DTCC

第十一届中国数据库技术大会

DATABASE TECHNOLOGY CONFERENCE CHINA 2020

架构革新 高效可控

IT168.com ChinaUnix.net ITPUB

北京国际会议中心 | 2020/12/21-12/23

元数据治理在企业中的实践

主讲人：何军

数据平台开发部负责人

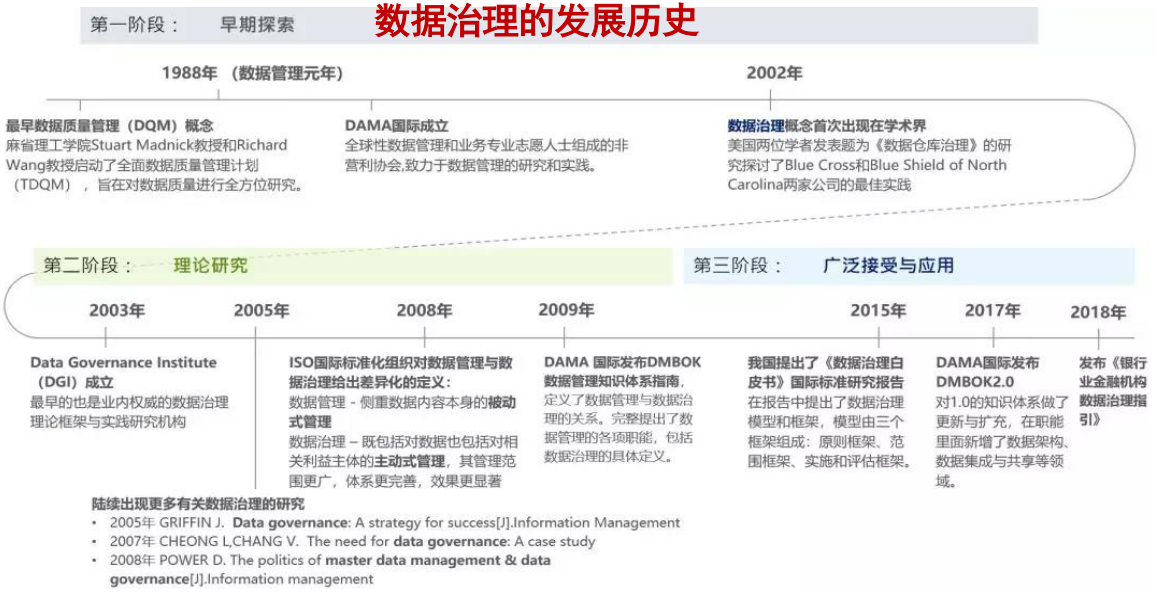
目录

数据治理简介和解读

数据治理的内涵与外延

狭义：数据治理是指对数据质量的管理、专注在数据本身。

广义：数据治理是对数据的全生命周期进行管理，包含数据采集、清洗、转换等传统数据集成和存储环节的工作，同时还包含数据资产目录、数据标准、质量、安全、数据开发、数据价值、数据服务与应用等，围绕数据生命周期而开展的业务。技术和管理活动都属于数据治理范畴。有的专家干脆把广义的数据治理称为数据资产管理。



