

以大数据为核心的线网指挥中心建设方案

陈莉莉^{1,2} 胡波^{1,2} 狄颖琪^{1,2}

(1. 南瑞集团(国网电力科学研究院)有限公司,210061,南京;

2. 国电南端科技股份有限公司南京轨道交通技术分公司,210061,南京//第一作者,高级工程师)

摘要 提出了围绕大数据分析处理的轨道交通线网指挥中心的建设方案,扩充了数据接口平台的功能,详细阐述了数据中心的混合架构建设方案,提出了基于实时流处理技术的实时线网监察方案。线网指挥中心的功能向以大数据为中心的转移,在技术上已经完全成熟并可实现。

关键词 城市轨道交通;线网指挥中心;数据仓库;流处理

中图分类号 F530.7; U284.59

DOI:10.16037/j.1007-869x.2020.01.012

Network Command Center Construction Plan Based on Big Data

CHEN Lili, HU Bo, DI Yingqi

Abstract A construction plan of a rail transit network command center around big data analysis and processing is proposed. The functions of the data interface platform is expanded, a hybrid architecture construction scheme for the data center is elaborated in detail, and a new method to realize the real-time network monitoring based on real-time stream processing technology is introduced. The transfer of the network command center core functions to big data center is fully mature in technology and therefore will be achieved.

Key words urban rail transit; network command center; data warehouse; stream processing

First-author's address NARI Group (State Grid Electric Power Research Institute) Co., Ltd., 210061, Nanjing, China

早期建设的线网指挥中心,都是围绕线网监察为核心建设的。随着城市轨道交通网络化进程的发展,监控和运营的复杂度大大增加,线网指挥中心的大数据分析和处理工作越来越重要。

目前,线路综合监控系统的监控细节和实时处理都已经在线路层执行,线网指挥中心的功能应该更偏向于整体监察和宏观管理。其不仅要对整个城市各条线路综合运营情况进行监察,还要在实时数据输入统计和分析的基础上,进行预测、运营指

挥和决策。可见,大数据的综合分析和处理已成为线网指挥中心的核心功能,而且将会越来越重要。

近年来,大数据技术,特别是实时流处理技术日益成熟,完全可以考虑将线网指挥中心从以线网监察为核心转为以大数据分析处理为核心。

1 系统架构

线网指挥中心系统主要包括运营指挥、应急指挥、统计分析、运营评估、信息发布与服务等功能模块。线网指挥中心同视频监控中心、PA系统(公共广播)、PCC(信息播报中心)、ACC(自动售检票清分中心)、消防与安防中心等线网其它机构交换数据和信息,协同工作,从而实现城市轨道交通的线网级控制和协调功能。

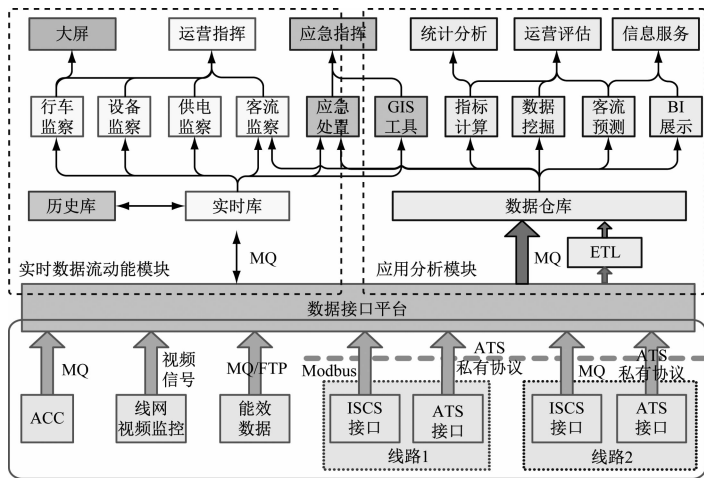
线网指挥中心系统的整体架构框图如图1所示。

线网指挥中心系统的功能复杂而庞大。在其软件体系内,B/S(浏览器/服务器)架构和C/S(客户端/服务器)架构并存,每个模块都有独立的架构,其功能也完全独立。因此,需要通过集成的方法,采用事务驱动方式,来协调各模块之间的工作,使之组成一个有机整体。

集成各模块时要充分考虑将来需求扩展的不确定性。可见,线网指挥中心系统要具有包容性和异构性,当其中一个模块升级或功能扩展时,不能影响其它模块和系统的整体功能。

ESB(企业服务总线)等使用SOA(面向服务的架构),采用消息中间件技术,其服务的提供方和需求方以松耦合关系连接,而且采用开放式标准。这样线网指挥中心系统中集成的各模块,具有充分的弹性和可扩展性,能很好地满足线网指挥中心系统集成需求。

