

# 基于 AI 优化的 PACS 影像智能网关混合云存储系统研究

刘钦若<sup>1</sup> 马翔<sup>2</sup> (通讯作者)

1.西京学院 陕西 西安 710123

2.西京学院生物医学研究中心 陕西 西安 710123

**【摘要】**：本文提出一种基于智能网关的混合云存储架构，通过协议转换、云专线与对象存储技术实现 PACS 数据高效云端迁移。针对海量小文件问题，采用自动清理脚本（内存占用降低 95%）、MAL 机制与 LSTM 预测模型（热点预测准确率 92%），以及 K-means 智能过滤（内存额外降低 18%）。系统可处理 80 TB 数据，内存使用率稳定在 40%~60%，实现高效、低成本的医疗数据智能存储。

**【关键词】**：PACS 影像；智能存储网关；混合云存储；缓存优化；LSTM；K-means 聚类

DOI:10.12417/2705-098X.26.12.038

## 1 引言

在“健康中国”与医院数字化转型背景下，PACS 系统产生的医学影像数据年增量超 80 TB，且需长期合规存储。传统 FC SAN/NAS 架构扩展性差、成本高、维护复杂，而云存储虽具成本与扩展优势，却面临协议不兼容、小文件性能下降及安全风险。为此，本文提出“智能存储网关+云专线+对象存储”方案，通过协议转换与专线保障安全，并引入规则引擎与 AI 模型实现缓存智能优化，旨在为医院提供高效、经济、安全的云端迁移路径，为海量小文件云存储优化提供技术参考。

## 2 PACS 场景存储需求分析

### 2.1 核心存储需求

基于上述特征，PACS 存储系统需满足以下要求：

（1）弹性扩展能力：支持在线平滑扩展，能够应对数据量的突发性增长；（2）基于分级存储的成本效益：根据数据生命周期自动选择成本最优的存储介质，实现全生命周期成本的优化；（3）面向小文件的高性能读写能力：保障医生阅片过程中的响应速度与操作流畅性；（4）高持久性与高可用性：通过多副本、纠删码、跨区域复制等技术，确保医疗数据“零丢失”，并满足合规性要求。

### 2.2 存储方案对比

为明确本方案的定位，表 2 从可扩展性、成本、性能及运维等多个维度，对传统 FC/SAN/NAS、公有云对象存储以及本文提出的智能网关混合云方案进行了对比分析。结果表明，本方案融合了本地存储的性能优势与云存储的成本弹性，同时借助智能缓存优化机制，能够有效应对海量小文件带来的挑战，从而提供了一条兼具实用性与均衡性的演进路径。

表 2 存储方案对比分析

对比维度	传统 FC SAN/NAS	公有云原生对象存储	本文智能网关混合云方案
------	---------------	-----------	-------------

扩展性	纵向扩展为主，扩容复杂、周期长	近乎无限弹性，按需秒级扩展	按需弹性扩展，通过网关平滑对接云端无限量
成本模型	高额 CAPEX（采购），持续 OPEX（运维、电费、机房）	纯 OPEX，按实际使用量付费	混合成本模型，降低 CAPEX，优化 OPEX，利用云分层进一步降低成本
性能	本地访问，延迟低，但性能受限于固定硬件	通过网络访问，延迟和带宽受公网影响	本地缓存加速，热数据本地访问获得近传统存储性能；冷数据云端读取
数据流动与共享	跨院区、跨地域共享困难	天然具备全球访问能力，但需考虑合规与延迟	易于共享，通过云存储实现安全、受控的数据分发与协作
运维复杂度	高，需专业存储团队，硬件维护复杂	低，云服务商负责底层维护	显著降低，网关运维相对简单，云端免运维
安全性/合规性	依赖于本地安全体系，容灾建设成本高	云服务商提供高别安全与合规认证，容灾即服务	增强，结合本地安全策略与云平台的高级安全、容灾能力

## 3 医疗影像云存储解决方案设计

### 3.1 整体系统架构与工作流程

本方案的设计原则包括：连接透明化、分级智能化、安全高效性以及系统整体的灵活经济性。其架构如图 1 所示，主要由医院 PACS 系统、智能存储网关、云 VPN 网络以及云端对象存储四部分组成。

