

数据治理建设管理方案（参考）

数据治理背景

- 1、多数企业内部系统建设多，数据分散，各系统数据共享困难，信息化建设“烟囱式”严重；
- 2、企业各部门系统口径不一致，数据标准不统一，数据质量无法整体评估；
- 3、数据不可知，数据挖掘不到位。数据使用者未知数据多，无法得知这些数据与业务的联系，不知如何管理数据；
- 4、数据不可控，数据使用者对汇集数据、处理数据、数据服务过程模糊
- 5、数据不可取，数据使用者知晓自己业务所需数据，但不知如何获取或者轻松获取；
- 6、数据关联性差，无法对数据进行自主探索和挖掘，数据价值停留在表面，没有形成知识图谱。

数据治理概念

数据治理是根据数据全生命周期、数据整体流向，将数据作为企业资产进行整体管控、人员绩效评判和风险管理工作的一套治理体系。保障企业数据及其应用过程中的合理运营、风险可控以及数据价值实现。将从企业组织架构、人员方案、规章制度、技术架构、人员绩效等多维度构建数据架构管理、元数据管理、数据质量管理、数据标准管理、数据安全治理、数据生命周期管理等模型，全面梳理整体数据，促进企业的数字化建设，是一个长期、复杂的工程。

数据治理目标

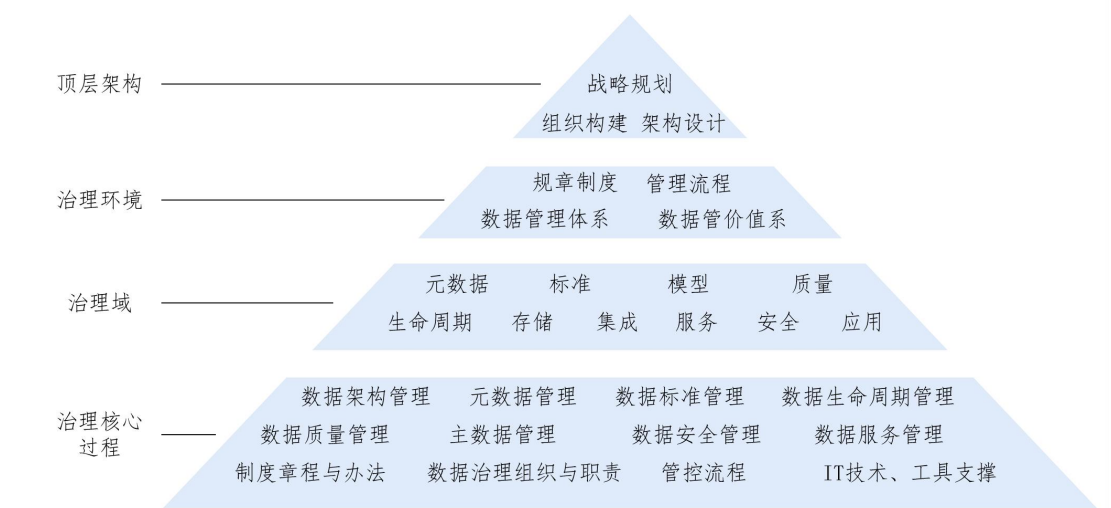
结合各行业组织信息化发展以及业务数据相关应用，《信息技术服务治理 第5部分：数据治理规范》明确了数据治理规范的实施方法和过程。数据治理框架以“运营合规、风险可控、价值实现”为目标，包括“顶层架构设计、数据治理

环境、数据治理域、数据治理过程”四大部分，分别对应数据治理实施基础、保障机制、治理对象和治理方法，实现数据治理与组织管理的融合。

运营合规：数据管理体系的建设必须符合法律、行业标准，并且通过数据评估、数据审计和优化保证数据合规，提高数据准确性和完整性。

风险可控：通过建立风险评估管理机制，严谨执行数据管理制度、约束组织架构、强化执行力度、明确管理流程，全局掌控组织内外数据治理环境、利益关系，降低经营风险，保障数据安全，将数据不确定性因素变为可知可控的风险。

价值实现：数据治理的最终结果，能否推进信息资源的整合，实现企业各部门数据资源共享，企业数字化转型持续发展都是数据治理价值的体现。



数据治理目标

基于数据治理体系，实现数据接入标准化、数据处理自动化、数据监控智能化、数据组织知识化、数据运行可视化、数据应用自助化。

- 1、数据接入标准化：**制定数据接口规范，统一数据标准规范；
- 2、数据处理自动化：**合理设计 ETL 过程或者使用 ETL 开发工具，提高数据抽取、转换、加载效率；
- 3、数据监控智能化：**通过定义多种接口规范，实现运维监控管理，以短信、邮箱等多种方式实现监控预警；
- 4、数据组织知识化：**各部门数据共享，提取核心数据，形成主数据模型，确定主题域，整合关联数据，形成以业务实体为单位的数据关系网络，即知识图谱；

5、数据运行可视化：合理使用 ETL 工具、自研产品，与数据治理平台相结合，任务调度和作业流程可视化；

6、数据应用自助化：建设企业综合服务平台，通过服务权限和数据权限控制实现数据便捷服务和应用。

需求分析

多数企业已经意识到数据资产管理的重要，但是对数据治理整个流程管控依旧存在不足，不仅限制组织数据质量的进一步提高，同时也限制了数据的价值实现。对于数据质量的需求主要分为：

1、数据标准和数据模型规范统一。

企业各组织机构均有一套独立的信息系统，各部门在各自的业务范畴内生产、使用和管理数据，使数据分散在不同的部门和信息系统中，缺乏统一的数据规范、数据来源和数据标准，导致数据格式、内容、含义上的不统一、不规范、冗余、无法在各部门共享。

2、需要规范统一的主数据。

组织机构核心系统的数据信息并不是存储在一个独立的系统中，或者说不是一个统一的业务管理流程在维护，缺乏主数据管理，导致整个业务范围内的主数据无法保证一致、完整，影响数据准确性。

3、数据质量企业化管控。

目前大部分的数据质量管理均由企业个部门分头进行，每个部门对数据质量的评估机制不一致，缺乏清晰的跨部门、跨机构的数据质量管控标准与规范，数据质量分析随机性强，存在业务需求不清的情况，影响数据质量；部门存在数据质量管理人员不足、知识与经验不够、监管方式不全面等问题；缺乏完善的数据质量管控流程和系统支撑能力。

4、基于数据全生命周期的治理。

大型企业或者政务单位，数据的产生、使用、维护、备份到过期销毁的数据生命周期管理规范 and 流程还不完善，对于数据何时过期以及是否有效无法准确识别，并且在多数企业非结构化数据未纳入数据生命周期的管理范畴；未有效利用

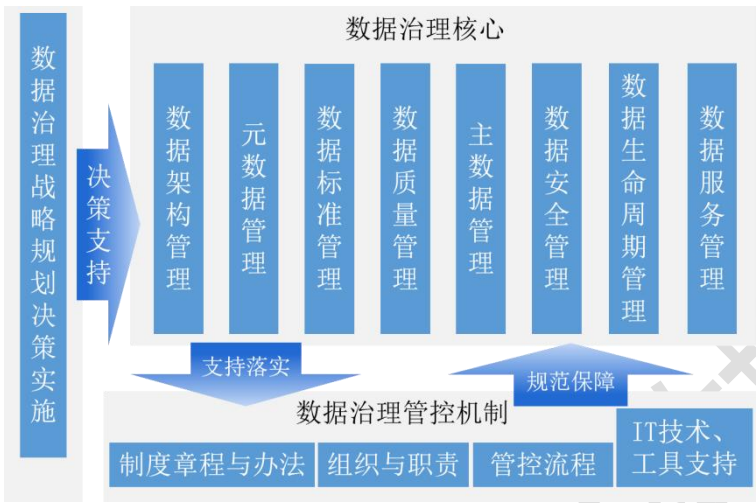
元数据管理。

5、建立数据治理监督 and 控制的专职组织。

上述四点均可归结于数据存储、数据管理的分散，权责不明确，各个部门关注数据的角度不一致，缺少一个能从全局视角，纵观整体数据管理的组织，导致数据管理规程、标准建立困难，相应的数据监管措施无法得到落实，考核体系也无从建立，无法保障数据管理规章制度的有效执行。

