

# 标签交流分享会

数据运营部

2017.6.19

# 目录

---

## 1 标签体系

## 2 打标签流程

## 3 标签结果处理

## 4 标签运营优化迭代情况

# 标签体系

维度	行业	统计	体系建设参考
5维 <ul style="list-style-type: none"><li>➢ URL标签</li><li>➢ APP标签</li><li>➢ SITE标签</li><li>➢ 临时标签</li><li>➢ 关键字标签</li></ul>	8大类 <ul style="list-style-type: none"><li>➢ 基础（国信）</li><li>➢ 金融（亚信）</li><li>➢ 房产（亚信）</li><li>➢ 汽车（泰岳）</li><li>➢ 电商（数据研发）</li><li>➢ 区域房产</li><li>➢ 微信</li><li>➢ 旅游</li></ul>	URL： <ul style="list-style-type: none"><li>➢ 23个一级分类；</li><li>➢ 188个二级标签；</li><li>➢ 1277个三级标签；</li><li>➢ 1944个四级标签；</li><li>➢ 299个五级标签；</li><li>➢ 52个六级标签；</li></ul> APP： <ul style="list-style-type: none"><li>➢ 3648个APP；</li></ul> SITE： <ul style="list-style-type: none"><li>➢ 84260网站</li></ul>	<ul style="list-style-type: none"><li>➢ 市场部红头文件；</li><li>➢ 各厂家标签体系；</li></ul> <div><p>兴趣标签体系及标 码--内部使用（5月</p></div> <div><p>APP标签体系及标 码--内部使用（5月</p></div> <div><p>Site标签体系及编 码内部使用版（5月份</p></div>

