



# 2012云计算架构师峰会

Cloud Computing Architects Summit China 2012

揭示企业级IT架构转型 分享最新技术的应用落地

主办单位：**51CTO.com**  
技术成就梦想

联合主办：**UBM**

# 大数据产品架构设计

张茂森

Weibo:@maosen  
Email:xiaosen.zhang@gmail.com

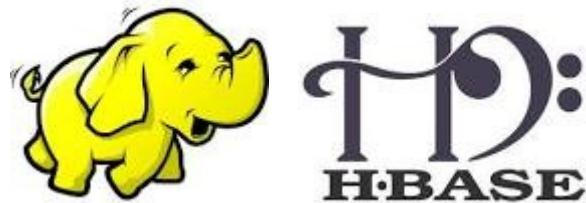
# Agenda

- 什么是大数据产品
- 数据产品的业务场景
- 技术架构
  - 高性能网站
  - 应用计算
  - 云计算
- 走过的一些弯路
- Q & A

# 大数据时代的新命题

# 大数据时代的新命题

- 大规模数据处理能力更易获得
  - 众多的海量数据处理工具、平台
  - 云计算服务

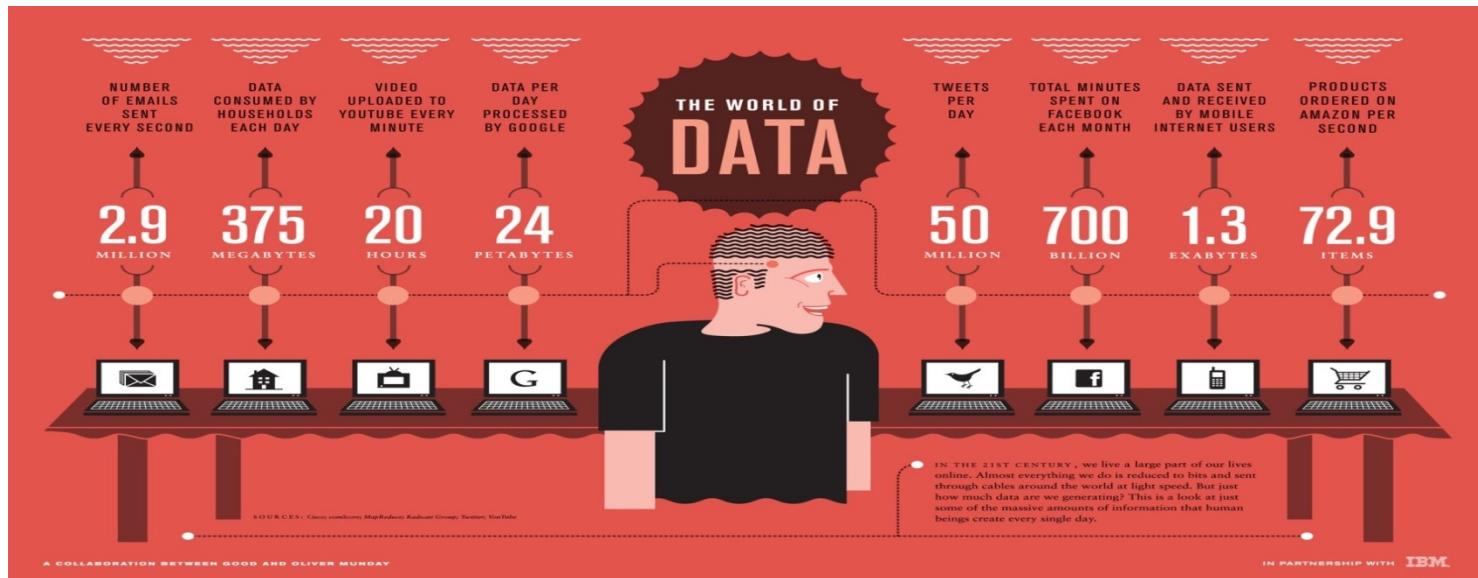


# 大数据时代的新命题

- 大规模数据处理能力更易获得
  - 众多的海量数据处理工具、平台
  - 云计算服务

# 大数据时代的新命题

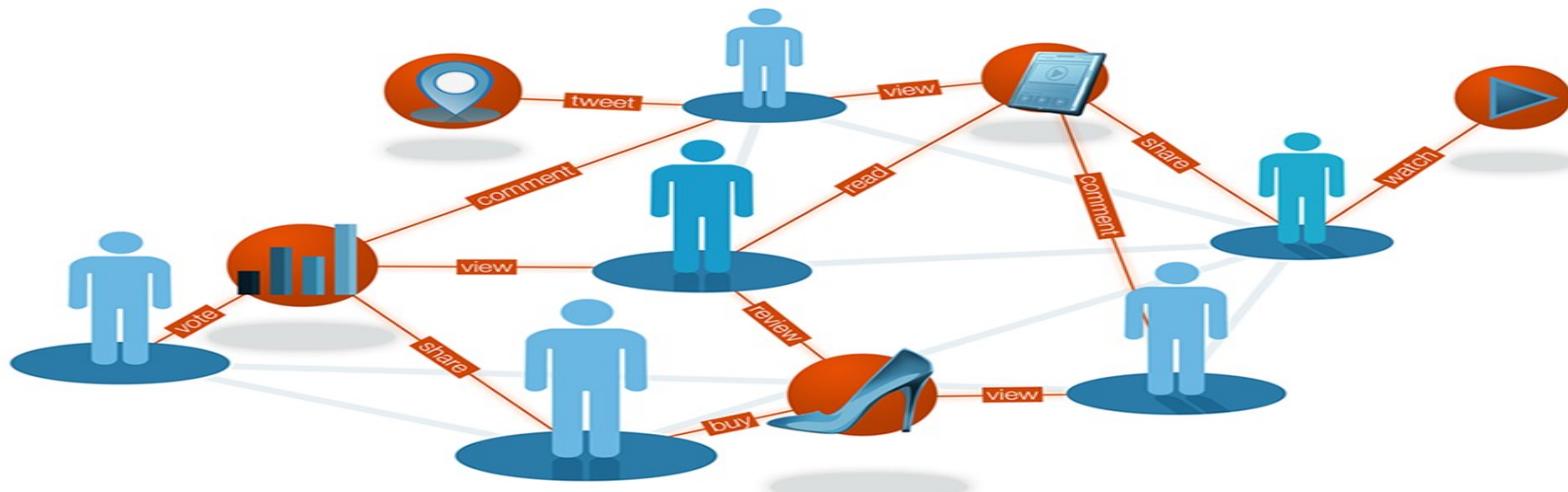
- 大规模数据处理能力更易获得
  - 众多的海量数据处理工具、平台
  - 云计算服务
- 用户数据在爆炸
  - 行为数据
  - UGC数据



# 大数据时代的新命题

- 大规模数据处理能力更易获得
  - 众多的海量数据处理工具、平台
  - 云计算服务
