

关于数据中台建设思路的探讨

Discussion on Construction of Data Middle Platform

张 洁,许建宏,肖 伟(中国联通研究院,北京 100048)

Zhang Jie,Xu Jianhong,Xiao Wei(China Unicom Research Institute,Beijing 100048,China)

摘 要:

复杂大型企业传统的IT架构,在当前市场快速发展变化的环境下,已无法满足对业务需求的高效支撑,许多先进企业逐渐向前中后台的平台式架构转型。数据中台旨在全面汇聚并共享企业信息,以数据智慧化驱动业务变革。通过分析数据中台的建设意义、适用场景、先进企业实践情况,提出了一种数据中台的IT建设方法,并找出了一系列促进数据中台发挥效果的非技术性关键因素。

关键词:

数据中台;数据开放;IT架构

doi:10.12045/j.issn.1007-3043.2021.08.015

文章编号:1007-3043(2021)08-0074-06

中图分类号:TN919

文献标识码:A

开放科学(资源服务)标识码(OSID):



Abstract:

With the rapid development of economy and market environment, the traditional IT architecture in complex enterprises is hard to support business requirements effectively. Many advanced enterprise are gradually transforming to the platform architecture of the front, middle and back platform. The data middle platform aims to gather and share the enterprise information comprehensively, and drive business change with data intelligence. By analyzing the construction significance, applicable scenarios and the practice of advanced enterprises, it puts forward an IT construction method of data middle platform, and finds out a series of non-technical key factors to promote the effect of data middle platform.

Keywords:

Data middle platform; Data opening; IT architecture

引用格式:张洁,许建宏,肖伟. 关于数据中台建设思路的探讨[J]. 邮电设计技术,2021(8):74-79.

0 前言

当市场不再处于空白期、快速发展期,而是进入了增长放缓、存量维系阶段时,企业之间的竞争力就主要体现在服务质量和运营成本等方面,此时,IT系统的智能化支撑尤为重要,而这严重依赖于企业各方面数据信息的掌握、贯通、应用甚至预测。数据中台应运而生,很多知名先进企业早已开始了数据中台的尝试。本文将从建设数据中台的意义、数据中台在IT架构中的定位、数据中台功能架构及数据中台建设的

关键要素等几方面,论述数据中台建设思路。

1 建设数据中台的意义

在传统的烟囱式IT架构中,各系统从后端到前端相互独立,紧耦合开发,导致系统臃肿、建设效率低、无法快速响应业务,且存在大量重复建设工作。为解决上述问题,需要整合出一个中间组织,为所有的项目提供一些公共资源。这个中间组织即为“中台”,其中,汇聚数据资源,统一提供数据及相关服务的就是“数据中台”。

数据中台是企业级公共的、可复用的数据及其衍生能力的组合,相比数据仓库或数据平台,数据中台

收稿日期:2021-07-16

更强调面向业务的赋能,其良性结果是平台化、生态化企业的构建,以数据为核心、平台为支撑,来驱动前线业务变革。

1.1 数据中台适用场景

构建中台必然需要建设成本、组织成本、时间成本,并非所有企业都适合建设中台。创业型公司,以生存为首要目的,需以最快速度打造产品,野蛮生长是其必然选择;发展到一定规模的企业,产品得到市场认可,公司的目标从“生存”走向“发展得更好”,可以考虑向中台架构发展;大规模的企业,各类产品、服务、部门、系统错综复杂,为了满足市场竞争要求,即便困难重重,也需调整架构,打破以往的职能模式,向“前中后台”的平台式组织过渡,进而向更加灵活的形态发展。其中,数据中台的价值在于加速从原始数据到价值数据的过程,提高企业的响应力,满足当前以“提高灵活性和速度”为重的新运营模式的支撑需求。

