

面向火箭数据分析和故障诊断的数据仓库设计

陈卓^{1,2}, 汪灏³, 平佳伟¹, 邓闯¹, 彭炳峰¹, 徐昕^{1,4}

(1. 上海航天电子技术研究所, 上海 201109; 2. 上海交通大学 计算机科学与工程系, 上海 200240;
3. 上海宇航系统工程研究所, 上海 201109; 4. 南京航空航天大学 航天学院, 南京 210016)

摘要: 针对运载火箭测试发射数据分散存储且数据利用不充分的问题, 提出了以数据仓库技术对火箭的测试发射数据进行组织和存储管理, 实现对火箭数据价值的深层次发掘; 根据运载火箭数据的特点, 提出了数据仓库基础层、源数据层、数据交互层、数据架构层、应用层、显示层 6 个层次的框架设计; 通过面向火箭数据分析和故障诊断的数据清洗规则设计, 对不同来源、不同格式的数据进行清洗和加载; 对火箭的数据应用方向进行了探索, 设计了包络分析、相关性分析、故障诊断等数据分析模块; 用某型号火箭测试发射数据进行了验证, 结果表明, 数据仓库方案切实可行, 为后续火箭数据的持续性管理和分析奠定了基础。

关键词: 数据仓库; 运载火箭; 结构化数据; 数据清洗; 故障诊断

Design of Data Warehouse for Rocket Data Analysis and Fault Diagnosis

CHEN Zhuo^{1,2}, WANG Hao³, PING Jiawei¹, DENG Chuang¹, Peng Bingfeng^{1,4}, XU Xin^{1,4}

(1. Shanghai Aerospace Electronic Technology Institute, Shanghai 201109, China;

2. School of Computer, Shanghai Jiao Tong University, Shanghai 200240, China;

3. Aerospace System Engineering Shanghai, Shanghai 201109, China;

4. School of Astronautics, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China)

Abstract: In response to the problems of decentralized management and insufficient utilization of launch data for rocket testing, a data warehouse technology is proposed to organize and store the test and launch data of rockets, achieving the deep exploration of the rocket data. According to the characteristics of the carrier rocket data, it proposes six levels of framework design such as the data warehouse foundation layer, source data layer, data interaction layer, data architecture layer, application layer, and display layer; The different sources and format data are cleaned and loaded by the data cleaning rule design for the rocket data analysis and fault diagnosis; The direction of rocket data application is explored, which designs data analysis modules such as the envelope analysis, correlation analysis, and fault diagnosis. The validation is conducted using launch data from a certain type of rocket, and the results show that the data warehouse solution is feasible, and it provides a foundation for the continuous management and analysis of the rocket data in the future.

Keywords: data warehouse; carrier rocket; structured data; data cleaning; fault diagnosis

0 引言

运载火箭系统是一项极为复杂的大规模系统工程, 横跨计算机、电子、通信、控制、材料、结构以及机械等领域, 每一枚火箭的研制均需数以千计的单位参与, 其发射以及飞行过程中会经历高温、强冲击、以及复杂恶劣的太空环境, 任何一部分构件的异常都可能导致发射失利, 这对于需要高投入、长周期研制的火箭损失难以承受, 其上所搭载的航天器及卫星价值更是难以估量^[1]。2010 年 12 月, 印度 GSLV-F06 运载火箭在“萨迪什·达万”航天中心执行发射, 在发射两分钟后偏离轨道爆炸, 带来数亿美元的损失, 2013 年 2 月, 俄罗斯运载三颗“格洛纳斯-M”导航卫星的质子-M 运载火箭在发射升空后一分钟内坠落并

爆炸, 事故原因是关键的角速度传感器被颠倒安装^[2], 2016 年 9 月, 猎鹰 9 号在执行静态点火测试时发生爆炸, 价值 2 亿美元的卫星损毁, 在国内, 长征系列多款运载火箭均出现过较为严重的航天事故。因此, 可靠性的提升是运载火箭领域最为关注的重点工作之一, 其中, 最为重要的手段就是故障诊断, 通过前期的测试和监测发现潜在的故障, 及时进行处理, 避免灾难性的损失出现, 其中火箭的长期测试数据以及专家知识是进行数据分析和故障诊断的重要依据^[3]。

