



DTCC

2016中国数据库技术大会

DATABASE TECHNOLOGY CONFERENCE CHINA 2016

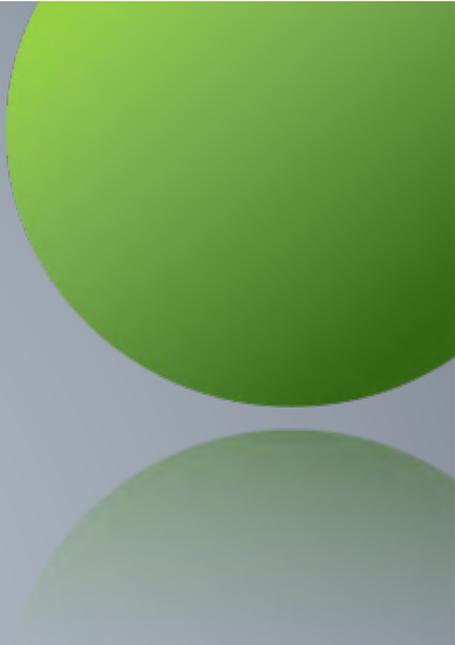
数据定义未来

SequeMedia
数据工场

IT168

ChinaUnix

ITPLUS



数据治理之

数据模型管控方案

演讲人：郑保卫 博士

EN·CORE

恩核（北京）信息技术有限公司

EN-CORE as a group of data experts provides a variety of services, including consulting, training, solutions, technical information is based on data from the planning of the company's IT system to the analysis, design, development, implementation, management, and governance.

个人介绍



- 郑保卫，工学博士，恩核（北京）信息技术有限公司创始人兼技术总监

- 出版书籍《海量数据库解决方案1》

- 《海量数据库解决方案2》及《数据架构实战》正在准备中

- 于2013年12月被北京市朝阳区认定为“凤凰计划”海外高层次人才

- 参与过大量关于数据架构、数据建模、数据分析等方面项目长期致力于数据架构及数据治理技术方面的研究和实践。

- 自2011年至2015年持续被数据库技术大会邀请为演讲嘉宾，自2012年至2014年持续被ORACLE嘉年华邀请为演讲嘉宾

- 荣获2015年中国大数据领域领军人物奖，由国家信息公共服务平台及国家软件公共公共服务平台颁发





CONTENTS

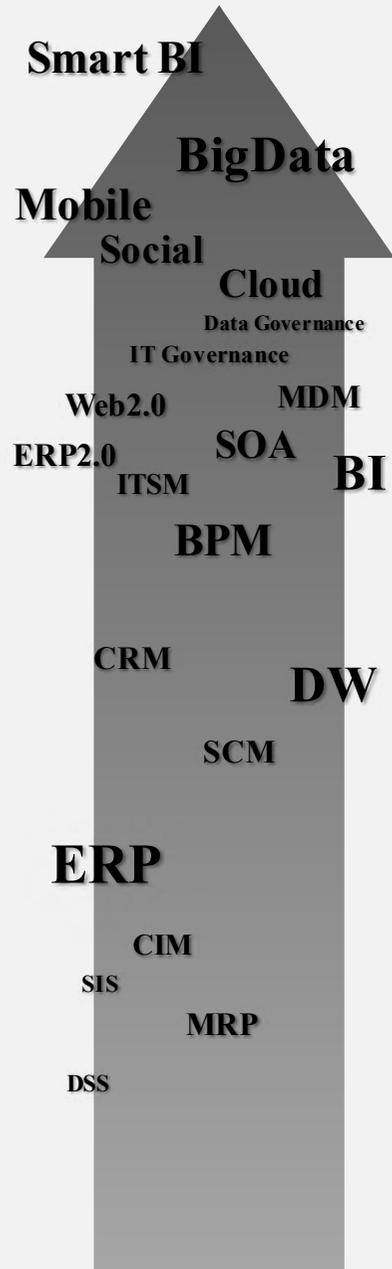
- I. 数据治理需求发展趋势
 - II. 数据治理成功关键要素
 - III. 数据模型管控方案
 - IV. 成功案例
-

A photograph of a building facade with a white sign. The sign features the text "Data-Driven World" in a smaller font above the large, bold text "EN·CORE". A small green dot is positioned between "EN" and "CORE".

Data-Driven World
EN·CORE

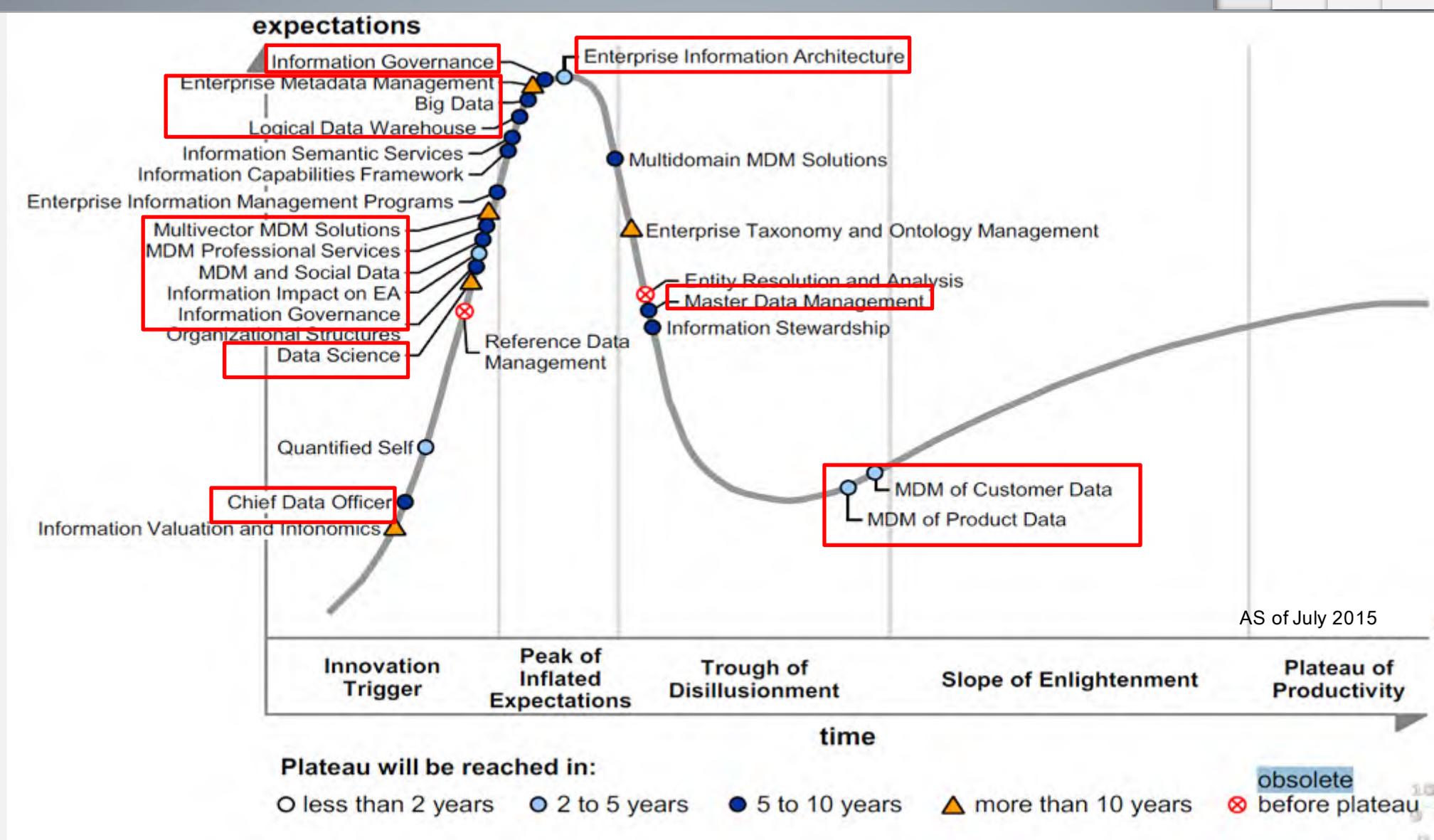
I. 数据治理发展趋势

I. 数据治理发展趋势



<p>4 2010~2020</p> <ul style="list-style-type: none"> ▪ 移动应用及互联网等应用扩大，数据量暴增，数据分析需求增加 ▪ 大数据处理技术增多 ▪ 企业数据质量要求提高 	<p>数据治理范围 应用 分析 管理 统计 整合 挖掘 存储</p>
<p>3 2000~2010</p> <ul style="list-style-type: none"> ▪ 00年初：系统数量增多，数据量暴增 ▪ 数据仓库建设开始(DW) ▪ 数据统计及简单分析 	<p>数据治理范围 应用 分析 管理 整合 存储</p>
<p>2 1990~2000</p> <ul style="list-style-type: none"> ▪ 90年代初：电子商务开始，大规模系统建设开始(ERP) ▪ 90年代末：综合业务支撑系统建设开始 	<p>数据治理范围 应用 整合 存储</p>
<p>1 1960~1990</p> <ul style="list-style-type: none"> ▪ 60年代：以简单的业务处理为目的 ▪ 70年代：以信息提供为目的 ▪ 80年代：以提升企业市场竞争力为目的 	<p>数据治理范围 应用 存储</p>