1.2 先进企业的数据中台实践

不少先进企业早已开始进行数据中台的实践,并取得了理想效果,为企业创造了价值,提升了竞争实力。

a) Supercell:基于中台,打造快速业务试错的核心竞争力。Supercell(芬兰移动游戏巨头),以《部落冲突》等全球热门游戏闻名,经过6年时间将游戏开发过程中公共、通用的游戏开发素材、算法做了很好的沉淀,支撑2~7个员工组成的开发团队在此基础上进行快速创新、试错,从而用最快时间找到用户真正喜爱的游戏。这种强大的业务试错能力是Supercell的核心竞争力。阿里的中台战略也是由此受到启发。

b) 阿里:赋能业务前台快速构建,促进企业线上线下业务线的扩张。阿里继共享事业部成功建设业务中台后,于2014年成立“公共数据平台及产品部”,将原来各自为战的烟囱式业务线支撑模式转变为数据共享战略模式。阿里数据中台的发展经历了2012年以前的“追随业务期”,2012—2014的“顶层设计初探期”,2014—2015的“顶层设计质变期”,2015年3月底,已实现同时服务20多个BU,与业务中台一起,赋能了应用前台的快速构建,促进企业线上线下业务线的扩张,并成功支撑双十一等大型活动。

c) 百度:企业级的组织与架构,快速响应专业线的需求。由全企业统一的“数据工厂”组织团队,负责公共数据的建设及服务,基于公司级通用数据仓库,形成了全公司数据统一展示、管理、使用的运营模式,

同时进行数据拉通、建模,快速响应公司各专业线的业务需求。

经研究发现,先进企业的数据中台有以下几点共同特点。

a) 形成过程:中台战略并不是一个新型产品或概念,都是在业务发展过程中,逐步演进出来的,都经历了“部门各自建设—出现问题(重复建维、系统壁垒、数据多份等)—公共沉淀”的过程。

b) 团队组织:数据中台都是由一个独立的团队统一建设,上层应用可由不同的业务支撑团队搭建。团队成员首先要熟悉业务,需要技术与业务专家从业务视角建设数据中台,保证中台对业务的赋能。

c) 建设模式:企业级公共数据在源头统一,确保大家使用同一份公共数据;公共数据统一加工,以业务需求为导向,遵循同一套标准规范;上层数据应用依赖公共数据可以百花齐放。

d) 服务机制:数据中台所有数据服务必须通过统一的出口提供,并保证服务的一致性、可靠性,供上层应用一点调用。数据中台建设过程中,尽量不对现有业务服务造成影响。建设完毕后对业务前台赋能效果有很大提升。

e) 高效运营:中台团队必须要找到高效的业务运营模式,以主动服务为导向提升自身重要性,才能够真正转型成为企业的“共享业务事业部”,而不是某种换汤不换药的“共享IT支撑部门”。

2 数据中台IT架构

由于每家企业的业务和组织情况不同,不存在通用的数据中台标准架构,本文旨在根据数据中台的使命及必要需求,设计一个基础的数据中台参考框架,企业可在此基础上定制出符合自身条件的数据中台架构。

2.1 数据中台在企业IT架构中的定位

如图1所示,在企业前中后台的整体架构设计中,数据中台与业务中台并列处于中间层,基于计算后台,面向业务前台。两类中台各司其职,业务中台面向业务运营,进行共性流程的服务化沉淀,如将支付流程、订单流程沉淀为支付中心、订单中心;数据中台负责数据管理,进行原始数据到数据资产的加工,并进一步将加工结果包装为可调用的数据服务,为企业的业务及管理进行智慧化赋能。

