中数国科日志收集与分析系统
产品白皮书

（中数国科集团）

【中数国科】

■ 文档编号	■ 密　　级
■ 版本编号	■ 日　　期
■ 撰 写 人	■ 批 准 人

■ 版权声明

目录

一、前言

网络安全的建设在我国一直都是党和政府关注的重点，更是国家战略。提升网络安全要在基础网络防护中加强建设。基础网络安全建设离不开防火墙、网闸、入侵检测、防御检测、防病毒网关、Web 应用防护系统等各类网络安全产品。

在国内外网络安全环境日益严峻的形势下，面对各种网络攻击的风险与威胁，网络系统的日志正是这些网络行为的具象化表达，是事中记录、事后溯源分析与总结的重要手段和方法，同样也是国家法律法规与标准的要求。

2017 年 6 月 1 日起施行的《中华人民共和国网络安全法》，第三章第二十一节的第三条规定：采取监测、记录网络运行状态、网络安全事件的技术措施，并按照规定留存相关的网络日志不少于六个月。

GB/T 22239-2019《信息安全技术 网络安全等级保护基本要求》中二级及以上系统对于日志审计的要求如下：

- 对重要的用户行为和重要安全事件进行审计;应对审计记录进行保护，定期备份，避免受到未预期的删除、修改或覆盖等;
- 根据安全审计策略对审计记录进行存储、管理和查询等；
- 应对分散在各个设备上的审计数据进行收集汇总和集中分析，并保证审计记录的留存时间符合法律法规要求。

各种网络安全产品、操作系统、应用软件等，往往来源于不同的产品厂商，为用户提供不同的管理界面以及多种多样的数据交互接口。每种产品设备类型的日志格式也不尽相同，各有各的表达，即使是表达同一件事情，也都有各自的表达方式。这些日志形成了一种数量庞大、种类繁多、形式各异、无法理解的“日志信息孤岛”局面。

因此，将网络系统中各类型的日志进行汇总、审计和检查可以系统全面的了解网络运

行状况，从而解决各类安全产品之间的“日志信息孤岛”问题，还可以满足各企业对信息安全系统的审计合规需求。

二、产品概述

中数国科日志收集与分析系统（以下简称“日志审计系统”）是一套以日志收集为基础，分类审计为核心，利用独立自主研发的解析引擎对日志进行解析，并通过规则和算法关联生成可被理解事件的信息收集与分析系统。支持各类日志不同业务场景下的审计、全文检索、异常告警、报表导出等功能。另外，通过可视化图表辅助，帮助用户从全局视角进行网络安全审计与检查，满足用户对信息安全系统的审计合规需求。

日志审计系统主要由采集器、解析引擎、分析引擎、关联引擎、搜索引擎作为核心数据处理与驱动组件，提供日志采集模块、管理分析模块、分类审计模块、关系分析模块、全文检索模块、异常告警模块、数据报表模块以及系统管理模块等各种模块功能。