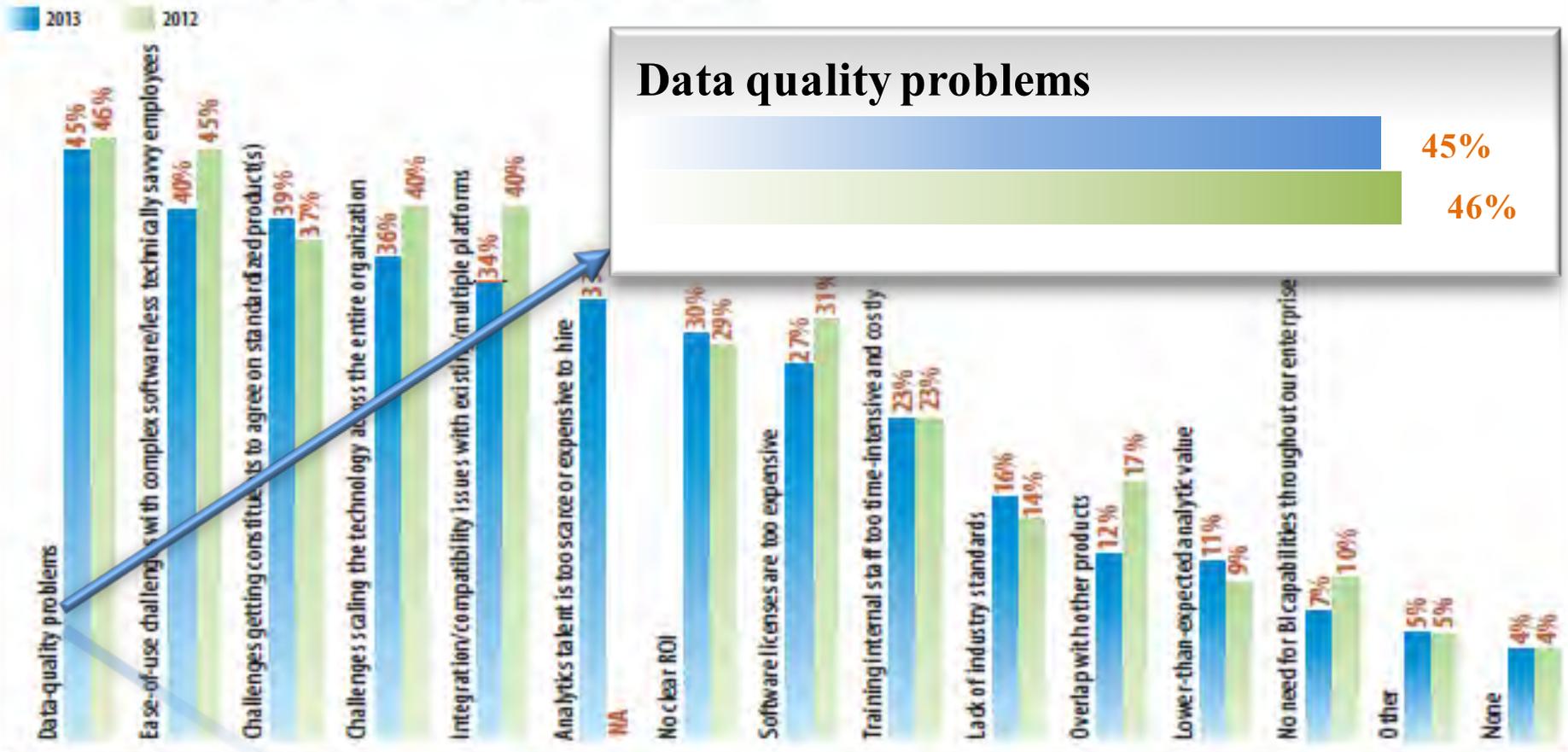
I. 数据治理发展趋势



I. 数据治理发展趋势

Barriers to Enterprise-Wide Analytics or BI Adoption

What are the barriers to adopting analytics or BI products enterprise-wide?



Note: Multiple responses allowed

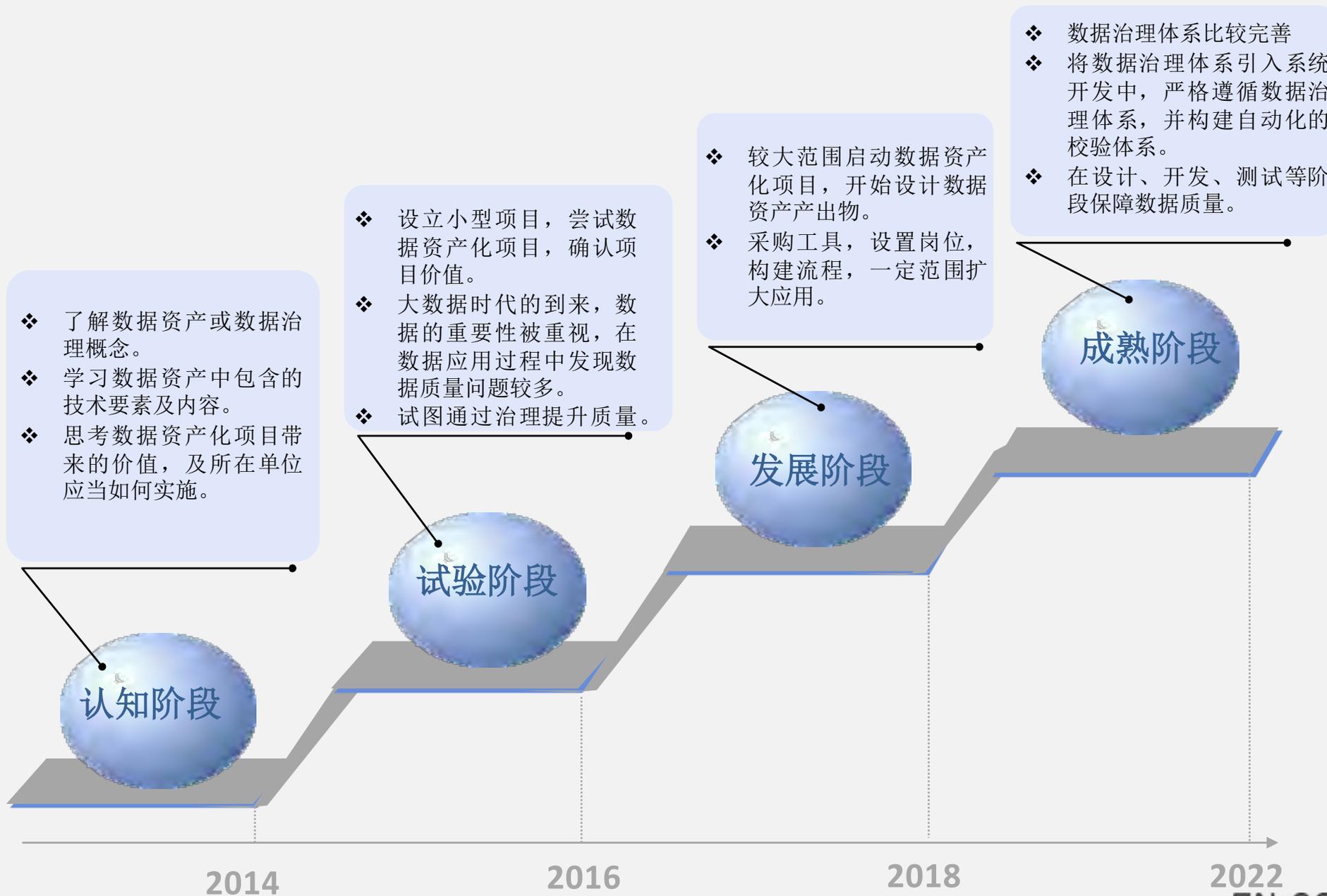
Base: 417 respondents in October 2012 and 414 in October 2011 at organizations using or planning to deploy data analytics, BI or statistical analysis software

Data: InformationWeek Analytics, Business Intelligence and Information Management Survey of business technology professionals

R06011127

I. 数据治理发展趋势

I II III IV



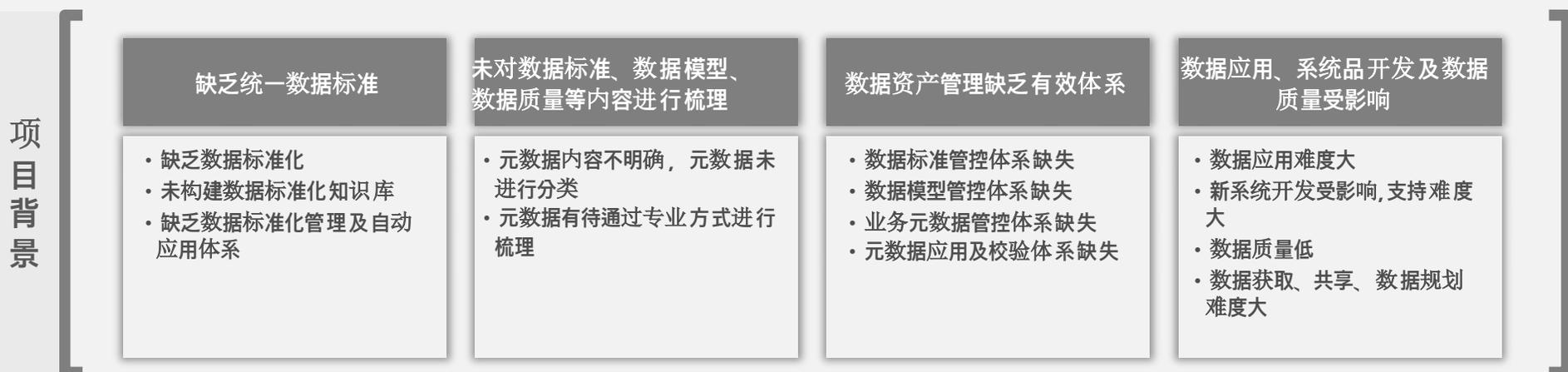
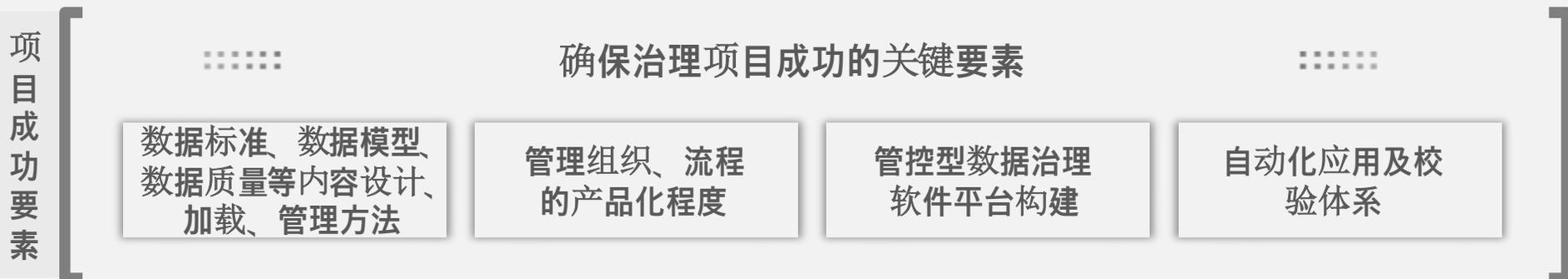
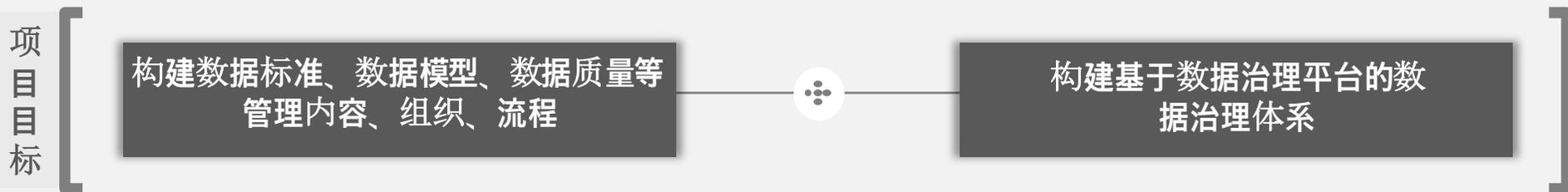
Data Driven World

EN·CORE

II. 数据治理成功关键要素

II. 数据治理之我见

为数据应用、数据管理、项目开发提供支持，提升数据获取、共享和数据规划能力



数据治理项目成功的核心要素



II. 数据治理成功的核心要素

方案1

实施人员必须具备丰富项目经验，提供可落地方案

成功案例

具有丰富的项目成功实施经验

项目名称	项目内容	效果
某央企	<ul style="list-style-type: none"> 元数据管理，数据模型管控，影响度分析，数据质量，血缘关系分析，信息资源目录管理。 	<ul style="list-style-type: none"> 构建了元数据知识库 实现了数据的标准化应用及管理 实现了对数据模型的管控 提升了数据质量 构建了企业级数据资产管控体系
某通信公司	<ul style="list-style-type: none"> 数据标准化，数据模型管控。 	
ETC电子收费系统	<ul style="list-style-type: none"> 元数据管理，数据标准化，数据模型管控，影响度分析。 	
国内10家以上，国外50家以上银行案例	<ul style="list-style-type: none"> 数据模型管控，数据标准化，元数据管理。 	

