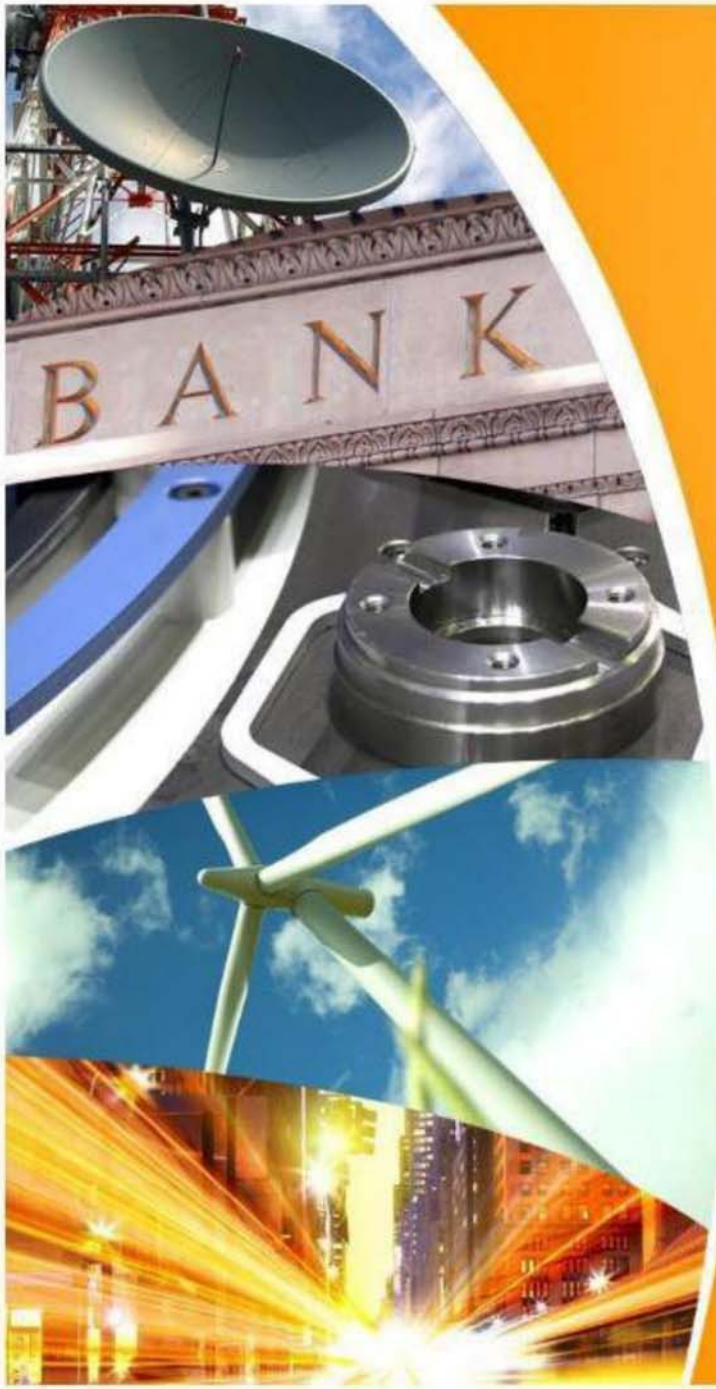
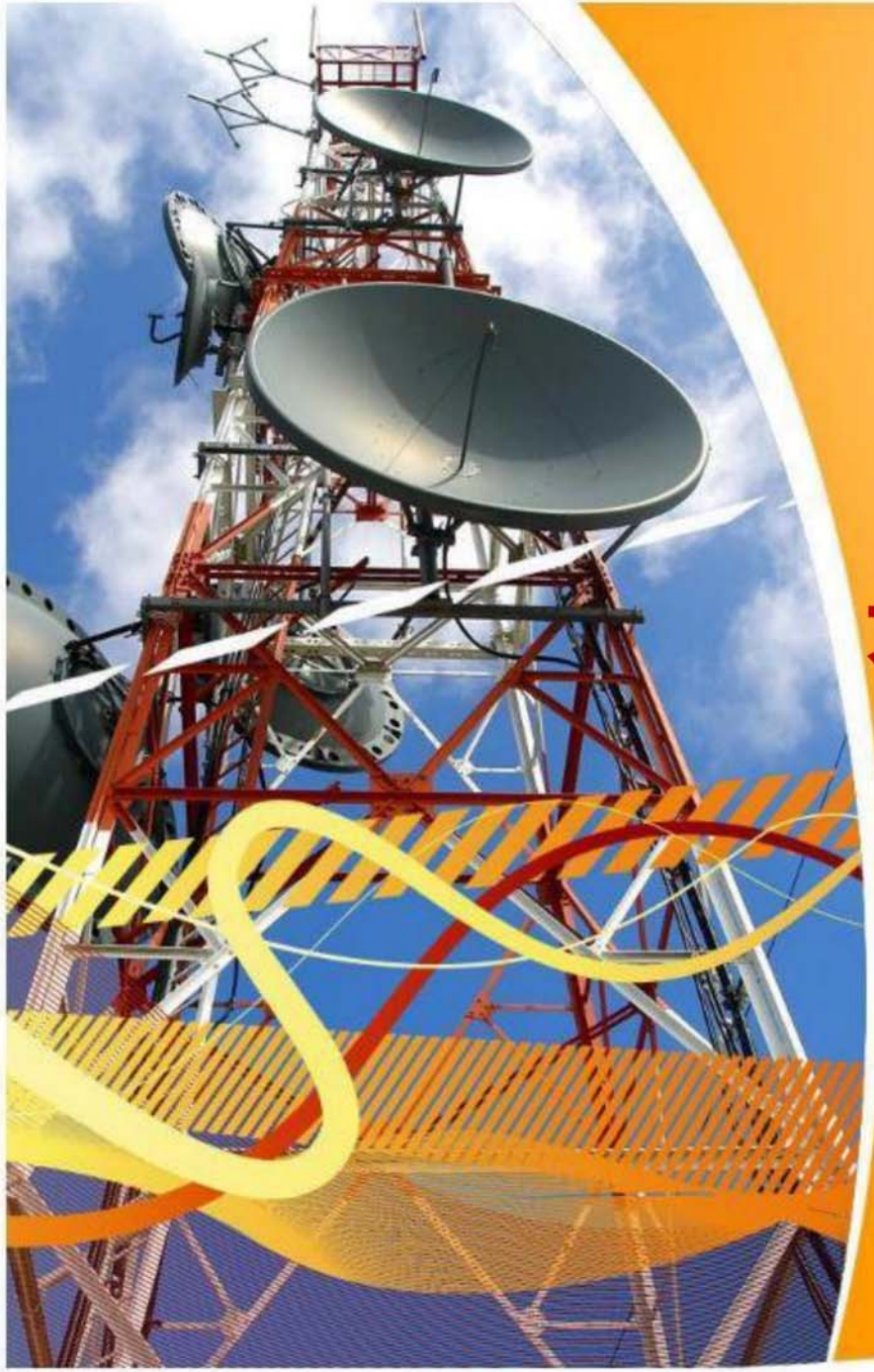



