



基于流程全生命周期的大型企业数据治理体系研究与实践

王倩, 徐丽超, 汪莹, 金多多, 杜婷
(中国电信股份有限公司上海分公司, 上海 200120)

摘要: 基于数字化转型过程中, 运营商企业数据治理的问题、难点进行了分析, 提出了对于大型企业数据治理需要抓住源头、面向流程全生命周期开展数据治理的理论, 并通过具体实践加以运用总结了数据治理的流程、方法、系统建设的关键方案, 同时对主数据管理的技术方案进行了比较和选型, 阐述了软件周期与数据周期各阶段的关系, 给出了企业数据运营质量提升的实际解决方案。

关键词: 数字化转型; 数据治理; 流程全生命周期; 主数据管理; 企业数据运营

中图分类号: TP393

文献标志码: A

doi: 10.11959/j.issn.1000-0801.2025090

0 引言

随着企业的数字化转型, 越来越多的企业意识到数据驱动是实现数字化转型的关键能力, 通过对现有的大数据平台架构进行优化, 从而实现数据驱动能力的提升。然而很多大型企业平台中数据体量庞大, 数据存在冗余和错乱, 应用中存在主数据管理职责不清晰、标准不统一等问题, 导致数据分析时产生误差。因此, 亟须开展数据治理工作, 强化主数据应用管理机制, 保障数据语义的一致性和准确性。同时, 面对系统数量多、接口稳定性凸显的问题, 需要强化上下游两级运营, 构建良好的治理机制来降低日常数据运营中的内耗, 提升企业数据运营质量。

1 数据治理问题的源头

数据产生的问题往往会在系统运营阶段暴露出来, 但是问题源头却不在运营。在建设规划阶

段, 往往侧重功能和流程的实现, 却忽视了系统、新建模块、新建流程的数据汇聚、分析、呈现、沉淀和能力开放的数据规划。在需求阶段, 往往只侧重功能、流程、用户和权限, 忽略了对数据支撑的分析和对数据对象的说明。在设计阶段, 目前多数关心功能、流程和系统割接, 缺乏数据接口设计、数据沉淀和共享方面的设计。在开发阶段也需要遵循数据接口规范、数据存储结构和共享开发实施, 并进行检查、测试和验证。在上线评审阶段, 往往忽视了数据方案的评审, 只注重功能上线。例如, 如果在开发跨系统接口时, 未完整考虑对于失败的调用上游是否重发、下游是否有幂等处理、重发几次、仍旧失败后采取哪种运维手段进行分析和补救达到日清日结等要素, 就会直接导致运营阶段出现系统间数据不一致的问题。因此, 数据治理的问题源头不在数据, 而在软件, 必须建立“面向流程全生命周期”的数据治理体系, 才能从源头解决问题。



2 开展面向流程全生命周期的数据治理工作

2.1 建立“数据三同步”的概念和工作方法，建立源头治理流程

目前，软件开发要遵循安全三同步的概念已深入人心。在开发建设过程中忽视数据工作，给交付运营带来了极大隐患。实际上，数据工作也与安全工作一样，要开展同步规划、同步建设和同步运营的三同步工作，并且在软件生命周期的各个阶段都必须有相应输出物来监督，确保执行到位，才能为企业数据运营质量提升打下基础。软件周期的规划建设、需求分析、方案设计、开发实施、系统上线、交付运营的各阶段和需要输出的数据输出物流程如图1所示。

2.2 从数据流程全要素出发，构建数据治理体系

要从实际工作需求出发，从岗位配置、规范、标准、流程、制度、细则和工具平台等要素角度梳理数据的薄弱点，明确重点。从完整的数

据流程全要素视角出发，在工作中全面覆盖，形成完善的治理体系，数据治理体系流程如图2所示。

根据数据管理成熟度初步测算，目前运营商的数据管理成熟度阶段多为已定义级，需要确定目标，以完善主数据管理体系来驱动企业的数据治理，并将数据管理成熟度普遍提升到已管理级、部分提升为优化级。各级数据管理成熟度如图3所示。

2.3 从数据周期出发，建立数据技术框架

应用软件从规划、需求、设计、上线到运营、迭代、下线的全生命周期管理，建立用户自服务体系。以企业级数据平台为基础，构建数据治理的技术工具，逐步建立全局数据治理工作框架，如图4所示。

