

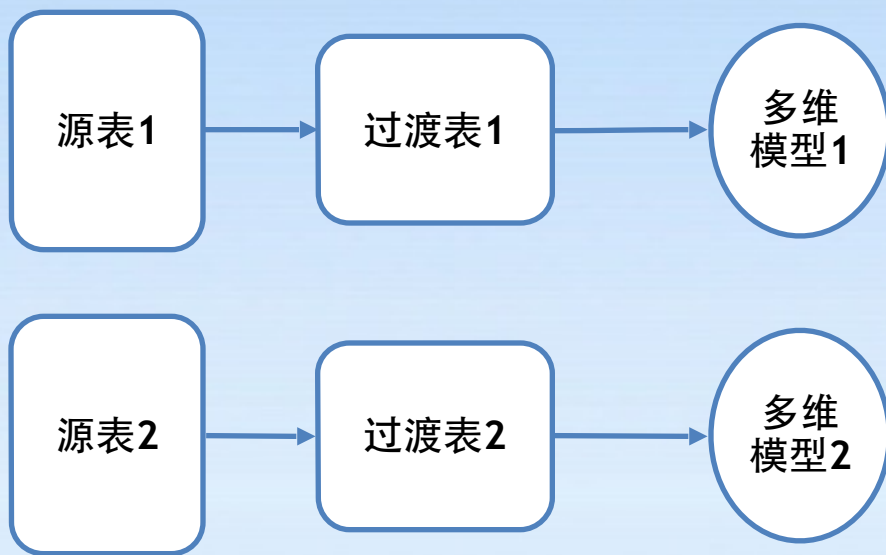
高性能，高可用，可扩展 在途牛旅游网的实际经验

2014.4.26

- 大数据
- 搜索
- 价格计算

途牛大数据模型架构

传统数据模型架构：需求驱动



- 需求驱动模型架构：

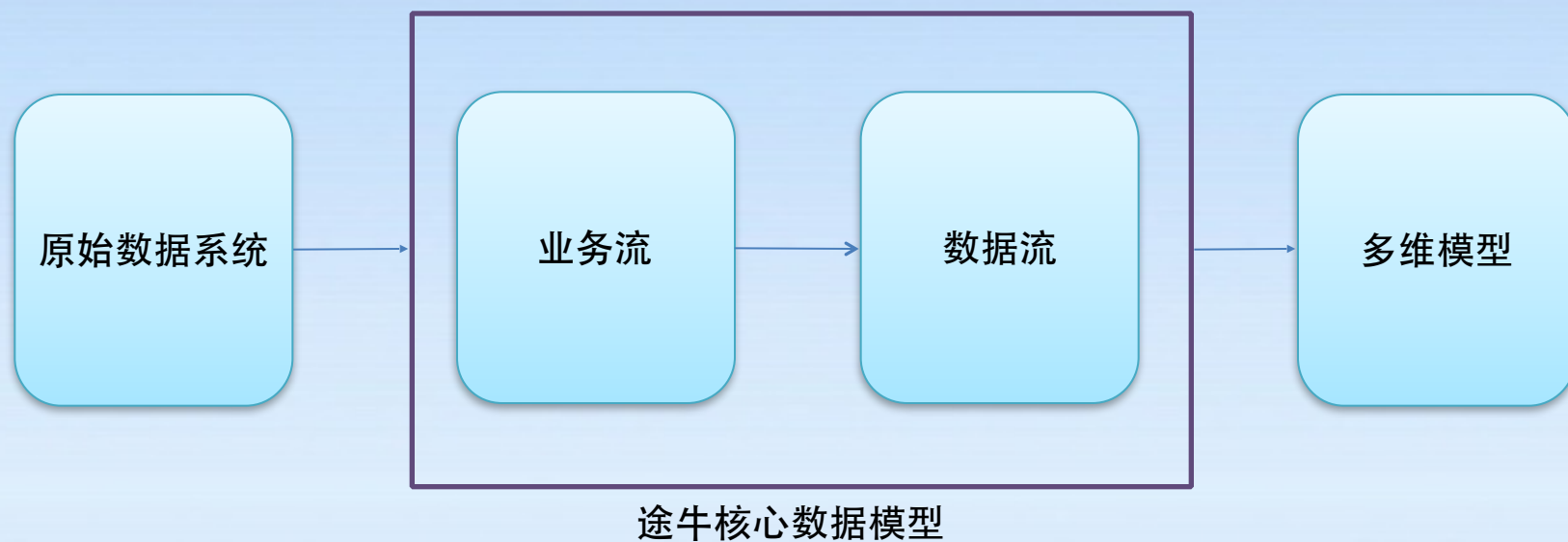
随需求变化快，模型的重构性低，大数据的处理和计算差性能

途牛业务特点：变化非常迅速，对数据的需求变化也非常快；

途牛模型架构：数据驱动模型架构，非需求驱动模型架构；

途牛大数据模型架构

途牛数据模型架构：数据驱动

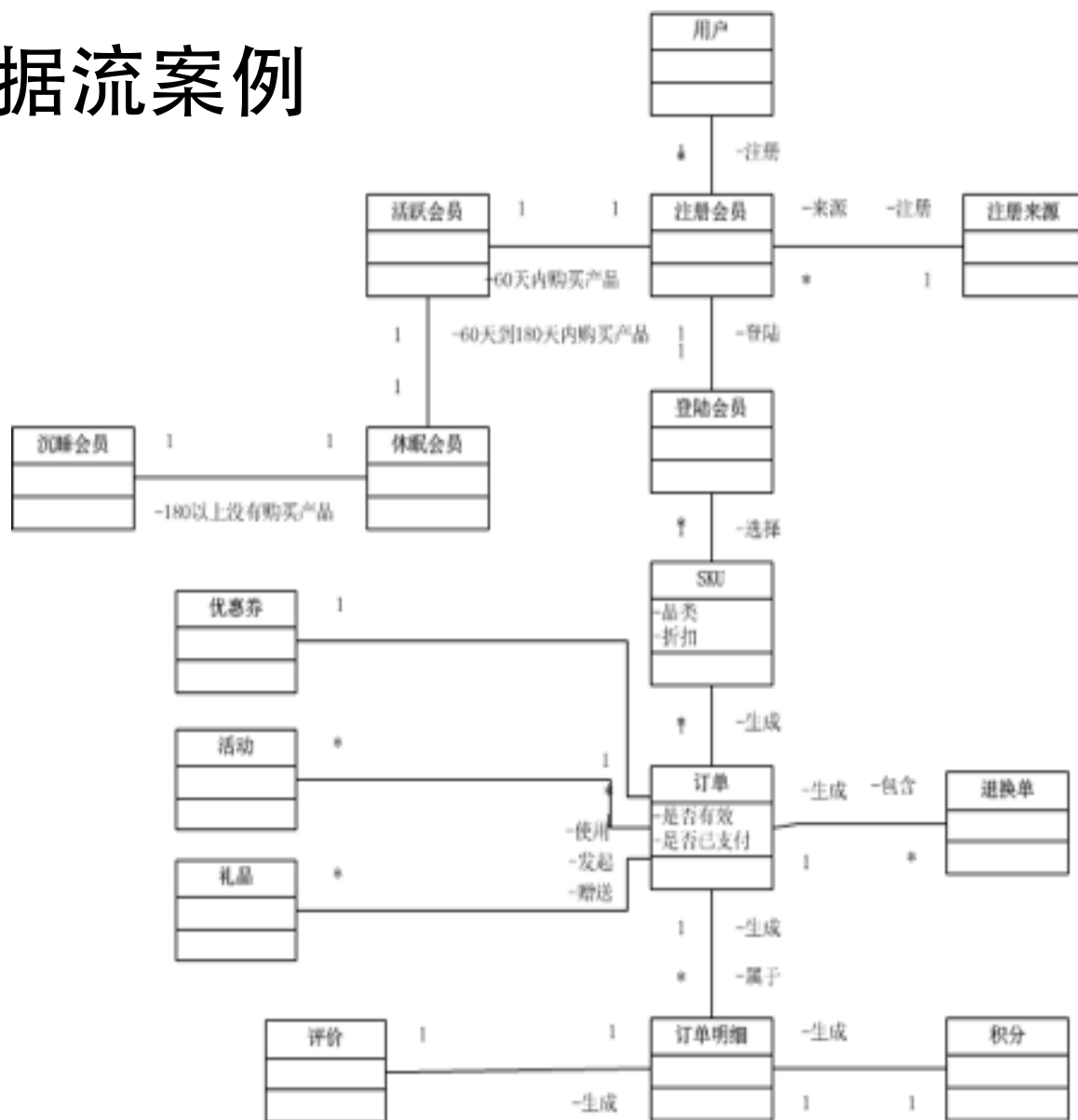