图1中的实时数据流功能模块要求处理延迟低、反应迅速。应用分析模块负责分析处理历史数



注:MQ 为消息中间件;GIS 为地理信息系统;ETL 为数据提取转换加载工具;BI 为商业智能工具

图1 线网指挥中心系统整体架构示意

据,一部分数据处理工作(如统计、分析和报表)没有实时性需求;另一部分数据处理工作以流处理方式进行(如图1中指向“应急处置”和“客流监察”的数据流),要产生准实时的预测和报警信息,有实时性需求。

从软件的整体架构设计来看,线网指挥中心系统从消息流和数据流两方面串起了各模块的功能。

在消息流方面:

1) 线网指挥中心系统采用事务驱动的方式来管理消息流。服务提供方将服务注册到系统中,并定义了统一的服务请求响应消息格式。请求服务模块能针对具体事务,查找服务列表并请求服务;服务提供方能响应请求,完成该事务。

2) 采用消息中间件技术,定义了标准的消息报文格式,以实现在 B/S 架构和 C/S 架构的软件模块之间共享消息。例如,采用 JSON 串在系统内定义统一的报警事件消息格式。

在数据流方面:

1) 建设数据接口平台,实现了数据的接入、数据的标准化及数据质量管理。数据接口平台能支持各种接口的接入,并把处理后的标准化数据发送给线网监控中心和数据管理中心。

2) 数据经过数据接口平台的数据标准化、数据比对和清洗处理后,分别送入线网监察系统的实时库和数据中心。实时库保存当前线网接入数据的当前状态断面,而数据中心则保存线网接入的历史数据。

3) 实时库、历史库及数据仓库,均提供了标准 API(应用程序编程接口)、JPI(JAVA 程序接口)、

Web Service(web 服务),并提供通过消息中间件的数据查询服务方式,可进行小规模数据读取。

2 数据接口平台

数据接口平台(见图2)是面向线路统一数据采集的接口,负责线路端的实时数据,不仅兼具数据规整功能和向线路端提供大数据访问接口的功能,而且还承担了线网指挥中心数据质量管理的一部分功能,是线路和线网指挥中心,以及其它线网机构间数据交换的桥梁。

2.1 数据接入

数据接口平台负责对各源系统进行接口管理和数据的预处理。接入的数据包括:ISCS(综合监控系统)、ATS(列车自动监控)等专业的实时状态数据及告警数据,线网 ACC、PCC 等的实时数据、离线数据及统计信息。

数据接口平台的硬件由接口服务器集群和内存服务器集群组成。接口服务器既可作为接口设备进行信息的通信与采集,又可作为各线路以及线网监控平台之间的网络隔离区。集群的负载均衡技术可满足接入数据吞吐量及业务的高可用性要求。

数据接口平台提供的接口类型丰富,可以支持各种数据源的接入。数据内容需遵循数据处理平台定义的接口格式或采用双方共同协商的接口格式(如:和原有综合监控系统兼容的 Modbus 协议、消息中间件数据传输协议、FTP 文件传输协议及 ATS 私有协议等)。

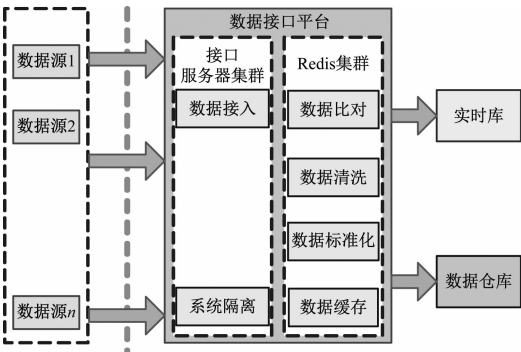


图 2 数据接口平台结构图

2.2 数据规整及标准化

数据规整即把生数据转为熟数据,进行数据比对、数据清洗和数据转换。

数据接口平台作为数据进入系统的输入端口,线网指挥中心系统的数据质量管理也从这里开始。接口软件收到的通信约定格式数据,经过数据比对、清洗和转换,进行模型标准化处理,转化为系统支持的标准数据格式。在数据标准化阶段需要先建立数据标准。模型标准化包括长度、字段、类型、意义、表述等方面的标准化。

以建立线网级统一的数据模型为例,要支持线网级的全数据接入,并保证全线网唯一索引,采用 line1. station1. app1. dev1. dig. point1 的格式,其字