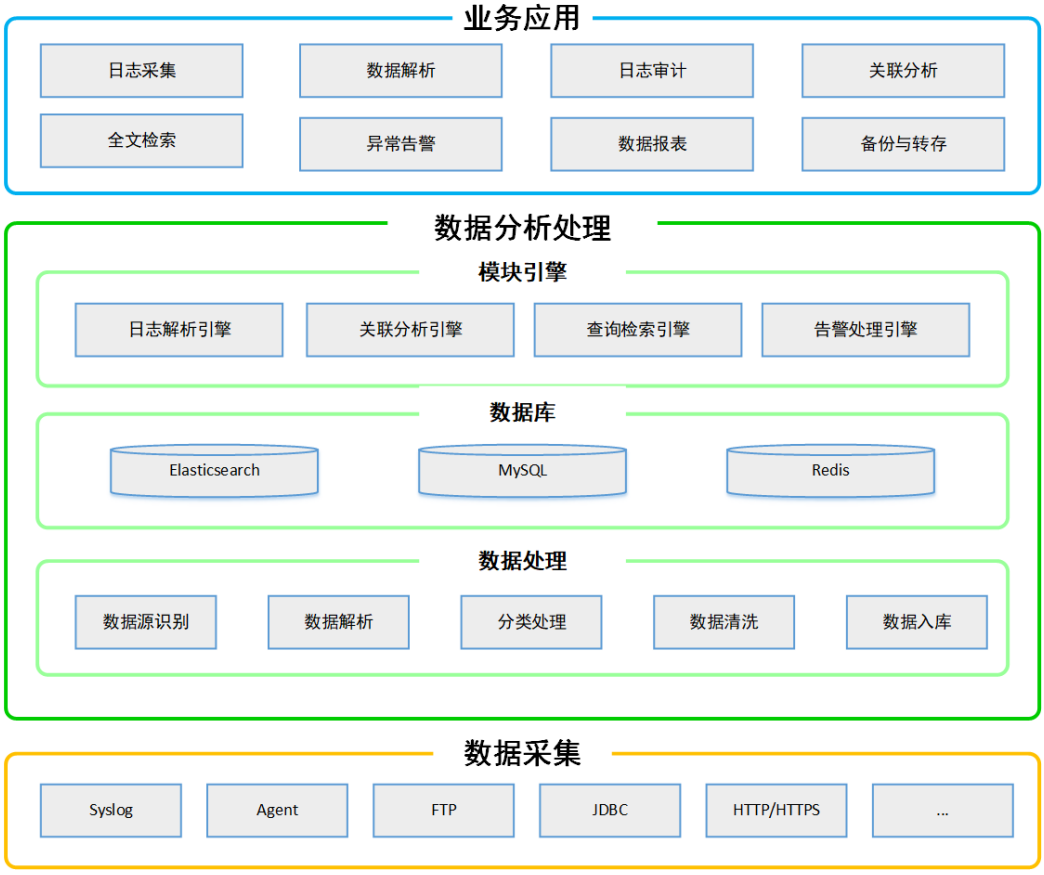

图 1 产品架构图

三、产品功能

3.1 日志采集

日志审计系统的日志采集功能主要由数据源配置、采集器、采集任务配置三个功能模块共同实现。

支持采集和汇聚各种主流网络设备、安全设备、主机设备、操作系统、中间件、应用软件、虚拟化平台等异构日志。支持 Syslog、WMI、SNMP Trap、JDBC、FTP、Agent、HTTP/HTTPS、API、Kafka 等多种数据采集方式。支持数据源配置，支持采集任务的实时定时设置管理等。

3.2 日志解析

采集的日志数据将会通过系统预设的解析引擎进行解析规则匹配。支持过滤丢弃规则设置，对属于丢弃规则的日志不予入库。支持对不同类型、不同格式的异构数据进行实时解析和分析，满足高吞吐量低延迟的数据处理要求；支持基于 Key-Value 的简单解析规则、基于逻辑表达式的复杂解析规则的自定义配置，快速实现未知日志的适配。解析后的日志落入数据库，为日志的分类审计提供支持。

3.3 分类审计

不同类型的数据源的日志格式通常不统一，本日志审计系统针对各类型的数据源进行分类分析审计，使用功能完善的审计框架，以实现日志的快速扩充、适配。

分类审计开放全字段作为查询条件，并可以将任意字段作为统计维度进行统计，统计图附带数据说明，并可切换柱状图、折线图、饼状图三种方式，支持同一维度内不同类型数据的对比。

分类审计包括主机事件审计、网络设备审计、网络安全设备审计、数据库审计、Web服务器审计、虚拟化平台审计、网络安全事件审计、网络应用行为审计、运维管理审计、其他杂项审计等。

3.4 全文检索

经过解析后的日志不仅可通过分类审计快速定位到目标日志，还可通过全文检索方式进行全局日志查询。

全文检索支持海量数据秒级反馈结果，支持通过特定条件进行高级搜索，支持基于原日志、时间、数据源、采集任务等特定字段进行全库日志检索。检索结果体现原日志、日志事件信息、数据源信息、采集任务信息等信息进行展示。

3.5 关联分析

日志审计系统内置丰富实用的日志事件定义规则，被采集的日志匹配对应的数据源类型事件库，进行事件规则匹配，提高事件匹配效率。

各类日志事件被关联形成潜在危害、异常行为等关联分析事件，并可通过告警设置将关联后的事件进行告警通知，并支持关联事件的整体与具体统计分析与事件溯源分析。

支持基于多重复合逻辑表达式的与、或及自定义规则定义，并可将多条日志事件规则合并组成关联分析规则，关联分析规则可进行频率、间隔、筛选的设置。

3.6 异常告警

日志审计系统的异常告警功能模块支持通过自定义告警规则进行告警配置，并可选择站内、邮件、短信等通知方式进行提示。自定义规则的配置支持等于、不等于、大于、小于、包含等多种数据关系定义，支持与、或两种不同的逻辑关系，支持通过时间、频率维度进行规则定义，支持自定义统计维度、统计范围和图表样式。

3.7 数据报表

日志审计系统内置报表模板，用户可以选择数据内容执行定期报表计划，以此实现工作汇报、业务展示等场景需求。支持周期性报表和单次报表两种报表生成方式，支持统计时间、数据范围、报表模板配置；支持报表转存导出 DOC、PDF 文件，并可将报表发送到指定邮箱。

3.8 备份与转存

基于大数据架构的数据存储与备份功能，利用 NFS 技术实现海量数据的无缺失、完整备份，实现数据的持久化保存与存储空间的合理利用。支持存储容量、备份数据的可视化监测与维护，支持磁盘空间耗尽时的自动异地转存策略配置。

