MediaV 聚合分析实时数据处理

肖 波

xiao_bo@mediav.com

http://weibo.com/vxiaobo

Outline

- 聚合分析总体架构
- 实时数据业务需求
- 技术框架选型
- Cassandra简介
- 性能测试
- 实际应用

聚合分析概览

为独立网站的 B2C 电商企业打造的专业数据统计分析系统。通过分析其网站用户体验、点击流、电商绩效等指标，形成网站商业分析报表，最终实现优化其线上业务表现的目标。

聚合分析
MediaV Business Analytics

产品使命

- 我们只分析对电商有价值的数据。

产品优势

- 可视化：信息高度图形化，解读数据更容易
- 准确性：对订单来源进行多渠道归因，更准确
- 便利性：预置统计代码，一键开通，安装更便利
- 垂直性：针对电商贴身设计分析模型，更垂直
- 实时性：分钟级延迟，让商业分析变得更实时

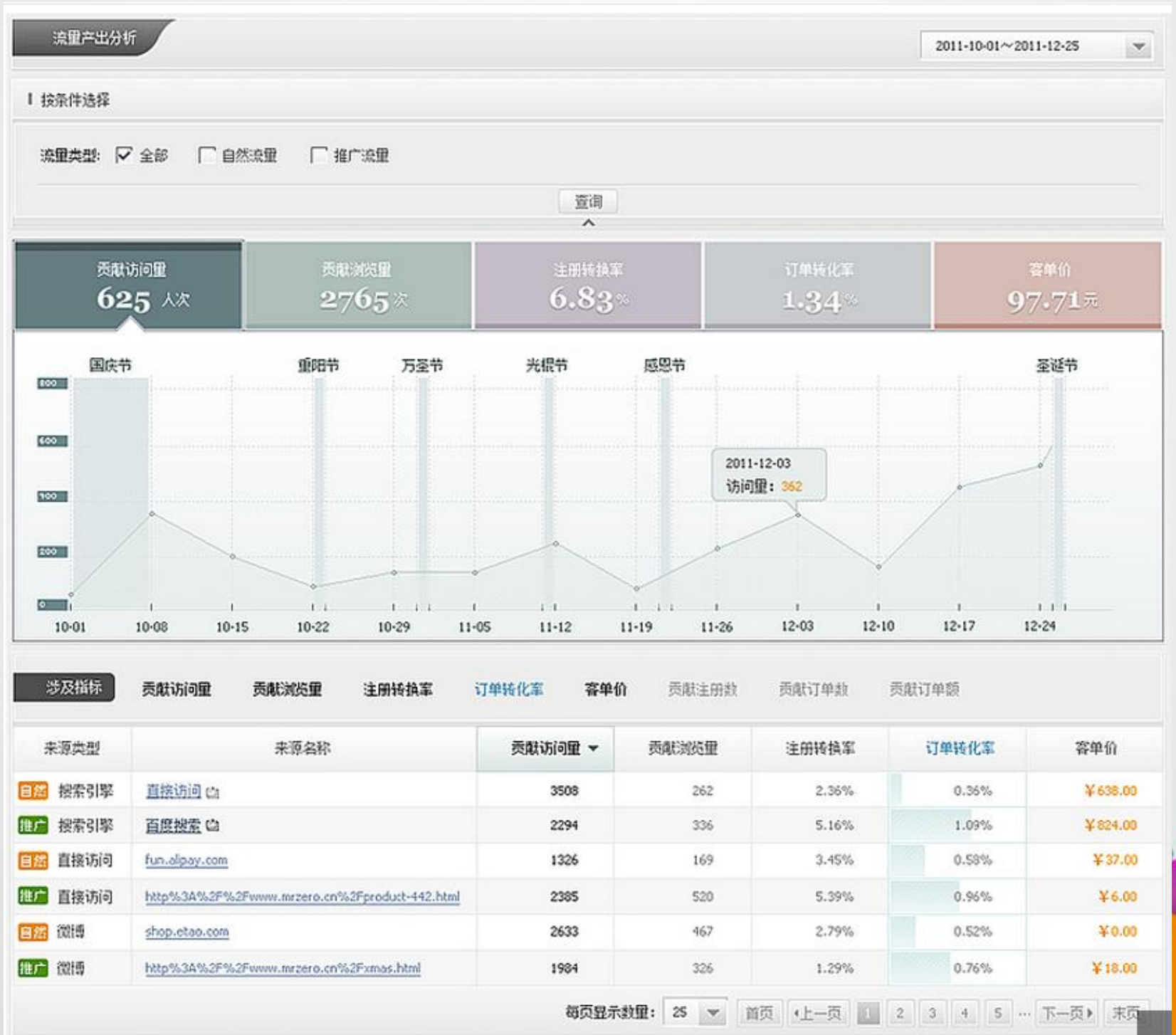
服务对象

- 内测阶段，面向使用商派（shopex）易开店和ECstore系统的中小电商企业。2012年4月起，开始服务独立电商B2C网站。

数据联播


访客在线趋势
今天
昨天
100
50
0
00:00 01:00 02:00 03:00 04:00 05:00 06:00 07:00 08:00 09:00 10:00 11:00 12:00 13:00 14:00 15:00 16:00 17:00 18:00 19:00 20:00 21:00 22:00 23:00
访问量
625 人次
昨日此时：522 人次
↑25.46%
浏览量
2736 次
昨日此时：1573 次
--
交易总额
3220 元
昨日此时：4638 元
↓27.92%
订单量
132 笔
昨日此时：107 笔
0.00%
16 : 27 : 42
2011年11月23日 星期三
历史最高
上周
本周
1200
800
400
0
1160
520 360 380
630 330 420
760 540 680
720 460 460
740 410 470
1160 620 750
530 230 460
星期一
星期二
星期三
星期四
星期五
星期六
星期日
订单转化率
1.1682%
昨日此时：1.465%
↓ 0.324%
今日热销商品
销售价:￥3899.00

