

# 航天器试验大数据系统实现与典型应用

王建军<sup>1</sup>, 向永清<sup>2</sup>, 何正文<sup>1</sup>

(1. 北京空间飞行器总体设计部, 北京 100094; 2. 北京三七数据技术有限公司, 北京 100085)

**摘要:** 航天器试验数据具有大数据特征, 对于航天器研制中所做的评价和决策具有至关重要的作用; 为实现航天器试验数据的高效且有效采集、存储、处理、分析和应用, 提高航天器研制水平, 分析了航天器试验数据的典型特征; 借助大数据技术成果构建了用于航天器试验的大数据系统, 描绘了系统部署位置, 基于 Hadoop、Flume、Kafka、MongoDB、HDFS、Spark Streaming、Presto、Kylin 等技术构建了由数据采集层、数据存储层、数据处理层、数据分析层、数据访问层和数据应用层共 6 层构成的系统架构, 阐述了系统各层的主要技术实现途径; 探讨了系统在实时监测、健康评估、试验有效性评价、可靠性评估与寿命预测、发现知识、虚拟试验等方面的典型应用与展望。

**关键词:** 航天器; 试验; 大数据; 应用

## Implementation and Typical Application of Big Data System for Spacecraft Test

Wang Jianjun<sup>1</sup>, Xiang Yongqing<sup>2</sup>, He Zhengwen<sup>1</sup>

(1. Beijing Institute of Spacecraft System Engineering, Beijing 100094, China;

2. Beijing Sanqi Big Data Technology Co., Ltd., Beijing 100085, China)

**Abstract:** Spacecraft test data has typical big data characteristic, and is critical to evaluations and decisions made in spacecraft development. In order to achieve efficient and effective acquisition, storage, processing, analysis and application of spacecraft test data, the characteristic of spacecraft test data was analyzed, the big data system for spacecraft test was built, the deployment position was given. The system architecture is based on technology of Hadoop, Flume, Kafka, MongoDB, HDFS, Spark Streaming, Presto, Kylin, etc. and composed of data acquisition layer, data storage layer, data processing layer, data analysis layer, data access layer and data application layer. The implementation approach of each layer was investigated. Typical application and prospect were discussed, such as real-time monitoring, health evaluation, test validity evaluation, reliability assessment and life time forecast, knowledge discovery, virtual test, etc.

**Keywords:** spacecraft; test; big data; application

## 0 引言

航天器试验是指为验证和检验航天器功能、性能、品质和可靠性所进行的各种试验; 为保证航天器任务圆满成功, 航天器研制一直在不断完善和实施精细化管理, 围绕着信息尤其是试验信息进行科学严密的评价和决策<sup>[1]</sup>。

大数据技术是当前信息领域的发展新趋势, 大数据是支撑国家安全和国防建设的战略资源<sup>[2]</sup>。《中国制造 2025》、《促进大数据发展行动纲要》、《国家信息化发展战略纲要》等均提出实施国家大数据战略, 要求突破工业大数据处理等核心技术, 并把航天装备列为重点领域突破发展<sup>[3-4]</sup>; 《2016 中国的航天》白皮书也提出要加快推进航天工业化与信息化的深度融合。

航天器研制应当抓住工业化与信息化融合发展这一重大历史机遇, 分析航天器试验数据特征, 运用当前先进的大数据技术, 建立科学的用于航天器试验的大数据系统(以下简称“航天器试验大数据系统”), 实现航天器试验数据采集、存储、处理、分析和应用的数字化和统一管

理, 提升航天器研制水平。

## 1 航天器试验数据特征分析

航天器试验数据是指在航天器试验设计、准备、实施、撤收、总结等各阶段所产生的数据资源。航天器试验数据具有大数据的 4V 特征, 即 variety (数据类型多样)、volume (数据量巨大)、velocity (速度要快)、value (价值巨大)<sup>[5]</sup>, 主要表现在:

1) 类型。从试验类型上来分, 包括电性能测试、环境模拟试验、飞行试验等一系列试验, 其中电性能测试包括常规综合测试和各种环境模拟试验下的测试, 环境模拟试验包括力学环境试验、真空热环境试验、电磁环境试验等; 从数据接口上来分, 包括 CAN、1553B、RS-422、以太网、VXI、GPIB、无线链路等; 从数据编码格式上来分, 包括 PSK、BPSK、DPSK、QPSK 等; 从数据类型上来分, 包括试验方案、大纲、细则、总结等文本, 试验监视、测量等数值, 试验图像、视频、音频等记录, 还有三维模型、试验用例、试验统计分析和报表、试验规章制度、试验日志等数据, 涵盖了结构化、半结构化和非结构化数据;

