大数据科学与应用丛书

**云端数据治理**

**刘小茵李尧程广明** **等编著**

**中工信版架团**

電子工紫出版社.

**大数据科学与应用丛书**

**云端数据治理**

刘小茵 李 尧 程广明 高智伟

编 著 谢灵群 钟世敏 张寒坤 朱楠楠

**電** **子** **工** **業** **出** **版** **社**

**Publishing House of Electronics Industry**

北京·BEIJING

**内** **容** **简** **介**

本书的编写团队结合最新的科研成果以及多年来在成熟度评估方面的实战经验，在借鉴国际 先进数据治理理论和方法的基础上，针对云端数据特点，开发了云端数据治理模型，构成了本书 的主要内容。本书旨在帮助读者和有数据治理需求的组织了解云计算环境下的数据治理方法，为 云端数据治理体系建立和云端数据治理实施提供指导。

未经许可，不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有，侵权必究。

**图书在版编目(CIP)** **数据**

云端数据治理/刘小茵等编著.一北京：电子工业出版社，2017.6

(大数据科学与应用丛书)

ISBN 978-7-121-30374-[6](#bookmark1)

I.① 云… Ⅱ.①刘…Ⅲ.①数据处理 IV.①TP274

中国版本图书馆 CIP 数据核字(2016)第276348号

策划编辑：牛平月

责任编辑：王敬栋

文字编辑：牛平月

印 刷： 北京季蜂印刷有限公司

装 订：北京季蜂印刷有限公司

出版发行：电子工业出版社

北京市海淀区万寿路173信箱 邮编100036

开 本： 720×1000 1/16 印张：21.75 字数：520千字

版 次： 2017年6月第1版

印 次： 2017年6月第1次印刷

定 价： 68.00元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行 部联系，联系及邮购电话： (010)88254888,88258888。

质量投诉请发邮件至zlts@phei.com.cn, 盗版侵权举报请发邮件至 dbqq@phei.com.cn。 本书咨询联系方式： (010)88254454, niupy@phei.com.cn。



信息技术与经济社会的交汇融合引发了数据的爆发式增长，数据蕴含着重要的 价值，已成为国家基础性战略资源。数据正日益对全球生产、流通、分配、消费活 动以及经济运行机制、社会生活方式和国家治理能力产生重要影响。

数据治理对于确保数据的准确、适度分享和保护是至关重要的，也是发挥数据 在各种决策中支持作用的必由途径。近年来，数据治理理论得到广泛关注，在一些 行业也得到了广泛应用。然而随着云计算、大数据等技术的发展与应用，数据本身 的特点也发生了诸多变化，如数据来源的多样性、数据的远程存储与传输、数据的 合规要求更具复杂性、数据面临的安全威胁更多等，这些变化让传统数据治理难以 全面深入地应对。因此针对云计算环境下数据的特点和治理需求，研究开发一套针 对云端数据治理的理论体系和实践指导是很有必要的。

编写团队结合最新的科研成果以及所在机构多年来在成熟度评估方面的实战经 验，在借鉴国际上先进数据治理理论和方法的基础之上，针对云端数据特点，开发 了云端数据治理模型，构成了本书的主要内容，旨在帮助读者和有数据治理需求的 组织了解云计算环境下的数据治理方法，为建立云端数据治理体系和云端数据治理 实施提供指导。

本书共11章，分为3篇：第1篇云端数据治理成熟度概述；第2篇云端数据治 理成熟度模型解读；第3篇云端数据治理体系与实施。围绕“概念—内容—实施” 循序渐进的总体思路，本书有望帮助读者了解数据治理的概念，建立对云端数据治 理的初步认识；帮助读者深入学习云端数据治理模型及其内涵，掌握开展云端数据 治理活动所需的理论知识；指导读者构建云端数据治理体系，开展具体的云端数据 治理活动。

**第** **1** **篇** 云端数据治理成熟度概述(第1～3章),通过对比云计算环境和传统 IT环境下数据的特点，结合数据治理发展趋势，有针对性地提出了云端数据治理成 熟度模型，帮助读者建立云端数据治理的初步认识。

**第1章** 数据治理概述，介绍了数据治理的相关概念，讨论了数据治理的定义， 并简要回顾了数据治理理论的研究历程以及数据治理面临的挑战。

**第** **2** **章** 大数据与云端数据治理，介绍了云计算和大数据的特点以及他们对数 据本身带来的影响，针对云计算环境下数据的特点，提出了云端数据治理的定义并 解释了其内涵，对比分析了云端数据治理与数据治理的关系，论述了开展云端数据 治理的作用和价值。

**■云端数据治理**

**第** **3** **章** 云端数据治理模型与治理成熟度评估模型，从治理内容、治理原则、 治理实施3方面解读了云端数据治理模型，帮助读者建立对云端数据治理的总体认 识；定义了云端数据治理能力成熟度等级并描述了相应等级的特征，将云端数据治 理能力成熟度分为5个等级，用于衡量组织的云端数据治理水平；介绍了云端数据 治理成熟度评估模型与评估方法。

**第** **2** **篇** 云端数据治理成熟度模型解读(第4～9章),从云端数据战略、云端 数据管理、云端数据质量、云端数据操作、云端数据架构、安全与隐私6个职能域 分别介绍云端数据治理模型的具体内容，帮助读者掌握云端数据实施的基本知识和 开展成熟度评估的依据。

**第** **4** **章** 云端数据治理战略，从战略、组织与角色、业务案例、资源保障和沟 通5个过程域提出了开展云端数据治理战略管理的要求，帮助组织制订一个能够全 局指导云端数据治理活动开展的顶层规划。

**第** **5** **章** 云端数据管理，从业务词汇表、元数据管理和主数据管理3个过程域 提出了对云端数据进行管理的要求，为云端数据管理应该达到的水平指明了方向。

**第** **6** **章** 云端数据质量，从数据质量战略、数据概要分析、数据质量评估和数 据清洗4个过程域提出了对云端数据质量进行管理的要求，能够指导组织如何提高 数据质量。

**第** **7** **章** 云端数据操作，从数据提供者管理、数据集成与互操作和数据生命周 期管理3个过程域提出了对云端数据操作进行管理的要求，帮助组织建立一个完整 的数据生命周期管理体系，提高数据全流程的管理水平。

**第** **8** **章** 云端数据架构，从架构方法、架构标准、数据管理平台和历史数据管 理4个过程域对云端数据架构建设提出了要求。

**第** **9** **章** 安全与隐私，从风险管理、数据安全、隐私保护和合规管理4个过程 域对数据安全与隐私保护提出了要求，帮助组织提高数据安全保护能力，降低数据 安全风险，指导组织如何在符合法律法规等要求的前提下开展数据治理。

**第** **3** **篇** 云端数据治理体系与实施(第10～11章),围绕云端数据治理体系构 建和云端数据治理实施论述，指导组织如何具体开展云端数据治理活动。

**第10章** 云端数据治理体系，介绍了体系的必要性和体系的基本框架，并从宏 观上介绍了云端数据治理体系的实施过程。

**第11** **章** 云端数据治理体系实施，系统全面地介绍了云端数据治理的实施过 程，能为读者提供一个清晰的云端数据治理实施指南。

本书各章执笔分别是：第1章由刘小茵、程广明、朱楠楠编写，第2、3章由刘 小茵、程广明编写，第4章由程广明编写，第5章由李尧编写，第6章由高智伟、 李尧编写，第7章由张寒坤、李尧编写，第8、9章由钟世敏、程广明编写，第10、 11章由刘小茵、谢灵群、程广明编写。刘小茵、程广明负责全书的组织、策划、汇 总和校审工作，其他执笔人分别负责了相关章节的审阅工作。

**■前** **言■**

本书得到了国家科技支撑计划项目、广东省科技计划项目和广州市科技计划项 目产学研协同创新重大专项资金等的支持，在此表示感谢。本书在撰写过程中，得 到了来自学术界和产业界众多专家的帮助，感谢他们给出了很多非常有价值的意见 和建议。

本书的宗旨是为读者提供最新的云端数据治理的参考，但由于云端数据治理研 究刚刚起步，本书提出的云端数据治理模型难免会有一些偏颇和不当之处。由于编 写时间仓促，加上编写水平有限，书中难免会有错误和不足之处，请读者不吝赐教， 提出意见和建议，以便我们不断进步。

编著者



**第1篇** **云端数据治理成熟度概述**

**第** **1** **章** **数据治理概述** ([2)](#bookmark2)

1.1 数据治理相关概念 ([2)](#bookmark3)

1.1.1 数据、信息、知识 ([2)](#bookmark4)

1.1.2 数据治理的定义 ([4)](#bookmark5)

1.2 数据治理理论研究历程 ([6)](#bookmark1)

1.2.1 数据治理理论的研究进展 ([6)](#bookmark6)

1.2.2 大数据治理理论的研究进展 ( [16)](#bookmark7)

1.3 数据引发的典型事件及原因归类 ( [19)](#bookmark8)

1.3.1 数据引发的典型事件 ( [19)](#bookmark9)

1.3.2 数据事件原因归类 ([21)](#bookmark10)

1.4 数据治理意义和面临的挑战 ([22)](#bookmark11)

1.4.1 数据治理对组织的意义 ([22)](#bookmark12)

1.4.2 数据治理面临的挑战 ([24)](#bookmark13)

**第** **2** **章** **大数据与云端数据治理** ([25)](#bookmark14)

2.1 云计算与大数据环境下的数据特点 ([25)](#bookmark15)

2.1.1 云计算的特点 ([25)](#bookmark16)

2.1.2 大数据的特点 ([28)](#bookmark17)

2.1.3 云计算环境下数据的特点 ([32)](#bookmark18)

2.2 云端数据治理介绍 ([36)](#bookmark19)

2.2.1 云端数据治理的定义 ([36)](#bookmark20)

2.2.2 云端数据治理与数据治理的关系 ([39)](#bookmark21)

2.3 云端数据治理的作用和价值 ([41)](#bookmark22)

**第** **3** **章** **云端数据治理模型与治理成熟度评估模型** ([43)](#bookmark23)

3.1 云端数据治理模型概述 ([43)](#bookmark24)

3.1.1 云端数据治理模型 ([43)](#bookmark25)

**■云端数据治理■**

3.1.2 成熟度等级定义与特征 ([47)](#bookmark26)

3.2 云端数据治理成熟度评估模型 ([49)](#bookmark27)

3.2.1 成熟度等级判别标准 ([49)](#bookmark28)

3.2.2 成熟度评估方法 ([50)](#bookmark29)

**第2** **篇** **云端数据治理成熟度模型解读**

**第** **4** **章** **云端数据战略** ([54)](#bookmark30)

4.1 概述 ([54)](#bookmark31)

4.1.1 目的和意义 ([54)](#bookmark32)

4.1.2 内容与关联性 ([55)](#bookmark33)

4.2 活动及要求 ([55)](#bookmark34)

4.2.1 战略 ([55)](#bookmark35)

4.2.2 组织与角色 ([64)](#bookmark36)

4.2.3 业务案例 ([73)](#bookmark37)

4.2.4 资源保障 ([81)](#bookmark38)

4.2.5 沟通 ([87)](#bookmark39)

**第** **5** **章** **云端数据管理** ([94)](#bookmark40)

5.1 概述 ([94)](#bookmark41)

5.1.1 目的和意义 ([94)](#bookmark42)

5.1.2 内容与关联性 ([95)](#bookmark43)

5.2 活动及要求 ([95)](#bookmark44)

5.2.1 业务词汇表 ([95)](#bookmark45)

5.2.2 元数据管理 ( [105)](#bookmark46)

5.2.3 主数据管理 ( [119)](#bookmark47)

**第** **6** **章** **云端数据质量** ( [132)](#bookmark48)

6.1 概述 ( [132)](#bookmark49)

6.1.1 目的和意义 ( [132)](#bookmark50)

6.1.2 内容与关联性 ( [133)](#bookmark51)

6.2 活动域及要求 ( [134)](#bookmark52)

6.2.1 数据质量战略 ( [134)](#bookmark53)

6.2.2 数据概要分析 ( [145)](#bookmark54)

6.2.3 数据质量评估 ( [155)](#bookmark55)

6.2.4 数据清洗 ( [166)](#bookmark56)

**■目录■**

**第** **7** **章** **云端数据操作** ( [176)](#bookmark57)

7.1 概述 ( [176)](#bookmark58)

7.1.1 目的和意义 ( [176)](#bookmark59)

7.1.2 内容与关联性 ( [177)](#bookmark60)

7.2 活动及要求 ( [177)](#bookmark61)

7.2.1 数据提供者管理 ( [177)](#bookmark62)

7.2.2 数据集成与互操作 ( [188)](#bookmark63)

7.2.3 数据生命周期管理 ( [198)](#bookmark64)

**第** **8** **章** **云端数据架构** ([210)](#bookmark65)

8.1 概述 ([210)](#bookmark66)

8.1.1 目的和意义 ([210)](#bookmark67)

8.1.2 内容与关联性 ([211)](#bookmark68)

8.2 活动及要求 ([211)](#bookmark69)

8.2.1 架构方法 ([211)](#bookmark70)

8.2.2 架构标准 ([222)](#bookmark71)

8.2.3 数据管理平台 ([234)](#bookmark72)

8.2.4 历史数据管理 ([243)](#bookmark73)

**第** **9** **章** **安全与隐私** ([251)](#bookmark74)

9.1 概述 ([251)](#bookmark75)

9.1.1 目的和意义 ([251)](#bookmark76)

9.1.2 内容与关联性 ([251)](#bookmark77)

9.2 活动及要求 ([253)](#bookmark78)

9.2.1 风险管理 ([253)](#bookmark79)

9.2.2 数据安全 ([263)](#bookmark80)

9.2.3 隐私保护 ([273)](#bookmark81)

9.2.4 合规管理 ([280)](#bookmark82)

**第3篇** **云端数据治理体系与实施**

**第10章** **云端数据治理体系** ([288)](#bookmark83)

10.1 云端数据治理体系的必要性 ([288)](#bookmark84)

10.2 云端数据治理体系基本框架 ([289)](#bookmark85)

10.3 云端数据治理体系过程模型 ([290)](#bookmark86)

**■云端数据治理■**

**第11章** **云端数据治理体系实施** ([292)](#bookmark87)

11.1 统筹和规划 ([292)](#bookmark88)

11.1.1 制订战略与目标 ([292)](#bookmark89)

11.1.2 确定体系范围 ([295)](#bookmark90)

11.1.3 建立组织与角色 ([296)](#bookmark91)

11.1.4 制订体系建设和运行计划 ([299)](#bookmark92)

11.1.5 形成体系文件 ([300)](#bookmark93)

11.2 构建和运行 ([303)](#bookmark94)

11.2.1 体系宣贯 ([303)](#bookmark95)

11.2.2 实施策划 ([303)](#bookmark96)

11.2.3 项目启动 ([304)](#bookmark97)

11.2.4 项目实施 ([307)](#bookmark98)

11.2.5 项目过程管理 ([323)](#bookmark99)

11.3 监督和评估 ([326)](#bookmark100)

11.3.1 绩效评估 ([326)](#bookmark101)

11.3.2 内部审计 ([327)](#bookmark102)

11.3.3 安全审计 ([329)](#bookmark103)

11.4 改进和优化 ([330)](#bookmark104)

11.4.1 制订改进计划 ([330)](#bookmark105)

11.4.2 实施改进措施 ([333)](#bookmark106)

参考文献 ([334)](#bookmark107)





**第1篇**

**云端数据治理成熟度概述**

**第** **1** **章** **数据治理概述**

从 IT (信息技术)时代到 DT (数据技术)时代，都离不开一个关键词——数 据。数据已成为21世纪最重要的战略资源之一。数据成为可以变现交易的资产，但 又不同于传统的财务资产，数据的可拷贝、可重用以及数据的收集、存储、使用都 有其特殊性，数据还涉及个人隐私、运行的安全。数据治理是以数据为研究对象的 一门学科，它涉及对组织内的人员、流程、技术和策略的分配，让组织能够从数据 中获取更优的价值。

本章首先介绍数据治理相关的概念，以及这些概念之间的关系；其次，介绍数 据治理理论的研究历程；再次，分析了因数据引发的典型事件以及数据面临的主要 问题；最后，阐述数据治理的意义和面临的挑战。

**1.1** **数据治理相关概念**

**1.1.1** **数据、信息、知识**

数据 (Data) 是客观事实经过获取、存储和表达后得到的结果，用于表示客观 事物未经加工的原始素材。通常以符号、文本、数字、图形、图像、声音和视频等 表现形式存在[1]。

信息 (Information) 不是具体的事物，也不是某种物质，而是客观事物的一种 属性。它是包含上下文语境的数据 (Data with Context), 必须依附于某个客观事物 (媒介)而存在。同一个信息可以借助不同的信息媒介表现出来，如：文字、手势、 表情、图形、图像、声音、影视和动画等。

知识 (Acknowledge) 是对情境的理解、意识、认识和识别，以及对复杂性的把 握。知识只有在经过广泛深入的实践检验，被人类消化吸收，并成为个人的信念和 判断取向之后才能成为知识[2]。

表1-1为数据、信息、知识之间的关系实例。

**第1章** **数据治理概述**

**表1-1** **数据、信息、知识之间的关系实例**

|  |  |  |
| --- | --- | --- |
| 姓名：李吴天 性别：男  年龄：11岁 年级：六年级  数学考试成绩：40分  自述：在2016年9月末考试中李吴天数学成绩40分，满分为100分。 | | |
| 数据 | 40 | 单拿数据40来说，在没有上下文语境的情况下，它没有任何意义。 |
| 信息 | 数学考试成绩： 40分 | 这时40就有其代表的意义，代表六年级小朋友在一次月末考试中，数学成绩为 40分。在这样的背景下，40成为信息中的一个关键指标。 |
| 知识 | 孩子是不是厌学 了，家长该如何解 决问题 | 总分100分，考试成绩40分为不及格，可认为孩子这次没有考好。家长可以从 以下3个方面判断：(1)孩子最近的表现是不是厌学，情绪低落，如果是近期厌  学或心情不好引起的成绩下滑，要及时调整；(2)是不是考试当天身体不舒服以 至于影响考试成绩，如果是这种情况就没必要太担心，毕竟只是一次考试；(3) 孩子以往学习成绩都是在40分左右，如果是这种情况，需要家长和学校都给予高 度重视，及时帮孩子补习，找准学习方法，培养学习兴趣。  根据多方面观察和分析，找出孩子数学成绩差的主要原因，并及时解决问题。 |

通过以上分析可以看出，数据、信息与知识之间既有联系，又有区别。数据是 记录下来可以被鉴别的符号。它是最原始的素材，未被加工解释，没有回答特定的 问题，没有任何意义；信息是已经被处理、具有逻辑关系的数据，是对数据的解释， 这种信息对其接收者具有意义。知识是在对信息进行了筛选、综合、分析等过程之 后产生的，它不是信息的简单累加，往往还需要加入基于以往的经验所作的判断。 因此，知识可以解决较为复杂的问题，可以回答“如何”的问题，能够积极地指导 任务的执行和管理，进行决策和解决问题。

数据是信息的表现形式和载体，数据和信息是不可分离的，数据是信息的表达， 信息是数据的内涵。如果没有数据和信息，知识就难以发挥作用。数据和信息的获 取则相对简单，而只有知识能够帮助解决问题。

数据信息知识智慧模型 (DIKW) 将数据、信息、知识纳入到一种金字塔形的 层次体系，每一层比下一层都赋予了一些新的特质。原始观察及量度获得了数据， 分析数据间的关系获得了信息。在行动上应用信息产生了知识3]。图1-1所示的是数 据、信息和知识之间的关系。

**■云端数据治理**

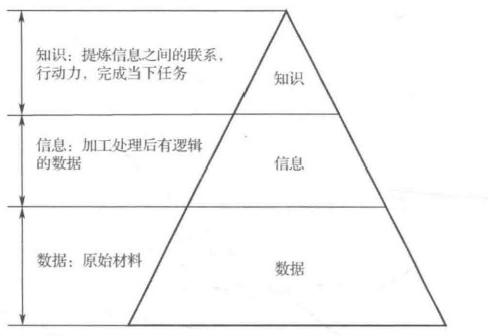


图1-1 数据、信息和知识之间的关系

**1.1.2** **数据治理的定义**

随着以移动互联网、物联网、社交网络为代表的新型信息发布方式的不断涌现， 数据规模不断增长，非结构化的数据所占比重将越来越高，有价值的数据密度会越来 越低。应用信息系统建设发展到一定阶段，数据资源将成为战略资产，而有效的数据 治理才是数据资产形成的必要条件。由于侧重点和切入视角的不同，目前业界对数据 治理的定义尚未形成统一的标准。下面主要介绍6个典型数据治理的定义。

**1.ITSS** **给出的数据治理的定义**

ITSS (中国电子工业标准化技术协会信息技术服务分会)指出，数据是资产，通 过服务产生价值，数据治理是在数据产生价值的过程中治理团队对其的评价、指导、 控制，是数据治理的最基本概念。ITSS 认为数据治理应包含以下3个方面内容4。

(1)确保信息利益相关者的需要评估，以达成一致的企业目标，这些企业目标 需要通过对信息资源的获取和管理实现；

(2)确保有效助力业务的决策机制和方向；

(3)确保绩效和合规进行监督。

ITSS 认为数据治理是一种体系，是一个关注于信息系统执行层面的体系。这一 体系的目的是整合IT 部门与业务部门的知识和意见，通过一个类似于监督委员会或 项目小组的虚拟组织对企业的信息化建设进行全方位的监管，这一组织的基础是企 业高层的授权和业务部门与IT 部门的建设性合作。从范围来讲，数据治理涵盖了从 前端事务处理系统、后端业务数据库到终端的数据分析，从源头到终端再回到源头 形成一个闭环负反馈系统(控制理论中趋稳的系统)。从目的来讲，数据治理就是要 对数据的获取、处理、使用进行监管(监管就是我们在执行层面对信息系统的负反

**第1章** **数据治理概述**

馈),而监管的职能主要通过五个方面的执行力来保证——发现、监督、控制、沟通、 整合。

**2.DAMA** **给出的数据治理的定义**

DAMA(国际数据管理协会)在其出版的 DMBOK(The DAMA Guide to the Data Management Body of Knowledge)[S1 中认为数据管理 (Data Management,DM) 是 规划、控制和提供数据及信息资产的一组业务职能，包括开发、执行和监督有关数 据的计划、政策、项目、流程、方法和程序，从而控制、保护、交付和提高数据资 产的价值。该定义突出了数据管理的职能、过程和规范三个关键词。在职能上认为 数据管理是业务数据管理专员和技术数据管理专员共同承担的责任，业务数据管理 专员是企业数据资产的托管人，技术数据管理专员是这些资产的专业管理人和监护 人。对数据管理职能治理可以协调IT 和企业之间的协同合作。在过程上，数据管理 是数据资产管理的权威性和控制性活动，是在数据管理和使用层面上进行规划、监 管和控制。在规范上，数据管理必须遵守相关的规则和规范，才能确保数据管理过 程能够顺利进行。

**3.ISACA** **给出的数据治理的定义**

ISACA (国际信息系统审计协会)在其出版物 COBIT(Control Objectives for Information and related Technology)101 中认为数据治理是一个由关系和过程所构成的 体制，用于指导和控制企业，通过平衡信息技术与过程的风险、增加价值来确保实 现企业的目标。数据治理通过评估利益相关者的需求、条件和选择以达成平衡一致 的企业目标，通过优先排序和决策机制来设定方向，再根据方向和目标来监测绩效 和合规性问题。

COBIT 认为数据治理的关键在于授权和控制并举，应该做出什么决策、谁来做 决策、如何做出决策和监督，促成IT 创造有利于战略的价值。例如，企业要不要上 ERP 一事是要按流程体系办事，经过评估分析其战略价值，按照企业实际状况来决 定的，而不是由领导一人决定的。

**4.DGl** **给出的数据治理的定义**

DGI(数据治理研究所)1认为数据治理是针对数据信息相关过程的决策权和职 责体系，这些过程遵循“在什么时间和什么情况下、用什么方式、由谁、对哪些数 据、采取哪些行动”的方法来执行。

**5.IBM** **给出的数据治理的定义**

IBM 认为数据治理是一门将数据视为一项企业资产的学科。它涉及到以企业资 产的形式对数据进行优化、保护和利用的决策权利。它涉及到对组织内的人员、流

**■云端数据治理**

程、技术和策略的编排，以从企业数据获取最优的价值8。从一开始，数据治理就 在协调不同的、孤立的且常常冲突的策略(可能导致数据异常)的过程中扮演着重 要角色。类似于客户关系管理 (CRM) 诞生之初，组织开始任命全职或兼职数据治 理负责人。与任何新兴学科一样，数据治理有许多定义，但市场已经开始围绕将数 据视为资产的定义进行具体化行动。

**6.Oracle** **给出的数据治理的定义**

Oracle 认为数据治理是制订决策权和问责的框架，以规范企业在估值、创建、 存储、使用、归档及删除数据和信息的行为。它包括流程、角色、标准和指标，以 确保组织能有效和高效地利用数据和信息以实现其目标[9]。

**1.2** **数据治理理论研究历程**

**1.2.1** **数据治理理论的研究进展**

在数据治理理论研究领域，以ISACA 、DAMA 、DGI 、IBM等组织为代表，提 出了自成体系的数据治理框架或与数据治理相关的治理理论模型。

**1.ISACA 提出的COBIT 5**

COBIT 是 ISACA 制订的面向过程的信息系统审计和评价的标准，以支持企业 实现其企业IT 治理和管理的目标。COBIT5 是以五项关键原则为基础进行企业IT治 理和管理，如图1-2所示。这五项原则结合在一起能够使企业构建一种能优化信息 和技术投资的、用于利益相关者收益的、有效的治理和管理框架。其中有一条原则， COBIT5 将治理和管理明确区分开来。这两种科目包含不同类型的活动，需要不同 的组织结构，并服务于不同的用途。从 COBIT5 的角度来看，治理和管理之间的关 键区别在于：治理确保利益相关者的需要、条件和选项得到评估，以决定平衡、协 商一致、需要实现的企业目标；通过优先等级和决策来设定导向；并监控商定的导 向和目标的绩效以及合规性。在大多数企业中，整体治理是董事长领导下的董事会的 责任。具体的治理责任可能授予适当级别的特别组织结构，尤其是在较大型综合性企 业中更是如此。管理层计划、构建、运行和监控与治理机构设立导向一致的活动，以 实现企业目标。大多数企业中，管理是首席执行官领导下的行政管理层的责任。

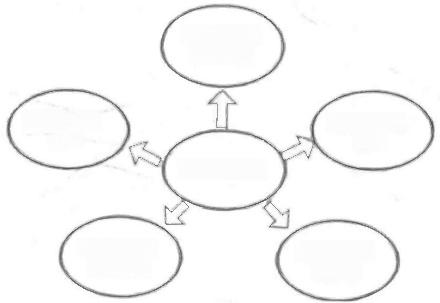
COBIT5 以持续改进的生命周期为基础提供了实用和广泛的实施指南。COBIT5 实施生命周期被划分为七个阶段，形成持续改进的生命周期循环，如图1-3所示。

**第** **1** **章** **数据治理概述**

2.我们现在在什么位置?

6.我们达到目标了吗?

实 现 收



满足利益相

关者需求

区分治理和

管理

COBITS原则

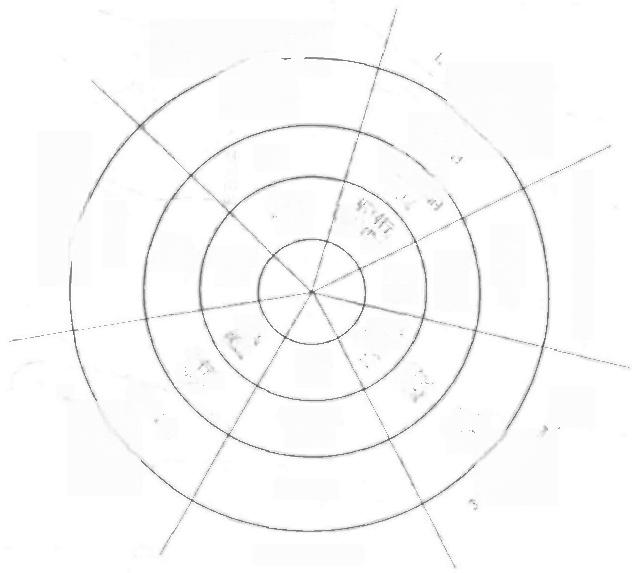
采用单一集 成框架

端到端覆盖 企业

启用一种综

合的方法

图1-2 COBIT5 数据治理基本原则

*7.* *我们如何保持这种势头?*

*评估有效性*

嵌入 新方法

保持

控 和

评估

启用新

方法

、







|  |  |
| --- | --- |
| 评估  现状 | 建立 执行 小组 |

定义

问题

和机

会

上府

实施

乡 

在 下

改进

学

√

识别角色

V

扮演者

规划方案 。

4. 需要做什么?

图1-3 COBIT 实施生命周期的七个阶段

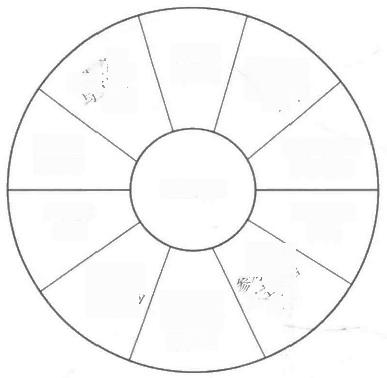
生命周期的三个互为联系的组成部分分别是：(1)生命周期不是一次性项目，

而是持续改进的过程(内环);(2)通过启动变更来解决行为和文化方面的问题(中 环);(3)项目集管理(外环)。

 **■云端数据治理**

**2.DAMA** **提出的数据管理知识体系**

DAMA总结了数据治理的十大功能(见图1-4)和七大环境要素(见图1-5)[113]。 数据治理的职能是在数据管理和使用层面之上进行规划、监督和控制。数据治理的 十大管理功能包括数据架构管理、数据建模与设计、数据存储与设计、数据安全管 理、参考数据和主数据管理、数据仓库与商业智能管理、数据集成与互操作管理、 文 档 和 内 容 管 理 、 元 数 据 管 理 及 数 据 质 量 管 理 。 其 对 应 的 功 能 见 表 1 - 2 。



数

*代≠*

数据存储与 设计管理

数据治理

管理



x

数据仓库

与商业智

能管理



数据质 量管理

数据架 构管理

元数据 管理



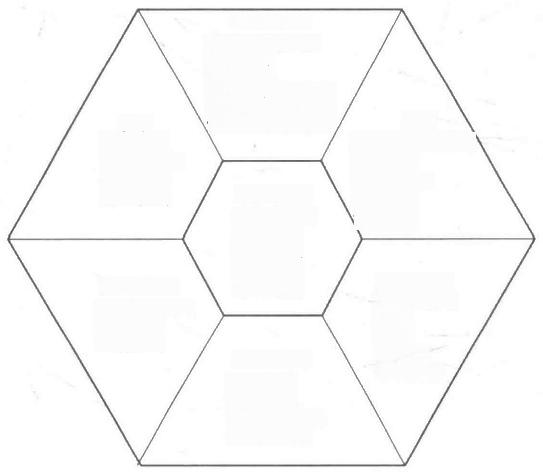
数据安全



或

称

**图1-4** **DAMA** **数据治理十大功能框架**

技术

·工具分类 ·标准和协议 ·选择条件 ·学习曲线

实践与方法

·识别最佳实践 ·通用方式

·可选方式

组织与文化

·成功的关键因素

·汇报结构

·管理矩阵

·价值、信仰、期望

·态度、风格、偏好

、仪式、标志、传 阶段、任务骤

依赖关系

·顺序和流程

目标与原则 ·用例场景

·愿景和使命 \触发事件

·商业利益

|  |  |
| --- | --- |
| ·战略目标  ·具体目标  ·指导原则  角色与职责  ·个人角色  ·组织角色  ·业务和IT角色  ·资格和技能 | 主要交付物 ·输入和输出 ·信息  ·文档  ·数据库 ·其他资源 |

图1-5 七大环境要素

**第1章-数据治理概述**

**表1-2** **DAMA 数据治理十大功能的职能**

|  |  |
| --- | --- |
| 数据治理功能 | 职 能 |
| 数据架构管理 | 定义企业的数据需求，并设计蓝图以便满足这一需求 |
| 数据建模与设计 | 数据的分析、设计、实施、测试、部署、维护等工作 |
| 数据存储与设计管理 | 提供从数据获取到清除的技术支持 |
| 数据安全管理 | 确保隐私、保密性和适当的访问权限等 |
| 数据质量管理 | 定义、监控和提高数据质量 |
| 参考数据和主数据管理 | 管理数据的黄金版本和副本 |
| 数据仓库与商务智能管理 | 实现报告和分析 |
| 文档和内容管理 | 管理数据库以外的数据 |
| 元数据管理 | 元数据的整合、控制以及提供元数据 |
| 数据集成与互操作管理 | 采集、提取、变换、移动、交付、复制、联合、虚拟化和运营支持 |

七大环境要素包括目标与原则、活动、主要交付物、角色与职责、实践与方法、 技术、组织与文化。每一环境要素包含如下内容。

●目标和原则：每个职能的方向性业务目标以及指导每个职能绩效的基本 原则。

●活动：每个职能都是由低一级的活动组成，有些活动被分为子活动。活动能 被进一步分解成任务或步骤。

●主要交付物：信息、物理数据库及各职能在管理过程中最终输出的文档。有

些交付物是必需的，有些是建议性的，还有些则是视情况而定可选的。

●角色和责任：在执行一个职能和对功能进行监督过程中所涉及的业务角色和 IT 角色，以及在该职能中每个角色所涉及的具体责任。很多角色可能会参与 多个职能。

●实践与方法：用于执行这些过程和生成交付结果的常见及流行的方法和具体 技术，也可能包括常见的约定、推荐的最佳实践方法和简要介绍的候选方法。

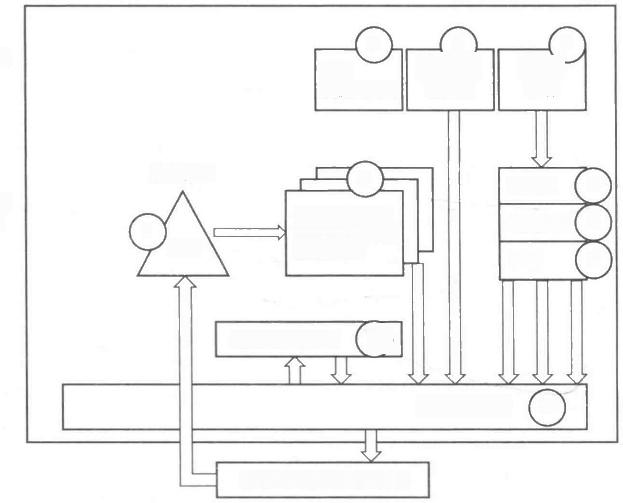
●技术：各种支撑技术(主要是软件工具)、标准和协议、产品选择标准和常 见学习曲线。

●组织与文化：主要包括管理度量标准、成功和商业价值的度量标准等因素。

**3.DGl** **提出的数据治理框架**

DGI 认为数据治理应建立独立的数据治理理论体系。DGI 从组织、规则、过程 三个层面总结了数据治理的十大关键要素，创新性地提出了数据治理框架。组件按 照职能划分为三组：人员与组织结构、规则与协同工作规范、过程?,如图1-6所示。

**■云端数据治理**



7

数据

利益相关人

2

4

关注域(目标、

指标/绩效度量、

财务策略)

6

数据规则与定义 3)

过程

10

数据相关的业务与IT过程

决策权

职责分工

控制

8

数据管理

委员会

人员与组织

结构

规则与协同

工作规范

数据管

理者

数据治理过程

数据治理

使命

9

5

1

**图1-6** **DGl 数据治理框架**

人员与组织结构即制订和执行数据治理规则和规范的组织结构，包含数据利益 相关人、数据治理委员会和数据管理人。

规则与协同工作规范即建立、协调和规范数据治理工作的规则(包含政策、需 求、标准、责任、控制和数据定义等),以及指导不同部门共同制订和执行规则的协 同工作规范。包含使命、关注域(目标、指标/绩效度量、财务策略)、数据规则和 定义、决策权、职责分工、控制六个组件。

过程即数据治理应遵循的工作步骤和流程，它应该是正式的、书面的、可重复 的和可循环的。主要包括主动、被动和正在进行的数据治理过程。

DGI 数据治理路线图指出数据治理实施生命周期应包含七个阶段：建立价值目 标、制订路线图、规划并提供资金、设计方案、部署方案、治理数据及监督、测量 和报告。

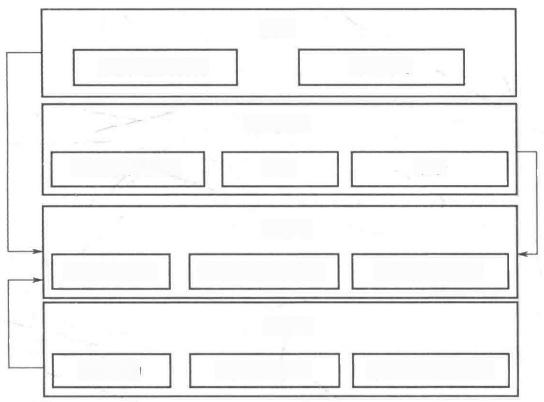
**4.IBM** **提出的数据治理成熟度模型**

IBM 在构建数据治理统 一 框架方面，提出了数据治理的要素模型(见图1- 7),并 认为业务目标或成果是数据治理的最关键命题。整个数据治理统 一流程需要以持续循 环的形式操作。该流程需要度量结果并循环回到高层支持者，以获得数据治理计划的 持续支持。在要素模型中组织结构和认知度、政策和数据相关责任者三个促成因素会 影响业务目标的实现。除此之外，应当重点关注数据治理的核心要素和支撑要素。

40

[U

**第1章** **数据治理概述**



成果

数据风险与合规

促成因素

组织架构与意识 照管 策略

核心域

数据质量管理 数据生命周期管理 信息安全与隐私

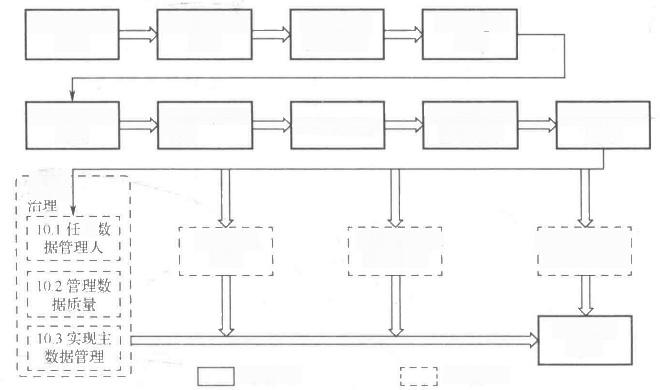
支撑域

数据架构 分类与元数据 审计、日志与报告

价值创造

图1-7 IBM 数据治理要素模型

每个组织应该执行一些步骤来治理自己的数据。IBM 数据治理统一流程列出了 14个主要步骤(10个必需步骤和4个可选专题),以及支持有效的数据治理计划的 IBM 软件工具和最佳实践[8如图1-8所示。



1. 定义业务

问题

5. 建立组织

蓝图

主数据

命娄

14. 度量 结果

必要步骤 可选步骤

113. 治理信 i 息生命周期1

12. 管理安| 全和隐私

9. 定义度量 指标

8. 创建元数 据存储库

6. 创建数据 字典

2. 获取高层 支持

4. 创建路 线图

3. 执行成 熟度评估

11. 治理 分析

7. 理解数据

**图1-8** **IBM 数据治理统一流程**

图1-8中10个必要步骤是建立有效的企业治理计划所不可或缺的步骤。然后从 4 个可选专题(主数据治理、治理分析、管理安全和隐私、治理信息生命周期)中 选择一个或多个步骤。最后，需要定期度量数据治理统一流程，将结果传送给管理 层 支 持 者 。 数 据 治 理 统 一 流 程 及 主 要 内 容 见 表 1 - 3 。

 **■云端数据治理**

**表1-3** **数据治理统一流程及主要内容**

|  |  |  |
| --- | --- | --- |
| 步骤 | 活 动 | 主 要 内 容 |
| 1 | 定义业务问题 | 围绕一个特定的业务问题定义数据治理计划的初始范围 |
| 2 | 获取高层支持 | (1)创建虚拟数据治理工作组；(2)获取IT和业务部门内高级管理人员的支持； (3)识别数据治理的负责人 |
| 3 | 执行成熟度评估 | (1)定义评估的组织范围；(2)定义想要的数据治理未来状态的时间范围；  (3)定义要评估的数据治理类别；(4)确定业务和IT部门中正确的研讨会参与者； (5)执行数据治理成熟度评估研讨会；(6)与高层管理人员沟通评估结果 |
| 4 | 创建路线图 | (1)总结数据治理成熟度评估的结果；(2)列出填补评估中强调的差距所需的关 键人员、流程和技术计划：(3)基于关键计划的优先级创建路线图 |
| 5 | 建立组织蓝图 | (1)定义数据治理章程；(2)定义数据治理的组织结构；(3)建立数据治理委员 会：(4)建立数据治理工作组；(5)确定数据监管人；(6)举行数据治理委员会和 工作组定期会议 |
| 6 | 创建数据字典 | (1)选择一个数据领域；(2)安排数据照管人来维护关键业务词汇；(3)识别关 键数据元素(4)从现有的词汇术语表创建数据字典；(5)填充数据字典；(6)链  接业务词汇与技术工件；(7)支持数据治理审计、报告和日志需求；(8)整合数据 字典与应用程序环境 |
| 7 | 理解数据 | (1)理解范围内的每个数据源(a.执行列和表级别分析；b.通过逆向工程主一  外键关系查询遗留模式；c.识别每个来源中的关键数据元素的位置；d.识别每个 来源中的敏感数据的位置);(2)理解来源之间的关系(a.理解关键数据元素在各  个数据源之间的数据重叠情况；b.发现来源之间的数据连接和复杂转换逻辑；c.发 现数据不一致性和异常) |
| 8 | 创建元数据  存储库 | (1)合并来自数据字典的业务元数据和来自发现流程的技术元数据；(2)确保合 适的数据血统；(3)执行影响分析；(4)管理操作元数据 |
| 9 | 定义度量指标 | (1)理解业务的整体关键绩效指标(KPI);(2)定义数据治理的业务驱动KPI: (3)定义数据治理的技术KPI;(4)建立数据治理成熟度评估的仪表板 |
| 10 | 主数据治理 | (1)任命数据照管人(a.任命首席数据照管人；b.确定数据照管计划的配置； c.确定每个数据领域的高层支持者；d.招聘每个数据领域的数据照管人；e.授 权数据治理委员会监督数据照管计划。);  (2)管理数据质量(a.建立数据质量策略，包括高价值数据属性的识别；b.设 置数据质量基准；c.创建业务案例；d.清理数据；e.持续监控数据质量。);  (3)实现主数据管理(a。识别业务问题；b.定义主数据主题区域；c。识别使  用数据的系统和业务流程；d.识别当前的数据源；e.定义记录系统的数据属性；  f.为每个记录系统任命数据照管人；g.建立主数据治理策略；h.为手动干预和  监控实现数据照管控制台；i.管理潜在的重叠任务；j.匹配来自相同来源或多个  来源的可疑重复内容，创建一个新主记录；k.链接来自多个来源的相关记录；1.检  查唯一标识符是否重复；m.管理关系；n.管理层次结构；o.管理分组；p.构 建主数据管理解决方案。) |

**第1章** **数据治理概述■**

续表

|  |  |  |
| --- | --- | --- |
| 步骤 | 活 动 | 主 要 内 容 |
| 11 | 治理分析 | (1)定义BICC的目标；(2)准备BICC的业务案例；(3)确定BICC的组织结 构；(4)协商BICC的关键功能 |
| 12 | 管理安全和隐私 | (1)与关键利益相关者协调一致；(2)收集企业安全架构蓝图；(3)加强数据库  变更控制；(4)自动化合规性工作流程：(5)定义敏感数据；(6)发现敏感数据： (7)分类和标记敏感数据；(8)加密敏感数据；(9)保护非生产环境中的敏感数据；  (10)监控应用程序中的欺诈；(11)预防计算机攻击；(12)编校非结构化文档中 的敏感信息 |
| 13 | 治理信息生命 周期 | (1)建立信息架构；(2)发现业务对象；(3)分类数据和定义服务水平；(4)归 档数据和非结构化内容；(5)建立管理测试数据的策略；(6)定义电子文档法律查  询策略；(7)分析内容 |
| 14 | 度量结果 | 依据预先定义的KPI集度量结果，确保计划继续催生业务价值。这些结果需要定 期传达给数据治理委员会和高层管理人员。进度度量指标将确保对数据治理计划的 持续支持和资金支助 |

**5.** **数据治理成熟度模型** **(DMM)**

由 CMMI 研究所推出的数据管理成熟度 (DMM)TM 模型9,该模型的总体目 标是帮助组织提高其关键数据资产管理的熟练程度，并提供适合持续改进、合规性 和审计的基准。该模型包含了确立、建立、维持和优化整个数据生命周期的有效数 据管理的最佳实践，该周期包括从数据的产生到交付、维护和归档。

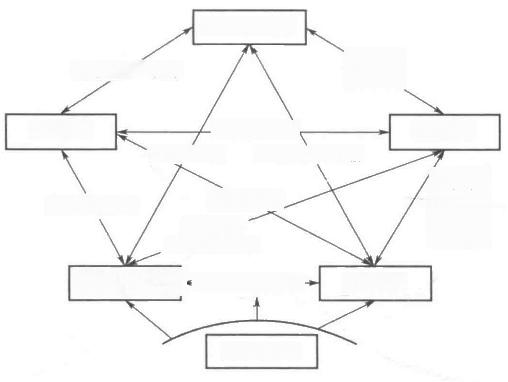
DMM模型由20个数据管理过程域和5个以CMMI过程域为基础的支持过程域 构成。该模型分为五个大类，如图1-9所示，每个类别都包含了若干过程域。图1-9 显示了这五个类别(数据管理战略、数据治理、数据质量、数据操作和数据平台及 架构)是如何共同作用，形成一个有效的数据管理方案的。DMM 也包括支持过程 域，这些支持过程域包含了为执行过程域提供支持的最佳实践方案。

数据管理战略的目的是定义数据管理方案的愿景、目标和目的，确保所有的利 益相关者就优先事项与计划的执行和管理达成一致。数据管理战略包含沟通、数据 管理功能、业务案例和方案资金过程域。

数据治理的目的是对所有权、管理工作和操作结构进行开发，从而确保公司数 据作为关键资产受到管理并以有效持续的方式实施。数据治理包含数据管理、业务 术语表和元数据管理3个过程域。

数据质量的最佳实践是关于如何定义以及实施一个协作的方法，从而检测、评 估以及清理数据缺陷，确保数据能够实现其在业务运营、决策制订以及策划上的预 期用途。数据质量包含数据质量战略、数据概要分析、数据质量评审和数据清洗4 个过程域。

**■云端数据治理**



数据管理战略

质量需求 质量战略

数据治理 元数据监督一

业务π联盟 利益相关者联盟

基础设施监督 方向&合规

质量需求

质量解决计划

数据平台及架构一一数据基础设施一 数据操作

支持过程域

业务流程 数据需求 质量规则 质量准则

实施监管：协作

数据质量

**图1-9** **DMM** **结构图**

数据操作的目的是明确数据需求，管理数据生命周期与供应链，包含数据要求 定义、数据生命周期管理和供方管理三个过程域。

数据平台及架构的目的是建立一种方法和标准，来确保已实施的数据管理平台 能成功地集成、归档和保存公司的数据资产以支持业务目标的最佳实践方案。其包 含架构方法、架构标准、数据管理平台、数据集成和历史数据、归档和保留5个过 程域。

**6.PWC** **数据治理能力框架**

PWC 数据治理能力框架是由普华永道公司根据数据治理的时间经验制订的，共 包含数据治理战略、数据治理组织架构、数据治理流程、数据治理策略、数据分析 与监测、技术与架构六大方面的内容[10],如图1-10所示。

**7.Gartner 数据治理模型**

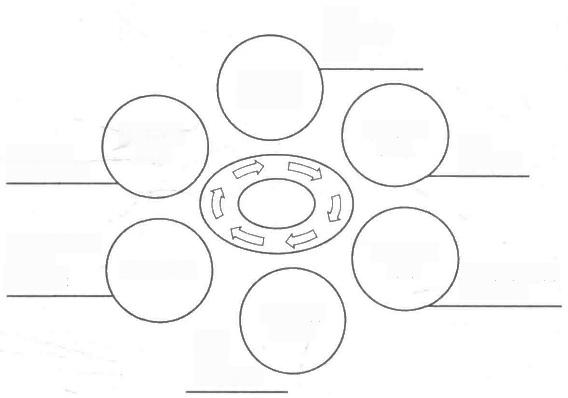
Gartner 提出了一个数据管理参考架构[,描述了数据管理的所有构成要素及要 素之间的层级关系，数据治理和信息管理仅是该架构的组成部分，如图1-11所示。 同时，Gartner 还建立了一个数据治理与信息管理的要素模型来描述支撑数据治理方 案的基本要素，该模型包含四个部分：规范、计划、建设和运营，它适用于实施数 据治理计划的任何组织，如图1-12所示。

Gartner 认为数据治理是 IT 治理的一部分，同时也属于公司治理范畴。Gartner 数据治理的四个实施阶段的主要内容如下。

(1)规范主要内容包括策略管理、过程建模与设计、学习管理；

(2)计划主要内容包括需求分析、影响分析和存储库与元数据管理；

**第1章** **数据治理概述**

职责

技能

数据治理

组织架构

愿景/范围

方法/线路图

数据治理

战略

数据治理

流程

交流

变革管理

技术架构设计 元数据

数据安全

技术&架构

数据分析

数据治理 策略

数据标准/评估

控制/规划

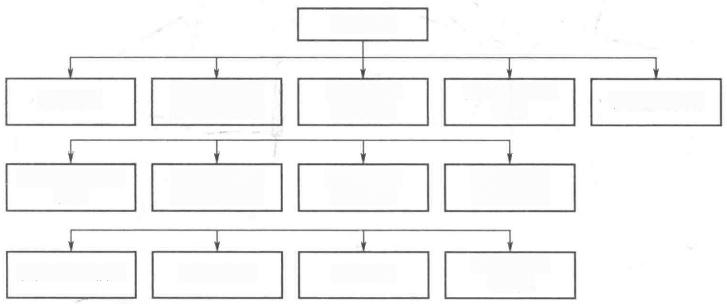
发现 分析

跟踪/报告

图1-10 PCW

&监测

数据治理能力框架



数据管理

主数据管理

基础设施

数据安全与

活动监控

数据库管理 与监控

数据治理与信息 管理

信息生命周期 管理基础设施

商务智能基础 设施

业务分析、报 告与数据访问

数据集成与 主数据管理

存储元数据模型

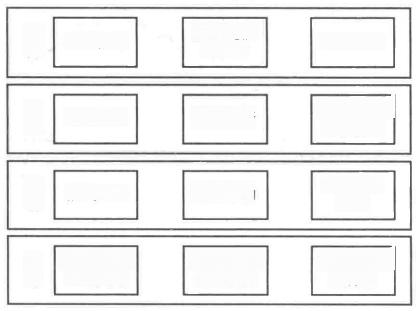
移动基础设施

数据库架构

信息资产

数据建模

图1- 11 Gartner 数据管理参考架构



过程建模

与设计

影响分析

数据架构

数据访问审 计与报告

存储库与元 数据管理

数据质量

规范

信息生命周 期管理

规 范 计 划 建 设 运 营

数据质量监 控与实施

学习管理

需求分析

模式设计

策略管理

图1- 12 Gartner 数据治理和信息管理要素

**■云端数据治理**

(3)建设主要内容包括模式设计、数据架构和数据质量规范；

(4)运营主要内容包括数据质量监控与实施、数据访问审计与报告和信息生命 周期管理。

以上列出了目前在业内具有较高知名度的数据治理能力框架，它们分别从不同 的角度定义了数据治理能力维度，拥有各自的优点和不足，表1-4 简要对它们进行 了对比分析。

**表1-4** **数据治理能力框架对比**

|  |  |  |
| --- | --- | --- |
| 数据治理模型 | 优 点 | 不 足 |
| COBIT 5 | 作为面向过程的信息系统审计和评价的 标准，给出了信息治理的较为完善的模型 | 虽然能为数据治理提供帮助，但其并不是针 对数据治理的指导准则 |
| DAMA | 数据治理理论较为完善，也给出了数据 治理实施的指导活动 | 数据治理理论偏通用规则，部分内容在云计 算环境下对数据治理不适用 |
| DGI | 给出了6个数据治理关注领域、10个数 据治理要素，为数据治理实践提供了支持 | 更多地关注数据治理的实施过程，数据治理 理论相对较少 |
| IBM | 以4个维度11个度量项作为数据治理能 力成熟度的衡量指标，并给出了数据治理 流程的14个主要步骤，能够成为数据治理  的主要参考之一 | 该模型主要偏重于数据治理的具体实施，理 论指导相对偏少，部分内容在云计算环境下对 数据治理不适用 |
| DMM | 将能力成熟度评估模型引入了数据治理 领域，具有很大创新性，比较详细地列举 了每个等级需要达到的基本条件，便于组 织评估 | 该模型重在给出评估依据，但没有关于数据 治理如何实施的详细指导，部分内容在云计算 环境下对数据治理不适用 |
| PWC | 包含了数据治理战略、组织架构、流程、 策略、数据分析与监测等内容，具有清晰 的管控逻辑 | 过于注重组织、策略和流程的重要性，数据 治理内容深度不足 |
| Gartner | 从13个方面支持了数据管理，并给出了 数据治理实施的规范、计划、建设和运营 四个阶段 | 强调数据治理过程的重要性，而对数据治理 策略策划重视不足 |

**1.2.2** **大数据治理理论的研究进展**

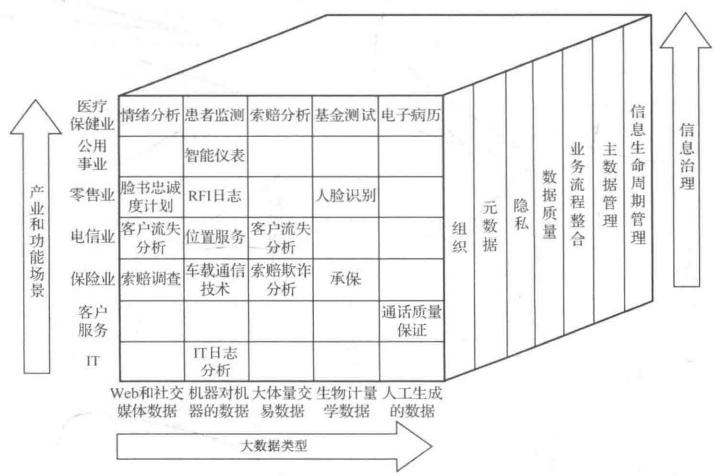
随着信息技术与人类生活方式的交互，数据产生的速度变得更快，数据体量变 得更大，数据内容更加复杂，传统数据治理难以应对大数据治理需求，因此大数据 治理一词应景提出。但由于这是一个崭新的领域，国内外研究成果不多。目前，关 于“大数据治理”定义最早由Sunil Soares(桑尼尔·索雷斯)在其出版的专著《Big Data Govemance:An Emerging Imperative》中提出：大数据是广义信息治理计划的 部分，即需要制订与大数据有关的数据优化、隐私保护与数据变现的政策[12]。该定

**第1章■数据治理概述**

义将大数据作为广义信息治理计划的一部分，定目标是大数据变现，而过程则包括 大数据优化和隐私保护。除此之外，我国也在大数据治理方面取得了一些研究成果。

**1.** **索雷斯大数据治理框架**

索雷斯大数据治理框架由大数据类型、信息治理准则、产业和功能场景三个维 度组成。大数据类型分为五种： Web 和社交媒体数据、机器对机器的数据、大体量 交易数据、生物计量学数据和人工生成的数据。信息治理准则包括七则：组织、元 数据、隐私、数据质量、业务流程整合、主数据整合和信息生命周期管理。产业和 功能场景分为七类：医疗保健业、公用事业、零售业、电信业、保险业、客户服务 和 IT。索雷斯大数据治理模型如图1-13所示。



**图1-13** **索雷斯大数据治理模型**

索雷斯认为信息治理的七个核心准则同样适用于大数据治理，因此在大数据治 理模型中通过组织、元数据、隐私、数据质量、业务流程整合、主数据管理、信息 生命周期七个方面重点描述大数据治理要重点关注的内容。表1-5 对大数据治理模 型的七个治理准则进行了简要解释。

**表1-5** **索雷斯大数据治理模型治理准则**

|  |  |
| --- | --- |
| 治 理 准 则 | 优 点 |
| 组织 | 信息治理组织需要将大数据纳入其总体框架中，包括章程、组织架构、角色和责任 |
| 元数据 | 大数据治理计划需要将大户数与企业的元数据库进行整合 |

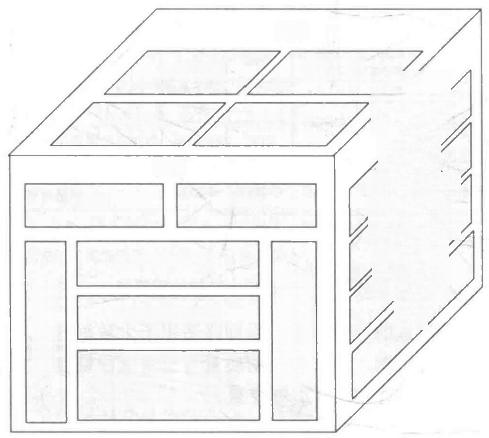
 **■云端数据治理**

**续表**

|  |  |
| --- | --- |
| 治 理 准 则 | 优 点 |
| 隐私 | 大数据治理组织需要识别敏感数据，并制订有关敏感数据的可接受使用的政策 |
| 数据质量 | 包括测量、提升和论证质量及整合组织数据的方法 |
| 业务流程整合 | 大数据治理计划必须识别需要大数据的核心业务流程，因而必须识别支持大户数治 理的关键政策 |
| 主数据管理 | 大数据治理需要制订有关将大数据整合到主数据管理环境的政策 |
| 信息生命周期管理 | 决定何种数据应该保留在运营和分析系统中、何种数据要予以存档、何种数据应予 以删除 |

**2.** **中国数据治理标准研究团队提出的大数据治理框架**

随着我国不断推进大数据战略，数据治理理论和实践得到了快速发展，尤其是 数据治理标准研究团队提出了新的数据治理模型，并向 ISO IT 治 理 和 IT 服务管理 分技术委员会提交了《数据治理白皮书》国际标准研究报告[13]。数据治理标准研究 团队也将大数据治理的最新研究成果集中体现在了《大数据治理与服务》一书中， 该书提出了 一 个从全局视角描述的大数据治理框架，主要从原则、范围、实施与评 估三个维度展现大数据治理的内容14]。



原则

战略一致

运营合规

范围

战略

大数据服务创新

大数 据安

大数据生命周期

大数据架构

大数据存储 元数据

主数据 数据仓库



大数据质量

绩效提升

风险可控



组织

**图1-14** **中国数据治理标准研究团队提出的大数据治理框架**

该模型给出了大数据治理的四个指导性原则，包括：战略 一 致、风险可控、运 营合规和绩效提升。战略一致是指大数据战略应与组织的整体战略保持一致，风险 可控是指在大数据治理过程中注意数据的安全和隐私问题，运营合规是指大数据治

**第1章** **数据治理概述**

理活动应符合国内外法律法规和行业规范，最后绩效提升体现数据治理的主要目的 是提供高质量的大数据服务。

大数据治理范围共包括七个关键域：战略、组织、大数据质量、大数据安全与 隐私及合规、大数据服务创新、大数据生命周期、大数据架构。

大数据治理的实施与评估描述了大数据治理实施和评估过程中需要关注的内 容，包括促成因素、实施生命周期、成熟度评估和审计四个方面。

**1.3** **数据引发的典型事件及原因归类**

**1.3.1** **数据引发的典型事件**

**1.** **“5·28携程网瘫痪”事件**

2015年5月28日11时许，携程网官网出现大面积瘫痪，网页版和手机 APP均 不能正常使用5I。据了解，点击进入携程网，页面显示“404报错”,虽然点击“返 回首页”后依然可以进入携程网，但其功能和其他链接均无法使用。事件发生后， 有传闻称，携程官方网站及APP 遭数据库物理删除，数据全部丢失。虽然最后携程 解释此次事件是由于员工错误操作，删除了生产服务器上的执行代码导致，并表示 经过排查，携程数据没有丢失，预订数据也保存完整。但此事件还是给业界带来了 很大的震撼，让人们再一次认识到数据容灾的重要性。

5·28携程网瘫痪事件是继网易“5·11 服务器瘫痪”和支付宝“5·27光纤被 挖断”之后，2015年5月短短数天时间内连续发生的与数据相关的信息安全事件， 充分暴露出了一些企业业务连续性保障能力不足，数据容灾能力不够0。

**2.** **“徐玉玉”事件**

2016年8月21日，徐玉玉因被电信诈骗电话骗走上大学的9900 元费用，伤心 欲绝，郁结于心，最终导致心脏骤停，不幸离世7]。该事件起因在于8月19日下午 4点30分左右，徐玉玉接到了一通陌生电话，对方声称有一笔2600 元的助学金要 发放给她。由于当天是贷款事项的最后一天，且因为之前曾接到过教育部门发放助 学金的通知，徐玉玉信以为真，就按对方的要求赶到附近一家银行，通过自动取款 机领款。她通过自动取款机操作后并未成功。对方得知她带着交学费的银行卡后， 要她取出卡上的9900元，把钱汇入指定账号，对方再把她的9900元连同助学金2600 元一起打过来。毫无戒备的徐玉玉按照对方的说法操作，结果所有学费被骗走。骗 子能精准运用场景诈骗，其根本原因在于事先掌握了徐玉玉的一些数据。

数据泄露越来越受到人们的重视，然而，数据泄露的现象却有增无减。金雅拓

**■云端数据治理■**

公布的《数据泄露水平指数调查报告》显示，与2015年下半年相比，2016 年上半 年的数据泄露总数增长了15%。在全球范围内，2016年上半年已曝光的数据泄露事 件高达974起，数据泄露记录总数超过了5.54亿条，而2015年下半年数据泄露事 件和数据泄露记录总数分别为844起和4.24亿条。此外，2016年上半年曝光的数据 泄露事件中，有52%的数据泄露事件都未公布数据泄露记录数量[18]。

**3.** **数据标准不统一，增加公积金监管难度**

《住房公积金管理条例》是我国住房公积金管理法规的顶层文件，具有引领性， 但不进行具体操作指引。各个地方在对公积金进行管理时，业务标准不一样，业务 流程不一样，管理的力度不一样，尤其是分散的管理，对国家的监管也带来很大的 难度。据报道，目前全国设置的公积金管理机构有600 多家，全国各地住房公积金 的信息化系统林林总总，有些是基于银行系统开发的，有些是中心委托软件公司开 发的，数据结构及数据定义缺乏统一的标准和规范。这种发展状况严重制约了公积 金的信息化管理，也带来了一系列的管理隐患，出现很多数据口径的偏差，直接导 致统计口径偏差和统计数据失真。2014年，中华人民共和国住房和城乡建设部发布 《JGJ/T320-2014住房公积金基础数据标准》[19,并要求当年10月1日起实施。据标 准规定，住房公积金基础数据由公共、归集和提取、个人住房贷款、建设项目贷款、 财务、结算六部分组成，包含29张数据表、394个数据项、241个代码。该标准适用 于住房公积金信息系统建设中基础数据的创建、检查、存储、传输、共享和集成。

**4.** **谷歌流感预测连续三年“不靠谱”**

2008年，谷歌推出了一款名为“谷歌流感趋势”(Google Flu Trends) 的产品， 通过寻找与流感相关的词语，比如“咳嗽”和“发烧”等，根据汇总的搜索数据， 近乎实时地对全球当前的流感疫情进行估测。2008年， Google 第一次开始预测流感 取得了很好的效果，比美国疾病预防控制中心提前两周预测到了流感的爆发。但是， 几年之后，该系统一直高估与流感相关的就医量，在这类数据最有用的流感季节高 峰期尤其预测不准确。自从2011年8月以来，Google 流感趋势在108周的时间里有 100周出了错。在2011—2012流感季节， Google 的预测比实际情况(由美国疾病预 防控制中心根据全美就诊数据推算得出)高出了50%。在2012—2013 流感季节， Google 预测的就医量是美国疾病预防控制中心最终记录结果的两倍。2014年预测结 果与美国疾病预防控制中心最后的记录结果也存在较大差异，于是2015年谷歌流感 趋势不再对公众开放120]。

**5.** **美国驻华使馆监测并发布我国** **PM2.5 数据**

2012年，美国驻华使馆监测并发布了北京 PM2.5 数据。2012年6月5日环保 部在召开的新闻发布会上表示，外国驻华使领馆开展对我国PM2.5 监测并且发布数

**第1章** **数据治理概述**

据，既不符合国际通行的要求、法规及国际公约，也不符合中国的要求。环保部副 部长吴晓青说，中国空气质量监测及发布，涉及社会公共利益，属政府的公共权力， 个别国家驻华领事馆自行开展空气质量监测并由互联网发布空气质量信息，既不符 合维也纳外交关系公约、维也纳领事关系公约的精神，也违反了中国环境保护的有 关法规[21]。

**1.3.2** **数据事件原因归类**

**1.** **数据安全**

信息安全保护的主要目标之一是保护系统和应用程序的基础数据。在云计算环 境下，传统的数据安全方法将遭到云模式架构的挑战。弹性、多租户、新的物理和 逻辑架构，以及抽象的控制需要新的数据安全策略。由于云服务的远程特性，云端 数据往往会传输到外部甚至公众的环境中，这对用户掌控数据提出了很大挑战，但 这种部署方式相对于传统的单点本地存储数据方式却有助于数据的容灾。

数据容灾是保障数据安全的一个重要手段，是指通过建立一个异地的数据系统， 保护数据安全和提高数据的持续可用性。服务供应商应保证数据中心的运行连续性， 保障服务连续性(尤其是在出现一些严重问题时，如火灾、长时间停电以及网络故 障等)。对于信息技术服务提供商，数据往往是服务的主要支撑之一。服务提供商需 要进行业务连续性管理，制订相应的业务连续性规划，并且能够得以落实和实施， 使得当出现灾难时，可以快速恢复业务，继续为用户提供服务。

**2.** **数据泄露**

云服务具有无位置特异性的特点，这有利于云服务的广泛使用。然而，无论是 用户还是云供应商都不可能直接确认在特定云中计算资源的详细位置，这就产生了 不安全的数据复用、数据跨境流动、用户数据安全和隐私保护等问题，从而对数据 存放位置、数据删除或持久性、数据备份与恢复和不同客户数据的混合等方面提出 了新的安全要求。大规模数据泄露的原因有很多，其中比较重要的原因是数据保管 者对数据资产的管理不足，例如数据传输过程中被窃听或主动丢失、存储数据的访 问控制措施不足以及对历史数据管理不重视等。

**3.** **数据标准**

数据标准化主要体现在对数据信息的分类和编码。对数据信息的分类是指根据 一定的分类指标形成相应的若干层次目录，构成一个有层次的逐级展开的分类体系。 数据的编码设计是在分类体系基础上进行的，数据编码要坚持系统性、唯一性、可 行性、简单性、 一致性、稳定性、可操作性和标准化的原则，统一安排编码结构和

**■云端数据治理。**

码位。数据标准是数据共享和系统集成的重要前提，数据标准化可以节省费用、提 高效率和方便应用，有利于系统推广应用，实现数据共享，减少数据采集费用。

**4.** **数据质量**

数据是组织最具价值的资产之一。数据质量是保证数据应用的基础，其决定决 策的风险。组织的数据质量与业务绩效之间存在着直接联系，高质量的数据可以使 组织保持竞争力并在经济动荡时期立于不败之地。不合理的使用以及使用不正确的 数据(即差的数据质量)可导致决策的失败，正可谓差之毫厘、谬以千里。数据质 量可以从数据本身的数据质量和数据的过程质量两方面理解。数据本身的质量包括 数据的真实性、完备性和自洽性，是数据本身应具有的属性，称为数据的绝对质量， 是保证数据质量的基础。数据的过程质量是指在利用和存贮数据的过程中所产生的 数据质量，包括使用质量、存储质量和传输质量。数据质量管理是对数据全生命周 期过程中可能引发的各类数据质量问题进行识别、度量、监控、处置等一系列管理 活动。

**5.** **合规性管理**

合规对于数据治理来说具有重要意义。在云计算环境下，物理数据存储在哪里， 运行在哪里以及从哪里获得数据都是可变的，数据的流动也是跨区域甚至跨国界的。 当组织将其业务从传统数据中心迁移至云计算数据中心时，将面临新的安全挑战， 其中最重要的挑战即是需要遵从各种合规性要求。例如：根据财政部、证监会、审 计署、银监会、保监会联合制订并发布的《企业内部控制基本规范》[22]的要求，“企 业应当在信息系统中设置操作日志功能，确保操作的可审计性。对异常的或者违背 内部控制要求的交易和数据，应当设计由系统自动报告并设置跟踪处理机制。”另外， 针对数据安全和个人信息保护，许多国家已经发布相关的法律法规要求，部分行业 也发布了针对性的数据安全管理要求，组织在处理涉及数据事务时，首先必须识别 出所有适用的法律法规要求，然后根据要求和组织实际，采取合理的措施以满足合 规性管理。

**1.4** **数据治理意义和面临的挑战**

**1.4.1** **数据治理对组织的意义**

在信息化高度发展的大环境下，数据对组织的价值日益增大，数据已经成为与 资金和人力资本同等重要的资产。数据价值的利用不是简单就能得到的，需要组织 去管理去挖掘，而如果管理不好，不仅可能无法得到数据价值，反而遭受数据之灾，

**第1章■数据治理概述**

例如数据泄露、决策误判等。数据治理定位于组织对数据资产的管控与利用，发挥 数据及其背后蕴藏的价值。

**1.** **数据治理提高组织数据管理能力**

数据价值发挥的前提是管理好数据，数据规模的增加、数据格式的复杂化等都 给组织管理数据提出了挑战。数据治理就是在整个数据生命周期过程中制订由业务 推动的数据政策、数据所有权、数据监控、数据标准以及指导方针，指导组织开展 数据管理活动。例如数据治理要求组织建立专业的数据治理组织与角色，并在方针 政策上明确数据治理的目标和方向，这让数据管理有操作的主体，目标也更加清晰； 数据操作帮助组织弄明白数据管理需要什么技术、如何去操作、怎么去衡量结果等， 真正实现对数据的有效管理。

**2.** **数据治理降低组织数据管理风险**

组织为业务开展通常会在一些关键领域收集、分析和使用各种有关客户、产品、 业务环境变化等方面的信息，但是意外的数据泄露无疑会对组织的业务产生较大的 负面影响。由于缺乏正确的数据治理策略以及正确使用数据的规则，组织的敏感数 据可能丢失。有效的数据治理可以为数据的访问管理、评估和管控风险、明确数据 利益相关人、建立决策权分配制度、明确岗位职责并提供全面支持，这必然会有助 于组织保护敏感数据。

除了数据泄露的风险，组织还要面对合规监管和安全控制风险，这关系到数据 隐私、存取管理和许可、信息安全控制，以及规范、合同或内部要求的遵守和执行。 一方面，数据治理应在业务的法律框架内进行；另一方面，数据治理政策和规则的 制订应与政府和行业相关标准相一致。数据治理可以降低因不遵守法律和规范所带 来的风险。

**3.** **数据治理提高数据质量，提升组织决策水平**

在数据领域最流行的一个说法是“更好的数据意味着更好的决策”。单纯的数据 量的积累不会对组织产生任何益处，拥有数据不代表能够用好数据。只有建立适当 的分析模型，并运用相应的技术手段，对大量的数据进行有效的深加工，提高数据 质量，才能发现隐含在大量数据中的信息并加以利用，进而指导组织做出相关决策。 数据质量管理是组织变革管理中一项关键的支撑流程，例如组织业务重点的变化、 组织业务整合、资产重组与并购等都对IT 支撑提出了更高的要求，包括整合数据源、 创建一致的数据副本、交互提供数据等。数据质量管理就是去伪存真，发现数据之 间关联的过程。数据治理要求建立规则和标准对数据进行整合、清洗等以满足组织 的业务需要，有效的数据治理可以产生高质量的数据，更高的数据可信度又会产生 与业务目标保持一直、更有洞察力和更为高效的决策。

**■云端数据治理**

**1.4.2** **数据治理面临的挑战**

数据需要治理，组织开展数据治理需要什么样的治理理论、如何去治理、由谁 来治理、治理到何种程度是组织需要考虑的问题。

**1.** **数据治理所处的环境更复杂**

组织的数据管理需要应对组织内部和外部的环境变化。内部环境来自于组织是 否对数据有足够的重视、管理者是否有意识地参与到数据治理活动、组织的人员能 力是否能够满足数据治理的需要、数据治理过程是否满足设定的计划等；外部环境 来自于监管部门的要求、合作方的要求、数据治理理论的发展水平等。数据治理是 一项庞大的工程，不是个别领导、个别部分或者少数人的活动，是需要各部门和各 层级之间的相互支持和协作。然而现状是很多组织缺少数据治理的组织文化，并未 从根本上充分认识数据治理的难度和价值，自然没有完善的组织和制度。随着组织 数据的增加，数据管理难度增大，数据面临的安全威胁更多，在没有完善的组织去 应对的情况下，数据安全威胁终有可能变成数据安全事件。

当前全球网络与信息技术不断演进，以云计算、大数据、移动互联的新技术在 推动产业发展的同时，也带来了更多新的挑战，数据治理所处的外部环境更加复杂。 移动化、互联网化变革使得信息的新增量大、开放性高、流动性强，数据治理的对 象——数据自身特性发生了重大的变化。虽然数据治理理论取得了很多进展，不同 研究组织从不同视角下推出了不同的数据治理理论体系，然而信息技术的发展是迅 速的，尤其是云计算和大数据技术的发展和应用， 一些数据治理理论难以完全适应 新环境下数据治理的需求。因此，这给组织在选择数据治理理论时带来了困难。

**2.** **数据治理持续性没有得到足够重视**

在传统数据治理时期，数据体量小，数据结构简单，最关键的是数据本地化存 储，数据是掌控在组织内部的，数据治理的对象可以看成一个处于静态的数据体。 那么数据治理就是分段化、不连续的活动，关注对某一时段数据的管理。因此，数 据治理强调的是一个项目，而不是一个持续改进的过程。

随着大数据时代的到来，云计算在数据的生命周期中所起的作用至关重要。数据 不再是静态的，而是数据体量不断增加、数据源不断扩充、数据结构更加复杂、数据 面对的威胁更多，数据治理也不再是一个单一项目。对待数据治理要像对组织的信息 治理一样，不能一劳永逸，应该将其看成一个需要不断创新、持续改进的过程。

**第** **2** **章** **大数据与云端数据治理**

**2.1** **云计算与大数据环境下的数据特点**

随着云计算、物联网、移动互联网等新技术的快速发展和应用，尤其是大数据 概念的出现，数据在其自身的体量、种类等方面发生了重大变化，这也迫使数据治 理必须面对更多、更复杂的外部环境，必须不断丰富其内涵。鉴于此，有必要先弄 清楚云计算、大数据的特点以及它们对数据本身和数据治理的影响。

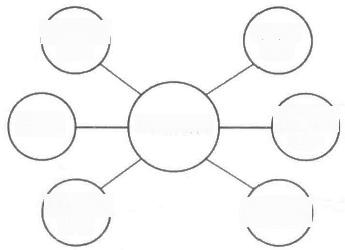
**2.1.1** **云计算的特点**

云计算是基于互联网的IT 资源交付模式，通常涉及通过互联网来提供动态易扩 展且虚拟化的资源。对云计算的定义有多种说法，到底什么是云计算，至少可以找 到上百种解释。现阶段广为接受的是美国国家标准与技术研究院 (NIST) 给出的定 义，即：云计算是一种按使用量付费的模式，这种模式提供可用的、便捷的、按需 的网络访问，进入可配置的计算资源共享池(资源包括网络、服务器、存储、应用 软件、服务),这些资源能够被快速提供，只须投入很少的管理工作，或与服务供应 商进行很少的交互。

云计算作为IT 第三次里程碑式的变革，它的出现使计算机提供计算能力的方式 发生了巨大的变化。云计算通过利用已有技术，尤其是虚拟化技术、分布式技术， 把所有IT 基础设施资源都集中化到一个数据中心，通过云系统建立一台超级大规模 的计算机，然后再通过虚拟化和分布式计算技术，把资源按照用户的需求动态地分 配给每个用户。使得建成的超大规模“云计算机”像发电厂一样，多用户可以从这 座“电厂”中随时获取所需要的计算资源。云计算采用按需计费的方式，就像生活 中用电一样，用户无须考虑如何“发电”,也不用担心“电力”调配、维护、更新问 题。这种集中管理的方式不但节省资源，提高资源利用率，而且可以大大提高运算 能力，带来更好的用户体验。

根据ISO/IEC 17788l23|对云计算的描述，将云计算的特点进行了总结，如图2-1 所示。通过这些特点来体现云计算的核心竞争力124]。

**■云端数据治理**



(泛在网络)

接入

多租户 云计算特点

快速弹性)

伸缩

服务可

度量

按需自动) 服务

(资源池化)

**图2-1** **云计算的特点**

**1.** **泛在网络接入**

让用户可以在任何地方通过任何设备来接入需要的信息服务是人们追求的一个 理想，云计算的发展让我们离这个目标更近了一步。资源的集中供应必然要求用户 通过网络来接入，云服务的供给就相当于电力和自来水集中供应之后通过电网和自 来水管道网络来提供服务一样，用户无须购买服务器，无须安装和维护操作系统、 数据库等复杂环境，轻松完成集群应用的部署操作。无须复杂的架构设计，即可应 对千万级的日访问量。用户可以使用不同的设备，包括 PC、手机、平板电脑等通过 网络接入云服务，实现无缝切换办公地点。同时，用户所有的业务和应用都在数据 中心处理，强大的机房充分保障了业务的连续性，使各类应用稳定进行，有效降低 了办公环境的管理维护成本。

**2.** **服务可度量**

管理学之父德鲁克曾经说过一句名言：“如果您不能测量它，您就不能管理它”。 云计算作为服务提供的方式，需要对服务进行度量。 一般服务提供方和用户之间需 要有一个服务水平协议 (SLA) 。这对于私有云来说，可以根据服务情况进行内部费 用核算；而对于公有云来说，服务可度量就是计费的前提，然后可以根据实际使用 量来进行计费。云计算支持对用户的云服务使用情况进行监控、报告和计费，此特 征能够使云服务的交付实现最优化。用户可以只对实际使用了的资源进行付费。云 计算的服务可度量使得用户在很大程度上不再依赖通过自建计算能力满足需求的模 式，转向按需购买云服务，这样就节约了用户自建成本并提高了资源使用率。而对 于云服务提供商，通过服务度量项衡量云计算的服务能力，可以让云服务提供商事 先权衡成本投入与预期收益的准确预测。

**3.** **多租户**

多租户技术 (Multi-tenancy Technology) 或称多重租赁技术，是一种软件架构技 术，它是在探讨与实现如何在多用户的环境下共用相同的系统或程序组件，并且仍

**第2章■大数据与云端数据治理**

可确保各用户间数据的隔离性。与传统的软件运行和维护模式相比，云计算要求硬 件资源和软件资源能够更好地共享，具有良好的可伸缩性，任何一个用户都能够按 照自己的需求获取云计算资源而不影响其他用户的使用。多租户技术是一项云计算 平台技术，该技术使得大量的租户能够共享同一堆栈的软、硬件资源，每个租户能 够按需使用资源，能够对软件服务进行客户化配置，而且不影响其他租户的使用。

**4.** **按需自助服务**

用户能够按照自己的需求对云服务进行使用和操作，在使用和操作中，用户能 够自动完成或者仅需与云服务提供商进行少量交互。这项核心特征主要是指云服务 能够提供给用户他们所需要的能力，减少用户使用和操作的花费、时间和投入，不 需要额外的或超量的人力方面的交互。

采用云计算自助服务可以将硬件和基础设施建设资金投入转变为按需支付服务 费用，客户无须承担建设和维护基础设施的费用，只对使用的资源付费，避免了客 户自建数据中心的资金投入。云服务商使用多种技术提升资源利用效率，如云基础 设施使用虚拟化、动态迁移和工作负载整合等技术，关闭空闲资源组件，使运行资 源利用效率提高并降低能耗；多租户共享机制、资源的集中共享可以满足多个客户 不同时间段对资源的峰值要求，避免按峰值需求设计容量和性能而造成的资源浪费。 资源利用效率的提高能有效降低云服务的运营成本，减少能耗，实现绿色IT。

**5.** **快速弹性伸缩**

弹性伸缩是指对一个系统适应负载变化进行调控的能力，即短时间内应对突发 性的高流量、高密度业务需求(例如，炎热夏季的午后，电力需求在用户打开空调 后出现峰值)。企业内部物理基础设施资源往往无法满足短时间内迅速获取所需资源 的需求，以及高峰期过后资源闲置的问题。虚拟化技术的应用使得弹性伸缩控制技 术的优势发挥得更加充分。

弹性伸缩能够对物理和虚拟资源进行迅速和灵活的调整，能够迅速地增减资源。 对云服务用户来说，用户能够使用的物理和虚拟资源看起来是没有限制的，而且能 够在满足服务协议的前提下，自动化地在任意时间购置任意数量的资源，能避免因 需求突增导致客户业务系统的异常或中断。这样就能提高资源利用率和用户满意度， 较好地解决资源利用率和应用系统之间的矛盾。

**6.** **资源池化**

只有构建了合理的资源池，才能实现云计算的最终目的——按需动态分配资源。 构建合理的资源池，是实现从传统的“烟囱式IT” 迈向云计算基础架构的第一步。 在传统的“烟囱式”IT 基础架构中，应用和专门的资源捆绑在一起，为了应对少量 的峰值负载，往往会过度配置计算资源，导致资源利用率低下。据统计，在传统的

**■云端数据治理■**

数据中心里， IT 资源的平均利用率不到20%。

构建资源池也就是通过虚拟化的方式将服务器、存储、网络等资源全面形成一 个巨大的资源池。云计算就是基于这样的资源池，通过分布式的算法进行资源的分 配，从而消除物理边界，提升资源利用率，统一资源池分配。

云计算资源池化的特点包含以下3个方面的内容。

(1)平台对不同的设备和虚拟化资源进行整合管理，上层业务感知不到物理设 备的更换和升级以及虚拟化平台的切换，下层硬件设备的更换、升级亦对用户和业 务零感知；

(2)虚拟化和物理服务器统一管理平台资源，兼容业界主流的虚拟产品和操作 系统，以及兼容客户现有的 IT 资源，提供基于资源池的统一编排调度；

(3)多资源池管理可实现设备自动发现、资源的快速发放，有效缩短业务上线 时间，提高维护效率。

**2.1.2** **大数据的特点**

世纪中叶以来，随着各种信息技术和设备深度融入人们的生活，以及各类工业 设备的复杂化和智能化发展，由此产生了数据规模的爆炸式增长。数据体量的增加 一方面是由于数据产生的速度不断变快，另一方面是数据的种类不断增加。面对大 体量、快速产生的数据，人们也在思考如何利用这些数据的价值，不再仅仅将处理 数据作为数据管理的目的和终点，而是将揭示其进一步的价值作为数据管理或数据 治理的新目的。

大数据是数据发展到一定程度的体现，大数据具备数据的一些基本特点，然而 到目前为止仍然没有对大数据的定义形成统一的意见。麦肯锡是研究大数据的先驱 之一，在其报告《Big data:The next frontier for innovation,competition,and productivity》129中给出的大数据定义是：大数据指的是大小超出常规的数据库工具 获取、存储、管理和分析能力的数据集。但它同时强调，并不是说一定要超过特定 TB 值的数据集才能算是大数据。美国国家标准技术研究所 (NIST) 的大数据工作 组在《NIST Big Data Interoperability Framework:Volume 1,Definitions》20中认为： 大数据是指那些传统数据架构无法有效地处理的新数据集。因此，采用新的架构来 高效率完成数据处理，这些数据集特征包括：容量、数据类型的多样性、多个领域 数据的差异性、数据的动态特征(速度或流动率，可变性)。Gartner 在最早给出了 这样的定义，“大数据”需要新处理模式才能具有更强的决策力、洞察发现力和流程 优化能力的海量、高增长率和多样化的信息资产。2001年Gartner 分析师 Doug Laney 提出了“3V” 模型，包括数量 (Volume) 、 速度 (Velocity) 和种类 (Variety) 。 在此 之后，业界很多人把“3V” 扩展到了“4V” 、“5V” 甚至“11V”, 其中还包括有效 性、真实性、价值和可见性等特点。维基百科给出的定义为：“大数据”是指一些使

**第2章-大数据与云端数据治理**

用目前现有数据库管理工具或传统数据处理应用很难处理的大型而复杂的数据集， 其挑战包括采集、管理、存储、搜索、共享、分析和可视化。

除了上述主流的定义，还有人使用3S 或者31描述大数据的特征[?]。 3S指的是：大小 (Size) 、速度 (Speed) 和结构 (Structure)。

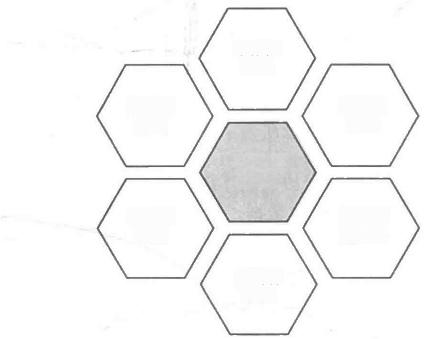
3I指的是：

(1)定义不明确的 (In-defined): 多个主流的大数据定义都强调了数据规模需要 超过传统方法处理数据的规模。而随着技术的进步，数据分析的效率不断提高，符 合大数据定义的数据规模也会相应不断变大，因而并没有 一 个明确的标准。

(2)令人生畏的(Intimidating): 从管理大数据到使用正确的工具获取它的价值， 利用大数据的过程中充满了各种挑战。

(3)即时的 (Immediate): 数据的价值会随着时间快速衰减，因此为了保证大 数据的可控性，需要缩短从数据搜集到获得数据洞察的时间，使得大数据成为真正 的即时大数据，这意味着用户能尽快地分析数据，这对获得竞争优势至关重要。

虽然学术界和产业界对大数据的定义没有形成统 一 的意见，但对大数据的特点 基本形成了 一 致的观点，尤其以“5V” 模型为代表，即大数据的特点包括：数据体 量大、种类多样性、处理速度快、价值密度低和真实有效性等，如图2-2所示。

数据体

量大

处理速

种类多 样性

度快

大数据

特点

价值密

其他特 点……

度低

真实有

效性

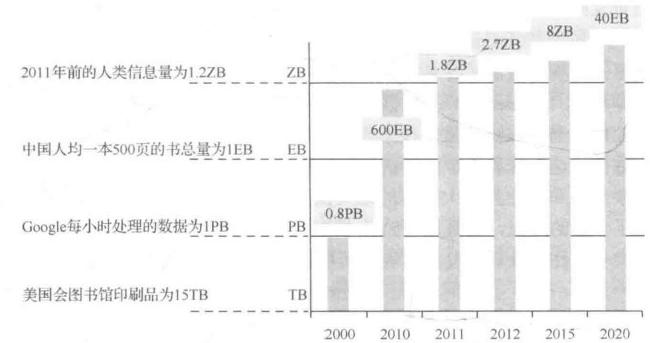
图2-2 大数据特点

**1.** **数据体量大**

据IDC(国际数据公司)研究结果表明，2008年全球产生的数据量为0.49ZB128, 2009年的数据量为0.8ZB,2010 年增长为1.2ZB, 人类正式进入ZB 时代。2011年 的数量更是高达1.82ZB, 相当于全球每人每年产生200GB 以上的数据。而到2012 年为止，人类生产的所有印刷材料的数据量是200PB, 全人类历史上说过的所有话 的数据量大约是5EB 。 更为惊人的是，2020年以前全球数据量仍将保持每年40%多

 **■云端数据治理■**

的高速增长，大约每两年就翻一倍，这与IT 界人尽皆知的摩尔定律极为相似，姑且 可以称之为“大数据爆炸定律”。据 IDC 预测，2020 年全球产生的数据量将达到 40EB, 是2008年的70倍、2011年的29倍，如图2-3所示。因而，数据拥有巨大 的体量是人们对大数据最直接、最直观的理解和认识。另一方面，人类接触到的数 据的种类也在不断地增加。最初人们对数据的理解局限于可以表示成数字的数据， 以结构化数据为主。而新技术的应用和生活方式的转变，人类处理的数据不再局限 于结构化数据，而是将图片、邮件、文档、音频、视频等非结构化数据纳入其中。



**图2-3** **全球产生的数据量变化**

**2.** **种类多样性**

随着传感器、智能设备以及社交协作技术的飞速发展，组织中的数据也变得更 加复杂，因为它不仅包含传统的结构化数据，还包含来自网页、互联网日志文件(包 括点击流数据)、搜索索引、社交媒体论坛、电子邮件、文档、主动和被动系统的传 感器数据等原始、半结构化和非结构化数据。

在大数据时代，数据格式变得越来越多样，涵盖了文本、音频、图片、视频、 模拟信号等不同的类型；数据来源也越来越多样，不仅产生于组织内部运作的各个 环节，也来自于组织外部，甚至各型各色的组织。

例如，在交通领域，北京市交通智能化分析平台数据来自路网摄像头/传感器、 公交、轨道交通、出租车以及省际客运、旅游、化危运输、停车、租车等运输行业， 还有问卷调查和地理信息系统数据。4万辆浮动车每天产生2000万条记录29,交通 卡刷卡记录每天1900万条，手机定位数据每天1800万条，出租车运营数据每天100 万条，电子停车收费系统数据每天50万条，定期调查覆盖8万户家庭等，这些数据 在体量和速度上都达到了大数据的规模。

人们日常工作中接触的文件、照片、视频，都包含大量的数据，蕴含大量的信

**第2章-大数据与云端数据治理**

息。这一类数据有一个共同的特点，大小、内容、格式和用途可能完全不一样。以 最常见的 Word文档为例，最简单的 Word 文档可能只有寥寥几行文字，但也可以混 合编辑图片、音乐等内容，成为一份多媒体的文件，来增强文章的感染力。这类数 据通常称为非结构化数据。

但不幸的是，企业中和人们日常接触到的数据绝大部分都是非结构化的。有的 咨询机构认为非结构化数据占企业总数据量的80%(也有机构认为占95%),总之， 没有一个权威、准确的统计。如何像处理结构化数据那样方便、快捷地处理非结构 化数据，是信息产业一直以来的努力方向之一。

**3.** **处理速度快**

在大数据时代，效率就成了大数据发挥应用价值的重要途径。面对体量庞大且 结构十分复杂的数据，如何快速准确地处理、分析，提取有价值的信息，是应用大 数据的必然之路。而究竟多快才能算快，这是一个相对的概念。举个例子，对于谷 歌来说，必须要在1秒之内完成搜索任务，而搜索结果还必须尽可能地满足搜索者 的需求130]。谷歌等于为大数据搜索应用确立了一个标杆，因为超过1秒钟的数据搜 索应用，会给用户带来不良的使用体验。甚至在某些情况下，如果应用速度达不到 “秒”级，其获得的结果可能会与目标大相径庭。以信息化背景下的军事战争为例， 战场数据集成是大数据技术军事领域运用的关键，随着云计算、通信、媒体和移动 计算的快速发展和深入应用，战场的数据量还将快速增长。在一场空中战斗中， 一 方如果能够从对方战术布置中及时破译信息，并及时通知到正在执行任务的飞行员， 可能会让他及时调整策略，从而出其不意地将对方歼灭。再举一个信用卡消费提醒 的例子：当用户刷卡消费的同时， 一般会收到银行的提示短信，会感到很安全，也 不会认为被打扰，因为当时正在处理跟消费支付相关的事情。如果过了几个小时， 才收到相同内容的短信，情况就不同了，用户也许正在跟朋友聊天，也许正在写一 篇文章，这条短信就成了打扰用户的垃圾信息，客户体验的好坏就在这短短的分秒 之间。

**4.** **价值密度低**

大数据最宝贵之处就在于它蕴藏着巨大的价值，但往往会被淹没在数据大海中， 就像钻石藏身于一堆废石之中，只有挖掘才能发现最耀眼的结果。价值密度低指的 就是从海量价值密度低的数据中挖掘出具有高价值的数据的大数据特点。这一特性 突出表现了大数据的本质是获取数据价值，更多时候体现在商业价值，即如何有效 利用好这些数据获取更大的收益。

举个视频监控的例子：在银行、地铁等一些敏感的部门或者地点，摄像头都是 24小时运转的，这会产生大量视频数据。 一般情况下，这些视频数据非常枯燥、乏 味，并不会引人注目。但是如果恰巧拍到有图谋不轨的人，那么这一帧图像对公安

**■云端数据治理■**

人员来讲，就是非常有价值的了。问题是我们无法在事前知道哪一帧会有用，只好 把所有的视频数据都保存下来，甚至保存了一年的数据，只有那一秒对破案有用。 但是在研究人类行为的社会学家眼中，这些视频可能就是难得的第一手资料，也许 可以借此窥探人类的某些行为模式。

大数据价值的另一体现方式在于关联分析，很多看似没有任何价值的数据通过 分析它们之间关联，便能够发现其中的价值。Facebook、微博为代表的社交网络应 用，构建了普遍关联用户行为数据。本来大家在网络上浏览网页、购买商品，游戏 休闲等，都是互不关联的。尤其是智能手机的普及，大家的网络行为更趋向于碎片 化。这些碎片化数据如果没有关联，是难以进行分析并加以利用的。但是社交网络 提供了统一的接口，让大家无论是玩游戏还是买商品，都能够轻松方便地分享到微 博上。微博扮演了用户行为数据连接器的角色。用户在网络上的碎片化行为，经由 社交网络，就能完整地勾勒出一幅生动的网络生活图景，可以真实地反映用户的偏 好、性格、态度等特征，这其中蕴育了大量的商业机会。

因此基于以上论述，可以得出以下结论：第一，数据是无价之宝；第二，价值虽 有，但却如沙滩中的黄金；第三，数据融合的价值，要远大于种类单一的数据价值。

**5.** **真实有效性**

数据的重要性就在于对决策的支持。数据的规模并不能决定其能否为决策提供 帮助，数据的真实性和质量才是获得真知和思路最重要的因素，是制订成功决策的 最坚实的基础。

追求高数据质量是一项重要的大数据要求和挑战。即使最优秀的数据清理方法 也无法消除某些数据固有的不可预测性，例如，人的感情和诚实性、天气形势、经 济因素以及未来。在处理这些类型的数据时，数据清理无法修正这种不确定性。然 而，尽管存在不确定性，数据仍然包含宝贵的信息。我们必须承认、接受大数据的 不确定性，并确定如何充分利用这一点。例如，采取数据融合，即通过结合多个可 靠性较低的来源创建更准确、更有用的数据点，或者通过鲁棒优化技术和模糊逻辑 方法等先进的数学方法。

**2.1.3** **云计算环境下数据的特点**

云计算和大数据就像一枚硬币的正反面一样密不可分，无论在学术界还是产业 界，在谈论其中一个话题时都无法抛开另一个话题，这种紧密的关系，使得云计算 对数据产生了重要的影响。从前文分析可以看出，云计算和大数据的特点存在诸多 的关联，例如云计算的泛在访问会给云服务提供商产生巨大的访问数据，例如访问 日志，绝大多数情况下，访问日志可能没有多少价值，而一旦在服务器发生故障或 数据泄露时，某一段的访问日志就有可能发挥极大的作用。这就说明虽然访问日志

**第2章■大数据与云端数据治理**

包含巨大的价值，但价值的密度是非常低的。又例如，泛在访问会允许庞大的用户 在同一时间段内同时访问相同的数据，这就要求服务提供商具备较快的数据处理能 力，而一旦能力达不到，便会降低用户的访问体验甚至引起服务的中断。例如早期 的12306网络订票服务和某大型电商的双十一促销活动。再如，云计算的多租户和 按需自助服务特点，允许用户个性化需求的表达。出版社可能更喜欢文字素材，平 面广告商可能更喜欢图片格式的信息，而多媒体广告商则可能更喜欢视频材料。因 此云计算平台处理的数据就会包括文字、图片、视频等多种格式，而随着数量的积 累，就造成了大数据的种类多样性的特点。

云计算和大数据的关系又可以这样简要描述：没有云计算就没有大数据，没有大 数据就无法体现云计算的价值。云计算的发展加快了大数据的形成，而大数据的处理 需求推动着云计算的快速发展。在云计算环境下，数据体现了以下三个主要特点。

第一，云计算环境下数据的体量变得更大、种类更复杂。

IDC 报告显示，预计到2020年全球数据总量将超过40ZB(相当于4万亿GB), 这一数据量是2011年的22倍。2013年中国产生的数据总量超过0.8ZB (相当于8 亿TB),2 倍于2012年，相当于2009年全球的数据总量。预计到2020年，中国产 生的数据总量将是2013年的10倍，超过8.5ZB。 云计算产业的发展速度相对于数 据体量的爆发式增长则慢很多。根据 Gartner报告显示，2011年全球云服务市场952 亿美元，2015年全球云服务市场约为1804亿美元，年增长率约为18%,而到2020 年前，都将维持约15%的增长速率。虽然两则数据没有非常直接的关联，但从云计 算与大数据的融合关系上可以看出，数据的爆发式增长给云计算产业发展带来了巨 大的动力和潜在市场。

