



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

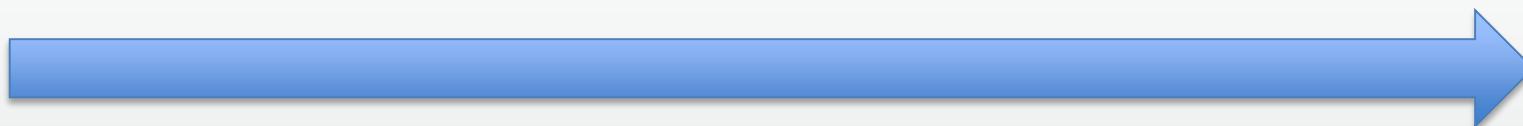
- 引题
- 互联网+数据仓库
- 点评网的数据平台
- 案例
 - 平台中的取舍
 - 建模中的平衡
 - 数据产品化
- 数据中心的未来

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

Big Data
Warehouse

OLTP技术



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, each with a unique face, standing in neat rows on a green field. The sheep are facing forward, and their shadows are cast on the grass. The image is used to illustrate the concept of horizontal scaling, where data is distributed across multiple servers or nodes.

数据规范化

用户触达

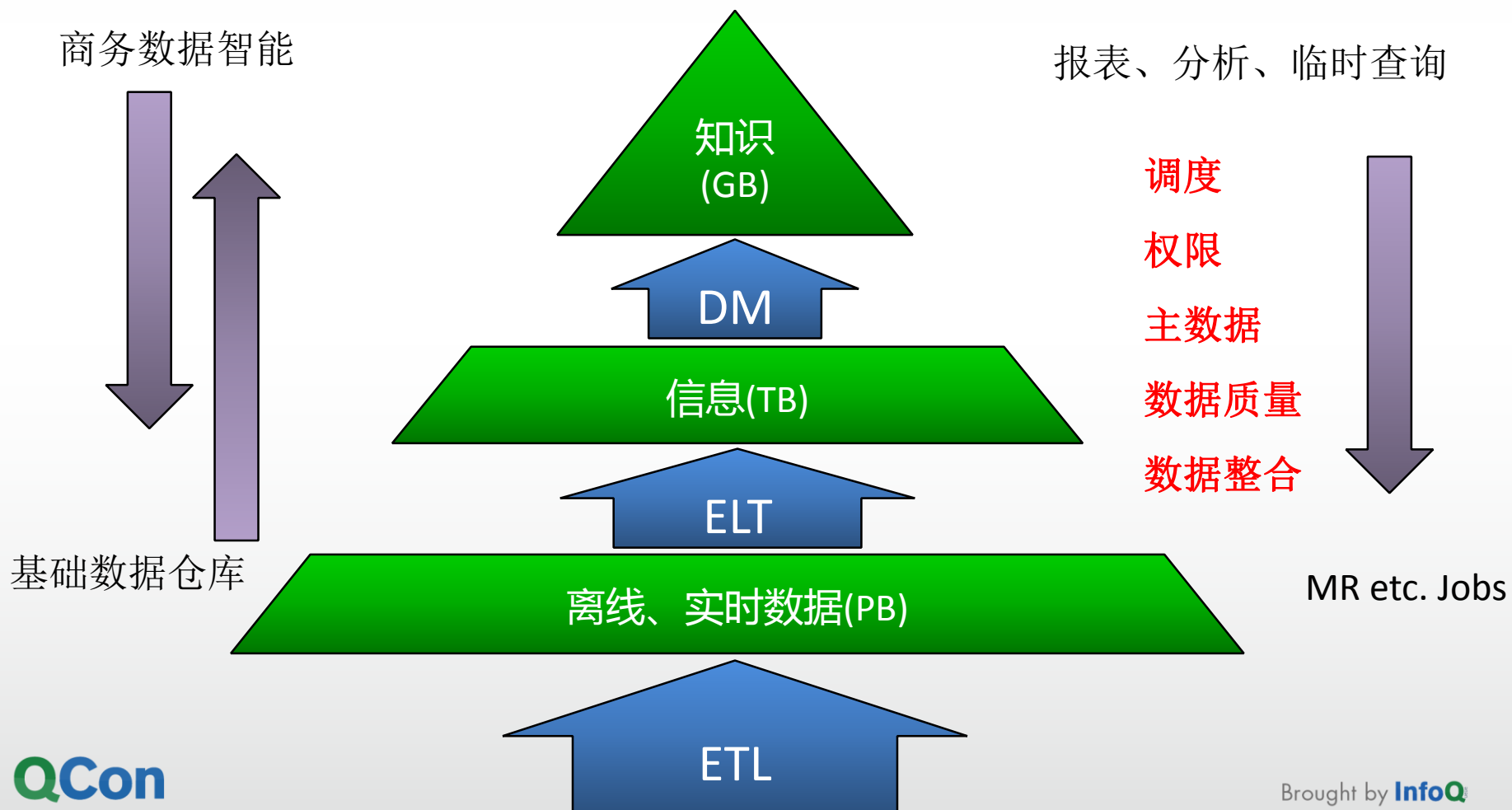


数据可闭环

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



平台的发展

数据中心



朴素报表

销售员	销售员ID	销售员姓名	销售员电话	销售员地址	销售员备注
销售员1	5744	销售员1	销售员1	销售员1	销售员1
销售员2	7863	销售员2	销售员2	销售员2	销售员2
销售员3	11498	销售员3	销售员3	销售员3	销售员3

