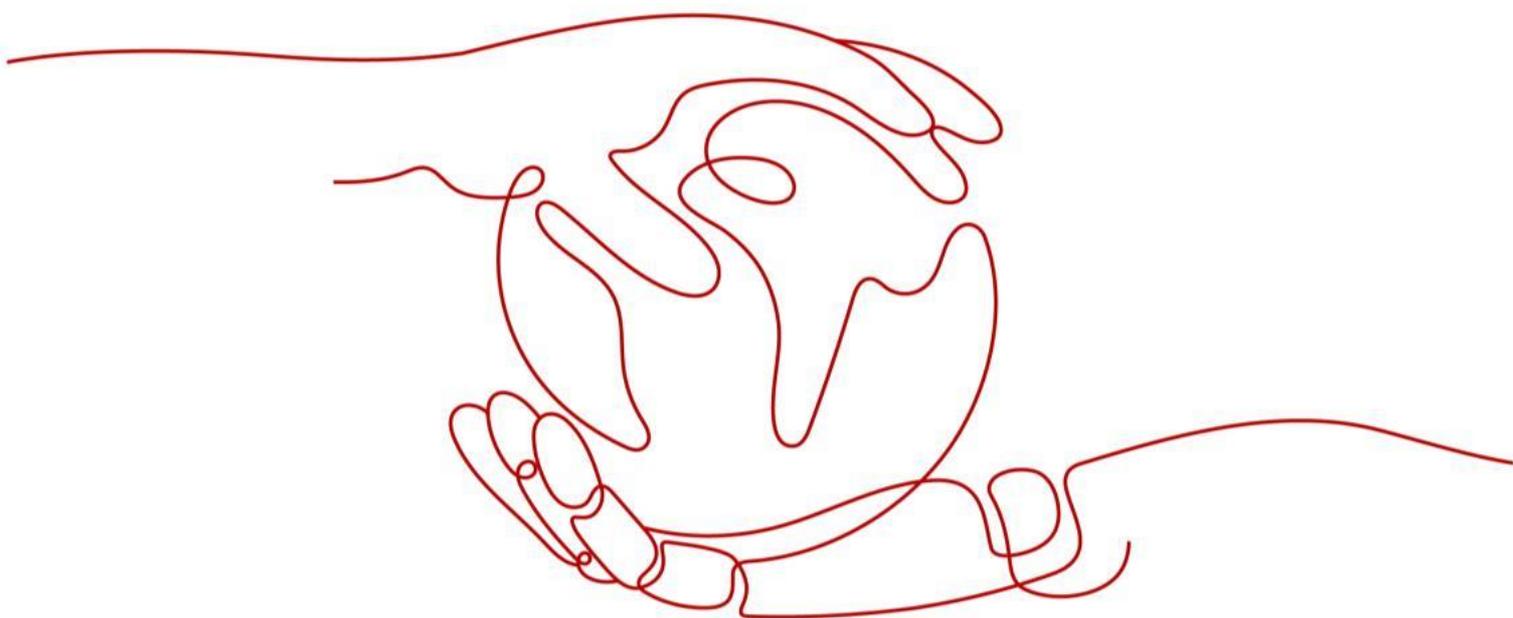


数据湖治理中心

# 数据治理方法论

文档版本 02  
发布日期 2021-03-31



华为技术有限公司



# 目录

<b>1</b>	<b>目的</b>	<b>1</b>
<b>2</b>	<b>面向的读者</b>	<b>2</b>
<b>3</b>	<b>数据治理框架</b>	<b>3</b>
3.1	数据治理框架	3
3.2	数据治理模块域	3
3.3	数据治理各模块域之间的关系	4
<b>4</b>	<b>数据治理组织架构</b>	<b>7</b>
4.1	数据治理组织架构框架	7
4.2	数据治理组织职责	7
<b>5</b>	<b>数据治理度量评估体系</b>	<b>10</b>
5.1	数据治理实施方法论	10
5.2	数据治理度量维度	11
5.3	数据治理度量评分规则	11
<b>6</b>	<b>华为数据治理案例</b>	<b>13</b>
6.1	华为数据治理思考	13
6.2	华为数据治理实践	14
6.3	华为数据治理效果	15
<b>7</b>	<b>新冠疫情数据治理思考</b>	<b>16</b>
<b>8</b>	<b>DAYU方法论产品落地</b>	<b>17</b>
<b>A</b>	<b>修订记录</b>	<b>19</b>

