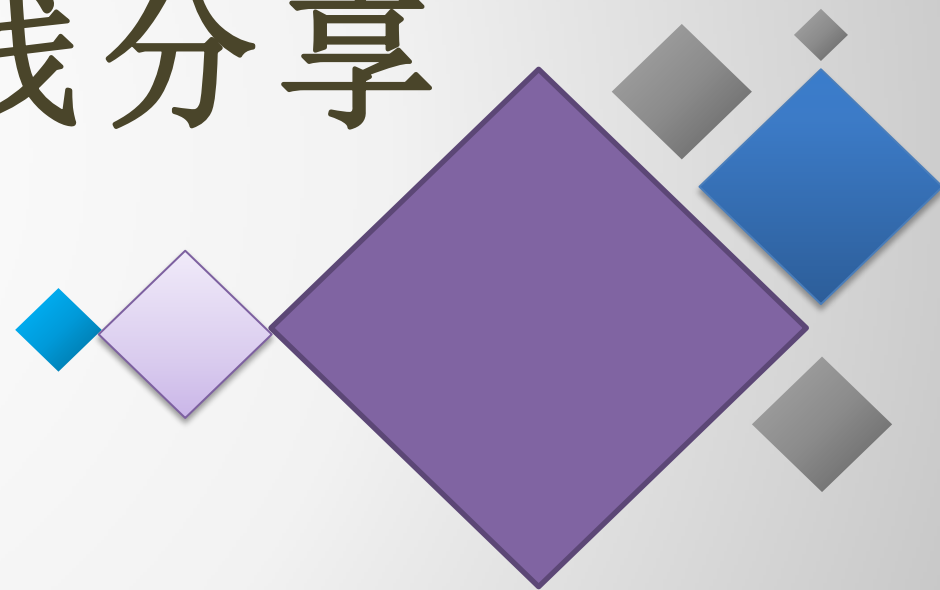


Druid 之 Metis BI 实践分享

许飞 360手机卫士服务端





——CONTENTS——



目录



Metis BI 系统介绍



Metis-Druid 实践案例



如何快速接入Metis-Druid?



PART ONE

Metis BI 系统介绍

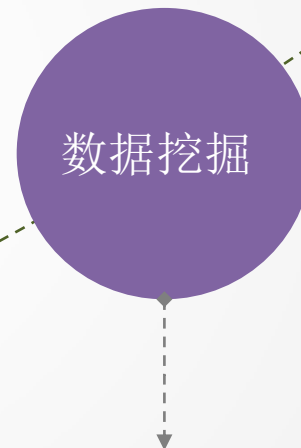
什么是Metis BI ?

每日处理的日志种类很多，有150多种，日志格式也很多，目前日志已经可以自动化接入处理

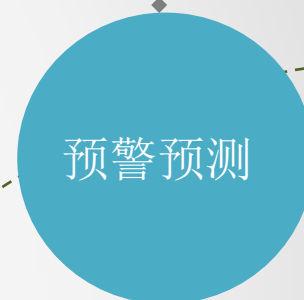


数据统计采用的技术很多，MR、spark、hive、storm、druid

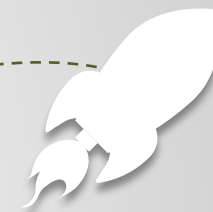
数据分析的方式多样化，支持多维图表展示，支持A/B 数据对比，支持top排名等功能分析