存储计算
/通用展
示平台

2012

规范/数
据仓库

2013

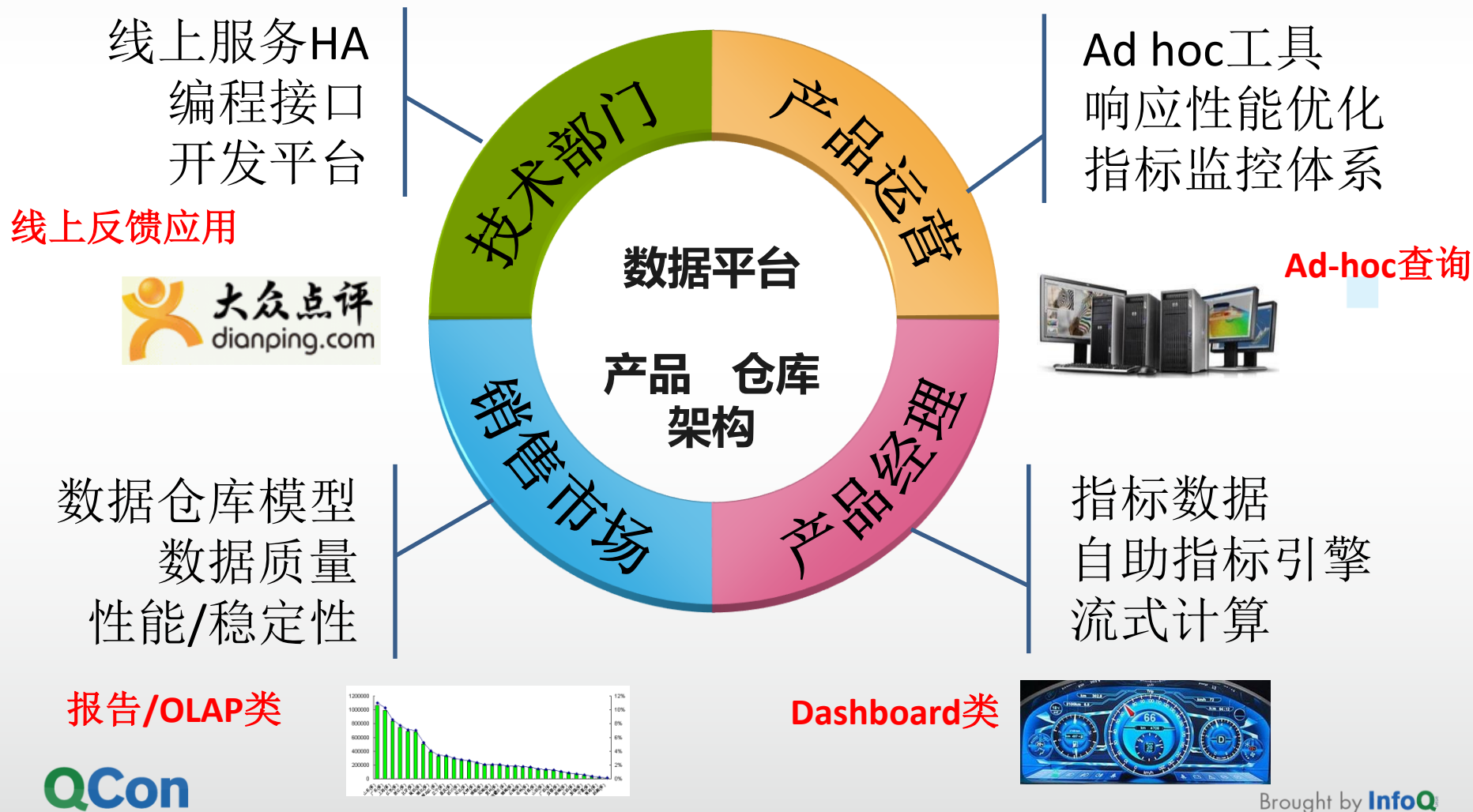
数据
产品

2014

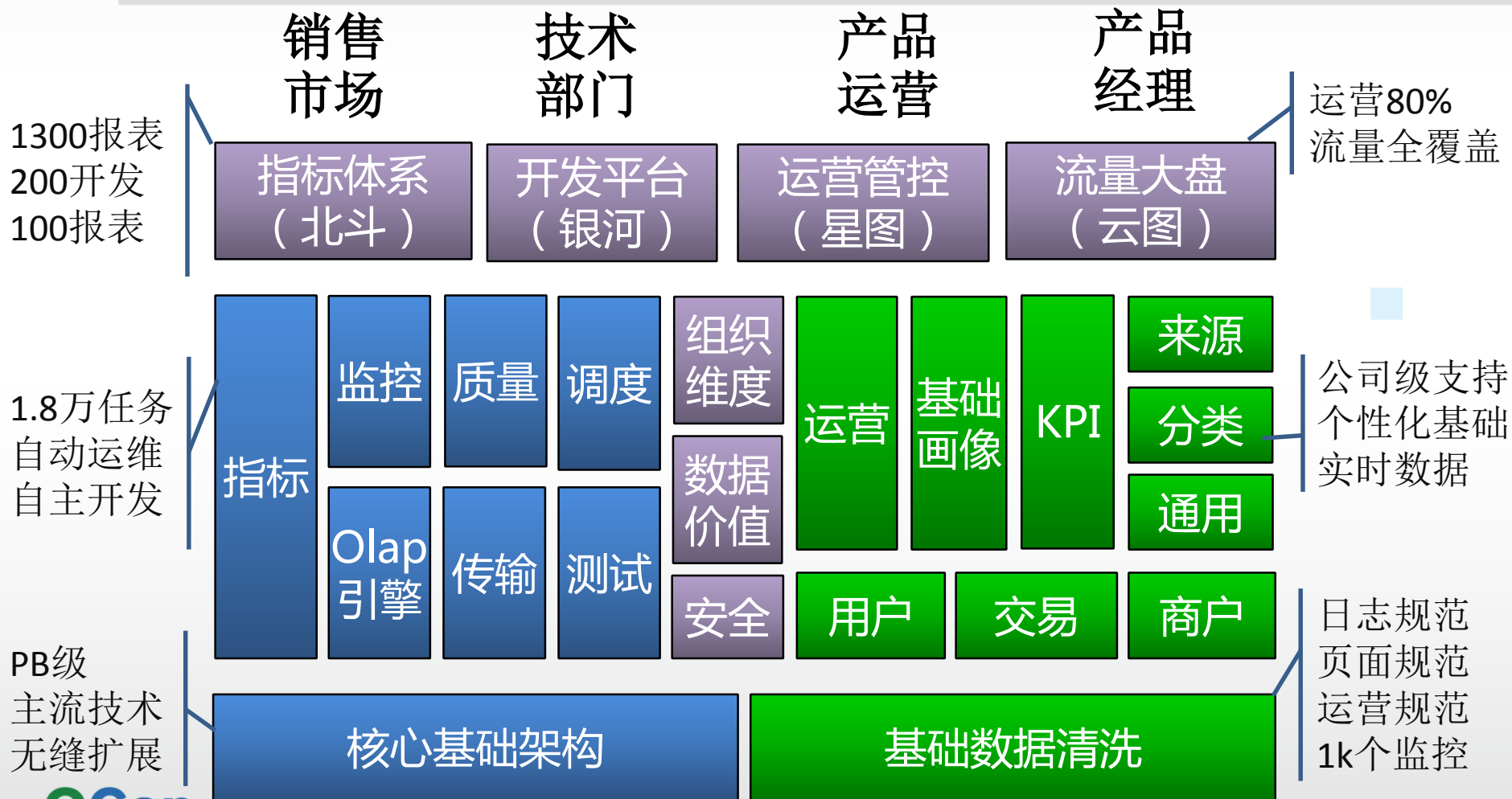
流式