- 用户数据在爆炸
  - 行为数据
  - UGC数据

# 大数据时代的新命题



- 数据间的关联性前所未有的复杂
  - 社交网络
  - 购买关系
  - 各个业务系统IT化整合

# 大数据时代的新命题

- 大规模数据处理能力更易获得
  - 众多的海量数据处理工具、平台
  - 云计算服务
- 用户数据在爆炸
  - 行为数据
  - UGC数据
- 数据间的关联性前所未有复杂
  - 社交网络
  - 购买关系
  - 各个业务系统IT化整合



- 数字化营销大行其道
  - RTB+DSP+DMP
  - SEO+SEM

# 大数据时代的新命题

- 大规模数据处理能力更易获得
  - 众多的海量数据处理工具、平台
  - 云计算服务
- 用户数据在爆炸
  - 行为数据
  - UGC数据
- 数据间的关联性前所未有复杂
  - 社交网络
  - 购买关系
  - 各个业务系统IT化整合
- 数字化营销大行其道
  - RTB+DSP+DMP
  - SEO+SEM

# 大数据产品

- 以海量数据为基础
- 以数据计算为核心
- 对N个数据源进行整合
- 通过复杂关联获取价值，面临维度爆炸的挑战
- 通过分析界面、报表操作等完成人机交互
- 与其它在线生产系统进行数据对接

# 数据产品业务场景

- 产品场景
  - 百万级用户
  - 千万级页面访问
  - 亿级数据接口请求
  - PB级后台数据处理吞吐
- 软实时计算的需求
  - 数据从产生到完成结果计算延迟要求到秒级
- 数据任务依赖关系复杂

