

浅谈液体火箭发动机试验数据解析入库技术

董 冬, 乔江晖, 朱成亮, 刘 晓, 白文义
(西安航天动力试验技术研究所, 陕西 西安 710100)

摘 要: 液体火箭发动机试验数据解析入库技术通过分析试验数据的结构特点, 选取适合异构试验数据的存储方式, 采用关系数据库存储结构化数据和文件数据库存储非结构化数据相结合的方式数据进行存储。基于数据入库通用模型建立试验数据解析入库模型, 包括数据分类与结构标准化, 数据解析或者附件上传, 导入关系数据库或存入文件数据库。模型建立完成后根据模型进行液体火箭发动机试验数据解析, 解析完成后的数据存入文件数据库或关系数据库中。液体火箭发动机试验数据解析入库关键技术, 尤其是试验数据解析入库模型可用于其它领域试验数据管理, 为相应的试验数据解析入库模型建立提供设计参考。

关键词: 试验数据; 数据解析; 数据入库; 关键技术

中图分类号: V433.9-34 **文献标识码:** A **文章编号:** 1672-9374 (2015) 04-0105-04

Discussion of analytic technique for LRE test data storage

DONG Dong, QIAO Jianghui, ZHU Chengliang, LIU Xiao, BAI Wenyi
(Xi'an Aerospace Propulsion Test Technique Institute, Xi'an 710100, China)

Abstract: The analytic technique of LRE test data storage was achieved through the analysis of structural characteristics of the test data, selection of the storage method suitable for heterogeneous test data, and combination of the relation database storage mode of structured data and file system storage mode of unstructured data. A test data analytic storage model was established based on the general data storage model analysis, including the data classification, data structure standardization, data analysis or annex upload, data imported in storage or stored in file system. The LRE test data analysis is completed according to the established data model, and then the analyzed data is stored in file database or in relational database. The LRE test data analytic storage is a key technology. The test data analytic storage model can be used in test data management in other areas, and can provide a reference model for the analysis of the corresponding test data storage.

Keywords: test data; data parsing; data warehousing; key technology

收稿日期: 2015-01-13; 修回日期: 2015-05-12

作者简介: 董冬 (1981—), 男, 工程师, 研究领域为发动机试验技术

0 引言

由于国外在信息化、云计算、海量存储及互联网络技术的领先发展,在试验数据解析入库领域已成熟应用上述技术,实现海量试验数据分布式管理,试验数据快速检索,大数据实时传输,而国内起步较晚,在该领域水平较低。近年来,国内液体火箭发动机研试人员也将试验数据如何深入利用、挖掘这些试验数据作为重要的研究方向,并建立相关试验数据管理平台。而这些数据如何存储和解析入库是建立试验数据管理平台需首要解决的问题^[1]。

1 试验数据特性

数据有 2 类:一类有统一的结构,可以用数字或文字来描述,具有类似的层次或网络结构,称之为结构化数据^[2],结构化数据包括完全结构化数据和半结构化数据;另一类信息则无法用数字或者统一的结构表示,内容结构不固定无法解析的数据,即为非结构化数据。既包含结构化数据,又包括非结构化的数据统称为异构数据。发动机试验数据符合异构数据特点,可称为异构试验数据^[3]。发动机试验数据的存储特点与传统意义的数据相比,有其自身的特点。从存储的角度来考虑,发动机试验数据具有存储容量大、媒介类型复杂、增长快速的特点。

2 数据存储方式选取

2.1 结构化数据

目前常用的结构化数据存放在数据库中^[4],以字段形式存放在表结构组成的数据库某个表单记录中。

2.2 非结构化数据

- 1) 直接存储在结构化数据库的 BLOB 字段中;
- 2) 以 FTP 上传的方式保存到文件服务器中;
- 3) 通过文件系统直接存储在文件服务器中^[5]。

2.3 液体火箭发动机试验数据采用方法

液体火箭发动机试验数据属于结构化和非结构化数据结合的构成形式,因此采用结构化数据

的关系数据库存储方式,以及非结构化数据的文件系统存储方式相结合的方式进行存储。这种存储方式既集成了结构化数据便捷存储的优点,同时又解决了非结构化数据文件大、结构复杂及不易解析查询的问题。

3 入库模型设计

3.1 入库通用模型

数据解析入库完成将各种海量异构试验数据导入数据库,它是整个数据管理平台建立的关键环节,根据数据分类建立解析入库模型是进行数据导入设计的首要前提,入库模型对于结构化数据解析的存储量占整个数据比重越高,数据利用率越高,入库模型设计越难。数据入库通用模型的工作原理是:将数据源进行分类,按照数据结构化程度分为结构化数据和非结构化数据,针对结构化数据进行数据解析后导入数据库,非结构化数据经附件上传后存入文件数据库^[6]。