根据用户行为，挖掘用户兴趣，对用户进行归类，实现用户画像等内容



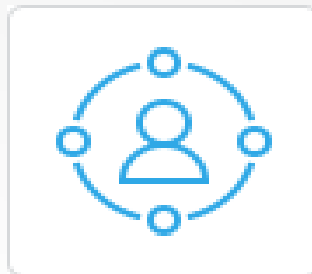
可以根据用户的阈值配置，对监控数据预警，同时可以对后续的一些行为进行预测





商业化广告分析

各种广告转化率的全方位监控分析，为用户提供全面细致的广告投放效果分析和投放策略制定的依据



事件跟踪分析

通过用户行为事件、业务计算事件统计，多维度下事件流操作统计，提升产品精细化推广运营效果，大幅提高产品质量



渠道质量分析

通过新增、活跃、留存、归因计算等模型，为用户提供全面的渠道推广数据分析，为渠道结算提供依据、同时提升渠道推广的质量



用户画像分析

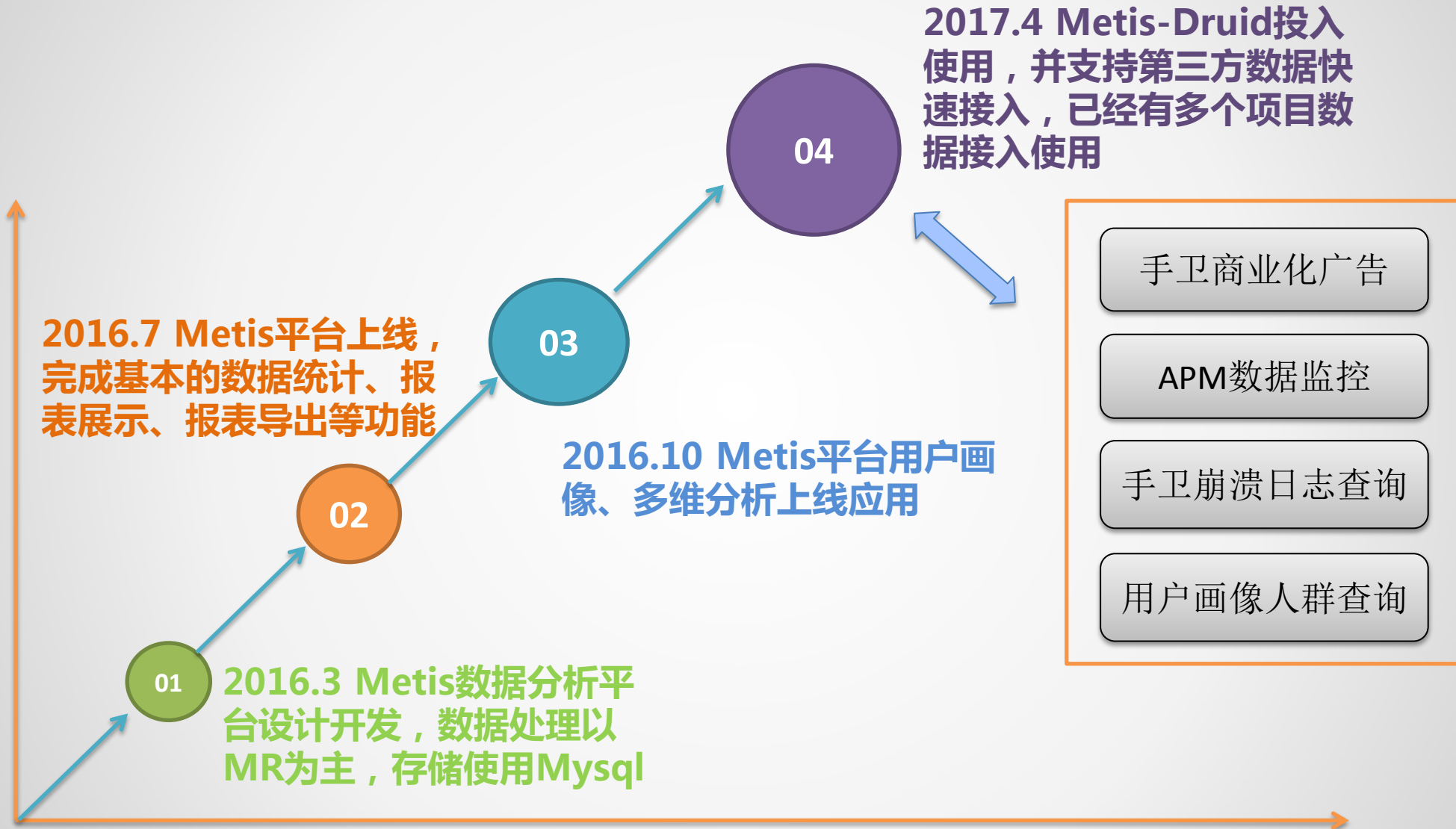
基于手机卫士2亿存量用户数据，形成完善可依赖的用户画像体系，为商业化广告的人群定向和精准发送push提供筛选依据