流量产出分析


流量产出分析
2011-10-01～2011-12-25
按条件选择
流量类型: 全部 自然流量 推广流量
查询
贡献访问量
625 人次
贡献浏览量
2765 次
注册转换率
6.83%
订单转化率
1.34%
客单价
97.71元
国庆节
重阳节
万圣节
光棍节
感恩节
圣诞节
2011-12-03
访问量: 362
10-01
10-08
10-15
10-22
10-29
11-05
11-12
11-19
11-26
12-03
12-10
12-17
12-24
涉及指标
贡献访问量
贡献浏览量
注册转换率
订单转化率
客单价
贡献注册数
贡献订单数
贡献订单额
来源类型
来源名称
贡献访问量
贡献浏览量
注册转换率
订单转化率
客单价
自然 搜索引擎 直接访问 3508 262 2.36% 0.36% ¥638.00
推广 搜索引擎 百度搜索 2294 336 5.16% 1.09% ¥824.00
自然 直接访问 fun.alipay.com 1326 169 3.45% 0.58% ¥37.00
推广 直接访问 http%3A%2F%2Fwww.mzero.cn%2Fproduct-442.html 2385 520 5.39% 0.96% ¥6.00
自然 微博 shop.etao.com 2633 467 2.79% 0.52% ¥0.00
推广 微博 http%3A%2F%2Fwww.mzero.cn%2Fxmas.html 1984 326 1.29% 0.76% ¥18.00
每页显示数量: 25
首页
上一页
1 2 3 4 5 ... 下一页
末页

站内运营分析

来源名称	访问量	浏览量	订购量	订购转化率	订购额
首页	3508	262			
类目-耐克鞋	2294	336			
新款正品耐克/nike SB韩版冬季高帮休闲鞋	1326	169	3.45%	0.58%	￥367.00
耐克正品 NIKE Air Force 1空军高帮板鞋	2385	520	5.39%	0.96%	￥426.00
限量 NIKE DUNK SB 耐克板鞋	2633	467	2.79%	0.52%	￥539.00
冬季新款正品NIKE耐克加毛保暖鞋高帮板鞋	1984	326	1.29%	0.76%	￥168.00

每页显示数量: 25 首页 ‹上一页 1 2 3 4 5 … 下一页› 末页

Outline

- 聚合分析总体架构
- 实时数据业务需求
- 技术框架选型
- Cassandra简介
- 性能测试
- 实际应用

实时数据业务需求

- 报表VS实时？
  - –实时运营：用户购买意愿的不可持续性
  - –实时反馈：站内运营活动及时监控与调整
  - –实时监控：流量异常及时报警
- 实时数据的特点
  - –资源占有度高
  - –计算复杂度高
  - –容错空间小

聚合分析实时业务指标

- 数据联播
  - –访问数
  - –浏览量
  - –订单数
  - –毛订单额
  - –订单转化率
- 店铺摄像头
  - –用户在线趋势
  - –店内页面访问者详情

聚合分析实时数据的大数据特征

- 20K 独立网店
- 单日峰值pv 5亿
- 单日峰值处理~200G日志
- 单日峰值实时数据增长30G+
- 大量的写操作，尤其是counter类型
- 读相对较少

Outline

- 聚合分析总体架构
- 实时数据业务需求
- 技术框架选型
- Cassandra简介
- 性能测试
- 实际应用

Realtime BigData DB

Memcached

Redis

MongoDB

Hbase

Cassandra

古希腊神话的杯具预言家

- Digg 的Cassandra杯具
  - 工程副总裁John Quinn在Digg V4中使用Cassandra取代Mysql，导致上线后网站经常宕机。
  - Quinn遭遇重大反对，至少遭遇了严重的短期问题，他也因此丢掉了在Digg的工作。
  - “Cassandra数据库速度更快，但或许它仍然处于实验期，也或者是Digg正在对Cassandra数据库进行测试，总之Cassandra的运行状况并不能令用户满意。” CEO Kevin Rose

Twitter

摘自:

http://www.slideshare.net/kevinweil/rainbird-realtime-analytics-at-twitter-strata-2011

MediaV

Best used

- **Kristóf Kovács CTO, partnerSYS**
  - http://kkovacs.eu/cassandra-vs-mongodb-vs-couchdb-vs-redis
- Write more than you read (logging)
- One natural niche is real time data analysis

Outline

- 聚合分析总体架构
- 实时数据业务需求
- 技术框架选型
- Cassandra简介
- 性能测试
- 实际应用

Cassandra概览

分布式无中心

弹性可扩展

高可用与容错

可调节的一致性

面向行

高性能

CAP

- CAP
  - Consistency
  - Availbility
  - Partition Tolerance

CAP理论指出，同时只能具有这三个特性中的两个。

Cassandra ： AP，最终一致性，拥有跨Data Center同步的能力

数据模型

Cluster

Keyspace：数据的最外层容器，类似关系型数据库

Column family：容纳一组有序行的容器，每行包含一组有序列

Column：最基本数据结构单元，名称、值、时钟构成的三元组

Super column：value是一个子列的映射（一起查询的内容放一起）

五维哈希：

[Keyspace][Column family][Key][Super column][Column]