火箭在长期测试过程中, 产生了大量的原始数据和各种业务数据, 这些数据分散在各承制单位, 各试验现场, 数据包括全箭各系统数据及单机数据。但是长期以来, 运载火箭试验过程中所采集的数据只是作为型号任务的产品,

收稿日期: 2023-08-27; 修回日期: 2023-10-30。

作者简介: 陈卓(1993-), 男, 硕士研究生, 工程师。

通讯作者: 汪灏(1991-), 男, 硕士研究生, 工程师。

引用格式: 陈卓, 汪灏, 平佳伟, 等. 面向火箭数据分析和故障诊断的数据仓库设计[J]. 计算机测量与控制, 2024, 32(1): 51-56.

各型号任务所获取的数据之间是隔离的,大量的历史数据处于脱机状态,绝大多数以原始数据或文本形式存储于型号电脑或光盘中,缺乏集中的存储和管理,设计师在数据分析过程中无法方便的获得历史数据^[4-5]无法通过一套软件快速的进行各阶段数据统计、分析,使得对产品的稳定性和包络分析较困难。

随着型号任务的持续增加和新型火箭的不断开发,火箭系统硬件设备逐渐呈现出设备配套数量多、种类多、部署场地多的特点,火箭系统测试数据呈现出格式不统一、数据种类多样的特点^[6]。现阶段对于火箭的数据分析仅限于同型号有限批次的比对,比对分析方式以理论知识为主,历史数据为辅,跨型号、跨时间的数据分析较少、不够全面,这样导致火箭产品在设计阶段缺乏足够的先验数据支撑;另一方面火箭测试数据的分散存储和格式不统一,给数据进行的全面的分析和利用带来了困难,使得火箭的故障诊断成为无源之木^[7]。

运载火箭领域历史积累的测试数据量是巨大的,具备很大的潜藏价值,但目前这些数据只是被分散存储起来,没有形成完整的数据资源目录、数据视图和数据共享机制,在数据资源深层次分析开发上也缺乏必要的手段,试验数据未能进行有效的统计、分析和评估,无法将这些测试和发射数据转换成对火箭有用的信息,这些宝贵的试验和发射数据价值未能发挥。

基于现状,为了解决运载领域测试数据利用率不高的问题,提高分析和决策效率和有效性,有必要对数据进行组织和管理,发挥运载领域测试数据应有的价值,为后续的任务服务,数据仓库正是为了构建这种新的分析型环境而出现的一种数据存储和组织技术^[8-11],本文提出将数据仓库技术应用于运载领域的的数据汇总,对火箭测试发射数据的进行综合管理和深层次的挖掘分析,辅助型号任务工作,为火箭的故障诊断工作打下基础。

火箭的故障诊断技术已经经过多年长期的发展,取得了不少理论成果并得到了实际应用,从 20 世纪 70 年代开始,美国开始在运载火箭上应用健康管理,监测发动机运行状态,此后,故障诊断和状态监测的理论和技术得到了迅猛的发展^[12]。可以将目前主流的故障诊断方法分为如下三类:基于信号处理的故障诊断方法、基于专家的故障诊断方法和基于数据挖掘的故障诊断方法。其中,基于信号处理的故障方法通过对采样信号进行降维升维和时频域变换提取特征信号进行故障诊断,常见的有小波变换和傅里叶变换^[13],不需要了解设备的工作原理,但在面对故障模式多样的火箭系统存在局限性;基于专家的故障诊断方法是根据火箭的理论知识以及经验知识,对可能出现的故障模式进行预测,常见的有知识库和知识图谱^[14],此种方法对于知识库内已有的故障模式有着较好的识别效果,但对于知识库外的故障是无法识别和预测的;基于数据挖掘的故障诊断方法成为近年来的主流研究方向,以大量的测试数据为基础,通过深度学习对数据样本进行训练,对

于复杂系统具有较高的适应性,但火箭的故障样本较少使其诊断精准度不高^[15]。

针对复杂运载火箭系统在故障数据不充分、故障模式不确定情况下的故障诊断需求,提出了一种基于数据挖掘与专家知识相结合的故障诊断方法,以专家知识作为深度学习的约束条件解决数据挖掘方法的结果不稳定问题,面向火箭数据分析和故障诊断的数据仓库是火箭故障诊断的基础。

