Universidad de La Habana
Facultad de Matemática y Computación

Implementación de un sistema de autenticación de usuarios basado en LDAP

Autor: Eric Nordelo Galiano
Lian Ulloa Mc-kion

Tutor: **Lic. Gilberto Garcia**

Tutor: **Lic. Darío Álvarez Arteaga**

Trabajo de Diploma
presentado en opción al título de
Licenciado en Ciencias de la Computación

Junio de 2019

Agradecimientos

Opinión del tutor

Resumen

Abstract

Índice general

Índice de figuras

Introducción

La necesidad de desarrollar los medios de comunicación y la transmisión de información es una de las características que distingue a nuestra civilización desde hace milenios. Desde que se inventó la imprenta hace aproximadamente 6 siglos, la humanidad ha generado incontables volúmenes de texto para almacenar conocimiento de todo tipo. No obstante, no es hasta la década de 1960, con la adopción y proliferación de las computadoras y el mantenimiento de registros digitales, que se alcanzaría una verdadera revolución en los medios de almacenamiento, transmisión y en la accesibilidad a la información a escala global. Esta revolución conocida como Revolución Digital.º .Era de la Información"gira en torno a las nuevas tecnologías e Internet, siendo la catalizadora de una enorme explosión tecnológica, produciendo grandes transformaciones en una sociedad donde la automatización de los procesos ya no solo industriales, tambien cotidianos, mejora considerablemente la calidez, rapidez y robustez de los mismos, y donde el surgimiento de las redes de dipositivos facilita la comunicación y la organización de personas y organismos. Una de las ventajas que trae consigo el establecimiento de redes de dispositivos es la posibilidad de descentralizar el almacenamiento de información sin comprometer el acceso a la misma. Desde la propuesta de Leonard Kleinrock en el año 1961 en un artículo titulado `"Information Flow in Large Communication Nets"` (traducido ha `"Flujo de Información en Grandes Redes de Comunicación"`), y el uso del término `"paquete"` en 1965 por Donald Davies para describir datos enviados entre computadoras en una red, que impulsarían el desarrollo de ARPANET entre 1966 y 1969, el planeta entró en una nueva etapa que revolucionó el desarrollo en todos los campos tecnológicos: la `"Era Digital"` o `"Era de la Información"`. Esta era gira en torno a las nuevas tecnologías e Internet y está llevando a cabo cambios profundos y transformaciones en una sociedad donde la automatización de los procesos mejora considerablemente la calidad, rapidez y robustez de los mismos, y donde la conectividad mediante las redes de dispositivos (no puede hablarse solo de computado-

Lleva referencia

Lleva referencia

Fue la invención del término lo que provocó el impulso de ARPANET? Además, esto realmente viene al caso?

ras) facilitan la comunicación y la organización tanto de personas como de empresas y organismos.

Las oraciones en el párrafo anterior son largas y la idea detrás de ellas es dificil de seguir. Todo lo que este en un formato similar al del párrafo anterior debe ser separado en oraciones cortas y conexas.

Una de las ventajas que trae consigo el establecimiento de redes de dispositivos es la posibilidad de descentralizar el almacenamiento de información manteniendo un adecuado acceso a la misma. Actualmente la red de la Universidad de La Habana cuenta con un gran número de servicios consumidores o fuentes de diferentes tipos de información, pero carece de un servicio que `"administre"` el flujo de comunicación con los mismos. De estos servicios se benefician no solo las facultades pertenecientes a la Universidad, también se benefician otras instituciones asociadas a la misma como el Instituto Superior de Diseño (ISDI) y el Instituto Superior de Ciencias y Tecnologías Aplicadas (InSTEC). La gran mayoría de estos servicios requiere que el la verificación previa del usuario que solicita acceder o agregar información a los mismos, por lo que con el objetivo de centralizar el proceso de verificación se han ido acumulando sobre un dominio web (directorio.uh.cu) varias actualizaciones aisladas que conforman lo que se conoce como "Directorio Único de la Universidad de La Habana"para poder integrar y administrar servicios que funcionan sobre diferentes tecnologías. Esto trae consigo dificultad a la hora de integrar nuevos servicios al directorio, así como inestabilidad y falta de robustez en el servicio de verificación y solicitud de información almacenada.