成功案例

数据标准化，数据模型，数据质量等共享统一知识库的专业化工具

数据标准和数据质量收集及分析，



管控和验证方案



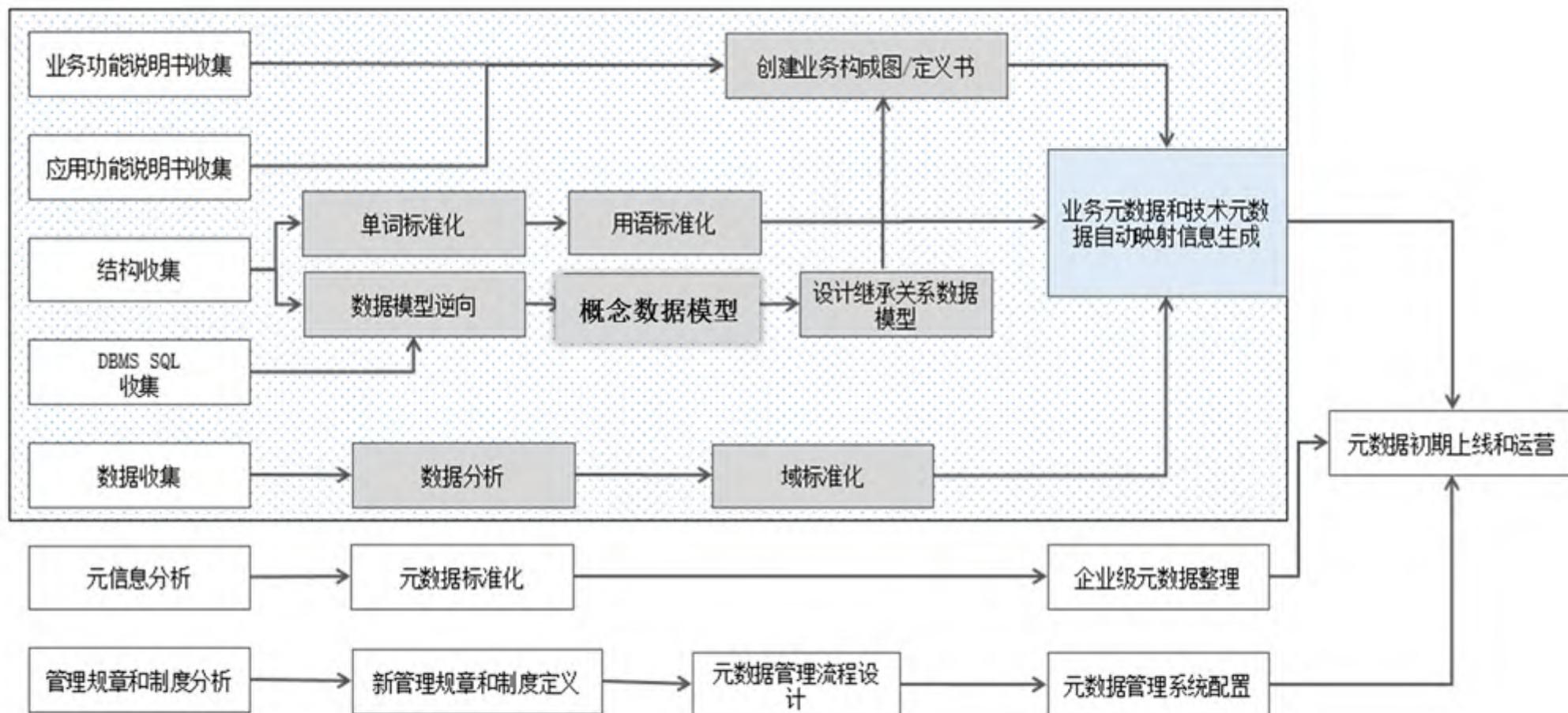
效果

- 通过使用了搜索引擎实现元数据收集自动化
- 基于专业的数据建模工具实现数据标准应用自动化
- 基于元数据工具和建模工具的统一应用，实现数据标准、数据模型等元数据校验自动化

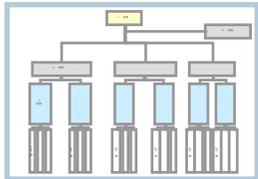
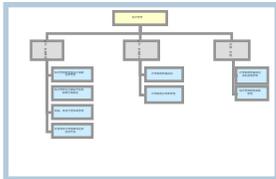
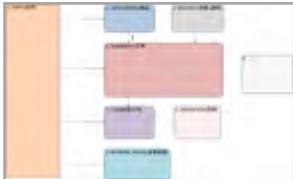
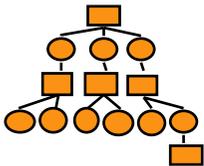
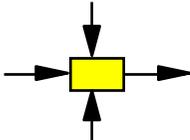
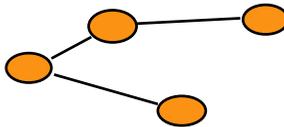
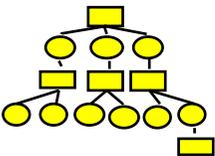
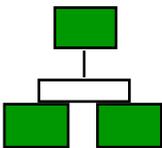
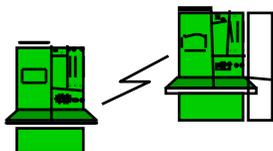
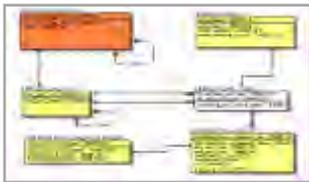
II. 数据治理成功的核心要素

I II III IV

措施2 提供基于数据架构的数据治理体系



II. 数据治理成功的核心要素

层级 \ 构成	业务架构	应用架构	数据架构	技术架构
概念层 <i>Planner</i>	 <p>企业整体业务架构</p>	 <p>企业整体应用架构</p>	 <p>继承关系模型</p>	 <p>技术整体模型</p>
	 <p>子模块业务架构</p>	 <p>子模块应用架构</p>	 <p>概念模型</p>	 <p>基础构造关系图</p>
逻辑层 <i>Designer</i>	 <p>详细业务流程</p>	 <p>详细开发流程</p>	 <p>逻辑模型</p>	 <p>基础构造设计图</p>
物理层 <i>Builder</i>	 <p>业务规划说明手册</p>	 <p>应用设计说明手册</p>	 <p>物理模型</p>	 <p>软硬件产品目录</p>

II. 数据治理成功的核心要素

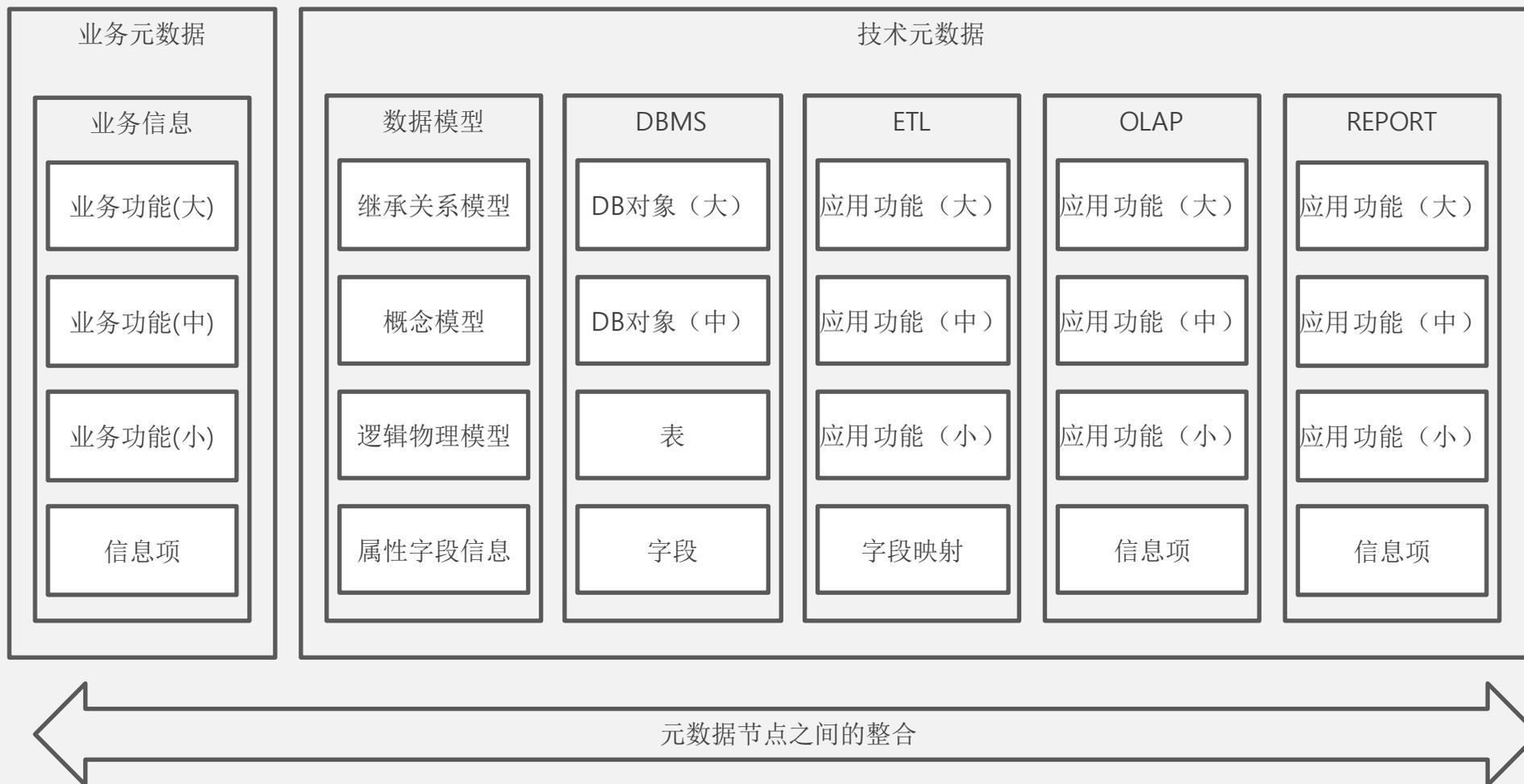
基于数据治理架构管理策略，构建元数据管理体系，其中包括组织、流程、元数据管理策略和内容。