公众号·志明与数据

数据治理体系构建



数据治理体系架构

数据治理核心是数据治理架构的主体，包括数据架构管理、元数据管理、数据标准管理、主数据管理、数据安全生命周期管理、数据服务管理。以上管理基于数据治理管控机制，如制定规章制度、建立组织体系、明确组织职责和管控、IT 技术和工具的实行。同时，整个数据治理整套体系需要高层的战略决策支持。

数据治理核心领域

数据架构管理

数据架构是企业架构的一部分，数据模型是数据架构的核心，企业数据架构整合整个企业的数据并标准化，是企业架构的一部分，一个企业的数据架构是一套规范和文档的集合。数据架构是用于定义数据，指导对数据资产的整合和控制、使数据投资与业务战略相匹配的一套整体构建规范，包括正式的数据命名、全面的数据定义、有效的数据结构、精确的数据完整性规则以及健全的数据文档。数据模型是定义业务实体以及运营和直到业务所需的那些事实。数据模型是一种分析和设计方法，包括概念数据模型、逻辑数据模型和物理数据模型，是整个数据治理的重点。理想的数据模型应该具有非冗余、稳定、一致、易用性特点。一般

情况下，概念数据建模和逻辑数据建模是需求分析活动，物理数据建模是设计工作。

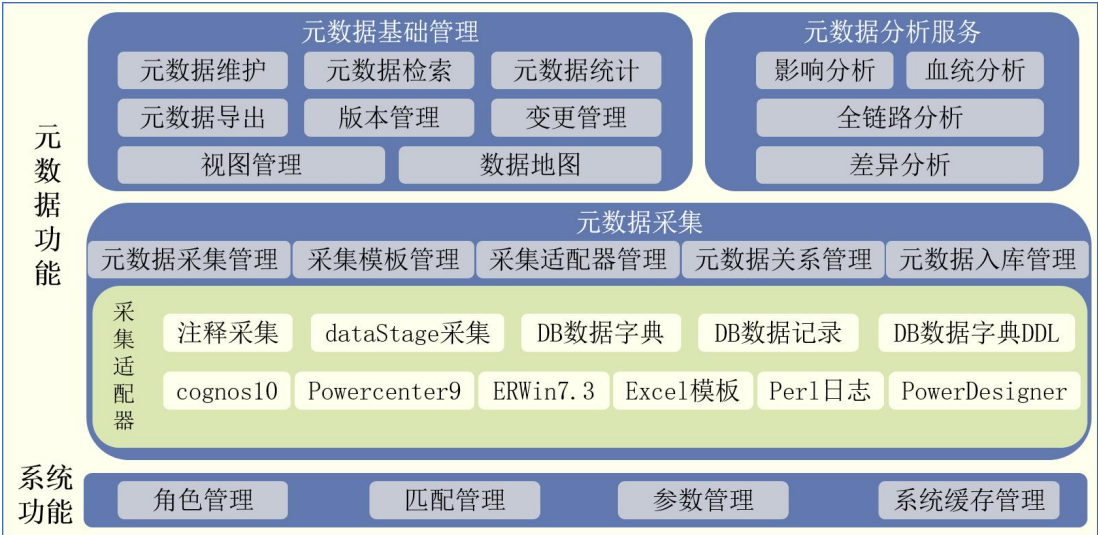
概念数据模型是一系列针对于每一个主题域的概念数据模型图表。用来定义业务实体以及这些业务实体之间的关系，业务实体是主要组成部分，是企业熟悉并感兴趣的事物、人员、地点的概念和类别。业务实体的例子是实例，概念数据模型图一般不描述业务实体的数据属性，可能会包括实体之间多对多的业务关系，促进人们对业务的理解，有利于语义上的一致性。

逻辑数据模型是在概念模型之下，增加了更多细节来反映每个实体的关键数据属性，企业逻辑模型识别每个业务实体实例所需的数据，关键数据属性代表了通用的数据需求以及那些被广泛共享的数据属性的标准定义，关键的数据属性是指如果缺失则导致企业无法正常运作的属性。

物理数据模型是逻辑数据模型在数据库中的具体实现，是根据技术约束、应用方法、性能需求和建模标准等优化详细的数据需求和业务规则的实施工作。内容包括数据库内所有的表、视图、字段及其相关主键和外键的定义，以及系统内数据流向及系统间的数据交换关系。

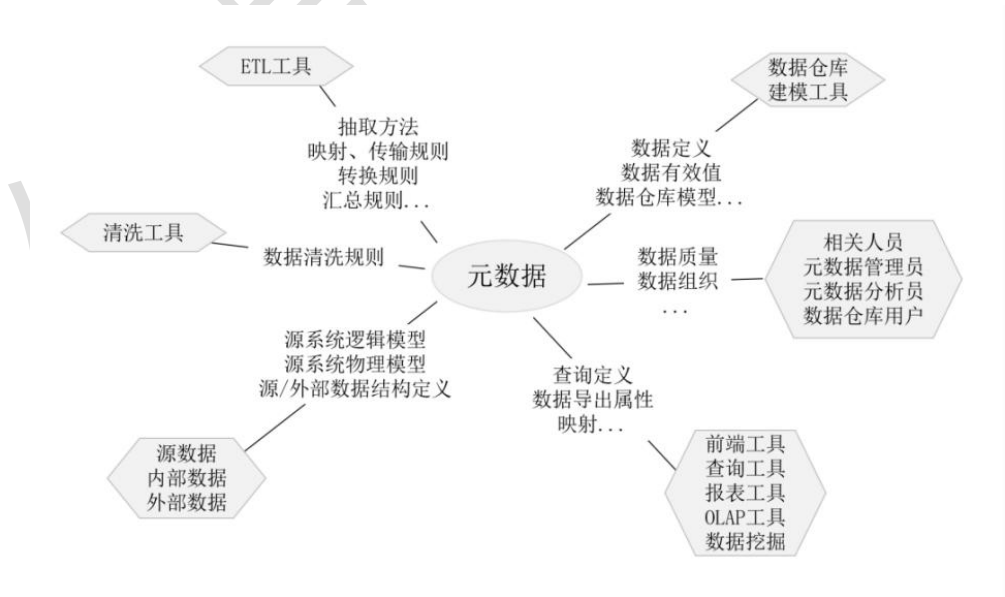
元数据管理

元数据是描述数据的数据，按用途分为技术元数据、业务元数据和管理元数据。技术元数据（Technical Metadata）用于描述数据系统中技术领域相关概念、关系和规则的数据；包括数据平台内对象和数据结构的定义、数据源到目的数据源的映射、数据转换的描述等；业务元数据（Business Metadata）用于描述数据系统中业务领域相关概念、关系和规则的数据；包括业务术语、信息分类、指标、统计口径等；管理元数据（Management Metadata）用于描述数据系统中管理领域相关概念、关系、规则的数据，主要包括人员角色、岗位职责、管理流程等信息。