数据的爆发式增长，主要在于数据产生者的转变。工业时代以前，数据的主要 来源在于人类自身活动产生的数据。而随着工业社会的到来，机器的大规模应用， 产生了越来越庞大的数据，可以将数据的增长曲线理解为指数函数的初级阶段。而 到了信息时代，数据的增长速度进一步变大，处在了不停翻倍的阶段，即指数函数 的中级阶段。这种数据增速转变的主要原因在于数据的产生者由人类自身变成了机 器，包括机器传感器以及与其他设备通信的智能设备。

数据的爆发式增长，从另一方面看，主要归功于非结构化数据的增长。广义的 非结构化数据也包括了半结构化和非结构化数据，主要是指那些无法用固定结构来 逻辑表达实现的数据。比如说现阶段用户散落在论坛、微博、微信或其他渠道发表 的关于产品的各种评价或吐槽，可能是文字、图片、表情符号，甚至是语音和视频。 目前，非结构化数据普遍被认为占到总量的85%以上，而且增速比结构化数据快得 多。比如 Twitter 刚出来的时候，每天就可以产生一亿条的信息量，新浪微博的信息 量更大，可以达到数十亿条信息，这种产生速度是结构化数据根本无法比拟的。

**■云端数据治理**

第二，云计算环境下数据远程存储和远程传输更常见。

云和数据中心是两个时常被人们互换使用的词，这主要是因为他们在对数据储 存时所起到的作用类似。云和数据中心之间的主要区别是，云是一个数据存储在互 联网公共区域的系统，而数据中心是在本地网络中的存储数据的物理设备的集合。 虽然这两种类型的计算机系统存在一定的差异，但都可以存储数据，尤其是云服务 提供商使用数据中心来支撑云服务和其他云计算资源。过去十多年，数据中心无论 是在技术上还是部署上都极速发展，变化也快，但粗略来说可以将其划分为三个阶 段[31]。

第一阶段是以数据中心大集中为主的 DC 1.0;

第二阶段是利用资源虚拟化和服务动态管理为手段，以增加资源共享利用率和 资源部署灵活度的DC 2.0, 上述两个阶段技术上最大的分界点为云计算技术的日趋 成熟和大批量部署；

第三阶段的DC3.0 是为了适应目前大型网站数据处理急速增长的业务需求而产 生的。

据 Gartner 统计，截至2010年底，全球的数据中心总量为339万个[321,这个数 字包括了为第三方服务的IDC 数据中心、行业和企业自建的数据中心，还包括大量 企业自用的机房。其中美国和西欧的数量分别达到了96万和68 万，占全球总量接 近一半；中国约有45.6万个，占全球总量8%。2012年中国数据中心总数量已经达 到51万个。据预计，到2017年，全球数据中心数量将达到860万个。如此庞大的数 据中心规模， 一方面反映出数据急剧增长后对存储空间的需求，另一方面也反映出数 据的远程存储规模越来越大，毕竟不是每个组织都能够为自己建立一个数据中心。

根据Gartner 的统计，2012年全球共有329EB 个人数据，只有不到7%的个人数 据存储在云端33]。随着消费者使用智能手机和平板电脑产生的数据越来越多(例如 拍照和视频录制),消费者将开始选择本地存储(例如计算机硬盘和移动硬盘)之外 的存储方式。2016年全球个人数字将急剧膨胀到4.1ZB, 随着数据自动化同步到云 端的技术不断成熟，将有36%的个人数据将被转移到云端。另据 Seagate 预计，到 2020年，70%的数据将存储在云端，而在2014年，60%的数据存储在个人端的设备 中[34]。

数据在云端的存储必然对数据远程传输提出更多的要求，例如数据传输的安全、 速率、完整性等。因此，在数据本地化存储空间不足和成本高昂的背景下，数据存 储在云端或异地数据中心将越来越多，远程存储和远程传输成为云计算环境下数据 的特点之一。

第三，云计算环境下数据的安全威胁更多、合规要求更复杂。

在 ISACA 一项调查中，约45%的 IT 专业受访者表示，云计算所涉及的风险已 高于任何利益的回报35]。美国IDG 公司2013年对1300多位企业高管进行调研，66% 的企业高管认为安全问题(包括访问控制、数据保护等)是云计算中的关键因素。

**第2章■大数据与云端数据治理**

近年来，云服务上的数据安全事件频发。2013年，斯诺登“棱镜门”事件曝光了美 国国家安全局的秘密监听计划，事件表明美国政府通过技术手段一直在对各大网络 服务商的服务器进行监听，并对获取的用户数据进行分析。2012年8月，苹果公司 的 iCloud 云服务受到黑客攻击，黑客暴力破解用户密码后，删除了部分用户资料。 而云平台并未备份用户数据，导致了用户数据的丢失。

云端数据安全是数据安全的一个重要研究对象，但云端数据的海量、远程、多 源、异构和动态性等特征导致其与传统封闭环境下数据安全有所区别，必须分析研 究传统数据生命周期各阶段的安全防护技术和产品在开放环境下、云端数据应用场 景下的适用性问题。首先，云端数据与云计算关联十分紧密，云计算本身面临的安 全问题也会转移到其承载的数据上。云计算相关技术仍在快速发展当中，云计算系 统的架构和特性还在发展变化过程中，其脆弱性和安全需求也随之而变化，这为其 安全性的保障带来了困难。

云端数据自身特点带来的安全威胁，可以分为技术和管理两方面。

技术上，云端数据的远程存储对传输性能和安全性提出了更高的要求。其本身大 体量数据的远程传输就是一个难点，而且目前对数据传输的及时性和数据处理的时效 性都有非常高的要求，所以如何既保证数据的高效传输又保证数据的安全是一个技术 问题。数据可用性和完整性的破坏也是云端数据所必须面对的威胁。例如删除大量原 始数据，可能导致永久性的损失；或是攻击者对部分数据进行加密，然后再进行勒索； 对部分数据进行篡改或加入伪造数据，也可能对分析结果产生巨大影响。

在管理上，与本地数据相比，云端数据要面对的问题更多。首先，数据的归属权 问题。 一方面来自于合规要求的差异性，在不同国家和地区会有不同的规定；另一方 面则是云端数据的易获取，数据容易被修改并变成他人的数据。其次，数据的删除。 由于远程存储或者数据分散存储在不同的数据中心上，当数据主体试图删除这些数据 时，如何保证各数据中心的数据都能够真正被删除是个值得重视的问题。再次，数据 收集前的告知形式和普遍存在的过度收集问题，也带来了隐私泄露的威胁。

云计算和数据中心的搭配可以让数据的流通范围更广，很多跨国公司在全球范 围内建立了数据中心。 一般情况下用户并不会知道数据的存放地，敏感数据是否会 被带出国也未可知。不过好在每个国家对数据的存储都会有法律要求，例如我国《地 图管理条例》136],其中一章专门与互联网地图服务有关。《地图管理条例》称，互联 网地图服务单位需要将存放地图数据的服务器设在中国境内，“互联网地图服务单位 用于提供服务的地图数据库及其他数据库不得存储、记录含有按照国家有关规定在 地图上不得表示的内容。互联网地图服务单位发现其网站传输的地图信息含有不得 表示的内容的，应当立即停止传输，保存有关记录，并向县级以上人民政府测绘地 理信息行政主管部门、出版行政主管部门、网络安全和信息化主管部门等有关部门 报告。”以及“任何单位和个人不得通过互联网上传标注含有按照国家有关规定在地 图上不得表示的内容”。在美国和欧盟之间存在着一个与数据跨境流动的“美国-欧

 **■云端数据治理■**

盟安全港协议”。由于美国缺乏统一的数据保护法，未能达到欧盟指令的充分保护要 求，这将妨碍个人数据在美欧之间跨境转移。为了解决这一问题，美国和欧盟于2000 年达成了一个折衷的安全港协议，安全港包含一系列隐私原则，包括通知、选择、 数据转移、数据获取、数据安全、数据完整以及执行等七大原则，用以保护个人数 据。2015年，欧盟单方面废止了欧美数据安全港协议，让欧美之间数据交换产生更 大的困难。2016年4月14日，欧洲议会通过了《一般数据保护条例》,该条例将在 2018年生效，届时在欧洲从事数据业务的组织必须符合该条例要求。因此，云计算 环境下数据的存储和流动合规要比传统的数据面临更复杂的要求。

**2.2** **云端数据治理介绍**

云计算环境下数据特点发生了很多的变化，尤其是数据在云端的操作给数据治 理带来了很多新的内容，如云端数据来源的多样性、云端数据的远程存储与传输、 云端数据的合规要求更复杂性、云端数据面临的安全威胁更多等，都超越了传统数 据治理的范畴。即便个别现有数据治理理论能够兼顾云计算环境下数据的特点，但 也无法覆盖到云端数据治理的全部需求。因此开发针对云端数据的治理理论和实践 方法是对云端数据进行管理并发挥其应用效果的必然选择。

**2.2.1** **云端数据治理的定义**

云端数据治理是以云端数据为主要治理对象，制订与云端数据战略、数据管理、 数据优化、数据安全与隐私保护等相关策略，指导组织规划、构建、评估和优化数 据治理体系的活动集合。

我们将上述云端数据治理的定义分解为以下部分。

(1)云端数据治理是以云端数据为主要治理对象

数据按存储地可以分为本地数据和云端数据。数据本地存储是互联网发展起来 之前数据存放的主要形式，此时的数据量是相对较小的，组织对其管理也相对容易。 随着互联网、云计算等技术的发展和广泛应用，越来越多的用户选择把数据存储在 云平台上，而放弃了原有的模式，也就是把数据存在企业内部本地的数据中心机房 里面。这一方面是成本因素，本地存储的前期成本远远高于云存储。随着组织数据 量不断增大，必须部署的基础设施和硬件就会越多，维护和更新成本也越来越大。 提前预算足够的存储空间，在组织还没有足够多的数据完全占用之前，这些设备因 为空闲会造成资源的浪费；而当数据占据了现有存储空间后，组织又不得不对数据 中心进行升级。另一方面，组织的业务覆盖的地理范围越来越大，尤其是跨国型组 织，本地化数据存储方式不利于跨区域业务中心对数据的实时调取要求。

**第2章=大数据与云端数据治理**

云端数据的产生主要来源于两大部分， 一是组织本地数据的云端迁移，二是云 端自身数据的聚合。从数据成分上来看，云端数据除了是传统格式化数据外，更多 的是半结构化和非结构化数据，因为非结构化数据的爆炸性增长是迫使组织将数据 迁移至云端的主要原因之一。同时，源于互联网、物联网、移动互联网等技术的数 据中非结构化数据已经占据绝大部分，这些数据可以不经本地存储而直接在云端汇 聚，成为组织可处理的数据。

(2)云端数据治理要制订与云端数据战略、数据管理、数据优化、数据安全与 隐私保护等相关的策略。

云端数据治理关乎策略的制订，这些策略成为开展数据治理的指导文件和措施 依据，包括数据治理战略、数据管理策略、数据优化策略、数据安全与隐私保护策 略等。

战略是一项活动的宏观性纲领，是一种从全局考虑谋划实现全局目标的规划。 对于一个组织，大的战略可以是一个从组织层面考虑的全局战略，也可以是针对某 项活动制订的战略，如数据治理战略、信息安全战略等。数据治理战略是对数据治 理活动的顶层设计和远程规划，其目的是在数据治理活动得到数据治理相关者的理 解和认同的前提下，对未来一段时间内数据治理活动进行规划。

数据管理是对数据进行基础规范以支撑数据治理整个活动的实施，涉及业务术 语、元数据管理和主数据管理等。业务术语是数据治理的基础，是为了获得组织对 某项事情的统一认识，不然可能出现众说纷纭的现象从而造成重大错误。例如对数据 治理成本进行估算时，究竟还是以“万元”为单位还是以“元”为单位，又或是当进 行数据整理时，小数点后数值保留几位，是要“四舍五入”还是“取整去零”等。

云端数据必须优化，不是所有数据都可以直接使用的。 一方面，云端数据拥有 大数据所具有的特点，数据体量大、结构复杂，造成数据价值密度低，所以必须对 云端数据进行优化。另一方面，为了最大程度上发挥数据的价值，也必须对数据进 行整理优化，通过数据聚合分析发挥单体数据所不能体现出来的价值。数据质量管 理、数据操作、数据架构都是为了实现数据优化而进行的数据治理措施。

数据安全与隐私保护必须重视。在大数据和云计算环境下，数据面临的安全威 胁更多，破坏力更强，因此云端数据治理的一项重要内容就是数据的安全与隐私的 保护。例如在处理社交媒体、地理定位、生物计量学和其他形式的个人可识别信息 时，必须考虑其涉及的声誉和法律法规风险。另外，不同地域或国家对数据安全管 理和隐私保护的规定是不一样的，所以保证合规是数据治理的重要前提。

(3)云端数据治理是规划、构建、评估和优化数据治理体系的活动集合

云端数据治理是依据制订的各项策略对云端数据进行治理的活动集合，为此组 织需要建立一个完整的数据治理体系，覆盖统筹和规划、构建和运行、监督和评估、 改进和优化四大阶段，促进云端数据治理的落地以实现既定目标。

统筹和规划阶段是云端数据治理活动的前期准备阶段，其目标是要统筹和规划

**■云端数据治理■**

数据治理目标和任务，谋划必要的治理环境，做好数据治理实施的准备。统筹和规 划阶段的主要内容包括：

a) 识别云端数据治理需求和任务；

b) 制订云端数据治理战略，规划数据治理实施的目标、方法和范围；

c) 明确云端数据治理的组织机构、治理主体、职责分工以及利益相关方的沟通 机制；

d) 统筹规划资源，制订业务案例，明确实施内容、程序和机制；

e) 根据组织战略、业务要求和利益相关方期望，建立与云端数据治理相适应的 路线图和绩效考评体系；

f) 规划数据治理的绩效评价和审核机制，明确审计监督的相关要求。

构建和运行阶段是云端数据治理策略的实施阶段，其目标是构建数据治理实施 的管理机制，确保数据治理实施的有序运行。构建和运行阶段的主要内容包括：

a) 建立云端数据治理实施相关的管理机制，确立实施要点；

b) 制订满足整体规划的构建和运行计划，并按计划实施；

c) 制订数据治理实施的工作标准与方法；

d) 建立信息交流与共享途径；

e) 对构建和运行过程进行管理，确保过程可追溯，结果可计量或可评估。

监督和评估是云端数据治理实施效果评估的阶段，其目标是监督云端数据治理 的实施过程，评估云端数据治理实施的符合性和实施的质量。监督和评估阶段的主 要内容包括：

a) 依据绩效评价和审核机制，对治理实施过程进行监督；

b) 依据云端数据治理目标，对治理实施结果达成程度进行评估；

c) 评估治理组织与角色的能力；

d) 监督和评估云端数据治理实施的适宜性和有效性；

e) 监督和评估数据治理体系的实施和落地。

改进和优化阶段是云端数据治理的过程和结果提升阶段，其目标是促进组织持 续改进云端数据治理实施的过程，提升治理实施的有效性。改进和优化阶段的主要 内容包括：

a) 建立云端数据治理实施的改进和优化机制，并对改进和优化过程进行监督；

b) 依据统筹规划的目标和要求，对未达成的指标进行检查和分析；

c) 确定改进措施，制订治理实施改进计划；

d) 持续推进云端数据治理实施的改进和优化。

通过策划、实施、检查和调整 (PDCA) 的活动循环，持续改进云端数据治理体 系，因此云端数据治理是一系列活动的集合。

**第2章■大数据与云端数据治理**

**2.2.2** **云端数据冶理与数据冶理的关系**

DAMA 将数据治理定义为对数据资产管理行使权力和控制的活动集合(规划、 监控和执行)。桑尼尔·索雷斯在《大数据治理》一书中将大数据定义为广义信息治 理计划的一部分，即制订与大数据有关的数据优化、隐私保护与数据变现的政策。 《大数据治理与服务》则将大数据治理定义为对组织的大数据管理和利用进行评估、 指导和监督体系的框架，通过制订战略方针、建立组织架构、明确职责分工等，实 现大数据的风险可控、安全合规、绩效提升和价值创造，并提供不断创新的大数据 服务。由此可见，我们定义的云端数据治理与传统数据治理、大数据治理的定义都 存在紧密的关联，这绝非特意使然，而是它们之间在本质上存在诸多的关联。

云端数据治理与数据治理和大数据治理的相同点至少包括以下三点。

●首先，它们的治理的对象都是数据。无论是云端数据还是大数据，都可以看 成数据发展到一定阶段的体现。

● 其次，它们都关乎数据管理的策略。对比三者主要内容可以发现，它们都包 括了治理的组织、数据质量、数据生命周期管理、数据操作、数据安全等方 面内容，通过制订这些内容相关的策略以指导数据治理的实施。

●再次，它们具有相似的治理过程。大致的治理流程都可以概括为：数据治理 策划、实施、检查、改进四大阶段。

所以，云端数据治理是数据治理理论在新环境下发展的产物，原有的数据治理 基本理论仍然是云端数据治理的基础。表2-1 概述了云端数据治理与数据治理和大 数据治理在概念上的关系。

**表2-1** **云端数据治理与数据治理和大数据治理在概念上的关系**

|  |  |  |
| --- | --- | --- |
| 内容 | 类 别 | 要 点 |
| 概念 | 云端数据治理 | (1)云端数据治理以云端数据为主要治理对象  (2)制订与云端数据战略、数据管理、数据优化、数据安全与隐私保护等相关的 策略  (3)指导组织规划、构建、评估和优化数据治理体系 |
| 大数据治理 | (1)大数据治理是广义数据/信息治理计划的一部分  (2)大数据治理制订与大数据优化、隐私和货币化相关的政策  (3)大数据治理需要协调多个职能部门的目标和利益  (4)大数据治理是对组织的大数据管理和利用进行评估、指导和监督的体系框架 |

 **■云端数据治理**

续表

|  |  |  |
| --- | --- | --- |
| 内容 | 类 别 | 要 点 |
|  | 数据治理 | (1)数据治理是指“决定如何做决定”  (2)数据治理是评估数据利益相关者的需求、条件和选择以达成一致的数据资源管理 的目标，通过优先级排序和决策机制来设定数据管理职能的发展方向，然后根据方向  和目标来监督数据资源的绩效与合规  (3)数据治理的业务职能是评估、指导和监督  (4)数据治理的核心是数据资产管理的决策权分配和职责分工  (5)数据治理需要明确组织架构、控制、政策和过程，并制订相关规则和规范  (6)数据治理应遵循标准的、成熟的、获得认可的过程，并且严格遵循相关规范 |
| 关系 | 云端数据治理 与数据治理 | (1)云端数据治理本质也是数据治理，是数据治理在云计算环境下的一个新发展方向 (2)数据治理的方法论仍然适用于云端数据治理  (3)数据安全和隐私保护在云端数据治理体系中变得更加重要  (4)云端数据治理不仅要管理好数据，也强调云端数据的应用创新 |
| 云端数据治理 与大数据治理 | (1)云端数据治理和大数据治理都是数据治理发展的一个新方向或新阶段  (2)云端数据治理的对象可能是传统的数据，也可能是大数据，因此大数据治理体系  的方法论也适用于云端数据治理  (3)数据安全和隐私保护在云端数据治理和大数据治理体系中都得到了重视 |

然而，云端数据治理与传统数据治理和大数据治理也存在诸多差异。

首先，治理对象的结构不一样，虽然云端数据本质上仍然属于数据，但正是由 于数据内容发生极大的变化，使得传统数据治理难以适应云计算环境下的数据治理 需求。云端数据无处不在，各种数据并存且相互关联，尤其是数据聚合分析给数据 相关者带来了更大的威胁，这些现象超出传统数据治理的范围。

其次，面临的环境不一样。传统数据治理所处阶段数据体量小，结构简单，数 据治理的主要目的是管理好数据。云端数据治理是在云计算和大数据特殊环境下产 生的新理论和新实践，云计算的发展给数据体量和安全等方面造成很大影响，大数 据的到来也使得传统数据治理理论难以完全适用，促使数据治理向大数据治理和云 端数据治理等发展。

再次，治理的内容有很大差别。云端数据面临的环境不同，迫使云端数据治理 要关注更广泛的内容，需要在传统数据治理的基础上，考虑云计算和大数据环境对 数据治理的影响。例如，云端数据的远程分布式存储和远程传输给数据安全带来了 更多的挑战，传统数据治理可能还无法完全适应当前新环境下的治理需求。表2-2 简要描述了云端数据与传统数据治理部分内容的对比。

**第2章■大数据与云端数据治理**

**表2-2** **云端数据与传统数据治理内容的对比表**

|  |  |  |
| --- | --- | --- |
| 数据治理内容 | 传统数据治理理解 | 云端数据治理区别点 |
| 数据管理目标 | 主要目标是管理好数据，避免出现问题 | 既要管理好云端数据，也要利用好云端数 据，发挥数据价值 |
| 数据管理策略 | 定义数据管理方案的愿景、目标和目的 | 云端数据治理需求和范围发生很多变化 |
| 沟通 | 更多是关注组织范围内的沟通 | 数据供应链的扩展，沟通范围扩大，内容更 广 |
| 方案资金 | 主要关注资金保障 | 组织面临的环境更复杂，除了保障资金以  外，还要保障人员、技术、设施、场地等资源 |
| 数据存储 | 本地化储存为主 | 云端远程分布式存储为主 |
| 数据治理 | 原始数据是基础，将数据抽取过来，清洗 完，整合到主题域和多维模型里，然后商业 智能基于主题域和多维模型做各种分析 | 数据控制权限的丧失(特别是在公有云环境 下)导致数据的采集、清洗、删除等变得困难 |
| 平台和架构 | 数据操作和管理的相关文档要求和形成 的记录 | 云平台的管理权限受限，影响了记录的获取 和分析 |
| 数据质量 | 对支持业务需求的数据进行全面质量管  理，通过数据质量相关管理办法、组织、流 程、评价考核规则的制订，及时发现并解决 数据质量问题，提升数据的完整性、及时性、 准确性及一致性，提升数据价值 | 海量、异构、不确定(随时变化)的数据带 来的数据清洗困难 |
| 数据操作 | 本地化数据的全生命周期管理 | 云端数据的生命周期管理，数据来源有效复 杂，数据存储更分散，数据彻底删除难以保障 |

**2.3** **云端数据治理的作用和价值**

数据已成为组织的一项资产，不仅需要管理好，还要利用好，以实现从数据中 挖掘更大的价值。然而，目前大部分组织的数据治理水平总体较低，普遍存在“重 采集轻管理、重利用轻安全”的现象，组织数据管理能力不足，数据质量不够高， 难以支撑数据驱动的智能判断，且数据的安全存在诸多隐患。

随着数据体量的爆发式增长，以及数据格式的复杂化，传统数据治理理论难以 满足大规模数据的治理需求。加上云计算、大数据等新技术的发展，给数据治理带 来了更大的挑战。云端数据治理针对云计算环境下数据治理的需求，通过优化数据 治理战略、建立更加完善的数据治理流程、强化数据治理保障，提升云端数据管理 水平。开展云端数据治理的重要作用和价值除了一般数据治理能带来的好处外，还 可以概括为以下三点。

**■云端数据治理■**

(1)治理对象更明确，帮助组织提升云计算环境下数据治理的能力

云计算快速发展和应用，促进数据体量的快速增加，无论是现有数据在云端存 储还是从云端直接获取数据，组织对云端数据的关系正在从接触向依赖过渡，将会 有越来越多的组织要处理云端数据，因此云端数据治理将成为众多组织无法回避的 话题。云计算环境下数据的体量变得更大、种类更复杂，数据远程存储和远程传输 更常见，数据安全威胁更多、合规要求更复杂，传统数据治理已经无法满足云端数 据治理的需求。数据云端数据治理是以云端数据为主要治理对象，是在数据治理基 本理论上建立起的新的数据治理理论，目标更加清晰，更具有针对性，帮助组织提 升云计算环境下数据治理能力。

(2)为组织云端数据治理建立清晰的决策规划和决策流程

数据治理是一项复杂的持续改进过程，而不是一个一蹴而就的工程，需要不断 评估现状，不断改进。云端数据治理从治理战略、组织架构、数据质量、数据操作、 管理平台等多个方面，通过策划、构建、评估和改进，建立一个完善的治理管理体 系，为组织海量的云端数据治理工作提供清晰的路线规划，促进各部门人员共同参 与决策。

(3)以成熟度等级方式帮助组织准确评估云端数据治理现状

数据管理是对数据的规划、控制，发挥数据和信息的价值，数据治理通过制订 正确的原则、政策、流程、操作规程等，保证数据管理的有效执行，解决企业对数 据概念理解不一致、统计口径不协调、管理过程不规范等问题。数据治理是一个涉 及管理与技术同步协调进行的活动，治理的组织架构也强调不同领域业务人员的深 度参与和决策。通过数据治理的开展，帮助组织构建一个自下而上的立体化治理组 织架构，涵盖部门基层员工到首席数据官，统一分配职责和权限。数据治理模型以 成熟度等级形式展示，通过全面分析组织数据管理现状，进行差距分析，给出改进 建议，解决组织因数据管理不善引起的难题。

第 3 章 云端数据治理模型与

治理成熟度评估模型

**3.1** **云端数据治理模型概述**

**3.1.1** **云端数据冶理模型**

**1.** **云端数据治理模型的内容**

云端数据治理是数据治理理论的发展，是传统数据治理的传承。基于原有的成 熟方法论，结合治理对象——云端数据的特点，通过加入新理论，创新治理理念， 本章提出了云端数据治理模型。云端数据治理模型的由来如下。

(1)借鉴CMMI 数据管理成熟度 (DMM)

DMM 是一个用于管理组织数据资产和相应活动的过程改进成熟度模型，该模 型包含了确立、建立、维持和优化整个数据生命周期的有效数据管理的最佳实践， 覆盖了数据的产生到交付、维护和归档整个阶段。

DMM 由五大职能域和一个支持过程组成。五大职能域分别为战略、治理、数 据质量、操作、平台及架构， 一共包括20个数据管理过程域。支持过程包括5个过 程域，用于保障五大过程域的实施。

由上可见，DMM 覆盖了数据治理的主要内容，尤其是提出了以成熟度模型来 评价数据治理效果的思想，为组织评价数据治理实施效果，明白所处的治理水平提 供了重要参考标准。基于此，云端数据治理也采用了成熟度评价模型。

(2)针对云端数据，扩展治理范围

如前文所述，云计算环境下，数据本身及外界环境发生了很多变化，如云计算 环境下数据的体量变得更大、种类更复杂，数据远程存储和远程传输更常见，面临 的安全威胁更多，合规要求更庞杂等，使得数据治理的边界和内涵进一步扩大。

尽管 DMM 对有效数据管理的要求和活动进行了定义，但只是针对之前的数据 环境。由于云计算和大数据等技术的快速发展，数据面临的环境和数据治理环境都 发生了诸多变化，原有的 DMM 无法完全适应现有环境下数据治理的需求，例如云 端数据跨地域流通面临的合规要求更多，云计算自身安全问题给云端数据带来了更 多的新风险，因此需要在 DMM 基础上增加新的内容。除了内容上无法完全适应云

**■云端数据治理**

战 略

沟 通

风险管理

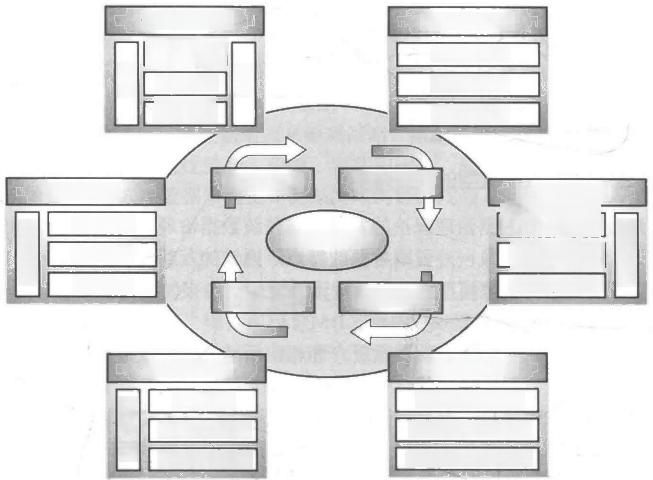
数据清洗

架构方法

端数据， DMM 也没有给出指导组织实施数据治理的方法。

云端数据治理模型是在充分吸收 DMM 等国际上最被认可的数据治理模型的基 础之上，结合专业的评估理论和经验开发而来的，旨在为组织提供理论和实践指导。 云端数据治理模型的结构非常清晰，组织不仅可以利用该模型来评价自己目前的能 力状态，还可以为数据治理执行建立一个清晰的路径图。

云端数据治理模型从全局视角描述了云端数据治理的主要内容和实施过程，如 图3-1所示。

云端数据战略

组织与角色

业务案例

资源保障

云端数据管理

业务词汇表

元数据管理

主数据管理

安全与隐私

数据安全

隐私保护

合规管理

统筹和规划 构建和运行

云端数据治理

改进和优化 监督和评估

云端数据质量

数据质量战略

数据概要分析

数据质量评估

云端数据架构

云端数据操作

架构标准

数据管理平台

历史数据管理

数据提供者管理

数据集成与互操作

数据生命周期管理

**图** **3** **-** **1** **云** **端** **数** **据** **治** **理** **模** **型**

云端数据治理模型由云端数据战略、云端数据管理、云端数据质量、云端数据 操作、云端数据架构、安全与隐私六大职能域构成，云端数据治理模型过程域及定 义见表3-1。而云端数据治理体系则包括了统筹和规划、构建和运行、监督和评估、 改进和优化四个阶段，构成了一个完整的闭环体系。

**表** **3** **-** **1** **云** **端** **数** **据** **治** **理** **模** **型** **过** **程** **域** **及** **定** **义**

|  |  |  |
| --- | --- | --- |
| 职 能 域 | 过 程 域 | 定义&功能 |
|  | 战略 | 定义组织云端数据治理的愿景、目标和目的，说明组织为什么实施云端  数据治理项目，该项目要达到什么样的目标，以及需要做什么 |
| 组织与角色 | 明确云端数据治理组织架构与成员，分配相关职责 |
| 业务案例 | 对云端数据治理方案、倡议或具体项目进行资助的提案，描述治理的范 围、活动和预期结果 |

**第3章■云端数据治理模型与治理成熟度评估模型**

**续表**

|  |  |  |
| --- | --- | --- |
| 职 能 域 | 过 程 域 | 定义&功能 |
| 数据战略 | 资源保障 | 为云端数据治理活动提供所需的人力、财力、物力等资源，规划支持方案 |
| 沟通 | 确保战略、标准、过程等云端数据治理沟通能够有效发布、理解和反馈 |
| 数据管理 | 业务词汇表 | 为云端数据治理相关的词汇和定义提供统一的解释 |
| 元数据管理 | 为获得高质量的、整合的元数据而进行的规划、实施与控制活动 |
| 主数据管理 | 指一整套的用于生成和维护企业主数据的规范、技术和方案，以保证主 数据的完整性、 一致性和准确性 |
| 数据质量 | 数据质量战略 | 为实现所需求的质量数据而制订的支持策略 |
| 数据概要分析 | 通过对数据自身统计值分析来反映出数据质量问题 |
| 数据质量评估 | 依据过程、技术和数据质量规则采取系统性的方法对数据质量进行评估 |
| 数据清洗 | 定义机制、规则和方法以用于数据的验证和纠错 |
| 数据操作 | 数据提供者管理 | 定义数据采购需求、获取数据、管理协议以及与提供者沟通的实践，通 过优化内部和外部数据源来提供数据质量以满足业务需求 |
| 数据集成与互操作 | 把不同来源、格式、特点性质的数据在逻辑上或物理上有机地集中，从 而为组织提供全面的数据共享，并确保数据在云服务器间的互操作性 |
| 数据生命周期管理 | 依据不同数据在不同阶段的价值实施不同的管理策略 |
| 数据架构 | 数据架构方法 | 设计实施最优的数据分层使得数据的获取、生成、存储和传输满足业务 和技术目标 |
| 数据架构标准 | 定义一套用于开发、批准和实施数据治理合规的方法和实践，用于数据 表示、数据访问和数据分布的标准 |
| 数据管理平台 | 确保实施并管理了一个有效的平台来满足业务需求 |
| 历史数据管理 | 对历史数据进行管理，以满足组织调用等需求，并确保满足合规要求 |
| 安全与隐私 | 风险管理 | 识别和分析潜在的问题，以便采取适当的措施确保目标可以实现 |
| 数据安全 | 通过技术或非技术方式保证数据的访问受到合理控制，并保证数据不被 人为或意外的损坏而造成泄露或更改 |
| 隐私保护 | 确保数据治理过程中的隐私保护以及以数据为手段增强隐私保护 |
| 合规管理 | 为确保云端数据治理活动与其所适用的国家法律、监管规定、行业规则、  自律准则等观念相一致而进行的管理 |

**2.** **云端数据治理的原则**

云端数据治理原则是指导云端数据治理活动的基本准则，治理活动要在符合治 理原则的前提下进行。根据云端数据治理活动特点，结合国家标准《信息技术服务 治理 第1部分：通用要求》等指导规范，云端数据治理活动的原则包括全局统筹、

**■云端数据治理■**

风险可控、运营合规、创新应用四大原则。

(1)全局统筹

云端数据治理是一项涉及多个部门、多种角色的信息治理工程，与组织业务息 息相关，因此数据治理是组织治理的一个重要组成部分。因此，云端数据治理战 略应与组织的整体战略保持一致，云端数据治理目标的制订要充分考虑组织的业 务目标，云端数据治理活动路线图应与组织对数据要求紧密关联在一起。组织应 明确云端数据治理责任人、建立云端数据治理机构、明确职责与权利、保证利益 相关方理解。

(2)风险可控

数据既是组织的一项重要资产，因其重要性，也成为风险的来源。好的数据有 助于组织做出更合适的决策，坏的数据则可能误导组织判断。云计算环境下，数据 面临更多的安全威胁，例如云端数据治理会涉及到远程数据的解密和传输等安全问 题，组织应确保数据治理的初衷是管理好数据、利用好数据，不能让数据暴露于因 数据治理可能带来的安全风险之中。组织在开展云端数据治理项目时，要在数据治 理前评估数据治理可能带来的安全风险，采取有效的措施降低或避免风险。在数据 治理过程中，要有计划地开展风险评估工作，重点关注安全和隐私问题，防止未授 权或不恰当访问、获取和使用数据。

(3)运营合规

数据会涉及敏感信息，不同国家和地方对敏感数据的定义也不尽相同，因此对 数据的收集、处理、存储等数据生命周期活动有很多不相同的要求，尤其是云端数 据的远程和分布式存储，组织的同一份数据可能存储在不同的地方，而在组织要收 集数据时，数据可能来源不同的地方、不同的行业等，数据采集规范存在很多差异。 因此，组织在进行云端数据治理时，要建立长效机制来及时不间断地收集和掌握云 端数据的监管要求，制订相应的沟通内容，将合规性传达给相关人员，让合规运营 成为云端数据治理的一项重要前提。

(4)创新应用

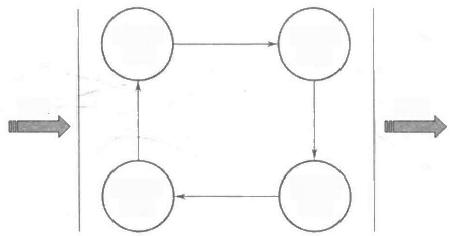
数据管理与数据治理的一项重要区别在于是否重视数据的创新应用，开展云端 数据治理的一项重要目的是管理好数据、使用好数据，在管理好数据的基础上，发 挥数据治理对组织分析和决策的重要作用，倡导数据驱动创新战略。创新应用的途 径有很多，例如通过数据治理提升组织工作流程，提高组织预判能力，改善用户体 验，促进资源合理配置，创造新产品、新服务等。因此组织在开展云端数据治理时， 要将创新应用作为数据治理的目标之一。

**3.** **云端数据治理的实施**

为保障云端数据治理的实施，组织需要建立一个完整的数据治理体系，覆盖统 筹和规划、构建和运行、监督和评估、改进和优化四大阶段，云端数据治理实施流

**第3章■云端数据治理模型与治理成熟度评估模型**

程如图3-2所示。



统筹和 规划

输入

改进和 优化

监督和 评估

构建和 运行

输出

**图3-2** **云端数据治理实施流程**

云端数据治理是一系列流程和活动的集合，不同类型和规模的组织开展数据治 理的流程可能存在一些差异，但总体流程应该是相似的，以下简要描述了数据治理 实施的基本过程，组织可以参照该过程完善数据治理流程。

首先，组织要做的是识别关键问题，确定数据治理目标，规划数据治理的总体 流程。在识别关键问题时，组织可以从对特定问题或领域进行数据治理的紧迫程度 以及数据治理能够带来的价值来综合衡量，对要解决的问题或治理领域排序，设定 优先级顺序。

其次，围绕数据治理目标，制订数据治理实施的管理机制和配套资源。组织需 要组建有足够代表性和权限的数据治理组织架构，使之可以贯穿整个组织，以便各 业务、技术和管理部门对整个信息供应链进行治理。建立沟通机制，确保数据治理 的沟通能够有效传达、理解和反馈。

再次，开展云端数据治理的各项具体措施，例如定义数据词汇表，制订数据架 构标准，对数据进行管理和操作，采取安全防护措施保证数据安全，识别法律法规 和相关要求，保证数据治理的合规性。

最后，组织要定义考核和评价指标体系，针对已开展的数据治理的效果进行评 估，确定数据治理结果与目标的符合程度，发现策略漏洞和治理流程不完善之处。 评估的方式包括组织自评和独立第三方评估，组织依据评估结果和改进建议，调整 数据治理策略。

**3.1.2** **成熟度等级定义与特征**

**1.** **初始级**

1 ) 定 义

云端数据治理措施未经严格的计划和跟踪，而是依赖于数据管理人员的经验及 个人意识进行，所开展的数据治理活动的一致性、性能和质量存在不稳定性或不可

 **云端数据治理**

重复性，开展数据治理的范围较小。

2)特征

达到云端数据治理成熟度初始级应具备以下特征。

a) 组织开展了临时性的云端数据治理项目，但涉及范围在部门级；

b) 云端数据治理的主要目的是预防数据管理不善可能给组织带来的风险； c) 管理层对实施云端数据治理基本没有意识；

d) 各部门间数据治理的流程有很大差异。

**2.** **管理级**

1)定义

云端数据治理措施得到了良好的规划和确定，初步建立了云端数据治理管理制 度、治理流程制度、实施指南等制度性文件。注重于数据治理管理的制度化，这些 规划和制度符合相关的标准和需求，使组织的数据治理工作有据可依。管理级与初 始级之间的主要区别是前者云端数据治理措施被良好规划和确定。

2)特征

达到云端数据治理成熟度管理级应具备以下特征。

a) 组织开展了云端数据治理项目，涉及范围跨部门级；

b) 初步建立了云端数据治理管理制度、治理流程制度、实施指南等制度性文件；

c) 管理层对实施云端数据治理具备一定的意识，并予以支持； d) 定义了有限的管理角色和管理职责；

e) 制订了跨部门的统一数据治理流程，建立了广泛的协作机制。

**3.** **定义级**

1)定义

组织建立了较完善的云端数据治理体系，形成了较为完善的人员、过程、技术 和资源方面的管理制度，能够切实保障云端数据治理工作的实施。在数据治理体系 中，制度文件充分定义，能够在组织级指导数据治理相关人员开展具体工作，数据 治理流程是组织级统一的，治理活动是可重复执行的。定义级与管理级的主要区别 在于对云端数据治理体系所规定和要求的工作切实进行，而且具有完整的实施记录 可追踪。

2)特征

达到云端数据治理成熟度定义级应具备以下特征。

a) 建立了较为完善的云端数据治理体系；

b) 充分定义了角色和职责；

c) 管理者能够积极推动数据治理工作；

d) 数据治理过程被记录、可追踪。

**第3章■云端数据治理模型与治理成熟度评估模型**

**4.** **量化级**

1)定义

组织为云端数据治理制订了可测量的评价指标，并收集、分析云端数据治理过 程的详细记录数据，以衡量数据治理效果。量化级与定义级的主要区别在于对云端 数据治理活动的实施进行定性或定量的评估。

2)特征

达到云端数据治理成熟度量化级应具备以下特征。

a) 建立了可执行的云端数据治理评价指标体系；

b) 云端数据治理策略和流程能够基于量化指标进行修正。

**5.** **优化级**

1)定义

基于组织的业务目标建立了云端数据治理有效性和效率的量化执行目标。过程 的执行以及试验性的新概念和新技术产生了量化反馈，从而使基于这些目标的持续 性的过程改进得到了实现。组织对云端数据治理过程进行总结，形成最佳实践。优 化级与量化级的主要区别在于基于测量的评估结果，对已定义的策略和标准进行不 断地改进和优化。

2)特征

达到云端数据治理成熟度优化级应具备以下特征。

a) 基于可执行的云端数据治理评价指标体系，对数据治理过程持续评估；

b) 基于评估结果，对数据治理流程持续优化改进；

c) 组织能够根据数据治理成果进行决策。

**3.2** **云端数据治理成熟度评估模型**

**3.2.1** **成熟度等级判别标准**

评估将采取定性的方法进行判定。针对云端数据治理的某一过程域，分析各条 成熟度等级要求及与之关联的管理过程中的管理、测量和制度化，判定其表现出的 特征是否满足某一成熟度级别要求，如果满足，则可判定此项活动处于对应的成熟 度等级。

依照此种办法，对于云端数据治理职能域中的每一个过程域，都给出一个对应 的成熟度级别。由于一个模块中，可能存在处于不同能力级别的过程域，最终该模 块的整体成熟度级别，取该模块中各过程域能力级别的最低值。

**■云端数据治理■**

**3.2.2** **成熟度评估方法**

治理成熟度评估是对云端数据治理进行监督、评价和改进的重要途径，通过综 合检查和评价，发现数据治理过程的不规范项，提出评价意见和改进建议，促进组 织实现云端数据治理的目标。

云端数据治理成熟度评估方式可以分为组织自评和第三方独立评估两种。组织 自评是组织建立的一种自我检查、自我完善的持续改进活动，以组织内部人员为评 估主体，对数据治理活动进行评估，能够为数据治理体系的持续优化提供改进依据。 第三方评估是由独立于组织且不受其经济利益制约或不存在行政隶属关系的第三方 机构遵循云端数据治理成熟度评估依据，按规定的程序和方法对组织的云端数据治 理项目进行评估。第三方评估是中立性的评估模式，且由于其专业性，评估结果更 加客观、真实，能够更好地发现数据治理的不足之处，帮助组织优化数据治理。

组织自评可以采用下列流程进行。

(1)定义评估范围。组织可以从某一个特定业务部门或某一个业务领域来启动 数据治理的成熟度评估，随着评估的深入，逐渐扩展到组织的整个数据治理活动。

根据定义的评估范围，确定开展评估所需要的时间，时间太短不能达到预期的目标， 时间太长又会因为没有具体成果而失去目标。

(2)建立评估工作组。评估工作组是实施评估的主体，为了保证评估的客观性 和便于沟通，评估工作应当包括业务和IT 部门的参与者。业务参与者应该包括技术、 市场、财务等业务和职能部门， IT 参与者应该包括数据管理团队、数据治理专家、 流程管理者、安全和隐私专家等。评估工作组的主要工作是建立策略、设计流程、 执行分析、产生报告、结构分析和提供建议等。

(3)定义评估指标。评估指标是评估活动的指向标，指引着评估需要关注的重 点和方向。在建立评估指标的过程中需要考虑组织的人员、流程和数据治理策略等 相关内容。

(4)与利益相关者沟通评估结果并予以发布。在完成云端数据治理成熟度评估 后，需要将评估结果汇报给业务和IT 的利益相关者，这样可以在组织内关键问题建 立公示，进而与管理者讨论后期计划。

(5)总结成熟度评估成果。组织依据发布的评估结果，分析每个评估类别的状 态，发现问题，提出解决和改进的建议，为数据治理后续活动提供优化依据。

第三方评估机构依据云端数据治理模型及治理成熟度要求，对组织开展的数 据治理活动进行全面的评估。针对每一个过程域，分析组织表现的特征是否符合 云端数据治理成熟度的等级要求，如果满足，则判断此过程域措施达到此成熟度 等级。如果评估人员发现组织呈现的证据无法证明其满足某一等级时，组织应当 继续提供有力的证据证明其已经达到该等级水平，否则评估人员可以判定其不符

**第3章=云端数据治理模型与治理成熟度评估模型**

合此等级。

无论是组织自评还是第三方独立评估，都可以采用定性评估的方式，给每一个 过程域措施依据成熟度等级要求打分，最后综合所有过程评分结果，给出组织云端 数据治理成熟度等级，这个综合成熟度等级采取“木桶定律”,即取各个过程域的最 低成熟度等级。表3-2为云端数据治理成熟度等级评分表、图3-3为云端数据治理成 熟度柱状图和图3-4 为云端数据治理成熟度雷达图，这是评估结果的三种不同呈现

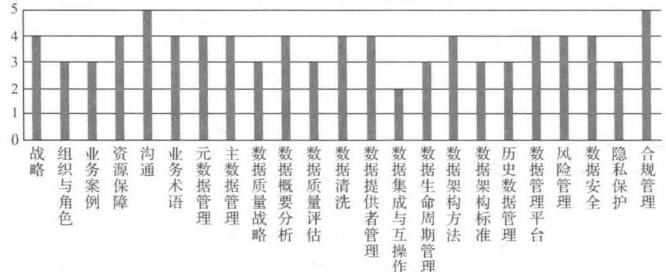
*方式。*

**表3-2** **云端数据治理成熟度等级评分表**

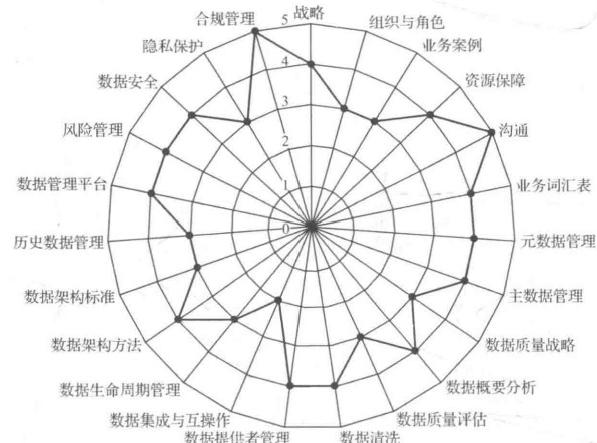
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 序 号 | 职 能 域 | 成熟度等级 | 过 程 域 | 成熟度等级 |
| 1 | 数据战略 | 3 | 战略 | 4 |
| 2 | 组织与角色 | 3 |
| 3 | 业务案例 | 3 |
| 4 | 资源保障 | 4 |
| 5 | 沟通 | 5 |
| 6 | 数据管理 | 4 | 业务术语 | 4 |
| 7 | 元数据管理 | 4 |
| 8 | 主数据管理 | 4 |
| 9 | 数据质量 | 3 | 数据质量战略 | 3 |
| 10 | 数据概要分析 | 4 |
| 11 | 数据质量评估 | 3 |
| 12 | 数据清洗 | 4 |
| 13 | 数据操作 | 2 | 数据提供者管理 | 4 |
| 14 | 数据集成与互操作 | 2 |
| 15 | 数据生命周期管理 | 3 |
| 16 | 数据架构 | 3 | 数据架构方法 | 4 |
| 17 | 数据架构标准 | 3 |
| 18 | 历史数据管理 | 3 |
| 19 | 数据管理平台 | 4 |
| 20 | 安全与隐私 | 3 | 风险管理 | 4 |
| 21 | 数据安全 | 4 |
| 22 | 隐私保护 | 3 |
| 23 | 合规管理 | 5 |

**■云端数据治理**

云端数据治理能力成熟度级别



**图3-3** **云端数据治理成熟度柱状图**



**图3-4** **云端数据治理成熟度雷达图**





**云端数据治理成熟度模型解读**

**第** **4** **章** **云端数据战略**

**4.1** **概述**

**4.1.1** **目的和意义**

俗话说：“三年发展靠机遇，十年发展靠战略。”明晰的战略对于一个组织的发 展至关重要，它是航标、是指南针，能够指引组织在正确的发展轨道上前进。战略 就是一种从全局考虑并谋划实现全局目标的规划，是一切活动的指导性纲领。战略 可以是一个从组织层面考虑的全局战略，也可以是针对某项活动制订的战略。大的 战略包括组织的竞争战略、营销战略、发展战略、品牌战略、融资战略、技术开发 战略、人才开发战略、资源开发战略等，而具体到某项活动，可以包括数据治理战 略、信息安全战略等。 一般来说，战略往往包括战略目的、战略方针和战略支持要 素等。

**例子1:阿里巴巴转型数据服务**

一直以来，公众对阿里巴巴的印象一直是一个大型的电商平台，而作为掌门人 的马云却称阿里巴巴未来是数据公司，而不是一个电商平台。马云为什么会如此定 位阿里巴巴，是基于他对数据的重视。未来阿里巴巴集团的三大战略分别是：全球 化、农村电商和大数据。

**例子2:美国一信用卡公司利用数据分析建立客户流失预警机制**

美国的一家信用卡公司，在对客户流失问题进行分析的时候，发现借助一些变 量指标对客户进行分类，不同项的指标分组下客户的流失倾向是不同的。比如同时 满足 A1 、B1 、X1条件的用户流失概率为66.7%,而满足A1 、B3 的客户流失率仅 为5%,从而建立起了客户流失预警的内部机制，对新客户开发的资源分配也起到了 重要的指导作用。

在云计算和大数据时代，组织的数据面临着更加复杂的环境，数据来源、传输、 存储、处理等环节面对的挑战更加严峻。并且随着数据对组织资产的贡献程度的不 断提升，开展数据治理是组织管理数据的必然选择。

数据治理战略是组织对数据治理活动的顺利开展制订的具有全局指导性的规 划，对数据治理目标的实现和过程管理具有重要的指导意义。

(1)数据治理战略是组织对数据管理成功与否的关键性因素；

**第4章■云端数据战略■**

(2)数据治理战略是保持数据管理融入组织其他战略的重要途径。 云端数据治理战略主要内容包括：

(1)为数据治理活动制订一个愿景， 一个长远目标；

(2)制订数据治理方针；

(3)制订数据治理指导原则；

(4)制订数据治理业务案例；

(5)制订组织架构并分配人员角色；

(6)规划数据治理保障资源；

(7)数据治理实施路线图等。

**4.1.2** **内容与关联性**

云端数据治理战略职能域包括五大过程域：战略、组织与角色、业务案例、资 源保障和沟通。战略过程域是对数据治理活动的顶层设计和远程规划，其目的是在 数据治理活动得到数据治理相关者的理解和认同的前提下，对未来一段时间内数据 治理活动进行规划。组织与角色过程域定义组织开展数据治理的管理团队和实施团 队等，并分配职责与权限以支撑数据治理活动的开展。业务案例过程域为组织实施 数据治理活动的子活动的优先级进行排序，并希望活动能在低成本条件下尽可能实 现输出成果。资源保障过程域是对组织实施数据治理活动所需要的各类资源提供能 力进行规划，包括人力和物力资源等。沟通过程域规划了数据治理活动中与数据治 理相关的策略、规范、规章、通知等文件和活动，使其能够及时顺畅地传达到数据 治理相关者。

战略过程域是数据治理整个活动的规划，涉及到组织与角色、业务案例、资源 保障、沟通过程域的顶层要求。组织与角色、业务案例和资源保障过程域是战略过 程域实施的组织与人员保障、业务指导、资金和其他资源保障，沟通过程域是保障 战略有效传达和被理解的基础。这五大过程域共同组成了云端数据战略职能域。

**4.2** **活动及要求**

**4.2.1** **战略**

**1.** **简介**

1)目的

为数据治理活动制订顶层规划，包括定义数据治理方案的愿景、长期目标与短

 **■云端数据治理**

期目标，确保所有的数据利益相关者就优先事项和计划的执行和管理达成一致。 2)介绍

战略是组织开展数据治理决策的集合，这些决策共同构成一个高层次的行动方 案，目的是指导组织实现数据治理的高层次目标。

数据治理战略的制订一般由数据治理委员会提出，并在数据治理高层管理者的 指导和授权下由数据治理委员会制订。为了获得数据利益相关者支持和认可，数据 治理战略必须能够准确地反映数据提供者和数据消费者的业务期望，将数据价值链 上的所有相关参与者的关切考虑在内，从而让利益相关者相信，数据治理的方案是 有价值的。

通常来说，制订数据治理战略时应至少考虑以下方面。

(1)使命和愿景，包括核心运作原则，治理活动的目标和目的；

(2)项目优先级，如依据项目对组织的重要程度来确定，衡量因素可能包括组 织对项目的依赖程度、组织商业价值的体现、与组织战略提倡活动的一致性等；

(3)活动范围，包括哪些重要的业务领域、数据治理的重点(如数据质量等), 还有关键数据集(如客户主数据);

(4)商业效益，依经济指标对活动价值进行预评价；

(5)数据治理架构，架构设计方法与标准以及架构的使用；

(6)组织、角色与责任，关键利益相关者名单；

(7)治理的需求和范围；

(8)合规要求；

(9)成功措施与度量项；

(10)衡量进度的基准工具；

(11)数据治理实施路线图；

(12)治理风险；

(13)数据安全与隐私保护。

数据治理战略不应该是一个一成不变的方案，应该是随着组织的需求以及外部 环境的变化而变化的。在制订或调整数据治理战略时，还应考虑数据治理执行方面 的重大注意事项，保证治理战略具有可实施性。如在制订战略时，要明确支撑战略 实施的资源要求，以及用于评价计划有效性的准则；在执行过程中，要对实施措施 进行监督，从而依据计划目标评估数据治理活动的进度。

3)相关过程域

组织与角色过程域提供了数据治理活动实施的主体——人员与角色，确保数据 治理战略内容的实施、评估和改进；

业务案例过程域提供了数据治理项目的优先级，有助于数据治理战略的有序推进；

资源保障过程域提供了数据治理活动实施所需的资源，包括资金、场地、基础 设施等；

**第4章■云端数据战略■**

沟通过程域有助于数据治理活动通知的有效传达(如策略、标准、过程及进度 的制订),从而使相关活动获得数据治理参与者的理解。

**2.GOM 表**

|  |  |  |  |
| --- | --- | --- | --- |
| 目 标 | 问 题 | 度 量 项 | 度量项说明 |
| G1:制订完整 的云端数据治理 战略 | Q1:数据治理战略是否 完整 | M1:数据治理战略完整性 | 数据治理战略覆盖了愿景、 目的、范围、组织与角色、需 求、合规、项目优先级等内容， 包括了管理和技术上的策略  以支持业务目标 |
| M2:文件化的数据治理战略 | 数据治理战略文件化，也便 于宣贯及指导数据治理实施 |
| Q2:数据治理战略是否 得到数据利益相关者的认 可 | M3:审批记录 | 制订的数据治理战略应当得 到数据利益相关者的认可，组 织应当保留评审及审批记录 |
| Q3:数据治理战略与组 织其他战略的一致性 | M4:战略一致性 | 数据治理的目标应与组织业 务目标相符 |
| Q4:数据治理战略是否 能够指导数据治理活动的 开展 | M5:可操作性 | 数据战略的制订不能仅从理 论上的最优化出发，而应考虑 到组织的实际情况，包括制度、 人员能力、资金、设施等因素， 保证战略的可操作性，能够实 际指导数据治理活动的开展 |
| G2:云端数据 治理战略得到贯 彻执行 | Q5:组织是否具备必要 的人员 | M6:数据治理组织 | 数据治理组织是治理人员及 其相互关系的集合，是数据治 理实施的主体 |
| M 7 : 人 员 | 配备了数据治理人员并分配 了相适应的角色与职责 |
| M8:人员能力 | 开展了人员培训和持续教育 项目，让数据相关者理解、参 与并遵从数据治理战略 |
| Q6:组织是否具备必要 的资源 | M9;资金、场地、设施 | 为数据治理活动的开展提供 足够的资金、场地、设施等资 源 |
| G3:云端数据 治理战略得到评 审与改进 | Q7:是否制订了对战略 进行测量评审的指标体系 | M10:指标体系 | 定义了数据治理活动的输出 衡量指标 |
| Q8:是否实施了战略的 评审与改进 | M11:统计学方法 | 根据组织已定义的标准过  程，对组织的数据管理战略进 行记录、维护、评审和沟通 |
| M12:改进计划 | 对数据治理战略执行的顺序 进行监督并更新 |
| M13:优先级执行情况 | 对数据治理战略执行的顺序 进行监督并更新 |

 **■云端数据治理**

**3.** **成熟度等级介绍**

**Level 1: 初始级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 1.1 | 数据治理的目标、优先级和范围能够反映 利益相关者的业务目标 | ●数据困境与需求分析  ●文件化的业务目标  ●数据治理目标、优先级和项目范围  ●数据治理结果报告与项目目标差距分析 |

【成熟度等级说明】

组织内部意识到了数据治理的必要性，并进行了初步的数据治理活动的实施， 如制订了数据治理的基本目标、治理项目的优先级和治理的范围。具体来说，业务 发起人、业务代表和项目组首先确定组织数据集的目标和项目优先级，可能包括如 何确定数据范围和要求，如何与组织业务过程相关联，例如创建、读取、更新和删 除数据过程，以及采用数据提供者和数据消费者管理的标准，开始进行数据质量检 查等。

**Level 2: 管理级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.1 | 数据治理目标、优先级和范围得到定义并 获得批准 | ●文件化的业务目标  ●数据治理目标、优先级和项目范围  ●数据治理战略评审纪要 |

【成熟度等级说明】

组织定义了数据治理范围内与业务单元或跨业务部门相关的主题域，且这些主 题域得到了所有利益相关者的批准。当多个业务单元有需求开展数据治理时，组织 整体的数据治理活动应该开展，组织有必要对主题域的优先级进行排序。另外，与 业务过程相关的外部数据及合规要求等应该在数据治理范围内。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.2 | 数据治理的目标和优先权与组织业务目标 相匹配 | ●文件化的业务目标  ●数据治理目标、优先级和项目范围  ●组织的业务目标 |

【成熟度等级说明】

数据治理的目标不是越大越好，也不是越具体越好。数据治理范围内的主题域 应当映射业务单元的目标、功能和过程。要充分调研业务单元的利益相关者的诉求， 并以此作为输入之一，对数据治理的优先权和目标进行评审和修改。

**第4章“云端数据战略■**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.3 | 应当建立并遵循对跨业务单元的项目优先 级从数据的角度进行排序的规则，以及业务 目标的可溯源性 | ●数据治理优先级排序规则 |

【成熟度等级说明】

基于业务单元的重要因素，如商业价值、紧迫性、工作量水平及过程对数据的 依赖程度，对项目优先级进行排序，并在数据治理活动中执行。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.4 | 利用度量指标来评估数据治理的成果 | ●数据治理的目标和相应度量指标  ●数据治理范围的定义  ●映射到创建、更新和删除数据功能的主题域  ●数据治理优先权授权列表  ●映射到业务目标的数据治理优先权  ●项目优先级列表  ●能力实现的顺序规划 |

【成熟度等级说明】

组织应当制订一些针对数据治理活动输出的评价指标，可以包括动态指标和静 态指标，也可以包括经济指标、技术指标、社会指标等。

**Level 3: 定义级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.1 | 在整个组织内建立了数据治理战略，并得 到批准、传达和维持 | ●数据治理战略  ●数据治理战略的批准记录  ●数据治理战略版本的更新记录 |

【成熟度等级说明】

数据治理战略描述了组织数据治理的长期规划，可以是全新制订的，也可以是 从原有管理方案更新完善的，包括数据治理的目标、范围、角色与职责、框架等。 数据治理战略对组织的关键目标和数据资产的业务优先权，以及数据资产的建立、 构建和完善进行了描述。制订战略的目的是增强对数据作为公司关键资产的认知和 意识，从所有利益相关者和决策者那里获得认可和承诺。

数据治理战略通常至少包括以下内容。

●使命和愿景，包括核心运作原则，治理活动的目标和目的；

●项目优先级，如依据项目对组织的重要程度来确定，衡量因素可能包括组织 对项目的依赖程度、组织商业价值的体现、与组织战略提倡活动的一致性等；

●活动范围，包括重要的业务领域、数据治理的重点(如数据质量等),还有 关键数据集(如客户主数据);

 **■云端数据治理**

●商业效益，依经济指标对活动价值进行预评价；

●数据治理架构，架构设计方法及标准，以及架构的使用；

●组织、角色与责任，关键利益相关者名单；

●治理的需求和范围；

●合规要求；

●成功措施与度量项；

● 衡量进度的基准工具；

●数据治理实施路线图；

●治理风险；

●数据安全与隐私保护。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.2 | 根据业务需求和目标，对组织的数据治理 目标进行评价和优先级排序，使之符合业务 战略 | ●数据治理需求调研报告  ●数据治理目标及优先级列表  ●数据治理项目优先级列表 |

【成熟度等级说明】

组织应当充分调研数据治理的需求，并关联到组织的业务发展方向和目标，衡量 数据作为一种资产对组织业务发展的重要性。依据需求调研结果，制订数据治理的目 标以及优先级排序。 一个比较通用的方法是：组织在制订数据治理战略时，要参考组 织的业务战略目标和目的以及行动方针，以此来推动数据治理战略和架构的升级。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.3 | 业务和技术共同协作，开发组织的数据治 理战略 | ●数据治理战略 |

【成熟度等级说明】

数据治理是一个通过技术和管理手段，实现以数据应用促进组织业务发展的过 程。数据治理战略在目标上要与组织业务目标相一致，这是开展数据治理活动的出 发点和落脚点。数据治理能否顺利实施、实施的效果如何，很大程度上取决于组织 所采取的技术路线。因此，数据治理战略应充分考虑到业务和技术两个角度。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.4 | 对数据治理战略执行的顺序规划进行监督 和更新 | ●数据治理战略优先级排序  ●数据治理战略执行记录  ●数据治理战略的更新记录 |

【成熟度等级说明】

数据治理活动的持续监督是必不可少的，好的开始不代表一定会有好的收益， 随着数据治理活动的开展，组织内部和外部环境会发生变化。 一些变化可能会对数

**第4章=云端数据战略**

据治理活动产生重大影响，如国家新通过了某项数据相关的法律法规，会对数据管 理过程中的多个环节提出法律要求，此时数据治理战略必须调整以满足合规要求。 另一方面，组织架构、人员、资金和技术等方面的变化也会对数据治理活动产生影 响，因此必须对数据治理战略的执行进行监督。监督的依据则是既定的数据治理活 动执行顺序，评估其在规划时间节点上完成的成果。

评审数据治理方案执行的顺序规划应考虑以下内容。

●具有里程碑意义的进度；

●风险识别；

●资源利用；

●与数据治理重点保持一致；

●实现数据治理目标。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.5 | 根据组织已定义的标准过程，对组织的数 据管理战略进行记录、维护、评审和沟通 | ●数据治理战略评审和更新记录 |

【成熟度等级说明】

根据组织与角色划分，数据治理机构代表所有业务利益相关者，负责确保数据 治理战略的有效沟通。当有业务需求或战略需要更改时，数据治理机构要负责对优 先事项、目标和路径图进行评审，并对以上内容进行更新，但与此同时也要遵循已 定义的批准过程。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.6 | 利用度量指标来评估和监督数据管理目标 的实现 | ●数据治理战略  ●数据治理目标和优先权列表  ●数据治理策略  ●利益相关者参与和批准的文件  ●描述数据治理范围的文档(如主题域、业务领域、  关键数据元素、重要规则等)  ●数据治理战略顺序规划  ●数据治理方案度量指标  ●方案成本一效益分析结果  ●数据治理方案评审  ●数据治理战略进度播报表 |

【成熟度等级说明】

组织应当制订一些度量指标来衡量数据治理目标的实现程度，例如数据治理关 键阶段的指标完成情况，以及对成本和效益的评价。组织在制订度量指标时，应充 分考虑指标的可操作性，能够让所有的利益相关者较方便地使用这些度量指标去评

 **■云端数据治理。**

估数据治理目标的实现情况。

**Level** **4:** **量化级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 4.1 | 利用统计学或其他量化手段来评价数据治 理战略目标在实现业务目标方面的有效性， 并基于度量指标进行修改 | ●数据治理战略评审和更新记录 |

【成熟度等级说明】

对数据治理战略目标的实现度量不再仅仅依靠定性评价，而是增加了统计学或 其他专业度量手段来衡量，例如通过建模分析的方式，全面透彻地分析数据治理目 标的达成对组织业务的贡献。

对数据治理战略和目标进行修改，通常包括以下两个主要方面。

●为了更有利于业务目标的实现，完善数据治理的目标、优先权以及范围；

●调整统计学评估方法或数据收集方法。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 4.2 | 组织研究创新性的业务过程和新的法规要 求，确保数据治理方案与未来的业务需求相 匹配 | ●基于度量指标开展的数据治理管理方案评估报告  ●监督新兴行业或法规新要求的计划和文档  ●数据治理策略 |

【成熟度等级说明】

总结数据治理经验，创新数据治理过程，识别和预判未来法规新要求，做到未 雨绸缪，尽早调整治理战略，确保数据治理方案与未来的业务需求相匹配。

**Level 5: 优化级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 5.1 | 组织搜集数据治理行业最佳实践方案，并 深入分析与组织的适用性，然后作为制订战 略和目标的重要依据 | ●数据治理最佳实践集  ●数据治理战略 |

【成熟度等级说明】

最佳实践方案是行业从多个维度总结出来的最优异的行动方案，能够指导组织 以最佳的方式实现最大的目标。组织在进行数据治理时，应当主动收集行业最佳实 践方案，深入分析最佳实践方案的内涵，总结成功经验，并作为制订数据治理战略 的主要输入。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 5.2 | 总结组织数据治理经验，为制订和执行数 据治理战略的行业最佳实践方案做贡献 | ●行业最佳实践方案的外部出版物和演示文档  ●最佳实践的对比分析报告 |

**第4章■云端数据战略**

【成熟度等级说明】

组织在数据治理活动上取得重要成功时，也应该积极分享实践成功经验，为行 业最佳实践做贡献。

**4.** **实施案例**

(1)文件名称列表

《云端数据治理章程》

(2)实施案例

下表展示了一个云端数据治理战略章程，仅供参考。

|  |
| --- |
| 云端数据治理章程  1 数据治理的定义  数据治理是对组织内的人员、流程、技术和策略进行编排，以企业资产的形式利用、优化和最大化数据的过程。 就像我们的董事会治理企业的过程中最大化股东价值一样，数据治理计划旨在最大化数据对于关键IT部门和业  务部门利益相关者的价值。  2 业 务 目 标  组织拥有一个稳定的SAP环境，但不是我们的所有数据都位于SAP中。数据治理的范围应该不断扩大，只要 该计划能够成功实施。但是，在接下来的12个月内，数据治理计划将重点关注对以下业务目标的支持。  (1)客户中心性；  (2)供应商管理；  (3)供应链优化；  (4)财务数据质量；  (5)电子文档和电子邮件生命周期的管理。  3 高层支持者  首席信息官将充当计划的总体高层支持者，与业务经理和关键职能区域紧密合作。  4 组 织  数据治理计划的中心角色将是数据治理主管。  数据治理委员会将对计划拥有最高监督职责。委员会将由首席信息官主持，将包含营销部门高级副总裁、供应  链部门高级副总裁、首席财务官和总顾问。委员会将每月举行一次会议，至少在最初是这样。数据治理主管将负 责连同首席信息官和其他成员一起设置数据治理委员会的议程。  数据治理工作组将由数据治理主管主持。工作组将包含来自营销、供应链、财务、法律、数据架构和内容管理 部门的成员。这些成员将对各自职能区域内的数据相关问题进行日常处理。数据治理工作组将每周举行一次会议，  至少在最初是这样。  5 度 量 指 标  数据治理工作组负责建立关键度量指标计分卡来监控数据治理计划的性能。数据治理主管负责定期向数据治理 委员会报告这些度量指标。  6数据管理员  数据管理员拥有对其领域内的数据的保管责任。关键数据领域的高层支持者如下所示。  (1)客户数据：营销部门高级副总裁：  (2)供应商和材料数据：供应链部门高级副总裁；  (3)财务数据：首席财务官。  高层支持者将任命数据管理员来负责每天的数据质量。数据治理主管将负责监督，确保数据管理计划在整个组 织内一致地实施。数据管理员将负责每月定义、收集和报告与他们各自的数据领域相关的关键度量指标。 |

**云端数据治理**

**4.2.2** **组织与角色**

**1.** **简介**

1)目的

为数据治理活动的实施提供组织架构和人员保障，通过组织架构设计去督促员 工完成组织的战略和目标。

2)介绍

云端数据治理组织是组织在云计算和大数据环境下，为实现大数据有效管理和 提高服务能力的目标，按照一定的结构形式、活动规律结合起来的，期望完成数据 治理活动的人员的集合。

在云计算和大数据环境下，数据治理战略通过授权、决策权和控制权影响组织 架构，其中授权和决策权影响组织架构的形式，而控制权是通过组织架构设计来督 促员工完成组织的战略和目标。组织一般由四大要素组成，即人、目标、结构和管 理。在数据治理组织中，人可以包括数据治理委员会、大数据利益相关者和大数据 专家等；目标就是数据治理战略中制订的想要实现的数据治理目标和愿景；结构则 描述了组织内部的相互关联，具体到部门、岗位、职责和从属关系等；管理则是组 织要持续运转的保障条件。

云端数据治理组织的设立应该随着组织规模、性质、业务目标等情况不同而不 同，但一般包括以下治理活动。

(1)根据业务情况，绘制关键流程图并建立职责分配模型 (RACI), 以识别云 端数据治理中的利益相关者；

(2)扩展传统数据治理或IT 治理的章程，明确云端数据治理的重要性；

(3)扩展传统数据治理委员会或IT 治理委员会的成员角色和职责，增加云端数 据治理的职责和角色，如将云端数据利益相关者和大数据专家纳入进去。

职责分配模型可以用来明确组织业务过程中的各个角色及其相关责任，将所有 参与者分成四大类，分别赋予相应的职责。结合云端数据治理项目，职责分配模型 的角色可以表述如下[12]。

●应负责方(Responsible,谁负责),指授权管理某属性的人。即负责执行任务 的角色，具体负责操控项目、解决问题，例如数据管理员。

●最终负责方 (Accountable,谁批准),指对数据属性承担最终责任的人。即 对任务负全责的角色，只有经其同意或签署之后，项目才能得以进行，如数 据治理首席执行官。

●咨询方 (Consulted, 咨询谁),指通过双向沟通接收咨询的某人或某些人。 即在任务实施前或实施过程中提供指定性意见的人员，如数学科学家。

**第4章■云端数据战略■**

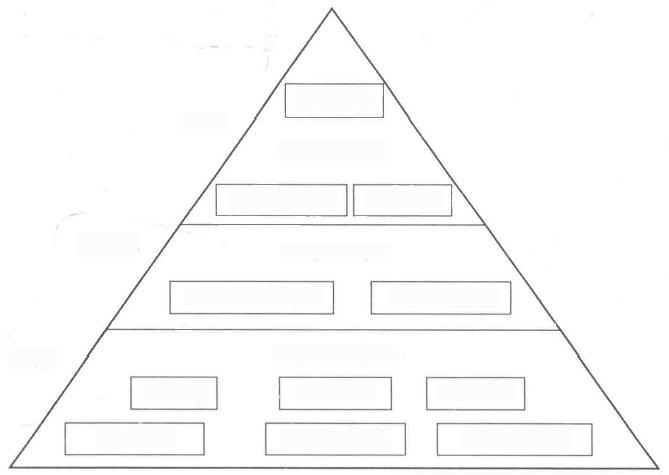
●被告知方 (Informed, 告知谁),指通过单项沟通被告知的某人或某些人。即 被通知结果的人员，不必向其咨询或征求意见。

数据治理是高层次的、规划性的数据管理活动，需要众多角色的协调配合才能 得以完成，按照管理职责权限的大小，数据治理人员大致可以分为三个等级，即数 据治理委员会、数据治理工作组和基层数据治理专员。

数据治理委员会通常是组织中的高级管理人员，也就是数据最终负责方，将由 首席信息安全官、首席信息官等组成。此外，数据治理委员会也将包含来自财务、 法律、HR 团队的职能部门代表，以及来自各种将数据视为组织资产的业务线的代表。 数据治理委员会定义数据治理愿景和目标，在组织内跨业务和IT 进行协调，设置数 据治理计划的总体方向，在发生策略分歧时进行协调。

数据治理工作组负责数据治理活动中的沟通环节，向数据治理委员会汇报并与 基层数据治理专员沟通，承担协调角色，对应职责分配模型中的应负责方。数据治 理工作组由数据治理领导组成，这里的领导也可能在数据架构、信息安全或风险小 组内拥有其他角色。

基层数据治理专员负责数据治理某项具体工作的实施，具有最低的管理和决策 权限，向数据治理工作组汇报。 一个相对完整的数据治理组织架构如图4-1 所示。



首席数据官

组织级

数据治理委员会

首席信息安全官 首席信息官

跨部门级 数据治理工作组

部门数据治理委员会 项目指导委员会

部门级

数据主管

数据分析师

基层数据治理专员

数据架构师

数据仓库管理员

数据安全管理员

数据收集员

图4-1 数据治理组织架构

数据治理团队的构建可以从两个方面着手， 一是确定既有角色能否承担云端数 据治理的责任，二是直接任命新的数据角色。例如，需要时可以任命新的数据主管， 组织根据数据规模或分类，酌情任命一定数量的数据主管。另外，组织还可以聘请

**■云端数据治理**

知识和经验丰富的数据科学家，能够及时咨询相关问题。 一些传统组织在进行信息 治理时，已经具备了很多信息治理角色，例如在金融行业，许多组织已经任命了首 席数据官，在进行数据治理时，只需要将数据治理相应角色的任务纳入其职责范围。 信息治理委员会由信息治理计划的高级管理人员组成，负责制订信息治理计划的总 体方向。在组建数据治理委员会时可以借鉴信息治理委员会，将数据治理对应的职 责分配给信息治理委员会相对应的角色。如首席信息安全官应当承担起数据安全官 的职责，首席信息官则应当将首席数据官的职责纳入麾下。

3)相关过程域

●战略过程域为数据治理活动制订顶层规划，提出了实现数据治理目标所需的 组织结构和人员角色；

●沟通过程域有助于数据治理活动通知的有效传达(如策略、标准、过程、进 度的制订),从而使相关活动获得数据治理参与者的理解。

**2.GOM 表**

|  |  |  |  |
| --- | --- | --- | --- |
| 目 标 | 问 题 | 度 量 项 | 度量项说明 |
| G1:建立云端 数据治理组织 | Q1;组建的数据治理组 织是否能够支撑数据治理 目标和愿景的实现 | M1:数据治理组织完整性 | 成立了专业的数据治理组  织，包括数据治理委员会、数 据治理工作组和数据治理专员 |
| Q2:数据治理组织与角 色是否符合数据治理战略 的要求 | M2:数据治理组织与战略 符合度 | 数据治理组织与角色明确了 战略赋予自身的目标和义务 |
| Q3:组织与角色是否得 到数据利益相关者的认可 | M3:评审记录 | 数据治理组织得到了数据 利益相关者的理解和支持，具 有评审记录 |
| Q4:数据治理组织与角 色能否指导数据治理活动 开展所有工作 | M4:组织完整性与角色能 力 | 数据治理组织覆盖了数据  治理所有活动环节，数据治理 角色具备相符合的职责与权  限，具备相应的能力 |
| G2:配备数据 治理角色，并分 配相应的职责 | Q5:数据治理角色是否 配备了合适的职责 | M5:职责与权限 | 数据治理角色得到了充分 定义被分配了相应的职责 |
| Q6:数据治理组织内的 角色是否具有代表性 | M6:代表性 | 数据治理委员会和数据治理 工作组成员具有广泛代表性 |
| Q7:数据治理角色是否 安全 | M7:资格认证 | 数据治理角色能力得到考 核和确认 |
| M8:能力提升计划 | 开展了人员培训和持续教  育项目，不断提升数据治理角 色能力 |
| M9:持续监管计划 | 对数据治理角色进行安全 检查和持续管理 |

**第4章■云端数据战略**

**3.** **成熟度等级介绍**

**Level 1: 初始级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 1.1 | 数据治理由部分员工根据工作需要开展， 管理层对实施云端数据治理具有基本的意识 | ●组织角色分配 |

【成熟度等级说明】

出于工作需要，部分员工对数据开展了基本管理，但并不全面。组织没有针对 数据治理建立专门的数据治理委员会，人员也不是固定或全职从事这项活动的，管 理层对数据治理的重要性和必要性认识不足。

**Level 2: 管理级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.1 | 组织具备开展云端数据治理的基本人员条 件 | ●人员配置管理办法  ●人员培训计划和培训记录  ●人员考核与能力评估记录 |

【成熟度等级说明】

组织具备一些从事与数据相关的人员，能够开展针对各自工作需要的数据管理 并实施人员配置管理措施，但对跨部门或跨业务的数据管理缺乏必要的沟通和统一 的认识。关键岗位有人员储备，并定义此类人员变更的应对方法。组织能够识别数 据治理所需要的基础知识、专业知识和综合知识，并采取措施使数据治理主要人员 能够具备这些知识，例如对相关人员开展能力提升培训和管理培训，并定期对人员 进行绩效考核。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.2 | 高层管理者意识到数据治理的存在 | ●会议纪要 |

【成熟度等级说明】

组织基于业务或其他原因，被动地实施了数据治理项目。高层管理者通过职务 上的管理职责意识到数据治理的存在，但对数据治理项目没有充分的认识，不会主 动积极推进数据治理工作的开展。

**Level 3: 定义级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.1 | 组建了数据治理负责团队，明确了数据治 理相关岗位的职责以及任职资格要求 | ●组织章程  ●数据治理组织架构  ●数据治理角色与职责分配 |

 **■云端数据治理■**

【成熟度等级说明】

组织组建了包括数据治理委员会、数据治理工作组和基层数据治理专员的三层 数据治理团队，这些角色可以由新任命的人员来充当，也可以来源于既有的相关岗 位人员(作为补充职责)。如首席数据官可能是原有信息治理委员会的首席信息官， 不过其职责已经新增了数据治理的相关权限和义务。组建的三层数据治理团队，每 个岗位的职责都被明确界定，并给出了任职的要求。

下表给出了一个组织数据治理角色与职责分配的简要范例。

**表4-1** **数据治理角色与职责分配**

|  |  |  |
| --- | --- | --- |
| 组 织 架 构 | 角 色 | 主 要 职 责 |
| 数据治理委员会 | 首席信息官 | (1)挖掘组织信息资源、制订组织信息化战略、合理布局组织信息 化、评估信息化对组织的价值；  (2)统筹数据治理全面工作，协调数据治理实施；  (3)评估数据治理输出，调整数据治理战略等。 |
| 首席信息安全官 | (1)指派和领导IT安全专家团队；  (2)监督开发和维护组织信息安全政策、标准和程序；  (3)建立一个IT安全风险管理体系；  (4)开发安全策略、处理安全事故和协调调查活动；  (5)将数据作为组织的一项重要资产，制订并实施数据安全保护策 略等。 |
| \* | |
| 数据治理工作组 | 信息管理总监 | (1)参与制订组织信息管理战略；  (2)主持实施IT基础设施建设和升级改造；  (3)具体负责数据治理实施工作；  (4)协调部门间数据治理沟通等。 |
| 合规总监 | (1)负责组织活动涉及的法规及规范的收集、宣贯，并负责组织运  营合规性的监督检查；  (2)数据安全与隐私保护相关法律、法规、标准等识别与贯彻等。 |
| 信息安全总监 | (1)参与制订组织信息安全策略  (2)具体负责组织信息安全监管；  (3)具体负责组织数据安全保护等。 |
| 数据科学家 | (1)参与数据治理活动；  (2)识别组织数据治理问题并提供解决思路；  (3)为数据治理活动提供咨询建议和意见等。 |
|  | |
| 基层数据治理专员 | 生产数据主管 | (1)负责组织生产和经营性数据的收集和管理；  (2)制订数据质量的业务规则；  (3)定义、捕获并保留业务元数据等。 |

**第4章■云端数据战略■**

续表

|  |  |  |
| --- | --- | --- |
| 组 织 架 构 | 角 色 | 主 要 职 责 |
| 基层数据治理专员 | 财务数据主管 | (1)管理组织日常财务信息；  (2)将财务数据映射到业务数据；  (3)制订数据治理等财务目标与评估指标等。 |
| 数据安全管理员 | (1)制订数据安全管理规章；  (2)识别组织敏感数据，并在元数据库中作标记；  (3)制订历史数据归档、销毁策略，满足合规和业务需求等。 |
|  | |

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.2 | 能够识别数据治理所需要的基础知识、专 业知识和综合知识，并采取措施使数据治理  主要人员能够具备这些知识 | ●培训课件  ●培训记录 |

【成熟度等级说明】

数据治理涉及管理和技术两大类知识，尤其在技术方面，需要相关人员熟悉软 件设计、数据库技术、统计学、信息学等学科知识。组织在任命角色和分配职责时， 一方面要识别每个岗位所需的基础知识、专业知识和综合知识；另一方面，组织要 对于拟任命的角色能力进行评估，并根据需要，采取措施使相关角色具备从事数据 治理所需的知识技能。可采取的措施主要包括开展培训、咨询、自学等途径。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.3 | 组建的数据治理团队与角色分配得到了数 据利益相关者的理解和认同 | ●培训课件  ●培训记录 |

【成熟度等级说明】

数据治理的主要目的是实现数据价值的提升，数据治理团队和角色直接决定了 数据治理的方向和重点，对数据治理目标能否实现起到了决定性的作用，因此数据 治理团队必须征求数据利益相关者的理解和认同。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.4 | 数据治理委员会和数据治理工作组成员具 有广泛代表性 | ●数据治理委员会成员名单  ●数据治理工作组成员名单 |

【成熟度等级说明】

数据治理委员会负责顶层设计和监督数据治理计划，设置数据治理计划的总体 方向，在发生策略分歧时进行协调， 一般由高层组成。为了保障数据治理的顺利实 施，数据治理委员会应当包括管理、财务、运营、供应链等负责人。同样的，数据 治理工作组成员也应具有广泛代表性。

 **■云端数据治理■**

**Level** **4:量化级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 4.1 | 制订培训计划及管理制度，持续提升数据 治理人员知识技能 | ●培训计划  ●培训管理制度  ●培训记录文件  ●培训教材/课件 |

【成熟度等级说明】

组织应识别数据治理各岗位、各级别的培训要求，制订培训计划，培训内容包 括管理、技术、操作及综合知识等。应制订培训管理制度、保存培训记录文件，如 培训需求、培训计划、培训课程体系及教材/课件、培训实施记录、培训效果评价等。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 4.2 | 为数据治理组织制订绩效管理办法，建立 考核体系 | ●绩效管理办法  ●管理手册 |

【成熟度等级说明】

制订绩效管理办法，建立考核体系，考核体系应与数据治理组织和业务目标相 匹配；根据不同岗位制订具体的、差异化的考核指标。

**Level 5: 优化级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 5.1 | 建立培训讲师库和培训教材库 | ●培训讲师人才库  ●培训教材  ●培训记录 |

【成熟度等级说明】

为持续提升数据治理人员职业技能建立培训讲师库和培训教材库，培训教材覆 盖了管理、工具、技术、过程等内容。并对培训效果进行了总结分析和改进，分析 至少包括培训目标实现率、培训覆盖率、培训有效率等。

**4.** **实施案例**

(1)文件名称列表

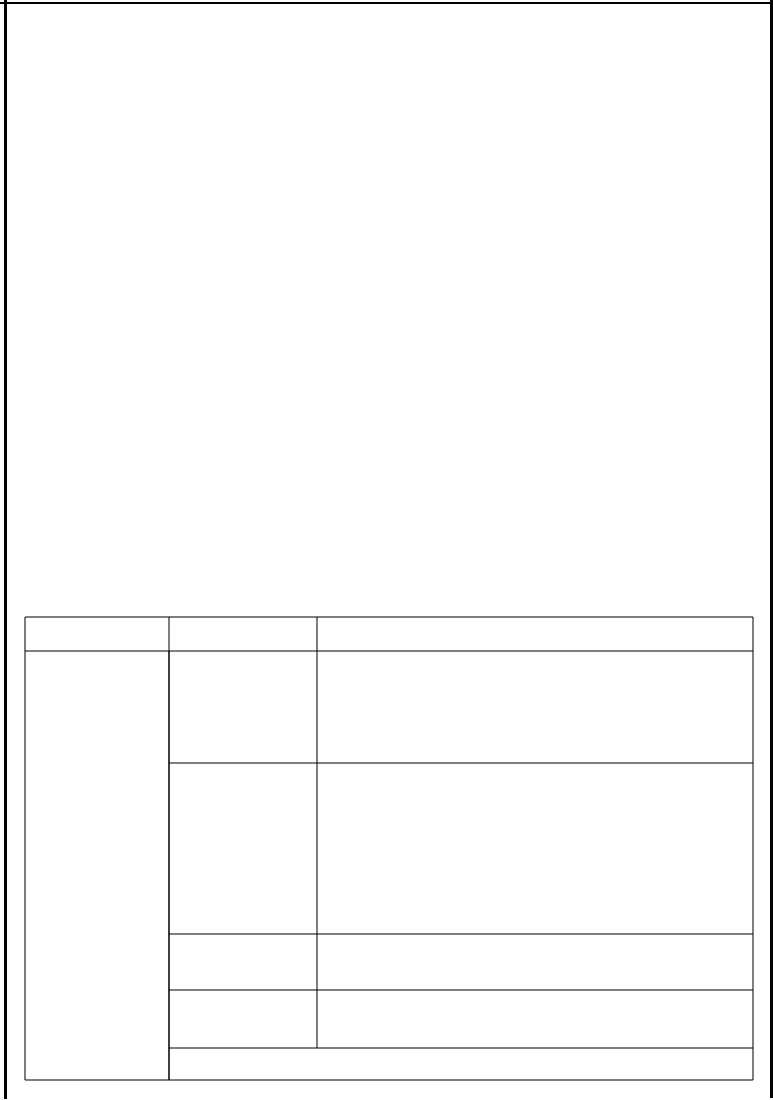
《基于角色分配的云端数据治理组织与角色职责》

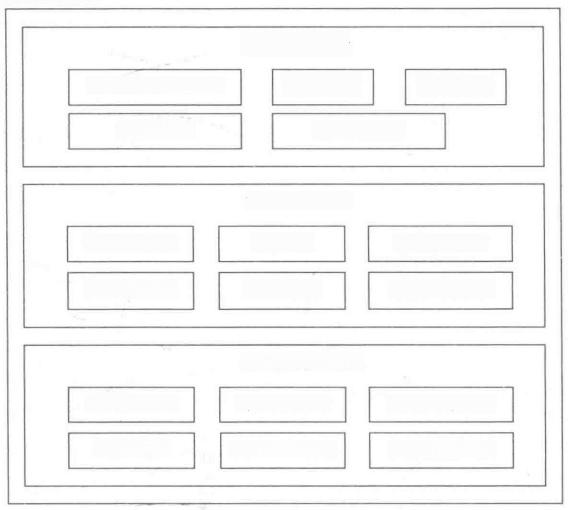
《XX 公司人员管理办法》

(2)实施案例

下表展示了基于角色分配的云端数据治理组织与角色职责。

**第4章■云端数据战略**

基于角色分配的云端数据治理组织与角色职责

1 云端数据治理组织架构

数据治理委员会

首席财务官 首席安全官

首席信息(数据)官

供应链副总

相关业务副总

数据治理工作组

信息安全总监

信息管理总监

供应链负责人

合规总监

数据科学家

首席数据治理员

基层数据治理专员

顾客数据管理员

生产数据主管

元数据专家

财务数据主管

数据质量分析师

数据安全管理员

2 角色与职责

主 要 职 责

角 色

组 织 架 构

(1)挖掘组织信息资源、制订组织信息化战略、合理布局组

织信息化、评估信息化对组织的价值

首席信息官

(2)统筹数据治理全面工作，协调数据治理实施

(3)评估数据治理输出，调整数据治理战略等

(1)指派和领导IT安全专家团队

(2)监督开发和维护组织信息安全政策、标准和程序

(3)建立一个IT安全风险管理体系

首席安全官

(4)开发安全策略、处理安全事故和协调调查活动

数据治理委员会

(5)将数据作为组织的一项重要资产，制订并实施数据安全

保护策略等

(1)为数据治理活动提供资金支持计划

首席财务官

(2)组织开展数据治理TCO分析

(1)负责数据提供者管理

供应链副总

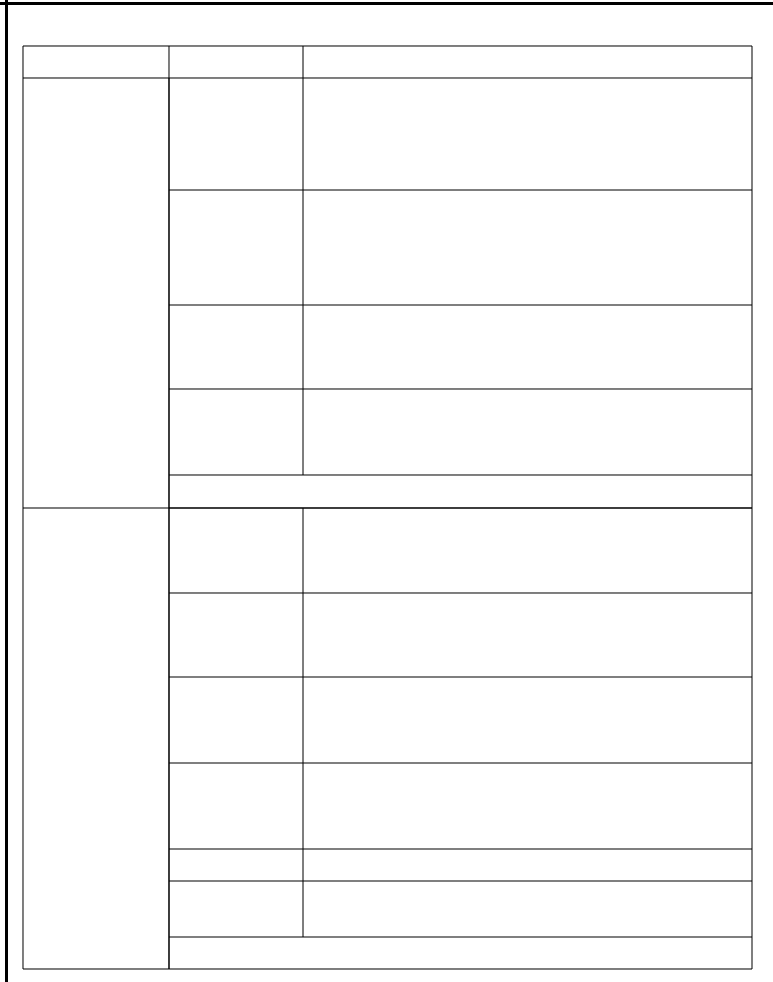
(2)组织制订数据流通规则

■

74

|

**云端数据治理**

续表

组 织 架 构 角 色 主 要 职 责

(1)参与制订组织信息管理战略

(2)主持实施IT基础设施建设和升级改造

信息管理总监

(3)具体负责数据治理实施工作

(4)协调部门间数据治理沟通等

(1)负责组织活动涉及的法规及规范的收集、宣贯，并负责组

织运营合规性的监督检查

合规总监

(2)数据安全与隐私保护相关法律、法规、标准等的识别与贯

数据治理工作组 彻等

(1)参与制订组织信息安全策略

信息安全总监 (2)具体负责组织信息安全监管

(3)具体负责组织数据安全保护等

(1)参与数据治理活动

数据科学家 (2)识别组织数据治理问题并提供解决思路

(3)为数据治理活动提供咨询建议和意见等

…

(1)负责组织生产和经营性数据的收集和管理

|  |  |
| --- | --- |
| 生产数据主管 | (2)制订数据质量的业务规则  (3)定义、捕获并保留业务元数据等  (1)管理组织日常财务信息 |
| 财务数据主管 | (2)将财务数据映射到业务数据 |

(3)制订数据治理等财务目标与评估指标等

(1)制订数据安全管理规章

数据安全管理员 (2)识别组织敏感数据，并在元数据库中标记

基层数据治理专员

(3)制订历史数据归档、销毁策略，满足合规和业务需求等

(1)识别顾客来源

顾客数据管理员 (2)协助制订元数据库

(3)管理顾客数据

元数据专家 (1)负责元数据整合、控制和交付，包括元数据存储库的管理

(1)负责确定数据的适用性

数据质量分析师

(2)依据制订的规则开展数据清洗

… ·

**第4章■云端数据战略■**

**4.2.3** **业务案例**

**1.** **简介**

1)目的

为组织提供一个数据治理项目决策的工具，决策的依据多以数据治理项目未来 的经济收益和对组织的整体收益为主，并确定治理项目的优先级，保证数据治理项 目的持续性。

2)介绍

业务案例是对数据治理方案、倡议或具体项目进行资助的提案，描述了治理的 范围、活动和预期结果，以及由“总体成本VS 投资收益”分析构成的理论阐述。

业务案例是一种决策工具，用来确定某个项目将对组织盈利能力产生的影响。 业务案例应该展示这个决定在一段时间内是如何改变取值现金流的，以及成本和收 入的改变。具体关注点包括组织内部收益率、现金流、投资回收期和投资回报率等。

业务案例可以是整个数据治理方案，或需要持续资助的方案要素，或具体数据 管理项目制订，包括战略上和战术上的内容。例如，对于数据治理总体方案，从战 略层面上，业务案例需要关注如何减少跨功能数据的冗余。就这一点而论，业务案 例需要解决一些组织上面临的挑战，包括组织为什么需要重点关注减少冗余、受冗 余影响的业务功能，以及方案将要解决的现存主要问题等。在战术上，该业务案例 需要定义如何满足战略目标，包括以下内容。

● 识别关键方案的交付物；

●定义顺序规划中的相关项目；

● 定义衡量指标；

● 确定所需的人力资源；

●为成本、收益和风险提供依据。

业务案例需要识别数据治理活动的关键交付物，并为数据治理活动及交付物建 立评估模型，即成本一效益原理。对数据治理总体方案，业务案例通常会对以下内 容进行评估。

●因数据资产管理欠佳而产生的风险(财务风险、操作风险和声誉风险等);

●运作成本(例如数据整合、操作中断和合规要求等引起的无形成本);

●期待通过数据治理投资获得的有形或无形收益；

●与组织战略计划的一致性。

数据治理活动是多个子活动的集合，业务案例则要为这些子活动的实施顺序提 供规划和管理。大多数组织采用了标准业务案例模板，该模板能够用于符合顺序规 划的数据治理业务案例。方案级别业务案例的关键输入是所有阶段中执行和支持数

**■云端数据治理■**

据治理战略所要求的总成本，总成本的确认和批准可以确保数据治理得到充足的资 金来满足其近期和长期目标。

总体拥有成本 (Total Cost of Ownership,TCO), 也指总体生命周期成本，是一 项帮助组织来考核、管理和削减在一定时间范围内与组织获得某项资产相关联的所 有成本的技术。尽管TCO 分析方法经常用于重大开支和重大方案，这些方法并不常 用于数据治理策划，因为由此产生的一些成本通常被分配到单个的项目，给评估总 成本带来了挑战。

TCO 没有普遍的计算公式，与行业性质有很大差别，但都要考虑独立成本之间 的关联。对于数据治理项目的成本计算，可以从有形成本和无形成本两个方面入手， 有形成本如数据搜寻、工具、人员资源、IT 和手动矫正过程等支出；无形成本如支 撑业务过程的数据价值或为组织带来竞争优势的数据价值。无形成本由于缺乏量化 指标，使得其成本难以计算，因此给数据治理TCO 计算带来了很多挑战。不过，TCO 这种思想还是为高级管理层提供了一个用于评价方案目标的实现情况和评估目前资 金是否充足的基线。此外，开发TCO 时，更倾向于利用现有可用的行业标准来协助 管理层参照同行组织进行基准测试。

3)相关过程域

●战略过程域为数据治理活动制订顶层规划，规划了数据治理目标和资源需 求，以及数据治理活动集；

●业务案例过程域为数据治理活动的成本进行估算，保障资金支持；

●资源保障过程域提供了数据治理活动实施所需的资源，包括资金、场地、基 础设施等，与业务案例过程域的TCO 直接关联；

●沟通过程域有助于数据治理活动通知的有效传达(如策略、标准、过程、进 度的制订),从而使相关活动获得数据治理参与者的理解。

**2.GQM 表**

|  |  |  |  |
| --- | --- | --- | --- |
| 目 标 | 问 题 | 度 量 项 | 度量项说明 |
| G1:利益相关 者批准和采纳的 跨业务范围的业 务案例 | Q1:组织如何决定数据 治理方案所要求的投资水 平 | M1:数据治理方案水平 | 为数据治理定义了业务案 例，能够指导组织开展具体的 数据治理活动 |
| Q2:组织如何决定是要 开发一个总体的业务案例 还是多个相互联系的业务 案例 | M2:业务案例范围 | 业务案例重点分析了数据 治理的收益和成本 |
| Q3:如何评价业务案例 | M3:评审记录 | 业务案例得到了数据利益 相关者的理解和支持 |
| Q4:需要谁参与其中， 需要谁批准 | M4:批准记录 | 业务案例得到了数据治理 高层的批准 |

**第4章■云端数据战略**

**续表**

|  |  |  |  |
| --- | --- | --- | --- |
| 目 标 | 问 题 | 度 量 项 | 度量项说明 |
| G2:业务案例 符合数据治理战 略需要 | Q5:业务案例是否反映 数据管理战略的目标和优 先事项 | M5:战略符合性 | 业务案例目标与业务目标 和数据治理目标相符 |
| M6:总体拥有成本 | 对包含了数据治理TCO的 成本要素进行管理和追踪 |
| Q6:业务案例是否反映 数据治理顺序规划 | M7:数据治理项目优先级 | 业务案例定义了数据治理 项目的顺序和优先权 |
| Q7:业务案例制订是否 考虑到资源保障的能力 | M8:业务案例与资源协调 性 | 业务案例的制订考虑到资 源保障的能力 |

**3.** **成熟度等级介绍**

**Level 1: 初始级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 1.1 | 为数据治理开发了一个基本的业务案例 | ●数据治理策略  ●业务案例  ●成本分析记录 |

【成熟度等级说明】

数据治理活动在小范围内开展，部分成员因工作需要，对数据开展了基本的管 理。为此分析了数据管理的成本，并对数据管理子活动的优先级进行了排序。由于 业务案例只是针对部分活动规划的，并不能完全适用于整个组织的数据治理活动。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 1.2 | 记录了数据治理活动的收益和成本，并用 于资金资助决策 | ●项目文档(会议记录、讨论文档)  ●项目级业务案例 |

【成熟度等级说明】

对数据治理活动的成本和收益进行了考量，并以文件记录的形式保存下来，为 数据治理活动所需的资源保障提供决策依据。

**Level 2: 管理级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等级要求 | 实例 |
| 2.1 | 业务案例方法论得到定义和遵循 | ●业务案例制订方法  ●业务案例 |

【成熟度等级说明】

大多数组织采用一些方法为业务单元里的一些关键项目或交叉项目制订了结构 良好的业务案例，通过总结提炼这些方法，可以为其他业务单元制订业务案例提供

 **■云端数据治理**

参考。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.2 | 业务案例为决策者审核数据治理的资助提 供支持 | ●业务案例 |

【成熟度等级说明】

业务案例与资源保障一起给数据治理项目制订了支持优先级和所需的资源，这些 信息将成为领导层对数据治理活动的理解和判断的依据，帮助他们做出科学的决策。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.3 | 为新倡议制订的数据治理业务案例应与业 务目标和数据治理目标相符 | ●业务案例标准方法论  ●数据治理业务案例倡议  ●业务案例批准或拒绝的文档或记录 |

【成熟度等级说明】

随着数据治理项目的进行，需要对某些治理活动进行调整，后者由于需要，开 发新的数据治理活动，在为这些新的数据治理活动制订业务案例时，应当考虑数据 治理活动之间的协调性，尤其是新的数据治理活动要与业务的总体目标和大的数据 治理项目的目标相符合。例如业务术语，新的数据治理倡议可能会产生新的业务术 语，这就要求新的业务案例需要对新项目提供新的或修改过的术语。再如，新的数 据治理项目需要新的资源，在数据治理整体的资源保障计划中，在有限的资源中， 如何协调新项目与原项目所用资源的分配，要求新倡议制订的业务案例要考虑原有 数据治理活动的资源保障。

**Level 3: 定义级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.1 | 根据组织的标准方法论开发数据治理案例 | ●业务案例标准方法论  ●数据治理业务案例倡议 |

【成熟度等级说明】

组织从实践中总结或引进较为成熟的方法论，作为组织开发数据治理业务案例的 指导方法，并在组织范围内得到确定，在后续数据治理案例制订时作为标准的方法。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.2 | 数据治理业务案例得到批准并执行 | ●业务案例批准或拒绝的文档或记录 |

【成熟度等级说明】

数据治理业务案例为组织提供了一个数据治理项目决策的工具，决策的依据多 以数据治理项目未来的经济收益和对组织的整体收益为主，并确定治理项目的优先 级，保证数据治理项目的持续性。因此为保障数据治理的顺利实施，决策者应参与

**第4章■云端数据战略**

到业务案例的开发当中，并审核后批准该业务案例，以在组织内得到贯彻执行。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.3 | 业务案例的制订考虑到资源保障的能力 | ●业务案例 |

【成熟度等级说明】

资源保障是数据治理活动的基本要素，组织必须提供一定的资源来支持数据治 理的实施。业务案例则需要考虑如何将这些资源分配到各个数据治理的子活动中， 协调各子活动的资源占用，在有限的资源内，发挥最大效益。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.4 | 业务案例反映了对数据治理方案的总体拥 有成本的分析，并根据组织的财务会计方法 向组织、方案和项目分配成本要素 | ●业务案例  ●总体拥有成本分析文件与结果 |

【成熟度等级说明】

总体拥有成本是一项帮助组织来考核、管理和削减在一定时间范围内与组织获 得某项资产相关联的所有成本的技术。业务案例与资源保障密不可分，制订的业务 案例必须对数据治理的各种经济成本加以考量，考量的方法应当与组织的财务会计 方法保持大体一致。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.5 | 在整个数据治理生命周期，对包含了数据 治理TCO的成本要素进行管理和追踪 | ●业务案例  ●总体拥有成本分析文件与结果 |

【成熟度等级说明】

数据治理生命周期包括数据治理项目规划、启动、实施和优化等多个环节，后 一环节是前一环节的顺承，后者受前者较大影响，且每个环节的工作重点不一样， 对成本需求也不一样，因此在整个数据治理生命周期，要不断地对成本进行跟踪和 管理。收集成本的数据可能来自整个组织，收集的方法可以利用标准的财务数据收 集方法。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.6 | 成本与收益衡量成为制订数据治理优先事 项的指导 | ●数据治理业务案例的批准文件  ●数据治理TCO和方法论的文档  ●数据治理的成本—收益分析结果  ●数据治理业务案例绩效度量项  ●数据治理TCO的追踪矩阵，包括关系和依赖  ●收集关于数据治理成本信息的标准过程  ●数据治理TCO计算的成本分配方法论 |

 **云端数据治理**

【成熟度等级说明】

定义成本与收益衡量指标，是对数据治理活动性价比的具体评价。例如， 一个 组织可能会有一些员工专门负责为财务报告进行数据清洗，综合度量项和成本追踪 会清晰地显示重复清洗的工作量和成本，并与更改源系统的成本相比较，从而在源 头上提高数据质量。

**Level 4: 量化级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 4.1 | 采用数据治理TCO,对数据治理倡议和基 础设施变更进行度量、评价和资助 | ●业务案例  ●总体拥有成本分析文件与结果 |

【成熟度等级说明】

参考资源保障获得以下信息：确保有充足、持续的资金来支撑数据治理方案， 需要进行追踪和报告用于成本比较与业务案例开发的成本分配模型。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 4.2 | 利用统计学和其他量化手段来分析数据治 理成本度量项，从而评估数据治理TCO和收 集方法，并将TCO度量项纳入数据治理方案 绩效体系 | ●数据治理成本度量项  ●总体拥有成本分析文件与结果 |

【成熟度等级说明】

采用统计学方法，定量地评估数据治理成本与效益，并将TCO 度量项纳入数据 治理方案绩效体系。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 4.3 | 对组织数据治理TCO模型的准确性进行确 认和检查，通过定期评审和分析完善模型 | ● TCO管理报告  ●基于成本度量项的方案变更建议  ●基础设施预算  ●数据治理TCO度量项  ●数据治理TCO方法论文档  ●关于结构、方法和计算的数据治理TCO变更建议  ●方案绩效记分卡  ●审核结果 |

【成熟度等级说明】

组织针对数据治理 TCO 模型的准确性开展论证，确保模型的准确性，论证的范 围可以来自组织内部，也可以邀请业内专家指导。要采取定期评审方式，不断完善 数据治理TCO 模型。组织应记录评审过程和 TCO 模型的演变过程。

**第4章\*云端数据战略**

**Level 5: 优化级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 5.1 | 统计学结果和利益相关者的反馈为数据治 理TCO的持续改进提供指导 | ●数据治理成本度量项  ●总体拥有成本分析文件与结果 |

【成熟度等级说明】

基于制订的成本度量项，对数据治理成本分析进行验证并修改，为TCO 持续改 进提供指导。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 5.2 | 组织分享TCO的最佳实践，通过出版物或 举办会议为行业成熟度做贡献 | ●数据治理成本分析最佳实践  ●出版的行业文章、白皮书、会议 |

【成熟度等级说明】

组织在数据治理活动上取得重要成功时，也应该积极分享实践成功经验，为行 业最佳实践做贡献，例如发表行业论文，发布白皮书、举办经验交流会议等。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 5.3 | 采用优化手段和预测模型来预测执行前拟 变更的结果 | ●对数据治理TCO模型的建议变更和更新  ●发表行业文章、发布白皮书、举办经验交流会议  ●预测分析工具和模型 |

【成熟度等级说明】

成本效益分析是持续的过程，随着数据治理收益的变化，成本效益分析结果也 是变化的，进而会影响业务案例的变更，尤其是新的数据治理倡议的规划。组织采 用预测模型，预判可能因为业务案例改变给数据治理带来的影响。预判模型可以是 专业的方法论和工具，可以是行业内已有的最佳实践，也可能是借鉴类似活动得到 的经验总结。

**4.** **实施案例**

(1)文件名称列表

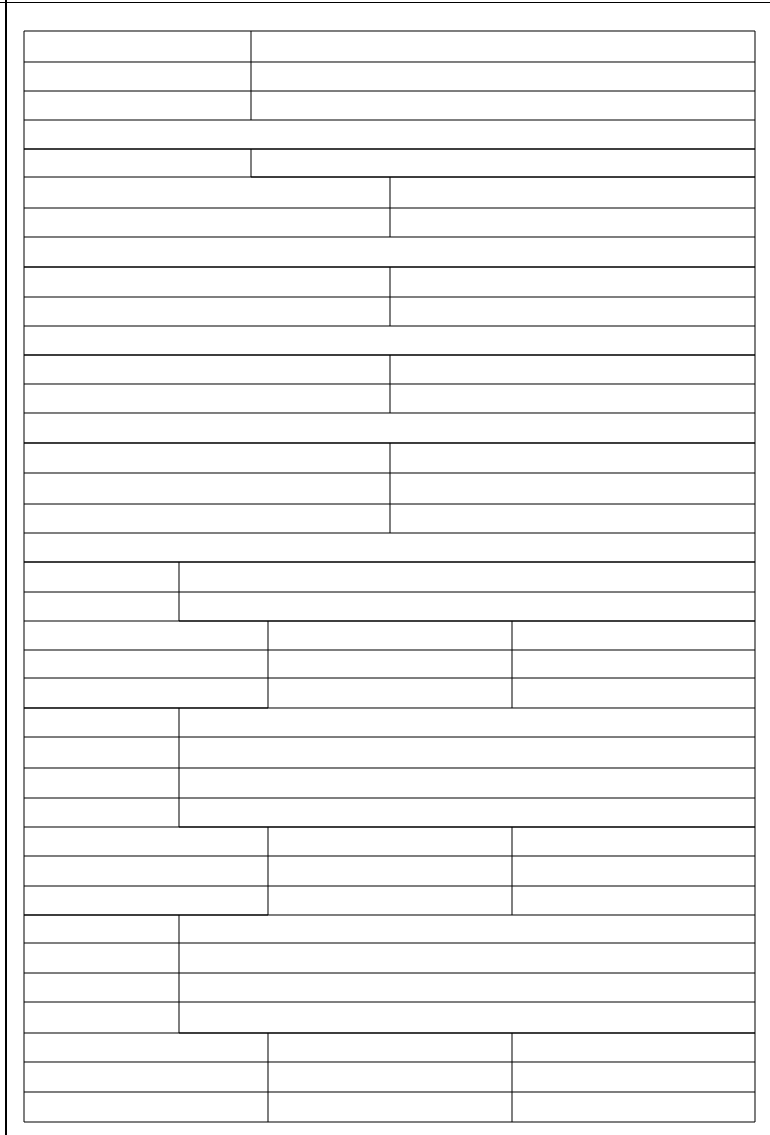
《云端数据治理业务案例》

(2)实施案例

下表展示了一个业务案例的模板。

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 业务案例模板   |  |  | | --- | --- | | 一、项目名称 |  | | 二、项目概述 |  | | 三、业务机会 |  | |

**■云端数据治理**

续表

四、业务结果潜在效应

五、活动倡议

六、假设条件

七、倡议分析

倡议1.

