
This is the **published version** of the bachelor thesis:

Soler Vega, Genís; Yang, Xiaoyuan, dir. Sistema de Seguridad de una escuela que proporciona estadísticas para ver el impacto de la navegación en la nota de los alumnos. 2021. (958 Enginyeria Informàtica)

This version is available at <https://ddd.uab.cat/record/257837>

under the terms of the  license

Sistema de Seguridad de una escuela que proporciona estadísticas para ver el impacto de la navegación en la nota de los alumnos

Genís Soler Vega

Resumen— Con el incremento del uso de Internet, cada vez son más comunes los ataques informáticos a través de la red. Es por este motivo que, en este artículo, se expone como se ha desarrollado un sistema para proteger la red interna de una escuela y a todos los alumnos. El sistema, así mismo, pretende positivar el uso de la seguridad informática, hasta ahora un término que se refiere a algo negativo o de peligro, para ver la repercusión que tienen los distintos sitios webs visitados en la nota de los diferentes alumnos, ya sea positiva o negativamente, y generar estadísticas para los profesores.

Palabras clave— Aplicación web, ciberseguridad, Proxy, React, Go, Pentaho

Abstract— With the incrementation in the use of the Internet, computer attacks through the network are becoming more common. It is for this reason that, in this article, it is exposed how a system has been developed to protect the internal network of a school and all students. The system, likewise, aims to make the use of computer security positive, until now a term that refers to something negative or dangerous, to see the impact that the different visited websites have on the grade of the different students, whether it is positive or negatively, and generate statistics for teachers.

Index Terms— Web application, cybersecurity, Proxy, React, Go, Pentaho



1 INTRODUCCIÓN – CONTEXTO DEL TRABAJO

Con el aumento del uso de la tecnología en el sector educativo y el aumento del número de ciberataques surge la necesidad de un sistema que nos proteja y, a su vez, nos permita sacar el máximo provecho a la tecnología con Big Data para poder mejorar el nivel de la enseñanza.

Con la llegada de la pandemia, todos los centros se han visto obligados a adaptarse a utilizar todo tipo de herramientas para poder hacer clase a distancia y, en el momento de volver a las clases presenciales, van a aparecer nuevos métodos de enseñanza mediante el uso de la red.

Para conseguirlo, en este proyecto se propone un sistema capaz de mantener la seguridad de la red y, a su vez, utilizar todos los datos que esta genera para mejorar el nivel de docencia.

Este documento se divide en nueve secciones. Empieza con un listado de los objetivos a cumplir. Seguidamente se

comenta el estado del arte en la sección 3.

En las secciones 4 y 5 se explica cómo es la arquitectura de todo el sistema y su diseño y implementación.

Finalmente, se detalla un test de rendimiento del sistema, las conclusiones finales y los planes futuros en las secciones 6, 7 y 8.

2 OBJETIVOS

Los objetivos de este proyecto se pueden agrupar en dos grupos: objetivos de desarrollo y objetivos del producto final.

2.1 Objetivos de desarrollo

1. Tener una interfaz web para poder interactuar con el sistema. Esta se divide en dos grupos. Interfaz para el administrador e interfaz para el profesor.
 - a. **Interfaz de administrador:** Permitir hacer la configuración del sistema: gestión de usuarios, gestión de cursos y visualización de estadísticas.
 - b. **Interfaz de profesor:** Poder gestionar las actividades de los cursos asignados al

-
- E-mail de contacto: genis.soler@e-campus.uab.cat
 - Mención realizada: Ingeniería de computadores
 - Trabajo tutorizado por: Xiaoyuan Yang (CAOS)
 - Curso 2021/22

profesor, gestionar las calificaciones de los alumnos y visualizar las estadísticas del impacto de las webs visitadas.

2. Hacer un proyecto capaz de adaptarse a todo tipo de escuelas, desde primaria hasta universidades.
3. Hacer una interfaz intuitiva y fácil de mantener.
4. Los diferentes accesos de la página web deben poderse configurar a través de la base de datos, sin necesidad de cambiar código.
5. Establecer un entorno que permita configurar y administrar la red interna de la escuela:
 - a. Permitir o bloquear el acceso a las distintas páginas web.
 - b. Predefinir las direcciones IP de los distintos dispositivos de la escuela.
 - c. Permitir el acceso al sistema mediante dispositivos externos. (Tablet o móvil del usuario)
 - d. Guardar un registro de los diferentes accesos a Internet.
 - e. Gestionar los accesos permitidos a Internet.
 - f. Generar los datos del impacto en la nota de las distintas webs en momentos de menos actividad para acelerar la obtención de las estadísticas.

2.2 Objetivos de ayuda a la docencia

1. Poner a disposición de los profesores un conjunto de datos que le permitan visualizar de forma rápida las mejores y peores páginas web, así como la evolución y el impacto en las diferentes asignaturas de una web seleccionada, para maximizar el proceso de aprendizaje.

3 ESTADO DEL ARTE

Hoy en día, en situación de pandemia, se ha acelerado el concepto de digitalización de las aulas, el concepto de la evolución de la educación tradicional a la e-educación.

En España, durante el 2021, se produjeron una media de 40.000 ataques cada día, que supone un aumento del 125% respecto 2020. [21]

Actualmente, en España se ha destinado un presupuesto de casi 1.000 millones de euros [20] para la digitalización del sistema educativo ya que, según PISA (programa para la evaluación internacional de los estudiantes), en 2018 España era el país que menos fomentaba el uso de la tecnología y solo el 11% de los profesores incentivaban su uso. [22]

Con esto en mente, se prevé que haya un aumento considerable del uso de tecnologías en las aulas y, por lo tanto, surge la necesidad de tener un sistema que cuente con protección ante el aumento de ciberataques y que incentive a los profesores al uso de la tecnología, facilitando datos y estadísticas para mejorar la calidad de la formación.

El uso responsable de la red es un tema fundamental en la era post-covid. Un uso no controlado y mal orientado puede impactar negativamente a los estudiantes y a su rendimiento escolar.

Las noticias de ciberacoso o cyberbullying son fenómenos constantes en nuestras escuelas. Por desgracia, los profesores no disponen de ninguna herramienta que indique el uso que están realizando sus alumnos. [26][27]

4 ARQUITECTURA DEL SISTEMA

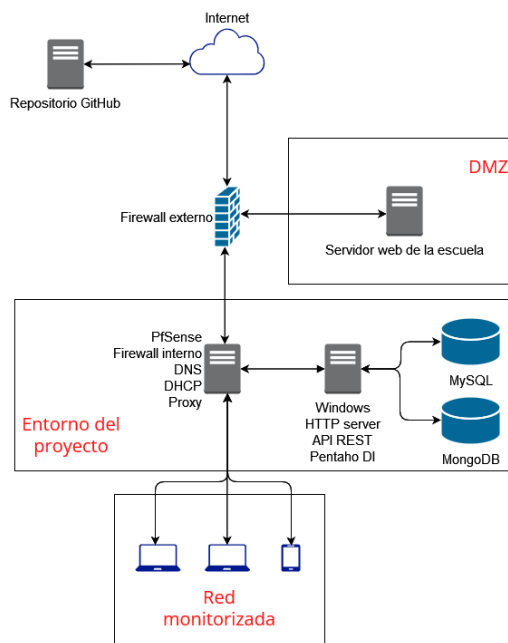


Fig. 1: Estructura del sistema.