Metis BI 图表展示





Metis BI 发展历程



如何解决多维查询问题？

➤ 多维查询下，如果多达20维以上，如何更快的查询？

维护麻烦查询效率一般

mysql

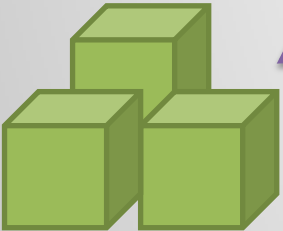
HBase

dap_activity_page_activity_id_day	dap_app_page_pkgname_day	dap_game_book_basic_day	dap_news_basic_day
dap_activity_page_day	dap_app_page_r_id_channel_day	dap_game_book_page_appkey_day	dap_news_channel_day
dap_activity_page_r_id_activity_id_day	dap_app_page_r_id_day	dap_game_book_page_day	dap_news_gated_launch_minute
dap_activity_page_r_id_day	dap_app_page_r_id_gray_minute	dap_hadoop_dir_size_day	dap_news_page_channel_day
dap_activity_page_r_id_gray_minute	dap_app_page_r_id_minute	dap_inmobi_basic_day	dap_news_page_day
dap_activity_page_r_id_minute	dap_app_page_r_id_pkgname_day	dap_inmobi_page_day	dap_news_page_r_id_channel_day
dap_activity_r_id_activity_id_day	dap_app_pkgname_day	dap_inmobi_page_r_id_day	dap_news_page_r_id_day
dap_activity_r_id_day	dap_app_r_id_channel_day	dap_inmobi_r_id_day	dap_news_page_r_id_gray_minute
dap_android_newuser_hourly	dap_app_r_id_day	dap_ios_active_user_event_day	dap_news_page_r_id_minute
dap_app_basic_day	dap_app_r_id_pkgname_day	dap_ios_channel_day	dap_news_r_id_channel_day
dap_app_channel_day	dap_app_recommend_app_id_day	dap_ios_channel_day_del20170615	dap_news_r_id_day
dap_app_cpm_app_id_day	dap_app_recommend_day	dap_ios_newuser_channel_hour	dap_push_client_day
dap_app_cpm_day	dap_app_recommend_page_app_id_day	dap_ios_newuser_channel_hourly	dap_push_session_upload_day
dap_app_cpm_page_app_id_day	dap_app_recommend_page_day	dap_ios_newuser_channel_retention_day	dap_push_session_upload_hour
dap_app_cpm_page_day	dap_cleandroid_cn_o2o_basic_day	dap_ios_newuser_detail_day	dap_push_session_upload_minute
dap_app_gated_launch_minute	dap_data_information	dap_ios_newuser_source_day	dap_shakeoff_ios_basic_day
dap_app_micro_total_basic_day	dap_game_base_basic_day	dap_miv_basic_day	dap_third_data
dap_app_micro_total_page_day	dap_game_base_page_day	dap_miv_page_day	dap_tongcheng_basic_day
dap_app_micro_total_page_r_id_day	dap_game_base_page_pkgname_day	dap_miv_page_r_id_day	dap_tongcheng_page_day
dap_app_micro_total_r_id_day	dap_game_base_pkgname_day	dap_miv_page_r_id_gray_minute	dap_tongcheng_page_r_id_day
dap_app_page_channel_day	dap_game_book_appkey_day	dap_miv_page_r_id_minute	dap_tongcheng_r_id_day
dap_app_page_day	dap_game_book_appkey_dto_day	dap_miv_r_id_day	dap_v5_activeuser_detail_day



Druid使用之后

HISTORICAL



REALTIME

OLAP查询平台

最近7天

Filter product province

Group By

order : pv limit : 10

Table

自动查询 查询 保存为模板 我的模板

筛选维度

- ★ province
- ★ product
- ★ channel
- ★ r_id
- ★ client_ma
- ★ plugin_ve
- ★ plugin_ne
- ★ plugin_ch
- ★ page_id
- ★ sub_page
- ★ pkgName

指标展示

- ✓ adshow
- ✓ 展现PV

OLAP查询平台

最近7天

Filter product province

Group By client_major_ver plugin_ver pkgName

order : pv limit : 50

Table

自动查询 查询 保存为模板 我的模板

DATE	CLIENT_MAJOR_VER	PLUGIN_VER	PKGNAME	ADSHOW	展现PV	展现UV	点击UV	点击PV	下载完成数	安装完成数	打开数
2017/07/28	7.7.1	171	com.Qunar	36,986	37,754	34,527	850	915	765	100	24
2017/07/28	7.7.1	171	com.kongzhong.c1.qh	39,837	38,965	36,574	164	164	122	18	5
2017/07/28	7.7.2	171	com.huajiao	41,827	46,740	36,543	758	744	606	130	36
2017/07/28	7.7.1	169	net.qihoo.mobilesafe.syncassistant	43,583	46,147	32,074	161	165	126	63	18
2017/07/28	7.7.1	169	com.qihoo.browser	45,020	45,524	31,210	826	841	610	204	22
2017/07/28	7.7.2	171	com.ss.android.article.news	48,622	48,635	36,473	1,679	1,788	1,623	901	787
2017/07/28	7.7.2	1002006	com.qihoo.magic	48,633	48,165	31,641	202	223	197	1,995	43
2017/07/28	7.7.0	171	net.qihoo.mobilesafe.syncassistant	50,732	50,603	37,016	45	50	47	17	4
2017/07/28	7.7.2	1002006	net.qihoo.mobilesafe.syncassistant	53,978	54,198	34,906	363	367	347	2,068	116
2017/07/28	7.7.0	171	com.qihoo.video	56,786	58,726	40,459	153	161	147	43	9

... < 1 2 3 4 5 > ...

