







数据质量和数据治理的关系 | 京东云技术团队

京东云开发者 2024-01-30 ◎ 1,765 ⑤ 阅读8分钟

▼ 団 智能总结 ① □ 复制 C 重新生成 这篇文章主要探讨了数据质量和数据治理的关系。指出数据治理不是单纯的数据清洗,数据质量有多个维度。数据 治理为数据管理提供指导,二者是共生关系,存在重叠也有各自侧重。无论有无显性治理,都可能存在数据质量问 题或隐形治理。数据质量常是数据治理的驱动因素,二者并非相同。 关联问题: 数据清洗如何进行 数据治理如何起步 怎样衡量数据质量 基于该文章内容继续向AI提问

很多不太了解的人会认为:数据治理就是干数据清洗的。

近两年,在我们公司,数据治理团队在数据降本方面做的比较多,效果还不错,我们很多人可能以为:数据 治理就是做数据清理的。

在京东科技集团数据治理工作组第一次全体会议上,我就讲过数据治理和数据清洗之间的关系:

数据清洗只是指通过识别和纠正数据中的错误、缺失、重复或不一致等问题,确保数据质量的过程。

我的观点是数据质量和数据治理是一体的,没有数据治理,就不可能有良好的数据质量。通过进行数据治 理,我们实现了数据质量。怎么会这样? 那是什么意思? 让我们来了解一下。

1数据质量管理

更准确地说,数据质量或数据质量管理侧重于确保数据符合我们的数据质量维度。数据质量有很多维度,比 如:

•完整性:是否填写了所有相关字段?

•有效性: 所有值都符合吗? 地址字段是否在正确的顺序和正确的拼写?

•准确性:数据是否反映了现实世界中的人或物体?米老鼠可能不是一个真正的销售对象。

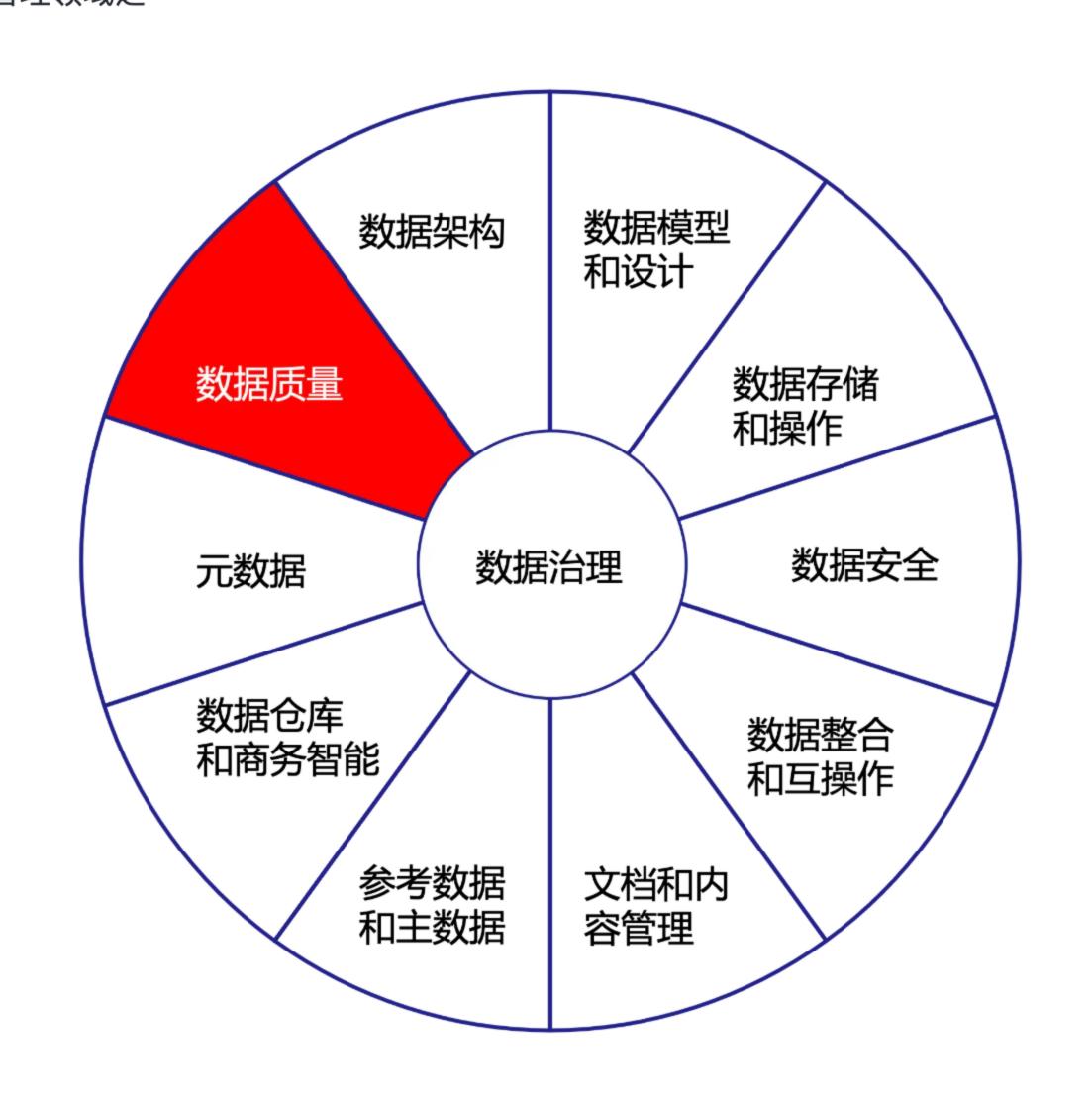
•一致性:数据是否与理解的模式一致?例如,出生日期的数据格式应为 YYYY/MM/DD/。