# 1 目的

数据是企业核心资产，企业需要建立起数据字典，有效管理其日益重要的数据和信息资源；同时建立数据持续改进机制，来不断提升数据质量。数据的价值和风险应被有效管理，以支撑企业管理简化、业务流集成、运营效率提升和经营结果的真实呈现。数据准确是科学决策的基础，数据架构和标准的统一是全流程高效运作、语言一致的前提。

当前企业数据面临很多的问题：没有统一的数据标准，各业务系统间数据无法充分共享，关键核心数据无法识别及跨系统无法拉通等。为有效管理企业数据资产，实现数据价值的最大化，急需建立一个完善的数据治理框架体系，为企业数字化转型打下坚实的数据基础。

DAYU数据治理方法论是华为数据管理方法论的精华总结，参考了业界数据治理最佳实践，并结合华为数字化转型成功经验优化而成。本文旨在帮助客户持续完善数据管理体系，沿企业主业务流打通信息链和数据流，提升数据质量，实现数据“清洁”，以支撑运营效率提升和经营结果的真实呈现，实现智慧数据驱动有效增长，充分实现数据资产价值。

本文先概要介绍了数据治理框架、数据治理组织架构和数据治理度量评估体系，然后结合华为数据治理案例和新冠疫情简单描述了数据治理的应用，最后描述了DAYU方法论在DGC产品上的落地以及详细的落地指导文档。

# 2 面向的读者

本文面向的读者包括：企事业各级管理人员、参与数据工作的相关人员、工具平台开发维护的相关人员等。

# 3 数据治理框架

## 3.1 数据治理框架

数据治理框架制定如下：

图 3-1 数据治理框架



## 3.2 数据治理模块域

数据治理主要专注于如下模块域：

- **数据集成**  
数据集成用来完成数据入湖动作，不是简单的数据搬家，而是按照一定的方法论进行数据备份。数据入湖的前提条件是满足6项数据标准，包括：明确数据 Owner、发布数据标准、定义数据密级、明确数据源、数据质量评估、元数据注册。此标准由数据代表在入湖前完成梳理并在数据治理平台上进行资产注册。
- **数据标准**  
数据标准管理着重建立统一的数据语言，L1到L5数据层级业务对象的定义是数据标准的载体，并对应发布包括L1到L5数据层级的数据标准。各业务对象对应物理

