

大数据分析方法与数据化决策

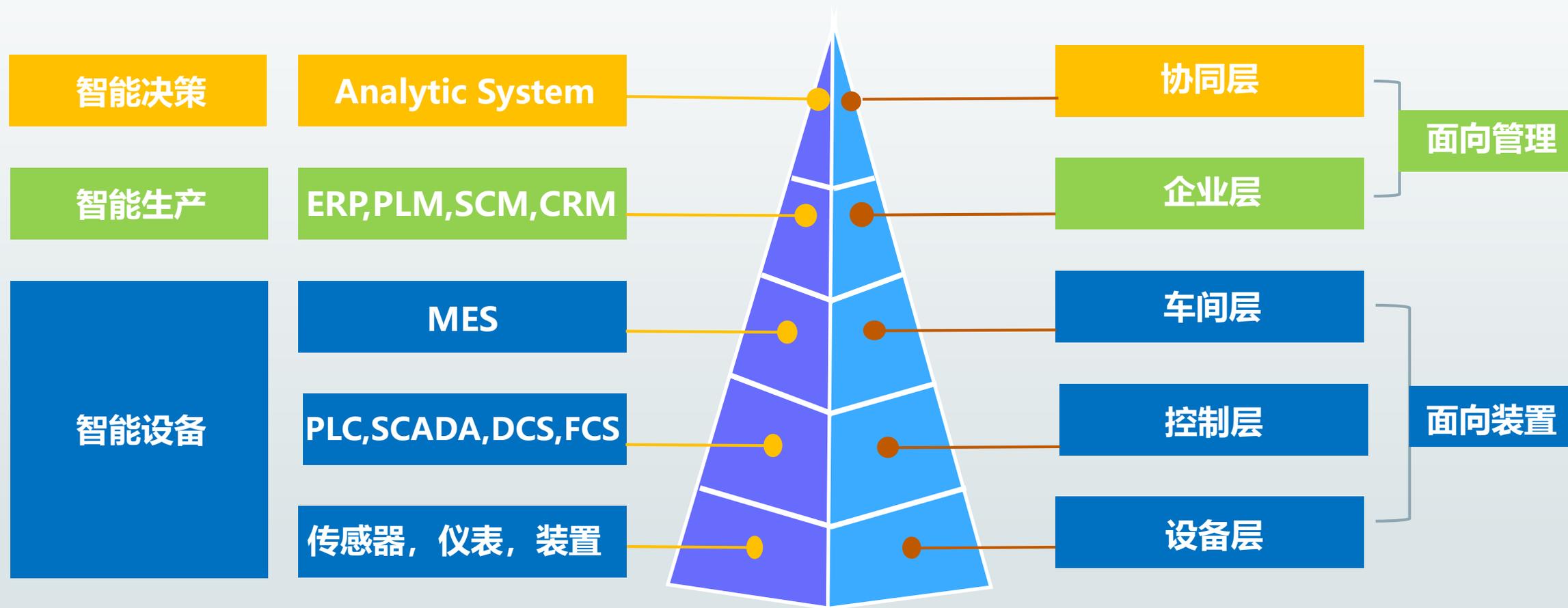
赵峰

甲骨文大数据事业部

Feng.z.zhao@oracle.com

Mobile:13701250400

大数据在企业中的应用领域 – 数据层两化融合

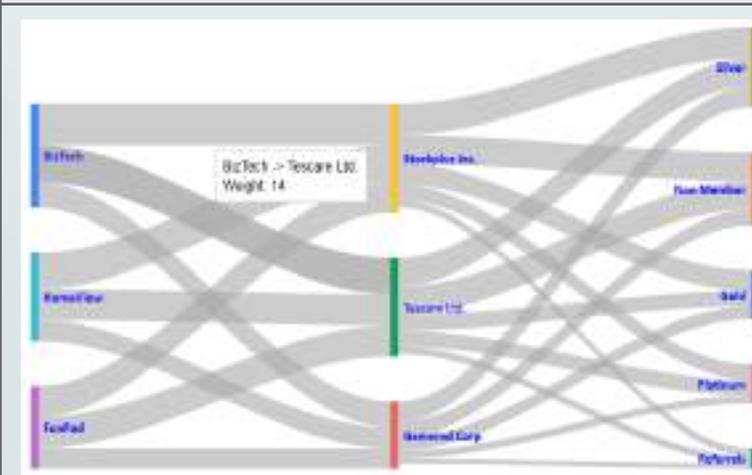
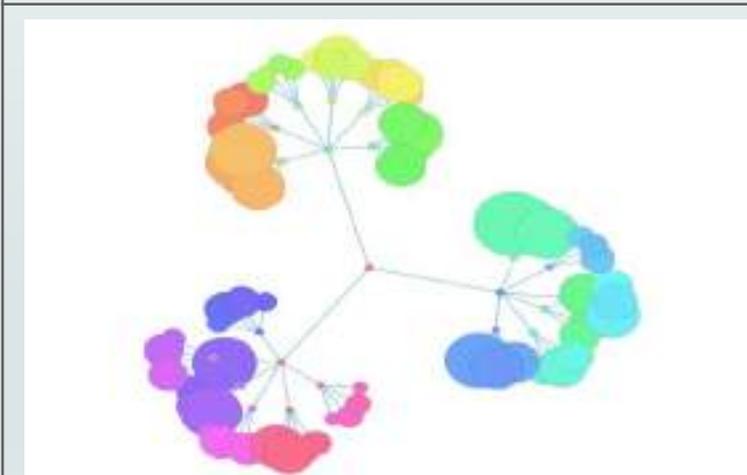
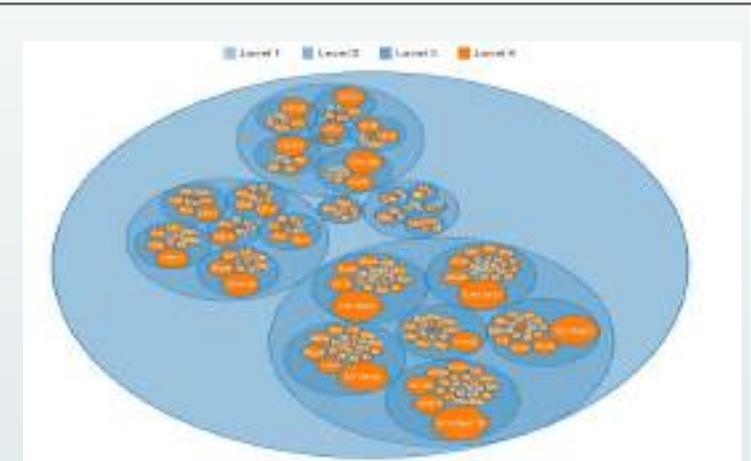
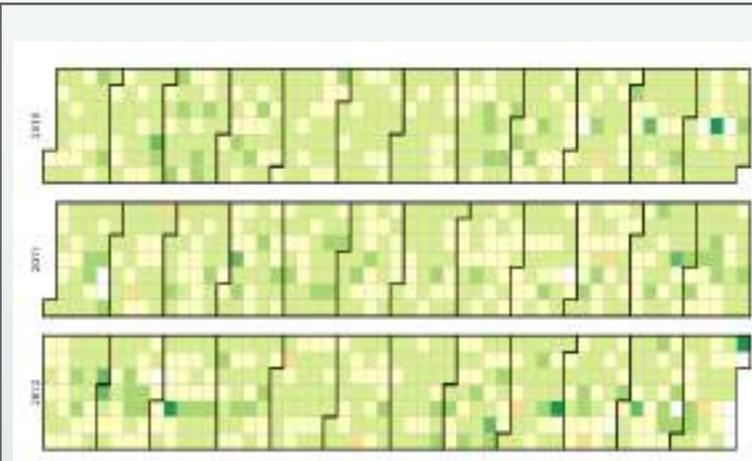
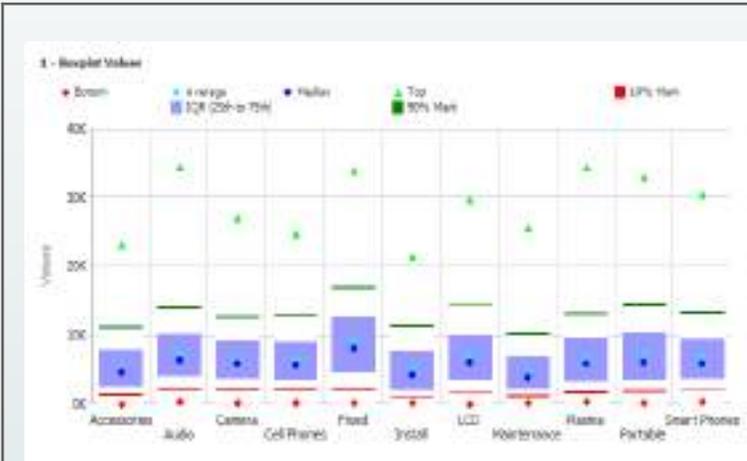


《国家智能制造标准体系建设指南（2015年版）》

大数据分析事关IT还是事关业务部门?