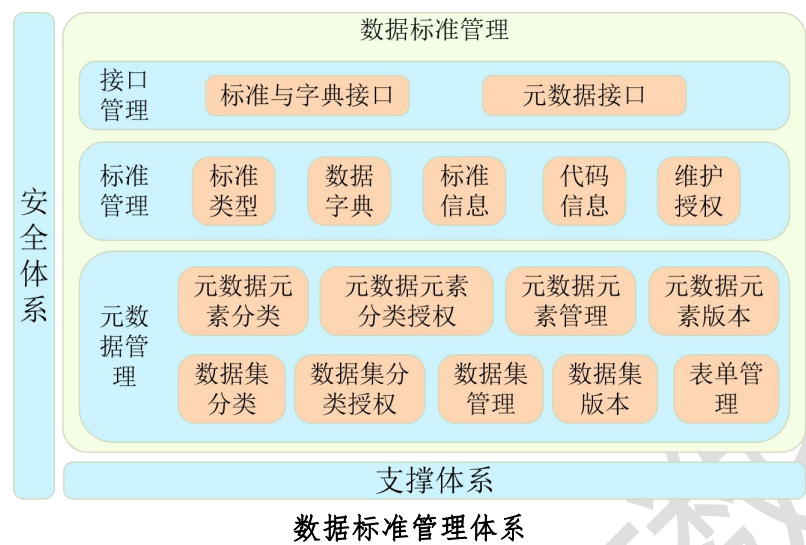
元数据管理功能架构

目前绝大数元数据管理系统都是按照技术元数据、业务元数据、管理元数据来开发和设计的，尤其强调了血缘关系；数据仓库是天然的靠血缘关系支撑的系统，但是业务系统的血缘关系并没有那么明显，即使在数据仓库系统中，血缘关系并非想象的那样唾手可得，绝大多数元数据管理工具的血缘关系大多来自于表的主外键依赖关系和对存储过程的解析，对于ETL工具而言则来自ETL开发过程。单纯的元数据管理存在的价值还是极其有限的，必须把元数据管理和数据标准、数据质量、数据资产、数据安全、数据认责等管理结合在一起，并通过各种服务向外提供给业务系统，才能真正发挥元数据的价值。



元数据体系

数据标准管理



数据标准是指研究、制定和推广应用统一的一套符合企业自身实际业务，包括数据的定义、数据分类、记录格式、操作、数据的应用多个方面的标准化体系，实现业务、技术、管理三方面的规范化。

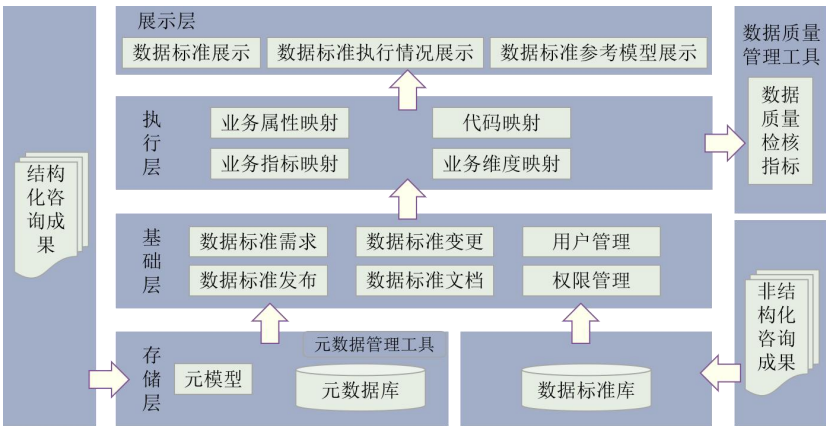
数据治理对标准的需求可以划分为三类：基础类数据标准、指标类数据标准和专有类数据标准。基础类数据是指组织日常业务开展过程中所产生的具有共同业务特性的基础性数据。

基础数据可分为客户、资产、协议、地域、产品、交易、渠道、机构、财务、营销等主题。

指标类数据是指为满足组织内部管理需要及外部监管要求，在基础性数据基础上按一定统计、分析规则加工后的可量化的数据。

专有类数据标准是指公司架构下子公司在业务经营及管理分析中所涉及的特有数据。

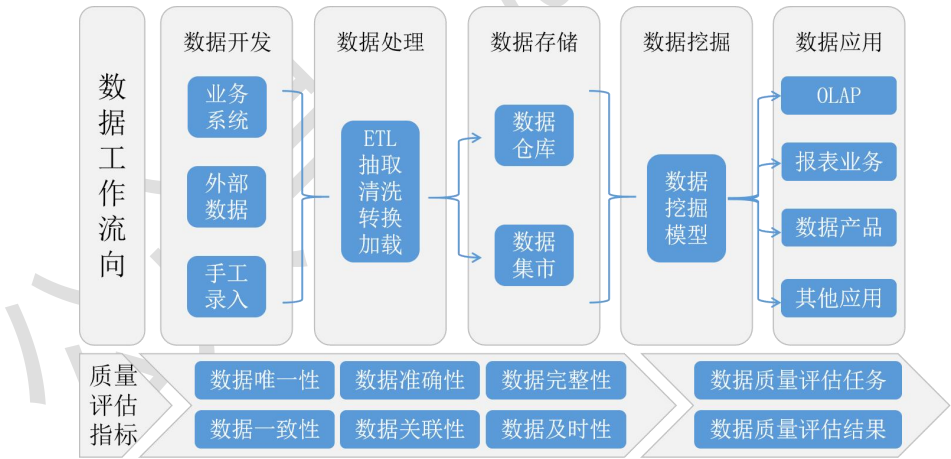
数据标准定义统一的数据标准，减少数据定义的二义性；各部门需要一个统一的数据接口，包括编码标准、数据内容格式上的统一；建立稳定的企业数据模型，统一所有数据，制定完整的数据地图，统一存储格式和压缩方式，提高数据存储效率和资源使用率。



数据标准管理功能架构

数据质量管理

数据质量管理是对数据从采集、存储、使用、维护、共享、过期全生命周期的各个阶段里可能会出现的问题，进行识别、监控、预警等管理活动。数据质量管理是一个将技术、业务和管理集合的解决方案。用有效的数据质量管理手段，降低甚至消除数据质量问题从而进一步提升企业数据变现的能力。数据质量管理包括对数据的绝对质量管理和过程质量管理。绝对质量即数据的准确性、完整性、一致性等数据本身的质量，过程质量是指数据应用、存储和传输的质量。



数据质量管理体系

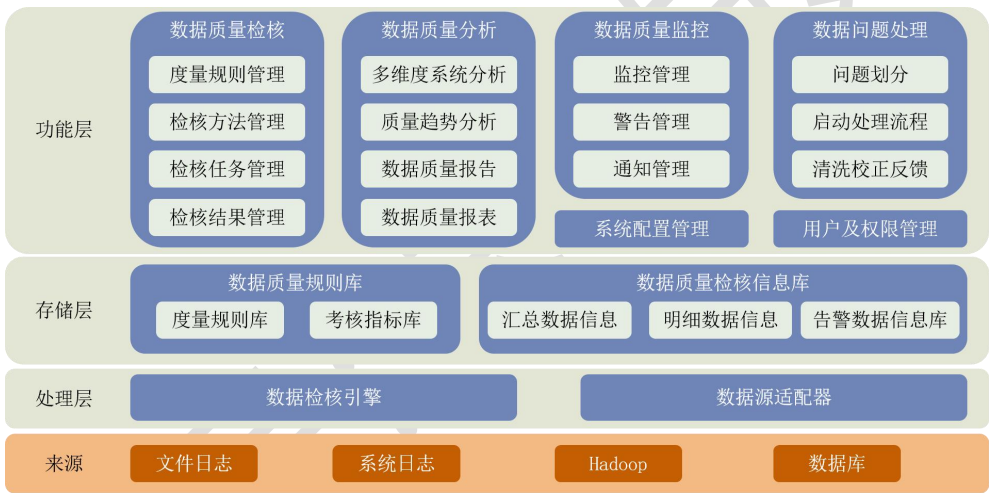
一般企业所遇到的数据质量问题如下：

- 1、数据真实性：数据必须反应客观的实体存在或者真实业务；
- 2、数据可靠性：分析和识别哪些数据不是准确的或者无效数据；
- 3、数据唯一性：用于识别和度量重复数据、冗余数据；