特点：

- 途牛核心层是基于公司的业务流设计的，不随需求的变动而变动；
- 需求的变动仅仅影响多维层模型；
- 模型架构复杂性低，重构性强；

途牛大数据模型架构

业务流落地成数据流案例



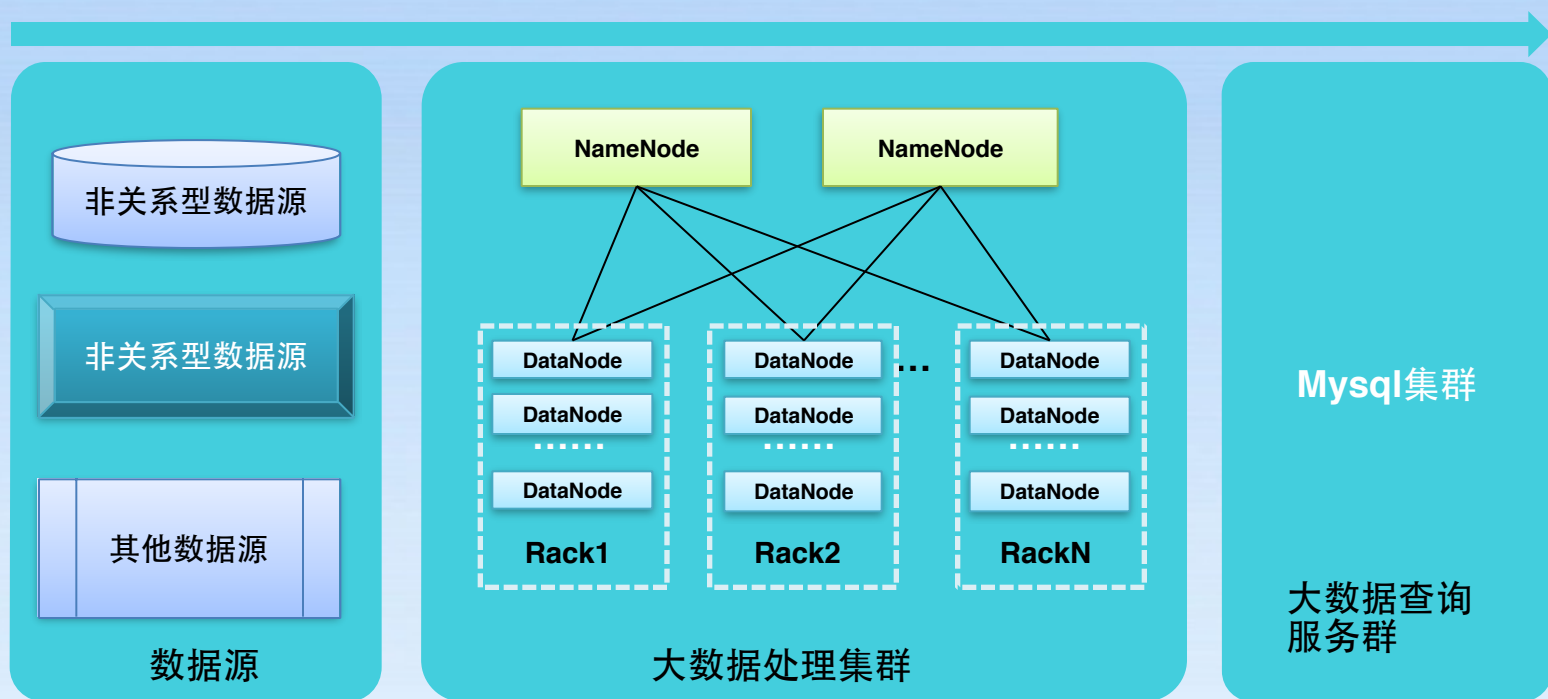
途牛大数据平台演变

老数据平台



- 数据架构：服务器架构伸缩性差，1+1的服务器搭建模式（1台数据仓库服务器+1台报表服务器）；
- 数据存储：数据量不到1T，没有包含用户行为数据；
- 系统并发：并发比较差，几十个人的并发，系统就会卡；
- 数据处理：处理时间长，每天数据处理需要8-10个小时；
- 数据运行：系统负荷高，系统异常频发；