实现的IT系统需发布相应的数据字典并进行数据源认证。而对于梳理出来，但没有落IT系统的业务对象，需在后继的开发中进行数字化落地。

- 数据开发  
数据开发是编排、调度和运维的中心，数据开发是一个提供分析、设计、实施、部署及维护一站式数据解决方案，完成数据加工、转换和质量提升等。数据开发屏蔽了各种数据存储的差异，一站式满足从数据集成、数据清洗/转换、数据质量监控等全流程的数据处理，是数据治理实施的主战场。
- 数据质量  
数据质量管理的目标在于保证数据满足使用的要求。数据标准是衡量数据质量最基本的基准。数据质量要求各业务部门对相应数据领域的的数据质量全权负责，按业务需求设计数据质量标准，制定数据质量管控目标，并遵循企业数据治理要求进行数据质量度量，制定符合各自业务情况的数据质量政策及数据质量相关的改进计划，持续进行数据质量管控。
- 数据资产  
数据资产包括业务资产、技术资产、指标资产等。数据资产管理是数据治理的重要支撑手段，核心是构建企业级的元数据管理中心、建立数据资产目录、建立数据搜索引擎、实现数据血缘和数据全景可视。其中元数据包括业务元数据、技术元数据和操作元数据，要求将企业所有概念数据模型、逻辑数据模型以及物理数据模型系统化地管理起来，同时建设企业数据地图及数据血缘关系，为数据调用、数据服务、数据运营及运维提供强有力的信息支撑。
- 数据服务  
数据服务通过在整个企业范围统一数据服务设计和实现的规范并进行数据服务生命周期管理，集约管理数据服务并减少数据调用和集成的开发成本。
- 数据安全  
由于企业使用的数据资源，既有来自于内部业务系统，所有权属于企业的数据，同时也有来自外部的数据，必须将数据安全纳入数据治理的范畴，对所有企业数据要求依据数据安全等级定义进行数据安全定级，在数据产生、传输、存储和使用的过程中进行必要的数据安全访问控制，同时对数据相应的CRUD活动均需产生日志以完成安全审计。
- 主数据  
主数据管理是数据标准落地和提升数据质量的重要手段，是企业级数据治理的重要范畴，其目标在于保证在企业范围内重要业务实体数据的一致(定义和实际物理数据的一致)。主数据管理首先进行企业主数据的识别，然后对已识别主数据按照主数据规范要求数据进行治理和IT改造，以支撑企业业务流和工具链的打通和串联。
- 管理中心  
数据治理的开展离不开组织、流程和政策的建设，管理中心也管理着数据治理过程中公共核心的统一数据源、数据驾驶舱等，满足不同角色的用户拥有个性化的工作台。

