



面向对象流程驱动的大数据集中管控模式

张 青

(中国电信股份有限公司广州研究院 广州 510630)

摘 要:随着大数据时代的来临,数据快速增长,大规模并行处理和访问要求日益迫切,分区分域的数据存储、处理和应用架构已逐渐形成。企业迫切需要通过实现全网数据一体化运营管理,有效开展数据应用,保障数据质量,充分发挥数据对经营管理的价值。为此,提出了一种面向对象流程驱动的大数据集中管控模式,应用数据的生命周期方法,梳理出几大典型的数据对象,并通过规范化的流程来驱动数据的协同管理。

关键词:数据生命周期;面向对象;流程驱动;集中管控

doi: 10.11959/j.issn.1000-0801.2015254

Centralized Big Data Management Model of Object-Oriented Process Driven

Zhang Qing

(Guangzhou Research Institute of China Telecom Co., Ltd., Guangzhou 510630, China)

Abstract: Along with big data era, the rapid data growth, large-scale parallel processing and access requirements of increasingly urgent, partition the domain data storage, processing and application architecture has been gradually formed. Enterprises are urgent to achieve full network data integration management, effective data application, guarantee the data quality, give full play to the value of data management. Therefore, a centralized big data management model of object-oriented process driven was proposed. The model uses a life cycle method to tease out several typical data objects, and through a standardized process to drive the collaborative management of data.

Key words: data lifecycle, object-oriented, process driven, centralized management

1 引言

随着大数据时代的来临,数据快速增长,对数据快速处理和访问的需求日益迫切。传统面向少数分析人员的数据分析模式以及数据仓库架构都不能满足海量数据、大规模并行访问以及模型快速响应等需求,分区分域的数据存储、处理和应用的分分布式处理架构已逐渐形成,对数据的集中管控难度随之加大。

目前企业对精确化管理及精准营销能力的要求不断提升,数据运营中自动化、标准化、集约化和系统化能力不足的问题越来越突出,难以满足电信运营商对集团和省两级架构集约化运营的要求。问题主要表现在以下几个方面:

- 缺少对业务对象,尤其是对指标、维度的系统管理,对业务部门的指标发布支撑效率低;
- 对数据模型与业务对象关系缺乏动态管理,导致业

收稿日期:2015-05-20;修回日期:2015-10-09

论文引用格式:张青. 面向对象流程驱动的大数据集中管控模式. 电信科学, 2015254

Zhang Q. Centralized big data management model of object-oriented process driven. Telecommunications Science, 2015254



务需求不能快速实现,口径难以统一,影响支撑及时性和数据应用效果;

- 统一的数据作业调度能力较弱,作业调度的灵活性、透明性不足,缺乏统一规范,导致系统间交互困难;
- 数据质量稽核手段多样,但缺乏统一标准,稽核规则散布在多个系统中,难以实现稽核结果的共享和交互;
- 数据安全急需强化,缺乏字段级的数据安全策略定义、安全检测、安全预防及告警的集中管理、快速部署及共享能力;
- 全网一体化的运营能力较弱,缺乏业务的快速加载、快速部署能力,数据模型、调度策略、质量标准、数据安全方面的交互和控制能力弱,难以高效地支撑公司集约化运营。

2 概念与思路

2.1 数据生命周期

生命周期的概念应用广泛,特别是在政治、经济、环境、技术、社会等诸多领域经常出现,其基本涵义可以通俗地理解为“从摇篮到坟墓”的整个过程。对于某个产品而言,就是从产品的需求到产品设计、开发、上市、消亡的全过程,既包括制造产品所需要的原材料的采集、加工等生产过程,也包括产品储存、运输等流通过程,还包括产品的使用以及产品报废或处置等废弃回到自然的过程,这个过程构成了一个完整的产品生命周期。

在信息化时代,尤其是随着移动互联网、物联网的发展,数据成为当今各行各业一个重要资产,很多企业特别是互联网、IT企业,数据本身就成为了一种重要的产品。数据从产生、加工到应用构成一个完整的生命周期,在生命周期的各个阶段需要有相应的动作来保证数据的可用性、可控性和安全性^[1]。

2.2 面向对象的管理

面向对象是当今主流的软件开发方法,随着信息技术的发展,面向对象的概念和应用已超越了程序设计和软件开发的范畴,扩展到其他领域,如数据库系统、交互式界面、应用结构、应用平台、分布式系统、网络管理结构、人工智能等^[2]。