En cuanto a la arquitectura del sistema, se ha establecido un entorno jerárquico donde se diferencian dos entornos: externo y propio del proyecto.

4.1 Elementos de la red externa

En este entorno se encuentran todos los sistemas que no pertenecen al desarrollo del proyecto pero que nos sirven para poder situar nuestro sistema dentro de la escuela.

4.1.1 Repositorio GitHub

Repositorio de programario en la nube donde se aloja el código del programa para asegurar no perder todos los cambios si nuestro dispositivo falla.

4.1.2 Internet

Red externa a la infraestructura de la escuela y que no podemos controlar.

4.1.3 Firewall externo

Elemento que separa red interna y Internet. Controla el paso de datos entre las dos redes y permite acceder a la web de la escuela.

4.1.4 DMZ

Red local aislada situada entre la red de la escuela y Internet que almacena aquellos archivos y recursos accesibles desde Internet, en este caso la web informativa de la escuela.

4.2 Entorno del proyecto

En este entorno se encuentran todos los sistemas que pertenecen al proyecto y que se han desarrollado y configurado.

4.2.1 *PfSense*

Contiene los siguientes servicios:

1. **Firewall interno:** firewall que controla el paso de datos para proteger la red interna, donde se encuentran los datos y recursos que no deben ser accesibles desde Internet.
2. **DNS:** Servicio que permite obtener la información asociada a un nombre de dominio. Comúnmente traduce un nombre de dominio a su dirección IP. Por ejemplo www.google.es → 172.217.168.67
3. **DHCP:** Permite obtener automáticamente toda la configuración de red a los dispositivos que se conecten. Además, permite establecer configuraciones estáticas para los dispositivos que siempre tienen la misma dirección. (por ejemplo, servidores internos)
4. **Proxy:** Servicio que hace de intermediario en las conexiones a internet. De esta manera se consigue que el destinatario no sepa de donde proviene la petición. Al ser el mediador, permite definir una lista de direcciones para denegar el acceso, así como tener un registro de todas las peticiones realizadas.

4.2.2 *Servidor Windows*

contiene los siguientes servicios:

1. **HTTP server:** servidor Node.js que contiene la web que sirve como interfaz para los usuarios.
2. **API REST:** servicio mediador entre interfaz y bases de datos que proporciona y guarda los datos solicitados por el usuario.
3. **Pentaho DI:** servicio encargado de leer los datos de los accesos a Internet y generar las estadísticas.
4. **MySQL:** servidor de bases de datos relacional encargado de almacenar los datos de configuración del portal web.
5. **MongoDB:** servidor de bases de datos orientada a documentos encargado de almacenar todos los datos relacionados a las estadísticas.

4.2.3 *Red monitorizada*

Red de dispositivos de la escuela: dispositivos monitorizados por el sistema y usados por los alumnos.

4.2.4 *Distribución de los servicios*

Para este proyecto se ha decidido englobar los servidores HTTP, API REST y de bases de datos en un solo dispositivo

Windows para facilitar el acceso y la programación.

En caso de poner el proyecto en marcha, el Servidor HTTP debe estar separado del servicio API y de las bases de datos para permitir el acceso a estos últimos solamente desde el servidor HTTP y proteger, de esta manera, el acceso a los datos desde el exterior.

5 DISEÑO Y IMPLEMENTACIÓN

5.1 Módulos y funcionalidades

Los módulos que forman el proyecto son la arquitectura y configuración del sistema, Interfaz web, API REST y el programario de generación de estadísticas.

5.1.1 *Arquitectura y configuración del sistema*

Conjunto de servicios que permiten hacer un control del tráfico interno de la escuela, así como de las conexiones que conectan con el exterior. Es importante tener una buena seguridad para proteger todos los datos de la organización y obtener los datos necesarios con el objetivo de proporcionar datos estadísticos para la mejora de la formación.

Toda la parte de gestión de seguridad se realiza desde el servidor PfSense, indicado en la fig. 1

5.1.2 *Interfaz web*

- Permite dar de alta, editar, eliminar y filtrar a todos los usuarios de la escuela, incluidos administrador, profesor y alumno.
- Establecer toda la organización de la escuela, permitiendo:
 - Dar de alta, editar y eliminar los distintos cursos.
 - Dar de alta, editar y eliminar las asignaturas asignadas a los cursos.
 - Asignar los profesores que imparten las asignaturas.
 - Asignar los alumnos que realizan las asignaturas.
- Visualizar estadísticas sobre:
 - Las 5 webs con mejor impacto en una asignatura.
 - Las 5 webs con peor impacto en una asignatura.
 - La evolución del impacto de una web en función del número de visitas que recibe.
 - El impacto de una web en las distintas asignaturas donde se han registrado visitas.
- Permite al profesor organizar toda la estructura de las asignaturas en las que esté asignado:
 - Establecer los apartados que forman la asignatura, por ejemplo, teoría y prácticas, así como el peso que tienen en la nota final.
 - Dar de alta, modificar y eliminar actividades, así como establecer el peso en la

nota dentro del apartado.

- Asignar la nota de una actividad de todos los alumnos de la asignatura.

5.1.3 API REST

Sistema que permite a la interfaz web obtener y generar todos los datos. Los distintos usuarios no interactúan de forma directa por lo que no tienen acceso.

5.1.4 Generación de estadísticas

Las estadísticas se generan de forma periódica en momentos de baja actividad para no tener un impacto negativo en el rendimiento de todo el sistema y para agilizar el proceso de consulta.

El programa se ejecuta mediante un script que puede configurar cada administrador mediante “cron”, una herramienta que permite planificar la ejecución de tareas de forma automática, por lo que los usuarios no tienen acceso al programa.

El programa obtiene los accesos a internet desde la última ejecución, todas las calificaciones de los alumnos en las actividades cuya fecha de entrega esté entre la fecha de la última ejecución y la fecha actual y, por último, toda la información generada anteriormente.

Las estadísticas generadas se guardan en una base de datos MongoDB para que puedan ser consultadas de forma rápida.