1 数据仓库方案总体设计

数据仓库由数据库发展而来,但两者在很多方面都存在很大的差异性,数据仓库已经脱离了软件产品的范畴,是一种综合性的解决方案^[16]。数据仓库是面向分析主题的、历史数据、多维的数据集合^[17-18],主要面向数据分析以及管理决策,需要对数据根据主题进行清洗、加工、汇总,形成规范、统一的全局信息;而数据库则是面向业务的、来源单一、查询量小,可以进行实时的数值处理,主要用于面向局部业务的数据查询处理工作^[19]。

具体到运载火箭领域,数据库更多的面向于火箭系统某一任务,某一型号,为具体的某一发测发业务提供查询、判读、存储等服务;数据仓库则是多维度的,可以综合管理运载火箭领域的测试数据,将不同型号、不同发次的数据,抽取数据特征信息,进行统一的治理,为火箭的数据分析和故障诊断等业务提供服务。

运载火箭数据仓库主要功能是根据数据的不同来源和信息特征,形成多维数据存储空间,为用户提供一个综合的、面向分析决策的支持环境,结合数据仓库一般结构和运载领域测试数据的特点,数据仓库采用六层分层框架进行设计,如图 1 所示。

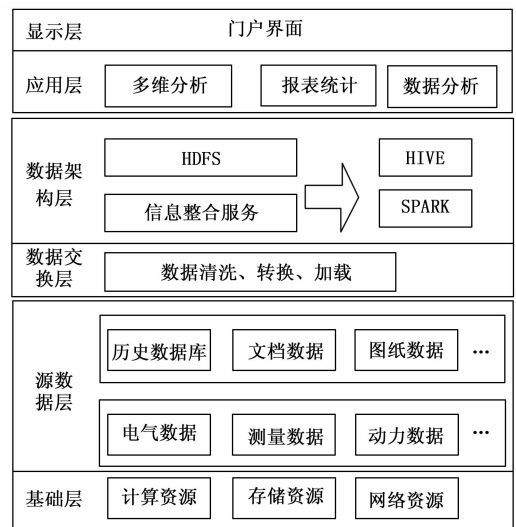


图 1 数据仓库分层结构

1) 基础层部分:由于火箭数据物理来源多样,位于各分系统实验室、综合测试厂房、试验场,各来源地之间硬件平台的多样性,因此数据仓库基础层不仅要包含支撑运

行的计算资源和存储资源, 还要包含网络资源, 预留数据持续维护更新的接口, 为数据的自动录入奠定基础。

2) 源数据层部分: 数据的来源不局限于测试发射生成的数据, 还包含产品在设计过程中的图纸、文档、记录等数据, 其来源可以从数据类型和格式两个维度进行分类。按数据类型分, 可以分为电气数据、测量数据、动力数据等; 按照数据格式可以分为结构化和非结构化数据, 结构化数据主要包括存储在光盘、硬盘中的“dat”格式源数据文件, “txt”格式文本文件及 SqlServer、Mysql 数据库中的测试数据, 非结构化数据包括文档数据、图像数据。

3) 数据交换层部分: 设置数据交换平台, 采取松耦合连接方式, 将本应在数据仓库进行的数据清洗、转换工作转移到数据交换平台, 避免源数据与数据仓库的直接交互, 提升数据获取、加载效率。

4) 数据架构层部分: 主要是完成数据的加载工作, 实现对各种同构、异构试验数据的整合, 以灵活、高效的方式进行存储, 以数据标签的方式对数据进行分类, 数据管控从元数据管理、数据质量和数据标准等三个方面统一对存入的数据进行标准化处理。其中 HDFS (hadoop-distributed-file-system) 为整个平台存储数据的基础, 用来存储导入进来的原始数据; HIVE 为火箭数据提供的基于大数据的数据仓库, 用来存储清洗后、数据映射后的范式化数据; Spark 平台为离线的数据分析与处理提供底层支撑。

5) 应用层部分: 主要负责具体业务逻辑的实现, 对数据进行清洗、映射, 实现数据仓库的维护和管理, 避免用户对数据仓库直接访问, 提供多维度数据分析、查询、数据统计等服务, 支持定制化统计分析以及更进一步的数据挖掘。