Showing 1 to 10 of 50 entries

指标展示

- adshow
- 展现PV
- 展现UV
- ✓ 点击UV
- ✓ 点击PV
- ✓ 下载完成数
- ✓ 安装完成数
- ✓ 打开数
- ✓ 广告点击率

Druid如何减少数据存储空间？



用很少的 **Money** 解决存储问题
减小数据存储开销

DRUID





Metis-Druid 系统主要解决问题场景





PART TWO

Metis-Druid 实践案例

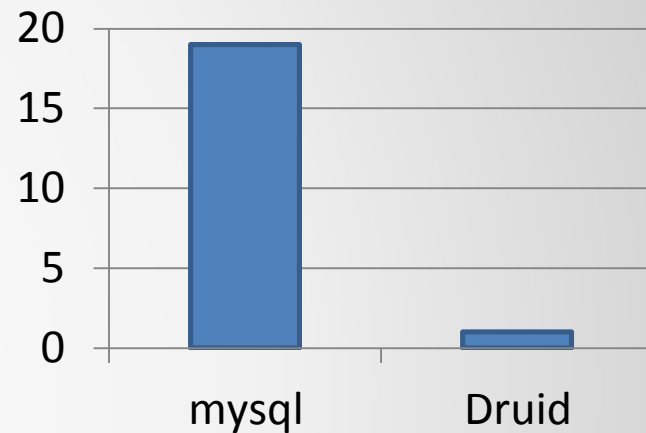


手卫商业化广告

查询效率

- 老系统：查询一个月渠道转化率趋势数据需要19秒
- Druid：同样的查询案例，仅需要不到1秒

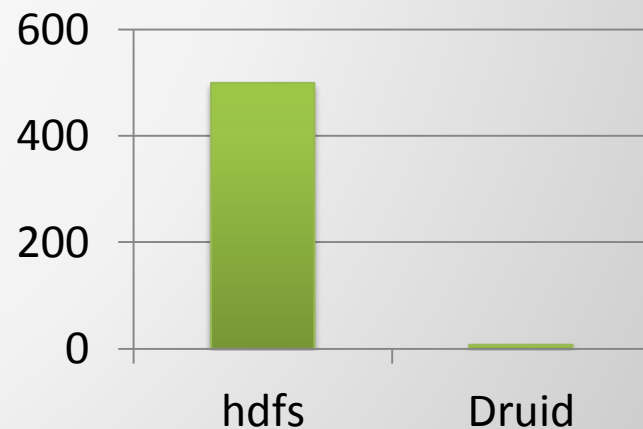
请求时间(s)