### 3.3数据治理各模块域之间的关系

数据治理主要模块域之间的关系如下图所示：

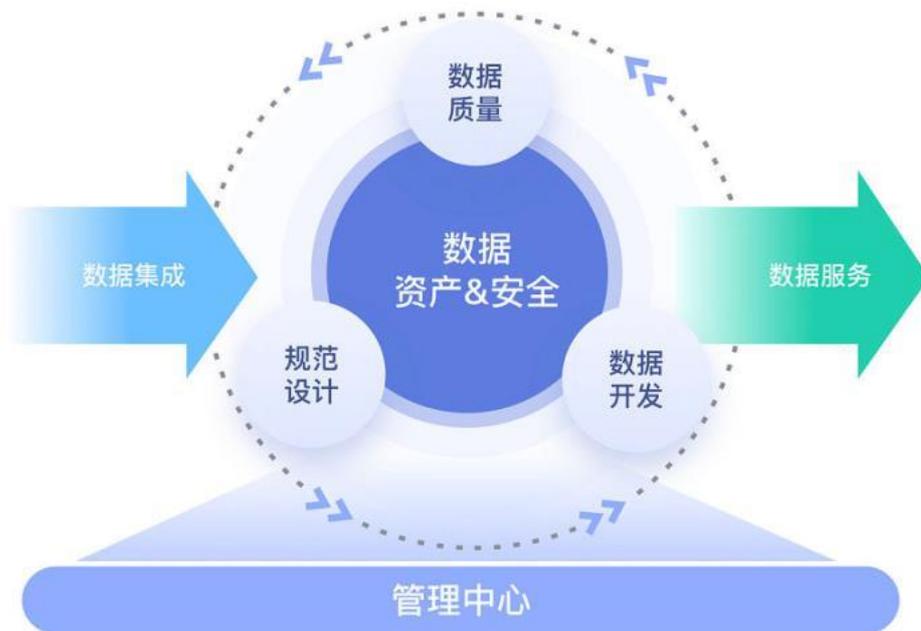
图 3-2数据治理各模块之间的关系



- 数据标准为数据开发和设计工作提供核心参考，同时也通过数据开发实现数据标准化落地。数据标准为数据应用提供“一致的语言”；数据标准为主数据管理提供主数据定义的基准，数据标准也是数据质量管理策略设计、数据质量规则制定的核心依据；数据标准所定义的密级分类和责任主体，为数据安全提供依据；数据标准也是数据资产管理重要输入。
- 主数据管理通过数据开发实现核心数据的一致性地记录、更新和维护，是数据质量提升的重要手段。主数据管理保障了数据应用和运营过程中核心数据的存在和一致性。
- 数据质量管理是数据应用和运营过程中数据准确性、一致性、完整性、及时性、唯一性、有效性的重要保障，是数据业务价值创造的重要前提。
- 数据资产管理模块完成元数据的采集和注册，数据资产管理为数据应用和数据消费提供了解数据的窗口。
- 数据服务管控实现在数据服务开发过程中服务标准、规范、要求和管理的落实，数据服务打通数据应用和数据消费的物理通道。
- 数据安全在数据开发过程中完成数据安全的IT实现，以达成数据应用过程中安全管理规范的要求。

以上工作的有效开展，离不开组织管理、明确的责任人、考核体系、流程制度、数据治理政策和数据治理平台的支撑。

图 3-3 DGC各模块之间的示意图



DGC平台提供了上述各个功能模块，满足数据接入、数据建模、标准化、质量监控、数据服务等全流程的数据管理动作。

# 4 数据治理组织架构

## 4.1 数据治理组织架构框架

数据治理可以采用集中化（全时投入）和虚拟化（部分投入）混合的组织模式。结合具备专业技能的专职数据治理人员和熟悉业务和IT系统的已有人员，在运作上实现数据治理团队的快速构建和能力导入，捆绑业务、IT开发和数据团队并利用已有人员熟悉度快速切入重点工作。

在工作内容和责任上具体到三层的工作组织：

图 4-1 数据治理工作组织



## 4.2 数据治理组织职责

在战略层面，由数据治理Sponsor和各部门负责人组成的数据治理领导小组制定数据治理的战略方向，以构建数据文化和氛围为纲，整体负责数据治理工作的开展、政策的推广和执行，并作为数据治理问题的最终决策组织解决争议，监控和监督数据治理工作的绩效，并确保数据治理工作预算支持。

数据治理委员会和各领域数据治理工作组是数据治理战略在运作层面具体的实施团队。其中：

- 数据治理委员会：由数据治理负责人、数据治理专家和数据架构专家团组成，面向企业进行数据治理工作的统筹并提供工作指导，在整个企业范围定期沟通数据治理工作，形成数据质量精细化管控文化。根据数据治理领导小组的愿景和长期目

标，建立和管理数据治理流程、阶段目标和计划，设计和维护数据治理方法、总则、工具和平台，协助各数据领域工作组实施数据治理工作，对整体数据治理工作进行度量和汇报，并对跨领域的的数据治理问题和争议进行解决和决策。

- 各领域数据治理工作组：在各领域数据范围内进行数据治理的工作，依据数据治理委员会制定的数据治理方法和总则，制定本领域数据治理目标和工作计划，负责领域数据资产的看护，维护更新相应数据标准和及相关元数据，设计本领域数据度量和规则，监控收集数据质量问题并持续改进提升，主动升级数据相关问题。最终完成领域内数据资产的看护，并支撑数据治理目标的达成。