El párrafo del la explicación del problema debería ser re escrito para su mejor comprensión.

Dicho esto, el objetivo de este proyecto es proveer a la Universidad de La Habana de un servicio que se encargue de administrar estos procesos de manera confiable y eficiente, que además sea fácil de integrar y fácil de modificar (implementar mejoras).

Para lograr esto surgen varias preguntas a responder como por ejemplo: ¿qué método se debería utilizar para almacenar la información?, ¿qué método se debería utilizar para acceder a la misma?, ¿qué método usar para mantener la informacón actualizada?...

Como respuesta a la primera pregunta, después de analizar distintos métodos para gestionar información, debido a las características del problema que enfrentamos llegamos a la conclusión que lo mejor es almacenarla en una estructura de directorios. La principal ventaja de esto sobre las bases de datos relacionales (y no relacionales) es la velocidad para realizar consultas

Recomiendo cambiar por: 'sin comprometer el acceso'

gran número?

Ahora mismo no lo hacen, parte del objetivo de esta Tesis es sentar las bases para que lo hagan.

Agregar informacion no requiere de esta autenticación

Reescribir por favor.

Si vas a usa esto como un objetivo general, 'estos procesos' debe ser claramente

sobre grandes volúmenes de datos (perdiendo en velocidad de modificación de los mismos, tanto en actualizaciones, como en inserciones y eliminaciones).

Decidido esto y después de analizar diversas maneras de manejar una estructura de directorios optamos por el LDAP (Lightweight Directory Access Protocol o Protocolo Ligero de Acceso a Directorios) como protocolo de acceso y OpenLDAP como implementación del mismo.

Este protocolo brinda un esquema similar al de una guía telefónica, implementando un enfoque jerárquico para el almacenamiento de la información. Además existe una interfaz de autenticación para la mayoría de los servicios que administran usuarios y permisos, que utiliza como fuente de información a servidores que implementen este protocolo. Partiendo de la problemática existente la autenticación de los servicios que se brindan en el recinto universitario, se ha formulado la siguiente hipótesis de investigación: mediante el sistema de autenticación basado en protocolo LDAP se solucionará la administración eficiente de la información y los servicios en la Universidad de La Habana.

Con esta tesis queremos sustituir el actual sistema de directorio único por un sistema de autenticación basado en el protocolo LDAP.

Para esto es necesario investigar sobre el estado del arte de los sistemas que implementan este protocolo, así como la disponibilidad que brindan. Además es necesario automatizar el proceso de recopilar la información provista por estas fuentes, para lo cual se diseñará una esquema de base de datos, guiado por el protocolo LDAP, que se adapte a la estructura de los datos almacenados actualmente. Por último, se plantea implementar una *API* acorde con el protocolo OpenId para facilitar la autenticación para futuros servicios.

Evitemos usar **Debido a esto** y **Decido esto**, se puede reformular como **Habiendo analizado las ventajas de usar una estructura de directorio se decidió ...**

En la introducción no toca explicar el protocolo LDAP. Tiene que haber una sección del capitulo 2 para explicar el concepto de LDAP y como funciona.

Esto esta aqui para plantear algun objetivo?

Generar la hiposteis a partir

Notes

Capítulo 1

Directorio Único

1.1. Surgimiento del Directorio Único

Como resultado de la necesidad de acceso a la información de los usuarios de la red de la Universidad de la Habana, tanto de estudiantes, como profesores y externos, varios servicios han sido implementados de manera aislada, durante los últimos años con el fin de suplir esta necesidad. A raíz del proceso de centralización de los servicios llevado a cabo hace ya tres años, dichos servicios, relacionados con la creación, almacenamiento y la actualización de la información referente a los usuarios, fueron integrados en lo que se conoce hoy como Directorio Único.