处理

2015

数据用户群



体系架构





平台中的取舍

实时计算 VS 离线统计

Brought by **InfoQ**

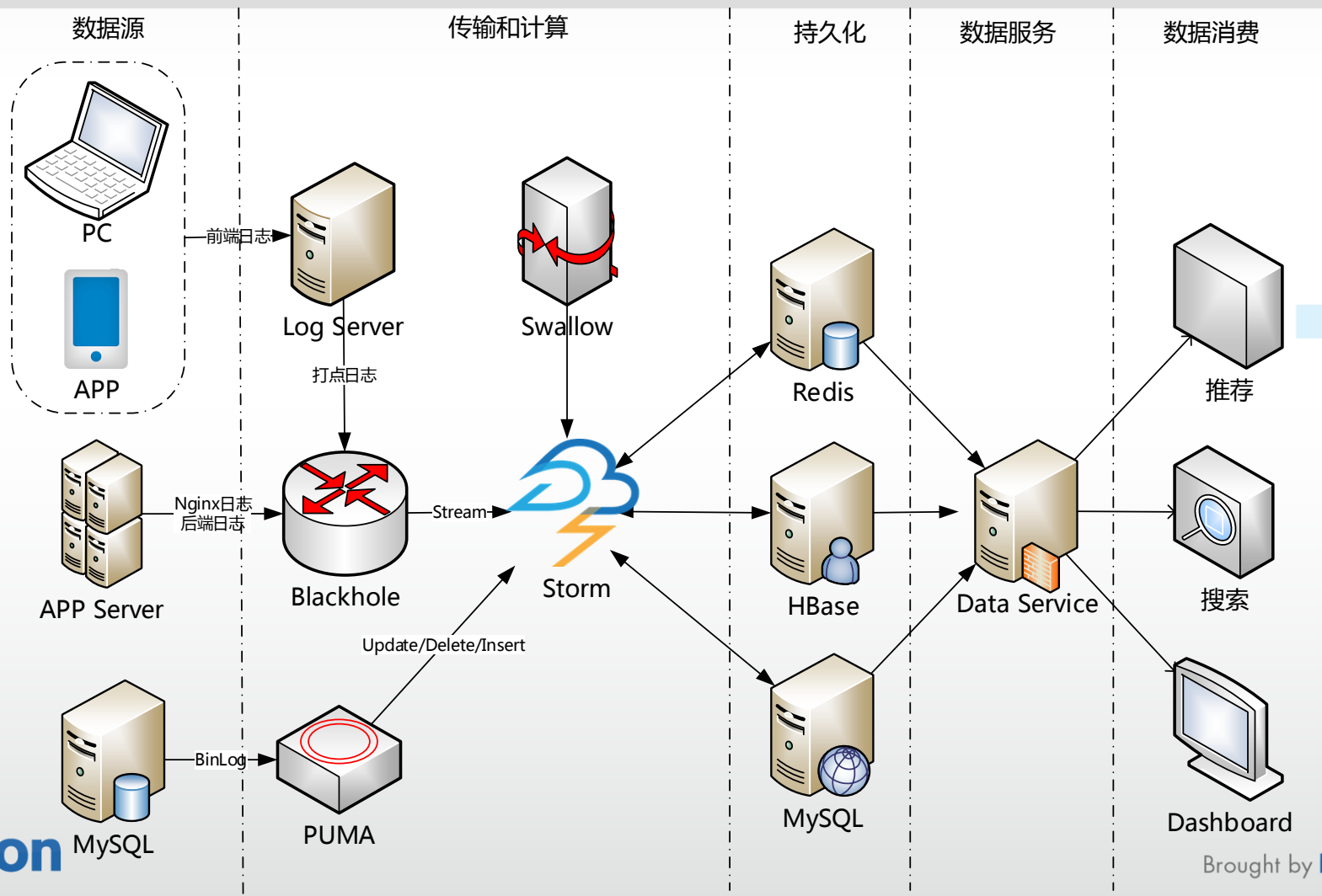
实时计算 VS 离线统计