数据入库通用模型适用于比较简单的数据库构成,对于大数据,尤其对于诸如振动等高频采样率数据只能通过原始文件形式存入,不能解析文件内容,无法进行数据分析;并且对于半结构化数据也无法解析,导致分析数据只能打开文件进行查看,无法针对文件内容进行分析,数据利用价值降低。同时这些不能利用的非结构化数据入库占用大量的存储空间,浪费存储资源。

3.2 试验数据入库模型

试验数据专有的数据特性和存储方式的复杂特点,共同决定了仅依靠数据入库通用模型不能达到试验数据入库要求。因此,在上述模型基础上进行改进设计,经数据结构标准化实现对半结构化数据进行结构标准统一,可作为结构化数据数据进行解析;同时大数据块也可通过时域、频域数据存储模块解析成多个 *.tdmdf 文件。这样诸如高采样获得的振动频域大数据就可进行分析比对^[7]。

试验数据入库模型基于数据入库通用模型,包括数据分类与结构标准化、数据解析及入库存储 3 个部分^[8]。其中,数据分类与结构标准化部分,数据源按结构可分为非结构化、半结构化及

结构化数据, 经过结构标准化模块后数据转换成结构化数据和非结构化数据; 数据解析部分完成结构化数据通过附件存储模块进行原始文件上传, 索引信息生成模块进行参数索引信息提取, 特定规则下解析规则模块实现数据 (非时域、频域数据和时域频域数据) 的解析, 以及实现非结构化数据通过附件存储模块进行文件上传, 并通

过索引信息生成模块机型文件索引信息提取; 入库存储部分完成非结构化数据的原始文件存储、文件索引信息存储, 以及结构化数据的参数索引信息存储、结构化数据内容以关系数据库表单形式存储, 大数据文件解析成多个 *.tdmdf 文件并存储在文件数据库中。其中试验数据解析入库模型, 如图 1 所示。

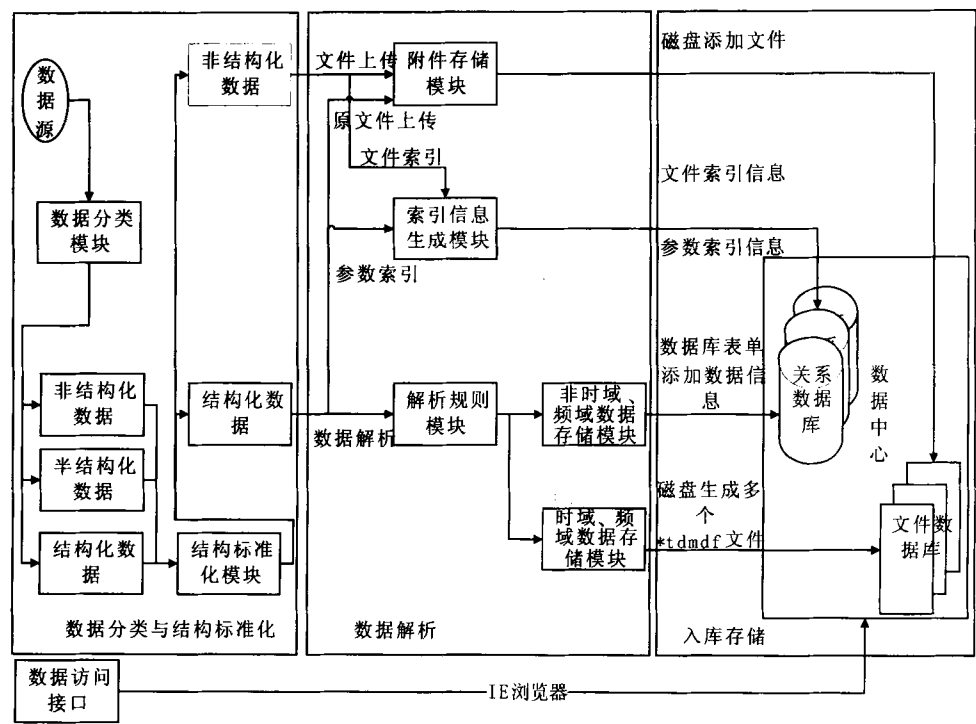


图 1 试验数据解析入库模型

Fig. 1 Model of test data parsing and warehousing

4 入库关键技术

4.1 数据分类与结构标准化技术

试验整个周期 (试验准备、试验及试验结束后工作) 获得数据经统计整理后, 按照业务分类试验数据一般包括缓变数据、速变数据、附件数据、资源数据及试验过程数据。文件类型可分为 *.UFF、*.txt, *.xls, *.doc, *.mdb, *.rar 及 *.jpg 等格式。数据按照文件结构可分为结构化和非结构化数据。结构化数据包括时域数据、频域数据及非时域、频域数据; 非结构化数据包括文档上传数据和原始数据 (图片、视频、附件)。试验数据通过数据结构标准化统一转换模块统一数据格式后, 转换后形成统一的文件结构标准, 用于