存储空间

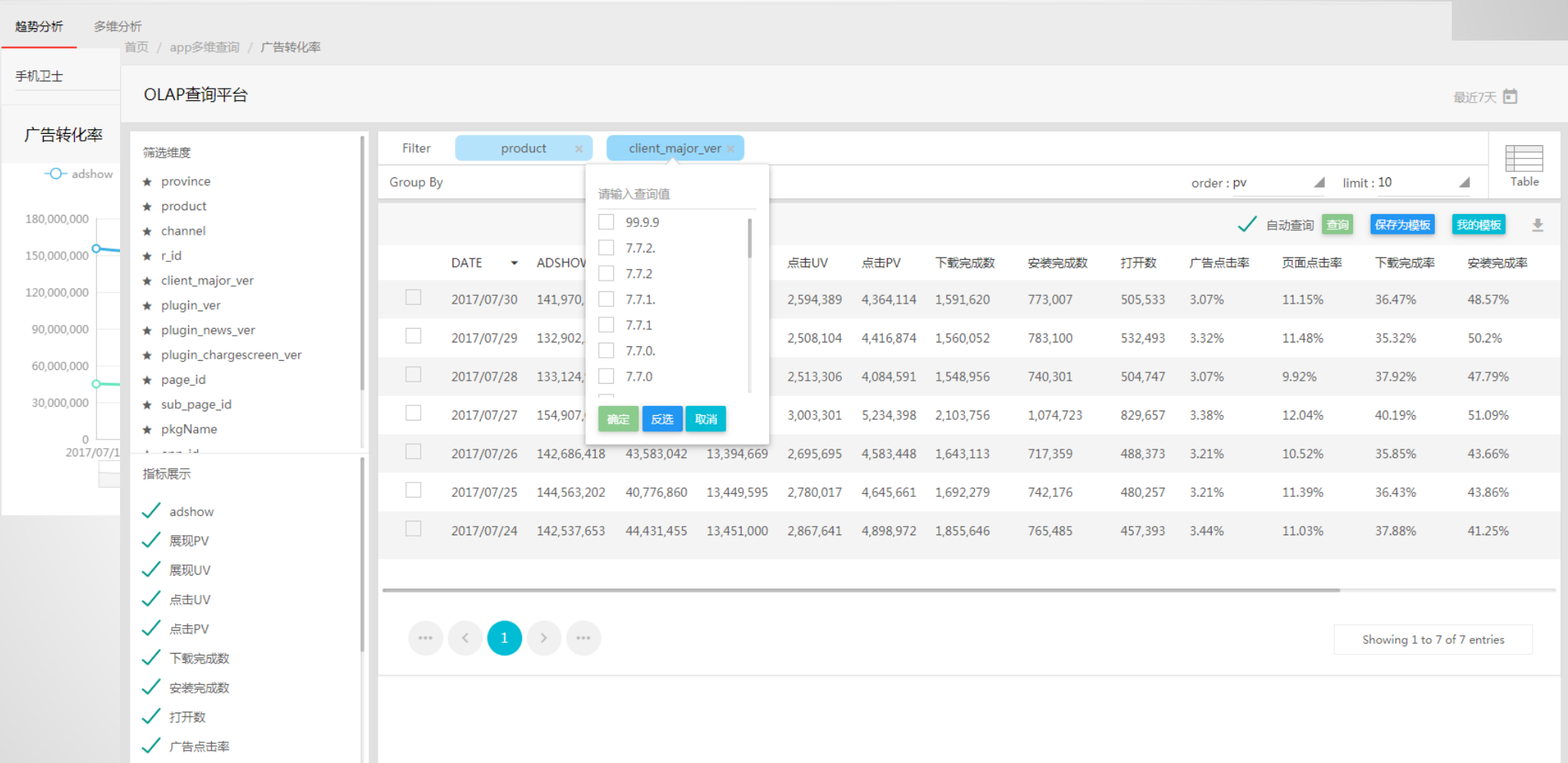
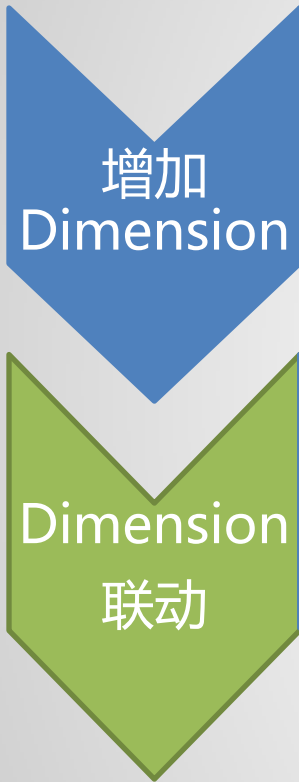
- 老系统：每天hdfs需要3*120G的存储空间
- Druid：8G存储空间

存储空间(G)





手卫商业化广告

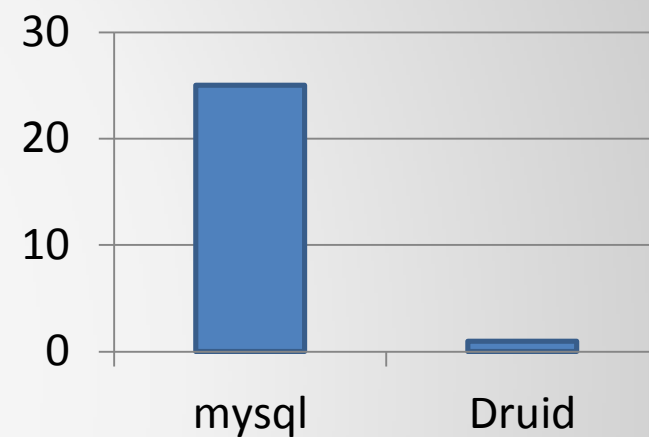


崩溃日志查询服务

查询效率

- 老系统：60%的查询都是慢查询，每次查询耗时25秒以上
- Druid：查询基本秒级返回

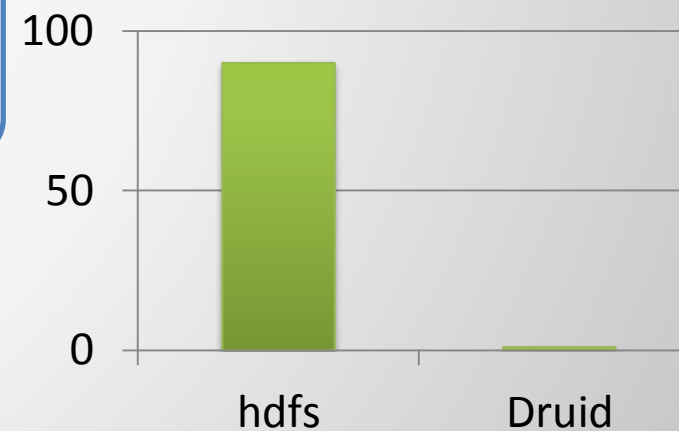
请求时间(s)



