

# 大数据时代的数据资产管理方法论与实践

程永新

(上海新炬网络信息技术股份有限公司 广东 广州 510623)

**摘要** 为了帮助企业处理海量数据并从中快速获取真正有用的数据,数据资产管理应运而生。通过介绍现阶段传统企业数据资产管理的六大问题,并结合在运营商数据管理经验阐述数据资产管理五星模型,以及在五星模型落地过程中使用到的七个实践工具。目前数据资产管理五星模型方法论及其工具已在运营商、航空、金融等多个行业得到了大量应用,并取得良好的效果。

**关键词** 大数据 五星模型 数据资产管理

中图分类号 TP3 文献标识码 A DOI:10.3969/j.issn.1000-386x.2018.11.055

## METHODOLOGY AND PRACTICE OF DATA ASSET MANAGEMENT IN THE ERA OF BIG DATA

Cheng Yongxin

(Shanghai New Centurion Network Information Co., Ltd., Guangzhou 510623, Guangdong, China)

**Abstract** In order to help enterprises to manage massive data and quickly obtain useful data from it, data asset management comes into being. Combining with the data management experience in operators, we expounded the five-star model of data asset management and seven practical tools used in the landing process of five star models by introducing the six major problems of data assets management in traditional enterprises at present stage. Now the five-star model methodology and tools in data asset management have been widely applied in many industries such as operator, aviation, finance and so on, and have achieved good results.

**Keywords** Big data Five star model Data asset management

## 0 引言

国际数据公司 IDC 的研究结果表明,2008 年全球产生的数据量为 0.49 ZB,2011 年的数据量是 1.82 ZB,预计 2020 年产生的数据量将是 44 ZB。随着信息技术特别是物联网的广泛应用,网络购物、移动支付、共享经济、智能家居等新业态新模式的蓬勃发展,全球数据呈现爆发增长、海量聚集的特点。每年都产生比以往更大量、维度更丰富的海量数据,采取更好的数据管理方式,更好地利用数据,构建以数据为关键要素的数字经济,核心就是数据资产管理。

## 1 企业数据资产管理现状

根据中国计算机协会大数据专家委员会评估,

50% 以上的企业大数据应用尚处于极为初级的阶段。造成这种现状归根到底在于传统企业中原有粗放式的信息化建设模式以及企业数据文化的缺乏,导致数据相关的管理体系、职责、角色、流程、工具和制度等各方面都有不同程度缺失,出现大量数据资产管理乱象。具体如下:

1) 数据黑盒化。由于企业缺乏数据文化和数据管理意识,信息化过程中未形成有效的数据规划和设计说明,IT 管理和业务人员无法完全理解这些数据,形成各种各样的数据“黑盒”。

2) 数据多头管理。由于信息系统的建设和管理职能分散在各个部门,缺少专门对数据资产进行管理的组织,使数据资产管理的职责分散、权责不明确。

3) 数据标准不统一。各部门缺乏统一的数据规划、可信的数据来源和数据标准,站在各自的立场生产、使用和管理数据,导致数据不规范、不一致、冗余且

难于共享。

4) 数据缺乏有效治理。与数据相关的元数据、质量等数据治理流程没有统一的流程管理,无法保障数据治理活动在整个企业范围内保持一致,企业无法提供一个有效、完整且准确的高质量数据环境。

5) 数据生命周期运营不完整。数据的产生、使用、维护、备份到过时被销毁的数据生命周期管理规范 and 流程还不完善,数据安全问题也十分突出,生命周期运营不完整导致的数据敏感数据泄漏的安全事件屡见不鲜。

6) 数据流通渠道不通畅。数据在不同的应用之间缺乏有效的平台或者工具进行快速方便的数据交互,缺少渠道引入更多的数据以补全企业数据之不足,数据形成事实孤岛。

## 2 企业数据资产管理方法论

企业进行数据资产管理的时候,往往面临如下的三个问题:1) 应该做什么;2) 怎么落地;3) 怎样持续落地。企业需要一个完整的数据资产管理体系,以此为基础制定企业科学的数据资产管理规划,指导自身数据资产管理活动实施。数据资产管理体系,不仅是定义清晰的数据资产管理工作内容和范围,以及数据资产管理组织架构和资源,还需要建立一整套流程和制度,并在相应的平台上去落地实施。

