

危废处置企业的工况在线监控

陈媛¹, 徐洁¹, 毛佳茗²

(1. 江苏省生态环境监控中心, 江苏 南京 210036; 2. 江苏省环境信息工程研究中心, 江苏 苏州 215021)

摘要:阐述了 HBase 数据库用于危废处置企业工况在线监控系统的特点和实现过程, 并对系统作了测试分析, 结果表明, 该系统有很高的吞吐率, 并且具有很好的扩展性。

关键词:危险废物; 在线监控; HBase

中图分类号: X830.3

文献标志码: B

文章编号: 1674-6732(2016)05-0064-04

The Online Condition Monitoring of Hazardous Waste Disposal Enterprise

CHEN Yuan¹, XU Jie¹, MAO Jia-ming²

(1. Jiangsu Ecological Environmental Monitoring Center, Nanjing, Jiangsu 210036, China; 2. Jiangsu Research Center of Environmental Information Engineering, Suzhou, Jiangsu 215021, China)

Abstract: This paper described the characteristics of HBase database for online condition monitoring system and implementation process of waste disposal enterprises, and a systematic test was performed, which showed that the system had a high throughput and good expansibility.

Key words: Hazardous waste; Online monitoring; HBase

危险废物因其对环境、人群会造成极大危害, 因此对其的无害化处理是一项重要的环保工作。对危废处理企业的工况在线监控, 实现了环保由“点末端监控”向“全过程”监控的转变, 是提高环境监管准确性和可信度的一项重要技术手段。

危废处置工况监控信息的特点是时序性强、采集密度高(≤ 1 s)、高并发性、高延展性。工况在线监控系统须在省级或更高层面建立一套独立于企业现场端承建厂商的统一监控平台, 须满足大量企业、大量点位数据的同时接入^[1], 传统利用实时数据库技术建设成本极高, 且实施难度大, 而如果利用关系型数据库存储实时数据, 则会导致性能效率低下、系统运转不畅。为了真实反映工况现场的情况, 提供工况监控数据的高效稳定传输、检索、分析等服务, 采用开源的 Hadoop 架构及 HBase 数据库^[2], 在廉价的机器上对海量数据进行分布式存储和管理^[3], 可以获得极高的性价比。

见图 1。

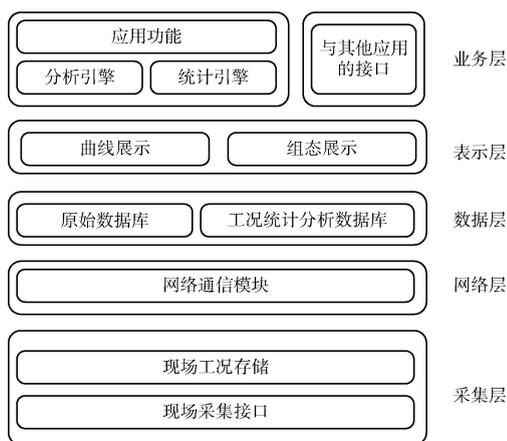


图 1 系统架构

(1) 采集层: 采集层位于在线监控前端, 由相关硬件和软件两部分组成, 主要负责在线监控数据的采集、存储和转发、补发、重发。危废处置设施采

收稿日期: 2016-05-25; 修订日期: 2016-07-04

作者简介: 陈媛(1975—), 女, 高级工程师, 硕士, 从事环境统计、环境信息管理、环境质量综合分析工作。

1 系统设计

1.1 总体框架设计

系统采用多层 B/S 应用结构体系, 系统架构

集单元主要负责采集各类控制系统中的相关参数,并通过隔离器、采集交换机存储到现场工况数据存储服务器中;

(2) 网络层:网络层位于工况前端(危废处置设施)与监控中心之间,由网络通信模块(包括 VPN 环保专网、3G 无线网络、数据采集传输软件)组成。主要负责工况前端所有过程数据、监测数据发送和监控中心的数据接收;