综合数据治理体系架构



II. 数据治理成功的核心要素

业务元数据和技术元数据矩阵图



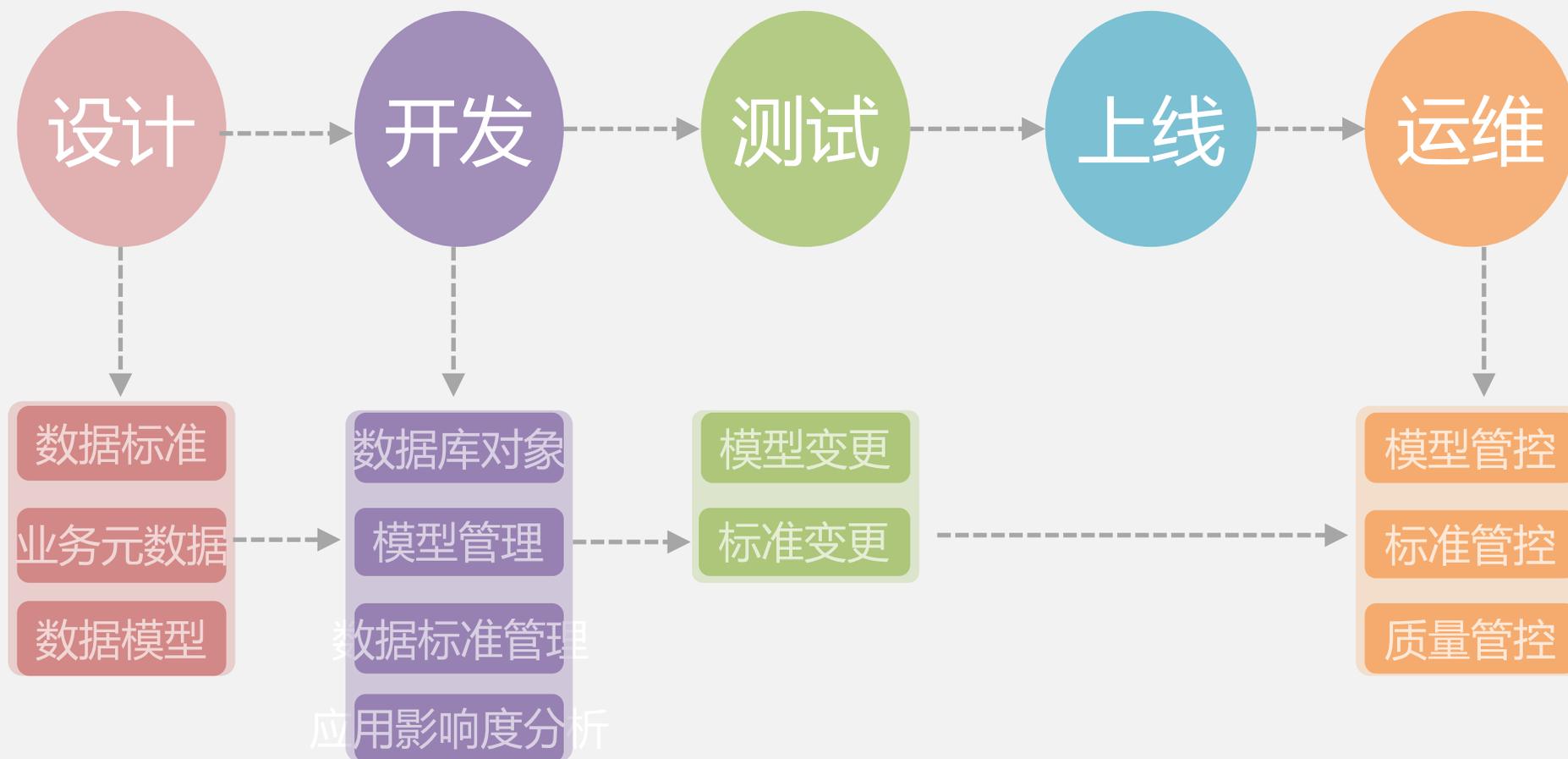
II. 数据治理成功的核心要素

I II III IV

措施3

提供管控型管理流程和自动化应用数据治理系统

支持型数据治理管理体系



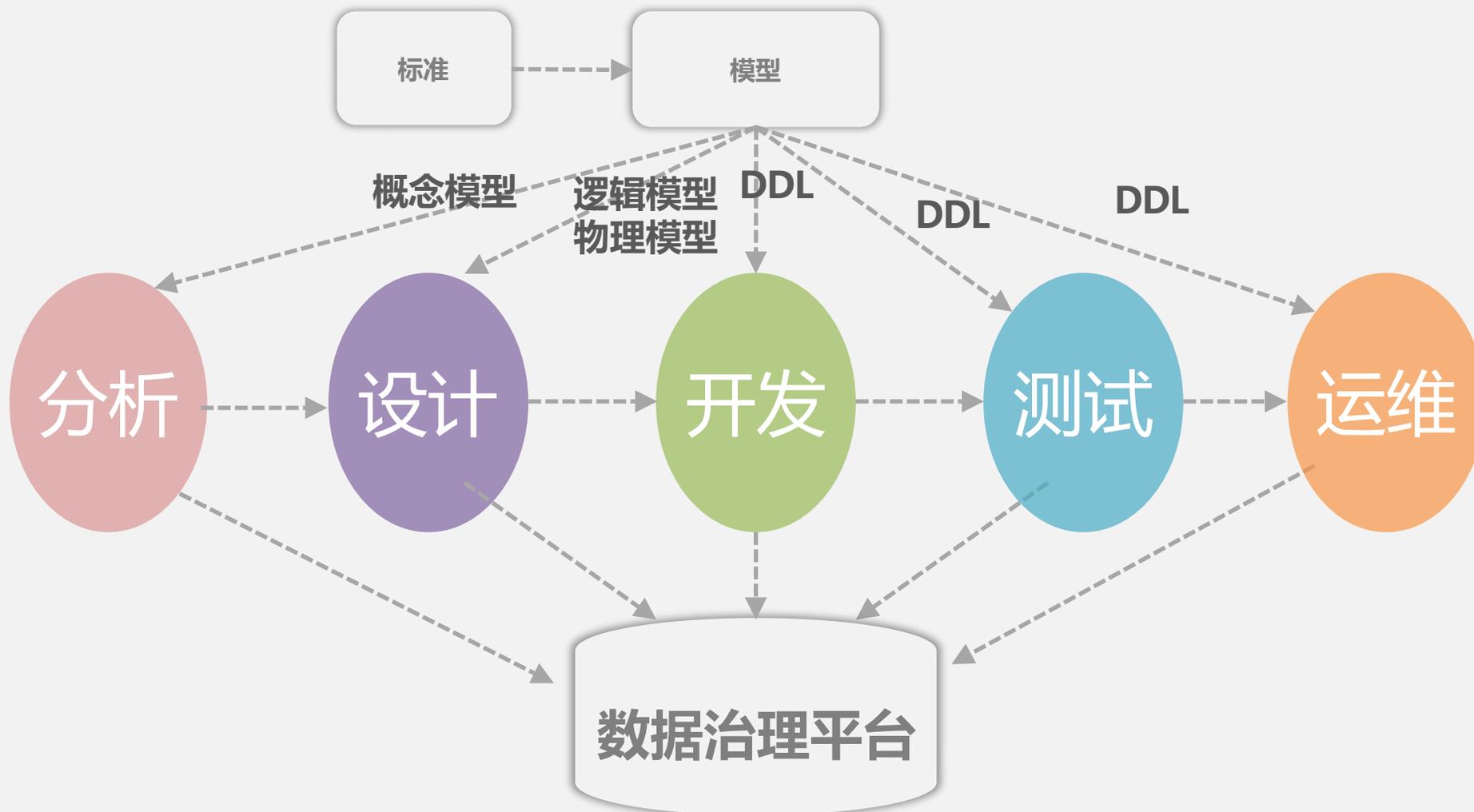
II. 数据治理成功的核心要素

I II III IV

措施3

提供管控型管理流程和自动化应用数据治理系统

管控型数据治理管理体系



II. 数据治理成功的核心要素

I

II

III

IV



II. 数据治理成功的核心要素

逻辑数据模型

单位信息

- # 单位编号
- * 名称
- * 上级单位

员工信息

- # 员工编号
- * 姓名
- * 性别
- * 年龄
- * 出生年月
- * 入职时间
- * 单位编码(FK)

岗位变更

- # 岗位名称
- * 员工编号(FK)
- * 开始时间
- * 结束时间

物理数据模型

DEPT_INF

- # DEPT_NO
- * DEPT_NM
- * HGH_LVL
- * DEPT_NO

EMP_INF

- # EMP_NO
- * NM
- * SEX
- * AGE
- * BRTH_DATE
- * ETY_TIME
- * DEPT_NO(FK)

PSTN_CHG

- # PSTN_NO
- * EMP_NO(FK)
- * STRT_TIME
- * END_TIME



基于数据标准知识库自动转化

数据标准及模型知识库

数据标准	标准单词		标准域		标准转换规则	单词拼接规则	数据模型
	产品	PROD	地址型	VARCHAR2(50)		域指定规则	
	分类	CTG	金额型	NUMBER		单词新增规则	
	编号	ID	编号型	VARCHAR2(50)			
	渠道	CHN	日期型	DATE			
	名称	NM	名称型	VARCHAR2(50)			
日期	DATE	数量型	NUMBER				