- 4、数据完整性：包括模型设计、数据条目、数据属性不完整一系列问题；
- 5、数据一致性：多源数据的数据模型、数据实体等不一致，相同的数据有多个副本的情况下的数据不一致、数据内容冲突；
- 6、数据关联性：数据关联关系存在缺失或者错误会直接影响数据分析结果；
- 7、数据及时性：及时性是指数据能否在需要的时候迅速获取到数据，与企业的数据处理速度和效率直接关联，是业务处理和管理效率的关键指标。

数据质量管理的规划和实施包括：

- 1、建立数据质量管控体系、数据质量评估体系，定期对数据质量进行评估；
- 2、各部门系统落实数据质量检查规则，持续改进；
- 3、结合数据质量管理与业务稽核，发现数据质量更深层次的问题；
- 4、制定数据质量考核机制，加强业务人员对数据质量问题的认识。



数据质量管理功能架构

主数据管理

主数据即企业各部门业务系统中最核心，最需要共享的数据。主数据管理需要将此类数据集中清洗和扩展，将其传输给组织范围内需要使用的操作性应用系统和分析性应用系统中。主数据管理需要遵循“六统一”原则

统一管理、统一标准、统一平台、统一建设、统一运营、统一应用。

主数据统一管理以确保方便、安全、快速、可靠地利用数据辅助决策和业务执行；

主数据统一标准属于数据标准之一，需结合企业信息数字化转型，对企业全业务领域的主数据进行体系化研究，由上而下开展主数据标准化工作；

主数据统一平台是以主数据全生命周期为定位，构建企业主数据代码管理体系、主数据库、标准规范等主数据信息的资源共享平台；

主数据统一建设需要对各类业务主数据制定分类标准，根据主数据收集、清洗、转换和编码过程，建立企业共用的主数据库；

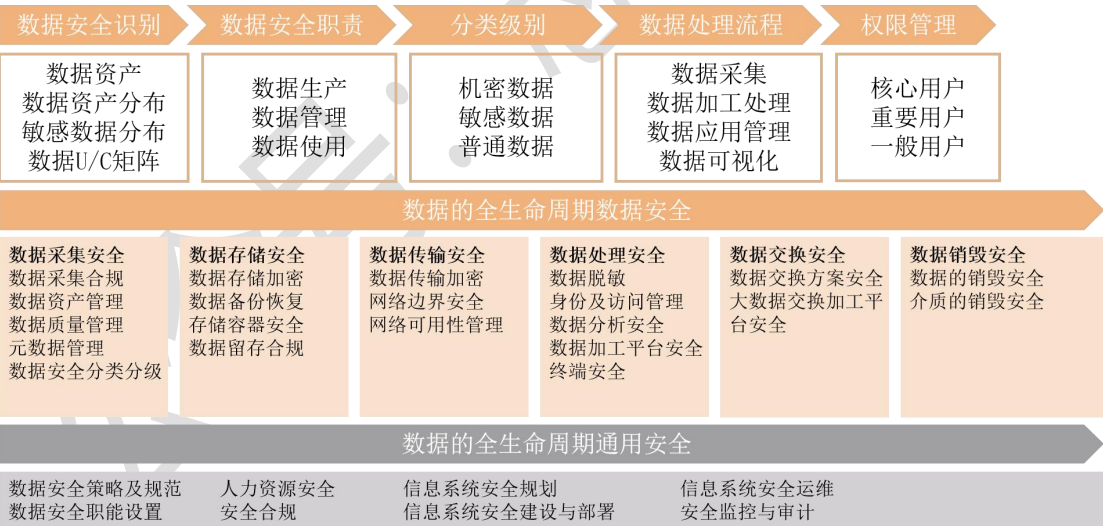
主数据统一运营，建立企业范围内主数据代码维护细则，对各部门生产的主数据进行统一审核、发布、维护等运营管理，保证主数据的一致性；

主数据统一应用是整个企业信息数字化的集中体现，主数据统一使各类业务数据抽取更加便捷，数据使用者能将各类主数据关联应用，提高企业整体业务数据价值。

数据安全

数据安全是指保护数据免受未经授权的访问、使用、更改和破坏。主要解决的是数据在存储、使用和传输交换过程中的安全问题。

数据安全管理流程可分为数据资产梳理与敏感数据识别、数据安全职责、数据分类分级、数据访问授权。



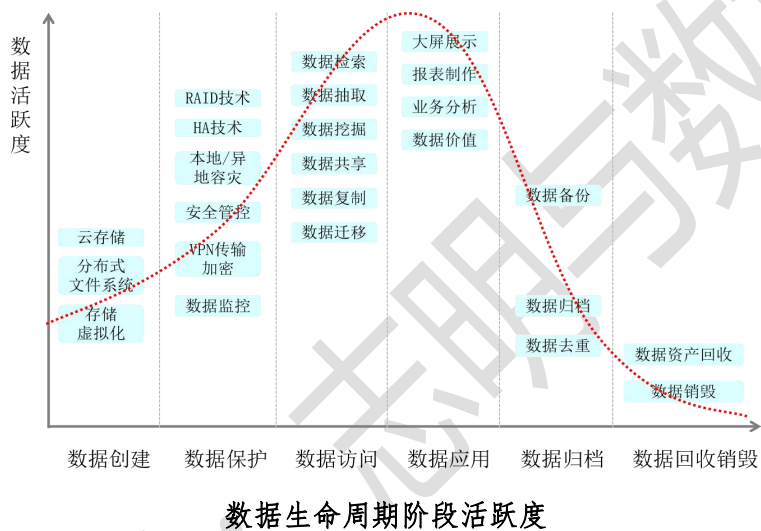
数据安全管理体系

数据使用的安全性包括基础数据的存储、访问和权限管理；对系统中采集的敏感信息在下游分析系统和内部管理系统中需要进行加密处理，避免数据被非法访问；建立访问权限管理体系，包括单点登录问题及用户名、数据和应用的访问授权统一管理，可通过数字水印等技术实现企业级权限系统；为数据修改、使用

等操作制定安全审计方法和责任追究制度，包括数据安全管理办法、系统开发规范、数据隐私管理办法、管理决策和分析系统审计管理办法等。

数据生命周期管理

数据全生命周期包括数据创建、数据保护、数据访问、数据应用、数据归档、数据回收销毁。在数据的生命周期中，数据自身价值决定了数据全生命周期的长度，并且数据价值在不同阶段是不同的，具体可以用数据不同阶段的活跃度来反应数据价值。



数据创建：数据应该能够按照数据质量标准和企业业务发展的需要生产，新的数据量增长速度快，需要存储环境以利于及时的处理、管理和保护，需要稳定、可靠、高可扩展能力的存储设备，如云存储服务器、分布式文件存储系统。

数据保护：企业经济效益与业务数据息息相关，越来越多企业信息向数字化方向转变，对于企业重要数据而言，需要对此类数据进行保护，本地磁盘、云服务、远程复制、集群备份、数据加密，采用不同的数据保护措施和技术，以保证各类数据和信息得到及时的和有效的保护。

