

航空发动机产品数据库架构研究与构建

Research and Construction of Aero Engine Product Database Architecture

■ 何为 孙小欢 潘若痴 张德志 / 中国航发动所

航空发动机研发过程中会产生大量产品数据，涉及发动机从论证到使用保障阶段的全生命周期。基于数据中台和大数据技术开展产品数据库系统架构的研究和构建，能够为产品数据库建设打好基础，为产品数据的管理和高效利用提供支撑。

航空发动机是典型的研制周期长、技术难度大、投入成本高的复杂系统，在研制过程中会产生大量数据，涵盖发动机的全生命周期，包括设计、制造、试验、使用保障等各个领域，对这些数据的有效存储、管理与应用对于提升航空发动机自主研发能力、提高航空发动机研发效率、保证发动机成品研发质量具有至关重要的作用。在实际的研发过程中，产品数据管理还存在着由于过程数据缺乏统一管理造成的数据浪费和丢失，数据源不唯一与数据版本未得到有效控制，数据挖掘、分析能力不足等问题。针对上述情况，基于数据中台架构和大数据技术，开展产品数据库系统的架构研究，为产品数据库系统构建打好基础，对早日实现航空发动机产品数据的管理和高效利用具有重要意义。

产品数据库系统架构设计原则与理念

产品数据库系统具有持续建设、长期应用的特点，需要从长远的业务与技术需求出发，考虑系统架构的总体设计。

一是基于数据中台理论实现数

据的充分复用。数据中台是一套可持续地让数据用起来的机制，是依据特有的业务模式和组织架构，通过有形的产品和实施方法论支撑，构建一套持续不断地把数据变成资产并服务于业务的机制。数据来源于业务并反哺业务，不断地迭代循环，实现数据可见、可用和可运营。建设数据中台的本质就是构建企业的公共数据层，把原先分散的、“烟囱式”的、杂乱的数据仓库，合并成一个可共享、可复用的数据中台。产品数据库架构设计过程充分借鉴数据中台理念，打通研发系统过程各环节、各流程数据共享链路，为协同研发提供统一权威数据源，充分实现各类数据充分复用。

二是具备多源异构数据的统一整合能力。产品数据库系统需要与需求数据管理系统（DOORS）、产品数据管理系统（PDM）、试验数据管理系统（TDM）、协同设计管理系统（CDM）进行集成，因此平台提供多源异构的数据整合能力，支持对达梦、人大金仓、Oracle、MySQL、SQLserver等数据库的统一管理，后续通过扩展也支持Hive、Hbase等大数据的统一管理。平台支持通过各类接口服务、应用程序界面（API）

函数实现系统的集成对接。平台可通过数据录入、文件导入、集成接入、数据库同步、数据托管等多种策略实现工程数据库系统数据的集中管理与透明化管控，支持数据的查询及访问权限控制。

三是具备大数据处理能力。产品数据库系统应提供全面的大数据存储和计算能力，根据大数据系统定位和任务类型衡量性能指标进行系统性能设计，支持100TB级别无序排序单节点服务器计算性能，每分钟处理不低于100GB，单节点读写性能每分钟吞吐不低于100GB。系统应提供强大的流式计算能力，并与其他功能进行融合。提供高速的列数据库，用于处理现场实时采集的大量序列化数据（来自设备、控制系统等的实时数据）。单节点服务器可支持每秒万级请求量，查询结果毫秒级返回，离线分析任务支持同时1000个以上并发任务运行。

四是具备轻量级的自助式数据处理及可视化技术。为了应对复杂多变的数据分析需求，系统应提供灵活的轻量级数据处理技术，针对数据加工工程，提供网络（Web）端拖拽式、可视化的数据融合加工服务，简洁易用，帮助用户摆脱数据

