

文章编号: 1005-8451 (2015) 08-0023-04

# 铁路总公司舆情监测管理系统设计

黄肖峰<sup>1,2</sup>

(1. 中国铁道科学研究院 研究生部, 北京 100081)

2. 中国铁路信息技术中心, 北京 100844)

**摘要:** 介绍铁路总公司舆情监测管理系统的设计与实现, 介绍系统建设目标, 给出系统的总体结构和系统功能, 并简要介绍系统使用的HADOOP和全文检索这两大关键技术, 对系统上线后的效益做出展望。

**关键词:** 门户网站; 系统设计; 虚拟化; HTML5

**中图分类号:** U29 : TP39 **文献标识码:** A

## Public Opinion Inspection Management System of China Railways

HUANG Xiaofeng<sup>1,2</sup>

(1. Postgraduate Department of China Academy of Railway Sciences, Beijing 100081, China;

2. Information Technology Center of China Railways, Beijing 100844, China)

**Abstract:** This article introduced the design and implementation of Public Opinion Inspection Management System of China Railways, described the building objective, the general structure and function of the System, gave a brief introduction of HADOOP and full text search which were the main techniques used in the System. Finally the article gave a prospective of the benefit.

**Key words:** portals; system design; virtualization; HTML5

现阶段随着我国铁路改革, 铁路实现了“政企分开”, 铁路总公司全方位走向市场, 为顺应市场潮流, 总公司推出了一系列利民措施, 例如: 网络购票、电话订票、实名制、货运电子商务等一系列服务措施, 但同时也带来一系列的诟病, 在社会中产生不同舆论的声音。因此在改革的过程中, 为了更好地了解社情民意, 根据民意不断改进铁路工作和服务质量, 亟须加强对互联网铁路相关舆论的舆情监测管理。

充分运用现代信息技术, 构建基于大数据的铁路总公司舆情监测管理系统, 通过对互联网海量信息中与中国铁路相关的舆情监测、舆情分析、舆情处置、效果评估等, 可以为综合实现舆情采集监测分析处置与导控全生命周期的闭环管理提供重要支持。

## 1 系统建设目标

充分运用现代信息技术, 构建基于大数据的舆情服务平台, 通过对互联网海量信息中与中国铁路相关的舆情监测、舆情分析、舆情处置、效果评估、

收稿日期: 2015-01-08

作者简介: 黄肖峰, 在读硕士研究生。

舆情报告、舆论引导、新媒体宣传、新闻宣传以及网评员队伍管理, 提高舆情采集监测的及时性和准确性, 提高分析互联网舆情问题的能力, 规范舆情处置协同工作的流程, 提升舆情处置与舆论引导、新媒体宣传的成效, 加强舆情人才队伍建设, 为中国铁路总公司及时了解互联网铁路相关社情民意、引领舆论向导、进行科学决策, 综合实现舆情采集监测分析处置与导控全生命周期闭环管理提供重要支持。

建设目标包含以下几个方面。

### 1.1 提升舆情信息采集的广度和深度

实现包括新闻网站、论坛、博客、微博、微信公众账号、移动新闻客户端、QQ群等媒体中舆情信息的全面采集。

### 1.2 提升舆情信息挖掘分析的水平

全面提升铁路舆情信息的深度挖掘和分析的能力, 实现舆情事件多维度的、全面的、深度分析与持续的追踪, 同时能够及时有效的提供预警通知服务。

### 1.3 建立舆情处置工作协同协作与规范化体系

建立一套科学的、高效的、标准的、规范的舆

情事件处置流程体系和覆盖全路两级的立体式舆情事件监测、上报、研判、分析和协调处置的规范化流程。同时对于整个过程进行全面监控和审计,为日后人员绩效考核和相关责任人的追究提供依据。

#### 1.4 提升重大舆情事件应急指挥的能力

以实际业务为主线,建设符合我国铁路舆情分析导控工作信息的展示中心、应急指挥中心,实现所有舆情监测分析信息集中展示、线上联合研判、协调指挥、应急处置过程审计等功能,从而全面提升全路在面对重大负面舆情事件时应急协调指挥与处置的效率。

#### 1.5 建立舆情处置与舆论引导的评估体系

整体提升全路舆情处置与舆论引导、新媒体宣传的效果,建立对中国铁路相关舆情事件应急指挥与处理、舆论引导、新媒体宣传等工作成效,以及舆情从业人员队伍工作绩效的综合评价体系。

