

# Lambda 架构在轨道交通车辆 智能运维系统中的应用

胡卫民

(株洲中车时代电气股份有限公司, 湖南 株洲 412001)

**摘要:** 针对轨道交通车辆智能运维系统中海量车载数据处理需求, 为解决单点数据处理模式下海量数据在接收、处理、存储、查询、分析等各环节的性能瓶颈问题, 文章提出一种基于 Lambda 架构原理搭建的数据处理架构方案。其以分布式服务器集群作为底层硬件支撑, 采用分布式处理框架, 大幅提高了运维系统的数据并发处理能力。实验结果证明, 该方案相对于单点数据处理模式, 在各环节的处理性能都大幅提升。在多个车辆智能运维项目的实践也证明, 基于该方案的轨道交通智能运维系统能有效满足车辆下发的海量数据的快速处理需求。

**关键词:** Lambda 架构; 智能运维系统; 大数据; 轨道交通

**中图分类号:** TP392

**文献标识码:** A

**文章编号:** 2096-5427(2021)02-0067-05

**doi:**10.13889/j.issn.2096-5427.2021.02.200

## Application of Lambda Architecture in the Intelligent Operation and Maintenance System for Rail Transit Vehicles

HU Weimin

(Zhuzhou CRRC Times Electric Co., Ltd., Zhuzhou, Hunan 412001, China)

**Abstract:** In view of the massive on-board data processing requirements in the intelligent operation and maintenance system for rail transit vehicles, this paper proposes a data processing architecture scheme based on the principle of Lambda architecture in order to overcome the performance bottlenecks in the reception, processing, storage, query, analysis and other links of massive data under the single-point data processing mode. It adopts distributed server cluster as the underlying hardware support and uses distributed processing framework, which greatly improves data concurrent processing ability of the operation and maintenance system. Experimental results show that compared with the single point data processing mode, processing performances are greatly improved in every link by this scheme. The application of many railway intelligent operation and maintenance projects also proves that the rail intelligent operation and maintenance system based on this scheme can effectively meet the demand of rapid processing of massive data issued by vehicles.

**Keywords:** Lambda architecture; intelligent operation and maintenance system; big data; rail transit

## 0 引言

智能运维有利于提升城市轨道交通装备的智能

化应用与管理水平, 是近几年来在轨道交通行业内发展起来的一种新运维模式, 其核心是利用各种传感装置获取车辆设备运行的实时状态和故障数据, 并借助大数据、云计算和人工智能等技术对设备系统进行故障诊断和状态管理<sup>[1]</sup>。如何在保障轨道交通系统安全可靠运营的基础上最大限度地降低维修成本, 并在满足可持续发展战略要求的同时提升轨道交通

收稿日期: 2020-10-10

作者简介: 胡卫民 (1989—), 男, 工程师, 主要研究大数据、数据集成等技术在轨道交通中的应用。

基金项目: 国家重点研发计划 (2016YFB1200401)

设备智能化管理水平,成为轨道交通行业广泛关注和研究的热点之一<sup>[2]</sup>。国外,欧盟于2014年开始实施Shift2Rail科技创新项目<sup>[3]</sup>,推进状态修、自动驾驶等技术的研究,以支撑欧洲轨道交通企业保持全球竞争优势。美国IBM公司在2009年就SmarterRailroad概念提出“用更全面的互联互通、更透彻的感知和度量、更深入的智能化,实现智能信息的网络化”<sup>[4-5]</sup>。阿尔斯通公司推出基于云计算的HealthHub预测性维护智能系统,其涵盖车载设备、地面检修装置、数据处理中心等子系统<sup>[6]</sup>。国内,轨道交通装备生产及应用单位、核心部件供应商也都在积极推进车辆智能运维系统的建设。上海地铁公司较早地开展了车辆智能运维系统国家示范工程项目,上海地铁17号线在车辆运维工作中全面应用了车辆智能运维系统,目前其运维人车比下降了20%,达到0.4<sup>[7-8]</sup>。青岛地铁公司在城市轨道交通车辆智能运维方面也开展了大量的工作,所上线的智能运维系统可以显著降低维修保障费用,大幅提高维修保障效率及设备的可靠性<sup>[9]</sup>。