对象表示静态的、结构化的数据性质,描述系统的静态结构,包括对象的结构、属性和操作。同样,数据在生命周期各个环节的存在形态也可以被梳理成特定的对象,通

过面向对象的方法对数据的全生命周期进行管理,可以给企业数据的集约化管控带来便捷。

2.3 流程驱动

流程化管理是指通过构建一套企业各个部门、单位、层级共同遵守的,各个业务环节整合的全面管理模式,从而实现跨应用、跨部门、跨层级、跨合作伙伴的合作管理。

流程无处不在,在企业应用的各种管理系统中,包含大量的流程管理工作。

在互联网的业务创新时代,尤其是随着大数据技术的发展,企业的各项业务都需要企业全面的数据支撑,数据信息的生产管理已经不再是IT部门独自面对的问题。企业的各个部门、各个产品线、业务平台都有自己的信息平台,梳理和完善各部门之间、上下级之间的数据管理流程成为企业数据集约化管控的必经之路。

2.4 面向对象流程驱动的大数据集中管控模式

面向对象流程驱动的大数据集中管控模式是指利用数据生命周期管理的概念,梳理数据在全生命周期各个环节的存在形式,形成特定的对象,从而有利于界定数据在不同环节的特征属性以及对应的处理动作^[3]。然后,通过面向对象的方法来管理各个环节的不同数据。通过制定数据对象在各系统之间、各部门之间以及集团与省级之间的标准化闭环流程,加强对全网数据的集约化协同管理。

3 基于生命周期的数据对象梳理

3.1 电信数据生命周期

数据伴随着IT系统的规划、建设、运营和应用4个阶段而产生,在系统的生命周期中,数据也有它的生命周期及生命周期各个环节中关于数据质量、数据安全操作的应用^[4]。目前电信运营商系统的数据全生命周期及在数据加工、质量检查、数据安全稽核的大致流程如图1所示。

3.2 关键数据对象及对象信息

为了实现数据的统一集中管理,首先需要根据数据的全生命周期提炼归纳出数据对象以及各个对象的关键属性信息和操作信息。通过对电信运营商系统及数据全生命周期的梳理,总结出管理全网数据的5个关键对象,如图2所示。