数据访问：数据在整个生命周期中的活跃点，企业数据集中管理的重要目的之一，企业信息化相关部门需要对数据进行分析处理，以挖掘出对于管理及业务开展有价值的信息。此时数据在经过抽取、转换、加载之后，数据价值提高。在这个过程中就需要关注数据提取操作是否可能对数据库造成破坏、提取出的数据在交付给分析处理人员的过程中其安全性是否会降低、数据分析处理的环境安全

性等。

数据应用：数据活跃度至高点，此时的业务数据频繁使用在大屏展示、报表制作、业务分析中，数据价值达到最高，以驱动企业业务的发展，精准营销，决策辅助，所以数据传输过程中需要考虑保密性和完整性的问题，对不同种类的数据分别采取不同的措施防止数据泄漏或数据被篡改。

数据归档：维持一个数据备份和归档系统可以从多个方面支持企业的业务运作，可以防止这些记录被无意破坏，确保那些仍然对于一个组织具有一定作用的数据可以得到妥善的保存，让企业可以在原始信息因为某种原因被损坏或破坏时进行恢复。

数据回收销毁：这个阶段主要涉及数据的保密性。应明确数据销毁的流程，对没有保留或保存必要的数据进行销毁或回收，采用必要的工具，数据的销毁应该有完整的记录。对于需要送出外部修理的存储设备，送修之前必须对数据进行可靠的销毁。

数据服务管理

数据整合归集最终目的就是要服务于各机构部门、人员等，能更准确更快更方便的服务是数据服务管理的目标。数据服务管理是指针对内部积累多年的数据，研究如何能够充分利用这些数据，分析行业业务流程，优化业务流程。

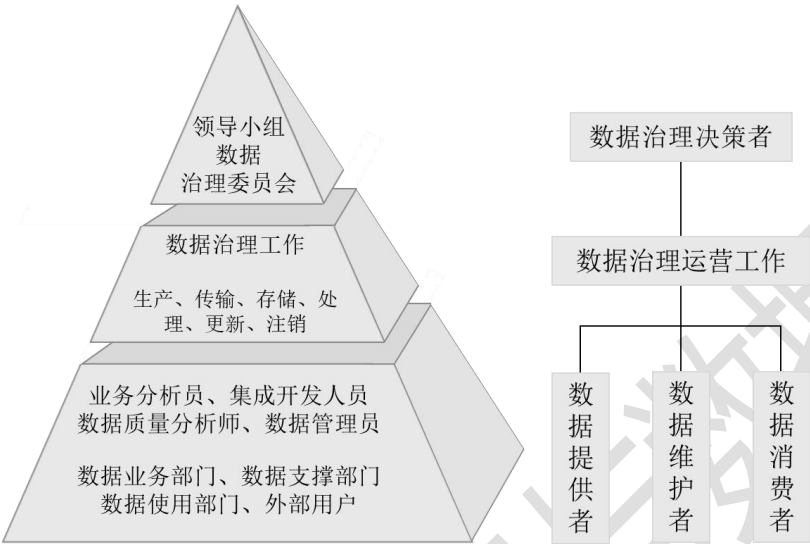
数据使用的方式通常包括对数据的深度加工和分析，包括通过各种报表、工具来分析运营层面的问题，还包括通过数据挖掘等工具对数据进行深度加工，从而更好的管理者服务。通过建立统一的数据服务平台来满足针对跨部门、跨系统的数据应用。通过统一的数据服务平台来统一数据源，变多源为单源，加快数据流转速度，提升数据服务的效率。

数据治理管控机制

组织构建

数据治理组织的构建旨在通过建立数据治理组织架构明确各级角色和职责，保障数据治理的各项管理办法、工作流程的实施，推送数据治理工作的有序开展，

如建立数据治理委员会、数据管理委员会、数据管理团队、数据质量团队、数据安全团队、数据技术团队等。



数据治理组织架构图

组织结构	角色	角色描述	角色主要职责
数据治理委员会	数据管理决策者	由公司主管领导和各业务部门领导组成	负责牵头数据治理工作；制定数据治理的政策、标准、规则、流程，协调认责冲突；对数据事实治理，保证数据的质量和隐私；在数据出现质量问题时负责仲裁工作。
数据治理工作中心	数据平台运营工作者	数据管理中心机构的平台运营人员	负责提交数据标准的要求及数据质量规则和业务规范，解释数据的业务规则和含义；监督各项数据规则和规范的约束的落实情况；负责数据治理平台中整体数据的管控流程制定和平台功能系统支撑的实施；负责平台的整体运营、组织、协调。
各业务部门	数据提供者	相关数据开发提供人员	负责数据及相关系统的开发，有责任执行数据标准和数据质量内容；负责从技术角度解决数据质量问题；作为数据出现质量问题时的主要责任者。
	数据维护者	数据维护人员	制定相关数据标准、数据制度和规则；遵守和执行数据标准管控相关的流程，根据数据标准要求提供相关数据规范；作为数据出现质量问题时的次要责任者。
	数据消费者	数据使用人员，包含内部用户和外部用户	作为数据治理平台数据管控流程的最后参与使用者；是数据资产价值的获益人；作为数据治理平台数据闭环流程的发起人。

规章制度

数据治理章程类似于企业的公司条例，该章程用以明确数据治理相关战略规划、合规管理和控制、标准和规范，如《数据治理工作管理办法》，加强组织数据治理管理能力，满足组织数据治理的要求，统一数据标准，提高数据质量，深化数据应用，支撑业务发展，提升管理水平及合规监管要求。

管控办法

数据治理管控办法是基于规章制度与工具的结合，可落地的操作的办法，用以完善数据管控相关的规章制度起草、审核、落实、监督的部门和管控机制，明确各数据管理活动的人员角色和职责、管理内容和相关流程，一般组织会设置《数据标准管控办法》、《数据质量管控办法》、《数据质量评估办法》、《元数据管控办法》等。

管控流程

数据管控流程是基于数据治理各核心领域的工作目标，在数据治理组织下，明确各参与方的工作步骤、工作任务以及流转方向。数据治理流程的设立是数据治理核心领域内容得以贯彻执行的基础，通过将数据治理任务进行分解，最大程度提高数据治理的执行效率，降低管理成本。

数据治理管控流程包括流程目标、流程任务、流程分级，根据数据治理的内容，建立相应流程，且遵循组织数据治理的规则制度。实际操作中可结合所使用的数据治理工具，与数据治理工具供应商进行协商，建立符合组织的流程管理。

数据管控流程的设计应从组织实际出发，充分考虑组织的业务特点、管控模式以及数据管控的应急响应，同时将数据管控流程固化在相应的管理工具或平台，实现各数据管理流程的自动化、可视化和实时监控。

考核机制

人员考核是保障制度落实的根本，建立明确的考核制度，实际操作中可根据

组织情况，建立数据治理管控相关的考核指标、制度和办法，不断优化相关机制。一般组织都会针对数据标准管理、数据质量管理、元数据管理等核心治理域设置相关的考核机制与指标，并与个人绩效相关联。

IT 技术工具支持

架构支撑



数据治理架构支撑

技术规范

技术规范是保障数据治理平台可持续管理的基础，随着数据量的增长、技术水平的发展，为更好、可持续的实现数据资产的管理、应用，需建立明确的技术规范。结合数据治理相关管控制度和流程，构建组织的数据技术规范，包括：数据应用研发规范、数据架构规范、门户数据整合规范、数据存储规范等。

选型规范

在数据治理体系建设和完善的基础上，调研分析数据治理产品和工具，结合组织实际应用需求，形成相关的数据治理产品选型标准和建议。

数据治理管控宣介

结合组织数据治理现状、数据治理体系建设、数据治理未来愿景、组织团队建设以及信息系统建设等，构建具有针对性的数据治理管控宣介方案，形成数据治理的全员意识，为组织数据治理的成功实施提供保障。

