

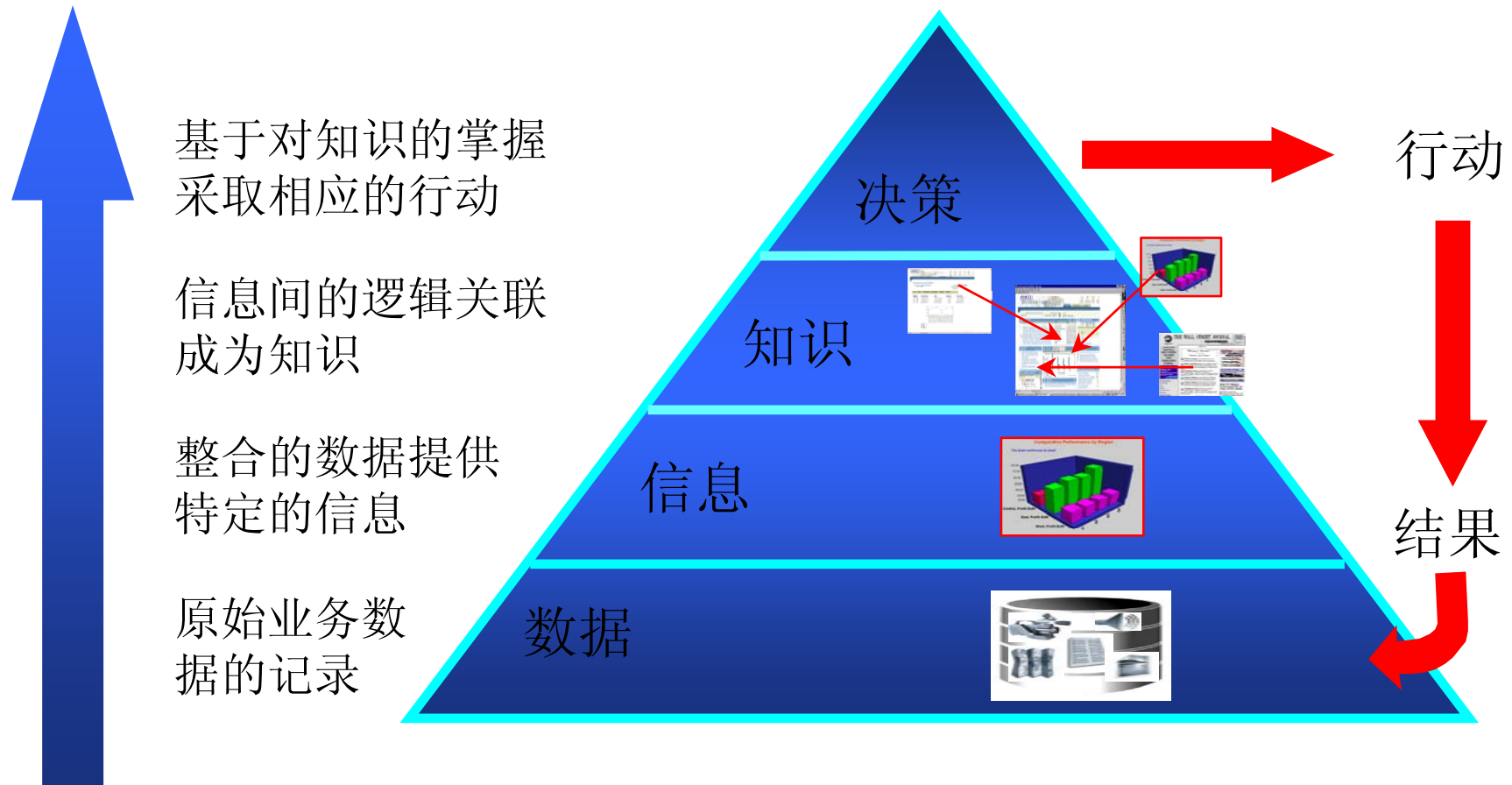


数据仓库方法论

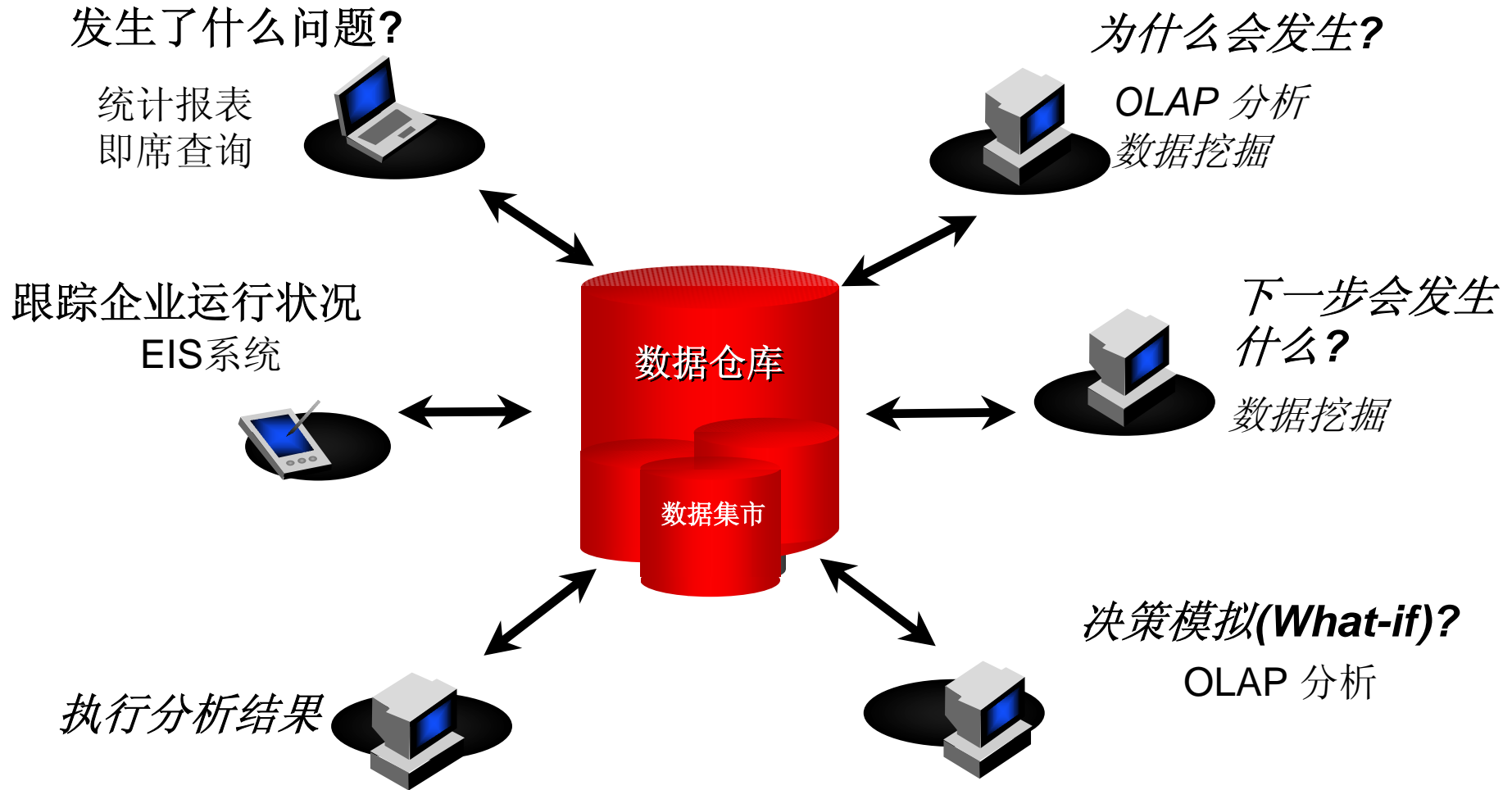
成功引导您构建数据仓库

ORACLE®
甲骨文

数据仓库系统的建设目标—从数据到知识



完整的决策分析过程



数据仓库项目失败的标志

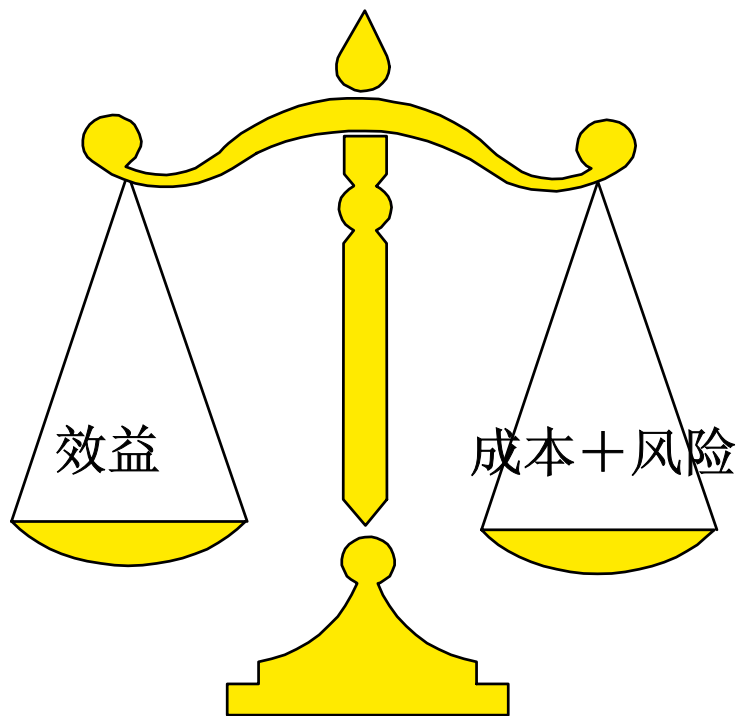
- 业务人员日常工作不依赖于数据仓库
- 业务人员不信任数据仓库
- 项目周期延长
- 费用超出预算

数据仓库项目普遍存在的问题

- 平台问题：
 - 高性能，可扩展
 - 稳定可靠
- 数据问题：
 - 数据整合
 - 数据质量
- 应用问题：
 - 产品易用性
 - 分析功能—提供决策支持能力

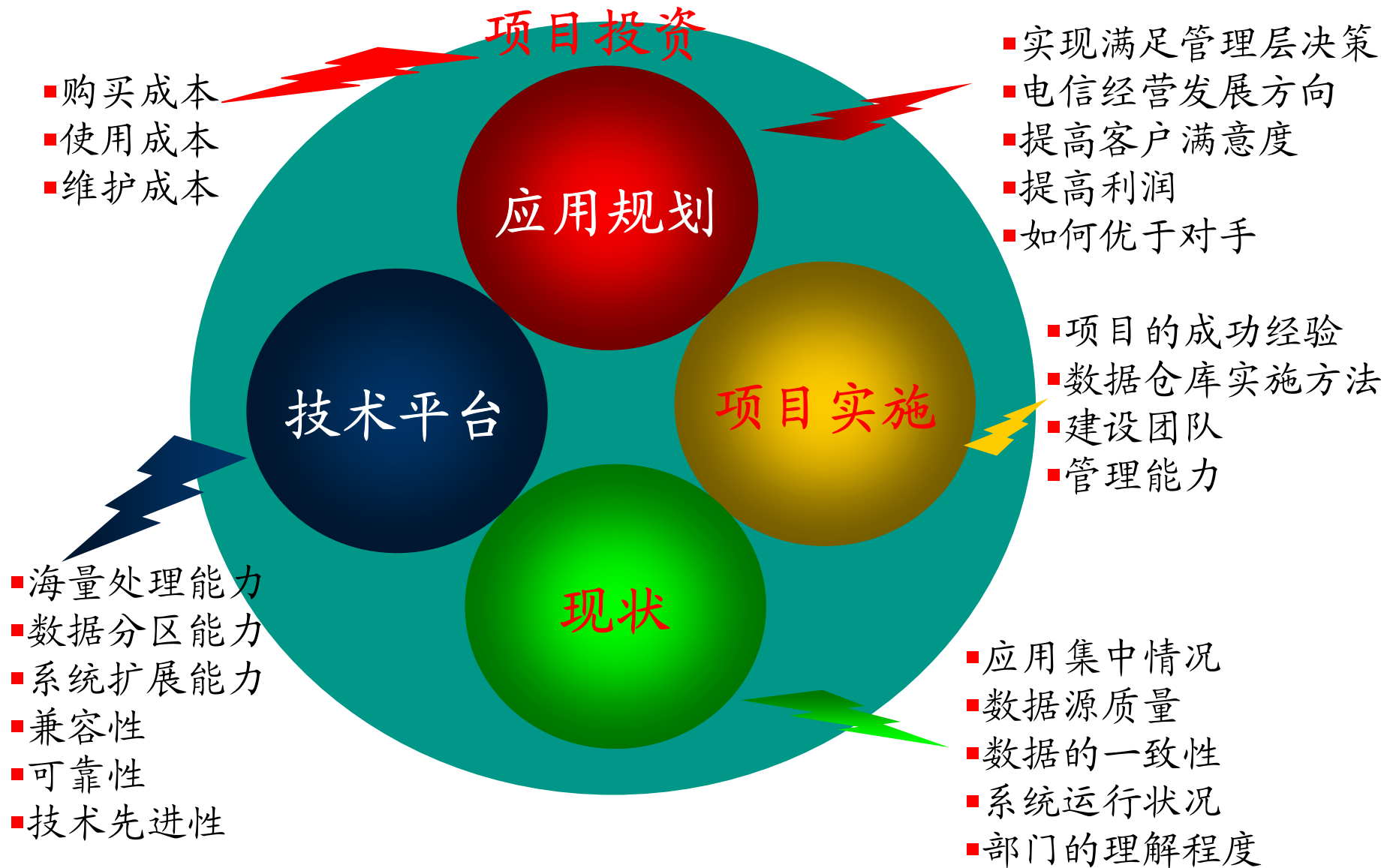
数据仓库项目成功的关键因素

在最短时间内看到数据仓库的效益



