

天下营销，唯快不破

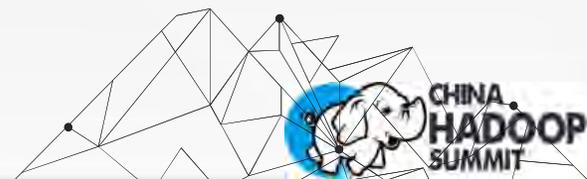
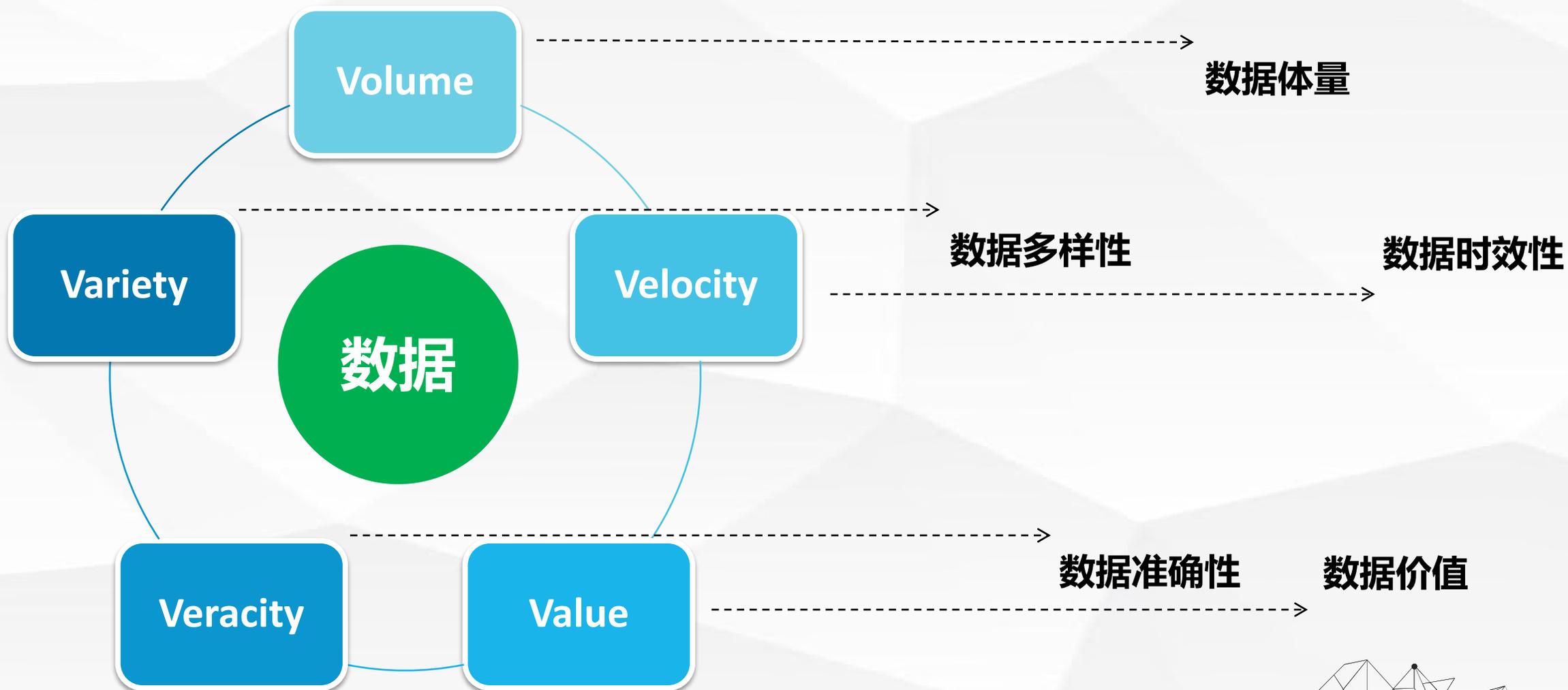
TalkingData营销云实时计算框架

主讲人：王福胜





智能数据时代的来临





不“快”不“行”



