

信息系统架构发展展望——以国家自然科学基金委员会信息系统为例

姚畅, 郝艳妮, 彭升辉, 牛志昂, 谢勇, 赵世振

引用本文

姚畅, 郝艳妮, 彭升辉, 牛志昂, 谢勇, 赵世振. 信息系统架构发展展望——以国家自然科学基金委员会信息系统为例[J]. 计算机科学, 2025, 52(4): 14-20.

YAO Chang, HAO Yanni, PENG Shenghui, NIU Zhiang, XIE Yong, ZHAO Shizhen. Prospects for the Development of Information System Architecture—Taking the National Natural Science Foundation's Information System for Example [J]. Computer Science, 2025, 52(4): 14-20.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于可编辑医疗联盟链的数据安全管理方案](#)

Data Security Management Scheme Based on Editable Medical Consortium Chain
计算机科学, 2024, 51(6A): 240400056-8. <https://doi.org/10.11896/jsjcx.240400056>

[基于Transformer的司法文书命名实体识别方法](#)

Named Entity Recognition Approach of Judicial Documents Based on Transformer
计算机科学, 2024, 51(6A): 230500164-9. <https://doi.org/10.11896/jsjcx.230500164>

[基于微服务的预分配额度限流设计研究](#)

Pre-allocated Capacity Quota Limiting System Based on Microservice
计算机科学, 2024, 51(6): 346-353. <https://doi.org/10.11896/jsjcx.231100125>

[跨域数据管理](#)

Cross-domain Data Management
计算机科学, 2024, 51(1): 4-12. <https://doi.org/10.11896/jsjcx.yg20240102>

[基于数据库表的微服务拆分方法](#)

Microservice Splitting Approach Based on Database Table
计算机科学, 2023, 50(11A): 230200102-7. <https://doi.org/10.11896/jsjcx.230200102>

信息系统架构发展展望

——以国家自然科学基金委员会信息系统为例

姚畅¹ 郝艳妮¹ 彭升辉¹ 牛志昂¹ 谢勇² 赵世振³

1 国家自然科学基金委员会信息中心 北京 100085

2 南京邮电大学计算机学院 南京 210023

3 上海交通大学约翰·霍普克罗夫特计算机科学中心 上海 200240

(yaochang@nsfc.gov.cn)

摘要 近年来随着互联网、大数据、人工智能等技术的广泛普及与迅猛发展,科研管理正逐步从传统模式转向数据驱动、信息化、智能化的新范式。这一变革使得传统信息系统架构越来越难以满足海量数据交换、数据共享、数据安全等方面的需求。文中以国家自然科学基金委员会(以下简称“自然科学基金委”)的信息系统架构为例,探讨其发展演进方向。面对新时期科学基金改革的推进,自然科学基金委现有信息系统在网络和数据安全保障、智能化服务水平提升以及系统开发与管理效能优化等方面面临诸多挑战。首先介绍了两种传统信息系统架构在自然科学基金委的应用现状,接着,分别从业务量和业务系统数量增长、信创替代、智能服务和数据管理4个方面对现有信息系统面临的挑战进行了深入分析。最后,从数据管理和微服务两个维度给出了对信息系统架构后续发展的思考。

关键词: 信息化;信息系统架构;数据中台;数据管理;微服务

中图分类号 TP315

Prospects for the Development of Information System Architecture

— Taking the National Natural Science Foundation's Information System for Example

YAO Chang¹, HAO Yanni¹, PENG Shenghui¹, NIU Zhiang¹, XIE Yong² and ZHAO Shizhen³

1 Information Center, National Natural Science Foundation of China, Beijing 100085, China

2 School of Computer Science, Nanjing University of Posts and Telecommunications, Nanjing 210023, China

3 John Hopcroft Center for Computer Science, Shanghai Jiao Tong University, Shanghai 200240, China

Abstract In recent years, with the widespread adoption and rapid development of technologies such as the Internet, big data and artificial intelligence, scientific research management is gradually shifting from traditional models to new paradigms driven by data, informatization and intelligence. This transformation has made traditional information system architectures increasingly inadequate to cope with the demands of massive data exchange, data sharing and data security. This paper takes the information system architecture of the National Natural Science Foundation of China (hereinafter referred to as “the NSFC”) as an example to explore its development and evolution directions. Faced with the advancement of science fund reforms in the new era, the existing information systems of the NSFC are facing numerous challenges in terms of network and data security, enhancement of intelligent service levels, and optimization of system development and management efficiency. This paper first introduces the current application status of two traditional information system architectures in the NSFC, and then conducts an in-depth analysis of the challenges faced by the existing information systems from four aspects: the growth of business volume and the number of business systems, the replacement of information technology with domestic innovation, intelligent services, and data management. Finally, it provides thoughts on the subsequent development of the information system architecture from two dimensions: data management and microservices.

