

doi: 10.11823/j.issn.1674-5795.2016.05.06

电子测量仪器数据服务平台研究

赵卫芳, 王晓飞, 何昭, 张亦弛
(北京信息科技大学光电学院, 北京 100192)

摘要: 简述了搭建电子测量仪器数据服务平台的背景意义及电子测量仪器数据服务平台现状和挑战; 描述了电子测量仪器数据服务平台的功能、架构。文章在阐述平台功能和结构的基础上, 重点阐述了数据服务平台的搭建过程, 包括 Hadoop 大数据架构搭建以及 Eclipse 软件开发平台搭建, 最后介绍了网页化的人机界面和微信公众平台。

关键词: 电子测量仪器数据; 服务平台; Hadoop; Eclipse

中图分类号: TB97

文献标识码: A

文章编号: 1674-5795(2016)05-0023-03

Research on Data Service Platform of Electronic Measuring Instrument

ZHAO Weifang, WANG Xiaofei, HE Zhao, ZHANG Yichi

(Beijing University of Information Science and Technology, Photovoltaics, Beijing 100192, China)

Abstract: This paper summarizes the background significance of setting up the electronic measuring instrument data service platform and describes the present situation and challenge of the electronic measuring instrument data service platform. And this paper introduces the function and structure of the data service platform of the electronic measuring instrument. On the basis of explaining the function and structure of the platform, this paper focuses on the building process of data service platform, including the Hadoop big data architecture and Eclipse software development platform. Finally, the paper gives a brief introduction of the web page of the human-computer interaction interface and WeChat public platform.

Key words: electronic measurement instrument data; service platform; Hadoop; Eclipse

0 引言

随着信息时代的发展, 电子测量仪器在我国科技进步、经济建设以及国防工程等领域发挥着重要的作用, 各类电子产品已经成为通信、医疗、娱乐等领域不可或缺的元素。对于厂家而言, 无论是研发环节还是生产环节, 都需要采用大量的测试设备(如信号源、接收机、示波器、网络分析仪、综测仪)来保证产品质量。

通过对京津冀、长三角、珠三角等电子产业发达地区大型生产制造企业的调研, 汇总的数据量可达 PB 级别。数据类型涵盖文本、数字、图表、检定过程的视频等, 既包括结构化数据, 又包括非结构化数据。传统的检定校准平台受限于过去的数据存储以及信息处理能力, 只能生成检定证书。大量有价值的测试数

据无法为后续的检定校准、产品分析、市场评估、行业监管等提供参考, 造成了信息资源的巨大浪费。本文基于日常检定校准的测试数据, 建立电子测量仪器大数据分析系统, 形成了电子测量仪器数据服务平台。从而充分保留和利用检定校准环节的大量信息, 建立不断更新的“电子测量仪器”大数据库, 为不同的潜在服务对象提供更有参考价值和指导意义的分析报告。

1 电子测量仪器数据服务平台的现状与挑战

1.1 平台现状

目前市面上电子测量仪器服务平台多种多样, 常见的有 IT168, RS, Keysight 等服务平台, 这些平台的服务涉及到简单的产品介绍、产品对比及采购意见, 类似于一般的电子商务平台, 其中 RS 也提供检定校准服务。然而这些服务平台多数是面向顾客进行产品销售, 并没有对检测到的数据进行大数据分析挖掘, 不能给出更深层次的有价值信息。

近年来, 随着大数据产业的兴起, 越来越多的数

收稿日期: 2016-06-06

作者简介: 赵卫芳(1989-), 女, 硕士研究生, 从事电子测量仪器大数据研究。

据服务平台应运而生，它们分析着各行各业的数据，随着电子测量仪器的应用越来越广泛，随之产生的数据量也越来越庞大，待挖掘的有价值的信息也越来越多，然而，受限于数据服务平台的不专业性，经过数据分析和挖掘而得到的信息比较表面，并不能有很好的使用价值，因此开发一套单独针对电子测量仪器数据的数据分析平台很有必要。

1.2 平台面临的挑战

电子测量仪器数据服务平台不但可以借鉴一般的数据服务平台的架构及分析角度，平台自身的功能要求及数据挖掘方法还有它独特的方面。经过前期调研，电子测量仪器数据依然存在很多问题，给搭建电子测量仪器数据服务平台带来了许多挑战。电子测量仪器数据主要问题如下：