在企业IT架构设计中,计算后台、业务中台、数据

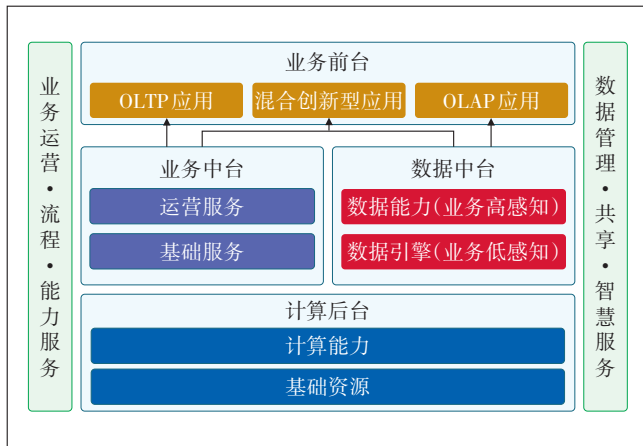


图1 数据中台在企业整体IT架构中的定位

中台与前端应用的分工及关系如下。

a) 计算后台负责基础设施的统一管理规划配, 统筹管理运维前中台所使用的硬件资源、中间件、数据库软件、框架组件等资源。

b) 业务中台与数据中台各司其责, 统一面向前台提供不同类型的专业服务。

c) 随着业务模式的变更, 前台应用向 OLTP & OLAP 的混合形式发展, 应用在充分了解中台能力的情况下, 自由组合各中台的服务能力。

2.2 数据中台的基本构成

数据中台的典型构成包括3个基本层级, 从下到上依次为数据湖、数据仓库及数据超市(见图2)。

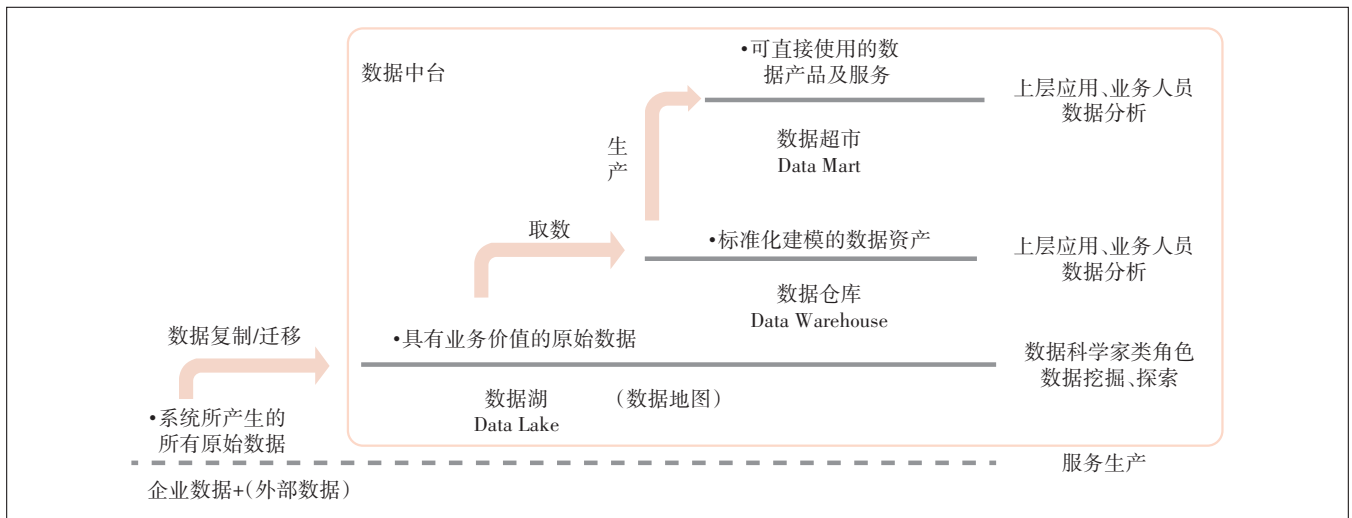


图2 数据中台的基本构成

a) 企业数据: 企业所有IT系统产生的数据, 以服务业务生产为目的, 标准化程度差, 存储周期短, 包含无业务价值、冗余数据, 无法直接提供数据服务。

b) 数据湖: 从企业基础数据中, 选取具有业务价值的原始数据, 恢复数据完整性、可读性, 保留数据原始结构, 形成数据湖。数据湖可长期存储历史数据, 并且具备统一权限、安全等管理功能, 以支持数据科学家类角色进行数据挖掘、探索。