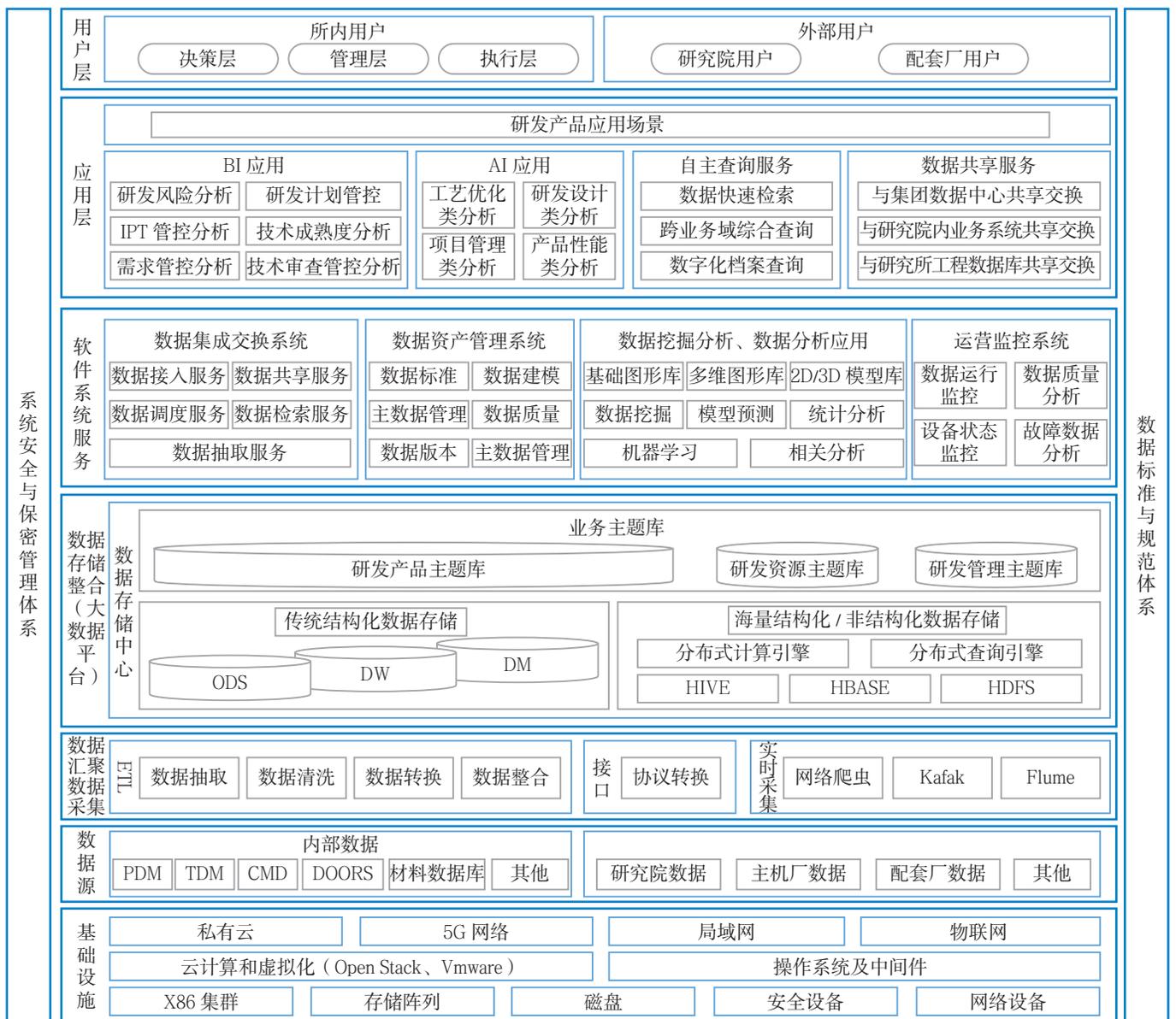
库结构化查询语言 (SQL) 脚本、数据库技术 (ETL) 工具等繁琐、技术难度高、难维护的数据处理方式, 提高工作效率。同时, 支持通过前台配置的方式实现不同数据库的迁移与同步。系统应提供自助式数据处理及数据可视化组件技术, 基于零代码快速实现展示数据的预处理, 通过拖拉拽方式可以定义各类决策分析看板, 可以快速响应产品数据

库数据应用方面存在的灵活多变的需求。

五是全面的国产化适配能力。考虑到产品数据库对国产化及安全保密的要求, 系统支持麒麟、中科方德等国产化服务器操作系统, 支持东方通等主流商用应用服务器软件, 支持达梦、人大金仓等国产数据库系统, 支持主流国产中央处理器 (CPU) 和兼容主流国产操作系统。

产品数据库系统总体架构

结合大型军工企业工程数据库系统、数据中台、数据中心实施经验, 产品数据库系统总体架构采用目前主流的关系库与大数据混合架构的方式。数据库系统功能主要由基础设施层和系统软件层提供物理软硬件支撑, 数据从数据源层被采集到数据存储层, 同时利用数据服务层的数据管理服务、数据应用服务和数



产品数据库系统总体架构

据共享服务处理之后，在数据应用层得到应用。

总体架构包括基础设施层、数据源层、数据采集层、数据存储层、软件工具层、数据应用层、数据标准与规范体系、系统安全与保密管理体系。

其中，基础设施层是产品数据库系统运行的基础设施，包括机房运行环境，以及计算、存储、网络、安全等设备，是产品数据库系统运行所需要的物理支撑。

数据源层是产品数据库系统数据的来源。内部数据源包括PDM、TDM、CDM、DOORS等；外部数据源包括集团各主机厂所、成附件厂、飞机方、部队方等。数据采集层的作用是使用工具和技术手段从数据源采集不同类型数据传输到数据存储层。