为什么要“快”？



什么叫“快”？



“快”的代价是什么？





智能数据在广告行业的实践-营销云



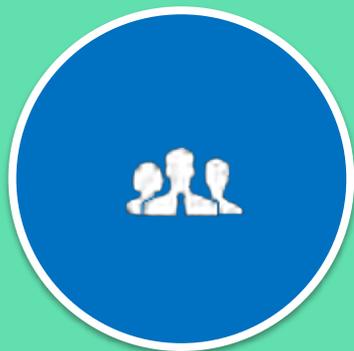
客群构建

- 上传构建人群
- POI构建人群
- 标签构建人群
- Lookalike



画像洞察

- 人口属性
- 终端属性
- 地理位置属性
- 行业指数



客群营销

- 广告投放平台
- EDM
- SMS
- Call Center



效果评估

- 触达分析
- 效果监测





营销云-挑战

数据特征

体量大

维度多

50亿覆盖

6.5亿月活

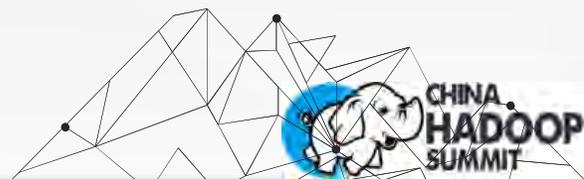
2.5亿日活

多种设备ID

800+标签

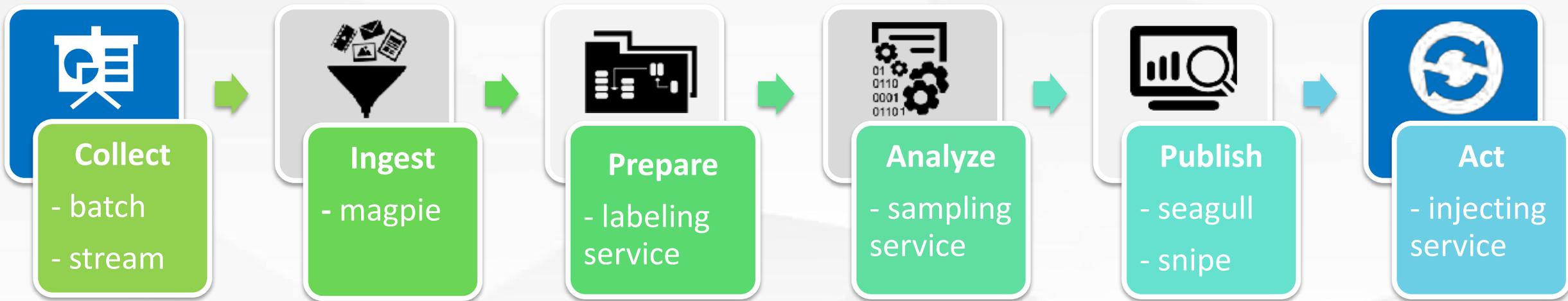
地理位置

时间周期





数据的组织和规划





TalkingData 标签体系

- 人口属性：包含性别、年龄等人的基本特征
- 资产情况：车辆、房产、收入等资产特征
- 兴趣特征：阅读资讯、运动健康等兴趣偏好
- 消费特征：网上/线下消费类别品牌等特征
- 位置特征：常驻城市、职住距离等
- 设备属性：所使用终端的特性等