c) 数据仓库: 以OLAP数据关联分析为目的, 采用数据仓库建模方法对分析所需数据进行标准化分层建模, 供前台和业务人员使用。

d) 数据超市: 基于数据仓库明细, 汇总数据, 沉淀常用的数据产品, 并封装为各种形式的 DataService, 供前台应用直接调用、业务人员直接使用, 进行高效数据分析等工作。

2.3 数据中台的架构设计

根据数据中台的基本构成, 将其IT架构细化设计, 整体框架如图3所示。



图3 数据中台IT架构-level0

2.3.1 数据汇聚

为了进行全面的业务分析预测, 需要广泛地汇聚企业内外部系统的数据, 以通信运营商为例, 数据源可以分为如图4所示的四大类。

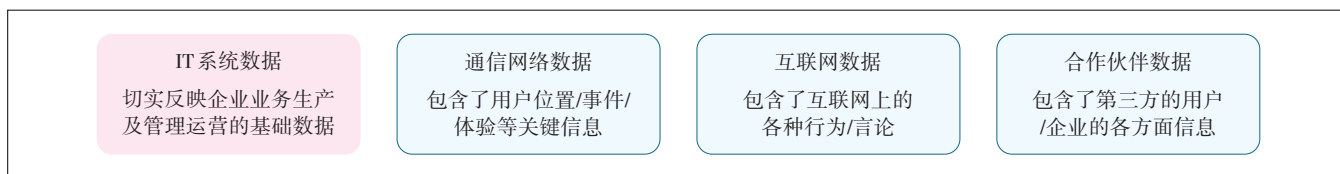


图4 通信运营商的数据采集范围

除了广泛覆盖企业内IT系统数据之外,根据企业业务,还需采集相关的非IT系统数据,如基站信令、互联网站数据、第三方合作伙伴的用户数据等,打通整合后,为公司业务发展、决策提供支撑。

完成数据汇聚后还需要做好数据地图的展示,让企业人员了解公司的数据资产及分布情况。

由于数据源的种类、存储类型不同,数据中台的采集层需要具备满足多种典型场景的采集技术,如数据库直连采集、文件采集、爬虫、实时消息采集、日志采集等。

2.3.2 数据加工

2.3.2.1 数据加工技术

根据主要业务场景,混搭使用多种业界先进数据处理技术,快速、高效、稳定处理各类结构化和非结构化数据。典型的数据处理场景分为以下3类。

a) 批量处理:大数据分布式处理技术,以HDFS为基础,具备常见的Hadoop生态圈中的常用框架组件,用于处理实时性要求低的大批量数据。

b) 流式处理:具备Flink、SparkStreaming、Storm等典型框架,满足不同场景的流式数据处理要求,用于处理实时性要求高的数据。

c) 机器学习:具备成熟的机器学习框架、算法库,提供复杂业务模型的训练环境。

2.3.2.2 数据仓库构建

“数据智能”(Data Intelligence)中一个必要且基础的环节,就是数据仓库的建设,同时,数据仓库也是公司数据发展到一定规模后必然会提供的一种基础服务。以主流的数据仓库建模理论(如Kimball模型等)为指导,做好数据标准化分层建设。按照企业业务总线,梳理出各类维度、事实,再以常用业务域分类,进行汇总、整合拉通,达到可复用、快速查找的目的。