优势分析 劣势分析

倡议2.

优势分析 劣势分析

倡议3.

优势分析 劣势分析

八、风险分析与处理计划

风险

处理计划

九、业务实施计划

阶段一

活动概述

交付物

责任人

截止时间

资源需求

成本效益分析

阶段二

活动概述

交付物

责任人

截止时间

资源需求

成本效益分析

阶段三

活动概述

交付物

责任人

截止时间

**第4章■云端数据战略**

**4.2.4** **资源保障**

**1.** **简介**

1)目的

为数据治理活动提供所需的各类资源提供能力进行规划，包括人力和物力资源等。

2)介绍

资源是保障数据治理活动实施最基本的条件，包括人力和物力资源等。组织应 当确保足够的人力资源投入到数据治理活动中，并让其他数据利益相关者参与到数 据治理过程中。

人力资源保障。数据治理是高层次的、规划性的数据管理活动，需要众多角色 的协调配合才能得以完成，按照管理职责权限的大小，数据治理人员大致可以分为 三个等级，即数据治理委员会、数据治理工作组和基层数据治理专员。

对于数据治理人员，组织要提供管理、培训和绩效等办法，通过管理使其各司 其职，推进数据治理活动的井然有序进行；通过培训活动，持续提升数据治理人员 能力；通过绩效办法，提高人员积极性。

物力资源保障。物力资源包括资金、场地、设备等资源，其中资金是最为关键 的一部分。组织制订资金支持方案来满足业务、IT 、数据治理功能等要求，并将方 案资金传达给数据治理，获得数据治理的认可，从而保证监督和责任。 一个结构清 晰的资助方案有助于促进协作者和利益相关者之间的一致性。

组织需要建立对数据治理长期投入的资金支持方案，并通过定量指标衡量资金 的使用和绩效，确保资金的合理性、适度性、及时性和有效性等。

场地和设备等资源是资源保障的重要组成部分，在制订数据治理战略时，要考 虑组织的物力资源保障能力，例如数据的存储工具及场地。

3)相关过程域

●战略过程域为数据治理活动制订顶层规划，对数据治理活动所需的资源进行 总体要求和规划；

●组织与角色过程域为数据治理活动的实施提供组织架构和人员保障，是资源 保障过程域的具体细化；

●业务案例过程域是对数据治理方案、倡议或具体项目进行资助的提案，涉及 资金的支持和使用跟踪。

 **云端数据治理**

**2.GQM 表**

|  |  |  |  |
| --- | --- | --- | --- |
| 目 标 | 问 题 | 度 量 项 | 度量项说明 |
| G1:建立一个 完善的资源保障 策略 | Q1:是否有一套经批准 的数据治理投资准则和重 点 | M1:资源保障策略完整性 | 为数据治理制订了完善的 资源保障策略 |
| Q2:是否能提供稳定充 足的资金保障 | M2:总体拥有成本 | 资金保障重点分析了数据 治理的收益和成本 |
| Q3:资源保障策略是否  由数据利益相关者制订、评 价和批准 | M3:审批记录 | 资源保障策略得到了数据 治理高层的批准 |
| M4:评审记录 | 资源保障策略得到了数据 利益相关者的理解和支持 |
| G2:资源保障促 进数据治理实施 | Q4:资源保障策略是否 具备动态调整的能力 | M5:策略调整能力 | 建立资源保障的效果评价  机制，动态地调整策略以促进 数据治理活动的实施 |
| G3:资源保障反 映了业务目标和  组织级优先事项 | Q5:资源保障策略是否  反映了组织的业务模型、优 先事项和财务决策过程 | M6:战略一致性 | 资源保障策略符合组织的 投资战略和总体规划 |
| 资源保障策略符合数据治 理优先事项和目标 |

**3.** **成熟度等级介绍**

**Level 1: 初始级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 1.1 | 为组织数据治理活动提供了基本的人员 资金、场地等资源 | ●资金支持计划  ●场地批准记录 |

【成熟度等级说明】

数据治理活动是小范围内的，组织没有建立完善的数据治理架构，因此为数据 治理活动提供的资金和场地等资源也是根据需要估算而来的，没有进行TCO 分析， 也没有持续改进资源保障策略的计划。

**Level 2: 管理级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.1 | 利益相关者参加并支持数据治理资源保障 策略 | ●数据治理资源保障策略  ●数据治理资源保障策略资助批准文件和记录 |

【成熟度等级说明】

数据治理是一个被广泛参与的活动，数据利益相关者的关注和参与是保障数据 治理实施的重要条件，为数据治理提供人力、物力等条件的资源保障策略也必须得

**第4章■云端数据战略■**

到数据利益相关者的参与和支持。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.2 | 基于成本一效益分析对数据治理项目进行 资助 | ●数据管理项目预算  ●包含成本—效益分析的资助申请 |

【成熟度等级说明】

组织有计划有目的地科学分配资源，基于成本一效益模型，最大化发挥资源的 作用。资源分配的方式根据资源种类的不同而不同。

第一，人力资源的分配。人力资源的分配一般有三个主要内容。

(1)为数据治理岗位配备管理和技术人才，特别是对关键岗位的关键人物的选择。

(2)为数据治理战略实施建立人才及技术的储备，不断为战略实施输送有效人才。

(3)在数据治理实施的过程中，注意数据治理团队的综合力量的搭配和权衡。

第二，资金的分配。 一般采用预算模型来为数据治理活动分配资金资源，而预 算是一种通过财务指标或数量指标来显示组织目标、战略的文件。通常采用这四种 预算方式。

(1)零基预算。它不是根据上年度的预算编制，而是将一切经营活动都从彻底 的成本一效益分析开始，以防止预算无效。

(2)规划预算。它是按规划项目而非职能来分配资源。规划预算的期限较长， 常与项目规划期同步，以便直接考察一项规划对资源的需求和成效。

(3)灵活预算。它允许费用随产出指标而变动，有助于克服“预算游戏”及增 加预算的灵活性。

(4)项目生命周期预算。在项目的不同生命周期中有着对资金的不同需求，而 且各阶段的资金需求有不同的费用项目。这时项目生命周期预算就根据不同阶段的 特征来编制各项资金的支出计划及原则。

在资金的分配中，应该遵循两项基本原则。

(1)根据各子活动对整个数据治理战略的重要性来设置资金分配的优先权，以 实现资源的有偿高效利用。

(2)兼顾数据治理整个项目的各个单元，发挥资金促进数据治理活动的协同功能。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.3 | 基于业务案例中论述的资助准则对数据治 理倡议进行资助 | ●数据治理业务案例  ●数据治理资源保障方案  ●数据治理资助准则  ●数据治理资助记录 |

【成熟度等级说明】

基于业务案例对资金支持策略进行评审，主要内容为资金对数据治理战略的支

**■云端数据治理■**

持程度，同时也可以考虑其他因素。业务案例给出了数据治理活动的优先级，并以 成本—效益模式计算所需资金的预算。在项目结束后，对资金收益进行追踪，并利 用这些收益来完善收益估算。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.4 | 数据治理成本估算扩展到组织的业务领 域、日常操作和IT领域 | ●数据治理预算  ●与资助相关的治理文件 |

【成熟度等级说明】

数据治理覆盖部门多，参与人员广泛，因此对其成本估算是一项复杂的工作， 是跨部门、跨层级、跨环节的。例如，数据治理需要业务部门、财务部门、行政部 门、人力资源部门、供应链部门等，需要将每个部门参与数据治理所需要的成本从 部门整体成本中剥离出来。数据治理生命周期覆盖数据治理项目规划、启动、实施 和优化等多个环节，各个环节对资金的需求不一样，支持方式也可能发生变化，因 此对数据治理成本估算是一个跨周期的活动。数据治理成本的来源多种多样，包括 数据采集成本、工具使用许可证的购买、IT 系统的执行和维护以及管理所有活动的 员工。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.5 | 制订了资源支持过程，并以此过程开展对 数据治理活动的支持 | ●数据治理业务案例  ●数据治理资源保障方案  ●数据治理预算  ●关于数据治理成本对总体方案、业务单元和项目 映射的管理报告  ●与资助相关的治理文件 |

【成熟度等级说明】

资源的分配和安排过程需要制度化，并进行全过程跟踪管理，这样做一来可以 保证资源支持的连续性、及时性和针对性，二来可以通过规范流程控制资源支持的 差错率，提高资源使用效率和收益。

**Level 3: 定义级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.1 | 为数据治理制订了完善的资源保障策略 | ●数据治理资源保障策略 |

【成熟度等级说明】

完整的资源保障策略应该至少包括资源支持标准、资源支持过程管理流程、资 源投资收益测算方法、资源支持过程记录等内容。

**第4章■云端数据战略**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.2 | 资源保障策略被确定为组织数据治理活动 投资的决策标准 | ●数据治理资源保障策略  ●数据治理资助准则 |

【成熟度等级说明】

用于映射成本和计算数据治理收益的标准方法在组织内得到建立和遵循，成为 其他数据治理新倡议资源支持的参考标准。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.3 | 资源保障策略支持的重点与数据治理的目 标和重点相符 | ●数据治理资源保障策略 |

【成熟度等级说明】

资源保障是数据治理活动的基本需求，在资源总数有限的情况下，应当参照数 据治理的目标和重点，合理地分配资源，提高资源支持的针对性，提高资金投资回 报率，也就是人们常说的“把钱花在刀刃上”。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.4 | 定义度量指标来衡量资源支持的目标和预 期收益的有效性 | ●数据治理资助准则  ●预算策划过程  ●衡量投资和资助目标的度量项  ●文档化的数据治理资助模型  ●度量数据治理收益的报告  ●优先次序准则和对数据治理战略的映射 |

【成熟度等级说明】

资源投资有效性需要具体的衡量指标，尤其是资金的有效性， 一般的投资衡量 指标包括财务净现值、财务内部收益率、动态投资回收期、投资利润率等经济指标。 对于数据治理，还可以将技术指标纳入进去，例如得益于数据治理对某项设备的精 度的提高，以及某项无形资产的评估价值提升等。

**Level** **4:量化级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 4.1 | 对度量项进行了定义和统计学分析，从而 评价资源保障策略在满足组织级目标方面的 有效性和准确性 | ●资源保障有效性的度量项  ●源保障有效性分析结果及记录 |

【成熟度等级说明】

不同行业可以制订不同类型的个性化度量指标，从而评价资源支持对数据治理 的有效性，以统计学的方法，在数据治理的不同阶段，动态地对资源支持的有效性

**■云端数据治理■**

进行评价，并跨环节比较分析有效性差异。

**Level** **5:优化级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 5.1 | 与同行分享从组织级数据治理资源保障获 得的经验教训 | ●同行交流会议  ●关于数据治理资源保障的公开陈述或白皮书 |

【成熟度等级说明】

沟通交流是促进知识和经验共享的最佳途径，云计算环境就是一个开放包容的 大环境，云端数据治理不是一个只属于组织内部的活动，它涉及外部供应链等整个 生态圈，尤其和云服务提供商和监管方存在紧密的关联。组织对外分享云端数据治 理经验是其数据治理战略和目标的对外宣告，在为其他组织提供参考的同时，也能 增强其在生态圈中的声誉，在能够获取外部反馈意见的同时，也能向客户展示其信 息治理能力。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 5.2 | 采用优化手段和预测模型，在实施之前  分析对资源保障方法所做的拟修改的预测结 果 | ●关于数据治理资源保障的公开陈述或白皮书  ●基于预测分析对资源保障做出的核准更改 |

【成熟度等级说明】

资源保障是一个动态调整的过程，资源提供能力会随组织的变化而变化，例如 人员的流动、组织经济效益的变化、组织发展策略的调整等都可能影响组织资源的 保障能力。组织采用预测模型，预判可能因为资源保障方法改变给数据治理带来的 影响。预判模型可以是专业的方法论和工具，可以是行业内已有的最佳实践方案， 也可能是借鉴类似活动得到的经验总结。

**4.** **实施案例**

(1)文件名称列表

《云端数据治理资金管理总体方案》

《云端数据治理资金支出管理办法》

《设备采购管理办法》

(2)实施案例

下表展示了一个简要的云端数据治理资金管理总体方案。

|  |
| --- |
| 云端数据治理资金管理总体方案  1 目的  制订云端数据治理项目实施的资金保障措施，以满足数据治理活动要求的流动资金，最大程度发挥数据治理专 项资金的效益，为数据治理战略制订的任务完成提供资金保障，提高公司数据管理和利用能力。 |

**第4章■云端数据战略**

|  |
| --- |
| 2 资 金 来 源  坚持必要资金需求必须到位、可调节资金部分优先保证等原则，拓宽多渠道资金筹措方案，强化流动资金筹措。 随着数据治理活动的深入，公司原制订的资金方案必要时可以修改，确保满足数据治理活动对资金的需求。  3 资 金 使 用  贯彻执行节约成本，降低消耗措施。项目要采取有效措施，控制成本消耗、节约各项费用。要正确编制业务案  例与资金预算方案，严格按计划开展数据治理活动。要把成本指标和使用效率作为数据治理成果评价的指标之一  4 资 金 管 理  数据治理在规划时，要正确分析、预判资金需求计划，做好成本效益分析，合理安排资金。对数据治理子活动， 应区分轻、重、缓、急，统筹安排，坚决保证重点数据治理倡议的资金保障。  为更好地管理数据治理资金使用效率，公司会安排专人主管资金使用，优化资金审批流程，记录资金使用过程， 并根据相关指标评判资金使用效率。  5 其 他  数据治理过程中，对资金到位情况、项目的运作情况、进度、成本的控制等进行实时监控，从而做出科学的计  划调整决策，以保证施工的高效、安全。财务部要以目标成本为基础，进一步整合规划与进度计划，实时统计项 目实际发生成本并与目标成本进行对比分析，实现成本的动态控制，有效控制资金使用风险。 |

**4.2.5** **沟通**

**1.** **简介**

1)目的

提供一个用于组织和数据治理项目的政策、标准、过程、进度通告以及其他数 据治理事宜发布和反馈的渠道，以便数据利益相关者获知和理解。

2)介绍

数据治理是一个跨部门、跨层级的活动，参与成员多，事务复杂，因此必须做到 各成员之间信息的畅通交流。沟通的内容包括与数据治理相关的政策、战略、角色与 职责、计划、机制和变更等，沟通的范围涉及所有利益相关者。沟通的渠道包括会谈、 电话、即时通讯工具、电子邮件、公告板、电子屏、报刊、杂志、媒体等。

公布指的是，在制订了一项新的方针后对其进行正式公告或宣布。对于许多组 织而言，在方针生效之前，是有必要进行公布的。应通过已建立的方针管理机制论 述对数据治理标准、政策和过程的公布，或将其纳入到组织的沟通战略当中。

对于一些利益相关者和组织，如果沟通渠道囊括了需要他们亲自参加的活动(如 常规会议、制度化的工作坊、市政会议等),那么这些渠道就更加有效。沟通的机制、 频率和时机也是数据治理沟通战略的重要部分。除此之外，如果沟通是来自高级管理 层，要求采取措施或给予回应的沟通可能更加有效，也就是应当采用反馈确认机制。

为了确保组织可以实现其目标，必须在预计采纳日期和合规检查开始之前，对 政策、标准和过程进行公告。应建立一个反馈过程来支持利益相关者的问题、建议 等，可以利用采纳度量项来评估对公布过程有效性的度量。

已经发表的政策、标准和其他指南应当让组织成员随手获得，例如集中化的电

**■云端数据治理**

子图书馆可以做到这一点。要对电子图书馆进行及时维护，从而确保可以随时更新 资产，还要有向相关利益有效传达更新内容的方法。 一般的沟通可以很简单，例如 发布一个通告，或更全面一点的，进行一次培训。

3)相关过程域

所有过程域都会涉及需要发布的内容，因此云端数据治理所有的过程域均与沟 通过程域相关。

**2.GQM 表**

|  |  |  |  |
| --- | --- | --- | --- |
| 目 标 | 问 题 | 度 量 项 | 度量项说明 |
| G1:建立了数 据治理沟通战略 | Q1:如何发布数据治理 的政策、标准和过程 | M1:沟通战略 | 建立了数据治理沟通战略  包括沟通方式、途径、要求等， 支持数据治理策略的有效传达 |
| Q2:组织如何向利益相 关者通知有关数据治理计 划和项目的消息 | M2:沟通计划 | 数据治理沟通战略能够满  足数据治理方针要求，制订沟 通计划 |
| Q3:沟通战略是否得到 数据利益相关者认可 | M3:评审记录 | 沟通战略得到了数据利益 相关者的理解和支持 |
| G2:数据治理  沟通战略确保了 相关人员能够在 恰当的时间了解 关于方案的正确 消息 | Q4:沟通的有效性怎么 评判 | M4:及时性 | 衡量沟通内容传达的速度 和反应时间 |
| M 5 : 响 应 率 | 衡量内容被反馈的比例 |
| Q5:如何在业务、IT、  数据治理和高管之中实现  数据治理优先事项、方式和 工作输出的双向沟通 | M6:沟通记录 | 记录沟通的时间、方式、内 容等，做到沟通可追溯 |
| M7:改进措施 | 对沟通战略进行评审，制订 沟通战略改进计划 |

**3.** **成熟度等级介绍**

**Level 1: 初始级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 1.1 | 因数据治理项目开展临时的沟通 | ●通告、邮件、会议记录或网页门户 |

【成熟度等级说明】

在不同的组织区域和不同的项目，沟通的级别和频率会有很大的差异。沟通策 略不够健全，往往会是临时的。

**Level 2: 管理级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.1 | 数据治理沟通计划得到了定义、记录和安 排，并获得利益相关者的认可 | ●数据治理沟通计划  ●数据治理沟通计划得到评审和授权的记录 |

**第4章■云端数据战略■**

【成熟度等级说明】

组织对数据治理沟通进行了策划，并确定作为数据治理活动的正式沟通渠道。 沟通战略的内容包括沟通的主题(如数据质量)、类型(如剖析过程、数据质量方针)、 受众(如数据管理员)、频率(如每季度、由事件决定)、渠道(公告板、邮件、书 面通知)、确认方式(需要反馈、不需要反馈)等。制订的沟通战略应该得到管理者 批准，且获得数据利益相关者的认可。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.2 | 对数据治理的标准、策略和过程进行沟通， 并基于反馈进行调整 | ●沟通方针  ●通告、邮件、会议记录或门户  ●沟通战略  ●沟通实例 |

【成熟度等级说明】

沟通的目的是保证策略、决议等得到有效传达、理解。数据治理的标准、策略 和过程是指导开展数据治理的纲领，为了保证数据治理相关者能够获取并理解数据 治理策略，应当进行必要的沟通，并基于反馈对治理策略进行调整。沟通的渠道是 多样的，可以是通告、邮件、会议记录或门户等。

**Level 3: 定义级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.1 | 沟通方针确定了对不同类型的数据治理沟 通进行传播或公告的标准，定义了沟通渠道 策略 | ●沟通方针  ●通告、邮件、会议记录或门户  ●沟通渠道策略 |

【成熟度等级说明】

有效的沟通渠道是组织中维系人与人之间良好工作关系的关键因素，是组织的 血管，是一种信息或思想的双向传递过程，这就是所谓的“上传下达”。

沟通具有五大要素：目的、对象、内容、方式和时间。

沟通的目的是首要考虑的，不同的沟通目的会选择不同的沟通方式发送不同的 沟通内容。 一般包括为了获得某种反馈信息、说服他人、对策略施加影响、咨询等。

沟通对象是需要获知信息的那一方，对于数据治理活动，沟通的对象可能是数 据治理负责人、跨部门人员、甚至组织外部数据利益相关者。

内容是沟通的关键，有时会因为信息的属性采用加密或不加密方式。

沟通方式，即沟通渠道，是描述如何进行沟通，常用的沟通方式包括：面谈、 电话、命令、文件、会议、报刊、广播、宣传栏、意见箱、活动、即时通讯、网络 等。不同的沟通方式会对沟通效果产生巨大影响。

组织需要制订一套策略，支持数据治理沟通渠道的选择策略。

**■云端数据治理■**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.2 | 沟通战略受组织方针指导，反映了数据治 理战略 | ●沟通方针  ●沟通渠道策略 |

【成熟度等级说明】

沟通战略不是独立于组织方针的一套体系，必须与组织的方针保持一致，同时 沟通战略是服务于数据治理的，要有针对性。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.3 | 在组织内对标准、政策和过程进行公告， 并基于反馈进行调整 | ●沟通方针  ●沟通记录  ●沟通反馈  ●标准、政策和过程的调整 |

【成熟度等级说明】

沟通的方式有多种，组织在针对某一内容开展沟通时，至少应当选择一种便于 反馈的方式。反馈的次数可以是多回合的，组织要基于反馈的信息，对数据治理相 关的标准、政策和过程进行调整。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.4 | 利益相关者根据已定义的标准所要求的过 程以及其他过程对沟通进行评审 | ●沟通方针  ●沟通评审记录  ●标准、政策和过程的调整 |

【成熟度等级说明】

为了保证数据利益相关者能够及时了解数据治理的过程，沟通是数据利益相关 者获取相关信息的最佳渠道。数据利益相关者根据已定义的标准对沟通策略进行评 审，提出对沟通策略改进的建议。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.5 | 制订度量项，并利用这些度量项来衡量数 据治理沟通的有效性 | ●沟通方针  ●通告、邮件、会议记录或门户  ●沟通有效性度量项  ●沟通计划  ●关于沟通的同行反馈 |

【成熟度等级说明】

制订度量项来评估沟通的有效性是改善沟通的重要方式，可以根据不同的沟通 方式制订相对应的度量项。例如，进行问卷调查时，回收份数、回收率、有效回收 率、调研对象代表性等指标可以用来作为度量项。

沟通度量项取决于沟通战略中定义的目标、目的和渠道。这些度量项应衡量数 据治理沟通的有效性和效率。沟通度量项的实例包括已确定的利益相关者的比例；

**第4章■云端数据战略**

传达给利益相关者的数据管理政策和程序的采用率；参与制订数据治理沟通战略的 利益相关者的比例。

**Level** **4:量化级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 4.1 | 依据沟通战略，对与外部利益相关者进行 的数据治理沟通进行策划和实施 | ●沟通方针  ●通告、邮件、会议记录或门户  ●与外部的沟通记录  ●沟通成果 |

【成熟度等级说明】

数据治理不是组织的完全封闭活动，尤其是云计算环境下数据的远程分布式存 储，数据流通站更多，数据来源更广泛，数据安全保护要求更多，因此数据治理必 须是一个与外部不断沟通的活动。与数据治理相关的外部机构一般包括监管机构、 工业与贸易组织、标准化机构等， 一个完备的沟通战略应该包括与外部机构的沟通 策略。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 4.2 | 采用统计学手段或其他量化手段来完善数 据治理沟通 | ●对与沟通有效性度量项相关的数据治理沟通计划 进行的更改  ●监管类通信记录(即对问询的回应、报告和备忘录) |

【成熟度等级说明】

统计方法是指有关收集、整理、分析和解释统计数据，并对其所反映的问题做 出一定结论的方法。通过对用来度量沟通有效性的度量项进行分析，评价沟通策略 和过程的有效性与效率。例如，进行问卷调查时，回收份数、回收率、有效回收率、 调研对象代表性等指标可以用来作为度量项。

**Level** **5:** **优化级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 5.1 | 为促进行业发展，提高对外影响，进行了 数据治理的外部沟通 | ●外部沟通  ●受影响的公共政策和行业最佳实践  ●对外发布的白皮书、标准、论文、报告等 |

【成熟度等级说明】

组织具备高度的社会责任感，将数据治理成果惠及客户、监管机构、产业链等， 公开或定向公开数据治理经验，帮助外部组织提高能力等。

 **■云端数据治理**

**4.** **实施案例**

(1)文件名称列表

《信息发布管理办法》

《微信群管理办法》

(2)实施案例

下表展示了某组织对于发布公告的管理规定，以及某公司微信群管理办法，仅 供参考。

|  |
| --- |
| 关于发布公告的管理规定  各部门/各单位：  为了加强我司OA系统内电子公告栏的信息管理，规范电子公告栏信息发布行为，确保公告发布的严肃性，现对 公司的公告发布规定如下，请各部门/单位遵照执行。  1 公告发布管理职能部门  行政及人力资源部  2 发 布 范 围  规章制度、通知、决定、会议纪要、值班表、快讯等以公司或部门名义发布的公告。  3 发 布 格 式  所有需发布公告，由行政及人力资源部制订标准格式、统一编号。  4统一发布公告程序  1)以部门名义发布公告程序  (1)以部门名义发布公告，由有关人员起草文稿后，填写《公告发布审批单》,由部门经理审核后，报分管领导 审批。  (2)承办人员将《公告发布审批单》交行政及人力资源部存档，将电子文稿发送至行政及人力资源部局域网信 箱，并对文件需发送的范围等要求加以说明。  (3)行政及人力资源部根据公告内容统一行文格式并按承办人要求发布。  2)以公司名义发布公告程序  (1)以公司名义发布公告，正式发文的填写《发文稿纸》,经有关人员会签后，报总经理签发，由行政及人力资  源部按行文格式统一发布。  (2)没有填写《发文稿纸》的文件，如需发布公告，必须填写《公告发布审批单》,报总经理批准。  5公告栏实行账户管理，信息管理中心负责公告栏的维护管理工作，对不遵守电子公告栏管理规定的用户，可 给予警告、封闭或删除其账户。  6任何人不得在电子公告栏上随意制作、发布、复制。违者必究。  行政及人力资源部 二0一五年四月一日 |

**第4章■云端数据战略**

|  |
| --- |
| 微信群管理办法  1 总 则  为更好地利用互联网技术，方便公务交流和提高办公效率，进一步规范微信群的管理工作，特制订如下管理办 法。  2 微 信 群 设 置  (1)微信群名：XXXX;  (2)公司全体管理人员(中心领导、部门领导)必须加入本群，其他群员须为公司正式员工(不包含兼职人员);  (3)所有群员在进入本微信群后，须采用实名制，以“姓名部门”为群昵称。具体操作步骤：进入微信群后，点 击屏幕右上角的图标，在打开的菜单中选择“详细设置”对“我的群昵称”进行修改即可。  3信息发布要求  (1)发布内容应正面、积极、正能量；  (2)群内交流须遵守国家法律法规及相关网络信息管理规定，严禁出现不良政治倾向、宗教、色情、暴力等内 容 ；  (3)严禁在群内诋毁公司、公司各部门及员工形象，禁止出现有违社会公德、不文明、侮辱性及涉及人身攻击 性语言 ；  (4)严禁在群内发布有煽动性、过激性的语言；  (5)严禁在群内发布或谈论涉密信息；  (6)严禁在群内发布色情、赌博、病毒链接；  (7)严禁在群内进行与工作或集体荣誉无关的拉选票、发布广告等商业活动  (8)严禁在工作时间发布与工作无关的内容；  4微信群管理  (1)微信群统一由行政部管理，执行微信群管理规定，负责群成员实名制、聊天监管、违规处理；  (2)负责审核群成员发布信息的真实性、可靠性和合规性，对非本群应发内容有权制止，并指导其发布相应内 容 ；  (3)定期检查群成员，对不应加入人员予以剔除。  5本办法自发布之日起实施，最终解释仅归行政部所有。  行政部  二○一六年六月六日 |

**第** **5** **章** **云端数据管理**

**5.1** **概述**

**5.1.1** **目的和意义**

数据和信息蕴含着巨大的商业价值，已有越来越多的组织将其视为具备战略意 义的组织资产。随着组织不断将业务和应用搬移到云端，云端数据的有效管理逐渐 成为组织能否充分利用数据、能否挖掘其商业价值的核心。云端数据管理是组织在 云计算环境下开展数据治理的基础保障和有效手段。

**例1:数值单位不统一造成** **NASA** **火星气候探测器事故**

1999年，价值1.25亿美元的火星气候轨道器在到达火星之后，刚刚准备开始工 作便失去了与地球的联系，后来被证实轨道器坠入火星大气层之中烧毁了。根据美 国NASA 火星气候轨道器事故调查委员会发布的报告显示：火星气候轨道器任务失 败的主要原因是因为火星气候探测者号上的飞行系统软件使用公制单位(牛顿)计 算推进器动力，而地面人员输入的方向校正量和推进器参数则使用英制单位(磅力), 导致探测器进入大气层的高度有误，最终瓦解碎裂。

**例2:数据管理有助于实现组织的全局视角**

随着 SaaS 服务逐步普及，云端的客户数据是目前企业级客户普遍面临的一个问 题。在大多数组织中，客户信息通常分散于CRM 、ERP 、OA 等各个业务系统中， 而每个业务系统中都只有客户信息的片断，即不完整的客户信息，但缺乏企业级的 完整、统一的单一客户视图，结果造成组织不能完全了解客户，无法协调统一的市 场行为，导致客户满意度下降，市场份额减少。因此，开展云端数据管理，整合并 存储所有业务系统和渠道的客户及潜在客户信息，并为相关的应用系统提供联机交 易支持，提供客户信息的唯一访问入口点，为所有应用系统提供及时和全面的客户 信息将会在很大程度上解决这一问题。

对于组织开展云端数据治理来说，云端数据管理作为基础保障，使得组织能够 集中化管理迁移至云端的各类数据以及直接从云端收集的与组织相关的数据，在分 散的系统间保证主数据的一致性，改进数据合规性、快速部署新应用、充分了解客 户、加速推出新产品的速度。从组织自身的IT 建设角度来说，云端数据管理可以增 强IT 结构的灵活性，构建覆盖整个组织范围内的数据基础和相应规范，并且更灵活

**第5章■云端数据管理**

地适应组织业务需求的变化。

**5.1.2** **内容与关联性**

云端数据管理职能域中包含的过程域有：业务词汇表、元数据管理、主数据管 理。通过这三个过程域，使组织能够对其管理的云端数据在内、外部形成统一的理 解，从而支撑云端数据治理的有效开展。

本职能域为以下方面提供了最佳实践：实施能够在共同职责的范围内持续运作 的数据治理方案和结构；拓宽和管理组织对经核准的业务词汇的收集，这些词汇被 用到目标数据架构、分类法和本体论当中；开发和执行元数据，从而完整地描述组 织的数据资产。数据管理对促进协作决策和有效执行治理机构的建立、维持以及合 规功能的建设进行了论述。

业务词汇表实践，针对组织内经核准的业务词汇，使其意义得到理解，有助于 组织对其开发进行优先排定和次序排列，并对后续词汇的创建和更改进行管理。元 数据管理提供了一个构建、策划、推广和管理元数据库的由上而下的方法，从而完 整地描述组织的数据资产。主数据管理可以帮助组织构建并维护贯穿整个信息供应 链的主数据单一视图，提供主数据的质量管理和统一业务实体定义，简化并改进业 务流程并提高业务响应速度。

本职能域的实践成熟度有助于加强数据管理方案的形成，从利益相关者处获得 不间断的拥护和支持，并凭借通过全面而周到地实施过程所获得的信心，帮助组织 实现其战略目标。

**5.2** **活动及要求**

**5.2.1** **业务词汇表**

**1.** **简介**

1)目的

云端业务词汇表的主要目的包括

(1)在业务涉及的范围内统一核心业务用语和概念的理解，便于云端数据的应 用和管理(特别是针对云端数据应用涉及多组织/用户的情况下);

(2)降低因为对于业务词汇和概念理解不一致带来的数据误用；

(3)为组织内精确查找业务相关信息提供基本保障。

**■云端数据治理**

2)介绍

业务词汇表包含与组织相关的词汇、词汇业务含义以及词汇与信息资产(技术 元数据)的关系，可以有效帮助组织用户了解其业务元数据和技术元数据的对应 含义。

业务词汇表是组织数据基础的核心，是组织开展云端数据治理的最基础支撑。 尽管词汇表是个很简单的概念，但要对用于表达业务含义的共享词汇进行有效的管 理，并进行定义、统一、协调、挑选、批准和维护，事情就可能变得很难。保持业 务词汇含义的一致性非常重要，尽管在组织中对于不同的业务线而言也许并非都十 分明显，因为业务单元和组织级使用的词汇之间的区别通常没有明确的定义或清楚 的记录。此外，多个业务单元使用的词汇通常包含了隐含意义，这需要通过探索研 究、合理解释和协议等方式进行确定。为业务词汇建立标准，包括命名惯例、缩写 词、标准的定义表示法以及标准元数据，为组织内表示统一意义奠定了基础。这些 标准应通过数据治理进行应用，还应建立、遵循和维护相应的批准和更改过程。应 实例化合适的沟通或反馈环路，从而确保业务过程中的变更和建议得到恰当传达， 以确保词汇表的准确性。

在复杂的分析、建模和开发计划涉及多个项目的大型组织中，业务词汇的标准 化至关重要。如果没有一种方法来标准化业务概念的含义和定义，显然，每个分析、 建模或开发线程都会建立自己的词汇。这些完全不同的词汇可能会加深用户对IT 资 产和它们支持的业务概念之间的关系理解不一致的情况。

经核准的标准业务词汇表强调了向目标数据架构的有效过渡。没有这个词汇表， 组织数据资产的重新架构、整合和有效分享的速度就会更慢、更加复杂、更加昂贵。 新数据存储的开发、整合和重新设计通常是由事件驱动的，经常会导致业务词汇特 殊命名和定义的出现，这些词汇之后会被实例化成逻辑属性和物理数据元素。每一 次都会出现这种情况，由此产生了对经核准标准词汇的组织机构进行创建和建设的 另一个障碍，这种障碍必然会导致未来的返工。

在组织的数据层，经审核批准的统一业务词汇也是完善组织数据接口的奠基石， 这些数据接口总是因目的相同而名字和表示方式不同而饱受困扰，带来了高昂的成 本。同样，统一的业务词汇和定义对于数据整合、集合、准确分析、趋势和预测分 析、语义建模等的实现十分关键。制作出让所有应用程序、系统和过程都理解和应 用的业务词汇综合概要——这种活动的广阔范围，需要从执行管理层那里获得政策 指示，强调了建立和控制数据统一性的重要性。项目可能会试图避免采用经核准的 词汇，从而完成短期目标。强有力的治理对于在早期拦截执行工作从而确保组织级 方针得到遵循是必要的。

开发一个已定义的过程，用来创造、更新、维护和公告业务词汇及其相应的定 义，这一点十分关键。组织业务词汇表的管理需要采纳这样的观点：意义的综合对 于数据管理的长期成功十分必要。因为现有系统的数据名称、定义、容许值和其他

**第5章■云端数据管理**

相关元数据在缩减，建议按照数据管理战略中确定的优先级排列次序。对业务词汇 表的各部分排列顺序和确定优先级的最好办法，取决于该工作内容的性质。数据库 存项目(如设计和创建一个操作型的数据存储)倾向于对由主题领域(如交易、产 品、顾客等)定义的数据集进行顺序和优先级排列。为了实现应用程序和商务智能 方面的目的，要进行业务过程的数据经常会跨主题领域水平拓展。理想的话， 一旦 组织确定了要增加一个完整全面的业务词汇表的目标，就会通过多个途径取得进展： 由上而下，即由受业务线维护的多个词汇表对新的词汇表进行补充；由下而上，主 要是获得数据模型或数据库脚本。

词汇对于大多数组织来说，业务词汇表是一份经批准并受控的业务词汇名称及 定义的汇编文档。本过程域定义了组织如何创造、批准、更新和公布统一的词汇和 定义，促进在组织内对共享数据的使用。在统一的业务词汇和定义下，结合对应的 元数据开展组织数据管理才显得有意义。业务词汇表为在组织内理解和整合数据提 供了一个扎实的基础。组织编制的业务词汇表应当确保每一个定义的词汇都能明确 代表一个特定的意思，没有歧义或重复。

3)相关过程域

●数据生命周期管理过程域有助于确定所有与数据管理相关的当前活动，从而 协助制作和管理业务词汇表；

●元数据管理过程域有助于确定业务词汇表中定义的数据用途，并确定哪些系 统和模型正在使用数据；

●数据集成与互操过程域为业务词汇表中的定义提供修改和描述的依据，并提 供业务词汇的应用场景。

**2.GQM 表**

|  |  |  |  |
| --- | --- | --- | --- |
| 目 标 | 问 题 | 度 量 项 | 度量项说明 |
| G1:描述数据 的语言要清晰 明确地与业务 语言保持一致 | Q1:是否有指定要使用 或参考业务词汇表的指令 | M1:业务词汇 | 定义业务词汇时须与其用 途相结合 |
| M2:内容完整性、内容一致 性 | 业务词汇表应保持其完整  性(是否包含业务规则和本体 结构),并确认在组织范围内 的业务词汇表是否一致，不应 出现定义冲突 |
| G2:组织已经 创建了一个经 核准且全面的 业务词汇表 | Q2:组织的业务词汇、 定义和相应的元数据是如  何创造、批准、验证和管理 的 | M3:规范化程序 | 需要建立文档化的管理程  序，对业务词汇表进行定义、 使用和维护管理等操作进行 指导 |

 **■云端数据治理**

**续表**

|  |  |  |  |
| --- | --- | --- | --- |
| 目 标 | 问 题 | 度 量 项 | 度量项说明 |
| G2:组织已经 创建了一个经 核准且全面的 业务词汇表 | Q2:组织的业务词汇、  定义和相应的元数据是如 何创造、批准、验证和管理 的 | M4:经认可的业务词汇表 | 业务词汇表须首先经组织  认可，并且在所有相关的业务 活动中按照业务词汇表的定  义来统一、规范业务的理解和 描述 |
| M5:行业一致性 | 业务词汇表中的定义在不 冲突及出现歧义的情况下可 参考行业标准的描述 |
| G3:组织遵循  与业务词汇相关  的命名、定义和  元数据的标准 | Q3:业务词汇是否作为  应用程序数据存储和数据 仓库设计的第一步被引用 | M6:逻辑数据模型 | 逻辑数据模型的描述和理 解须与业务词汇保持一致 |
| G4:整个组织  都可以获得业 务词汇表，使得  所有的利益相  关者都能够了  解标准业务词 汇 | Q4:是否将业务词汇表 向所有的利益相关者公告  并让他们获得词汇表 | M7:业务词汇可用性 | 保持业务词汇的更新和有 效，并且应在组织内部发布， 确保组织内部对业务词汇的 理解一致 |
| Q5:组织是否进行了交  叉引用并将业务的专门词 汇(同义词、业务单元词汇 表、逻辑属性、物理数据元 素等)映射到标准化的业务 词汇 | M8:词汇唯一性 | 业务词汇表中的定义须保 持其唯一性，不出现重复定义 或同一词汇出现多个定义 |
| M9:业务词汇指导性 | 数据应用的全生命周期(开 发、集成、整合等环节)均应 按照业务词汇的标准定义来 描述 |
| G5:数据治理  促进了对业务 词汇的评审、批  准和持续使用 | Q6:组织的业务词汇表  如何得到完善和维护，从而 反映变更和补充内容 | M10:变更控制 | 业务词汇的变更应当进行  影响评估，以确定需要开展的 活动来降低影响 |
| M11:出版物 | 根据组织内外部的实践，把 业务词汇管理的经验形成白 皮书 |
| Q7:数据治理在创造、  批准、管理和更新业务词汇 方面扮演了什么角色 | M12:业务词汇表一致性 | 应当在组织级以上的层面  开展云端数据治理，治理的前 提是业务词汇的定义和描述  的一致性 |

**第5章■云端数据管理**

**续表**

|  |  |  |  |
| --- | --- | --- | --- |
| 目 标 | 问 题 | 度 量 项 | 度量项说明 |
| G6;随着新数 据要求和项目 的出现，合规和 执行过程确保 了业务词汇的 持续应用 | Q8:是否实施了合规过  程来确保业务单元和项目 正确使用了业务词汇 | M13:业务词汇转换规则 | 组织应形成机制，推动业务 词汇定义的流程化和细化 |
| M14:完整性校验 | 对于组织建立的业务词汇  表，通过量化度量来评价其完 整程度 |
| M15:使用一致性 | 对于业务词汇通过监督过  程来确保其正式使用，并针对 不合规的使用进行处理 |
| M16:访问权限控制 | 业务词汇表作为元数据库 的组成部分需要被集成使用， 并且针对不同的业务应用进 行分级管理 |
| M17:优化技术 | 优化业务词汇表以精简、准 确、全面表述和定义 |
| G7:组织拥有 一个沟通计划 和过程，持续反 馈词汇表对数 据使用者和其 他利益相关者 的帮助情况 | Q9:组织是否采用了一 个已定义的过程让利益相 关者提供关于业务词汇的 反馈 | M18:改进计划 | 对于业务词汇的管理过程  使用定量技术监控和优化，并 根据定量管理的分析结果向 高层汇报 |

**3.** **成熟度等级介绍**

**Level 1: 初始级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 1.1 | 根据特定用途定义业务词汇 | ●业务场景描述  ●业务词汇需求分析  ●业务词汇定义(表) |

【成熟度等级说明】

业务词汇定义可能供项目开发使用或者供外部发布使用。

业务词汇可能在开发项目的文档中得到定义。部分词汇表仅供特定业务范围使用； 针对某一业务制订的词汇表中可能包含与其他业务相互重叠或冲突的词汇和定义。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 1.2 | 参考经认可的已定义业务词汇创建逻辑数 据模型 | ●标准词汇与定义  ●业务词汇定义(表)  ●业务词汇评审记录  ●逻辑数据模型 |

**■云端数据治理■**

【成熟度等级说明】

如果存在经认可的标准业务词汇，则在衍生其他属性名称时应参考标准词汇。 每个项目在物理数据存储区实施该实践，可防止项目独立创建非标准词汇，也满足 了阐明词汇并改进其业务使用的目的，而非增加复杂性。业务词汇是描述实体相关 事实(属性)的基础，然而命名规则以及数据库设计原则的不同可能会导致不同的 命名(例如，标准行业做法是每个属性名称都以类单词结尾)。例如，描述公积金计 划的业务词汇可能是“日期：关于计划制订”(Date of Origination Plan), 其对应的 逻辑名称是“计划制订日期”(Plan Origination Date)。

**Level 2: 管理级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.1 | 建立、文档化过程并根据过程来定义、管 理、使用以及维护业务词汇表 | ●业务词汇表管理程序  ●业务词汇表使用说明  ●业务词汇元数据定义 |

【成熟度等级说明】

业务词汇过程的建立由业务部门负责人发起，主要运用数据管理功能，需要相 关者和业务单元专家的参与，涉及云端数据治理。该实践能促进多个业务单元的共 同持续参与，建立共同目标，并达成对期望结果和相应时间线的一致意见。

为词汇、定义以及元数据建立标准也需要各方的参与以及共同的决策来推动。 例如，作为数据需求定义过程的一部分，新方案应使用标准业务词汇来确保语言的 一致，使文本内容之间具备可比性，推动整个组织共同分享数据。同样的，应根据 整个组织的共同意见来更新业务词汇表，通过数据治理的方式来管理词汇表。

获取并维护业务词汇的元数据是跟踪和管理业务词汇表的基础。业务词汇的元 数据例子包括：

●词汇名称 ● 定义

●替换形式(近义词、逻辑形式、物理形式等)

●业务词汇具体信息(例如，状态、保密级别、创建日期等)

●治理(例如，等级、所有者等)

●使用(例如，记录系统、有效值等)

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.2 | 标准业务词汇可用，且发布给了内部相关 者 | ●业务词汇表管理程序  ●业务词汇表使用说明 |

【成熟度等级说明】

业务词汇创建且经认可后，应得以发布并推广。例如，以适当频率通过邮件和

**第5章■云端数据管理**

组织门户网站来告知词汇的发布。这对于成功执行变更管理十分重要。组织可能选 择按阶段实施该任务，例如组织单元针对即将进行的主要开发工作按阶段实施等。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.3 | 业务词汇表里添加的每个词汇都有其唯一 的命名与定义 | ●业务词汇表管理程序—变更  ●业务词汇命名规则 |

【成熟度等级说明】

当“基本词汇”(例如“交易价格”或“产品名称”)衍生出其他词汇时，会在 这些衍生词汇上添加一个或多个定性词，例如“最终交易价格”或者“产品简称”, 并附上相应的唯一定义。这些在基本词汇定义上添加的定性词和修饰词应通过数据 治理的审核。当存在组织必须遵守的行业流行标准时，应对照权威来源对业务词汇 以及定义进行确认。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.4 | 作为数据需求定义过程的一部分，在数据 获取、集成以及整合过程中应当使用标准业 务词汇 | ●云端数据治理需求分析  ●业务数据开发程序  ●过程裁剪指南  ●业务词汇表管理程序 |

【成熟度等级说明】

业务词汇管理过程应当包含一致性过程的步骤。从本质上看，此任务是将参考 业务词汇表的重要性渗透到开发生命周期当中。应当创建过程裁剪指南(例如，组 织可能选择豁免本地数据存储区或限定用户群的数据存储区)。

**Level 3: 定义级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.1 | 在创建共享库、建立数据传送标准(例如 XML)、构建本体、建立语义模型以及进行其 他相似的涉及组织云端数据的活动时，组织 使用经认可的业务词汇表 | ●业务词汇表使用说明  ●语义模型生命周期管理指南  ●业务词汇表审批记录 |

【成熟度等级说明】

业务词汇表明确了业务与云端数据治理过程中的相关词汇定义。针对某一云服 务业务制订的词汇表中往往包含与其他类似业务相互重叠或冲突的词汇和定义，以 帮助识别所有的数据管理活动，更好地理解组织层面上的词汇以及业务目标。因此， 对于共享库、语义模型等建立时应当充分考虑经认可的词汇解释，避免因为理解不 一致带来的错误，这些错误产生于海量数据交互、应用和分析的过程中。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.2 | 实施并跟踪组织级的云端数据治理，保证 其与业务词汇表保持一致性 | ●云端数据开发利用程序 |

 **■云端数据治理**

【成熟度等级说明】

尽管组织可能针对某个业务单元、主题领域或主要项目进行了该实践的标准化， 但只有当该实践在合适的阶段都得以实施时，遵守该实践的完全价值才能显现。要 求利用资源对开发生命周期进行一致性监督。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.3 | 通过确定业务词汇、属性、物理数据的元 素名或者近义词之间的关系，组织运用机制 来推动词汇的转换 | ●业务词汇表管理程序  ●业务词汇对照表  ●业务词汇转换指南 |

【成熟度等级说明】

一般来说，业务词汇的转换在数据治理的不同阶段通过使用一些支持技术来实 现(例如元数据库或者客户数据库)。这些支持技术的最大作用是帮助相关者更加简 单地搜索查看业务词汇表的内容。对于业务词汇的转换，组织需要形成相应的机制， 来确保转换的一致性和及时性。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.4 | 在变更业务词汇前执行，影响评估并获得 治理认可 | ●业务场景描述  ●业务词汇表变更控制程序  ●业务影响分析报告及建议 |

【成熟度等级说明】

云端数据治理中使用业务词汇表的过程中，针对某一云服务业务制订的词汇表 中往往包含与其他类似业务相互重叠或冲突的词汇和定义，以帮助识别所有的数据 管理活动，更好地理解组织层面上的词汇以及业务目标。因此，对于业务词汇变更 须通过影响评估分析来预防可能产生的业务影响，并在统一的治理体系下实现这一 过程。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.5 | 获取并使用度量数据来评估组织建立完整 业务词汇表的进度 | ●业务词汇表完整度控制程序 |

【成熟度等级说明】

度量不仅仅能估算目标和进度的完成情况，还能使参与者保持热情，持续地进 行与词汇表相关的工作。度量的内容和范围应以云端数据治理涉及的业务范围作为 主要的对象，开展有效的度量。度量指标项应经相关业务部门主管评审，并与部门/ 业务绩效相关联，以确保度量工作开展的有效性。

**第5章■云端数据管理**

|  |  |  |
| --- | --- | --- |
| 序号 | 等级要求 | 实例 |
| 3.6 | 使用一致性监督过程来确认业务词汇的正 确使用，发现例外情况，并确保例外情况得 到处理 | ●业务词汇一致性监督控制程序  ●业务词汇表管理程序  ●业务影响分析报告及建议 |

【成熟度等级说明】

根据一致性报告制订针对不符合要求词汇以及例外词汇的修订方案。修订方案 的执行日期可能与预定的发布日期相同。针对一些对组织至关重要的业务词汇(例 如共享库、重要标准报告、管理要求等),可以确切地策划发布以解决问题。所有相 关者都要收到例外报告，包括相应的影响分析以及修订活动。

**Level** **4:** **量化级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 4.1 | 使用数据或其他定量技术来管理过程，并 向高层报告并预测业务词汇的集成情况 | ●业务词汇集成度量化评价程序  ●业务词汇集成度报告 |

【成熟度等级说明】

通过线性规划、非线性规划、重置问题模型、等候线问题模型、马尔可夫过程、 存量控制、网络分析和博弈问题模型等定量决策技术，对业务词汇表这一过程域进 行量化管理，以实现对于业务词汇管理的绩效量化体现和针对性改进措施的制订。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 4.2 | 将业务词汇表集成到组织的元数据库中 并设定适当的访问权限 | ●元数据库管理程序  ●元数据库访问控制程序 |

【成熟度等级说明】

元数据库中的大部分定义与业务有直接关系，因此需将业务词汇表与组织的元 数据库紧密结合，形成有效的定义/解释传递。考虑到业务词汇表是云端数据治理过 程中与业务直接对接的第一层，其定义的准确和有效往往在一定层度上决定了数据 治理的有效性，因此对于其访问权限(特别是删除和修改的权限)需要进行控制。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 4.3 | 业务词汇表适当使用标准行业业务词汇和 定义 | ●XX行业词汇与定义  ●业务场景描述 |

【成熟度等级说明】

考虑到云端数据的跨边界、跨业务、跨访问区域(不同租户)的交换、共享等 需求，对于业务词汇的定义可参考行业标准。但直接采用行业标准词汇和定义时， 须先明确其业务场景的一致性。

 **■云端数据治理**

**Level** **5:** **优化级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 5.1 | 完善业务词汇表，使其包含相关的业务规 则、本体结构，并在组织内部保持一致 | ●业务规则说明  ●业务词汇表管理程序  ●业务及词汇表一致性评审程序 |

【成熟度等级说明】

本体定义了组成主题领域的词汇表的基本词汇及其关系，并结合这些词汇和关 系来定义词汇表外延的规则。根据业务的变化及云端数据快速变化的场景，对于业 务词汇表中的内容须及时更新和校准，实现其与业务的准确对应，并至少应在组织 内部各部门、业务、岗位之间保持一致。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 5.2 | 采用优化技术来提高使用业务词汇表推动 分类法、开发本体或使用语义表征的过程 | ●业务词汇表使用说明  ●组织知识系统优化方案 |

【成熟度等级说明】

知识组织系统是对人类知识结构进行表达和有组织地阐述的各种语义工具 (Semantic Tools) 的统称，包括分类法、语义表征、概念本体以及其他情报检索语言 与标引语言。

在传统的知识组织体系里， 一个类目、或者一个主题词被认为是一个概念，而 在新型体系中，概念含义以及相关关系的确定有了更为严格和规范的要求，原先的 类目和主题词很可能不再被当成一个概念来对待，而成为表示概念含义的一个属性 或注释等。概念、概念与概念之间的关系以及不同关系类型的概念之间的关系成为 核心内容。

通过使用业务词汇表统一词汇与定义，来开展特征优化、粒度和命名优化、结 构优化等环节从而推动分类法、开发本体或使用语义表征的过程。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 5.3 | 组织发布有关业务词汇有效管理方法的白 皮书及案例研究 | ●业务词汇管理白皮书  ●业务词汇管理最佳实践  ●知识库管理程序 |

【成熟度等级说明】

根据组织不断优化业务词汇管理的经验和积累，总结提炼管理方法白皮书及案 例，形成知识库，并不断更新内容。编制白皮书及案例研究时可参考行业标准或最 佳实践方案，以及对历史数据和相关公开信息的整理，运用案例研究与数据统计分 析方法，编制形成科学并有指导意义的案例集合白皮书。编制方法可形成指南，用 于指导白皮书和案例集的不断完善、改进。

**第5章■云端数据管理**

**4.** **实施案例**

(1)文件名称列表

《业务词汇表》

(2)实施案例

|  |
| --- |
| **业务词汇表**  1 简 介  [业务词汇表的简介应提供整个文档的概述。请在本节提供便于读者理解此文档所需的任何信息。此文档用于定  义问题领域的专用词汇，并解释“用例说明”或其他项目文档的读者可能尚不熟悉的词汇。此文档通常可用作非  正式的数据字典来收集数据定义，以便让“用例说明”和其他项目文档重点阐述系统利用这些数据信息需要执行 的操作。此文档应保存在名为“业务词汇表”的文件中。]  1)目的  [阐明该业务词汇表的目的。]  2 ) 范 围  [简要说明该业务词汇表的范围：它的相关项目，以及受到此文档影响的任何其他事物。]  3)参考资料  [本小节应完整列出此业务词汇表中其他部分所引用的任何文档。每个文档应标有标题、报告号(如果适用)、 日期和出版单位。列出可从中获取这些参考资料的来源。这些信息可以通过引用附录或其他文档来提供。]  4 ) 概 述  [本小节应说明该业务词汇表中其他部分所包含的内容，并解释此文档的组织方式。]  2 定 义  此处定义的词汇形成了此文档的基础。本文档按字母顺序定义最便于查找。  1)需求管理  我们所指对需求过程和需求内容进行管理，是把两者有机结合起来管理，而不仅仅是管理需求内容。  2)需求修订  需求未确定基线前发生了变化，我们进行修订；这种变化是不需要控制的。  3)需求变更  确定基线后，需求发生变动；这种变动是需要通过一定程序进行控制的。需求变更包含三种情况：需求改变(原 来需求发生变化)、新需求(原来不存在的需求)、 需求建议(对原来需求提出更好的看法)。 |

**5.2.2** **元数据管理**

**1.** **简介**

1)目的

元数据管理作为组织数据资产管理中的一项核心技术，也将为主数据管理提供 有力支撑，具体表现在以下三方面。

(1)为业务实体的定义、关系和业务规则的IT 实现提供清晰、标准的语义转换， 提高业务和IT 之间的一致性，保障IT 系统能够真实反映业务事实；

**■云端数据治理■**

(2)为主数据管理的各个数据处理阶段提供数据标准、数据映射关系和数据规 则的描述，保障主数据管理的数据质量；

(3)为主数据库和其他系统之间的数据交互提供有关数据标准、数据映射关系 和数据规则的描述。

2)介绍

元数据就是关于数据的数据或者用来描述数据的数据。Forrester Research 将元 数据定义为“用于描述数据、内容、业务流程、服务、业务规则以及组织信息系统 的支持政策或为其提供上下文的信息”。元数据是识别、描述、解释以及提供与组织 数据资产相关的内容、语境、结果以及分类的信息，元数据保证这些资产的检索、

使用以及管理得到有效执行。生活中我们填写的《个人信息登记表》中的各项内容， 包括姓名、性别、民族、政治面貌、 一寸照片、学历、职称等，这些就是锁定一个 人的元数据。

数据治理成功的关键在于元数据管理，即赋予数据上下文和含义的参考框架。 经过有效治理的元数据可提供数据流视图、影响分析的执行能力、业务词汇表以及 其词汇和定义的可问责性，最终提供用于满足合规性的审计跟踪。元数据管理成为 一项重要功能，让IT 部门得以监视复杂数据集成环境中的变化，同时交付可信、安 全的数据。因此，良好的元数据管理工具在全局数据治理中起到了核心作用。

通过开展元数据管理，可以使管理下的结构化和非结构化数据资产信息清晰、 有条理，并得以明确和扩充；支持组织开展数据分享，确保数据使用的一致性；提 高对业务变更的反应能力；减少数据相关的风险。有效的元数据管理以及组织元数 据目录的创建能够支持并推动关键数据管理活动的执行以及目标的达成。

元数据至少包含三类，业务元数据、技术元数据和操作元数据。

(1)业务元数据：用于理解、搜索、定位并控制包含了词汇、定义、值、作者、 关键词以及发布者等内容的描述性信息。如业务词汇名称、定义、责任人或管理员 以及相关的参考数据。

(2)技术元数据：描述物理数据层的数据资产实例以及他们通过自动化过程所 进行的转换，还描述了数据存储区以及数据界面的内容和位置，包括表格、场结构、 列、与相关文件的联系、指数等信息。如数据源表格名称、数据源表格栏目名称及 数据类型(如字符串、整数等)。

(3)操作元数据：提供管理信息以管理数据资产，管理信息包括创建时间、文 件类型、数据用途、存档、集成以及日程更新所需信息、访问权限以及资格限制等。 如最近更新的数据、访问次数或最后访问的数据。

3)相关过程域

●数据生命周期管理过程域提供更多相互依赖的相关信息。

●数据提供者管理过程域包含了开发数据需求以及相应的元数据信息。

●架构方法过程域提供了与数据平台架构相关的活动的信息。

**第5章■云端数据管理**

●业务词汇表过程域包含了开发经认可的组织级共享词汇表以及对应元数据 的相关信息。

●数据集成与互操作过程域包含了生成以及管理数据映射和转换元数据的相 关信息。

**2.GQM 表**

|  |  |  |  |
| --- | --- | --- | --- |
| 目 标 | 问 题 | 度 量 项 | 度量项说明 |
| G1:元数据的 表达使用高管能 理解的语言，因 此他们能够理解 元数据的价值 | Q1:是否所有内部相关 者都参与定义元数据分类  及属性的过程 | M1:元数据管理策略 | 要形成元数据管理策略并  在内部发布，数据治理对于元  数据的使用和管理过程中所  遇到的问题，应作为反馈用于  优化元数据管理策略，维护过  程应引入内部相关者的意见 |
| G2:数据治理 监督指导元数据 策略、分类和标 准的制订与执  行，保证它们被 整个组织采纳并 一致使用 | Q2:使用什么方法制订 和评估元数据管理的标准 和过程 | M2:元数据类型和数据定义 | 在定义元数据类型时应当  考虑输入、查询和使用等活动 的要求 |
| M3:变更管理 | 通过预测模型对于元数据  管理过程的变更进行预测，以  降低变更带来的不可控影响 |
| Q3:使用什么方法维护 和更新元数据库 | M4:元数据完整性 | 元数据库的增强取决于业  务重点、顺序计划以及可用的 资源。初始元模型通过模型变 更得以增强和扩展，反映了不 断提升的完整性 |
| M5:元数据库变异分析 | 对于发生的元数据库信息  及描述的变异，开展根本原因 分析，用结构化的问题处理法  逐步找出云端元数据管理过  程中出现问题的根本原因并 加以解决，避免仅仅关注问题 的表象，同时避免问题的重复 出现 |
| M6:变更分析 | 数据变更对元数据库产生  的影响需要进行提前评估，并  制订规范的评估计划来保证 评估结果的有效性和价值 |

 **■云端数据治理**

**续表**

|  |  |  |  |
| --- | --- | --- | --- |
| 目 标 | 问 题 | 度 量 项 | 度量项说明 |
| G2:数据治理  监督指导元数据 策略、分类和标 准的制订与执 行，保证它们被 整个组织采纳并 一致使用 | Q3:使用什么方法维护 和更新元数据库 | M7:反馈机制 | 记录元数据管理的过程，将 数据输入、查阅以及使用的统 一实践文档化，对交换数据所 遵循的标准进行文档化，形成 整合记录，并保存在元数据库 中。利用这些记录和分析结 果，在新的元数据管理过程中 支持管理决策 |
| Q4:元数据管理过程是 否得到定义与执行 | M8:策略一致性 | 针对元数据的主要分级(例 如业务级、技术级),组织应 提出一系列分类(例如业务类 词汇、数据库) |
| M9:管理过程标准化 | 在元数据的生命周期中(即 创建、更新以及删除)定义一 个标准管理过程 |
| M10:变更控制 | 当纳入一类元数据到库中  时，所有范围、类别、属性以 及标准上的增添或变更都要 由数据治理来管理 |
| 基于现有的成熟架构，对元 数据的创建以及变更进行确认 |
| 利用元数据来对数据变更 做影响分析 |
| M11:目标一致性 | 元数据的改进目标应当与 组织的业务和数据治理策略  相关联 |
| M12:元数据管理历史信息 的使用 | 元数据矩阵上各元素间的  关系通过数据定量化表示，可 更准确地整理和分析元数据  管理 |
| M13:量化目标 | 建立量化管理的规范和制 度，对元数据的管理过程进行 量化考核和评价，细化管理过 程的改进，不断提升元数据管 理能力和水平，促进云端数据 治理的绩效 |
| Q5:元数据获取、更新 及使用过程中的分工与义 务是否得到清晰的定义 | M14:职责与分工 | 将数据管理功能进行分配， 以推动人员间的共同合作努 力；整个组织的代表人员参 与，执行审批并确保一致意见 的达成 |

**第5章■云端数据管理**

**续表**

|  |  |  |  |
| --- | --- | --- | --- |
| 目 标 | 问 题 | 度 量 项 | 度量项说明 |
| G3:元数据库 的内容涵盖管理 下的数据资产的 所有相关类别与 级别，准确反映 组织已实施的数 据层 | Q6:被纳入元数据库的 元数据范围是如何定义的 | M15:元数据文档 | 组织应识别现存元数据的  来源，评估他们的完整性、类 别以及属性，将其完善、整合  到一个反映它们需求的聚合 元模型中 |
| M16:元数据文档记录了数 据关联 | 对业务元数据、技术元数据 以及操作元数据中的关键数  据集进行确认和相互关联 |
| M17:元数据管理模型完整 性 | 对于云端数据涉及的内外  部所有平台，需要采用统一的 模型来规范和管理元数据 |
| G4:将对组织 重要的内部标准 以及经选择的外 部标准纳入元数 据中，并与组织 的过程和标准保 持一致 | Q7:元数据策略是否已 确定?与内部和经选择的 外部标准是否一致 | M18:改进计划 | 应在元数据策略中定义关  键的度量和要达成的目标，这  些可以随着元数据资产的延 伸和扩展得到进一步完善 |

**3.** **成熟度等级介绍**

**Level 1: 初始级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 1.1 | 生成、储存可获取的元数据文档 | ●云端数据资产清单  ●业务词汇表  ●元数据库  ●文档化的元数据列表 |

【成熟度等级说明】

组织通常会在多个位置根据数据的不同用途对数据资产信息进行维护。可在数 据模型、项目文档和操作文档或业务词汇表中获取、储存元数据。这些来源组合在 一起构成了虚拟的元数据库。组织应识别现存元数据的来源，评估他们的完整性、 类别以及属性，将其完善、整合到一个反映它们需求的聚合元模型中。

基于数据模型的来源，其作为可使用的现存元数据资产的例子，通常包含以下 内容。

●实体类型名称

●属性名称

 **■云端数据治理**

●表格名称 ● 列名

●数据类型 ●长度

●容许值

●默认值

●指定/可选指示器

●实体类型、属性、表格、列的数据定义

**Level 2: 管理级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.1 | 建立并执行元数据管理过程 | ●元数据生命周期管理程序  ●元数据角色和分工  ●元数据规划程序 |

【成熟度等级说明】

在元数据的生命周期中(即创建、更新以及删除)定义一个标准过程能使组织 获益，该过程用于执行主要的步骤，通常涉及以下内容。

● 识别内部相关者及他们的角色；

●经业务批准的数据概念的定义；

●决定所需的元数据组件及类别；

●选择、建立公共库以进行数据的储存、维护及检索；

●配置管理和维护的规则与标准。

业务相关者的参与能保证元数据清晰地描述终端用户所需的信息并且支持数据 生命周期中关键过程的实施，例如：

●数据搜索

●数据迁移

●目标设定与分类

●使用(例如，供汇报使用或在系统开发生命周期中使用)

●治理与控制

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.2 | 元数据文档记录了数据间的相互依赖关系 | ●元数据关系图  ●云端数据沿袭记录/日志 |

【成熟度等级说明】

组织将业务元数据、技术元数据以及操作元数据中的关键数据集进行确认是十 分重要的，这使组织能够从数据源追溯到目标库。

**第5章■云端数据管理**

组织通常缺少足够的元数据来建立数据沿袭，数据沿袭通常包括以下内容。 ●数据源

●权威来源或首选来源的指定

● 已应用的转换或聚合

●数据目的地

●数据元素层面的集成映射，反映了来源在集成时是如何结合在一起的

●新数据的接收时间

● 数据的变更时间等

例如，当监管组织收到一份报告时，组织需要在多个过程的所有来源中，通过 对转换、计算或聚合进行互相衔接的检查来确定数据的起源。这提供了完整的审计 跟踪，跟踪信息包括数据的内容、数据是如何变更的以及数据是如何在报告中使用 的。数据拼凑在一起供特定用途使用是一个十分耗时的过程，这个过程会在组织的 多种情况下重复进行。而这种稳定的相互衔接的元数据则简化了这个过程并推动了 其自动化的进程。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.3 | 开发并使用元数据来对可能的数据变更进 行影响分析 | ●云端数据变更影响分析报告  ●元数据应用说明 |

【成熟度等级说明】

评估数据变更的影响需要先了解数据在组织里的使用范围(例如，由相关者使 用、在应用数据的存储中使用、在数据仓库环境中使用等)。除了描述数据内容的元 数据外，还须记录与数据库中每个数据集相关联的系统和相关者(生成、维护和使 用环节)。初始范围可能由数据集决定，但最终的目标是在属性和列的层面上达到映 射。例如，当业务在高度分享的类型代码(例如组织类型)中添加了一个新值时， 多个数据存储都需要更改他们的结构。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.4 | 建立并遵守元数据的分类、属性以及标准 | ●元数据管理程序  ●元数据分类分级指南 |

【成熟度等级说明】

针对元数据的主要分级(例如业务级、技术级),组织应提出并同意一系列分类 (例如业务类词汇、数据库)以及针对每个分类的对应属性(例如数据管理员、更新 日期)。应明确说明每个元数据的元素表示惯例。应根据这些决策编写元数据标准， 并对其进行发布、实施和监督。

 **■云端数据治理**

**Level 3: 定义级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.1 | 建立和发布组织的元数据管理策略，该策 略由数据治理进行维护，内部相关者参与相 关过程 | ●元数据管理策略  ●元数据管理目标  ●元数据角色和分工  ●元数据管理控制程序 |

【成熟度等级说明】

一个全面、完整、有序的元数据库是时刻在完善的。与向组织建议的数据质量 计划相似，我们建议组织制订策略，在自上而下的愿景、战略目标和战术目标的实 现中进行协作，并且为需求、设计和开发(或平台开发)制订步骤，这其中包含了 一个调解依赖关系以及解决顺序冲突的阶段执行计划。

元数据管理策略应解决以下问题。

●组织在元数据管理上的共同愿景；

●相关者的角色和分工；

●元数据特性(范围、类别、属性、使用等);

●元数据资产的概念元模型；

●元数据创建和变更中数据治理的角色和职责；

●元数据管理中的一致性计划；

●指导实施的顺序计划。

随着系统的更加完整以及耦合，并且包含了外部系统(在这些外部系统中，依 赖取决于合同),元数据策略应论述所有的数据源以及相关的交换标准。适当地采用 外部公认的行业标准能降低内部的差异性并且增加组织与顾客、客户、同行与管理 部门间的互操作性。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.2 | 组织的元数据库根据阶段实施计划添加了 额外的元数据类别与级别，并与架构层相互 联系 | ●元数据变更管理程序  ●元数据库 |

【成熟度等级说明】

根据需求范围的不同，元数据库可能是由内部构建的，也可能是向第三方供应 商购买的。在实际情况中，人们更倾向于选择购买而非内部构建元数据库，因为功 能开发的成本比功能购买的成本要高。元数据库的增强取决于业务重点、顺序计划 以及可用的资源。初始元模型通过模型变更得以增强和扩展，反映了元数据的完整 性在不断提升。以下元素通常构成了元数据的扩充和扩展。

●数据质量规则，包括与来源真实性以及出处相关的规则；

●业务规则；

**第5章■云端数据管理**

●扩展的分级(由组织添加);

●组织级分级和业务领域分级的差别；

●权威数据来源(可信来源);

●为业务词汇分配数据管理员；

●为物理数据对象分配数据监管员；

●包含转换逻辑的过程元数据(例如，提取转换加载);

●为静止数据构建以下架构层的联接；

●业务词汇；

●组织逻辑数据模型；

●逻辑数据模型；

●物理数据模型；

●数据库方案；

● BI层；

●物理数据沿袭(来源—转换—目标);

● 元数据历史和版本管理(以进行审计和变更控制)。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.3 | 元数据管理功能以数据管理工作为中心， 由数据治理进行监督 | ●元数据分类分级控制程序  ●元数据管理程序  ●元数据角色和分工 |

【成熟度等级说明】

元数据相关工作包含分类、属性确定、标准制订以及组织级的实施。

元数据因其所涉及的问题范围更广、更具技术性而不同于普通数据，并且涵盖 了一些与普通数据生命周期过程无关的内容。元数据更多地受项目的影响，而从数 据治理的角度上看，元数据的范围与深度使其治理具备很大难度。尽管元数据的建 立和元数据的维护具备重要的差异性，能够将基于项目的元数据纳入整体数据治理 活动中也还是非常重要的，特别是对主要的数据管理、数据仓储、元数据库实施活 动至关重要。

由于元数据在业务与技术环境中的应用范围广、程度深，角色与职责会涉及到 不同领域的许多职员。通常来说，最好是将数据管理功能进行分配，以推动人员间 的共同合作努力，另外，由于需要相关人员投入时间进行工作，因此要将数据管理 功能分配情况完全告知相关者并获取他们的同意。数据过程由于有整个组织的代表 人员参与，因此能够执行最后的审批并确保一致意见的达成。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.4 | 数据治理审批元数据的增添以及更改 | ●元数据变更管理程序  ●变更管理的角色和职责 |

**■云端数据治理**

【成熟度等级说明】

当纳入一类元数据到库中时，所有范围、类别、属性以及标准上的增添或变更 都要由数据治理来管理，并且在组织层面上进行沟通。通过制订规范的流程，对不 同业务阶段的元数据实行变更管理，保障设计、开发、实际上线等阶段变更的一致 性，来推进变更管理、IT 流程管理规范化。

以下是参考流程：

业务系统发生需求变更或功能模块升级导致表结构变更时，需要在元数据管理 系统中发起申请流程，提交设计态元数据进行备案。功能开发完成后，部署到测试 环境，由待发布环境元数据系统获取测试态元数据，将元数据变更在生产环境部署 前与设计态元数据对比，比对一致上线部署，不一致驳回上线申请流程。元数据变 更在投产上线完成后，生产环境元数据系统会采集生产态元数据，并再次与设计态 元数据进行对比，实现对元数据变更的事后核查。通过三个不同形态元数据的比对 分析，达到事前风险预警+事后变更核实的管理效果。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.5 | 使用措施和度量来评估元数据的准确度并 采纳 | ●元数据策略及目标  ●元数据准确度评估程序  ●元数据应用情况分析报告 |

【成熟度等级说明】

应在元数据策略中定义关键的度量和要达成的目标。这些可以随着元数据资产 的延伸和扩展得到进一步完善。

度量定义了需要被跟踪的内容(并从根本原因上解释了其重要性),有以下例子：

●元数据管理的相关值(成本控制、工作效率、过程有效性);

●关键过程和流程的性能，例如每个特性的使用频率、每个特性储存的数据量 以及每个特性的应用数量；

●应用数据特性的临界(哪些是核心、针对什么过程、在哪些应用中使用、包 含在哪些计算过程中);

●生命周期中数据迁移的相关成本(有多少存在风险，特别是当数据沿袭断裂 时);

●对单一权威来源的跟踪过程；

●元数据广度、深度、范围、可用性、适时性、准确性、复制、符合性、联接 和清晰度的质量。

**第5章■云端数据管理**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.6 | 根据现有架构确认元数据以及元数据进行 的变更 | ●元数据管理策略  ●元数据角色和分工  ●元数据扩展的库报告  ●元数据管理组织标准  ●元数据元模型图  ●平台与元数据之间的差距分析结果  ●元数据度量报告(采纳、完成比等)  ●元数据顺序计划的里程碑以及进度报告 |

【成熟度等级说明】

云端往往通过双向的元数据架构实现其元数据管理，需要让信息包含于一个可 以转化的存储中，确保这些信息随后可以提供其初始来源。基于现有的成熟架构， 对元数据的创建以及变更进行确认。例如，应当确认关于物理模型、接口规范、提 取转换、加载脚本、数据提供服务等的信息来确定他们准确描述了现有的环境。

**Level** **4:** **量化级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 4.1 | 组织开发了部署在所有平台的完整元模型 | ●元数据模型 ●元模型  ●元模型管理规则 |

【成熟度等级说明】

模型是用来描述特定的系统、过程、事物或概念的准确而抽象的表示。例如软 件架构师可以用概要设计的形式建立一个应用系统模型。本质上来说，元数据是数 据的形式化模型，是数据的抽象描述，这一描述准确地描述了数据。元模型也就是 模型的模型，是用来描述元数据的模型。因此，对于云端数据涉及的内外部所有平 台，需要采用统一的模型来规范和管理元数据，完整的元模型则提供了很好的基础。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 4.2 | 元数据类型和数据定义支持输入、查阅和 使用活动 | ●元数据管理程序  ●元数据类型定义  ●元数据操作指南  ●业务词汇表 |

【成熟度等级说明】

通过构建元数据类型定义为内部团队、合作伙伴和客户创建自己的元数据管理 模式。对于元数据的数据查阅和使用等活动，需要首先进行标准化的元数据类型定 义，以使得用户对元数据的使用顺畅。

**■云端数据治理**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 4.3 | 根据元数据矩阵和元数据的相关历史信息 指导元数据管理活动 | ●元数据管理程序  ●元数据库矩阵  ●元数据操作日志 |

【成熟度等级说明】

元数据矩阵上各元素间的关系通过数据定量化表示，可帮助组织更准确地整理 和分析元数据管理。结合对于元数据使用和变更过程中的历史数据，加强对元数据 活动的管理。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 4.4 | 制订了量化目标，并能够指导元数据的管 理并支持过程的执行 | ●元数据管理程序  ●元数据管理工具  ●元数据管理过程量化评价指南 |

【成熟度等级说明】

通过利用 SMART 原则，建立量化管理的规范和制度，对元数据的管理过程进 行量化考核和评价，细化管理过程的改进，不断提升元数据管理能力和水平，促进 云端数据治理的绩效。量化管理工作重视结果，强调自主、自治和自觉。形成量化 目标体系后需要进行定期检查，利用有效信息反馈渠道进行定期评价；同时要向被 考核对象通报，结合考核情况调整元数据管理工作，当出现意外、不可测事件导致 严重影响管理目标的实现时，也可以通过一定的手续，修改原定的目标。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 4.5 | 元数据库包含了过程、汇报和执行的数据 分析报告，使用这些报告制订与新的元数据 管理活动相关的决策，且决策是基于事实制 订的 | ●已记录的元数据量化目标  ●度量方法的文档化，包括使用的数据和其他量化 技巧，以及性能阈值  ●全面的元数据库报告包括历史报告  ●元数据管理的过程效率报告  ●统一或集成的元模型 |

【成熟度等级说明】

记录在元数据管理的过程、分析和报告，将数据输入、查阅以及使用的统一实 践文档化，对交换数据的表示所遵循的标准进行文档化，形成整合记录，并保存在 元数据库中。利用这些记录和分析结果，在新的元数据管理过程中支持管理决策。

**Level 5: 优化级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 5.1 | 执行根本原因分析以降低库信息和它所描 述信息的变异 | ●元数据库信息变异根本原因分析措施  ●基于标准定义的元数据管理持续性报告和证据 |

**第5章■云端数据管理**

【成熟度等级说明】

对于发生的元数据库信息及描述的变异(由误差、错误、篡改等导致),通过开 展根本原因分析，用结构化的问题处理法逐步找出云端元数据管理过程中出现问题 的根本原因并加以解决，避免仅仅关注问题的表象而不关注问题的本质，同时避免 问题的重复出现。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 5.2 | 使用预测模型指导元数据管理过程的变更 | ●元数据变更管理程序  ●预测模型文档 |

【成熟度等级说明】

元数据管理过程产生变更的同时会带来业务的影响。通过采用预测模型可在变 更前分析其影响并支撑决策。例如：使用决策树、支持向量机、粗糙集、模糊聚类、 基于范例推理、统计方法、神经计算等多种数据挖掘算法，支持特征抽取、分类、 聚类、预测、关联规则发现、统计分析等数据挖掘功能，并支持高层次的决策分析 功能。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 5.3 | 根据元数据策略确定量化元数据的改进目 标 | ●元数据策略管理程序  ●元数据管理方针、目标 |

【成熟度等级说明】

元数据的改进目标应当与组织的业务和数据治理策略相关联，元数据策略很好 地体现了相关的内容。量化管理能从更细化的层面描绘元数据管理过程中需要或值 得改进的方面以及所要达到的程度，便于确定改进的目标和方向。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 5.4 | 评估计划的数据变更对元数据库的影响； 元数据的获取、变更以及精简过程得到持续 改进 | ●元数据库变更计划  ●元数据库管理程序  ●数据变更影响分析 |

【成熟度等级说明】

数据变更对元数据库产生的影响需要提前进行评估，并制订规范的评估计划来 保证评估结果的有效性和价值。对于这一过程，通过PDCA 管理方法来不断提升其 效果，以保证改进的有效性。

**4.** **实施案例**

(1)文件名称列表

《元数据管理最佳实践》

 **■云端数据治理**

(2)实施案例

|  |
| --- |
| 元数据管理最佳实践  1 显示成效，小处着手，真抓实干。如能显示成效，元数据管理项目就成功了一半。选择一个有把握取得成功 的小型项目并迅速展现确凿成果。然后以此为开端拓展项目规模。  2全面量化。随时做好对您取得的成果进行量化的准备。很难说什么时候您会被要求为元数据项目提供一个合 理的存在原因。由于最终体现出来的效益(包括基于良好数据而实现的出色业务决策、高效的业务流程以及得到 优化的利益相关方交互活动等)均看上去与IT部门的元数据项目相距甚远，因此往往难以对其成果进行量化并确 认可通过您的举措获益的几项关键业务流程、决策和交互活动，据此应限制项目范围并仅对此范围内的项目进行 量化，同时为所有其他项目构建可量化的业务案例。例如：由于在投资模型中使用通过您的数据仓库提供的坏数 据而导致不良投资决策的代价是什么?  3设立一个有重点而又合理的项目范围。许多元数据和数据治理方案的失败原因是好大喜功又急于求成。范围  渐变和丧失分析能力使得项目实施人员很快就穷于应付，并最终在无法取得任何有意义的结果时使项目遭到撤消。 没有谁愿意把时间或金钱浪费在漫无止境却不能带来任何业务成果的会面和会谈上。  4争取高层人员的支持。如果缺乏适当的高层支持，您的项目就无法成功。切勿低估从业务部门和IT部门获  得适当支持的重要性。多位客户告诉我们，这是他们在初次失败之后学到的教训。此外，向其他团体寻求对项目 的支持也是必不可少的。  5筹建数据治理方案。对于数据治理方案，组织必须认识到，元数据不仅是一项IT生产力工具，它还可以在  使用过程中给整个组织带来更广泛的利益以及更可观的价值。早期的目标之一应是组建一个数据治理执行委员会 或指导委员会，协助为数据驱动型机遇设立长远的整体目标和优先考虑事项。此外还需要一个用于业务与业务或 业务与IT协作的整体框架来确保项目成功。  6选择高价值、低复杂度的目标作为试点。从数据容量、机构或体系架构复杂性等角度出发，选择一个可取得 一定实际回报，但无须兴师动众就能解决的特定问题，然后致力于在进军其他领域之前解决该问题。例如，针对 新的数据仓库或主数据管理实施提供完整的数据沿袭和全套业务词汇及定义。重点在于选择一个对总体业务具有 高价值和战略重要性的项目，却又不会在逐渐完善数据治理流程时，被其复杂性压垮。打赢头一仗是关键所在。 (同时请参见第2条)。  7分配责任人并让业务用户参与进来。没有业务上下文的数据毫无价值。包含NW\_Net\_Revenue列的表格没有 具体含义，除非与诸如以下对象挂钩。  (1)业务词汇  (2)词汇定义  (3)词汇责任人  (3)参考数据  (4)其他记录和评注  一旦有了此类业务上下文，您就能将业务词汇及定义与技术元数据相关联，为业务部门和IT部门创建通用语言，  从而改善其交流与协作。其中要素之一是确保为数据和元数据分配责任人，可问责性是关键。  8奖励与惩罚并用。考虑可提供什么样的刺激奖励，例如  (1)更有效、更轻松地访问业务词汇定义，鼓励业务用户参与内容创建，因为他们可从成果中受益； (2)投稿比赛，对贡献新内容者给予奖励；  (3)定义内容(利用博弈技术),鼓励管理员之间的良性竞争，激励他们创建优秀内容；  (4)公开表扬，可通过对提供最多新颖且优秀内容的贡献者进行简单地评级或表彰。  另一方面，惩罚也必不可少。拥有高级管理层和强大的数据治理委员会的支持对于确保各方支持和接纳助益匪 浅。但双管齐下也很重要。任何惩罚措施都可能因为过于严厉而造成口服心不服，而奖励措施则可能因为可有可 无、好处有限而难以激发起参与者的热情。 |

**第5章·云端数据管理**

**续表**

|  |
| --- |
| 难题在于怎样让业务部门乐于参与。除极少数情况外，没有全职的业务人员负责提供这些上下文资料，他们都 有其他职责，对他们来说这仅仅属于“副业”。问题是：这对他们有什么好处?您将必须提供激励措施。总体来 讲，最好的做法是确保将参与情况在员工绩效管理目标中正式记录备案。这既能鼓励他们积极参与，又能防止他 们在出现不可避免的紧急情况时分心。  9将元数据管理与业务方案绑定。两家客户称，它们各自尝试实施数据治理方案，但均以失败告终。这两家组  织采用的新办法不再是从上而下的方案，而是当重要的新项目涉及IT部门时，将元数据管理和数据管理员最佳实 践与业务方案挂钩。该方法非常实用，旨在取得成功且成功率极高。同时，快速显示可衡量成果的可能性也变大。  10寻找潜在的数据危机，并在不可避免时做好准备。如果组织管理层就是不同意为元数据管理或数据管理方  案拨款该怎么办?一家金融服务公司高级经理的建议是：“只须寻找数据危机。”并随时准备提交应对方案。危 机迟早会发生，重要的是您已为此做好了准备，并可呈交如何在未来防止此类危机再次发生的方案。 |

**5.2.3** **主数据管理**

**1.** **简介**

1)目的

主数据关注的重点是被描述的数据对象反映了业务中的哪个实体，是否真实的 反映了该实体。主数据管理是解决以上问题的有效手段，其价值主要体现在：

(1)消除有关业务实体定义、格式、数据标准的冲突和不一致性；

(2)根据业务需求，提供统一的、标准的业务实体的数据标准以及真实的数据；

(3)统一标准的数据和信息可供重复调用，进一步保障各系统中主数据来源的 一致性，并提高系统间的“沟通效率”。此时，主数据已经变为一种规范化、可重用 的“信息服务”并按需提供给相关系统。

2)介绍

主数据是指在整个信息供应链中各个业务系统之间都需要共享的数据、业务规 则和策略等。常见的主数据例如：对于电信行业客户而言，电信运营商提供的各种 服务可以形成其产品主数据；对于航空业客户而言，航线、航班是其组织主数据的 一种。对于某一个组织的不同业务部门，其主数据也不同。例如市场销售部门关心 客户信息，产品研发部门关心产品编号、产品分类等产品信息，人事部门关心员工 机构、部门层次关系等信息。

主数据管理是指一整套的用于生成和维护组织主数据的规范、技术和方案，以保 证主数据的完整性、 一致性和准确性，是构建组织信息单一视图的重要组成部分，可 以保证在整个组织范围内跨业务协调和重用主数据。描述了一组约束(规程)、方法 和技术解决方案用来保证整个信息供应链内主题域和跨主题域相关主数据的完整一 致性。主数据管理是应用流程的补充，为应用提供精确、完整的关键业务实体数据。

主数据管理不会创建新的数据或新的数据纵向结构，而是提供一种方法使组织 能够有效地管理分布在整个信息供应链中的各种主数据(由信息供应链各业务系统

**■云端数据治理**

产生)。主数据管理可以帮助组织构建并维护贯穿整个信息供应链的主数据单一视 图，提供主数据的质量管理和统一业务实体定义，简化并改进业务流程并提高业务 响应速度。

主数据管理问题的存在，是由组织业务发展的渐进性、IT 技术发展的渐进性、 业务系统自下而上而不是自上而下、缺乏统一的数据治理和元数据管理共同造成的。 正是由于这种渐进性，各组织的业务系统都大体经历了从无到有、从简单到复杂， 直到形成了一个个业务。从根本上来说，组织很难只用一个业务系统覆盖所有的业 务，特别是大型跨国组织，同一个业务系统也可能会在不同的国家或地区部署多套， 加上组织信息化建设缺少统一规划，从而造成了需要在各业务系统中共享的主数据 被分散到了各个业务系统分别进行管理。分散管理的主数据由于不具备一致性、准 确性和完整性，使得组织普遍存在着产品、供应商和订单管理不力的现象，解决这 一问题的根本方法就是引入主数据管理。