数据采集层主要包括3类场景：对小批量的结构化数据，可以通过ETL、网络服务接口等方式从业务系统定期获取；结构化数据通过Sqoop工具采集，非结构化数据通过Flume工具采集；设备采集的实时数据通过Kafka工具进行采集。

数据存储层就是对采集的所有数据进行存储。存储包括关系数据库存储和大数据存储：关系数据库存储对象包括的元数据、主数据、业务数据及主题分析数据；大数据存储对象主要包括大量的非结构化数据以及对于采集存储时效要求极高的设备物联网数据。

软件工具层为存储区的数据提供管理和应用服务，支撑数据仓库的正常有序运转和数据综合分析，主要包含数据版本比对数据质量管理、元数据管理、数据的对外接口管理，以及数据全文检索、数

据追溯服务、数据共享服务、数据可视化服务、数据挖掘服务、场景定制服务等功能。

数据应用层主要基于发动机研发业务应用实际需求，提供商业智能（BI）分析、人工智能（AI）分析、自主查询服务及数据共享交换等多形式应用，满足产品数据库系统中需求、设计、试验、实物、外场等产品数据的分析展示场景。

为实现统一数据源，产品数据库系统需建立一套顶层数据标准，建立并落实信息分类编码标准体系和信息分类编码管理办法，保障信息分类编码管理工作的有效开展，实现产品数据的统一规范管理。产品数据库系统整体系统架构构建、数据集成及项目实施过程，须严格遵守系统部署所在单位的各项规章制度，确保数据安全。

产品数据库系统功能架构

产品数据库系统基础平台建设包含数据汇聚、数据资产管理、 workflow引擎、数据挖掘分析、数据应用、运维监控等6大系统建设内容，主要提供数据存储、数据计算、数据分析、数据管理及运维监控等应用服务，有效地支撑产品数据库系统正常运转和数据管理工作。

数据汇聚。根据不同来源、存储类型、数据形态特点及数据使用要求，提供库表数据批量同步、线下数据导入、接口集成、文件上传、逻辑数据接入等多种数据归集方式，高效地实现多源异构、分散数据的统一归集。

数据资产管理。产品数据库系统应支持关系型数据库、HIVE库的前台建模，可由用户根据需要进行

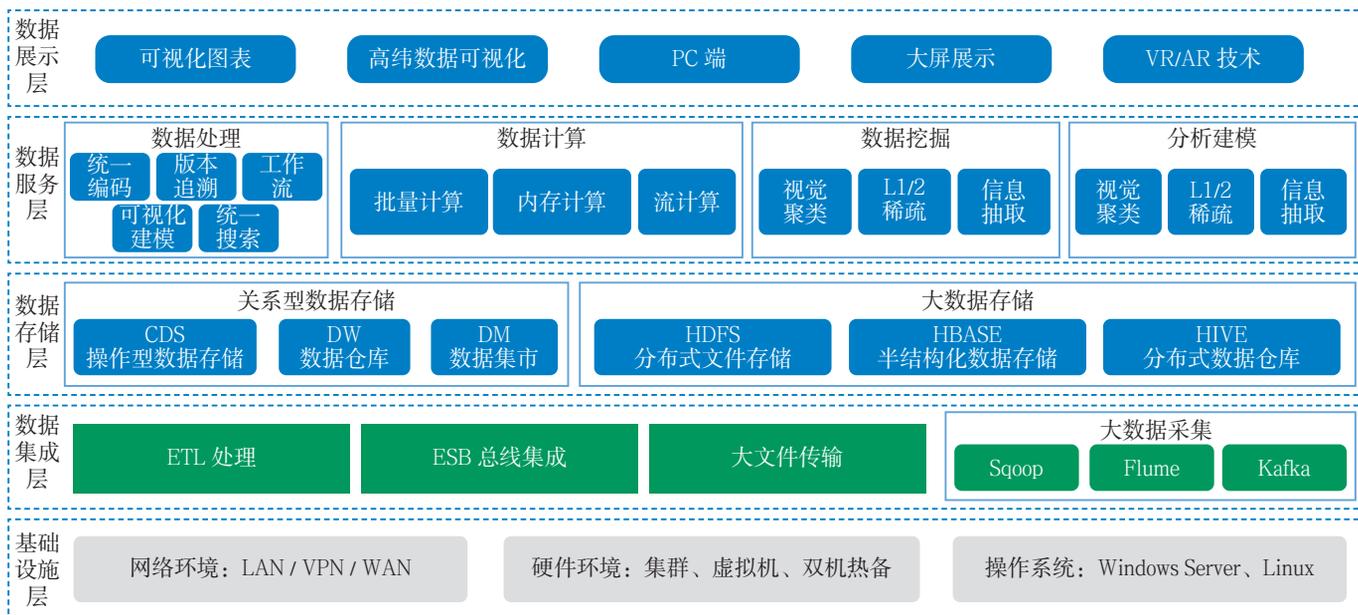
自定义，支持业务库模型抽取的方式进行数据的注册，可实现企业数据资产的集中管理，同时支持用户根据业务规则自定义数据质量规则，系统可按照用户设定的质量规则对数据进行前置质量规则校验以及周期性稽查，自动发现数据问题。生成质量报告，还可设定阈值当数据质量过低时进行预警，在第一时间发现质量问题。