2) 规模。航天器自身功能越来越强大使其产生的数据量加大, 包括遥测数据、遥控数据、载荷数据、总线数据、

收稿日期: 2018-10-16; 修回日期: 2018-11-03。

作者简介: 王建军(1982-), 男, 河北张家口人, 硕士, 高级工程师, 主要从事航天器研制、航天器系统工程与项目管理方向的研究。

信号波形和频谱数据等; 试验的精细化使试验仪器设备产生的数据量加大, 包括示波器、频谱仪、供配电、测控、载荷等测试设备产生的激励和采集的数据, 采集到的数据经过解析处理后产生大量数据; 航天器要经历大量试验使数据量加大, 包括温度、湿度、洁净度、真空度、热环境、力学和噪声等环境模拟和监测数据、环境模拟分析数据、试验现场视频监视数据等; 航天器试验历时周期长(1—3年), 系统通电时间长(正样飞行产品系统级电性能测试正式通电时间约 2000 h); 每年同时在试验航天器可以达到十余颗甚至数十颗;

3) 速度。每颗航天器自身高速产生遥测数据(量级  $10^3$  bps, 射频调制)、载荷数据(量级  $10^3$  bps, 射频调制)、总线数据(量级  $10^3$  bps)、脐带数据(量级  $10^3$  bps), 试验用各类传感器高速产生数据流, 试验用仪器设备同样高速产生数据; 而试验数据实时采集、分析、应用要求不断提高;

4) 价值。航天器数据中不仅包含着试验设备和受试产品状态信息, 还包含着航天器功能、性能、品质和可靠性信息, 还可以通过大数据技术进一步挖掘, 摸清航天器试验过程中所体现的特点, 为设计改进、试验技术的改进和领导决策等提供支持, 具有极高的应用价值<sup>[6]</sup>。

除此之外, 航天器试验数据还具有更强的多源性、专业性、时序性、关联性、噪声性和更高的采集、分析和处理准确性要求等特征。

## 2 系统实现

### 2.1 部署位置

由于航天器试验具有试验意义重、系统和过程复杂、采用技术新、多学科交叉、规模和费用大、环境和可靠性高等特点, 航天器试验大数据系统的部署应充分继承现有成熟试验系统技术并与其有机结合, 不影响试验任务的顺利开展, 在此前提下逐步完善替换原有系统功能。航天器试验大数据系统在整个航天器试验系统中的部署位置如图 1 所示, 大数据系统直接接入综合试验网络, 读取原有试验系统采集到的数据, 充分利用大数据技术实现试验数据存储、分析和应用。

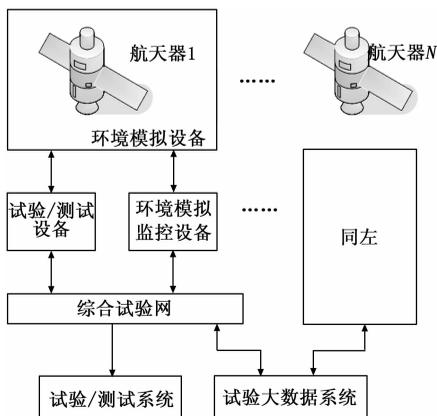


图 1 系统部署位置

### 2.2 系统架构

航天器试验大数据系统架构如图 2 所示, 按照数据流向自底向上分别是数据采集层、数据存储层、数据处理层、数据分析层、数据访问层和数据应用层, 此外数据管理平台贯穿全部层级。

1) 数据采集层负责将航天器试验过程产生的各类结构化、半结构化和非结构化数据通过离线 ETL 和实时采集等方式进行归集汇总;

2) 数据存储层根据航天器试验数据的不同特点, 将数据采集层采集到的数据分别存储到文档数据库、关系数据库、图数据库等数据库中;

3) 数据处理层和数据分析层负责航天器试验数据的处理和分析, 数据处理可采用批处理、实时处理及交互式查询等方法, 数据分析可采用传统统计分析、深度学习及数据挖掘等方法;

4) 数据访问层是航天器试验数据与应用之间的桥梁, 可提供命令行、Restful API、RPC、Web Service、SQL 等接口, 根据应用层的应用需求通过不同接口从数据分析层获取所需数据;

5) 数据应用层是航天器试验数据各种应用的集合, 包括实时监控类应用、统计分析类应用、交互式查询类应用及日志分析类应用等;