在云计算时代，通过建立云端数据与主数据之间的映射关系可以有效地提高客 户关系管理水平，提高客户满意度和忠诚度，提升销售业绩。例如通过从微博、微 信、交友网站以及呼叫中心语音记录中获取数据，进行更精确的客户流失建模，可 以有效地提升客户流失预测的准确率。再比如从社交媒体、多媒体、电话语音记录 等多种数据源获取数据用于客户细分、交叉销售、提升销售、客户维护挽留、客户 偏好管理等，都可以有效地提升客户关系管理水平。

3)相关过程域

数据生命周期管理过程域提供更多相互依赖的相关信息；

架构方法过程域提供了与数据平台架构相关的活动信息；

元数据管理过程域提供了主数据的描述信息以及定义；

数据集成与互操作过程域提供了主数据管理的基础，包含了主数据标准化的参 考信息。

**2.GQM 表**

|  |  |  |  |
| --- | --- | --- | --- |
| 目 标 | 问 题 | 度 量 项 | 度量项说明 |
| G1:主数据能 充分表达所描述 的业务对象或场 景 | Ql:主数据是否与业务 相关 | M1:业务问题和主题区域 | 需要确定与主数据管理相  关的业务问题，从而开展具有 针对性的主数据管理工作；同 时，对于主数据涉及的领域明 确定义有助于开展主数据管 理工作，并在海量云端数据环 境下有利于快速寻找与主数 据管理目标相一致的数据内 容 |

**第5章■云端数据管理**

**续表**

|  |  |  |  |
| --- | --- | --- | --- |
| 目 标 | 问 题 | 度 量 项 | 度量项说明 |
| G2:主数据策 略、分类和标准 的制订与执行， 保证它们被整个 组织采纳并一致 使用 | Q2:主数据管理过程是 否得到定义与执行 | M2:主数据标准体系 | 云端数据的体量更大、种类 更复杂，需要构建主数据标准 体系，明确数据分类、描述、 编码等标准规则，形成主数据 管理的基础 |
| M3:主数据管理过程 | 针对主数据管理需求，制订 管理规范/程序 |
| M4:历史信息指导主数据管 理活动 | 结合对于主数据使用和变  更过程的历史数据，加强对主 数据活动的管理 |
| M5:量化目标 | 建立量化管理的规范和制 度，对主数据的管理过程进行 量化考核和评价，细化管理过 程的改进，不断提升主数据管 理能力和水平，促进云端数据 治理的绩效 |
| G3:管理主数 据质量，确保主 数据的应用基于 准确的基础 | Q3:纳入管理前的主数 据是否经过分析和筛选 | M6:主数据共享 | 列表的维护(增删改和冲突 处理)由各部门通过一系列讨 论和会议进行处理，以确保主  数据的共享符合相关部门要 求和需求 |
| M7:主数据自动存储管理 | 当需要对主数据进行增删  改时，外部应用(请求发起端)  将请求更新主数据主机中的 数据，并调用数据所有者所在  的应用(通过接口)来更新对 应的数据 |
| M8:唯一标识 | 通过数据质量控制和数据  清洗等手段匹配相同源或多 个源中的重复可疑项，解决错 误和不完整的数据，将云端相  同数据的多个版本合并为一 个真实版本 |
| M9:主数据管理质量 | 组织需要高质量的主数据  提升云端数据分析；同时，也  可以利用云端海量数据提高 主数据的质量 |

 **■云端数据治理**

**续表**

|  |  |  |  |
| --- | --- | --- | --- |
| 目 标 | 问 题 | 度 量 项 | 度量项说明 |
| G3:管理主数 据质量，确保主 数据的应用基于 准确的基础 | Q3:纳入管理前的主数 据是否经过分析和筛选 | M10:统一的制订和维护标 准 | 构建通用的主数据管理基  础，使用集中和标准化的方法 转换异构数据，通过使用统一 的数据标准，解决数据不一致 的问题 |
| G4:将对组织 重要的内部标准 以及经选择的外 部标准纳入主数 据中，并与组织 的过程和标准保 持一致 | Q4:主数据管理策略是 否已确定?与内部规范和 经选择的外部要求是否一 致 | M11:主数据监管策略 | 主数据监管是个持续的过  程，组织管理者通过主数据监 管管理其主数据的质量，定义 准则、策略、流程、业务规则 以及度量值 |
| M12:变更分析 | 通过采用预测模型在变更 前分析其影响并支撑决策 |
| M13:改进目标 | 主数据的改进目标应当与  组织的业务和数据治理策略 相关联，主数据策略很好地体 现了相关的内容 |
| M14:持续改进 | 云端数据治理是一个活动  集合，覆盖了统筹和规划、构 建和运行、监督和评估、改进 和优化四大阶段，通过规范的 程序确保主数据管理的持续 改进 |

**3.** **成熟度等级介绍**

**Level 1: 初始级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 1.1 | 建立并执行主数据管理过程 | ●主数据管理程序 |

【成熟度等级说明】

组织的主数据主要指用来描述组织核心业务实体的数据，比如客户、合作伙伴、 员工、产品、物料单、账户等；它是具有高业务价值的、可以在组织内跨越各个业 务部门被重复使用的数据，并且存在于多个异构的应用系统中。需要有针对性制订 管理规范/程序，进行管理。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 1.2 | 通过手工方式维护一个逻辑或物理的列表 用来共享主数据 | ●主数据列表  ●主数据列表管理指南  ●主数据共享机制 |

**第5章■云端数据管理■**

【成熟度等级说明】

当各个系统或用户需要某些数据时可以索取该列表。列表的维护(增删改和冲 突处理)由各部门工作人员通过一系列讨论和会议进行处理。虽然各部门依然独立 维护各自关键数据，但已经开始使用列表方式维护一个松散的主数据列表，满足各 部门的主数据需求。由于缺乏集中的基于规则的主数据管理，在数据量比较小时列 表管理的方式是可行的，但当数据量较大时，数据维护的成本会很高，效率比较低。

**Level 2: 管理级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.1 | 通过引入统一存储实现主数据的自动存储 管理 | ●统一存储管理规范  ●主数据自动存储管理规范  ●接口规范 |

【成熟度等级说明】

统一存储往往也被称为主数据主机(Master Data Host), 并通过一个应用程序对 外提供数据访问功能。当需要对主数据进行增删改时，外部应用(请求发起端)将 请求更新主数据主机中的数据，并调用数据所有者所在的应用(通过接口)来更新 对应的数据。各个外部应用需要能够了解基本的业务规则(元模型)以便访问主列 表并与主列表进行交互，并且各个外部应用有责任坚持数据管理的原则和规程。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.2 | 建立并遵守主数据的分类、属性以及标准 | ●主数据分类规则  ●主数据管理程序  ●主数据标准体系  ●编码信息库  ●编码对照表 |

【成熟度等级说明】

通过主数据的标准化和信息化建设构建适用于全组织各业务模块的主数据分类 体系，以满足各业务节点对主数据分类管理以及管理精细度的需求。同时，通过逐 步标准化主数据，形成主数据管理的基础能力，有助于固化主数据管理组织和流程。

相比于传统的数据环境，云端数据的体量更大、种类更复杂。如果在分类划分 标准、特征项设置、编码规则等方面不专业、不一致、不合理，极易造成数据归类 错误，影响云端数据治理的有效性。因此，需要打造主数据标准体系，明确数据分 类、描述、编码等标准规则，形成主数据管理的基础。

可考虑从以下方面开展：

●建立标准的数据分类体系，实现分类信息的统一规范化；

●借鉴其他组织的成熟规范编码，建立适合本组织的编码信息库，最大限度地 降低工作量，保证信息的规范性；

**云端数据治理**

●可提供编码对照功能，与其他信息系统进行编码基础信息的共享和交换，实 现应用的集成；

●自动生成编码，保证唯一性、规范性、标准性。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.3 | 确定当前数据源，确定处理数据的系统和 业务流程 | ●数据源追溯及管理规范  ●主数据处理系统  ●主数据管理流程 |

【成熟度等级说明】

不是每个数据源都能提供高质量的查询结果。由于数据源返回结果之间存在大 量冗余，查询的数据源数量越多，冗余度也会越大，对大量不能保证准确来源的数 据进行分析很可能会出现偏颇。基于以上原因，数据源的管理成为一个关键问题。 所以在做云端数据治理时，把查询提交给很少量的数据源，但又要求返回的结果能 够很好地满足用户的特定需求，是数据源选择的理想目标。通过建设适当的数据处 理系统，并制订规范的业务流程，能够促进实现云端数据治理的价值。

**Level 3: 定义级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.1 | 基于统一的制订和维护标准，形成各业务 系统能够接收的主数据 | ●主数据管理规范  ●主数据标准 |

【成熟度等级说明】

主数据的统一管理意味着构建了一个通用的面向所有业务的平台，此时主数据 主机作为一个集线器 (Hub) 从多个业务系统整合主题域数据，使用集中和标准化 的方法转换异构数据。通过使用统一的数据标准，有效地解决了数据不一致的问题， 避免了数据在不同的地方代表不同的含义，大大降低了外部应用访问数据的复杂性。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.2 | 确定业务问题、定义主数据的涉及领域 | ●业务需求分析报告  ●主数据分类规则 |

【成熟度等级说明】

主数据管理主要是通过一组规程、方法和技术解决方案来保证与业务相关的整 个信息供应链内涉及的领域以及领域间相关主数据的完整一致性。因此，首先需要 确定与主数据管理相关的业务问题是什么,从而开展具有针对性的主数据管理工作。 同时，对于主数据涉及的领域明确定义有助于开展主数据管理工作，并在海量云端 数据环境下有利于快速寻找与主数据管理目标相一致的数据内容。主数据不再按照 业务条块分开存储，而是按照涉及领域集中存储和管理，打破了各业务“竖井”的 组织边界。

**第5章■云端数据管理**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.3 | 匹配相同源或多个源中的重复可疑项，创 建新的主记录，链接多个源中的相关记录，  检查唯一标识的重复项 | ●数据质量控制要求  ●数据清洗指南  ●数据源匹配规则 |

【成熟度等级说明】

云端数据生命周期过程中的分散特性带来了数据源的不一致和重复的主数据， 在云端数据治理过程中需要抱着发现重复、错误和不一致的宗旨检查数据的一致性 和结构，通过数据质量控制和数据清洗等手段匹配相同源或多个源中的重复可疑项， 解决错误和不完整的数据，将云端相同数据的多个版本合并为一个真实版本。通过 集中组织业务相关数据实现冗余数据存储数量的减少甚至消除，并且可以淘汰保存 有重复数据的实例，从而显著降低成本。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.4 | 制订主数据监管策略 | ●主数据监管策略  ●数据管理员职责和权限  ●数据质量控制要求 |

【成熟度等级说明】

主数据监管是个持续的过程，组织管理者通过主数据监管管理其主数据的质量， 定义准则、策略、流程、业务规则以及度量值，从而实现业务目标。包括委派数据 管理员、管理数据质量。

委派数据管理员：主要包括委派首席数据管理员、确定数据管理工作计划的配 置(根据成熟度的不同，从低到高分为按IT 系统、按组织和按主题区域调整数据管 理工作)、确定每个数据领域的主管发起人以及为每个数据领域分配数据管理员等。

管理数据质量：通过使用度量、提高和证明组织数据质量和完整性的各种方法， 提高主数据的质量，防止数据质量随着时间的推移而降低。数据质量包括数据标准 化、匹配、存活力和长期的质量监控。数据治理组织需要制订用于确定高价值数据 属性的策略和用于度量数据质量逐渐改善情况的机制。

**Level** **4:** **量化级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 4.1 | 制订了量化目标，能够指导主数据管理并 支持过程的执行 | ●主数据管理程序  ●主数据管理工具  ●主数据管理过程量化评价指南 |

【成熟度等级说明】

通过利用 SMART 原则，建立量化管理的规范和制度，对主数据的管理过程进 行量化考核和评价，细化管理过程的改进，不断提升主数据管理能力和水平，促进

**■云端数据治理**

云端数据治理的绩效。目标管理是使管理者的工作由被动变为主动的一个很好的管 理手段，实施目标管理不仅是为了利于员工更加明确高效地工作，更是为管理者将 来对员工实施绩效考核提供了考核目标和考核标准，使考核更加科学化、规范化， 更能保证考核的公正、公开与公平。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 4.2 | 提升主数据管理质量以便更好地支持云端 海量数据分析 | ●主数据管理程序  ●数据质量控制要求 |

【成熟度等级说明】

如同很多组织利用主数据管理清洗即将装载进数据仓库的数据一样，组织同样 需要高质量的主数据提升云端数据分析；同样地，也可以利用云端海量数据提高主 数据的质量，如从半结构化、非结构化数据中提取数据丰富的主数据。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 4.3 | 根据主数据的相关历史信息指导主数据管 理活动 | ●主数据管理程序  ●主数据操作日志 |

【成熟度等级说明】

主数据库中各元素间的关系通过数据定量化表示，可更准确地整理和分析主数 据管理。结合对于主数据使用和变更过程中的历史数据，加强对主数据活动的管理。 主数据的历史变更包括如客户名称变更、组织的新增地址及如何合并重复记录的沿 袭等信息，在主数据管理平台内可调取和查阅。

**Level** **5:** **优化级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 5.1 | 使用预测模型指导主数据管理过程的变更 | ●主数据变更管理程序  ●预测模型文档 |

【成熟度等级说明】

主数据管理过程产生变更的同时会带来业务的影响。通过采用预测模型可在变 更前分析其影响并支撑决策。例如主数据迁移或更改可能给业务带来的中断影响， 组织可以采用基于业务连续性方案制订的恢复时间和恢复目标来评估主数据管理过 程变更带来的影响，并采取相应措施避免或减小其影响。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 5.2 | 根据主数据策略，量化主数据的改进目标 | ●主数据策略管理程序  ●主数据管理方针、目标 |

【成熟度等级说明】

主数据的改进目标应当与组织的业务和数据治理策略相关联，主数据策略很好

**第5章■云端数据管理**

地体现了相关的内容。量化管理能从更细化的层面描绘主数据管理过程需要/值得改 进的方面以及程度，便于确定改进的目标和方向。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 5.3 | 通过规范的程序控制主数据的获取、变更 以及精简过程得到持续的改进 | ●主数据管理程序  ●主数据生命周期管理优化程序 |

【成熟度等级说明】

云端数据治理是一个活动集合，覆盖了统筹和规划、构建和运行、监督和评估、 改进和优化四大阶段。主数据管理是云端数据治理的一项重要内容，组织要将主数 据管理的持续改进纳入云端数据的治理过程。

**4.** **实施案例**

(1)文件名称列表

《主数据管理办法》

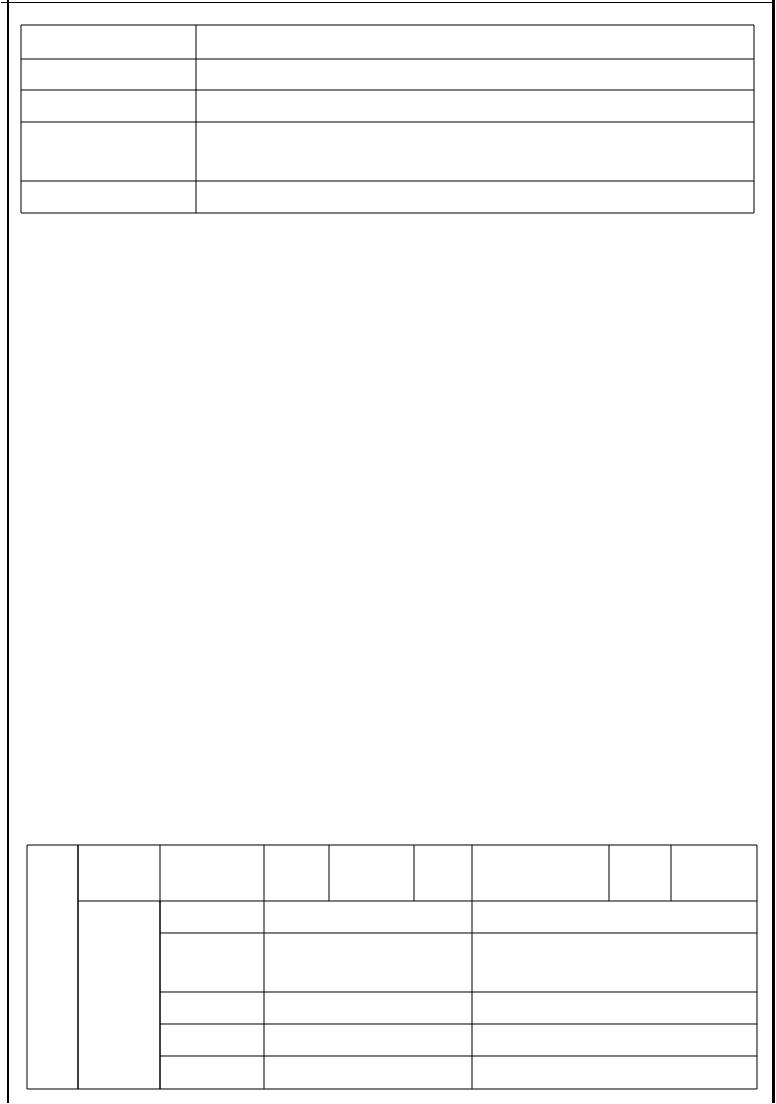
(2)实施案例

|  |
| --- |
| 主数据管理办法  1 总则  (1)为逐步形成完善的供应链管理体系，为组织运营和各业务发展提供唯一、准确的供应商基础数据，实现供 应商基础数据的单点录入、全局共享，制订本办法。  (2)本办法所称供应商，是指直接向公司提供物资和服务的组织及其分支机构、事业单位和个人。个人包括个  体工商户和其他自然人。  (3)本办法所称供应商主数据，是指在整个组织范围内各个信息系统需要共享的、长期稳定存在的、描述供应 商自然属性的相关数据。  (3)供应商主数据管理的原则： 一级平台、两级管理、三级操作。  2供应商主数据管理范围  (1)供应商主数据按照组织供应商和个人供应商分别管理。  (2)供应商信息主要包括基本信息、业务地点信息、联系人信息、采购信息和财务信息等五类信息。供应商基  本信息是供应商的自然属性，由供应商主数据系统管理。  (3)供应商的其他业务属性，由各业务属性的归口部门负责，通过各专业应用系统创建和维护。供应商的业务 地点信息和财务信息由财务部门归口负责，在ERP系统中维护；联系人信息和采购信息由采购管理部门归口负责， 在采购管理系统中维护。  3供应商主数据管理职责  (1)建立全集团统一的供应商主数据管理平台，建立全集团集中的维护工作组，统一负责供应商编码、数据质  量、数据安全等管理和日常维护工作。  (2)总部和分公司两级采购管理部门是供应商主数据的业务管理部门，负责制订供应商主数据管理制度、规范、  编码规则和操作手册，负责指导下级公司的供应商主数据业务操作工作。  (3)总部、分公司、子公司三级采购管理部门是供应商主数据的业务操作部门，负责受理各级供应商主数据创 建的申请、审核、创建、维护和分发等工作。各级采购管理部门的操作权限如下。 |



申 请 人 填 写

**■** **云** **端** **数** **据** **治** **理**

续表

职 责

单 位

总部采购管理部门

分公司采购管理部门

负责公布和全国供应商的审核、创建、维护和分发

负责本公司涉及范围内的供应商审核、创建、维护、分发

负责本公司涉及范围内供应商创建申请、更新申请和初审，并提交所属分公司/

子公司采购管理部门

总部采购管理部门集中管理

负责审核供应商信息的创建申请、更新申请

各级业务部门

(4)主数据操作实施属地原则，供应商原则上由供应商首次登记单位进行管理与维护。特殊情况下，可由集团 总部直接予以维护更新。其他分公司具备使用权，无更新维护的权限，可依据管理需要对供应商资质材料进行备 案管理。总部采购管理部门可以直接变更所有供应商的基本信息，可以下放或回收供应商主数据信息的更新维护 权限。

4 供应商主数据质量要求

(1)供应商主数据系统以真实性和唯一性为最高质量管理目标。

(2)供应商应严格按照供应商编码申请表要求规范填写申请事项。供应商在填写或更新其申请时，须承诺所填 写的基本信息是真实、完整、准确和唯一的。与供应商往来的各业务前端部门负责其基本信息进行初步审核。

(3)为保证数据的唯一性，供应商主数据业务操作部门按照如下原则进行数据查重。

a) 按照查重优先级别顺序，组织供应商查重规则为：

●供应商名称完全一致的，认为是同一供应商；

●供应商组织机构代码一致的，认为是同一供应商；

●国税登记号一致的，认为是同一供应商；

●地税登记号一致的，认为是同一供应商。

b) 按照查重优先级别顺序，个人供应商查重规则为：

身份证号码一致的，为同一供应商。

(4)建立供应商主数据定时查重的运行机制，加强主数据质量的日常管理。根据查重结果，各级采购管理部门 有权对可能重复的供应商数据进行合并或失效处理。同时将失效供应商信息根据失效流程发布给相关业务系统。

5 附则

(1)各子公司的供应商主数据管理应按照采购管理和IT 规划要求，逐步纳入供应商主数据管理平台统一管理。

(2)本暂行办法由集团采购与管理部负责解释，自发布之日起施行。

供应商基本信息创建申请单

供应商

电子

邮件

联系 电话

申请

日期

申请人

填写说明

内容

项 目

工商注册名称全称，按照工商营业执照

供应商名称

上的名称，填写全称

信息内容

英文名称

英文名称

法人代表

所在国家

工商注册法人代表

工商注册所在国家

**第5章■云端数据管理**

**续表**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **续表**   |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | | 申 请 人 填 写 | 信息内容 | 所在省份 |  | | 工商注册所在省份名称，例如：北京、 河北、澳门等 | | | 所在城市 |  | | 工商注册所在城市，按照营业执照上的 地址填写 | | | 工商注册地址 |  | | 工商注册所在地址，按照营业执照上的 地址填写 | | | 国税登记号 |  | | 国税登记编号，按照税务登记证上的编 号填写(个人供应商可不填) | | | 地税登记号 |  | | 地税登记编号，按照税务登记证上的编 号填写(个人供应商可不填) | | | 供应商类型 |  | | 类型：  1.企业供应商  2.个人供应商 | | | 组织机构代码 |  | | 企业法人组织机构代码填写9位组织机 构代码，前8位和最后1位之间以中划线  连接；个人填写个人身份证号码 | | | 营业执照 登记号 |  | | 供应商营业执照登记号码，按照营业执 照上的号码填写 | | | 上级公司名称 |  | | 上级公司名称，按照工商营业执照上的 名称，填写全称 | | | 上级公司组织 机构代码 |  | | 上级公司的企业法人组织机构代码，填 写9位组织机构代码，前8位和最后1位  之间以中划线连接 | | | 主页地址 |  | | 公司网站主页地址 | | | 供应商承诺：  以上所提供的信息和证明材料真实准确，对因提供有关信息、证件不实或违反有关规定造成的后果，责任 自负。  (供应商盖章/签字) 年 月 日 | | | | | | | | 推荐部门： | | | | 推荐部门审批： | | | | 采购管理部门审批意见： | | | | | | | | 受理情况 | | | | | | 操作人员签字 | | 口供应商已存在，已向省级单位提交分发申请  □供应商不存在，已向省级单位提交创建申请  口供应商已存在，已分发至本级 系统  系统  口供应商不存在，已创建供应商并分发至本级 | | | | | | 签名：  日期： | | 备 注 | | | 企业供应商需提交盖企业公章的营业执照、组织机构代码、国税、地税 复印件。  个人供应商需提供由本人签字的身份证复印件。  提示：本表单受理之日起2个工作日完成 | | | | |

 **■云端数据治理**

续表

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 供应商基本信息更新申请单   |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | 申 请 人 填 写 | 供应商 申请人 |  | 申请人  电话 |  | | 电子  邮件 |  | | 申请  日期 |  | | 信息内容 | 项 目 | 原内容 | | 更新内容 | | | 填写说明 | | | | 供应商编号 |  | |  | | | 填写供应商在集团的唯一编号 | | | | 供应商名称 |  | |  | | | 工商注册名称全称，按照工商营业 执照上的名称，填写全称 | | | | 英文名称 |  | |  | | | 英文名称 | | | | 法人代表 |  | |  | | | 工商注册法人代表 | | | | 所在国家 |  | |  | | | 工商注册所在国家 | | | | 所在省份 |  | |  | | | 工商注册所在省份名称 | | | | 所在城市 |  | |  | | | 工商注册所在城市，按照营业执照 上的地址填写 | | | | 工商注册 地址 |  | |  | | | 工商注册所在地址，按照营业执照 上的地址填写 | | | | 国税登记号 |  | |  | | | 国税登记编号，按照税务登记证上 的编号填写(个人供应商可不填) | | | | 地税登记号 |  | |  | | | 地税登记编号，按照税务登记证上 的编号填写(个人供应商可不填) | | | | 供应商类型 |  | |  | | | 类型：  1.企业供应商  2.个人供应商 | | | | 组织机构 代码 |  | |  | | | 企业法人组织机构代码填写9位组 织机构代码，前8位和最后1位之间 以中划线连接(个人填写个人身份证 号码) | | | | 营业执照 登记号 |  | |  | | | 供应商营业执照登记号码，按照营 业执照上的号码填写 | | | | 上级公司  名称 |  | |  | | | 上级公司名称，按照工商营业执照 上的名称，填写全称 | | | | 上级公司组 织机构代码 |  | |  | | | 上级公司的企业法人组织机构代 码，填写9位组织机构代码，前8位 和最后1位之间以中划线连接 | | | | 主页地址 |  | |  | | | 公司网站主页地址 | | | |

**第5章■云端数据管理**

续 表

续表

|  |  |  |  |
| --- | --- | --- | --- |
| 供应商承诺：  以上所提供的信息和证明材料真实准确，对因提供有关信息、证件不实或违反有关规定造成的后果，责任 自负。  (供应商盖章/签字) 年 月 日 | | | |
| 推荐部门： | | 推荐部门审批： | |
| 采购管理部门审批意见： | | | |
| 受理情况 | | | 操作人员签字 |
| □已提交更新申请至省级公司  口本省注册供应商，信息已更新  口非本省注册供应商，已提交申请至供应商注册省处理 | | | 签名：  日期： |
| 备 注 | 企业供应商需提交盖企业公章的营业执照、组织机构代码、国税、地税 复印件。  个人供应商需提供由本人签字的身份证复印件。  提示：本表单受理之日起2个工作日完成 | | |

**第** **6** **章** **云端数据质量**

**6.1** **概述**

**6.1.1** **目的和意义**

在云计算环境下，云端数据量成指数倍激增，如何管理这些云端数据，尤其是 提升数据质量成为很多组织颇为头痛的问题。在这一堆庞大的数据面前，垃圾数据 以及很多无用的数据不可避免，然而它们也会给我们的机房乃至数据中心带来数据 处理上的压力和负担。

垃圾数据一旦产生，就需要我们在数据处理的过程当中，对垃圾数据进行过滤 和清洗，并且自动决策这些数据的去留。这样的环境下，如果选择了一个坏的数据， 就会像感染病毒一样，可能引发连续的错误决策，甚至让企业蒙受经济损失。

**例子1:未经治理的数据导致商业行为失败**

这个事例源于一个大型金融服务机构的客户呼叫中心。就像几乎所有的客服柜 台一样，这里的客户服务代表们要做的就是接听电话，并把客户信息输入到一个共 享数据库里。

这个特殊的数据库里有一列是用来记录称谓的，并且是可编辑的。但是数据库 管理员并没有对这一列的输入规则进行约束，例如只能输入“某某先生”、“某某女 士”之类的称谓，反而可以接受客服代表输入的任何长达20或30字符的内容。在 倾听一些客户愤怒的投诉时，部分客服代表就会给每条记录添加一些他们自己想出 来的不完全友善的注释，例如“这个客户真是个白痴”这类的注释。

这种情况持续了很多年，因为机构里的其他系统都不会从这个称谓列中提取数 据，所以没有人注意到这一情况。其后某天，市场部决定发起一次直接邮寄活动来 推广一项新服务。他们想出了一个绝妙的点子：与其花钱购买一份名单，不如利用 客服柜台的数据库。于是，以诸如“亲爱的白痴客户林玲”这样的措词抬头的邮件 开始源源不断地发到客户邮箱里。当然没有任何客户会签约使用这项新服务。该机 构直到开始检查他们所发出的邮件时，才弄清楚前因后果。

**例子2:脏数据给带来更大的组织风险**

对于诸如保险公司、投资公司、基金公司、通讯公司等拥有大量客户的服务类 企业而言，客户数据是其重要的财富来源。然而，客户数据质量问题却一直是困扰

**第6章■云端数据质量**

企业开发新服务项目的绊脚石。在一项关于客户数据质量的调查研究中发现，平均 有8%～15%的客户数据记录存在各种问题。例如各种证件号码输入错误、联系方式 过期等，其中有五分之一的数据问题是由于客户的死亡造成的。这些账号 ID 如果被 欺诈和盗窃，必会导致更为严重的现实问题，例如已故客户的商业利益被陌生人兑 现，资产继承权被剥夺，公司机密泄漏等。

在独立研究机构 Ovum 公布的题为“优化企业应用：数据连接”的研究报告指 出：数据蔓延和劣质数据是导致不良应用程序管理和性能的罪魁祸首，直指低劣的 数据与实施或运行应用程序的初衷背道而驰。“无论应用平台的架构有多么精良，或 开发团队的效率有多高，在基础数据及其管理策略存在缺陷的情况下，所有为改进 应用管理和交付所做的努力都将是徒劳无功的。”Ovum 也对位于北美、澳大利亚和 英国的146家大型企业中的高层 IT 主管人员进行了调查。调查结果显示，影响应用 可持续性的重要原因在于数据方面的问题。数据质量的主要挑战包括以下四点。

(1)应用交付。提供清洁、可信的数据，合规性与及时访问被列为最主要的 挑战；

(2)应用性能。性能问题令将近85%的受访公司困扰不已，元凶是缺乏标准化 数据、低效(或缺乏)归档政策，其次是依赖过多的点对点界面；

(3)应用数据管理。数据可靠性牵涉到所有应用程序，平均20%～30%的数据 是重复的，这导致了应用程序维护成本的持续增长；

(4)此外，与数据迁移、同步和保留相关的问题亦在其中。

**6.1.2** **内容与关联性**

数据质量职能域通过定义和实施检测、评估以及清理数据缺陷，确保数据能够 满足其在业务运营、决策制订以及商业策划上的预期用途。整个职能域对云端数据 从需求、采集、清洗、分类、应用、存储、共享、维护、归档、消亡等生命周期的 每个阶段里可能引发的各类数据质量问题，进行识别、度量、评估、监控、预警等 一系列管理活动，并通过改善和提高组织的管理水平使得数据质量获得进一步提高。

数据质量职能域中包含的过程域有：数据质量战略、数据概要分析、数据质量 评估和数据清洗。通过这四个过程域，使组织能够充分理解其管理的数据属性和质 量，还包括评估、阻止和修正缺陷的机制，并确保数据质量能够达到组织的业务目 标和整体的战略目标。简而言之，这些过程域共同组成了由数据质量战略驱动的完 整数据质量计划。

(1)数据质量战略是所有数据质量管理活动的基础。它描述了数据质量的相关 活动，这些活动帮助组织确保数据质量满足业务需求。数据质量战略过程域定义了 一系列数据质量目标；

(2)数据概要分析和数据质量评估过程域帮助组织对照以上质量目标来评估组

**■云端数据治理：**

织管理的数据；

(3)数据清洗过程能有效减少以上各过程域的工作量、降低成本，并且确保组 织能够按需使用自身承载的数据资产。

本职能域的实践成熟度有助于将业务目标和重点转化为可实施的计划，并以组 织级项目的形式进行执行，使组织能够对数据质量的各个维度进行主动地管理。并 使组织能够从其数据资产中获取最大价值，能够从精准的、可信任数据中获取更大 的商业机遇。

**6.2** **活动域及要求**

**6.2.1** **数据质量战略**

**1.** **简介**

1)目的

定义一个完整的组织级策略以达到并维持所要求的数据质量水平，从而支持业 务战略目标和战术目标的实现。

2)介绍

数据质量战略定义了提高数据完整性的战略目标、战术目标和计划。它如同一 幅蓝图，强调了人员在数据质量上共同承担的责任。数据质量战略应涉及以下内容： 意义、数据储存的设计(例如，参照完整性、标准化、基数、分级管理、选择限制) 以及业务过程。四个部分都应该得到解决，另外，数据质量战略需要与目标数据架 构保持一致。完整的数据质量战略的另一个目标应该是要能减少和处理数据多余的、 过时的和无用的信息。

数据质量战略是基于可信任数据的分析建立的，分析内容为现存的主要质量问 题和业务目标。数据质量目标无法仅通过技术和技巧来满足。高质量主要得益于持 续的审查以及组织里所有相关责任人对审查信息的分享与沟通。可实施的数据质量 战略要求进行文化转移、获取高管的强力支持和进行持续的推广、教育以及对数据 资产的关注。

为了能持续地营造数据质量文化，组织必须要制订一个全面的可度量的策略， 且该策略要适用于所有的业务单元、业务过程和应用。数据质量战略的采用能帮助 相关责任人理解不同组织目标间的相通性，例如改进分析方法、更加精准的管理风 险以及改进操作。

当定义数据质量战略时，组织应纳入标准的使用指南，为质量的策略定义需求 维度，数据质量对照这些维度进行度量。以下是一个参考维度集。

**第6章■云端数据质量**

**表6-1** **数据质量战略维度**

|  |  |  |  |
| --- | --- | --- | --- |
| 序号 | 维 度 | 描 述 | 评 估 标 准 |
| 1 | 准确性 | 满足业务需求的程度 | 与原始内容相近性、与权威来源对比的精确性以及  度量精度的评判标准 |
| 2 | 完整性 | 必须的数据项已被记录 | 业务所需的数据是否有缺失，例如数据源是否完 整、记录维度是否完整、数据取值是否完整 |
| 3 | 时效性 | 数据被及时更新 | 数据能否及时更新以反映当前事实，例如数据处理 的及时性 |
| 4 | 唯一性 | 数据不存在重复值 | 在指定的数据集中是否存在重复值 |
| 5 | 正确性 | 数据体现了真实情况 | 数据内容和定义是否一致 |
| 6 | 可信性 | 数据的可信赖程度 | 数据能够使用的可信程度，例如是否经过确认无误 的数据 |
| 7 | 可用性 | 数据在需要时能够使用 | 数据可用的时间和数据需要内访问时间的比例 |
| 8 | 可访问性 | 数据能被访问 | 数据是否便于访问，例如访问前是否需要传输、解 密、解压缩等 |
| 9 | 业务有效性 | 数据符合已定义的业务规则 | 数据项是否符合已定义的业务要求 |
| 10 | 技术有效性 | 数据符合已定义的技术规则 | 数据项是否符合已定义的技术要求，如格式、取值 范围等 |

3)相关过程域

● 数据提供者管理过程域为制订数据质量目标和具体标准提供了帮助；

●战略过程域提供了协助数据质量战略制订的相关信息，并与组织级的数据管 理策略保持一致；

●数据质量评估过程域、数据概要分析过程域和数据清洗过程域包含了帮助提 高数据资产质量的相关实践。

**2.GQM 表**

|  |  |  |  |
| --- | --- | --- | --- |
| 目 标 | 问 题 | 度 量 项 | 度量项说明 |
| G1:数据质量 战略与业务目标 一致 | Q1:组织有哪些数据质 量战略 | M1:数据完整性、可用性、 保密性策略 | 组织的数据质量战略应包 含数据层面的质量战略 |
| M2:项目级、业务级和组织 级策略 | 组织的数据质量战略还应 包含不同组织级别的战略 |
| Q2:业务过程中如何保 证数据质量 | M3:数据质量专职人员及其 能力 | 设置专职人员，考核和度量 其能力，保证业务过程的数据 质量 |
| M4:业务流程及子过程的目 标 | 定义明确的业务流程及其  子过程的质量目标 |

 **■云端数据治理**

**续表**

|  |  |  |  |
| --- | --- | --- | --- |
| 目 标 | 问 题 | 度 量 项 | 度量项说明 |
| G1:数据质量 战略与业务目标 一致 | Q2:业务过程中如何保 证数据质量 | M5:子过程中的数据质量 | 业务流程是由众多子过程  组成，度量子过程的数据质 量，能够有效反应业务过程的 数据质量 |
| Q3:数据质量战略中是 否清晰描述了目标、方针 和过程 | M6:数据质量目标、方针、 过程 | 数据质量的目标应量化和  细化，能够反应组织业务过程 的数据质量，方针、过程也应 有量化的定义，并通过度量方 法得到量化的指标 |
| G2:数据质量 目标必须是可实 现的 | Q4:数据质量目标与组 织的哪些过程有关联 | M7:与质量目标关联的组织 过程 | 组织的质量目标，必须得到 组织相关过程的支撑 |
| M8:过程对于实现目标、方 针的关键性程度 | 对于不同的组织过程，应对 其关键性进行排序，以识别关 键性的过程。 |
| M9:量化的过程质量模型 | 组织的数据目标可以通过 过程质量的模型进行测量和 实现 |
| Q5:组织有哪些数据质 量需求 | M10:传输、存储等完整性 比率 | 组织应考虑数据在传输、储 存等过程中的质量 |
| M11:数据、应用、服务等 可用性比率 | 组织应考虑数据的可用性， 并通过手段进行测量 |
| M12:保密性等级 | 组织应考虑数据的保密性， 并设置不同等级，以满足不同 业务的需求 |
| G3:数据质量 战略必须贯彻到 组织的项目过程 中 | Q6:数据质量战略如何 得到广泛的推广、沟通和 使用 | M13:通过会议、培训等 人员普及率 | 从人员的角度进行普及，应 保证人员培训和沟通的效率 |
| M14:过程工具、流程普及 率 | 通过工具在业务过程和流  程中，进行数据质量战略的推 广和使用 |
| Q7:是否将数据质量与 系统的开发生命周期进行 了结合 | M15:与数据质量相关的子 过程 | 业务部门应结合业务特点 和组织系统的开发生命周期， 识别所有与数据质量相关的 过程及其子过程 |
| Q8:数据质量改进与业 务过程改进是如何结合在 一起的 | M16:数据质量和业务过程 改进的目标 | 数据质量目标和业务目标  应相符，并且能够支撑业务目 标的达成，目标应是量化的， 可测量的，才能实现过程改进 的目的 |
| M17:实现改进的子过程 | 应识别和通过度量，找到能 够改进数据质量和业务过程  的子过程 |

**第6章■云端数据质量**

**续表**

|  |  |  |  |
| --- | --- | --- | --- |
| 目 标 | 问 题 | 度 量 项 | 度量项说明 |
| G4:组织必须 掌握数据质量项 目的进度 | Q9:组织如何度量数据 质量项目的进度 | M18:质量度量项 | 数据质量的度量项应能够 反应质量项目的进度和数据 质量情况 |
| M19:质量度量工具 | 数据质量的测量需要相应  工具的支撑，以减少人工度量 的误差 |
| Q10:哪个组织单元负责 维护数据质量战略 | M20:产品部、运维部、开 发部 | 数据质量战略应及时更新  和优化，调整不适应的目标和 过程，与数据质量直接相关的 部门包括产品部、运维部和开 发部，应由其负责维护数据质 量战略 |
| Q11:哪个组织单元负责 数据质量的相关活动 | M21:组织与职责 | 数据质量的活动应由组织  全面考虑和统一部署，因此需 要在组织层面负责数据质量  的相关活动 |
| Q12:在标准、方法、技 能上如何做决策 | M22:综合决策的数据模型 | 决策需要综合考虑成本、技 术等因素，可以建立相应的决 策模型 |
| Q13:数据相关的角色、 分工和责任定义是怎样的 | M23:定义与业务过程相关 | 应从业务过程的角度，定义 数据相关的角色、分工和责  任，保证数据质量与业务过程 的一致性 |
| M24:定义与质量目标相关 | 应保证数据相关的角色、分 工和责任，能够支撑数据质量 的战略目标、方针和策略 |
| Q14:如何体现数据资产 质量的提升 | M25:数据热力图 | 建立组织数据资产的数据  热力图，反应数据资产的使用 效率和频率 |
| M26:数据分类、筛选、检 索效率 | 通过数据质量效率相关的度 量项，不断改进数据资产质量 |

**3.** **成熟度等级介绍**

**Level 1: 初始级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 1.1 | 记录数据质量的目标、规则和准则 | ●文件化的数据质量目标  ●文件化的数据质量规则和准则 |

 **■云端数据治理■**

【成熟度等级说明】

数据质量的目标一般包括完整性、 一致性、准确性、及时性等。评估数据是否 达到预期设定的质量目标，就可以通过以上指标来进行判断。

数据质量的规则包括数据字典、业务规则、经验规则和标准规则等，用于判断 各个阶段的数据是否符合相应的规则，以检测出数据的异常和错误。

数据质量的项目级准则包括已建立的标准、控制处理流程以及一些度量项，例 如错误率和质量阈值。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 1.2 | 业务相关责任人参与制订数据质量准则和 目标 | ●文件化的质量管理策略 |

【成熟度等级说明】

云端数据治理与传统数据治理的不同之处：云端数据的汇聚特性更弱，云端数 据存在于大量的业务节点和过程之中。因此需要基于业务或者过程来明确相关的责 任人，责任人应参与数据质量准则和目标的制订过程中，使数据质量在整个业务过 程中保持一致性和有效性。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 1.3 | 按照数据质量计划执行，实施数据质量规 则；监督数据质量的准则是否得到有效遵循 | ●文件化的质量管理策略  ●数据质量度量方法和记录 |

【成熟度等级说明】

可以通过以下工作实现数据质量战略。

●制订数据质量计划、准则、规则

●相关责任人的会议记录和评审记录

●对数据质量的度量项进行度量

●制订指导数据质量处理的文档

●数据库和软件所实施的数据质量规则，记录规则实施的过程

**Level 2: 管理级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.1 | 定义、评审以及管理数据质量战略 | ●文件化的数据质量战略  ●会议或评审的记录 |

【成熟度等级说明】

数据质量战略的根本目标是将组织的计划纳入其中，以确保数据能够满足预期 用途以及业务需求。策略的制订应当要能够推动当前状态进入目标状态。它应明确 地与业务目标、业务驱动因素以及组织数据管理策略保持相符。

**第6章■云端数据质量**

一个数据质量战略至少要包含以下内容。

●质量的战略目标和战术目标；

●业务效益与影响(依据范围内的业务范围分类);

●实施重点；

●质量准则；

●方针和治理；

●相符性过程；

●指导实施的顺序计划。

数据质量战略还应当包含以下内容的相关指南：数据概要分析、数据质量的期 望以及帮助指导数据清理类项目的规则。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.2 | 业务相关责任人参与制订数据质量战略 | ●文件化的数据质量管理策略 |

【成熟度等级说明】

业务相关责任人应参与制订数据质量目标、规则和准则等策略，使数据质量在 整个业务过程中保持一致性和有效性。数据质量战略建议应由组织的业务部门主导， IT 部门参与。这样，业务部门的相关责任人能由始至终地跟踪数据质量，不断优化 和改进数据质量战略。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.3 | 组织建立方针、过程和指南来实施数据质 量战略 | ●文件化的数据质量方针、过程和指南 |

【成熟度等级说明】

数据质量战略的实施由方针、计划和过程来支持，这些共同定义了达成组织质 量目标所必要的活动执行标准和指南。过程和方针中应包含项目中所应用的数据概 要分析、数据清洗、评估以及监控等活动指南。这些活动能应用于很多领域，例如 数据存储合并、数据仓储、源到目标的转换以及数据转换等。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.4 | 通过采用组织选择的数据质量维度来确定  数据质量需求 | ●数据质量维度及说明  ●数据质量需求  ●数据质量水平的评估记录 |

【成熟度等级说明】

使用多样的方式和形容词来描述质量维度。行业供应商和数据质量度量的专家 会从七到九个维度来进行描述。最重要的是组织决定哪些维度对他们来说是最重要 的，描述这些维度并使用他们来指导需求的开发，建立指导活动的准则，并使用他

 **■云端数据治理**

们对质量水平进行标准化的评估。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.5 | 根据业务目标与计划制订数据质量战略， 并通过了高级管理层的评审 | ●文件化的数据质量战略  ●高层会议或评审的记录 |

【成熟度等级说明】

为实现组织所需的数据管理能力，组织应定义相关的操作模型，并基于操作模 型建立数据质量战略。获取质量上的提升需要大量的工作，因此必须要识别出满足 目标的有效质量级别并平衡所需的工作，这些工作与业务计划也要保持相符。而业 务计划是制订数据质量战略的关键驱动因素。以上这些内容，必须要通过高级管理 层的评审和批准，得到高层支持能够有效保证数据质量战略的实施。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.6 | 对计划进行监督以评估进度 | ●质量工作计划的监督记录  ●数据质量战略  ●已识别出关键里程碑的数据质量顺序计划  ●方针、过程和指南 |

【成熟度等级说明】

这些计划的制订是为了达成数据质量战略的战略目标和战术目标。在计划制订 中，将组织的战略、战术目标作为计划的里程碑，说明每步计划与目标之间的关联 性。为实现目标和落实计划，计划中应给出组织应掌握的方针和过程方法，对于工 作计划应给出相应的指南和建议。

**Level 3: 定义级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.1 | 整个组织都遵循数据质量战略，并且根据 数据质量战略相应地制订了方针、过程和指 南 | ●文件化数据质量方针和过程  ●数据质量实施指南 |

【成熟度等级说明】

数据质量管理体系应当包括建立规范与制度体系、明确管理流程、建立管理系 统与工具体系、建立检查与跟踪机制、建立考核体系等。例如质量管理流程包括问 题收集、问题分析、制订实施方案、实施治理、考核评估和优化改进等环节。

数据质量战略应考虑以下内容。

●与业务策略保持相符；

●与业务目标相关的目标和阈值；

●整体的数据管理策略和目标；

● 主题领域的优先次序；

**第6章■云端数据质量**

● 与组织数据和架构标准保持相符；

●要求使用数据质量维度；

●审批数据质量规则的过程；

●数据概要分析、评估、清理活动的指南和准则；

● 指导实施的顺序计划；

●识别关键项目，并映射依赖。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.2 | 定义数据质量治理、实施以及管理活动中 的角色与分工 | ●文件化的数据角色和分工 |

【成熟度等级说明】

为了保证数据质量实施的顺利开展，必须为管理活动、实施和治理流程中指派 相应的角色和分工。下面给出一般组织的数据质量角色和职责说明。

**表6-2** **数据质量组织与角色分配**

|  |  |  |
| --- | --- | --- |
| 组 织 | 角 色 | 分 工 |
| 管理层 | 最高管理者 | 数据质量战略的评审 |
| 管理者代表 | 实施数据质量战略，管理数据质量团队 |
| 治理流程 | 数据治理组长 | 组织数据质量的整体策划和治理 |
| 数据架构师 | 数据架构和整体模型的设计、维护 |
| 数据管理员 | 数据生命周期的治理(如采集、清洗、分类、存储、归档和销毁等) |
| 实施过程 (项目层面) | 项目经理 | 项目规划、包含工作范围和时间 |
| 数据构建者 | 数据模型、查询语句 |
| 产品开发者 | 实现数据跟踪 |
| 分析者 | 提供新的数据业务问题 |
| 报告制作者 | 为问题和解决方案提供报告 |

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.3 | 使用已定义过程确定数据质量活动的效益 与成本，以指导数据质量战略的实施 | ●数据质量活动的效益和成本分析 |

【成熟度等级说明】

数据质量战略应当要能证明策略实施的价值与重要性。应建立清晰的价值定位 以执行策略。数据质量效益与成本的确定包含了对投资回报率 (ROI) 的分析、缺 陷对成本的影响分析以及与改进相关的业务目标。

**■云端数据治理**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.4 | 数据质量战略中的方针、过程和治理在整 个数据生命周期中都得到贯彻，并且相应的 过程在系统开发生命周期的方法学中都得到 强制执行 | ●文件化的数据质量战略实施记录  ●数据生命周期的数据质量检查  ●系统开发过程中的数据质量检查 |

【成熟度等级说明】

数据质量战略制度化的要求之一是确保随着新的数据管理系统和数据仓储的开 发，原则和需求能够嵌入到开发生命周期当中，成为其中的一部分。同样重要的是需 要对开发活动的计划和执行进行监督和适当调整以反映数据质量战略的执行情况。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.5 | 数据质量项目，例如数据概要分析、数据 评估、数据清洗和风险评估，与数据质量战 略识别出来的业务需求和成本效益分析相符 | ●数据质量业务需求和成本效益分析  ●文件化的数据质量项目过程记录 |

【成熟度等级说明】

组织开展和实施的质量项目，必须基于自身业务需求和成本效益分析的结果。 对于任何一个组织来说，数据质量工作是一个额外的、短期难以见到效益的、投入 成本较大的工作。目前有众多互联网公司开始注意到了数据的重要性，并为此投入 了大量的人力、物力和财力，但取得良好效益的组织并不多。因此，组织在实施数 据质量项目的过程中，为避免数据质量投入难以为继的情况，必须基于自身业务需 求和成本效益分析的结果，选择适应的人员、工具和流程，来开展数据质量项目。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.6 | 为整个组织的数据质量改进工作制订计划 和安排工作顺序，并对计划进行监督和维护 | ●对数据质量战略的评审  ●提供质量准则和指南的数据管理标准  ●经认可的方针和过程  ●经认可的数据质量度量项  ●已嵌入的软件开发生命周期(SDLC)数据质量过 程  ●主题领域相关的数据质量业务规则  ●标准的数据质量过程 |

【成熟度等级说明】

数据质量改进工作是一项长期、循序渐进的工作，为实现数据质量的目标和项 目目标，必须制订周详的质量改进计划，并针对计划工作提供资源，安排工作的顺 序，以保障项目目标和阶段目标的达成。工作计划和顺序由数据治理小组进行监督， 由治理主体提供相关的投入支持。

**第6章■云端数据质量**

**Level** **4:** **量化级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 4.1 | 使用数据质量度量项来分析所提出的关于 数据质量战略的变更 | ●数据质量的度量项  ●数据度量和分析记录  ●数据质量战略的变更记录 |

【成熟度等级说明】

通过制订策略来达到组织对数据质量的需求无法一次性完成。 一旦制订了策略， 就必须对策略的执行情况以及业务需求的变化进行密切跟踪。这两方面都必须定期 得到评估以保证策略的成功执行，并保证业务对质量的需求持续得到满足。在识别 出不足和差距后，应当对策略进行完善升级，这个过程可能会涉及到工作重点的转 移或对正在进行的项目的调整。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 4.2 | 对需要修正和预防的数据质量问题进行优 先次序的排列，对排列过程进行量化评估。 对优先次序进行定期评审和调整以满足变化 的业务目标 | ●数据质量问题列表  ●质量问题排序的量化评估  ●质量问题排序的定期评审 |

【成熟度等级说明】

解决质量问题所需的工作量和成本量可能是巨大的。因此，很重要的一点是确保 与这些活动相关的优先排序以及决策制订的有据可依，以保证成本与业务目标的价值 和成果相符。因而，量化评估质量问题并使用这些信息协助决策的制订至关重要。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 4.3 | 系统性地收集相关责任人对数据质量问题  的需求，相关责任人对数据质量改进的期望被 纳入数据质量战略当中，并得到度量和监督 | ●相关责任人的数据质量需求和期望  ●数据质量管理策略中应增加相关责任人的角色职 责 |

【成熟度等级说明】

记录的业务需求、数据质量准则和相关责任人的期望并非总是相符的。相关责 任人很容易认为他们对质量的期望没有获得满足，除非他们自己参与到数据质量战 略的制订当中。为了抵制这种情况的发生并保证质量级别满足业务需求，我们必须 要和相关责任人一起制订数据质量战略的衡量准则及影响成功因素。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 4.4 | 根据性能度量项的分析结果，可以用以调 整和修改数据质量战略的方针、过程以及指 南 | ●数据质量战略上经认可的变更  ●方针、过程和度量项上经认可的变更  ●根据数据质量进度的分析报告所确定的标准度量项  ●对数据质量战略、顺序计划或者配套的方针、过 程和计划进行的经认可的修订 |

**■云端数据治理**

【成熟度等级说明】

数据质量度量项直观地反映了数据质量的好坏，针对数据度量项的分析结果， 可以反映出数据过程的能力、数据过程的达成情况、数据操作的规范性等。据此可 以分析组织执行的数据质量战略方针、过程和指南的适用性，针对不适用的方针、 过程和实践指南，可以依据度量项的分析结果，进行调整和修改。

**Level** **5:** **优化级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 5.1 | 由高管对数据质量方案的关键节点和度量 项进行定期评审，并实施持续性的改进 | ●数据质量方案的评审记录  ●质量方案的持续性改进记录 |

【成熟度等级说明】

此处所指的度量项，并非一些项目级的度量项，诸如代码缺陷率、数据相关性 等。而应针对数据质量战略，围绕数据质量的目标提供相应的数据度量项，如可用 性、准确性、完整性等度量项。结合组织制订的数据质量方案关键节点，提交给组 织高管进行评审，识别可以改进的策略和措施。组织高管评审，能够体现高层对数 据质量工作的支持，提供所需的工作资源，保证数据质量实施的可持续性。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 5.2 | 组织与同行共享提高数据质量的最佳实践 与成功方法 | ●交流数据管理策略最佳实践的展示。(白皮书或文 章 ) |

【成熟度等级说明】

标准的实施(包括质量标准和架构标准)对达成质量目标是十分重要的。标准 包含以下内容：

●数据的架构标准

● 数据的交换标准(例如XML )

●针对质量规则的性能度量项

●数据质量共享服务

●本体和分类学

**4.** **实施案例**

1)文件名称列表

●《数据质量战略文件》

●《数据质量管理手册》

●《数据质量规则库》

2)记录文件列表

●《数据质量战略评审会议纪要》

**第6章■云端数据质量■**

●《数据质量目标测量记录》

●《质量目标达成情况评审记录》

3)实施案例模板

数据质量战略

|  |  |  |  |
| --- | --- | --- | --- |
| 发布部门 |  | 生效时间 | 2016年04月01日 |
| 批准人 |  | 文件编号 |  |
| 介绍 | 定义数据质量战略，制订组织统一的数据质量战略文件 | | |
| 目 的 | 该策略的目的是为了形成对数据质量的一致性理解，定义数据质量的目标和方针，规范 化我们对各类数据的操作，从而提高组织各类数据的质量 | | |
| 适用范围 | 该策略适用于组织内所有人员 | | |
| 目标 | 完 整 性 9 9 %  可 用 性 9 0 %  有 效 性 9 5 % | | |
| 方针 | 提高业务数据质量，实现跨越式发展 | | |
| 数据质量  策略 | ■角色和职责定义：数据治理小组、业务数据管理员、信息系统管理员、数据专家、财 务专家 ……  ■质量过程的组织策划：数据需求分析、数据清洗、质量监控、成本核算……  ■过程目标：  需求分析目标：错误解决率90%  数据清洗：数据质量改进效率20%  ………  ■组织内数据质量目标：  项目级目标：数据可用性95%,数据完整性99%,数据有效性95%  业务级目标：数据可用性95%,数据完整性99%,数据有效性95% | | |
| 惩罚 | 违背该策略可能导致：员工以及临时工被解雇、合同方或顾问的雇佣关系终止、实习人 员和志愿者失去继续工作的机会、学生被开除；另外，这些人员还可能遭受信息资源访问 权以及公民权的损失，甚至遭到法律起诉 | | |
| 引用标准 | 略 | | |

**6.2.2** **数据概要分析**

**1.** **简介**

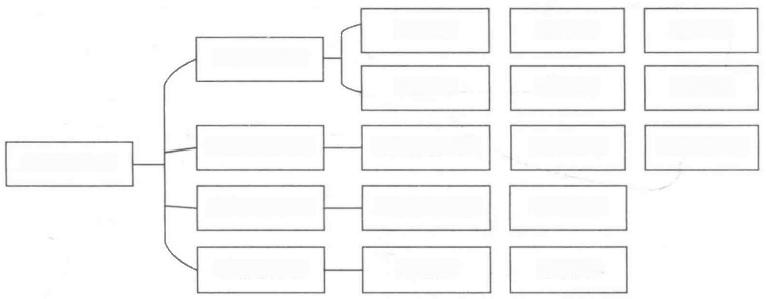
1)目的

了解管理下的特定数据集的内容、质量和规则。通过对数据自身统计值的分析 来反映数据质量问题。

**云端数据治理**

2)介绍

数据概要分析推动数据质量方案的有效进行，并且是许多信息技术活动的重要 起始步骤。数据概要分析应当针对关键的数据存储定期进行，同时也应当作为一个 由事件驱动的重要活动来协助数据质量的评估，以满足特定目的。例如，数据仓库 的定型、元数据库的填充、数据转换或迁移的执行、数据存储合并的策划或者数据 存储的修正之前，组织都会对数据进行概要分析。



值域分析

字段概要分析

数据分布

功能相关性分析

数据概要分析

主键唯一性分析

外键分析

基线分析

波动监测

关联性分析

表基础分析

血缘分析

字段间概要分析

数据表概要分析

表间概要分析

类型监测

异常监测

可追溯性分析

**图6-1** **数据概要分析架构图**

数据概要分析的任务是进行探索发现，展示数据库中存储的内容以及物理值是 如何与元数据库或数据存储文档中记录的预期允许值产生差异的。概要分析通常会 检查数值、范围、频率分布、出现偏差的元数据以及不标准的记录格式等。它也包 括对业务规则准确性的检查和对一些已知问题的分析。

根据数据概要分析的结果，组织可以进行一些修正活动。例如，修正数据、更 新元数据的描述、加强提取转换加载 (ETL) 脚本、变更数据结构、对质量规则进 行添加或完善、应用业务规则等。数据概要分析的结果通常作为关键的输入推动了 规则、统计、度量项、内容标准的确定或业务过程的重设计，重设计可以提高组织 数据资产的质量。

数据概要分析和数据质量评估不同，概要分析活动产出的是与数据集相关的一 系列结论，而评估过程的评估内容反映数据在多大程度上满足特定的质量需求。数 据概要分析通常是数据质量评估执行过程中的第一步。

建议组织分析正在执行的数据概要分析项目和数据存储内容；建立数据概要标 准的准则以确定应当执行数据概要分析的范围和对象；实施标准的数据概要分析过 程。根据业务需求和数据质量战略中定义的目标对需要执行数据概要分析的数据存 储和数据集(内部或外部)进行优先次序的排列。

3)相关过程域

●业务词汇表过程域覆盖了数据概要分析工作所要参考到的业务术语的标准

**第6章■云端数据质量**

化和评审；

●元数据管理过程域论述了与数据概要分析工作中所参考的特性和数据属性 相关的分类与具体信息；

●数据架构标准过程域包含了支持数据概要分析过程中所参考的设计和数据 表示标准的最佳实践。

**2.GOM 表**

|  |  |  |  |
| --- | --- | --- | --- |
| 目 标 | 问 题 | 度 量 项 | 度量项说明 |
| G1:建立并执  行数据概要分析 相关的标准方 法、工具、过程 集 | Q1:组织有哪些标准的 方法来执行数据概要分析 | M1:概要分析的标准 | 组织应参考相关的标准，支 撑自身的数据概要分析 |
| M2:概要分析的技术方法 | 组织应参考成熟的技术方  法和最佳实践，支撑自身的数 据概要分析 |
| M3:概要分析的流程管理 | 通过对数据概要分析流程  的管理，推进组织的数据概要 分析活动 |
| Q2:组织拥有哪些数据 概要分析工具和技能，以 及是否拥有该专业知识的 人员资源 | M4:概要分析的支撑软件和 硬件 | 数据的概要分析应配备足 够的硬件能力和自动化的软 件来执行 |
| M5:概要分析的文件工具 | 数据概要分析的指导文件 记录和分析工具等。 |
| M6:概要分析的专业人员及 团队 | 应配备专业的数据概要分 析人员和团队 |
| G2:提供建议 以提高数据资产 中的数据质量 | Q3:组织应用哪些统计 模型来实施数据概要分析 | M7:数理统计模型 | 组织应选择符合业务目标 和数据质量目标的数理统计 模型 |
| M8:模型优化和监测 | 对模型进行优化和监测，以 判断是否满足数据概要分析 |
| Q4:方针和过程定义了 哪些数据概要分析的准则 | M9:数据画像 | 通过数据画像，组织能够掌 握数据的整体概貌 |
| M10:数值分析 | 通过数值分析，组织能够掌 握数据的分布和特性 |
| M11:信息结构分析 | 通过信息结构，组织能够知道 数据的关联性，弥补数据缺失 |
| G3:物理数据 的表示是真实且 可理解的，以提 高对所辖数据集 的业务理解 | Q5:概要分析的进度安 排是基于哪些定义事件、 考虑事项和触发条件 | M12:触发概要分析的事件 | 当产生不满足组织业务目 标的事件时，应对事件进行记 录和分析，并进行数据概要分 析 |
| M13:事件产生的数据 | 大量事件产生的数据，能够 指导数据概要分析活动 |
| M14:数据和业务的关联性 | 数据的概要分析，能够解释 所属业务的流程，能够分析业 务流程的能力 |

 **■云端数据治理■**

**3.** **成熟度等级介绍**

**Level 1: 初始级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 1.1 | 为数据存储进行基本的概要分析 | ●数据概要分析报告  ●数据概要检查清单 |

【成熟度等级说明】

基本概要分析的内容包括分析列中不同值的种类和数量、零的数量或百分比、 空白值或无效值、字符串长度、日期范围、模式。这与高级分析不同，高级分析的 对象为基数、频率分布等。

**Level 2: 管理级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.1 | 建立并遵循数据概要分析方法学 | ●概要分析方法学  ●概要分析过程记录 |

【成熟度等级说明】

组织建立或采用的方法学描述了进行数据概要分析的方法。方法学通常会论述 工作的策划和范围的定义、概要分析的技巧、报告的模板以及结论的展示格式。另 外，分析过程在多个数据存储和共享的数据库中应该是可重用和可利用的。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.2 | 为项目建立数据概要分析的计划 | ●项目的数据概要分析技术 |

【成熟度等级说明】

数据概要分析计划当中要包含如下内容。

●选择数据存储进行检查；

●识别需要进行概要分析的数据集；

●相关责任人的清单以及他们的角色定义；

● 概要分析活动的目标；

●基于目标制订的数据质量准则，包括数据的参照完整性(父与子),元数据 记录中数据的一致性、与已建立的规则和模式之间的一致性、标准的数据质 量维度；

●概要分析活动中要应用的规则；

● 数据概要分析使用的方法和工具；

●记录结果的模板；

●活动进度，包括资源。

**第6章■云端数据质量**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.3 | 与相关责任人和数据治理团队共享数据的 概要分析计划 | ●概要分析计划的分发记录 |

【成熟度等级说明】

不应该无实质性地对概要分析活动进行策划和执行。相关责任人和数据治理方 都有其特定需求，应该将这些需求纳入考虑范围内。另外，要认识到执行概要分析 活动的成本和工作量并不少，因此，保持概要分析活动与业务目标相符是十分重要 的。分享概要分析的计划能保证一致意见的达成以及与业务目标的持续相符。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.4 | 根据计划执行数据概要分析活动，并且当 实际情况与计划出现明显偏差时对计划执行 工作进行调整 | ●概要分析活动的监控  ●概要分析计划的偏差分析  ●概要分析计划的调整说明 |

【成熟度等级说明】

由于概要分析工作经常会出现出乎意料的结果，因此组织需要足够灵活，才能 够在概要分析活动过程中决定是否需要针对初步执行的结果投入额外时间或工作量 来扩展范围。因此，在概要分析的活动中需要不断监控概要分析的过程结果，以便 对概要分析做出判断和调整。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.5 | 将数据概要分析的结果和建议报告给相关 责任人 | ●将数据概要分析方法学的文档  ●经认可的数据概要分析计划和进度  ●数据概要分析的调查报告和度量项  ●根据数据概要分析提议的业务附加规则  ●定义技能集和为负责数据质量工作的员工制订的  培训计划 |

【成熟度等级说明】

应使用分析结果作为数据质量评估和数据清洗工作的输入信息，并通过分析结 果得出数据质量战略。例如，组织可能认为其产品存在无法接受的错误率，相关责 任人则需要对这些错误率的影响程度进行判断，并负责决定是否需要采取替代以及 修正方案。

**Level 3: 定义级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.1 | 定义并标准化数据概要分析的方法学、过 程、实践、工具以及结果模板 | ●组织级的概要分析标准  ●标准化的报告模板和度量项、度量过程  ●概要分析的集中存储和发布 |

**■云端数据治理■**

【成熟度等级说明】

在组织里识别并持续使用标准概要分析工具以提高效率。应该定义并遵循组织 级的标准方法来分析和展示数据概要分析在业务和技术上对修正活动的影响。报告 模板和度量应当进行标准化，并进行集中储存和发布，以确保在组织级别上得到统 一使用。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.2 | 所有识别出来的用来满足概要分析目标的 技能都得到了使用 | ●支持概要分析的技能和工具 |

【成熟度等级说明】

数据概要分析一般被认为是对现有数据进行探索发现的活动，通常进行概要分 析的目的是为了满足特定的目标。组织为实现特定目标，应配备相应的概要分析技 能和工具。概要分析工作通常包括几个阶段，例如，数值概要检查(值域、ID 唯一 性等)、数值标准化分析(地址、语法等)、对所选业务规则的检查(业务规则需要 在组织内部确定)。

数据概要分析组应拥有为数据概要分析活动所挑选的技能和对应的工具。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.3 | 对数据需求、记录的元数据、物理数据以 及数据质量规则之间的关联性和可追溯性进 行捕捉和维护 | ●由数据专家提供的数据关联性分析  ●数据的可追溯性分析 |

【成熟度等级说明】

数据概要分析活动应当由对数据需求、数据质量规则、数据内容以及数据结构 有所认识的数据概要分析专家来执行。通过建立数据需求、物理数据以及元数据之 间的可追溯性来达成此实践。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.4 | 应识别核心的共享数据集以及对应的数据 存储，应当定期剖析和监督这些数据集和数 据存储 | ●数据集的重要性分类  ●数据概要分析的时间规则 |

【成熟度等级说明】

组织应当定义规则来确定不同数据集进行概要分析的时间(例如当获取到数据 时或者在数据合并、迁移、输出、分析、为保持相符性而做报告之前、或者进行结 构转换时)。对于组织数据基础的核心共享数据集，组织也应定期进行剖析和监督。

**第6章■云端数据质量**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.5 | 概要分析过程是可重用的，并且在多个数 据存储和共享数据库中进行部署了该过程。 | ●数据概要活动的最佳实践集  ●数据概要分析活动的复用性设计  ●数据概要实践的分发和共享 |

【成熟度等级说明】

因为数据质量活动通常价值很高却耗费很多资源，所以组织内的同行应当进行 共享和相互指导以建立数据概要活动的最佳实践。因此，提高效率并避免返工是很 好的方式。有些组织发现最有效的方式是由一支拥有技能人才的数据概要分析队伍 来进行分析活动。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.6 | 软件开发生命周期(SDLC)包含了数据概 要分析的任务以及裁剪准则、指南和治理 | ●数据概要分析标准，包括过程准则、标准、最佳 时间准则、裁剪以及报告格式  ●从组织级标准中裁剪的数据概要分析方法学  ●展示数据内容和特点的概要分析结果以及展示数 据需求可追溯性的报告  ●已记录的数据相关的决策和原理的裁剪  ●关于实践者对概要分析技能需求的文档记录  ●数据概要分析的度量项  ●通过数据概要分析活动总结出的建议报告  ●业务和技术的影响分析结果的模板  ●标准数据概要分析报告的要求  ●经认可的数据概要分析标准工具 |

【成熟度等级说明】

大部分的数据存储开发工作(例如新的数据仓库的创建)应该将数据概要分析 活动纳入项目计划。数据概要分析实践的制度化要求软件开发生命周期 (SDLC) 包 含活动的参考和指南，并且裁剪标准得到定义和遵守。

**Level 4: 量化级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 4.1 | 度量数据概要分析过程的性能，并根据性 能对整个组织的活动进行管理 | ●数据质量度量项的记录  ●概要分析内容和步骤的记录  ●数据概要分析的性能基线 |

【成熟度等级说明】

根据数据质量度量的反馈对数据概要分析的计划和进度进行管理。度量结果应该 要表明这个活动的输出在多大程度上满足并与业务需求和重点保持相符。应根据质量 和临界的提示决定什么时候、如何执行概要分析、分析内容是什么,这可能因业务应

**■云端数据治理**

用的不同而不同。被认为对业务过程十分重要的高度共享数据或数据集，应对其进行 定期的概要分析，这主要是因为其数据质量至关重要，需要对其进行频繁监督。

数据质量度量应该要能展示员工在多大程度上完成好了数据概要分析工作。计 划和执行的评估(实际和预估的对比)应该考虑以下因素：技能的使用、结果和决 策的影响、与方法和标准的相符性、质量的输出以及工作的完成水平。

建立数据概要分析过程性能基线，并使用该基线来告知数据概要分析活动的策 划、执行和结果。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 4.2 | 数据概要分析工作包含了对数据内容与其 经认可的元数据和标准间一致性的评估 | ●数据质量的基准  ●组织质量文档中规定的数据质量基准  ●数据质量基准的更新和修改记录  ●数据质量基准的评审记录 |

【成熟度等级说明】

数据概要分析应依据事先定义好并得到认可的数据质量基准。数据质量的基准 应采用认可的标准业务术语、含义、值和范围等，应采用基准对数据存储中的数据 内容进行概要分析。除此之外，还应考虑组织数据字典、数据模型、系统需求等文 档中规定的数据质量基准里。当需要时，应依据最佳实践来更新这些数据基准文档。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 4.3 | 在数据概要分析活动过程中，将实际问题与 根据历史概要分析结果采用统计方法预测出来 的问题进行对比 | ●数据概要分析的历史记录  ●数据概要分析结果的统计分析  ●概要分析引起的纠正和改进措施 |

【成熟度等级说明】

应当将概要分析的当前结果与对应的历史结果进行系统对比，以评估概要分析 活动对纠正措施和质量改进的影响。在实践中，往往会碰到许多无法事先预测的结 果。组织应根据实际问题与预测问题的偏差，不断修正和改进数据概要分析活动， 以实现数据概要分析的精准化和对实际工作的指导作用。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 4.4 | 对结果进行集中储存、系统监控，并且在统 计和度量项上进行分析，以此为数据质量的持 续改进提供理解分析 | ●已记录的数据概要分析方法学、最佳实践以及 标准  ●显示了将概要分析结果应用到数据质量治理的 项目报告  ●展示数据概要分析工作结果的数据质量支持工 具，例如仪表盘、积分卡等  ●数据质量门户网站，展示了用于建立性能基线 的数据质量模型和结果 |

**第6章■云端数据质量**

【成熟度等级说明】

使用同一个影响分析方法对修正活动的业务、技术、成本影响进行评估。将概 要分析的总结结果提供给数据治理主体以及高管。结果用以报告数据治理和数据架 构决策的制订，特别是针对高度共享的数据信息。

**Level** **5:优化级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 5.1 | 基于数据含义、技术特征和行为等知识的长 期累积，组织能够找到缺陷的根本原因以及其 他问题 | ●数据/业务缺陷的根本原因分析 |

【成熟度等级说明】

数据的概要分析是一个长期的、持续改进的过程，在实现组织的战略、战术目 标是不可避免的存在各种数据问题和缺陷。而通过长期累积的专业知识和工作经验， 使组织能够根据自身对业务的理解，找到数据缺陷存在的根本性原因。这些原因可 能是数据定义、数据存储格式、数据关联性、数据操作流程等。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 5.2 | 根据预期的量化效益、预估的成本以及业务 目标对数据概要分析过程和其他活动进行分  析，以识别缺陷并做出改进 | ●从效益、成本和目标的角度识别数据概要活动 的缺陷和改进方向 |

**【**成熟度等级说明】

第五级的成熟度要求组织对数据概要分析过程应有具体的量化分析，这种分析 不仅体现在概要分析过程的性能上，还应体现在数据概要的效益、成本和目标实现 程度等方面。可以对相同数据在不同时间里进行的数据概要分析结果进行定期的统 计分析，以度量概要分析活动的效能。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 5.3 | 针对所有的关键数据和数据库自动生成和推 送实时或近实时的概要分析报告 | ●数据概要分析报告  ●自动化生成和推送机制 |

【成熟度等级说明】

自动化生成和推送的性能分析和概要分析报告，能够提高数据质量项目的效率 以及应对计划或非计划事件的能力。通过及时的概要分析报告和便利的推送手段， 让数据相关责任人能够掌握关键数据和数据库的分析结果。

**4.** **实施案例**

1)文件名称列表

●数据概要分析的参考标准

 **■云端数据治理**

●数据概要分析的流程定义和职责说明 2)记录文件列表

●相关责任人使用概要分析结果的记录

●控制图，证明数据存储所使用的过程是稳定的(数据存储经过了充分的概要 分析)

●数据概要分析过程的改进目标包含在标准数据管理策略、方案以及报告中

●按时生成实时数据概要分析报告

●从数据概要过程分析中获得的结论以及针对改进提出的建议 3)实施案例模板

下表展示了一个数据概要分析说明。

|  |
| --- |
| 数据概要分析说明  数据的统计信息除了基本的记录条数、最大值、最小值、最大长度、最小长度、唯一值个数、NULL值个数、 平均数和中位数、针对字段的枚举和分布频率外，还包括相关性分析、主键相关分析和血缘分析等。  概要分析的定义和标准：  1 值 域 分 析  对于所有字段都适合，分析字段的值是否满足指定阀值、分析字段值的统计量(最大、最小、中位数、均值、 方差等);  常见BUG:最大、最小值越界、发现有NULL值。  2 基 数 分 析  分析字段中不同值的个数，这种方法更适合于维度类指标；  常见BUG:对于度量类指标数据比较集中，基数下记录个数过大或过小等。  3 类 型 监 测  分析字段真实值是否符合定义的数据类型， 一般在数据预览时对数据内容做评估。  4 数 据 分 布  分析各个维度值在总体数据中分布情况，根据经验值能从总体上看出数据的质量情况；  常见BUG:数据分布不符合预期；度量类指标数据过分集中，发现有NULL值过多。  5 波 动 监 测  分析检测值在一定周期内的数值波动是否在指定阀值内，这个更多用在线上数据监控中，可分析检测值在一定 周期内的数值波动是否在指定阀值内，如出现大幅波动需要关注。  6 异 常 值 监 测  分析字段中是否包含异常数据，例如空、NULL,分析一些约定异常值的数据的数量；  常见BUG:col存在异常值，约定异常值数据过多，如-99,-1,NULL等。  7功能相关性分析  分析字段或字段之间是否满足指定的业务规则，这个需要理解本表内字段间的业务逻辑关系，如pv>=uv(pv;  PageView;uv:UniqueVistor);  常见BUG:字段间的业务逻辑未满足预期。  8 主键唯一性分析  分析表数据中主键是否唯一，这个非常重要，如果主键不唯一会给下游表的计算带来无穷的困扰； |

**第6章■云端数据质量**

**续表**

|  |
| --- |
| 常见BUG:主键不唯一；主键为NULL  9 表基础分析：  分析表的基础统计量如分区、行数、大小，如果有对照表，最好参考对照表，如无对照表这个更多是靠经验值， 分析此表的业务对于表的基础统计量是否符合预期进行判断；  常见BUG:表基础统计量不符合预期。  10外键分析  分析事实表中的外键是否都能在维表中关联到；  常见bug:表里的外键没有在对应维表中找到对应，外键本身为NULL或异常值。  11血缘分析  分析表和字段的从数据源到当前表的血缘路径，以及血缘字段之间存在的关系是否满足，关注数据的一致性以 及表设计的合理性。 |

**6.2.3** **数据质量评估**

**1.** **简介**

1)目的

对照数据质量规则、根据过程和技巧，为度量和数据质量的评估提供系统的

*方法。*

2)介绍

数据质量是保证数据应用的基础。它的评估标准主要包括四个方面：完整性、 一致性、准确性、及时性。评估数据是否达到预期设定的质量要求，就可以通过这 四个方面来进行判断。

(1)完整性

完整性指的是数据信息是否存在缺失的状况，数据缺失的情况可能是整个数据 记录缺失，也可能是数据中某个字段信息的记录缺失。不完整的数据所能借鉴的价 值就会大大降低，也是数据质量最为基础的一项评估标准。

数据质量的完整性比较容易去评估， 一般我们可以通过数据统计中的记录值和 唯一值进行评估。例如，网站日志日访问量就是一个记录值，平时的日访问量约1000, 突然某一天降到100了，需要检查一下数据是否存在缺失了。再例如，网站统计地 域分布情况的每一个地区名就是一个唯一值。我国包括了32个省和直辖市，如果统 计得到的唯一值小于32,则可以判断数据有可能存在缺失。

(2)一致性

一致性是指数据是否遵循了统一的规范，数据集合是否保持了统一的格式。

数据质量的一致性主要体现在数据记录的规范和数据是否符合逻辑。规范指的 是， 一项数据存在它特定的格式。例如手机号码一定是13位的数字， IP 地址一定是

**■云端数据治理**

由4个0到255间的数字加上”.”组成的。逻辑指的是，多项数据间存在着固定的 逻辑关系，例如PV 一定是大于等于UV 的，跳出率一定是在0～1之间的。

一般的数据都有着标准的编码规则，对于数据记录的一致性检验是较为简单的， 只要符合标准编码规则即可。例如地区类的标准编码格式为“北京”而不是“北京 市”,我们只须将相应的唯一值映射到标准的唯一值上就可以了。

(3)准确性

准确性是指数据记录的信息是否存在异常或错误。和一致性不一样，存在准确 性问题的数据不仅仅是规则上的不一致。最为常见的数据准确性错误就为乱码。其 次，异常大或者异常小的数据也是不符合条件的数据。

数据质量的准确性问题可能存在于个别记录，也可能存在于整个数据集，例如 数量级记录错误。这类错误则可以使用最大值和最小值的统计量去审核。

一般数据都符合正态分布的规律，如果一些占比少的数据存在问题，则可以通 过比较其他数量少的数据比例，来做出判断。

当然如果统计的数据异常并不显著,但依然存在着错误，这类值的检查是最为 困难的，需要通过复杂的统计分析对比找到蛛丝马迹。这里可以借助一些数据分析 工具，具体的数据修正方法就不在这里介绍了。

(4)及时性

及时性是指数据从产生到可以查看的时间间隔，也叫数据的延时时长。及时性 对于数据分析本身要求并不高，但如果数据分析周期加上数据建立的时间过长，就 可能导致分析得出的结论失去了借鉴意义。

数据质量评估帮助推动质量期望的确定与达成，为重要的业务数据特性度量其 数据质量。评估活动的发起由不同优先级的业务推动，关注的重点通常是多个业务 领域所要求的高度共享的数据、精准的财务报表所需要的数据以及支持关键业务过 程的数据集。

应该对照那些对组织至关重要的数据元素进行质量评估。例如，那些对组织目 标的成功实现有直接影响的数据；监管汇报或内部符合性汇报中的数据、设定的关 键雇员数据、业务决策制订的相关必要数据等。重要的一点是确保方针中提供了指 南以决定哪些是关键数据集和数据元素。

为所选的数据集的特性建立目标(期望的质量级别)、阈值(可接纳的质量级别) 以及度量项。通常是以积分卡和仪表板的形式捕获和发布这些度量和度量项的。评 估的结果推动了根本原因的分析并且是组织数据质量改进计划的输入信息。

作为评估过程的一部分，将数据质量影响进行分类能够帮助相关工作的进行并 且确定效益。将影响对象进行分类(例如成本、风险、符合性、生产率)也能够协 助相关人员根据优先次序制订数据清洗计划。

随着治理功能和目标数据架构的逐渐成熟，组织倾向使用以基于主题领域的方 法来进行数据质量评估。例如，在基于主题领域的方法中，管理客户信息的数据人

**第6章■云端数据质量**

员代表组织，负责为共享的客户数据制订数据质量规则并进行监督，并成为数据质 量仪表板中那部分对应信息的“所有者”。应对生成并更新客户信息的关键数据存储 进行定期监控以确定可接受的质量阈值和目标是否达成。如果未达成，客户数据管 理人员应发起对根本原因的分析并发起对修正活动和缺陷预防活动的相应改进。

3)相关过程域

●数据质量战略过程域包含了与数据质量维度相关的额外信息，包含准确性、 适时性和唯一性等；

●数据概要分析过程域包含了数据质量评估中决定数据质量实践的相关的额 外信息；

●元数据管理过程域包含了与元数据管理信息和期望相关的额外信息。

**2.GOM 表**

|  |  |  |  |
| --- | --- | --- | --- |
| 目 标 | 问 题 | 度 量 项 | 度量项说明 |
| G1:建立并维 护以业务事件为 驱动的数据质量 评估功能，用以 改进数据资产的 质量 | Q1:如何进行数据质量 评估活动的 | M1:数据质量评估的开展形 式 | 数据质量评估，可以通过会 议、研讨以及日常例会的形  式，对数据质量和产生事件进 行评估 |
| M2:数据资产质量改进目标 | 评估应面向数据质量改进 的目标，不断改进数据质量 |
| Q2:评估活动是预先安 排好的还是由事件驱动的 | M3:数据质量评估的能力 | 数据质量评估应由具备相 应能力的人员和过程来执行 |
| M4:由事件驱动的评估 | 建立由事件驱动的质量评 估能力 |
| G2:根据行业  公认的技术和过 程，将数据质量 评 估 的 战 略 目 标、战术目标以 及阀值进行标准  化 | Q3:有哪些标准的数据 质量评估技术和方法 | M5:数据质量评估的技术和 方法 | 应识别数据质量评估的技  术和方法，并配备相关的工  具，以支撑数据质量评估过程 |
| M6:数据质量评估的过程 | 定义数据质量评估的过程 以推进数据质量评估工作 |
| Q4:领域里有哪些标准 的数据质量评估的维度、 目标和阈值 | M7:评估的质量目标和阈值 | 应定义数据质量评估的维  度、目标和阈值，对每次评估， 应详细记录数据质量的数值  分布 |
| M8:评估目标与过程数据、 度量项的关系 | 应建立评估目标一度量项  之间的关系模型，才能有效评 估数据质量 |
| M9:评估目标的改进和提升 | 数据质量评估的结论需要 指导质量目标的改进和提升 |

 **■云端数据治理**

**续表**

|  |  |  |  |
| --- | --- | --- | --- |
| 目 标 | 问 题 | 度 量 项 | 度量项说明 |
| G3:为数据质 量的统计评估建 立经验方法 | Q5:核心的数据特性有 什么标准的数据质量规则 | M10:数据分布的正态性等 | 数据的特性和规则，应满足 正态性分布，其余业务相关的 质量特性需要结合具体业务  进行分析 |
| Q6:领域内有哪些经验 方法和最佳实践 | M11:行业最佳实践和经验 参考 | 结合数据治理和行业实践 经验，进行数据质量评估 |
| M12:国际/国内行业标准 | 参考国际/国内的标准进行 数据质量评估 |
| G4:使用积分 卡、仪表板或其 他分析报告建立 标准的数据质量 评估汇报 | Q7:使用了哪些数据规 则引擎或者评估工具 | M13:业务规则、标准规则 和经验规则 | 数据的规则应满足业务对  数据的要求、满足标准的要  求，以及经验规则对数据的要 求等 |
| M14:软件及辅助文件工具 | 大量的组织数据，需要通过 软件和支撑工具，才能实现有 效评估 |
| Q8:是否评估适合组织 的工具和方法 | M15:考虑行业、技术以及 成本等因素 | 采用的工具和方法，应考虑 行业特色、考虑技术能力和经 济成本等因素 |
| G5:使用数据 质量评估的结果 和总结来制订并 完善数据质量改 进报告 | Q9:数据质量问题对组 织哪些方面产生影响 | M16:业务、成本、服务质 量、满意度、过程质量 | 数据质量问题和事件为组  织提供的服务对目标的达成  有着直接影响，需要具体地分 析其中的数值关系 |
| Q10:是否分析数据质量 问题并将其作为数据质量 改进的重要输入 | M17:过程、技术、方法等  问题 | 影响数据质量的因素很多 必须不断改进和优化数据质 量 |
| M18:改进方案中重点考虑 现有的数据质量问题 | 必须分析得出现有数据质  量问题，并进行优先级排序  以找寻改进的方向，并明确改 进方案 |

**3.** **成熟度等级介绍**

**Level 1: 初始级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 1.1 | 执行数据实例评估活动并记录结果 | ●数据质量评估模型  ●数据质量评估结果和记录 |

**第6章■云端数据质量**

【成熟度等级说明】

数据质量的评估模型一般包括评估对象、评估指标、评估规则和期望值。数据 质量评估的1级仅需要在项目过程中，针对评估数据集制订相应的评估指标，可以 通过形式化语言或者自然语言描述评估规则，执行数据质量评估工作获得评估值， 并记录评估结果。

**Level 2: 管理级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.1 | 根据标准的技术和过程建立、使用和维护 数据质量评估的战略目标、战术目标和阈值 | ●数据质量评估的目标和阈值  ●数据质量评估的方法论  ●数据质量评估的测量记录 |

【成熟度等级说明】

在传统的质量管理体系中，组织质量评估的战略目标、战术目标和阈值，是组 织建立质量管理体系过程中一个极其重要的内容。在云计算时代，数据成为质量的 依据和来源，建立数据质量评估的目标和阈值，就成为云端数据质量管理的首要前 提。标准的过程方法包括现状调查(输入)、分析研究(活动)和确定目标(输出) 等三个步骤。应先调查组织的数据质量现状，通过发散的思维，充分分析组织面临 的数据质量要求、客户需求以及存在的问题；在此基础之上，通过德尔菲法、层次 分析法等方法，对质量要求、客户需求和现实问题，提出相应的解决方案和优先级 排序；最后确定数据质量评估的各项战略、战术目标以及阈值，这些目标必须是量 化的和可测量的。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.2 | 数据治理宜按主题域为数据质量评估确定 关键的特性集 | ●组织业务相关的主题域  ●主题域相关的数据特性集  ●数据质量度量和评估的维度 |

【成熟度等级说明】

在组织管理共享数据资产质量的过程中，按主题领域定义关键特性集是十分必 要的。数据治理小组评审了关键特性，并预估了这些特性的价值，其预估的价值会 超过其管理和评估成本。同时，数据治理建立术语和定义的管理工作，包括了如何 定义和度量数据质量以保证组织内部的一致性。这个功能还包括数据质量维度的选 择，数据将会根据这些维度得以评估。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.3 | 根据每个数据质量评估方针所制订的周 期，定期执行对数据质量的评估 | ●数据质量评估方针中确定的评估周期  ●定期执行的数据质量评估记录 |

**云端数据治理■**

【成熟度等级说明】

数据质量评估从业务单元的角度对数据质量进行评估，判断其与已定义质量维 度和准则间的一致性。数据质量方针中会确定评估的周期，针对不同的数据集，周 期应有所不同。开展定期的数据质量评估，便于组织掌握自身的数据质量情况，以 执行改进和优化策略。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.4 | 数据质量评估结果包括了对修正活动的建 议以及相关的理由 | ●数据质量评估的结果 |

【成熟度等级说明】

业务相关责任人对数据质量的改进活动进行了优先次序的排列，并且对数据质 量的度量项进行评估、检查以及评审。

支持依据通常包括根本原因分析以及影响评估，这些依据推动了纠正措施的确 定以及数据质量改进活动优先次序的确定。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.5 | 影响分析包括以下内容：修正成本的预估、 工作量、业务影响的特征、有形及无形的效  益 | ●业务影响分析及说明 |

【成熟度等级说明】

执行影响分析的目的是为了修正数据质量缺陷。不同的修正决策产生的成本不 同，因此为了实现最小成本一最大效益的目的，必须要实施影响分析。

对业务影响进行严重级别的定性十分有用，这能估算缺陷对业务过程的影响(例 如高、中、低),并提供简要的依据说明。同时，技术影响也应得到估算，并根据高、 中、低得以定性，因为技术影响论述了修正缺陷的难度、时间和工作量。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.6 | 数据质量评估报告中的分析，需要有支撑 的具体数据度量和评估的记录证据，以便对 问题进行深入挖掘和影响分析，找寻改进的 机会 | ●战略目标、战术目标以及阈值的文档  ●已记录的数据质量维度和特性  ●已记录的数据质量评估的度量项  ●已记录的业务和技术的影响分析  ●数据质量改进活动的工作量预估  ●业务相关责任人审评数据质量评估活动的证据  ●提出进行修正活动的建议 |

【成熟度等级说明】

数据质量问题往往很细小，为了支持组织整体的数据质量改进，必须要记录所 有相关的数据度量和评估的记录，建议培养对实际价值的挖掘能力以推动对数据的

**第6章■云端数据质量**

理解并且隔离缺陷的来源。

**Level 3 定义级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.1 | 根据已通过的关于进度安排的数据质量方 针或根据具体的原因事件定期执行数据质量 评估活动 | ●数据质量评估的方针  ●标准化的过程、方法学和质量规则  ●数据质量评估的触发原则  ●组织评审记录和实施记录 |

【成熟度等级说明】

数据质量评估方针应定义如何将方法学、过程和数据质量规则标准化并在组织 层面沟通与一致地应用这些内容。启动评估活动的准则应该得到详细说明。结合定 期的评估和事件触发的评估，对数据质量进行有效跟踪并对事件原因进行定位。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.2 | 定义、评审业务影响(包括成本和风险) 的评估方法，并在组织里保持评估方法的一 致性 | ●业务影响的评估方法和评估技术  ●数据质量问题的根本原因分析 |

【成熟度等级说明】

在评估影响时，数据质量专家应保证活动使用了合适的技术来确定数据质量问 题发生的根本原因，这可能需要分析组织数据集范围以外的信息。在较大的组织机 构中，需要各业务部门遵循一致的评估方法，如对成本和风险的评估，这样才能从 细小的数据规律中定位组织数据的质量问题。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.3 | 组织层面在制订改进方案时，应考虑数据 质量评估所制订出的改进计划 | ●数据质量改进方案  ●组织对数据质量改进方案的评审记录  ●数据质量评估的结果和改进计划 |

【成熟度等级说明】

根据评估活动而非数据清洗活动所制订的改进方案应当能解决数据质量的根本 问题。这有可能会影响到多个相关责任人以及数据存储的使用。重要的一点是确保 从组织层面上对改进活动进行评估和策划。另外，应当要识别一个主题领域数据集 对另一个数据集的依赖，并协调两者间的关系以减少返工。高管应根据数据质量报 告所提供的影响评估结果确定这些活动的重点。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.4 | 使用已建立的阈值和目标并按照每个所选 的质量维度对数据质量进行评估 | ●数据质量评估报告  ●经组织评估和定义的数据质量维度 |

 **■云端数据治理**

【成熟度等级说明】

当进行评估活动时，提供清晰的方法来讨论评估对象的重要性并展示其对组织 的相关影响都是很重要的。应当根据标准的、经认可的数据质量维度来对评估结果 进行整理。按照已定义的阈值和目标进行打分，以清晰地表明质量是否满足要求。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.5 | 将数据质量度量的报告标准融合到系统开 发生命周期和相符性过程当中 | ●为每个标准数据质量维度记录的分数、目标和阈 值  ●针对经认可特性发布的可理解的组织级数据质量 规则  ●组织级数据质量评估方针  ●标准的组织级数据质量评估过程 |

【成熟度等级说明】

数据质量度量应遵循标准的过程，采用标准的目标和阈值作为度量的参考。数 据质量度量的规则也应该是标准化的，能够被组织中的所有业务部门和业务过程所 接受，这样才能保证数据质量度量的一致性。数据质量的好坏很大程度上取决于业 务信息系统，因此必须将标准化的过程和目标阈值融合到系统的全生命周期中，保 证系统数据能够符合标准化的要求。

**Level 4: 量化级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 4.1 | 根据特性和数据波动，系统地生成数据质 量度量报告 | ●数据质量的波动特性  ●数据质量的性能基线 |

【成熟度等级说明】

第4级的最大特点，是数据已得到有效度量，通过量化的手段对数值进行评估， 可以直观地体现数据的作用。例如可用性、准确性、完整性等数值，在3级中仅须 评估是否满足要求。而在4级中，则需要进一步评估得到其准确的数值，以此分析 对目标和阈值的偏离程度，从而识别改进的方向。因此4级的要求不仅包括准确的 数值，还要包括度量的方法和手段。可以根据标准机制，例如仪表板、记分卡等来 生成度量报告。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 4.2 | 使用统计或其他量化技巧对数据质量的操 作元数据进行标准化、获取及分析，以指导 改进活动 | ●审计和控制报告  ●数据质量评估的进度报告，供改进活动使用  ●数据质量的信任调查(例如，使用者对给定数据  集的信任程度) |

**第6章■云端数据质量**

【成熟度等级说明】

如上所述，数据量化的目标是用来分析和获取改进方法，指导改进活动，而操 作元数据则是识别改进方向和手段的重要依据。通过操作元数据的标准化、获取和 分析，可以识别出哪些关键操作活动、操作步骤和操作流程能够影响数据质量，通 过一般的统计方法和关联性计算，很容易得到想要的结果。

**Level 5: 优化级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 5.1 | 组织能够量化评估数据策略变更所带来的 效益，并根据数据治理实践调整管理的重点 和优先级 | ●对数据策略变更进行效益和成本分析  ●组织参考的数据治理实践  ●组织识别的数据质量问题 |

【成熟度等级说明】

已优化级要求数据质量评估的结果，能够量化地反映数据质量和效益的关系，

并且能够通过数据让管理层明确管理的重点。不断优化的量化管理是管理体系的最 高级别，让组织能根据制订的目标，不断识别面临的问题，不断调整自身的管理策 略和手段，从而实现最大的商业价值和战略目标。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 5.2 | 持续地对数据质量评估及汇报过程进行审 批和改进 | ●评估分析结果  ●过程评审文档  ●过程改进的提议和评审 |

【成熟度等级说明】

已优化级是一个不断改进和优化的过程，因此需要通过评估和报告的形式，对 过程改进提出相应的建议，对数据质量进行持续的评估和改进，以持续满足组织内 部和外部的业务需求和市场需求。

**4.** **实施案例**

1)文件名称列表

●数据质量评估管理流程

●数据质量评估标准

●数据质量评估目标和阈值 2)记录文件列表

●数据质量评估记录及报告

●事件驱动的数据评估过程记录

● 数据质量改进方案

 **■云端数据治理**

3)实施案例模板

下表展示了一个数据质量评估管理流程。

|  |
| --- |
| 数据质量评估管理流程  1数据质量管理的组织架构  目前企业数据中心的组织架构中，有专门的数据质量管理小组。建立数据质量管理小组是保证数据质量管理体 系正常工作的前提和保证，对于企业数据中心这样的大型项目尤其如此，必须设立专门的数据质量管理小组和数 据质量管理组织体系。数据质量管理小组专门负责整个企业数据中心项目的数据质量标准的建立、数据质量监督 和数据质量保证。根据项目规模，企业数据中心还在几个核心的分系统建设和维护过程中设立专职的数据质量管 理人员，从而组成专门的数据质量管理小组。数据质量管理小组的工作职责主要包括以下6个方面。  (1)负责制订数据质量的评估指标；  (2)负责制订数据质量的评估方法和评估形式  (4)负责制订数据质量可信等级管理策略；  (5)负责制订数据质量管理操作流程规范；  (6)负责定期发布数据质量报告。  2数据质量校验流程  企业数据中心平台每天都有很多ETL任务定时执行加载数据，确保ETL加载数据的完整性、准确性是数据质量 管理的基本要求。  2.1日常数据校验  数据质量管理人员每天要对ETL加载任务执行情况进行检查。  (1)数据校验方法选择  ETL任务数据质量校验要求必须采用以下三类方法中的至少一种来进行判断：记录数检查法；关键指标总量验  证法；值域判断法。  (2)数据校验周期  每天ETL加载任务比较多，如果全部执行数据校验需要的时间过长，因此根据每个主题数据的可信等级确定校 验频率。可信等级与校验频率的对应关系如下。  一级：每次加载都必须执行数据校验  二级：每三次加载执行一次数据校验  三级：每六次加载执行一次数据校验  对于需要特别保障的主题数据，可调整校验频率并额外增加经验审核法。  2.2定时数据抽查  ETL数据校验可以确保每天加载的增量数据的完整性、准确性，在此基础上，数据质量管理小组必须每季度组织一 次企业数据中心平台的定期抽查。定期抽查的范围必须包括可信等级为一级的所有主题数据，可信等级为二级的两个 主题的数据，可信等级为三级的一个主题的数据。定期抽查必须采用数据质量评估方法中定义的所有方法。  2.3全面数据检查  数据质量管理小组必须每年组织一次企业数据中心平台的全面检查。全面检查的范围包括企业数据中心平台所 有主题的数据。全面检查必须采用数据质量评估方法中定义的所有方法。  3 数据异常处理流程  (1)数据质量管理人员发现数据差错应及时核对核实，根据核对核实的情况填写数据问题处理单，描述数据质 量问题的现状、原因和改正、预防措施。  (2)数据质量管理小组组长审批后，报数据中心主管部门批准后执行数据修正任务。 |