数据解析入库。

4.2 基于 XML 技术的数据解析

试验数据复杂的文件结构决定了试验数据解析难度大, 解析规则多样性的特点。专门为 Web 应用程序而设计的 XML (ExtensibleMarkup Language), 作为一种可扩展性标记语言, 适用于不同应用间的数据交换^[9]。XML 由于具有数据描述和数据传送能力, 有很强的开放性, 使其成为不同应用系统之间的数据接口标准^[10]。数据解析通过基于 XML 技术建立不同的解析模板, 实现解析规则描述, 完成数据在模板选择后数据内容对应解析, 与页面程序完成数据交互, 对相应数据进行页面显示, 同时上传至数据库或者解析成多通道文件^[11]。非结构化数据直接以附件形式上传。

针对结构化数据制定相应的解析规则,进行分类解析。结构化数据即需要解析数据进行规则选择,解析规则主要有2大类:一类是时域、频域的参数数据解析规则;另一类是非时域、频域数据解析规则。

结构化数据中的时域、频域数据,主要来源为*.txt等文件类型的格式化可解析文件,具有特定统一的文件结构。时域数据或频域数据均使用统一的XML模板,在进行数据结构统一标准后均可使用XML技术进行数据解析入库。根据文件参数个数生成多个文件名称(原文件名、参数代号及日期),在文件数据库所在磁盘建立一定存储空间,文件数据库建立多个单参数数据文件以文件数据库所在路径上传数据,并为所上传数据分别生成参数总数以时间列和每个参数列构成的*.tdmndf格式文件。

结构化数据中的非时域数据、频域数据,主要来源于以*.doc、*.mdb或*.xls文件类型的表格数据,可解析但无统一格式,每个数据文件建立专有的XML模板,使用XML技术建立与之对应的一种解析规则,该解析具有专有特性。解析完成数据内容提取后写入数据库表空间的对应表单中,创建相应关联表单进行数据存储。

非结构化数据,主要数据来源为原始数据或文档上传数据。包括文档、视频、图片及附件类型数据;它们均有相对应的XML上传模板,不需要进行解析,选择不同的上传模板以附件方式上传存储在文件数据库中,同时在数据库中将文件索引信息添加在数据库表记录中。

4.3 入库存储技术

数据库用于存放异构试验数据,根据试验数据的分散特点建立分布式数据库比较适合,同时数据存储方式决定了数据库的构成形式,即分布式数据库由关系数据库和文件数据库共同构成^[12]。关系数据库用于存储结构化数据及其对应的索引信息,文件数据库用于存储原始数据,附件数据,多通道参数文件数据(格式为*.tdmndf)。系统通过统一接口(即WebService+Http)对数据中心的数据进行访问或对数据中心进行数据存储、数据解析导入、数据分析等操作^[13]。

结构化数据经过数据解析后,将非时域、频域数据数据内容和数据索引信息分别存放在数据库表记录中,同时将原始数据文件以附件信息存放在文件系统中;将时域、频域数据数据解析后多个单参数文件和原始文件(以附件形式)存放在文件系统中,同时将索引信息分别存放在数据库表记录中^[14]。

非结构化数据,包括视频、图片等的原始数据和文档数据,均以附件形式上传至文件数据库,系统在磁盘中开辟一定的存储空间,以附件文件数据形式存储在文件数据库中,进行文件名称的创建和数据写入。同时数据的索引信息将被提取并与试验任务关联。非结构数据同时在数据库中将文件索引信息添加在数据库表记录中。

5 解析入库技术应用

5.1 批量导入存储

为了保证大量的、连续的、长时间的数据准确、快速地入库,同时数据模板类型多样,且每个模板下对应大量数据文件,因此针对同一个模板的数据文件,系统采取批量导入存储技术。数据经过解析后形成多个数据文件,把多个数据文件批量同时传输给服务器端,服务器端及时解析成多个通道参数数据形式,以批量的方式一次性插入到数据库中。这样解决了重复导入同样模板的数据文件问题,提高了工作效率。

5.2 大数据解析入库能力

试验速变高频大数据的管理能力决定了试验数据管理能力水平,因为速变高频大数据具有占存储空间大,不易解析,解析耗费内存等特点^[15]。因此,一般针对试验速变高频数据的入库方式是仅存放在文件数据库中,而试验数据管理平台对于速变高频数据进行了解析导入,它通过时域、频域数据存储模块将速变高频数据解析成多个*.tdmndf文件,每个*.tdmndf文件为单通道参数文件,存放在文件数据库中,可用于数据管理平台参数显示比对,这样在选取该数据文件的某个参数时就不再调用源数据文件,仅调用后续的单通道文件,数据显示效率提高。