存储空间

- 老系统：月存储 $30\text{G} \times 3 = 90\text{G}$
- Druid： $36\text{M} \times 30 = 1\text{G}$,
- 压缩倍数 $= 90 / 1 = 90$ 倍

存储空间(G)





留存率分析

查询方式

Dimensions

在线活跃留存分析

Search

10



日期

1日留存率

3日留存率

7日留存率

30日留存率

2017/06/28

0%

0%

0%

0%

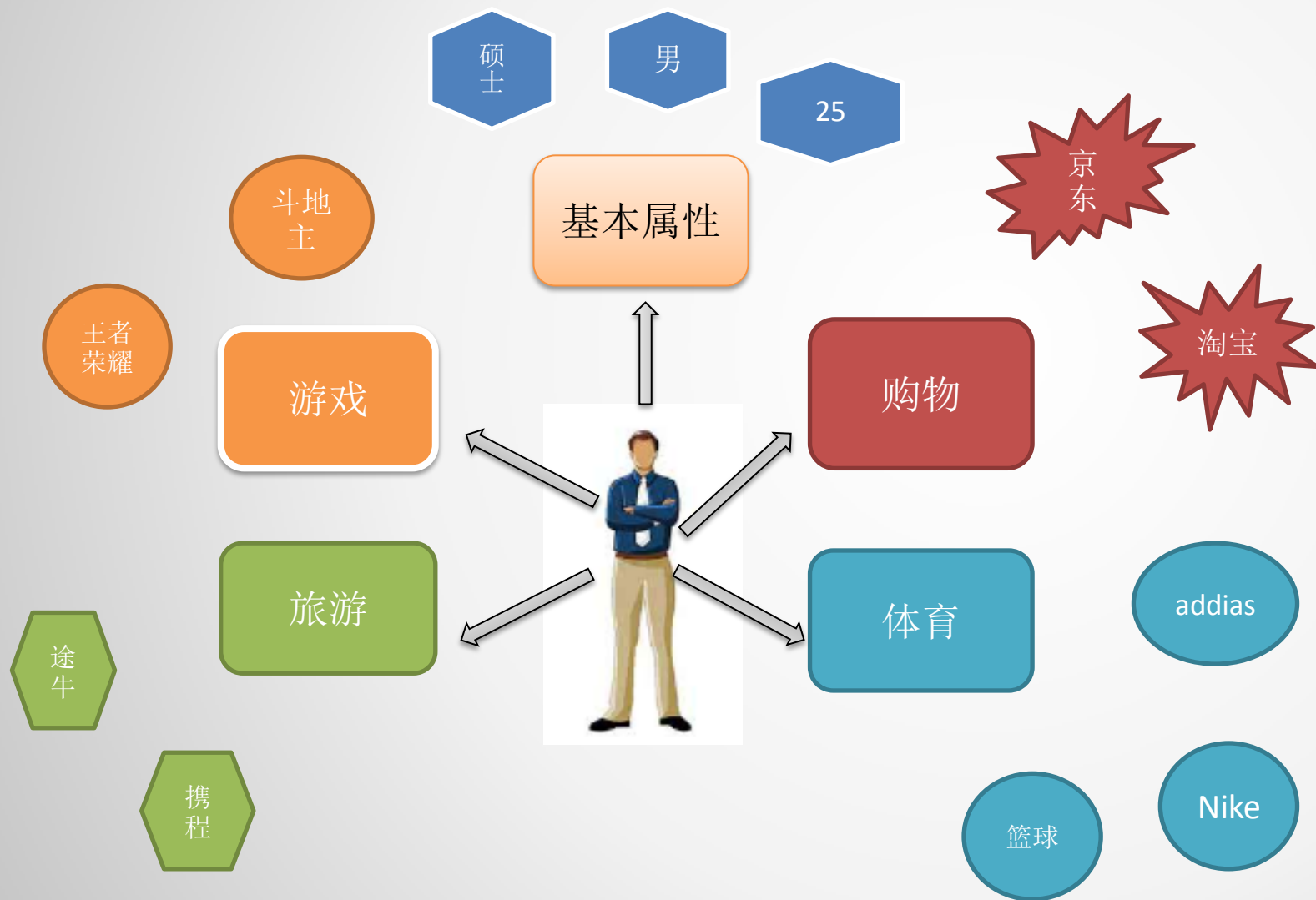
Breakdown

Export CSV

Search

	Segment	Users	Day 0	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Day 8	Day 9	Day 10	Day 11	Day 12
	All Users	36,069	100%	26.6%	19.4%	16.0%	13.6%	11.8%	10.6%	9.66%	8.90%	8.09%	7.49%	7.00%	6.61%
	Jul 26	678	*100%												
	Jul 25	1,212	*100%	*2.97%											
	Jul 24	387	100%	*34.1%	*8.01%										
	Jul 23	417	100%	36.0%	*23.7%	*3.84%									
	Jul 22	434	100%	34.3%	23.7%	*18.0%	*4.15%								
	Jul 21	433	100%	32.6%	25.2%	18.7%	*14.8%	*3.93%							
	Jul 20	484	100%	32.2%	21.3%	17.8%	15.3%	*12.6%	*3.31%						