Keywords Informationization, Information system architecture, Data-centric architecture, Data management, Microservice

1 背景与意义

随着互联网、大数据、人工智能等技术的飞速发展,信息系统所处理的数据量呈指数级增长,科研管理领域因此经历

了一场深刻的范式转变。通过深入分析海量科研数据,管理者能够更准确地预测科研发展趋势,优化资源配置。同时结合海量科研数据和大模型技术,还能够构建智能科研管理工具,实现管理自动化和服务效率的提升。然而,海量科研数据

的引入也给底层的信息系统架构带来了严峻的挑战。在传统的以业务为中心的信息系统架构中,不同的业务系统各自负责数据的存储与管理,引入了一个个数据孤岛,给数据的交换、同步、共享以及安全管理带来了不便。因此,构建一个以数据为中心的新型信息系统,对于推动科研资助管理的现代化发展具有重要意义。

事实上,在大数据时代,数据的高效整合、共享和利用已经在国内外各个领域成为发展趋势。国家图书馆为了释放海量数据资源的价值、降低维护成本和提升快速创新能力,探索基于数据中台、业务中台的国家图书馆中台架构,实现数据资源统一管理 and 应用快速创新,为智慧图书馆建设提供了保障^[1]。国家开放大学出版社构建数据中台,将出版企业的不同销售渠道统一管理,大幅降低了系统维护的人力和物力成本^[2]。亚马逊云科技面向企业级市场构建数据中台,根植于智能数据湖仓技术,核心功能涵盖数据采集、存储、计算、治理与服务五大环节,通过 API 标准接口无缝对接各类业务场景,目前已经被应用于教育、金融、医疗等多个领域,如麻省理工学院、美国国家航空航天局、摩根大通等。

本文以自然科学基金委信息系统为例,探索信息系统架构未来的发展方向。自然科学基金委作为支持基础研究和应用基础研究的主渠道,为我国基础研究高质量发展提供了重要支撑。其成立 38 年以来,在科学基金项目、数据开放共享和知识服务,以及智能化行政办公、信息安全、数据中心基础设施建设等信息化方面取得了显著成效,在促进信息化与科研活动、科研管理体系融合,充分挖掘人工智能和大数据等新一代信息技术对项目管理和日常行政办公的主动支

撑方面起到了重要作用^[3]。自 1992 年启动科学基金项目管理工作以来,随着业务需求的不断拓展,自然科学基金委的信息系统先后经历了小型机、客户端/服务器架构、浏览器/服务器架构的演进,建立了 20 余个信息系统,有效提升了“智能感知、精细管理、科学决策、高效服务”的数字化治理能力和便捷化服务能力,目前正逐步推进以数据的深度挖掘和智能融合应用为主要特征的智能化转型。

自然科学基金委现有信息系统在构建时主要采用了单体式架构和面向服务的架构(Service-Oriented Architecture, SOA)。随着业务系统越来越庞大,数据同步、交换、共享的需求越来越大,系统的整体性能和数据安全保障面临挑战。此外,为适应人工智能和大数据时代的发展,进一步向自动化与智能化转型,也需要数据和信息系统的支持。为了应对上述挑战,自然科学基金委的信息系统正逐渐从“以业务为中心”转变为“以数据为中心”,为提升科学基金管理和资助效能奠定基础。

2 自然科学基金委信息系统的建设现状

为支撑和保障科学基金管理工作的 efficient 运转,自然科学基金委信息中心(以下简称“信息中心”)持续应用信息化技术落实科学基金改革的具体任务,以实现规范简约、公平公正和科学智能的科学基金信息化的管理目标。通过二十多年的建设,自然科学基金委已建成了以科学基金项目管理为核心,以“国家自然科学基金委员会网站”“科学基金网络信息系统”“办公自动化平台”“科学基金大数据知识管理服务平台”和“自然科学基金委基础环境管理平台”为主的信息系统平台^[4-5]。平台整体架构如图 1 所示^[3]。

