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

Directorio Único es un servicio, que fue implementado con el objetivo de acceder a los datos del personal de la Universidad de la Habana. En el momento de su surgimiento, se hacia necesario disponer de un sistema que unificara las principales fuentes de datos de trabajadores y estudiantes. De esta manera sería factible presentarlas a través de una misma interfaz. Desde su creación se han implementado, sobre dicha interfaz, varios servicios de manera escalonada. Entre los principales servicios implementados, podemos destacar la implementación de un mecanismo de autenticación OAuth para los sitios de la intranet. Todos los servicios implementados sobre directorio funcionan hasta hoy. Otros servicios permiten denegar o permitir el acceso de los usuarios a determinados recursos brindados por la Universidad. Se puede tomar como ejemplo, el acceso al servicio que administra la asignación de viajes internacionales al personal de la Universidad. Otro ejemplo, un poco más palpable, lo tenemos en el servicio que decide la cuota de internet asociada a cada usuario. Este se basa tanto, en el año que cursa, en caso de ser estudiante, como en el cargo que ocupa o departamento en que trabaja, en el caso de los trabajadores.

Todos estos servicios, se implementaron sobre una misma base, debido al proceso de centralización iniciado hace ya 4 años. Dicha centralización nos provee de ciertas ventajas, ya que representa una fuente de información de todo el personal de la Universidad de La Habana. Con este enfoque, la información queda más accesible y más fácil de administrar.

Sin embargo, esta información no se genera de forma centralizada, sino que se encuentra esparcida entre distintas fuentes. Primeramente contamos

con el Sistema de Gestión para la Nueva Universidad (SIGENU). La información almacenada en esta, es administrada por las secretarias de las respectivas facultades y es referente solo a los estudiantes. También se consume información de los trabajadores de la Universidad de La Habana, la cual se maneja por los departamentos de recursos humanos de cada una de las unidades presupuestadas de la Universidad (UH, IFAL, JBN, UPA, ISDI, INSTEC). Sobre esta 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.

Que actualizaciones? Assets es un software para administracion de RH, que otras funcionalidades quieren implementar sobre esta db?

Uno de los principales problemas que provoca esta falta de acceso, es la incapacidad de agregarle directamente a las fuentes de datos, un campo que indique en que momento una instancia de usuario ha sido modificada.

Es verdad, pero nuevamente, esta idea no parte del uso del sistema original sino de adaptarlo a nuestras condiciones especificas. Bien las secretarias de RH podrian decir que la información de un trabajador se actualizó en un sistema aparte a Assets luego de haber introducido el cambio en Assets

Entre los datos más relevantes, almacenados en el Directorio Único, se encuentran 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 nivel de privilegio de un usuario ,permitir o no el acceso a ciertos servicios.

Una parte de la información almacenada, contiene información personal de los usuarios, estando esta más asociada a los trabajadores y externos. 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)

A los datos que actualmente se mantienen en el sistema, se le asocia información adicional para conocer la actividad realizada por el usuario. Dicha información, la almacenad el nodo de la Universidad en otro sistema aparte. Durante su estancia en la red, podemos registrar los momentos en que se autentica , en que sistema lo hace, la cantidad de cuota de internet consumida, etc.

En el listado siguiente, se encuentra un ejemplo de una posible respuesta ofrecida por el directorio cuando se consultan los datos de un trabajador de la UH.

```
<TrabajadorInfoCuote>
<Id>15869</Id>
<CatOcupacional>técnicos docentes principal</CatOcupacional>
<Docente>Si</Docente>
<CatDocenteInvestigativa>Instructor</CatDocenteInvestigativa>
<Contrato>Indeterminado</Contrato>
<Cargo>INSTRUCTOR</Cargo>
<Adiestrado>No</Adiestrado>
<AdministradorArea>No</AdministradorArea>
<Tecnico>Si</Tecnico>
<TecnicoInformatico>No</TecnicoInformatico>
<EspecialistaPrincipal>No</EspecialistaPrincipal>
<Cuadro>No</Cuadro>
<Asset>1</Asset>
<Departamento>DIRECCION DE INFORMATIZACION</Departamento>
</TrabajadorInfoCuote>
```

1.2. Problemas del sistema actual

El sistema, tal y como existe en este momento, presenta varios problemas. Esta tesis pretende brindar una propuesta de solución, así como su implementación. 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 de manera escalonada sobre la idea inicial, 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. 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. De hecho, actualmente el personal encargado del mantenimiento de Directorio no puede responder a las necesidades de actualización. La principal causa de esta desatención, es que los desarrolladores de Directorio perdieron muchos de los cambios en su historial de código. Tan importante es la pérdida de este historial que imposibilita la recuperación de la lógica del Directorio actual partiendo solamente del código almacenado.