- 效益
 - 降低客户流失率
 - 增加收入
 - 提高客户满意度
- 成本
 - 软硬件投资
 - 管理维护成本
- 风险
 - 技术风险
 - 业务风险

数据仓库项目规划考虑的因素



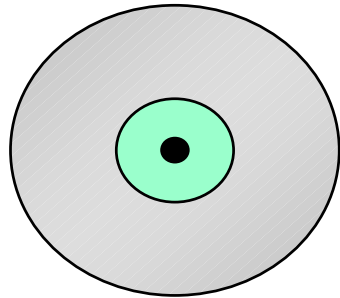
数据仓库建设整体思路

- 整体规划，分步实施
 - 整体规划
 - 设立分阶段目标，逐步实施
- 需求驱动
- 注重分析型结果和操作型应用的交互，形成闭环操作流程

DWM—数据仓库实施方法

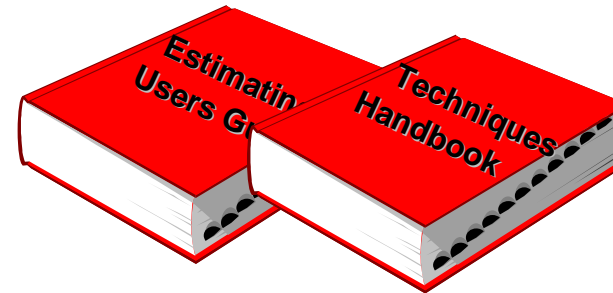
- Oracle数据仓库实施方法论(DWM-- Data Warehouse Method)
 - 多年数据仓库系统实施的提炼;
 - 预定义好数据仓库系统不可缺少的步骤和任务;
 - 确定正确的系统范围和用户需求、建立灵活的系统架构;
 - 数据仓库系统的实施分为13个过程，7个阶段。