5.2 Metodología de desarrollo

Para el desarrollo del proyecto y con la finalidad de tener un buen control del flujo de trabajo y del cumplimiento de los tiempos de entrega, se ha escogido seguir la metodología ágil Kanban.

El principal argumento para escoger este modelo es que no presenta grandes resistencias a posibles cambios. Esto es importante ya que el sistema se compone de muchos componentes que han de trabajar de forma conjunta y es muy probable que, durante el proceso de desarrollo, surja un posible cambio de funcionalidad o de software utilizado para el funcionamiento de uno de los componentes.

Este modelo es también muy visual, ya que permite ver el estado de todo el proyecto de forma muy rápida.

Finalmente, debido a que no es una metodología específica para el desarrollo de Software, permite hacer un seguimiento de la parte de desarrollo de la aplicación web y de toda la preparación del sistema de seguridad.

Cabe destacar que se han realizado reuniones de forma periódica con el tutor para supervisar el trabajo realizado y establecer los siguientes desarrollos a realizar.

Por último, se ha utilizado un controlador de versiones del código con GIT [8] para tener actualizados los diferentes cambios realizados en un repositorio en la nube.

5.3 Consideraciones

Durante el desarrollo del proyecto se han tomado decisiones importantes relacionadas con aspectos legales y herramientas utilizadas.

5.3.1 Aspectos legales

Todos los datos recogidos sobre las webs que visitan los alumnos son de uso exclusivo para generar estadísticas propias y, en ningún caso, son proporcionados a terceros.

Las estadísticas sobre el impacto se muestran siempre de forma general, de manera que el profesorado y el personal técnico nunca tiene acceso a ver qué sitios web visita un alumno en concreto, cumpliendo, De esta manera, con el artículo 5 de la Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y Garantía de Derechos Generales (LOPDGDD). [23]

Así mismo, solamente se podrán recoger dichos datos con previo consentimiento del afectado cumpliendo, De esta manera, con los artículos 6 y 7 de la LOPDGDD. [23]

5.3.2 Tecnologías utilizadas

5.3.2.1 Front-end

React:

Se ha utilizado el framework de JavaScript, desarrollado por Facebook, React para el desarrollo de la parte Front-End del portal web. Aunque existen otras tecnologías, react es actualmente el framework más usado. [19]

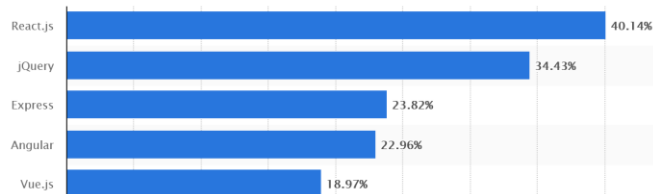


Fig. 2: Frameworks de JavaScript más usados.

A diferencia de HTML, React trabaja con un DOM virtual de manera que no es necesario renderizarlo entero por cada cambio. Es decir, el DOM virtual utiliza el DOM real, pero solamente actualiza los cambios (de forma automática), consiguiendo una rápida renderización.

Por otra parte, se ha obtenido por React debido a que, en comparación de sus competencias, es el más recomendado. El 64.6% de la gente que lo ha usado lo volvería a usar, mientras solamente el 7.1% no lo volvería a usar. [24]



Fig. 3: Resultados encuesta de los usuarios de react, vue y angular.

MaterialUI:

Material UI es una librería para React que permite importar 40 componentes predefinidos a nuestra web. Los componentes constan de un buen acabado visual y permiten trabajar con CS mediante programación con JavaScript. [4]

Aunque Bootstrap es el framework mas utilizado, tiene un tamaño mínimo de 49KB en javascript y 137KB en código CS y es necesario vigilar las librerías importadas y no utilizadas ya que pueden aumentar mucho el tamaño final.

Material ui cuenta con Material Design Lite que tiene un tamaño menor de 27KB y ayuda a hacer las animaciones mas fluidas. [13]

Recharts:

Actualmente la librería mas usada para hacer gráficos con react, con 24.9 mil usuarios y 410 mil descargas semanales.

Permite crear gráficos de forma muy rápida y sencilla, aunque existen alternativas como Nivo que tiene menor tamaño (120.2KB de Recharts y 83.1KB de Nivo) y más rapidez de renderización.

Finalmente me decanté por Recharts por ser el más usado, ya que nivo tiene 7.2 mil usuarios y 605 descargas semanales. [11]

5.3.2.2 Back-end

GO:

Go es un lenguaje que está creciendo mucho en cuanto a desarrolladores que quieren aprender-lo. En el 2019, más de un 70% de los desarrolladores tenían a GO como primera preferencia en cuanto a experiencia.

El lenguaje es similar a C y los desarrolladores que programan en este lenguaje son los que más facilidad han tenido a la hora de aprender GO.

El uso de go en el desarrollo de API ha aumentado mucho siendo utilizado en un 59% de los casos en 2016 a un 73% en 2018.

Además, el lenguaje destaca por tener un muy buen rendimiento al ejecutar todo tipo de funcionalidades sin fallar, en la concurrencia i en la seguridad, aspectos muy importantes para la ejecución de una API. [14]

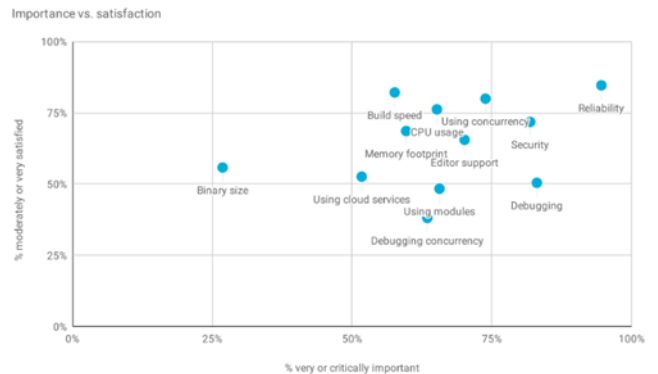


Fig. 4: cualidades de GO categorizadas por su importancia y la satisfacción de los usuarios.

5.3.2.3 Capa de persistencia

MySQL:

Este motor de bases de datos relacionales y de código abierto ha sido el elegido para almacenar todos los datos relacionados con el portal web, principalmente debido a que permite almacenar una gran cantidad de datos y tiene una buena escalabilidad y flexibilidad.

La principal competencia de MySQL en cuanto a programario de código abierto es PostgreSQL, que dispone de mejores prestaciones en cuanto a rendimiento en aplicaciones muy grandes, en integridad y fiabilidad de datos e, incluso, a la cantidad de tipos de datos que permite almacenar, siendo esta superior a los ofrecidos por MySQL.