5个关键对象的信息如下。

(1) 模型管理对象信息

模型属性包含模型名称、模型编码、模型类型、所属主题域、分类等;模型管理主题域包括参与人域、产品域、账

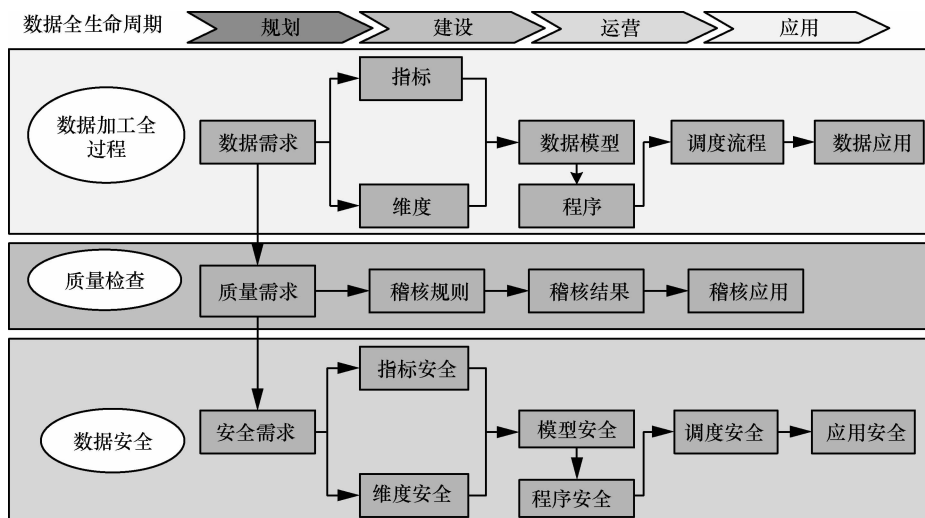


图1 数据全生命周期流程

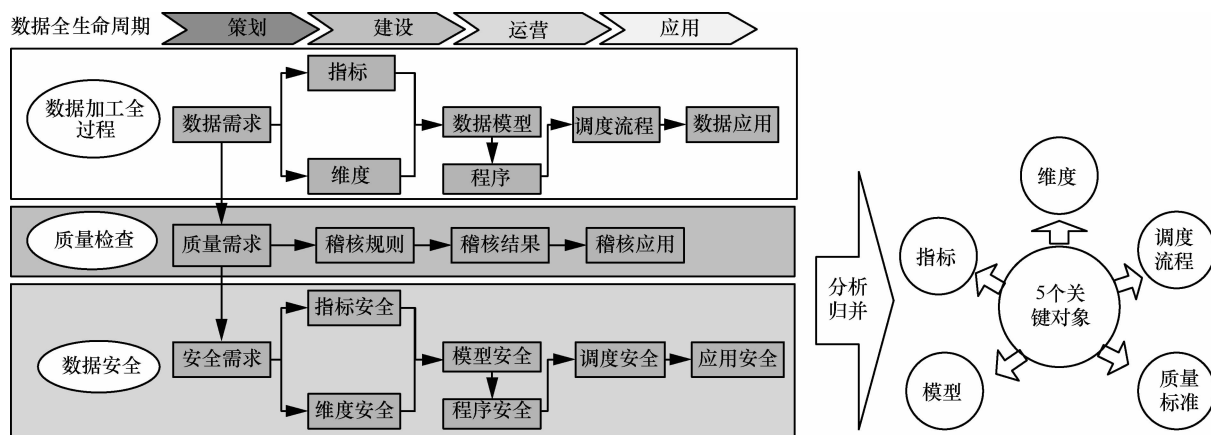


图2 关键数据对象的提取

务域、事件域和市场营销域；模型分类主要包括日表、月表两套体系。

(2) 指标管理对象信息

指标属性包含指标名称、指标解释、指标口径说明、指标和维度关系、指标类别、指标层级等；指标类别主要分为基础指标和衍生指标；指标层级主要分为收入分析类、资源成本类、结算收入类等几大类。

(3) 数据质量管理对象信息

数据质量稽核规则包含层次、稽核性、接口类型、接口名称、字段名称、稽核规则、规则标识、告警对象、运行平台；稽核的层次主要包括接口层、前置机、文件层、整合层和中间层等；稽核标准主要包括波动性、合规性、及时性、逻辑性、完整性和一致性等。

(4) 维度管理对象信息

维度的定义包括维度含义描述、维度值编码及编码的

含义等；维度的属性包括维度的区域属性、时间属性、业务属性等；维度的版本是指维度在不同时间段的变化。

(5) 调度流程对象信息

调度的任务包括任务对应的动作和程序信息、任务的关联部门和人员信息等；调度的规则主要是指调度的频率、优先级、触发条件等。

4 面向关键对象的业务流程和功能梳理