领域数据治理工作组由数据Owner、数据代表、数据管家、数据专员和数据架构师组成。其中：

- **数据Owner (Data Owners)**：数据Owner是领域数据治理工作的责任人。
  - 制定本领域数据治理的目标，工作计划和执行优先级。
  - 建立数据治理责任机制，将本领域的的数据治理工作分解到工作组成员,并跟进及管理工作组工作。
  - 设计数据质量规格，承接数据需求，对数据问题及争议进行裁决。
  - 建设和维护本领域信息架构。
  - 建立和推动领域数据文化和氛围。
- **数据代表 (Data Representatives)**：数据代表是领域数据治理工作的专家带头人。
  - 深刻理解数据工作的目标、方法、规则、工具，并通过识别关键业务流程和IT系统，对本领域数据治理的路标和工作计划进行细化并排序，最终管理执行。
  - 作为本领域数据治理专家，管理并解决问题和争议，必要时提交数据Owner进行裁决。
  - 对业务环节数据的完整性、及时性、准确性、一致性、唯一性、有效性负责，确保行为即记录，记录即数据，并依据数据质量规格对本领域数据进行度量和报告。
  - 落实本领域信息架构的建设和遵从，负责本领域数据资产的看护，维护相应数据标准和数据目录并更新发布。
  - 承接上下游数据需求，并主动根据领域内应用场景和业务需求识别数据需求，对需求的实现进行推动和管理。
  - 依据相关规定定义本领域数据安全等级，并进行数据授权管理。
- **数据管家 (Data Stewards)**：数据管家是领域数据治理工作的协助者。
  - 确保领域治理工作的流程和内容规范，符合数据治理要求。
  - 协助数据代表进行问题跟踪和解决。
  - 梳理、维护并更新领域数据元数据（业务对象、数据标准、数据模型）。
  - 推广和维护数据治理工具和平台在本领域的应用。
- **数据专员 (Data Specialists)**：数据专员是领域数据治理工作的专家团队。基于本领域数据治理的工作计划，利用数据专项技能，支撑数据代表完成数据Owner分配的各类数据治理工作。
- **数据架构师 (Data Architects)**：数据架构师是领域数据治理工作在IT层面的代表。
  - 开发和维护本领域的的数据系统或子系统，确保数据在系统中得以记录，数据标准、数据质量规则、数据安全、主/参考数据管理、数据服务在系统中得以实施。

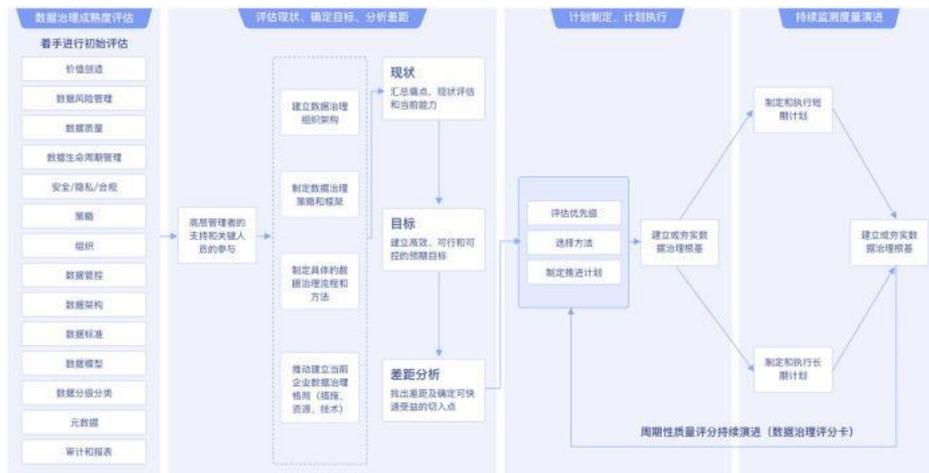
- 提供数据系统的数据相关信息（元数据、数据字典、数据血缘）。
- 协助执行IT相关的数据治理工作。
- 确保数据系统的技术方案符合本领域的信息架构，技术选择能够满足数据发展中长期的需求。

# 5 数据治理度量评估体系

## 5.1 数据治理实施方法论

数据治理实施方法论按照数据治理成熟度评估->评估现状、确定目标、分析差距->计划制定、计划执行->持续监测度量演进的关键实施方法形成数据治理实施闭环流程。

图 5-1 数据治理实施方法论



这也遵循了PDCA（Plan-Do-Check-Action）循环的科学程序，同时结合数据治理工作的特点设计了两个层面的度量评估：

度量评估工具	度量评估对象	度量评估方法	度量评估频次
数据治理成熟度评估	企业整体	调查问卷	年度
数据治理评分卡	各业务及IT部门	数据治理工作组与各业务及IT部门共同打分	季度

## 两个层面的数据治理度量评估工具

通过年度的整体数据治理成熟度评估，了解各维度数据治理现状，并制定可操作性目标，分析差距，制定切实可行的计划，在推进落实计划的过程中，利用季度性实施的数据治理评分卡，针对性地监测度量各业务/IT部门的数据治理情况，持续推进各部门的数据治理水平提升，进而提高整体数据治理成熟度。