公众号：志明与数据

附录

数据治理工作管理办法

为逐步提高数据资产利用效果，推动信息化建设向标准化、信息化和数字化方向发展，数据治理应遵循以下原则：

- （1）统一规范。数据标准要严格执行组织的统一标准。
- （2）分级管理。实行分层级的数据管理模式，明确职责分工，层层落实责任。
- （3）过程控制。建立数据从采集、报送、审核到应用、维护全过程的控制规范，保证数据质量，提高应用效果。
- （4）保障安全。建立数据访问的身份验证、权限管理及定期备份等安全制度，规范操作，做好病毒预防、入侵检测和数据保密工作。
- （5）数据共享。整合应用系统，做到入口唯一，实现数据一次采集，集中存储，共享使用。

数据治理的规范工作包括：数据标准、数据采集、数据审核、数据维护、数据分析、数据应用、数据发布、数据传输、数据存储（备份、恢复）、数据安全、数据质量监控、数据管理考核等。

（一）数据标准

第一条组织数据必须按统一制定的数据标准和代码规范执行。

第二条在数据采集工作中，组织依据制定的标准进行数据的采集录入、审核修正、数据质量监控和管理考核，确保数据质量。

第三条数据指标的设定和修改必须严格执行软件的管理规定和程序。数据治理管理小组根据业务管理部门的需求，及时制定（修改）数据采集的技术规范，包括数据采集格式、权限设定和方法步骤等，并按时向业务管理部门通报所采集数据的指标内容变化情况，以便业务部门根据采集到的指标提出数据加工需求。

第四条数据指标的确定应以有效实用和优化服务为目标，不断提高数据集中度和信息共享度，科学归并各项业务的同类、同属性指标，避免业务部门重复上报数据。

（二）数据采集

第一条数据采集重点是原始数据的采集，原始数据主要包括业务部门通过纸质资料报送的数据，通过电子传输报送的数据（含电磁介质报送的数据和网络传输报送的数据，下同）；相关部门通过纸质、电子传输提供的数据；管理业务处理结果（检查、调查、核实、认定、审批等）产生的数据。

第二条数据采集主要依靠应用系统进行录入（导入），现用应用系统不能满足数据采集需要的，通过相关系统的补录模块进行数据录入（导入）。

第三条数据采集方式分为手工采集、数据电文导入和光学字符识别（OCR）、图形扫描等其他方式。

为保障数据采集的质量，提高数据采集的效率，应充分应用信息化技术，最大限度地减少手工录入。

第四条为提高数据录入质量，组织各有关机构部门要认真执行资料数据的核对制度，严格以原始资料为依据，确保录入数据的一致性和准确性。数据采集录入前，操作人员应先将准备录入的原始资料或有关表证单书进行完整性、逻辑性、真实性审核，发现漏填、错填和逻辑关系不符的，应于做出修正或补正；

第五条对通过电子传输方式报送数据的，数据必须通过接收软件相应设置的完整性、逻辑性审核。

第六条各部门机构应建立纸质数据和电子数据的核对制度，明确职责分工和工作程序，落实岗位责任，发现问题及时处理。

第七条对于未规定的数据采集标准规范，组织相关责任部门应根据不同的数据采集岗位和软件覆盖面，细化工作职责，量化作业标准，结合业务操作规程，建立统一的数据采集录入操作规范和管理制度。数据采集规范应包括数据来源、采集时间、操作岗位、操作步骤、操作内容和采集内容的逻辑校验等要素。

第八条各级数据采集人员应强化数据质量意识，熟练掌握正确的采集和审核方法，按照部门职责分工，严格按照相关操作规范采集各类数据。

第九条任何人不得擅自委托他人以本人用户名录入数据。确因工作需要委托他人以自己名义录入的，必须报经主管领导批准，并及时更改口令。

（三）数据审核

第一条数据审核是确保数据质量的重要环节。审核的重点是数据的及时性、

完整性、真实性、准确性和规范性。

(1) 及时性是指数据要在规定的时间内采集，确保应用系统数据与实际业务同步；

(2) 完整性是指数据采集应按照规定及相应表、证、单、书采集的要求进行，不得缺表或漏项，杜绝数据的机外操作和循环；

(3) 真实性是指数据采集要如实反映组织管理实际，以合法、真实的原始资料为依据，不得随意捏造；

(4) 准确性是指数据采集应准确反映组织管理实际，与纸质资料数据一致，数据之间逻辑相符，不得出现运算错误或逻辑错误；

(5) 规范性是指数据采集应按照统一的业务标准及技术规范进行。

第二条各部门及相关岗位人员按照自身工作职责范围，负责相关数据的日常审核。

第三条为加强数据质量管理，建立数据审核制度，数据管理员定期或不定期对已采集数据进行审核。对审核中发现的数据质量问题应及时反馈至具体的原始数据采集岗，限期补录修正。

第四条各部门机构可采取人工抽样、数据软件检测等方式进行数据审核。对检索出不符合标准的数据，反馈至原数据采集岗进行补录修正。在数据应用环节发现的数据采集差错，应通过相关业务管理部门的数据管理员进行审核确认，然后反馈至原数据采集岗进行变更修正。

第五条审核中发现的差错应及时通知相关部门。

第六条数据审核岗人员与数据采集岗人员对数据问题有异议的，由上级业务主管部门负责对争议事项做出裁定。

(四) 数据维护

第一条数据维护包括对数据中错误数据的修正、不完整数据的补充、垃圾数据的清理及历史数据的迁移等。

第二条各业务管理部门应结合各应用系统制定详细的数据维护工作制度，明确数据维护的权限、职责，严格按照工作制度进行数据维护。已经进入应用系统的数据，不得擅自修改、删除。

第三条对错误数据的修正和不完整数据的补充，组织统一按以下程序处理：

(1) 对尚未录入应用系统或已录入尚保留修改权限的数据，需修改的，可由前台原数据采集岗根据审核意见纠正后重新录入（或直接纠正）；

(2) 对已录入应用系统的数据，经审核发现错误，前台原数据采集岗已无权限修改的，应根据具体业务工作流程和要求实行层级审批修正。

第四条对已录入应用系统需层级审批修正的数据，信息部门有权限处理的，应由数据修正申请机构部门及时提出数据修正申请，报数据所属系统的责任部门审核审批。

对经确认不能予以修正的，有权修正单位应及时向申请单位或相关人员说明原因。

第五条对需由数据中心办理的数据修正事项，应由申请部门机构及时提出数据修正申请，分别报相关责任部门审批。数据中心在接到同意修正的审批后办理。

对经确认不能予以修正的，有关单位应及时向数据修正申请单位及有关人员说明原因。

第六条对不按规定程序上报审批的数据修正申请，各部门机构数据管理的技术管理团队不予以修改。

第七条对经查实，属有关单位弄虚作假、人为调整指标等主观因素造成其无法自行恢复的数据错误，各部门机构数据管理的技術管理团队有权不予修正，并向相关责任部门报告。

第八条各部门机构数据管理的技術管理团队应按照系统初始维护设置和上级部门分配的权限实施数据修正业务。执行过程中如遇权限调整，各部门机构数据管理的技術管理团队应在权限变动后十五日内向相关责任部门公告通知。