四、关键技术

4.1 完善的数据采集与治理

采集网络空间中海量、异构、高速的日志数据。日志采集支持分布式采集，提供分布式消息队列保证数据采集性能。

通过各种渠道和途径汇聚的海量数据存在以下问题：

- 不同业务数据的数据格式定义并不完全相同；
- 不同途径获取的数据存在重复、包含，甚至矛盾的情况；
- 非结构化数据中存在许多可用于关联分析的线索，但因其存储空间大、保存时间短，难以充分有效发挥作用；
- 海量数据中存在不少无法处理或者没有价值的垃圾数据，降低了整体数据的利用率。

针对以上情况需要进行数据预处理，即对数据进行规范化、归一化、去重、补全、过滤和归并等数据处理过程，提取其中有效的信息，剔除无用的数据。

数据的**清洗过滤**包括三个方面：

- 清洗：针对数据格式的不一致、数据输入错误、数据不完整等问题，支持对数据进行转换和加工。常用的数据转换组件有字段映射、数据过滤、数据清洗、数据替换、数据计算、数据验证、数据加解密、数据合并、数据拆分等；
- 修改：错误数据，产生原因是业务系统不够健全，在接收输入后没有进行判断直接写入后台数据库造成的，比如数值数据输成全角数字字符、字符串数据后面有一个回车、日期格式不正确、日期越界等；
- 删除：重复性数据。

对于安全事件数据清洗与过滤功能包括但不限于：

- 不属于大数据平台数据源中的数据；
- 重复数据；
- 噪音数据；
- 数据不完整或不合理性的数据；
- 低于业务需求的最低级别以下的数据。

对异构原始数据进行统一格式化处理，以满足大数据平台存储层数据格式定义的设计对于被标准化的数据应保存原始日志。

数据标准化的原则包括但不限于：

在保证基本扩展能力的基础上，根据每种类型数据的标准库规则，实现相关字段的标准化；

对于常用的字段，保证字段内容的一致性，消除不同威胁对于相似问题描述的不一致性，满足依赖于这些字段的规则的可移植性。未被标准化的数据应保存原始日志。可用于事后为该特定数据再定义标准化规则。

4.2 快速展示日志数据的方法（专利）

在日志采集过程中，不同类型的设备，所发出的日志格式千差万别，这给前端界面对他们的展现、查询等处理增加了难度。传统的做法是，给每一种类型的数据专门做定制的展现和查询页面，这种开发方式有以下弊端：

（1）极大的增加了前端程序开发的工作量和难度，有多少种数据类型，就需要开发多少种页面；

（2）由于前端页面的增加，前端页面的体积就会较大，浏览器页面加载及渲染速度就会变的缓慢，用户体验不佳；

（3）缺乏灵活度，在客户现场实际使用的过程中，遇到新的没有被支持的数据格式，是比较常见的，这样就定制开发新页面，开发及部署周期较漫长，极大的影响了客户的体验。

为此，日志审计系统采用一种快速展示页面的方法，通过创建一个数据分析处理页面组来分析和处理各种类型日志数据的特征，然后将他们的特征、数据逻辑关系自动存储在后台数据库中，前端页面在展现时，根据这种逻辑关系就能自动生成展现页面和查询页面，如果遇到新的未处理过的数据，只需要将他们的特征数据通过界面设置，之后对他们的前端处理，就能够自动生成，完全不需要再次进行前端开发。

通过使用此技术，日志审计系统取得了如下有益的技术效果：

1、减少前端工作量：页面的数据展示都通过后端数据库存储的逻辑关系自动生成 ，不需要每种日志开发一个页面；

2、减少了前端页面的体积，提高了浏览器的加载及渲染速度；

3、灵活便捷：遇到新的数据，只需要进行简单的页面设置，就能够完成对新数据的支持，甚至经过简单培训，用户自己都可以完成配置，极大的增加了用户好的体验；

4、减少了售后支持的工作量和难度，降低售后支持的成本。

4.3 基于大数据技术的日志数据存储

数据存储层将采集的日志数据根据数据处理的需要保存在相应的数据库中。

数据存储支持不同类型的大数据存储，这些数据包括结构化和非结构化数据，关系型和非关系型数据，实时数据和历史数据。服务于后续的监测分析，系统使用多种数据存储技术，使用非关系型数据库 ElasticSearch 存储采集数据，使用非关系型数据库 Redis 存储缓存数据。

ElasticSearch 提供了一个分布式多用户能力的全文搜索引擎，基于 RESTful web 接口。可作为 Restful API 标准的可扩展和高可用的实时数据分析的全文搜索工具使用。分布式实时文件存储，可将每一个字段存入索引，使其可以被检索到。实时分析的分布式搜索引擎，分布式：索引分拆成多个分片，每个分片可有零个或多个副本。集群中的每个数据节点都可承载一个或多个分片，并且协调和处理各种操作；可以扩展到上百台服务器，处理 PB 级别的结构化或非结构化数据。