6) 显示层部分: 主要实现统一的交互界面, 将查询报表、统计分析、数据挖掘的结果展示在用户面前。

2 运载火箭数据仓库主题及模型设计

2.1 业务层主题

数据仓库的建立是为业务层的分析和决策提供服务的^[20], 对业务需求和特点的充分理解是后续数据模型是否准确和有效的关键, 根据运载火箭数据仓库的业务类型, 设计了三大主题, 如图 1。

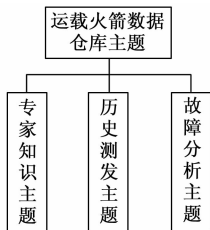


图 2 运载火箭数据仓库主题

专家知识主题: 运载火箭在产品研制过程中会产生大量的设计数据, 包括原理图、源代码、文档报告等原始知识, 可以分为硬件和软件两大类, 硬件包含图纸、物料和结构等内容, 软件包含源码、工程化文档和等内容, 这些

知识对于新产品的设计以及现有产品的优化改进有着重要的参考意义^[21], 通过查阅功能相似的产品设计图纸、软件功能模块和相关产品文档, 能够有效提高新设计产品的可靠性、功能完备性。

历史测发数据分析主题: 火箭在不同批次和不同阶段试验中产生了大量的测发原始数据, 这些数据全面记录了火箭在不同批次、不同流程、不同地点下的状态, 这些数据对于现有火箭产品的设计、状态分析和发射决策具有重要的指导意义。运载火箭的原始数据分散存储于各个系统, 数据格式未统一, 有必要将数据提取出来以变数据挖掘分析, 可以跨时间维度、跨型号维度、跨靶场维度等多角度对测发数据进行比对分析, 对火箭设备的状态把控以及产品设计提供数据支撑。

故障数据分析主题: 在火箭测发过程中, 会对各单机电压、电路和加注过程中的压力、液位、流量等数据进行关注, 在不同的测试阶段其数值会呈现规律性变化, 当火箭系统设备故障或者环境变化时, 这些参数会出现异常数值, 往往代表着火箭设备隐藏的异常, 需要对其发生机理进行分析和确定, 判断是否继续执行发射, 因此需要建立故障数据分析主题, 实现故障机理分析和故障趋势分析, 根据故障类别, 可以分为供配电故障、控制单机故障、电缆故障、软件故障、库房故障、配气台故障等多个细分子主题。

2.2 面向火箭生命周期的数据模型

数据模型的建立是对火箭进行数据抽取以及主题分析的基础。能否充分理解和透彻分析火箭数据以及业务需求直接影响到后面数据模型的设计是否合理、有效, 面向火箭生命周期建模, 构建统一的数据表示模型, 模型需要能够表征专家知识主题、历史测发主题和故障分析主题所需要的多个维度信息, 这些维度信息对应原始数据映射后的存储结构。

建立数据模型的目的是帮助对离散独立的原始数据进行规范统一的处理, 以清洗的数据逻辑关系实现主题分析工作。结合火箭的分析需求, 以历史测发主题分析为例, 当要对某一单机设备特性进行分析的时候, 该单机的历史数据反映的是该关于该单机设备方方面面的信息, 能够对该单机的特性、质量和极限情况做出评估, 可以从 9 个维度对数据仓库中的信息进行分类, 分别是设备参数、时间、任务、阶段、地点、流程、型号、设计信息和故障信息维度。