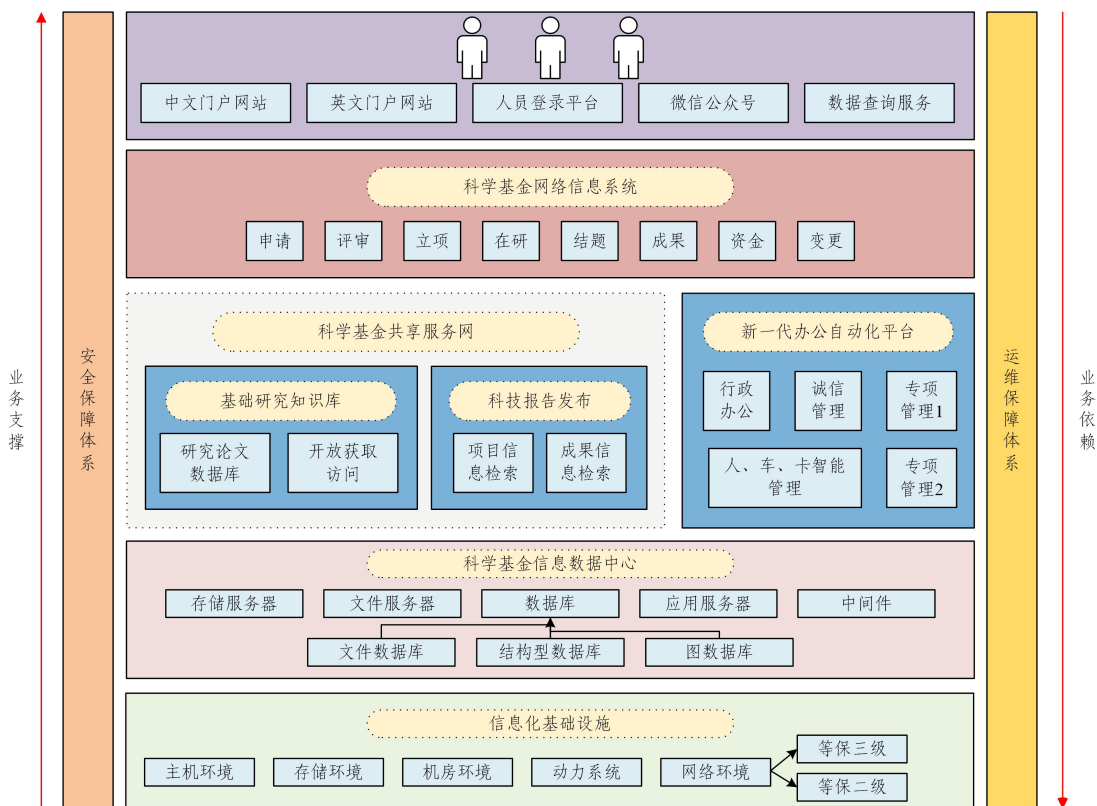


图 1 自然科学基金委信息系统的整体架构^[3]

Fig. 1 Overall architecture of the NSFC's information system^[3]

自然科学基金委信息系统平台涉及大小 25 个系统,分别

归属于官网服务、项目管理、行政办公、数据服务和基础保障

5个方面。由于系统建设的时间不一致和受限于建设时的技术水平,不同信息系统采用的系统架构和技术路线有所不同。整体而言,自然科学基金委现有信息系统中,14个系统采用了单体式架构(见图2),11个系统采用了SOA架构(见图3)。这两种架构均属于“以业务为中心”的架构。单体式架构易于开发、测试和部署,开发人员可以迅速开发出满足业务初期需求的系统版本。但随着业务范围的不断扩大,系统的代码耦合度越来越严重,系统的可维护性和可扩展性逐渐降低^[6]。SOA架构通过将系统的核心功能封装为一系列的标准服务,能够实现一定程度的复用。但是随着业务需求趋于多样化,SOA架构越来越难以在服务稳定集成和需求灵活适配之间找到平衡^[7]。随着自然科学基金委业务的发展,这种“以业务为中心”的架构逐渐带来了一系列挑战(具体见第3章),进而促使自然科学基金委的信息系统架构向数据中台+微服务这种“以数据为中心”的模式演进。

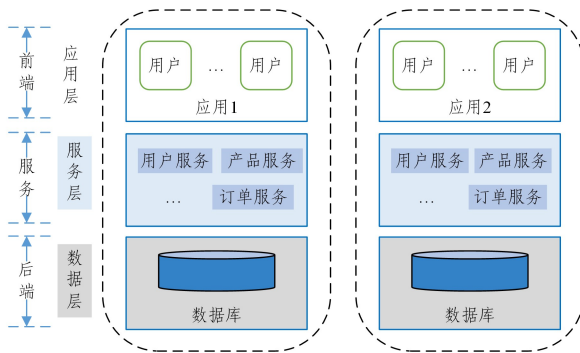


图2 单体式信息系统架构的层次结构

Fig. 2 Monolithic information system architecture

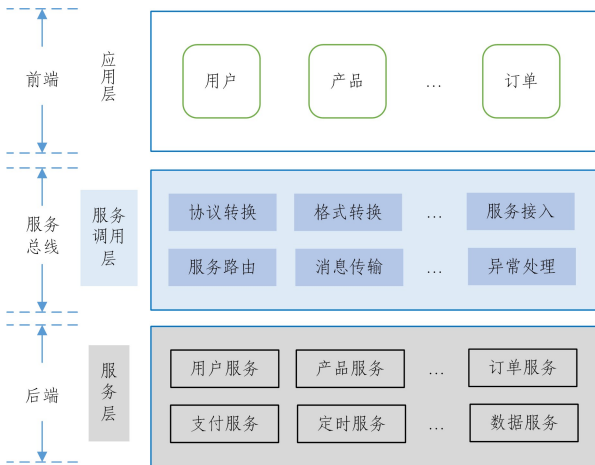


图3 SOA架构的层次结构

Fig. 3 Architecture of SOA system

2.1 单体式架构

单体式系统架构是早期信息系统开发的主流架构模式,其层次结构如图2所示。基于单体式系统架构的项目立项和开发过程大致如下:业务部门提出需求,信息中心根据需求编制建设任务书,由第三方进行造价评估,通过政府采购方式招标后,进入到设计、开发、测试和上线的项目建设周期中。在单体式架构中,每个新的信息系统上线运行都预示着一个新的独立系统建成。因此,采用单体式架构建设的信息系统

可能面对各自独立、互不相通的局面。概括而言,基于单体式系统架构进行系统开发将带来以下几个方面的弊端:

1)重复开发导致开发周期长和运维困难。不同信息系统处理的业务虽然不同,但很多功能模块之间存在重合之处,例如数据库访问、人员信息和大数据处理等。基于单体式系统架构造成的系统重复开发不仅浪费资源,而且会带来额外的系统运维负担。