车辆智能运维系统是一种建立在设备基础上、以状态修模式为主要发展目标的智能化、数字化系统。其依托大数据、人工智能、故障预测与健康管理的(prognostics and health management, PHM)等新技术,结合设备履历数据,并借助实时监控设备,采集和分析城市轨道交通车辆的运行和检修数据,判断设备故障趋势,诊断设备的运用健康状态,从而实现故障预警和分级报警,指导关键设备现场维修作业的智能化

管理<sup>[9]</sup>。由于需要实时处理来自车辆设备终端、大量高密度大体量的数据,原有的单点数据处理模式在接收、处理、存储、数据接口服务、数据分析等各个方面均已无法满足现有数据规模要求。为此,本文提出一种基于Lambda架构原理搭建的、针对海量车载数据的数据处理架构方案,并将其与单点架构方案进行了比较。

## 1 基于Lambda架构的车辆智能运维系统总体架构

车辆智能运维系统承担着确保“车辆运营可靠性”的重大责任<sup>[10]</sup>。系统应用大数据、人工智能、PHM等新技术,通过汇聚线路或线网所有车辆实时运行状态数据、故障数据、离线文件数据、检修作业数据以及音视频文件等数据,经过数据处理平台,最终给应用层做数据呈现。图1示出基于Lambda架构的车辆智能运维系统总体架构。

### 1.1 Lambda架构简介

Lambda架构(图2)可以实时处理海量高并发数据,其首先将大数据系统架构分为批处理层(batch layer)、实时处理层(speed layer)和服务层(serving layer)3个层次,然后分别对这3个层次进行操作,以便缓解数据库的读写压力和降低实时数据处理的复杂程度<sup>[11]</sup>。该架构整合离线计算和实时计算,融

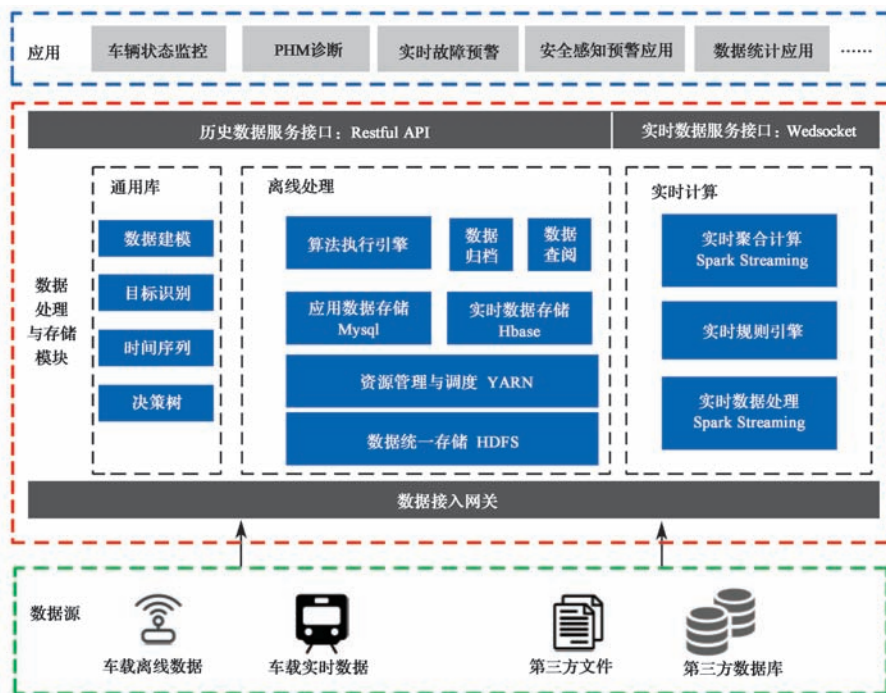


图1 基于Lambda架构的车辆智能运维系统总体架构

Fig. 1 Architecture of the vehicle intelligent operation and maintenance system based on Lambda architecture