3 数据治理需要构建的关键技术平台

3.1 建设主数据管理平台

构建主数据管理工作台，统一数据管理入口。完善统一视图，主数据变化的实时捕捉，管理流程实时触发。例如，通过工作台进行操作捕



图1 软件周期治理流程

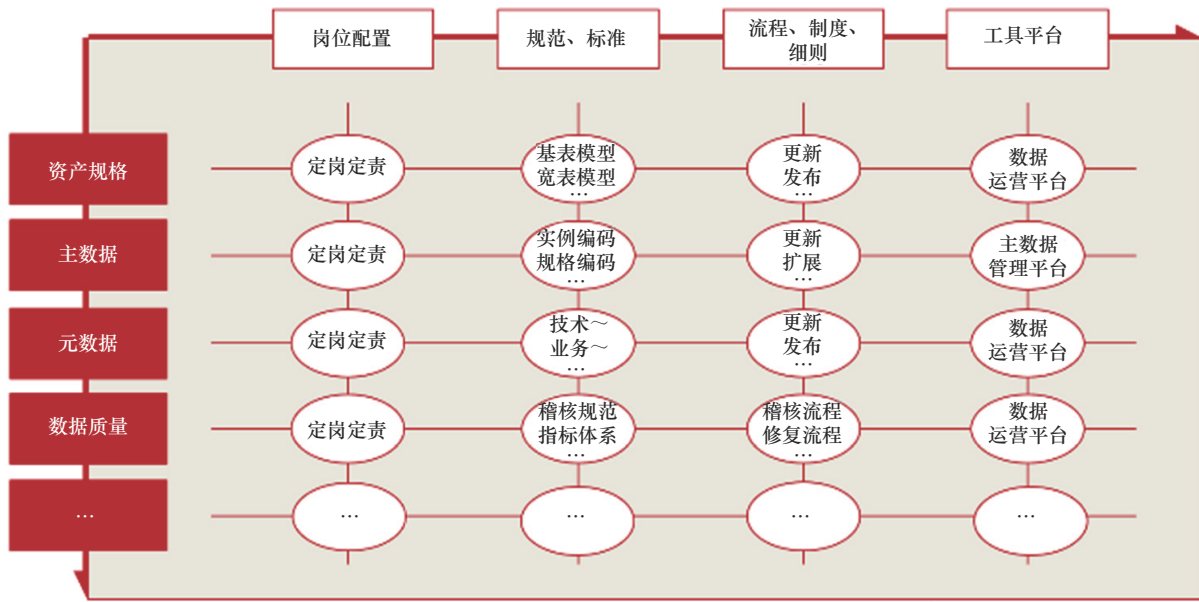


图2 数据治理体系流程



图3 各级数据管理成熟度

捉、后台变化捕捉。上下游主数据编码完全一致，使用系统原则上应直接使用属主系统的ID，若使用系统必须作映射，映射ID仅作系统内部使

用。在传递给下游系统时，务必携带属主系统的原始ID，数据管理提供监测分析能力。主数据管理平台如图5所示。

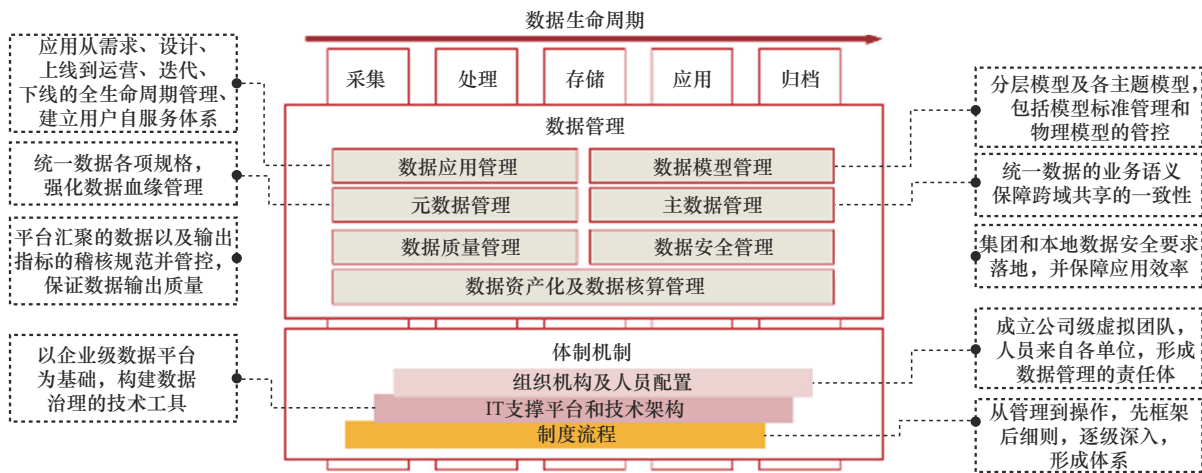


图4 数据治理工作框架

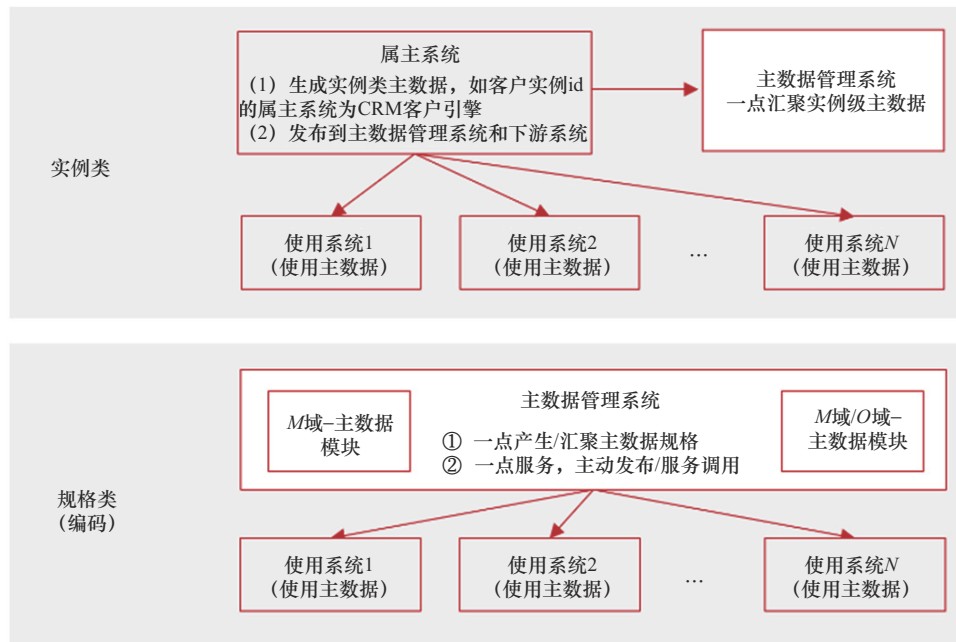


图5 主数据管理平台

在实际建设时，首先要统一与主数据、数据管理有关的术语，减少建设沟通时的歧义，统一术语如图6所示。

其次，要厘清运营商内部的业务元数据和物理元数据，完成管理模型设计，如图7所示。

同时，确定主数据对象、主数据实例，形成数据资产卡片，如图8所示。

主数据平台需要从开发设计、业务准备、生产运营各阶段构建起完善的流程环节，把技术部

门、业务部门、生产部门人员的积极性都调动起来。

(1) 开发设计阶段。技术人员需要通过平台对属主系统和业务方进行数据模型设计过程（同时识别是否为主数据，定义主数据对象、属性、属性值的规格）。业务描述举例：增加“产品”主数据对象、属性、属性值的业务元数据描述。物理实现举例：通过 product、prod_attr、prod_attr_value3 张表描述物理元数据。通过各域

统一术语	曾用名	解释
业务元数据	业务元描述	数据的业务内涵、指标口径等，例如：宽带、速率、接入方式
物理元数据	数据模型、数据表结构	刻画数据的物理方式，一般到Table、column这个级别
主数据对象	主数据、主数据名称	具体的实体对象，例如：支局、物理网络、产品
主数据实例	实例主数据	具体的数据记录，Row这个级别
数据资产卡片	数据资产、数据模型	System、Dbbase、Table、column，数据模型的颗粒度同物理元数据