数据仓库方法论（DWM）的组成



软件工具

- 方法指定原则
- 评估模型
- 工作计划生成器
- 项目框架



文档

- 技术手册
- 用户指南

Professional Community

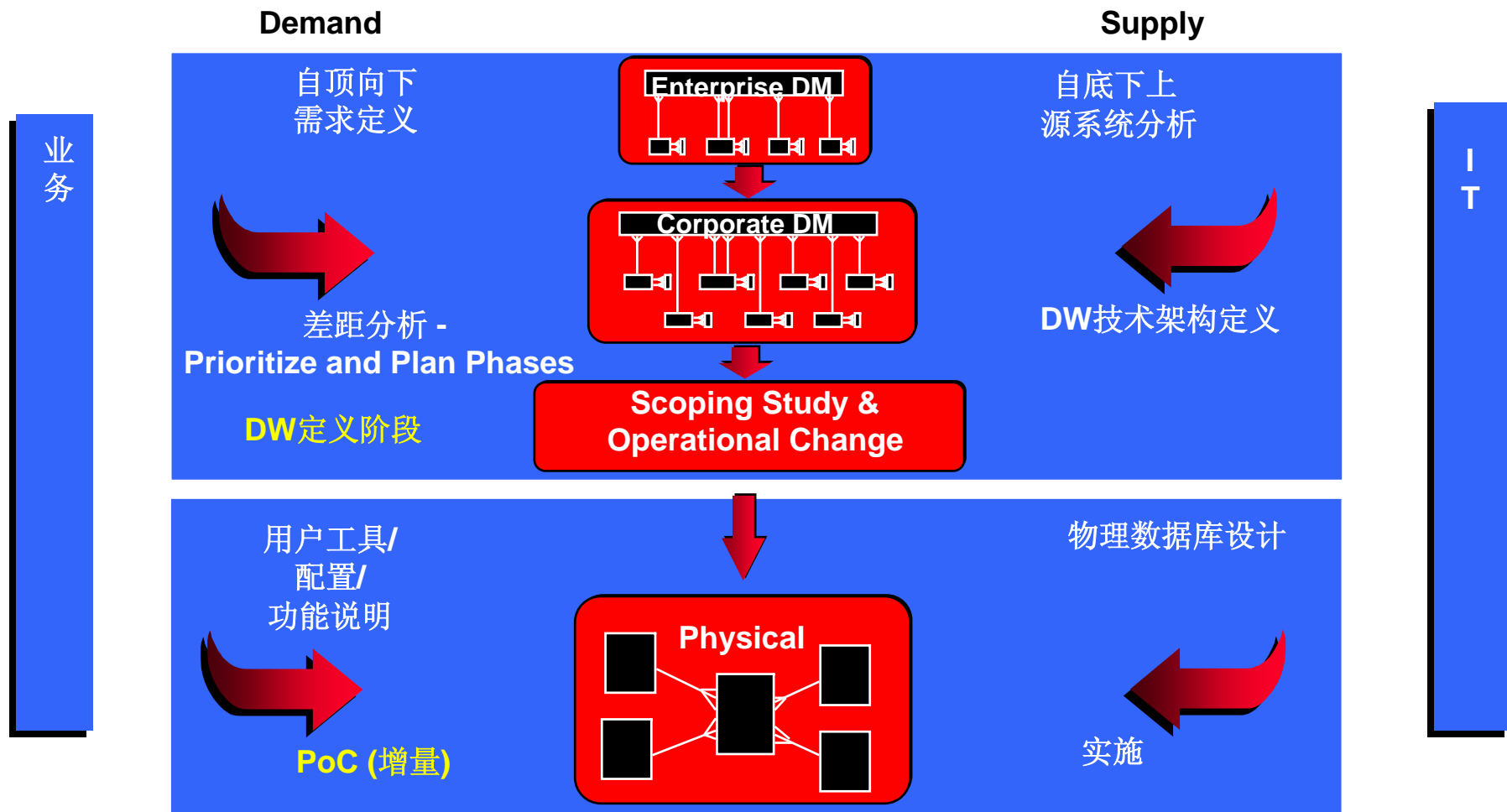
Oracle数据仓库实施方法论

过程 阶段	实施策略	系统定义	系统分析	系统设计	系统建立	系统应用	系统维护
业务需求定义	34.6%	9.3%	13.3%				
数据获取	8.4%	8.5%	23.1%	16.4%	17.2%	21.4%	
系统结构定义	11.5%	22.2%	14.3%	5.4%	14.5%		
数据质量控制	2.2%	12.1%	6.5%	7.9%	0.6%		
数据仓库管理	3.4%	4.4%	3.9%	11%	16.1%		
元数据管理	3.7%			4.7%	4.9%		
数据访问	6.3%			4.7%	4.9%		
数据库设计与建立				4%	2%		
文档设置	1%	1.1%	2%	4.5%	3.9%		
系统测试	1.2%		7.4%	15%	19.8%	19.7%	
培训	0.5%	7.7%	2.7%	4.1%	4.7%	27.8%	
系统上线			1.7%	0.2%	0.4%	17.8%	
技术支持							45%

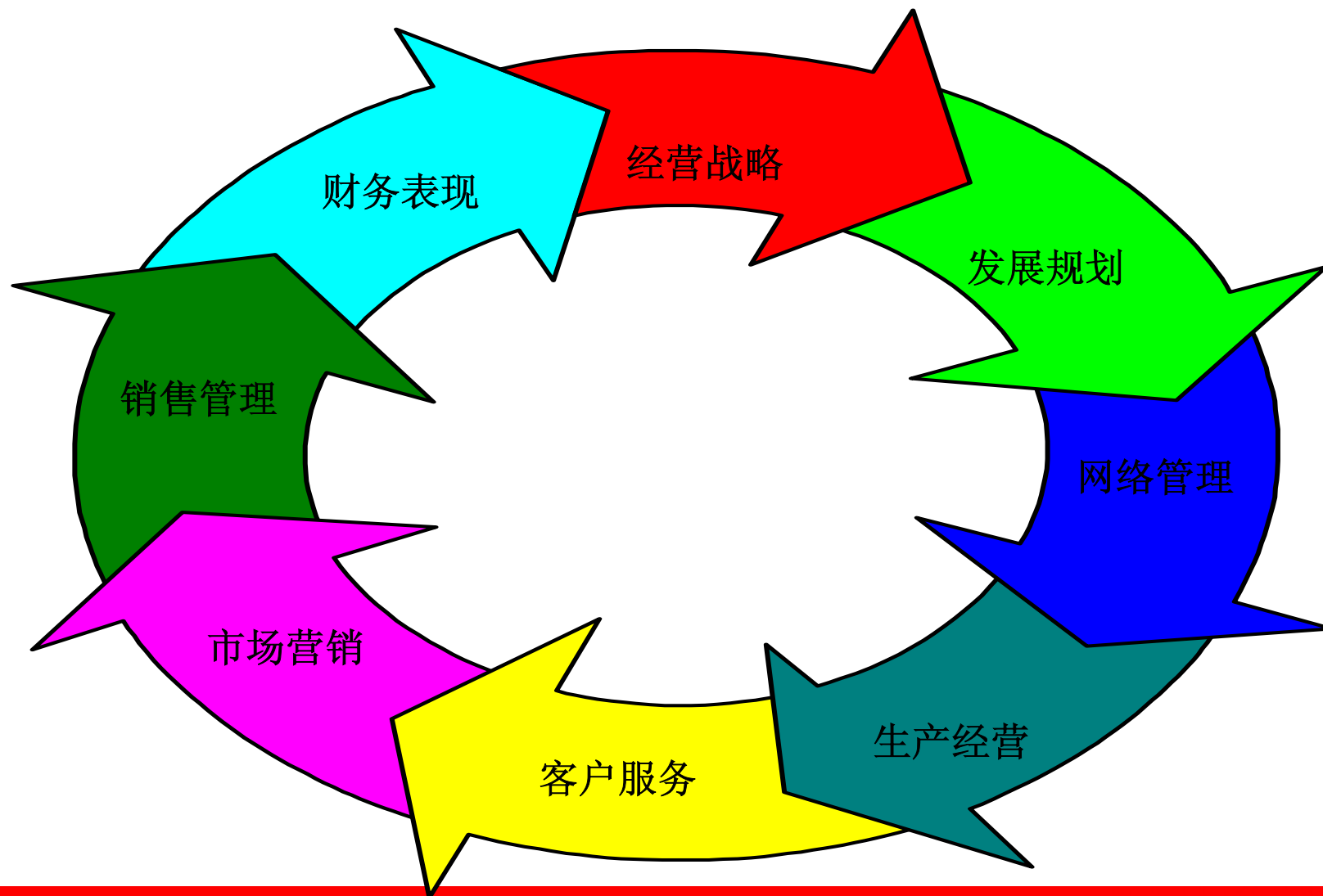
Oracle数据仓库方法论的提交物

阶段	No.	主要任务	提交成果	完成时间
系统准备、	1.	项目计划、项目管理	项目计划、报告	
项目管理	2.	测试环境产品安装与调试	系统安装文档	
系统定义设计阶段	1.	系统实现方案	系统方案文档	
		✓ 详细调研		
		✓ 分析模型的定义		
		✓ 维值、维值层次和维值编码规则的定义		
		✓ 分析报表的格式定义		
		✓ 扩展指标的定义		
		✓ 用户权限的定义		
	2.	数据接口方案	数据接口文档	
		✓ 接口格式的定义		
✓ 数据抽取、转换和装载放、方案				