第九条数据维护前应做好相应备份工作。数据维护工作应严格备案，对每项数据维护的内容、时间、维护原因、责任人等进行详细记录，涉及的书面材料必须登记存档。

第十条应用系统因数据平台转换、系统升级等原因需对历史数据转储、迁移的，由信息技术部门会同业务管理部门确定历史数据的处理方法，保证新旧数据的衔接和系统的平稳过渡。

第十一条信息技术部门应加强对数据的监控，定期检测数据的存储，分析数据的构成，提出数据清理优化的方案，经业务管理部门确认后，定期进行垃圾数

据的清理。垃圾数据主要指过期的、重复的、没有使用价值的数据，不包括按规定应保留的历史数据。

（五）数据分析

第一条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，充分发掘数据中蕴涵的信息，形成有用指标数据值的过程。

第二条数据分析一般按照选题、分析、应用、反馈、再选题的方式进行，积极研究、探索科学实用的分析方法，同时依据数据应用过程中暴露出的问题，改进数据管理方法和手段，促进数据管理质量不断提高。

第三条数据分析分为固定分析和专项分析两种。固定分析指标在一定时期内相对固定，通过编制应用软件自动生成。专项分析的选题指标根据组织所处行业的新形势和新问题专门设计，临时从前台或后台抽取相关数据进行分析处理。

第四条各部门机构应充分利用数据资源，规范数据分析程序，积极探索数据获取、加工、计算、展现等过程的规律，不断探索和创新数据分析方法，提高数据分析质量。

（六）数据应用

第一条数据应用分为直接使用和加工后使用。数据加工是指根据工作需要，用数理统计、多维分析等科学方法对数据进行抽取、汇集、归类、挖掘、比对，并以报表、图形、文字等形式展现数据处理结果。

第二条各部门机构应充分利用数据资源，深化数据应用，用数据描述现状，预测趋势，规范执法，加强管理，优化服务，提供决策支持。数据应用的任务一般包括：

（1）整合应用系统的数据，建立数据关联关系，满足各级、各部门、各岗位业务管理的需要。

（2）分析行业管理现状，研究存在问题，提出改进措施，提高管理水平，规避管理风险。

（3）监控执法过程，跟踪执法结果，检查执法质量。

（4）建立合理的指标体系，全面考核和监控各部门机构管理的工作质量和效率。

(5) 统筹规划、合理分布数据，实现各类数据的全面共享，简化组织审批程序，优化数据服务。

第三条组织数据加工统一由数据中心承担，信息部门提供技术支持。业务管理部门有特殊需求时，报相关责任人批准后及时向数据管理中心提出，数据管理中心应及时提供已加工数据列表和数据使用授权，以保证各部门业务管理需要。

第四条数据加工分两个层次，一个层次是为满足日常业务需要，利用数据加工工具进行数据加工，供各部门日常管理使用，可直接展示数据处理结果；另一层次是满足评估分析和辅助决策等数据应用系统的需要，利用数据加工工具或专用软件对数据加工或进行系统间的数据衔接。

第五条数据中心负责数据加工软件的日常管理工作。

第六条数据使用单位和人员必须严格按照授权使用数据，负责管理组织、本人口令，不得越权使用数据；不得采取任何方法破坏数据；对所使用的涉密数据负有保密责任。

(七) 数据发布

第一条数据发布是数据管理的重要内容。各部门机构必须建立数据发布制度，明确数据发布职责，依法发布数据。

第二条各部门机构对外公布数据必须报领导或分管领导批准后对外公布，否则不得擅自对外公布和发布数据。

(八) 数据传输

第一条各部门机构应针对各项数据传输工作，设置专门岗位，明确职责分工，制定相应考核制度。

第二条数据传输工作严格遵照相应的操作规程和时间要求，不得延误。由于特殊原因，数据发送方不能按时完成数据传输任务时，应及时通知数据接收方，双方应积极采取措施，恢复正常传输。数据传输完成，双方应及时进行数据对账。

第三条数据传输应当使用组织机构内部计算机网络完成，未经批准不得借助其它公共计算机网络平台进行数据传输。使用可移动载体进行数据传输的，传输完毕后，必须从载体上完全清除数据。

第四条各部门机构内部信息的传递，如任务下达、信息交换、调查统计、报表报送、成果报告、情况通报等，应采取电子数据传输方式。

第五条信息技术部门要定期检测网络运行状况，及时发现解决网络传输中的问题，确保数据传输网络畅通。

（九）数据存储、备份和恢复

第一条信息技术部门应加强对各类数据存储和备份的管理，以保障应用系统的正常运行，保存完整的历史数据。

第二条信息技术部门应定期对存储和备份的数据进行整理优化，以提高系统运行和数据处理的效率。

第三条各类数据由信息技术部门统一集中存储和备份。

第四条信息技术部门应根据不同类型数据的更新频率、数据量、重要程度、保存期限，制定相应备份、恢复策略和操作规范。

第五条已做备份的数据任何部门不得擅自更改。

第六条数据备份文件必须存储在非本机磁盘的其它介质中，建立登记制度，由专人保管，备份介质必须保存在符合条件的环境中，对应用系统中存储时间长、使用频度低的历史数据，可按有关规定予以结转、存储，长期妥善保管，并随时可供调取应用。重要数据应异地存放。

第七条数据备份文件应定期进行恢复测试，以确保所备份的数据能够及时、准确、完整地恢复。

（十）数据安全

第一条数据管理中心应建立健全数据安全管理制度和办法。数据安全管理的内容包括：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、网络安全等。

第二条为统一规范操作权限，各部门机构操作人员的录入权限、访问权限、维护权限应先由各业务管理部门按照部门、岗位的职责分工，提出权限设定规则，报数据管理领导小组批准后由信息技术部门予以维护，任何人不得擅自设立、变更和注销。

第三条各类应用系统的使用必须实行用户身份验证。应用系统应按规定设置相应的用户名、密码，并按不同权限级别，对用户口令加密保护。操作人员应注意自己用户名和口令的保密，并定期或不定期修改口令。

第四条数据安全实行专人负责制。信息技术部门要指定专人负责系统数据及

磁介质资料的的安全管理工作。对数据库的管理实行数据库管理员制度，制定和明确管理员用户和数据查询用户的操作权限及规程，关键数据库管理岗位应设两人或两人以上。

第五条对涉密数据的传输、存储，应严格按照相关规定进行加密处理。

第六条对各类数据应严格执行保密制度，不得泄漏。对涉及从业人员隐私而未征得本人同意或按有关保密规定不能公开的数据，不得对外公开，也不得用于业务以外的其它目的。

第七条对数据的各项操作应实行日志管理，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。

第八条各部门机构要加强网络安全管理，采取严格措施，做好计算机病毒的预防、检测、清除工作，防止各类针对网络的攻击，保证数据传输和存储安全。管理员用户应掌握和运用数据库访问跟踪布控技术，加强对数据库的核查与监控。

第九条对需要长期保存的数据磁带、磁盘，应在质量保证期内（一般为一年）进行转储，以防止数据失效造成损失。

第十条系统维护和数据管理责任单位应当采取以下措施，保证信息系统应用数据的安全：

- （1）明确岗位职责，严格操作规范，完善内部控制；
- （2）加强应用系统的运行维护，定期进行数据备份；
- （3）实时监控数据库的访问情况，定期审核和更新数据库的口令等

（十一）数据质量监控

第一条数据质量监控是数据管理中心以组织业务规程为基础，通过建立数据质量指标体系和设置过错类别，运用一定的手段和方法，对一定时期的数据质量进行检查、认证、鉴定与评价等一系列管理活动。