图6 统一术语

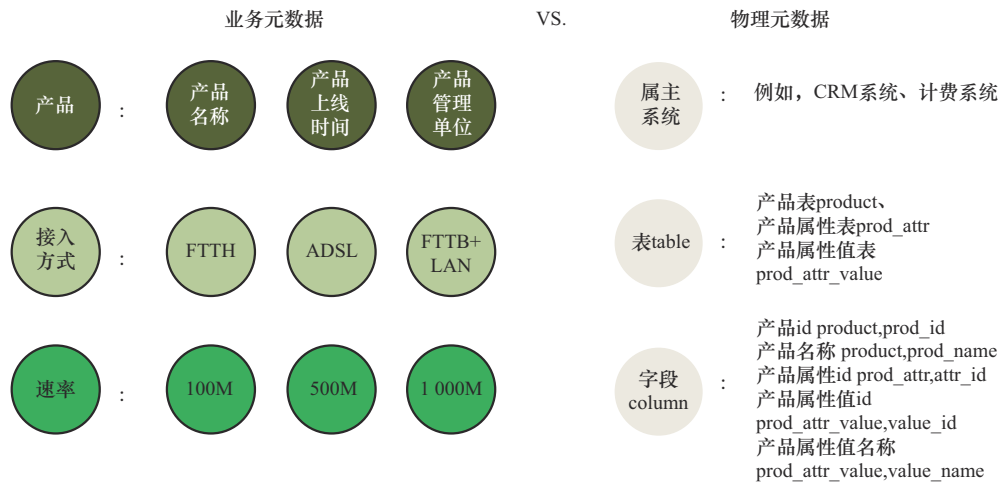


图7 运营商内部的业务元数据和物理元数据

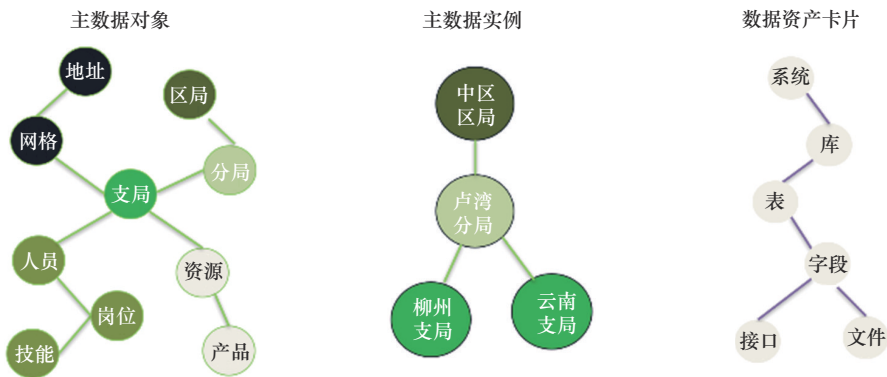


图8 确定数据对象、主数据实例,形成数据资产卡片

元数据管理平台或后台完成模型设计，如果识别为主数据对象，则该数据对象需要归集到主数据管理平台。

(2) 业务准备阶段：业务人员作为数据归口部门要通过平台来配置增加主数据对象编码。例如，增加“宽带”，需要在 product 表中增加



1000001038; 在 Prod_attr 表中增加速率、付费方式两个属性; 在 Prod_attr_value 中增加 1 000 Mbit/s、500 Mbit/s 两个速率; 后付费、预付费两个付费方式。通过各域主数据管理界面配置主数据编码, 需要归集到主数据管理平台。

(3) 生产运营阶段: 生产部门要合规使用主数据(形成实例的过程)。例如, A 用户申请宽带, 形成宽带实例 1; B 用户申请宽带, 形成宽带实例 2。各域生产过程中产生实例, 其中关键实例需要归集, 非关键实例不属于主数据管理范围, 不需要归集。可有效减少主数据平台与属主系统交互压力, 提高效率。

在具体功能实现方面, 要提供一屏展示、数据出生证管理、发现管理、订阅管理等功能。

(1) 主数据管理平台建立一屏展示功能, 采集各域主数据、完成全局主数据对象、编码、属性、属性值的一屏展示, 明确各部门的工作职责。主数据采集管理的具体方案需要进行研究, 目前已尝试了 3 种可行的方案, 并在图 9 中从部署效率、便捷性、完整性、安全性各方面进行了方案比较, 大家可以从自身的需要出发进行选型。

(2) 实现各属主系统在新增主数据对象和核心主数据编码时, 先向主数据管理平台申请统一编码审核后方可在属主系统创建, 然后属主系统将主数据规格同步至主数据管理平台。避免重复构建主数据, 下游跟随属主系统, 实现事前管控。

(3) 实现发现管理功能。当发现缺失的主数据信息, 则开展数据治理工作, 主动提示数据归口管理部门人员, 由属主系统配合补全。

(4) 提供订阅功能, 可通过界面/接口订阅, 并向订阅了主数据、主数据映射的系统分发数据。使用系统根据各自需求向数据资产管理平台发起主数据订阅, 主数据管理平台对使用系统提供订阅查询功能。