2)系统之间交互协作困难。不同的信息系统是由不同的第三方在不同的时间分别开发而成,系统之间的数据交互接口难以统一,造成系统之间的交互和协作困难。另外,相同的数据在不同的信息系统中维护,如何保持这些数据的同步也是一个技术难点。虽然应用系统之间可通过数据库的数据导出、汇总和再导入方式进行同步,但是这种方法无法保证实时的数据同步,并且会随着信息系统数量的增加而愈发容易出错。

3)系统迭代优化困难。鉴于业务新增和信创替代等方面的需求,信息系统开发完成后可能还需要进行软硬件替代、功能新增和性能优化等。在单体式系统建设模式下,运维人员通常需要配合第三方来完成上述系统的升级操作。但是随着业务系统数量的不断增加,运维人员的工作量越来越大,如果没有及时完成系统的迭代更新,可能会降低系统的服务能力,甚至导致严重的安全和稳定问题。另外,单体式系统建设模式也不利于底层技术的长期积累。

2.2 面向服务的SOA架构

为了解决单体式系统架构带来的诸多问题,业界提出采用面向服务的SOA架构。该系统架构的层次结构如图3所示。该架构的核心是将系统的核心功能打包封装为一系列标准化的服务并对外提供调用接口,信息系统可以采用上述标准化的服务来进行构建,不同信息系统之间通过服务调用的方式实现功能复用。SOA架构最核心的价值在于松耦合的服务调用实现可复用业务,通过服务的编排能有效实现业务的快速响应和创新。SOA虽然能够有效解决单体式系统建设模式的前两个弊端,但在系统迭代优化和新信息系统开发时,该架构仍将面临以下几个方面的问题:

1)服务功能固化导致的自适应困难。当系统开发完成上线后,打包好的标准化服务进入稳定状态,即服务对应的功能固定下来,在系统后续的使用和运维期间,负责服务开发的第三方在服务增加或者服务性能优化方面的动力不足,从而导致服务所提供的功能固化和系统性能弱化。

2)高性能和可扩展服务的实现困难。采用SOA架构进行系统开发时,一种常见的场景是首先确定横跨多个信息系统的主业务流程和核心功能,然后对上述业务流程和核心功能进行拆分,进而封装和改造成标准化的服务。这是一种典型的“自顶向下”的系统开发模式。然而,由于不同的服务分属于不同的信息系统和开发团队,特别是进入到系统运维期以后,开发团队对于服务再开发和改造往往采取的是配合和协助的应对方式,因此难以实现高性能和可扩展的服务。

3)服务升级和改造困难。在新信息系统的开发过程中,当原有的标准化服务不可用或者无法满足系统的性能需求时,如果原服务提供者不能提供改造支持,则新信息系统的

开发可能需要重新实现该服务,这使得新信息系统的开发又重新陷入单体式开发模式的困境之中。另一种可能的情况是原服务提供者能够提供升级和改造支持,但受限于服务定义时对性能、通用性等方面的前瞻性考虑不足。为满足新信息系统的需求,需对原服务的数据模型、业务逻辑和流程等做出改动,这将带来较大的成本开销和项目风险。因此,为保证现有信息系统的稳定性,更多时候是选择维持现有服务不变。

2.3 信息系统架构的演进

由于单体式架构和 SOA 架构难以应对日益增长的业务需求,信息系统架构的转型逐渐成为大势所趋。下面介绍两个成功案例。

Netflix 在 2008 年经历的服务宕机事件推动了其信息系统架构的转型。通过构建数据中台,Netflix 实现了从单体式架构到微服务架构的转变,为高并发请求提供了强大的支撑,同时内部团队的开发模式也得到了改进。基础工具链团队提供工具支撑,解决了所有开发团队的通用问题,使得业务团队只需关注业务功能的实现和创新,提升了客户满意度的同时,也带来了运营效率的提升和成本的降低。

阿里巴巴在电商业务开展的早期采用了单体式和 SOA 架构。随着多个业务团队(如淘宝、1688、菜鸟、高德)纷纷涌现,每项业务都基于数据驱动,数据孤岛现象愈发严重,统一数据平台建设变得迫切。阿里巴巴于 2015 年提出“中台战略”,开始构建“大中台,小前台”的组织机制,拉开了数据中台建设的大幕。到 2018 年之后,阿里巴巴的中台建设逐渐成熟,业务方和技术方达成了良好的配合状态。阿里巴巴从 2018 年开始所有系统上云,到 2021 年实现了云上数据中台与业务伴生。

显而易见,Netflix 和阿里巴巴通过信息系统架构的转型,支撑了业务的进一步增长。然而,推动信息系统架构转型的关键并不是技术,而是业务的需求。本文着眼于自然科学基金委业务需求的变化,探讨其信息系统架构向数据中台和微服务架构演进的方案,旨在应对自然科学基金委业务发展面临的挑战,为科研管理的信息化发展提供解决方案。

3 自然科学基金委信息系统发展面临的挑战

3.1 业务量增长面临的挑战

根据信息系统的数据统计显示,自然科学基金项目的业务量呈现逐年增加的趋势。图 4 给出了近六年自然科学基金项目的申请数量和获批数量。相较于 2019 年,2024 年业务系统中评审专家数为 31.5 万人,同比增加 18.6%;项目负责人 40.70 万人,同比增加 3.8%;基础研究人员 111.34 万人,同比增加 16.2%;依托单位管理用户 4 230 个(含注册和注销),同比增加 39.4%。近年来,自然科学基金委积极推进各项业务改革,新增面向本科生和博士生的青年学生基础研究项目、国家杰出青年科学基金(延续资助)项目、面向全球的科学研究基金项目等 10 余项新业务,取消了面上项目连续两年申请未获资助后暂停一年申请的限制等。2024 年集中接收期间,自然科学基金委共受理全国 2 485 个依托单位提交的 38.5 万份申请,较 2023 年同期增长 26%。