1) 数据杂乱，价值低，可机读比例低

电子测量仪器应用于各行各业，由于没有统一的数据存储模板，造成数据存储格式多种多样，应用价值低，不能统一进行机读。

2) 数据多为静态数据

电子测量仪器数据一般采用静态记录，缺乏实时更新，给数据分析挖掘造成延迟。

3) 缺乏便捷的数据获取渠道

电子测量仪器应用十分广泛，且地域分布不集中，又由于数据存储不统一，难以形成便捷的数据采集渠道。

4) 缺乏高质量的数据应用

一般的数据服务平台有统一的数据分析方法，但是分析结果比较表面，达不到深度挖掘，没有实现高质量的数据应用。

5) 缺乏便捷、及时、有效、公开的互动交流

一般的数据服务平台比较机械化，只能满足简单的数据分析挖掘要求，并不能提供便捷、及时、有效、公开的互动交流。

2 基于大数据分析的电子测量仪器数据服务平台的功能、架构及构建方法

2.1 平台功能

搭建电子测量仪器数据服务平台，是想通过大数据系统的整理分析，为客户需求提供有力的数据支撑，使得用户在选择相应业务，选购相关产品时，有精确的数据作为保证。

2.1.1 数据查询及产品分析

电子测量仪器数据服务平台在接收大量测试数据

后，可以进行相应的数据查询及跟踪；通过对同型号产品测试数据进行分析，给出同型号产品性能分析；通过对同类型产品测试数据进行分析，给出同类型产品性能对比，多方位、多角度对产品进行精准分析。通过对产品的精确度、稳定性等进行数据定量分析，进而给用户提出合理的推荐意见。

2.1.2 检定校准建议

电子测量仪器数据服务平台通过数据整理、分析、挖掘，能给送检用户提供准确的检定周期、需要检定的仪器特性以及需要检测特性的检测点等。以往的送检用户针对这些检定项目都是凭借经验或者检定机构的建议，现在通过数据服务平台，能为这些检测项目提供准确的数据支持，更有利于仪器的检测及使用。

2.1.3 市场评估

电子测量仪器数据服务平台会接收不同地域、不同行业的电子测量仪器数据，通过对数据的分析挖掘，能得出电子测量仪器使用的行业划分、地域分布以及厂家占比、使用率等，进而可以为计划进军电子测量仪器行业的厂家提供数字化的市场分析，给出合理的选址建议、服务对象等信息。

2.2 平台架构

电子测量仪器数据服务平台以底层的数据采集为基础，数据采集通过制定统一数据采集模板来进行；采集到的数据存储到分布式文件系统 HDFS (Hadoop Distributed File System)；数据服务平台的核心是数据计算，采用 MapReduce 计算框架，并应用 Apache Mahout 相关算法进行数据分析挖掘；处理好的数据存储到 HBase 数据库，便于数据调用；然后通过 Eclipse 软件开发平台开发动态网页，实现数据可视化，整体架构由 Hadoop 大数据框架进行集群配置管理。电子测量仪器数据服务平台整体架构如图 1 所示。