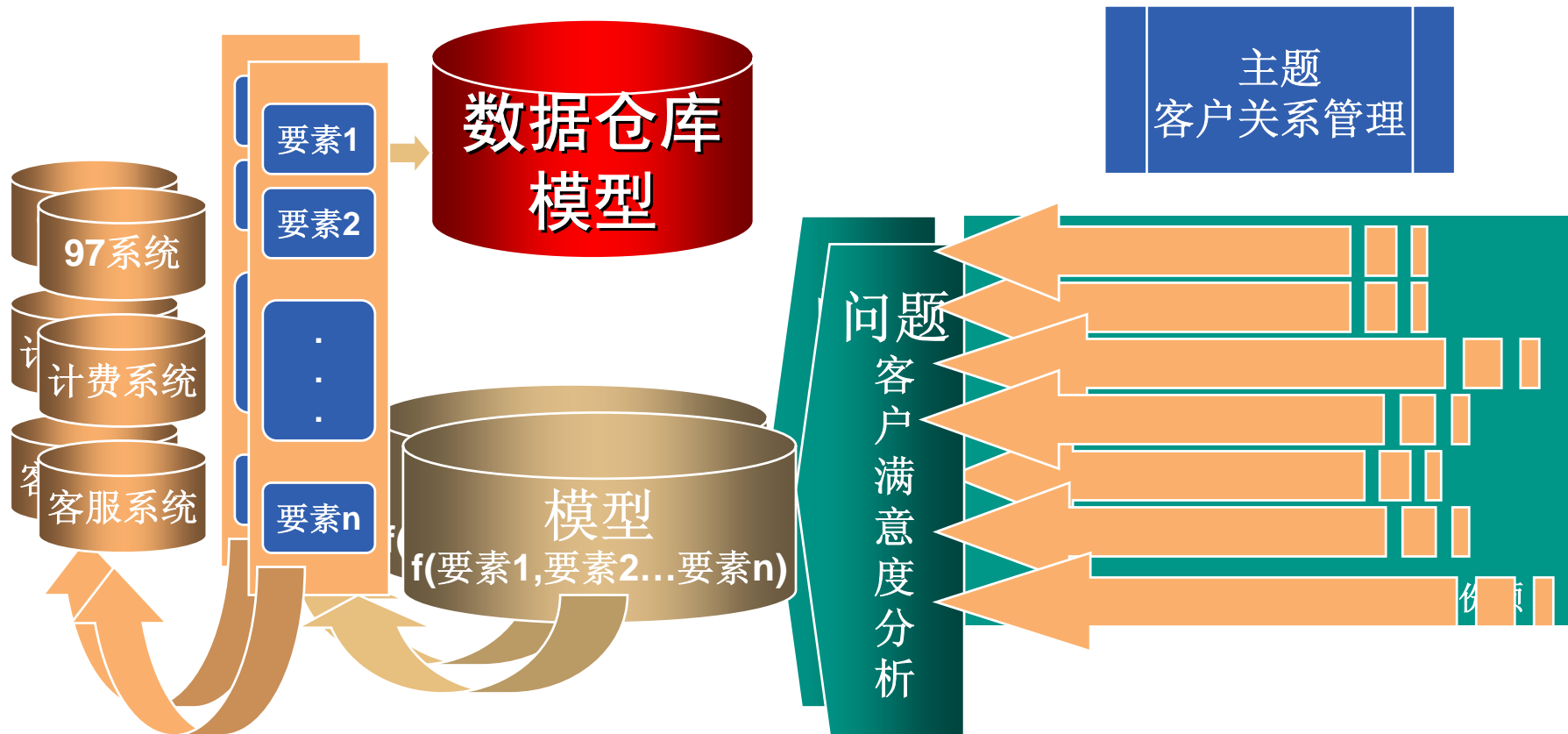
数据仓库系统设计和开发方法



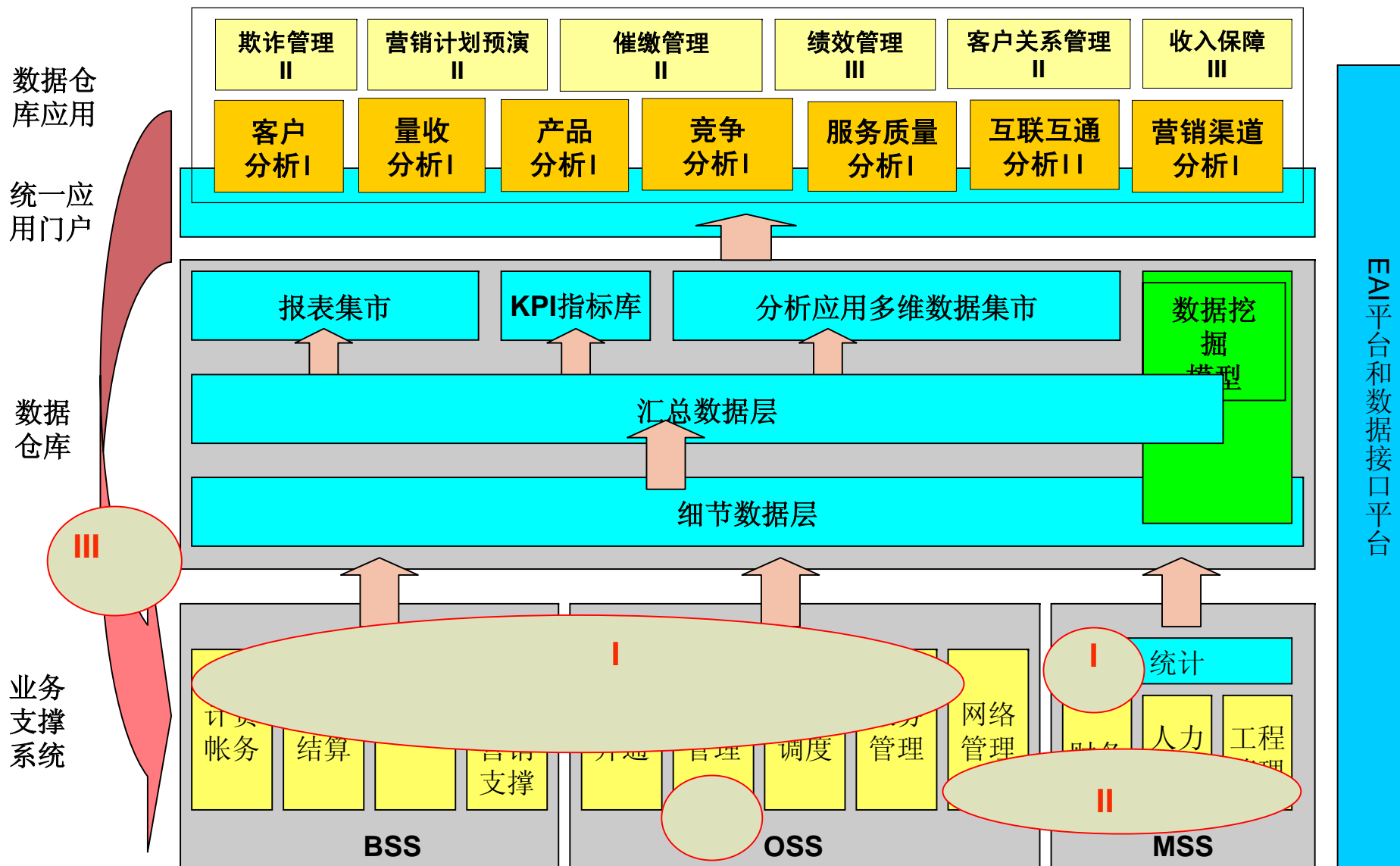
整体规划，分步实施



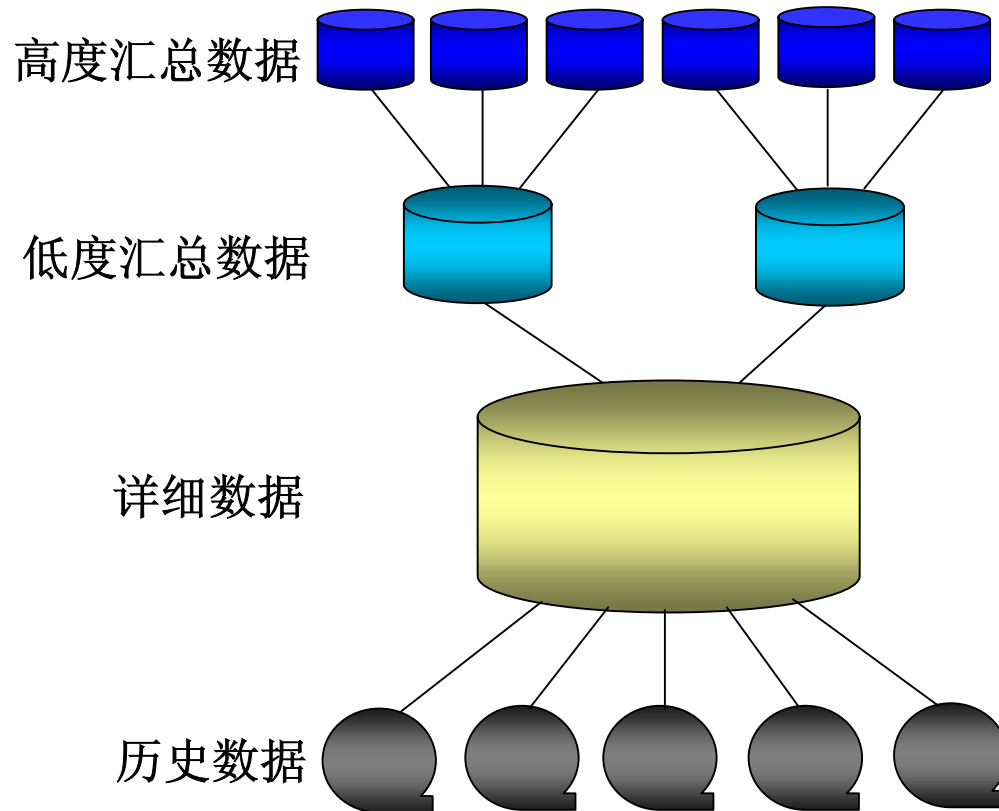
应用驱动的建设模式



规划系统应用架构



规划系统数据架构



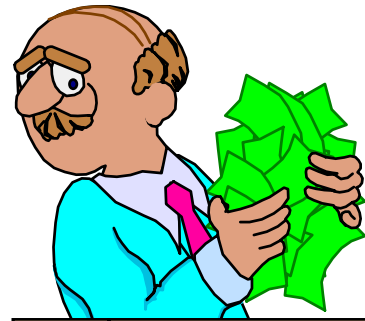