1) 设备参数维度包含该单机设备对应的可表征该设备状态的信息, 如单压、电流、温度和压力等信息。

2) 时间维度包含该设备历史参数记录的日期。

3) 任务维度包含该单机设备历史测发数据分别对应的某一型号的次发信息。

4) 阶段维度包含历史测发数据所对应的研制阶段, 比如分系统阶段、集成试验阶段、合练阶段和正式任务阶段。

5) 地点维度指当前历史测发记录生成的地点, 包含分系统实验室、总装厂房以及各大靶场。

6) 流程维度包含当前记录所对应的试验流程, 不同的试验流程对单机的特性影响比较大, 包含总检 II、总检 I 等测试流程。

7) 型号维度包含了当前测试记录所对应的型号, 多数单机设备可通用于各个型号, 通过比对单机设备在不同型号中的表现对于单机特性分析具备很大的意义。

8) 设计信息维度包含单机设备的原理图、结构、物料和软件等信息, 可以从顶层统筹了解单机设备特性。

9) 故障信息维度包含单机设备相关的软硬件归零信息, 以及各项更改和偏离设计。

3 系统软件设计

3.1 数据清洗方案设计

运载火箭试验数据资源相对稳定, 在不同阶段的测试数据均存在备份, 但这些数据未经整理, 来自于不同型号、不同阶段的试验, 存在着格式不统一、重复数据、无效数据等问题, 需要通过数据集成之后传输进数据仓库。数据集成的一般过程主要是将异构分布式数据源数据通过一定的规则映射复制到数据仓库, 如图 3 所示, 将不同来源的数据按统一的格式处理, 主要完成的工作包括数据的抽取、转换和加载, 即数据的 ETL (extract-transform-load) 过程^[22-25]。

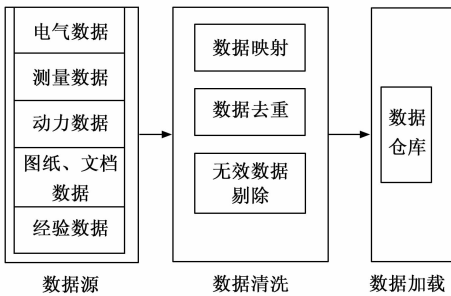


图 3 数据清洗流程

ETL 过程就是将火箭试验过程中产生的数据转化为能够存储在数据仓库的决策支持型数据, 设计到多个数据源的数据清洗及合并。对于试验数据的清洗, 分为结构化数据和非结构化数据两类, 对于已经结构化的数据, 比如火箭各测试阶段源数据、存储在数据库中的数据, 可以通过制定 ETL 规则, 直接抽取所需要的数据, 对于文档, 图纸、图像等非结构化的数据, 可以人工预定分类和编码准则, 对文件进行梳理, 提取有用信息。

数据去重和无效数据剔除后, 需要抽取有效数字字段并进行数字映射到数据仓库中, 主要是根据试验数据的分析和故障诊断需求, 对数据字段进行拆分, 建立数据字段之间的映射^[26], 进行相关字段的合并和补充, 对于运载测试数据, 建立时间序列、数据来源等多维度视角对数据进行梳理合并。

3.2 数据仓库应用层设计

运载火箭数据仓库系统采用分布式软件架构, 建成一个以数据存储、管理、统计分析为核心功能的分布式数据

管理系统, 如图 4 所示, 它基于 Hadoop 分布式系统进行架构使用 Hive 数据仓库、Flume 数据采集工具、Spark 分布式计算框架形成后台计算系统, 前端可视化系统, 最后进行联合开发, 以满足数据可视化显示, 数据存储、分析任务管理等功能。

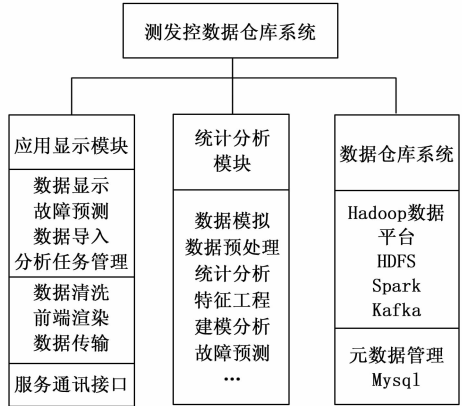


图 4 测发控数据仓库软件结构

应用层软件架构主要分为三个模块, 数据仓库模块、应用显示模块、统计分析模块。

1) 数据仓库模块: 负责管理火箭数据的存储、查询和计算, 将大量的设计数据和测试数据进行科学的保存和管理, 让设计人员能够便捷的获取全面多维度的测试数据和设计数据, 对数据进行分析利用。包括 Hadoop、Hive、Spark、Kafka、Flume 等组件, 使用 Mysql 进行元数据管理。

该模块主要提供以下功能点: ①历史火箭原始数据的存储、新增测试数据的导入; ②结构化和非结构化数据的存储与关联; ③完成清洗合并后的数据上载; ④元数据的底层操作管理。

2) 应用显示模块: 负责数据显示、导入和任务管理等与用户直接交互的功能。前端显示界面包括数据展示、数据导入导出接口、数据定制化分析界面、故障预测界面。后端处理负责数据查询、计算、传输和前端渲染工作。此外该模块还将通过服务通讯接口与数据仓库模块, 统计分析模块通讯, 以实现数据交互和任务交互功能。