**第6章-云端数据质量**

续表

|  |
| --- |
| **数据质量的评估报告**  1最终根据数据质量检查的情况，数据质量管理小组会定期或不定期生成相关的数据质量评估报告，数据质量 报告分为两类。  (1)每月定期提交的数据质量报告，即每月数据质量问题的陈述；  (2)数据质量抽查或全面检查后提交的数据质量报告。  2每月数据质量评估报告  每月的数据质量报告是在每月末或者下月初时编制的月度执行报告，数据质量管理小组会对当月数据质量情况 进行汇总统计，并根据“数据质量问题频率”的变化情况对各主题的等级进行检查策略的调整。报告格式如下。  A 非 系 统 问 题  (1)数据质量问题的现象(2)业务数据范围(3)问题陈述  B . 系统问题  (1)数据质量问题的现象(2)业务数据范围(3)问题陈述  3数据质量抽查或全面检查后提交的数据质量报告  数据质量抽查或全面检查报告是在每次进行完整的企业数据中心抽查或全面检查后编制的质量报告，同每月数 据质量报告相比，除了要统计数据质量情况并重新划分主题等级外，还需要对整个质量体系运作的情况进行评估 和改进。报告格式如下。  A 基 本 概 况  参与数据质量检查的相关组织与人员；数据质量检查的时间；数据质量检查的地点；数据质量检查的形式；数 据质量检查的范围等方面的内容。  B数据质量的检查与评价  (1)评价过程与步骤；  (2)数据检查方式：  主要分为全部检查和抽查。全检必须说明检查的范围、内容及方法；抽查必须说明抽样方案、过程及数据检查 的范围、内容及方法。  (3)数据质量评价方法。  C 数据质量评述与结论  (1)评述  对数据质量进行的综合描述(包括存在的问题)。  (2)结论与建议  包括检查结果、可信等级调整建议等。  4总结及展望  在企业数据中心平台搭建完成后，严格执行项目建设的流程管理，进行质量评估是非常重要的。企业数据中心 将来在规范流程管理的同时，应进一步加强数据质量的评估，评估操作流程的优化，形成良性循环。企业数据中 心的建设必须支持企业发展的战略重点，成为企业新一轮发展的重要支撑。 |

**■云端数据治理■**

**6.2.4** **数据清洗**

**1.** **简介**

1)目的

定义机制、规则、过程和方法，并根据预定义的业务规则对数据进行确认和修改。

2)介绍

数据清洗从名字上也看得出就是把“脏”的数据“洗掉”,指发现并纠正数据文 件中可识别的错误的一道程序，包括检查数据一致性、处理无效值和缺失值等。因 为数据仓库中的数据是面向某一主题的数据的集合，这些数据从多个业务系统中抽 取而来而且包含历史数据，这样就避免不了有的数据是错误数据，有的数据之间有 冲突，这些错误的或有冲突的数据显然不是我们想要的，被称为“脏数据”。我们要 按照一定的规则把“脏数据”“洗掉”,这就是数据清洗。而数据清洗的任务是过滤 那些不符合要求的数据，将过滤的结果交给业务主管部门，确认是过滤掉还是由业 务单位修正之后再进行抽取。不符合要求的数据主要有不完整的数据、错误的数据、 重复的数据三大类。数据清洗与问卷审核不同，录入后的数据清洗一般是由计算机 操作而非人工完成。

数据质量业务规则阐明了所有可应用的质量维度，而数据清洗活动主要对数据 进行更正以达到这些质量业务规则所制订的终端客户的准则。业务规则提供了标准 基线以识别数据缺陷和影响业务运营的异常情况。

应当在数据源或接近最初创建点处执行数据清洗，应尽量遵循数据概要分析以 及数据质量评估分析的结果。组织应制订数据清洗策略以指导清洗规则的分享和重 用，并将数据生命周期中重复或冲突的清洗过程降到最低。必须向所有上游和下游 系统发布数据的变更记录和说明。应当记录统一的问题的升级过程。应与内外部的 数据提供者确认数据的变更。

在制订数据质量战略时，与数据概要分析和数据质量评估相似，我们建议组织 分析事件并建立标准，以决定数据清洗活动的触发条件和执行时间。通常情况下， 数据清洗活动会更多地在关键的共享数据库或识别出的关键数据元素中执行。但我 们建议根据临界和预算，把数据清洗活动作为标准实践扩展到一些操作系统中，这 些操作系统可以将数据提供到其他目的地。和数据概要分析一样，我们建议对数据 清洗标准进行裁剪。裁剪的考虑因素包括数据库规模、数据库临界、特定的业务或 符合性需求，以及事件，例如在重设计或合并之前的工作等。

数据仓库领域的权威 W.H.Inmon 博士把数据清洗的过程分为六个步骤。

**步骤一：**元素化(将非标准的数据统一格式化，成为数据元素)。

**步骤二：**标准化(将元素标准化，根据数据字典消除不一致的缩写等)。

**第6章■云端数据质量**

**步骤三：**校验(对标准化的元素进行一致性校验，即在内容上修改错误)。 **步骤四：** 匹配(在其他记录中寻找相似的记录，发现重复异常)。

**步骤五：** 消除重复记录(根据匹配结果进行处理，可以删除部分记录或者把多 个记录合并为一个更完整信息的记录)。

**步骤六：** 档案化(将结果写入元数据存储中心。这样可以更好地进行后续的清 理过程，使得用户容易理解数据库以及更好地进行切片、切块等操作)。

3)相关过程域

●数据质量评估过程域和数据概要分析过程域提供了可以用于指导数据清洗 活动以及对如何其进行优先次序排列的信息；

●元数据管理过程域提供指导以确保对必要的信息进行管理，从而协助捕获和 记录数据变更；

●数据提供者管理过程域包含了与需求开发和管理相关的信息，把从内部数据清 洗活动所获得的信息与提供者共享，以协助其创建更佳的数据源的相关信息。

**2.GQM 表**

|  |  |  |  |
| --- | --- | --- | --- |
| 目 标 | 问 题 | 度 量 项 | 度量项说明 |
| G1:建立并 一 致地遵循数据清 洗策略 | Q1:组织是否有可重用  的数据清洗工具和过程 (自动方式以及人工方式) 以解决数据质量问题 | M1:数据清洗软件 | 组织的数据清洗工具，应满 足组织云端数据规模和数据  特性 |
| M2:自定义的数据规则库 | 组织的数据清洗规则， 一般 包括行业标准、业务规则，以 及数据关联性等 |
| M3:数据清洗流程 | 组织应明确定义数据的清  洗流程，从源头上提升组织拥 有的数据资产质量 |
| Q2:支持数据清洗，组 织部署了哪些一致性工作 | M4:一致性的数据规则和 工具 | 组织应在所属范围内，定义 一致性的数据规则，采用统一 的数据工具，确保数据的一致 性 |
| M5:相同的工作原则 | 组织数据清洗，应遵循相同 的工作原则和业务规则 |
| M6:一致性的清洗步骤 | 组织数据清洗，应定义明确 的操作顺序，遵循一致性的清 洗步骤 |
| Q3:组织是如何定义、 制度化和监督数据清洗过 程的 | M7:基于数据质量需求进 行定义 | 数据清洗应满足组织对数  据质量的需求，应基于数据质 量需求定义数据清洗过程 |

 **■云端数据治理**

**续表**

|  |  |  |  |
| --- | --- | --- | --- |
| 目 标 | 问 题 | 度 量 项 | 度量项说明 |
| G 1 : 建 立 并 一 致地遵循数据清 洗策略 | Q3:组织是如何定义、 制度化和监督数据清洗过 程的 | M8;基于IT治理的流程实 施制度化 | 数据质量清洗的流程应基 于最佳实践，例如对组织IT 治理的实践，以及参考国际/ 国内的标准 |
| M9:基于业务流程实施制 度化 | 组织应基于业务流程，实施 数据清洗的制度，确保数据质 量得到贯彻和实施 |
| M10:成立数据治理小组 | 组织应建立组织治理小组 整体负责数据清洗流程和数 据质量 |
| Q4:业务范围内建立了  哪些质量阈值和公差范围 | M11:主数据的质量阈值和 公差 | 组织应定义主数据的质量  阈值和公差，应掌握主数据的 数值分布情况和数据差异 |
| M12:关系数据的质量阈值 和公差 | 组织应定义关系数据的质量 阈值和公差，应掌握关系数据 的数值分布情况和数据差异 |
| M13:产品或服务数据的质 量阈值和公差 | 组织应定义服务数据的质量 阈值和公差，应掌握服务数据 的数值分布情况和数据差异 |
| M14:过程数据的质量阈值 和公差 | 组织应定义过程数据的质量 阈值和公差，应掌握过程数据 的数值分布情况和数据差异 |
| G2:建立并维 护标准的数据清 洗过程 | Q5:组织如何清洗重复 记录 | M15:确认后，进行合并或 者清除 | 组织应定义针对重复记录  的处理规则 |
| M16:改进数据结构或规则 库 | 组织应针对重复记录的产生 机制，改进数据结构或规则库 |
| Q6:是否在发现问题时 就执行修正活动 | M17:基于事件触发的数据 清洗流程 | 组织应建立基于事件的数  据清洗机制，可以是自动的， 也可以是手动的，以减少事件 带来的数据质量影响 |
| M18:数据清洗的结果，用 于数据质量改进流程 | 将数据清洗的结果，作为数 据质量改进流程的输入，以指 导数据质量不断优化和改进 |
| Q7:是否将数据清洗过 程贯彻到根本原因分析活 动中 | M19:基于数据清洗的过程 和结果，进行根本原因分析 | 针对数据清洗的结果和过  程，进行原因分析，找到影响 数据质量的根本原因 |

**第6章■云端数据质量**

**续表**

|  |  |  |  |
| --- | --- | --- | --- |
| 目 标 | 问 题 | 度 量 项 | 度量项说明 |
| G2:建立并维 护标准的数据清 洗过程 | Q8:在哪些过程中确认 修改以及评估数据清洗的 有效性 | M20:监督过程 | 在对数据的处理过程中，通 过监督数据使用的效率和数 值分布，确认数据清洗的有效 性 |
| M21:测试过程 | 通过测试数据应用的性能  和测量数据质量的度量项，确 认数据清洗的有效性 |
| M22:产品/服务提供过程 | 通过服务满意度、产品提供 的有效性等确认数据清洗的 有效性 |
| G3:数据清洗 过程有效支撑业 务和组织的发展 | Q9:投资回报率是否被 纳入到数据清洗的成本计 算 | M23:计算投资回报率与数 据清洗成本的关系 | 组织数据清洗需要投入人  力、物力和财力，应建立投资 回报和数据清洗成本的关系， 让组织管理层了解数据清洗 的回报，提高对数据质量的认 证 |
| Q10:组织的哪些数据清 洗活动考虑了操作和信誉 风险 | M24:数据合法性、合理性 判断 | 组织数据清洗应首先考虑  合法性，其次考虑合理性以及 操作的各种风险 |
| M25:脏数据替换 | 组织数据清洗应评估脏数  据带来的风险，采取相应的处 置措施 |
| M26:数据的可用性范围 | 组织应综合考虑数据的可 用性、安全性和保密性的要 求，定义数据的可用性范围 |

**3.** **成熟度等级介绍**

**Level 1: 初始级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 1.1 | 定义需求并执行数据清洗活动。 | ●数据清洗的需求  ●数据清洗的指南 |

【成熟度等级说明】

数据清洗是实现高质量数据的必经之路，数据质量的规则、数据之间的关联性， 以及云端数据带来的多副本复杂性等都直接影响了数据清洗的效率。因此需要明确 定义数据清洗的需求，依据具体的需求来执行数据清洗活动。

 **■云端数据治理**

**Level 2: 管理级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.1 | 数据清洗活动是根据数据清洗需求确定  的，数据清洗需求与过程改进相互作用以支 持业务目标的达成 | ●数据清洗需求的识别  ●数据目标状态的描述 |

【成熟度等级说明】

数据清洗需求通常来自于数据概要分析和数据质量评估的结果、变更申请、报 告中矛盾之处的识别以及常规的审计。需求通常会识别出以下内容：

●需要解决的问题(即由于数据问题导致的无法达成的业务目标);

●需要清洗的数据集(数据或元数据);

●数据的目标状态将如何协助解决问题。

数据目标状态的描述需要涉及对业务规则、数据定义、现在以及未来的使用、 质量标准规则、业务目标、成本与效益分析、数据、软件以及系统架构等的理解。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.2 | 数据清洗活动与数据质量需求保持一致  (例如， 一致性、精确性、唯一性等质量维度),  也与质量准则保持一致 | ●数据清洗活动和过程的记录  ●垃圾数据的分析  ●数据清洗结果的评估 |

【成熟度等级说明】

应当将数据清洗过程和规则进行记录，并与业务目标和需求联系起来。应当对 被清洗的数据进行评估，以检查其与数据定义间的一致性。而数据字典也应根据数 据清洗工作中发现的不足进行相应更新。

使用数据清洗规则以协助达成数据清洗目标和执行过程。规则通常涉及以下 内容。

●按照业务定义确认数据清洗规则；

●将数据清洗规则与业务规则联系在一起；

●与质量规则保持相符；

●让业务专家进行规则的创建；

●数据清洗规则和其确认的文档及审批；

●添加新规则的过程；

●采取措施以确保库一致应用了清洗规则。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.3 | 定义数据清洗的范围 | ●数据清洗的数据集、数据原则和规则  ●数据清洗活动的排序 |

**第6章■云端数据质量**

【成熟度等级说明】

范围包括定义清洗活动的数据集、数据元素和规则的识别，以及对数据清洗活 动进行优先次序排列的准则。需要根据时间限制以及可能存在的资源限制对范围进 行平衡。数据质量评估过程所得出的影响分析是估算数据清洗工作活动量和确定工 作范围的输入信息。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.4 | 根据事先定义的计划执行数据清洗活动 | ●数据清洗对数据和业务的影响分析  ●数据清洗的策划报告  ●数据清洗计划 |

**【**成熟度等级说明】

策划数据清洗的步骤：

●评估数据变更对业务的影响；

●决定需要进行变更的数据范围；

●识别数据的原始用途，识别数据清洗对数据含义和用途的影响，避免因对数 据含义无意识地更改而导致的低效与混乱。

数据清洗计划通常包含以下内容。

●用于修改数据的工具和方法(即按照有效来源和逻辑检查等对多个库进行对 比和确认);

● 定义过程以处理在数据有效性上的观点差异；

●对修改、审计跟踪和版本控制(配置管理)进行记录；

●对数据存储维护计划所需的调整进行反馈；

●定义如何在执行数据变更后维护参考完整性。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.5 | 建立并维护数据清洗方针 | ●数据清洗的方针 |

【成熟度等级说明】

制订过程中，确保适当的相关责任人(从治理到执行)参与其中是十分重要的。 如果有规定使用的支持数据清洗活动的工具和方法，则应该在方针中说明。

方针中应当包括一个与受影响的相关责任人进行沟通的机制，以及管理他们反 馈的方法。另外，方针当中应包含触发数据清洗活动的一些明确准则和阈值。方针 中应清晰阐明对根本原因分析、业务和技术影响的预期。方针也应阐明对数据清洗 项目进行优先次序排列和进度安排的准则，以及资源分配的指南。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.6 | 建立数据变更的方法，并在计划中进行定义 | ●数据修改和变更的方法学  ●数据清洗计划采用的方法说明 |

 **■云端数据治理**

【成熟度等级说明】

方法包括多个库的对比、参照有效来源、逻辑检查、参照完整性或公差范围进 行确认。

方法通常会涉及一些功能以确保以下内容。

●标准化、确认以及维护内容，必要时使用外部资源；

●有过程来处理在数据有效性上的不同观点；

●定义修改记录、审计跟踪以及版本控制(配置管理);

● 为数据存储的维护计划提供反馈；

●在执行数据变更后维护参照完整性。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.7 | 数据清洗的相关问题得到沟通和解决，尽 可能在内部或外部数据源头处进行清洗 | ●数据清洗方针  ●数据清洗过程和规则  ●数据清洗的度量项  ●数据清洗的计划  ●数据清洗的问题 |

【成熟度等级说明】

避免在多个场景多个实例中花费时间与精力清洗相同的数据集，最佳的实践方 法是从源头进行预防。在云计算场景中，数据的源头包括组织内部和外部的多个来 源，对来源进行存储和实施 ETL过程，匹配以数据清洗机制，这样尽可能地从源头 保证数据的质量。

**Level 3: 定义级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.1 | 在所有清洗活动中维护数据更改的历史记录 | ●数据变更记录  ●数据规则的变更  ●数据规则之间的联系性分析 |

【成熟度等级说明】

清洗规则的可追溯性，能够反映出数据变更的历史，即从规则需求的生成到变 更发生再到规则之间的相互联系，是具备可追溯性的。云端环境下，多源数据的可 追溯性会变得更加复杂，因此需要针对可追溯性建立相应的参考表结构和关联模型。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.2 | 通过策略、过程和程序，从源头上保证数 据清洗活动与发布的规则一致 | ●数据清洗过程的监督  ●数据清洗结果与数据质量规则的一致性分析 |

【成熟度等级说明】

数据清洗的规则是一套统一的、具有高度一致性的、公开发布的指导性文件，

**第6章■云端数据质量**

组织任何的数据清洗活动，不应违背该规则，否则产生垃圾数据和脏数据，直接影 响业务运行。应通过有效的管理策略、管理过程和管理程序，从业务需求的源头保 证数据清洗活动与规则一致。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.3 | 整个组织使用一致的数据清洗规则 | ●数据清洗过程的监督和抽查  ●组织统一发布的数据清洗规则 |

【成熟度等级说明】

数据清洗的规则直接影响到数据的质量。在云计算环境下，多种多样的数据来 源，可能会导致数据质量的不一致性。因此组织应遵循一致的数据清洗规则，多个 业务领域所共享的质量规则和业务规则不应有差异，除非存在经过深思熟虑且经过 审批同意的例外情况。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.4 | 建立数据治理小组，维护并且确认是否遵 守数据治理规则 | ●数据治理小组的角色定义和分工  ●抽查数据治理规则的遵循情况 |

【成熟度等级说明】

组织需要设置专门的数据治理小组，数据治理小组的职责与数据质量实施部门 的区别在于，数据治理小组需要从组织的业务层面和战略层面对数据治理的规则进 行维护和确认，并保证各数据质量相关的部门有效遵守了该准则。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.5 | 使用数据清洗结果的标准报告模板，包括 细节版和总结版 | ●数据变更记录  ●跟踪矩阵  ●数据清洗反馈  ●数据清洗治理、活动和规则制订的责任分配  (RACI)矩阵  ●数据清洗结果的报告模板 |

【成熟度等级说明】

组织能够通过报告获知重复进行的工作(在多个数据存储中清洗相同的数据集)。 在使用相同共享数据(例如客户数据)的情况下，允许人员花费时间同时对多个数据 库进行各自修改的效果，不如使用资源对系统和过程进行修改的效果好。后者能够实 现仅使用一个单一的标准数据集，该数据集能够满足所有使用者对质量的需求。

**Level** **4:** **量化级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 4.1 | 服务水平协议包含了让数据提供者为已清 洗的数据负责的数据质量准则 | ●服务水平协议  ●反馈文档 |

**■云端数据治理■**

【成熟度等级说明】

在云环境下，组织数据的来源不仅仅是自身业务，更多的是来自于内部、外部 的众多数据源。这些数据提供者也应遵循一致的数据质量准则，最佳的做法是通过 服务水平协议 SLA, 来约束数据提供者的数据质量，将数据确认的结果提供给内部 和外部供应商以实现他们在数据清洗过程上的改进。

**Level** **5:** **优化级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 5.1 | 组织参与建立和维护外部或行业标准以改 进数据内容的质量 | ●数据质量的参考标准  ●标准的有效性评审记录 |

【成熟度等级说明】

标准化是实现组织商业价值和社会价值的重要体现，组织的数据能够有效支撑 标准化的实践过程；另一方法，国际/国内以及行业的标准化活动，能够有效指导企 业建立符合国际/国内标准和行业标准的高质量数据，对于企业扩大影响力，推广商 业应用尤为重要。当组织参与行业活动，对多个组织生成或使用的数据进行标准化 时，该实践尤为适用。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 5.2 | 根据标准过程来管理数据提供者的数据清 洗需求 | ●通过会议纪要或备忘，展示标准的包含内容  ●服务水平协议，包括清洗过程和对数据提供者的 期望 |

【成熟度等级说明】

组织数据清洗过程可以作为一个商业产品，服务于众多数据提供者。数据提供 者的清洗需求则多种多样，有不同的商业目的和业务规则，通过标准的过程提供数 据清洗过程能够最大化满足不同数据提供者的商业需求。此处的标准可以包括行业 标准、国际/国内标准等。

**4.** **实施案例**

1)文件名称列表

●数据清洗规则库

● 数据清洗管理手册

●数据质量阈值 2)记录文件列表

●数据清洗的记录和问题分析报告

● 数据清洗的有效性分析报告

●数据清洗和成本效益的分析报告

**第6章■云端数据质量**

●事件触发的数据清洗过程记录

3)实施案例模板

下表展示了一个数据清洗管理手册。

|  |
| --- |
| 数据清洗管理手册  1 定义和确定错误的类型  1.1数据分析  数据分析是数据清洗的前提与基础。通过详尽的数据分析来检测数据中的错误或不一致情况，除了手动检查数 据或者数据样本之外，还可以使用分析程序来获得关于数据属性的元数据，从而发现数据集中存在的质量问题。  1.2定义清洗转换规则  根据上一步进行数据分析得到的结果来定义清洗转换规则与工作流。根据数据源的个数，数据源中不一致数据  和“脏数据”多少的程度，需要执行大量的数据转换和清洗步骤。  要尽可能地为模式相关的数据清洗和转换指定一种查询和匹配语言，从而使转换代码的自动生成变成可能。  2搜寻并识别错误的实例  2.1 自动检测属性错误  检测数据集中的属性错误，需要花费大量的人力、物力和时间，而且这个过程本身很容易出错。所以需要利用 高的方法自动检测数据集中的属性错误，方法主要有：基于统计的方法、聚类方法、关联规则的方法。  2.2检测重复记录的算法  消除重复记录可以针对两个数据集或者一个合并后的数据集，首先需要检测出标识同一个现实实体的重复记录， 即匹配过程。检测重复记录的算法主要有：基本的字段匹配算法，递归的字段匹配算法，Smith—Waterman算法，  Cosine相似度函数。  3纠正所发现的错误  在数据源上执行预先定义好的并且已经得到验证的清洗转换规则和工作流。当直接在源数据上进行清洗时，需  要备份源数据，以防需要撤销上一次或数次的清洗操作。清洗时根据“脏数据”存在形式的不同，执行一系列的转  换步骤来解决模式层和实例层的数据质量问题。为处理单数据源问题并且为其与其他数据源的合并做好准备， 一  般在各个数据源上应该分别进行多种类型的转换，主要包括三种。  3.1从自由格式的属性字段中抽取值(属性分离)  自由格式的属性一般包含着很多的信息，而这些信息有时候需要细化成多个属性，从而进一步支持后面重复记 录的清洗。  3.2确认和改正  这一步骤处理输入和拼写错误，并尽可能地使其自动化。基于字典查询的拼写检查对于发现拼写错误是很有用 的 。  3.3标准化  为了使记录实例匹配和合并变得更方便，应该把属性值转换成一个一致且统一的格式。 |

**第** **7** **章** **云端数据操作**

**7.1** **概述**

**7.1.1** **目的和意义**

随着云计算技术的飞速发展，企业在收集、分析和存储数据方面的能力较以往 呈几何倍数增长。根据权威机构调查显示，未来数据中心将进一步向云端偏移。云 端数据具有数量庞大、结构多样、移动性强等特点，针对以上特点，对数据的操作 如果没有实施规范化管理，未来将会影响到组织资源利用率乃至整体效率。

**例子1:数据采集板缺陷导致甬温线特别重大铁路交通事故**

2011年7月23日20时30分05秒，甬温线浙江省温州市境内，由北京南站开 往福州站的 D301 次列车与由杭州站开往福州南站的 D3115 次列车发生动车组列车 追尾事故，造成40人死亡、172 人受伤，中断行车32 小时35分，直接经济损失 19371.65 万元。经调查认定，导致事故发生的原因是甬温线温州南站提供的 LKD2 —T1型列控中心设备数据采集板软件存在严重设计缺陷和重大安全隐患。当温州南 站列控中心采集驱动单元采集电路电源回路中保险管F2 遭雷击熔断后，采集数据不 再更新，错误地控制轨道电路发码及信号显示，使行车处于不安全状态。

**例子2:商业银行数据生命周期管理面临的问题**

近年来，银行业务发展对信息科技的依赖程度日益增加。为满足不断增长的业 务规模与不断发生变化的业务需求，各商业银行不断提升业务系统的建设与改造力 度，逐步实现了全辖账务、信贷、客户信息等业务数据的总行集中管理。目前商业 银行数据集中管理，实现了技术上高度集成、营销上高度集约，但数据统一处理也 造成了数据量的急剧膨胀，给数据的管理和维护带来诸多问题，如存储空间、数据 保存、存储维护成本、大数据处理效率及数据安全等。商业银行积累了海量数据信 息，但不是所有的数据都具有相同的使用价值。随着时间的推移，有些数据被频繁 访问或更新，而有些数据则很少被访问。快速产生的、大量的、低价值的数据不仅 严重消耗了系统资源、拖累了系统处理性能，还可能影响系统的稳定性从而导致未 知的系统风险。

对于组织开展云端数据治理来说，云端数据操作管理的目的是对组织的数据资 产在整个数据生命周期内(从数据的产生、采集到存档和销毁等)进行规划、控制

**第7章■云端数据操作**

与支持，以实现对数据的管控从而确保达到组织的业务目标。

云端数据操作将有助于组织确保以下内容。

(1)数据需求得到完整的明确说明；

(2)数据在其所有产生以及使用的业务过程中都是可追溯的；

(3)数据变更及其过程得以管理；

(4)根据需求选择数据源，并将其视为权威来源进行控制和确认。

**7.1.2** **内容与关联性**

云端数据操作职能域包含的过程域有：数据提供者管理、数据集成与互操作、 数据生命周期管理。通过这三个过程域，使组织能够充分理解数据在生命周期内从 数据源选取到操作维护的机制，确保达到业务目标。

数据提供者管理描述了选择和控制数据源以及与内外部提供者进行双向互动沟 通的最佳实践。数据集成与互操作帮助组织通过共享数据库的设计来创建和维护与 业务需求的相符，并且建立标准并执行。数据生命周期管理协助组织确保在所有生 命周期阶段中数据可以从数据流很好地映射到业务过程。

本职能域的实践成熟度能够帮助组织增加对其数据资产的了解；允许业务使用 者识别最佳来源以满足需求，使上游供应商和下游的使用者能够通过映射依赖来预 期变更的影响；支持数据集成活动以及建立更为精准和有效的数据存储。

**7.2** **活动及要求**

**7.2.1** **数据提供者管理**

1. 简 介

1)目的

确保收集和使用的数据能满足业务目标，优化内部和外部数据源以提升数据质 量、满足业务需求并持续管理数据供应协议。

2)介绍

数据提供者管理的第一步是数据需求定义，其建立了相关过程，用以识别和精 确定义达成业务目标所需要的数据，并对这些数据进行优先次序的排列、记录和确 认。数据需求的描述必须使用经认可的标准业务术语，这样可降低其复杂性，对组 织内部有效地分析数据至关重要。数据提供者管理是定义数据采购需求、获取数据、 管理协议以及与提供者沟通的实践。除了将数据需求和合适来源进行映射外，这个

**■云端数据治理**

过程还包含了与内外部数据提供者建立服务水平协议。

(1)数据需求定义

数据来源的映射依赖于精准的术语、参照的数据定义、相符的计算方法学、数 据概要分析以及定义好的数据需求。目的是为了创建能识别、对比和选择替代来源 的参数，并通过业务规则判定属性可以结合、分离和解析的位置，以满足业务需求。

需求定义应当遵循一个有序的发现以及分解过程，这个过程阐明了业务概念和 需求，并且与现有的经认可的业务术语和定义保持一致。业务规则的制订和阐释应 当与支持目的地数据存储的应用逻辑设计保持并行。这个过程是双向且迭代的。完 成后，记录的功能数据需求应当以每个数据存储的逻辑设计形式完全展现出来，并 标准化。数据范围要到业务使用者具体确认完需求后才能最终确定，以此来确保每 个业务需求和其对应的业务规则都能获得审批，并且如果项目开发方法允许的话， 进行相关的测试。

应当在属性的层面上执行分析，并考虑关键的绩效指标、替代数据源以及服务 水平需求。将数据质量准则定义纳入数据采购过程，并使用数据质量准则来对比数 据采购的替代方案，并和经选择的内外部数据提供者一起支持服务水平协议。

应根据已建立的准则来评估源数据的质量，例如数据错误处理，权威来源的使 用、测试，管理对变更的反应和升级问题等，并通过适当的控制来确保数据的完整 性。应当对已交付的数据进行定期监督以确保已确立的准则持续得到满足。

(2)数据提供者管理策略

在大数据生态系统中，数据提供者将机构外部公共网络数据资源、外部合作组 织私有数据资源、机构内部数据资源或大数据服务提供者系统运行过程中的各种日 志、事件等系统行为数据资源进行抽象和建模。依据数据来源不同，数据提供者分 为外部公共数据资源提供者、外部机构私有数据提供者、内部数据提供者、机器或 系统数据提供者等数据提供角色。由此可见，数据提供者的种类是比较多的。为了 保证数据能够满足组织业务需求，减小不同数据来源的差异性，必须对数据提供者 进行管理。

对数据提供者进行管理确保了通过服务水平协议持续地满足已建立的准则。为 了达到这个目标，重要的一点是了解数据提供者是如何生成、合并或聚合以及交付 数据的。将此经验结合到数据提供者，以增加对数据的信心以及对质量问题的评估 方式作出评价。

3)相关过程域

业务词汇表过程域为数据提供者管理提供了基本的需求信息；

数据质量战略过程域包含了应用于指导与提供者签订协议的数据质量目标，因 此支持了该过程域；

数据概要分析过程域和数据质量评估过程域通过提供数据评估的方法和评估的 结果来支持该过程域。

**第7章■云端数据操作**

**2.GOM 表**

|  |  |  |  |
| --- | --- | --- | --- |
| 目 标 | 问 题 | 度 量 项 | 度量项说明 |
| G1:根据已记录 的过程，对数据在 采购、获取以及提 供者管理上的需求 包括数据需求准则 进行评估 | Q1;如何捕获、确认数据采购需 求并对他们进行优先次序的排列 | M1:数据搜寻 规格说明书 | 数据需求形成规范文档，便于 进行优先次序排序 |
| Q2:数据采购的需求是否明确、 清晰、由业务需求推动并可行 | M 2 : 分 析 和 测 试 | 来确认采购的数据满足已陈述 的需求 |
| Q3:是否有机制来确保对采购需 求进行业务上的审批 | M3:数据质量  准则 | 与内外部提供者共同定义数据 质量准则并将其纳入服务水平协 议中 |
| G2:根据标准数  据源选择和控制过 程，执行对数据提 供者的选择、缔约、 监督以及管理 | Q4:数据属性是如何映射到数据  源和下游应用的 | M 4 : 过 程 | 定义并遵循一个过程来为数据 搜寻规格说明书分析数据需求， 并将需求映射到已提供的数据元 素上 |
| Q5:如何管理数据源的选择过程 | M 5 : 度 量 项 | 建立、维护和使用数据搜寻管 理过程的度量项 |
| M6:数据搜寻 的评估和选择过 程 | 定义并在整个组织内使用数据 搜寻的评估和选择过程 |
| M 7 : 提 供 者 服 务水平协议 | 根据标准模板和过程来制订提 供者服务水平协议，并在组织内 实施、跟踪以及执行该协议 |
| G3:识别、记录 和了解潜在来源和 数据提供者，包括 他们的服务、数据 范围、过程和技术 | Q6:如何对数据提供者提供的服 务和内容质量进行监督 | M 8 : 服 务 水 平 协议 | 定期评审服务水平协议以确保 业务目标和需求的满足 |
| M9:计划变更 | 定期和数据提供者评审对数据 内容、过程、格式和质量等的计 划变更 |
| G4:标准服务水 平协议论述了所有 的业务需求，并使 用这些协议来管理 数据提供者 | Q7:提供者是否符合适用标准 | M10:关键性能 度量项 | 使用统计和其他量化技巧对与 服务水平协议相关的关键性能度 量项进行分析和评审，并使用这 些度量项来识别和解决问题 |
| Q8:是否存在一个可重复的过程 来管理问题，且这个过程包含了可 靠的连接点 | M11:过程改进 | 使用数据提供者过程的统计和 量化分析来改进过程并保证业务 目标得到充分的支持 |
| M12:经验 | 与同行共享学习到的搜寻经验 以及改进后的最佳实践 |

 **■云端数据治理**

**3.** **成熟度等级介绍**

**Level 1: 初始级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 1.1 | 将数据利益相关者评审并同意的数据需求 转化为数据采购规范 | ●业务词汇目录和它们的定义  ●需采购数据的系统的名称和功能  ●数据采集范围及方法  ●数据提供时间  ●双方责任  ●其他相关事项 |

【成熟度等级说明】

数据需求的描述应当与业务用语保持一致，并且在可获得的情况下使用经认可 的业务词汇。经认可的业务需求通常在业务词汇表的最初定义时和更改时产生影响。 每个项目的需求得到批准后，将其对应的业务词汇表进行更新是十分必要且重要的。 组织的数据需求根据提交对象的不同，可分为内部数据需求说明书和外部关联数据 需求说明书，分别提交相关方。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 1.2 | 执行分析和测试来确认采购的数据满足已 陈述的需求 | ●数据需求文档  ●审查委员会的会议记录  ●相关者需求审评决策的记录  ●数据采购需求  ●数据源选择准则  ●外部提供者的合同覆盖清单(包含格式、用法、  成本、时机、交付方法等)-  ●数据提供者的质量检查(清单)  ●数据反馈评估报告  ●与内外部数据提供者达成的协议  ●经审批的供应商发票  ●与数据提供者的交流沟通 |

【成熟度等级说明】

组织使用系统方法收集、记录、评估数据需求并且根据优先次序对其排列并确 认数据，使需求与业务目标、分析需求相符。组织定义了数据的相关需求，如采购 需求、数据源选择准则等，对采购的数据需要通过一系列验证手段(如第三方评估 等)来确认其满足需求。

**第7章■云端数据操作**

**Level 2: 管理级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.1 | 定义并遵循数据需求过程，并将需求映射到 已提供的数据元素上 | ●数据需求规格说明书文档  ●数据字典 ●元数据 |

【成熟度等级说明】

组织需要根据自己业务以及相关需求定义一个需求分析的标准过程，并将此过 程规范化、固定化。过程应包括识别对业务至关重要的数据需求以及对其进行的优 先排序，有时还包括关键的数据元素，已提供的数据需要与数据需求对应。必要时 还应包含需求和源特性间映射的数据文档(例如数据字典、元数据库等)。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.2 | 定义并遵循一个从外部提供者处获取数据 的数据采购过程 | ●评估准则  ●决策准则 |

【成熟度等级说明】

组织内一个已定义的采购过程通常包含以下全部或部分内容：检查所需的数据 是否已由组织的其他领域进行采购、评估准则、决策准则、识别授权源、审批权、 成本评估等。 一旦过程形成，应按要求遵循。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.3 | 维护数据需求与业务需求和目标间的可追 溯性 | ●数据需求管理工具及使用说明  ●数据模型 |

【成熟度等级说明】

组织通常会使用需求管理的工具来确保需求变更得到完整的跟踪。有些数据需 求也许能够通过这种能力得以捕获。另外一些需求则通过使用数据模型、特定的模 板或元数据库得以捕获和维护。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.4 | 与内外部提供者共同定义数据质量准则并 将其纳入服务水平协议中 | ●数据质量维度准则 |

【成熟度等级说明】

组织应与相关数据分析师共同定义主要的内外方标准数据质量维度准则 (例如准确性、完整性、 一致性、适时性),定期审核与数据源质量相关的关键 性能度量项。

**■云端数据治理**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.5 | 与数据提供者按计划举行讨论，以处理在服 务水平协议中建立的数据质量阈值与目标之 间的偏差 | ●采购方针  ●数据源选择准则  ●数据采购需求  ●展示来源、需求和特性之间相符性的数据定义  ●从数据需求到来源的映射  ●提供者服务水平协议  ●采购过程  ●数据源评估  ●与数据提供者会议的纪要 |

【成熟度等级说明】

组织应让数据分析师与提供者进行联络沟通，从而在采集需求和数据质量确认 过程上达成一致意见，并持续提供反馈，以减少偏差，确保达到业务目标。

**Level 3: 定义级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.1 | 对用于开发数据采购需求的组织级标准过 程进行监督 | ●数据治理方案  ●数据采购需求标准过程 |

【成熟度等级说明》

数据治理活动的持续监督是必不可少的，组织需要对数据的获取、处理、使用进 行监管，而监管的职能主要通过以下五个方面的执行力来保证——发现、监督、控制、 沟通、整合，其中监督一项包含对制订数据采购需求的组织范围内的过程监督。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.2 | 建立、使用和维护数据采购管理过程的度量 项 | ●信息格式  ●交付时间  ●更新的及时性 |

【成熟度等级说明】

组织需要不断完善数据采购管理过程的度量项，其中技术性能的度量项可能与 成本、进度、投入的工作量以及特定目标相关。组织对关键的性能度量项(即信息 格式、交付时间、更新的及时性、有效值、属性的精度和问题管理等)进行定义、 实施，并以它们为度量标准评估数据提供者。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.3 | 定义并在整个组织内使用数据采购的评估 和筛选过程 | ●数据采购评估准则 |

**第7章■云端数据操作**

【成熟度等级说明】

作为筛选过程的一部分，组织通常使用外部数据提供者的审计报告来体现他们 与控制标准和过程的符合性。审计报告可包含与外部数据提供者相关的基本情况， 业务情况，评价方法，存在问题，结论意见等内容。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.4 | 根据标准模板和过程来制订提供者服务水平 协议，并在组织内实施、跟踪以及执行该协议 | ●数据提供者服务水平协议 |

【成熟度等级说明】

服务水平协议通常包含已达成共识的并与数据质量、交付和及时性相关的补救 项选择。关键性能度量项被纳入标准服务水平协议当中，数据源的显示板、记录板 以及性能报告也反映了这些度量项指标。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.5 | 定期评审服务水平协议以确保业务目标和 需求的满足 | ●评审记录  ●服务水平协议模板  ●与提供者签订的服务水平协议 |

【成熟度等级说明】

组织可根据自身实际情况，按照固定周期组织相关方对服务水平协议进行评审 并记录，从而最大程度地减少多种的以及有些时候相互抵触的需求，以达到持续改 进的目的，确保满足业务目标和相关需求。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.6 | 定期和数据提供者举行会议，评审对数据内 容、过程、格式和质量等的计划变更 | ●标准数据采购过程  ●已定义的数据采购质量准则  ●已定义的数据采购度量项  ●根据相关方的反馈和最佳方案对数据采购过程进  行的更新  ●标准、流程、方针和工作流程图  ●数据提供者的会议纪要  ●与外部数据提供者相关的行业审计报告 |

【成熟度等级说明】

组织可根据自身实际情况，按照固定周期与数据提供者对数据内容、过程、格 式和质量等的计划变更进行评审并记录，从而避免计划实施后与预期产生不一致的 情况，以达到持续改进的目的，确保满足业务目标和相关需求。

 **■云端数据治理**

**Level** **4:量化级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 4.1 | 依据度量项评估数据需求定义是否满足了 业务目标，并对无法满足业务需求的数据需求 定义进行纠正 | ●维护业务需求和数据需求间映射和可追溯性的标 准工具集  ●数据需求定义框架所采用的行业最佳实践的选择 标准 |

【成熟度等级说明】

组织开展云端数据治理的目的是要提高数据管理能力，实现业务目标。数据源 是数据治理的重要内容，良好的数据有助于提高组织决策水平，而差的数据可能影 响组织决策。在对数据需求进行定义时，要充分考虑到业务目标，例如提高客户满 意度、提高产品质量、增强售后服务、开发精准营销等，组织应当针对不同目标采 取最合适的数据。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 4.2 | 使用统计和其他量化方法对与服务水平协  议相关的关键性能度量项进行分析和评审，并 使用这些度量项来识别和解决问题 | ●服务水平协议评审记录 |

【成熟度等级说明】

对与服务水平协议相关的关键性能度量不再仅仅依靠定性评价，而是增加了统 计学或其他专业度量手段来衡量，全面透彻分析其对组织业务的作用影响。关键性 能度量项可以从成本、实施风险和维护等方面来考虑选取。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 4.3 | 根据提供者的评估结果和预期的数据需求， 与已选择的外部提供者建立伙伴关系 | ●性能报告、仪表板、记分卡和热图  ●对数据提供者打分的准则  ●提供者绩效的分析报告  ●对提供者服务水平协议进行变更的建议 |

【成熟度等级说明】

组织根据预估数据需求对数据提供者进行评估，并追溯到潜在的数据源，建立 合作关系，以确保达到业务目标。

**Level 5: 优化级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 5.1 | 使用提供者过程的统计和量化分析来改进 过程并保证业务目标得到充分支持 | ●数据提供者过程记录 |

**第7章■云端数据操作**

【成熟度等级说明】

组织对数据提供者过程的统计结果进行分析，找出存在的不足并制订相应的改 进措施来确保对业务目标的支撑及实现。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 5.2 | 与同行分享数据提供者管理经验以及改进 后的最佳实践 | ●分析结果  ●与数据采购绩效相关的建议  ●数据源与业务目标相符的机制  ●演示、文章和白皮书 |

【成熟度等级说明】

组织在相关活动上应该积极分享实践成功经验，例如对外发布白皮书、行业最 佳实践、论文等资料，为行业最佳实践做贡献。

**4.** **实施案例**

(1)文件名称列表

《数据采集需求说明书》

(2)记录文件列表

《数据反馈评估报告》

(3)实施案例

下表展示了数据采购需求的样例，仅供参考。

|  |
| --- |
| 数据采集需求说明书  1 引 言  1.1编写目的  这部分说明文档编写目的，描述本系统特点及使用数据仓库技术实现的业务目标。  1.2背景  这部分是项目背景描述。  1.3参考资料  这部分列出本文档引用资料的名称，并说明文档上下级关系。  1.4术语定义及说明  这部分列出本文档中使用的术语定义、缩写及其全名。  1.5一般约束  这部分描述本系统开发过程中的各种可预见的约束条件，例如工具性能约束、开发技术支持约束、软件范围约  束等。  2 数 据 源 分 析  2.1 数据源概述  这部分将对本项目中涉及的数据源名称及其业务功能做出简单描述，如果多个数据源之间有关联，那么需要说  明他们之间的关联关系。 |

 **■云端数据治理**

续表

|  |
| --- |
| 2.2数据源物理环境描述  (1)<硬件平台>  这部分记录所有数据源程序运行环境需要的硬件平台，包括  ●每一个数据源所在的硬件平台环境  ●数据源服务器所在的硬件平台环境  ●数据仓库服务器的硬件平台环境  ●数据源程序运行监控硬件平台环境  这里的硬件环境包括机器的型号、硬盘的容量及其速度以及网络环境等。  (2)<软件平台>  这部分内容是记录所有与数据源程序相关的软件平台，包括：  ●每一个数据源所在的软件平台环境  ●数据源服务器所在的软件平台环境  ●数据仓库服务器的软件平台环境  ●数据源程序运行监控软件平台环境  这里的软件环境包括操作系统的种类和版本、数据库的种类和版本等，对监控平台还要关注其应用软件。 (3)<网络平台>  这部分记录与数据采集程序有关的所有网络环境的定义。  2.3数据源数据更新特征  这部分对每一个数据源中本次项目中涉及的主表的数据更新状态进行分析，例如是日更新还是月更新以及更新 的标志，以作为以后编写增量采集程序的基础。  2.4数据源可以提供的文件格式  这部分对本项目中涉及的数据源可以提供的数据格式进行记录，供确认采集需求使用。  3数据仓库环境分析  3.1数据仓库数据量预估  这部分评估本次项目中所有与数据量有关的参数，包括数据源历史数据量、数据源日增量、数据仓库历史存储 时间，并根据这些参数估算本次项目中需要向数据仓库增加的数据量。  3.2数据仓库粒度定义  这部分根据业务需求和数据源数据分析确定本次项目中涉及的业务实体的粒度。这里讨论这个问题，是因为粒 度的高低关系到数据采集的难度。  3.3数据仓库运行环境描述  (1)<硬件平台>  这部分记录所有数据源程序运行环境需要的硬件平台，包括：  ●每一个数据源所在的硬件平台环境  ●数据源服务器所在的硬件平台环境  ●数据仓库服务器的硬件平台环境  ●数据源程序运行监控硬件平台环境  这里的硬件环境包括机器的型号、硬盘的容量及其速度以及网络环境等。  (2)<软件平台>  这部分记录所有与数据源程序相关的软件平台，包括： |

**第7章■云端数据操作**

续表

|  |
| --- |
| ●每一个数据源所在的软件平台环境  ●数据源服务器所在的软件平台环境  ●数据仓库服务器的软件平台环境  ●数据源程序运行监控软件平台环境  这里的软件环境包括操作系统的种类和版本、数据库的种类和版本等，对监控平台还要关注应用软件 (3)<网络平台>  这部分记录与数据采集程序有关的所有网络环境的定义。  (4)<数据仓库工具平台>  在这里介绍数据仓库工具的使用限定，包括工具的版本信息、使用模块的名称及其作用、在本文中的缩略语等。 (5)<产品供应商支持能力描述>  这部分详细记录产品供应商提供的技术支持承诺内容、指定的技术支持人员及其联系方式。  (6)<数据仓库工具测试记录>  如果在“软件需求分析”阶段实施了数据仓库工具测试，那么这里将测试的结果记录下来。 4数据采集程序需求分析  4.1数据采集程序运行平台描述  (1)<硬件平台>  这部分记录所有数据源程序运行环境需要的硬件平台，包括：  ●每一个数据源所在的硬件平台环境  ●数据源服务器所在的硬件平台环境  ●数据仓库服务器的硬件平台环境-  ●数据源程序运行监控硬件平台环境  这里的硬件环境包括机器的型号、硬盘的容量及其速度以及网络环境等。  (2)<软件平台>  这部分记录所有与数据源程序相关的软件平台，包括：  ●每一个数据源所在的软件平台环境  ●数据源服务器所在的软件平台环境  ●数据仓库服务器的软件平台环境  ●数据源程序运行监控软件平台环境  这里的软件环境包括操作系统的种类和版本、数据库的种类和版本等，对监控平台还要关注应用软件。  (3)<网络平台>  这部分记录与数据采集程序有关的所有网络环境的定义。  (4)<数据仓库工具平台>  在这里介绍数据仓库工具的使用限定，包括工具的版本信息、使用模块的名称及其作用、在本文中的缩略语等。  (5)<产品供应商支持能力描述>  这部分详细记录产品供应商提供的技术支持承诺内容、指定的技术支持人员及其联系方式。  (6)<数据仓库工具测试记录>  如果在“软件需求分析”阶段实施了数据仓库工具测试，那么这里将测试的结果记录下来。  4.2数据采集程序性能要求  这部分描述数据采集程序运行时的各种性能要求和管理要求，例如计划多少时间完成数据采集程序的运行、对 数据采集程序的各个组成部分怎样定义、需要对数据采集程序提供怎样的维护和管理手段等。 |

 **■云端数据治理**

**续表**

|  |
| --- |
| 4.3 数据采集程序组件命名规则  这部分详尽说明数据采集程序的每一个组件的命名规则，集体内容因数据采集程序运行方式的不同而不同。 5 附件  这部分将与本系统相关的数据源字典列在此处。 |

**7.2.2** **数据集成与互操作**

**1.** **简介**

1)目的

减少组织从多个来源获取数据的需要，提高数据在需要整合和聚合的业务过程 中的可用性，并确保数据在不同业务过程或云服务器间的互操作性。数据集成可带 来源数据的优化，通过集中来实现成本节约，并提高数据质量。

2)介绍

数据集成与互操作是把不同来源、格式、特点性质的数据在逻辑上或物理上有 机地集中，从而为组织提供全面的数据共享，并确保数据在云服务器间的互操作性。

数据集成说明了从多个源到目标环境的数据传输和处理(连接、结合、重复数 据删除等)。数据集成处理的主要内容包括多个来源的不同的数据表示、业务含义的 合理化以及转换逻辑的复杂性。

通常情况下，数据集成关注于应用程序之间的数据传输和整合，以及数据从交 易系统或外部来源集成到共享存储库。数据集成的范围广泛而且详细。利益相关方 需要了解项目交付的组件和依赖性，以及所计划的集成工作的迭代阶段。业务利益 相关方需要采取长远的视角，这意味着处理集成障碍的灵活性。

成功的集成努力取决于持续的合作。业务用户需要确定并验证用于转化或整合 的业务规则，并且IT 需要了解预期的业务使用，并在组织的集成和数据标准范围内 实施业务规则。这还涉及到数据管理职能、标准应用的表现一致性、业务术语表的 更新以及其他组织范围内的考虑事项。由于不断变化的业务需求，应定期评审现有 集成环境的表现一致性及其与业务要求的一致性。

互操作性又称互用性， 一般是指不同的计算机系统、网络、操作系统和应用程 序一起工作并共享信息的能力。对于数据治理活动中，数据互操作性正是基于数据 来源广泛、数据格式复杂、数据处理平台差异化等特点的考虑，保证数据在不同平 台间的可用性。

3)相关过程领域

数据质量策略过程域能够提供数据集成期望和标准方面的指导。

**第7章■云端数据操作■**

元数据管理过程域提供了关于过程和基础设施建立的更多信息，以指定和扩展 关于所管理的数据资产的有组织的明确信息，增加数据共享，确保数据的合规使用， 提高对业务变化的响应，并减少数据相关的风险。

架构标准过程域给出了关于提供一组经批准的预期的更多信息，以管理关于经 批准的数据表示、数据访问、数据集成和数据分发的架构元素，这对于数据资产控 制和信息的有效利用和交流至关重要。

数据生命周期管理过程域提供了关于整个组织内的数据使用和需求的信息。

数据概要分析过程域提供了关于理解所管理的一组指定数据的内容和质量的更 多信息。

**2.GQM 表**

|  |  |  |  |
| --- | --- | --- | --- |
| 目 标 | 问 题 | 度 量 项 | 度量项说明 |
| G1:建立并遵  循一致的过程 以确保用于数据 集成的业务和技 术的持续一致性 | Q1:如何评估数据整合需要 | M 1 : 文 件 | 将数据集成计划形成文件 |
| Q2:如何在最大程度上减少未 来冗余 | M2:数据集成性 能模型 | 定期评审数据集成性能模型，并 将其作为改进的信息输入 |
| Q3:在存在冗余的情况下，组  织如何有效整理数据 | M3:管理数据 | 完全整合并集中管理所选的高  度共享的数据，并根据需要将其传 输至集成数据存储区 |
| G2:利用可确 保遵循数据架构 标准和数据质量 要求的标准过程 和工具集，进行 数据集成 | Q4:是否存在数据集成标准，  是否对其进行了评审、监督、批准  和实施 | M4:集成验证 | 执行集成验证，以确保将架构和 接口规范形成文件，并在发布生产 之前满足这些规范 |
| Q5:描述用来实施集成标准的 合规过程 | M 5 : 规 则 | 组织遵循一套标准的惯例和规 则进行数据集成活动 |
| M6:数据集成原 则和工具 | 组织所使用的一套数据集成原  则和工具提供了成批数据传输和 载入、变更数据采集、版本控制和 配置、元数据采集和管理以及在线 数据质量检查和控制 |
| M7:数据优先级 规则 | 建立并遵循标准的过程，以基于  使用情况、要求和选择的触发因  素，与业务用户一起创建并验证数 据优先级规则 |
| M8:架构标准 | 根据支持重复使用的架构标准， 指定集成接口的开发和部署 |

 **■云端数据治理■**

续表

|  |  |  |  |
| --- | --- | --- | --- |
| 目 标 | 问 题 | 度 量 项 | 度量项说明 |
| G3:主动研究  并评估集成技  术，以进行应用 和采纳 | Q6:在数据获取和整合时，如 何将数据质量阈值和目标应用于 数据源 | M9:质量检查 | 质量检查被定义为组织集成标  准的一部分，并作为数据集成过程 的一部分执行 |
| M10:统计分析 | 集成测量指标的统计分析指导 了接口和集成变更的决策 |
| Q7:识别丢失数据的过程是否  为自动过程，对缺陷或差距的跟踪 是否支持修复 | M11:接口和集成 性能测量指标 | 收集并分析接口和集成性能测 量指标，以确定不符合标准和指标 的情况 |
| M12:变化控制过 程 | 建立并遵循变化控制过程，以确 保控制和协调集成环境的变化，包 括上游资源和下游目标 |
| M13:修复过程 | 建立并遵循修复过程，以解决选 定的异常情况 |
| G4:建立并管 理数据转换、转 化和充实规定， 以便使数据在进 入集成环境之前 得到全面处理并 达到质量标准 | Q8:如何确保足够的工作人员  来监测、管理和维持数据质量，以 进行获取和集成 | M14:数据集成性  能模型 | 定期评审数据集成性能模型，并 将其作为改进的信息输入 |
| M15:最佳实践 | 该组织在其行业内发布并分享 整合最佳实践 |
| M16:数据源和目 的地变更 | 组织通过数据治理过程，来记录 并管理数据源和目的地变更 |

**3.** **成熟度等级介绍**

**Level 1: 初始级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 1.1 | 对不同系统之间的数据进行了集成操作 | ●数据集成脚本  ●来源的集成序列 |

【成熟度等级说明】

数据集成是将组织内不同来源与格式的数据从逻辑上或物理上进行集成的过 程。传统上，数据集成可以分为两大类方法，即数据仓库和联邦数据库。数据库仓 库技术在物理上将分布在多个数据源的数据统一集中到一个中央数据库中；而联邦 数据库则仅通过将用户查询翻译为数据源查询来进行逻辑上的数据集成。例如，生 产数据库管理员 (DBA) 按预定顺序手动启动多个集成脚本，或进行手动更改来纠 正缺陷数据。

**第7章■云端数据操作**

**Level 2: 管理级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.1 | 将数据集成计划形成文件 | ●数据范围的识别  ●供应商和源系统识别(以及SLA)  ●数据增长和数据保留期限的识别  ●适用的数据标准的识别  ●来源的集成序列  ●与数据信任相关的决策指导  ●使用质量维度的分析  ●变更数据采集规则  ●分析指导  ●一致性及转换规则  ●导入异常处理  ●利益相关方和数据管理人员角色和责任  ●针对业务要求的验证  ●数据备份和恢复的风险缓解规划  ●数据归档过程  ●数据隐私保护  ●用于不同用户组的数据访问矩阵 |

【成熟度等级说明】

组织需要将数据集成的相关内容的格式予以明确规定，并形成统一文件，并作 为未来实施的依据，数据集成计划可包括对组织数据现状的概述、需求分析、集成 目标、实施方案等内容。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.2 | 组织所使用的数据集成规则和工具规定了  批量数据传输和载入、变更数据采集、版本控 制和配置、元数据采集和管理以及在线数据质 量检查和控制 | ●数据集成规则  ●一致性及转换规则  ●变更数据采集规则  ●使用质量维度的分析 |

【成熟度等级说明】

从来源到目标的数据传输通常需要复杂的规则、数据转换和标准化。为了以最 小的风险持续完成这些任务，组织应使用可支持移动数据采集的专用工具，并且这 些工具也可以用于数据和转换的基本分析。组织使用这些工具促进了与ETL 过程相 关的元数据的维护，并提供了以受控的方式进行变更的方法。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.3 | 建立并遵守变更管理过程，以确保协调和控 制集成环境的变化，包括母资源和子目标 | ●变更管理规则  ●数据集成资源清单 |

 **■云端数据治理■**

【成熟度等级说明】

操作系统的变更通常对系统存储或管理的数据和其他子系统使用的数据有着直 接影响。子系统的实例包括依赖于源系统数据的数据库(例如数据仓库)和其他操 作系统。组织应根据自身业务需求建立变更管理过程并执行。

|  |  |  |
| --- | --- | --- |
| 序号 | 等级要求 | 实例 |
| 2.4 | 建立并遵循处置过程，以解决异常情况 | ●风险处置计划  ●数据备份和恢复的风险缓解规划 |

【成熟度等级说明】

组织需要有信心对其系统进行有效管理，通过实施风险事件(包括负载中断或 数据验证失败等)处置计划来保持数据的可用性和完整性。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.5 | 实施集成验证并确保将架构和接口规格要 求形成文件，并在发布之前满足这些规范 | ●数据集成标准  ●验证与确认计划  ●集成测试环境  ●应用程序编程接口  ●数据集成政策 |

【成熟度等级说明】

API(Application Programming Interface, 应用程序编程接口)是一些预先定义 的函数，目的是提供应用程序与开发人员基于某软件或硬件的以访问一组例程的能 力，而又无须访问源码，或理解内部工作机制的细节。API 本身是抽象的，它仅定 义了一个接口，而不涉及应用程序如何实现的细节。API 文档则是对 API 的详细说 明，用于用户的二次开发，或者运行维护。组织应建立数据集成的技术文档，通过 实施数据集成验证确保达到规范要求。

**Level 3: 定义级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.1 | 组织应遵循一套标准的实践方法和规则进 行数据集成活动 | ●数据集成标准  ●数据集成实践方法 |

【成熟度等级说明】

在建立了数据集成的相关实践方法，比如对集成模式、集成建模、集成架构的 标准化操作模式和规则后，组织应按要求执行，以确保达到业务目标。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.2 | 质量检查被定义为组织集成标准的一部分， 并作为数据集成过程的一部分执行 | ●数据集成标准  ●质量检查目标 |

**第7章■云端数据操作■**

【成熟度等级说明】

质量检查是数据集成过程中的重要环节。例如， ETL 过程应执行基本级别的质 量控制检查，特别是对于后续的变更数据采集载入，以确保集成的数据满足由组织 定义的质量属性。这些检查应定义在异常状态下数据被拒绝、标记或排序的标准。

在集成时未能执行检查，可能会导致数据长时间未知并在后续需要执行更多的 清理工作问题。组织可能执行验证以下任务。

●数据是被分析的；

●根据属性，数据是完整和准确的；

●数据丢失处理的逻辑是建立的；

●如果数据不能从首选来源处获取，存在着其他来源；

●生成并监控了数据质量和错误处理报告。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.3 | 建立并遵循标准的过程，以基于使用情况、 要求和选择的触发因素，与业务伙伴一起创建 并验证数据优先级规则 | ●数据优先级规则 |

【成熟度等级说明】

通常情况下，多来源有助于组织的数据集成到一个单一的逻辑目的地，比如有 助于数据仓库所归属的操作系统。在这些情况下，可以预估一个以上的数据来源所 提供的数据可能会与来自组织其他来源的数据相同或类似。在实践中，在提供单个 存储库的最终 ETL 过程中，有必要建立和使用优先级和一致性规则，以指导最终的 集成。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.4 | 根据支持重复使用的架构标准，指定集成接 口的开发和部署 | 数据集成架构标准 |

【成熟度等级说明】

组织应将接口纳入完整的数据包内，并根据组织的配置管理过程和标准进行管 理。实践中，包括接口集成和公用数据服务，组织需要明确规定数据集成规则和性 能要求。组织应记录系统之间的接口，以支持未来的变化，并提供适合集成到另外 的系统或在另外的系统进行集成的方法。这些接口的开发应遵循用于定义和记录应 用程序编程接口(API) 的通用编程实践标准。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.5 | 收集并分析集成接口性能测量指标，以确定 不符合标准和指标的情况 | ●数据集成报告 |

 **■云端数据治理■**

【成熟度等级说明】

组织应根据业务需求收集接口相应指标并进行分析，并对分析结果进行统计。 例如，执行缺陷跟踪，以监测缺陷严重程度、原因、统计发生和计划决议。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.6 | 组织通过数据治理过程，来记录并管理数据 源和目的地变更 | ●验证和确认结果  ●性能要求  ●性能测量指标和分析结果  ●数据质量持续改善的措施和测量指标  ●数据传输策略和SLA  ●集成方法标准  ●集成最佳实践指导  ●标准接口规范  ●集成环境变更管理过程 |

【成熟度等级说明】

组织在数据治理过程中，应对变更管理加强监督，记录相关数据，包括数据源 和目的地变更。组织需要建立变更管理策略，在对数据源和目的地进行变更时要遵 循相应策略，取得授权，按着规划操作，并予以记录。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.7 | 所有结构化和非结构化数据应对其他平台 可用，确保不同平台间能够对数据进行处理 | ●数据互操性要求 |

【成熟度等级说明】

非结构化数据包括文本、交互数据(例如用户行为记录，表达的对人、产品的 看法和观点)图像、视频、音频等。

数据创造的主体包括组织和用户个体，而个体所产生的绝大部分数据均为图片、 文档、视频等非结构化数据；另一方面，信息化技术的普及使得组织更多的办公流 程通过网络得以实现，以往纸质的表单、票据等现在都实现了数字化存档，而这方 面产生的数据也以非结构化数据为主，非结构化数据也在客户关系管理 (CRM) 系 统中普遍存在，特别是客户服务代表和呼叫中心的工作人员的笔记记录。再例如 Web 页面，它通常被认为是一个典型的非结构化数据。尽管基本上所有的网页都是由 HTML 语言组成的，具备丰富的结构定义。但是Web 页面还包含链接和引用外部的 内容，而这些内容往往是非结构化的内容，如图像、XML 文件、动画等。

对于结构化数据，要想实现互操作性通常比较容易。非结构化数据的来源包括 用户使用云服务过程中产生的非结构化数据、存放在云平台上的非机构化数据等， 数据格式差异较大，在不同的云平台间处理，需要制订互操作性要求，通过预处理

**第7章■云端数据操作**

实现数据的大体一致。

**Level** **4:量化级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 4.1 | 集成度量的统计分析影响集成接口更改的 决策 | ●变更标准 |

【成熟度等级说明】

组织应在接口性能分析中指定并使用变更标准(即容差极限)。例如，数据概要 分析结果可表明与集成相关的缺陷量在可接受的容差范围内。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 4.2 | 完全整合并集中管理所选的高度共享的数 据，并根据需要将其传输至集成数据存储区 | ●统计分析结果  ●数据概要分析  ●集成高度共享的数据并持续改进 |

【成熟度等级说明】

对于依赖于高度共享的数据(例如参考数据、主数据等)的组织业务过程，关 键实体类型(例如客户、帐户、产品、组织等)的重要信息差异(时效性、代表性 和价值方面的差异)会带来很大影响。数据集成的具体目标是简化、标准化、集中 化提供多种业务和多个业务目标的共享数据。

**Level** **5:** **优化级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 5.1 | 定期评审数据集成模型，并将其作为改进的 信息输入 | ●数据集成模型  ●数据改进方案 |

【成熟度等级说明】

组织可根据自身实际情况，按照固定周期组织相关方对数据集成模型进行评审， 以达到持续改进的目的，确保满足业务目标和相关需求。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 5.2 | 组织在其行业内发布并分享整合最佳实践 | ●定量模型  ●性能触发因素和阈值  ●根本原因分析结果  ●演示材料、白皮书或发表的文章 |

【成熟度等级说明】

组织在数据集成活动上应该积极分享实践成功经验，为行业最佳实践做贡 献。

 **■云端数据治理**

**4.** **实施案例**

(1)文件名称列表

《数据集成标准》

(2)记录文件列表

《数据测试报告》

(3)实施案例

|  |
| --- |
| 数据跨域互操作技术规范  1 范围  本规范规定了××项目内数据跨域互操作接口采用的协议、连接方式、调用参数以及数据的返回格式。  2规范性引用文件  下列文件对于本文件的应用是必不可少的。凡是注日期的引用文件，仅所注日期的版本适用于本文件。凡是不 注日期的引用文件，其最新版本(包括所有的修改单)适用于本文件。  GB 18030—2005信息技术中文编码字符集  3 术 语 和 定 义  下列术语和定义适用于本规范。  (1)资源(resource)  可以被标识的实体对象或服务。  在本规范准中，特指可被标识的数据集、数据或服务。  (2)数据集(dataset)  由相关数据组成的可标识集合。 一个数据集可能是一个较小的数据集合，在物理上或逻辑上位于一个较大的数 据集之内；反之， 一个数据集也可能由若干数据集组成，是这些子数据集的父数据集。  本规范中，数据集指XX项目的主题数据库、参考数据库、专题数据库、专业数据库及其各级子库。  (3)数据(data)  对事实、概念或指令的一种形式化表示，适用于以人工或自动方式进行通信、解释或处理。  本规范中，数据是指所有能输入到计算机并被计算机程序处理的符号介质的总称，是用于输入电子计算机进行 处理，具有一定意义的数字、字母、符号和模拟量等的通称。  (4)服务(service)  本规范中，服务是指××项目中数据集满足应用系统或者人的需要时，与之发生的“直接”接触行为及有关结果。  (5)标识符(identifier)  用于标识数据对象或服务项目的一组字符。  4符号与缩略语  (1)URI统一资源标识符(Uniform Resource Identifier)  (2)URL统一资源定位符(Uniform Resource Locator)  (3)XML可扩展标记语言(Extensible Markup Language)  5 数据格式定义  1)接口的编码方式及响应格式  (1)接口编码及响应方式 |

**第7章■云端数据操作**

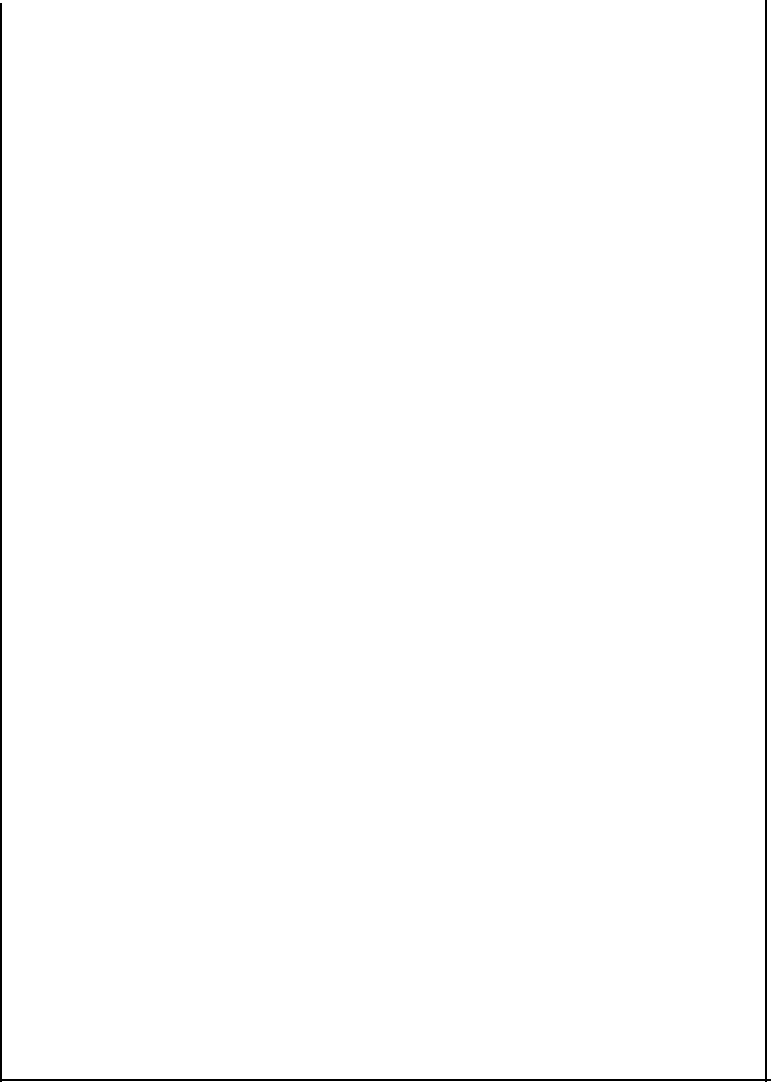
续表

接口响应的数据基于xml 语言格式，编码方式采用自定义编码的方式(常用编码方式包括：UTF-8、GBK、GB2312、 ISO-8859-1等),接口响应数据所使用的编码方式必须在返回XML 头信息的encoding属性中说明。

(2)接口响应请求状态码

以下状态码只列出接口中遇到的普遍问题，可根据数据库建库规范中规定的范围增加。

|  |  |  |
| --- | --- | --- |
| 序 号 | 通用状态码 | 相 关 说 明 |
| 1 | 200 | 服务调用成功 |
| 2 | 400 | 请求格式错误 |
| 3 | 401 | 未授权访问 |
| 4 | 402 | 不可识别的verb |
| 5 | 403 | 服务器拒绝访问 |
| 6 | 404 | 指定的资源未找到 |
| 7 | 408 | 请求超时 |

2)科学数据交换格式

(1)科学数据交换格式定义

科学数据交换格式：SDEF(Scientific Data Exchange Format),SDEF 使用统一的格式对数据进行封装，为资源 共享与透明访问提供保障。

(2)科学数据交换格式支持的数据类型

<value>元素与<title>元素中的数据值相同， <tile> 元素可省略。

6 采用协议

1)HTTP1.1 (超文本传输协议)

超文本传输协议 (HTTP) 是一种分布式、合作式、超媒体信息系统。它是一种通用的、无状态 (stateless) 的

协议，除了应用于超文本传输外，它也可以应用于诸如名称服务器和分布对象管理系统之类的系统，这可以通过

扩展它的请求方法、错误代码和报头来实现。HTTP的一个特点是数据表现形式是可输入的和可协商性的，这就允 许系统能被建立而独立于数据传输。

HTTP 是一个客户端和服务器端请求和应答的标准(TCP)。客户端是终端用户，服务器端是网站。通常，由HTTP 客户端发起一个请求，建立一个到服务器指定端口(默认是80端口)的TCP 连接。HTTP 服务器则在那个端口监 听客户端发送过来的请求。 一旦收到请求，服务器(向客户端)发回一个状态行，比如 “HTTP/1.1200OK”, 和 (响应的)消息，消息的消息体可能是请求的文件、错误消息、或者其他一些信息。

2)接口安全

采用基于IP 地址的身份验证方式。在服务调用的过程中，服务提供者获取调用者的ip 地址，在本地保存的授权

访问 ip 地址列表中查询，认证通过执行服务返回数据，不通过则拒绝服务。

3)连接方式

(1)POST 方式(必须实现):

POST 方法被用于请求源服务器接受请求中的实体作为请求资源的一个新的从属物。

POST方法的实际功能是由服务器决定的，并且经常依赖于请求URI(Request-URI) 。POST提交的实体是请求URI 的从属物，就好像一个文件从属于一个目录， 一篇新闻文章从属于一个新闻组，或者一条记录从属于一个数据库。

 **■云端数据治理**

续表

|  |
| --- |
| POST方法的响应是可缓存的。  (2)GET方式(可选):  GET方法意思是获取被请求URI(Request-URI)指定的信息(以实体的格式)。如果请求URI涉及到一个数据  生成过程，那么这个生成的数据应该被作为实体在响应中返回，但这并不是过程的资源文本，除非资源文本恰好 是过程的输出。  GET请求的响应是可缓存的。  4)技术实现  接口采用OpenURL技术实现。 |

**7.2.3** **数据生命周期管理**

**1.** **简介**

1)目的

确保从创建或采购到有效删除的整个数据生命周期中，组织理解、映射、清算 以及控制业务过程中的数据流。数据生命周期管理能够更好地帮助组织对数据风险 的管理以及支持数据质量改进，特别是当涉及到大量数据、高速的云端数据迁移以 及复杂又相互依赖的数据分享过程时。

2)介绍

数据是流动的，在其存储空间流进和流出，并被包装在信息产品中交付使 用。数据本身存在着从产生到消除的生命周期，在数据的生命周期中，数据流 转才能产生价值，而且在不同阶段数据的价值会发生变化，数据的采集粒度与 时效性、传输过程、存储方式、数据处理、数据分享等都会对数据的价值体现 产生影响。云端数据生命周期管理是依据不同数据在不同阶段的价值实施不同 的管理策略，以用于管理数据在整个生命周期内的流动。数据生命周期管理是 一种信息管理模式，通过它可以有效控制云端数据规模，提高数据访问效率， 从而提高系统运行的整体效率，帮助组织在数据生命的各个阶段以最低的成本 获得最大的价值。

数据生命周期中数据的生命过程没有统一的顺序规定，尤其是在数据流转更 加复杂多变的大环境下，数据不再单纯地遵循一个固定的顺序过程。在整个生命 过程中，数据可能被提取、导入、迁移、验证、编辑、更新、转型、转换、整合、 隔离、汇聚、引用、评审、报告、分析、挖掘、备份、恢复、归档和检索，最后 被删除。

通过将数据处理过程归类，可以将数据生命周期划分为六个主要阶段，即产生、 存储、使用、传输、共享和销毁组成。需要指出的是，并非所有的数据都会经历生 命周期的每一个阶段。各阶段的定义如下。

**第7章■云端数据操作■**

(1)数据产生：指新的数据产生或现有数据内容发生显著改变或更新的阶段；

(2)数据存储：指非动态数据以任何数字格式进行物理存储的阶段；

(3)数据使用：指数据被查看、处理以及不包括修改在内的其他各种使用方式；

(4)数据传输：指数据从一个实体通过网络或其他途径流动到另一个实体的 过程；

(5)数据共享：指数据在不同实体之间发生交互的阶段；

(6)数据销毁：指利用物理或者技术手段使数据永久或临时性不可用的过程。

数据是组织的一项重要资产，对提高组织决策水平起到重要作用。但是随着时 间的推移，数据规模日益庞大，给系统存储、访问效率造成了越来越大的压力。 一 般情况下，数据生成时间越长，其获得的访问频率则越低。尤其是在大数据时代， 数据体量非常之大，数据访问效率会变得更低。如果将不活跃的数据进行迁移，则 可以降低组织的拥有成本。

数据生命周期是以成本与数据不断变化的价值相对应的办法，从数据的创建到 最终销毁进行全程的管理，从而允许组织在满足业务需求以及法律法规的前提下制 订明确的保留时间表，对实时云端数据进行管理识别出哪些数据是有价值的，哪些 数据需要被永久保存，有效控制数据在线规模，不断提高数据的访问效率，降低存 储成本。

3)相关过程域

数据提供者管理过程域提供了从各个层面上定义数据需求的相关信息；

数据质量战略过程域为指导哪些数据需要特别管理提供了依据；

主数据管理过程域提供了数据生命周期管理的范围与重点。

**2.GQM 表**

|  |  |  |  |
| --- | --- | --- | --- |
| 目 标 | 问 题 | 度 量 项 | 度量项说明 |
| G1:定义和维护 与所选业务过程相 关的数据生命周期 以反映变更 | Q1:从业务过程映  射到为支持这些过程 而创建和维护的数据 时，定义了什么活动、 里程碑还有产品 | M1:数据生命周期 | 定义和维护与所选业务过程相关的 数据生命周期以反映变更 |
| M2:度量和度量项 | 定义度量和度量项，收集相关信息以 评估过程到数据映射的进度以及对权  威数据来源的采纳 |
| M3:数据生命周期管  理过程 | 定义数据生命周期管理过程，并由相 关者评审、由数据治理主体和过程进行 管理 |
| M4:符合性 | 定义和维护业务过程到数据的映射 并定期评审符合性 |

 **■云端数据治理**

**续表**

|  |  |  |  |
| --- | --- | --- | --- |
| 目 标 | 问 题 | 度 量 项 | 度量项说明 |
| G2:基于一个识 别共享数据流并对 数据流进行优先次 序排列的框架，将 业务过程映射到数 据流，这种映射在 特性层面延伸到整 个数据生命周期 | Q2:组织是否建立 了清晰的角色与分工 以创建和维护业务过 程到数据的映射 | M 5 : 需 求 | 数据生成者和使用者的需求得到映 射和调准 |
| M 6 : 映 射 | 基于一个识别共享数据流并对数据  流进行优先次序排列的框架，将业务过 程映射到数据流，这种映射在特性层面 延伸到整个数据生命周期 |
| M 7 : 改 进 | 定期分析度量项和相关者的反馈，从 而将改进引入到数据依赖的管理中 |
| M8:权威数据源 | 由数据治理人员来评审和审批权威 数据源的指定 |
| M9:数据流映射 | 在组织级别上，针对每个主要业务过 程，为共享数据实施数据流和完整数据 到过程生命周期的映射 |
| M10:职责 | 项目在系统开发生命周期活动上的  职责包括将数据特性映射到对组织重  要的业务过程、共享数据集、来源以及 目标数据集上 |
| M11:共享数据表示法 | M11:系统开发生命周期过程要求有  参考、采用经认可的共享数据表示法以 及从权威来源处获取数据 |
| Q3:是否使用了标 准的过程建模方法和 工具来为业务过程进 行定义和建模 | M12:标准过程 | 在组织里采用标准过程进行数据生  命周期的影响分析，以识别和预估接口  和数据集的变更，并为其进行进度安排 |
| M13:度量项 | 使用度量项来扩充经认可的共享数 据的重用并消除过程的冗余 |
| G3:定义和维护 数据影响、依赖以 及相互依赖 | Q4:治理人员是否 负责业务过程数据需 求、映射和优先次序 排列的管理和它们之 间的协调 | M14:变更 | 由数据治理机构来管理为特定业务 目标实施的共享数据集或目标数据集 的变更，并需要内部相关者的参与 |
| M15:数据生命周期度 量项 | 定期完善数据生命周期度量项并由 高管进行评审 |
| M16:经验 | 组织就数据管理生命周期过程与行 业和同行分享经验 |

**第7章■云端数据操作**

**3.** **成熟度等级介绍**

**Level 1: 初始级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 1.1 | 定义了业务过程中的数据生命周期，并实 施了数据生命周期管理 | ●数据生命周期划分依据 |

【成熟度等级说明】

为更好地管理数据，组织应定义业务过程中的数据生命周期。定义包含了从元 数据创建或获取到转换再到完成目标的可追溯性。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 1.2 | 识别了数据生态中数据流通途径，定义了 数据产生者和数据消费者 | ●数据产生者  ●数据消费者  ●数据生产者与数据消费者框架图 |

【成熟度等级说明】

数据是流动的，从数据产生到数据消费经过多个角色，如数据产生者、数据服 务提供者、数据框架提供者和数据消费者等。组织在进行数据生命周期管理前，应 该识别数据的流通途径，区分数据产生者和数据消费者。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 1.3 | 利益相关方在数据元素的范围以及权威的 数据来源上都达成一致意见 | ●数据生命周期框架和项目计划  ●数据集的权威数据源和已批准属性的列表 |

【成熟度等级说明】

组织应该至少在一个关键的共享实体或对业务很重要的数据集(例如使用实体、 交易历史等)上定义范围和权威来源。

**Level 2: 管理级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.1 | 依据数据消费者需求，定义数据源需求 | ●数据消费者需求列表  ●数据源需求列表 |

【成熟度等级说明】

组织应建立一个能够准备快速识别数据消费者需求的机制和过程，从而将数据 消费者需求映射到数据源上，有针对性地对数据进行管理。

**云端数据治理**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.2 | 定义和维护业务过程到数据的映射，并定期 评审符合性 | ●业务过程数据映射  ●数据产生者和消费者间的映射  ●分享特性的业务过程目录和映射图 |

【成熟度等级说明】

组织应根据业务需求定义业务过程到数据的映射并不断改进。组织应在固定周 期内对数据映射进行评审，以保证达到业务目标。使用有效的工具包括元数据库、 数据建模工具、业务过程建模工具，以及他们相关的组件和模型。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.3 | 在项目级别上，制订数据生命周期的变更控 制和管理流程 | ●数据生命周期管理流程  ●数据生命周期变更控制策略  ●数据变更管理过程文档 |

【成熟度等级说明】

数据生命周期过程管理是数据管理活动的集合，数据在不同阶段需要的管理方 式存在差别。在数据产生阶段，管理的主要目的是确保能够收集到符合业务目标的 数据集，在处理阶段数据管理的主要目的是提高数据质量。在数据归档和销毁阶段 除了要满足业务需求外更多地关注合规要求。而在相同的数据生命周期阶段，不同 类型的数据管理方式也存在很大差别。针对各式各样的数据，组织需要建立普遍适 用的管理策略，并且能够基于不同需求调整数据生命周期管理流程，但应当制订相 应的变更控制措施。

**Level 3: 定义级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.1 | 定义数据生命周期管理策略，并由相关方评 审，由数据治理主体对其进行管理 | ●数据生命周期管理策略  ●数据生命周期管理策略评审记录  ●数据生命周期管理策略维护记录 |

【成熟度等级说明】

数据生命周期管理是依据不同数据在不同阶段的价值实施不同的管理策略 的，以简单、可靠、经济、有效的方式使企业获取数据的最大价值，从而有效 降低组织 IT 的总体拥有成本。数据生命周期管理是一种数据管理模型，即对数 据进行贯穿其整个生命周期的管理，从创建和使用到归档和处理，它是一种针 对数据进行主动管理的过程策略，组织应根据自身需求定义数据生命周期管理 过程。

**第7章■云端数据操作■**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.2 | 在组织级，建立和维护全数据生命周期的变 更管理过程 | ●数据生命周期管理策略  ●数据生命周期管理策略评审记录  ●数据生命周期管理策略维护记录 |

【成熟度等级说明】

组织应建立全数据生命周期变更管理过程并保持不断改进。举例来说，保管人 (物理的)和数据管理员(逻辑的)使用一种标准的评估和影响分析过程来处理数据 元素的变更——添加、删除和修正。对源到目标的映射也进行相应变更。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.3 | 明确定义的数据生命周期管理角色和职责 | ●数据生命周期管理角色和职责  ●数据生命周期管理角色培训计划 |

【成熟度等级说明】

组织已经配备了专门的人员负责数据生命周期管理工作，明确定义了其职责范 围。组织应采取措施保证人员能力能够持续满足数据生命周期管理职位需求，例如 采用培训教育、经验交流等方式。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.4 | 在组织级别上，针对主要业务过程，实施从 数据流和完整数据到共享数据的全生命周期 过程的映射 | ●过程到数据映射的模板  ●数据映射的项目计划  ●关键业务过程的生命周期数据映射 |

【成熟度等级说明】

组织在主要业务过程中，实施从全数据到共享数据的全生命周期过程的映射。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.5 | 由数据治理机制来管理为特定业务目标实  施的共享数据集或目标数据集的变更，并需要 相关方和内部人员的参与 | ●数据变更需求  ●已识别的数据特性源  ●数据特性的权威数据源 |

【成熟度等级说明】

组织应通过数据治理机制，组织相关方(包括内部人员),管理为达成特定业务 目标而实施的对共享数据集或目标数据集的变更。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.6 | 定义数据生命周期管理过程的度量项，并用 来指导管理过程 | ●数据生命周期管理度量项 |

**■云端数据治理**

【成熟度等级说明】

组织应当根据自身需求制订一些度量指标来衡量数据生命周期管理的规范程 度，并根据这些度量项指导数据生命周期管理过程。

**Level 4: 量化级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 4.1 | 定义目标和指标，并用来提高数据生命周期 管理过程的有效性 | ●数据生命周期分析标准 |

【成熟度等级说明】

为确保对数据生命周期的管理有效，组织应使用标准的过程方法，包括编写影 响分析报告，并根据分析结果实施针对接口和数据集变更的相应措施。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 4.2 | 定义度量项，并采用统计学方法分析数据生 命周期管理 | ●数据生命周期管理度量项  ●度量项的记录和结果  ●数据生命周期管理改进计划  ●数据生命周期管理改进过程记录 |

【成熟度等级说明】

组织应根据自身业务实际情况，使用度量项来改进共享数据在使用过程中出现 的冗余情况，降低数据拥有成本，提高数据访问效率。

**Level 5: 优化级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 5.1 | 定期分析度量项和相关方的反馈，从而改进 数据生命周期管理 | ●数据生命周期度量项  ●相关方反馈  ●数据生命周期管理改进计划  ●数据生命周期管理改进过程记录 |

【成熟度等级说明】

组织应根据自身业务实际情况，按照固定周期将数据生命周期的度量项和相关 方反馈作为输入内容进行分析评审。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 5.2 | 定期完善数据生命周期度量项并由高管进 行评审 | ●数据生命周期度量项  ●根据统计分析结果向高管进行汇报 |

【成熟度等级说明】

组织根据自身实际情况，按照固定周期不断更新完善数据生命周期度量项。此

**第7章■云端数据操作■**

过程需要组织高层亲自参与评审，以确保符合组织自身实际情况，并与数据管理目 标相一致，最终实现业务目标。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 5.3 | 组织就数据管理生命周期过程与行业和同 行分享经验 | ●交流数据生命周期管理过程经验的公共展示、白 皮书或相似的报告 |

【成熟度等级说明】

组织在数据生命周期活动上应该积极分享实践成功经验，为行业最佳实践做 贡献。

**4.** **实施案例**

(1)文件名称列表

《数据备份与恢复管理制度》

(2)记录文件列表

《数据依赖报告》

(3)实施案例

|  |
| --- |
| 数据备份与恢复管理制度  1 总 则  1.1 目的  为了规范公司数据备份清单的建立、备份的职责、备份的检查、以及系统受到破坏后的恢复工作，合理防范计 算机及信息系统使用过程中的风险，特制订本制度。  1.2适用范围  本制度适用范围为公司总部及下属所有分(子)公司所有数据的备份与恢复管理工作。  2 职 责 分 工  2.1 信息管理室是公司信息备份的归口管理部门，负责公司所有重要信息备份的管理和协调。  2.2各部门负责人是确定本部门重要备份数据的责任人，部门负责人应定期向信息管理室提供本部门重要数据 的备份或最新的备份。  3备份清单建立和备份方式  3.1 各部门人员将需要备份的数据清单填到《部门数据备份清单》中，然后提交到信息管理室。信息管理室汇 总各部门上报的《部门数据备份清单》,并核对网络上的存储数据，形成《公司数据备份清单》。  3.2根据公司情况和《公司数据备份清单》将数据分为一般数据和重要数据两种。  3.2.1一般数据主要指：个人或部门的各种信息及办公文档、电子邮件、人事档案、考勤管理、监控数据等。  3.2.2重要数据主要包括：各部门日常表单记录、生产数据、财务数据、技术部门图纸、标书、合同等。  3.3 备份方式：信息管理室负责根据数据的重要性及保存期限等情况，为各类数据设定适当的备份方式。备份 方式有服务器备份、在线备份、光盘自动归档备份等。  3.4一般数据由各部门每月自行备份，部门负责人整理归档后刻盘，信息管理员每半年对一般数据资料进行选 择性收集归档。 |

 **■云端数据治理■**

续表

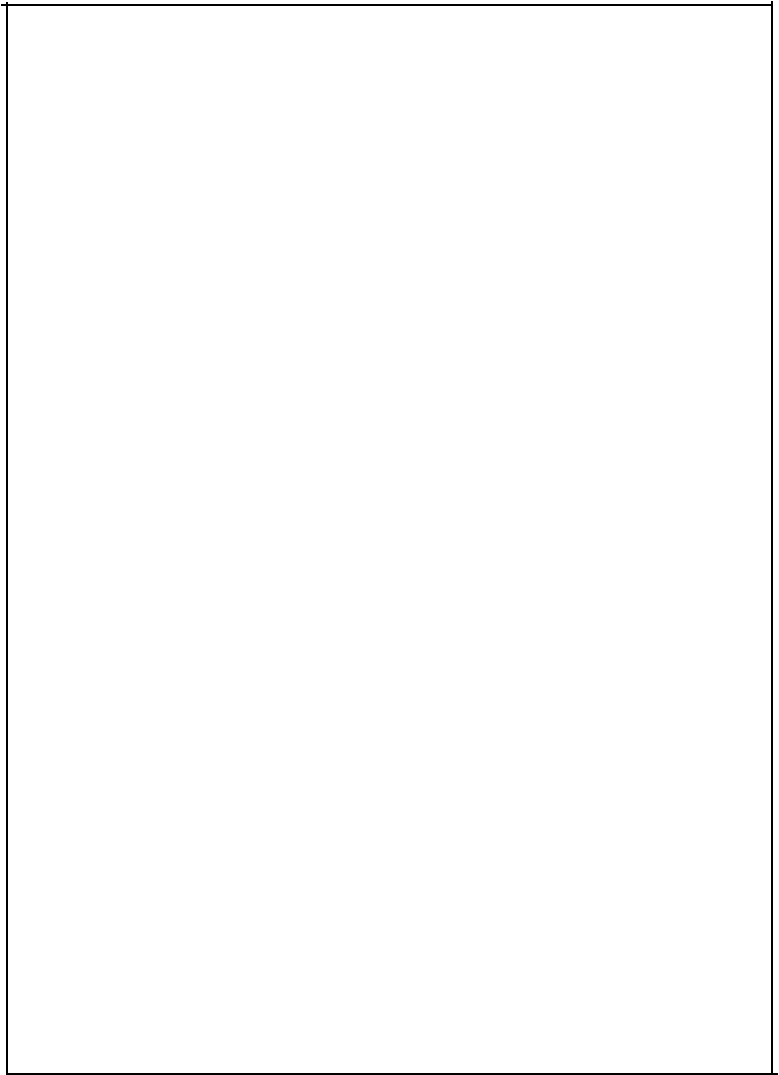
|  |
| --- |
| 3.5 重要数据由各部门和信息管理室共同负责，实行双重备份和异地备份机制，重要数据每日0:00进行自动 备份。服务器的数据库由信息管理员在服务器硬盘每周做软件备份。当服务器、交换机及其他系统主要设备配置 更新变动，以及服务器应用系统、软件修改后均要在改动当天进行备份。  3.6 为了加大数据备份的相对安全信息管理室应定期(每年至少1次)通过服务器或光盘做永久性备份。不同 时期的各备份介质做好相应的编号及相关说明，并移交公司档案室存档。  3.7 自动备份失败情况下设定邮件通知，并采取手工备份或其他方式及时备份；定期对备份日志的进行检查， 对重要数据实现异地备份(或备份介质异地存放)。  4 备份介质的存放和管理  4.1 备份介质的保存：备份介质应统一保存在安全可靠的地方，保存地点应能防火、防盗、防潮。备份介质必 须有由专人负责进行存取，其他人员未经批准不能操作。  4.2 备份介质的检查：信息管理室应定期(每年至少1次)检查各类备份资料的情况，及时更新或销毁过期的 资料，并做好检查记录；对损毁的备份资料，要及时与使用部门联系，共同研究补救措施和方法；对于到达资料 保存期限的备份资料，应在得到相应部门负责人批准后及时销毁。  4.3 所有备份介质一律不准外借，不准流出公司，任何人员不得擅自取用，若要取用须经分管领导批准，并填 写《备份介质借用登记表》。借用人员使用完后，应立即归还。由信息管理员检查，确认完好。信息管理员及借用 人员须分别在《备份介质借用登记表》上签字确认介质归还。  4.4 长期保存的备份介质，必须按照制造厂商确定的存储寿命定期转储。需要长期保存的数据，应在介质有效 期内进行转存，防止存储介质过期失效。  4.5 存放备份数据的介质必须具有明确的标识；标识必须使用统一的命名规范，注明介质编号、备份内容、备 份日期、备份时间、保留期限等重要信息。  4.6 恢复性测试：对于备份系统，应至少每年测试一次(模拟灾害发生时进行演练)。  5 灾害恢复  5.1 硬件的灾害恢复：信息管理室应保存公司所有计算机、服务器及其辅助设备清单；信息管理室应保留所有 硬件的供应商目录及详细资料。 一旦发生灾害，供应商在接到公司的通知后应在两个工作日将设备配备到位；在 与信息系统供应商签定的采购合同中，必须有相关备份、灾害恢复与技术支持条款，并要求供应商出具安全保密 承诺书；计算机系统硬件受到破坏，使得系统无法正常工作时，各部门应及时将情况反馈给信息管理室，并协助 信息管理室做好原因调查工作，属于设备供应商问题的，应追究设备供应商的赔偿责任；同时，信息管理室及时 提出所需设备的采购申请，并将这些设备安装到公司事先指定备用办公地点或依据当时情况指定临时地点，以便 尽快恢复正常业务。  5.2 软件的灾害恢复：公司信息系统及计算机所使用的软件及数据备份的保存按数据备份的要求进行， 一旦发 生系统灾害，负责系统管理的人员通过灾害发生前或其他信息系统所保存的各种备份介质，依据备份标记进行目 的性的恢复；对所有发生的硬件和软件灾害，必须对灾害发生及恢复情况进行书面记录；对于重大灾害，应立即 报告分管副总和总经理。  5.3 一般数据发生灾害时能恢复到上月最后一天，重要数据发生灾害时能恢复到前一天0:00。  5.4 公司应定期(每年至少1次)对公司信息系统的相关风险，及所采取的风险管理实施是否充分适当进行评 估，该项工作由信息管理室牵头，各部门相关人员参加。  6 附则  6.1 本制度由公司总部信息管理室负责解释和修订。  6.2 本制度自发布之日起开始执行。 |

**第7章■云端数据操作**

续表

|  |
| --- |
| 数据管理办法  1 总则  1.1为适应集团信息化发展要求，充分利用数据资源为生产、经营、管理和决策服务，保证各类信息合理、有 序流动和信息安全，确保集团信息化建设快速协调有序安全发展，根据国家有关法律法规以及《集团信息安全管 理办法》等规定，特制订本管理办法。  1.2本办法适用于集团各职能部室、直属和特设机构、专业化公司、事业部、区域公司及其所属各单位(以下 简称各单位)。  2 管 理 范 围  本办法管理范围包括：各单位与生产、经营、办公、安全等相关的应用系统和数据，以及为其提供支撑的基础 设施资源、计算存储资源和办公终端资源等。  3组织机构和工作机制  3.1集团信息化领导小组是集团数据资源管理体系的最高层，负责审定集团有关数据资源管理的规章、制度、 办法，负责审核有关标准、规范、重要需求等。  3.2集团信息化领导小组办公室(以下简称集团信息办)负责集团数据管理的监督、检查和考核，指导集团数  据管理工作，查处危害集团数据安全的事件。各单位负责本单位数据的采集、传输、使用、安防、备份等管理工 作。  4数据分级管理  4.1根据数据在生产、经营和管理中的重要性，结合有关保密规定，按照集团级应用系统和数据、厂矿级应用 系统和数据、区队(车间)级应用系统和数据分别制订管理标准。  4.2 集团级应用系统和数据，技术管理由集团信息办负责，业务管理由相关业务处室负责，运维管理由XX公 司负责。厂矿级应用系统和数据由各单位信息管理部门管理，集团需要利用的管理数据和生产数据要同步上传到 集团数据中心。区队(车间)级应用系统和数据由各单位信息管理部门管理和维护。  5数据标准管理  5.1集团信息办负责集团数据编码和接口标准的统一规划和标准制订，负责对集团及各单位应用系统的数据标 准管理进行引导和考核。各单位新建应用系统应严格执行集团下发的数据编码和接口标准，应用系统应根据自身  实际逐步按照集团标准进行完善。  5.2数据编码和接口标准应符合以下要求  (1)数据编码应能够保证同一个对象编码的唯一性及上下游管理规范的一致性  (2)接口应实现对外部系统的接入提供企业级的支持，在系统的高并发和大容量的基础上提供安全可靠的接入； (3)提供完善的数据安全机制，以实现对数据的全面保护，保证系统的正常运行，防止大量访问，以及大量占  用资源的情况发生，保证系统的健壮性；  (4)提供有效的系统可监控机制，使得接口的运行情况可监控，便于及时发现错误并排除故障；  (5)保证在充分利用系统资源的前提下，实现系统平滑的移植和扩展，同时在系统并发增加时提供系统资源的  动态扩展，以保证系统的稳定性；  (6)在进行扩容、新业务扩展时，应能提供快速、方便和准确的实现方式。  6数据资源管理  6.1基础设施资源集中管理。为了避免信息机房等基础设施资源重复投资建设，造成资金浪费、设施利用率低 等问题，各单位应充分利用集团数据中心资源，集团信息办负责统一协调集团及各单位的基础设施资源。  (1)各单位未经集团批准不得私自新建、改建、扩建信息机房。 |

**■云端数据治理**

续表

(2)集团数据中心要按照《集团机房建设技术规范》建设，满足各单位应用系统及数据统一到集团数据中心所

需的各项使用要求。

(3)各单位现有机房自行管理、统一管控。各级信息管理部门作为主要责任部门，要保证信息机房各项运行指

标达到集团要求。

6.2 计算存储资源集中管理

为了消除“信息孤岛”,实现集团数据共享和集成，提升数据安全防护等级，各单位所需计算和存储资源，要统

一使用集团数据中心的云计算资源，做到资源集中、高效利用。

(1)现有的集团级应用系统及数据(安全监测系统除外)、各单位应用系统及数据(直接用于生产安全、自动化 控制和监测监控的系统除外)要按照在用服务器、存储的服务年限和系统生命周期科学制订迁移到集团数据中心

的计划和方案，并报集团信息办批准后实施。

(2)新建应用系统原则上不再购置新的服务器和存储，所需计算和存储资源应使用集团数据中心的云计算资源。 各单位如有特殊生产要求，需要购置服务器或存储的，应报请集团领导批准，由集团信息办备案后，按集团采购