架构设计

P2P：对等结构，可用性和可扩展性

Gossip：流言协议用于故障检测（增量），故障节点计入列表

Anti-entropy：逆熵，副本同步机制，邻居交换Merkle树比对

Memtable、SSTable、Commit log：数据写入Commit log则认为写成功，Commit log可用于数据修复。

Hinted handoff：提示移交，提升弱一致性级别的写性能（ANY）

Bloom filter：判断元素是否存在于集合的超快速、不确定的判断算法，可看做查询的缓存，假阳性结果

Tombstone：删除标记，合并SSTable时清理

为什么写快

写优化是Cassandra的设计决策。

Memtable和Commit log的存在，写一个值不需要任何的读或者定位操作，所有的写都是以追加方式顺序写入的。

Outline

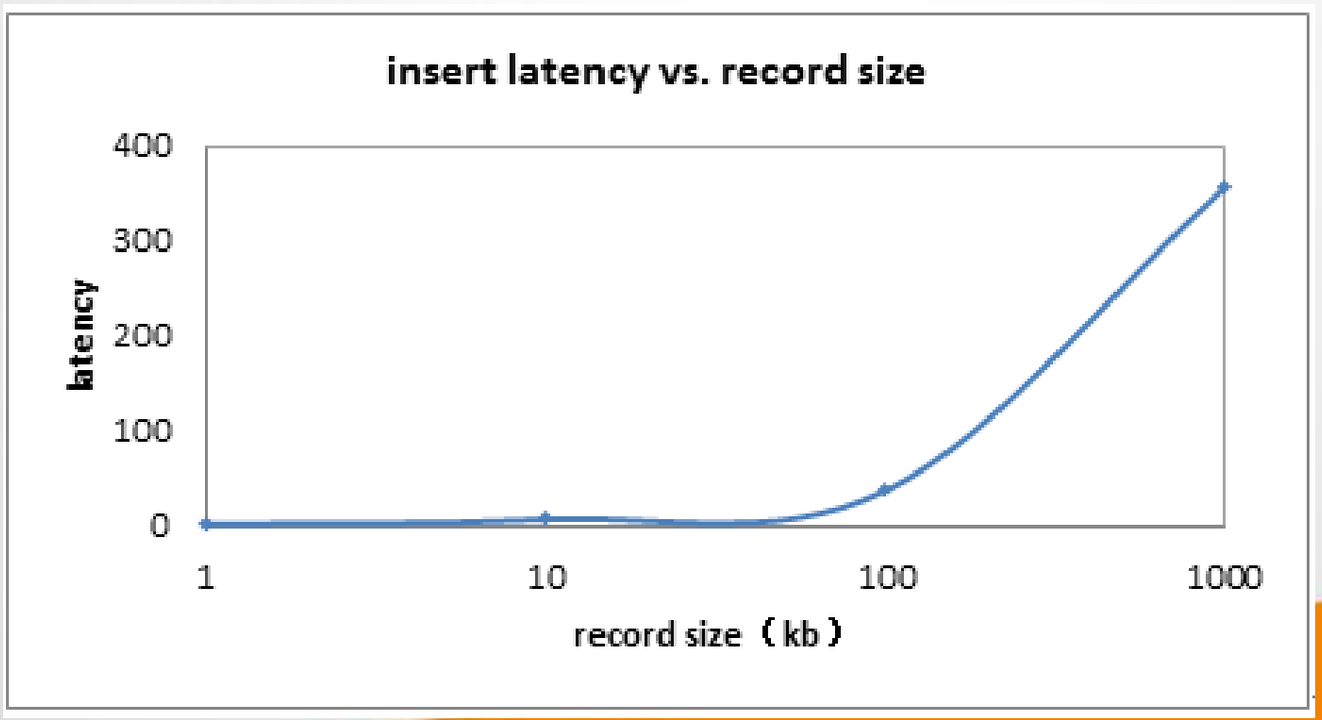
- 聚合分析总体架构
- 实时数据业务需求
- 技术框架选型
- Cassandra简介
- 性能测试
- 实际应用

测试环境

- 硬件
  - 6 node
- 软件
  - Cassandra v0.6.6，Thrift API
- 数据
  - Num of key space = 1, num of column family = 1, replication factor = 2
  - Byte Ordered Partition，共约60G（平均分布于6个结点）
    - 1kb/行的数据共15000000行
    - 10kb/行的数据共1500000行
    - 100kb/行的数据共150000行
    - 1000kb/行的数据共15000行
- 测试方式
  - Yahoo! Cloud Serving Benchmark