# 整体架构



用户端

分析类产品UX

报表类产品UX

推荐类产品UX

外部系统

数据展现计算

应用计算

数据中间层

实时效果分析平台

实时计算&查询引擎系统

历史数据查询&存储层

推荐&挖掘系统

云计算服务

流计算框架

即时计算框架

专用分布式计算平台

公用分布式计算平台

任务管理&调度系统

元数据分析&优化系统

数据IDE工具



实时数据收集总线

实时数据源

离线数据源

# 整体架构



用户端

分析类产品UX

报表类产品UX

推荐类产品UX

外部系统

Javascript新一代的编程语言

应用计算

数据中间层

实时效果分  
析平台

实时计算&查询引  
擎系统

历史数据查  
询&存储层

推荐&挖掘系  
统

云计算服务

流计  
算框  
架

即时  
计算框  
架

专用分布  
式计算平  
台

公用分布  
式计算  
平台

任务管理&调  
度系统

元数据分  
析&  
优化系  
统

数  
据  
IDE  
工  
具



实时数据收集总线

实时数据源

离线数据源

# 整体架构



用户端

分析类产品UX

报表类产品UX

推荐类产品UX

外部系统

Javascript新一代的编程语言

应用计算

数据API：基于非阻塞IO模型&轻量级线程式编程

实时效果分析平台

实时计算&查询引擎系统

历史数据查询&存储层

推荐&挖掘系统

云计算服务

流计算框架

即时计算框架

专用分布式计算平台

公用分布式计算平台

任务管理&调度系统

元数据分析&优化系统

数据IDE工具



实时数据收集总线

实时数据源

离线数据源

# 整体架构



用户端

分析类产品UX

报表类产品UX

推荐类产品UX

外部系统

Javascript新一代的编程语言

应用计算

数据API：基于非阻塞IO模型&轻量级线程式编程

实时效果分析平台

实时计算&查询引擎系统

Sharding or Cluster

推荐&挖掘系统

云计算服务

流计算框架

即时计算框架

专用分布式计算平台

公用分布式计算平台

任务管理&调度系统

元数据分析&优化系统

数据IDE工具



实时数据收集总线

实时数据源

离线数据源

# 整体架构



用户端

分析类产品UX

报表类产品UX

推荐类产品UX

外部系统

Javascript新一代的编程语言

应用计算

数据API：基于非阻塞IO模型&轻量级线程式编程

实时效果分析平台

实时计算&查询引擎系统

Sharding or Cluster

推荐&挖掘系统

云计算服务

流计算框架

即时计算框架

专用分布式计算平台

分布式计算&存储&K/V服务

任务管理&调度系统

元数据分析&优化系统

数据IDE工具



实时数据收集总线

实时数据源

离线数据源

# 整体架构



用户端