管理相关规定执行。

(3)对于当前集团网络不具备实施条件的单位，可向集团提出申请建设集团区域性数据分中心，并根据建设进

度制订应用系统和数据迁移计划。集团区域性数据分中心建成后，新建系统需要集中部署、分级管理。

6.3 办公终端资源集中管理。为了提高办公效率、降低办公成本、实现节能降耗，集团级应用系统要统一使用

集团数据中心云桌面，并在厂矿和区队(车间)级应用系统中逐步实现全面使用。

(1)各单位新建系统所需计算机和新增办公用计算机要使用集团数据中心云桌面。

(2)原有集团推广的应用系统所使用的计算机，以及各单位在用的计算机，分别由应用系统主管部门和各单位

按年度提出云桌面更换计划，逐步完成云桌面更换工作；集团信息办负责协调和监督。

(3)各单位申请云桌面使用，应与××公司签订租用协议，由××公司负责云桌面运维，各单位信息管理部门 负责本单位云桌面管理。

(4)对于当前集团网络不具备实施云桌面替换条件的单位，应协同集团相关部门接入集团网络或建设集团区域 性数据分中心。在网络接入后或集团区域性数据分中心建成后，按计划完成云桌面的部署工作。

6.4 各单位使用资源应按集团规定支付相关费用。 7 数据分析管理

(1)数据分析是采取科学合理的方法，利用现代信息技术手段，对计算机应用系统生成的数据进行分析，充分

发掘数据中蕴含的信息，用数据描述现状，预测趋势，规范生产行为，优化管理流程，加强经营监管，提供决策

支持。

(2)集团信息化领导小组应加强对各单位数据分析的指导，鼓励各单位结合自身实际，充分利用“大数据”技

术，自行组织开发业务选题和数据模型，组织经验交流，提高分析水平。集团信息办要做好数据分析引导和管理

工作，为集团安全生产、经营管理工作服务。基层各单位要充分挖掘和利用现有数据资源，不断探索和创新数据 分析方法，规范数据分析程序，提高数据分析质量，做好本单位各项应用的数据分析工作。

(3)集团级数据分析、处室级数据分析和厂矿级数据分析分别由集团信息办、相关业务处室和各基层单位负责

策划和实施，集团信息化领导小组负责监督和考核。

8 数据应用管理

(1)数据应用是指利用数据分析的成果，查找存在问题，开展业务运转状况评估，提出改进措施，提高管理水

平，规避管理风险。

**第7章■云端数据操作**

续表

|  |
| --- |
| (2)各级信息管理部门应加强数据应用。集团信息办负责代表集团对各单位以及单位之间数据共享应用的统一 规划并制订标准。各单位要严格按部门、按层级落实数据应用工作，对数据进行科学统计、分析、挖掘和应用， 为各级领导决策提供依据。  9数据安全管理  (1)各级信息管理部门应建立数据安全管理制度及相关措施，主要包括：数据访问的身份验证、权限管理及数 据的加密、保密、日志管理、网络安全、容灾备份等。  (2)为统一规范操作权限，各单位应明确工作人员的录入权限、访问权限及维护权限的管理部门，任何人不得 擅自设立、变更和注销。  (3)各级信息管理部门要指定专人负责系统数据及介质资料的安全管理工作。要加强数据库的安全管理，制订 和明确管理员用户和数据查询用户的操作权限及规程。  (4)对数据的各项操作至少要建立运行日志，严格监控操作过程，对发现的数据安全问题，要及时处理和上报。 管理员应掌握和运用数据库访问审计技术，实现对数据库操作的监测和追溯。  (5)各级信息管理部门要加强用户身份验证管理、网络安全管理，采取严格措施，做好计算机病毒的预防、检 测、清除工作，建立针对网络攻击的防范措施，保证数据传输和存储安全。  (6)各级信息管理部门要加强数据的容灾备份工作，建立数据容灾备份机制，保障系统应急恢复和数据溯源。  重要数据要上传至集团数据中心备份。  10 附则  (1)本办法解释权归集团。  (2)本办法自本文印发之日起执行。 |

**第** **8** **章** **云端数据架构**

**8.1** **概述**

**8.1.1** **目的和意义**

IT 架构是由数据架构、应用架构和技术架构共同构成的。其中，数据架构是组 织IT 架构的核心，因为信息系统支撑下的组织业务运作状况，是通过信息系统中的 数据反映出来的，数据是信息系统管理的重要资源。因此构建组织IT 架构时，首先 要考虑数据架构对当前业务的支持。理想的IT 架构规划逻辑是数据驱动的，即：首 先根据业务架构分析定义数据架构；然后根据数据架构结合业务功能定义应用架构； 最后根据数据架构与数据定义，来设计技术架构。

**例子1:数据架构研究程序与数据库的关系**

随着信息技术的发展，银行信息系统越来越多，各种数据库越来越多，访问某 个数据库的应用(程序)也越来越多。如国内一些大型银行，信息系统主机侧的核 心银行应用系统，数据库表有数千个，文件近万个，程序数近万个。如此巨量的信 息系统数据资源，如何配置，如何管理，程序与数据库的关系又是如何，是一个数 据架构研究的问题。

**例子2:数据架构的修改影响应用架构、流程架构**

信息系统设计通常可分为应用架构设计、数据架构设计和流程架构设计等。其 中最关键的设计是数据架构的设计，数据架构一经定下，应用架构基本也能定下来。 对于信息系统，最不希望的是对数据架构的修改。因为数据架构的修改几乎肯定会 涉及其他两方面架构的修改，代价巨大。

现实中，通常信息系统的数据架构和流程架构都没能得到系统设计人员的足够 重视。实际上，没有一个好的数据架构，就没有一个好的应用架构。数据的松耦合 是应用松耦合的基础和前提。如国内的一些大型银行，程序与数据库的数量都是万 级的。如果程序与数据库是多对多的关系，那么其关系的数量就是上百万的。在系 统维护中，要管理和维护好这种数量级的关系，无论投入多少成本，可以说基本上 是不可能的。正因为这样，系统的修改、升级往往千虑一失。由于考虑不周(实际 上是不可能考虑周到),结果是千里之堤毁于蚁穴， 一丝疏忽，造成重大故障，给服 务带来重大损失。

**第8章■云端数据架构**

数据架构是组织重要的IT 资产。组织需要建立一套相对完整的架构管理机制， 实现数据架构标准化、规范化，得到持续更新，并推动架构标准在系统建设过程中 得到应用和遵从。当系统更新、整合、扩充时，能够准确而快速地找出系统、数据 库、平台等相关点，确定各变化的关联关系，是保证架构延续性、系统稳定性、应 对系统频繁变动带来高风险的重要手段。

**8.1.2** **内容与关联性**

云端数据架构职能域包含的过程域有：架构方法、架构标准、数据管理平台、历 史数据管理四个过程域。数据架构所包含的过程域能够帮助组织设计一个最佳的数据 层来满足当前和未来的业务目标；建立和实施经审批的可执行标准；选择平台和支持 技术来满足范围和性能需求；集成不同的信息源并有效管理历史数据和旧数据。

架构方法帮助组织开发一个经审批的方法来审视和设计一个可使用的数据和技 术架构，这个架构强调要减少重复数据并将数据分享最大化。架构标准论述了数据 表示、数据存取和数据分布标准的制订和评审。数据管理平台强调利益相关方要参 与并且治理影响平台选择和实施的决策。历史数据管理解决了版本控制、记录保存 和归档问题，确保了数据满足适用的可用性、业务需求和管理需求。

**8.2** **活动及要求**

**8.2.1** **架构方法**

**1.** **简** **介**

1 ) 目 的

设计并实施一个最佳的数据层来推动云端数据的采集、储存、控制和交付活动， 以达成业务和技术目标。

2 ) 介 绍

架构方法对于设计一个良好的目标数据架构是十分关键的，良好的架构可以满 足当前和未来的业务目标。数据模型是数据架构规划中最重要的内容，定义良好的 数据模型可以反映业务模式的本质，确保数据架构为业务需求提供全面、 一致、完 整的高质量数据，且为划分应用系统边界，明确数据引用关系，定义应用系统间的 集成接口，提供分析依据。良好的数据建模与数据标准的制订才是实现数据共享、 保证一致性、完整性与准确性的基础。有了这一基础，企业才能通过信息系统应用 逐步深入，最终实现基于数据的管理决策。

 **■云端数据治理**

为实现关键业务数据的分享，应对数据架构进行精简、减少数据冗余、通过接 口和数据服务来改进数据的集成。目标架构应支持数据质量目标，保证使用高质量 且可信任的数据来满足关键业务目标，并且需要的数据协调最少。

架构方法应考虑可扩展性，以保证组件、服务和代码库等形式的资源性能可以 伴随着数据加载得到提升。

架构方法应考虑弹性原则，以确保系统具备容错能力及从反常使用或毁坏中的 数据恢复。

架构方法应考虑安全需求，对公司数据和个人信息进行保护，同时能够保证架 构适当反映已识别的法律要求和管理要求。

目标架构要求要有一个从现状到未来式的转换计划： 一个IT 的实施计划，并定 义度量架构蓝图及组件有效性的度量项和关键绩效指标。度量项可能包括新数据组 件或服务的采用数量、采用率、设计新组件对特定业务需求的满足、在设计之前解 决数据质量问题的数量等。

架构方法必须足够灵活，以应对不断发展的新技术的组织和过程需求，且必须 阐述数据迁移的内容，确保数据治理和利益相关方采用转换计划以及构架的目标组 件，确认业务运营受到影响的程度达到最低。

在进行设计和实施之前，应当对架构方法进行评审和审批以确认它满足业务需 求。因为架构方法属于长期的指导活动，所以应当根据变化的业务需求和组织环境 对其进行维护和修正。

3)相关过程域

●参考主数据管理过程域了解业务目标与数据管理方案和架构活动保持相符 的相关信息；

●参考战略过程域了解协作的一致意见和组织对数据资产的控制；

●架构标准过程域包含了适用于架构方法的标准的相关信息；

●数据概要分析过程域包含了在任何架构中设计和合并之前应当执行的概要 分析活动的信息。

**2.GQM 表**

|  |  |  |  |
| --- | --- | --- | --- |
| 目 标 | 问 题 | 度 量 项 | 度量项说明 |
| G1:经审批的  架构方法与业务 需求、架构标准保 持一致 | Q1:业务信息需求 是否真实、完整 | M1:定义需求 | 为创建企业数据架构，企业应首先定  义其信息需求 |
| M2:架构方法 | 一个信息系统存在的价值就看它是  否能满足业务系统的各种需求，组织应 建立并遵循一个使IT架构满足业务需  求的架构方法 |

**第8章■云端数据架构**

**续表**

|  |  |  |  |
| --- | --- | --- | --- |
| 目 标 | 问 题 | 度 量 项 | 度量项说明 |
| G1:经审批的  架构方法与业务 需求、架构标准保 持一致 | Q1:业务信息需求 是否真实、完整 | M3:评估计划 | 数据架构应得到有效的执行，并定期 进行评估。应特别注意企业数据模型与 关键业务战略、流程、组织和系统保持 一致性 |
| Q2:架构标准的具 体要求有哪些 | M4:数据标准评估 | 架构蓝图的制订和组件的设计应考 虑已经存在的数据标准和数据模型 |
| G2:持续监督  从“原样”到“未 来式”的转换方案 的执行，确保项目 与长期目标保持 相符 | Q3:同步、合并或 清除复制的数据时使 用了什么合理化方法 | M5:建立并执行合规流 程 | 建立流程对各项目的数据架构进行  评估，确保概念、逻辑和物理数据模型 一致，并对组织数据架构的长期实施有 所贡献 |
| M6:进行数据质量分析 | 由于大约一半的数据质量问题为结  构性问题，全面的分析频繁表明了新的  数据存储区中具有可以修复的设计缺  陷。当一个组织着手于一项重要的重新  设计或合并时，宜进行数据质量分析 |
| M7:数据接口规格追溯 | 整合接口和公用数据服务的接口的  开发应遵循通用编程惯例，应提供指导 这些活动的组织标准，还应将接口包括 在完整的技术数据包内，并根据组织的 配置管理过程和标准进行管理 |
| M8:公用接口设计 | 定义目标并识别在未来可以进行合  并、提取、消除或替代的接口，考虑的 内容包括数据迁移、统一的导入机制、 数据表示的标准化、数据发布方法、合 并视图、候选的公用数据服务等 |
| M9;执行数据存储的合 理化过程 | 合理化过程评估确保新数据组件开  发方向的信息需求之前未被提及，合理 化过程应评估现有的数据资产、冗余级 别、现有业务需求的满足程度、所使用 的平台和技术 |
| Q4:在截至日期、 紧凑的日程以及其他 压力状况下，组织如 何保证转换到目标状 态的计划进展正常 | M10:架构层转换计划 | 转换计划应包含现有的数据存储状  态、退役以及重设计的目标日期、人员 计划、新技术的培训计划以及未来组件 实施的目标日期等 |

 **■云端数据治理**

**续表**

|  |  |  |  |
| --- | --- | --- | --- |
| 目 标 | 问 题 | 度 量 项 | 度量项说明 |
| G3:所有内部  利益相关方都参 与审批并采纳了 架构方法 | Q5:架构方法是否 得到一致的遵循，并 且项目级别的决策是 否与这方法相符 | M11:利益相关方参与到 架构决策当中 | 为了让信息系统能真正满足业务需  求，数据治理委员会、企业数据架构师  业务专家等应积极参与到架构决策中 |
| M12:业务单元、IT和  数据治理协作开发目标架 构并对其进行审批 | 组织设计数据层以解决业务需求和  阐述最佳实践是十分重要的，让所有内  部利益相关方参与是采用以及维持长 期支持的关键 |
| M13:建立并遵循治理过 程 | 合理化活动确保目标数据架构能解  决数据资产当前和未来的需求，活动内 容通常包括：多个的数据模型、冲突的 业务术语定义、值的相似性和差异、未  记录的业务逻辑、多样的ETL与重复 的数据内容映射、多个数据存储中的多 余数据、普通数据使用的例外情况、业  务对特定数据存储的依赖程度以及计 划的应用合并工作等 |
| G4:基础平台、  技术需求与架构 方法保持相符，并  经过了利益相关 方的审批 | Q6:组织是否记录 并审批技术能力和需 求以满足经营业务的 持续性 | M14:目标数据架构与数 据战略一致性 | 定义数据架构的决定是数据战略的 一部分，实施数据架构组件的决策是战 略决策，架构必须支持战略 |
| M15:架构、技术需求以 及支持性的基础设施能力 的一致性 | 架构设计应当与组织预期的目标基  础设施相符。组织需要决定如何管理数 据迁移、数据集成、数据质量和数据安 全，以达到在可扩展性、性能、可靠性 等方面的目标 |
| G5:业务和IT 的利益相关方使 用了度量架构蓝 图及其组件有效 性的度量项和关 键绩效指标 | Q7:组织是如何制 订度量项和度量指标 的 | M16:性能统计分析和数 据质量改善在架构设计过 程中被用作输入信息 | 新架构组件的性能测量指标、成本效 益分析是架构设计决策的输入信息。识 别并跟踪现有架构的限制，并将其用于 未来架构计划的输入信息。新架构组件  的测量指标也被用于评估是否达到业 务目标和数据质量目标 |
| M17:架构评估与改进计 划 | 架构转换的度量项使主要项目在里  程碑成果的汇报上能准确展示在架构 采用上的进度，并强调从新组件和已修 正组件中实现的能力改进 |
| M18:出版物 | 随着数据架构管理的进步，组织需要 与各种数据管理组织保持密切接触，以 了解、学习目前业内的最佳实践，持续 的更新数据架构活动中的最佳实践 |

**第8章■云端数据架构**

**3.** **成熟度等级介绍**

**Level 1: 初始级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 1.1 | 建立并遵循一个使IT架构满足业务需求的 架构方法 | ●用于实施的架构设计 |

【成熟度等级说明】

一个信息系统存在的价值在于它是否能满足所面对的业务系统的各种需求。为 了这个目标，首先要把业务系统的一些需要计算机处理的主要内容，在概念上从计 算机处理的角度来归纳和分析，以对其加深了解，才有可能满足其需求。对信息系 统涉及的各种主要的业务内容在概念上进行研究并准确定义，是建立数据处理系统 的数据库、数据仓库的数据主题、数据模型、数据架构的基础。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 1.2 | 识别业务和IT利益相关方，并且利益相关 方参与到架构决策当中 | ●架构审批的相关者清单  ●针对架构的业务和技术审批 |

【成熟度等级说明】

信息系统支撑下的企业业务运作状况，是通过信息系统中的数据反映出来的。 数据是信息系统管理的重要资源，数据架构是企业IT 架构的核心。为了让信息系统 能真正满足业务需求，数据治理委员会、企业数据架构师、业务专家等应积极参与 到架构决策中。企业可以通过定期举行培训课程、研讨沙龙以及聘请外部专家做相 关分享，来不断培养员工的数据分析能力。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 1.3 | 定义需求以指导实施 | ●数据架构需求文档 |

【成熟度等级说明】

为创建企业数据架构，企业应首先定义其信息需求。确定企业信息需求的一个 方法是基于组织内部和外部目标评估目前组织所需要的输入和输出。利用真实的系 统文档和报告采访参与者，形成数据材料这一材料提供重要的数据实体、数据属性 和数据计算的一览表，该材料就成为企业数据模型的基本需求。

**Level 2: 管理级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.1 | 目标数据架构与数据战略相符，并对其进 行补充 | ●定义目标数据架构 |

**■云端数据治理**

【成熟度等级说明】

定义数据架构的决定是数据战略的一部分，实施数据架构组件的决策是战略决 策。架构必须支持战略。

由于战略是自上向下制订的，因此它描述了经广泛认可的愿景、战略目标、战 术目标以及组织级数据资产的重点。目标数据构架应使用经广泛认可的愿景作为重 要的输入信息，使用数据战略中的高级别需求来构建和制订未来状态的蓝图。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.2 | 建立并遵循一个治理过程以确保目标数据 架构由业务和IT利益相关方共同进行合理化 并审批 | ●对架构设计的合理化活动及审批过程 |

【成熟度等级说明】

合理化活动确保目标数据架构能解决数据资产当前和未来的需求。逻辑层面的 考虑内容通常包括：多个的数据模型(有可能不同)、冲突的业务术语定义、值的相 似性和差异、未记录的业务逻辑、多样的ETL 与重复的数据内容映射等。物理级别 映射和评估包括：多个数据存储中的多余数据、普通数据使用的例外情况、业务对 特定数据存储的依赖程度以及计划的应用合并工作。在架构层面参与数据的合理化 工作使组织能够完整地描述现有的架构，从而支持未来状态蓝图的制订。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.3 | 根据现有数据层组件和未来状态蓝图建立 架构层的转换计划 | ●数据架构层转换计划 |

【成熟度等级说明】

转换计划包含了现有的数据存储状态、退役以及重设计的目标日期、人员计划、 新技术的培训计划以及未来组件实施的目标日期等。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.4 | 建立和遵循过程以确保共享数据的数据接 口规格得到文件化记录，并且可以对所有使 用接口的行为进行端到端追溯 | ●共享数据接口的可追溯性图 |

【成熟度等级说明】

包含整合接口和公用数据服务的实例应明确规定数据集成规则和性能要求，应 记录系统之间的接口，以支持未来的变化，并提供适合系统间集成的方法。这些接 口的开发应遵循通用编程惯例，应提供指导接口开发的组织标准，还应将接口包括 在完整的技术数据包内，并根据组织的配置管理过程和标准进行管理。

**第8章■云端数据架构**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.5 | 建立并执行合规流程确保项目参考并使用 经认可的目标架构 | ●关于实施与目标架构一致性的记录 |

【成熟度等级说明】

企业数据架构提供个别项目的企业数据需求，加速这些需求的识别和定义过程。 对项目的数据设计进行评估，可确保概念、逻辑和物理数据模型一致，并对企业数 据架构的长期实施有所贡献。

**Level 3: 定义级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.1 | 整个组织遵守为实现目标数据架构而制订 的架构方法 | ●关于数据相关的架构方法的标准文件 |

【成熟度等级说明】

组织要贯彻目标架构的一致性是有难度的，因为一些短期的、事件驱动的重点 可能会导致一些例外情况的发生。这些例外情况会造成长期的影响后果，导致在实 现目标架构上相关复杂性、成本以及返工的增加。清晰、明确的偏离例外过程有助 于合规性，因为项目团队有时会以计划或预算为理由，抵制使用架构标准。通过利 用数据治理监督，架构评审委员会应裁定偏离例外情况。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.2 | 执行数据存储的合理化过程 | ●数据资产合理化过程评估报告 |

【成熟度等级说明】

组织需要合理化地评估现有的数据资产、冗余级别、现有业务需求的满足程度、 所使用的平台和技术。合理化过程评估确保新数据组件开发方向的信息需求之前未 被组织里的其他数据存储所提及。现有的数据存储可以通过添加抽象层(例如， XML) 或提高数据质量过程来达成长期的架构目标。合理化过程的结果为架构转换 计划提供了关键的输入信息。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.3 | 业务单元、IT和数据治理协作开发目标架 构并对其进行审批 | ●数据架构开发规范 |

【成熟度等级说明】

组织设计数据层以解决业务需求和阐述最佳实践是十分重要的。组织应当避免 仅把它授权给 IT, 因为这样会导致业务的抵触。例如，没有参与其中的业务发起人 可能并没有收到精简计划以将他们的数据存储转换到目标状态的相关信息。让所有

**■云端数据治理**

内部利益相关方参与开发架构是采用以及维持长期支持的关键。

我们建议使用协作的方法来开发架构。如果能确保集体参与，架构收益将比预 期时间更长。随着转换计划的执行，有可能每个主要业务发起者所拥有的一个或多 个数据储存最后都会受到影响。随着数据存储进行重新设计、合并或替代，意见上 的一致接受能推动决策的制订并减少阻碍。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.4 | 组织创建并维护度量项以评估状态转换的 进度 | ●架构转换度量项及指标 |

【成熟度等级说明】

架构转换的度量项使主要项目在里程碑成果的汇报上能准确展示架构采用上的 进度，并强调从新组件和已修正组件中实现的能力改进。例如，关于改进的业务反 馈能够以数据信心指数的方式进行表示，这个指数与目标架构达成业务目标的情况 成正相关。制订并监督可追溯图以评估开展的实施满足业务需求的程度，并确认现 有数据源按计划进行了成功的转换。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.5 | 对内部和经选择的外部数据标准进行评  估，并将它们应用于架构蓝图的制订和组件 的设计上 | ●外部数据标准评估报告  ●外部数据标准应用建议 |

【成熟度等级说明】

尽管转换到目标数据层是一项长期的活动，但只要在一开始的时候遵守标准就 能够防止返工并避免其带来的成本消耗。在许多组织里，数据标准和数据模型(例 如，企业数据模型、共享的存储数据模型等)是已经存在的，且属于目标数据架构 开发前的符合性过程的一部分。如果这些构件不存在、不完整或没有被遵循，那么 目标数据层的构建工作刚好提供一个机会去创建这些构件并制订符合性过程。不断 发展的技术能力以及行业内的交流创造了对新标准的需求。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.6 | 架构、技术需求与支持性的基础设施能力 是一致的 | ●云端数据基础设施需求列表 |

【成熟度等级说明】

架构设计应当与组织预期的目标基础设施相符。组织可能已有一个合适的企业 架构、关于平台和技术的决策以及对相应的产品进行审批。如果缺乏这些，恰好为 开发数据层架构方法的工作提供了一个机会，以进一步推动组织层面在这方面的进 度。组织需要决定如何管理数据迁移、数据集成、数据质量和数据安全，以达到在

**第8章■云端数据架构**

可扩展性、性能、可靠性等方面的目标。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.7 | 架构包含公用接口设计 | ●公用接口设计规范 |

【成熟度等级说明】

一些组织体验过在数十年里因接口增加而带来的高复杂性和成本负担。为现有 接口的总体维护费用统计出一个准确的数字是十分困难的，但对一些典型的情况进 行总结反思也是次优的方法，例如每次主要的应用发布中所执行的集成测试的次数 以及每年所创建的点到点接口的数量。许多组织相对比较少关注这个长期的负担， 因为这方面的成本通常都被分配在了项目的进度和预算之中。

在谈及架构方法的情况下，很有利的做法是定义目标并识别在未来可以进行合 并、提取、消除或替代的接口。考虑的内容包括数据迁移、统一的导入机制、数据 表示的标准化、数据发布方法、合并视图、候选的公用数据服务等。

许多组织对他们所选择的数据，包括其来源和质量都缺乏全面的认识。收集并 分析接口和集成性能测量指标，以确定不符合标准和指标的情况。例如，执行缺陷 跟踪，以监测缺陷严重程度、原因、统计发生和计划决议。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.8 | 在进行数据存储区组件设计之前，进行数 据质量分析 | ●应用于设计的数据质量分析报告 |

【成熟度等级说明】

当一个组织着手于一项重要的重新设计或合并时，宜进行数据质量分析，以作 为初步措施。由于大约一半的数据质量问题为结构性问题，全面的分析有助于发现 新的数据存储区中具有可以修复的设计缺陷。该组织应对标准进行定义，以便在架 构更改之前，为何时需要进行分析提供决策指导。

**Level** **4:** **量化级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 4.1 | 性能的统计分析和数据质量改善在架构设 计过程中被用作输入信息 | ●设计的组件的定量性能标准  ●可量化的架构评估指标 |

【成熟度等级说明】

新架构组件的性能测量指标(针对定义的业务目标)、成本效益分析是架构设计 决策的输入信息。识别并跟踪现有架构的限制，并将其用于未来架构计划的输入信 息。同样，随着时间的推移，架构方法的优化包括前瞻性规划，以确保将来可有效 实现下一代的能力。新架构组件的测量指标也被用于评估是否达到业务目标和数据 质量目标。这可能包括影响业务的KPI 指标，通常包括以下指标：

 **■云端数据治理**

● 时间

●由于较少的数据修复所带来的金钱节省

●增加的收入

●改善的客户满意度

●节省的时间

●改进的数据

●增强的分析

**Level** **5:** **优化级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 5.1 | 定期对目标模型进行评估，按需进行调整 | ●架构方法修改提议及反馈  ●说明了实际性能与预测性能之间的相关性数据集 的分析 |

【成熟度等级说明】

数据架构必须得到有效的执行，并定期进行评估。应特别注意企业数据模型与 关键业务战略、流程、组织和系统保持一致性。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 5.2 | 该组织通过出版物和会议分享、学习架构 方面的经验教训 | ●关于该组织的架构方法的演示材料或出版物 |

【成熟度等级说明】

随着数据架构管理的进步，组织需要与各种数据管理组织保持密切接触，以了 解、学习目前业内的最佳实践，持续地更新数据架构活动中的最佳实践。

**4.** **实施案例**

下表为某集团关于关系数据模型建模方法的说明示例。

|  |
| --- |
| 集团关系数据模型建模方法说明  1 目标  形成一套统一的建模方法论，指导各子公司在此基础上完善和扩展逻辑数据模型，为公司信息系统建设提供进 一步有力保障；通过统一建模及模型的管理，规范和增强公司信息系统的数据管理。  2 原 则  建模过程中以业务运营支撑系统为主要参照数据源，遵照第三范式分主题进行实体关系模型的设计，客观、抽 象、清晰、较全面地反映各子公司的业务及其系统的实际，并在此基础上逐步加强系统的数据管理。  3适用范围  本规范适用于各子公司关系型数据库。 |

**第8章■云端数据架构**

**续表**

|  |
| --- |
| **4** **建模概述**  数据模型是对客观事物及其联系的数据描述，包括层次数据模型、网络数据模型、关系数据模型、基于逻辑的 数据模型、多维数据模型等。公司信息系统底层细节数据的存储采用关系数据模型，即关系数据库。除非特别说 明，本文所指的数据模型将特指关系数据模型，数据仓库将特指系统的底层关系数据库。  4.1 第三范式  范式理论是关系数据模型设计的基础，关系数据模型可以从第一范式到第五范式进行无损分解，该过程也称为 规范化。规范化的基本思想是逐步消除数据依赖中不合适的部分，使各关系模式达到某种程度的分离，即“一事 一地”的设计原则，它有非常严格的数学定义。  规范化为数据库的设计和实施、应用程序的开发提供了诸多好处，其主要优点体现在以下4方面。  (1)可以得到更全面的数据组织；  (2) 减少不必要的数据冗余；  (3)易于维护数据的完整性：  (4)模型设计更灵活，便于系统今后的扩展数据库.  数据库设计一般采用第三范式，从所表达的含义看， 一个符合第三范式的关系必须满足以下三个条件。  (1)每个属性的值唯一，不具有多义性；  (2)每个非主属性必须完全依赖于整个主键，而非主键的一部分；  (3)关系模式中不存在传递依赖。  考虑到底层关系数据库的可扩展能力、投资成本和易于管理等多种因素，逻辑数据模型应遵照第三范式进行设 计。  4.2 实体—关系模型  实体—关系方法是一种具有代表性的语义数据建模方法，其图形化表示称为实体一关系图(简称 ER图)。按照 实体一关系方法建立的系统数据模型称为实体一关系模型 (Entity Relationship Model,简称 ER模型)。ER 模型经 过多年的发展完善，逐渐形成了完整、统一的建模标准，同时许多工具软件支持用ER 模型完成数据库的概念、逻 辑和物理建模过程。IE 和IDEF1X是两种常见的基于实体一关系模型的图形化建模方法及符号体系，它们所表达 的概念都基本相同。底层关系数据库逻辑数据模型应是一个采用IE 符号体系的实体—关系模型。  4.3 建模的三个阶段  建模时通常采用“自顶向下、逐步求精”的方法，因此建模从设计过程上可以分为以下三个阶段。  (1)概念模型设计 (Concept Data Modeling):这一阶段之前的首要工作是通过需求分析，明确需求所涵盖的业 务范围。然后再对需求范围内的业务及其间关系进行高度概括性的描述，把密切相关业务对象进行归类，即划分 主题域。概念模型的设计是为逻辑模型的设计做准备的，它没有统一的标准，主要根据设计者的经验完成。  (2)逻辑模型设计 (Logical Data Modeling):分别对概念模型的各个主题进行细化，根据业务定义、分类和规 则，定义其中的实体并描述实体之间的关系，并产生实体关系图 (ERD), 然后遵照规范化思想在实体关系的基础 上明确各个实体的属性。实体产生于公司开展的业务、服务及其涉及的对象(如客户、帐户、员工、机构、资源), 实体间的对应、约束关系则来自于各业务过程中的规则。可以说，这一阶段面对的是业务。  (3)物理模型设计 (Physical Data Modeling): 主要依据逻辑模型针对具体的分析需求和物理平台采取相应 的优化策略。此时会在一定程度上增加数据冗余或者隐藏实体之间的关系，是一种反规范化的处理，目的是 提高数据分析的速度，适应具体数据库的容量、性能等限制。可以说，这一阶段面对的是具体软硬件平台和 性能要求。 |

 **■云端数据治理**

**续表**

|  |
| --- |
| 一旦逻辑模型到位，物理模型就有了可参照的依据，开发工作内容也同时得到明确。鉴于各子公司信息系统可 能采用不同的数据库平台、不同的配置方案并具有不同的分析需求侧重点，各子公司可进行本地化补充并经集团  公司审批之后自行进行物理模型的设计。  5 建模中所考虑的关键问题  5.1 逻辑模型和业务的关系  业务是指公司开展或参与的与市场、客户、合作伙伴等相关的各种服务活动，它是逻辑模型设计的出发点和依 据。按照实体—关系方法设计的逻辑数据模型就是业务的数据逻辑视图，图形化的结果就是实体关系图。业务过 程中涉及的对象表现为实体或属性，业务过程中的规则表现为实体间的关系，而后者通过实体的主外键引用体现 出来。  建立逻辑数据模型的第一步就是要进行业务分析，这一过程参与的人员主要是业务人员(或者是熟悉业务的技  术人员)。在划定需求要求的业务范围之后，业务分析首先要弄清楚企业主要开展或参与哪些活动，每项活动涉及 哪些对象和资源，为了开展这些活动需要哪些相关活动等。通过分析这些情况，识别并列出业务活动中涉及的每  一个实体及其主要属性，并逐步完善其属性。  设计出逻辑数据模型的大致框架后，可以在概念模型主题域划分的基础上对实体再次进行归类，以便将来理解  和改善逻辑数据模型。  5.2逻辑模型和需求的关系  首先，确定需求的范围可以从“水平”上划定逻辑模型的设计范围，决定了建模时要考虑已有的哪些业务、不 需考虑哪些业务。这样就可以不漏掉数据建模时应该分析的业务，同时可以避免在非关键业务上浪费精力。  其次，需求分析可以从“深度”上明确数据逻辑模型设计时要把握的数据最细粒度是多少。由于底层细节数据 决定了将来经营分析系统可以分析到的最细程度，因此其逻辑模型设计阶段要根据当前及近期可能的最细分析程 度决定其实体或属性的最细粒度。  5.3 逻辑模型设计的难点  为避免在业务定义上的歧义，应建立公认的标准定义。同时，逻辑数据模型必须具有扩展性，能够灵活适应未 来业务模式的变化。因此，逻辑模型不应该完全受原有业务系统对业务理解和实现的限制。  6建模工具软件说明  建模工作采用冠群公司的Erwin Data Modeler。该工具是目前较流行的数据模型建模工具，支持第三范式和星型 模型，支持逻辑数据模型和物理模型的设计，支持不同目标数据库之间物理模型的转换。在逻辑数据模型设计中，  ERWin支持两种IDEFIX和IE两种逻辑建模符号体系，支持IDEF1X、IE和DM(Dimensional Modeling,多维建 模)三种物理建模符号体系，其中DM符号体系用于星型模型的设计。建议统一采用IE建模符号体系。 |

**8.2.2** **架构标准**

**1.** **简介**

1)目的

提供一套经批准的数据表示、数据访问和数据分发标准，以利于实现云端数据 共享、保证一致性、完整性与准确性。

**第8章■云端数据架构**

2)介绍

架构标准定义了制订、批准以及遵从数据表示、数据访问和数据分发标准，其 可能包括以下方面的标准：

(1)数据表示，包括业务词汇、逻辑、物理、XML 、建模标准、模型管理等。

(2)数据访问，包括普通数据业务、适用的信息交换标准、点到点数据中转、 批量数据移动和数据集成的标准

(3)数据分发，包括内部和外部数据供应：分发控制和管理、请求及批准访 问、访问限制、分发模型(例如推/拉、发布和订阅)、所有权和权限、监管部门 和审核等。

标准的制订是必要的。业务部门在评估和审批标准以及区分目标优先次序方面 应发挥积极的作用，最好的方式是在业务部门和IT 之间实施一种协作方法，以建立 一个系统的、受管理的方法来制订、发布并实施可同时满足业务需求和该组织的架 构方法的标准。 一旦在整个组织范围内制订标准，则标准应支持整个组织向目标数 据架构的演变，以实现数据资产的技术堆栈的选择和部署的改进、开发周期的简化。 通过尽可能减少项目人员的不一致工作，并通过组件的重新利用等方法来减少成本。 为了实现这些目标，有效和积极的架构管理是必需的。

把数据架构管理融入一个组织要遵循以下指导原则：

(1)数据架构是一系列规范架构的整合，用于定义数据需求、制订数据整合、 控制数据资产，是数据投资与业务战略相一致。

(2)企业数据架构与流程架构、业务架构、系统架构和技术架构一起，是整个 企业架构的一部分。

(3)企业数据架构不仅仅涉及数据，还采用通用的业务术语来建立企业内的语义。

(4)企业数据架构包括3个主要的规范：企业数据模型、信息价值链分析和数 据交付架构。

(5)企业数据模型是整合的面向主题的数据模型，定义了跨越整个组织的关键 数据。

(6)信息价值链分析定义数据、流程、角色、机构以及其他企业元素之间的关 键关系。

(7)数据交付架构定义数据如何在数据库和应用之间流转的主蓝图。它保障数 据质量和完善性，以支持事务的业务处理和商务智能报告分析。

(8)业务部门在评估审批标准以及对目标优先次序制订方面应发挥积极的作用， 在业务部门和IT 之间实施一种协作方法，以建立一个系统的过程来制订、发布并制 度化实施可同时满足业务需求和该组织架构方法的标准。

3)相关过程域

●战略过程域和数据质量战略过程域是标准定义过程的重要考虑因素，以确保 实现的架构将支持业务目标；

 **■云端数据治理**

●数据提供者管理过程域提供了关于要求的信息，这将会影响组织的标准，尤 其是诸如安全要求等事宜，在各种平台之间必须持续满足这些事宜；

●业务词汇表过程域包含有关业务术语标准的信息；

●元数据管理过程域包含有关利用元数据标准的信息。

**2.GOM 表**

|  |  |  |  |
| --- | --- | --- | --- |
| 目 标 | 问 题 | 度 量 项 | 度量项说明 |
| G1:制订一整 套与架构方法和 数据管理策略相 一致的数据标准 | Q1:该组织的目标 数据架构需要哪些标 准类别，以及如何确 定其范围并定义 | M1:架构标准 | 架构标准包括：数据表示、安全、数 据访问和数据供应 |
| M2:架构方法、蓝图和 组件设计与所选的标准 一致性 | 在标准被认可和通过之后，应在整 个组织内进行宣传和发布，以确保在 设计、工具选择等方面使用新的或改  进的标准 |
| M3:数据架构标准对外 部要求的适应性 | 外部要求用于监管报告、信息交换或 合规性目的。组织应根据业务需求采纳 或调整由外部要求(例如监管机构、行 业交换标准)推动的标准 |
| M4:架构标准评审与审 批 | 数据治理委员会、企业数据架构师、 业务专家等应积极参与评审和批准标准 |
| G2:实施涉及 到业务和IT利益 相关方的可持续 的标准制订和维 护过程 | Q2:如何对数据模 型进行批准、维护和 治理 | M5:架构标准实施效果 | 在整个组织内遵循架构标准，包括新 的开发、重新设计、整合以及现有数据 存储区的改进等活动 |
| M6:角色和职责 | 明确架构标准制订过程的角色和职  责，包括根据需要制订、维护、测量、 审核并修订标准等过程 |
| M7:架构标准与业务需 求一致性 | 应由数据治理部门评审和批准架构标 准的制订及修订，以确保架构标准与业 务需求保持一致 |
| M8:架构标准的采用和 合规性的测量指标 | 测量指标可以协助各个项目团队评估 其对数据资产的质量和结构的贡献；同 时，合规性指标可以向组织表明标准的 覆盖面和有效性 |
| M9:架构标准的合规性 评估 | 用于支持标准制订和管理的“使能技 术”的清单包括数据建模工具、用于评 估合规性或偏离的数据发现工具，以及 必要的外部转换脚本等 |

**第8章■云端数据架构**

**续表**

|  |  |  |  |
| --- | --- | --- | --- |
| 目 标 | 问 题 | 度 量 项 | 度量项说明 |
| G3:建立标准 遵守和例外情况 的有效的治理和 审核过程 | Q3:为了确定哪些 方法适用，该组织是 否定义了与架构一致 的标准数据访问方法 和标准 | M10:架构标准偏离请 求、批准和管理的过程 | 应对由事件驱动的新标准、现有标准 的增加和更改的请求，进行协调、同意 并批准，以确保朝向长期架构目标的进 展不受影响 |
| M11:架构标准改进 | 组织应根据合规性评估、数据偏离发 现工具及各数据资产测量指标结果，决 定是否需要变更数据架构标准，或增加 对数据管理专员、软件开发工程师、项 目经理等利益相关方的有关数据架构的 使用培训 |
| M12:架构标准变更 | 组织应基于测量指标和风险因素，形 成对提议的标准修改的影响进行分析评 估 |

**3.** **成熟度等级介绍**

**Level 1: 初始级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 1.1 | 定义了数据架构部分标准，但并未全部或 全程得到遵循 | ●数据架构标准 |

【成熟度等级说明】

通常情况下，组织会制订架构标准，但其可能并不会始终如一地实施。例如， 在数据仓库环境中可能会严格遵循标准，但对于操作数据存储区，标准未得到严格 遵循，或者一些业务部门并未采用和实施。

**Level 2: 管理级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.1 | 定义并遵循架构标准，包括数据表示、安 全、数据访问和数据供应 | ●包括数据表示、安全、数据访问及数据供应的数 据架构标准 |

【成熟度等级说明】

数据架构标准的主体通常会基于更广泛的架构活动进行演变，例如整个组织范 围内的应用程序或技术架构的转换、重要的应用集成等。制订一套有效的标准需要 时间和努力，根据资源和预算因素，建议扩展标准基础，以覆盖影响组织的数据层 的各个方面。

**■云端数据治理**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.2 | 与业务利益相关方一起评审架构标准，并 通过数据治理进行批准 | ●数据架构标准审批记录 |

【成熟度等级说明】

数据架构标准的制订通常由IT 发起，并使用数据需求作为输入信息。业务相关 人应通过数据治理来参与评审和批准标准。经批准的标准通常由数据管理职能进行 维护，并使用业务战略和数据战略作为输入信息。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.3 | 定义并遵循适用于架构标准偏离请求、批 准和管理的过程 | 架构变更审批记录 |

【成熟度等级说明】

由于标准的主体不断演变，最终会影响整个数据层和所采用的数据访问机制， 对于由事件驱动的新标准、现有标准的增加和更改的请求，必须认真协调、同意并 批准，以确保朝向长期架构目标的进展不受影响。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.4 | 架构方法、蓝图和组件设计与所选的标准 保持一致 | ●新项目引入标准架构 |

【成熟度等级说明】

在标准被批准之后，在整个组织内进行宣传和发布是一项挑战。如果目标数据 架构正处于开发之中，这将会需要关注和努力，以确保在设计、工具选择等方面使 用新的或改进的标准。

**Level 3: 定义级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.1 | 在整个组织内遵循架构标准 | ●架构标准合规性或监管报告  ●架构标准例外规程 |

【成熟度等级说明】

将标准应用于新的开发、重新设计、整合以及现有数据存储区的改进，这通过 整个生命周期(从建模到任何正常的生产变更或发布)表现出来。承诺创建和实施 标准是一项原则，会随着时间的推移带来回报；遵守标准是必不可少的。清晰、明 确的例外过程有助于合规性，因为项目团队有时会以计划或预算为理由，抵制使用 架构标准。通过利用数据治理监督，架构评审委员会应裁定例外情况。

**第8章■云端数据架构**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.2 | 数据架构标准的制订应包括适用于组织的 外部要求 | ●架构标准使用法律法规列表 |

【成熟度等级说明】

外部要求可用于监管报告、信息交换或合规性目的。根据组织的业务需求采纳 或调整由外部要求(例如监管机构、行业交换标准)推动的标准，标准采用和集成 的方法可以利用以下任何一项：

●当不矛盾时，纳入到内部制订的标准；

●创建面向外部的转变，以满足由评审委员会和相关治理部门所定义并批准的 要求。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.3 | 明确利益相关方在架构标准方面的角色和 职责，包括合规性责任、所有权和培训等 | ●标准制订规程包括制订、维护、测量、审核、修 订标准等内容 |

【成熟度等级说明】

完整的标准制订过程包括根据需要制订、维护、测量、审核并修订标准的活动 步骤、里程碑、明确的角色和职责以及治理结构。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.4 | 数据治理确保架构标准与业务需求保持  一致，并与该组织的高级架构治理部门保持 一致 | ●架构评审委员会会议记录 |

【成熟度等级说明】

提出并制订标准的过程通常由数据管理职能或 IT 领导，例如 IT 或业务架构通 常制订数据访问标准，数据管理职能经常领导制订关于业务数据以及将业务考虑因 素纳入到IT 环境中的标准，例如数据建模标准。 一旦制订，应由数据治理部门评审 和批准其补充和修订内容。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.5 | 定义并实现用于监督和控制架构标准的采 用和合规性的测量指标 | ●架构标准测量指标  ●架构标准开发和改进过程记录 |

【成熟度等级说明】

标准的一致性和对标准的正确使用可以为设计实践，以及应对复杂性日益增加 的主动防御系统提供大力支持，因此有效的合规性方案是必不可少的。测量指标可 以协助项目团队评估标准对数据资产质量的提高和结构的优化所做的贡献；同时，

 **■云端数据治理**

合规性指标可以向组织表明标准的覆盖面和有效性。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.6 | 制订、记录并实施审核程序，以评估对于 架构标准的合规性 | ●架构标准审核程序  ●架构标准审核记录 |

【成熟度等级说明】

用于支持标准制订和管理的“使能技术”的清单可以包括数据建模工具、用于 评估合规性或偏离(作为审核活动的一部分)的数据发现工具，以及必要的外部转 换脚本。

在系统开发生命周期内，数据模型、数据访问方法和数据分发设计的审核应基 于统一适用的标准。

**Level** **4:** **量化级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 4.1 | 审核结果、测量指标和内部偏离模式表明 了何时需要变更数据结构标准以及增强标准 应用指导 | ●标准评审记录  ●架构标准符合性测量指标 |

【成熟度等级说明】

组织应根据合规性评估、数据偏离发现工具及各数据资产测量指标结果，决定 是否需要变更数据架构标准，或增加对数据管理专员、软件开发工程师、项目经理 等利益相关方的有关数据架构的使用培训。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 4.2 | 组织对数据架构标准提议进行基于风险的 影响分析 | ●关于提议的标准修改的影响分析 |

【成熟度等级说明】

该组织中负责领导标准制订的小组应基于测量指标和风险因素，形成对标准的 修改提议，并对这一修改的影响进行分析评估。

**Level** **5:** **优化级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 5.1 | 对于新的或提议的数据标准更改，向外部 利益相关方提供反馈 | ●外部标准机构参与的证据 |

【成熟度等级说明】

例如，监管机构可能会提供用于评审的新要求。该组织将评估这些要求，并提 供可行性和责任信息输入，提出修改建议。

**第8章■云端数据架构**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 5.2 | 该组织在其行业内推进数据架构标准举措 | ●可能被采用的未来技术的提议标准 |

【成熟度等级说明】

组织加入国际性或行业性的数据管理协会，积极参加各种数据管理研讨会，以 演讲人身份介绍数据架构标准相关主题。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 5.3 | 该组织对潜在采用的创新的数据技术和方 法进行研究，并为所采用的这些技术和方法 制订适当的新标准 | ●新兴技术的研究 |

【成熟度等级说明】

组织数据管理专员需要在业务工作中保持业务知识更新，加入数据管理组 织，接受持续的教育，积极学习、研究新的数据管理技术，为组织采用新技术做 好准备。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 5.4 | 该组织通过出版物、会议和白皮书分享所 获得的最佳标准实践和经验教训 | ●与数据标准相关的演示材料和其他作品 |

【成熟度等级说明】

组织数据管理专业人士应订阅专业杂志并通过网上各种论坛、专业讨论群分享 数据架构管理相关主题。

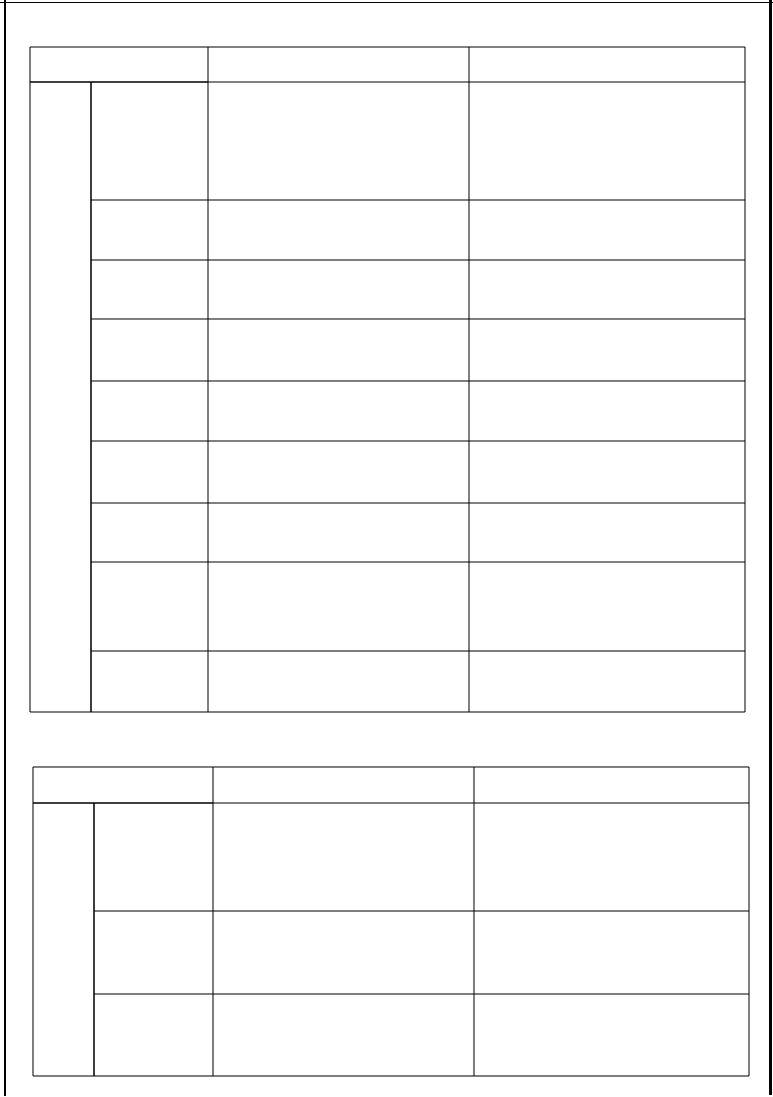
**4.** **实施案例**

某运营商数据分类规范示例。

|  |
| --- |
| 某运营商数据分类规范  企业数据分类是企业数据标准的一个组成部分，即全部企业数据执行同一个数据分类标准。企业数据分类需要  满足各种数据需求对数据组织的要求，并独立于具体的数据模型和数据分布。企业数据分类还要有利于数据的维 护和扩充。  公司企业数据根据以下指导原则进行分类。  ●国际电信管理论坛(TMF)的SID分类框架  ●与IT目标架构相互配合  ●公司实际情况，具体的数据分类可灵活处理  公司数据分为客户、产品、服务、资源、市场营销、合作伙伴和企业管理等七类数据，具体分类如下： |

**■云端数据治理**

续表

1客户主题域

详 细 注 释

数 据 分 类

所覆盖的主要实体/数据

客户是指已经购买或使用公司及

业务合作伙伴所提供的产品的个人

基本客户信息、集团客户资料、个人

客户

或组织。客户还包括使用公司产品的

客户资料、客户关系、客户文档资料

潜在个人或组织用户

客户与公司签订的对某些(种)产

客户订单

订单及其处理

品的购买意向

客户由于所购买或使用的产品出

客户问题单及其处理

客户问题

客户优惠

现问题，而向公司反映所产生的信息

在产品资费的基础上，对客户所进

客户/账户级费率

行的优惠

通过各种渠道，与客户进行联系而

客户

客户交互事件

客户交互

客户评估与

统计

产生的所有信息

对客户相关信息收集、处理、分析、

黑名单、客户分级、客户离网预测、购

买倾向、客户价值、客户生命周期统计

客户服务水平协议/客户层面服务的

加工得到的结果

客户与公司达成的关于购买和使

客户服务

用产品的权利和义务

使用记录

欠费数据、催缴处理、账务争议、调

对客户异常账务场景处理及其结

客户异常账务

处理

账处理、错单、无主账单、无局向账单、

果

欺诈报告

基本账户信息、账户关系、账单及其

客户因购买或使用公司及合作伙伴提

供的产品而产生或关联的财务信息

客户账务

处理、详单、收费处理、发票、押金

2 产品主题域

数 据 分 类

详 细 注 释

所覆盖的主要实体/数据

业务实例就是客户购买或使用产

业务实例

品后，电信运营商提供相应的业务资 源支撑，业务开通后形成的实例。客

业务实例

户可以通过该实例获得相应的功能

业务是指公司利用自身或者第三

方资源，为客户提供具有市场价值的

产品

业务基本信息、业务成本、业务关系、 业务特征

业务/产品战略、业务/产品组合管理、 业务/产品生命周期管理、业务/产品目 录

业务

基本元素

它是指向每个细分市场提供产品/

产品战略与 计划

业务的计划，开发和部署产品/业务

的计划，以及产品/业务淘汰的计划

**第8章■云端数据架构**

续表

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 续表   |  |  |  |  | | --- | --- | --- | --- | | 数 据 分 类 | | 详 细 注 释 | 所覆盖的主要实体/数据 | | 产品 | 产品 | 产品是公司针对不同的营销渠道、 客户细分、地域细分和销售目标，对 业务、资费计划进行必要的组合、包 装后的产物 | 产品基本信息、产品特征、产品关系、 资费计划、产品优惠 | | 产品评估和 统计 | 对业务/产品相关信息收集、处理、 分析、加工得到的结果 | 业务/产品的用户统计、使用量统计、 成本收益分析、产品域汇总数据 |   3 合作伙伴主题域   |  |  |  |  | | --- | --- | --- | --- | | 数 据 分 类 | | 详 细 注 释 | 所覆盖的主要实体/数据 | | 合作  伙伴 | 合作伙伴 | 合作伙伴是指和公司存在伙伴关 系的实体 | 基本合作伙伴信息、合作伙伴关系、 合作伙伴文档资料、合作伙伴识别信息 | | 合作伙伴战略 和计划 | 同合作伙伴商业联系中的战略和 规划 | 合作伙伴战略、合作伙伴计划、合作 伙伴生命周期管理 | | 合作伙伴交互 | 公司通过各种渠道与合作伙伴进 行联系而产生的所有信息 | 合作伙伴联系事件 | | 合作伙伴业务 | 表示公司可以向合作伙伴定购的 业务 | 基本业务信息、业务关系、业务目录、 业务特征、资费计划 | | 合作伙伴订单 | 公司与合作伙伴签订的购买产品 的意向 | 订单及其处理、采购招标、交易合同 | | 合作伙伴服务 | 公司与合作伙伴达成的关于购买 和使用产品的权利和义务 | 合作伙伴服务水平协议 | | 合作伙伴评估 与统计 | 是对合作伙伴相关信息的收集、处 理、分析和加工的结果 | 黑名单，合作伙伴信用度、分级、价 值，合作伙伴绩效评估，绩效警报 | | 合作伙伴问题 | 该域处理和合作伙伴相关的问题 (双向) | 合作伙伴对公司的故障通知和投诉、 公司向合作伙伴的故障通知和投诉 | | 合作伙伴账务 | 公司和合作伙伴之间产生或关联 的财务信息 | 基本账户信息、账户关系、账单及其 处理、账务查询、收费付费处理、保证 金、押金、佣金 | | 合作伙伴异常 账务处理 | 是指对和合作伙伴相关账务中出 现的异常及进行的处理 | 欠费数据、催缴处理、账务争议、调 账处理、欺诈报告 | |

 **■云端数据治理**

续表

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 4服务主题域   |  |  |  |  | | --- | --- | --- | --- | | 数 据 分 类 | | 详 细 注 释 | 所覆盖的主要实体/数据 | | 服务 | 服务 | 服务是对利用资源向业务提供支 撑的某一实体。这些实体是服务的实 例化结果 | 服务 | | 服务规格 | 服务规格是一组服务的不变特性  和行为。 一个服务可实例化为多个服 务，每个服务使用服务中定义的不变 特性和行为 | 服务规格、服务规格特征 | | 服务配置 | 服务/服务规格、资源规格/资源及 参数之间的依赖关系。它同时包括配 置变更的信息 | 服务工单/服务规格配置/资源规格配 置/服务配置/资源配置 | | 服务评估与 统计 | 服务评估是指收集、处理、分析和 加工各种服务绩效统计数据和运营  特征的结果 | 服务使用量的统计、服务的质量统 计、服务域汇总数据 | | 服务使用 | 服务使用的记录及其性能信息。 | 通信服务使用记录、服务性能数据、 预处理规则 | | 服务战略和 计划 | 它定义了优化现有服务、推出新服 务，淘汰旧服务的战略和计划 | 服务战略、服务计划、服务生命周期 管理、服务目录 | | 服务故障 | 服务的警告、中断和服务缺陷数据 | 服务告警、服务故障/故障单及其处 理 | | 服务测试 | 服务测试用来在服务创建过程中， 故障诊断或故障修复后检测资源的 运行特性 | 服务测试信息 |   5 资源主题域   |  |  |  |  | | --- | --- | --- | --- | | 数 据 分 类 | | 详 细 注 释 | 所覆盖的主要实体/数据 | | 资源 | 资源 | 为业务和服务提供支撑的、具有价 值的、有形或无形的物件 | 纯逻辑资源，号码、逻辑连接等；纯 物理资源，局站/机房、配线设备、线  缆、局外设施等 | | 资源规格 | 资源规格定义了一组资源的不变 特性和行为 | 资源规格、资源规格关系 | | 资源拓扑 | 资源拓扑定义了物理、逻辑和网络 拓扑信息 | 以上资源的拓扑信息 | | 资源评估及 统计 | 收集、处理、分析和加工资源相关 数据得到的结果 | 资源的使用量统计、资源的性能统 计、资源域汇总数据 | |

**第8章■云端数据架构** 

续表

续表

|  |  |  |  |
| --- | --- | --- | --- |
| 数 据 分 类 | | 详 细 注 释 | 所覆盖的主要实体/数据 |
| 资源 | 资源战略和 计划 | 资源战略和计划负责规划初始和 适应未来增长的网络和资源元素，协 调逻辑和物理资源的增长以及管理  逻辑和物理资源的生命周期 | 资源的战略、资源计划、资源生命周 期管理 |
| 资源使用 | 资源使用记录及其性能信息 | 资源使用状态、资源性能数据 |
| 资源故障 | 资源的告警、中断及资源缺陷数据 | 资源故障/资源故障单及其处理 |
| 资源测试 | 资源测试用来在安装过程中，故障 诊断或故障修复后检测资源的运行  特性 | 资源测试信息 |

6 市场营销主题域

|  |  |  |  |
| --- | --- | --- | --- |
| 数 据 分 类 | | 详 细 注 释 | 所覆盖的主要实体/数据 |
| 市场  营销 | 市场战略和 计划 | 是指为了一定的市场目标而制订 的一系列市场战役 | 市场战略、市场计划、品牌 |
| 市场细分 | 市场细分是指满足某些要求或具  备某些特征的市场子集 | 市场细分、客户分群相关信息、维系  挽留计划 |
| 市场活动 | 是指为了完成市场目标而进行的 一组有明确市场针对性的市场活动  集合 | 市场活动，包括广告、调查、促销、 测试 |
| 竞争者 | 竞争者指向市场提供替代性产品  的实体。 | 市场竞争信息、竞争者信息 |
| 销售机会 | 销售机会是对客户购买意向的评  估，该域具备跟踪整个生命周期销售 机会的能力 | 销售机会信息及其联系方式，以及进 展监控信息 |
| 营销评估与  统计 | 对营销相关信息收集、处理、分析、  加工得到的结果。 | 市场活动/销售渠道的预测、绩效评 估、统计分析、市场营销域汇总数据 |
| 销售渠道 | 销售渠道是客户购买公司电信业  务的场所，包括自有渠道和第三方渠  道 | 销售渠道整体视图、渠道管理规则、 销售活动管理、渠道绩效、外部渠道结  算等 |

**■云端数据治理■**

续表

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 7 企业管理主题域   |  |  |  |  | | --- | --- | --- | --- | | 数 据 分 类 | | 详 细 注 释 | 所覆盖的主要实体/数据 | | 企业  管理 | 人力资源 | 人力资源管理所涉及的各方面数 据。 | 组织结构、岗位信息、员工信息、员 工管理、绩效、薪酬福利、培训、成本 核算、人力资源汇总数据 | | 工程项目 | 从规划、立项、设计、采购、施工 到竣工验收的整个项目生命周期的 相关信息。 | 立项管理、工作分解、进度计划、项 目文档、项目预算、项目资源、成本核 算、工程项目汇总数据 | | 采购和库存 | 库存管理相关的信息。 | 采购需求、采购计划、采购合同、入 库信息、出库信息、库位、库存状态、 调拨、库存盘点、采购和库存汇总数据 | | 决策支持信息 | 一系列系统和能力再次组织，使得 高级管理层能够在合适的时间、合适 的地点接触和分析关键的公司信息。 | 营销决策支持信息、运维决策支持信 息、计划决策支持信息 | | 财务和会计 | 记录、核算、反映和分析资金在公 司运行活动中的变动过程及其结果  的相应信息。 | 总账、应收账、应付账、现金、资产、 利润、成本、预算、财务报表、财会汇 总数据 | | 企业协同 | 企业内部各个部门的协作办公及 知识管理所涉及的信息。 | 文档流转信息、个人事务管理、工作 计划及分配、发布信息、通信服务信息、 日常事务管理信息、知识、知识生命周 期管理信息 | | 通用接入服务 | 通用接入服务所涉及的数据 | 系统用户鉴权数据、系统用户个性化 数据 | |

**8.2.3** **数据管理平台**

**1.** **简介**

1)目的

实施和管理一个有效的云端数据管理平台，以满足业务需求。

2)介绍

数据管理平台作为权威的数据源，通常包括用于数据分发、转换和集成到应用 程序的资产。尽管通常存在多个物理平台，但组织应将数据管理平台视为整体技术 生态系统，并在共同授权下管理(人员、小组、标准、SLA、政策)。

一个组织应拥有一套应用程序和技术实现跨业务部门的数据管理，并实现数据

**第8章■云端数据架构**

管理目标。数据管理平台会显著影响数据管理性能。因此，数据管理平台选择过程 应与数据管理和数据质量策略一致，以确保所需要的特点和能力将得到满足。

为了支持“自建或购买”的决策，应足够详细地对现有系统和能力进行差距分 析、考虑扩展现有平台的使用。此外，为遵循组织标准和蓝图，组织应计划在较长 时间内分阶段进行实施目标平台。

建立良好的数据架构就是为了能实现应用的松耦合。数据库分类分层是数据架 构最重要的规划，也是与应用架构关系最密切的系统规划。数据库分层分类的原则： 该数据库应该由哪一层的哪一个程序的集合访问，就把该数据库配置在相应程序集 合中，而且只能由封装自己的相应程序访问，其他程序不能跨界访问。在数据封装 的前提下，通过适当的数据冗余来保证数据流转与共享的效率。数据冗余要遵循以 下原则。

(1)兼顾效率

数据冗余可以提高效率，但数据冗余增加了系统的复杂度，带来额外的空间开 销和管理维护开销，不应该滥用。在兼顾效率时要平衡效率与复杂度的关系。

(2)数据同步

源数据的增、删、改的权限在源数据归属的应用。当源数据发生变化时，辅数 据要及时同步变化。以保证数据的一致性和准确性。至于怎么样算“及时”,用什么 手段进行数据同步，要视具体冗余数据的性质而定。

(3)冗余数据的选择

由于存在数据同步的管理，为了减少同步的频繁度和复杂度，冗余数据通常应 该是一些静态数据或者同步实时性要求不高的数据。

3)相关过程域

战略过程域中的实践执行结果；

业务词汇表过程域和元数据管理过程域提供了该类型信息的识别、分类和管理指导；

数据生命周期管理过程域提供了关于数据用于何处的信息；

数据质量策略过程域提供了有关质量方面的信息。

**2.GQM 表**

|  |  |  |  |
| --- | --- | --- | --- |
| 目 标 | 问 题 | 度 量 项 | 度量项说明 |
| G1:平台满足 经批准的要求和 架构 | Q1:平台是“建  构”还是“购买”, 支持什么形式的数 据、数据交换和接 口 | M1:数据管理平台策略 对数据治理目标的支持 | 数据管理平台策略的目标应包括性能  指标、可扩展性要求、可维护性要求、数 据质量要求、安全要求等内容 |
| M2:数据管理平台具有 选择过程 | 数据管理平台的选择需要数据管理专员 及相关领域的专家积极参与，软件项目经理  也应投入足够的时间参与到持续不断的沟 通中，以确保“建构或购买”决策会考虑到  目标数据架构并支持数据管理策略 |

 **■云端数据治理**

**续表**

|  |  |  |  |
| --- | --- | --- | --- |
| 目 标 | 问 题 | 度 量 项 | 度量项说明 |
| G1:平台满足 经批准的要求和 架构 | Q1:平台是“建  构”还是“购买”, 支持什么形式的数 据、数据交换和接  口 | M3:数据管理平台具备 可扩展性、灵活性 | 数据管理平台解决了所需的可扩展性、 灵活性，能适应预期复杂性的变化以及数 据量和用户数量 |
| G2:存在并遵  循有效的平台管  理过程，以满足业 务需求 | Q2:该组织如何 根据数据、用户以 及整体复杂性的预 期增长解决平台的 可扩展性、安全性 和灵活性 | M4:平台支持该组织的 安全性和访问要求 | 安全性要求包括数据安全、认证安全、 系统安全、网络安全等内容，平台应采用  由各个业务部门推动并在职责划分的原 则下进行多级别访问管理及访问限制 |
| M5;工作流程和服务水 平要求 | 组织应识别工作流程和服务水平要求  然后确认并访谈适当的业务人员，记录他 们提出的具体问题，并进行分析 |
| M6:平台性能数据与业 务绩效需要和容量要求的 关联 | 数据交付活动的状态数据包括平均查  询响应时间、每天访问的用户数量、平台 用户的反馈等 |
| M7:平台变更控制 | 统计数据包括数据、查询、报表活动的 频率和资源占用情况等，组织应对数据管 理平台的定性和定量性能指标进行分析， 以支持平台变更决定 |
| M8:统计学方法 | 组织应将平台变化的影响与预测模型  进行比较并分析，以改进预测模型，提高 预测模型的有效性，实现预纠错 |
| G3:该平台提  供可信的数据 | Q3:如何定义、  选择权威数据源， 并将其集成到平台  的特定部分?如何 解决数据重复 | M9:关键元数据 | 数据管理平台是关键数据元素的权威  源和记录系统，组织应将关键元数据形成 书面文件 |
| M10:数据重复记录 | 组织在特意采用数据重复的情况下，应将 其形成文件，并且应针对该组织的预期收 益，说明保持重复源同步有关的成本的合理 性(包括数据不能保持同步的情况的影响) |
| M11:平台元数据对组 织元数据存储库的支持 | 数据治理成功的关键在于元数据管理，  元数据管理组织应监视复杂数据集成环 境中的变化，同时交付可信、安全的数据。  元数据存储库作为对象存储体系的一个 核心部分，是实现分布式对象分布存储的 关键。平台应将其元数据提供给该组织的  元数据存储库 |

**第8章■云端数据架构**

**3.** **成熟度等级介绍**

**Level 1: 初始级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 1.1 | 将数据管理平台和组件记录下来 | ●数据管理平台及组件列表 |

【成熟度等级说明】

目前，云端数据管理平台主要有 SQL、N oSQL 和 NewSQL 三种。其中，NewSQL 具有海量数据的存储管理能力，还支持ACID 和 SQL 等特性。应记录云端数据管理 平台和组件。例如，确定并记录在项目中用于管理关键数据的自动化系统和技术。

**Level 2: 管理级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.1 | 数据管理策略中列出的指标和目标通过平  台实施来支持 | ●数据管理策略 |

【成熟度等级说明】

数据管理策略的目标一般包括性能指标、可扩展性要求、可维护性要求、数据 质量要求、安全要求等内容。在平台开发期间以及发布之前，根据数据管理策略中 指定的目标对平台进行验证和确认。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.2 | 数据管理平台具有选择过程 | ●数据管理平台技术实现管理程序 |

【成熟度等级说明】

数据管理平台的选择需要数据管理专员及相关领域的专家积极参与，软件项目 经理也应投入足够的时间参与到持续不断的沟通中，以确保“建构或购买”决策会 考虑到目标数据架构并支持数据管理策略。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.3 | 平台支持该组织的安全性和访问要求 | ●数据管理平台访问控制程序 |

【成熟度等级说明】

安全性要求包括数据安全、认证安全、系统安全、网络安全等内容，平台应采 用由各个业务部门推动并在职责划分的原则下进行多级别访问管理及访问限制。如 对数据管理平台与应用系统之间的信息通讯进行加密，使用基于SSL 的 https 的加密 传输等。

**■云端数据治理**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.4 | 数据治理机构提供建议并同意重大的平台 决策 | ●数据管理平台决策标准 |

【成熟度等级说明】

通常情况下，治理使用一个框架以记录并评估具有竞争性的平台解决方案和选 项。该框架应包括数据库架构、数据仓库、稳定和内容架构以及元数据架构的指导 决策的标准。组织制订数据管理平台决策标准时应考虑组织未来业务的发展需求、 技术水平、研发人员、资金投入等情况。

**Level 3: 定义级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.1 | 对跨系统的重复数据集进行记录和计划， 并说明理由 | ●数据重复统计记录 |

【成熟度等级说明】

数据冗余会让用户产生困惑，对业务操作产生负面影响，带来更高的代价，并 显著增加数据管理环境的复杂性。然而，数据重复具有合理的理由，以满足特定的 功能性和性能要求。例如，为数据挖掘应用程序和数据仓库预处理大型数据集。在 特意采用数据重复的情况下，应将其形成文件，并且在各种情况下，应针对该组织 的预期收益，说明保持重复源同步有关的成本的合理性(包括数据不能保持同步的 情况的影响)。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.2 | 制订平台实施计划 | ●平台实施计划 |

【成熟度等级说明】

平台实施计划应考虑所需的可扩展性、灵活性和安全性，以适应预期复杂性的 变化以及数据量和用户数量。实施计划应包含一系列支持系统建设、测试及部署的 管理活动，包括：

●管理平台实施及在开发、测试环境中的变更管理；

●创建测试数据，包括必要的安全程序，如数据脱敏；

●开发数据迁移及转换程序；

●数据质量需求验证；

● 用户培训的建立及交付；

● 有效文档的开发。

**第8章■云端数据架构**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.3 | 平台设计和平台能力应确保可满足工作流 程和服务水平要求 | ●工作流程和服务水平要求 |

【成熟度等级说明】

项目一般是由项目需求驱动的，项目需求包括项目需求的引导、组织、记录、 评审、完善、批准和变更控制。需求可以同时使用文字和图形来表述。概念数据模 型主体域是一个很好的工具，它可以反映与业务流程和应用功能相关联的数据。

应识别工作流程和服务水平要求，然后确认并访谈适当的业务人员，记录他们 提出的具体问题，并进行分析。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.4 | 收集、存储并使用平台性能数据，以验证 该平台符合业务绩效需要和容量要求 | ●平台性能数据列表 |

【成熟度等级说明】

透明性和可见性是数据管理平台监控的关键原则，监控的一种最佳实践是仪表 盘，仪表盘可以展示数据交付活动的状态，让相关方看见并理解数据管理工作，比 如，监控平均查询响应时间、每天访问的用户数量。另外定期收集平台用户的反馈 也很有效。

**Level 4: 量化级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 4.1 | 对数据管理平台的定性和定量性能指标进 行分析 | ●数据管理平台性能指标分析记录 |

【成熟度等级说明】

对使用情况的统计数据和使用模式进行定期分析是至关重要的。对数据、查询、 报表活动的频率和资源占用情况进行统计分析，有助于对性能进行优化。

应使用统计和其他定量技术，对数据管理平台的定性和定量性能指标进行分析， 以支持平台变更决定。

**Level** **5:** **优化级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 5.1 | 以定量方式表达平台改进目标，并通过治 理批准 | ●资源占用、响应时间统计报表 |

【成熟度等级说明】

基于使用模式和统计数据而创建索引和聚集是最佳的优化手段。

 **■云端数据治理**

应通过对资源占用情况、响应时间进行统计分析，了解瓶颈在哪里，应该在哪 里进行优化，并经领导批准，形成优化方案。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 5.2 | 组织基于统计性能数据和因果分析持续改 进平台，同时改进预测模型 | ●性能预测模型  ●预测模型改进措施  ●预测模型改进过程记录 |

【成熟度等级说明】

通过使用性能预测模型，对数据管理平台进行优化。运行中的数据管理平台， 由于受系统环境的制约，其性能会慢慢下降，可利用kalman 等原理预测数据库的性 能。实际应用时应调整网络、服务器等相互依存的组成部分，使其性能同步达到最 佳。同时，应考虑在提高数据库性能的同时，兼顾安全的平衡。

将平台变化的影响与预测模型进行比较并分析，以改进预测模型，提高预测模 型的有效性，实现预纠错。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 5.3 | 组织在其所在的行业内分享与数据管理平 台的设计、开发、部署和运作相关的经验 | ●与数据管理平台相关的发表作品 |

【成熟度等级说明】

组织数据管理专业人士应订阅专业杂志，加入网上各种论坛、专业讨论群，分 享或阅读数据管理平台相关主题的文章。还可以公开发布白皮书、论文、论著、最 佳实践等。

**4.** **实施案例**

下表为某集团用户数据库管理规范示例。

|  |
| --- |
| 某集团用户数据库管理规范  1 适用范围  本规范规定了用户数据库(UDB)进行基础用户数据进行集中管理、应用、开放的功能、接口、流程、集成方 面内容。  本规范适用于集团针对个人基础用户帐户进行宽带接入、业务网络使用、外网服务使用。  本规范设计与开发过程应保持“松耦合、模块化、可配置”原则，对于统一版本应在模块和代码中予以标识，对于 各子公司定制开发部分，应在程序名称、模块区隔中提前设计明确位置，并在自开发文档说明中给出明确标识。  2 需 求 约 束  (1)此规范作为架构人员设计架构的依据；  (2)系统提供验证映射帐户的接口，可以通过配置是否启用该功能；  (3)系统采用开放式的设计思路，对服务方法，协议适配采用插件的方式，以方便后续的扩展。 |

**第8章■云端数据架构**

续表

|  |
| --- |
| 3 UD B发展规划  用户数据库(UDB)定义：集团统一账号数据的安全状态服务平台。  1)第一阶段  用户数据库(UDB)主要数据：用户号码数据、状态数据、密码数据、配置数据。  用户数据库(UDB)主要功能：对于业务系统的管理平台进行用户数据安全及状态寄存服务，账号逻辑层面将 号码落实到个人，业务层面拓展到接入和应用多层面全业务。  用户数据库(UDB)后台支撑：UDB统一产生一次性密码，并经两种渠道下发：短信一次性密码方式和令牌方 式。  2)第二阶段扩展  在功能上，拓展群组管理，并增加个人通信录、自服务等用户能力，与支付、门户等业务能力进行集成，针对  用户验证工作进行细化和拓展。  在架构上，逐步发展成为业务层和接入层提供核心号码及密码的安全和状态服务的综合认证平台。  4 UD B功能描述  UDB系统的主要工作是完成用户统一账户(172账户)的集中服务。它维护各个业务系统的账户到用户统一账  户的映射关系，将各个独立的业务整合成以172账号为中心的一个大的平台，用户只需要一个172账号和密码， 即可使用诸各种业务。  我们将UDB系统划分为四大部分，即：协议适配，服务，后台管理，基础数据。  (1)协议适配：该部分的组件主要用来和各种业务系统进行交互，从不同网络类型中接收不同的种类的协议封 装的数据，统一转换成UDB可以识别的格式。当服务等操作完成后，又将数据封装成特定的协议，通过不同的  网络类型返回给各应用系统。  (2)服务：该部分是UDB系统的核心部分，它对各种业务系统的服务请求进行服务。这部分可以执行严谨的  服务策略，可以实现多种服务方式。  (3)后台管理：该部分包括系统管理，和CRM系统的接口封装，用户管理，登录审计等多种操作。  (4)基础数据：该部分是服务和用户管理联系的桥梁，用户管理部分所操作的服务策略、操作日志记录等都保 存到数据库中。服务部分需要使用这部分的数据。  5 UD B接口及描述  UDB和周边系统接口如下：  (1)子公司UDB与业务支撑系统之间的接口：包括与归属地CRM之间的接口，涉及业务受理时的账号开通、 用户账号信息更新的通知等。  (2)子公司UDB与集团UDB之间的接口：包括集团UDB到子公司UDB的账号信息查询、相互维护的账号信 息更新；集团UDB平台做漫游认证的转发等服务。  (3)UDB与业务系统之间的接口，主要是账号认证、用户业务注册和上线状态通知、业务密码设置、短信密码 获取请求等消息交互。  (4)UDB自身管理配置的接口：主要传递系统管理操作员的登录认证，以及对UDB的配置、统计、管理维护 操作信息交互。  6 UDB接口通信实现框架  通信协议层包括：Socket、FTP、HTTP等。支持RADIUS和SAML协议族，并通过UDB协议适配组件进行 适配。 |

 **■云端数据治理**

续表

|  |
| --- |
| 消息处理层包括两种情况： 一种是对通信协议交互关系的处理；另一种是对应用数据提供基本的解析机制。为 了提高系统的可扩展性，消息数据采取XML格式进行封装。  应用层：提供应用语义处理，实现各自接口的业务逻辑。  7 UDB接口安全方案  由于UDB系统需要和互联网业务相连接，我们需要从多个层次来保护UDB系统的安全性。我们从传输层、会 话层、应用层三个层面来考虑系统的安全性。  (1)传输层：  在传输层我们可以考虑使用如下的一种或者多种方法的组合：  a)考虑通过专线连接外网SP登录服务器与内网UDB服务器  b)不采用专线，连接通过intermet,但系统不直接暴露在intemet上，而是放到受防火墙保护的DMZ区域，在 防火墙本身设置策略，只是允许指定的SP服务器的IP和PORT连接到UDB的指定端口；  c)在各业务系统和UDB的连接中采用隧道加密形式。  (2)会话层：  在会话层我们可以考虑使用如下的一种或者多种方法的组合：  a)UDB系统和外围系统的数据交互，建议采用加密协议进行(如SSL加密):  b)SP的业务系统连接UDB系统的时候采用双向数字证书服务。确保正确的SP服务器连接到正确的UDB服 务器。  (3)应用层  在应用层我们可以考虑使用如下的一种或者多种方法的组合。  a)数据中账户的密码全部加密保存到数据库中；  b)由于internet上的业务很容易受到攻击，所以对外网业务使用OTP(one time password);  c)对用户提供的密码实行强密码验证，系统管理员可以设置强密码规则。如：密码必须由字母、数字、特殊符 号组成，密码位数在8～18位之间等；  d)对用户服务，系统管理员的操作都记录详细的操作日志，记录访问人、访问时间、访问原始IP、访问目标IP、 访问的动作、访问的结果等。  8数据编码规范(略)  9 业 务 场 景  (1)账号申请流程  在UDB中需要维护账户和其他业务系统账户的关联关系，以172账号作为主ID账号，其他业务系统账号和这 个主ID进行关联。对于新开通172账号的用户，可能还没有或者部分拥有其他服务账号，可以在开通172业务的  时候， 一起开通其他业务系统账号，所以在新开172账户的时候可以能有如下两种情况： a)用户已经有其他业务系统账号  在开通172账号的时候，用户提供其他业务系统的账号和密码。为了防止虚假信息，在绑定业务系统的账  号的时候，需要将用户提供的业务系统的账号和密码交给各个业务系统进行验证，只有验证通过了才可以实 施绑定。  b)用户没有其他用户业务系统账号  如果开通172账户的时候，用户没有开通其他业务系统账户，可以在开通172账户的同时选择为用户开通其他 业务系统账号。系统将实现账号间的映射关系。 |

**第8章■云端数据架构**

续表

|  |
| --- |
| (2)账号绑定流程  用户可以在开通172账户的时候选择同时开通相关的业务系统账户，也可以在后续自己来绑定业务系统相关的 账户。  (3)账号解绑流程  用户可以自己解除业务系统账号和172账号的绑定关系，如果解除了绑定关系，用户将不能使用172账户登录 相应的业务系统。  (4)密码修改流程  用户可以登陆UDB系统修改密码，也可拨打集团客户电话修改密码。  (5)账号升级流程  172账户是可以作为通讯号码和服务号码的。UDB需要维护172账户的这种状态。当一个172用户从服务号码  升级成通讯号码的时候，UDB需要更新此状态信息。维护此状态的好处是：通过该识别，UDB可以判断是否可  以将信息发送到用户的手机。账号升级的过程很简单，更新172用户在UDB数据库中的一个状态字段即可。为  了保证账号和密码的安全性，我们在服务方法上做了一个约定：外网服务的必须通过OTP(One Time Password) 进行服务，如果是内网服务可以通过OTP、用户的固定密码、手机短消息发送动态密码的方式进行服务。  (6)用户使用基础账户登录业务系统流程  用户使用172账号登录业务系统，业务系统识别出172用户，并且把用户服务发送给UDB系统进行服务，UDB  服务用户，将服务结果(172账户关联的业务系统账号传送给业务系统),业务系统认证UDB服务后，用户按照 原来的情况使用业务系统。 |

**8.2.4** **历史数据管理**

**1.** **简介**

1)目的

确保数据保存满足组织和监管对历史数据可用性的要求，并确保数据归档和销 毁满足法律和监管要求。

2)介绍

云计算环境下数据本身发生了很多变化，最明显的特征是数据体量暴增，组织 面对大体量的数据越来越多地使用云计算技术。例如租赁云存储空间以保存组织数 据，租赁云主机对组织数据进行处理等。云计算资源的租赁会占用组织成本，因此 组织必然会在意云计算资源的使用效率，其中云储存空间的高效合理使用就是重要 的一项内容。为了提高存储空间的利用率，降低组织因存储和处理将要过时或不再 产生价值的数据而消耗的成本，那么对历史数据的正确管理是必须要开展的。除了 节约成本的原因，外部监管对历史数据的处理要求也是一个非常重要的原因。

历史数据管理定义了如何管理历史数据(例如数据变更、版本控制)以及生命 周期后期阶段的数据管理，包括如下内容。

(1)数据保存和归档；

 **■云端数据治理■**

(2)压缩数据存储空间，提高利用率，降低存储成本；

(3)恢复数据到某个时间点的状态；

(4)保证内部和外部存放要求得到满足。

历史数据管理的主要目标是确保历史数据由组织以一致的方式进行管理，以满 足业务需求。要做到这一点，组织需要定义数据保存标准，以及用于数据归档的标 准过程和一套准则。通常情况下，监管要求推动着这些准则的制订和完善，以规定 存档数据的保存期限、保存方式、加密方式等。这些准则也体现了组织内部和外部 对其数据保存的要求，这通常结合了监管和业务要求。对于现有的数据存储区和计 划的举措，明智的做法是确保各个业务部门了解并明确定义其当前保存要求。这是 制订归档政策和策略的先导。

历史数据可以在两大模式下进行保存：在线可访问或归档。建议组织定义一套 用于说明和实施数据历史的核心规则(例如数据建模规则、历史更新等),在符合删 减要求的情况下，随着时间对数据进行处理来获得效率。组织应收集、发现并评审 历史业务过程(例如预测、报告和操作过程)以及合规性的主要要求，以指导对历 史数据的有效管理。

所定义的归档过程应制订适用的业务规则，确定何时归档数据以及如何存储、 访问或恢复数据。为了确定哪些记录满足该标准，应定期监测归档记录的数据存储 区。最后，性能改进很可能会通过旧数据的存档得到支持。对数据归档时，应先确 定旧数据的权威来源，还应确定数据是否可以采用可写模式(相对于只读模式),并 按照数据优先性和使用方式(例如近线与远程等)以及其他安全规则考虑可访问性。

3)相关过程域

数据提供者管理过程域定义提供了关于收集和说明数据要求的信息；

架构方法过程域包含了关于定义数据资产和能力的高层次的业务要求信息。

**2.GQM 表**

|  |  |  |  |
| --- | --- | --- | --- |
| 目 标 | 问 题 | 度 量 项 | 度量项说明 |
| G1:利用通用  标准，对历史数据 进行一致的管理 | Q1:适用于历史数据管  理的架构标准和最佳实践 是什么,如何定义和治理相 应的业务规则 | M1:历史数据管理 策略 | 组织应公布并遵守数据保存策略， 所定义的归档过程应要求指定适用的 业务规则，以确定何时归档数据以及 如何存储和访问或恢复数据 |
| G2:满足收集 和存储历史数据 的业务需求和监 管需求 | Q2:如何保证数据保存 达到所要求的时长 | M2:历史数据访问 策略 | 组织应基于业务需求和技术策略制  订历史数据使用管理办法。不同的数据 有不同的归档场景，制订数据归档策略 时，应对数据使用的需求进行分析，根 据分析的结果，结合法规、技术、访问 成本等，梳理数据的归档场景 |

**第8章·云端数据架构**

续表

|  |  |  |  |
| --- | --- | --- | --- |
| 目 标 | 问 题 | 度 量 项 | 度量项说明 |
| G2:满足收集 和存储历史数据 的业务需求和监 管需求 | Q2:如何保证数据保存 达到所要求的时长 | M3:历史数据存储 | 组织应具有规定的数据仓库，提供 历史数据访问，以支持业务过程的分 析需求 |
| Q3:如何保持归档的数 据的完整性 | M4:历史数据恢复 指南 | 组织应制订备份和恢复功能的指  南，明确恢复时间目标(RTO),以确 保任何具体时间点的数据文本可以被  重新创建 |
| G3:数据归档 满足了组织和监 管要求 | Q4:用于确定何时可以 删除存档数据的考虑因素 有哪些 | M5:归档数据恢复 测试 | 组织应建立一套定期测试的策略， 以确保数据恢复程序准确性 |
| M6:方案审核 | 方案应包括保存于各个特定时期的 数据或日志类型的备份和归档的恢复 测试方案，访问归档数据的技术使用 方法和维护方法等内容 |
| M7:历史数据归档 策略改进与优化 | 组织应制订面向用户的满意度调查 制度，定期收集相关方的反馈，以作 为优化归档策略的根据 |

**3.** **成熟度等级介绍**

**Level 1: 初始级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 1.1 | 历史数据应是可用的，能够支撑业务决策 | ●归档程序  ●历史数据清单 |

【成熟度等级说明】

组织对历史数据的管理不是简单地将其删除，而是在降低成本的目标下还要保 证历史数据在被需要时能够及时获取得到。随着组织业务的进行，较早前的数据可 能无须直接使用，组织需要对这些早期数据进行处理，以腾出空间给需要处理的数 据。例如，通常情况下，组织的发展历程数据一般不是业务需要的，但是当组织需 要对组织的发展历程进行梳理或者举办庆祝活动时，历史数据就是非常有价值的数 据。再例如，执法部门可能会拥有大量的视频监控材料， 一般情况下这些数据可能 不会产生价值，但当需要对视频材料进行回顾以发现某些可用线索时，这些视频材 料又会变得很有价值。因此一些数据的价值不是立马体现的，组织需要对这种数据 进行处理，以降低维护成本。但同时也得保证在组织需要这些数据时这些数据还是 可用的，以帮助组织做出业务决策。

**■云端数据治理**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 1.2 | 对所有的数据存储进行备份，并且按照策 略规定进行数据归档 | ●数据归档或销毁程序 |

【成熟度等级说明】

基于策略和连续性运行的要求， 一些数据备份可以存储在一个独立的位置，以 减轻与主数据位置相关的风险。此外，根据备份指南备份数据变更日志也很重要。 数据变更日志应收集变更用户或过程的ID 以及变更时间等事项。

**Level 2: 管理级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.1 | 制订历史数据的管理策略 | ●历史数据管理策略 |

【成熟度等级说明】

制订历史数据的管理策略，包括保存、恢复、销毁和审核跟踪等要求。在制订 管理策略是应特别注意：在云计算时代，数据急剧增长，与传统的数据备份、数据 归档不同，云端数据的归档更需要关注数据选择性恢复的功能。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.2 | 定义方法以确保业务需要时能对历史数据 进行必要的访问 | ●历史数据使用管理办法 |

【成熟度等级说明】

不同的数据有不同的归档场景，制订数据归档策略时，应对数据使用的需求进 行分析，根据分析的结果，结合法规、技术、访问成本等，梳理数据的归档场景。 基于业务需求和技术策略制订历史数据使用管理办法。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.3 | 对选定的归档数据进行恢复测试 | ●归档数据恢复管理程序 |

【成熟度等级说明】

一般情况下，不需要对所有归档和备份的数据进行恢复，但需要建立一套定期 测试的策略，以确保数据恢复程序准确性。恢复测试应按照由组织实际使用的技术 类型来执行。

**Level 3: 定义级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.1 | 设立合适的数据仓库 | ●支持历史数据访问的数据仓库 |

**第8章■云端数据架构**

【成熟度等级说明】

数据仓库中的每一条数据在某个时间点都是准确的，并常常在其结构中体现时 间元素。因此，可以将数据仓库想象为数据快照的历史记录，在每一快照对应的时 间点上进行记录。组织应具有规定的数据仓库，提供历史数据访问，以支持业务过 程的分析需求。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.2 | 任何具体时间点的数据文本可以被重新创建 | ●数据恢复指南 |

【成熟度等级说明】

定义目标有助于指导资源和任务的合理分配，制订必要的战略，以及对计划和 程序的整体经济性做出合理判断。 一旦设定目标，就能为计划的实际制订过程提供 指导。因此，当制订备份和恢复功能的指南时，组织应明确恢复时间目标 (RTO)。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.3 | 通过数据治理来定义和批准政策，并在要求 数据更改记录和日志保存的组织层面上实施 | ●数据记录政策 |

【成熟度等级说明】

哪些数据需要归档，主要与监管要求及企业战略有关。组织根据监管要求和企 业战略，明确热数据、温数据和冷数据之间的界限，确定数据的归档政策，并根据 归档政策进行归档。因此，组织的数据保存和归档政策应与监管要求保持一致。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.4 | 审核方案应确保符合组织的数据记录、保 存和归档策略 | ●审核记录 |

【成熟度等级说明】

该方案应包括保存于各个特定时期的数据或日志类型的备份和归档的恢复测试 方案，访问归档数据技术的使用方法和维护方法等内容。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.5 | 与数据利益相关者保持沟通，建立意见 反馈机制 | ●沟通策略  ●反馈机制 |

【成熟度等级说明】

提高利益相关方的认可及参与程度，加强问题管理，将有效提高历史数据管理 的成功率。组织可以通过如下方法加强与利益相关方的沟通。

●为数据管理项目维护一个内部网站；

● 在企业内部门户发布公告；

 **■云端数据治理**

●出版简讯或电子邮件分发。

应积极利用已建立的沟通渠道，接受利益相关方的反馈，以确认现有数据保存 和归档策略的效果。

**Level** **4:** **量化级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 4.1 | 使用定量分析技术或指标，对历史数据进 行分析 | ●历史数据分析记录 |

【成熟度等级说明】

对使用情况的统计数据和使用模式进行定期分析是至关重要的。对数据、查 询、报表活动的频率和资源占用情况进行统计分析，有助于对历史数据管理效能 进行优化。

使用统计方法学等定量技术分析历史数据，以改善数据质量，同时有助于加强 对历史数据的管理。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 4.2 | 利用测量指标结果和利益相关方的反馈， 来改进数据归档策略 | ●用户满意度调查表  ●历史数据信息测试记录 |

【成熟度等级说明》

制订面向用户的满意度调查制度，定期收集相关方的反馈，以作为优化归档策 略的根据。满意度调查可利用互联网技术，建立满意度调查问卷的模板，客户只需 要将问卷稍作调整便可以直接使用。对于满意度的结果，系统自动生成调查结果。 为了便于对比，可以对历届数据自动对比，生成简单的统计分析报告。

**Level 5: 优化级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 5.1 | 组织应在其行业内分享关于历史数据和归 档的经验和最佳实践 | ●与历史数据相关的演示材料和其他发表的作品 |

【成熟度等级说明】

组织负责数据管理的专业人士应订阅专业杂志，加入网上相关论坛、专业讨论 群，分享或阅读历史数据管理相关主题的文章。组织应当发布历史数据管理的最佳 实践，或发表论文、专著、白皮书等出版物为其他组织提供参考。

**4.** **实施案例**

组织应了解业务对历史数据归档、保存、删除和检索等要求，并通过制订历史 数据管理政策、标准以其满足业务及监管的要求，并对实施过程进行一致性监测。

**第8章■云端数据架构■**

实施过程中可借助物理数据控制、存储检索方法、数据归档和检索的自动化工具实 现历史数据归档管理的标准化。

下面以 SAP 系统提供的归档功能为例，介绍其数据归档流程。

|  |
| --- |
| SAP系统提供的归档功能  SAP数据归档流程分为三个阶段：  ●创建存档文档——归档程序将R/3数据库的归档数据写到归档文档。  ●转移归档文档——将归档文档移至第三方存储介质。  ●删除数据——删除程序先从归档文档中读取数据，再从数据库删除它。  通常而言，仅完成以上三个阶段完成并不足够，归档文档还需要被安全存储和管理，以便日后需要的时候 读取。  1 归档步骤一：归档数据  开始归档活动前，可以选择是否运行预处理程序。在第一步中，归档程序创建一个归档文档。将被归档的数据 从数据库中后台读取并且写入归档文档。这个过程持续到以下结果发生：  (1)归档完成：所有选择数据被读取和写入。  (2)归档还未完成，但是归档文档内容已满。  (3)归档还未完成，但是归档文档包含客户定制的特定对象数量。  只有来自完结了的业务流程的数据和所选财年的数据可以被归档。  主数据只有在被标记了删除标记后才能被归档。 一些归档对象在一个特殊流程程序设置了删除标记，例如  CO\_ORDER。  2 归档步骤二：转移归档  对于归档文档在第三方介质上的存储和管理有许多解决方案。你可以在归档管理中决定归档文档现在放在什么 位置。  连接到分级存储管理系统(HSM):一个HSM系统模拟一个无限大的文件系统。归档文档在HSM系统中，数  据归档被存储时立即创建一个归档文档。在客户定制时，足以为归档对象指定HSM文件系统的路径，也就是归档  文档的目标路径。  手动管理：如果你不想将归档文档存储在一个外部存储系统或者HSM系统，IT部通过在磁带或者本地制作的 CD管理文档，因此它们没有对SAPR/3系统的接口。  确保你为所选择的第三方存储介质遵循了恰当的备份策略。几年后，也许将这些CD和WORM上的存储内容再  次拷贝到新的CD和WORM上。  3归档步骤三：删除数据  (1)方式一：归档和删除任务同时运行  删除工作在归档文档关闭是自动开始。前提是归档对象在技术定制选择了“自动开始”。  关闭归档文档后，ADK打开一个新文件，继续开始归档过程。删除程序则同时开始，它从关闭了的归档文档  读取数据，然后删除数据库中的数据记录。这一场景，归档和删除程序同时进行。每个归档文档都有一个删除  过程。  (2)方式二：手动开始删除任务  归档任务完成后，管理员开始使用数据归档事务手动执行删除任务。前提是归档对象的技术定制选择了“没 安排”。 |

 **■云端数据治理**

续表

|  |
| --- |
| 所有的归档文档关闭后，管理员开始手动删除程序。归档任务一旦完成就关闭。每个归档文档都有一个删除 过程。  (3)方式三：自动开始删除任务  归档任务完成后，删除任务自动开始。前提是“结束后开始”在事务AOBJ的归档对象定义中被标记。  所有的归档文档关闭后，删除程序按照计划自动开始。归档任务一旦完成就关闭。每个归档文档都有一个删除 过程。 |

**第** **9** **章** **安全与隐私**

**9.1** **概述**

**9.1.1** **目的和意义**

安全应从战略、战术、操作角度来解决问题，并集成到组织的每个层次。安全 管理涵盖了组织安全开发、风险管理、合理的对策选择与实现、治理和绩效考核。 云端数据安全管理包括计划、制订、执行相关安全策略和规程，确保数据和信息资 产在使用过程中有合适的认证、授权、访问和审计等措施。有效的数据安全策略和 规程须确保合适的人以正确的方式使用和更新数据，并限制所有不适当的访问和更 新数据。

隐私和合规是信息安全非常重要的部分。理解并符合所有利益相关者的隐私 及保密需求，是任何组织的最佳利益所在。组织安全人员应该知晓数据安全法律 法规，这样他们才可以把预期的责任告知管理层等利益相关者，同时指导自己的 工作。

数据安全与隐私保护不仅仅是一个业务问题，应该作为组织的一项责任和义务 被正确对待。由于安全管理依赖于组织资源的投入，因此，安全必须被合理地融入 组织的整体业务目标。实际中，有越来越多的组织因为没有对安全及隐私问题给予 足够的重视、支持和投资而付出了惨痛代价。

隐私保护涉及的方面很多，如果组织想处理好隐私问题，应该清楚相关的法律 法规，确认自己可以做什么不可以做什么,确保自己的行为是合法的。组织应告知 员工什么行为是可接受的，可能会遭遇到什么样的监控，不满足这些要求可能会导 致什么样的后果。组织应通过策略、标准和安全培训清楚说明这些问题。

**9.1.2** **内容与关联性**

数据安全管理的最终目标是在符合隐私保护及法规要求的前提下，满足业务需 求。组织必须认识到客户、供应商、合作伙伴等利益相关者的隐私及保密需要，利 益相关者是其数据的最终拥有者，组织是数据的托管人。政府会制订一些法规限制 对相关信息的使用，以保护利益相关者的安全权益，同时明确组织的责任。数据安