日志审计系统具备快速自定义的各种形式搜索，而不局限于固定几种的字段，系统可以自由选择搜索策略，保存搜索策略，可以指定字段及条件进行搜索，同时支持字段组合搜索。系统支持普通搜索、高级搜索两种方式，通过即时查询，立即产生搜索结果，操作简洁易用。

Redis 数据库是非关系型数据库，将数据存储在缓存中，读取速度快是其最大的优点。

Redis 适用于存储使用频繁的数据，这样减少访问数据库的次数，提高了运行效率。

4.4 基于分布式架构的日志关联分析

分析计算提供计算模型和计算方法。分析方法包括关联分析、数理统计分析、数据挖掘等。其中日志审计关联分析方法包含：基于规则的关联分析、基于统计的关联分析、基于行为的关联分析、数量统计分析等。

- **基于统计的关联分析**

基于规则的关联分析是指将可疑的安全活动场景（例如某潜在安全攻击行为的一系列安全事件序列）加以预先定义，系统能够使用定义好的关联性规则表达式，对收集到的安全事件进行检查，确定该事件是否和特定的规则匹配。基于规则的关联分析即可用于识别单个安全事件的场景，也可用于识别多个安全事件组成的场景。

在基于规则的关联分析过程中，安全专家制定规则形成规则库；规则驱动关联引擎分析安全事件，形成关联事件。规则采用如下产生式规则形式：

IF　条件　THEN 结果

其中，条件为安全事件中某些属性的限制条件，即规则的激活条件，具有检测事实存在与否、比较事实、根据标志检验事实等功能。条件可以由单个检测属性组成，也可以由多个检测属性组成，且各属性用逻辑符号 OR、AND、NOT 来表示多属性的逻辑关系。结果是新证据的断言或某个用户行为的可疑度，具有产生一条高优先级关联威胁的功能。

- **基于统计的关联分析**

统计关联是指在给定的时间范围内，发生符合某种条件的威胁次数超过设定值而产生报警的过程。统计关联主要是针对系统中的事件计数，通过定义门限值，统计在一定时间间隔事件发生次数。如果系统发生事件超出了正常设定的门限值，就认为系统出了异常。

在统计关联分析过程中，首先定义一些大的安全事件类别，对每个类别的事件设定一个合理的阀值，将出现的事件先归类，然后进行缓存和计数，当在某一段时间内，计数达到该阀值，可以产生一个级别更高的关联事件。

在基于统计的关联分析过程中，统计关联引擎统计固定时间长度中安全事件数量是否达到统计阀值 N，如果超过 N，生成关联威胁。

- **数量统计分析**

数量统计分析是指采用统计学方法，对各种威胁的状态、频次、发生周期等数据量化特征进行计算、得出威胁数据的分布状况、主要特征、时间序列的趋势性、是否存在异常值、威胁汇总结果等内容，威胁统计分析结果可直接用于威胁性质的判定、解释和决策。

- **基于行为的关联分析**

基于行为的关联分析用于在海量告警数据中发现攻击活动背后逻辑，发现其攻击步骤与预测其下一步攻击行为。

网络入侵行为通常是一系列的攻击，或者是组合式攻击，这些攻击都不是孤立的，而是表现为不同的阶段，前一阶段为后一阶段做准备，也就是攻击之间存在某种因果关联关系。因此，要对安全监控告警信息进行处理，以进行全局性分析。

大量的入侵案例表明入侵通常分为 4 个阶段：收集目标系统信息、进入系统、提升权限、放置后门和清理日志。通过行为关联的方法找出报警信息之间的这些关系，对于存在关联关系的告警信息就要提取出攻击步骤和策略，从而化简告警信息，提供给管理员更加直观的、深层次的信息。

4.5 使用 NFS 与快照技术（Snapshot）对日志数据进行迁移与备份

通过使用 NFS，用户和程序可以像访问本地文件一样访问远端系统上的文件，使得每个计算机的节点能够像使用本地资源一样方便地使用网上资源。换言之，NFS 可用于不同类型计算机、操作系统、网络架构和传输协议运行环境中的网络文件远程访问和共享。

日志审计系统用到的 NFS 的工作原理是使用客户端/服务器架构，由一个客户端程序和服务器程序组成。服务器程序向其他计算机提供对文件系统的访问，其过程称为输出。NFS 客户端程序对共享文件系统进行访问时，把它们从 NFS 服务器中“输送”出来。文件通常以块为单位进行传输。其大小是 8KB（虽然它可能会将操作分成更小尺寸的分片）。NFS 传输协议用于服务器和客户机之间文件访问和共享的通信，从而使客户机远程地访问保存在存储设备上的日志数据。