分析类产品UX

报表类产品UX

推荐类产品UX

外部系统

Javascript新一代的编程语言

应用计算

数据API：基于非阻塞IO模型&轻量级线程式编程

实时效果分析平台

实时计算&查询引擎系统

Sharding or Cluster

推荐&挖掘系统

云计算服务

流计算框架

即时计算框架

高优先级计算任务

分布式计算&存储&K/V服务

任务管理&调度系统

元数据分析&优化系统

数据IDE工具



实时数据收集总线

实时数据源

离线数据源

# 整体架构



用户端

分析类产品UX

报表类产品UX

推荐类产品UX

外部系统

Javascript新一代的编程语言

应用计算

数据API：基于非阻塞IO模型&轻量级线程式编程

实时效果分析平台

实时计算&查询引擎系统

Sharding or Cluster

推荐&挖掘系统

云计算服务

流计算框架

多维查询等

高优先级计算任务

分布式计算&存储&K/V服务

任务管理&调度系统

元数据分析&优化系统

数据IDE工具



实时数据收集总线

实时数据源

离线数据源

**网站**

# 如何做到高性能

# 如何做到高性能

- 数据产品的特性会放大“并发”对网站的压力

# 如何做到高性能

- 数据产品的特性会放大“并发”对网站的压力
  - 举个银行柜台的例子



# 如何做到高性能

- 数据产品的特性会放大“并发”对网站的压力
  - 举个银行柜台的例子
- 解决方案

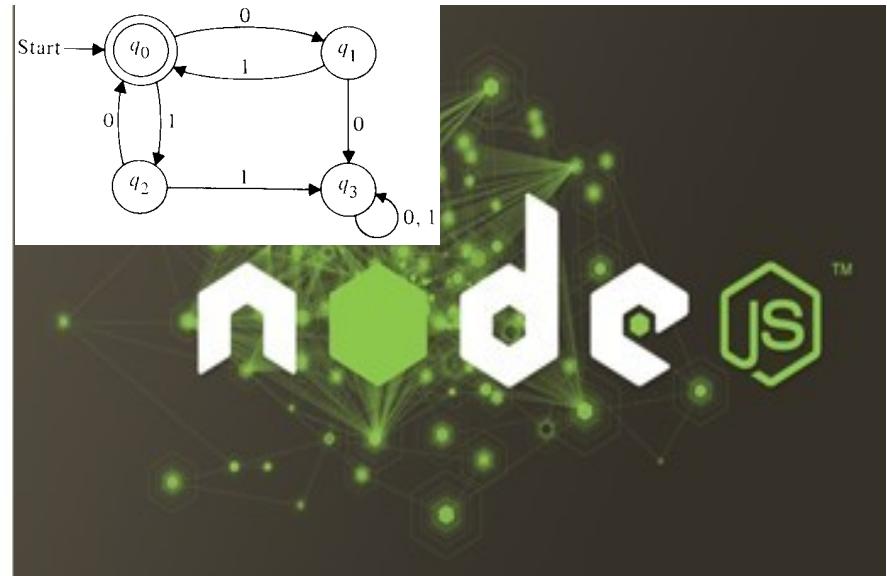
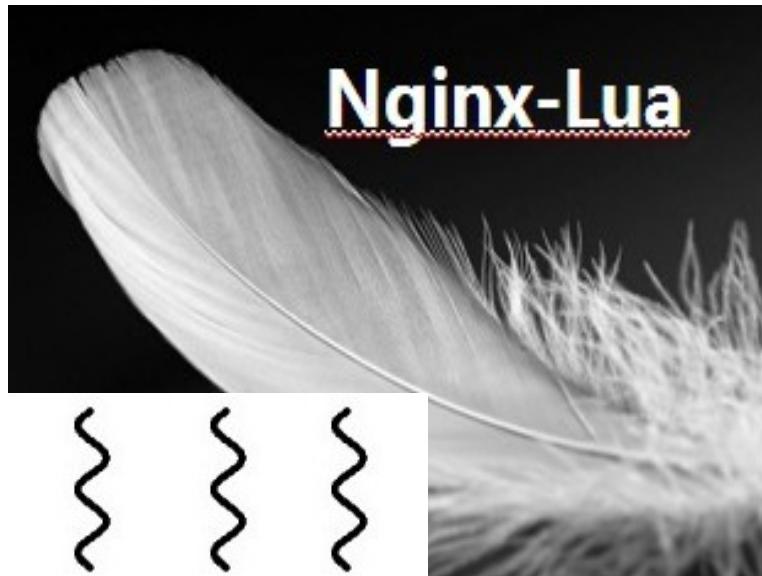
# 如何做到高性能

- 数据产品的特性会放大“并发”对网站的压力
  - 举个银行柜台的例子
- 解决方案
  - 将“柜台”和“后台”解耦
    - 用数据中间层将前后台各司其职
    - 用Javascript让“柜台”更强大

# 如何做到高性能

- 数据产品的特性会放大“并发”对网站的压力
  - 举个银行柜台的例子
- 解决方案
  - 将“柜台”和“后台”解耦
    - 用数据中间层将前后台各司其职
    - 用Javascript让“柜台”更强大
  - 使用更好的“排队”模型
    - 异步或者非阻塞IO模型
    - 轻量级多线程 or 事件驱动

# 如何做到高性能



- 使用更好的“排队”模型
  - 异步或者非阻塞I/O模型
  - 轻量级多线程 or 事件驱动

# 如何做到高性能

- 数据产品的特性会放大“并发”对网站的压力
  - 举个银行柜台的例子
- 解决方案
  - 将“柜台”和“后台”解耦
    - 用数据中间层将前后台各司其职
    - 用Javascript让“柜台”更强大
  - 使用更好的“排队”模型
    - 异步或者非阻塞IO模型
    - 轻量级多线程 or 事件驱动