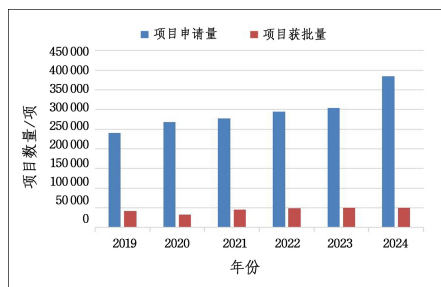


图 4 近六年自然科学基金项目申请量与获批量的变化趋势

Fig. 4 Trend in the number of applications and approvals of NSF projects in the past six years

业务类型的不断增加和业务量的急剧增长,给自然科学基金委的信息系统带来了性能优化和运维上的挑战。自然科学基金委信息系统对外服务的瓶颈在于数据库的访存。由于采用了单体式和 SOA 架构,面对不同的业务系统,自然科学基金委需要维护多个数据存储系统,导致运维难度大幅增加。此外,由于不同的访存系统存在差异,自然科学基金委也很难开发设计统一的优化方案来提升数据访存的性能。

3.2 信创替代面临的挑战

近几年,国际环境风云变幻,大国之间的博弈和竞争日趋激烈,关键信息基础设施高度依赖国外厂商,西方技术封锁和产品断供等事件推动信创产业快速发展,网络安全已上升到国家安全高度。只有加快构建信息技术应用创新产业,实现自主可控的软硬件环境,才能保证社会经济的正常运转和国家安全。自然科学基金委正积极推进以计算机终端、数据库和服务器等为代表的信息基础设施的信创替代。

然而,虽然国内信创产业取得了长足发展,但是相比国外同类型产品,信创软硬件产品及整体解决方案在性能、功能、易用性和兼容性等方面仍存在一定差距。业务量的增长,对自然科学基金委信息系统在响应速度、并发度、自适应性和算力等方面提出了更高要求。自然科学基金委只能通过信息系统架构的转型升级,来弥补国内信创产品与国外同类型产品之间的差距,以实现高性能、及时稳定的业务支持和智能服务。

3.3 智能服务面临的挑战

近年来,人工智能技术在多个行业和领域取得了突破性进展,特别是以 ChatGPT 为代表的大型语言模型提出后,其在诸多行业中得到了广泛的关注和应用。自然科学基金委积极推进人工智能和大数据等新一代信息技术在科学基金项目管理 and 行政办公自动化等过程中的应用和落地,探索建设人工智能辅助指派系统^[8]。随着科学基金业务量的急剧增长,评审专家指派的难度不断增加,采用人工智能技术有助于实现快速、精准和高效的智能评审专家指派^[9]。除此之外,生成式人工智能技术可实现图片、文字、代码和音视频的自动识别和生成等一系列智能化服务,可以有效降低信息处理的错误率和提升日常行政办公的工作效率等,这对于支撑自然科学基金管理和行政办公的协调和发展起到了重要作用。

然而,这些新型信息技术需要海量高质量数据的支撑。在单体式和 SOA 架构中,数据通常以不同的形式分散管理,难以以统一的接口、统一的格式对应用提供数据服务。此外,

基于人工智能技术生成的内容难以保障其准确性和可靠性,并且可能侵害个人权益、泄漏隐私数据。数据的分散管理难以提供数据的隐私和安全保障,因此自然科学基金委亟需构建新型信息系统架构,以应对新型信息技术带来的挑战。

3.4 数字化转型面临的挑战

自然科学基金委数字化转型面临着数据分散、数据管理能力偏弱和数据应用能力不强等挑战。数据作为关键生产要素,如何发挥其在自然科学基金委改革中的关键作用已成为自然科学基金委数字化转型需解决的关键问题之一^[10]。自然科学基金委通过38年的发展,积累了宝贵的科学基金大数据资产,但是受限于数据保存方式多样、系统架构老旧、信息系统间数据流通不畅和数据安全保护等方面的约束,现有科学基金数据在应用端的红利未能得到充分释放。因此,需优先对数据进行标准化和有效封装,以提升数据质量。在此基础上,构建以数据为中心的新型信息系统架构,有助于打通不同信息系统之间的壁垒,并建立有效的数据分类分级保护机制,实现数据在不同信息系统之间的安全可信流通。进一步通过对海量数据的挖掘和分析,促进科学基金数据在项目、行政管理、数据共享和知识服务等多个场景下的智能应用,培育基于科学基金数据的新应用和新服务,实现知识扩散、价值倍增,以充分提升科学基金数据在推动基础研究进步