# 标签架构

---

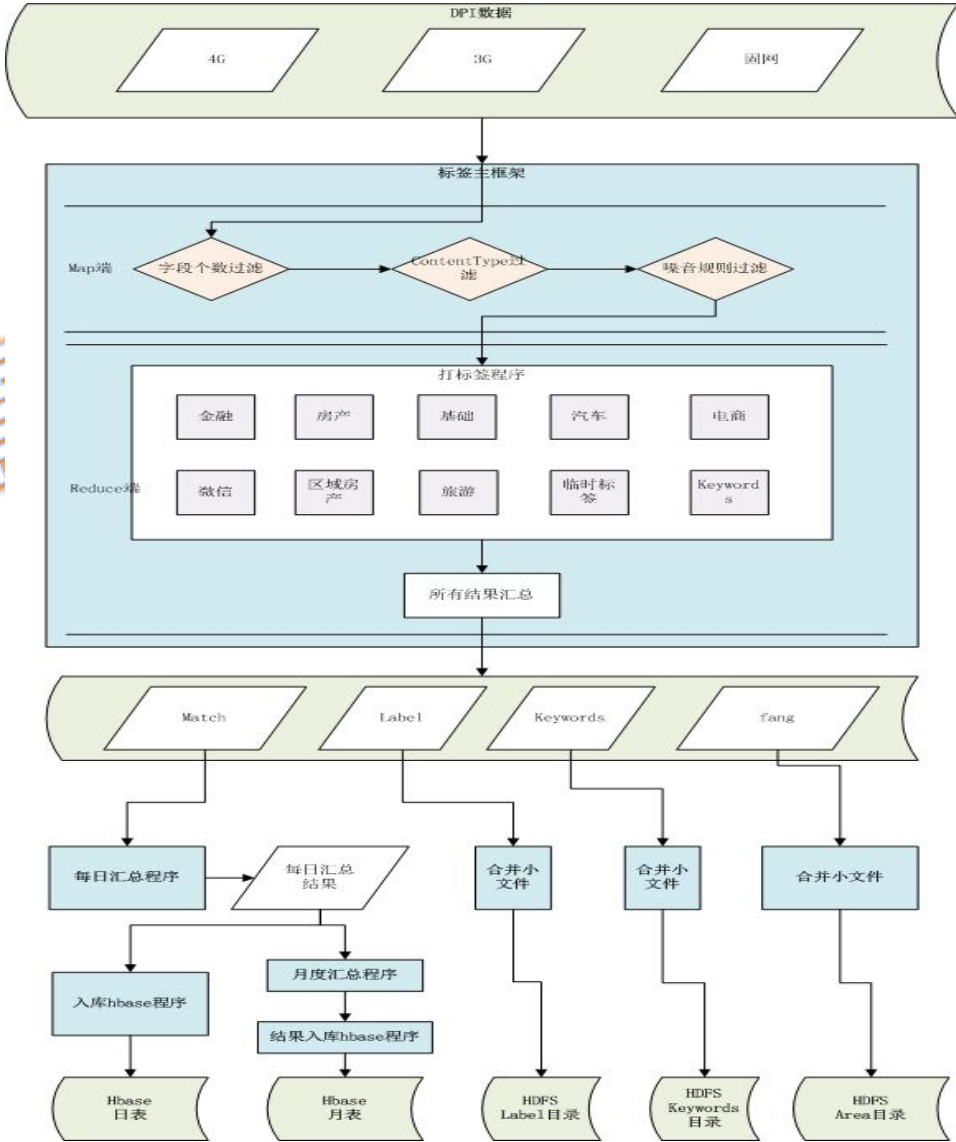
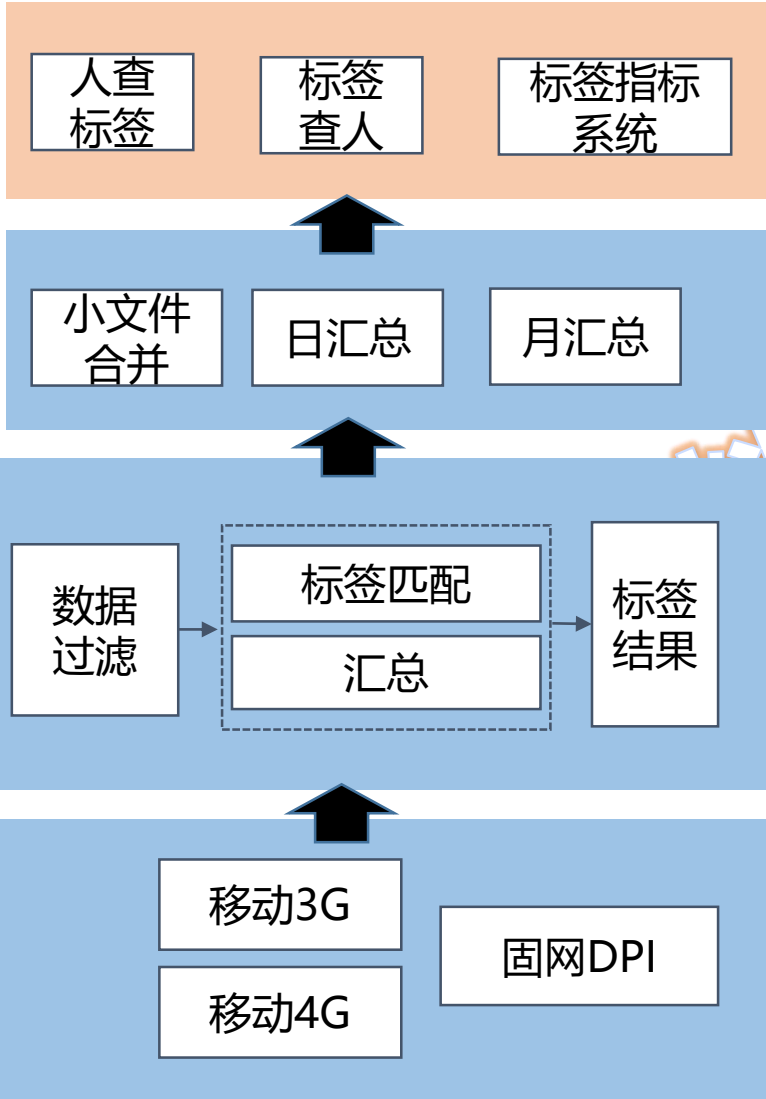
1 标签体系