# 如何做到高性能

- 数据产品的特性会放大“并发”对网站的压力
  - 举个银行柜台的例子
- 解决方案
  - 将“柜台”和“后台”解耦
    - 用数据中间层将前后台各司其职
    - 用Javascript让“柜台”更强大
  - 使用更好的“排队”模型
    - 异步或者非阻塞I/O模型
    - 轻量级多线程 or 事件驱动
  - 让“后台”更强大
    - 基于sharding的数据库集群
    - 支持Like-SQL的分布式的NoSQL数据库

# 应用计算

# 绕开CAP理论

# 绕开CAP理论

- 分布式系统的CAP理论
  - Consistency
  - Availability
  - Partition tolerance

# 绕开CAP理论

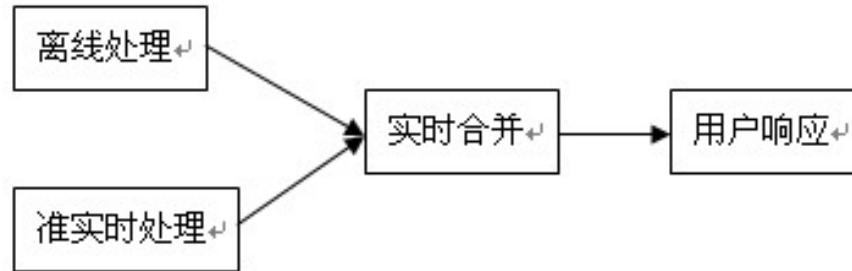
- 分布式系统的CAP理论
  - Consistency
  - Availability
  - Partition tolerance
- 重新审视数据
  - 数据是不可变的，变化的是事件

# 绕开CAP理论

- 分布式系统的CAP理论
  - Consistency
  - Availability
  - Partition tolerance
- 重新审视数据
  - 数据是不可变的，变化的是事件
- 分析业务，将系统拆分

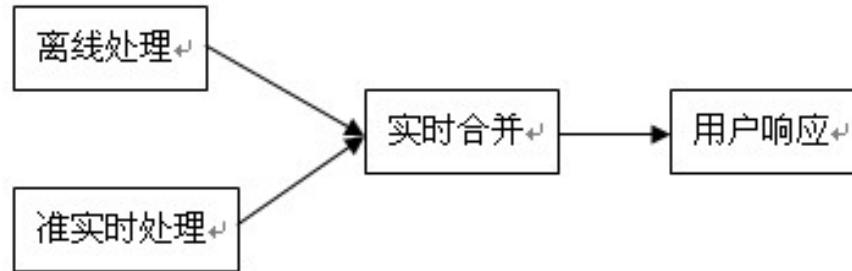
# 绕开CAP理论

- 分布式系统的CAP理论
  - Consistency
  - Availability
  - Partition tolerance
- 重新审视数据
  - 数据是不可变的，变化的是事件
- 分析业务，将系统拆分



# 绕开CAP理论

- 分布式系统的CAP理论
  - Consistency
  - Availability
  - Partition tolerance
- 重新审视数据
  - 数据是不可变的，变化的是事件
- 分析业务，将系统拆分



- 分别降低两个系统对于CAP的需求：
  - 离线计算，批数据因为没有变更，所以不存在一致性的问题
  - 准实时计算，可以容忍一致性，通过合并能够做到最终一致

# 实时计算

# 实时计算

- 流式计算
  - 实质上是增量计算
  - 弱状态保存需求
    - 异常检测
    - 攻击分析
  - 强状态保存需求
    - 数据统计、效果跟踪
    - 分析挖掘

# 实时计算

- 流式计算
  - 实质上是增量计算
  - 弱状态保存需求
    - 异常检测
    - 攻击分析
  - 强状态保存需求
    - 数据统计、效果跟踪
    - 分析挖掘
- 即时计算
  - 实质上是计算能力换时间
  - 索引和存储结构需重新设计
  - 并发响应有较大挑战
  - 多维分析场景

# 云计算服务

# 云计算服务

- 分布式处理开源工具
  - Hadoop HBASE Cassandra MongoDB Storm Scribe .....