离线架构 - 物理架构



- 数据架构：可伸缩性的服务器架构，支持多台服务器扩充，目前20多台服务器，集群可以扩张到上千台以上；
- 数据存储：包含了用户行为数据，上百T的数据容量；
- 数据并发：支持上百人同时在线并发；
- 数据处理：离线数据处理时长控制在5小时以内，实时数据处理达到5秒以内；
- 数据运行：充分利用各台服务器资源，每台服务器负荷均匀；

离线架构 - 处理架构



- 数据逻辑层次分明，控制系统的高可用性和大数据处理能力；
- 降低系统的数据耦合性，降低数据出错风险，提供数据的可用性；
- 结构化和非结构化数据进行大数据平台，进行快速的批量处理；

离线架构 - 数据处理

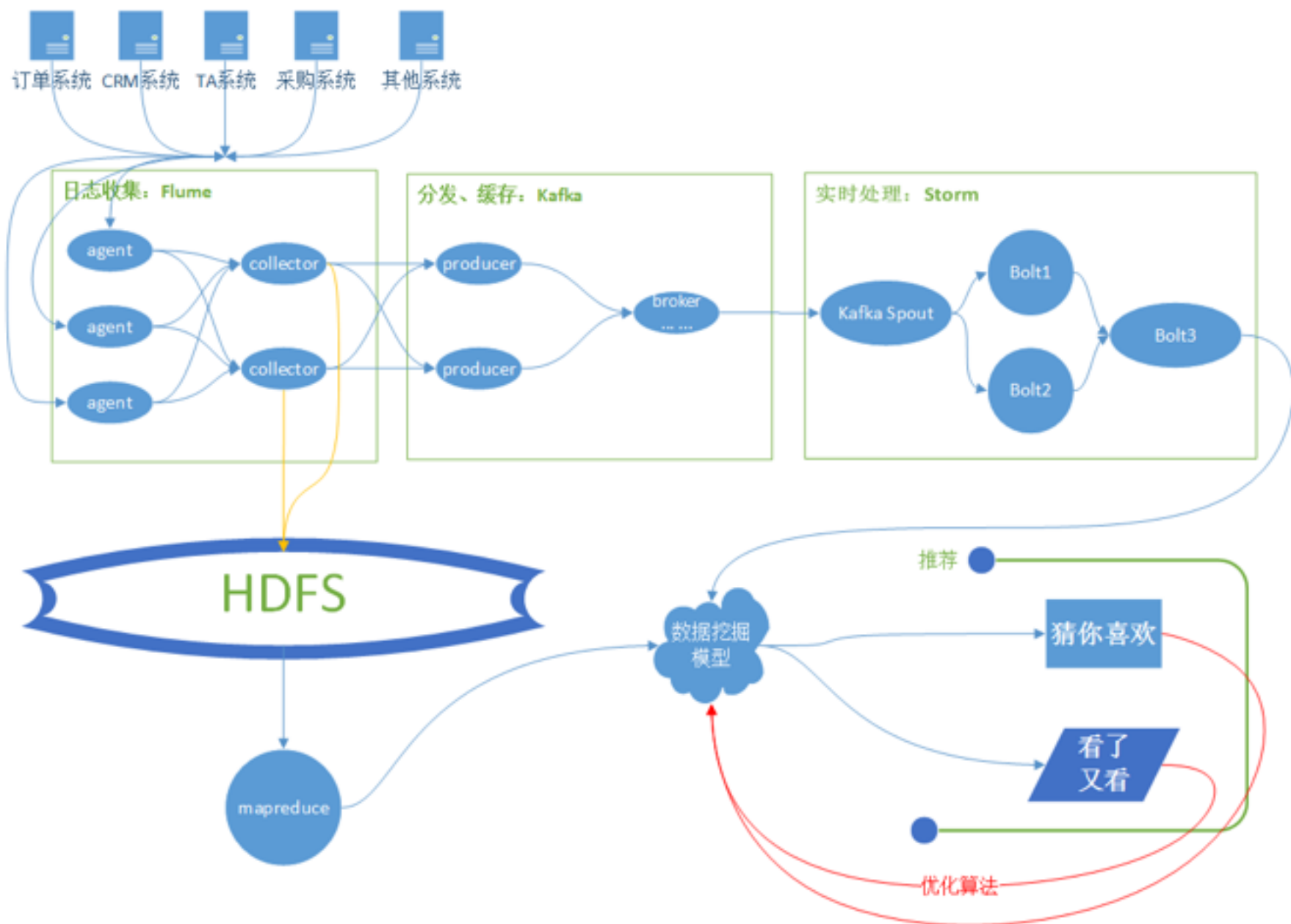
多数据系统采集处理能力：通过跨平台的数据采集服务引擎

1. 数据源系统繁多：由于各种历史缘故，途牛底层数据系统繁多，共达60多个；
2. 数据源跨平台：数据源系统分部在各不同的平台，linux、windows等；
3. 数据源格式多样：有文本数据、有各种关系型数据库（Oracle、Mysql、Sqlserver等）；