Entre las principales funcionalidades que componen el Directorio Único, se encuentra la de constituir una fuente, centralizada, de información sobre toda aquella persona que pretenda hacer uso de las facilidades disponibles en la red universitaria. Dicha centralización representa un elemento necesario para garantizar el control apropiado del uso de dichas facilidades, ya que la fragmentación de los datos, inherente a la estructura actual de la Universidad y a sus actividades, dígase la composición por facultades, departamentos de investigación, etc, así como la continua realización de eventos para fomentar la investigación y el intercambio de experiencias, aumenta considerablemente la complejidad a la hora de administrar la información almacenada y de garantizar cierta restricción o concesión de acceso a los servicios brindados.

Sin embargo, esta información no se genera de forma centralizada, sino que se encuentra esparcida entre distintas fuentes como son el SIGENU y una base de datos conocida como 'Assets', la cual se encuentra esparcida en varios servidores y sobre la cual no se tiene total acceso debido a que, para su creación y luego su administración, se utiliza un software privativo,

el cual impide el acceso requerido para implementar las actualizaciones que se hacen necesarias en este momento. Por esta razón, para poder actualizar esta información centralizada, todos los datos existentes en las fuentes antes mencionadas, son procesados a través de un sistema que se compone de alrededor de 3 o 4 capas de procesamiento. Entre las principales razones que justifican la existencia de estas capas, se encuentra la incapacidad de agregarle directamente a las fuentes de datos, un campo que indique en que momento una instancia de usuario ha sido modificada.

Cuáles son los nombres oficiales de las fuentes de datos?, Qué significa SIGENU? si es que son siglas

Entre los datos más relevantes, almacenados en este sistema, se encuentra aquellos que permiten la restricción y/o concesión de acceso a los servicios brindados, es decir, las credenciales de los usuarios de la red. Esta información permite:

1. **Autenticar al usuario:** Comprobar que la persona que solicita un servicio es quien dice ser.
2. **Administrar el acceso:** En correspondencia del rol de la persona autenticada, permitir o no el acceso a ciertos servicios

Del resto de los datos que actualmente se mantienen en el sistema, una parte se reserva para brindar información acerca de la actividad realizada por el usuario durante su estancia en la red, es decir, los momentos en que se autenticó en la red, en que sistema lo hizo, la cantidad de cuota de internet consumida, etc. La otra parte, estando esta última más asociada a los trabajadores y externos, contiene información más personal de los usuarios. Entre la misma sobresalen:

1. Año que cursa (en caso de ser estudiante)
2. Dirección Particular
3. Departamento al que pertenecen
4. Datos sobre la nomina
5. Puesto que ocupa
6. Cargos importante (si es que los posée)

Esta información es requerida principalmente por los departamentos responsables de la administración de los recursos humanos de la Universidad de la Habana.

1.2. Problemas del sistema actual

El sistema, tal y como existe en este momento, presenta varios problemas, a lo cuales esta tesis pretende brindar una posible solución. Los problemas son presentados a continuación:

1.2.1. Mantenimiento y extensión del Sistema

Debido a la naturaleza del surgimiento del Directorio Único, es decir, el acoplamiento de varios servicios aislados en uno, cada componente es demasiado dependiente de la forma en que las demás brindan sus correspondientes funcionalidades. Esto se debe a que la interacción entre las mismas ha sido configurada mediante un enfoque estático , y que el propio sistema no dispone de herramientas que permitan su modificación de una manera cómoda para los encargados de su mantenimiento. Dicho enfoque dificulta enormemente las tareas de actualización del sistema, las cuales son necesarias para poder adecuar el mismo a las nuevas condiciones y necesidades que van surgiendo en la red a través de los años.

Mi intención aquí es decir que estan cableadas pero realmente no se como decirlo

1.2.2. Desuso de datos almacenados

Desde el surgimiento del Directorio Único, se han ido incorporando nuevos campos al diseño de las bases de datos que representan las fuentes de información del sistema, con el fin de suplir las necesidades que ocupan a la Universidad en cada nuevo período escolar.

Actualmente muchos de esos campos han dejados de ser útiles para la Universidad, pero a pesar de esto se mantiene en el sistema debido a que, como consecuencia de la rigidez del mismo, cualquier cambio, sobre todo aquellos cuya repercusión y alcance no se conocen, podrían significar la caída del sistema por tiempo indefinido, de ahí que se siga la filosofía de que "... lo que funciona no se toca...", provocando una sobre carga innecesaria para el sistema, que aunque pueda ser pequeña, no deja de ser significante, reflejándose sobre todo en el espacio ocupado por la información en disco.