3.2 建设系统间接口数据传递方案

首先要建立日常闭环接口机制, 这一点往往容易被忽视。所有系统要进行信息交互, 要按具体业务场景来确定哪一种交互方式最合理。常使用的交互技术一般分为文件共享技术、数据库中间表技术、点对点接口技术、消息队列技术、服务总线交换技术、ETL 数据交换技术和数据采集交换技术。这 7 类数据接口的数据传递规范, 应

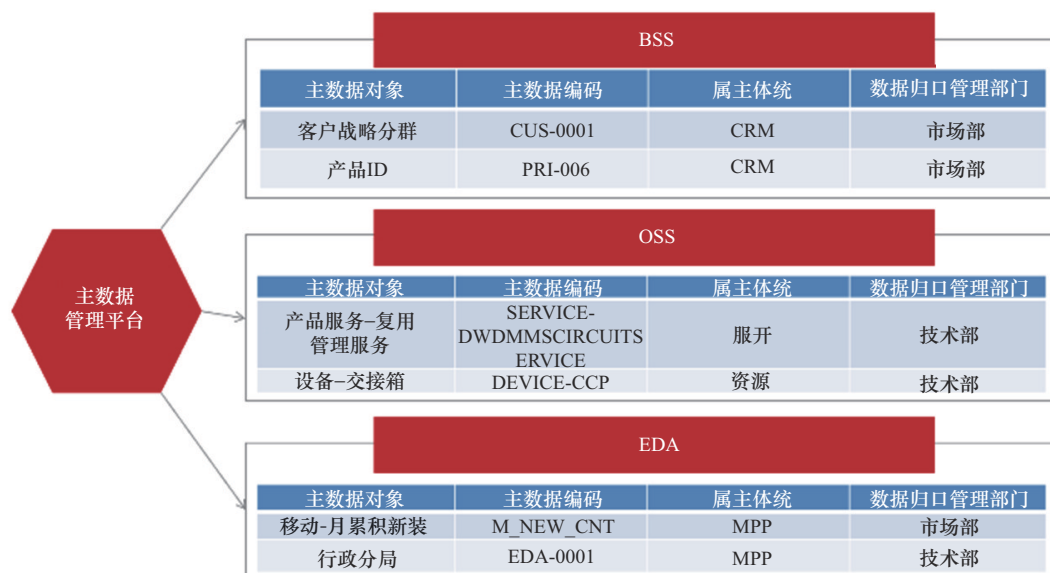


图9 3种主数据采集管理的可行方案

该同时设计在每个业务场景下，何种数据传递适合哪一种接口方式、以及闭环运营方式。

对接收数据的下游系统而言，要增加唯一性、幂等性、完整性校验机制、拒收机制、单笔回滚机制。对提供数据的上游系统而言，要完善重发机制。并要参照通信协议，形成系统间的单笔交易闭环，保持全程一致性的系统闭环、握手流程。同时，生产运营部门要建立日清日结的工作标准和工作岗位。以运营商的计费系统接口举例说明数据交互方式如图10所示。

上游系统	下游系统	数据内容	数据交互方式
计费	营服协同	实时短消息	点对点实时接口
计费	营服协同	实时短消息	文件共享方式
计费	OSS	停复机指令、断网指令	实时接口调用
计费	移动核心网	批价健全	在线消息方式
计费	CQD	免费量话单(历史月)	文件共享方式
计费	CQD	共享免费量(当月)	文件共享方式
计费	CQD	用户资料(所有)	消息队列方式
计费	CQD	清单外送(6+1个月)	文件共享方式

图10 举例说明数据交互方式

构建数据质量稽核修复平台。由于数据体量庞大，要构建基于云和大数据平台的数据稽核平台，实现自动化比对和修复功能，才能提升数据质量，主要功能包括以下5点。

(1) 基于大数据架构，将数据源模型导入组

件模型(通过大数据预处理平台统一入库)。

(2) 基于云分布式架构核心自动稽核比对流程管理功能，包括自动调度比对任务、自动执行比对流程。

(3) 需要基于比对模型构建知识库分析管理功能，包括知识库模型可配置，智能分析分类比对结果等。

(4) 大数据比对模型要具备时间维度差异分析功能，来追溯重点差异发生历史时间段和批量差异波动历史。

(5) 需要有数据源配置、稽核点配置、稽核任务配置，基AI的配置、审核、发布功能。

在实践中，还需要扩大数据比对范围、加深数据比对精细度，建立跨多个团队的数据比对修复流程(包含业务原因、系统原因的差异)，提高工作效率。数据差异修复流程如图11所示。