用户画像兴趣标签应用



如何根据标签
快速查询人群
数量？

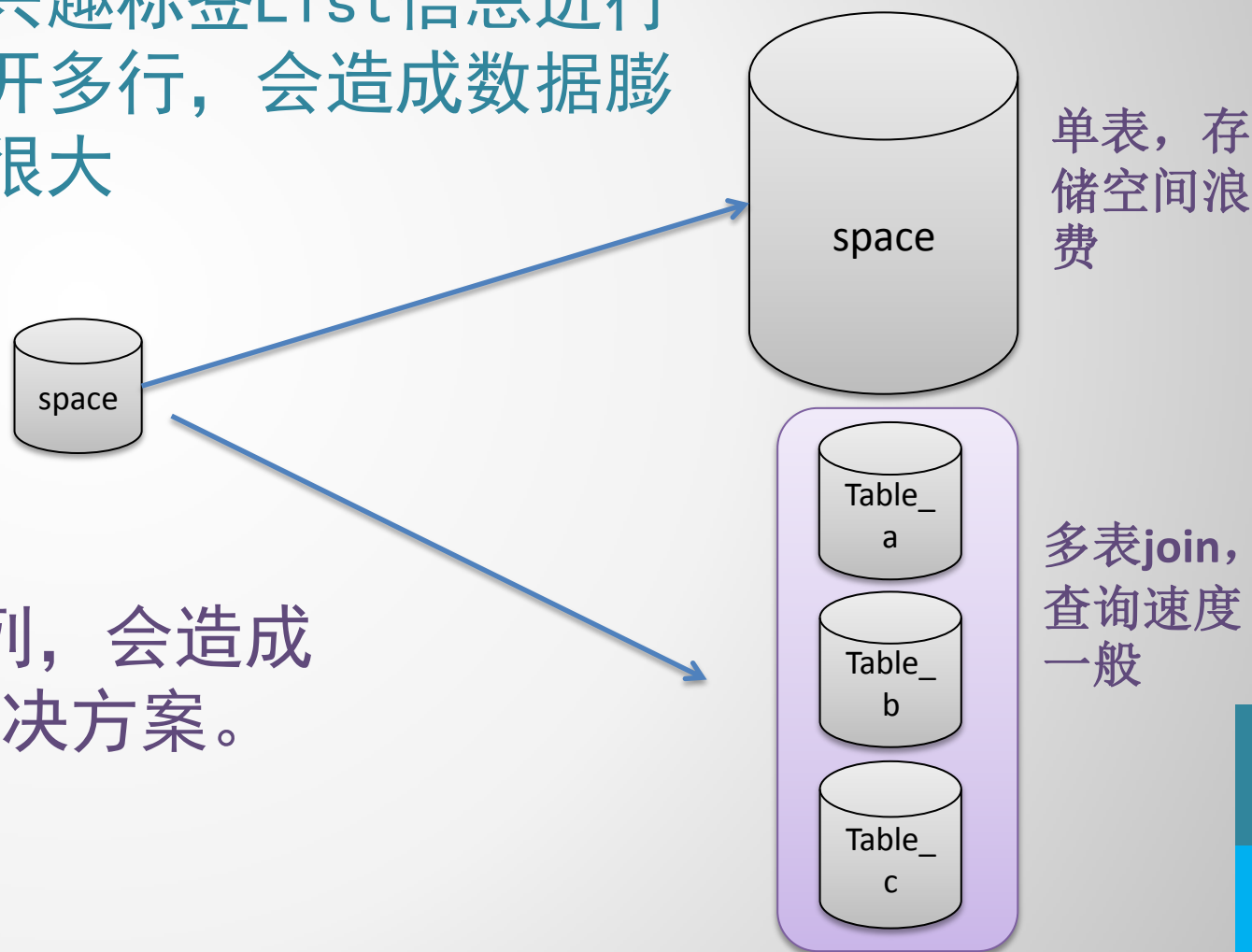
如何查询人群数量

Greenplum



将兴趣标签List信息进行展开多行，会造成数据膨胀很大

如果将兴趣标签List信息展开多列，会造成dimension膨胀，都不是很好的解决方案。



Druid 如何做到的？

Druid supports “multi-value” string dimensions , Druid支持list存储，例如某个人的兴趣标签list有很多，在存储的时候，将兴趣标签list存成一组数据，作为一个维度，与其他dimension平级，查询的时候，和其他dimension查询方式一样。

```
{ "timestamp": "2011-01-12T00:00:00.000Z", "tags": ["t1","t2","t3"]} #row1  
{ "timestamp": "2011-01-13T00:00:00.000Z", "tags": ["t3","t4","t5"]} #row2  
{ "timestamp": "2011-01-14T00:00:00.000Z", "tags": ["t5","t6","t7"]} #row3
```



PART THREE

如何快速接入 Metis-Druid?

How to access ?

Metis系统支持第三方数据源的快速接入，接入的数据通过项目进行权限隔离，接入方式通过Metis的申请接入平台，



个人设置

全屏

清除本地数据

用户设置

审批浏览权限

申请浏览权限

申请接入平台

项目管理

频道管理

配置管理

退出

首页

app多维查询

广告转化率

news多维查询

mv多维查询

申请接入

接入频道名称: 请输入频道名称

所属项目: 请选择项目

所属分类: 请选择分类

接口数据源: ☐ 本地数据 ☐ 第三方数据 ☒ Olap数据

表名: druid表名, 在druid集群的表名

维度: 筛选项的字段名称, 逗号分隔

指标: {'指标1': '类型', '指标2': '类型'}

orderby: 按照指标降序显示, 如pv、uv, 逗号分隔

limit: 10,30,50,100最大不超过1000, 逗号分隔

集群地址: 集群的访问地址

提交

chart模板

OLAP查询平台

Filter: product

Group By: order (pv)

Level: 10

Table

Time Series

Bar Graph

DATE	ADSHOW	展示PV	展示UV	点击PV	点击UV	下载次数	安装次数	打开数	点击率	下载率
2017/06/12	171,818,274	16,324,076	16,054,094	2,826,478	4,716,173	1,916,163	846,167	401,684	8.42%	40.68%
2017/06/11	184,078,738	40,030,062	33,959,311	1,357,328	5,225,789	2,043,398	737,174	458,520	6.73%	19.34%
