



数据仓库技术在大众点评网的实践和案例分享

互联网数据仓库之路

Who am I

- 闫剑锋
 - 数据管理行业老兵，15年
 - 互联网行业新兵，5年
 - 执着于数据行业，并相信：
 - One size does NOT fit all (M. Stonebraker)
 - CLOUD is just automation (James Hamilton)
 - 18016286939
 - QQ: 1449867374
 - a_zha@126.com



Agenda

- 引题
- 互联网+数据仓库
- 点评网的数据平台
- 案例
 - 平台中的取舍
 - 建模中的平衡
 - 数据产品化
- 数据中心的未来

数据技术的历史反思

OLAP和BI
技术

OLTP技术



1980-1990

1990-2000

大数据技术的状态

OLAP和BI
技术

OLTP技术

Big Data
Warehouse



2000-2010

2010-2020



互联网+数据仓库

互联网数据挑战 数据仓库

Brought by **InfoQ**

水平扩展

A large flock of white sheep with black faces and legs is standing in rows on a green grassy field under a blue sky. One sheep in the foreground is highlighted with a white rectangular box containing the text.

数据规范化

用户触达

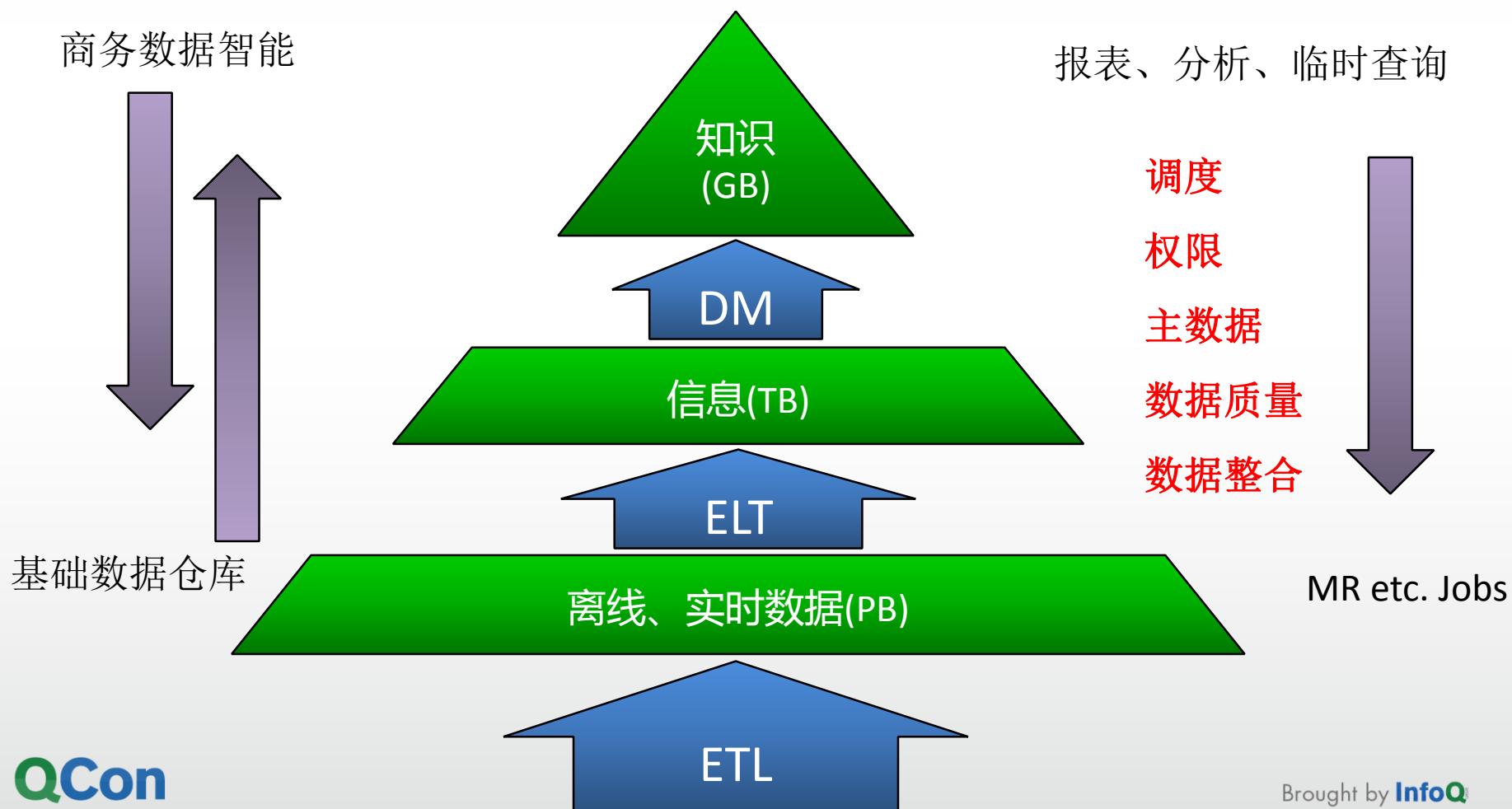


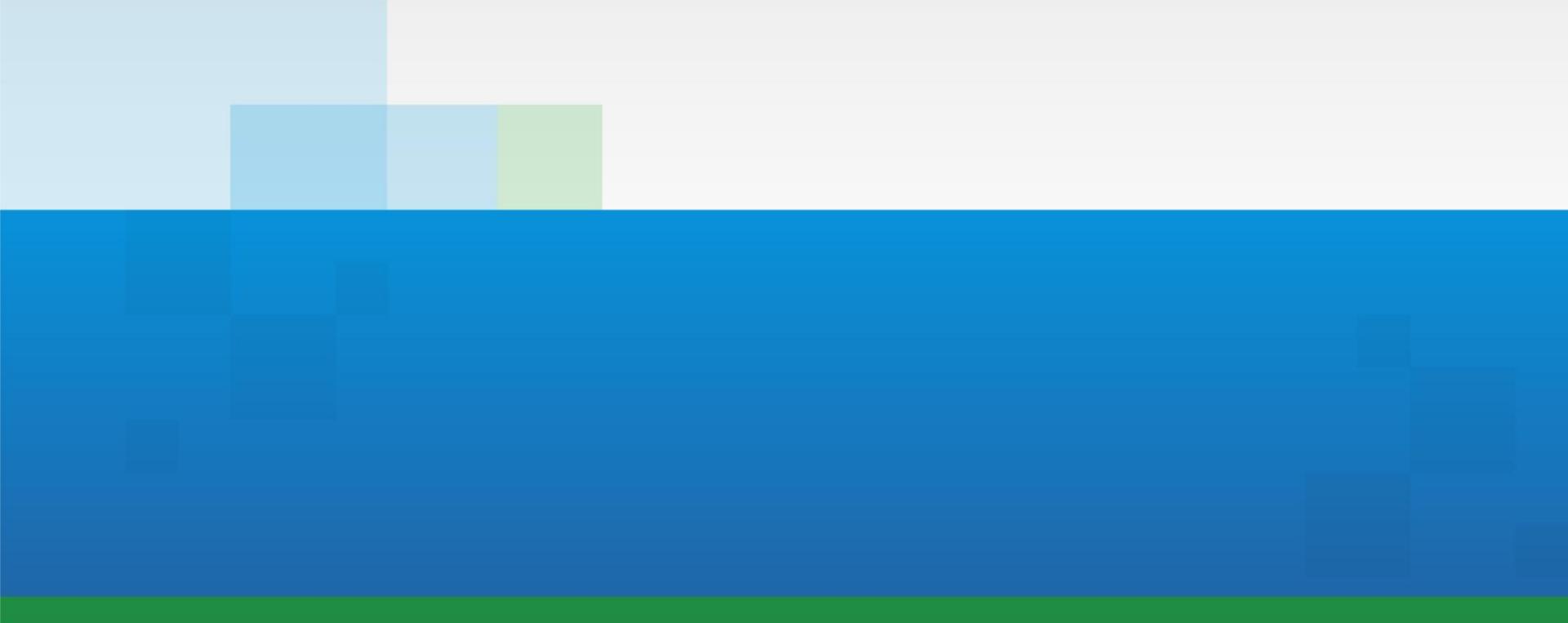
数据可闭环

马太效应



数据仓库





点评网的数据平台 平台的发展 数据用户群 体系结构

Brought by **InfoQ**

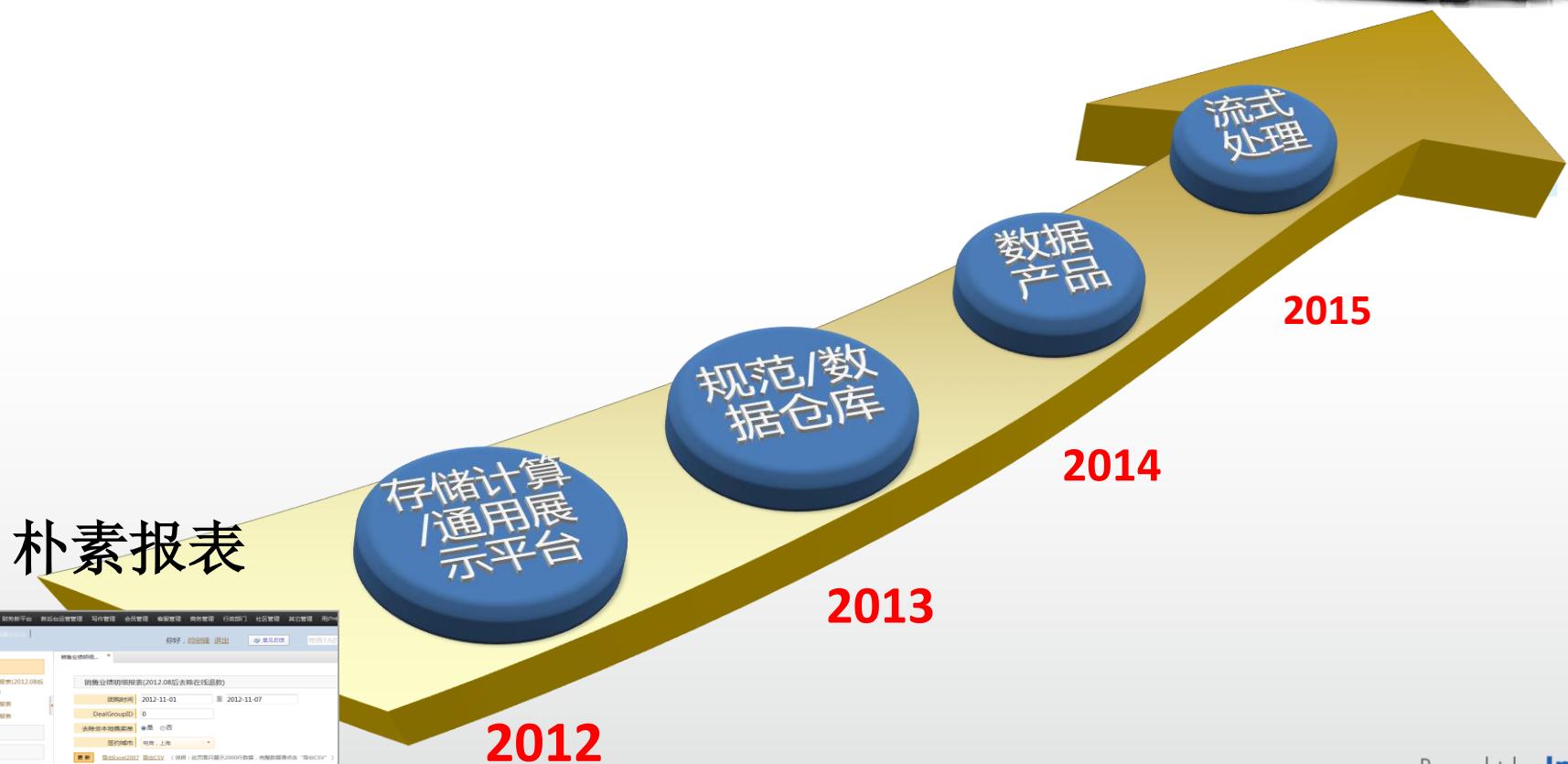
2012年的点评数据

技术部

~30张正式报表
各种临时抓

产品运营部
销售部

平台的发展



数据用户群

