



PONTIFICIA UNIVERSIDAD JAVERIANA
FACULTAD DE INGENIERIA
DEPARTAMENTO INGENIERIA DE SISTEMAS

Proyecto Introducción a Sistemas Distribuidos

Período Académico 2021-3

Sistema Distribuido de Bolsa de Empleo

Objetivos

- Poner en práctica conceptos de sistemas distribuidos en un problema práctico.
- Utilizar patrones de comunicación síncronos y asíncronos.
- Resolver problemas presentes en sistema distribuidos, tales como fallas de los componentes y actualización de réplicas.

Descripción del Sistema a Desarrollar

El objetivo del proyecto es desarrollar un sistema de empleo en el que los empleadores y los aspirantes puedan suscribirse para recibir notificaciones de la siguiente manera:

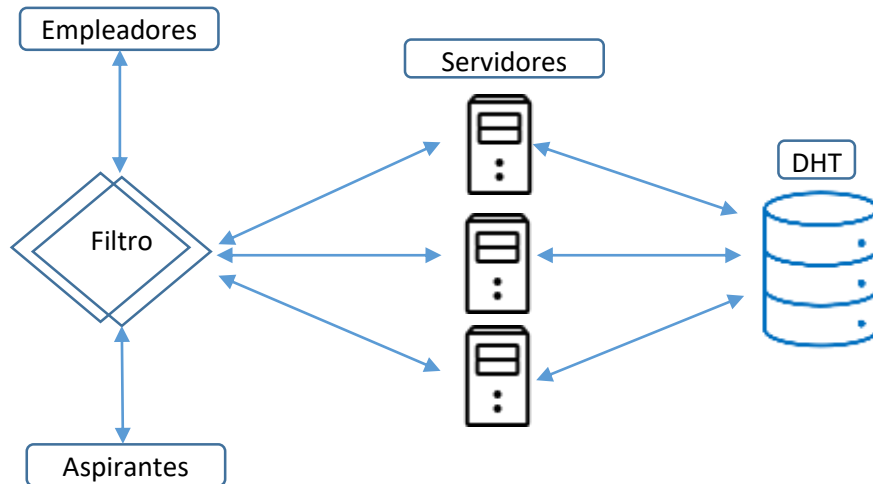
- Un empleador generar una o varias ofertas de trabajo. Se debe almacenar el código del empleador y el código de la oferta, para poder darle respuesta.
- Cada oferta de trabajo será filtrada de acuerdo con el sector al que corresponda, según la “Clasificación única de ocupaciones para el trabajo”¹ Se puede limitar a cinco sectores.
- Después de filtrar el tipo de empleo ofertado, se almacena en la base de datos utilizando una DHT²
- Los aspirantes se suscriben a máximo dos sectores diferentes con el fin de recibir ofertas laborales. Se deben simular mínimo 10 solicitudes de diferentes aspirantes (suscripciones)
- En cuanto se reciben las solicitudes de empleo (suscripciones de aspirantes), se filtran de acuerdo con los sectores y se realiza la búsqueda en la base de datos

¹

<https://www.mintrabajo.gov.co/documents/20147/61040102/20201223+Clasificacion+CUOC.pdf/48d6bd59-20fc-4f68-d829-c74ce2e90f80?t=1609861773720>

² *Distributed Hash Table*

- Los aspirantes reciben de forma automática, ofertas laborales de acuerdo con sus capacidades. Cada vez que el aspirante recibe una oferta, debe aceptarla o rechazarla. Si recibe más de una oferta debe indicar cuál de ellas acepta.
- Se puede usar un generador de Carga como JMeter (<https://jmeter.apache.org/>) para generar las vacantes y/o solicitudes de empleo.



Las funciones del componente “Filtro” son

- Recibir las publicaciones de empleos y clasificarlos. Se deben recolectar mínimo 10 ofertas de empleo para distribuirlas entre los servidores
- Recibir las solicitudes de los aspirantes y clasificarlas de acuerdo con las capacidades y habilidades. Se deben recolectar mínimo 10 solicitudes de empleo para distribuirlas entre los servidores
- Si no hay vacantes que correspondan a las solicitudes de los aspirantes, estas deben quedar en espera
- El componente filtro debe tener un backup para proveer disponibilidad, en caso de que la máquina falle

Las funciones de los servidores son las siguientes:

- Recibir las ofertas de empleo y almacenarlas en la base de datos utilizando una DHT para facilitar su búsqueda
- La DHT debe estar distribuida en los servidores y se debe hacer de forma equitativa
- Consultar la base de datos para verificar si existe una vacante, de acuerdo con la solicitud de los aspirantes. También deben almacenar las vacantes publicadas por los empleadores.

- Se deben cumplir unos criterios mínimos para que el empleador contrate al aspirante (p.ej. experiencia, formación académica, habilidades específicas), si cumple con dichas condiciones, automáticamente se da el visto bueno
- Si el aspirante cumple con los requisitos, se debe notificar al aspirante. Tan pronto el aspirante acepte la oferta se debe notificar al empleador y eliminar la vacante de la base de datos
- Si el aspirante es rechazado, se le debe notificar la decisión y este podrá seguir en espera de otra vacante (se debe poner en cola)
- Se debe tener en cuenta que dos o más aspirantes pueden estar solicitando el mismo tipo de empleo al mismo tiempo y estar siendo procesada por servidores diferentes.