和科技自立自强方面的价值。

4 信息系统架构发展展望

前文讲述了自然科学基金委近年来在科研管理中面临的4个方面的挑战,分别涉及业务规模、国产化支撑、数据安全和数据共享。应对该挑战的核心在于数据的统一化管理。作为先驱,以阿里巴巴为代表的各大互联网企业提出了数据中台^[11]的概念。数据中台通过将数据以及核心业务中心化,完全破除了应用之间的壁垒,使数据的管理、系统之间的交互更加便利。此后,阿里云、腾讯云、亚马逊云等纷纷推出各自的公有云中台产品,用于支撑金融、出行、电商、工业、医疗等行业。接着,各个高校,包括上海外国语大学^[12]、北京化工大学^[13]、苏州大学^[14]等,陆续建立私有的数据中台,用于服务高校的信息化管理。在此背景下,自然科学基金委也开始思考数据中台在科研管理方面的应用^[15]。然而,与大多数行业不同,自然科学基金委的科研管理对数据安全有更高级别的要求,因此难以采用公有云的中台方案;各高校自建的私有数据中台具有一定的参考价值,但自然科学基金委由于数据规模、数据服务的需求更大,因此对数据中台系统的要求更高。本章结合自然科学基金委在科研管理中面临的挑战,提出“以数据为中心”的信息系统架构设想,如图5所示。

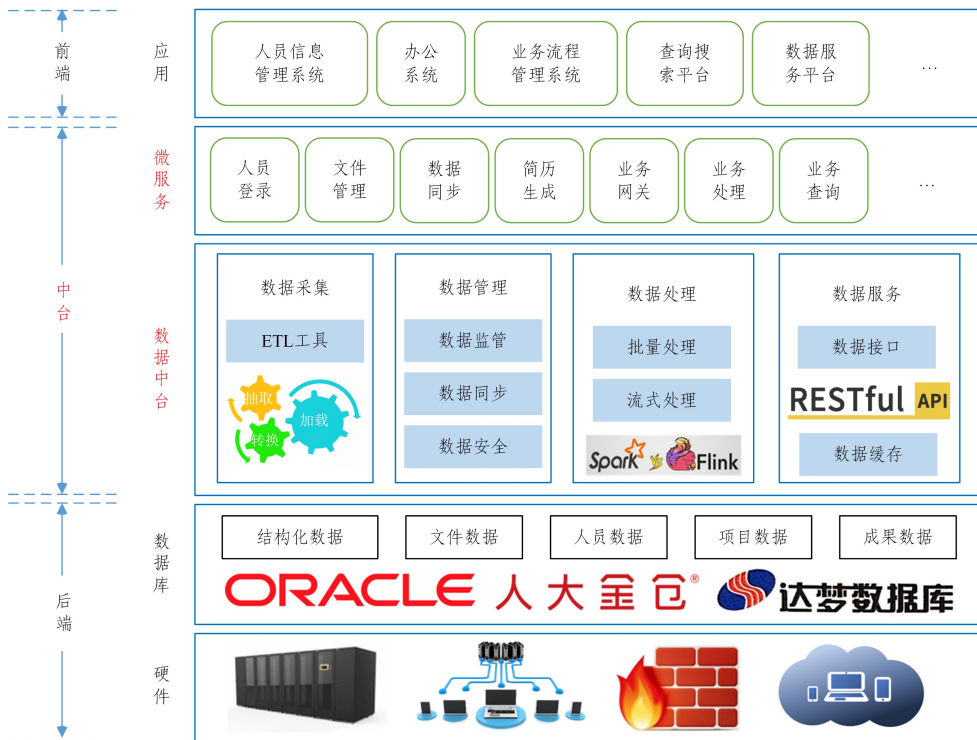


图5 基于中台的下一代信息系统架构

Fig. 5 Data-centric information system architecture

4.1 下一代信息系统的数据管理

下一代信息系统架构的核心是引入数据中台层,将数据规范统一管理,并将数据从之前各应用系统的属性重新规划成整体数据域,数据与应用系统解绑,以便数据充分共享与利用。对此架构中的数据中台层的规划是将前端业务开发与后端数据存储隔离开来,使得系统开发方可以专注于业务逻辑,无需关注底层的硬件与数据。数据中台层通过数据采集、

数据管理、数据服务和数据服务四大类服务,实现对自然科学基金委相关业务的支撑。

4.1.1 标准化的数据采集

目前,自然科学基金委各类信息系统中采集的多种数据,包括人员信息、项目申请、科研成果等,可能通过不同的方式重复采集进入不同的信息系统,例如科研成果数据既可以从合作出版商直接获取数据,也可以由科研人员直接手工录入。

不同数据源的数据格式不尽相同,给数据的存储、管理和使用带来了不便。为解决这一问题,下一代信息系统的数据中台专门设计了数据采集处理区,将原始结构化或非结构化的数据通过特征提取成符合表格数据模型或关系数据库的格式,典型的例子包括 XML 文档和 JSON 数据格式等。再采用特定规则和方法对提取后的数据进行筛选、补充、合并、格式转化、拆分和计算。最终将符合规范的数据加载到目标存储数据仓库中,供相关信息系统使用。

标准化的数据采集为自然科学基金委整合不同业务系统的数据资产奠定了基础,有助于打破数据孤岛,实现数据共享,解决自然科学基金委数字化转型面临的挑战。此外,数据标准化对人工智能技术的应用也有一定帮助。