同时，要明确各业务部门、技术部门、生产部门各岗位职责，形成企业数据运营管理办法，才能长效提升企业数据运营质量。主数据采集管理技术方案比较见表1。

4 结束语

产业数字化正以前所未有的速度发展，千行百业都在通过软件系统大量产生数据，企业数据

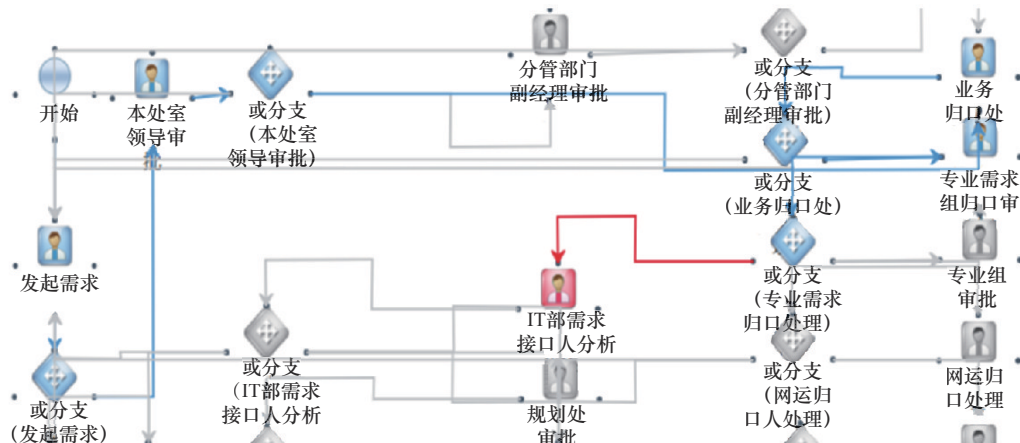


图11 数据差异修复流程



表1 主数据采集管理技术方案比较

	方案一	方案二	方案三
主数据采集管理技术方案简述	采取属主系统对接方式，有针对性地对接各域中的各个属主系统，仅获取主数据对象的物理元数据和主数据实例	采取全量系统对接方式，对接各域中的各个系统，获取物理元数据、抽样的数据实例。并从数据安全考虑，采集接口按需开启，采集后及时关闭	采取各域本机部署脚本采集方式，由主管理系统开发统一脚本（通过安全测评），而脚本由各域系统在本机部署，扫描后仅将物理元数据结果上传平台，且由各域系统自行控制定时任务执行计划
比较项1：部署效率	需要开通与属主系统对接网络，有接口开发工作量，部署较慢	需要开通与全量系统对接网络，有接口开发工作量，部署最慢	不需要开通网络，部署最快
比较项2：采集便捷性	主数据管理平台主动采集，且仅采集物理元数据，较为便捷	主数据管理平台主动采集，可自由决定采集的物理元数据和抽样实例范围，且接口可按需控制开关，最为便捷	由各域系统决定采集执行计划，较为不便
比较项3：数据完整性	较为完整	最为完整	较为完整
比较项4：安全性	网络打通且直连，安全风险相对较高	网络打通，但设置开关，较为安全	网络隔离，最为安全
比较结论		从数据完整性、灵活性考虑，优选方案二	从安全性和部署效率考虑，优选方案三

的高效、高质量运营已经时不我待。如今我们津津乐道的产业数字化是为了明天的数字产业化打基础。运营商作为先行涉足数据的大型企业之一，有着类别丰富的、规模可观的软件系统的实战经验。因此，要积极探索各类软件系统规律和数据运营规律。

今后还将进一步总结在线交易性、计算密集型、数据分析型、结构一体化等各类软件系统的特性、数据沉淀的特性，继续强化社会数据治理体系，为社会数字交易的发展做出贡献。

[作者简介]

王倩（1977-），女，中国电信股份有限公司上海分公司工程师，主要从事运营商MBOSS及大数据系统的架构设计与开发技术的研究和实践工作。

徐丽超（1987-），女，中国电信股份有限公司上海分公司会计师，长期从事事业财融合数据治理方案的研究工作。

汪莹（1983-），女，中国电信股份有限公司上海分公司会计师，长期从事事业财融合数据治理方案的研究工作。

金多多（1992-），女，中国电信股份有限公司上海分公司会计师，长期从事事业财融合数据治理方案的研究工作。

杜婷（1991-），女，中国电信股份有限公司上海分公司会计师，长期从事事业财融合数据治理方案的研究工作。