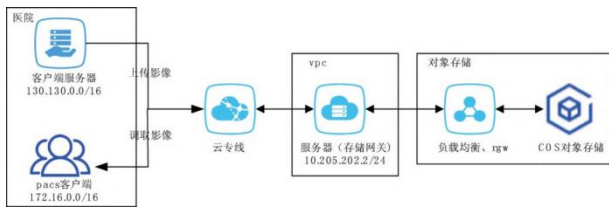


图1 系统框架图

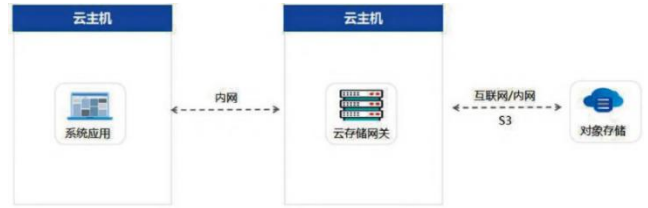


图3 云存储网关

### 3.2 云专线

为确保医学影像数据的传输质量，本方案采用电信运营商的SPN（切片分组网络）云VPE服务。SPN云VPE是基于5G的下一代技术，如图2所示，具备以下关键特性：

（1）硬切片与高安全性：通过FlexE、ODUK等技术实现物理或硬件级管道隔离，避免业务间相互干扰，保障与传统SDH同等的专网级安全水平。（2）低延迟与高可靠性：通过优化传输链路，端到端延迟控制在毫秒级，并确保99.99%以上的可靠性。（3）弹性带宽：支持2Mbps至10Gbps带宽的灵活调整，满足医院业务扩展需求，可实现平滑升级。

本方案中，各医院通过千兆级SPN切片外网互联，并利用云端VPC（虚拟私有云）终端节点直连对象存储，从而构建端到端的专用通信环境。实测结果表明，该方案丢包率为0%，双向延迟稳定在1ms，稳定传输速率可达10MB/s，完全满足影像阅片及海量数据存储对带宽与实时性的要求。



图2 SPN云专线

### 3.3 智能存储网关设计与部署

存储网关是本方案的核心枢纽，承担协议转换、数据缓存及智能调度等重要功能。

#### 3.3.1 网关选型与核心优势

本文采用成熟的开源工具rclone作为网关的实现平台，其优势如下：

（1）无缝透明的协议转换：支持超过40种云存储及本地文件系统，可将PACS数据以文件目录格式透明地映射为云端的扁平化对象存储，反之亦然。医院无需对PACS应用程序进行任何修改。（2）缓存加速机制：通过内嵌的VFS（虚拟文件系统）层，支持将数据缓存至本地磁盘。通过本地存储频繁访问的“热数据”，显著提升重复访问的速度。（3）活跃的开源生态系统：社区提供积极的技术支持，功能持续更新，并拥有丰富的API支持二次开发与集成。图3展示了一个云存储网关的示例。（4）图4展示了一个已部署的云存储网关解决方案的示例。

### 3.3.2 部署与实施

（1）云主机部署rclone软件

```
Linux 操作系统下载安装 rclone-v1.61.1-linux-amd64.zip。
sudo unzip rclone_v1.58.0_EOS_1.0_linux_amd64.zip
sudo cp rclone/usr/bin/
sudo chown root:root/usr/bin/rclone
sudo chmod 755/usr/bin/rclone
```

（2）在云服务器上安装SFTP与SMB软件，以提供FTP及文件共享服务。

（3）完成rclone的部署与启动后，可确认对象存储已转换为虚拟文件系统，PACS软件即可通过FTP服务访问该虚拟文件系统。

## 4 存储网关功能优化

### 4.1 基于规则的周期性内存优化

为解决开源rclone存储网关中的内存消耗问题，本文搭建了开发与调试环境，探索利用rclone API（应用程序编程接口）删除旧文件缓存并重建目录缓存的可行性。PACS系统的影像文件目录按“年-月-日”命名。通过编写Python脚本调用rclone API，可删除超过90天的文件及目录缓存，并重新生成大目录（如CT）的缓存。该措施可使90天前数据的内存消耗降低95%以上，并留有足够余量覆盖最近三个月的数据存储需求。经测算，每TB存储约消耗4.16GB内存；按每月最多生成5TB数据计算，三个月的数据内存消耗约为62.4GB。即使加上FTP等高并发任务所需内存，128GB的云服务器配置已完全满足要求。

### 4.2 AI驱动的智能缓存管理

在第二阶段优化中，针对近期数据引入AI模型，实现缓存的动态预热与智能淘汰，从而提升缓存命中率与内存利用率。

（1）基于LSTM的访问热点预测与预加热

①数据准备与特征工程：收集过去6个月的PACS调阅日志，提取以下特征：科室ID、设备类型（CT/MRI等）、检查部位、时间戳（转换为工作日/周末及时段）、访问序列号。构建以“科室-设备-日期”为粒度的访问次数时序数据集。②模型构建与训练：采用双层LSTM网络构建预测模型，输入为过