# 云计算服务

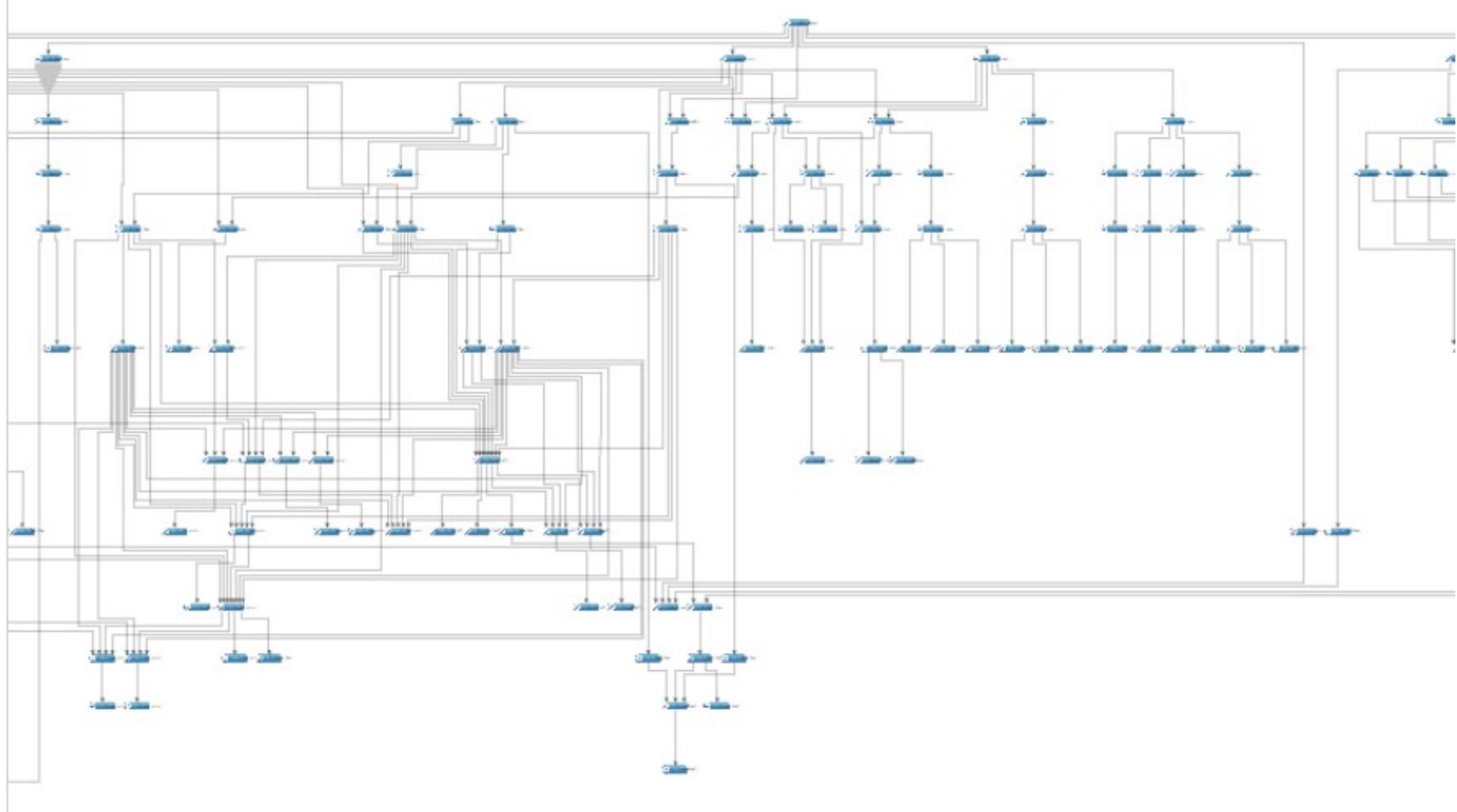
- 分布式处理开源工具
  - Hadoop HBASE Cassandra MongoDB Storm Scribe .....
- 更棘手的反而是

# 云计算服务

- 分布式处理开源工具
  - Hadoop HBASE Cassandra MongoDB Storm Scribe .....
- 更棘手的反而是
  - 任务调度
  - 元数据分析应用