传统上一直采用数据复制、备份、恢复等技术来保护重要的数据信息，定期对数据进行备份或复制。由于数据备份过程会影响应用性能，并且非常耗时，因此数据备份通常被安排在系统负载较轻时进行(如夜间)。另外，为了节省存储空间，通常结合全量和增量备份技术。显然，这种数据备份方式存在一个显著的不足，即备份窗口问题。在数据备份期间，企业业务需要暂时停止对外提供服务。随着企业数据量和数据增长速度的加快，这个窗口可能会要求越来越长，这对于关键性业务系统来说是无法接受的。降低数据保护的代价，提高数据保护过程中的应用感知能力，逐步成为客户的核心诉求，因此需要将数据备份窗口尽可能地缩小，甚至缩小为零。而数据快照(Snapshot)技术，就是为了满足这样的需求而出现的数据保护技术。

快照是指关于指定数据集合的一个完全可用拷贝，该拷贝包括相应数据在某个时间点（拷贝开始的时间点）的映像。快照可以是其所表示的数据的一个副本，也可以是数据的一个复制品。从更具体的技术细节来讲，快照是指向保存在存储设备中的数据的引用标记或指针。日志审计系统使用 Snapshot 快照首先将原有的内容读取出来，写到另一位置处（为快照保留的存储空间，此文中我们称为快照空间），然后再将数据写入到存储设备中。而下次针对这一位置的写操作将不再执行复制操作，从 COW 的执行过程我们可以知道，这

种实现方式在第一次写入某个存储位置时需要完成一个读操作（读原位置的数据），两个写操作（写原位置与写快照空间），如果写入频繁，那么这种方式将非常消耗 IO 时间。因此可推断，如果预计某个卷上的 I/O 多数以读操作为主，写操作较少，这种方式的快照实现技术是一个较理想的选择，因为快照的完成需要较少的时间。

五、产品亮点

5.1 全面的日志采集能力

采用 Syslog、WMI、SNMP Trap、JDBC、FTP、Agent、HTTP/HTTPS、API、Kafka 方式，实现对各种主流操作系统、网络设备、安全设备、Web 服务器、数据库、虚拟平台等异构日志的采集，适用于传统网络，也适用于工业网络的日志审计。

5.2 灵活的数据分析与可视化呈现

日志审计系统将各类型数据源的日志数据收集后，面临着海量非结构化数据需要转换成用户易懂、可分析、可汇报、可展示的需求。因此，日志审计系统搭载了分类审计模块与数据分析模块。即时呈现隐藏在瞬息万变且庞杂数据背后的业务洞察。

分类审计中的全字段开放检索筛选功能配合时间、数据源、日志类型可以精准定位目标数据，将筛选后的日志以列表的方式呈现，并可将列表内容以统计图的方式表达，以更直观、生动的形式展示检索内容。并可根据不同的数据维度进行统计作图，柱状图表达各数据间的对比情况，折线图表达所选范围内的数据变化趋势，饼状图表达数据在整体范围内的占比情况，支持同一维度下不同数据之间的对比，具有灵活的数据分析能力。

配合使用数据报表功能，方便用户进行网络安全主题会议的材料收集与分析。

5.3 满足等保合规要求

随着企业规模的不断扩大，暴涨的数据量和急速拓展的业务量使得用户的业务体系变得更加复杂，网络环境也变得愈发复杂，随之而来的网络安全事件也不断增加，这对企业的网络安全管理员的业务水平提出了更高的要求。针对上市公司、大中型企业（尤其是央企），国家和各行业主管部门都出具了大量合规管理标准、规范、规定，对网络信息系统的安全审计提出了要求。

日志审计系统在充分领会相关标准要求后，满足《中华人民共和国网络安全法》，第三章第二十一节的第三条规定：采取监测、记录网络运行状态、网络安全事件的技术措施，并按照规定留存相关的网络日志不少于六个月，配备大容量存储空间，用于原始日志保存，最长可保存 180 天，便于安全事件的溯源管理与取证调查。

日志审计系统满足《GB / T 22239-2019 信息安全技术网络安全等级保护基本要求》中针对二级及以上信息系统安全边界防护、安全计算环境层面与日志审计相关的安全要求。

5.4 基于海量日志特征的实时关联分析

日志审计系统内置海量日志事件动作库与关联分析规则，通过日志特征条件自定义日志事件策略形成动作库，作为关联规则的基础元素，可被重复使用，支持将异构日志的实时解析与事件匹配。

支持针对 Windows、Linux/Unix、网络设备、安全设备、数据库、Web 服务器、虚拟平台等数据源产生的日志事件进行多维度的关联分析，通过建立异常行为分析模型与系统潜在危害分析模型，将网络系统中的登录、访问、操作、攻击威胁、异常情况等事件进行关联匹配，形成关联分析事件，并支持生成告警。

