 DOI: https://doi.org/10.12349/iser.v6i4.5290

Construction of Unstructured Data Resource Management Platform in University

Cong Liu

Xi'an Medical University, Xi'an, Shaanxi, 710021, China

Abstract

As an educational and research institution, colleges and universities will produce a large amount of data in daily scientific research, education and teaching, administrative management and other aspects. Because data is scattered and lack of unified management, it is difficult to use data, so it is necessary to establish an unstructured data resource management platform for data management. Based on the actual situation of the school, this paper designs and implements an unstructured data resource management platform that meets the functional and performance requirements of the school from four levels: hardware environment, software configuration, management side, and user side. This platform eliminates data barriers, avoids resource waste, and provides a strong impetus for resource integration, strengthening governance, and future scientific research breakthroughs, teaching innovation, and management upgrading.

Keywords

Smart Campus; Unstructured Data; Data Islands; Hierarchical Authorization

高校非结构化数据资源管理平台的构建

刘聪

西安医学院，中国 · 陕西 西安 710021

摘　要

高校作为一个教学与科研机构，在日常的科学研究、教育教学、行政管理等方面会产生大量数据，由于数据分散且缺乏统一管理使得数据难以利用，因此需要建立非结构化数据资源管理平台进行数据管理。本文结合学校实际情况，分别从硬件环境、软件配置、管理端、用户端等四个层面进行设计与实现，搭建了满足学校功能、性能需求的非结构化数据资源管理平台，该平台消除了数据壁垒，避免了资源浪费，为资源整合、强化治理及日后的科研突破、教学革新和管理升级提供了强劲动力。

关键词

智慧校园；非结构化数据；数据孤岛；分级授权

1 引言

随着数字化转型的深入，高校在科研、教学和管理中产生的非结构化数据呈现爆发式增长，但其存储和管理仍面临诸多挑战：数据类型多样化，数据规模大，存储分散化，管理粗放，检索困难等导致“数据沉睡”，为消除数据壁垒，避免资源浪费，构建非结构化数据资源管理平台迫在眉睫，通过进行资源整合，资源强化治理，不仅能提升数据价值，还能推动高校数字化转型。未来，随着技术的演进，非结构化数据资源管理平台将推动高校从“数据存储”向“数据驱动”跨越，为科研突破、教学革新和管理升级提供持久动力 [1,2]。

【作者简介】刘聪（1995–），女，中国陕西西安人，硕士，助理工程师，从事高校智慧化建设研究。

2 需求调研

非结构化数据和结构化数据不同，非结构化数据是指视频、图片、文档、音频等不宜用数据库二维表来存储的数据，高校在日常的科研、教学、管理中会产生大量的此类数据，因此，构建高校非结构化数据资源管理平台十分必要，而平台如何构建、数据如何存储，前期调研是关键。通过前期调研来完善需求、理清思路，前期调研主要包含六个方面：数据类型和规模、业务需求、技术架构、数据安全和隐私保护、用户管理、基础设施 [3]。

数据类型和规模：高校在日常教学、科研、行政管理过程中会产生海量数据，包括视频、音频、图片、文档等，当前我校已汇聚到的非结构化数据资源总量为 27 万 +，存储容量达 9.93TB，因此平台构建需要注意存储容量，同时考虑到未来数据的持续增长，平台构建时需要支持横向扩展。

业务需求：前期调研需要收集各二级单位关于非结构化数据资源管理平台的各类需求，通过调研明确了平台需能够进行各类非结构性数据资源的上传、下载、删除、分享、搜索、浏览、资源定位、权限分层分级管理，同时要求图片能够进行人像查找、视频能够进行视频剪辑以及视频合成等。

技术架构：构建非结构化数据资源管理平台需要综合考虑存储、管理、安全、性能、可扩展性和合规性等多个维度。在高校中，数据的安全与合规是首要强调的因素，它涉及高校的科研成果和师生的隐私，因此，高校对数据做了防火墙设置，对平台进行监控，对进入平台的人进行身份认证，以确保数据的安全合规问题。高校数据存储量大，因此采用分布式对象存储，同时基于 Ceph 的分布式文件系统，能够满足高校的数据增长需求。平台基于 AI 算法进行图像识别，可实现数据快速批量查找。