合不可变性、读写分离和复杂性隔离等一系列架构原则，可集成 Hadoop，Kafka，Storm，Spark，Hbase 等各类大数据组件<sup>[12]</sup>。

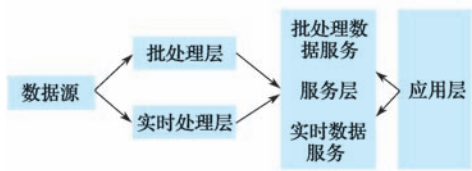


图 2 Lambda 三层架构  
Fig. 2 Lambda three-tier architecture diagram

1.2 基于 Lambda 架构的车辆智能运维系统总体架构

基于 Lambda 架构的车辆智能运维系统总体架构（图 1）分为 3 层，即数据源、数据处理与存储、应用，符合 Lambda 架构的 3 层设计思路。在该架构中，数据源涉及车载离线和实时数据、第三方文件及第三方数据库数据；数据处理与存储层包括数据接入网关、数据处理与存储模块、数据接口模块；应用层涉及车辆状态监控、关键部件 PHM、安全感知预警、智能分析（载荷、能耗、旅速、空转滑行、司机行为分析、网压异常）等应用场景，其支撑数据由数据处理与存储层提供。

2 数据处理与存储

数据处理与存储层是 Lambda 架构方案的核心，其底层由分布式服务器集群类硬件支撑，软件集成了负载均衡、高性能通信、非关系型数据库及分布式消息队列等组件，采用流处理和批处理框架，可以充分利用底层硬件计算存储能力来提高数据处理效率。该层从功能上可以分为数据输入网关、数据处理与存储、数据接口和离线数据分析几个模块。

2.1 数据接入网关

由于线路或线网车辆多、数据下发频率高，按照每条线路 30 辆车、单车每 500 ms 下发 10 000 个数据的点位信息到地面的数据量进行计算，单条线路数据网关需要具备 60 万点/s 的接收能力；如果按照 6 到 8 条线路规模进行线网规划的话，网关的数据接收能力必须达到 360 万~480 万点/s。若采用以往的单点架构，在数据接入层会出现数据接收不及时的情况，这会导致数据丢失。对此，本文设计了一种数据接收网关架构，具体如图 3 所示。图中，车载终端设备通过 3G/4G 网络将数据直接下发到地面具备固定 IP 或者域名的服务器；在服务器端，部署了一套 Nginx 负载均衡软件，数据被均衡转发至后端的多个

服务器；每个服务器上会运行一套基于 Netty 的数据接收和预处理程序，其首先完成对车的应答并对数据进行相应的预处理，然后再将数据写入到下游的 Kafka 集群。该网关在数据接入性能方面能够满足线性扩展需求，且基本能够满足城轨车辆秒级、毫秒级数据接入的需求。