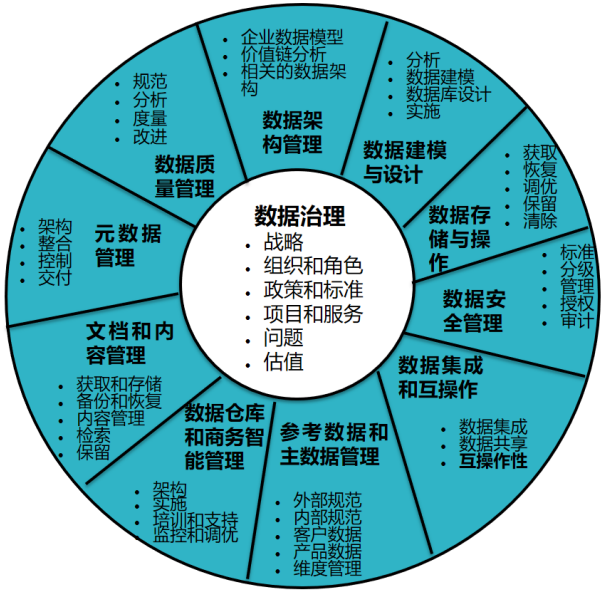
利用比喻介绍数据治理各管理域

- ❖数据架构和模型管理------要把水果摆放的紧凑、稳定、便于取放
- ❖数据标准管理------水果规格一致
- ❖数据质量管理------保证没有烂水果
- ❖元数据管理------描述水果的品种、价钱和产地
- ❖主数据管理------往往是高价值、受欢迎而放在专区的水果
- ❖数据生命周期管理------水果保质期管理、下架管理
- ❖数据安全管理------防止被人偷吃、损坏
- ❖数据服务与应用------导购精准服务
- ❖数据资产管理------盘点有哪些水果，数量多少，价值多少

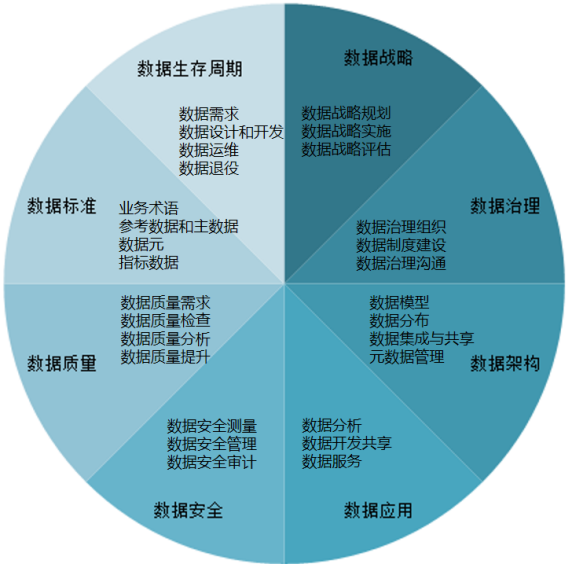
数据治理范围的权威参考

- DAMA是目前国际上权威的数据治理框架，2018年我国发布数据管理能力成熟度评估模型，提供数据管理能力的重要标准。

国际DAMA数据治理框架



国内数据管理能力成熟度评估模型



其他比较权威的参考：中国信通院2019年发布的《数据资产管理白皮书3.0》

企业中的数据治理工作范围

数据治理流程

- 组织架构
- 职责分工
- 体系规划
- 管理制度
- 策略执行
- 考核评估

数据治理专项工作

- 数据标准管理
- 元数据管理
- 数据架构管理
- 主数据管理
- 数据安全管理
- 数据质量管理
- 数据生命周期管理

数据治理工具与技术

- 数据标准管理工具
- 元数据管理工具
- 数据架构工具
- 主数据管理工具
- 数据安全管理工具
- 数据质量管理工具
- 数据生命周期管理工具
- 流程管理工具

企业开展数据治理工作的切入点

很多企业在投资建设新的数据仓库、核心系统、大数据平台、数据中台时或者开展主数据管理时，会考虑开展数据治理工作，建立数据标准，进行数据质量检查，盘点各系统现状，保证数据迁移有序，保证新平台从建设之初数据就规范可控。