大数据分析统计，展示

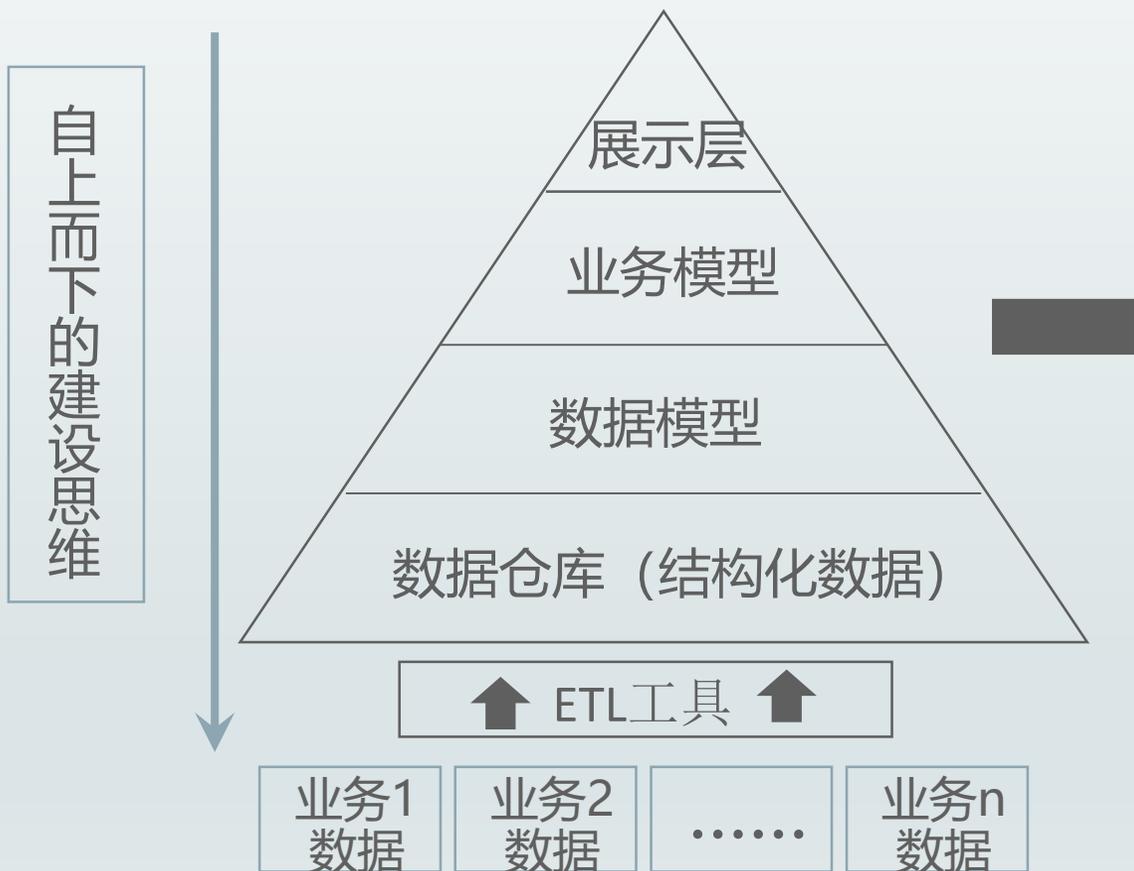


	Gold	Non Member	Platinum	Safaris	Silver
2010 Q1	410271.90	860818.84	492330.81	433405.38	182,094.88
2010 Q2	482719.27	175820.90	123813.14	658415.57	2099025.01
2010 Q3	1414605.69	1752953.90	1971737.58	521396.66	1979482.85
2010 Q4	301062.03	1320200.03	894771.02	264663.37	1010400.70
2011 Q1	682099.99	989054.49	810770.23	363207.31	1004982.35
2011 Q2	148887.73	1722116.70	136282.83	812888.74	1888887.43
2011 Q3	1492951.01	1846335.97	1264832.43	959204.37	2087193.26
2011 Q4	884170.65	1151151.10	757350.26	382348.34	1235413.83
2012 Q1	782220.88	1300889.18	808709.46	468864.22	817482.33
2012 Q2	1310910.28	1484871.52	1113801.48	727707.81	1018227.82
2012 Q3	1808815.73	2584250.24	1478042.88	1891282.89	2387825.84
2012 Q4	934287.79	1102054.44	804843.78	289326.22	1244938.78

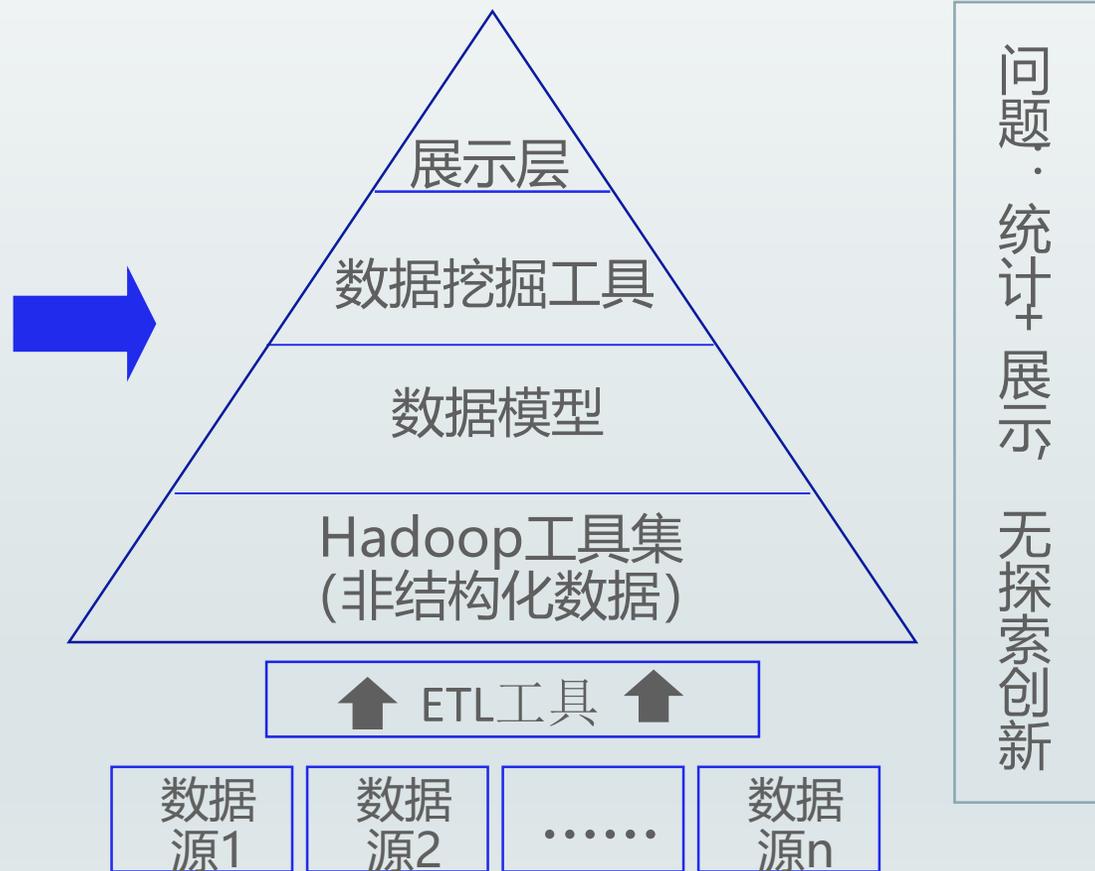


常见的大数据建设思路 – “BI路径”

BI系统建设层次



大数据系统建设层次



从业务改进的循环看分析

1. 界定

2. 测量

3. 分析

4. 改进

5. 控制

过程改进

- 识别问题
- 界定需要
- 设定目标

- 确认问题/过程
- 精炼问题/目标
- 测量关键阶段/投入

- 提出问题假设
- 识别少数关键原因
- 验证假设

- 解决根本问题的方法
- 验证解决方法
- 标准化解决方法/测量结果

- 建立维持绩效的标准测量
- 及时纠正问题

过程设计/再设计

- 确定具体/主要问题
- 定义目标/改变的愿景
- 明晰改变的范围和内外顾客需要

- 根据需要测量绩效
- 收集过程效率数据

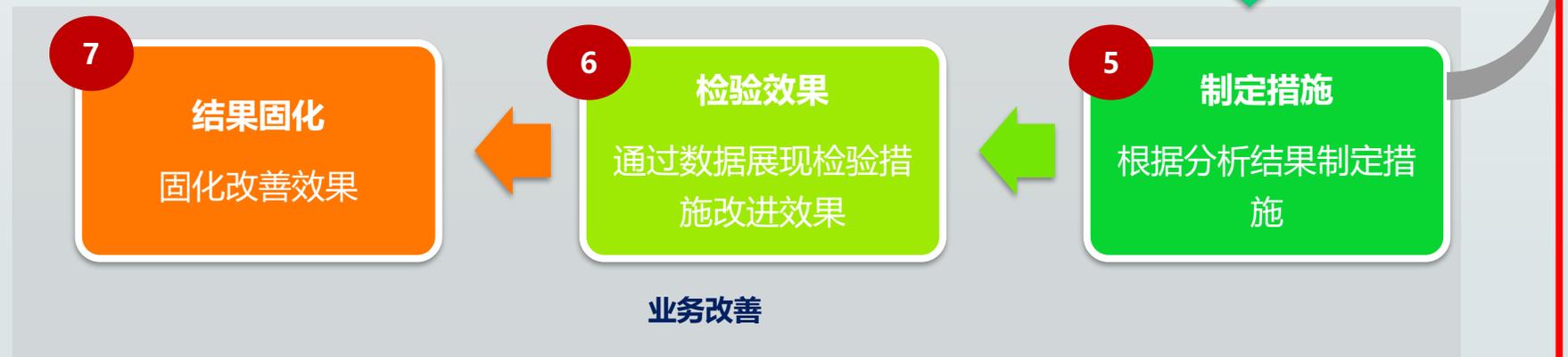
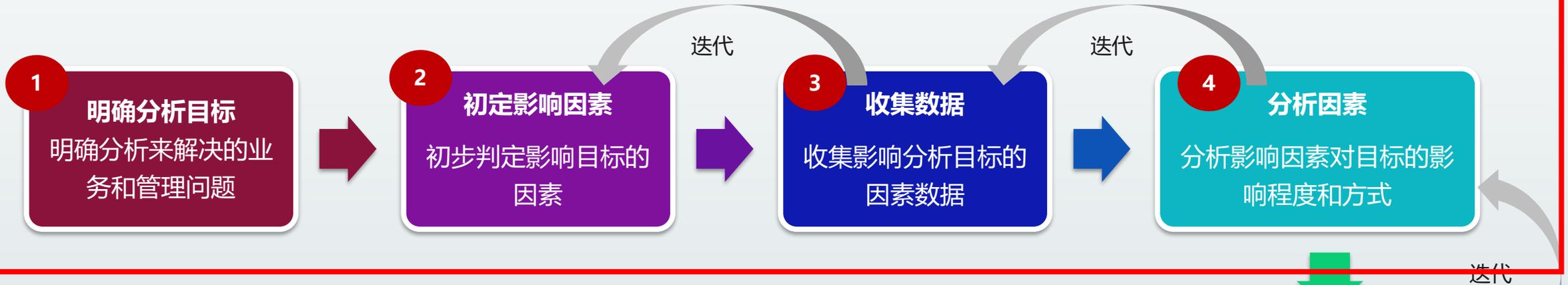
- 识别“最优方法”
- 评估过程设计
 - 增值/非增值
 - 瓶颈问题/脱节问题
 - 备选方案
- 精炼客户需要