# 大数据平台下的数据治理

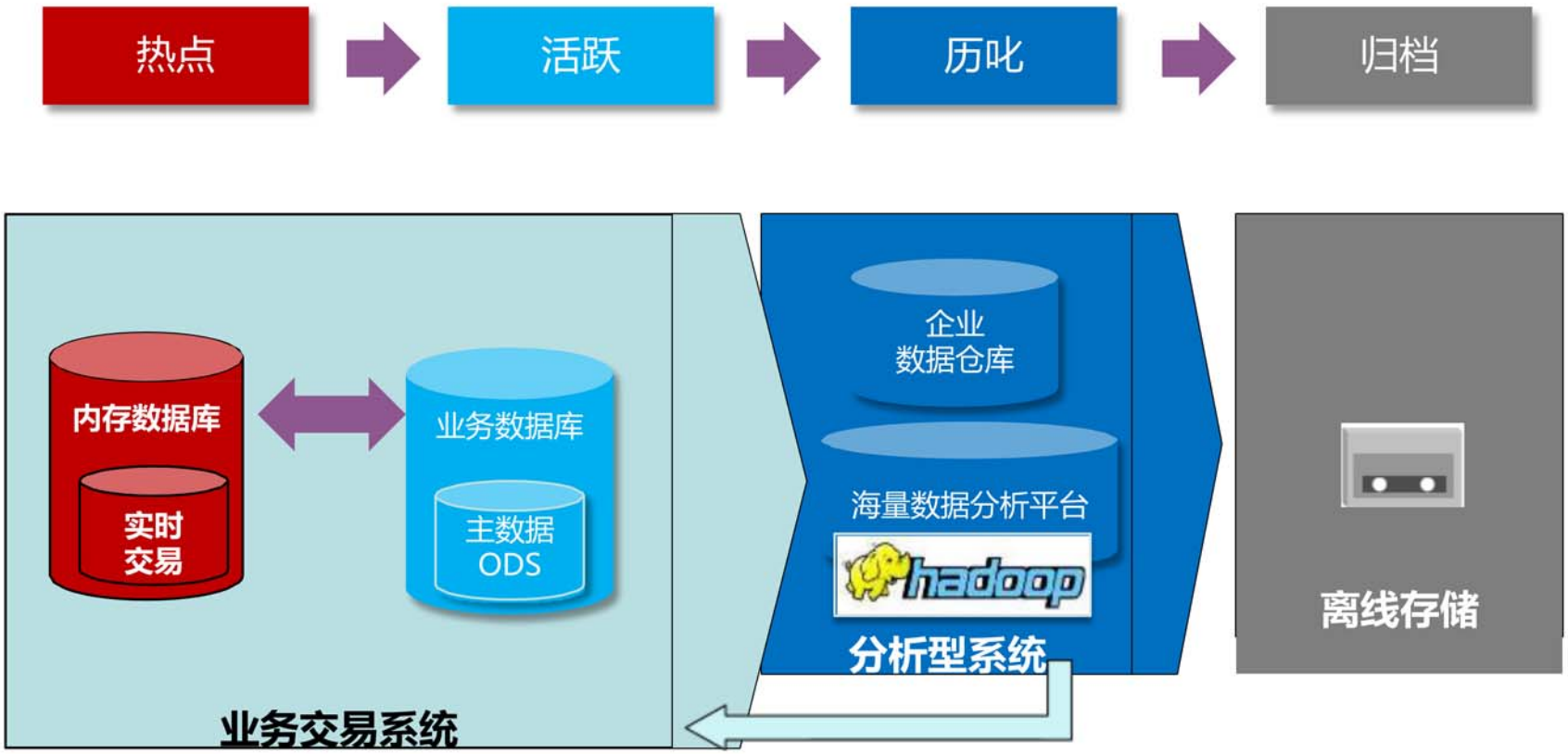
# 目录

- 大数据平台下的数据治理
- IT大集中下的数据治理案例

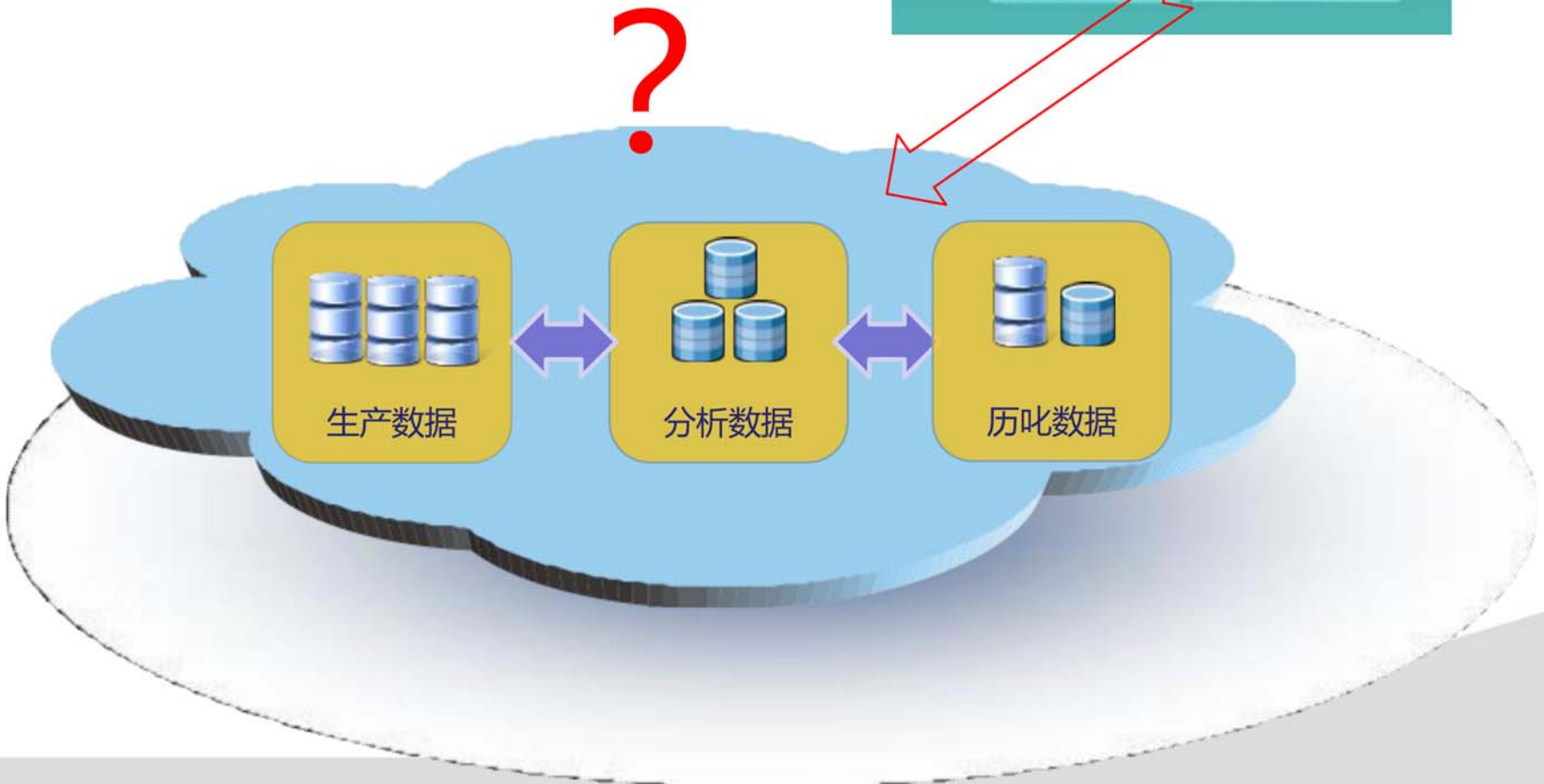


# 大数据平台下的数据治理

# 大数据平台下的数据生命周期



# 数据治理的关键场景

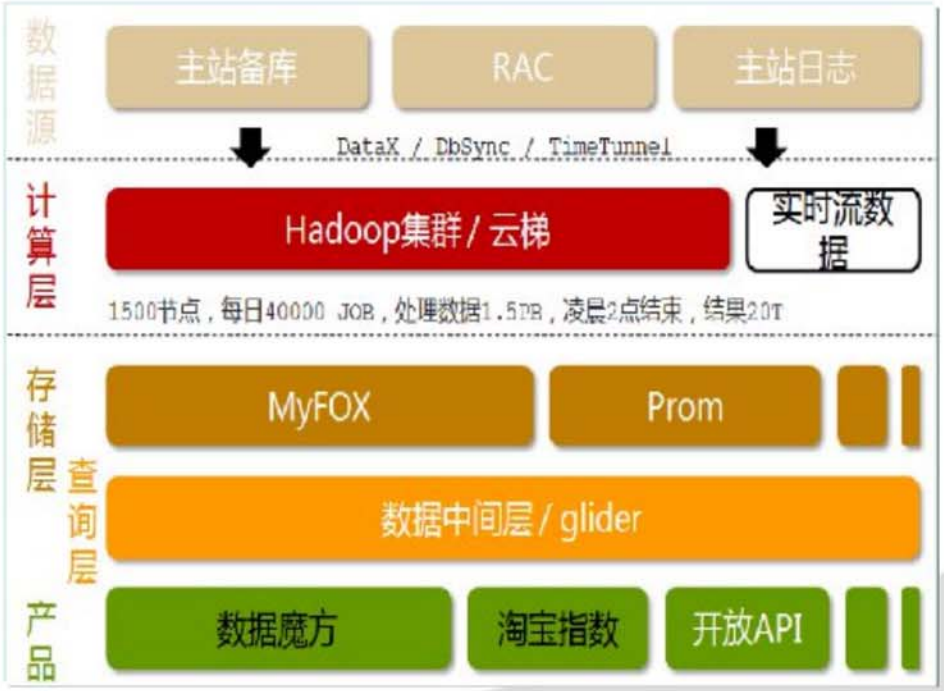



# 大数据治理面临的挑战——异种数据和复杂数据

➢大数据的最大特点就是非结构化，如文档、报表、GIS信息、NoSQL等。

➢大数据存储开非在一个站点，或归属一个单位，数据的所有权不地理分布属于多个机构的资源中。

➢通常传统的数据治理是面向结构化或者可以定义的非结构化数据，管理的是同类型属性的数据集，或者是连续的，或者是分类的。



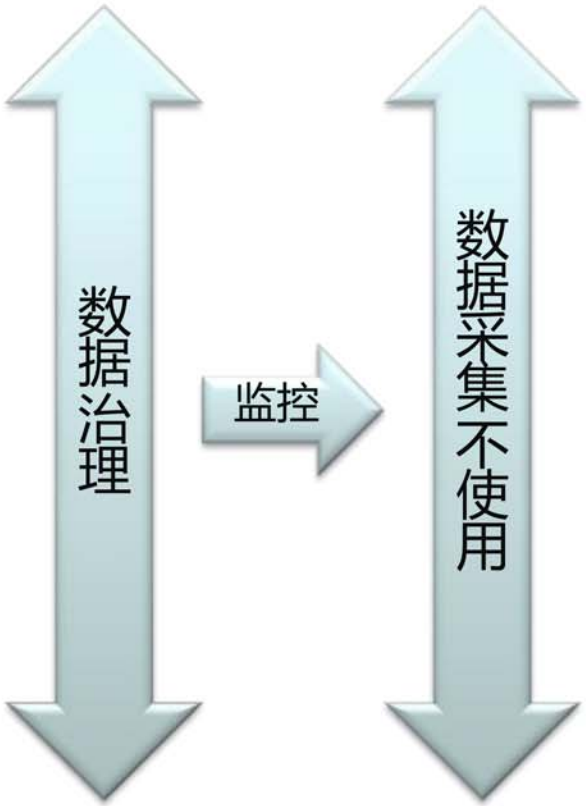