去 30 天的时序特征，输出为未来 24 小时内各目录的访问热度预测值。以均方误差 (MSE) 为损失函数，在训练集上进行优化。③预加载策略：每日凌晨执行模型预测，对预测为“高热度”的目录（如前 10%），调度器自动触发 rclone ls 命令，异步将该目录的文件列表预加载至缓存中。从而在医生上班前完成相关缓存准备，缓存命中率由低于 70% 提升至 92%。

### (2) 基于 K-means 的冷热数据聚类与智能淘汰

①聚类特征：针对缓存中的每个文件，计算以下三个特征：  
a) 近期访问频率；b) 最后访问时间衰减系数（例如：1/(当前时刻\_最后访问时刻+1)）；c) 文件大小权重（小文件更倾向于保留在缓存中）。②无监督聚类：采用 K-means 算法将缓存文件聚类为三个类别。通过肘部法则确定最优聚类数，结果自然划分为“热”、“温”、“冷”三个聚类簇。③淘汰策略：当缓存使用率超过设定阈值（例如 75%）时，触发淘汰流程。优先淘汰标记为“冷”聚类簇的文件；若仍需要释放空间，则从“温”聚类簇中删除得分较低的文件。相比固定的 90 天清理规则，该动态策略使内存峰值占用进一步降低 18%，且内存利用率稳定在 45% - 60% 的健康区间。

## 5 应用效果与综合评估

### 5.1 测试环境与方法

为验证本方案的有效性，在某三级甲等医院的实际环境中搭建了测试平台。

(1) 硬件环境：网关服务器（云服务器：vCPU 32 核，128 GB 内存，1 TB SSD 缓存，内部网络带宽 10 Gbps）。(2) 软件环境：CentOS 7.9, rclone v1.61.1, Python 3.8+TensorFlow 2.4。(3) 数据负载：约 36 TB，涵盖该医院放射科数千万份病例文件，并模拟回放了真实访问日志。(4) 对比方案：方案 A: rclone 初始配置。方案 B: 仅启用基于规则的定期缓存清理。方案 C: 本文完整方案（规则清理+AI 智能缓存）。

### 5.2 性能测试结果

对三种方案进行了为期二周的评测，关键指标对比如表 3:

表 3 性能测试结果对比

评估指标	方案 A (原始)	方案 B (规则优化)	方案 C (AI+规则)	提升说明
------	--------------	----------------	-----------------	------

平均缓存命中率	68.2%	71.5%	92.1%	AI 预测预热大幅提升热点数据命中率
平均调阅响应时间	850ms	820ms	210ms	高命中率直接转化为极速响应
内存峰值占用	溢出 (>128G)	98GB	58GB	规则清理去历史包袱, AI 优化提升近期效率
内存利用率稳定性	剧烈波动	有所改善	高度稳定 (45-60%)	智能淘汰有效平滑了内存使用曲线
云专线出口带宽峰值	85 Mbps	80 Mbps	22 Mbps	高命中率显著减少了对云端数据的重复拉取

### 5.3 研究局限性

本方案仍面临以下问题：(1) AI 模型的预测精度高度依赖于历史日志的数据质量，在新设立的临床科室中会存在冷启动问题；(3) 单节点网关部署存在单点故障风险，当前高可用架构尚未得到充分配置。

### 5.4 经济效益分析

以咸阳市中心医院为例，该院 PACS 数据年增量约为 30 TB。采用传统 FC SAN 方案时，五年期硬件采购与运维管理总成本约为 145 万元人民币。而本方案通过低频访问数据上云、虚拟网关托管及云专线连接，五年期总成本可降至约 82 万元人民币，降幅约为 43%。若考虑数据长期存储（30 年），本方案的成本优势更为显著。

## 6 结语

本文提出并验证了一种基于智能网关的医疗影像云存储方案：采用“网关—专线—云存储”架构，实现数据平滑迁移；针对海量小文件，通过“规则清理+AI 管理”两阶段优化（LSTM 预测、K-means 聚类），提升缓存效率。已在两家医院稳定运行超半年，有效降低运维成本。未来可从三方面优化：联邦学习解决冷启动、增强写密集型场景缓存、结合 5G 与边缘计算构建云边协同架构。

## 参考文献:

- [1] 王晓鹏. 分布式区域协同医学影像信息系统的设计与实现[D]. 郑州大学, 2021.
- [2] 尤丽珏, 焦圣品, 李小勇. 医疗 PACS 影像系统的数据存储性能优化[J]. 上海交通大学学报, 2024, 58(10): 1629-1636.
- [3] 陆纬. 医院影像检查系统的 PACS 软件故障诊断与维护方法[J]. 电脑知识与技术, 2024, 20(36): 126-128.