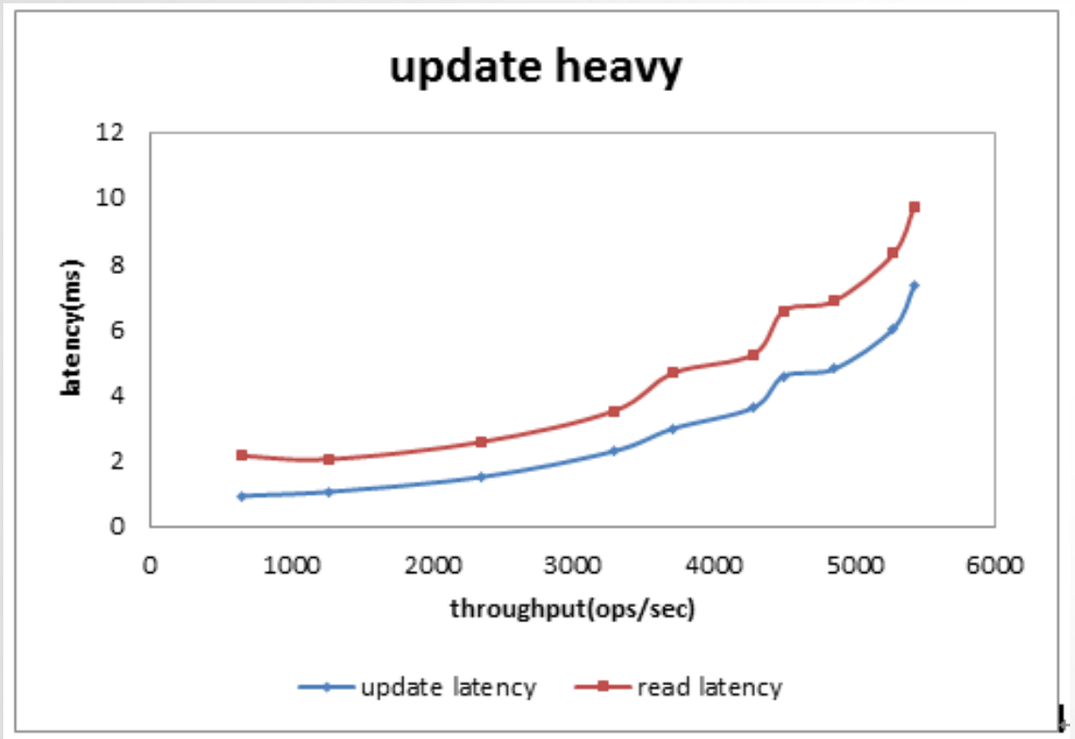
Load Data

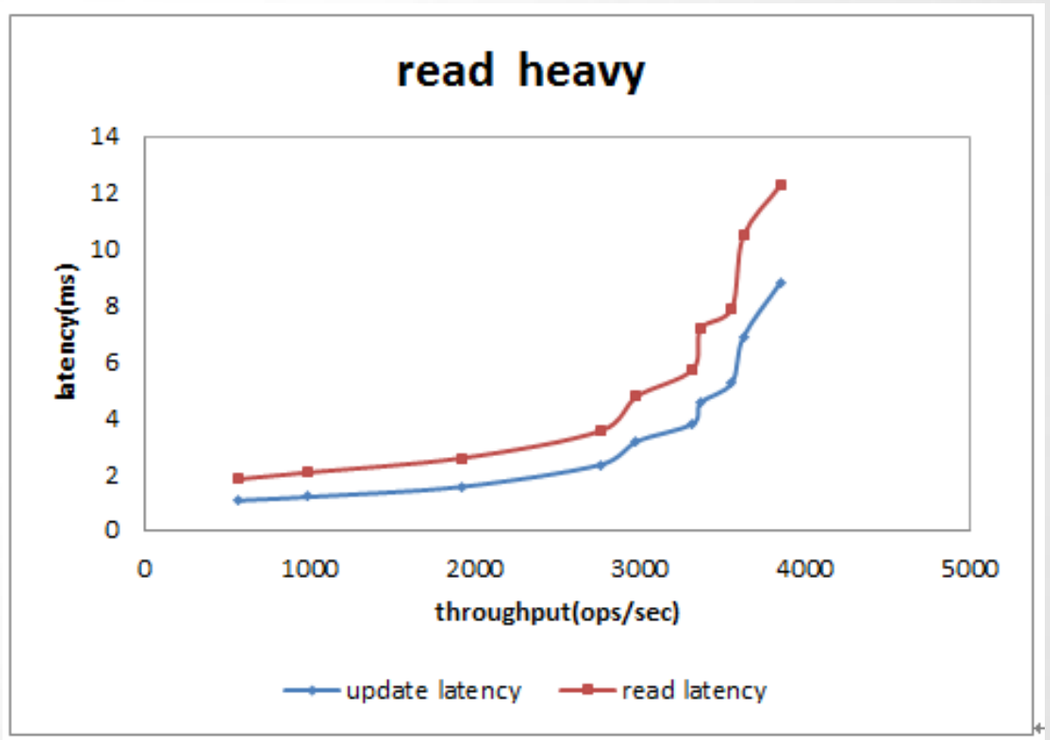
- Cassandra load数据时间
  - –1kb/行的数据共15000000行： 3662 seconds
  - –10kb/行的数据共1500000行： 1531seconds
  - –100kb/行的数据共150000行： 1057seconds
  - –1000kb/行的数据共15000行：910seconds