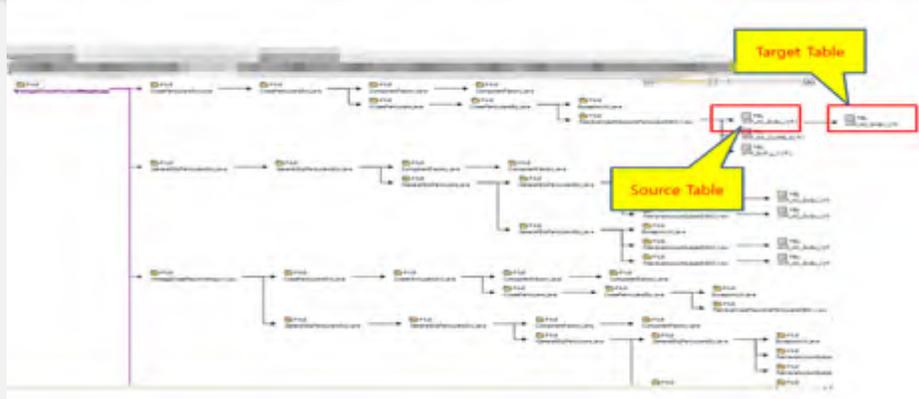
数据模型

- 实体-表
- 属性-字段
- 关系-主外键
- 模型差异管理

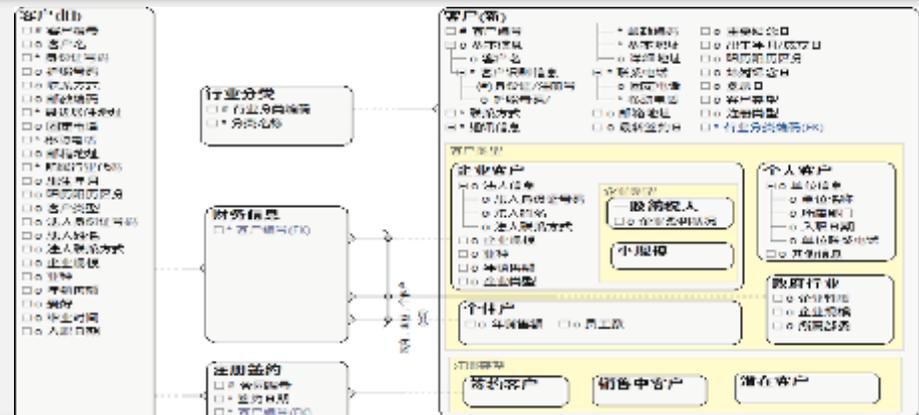
II. 数据治理成功的核心要素

措施4 提供可视化和共享知识库的数据治理系统

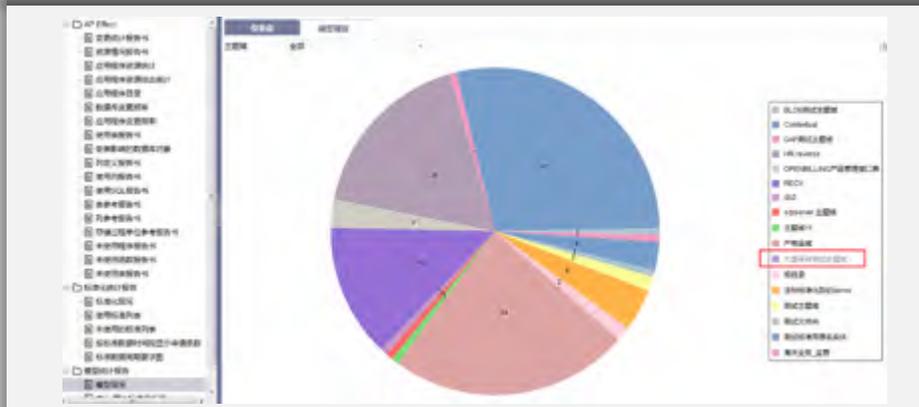
可视化



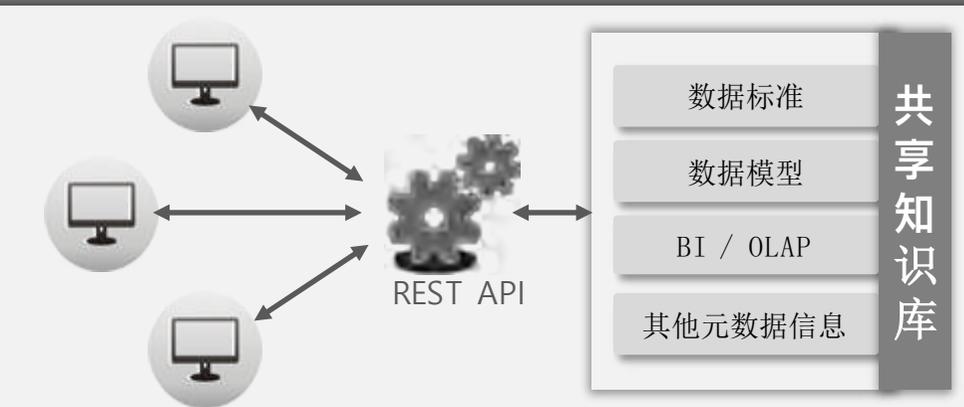
模型视图



综合分析



API



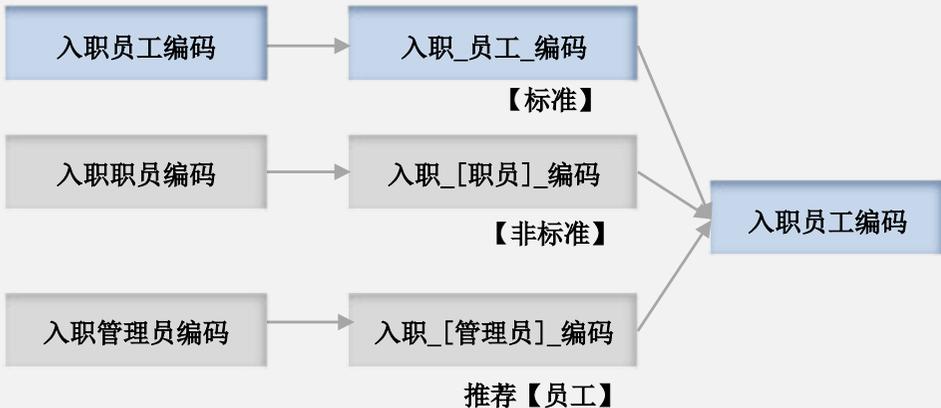
II. 数据治理成功的核心要素

数据标准化的自动校验及应用

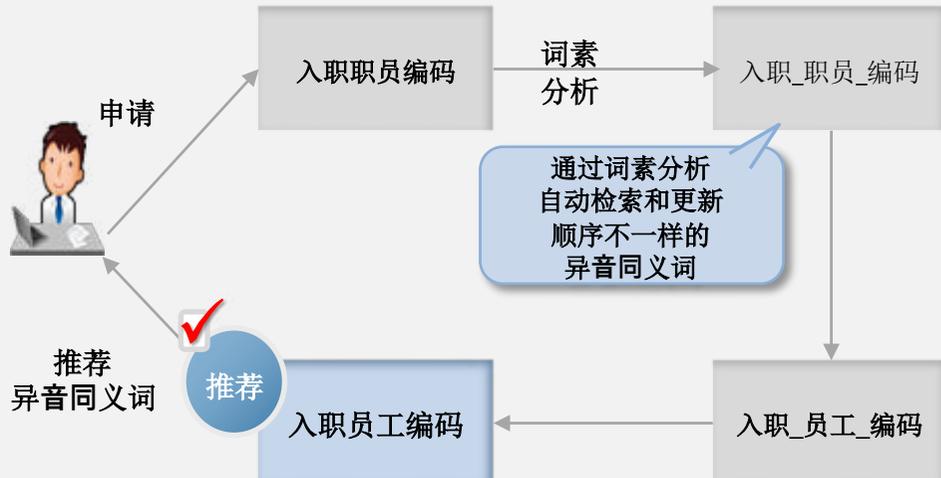
属性名校验

标准词典	单词	标准	同义词
	员工	Y	职员, 管理员
	职员		
	入职	Y	
	服务员		

支持同义词



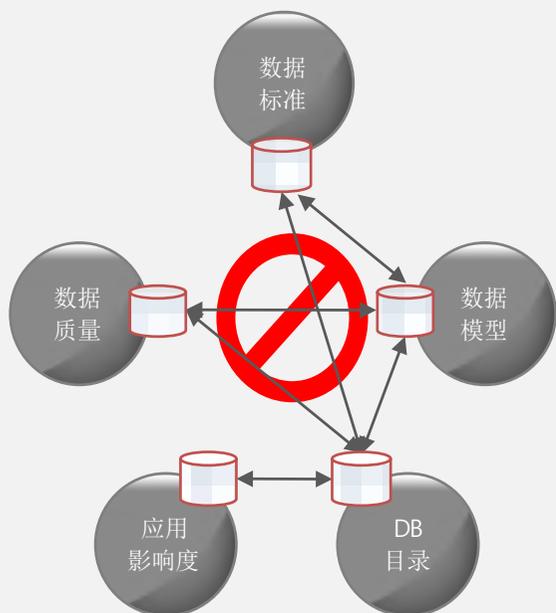
词素解析及校验



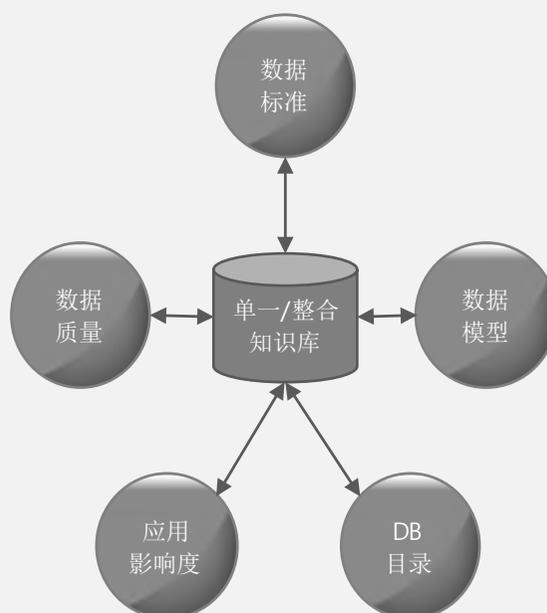
II. 数据治理成功的核心要素

I II III IV

独立型知识库架构



整合型知识库架构



要点

- 1 以数据模型和DB为中心的设计
- 2 数据库信息集中管理
- 3 元数据相关关系分析和应用
- 4 专业数据建模技术应用



Ⅲ. 数据模型管控

1. 数据模型管控必要性及问题分析
2. 数据模型管控解决方案
3. 数据模型管控核心价值

III. 数据模型管控— 数据模型管控必要性及问题分析

数据模型变更缺乏事前审计、事中监控、事后管理等体系化的管控措施，致使数据模型逐渐变成“黑盒子”，给系统建设和数据应用带来严重影响。

1 生产库中存在大量字段和表没有注释、含义模糊不清、同名不同意、同意不同名、冗余字段和表、枚举型字段中的值使用不统一等现象，直接影响对数据的识别和应用。

2 数据模型变更变更前的合理性缺乏专业人员评审，大部分企业缺乏专业的数据架构师，对不同系统的数据模型，在变更时从数据设计、业务合理性、数据治理、数据库性能等方面进行综合性评审。

3 修改过程中缺乏监控和管理，修改操作是否符合规范，修改脚本是否按照要求编写，修改时是否先修改模型再编写脚本，是否及时保证数据模型与数据库的同步等缺乏监控和管理。

4 修改数据模型后未及时将修改的部分公开，修改的内容仅限于其内部或较少几个人知道，其他人员均不知晓，同时也未对修改的内容进行管理，致使系统出现故障时排查问题难度较大，数据模型逐渐变成“黑盒子”。