Al no necesitar de transacciones muy complejas y con una gran cantidad de datos, he decidido utilizar MySQL ya que, actualmente, es el mas usado con el 38.9% (datos del 2019). En 2022 es el primer motor de bases de datos de código abierto con una puntuación de 1206.05 seguido por postgres con una puntuación de 606.56 [12]

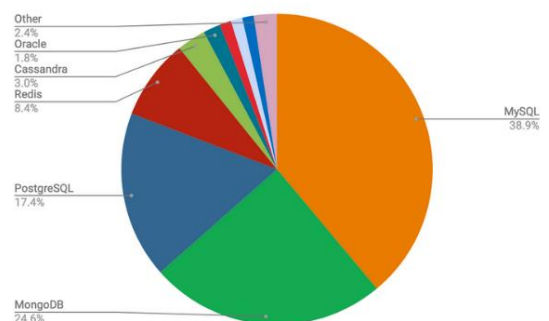


Fig. 5: Uso de los principales motores de bases de datos.

Cabe destacar que, en caso de ampliar las funcionalidades a un proyecto final en la nube, sería preferible utilizar PostgreSQL por su mayor escalabilidad y mejor soporte para la concurrencia. Además, MySQL es propiedad de Oracle y, aunque también es de código abierto, podrían cambiar las condiciones en un futuro.

En cuanto a combinaciones de uso de motores de bases de datos, MySQL es el que más veces se combina con

MongoDB, motor de bases de datos escogido para las estadísticas. [25]

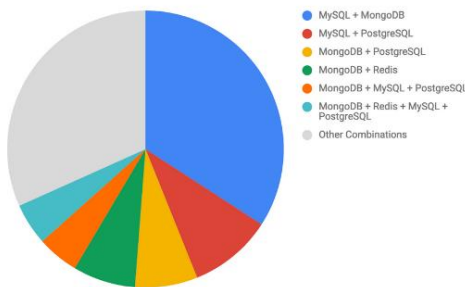


Fig. 6: Principales combinaciones entre dos motores de bases de datos.

Para la interacción con la base de datos se ha utilizado MySQL Workbench, una aplicación gratuita de Oracle que permite realizar todas las acciones de manera visual.

MongoDB:

Mongo DB es el motor de bases de datos orientado a documentos más utilizado con una puntuación de 488.57 seguido de Amazon DynamoDB con 79.85.

Además, es el segundo motor de bases de datos con más proyección, teniendo un 17.89% de la gente que quiere aprenderlo. El primero es PostgreSQL con un 17.99% y el tercero Redis con un 12.58%. [12]

5.3.2.4 Arquitectura del sistema

PfSense:

PfSense es una distribución basada en FreeBSD que contiene todo tipo de servicios de administración de redes preinstalados. Aunque en un principio están todos desactivados, es posible configurarlos e iniciarlos de una forma sencilla y rápida a través de su portal web.

5.3.2.5 Pipeline de generación de estadísticas

Bash:

Dado que PfSense genera los registros en su propio formato, se utiliza este intérprete de órdenes para dar un formato parseable y copiar el archivo resultante en el equipo donde se tratan todos los datos. Finalmente, elimina el contenido de éste para tener siempre los datos actualizados.

Para el tratamiento de los datos, se usa Pentaho DI, una plataforma de BI.

Pentaho DI

Se ha decidido utilizar Pentaho debido a la capacidad de creación de informes y, aunque está principalmente orientado a empresas, existe una distribución comunitaria.

Por otro lado, es un sistema muy flexible en cuanto a la integración de datos desde casi cualquier fuente y tiene un alto rendimiento, permitiendo trabajar con un gran volumen de datos.

Por el contrario, no existe mucha documentación más allá de manuales de usuario básicos.

5.3.2.6 Entorno de desarrollo y testing

VirtualBox:

Permite simular múltiples dispositivos con sistemas operativos distintos en un mismo host.

Esta herramienta soporta múltiples plataformas y está disponible de forma gratuita bajo licencia GNU.

Permite asignar los recursos de nuestro host de forma personalizada a las distintas simulaciones de forma muy sencilla. Además, cuenta con un driver instalable que permite una comunicación y compartición de recursos eficiente entre las simulaciones y el host principal.

Existen otras aplicaciones que permiten crear estas simulaciones, como por ejemplo VMWare, pero VirtualBox es la única disponible gratuitamente.

5.4 Implementación

5.4.1 Estructura de la base de datos para la configuración de los accesos

Para este proyecto se ha querido hacer que la gestión de accesos permitidos para cada tipo de usuario sea muy flexible y sin necesidad de tocar el código fuente.

La configuración se puede realizar desde la base de datos debido a la siguiente estructura:

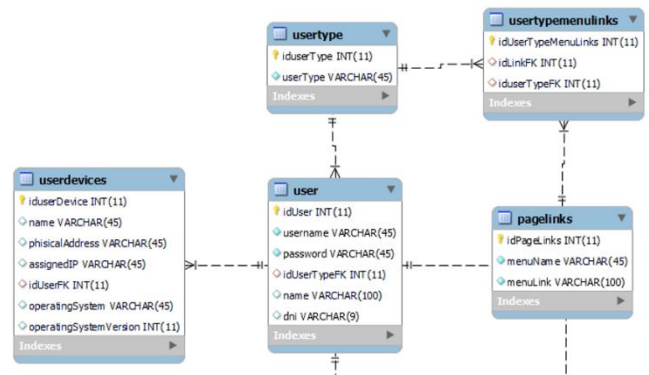


Fig. 7: Estructura para la configuración de accesos al portal por cada tipo de usuario.

PageLinks contiene todos los diferentes accesos. El portal web cargará de userTypePerMenyLinks los accesos disponibles para el tipo de usuario que haya iniciado sesión.

deliveriesdeliveries5.4.2 Estructura de bases de datos MySQL

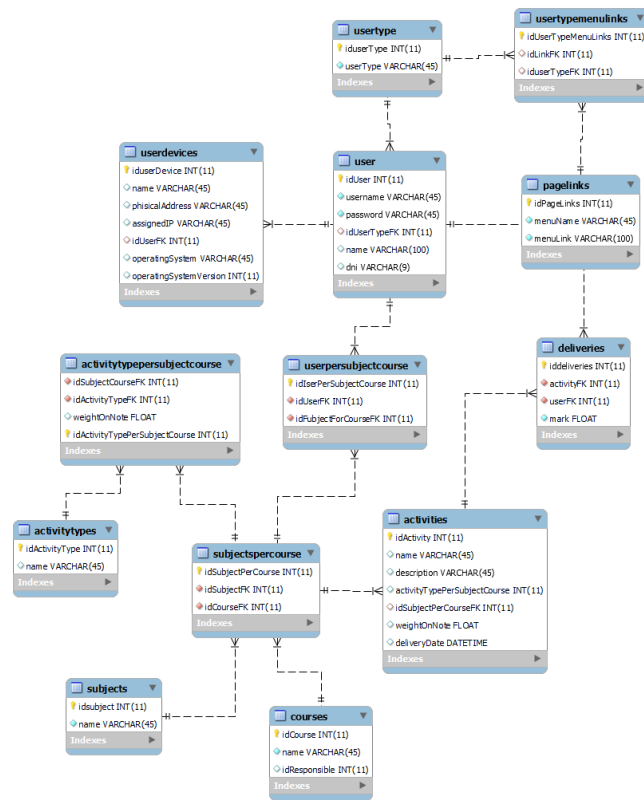


Fig. 8: Estructura completa de la base de datos MySQL.