经过多年在大型传统企业数据资产管理方面的实践和摸索,本文总结了一套适用于传统企业的数据资产管理方法论:数据资产管理五星模型。该模型由数据架构、数据治理、数据运营、数据共享和数据变现五个部分组成。它们由数据资产管理团队实施并推动数据资产管理落地。模型如图 1 所示。



图 1 数据资产管理五星模型

### 2.1 数据架构管理

数据资产管理团队首先从数据架构入手打破数据“黑盒”,建立规范化数据架构管理制度和流程,实现数据从“黑盒”向“白盒”转化。打破数据“黑盒”,从现

有数据架构梳理开始,通常分为三个层面:(1) 物理模型层面。物理模型梳理的实质在于从关系数据库中进行数据模型梳理、合并和清理,形成一份稳定的物理模型设计。(2) 逻辑模型层面。在物理模型梳理的基础上,透过数据实体关系分析、实体业务属性梳理、字段业务属性和取值分析等手段,形成完整和清晰的系统逻辑模型。(3) 业务逻辑层面。业务与数据模型关系梳理,从企业业务层面全面梳理业务流程与数据之间的数据流向与相互关联关系,进而实现可视化展示,便于企业人员进一步理解数据。

数据架构还需要进一步的变更管理,提升数据架构设计质量以及确保数据架构与实际应用的高度一致。数据架构变更管理包括:(1) 数据架构基线管理。对各应用系统的数据架构实现基线化管理,企业数据使用者根据基线检索当前企业中最新的数据架构,回溯任意时间点的数据架构状态以及跟踪不同数据实体变更的原因和内容。(2) 数据架构变更评审。对由于业务需求变化而引发的各种 IT 应用中涉及到数据模型和架构的变化进行必要性、合理性和质量等方面的综合评审。(3) 数据架构稽核。定期或不定期对数据架构和实际应用进行差异性分析,找出两者的差距,并制定有效的计划进行整改,保持数据架构设计的权威性和准确性。

### 2.2 数据治理

数据治理从组织架构、管理制度、操作规范、应用技术、绩效考核支持等多个维度对组织的元数据、数据标准、数据质量等领域进行全面的管理和持续改进的过程。

元数据管理是企业数据相关信息的管理过程,它是企业数据资产管理中重要的一环。企业数据资产团队建立从元数据的需求收集、获取、加工、分析到元数据维护闭环的元数据管理流程,逐步收集并管理各个业务域数据定义与加工规则,进而追踪系统实现过程,形成元数据管理体系,推进数据管理的落地。数据资产团队组织制定元数据管理规范、元数据采集计划、元数据应用方案,并对元数据规范工作执行结果进行评估和考核。业务部门和 IT 部门负责根据元数据管理要求提供元数据,并保障元数据更新的及时性、准确性、完整性。

数据标准是企业根据自身特点、行业规范以及国家、国际通用标准建立的一套符合自身实际,涵盖定义、设计、应用多层次数据的标准化体系。为了实现数据标准管控目标,数据资产管理团队针对数据库建模中所涉及的数据表字段、字段属性等信息的规范化

管理,实现企业数据标准化设计和应用。

数据质量是通过企业数据采集和处理过程中,对企业数据完整性、一致性、准确性和处理时效等不同质量维度进行有效的管理,及时发现、分析以及解决数据处理过程中在的质量问题并持续改进,主要包括:(1)建立企业数据质量评估体系,可以量化评估当前企业数据质量状况和质量改进效果;(2)建立数据质量管理体系,通过数据质量管理规则来集中化地发现问题并持续改进;(3)将数据质量管理与业务稽核结合,通过业务规则的稽核来发现数据质量深层次的问题,方便业务人员准确清晰知晓数据质量问题。

### 2.3 数据运营

数据运营管理是通过对数据的开发、维护和支持活动使企业数据资产达到最佳的利用价值。通过数据运营,企业数据使用者可以安全、高效、便捷地使用数据做出各种各样的数据分析和应用。数据运营重点包含数据安全运营、数据生命周期管理和 SQL 代码审核等方面。

数据安全运营是建立适合企业自身的数据安全管理制度、开发规范和相关管理办法等。数据资产管理团队根据国家、行业和企业对数据安全管控需求,对现有数据进行敏感分级分类,形成敏感数据目录。以此为基础,还需要针对每种数据制定相应的安全管理策略,将安全策略整合到敏感数据目录当中。安全管理策略可以包括:敏感数据访问权限控制、敏感数据访问审计策略、敏感数据迁移策略、数据安全控制策略、数据安全备份策略等。数据资产管理团队数据运营工作中,持续对敏感数据目录及定义的策略进行必要的管理以及维护,以保证数据安全运营的持续有效。