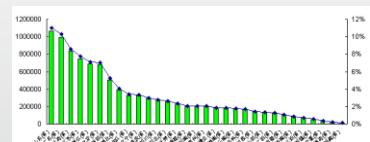
线上服务HA
编程接口
开发平台

线上反馈应用



数据仓库模型
数据质量
性能/稳定性

报告/OLAP类



QCon

数据平台

产品仓库
架构

销售市场

产品运营

产品经理

技术部门

Ad hoc工具
响应性能优化
指标监控体系



Ad-hoc查询

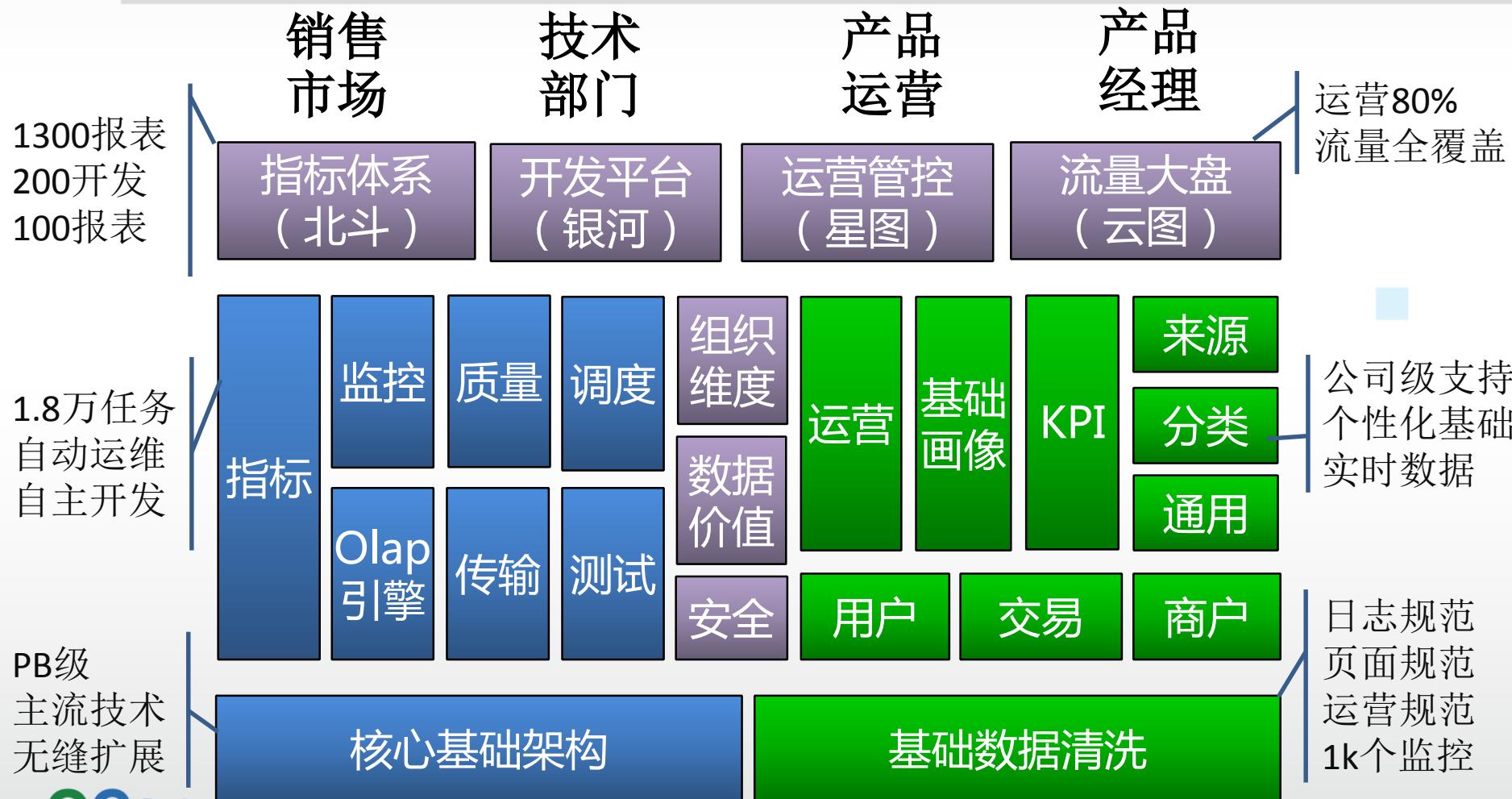
指标数据
自助指标引擎
流式计算

Dashboard类



Brought by InfoQ

体系架构





平台中的取舍 实时计算 vs 离线统计

Brought by **InfoQ**