MediaV

Update Heavy vs Read Heavy





Cassandra VS Hbase

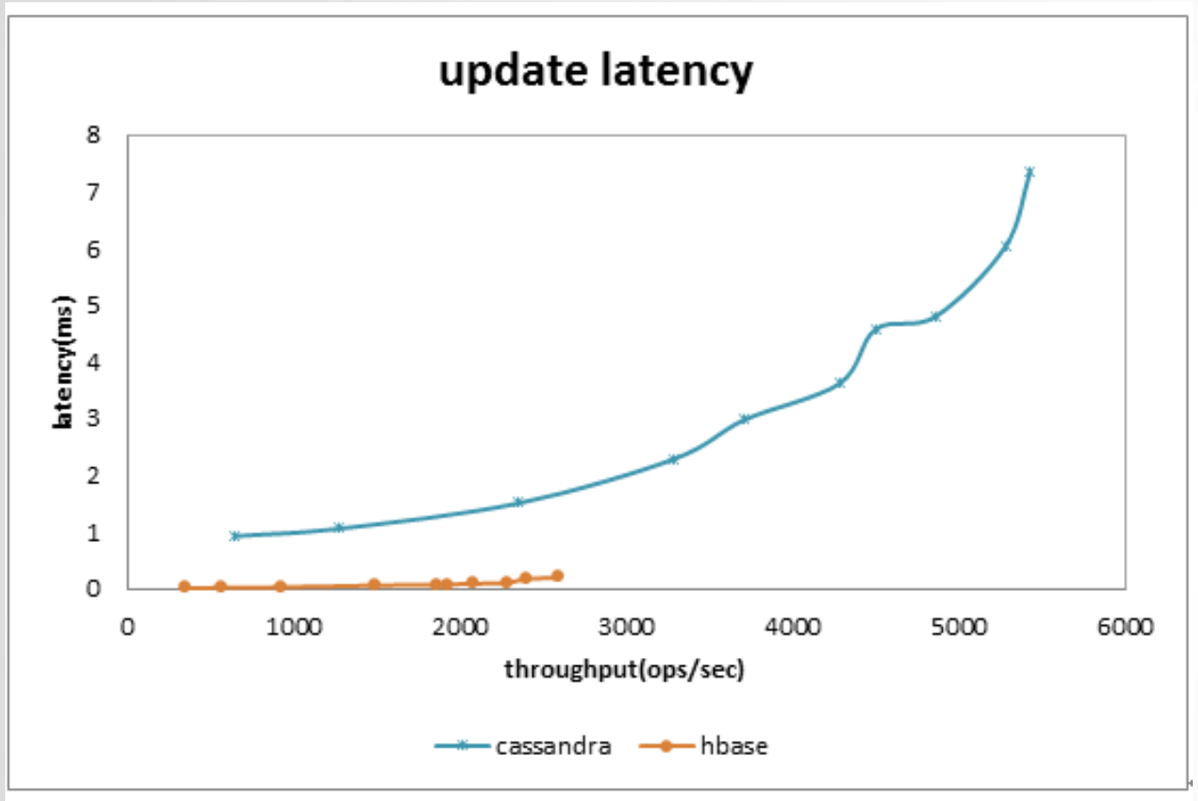
- Hbase Introduction
  - –Yet Another NoSQL database
  - –A Key-Value database
  - –A kind of BigTable implementation
  - –Built-in MapReduce processing
  - –HDFS data storage/Fault tolerance
  - –Great Scalability