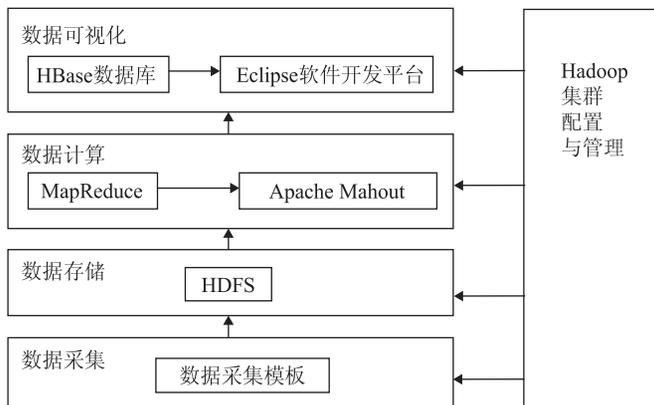


图 1 电子测量仪器数据服务平台整体架构

2.3 平台构建方法

2.3.1 平台整体架构搭建

当前主流的大数据框架包括 Storm, Spark, HPCCC 和 Hadoop 等^[1], 其中 Hadoop 以其分布式计算的独特优势被广泛采用, 因此, 本电子测量仪器数据服务平台选用 Hadoop 大数据框架。Hadoop 大数据框架的核心是分布式文件系统 HDFS 和 MapReduce; HDFS 是大数据底层存储数据的分布式文件系统; MapReduce 是 Hadoop 的编程模型, 对大量数据进行分析计算。HBase 数据库是 Hadoop 大数据项目下的子项目, 能很好地存储非结构化数据如视频类型数据, 满足电子测量仪器数据服务平台数据存储功能^[2]。制定合理的数据采集模板进行数据收集, 将收集的数据导入到 Hadoop 的分布式文件系统 HDFS 中, 利用 MapReduce 进行简单处理, 再通过 Apache Mahout 相关算法进行深度数据挖掘, 实现相应的数据服务功能^[3]。项目使用 Eclipse 软件开发平台进行程序开发, 配合使用 TomCat 服务器、Servlet 软件进行 JSP 动态网页制作, 进而满足客户与数据的动态交互。

2.3.2 数据采集

在数据服务平台搭建的同时, 科研人员对京津冀、长三角、珠三角等电子产业发达地区的大型生产制造企业进行了数据采集工作, 经过汇总多家科研机构及生产厂家普遍关心的问题及希望得到的数据深层数据信息, 比如行业划分、地域分布、市场评估等, 确定了电子测量仪器数据的采集模板及可挖掘信息的数据源。以信号发生器为例, 初期数据采集模板罗列了信号发生器检定结果的各种数据, 采集模板分为几个独立的工作表, 包括基本信息、频率特性、绝对电平、相对电平、波形、模拟调制、数字调制、参数信息等, 各个工作表又包含相关内容。项目合作单位按照采集模板要求填写数据, 来进行数据采集。

2.3.3 数据存储

数据存储是电子测量仪器数据服务平台的基础, 经调研, 电子测量仪器数据量大, 并且涵盖多种数据类型, 既包括数字、文本等结构化数据, 又包括音频、视频等非结构化数据, 因此选取合适的存储设备至关重要。传统的数据库如 SQL 数据库, 只能存储结构化数据, 而目前兴起的 HBase 数据库既能存储结构化数据, 又能存储非结构化数据, 并且能以分布式的方式进行存储, 满足高效和安全的要求。

2.3.4 数据计算

数据收集后, 通过存储设备对数据进行简单排序

整理, 可以实现数据档案管理和追踪; 通过大数据架构下的计算模型可以对存储设备中同型号产品及同类别产品数据进行分析处理, 实现同型号产品及同类型产品性能对比, 给出精确的产品分析; 至于更深层次的市场评估报告, 则需要使用复杂的机器算法对数据进行深度分析挖掘, 得出有价值的信息。

2.3.5 数据可视化

电子测量仪器数据服务平台最终要服务于客户, 数据经过分析挖掘后得出的有价值的信息要通过网页直观的呈现给用户, 网页要做到简单明了, 功能模块便于应用。

3 网页化的人机界面及微信公众平台

3.1 网页化的人机界面

网页是电子测量仪器数据服务平台的外在表现, 网页要有数据入口功能, 能为后台工作人员以及用户提供输入接口; 还要有业务需求选择功能, 包括数据档案管理与追踪、同型号产品及同类型产品性能对比、市场评估报告以及其他企事业单位普遍关心的计量、校准、测试问题。网页设计要简洁、方便、人性化, 更多的功能接口有待开发。

通过 Eclipse 软件开发平台的连接功能, 在连接 Hadoop 大数据框架的同时^[4], 构建动态网页, 将处理后的数据显示在网页上, 使用户通过网页上相应功能模块获取相应的数据^[5], 以满足不同的需求, 比如有些客户只需要查询历史数据; 而有些客户需要对比各个型号的产品, 进而选取合适的型号进行购买。

3.2 微信公众平台

在搭建电子测量仪器数据服务平台的同时, 建立了微信公众号: 一方面是宣传电子测量仪器数据服务平台以及介绍其工作进展; 另一方面是积极汲取各生产厂商及科研机构用户的合理意见, 尽可能完善数据服务平台功能, 使其更人性化, 便于用户使用。

4 结论

大数据框架搭建完成后, 通过数据导入导出到 HDFS 分布式文件系统对数据进行简单排序整理, 进而实现数据档案管理和追踪; 通过 MapReduce 对 HDFS 分布式文件系统中的同型号产品以及同类型产品数据进行分析计算处理, 进而实现同型号产品性能对比、

(下转第 28 页)