根据数据治理工作的特点，由哪个部门来做？

缺点

01 **监管性**
数据治理工作具有监管特性，类似安全部门或质量监管部门，具有执法权

→ 质量与效能部

1.缺乏规划能力
2.缺失数据知识
3.缺少业务知识

02 **规划性**
数据治理工作具有规划特性，类似系统规划部门，具有立法权

→ 系统规划部

1.缺少运营经验
2.缺失数据知识
3.不是工作重点

03 **专业性**
数据治理工作需要数据的专业知识和业务知识，尤其要求对于数据技术有一定积累

→ 大数据部

1.事关私利，很难执行
2.缺失行业业务知识
3.对于业务系统统筹力度小

元数据管理实践分享

元数据的定义

元数据 是描述数据的数据，根据描述内容，可分为业务元数据、技术元数据、管理元数据。

业务属性

- 中文名称
- 英文名称
- 业务定义
- 业务规则
- ……

技术属性

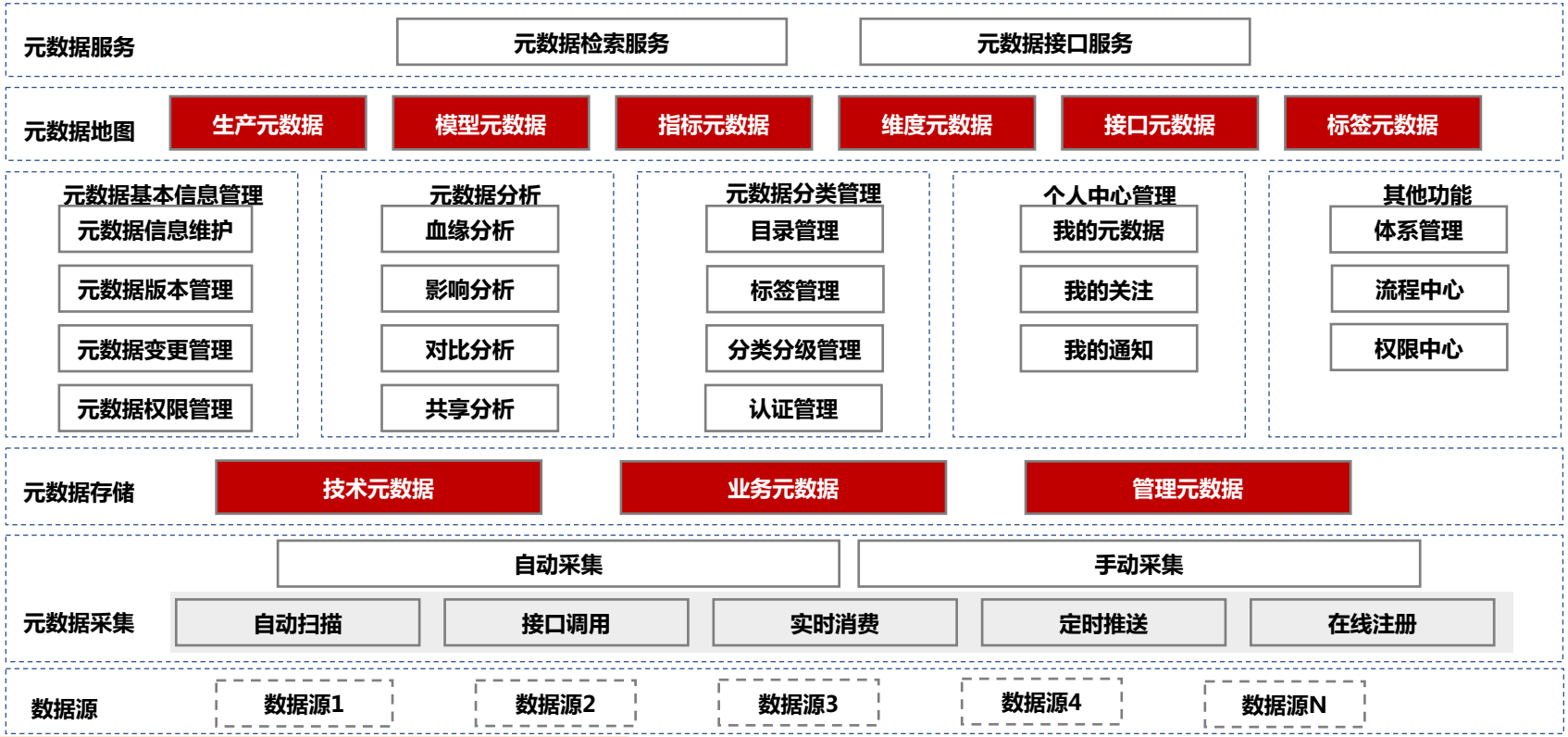
- 存储位置
- 存储周期
- 数据类型
- 数据长度
- ……

管理属性

- 数据开发负责人
- 数据业务负责人
- 数据使用部门
- 是否公司统一
- .....

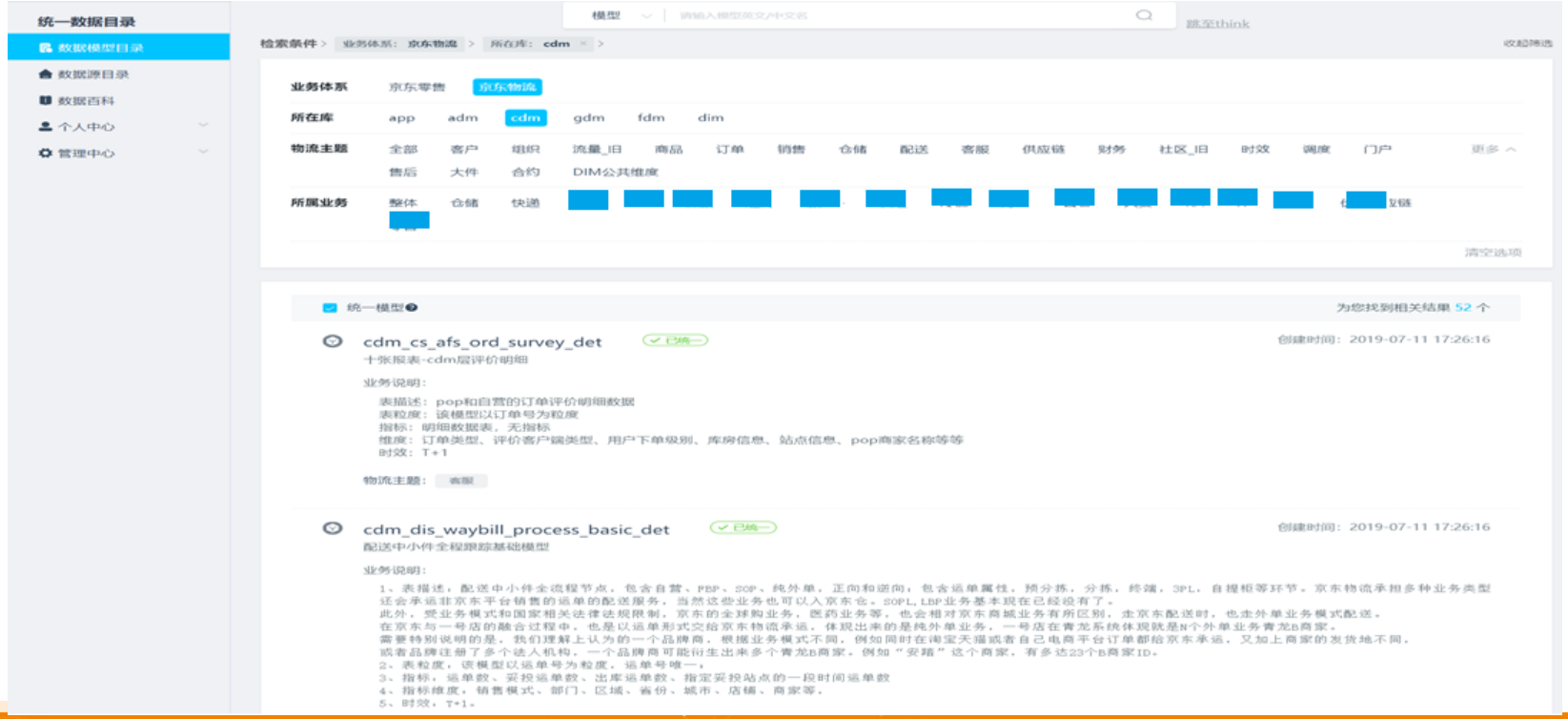
元数据是帮助查找、存取、使用和管理信息资源的信息。元数据是业界公认的数据管理中的核心要素，做好元数据管理，更容易的对数据进行检索、定位、管理、评估。