1.2.2. Desuso de datos almacenados

Desde el surgimiento del Directorio Único, se han ido incorporando nuevos campos a las fuentes de información del sistema. Estos cambios han tenido como objetivo 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. Como consecuencia de la rigidez del Directorio, 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...". Pero mantener este enfoque, provoca una sobrecarga innecesaria para el sistema, que aunque pueda ser pequeña, no deja de ser significante. Dicha sobrecarga se refleja 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 protoco-

los usualmente implementados en estos casos. Generalmente se debe espera una cierta cantidad de días para implementar la eliminación total de las cuentas. De esta manera se puede prevenir la pérdida de acceso a servicios críticos, como son el correo, el proxy y la nube recientemente despliega en la intranet de la Universidad. Muchas veces estos servicios son desarrollados, administrados y mantenidos por agentes externos a la Universidad.

1.2.4. Carencia de protocolos comunes para servicios externos

A menudo, se implementan nuevos servicios en la red de la Universidad. Generalmente estos servicios necesitan tener control de acceso sobre los recursos que brindan a sus usuarios. Esto implica el tener que desarrollar para cada nuevo servicio, un mecanismo de autenticación de usuarios. Este mecanismo además tiene que ser capaz de brindar una funcionalidad para administrar los roles o grupos a los que pertenecen dichos usuarios. Un enfoque más útil, es el de delegar esta tarea a un sistema externo y centralizado. De esta forma se evita el tener que repetir el desarrollo de la misma funcionalidad para cada servicio.

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

Esta nota es de conexion entre el capitulo anterior y este. Se necesita una idea lógica para conectar los dos capítulos. Al final del capítulo anterior se describió la situación del directorio único en la UH y de pronto en este capítulo caemos en el análisis de las herramientas para dar la solución ... falta algo: nunca se habló de cual iba a ser la propuesta de solución (no cuenta la breve descripción dentro de la introducción). Es necesario destinar un epígrafe al final del capítulo anterior proponiendo la idea de la solución a implementar utilizando un protocolo de acceso a directorios. En este capítulo se discuten las herramientas con las que se puede llevar a cabo: LDAP y active directory como candidatos del protocolo de acceso a directorios; python y adversarios como lenguajes de base para la implementacion del api; django vs flask como frameworks alternativos para la implementacion del api. Así, cuando concluya este capítulo le debe quedar claro al lector cuales son las herramientas que se van a utilizar para orquestar la solucion al igual que los conceptos fundamentales de dichas herramientas. OJO-nota: en el párrafo de arriba utilicé oraciones impropias para una tesis, son tan largas que la idea se pierde. Son suficientes para transmitirles una idea de manera informal a Uds como estudiantes, pero sirven para hacer llegar el mensaje de manera formal a un tribunal científico. Finalmente, el capitulo 3 es la implementación en detalle de la solución.

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.

Poner referencia al tanembaum por hablar de las capas del modelo OSI.

Poner ejemplos de servicios de directorios para que quede claro

Me falta poner referencias para DAP y X.500. Acaben de empezar a manejar las referencias, para luego es muy tarde.

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

Referencia al articulo de Codd para las 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)

Usar solo Nombre Distinguido para ser consecuente con el resto de los encabezados de sección.

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 .

No me queda claro como se ve el identificador. Una imagen vale mas que mil palabras. Toda esta explicación se vería mejor a través de un ejemplo.

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).

Es importante que en el capítulo de implementación se explique como solicitamos nuestro identificador global a IANA (Lian hizo esto si no recuerdo mal.)

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.

poner referencia a la pagina de ibm
https://www.ibm.com/support/knowledgecenter/en/SSVJJU_6.3.1/com.ibm.IBMDS.doc_6.3.1/admin_
... Acaben de poner las referencias.

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.

Active Directory

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

Poner al menos una referencia a la página de manual. Esta explicación esta pobrecita, hay qye hablar un poco mas de las diferencias de AD de Microsoft con respecto al resto de las alternativas.

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

que utilizaremos ... pero esto no se pone aun, aca solo estas hablando de las alternativas y de las diferencias entre ellas .... y hay que demostrar conocimiento. .

, 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

no solo vale decir la estructura que presentan los DNS ... una estructura de delegación de zonas, arborea, similar a la del DNS... sería una variante mas acorde para describir la relación

. 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.

TODO: Revisar la longitud y ajuste al tema de las oraciones. Argumentar con ejemplos, idealmente con imagenes también. Argumentar la selección de la herramienta OpenLDAP. Abordar el análisis del lenguaje para la implementación del API. Argumentar cual es el framework utilizado para implementar el API. Descripción completa de la selección de herramientas para implementar la solución.

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

4.1. Configuración

4.2. Experimentación

4.3. Metodología para medir rendimiento de Squid

Recomendaciones

Conclusiones