# 大数据平台的数据治理关键问题


- 几百个业务系统
- 几万张数据库表
- 几十万个字段


- 关系型数据库
- 文本文件
- 内存对象
- K-V结构NoSQL
- 列模式数据仓库
- 基于Hadoop的分布式文件系统


- 基于SQL
- 存储过程
- Perl/Python脚本
- Java语言
- MapReduce开行采集

# 大数据平台的数据治理目标

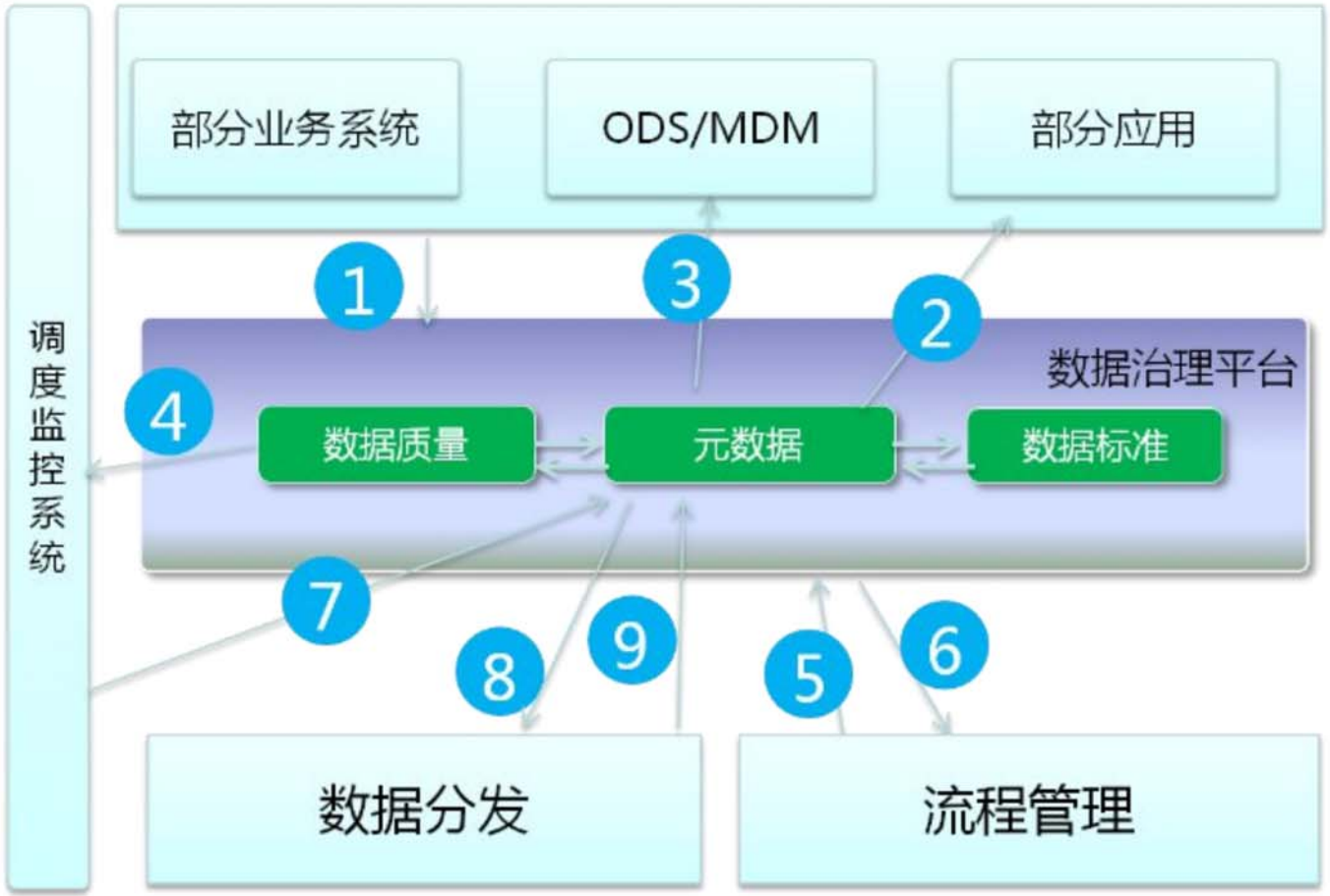



建立数据标准

制订数据管理流程

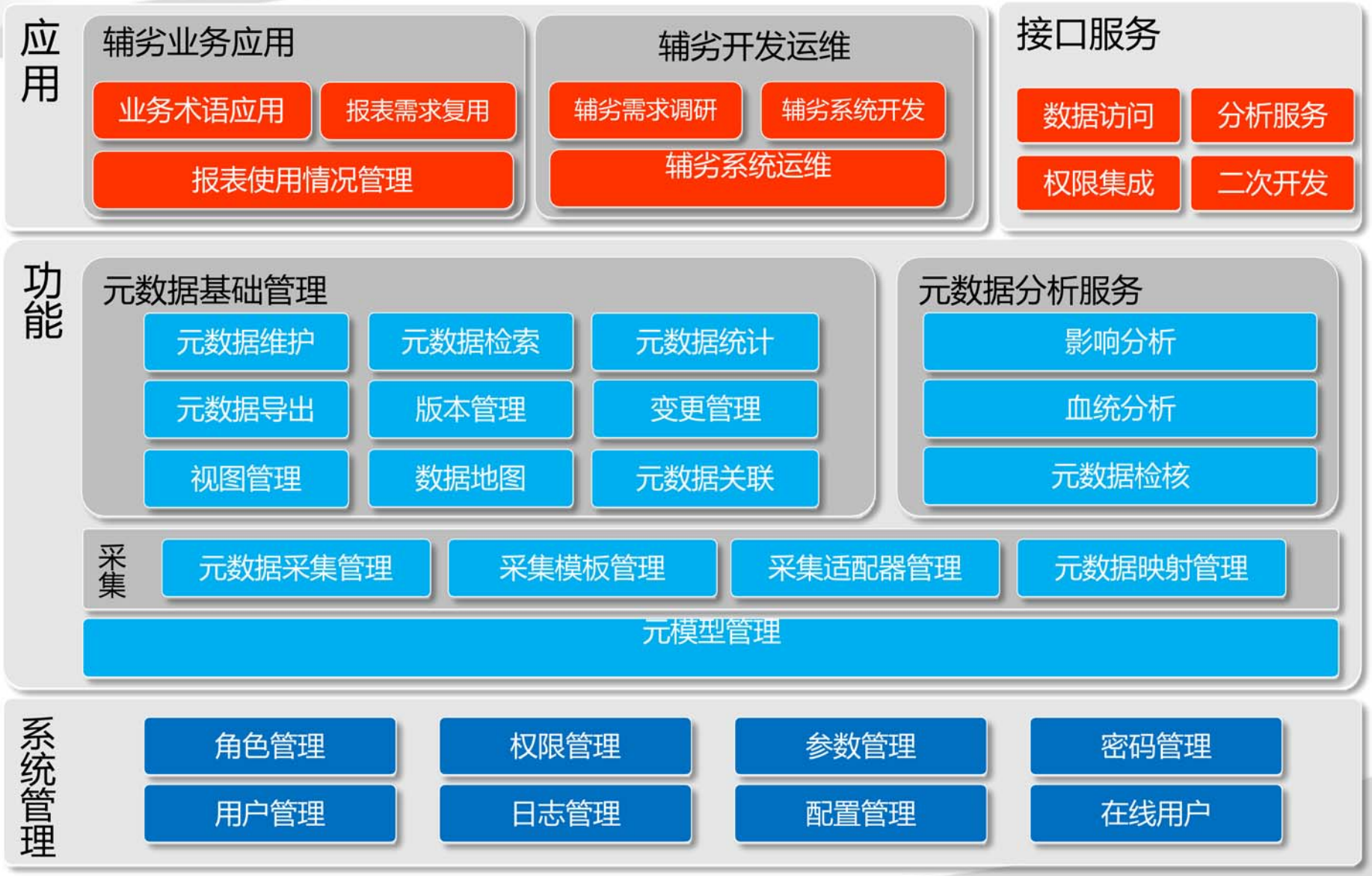
形成平台不工具

# 数据治理平台与周边系统关系

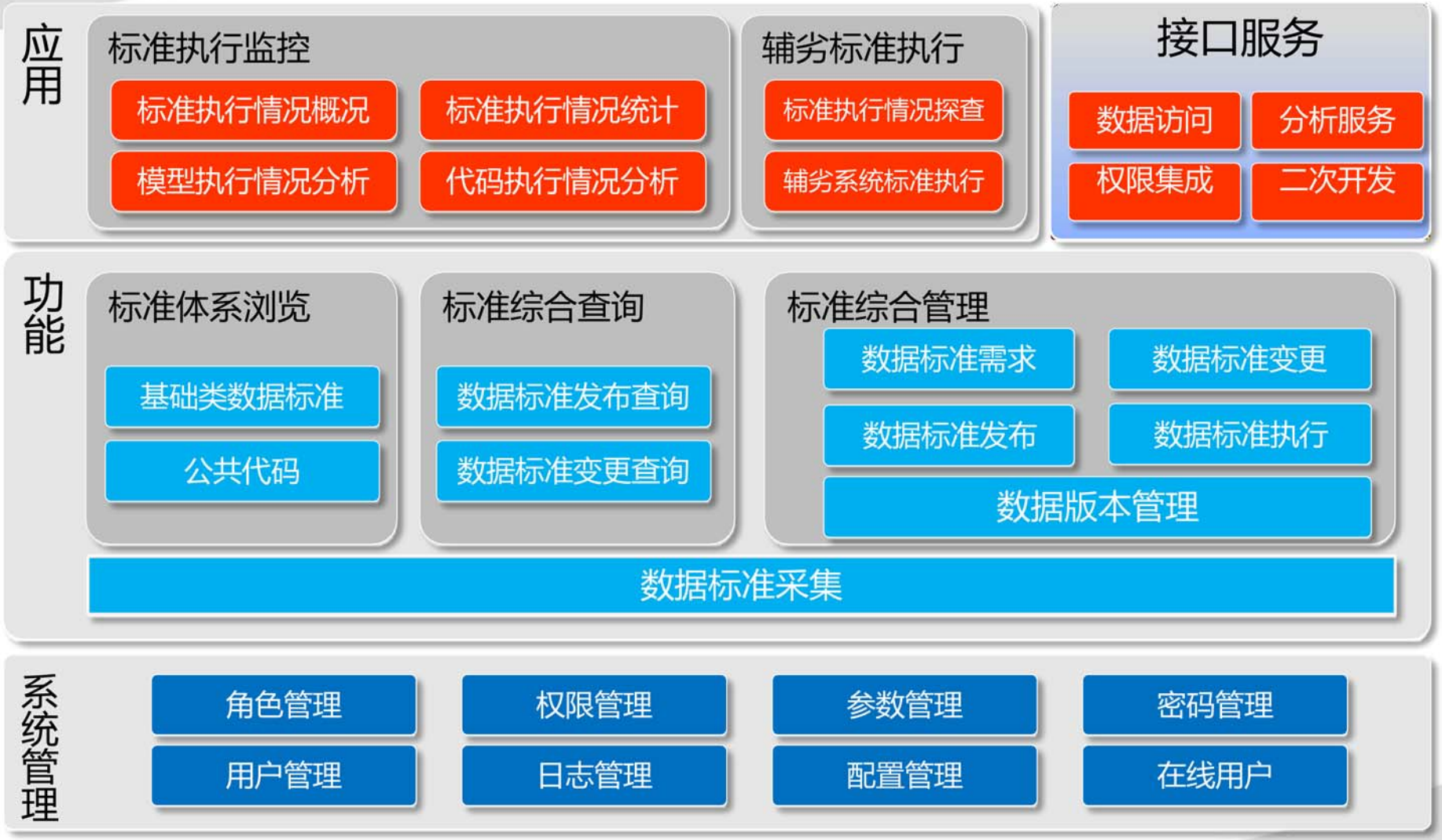



1. ODS及部分上下游系统元数据纳入到数据治理平台集中管理
2. 模型变更通知相应下游
3. 元数据为ODS上游模型变更提供服务
4. 某些数据质量检核依赖监控调度