海量数据计算能力：

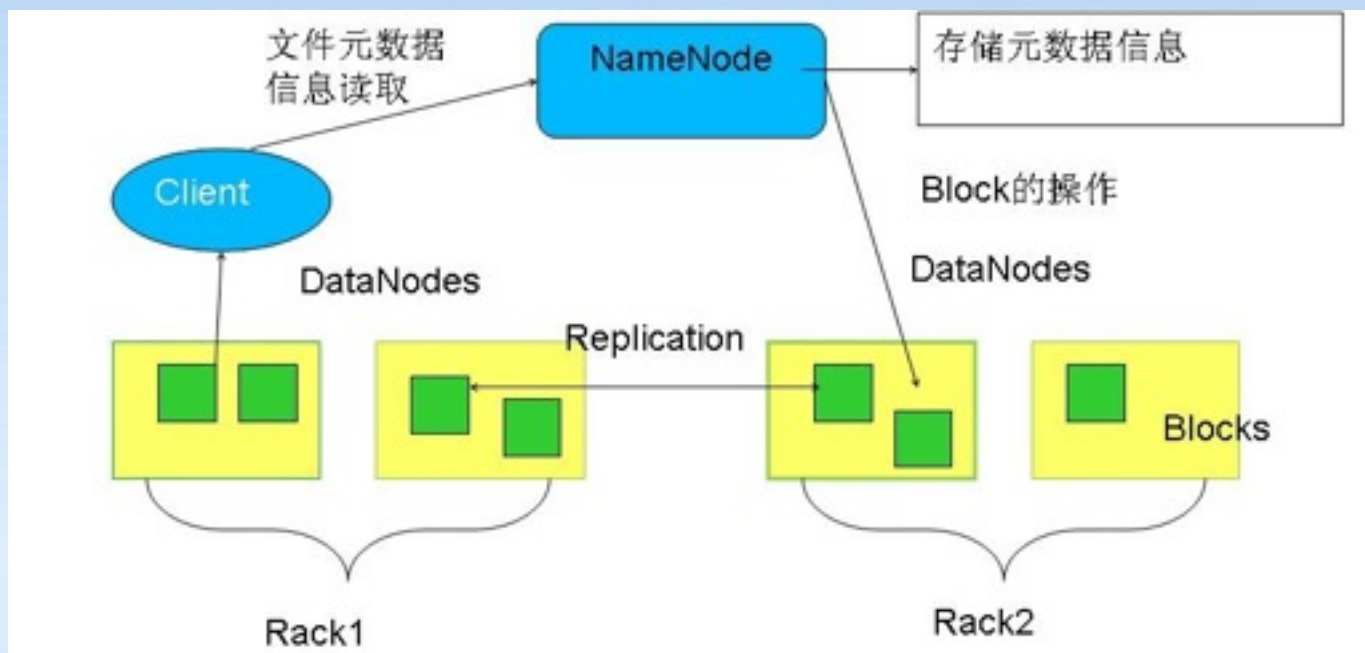
1. 大数据平台拥有处理百T级别数据的能力；
2. Hadoop分布式处理能力，各个数据节点并发处理海量数据；

实时架构



数据平台高可用性

- 软硬件设计时避免SPOF;
- 自动化监控系统

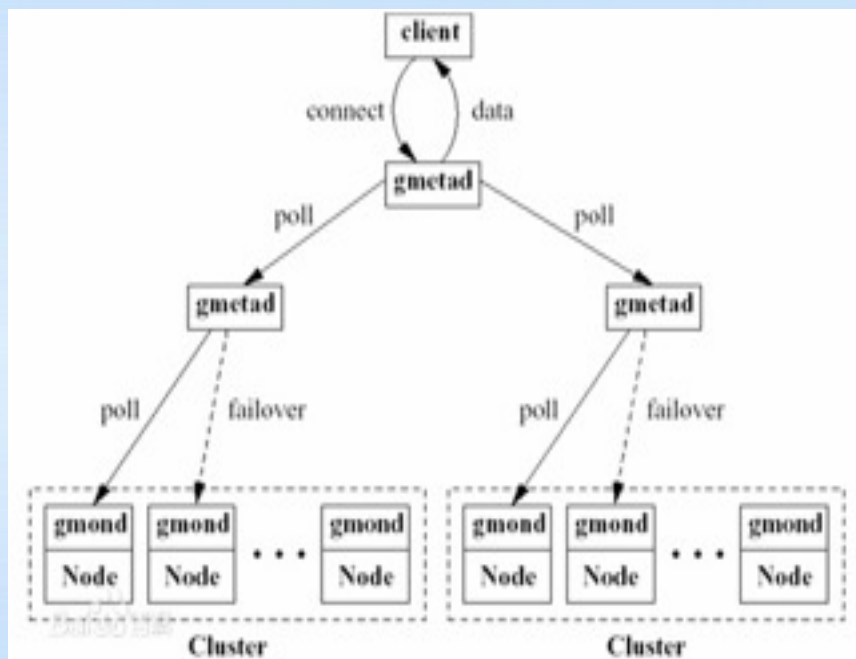


- 采用双NameNode，共享元数据库，当Active的NameNode异常后，系统能快速切换到Standby的NameNode，使得Standby的NameNode成Active，系统可以正常运行；
- 多硬盘的DataNode，每个Data Block有三份，当数据损坏（比如模块硬盘坏了或者某台服务器挂了），集群可以从同个Rack或者最近的Rack中拷贝数据Block，系统可以正常运行

数据平台高可用性

- 从机制上保证系统的可用性，尽量降低系统不可用的时长和损失；
- 建立Ganglia监控中心，建立实时报警机制，通过短信和邮件及时通知；

◆ Ganglia监控系统：



搜索

产品类型: 自助游

出发城市: 北京 重庆 广州 香港 南京 上海 成都 杭州

行程天数: 3天 5天 6天 7天

选岛: 基础 优质 豪华 奢华 至尊

欢乐岛 艾阿胡岛 美人蕉岛 梦幻岛 假日岛 马累岛 绚丽岛 天堂岛
蕉叶岛 白金岛 圣塔拉富士岛 安娜塔拉笛古岛 安娜塔拉薇莉岛 太阳岛 皇家岛

含餐: 一价全含 (AI) 早中晚餐 (FB) 早晚餐 (HB) 早餐 (BB)

上岛: 快艇 内陆飞机+快艇 水上飞机

产品特点: SPA尊享 潜水圣地 美食 亲子 蜜月 海滨海岛

酒店品牌: 希尔顿 凯悦 喜来登 卓美亚集团 唯一集团 安娜塔拉集团 泰姬集团 四季
香格里拉酒店 悦榕集团Banyan Tree 皇冠 阿达兰集团Adaaran 艾登尼兹集团Aydeniz

筛选条件

综合

请选择价格

yyyy-mm-dd — yyyy-mm-dd



<马尔代夫康杜玛岛Kandooma4晚6日自助游>广州出发

蓝白色的浪漫 优质浮潜环境 最美海底世界 全岛WIFI覆盖 细致中文服务

编号: 5112744 满意度: 100% 13人点评

团期: 05/11,05/18,05/25,05/27,06/01,06/08,06/15,06/22...