2 打标签流程

3 标签结果处理

4 标签运营优化迭代情况

# 标签架构&流程图



# 数据过滤准则

## 异常数据过滤

### □ 字段不够

- 移动DPI：38、40字段；
- 固网DPI：12字段

### □ URL长度异常

- 长度小于9

### □ 主标识为空

- 移动DPI：mdn、imei/meid、imsi字段全部为空；
- 固网DPI：useraccount字段为空；

异常数据过滤

ContentType  
类型过滤

噪音库文件

## ContentType类型过滤

image、application等7种类型；

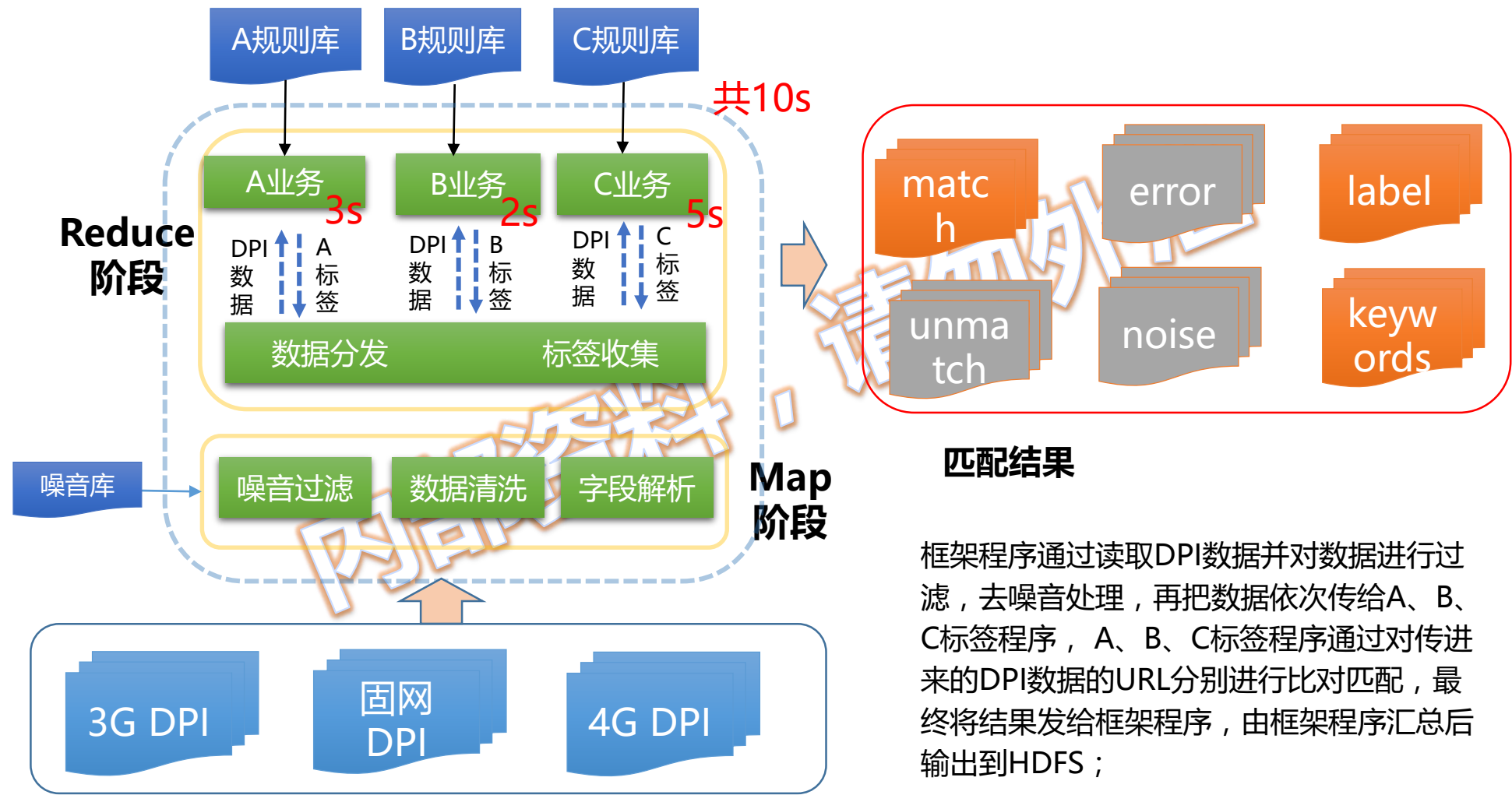
## 噪音库文件过滤

.jpg、.gif、.jpeg等51种后缀；



rule\_noise

# 打标签匹配流程



  
基础标签打标流程.docx

框架程序通过读取DPI数据并对数据进行过滤，去噪音处理，再把数据依次传给A、B、C标签程序，A、B、C标签程序通过对传进来的DPI数据的URL分别进行比对匹配，最终将结果发给框架程序，由框架程序汇总后输出到HDFS；