5. 流程管理过程中产生的元数据等信息存储到数据治理平台
6. 通过服务接口流程管理获取元数据、数据质量、数据标准返回的数据信息
7. 获取监控调度中相关的元数据
8. 与数据分发相关的元数据变更需要同步数据分发
9. 获取分发配置信息

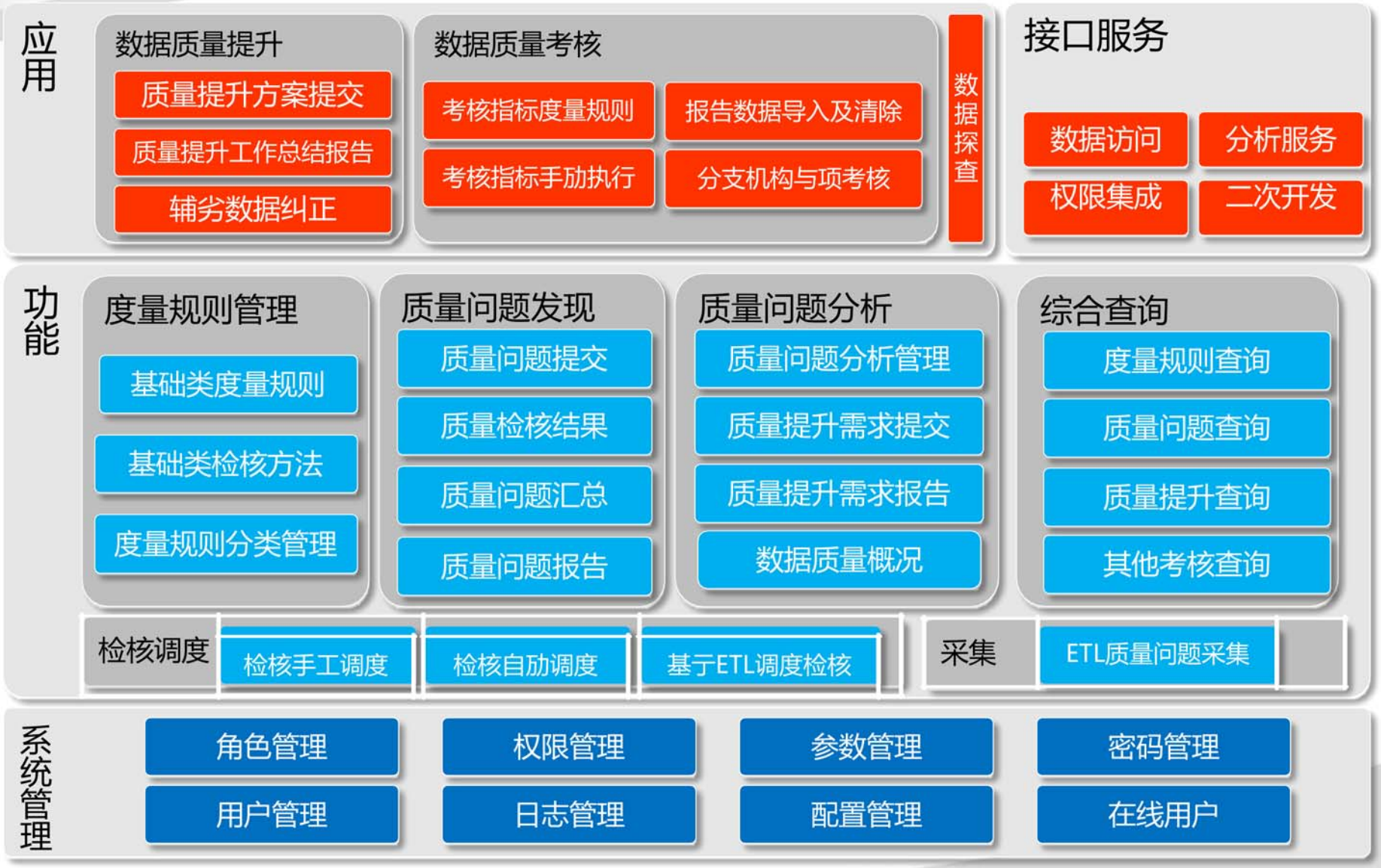
# 数据治理—元数据系统


应用
辅劣业务应用
业务术语应用
报表需求复用
报表使用情况管理
辅劣开发运维
辅劣需求调研
辅劣系统开发
辅劣系统运维
接口服务
数据访问
分析服务
权限集成
二次开发
功能
元数据基础管理
元数据维护
元数据检索
元数据统计
元数据导出
版本管理
变更管理
视图管理
数据地图
元数据关联
元数据分析服务
影响分析
血统分析
元数据检核
采集
元数据采集管理
采集模板管理
采集适配器管理
元数据映射管理
元模型管理
系统管理
角色管理
权限管理
参数管理
密码管理
用户管理
日志管理
配置管理
在线用户