4.1 概述

为了对上述提炼出的管理全网数据的5个核心对象进行集中管控和运营保障,确保核心对象在产生和使用过程中具有良好的质量、透明的关系,为业务运营的标准化、智能化提供坚实的基础。下面描述5个面向对象的管理功能,具体包括模型管理、指标管理、质量管理、维度管理、流程调度管理。



- 模型管理实现对数据模型的统一版本管理和模型落地监测,实现对模型的定义、发布到下线的全生命周期集中管理,确保集团和省分数据模型一致。
- 指标管理是指对全网重点经营指标和省分经营指标进行管理,包括指标定义、指标发布、指标下线、版本管理和指标视图。
- 质量管理提供数据质量标准的全生命周期管理,包括稽核规则定义、稽核规则视图、稽核规则发布、稽核规则下线 and 版本管理。
- 维度管理提供维度的全生命周期管理,包括维度定义、维度下线、维度视图和维度发布等。
- 流程调度管理提供数据系统的调度管理,包括流程定义、流程调度和流程监控。

4.2 管理关键对象的业务流程和功能细化

4.2.1 模型管理流程及功能

模型管理流程如图 3 所示。

模型管理流程由模型定义、模型发布、模型导出、模型下线等功能实现。模型定义提供对模型的增加、删除和修改等基本操作;模型发布是指对模型进行版本定义、发布对象定义、发布信息通知、版本内容共享下载的管理功能,模型发布的内容包括模型的定义信息、版本差异报告;模型导出功能是指将物理模型定义按文件方式导出;最后,通过模型变更审批流程对数据域内所引用该版本模型的系统进行模型下线。