企业绩效分析

- 固定报表
- 定制报表

即席查询

数据挖掘

确定数据仓库应用的优先级

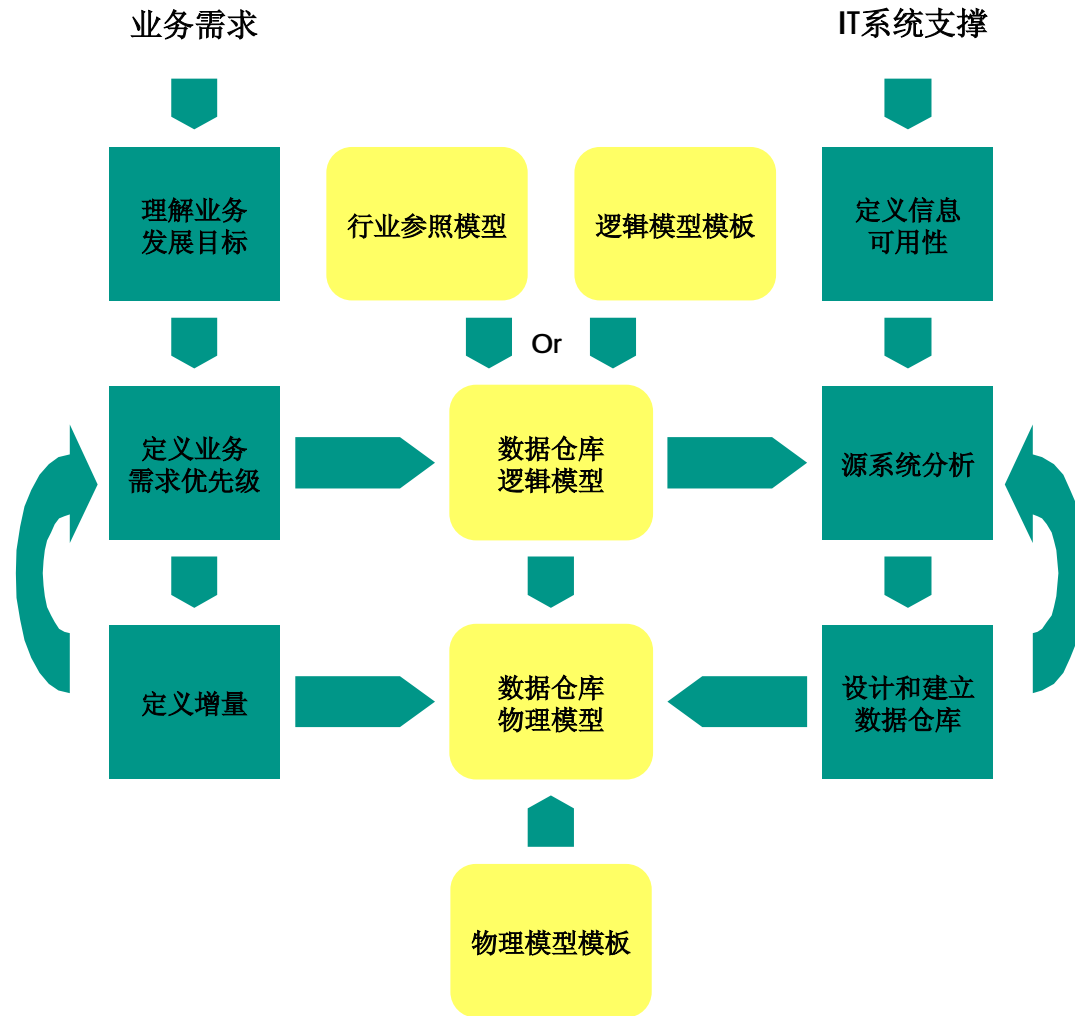


		效 益	
		低	高
成本和风险	高	不考虑	下次考虑
	低	没有价值	<u>目标应用</u>

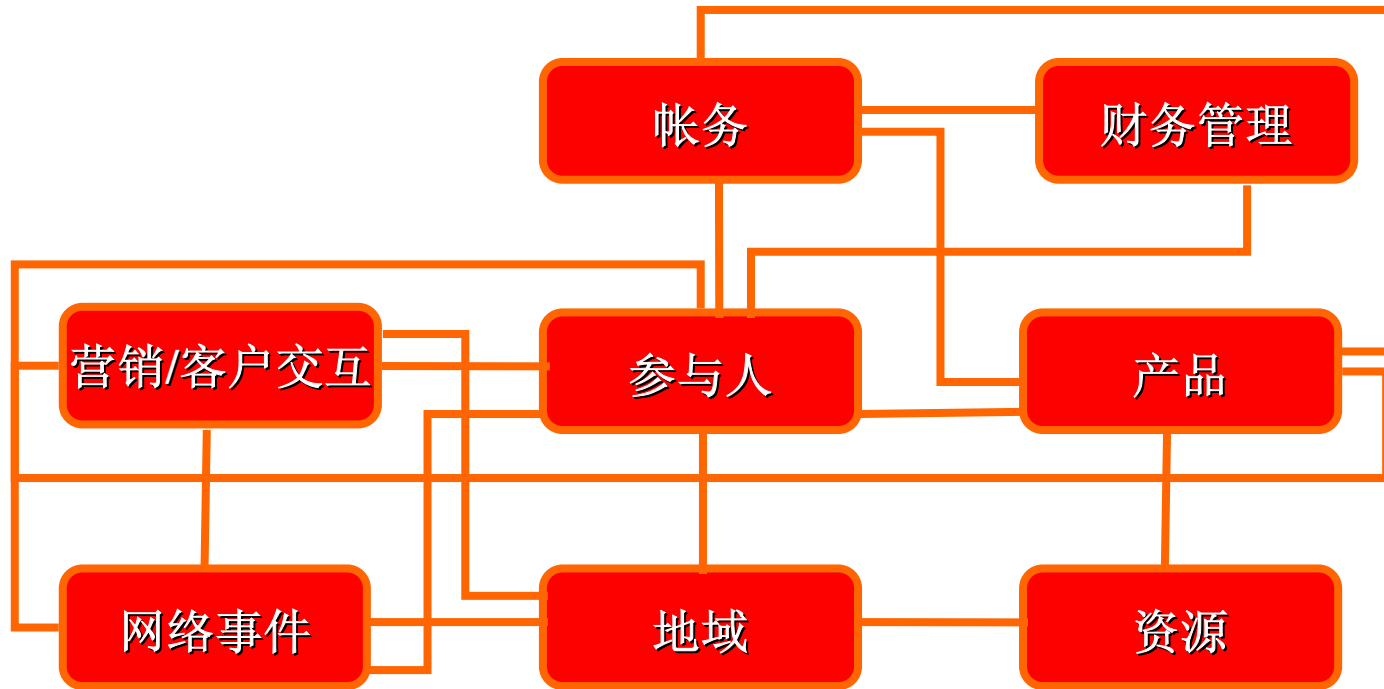
数据模型的作用

- 统一企业的数据视图（业务概念和关系）
- 定义业务部门对于信息的需求