2.3.2.3 能力沉淀

在进行一段时间的数据需求服务后,可将加工过程中常用的数据结果及功能沉淀下来,避免重复开发,同时,可以缩短需求响应时间。在大数据服务中,

常见的数据能力有以下几类。

a) 数据产品:大部分数据加工结果可通过标签、指标、业务挖掘模型来沉淀、共享,数据产品需要根据业务发展不断补充并调优,保持产品的更新及准确性。

b) AI能力:人工智能已在各行各业逐渐普及,基础的AI能力需要以低门槛的形式提供,方便使用,降本增效。

c) 工具:常用的数据分析工具可沉淀在数据中台,用户直接云化使用,避免各个前台应用分别部署维护,浪费资源。

2.3.4 开放共享

数据开放共享是数据中台最重要的目的。做好数据开放,需从清晰展示数据目录、多样化开放手段、多角色多场景用户交互及良好的平台运营维护方面努力,让数据使用者“了解”数据、“会用”数据、“敢用”数据。

a) 数据目录:以数据商店的形式,直观展现大数据平台加工的数据及能力内容,统一数据服务门户;设计友好的交互界面,以业务语言解释说明数据及能力的内容、使用方法和场景;让用户“了解”数据及能力。

b) 开放方式:以多种技术手段开放数据及能力,包装成完整的产品及服务,以支持各种不同场景的数据开放。

c) 交互场景:面向不同层级的用户,提供多种场景的数据服务,让用户“会用”数据及能力。目前数据服务主要分为自助式服务和主动式服务2类,前者面向技术人员以租户和交互式探索形式提供服务,面向非技术人员提供可视化自助分析工具,如拖拉拽配置组合各类数据标签指标等;后者有专人对口不同业务部门,了解业务后深度服务业务分析,如建模、分析报告等。

d) 开放运营:展示产品质量和使用情况,让用户“敢用”数据;做好平台运维,保证使用稳定性、及时性;统计分析使用效果,反向促进能力提升;建立并维

护问题反馈渠道,提升使用感知。

2.3.5 数据治理

数据治理目标:保证数据质量,促进数据开放,保持大数据工作可持续发展。

数据治理原则:轻量化、便捷化,以解决实际问题为导向,对数据生产起正向作用。

根据数据中台常见痛点问题,提出以下几个必要的治理要求。

a) 基础数据规范:制定基础模型、编码、指标及标签的标准规范,并系统化支撑,可编辑可查询,灵活更新。在各类系统开发建设过程中,所创建的数据如果在标准规范内已有定义,则直接引用标准,如果要创建标准规范以外的新型数据,需要先修订标准再建设,保持数据规范、可复用、口径清晰且一致。

b) 数据开发规范:制定轻量化开发规范,标准化数据命名、加工脚本等常用关键环节;规范各场景开发/测试流程及路径区域,避免相互影响,节省算力;同时,需通过统一开发平台,系统化支撑标准落地,对开发过程进行监测、预警,维持有序的开发环境。

c) 生命周期规范:制定各类数据的存储策略,并写入系统,到期前通知负责人,确认下线;定期清理长时间无访问量的数据表,降低存储计算成本。

d) 数据监控:平台元数据定期扫描比对,发现差异及时通知负责人;资源画像、作业画像、存储画像、计算画像、用户操作审计,多方位监控。

e) 数据安全:主要分为内部数据安全和外部数据安全,前者涉及事前权限控制,事后操作日志可追溯;后者要保证数据不出门(在内部完成数据加工,输出数据分析结果),并对关键信息加密、脱敏。

3 数据中台建设关键要素

数据中台的建设,除了满足IT系统技术条件之外,还有一些非技术性的关键要素,如果这些要素不能满足,将会直接影响数据中台的使用结果。

3.1 团队组织

目前大型企业内部组织结构复杂,常见情况有本专业数据为部门私有资产,不愿意共享;各专业线人员对其他专业不理解,即使有业务场景,也不能发挥最优价值。这样就导致数据中台难成体系、难于推广、难以发挥实质作用;数据拉通受阻,不能从根本上解决数据壁垒问题。