元数据管理功能架构


元数据服务
元数据检索服务
元数据接口服务
元数据地图
生产元数据
模型元数据
指标元数据
维度元数据
接口元数据
标签元数据
元数据基本信息管理
元数据信息维护
元数据版本管理
元数据变更管理
元数据权限管理
元数据分析
血缘分析
影响分析
对比分析
共享分析
元数据分类管理
目录管理
标签管理
分类分级管理
认证管理
个人中心管理
我的元数据
我的关注
我的通知
其他功能
体系管理
流程中心
权限中心
元数据存储
技术元数据
业务元数据
管理元数据
元数据采集
自动采集
手动采集
自动扫描
接口调用
实时消费
定时推送
在线注册
数据源
数据源1
数据源2
数据源3
数据源4
数据源N

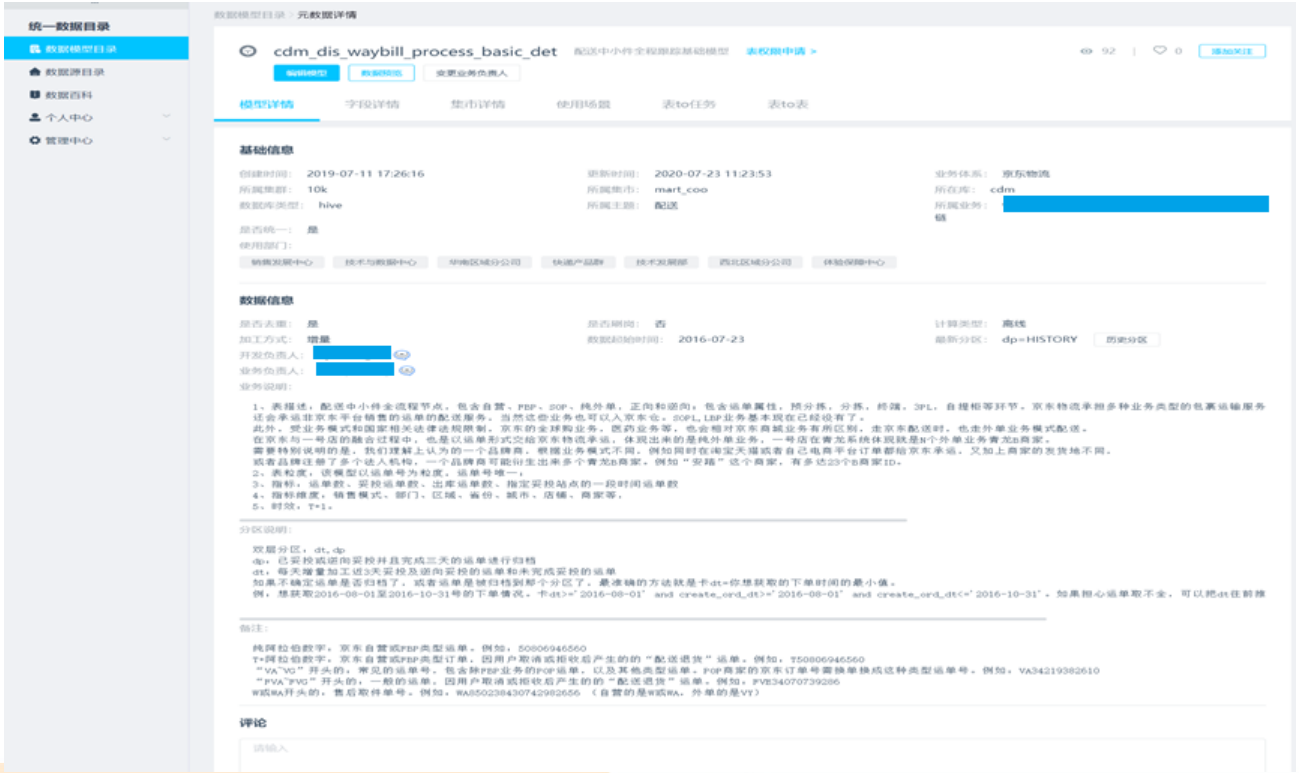