部分产品属性

10156元起

搜索

济州岛旅游

简介

攻略

景点大全

风景图片

游记

资讯

点评

地图

重置筛选条件

产品类型:

产品推荐 跟团旅游 自助游 邮轮旅游

出发城市:

全部 上海 厦门 广州 深圳 徐州 扬州 济州市

+ 多选

出游时间:

yyyy-mm-dd - yyyy-mm-dd

行程天数:

全部 1天 4天 5天 6天 7天

+ 多选

线路玩法:

全部 首尔+济州岛 济州岛一地

+ 多选

高级选项:

产品特点

- ☐ 美食
- ☐ 摄影采风
- ☐ 毕业旅行
- ☐ 游学
- ☐ 亲子
- ☐ 民俗风情
- ☐ 蜜月
- ☐ 海滨海岛
- ☐ 乐园
- ☐ 购物

综合 销量

推荐 特价 热卖

1/1

上一页

下一页

济州岛-上海3晚4日游>上海往返 皇家加勒比邮轮

编号: 2232438 满意度: 100% 新品推荐

3100元起

/27 查看团期

济州岛-上海3晚4日游>世界十大邮轮之一, 感受前所未有的海上梦幻假期

精致享受

编号: 2324261 满意度: 100% 新品推荐

7999元起

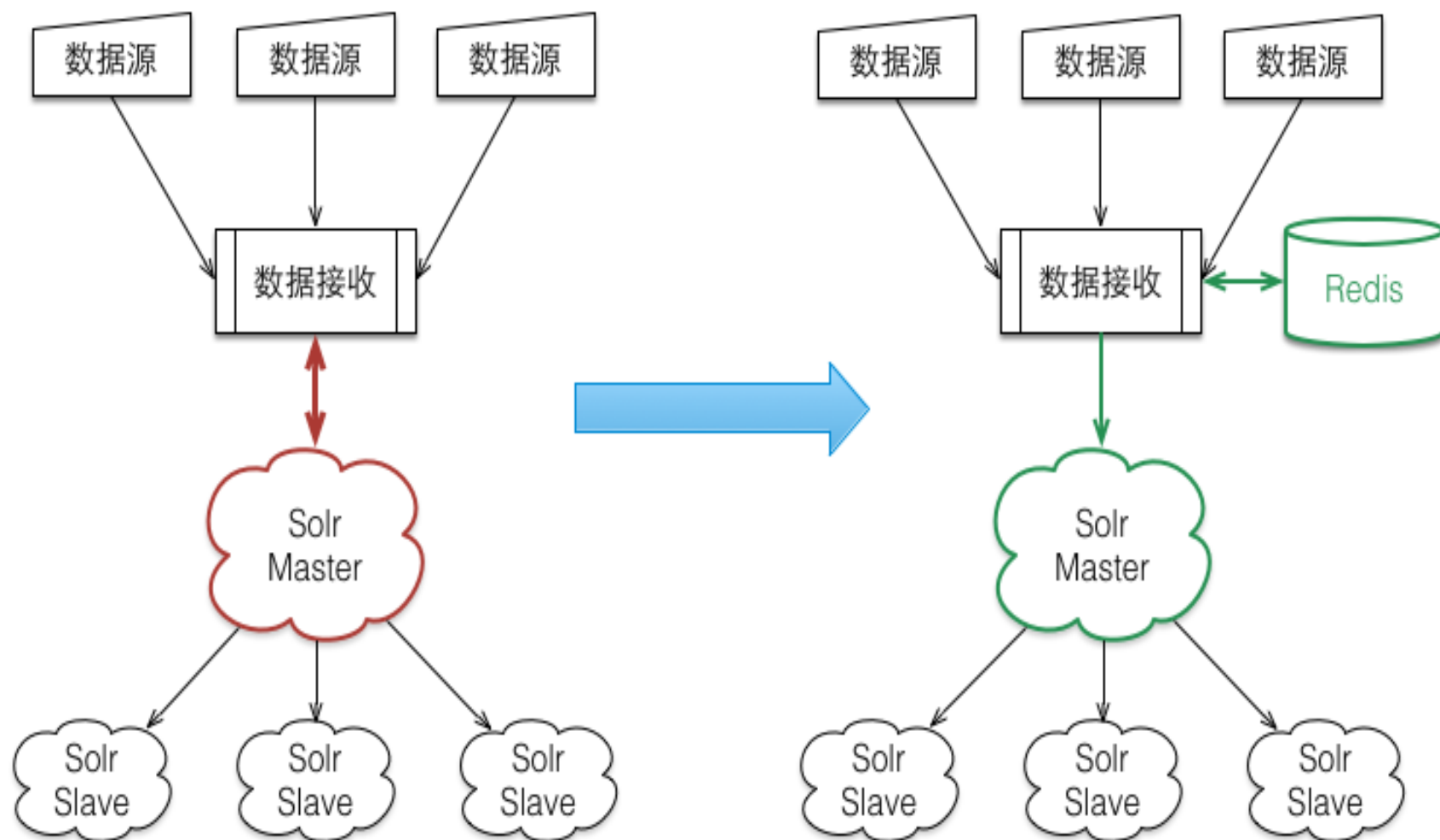
高频局部更新 – 现象

- 产品属性构成复杂
- 数据来源众多
- 大批量更新请求
- 生产使用低版本**Solr**方案

高频局部更新 – 方案

- 根源：低版本Solr(<4.0)不支持局部更新
- 原方案
 - 查询Solr获得需更新的产品的完整copy
 - 更新相应字段后重新提交新的copy
- 改进方案
 - 在高性能KV存储中保存产品信息
 - 产品信息局部更新时从KV中获得当前完整产品数据
 - 更新相应字段后提交给Solr并更新KV
- 优缺点
 - 优：每次更新减少一次Solr查询，大幅性能提升
 - 弱：非Solr原生支持，在KV失效后仍会对Solr产生压力