应用层软件采用 B/S 架构设计, 因此本模块主体分为两个部分, 一个网页端、一个是服务端, 网页端作为与用户的主要交互界面, 服务端则负责网页的渲染, 数据传输、缓存以及外部通讯的功能。网页端则, 位于上层, 通过网页展示相关性分析等数据展示、故障预测功能, 进行导入数据和分析任务管理等。

3) 统计分析模块: 负责具体的数据模拟, 预处理、统计分析、特征工程, 建模分析、故障诊断等功能, 是数据分析应用的核心模块。统计分析模块是发挥火箭宝贵的数据价值的关键所在, 通过深层次的数据分析, 挖掘不同测试参数之间存在大量的内在的、潜在的数据和信息^[27], 为火箭的设计和测试工作提供帮助。

该模块主要提供以下功能点: ①具备相关性分析能力,

可自由选择测试参数进行分析并生成结果; ②具备基于历史数据学习的故障诊断能力, 能够实现对火箭常规参数的拐点识别, 异常识别; ③具备数据包络分析能力, 可跨型号、跨时域对参数实现包络分析④提供定制化数据分析模块, 满足不同的数据分析需求。

3.3 基于数据和知识融合的故障诊断方法

现有基于统计学习的模型在独立同分布的情况下有不错的结果, 但是火箭的数据样本大多不满足独立同分布的样本且数量较少, 诊断结果就会变得不稳定, 缺乏可解释性, 即使诊断结果是正确的, 也无法将该结果作为火箭决策的依据, 为了提高统计模型诊断的效果, 需要进行数据和知识融合的故障诊断, 以知识因果关系训练统计学习模型, 充分利用专家知识。

首先对数据和知识进行相关性分析以确定数据的因果方向, 针对时序数据、非时序数据和独立同分布数据的每一对节点进行独立性检验, 将其中一个节点去掉保持其余不变, 计算因果图的性能, 如果性能接近保持不变, 则认为该边可以被去掉, 重复这一过程, 直至任何相连的节点都不独立, 随后对图的方向进行检验, 确保因果图的方向是由因到果, 针对得到的因果图, 使用干预推理对遗漏的因果关系进行调整补充, 最后采用反事实推理方法找出给定结果的原因, 最后得到基于因果理论的结构因果模型。

随后进行基于强化学习的因果建模, 使用结构因果模型对数据仓库切片数据得到因果图, 使用合适的评价函数进行训练, 寻找最优的因果图。

将因果推理嵌入到机器学习模型中, 通过数据仓库中的知识因果关系, 使得机器学习模型在受到因果关系约束的情况下对样本数据进行学习, 从而提升模型的准确率和稳定性。

4 实验结果与分析

4.1 实验环境及验证方法

为了验证面向火箭数据的数据仓库实际使用性能, 数据仓库的构建包括 4 台服务器 (CPU: Intel Xeon E5-2640 六核心, 主频 2.5 GHz, 内存 32 G, 存储 4 T) 和多台显示终端。数据仓库的构建需要以 Hadoop 大数据平台为基础, 使用 3 台服务器作为 Hadoop 节点组成数据仓库, 1 台服务器作为应用服务器负责应用显示模块和数据分析模块, 4 台服务器配置相同, 数据仓库系统通过交换机与外部访问终端连接, 访问终端以网页的形式与数据仓库进行交互, 进行数据的导入、导出, 数据的分析处理。

原始数据选用某型号 10 发火箭历史数据, 数据包包含老的测发控系统测试发射数据和新型一体化测发控系统测试发射数据, 数据生成周期包含集成实验、发射场实验数据, 数据维度包含电气数据、测量数据、动力数据。

为测试运载火箭数据仓库能力, 从数据导入和历史测发数据查询两个方面对数据仓库进行测试。

4.2 数据导入测试

新型一体化测发控系统在软件设计和通信协议上较老

的测发控系统有很大不同, 对应的火箭历史数据也有很大不同, 需要对火箭数据从顶层结构上进行预处理, 将同单机同硬件结构所产生的数据进行数据抽取, 映射到同一数据标签下, 将同单机不同硬件结构所产生的数据之间进行关联, 映射到不同的数据标签下。在进行数据导入工作时, 设计师需选择导入的数据类型以及对应的测试属性, 如图 5 所示, 上传成功后, ETL 工具会对数据进行抽取映射, 按统一的数据格式存储于数据仓库中。