数据安全和隐私保护：防火墙设置、平台监控、身份认证都能在一定程度上保护数据，平台可通过划分数据访问权限，进行数据分级分层，保证各二级单位的数据具有私密性。

用户和管理：平台需要有简单、易操作的 Web 界面、需要满足数据的高频访问和低频归档、响应速度快、功能丰富。

基础设施：根据学校自身情况进行存储容量的分配及网络带宽、服务器、虚拟化环境的配置等。

通过上述调研及实施方案，基本可以制定科学的非结构化数据资源管理平台建设方案，确保平台贴合学校实际需求。

3 架构设计

本文设计了一个非结构化数据资源管理平台的整体架构，总体上分为四个层次：硬件支撑、软件配置、管理设置、用户操作，整体架构如图 1 所示。



图 1 非结构化数据资源管理平台架构图

硬件环境：该层是构建非结构化数据资源管理平台的基础，需要服务器、存储介质、网络、负载等硬件支撑。由于学校资源丰富、数据量大、资源类型多样，数据传输和访问需要高宽带和低延迟，数据资源持续增长，因此网络架构和存储介质是很重要的一点；由于资源量大，需要元数据处理、检索、分析能力较强，因此对 CPU 和内存也有一定要求；同时为了避免单点故障，需考虑负载均衡和冗余设计；服务器长时间运行也需要稳定的供电和散热 [4]。

软件配置：该层可进行统一身份认证，进行校园门户集成，进行各业务系统的数据对接，需要对密码复杂度有要求，对安全等级有要求等。由于学校建设了“一网通办”平台来管理所有应用，因此非结构化数据资源管理平台也需要集成到学校统一门户，统一身份认证是集成的基础，非结构化数据资源管理平台需要进行过统一身份认证，为避免重复登录，要进行单点登录的对接；通过实时监控网络流量，拦截恶意攻击，同时启用防火墙，持续进行漏洞扫描以确保数据安全性；非结构化数据资源管理平台需要获取并分析课程视频信息，因此该平台需要和教务管理系统、数据中心、教学平台等进行课程数据的同步 [5]。

管理设置：对于管理员来说，非结构化数据资源管理平台的管理设置功能最重要。该平台可建立资源栏目、划分资源浏览和下载权限，分级分层展示、可进行数据的统计与分析。平台以栏目形式存在，资源以栏目创建和展示；平台根据不同角色来划分资源的浏览和下载权限以保证资源的安全性和私密性，资源的浏览权限只局限于本部门和本学院，下载权限只局限于该部门或该学院的一个人；资源对于不同等级的人展示不同，领导要尽可能的掌握全部数据，因此权限更大；同时，数据统计与分析功能可清晰直观地看到数据资源情况，可供领导层做出下一步决策。

用户操作：使用者通常要求平台简单好用，要求浏览检索时间短，要求平台功能丰富，因此该平台能够进行各类资源的上传、下载、删除、安全分享、搜索、浏览，能够进行人像查找、视频能够进行视频剪辑以及视频合成。

4 基础环境

基础环境包括该平台在构建过程中的软硬件要求，硬件方面需考虑服务器、存储、网络、运维、成本等要求，软件方面该平台需要与数据中心及各类业务系统进行对接，

需进行统一身份认证，同时需要将该平台集成到校园门户网站中。

根据实际部署环境，结合高校场景，该平台所使用的服务器包括 1 台 64G 的主服务器、6 台 8G 的转码服务器、1 台 64G 的数据库服务器、1 台 64G 的 AI 服务器、1 台 64G 的应用服务器和 1 台 16G 的维护服务器。结合我校数据需求，数据存储容量为 120TB，满足科研，教学需要。学校对网络进行全面升级，配备专业网络运维人员实时保障网络。为了师生能更好地使用该平台，将该平台集成到了学校的门户网站，通过统一身份认证来验证用户身份，确保资源仅师生可见，通过进行单点登录的对接，使师生既能安全获取资源又能避免重复登录。