年度进行的整体数据治理成熟度评估，可以结合主观及客观调查度量，综合反映企业当前数据治理成熟度水平，帮助制定切实可行的数据治理整体目标。季度性实施的数据治理评分卡是一个重要的工具手段，从一些关键核心维度进行季度性迭代评分，持续促进所有部门数据治理工作的落实提升。

## 5.2数据治理度量维度

数据治理成熟度评估使用调查问卷方式，包括11个治理模块，60多个调查问卷事项，全方面多维度度量了数据治理工作的水平。

数据治理评分卡使用评分卡形式，由数据治理组织和各业务IT部门共同针对各部门自身数据治理情况进行打分。评分卡是季度性打分，作为一个工具手段持续推动促进各部门的数据治理工作，改进数据质量，提升数据治理水平。

图 5-2数据治理评分卡



## 5.3数据治理度量评分规则

数据治理成熟度评估分5个级别：

- 级别1是最低级别，数据治理处于初始级别，空白状态，没有相关流程及数据治理管控。
- 级别2是开始管理级别，仍处于匮乏状态，没有形成正式一致的数据治理方法。
- 级别3是明确定义级别，处于可改进状态，已经形成了主动治理的流程和正式一致的数据治理方法。
- 级别4是量化管理级别，已经步入适宜匹配状态，贯穿组织采用的正式一致的数据治理方法都是可量化管控的。
- 级别5是最高级别，专注于持续优化提升，已经步入卓越状态。

图 5-3数据治理成熟度级别分层



数据治理评分卡对各维度度量是按5分制进行评分，每一个度量细项只有得分0或得分1，如果满足度量细项要求得分1，否则得分0。

图 5-4数据治理评分卡数据执行维度度量评分计算范例

流程域	数据质量	数据质量	数据质量	数据质量	数据质量	数据安全	对象得分
7.10交付项目管理	1	1	1	0	0	1	3.3

Cover | 度量结果 | 组织 | 数据架构 | 数据标准 | 平台工具 | 数据执行 | ...

如上示例，针对流程7.10交付项目管理，数据执行维度有6个度量细项，则这一度量维度的总体得分就是SUM(度量细项得分)\*5/6。所以这一度量维度的实际得分就是(1+1+1+0+0+1)\*5/6=3.3。

# 6 华为数据治理案例

## 6.1 华为数据治理思考

华为在发展中也遇到了如下的数据问题：

- 数据管理责任不清晰，造成数据问题无人决策解决；
- 数据多源头，造成数据不一致，不可信；
- 数据大量搬家造成IT重复投资；
- 数据无定义造成难于理解、难于使用；
- 各部门发布报告，统计口径不一致，困扰业务决策；
- 数据形态多样化，数据量迅猛增长，数据处理逻辑复杂，投资大；

华为在数字化转型过程中，解决了上述问题，因为华为认识到只有建立了完整的数据治理体系，保证数据内容的质量，才能够真正有效地挖掘企业内部的数据价值，对外提高竞争力。

### 高质量数据是业务创新的基础

企业在市场中的竞争领域已经从同一领域市场份额争夺，发展到开发新竞争领域的创新性竞争阶段，这从客观上对企业的创新能力提出了更高的要求，现在企业的创新在很大程度上要借助科技的手段，在业务数据的开发和利用基础上进行创新，数据为企业实施有效的创新提供了丰富强大的动力。

企业数据繁杂无标准质量低

企业的IT系统经历了数据量高速膨胀的时期，这些海量的、分散在不同角落的数据导致了数据资源利用的复杂性和管理的高难度。企业无法从统一的业务视角去概览整个企业内部的数据信息。暴露出来的只是一个个独立的系统，系统与系统之间的关系、标准数据从哪里获取都无从知晓。