图 5 数据导入界面

为验证数据仓库对于火箭原始数据的导入速度, 选择某型号 3 组原始测试数据进行测试, 测试数据均选用在相同测试流程下生成的原始记录数据, 三组测试数据分别包含 1 条测试流程、5 条测试流程和 10 条测试流程, 数据导入测试结果如表 1 所示。

表 1 数据导入测试结果

组别	测试流程数/条	原始大小/G	导入时间/分
分组 1	1	0.45	11
分组 2	5	2.31	40
分组 3	10	4.59	67

从上数试验结果可以看出, 导入数所需要的时间并非线性随数据大小而增长, 而是逐步减少的, 因为在数据导入过程中伴随着清洗与合并操作, 当数据量较大时, 该部分工作所占用时间会被均摊。

4.3 历史测发数据查询测试

对于数据的统计分析功能, 数据仓库支持对数据的单任务多参数、多任务单参数、跨型号参数查询、不同参数关联查询, 支持多个维度的定义查询规则, 同时支持参数的包络分析, 拐点识别等功能。

以地面电源参数为例进行数据查询和分析, 地面电源是火箭系统的重要设备之一, 为火箭的测试和发射提供能源支持和保障, 其稳定性是整个火箭能否正常工作的关键, 通过电压和电流等参数间接判断地面电源状态是每次测试必做的内容之一。在本型号直流电源共计 4 台, 选择“D96-1 电源电压”作为查询参数, 如图 6 所示, 数据仓库提供了“D96-1 电源电压”参数的全周期可视化展示, 蓝色的曲线为电压变化情况, 绿色曲线为数据分析服务自动计

算出在该测试周期的参数包络值，黄色数据点为数据分析服务识别到的参数拐点。

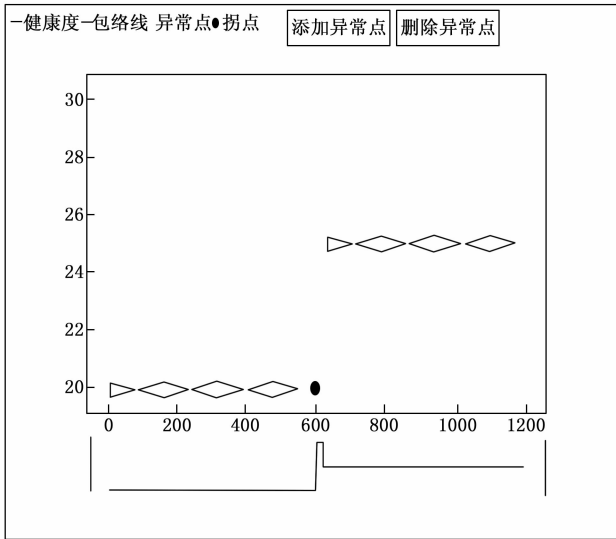


图 6 “D96-1 电源电压”数据查询

为验证历史数据查询性能，使用“D96-1 电源电压”参数，分别选择查询不同时间区域的参数，观察其结果响应时间。

表 2 历史数据查询响应结果

项目名称	参数/万条	响应时间/s
查询 1	0.78	2.3
查询 2	5.67	4.5
查询 3	45.32	27.93

从上述查询结果可以看出，数据仓库对于历史测发主题的参数查询响应时间为秒级，该响应时间可以支撑在实际型号任务中的应用。

5 结束语

本文研究了面向火箭数据分析和故障诊断的数据仓库设计。对火箭数据进行了跨型号、跨批次的整合和管理，为火箭测试和设计数据的长期维护搭建了框架，目前已在某型号的数据管理中实现应用，能够对火箭测试数据进行分析 and 展示。但本文在对火箭的数据利用方面只进行了相关性分析、简单故障识别，后续将通过机器学习等手段，进一步挖掘火箭历史数据中的价值。

参考文献:

[1] 向刚, 韩峰, 周虎, 等. 数据驱动的航天器故障诊断研究现状及挑战 [J]. 电子测量与仪器学报, 2021, 35 (2): 1-15.