- 设计新过程
 - 挑战假设
 - 运用创造力
 - 流程法则
- 实施新的过程, 结构, 系统

- 建立维持绩效的测量和评估
- 及时纠正问题

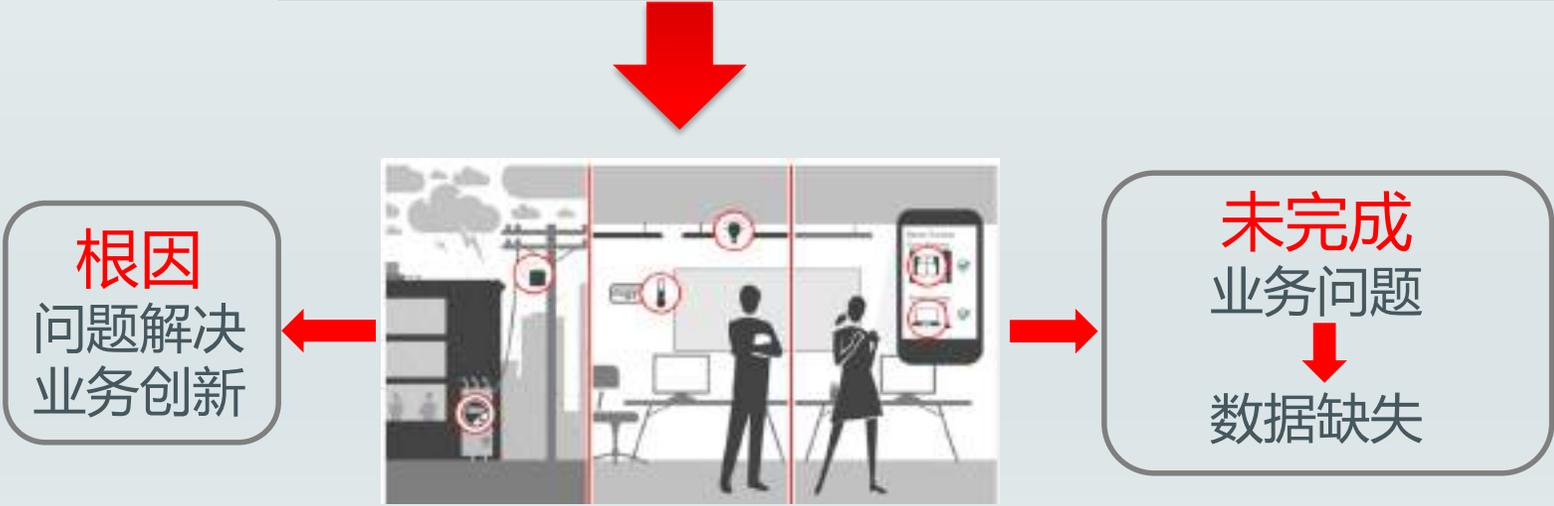
数据分析与业务优化流程

大数据分析

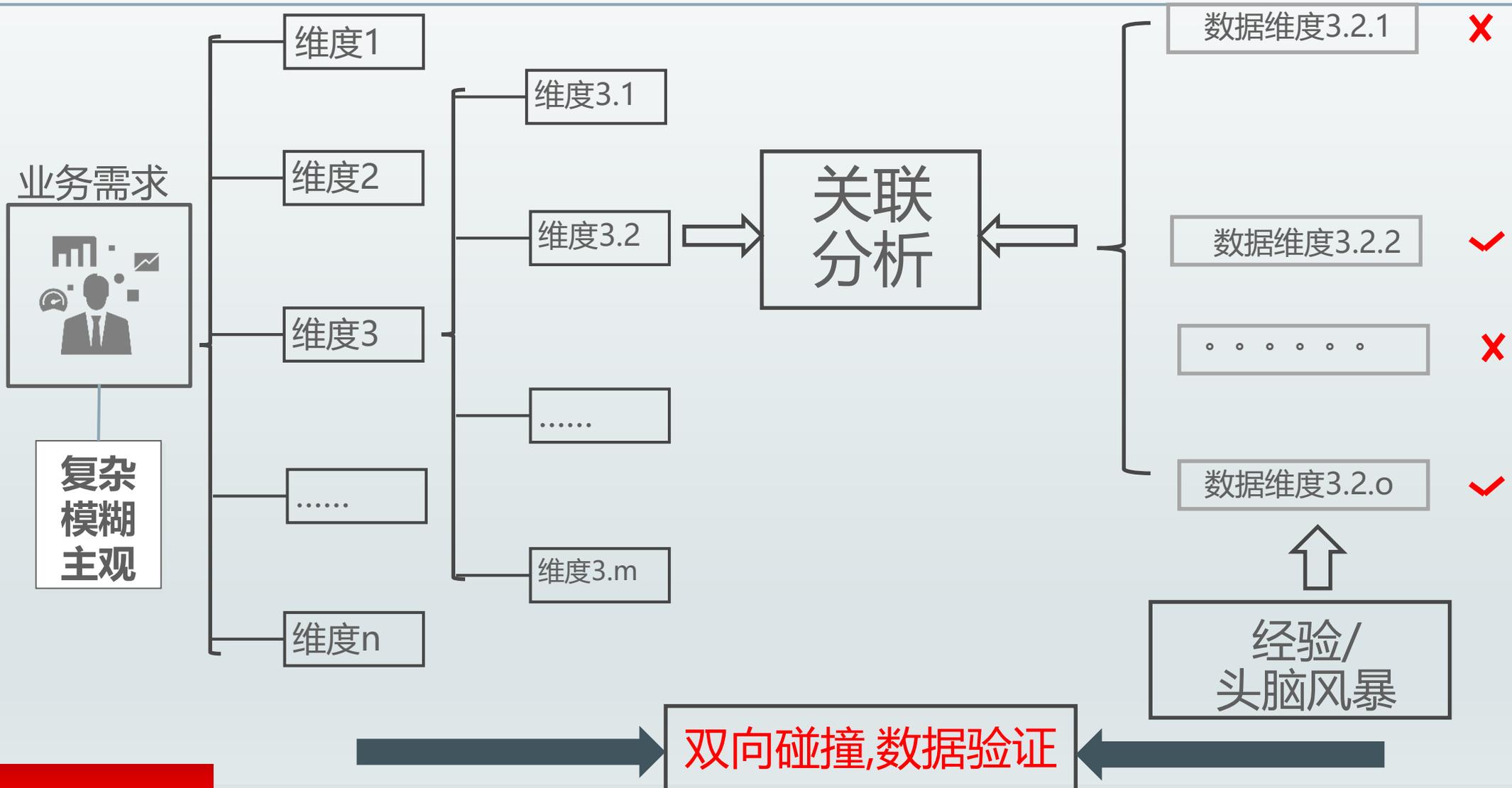


管理与业务流程优化

为业务提升而分析：敏捷的数据探索 – 快速，迭代



业务价值与数据之间的一见钟情 – 挖掘“根因”

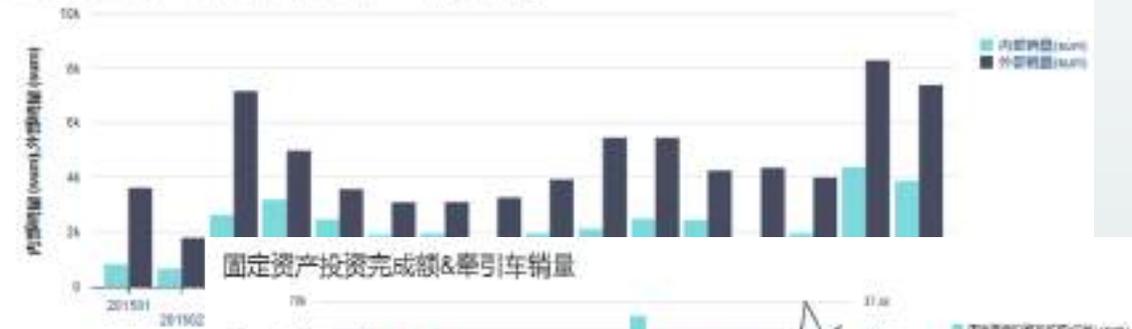


业务问题拆解 – 麦肯锡MECE分析原则 (相互独立, 完全穷尽)

分析内容	关注点	内容	关注点	分析方法/模型
		数据源验证	验证外部数据源可信度	分析内外部数据源的相关性
市场	GDP	宏观环境	GDP	趋势, 牵引车销量相关性分析
	SUV销量		全社会固定资产投资总额	趋势, 牵引车销量相关性分析
	X车型销量		零售商品消费额	趋势, 牵引车销量相关性分析
	出口额		趋势, 牵引车销量相关性分析	
竞争	数据源可靠性分析	货物运输	公路货运量	趋势, 牵引车销量相关性分析
	分析锁定竞品	市场与销售	牵引车销售市场份额	市场份额趋势
	销量比较		大马力牵引车销量	大马力牵引车在牵引车市场中的消长
	锁定主要竞争市场		确定竞品	通过累计销量排名确定竞争厂商和竞品
生命周期	产品生命周期分析	竞品销量	相关竞品销量表现	
产品	价格分析	竞品分析	价格分析	分析竞品价格差异
	低配置车型占比分析		油耗分析	分析竞品油耗差异
	基于口碑模型的竞品对比		营销特点分析	解放营销方式(大单)分析
	基于质量状况模型的竞品		技术水平分析	按照技术水平评价模型比较竞品间差异
营销	与竞品在主要市场的促销	预测	预测未来销量	根据预测模型预测未来销量
销量预测	短期销量预测模型	总结	从环境、市场需求、销售、竞争综合	逐层递进
总结	从宏观环境、竞品、营销、产品方面综合分析销量不佳原因		从宏观到微观, 逐层递进分析	

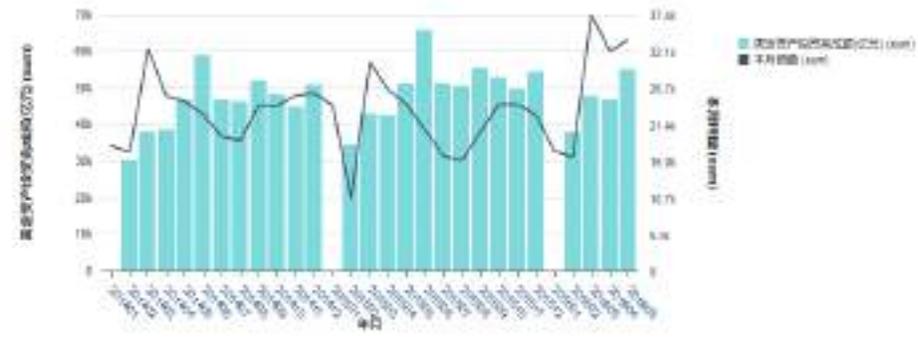
数据关联度检验

内部销量 (sum), 外部销量 (sum) by 年月



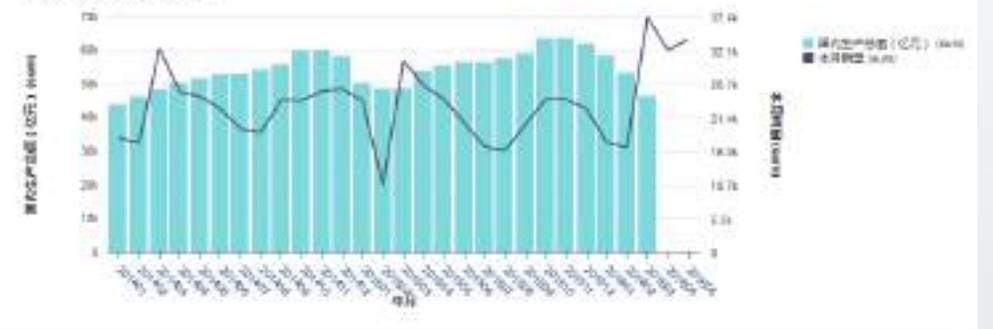
0.87

固定资产投资完成额&牵引车销量



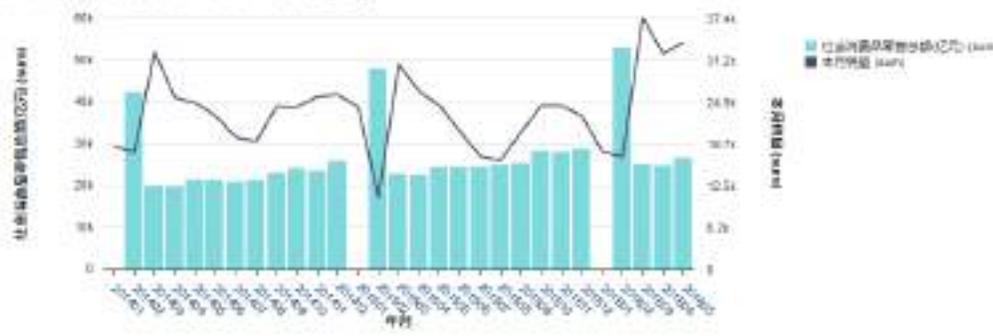
0.23

GDP&牵引车销量



-0.44

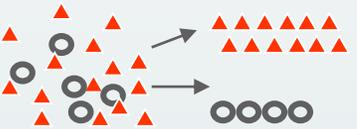
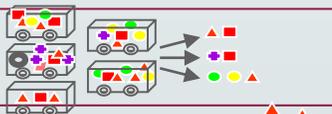
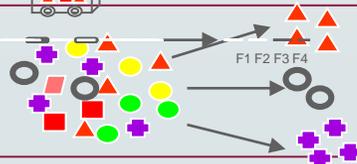
社会消费品零售总额&牵引车销量