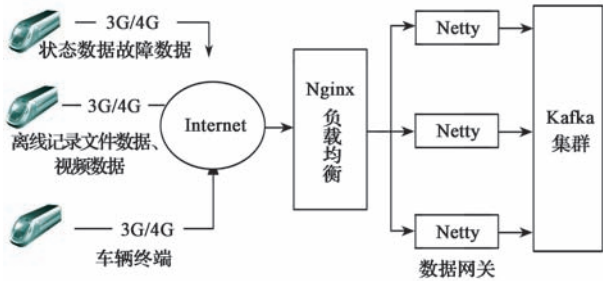


图 3 Nginx+Netty 的分布式数据网关  
Fig. 3 Distributed data gateway of Nginx+Netty

2.2 数据处理与存储模块

数据被网关接收后，系统根据数据所支撑的业务类型选择相应的数据处理流程，并将处理后的数据进行持久化存储（图 4）。

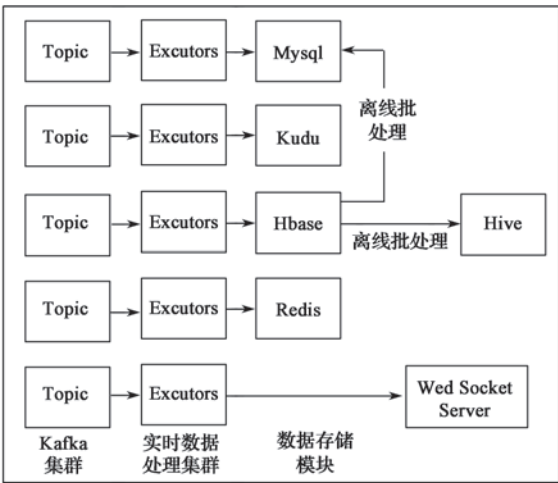


图 4 数据处理与存储架构  
Fig. 4 Data processing and storage architecture diagram

数据进入到 Kafka 集群之后，在实时处理环节按照不同的业务特点对数据进行处理和存储。如对于故障、预警、PHM 数据，使用 Kudu 数据库进行存储；对于状态数据，使用 Redis 和 Hbase 数据库进行存储；对于关键部件寿命数据，用 Mysql 数据库进行存储。

等到全部数据均被写入存储模块之后，离线批处理程序会从 Hbase 或者 Hive 数据库中抽取相应的数据进行智能分析作业，如载荷分析、能耗分析、旅速分析、空转滑行分析、司机作恶行为分析及网压异常分析等，并将分析结果写入到 Mysql 数据库进行存储。



2.3 数据接口模块

数据经过流处理和批处理后，会生成一些结果性质的数据并被持久化到存储系统中，上层应用系统需要对数据进行展示。数据服务接口主要分为实时数据服务接口和历史数据服务接口（图 1）。目前，实时数据接口，即 WebSocket 协议数据接口，主要用来满足实时状态监视、故障实时推送、故障预警推送、故障诊断结果推送等功能需求；历史数据接口，即 Restful API 接口，主要用于非实时应用场景下的各种数据服务。

2.4 数据分析

在业内，数据分析的技术路线一般有两种，一种基于业务规则，另一种基于数理统计。本文主要讨论基于业务规则的数据分析方法在 Lambda 架构下的分布式实现。

对于离线数据，采用 Spark 对存储到 Hbase 或者 Hive 中的车辆状态数据进行定时分布式分析计算，并将数据分析结果存储到 Mysql 数据库中以便于应用端直接获取。假设一列车一天的数据量为 20 GB，针对这些数据需要进行 50 个分析顶点的计算，一条地铁线路若有 40 列车，那么一天则需进行 2 000 次大规模的数据计算。在以往的模式下，受限于单机的内存容量，只能进行单台车、单项点的分析；而采用 Spark 和 Hbase 等分布式技术组件后，则数据按照车辆进行分区后可同步开展分析计算，同时还可对各台车的分析顶点进行并行计算，大大提高了数据分析的效率。

对于实时数据，采用 SparkStreaming 或者 Flink 对写入 Kafka 消息队列的实时数据按车进行分区消费，保证数据按照单台车的时间有序排列；按照特