•唯一性:是否有重复的记录?

•及时性:它是最新的吗?

数据质量确保我们的数据符合这些维度。或者简单地说,数据质量管理确保我们拥有高质量的数据,干净的 数据。

让我们扩展我们对数据质量管理的理解,并从DAMA国际的角度来看待它。数据质量是国际数据管理协会确 定的11个数据管理领域之一:



1.数据治理通过建立一个考虑到企业需求的数据决策系统,为数据管理提供指导和监督。

2.数据架构定义了管理数据资产的蓝图,通过与建立战略数据需求的组织策略保持一致,并设计满足这些需 求。

3.数据建模和设计是一个发现、分析、表示和交流精确形式的数据需求的过程。

4. 数据存储和操作包括设计、实现和支持使其价值最大化。操作在整个数据生命周期中提供支持,从规划 数据到处理数据。



目录 收起 へ 1数据质量管理 2 无论治不治理,数据质量都在那里 2.1 真的没有数据治理 2.2 有隐形的数据治理 3 数据质量和数据治理关系 4 总结

相关推荐 自学Flink之路----flink的优势 1.2k阅读·0点赞 只需三步实现Gateway结合Sentinel实现... 1.6k阅读 · 8点赞 数据中台建设实践(二)-数据治理之数... 2.0k阅读·25点赞 Kafka 如何避免重复消费? 277阅读 · 17点赞 spring boot 启动流程分析 24k阅读 · 40点赞

精选内容 【圖分布式锁通关指南 02】基于Redis... 别惹CC·44阅读·1点赞 CompletableFuture#allOf、依次 join、... 桦说编程 · 65阅读 · 3点赞 ip 命令创建 vlan bond 子接口 bobz965 · 15阅读 · 0点赞 多租户的 4 种常用方案 苏三说技术 · 608阅读 · 13点赞 别再瞎摸索! POI 操作 Excel 轻松添加... isfox · 43阅读 · 1点赞

找对属于你的技术圈子 回复「进群」加入官方微信群 11.