1.2.3. Incumplimiento de políticas de baja de usuarios

Este es otro problema en el cual se incurre con bastante frecuencia en la Universidad. Debido a la volatilidad de algunos contratos concertados con personal ajeno a nuestro centro de altos estudios, sucede frecuentemente que al dar de baja a estos usuarios, sus cuentas son eliminadas con efecto casi inmediato, lo cual va en contra de los protocolos usualmente implementados

en estos casos, los cuales son necesarios para prevenir la pérdida de acceso a servicios críticos, muchas veces desarrollados, administrados y mantenidos por agentes externos a la Universidad.

1.2.4. Carencia de réplicas sincronizadas

Actualmente el Directorio Único no cuenta con réplicas que puedan garantizar el continuo funcionamiento del sistema en caso de que el servicio principal se detenga por causa de algún problema en el mismo. Esto causa la imposibilidad de acceder a gran parte de los servicios pertenecientes a la red universitaria, afectando la productividad de la misma. Este problema no se debe solo a la falta de réplicas del servicio actual, si no también a la falta de alternativas para este servicio, ya que actualmente no existe ninguna funcional, a pesar de que el Directorio Único lleva fallando ya un buen tiempo.

1.2.5. Carencia de protocolos comunes para servicios externos

Resulta natural que, con el paso del tiempo, se haga necesaria la disponibilidad de nuevos servicios en la red de la Universidad. Generalmente estos servicios, presentan como requerimiento común, aquel asociado a la autenticación de usuarios, así como a la administración de roles para los mismos. Por esta razón, esta tarea se suele asignar a un sistema externo y centralizado con el objetivo de evitar que se repita este proceso en el desarrollo de cada servicio. Generalmente es implementado utilizando uno de los protocolos desarrollado por las entidades encargadas de estandarizar ciertos procesos en el amplio campo de la ciencia de la computación. Teniendo esto en cuenta, es que se pretende implementar una API Rest que permita modificar la lógica detrás de la información brindada, sin que esto implique modificar todos los servicios que consuman información de nuestro sistema.

Capítulo 2

Estado del arte

2.1. LDAP

2.1.1. ¿Qué es el protocolo LDAP?

LDAP (Lightweight Directory Access Protocol) es un protocolo perteneciente a la capa de aplicaciones, tanto para servidores como para clientes. Es abierto y multiplataforma. Está pensado para la implementación de servicios de directorio, facilitando el acceso rápido a la información almacenada. Presenta una estructura arbórea, la cual organiza la información en ramas y permite realizar búsquedas de manera eficiente, debido a que la cardinalidad de las posibles repuestas se reduce a medida que se avanza por cualquiera de estas ramas. Es una versión ligera del protocolo DAP (Directory Access Protocol), el cual a su vez es parte del estandar para servicios de directorios en la re X.500.

Me falta poner referencias para DAP y X.500

2.1.2. Conceptos Importantes dentro de LDAP

Servidor de Directorio

Un servidor de directorio, no es más que un tipo de base de datos pensada para ser utilizada directamente en la red. A diferencia de las bases de datos tradicionales (Bases de Datos Relacionales) que representan los datos en tablas y cada instancia es una fila, en este cada entrada en el directorio es un árbol de entradas, donde cada árbol puede contener datos o ser una hoja (un árbol vacío)

Entradas

Cada entrada en un servidor de directorio representa una colección de información referente a cierta entidad. Está compuesta principalmente, por un nombre distinguido, que es el identificador unívoco de la misma. Además cuenta con un conjunto de atributos y de clases de objetos los cuales definen la estructura y el comportamiento de la entrada.

Distinguished Name (Nombre Distinguido)

Este es el identificador unívoco de la entrada. Esta compuesto por lo que se conoce en la literatura como 'Nombres distinguidos relativos' o 'RDN' por sus siglas en inglés. Estos RDN no son más que un conjunto ordenado de pares atributo-valor. Usualmente se escogen los atributos más representativos de cada entrada para la representación del DN.