[2] LI W, LI P, ZOU Y. Review and Future Trend of Space Propulsion Technique Using Hydrocarb On Propellants [J]. Journal of Astronautics, 2015, 36 (3): 243-252.

[3] 彭宇, 刘大同, 彭喜元, 等. 故障预测与健康管理技术综述 [J]. 电子测量与仪器学报, 2010, 24 (1): 1-9.

[4] 张晓妍, 沈利军. 测量数据仓库的概念研究 [J]. 遥测遥控, 2018, 39 (1): 1-5.

[5] 王家欢. 基于语义的数据仓库构建方法研究及应用 [D]. 上海: 上海交通大学, 2020.

[6] 慈颖. 基于航天装备数据知识图谱体系研究 [J]. 计算机测量与控制, 2023, 31 (5): 249-254.

[7] 何克磊. 飞行器健康评估和故障预测技术研究 [D]. 南京: 南京航空航天大学, 2012: 5-15.

[8] 罗红华. 基于数据挖掘及数据仓库技术的研究与应用 [J]. 通讯世界, 2019, 26 (10): 77-78.

[9] 卫一兰. 基于数据挖掘技术的民航企业决策支持系统的设计与实现 [D]. 成都: 电子科技大学, 2019.

[10] 吴纪龙. 中医药大数据资源数据仓库构建及处方分析应用研究 [D]. 北京: 北京交通大学, 2021.

[11] 吴振涛. 基于数据仓库技术的数据集成在数字化校园中的应用 [J]. 电子设计工程, 2016, 24 (9): 28-31.

[12] FIGUEROA F, WALKER M, UNDERWOOD L W. NASA Platform for Autonomous Systems (NPAS) [C] // San Diego: AIAA Scitech 2019 Forum, 2019.

[13] 戴屹梅. 面向突变信号的火箭发动机轴承故障特征提取方法 [D]. 北京: 北京交通大学, 2019.

[14] 陈潇萍, 王剑锋, 张虹, 等. 液氧液氢发动机领域知识图谱构建与应用 [J]. 北京航空航天大学学报, 2022, 3 (3): 1-11.

[15] 陈舒文. 基于深度学习算法的航空电源设备故障诊断技术研究 [D]. 南京: 南京航空航天大学, 2020.

[16] 任仲晟. 基于数据仓库的数据挖掘技术 [J]. 数字技术与应用, 2021, 39 (9): 59-61.

[17] 何龙祥, 葛继成, 王轻, 等. 基于数据仓库的医疗数据可视化系统的设计与实现 [J]. 软件设计研究与应用, 2021, 42 (1): 104-106.

[18] 刘海强, 陈晓晶, 张兴华, 等. 面向煤矿安全监控的数据仓库关键技术 [J]. 工矿自动化, 2022, 48 (4): 31-37.

[19] 蔡金芷, 皮喜田. 个人健康管理数据库的设计与实现 [J]. 计算机测量与控制, 2017, 25 (7): 222-224.

[20] 喻凡坤. 无人系统故障知识图谱的构建方法及应用 [J]. 计算机测量与控制, 2020, 28 (10): 66-71.

[21] 范满意. 基于知识图谱的航空发动机 PHM 仿真验证平台设计 [J]. 计算机测量与控制, 2023, 31 (6): 305-313.

[22] 张军, 王芬芬. 数据仓库技术在高校数据统计与分析系统中的应用研究 [J]. 智能计算机与应用, 2019, 9 (3): 122-125.

[23] 雷博文. 基于大数据的实时数据仓库的设计与实现 [D]. 北京: 中国地质大学 (北京), 2021.

[24] 周春红. ETL 在 BI 项目中的设计与实现 [D]. 上海: 东华大学, 2016.

[25] 谢峰. 军械装备保障数据仓库中的 ETL 过程优化研究 [J]. 计算机测量与控制, 2015, 23 (5): 1706-1709.

[26] PETROVIC M, VUCKOVIC M, TURAJLIC N, et al. Automating ETL processes using the domain-specific modeling approach [J]. Information Systems and e-Business Management, 2017, 15 (2): 425-460.

[27] 杨欣茹, 王猛, 高莉, 等. 固液捆绑火箭发动机健康诊断技术 [J]. 上海航天, 2022, 39 (5): 8-14.