(下转第112页)

计了一套卫星轨控发动机试验用质量流量计现场校验系统,使用真实试验推进剂,并采用发动机试验时的增压压力对推进剂管路和流量计进行增压,通过检验试验得到以下结论:

1) 这种校验方式可以应用于流量计的标定中,采用真实介质校验结论可信度更高。

2) 通过校验可以在符合精度要求的流量计中挑选精度较高的流量计作为试验用主流量计,其他流量计作为备份使用。

参考文献:

- [1] 周红玲,姜文龙,刘昌国.国内外卫星用液体远地点发动机发展综述[J].火箭推进,2011,37(5): 1-8.
- [2] 张惠军.科氏力质量流量计在凝胶推进剂火箭发动机试车中的应用[J].火箭推进,2004,30(6): 55-60.
- [3] 吴波. Promass83A 质量流量计在发动机试验中的应用

[J]. 火箭推进, 2005, 31(6): 57-60.

- [4] 罗凡,孙玉声.不同介质对质量流量计的校准误差影响[J].中国测试技术,2007,33(6): 52-54.
- [5] 纪纲.流量测量仪表应用技巧[M].北京:化学工业出版社,2005.
- [6] 李长武,陈平,袁明,等.液体流量仪表在线校准方法及趋势[J].计量技术,2009,(4): 58-61.
- [7] 周士虎,张永华.流量计校验装置在航天发射场加注系统中的应用[J].石油工业技术监督,2001,17(4): 28-33.
- [8] SMITH D M, CARVER D B. Flow calibration of two hypersonic nozzles in the AEDC Heat-H2 high-enthalpy arc-heated wind tunnel, AIAA-1996-0913[R]. USA: AIAA, 1996.
- [9] KIRTLEY D E, ENGELMAN S F, FIFE J M. Analysis of xenon flow calibration techniques for electric thruster testing, AIAA-2002-3817 [R]. USA: AIAA, 2002.
- [10] 钟爱民.质量流量计自动校验控制系统的设计研究[D].上海:上海交通大学,2007.

(编辑:王建喜)

(上接第108页)

6 结束语

通过对液体火箭发动机试验数据解析入库模型的设计以及对解析入库关键技术的研究与应用,解决了试验数据批量入库存储与大数据解析入库的问题,为数据挖掘利用打下坚实的基础,为相关领域的试验数据解析入库提供设计参考。

参考文献:

- [1] 王花梅,罗续业.海上试验场数据管理与应用系统的原型设计[J].海洋技术,2013,6(4): 06-08.
- [2] 邱丽丽,俞烽.异构数据动态交互平台设计与实现[J].计算机应用与软件,2013,12(3): 182-182.
- [3] 张冠军.基于XML异构系统间的数据交换技术[J].现代电子技术,2013,24(2): 45-47.
- [4] 陈硕颖,郑敏信,齐铂金,等.用于电池管理系统的数据存储系统设计[J].单片机与嵌入式系统应用,2014,14(1): 65-68.
- [5] 张恩,张广弟,兰磊.基于MongoDB的海量空间数据存储和并行[J].GEOSPATIAL INFORMATION, 2014, 12

(1): 22-23.

- [6] 韩晶,鄂海红,宋美娜,等.基于主体行为的非结构化数据模型[J].计算机工程与设计,2013,12(3): 04-08.
- [7] 韩翠峰.大数据带给图书馆的影响与挑战[J].图书与情报,2013,6(5): 37-40.
- [8] 林菲,张万军,孙勇.一种分布式非结构化数据副本管理模式[J].计算机工程,2013,12(4): 36-38.
- [9] 辛怀声,王鹏.基于XML的数据解析软件设计与实现[J].电脑知识与技术,2014,12(1): 18-20.
- [10] 刘波,苗彩,王国瓦,等.WITSML数据模式分析及应用[J].录井工程,2014,12(1): 27-29.
- [11] 武艳,谭献海.基于XML的异构数据转换的研究[J].铁路计算机应用,2013,12(10): 4-7.
- [12] 刘斌.基于元数据的海量试验数据管理系统研究与设计[J].微电子学与计算机,2012,12(4): 08-11.
- [13] 薛涛,刁明光,李建存,等.资源环境遥感海量空间数据存储,检索和访问方法[J].国土资源遥感,2013,12(2): 08-11.
- [14] 张蕊,李广云,王力,等.基于HDFS的海量激光点云数据分块存储方法研究[J].测绘通报,2014,12(3): 21-24.
- [15] 陈娜,张金娟,刘智琼,等.基于Hadoop平台的电信大数据入库及查询性能优化研究[J].移动通信,2014,12(7): 12-14.

(编辑:马杰)