**■云端数据治理**

全管理的主要活动如下：

(1)理解数据安全需求和监管要求

实现组织内部数据安全应对业务要求和需求进行透彻的了解，并在系统开发 项目的分析阶段就识别具体的应用安全要求。识别并整理出主要的隐私和安全法 规清单。

(2)定义数据安全策略

组织的安全管理人员在法律部门的协助下，基于数据安全要求定义数据安全策 略。作为 IT 安全策略的一部分，数据安全策略应以数据为中心，且更为精细。

(3)定义数据安全标准

定义数据加密标准和机制、互联网数据传输协议、远程访问标准、安全事故报 告规程等数据安全标准。

(4)定义数据安全控制机措施

设计组织自己的认证、授权、访问、审计等安全控制措施，以满足相关法律的 要求。

(5)管理用户、密码和用户组成员

由安全管理员创建、修改和删除用户帐户和用户组，并设立审批机制对用户 权限变更进行审批。在异构环境下，建议集中管理用户身份数据，以确保数据的 完整性。

(6)管理数据访问视图和权限

通过授予权限控制用户对敏感数据的访问，没有权限的用户无法执行任何操作， 以防止不适当的数据访问。

(7)监控用户身份认证和访问行为

通过主动或被动方式监控用户认证和访问行为，以检测异常或可疑的交易。

(8)划分信息密级

按密级分类模型，对组织数据和信息产品进行分类。安全管理员应负责数据密 级的评估和确定工作。

(9)审计数据安全

数据安全审计工作是无可代替的工作，应规律、高效地持续执行数据安全审计 工作。

风险管理过程域帮助组织识别云端数据面临的威胁，采取措施应对威胁，并持 续监控以确保风险级别处于可接受的级别。数据安全过程域强调通过访问控制、加 密技术等手段保护云端数据的安全，确保合适的人以正确的方式使用和更新数据， 并限制所有不恰当的使用和更新数据。隐私保护过程域论述了通过识别政策法规，

落实技术手段，以确保云端隐私数据不被滥用及未经授权使用。合规管理过程域强 调组织应建立和维护机制，保证自身的合规性。

第9章■安全与隐私

**9.2** **活动及要求**

**9.2.1** **风险管理**

**1.** **简介**

1)目的

识别云端数据面临的威胁，采取适当的措施将云端数据的风险降至组织接受的 范围。

2)介绍

风险管理是社会组织或者个人用以降低风险可能产生的消极结果的决策过程， 通过风险识别、风险评估、风险评价，并在此基础上选择与优化组合各种风险管理 技术，对风险实施有效控制和妥善处理风险所致损失的后果，从而以最小的成本收 获最大的安全保障。

百分之百安全的环境是不存在的。每种环境都具有某种程度的脆弱性，都面临 一定的威胁。正确的实施风险管理需要全面了解组织，了解它所面临的威胁并对威 胁进行分类和评估，知道应该采取什么样的应对措施来处理这些威胁，并持续监控 以确保风险级别处于可接受的范围。

在云端数据治理过程中，组织的管理层需要坚定支持风险管理，批准并发布风 险管理策略。组织应有计划地开展风险评估工作，重点关注安全和隐私问题，防止 未授权或不恰当地使用数据。

(1)制订风险相关的政策和策略，将风险控制在可承受范围内；

(2)监控和管理关键风险，降低其对组织的影响；

(3)通过风险管理策略来审查数据应用过程中可能产生的风险；

(4)建立风险管理效果评估指标，通过指标对数据应用过程中的风险进行全面 监控。

3)相关过程域

●参考云端数据战略过程域，了解组织风险管理的目标；

●参考数据生命周期管理过程域，了解云端数据生命周期管理内的管理策略；

●参考数据质量评估过程域，了解质量目标与质量控制活动保持相符的相 关信息。

 **■云端数据治理**

**2.GQM 表**

|  |  |  |  |
| --- | --- | --- | --- |
| 目 标 | 问 题 | 度 量 项 | 度量项说明 |
| G1:识别组织  面临的风险 | Q1:组织风险识别是否 完整 | M1:风险识别内容完 整性 | 风险的识别应包括识别需要保护  的资产并赋值，识别资产存在的脆弱 性并赋值，识别资产面临的威胁并赋 值，确认已有的安全控制措施等内容 |
| M2:风险识别和风险 分析流程 | 组织应建立方法、流程，用于在风 险事故发生之前系统地、连续地认识  所面临的各种风险以及分析风险事 故发生的潜在原因 |
| G2:管理好关  键风险 | Q2:组织是如何管理关 键风险的 | M3:风险监控 | 组织应根据风险要素识别的结果，  按照一定的计算方法计算安全事件  发生的可能性及安全事件造成的损 失，并根据风险计算结果综合判定风 险等级，排好优先次序 |
| M4:风险定量分析 | 组织应建立有效的模型，通过不同 来源的数据，使用数字化的尺度来描 述后果和可能性 |
| G3:将组织的 风险控制在可承  受的范围内 | Q3:组织是否有文件化 的风险管理程序 | M5:风险管理策略 | 组织应根据ISMS、行业特定因素  制订自己的风险管理办法：全面了解 组织及其面临的威胁，采取适当的措 施应对这些威胁，并持续监控以确保 风险级别处于可接受的级别 |
| M6:风险趋势分析 | 组织应建立对风险规划、识别、估 计、评价、应对全过程的监视和控制 机制，从而核对风险管理策略和措施 的实际效果是否与预见的相同，并寻 找机会改善和细化风险规避计划 |

**3.** **成熟度等级介绍**

**Level 1: 初始级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 1.1 | 识别并监控风险 | ●数据风险列表  ●现有安全控制措施列表  ●资产重要性评估列表 |

**第9章■安全与隐私**

【成熟度等级说明】

风险的识别应包括识别需要保护的资产并赋值，识别资产存在的脆弱性并赋值， 识别资产面临的威胁并赋值，确认已有的安全控制措施等四部分内容。

**Level 2: 管理级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.1 | 对已识别的风险进行分析、监控，并按优 先级进行排序 | ●数据风险等级列表  ●重要资产列表 |

【成熟度等级说明】

根据风险要素识别的结果，按照一定的计算方法，如矩阵法，计算安全事件发 生的可能性及安全事件造成的损失。

根据风险计算结果综合判定风险等级，并排好优先次序：引致最大损失及最可 能发生的事情优先处理、而相对风险较低的事情则次之。

**Level 3: 定义级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.1 | 建立和维护风险管理策略 | ●风险管理办法  ●风险处置计划 |

【成熟度等级说明】

安全预算有限，脆弱性的数量却很多，如何才能确定关键脆弱性的处理顺序、 保证组织优先解决最紧迫的问题并获得最大的投资回报，是风险管理的关键所在。

组织应根据 ISMS、行业特定因素，制订自己的风险管理办法：全面了解组织及 其面临的威胁，采取适当的措施应对这些威胁，并持续监控以确保风险级别处于可 接受的范围。

NIST 推荐的信息安全风险管理指南包括下列步骤：

● 系统特征描述

●威胁识别

●脆弱性识别

●控制分析

●可能性确定

●影响分析

●风险确定

●控制建议

●结果记录

**■云端数据治理**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.2 | 建立风险识别、分析的流程 | ●风险分析程序  ●风险分析记录  ●风险分析报告 |

【成熟度等级说明】

风险分析有下列主要目标：

● 识别资产和它们对于组织的价值；

●识别脆弱性和威胁；

●量化潜在威胁的可能性及其对业务的影响；

●在威胁的影响和对策的成本之间达到预算的平衡。

组织应建立方法、流程，用于在风险事故发生之前系统地、连续地认识所面临 的各种风险以及分析风险事故发生的潜在原因。

**Level 4: 量化级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 4.1 | 对风险进行定量分析 | ●数据风险分析记录 |

【成熟度等级说明】

建立有效的模型，通过不同来源的数据，使用数字化的尺度来描述后果和可能 性(而不是定性评估中所使用的描述性尺度)。在很多情况下，定量估算使用历史的 事件数据，优势是其直接与信息安全目标和组织所关心的问题相关。

**Level** **5:** **优化级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 5.1 | 建立风险监控机制，并对风险趋势进行分 析 | ●数据风险监控管理程序  ●风险监控记录  ●风险管理策略改善记录 |

【成熟度等级说明】

风险是随着内部/外部环境的变化而变化的，它们在经营活动的推进过程中可能 会增大或者衰退乃至消失，也可能由于环境的变化又生成新的风险。应建立对风险 规划、识别、估计、评价、应对全过程的监视和控制机制，从而核对风险管理策略 和措施的实际效果是否与预见的相同，并寻找机会改善和细化风险规避计划。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 5.2 | 密切关注最新风险管理理论及技术 | ●最新风险管理研究成果  ●监控、审计机制优化方案 |

**第9章■安全与隐私**

【成熟度等级说明】

密切关注国内外的风险管理的方法论，持续优化安全监控和审计机制。持续关 注国内外最新的风险管理技术方法，提升风险发现和跟进的能力。

**4.** **实施案例**

数据风险评估是风险管理的重点内容，风险管理的实施过程包括：

(1)风险评估与管理培训

公司请风险评估专家为公司进行风险评估与管理技术培训，培训对象为信息安 全管理小组成员及部门的信息安全员，以及数据管理员。通过培训，使大家掌握数 据资产识别和风险评估的方法、步骤、技巧。

(2)风险评估

公司信息安全管理小组同部门信息安全员在所确定的体系范围内，分析组织内 与云端数据安全有关的资产，对资产进行识别与评价，识别和评价所面临的安全威 胁、薄弱点及其对数据安全的影响，确认已有的安全控制措施，实施风险评估。最 终输出数据风险评估报告。

(3)风险处理计划的制订

根据风险评估结果制订风险处理计划，计划主要列出对目前存在的风险应采取 的安全控制措施、职责和优先级。

(4)文件的制订

组织的信息安全管理小组应制订风险管理相关的程序文件。编制的文件包括： 《风险评估管理程序》等。

(5)文件评审和发布

文件应经过管理层审批后才能发布。

(6)控制措施的实施

组织应建立监督机制，监控控制措施的运行。

与该职能域相关的程序文件主要包括《风险评估管理程序》。

|  |
| --- |
| 风险评估管理程序  1 目的  为了明确公司云端数据治理体系风险评估的准则，指导风险评估的实施和风险控制措施的改善工作。  2 适 用 范 围  适用于公司的数据资产风险评估过程。  3 定 义 和 缩 写  3.1资产(asset):即信息资产(与信息相关且对企业有价值的资产),是安全策略保护的对象，此处尤指数据  资源。  3.2资产组(Asset Team);又称资产组合，是指具有相同或相近的业务功能的资产组合；如由硬件、软件、配 置信息等组成的应用系统资产组，由电子文件、纸质文档组成的信息资产组。 |

**■云端数据治理**

**续表**

3.3 资产价值 (asset value): 资产的重要程度或敏感程度的表征。资产价值是资产的属性，也是进行资产识别 的主要内容。

3.4 机密性 (confidentiality): 数据所具有的特性，即表示数据所达到的未提供或未泄露给未授权的个人、过程 或其他实体的程度。

3.5 完整性 (integrity): 保证数据及数据系统不会被非授权更改或破坏的特性，包括数据完整性和系统完整性。

3.6 可用性 (availability): 数据或资源的特性，被授权实体按要求能访问和使用数据或资源。

3.7 残余风险 (residual risk): 采取了安全措施后，仍然可能存在的风险。

3.8 威胁 (threat): 可能导致发生对系统或组织产生危害的不希望事故的潜在原因。

3.9 脆弱性 (vulnerability): 是指可能被威胁所利用的资产或若干资产的弱点，又称弱点，脆弱性和资产本身

的特点和性能有关。

4 职责

|  |  |
| --- | --- |
| 部门/岗位 | 过 程 内 容 |
| 数据治理委员会 | 组织各部门的数据资产识别和风险评估工作 |
| 数据管理员 | 数据资产识别和风险评估，部门的相关人员要予以协助 |

5 管理规定

本公司数据资产的风险评估分为如下9个过程，每个过程内的详细活动如下表。

|  |  |  |
| --- | --- | --- |
| 序号 | 过 程 活 动 | 过 程 内 容 |
| 1 | 数据资产识别、分 组与登记 | 根据资产分类表，对评估范围内的资产进行识别、分组和登记；与数据 资产相关的信息资产分组一般分为系统、信息、人员、环境设施、外购服 务和无形资产6种 |
| 2 | 资产和资产组  赋值 | 从保密性、完整性和可用性三个方面对数据及其相关资产进行赋值。根 据组内资产价值的最高值确定整个资产组的价值 |
| 3 | 风险识别与分析 | 识别并登记与资产组相关的主要威胁、弱点和现有控制措施 |
| 4 | 风险评价 | 根据预先定义的赋值标准，对风险发生的可能性和风险的影响进行赋值 并通过资产组价值、风险发生可能性和风险影响计算出风险值- |
| 5 | 确定风险可接受 标准 | 根据风险计算结果，确定风险级别，确定企业可接受的风险值。 一般来 说，风险可接受标准为中、低风险的分界线 |
| 6 | 风险控制措施的 选择和实施 | 对于超过风险可接受标准的风险，企业要选择和实施管理或技术控制措 施，以降低风险发生的可能性或影响程度。选择风险控制措施要考虑预估 的残余风险值 |
| 7 | 控制措施有效性 测量 | 制订或利用一定的测量方法，对采取的全部或部分控制措施进行测量 以判断控制措施的有效性，对无效或效果不明显的措施进行整改 |
| 8 | 风险评估更新 | 根据风险评估周期，周期性进行风险评估与处置 |
| 9 | 残余风险处置 | 风险评估更新后，对于仍超过可接受标准的风险可以采取新的控制措施 也可以接受风险 |

**第9章■安全与隐私** 

续表

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 5.1 数据及其相关的信息资产识别、分组与登记  5.1.1 信息资产识别  (1)资产是公司直接赋予价值因而需要保护的有用资源，以多种形式存在；  (2)信息资产分为：硬件类、软件类、信息类、人员类、服务类、无形资产等六大类；  (3)各部门的信息资产识别由信息安全管理组统一组织进行，主要工作由信息安全管理员完成，部门或机构的 相关人员要予以协助。  5.1.2 信息资产分组  (1)本公司的资产组主要有六种类型：软件类资产组、信息类资产组、人员类资产组、硬件类资产组、服务类 资产组、无形资产组；  (2)对于识别出来的信息资产，部门数据安全管理员要根据资产的不同类型和价值进行必要的组合，形成资产组。  5.1.3 信息资产登记  识别出来的信息资产由数据安全管理员详细登记在《云端数据资产风险登记表》中。  5.1.4 资产和资产组赋值  (1)资产赋值  信息资产的赋值通过信息资产的机密性、完整性和可用性赋值确定，信息资产的机密性、完整性、可用性赋值 如下表。   |  |  |  |  | | --- | --- | --- | --- | | 等 级 | C-机密性 | 1一完整性 | A一可用性 | | 5(很高) | 包含组织最重要的秘密， 关系未来发展的前途命运 对组织根本利益有着决定 性影响，如果泄漏会造成灾  难性的损害 | 完整性价值非常关键，未经 授权的修改或破坏会对组织造 成重大的或无法接受的影响， 对业务冲击重大，并可能造成 严重的业务中断，难以弥补 | 可用性价值非常高，合法 使用者对信息及信息系统的 可用度在正常工作时间达到 99.6%以上 | | 4 ( 高 ) | 包含组织的重要秘密，其 泄露会使组织的安全和利 益遭受严重损害 | 完整性价值较高，未经授权 的修改或破坏会对组织造成重 大影响，对业务冲击严重，比 较难以弥补 | 可用性价值较高，合法使 用者对信息及信息系统的可 用度在正常工作时间达到  90%以上 | | 3 ( 中 ) | 包含组织的一般性秘密， 其泄露会使组织的安全和 利益受到损害 | 完整性价值中等，未经授权的 修改或破坏会对组织造成影响， 对业务冲击明显，但可以弥补 | 可用性价值中等，合法使用 者对信息及信息系统的可用度 在正常工作时间达到70%以上 | | 2 ( 低 ) | 包含仅能在组织内部或  在组织某一部门内部公开 的信息，向外扩散有可能对 组织的利益造成损害 | 完整性价值较低，未经授权 的修改或破坏会对组织造成轻 微影响，可以忍受，对业务冲 击轻微，容易弥补 | 可用性价值较低，合法使 用者对信息及信息系统的可 用度在正常工作时间达到  25%以上 | | 1(很低) | 包含可对社会公开的信  息，公用的信息处理设备和 系统资源等 | 完整性价值非常低，未经授 权的修改或破坏对组织造成的 影响可以忽略，对业务冲击可 以忽略 | 可用性价值可以忽略，合 法使用者对信息及信息系统 的可用度在正常工作时间低 于 2 5 % | |



**■云端数据治理**

续表

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| (2)采用以下公式进行资产赋值：  资产赋值=Round({Log2[(2C+2I+2A)/3]},1); (其中： “C” 为机密性赋值；“1”为完整性赋值： “A” 为可  用性赋值。)  注：资产组的赋值取组内资产价值的最大值。  5.2 风险识别与分析  5.2.1 威胁识别与分析  应根据资产组内的每一项资产，以及每一项资产所处的环境条件、以前曾发生的安全事件等情况来进行威胁识 别。 一项资产可能面临着多个威胁，同样一个威胁可能对不同的资产造成影响。可能的威胁如下表。   |  |  | | --- | --- | | 管理不到位 | 安全管理无法落实，不到位，造成安全管理不规范，或者管理混乱，从而破坏信 息系统的正常有序运行 | | 恶意代码和病毒 | 具有自我复制、自我传播能力，对信息系统构成破坏的程序代码 | | 越权或滥用 | 通过采用一些措施，超越自己的权限访问了本来无权访问的资源；滥用自己的职  权，做出破坏信息系统的行为 | | 黑客攻击 | 利用黑客工具和技术，例如侦察、密码猜测攻击、缓冲区溢出攻击、安装后门、 嗅探、伪造和欺骗、拒绝服务攻击等手段对信息系统进行攻击和入侵 | | 物理攻击 | 物理接触、物理破坏、盗窃 | | 泄密 | 机密泄漏，机密信息泄漏给他人 | | 篡改 | 非法修改信息，破坏信息的完整性 | | 抵赖 | 不承认收到的信息和进行的操作和交易 |   5.2.2 脆弱性识别与分析   |  |  |  | | --- | --- | --- | | 类 型 | 脆弱性分类 | 脆弱性示例 | | 技术脆  弱性 | 物理环境 | 从机房场地、机房防火、机房供配电、机房防静电、机房接地与 防雷、电磁防护、通信线路的保护、机房区域防护、机房设备管理 等方面进行识别 | | 服务器(含操作系统) | 从物理保护、用户账户、口令策略、资源共享、事件审计、访问 控制、新系统配置(初始化)、注册表加固、网络安全、系统管理等 方面进行识别 | | 网络结构 | 从网络结构设计、边界保护、外部访问控制策略、内部访问控制 策略、网络设备安全配置等方面进行识别 | | 数据库 | 从补丁安装、鉴别机制、口令机制、访问控制、网络和服务设置、 备份恢复机制、审计机制等方面进行识别 | | 应用系统 | 审计机制、审计存储、访问控制策略、数据完整性、通信、鉴别 机制、密码保护等方面进行识别 | |

**第9章■安全与隐私**

续表

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 续表   |  |  |  | | --- | --- | --- | | 类 型 | 脆弱性分类 | 脆弱性示例 | | 管理脆  弱性 | 技术管理 | 物理和环境安全、通信与操作管理、访问控制、系统开发与维护、 业务连续性管理 | | 组织管理 | 安全策略、组织安全、资产分类与控制、人员安全、符合性 |   脆弱性评估将针对资产组内所有资产，找出该资产组可能被威胁利用的脆弱性，获得脆弱性可以采用问卷调查、 访谈、工具扫描、手动检查、文档审查等方法。  5.2.3 现有控制措施识别与分析  在识别威胁和脆弱性的同时，数据安全管理员应对已采取的安全措施进行识别，对现有控制措施的有效性进行 确认。在对威胁和脆弱性赋值时，需要考虑现有控制措施的有效性。  5.2.4 风险评价  (1)本公司数据资产风险值通过风险各要素赋值相乘法来确定。  风险值计算公式： R=Round(a\*i\*p,1) (其中， R 表示安全风险值； a 表示资产组价值； i 表示风险影响； p 表 示风险发生可能性。);  (2)风险影响的赋值标准   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | 风险影响 | | | | | | | 赋值 | 等级 | 风险影响的不同维度 | | | | | 区 域 | 相 关 方 | 系 统 | 业务连续性 | | 5 | 很高 | 整个行业或  社会 | 全部或大部分客 户、监管机构、甚 至社会公众 | 核心业务系统不能使 用 | 公司大部分或全部 核心业务受到严重影 响 | | 4 | 高 | 公司办公区 域，或整个机房 区域 | 整个公司，或部 分客户 | 核心业务系统访问速 度缓慢，或部分功能不 能使用 | 公司部分核心业务 或日常办公活动受影 响 | | 3 | 中 | 多个部门(或 分支机构)的办 公区域，或部分 机房区域 | 多个部门，或个 别客户 | 核心业务系统偶尔不 能访问或访问速度缓  慢，或部分其他应用系 统不能正常工作 | 公司个别业务受影 响，或部分日常办公 活动中断 | | 2 | 低 | 某个部门或 分支机构的办 公区域 | 本部门或部门内 部人员 | 个别其他应用系统不 能正常工作 | 公司业务不受影  响，只是少部分日常 办公活动受影响 | | 1 | 很低 | 个人办公区  域 | 个人 | 个人终端 | 基本不影响公司业 务和日常办公活动 | |

**■云端数据治理■**

5.2.6 风险接受准则

(1)对于信息资产评估的结果，本公司原则上以风险值小于16 分(不包括16分)的风险作为可接受风险，大 于等于16分为不可接受风险，要采取控制措施进行控制；

(2)对于不可接受的风险，由于经济或技术原因，经信息安全管理委员会批准后，可以暂时接受风险，待经济 或技术具有可行性时再采取适当措施降低风险。

5.2.7 风险控制措施的选择和实施

(1)对于不可接受的风险，有四种风险处置方式：降低风险、转移风险、消除风险、接受风险；

(2)最常见的风险处置方式是降低风险，降低风险可以选择一项或数项控制措施，也可以根据以往的经验给出 合适的控制措施。

续表

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| (3)风险发生可能性的赋值标准   |  |  |  |  | | --- | --- | --- | --- | | 风险可能性 | | | | | 赋值 | 等级 | 风险发生可能性 | | | 时 间 频 率 | 发 生 机 率 | | 5 | 很高 | 出现的频率很高(或≥1次/日);或在大多数情况下几乎 不可避免；或可以证实经常发生 | 不可避免(≥99%) | | 4 | 高 | 出现的频率较高(或≥1次/周);或在大多数情况下很有 可能会发生；或可以证实多次发生过 | 非常有可能(90%~ 99%) | | 3 | 中 | 出现的频率中等(或≥1次/月);或在某种情况下可能会 发生；或被证实曾经发生过 | 可能(10%～90%) | | 2 | 低 | 出现的频率较小(或≥1次/年);或一般不太可能发生； 或在某种情况下可能会发生 | 可能性很小(1～10%) | | 1 | 很低 | 威胁几乎不可能发生，仅可能在非常罕见和例外的情况 下发生，或没有被证实发生过 | 不可能(≤1%) |   注：风险的影响或发生可能性根据赋值标准，可能适用两个或以上维度，这种情况下，赋值取这两个或多个维 度赋值的最大值。  5.2.5 确定风险可接受标准  (1)风险等级采用分值计算表示：分值越大，风险越高； (2)风险等级：   |  |  |  | | --- | --- | --- | | 赋 值 | 级 别 | 描 述 | | ≥36,<125 | 高 | 如果发生将使资产遭受严重破坏，组织利益受到严重损失 | | ≥16,<36 | 中 | 发生后将使资产受到较重破坏，组织利益受到损失 | | <16 | 低 | 发生后将使资产受到轻微的破坏，组织利益受到轻微损失 | |

**第9章■安全与隐私**

**续表**

|  |
| --- |
| 5.2.8 控制措施有效性测量  在风险控制措施全部或部分实施完成后，信息安全组可以组织人员对风险控制措施的有效性进行测量；可以测量全 部的控制措施，也可以测量部分的控制措施，出于时间和经济成本的考虑，建议选取主要的控制措施进行测量。  5.2.9风险评估更新  风险评估需要至少每年进行一次，重新进行风险评估时，各部门需要对本公司信息资产登记表和风险管理表进  行更新。在发生以下情况时，可以增加风险评估次数。  (1)当信息系统发生重大变更时；  (2)当公司业务或信息资产发生重大变化时；  (3)发生严重信息安全事件时；  (4)企业认为有必要时。  5.2.10残余风险处置  (1)对于低于风险可接受标准的残余风险，本公司接受这样的风险，也可以继续采取措施，进一步减少  该风险 ；  (2)对于高于本公司风险可接受标准的残余风险，由于费用或技术等原因不能增加新的控制措施，经云计算安  全管理委员会批准可以决定暂时接受此风险，并由云计算安全管理组组织人员积极跟踪相关技术和产品的发展情  况，以尽快采取新的技术或方法，降低风险；对于可以继续实施控制措施的，由各部门继续实施相应的信息安全 控制措施。 |

**9.2.2** **数据安全**

**1.** **简介**

1)目的

保护云端数据的安全，确保合适的人以正确的方式使用和更新数据，并限制所 有不恰当的使用和数据更新。

2)介绍

由于云计算服务模式下，虚拟化IT 资源是多租户共享的，不同的云计算用户可 以在任何地点、任何时间访问云服务。虽然企业会拥有一定的控制权，但不能拥有 完全的控制权，云端数据面临保密性和完整性风险。

数据安全的目标是保护云端数据的安全，确保合适的人以正确的方式使用和 数据更新，并限制所有不恰当的使用和更新数据。组织应该积极应用云计算技术 优势的基础上，清晰自身数据面临的安全威胁，计划、制订、执行相关安全策略 和规程，确保数据和信息资产在使用过程中有恰当的认证、授权、访问和审计等 措施。

下面是 CSA 给出的针对云端数据安全管理的最佳实践，组织可以借鉴：

(1)通过加密手段保护数据；

 **■云端数据治理■**

(2)通过预定义的安全策略对文件的访问进行授权；

(3)实现细粒度的访问控制；

(4)尽量用加密系统实现安全策略；

(5)考虑部分使用动态加密方案；

(6)在终端部署防病毒软件和恶意软件防护系统；

(7)提供及时的访问审计信息；

(8)提供基础设施的认证机制；

(8)严格控制特权用户(系统管理员)的接入；

(9)采用数据隔离、数据切分等方式保证数据的安全性。

3)相关过程域

●参考风险管理过程域了解风险评估的情况；

●参考数据生命周期管理过程域了解云端数据生命周期管理内的管理策略；

●参考数据质量评估过程域了解质量目标与质量控制活动保持相符的相关信息；

●参考合规管理过程域了解组织合规管理策略。

**2.GQM 表**

|  |  |  |  |
| --- | --- | --- | --- |
| 目 标 | 问 题 | 度 量 项 | 度量项说明 |
| G1:制订一整 套与安全需求相 一致的数据安全 策略 | Q1:组织如何确定数据  需求，以制订经批准的数据 安全策略 | M1:数据安全策略 | 数据安全策略应以数据中心为中 心，定义应用程序、数据库角色、 用户组和密码标准等 |
| M2:数据密级分类制 度 | 组织应制订组织数据密级管理办 法，并对云端数据划分密级。典型 的密级分类模式包括：国家秘密、 商业秘密、内部使用、受限使用、 公众级等 |
| G2:实施涉及 利益相关方的可 持续的标准制订 和维护过程 | Q2:如何对数据安全策 略进行批准、维护 | M3:利益相关人参与 数据安全控制措施制订 的程度 | 组织实施的数据安全活动应充分 考虑员工和客户等利益相关人的期 望，并邀请利益相关人参与数据安 全措施和过程的制订 |
| M4:数据安全控制措 施 | 组织应实施适当的控制，以满足 数据安全策略 |
| G3:建立标准 遵守的有效的治 理和审核过程 | Q3:组织如何确定数据 安全管理的效果 | M5:用户身份认证和 访问行为监控机制 | 监控认证和访问行为有助于检 测到异常或可疑的活动，方便进 一步调查和解决问题。组织应建 立机制，监控用户身份认证和访  问行为 |

**第9章■安全与隐私**

**续表**

|  |  |  |  |
| --- | --- | --- | --- |
| 目 标 | 问 题 | 度 量 项 | 度量项说明 |
| G3:建立标准 遵守的有效的治 理和审核过程 | Q3:组织如何确定数据 安全管理的效果 | M6:数据安全审计机 制 | 组织应建立高效、持续的数据安 全审计机制，对数据安全策略、数 据安全标准、变更请求、访问监控 日志以及其他电子和书面记录进行 审计，以确保数据资产在使用过程 中有恰当的认证、授权、访问和审 计等措施 |
| M7:数据安全意识培 训 | 企业应采用培训、绩效等方式鼓 励员工积极参与到数据安全活动  中，所有利益相关者均能在日常运 营过程中自然遵守这些制订 |

**3.** **成熟度等级介绍**

**Level 1: 初始级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 1.1 | 识别具体项目的数据安全需求 | ●项目数据安全需求列表 |

【成熟度等级说明】

对业务进行透彻地了解，在系统开发项目的分析阶段识别具体的应用安全要求。 但缺乏负责数据安全工作策略制订的团队，也未针对数据安全工作的开展制订相应 的策略、制度和规范，仅在部分项目中考虑到了数据安全的风险。

数据安全管理的最终目标是保护信息资产，使其符合隐私及法律法规要求，并 与业务要求相一致，这些要求主要依据以下四条。

●利益相关者的隐私及保密需要，利益相关者包括客户、最终用户、合作伙伴等；

●法律法规，法律法规保护利益相关者的安全利益；

●特定业务关注，每个组织、部门都有自己需要保护的特定数据，如知识产权；

●合法访问需求。

**Level 2: 管理级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.1 | 基于组织数据安全需求定义数据安全策 略，并获得评审和批准 | ●数据安全管理策略、程序  ●数据安全策略评审记录 |

【成熟度等级说明】

定义数据安全策略应基于组织数据安全需求，数据安全策略应以数据中心为中

 **■云端数据治理**

心，定义应用程序、数据库角色、用户组和密码标准等。定义的数据安全策略应当 得到数据相关者的评审和批准。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.2 | 基于数据安全维度制订数据分类分级策略 | ●数据密级分类规定  ●数据密级分类表 |

【成熟度等级说明】

制订组织数据密级管理办法，并对云端数据划分密级。典型的密级分类模式包 括：国家秘密、商业秘密、内部使用、受限使用、公众级等。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.3 | 定义了在确定的数据安全职能下的岗位和 职责 | ●数据安全职责  ●数据安全管理人员能力培训记录 |

【成熟度等级说明】

通过制订工作规范和流程，明确组织数据管理岗位和职责，既包括本地数据， 也包括云端数据。针对不同的岗位特点，安排合适的人员负责数据安全的日常管理 工作。组织也要针对数据管理人员能力制订提升计划，比如培训和外部学习等方式， 保证负责该项过程的人员具备对组织内数据安全关键工作的判断能力。

**Level 3: 定义级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.1 | 客户和员工等利益相关人参与数据安全控 制措施的制订 | ●数据安全控制措施评审记录  ●数据安全利益相关人联络表 |

【成熟度等级说明】

实施数据安全活动却不考虑客户和员工的期望，会导致他们的不满，并增加组 织的风险。

组织实施的数据安全活动应充分考虑员工和客户等利益相关人的期望，并邀请 利益相关人参与数据安全措施和过程的制订。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.2 | 设计组织自己的数据安全控制措施 | ●数据安全控制程序 |

【成熟度等级说明】

组织实施适当的控制，以满足数据安全策略。例如， 一个控制客体可以按月读 取、评审 DBA 和用户权限。为满足这一目标，组织实施一个控制流程，与负责跟踪 所有用户权限请求的变更管理系统相对照，来验证分配权限。此外，组织以审批流 程或签署文件的形式，记录和归档每一个请求。

**第9章■安全与隐私**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.3 | 定义控制措施防止非授权用户对数据的访 问 | ●数据访问控制程序  ●数据管理系统权限申请表 |

【成熟度等级说明】

通过数据隔离、数据加密、数据切分、数据视图等技术方式防止非授权用户对 数据的访问，保证数据的安全性。例如，在数据分类分级定义的基础上，明确提出 对各类各级别数据的加密要求，针对数据的加密要求应包含对数据加密算法的要求 和数据加密密钥的管理要求。建立组织内部统一的账号管理平台(如利用 SSO、联 合认证技术),实现对组织内部账号的统一管理策略。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.4 | 制订云端远程数据传输策略 | ●数据传输通道安全策略  ●数据传输加密与解密策略 |

【成熟度等级说明】

数据远程传输一般发生在以下三个场景。

第一，组织本地数据与云端数据的交互；

第二，组织与外界开展的数据共享行为；

第三，组织内部不同地域间的数据转移等多个场景。

无论哪种场景，组织都应当根据法律法规的要求以及业务开展的成本和需求， 明确组织远程数据传输的策略，例如对传输通道加密的场景采用统一的部署方案(如 HTTPS 加密、VPN 加密),以保证加密效果的一致性。

**Level** **4:** **量化级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 4.1 | 建立机制，监控用户身份认证和访问行为 | ●用户访问监控管理程序  ●用户异常登陆记录  ●用户认证和访问定期审计记录 |

【成熟度等级说明】

监控认证和访问行为有助于检测到异常或可疑的活动，方便进一步调查和解 决问题。 一般来说，自动化系统配合一定的人工检查是最佳监控方式。在身份认 证层面实现唯一的用户身份标识，能够通过账号追溯到组织内外部唯一的人员或 组织。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 4.2 | 建立数据安全审计机制 | ●数据安全审计策略 |

 **■云端数据治理**

【成熟度等级说明】

建立高效、持续的数据安全审计机制，对数据安全策略、数据安全标准、变更 请求、访问监控日志以及其他电子和书面记录进行审计，以确保数据资产在使用过 程中有恰当的认证、授权、访问和审计等措施。组织内部接入安全监控和审计管理 平台的信息系统的数量，同时对安全监控和审计系统发现的风险进行量化以持续改 进风险发现和跟进的规则，实现对组织内部的安全风险的量化管理。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 4.3 | 量化评估数据安全职能的运行效果，动态 调整数据安全职能 | ●数据安全职能效果统计报告  ●数据安全职能调整方案 |

【成熟度等级说明】

数据安全职能的运行效果可以通过量化指标进行定期统计和衡量，并根据量化 的结果定期调整数据安全职能的设置。数据安全的职能框架包括但不限于：

●数据安全规范及标准

●数据安全技术及产品

● 数据安全监控及审计

● 数据安全宣传与促进

**Level** **5:** **优化级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 5.1 | 数据安全理念融入到企业文化中 | ●数据安全培训演示材料  ●数据安全员工激励制度 |

【成熟度等级说明】

真正的数据安全管理是安全理念深深地融入到企业文化中。企业采用培训、绩 效等方式鼓励员工积极参与到数据安全活动中，所有利益相关者均能在日常运营过 程中自然遵守这些制订。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 5.2 | 组织对潜在采用的创新的数据安全技术和 方法进行研究，并为采用的这些技术和方法 制订适当的新标准 | ●新兴技术的研究 |

【成熟度等级说明】

组织安全管理员需要在业务工作中保持业务知识，加入数据安全管理组织， 接受持续的教育，积极学习、研究新的数据安全管理技术，为组织采用新技术做 好准备。

**第9章■安全与隐私**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 5.3 | 组织通过出版物、会议和白皮书分享所获 得的最佳数据安全管理实践和经验教训 | ●与数据安全相关的演示材料和其他发表的作品 |

【成熟度等级说明】

组织IT 安全管理专业人士应订阅专业杂志和网上各种论坛、专业讨论群，分享 或阅读数据安全管理相关主题。

**4.** **实施案例**

云端数据安全管理要围绕数据生命周期变化来实施，在其数据的采集、传输、 存储和使用的各个环节采取安全措施，制订覆盖数据存储、应用等环节的数据安全 控制要求，提高安全防护能力。

以下为某电信运营商数据安全管理规范和某商业银行数据安全管理规范，仅供 参考。

|  |
| --- |
| 数据安全管理规范  1 概 述  数据信息安全，顾名思义就是要保护数据信息免受威胁的影响，从而确保业务平台的连续性，缩减业务平台有 可能面临的风险，为整个业务平台部门的长期正常运行提供强有力的保障。  为加强数据信息的安全管理，保证数据信息的可用性、完整性、机密性，特制订本规范。  2数据信息安全管理制度  2.1 数据信息安全存储要求  数据信息存储介质包括：纸质文档、语音或其录音、输出报告、硬盘、磁带、光存储介质。  存储介质管理须符合以下规定。  (1)包含重要、敏感或关键数据信息的移动式存储介质须专人值守。  (2)删除可重复使用存储介质上的机密及绝密数据时，为了避免在可移动介质上遗留信息，应该对介质进行消 磁或彻底的格式化，或者使用专用的工具在存储区域填入无用的信息进行覆盖。  (3)任何存储媒介入库或出库须经授权，并保留相应记录，方便审计跟踪。  2.2 数据信息传输安全要求  (1)在对数据信息进行传输时，应该在风险评估的基础上采用合理的加密技术，选择和应用加密技术时，应符 合以下规范。  a.必须符合国家有关加密技术的法律法规；  b.根据风险评估确定保护级别，并以此确定加密算法的类型、属性，以及所用密钥的长度；  c.听取专家的建议，确定合适的保护级别，选择能够提供所需保护的合适的工具。  (2)机密和绝密信息在存储和传输时必须加密，加密方式可以分为：对称加密和不对称加密。  (3)机密和绝密数据的传输过程中必须使用数字签名以确保信息的不可否认性，使用数字签名时应符合以下 规范。  a.充分保护私钥的机密性，防止窃取者伪造密钥持有人的签名；  b.采取保护公钥完整性的安全措施，例如使用公钥证书； |

 **■云端数据治理**

续表

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| c. 确定签名算法的类型、属性以及所用密钥长度；  d. 用于数字签名的密钥应不同于用来加密内容的密钥。  2.3 数据信息安全等级变更要求  数据信息安全等级经常需要变更。 一般地，数据信息安全等级变更需要由数据资产的所有者进行，改变相应的 分类并告知信息安全负责人进行备案.。对于数据信息的安全等级，应每年进行评审，只要实际情况允许，就进行  数据信息安全等级递减，这样可以降低数据防护的成本，并增加数据访问的方便性。  2.4 数据信息安全管理职责  数据信息涉及各类人员的职责如下。  (1)拥有者：拥有数据的所有权；拥有对数据的处置权利；对数据进行分类与分级；指定数据资产的管理者/维  护人；  (2)管理者：被授权管理相关数据资产；负责数据的日常维护和管理；  (3)访问者：在授权的范围内访问所需数据；确保访问对象的机密性、完整性、可用性等。  3 数据信息重要性评估  3.1 数据信息分级原则  (1)分级合理性  数据信息和处理数据信息分级的系统应当仔细考虑分级范畴的数量以及使用这种分级所带来的好处。过于复杂  的分级规划可能很累赘，而且使用和执行起来也不够经济实用。  (2)分级周期性  数据信息的分级具有一定的保密期限。对于任何数据信息的分级都不一定自始至终固定不变，可按照一些预定  的策略发生改变。如果把安全保护的分级划定得过高就会导致不必要的业务开支。  3.2 数据信息分级  数据信息应按照价值、法律要求及对组织的敏感程度和关键程度进行分级，分级等级如下：   |  |  |  | | --- | --- | --- | | 等级 | 标识 | 数据信息价值定义 | | 5 | 很高 | 重要程度很高，其安全属性被破坏后可能导致系统受到非常严重的影响 | | 4 | 高 | 重要程度较高，其安全属性被破坏后可能导致系统受到比较严重的影响 | | 3 | 中 | 重要程度较高，其安全属性被破坏后可能导致系统受到中等程度的影响 | | 2 | 低 | 重要程度较低，其安全属性被破坏后可能导致系统受到较低程度的影响 | | 1 | 很低 | 重要程度很低，其安全属性被破坏后可能导致系统受到很低程度的影响 甚至忽略不计 |   4 数据信息完整性安全规范  数据信息完整性应符合以下规范。  (1)确保所采取的数据信息管理和技术措施以及覆盖范围的完整性；  (2)应能够检测到网络设备操作系统、主机操作系统、数据库管理系统和应用系统的系统管理数据、鉴别信息  和重要业务数据在传输过程中完整性是否受到破坏，并在检测到完整性错误时采取必要的恢复措施；  (3)应能够检测到网络设备操作系统、主机操作系统、数据库管理系统和应用系统的系统管理数据、鉴别信息  和重要业务数据在存储过程中完整性是否受到破坏，并在检测到完整性错误时采取必要的恢复措施； |

**第9章■安全与隐私**

续表

|  |
| --- |
| (4)具备完整的对用户访问、处理、删除数据信息的操作进行记录的能力，以备审计；  (5)在传输数据信息时，在经过不安全网络时(例如INTERNET网),需要对传输的数据信息提供完整性校验；  (6)应具备完善的权限管理策略，支持权限最小化原则、合理授权。  5 数据信息保密性安全规范  数据信息保密性安全规范用于保障业务平台重要业务数据信息的安全传递与处理应用，确保数据信息能够被安 全、方便、透明地使用。为此，业务平台应采用加密等安全措施开展数据信息保密性工作：  (1)应采用加密性措施以实现重要业务数据信息传输的保密；  (2)应采用加密措施实现重要业务数据信息存储的保密  加密安全措施主要分为密码安全及密钥安全。  5.1密码安全  密码的使用应该遵循以下原则。  (1)不能将密码写下来，不能通过电子邮件传输；  (2)不能使用缺省设置的密码：  (3)不能将密码告诉别人；  (4)如果系统的密码泄漏了，必须立即更改；  (5)密码要以加密形式保存，加密算法强度要高，加密算法要不可逆；  (6)系统应该强制指定密码的策略，包括密码的最短有效期、最长有效期、最短长度、复杂性等；  (7)如果需要特殊用户的口令(比如说UNIX下的Oracle),要禁止通过该用户进行交互式登录；  (8)在要求较高的情况下可以使用强度更高的认证机制，例如：双因素认证；  (9)(要定时运行密码检查器检查口令强度，对于保存机密和绝密信息的系统应该每周检查一次口令强度；其他 系统应该每月检查一次口令强度。  5.2密钥安全  密钥管理对于有效使用密码技术至关重要。密钥的丢失和泄露可能会损害数据信息的保密性、重要性和完整性。  因此，应采取加密技术等措施来有效保护密钥，以免密钥被非法修改和破坏；还应对生成、存储和归档保存密钥 的设备采取物理保护。此外，必须使用经过业务平台部门批准的加密机制进行密钥分发，并记录密钥的分发过程，  以便审计跟踪，统一对密钥、证书进行管理。  密钥的管理应该基于以下流程。  (1)密钥产生：为不同的密码系统和不同的应用生成密钥；  (2)密钥证书：生成并获取密钥证书；  (3)密钥分发：向目标用户分发密钥，包括在收到密钥时如何将之激活；  (4)密钥存储：为当前或近期使用的密钥或备份密钥提供安全存储，包括授权用户如何访问密钥 (5)密钥变更：包括密钥变更时机及变更规则，处置被泄露的密钥；  (6)密钥撤销：包括如何收回或者激活密钥，如在密钥已被泄露或者相关运维操作员离开业务平台部门时，应 当归档密钥；  (7)密钥恢复：作为业务平台连续性管理的一部分，对丢失或破坏的密钥进行恢复；  (7)密钥归档：归档密钥，以用于归档或备份的数据信息；  (9)密钥销毁：密钥销毁将删除该密钥管理下数据信息客体的所有记录， 一旦销毁将无法恢复。因此，在密钥 销毁前，应确认由此密钥保护的数据信息不再需要。 |

**云端数据治理**

续表

|  |
| --- |
| 6数据信息备份与恢复  6.1 数据信息备份要求  6.1.1 备份要求  (1)数据信息备份应采用性能可靠、不宜损坏的介质，如磁带、光盘等。备份数据信息的物理介质应注明数据 信息的来源、备份日期、恢复步骤等信息，并置于安全环境保管。  (2)一般情况下对服务器和网络安全设备的配置数据信息每月进行一次备份，当进行配置修改、系统版本升级、 补丁安装等操作前也要进行备份；网络设备配置文件在进行版本升级前和配置修改后需要进行备份。  (3)运维操作员应确保对核心业务数据每日进行增量备份，每周做一次包括数据信息的全备份。业务系统将进 行重大系统变更时，应对核心业务数据进行数据信息的全备份。  6.1.2 备份执行与记录  备份执行过程应有详细的规划和记录，包括备份主体、备份时间、备份策略、备份路径、记录介质(类型)等。  6.2备份恢复管理  (1)运维操作员应根据不同业务系统实际拟定需要测试的备份数据信息以及测试的周期。  (2)对于因设备故障、操作失误等造成的一般故障，需要恢复部分设备上的备份数据信息，遵循异常事件处理 流程，由运维操作员负责恢复。  (3)应尽可能地定期检查和测试备份介质和备份信息，保持其可用性和完整性，并确保在规定的时间内恢复 系统。  (4)应确定重要业务信息的保存期以及其他需要永久保存的归档拷贝信息的保存期。  (5)恢复程序应定期接受检查及测试，以确保在恢复操作程序所预定的时间内完成。  恢复策略应该根据数据信息的重要程度和引入新数据信息的频率设定备份的频率(如每日或每周、增量或整体)。 商业银行数据安全管理规范  1 总 则  1.1 本管理规范定义了××××商业银行数据安全管理方面的相关要求。  1.2本规范适用范围为涉及维护××××商业银行各类数据的安全管理员、系统管理员和业务管理员。  1.3××××商业银行的所有员工必须遵照数据安全管理规范，科技科加强管理，防止数据管理存在安全问题。 如由于未遵循以上规范导致出现数据安全问题，相关部门和人员负有责任。  1.4本标准由××××商业银行科技科发布，对标准具有解释、增加和修改的权力。本规范自下发之日起正式 执行。  2系统数据安全  2.1 ××××商业银行系统数据指××××商业银行中设备和系统的各种配置数据，如权限设置、存储分配、 网络地址、硬件配置及其他系统配置参数等。  2.2应制订系统数据管理制度，对系统数据实行严格的安全与保密管理。系统数据管理制度应遵循以下原则。 (1)重要的系统数据应实施严格的安全保密管理，参见本规范“数据保密”小节。  (2)对系统数据的操作应经过严格的身份鉴别与权限控制，防止非授权操作。  (3)应采取有效措施防止系统数据的非法生成、变更、泄漏、丢失与破坏。  (4)系统数据应进行核对审查，防止使用过程中产生误操作或被非法篡改。  (5)系统数据的备份和恢复应符合本规范“数据备份和恢复”小节的要求。  (6)对系统数据应实行专人管理。 |

**第9章■安全与隐私**

**续表**

|  |
| --- |
| 3业务数据安全  3.1 业务数据主要包括××××商业银行上各业务系统的用户数据、业务数据、统计数据及其他相关数据。 3.2各个业务管理部门应遵循以下原则制订具体的业务数据安全管理规定。  (1)重要的业务数据应实施严格的安全保密管理，参见本规范“数据保密”小节。  (2)业务数据在系统中的保存时间应有最低限制。  (3)业务数据应进行核对审查，防止使用过程中产生误操作或被非法篡改。  (4)业务数据应及时、完整、真实、准确地转储到不可更改的介质上，并规定保存期限。  (5)业务数据的备份和恢复应符合本规范“4 数据备份和恢复”小节的要求。  (6)对业务数据应实行专人管理。  4数据备份与恢复  4.1 ××××商业银行相关数据的备份与恢复应遵循原则。  (1)数据备份媒介应采用性能可靠、不宜损坏的介质，如磁带、光盘等。  (2)××××商业银行中的网络设备和主机系统的配置信息在每次更新后应及时备份，更新前的备份按需要保  存一定时间。  (3)××××商业银行中的网络设备和主机系统的数据库信息应进行备份。对重要数据应做到定期全备份和增  量备份，并配备可靠的备份设备。  (4)备份数据应进行检查，以保证其有效性和可用性。  (5)备份数据的物理介质应注明数据的来源、备份日期、恢复步骤等信息，并置于安全环境保管。 (6)备份数据的恢复须经主管人员签字认可后，方可进行。  (7)过期的备份数据应经主管人员认可后，方可销毁。  5 数据保密  5.1 ××××商业银行数据保密的范围包括XXXX商业银行网络组织、系统软件、业务数据及用户数据等重要 信息，另外还包括XXXX商业银行为生产通信、科研通信、办公通信、机要通信、党政专用通信提供的通信手段 及保障措施。  5.2 ××××商业银行维护和管理人员都应熟悉并严格遵守和执行本办法的以下安全保密规定。  (1)保密数据应仅用于明确规定的目的，未经批准不得它用。  (2)保密数据的多个副本应编号保存，防止丢失；保密数据以电子形式存在时，应尽量避免以明码形式存储和  传输。  (3)机密资料不得抄录影印或对外泄露。  (4)无正当理由和有关批准手续，不得查阅用户资料。经正式批件查阅数据时应登记，并由查阅人签字。  (5)各级有关部门应根据数据的保密规定和用途，确定各类保密数据使用人员的存取权限、存取方式和审批手续。 |

**9.2.3** **隐私保护**

**1.** **简介**

1) 目的

确保云端隐私数据不被滥用及未经授权使用。

**■云端数据治理**

2)介绍

隐私不等同于安全，尽管在概念上两者有重叠之处。隐私是指个人或群体控制 有关自身特定信息的能力。隐私涉及一个人决定什么样的信息愿意让其他人知道、 哪些人可以知道以及决定哪些人什么时候可以知道的权利。隐私权的执行须借助 “安全”。

随着当今社会人们对科学技术的依赖越来越强，个人隐私也受到了更多的威胁。 根据《2015年度数据泄露调查报告》称：2015年，全球61个国家出现79790起数 据泄露事件，其中2122起已经得到确认。

有数据统计，在个人信息保护方面，网民被泄露的个人信息涵盖范围非常广泛， 其中78.2%的网民个人身份信息被泄露过，包括网民的姓名、学历、家庭住址、身 份证号及工作单位等；63.4%的网民个人网上活动信息被泄露过，包括通话记录、 网购记录、网站浏览痕迹、IP 地址、软件使用痕迹及地理位置等。在个人信息泄 露带来的不良影响方面，82.3%的网民亲身感受到了个人信息泄露给日常生活造成 的影响。

保护隐私的方式包括政策法规和技术手段。其中，政策法规包括通用法规和行 业规章。通用法规影响所有行业，包括政府部门，行业规章则是针对特定行业制订 要求；技术手段主要包括基于数据失真的技术(掺沙子)、基于数据加密的技术和数 据脱敏的技术。

3)相关过程域

●参考风险管理过程域了解组织风险评估的情况；

●参考数据生命周期管理过程域了解云端数据生命周期管理内的管理策略；

●参考数据安全过程域了解安全目标与安全控制活动保持相符的相关信息；

●参考合规管理过程域了解合规控制机制与合规情况。

**2.GOM 表**

|  |  |  |  |
| --- | --- | --- | --- |
| 目 标 | 问 题 | 度 量 项 | 度量项说明 |
| G1:识别云端 所有的隐私数据 | Q1:将云端所有 隐私数据都识别出 来 | M1:法规识别方法 | 建立规程确保由组织了解相关法律法规  风险的人员参与对业务场景的评估，以确保 涉及个人数据的业务场景中均被识别 |
| M2:隐私数据识别与 分析方法 | 隐私数据的识别应包括隐私数据定义及 隐私数据分析等内容 |
| G2:建立有效 的隐私管理过  程，以满足业务 需求 | Q2:使云端所有 识别出来的隐私数 据都受到保护 | M3:护隐私数据保护 管理策略 | 隐私保护策略应该包括以下内容：在收 集、使用或披露个人信息时要获取对方同 意；在法律允许的范围内公正地收集信息： 制订明确、易懂的个人信息保护政策 |

**第9章■安全与隐私**

**续表**

|  |  |  |  |
| --- | --- | --- | --- |
| 目 标 | 问 题 | 度 量 项 | 度量项说明 |
| G2:建立有效 的隐私管理过  程，以满足业务 需求 | Q2:使云端所有 识别出来的隐私数 据都受到保护 | M4:隐私数据保护过 程 | 隐私保护应在法律法规要求和业务需求  之间找到平衡，解决隐私保护问题最好的方 法是将隐私保护规则与技术结合起来，采用 匿名保护技术、数字水印技术、密文搜索技 术等隐私保护技术 |
| M5:数据分析过程中 的隐私保护规范 | 组织应制订对使用的数据所涉及的个人  信息进行适当处理的隐私保护规范，规范应 包括如下内容：在数据分析前执行数据匿名 化处理，分析人员对匿名化处理后的数据进 行分析；明确数据分析过程中与个人隐私相 关的红线，并对涉及到个人隐私的数据分析 需求进行人工审核，针对具体的数据分析场 景制订相应的隐私保护方案 |
| G3:组织提供 可信的隐私数据 保护 | Q3:现有的隐私 数据保护策略是否 有效 | M6:首席隐私官 | 组织通过设立的首席隐私官落实梳理组  织所有的隐私保护要求清单，并通过定期跟 进监管机构，更新该要求清单。定期组织对 用户信息的有关保护措施进行全面评审，以 确保措施的合规性 |

**3.** **成熟度等级介绍**

**Level 1: 初始级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 1.1 | 识别组织处理的隐私数据 | 隐私数据列表 |

【成熟度等级说明】

隐私不同于安全，隐私涉及一个人决定什么样的信息愿意让其他人知道、哪些 人可以知道以及决定那些人什么时候可以知道的权利，安全则是用来执行这些隐私 权的“工具”。

各国对隐私的定义及隐私保护的法规都有不同，组织应识别所处理的隐私数据。 隐私数据的识别应包括隐私数据定义及隐私数据分析等内容，识别隐私数据时须注 意普通数据的组合等操作有可能将导致隐私数据的形成。

**Level 2: 管理级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.1 | 采取适当的措施对已识别的隐私数据进行 保护 | ●隐私数据保护技术列表 |

 **■云端数据治理■**

【成熟度等级说明】

云计算环境下，个人数据的收集、挖掘、分析变得高效、可靠，但组织如不能对 海量的敏感信息进行处理和保护，将面临诉讼风险。隐私保护应在法律法规要求和业 务需求之间找到平衡，解决隐私保护问题最好的方法是将隐私保护规则与技术结合起 来，可以考虑采用匿名保护技术、数字水印技术、密文搜索技术等隐私保护技术。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.2 | 按最小够用原则收集个人数据 | ●个人数据收集列表  ●个人数据收集评估制度 |

【成熟度等级说明】

在涉及个人数据收集的业务场景中建立固定的评估机制，由了解相关法律法规 风险的人员参与对业务场景的评估。业务场景中对个人数据的收集不应超出业务所 需的范围，应满足最小够用原则。

**Level 3: 定义级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.1 | 建立和维护隐私数据保护管理策略 | ●隐私数据保护策略  ●隐私数据技术保护操作指引 |

【成熟度等级说明】

鉴于云端用户个人信息的重要敏感性，组织应建立个人信息保护规程。相关内 容应包括：

●在收集、使用或披露个人信息时要获取对方同意；

●在法律允许的范围内公正的收集信息；

●制订明确、易懂的个人信息保护政策。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.2 | 执行个人数据收集的业务场景的安全风险 评估 | ●针对个人数据收集的风险评估报告 |

【成熟度等级说明】

在进行恰当保护之前，必须先研究和评估资产面临的脆弱性及其级别，这样才 能确定和实现真正有效的控制和防范措施。因此，组织应基于组织中个人数据保护 原则，由个人数据保护团队执行个人数据收集的业务场景的安全风险评估。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.3 | 组织制订了数据分析过程中的隐私保护 规范 | ●数据分析过程隐私保护规范 |

**第9章■安全与隐私**

【成熟度等级说明】

组织应制订对使用数据所涉及的个人信息进行适当处理的隐私保护规范，规范 应包括如下内容：

●在数据分析前执行数据匿名化处理，分析人员对匿名化处理后的数据进行分析；

●明确数据分析过程中与个人隐私相关的红线，并对涉及个人隐私的数据分析

需求进行人工审核，针对具体的数据分析场景制订相应的隐私保护方案。

组织应将该规范的内容在组织内部进行宣导，保证组织对相关要求内容理解的 一致性。

**Level 4: 量化级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 4.1 | 设立首席隐私官专门处理隐私问题，负责 监督是否以合法的方式处理敏感数据 | ●隐私策略评审记录  ●隐私保护法律法规清单 |

【成熟度等级说明】

组织应持续跟进组织需符合的隐私保护法律法规要求，以保证组织业务的发展 不会面临合规的风险。通过设立的首席隐私官落实梳理组织所有的隐私保护要求清 单，并通过定期跟进监管机构，更新该要求清单。定期组织对用户信息的有关保护 措施进行全面评审，以确保措施的合规性。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 4.2 | 隐私保护教育培训 | ●隐私保护培训教材  ●隐私保护考试试题  ●隐私保护培训制度 |

【成熟度等级说明】

固化针对员工和第三方人员的入职、在职的隐私保护意识教育流程，通过培训、 考试等手段提升整体的隐私保护意识水平。

基于云计算技术的“云课堂”是一种高效、便捷、实时互动的培训形式。使用 者移动互联网，进行简单易用的操作，便可快速高效地与各地学生、教师同步分享 语音、视频及数据文件。

**Level** **5:优化级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 5.1 | 组织通过出版物、会议和白皮书分享所获 得的最佳标准实践和经验教训 | ●与隐私保护相关的演示材料和其他发表的作品 |

【成熟度等级说明】

组织中IT 安全管理专业人士应订阅专业杂志和网上各种论坛、专业讨论群，分

**■云端数据治理■**

享或阅读隐私保护相关主题。

**4.** **实施案例**

组织应依照国家相关法律法规，建立完整的个人信息保护安全策略，包括技术、 管理等措施，对个人信息处理过程中可能出现的泄露、丢失、损坏、篡改、不当使 用等事件制订预案，比如建立风险应对策略，采取风险控制措施，监控风险变化等， 确保对云用户个人信息的有效保护，降低风险。组织应制订个人信息保护的教育培 训计划并组织落实，其次建立个人信息保护的内控机制，定期开展检查，并形成检 查评估报告。

(1)文件名称列表

《隐私声明/政策》

《隐私保护管理程序》

《隐私保护操作指南》

(2)记录性文件名称列表

《隐私保护操作记录》

(3)实施案例

|  |
| --- |
| 隐私声明/政策  隐私声明/政策是个人信息控制者保证个人信息保护知情权的重要手段，是约束自身行为和配合监督管理的重要 窗口。隐私声明/政策的内容应包括但不限于以下方面。  (1)适用范围。包含隐私声明所适用的产品或服务范围、隐私声明所适用的用户类型、隐私声明生效时间、更 新时间等。  (2)目前遵循的个人信息保护协议和取得的认证。包含个人信息控制者目前主动遵循的国际或国内的个人信息  保护标准、协议，以及个人信息控制者目前已取得的与个人信息保护相关的权威第三方机构认证。  (3)个人信息的收集。根据不同产品、服务类型详细列出收集的个人信息类型，并说明收集个人信息的目的，  个人信息包括但不限于：个人为正常使用功能所提供的数据、使用功能时产生的数据、统计分析得到的数据、使 用自动数据收集工具得到的数据、通过第三方获取的数据等。  (4)个人信息的使用。个人信息控制者根据收集到的不同数据类型，详细说明如何使用个人信息，如实现用户 请求、提供个性化服务、支持和改进服务、商业情报、广告支持等。  (5)个人信息的共享。详细说明需要共享的对象、需要共享的原因、需要共享的数据类型、对接收共享数据方 的约束和管理、共享前是否需要征求用户的同意等。此外，还需说明何种情况下个人信息控制者会不经过用户同 意共享数据，如响应执法机关和政府机构的要求、进行个人信息保护安全审计、保护用户避免遭受欺诈和严重人 身伤害、保护企业财产、权利及信息系统安全、企业的并购、出售和财产转移等，并说明如何确保个人信息在此 过程中得到有效保护。  (6)用户对个人信息的控制。详细说明个人对其个人信息拥有何种控制权，内容包括但不限于：信息收集、 使用和共享时允许用户选择的数据范围，用户所具备的访问、修改、删除等控制权限，用户个人信息保护偏好 设置，用户可以选择的通信和广告偏好，用户不再使用服务后要求删除个人信息的渠道、用户进行隐私维权的 有效渠道等。 |

**第9章■安全与隐私**

**续表**

(7)个人信息的安全保护。详细说明个人信息控制者对个人信息进行安全保护的措施。包括但不限于数据完整 性保护措施，数据传输、存储和备份过程的加密措施，数据访问、处理的授权机制，数据的保留和删除机制，数 据存储和处理的地理位置，确保第三方使用共享数据的安全保护机制等。

(8)个人信息的跨境传输。如果因业务需求、政府和司法监管要求存在跨境信息传输情况，需详细说明需要进 行跨境传输的数据类型，以及跨境传输遵守的标准、协议和法律机制(合同等)。

(9)自动数据收集工具说明。如果个人信息控制者或其授权第三方使用自动数据收集工具收集个人信息，则需 要对使用的技术机制做详细描述，同时，说明使用自动工具收集个人信息的目的，并向用户提供限制自动工具进 行数据收集的方法。常见的自动数据收集工具有： Cookie、脚本、Web信标、Flash Cookie、内嵌 Web 链接、本地 存储器等。

(10)变更通知。个人信息控制者在隐私惯例发生重大变化时，需及时更新隐私声明，并说明使用何种机制并及 时通知用户。

(11)法律、政策等规定的其他内容。如未成年人、老年人、残疾人使用系统注意事项等。

(12)联系方式。个人信息控制者需要明确给出处理个人信息保护相关反馈、投诉的渠道，如个人信息保护责任 部门的联系方式、地址、电子邮箱、用户反馈问题的表单等，并明确用户可以收到回应的时间。

个人信息示例

1 身份关联信息

身份关联信息，是指能够切实可行地单独或通过与其他信息相结合，识别出具体的个人信息或此等信息的集合。

|  |  |
| --- | --- |
| 信 息 分 类 | 包 含 成 分 |
| 用户身份信息 | 指能单独、准确识别用户真实身份的影印件及其他信息。包括但不限于身份证、 军官证、护照的影印件、身份证号信息、机动车登记号码、房产证号码等 |
| 生物识别信息 | 指与用户具有唯一对应关系的用户生理信息。包括但不限于指纹、声纹、虹膜、 脸 谱 、 D N A 等 |
| 用户基本资料 | 指对于用户社会属性和自然属性进行描述的信息。包括但不限于生日、性别、职 业、职位、民族、国籍、政党属性、邮编、姓名、宗教信仰、地址、工作单位等 |
| 病理及健康信息 | 指用户因生病医治而产生的相关记录或由用户身体健康状况产生的相关信息。包 括但不限于病症、住院记录、医嘱单、检验报告、手术及麻醉记录、护理记录、医 疗费用、用药记录等病理信息及体重、身高、肺活量等个人健康信息 |
| 身份标识和鉴权信息 | 指用户使用服务过程中区别于其他用户的标识信息以及对该标识信息进行识别确 认的信息。包括交易类服务身份标识和鉴权信息及普通服务身份标识和鉴权信息 |
| 普通服务身份标识和 鉴权信息 | 包括电话号码，QQ等社交类软件昵称、IP地址、邮箱地址及前述有关的密码、 口令、密码保护答案等 |
| 交易类服务身份标识 和鉴权信息 | 包括支付宝、财付通等交易类软件账号；银行卡账号；证券账号、支付宝、财付 通等交易类软件账号的密码、用户个人数字证书等 |

**■云端数据治理**

续表

续表

|  |  |
| --- | --- |
| 信 息 分 类 | 包 含 成 分 |
| 财产信息(金融信息) | 指用户在真实生活或虚拟服务中所拥有和使用的有关财产的信息，包括真实财产  信息和虚拟财产信息  真实财产信息，包括支付宝、财付通等交易类软件账号、银行卡等账户及其各账 户下的支付记录、交易记录、流水记录等信息  虚拟财产信息，包括Q币Q点、点券等虚拟货币，虚拟交易个人信息；游戏各类 兑换码；金币等相关记录虚拟财产信息 |
| 联系人和关系链信息 | 联系人信息指用户在电话、实时聊天工具、微博、邮件中的联系人信息，包括相 应的账号、用户名、头像、签名、用户备注等信息  关系链信息包括但不限于通讯录、好友列表、群列表、朋友圈列表、关注对象列 表、备注等 |
| 位置信息 | 包括经纬度、地区代码、小区代码、基站号 |
| 服务记录 | 指通过Text、SML、日志等储存的用户的操作记录、服务内容信息记录等，包括 社交信息记录、消费记录、游戏记录、视频操作流水记录、广告曝光、点击和转化 日志、业务日志等 |
| 设备信息 | 指包括硬件型号、设备MAC地址、操作系统(安卓、iOS、WM等)、软件列 表唯一设备识别码(如IMEIandroidID/IDFA/OPENUDID/GUID、SIM卡IMSI信 息)等在内的描述设备基本情况的信息 |
| 本地信息 | 指用户建立的、存储在移动设备上的，或通过移动设备访问的软件列表、日历、 地址簿、电话短信记录、相册、视频等信息 |
| 其他用户信息 | 指其他可被信息系统所处理、与特定用户相关、能够单独或通过与其他信息结合 来识别特定用户的个人信息 |

2 非身份关联信息

非身份关联信息，是指除身份关联信息之外的、与可被识别的自然人相关的个人信息。

**9.2.4** **合规管理**

**1.** **简介**

1)目的

组织建立和维护机制，保证自身的合规性。

2)介绍

每个组织都会受到政府和行业法规的约束，在这些法规中，有许多规定了数据 和信息如何进行管理。合规性管理，也称为符合性管理，就是在这些法规要求之下，

**第9章■安全与隐私**

开展的一系列达标活动的集合。应实施适当的控制措施，以确保数据安全相关的法 规得到遵从。

作为数据所有者，组织面临着其用户日益复杂的外部合规要求。这些要求包括 行业标准、规章制度和法律法规等。组织通过机制应能完成以下要求。

(1)组织应识别出所有相关的法律、法规、规章和合同要求，形成文件并保持 更新，例如，支付卡行业对数据安全标准的要求。

(2)组织应识别出跟记录相关的法律、法规、规章、合同和业务要求。根据要 求对记录进行保护以防其丢失、毁坏、伪造、未授权访问和未授权发布。例如，用 户的合同中包括对事务日志和审计日志的保存要求的条款。记录应根据法律、法规、 规章、合同和业务要求进行保护和保存。例如：作为组织在法令或法规规则下运行 证据的记录，以确保充分防御潜在的民事或刑事诉讼，或者和股份持有者、外部方 和审计员确认组织的财务状况，可以根据国家法律或规章来设置信息保存的时间和 数据内容。

(3)组织应识别出针对隐私和个人可识别信息保护相关的法律、法规和合同条 款的要求。例如：随着电子病历的应用，服务于卫生保健业务的云计算提供商应当 识别我国《医疗机构病历管理规定》(2013 年版)对个人可辨识的健康信息的保护 要求：“除为患者提供诊疗服务的医务人员，以及经卫生/计生行政部门、中医药管 理部门或者医疗机构授权的负责病案管理、医疗管理的部门或者人员外，其他任何 机构和个人不得擅自查阅患者病历。”

(4)组织应识别出针对密码控制的相关的协议、法律和法规。例如：国家对加 密技术的限制使用。

识别合规性要求的机制至少每年被评审一次，确保对自身业务流程造成影响的 事项都已被识别。

3)相关过程域

●风险管理过程域帮助组织识别合规风险；

●隐私保护过程域提供了组织隐私保护措施及落实情况；

●数据安全过程域提供了组织安全目标与安全控制活动保持相符的相关信息。

**2.GQM 表**

|  |  |  |  |
| --- | --- | --- | --- |
| 目 标 | 问 题 | 度 量 项 | 度量项说明 |
| G1:识别所需遵 从的法律 | Q1:组织是否将所 有需遵从的法律法规 都识别出来了 | M1:识别出的所有相关 的法律、法规、规章和合 同要求 | 组织负责合规管理的人员应具备对 合规要求解读能力和对业务情况的理 解力，相关人员应熟悉制订和采用法 律的原因以及它们的整体目标 |

 **■云端数据治理**

**续表**

|  |  |  |  |
| --- | --- | --- | --- |
| 目 标 | 问 题 | 度 量 项 | 度量项说明 |
| G2:建立有效的 合规性管理过程 以满足业务需求 | Q2:云端所有识别 出来的法律法规是否 都得到遵从了 | M2:已识别的法规管理 规范 | 组织通过信息安全教育培训等方  式，建立和提高相关人员的法规遵从 等信息信息安全意识及能力 |
| M3:法律法规识别、分 析流程 | 组织应建立法律法规识别、分析的 流程，搞清楚法规的相关性，该法规 需要什么样的企业策略和规程，我们 应该如何遵从等问题 |
| G3:组织提供可 信的合规型管理 | Q3:现有的合规性  管理策略是否有效 | M4:审计机构 | 组织应定期进行独立的审计与评  估，用于识别和消除任何可能违背现 有安全策略、程序，以及合同、法律 法规的不符合项 |
| M5:合规监控机制 | 组织应建立对合规性识别、监视和 控制的机制，从而核对合规性管理策 略和措施的实际效果是否与预见的相 同，并寻找机会改善 |

**3.** **成熟度等级介绍**

**Level 1: 初始级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 1.1 | 识别组织需遵从的法律法规 | ●遵从法规列表 |

【成熟度等级说明】

负责该项过程的人员应具备对合规要求的解读能力和对业务情况的理解力。项 目人员应熟悉制订和采用法律的原因以及它们的整体目标，必要时可以聘请或请教 律师、安全专家充当组织的顾问，从他们那里获得法律帮助。

**Level 2: 管理级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.1 | 对已识别的法规进行分析，确保遵守法律 法规，并采取适当的方式向利益相关者传达 这些法律法规 | ●法规遵从培训演示材料 |

【成熟度等级说明】

如果个人信息保护不当，容易出现个人信息的泄露、丢失、损坏、篡改、不当 使用，甚至触犯法律法规要求等。通过个人信息安全意识教育培训，可以建立和提

**第9章■安全与隐私**

高相关人员的个人信息保护意识，让他们意识到自身个人信息保护的职责以及履行 职责的方法。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 2.2 | 项目人员有较强的合规解读能力 | ●遵从法律解读材料 |

**【成熟度等级说明】**

负责该项过程的人员应具备对合规要求解读能力和对业务情况的理解力。 Level 3: 定义级

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.1 | 设立审计机构，负责法规遵从管理策略的 设立及维护 | ●法规遵从管理策略  ●设立法规遵从审计组织 |

【成熟度等级说明】

定期进行独立的审计与评估，用于识别和消除任何可能违背现有的安全策略、 程序，以及合同、法律法规的不符合项。从事这些评审的人员应该具备适当的技能 和经验，例如：内部审计人员需要经过云计算安全管理体系内审员的培训；外部审 计人员需要取得专业的认证资格。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 3.2 | 建立法律法规识别、分析的流程 | ●法律法规识别、分析程序  ●具体法律法规遵从建议 |

【成熟度等级说明】

组织知道需要符合什么样的法律法规很重要，同时还要知道如何恰当确保遵守 法律法规，以及如何正确地向必要的利益相关方传达这些法律法规。

组织应建立法律法规识别、分析的流程，搞清楚法规的相关性，如该法规需要 什么样的企业策略和规程、我们应该如何遵从等问题。

**Level 4: 量化级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 4.1 | 制订计算机和业务道德方面的指南 | ●计算机和业务道德方面指南  ●业务道德培训记录 |

【成熟度等级说明】

组织制订计算机和业务道德方面的指南，并成为员工手册的一部分。通过“云 课堂”等高效、便捷、实时互动的培训形式对员工进行培训，并通过在线测试等评 估方式跟踪培训情况，培训和测试均强制进行，并纳入绩效考核体系。

**■云端数据治理■**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 4.2 | 定期审核合规情况 | ●个人数据保护优化方案  ●涉及数据收集相关条款的合同模板 |

【成熟度等级说明】

定期审核个人数据收集场景的评估情况，归纳总结出恰当的个人数据收集的保 护措施，如根据不同的数据收集场景提供适当的合同模板、协议模板，并将相应的 保护措施及时同步给各业务团队。

**Level** **5:** **优化级**

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 5.1 | 建立合规监控机制，并对违规趋势进行分 析 | ●合规监控机制  ●违规趋势分析报告 |

【成熟度等级说明】

建立对合规性识别监视和控制机制，从而核对合规性管理策略和措施的实际效 果是否与预见的相同，并寻找机会改善。

|  |  |  |
| --- | --- | --- |
| 序号 | 等 级 要 求 | 实 例 |
| 5.2 | 组织通过出版物、会议和白皮书分享所获 得的最佳标准实践和经验教训 | ●与合规管理相关的演示材料和其他发表的作品 |

【成熟度等级说明】

组织IT 安全管理专业人士应订阅专业杂志，并通过网上各种论坛、专业讨论群 分享或阅读合规管理相关主题。

**4.** **实施案例**

要实现有效的合规性管理，首先应该了解组织需要符合什么样的法律法规，如 何确保遵守法律法规，以及如何正确地向必要的利益相关者传达这些法律法规。其 次，应该制订合规性规划，明确需要部署些什么内容来满足必要的内部和外部驱动。 再次，审计团队评估组织在满足需求方面做得如何。最后，将合规性审计活动纳入 到关键绩效指标中。

1)文件名称列表

《云端数据治理合规管理程序》

2)记录性文件名称列表

《云端数据法律法规清单》

**第9章■安全与隐私** 

3)实施案例

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 云端数据治理合规管理程序  1 目的  为保障公司的云端数据治理活动符合国家相关法律法规、标准及合同中相关要求。  2 适用范围  适用于公司合规性、知识产权保护和个人隐私保护等方面的管理。  3 术语和定义  云端数据：指公司通过传输或拷贝并存储在云端的数据，以及公司从云端直接获取的数据。 4职责   |  |  | | --- | --- | | 部门/岗位 | 主 要 职 责 | | 数据治理委员会 | 编制和维护本程序，牵头组织落实本程序，监督、检查本程序执行情况  组织人员识别出适用本公司的数据治理方面的法律法规，并确定具体适用的条款 维护《云端数据治理法律法规清单》,并在管理评审会议中报告有关法律法规对本 公司数据治理和业务的影响 | | 数据安全管理员 | 识别数据治理相关法律法规  负责组织、实施公司技术符合性检查及检查后的整改工作 |   5 管理规定  5.1 云端数据治理法律法规识别  1)识别流程  相关的安全法律法规的识别流程如下：  (1)云端数据治理法律法规的识别工作由负责法律事务人员进行，各部门数据管理员在法律法规的适用性识别 上予以协助；  (2)公司的各个部门的文件管理人员负责将接收到的云端数据治理相关法律法规转发给法律事务负责人员予以 识别；  (3)负责法律事务人员识别出适用本公司的云端数据治理方面的法律法规，确定具体适用的条款，并上报数据 治理委员会，数据安全管理员将识别出的法律法规登记在《云端数据法律法规清单》中；  (4)当法律法规出现修订、合并、删减、取消时，法律事务负责人员要及时获取最新的法律法规，由数据安全 管理员更新《云端数据法律法规清单》,并识别确定具体适用的条款；  (5)数据治理委员会针对有关法律法规对本公司云端数据治理活动和业务的影响，在每年的管理评审会议上进 行报告。  2)识别内容  应识别的云端数据治理方面的法律法规包括但不限于以下6方面。  (1)全国人大及人大常委会发布的法律和法律解释；  (2)国务院、最高人民法院、最高人民检察院批准和颁布的法律、法规、法律解释；  (3)国家职能部门批准和颁布的规章制度；  (4)省、市等部门颁布的规章制度；  (5)适用于我国的国际公约、国际条约、商业惯例和规范性文件等；  (6)涉及国外监管要求的还要识别当地数据治理相关的法律法规等。 |

 **■云端数据治理**

续表

|  |
| --- |
| 5.2数据保护和个人信息的隐私保护  (1)数据保护依公司保密管理相关要求执行；  (2)对于因客户投资、员工入职、合作伙伴进场、第三方来访等原因获得的个人隐私信息，除当事人允许，为 当事人获取福利或权益需要(且需要当事人允许)、司法部门依法调查和法律法规要求外，公司各部门不得向无关 的第三方泄露任何个人的隐私信息。  5.3符合安全策略和标准  云端数据治理团队通过以下方式确保各项数据治理活动和流程被正确地执行，进而符合相关管理标准和公司的 各项安全制度规定。  (1)云端数据治理制度和流程执行情况的督促和推进；  (2)云端数据安全技术和产品的实施和应用；  (3)云端数据治理培训和宣传；  (4)管理评审；  (5)纠正、预防措施的实施和落实效果跟进。 |





**云端数据治理体系与实施**

**第** **1** **0** **章** **云端数据治理体系**

**10.1** **云端数据治理体系的必要性**

虽然云计算的发展有力地推动了云端数据治理的工作进程，但由于缺乏满足治 理目标和要求的管理手段，在该进程中，许多组织仍然面临着一些问题和困扰。主 要表现在以下方面。

(1)云端数据治理方案失败，主要原因是方案没有围绕实际的业务问题定义云 端数据治理的范围。对大多数组织而言， IT 技术并不能完全配合和支撑各种业务目 标。为了实现组织的战略目标，首席信息官及各高级主管需要努力设法使IT 技术能 配合和支撑各种业务目标。只有云端数据治理方案开始解决已识别的业务问题时， 业务职能部门才会支持它将范围扩展到更多区域。

(2)云端数据治理方案低于预期结果，主要原因是数据识别不完整。例如，获 取云端数据的类型、使用云端数据的方式方法、要满足哪些业务目标，以及谁有权 拥有这些云端数据等方面并未明确界定。如果基本的云端数据不可靠，那么大多数 组织的云端数据治理方案可能会失败或者低于预期结果。

(3)云端数据治理方案实施困难，主要原因是无法得到业务部门的支持。云端 数据治理方案经常是由组织内的IT 部门或者数据管理员个人发起的，在实施过程中， 各业务部门没有给予充分的配合。为了推动云端数据治理规范的执行和落地，在建 立云端数据治理组织架构时，应该考虑由业务部门来主导云端数据治理活动， IT 部 门来执行，并且引入相应的考核体系。

任何云端数据治理方案都应该考虑到云端数据的各种特点：数量体量大、种类 多、产生速度快、数据来源广以及数据质量不可靠等。数据架构团队(数据分析师) 要完全识别、定义并分析这些云端数据，必须征询组织的相关方(包括数据及业务 流程拥有者)的意见。这样做能保证在方案实施的初期，让包括组织战略制订者、 业务流程、数据拥有者以及 IT 架构专家在内的云端数据治理部门弥补差距和不足， 从而确保云端数据治理方案的有效实施。

总体来看，有效推动云端数据治理，亟待解决的不是技术问题，更多的是各种 管理问题。云端数据治理是一套持续改善管理机制，通常包括了数据架构组织、数 据模型、政策及体系制订、技术工具、数据标准、数据质量、影响度分析、作业流 程、监督及考核等内容，建立云端数据治理体系是推进云端数据治理实施的重要实

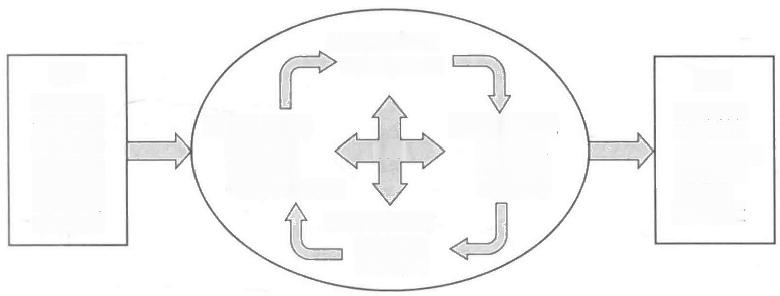
**第10章■云端数据治理体系**

现途径。

实施云端数据治理可通过建立云端数据治理的体系，即围绕云端数据治理的 过程、所需的准则和方法、过程所需的资源等与实施相关的因素，建立一个完善 的云端数据治理体系。这个体系包括支撑云端数据治理的战略、各个治理过程的 目标、岗位职责和组织文化、软硬件资源以及过程管理等。实施云端数据治理的 长期目标是通过云端数据治理，为组织的利益相关者带来价值，实现组织的战略 发展目标。

**10.2** **云端数据治理体系基本框架**

云端数据治理过程采用“策划 (Plan) 一实施 (Do)— 测评 (Check)— 处置 (Action)” 的循环作用和持续改进来提高所有过程有效性的 PDCA 方法，并结合借 鉴项目管理、数据治理、IT 治理等领域的实施方法论，形成云端数据治理管理体系。 其中，P 是统筹和规划，指根据云端数据治理的项目目标，对照业务需求，确定云 端数据治理实施的范围和战略，规划实现目标并确定必要的过程；D 是构建和运行， 指在受控条件下，运行云端数据治理的过程； C 是监督和评估，指对云端数据治理 体系进行评估与诊断、监视与测量、审核、考核，并报告结果； A 是改进和优化， 指采取措施，包括治理审计、绩效考核、纠正措施和预防措施等，以持续改进过程 绩效。组织云端数据治理体系基本框架如图10-1所示。



输出

·数据政策

·数据标准

·已解决问题 ·数据管理

·质量数

·数据价值

输入

·业务目标 ·业务战略 ·治理目标 ·治理战略 ·数据需求 ·数据问题 ·监管需求

A 改进和优化 ·合规管理 ·治理审计 ·绩效评估

C 监督和评估 ·改进措施 ·改进计划

P 统筹和规划 ·差距分析 ·确定范围

D 构建和运行 ·数据治理项目

·制定目标和方案

领导

**图10-** **1** **组织云端数据治理体系基本框架**

为使组织的云端数据治理体系有效运行，须确定和管理众多相互关联的活动。 通过使用资源和实施管理，将输入转化为输出的一项或一组活动，可以视为一个过 程。图10-1反映的是以过程为基础的云端数据治理体系。该图的左侧展示了云端数 据治理体系输入的具体内容，以启动相应的活动。应根据组织的业务战略和目标，

 **■云端数据治理**

制订云端数据治理体系的战略和目标。根据业务需求，识别组织的数据需求和数据 问题，同时需要考虑监管的合规性和一致性。有效的云端数据治理体系一定是与组 织的业务价值紧密相连的。组织使用IT 技术配合和支撑各种业务，目的都是为了达 成业务目标和实现业务战略。

图10-1 的右侧是云端数据治理体系的输出，也是体系的交付物。数据策略是 云端数据治理的基本规则，这些策略贯穿云端数据治理体系的所有过程，描述了什 么可以做，什么不该做。数据规程是云端数据治理的程序，是完成某项活动的详细 步骤。云端数据治理体系是识别、管理和解决云端数据问题的有效途径，最终目标 之一是为了管理并解决云端数据问题。在体系的运行过程中，云端数据治理专业人 员应不断沟通、培训和推广数据的重要性，提高利益相关者对云端数据治理体系的 认可。

**10.3** **云端数据冶理体系过程模型**

云端数据治理体系 (Cloud Data Governance System,CDGS) 过程模型是基于云 端数据治理体系的基本框架，借鉴项目管理、数据治理、IT 治理和管理体系的实施 方法论，并结合云端数据治理的特征，总结过程模型。CDGS 过程模型中包括了云 端数据治理的核心过程。

●首先就组织现状进行了解并进行差距评估分析；

●基于组织现有业务战略、IT 战略、规划、组织、资源的理解，进行云端数据 治理及治理范围的确认；

●制订整体性框架建议，经业务部门修正后据此作为实施的基本结构，进行详 细作业规划及工作任务分解；

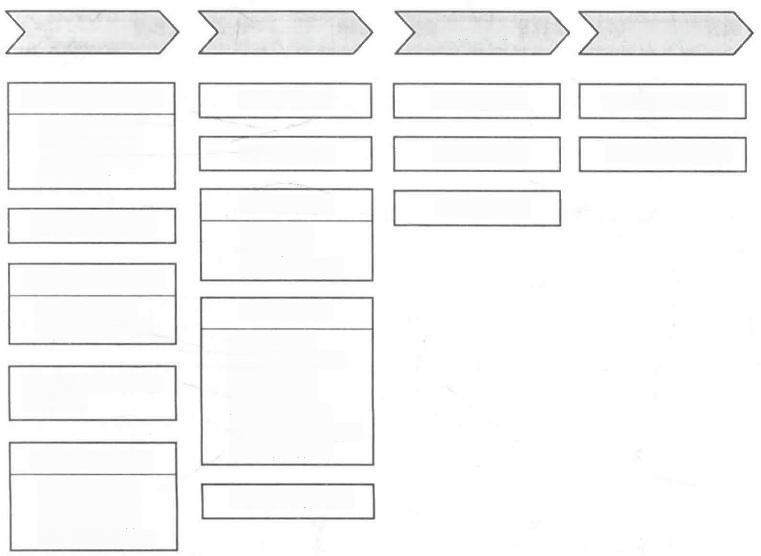
●重要关键点先进行培训以利于制度建立的符合性；

●文档撰写前共同讨论撰写大纲及关键点以利于可操作性；

●共同讨论制度文档的可行性，并帮助文档审查。 CDGS 过程模型由四个阶段组成，如图10-2所示。

过程模型是云端数据治理体系开展活动所需要的最完整模型，其中部分过程可 依据组织本身的范围或管理成熟度进行删减或简化。

**第10章■云端数据治理体系**

统筹和规划 构建和运行 监督和评估 改进和优化

SI1 制定战略和目标

·制定体系战略

·制定体系目标

·获取高层支持

·差距分析

S1.2 确定体系范围

S1.3 建立组织与角色

·建立组织架构

·任命管理者代表

S1.4制订体系建设和

运行计划

S1.5形成体系文件

·制定数据策略

·制定数据规程

·定义绩效指标

·文件评审与发布

S2.1 体系宣贯

S2.2 实施策划

S2.3项目启动

·项目目标

·项目范围

·项目组织架构

S2.4项目实施

·制定路线图

·定义业务词汇表

·管理元数据

·管理主数据

·提高数据质量

·数据生命周期管理

·数据安全管理

S2.5项目过程管理

S3.1绩效评估

S3.2内部审计

S3.3 安全审计

S4.1制订改进计划

S4.2 实施改进措施

图10 - 2 CDGS 过程模型

**第11章** **云端数据治理体系实施**

**11.1** **统筹和规划**

云端数据治理体系的统筹和规划阶段目标是要统筹和规划体系战略和目标， 谋划必要的治理环境，做好云端数据治理实施的准备。统筹和规划阶段的主要内 容包括：

●制订云端数据治理体系战略，规划云端数据治理体系的目标、范围和方法；

●明确云端数据治理体系的组织机构、治理主体、职责分工，以及利益相关方 的沟通机制；

●统筹规划资源，制订体系建设和运行计划，明确构建和运行、监督和评估、 改进和优化阶段的内容、程序和机制；

●根据体系战略、业务要求和利益相关方期望，制订云端数据策略和规程，定 义绩效指标，并形成云端数据治理体系文件。

**11.1.1** **制订战略与目标**

**1.** **制订体系战略**

云端数据治理体系战略是对云端数据治理活动的顶层设计和远程规划，其目的 是在云端数据治理活动得到数据治理相关者的理解和认同前提下，对未来一段时间 内数据治理活动进行规划。为了获得云端数据利益相关者支持和认可，云端数据治 理战略必须能够准确地反映云端数据提供者和数据消费者的业务期望，并考虑云端 数据价值链上的所有相关参与者的利益，从而让利益相关者相信，云端数据治理的 方案是有价值的。体系战略的制订一般包括以下三个阶段。

1)战略分析

总结影响组织云端数据治理体系的关键因素，并确定在战略选择步骤中的具体 影响因素，它包括以下三个主要方面。

(1)分析组织的业务战略和目标，作为制订和评估组织云端数据治理体系战略 的依据；

(2)对外部环境进行分析，包括国际、国内、地区和当地的各种法律法规、技 术、竞争、市场、文化、社会和经济因素，有助于理解外部环境；

**第11章■云端数据治理体系实施**

(3)对内部因素进行分析，要了解组织自身所处的相对地位，具有哪些资源、 关键成功因素及已识别的风险；了解与云端数据治理体系有关的利益相关者的利益 期望，以及在战略制订、评价和实施过程中，这些利益相关者会有哪些反应。

2)战略选择

战略选择阶段体现的是云端数据治理体系的发展方向。体系战略的制订一般经 过首席数据官的指导和授权，由组织级数据治理委员会制订。

云端数据治理体系的战略至少应包括以下内容。

(1)云端数据治理体系的愿景；

(2)指导原则、价值观和管理愿景；

(3)云端数据治理的长远目标；

(4)云端数据治理体系的成功衡量标准或者关键的成功因素。

通常从以下两方面评估战略备选方案： 一是考虑选择的战略是否发挥了组织的 关键成功因素，并克服了劣势，是否利用了机会，将风险削弱到最低程度；二是考 虑选择的战略能否被组织利益相关者所接受。

3)战略实施和控制

战略实施和控制就是将战略转化为行动，主要包括以下内容。

(1)为了实现云端数据治理体系目标，需要获得哪些资源以及如何使用这些 资源；

(2)为了实现既定的战略目标，需要对组织结构做哪些调整；

(3)如何通过对组织的文化管理来保证云端数据治理体系战略的成功实施。

**2.** **制订体系目标**

云端数据治理体系的目标由组织级数据治理委员会制订，能科学、全面地反映 组织的云端数据治理关键业绩指标。云端数据治理体系的目标按照作用的不同分为 管理目标和过程目标，管理目标包含云端数据治理成本、数据治理达成的目标、人 员变现、数据专业人士数量等；过程目标包含项目目标、主数据和元数据质量、数 据操作性能等指标。

云端数据治理体系的目标应该遵循 “SMART”(S=Specific 、M=Measurable、

A=Achievable 、R=Relevant 、T=Time-based) 原则，即该目标应是具体的、可测量的、 可实现的、现实的、就制订的目标时间范围来说是及时的，并与云端数据治理体系 的战略一致。云端数据治理体系的目标管理一般包括以下四个阶段。

1)确定目标责任和责任人

根据体系的战略确定目标责任和目标责任人，目标责任就是对目标达成与否的功 过承载，责任人就是承载这种功过的具体人。云端数据治理体系管理目标的责任人一 般为组织的决策层和管理层的管理人员，可能包括组织级和部门级的数据治理委员 会。过程目标的责任人一般为组织的执行层的人员，可能包括各类数据管理人员。

**■云端数据治理■**

2)设立和实施目标

根据 SMART 原则，将目标责任量化成可操作、可实现、可考量的具体目标， 并落实到目标责任人。例如，云端数据质量管理的目标可以量化为数据价值统计、 错误/需求违规、期望满足程度等。目标责任人可能包括数据质量分析师、数据管理 人员和数据分析师等。

3)监督和评价目标实现过程

在目标实施中，为了确保目标的达成，还必须对目标实现过程进行监督。监督 的目的在于强化对体系目标管理的执行力度。对各级目标的完成情况，要事先规定 出期限，依据已经设立的目标，定期进行统计和分析。对于最终结果，应当根据体 系目标进行评价，并根据评价结果采取措施持续改进云端数据治理体系。

4)评审目标的适宜性

可以从以下三方面评审目标是否适合云端数据治理体系： 一是评价实现体系目 标的各种资源使用情况；二是实现的目标是否还有弹性空间，如是否可以当作基准、 是否需要变更、是否可以保持相对稳定等；三是所实现的目标是否能促进实现组织 的战略。根据评审的结果修订体系的目标。

**3.** **获取高层支持**

云端数据治理体系的形成离不开高层管理者的广泛参与，体系的有效运行离不 开高层管理者的大力推进与协调。可以通过以下途径获取高层对体系的支持。

1)创建虚拟云端数据治理小组

云端数据治理需求经常由不同的数据管理团队提出，例如数据架构团队希望能 将 IT 技术更好地融合至业务中；风险管理团队希望优化数据隐私策略；记录管理团 队希望建立电子文档保留策略；数据库管理团队希望能优化对分类数据的受控访问； 首席数据官希望业务部门管理者支持数据治理活动。这些团队向组织的不同部门报 告，但他们都在尽力解决类似的数据问题。因此，通过成立虚拟云端数据治理小组， 定期召开会议，协调解决数据问题，可以促进云端数据治理体系的有效运行。

2)获取 IT 和业务部门内高级管理人员的支持

对一个企业来说，主要利益相关人士包括股东、管理人员和理事。其他利益相 关人士包括雇员、供应商、顾客、银行和其他贷款人、政府政策管理者、环境和整 个社区。企业治理的顶点是董事会，它代表着主要利益相关人的利益。董事会负责 设置策略来确保企业高层官员执行合适的数据治理。因此数据治理只有在其流程吸 引企业的利益相关者时才会成功。

在云端数据治理体系中，任何依靠数据来获得收益的职能部门人员都是利益相 关者。各级数据管理人员应不断地沟通并推广数据价值的重要性，汇报数据管理部 门的工作成绩，提高利益相关人士对数据治理及效益的重视和认可程度。组织应指 定专人负责与利益相关者进行沟通，可通过公告、电子邮件、纸质文档、培训等方

**第11章■云端数据治理体系实施**

式为利益相关者介绍云端数据治理体系的战略、效益、目标、原则、以及重要数据 管理问题的执行信息等。

**4.** **执行差距分析**

差距分析是指在战略实施的过程中，将数据治理现状与数据治理战略目标进行 对比分析，为策划数据治理项目提供依据。

组织可以通过现场调研、人员访谈、文件审阅及现场查看等方式，了解组织云 端数据治理体系现状，然后与云端数据治理成熟度模型进行对比分析，发现组织在 云端数据治理体系中存在的差距。执行差距分析的过程可以包括以下环节。

1)定义评估的组织范围

数据治理活动的参与范围是有边界的，不同规模和性质的组织参与到数据治理 的部门和人员数量存在较大差距。对于小型组织，数据治理可能就是面向整个组织 的，而对于大型的、跨地域的组织，数据治理可能是多个部门开展小范围数据治理 活动的集合。评估范围选择太大，会导致差距分析工作复杂；范围选择太小，会影 响评估结果的客观性。因此，执行差距分析时的评估范围应当采取适中原则，选取 参与程度较高的部门、与数据流程接触较为紧密的人员、与组织业务目标最贴近的 数据。

2)确定评估的时间范围

执行差距分析的目的是为后续数据治理活动的改进提供参考依据，评估时间过 长会贻误数据治理整体进度，时间过短则难以实现评估有效性。差距分析的评估时 间没有一个固定的长度，与组织数据治理的复杂性存在诸多关联，鉴于数据治理体 系的持续性，差距分析也可以是一个持续的、动态的过程。

3)选取合适的成熟度评价指标体系

本书的第2篇系统地介绍了云端数据治理过程域的成熟度等级评价依据，对数 据战略、数据管理、数据质量、数据操作、数据架构、安全与隐私等6个职能域提 出了很多具体的评估依据。组织应当根据定义的评估组织范围等因素，选择合适的 指标体系对数据治理活动进行评价。当定义的评估组织范围较小时， 一些指标体系 就可以不用纳入进去，例如对于部门级的数据治理评估，数据安全隐私职能域可以 不用选取，因为这是组织层面上应该考虑的问题。

**11.1.2** **确定体系范围**

云端数据治理体系可以覆盖组织的全部或者部分数据管理活动。无论是全部还 是部分，组织都必须明确界定体系的范围，如果体系仅涵盖组织的一部分数据管理 活动，这就变得更重要了。云端数据治理活动失败的根源之一是缺乏与业务价值的 链接。例如， IT 部门独自实施项目，没有和相关业务部门进行联动，由于在没有合

**■云端数据治理■**

适的业务支持的情况下治理数据，造成了组织不能从云端数据治理中获得任何价值。 因此，云端数据治理需要从业务角度(而不是从技术角度)定义数据治理体系的范 围，吸引相关业务领域高层领导的支持，从而保证云端数据治理体系可以获得更高 的业务收益。组织可围绕一个新的业务或者根据对旧业务的变更改造来定义云端数 据治理体系的范围。

例如， 一家保险公司的业务是提供人寿保险服务，因为企业的发展原因，业务 部门提出“要增加汽车保险业务”。目前企业的所有系统和数据都在云环境中，为了 能向已购买人寿保险的客户销售汽车保险，企业需要围绕这个新的业务来开展治理 工作。因此企业的云端数据治理体系的范围可以定义为“汽车保险业务相关的数据 治理”。如果监管要求发生了变化，业务部门则会提出“寿险业务的安全与隐私条例 发生了变化”。为了满足新的需求，企业要围绕这个旧的业务进行相关的云端数据治 理工作。因此企业的云端数据治理体系的范围可以定义为“人寿保险业务变更有关 的数据治理”。

**11.1.3** **建立组织与角色**

**1.** **建立组织架构**

组织需要组建具有足够权限的云端数据治理组织架构以便可以贯穿组织的各个 云端业务、技术和管理部门，从而对整个信息供应链进行治理，实现云端数据治理 体系目标。

组织架构包括三层：顶层为决策层，包括首席数据官和组织级数据治理委员会； 中间层是管理层，包括部门治理委员会和项目指导委员会；底层是执行层，包括数 据管理专员和业务数据管理专员，如图11-1 所示。

