



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



大数据分析 & 挖掘在淘宝的 应用与挑战 -毛波



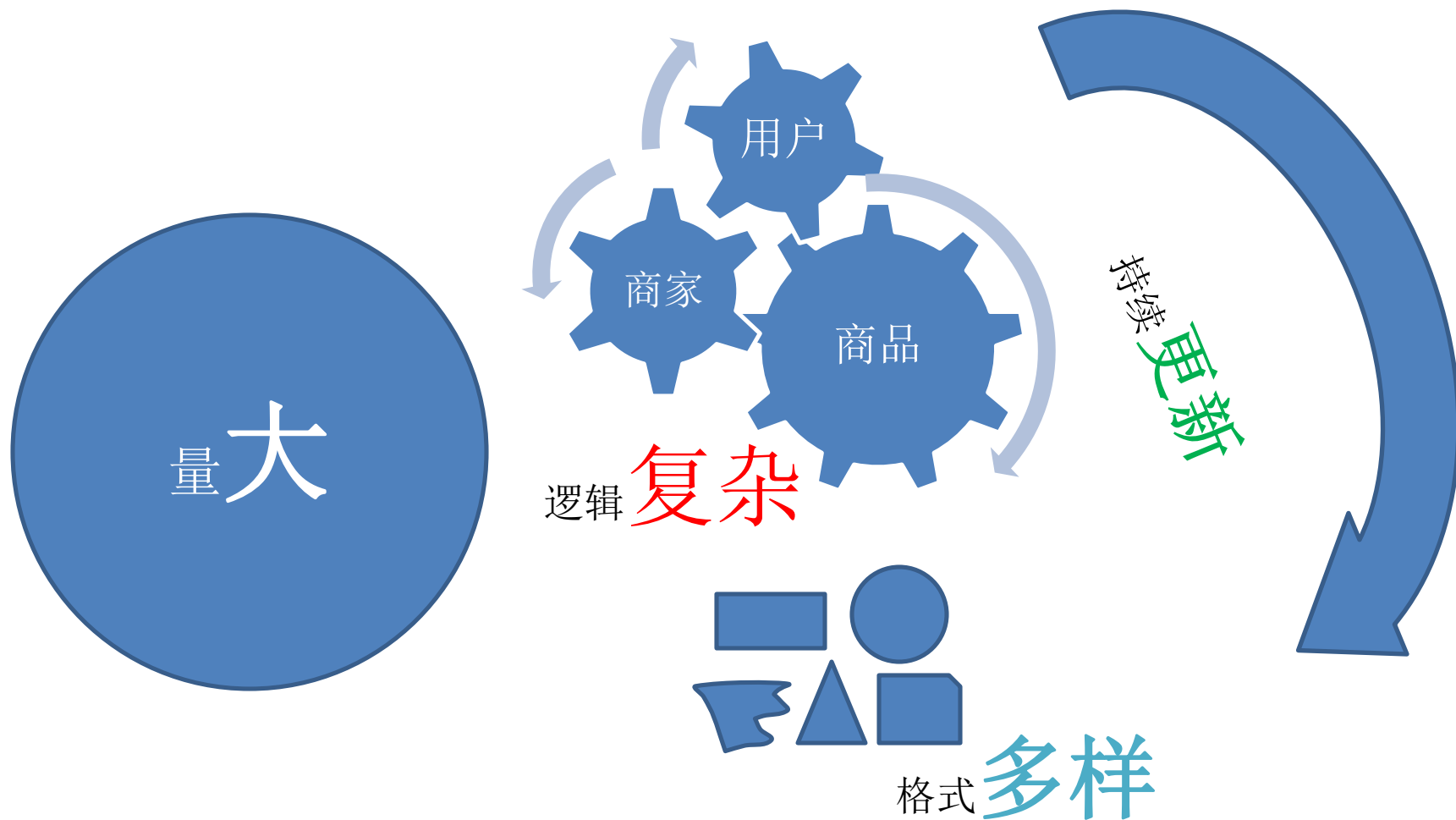


目录

- 大数据之我见
- 大数据-甜蜜的烦恼
- 淘宝数据技术变迁
- 淘宝数据架构
- 瞬间爆炸式增长数据的应对
- 现在及未来
- 一些观点



大数据之我见