高频局部更新 - 示意图



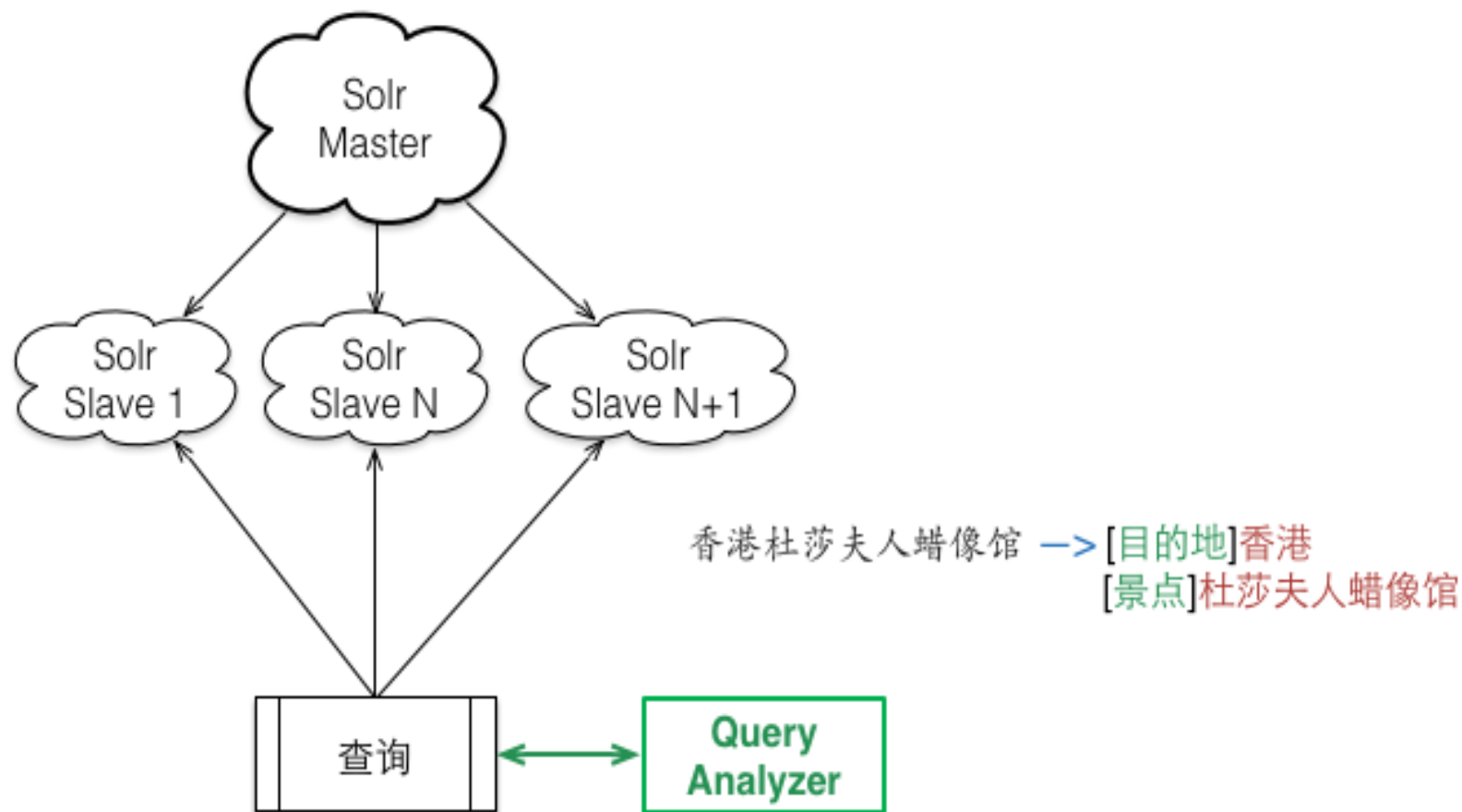
搜索前置解析- 现象

- 搜索框仅支持关键词查询
- 重度依赖分词和关键词匹配
- 相关性较差

搜索前置解析 – 方案

- 目的
 - “马尔代夫”仅呈现相关目的地产品
- 前置解析
 - 分词优化
 - 实体识别
 - 词性分类
- 香港杜莎夫人蜡像馆 -> [目的地]香港 + [景点]杜莎夫人蜡像馆