数据生命周期管理通过合理、完整的数据生命周期管理解决方案,针对不同类型的业务数据进行贯穿其整个生命周期的管理。数据资产管理团队按需求调研、策略制定、策略发布和技术实施四个步骤实现数据资产生命周期管理。首先,针对业务部门和 IT 部门对数据使用的需求进行调研和分析,同时分析企业中各类数据特点和趋势。其次,结合各方数据需求及数据特点,制定不同数据在不同阶段数据生命周期策略。再次,数据资产管理团队按既定策略跟各业务部门、IT 运维部门确认并达成一致后,形成企业中数据生命周期管理规范并发布。最后,应用数据生命周期管理工具,按既定策略实现数据的自动化加密、转储、归档和销毁的技术实现。

SQL 代码审核采取“优化前移”的方式,即将 SQL 代码的管理和优化从上线后提前到上线前的应用测试

验证阶段,极大提升生产应用的数据使用感知和减少性能故障的发生。完整的 SQL 代码审核管理体系,包括自动采集、问题分析、问题优化和持续跟踪四个方面工作。自动采集从测试环境收集并提交的本版本应用程序中涉及到的新增 SQL,并监控当前版本与历史 SQL 基线有变化的 SQL 及其执行计划。针对新增 SQL 代码审查和问题分析,生成审查报告交付开发 DBA。开发 DBA 根据分析初审报告对问题 SQL 进行深度优化以符合上线的性能要求。最后,持续跟踪上线后的新增 SQL 性能实际状况,开发 DBA 根据分析结果及时解决还潜在于应用环境中的 SQL 性能问题。

### 2.4 数据共享

在企业内部建立有效的数据共享机制,打破企业内部数据孤岛,让数据实现跨部门跨应用的打通,真正让企业数据资产在企业内部流动起来,产生更多的连接和关联,进而产生更大的价值。

企业数据共享体系自下而上应分为南向数据接入管理、数据关联中心管理、数据仓库管理和北向共享接口管理几个层面。南向数据接入管理,负责管理各种企业应用系统数据的接入,包括实时性接入和定时性接入。数据关联中心管理,负责对各种接入数据源进行数据规范化处理和数据关联。数据仓库管理,负责构建面向共享的共享数据模型,同时根据共享模型进行统一数据处理和存储。北向共享接口管理,负责数据对外开放共享接口管理。外部应用系统通过北向共享接口从数据共享中心获取数据。

通过建立统一的数据共享中心来满足针对跨部门、跨系统的数据应用。通过企业内集中的数据共享平台来规范共享数据源,由原来多源网状数据整合为单源集中式数据共享,提升数据共享和活性,提升数据共享服务的应用效率。

### 2.5 数据变现

有了数据架构管控、持续的数据治理运营和长效的数据共享,解决好数据资产管理的基础后,企业可以对内强化能力实现资产增值,对外可以完成数据场景变现。打通了各系统各部门的数据通道,将以前各个部门、系统盲人摸象似的局部数据分析方法变成全局的大数据分析,在企业内部重构产品、重构用户定义、重新审视主营业务的用户是否最佳,进一步实现精准营销和风险控制。企业在面对客户维度不够、产品信息不全等问题时,进行数据跨界合作,补充客户维度信息及产品流通信息,实现数据增值,最终驱动主营业务成长。

数据变现是实现数据外部价值的一系列活动。企

业通过对数据进行建模分析挖掘,把符合共享开放层级的数据作为商品,以符合国家规范安全的形式进行共享交换和开放发布,使得数据具有流通性,方便其他数据消费者使用。

在国家实施大数据战略的今天,企业之间应建立常态化数据交换和数据交易能力。仅仅通过数据租售或者是数据分析结果的变现(包括智库、咨询研究报告、征信评价等)不足以产生巨大价值。通过跨界战略合作,交换彼此应用场景需要的数据才能推动企业核心业务价值裂变。