# 标签架构

---

1 标签体系

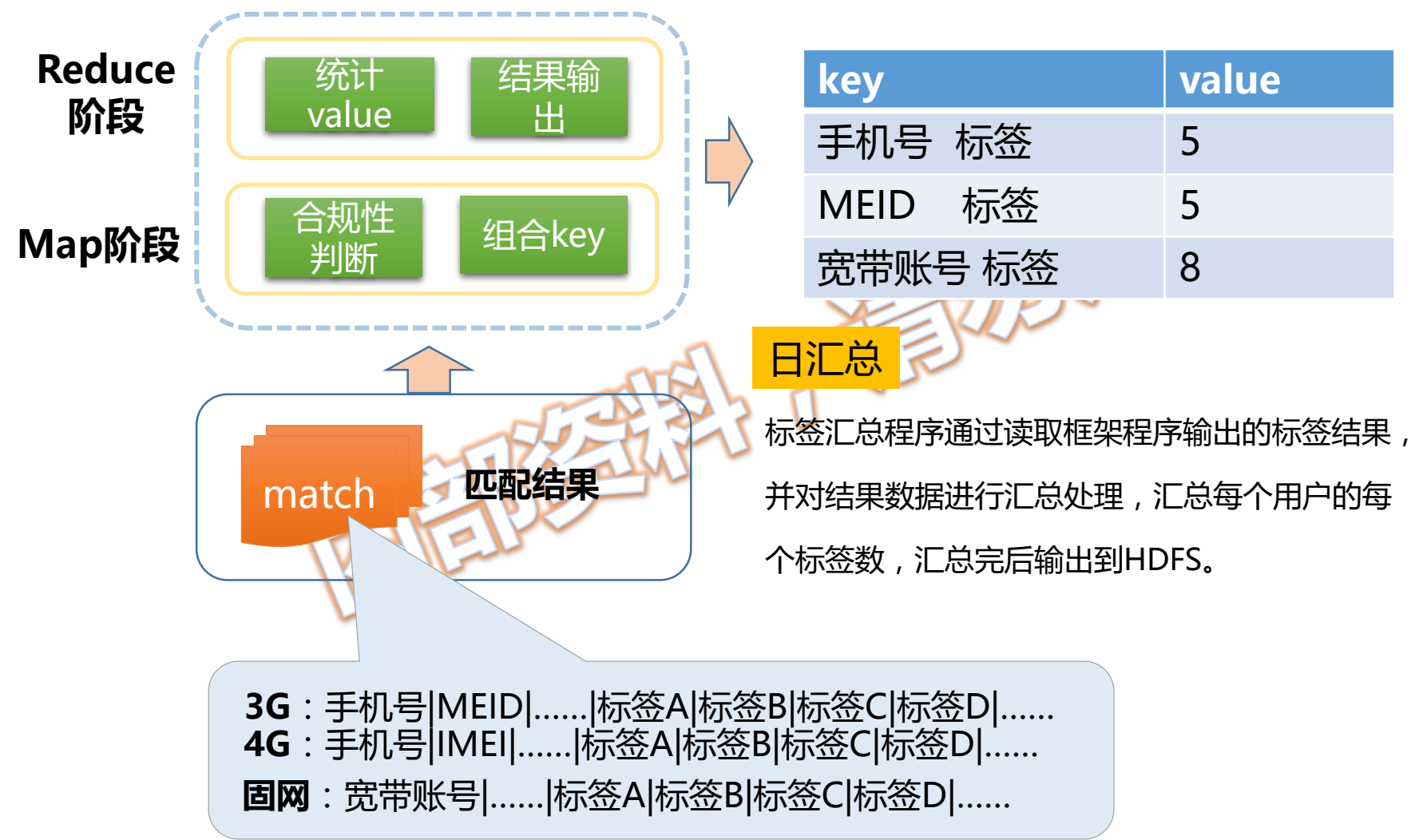
2 打标签流程

3 标签结果处理

4 标签运营优化迭代情况



# 标签汇总流程



## 月汇总

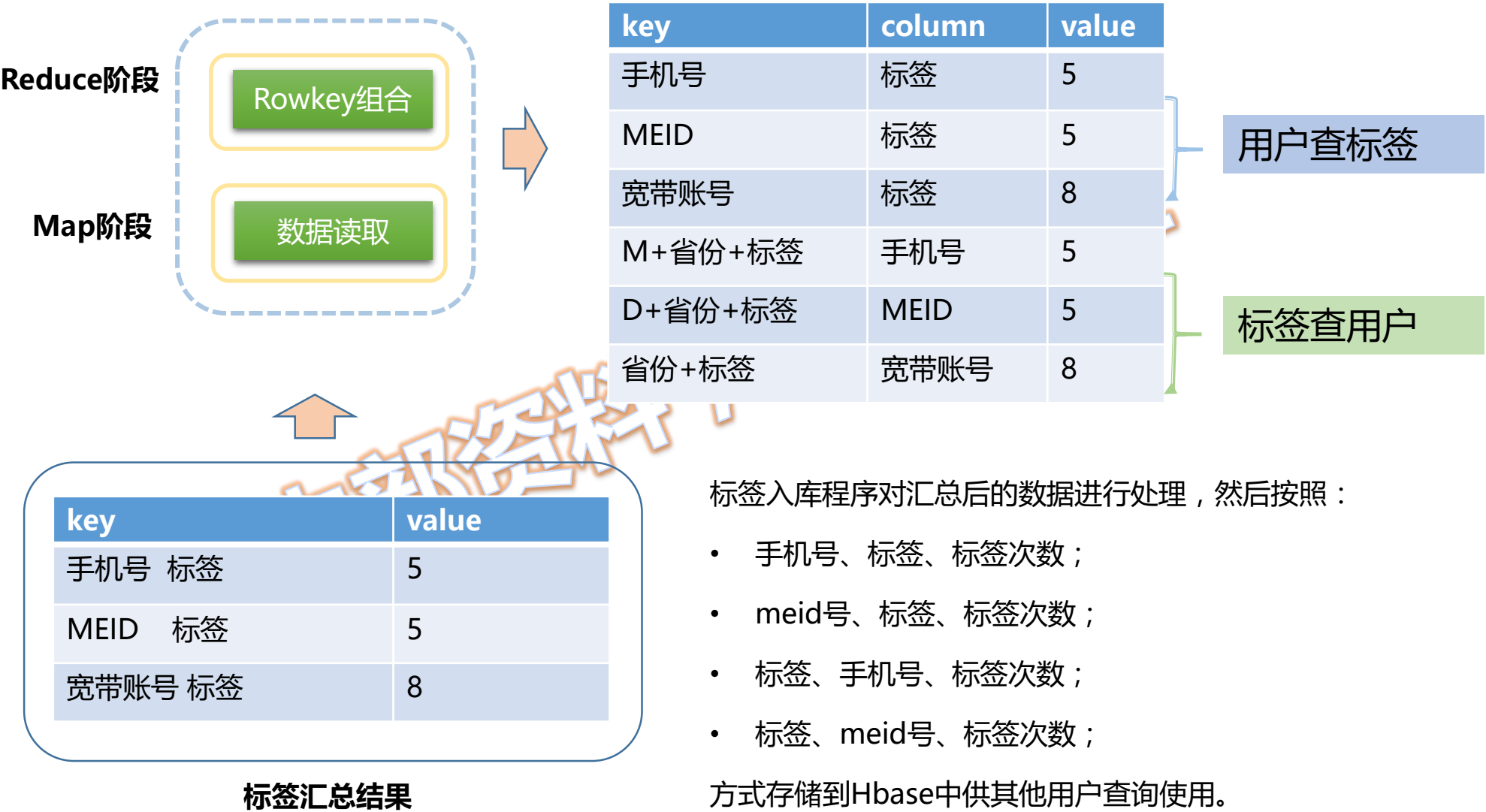
牛顿冷却算法



月度标签权值算法模型.docx

3G : 手机号|MEID|.....|标签A|标签B|标签C|标签D|.....  
4G : 手机号|IMEI|.....|标签A|标签B|标签C|标签D|.....  
固网 : 宽带账号|.....|标签A|标签B|标签C|标签D|.....

# 标签入库流程



# 标签架构

---

1 标签体系

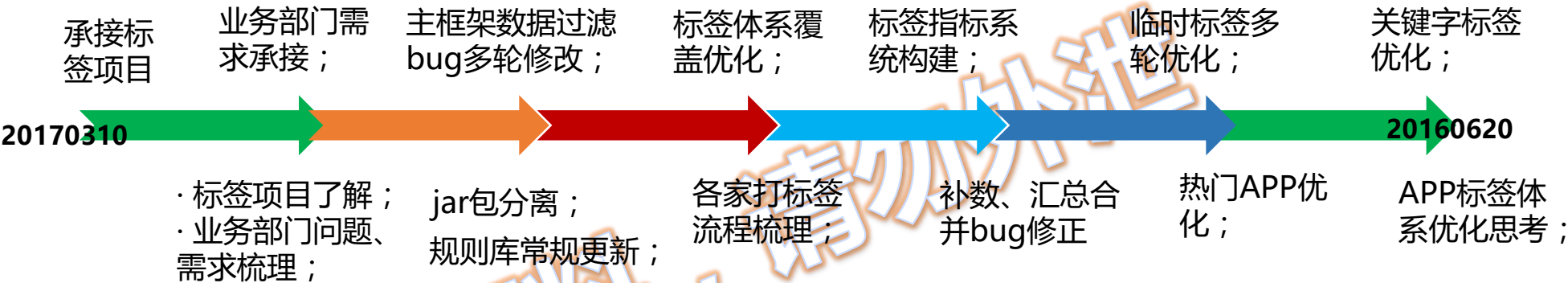
2 打标签流程