6) 数据管理平台是整个航天器试验大数据系统升级扩展、平稳运行、安全运维的基础, 具有对各类技术组件的透明访问能力, 完成数据设计、获取、应用、销毁等试验数据的全生命周期管理, 并把标准、质量规则和安全策略固化在平台上, 实现事前管理、事中控制、事后稽核与审计的全方位数据质量管理和安全管理。

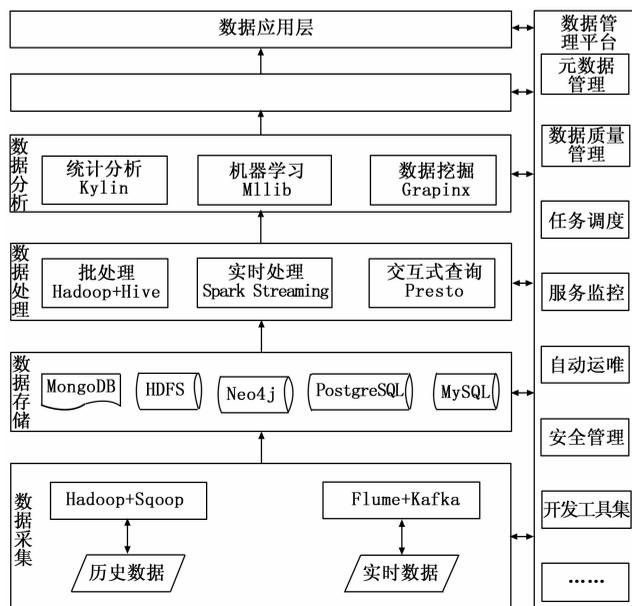


图 2 系统架构

### 2.3 主要技术实现

#### 2.3.1 数据采集层

航天器试验数据的采集分为历史数据的离线批量采集

和实时数据的流式采集。

1) 历史数据的离线批量采集采用 Hadoop 平台作为计算和存储引擎, 使用 Sqoop 工具通过 ETL 的方式进行数据的采集, 并实现 HDFS 与 MySQL、PostgreSQL、MongoDB 等传统数据库之间的数据交换;

2) 实时数据的实时采集采用 Flume 与 Kafka 结合的方式, 使用 Flume 作为数据的生产者, 使用 Kafka 作为数据的消费者, 这样可得到较高的吞吐量和可靠性。

### 2.3.2 数据存储层

航天器试验数据的存储选用多种数据库结合的方式: 选用 MongoDB 作为文档数据的存储工具, 选用 HDFS 作为大规模离线数据的存储工具, 选用面向网络的数据库 Neo4j 存储试验故障诊断相关的专家知识和决策树信息, 选用 PostgreSQL 和 MySQL 作为一些结构化数据的存储工具 (如测试数据记录、某些设备的关键参数等)。

### 2.3.3 数据处理层

航天器试验数据的处理分为批处理、实时处理和交互式查询。

1) 批处理选用 Hadoop 平台结合数据仓储工具 Hive, Hive 是传统数据仓储的一种分布式替代, 很适合传统 ETL 中数据的清洗、过滤、转化及直接汇总等场景;

2) 实时处理 (如遥测数据流的处理) 选用 Spark Streaming 工具。Spark Streaming 是一个基于 Core Spark API 的可扩展、高吞吐量、并具有容错能力的用于处理实时数据流的组件, 可接受 Flume、Kafka、HDFS、Kinesis 或 TCP 等各种数据源传递来的数据, 对接收到的数据可使用一些高阶函数进行封装的复杂算法做进一步处理, 最后处理好的数据可写入到文件系统和数据库中, 或直接用于实时展示, 此外还可在数据流上应用一些机器学习或图计算等算法<sup>[7]</sup>;

3) 交互式查询选用 Presto 工具。Presto 是一个运行在服务器集群上的分布式系统, 如图 3 所示<sup>[8]</sup>采用 Master-Slaver 架构, 主要由一个 Coordinator 和多个 Worker 组成 (Discovery Sever 通常内嵌于 Coordinator 中)。Coordinator 负责解析并分析 Presto CLI 提交的 SQL 语句, 生成执行计划分发给 Worker 执行; Worker 负责实际执行查询任务, 启动后向 Discovery Sever 注册; Coordinator 从 Discovery Sever 获得可以正常工作的 Worker 节点; 当读取 Hive 数据时, 需配置一个 Hive Metastore 服务为 Presto 提供 Hive 元信息; Worker 与 HDFS 交互读取数据。

### 2.3.4 数据分析层

航天器试验数据的数据分析主要包括统计分析、机器学习和数据挖掘。

1) 数据统计分析选用开源分布式分析引擎 Kylin, 它提供 Hadoop、Spark 之上的 SQL 查询接口及多维分析 (OLAP) 能力以支持超大规模数据, 其架构如图 4 所示, 由 REST Sever、JDBC/ODBC 接口、Query 引擎、Routing、Metadata、Cube 构建引擎等组成<sup>[9]</sup>;