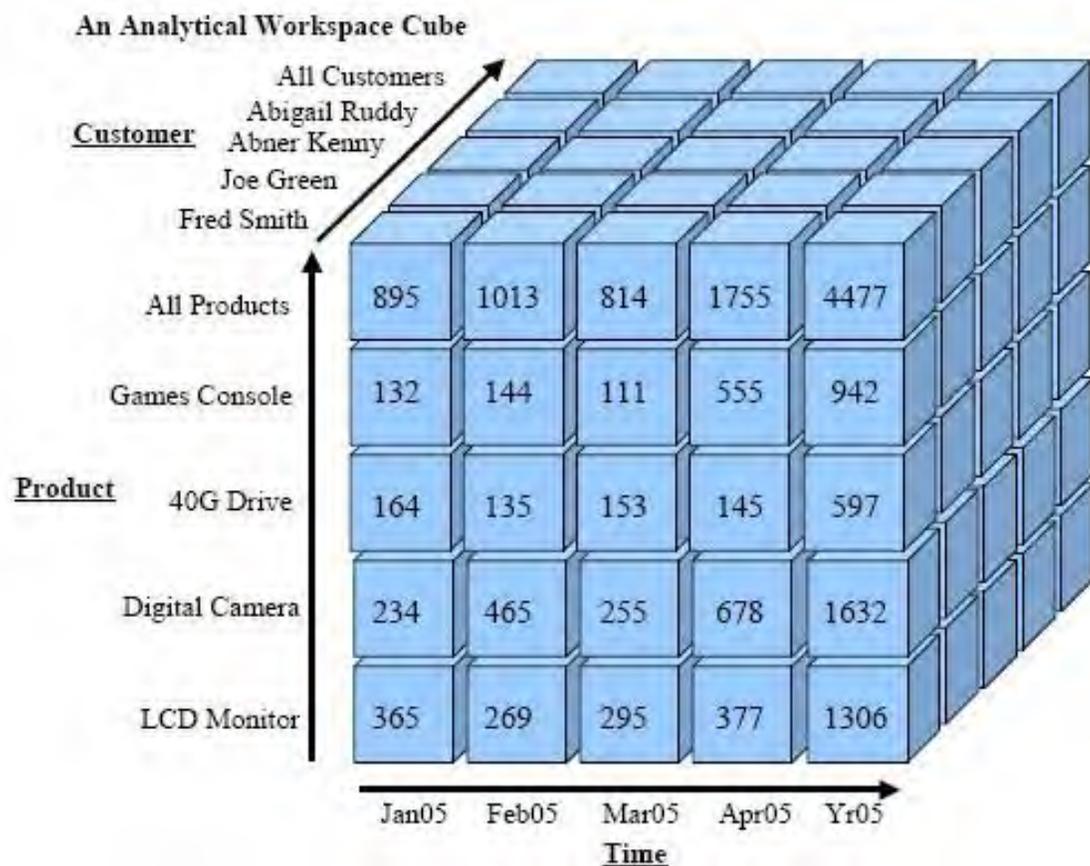


共有6大类，超过800个受众定向标签，这些标签关联的设备累加起来超过700亿。如何利用这些标签为用户提供快速的标签人群构建，对人群进行多维度的快速画像是一个挑战

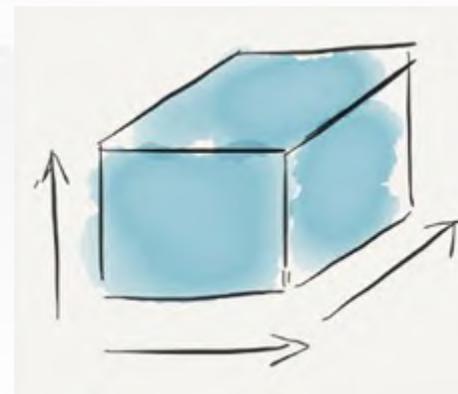


数据的BitMap化

OLAP-Cube: multi-dimensional array of data

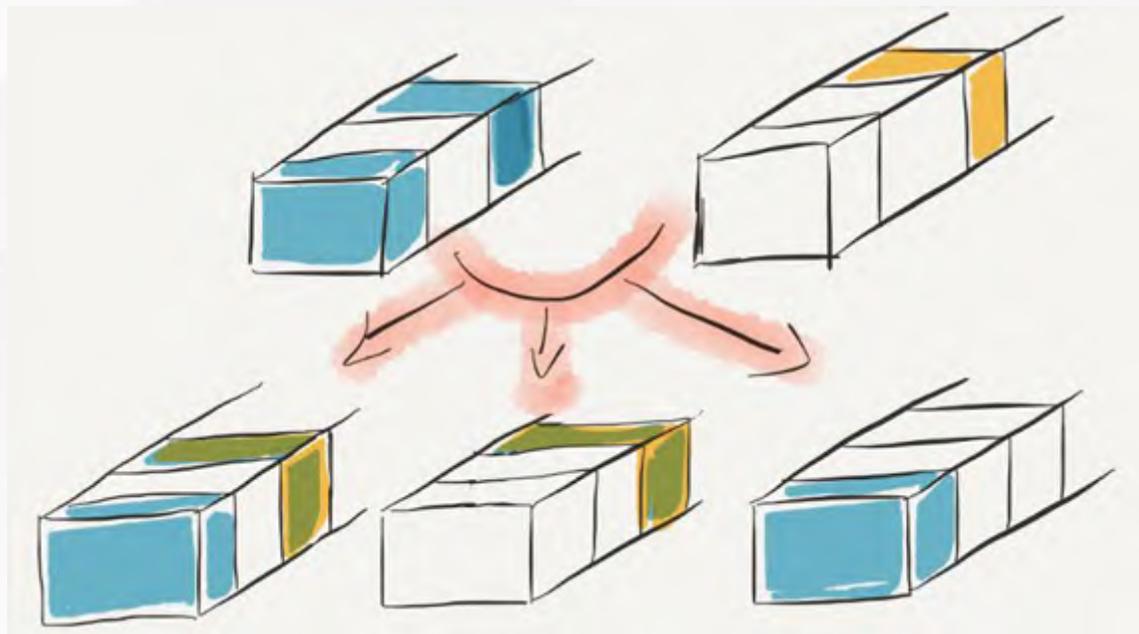


Atom OLAP Cube:





数据的BitMap化



计算：

	\cup	\cap	\oplus
交换	$A \cup B = B \cup A$	$A \cap B = B \cap A$	$A \oplus B = B \oplus A$
结合	$(A \cup B) \cup C = A \cup (B \cup C)$	$(A \cap B) \cap C = A \cap (B \cap C)$	$(A \oplus B) \oplus C = A \oplus (B \oplus C)$
幂等	$A \cup A = A$	$A \cap A = A$	



标签的BitMap运算

city	Bitmap
北京	bitmap1
天津	bitmap2

device	Bitmap
苹果	bitmap3
华为	bitmap4

计算：

SELECT Device Where city=北京 and device=苹果

运算转变成：Bitmap1 and bitmap3

优势：

1. 存储减少
2. 计算快
3. 支持join





使用Spark做人群构建，画像

TalkingData是把设备标签数据进行了BitMap化处理，利用BitMap的快速And, Or运算能力来做标签人群构建和人群画像。

虽然Bitmap对数据有很好的压缩能力，但是因为TalkingData积累的移动设备数量庞大，生成的Bitmap很多都是几百兆，用来构建、画像的的Bitmap数据依然很大。

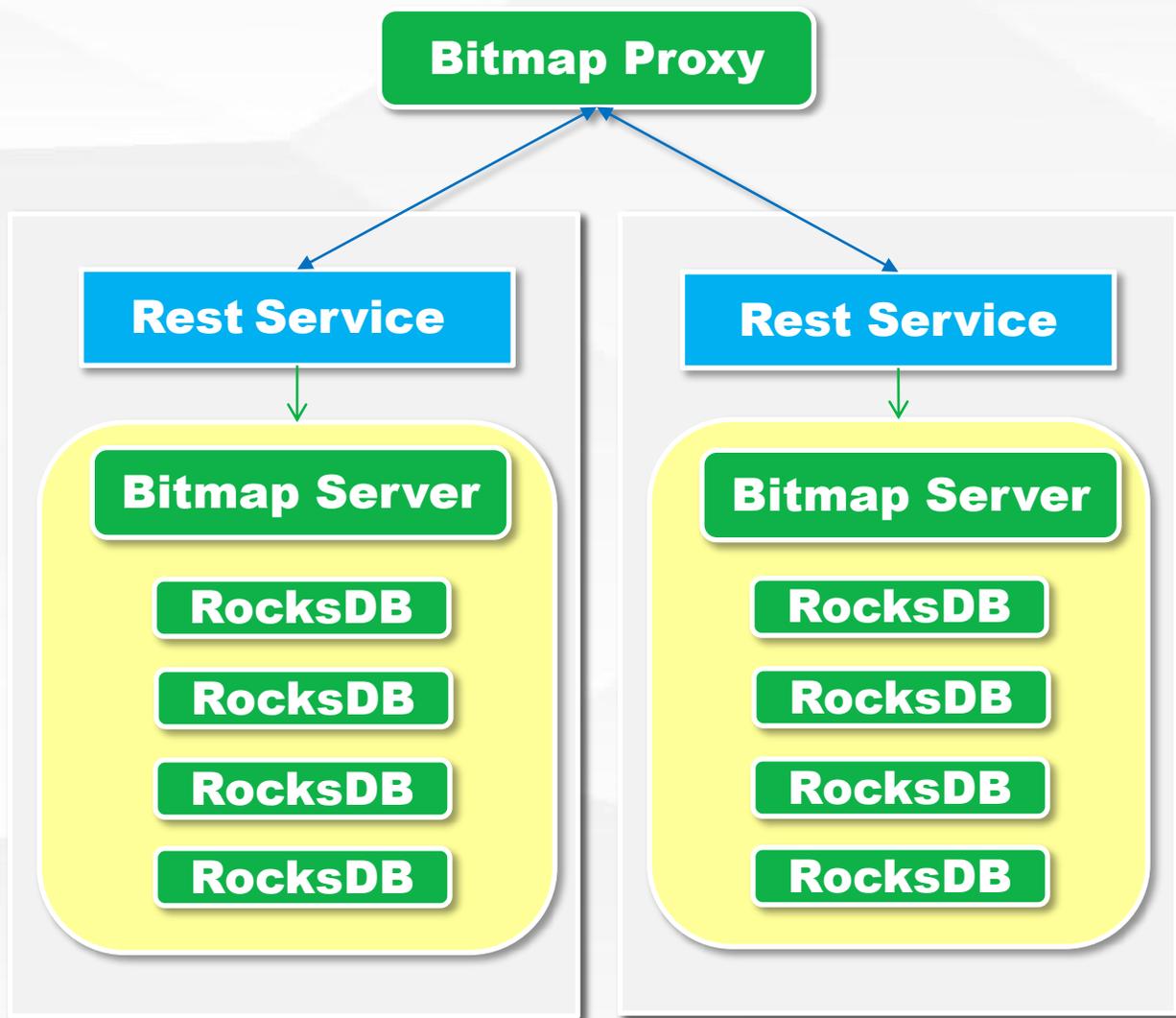
最开始的实现方案是把生成的Bitmap存在HDFS上，使用Spark集群来做这些运算，但是运算速递，特别是画像速递并不理想。一个一千万人群构建大概需要5分钟，画像常常需要1个小时才完成。