4.2.2 指标管理流程及功能

指标管理流程如图 4 所示。

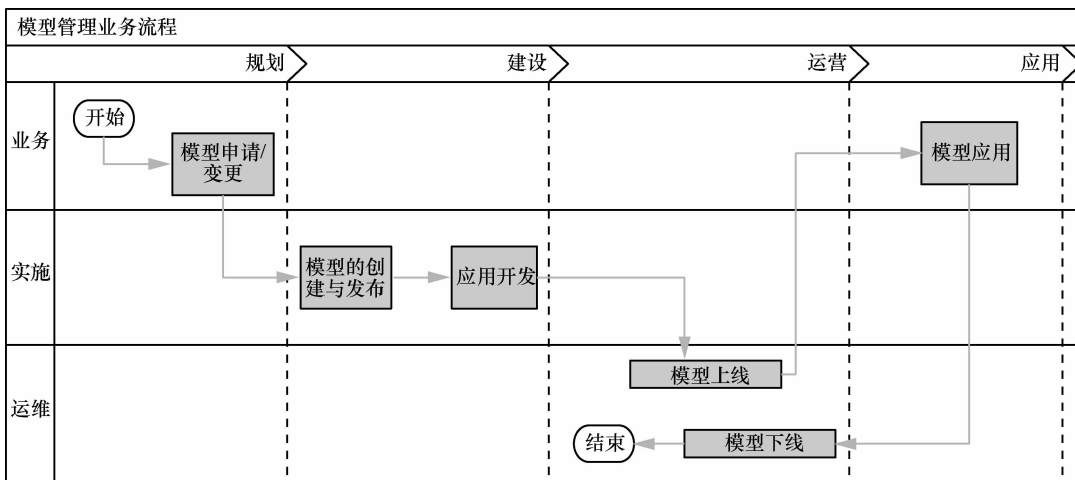


图 3 模型管理流程

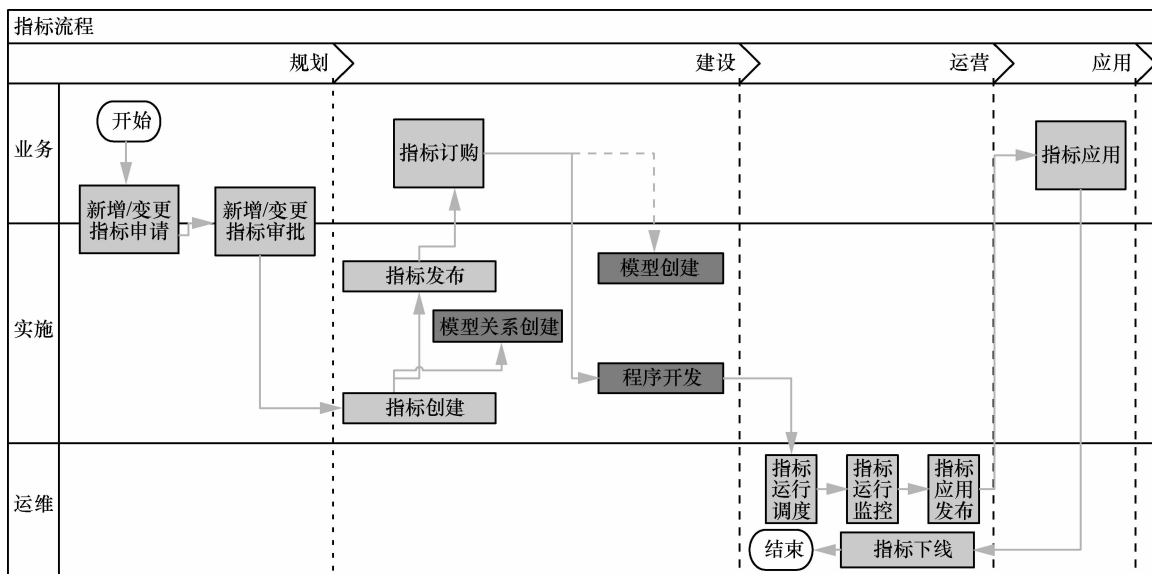


图 4 指标管理流程

指标管理流程由指标定义、指标发布、指标下线等功能实现。指标定义是对指标的基本属性进行管理,实现对指标名称、单位、业务口径、技术口径等描述性基本信息 and 指标支持维度、指标来源的模型、指标捆绑维度、指标计算方法等结构化信息的管理,包括指标的新增、修改、删除和查询等功能。全网重点经营指标发布,在指标定义、申请后,经过集团公司指标审批流程对指标相关属性进行审批、发布后,各省可以接收、查询、落地(不可修改)。全网重点经营指标之外由省分公司新增的指标,经过省分公司指标审批流程对指标相关属性进行审批、发布后,才能在省分公司生效,可以被省内应用引用。指标发布的内容包括指标定义的关键信息:指标名称、编码、单位、口径描述、指标的维度信息、计算规则、模型来源等。全网重点经营指标下线,评估下线影响,审核后,下线通知下发到省分公司;省分公司同样进行下线影响分析,分析评估后实施下线工

作,并在平台登记。对于省分公司在全网重点经营指标之外的指标下线,直接进行影响评估,实施下线,并登记。

4.2.3 质量管理流程及功能

质量管理流程如图 5 所示。

质量管理流程由稽核规则定义、稽核规则发布、稽核规则下线等功能实现。稽核规则是指对稽核的目标对象进行一些规则的判定校验,以实现数据稽核的相关业务逻辑;稽核规则管理是对稽核规则的新增、修改、删除和查询。稽核规则发布是对稽核规则的新建、变更、删除进行申请审批流程化管理,经过审批通过以后进行分发。稽核规则不再使用的,则进行下线操作。部分模型或者指标下线以后,对应的相关稽核规则也需要下线。