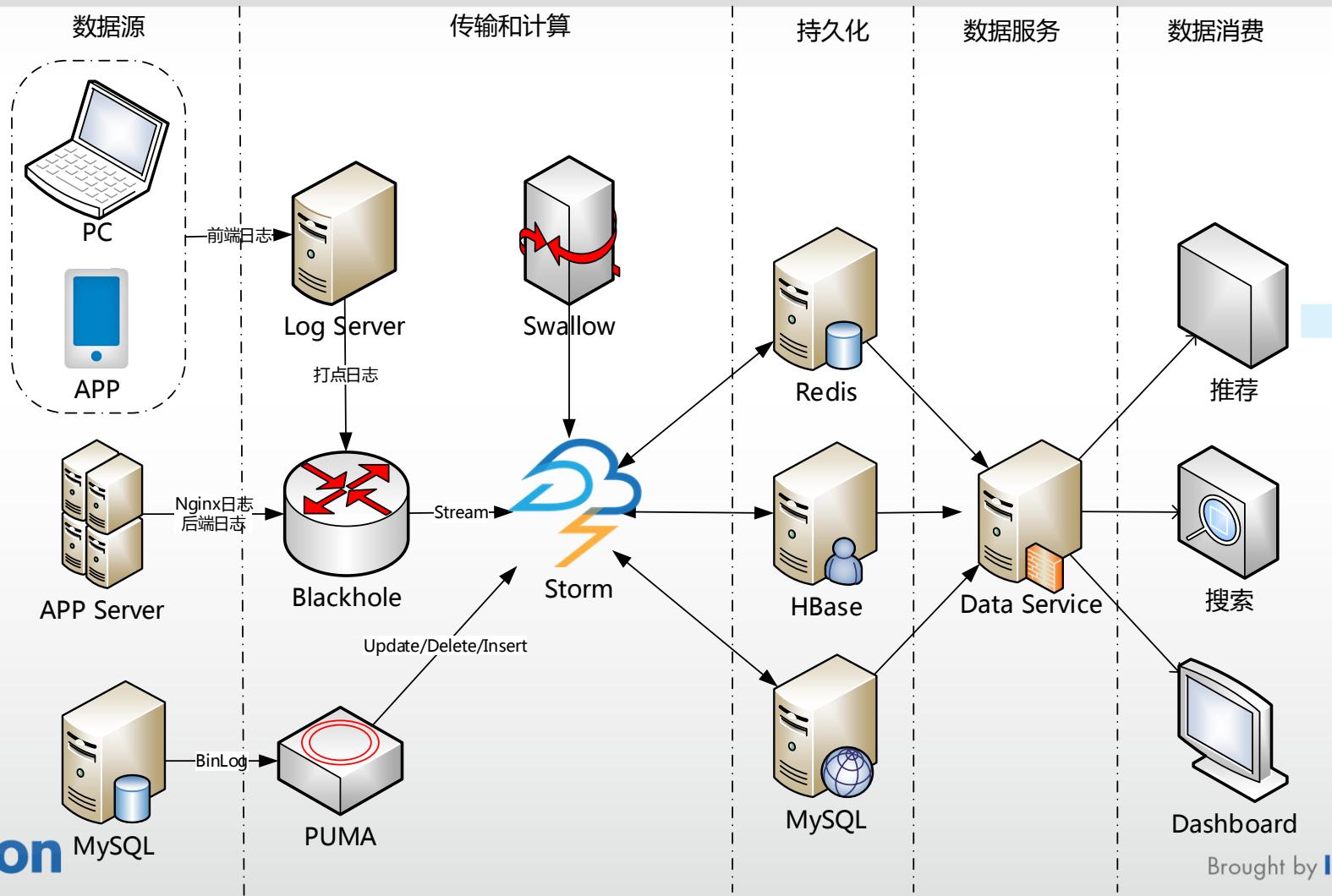
实时计算 VS 离线统计

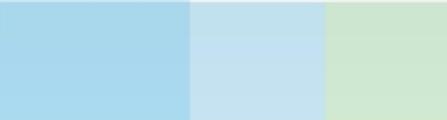
	实时计算	离线统计
时效性	强(毫秒到秒)	弱 (分钟)
处理方式	数据流 (单条/Micro Batch)	Batch
一致性	较弱	强
数据量	GB~TB	TB~PB
业务价值	10%~20%	80%~100%

CAP

CAP

实时计算架构





建模中的平衡

业务变化 **vs** 基础稳定
维度多样 **vs** 口径一致
运营精细 **vs** 指标全面

业务变化 VS 基础稳定

- 线上（自底向上）
 - 案例：流量数据体系
 - 策略侵入少，如流量、用户行为、支付等
- 线下（自顶向下）
 - 案例：销售数据体系
 - 策略侵入多、业务变化快，如销售、运营、市场等
- 线上线下串联（找到解耦点）
 - 案例：产品健康体系
 - 接耦点在流量和订单的挂接（下单？支付？退单？）

维度多样 VS 口径一致

- “挟天子以令诸侯”
 - 案例：核心数据体系（GMV）
 - 调研并统一业务口径，取得CFO级别支持
- 中路突破 + 政委下乡
 - 案例：商户集市与商家平台BU数据体系建设
 - 早期业务部门数据建设阶段
- 信念中心 + 落地
 - 案例：流量日志规范与解析
 - 部门自身影响力较大，上下游关系打通顺畅
- 数据产品化/服务化
 - 案例：dpid与用户集市
 - 部门自身开发能力，有合适的产品/服务切入点