- 是建立数据仓库原子层的基础
- 支持数据仓库的发展规划
- 初始化业务数据的归属

数据仓库建模方法

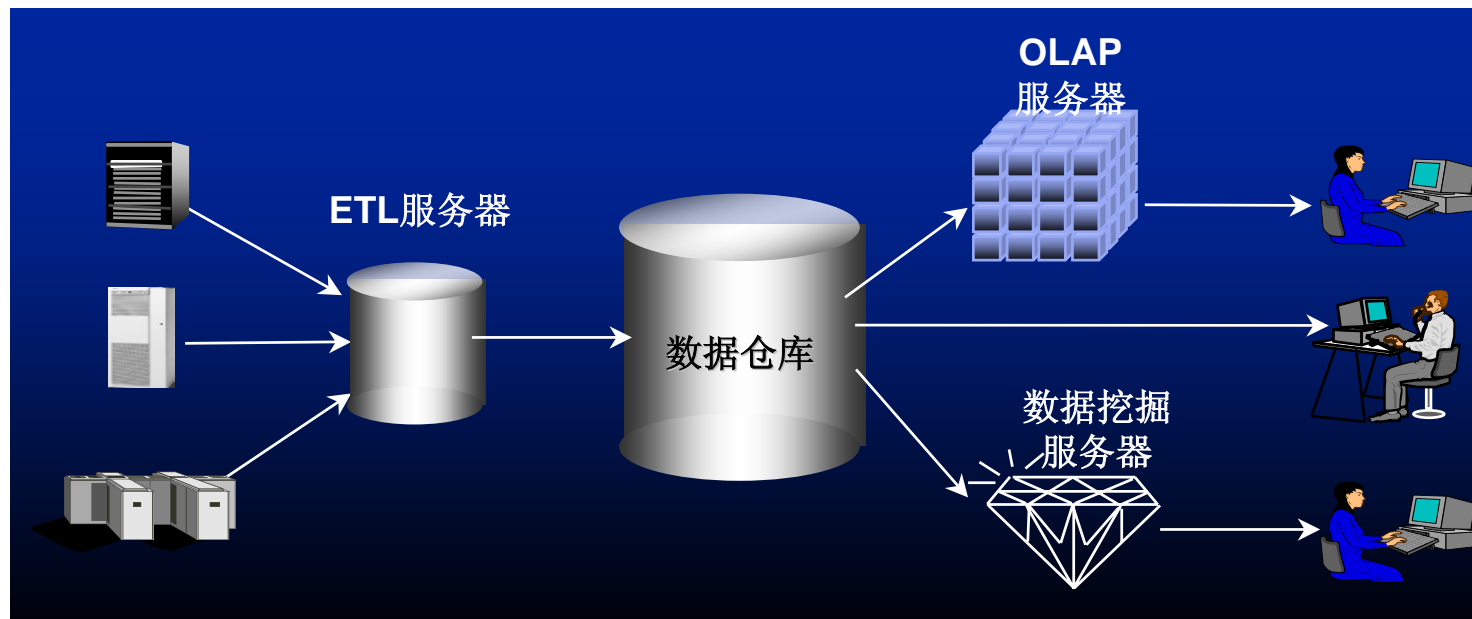


举例：Oracle 电信数据模型主题域



传统的数据仓库架构

—使用专门产品实现专门的功能

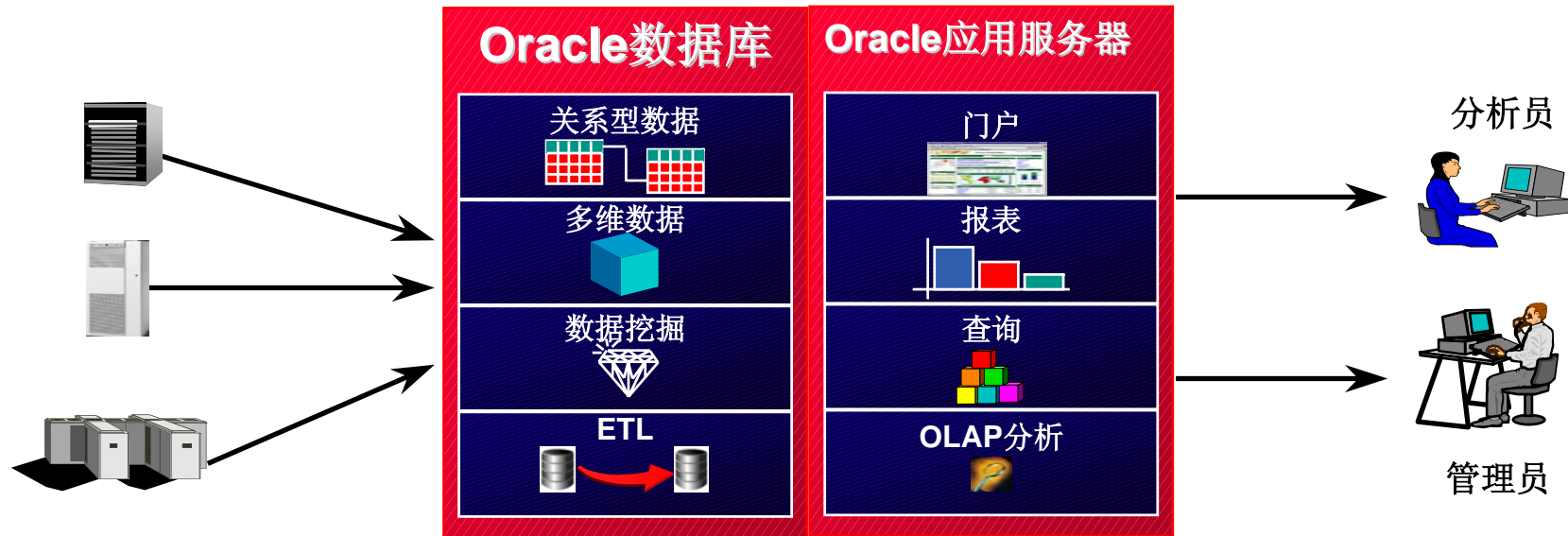


存在的问题：

- 系统可管理性差
- 接口复杂
- 系统灵活性差

Oracle 整合的 数据仓库架构