问题汇总

审计工作及指标缺失

管理流程缺失

辅助性工具体系缺失

事后弥补措施缺失

III. 数据模型管控- 数据模型管控必要性及问题分析

针对数据模型管控方面存在的诸多问题，我们应从相应岗位设置、管控工具构建、管理流程设计等三个方面着手，构建适合的数据模型管控体系。

改善目标

1 审计工作及指标缺失

- 设置数据架构师岗位，设计模型变更合理性审计指标

2 管理流程缺失

- 设计适合的管理流程及数据架构师、模型设计人员、数据标准管理、开发、运维等角色分工

3 辅助性工具体系缺失

- 构建数据模型管控平台，实现辅助管理

4 事后弥补措施缺失

- 设计事后弥补流程，对未按照流程发起的变更，设计及时发现及事后弥补方案。

改善方案

1 岗位设置

- 数据架构师（制定制度及流程）
- 数据模型（模型设计及管控）
- 数据标准（标准管理及应用）

+

2 管控工具

- 模型管控
- 数据标准
- 变更监控

+

3 管理流程

- 人员交互流程
- 事前、事中审计及管理流程
- 事后弥补及共享流程

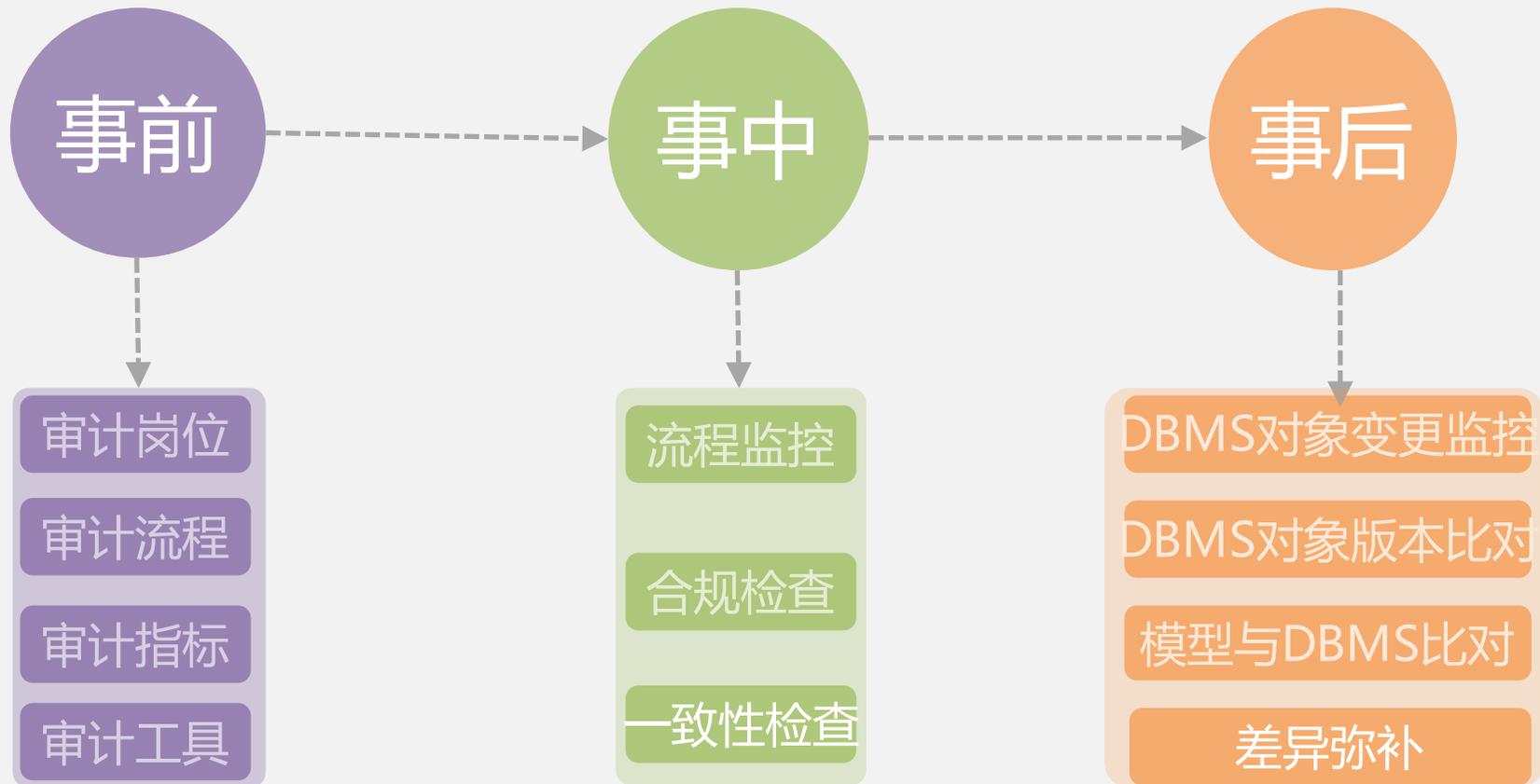
III. 数据模型管控- 数据模型管控方案

I

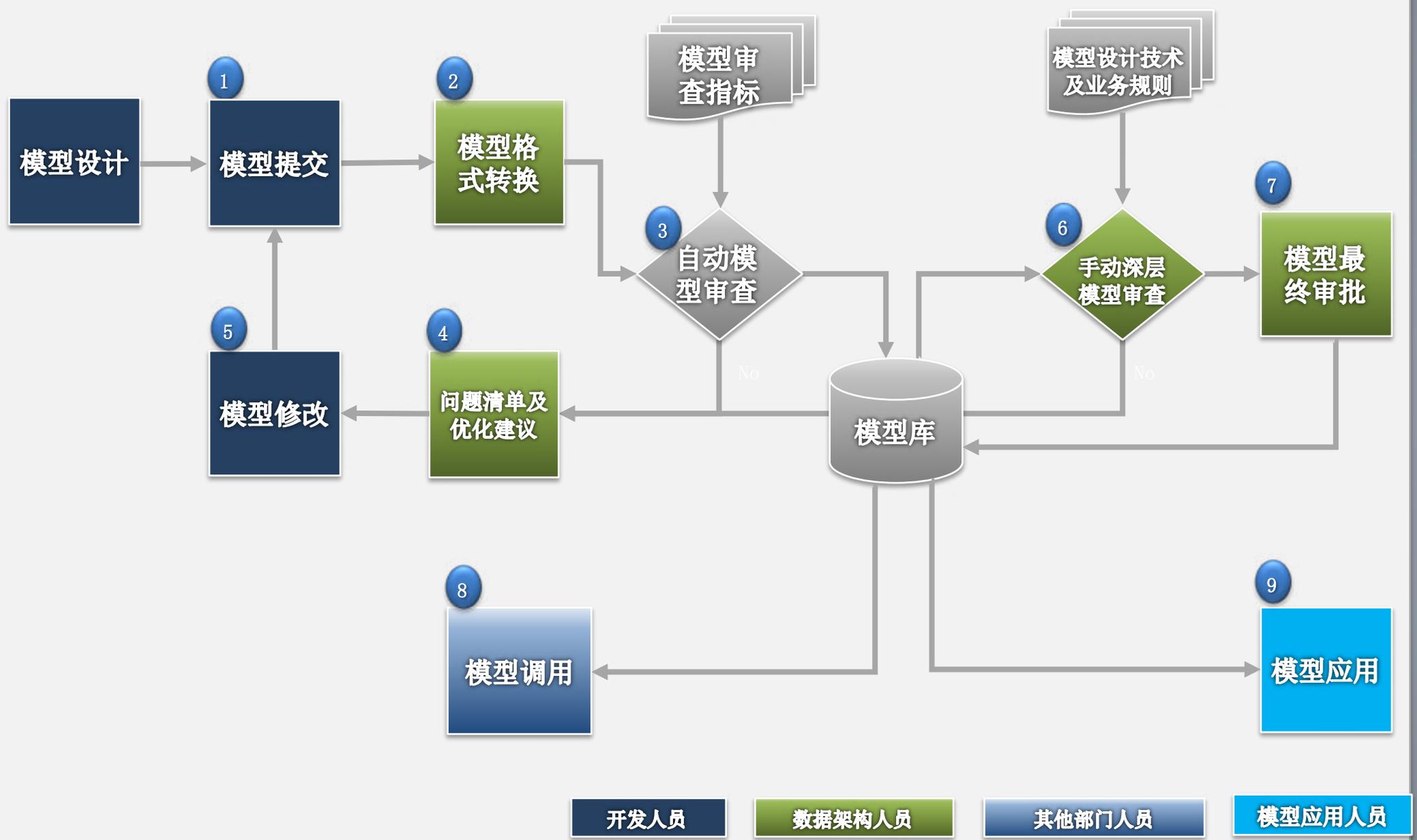
II

III

IV



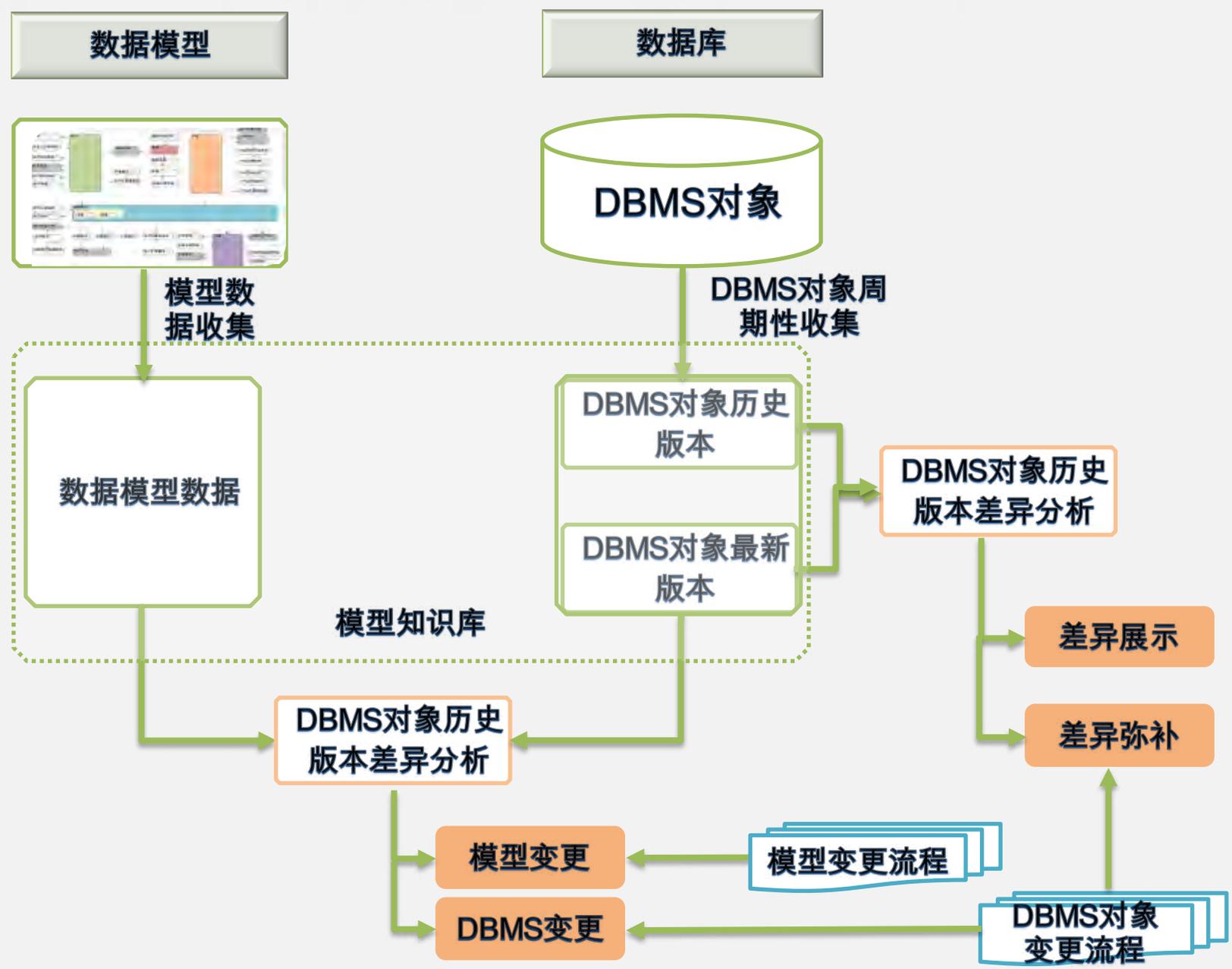
III. 数据模型管控- 事前数据模型管控方案



III. 数据模型管控— 数据模型审计指标

指标类别	审核指标	审查指标内容	审查方式
逻辑模型	中文实体名称	模型工具(DA#)根据已构建的单词词典对模型中的实体名称进行解析, 审查其构成是否符合数据标准化	工具
	实体定义	逻辑数据模型中的实体名称定义部分是否为空, 即是否对实体的定义、用途进行详细说明	工具
	中文属性名称	模型工具(DA#)根据已构建的单词词典对模型中的属性名称进行解析, 审查其构成是否符合数据标准化	工具
	属性定义	逻辑数据模型中的属性名称定义部分是否为空, 即是否对属性的定义、用途进行详细说明	工具
	主标识符	审查逻辑模型中是否有未定义主标识符的实体, 并自动列举这些实体的名称	工具
	继承标识符	审查从其他实体中继承的主标识符名称是否与自身实体中的属性重名	工具
	实体相似度	审查逻辑模型中的所有实体相互之间是否有属性个数及属性名称相似度非常高(70%以上认为重复)	工具
	实体名称相似度	审查逻辑模型中的所有实体相互之间在实体名称上是否相似度非常高	工具+人工
	主标识符相似度	审查逻辑模型中的所有实体相互之间在实体的主标识符上是否相似度非常高或相同(审查重复性)	工具
	范式	审查具有继承关系的实体中是否有除了继承的标识符之外的其他一般属性名称相同	工具
实体孤立	审查在逻辑数据模型中是否存在没有与任何实体具有关系的孤立实体	工具	
物理模型	表名称	模型工具(DA#)根据已构建的单词词典对模型中的表名称进行解析, 审查其英文定义是否符合数据标准化	工具
	字段名称	模型工具(DA#)根据已构建的单词词典对模型中的字段名称进行解析, 审查其英文定义是否符合数据标准化	工具
	主键定义	审查物理模型中是否有未定义主键的表	工具
	字段域定义	模型工具(DA#)根据已构建的单词词典对模型中的域进行解析, 审查字段域是否符合数据标准化	工具
	字段重复	审查物理模型中具有关系的表中是否有除了外键之外的其他一般字段名称相同	工具
	中文字段	审查物理模型中是否具有利用中文对字段名进行定义的情况	工具
	主键相似度	审查物理模型中是否具有主键名称相似或完全一致的表(审查重复)	工具

III. 数据模型管控— 事后数据模型管控方案



II. 数据模型管控-价值

I II III IV

设置数据模型 审计专职岗位和设计审计指标

- 设置专职数据建模或数据架构师岗位
- 培养数据架构师，掌握建模技术及业务知识
- 设计符合实际情况且可落地的审计指标

构建自动化强，落地性好的软件工具

- 数据架构师、数据标准、数据模型等管理流程配置功能
- 数据标准管理及应用、数据模型管控功能
- 审计校验功能，内置审计校验指标
- 数据库对象自动搜集及差异比对功能

设计及实施数据模型管控流程

- 设计数据架构师、数据标准、数据建模人员、开发人员交互流程
- 设计事前、事中、事后审计及校验流程
- 设计数据模型及数据库对象变更流程

核心价值

- 模型透明化管理
- 提升数据质量
- 数据资产化
- 数据变现保障
- 实现模型数据动态化管理

直接效果

- 模型变更有流程可遵循
- 模型审计有据可依
- 数据标准确保落地
- 模型事后变更及时弥补
- 变更数据共享并公开

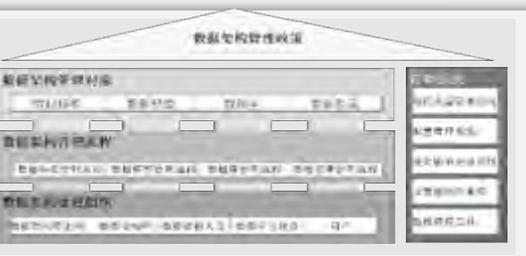
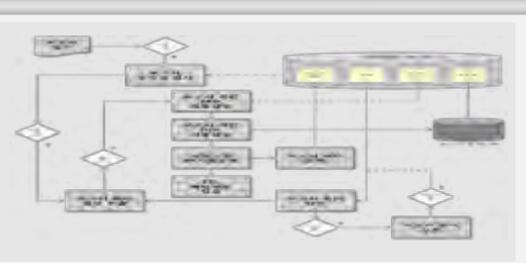