运营精细 VS 指标全面

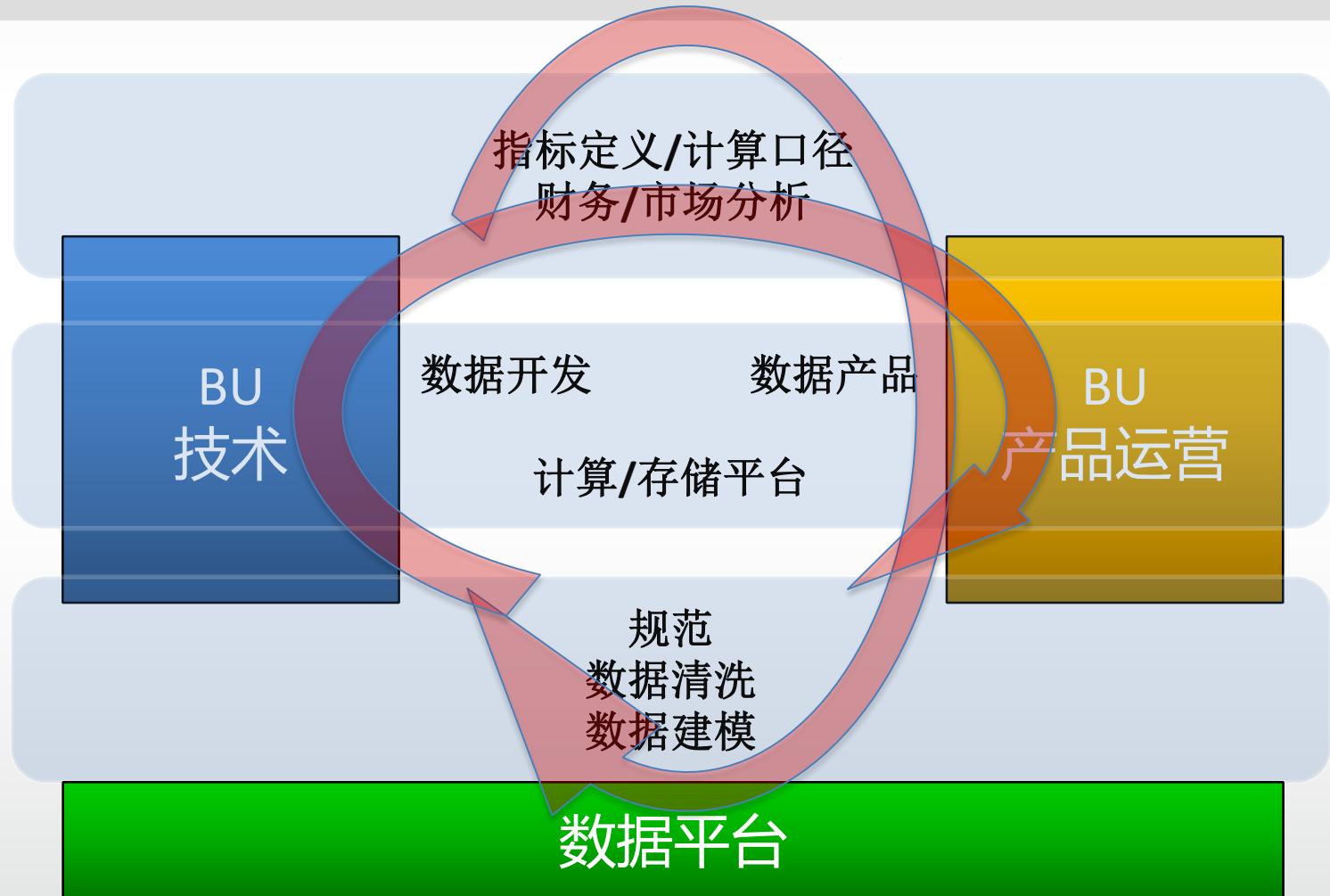
- 主旨
 - 数据开发相对滞后，业务真正成熟前很难预测
 - 业务先行，数据跟上，体系最后
- 婴儿月嫂模式：数据端到端产出由平台方全部承包
 - 案例：财务分析数据支持
- 小学家教模式：平台方与业务方合作，手把手培养业务
 - 案例：交易平台BU数据建设
- 中学放手模式：开放自助平台，具体开发无需平台方
 - 案例：BU业务报表/指标
- 大学教授模式：帮助制定整体框架，建设由业务负责
 - 案例：推广BU数据建设



