

# 工业物联网大数据平台在工程机械行业的应用

Application of Industrial IoT Big Data Platform in Construction Machinery Industry

■ 江苏徐工信息技术股份有限公司 张启亮 韩 键 姜丽萍

**摘 要** 针对当前物联网系统应用没有统一通信协议和系统平台的瓶颈问题，设计并实现了工业物联网大数据平台，以具有泛接入特性的 XM2M 协议为基础，涵盖了物联网系统数据接入、数据处理、数据存储、数据交换和数据分析等内容，并进行了示范应用。该平台是以 Hadoop、Kafka、Storm 等开源大数据处理技术为基础，以打造物联网大数据 PaaS 云平台为手段，目的是为客户提供 PaaS 云服务、开放数据接口服务以及行业化整体解决方案，降低物联网系统在跨行业、跨领域实施的成本，建设工业物联网大数据生态圈。

**关键词** 工业物联网 大数据 分布式 云平台 整体解决方案 泛接入

**Abstract:** In order to settle the crucial problem that there is not standardized communication protocol and platform, the big data platform of industrial internet of things is designed and realized. The big data platform of IoT, which covers the data access, data processing, data storage, data exchange and data analysis, is based on the XM2M protocol with the characteristics of extensive access, and has been applied in demonstration. The platform is based on Hadoop, Kafka, Storm and other open source data processing technology, using the big data internet of things PaaS cloud platform as the technology method. The purpose is to provide customers with PaaS cloud services, open data interface services and industry as a whole solution to reduce the cost of the internet of things in cross-industry, cross-domain implementation of building industrial networking large data ecosystem.

**Keywords:** industrial internet of things; big data; distributed; cloud platform; the overall solution; generic access

## 1 引言

目前，国内企业的物联网应用多以各自建设单独系统的方式，所提供的业务功能也仅限于设备的定位、监控等基本功能。伴随着企业需求的不断增加，原有物联网应用越来越成为企业转型升级的瓶颈，主要表现为两个方面：一是由于建设之初没有统一标准<sup>[1]</sup>，物联网应用系统无法适应业务的快速扩展，同时不同行业的差异化更增加了物联网系统的实施成本和复杂程度；二是当接入设备的数量达到一定级数后，传统的系统架构已经无法满足系统需求<sup>[2]</sup>，即使是简单的查询请求，传统平台的响应也是分钟级别，即使通过软硬件的升级，也无法解决性能

问题。

江苏徐工信息技术股份有限公司以打造跨行业、支持泛接入的工业物联网大数据整体解决方案为目的，通过制定标准化的泛接入协议解决不同行业设备的接入问题，通过搭建基于大数据的物联网云平台根除系统瓶颈，并且在工程机械行业的龙头企业徐工集团开展典型应用，极大地推动了国内工业大数据的应用与发展。