IV. 成功案例

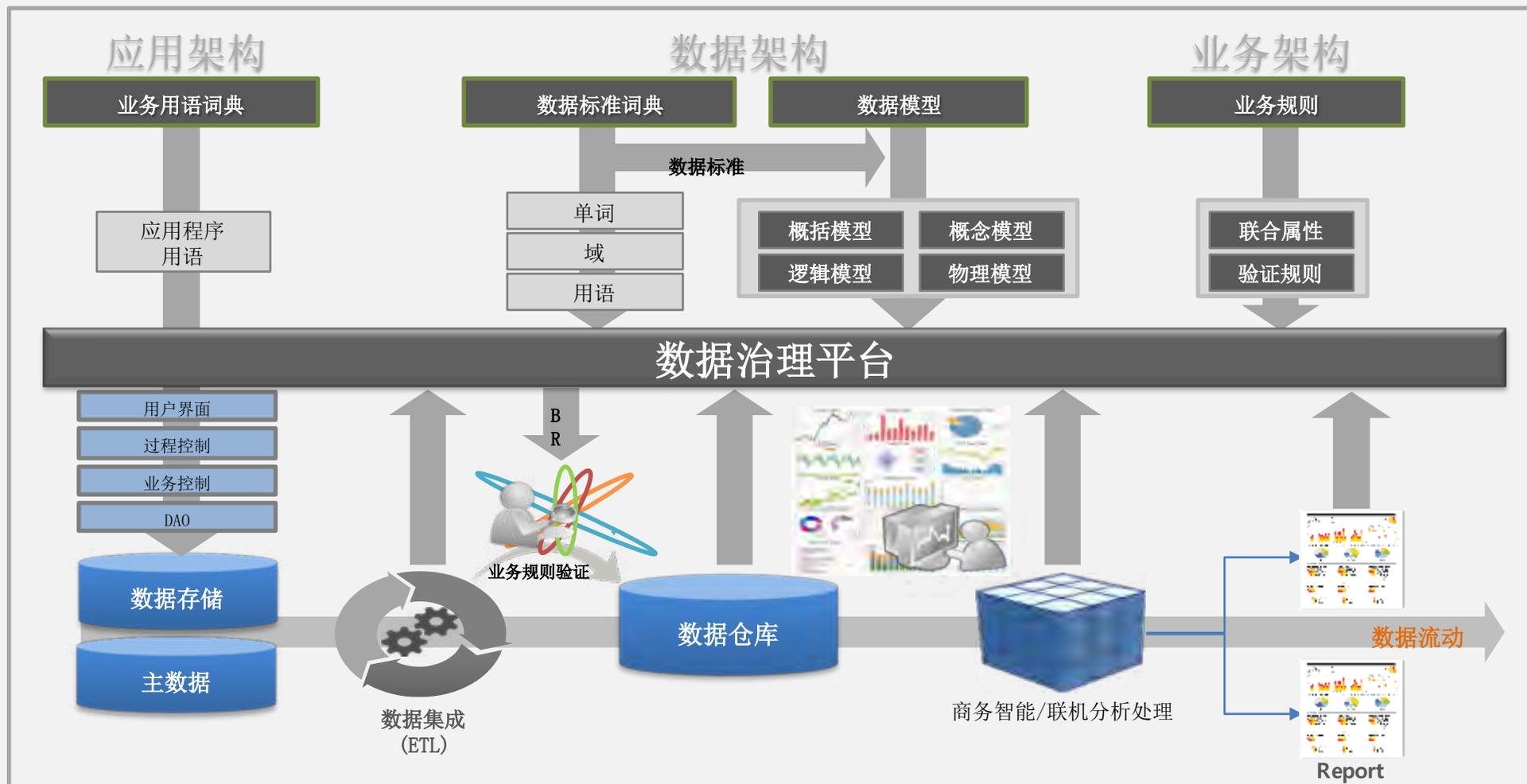
1. . 某移动公司案例
2. 金字工程案例



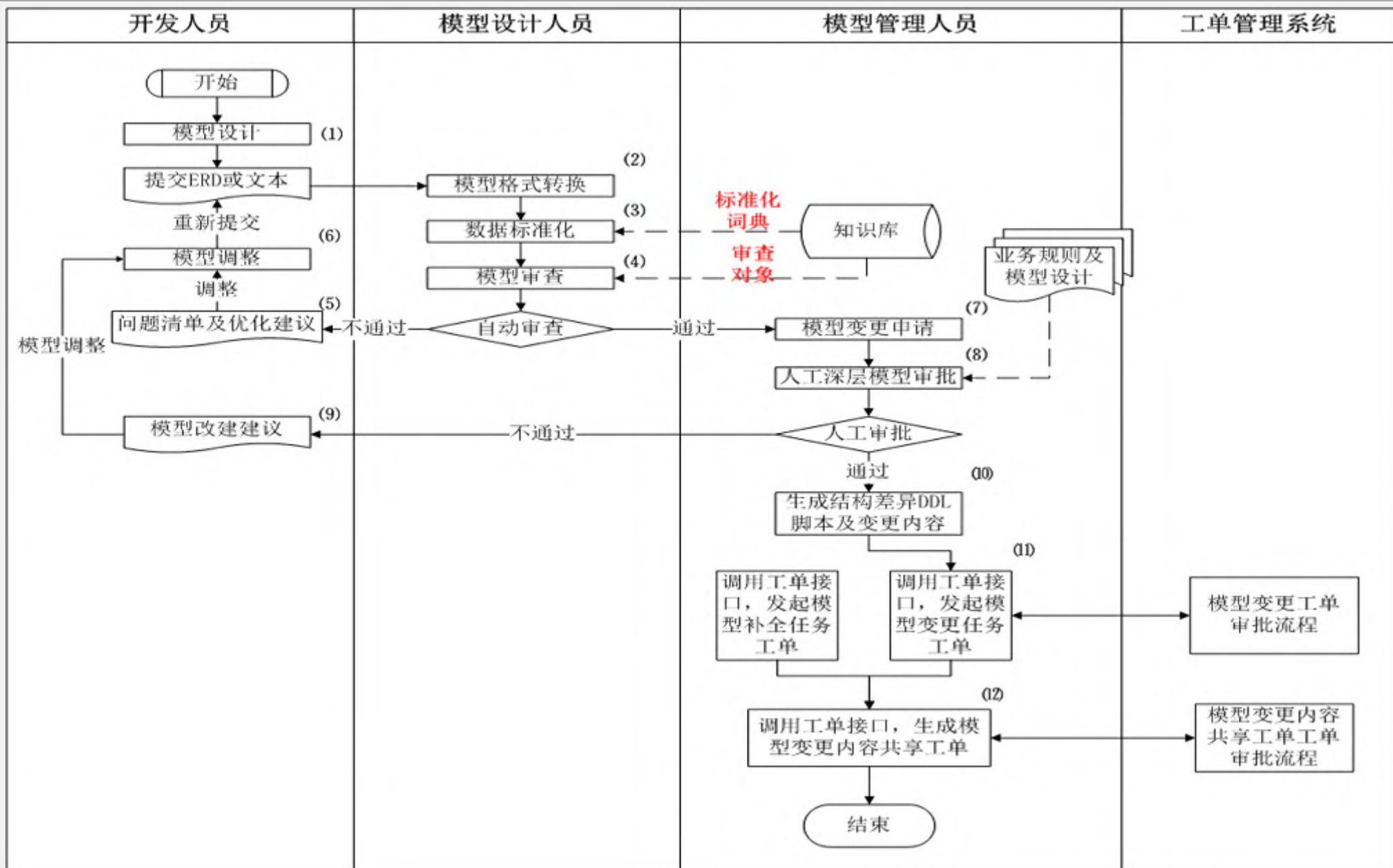
IV. 成功案例

<p>现况诊断及设计改善方案</p>	<ul style="list-style-type: none">• 现况问题搜集• 当前管理方式分析• 报表及关键字管理现况分析• 设计改善方案	
<p>韩国银行外汇管理系统数据标准及数据质量标准定义</p>	<ul style="list-style-type: none">• 标准单词、域、用语、及命名规则制定• 数据标准定义指南• 数据架构定义• DQI, CTQ设计• BR管理流程定义	
<p>数据质量管理体系统构建</p>	<ul style="list-style-type: none">• 质量检测原则设计• 标准流程定义• 组织、角色、职责定义• 系统需求定义	
<p>报表及关键字元数据管理系统构建</p>	<ul style="list-style-type: none">• 报表元数据管理指南定义• 报表元数据管理组织、角色、职责定义• 报表元数据流程定义• 关键字元数据管理流程定义的	

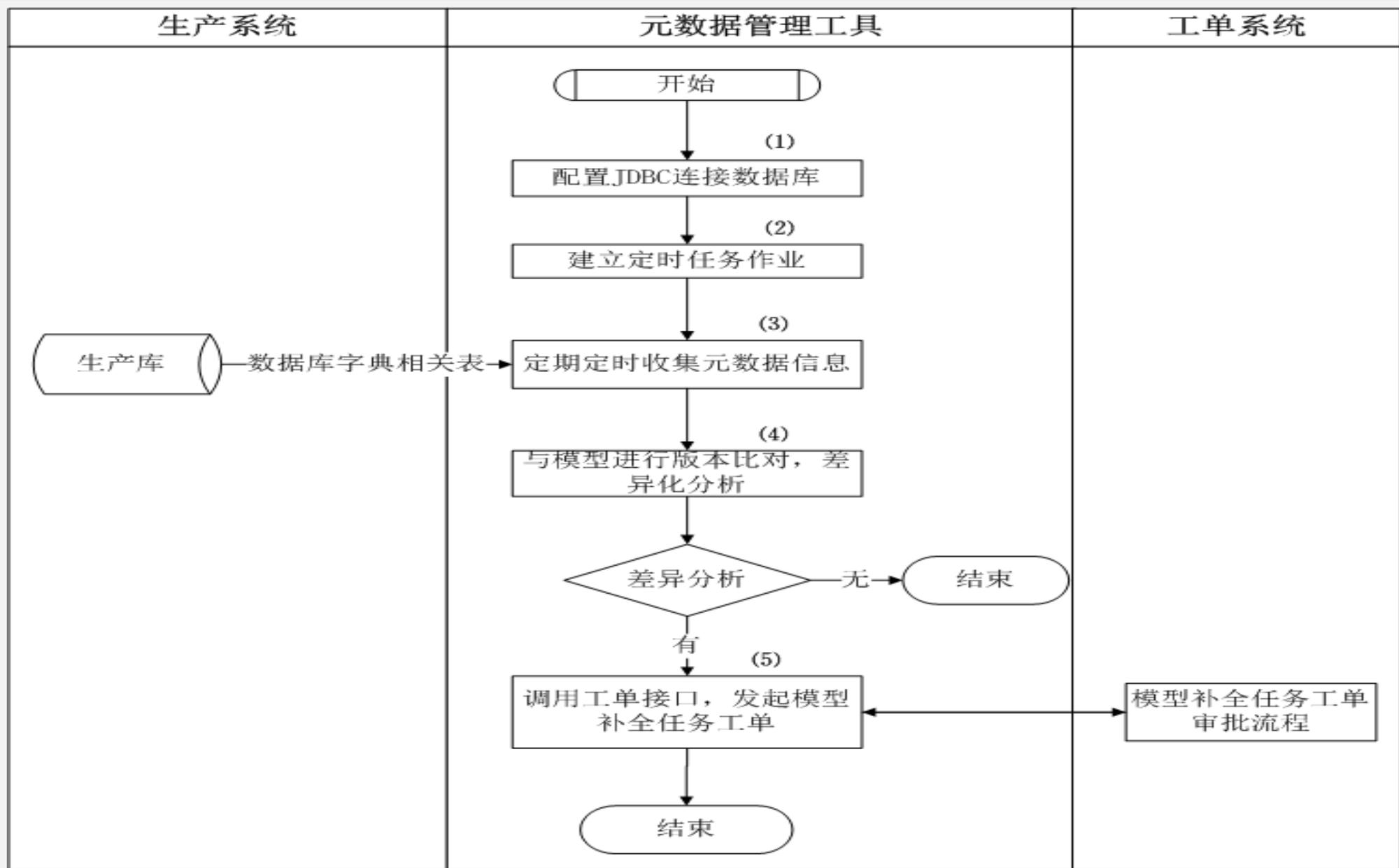
IV. 成功案例



IV. 成功案例



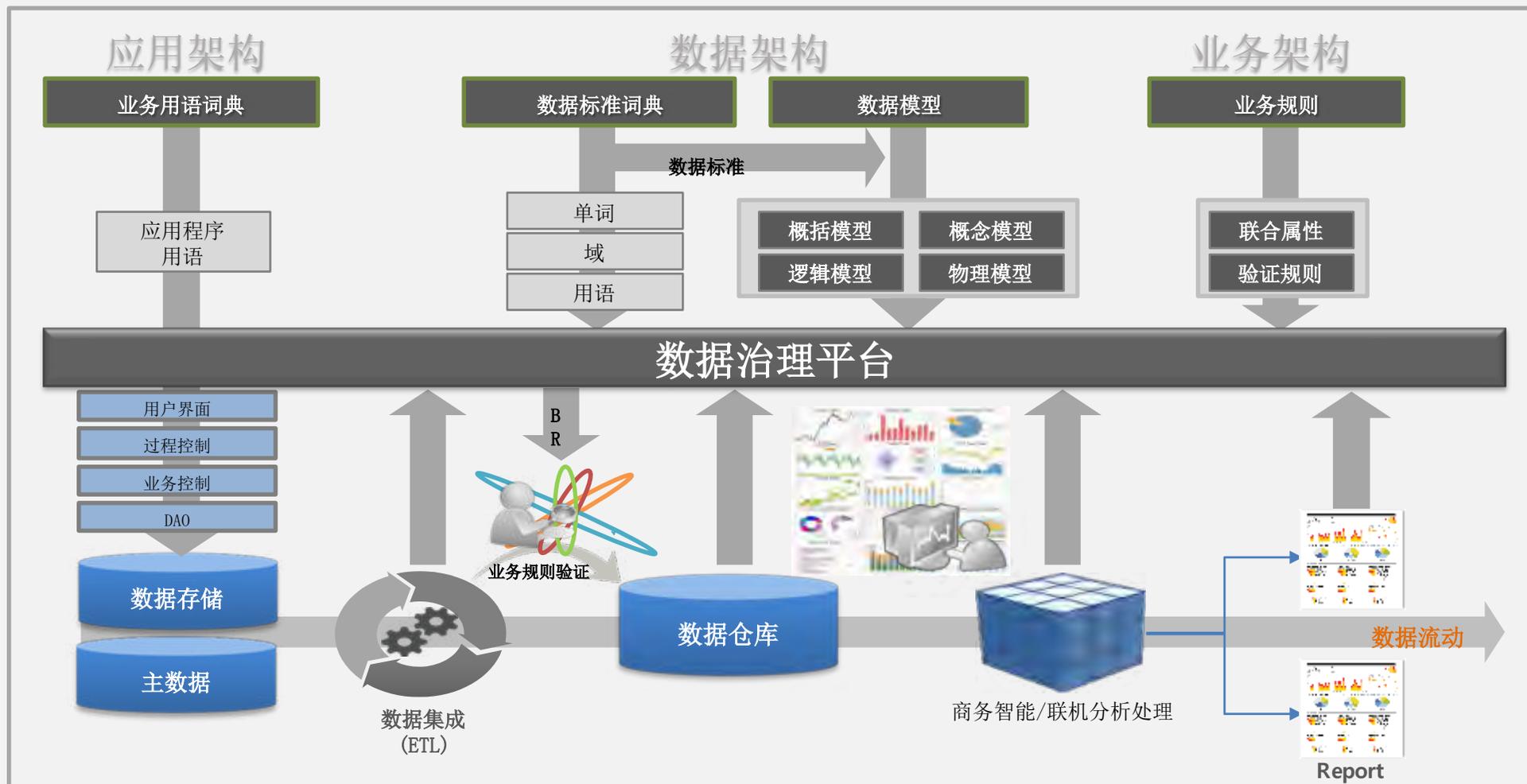
IV. 成功案例



IV. 成功案例

模型名词	类型	创建人	创建日期	修正日期	Locker	权限
 10085电商平台	文件夹	gaoxiaoyan	2015-11-27 09:...	2015-11-27 09:...		修改
 10086社区	文件夹	wangxufe...	2015-12-18 17:...	2015-12-18 17:...	wangxufe...	修改