符串的各段依次为线路、车站、应用、设备、点属性及点类型,以保证索引的清晰和有效。

在建立设备模板时,需先根据经验总结出设备的共性,进而形成设备模板,再通过设备模板派生出设备。为了保证时效性,有关设备模板的所有操作都在内存库中进行。

2.3 大数据访问接口

数据接口平台为线路、其它线网中心及 OA(办公自动化)等系统提供访问线网大数据中心的 Web Service 接口及 JDBC(java 数据库连接)接口。

3 数据中心

数据中心主要完成对行车、客流、设备状态等各类结构化数据,以及视频、图片、语音、电子文档和日志等非结构化数据的标准存储;汇总了各级信息数据,是整个线网指挥中心的基础数据平台,实现了数据的统一存储与共享服务;为应用程序提供统一的标准数据接口,为应用程序的下一步数据挖掘和分析服务。

数据中心平台有 Hadoop(分布式计算)、MPP(大规模并行处理)数据仓库及 Hadoop + MPP 混合架构等 3 种建设方案。经过充分的研究和比对测试,3 种方案的优缺点如表 1 所示。

表 1 数据中心平台建设方案比较

方案	优点	缺点
Hadoop	建设成本低;良好的可扩展性,可以实现不停机扩容;支持非结构化数据;支持实时流处理;可选的数据分析和挖掘组件很多	即席精准查询速度相对较慢;复杂 SQL 语句查询效率低;开发和运维的技术成本高;性能和运行稳定性与实施方水平关系大
MPP	稳定可靠;精准查询和复杂 SQL 语句查询效率高;索引结构进行数据统计,生成报表效率高;开发维护相对简单	不能进行实时流处理;不支持非结构化数据存储;MPP 一体机价格昂贵、扩容的成本很高,而软 MPP 集群扩容方便,但有最大节点数限制,存储容量上有限制
Hadoop + MPP 混合架构	结合上两个方案的优点,适应需求变化能力强	两个集群的数据交换没有高效接口,要根据应用设计合理方案;运维技术要求高

由表 1 可见, Hadoop + MPP 的混合架构方案,兼具了其他两种方案的优点。已有线网指挥中心如果采用了 MPP 架构的数据仓库,则随着业务量增长,平台的原设计容量会不足,为降低扩容成本,经常采用混合架构。

Hadoop + MPP 混合架构的数据常按数据热度集成或按数据模型层次集成。

3.1 按数据热度集成

按数据热度集成的 Hadoop + MPP 混合架构如

图 3 所示。根据数据的使用频度、存储时间及访问时间,数据可分为热数据,温数据和冷数据。一般来说,冷数据指半年以前的,很少被访问的数据;温数据指半年内的,需要经常访问以供数据分析和挖掘,并生成报表的近线数据;热数据是指近期生成的在线数据。

在数据生命期内,MPP 数据仓库存储热数据和温数据。这些数据被访问的频度比较高,其中热数据可根据数据生命期管理机制迁移为温数据。

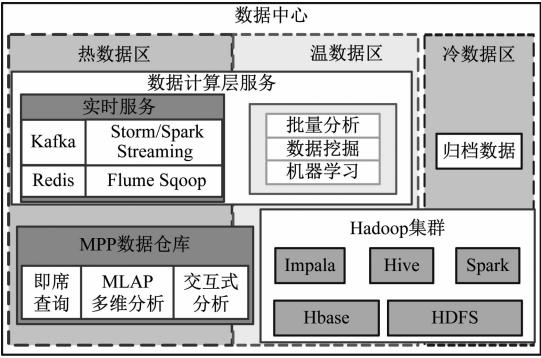


图3 按数据热度集成的混合架构

历史数据平台的数据架构主要搭建在近线集群上,分区分层分主题,为最终的数据深度利用打基础。在线集群以应用为目标,主要用于查询源数据区和数据集市区的数据。

3.2 按数据模型层次集成

按数据模型层次集成的 Hadoop + MPP 混合架构如图4所示。

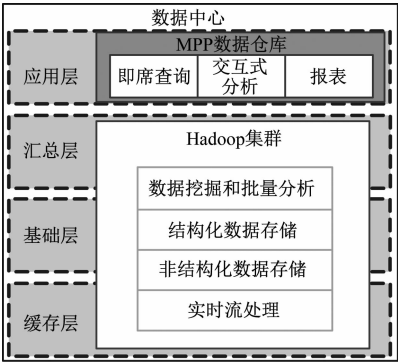


图4 按数据模型集成的混合架构

企业级的数据仓库按照数据模型层次划分。最下层是存储临时数据的缓存层;缓存层的数据经过清洗转换后,存入基础数据层;基础层的数据保留时间长,一般为10年;汇总层存储各数据集市需要的公共数据;最上面的应用层存储了面对具体应用的数据集市。