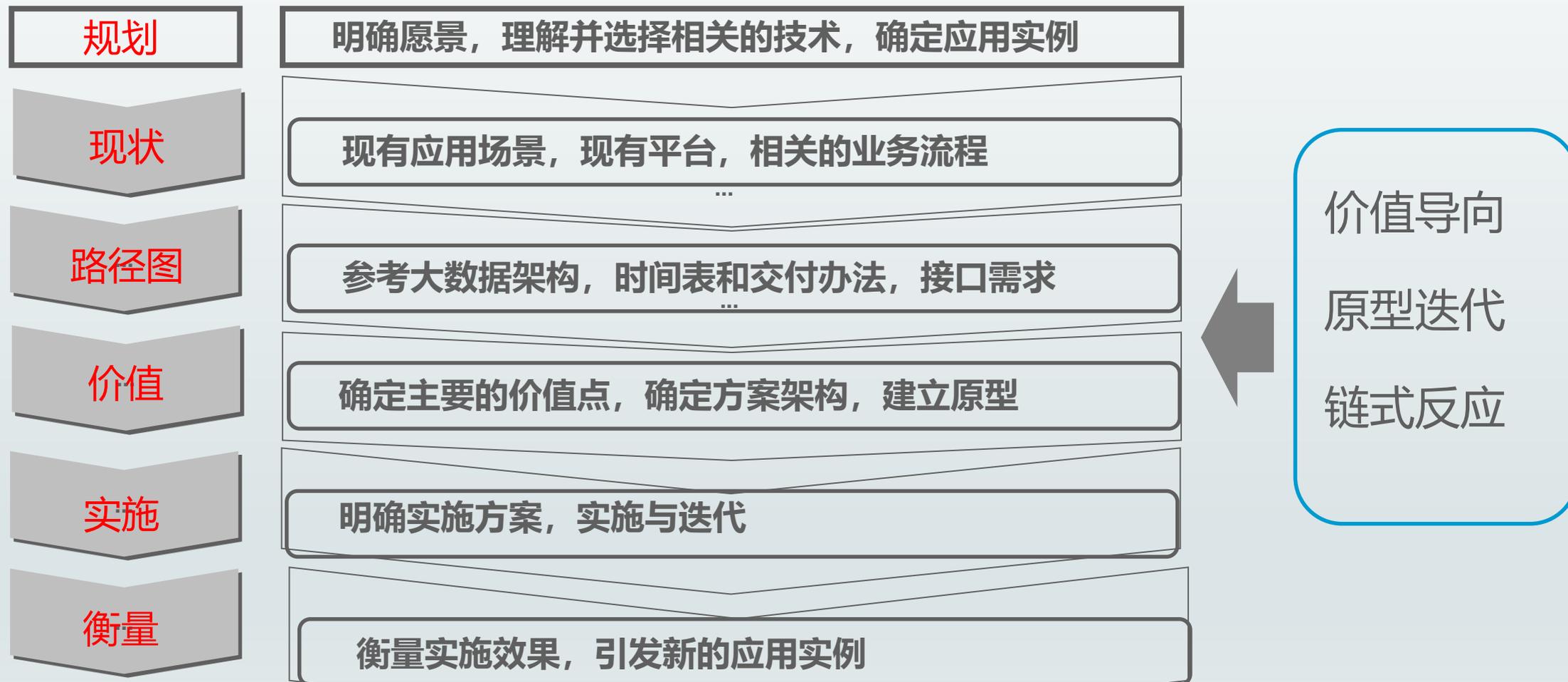
-0.19



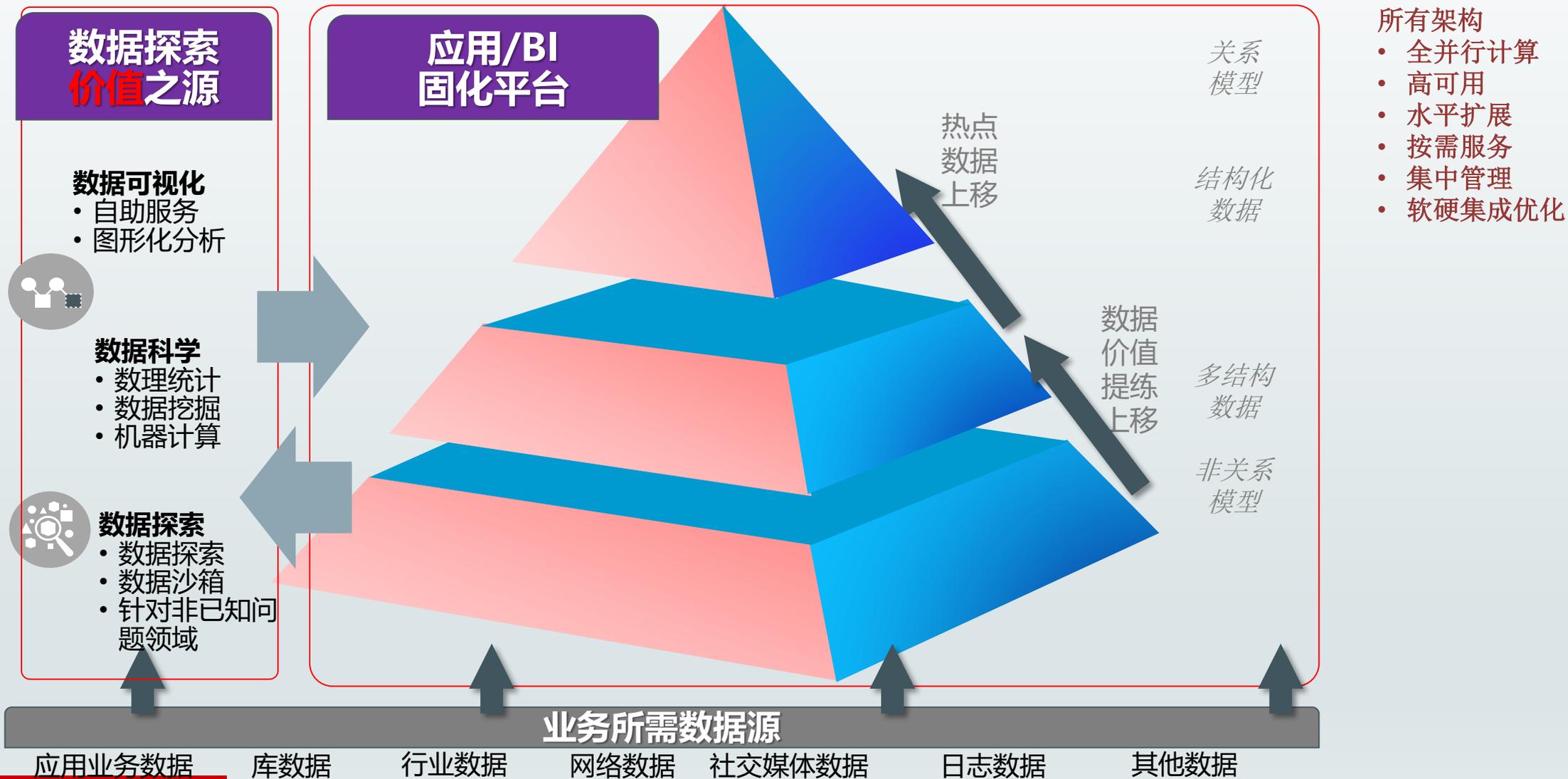
在强关联关系中建立算法模型 – 从关联走向因果

类型	算法	适用性
分类 	Logistic 回归 (GLM) 决策树 贝氏 支持向量机	传统统计技术 流行/规则/透明度 嵌入式应用程序 宽泛/狭窄的数据/文本
回归 	多重回归 (GLM) 支持向量机	传统统计技术 宽泛/狭窄的数据/文本
异常检测 	一类 SVM	缺少目标领域的示例
属性重要性	最短描述长度 (MDL)	属性精简 识别有用的数据 减少数据噪声
关联规则 	Apriori	购物篮分析 链接分析
聚类 	分层 K 均值 分层 O 聚类	产品分组 文本挖掘 基因和蛋白质分析
特性提取	非负矩阵因式分解	文本分析 特性精简

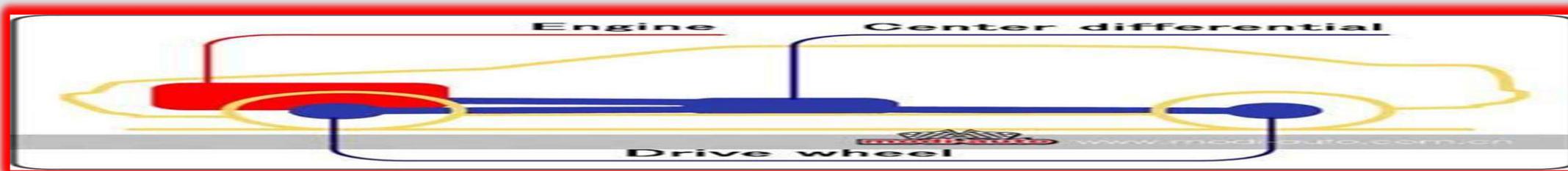
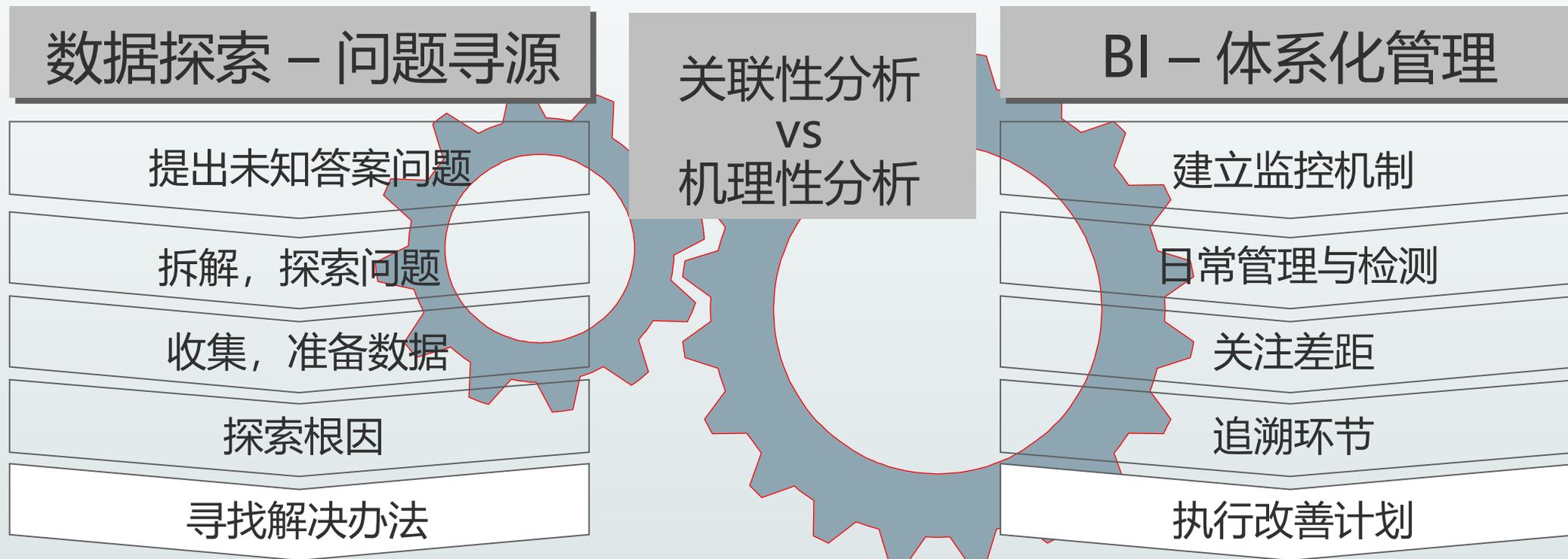
大数据建设：一个跨度较长的系统过程



数据探索与体系化建设



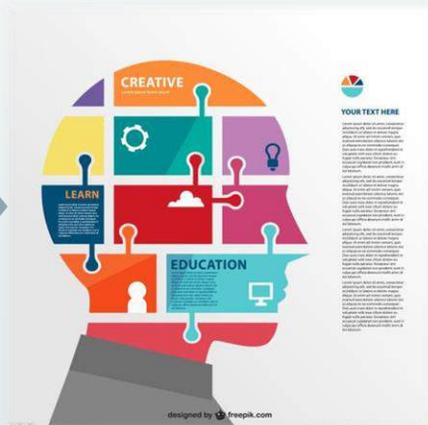
数据探索与数据化管理体系 - 双引擎驱动的业务持续优化



从数据到决策



数据报表



状态? 原因?