搜索前置解析 - 示意图

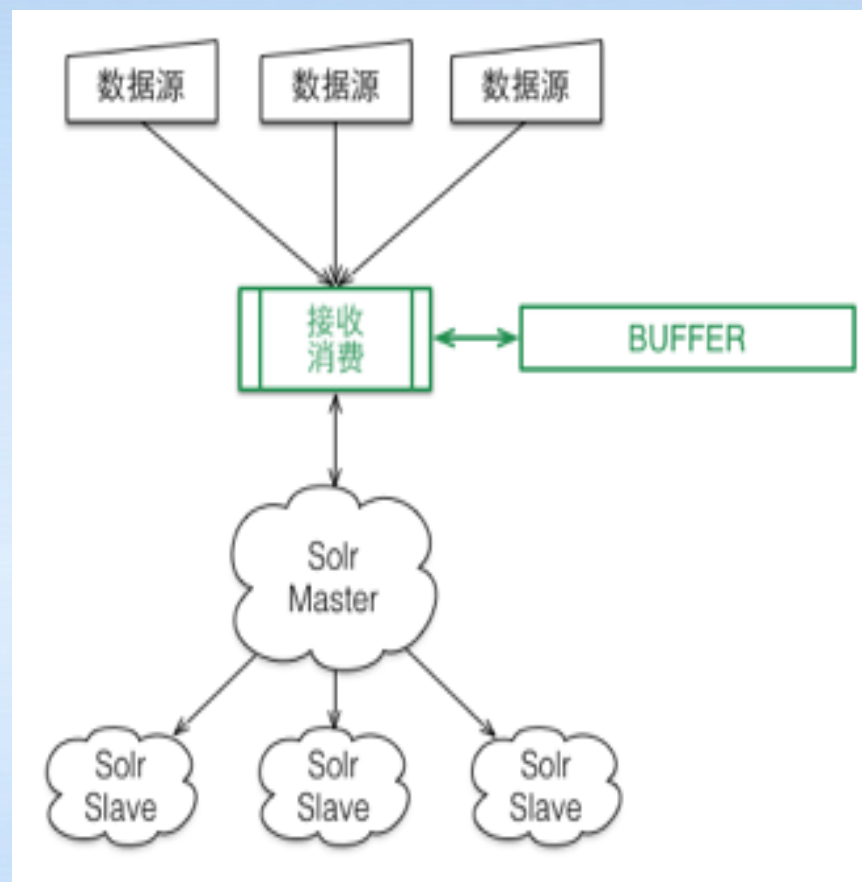


索引数据缓冲 — 现象

- 到达即消费，无记录
- 数据来源多
- 推送频率高
- 监控能力弱
- 问题定位难度大

索引数据缓冲 – 方案&示意图

- 支持**Failover**修复
- 方便定位
- 合并后更新



起价计算



**网上预订
专属优惠**

- 1. 早预订多人预订最高50元/每人
- 2. 旅游归来发表点评，每位成人获赠200元抵用券+5元现金
- 3. 旅游归来写游记，最高可获1000元！

促销价: **¥9926** 元起 ~~¥10076~~起 [起价说明](#)

满意度: 98% 已有19人出游 9人点评

参加活动: **网上预订抵 ¥100** **送 ¥200+5**
早鸟优惠

行程套餐: 6天 机+酒 从上海出发

包含服务: **早晚餐** **快艇**

产品经理推荐

适合: **美食**

★ 艾拉胡岛归属于Chaaya旗下，于2008年重新开幕

★ 可代为申请蜜月（需提供6个月内的结婚证明）

本起价是可选出发日期中，按双人出行共住一间房核算的最低单人价格。产品价格会根据您所选择的出发日期、出行人数、入住酒店房型、航班或交通以及所选附加服务的不同而有所差别。