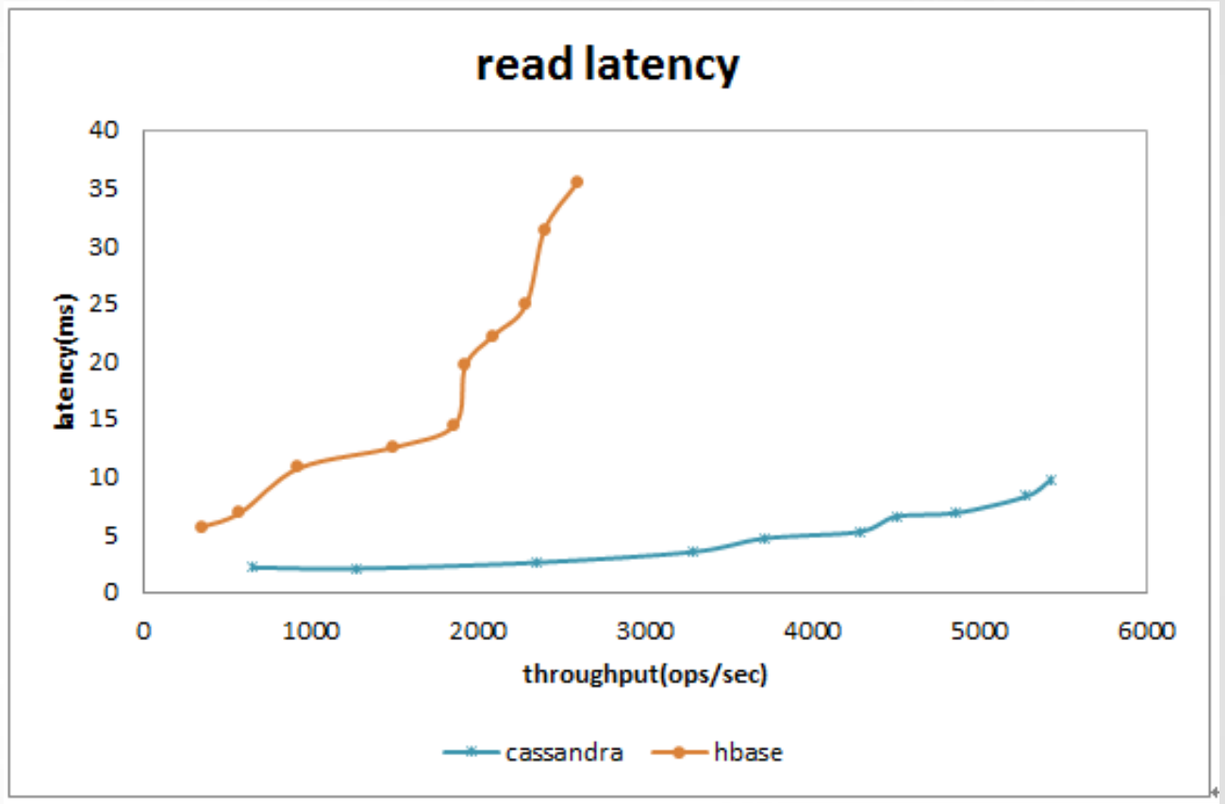
HBase = HDFS + Random read/write

测试环境

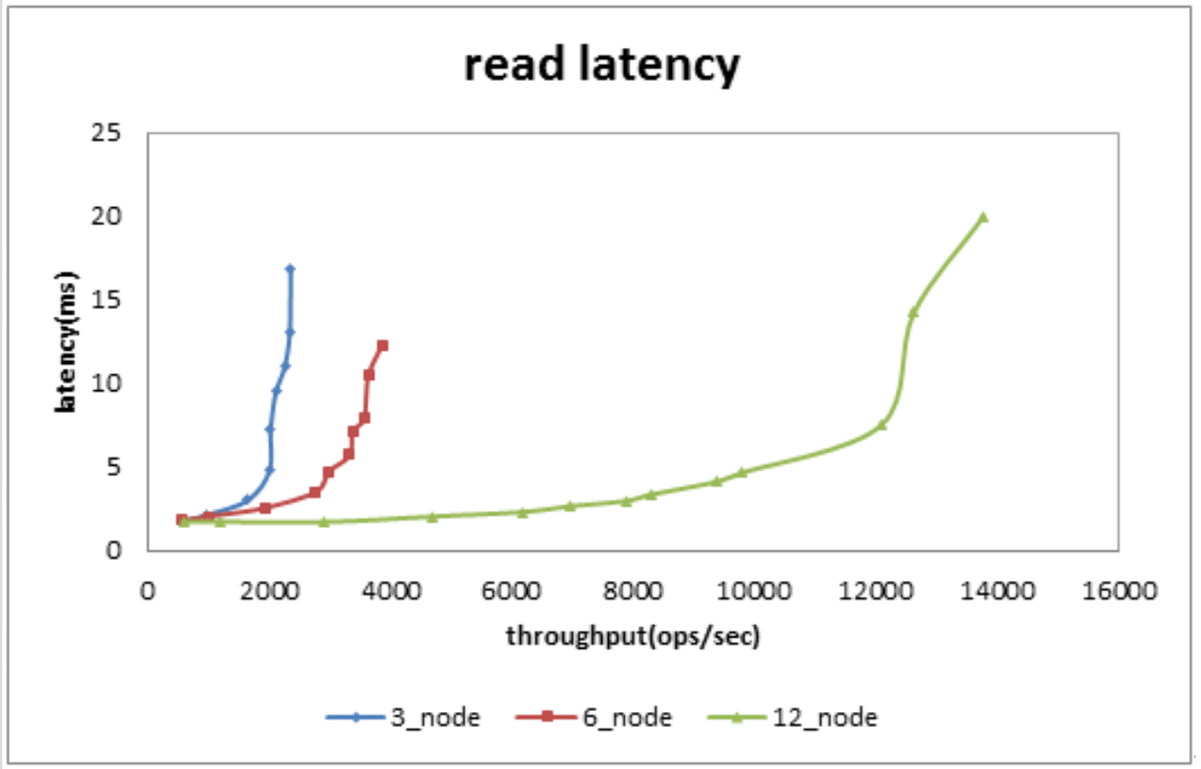
- **参数**
  - Cassandra：KeysCached设为0，其余参数均取缺省值
  - Hbase：缺省值
- **数据量**
  - 实验的数据总量均在60G左右
  - 各有六个节点，数据平均分布
- 测试方式
  - Yahoo! Cloud Serving Benchmark