3 标签结果处理

4 标签运营优化迭代情况

# 标签运营优化迭代情况

## □ 迭代优化



## □ 主要成果

- **数据过滤占比**
  - 90% -> 75% -> 70% ;
- **标签体系覆盖率**
  - Site : 4%~99.8% ;
- **临时标签优化**
  - 对比测试，提高5倍匹配率；
- **关键字标签**
  - 分隔符bug发现，修改，供客户正常使用；
- **热门APP优化**
  - 按照类别进行top20的持续优化；
- **标签指标体系**
  - 多维度评估打标签质量；



标签bug或升级需求\_v1.0.xlsx

# 标签现有主要问题及下一步方向

## □ 现有主要问题及下一步方向

### ➤ 资源不足

- 数据量增加，业务增加，但资源不变；

解决方案：资源申请；主框架业务剥离；

### ➤ 金融表标签APP规则不准

解决方案：协调厂商重新维护；

### ➤ 金融APP打标签逻辑错误

解决方案：修正逻辑错误；

### ➤ 房产打标签不准，无优化

- 抓包工具不好用，所打标签基本不可用，需要重建；

解决方案：与区域房产合并，厂商手工维护；

### ➤ APP标签体系不统一

- 历史原因，APP无体系，各厂商自行维护，纵观全局，非常凌乱；

解决方案：制定APP标签体系，将现有APP重新整理；

### ➤ 汽车标签黑盒解不开，且无APP

- 厂家自己维护的标签库与提供给云公司的标签库不一致；

- 实现逻辑与实际业务对不上；

解决方案：问题不解决，不验收或者作为遗留问题；

### ➤ 电商标签基本无人维护

- 项目结束，无运营合同；

解决方案：电商https率高，不予维护；或可安排专人学习和维护或者重建；

内部资料 请勿外泄

谢谢