# 任务调度

# 任务调度



# 任务调度

## Workflow/Data Pipelines 管理

- Flow Control
- Time trigger
- Data trigger

## 业务运维

- 智能重跑
- 补历史数据
- 查看日志/源代码
- 日志/代码搜索
- 业务分组

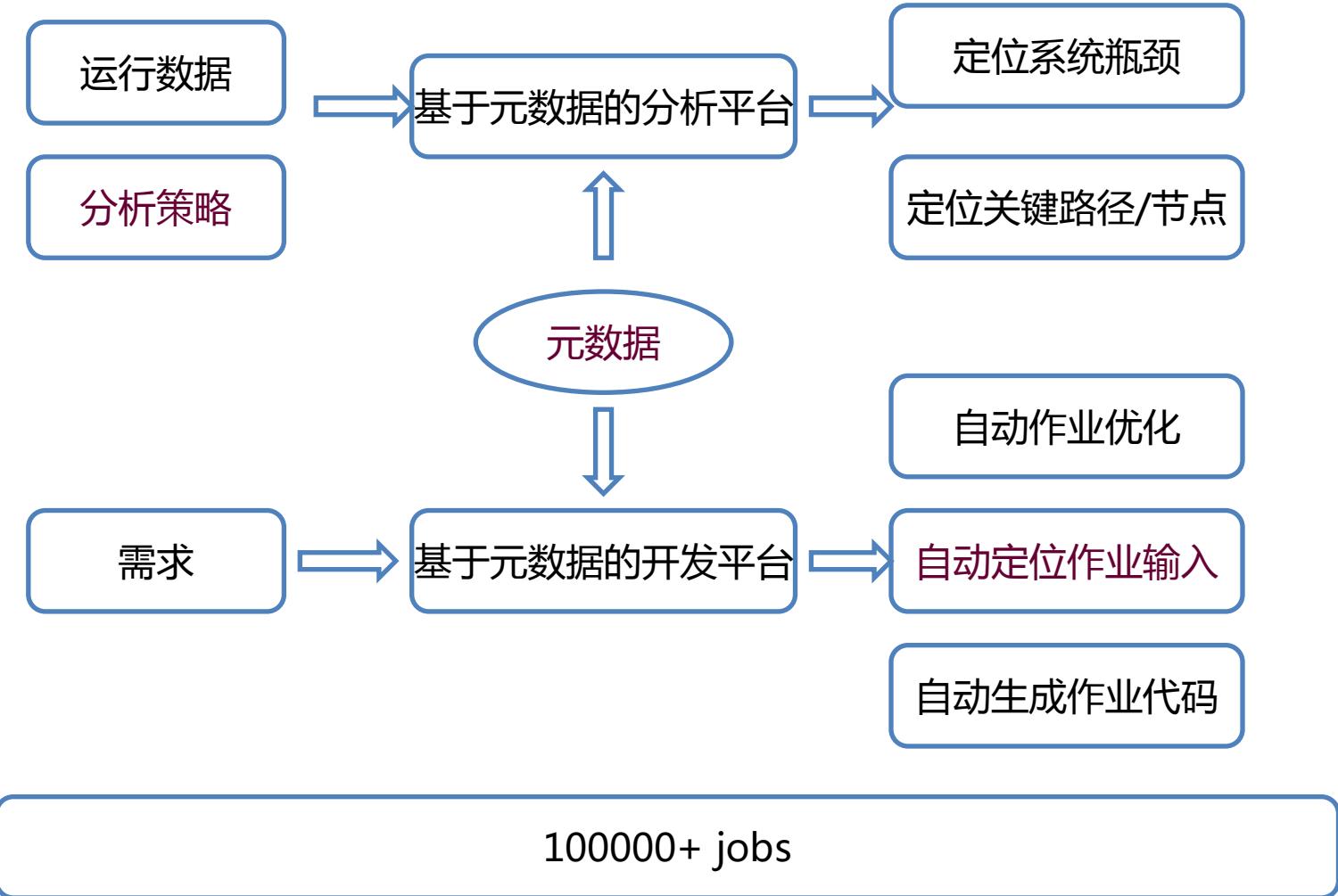
## 智能调度

- 异构系统支持
- 优先级设置
- Dynamic LB/Failover
- 自动重跑