#### 1.6 提升舆情人才队伍整体能力

借助全路综合评估体系,建立系统的、统一的从业人员人才队伍管理平台,全面的提升中国铁路行业舆情相关人才队伍的能力水平。

#### 1.7 建立舆情工作知识化无形资产的管理和有效利用

充分借助历史舆情工作中积累的成功理论、方法、处置方案和失败经历,借助数字化存储技术,实现舆情工作中知识化无形资产的管理,综合提高知识在全路范围内的快速传播和有效利用,为中国铁路新闻宣传与舆论引导提供有力支撑。

## 2 系统总体架构

铁路总公司舆情监测管理系统采用统一、安全、平台化设计,从实战的业务角度、考虑行业特色,构建一个覆盖全路的互联网舆情采集、监测、分析、处置与导控的全生命周期闭环管理平台。系统底层采用Hadoop技术架构,实现平台分布式的采集、存储、智能分析和数据库。系统整体技术架构如图1所示。

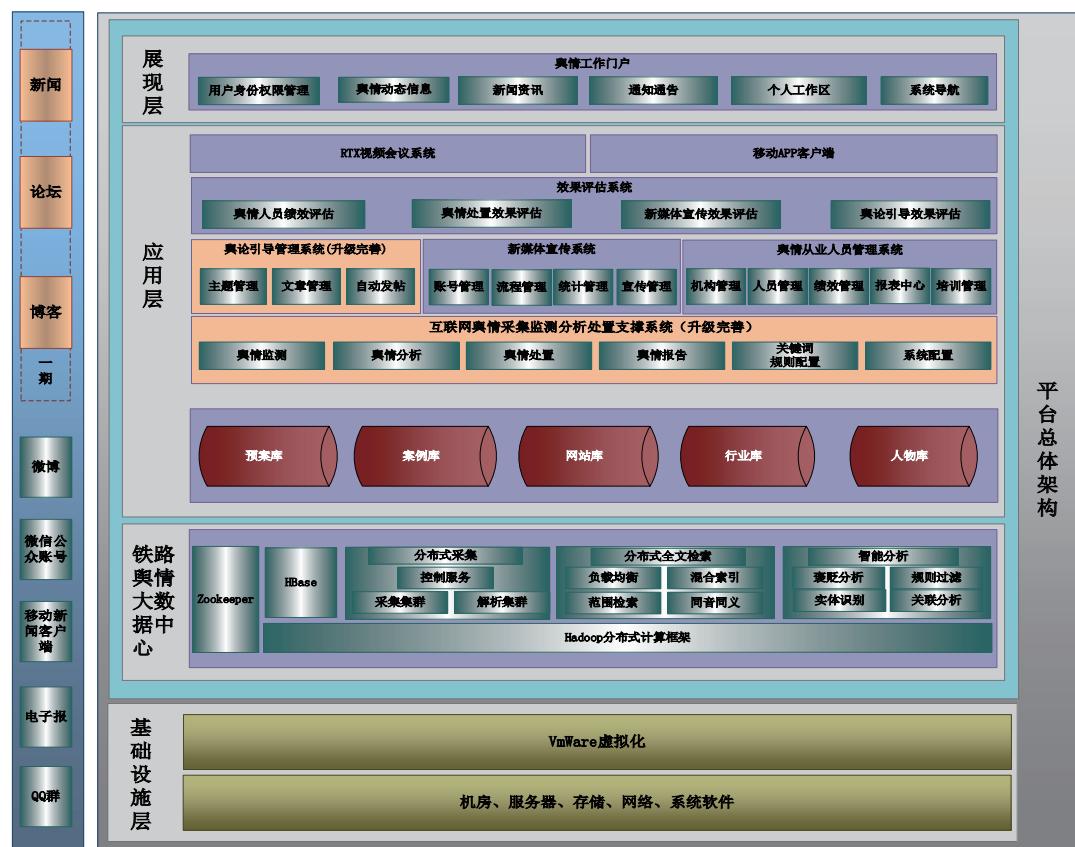


图1 系统整体技术架构

系统采用面向服务的思想,利用松散耦合的分层方式将系统整体总体架构分为4个层次,分别是基础设施层、铁路舆情大数据中心、应用层和展现层。各层之间的界限清晰,功能明确,具有较高的可配置性和伸缩性。