RocksDB是一个高性能的KV存储系统，读写性能很优越，使用磁盘做存储。比较适合我们这种Bitmap数量比较多，总数据量大，又需要快速读写的场景。于是我们开始调研RocksDB, 尝试使用RocksDB来解决人群构建和画像的性能问题。



RocksDB计算引擎 - 架构



➤ **Bitmap Proxy (app layer proxy)**
Round-Robin机制发送请求；

➤ **Rest Service (computing layer proxy)**
提供基于表达式的bitmap的and , or , xor , count等REST服务接口；

➤ **Bitmap Server (computing engine)**
提供分布式bitmap计算能力，计算完成时同时将操作结果进行存储；

➤ **RocksDB (persistence layer)**
Key-Value存储系统



RocksDB计算引擎 – bitmap数据分片

offset

0 200000000 400000000 600000000 800000000 1000000000 1200000000 1400000000 1600000000 1800000000 2000000000

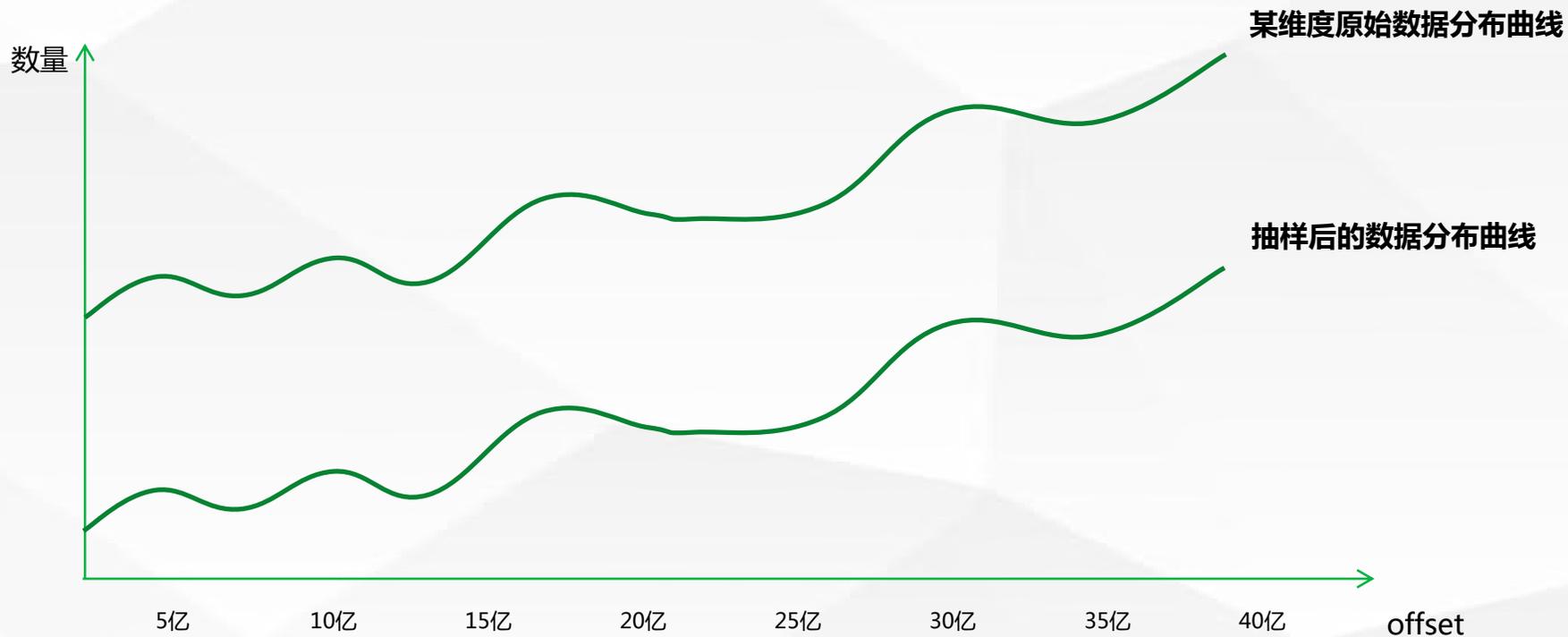


Horizontal Sharding





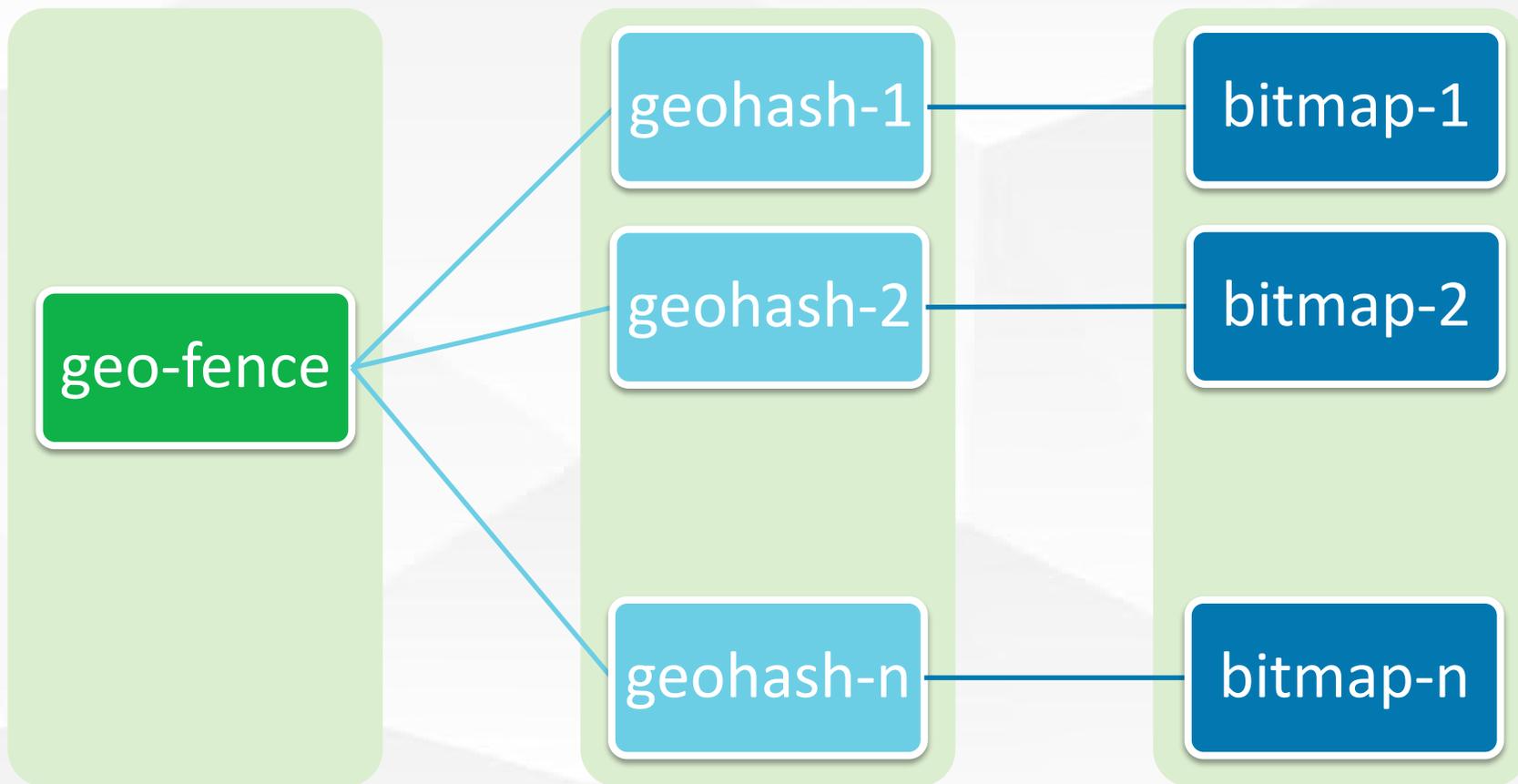
RocksDB计算引擎 – bitmap抽样能力



数据体量庞大，单一维度精度要求相对不高服务场景（例如数据交叉比对）

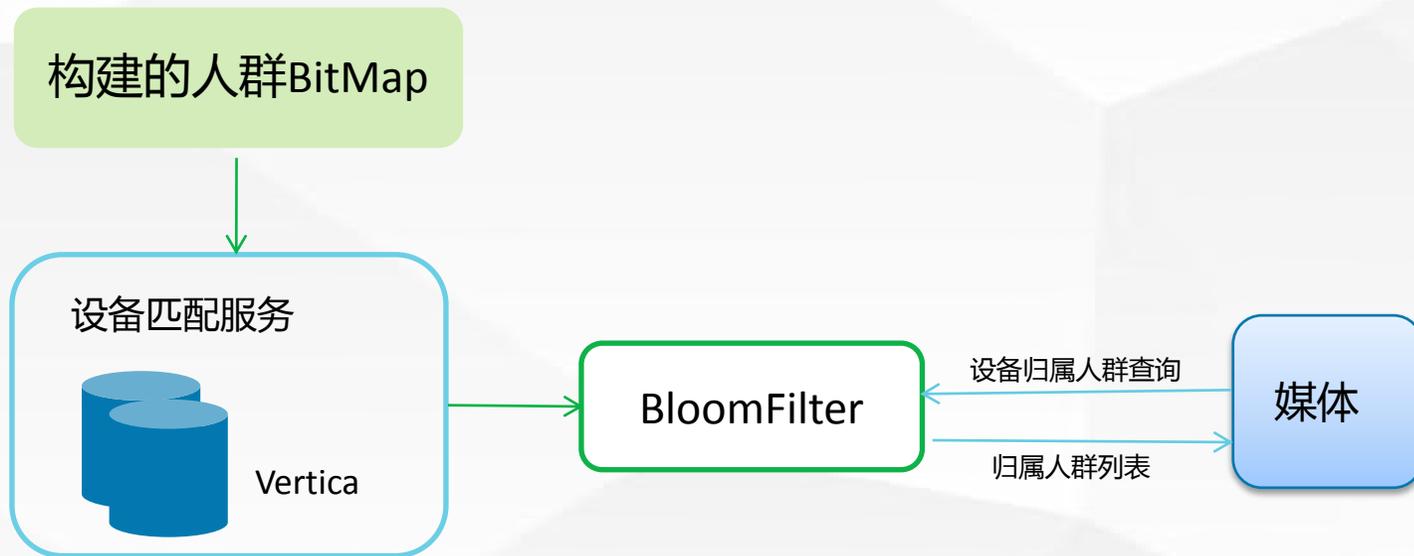


地理围栏人群预估能力





实时广告投放查询





营销云技术指标

功能模块	功能描述	数据量	处理时间
人群构建服务	标签构建	1 亿设备	< 5 secs
	上传构建	1 亿设备	< 5 mins
	Lookalike构建	1万->100万~2000万	< 15 mins
	人群预估	方圆5公里(1百万设备量)	< 1 sec
人群画像服务		15个维度 (1亿设备量)	< 2 mins
数据投放服务		5种ID类型 (1 亿设备量)	< 10 mins
前置机服务	RTB查询	5亿设备	< 5 ms (6000QPS)





远景技术规划

- **Bitmap computing engine open source;**
- **ML for segment optimization;**
- **Investigate more approach to improve “fast”;**