Con la estructura anterior se pueden crear todo tipo de cursos y asignar las asignaturas que se deseen. Además, permite adaptarse a entornos como las universidades al poder dar de alta a un alumno en diferentes asignaturas, aunque esté realizando más de un curso a la vez.

También se adapta a las escuelas de educación primaria y secundaria, ya que permite asociar una misma asignatura en más de un curso.

Se mantienen separadas las tablas de cursos, asignaturas y usuario de forma que se relacionan en una tabla donde cada registro pertenece a un curso-asignatura. De esta manera, se registran los usuarios sin depender de que estén en un curso concreto.

En referencia a las actividades, del mismo modo que los usuarios, se relacionan con curso-asignatura de manera que, cada usuario perteneciente a dicho curso-asignatura, pueda realizar una entrega y obtener una calificación personalizada.

Finalmente, se ha añadido una tabla para guardar los datos de los dispositivos de cada alumno con su correspondiente IP asignada en el DHCP para poder hacer la unión de página visitada con el impacto en la nota.

5.4.3 Generación de estadísticas

En la generación de las estadísticas inicialmente se ejecutan tres hilos de forma paralela:

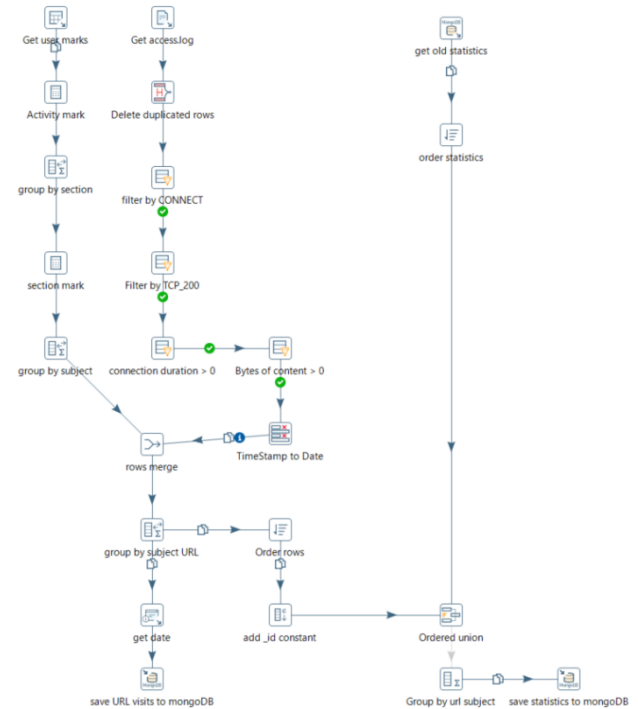


Fig.9: Proceso de extracción, transformación y carga (ETL) de las estadísticas.

Hilo 1

Se encarga de obtener de MySQL todas las calificaciones de los alumnos en las actividades cuya fecha de entrega se sitúe entre la última ejecución del programa y la actual.

Una vez obtenidos los datos, se calculan las calificaciones en base al peso de la actividad dentro de la sección a la que pertenece.

Seguidamente, se agrupan todas las actividades por usuario, asignatura y sección, generando un nuevo valor a partir de la suma de las actividades anteriores.

Se calcula la nota de cada sección en base al peso que tiene dentro de la asignatura.

Finalmente, se agrupa todos los resultados por url y asignatura sumando cada valor calculado anteriormente resultando en la nota final de cada alumno en cada asignatura.

Hilo 2

Este hilo es el encargado de obtener todos los datos de los accesos registrados por el proxy y filtrarlos de manera que elimine entradas duplicadas, seleccione solamente las conexiones de tipo "connect" y se hayan podido realizar con éxito (TCP_200), se descartan todas las entradas de las conexiones que no tienen duración y que hayan descargado algún tipo de contenido.

Finalmente se transforma la columna de la fecha de

acceso, guardada en formato timestamp.

Hilo 3

Obtiene todos los datos del impacto de cada web en cada asignatura calculados en una ejecución anterior.

Fusión del hilo 1 y 2

Una vez calculadas todas las calificaciones de los alumnos y obtenidos todos los accesos realizados, se agrupan los datos por url y asignatura y se calcula la media de las notas por cada web y asignatura junto con el número de entradas, que equivale al número de visitas.

Los datos resultantes se guardan en colecciones diferentes y se utilizarán para guardar dos tipos de estadísticas:

1. Estadísticas para medir la evolución de cada página visitada a lo largo del tiempo junto con el número de visitas.
Estos registros son únicos y no forman parte del proceso de obtención de estadísticas generadas en ejecuciones previas.
2. Estadísticas para medir el impacto en cada asignatura.
Para generar estos datos, es necesario agrupar con los datos obtenidos por el hilo 3, hacer la media y actualizar la base de datos.

5.4.4 Configuración de los servicios del sistema

Primero se ha instalado el sistema PfSense [1]. Se ha hecho mediante la descarga de la imagen ISO de la web oficial y instalado en un entorno virtual con VirtualBox.

Inicialmente, se deben crear las interfaces de red para poder acceder al portal web i configurarlo. Se han creado dos interfaces:

1. *Interfaz WAN*: configurada para pedir la IP mediante DHCP y conectada a la red "NAT" de la máquina virtual para que sea el host principal el que le asigne conexión a internet.
2. *Interfaz LAN*: se ha asignado una IP estática "192.168.1.1" a esta interfaz y conectado a la red "red interna" de la máquina virtual para simular una intranet virtual.

5.4.4.1 Configuración del servidor DHCP

Este servicio se ha establecido en el sistema PfSense - 192.168.1.1.

Tabla 1

Parámetros establecidos para el servidor DHCP

Parámetro	Valor
Subnet	192.168.1.0/32
Rango de direcciones	192.168.1.1 - 192.168.1.254
Rango reservado	192.168.1.1 - 192.168.1.9
Rango proporcionado	192.168.1.10 - 192.168.1.254
Clientes permitidos	Todo tipo de cliente
DNS principal	192.168.1.1
Direcciones estáticas	Windows - 192.168.1.2

Es necesario añadir todos los dispositivos de los usuarios para poder relacionar cada visita con la nota final del alumno.