 4G管家数据模型	文件夹	lizhanyang	2015-12-16 18:...	2015-12-18 14:...		修改
 中移在线-客服系统	文件夹	encore	2016-02-26 10:...	2016-03-11 15:...		修改
 分析平台	文件夹	mazhong	2015-12-19 14:...	2015-12-24 19:...		修改
 呼叫平台	文件夹	lishenghui	2015-12-19 17:...	2015-12-21 17:...	lishenghu...	修改
 外呼众包	逻辑		2015-12-25 15:...	2015-12-25 15:...		修改
 实名制	文件夹	liuguangxu	2015-12-18 16:...			修改
 微信平台	文件夹	gaoxiaoyan	2015-12-31 17:18:08			修改
 支付平台	逻辑		2015-12-30 08:...	2015-12-30 18:...		修改
 智能短信平台	文件夹	wangxufe...	2016-01-15 15:...	2016-01-15 15:...		修改
 精益扶贫业务管理系统	文件夹	gaoxiaoyan	2016-02-15 11:...	2016-02-19 10:...		修改
 能力管控平台	文件夹	liupengch...	2015-12-16 16:...	2015-12-17 18:...		修改
 语音专线管控平台	文件夹	loupeili	2015-12-31 11:...	2015-12-31 11:...		修改

IV. 成功案例



谢谢！

EN·CORE



THANKS

SequeMedia
www.sequemediacn.com

IT188

ChinaUnix

ITPUB

THANKS