4.2.4 维度管理流程及功能

维度管理流程如图 6 所示。

维度管理流程由维度定义、维度发布、维度下线等功

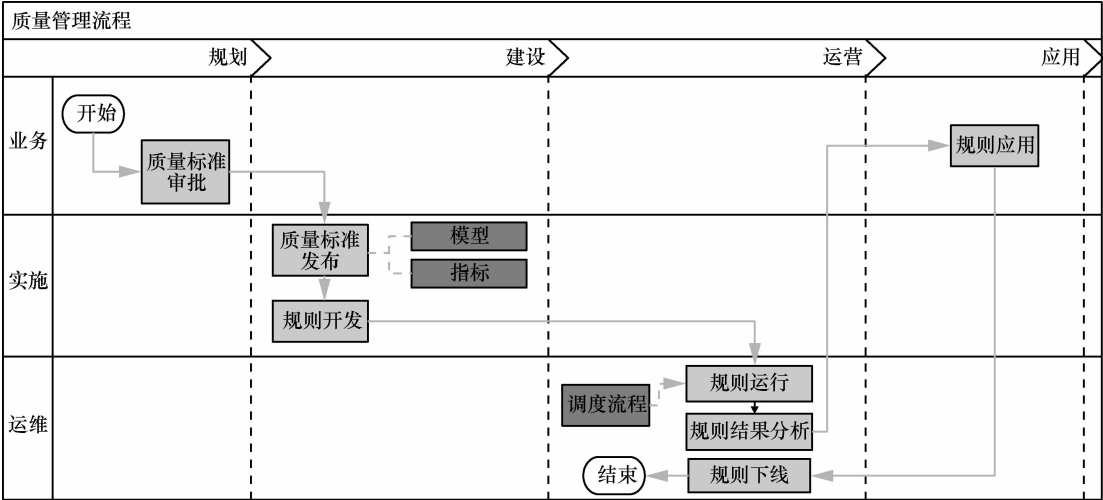


图 5 质量管理流程

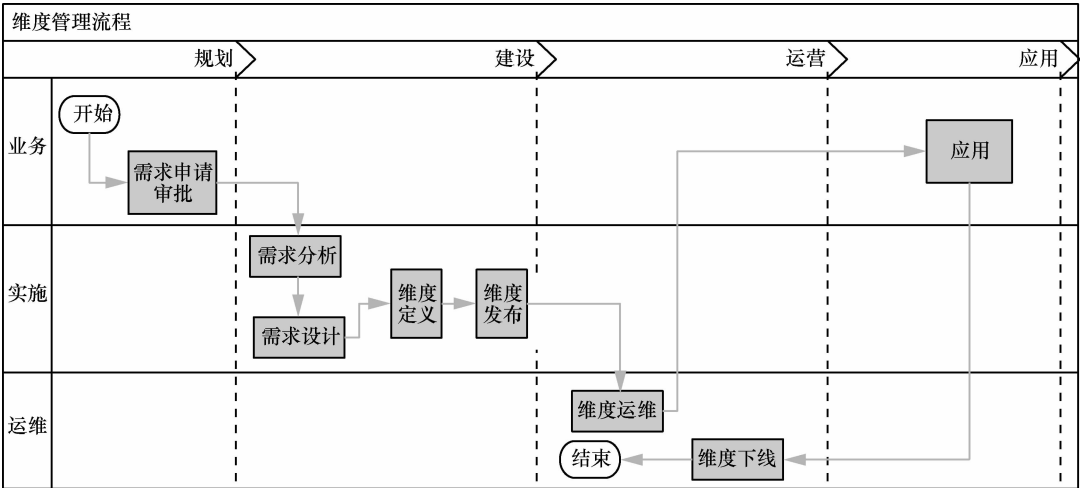


图 6 维度管理流程



调度管理流程如图 7 所示。

调度管理流程由流程定义、流程调度、流程监控等功能实现。流程定义是通过管理流程目录、流程版本、流程基本信息、流程调度信息、流程环节信息、流程环节路由信息实现整个调度流程管理；流程调度是根据流程定义的环节、路由信息调度执行，环节主要分为子流程、数据域系统ETL执行流程；流程监控是监控日志数据记录的接口，实现跟踪整个数据生产过程。

5 指标运营视图场景验证

如图 8 所示,对上述梳理出的指标管理流程进行图形

化(即指标运营视图),可应用到实际生产系统中。

指标运营视图围绕重点经营指标构建全方位的管理视图,提供从指标定义、指标来源、规划时间窗、加工进度运行状态、异常情况 etc 360°视角的指标运营数据组织和呈现,按照业务需求和运营管理要求对指标整个运行周期进行定义与跟踪,监控指标的数据源接口、关键处理环节、异常原因类型、责任单位等信息,采用甘特图等直观、简洁的界面形式展现整体指标的运行进度情况,并提供信息下钻等交互操作模式,同时对延误或异常的指标进行系统提示(短信、邮件、通知单)和监控对象明细的查询。

指标运营视图是对全网经营指标从定义、生产到下线的全过程动态监控,该视图综合展现了所监控指标的业务定义环节、动态加工环节、实时运行过程中的各方面信息,为业务/分析人员、系统监控人员、系统维护人员、信息管理人员等提供了一个全面综合、直观可视、实时互动的集