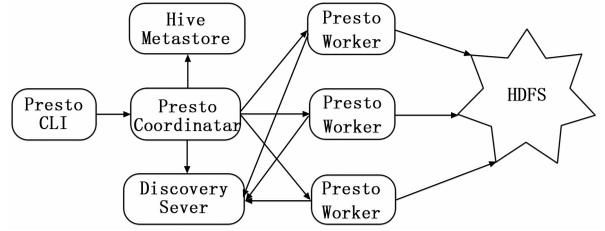


图 3 Presto 架构

2) MLlib 是 Spark 对常用机器学习算法的实现库, 同时包括相关的测试和数据生成器, 其主要功能包括特征提取、统计、分类、回归、聚类、协同过滤与推荐及降维等;

3) Graphx 是 Spark 中用于图形和图形并行计算的新组件, 基于 Spark 平台提供对图计算和图挖掘简洁易用而丰富的接口, 可用于试验数据知识图谱的构建、试验数据诊断、数据源分析等。

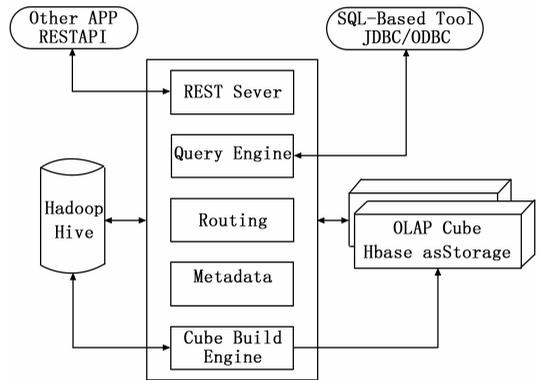


图 4 Kylin 架构

## 3 典型应用与展望

### 3.1 数据管理

航天器试验大数据系统核心模块接入现有综合试验网络, 通过网络获取原试验系统中来自于传感器采集、射频设备接收以及环境模拟和监测系统和试验设备自产生的各类航天器试验数据, 利用大数据技术提供的数据存储、交换、仓储、处理、分析与可视化工具完成试验数据的采集、清洗、筛选、解析、归一、融合、存储、分析和展现以及数据移动和备份等, 实现多学科综合过程中复杂数据传递和转换, 最大限度避免试验数据和精度损失, 提供纯净、可用的数据, 消除内部信息孤岛<sup>[10]</sup>。

### 3.2 试验实时监测

通过对航天器试验数据的采集、清洗、分析、处理和可视化, 实现试验运行状态的实时监视, 并按照规则进行数据实时判读, 及时发现问题并报警。

1) 试验运行状态实时监视。商用的大数据应用软件提供了雷达图、仪表盘、气泡图等大量优秀的可视化工具, 通过这些可视化工具对试验运行状态数据实时监视、统计分析和显示; 根据试验参数设置对试验设备响应和航天器响应进行判断, 以确认试验运行状态是否正常;

2) 数据实时判读。所有采集的数据经过解析处理后,同一数据纵向比,同类数据横向比,关联数据联合比,发现异常状态实时报警。同一数据纵向比是指对同一个参数(如用于监视设备内部电源状态的电源电压)在航天器研制各阶段的实际数据进行比对,判断数据是否一致;同类数据横向比是指对表征同一特征的同类参数(如粘贴在设备机壳内表面和外表面的温度测点)进行比对,判断数据是否合理;关联数据联合比是指对一些有关联的数据联合判断,确定数据是否正确,比如蓄电池电压与充放电时间、当前负载、太阳电池阵供电电源联合判读;

3) 试验数据网侦测。航天器和各试验设备数据均通过网络接口与试验数据网连接,通过对试验数据网通信状态的侦测,对于航天器和试验设备通信反应异常的判断其工作状态异常。

### 3.3 健康评估和故障预测

对航天器试验数据进行分析、建模和管理,评估航天器健康状态,提前预测故障避免问题发生。

1) 成功包络线建立和分析。通过对以往成功的各航天器历史试验数据进行分析,建立数据成功包络线,判断当前试验数据是否落入成功包络线内,对于超出包络线的数据即认为存在异常,进一步开展分析判断;