4.1.2 统一的数据管理

经过采集和标准化处理后的数据,在数据中台中实现了逻辑层面的集中控制,给数据监管、数据同步和数据安全带来了便利。数据监管是当今数字化时代至关重要的一项任务,通过记录数据的产生、访问、更新、销毁等,清晰追踪数据的来源和生成过程,及时发现潜在问题。借助数据监管,可以更好地管理和保护数据,提高数据治理的效率和可信度。数据同步的目标是确保不同业务系统之间的数据保持一致性和实时性。在之前的单体式开发管理模式中,数据同步需要通过人工导出、导入的办法来进行,导致数据的实时性难以保证。基于数据中台架构,能够获取所有业务系统提供以及依赖的数据,并从中构建出一幅“数据依赖关系图”。根据关系图,可以实现数据自动化同步,进而确保数据同步的实时性。数据安全管理的目的,一是防止敏感数据的泄漏,二是保证数据的可用性。

通过数据中台对数据进行集中化管理,能够更容易地对数据进行分类分级,并通过权限控制、隐私计算技术等方式加强敏感数据的安全管理,这对解决智能服务挑战中可能面临的数据安全和隐私泄漏问题具有重要意义。此外,统一的数据管理使得信息中心运维团队能够专注于单个系统的维护,不仅能够降低运维的难度,还能促进技术的积累和迭代。

4.1.3 分布式的数据处理

数据中台通过集中存储和管理数据,能够为各类信息系统提供高质量的数据支持。数据处理通过从海量数据中提取有价值的洞察,来支持决策和业务发展。对非实时性、大数据量的数据采用批量处理方法进行分析,例如学科热点分析、基金项目的整体申请和资助情况分析等。对注重实时性的数据采用流式处理方法进行分析,快速捕捉和响应正在发生的事件和变化,例如网络安全检测、实时信息检索等。另外,随着 AI 技术的发展,大模型等工具也将应用到科学基金的相关管理中,智能化大模型计算等新质数据处理方式也将有一定的应用。

基于信创产品当前的性能表现和业务量增长,自然科学基金委的部分业务需要采用更多硬件设备进行分布式部署。高效的分布式数据处理系统有助于解决自然科学基金委面临的业务量增长和信创替代的挑战。

4.1.4 标准规范的数据服务

为确保数据中台的易用性,制定统一标准的接口是关键,

可采用目前较为成熟的 RESTful API 作为数据服务的接口。此接口基于 HTTP 协议,使用统一的 URL 结构和 HTTP 方法,由此,不同系统之间的数据交互将变得简单、一致。接口统一化能够有效解决自然科学基金委业务增长过程中面临的系统接口日益复杂的问题。

此外,数据缓存也可在数据服务中广泛应用,以减少数据服务的响应时间,提高系统的可伸缩性和并发性能,提升用户体验和系统整体性能。这也能有效弥补信创产品与国外同类产品之间的性能差异,加快自然科学基金委的信创替代进程。

4.2 下一代信息系统的微服务化管理

随着科学基金业务需求的不断增长,现有信息系统架构的弊端已有显现。为解决这些问题,下一代信息系统整体架构中可引入微服务。在微服务架构中,每个微服务都是一个独立的实体,负责单个业务,可以独立部署、扩容和升级。不同微服务可以采用不同框架,利用不同语言进行软件开发。通过将一个复杂的业务系统拆分成多个微服务,一方面能够减少系统开发的工作量,另一方面能够灵活地根据业务需求调整系统中不同微服务的部署策略,提升系统的端到端性能。

数据中台的引入为微服务架构进一步提供了便利。一方面,数据中台为微服务提供了后台的数据访问、存储功能,使得应用系统开发者可以专注于业务逻辑。另一方面,数据中台为微服务提供了一个以数据为中心的交互平台。不同的微服务只需要和数据中台利用统一接口交互,就可以实现微服务之间的交互协作。数据中台与微服务的具体关系如图 6 所示。

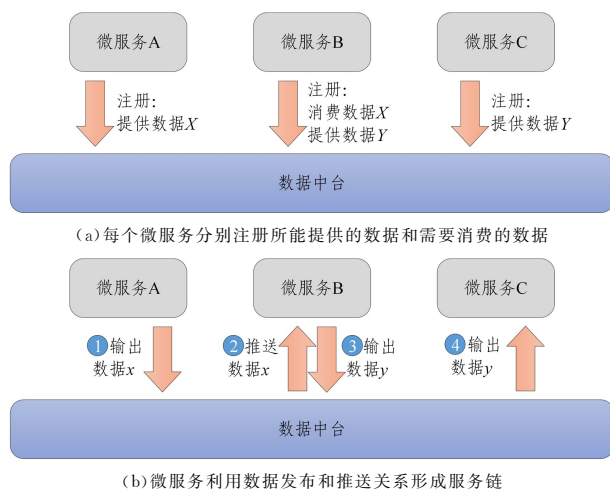


图 6 数据中台与微服务

Fig. 6 Data-centric architecture and microservice

所有数据的存储和访问由数据中台统一管理,所有应用系统都需要通过统一的数据接口服务来对数据进行访问或存储。每个应用的微服务既可以作为数据提供者向数据中台存储数据,又可以作为数据消费者从数据中台获取数据。如图 6 所示,微服务之间可以通过数据的提供和消费串联起来,构建出一个大的业务应用系统。例如,需要构建一个包含微服务 A、微服务 B、微服务 C 的服务链。首先,微服务 A 将其本身注册为其输出数据 X 的提供者,微服务 B 将其本身注册为数据 X 的消费者。然后,一旦微服务 A 产生了一个数据 X