开会, 汇报

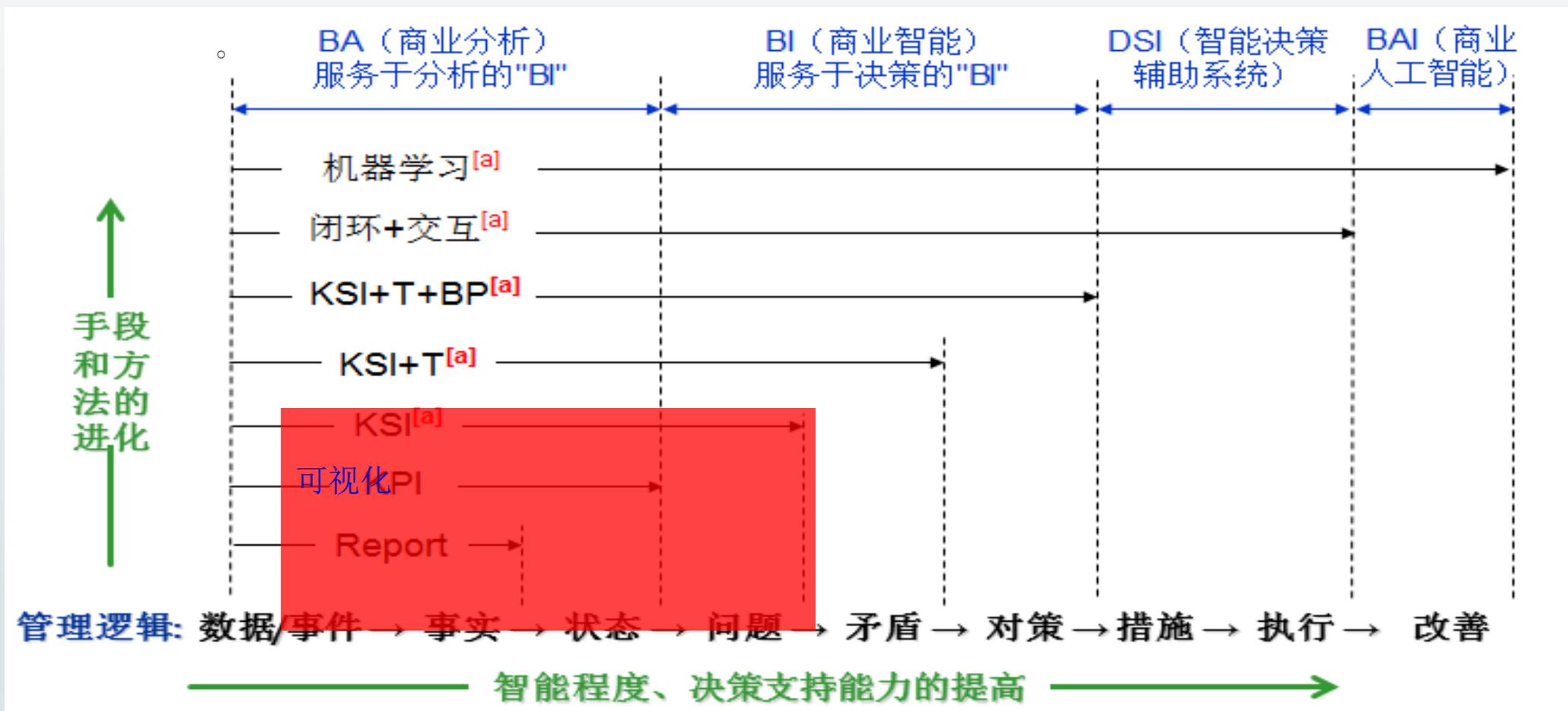


领导决策



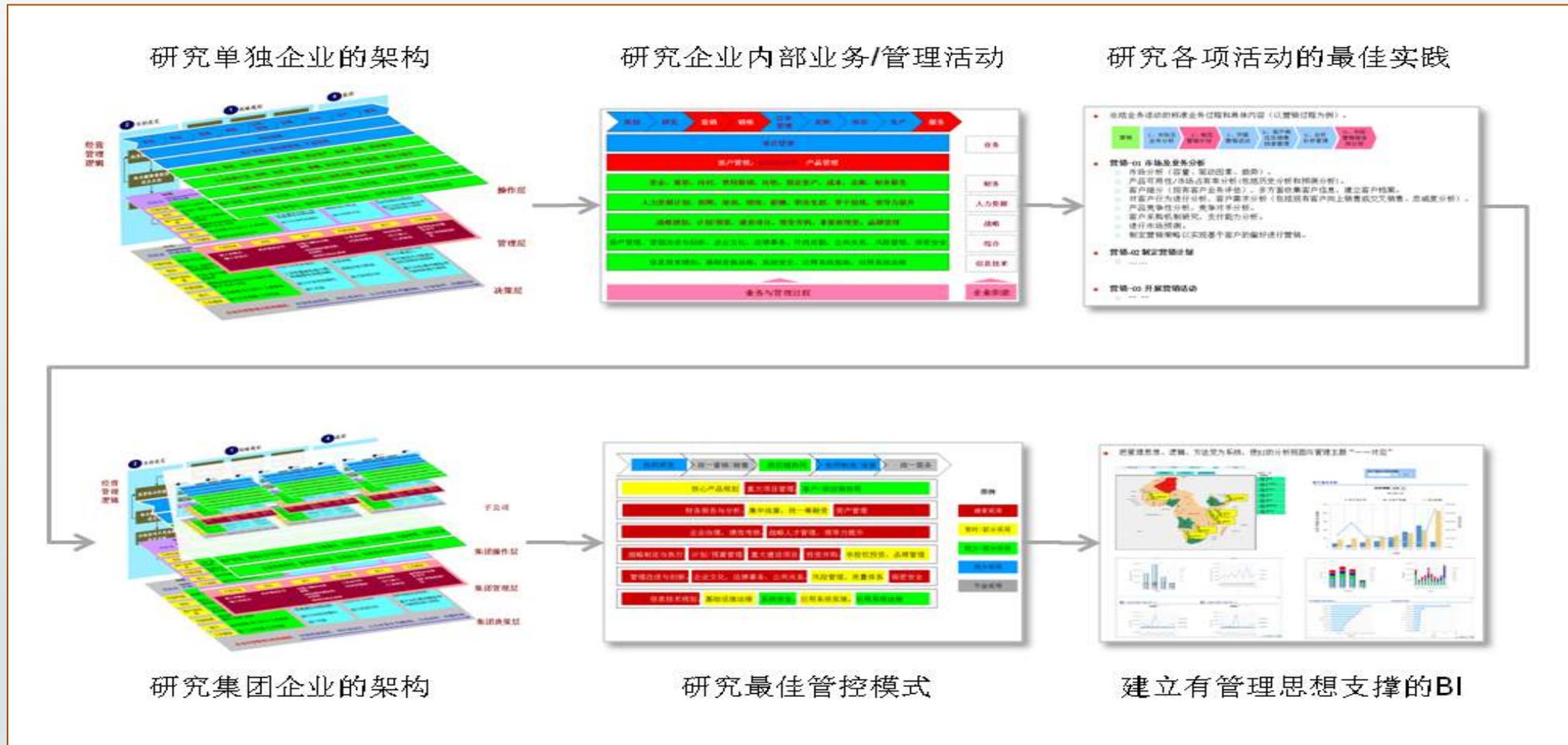
寻求解决问题的途径 *in-Sight* 决策支持系统价值提升模型

业界常用的主要方法（报表、KPI、可视化）均不能产生智能的决策支持系统

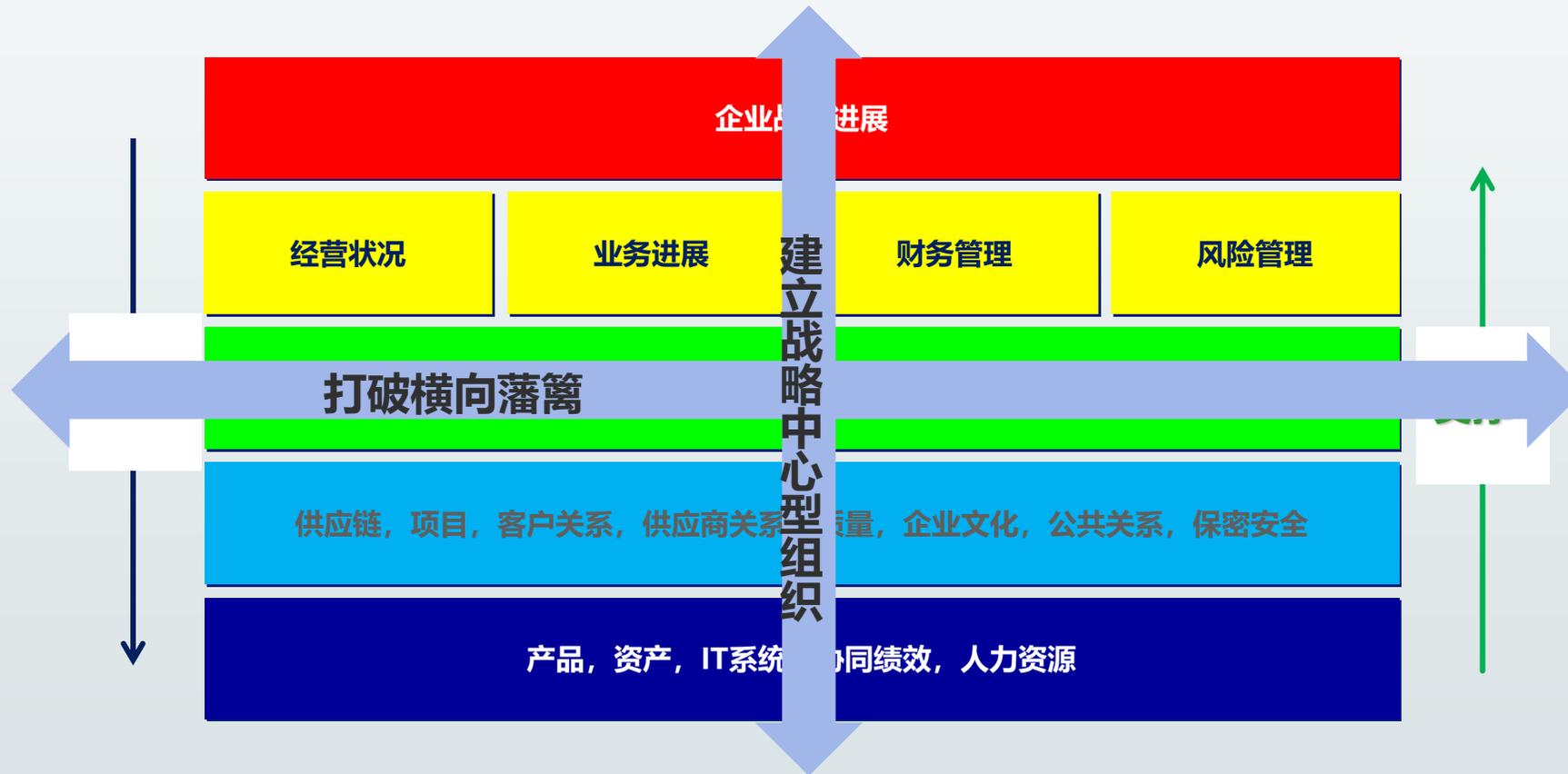


数据分析，管理先行

- 完整系统地认识组织的运行方式和管理逻辑，构建完整的决策支持系统