5.5 强大的检索查询

日志审计系统支持亿级（TB）日志查询秒级响应，支持基于关键字的日志内容全文检索，支持历史检索条件的重用，并且支持通过索引策略的配置进行条件组合查询，支持等于、不等于、大于等于、小于等于、包含、不包含等 6 种逻辑关系符的组合运用，完成精确查询。

六、应用场景

6.1 法院行业场景

- **需求**

法院业务专网属高度机密网络，如果遭受攻击、病毒、入侵造成数丢失、泄漏将会对社会造成不可弥补的损失。根据《网络安全法》、《人民法院非涉密重要信息系统安全等级保护定级工作指导意见》、《人民法院信息安全保障总体建设方案》、《信息系统等级保护基本要求》的相关要求，需通过安全建设，满足等保的相关要求。

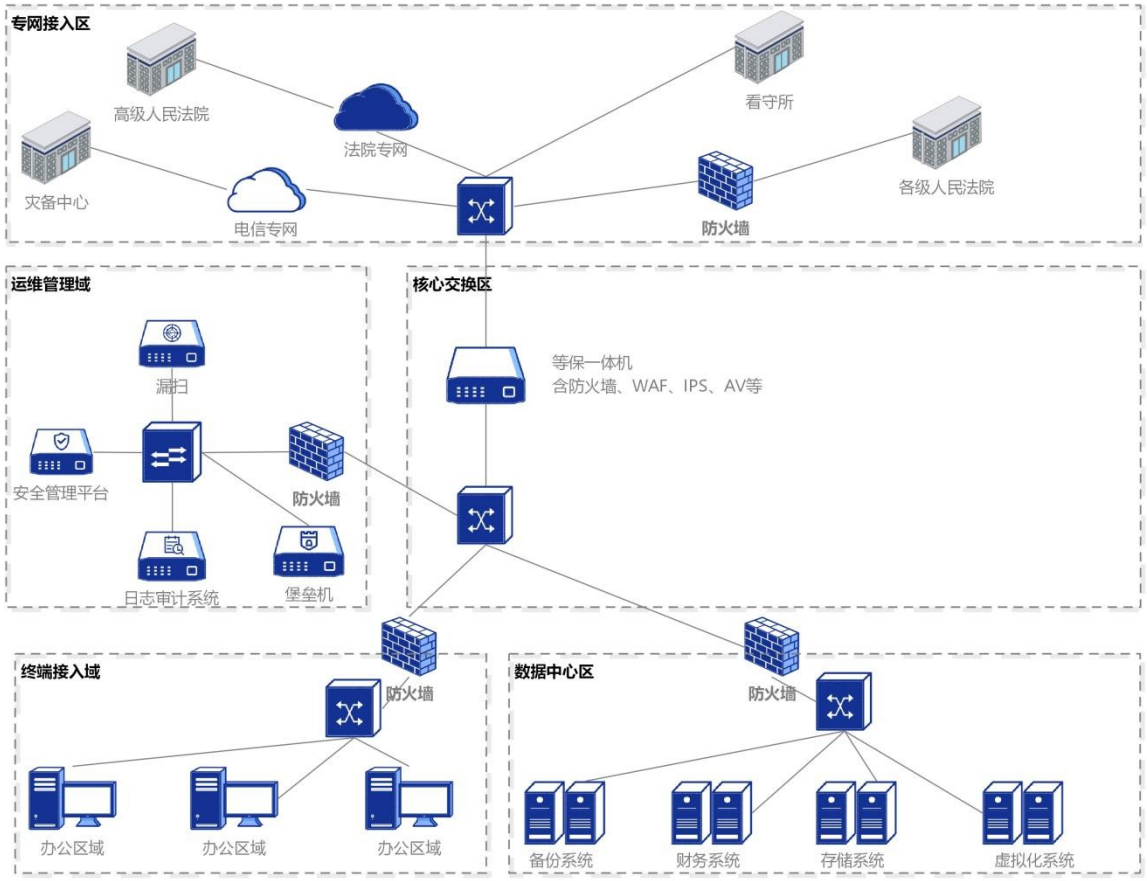

图 2 法院场景部署日志审计方案拓扑图

- **解决方案**

在运维安全域部署防火墙设备，做到管理区边界隔离；部署堡垒机，运维权限集中管理，运维行为全程审计；部署漏洞扫描系统，及时发现网络设备漏洞；部署**日志审计系统**，记录和查询网络安全日志，符合国家网络安全法和公安部 151 号令要求；部署安全管理平台，对网络设备进行统计监控与管理。

6.2 教育行业场景

- **需求**

计算机网络技术和管理信息化的发展，使我国的普通教育逐步进入新时代，数字化校

园成为未来的发展方向，普通教育实现管理网络化、教育手段现代化。伴随高校网络规模的扩大，稳定的校园网络和计算机系统成为重要的基础设施，计算机病毒、黑客、系统漏洞等网络不安全因素对数字化校园建设存在严重威胁。根据《网络安全法》、《教育信息化“十三五”规划》、《信息系统等级保护基本要求》的相关要求，需通过安全建设，满足等保的相关要求。