(3) 数据层:数据层位于监控中心,由工况监控原始数据库和统计分析数据库组成,主要负责工况数据的统一存储。考虑到未来几年监控企业逐步增加,监测数据的持续增长,存储实时采集数据的数据库需要有几个技术支撑:能够存储大量的数据;需要很高的写吞吐量;在大规模数据集中有很好性能的随机访问;能够动态扩展系统容量。鉴于这些技术要求,原始数据库采用 HBase,通过分布式的采集服务汇总各危废处置设施的工况数据,随着数据量的增加,HBase 可以自动水平切分扩展,它与 Hadoop^[4]的无缝集成保障了其数据可靠性和海量数据分析的高性能^[5]。根据采集数据的结构特征,通过时间键值 RowKey 构建索引,保证系统的高并发和高可扩展性。统计分析数据库采用关系型数据库 MS SQL Server,由分析统计平台对原始数据进行加工、分析及统计产生,提供给应用模块展示使用。综合 NoSQL 数据库与关系型数据库的优劣势,根据环境监测数据特点,构建适合于工况监控的数据存储模型;

(4) 表示层:表示层位于监控中心,采用曲线、组态等形式对工况数据进行各种展示,支持自动报警;

(5) 业务层:业务层位于监控中心,为最终用户提供一系列的统计分析功能,由应用功能模块和与其他应用的接口模块两部分组成,负责完成对工况数据的综合应用。

1.2 系统数据库设计

1.2.1 系统存储结构设计

系统存储结构设计见图 2。

系统将按照分布式多级数据库的方式设计数据的存储结构^[6],分为 3 个层次:前端工况数据库、中心工况原始数据库、中心统计分析数据库。

(1) 前端工况数据库:作用是在危废处置单位前端将全厂的工况数据作汇总和转发;

(2) 中心工况原始数据库:主要存储、汇总前端传来的工况原始数据,采用 HBase 分布式数据库;

(3) 中心统计分析数据库:主要是存储统计分析数据,包括各参数的总量统计、分析结果的统计等,同时也担当工况应用模块其他功能业务数据的存储,采用 MS SQL Server 关系型数据库。

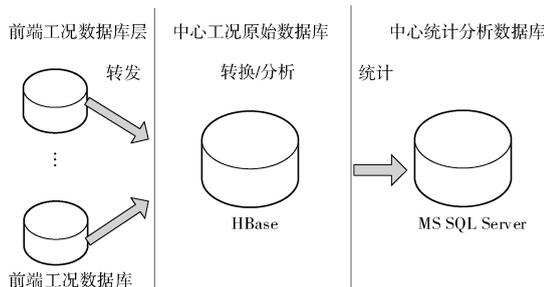


图 2 系统存储结构设计

2 关键技术实现

2.1 数据传输实现

在网络层中,通过 MetaQ 消息中间件^[7],以异步通信方式,实现工况监测数据断续传输,确保数据传输的完整和稳定。MetaQ 整体结构见图 3。其中:(1) MetaQ 使用了文件内存映射特性,免去每次都通过 I/O 去物理硬盘读取文件,所以效率上有很大的提升;(2) 支持 HA 复制,包括异步复制和同步复制,保证消息的可靠性^[8-9];(3) 提供事务支持,包括本地事务和 XA 分布式事务;(4) 利用 MetaQ 的这些技术特性,系统将每个接入企业的监测数据作为一个消息主题,同时增加对应的消费服务,保证了监测数据的采集效率,数据入库的时序性、可靠性,通过动态增加集群节点和对应的消费服务提高事务吞吐能力。

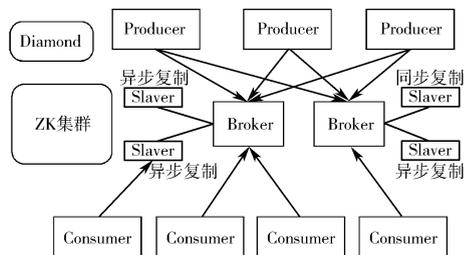


图 3 Meta Q 整体结构

2.2 分布式数据库实现

按照工况在线监控业务的实际需求,HBase 表结构设计包含 3 张主表^[10],分别为基本信息表、五分钟信息表及错误信息表。(1) 基本信息表:包含

工况数据接收数据,以“公司代码+时间+产线+点位”作为 RowKey;(2) 五分钟信息表:包含工况数据接收的 5 min 数据平均值,为了便于系统的查询,以“公司代码+时间+产线+点位”作为 RowKey;(3) 错误数据表:工况接收的异常和错误数据都放在表中,为了便于系统查询,以“公司代码+日期+产线+点位”作为 RowKey。