定的时间窗口对数据进行特定的业务规则计算，并将符合规则的数据进行实时推送，便于第一时间发出警报。

3 实验验证和对比

为了寻找单点数据处理模式架构的性能瓶颈点，并发现在 7 倍硬件规模下 Lambda 架构的性能相对单点数据处理模式的优势程度，本文利用 8 台相同配置的 X86 机架式电脑（单台 128 GB 内存，2 个物理 CPU，32 核，500 GB SSD 系统盘，8 TB SATA 硬盘）进行实验，其中 2 台用于搭建通信服务器，5 台用于搭建 5 个节点的大数据集群的 Lambda 整体验证环境，剩余 1 台用来搭建单点架构的数据处理对比环境。在 2 台通信服务器上部署了基于 Netty 的数据接收网关程序，并通过 Nginx 实现两个节点之间的负载均衡；在大数据集群上，部署 Kafka，Hbase，Kudu，Hive，Spark，Redis 等组件，并按照 Lambda 架构的设计原则，将相应的软件模块部署到大数据平台上进行数据处理作业。通过程序模拟不同车载终端设备进行数据发送，形成完整的车地数据处理链路。

单点数据处理模式架构由数据仓库和关系型数据库构成<sup>[13]</sup>。其在数据接入、数据处理和存储、数据接口方面均存在一定的性能瓶颈，容易出现数据丢包、数据积压和处理延时现象；由于无法线性扩展存储，当数据量增加到一定级别时，容易出现查询速度慢、数据实时显示不流畅等问题。单点数据处理模式下所遇到的问题，在基于 Lambda 架构的模式下都能迎刃而解。两种架构模式具体对比分析结果如表 1 所示。

表 1 单点数据处理模式架构和 Lambda 架构对比分析结果  
Tab. 1 Comparison and analysis results of single-point data processing mode and Lambda architectures

| 对比项目   | 单点数据处理模式架构                                                                        | Lambda 架构                                                            |
|--------|-----------------------------------------------------------------------------------|----------------------------------------------------------------------|
| 系统复杂度  | 低                                                                                 | 高                                                                    |
| 数据接入性能 | 低（数据并发低于 100 包 /s 时，不会出现丢包；数据并发达到 300 包 /s 时，会出现数据丢包现象；达到 600 包 /s 时，数据丢包率达到 10%） | 高（数据并发达到 5 000 包 /s 时，未出现数据丢包，由于实际项目目前远未达到 5 000 包 /s，所以没有进一步进行压力测试） |
| 数据处理性能 | 低（数据并发达 100 包 /s 时，数据处理不延时；到 200 包 /s 时，数据处理出现较长延迟）                               | 高（数据并发达到 5000 包 /s 时，数据处理未发现延时）                                      |
| 数据存储扩展 | 不可水平扩展（单表数据量达到千万级别后，只能分表或者分库）                                                     | 可以水平扩展（单表可以支持上千亿条数据）                                                 |
| 数据接口性能 | 低（单表数据达到 2 000 万条时，查询速度下降明显，查询 3 个点位，10 min 位数据，平均 20 s 左右）                       | 高（查询 30 个点位，10 min 左右数据，2 s 内返回数据）                                   |
| 数据分析   | 只能对数据进行单点分析（以一台车一天 20 GB 数据计，只能分析 50 个顶点中的 20 个）                                  | 可实现对数据的分布式分析，性能成倍数增加（一台车一天 20 GB 数据，可分析 50 个顶点中的所有顶点且只需要花费 1 h 时间）   |

新的数据处理架构虽能够满足车辆智能运维系统海量的数据处理、存储、查询、分析需求，但相较于传统的架构，其涉及的组件较多（如 Kafka，Hbase，Kudu，Hive，Spark，Redis 等），学习曲线相对陡峭，软硬件成本相应增加。

通过实验结果对比可以发现：在数据接入和处理方面，单点数据处理模式在数据流量达到 100 包 /s 情况下就会遇到性能瓶颈；在存储、接口性能、数据分析方面，7 倍硬件资源优势的 Lambda 架构性能方面远高于 7 倍单点模式的性能。因此，如果仅针对数据接入和处理，数据流量小于 100 包 /s 的项目或线路可以采用低成本的单点模式架构方案；但若对数据存储容量、数据接口查询性能、对数据分析项点有更高的要求，则建议一开始就采用 Lambda 架构作为整体的数据处理架构。