1 选择出游人数 成人 儿童(2-12周岁) **总价 23493 元 人均 11746 元** **立刻预订**

2 选择出游日期

	日	一	二	三	四	五	六
	5-11	5-12	5-13	5-14	5-15	5-16 11746起	5-17

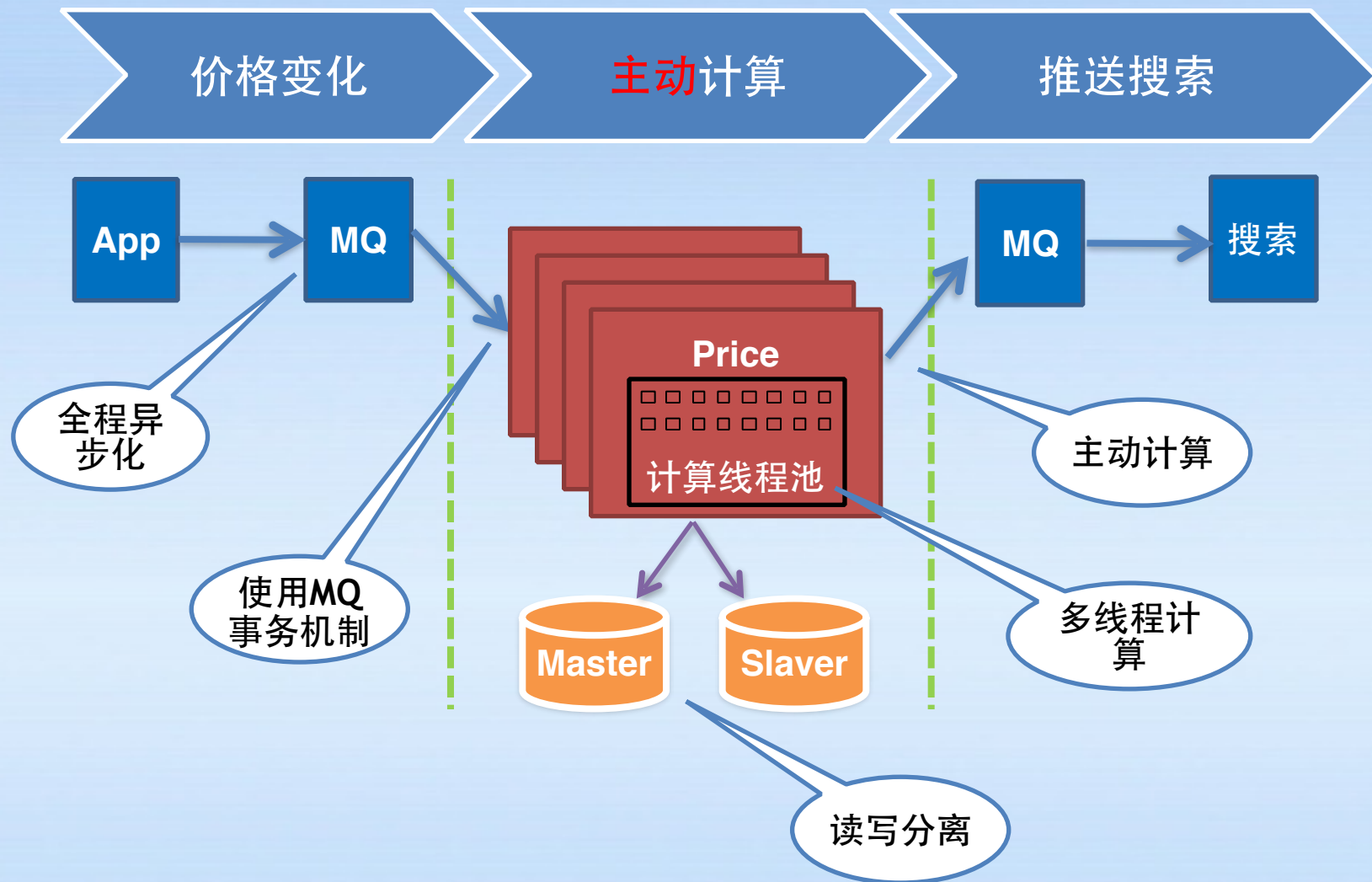
3 选择出游方案

✈ 交通

	日	一	二	三	四	五	六
	5-18 11746起	5-19 11646起	5-20	5-21 10596起	5-22 11746起	5-23 11746起	5-24 11946起
	5-25 11746起	5-26 11646起	5-27	5-28 10596起	5-29 11946起	5-30 11746起	5-31
2014-05-16 07:00	6-1 11176起	6-2 11176起	6-3	6-4 10076起	6-5 10076起	6-6 10076起	6-7 10076起
2014-05-20 18:50 05:40	6-8 11276起	6-9 11276起	6-10	6-11 10076起	6-12 10076起	6-13 10076起	6-14 10076起
仅显示当天可用机票，您	6-15 11276起	6-16 11276起	6-17	6-18 10076起	6-19 11276起	6-20 11276起	6-21

最低优惠价9926元 = 连牛价10076元 - 最大优惠金额50元 - 可使用抵用券金额100元
提前49天(含)以上完成签约和付清全款，每位成人优惠50元！

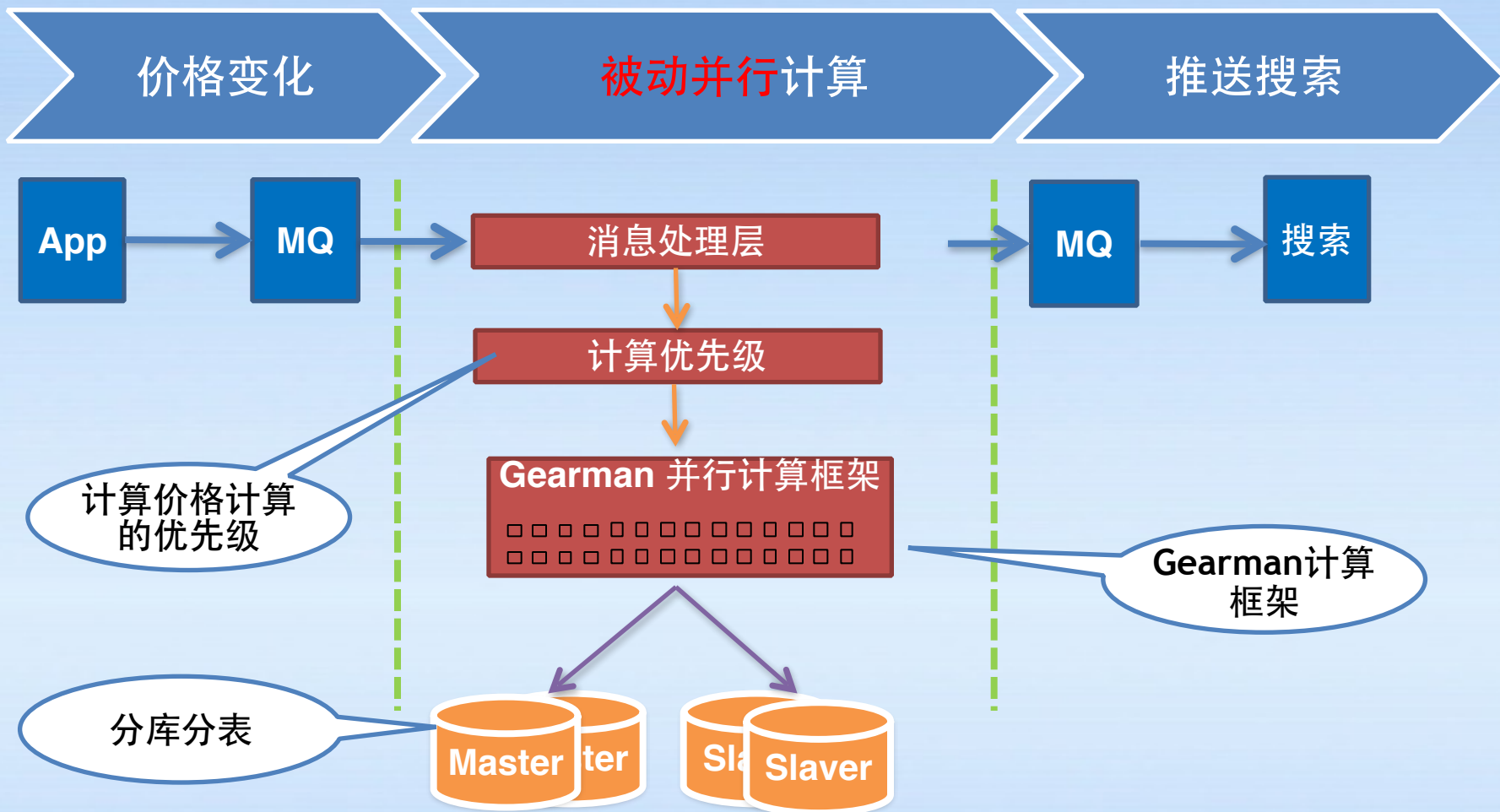
起价计算架构演进-V1



起价计算 - 我们需要更快响应

- 运维：你们的机器，经常**CPU**高负载！
- 业务：下周要引入新的业务，价格计算没有问题吧？
- 产品：调整价格都半天了，网站显示的还是老价格！
- **DBA**：你们单个表达到**2亿**级别了！

起价计算架构演进—V2



起价计算 - 经验教训

- 水平扩展：系统在设计阶段就要考虑如何扩展
- 容量规划：系统在设计阶段对自身服务能力的规划
- 监控报警：完善的监控报警机制，提前补漏
- 系统解耦：不要在高并发下压垮依赖系统

谢谢！