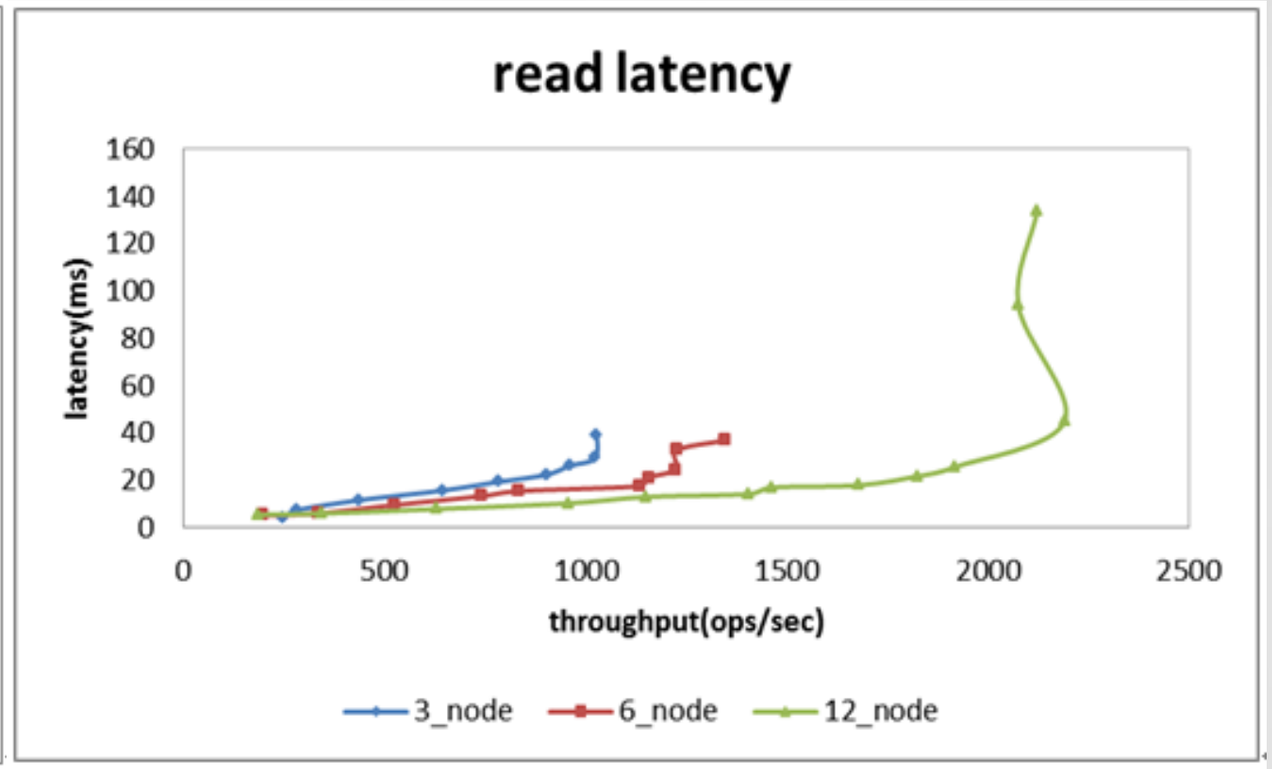
Update & Read


update latency
latency(ms)
0
1
2
3
4
5
6
7
8
0
1000
2000
3000
4000
5000
6000
throughput(ops/sec)
cassandra
hbase


read latency
latency(ms)
0
5
10
15
20
25
30
35
40
0
1000
2000
3000
4000
5000
6000
throughput(ops/sec)
cassandra
hbase