因此,跨域、互通的中台建设需要跨域、跨专业的

人员,如果人员分散在不同组织中,工作职责各不相同,接触的领域各不相同,那么,任何跨域系统层级的技术性调整都难以发挥预想的作用。

参考先进企业的中台团队模式,由统一、跨部门的专业化团队负责数据中台的建设。成立企业级数据中台团队,由业务领域的分析人员、技术相关的数据研发工程师、运营相关的服务团队构成,确保团队构成的全面性与专业性,同时借此环境重点培养自有专业人才。以业务需求为导向,各域需求方参与中台建设,相互支持依赖、监督要求。专业人员集中,在统一架构、规范下,逐渐优化公共数据能力。

前台由各专业线IT支撑团队自由建设,但所有数据分析类功能,必须经过数据中台,即明确是否调用中台能力或将数据能力沉淀至中台,中台团队权责匹配。

引入市场化、互联网化的绩效评价机制,对数据中台团队的建设及服务进行衡量,以使用效果为导向,包括所提供服务的频率、前台应用满意度等。

3.2 数据中台部署架构

需要建设数据中台的大型企业,一般都是发展到一定规模后,发现烟囱林立的IT结构不适合大型的改造、重构建设,影响业务支撑。对于这种情况,建议该企业采用逻辑上统一的中台架构,以服务业务前端为前提,遵循统一标准,统一服务运营,并支持分布式部署。

a) 统一需求入口:由各域业务人员、需求分析人员组成的专业需求团队对上层应用需求进行分析、梳理、分解、指派。

b) 统一服务出口:服务出口的责任方需要负责数据服务的统一封装、制定并执行服务标准,提供标准、一致的服务。

c) 数据研发:各域数据研发团队组织统一但工作内容不变,横向引入标准团队,负责牵头各类数据研发标准的制定,对现有模型、指标进行梳理,发现问题并牵头与各域研发人员一起解决问题。

d) 架构可扩展性:新业务或新技术(如5G、合作伙伴能力)带来的新的公共能力模块,可以按照目前的统一框架扩展分布节点。

3.3 互联网化服务运营

发挥数据中台价值的关键在运营。中台团队需要在建立系统化、规范化的开放机制基础上,转变思路,从被动服务转变为主动“营销式”服务,为业务前

端创造价值,营造良好的开放生态。

a) 明确统一的数据开放机制:数据是全企业的共有数据,让更多的人用起来,才能充分发挥数据的价值。明确规定什么角色可以用什么数据、开通什么服务,并将规则固化落地到统一开放平台中,系统化正规化,减少线下靠“刷脸”“打招呼”等人为交互。

b) 服务化的中台运营:以用户视角主动服务,培养数据中台使用习惯及模式,提升使用率(见图5)。

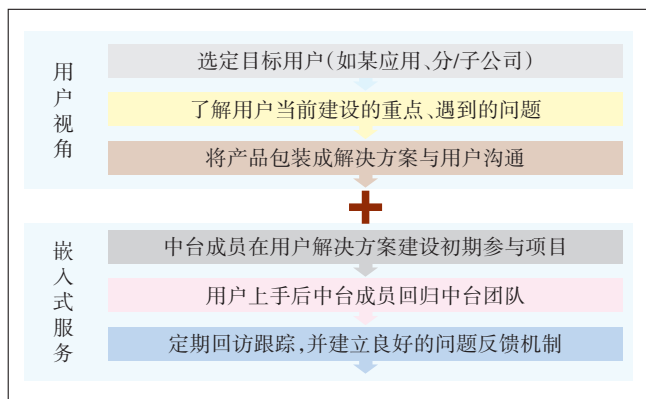


图5 服务化的中台运营

3.4 数据中台分阶段建设

由于需求企业的复杂情况(如业务支撑已稳定运行、IT系统烟囱众多、部门壁垒等),数据中台的建设无法从零开始,需要在不影响现有业务支撑的条件下,分阶段进行,一般可以分为以下四大阶段。