5.4.4.2 Configuración del servidor DNS

Este servicio se ha establecido en el sistema PfSense - 192.168.1.1.

Tabla 2

Parámetros establecidos para el servidor DNS

Parámetro	Valor
Interfaz asignada	LAN
Puerto	53
Hosts	windows-server - 192.168.1.2

El uso de un DNS Forwarder es necesario para el correcto funcionamiento, ya que no será él el que resuelva los nombres de dominios, sino un servidor DNS exterior.

5.4.4.3 Configuración del servidor Squid proxy

Este servicio se ha establecido en el sistema PfSense - 192.168.1.1.

Tabla 3

Parámetros establecidos para el servidor DNS

Parámetro	Valor
Interfaz asignada	LAN
Puerto	3128
Transparent proxy	Si
Acces log	Si, en la dirección /var/squid/logs

El uso del Acces log permite ver qué páginas web son visitadas desde los distintos dispositivos de la red. No guarda el nombre de usuario, sino la dirección IP.

5.4.4.4 Instalación de MongoDB

Este servicio se ha establecido en el sistema Windows-192.168.1.2.

Se ha creado una nueva base de datos "tfg" junto dos colecciones:

1. *webpagesBySubjectCourse*: colección que almacena el impacto de cada web visitada en cada asignatura. Solamente se guarda un registro por cada web-asignatura haciendo la media en la nota.
2. *webpagesEvolution*: colección que almacena los registros agrupados de URL-asignatura junto con el número de visitas y nota. En cada generación de estadísticas, se añaden todos los registros como nuevos datos.

6 RENDIMIENTO DEL SISTEMA

Se ha realizado un estudio de rendimiento de los diferentes sistemas para ver como se comportan en caso de recibir mucha carga de peticiones.

El comando para ejecutar el test es `ab -c Peticiones simultaneas -n TotalPeticiones direcciónURL`. [10]

Las pruebas se han realizado en un entorno simulado mediante la creación de máquinas virtuales. Las configuraciones se muestran a continuación:

Tabla 4
Configuración Windows 10 (HTTP y API server)

Parámetro	Valor
Procesador	Intel(R) Core(TM) i7-9750H
Velocidad del procesador	2.60GHz
Número de procesadores	4
Memoria HDD	50GB
Memoria RAM	5GB

6.1 Rendimiento del API con GO y servidor HTTP

Se ha optado por hacer diferentes pruebas combinando el número de peticiones totales al servidor y número de peticiones simultáneas (conurrencia). Los resultados obtenidos son los siguientes:

Tabla 4
Prueba de rendimiento de la API GO

Peticiones totales	Peticiones simultáneas	Peticiones por segundo	Tiempo total en segundos	Peticiones fallidas
100	10	14.49	6.9	0
500	10	15.35	32.5	0
100	100	75.13	1.33	0
500	500	205.15	2.4	0
1000	100	127.57	7.8	0
1000	1000	281.35	3.5	0
2000	2000	382.11	5.23	0
2500	2500	321	7.77	154

Tabla 4
Prueba de rendimiento del servidor HTTP

Peticiones totales	Peticiones simultáneas	Peticiones por segundo	Tiempo total en segundos	Peticiones fallidas
100	10	46.49	2.15	0
500	10	174.37	2.86	0
100	100	45.44	2.2	0

500	500	183.9	2.7	0
1000	100	310.76	3.2	0
1000	1000	245.55	4	0
2000	2000	313.98	6.3	0
15000	15000	503.15	39.8	0

6.2 Resultado

6.2.1 Rendimiento del Back-end

Las pruebas se han realizado con el endpoint para obtener las estadísticas, ya que es el que más datos obtiene y del que se reciben más peticiones.

El servidor es capaz de gestionar hasta 2.000 peticiones simultáneas en 5.23 segundos antes de que se sobrecargue. El número de peticiones a soportar para una escuela es apto, ya que el número de profesores es muy difícil que llegue a los 2.000 en un mismo centro.

Asimismo, para una universidad grande, el sistema debe ser capaz de soportar muchas más peticiones, ya que se trabaja con un volumen de datos mucho mayor. En esta situación, sería conveniente el uso de balanceadores de carga para distribuir las peticiones en diferentes servidores.

6.2.2 Rendimiento del Front-end

Las pruebas para el servidor HTTP se han hecho del portal de inicio de sesión, ya que es el único que se carga si necesidad de datos guardados en la base de datos.

En este caso, el sistema es capaz de soportar un mayor número de peticiones simultáneas, empezando a sobrecargarse a partir de las 15.000 necesitando un total de 39.8 segundos.

Con estos resultados, podemos decir que dedica $39.8/15.000 = 0.002$ segundos de media por petición así que se puede decir que, en caso de muchas peticiones, el servidor es capaz de adaptarse.

7 CONCLUSIÓN

Como conclusión, puedo decir que se ha desarrollado de forma satisfactoria un sistema capaz de adaptarse a cualquier escuela y que permite hacer un uso de la tecnología para dar todo tipo de información y así ayudar a mejorar el sector de la educación.

El término de “seguridad informática” no siempre tiene que estar asociado a aspectos negativos, ya que, como se ha hecho en este proyecto, la seguridad informática nos puede proporcionar una gran cantidad de datos muy útiles para poder mejorar nuestras tareas del día a día.

Se ha conseguido crear un proyecto que destaca por ser completo, ya que utiliza diferentes ámbitos trabajados durante el grado como seguridad, administración de redes, programación web, business intelligence, bases de datos y creación de API.

Todos los objetivos marcados al inicio del documento se han podido cumplir, aunque finalmente no ha sido posible la implementación de un portal cautivo para que los alumnos puedan utilizar distintos dispositivos con tan solo iniciar sesión.

Aunque se ha realizado todo el desarrollo, no se ha podido implantar en un entorno real para ver como se adaptaría finalmente tanto en recolección de datos, en la gestión de las actividades y en fluidez en el rendimiento.

8 PLANES FUTUROS

9.1 Almacenamiento en Cloud

Para el futuro, esta planeado mover los datos en la Nube para hacer posible instalar el sistema en múltiples escuelas y controlarlo desde un punto central.

Aunque el sistema se ha desarrollado como prototipo y no es un sistema en estado comerciable, se ha hecho pensando en los requisitos que podrían requerir la instalación de este en muchos centros educativos.

El hecho de tener los datos y sistemas en la Nube permite definir un modelo de alta disponibilidad para asegurar que los usuarios de todos los centros puedan acceder en todo momento a la plataforma web.