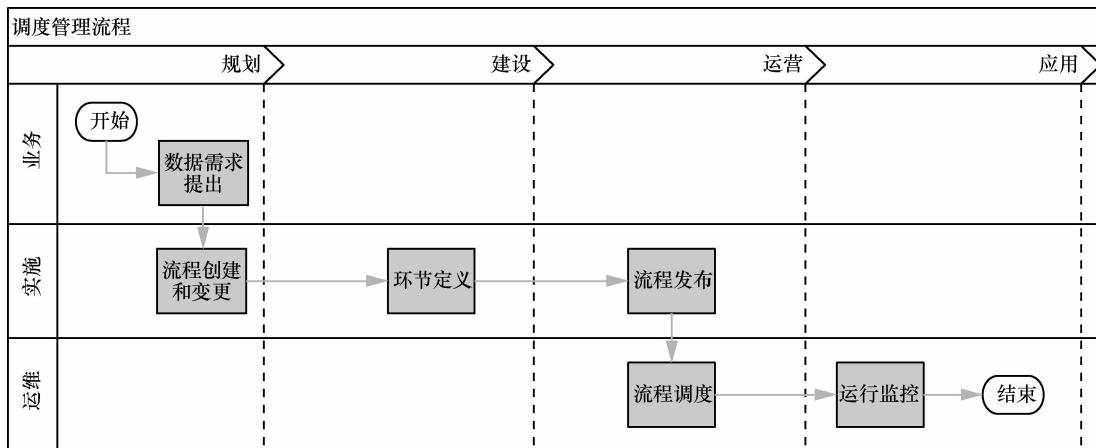


图 7 调度管理流程

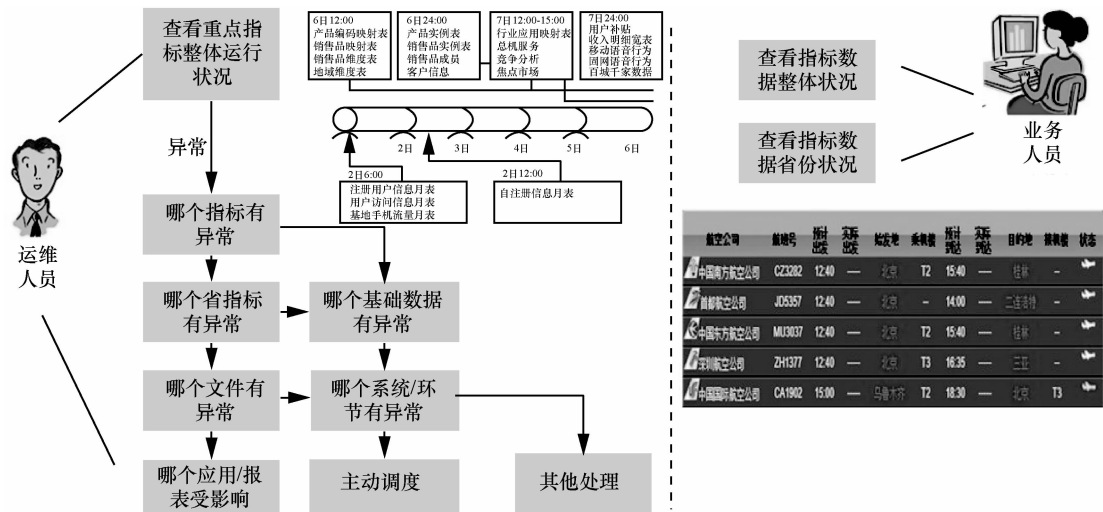


图 8 指标运营视图场景

成工作视窗。

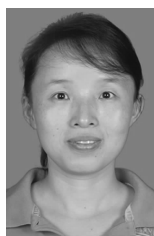
6 结束语

随着信息化的发展,企业的IT系统、数据逐步走向精确化、集约化的发展模式^[5]。电信运营商在数据的集中化管控过程中,面临着业务对象定义不明确、作业调度不统一、数据安全弱、数据质量差异大等问题。本文提出了一种面向对象流程驱动的大数据集中管控模式,采用数据的生命周期方法,梳理出几大典型的数据对象,并通过规范化的流程来驱动数据的协同管理。该模式已应用于某电信运营商的数据运营管理平台的建设中,初步实现全网作业的统一生产调度和管理、全网集约化经营指标的一点发布和快速响应、全网同口径系统数据上传下发的准确一致,从而基本实现全网一体化运营的目标。在平台实际运营过程中,还需配套相关的人力资源和考核等措施。

参考文献

- 何晓斌,朱建涛,郑翔等.一种面向对象的数据管理模型.计算机工程与科学,2009,31(11):73~76
He X B, Zhu J T, Zheng X, *et al.* An object-based data management model. Computer Engineering & Science, 2009, 31(11): 73~76
- 李中泉,徐金兰,郭怡晓等.关系数据模型与面向对象数据模型的关系探讨.计算机与应用化学,2013,30(7):112~114
Li Z Q, Xu J L, Guo Y X, *et al.* Relationship in the relation data model and object-oriented data model. Computers and Applied Chemistry, 2013, 30(7): 112~114
- 孙君意,李厚福,韩燕波.基于特征的业务流程数据可扩展管理机制.计算机集成制造系统,2011,17(8):1856~1863
Sun J Y, Li H F, Han Y B. Feature-based scalable management mechanism of business process data. Computer Integrated Manufacturing Systems, 2011, 17(8): 1856~1863
- 常润梅,孟利青.电信业云计算中心数据生命周期管理研究.信息通信,2013(2):227~228
Chang R M, Meng L Q. Study on data life cycle management telecommunication cloud center. Information & Communications, 2013(2): 227~228
- 张青,黄平.电信行业基于数据驱动的精准营销模式研究.电信科学,2014,30(7):143~148
Zhang Q, Huang P. Research on precision marketing method based on data-driven in telecom industry. Telecommunications Science, 2014, 30(7): 143~148

[作者简介]



张青,女,中国电信股份有限公司广州研究院高级工程师,主要从事大数据处理、数据仓库、数据挖掘及客户营销维系等方面的研究工作。