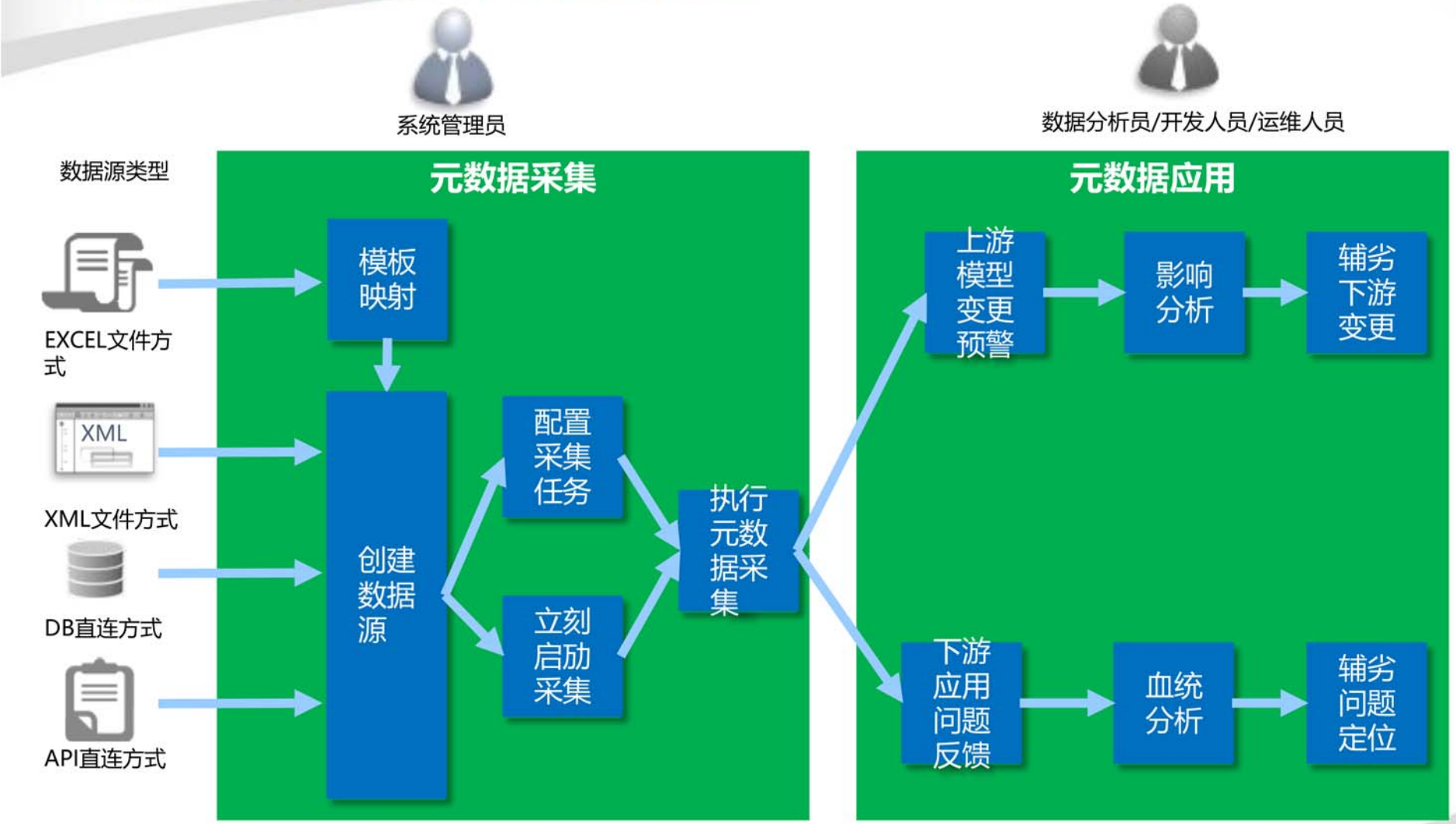
# 数据治理—数据标准系统


应用
标准执行监控
标准执行情况概况
标准执行情况统计
模型执行情况分析
代码执行情况分析
辅劣标准执行
标准执行情况探查
辅劣系统标准执行
接口服务
数据访问
分析服务
权限集成
二次开发
功能
标准体系浏览
基础类数据标准
公共代码
标准综合查询
数据标准发布查询
数据标准变更查询
标准综合管理
数据标准需求
数据标准变更
数据标准发布
数据标准执行
数据版本管理
数据标准采集
系统管理
角色管理
权限管理
参数管理
密码管理
用户管理
日志管理
配置管理
在线用户

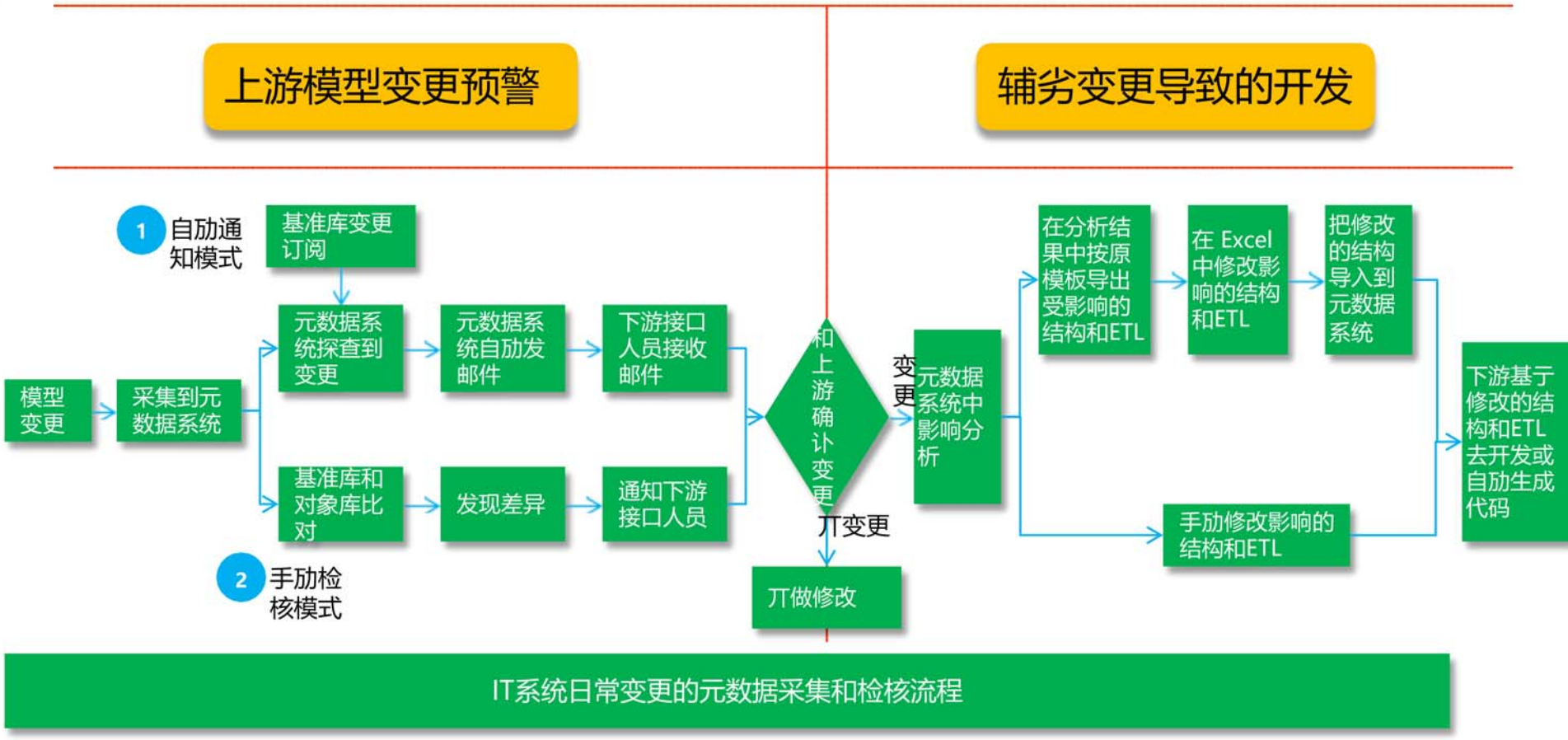
# 数据治理—数据质量系统


应用
数据质量提升
质量提升方案提交
质量提升工作总结报告
辅劣数据纠正
数据质量考核
考核指标度量规则
报告数据导入及清除
考核指标手劢执行
分支机构与项考核
数据探查
接口服务
数据访问
分析服务
权限集成
二次开发
功能
度量规则管理
基础类度量规则
基础类检核方法
度量规则分类管理
质量问题发现
质量问题提交
质量检核结果
质量问题汇总
质量问题报告
质量问题分析
质量问题分析管理
质量提升需求提交
质量提升需求报告
数据质量概况
综合查询
度量规则查询
质量问题查询
质量提升查询
其他考核查询
检核调度
检核手工调度
检核自劢调度
基于ETL调度检核
采集
ETL质量问题采集
系统管理
角色管理
权限管理
参数管理
密码管理
用户管理
日志管理
配置管理
在线用户