3 系统功能实现

3.1 实时监控

以工况原始数据库中的数据为主要数据源,以组态画面为主要展示方式,对前端采集到的各类工艺参数、状态数据进行显示和报警,实现对危险固废集中焚烧处置设施运行状况、监控设备运转情况、网络连通情况的实时和视频监控。

(1) 组态配置:提供组态的设置配置工具,满足各企业在工艺流程、位置布局等方面的差异化需求。提供数字量设置、开关量设置、流程线设置等功能;

(2) 监测信息实时监控:建立完备的监控数据显示界面,对各自动监测站点采集到的实时监控数据和设备运行状态实时在线显示,可以趋势曲线等形式实时显示连续自动采集的数据,可查询指定的任意时间段的数据,包括当日的实时数据、日数据、月数据和往日历史数据。当监测值超过报警阈值时,系统按指定的报警颜色进行显示;

(3) 系统运行状况监控:对系统的数据传输情况、网络通信状态等进行跟踪监控;

(4) 实时报警:基于采集到的实时数据,根据设定的报警规则,对监测数据和设备运行状态提供报警监控功能。分为报警设置、自动报警、报警信息查询三部分功能;

(5) 视频监控:包括实时视频查看和视频录像查看两个功能。

3.2 查询分析

以工况原始数据库中的数据为主要数据源,以图表、报表为主要展示方式,对工况数据进行汇总、统计、分析,并对工况数据中的关联关系进行分析。包含以下功能模块:

(1) 数据查询:以数据表格、曲线的方式显示各类监测统计数据,查询各参数的小时均值、日均值、月均值、年均值等数据,支持多参数同时查询显示;

(2) 数据统计:对指定时间段内处置设施的各

类数据进行多种形式的汇总统计,如年度/月度设施停运次数及停运时间统计、焚烧温度分布情况统计、污染物排放量统计、电力消耗统计等,支持同比、环比等分析,以反映其变化趋势和规律;

(3) 参数关联分析:根据设定的规则,对企业工况数据之间的关联性关系进行分析,调取实时/历史工况数据,分析这些关联参数变化是否一致,当比例参数超出正常范围时,系统将记录下起始时间,列出可能的原因;

(4) 多企业对比分析:根据选定的范围(类型、区域、规模等条件),对范围内的多个企业的同类参数进行比较,计算其平均水平及每个企业与平均值的差异程度,便于管理人员发现异常情况;

(5) 报表生成:利用 Report Services 报表组件,基于中心工况应用数据库,自动生成各类报表和报告,在系统中实现报表查询,也可根据设定的规则将生成的报表和报告向指定人员自动发送。

3.3 综合分析

以工况统计分析数据库的数据为主要数据源,结合其他相关业务数据(如污染源自动监控数据、危废转移和处置统计数据、建设项目审批数据等),对危废处置设施的整体状况作综合分析,对不同业务系统的数据关联分析。

(1) 危废处置企业综合查询:以“一厂一档”的形式,对危废处置企业的各类信息进行综合查询,包括企业基本信息、经营许可证信息、危废处置分类数量信息、项目审批信息、污染物排放监控信息、日常监管信息等;

(2) 工况监控数据和污染源自动监控数据对比分析:将同一危废处置企业同时段的工况监控数据和污染源自动监控数据进行对比,判断其是否一致,自动找出超过允许偏差范围的企业数据,提醒管理人员及时处理;

(3) 工况监控数据和危废处置台账信息对比分析:根据危废处置台账中的危废类别及处置数量信息,与工况监控数据进行比较,判断不同类别的危险废物是否按规定的工艺进行处置、炉温是否达到规定的焚烧温度,焚烧时间是否足够等,系统将自动划定存在违规嫌疑的企业,由管理人员进行确认;