Artibutos

Los atributos son los encargados de guardar la información de cada entrada y tiene asociados un tipo, un conjunto de opciones.

Los atributos representan una parte importante del esquema del directorio LDAP. A través de estos podemos definir nuevas clases de objetos para poder suplir las necesidad de almacenamiento de información. Para poder definir tanto atributos como clases de objetos es necesario proveerle a ambos un identificador, el cual presenta un formato similar al siguiente: 1.3.6.1.4.1.<Identificador global>.1.5 . El identificador global al que se hace referencia, no es más que un número de series que distingue a la implementación del protocolo LDAP utilizada a nivel global. Este se puede obtener realizada una solicitud a IANA(Internet Assigned Numbers Authority).

Clases de Objetos

Estos también representan una parte importante del esquema del protocolo LDAP. No son más que conjunto de atributos que definen la información almacenada en cada entrada. Pueden ser de dos tipos: estructurales o auxiliares. Cada entrada puede tener asociada una clase de objetos estructural y cero o más clases auxiliares.

Filtros

Los filtros representan el mecanismo utilizado para realizar consultas al directorio. La lógica utilizada para filtrar las entidades almacenadas en el

servidor se define a través de reglas de comparación, las cuales, a su vez se definen en los atributos.

2.1.3. Implementaciones más utilizadas

IBM Security Directory Server

Este servicio implementa las especificaciones de Internet Engineering Task Force (IETF) LDAP V3. Permite la comunicación con clientes basados en IETF LDAP V3. Esta alternativa presenta una amplia variedad de funcionalidades que facilitarían la integración con el sistema de la Universidad , pero esta herramienta es de pago, por lo que no podemos utilizarla.

poner referencia a la pagina de ibm https://www.ibm.com/supp

Active Directory

Esta es la implementación que brinda Microsoft del protocolo LDAP. Igualmente presenta una amplia variedad de funcionalidades pero también es de pago.

Oracle Internet Directory

Oracle Internet Directory is a general purpose directory service that enables fast retrieval and centralized management of information about dispersed users and network resources. It combines Lightweight Directory Access Protocol (LDAP) Version 3 with the high performance, scalability, robustness, and availability of an Oracle Database.

poner cita de la url https://docs.oracle.com/cd/

OpenLDAP

Esta es la implementación que estaremos, principalmente debido a que es totalmente gratis y se integra fácilmente al entorno de sistemas basados en Linux, el cual es la base de la mayoría de los servidores de la Universidad.

2.1.4. Modos de empleo usuales

DNS

LDAP es usualmente utilizado con una estructura de DNS. Las clases de objetos que existen por defecto en el esquema de OpenLDAP, permite simular la estructura de que presentan los DNS. Esto da la oportunidad de brindar las mismas funcionalidades de servicio de nombres de dominios y a la vez utilizar las ventajas de búsqueda y modificación de los LDAP.

Sistema de Autenticación

Esta implementación también brinda ventajas a la hora de implementar un sistema de autenticación de usuarios. Esto se debe principalmente al amplio soporte que tiene el protocolo LDAP para varios servicios. La posibilidad de agrupar a los usuario mediante unidades organizativas (Organizational Unit [OU]) y de representar su pertenencia a determinados grupos, permite fácilmente administrar el acceso que cada uno debe tener a los servicios ofrecidos por la universidad. Este modo de organizar la información de los usuarios se asemeja bastante a la manera en que se asigna permisos a un usuario en los sistemas operativos basados en Linux, de hecho, una de las funcionalidades implementadas para este protocolo permite autenticar un usuario en una máquina, ya sea virtual o física, siempre y cuando este exista en el servidor LDAP.

2.2. Docker

2.3. EA3

2.4. EA4

2.5. EA5

Capítulo 3

Implementación del sistema

3.1. Consideraciones previas

3.2. SI2

3.3. SI3

3.3.1. SI3.1

3.3.2. SI3.2

3.3.3. SI3.3

3.4. Desarrollo de la solución

Capítulo 4

Experimentación y Resultados

Recomendaciones

Conclusiones