a) 一阶段:制定标准规范、方法论并细化实施计划,统一整合基础数据层。

b) 二阶段:按照业界标准方法论,梳理企业业务总线,同时从梳理重点业务的底层数据入手,逐步优化构建数据中间层。

c) 三阶段:通过统一开放平台优先支持新建的应用,同时对老应用进行改造迁移。

d) 持续运维:数据中台人员深入业务场景,不断沉淀优化公共数据层。每年可针对公司战略重点,与业务、前端IT人员合力建设。

4 结束语

随着市场竞争的加剧,增速放缓,业务复杂的大型企业迫切需要建设信息共享中台,即数据中台,以打破部门间的信息壁垒,提升企业的智能化程度。而中台本质上是改革,是一种战略选择,是资源的再分配与流程的再造,需要新的组织、管理模式与之匹配,

而不仅仅是系统的建设。除了强大的数据汇聚、处理能力,有效的开放服务机制外,还需要配套的组织团队、灵活的部署落地、互联网化服务运营支撑,保证达到应有的效果。

参考文献:

- [1] 钟华. 企业IT架构转型之道:阿里巴巴中台战略思想与架构实战[M]. 北京:机械工业出版社,2017.
- [2] 阿里巴巴数据技术及产品部. 大数据之路:阿里巴巴大数据实践[M]. 北京:电子工业出版社,2017:132-133.
- [3] 邓中华. 大数据大创新:阿里巴巴云上数据中台之道[M]. 北京:电子工业出版社,2018.
- [4] 吕兆星. 企业大数据系统构建实战:技术、架构、实施与应用[M]. 北京:机械工业出版社,2017.
- [5] 钱岭,孙少陵,石在辉. 从点状应用到大数据统一平台[J]. 大数据,2017(3):84-93.
- [6] 孟祥飞,冯景华,赵洋,等. 应用驱动的大数据融合平台建设[J]. 大数据,2017(2):67-77.
- [7] 王勇. 企业大数据平台建设中的问题及对策探讨[J]. 信息系统工程,2019(6):39.
- [8] 程学旗,靳小龙,王元卓,等. 大数据系统和分析技术综述[J]. 软件学报,2014(9):1889-1908.
- [9] 马朝辉,聂瑞华,谭昊翔,等. 大数据治理的数据模式与安全[J]. 大数据,2016(3):83-95.
- [10] 王斌,张昱. 通信运营企业大数据平台建设思考[J]. 邮电设计技术,2017(8):73-78.
- [11] 董昭. 开放的企业级大数据平台建设思路[J]. 通讯世界,2017(14):48-49.
- [12] 彭昭,董斌,杨迪,等. 大数据PaaS平台架构的研究[J]. 互联网天地,2015(10):81-84.
- [13] 童晓渝,张云勇,房秉毅,等. 大数据时代电信运营商的机遇[J]. 信息通信技术,2013(1):5-9.
- [14] 范济安,李卫,魏进武. 电信运营商的大数据发展战略[J]. 信息通信技术,2015(6):5-12.
- [15] 李卫,魏进武,张云勇. 大数据发展助力运营商业务创新转型[J]. 企业文明,2015(9):42-44.
- [16] 杨洋,严晓明,邓晓晖. 基于云计算和服务化中台架构的医院信息系统构建[J]. 中国卫生信息管理杂志,2019,16(6):749-754.
- [17] 李巍巍. 数据中台技术在业务系统中的应用研究[J]. 现代信息技术,2019,3(21):108-110.
- [18] 冯立娜,朱旭. 整合资源提升效率构建运营商中台体系[J]. 通信企业管理,2019,381(1):29-33.

作者简介:

张洁,高级工程师,硕士,主要从事企业信息系统的规划、研究工作;许建宏,高级工程师,硕士,主要从事企业信息系统的规划、研究工作;肖伟,高级工程师,学士,主要从事企业信息系统的规划、研究工作。