(4) GIS 展示:利用地理信息系统,提供图形化展示功能,对在线监控数据进行动态显示,包括危废处置单位信息、实时监控数据、报警信息,生成监测数据统计专题图。

4 系统运行测试

4.1 数据传输性能测试

对系统数据传输性能进行单线程和多线程测试,结果见表 1、表 2。可见多线程模式下数据传输性能有较大幅度提高,提高幅度在 60% 以上。

表 1 单线程测试

数据条数	数据包格式	接收耗时/ms
数据量:30 000 (批量单条)	标准格式: (type + 10 个字段)	3 154 ~ 3 360
数据量:20 000 (批量单条)	标准格式: (type + 10 个字段)	2 050 ~ 2 154
数据量:15 000 (批量单条)	标准格式: (type + 10 个字段)	1 590 ~ 1 700
数据量:10 000 (批量单条)	标准格式: (type + 10 个字段)	1 050 ~ 1 090
数据量:5 000 (批量单条)	标准格式: (type + 10 个字段)	530 ~ 580
数据量:1 000 (批量单条)	标准格式: (type + 10 个字段)	120 ~ 180
	1 000 个	

表 2 多进程测试(3 个消费进程)

数据条数	数据包格式	接收耗时/ms
数据量:30 000 (批量单条)	标准格式: (type + 10 个字段)	800 ~ 910
数据量:20 000 (批量单条)	标准格式: (type + 10 个字段)	500 ~ 730
数据量:15 000 (批量单条)	标准格式: (type + 10 个字段)	265 ~ 460
数据量:10 000 (批量单条)	标准格式: (type + 10 个字段)	150 ~ 230
数据量:5 000 (批量单条)	标准格式: (type + 10 个字段)	80 ~ 120
数据量:1 000 (批量单条)	标准格式: (type + 10 个字段)	20 ~ 100
	1 000 个	

4.2 数据库性能测试

为测试实时工况数据的数据导入及数据查询性

能,对数据总记录数 10 亿条,单行记录 > 1Kb 的情况进行了导入测试及数据查询的性能测试,性能测试的结果见表 3。该实验结果表明,现提出的基于 HBase 的数据存储模型具有高效的存储效率和查询效率,满足实际需求对数据存储和访问的性能要求。

表 3 性能测试结果

操作类型	总记录数	线程	响应时间/ms
导入(多线程并发)	1 000 000 000	50	981 000
		100	968 000
单行记录查询	1 000 000 000		12.95
多条件多行查询	1 000 000 000	1 700	45.6
		40 800	51
		1 224 000	52.2
		7 344 000	54.25

5 结语

采用 HBase 数据库用于工况在线监控系统具有低成本、高性能的优势,测试结果表明,该系统有很高的吞吐率,并且具有很好的扩展性。在未来的工作中,将不断丰富该系统的功能,优化集群节点的数目,对收集到的监控数据作进一步的挖掘分析,为环境管理服务。

[参考文献]

- [1] 淘宝 GitHub. MetaQ 原理与应用[EB/OL]. (2013-1-13) [2016-6-11]. <http://taobao.github.io/metaq/document/design/design.htm>.
- [2] 郭艳霞,颜军.海量数据存储模式的研究[J].计算机与数字工程,2008,36(11):162-165.
- [3] 陈庆奎,周丽珍.基于 HBase 的大规模无线传感网络数据存储系统[J].计算机应用,2012,21(7):1920-1923,1977.
- [4] 刘鹏.实战 Hadoop[M].北京:电子工业出版社,2011.
- [5] TOM W. Hadoop 权威指南[M].2 版.周敏奇,王晓玲,金澈清,等,译.北京:清华大学出版社,2011.
- [6] Oracle Corporation. Oracle real application clusters[EB/OL]. [2013-12-11].
- [7] DIMIDUK N, KHURANA A. HBase 实战[M].1 版. Manning Publications, 2012.
- [8] GEORGE C, JEAN D, TIM K. 分布式系统概念与设计[M].金蓓弘,译.北京:机械工业出版社,2004.
- [9] 庄晓丹.消息的可靠性.顺序和重复[EB/OL].(2013-4-13)[2016-6-11]. <https://github.com/killme2008/Metamorphosis/wiki>.
- [10] CSDN 博客. MetaQ 初探 [EB/OL]. (2014-8-17) [2016-6-11]. <http://m.blog.csdn.net/blog/wtycoon1988/38868407>.

栏目编辑 周立平