### 数据是企业争夺优质客户的关键

数据是企业的生命线，谁掌握了准确的数据谁就获得了先机。在当前竞争日益激烈的市场上，企业都在不同的细分市场上争夺优质客户。如何在这样的市场环境中选择市场的经营策略？企业每一笔资金的来源与利用、每一次经营管理决策都必须基于准确的数据分析判断。只有基于准确的数字，才能够帮助企业在激烈的竞争中取得竞争优势。

图 6-1 华为数据治理工作思考



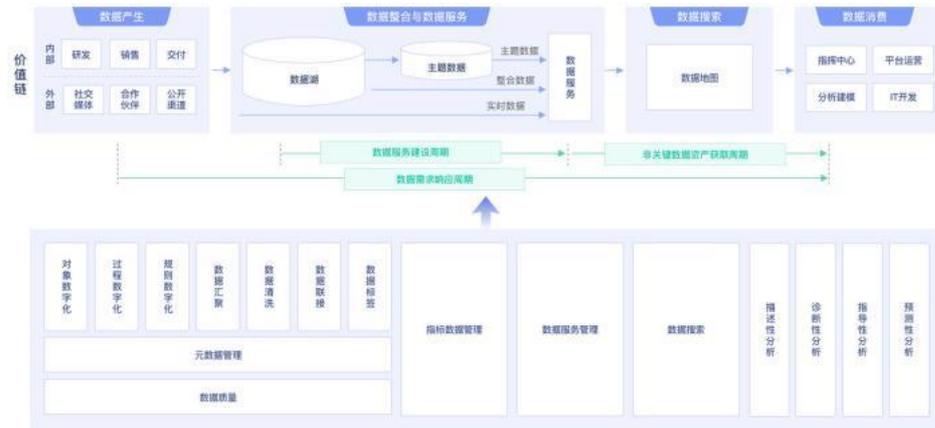
## 6.2 华为数据治理实践

华为数据治理的规范流程建设，完成了从数据产生、数据整合、数据分析与数据消费全价值流的规则制定。华为数据治理组织实践，建立实体化的数据管理组织，虚线向公司数据管理部汇报，同时组建了跨领域数据联合作战团队。华为已建立统一的数据分类管理框架，指导各领域进行分类管理。华为信息架构框架，通过政策发文明确信息架构的定义和构成要素，在公司层面建立统一的架构方法。基于ISO8000标准，华为建立了数据质量管理框架和运作机制，每年例行开展两次公司级数据质量度量，从“设计”与“执行”两个方面度量数据质量，由公司数据Owner定期发布公司数据质量报告，牵引各业务领域持续改进数据质量。

图 6-2 华为数据治理愿景与目标



图 6-3 华为数据治理能力



### 6.3 华为数据治理效果

以财经为例，在数据治理前存在很多问题，如由于IT系统的烟囱式建设，导致一个角色跨多个IT系统操作，效率低；数据获取难，手工处理多，单一个收入管理需要从5个系统导出数据，约11个人总共花费50小时完成分析。

通过数据治理，华为可以做到3天月度财务报告出初稿、5天月度财务报告出终稿、11天年度财务报告完成初稿，要把一家业务遍及170多个国家和地区的全球化公司的“总账”算清楚可真不容易。要做到财报的高效、准确、完整，离不开高度集成的全球结账管理系统，更离不开数据治理，通过交易核算自动化、ERP优化、数据调度优化、数据质量监控以及提升数据分析平台的性能，华为实现了全球核算实时可视，过程可跟踪、可管理。

图 6-4 华为数据治理实践



# 7 新冠疫情数据治理思考

中共中央政治局常务委员会2月3日召开会议强调：这次疫情是对我国治理体系和能力的一次大考。全民抗疫的过程中，缺乏有效数据支持就是缺乏对于疫情的快速响应机制、缺乏对于医疗资源合理调配的能力。

通过数据治理，我们可以构建开放通用的数据采集接口，提高数据采集效率；统一数据标准，轻松融合数据；建立跨平台的数据提取和数据追溯，实现开放共享，打通信息孤岛；保护隐私数据，构建可信数据。