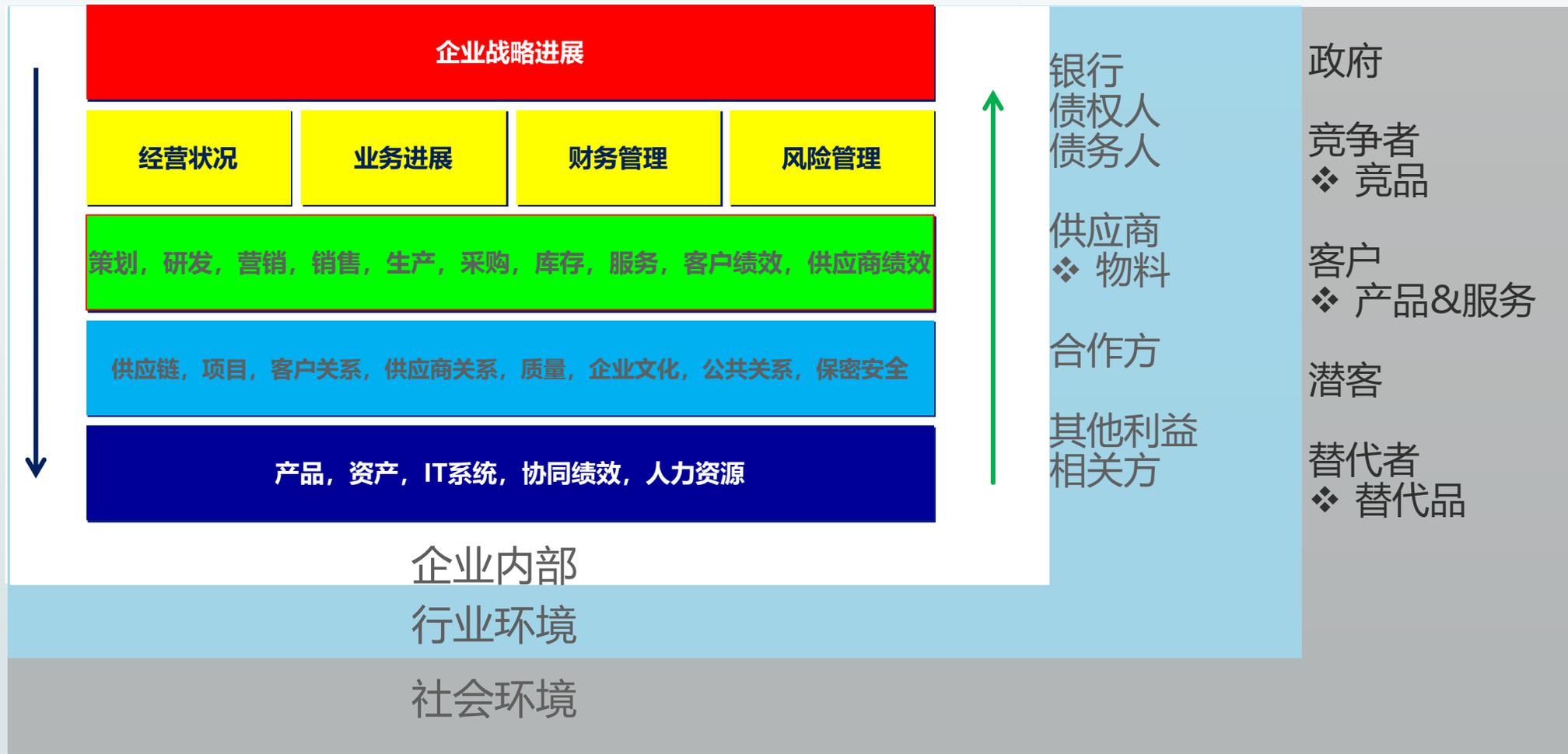


组织经营体系与数据分析领域



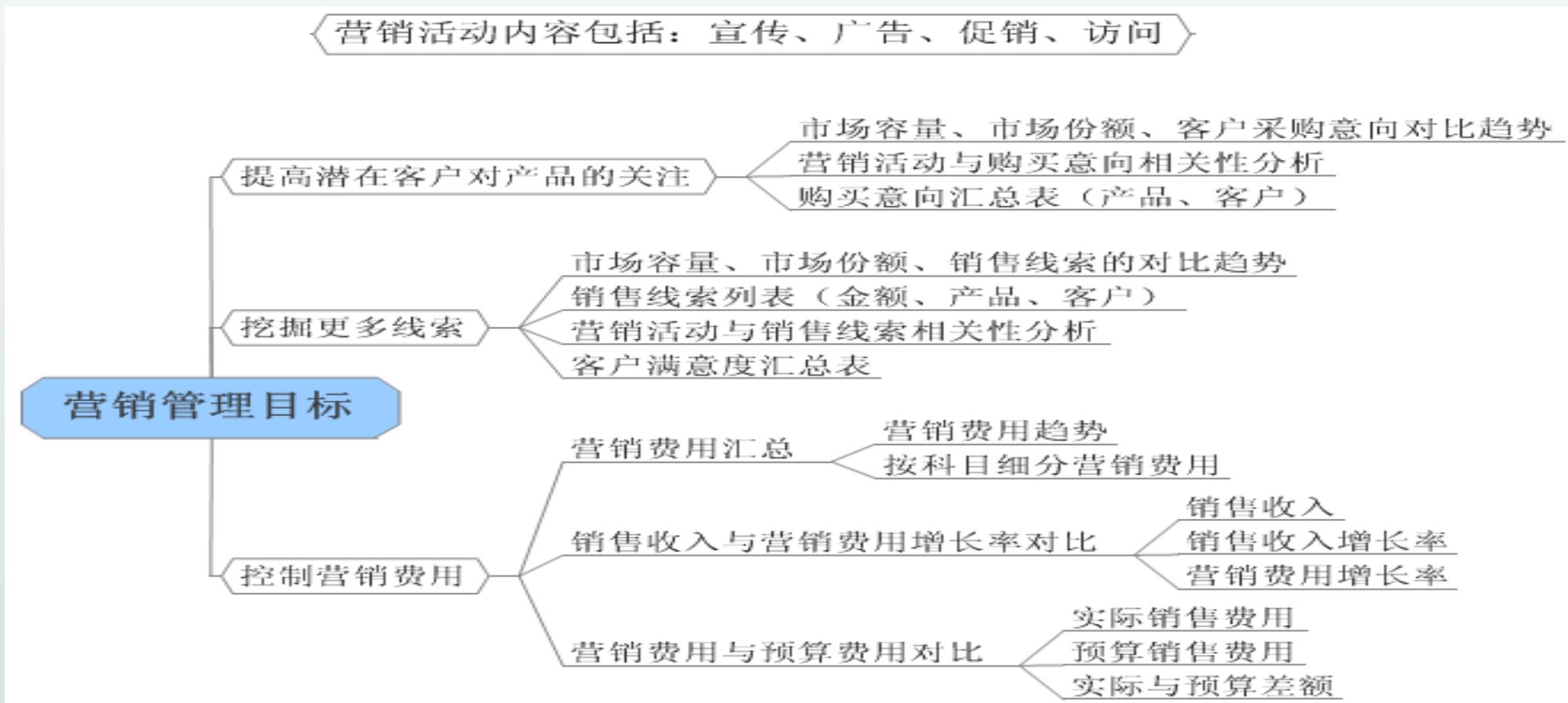
企业经营分析体系模型

组织内外环境与数据分析



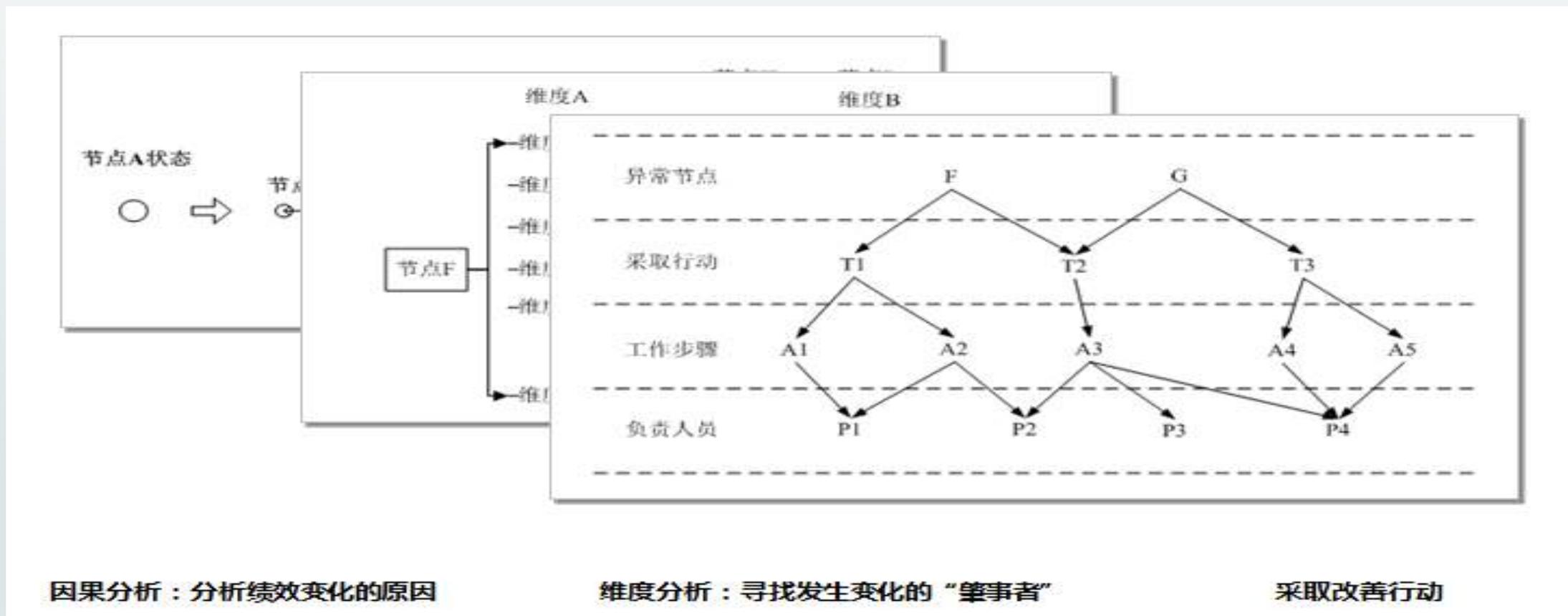
数据分析，管理先行

- 把管理目标系统地分解、落实到业务过程中的数据，构建决策分析模型（示意）



应用通过数据追踪运营问题根源的方法

- ◆ 通过分析，发现问题的根本原因



数据的种类和价值

价值稀疏数据

互联网“大数据”