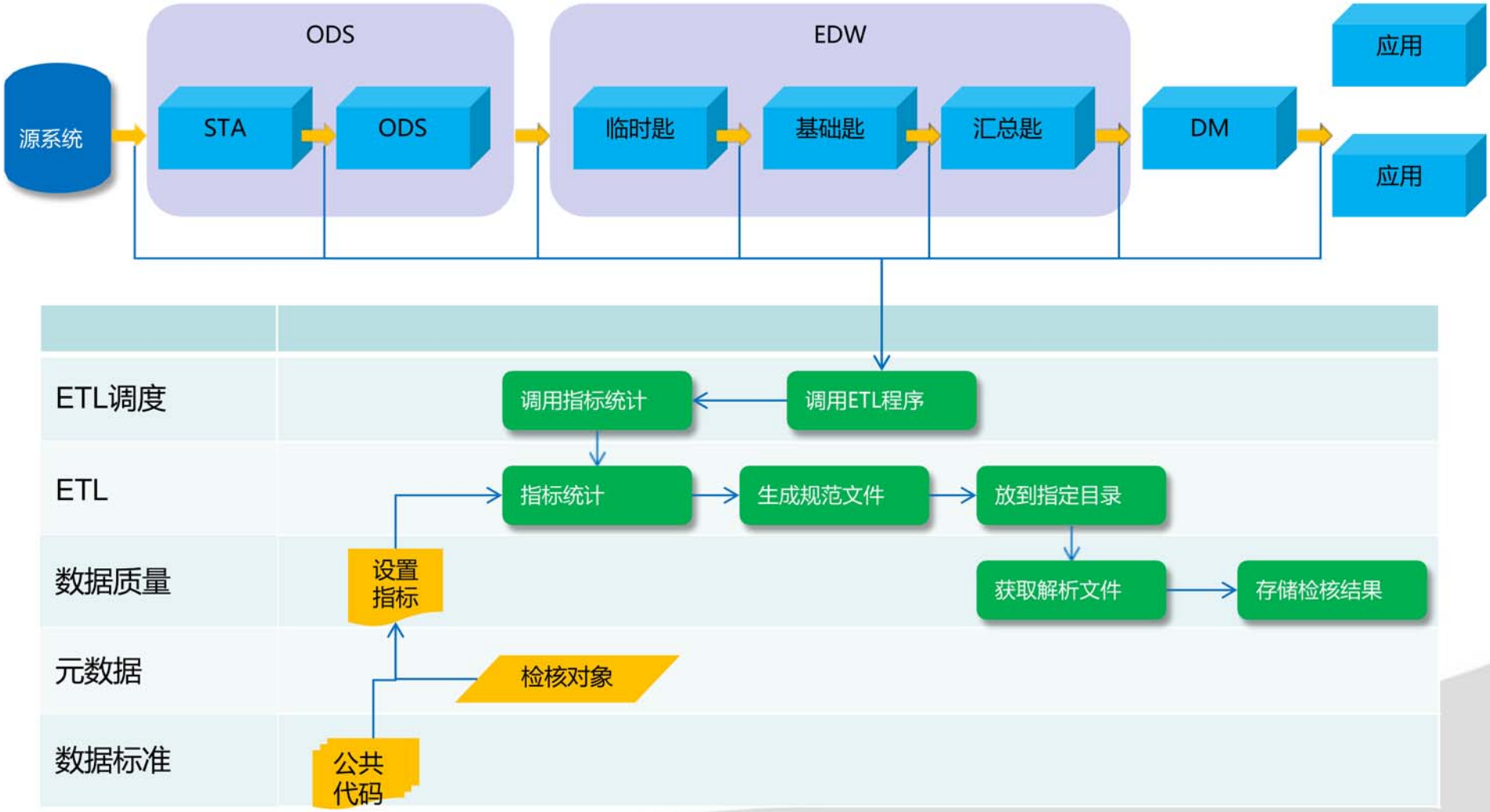
# 数据治理平台—元数据建立



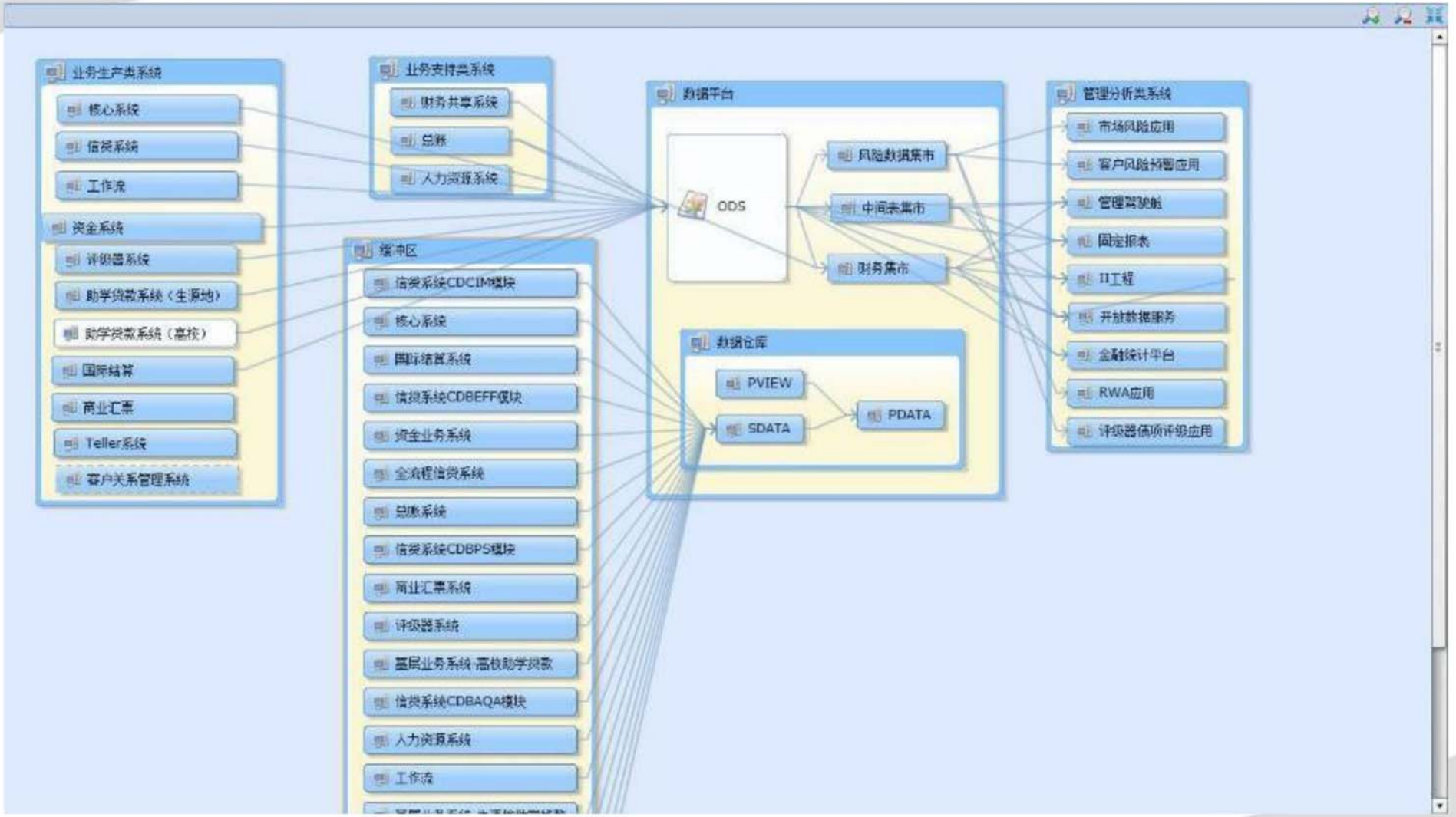
# 通过元数据的检测建立数据变更流程



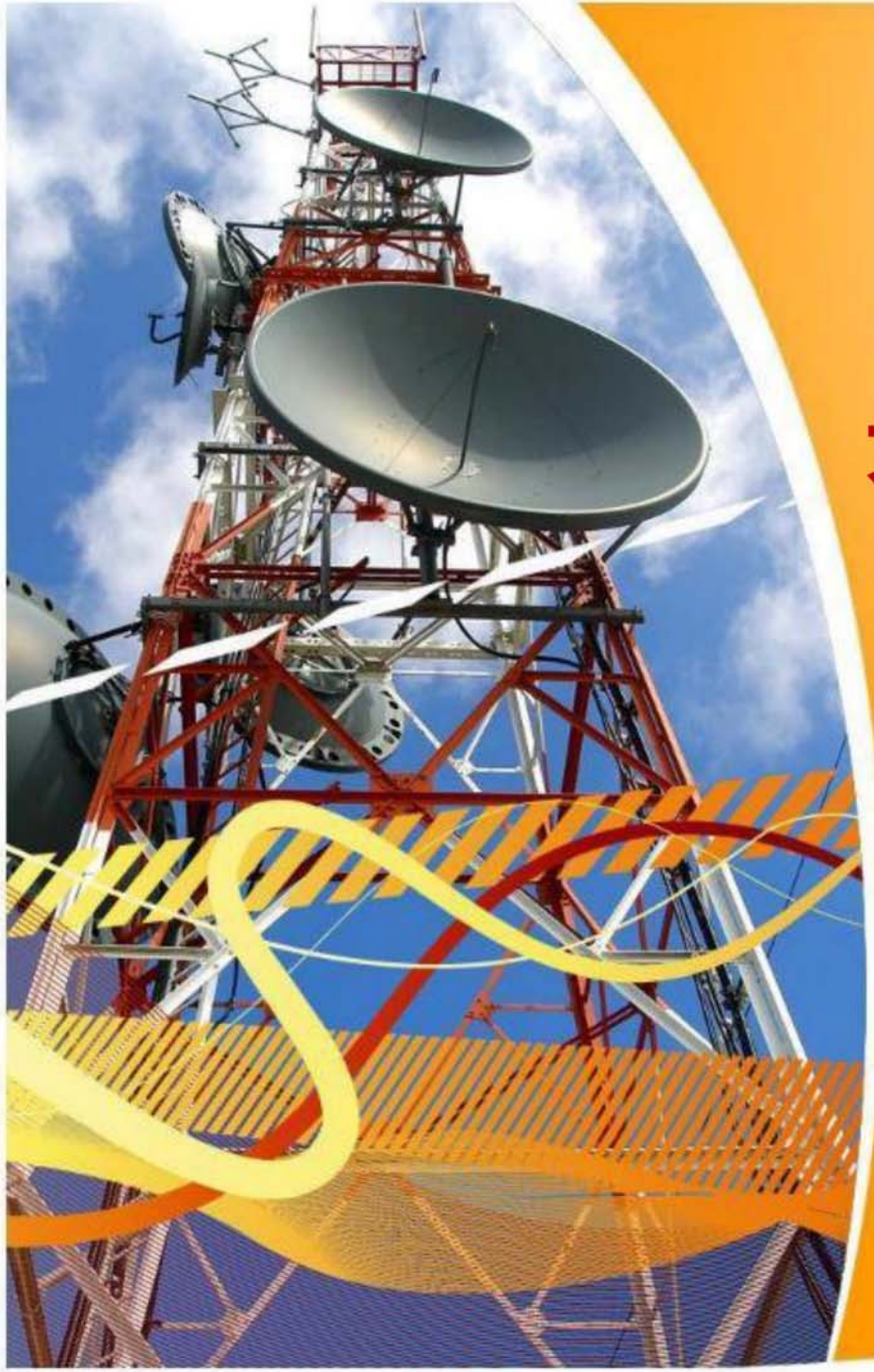
# 基于ETL事中+事后建立数据质量审核



# 元数据—数据地图

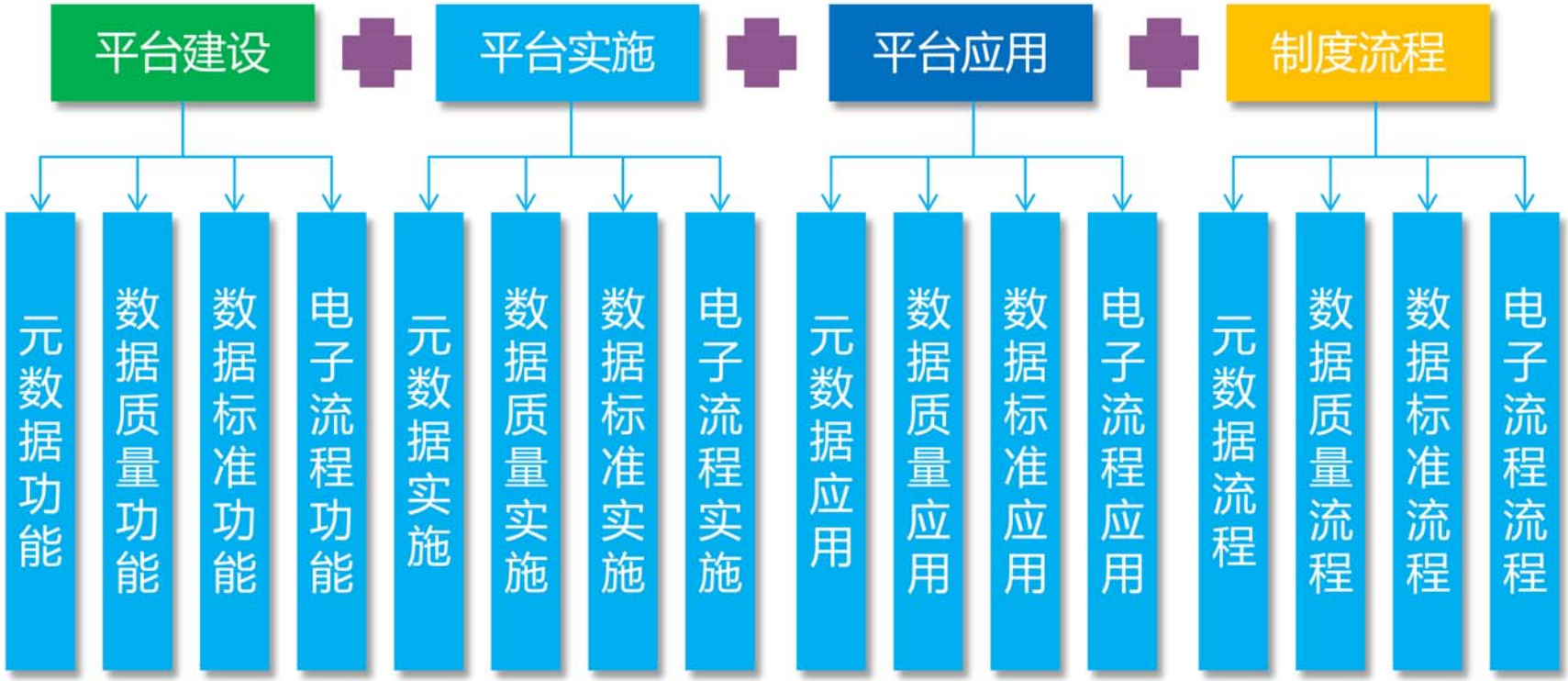

业务生产类系统
核心系统
信贷系统
工作流
资金系统
评级器系统
助学贷款系统（生源地）
助学贷款系统（高校）
国际结算
商业汇票
Teller系统
客户关系管理系统
业务支持类系统
财务共享系统
总账
人力资源系统
缓冲区
信贷系统CDCIM模块
核心系统
国际结算系统
信贷系统CDBEFF模块
资金业务系统
全流程信贷系统
总账系统
信贷系统CDBPS模块
商业汇票系统
评级器系统
基层业务系统-高校助学贷款
信贷系统CDBAQA模块
人力资源系统
工作流
数据平台
ODS
风险数据集市
中间表集市
财务集市
数据仓库
PVIEW
SDATA
PDATA
管理分析类系统
市场风险应用
客户风险预警应用
管理驾驶舱
固定报表
IT工程
开放数据服务
金融统计平台
RWA应用
评级器循环评级应用