Scalability





Cassandra

Hbase

MediaV

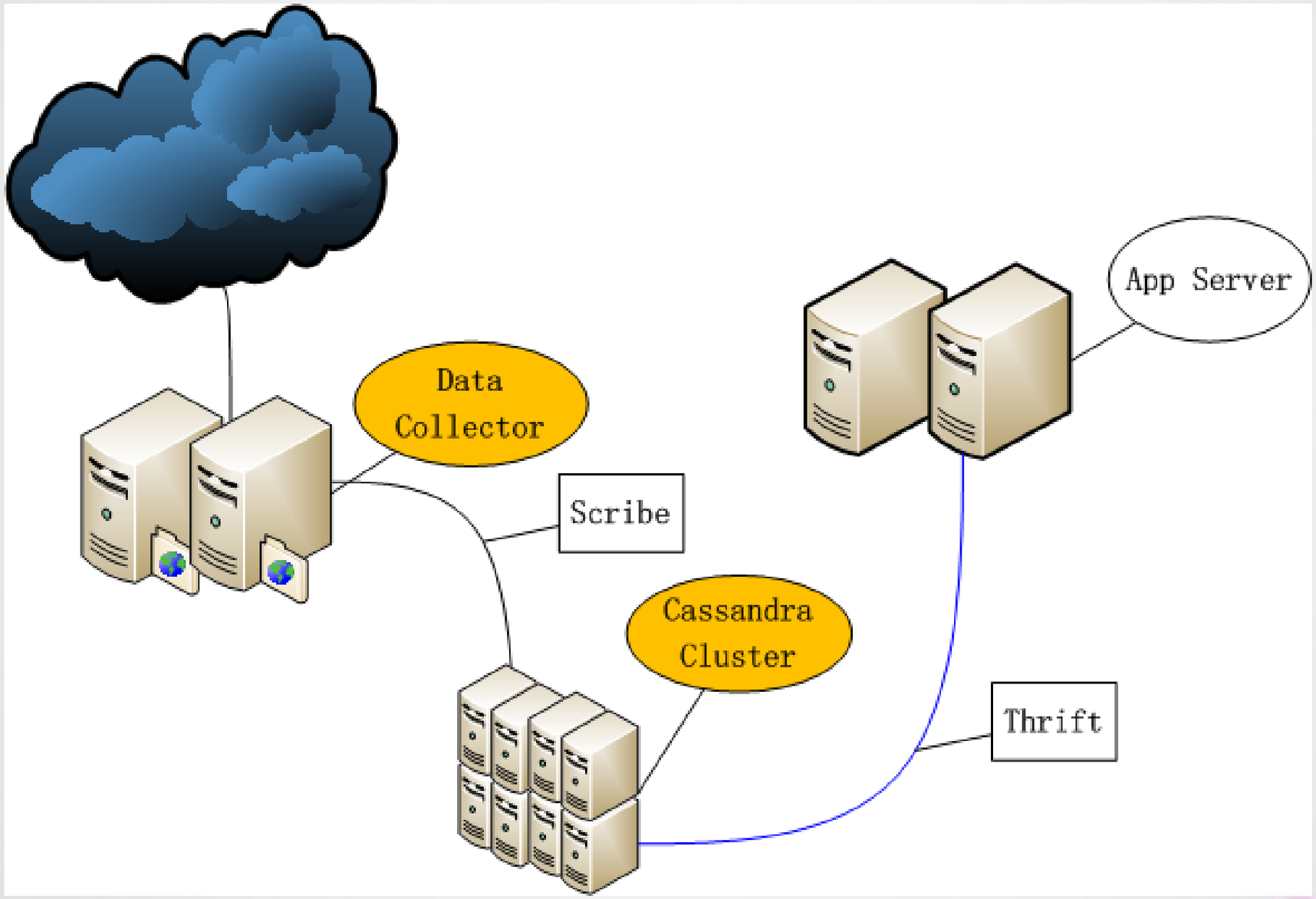
Cassandra vs. Hbase

- **Cassandra与Hbase写操作的延时都很小。**
- **Cassandra写操作的延时较Hbase大。**
- **Cassandra读操作的延时较Hbase小很多。**
- **Cassandra范围查询的延时较Hbase大，这在完全随机的请求下表现的尤为明显。**
- **Cassandra的Scalability较Hbase好。**

Outline

- 聚合分析总体架构
- 实时数据业务需求
- 技术框架选型
- Cassandra简介
- 性能测试
- 实际应用

Data Flow


Data
Collector
Scribe
Cassandra
Cluster
App Server
Thrift

Log Collection&Distribution

- MediaV第二代日志服务器
  - –服务于广告投放系统和其他业务系统
  - –数十台机器组成的集群
  - –跨DataCenter部署
  - –性能: 每天数十亿PV
- Scribe 日志分发模块
  - –Facebook开源
  - –负责在分布式系统间收集和分发日志
  - –基于Category，同一类日志可以分发到不同业务系统

Thrift Sever&Clinet

- Thrift
  - Facebook开源
  - 跨语言、跨机器应用之间的通信、RPC调用协议
- Thrift Server
  - LogProcess 负责接收Scribe Push过来的每一条实时日志，多线程并发入库，有多个实例
  - QueryProcess 负责相应应用层过来的查询请求，多线程并发查询，并缓存查询数据，多个实例
- Thrift Client
  - 应用层向QueryProcess 请求数据，并按照站点hash选取QueryProcess 实例

Cassandra Schema

例子：实时统计网店的今日访问数据

CF_Schema:

Create column family counter_cf

WITH default_validation_class=CounterColumnType

AND key_validation_class=UTF8Type

AND comparator= UTF8Type

Meta_data:

Rowkey = siteId

column name

visit#time

value

counter_value

数据更新：INCR counter_cf [‘siteId_0'][‘visit#1320310080000’] BY 1

Cassandra Java Driver - Hector

```
// init
Cluster cluster = HFactory.getOrCreateCluster("Test Cluster", "192.168.13.46:9160");
Keyspace keyspaceOperator = HFactory.createKeyspace("Keyspace1", cluster);
//InsertSingleColumn
Mutator<String> mutator = HFactory.createMutator(keyspaceOperator, StringSerializer.get());
mutator.insert("jsmith", "Standard1", HFactory.createStringColumn("first", "John"));
//InsertSuperColumn
 mutator.insert("billing", "Super1", HFactory.createSuperColumn("jsmith",
     Arrays.asList(HFactory.createStringColumn("first", "John")), stringSerializer, stringSerializer, stringSerializer));
// incr counter
mutator.incrementCounter(“rowkey”, “cfName”, “clName”, incr);
// ttl
mutator.addInsertion((“rowkey”, “cfName”, HFactory.createColumn(“clName”, Ttl, stringSerializer,
     stringSerializer));
// query
rangeSlicesQuery.setColumnFamily("Standard1");
rangeSlicesQuery.setKeys(“rowkey1", “rowkey2");
rangeSlicesQuery.setRange(“startClName", “endClName", true, 3);
```

MediaV

Cassandra Cluster

- 集群目前机器总数量<10台, replication factor设置为2
  - –单台Counter写负载约在 4000 ops/sec
  - –单台其他写负载约在 2000 ops/sec
  - –单台读负载约在 500 ops/sec
  - – Zabbix 监控流量及性能数据

参考内容

http://www.fenxi.com/

http://baike.baidu.com/view/7357471.htm

http://v.youku.com/v_show/id_XMzQyNTA1MDYw.html

Cassandra The Definitive Guide

Cassandra High Performance Cookbook

http://thrift.apache.org/

https://github.com/zznate/hector-examples

https://github.com/brianfrankcooper/YCSB/wiki

广告时间

- MediaV：
  - 创建于2009年6月，专注于精准互联网广告，目前在北京和上海设有研发中心。Mediav愿景：用互联网技术知识推动中国互联网产业创新，让互联网营销变得更有价值。在这里，你将与来自Google、Microsoft、Yahoo、百度、阿里巴巴的工程师一起工作，做分析建模型，用数字和逻辑改变互联网广告的未来。
  - 更多详情介绍请见：http://www.mediav.com
  - 简历投递 hr@mediav.com
  - 新浪微博 @mediavhr @mediav

MediaV