2) 长期走势分析和趋势预测。可以实现试验参数的长期走势分析,根据长期走势对发展趋势进行预测,如频率准确度是反映应答机的重要指标,通过分析发现其在各试验中的长期走势为一条逐渐下降曲线,那么有理由推测其性能下降并预测其后续趋势为进一步下降,于是根据故障树分析并结合其他试验数据逐步排查确定其内部晶体参数发生漂移,给出更换晶体的解决建议,并建议对后续晶体加强筛选和通电老练试验从而加速晶体应力释放;

3) 健康评估和故障预测。航天器在试验过程中健康评估和故障预测是极其复杂的,需要对各种手段获取的监测数据、历史数据等进行综合分析,利用各种预测诊断算法挖掘这些数据所反映的设备健康状态极其变化趋势,并推测出可能发生的故障模式,评估方法有多项式曲线拟合、ARMA模型、新陈代谢模型、层次分析法、模糊评判法、人工神经网络法、基于贝叶斯网络的方法、灰色理论、专家系统等方法,无论采用何种方法都需要丰富的专家知识和经验,需要一个长期的分析和论证过程<sup>[11]</sup>。大数据提供了一种全新的思维方式,采用数据驱动方法不再依赖于预设模型的精准性,而是收集大量的数据,从海量时间序列数据中寻找规律,然后用一些相对简单的模型去契合数据,消除信息的不确定性,实现航天器健康评估和故障预测<sup>[12]</sup>。另外,通过对试验过程中设备运行数据进行分析,可以对试验设备进行故障预测和健康评估,从而实现适时维修。

### 3.4 试验有效性评价

航天器试验有效性评价是一项复杂的综合性工作,有关评价指标的定义涉及很多因素,包括试验与故障的关系分析、试验技术有效性分析、模型与试验有效性指数计算、

航天器在轨早期故障预示等内容<sup>[13]</sup>。航天器试验大数据系统可以很好的将历史航天器和在试验航天器试验数据(含航天器和仪器设备状态、试验方法、条件、工况、结果等)、故障数据、试验标准规范、试验文件(大纲、细则和总结等)、试验模型等有关数据进行综合处理,给出试验有效性评价,同时避免过试验和欠试验。

通过对航天器试验数据的深入分析给出试验结论和建议。如通过航天器力学环境试验和真空热环境试验获得了大量的数据,对这些数据进行分析可以给出航天器抗力学和热设计是否合理的结论,并给出有关设计改进建议,完成航天器对力学和热环境适应性的评价。

### 3.5 可靠性评估与寿命预测

航天器可靠性评估与寿命预测是一项复杂的系统工程,无论是采用何种评估、预测模型和方法,都需要分析航天器故障模式、收集各层级产品多种来源的数据,试验数据更是重要的支撑。通过将航天器试验数据与其他所需数据结合,完成航天器可靠性评估和寿命预测。如为评估某航天器太阳翼帆板驱动机构可靠性,航天器试验大数据系统收集相似产品历史试验及在轨应用情况,在轨工况分析和故障模式分析数据,所开展加速寿命试验的考核标准、试验状态、试验环境条件、试验温度、转动次数、试验前后物理特性检验等数据,基于指数分布模型进行可靠性评估,针对主要失效模式进行寿命预测。

### 3.6 发现知识

航天器试验数据记录着航天器所有有关试验数据,透视着深邃的规律和知识。利用统计分析和人工智能,通过对历史航天器和在试验航天器海量异构多源试验数据以及试验中故障案例的汇总、分析和比较,对整个试验过程进行剖析和精细建模,开展关联度分析、时间序列分析、分类分析、聚类分析、概念分类、偏差检测、趋势预测、信息摘要提取等处理,从航天器试验数据中挖掘专业、工程和管理等维度的隐性知识,形成航天器试验标准规范、典型流程和模板、过程向导、重用库等,用于后续航天器设计、制造和试验,基于知识进行航天器系统研发设计和系统工程管理,从前端避免问题的发生,达到发现新知识、提升新能力的目的。

### 3.7 虚拟试验

航天器虚拟试验是一种基于数字模型的航天器试验数据产生、获取、分析和评价的系统工程过程,以建模仿真、虚拟现实和知识工程方法为基础。大数据提供了一个标准的数据采集、存储、分析、应用和展示平台,有力支撑虚拟试验体系的构建,解决虚拟试验海量数据、计算复杂度、实时性、开放性、资源分散、数据异构、数据可视化的问题,并可采用各类中间件构建“虚实结合”的数据融合机制,实现数字化设计、虚拟试验验证、实物试验的统一集成。通过虚拟试验,提前发现并纠正航天器设计和试验设计的薄弱环节,提高真实试验的成功率和有效性<sup>[14]</sup>。如对于

(下转第262页)