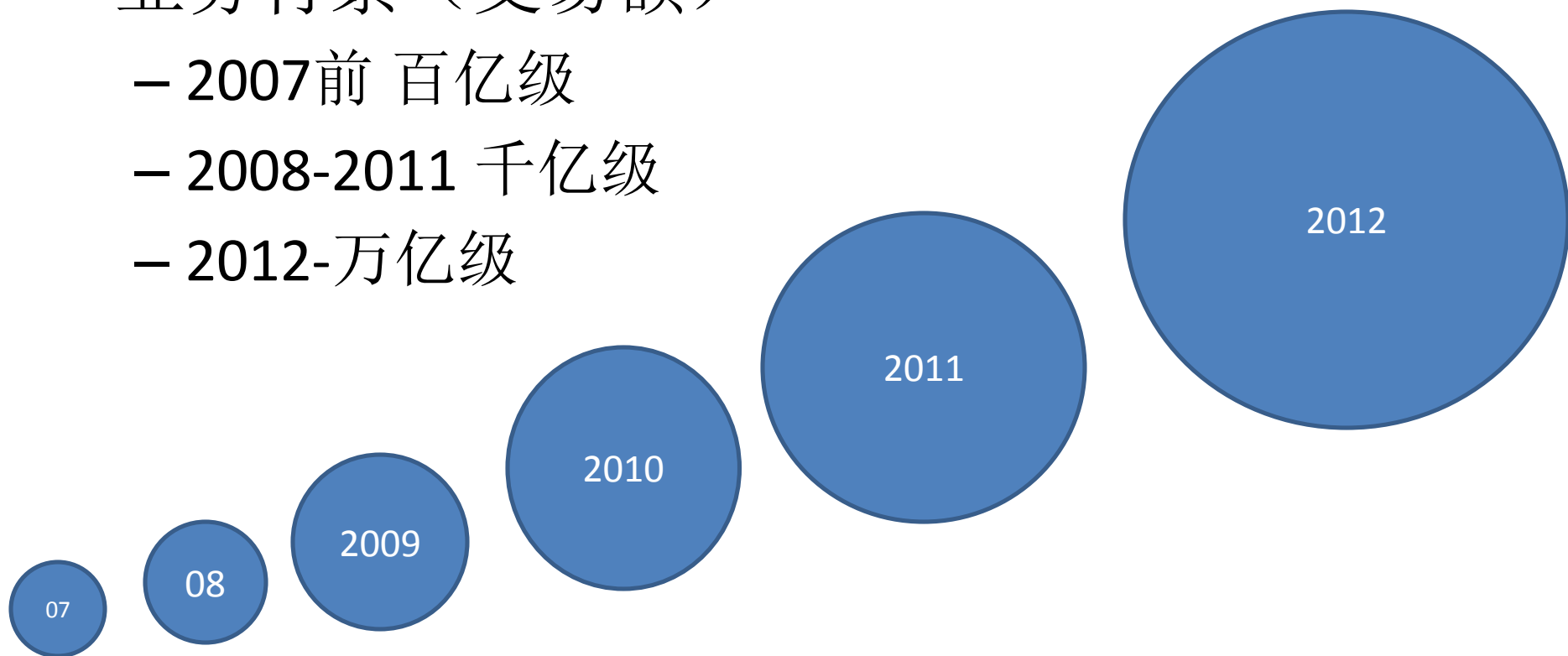
大数据-甜蜜的烦恼

- 手中有数，心中无术
- 数据处理失控
- 数据不一致



淘宝数据技术变迁

- 业务背景（交易额）
 - 2007前 百亿级
 - 2008-2011 千亿级
 - 2012-万亿级





淘宝数据技术变迁（续）





淘宝数据技术变迁（续）

数据库
脚本

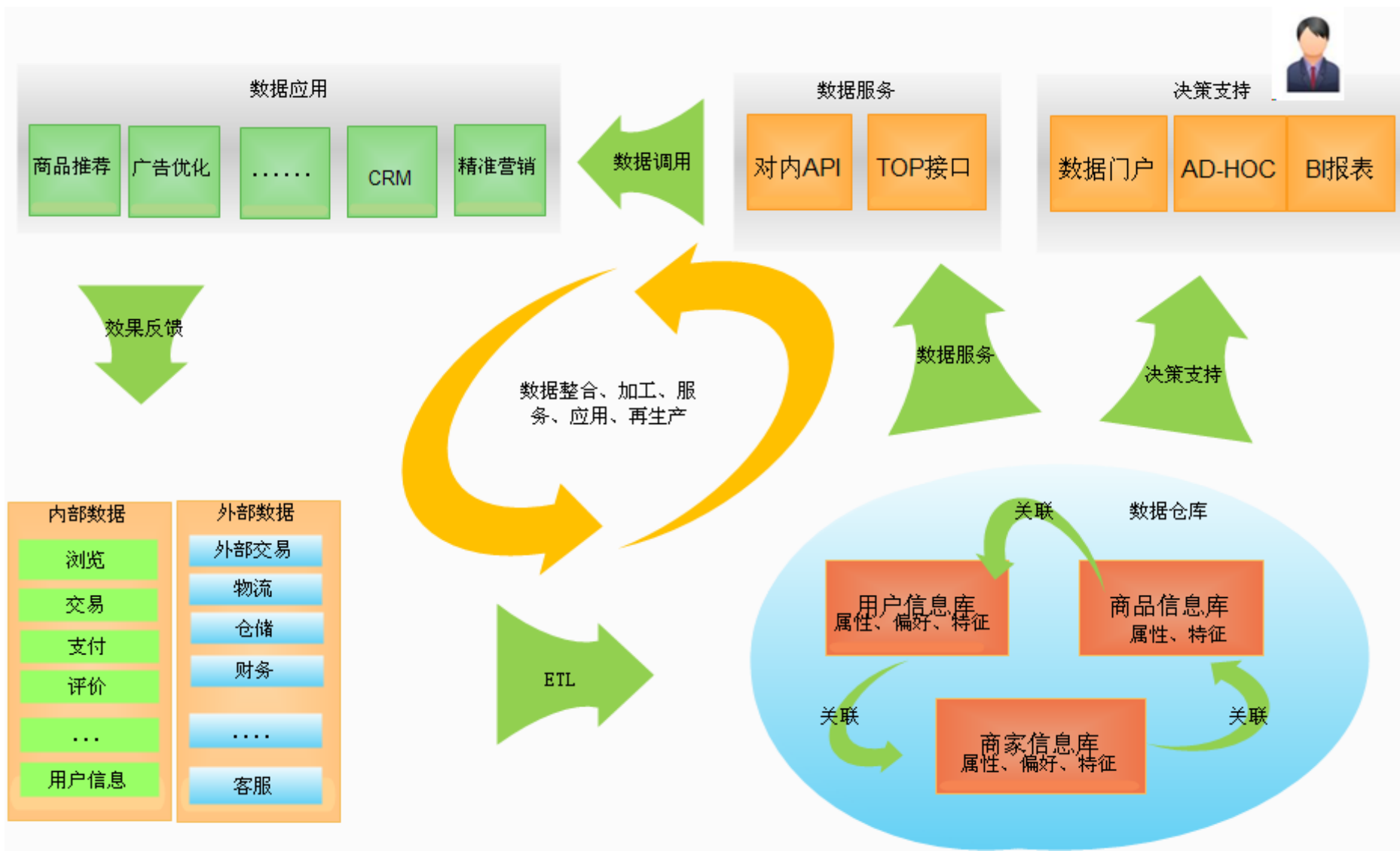
数据库集群
脚本
BO/MSTR