实时定价

用户建模与动态营销

语义搜索购物

视频感知配菜

基于顾客识别的服务

设备运行监测与主动维修

监测实时工况优化管理

预

金

消

机构“大数据”

犯罪率预测与预防

舆情监控

非医疗公众数据聚合

流行病预测

基于大样本病例的诊断

基因数据的分析和利用

群体行为相关分析与预测

交通

运动

土壤

食品

行政

治安

价值密集数据

企业“大数据”

市场分析

战略进展

采购管理

供应商关系

用户社区

经营状况

库存管理

质量管理

供应商协同

业务进展

客户服务

产品管理

在线MRO

财务管理

客户绩效

资产管理

用户识别

风险管理

供应商绩效

IT管理

精准营销

策划与研发

供应链

协同绩效

全过程感知

营销与销售

项目管理

人力资源

绩效改善

生产管理

客户关系

保密与安全

组织中的数据生态建设

- 通过大数据实验室工作推进重点应首先保证团队的大数据意识和思维，**目标保持一致。**
- 其次按照专业领域详细分工，利用相关工具集，借助相关分析方法论，进行反馈，改进和迭代。明确相关工作责任部门，实现**专业分工。**
- 发挥各部门积极性，责任到岗，建立起相应的考核和激励机制。**协同推进**配套工作。



数据分析岗位建设

4. 数据探索方法

在实际工作中探索数据的方法论，并探索方法论的应用案例，开发工具，开发流程，其他数据源，提高数据探索效率，实现数据探索。

数据探索探索的方法论

探索	应用	工具
<ul style="list-style-type: none"> 数据源 数据清洗 数据探索 数据可视化 数据建模 数据应用 数据反馈 	<ul style="list-style-type: none"> 数据源 数据清洗 数据探索 数据可视化 数据建模 数据应用 数据反馈 	<ul style="list-style-type: none"> 数据源 数据清洗 数据探索 数据可视化 数据建模 数据应用 数据反馈

中国一汽 | in-Sight | ORACLE

1. 工作内容与职责

数据实验室内部人员的职责如下：

1. 数据分析师

- 根据业务提出数据需求，并开发数据探索工具，开发流程，其他数据源，提高数据探索效率，实现数据探索。
- 根据业务提出数据需求，并开发数据探索工具，开发流程，其他数据源，提高数据探索效率，实现数据探索。
- 根据业务提出数据需求，并开发数据探索工具，开发流程，其他数据源，提高数据探索效率，实现数据探索。

2. 大数据平台运维人员

- 负责大数据平台的日常运维
- 负责数据的安全、备份等工作
- 从技术角度为数据分析师提供技术支持

中国一汽 | in-Sight | ORACLE

4. 数据探索方法

1. 工作内容与职责

数据实验室

2. 人员岗位要求

3. 实验室工作流程

3. 实验室工作流程

不同模式下的“工作流-职责”示意图

序号	任务名称	二合一模式		三合一模式	
		数据分析师	运维人员	数据分析师	运维人员
1	数据源探索与清洗	数据源探索	数据清洗	数据源探索	数据清洗
2	数据探索与可视化	数据探索	数据可视化	数据探索	数据可视化
3	数据建模与应用	数据建模	数据应用	数据建模	数据应用
4	数据反馈与优化	数据反馈	数据优化	数据反馈	数据优化

中国一汽 | in-Sight | ORACLE

2. 人员岗位要求

总体上，数据分析师的技术能力涉及了：统计、空间、语言学、交互式发现、数据可视化、利用数据能力可以便于对海量的数据进行分析和挖掘，从而发现数据的新含义。此外，良好的分析思维、行业经验、编程开发能力和相关的知识都是合格的岗位要求。

数据分析师：

1. 数据与编程

- 对数据可视化与挖掘过程的了解
- 了解数据可视化与交互的原理

2. 数据应用

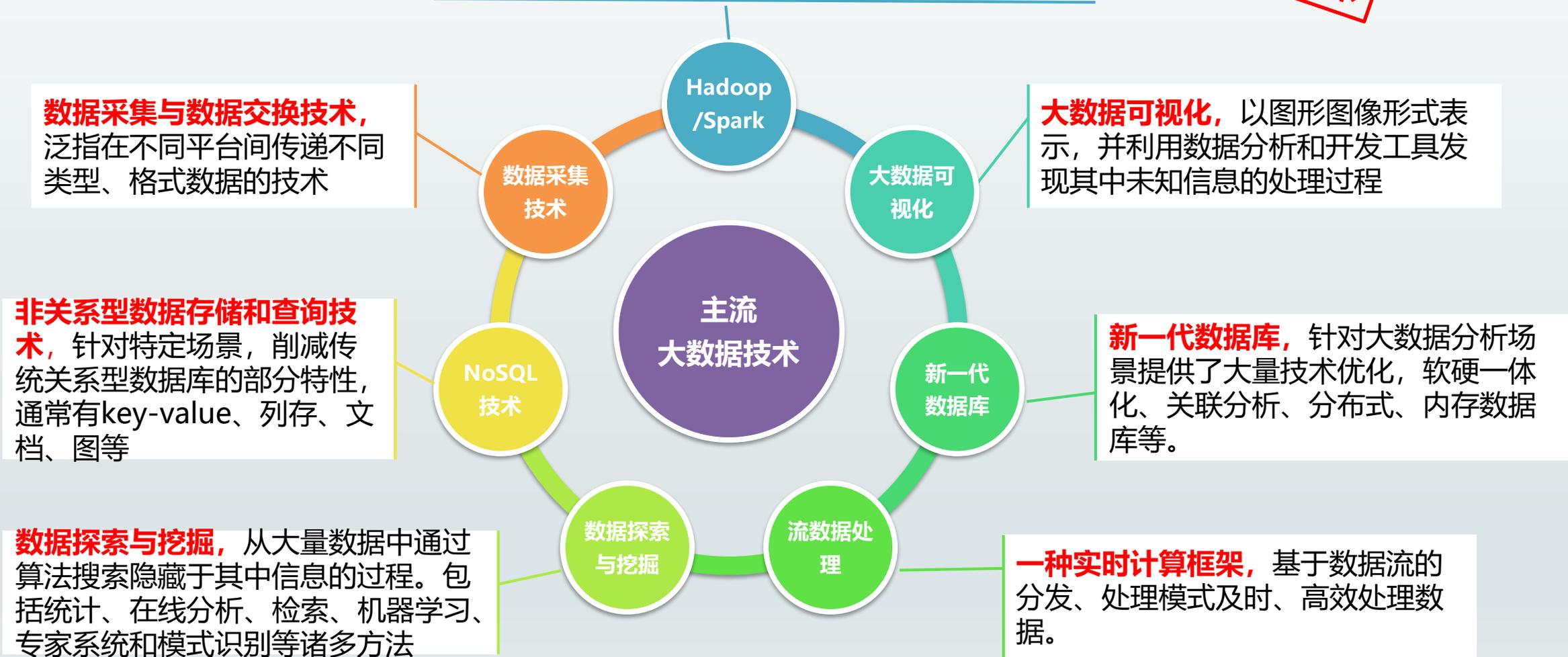
- 对业务数据有深刻的理解，以便能准确地应用正确的方法，解决于正确的数据和数据源的问题
- 拥有“发现问题”的能力，从而能基于数据进行分析和挖掘
- 行业与编程经验
- 行业工作经验对理解数据的意义和业务含义具有重要的价值