工作流引擎。针对产品数据库的维护、订阅和查看等业务提供轻量级的流程审批业务，针对不同人员分别提供不同的数据业务审批流程。同时可以灵活设定流程，控制流程的启动停止。

数据挖掘分析。搭建一体化大数据挖掘分析平台，有效使用数据分析方法指导企业运营与科学决策，加快企业实现从业务驱动到数据驱动的转型进程。系统基于大数据的框架，覆盖数据汇聚、处理、建模、可视化的数据分析全过程。

数据应用。产品数据库系统应提供数据检索、即席查询、在线分析处理、报表统计、产品数据包生成等数据应用，通过搭建一体化数据应用分析平台，帮助用户快速完成数据展示操作，能够随时更改观察数据的维度、指标，将数据以丰富的图表方式，进行迅速、直观的表达，同时具备联动、钻取、链接等交互操作，满足不同层面对数据应用的需求。

运维监控。产品数据库系统应提供运维管理功能，监控整个集群的健康状态、软硬件的运行状态及系统应用情况。当集群中存在运行任务时，需要对任务的参数进行控制，保证整个集群能够正常、高效



产品数据库系统技术架构

地完成任 务；当出现异常时，需要分析异常，调整参数，以确保集群的高效运行。

产品数据库系统技术架构

产品数据库系统技术架构包括5个部分，分别为基础设施层、数据采集层、存储层、服务层、展示层，每部分采用相应的技术支撑功能实现。

基础设施层。基础设施层应用技术包括网络环境（LAN/VPN/WAN等）、硬件环境（集群技术、虚拟机、双机热备技术等）、操作系统（Windows Server、Linux等）中使用的技术。

数据集成层。为保证数据源系统中的各类数据能够快速、稳定地接入到ODS缓冲区，针对不同类型的数 据制定数据采集技术。Service API接口针对PDM、TDM等系统数据采用接口抽取；对于大批量数据使用ETL作为定时数据抽取工具。对于非结构化数据平台以及外部

系统等非结构化数据，可直接采用Flume、Kafka等工具进行数据接入存储，结构化数据存储与分布式数据库数据迁移时可采用Sqoop进行实时数据同步抽取。

数据存储层。为支持快速的数据接入、导出能力，并支持快速数据检索及质量检测，产品数据库选择关系型数据库存储模式。同时也基于分布式系统基础架构Hadoop构建分布式数据存储，以用于存储的扩展及进行基于大数据的决策分析。ODS数据缓冲区、数据仓库及数据集市之间的数据同步，采用ETL进行同步，与Hadoop存储环境基于Sqoop进行数据交换。

数据服务层。为产品数据库系统在数据管理、数据计算、数据挖掘、分析建模等数据服务方面提供技术支持。数据管理主要采用自定义建模、全文检索等技术，数据计算采用处理框架，内含批量计算、内存计算、流计算，提供大数据处理能力。

数据挖掘及分析建模提供对应的大数据算法及模型，为决策层的数据分析及态势预测提供支撑。

数据展示层。分析结果最终通过可视化分析工具及各种可视化展示技术快速地发布到计算机客户端及大屏等设备。

结束语

航空发动机产品数据的管理和综合应用是一个持续建设、完善的过程，需要长期的积累和坚持，尤其需要架构先进、稳定的数据库系统作为工具支持。基于稳定的数据库架构，持续汇聚、存储产品数据，根据发动机机理开发确定性模型，开展全过程跨领域产品数据的协同应用，为发动机自主研发能力和产品研制质量的提升提供数据支撑。 **航空动力**

（何为，中国航发动 力所，高级工程师，从事发动机数据管理和系统工程方法应用）