MPP 数据仓库对 SQL(结构化查询语言)语句的支持度比 Hadoop 平台高,其与 BI(商业智能)工具之间的接口更容易实现,且即席查询效率高。由于 MPP 数据仓库的索引结构更适合统计和报表应用,所以 MPP 数据仓库被放在应用层。

Hadoop 平台的批量数据处理能力较强,故汇总层、基础层和缓存层通过 Hadoop 平台处理好数据,存入应用层。

当然,采用混合架构的线网指挥中心系统结构很复杂,增加了运维难度。

4 基于流处理的线网监察方案

在线网监察系统可应用实时流处理技术。首先,在线路综合监控系统配置转发程序,把收到的综合监控的消息总线数据变化报文和报警事件消息转发到 Kafka 系统的消息总线;然后,线网指挥中心的 Hadoop 平台通过 Kafka 消息总线来获取线路端的数据,并作为实时流处理的输入。实时流处理方案的流程如图5所示。

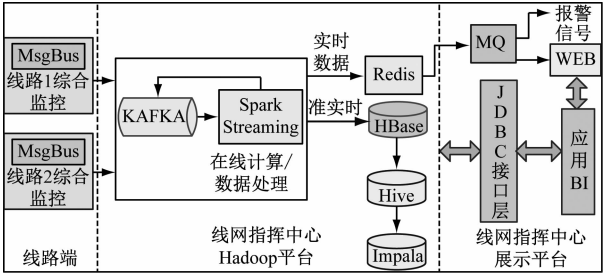


图5 实时流处理方案

实时流处理方案基于 Redis + HBase 组合数据库,其数据采用分层存储的方式,并基于关系型数据库来实现档案数据、模型数据及元数据的同步存储。采用 Kafka 系统,可实现时序消息的数据采集,解决测点数据高效稳定传输问题。基于 Storm、Spark Streaming、Flink 等3种流框架的实时组件,可实现秒级延时,从而解决实时数据的并发高速处理问题。

线路的数据变化和事件等,可以迅速接入线网指挥中心;经过实时计算和分析后,异常数据将直接生成报警信号,并在展示平台上展示出来。

5 结语

城市轨道交通的大数据分析应用,不仅能有效提高线网运营的经济指标,而且能有效提高运营管理水平 and 运营水平综合指标。

而随着技术的发展,原先以线网监察为中心的线网指挥中心,可以借助数据中心强大的数据分析和挖掘能力,提高对地铁数据的加工能力,实现数据的增值,帮助地铁企业充分发掘潜在的数据价值,全面提升运营管理、科学规划、应急辅助决策及公众信息服务能力。

当然大数据技术只是提供一种技术手段,数据价值的提升最终需要依靠从具体业务中总结和提炼的算法。而依靠实时流处理技术进行的实时线

(下转第 105 页)