中国一汽 | in-Sight | ORACLE

大数据主流技术

分布式计算存储平台，基于 PC Server 集群部署，提供分布式数据存储、分布式计算框架。同时整个生态圈提供了大量外围组件满足各类应用场景需求。

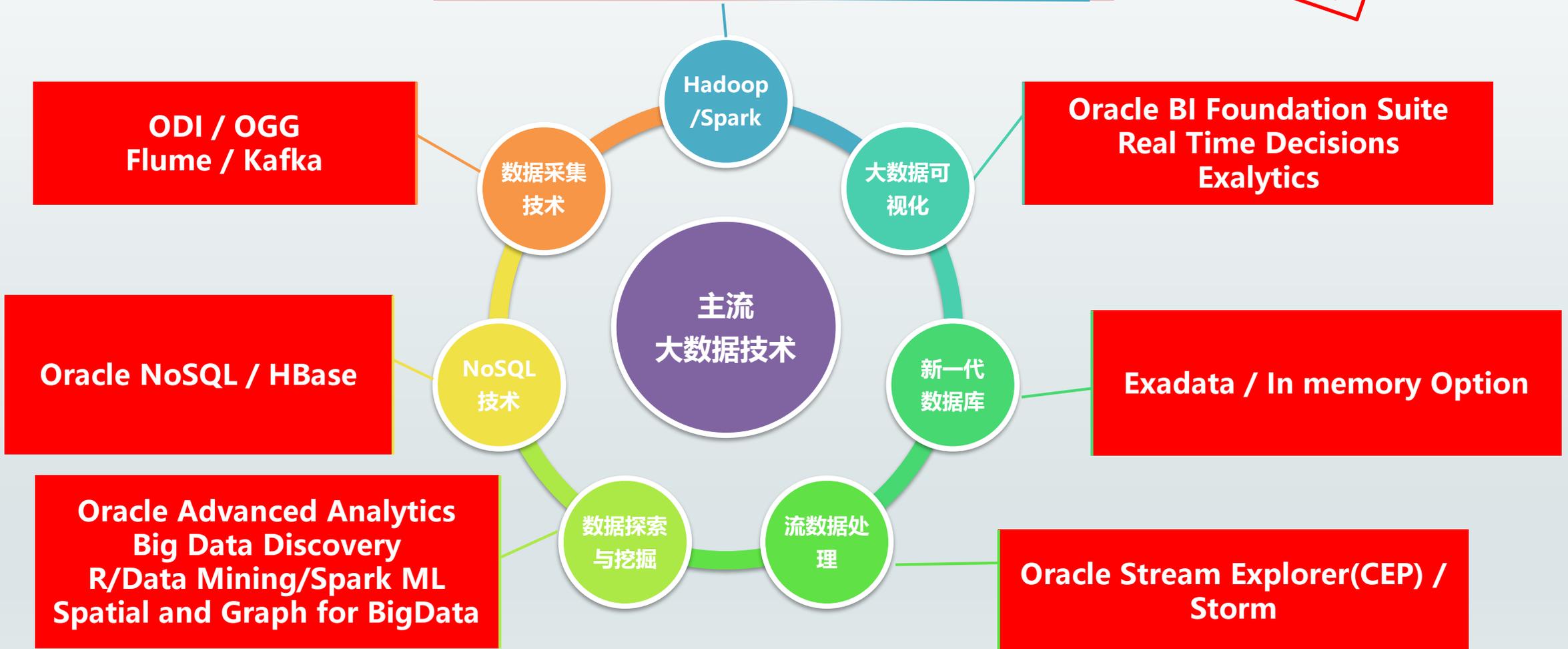
大数据七种技术



Oracle全覆盖大数据技术

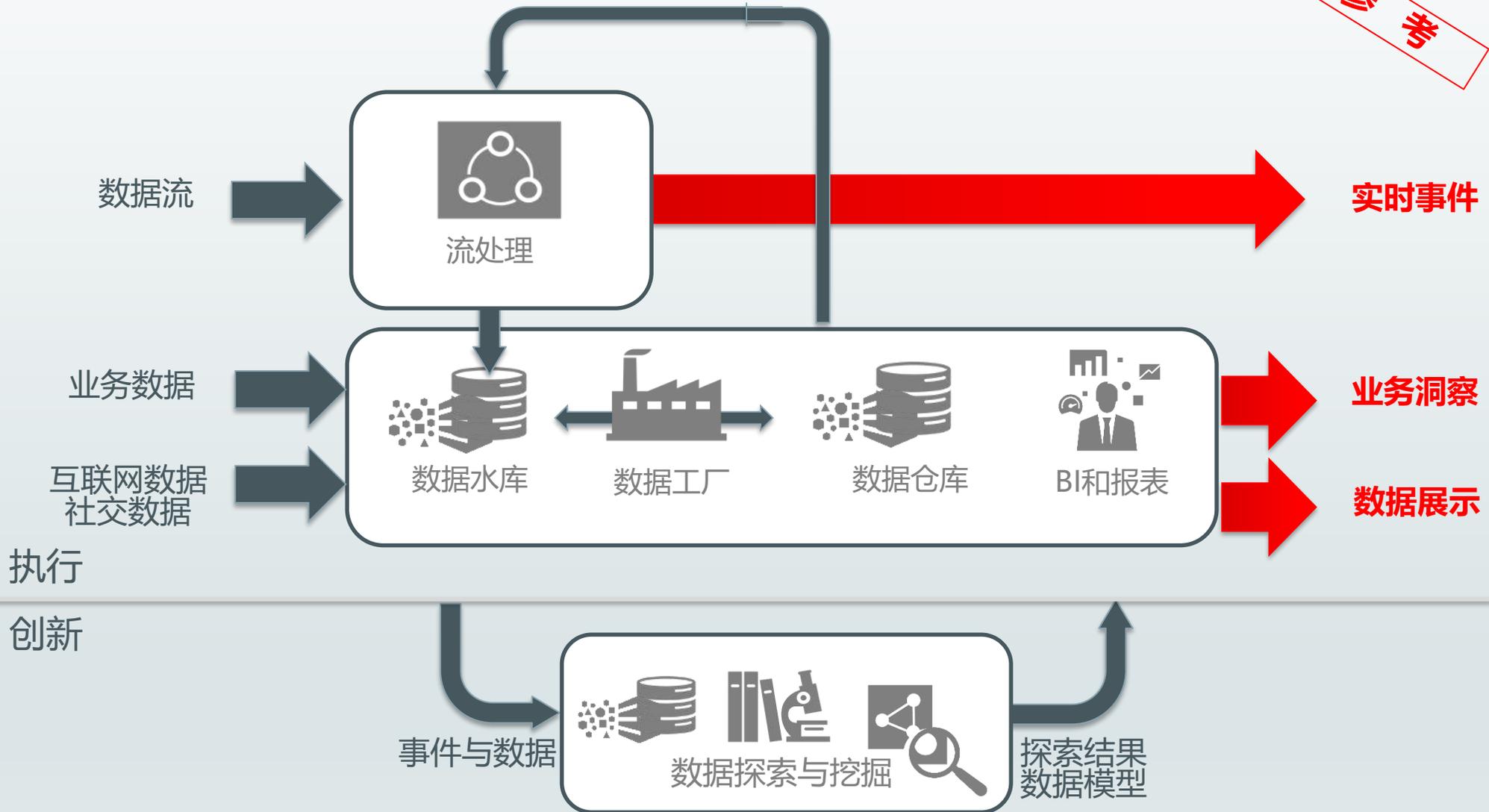
Oracle Big Data Appliance
Big Data Connector / Big Data SQL

大数据七种技术



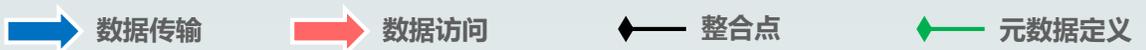
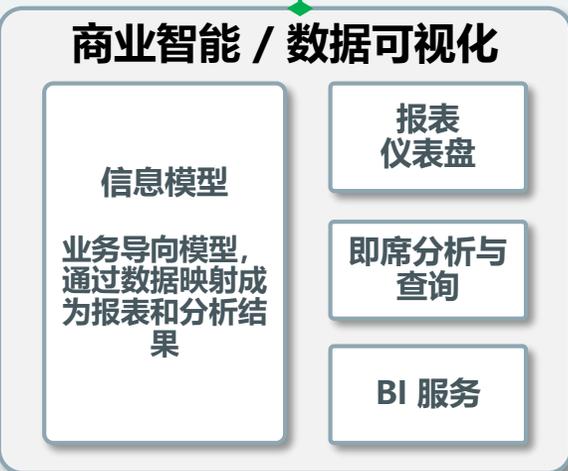
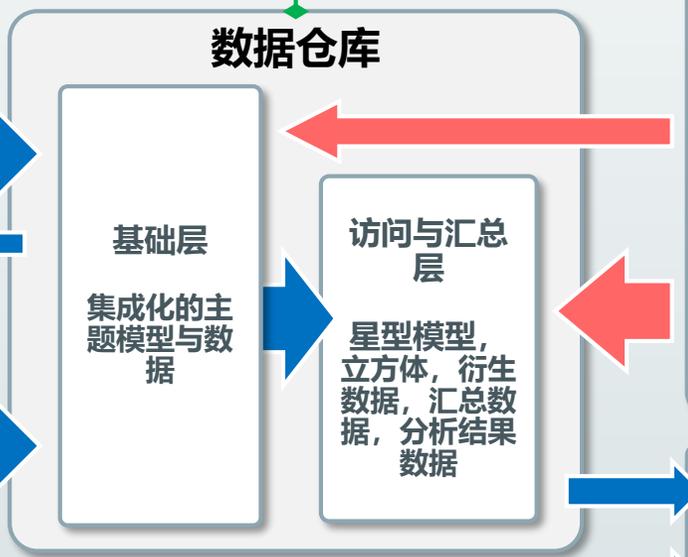
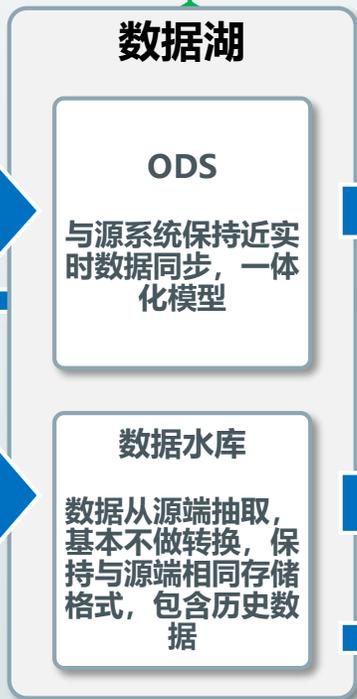
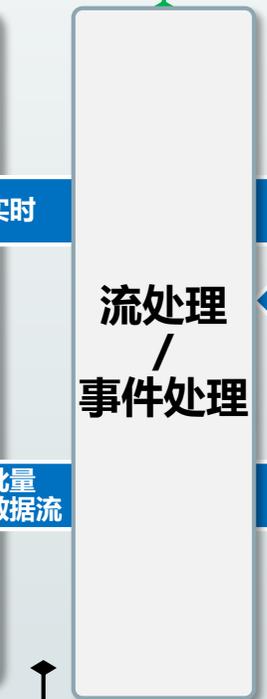
Oracle大数据参考架构

参考



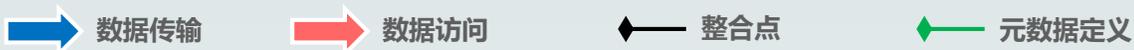
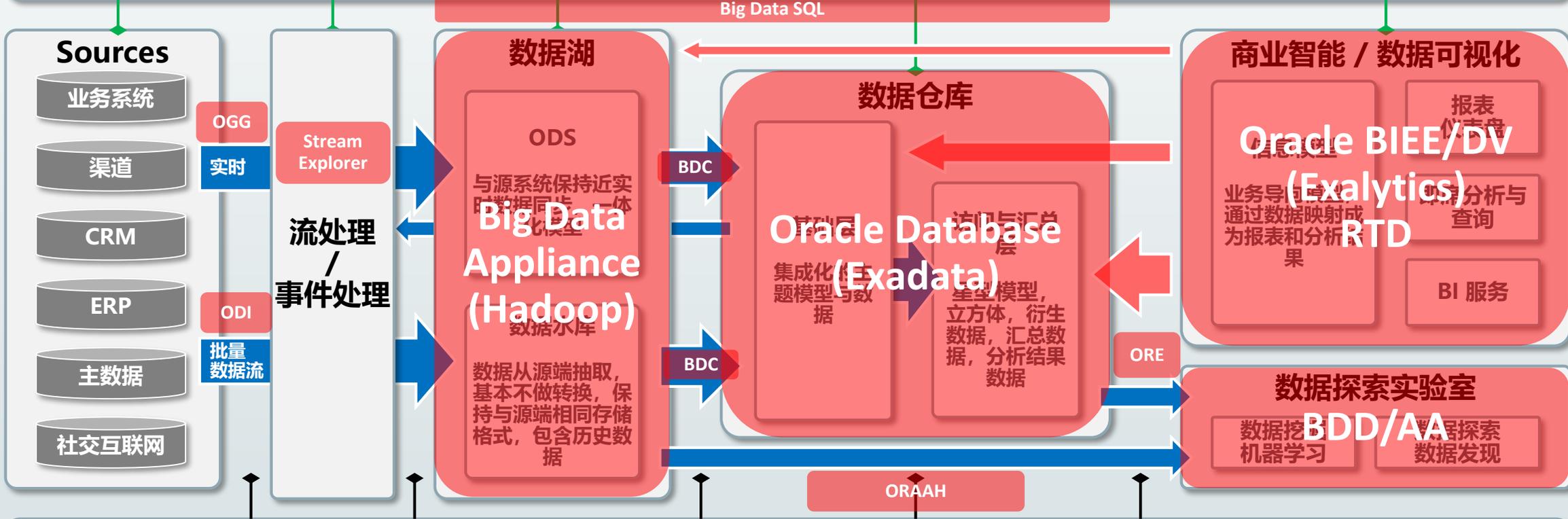
大数据通用架构

参考



大数据通用架构

参考



ORACLE®