@稀土掘金技术社区

5.数据安全确保数据隐私和机密性得到保护,数据不被破坏,数据被适当访问。

6.数据集成和互操作性包括与数据存储、应用程序和组织内部之间的数据移动和整合相关的过程。

7.文档和内容包括计划、实施和控制活动,用于管理在一系列非结构化媒体中发现的数据和信息的生命周期,特别是支持法律和法规遵从性要求所需的文档。

8.参考数据和主数据包括核心关键共享数据的持续核对和维护,使系统能够一致地使用关于基本业务实体的最准确、最及时和最相关版本的真相。

9.数据仓库和商务智能包括计划、实施和控制过程,以管理决策支持数据,并使知识工作者能够通过分析和报告从数据中获取价值。

10.元数据包括规划、实施和控制活动,以便能够访问高质量的综合元数据,包括定义、模型、数据流和其他对理解数据及其创建、维护系统至关重要的信息并访问。

11.数据质量包括质量管理技术的计划和实施,以测量、评估和提高数据在组织内的适用性。

根据DAMA的说法,数据质量包括质量管理技术的计划和实施,以测量、评估和提高数据在组织内的适用 性。

实际上,数据质量在大多数其他数据管理领域中都发挥着作用。要保证数据安全,就不可能没有好的数据质量,元数据与数据质量的关系也是双向的,数据架构也会对数据的质量起到一定的作用,反之亦然。处于中心地位的数据治理确保所有这些数据管理领域结合在一起。

2 无论治不治理,数据质量都在那里

到现在为止,你可能会说:"好吧,我有点明白了,但我们没有做数据治理,数据质量也挺好。"如果是这样的话,我认为有两种可能的现实:

- •真的没有数据治理
- •有隐形的数据治理

让我们更详细地看一下这两个案例。

2.1 真的没有数据治理

如果数据治理不存在,那么我们可能会遇到以下一种、部分或全部情况:

- •数据质量不是企业级全域的: 你也可能认为是覆盖企业范围的,比如ERP,这并不意味着数据质量是企业范围的。尤其是在像我们这样的大型组织中,因为可能存在数据质量规则,这些规则仅根据该企业系统的利益相关者的输入创建,但这可能会影响那些不是系统利益相关者或用户的人,比如有集成相关的系统用户,也可能是数仓或者下游数据用户。
- •缺乏通用标准:可能会创建数据质量标准,但前提是它们与特定部门、业务单位或系统的需求有关。一旦 其他部门或系统加入到数据质量计划的工作中,这些标准就会发生冲突,需要解决方法或完全更改。
- •没有明确的角色和职责:可能会分配负责清洗和维护数据的资源,但资源类型因团队而异。当标准、定义和优先级发生冲突时,也可能不清楚谁拥有所有权,谁负责解决。
- •数据质量管理大多是被动的: 这主要是数据质量管理处于早期阶段,也意味着缺乏数据治理。数据质量问题是以被动的方式识别和处理的,并不总是解决问题的根本原因。

2.2 有隐形的数据治理

就目前我们公司的现状开看,作为产研、数仓、数据集市、数据分析等团队成员,你实际上可能会说:"好吧,我们并没有像你上面描述的那样糟糕"。您实际上可能有:

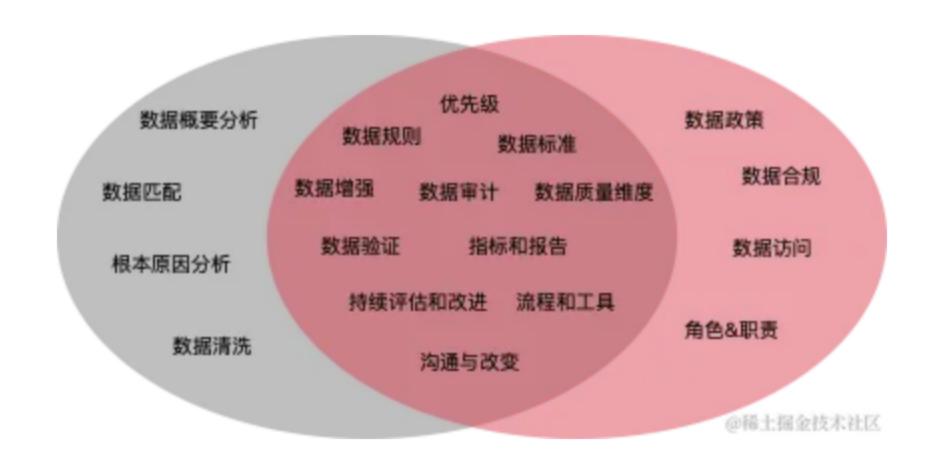
- •数据质量要求
- •数据质量校验规则
- •数据质量测试,甚至是监控
- •数据质量问题处理