## 2 云架构工业物联网大数据平台设计

### 2.1 系统架构

工业物联网大数据平台是基于分布式的高性能、

项目来源：徐工机械工业互联网化提升项目之子项目“徐工全球物联网统一平台”，项目编号：2016-320351-35-03-514440。

高可用、易扩展、易开发、易管理一体化物联网的大数据产品，涵盖了物联网系统中的各个环节，包

括数据接入、数据处理、数据存储、数据交换<sup>[3]</sup>。整体架构如图 1 所示。

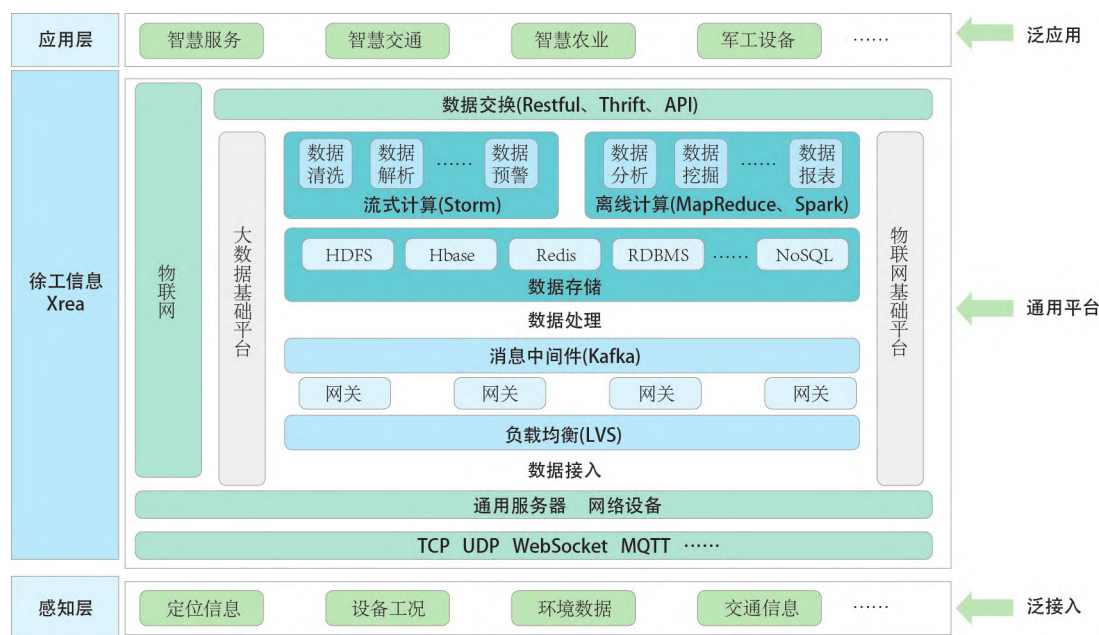


图 1 工业物联网大数据平台系统架构

平台架构<sup>[4]</sup>中，主要分为三大层：泛接入层、通用平台层和泛应用层。

泛应用层无需对数据进行计算和分析，只需要通过查询接口获取平台层中已经处理过的数据并进行展示。

通用平台层为核心，主要提供数据的接入、数据的存储、数据的计算（包含实时计算和批量计算）和数据的交换和监控管理等服务。

泛接入层全部采用 X86 通用服务器或者一体化设备，不需要 IBM 小型机等昂贵的计算设备，也不需要 EMC 等高端的存储设备，从整体上可以大幅拉低成本；此外，在泛接入层设计了面向不同设备类型的泛接入通信协议 XM2M，支持不同行业领域、不同类型设备的接入，从根本上解决了工业物联网大数据平台跨行业应用的问题，降低了应用的实施和推广成本。

## 2.2 通用平台层设计

通用平台层根据功能<sup>[5]</sup>可以分为五个子系统，

分别为：数据接入子系统、数据存储子系统、数据处理子系统、平台服务子系统、监控报警子系统。

### 2.2.1 数据接入子系统

#### (1) 负载均衡

平台使用 LVS (Linux Virtual Server, Linux 虚拟服务器) 负载均衡技术。LVS 工作在网络层，可以实现高性能、高可用的服务器集群技术。LVS 和网关服务器部署在同一个局域网内，它们的 IP 是相同的网段，且每个数据网关所在服务器的配置相同，使用 LVS 最高效的 DR 模式和 LC 调度算法。LVS 可以基于 IP 和端口进行负载均衡，因此，对于不同协议的网关通过分配不同的端口可以采用同一个 LVS 进行负载均衡。

为了保证 LVS 的高可用，避免出现单点故障，可以采用 Keepalived+LVS 这种轻量级的解决方案。Keepalived 是 Linux 下面实现 VRRP 备份路由的高可靠性运行件，在这里主要用来检测 LVS 的状态和 LVS 主备之间的故障切换。实际应用环境中可以

根据负载压力选择硬件负载 (F5、Radware) 来代替 LVS, 以提供更优的接入性能。

#### (2) 网关

数据网关主要负责接收终端上报的数据、指令下发和数据路由。在试运营和割接初期会存在新老系统同时存在的情况, 这时终端上传的数据要同时发往新老系统, 这就要求数据网关有数据路由的功能。数据路由功能可以通过开关配置。

#### (3) 消息中间件

平台使用 Kafka 作为消息中间件, 传递的消息主要包括: 原始数据 (由网关简单解析后的产物)、轨迹数据、工况数据、报警数据等 (由数据解析程序做完全解析后的产物)。

### 2.2.2 数据存储子系统

#### (1) HDFS

HDFS 是 Hadoop 最核心的设计之一, 能提供高吞吐量的数据访问, 非常适合大规模数据集上的应用<sup>[6]</sup>。平台使用 HDFS 来存储需要离线备份或者需要进行离线计算的数据, 使得平台具备了 HDFS 的三大特性, 即: 可以自动复制文件的新副本保证存储的高可用性、可以自动实时将新增服务器节点匹配进整体阵列实现不停机状态下的系统实时扩容、可以通过分布式算法均摊数据访问请求提高数据吞吐量。

#### (2) Hbase

HBase 是一个面向列的分布式数据库, 平台使用 HBase 来存储不断增长的数据, 如原始通信数据、轨迹数据、工况数据等, 以支持客户端高效的数据访问, 使平台具备动态无缝无宕机扩容、高并发用户数高速读写访问和高可用三大优点。

#### (3) RDBMS

HBase 虽然支持大数据量的存储, 但是不支持复杂的查询。本平台中业务数据在进行展示时需要支持复杂查询, 这部分数据主要存储在关系型数据库。90% 以上的存储已经被 Hadoop 和 HBase 承担, RDBMS 在本平台中负荷较低, 性能比较高。