数据产品化之路

业务闭环 vs 数据闭环 内部系统 vs 用户体验

业务闭环 vs 数据闭环

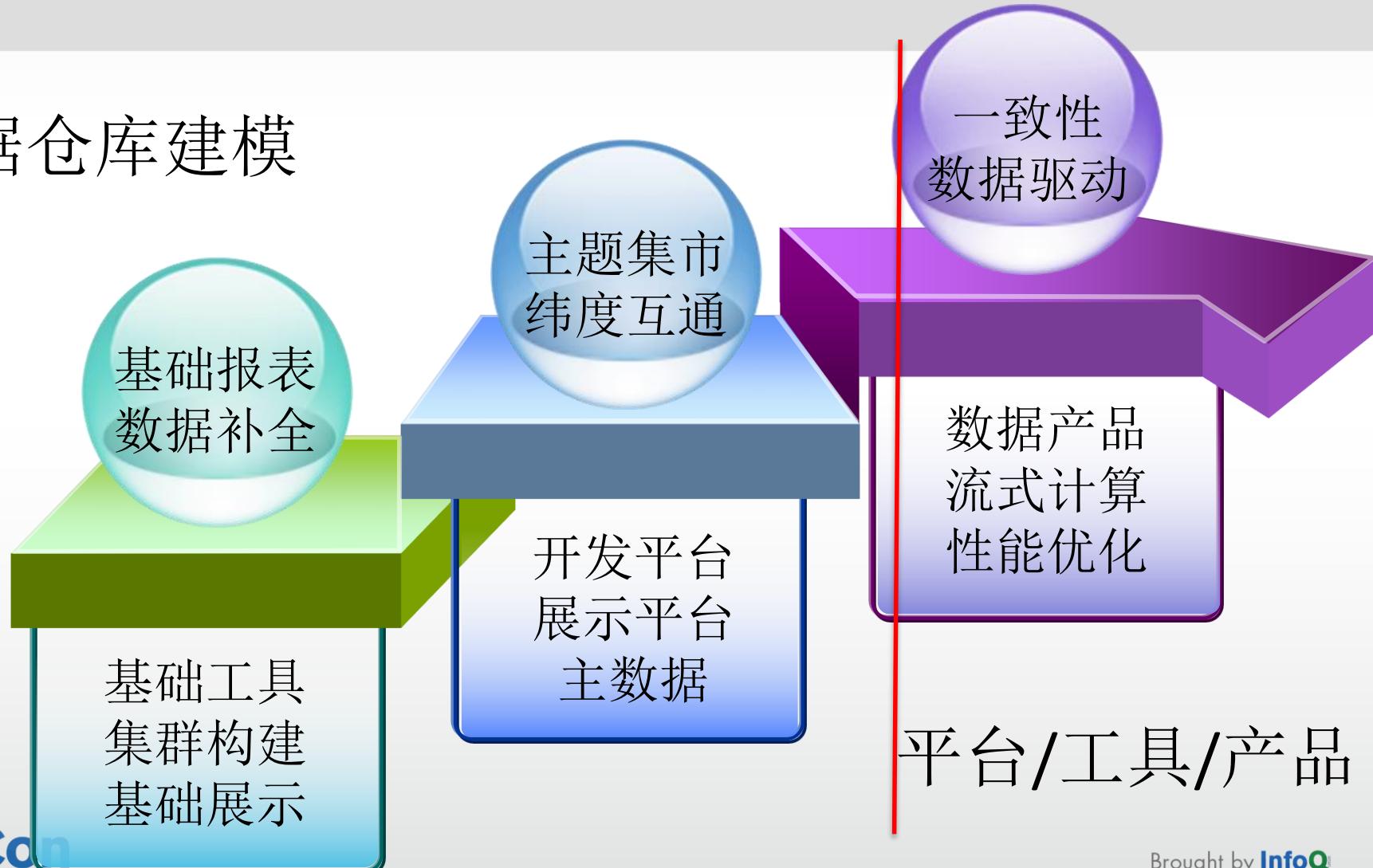


内部系统 VS 用户体验

- 运营活动平台
 - 团单、资源位、电影、红包
 - 分析和报表（Push分析）
- 数据部门的产品
 - Frame方式嵌入（报表平台）
 - Data Service复用（用户画像）
 - 数据模型的建立（留存分析）
 - 数据源标准的定义（运营效果拆分）

数据中心的未来

数据仓库建模



Q&A

谢谢大家！



@InfoQ



infoqchina