的实例 x 并发布给数据中台,数据中台就会自动推送 x 到微服务 B,微服务 B 则可以根据 x 完成下一步的操作。微服务 B 和微服务 C 也可以通过类似的操作完成串联。

管理不同的微服务,核心是要确保不同微服务之间的数据一致性,因此每个微服务必须包含一个数据注册模块。在每个微服务的数据注册模块中,需要列出该微服务提供的数据以及消费的数据。数据中台在收到所有微服务的注册信息后,需要对每类数据的提供者和消费者进行校验。数据中台根据每类数据的提供者或消费者的属性,提供不同的服务接口,从而保障微服务之间的数据依赖关系不存在环路。

微服务的划分同样是一个难点。在微服务架构中,服务的粒度并非越小越好。一个应用如果划分出的微服务过多,服务链就会过长。由于微服务之间的每次交互都需要经过参数序列化、通信、参数反序列化,因此过长的服务链会大大增加系统的负载以及端到端的延迟。此外,每个微服务都有可能出现故障,服务链过长会大大提高整个应用的故障率,降低整个应用的稳定性。因此,业务的合理拆分也是微服务架构部署的关键之一。

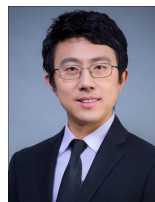
结束语 随着数据的爆炸式增长,以及信创替代和智能大数据服务的蓬勃发展,传统的单体式架构和 SOA 架构越来越难以满足科研管理对数据交换、数据共享、数据安全等方面的需求。本文提出基于数据中台和微服务的下一代信息系统架构的设计,通过对数据进行统一管理和微服务模式的引入,能够有效打通数据孤岛,实现数据安全共享,同时还能大幅降低业务系统开发、部署和运维的难度。最后,本文就管理模式的变化提出了后续发展的思考。

参考文献

- [1] ZHI Y Y. On the Application of Middle Platform in the National Library of China[J]. Library Tribune, 2021, 41(10): 76-82.
- [2] WANG J C. Research on the Construction of Data Platform for Publishing Companies in the Big Data Era: Taking the National Open University Press as an Example[J]. China Digital Publishing, 2024, 2(5): 70-75.
- [3] LI D, HAO Y N, PENG S H, et al. Information construction of the National Natural Science Foundation of China: today and intelligent development prospects[J]. Bulletin of National Natural Science Foundation of China, 2023, 37(2): 307-312.
- [4] YAO C, WANG X F, DU Y, et al. Overall scheme and key technologies: knowledge management and service platform based on NSFC big data[J]. Bulletin of National Natural Science Foundation of China, 2019, 33(1): 55-61.
- [5] LI D, YU X F, DU Y, et al. Status and prospects of the open sharing platform of the National Natural Science Foundation of China[J]. Bulletin of National Natural Science Foundation of China, 2021, 35(5): 808-814.
- [6] LI C X. Research Overview of Microservices Architecture[J]. Journal Software Guide, 2019, 18(8): 1-3, 7.
- [7] FENG Z Y, XU Y W, XUE X, et al. Review on the Development of Microservice Architecture[J]. Journal of Computer Research and Development, 2020, 57(5): 1103-1122.
- [8] JIANG H J, HAO Y N, WANG R, et al. The studies on the assigning reviewers for proposals received by National Natural Science Foundation of China[J]. Bulletin of National Natural Science Foundation of China, 2023, 37(5): 872-876.
- [9] LI D, HAO Y N, PENG S H, et al. Practice and exploration of artificial intelligence technology in sciencefund project management[J]. Bulletin of National Natural Science Foundation of China, 2022, 36(5): 790-797.
- [10] National Data Administration, Office of the Central Cyberspace Affairs Commission, Ministry of Science and Technology, et al. Data Element \times Three-Year Action Plan [EB/OL]. [2024-01-05]. http://www.cac.gov.cn/2024-01/05/c_1706119078060945.htm.
- [11] ZHONG H. The path to corporate IT architecture transformation: the strategic thinking and architectural practice of Alibaba's data-centric architecture[M]. China Machine Press, 2017.
- [12] CAI Y C, OUYANG J, YAN D. Research on Library Data Service Mode Based on the Concept of Data Middle Platform[J]. Library Journal, 2021, 40(11): 99-107, 63.
- [13] LIU S M, DU B, YUN G G, et al. Construction and practice of business middle-platform based on data middle-platform and process platform [J]. Modern Electronics Technique, 2022, 45(20): 65-68.
- [14] LU J J. Research on university data service architecture and rapid application construction based on data platform[J]. Cyber Security And Data Governance, 2024, 43(2): 73-77.
- [15] HAO Y N, LI D, HAN L C, et al. Full-cycle Data Security Management Research and Consulting Practice for Data Middle-ware System — Take National Natural Science Foundation of China Data Management as An Example[J]. Bulletin of National Natural Science Foundation of China, 2024, 38(4): 696-702.



YAO Chang, born in 1980, Ph.D, associate research fellow. His main research interests include scientific research management informatization and system construction.



ZHAO Shizhen, born in 1988, Ph.D, associate professor, is a member of CCF (No. E7035M). His main research interests include network system architecture design and implementation.

(责任编辑:柯颖)