	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

Brought by **InfoQ**

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

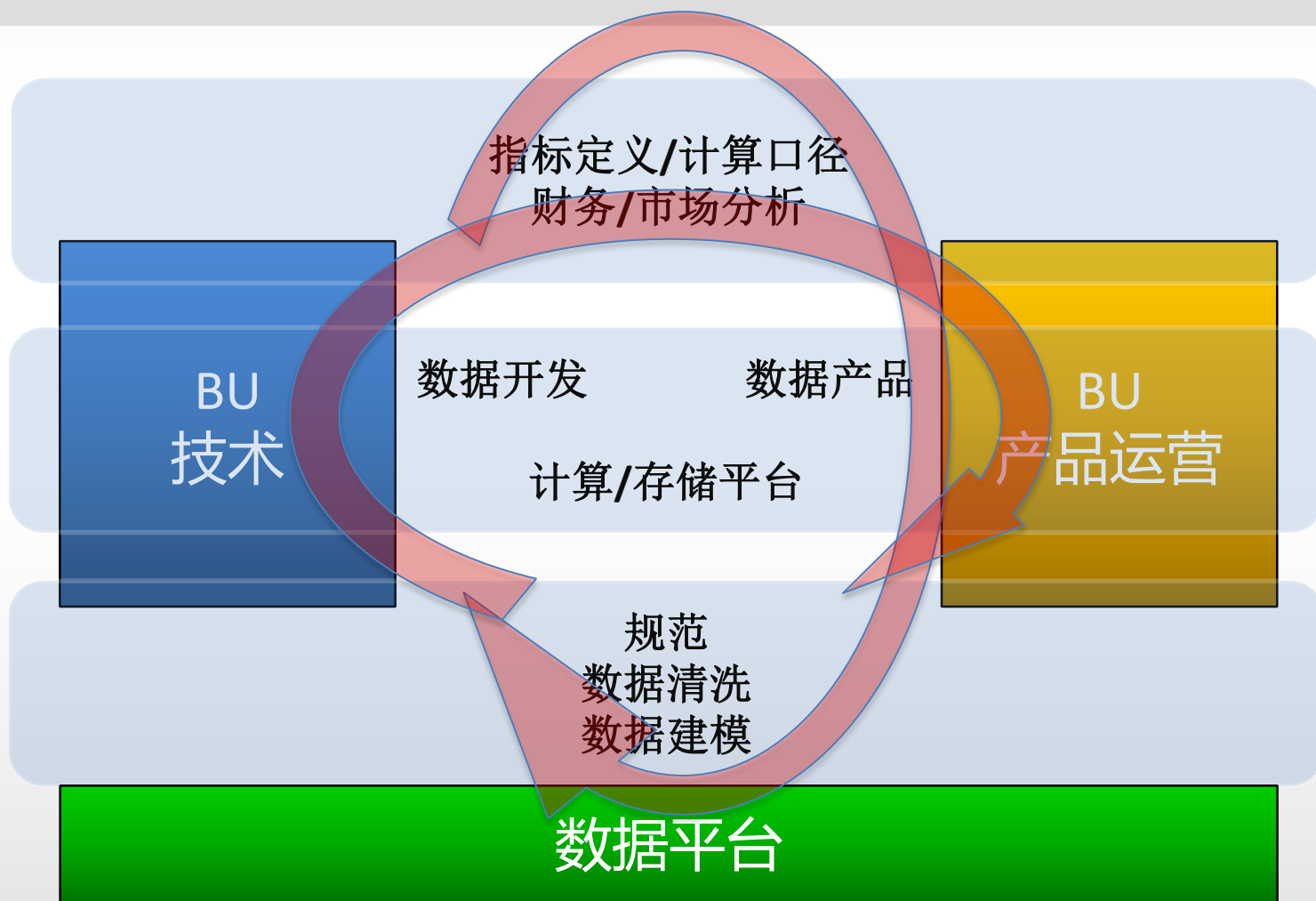


数据产品化之路

业务闭环 VS 数据闭环 内部系统 VS 用户体验

Brought by **InfoQ**

业务闭环 VS 数据闭环

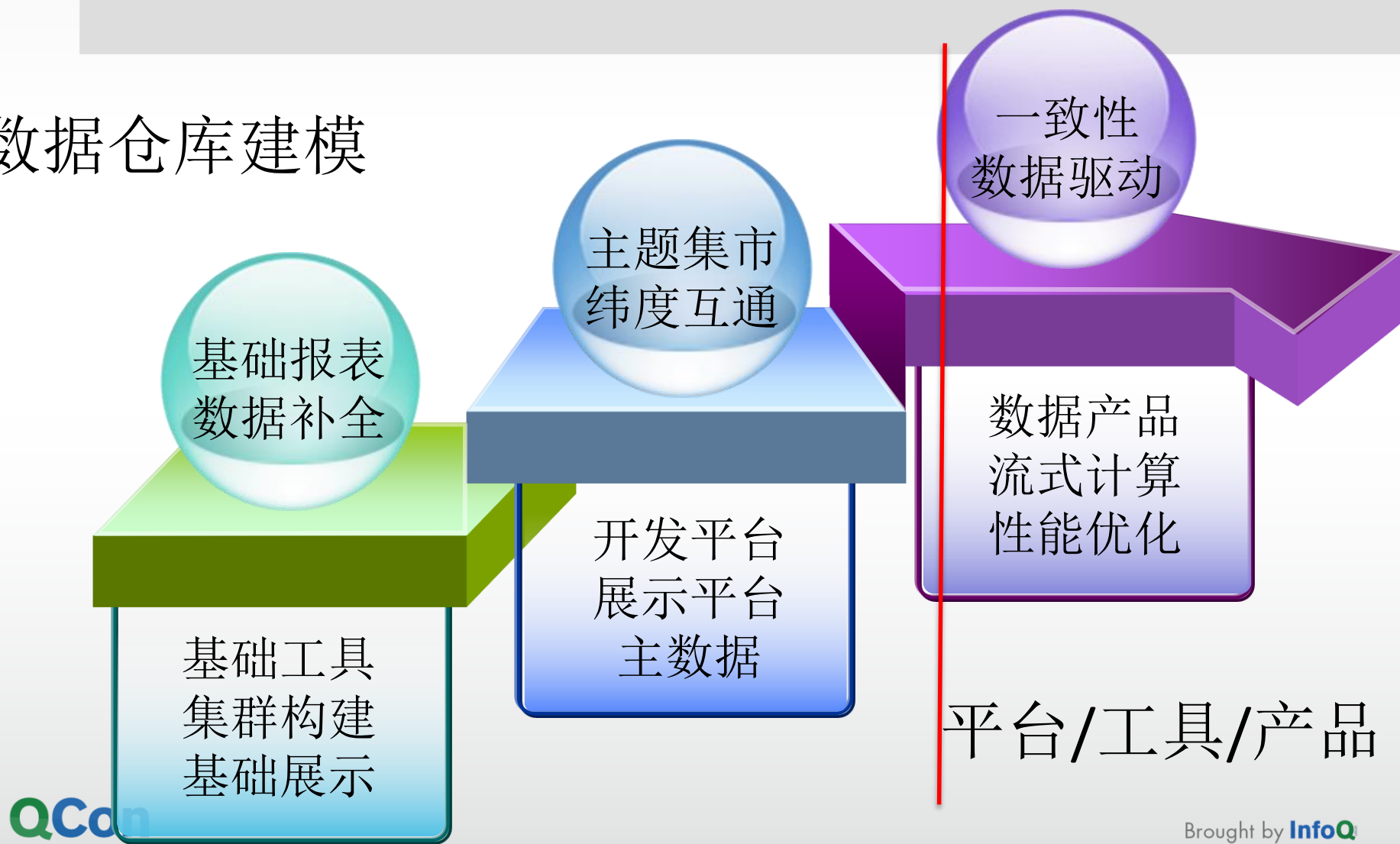


内部系统 VS 用户体验

- 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





@InfoQ



infoqchina

Q&A
谢谢大家！

Brought by **InfoQ**