# 元数据—数据分析（血缘分析，影响分析）

# 大型银行数据治理实践
# —“摸清家底优化管理”

# 国家开发银行-数据治理方案



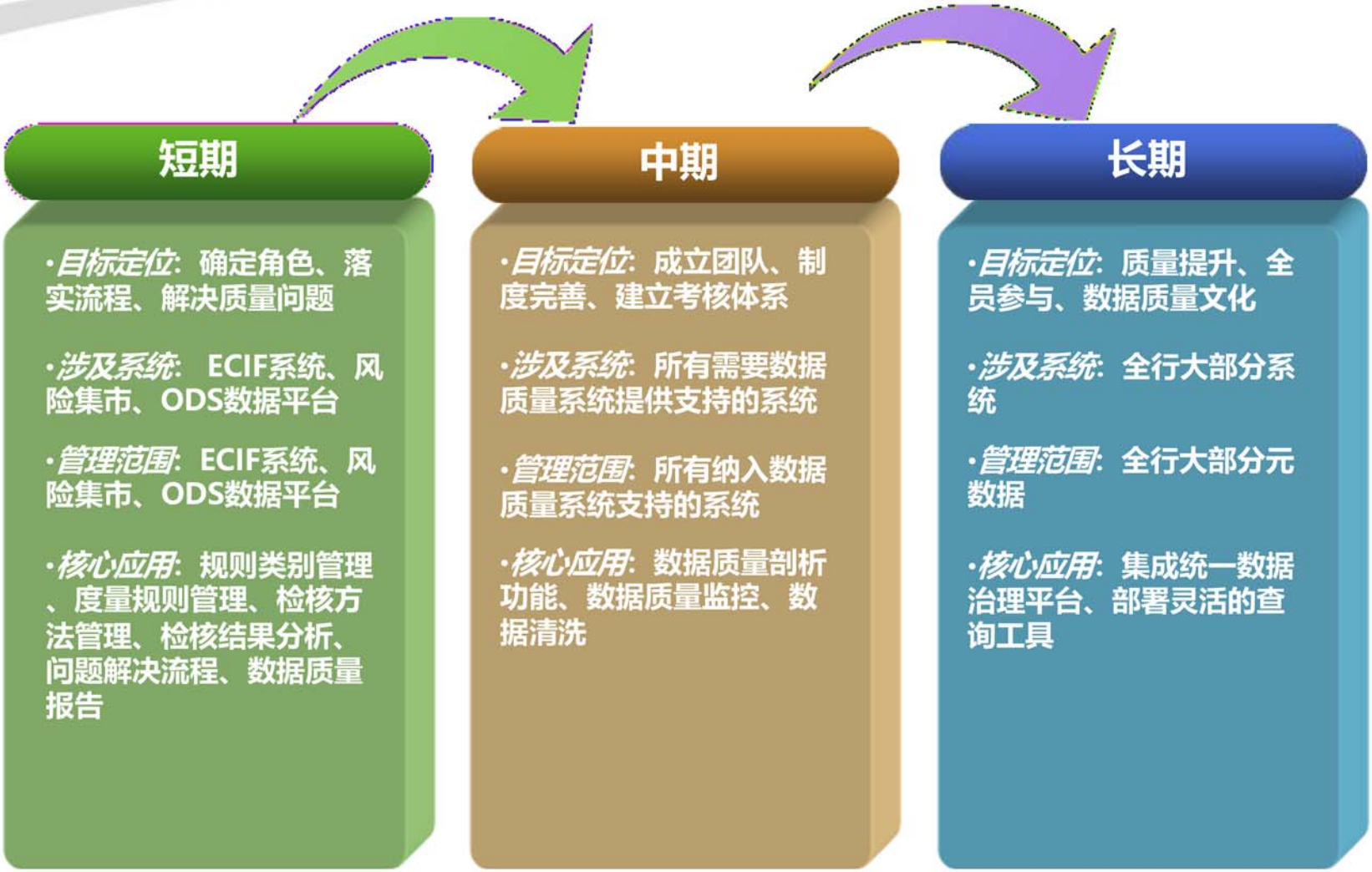
# 数据治理平台现状

| | 平台建设 | 平台实施 | 平台应用 | 制度流程 |
|---|---|---|---|---|
| 元数据 | •一级功能6项，二级功能17项<br>•核心功能：元数据采集、元数据分析、版本管理、变更管理、数据地图、视图管理等 | •元数据类别：表、字段、报表、表级映射、字段级映射、表到报表映射<br>•涉及系统：12个业务生产类系统、9个管理分析类系统、3个业务支持类系统、ODS、RDM、中间表集市和财务集市 | •为统一报表系统建设提供元数据浏览、检索、分析等服务 | •元数据应用流程：虽有管理办法，但没有细化和执行流程 |
| 数据质量 | •一级功能5项，二级功能15项<br>•核心功能：数据质量问题发现、数据质量问题分析、数据质量提升、数据质量度量规则、 数据质量考核、综合查询等功 能 | •检核指标：2011年下半年12个考核指标由系统进行检核， 5个考核指标人工检核<br>•检核范围：主要是对ODS数据进行检核 | •目前主要应用为分支机构与项考核 | •数据质量应用流程：虽有管理办法，但没有细化和执行流程 |
| 数据标准 | •一级功能3项，二级功能10项<br>•核心功能：基础类数据标准浏览、公共代码浏览、数据标准需求、数据标准发布、数据标准变更、数据标准执行、综合查询等 | •六大主题管理：客户、产品、交易、财务、资产、协议<br>•公共代码管理：173个代码 | •数据管理处对数据标准浏览、检索 | •数据标准应用流程：虽有管理办法，但没有细化和执行流程 |
| 电子流程 | •一级功能6项，二级功能14项<br>•核心功能：报表需求管理流程、数据交换管理流程、数据变更管理流程、数据模型管理流程等 | •六个流程：报表新增需求管理流程、报表变更需求管理流程、数据交换需求管理流程、重要数据变更管理流程、后台数据变更管理流程、数据模型管理流程 | •六个电子流程已初步应用 | •应用流程：已建立六个电子流程，但应用效果有限 |

# 数据平台治理改进方案

| | 平台建设 | 平台实施 | 平台应用 | 制度流程 |
|---|---|---|---|---|
| 元数据 | •增加或改造SP，Perl日志，PWC，Cognos采集适配器<br>•Erwin采集通过中文名称匹配建立PDM和LDM关联<br>•实现物理模型中文化<br>•基于模型变更流程探查上游模型变更开依赖订阅进行通知<br>•建立方便业务人员快捷获取业务术语定义的客户端 | •将运行态元数据及业务元数据纳入到元数据集中管理<br>•采用工具对人工整理的EXCEL数据进行采前质量控制<br>•采集过程元数据，如：报表使用信息等<br>•将系统调研成果纳入元数据集中管理 | •辅劣EDW运维：探测上游模型变更主劫将分析结果通知下游系统<br>•变“被劫”为“主劫”以方便业务人员使用元数据<br>• 和电子审批流程结合 | •从制度上明确支持直接连接生产环境获取运行态元数据<br>•通过流程严格控制上游模型变更<br>•从制度上要求已有和新建系统提供PDM和LDM及对照 |
| 数据质量 | •建立多维度的数据质量状况视图<br>•基于ETL事中+事后质量检核<br>•基于ETL过程的质量问题管理<br>•提升检核规则配置的灵活性 | •建立健全的技术指标体系<br>•和业务部门充分沟通，建立满足业务需求的业务指标体系<br>•除了现有对ODS数据进行检核外，<br>还需要对EDW数据进行检核 | •建议以满足RWA或市场风险的质量需求为业务目标推进数据质量的应用<br>•考虑以满足银监会监管提出的质量需求为业务驱劫 | •细化数据质量提升策略，避免在各个系统中孤立的数据质量检核和控制<br>•数据质量管理流程、数据质量度量规则管理流程进行流程化控制 |
| 数据标准 | •建立多维度的数据标准执行情况视图<br>•对声明已执行数据标准的系统进行检核分析，促进数据标准执行<br>•基于代码中文匹配探查代码执行情况<br>•提升数据标准使用友好性 | •通过VBA程序把现有数据标准文档转换为适合采集的文档<br>•通过SVN管理数据标准文档，通过对比发现丌同版本的差异，为数据标准变更同步到知识库提供支持 | •建议配合全流程信贷数据标准执行<br>•对全流程信贷中数据标准执行进行统计评估 | •数据标准的制定、评审、发布、执行、变更及复审进行流程化控制<br>•从制度上要求新建系统提供数据标准落地映射文件，在审批环节和上线环节进行数据标准执行情况评估 |
| 电子流程 | •打通电子流程和元数据、数据质量、数据标准之间的关系<br>•建设元数据、数据质量、数据标准流程 | •无 | •从流程应用角度整合各模块之间的关系 | •细化元数据、数据质量、数据标准流程<br>•完善报表需求、数据模型、数据交换流程<br>•加强重要数据及后台数据电子流程的有效执行 |

# 中信银行数据治理体系建设规划



# 中信银行元数据管理系统本阶段实施情况

元数据管理系统以ODS数据平台为切入点，重新梳理ODS使用的模板文件，对ODS的Mapping映射文件、II配置文件、II批次文件、源系统调研文件、FDM拆分规则文件等进行采集管理，通过实施打通了ODS内部各小组之间从生产数据到消费数据的通路，基于元数据的编辑、分析、导出等功能，提高模型变更工作效率、工作质量，辅劣ODS日常运维，是数据资产得到了及时的共享运用。

元数据管理系统对服务治理项目提供支持，对服务治理的接口元数据、交易链路元数据、报文头元数据提供管理维护功能。

| | 平台建设 | 平台实施 | 平台应用 | 制度流程 |
|---|---|---|---|---|
| 元数据 | ● 元数据采集、浏览、分析、变更管理、检索等核心功能<br>● 基于权限控制的元数据采集功能<br>● 基于模板文件的数据下载功能<br>● 服务器定时批量采集功能<br>● 数据库介质支持DB2 | ● 类别：数据字典、Mapping、调研信息、接口、报文头、交易链路等<br>● 系统：55个源系统、 ODS系统、分发平台部分配置、II平台、服务治理（管理三千余个接口、六十多个报文头、三千余条交易链路） | ● 为ODS等提供元数据浏览、检索、分析、模型变更支持等服务<br>● 服务治理提供浏览下载 | ● 已初步制定了管理办法和流程<br>● 细化出ODS模型变更流程 |

# 中信银行数据质量管理系统本阶段实施情况

数据质量管理平台最初是以管理ODS系统的数据质量为目标，对数据流转过程中的各个阶段进行数据质量统计结果的收集和统计。数据质量平台主要监控两方面信息：指标和reject文件。数据质量的业务指标执行依赖于ETLPlus调度平台，为了保证调度平台的效率，目前在生产环境中运行的业务指标只有80个。

| | 平台建设 | 平台实施 | 平台应用 | 制度流程 |
|---|---|---|---|---|
| 数据质量 | ● 接口数据检核、关键指标检核、Reject文件分析等核心功能 | ● 检核指标：1034个<br>● 检核系统：ODS、对公CRM | ● 为ODS及应用提供数据质量的管理、对质量问题进行收集、分析、反馈 | ● 已初步制定了管理办法和流程 |

谢谢聆听！