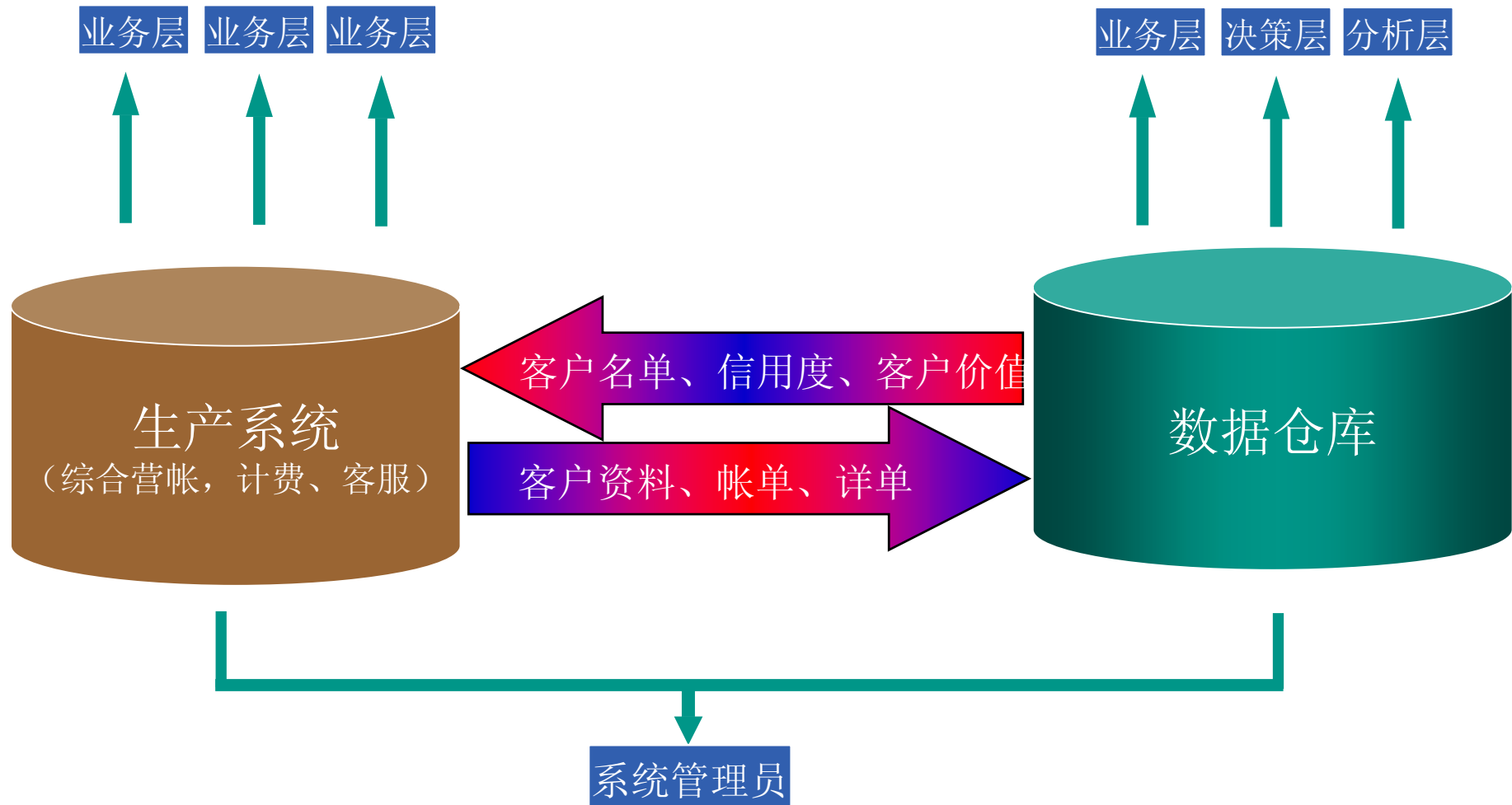
低成本，低风险



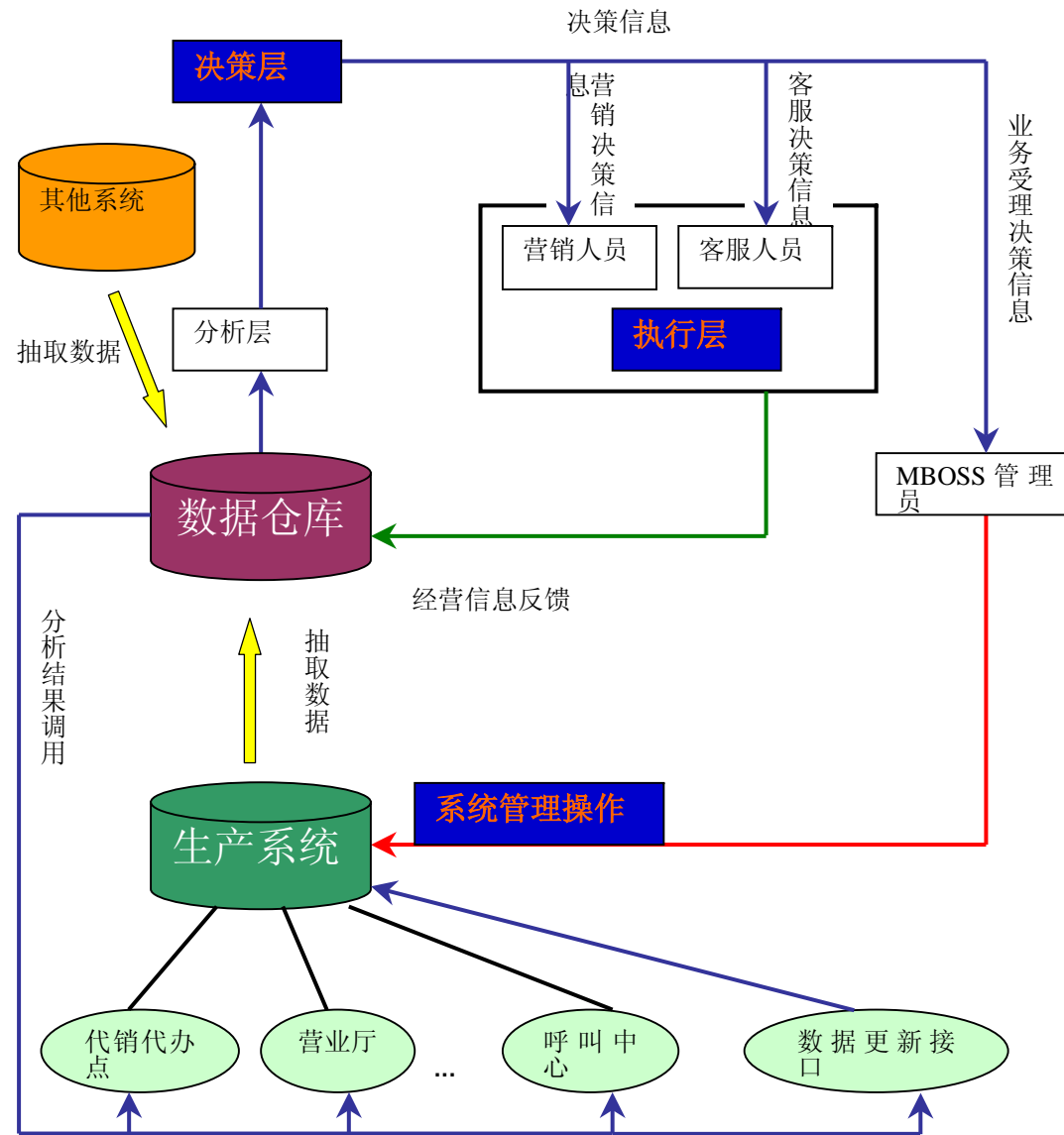
整合而不是集成

- 整合的 数据处理: ETL, OLAP, 数据挖掘
- 整合的 数据访问: 报表, 查询, 分析, 门户
- 整合的 元数据管理: 支持CWM 标准

数据仓库的直接使用和间接使用



对数据仓库中信息的使用方式



数据仓库建设思路:

- 1) 在分析层面上实现对业务和客户的分析, 可以帮助客户加深对运营状况的了解, 提高市场竞争能力;
- 2) 将数据仓库的分析结果反馈到生产系统, 实现分析型结果和操作型应用互动的闭环过程, 实现对业务运营的具体指导, 才能够真正发挥数据仓库的最大效能;
- 3) 基于数据仓库提供高端分析和咨询服务, 能够为运营商带来更高和持续的价值。

~~数据 + 元数据
=
信息
?~~

质量(数据 + 元数据)
=
信息

数据质量金字塔

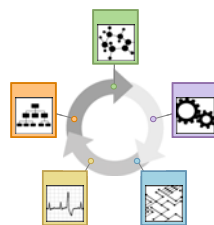


如何保证数据质量



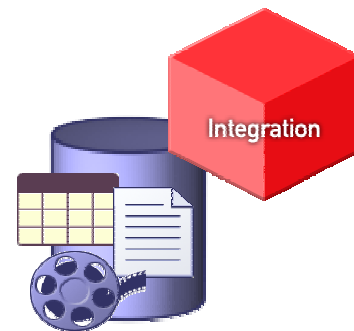
组织结构**People**

- 管理层对数据的认识和相关项目的支持
- 岗位职责
- 绩效考评



流程**Process**

- 建立贯穿全过程的流程管理体系
- 数据执行过程的监控
- 数据的审计追踪

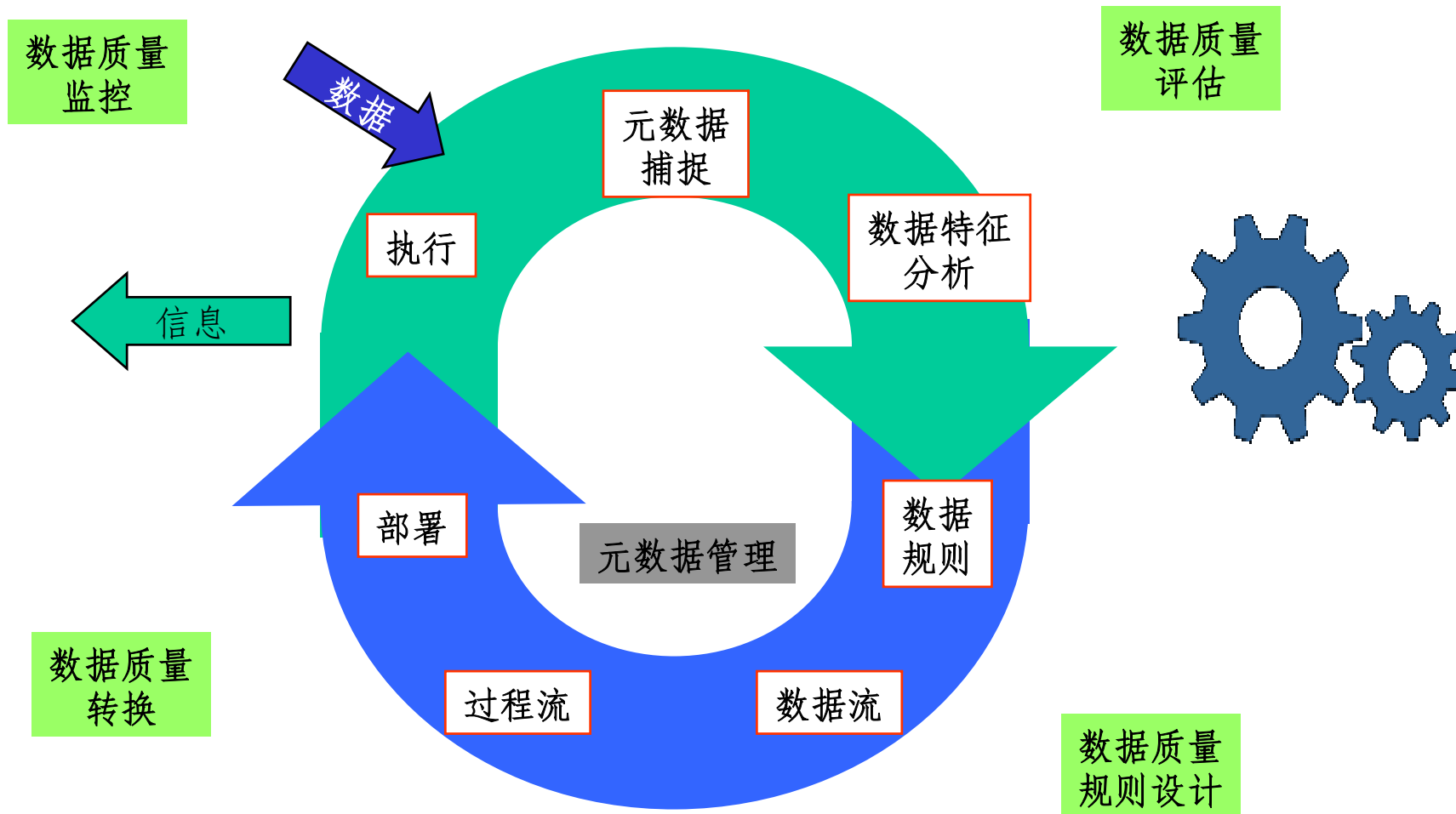


技术**Technology**

- 元数据-Meta data
- 模型设计的适应性和扩展性
- 数据转换过程中的保证

安全性和完整性

Oracle数据质量管理流程



OWB中的数据质量分析

- 数据类型验证
 - 数据类型（整数、浮点数、字符串等）是否正确
- 数据模式验证
 - 电话号码、身份证
- 唯一性验证
 - 重复的信息
- 数据有效性验证
 - 符合值域范围
- 一致性验证
 - 不同表之间、不同系统之间
- 完整性验证
 - 关键字段缺失

数据匹配/合并

- 源系统

97系统	姓名	联系人	地址	ID	客户性质
	工行济南分行	张晓东	经四路310号	123456789	C(金融)

10000	客户名称	联络人	地址	合同号	客户分类
	工商银行济南分行	张主任	槐荫区经四路310号	JNDKH-000558	12(商业大客户)

- 匹配/合并结果

数据仓库	客户	联系人	详细地址	ID	客户类别
	工商银行济南分行	张晓东	槐荫区经四路310号	123456789	12(商业大客户)

数据仓库项目成功关键因素

数据仓库项目成功的要素—应用方面

- 以应用为核心
 - 数据仓库不是目的，而是手段
- 明确系统的使用对象
 - 管理决策层
 - 专业分析员
 - 前台业务部门
- 提升管理水平
 - 不是统计报表系统
 - 最大限度发挥数据仓库的作用

数据仓库项目成功的要素—技术方面

- 信息集成是关键
 - 保证数据质量
 - 数据获取过程可靠，自动化
- 以服务用户为目标
 - 方便用户使用
- 提升系统的可管理性
 - 数据访问的性能
 - 系统的可靠性
 - 易于使用和维护

数据仓库项目成功的要素—实施方面

- 制订合理的项目目标
 - 必须与企业发展目标一致，才能获得领导层的支持
 - 项目目标必须明确而具体
 - 近期需求与长远规划相结合
- 沟通是关键
 - 业务部门和技术部门紧密合作
- 保持项目组成员的长期稳定

数据仓库项目成功的要素—推广方面

- 加强项目培训，提升员工的计算机应用水平和决策分析水平
- 建立系统推广配套体系与奖惩制度

Q&A

ORACLE®

甲骨文