De esta manera surge la necesidad de separar el perfil “administrador” en dos: Administrador general y administrador del centro.

El administrador general es el encargado del mantenimiento de los sistemas en la nube (mongoDB, Docker para separar los entornos de cada escuela y servidores) para asegurar el correcto funcionamiento de este. Este no pertenece a una escuela en concreto sino a la empresa creadora del proyecto, Security System.

El administrador del centro se encarga de establecer las configuraciones personalizadas de su centro y los profesores solamente pueden gestionar las actividades, calificarlas y ver las estadísticas de su centro.

AGRADECIMIENTOS

Quiero dar las gracias a mi tutor, Xiaoyuan Yang, por el seguimiento continuado durante todo el proyecto y por el apoyo y comprensión que me ha ofrecido en un mal momento personal.

BIBLIOGRAFIA

- [1] pfSense. (n.d.). PfSense. Retrieved March 14, 2021, from <https://www.pfsense.org/>
- [2] Getting Started -. (n.d.). React. Retrieved March 14, 2021, from <https://reactjs.org/docs/getting-started.html>
- [3] Documentation - The Go Programming Language. (n.d.). Go-Lang. Retrieved March 14, 2021, from <https://golang.org/doc/>
- [4] MUI: The React component library you always wanted. (s. f.).

- MUI. <https://mui.com/>
- [5] Stack Overflow - Where Developers Learn, Share, & Build Careers. (s. f.). Stack Overflow. <https://stackoverflow.com/>
- [6] Pentaho Data Integration. (s. f.). Hitachi Vantara Lumada and Pentaho Documentation. https://help.hitachivantara.com/Documentation/Pentaho/7.1/0D0/Pentaho_Data_Integration
- [7] Trello. (s. f.). Trello. <https://trello.com/b/hhSIik5L/tfg>
- [8] GitHub: Where the world builds software. (s. f.). GitHub. <https://github.com/>
- [9] 2019 MUI Developer Survey results - MUI. (s. f.). MUI Developer Survey. <https://mui.com/blog/2019-developer-survey-results/>
- [10] ab - Apache HTTP server benchmarking tool - Apache HTTP Server Version 2.4. (s. f.). Ab - Apache HTTP Server Benchmarking Tool. <https://httpd.apache.org/docs/2.4/programs/ab.html>
- [11] A Comparison of Data Visualization Libraries for React. (s. f.). Capital One. <https://www.capitalone.com/tech/software-engineering/comparison-data-visualization-libraries-for-react/>
- [12] DB-Engines Ranking. (s. f.). DB-Engines. <https://db-engines.com/en/ranking>
- [13] Dhaduk, H. (s. f.). Bootstrap vs. Material: Which is the True Frontend Maestro? Insights on Latest Technologies - Simform Blog. <https://www.simform.com/blog/bootstrap-vs-material/#section2>
- [14] Go Developer Survey 2019 Results - The Go Programming Language. (s. f.). GO Surveys. <https://go.dev/blog/survey2019-results>
- [15] How to Use Go With MongoDB? (s. f.). GeeksforGeeks. <https://www.geeksforgeeks.org/how-to-use-go-with-mongodb/>
- [16] PostgreSQL vs MySQL. (s. f.). Sumo Logic. <https://www.sumologic.com/blog/postgresql-vs-mysql/>
- [17] Smallcombe, M. (s. f.). PostgreSQL vs MySQL: The Critical Differences. Integrate.io. <https://www.integrate.io/blog/postgresql-vs-mysql-which-one-is-better-for-your-use-case/>
- [18] Statista. (s. f.-a). Global sought-after database skills for developers 2021. <https://www.statista.com/statistics/793854/worldwide-developer-survey-most-wanted-database/>
- [19] Statista. (s. f.-b). Most popular web frameworks among developers worldwide 2021. <https://www.statista.com/statistics/1124699/worldwide-developer-survey-most-used-frameworks-web/>
- [20] Estrategia de digitalización del sistema educativo. (s. f.). Ministerio de Educación y Formación Profesional. <https://www.educacionyfp.gob.es/prensa/actualidad/2021/07/20210730-cumplendosegundoinforme.html>
- [21] El año de los grandes ciberataques en España. (s. f.). ELMUNDO. <https://www.elmundo.es/tecnologia/2021/12/01/61a63b4ae4d4d8db5a8b4577.html>
- [22] García, C. (s. f.). España es el país en el que menos se fomenta el uso de la tecnología en el aula. elEconomista.es. <https://www.eleconomista.es/ecoaula/noticias/10853662/10/20/Espana-es-el-pais-en-el-que-menos-se-fomenta-el-uso-de-la-tecnologia-en-el-aula.html>
- [23] BOE.es. (s. f.). BOE. <https://www.boe.es/buscar/doc.php?id=BOE-A-2018-16673>
- [24] Abraham, A. W. |. (s. f.). Angular vs React vs Vue. Medium. <https://medium.com/somoswigou/angular-vs-react-vs-vue-cu%C3%A1l-es-la-mejor-opci%C3%B3n-941a207951c7>
- [25] 2019 Database Trends. (s. f.). ScaleGrid. <https://scalegrid.io/blog/2019-database-trends-sql-vs-nosql-top-databases-single-vs-multiple-database-use/>
- [26] Contact, M. (s. f.). Escuelas, las más afectadas por ciberataques, con un 44%. mundocontact. <https://mundocontact.com/escuelas-las-mas-afectadas-por-ciberataques-con-un-44/>
- [27] Prácticas, O. (s. f.). Crecen los ciberataques en los centros educativos. Delfos. <https://www.delfossistemas.com/crecen-los-ciberataques-en-los-centros-educativos/>
- [28] Frontend. Youtube. www.youtube.com/watch?v=Xqd26iCrEYc

APÉNDICE

A.1 Vista de la interfaz web

La vista del portal web contiene un gran número de ventanas así que en este documento solamente se recogen las principales.