元数据管理实践分享（以模型元数据为例）

通过筛选条件：业务体系、所在库、所属主题、所属业务，选择要关注的模型元数据



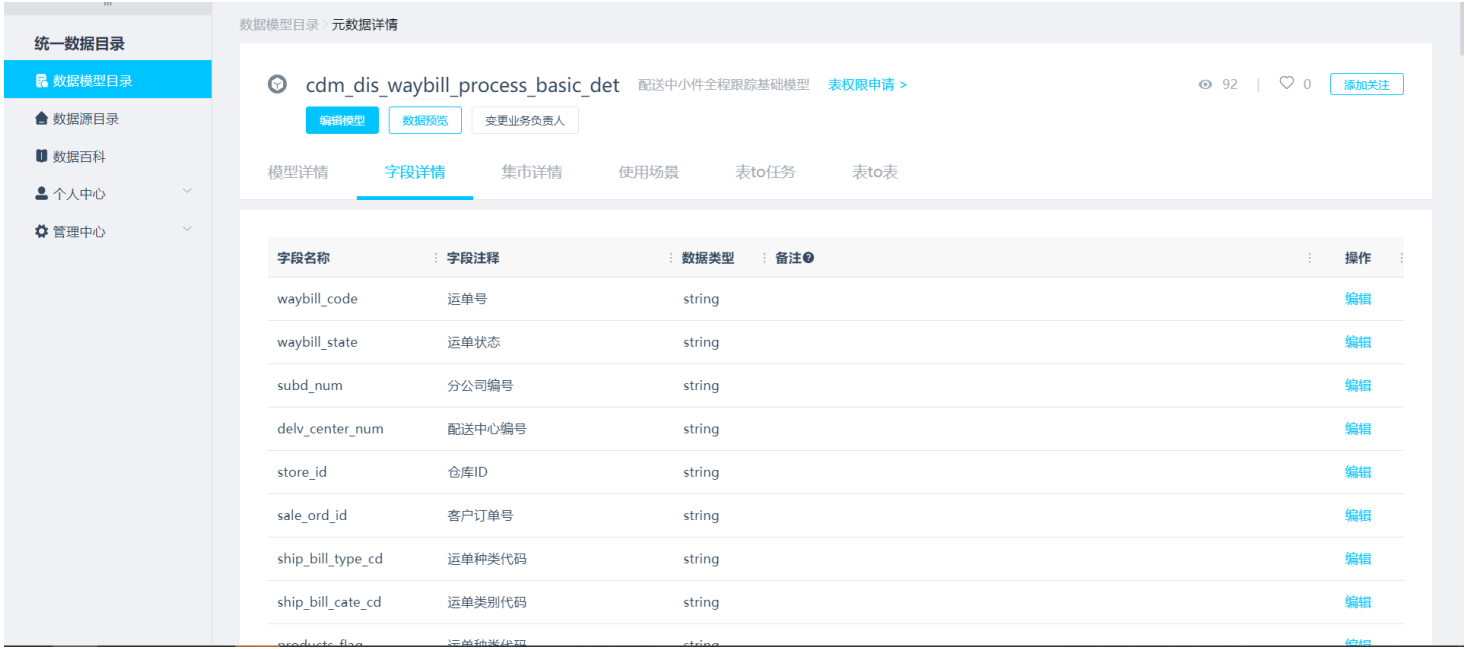
元数据管理实践分享

模型元数据详情-模型详情



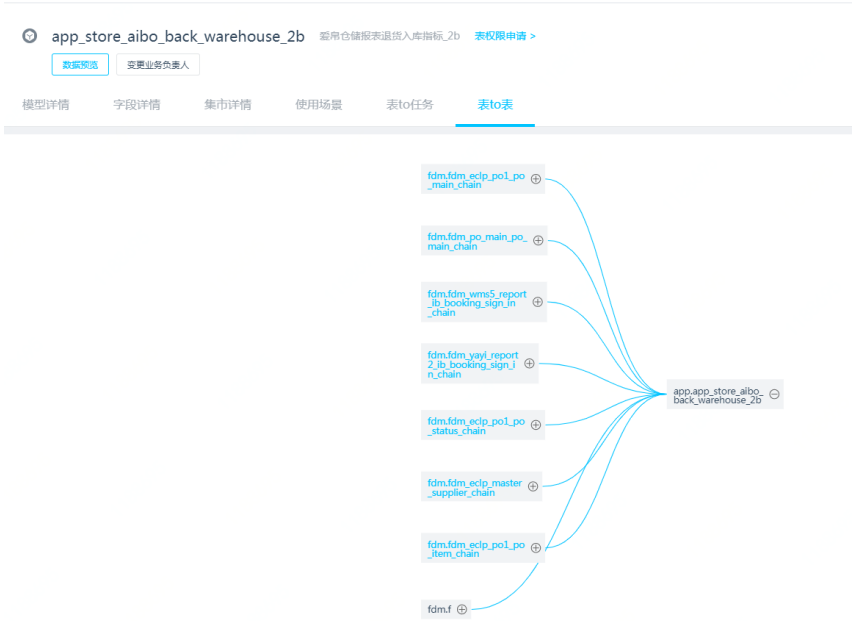
元数据管理实践分享

模型元数据详情-字段详情



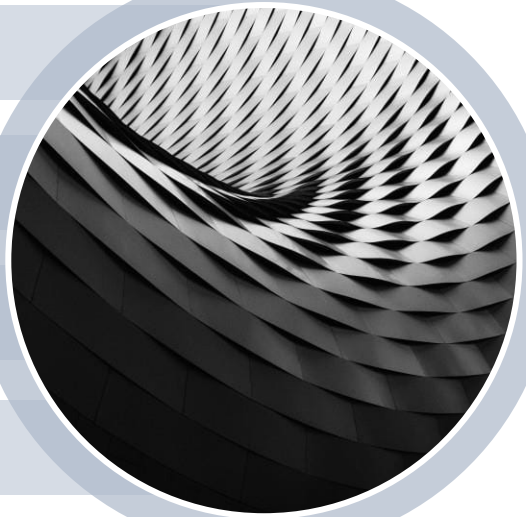
元数据管理实践分享

模型元数据详情-表To任务、表To表



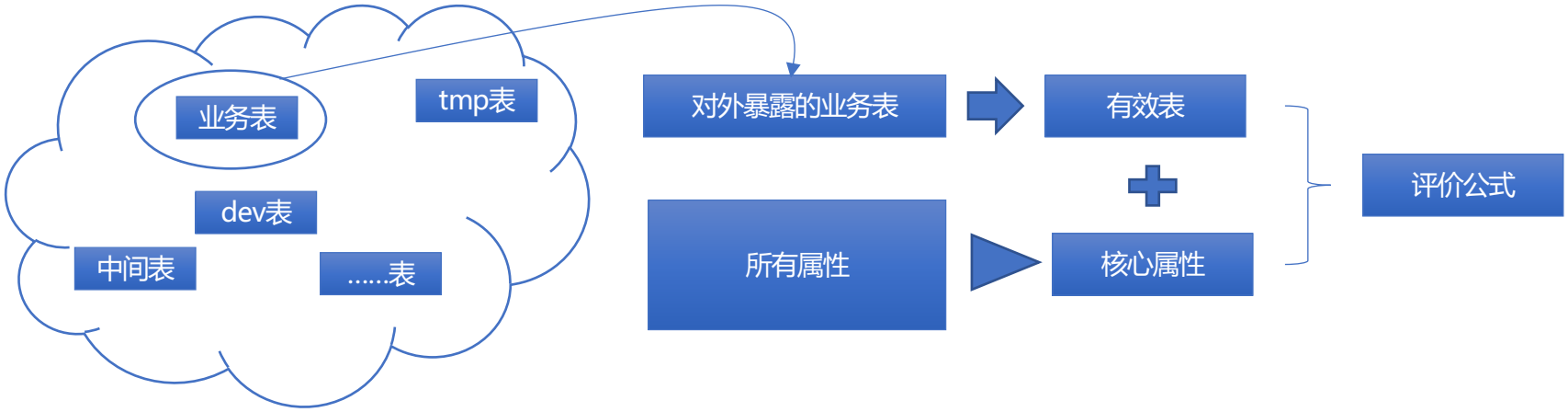
元数据在企业中的价值

1. 构建企业的数据知识库，可以跨产品线了解数据资源