云端数据治理组织架构各组成部分的职责描述如下。

(1)首席数据官指导和促进云端数据治理体系的实施，职责包括：

●向全员传达建立云端数据治理体系的重要性和必要性；

●制订云端数据治理体系的愿景和目标，并协调组织内各部门掌控云端数据治 理体系的总方向；

●为体系实施提供内部资源保障；

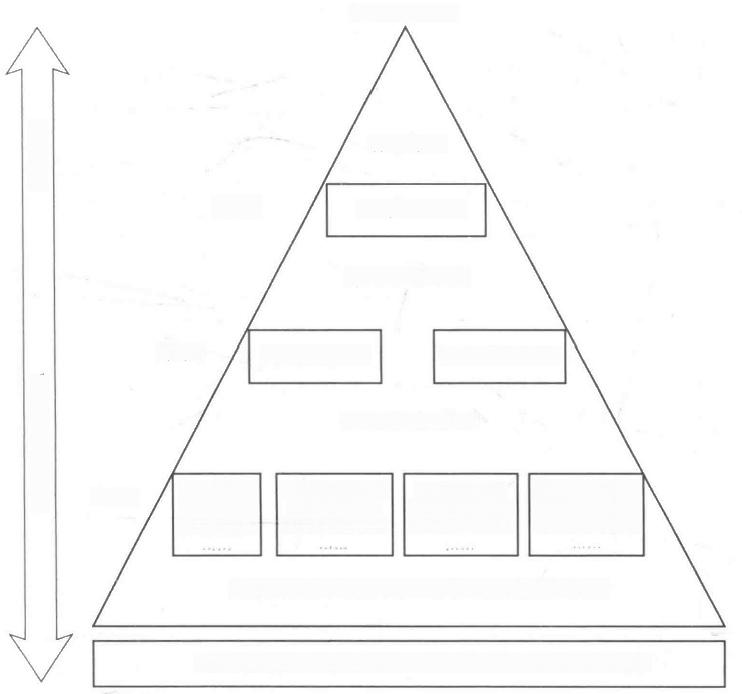
●参与体系的重大决策，并做出决定。 (2)组织级数据治理委员会，职责包括：

●关注云端数据治理体系进展状态，在体系建设重要里程碑处审查体系状态；

●批准影响到体系策划的主要变更；

●协调解决体系实施的重大问题；

●推动、监管体系运作，确保体系能运作成功；

**第11章=云端数据治理体系实施**

數 据 治 理

云端数据治理体系

数据架构师 数据分析师 数据建模师

本地级

●协调数据管理、数据管理制度和数据治理工作。 组织最高管理者

首席数据官

数据治理委员会

组织级

高级数据管理人员

部门治理委员会 子项目指导委员会

部门级

协调型数据管理专员

数据质量分析师 数据库管理员 数据安全管理员

元数据管理员 数据模型管理员 数据整合架构师

数据仓库架构师

数据整合架构师

商业情报分析师

数据治理体系的业务数据管理专员、其他业务主体域专家

数据战略、数据管理、数据质量、数据操作、数据架构、安全与隐私

**图11-1** **云端数据治理组织架构图**

(3)部门治理委员会和项目指导委员会，职责包括：

●实施组织级数据治理委员会的决策；

●负责体系在本部门的组织协调工作；

●对云端数据治理进行现状调研、评估、差距分析与规划；

● 流程梳理和设计；

●制订最佳治理流程规程和管理规范，及有关的技术方案；

●组织相关人员建立和维护体系过程；

●推动组织体系的实施，对相关过程提供过程培训和辅导；

●策划并监督体系的内部实施活动；

●促进组织过程持续不断改进；

●数据管理专业人员的招募和配备；

 **■云端数据治理■**

●数据管理技能开发；

●承包商管理；

●预算和资源分配；

●管理度量指标的制订、监督和评价；

●跨业务和IT 组织的协作；

●支持云端数据治理体系所需的组织和文化变革。

(4)数据管理是业务数据管理专员和数据管理专员的共同职责。业务数据管理

专员是组织数据资产的受托人，数据管理专员是这些资产的专家型保管人。 业务数据管理专员，职责包括：

●具有业务专家和兼职的云端治理管理体系职责的双重角色；

●识别和定义本领域的组织信息需求；

●进行其所负责的业务领域内的云端数据治理工作； 数据管理专员，职责包括：

●负责管理、监控和执行数据政策、标准和规程；

●协调、维护和实施数据架构；

●提供数据管理服务；

●通过实施数据库和应用程序推动数据交付；

●获取和保护数据资产；

●监控数据质量；

●审计数据质量和安全性。

数据管理专员可能包括数据架构师、数据分析师、数据建模师、数据质量分 析师、数据库管理员、数据安全管理员、元数据管理员、数据模型管理员、数据 仓库架构师、数据整合架构师、商业情报分析师、数据整合开发人员和分析报表 开发人员。

一些组织配备高级数据管理专员、协调型数据管理专员和业务数据管理专员。 高级数据管理专员通常是组织级和部门级委员会中的高级管理人员；协调型数据管 理专员在跨团队讨论和同高级数据管理专员讨论时，领导并代表业务数据管理专员 团队；业务数据管理专员是受认可的业务领域专家，与数据管理专业人员一同工作， 持续定义和控制数据。

**2.** **任命管理者代表**

云端数据治理体系管理者代表应是组织级数据治理委员会中的一员(可以兼 职)。管理者代表由首席数据官任命，在云端数据治理体系范围内，可直接代表首席 数据官协调、指导工作。

云端数据治理体系管理者代表的职责包括：

●建立、完善和推行云端数据治理体系，组织编制、修订、审核云端数据治理

**第11章■云端数据治理体系实施**

体系文件、开展实施纠正、预防措施活动；

●向首席数据官和组织级数据治理委员会报告组织目标的实现业绩及体系的 运作情况，作为云端数据治理活动改进的基础；

● 帮助理解业务信息需求；

●帮助部门管理者识别和任命业务数据管理员；

●管理和协调解决云端数据治理活动，例如云端数据质量问题；

●向业务部门或多个职能区域报告数据治理情况，确保体系的各个角色的一致性；

●组织各级人员培训，使全体员工意识到云端数据治理体系的重要性；

● 与云端数据治理体系有关事宜的外部联络。

**11.1.4** **制订体系建设和运行计划**

云端数据治理体系的建立和运行推进计划见表11-1。

**表11-1** **云端数据治理体系建设和运行计划表**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 阶段目标 | 项 目 | 工 作 内 容 | 负责人 | 日期 |
| 建立云端 数据治理体 系体系文件 | 学习云端数据治 理相关知识，统一 意识，建立组织架 构，提供资源 | 组织各部门负责人学习云端数据治理相关知识，统 一意识，明确公司导入云端数据治理体系的意义和必  要性 |  |  |
| 批准《云端数据治理体系建立、运行推进计划》 |  |  |
| 管理《云端数据治理体系建立、运行推进计划》的 实施过程，评价计划的实施结果 |  |  |
| 建立公司云端数据治理组织架构，批准《云端数据治 理体系组织架构图》,提供体系建立、运行的各项资源 |  |  |
| 编制、审核云端 数据治理体系文件 | 制订、批准公司的云端数据治理战略和目标 |  |  |
| 明确公司所有职能机构的职责和权限，确定《云端 数据治理体系组织架构》 |  |  |
| 审核公司云端数据治理战略、目标、组织架构 |  |  |
| 批准公司云端数据治理战略、目标、组织架构 |  |  |
| 颁布公司云端数据治理战略、目标、组织架构 |  |  |
| 组织学习培训公司云端数据治理战略、目标 |  |  |
| 编制云端数据治理策略 |  |  |
| 审核公司云端数据治理策略 |  |  |
| 批准公司云端数据治理策略 |  |  |
| 颁布公司云端数据治理策略 |  |  |

 **■云端数据治理■**

续表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 阶段目标 | 项 目 | 工 作 内 容 | 负责人 | 日期 |
| 云端数据 治理体系的 运行 | 云端数据治理体 系的运行 | 组织培训学习公司云端数据治理策略 |  |  |
| 选择业务问题进行云端数据治理试点 |  |  |
| 覆盖公司所选试点相关的职能机构和云端数据资  产，识别云端数据治理所需要的过程，并明确过程的  顺序和相互作用，确定《云端数据治理路线图》 |  |  |
| 制订和修订云端数据治理文件，包括策略和规程 (例如，数据需求管理制度、元数据管理制度、元数  据管理流程、数据质量管理制度) |  |  |
| 审核公司云端数据治理策略和规程 |  |  |
| 批准公司云端数据治理策略和规程 |  |  |
| 颁布公司云端数据治理策略和规程 |  |  |
| 组织培训学习公司云端数据治理策略和规程 |  |  |
| 实施云端数据治理 |  |  |
| 记录、修订、完 善体系文件 | 在云端数据治理体系运行过程中发现云端数据治  理体系文件存在的不足时，由相关部门记录，汇报项  目数据治理委员会，组织修订和完善 |  |  |
| 检查体系运行的 状况 | 对云端数据治理体系的运行(实施效果、符合性、 有效性)进行一般性监视和测量 |  |  |
| 云端数据 治理体系运 行的监督和 评估 | 绩效评估 | 依据云端数据治理策略和规程要求，对公司的工作 流程和云端数据治理过程绩效进行统计和分析 |  |  |
| 内部审计 | 依据相关的体系文件、云端数据治理制度、法律法 规等，验证公司云端数据治理体系的适宜性、充分性  和有效性 |  |  |
| 安全审计 | 依据安全策略和规程要求，验证云端数据和信息资 产的认证、授权和访问等过程的符合性 |  |  |
| 云端数据 治理体系运 行的改进 | 改进计划 | 制订体系的改进计划 |  |  |
| 改进计划的实施 | 对于绩效评估和审计活动所发现的问题提出改善 意见，制订对策，持续改善 |  |  |

**11.1.5** **形成体系文件**

组织需要制订适用于数据治理所有过程的数据策略，这些策略需要涉及云端数 据治理的各个方面，让各部门“有法可循”,从而帮助组织建立完善的云端数据治理

**第11章■云端数据治理体系实施**

机制。在流程方面，组织需要制订云端数据治理的一般流程，规范从云端数据需求 定义到云端数据质量检测和评估的一系列流程的具体操作。基于策划的结果，再结 合组织已有的管理文件或其他管理体系，形成组织的云端数据治理体系文件。

组织应对文件清单进行评审，并确定每个文件后期编写和维护的责任方。文件 正式开展编写前，应对所有文件的责任方开展文件编写的培训，介绍体系文件的知 识及如何编写文件。体系文件的编写主要包括以下内容。

**1.** **制订数据策略**

数据策略是云端数据治理的基本法规和行动准则，这些策略贯穿云端数据治理 体系的所有过程，描述了什么可以做，什么不该做。数据策略不仅为协调云端数据 治理体系有效运行提供了有效手段，也为云端数据治理体系的评价和审计提供了依 据。数据策略的制订一般包括以下内容。

(1)从组织的自身需要出发制订数据策略

不同的组织由于服务类型、IT 技术、组织规范、机构设置、管理方式、用户要 求和组织文化等的不同，云端数据治理体系存在着比较大的差异。数据策略应当如 实反映组织的特点，能够反映出组织对影响因素做出的系统的控制安排，并明确控 制的重点所在。

(2)从总结组织在云端数据治理经验的角度出发制订数据策略

既然制订数据策略的目的在于提高云端数据治理水平，那么在制订数据策略过 程中就应对以往云端数据管理工作加以总结，把以往先进的、成功的经验有机地结 合起来。

(3)从利用现有管理标准和工作标准角度出发制订数据策略

数据策略的制订是以数据规程文件为基础的，但数据策略不可能包罗一切。数 据策略常常涉及到所有方面，但是具体细节在规程文件中规定。

(4)从让数据管理专员积极参与的角度出发制订数据策略

制订数据策略最重要的事是便于实施，而数据策略是由数据管理专员实施。所 以在制订数据策略过程中应倾听组织内部全体数据管理专员的意见，并充分发挥其 积极性和创造性，

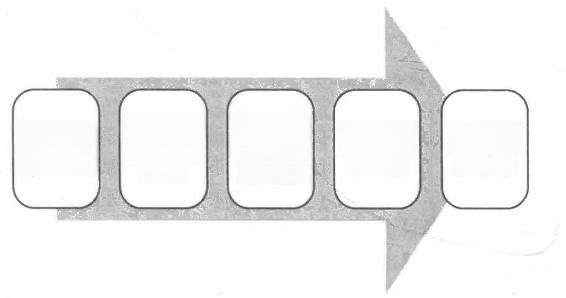
**2.** **制订数据规程**

规程文件提供了一个完成任务的基本平台，将日后成功的经验与失败的教训、 改进发现平台中的弱点和缺点的方法、新的创新技术融入规程文件中，并不断改进。 编写规程文件的目的是指导数据管理专员高效、高速、高质完成任务。每一项数据 治理任务是许多活动或过程的组合，它的数据流、信息流、人流、资金流需要经过 多个职能部门。

在云端数据治理体系中，数据规程是云端数据治理的程序，是完成某项具体

**■云端数据治理**

治理活动的详细步骤。编写数据规程文件应由过程中技术含量高、实施复杂、工 作量多的职能部门主管主导，由部门人员协助，以过程为主线，编写规程文件。 编写数据规程文件应从组织整体的视角出发，解决过程与职能之间的矛盾，理顺 过程中各个活动之间的流通节点。数据规程制订步骤如图11-2所示。

识别完成数 据治理任务 需要的过程

确定数据治

理所需过程

的排序及相

互作用

确保过程有

效运行和控

制所需的准

则和方法

确定可以获 得的必要资 源和信息

监控、测量

和分析数据

治理过程

**图11-2** **数据规程制订步骤**

云端数据治理体系的数据规程包括数据安全规程 、 数据管理控制规程 、 元数据 整合和使用规程、数据模型管理规程、数据质量评估规程等。

**3.** **定义绩效指标**

绩效指标必须与组织的云端数据治理体系战略和目标紧密联系。云端数据治理 体系的关键绩效指标主要分为两类：业绩指标和行为指标。业绩指标是对云端数据 治理效果的综合评价。业绩指标可量化，属于定量性指标。行为指标是影响云端数 据治理效果和基本职能的因素。行为指标侧重于管理行为的过程控制，行为指标属 于定性指标。业绩指标和行为指标相结合能够较完整反映评价对象的全貌。例如， 元数据管理的业绩指标包括元数据质量 、 主数据服务数据合规性 、 元数据文档质量 、 元数据管理成熟度、元数据存储库可用性等。数据管理专员的行为指标包括云端数 据策略和数据规程的符合性等。

**4.** **文件评审与发布**

组织体系文件在经过修改和评审后，如达到发布标准，可以根据组织本身的文 件管理程序 ， 正式发布 。

**第11章■云端数据治理体系实施**

**11.2** **构建和运行**

**11.2.1** **体系宣贯**

体系文件生效后，文件编制负责人必须不断地进行沟通、教育和推广云端数据 治理体系的重要性以及数据治理的业务贡献，提高利益相关者对云端数据治理体系 的效益意识和认可度。可通过以下三种方法对云端数据治理体系进行宣传贯彻。

(1)在内部网站或者公告栏发布公告；

(2)通过邮件或者信息分发相关信息；

(3)对体系内所有人员进行培训。

培训要取得成效，重点要把握以下要点。

(1)制订适宜的培训计划。体系内的所有员工必须进行云端数据治理体系知识 及数据治理意识普及教育，内容主要包括战略、目标和云端数据策略。根据云端数 据规程的要求，对影响规程执行的有效性的员工进行重点培训。

(2)与被培训员工不断沟通，收集培训反馈信息。每次培训结束后可评估培训 效果，对培训内容本身的可实践性、对员工工作绩效的影响程度进行全面的评价， 为以后的培训奠定基础。

(3)形成培训的相关记录，利用记录分析和归纳培训的有效性。

**11.2.2** **实施策划**

云端数据治理体系是识别、管理和解决云端数据相关问题的一种手段。组织可 以从对特定问题或领域进行数据治理的紧迫程度以及数据治理能够带来的价值来综 合衡量，对排名靠前的问题或领域优先进行试点云端数据治理体系验证。这样能充 分获得业务职能部门以及 IT 部门的支持，从而保证云端数据治理体系的成功实施， 等成功后再考虑扩展至其他领域。

云端数据治理的试点实施可以借鉴项目管理方法论，在有限的资源约束下，运 用系统的观点、方法和理论，对试点领域涉及的云端数据治理体系全部工作进行有 效的管理。即从项目的投资决策开始到项目结束的全过程进行计划、组织、指挥、 协调、控制和评价，以实现项目的目标。

云端数据治理实施项目包含筹建云端数据治理制度，以及不断优化云端数据治 理制度的过程。因此云端数据治理实施的过程包含了两个层面的工作： 一个层面是 具体的云端数据治理体系文件的实施过程；另一个层面是把成功实施的制度转化为 日常工作，并且进行持续改进的过程。

 **■云端数据治理■**

组织在实施云端数据治理项目时，须同步调查各阶层员工(如数据管理人员和 业务部门利益相关者)对云端数据治理体系的了解及掌握程度，查看公司内部审计 资源能否达到公司发展需求。

**11.2.3** **项目启动**

在项目启动阶段，组织应基于过程模型，确定云端数据治理项目目标、范围、 进度和质量要求，并制订项目总体计划。同时，基于项目要求，挑选合适的项目组 成员，成立项目组。组织可召开项目启动会，正式启动项目，向相关领导介绍实施 方法及总体计划。

**1.** **确定项目目标**

实施云端数据治理项目的目标是验证云端数据治理体系的统筹和规划，包括体 系目标、云端数据策略和规程等策划的适宜性、充分性和有效性。

例如，某组织的云端数据治理项目目标包括以下内容。

(1)数据质量：整合数据源、创建一致的数据副本、交互提供数据或整合数据、 实现可定义的数据质量检查和维度分析及问题跟踪。

(2)数据架构：消除数据的不一致性，定义标准的业务数据词汇表，有利于数 据共享。

(3)数据安全：定义数据安全标准、制订数据安全措施、管理访问权限等。

**2.** **确定项目范围**

云端数据治理的核心是为业务提供持续的、可度量的价值。组织可围绕一个 特定的业务问题定义云端数据治理的范围，对业务问题的识别和管理过程包括以 下内容。

● 识别、采集和记录业务问题；

●跟踪问题，寻找问题的根本原因；

●记录利益相关者的观点；

●获得云端数据治理委员会对方案的授权；

● 保留所有的相关记录。

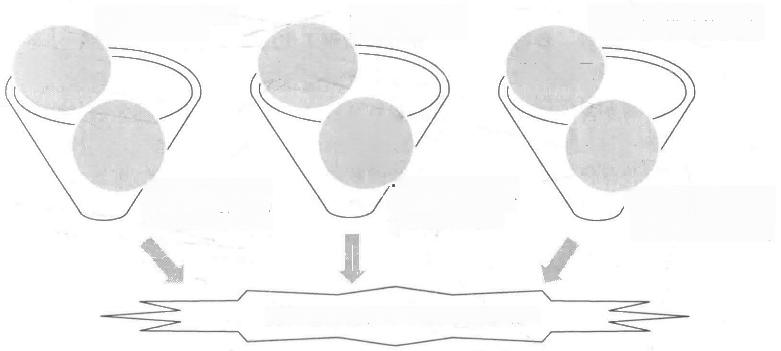
云端数据问题可能包括数据质量问题、业务规则冲突、数据命名和定义冲突、 数据安全问题、合规性问题等。当数据治理实施开始解决已识别的问题时，业务职 能部门便可以将数据治理范围扩展到更多区域。

例如，某银行的客户服务成本超出预算，需要对主数据进行整合。银行的呼叫 中心希望通过客服人员平均处理时间、自助服务和电话首问负责制等机制来降低运 营成本。数据治理部门可以对呼叫客服人员的呼叫记录和工作记录进行分析，获得

**第11章■云端数据治理体系实施**

趋势数据，如与投诉相关的高比例呼入电话，并将它与客户主数据进行整合，从而 获得呼叫者的属性。

例如，某收费系统的云端数据治理项目中识别的业务问题，如图11-3所示。



·缺少可控的数据管理系统 ·未建立数据管理操作指南

数据管理基

础设施不足

管理对象

内容不足

/缺乏详细的物理模型 ·缺乏真正的逻辑模型 ·缺乏当前的数据架构

复杂的数据结构导致系统效率和质量低下

·缺少数据管理组织 ·没有培养专业人才 ·沟通协调不畅通

.缺少数据智能专家 ·缺少数据建模人员 ·缺少数据管理专家

·未使用购机标准模型 ·模型拓展性能差

·缺少数据标准化 ·消极的管理心态

缺少专业 的数据管 理组织

核心模型的 质量不佳

欠缺数据 管理体系

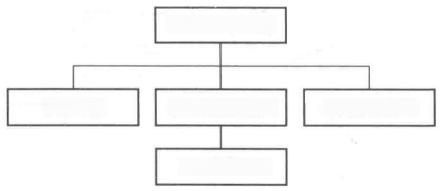
缺少数据 相关技术

**图** **1** **1** **-** **3** **某系统识别的业务问题**

**3.** **构建项目组织架构**

在项目管理中“人”的因素极为重要，因为项目中所有活动均是由人来完成的。 如何充分发挥“人”的作用，对于项目的成败起着至关重要的作用。建立适宜有效 的项目组织架构，是为了有效发挥每位参与项目人员的作用，让项目的所有相关人 员能够在可控状态下有条不紊地进行项目相关活动。云端数据治理体系是确保对数 据资产进行有效控制和使用的管理体系。虽然数据管理专员可以在各自的业务领域 保护和管理数据资产，但是建立项目组织架构，可以确保数据管理专员以整个组织 的视角来保证组织数据质量及其有效利用。

结合组织云端治理管理体系的管理架构，项目组织架构包括三层，如图11-4所示。



项目指导委员会

项目执行负责人

项目实施团队

项目质量管理

技术专家

**图** **1** **1** **-** **4** **项** **目** **组** **织** **架** **构** **图**

(1)项目指导委员会，职责包括：

●实施数据治理委员会的决策；

 **■云端数据治理**

●关注项目进展，在项目重要里程碑处审查项目状态；

●批准影响到项目计划的主要变更；

●协调解决本项目的重大问题；

● 推动、监管项目运作，确保项目能运作成功；

●协调数据管理、数据管理制度和数据治理工作；

●预算和资源分配；

●项目度量指标的制订、监督和评价。 (2)项目执行负责人，职责包括：

●对项目进行现状调研、评估、差距分析与规划；

●制订最佳治理流程规程和管理规范，及有关的技术方案；

●制订项目进度计划，监控项目进展；

● 领导项目成员达到目标；

●识别并控制本项目实施过程中的风险；

●管理项目范围和变更控制处理；

●制订并执行项目沟通计划；

●提供变更材料；

●数据管理专业人员的招募和配备。 (3)项目质量管理，职责包括：

●监督检查本项目的实施过程；

●对发现的实施中的问题，提出改进建议并督促问题的解决；

●必要时，向项目指导委员会汇报工作开展的问题；

●向项目实施组提供实施过程中的问题汇总分析结果。

(4)技术专家可以是外聘的或者组织内的高层数据管理专员，职责包括：

●为项目组提供行业和技术支持；

●对项目执行提出指导性意见。 (5)项目实施团队的职责包括：

●负责项目在本部门的组织协调工作；

●组织相关人员建立和维护组织数据策略和数据规程等文件要求；

●推动组织数据治理的实施，对相关过程提供过程培训和辅导；

●策划并监督本项目的内部实施活动；

●识别并控制本项目实施过程中的风险；

●促进数据治理过程持续不断的改进。

项目实施团队，建议由参与云端数据治理体系的部门，如体系管理部门、IT 部 门、财务部、销售部、营销部、风险部等相关人员组成。典型的 IT 参与者可能包 括信息管理团队、组织数据架构师、记录管理团队以及安全和隐私专业人员。业务 参与者可能包括来自需要使用云端数据来制订业务策略、性能分析、生成报告、开

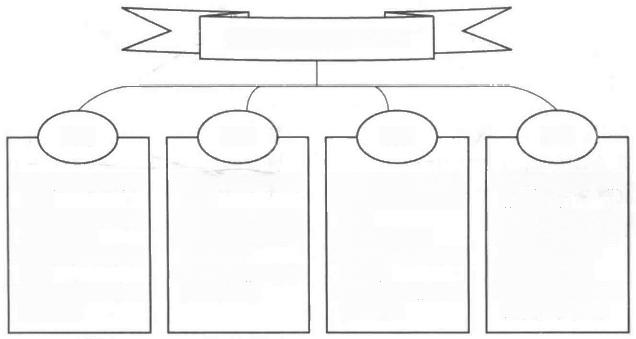
**第11章■云端数据治理体系实施**

发模型、设计业务流程和管理数据的部门代表。

**11.2.4** **项目实施**

**1.** **制订路线路**

技术路线图作为一种云端数据治理项目规划工具可以帮助组织确定数据治理项 目需要的技术能力和资源需求，制订项目实施计划，从而确保技术和资源在需要时 可以及时获得。技术路线图通常以时间规划的方式将项目规划、项目实施、资源利 用等结合起来，标明并重点关注最困难的治理问题以便及时展开有针对性的增强活 动，检查并解决项目中最需要解决的问题。对于云端数据治理，路线图就是关于人 员、过程、技术方案和工具的短期和中长期计划，人员、过程、技术方案和工具的 确定依据正是执行差距分析的结果。图11-5所示为数据治理差距分析报告部分结果。



数据治理差距分析部分结果

技术

·数据管理平台功能 存在缺陷

·数据加密设备陈旧 ·数据清洗工具效率

低下，易出错

·网络带宽不足

·存储设备老化，存

在数据丢失现象

·数据清洗能力不足

·数据安全管理知识 有限

·数据标准缺乏

·数据管理平台操作 不熟练

·主数据管理能力和 技术不足

·数据治理委员会不 完善

·数据安全管理员能 力需要提升

·需要任命新的数据 主管

·部分人员数据治理 意识缺乏

·数据治理策略尚不 完整

·业务案例不够细化 ·数据治理绩效指标

不够明确

·缺少数据治理持续 提升的计划

过程

人员

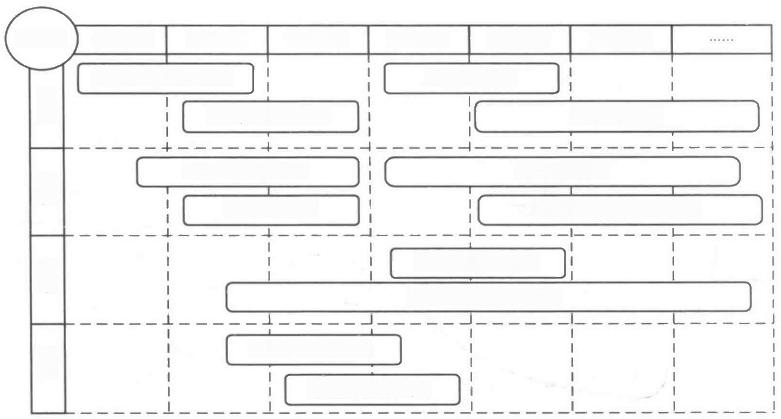
工具

**图11-5** **数据治理差距分析报告部分结果**

图11-5中，组织的云端数据治理体系存在一些不足。例如，在人员上，存在数 据治理委员会缺少必需的成员、数据治理人员能力需要提升等问题；在过程上，数 据治理策略不完整会导致部分数据治理活动没有可以遵循的规则、绩效指标不明确 难以支撑数据治理绩效评估等问题；在技术上，存在缺少必要的数据质量提升能力、 数据安全管理知识不够等问题；在工具上，缺少一些数据治理活动所需的工具和资 源等问题。

针对以上问题，组织建立的路线图必须能够覆盖解决上述问题的流程，例如， 组建云端数据治理工作组，选择合适的职位和角色组成数据治理委员会，然后在数 据治理委员会领导下开展后续工作。图11-6所示为了一个云端数据治理项目实施的 技术路线图示例。

**■云端数据治理**



路线图 第一个月 第二个月 第三个月 第四个月 第五个月 第六个月

完善数据治理委员会

人员能力培训

数据治理活动

数据治理活动测量和评估

数据清洗技术培训

数据安全管理知识培训

升级数据管理平台

数据相关设备更新

完善数据治理策略

制定业务案例

数据治理培训活动

人员能力评估

人 员

技 术

工 具

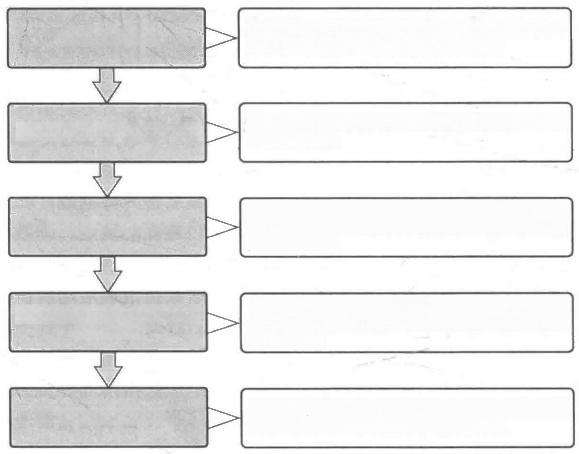
过程

图11-6 云端数据治理项目实施的技术路线图示例

**2.** **定义业务词汇表**

业务词汇表是组织数据基础的核心，是组织开展云端数据治理的基础支撑，是

组织对同一基础词汇的共同认识。业务词汇表的作用是服务组织的数据治理活动， 因此其虽然来源广泛，但应当与组织业务和数据治理活动紧密相连，有时是业务关 键元数据的逆向抽取结果。业务词汇表的定义和管理过程如图11-7所示。



确定数据治理的范围，有针对性地发起业务词汇 表制定，例如业务部门、财务、风险等

安排专人负责组织业务词汇表的修订工作，以及 日常管理工作

对业务数据定义的范围进行确定，例如业绩的统 计时间截止日、合同采购金额大中小的定义、客 户类型的定义

建立业务词汇表对元数据创立的支撑作用，使元 数据、主数据、数据治理、业务目标等以业务词 汇表为基础

依据云端数据治理成熟度模型中业务词汇表成熟 度等级要求对业务词汇表管理进行评估

建立业务词汇表与组织元数 据的映射

确定业务词汇表管理负责人

识别关键数据元素

执行成熟度评估

选择数据领域

**图11-7** **业务词汇表的定义和管理过程**

**第11章=云端数据治理体系实施■**

定义和管理业务词汇表的过程包括以下五个环节。

(1)选择数据领域

业务词汇表是组织数据基础的核心，是组织开展云端数据治理的最基础支撑， 因此业务词汇来源于组织需求。定义业务词汇首先应该选择一类对数据治理需求较 大的业务数据或者其相关部门，在确定一个业务范围或部门内业务词汇后再将业务 词汇来源扩展到整个与数据治理相关的环境，包括组织内部和组织外部。

(2)确定业务词汇表管理负责人

业务词汇表的定义不是一个自动化的过程，是需要有人负责收集相关需求并发 起定义的活动，因此组织需要确定业务词汇表管理负责人来定义和管理业务词汇表。 业务词汇表管理负责人的主要职责包括：识别业务词汇定义需求；确定数据定义范 围；发起业务词汇定义活动；汇总编制业务词汇表；分发业务词汇表；更新业务词 汇表；建立业务词汇表与元数据映射；执行业务词汇表管理成熟度评估等。

(3)识别关键数据元素

数据领域确定了业务词汇的来源范围，但无法确定哪些内容应该定义为业务词 汇，识别关键数据元素就是解决这个问题。例如一份合同中，合同名称、合同金额、 起止时间等都需要明确定义以免引起误解。再例如组织的财务报表，哪种数据应该 会被统计，哪个时间段的数据应该被统计进去等。

(4)建立业务词汇表与组织元数据的映射

业务词汇表的建立主要目的是服务于元数据管理和主数据管理等，因此要保证 业务词汇表对元数据创立的支撑作用，即定义的业务词汇不能是没有价值的或者从 来不会使用的，所以定义的业务词汇表要与组织元数据建立映射。

(5)执行成熟度评估

依据云端数据治理成熟度模型中业务词汇表成熟度等级要求对业务词汇表管理 进行评估，以便及时改进业务词汇表。

表11-2展示了某电子收费系统在数据治理项目中制订的业务词汇表。

**表11-2** **某电子收费系统业务词汇表**

|  |  |  |  |
| --- | --- | --- | --- |
| 序 号 | 来 源 | 术 语 | 描 述 |
| 1 | 处理标志 | 消息接收处理标注编码 | 对于接收到的消息包含0正常入库处理 |
| 2 | 注册时间 | 注册TIMES TEMP | “YYYY-MM-DD HHmm:ss”,注册时间，实际长度19 |
| 3 | 通行券类型 | 通行券类型编码 | 0手撕票，1二维条码，2国标IC卡，3一卡通IC卡 |
| 4 | 争议处理结果 | 争议处理结果编码 | 0争议应对，1争议拒付 |
| 5 | 卡类型 | 卡类型编码 | 0国标卡，1一卡通，2电子标签 |
| 6 | 状态标志 | 允许状态标志编码 | 0禁止，1允许 |

 **■云端数据治理■**

续表

|  |  |  |  |
| --- | --- | --- | --- |
| 序 号 | 来 源 | 术 语 | 描 述 |
| 7 | 客户类型 | 客户类型编码 | 0单位用户，客户编码对应单位用户的客户编码  1个人用户，客户编码对应个人用户表的客户编码 |
| 8 | 回收状态 | 回收状态编码 | 0当天回收，可以继续发行；1注销回收，必须退库到中 心；2其他 |
| 9 | 服务类型 | 服务类型编码 | 0服务费，1退卡购置费入账户，3服务费扣除，9购置 费，10记账卡押金，11挂失费，12挂失费冲正，13挂失 费被冲正，14更换费，15更换费冲正，16更换费被冲正， 99修改充值信息，不更新账户 |
| 10 | 绑定标志 |  | 0不绑定，1绑定，如果绑定，在0016文件的证件类型 是22,证件号是车牌号 |
| 11 | 模块授权 |  | 0不需要授权，1任务模块需要授权，2功能模块 |
| 12 | 主卡标识 |  | 0副卡，1主卡 |
| 13 | 所属标志 |  | 0非系统职工，1本系统职工 |
| 14 | 存储类型 |  | 0预付费，1后付费 |

**3.** **管理元数据**

元数据是关于数据的数据，是最小的数据单位。元数据可以为数据说明其元素或 属性(名称、大小、数据类型等),或其结构(长度、字段、数据列),或其相关数据 (位于何处、如何联系、拥有者),其作用类似于图书馆的书籍目录或中图分类号。图 书馆藏书上百万套，如果没有一个数据系统能够指引读者快速定位到读者所需书籍的 大致位置，读者是无法从百万级书籍中找到自己想要的那本的。图书标识码基于统一 的分类原则，将图书进行分类，采用二级目录便能够快速定位其所在位置。

元数据管理是关于元数据的创建、存储、整合与控制等一整套流程的集合，通 过计划、实施和控制活动以实现轻松访问高质量的整合的元数据。

元数据可以分为五类，业务元数据、技术元数据、操作元数据、流程元数据和 数据管理制度元数据，表11-3列出了这五种元数据的定义和示例。

**表11-3** **元数据的定义和示例**

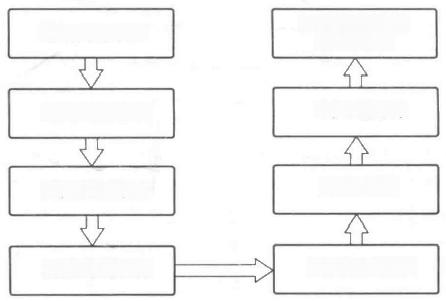
|  |  |  |
| --- | --- | --- |
| 类 型 | 定义或作用 | 示 例 |
| 业务元数据 | 从业务角度描述了数据仓库 中的数据 | 业务数据定义、业务规则、数据模型、数据质量描述、 数据更新周期、历史数据可用性、监管或合同约束 |
| 技术元数据 | 物理数据层的数据资产实例 以及他们通过自动化过程所进 行的转换 | 物理数据库表名和制度名、字段属性、数据库存储 特性 |

**第11章■云端数据治理体系实施**

续表

|  |  |  |
| --- | --- | --- |
| 类 型 | 定义或作用 | 示 例 |
| 操作元数据 | 提供管理信息以管理数据 资产 | 审计控制、数据归档和保留规则、清洗规则、备份和 恢复规则、物理数据模型、抽取的历史和结果 |
| 流程元数据 | 定义和描述数据治理过程的 数据 | 监管机构、数据利益相关者、数据治理组织架构与角 色、流程反馈文档、流程名称、流程顺序 |
| 数据管理制度元数据 | 关于数据管理专员、监督制 度流程和责任分配的数据 | 业务目标、数据增删改减规则、数据管理专员、角色 与职责、数据共享协议、管理与监管机构、治理组织架 构和职责 |

元数据管理的流程包括理解元数据需求、定义元数据架构、开发元数据标准、 构建元数据环境、创建整合元数据、交付元数据和维护元数据库等环节，如图11-8 所示。



执行元数据管理 成熟度评估

定义元数据架构 维护元数据库

开发元数据标准

创建整合元数据

理解元数据需求

构建元数据环境

交付元数据

图11-8 业务词汇表的定义和管理过程

具体的元数据管理流程如下。

(1)理解元数据需求

元数据管理的需求来源于组织业务目标，元数据管理战略必须能够反映对组织 元数据需求的理解，收集需求的目的包括：确认组织元数据管理环境；设定范围和 优先级；教育和沟通；指导工具评估和实施；指导元数据建模；指导元数据内部标 准；指导提供基于元数据的服务，以及预估和验证人员需求。元数据需求一般来源 于两个方面： 一是业务用户需求，二是技术用户需求。业务用户在与组织发生业务 交互时，需要理解组织提供的业务及其背景解释，例如签订的业务合同和服务水平 协议等文件中会涉及一些需要双方都能认可的条款，条款中涉及的专有词汇需要在 业务合同和服务水平协议中清晰地定义。技术用户包括数据库管理员、元数据专家、 架构师、IT 支持人员和开发人员等，通常这些人是组织信息资产的监管人，他们 必须完整地理解数据的技术实现方式，由此产生一些技术用户需求。技术需求一般

**■云端数据治理**

包括每日数据流量、已知和未知的数据源、物理与逻辑架构流程、非标准元数据需 求等。

(2)定义元数据架构

元数据架构定义了组织元数据管理的主要流程和层级，元数据架构应该具备能 够从多个元数据来源定期更新元数据存储库的能力，以支撑元数据管理，进而实现 从多种元数据来源抽取元数据的目的。元数据存储库一般有三种技术架构方法：集 中式、分布式和混合式。集中式架构是指采用单独的元数据存储库保存来自各个元 数据来源的最新元数据的副本，这种架构方式比较占用资源，对IT 资源有限的组织 来说不是很合算，但其优点在于能够帮助组织实现元数据存储库的统一性和一致性。 分布式元数据架构不需要永久的元数据库，只需要维护一个单一访问点，利用引擎 收集用户需求，这种架构可以保证元数据始终是最新且有效的。混合式元数据架构 是一种折中的架构，云数据依然从元数据来源系统进入数据库，但存储库的设计只 考虑用户增加的元数据、高标准化的元数据以及手工获取的元数据，优势之处在于 可以实现准时地获取元数据，以及提供更加完善的元数据信息以更好地满足元数据 用户的需求。混合式元数据架构能够减少定制化系统开发时间，并能满足用户使用 时效性，但混合架构无法提升系统可用性。

(3)开发元数据标准

元数据标准是指导组织制订元数据的依据。随着技术和产品的更新迭代频次加 快，以及数据交换共享更加频繁，数据间共同的语言——元数据管理变得非常重要。 元数据标准能够使数据生产者和用户一起着手处理有关元数据交换、共享和管理的 问题。组织制订的元数据标准应当遵循两个最基本的原则： 一是标准的制订需要参 照通用元数据标准和行业元数据标准，二是元数据标准需要满足业务需求。

成熟的元数据标准包括通用元数据标准和行业元数据标准。通用元数据标准例 如ISO 发布的ISO/IEC11179 系列标准、GB/T18391.1 系列标准、《GB/T30522—2014 科技平台元数据标准化基本原则与方法》等。行业元数据标准是根据各个行业特点 制订的元数据标准化指导规范，例如CWM(Common Warehouse Metamodel, 公共 仓库元模型)、OIM(Open Information Model, 开放信息模型)、MDIS (Meta Data Interchange Specification, 元数据交换标准)、《GB/T19710-2005 地理信息元数据》 等。组织在制订数据元数据标准时应当采用已经被业界认可、面向行业特定需要的 元数据标准，并作为所有元数据管理新技术的原则。

组织制订的元数据库应该是面向组织需求的，尤其是业务元数据和操作元数据， 应该紧密围绕组织的信息治理战略，能够覆盖数据治理体系所有环节的需求。

(4)构建元数据环境

元数据环境是指组织足够重视元数据的管理，在策略、操作指南、人员意识等 各个层级支持元数据管理。组织的数据治理战略应当包括元数据管理目标，并相应 地制订元数据管理指南，经过高层管理者批准后传达给数据治理相关者。

**第11章■云端数据治理体系实施**

(5)创建整合元数据

根据业务和数据治理需要，依据元数据标准创建元数据，做到元数据源于数据 治理并服务于数据治理。添加元数据等操作可以授予用户和程序以手工方式完成， 也可以通过元数据创建和更新工具定期扫描和更新元数据库。整合元数据是将不同 来源的元数据进行汇总存储，元数据来自数据内部，也可能来自组织外部。

(6)交付元数据

交付元数据负责将创建的元数据从元数据库中分发给数据治理组织使用，同时 建立元数据使用反馈机制，及时调整元数据管理策略。

(7)维护元数据

元数据管理是一个持续的过程，组织要根据数据治理的内外部环境变化，调整 元数据管理策略，及时更新元数据库，尤其是当新业务和新技术出现后，更新元数 据以支持数据治理。元数据是数据治理的一个基础，组织保证其正确性、可用性和 安全，元数据的更新应当采用变更管理机制，在变更前评估因此可能带来的影响。

(8)执行元数据管理成熟度评估

依据云端数据治理成熟度模型中元数据管理成熟度等级要求对元数据管理进行 评估，以便改进元数据管理策略和管理流程。

**4.** **管理主数据**

主数据是用来描述组织核心业务实体的数据，例如客户、合作伙伴、供应商、 地点、员工、产品、库存、物料单、账户等；它是具有高业务价值的、可以在组织 内跨越各个业务部门被重复使用的数据，并且存在于多个异构的应用系统中。所以 主数据是在进行组织业务架构分析中发现的核心业务对象，或者认为主数据是组织 已经存在的涉及到价值链核心业务流程的各个IT 系统的基础数据。不同行业的组织 拥有不同类型的主数据，例如对于电信行业组织，电信运营商提供的各种服务可以 形成其产品主数据；对于航空业组织而言，航线、航班是其主数据的一种。对于某 一个组织的不同业务部门，其主数据也不同，例如市场销售部门关心客户信息，产 品研发部门关心产品编号、产品分类等产品信息，人事部门关心员工机构、部门层 次关系等信息。

主数据管理是一套关于如何在组织内创建、整合、维护并使用主数据的活动 集合，通过制订主数据的规范、技术和方案，以保证主数据的完整性、 一致性和 准确性。

主数据管理的目标至少包括以下三点。

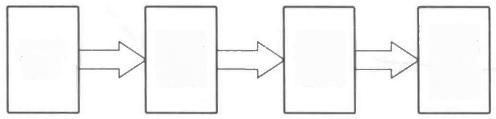
(1)建立集中统一的组织主数据编码规范和管理维护流程，实现主数据全生命 周期的过程管理；

(2)建立支撑主数据管理规范和管理维护流程的主数据管理平台，集中统一管 理主数据编码数据库；

 **■云端数据治理**

(3)为组织和各级单位提供集成、全面、准确和及时的主数据服务和信息化基 础工作的支持。

图11-9所示为主数据管理的大致过程，包括汇集主数据、定义主数据整合架构、 选择主数据管理工具和执行主数据管理成熟度评估。

汇集主

数据

定义主

数据整

合架构

选择主

数据管

理工具

执行主

数据管

理成熟

度评估

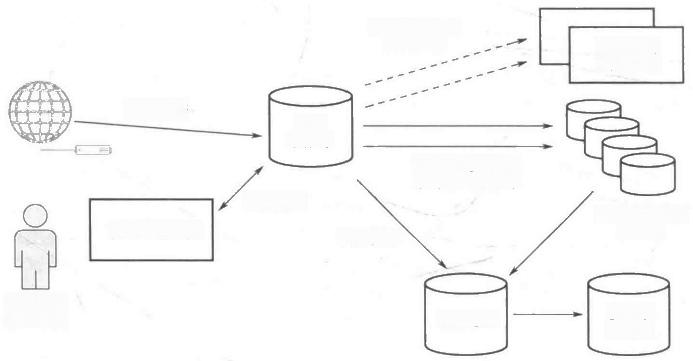
**图11-9** **主数据管理主要过程**

(1)汇集主数据

数据的价值不是靠单个数据就能完全体现的，而是多个关联数据或者看似毫不 相关的数据共同呈现的，主数据汇集就是把处在不同环节、不同部门之间的数据汇 集的过程。以客户主数据为例，对于大多数组织，客户信息通常分散于 CRM 等各 个业务系统中，而每个业务系统中都只有客户信息的片断，即不完整的客户信息， 缺乏组织级的完整、统一的单一的客户视图。因此，组织要对客户主数据进行汇集， 统一视角，通过数据关联对客户进行全面了解，协调统一的市场行为，提升客户满 意度，扩大市场份额。

(2)定义数据整合架构

数据架构是组织IT 架构的核心，因为信息系统支撑下的组织业务运作状况，是 通过信息系统中的数据反映出来的，数据是信息系统管理的重要资源。有效的数据 整合架构能够确保数据质量和数据一致性 ， 相反 ， 如果没有数据整合架构 ， 数据则 是分散的、孤立的。图11-10所示为一个主数据管理架构示例。



其他事物处理 数据系统

订阅更新

记录

数据库

近实时更新联机 事物处理数据库

受控更新

夜间抽取

业务数据

管理专员

联机事物处理 数据库

联机事物 处理系统

多维数 据集市

代码管理系统

数据仓库

图11-10 主数据管理架构示例

**第11章·云端数据治理体系实施**

(3)选择主数据管理工具

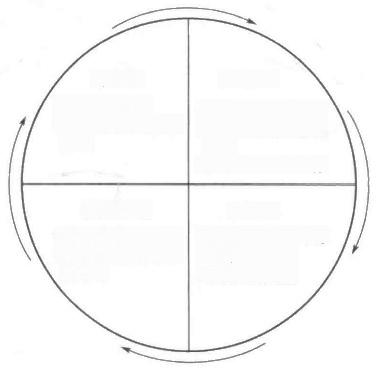
主数据的体量是很大的，尤其是新信息技术在组织业务过程中的应用，管理主 数据不能仅仅依靠人工处理。市场上主数据管理工具众多，组织应当选择适合业务 需要的主数据管理工具，并加以维护。

(4)执行主数据管理成熟度评估

依据云端数据治理成熟度模型中主数据管理成熟度等级要求对主数据管理进行 评估，以便改进主数据管理策略和管理流程。

**5.** **提高数据质量**

数据质量管理是数据治理项目中一项关键的支撑流程，决定了数据决策的准确 性和预见性，数据质量管理循环过程如图11-11 所示。数据质量管理的终极目标是 通过可靠的数据管理提升数据在使用中的价值，并最终为组织赢得收益。

处理Act

·数据质量修复

·数据质量管理方案 优化

检查Check

·数据质量操作监控 ·数据质量指标符合

性检查

计划Plan

·识别数据治理需求

·制定数据质量管理

方案

·定义数据质量指标

实施Do

·评估数据质量问题

·执行数据清洗

·修复数据质量

**图11-** **11** **数据质量管理循环过程**

影响数据质量的因素主要来源于以下四个方面。

(1)信息因素。例如元数据描述及理解错误、数据度量的各种性质(如数据源 规格不统一)得不到保证和变化频度不恰当等。

(2)技术因素。主要是指由于具体数据处理的各技术环节的异常造成的数据质 量问题。数据质量问题的产生环节主要包括数据创建、数据获取、数据传输、数据 装载、数据使用、数据维护等方面。

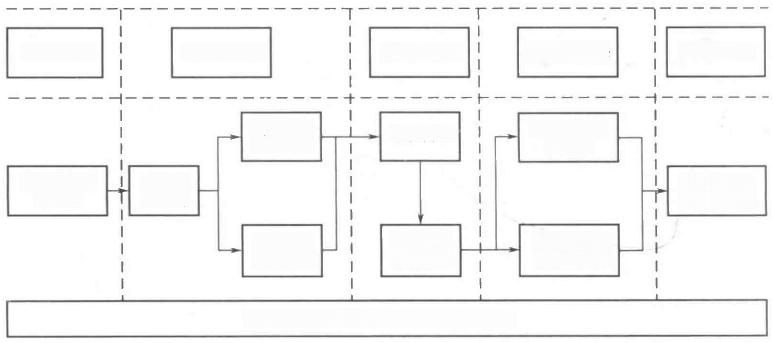
(3)流程因素。是指由于系统作业流程和人工操作流程设置不当造成的数据质 量问题，主要来源于系统数据的创建流程、传递流程、装载流程、使用流程、维护 流程和稽核流程等各环节。

(4)管理因素。是指由于人员素质及管理机制方面的原因造成的数据质量问题。

 **■云端数据治理■**

如人员培训、人员管理、培训或者由于奖惩措施不当导致的管理缺失或管理缺陷。

六西格玛是一种改善质量流程管理的技术，通过定义、测量、分析、改进和控 制过程对质量管理进行持续改进。图11-12 所示为六西格玛管理方法在数据质量管 理中的应用。



定义Define 度量Measure 分析Analyze 改进Improve 控制Control

定义数据

质量指标

数据概

要分析

定义数据

质量规则

沟通行动和结果，循环评估组织管理流程

改进数据质 量管理流程

定义数据质 量需求

数据质量管 理成熟评估

数据质量修 复方案

数据质量 指标验证

数据清洗

**图11-12** **六西格玛管理方法在数据质量管理中的应用**

(1)定义数据质量需求

数据质量需求一般来源于组织的业务需要。例如在电信实名制过程中，电信运 营商发现很多客户信息是不完整的，有些是重复的，缺少关联，因此某业务部门可 以提出如下质量需求。

●客户信息需要包括姓名、身份证号码、性别、住址等；

●同一身份证下通话卡的数量、是否在用；

●用户实名制登记时间、地点。 (2)数据概要分析

数据概要分析可以参考第6章图6-1所示的分析维度，选择工具对数据进行预处理。 (3)定义数据质量指标

数据质量指标主要包括四个方面：完整性、 一致性、准确性、及时性。

●完整性是指数据信息是否存在缺失的状况，数据缺失的情况可能是整个数据 记录缺失，也可能是数据中某个字段信息的记录缺失；

●一致性是指数据是否遵循了统一的规范，数据集合是否保持了统一的格式；

●准确性是指数据记录的信息是否存在异常或错误；

●及时性是指数据从产生到可以查看的时间间隔，也叫数据的延时时长。及时 性对于数据分析本身要求并不高，但如果数据分析周期加上数据建立的时间 过长，就可能导致分析得出的结论失去了借鉴意义。

组织针对不同业务数据，可以细化上述指标，使其具有可操作性。

**第11章■云端数据治理体系实施**

(4)定义数据质量规则

数据质量规则是指导数据质量提高的具体指南，例如数据质量管理的目标、需 求、范围、方法、管理维度、测量基准、计算规则、测量工具等。数据质量预警机 制是数据质量规则一项内容，预警信息的产生和发布需要明确定义的流程，例如，

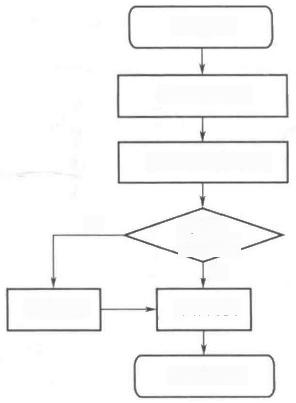
●将满足和不满足业务需求的数据值全部记录下来；

●简要分析满足和不满足业务需求数据比例和主要问题；

●生成通知事件，及时向数据资产管理员警示潜在的数据质量问题；

●发布预警通知，告知业务部门数据质量缺陷。 (5)数据清洗

数据清洗是指发现并纠正数据文件中可识别的错误的最后一道程序，包括检查 数据一致性，处理无效值和缺失值等。组织可以选用数据清洗工具，参考图11-13 所示的数据清洗流程进行数据清洗工作。



原始数据

数据概要分析

定义清洗转换规则

满足清

洗规则之

是

数据清洗

目标数据源

特殊处理

否

**图11-13** **数据清洗流程**

(6)数据质量指标验证

数据质量指标验证是数据清洗结果与定义的数据质量指标进行对比分析的过 程，该过程主要有两个目的： 一是确定清洗后的数据质量是否满足了定义的数据质 量目标；二是对定义的数据质量指标合理性进行调整。

(7)数据修复方案

数据符不符合数据质量指标不是衡量数据价值的唯一依据，只能说明数据暂时 不符合特定指标的期望。数据修复方案主要就是针对不符合数据，提出数据修正的 计划、措施和流程。针对不同的数据质量问题，制订相应的修复方案。

**■云端数据治理■**

(8)改进数据质量管理流程

通过对数据质量进行指标验证，发现数据质量问题原因，可能来自数据标准、 组织与人员、管理流程、技术等环节，制订针对性改进方案，尤其是要改进数据质 量管理流程。

(9)数据质量成熟度评估

依据云端数据治理成熟度模型中的数据质量管理成熟度等级要求对数据管理进 行评估，以便改进数据管理策略和管理流程。

(10)沟通行动和结果，循环评估组织管理流程

保持数据利益相关者持续地参与到数据质量管理的过程中，例如，数据质量管 理策略和流程的制订需要听取利益相关者的意见和建议，并得到批准，数据质量需 求应当包含利益相关者的诉求。数据质量操作过程应当被记录，并作为改进数据质 量管理策略和流程的输入。

**6.** **数据生命周期管理**

数据生命周期管理 (Data Life Cycle Management,DLM) 是一种基于策略的方 法，用于管理信息系统的数据在整个生命周期内的流动，从数据创建和初始存储，

到数据过时被删除。传统数据生命周期管理以成本节约为导向，保存有价值的数据， 而云端数据生命周期管理的目的是以数据的价值和安全为导向，基于云端数据价值 和面临的环境所出发。随着本地数据向云端迁移，加上信息化的普及让云端数据产 生速度更快，云端积累的数据体量越来越大。数据的处理会直接在云端开展，因此 云端数据的全生命周期管理都会与云计算紧密相关，加强数据生命周期管理是必然 的选择。由于云计算载体和云端数据自身的特点，云端数据面临的安全威胁更大， 保证云端数据安全是利用云端数据的前提。

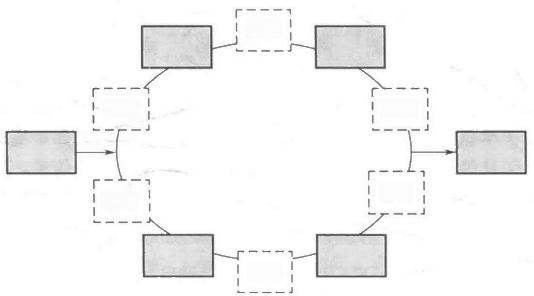
云端数据的生命周期与传统数据的生命周期并没有太大差别，仍然可以将数 据生命周期划分为六个主要阶段，即收集(对于数据生命周期管理)、存储、传 输、使用、共享和销毁，如图11-14所示。在整个生命过程中，数据可能被分析、 存储、更新、使用、转换、整合、传输、归档、共享和评审，最后被销毁。

云端数据生命周期管理应当围绕数据所处的不同生命周期阶段制订相应的管理 策略，确保数据从产生到销毁整个过程被正确地管理，并发挥其价值。下面对数据 生命周期的六个重要阶段进行展开说明。

1)收集

数据采集阶段组织需要关注包括数据采集策略、数据采集规范、数据采集范围 和安全与隐私等方面的内容，如图11-15所示。

**第11章■云端数据治理体系实施**



更新

使用

转换

收集

评审

共享

归档

存储

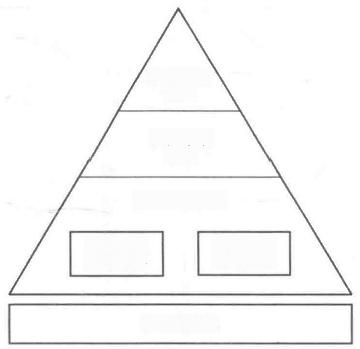
销毁

传输

分析

整合

图11-14 云端数据生命周期

数据采集’

策略

数据采集

规范

数据采集范围

本地数据

数据直接 云端产生

云端迁移

安全与隐私

**图11-15** **数据采集管理策略**

数据采集策略是指导数据采集活动的高级纲领，涵盖了对采集原则、采集要求、 采集目的、采集目标、采集权限等规定。例如采集原则， 一 般有两个方向： 一 是尽 可能多地采集数据，通过处理发现有价值数据；二是依据业务需求进行目的性较强 的采集。前者认为只要与组织相关的数据，都尽可能地采集到云端数据管理平台， 该原则的主要优点是无论如何总能发现对组织有价值的数据，但会消耗过多成本，

同时对数据处理能力提出了很高的要求。后者以需求为导向，只有当业务发出数据 需求时，才会有针对性地对数据进行采集，能够很大程度上节约成本。由于云端数 据采集的便利化，数据无时不刻不在包围着组织，因此组织应当更为积极地采用数 据采集策略，更多地收集数据。

为确保数据采集活动有序地进行，保障不同数据采集源之间的互操作，组织需 要根据数据采集策略制订相应的数据采集规范，包括数据采集方式、数据采集工具、

**■云端数据治理**

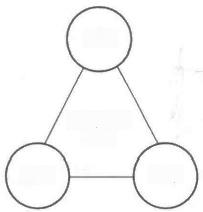
数据采集内容、数据采集依据标准等。

组织为满足业务需求，会广泛地采集数据，对于云端而言，数据可能来自两部 分： 一是本地化采集的数据迁移到云端的，二是组织直接在云端采集到的数据。

数据采集过程是与客户信息关联最紧密的活动，涉及客户与用户的隐私。传统 数据采集活动中泄露的用户隐私多是用户比较直观的信息，例如采集到的用户的身 份信息、财务信息等， 一般不需要进一步处理即可清楚地获知信息。然而由于大数 据的出现，数据关联分析效力体现更明显，原本采集的没有直接涉及用户信息的数 据，可能会因为它们之间的关联推理出用户隐私，因此云端数据采集必须更加注重 用户隐私的保护。

2)存储

随着云计算和云存储概念和技术的不断成熟，组织数据中心正经历着从传统数 据中心向云数据中心转变的新的变革，小型组织通过将数据中心向云存储端转移， 可以很好地控制数据中心成本，而大型组织使用云端进行数据存储能够更好地满足 跨区域业务需求和合规要求。云端数据存储管理需要注重安全、成本和可靠三项原 则，如图11-16所示。



安全

云端数据

存储

成本 可靠

**图11-16** **云端数据存储三原则**

(1)安全原则

云计算架构更加开放，多租户和泛在访问等特点使云计算应用更加方便，但也 给云计算安全带来了新的威胁点。那么存储在云端的数据安全就成了利用云计算时 需要重点关注的内容。数据是组织的一项重要资产，在将数据迁往云端时，组织应 当选择安全性高的云服务提供商，例如通过专业安全认证的服务商。除此之外，组 织要在安全平台上做好自身数据的安全管理，例如访问控制、法规要求等。云端数 据的管理权和所有权是分离的，比如公有云服务方面，是否给供应商提供一些高权 限的管理。在法律风险方面主要是地域性的问题，云计算应用引发了地域性弱、信 息流动性大的特点，在信息安全监管、隐私保护等方面可能存在法律风险。

(2)成本可控

随着数据体量的增加，尤其是依赖数据进行服务的组织，因数据存储需要支出 的成本在集聚增加，通过自建数据中心解决所有问题已不现实，云端存储是降低成

**第11章■云端数据治理体系实施**

本的一个重要途径。成本与服务等级是对应的，在组织选择高质量服务时必须支付 更多的成本，而为了降低成本选择服务质量没那么好的产品时则可能要面临安全和 性能上的折扣，因此组织需要根据业务需求制订对应的数据云端存储方案。降低成 本的一个重要途径是根据数据的活跃周期，为数据存储选择对应的服务级别，对于 活跃数据选择高质量的存储服务，对于处在生命周期后端的冷数据选择成本较低的 存储方案。

(3)性能可靠

数据云端存储，可能会让组织面临几个困扰：数据完整性如何保障、数据上传/ 下载能否满足需求、云服务器宕机时组织业务如何对数据进行调取等，因此云端数 据存储的可靠性是部署云数据中心时组织要重点考虑的原则。

为保证云端数据的可靠性，云平台需要采取冗余方式防止数据丢失，并根据数 据等级，采取不同的备份方式，例如可参照国际标准 SHARE78。

3)传输

传输是数据生命周期过程中必不可少的过程。从传输的范围来分，可以分为处 理系统内部传输和跨系统传输，其中跨系统传输可能包括跨组织内部部门之间传输、 跨组织间传输、跨地域传输等。对于大型组织，数据跨组织传输也可能是跨地域传 输。数据传输涉及传输信道、传输方式、传输设备、传输协议、传输性能等要求， 不同场景下，组织进行数据传输需要采取相应的策略。

4)使用

数据使用应当遵循“合法”、“合理”原则，所谓合法是指数据的使用必须遵守 外部的法律法规制度以及组织内部制订的有关数据使用规则，如数据使用审批、数 据使用过程有记录、数据使用完毕后处理等。所谓合理是指数据使用要根据需要进 行，不能浪费数据价值，也不能过分使用数据，尤其是当数据涉及敏感信息时，应 该采取最小权限和最小使用原则。

5)共享

共享是数据的基本特性特征之一。数据作为一种极具价值的资产，可以重复使 用，但价值体现必须在数据在传输、流通和共享中实现。数据共享遵循几个基本原 则，包括科学性、统一性、扩展性、安全性等。科学性是指数据共享的方式要科学， 例如数据共享格式、共享途径、共享时效等，这些会影响到数据能否满足使用方应 用需求；统一性是指数据分享的方式要统一，例如数据单位应当采用国际或行业公 认的格式；扩展性是指数据共享设计时需要充分考虑数据扩充、时间增量等问题； 安全性是数据共享的又一个最基本的原则，数据只有在安全使用的前提下才可以共 享。云计算的特点给数据流通带来了更加便捷的条件，相对开放的云环境，也更加 强调数据流通的必要性。但与此同时，在数据流通过程中，如何保障数据共享的效 率、安全是组织云端数据治理必须要考虑的话题。

**云端数据治理■**

6)销毁

销毁是数据生命周期的最后一个阶段，是指通过物理或者技术手段使数据永久 或临时不可用的过程。对传统的数据本地化储存，数据销毁往往通过技术处理后直 接销毁存储设施即可完成，是非常彻底的数据销毁方式。但在云计算环境下，数据 能否彻底删除是很多云服务使用者比较担忧的一个焦点。云端空间是可移动的，并 会频繁创建快照，还会定期进行备份，而且数据会在好几个数据中心间共享，这样 可以节约成本，提高可用性。数据可能会被数据所有者移动，或者是在公共云里被 服务供应商移动，这样才能在多租户环境中优化资源的使用情况。当数据被移动时， 原本位置的数据应该要销毁。如果有任何数据残留，就有可能产生安全问题，也可 能出现未经授权访问残留数据的问题。组织存储在自建云服务器上的数据，自己负 责数据的销毁，对于租用云服务，组织应当在服务等级协议中明确云服务商在数据 销毁阶段的责任，并在必要时，明确对于残留数据管理的主体责任。

**7.** **数据安全管理**

数据已经成为组织甚至全社会的重要资源，因此也成为了被“关注”的显著目 标，尤其是大数据的到来，数据所蕴含的价值已经超过了其自身的价值。虽然作为 一种新的技术，云计算给数据处理能力带来了巨大的变化，但云计算特点，如泛在 网络访问、快速伸缩弹性、多租户等，也给承载在云平台上的数据带来了新的安全 威胁和漏洞，因此加强云计算环境下数据安全管理是云端数据治理相对于传统数据 治理的一个很大区别点。数据安全管理是围绕数据整个生命周期而实施的一系列措 施，包括技术上和管理上的手段。

在云端数据治理体系中，数据安全管理是每个环节都必须要考虑的。在制订数 据战略时，数据安全应该是一个重要部分，并在数据治理组织与角色中任命数据安 全管理负责人和数据安全管理专员。在日常的数据生命周期管理过程中，组织需要 采取相应的措施保证数据安全。

在数据收集阶段，组织需要特别关注数据分类、数据采集、数据清洗、数据质 量等环节中的安全管理需求。例如，组织应对不同类别和级别的收集数据实施不同 的安全管理策略和相应的保障措施，并对数据分级分类变更的操作记录进行审计。 在数据收集时，定义采集数据的目的和用途，明确数据采集源和采集数据范围，以 满足合规要求。

在数据存储阶段，组织需要重点关注数据中心安全和数据自身安全。对于自建 数据中心，要关注数据中心的物理安全、支撑性资源安全和系统安全。物理安全是 数据安全的基础，主要包括物理位置的选择、防火、防水、防雷击、防盗、防破坏、 电磁防护等。支撑性资源安全主要包括水、电、通信等基础设施安全。系统安全主 要指存储系统的安全，可以采取用户身份鉴别、资源访问授权等方式控制对存储系 统造成的潜在威胁。对于租用其他云服务，组织需要在服务协议中明确各方责任，

**第11章\*云端数据治理体系实施**

告知组织对云端数据安全的关切。除此之外，组织自身仍然需要采取必要的措施保 护数据的安全，如数据加密、数据备份等。

在数据传输阶段，组织要保证数据在传输过程中的安全，涉及的主要措施包括 数据的加密和数据网络安全控制等。数据加密包括硬件加密和软件加密，硬件加密 是通过专用加密芯片或独立的处理芯片等实现密码运算，例如采用AES128 位或256 位硬件数据加密技术对产品硬件进行加密，具备防止暴力破解、密码猜测、数据恢 复等功能。软件加密就是用户在发送信息前，先调用信息安全模块对信息进行加密， 然后发送，到达接收方后，由用户使用相应的解密软件进行解密并还原。数据网络 安全控制是对数据传输的通道进行安全防护，主要的安全措施包括网络结构安全、 网络访问控制、网络设备防护等。网络结构安全指组织内部的网络安全架构设计和 组织内网与公网的连接段网络架构。网络访问控制通过控制方法对未授权的访问进 行隔离，如 MAC 地址过滤、VLAN 隔离、访问控制列表和防火墙控制等。网络设 备防护是保护实现网络结构设计和网络访问控制的支持设施，避免这些设施失效而 影响网络安全控制措施。

在数据使用阶段，组织要保证数据正确、合理地使用，要防止数据的泄露和过 度使用。在数据使用阶段，数据安全管理需要更多地依赖管理手段，所以建立完善 的数据安全管理体系是促进数据安全使用的有效途径。

**11.2.5** **项目过程管理**

**1.** **项目质量管理**

项目质量控制的目的是确保项目工作按照计划进行，并达到预定的目标；最终 的体系文件是符合云端数据治理成熟度模型的要求；确保项目能按预定计划完成。

因此，在项目进行中，需要设立若干控制点，主要包括以下六类。

●体系方案质量控制；

●业务质量控制；

●交付物质量控制；

●项目初始阶段质量控制；

● 项目进行中质量控制；

●项目结束阶段质量控制。

质量管理人员是独立于项目组的，作为项目成功的重要因素和保障，在项目执 行过程中始终与项目组保持密切联系。从另一个角度观察项目的开展，帮助项目组 发现和解决项目执行中的问题，确保项目的成功。

在项目实施阶段，项目管理与其他子过程并行执行，它从项目开始持续到项目 完成。它的步骤和任务侧重于项目小组的管理工作。项目管理是一个不断重复的过

**■云端数据治理■**

程，许多任务例如“过程控制”要贯穿于项目的始终，而另外一些任务会因为突发 事件而启动，例如：

● 用户改变需求；

●提前实现需求；

● 用户处于紧急关头；

●实际过程与计划过程有差异；

●标识新的风险。

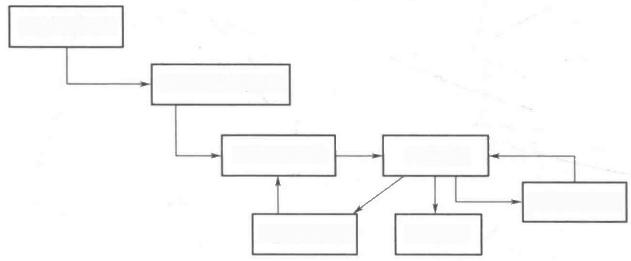
此过程在项目启动和项目结束之间与“实施方案”过程并行进行。它们是两个 不同的方面。

项目管理过程侧重于管理项目计划中定义的步骤，确保实现所有的承诺，及时 向组织业务部门或者高层管理反馈情况，保证项目质量。其中最关键的工作是对计 划外事件的管理(变更、风险等)。

**2.** **项目风险管理**

风险，多指对项目“不利”的不确定因素。这些不利的风险存在于任何项目中， 并往往会给项目的推进和项目的成功带来负面影响。风险一旦发生，它的影响是多 方面的，如导致项目服务的功能无法满足客户的需要、项目费用超出预算、项目计 划拖延或被迫取消等，其最终体现为客户满意度的降低。因此，识别风险、评估风 险并采取措施应对风险(即风险管理)有着十分重要的意义。

项目组将采用图11-17所示的风险管理过程对项目过程中的分析进行管理。



风险识别/记录

定性/定量分享分析

风险计划编制

风险计划落实

风险重新评估

风险监控

风险注销

图11-17 风险管理过程

具体到本项目，我们可以以项目计划为线索，识别项目在各方面的风险。在实 施过程中，应特别关注以下三方面的风险。

(1)项目范围的风险

若项目范围定义不清晰，可能导致各部门对项目范围的认知产生分歧。若双方 的分歧较大，不能达成一致，则必然会造成效率低下，相互推脱。

因此，应对项目的实施范围做尽可能清晰的界定，前期多花一些时间在项目实

**第11章■云端数据治理体系实施■**

施前的范围界定工作上。

(2)项目进度的风险

关于项目实施的周期，目前有时间进度上的紧迫性。但项目进度的控制绝非易事， 不仅取决于项目组的能力，同时也在很大程度上受到各个部门的配合、对范围控制是否 有效、对项目投入(包括人员时间的投入和资金等的投入)是否足够等方面的影响。

(3)项目人力资源的风险

人力资源是本项目实施过程中关键的资源。要降低项目的人力资源风险，就要 保证进入到项目中并承担角色的各类项目者满足项目要求。因此，实施方应对参与 人员进行认真的评估，包括项目组和贯标部门人员的评估。同时，应保证项目人员 对项目的投入程度。

**3.** **项目沟通管理**

沟通要求如下。

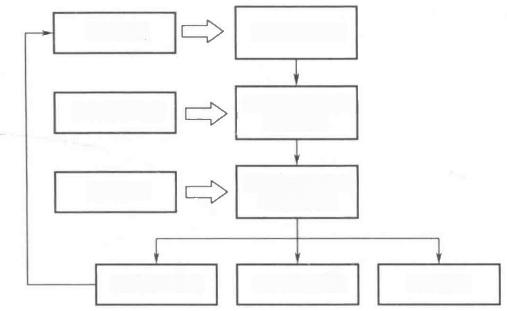
●项目实施团队将治理具体计划和需准备内容提交给项目执行负责人；

●项目执行负责人在项目关键里程碑处向项目领导委员会汇报项目进展情况，

●项目重大变更应得到项目领导委员会的同意。

**4.** **项目变更管理**

项目变更管理控制过程如图11-18所示。



项目计划 建立基准计划

评审项目状态和

进展情况

确认变更的发生

或将要发生

更新项目计划 采取纠正措施 共享教训

执行绩效报告

变更请求

图11-18 项目变更管理控制过程

项目变更审批流程如下。

●提出修改方将首先填写变更申请表(REQUES FOR CHANGE,以下简称RFC)。

●RFC 须提交评审小组确认。评审小组将就 RFC 的技术可靠性以及对整个项 目的影响作出评估。评审小组主席由组织数据治理委员会指定。

●评审通过后，授权并发放实施通知单。

●数据管理专员予以实施变更。

**■云端数据治理■**

**11.3** **监督和评估**

组织应在云端数据治理体系统筹和规划阶段考虑监督和评估的需求。监督和评 估的关键指标应涵盖与云端数据治理目标和实施框架相关的职能和层次，以及技术 和业务流程等相关方面。组织可通过绩效评估、内部审计和安全审计方式对云端数 据治理体系过程进行监督和评估。

**11.3.1** **绩效评估**

云端数据治理过程要素包括人员和流程，二者都是无形的。因此，拥有协商一 致的关键绩效指标(KPI) 集对于监督和评估数据治理体系的运行很重要。组织应根 据云端数据治理关键成功因素和职责确定部门和岗位的绩效指标。绩效评估的结果 应定期传达给数据治理委员会和高层管理人员，从而获得利益相关者的持续支持和 资金支助。绩效评估指标的建立和考核主要包括以下内容：

(1)分析业务整体绩效指标

每个业务部门都有一个 KPI 层次结构来运行其业务。可以用工作职能(比如销 售、营销、财务和风险)来进行描述。理解这些度量指标很重要，因为云端数据治 理的重要目标是使各部门的KPI 提高。

(2)定义数据治理的业务驱动绩效指标

云端治理项目需要制订可操作、可实现、可考量的具体 KPI, 以提高业务的绩 效指标。例如，某家银行的系统在全行范围内无法推广，各个系统对数据标准管理 要求不一致，业务分析人员不能快速找到正确的信息，银行的总体风险是一个业务 整体 KPI。数据治理的业务驱动 KPI 是元数据管理成熟度。云端数据治理项目组应 该使用此 KPI 每月跟踪绩效并向风险团队报告进度。

(3)定义数据治理的技术绩效指标

技术KPI 评估和测量云端数据治理在技术方面的绩效。可围绕以下云端数据治 理职能建立技术绩效指标。

●云端数据管理——主数据和元数据质量、主数据服务数据合规性、元数据存 储库分布、元数据文档质量、管理专员代表和覆盖范围、元数据使用/引用、 元数据存储库可用性；

●文档和内容管理——完整性、可用性、访问控制；

●数据架构管理——数据架构标准(如数据库、报告工具、分析工具等);

●数据操作管理——可用性、性能；

●数据安全管理——数据库漏洞、数据库的 SQL 错误数量、未授权统计。

**第11章■云端数据治理体系实施**

(4)指标考核

① 考核流程

组织应定期考核这些 KPI, 并向数据治理委员会和高级管理人员报告结果。业 务驱动和技术数据治理 KPI 建议每1～3个月评估和统计一次。

② 考核信息反馈

●对考评结果要做到全面剖析，对未达标的工作部分要加以分析，找出原因并 采取措施进行改进。如果绩效指标不适宜，应调整绩效指标和细化工作职责 标准。

●对考核成果要进行充分利用，要及时通过管理人员对有关的责任人进行沟 通。对于考核者存在的不足，应提出具体的建议要求及改进工作的途径，以 保证体系的有效运行。

**11.3.2** **内部审计**

组织应对云端数据治理体系进行内部审计，验证体系是否持续满足规定的要求 并且正在运行。在许多情况下，尤其在小型组织内，应该由与受审核活动无责任关 系的人员进行，以证实其独立性。

组织应制订内部审计规程文件。内部审计规程文件的内容包括目的、范围、定 义、审计的基本要求、审计人员的确定与责任、审计计划、审计的基本步骤、方法 及要求、审计的分析与记录、审计结果的跟踪验证要求等。内部审计规程文件是组 织内部审计各项活动的总指导和规定，可包含体系管理和数据治理过程的审计。内 部审计应对所有过程和部门进行，在规定时间内(通常为一年)覆盖100%;应对所 有云端数据治理过程和因素进行审计，确保数据治理过程符合数据策略和数据规程 文件的要求。

云端数据治理体系数据战略过程域的内部审计检查表参考模板见表11-4。

**表11-4** **数据战略过程域内部审计检查表**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 序号 | 职能域 | 过程域 | 检查项 | Levell | Level2 | Level3 | Level4 | Level5 | 审计  结果 |
| 1 | 数据  战略 | 战略 | Q1:数据治理战略是否完整 |  |  |  |  |  |  |
| Q2:数据治理战略是否得到数据 利益相关者认可 |
| Q3:数据治理战略与组织其他战 略的一致性 |
| Q4:数据治理战略是否能够指导 数据治理活动的开展 |

 **■云端数据治理**

续表

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 序号 | 职能域 | 过程域 | 检查项 | Levell | Level2 | Level3 | Level4 | Level5 | 审计 结果 |
| 1 | 数据  战略 | 战略 | Q5:组织是否具备必要的人员 |  |  |  |  |  |  |
| Q6:组织是否具备必要的资金 |
| Q7:是否制订了对战略进行测量 评审的指标体系 |
| Q8:是否实施了战略的评审与改 进 |
| 2 | 组织与 角色 | Q1:组建的数据治理组织是否能 够支撑数据治理目标和愿景的实现 |  |  |  |  |  |  |
| Q2:数据治理组织与角色是否符 合数据治理战略的要求 |
| Q3:组织与角色是否得到数据利 益相关者认可 |
| Q4:数据治理组织与角色能否指 导数据治理活动开展的所有工作 |
| Q5:数据治理角色是否配备了合 适的职责 |
| Q6:数据治理组织内的角色是否 具有代表性 |
| Q7:数据治理角色是否安全 |
| 3 | 业务  案例 | Q1:组织如何决定数据治理方案 所要求的投资水平 |  |  |  |  |  |  |
| Q2:组织如何决定是要开发一个 总体的业务案例还是多个相互联 系的业务案例 |
| Q3:如何评价业务案例 |
| Q4:需要谁参与其中，需要谁批 准 |
| Q5:业务案例是否反映数据管理 战略的目标和优先事项 |
| Q6:业务案例是否反映数据治理 顺序规划 |
| Q7:业务案例制订是否考虑到资 源保障的能力 |

**第11章■云端数据治理体系实施**

续表

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 序号 | 职能域 | 过程域 | 检查项 | Levell | Level2 | Level3 | Level4 | Level5 | 审计  结果 |
| 4 |  | 资源  保障 | Q1:是否有一套经批准的数据治 理投资准则和重点 |  |  |  |  |  |  |
| Q2:是否能提供稳定充足的资金  保障 |
| Q3:资源保障策略是否由数据利 益相关者制订、评价和批准 |
| Q4:资源保障策略是否具备动态 调整的能力 |
| Q5:资源保障策略是否反映了组 织的业务模型、优先事项和财务决 策过程 |
| 5 | 沟通 | Q1:如何公告数据治理的政策、 标准和过程 |  |  |  |  |  |  |
| Q2:组织如何向利益相关者通知 有关数据治理计划和项目的消息 |
| Q3:沟通战略是否得到数据利益 相关者认可 |
| Q4:沟通的有效性怎么评判 |
| Q5:如何在业务、IT、数据治理 和高管之中实现数据治理优先事 项、方式和工作输出的双向沟通 |

**11.3.3** **安全审计**

云端数据治理体系安全审计的目的是确保云端数据和信息资产的认证、授权和 访问等过程符合体系的相关安全策略和规程等策划要求。

安全审计包括四个基本要素：控制目标、安全漏洞、控制措施和控制测试。其 中控制目标是指组织根据云端数据治理体系的要求，结合实际的云端数据治理过程 而制订的安全控制要求。安全漏洞是指系统的安全薄弱环节，容易被干扰或破坏的 地方。控制措施是指组织为实现安全控制目标所制订的数据策略、数据规程、数据 隐私要求及IT 策略要求等。控制测试是将组织的各种安全控制措施与策划的安全标 准进行一致性比较，评价组织安全措施的适宜性、充分性和有效性。

安全审计是一个专门的审计项目，组织可根据自身的需求不定期开展。安全审

 **■云端数据治理**

计的内容可包括：

●访问权限审计——身份认证和访问行为、数据访问权限、特权用户行为、敏 感数据访问、密码标准等；

●变更控制审计——关键的数据库结构和数据值的变更是否授权等；

●数据安全审计——评估现有标准和规程与业务要求和技术要求是否一致、检 查安全审计数据的可靠性和准确性、外包项目数据安全等。

**11.4** **改进和优化**

**11.4.1** **制订改进计划**

持续改进是为了不断提升组织云端数据治理能力，并致力于实现云端数据治理 体系战略和目标所采取的循环活动。组织可制订云端治理体系的改进计划，实施对 体系的改进和优化。改进计划参考模板如下：

|  |
| --- |
| 云端数据治理体系改进计划  1 改进的背景和总体目标  云端数据治理已经在试点评估项目中得到成功实施，提高了公司的云端数据治理能力。基于自身能力提高的需  要和公司长期云端数据治理战略，公司计划在原有试点项目的基础上扩大云端数据的治理范围，提高公司云端数 据治理效率，保证其他系统治理项目能够有序进行。过程改进范围涉及公司五大部门，计划周期为12个月。  2 工 作 原 则  1)过程改进的工作要为实现过程改进目标服务；  2)过程改进要能解决公司的实际问题  3)改进工作小组首先要自己了解过程改进的意义；  4)过程要简洁、明了，具有可操作性。  3 改 进 目 标  通过过程改进项目，在本公司建立符合云端数据治理管理体系要求的云端数据过程框架，并在本公司所有云端  数据治理项目中应用；能够根据组织建立的云端数据治理战略，对过程进行不断完善和改进；最终目的是规范化 本公司云端数据治理过程，从而提高开发效率。  包括以下内容：  1)差距分析与项目启动阶段(2个月)  主要内容：组建工作组、差距分析访谈、项目启动会议召开。  2)改进策划阶段(4个月)  主要内容：体系培训、制订项目实施计划。  3)试点阶段(4月)  主要内容：云端数据治理试运行项目过程管理、文件制度化并正式发布、体系文件推广与内部培训、运行数据 收集。 |

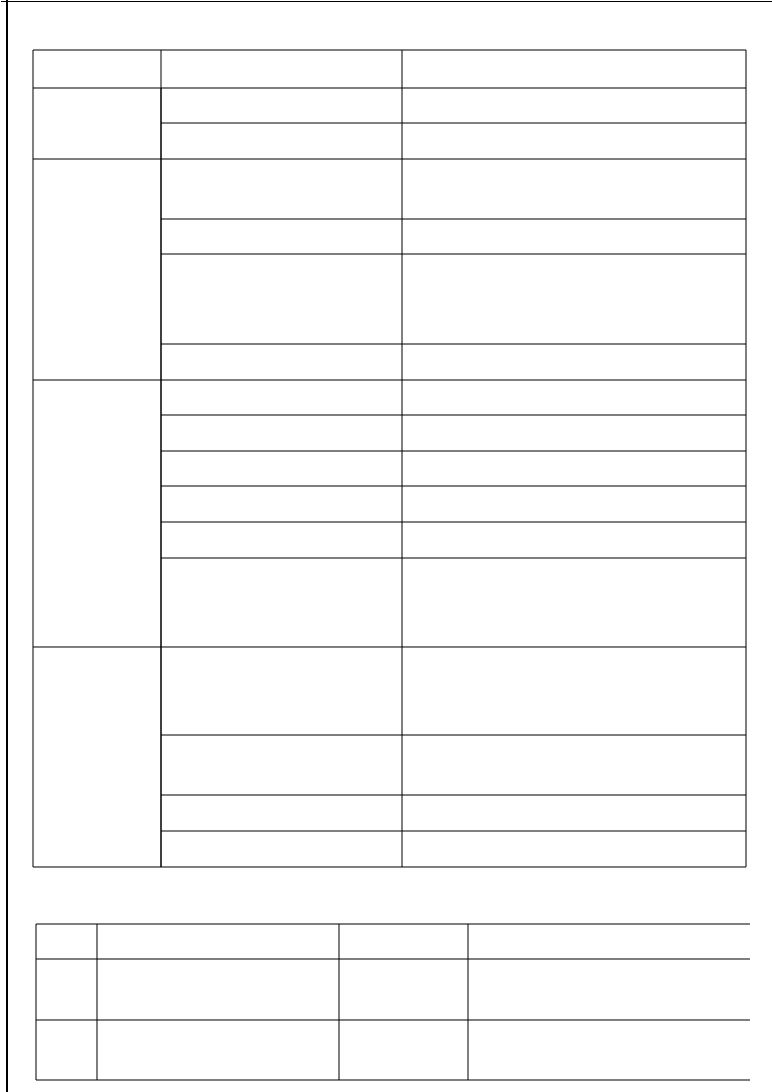
**第11章■云端数据治理体系实施**

续表

|  |
| --- |
| 4)评估阶段(2个月)  主要内容：试点项目过程数据整理、总结。  4改进机构与职责  本次过程改进的项目组的组织架构如下。  项目指导委员会  项目执行负责人 质量控制员  工作组1 工作组2 工作组3  各部分的职责如下：  项目指导委员会：  1)从最上层开始推动过程改进活动；  2)支持过程改进：  ——提供足够的资源；  ——允许项目计划作适当调整；  ——对过程改进执行较好的工作组给予奖励；  3)帮助解决冲突；  4)审批过程改进计划、过程评估报告。  项目执行负责人：  1)制订过程改进计划，并实施过程改进计划；  2)组织EPG工作，制订公司标准过程集ZDSSP;  3)新过程试用、评估，推动ZDSSP的实施；  4)为工作组提供培训和支持；  5)维护过程数据库；  6)阶段性评估过程改进状态，每月进行工作总结，并向最高管理者、管理者代表汇报工作进展。 质量控制员：  1)编制组织级质量保证计划；  2)实施组织级质量保证计划，主要对组织云端数据治理活动、组织培训及相关工作输出进行检查： 3)通过监控过程来保证工作质量；  4)保证云端数据治理过程符合相应标准与规程；  5)保证云端数据治理过程中存在的不符合问题得到处理，必要时将问题反映给项目执行负责人； 6)确保项目组制订的计划、标准和规程适合项目组需要，同时满足评审需要。  工作组成员：  要对自己负责的部分有较深入的研究，与项目执行负责人共同完成该部分过程文件的撰写和推动工作。 |

**■云端数据治理**

续表

5 进度计划

任 务

关键产出物

项 目 阶 段

差距分析

差距分析报告

云端数据治理基础培训教材

准备阶段

基础培训

云端数据治理项目实施计划

制订实施计划

云端数据治理项目实施进度安排

制订专题培训教材

专题培训教材

培训计划

改进策划阶段

专题培训

培训签到表

培训总结

云端数据治理实施流程文件

制订/修改云端数据治理实施流程

确定云端数据治理试点项目

试点实施

试点持续支持

试点总结

试点评审

试点项目列表

试点问题跟踪表

持续支持工作纪要

试点总结报告

试点阶段

试点评审问题清单

过程改进计划

过程改进报告

修订体系文件(如需)

修订后的体系文件

培训通知

培训教材

云端数据治理知识培训

培训合格证书

评估计划

评估阶段

项目评估

评估问题列表

持续支持纪要

指导问题解决

解决发现的问题

问题解决情况报告

6 沟通计划

参 与 人 员 一

沟通频率和时机

序号

沟 通 任 务

云端数据治理基础培训，过程

比对

项目执行负责人、质量控制员、工作

项目启动时

1

组成员

项目执行负责人、质量控制员、工作 组成员

2 过程改进实施计划制订

项目启动时

**第11章■云端数据治理体系实施■**

续表

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 续表   |  |  |  |  | | --- | --- | --- | --- | | 3 | 工作完成情况及遇到的问题 | 每周 | 项目执行负责人、质量控制员、工作 组成员 | | 3 | 云端数据治理过程改进中遇到 的疑难问题 | 每月 | 项目执行负责人、质量控制员、工作 组成员 | | 5 | 阶段评估工作准备 | 每次阶段评估前 | 项目执行负责人、质量控制员、工作 组成员 | | 6 | 项目总结 | 项目结束时 | 项目指导委员会、项目执行负责人、 质量控制员、工作组成员。 | |

**11.4.2** **实施改进措施**

组织也可通过绩效评估和审计等机制发现已经发生的或潜在发生的问题，然后 对问题的原因进行分析，并依托当前的资源针对导致问题的原因采取纠正措施或预 防措施，从管理体系的规定和执行两方面实施持续改进。

**参考文献**

[1]马欢，刘晨，译.DAMA 国际.DAMA 数据知识管理知识体系指南[M]. 北京： 清华大学出版社，2012.

[2]百度百科 . 知识[EB/OL].<http://baike.baidu.com/link?url=J-1FZ-YcEAI0d>

ZqI6hKvX68atuJTJ1-mrG0Kmqocrlm4hum2VnlIJOd\_ufYxyvPFqX3KgVVde-HzfodBn Y62uyX7YiP5zIF44B2McQz35Su.

[3]CIO 时代.DIKW: 数据、信息、知识、智慧的金字塔层次体系[EB/OL].http://

www.ciotimes.com/ProCase/85417.html.

[4]百度百科 . 数据治理[EB/OL].<http://baike.baidu.com/link?url=decS5NfcXA> AivvRQPXnVehb5yqVMcYB-SbYUdm8Dmrfy\_evPpwCqkFYmfkPeuah53TAq764Do9 wkCocH7Q\_70JX5tG4CaafebBRWtocOZaUX6q9ytpAQegBm2fbfkHOw.

[5]DAMA International.The DAMA Guide to the Data Management Body of Knowledge 1st Edition [M].USA:Technics Publications,2009.

[6]ISACA.COBIT5:A Business Framework for the Governance and Management of Enterprise IT [S].2012.

[7]Data Governance Institute.The DGI Data Governance Framework [R].2009.

[8]IBM Corporation.IBM Data Governance Council Maturity Model:Building a Roadmap for Effective Data Governance [R].2007.

[9]CMMI Institute.Data Management Maturity (DMM).[EB/OL].http:// cmmiinstitute.com/data-management-maturity

[10]商业银行数据治理能力成熟度评估应用研究.北京：北京理工大学， [EB/OL].<http://max.book118.com/html/2014/1101/10050845.shtm>

[11]Gartner Group.Big data[EB/OL].<http://www.Gartner.com/it-glossary/big-data>.

[12]桑尼尔·索雷斯.大数据治理[M]. 匡斌，译.北京：清华大学出版社.2014

[13]张明英，潘蓉. 《数据治理白皮书》国际标准研究报告要点解读[J]. 国际 提案，2015(6):54-57.

[14]张绍华，潘蓉，宗宇伟.大数据技术与应用.大数据治理与服务[M]. 上海： 上海科学技术出版社，2016.

[15]百度百科.5·28 携程网瘫痪事件 [EB/OL].<http://baike.baidu.com/link?url=> POW46wYrjm8tUk56dLPiqakk12989Ft2p-gYhg6LsOTYh5COOURCM-dX\_mTnAj\_YE 3gHqZO8HPPQ1TyWLp-hgrVv3Pmk⁷IitZE4jhDUQf8aUZ3Uov6tQ4vK4Ba3YyjCeZiox

**■参考文献■**

ZhlnAgQ6iOq7CaQkrYq3\_q4JsyD41lmsvpaGt5q

[16]TechWeb.携程瘫痪背后：国内企业数据管理混乱[EB/OL].http://www. techweb.com.cn/business/2015-05-29/2158525.shtml

[17]潘蓉.由“徐玉玉”事件引发的数据治理体系化思考.[EB/OL].http://www. 360doc.com/content/16/0831/23/31267089587401478.shtml

[18]DOIT. 金雅拓发布《2016年上半年数据泄露水平指数调查报告》[EB/OL]. <http://www.doit.com.cn/p/256787.html>

[19]宋秀明.住房公积金基础数据标准.中国建设信息，2014年21期

[20]新华网 NEWS.英媒称谷歌流感预测连续三年“不靠谱” .[EB/OL]. <http://news.xinhuanet.com/cankao/2014-03/15/c>\_ 133188416.htm.

[21]人民网.美使领馆发布 PM2.5 不合法规 [EB/OL].htp:/paper:people. com.cn/rmrbhwb/html/2012-06/06/content\_ 1063104.htm?div=-1

[22]财政部.企业内部控制基本规范.2008

[23]BS ISO/IEC 17788-2014 Information Technology -Cloud Computing - Overview And Vocabulary

[24]赵国祥，刘小茵，李尧.云计算信息安全管理—CSAC-STAR 实施指南[M]. 北京：电子工业出版社，2015

[25]McKinsey.Big data:The next frontier for innovation,competition,and

productivity.2015

[26]NIST.NIST Big Data Interoperability Framework:Volume 1,Definitions.2015

[27]大数据的特征.CDA数据分析师<http://cda.pinggu.org/view/453.html> [28]大数据时代.百度百科.

<http://baike.baidu.com/link?url=ADU2VFT>\_85opWM34nGACT4E7Nr7K0bQPCZ O8IGWeLUiAZtZo\_PJ32EWxUG\_jG542UGYZEXsXkFRLqVg3QbKbn4ijIAmimDfR0 R4ZaCSBOuFkj7Imle4dJ3senFZtZwTVELSSNzo6QviDHXQii9hkCa

[29]专访：大数据时代 人人都有洞察力.36 大数据 .http://www.36dsj. com/archives/1777

[30]赵国栋，易欢欢，糜万军等.大数据时代的历史机遇[M]. 北京：清华大学 出版社，2013.

[31]“云数据中心”发展现状及趋势.doit.<http://www.doit.com.cn/p/227504.html> [32]徐玉.全球数据中心发展趋势和特点[J]. 电信科学，2011,27(12):62-66. [33] Gartner.大数据观察.<http://www.shuju.net/article/MDAwMDI5M2DUy:html> [34] Rocky.搜狐.<http://it.sohu.com/20141212/n406884557.shtml>

[35] ISACA.IT专家网.<http://security.ctocio.com.cn/310/9407810.shtml> [36]中华人民共和国国务院令第664号，地图管理条例

[37]程广明.大数据治理模型与治理成熟度评估研究[J]. 科技与创新，2016

**■云端数据治理**

(9):6-7.

[38]良好的数据质量才能实现更好的业务成果，阿里云，<https://www.aliyun.com/> zixun/content/21115876.html

[39]数据分析前提：数据质量的基本要素， <http://www.68dl.com/bigdata>\_tech/ 2014/1020/13289

[40]李武，姚瑁.数据库原理及应用[M]. 哈尔滨：哈尔滨工程大学出版社，2010

[41]张淼，杨杭，李小军，等.一种搜索数据质量统计方法：CN,CN 103544314 A[P].2014.

[42]刘洪，黄燕.基于经典计量模型的统计数据质量评估方法[J]. 统计研究， 2009,26(3):91-96

[43]祝君仪.大数据时代背景下统计数据质量的评估方法及适用性分析[J].中 国市场，2015(29):41-42.

[44]张亮亮，辛金国.大数据背景下统计数据质量内涵探索[J]. 统计与咨询， 2016(2):6-9

[45]许鑫，冯诗惠.大数据领域的主题分析——基于WOS 和 Wikipedia 的互证 研究[J]. 情报杂志，2014(11):124-130

[46]刘益江.数据仓库的数据质量分析与评价[D]. 广东工业大学，2012.

[47]金翰伟.基于 Spark 的大数据清洗框架设计与实现[D]. 浙江大学，2016.

[48]丁小欧，王宏志，张笑影，等.数据质量多种性质的关联关系研究[J]. 软件 学报，2016(7) .

[49]Inmon W H. 数据仓库[M]. 王志海，译.北京：机械工业出版社，2000:5

[50]普元信息.[EB/OL]<http://www.primeton.com/read.php?id=2180&his=1>.

[51]Shon Harris.CISSP 认证考试指南(第六版)[M]. 张胜生，张博，付业辉译. 北京：清华大学出版社

[52]卢海红.基于 XML 数据仓库历史数据归档与管理策略的研究[D]. 哈尔滨 工程大学，2007

[53]Informatica. 用于全局数据治理的元数据管理，2012

[54]中国联通. 中国联通供应商主数据管理办法(试行),2013

**反侵权盗版声明**

电子工业出版社依法对本作品享有专有出版权。任何未经权利人书面许可，复 制、销售或通过信息网络传播本作品的行为，歪曲、篡改、剽窃本作品的行为，均 违反《中华人民共和国著作权法》,其行为人应承担相应的民事责任和行政责任，构 成犯罪的，将被依法追究刑事责任。

为了维护市场秩序，保护权利人的合法权益，我社将依法查处和打击侵权盗版 的单位和个人。欢迎社会各界人士积极举报侵权盗版行为，本社将奖励举报有功人 员，并保证举报人的信息不被泄露。

举报电话：(010)88254396;(010)88258888

传 真：(010)88254397

E-mail: dbqq@phei.com.cn

通信地址：北京市海淀区万寿路173信箱

电子工业出版社总编办公室

邮 编：100036



**云端数据治理**

本书的编写团队结合最新的科研成果以及多年来在成熟度评估方面的实战经 验，在借鉴国际先进数据治理理论和方法的基础上，针对云端数据特点，开

发了云端数据治理模型，构成了本书的主要内容。

本书旨在帮助读者和有数据治理需求的组织了解云计算环境下的数据治理方

法，为建立云端数据治理体系和云端数据治理实施提供指导。

**电** **子** **信** **息** **出** **版** **分** **社** **微** **博**

[**http://weibo.com/etpublish**](http://weibo.com/etpublish)

**策划编辑：牛平月**

**责任编辑：王敬栋**

**封面设计：汇众智捷**

官方微信平台