- 资源分组

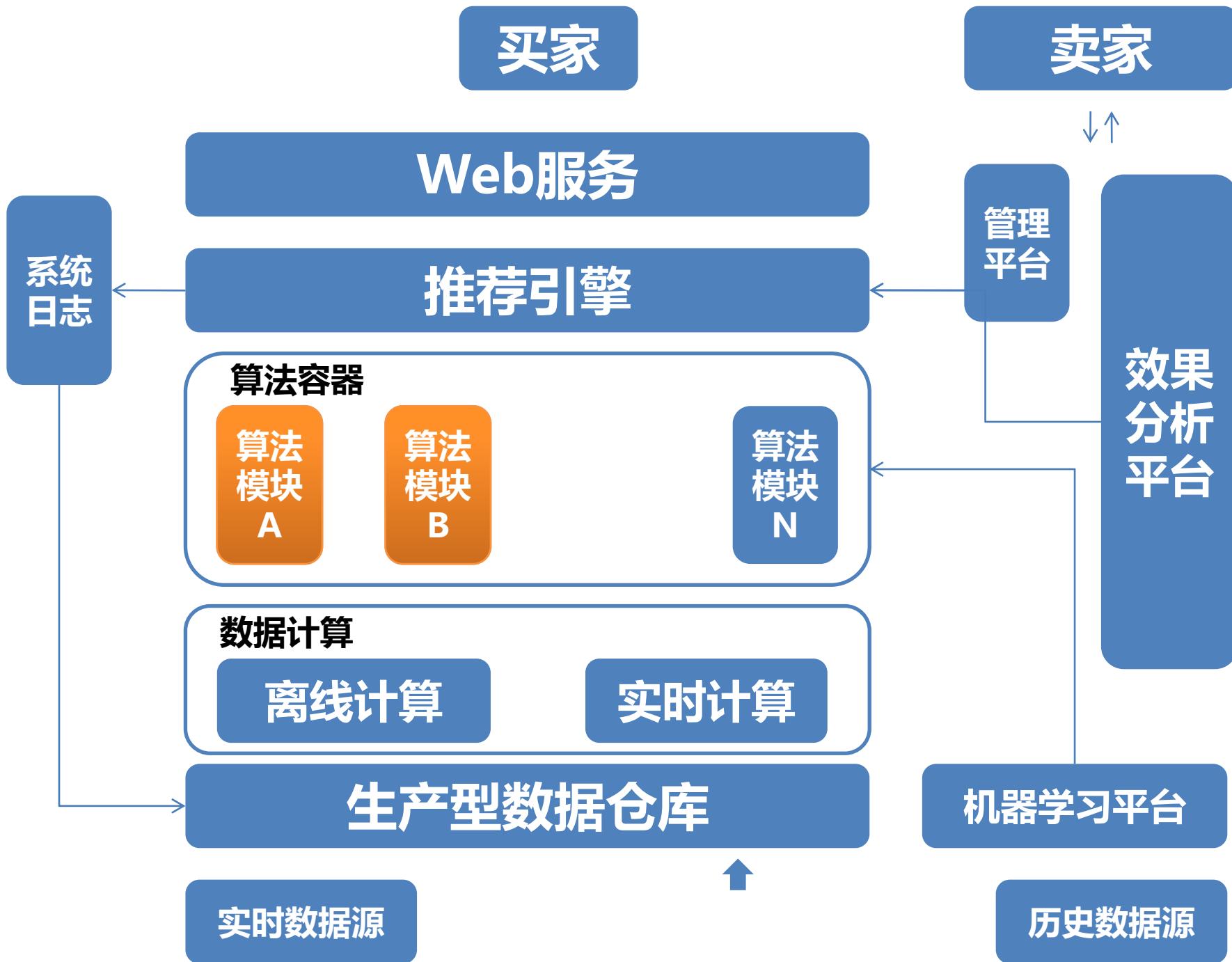
## 监控/告警/分析

- 运行信息收集
- 系统瓶颈分析
- 关键路径分析

# 发掘元数据价值



# 挖掘与推荐



# 山路十八弯

- 不重视系统运维、没有专职的运维人员
- 忽略部分数据标准
- 编程模式的改变带来的苦痛
- 调度系统环境分离
- 数据上下游协同
- 过分强调cluster，但是很多情况下  
Sharding是更简单的选择

# Q & A

Thanks !