4 结语

本文对轨道交通车辆智能运维系统中基于 Lambda 架构搭建的数据处理平台部分进行了详细的介绍，并和传统的单点数据处理架构进行了对比。基于该架构的轨道交通智能运维系统目前已被成功运用到高速铁路系统安全保障项目以及多条地铁线路中，如广州地铁 2 号线、郑州地铁 4 号和 5 号线、武汉地铁 5 号线等，为智能运维系统提供了稳定可靠的数据平台支撑。

针对目前 Lambda 架构相对复杂的情况，未来可以考虑在数据处理和存储层面寻找一些集成度更高的组件进行替代，以降低系统复杂度，用更少的组件就支撑起整个架构。

参考文献：

[1] 胡佳琦. 上海市轨道交通车辆智能运维系统研究与应用 [J]. 现代城市轨道交通 ,2019(6):16-21.

HU J Q.Exploration and research on intelligent operation and maintenance system of urban rail transit vehicles [J].Modern Urban Transit, 2019(6):16-21.

[2] 刘述芳. 城市轨道交通关键设备智能运维系统初步建构 [J]. 设备管理与维修 ,2018(2):22-23.

[3] 傅嘉俊. 上海轨道交通车辆智能运维系统 [J]. 城市轨道交通 ,2020(7):23-26.

[4] 张铭媛. 基于 Apache Spark 的新三板大数据监管平台构建研究 [D]. 北京 : 中国财政科学研究院 ,2017.

[5] 罗琦芳. 一种基于 Lambda 架构的电信数据平台解决方案 [J]. 电子技术与软件工程 ,2017(11):182.

[6] DIERKX K. The Smarter Railroad: An Opportunity for the Railroad Industry[EB/OL].[2020-09-09]. <http://ishare.iask.sina.com.cn/f/8387985.html>.

[7] Alstom launches HealthHub, an innovative tool for predictive maintenance[EB/OL].(2014-09-23)[2020-09-09].<https://www.alstom.com/press-releases-news/2014/9/innotrans2014-alstom-launches-healthhub-an-innovative-tool-for-predictive-maintenance->.

[8] 苏树鹏. 基于 Lambda 架构的移动互联大数据平台架构的设计与应用 [J]. 企业科技与发展 ,2016(6):66-68.

[9] 潘莹. 故障预测和健康管理技术在地铁车辆运维中的应用 [J]. 控制与信息技术 ,2020(4):91-95.

PAN Y. Application of Fault Predication and Health Management in Metro Vehicle Operation and Maintenance[J]. Control and Information Technology, 2020(4):91-95.

[10] 王同军. 智能铁路总体架构与发展展望 [J]. 铁路计算机应用 ,2018(7):1-8.

WANG T J. Overall framework and development prospect of intelligent railway [J]. Railway Computer Application, 2018(7):1-8.

[11] 刘丙林, 朱佳, 李翔宇. 城市轨道交通车辆智能运维系统探索与研究 [J]. 现代城市轨道交通 ,2019(6):16-21.

LIU B L, ZHU J, LI X Y. Exploration and research on intelligent operation and maintenance system of urban rail transit vehicles [J]. Modern Urban Transit, 2019(6):16-21.

[12] 韩嫻. Lambda 架构在处理海量高并发数据中的应用 [J]. 现代电视技术 ,2016(12):70-72.

[13] 曹勇, 张玉文, 龚艳. 基于大数据和云计算的车辆智能运维模式 [J]. 城市轨道交通研究 ,2020(4):69-73.

CAO Y, ZHANG Y W, GONG Y. Subway Vehicle Intelligent Operation and Maintenance Mode Based on Big Data and Cloud Computing [J]. Urban Mass Transit, 2020(4):69-73.