通过构建一个高性能、易扩展、低成本的存储平台，为教学、科研及行政管理提供坚实的数据基础设施支持。

5 平台管理与应用

作为全校非结构化数据资源管理平台，平台允许学校新闻媒体中心、各部门、各学院、附属医院、研究所、师生个人等各类人员进行资源的分层上传和分层浏览。同时，为了使业务系统互联互通，教学资源闭环管理，资源管理平台打通了录播系统和教学管理平台，使得课程资源能够实现自动录制和推送，即：以数据中心为核心，数据中心通过自动获取教务管理系统的数据，将数据推送至资源管理平台、网络教学平台、班牌系统等进行统计、分析、使用，推送到网络教学平台的个人中心的课程资源，教师可分享转发至其他师生或教研室，也可将视频进行剪辑来参加教学评比相关赛事，学校领导可以通过大数据屏，直观、清晰地获取到学校的科研、教学、管理等相关情况，非结构化数据资源管理平台与其他业务系统关联图具体如图 3 所示。

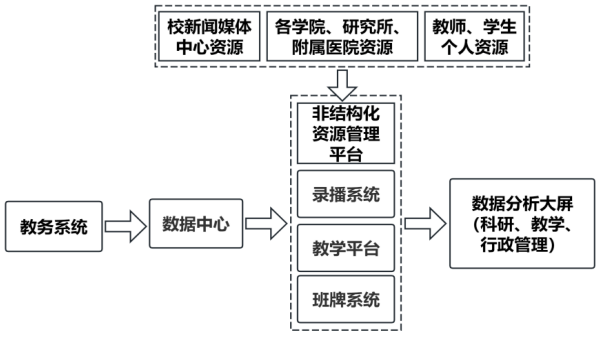

图 2 非结构化数据资源管理平台与其他业务系统关联图

当前，资源管理平台数据总量达 27.9 万 +，存储容量达 9.93TB，其中，视频资源 1.2 万 +，图片资源 26 万 +，其他资源共计 0.7 万 +，平台可进行各类资源的上传、下载、删除、分享、搜索、定位、人像查找、视频剪辑、视频合成等，部分展示如图 3- 图 5。



图 3 平台资源统计分析



图 4 视频剪辑与合成

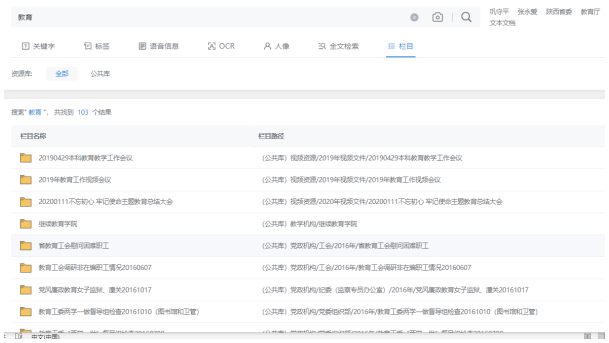

图 5 资源栏目查找

6 结语

本文设计了一个非结构化数据资源管理平台整体架构，从硬件层面、软件层面、管理层面、用户层面等进行阐述，明确了非结构化数据资源管理平台在构建过程中的各方面的问题。通过建设高质量非结构化数据资源库，能够更好挖掘科研、教学数据价值，进而为管理者的科学决策提供精准的数据支持。

参考文献

[1] 闫丽飞,褚宇宁,赵维伟,等.大规模非结构化数据资源快速存储方法研究[J].集成电路与嵌入式系统,2024,24(04):77-81.

[2] 李来存.基于物联网技术的信息系统数据存储系统[J].信息技术,2024,(05):120-126+132.

[3] 徐智,王岳,王欣.结合区块链的非结构化大数据云存储优化研究[J].计算机仿真,2021,38(07):304-307+354.

[4] 贾亚刚,安宁,齐瑞红.智慧校园环境下高校非结构化数据资源平台的构建[J].中国现代教育装备,2021,(11):1-3.

[5] 魏赫男.非结构化数据管理系统设计与开发[D].西安电子科技大学,2023.