Hadoop集群
调度监控
数据门户

Hadoop集群
Storm
调度、监控
实时日志传输
实时数据库同步
数据门户
数据自助查询工具

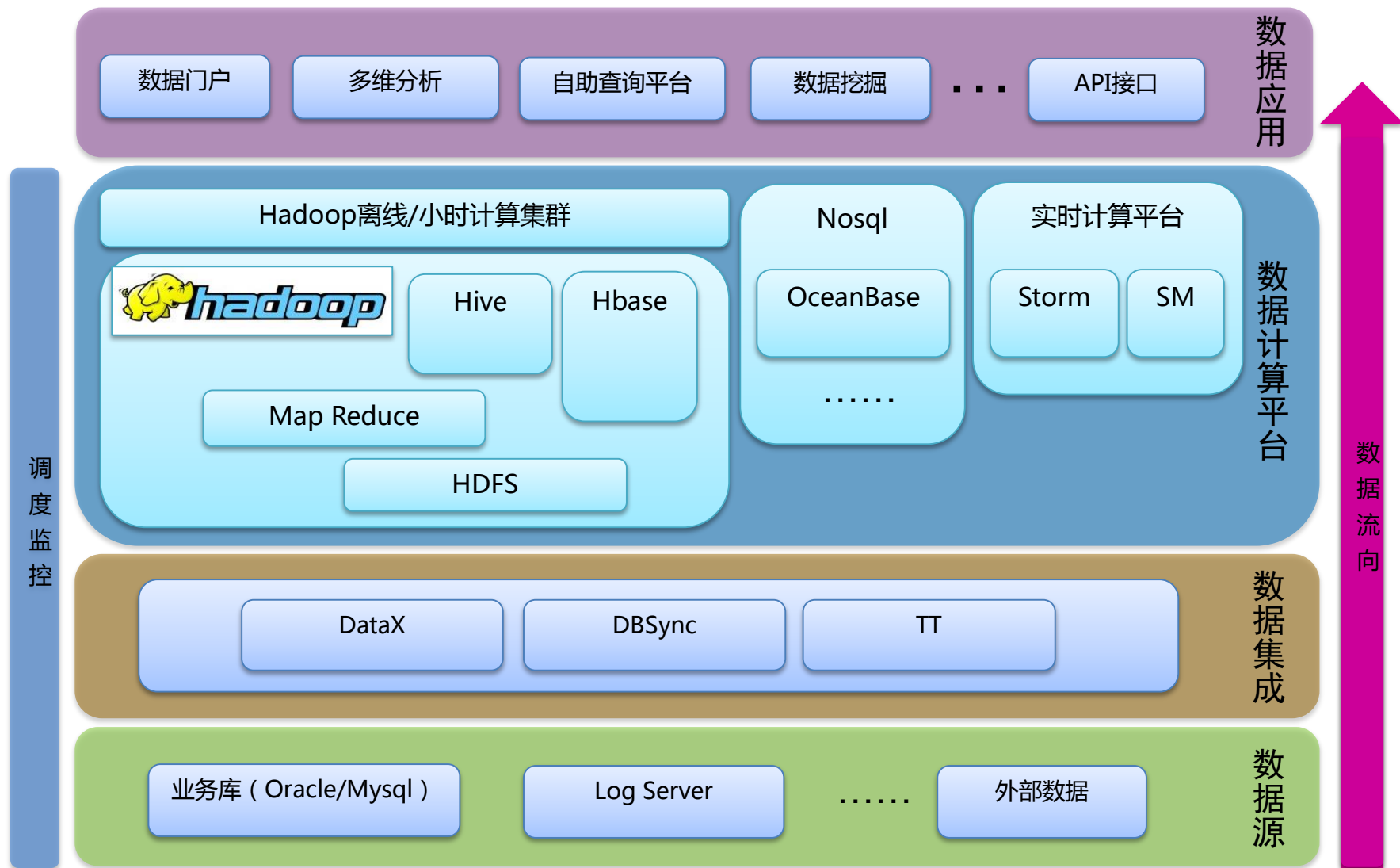


淘宝数据架构-逻辑





淘宝数据架构-物理





瞬间爆炸式增长数据的应对

- 双11购物狂欢节
 - 2011年11月11日，全网单日成交52亿
 - 高峰期10分钟1亿



瞬间爆炸式增长数据的应对（续）

- 提前做系统清理
- 硬件临时扩容
- 救急方案预备
 - 服务降级方案
 - 救火软硬件池
- 数据分散化
 - 实时传输
 - 实时计算
 - 阶段性汇总
- 随时待命人肉处理



现在及未来

- 实时化
 - 天->小时->分钟->秒
- 展现个性化
 - 同一数据，不同视角
- 即席化
 - 基于搜索引擎等技术
 - 多维度
 - 即席定制，秒级反馈



一些观点

- 结合事件看数据
- 别听TA的，给TA想要的
- 数据是血液，融在业务中

谢谢