#### 2.1 基础设施层

最底层主要为整个系统平台提供基础的硬件、网络、数据管理和操作系统等底层的软硬件支撑。在此之上通过虚拟化,实现对服务器硬件设备的最大化利用和有效的管理。具体包括对于中国铁路总公司以及各铁路局与应用系统相适应的机房消防改造、

网络升级、安全加固、海量存储建设、服务器等硬件的利旧和采购。

## 2.2 铁路舆情大数据中心

基础设施层之上，通过在 Hadoop 架构之上构建分布式的采集、分布式全文检索、分布式智能分析，为上层应用提供基础的应用支撑，同时通过 Hbase 数据库实现互联网所采集数据的有效存储和管理，从而构建铁路舆情大数据中心。

## 2.3 应用层

应用层作为整个业务平台系统的核心，采用面向服务的思想，由 7 个应用子系统（互联网舆情采集监测分析与处置支撑系统、舆论引导系统、新媒体宣传管理系统、舆情从业人员队伍管理系统、效果评估系统、视频会议系统、移动 APP 客户端）和舆情工作门户组成，同时利用 Web Services 技术为其他外部系统提供基于 SOAP 的远程方法调用接口，用以完成未来与中国铁路总公司其他系统之间的数据交换和同步，并基于多年业务数据的积累，构建铁路舆情预案库、案例库、网站库、行业库和人物库，形成属于铁路总公司的知识管理体系。

## 2.4 展现层

在展现层，通过舆情监测导控信息门户为中国铁路总公司及各铁路局舆情监测分析人员、高层领导、网评员等提供舆情信息、各类舆情分析图表、舆情分析报告、效果评估等信息的展示和图形化操作界面。

## 3 系统应用功能

整个系统包含 7 大模块，覆盖舆情信息的获取、浏览、研判分析和处置、舆论引导、新媒体正面宣传、成效评估和报告决策支持等环节，具体功能包括：

(1) 获取敏感信息：面向互联网实时收集包括新闻网站、论坛、博客、微博、微信公众账号、移动新闻客户端等媒体的信息。

(2) 舆情浏览：通过多级应用平台对涉及铁路相关敏感信息、舆情信息的监测浏览。对于负面敏感信息、重大舆情事件，可通过平台进行上报。

(3) 研判分析与处置：基于舆情监测分析处置支撑系统，实现舆情的深度分析，直观的为用户提供多维度各类舆情分析图表。通过舆情处置流程，实

现舆情事件处置指令下发、审批、处置反馈、上报，及整个过程全面监控与审计。

(4) 舆论引导：加强互联网舆论引导，实现舆论引导指令管理、网评主题管理、网评文章管理、网评文章的实时监测，以及对舆论效果的评估。

(5) 新媒体正面宣传：实现目前新媒体平台信息发布的人工管理、分散管理向信息化管理转变，建立新媒体宣传稿库，提高管理效率和科学化水平。

(6) 成效评估：通过舆情采集、监测和监控审计功能，对于日常舆情处置和舆论引导、新媒体宣传工作实施监测，并基于效果评估和人员绩效考核模型，实现综合评估。

(7) 报告决策支持：对于日常监测、处置和重大舆情事件应急指挥相关分析结论、处置过程、处置结论、引导效果评估、人员绩效评估，都可通过舆情综合报告功能自动生成相应汇总的报告。

## 4 系统关键技术和特点

### 4.1 Hadoop 系统架构

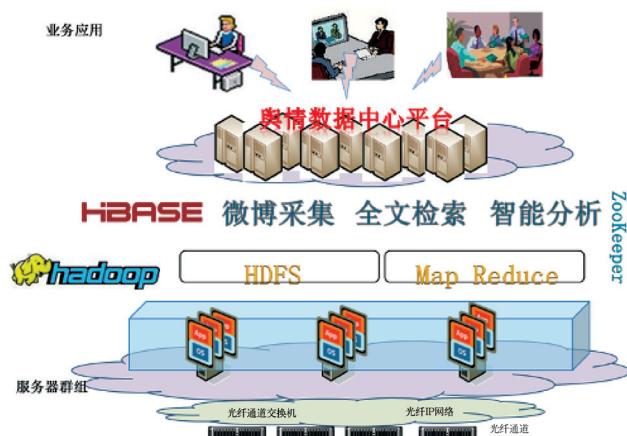


图2 Hadoop框架示意图

如图 2 所示，Hadoop 框架中最核心的设计就是：MapReduce 和 HDFS。MapReduce 用简单的一句话解释就是“任务的分解与结果的汇总”。Hadoop 分布式文件系统 (HDFS, Hadoop Distributed File System)，为分布式计算存储提供了底层支持。