以“健康码”为依托的各类政务小程序，正在加速提升市民信息化应用能力，基于数据治理、数据共享，很可能在不久以后，我们就可以实现“一码走天下”。将线上疫情应急治理体系转变为常态化管理体系，降低日常行政成本，发挥更大的数据流动与服务价值。通过自主申报健康信息，民众可获取反映健康状况的健康码，并凭码通行社区、办公楼、交通口等核验场景。在防疫初期，健康码有助于防止线下扎堆填报信息带来的病毒传染风险，一改政府人海战术排查病患的做法；随着疫情得到初步控制，健康码实现的人员行程追踪、同行密切接触人员自查等功能亦助力企业复工复产。随着各地复工复产加速，跨区域人口流动为疫情防控带来新的挑战，统一各层级政府和不同部门的数据标准，共享确诊、疑似病例、密切接触者、县域风险等级数据库是当下数据治理的关键。

# 8 DAYU方法论产品落地

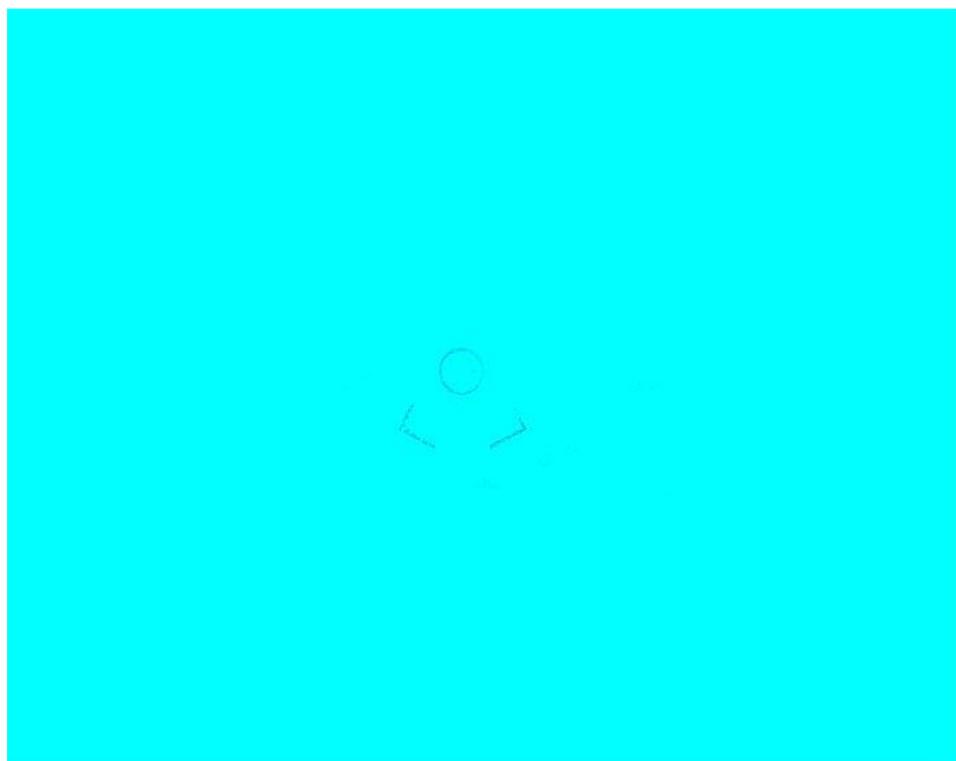
DAYU数据治理方法论已经在华为云云服务数据湖治理中心DGC上落地实现，包括流程落地和功能落地。流程落地是指有一套详细的流程规范（需求、设计、实施、验证、发布等阶段）指导用户使用DGC开展数据治理工作；功能落地是指DGC平台提供自动化、智能化的工具帮助用户高效完成数据治理工作。

DAYU数据治理方法论还有个完全版本，详细描述了流程落地和功能落地内容，欢迎咨

图 8-1 方法论流程落地



图 8-2 方法论功能落地



# A 修订记录

发布日期	修订说明
2020-03-25	第一次正式发布。