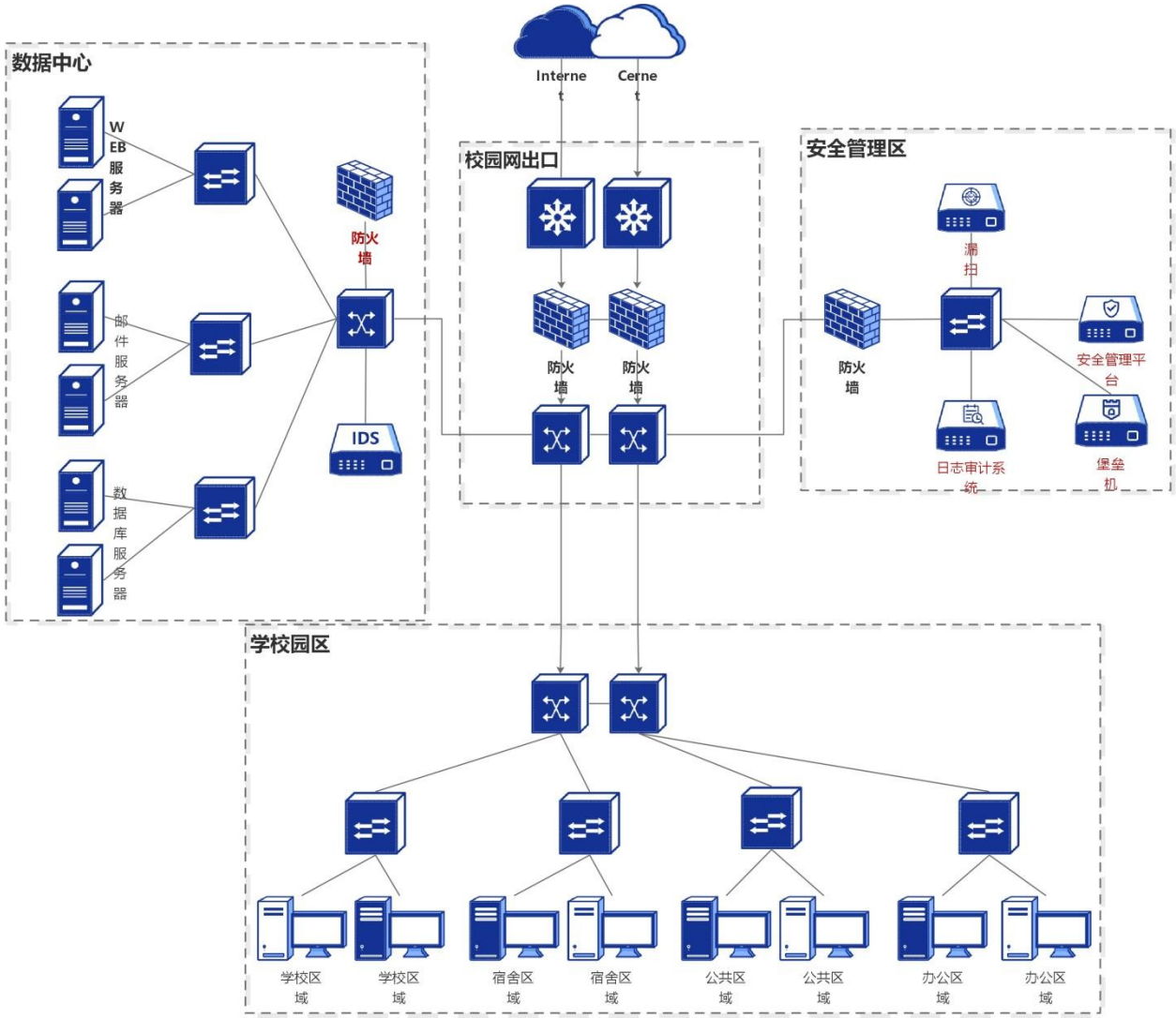

图 3 校园场景部署日志审计方案拓扑图

- **解决方案**

数据中心部署防火墙设备，做到数据中心边界的访问控制，部署 IDS 设备，有效检测针对信息系统的各种攻击，如病毒、木马等；

安全管理区部署防火墙设备，做到管理区边界隔离；部署堡垒机，运维权限集中管理运维行为全程审计；部署漏洞扫描系统，及时发现校园网络设备漏洞；部署**日志审计系统**，记录和查询校园网络安全日志，符合国家网络安全法要求；部署安全管理平台，对网络设备进行统计监控与管理。

6.3 广电行业

- **需求**

作为信息技术与广电行业技术相结合的产物，网络安全工作的重要性日益彰显，广电总局于 2011 年发布了《广播电视安全播出管理规定》（总局令第 62 号），明确提出要开展信息系统等级保护工作。