#### 4.1.1 MapReduce

MapReduce 是一个高性能的批处理分布式计算框架，用于对海量数据进行并行分析和处理。与传统数据仓库和分析技术相比，MapReduce 适合处理

各种类型的数据，包括结构化、半结构化和非结构化数据。数据量在 TB 和 PB 级别，在这个量级上，传统方法通常已经无法处理数据。MapReduce 将分析任务分为大量的并行 Map 任务和 Reduce 汇总任务后进行处理。

#### 4.1.2 HDFS文件存储

HDFS 是运行在通用硬件上的分布式文件系统，它提供了一个高度容错性和高吞吐量的海量数据存储解决方案。HDFS 通过一个高效的分布式算法，将数据的访问和存储分布在大量服务器之中，在可靠地多备份存储的同时还能将访问分布在集群中的各个服务器之上，是传统存储构架的一个颠覆性的发展。

### 4.2 全文检索

全文检索是对海量文本数据进行检索的重要而基础的技术手段。所有的数据检索都需要基于独立的全文检索系统来实现，不能只依靠数据库级的检索，全文检索的功能是实现各业务模块的重要资源。独立的检索引擎的数据处理能力可以实现 TB 级的检索，可以有效的满足业务支撑，其主要特点如下。

#### 4.2.1 简洁易用的开发接口

全文检索技术的核心是全文检索服务器，全文检索服务器以 Web Service 方式提供服务。应用系统按照 Web Service 格式标准向服务器发送信息检索和数据更新等请求，并接收返回结果。用户可以用多种形式进行广泛的应用开发，设计出与自身应用程序风格一致的界面或是满足各种特殊的应用需求，极大地降低和保护了用户的软件成本。

#### 4.2.2 采用XML进行数据交换

全文检索在内部和外部均采用 XML 进行交换数据的描述，有力地保证了系统的可扩展性与跨平台

(上接 P22)

临的问题：分布式内存数据库技术用于应对高并发，Hadoop 大数据存储技术用于应对海量存储，同时使用复制服务器、中间件 CTMSX 和 RabbitMQ 消息中间件，实现异构系统间顺畅、准确的进行消息传递的目的。在互联网售票系统中将来可以用作全路实名制数据分析、用户行为分析、系统日志分析等需求。

综上所述，分布式内存数据库技术和 Hadoop 大

性，易于应用开发商的使用。

#### 4.2.3 自动分词技术

全文检索引擎以基于中文分词的混合字词为索引单位，内嵌的分词系统采用以词典为基础的分词算法。系统自带一部通用的系统词典，用户可以通过建立用户词典来定义新的词汇，用户词典一般包含了某个领域的专业词汇。系统在自动分词时将同时参考缺省分词词典和用户词典中的词汇。

## 5 结束语

目前，系统正在紧张的建设当中，相信系统建成后，将有效提高铁路舆情采集监测的及时性和准确性，提高分析互联网舆情问题的能力，规范舆情处置协同工作的流程，为中国铁路总公司及时了解互联网铁路相关社情民意、引领舆论向导、进行科学决策提供更好的支持。

#### 参考文献：

- [1] 王青,成颖,巢乃鹏.网络舆情监测及预警指标体系构建研究[J].图书情报工作,2011(8):56-59.
- [2] 袁文英.从7.23甬温线动车事故看政府如何应对网络舆情[J].湖南行政学院学报,2013(1):20-23.
- [3] 范晨熙.基于 Hadoop 的搜索引擎的研究与应用[D].杭州:浙江理工大学,2013.
- [4] 郑明杭.基于 HDFS 的视频监控系统软件设计[D].杭州:浙江大学,2013.
- [5] 田胜利.针对 HBase 的 MapReduce 数据访问方式的优化[D].长沙:国防科学技术大学,2012.
- [6] 康毅.HBase 大对象存储方案的设计与实现[D].南京:南京大学,2013.

责任编辑 徐侃春

数据存储技术在铁路客票系统中或者铁路行业的其他系统中有着良好的发展前景。

#### 参考文献：

- [1] A Novel Real-time Database Memory Management Approach[C]. Wuhan:2010 The 2nd International Conference on Industrial Mecha-tronics and Automation, 2010.
- [2] 曹英.大数据环境下 Hadoop 性能优化的研究[D].大连:大连海事大学,2013.

责任编辑 徐侃春