2017/06/10	189,894,967	61,179,198	34,672,138	1,346,742	5,434,569	2,117,279	720,057	492,264	8.98%	18.90%
2017/06/09	186,495,209	61,862,112	34,151,069	1,357,401	5,188,811	2,078,809	685,847	424,401	8.70%	18.58%
2017/06/08	190,181,235	61,196,518	34,848,529	1,348,768	5,941,172	2,160,779	1,014,366	724,575	9.73%	42.93%
2017/06/07	185,941,214	62,527,651	34,551,767	1,082,851	5,180,142	2,011,291	662,235	411,974	8.67%	17.88%
2017/06/06	190,047,859	60,211,222	34,509,389	1,305,138	5,362,919	2,171,896	722,157	446,189	9.24%	19.08%



APM系统监控

APM平台接入后，展示其相关频道，以及筛选dimension和指标。

←

METIS

Hi, 许飞

个人设置

首页

APM监控数据统计

交互监控统计

内存监控统计

网络监控统计

传感器监控统计

首页 / APM监控数据统计 / 交互监控统计

OLAP查询平台

最近7天

筛选维度

★ apmid

★ m

★ s

★ rv

★ an

★ av

★ lc

★ st

★ n

★ fv

指标展示

✓ 总时间

✓ 总次数

✓ 平均耗时

Filter

apmid

Group By

order : t

limit : 10

Table

✓ 自动查询

Table

Time Series

Bar Graph

DATE	总时间	总次数	平均耗时
<input type="checkbox"/> 2017/07/25	297,976,625	3,161,897	94.24
<input type="checkbox"/> 2017/07/24	296,013,997	2,982,525	99.25
<input type="checkbox"/> 2017/07/23	295,675,825	2,772,894	106.63
<input type="checkbox"/> 2017/07/22	317,548,872	2,950,506	107.63
<input type="checkbox"/> 2017/07/21	331,510,125	3,684,288	89.98
<input type="checkbox"/> 2017/07/20	361,362,709	3,517,281	102.74
<input type="checkbox"/> 2017/07/19	367,909,159	3,490,089	105.42

联系我们



后续的计划

- Druid的跨表join操作，后续与系统部在这方面看看有没有很好的实践
- Metis BI系统的智能预测方向

Question & answer



联系方式