第二条数据质量监控应按照分途采集、集中比对、管住增量、强化责任的原则，对数据采集和处理的真实性、准确性、完整性、合法性、逻辑性、及时性进行监控。

第三条组织各系统建立数据质量监控联动机制，实行数据管理中心、基层数据管理员监控联动。

(1) 数据管理中心负责组织全部范围的数据质量监控，定期或不定期发布数据质量监控检测结果。

(2) 数据管理中心负责组织全部范围的数据质量监控，定期或不定期发布全市、区数据质量监控检测结果，指导相关部门落实数据质量监控检测结果。

(3) 基层数据管理员负责督导落实数据质量监控检测结果。

第四条数据管理中心应建立相应的数据质量监控机制，采用人机结合等方式，以一定形式对数据质量实施监控审核，并及时发布监控结果。

第五条数据管理中心应于年度开始10日内制定年度数据质量监控检测计划，经相关领导、责任人批准后有计划开展全年数据质量监控检测工作，并报送备案。

第六条数据管理中心负责对数据质量监控发现的问题数据限期督导相关部门机构完成修正，一般最长要求5日内完成，并将修正结果报送审核。

第七条对数据质量监控发现的问题数据严格按照本办法数据维护规定的程序开展修正工作，任何人不得擅自修改。

(十二) 数据管理考核

第一条数据管理是组织信息化建设的重要内容，数据质量直接关系到组织信息化的应用效果。各部门机构应根据本办法所涉及的内容制定具体的数据管理考核办法，建立切实有效的数据管理考核奖惩机制。

第二条数据管理考核应遵循客观公正、公开透明的原则。各级可采取日常考核和定期考核相结合、人工抽检考核和机器全面检测相结合的方法，强化数据管理考核。组织应对数据管理工作实施百分制考核，主要考核项目包括：

- (1) 组织机构是否健全，岗位职责是否明确；
- (2) 数据录入是否及时，数据采集是否完整；
- (3) 数据内容是否准确，抽取展示是否按时；
- (4) 审核差错是否修正，数据维护是否及时；
- (5) 管理机制是否建立，安全措施是否到位。

第三条数据管理考核办法应突出数据录入、审核、维护、备份、安全等重点环节，进行指标量化。依据标准，对必录项的考核指标，在采集录入环节，其及时率、准确率、完整率必须达到95%以上；经过审核、修正后，其“三率”的考核指标应达到99%以上。

第四条各部门机构和部门人员发生下述行为之一的，组织应根据其危害程度、过错大小、情节轻重等，依据执法责任制和目标管理责任制追究相应责任：

- (1) 数据录入不及时，造成下一环节工作延误的；
- (2) 数据录入不规范、差错率连续居高的；
- (3) 未认真履行数据审核职责，造成大量数据差错不能及时纠正的；
- (4) 未及时进行数据维护，造成应用环节数据严重失真的；
- (5) 未按规定存储和备份数据，造成数据丢失的；
- (6) 未按规定进行系统维护，或因过失引发信息系统运行障碍造成数据丢失的；
- (7) 未按规定实施安全管理防范措施，造成计算机系统瘫痪或病毒侵害严重，严重影响数据传输和存储的；
- (8) 未按规定执行保密制度，发生重大泄密事件的；
- (9) 擅自对数据、权限进行修改、删除，或擅自使用他人的用户名、密码进行操作访问的；
- (10) 其他过错行为。

对有意破坏、恶意攻击税收应用信息系统，造成重大损失，涉嫌刑事责任的，移送司法机关处理。

数据质量评估办法

下面将从数据质量评估核心指标、数据质量评估模式、数据质量评估管理流程三个方面介绍数据质量评估方法。

数据质量评估的核心指标

数据质量问题频率

指标定义：数据质量问题频率 = 数据质量问题发生次数/存储的总数据量

指标单位：次/GB

根据数据质量评估指标将各数据仓库中的主数据及其历史行为划分为三个等级：

数据质量等级	描述	统计口径
一级	数据质量差，需要重点监控	数据质量问题频率大于等于 1 次/GB
二级	数据质量一般	数据质量问题频率大于等于 0.5 次/GB，小于 1 次/GB
三级	数据质量好	数据质量问题频率小于 0.5 次/GB

通过对数据质量问题频率的考评和等级划分，就可以从数据仓库众多的数据中解放出来，集中精力把有限的资源投入到需要重点关注的主题数据。因此数据质量可信等级是数据质量提高的有效途径。与之相配套的，必须建立了一套相关的管理制度，管理制度主要包括：

（1）可信等级初始值确立流程

数据中心平台各主题域的数据质量可信等级初始值一般设定为一级，由数据质量管理小组牵头，执行全面检查后报数据中心主管部门批准后确立。

（2）可信等级变更流程

数据质量管理小组每季度、每年组织定时抽查、全面检查时，每半年根据检查结果对相关主题域的可信等级提出调整意见，报数据中心主管部门批准后确立。

数据质量评估具体方法：

对于具体数据的质量检查模式采用记录数检查法、关键指标总量验证法、历

史数据对比法、值域判断法、经验审核法及匹配判断法。通过这些方法方法，可以对单个数据点的数据准确性进行检查，及时发现数据质量问题。

（1）记录数检查法

通过比较记录条数，对数据情况进行概括性验证。主要是检查数据表的记录数是否为确定的数值或在确定的范围内。

适用范围：

对于数据表中按日期进行增量加载的数据，每个加载周期递增的记录数为常数值或可以确定的范围时，必须进行记录条数检验。

（2）关键指标总量验证法

对于关键指标，对比数据总量是否一致。主要是指具有相同业务含义，从不同维度统计的汇总逻辑的检查。

适用范围：

同表内对同个字段从不同的维度进行统计，存在汇总关系时，必须进行总量检验。

本表的字段与其它表中的字段具有相同的业务含义，从不同的维度统计，存在汇总关系，且两张表的数据不是经同一数据源加工得到。满足此条件时必须进行总量检验。

（3）历史数据对比法

通过历史数据观察数据变化规律，从而验证数据质量。通常以同比发展速度进行判断。评估时应根据各种指标发展特点，重点对同比发展速度增幅（或降幅）较大的数据进行审核。历史数据对比法包括同比和环比两种方式。

适用范围：

不能进行记录数检查法、关键指标总量验证法，且事实表的记录数小于 1000 万条时必须进行历史数据对比法。

（4）值域判断法

确定一定时期内指标数据合理的变动区间，对区间外的数据进行重点审核。其中数据的合理变动区间范围是直接根据业务经验来确定的。

适用范围：

事实表中的字段可以确定取值范围，同时可以判定不在此范围内的数据必定

是错误的。满足此条件必须进行值域判断法。

（5）经验审核法

针对报表中指标间逻辑关系仅靠计算机程序审核无法确认、量化，或有些审核虽设定数量界限，但界限较宽不好判定的情况，需要增加人工经验审核。

适用范围：

以上方法都不适用的情况下，可以使用经验审核法。

（6）匹配判断法

与相关部门提供或发布的有关数据进行对比验证。

适用范围：

与有相关部门提供或发布的有关数据口径一致的，可以使用匹配判断法。

数据质量管控流程

（一）数据质量校验流程

数据仓库每天都有很多 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。

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。

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。

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) 定时数据抽查

数据校验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数据仓库的定期抽查。

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，可信等级为二级的二个主题的数据，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。

3) 全面数据检查

数据质量管理小组必须每年组织一次数据仓库的全面检查。

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。

（二）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。

(三) 数据质量的评估报告

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的生成相关的数据质量评估报告，数据质量报告分为两类：

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。

数据质量抽查或全面检查后提交的数据质量报告。

1. 每月数据质量评估报告

每月的数据质量报告是在每月月末或者下月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”的变化情况对各个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

2. 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个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

包括：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（包括存在问题）。

(2) 结论与建议

包括检查结果、可信等级调整建议等。

欢迎扫码关注
#志明与数据



公众号



视频号