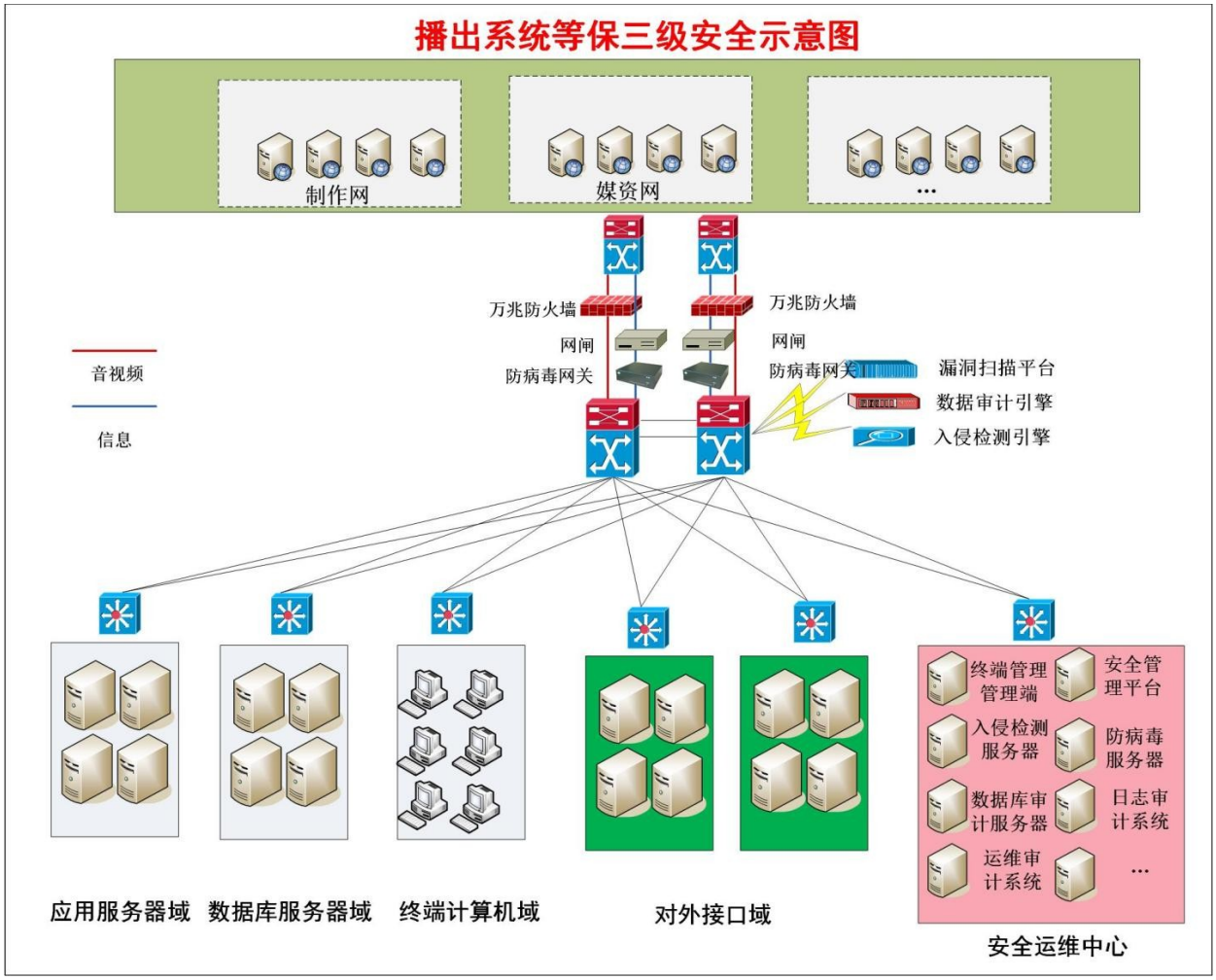

图 4　广电场景部署日志审计方案拓扑图

- **解决方案**

在系统边界部署防火墙，对进入系统的数据包进行细粒度的访问控制；部署隔离网闸与单向网闸（单向网络隔离与交换系统），实现相关系统内、外的安全隔离和文件的安全摆渡；在相关系统边界部署 IDS 系统，对进入相关系统的入侵行为进行监控，并将告警信息发送至安全管理中心；部署**日志审计系统**、堡垒机，对相关系统接入交换机、服务器、安全设备的运行状况、用户行为等重要事件进行安全审计；部署中数国科安全管理平台，对网络设备进行统计监控与管理。

方案根据国标与广电总局行标等级保护标准的要求，帮助用户从全局视角进行网络安全相关的运维支持和安全服务，进一步提高电视中心信息安全管理和运维的能力，同时满足用户对信息安全系统的审计合规需求。