2. 是数据民主化的重要载体，非研发同学也可以了解企业的数据
3. 血缘关系方便上下游数据问题定位，数据变更能精确通知下游同学
4. 助力企业盘点数据资产，是数据资产化的基础



元数据质量管理的思考

衡量元数据质量的评价标准



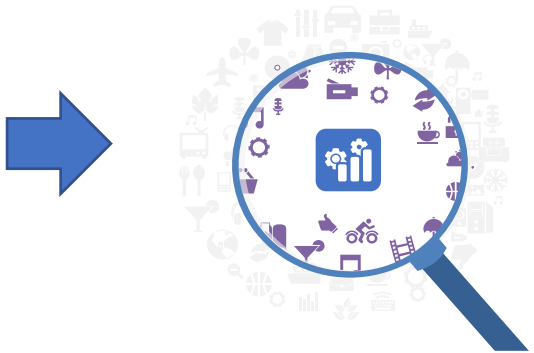
评价公式

AVG（有效表的已经覆盖元数据核心属性总量/有效表元数据核心属性总量）

针对核心属性，构建自动化检测工具

检查规则

- 缺少业务负责人
- 缺少所属主题
- 缺少所属业务
- 缺少是否统一标签
- 缺少字段注释
- 业务说明未按格式填写
- ……



整改
（会涉及到IT系统和DT系统）

保证元数据质量，引入长效机制

两个原则：

- **元数据采集尽量自动化**
- **元数据维护尽量源头化**

生产系统表设计阶段，明确表的中文名，表的业务说明，表字段的中文注释，表的开发负责人，表的业务负责人。

↓

生产系统表建表阶段，要求DBA通过工具检查建表脚本，来审核上述各项是否具备。

↓

数仓模型开发阶段：ODS表，继承了生产系统表相关的元数据属性，ODS以上各层表的开发，数据开发人员在大数据平台上，完成模型开发，建的表要包含表的中文名，字段注释等信息。之后，系统会自动把元数据信息，同步到元数据系统，开发人员在元数据系统中，维护业务说明等相关元数据信息。

↓

数仓模型测试阶段：测试人员检查元数据系统中，相关元数据信息的完整性，如果发现有元数据信息缺失的情况，提交问题单，驱动开发人员补齐元数据。

THANKS