



广告位CTR排名

最近一周CTR前20名



欢迎加入品友!

x@ipinyou.com