#### (4) Redis

Redis 是一款开源的、高性能支持持久化的键-值存储系统, 其性能极高, 能支持超过每秒 10 万的读写频率。平台主要使用 Redis 来存储读写频率较高, 数据量较恒定的数据。

### 2.2.3 数据处理子系统

系统主要包括实时处理和批量处理。实时处理主要针对终端定时上报数据进行解析、报警判断等时效性要求高的任务; 批量处理主要针对报表分析、BI 等对时效性要求不高的任务<sup>[7]</sup>。

#### (1) 实时处理

在本平台中, 对大数据处理的计算时效性要求很高, 要求计算能在非常短的时延内完成。目前, 实时计算比较优秀的框架主要有两个: Storm 和 Spark Streaming, 两者都可以提供很好的可扩展性和容错性, 区别主要在于 Storm 对消息的处理可以达到毫秒级的延迟, 而 Spark Streaming 则有秒级的延迟。依据本平台对时效性的要求, 选择 Storm 作为实时处理框架。

#### (2) 离线计算

本平台离线分析模块使用 MapReduce 计算框架, 是 Hadoop 最核心的设计之一, 适合离线处理超大规模数据集, 如日报表分析、月报表分析、年度报表分析等。在整合 MapReduce 高度可扩展、高容错能力、公平调度算法和动态灵活资源分配等优势的前提下, 本平台对 MapReduce 进行进一步的封装, 便于业务工程师进行开发和离线分析任务调度。

### 2.2.4 平台服务子系统

#### (1) 数据导入

网关接收到的终端原始数据和实时计算产生的部分数据分别需要存放在 HDFS 和 Hbase 里。如果网关在接收到终端的原始数据时直接写入 HDFS, 会造成网关逻辑复杂, 单条消息处理时延增大, 处理能力下降; 如果网关批量写入原始数据, 性能会有所提升, 但是当网关宕机时会有数据丢失。如果实时计算直接把数据写入 Hbase 中, 会有同样的问题。在本平台使用 Kafka 进行解耦, 网关把原始数据写入 Kafka 中, 实时计算把解析数据写入 Kafka 中,

使用 Flume NG 将 Kafka 中的数据导入到相应的存储中<sup>[8]</sup>。

## (2) 数据交换接口

本平台提供三种数据交换接口：跨语言的 Thrift RPC 接口、JAVA API 接口和 REST API 接口。

### ● 跨语言的 Thrift RPC 接口

本平台提供跨语言 (目前支持 C++、JAVA、Python、PHP、Ruby、Erlang、Perl、Haskell、C#、Cocoa、Smalltalk 和 OCaml) 的数据交换接口,旨在提供业务层和平台层的数据交互服务,包括实时数据获取、历史数据获取、终端指令下发及应答检查。该数据交换接口基于 Thrift 开发,特性如下:数据包比 XML 和 JSON 小,传输的效率较高;性能比较好,支持高并发,且技术成熟;支持多语言的客户端,使用 thrift 可以快速生成客户端代码。

### ● JAVA API

本平台提供与平台各模块直接交互的 JAVA API,为业务层提供了高性能的数据交互服务,包括实时数据获取、历史数据获取、终端指令下发及应答检查。

### ● REST API

本平台为三方平台提供统一的 REST API 接口,可以使用几乎任何客户端在任何编程语言与 REST API 进行交互,以及编写和测试应用程序。接口支持设备的管理,实时数据、历史数据的获取,终端的指令下发。

## 2.2.5 监控报警子系统

### (1) 系统监控

本平台使用 Ganglia 来监控每个节点的工作状态,通过长期监测系统运行状况可以对合理调整、分配系统资源提供参考。系统运行初期可以依据系统监控数据对平台进行优化。

### (2) 系统告警

本平台使用 Nagios 提供三个层面的监控和告警:系统层面,使用插件监控服务器的 CPU、内存、硬盘使用率、I/O 负载;软件层面,本平台的所有进程在启动后把自己的 PID 写入指定目录的文件内,文件名为进程名。Nagios 根据 PID 定时检查进程的状态,对挂掉的进程显示报警;业务层面,本平台中的各个模块可以自定义告警。平台提供多种告警通知方式,包含邮件或短信或 APP 等。

## 3 应用效果及分析

随着信息化和自动化技术的飞速发展,工业大数据将成为制造产业革命的核心。徐工信息自主研发的物联网信息终端及新一代工业物联网大数据平台对中国工业大数据应用进行了深入探索,很好地解决了泛接入和系统性能瓶颈的问题,并在徐工集团内进行了典型应用,效果非常好。