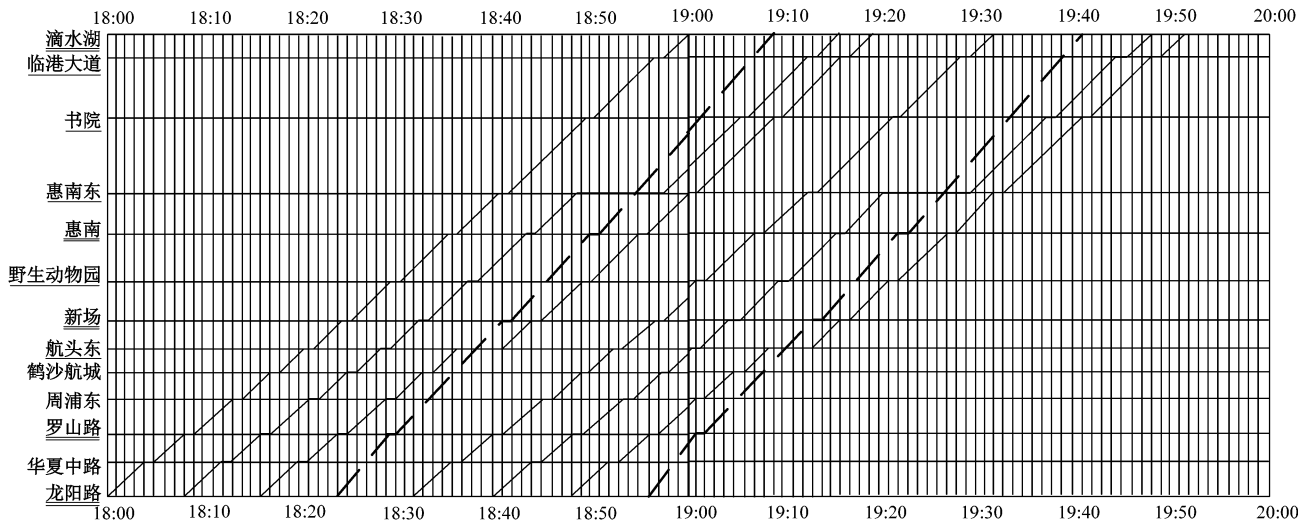


图3 18:00—19:00 16号线下行方向列车运行图截图

车时,慢车在惠南东站停站9 min。

3 结语

市域轨道交通线路采用快慢车模式可解决市郊载客量与客流不匹配的问题。本文以16号线为例,对客流分布特征进行拟合预测,并基于关联规则法确定了列车停站方案,提出了列车开行优化方案,使市域轨道交通线路的载客量和各时段客流能够较好地匹配。

参考文献

- [1] ANTHONY R N. Planning and controls system: a framework for analysis [D]. Boston: Harvard University. 1965.
- [2] MIGNONE A, ACCADIA G. Operations research models for programming support of cadenced timetables[J]. Ingegneria Fer-

rovriariv, 2010(1): 9.

- [3] BOWMAN J L, BEN AKIVA M E. Activity-based Disaggregate Travel Demand Model System with Activity Schedules [J]. Transportation Research Part A, 2001, 35(1): 1.
- [4] 谢军, 吴伟, 杨晓光. 用于短时交通流预测的多项式分布滞后模型[J]. 同济大学学报(自然科学版), 2011, 39(9): 1297.
- [5] 豆飞, 贾利民, 秦勇, 等. 铁路客运专线模糊k近邻客流预测模型[J]. 中南大学学报(自然科学版), 2014, 45(12): 4422.
- [6] 姚向明, 赵鹏, 禹丹丹. 城市轨道交通网络短时客流OD估计模型[J]. 交通运输系统工程与信息, 2015, 15(2): 149.
- [7] 周玮腾, 韩宝明, 李得伟. 城市轨道交通客流分布短时预测模型研究及应用[J]. 城市轨道交通研究, 2015(2): 24.
- [8] 丁小兵, 徐行方. 基于时段分布的市郊轨道交通车站滞留客流分布算法[J]. 交通运输系统工程与信息, 2016, 16(1): 116.
- [9] 许得杰. 城市轨道交通大小交路列车开行方案优化研究[D]. 北京: 北京交通大学, 2017.

(收稿日期: 2018-03-22)

(上接第54页)

网监察方案和以往线网实时监察的实现技术做对比,从技术角度来说可行性并不是问题,当然有系统实现的技术门槛高,以及后期维护的难度和成本等问题需要妥善考虑。

为进一步利用和管理好数据,还需要引入数据治理机制,将数据作为地铁企业的商业资产进行应用管理。数据治理涉及的范围广泛,涵盖数据结构设计、数据管理、数据安全性管理、数据质量管理、数据仓库和BI管理、文档记录和内容管理、元数据管理、主数据和参考数据管理等多个方面,需要成立数据治理委员会并配备专门的数据专员来推行和实施。

参考文献

- [1] 元进辉. 北京市轨道交通路网指挥中心指挥系统(TCC)的设

计和实施[J]. 智能建筑电气技术, 2010, 4(5): 27.

- [2] 杨承东, 徐余明. 基于综合监控系统的线网指挥中心构建方案研究[J]. 城市轨道交通研究, 2013(10): 25.
- [3] RAJDEEP C, BIKRAMJIT P. Proposed hybrid data warehouse architecture based on data model [J]. International Journal of Computer Science & Communication, 2010, 12(12): 211.
- [4] DAVID S, KEVIN S. Selecting a data warehouse architecture [C]// SCSUG. Proceedings 2000 South Central SAS Users Group (SCSUG) Conference. Austin, USA: SCSUG, 2000.
- [5] 朱东升, 徐石明, 李天阳. 基于Hadoop平台的地铁NCC数据中心方案研究[J]. 计算机测量与控制, 2017, 25(12): 224.
- [6] ARTIKIS A, WEIDLICH M, SCHNITZLER F. Heterogeneous Stream Processing and Crowdsourcing for Urban Traffic Management [C]// ACM. Proceedings of the 17th ACM International Conference on Extended Database Technology (EDBT). Athens: Association for Computing Machinery (ACM), 2014.

(收稿日期: 2018-05-11)