那么,你可能就有了我喜欢称之为"卧底数据治理"的东西。你可能有很多数据治理的工作内容,但没有正式、正规和专业化的数据治理。

3 数据质量和数据治理关系

我想我已经从上一节中明确指出,数据治理和数据质量是共生关系。它们是同一枚硬币的两面。**没有数据治理,就不可能有良好的数据质量,而数据治理实施必须非常有效,才能从根本上解决数据质量问题。**

实际上,数据质量和数据治理之间存在相当多的重叠,如下图所示:



数据治理描述了谁需要做什么、在什么条件下对什么数据执行操作,以及使用哪些流程、程序、工具和总体最佳实践。因此,很多因素都会对数据质量产生有益的影响,但不仅如此。业务规则、数据标准、指标、角色和职责等都有利于数据质量,因此是重叠的,但不仅仅是数据质量。主数据管理、数据可访问性、数据集成、元数据管理、BI,甚至数据安全等等都有直接的好处。

当然,也有一些领域只与数据质量有关,例如:数据概要分析、数据匹配、根因分析和数据清洗。

4总结

很多时候,数据质量是数据治理的驱动因素之一,这是数据治理计划的最初重点,因此两者之间可能会混 淆。但它们并不相同,它们是同一枚硬币的两面,而不是两枚硬币。

作者: 京东科技 李然辉

来源: 京东云开发者社区 转载请注明来源

标签: 数据分析



为你推荐	
如何理解数据治理的一点见解 Ros1n │ 3年前 │ ◎ 1.9k ြ 6 ♀ 2	大数据 后端
数据治理: 说起来容易,做起来难? 这个方法论值得收藏 IT技术管理的那些事儿 3年前 ② 2.2k ① 3	数据分析
数据治理流程中,最重要的3点都在这 马哥说数据 │ 3年前 │ ◎ 1.4k ௴ 点赞 ፡፡ 评论	大数据
数据治理怎么做? 这篇万字长文终于讲清楚了! □ ③ 2.0k □ 6 □ 评论	后端
火山引擎DataLeap一站式数据治理解决方案及平台架构 字节跳动数据平台	大数据 数据库 数据分析
攻据分析前端向的业务理解 ioletJack │ 10月前 │ ◎ 153 凸 点赞	数据分析 数据可视化
数据类型错误带来严重后果!一文读懂怎么正确选择数据类型 quids数据库云服务 1年前 □ 517 □ 点赞 □ 评论	大数据 数据库 MySQL
京东大数据治理探索与实践 京东零售技术实践 京东零售技术 2月前 ◎ 96 □ 2 □ 评论	大数据
十算性能提升百倍 火山引擎数智平台这款产品助力企业员工更好看数用数型 中部	数据库 数据挖掘 运营
基于主动元数据 构建智能数据治理体系 京东零售技术 │ 3月前 │ ◎ 73 心 点赞 ♀ 评论	算法
LT已死,EtLT才是现代数据处理架构的终点! 自鲸开源 1年前 ◎ 1.7k 心 4	大数据

数据质量和数据治理的关系 | 京东云技术团队很多不太了解的人会认为:数据治理就是干数据清洗的。 近两年,在我们公司,数据 - 掘金 https://juejin.cn/post/7329448423327875111

数据治理怎么做? 这篇万字长文终于讲清楚了! 云祁 │ 3年前 │ ◎ 201	后端
大数据之数据治理体系全面指南 五分钟学大数据 │ 8月前 │ ◎ 1.2k ⑥ 2 ፡ 评论	大数据
数据治理设计理论 LightGao │ 10月前 │ ◎ 819	数据库
治理要从数据建模抓起 京东云技术团队 京东云开发者 │ 1年前 │ ◎ 3.2k ⑥ 2 ᠍ 评论	数据库数据挖掘