### 3 数据资产管理实践工具

在为众多企业实践数据资产管理的过程中,本文研发了相关数据资产管理工具,使数据资产管理过程更为体系化、规范化、流程化和自动化。

(1) 数据架构管理工具 从全局入手,对企业不同数据模型之间设计、管理、实施、分析和稽核全程进行有效管理,解决企业数据模型管理分散、无统一数据架构视图、数据模型标准无法有效落地的困境,强化数据架构管控能力。

(2) 数据标准管理工具 针对数据标准管理职能构建,包含标准生成、标准映射、标准维护等数据标准管理能力,协助企业数据标准落地。

(3) 元数据管理工具 提供可视化数据资产相关信息的检索和应用分析能力,通过元数据采集、元数据应用、元数据展示,向数据使用者全面展示企业数据地图、数据血缘关系、影响性分析、数据链路分析等。

(4) 数据生命周期管理工具 对数据产生、应用、归档到销毁的整个生命周期全过程进行有效的自动化管理。针对不同生命周期的数据建立半自动或自动化的多层次数据存储体系。

(5) 数据质量管理工具 将数据质量检查规则规范要求,在工具中配置转换为数据质量检查规则。工具通过数据质量规则的自动化检测和稽核,确保数据在事前、事中、事后的数据质量管控,切实有效地提高企业数据质量。

(6) 数据共享平台 将企业多年陆续分散建设的异构应用信息系统进行整合,保证分布异构业务系统之间实现数据的互联互通,完成异构数据的抽取、装载、清洗和接口管理,构造统一的数据交换和共享平台

(7) 数据安全工具 保证数据资产使用、交换和共享过程中的安全。数据安全工具包括统一认证、数据脱敏、日志审计等模块。

## 4 结 语

数据被业界公认为是企业的最宝贵资产之一。大数据之父维克托则乐观预测,数据列入企业资产负债表只是时间问题。数据资产管理正是为企业提供了完整数据管理体系和路线图。通过数据资产管理,数据能更好地为企业和社会服务。例如:提供更优质的产品和用户体验、控制企业运营风险,以及为企业提供数据化决策支持。本文基于大数据时代背景下的数据资产管理问题,详细阐述了数据资产管理五星模型的方法论,为众多传统企业数据资产管理相关实践提供了参考。

### 参 考 文 献

- [1] 倪江波. 大数据的价值[J]. 中国建设信息化, 2014(13):1.
- [2] 中共中央关于制定国民经济和社会发展第十三个五年规划的建议[R]. 2015.
- [3] 李芬, 朱志祥, 刘盛辉. 大数据发展现状及面临的问题[J]. 西安邮电大学学报, 2013, 18(5):100-103.
- [4] 王珊, 王会举, 覃雄派, 等. 架构大数据:挑战、现状与展望[J]. 计算机学报, 2011, 34(10):1741-1752.
- [5] 胡敏. 大数据变现四种途径, 如何把海量数据变成现金[OL]. [2015-06-12] <http://www.tmtpost.com/1024274.html>.
- [6] 唐怀坤. 大数据变现的九种商业模式[OL]. [2015-06-11] [http://www.cac.gov.cn/2015-06/11/c\\_1115589180.htm](http://www.cac.gov.cn/2015-06/11/c_1115589180.htm).
- [7] 刘士峰. 银行管理与IT架构[M]. 上海:华东理工大学出版社, 2014.
- [8] 维克托·迈尔-舍恩伯格. 大数据时代[M]. 浙江:浙江人民出版社, 2013.
- [9] Data is giving rise to a new economy[OL]. [2017-05-06] <http://zhuankanxuanu.blog.163.com/blog/static/216292004201710673225105>.
- [10] 国务院. 促进大数据发展行动纲要[OL]. [2015-09-05] [http://www.gov.cn/zhengce/content/2015-09/05/content\\_10137.html](http://www.gov.cn/zhengce/content/2015-09/05/content_10137.html).

### (上接第 282 页)

- [11] 黄国顺. 保正域的决策粗糙集属性约简[J]. 计算机工程与应用, 2016, 52(2):165-169.
- [12] 史博文, 李国和, 吴卫江, 等. 基于强化正域的属性约简方法[J]. 计算机应用研究, 2017, 34(1):107-109.
- [13] 刘偲, 秦亮曦. 测试代价敏感的决策粗糙集正域约简[J]. 计算机科学与探索, 2017, 11(6):1014-1020.
- [14] Blake C. UCI repository of machine learning databases[DB/OL]. 1998. <ftp://ftp.ics.uci.edu/pub/machine-learning-databases>.