Escaneando el siguiente código se observa en un video todas las diferentes ventanas con sus funcionalidades:



Fig. 10 Código QR para ver toda la parte front-end [28]

5.1.1 Inicio de sesión

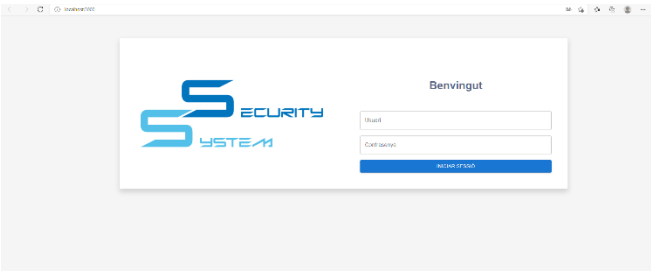


Fig. 11: Vista del inicio de sesión del portal web

5.1.3 Gestión de usuarios

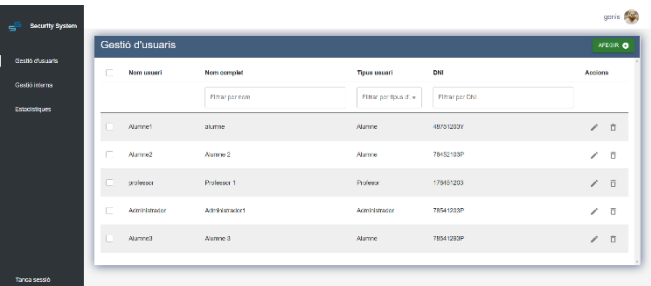


Fig. 12: Vista de la gestión de usuarios del portal web

5.1.3 Gestión interna de los cursos

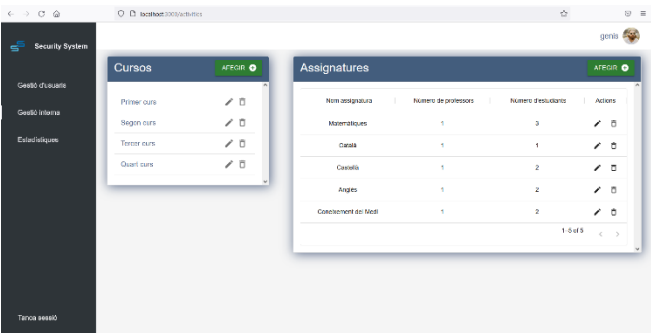


Fig. 13: Vista de la gestión interna de la escuela

5.1.4 Gestión de las actividades

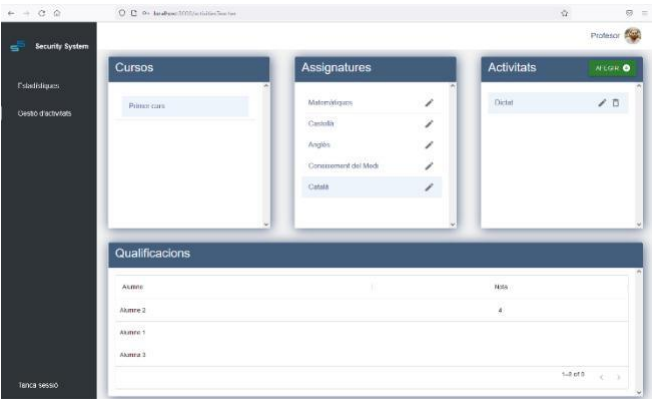


Fig. 14: Vista de la gestión de actividades y calificaciones

5.1.4 Visualización de las estadísticas



Fig. 15: Vista de las estadísticas

A.2 Flujo de datos

En el siguiente diagrama se muestra el flujo de actividades que sigue el programa:

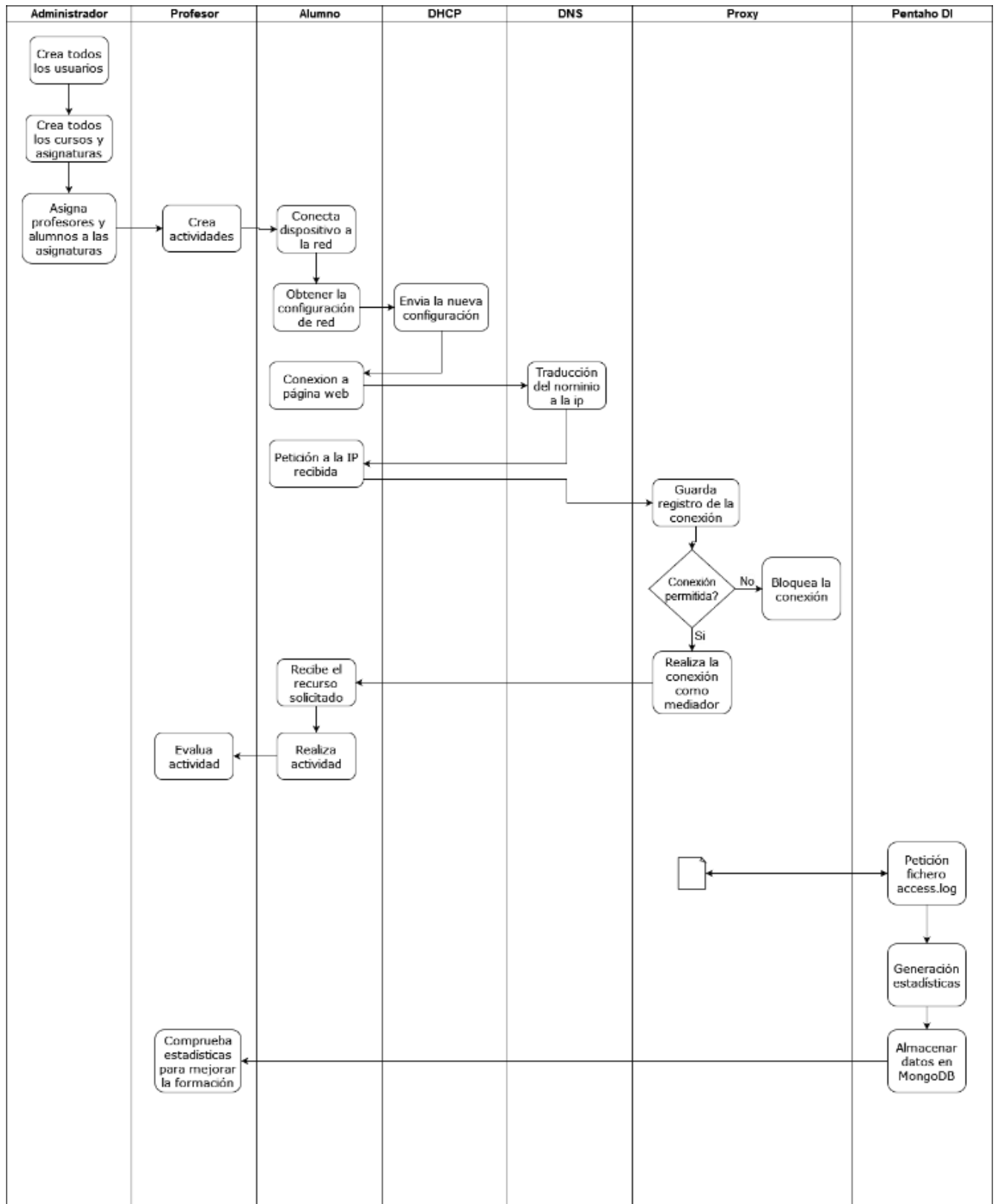


Fig. 16: Diagrama de flujo de las actividades en el sistema