### 3.1 解决方案

#### 3.1.1 工业大数据之源——工程机械黑盒子 (见图 2)



图 2 工程机械黑盒子



通过这些物联网智能终端,可以捕捉收集遍布全球的工程机械设备的的大数据,并将这些数据准确传输到工业物联网大数据平台进行深度的数据挖掘与分析。

### 3.1.2 工业物联网信息服务平台

通过徐工信息物联网终端的安装,机械设备的

实时位置信息可以在工业物联网信息服务平台上实时准确查询,用户可以通过平台看到每一台设备的实时位置等信息(如图3所示)。设备位置异常时也会推送预警信息,实现客户对设备的精准监控,预防设备偷盗或冒用现象。

安装了物联网信息终端的设备运转信息,包括



图3 设备定位显示

设备生产作业实时数据监测。当车辆发生故障时,会自动对故障进行诊断分析,找到故障解决方案,协助售后人员及时赶到现场维修,提高售后服务质量。

### 3.1.3 工业大数据——深度数据挖掘

通过数据挖掘,得到设备在线数量分布、开工率分析、平均工时分析及区域工作热度分析,如图4所示。可以对售后服务备件市场进行预测,优化生产计划及库存,指导企业进行精准营销和服务。

### 3.1.4 工业物联网大数据平台

工业物联网大数据平台具备全面的物联网设备接入功能、高效的物联网数据处理功能、灵活的物联网数据接口、多方位的物联网数据保护、丰富的物联网大数据分析能力,可以推动价值链整合与业务创新,是工业4.0、工业互联网的共性基础技术,为工业企业提供工业物联网大数据整体解决方案。

## 3.2 方案实施后的价值或成果

工业物联网大数据平台由底层设备、统一数据

中心、以及大数据分析模块构成,该平台已经在集团内及外部市场成功推广,面对不同的客户,实现了从底层设备接入管理到工业大数据分析,未来随着产品的推广,可根据客户需求和前沿技术的发展继续丰富完善工业物联网大数据平台,保持平台的鲁棒性和技术先进性。

## 3.3 方案的创新点

### ●全面的物联网接入功能

接入类型多:屏蔽协议差异性,支持接入多种物联网传感器和采集器,并且可以自定义扩展;兼容性强:兼容已有物联网文件和数据的接入;数据协议泛支持:SOAP、MQTT、JSON或自定义二进制协议;通讯协议泛支持:TCP、UDP、HTTP、Web Socket。

### ●高效的物联网数据处理功能

接入性能高:支持千万级物联网传感器和终端数据并发接入;实时计算效率高:实时处理吞吐量到达百万级,延迟只需数秒钟;离线计算效率高:



图 4 工业大数据应用

对 TB 级数据进行分布式处理，只需分钟级；高效查询：在千亿数据集内，达到秒级的查询响应时间；高效存储：支持结构化、半结构化和非结构化等数据的 PB 和 EB 级存储，并可动态按需扩容，服务不中断。

#### ● 灵活的物联网数据接口

支持 Restful 接口：灵活实现与 Restful 接口应用程序之间的数据交换和共享；支持 Thrift 接口：灵活实现跨语言环境下开发的应用程序之间的数据交换和共享；API 直连接口：灵活实现 API 直连接口应用程序之间的数据交换和共享。

#### ● 多方位的物联网数据保护

链路安全：通过 SSL 和 TLS 保证链路安全；网络安全：通过防火墙等硬件设备防止网络攻击；接入安全：通过密钥鉴权对数据的访问进行有效控制；存储安全：通过冗余副本保证数据的存储安全；数据安全：通过数据校验机制保证数据的防篡改。<sup>[5]</sup>

#### 参考文献

[1] 赵博龙, 杨洁. 物联网现状应用调查 [J]. 科技展望,

2015(21): 21.

- [2] 朱仲英. 传感网与物联网的进展与趋势 [J]. 微型电脑应用, 2010(1): 1-3.
- [3] 郭闻博, 林泓. 基于 Hadoop 平台的数据规范化处理研究 [J]. 空军预警学院学报, 2016(3): 65-68.
- [4] 王祎茜. 基于 Web 的物联网应用体系架构和关键技术研究 [J]. 硅谷, 2014(12): 61, 78.
- [5] 姜文. 基于 Hadoop 平台的数据分析和应用 [D]. 北京: 北京邮电大学, 2011.
- [6] 刘海龙, 宿宏毅. 利用 Hadoop 云计算平台进行海量数据聚类分析 [J]. 舰船科学技术, 2016(14): 152-154.
- [7] 王欣. 基于 Hadoop 的云安全存储系统的平台设计 [J]. 佳木斯大学学报 (自然科学版), 2016(4): 56-58.
- [8] 饶慧. 基于 Hadoop 的大数据计算技术分析 [J]. 科技展望, 2016(12): 14.

(收稿日期: 2017-03-17)