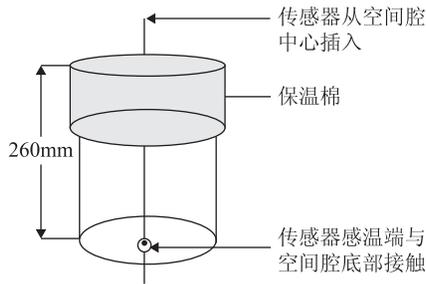


图4 金属浴试验箱测试示意图(侧面、单腔)

$$\Delta T_u = T_{\max} - T_{\min} \quad (6)$$

式中： ΔT_u 为空间均匀度， $^{\circ}\text{C}$ ； T_{\max} 为四个测试腔实测温度平均值最大值， $^{\circ}\text{C}$ ； T_{\min} 为四个测试腔实测温度平均值最小值， $^{\circ}\text{C}$ 。

实测值减去标准值，可获得金属浴试验箱各测试腔的示值误差，如表1所示。

表1 金属浴试验箱测试腔的示值误差 $^{\circ}\text{C}$

标准值	测试腔			
	①	②	③	④
50	-0.2	-0.2	-0.2	-0.2
100	-0.1	-0.2	-0.2	-0.2
150	-0.2	-0.2	-0.2	-0.3
200	-0.2	-0.4	-0.3	-0.1
220	-0.1	-0.3	-0.2	0.0

根据上述测试方法^[5]，对金属浴试验箱温度波动度及均匀性等指标进行测试，测试结果如表2所示。

表2 金属浴试验箱波动度及均匀性试验结果

设定温度/ $^{\circ}\text{C}$	波动度/ $^{\circ}\text{C} \cdot 30 \text{ min}^{-1}$	均匀性/ $^{\circ}\text{C}$
50	0.0	0.0
100	0.0	0.1
150	0.1	0.2
200	0.1	0.4
220	0.1	0.4

3 结论

通过分析可知，该金属浴试验箱测试腔内示值误差较小，温度波动度不超过 $0.2^{\circ}\text{C} \cdot 30\text{min}^{-1}$ ，温度均匀性不超过 2°C ，性能较好。随着温度的升高，腔体温度均匀性和波动度均有所下降，在 200，220 $^{\circ}\text{C}$ 时，由于周围环境温度与腔体温度相差较大，导致空间波动度和均匀性比 50，100，150 $^{\circ}\text{C}$ 时差，但总体温度波动度和均匀度符合设计要求。

参考文献

- [1] 杨世铭, 陶文铨. 传热学[M]. 北京: 高等教育出版社, 2006.
- [2] 国家技术监督局. JJG 364-1994 表面温度计检定规程[S]. 北京: 中国计量出版社, 1994.
- [3] 国防科工委科技与质量司. 热学计量[M]. 北京: 原子能出版社, 2002.
- [4] 国家质量监督检验检疫总局. JJF 1257-2010 干体式温度校准器校准方法[S]. 北京: 中国计量出版社, 2010.
- [5] 国家质量技术监督局. JJF 1030-2010 恒温槽技术性能测试规范[S]. 北京: 中国计量出版社, 2010.

(上接第25页)

同类型产品性能对比; 针对更深层次的市场评估功能, 平台使用 Apache Mahout 提供的相应算法来实现数据挖掘。就目前而言, 收集到的数据量较少, 无法达到深度挖掘的数据量要求, 随着数据量的增加, Apache Mahout 才能发挥数据价值挖掘的功效。

电子测量仪器数据服务平台构建选用 Hadoop 大数据框架, 搭建过程中对版本兼容性有严格要求。该软件平台还采用了 Hive, Zookeeper 等子项目辅助, 整体框架简洁高效, 通过 Eclipse 软件能够方便快捷地扩展功能。该系统目前仍在调试中, 后续还需要更多的研究才能确保平台更加稳定, 功能更加完整。

参考文献

- [1] 韩晶. 大数据服务若干关键技术研究[D]. 北京: 北京邮电大学, 2013.
- [2] 陆嘉恒. Hadoop 实战. [M]. 2 版. 北京: 机械工业出版社, 2014.
- [3] 费珊珊. 基于云计算 Hadoop 平台的数据挖掘研究[D]. 北京: 北京邮电大学, 2014.
- [4] 潘梦云. 基于 Hadoop 的数据处理系统的设计[D]. 邯郸: 河北工程大学, 2014.
- [5] 黄凯. 基于 Eclipse 体系的构件开发管理平台的设计实现[J]. 科学技术与工程, 2005, 5(14): 994-997.