DHT

Para construir la DHT, se puede utilizar un código por cada empleador y un código por cada empleo para generar la llave. Inicialmente se deben leer como mínimo 10 ofertas de empleo y almacenarlas en la DHT en espera de que lleguen solicitudes de los aspirantes.

El proyecto se deberá implementar con al menos tres máquinas diferentes: una posible configuración puede ser la siguiente:

- PC1: Empleadores y aspirantes, Servidor 1, DHT
- PC2: Filtro, Servidor 2, DHT
- PC3: Filtro, Servidor 3, DHT

Se pueden tener diferentes implementaciones, teniendo en cuenta que se deben cumplir con los requisitos planteados.

Detalles de Implementación

Comunicación: la comunicación entre los empleadores y el filtro, así como entre aspirantes y el *filtro*, se deben hacer con el mecanismo publicador/suscriptor. La comunicación entre el *filtro* y los servidores se debe hacer utilizando el mecanismo solicitud/respuesta (request/reply).

Fallas: en su implementación debe considerar una posible falla del componente *filtro*, por lo tanto, debe existir un backup que realice las mismas funciones y es el encargado de monitorear que el componente principal está o no en funcionamiento. Cuando se detecte que uno de estos procesos deja de funcionar, el otro debe asumir la carga. Cuando el nuevo proceso entra en operación, la reconexión y distribución de cargas debe hacerse de forma automática, es decir, será transparente para el cliente; las operaciones deben continuar en el punto en que se habían dejado y no iniciar desde el principio.

Necesario para la evaluación: el día de la sustentación es importante que se pueda observar:

- Estado inicial de la DHT después de la primera racha de ofertas de empleo y posibilidad de consultar su contenido.
- Operaciones que van realizando cada uno de los procesos y resultado de la operación, si aplica.

Medidas de Rendimiento

Una vez implementado el proyecto, se deben realizar las siguientes pruebas, para medir el rendimiento del sistema frente a la cantidad de requerimientos y las fallas de uno de sus componentes. Los experimentos a realizar se describen a continuación:

Parámetros (lo que se varía en la experimentación)

- Número de solicitudes enviadas por los aspirantes
- Número de ofertas laborales enviadas por los empleadores
- Falla del componente "Filtro"

Variables Respuesta

Las variables a medir serán:

- I. Tiempo de respuesta de las publicaciones de empleo después de almacenarlas en la DTH
- II. Tiempo que transcurre entre la publicación de una oferta de empleo y la asignación del empleo al aspirante
- III. Tiempo promedio de filtrado de datos por sector para publicación/solicitud de oferta laboral

Descripción del Experimento I

Consiste en realizar **dos corridas del sistema**, por un tiempo **T** definido por el grupo. En este tiempo se deben enviar ofertas de empleo de manera continua y posteriormente emitir las solicitudes de empleo.

En la primera corrida se mide el procesamiento normal de oferta y demanda en la que se asignan empleos de forma satisfactoria. En la segunda corrida del experimento se debe simular conflicto entre varios aspirantes al mismo cargo en los 5 diferentes sectores de trabajo.

Descripción del Experimento II

Realice las 2 corridas descritas en el experimento I, pero al tiempo **T/2** debe fallar uno de los componentes “Filtro”. Debe levantarse un proceso idéntico y el sistema no debe parar su funcionamiento. Durante cada corrida debe medir las variables descritas.

Cómo presentar los resultados

Después de realizar las 4 corridas del sistema descritas anteriormente rellene una tabla con los siguientes resultados:

	Experimento I		Experimento II	
Número de solicitudes de empleo	T. de publicación de oferta laboral promedio	T. Respuesta promedio (Asignación de empleo)	T. Respuesta promedio (Filtro de datos por sector)	T. promedio de búsqueda (consulta en la DHT)
5				
10				

Para los experimentos I y II, **realice gráficos de las variables respuesta en función del número de solicitudes. Analice y describa los resultados.** ¿Cómo afecta la cantidad de requerimientos a los tiempos de respuesta (solicitud de empleos)? ¿Qué pasa cuando falla un componente en el sistema?

PRIMERA ENTREGA

La primera entrega se realizará en la semana 10 a través de la plataforma. Las sustentaciones serán el día correspondiente a la asignatura de acuerdo con el grupo.

La primera entrega consta de un informe donde se debe especificar:

- Diseño del sistema: Diagrama de clases y diagrama de secuencia.
- El protocolo de pruebas que utilizará para la entrega final
- Implementación inicial: deben tener implementada al menos la generación de ofertas de trabajo.

Se deben implementar los requerimientos de acuerdo con el enunciado y la rúbrica.

El día de la sustentación, cada equipo tendrá 10 minutos para mostrar sus resultados y responder las preguntas que tengan los profesores.

SEGUNDA ENTREGA

La entrega se realizará en la semana 17 a través de la plataforma. Las sustentaciones se harán los días que correspondan para cada curso. El día de la sustentación los integrantes del equipo deben **mostrar la funcionalidad del proyecto**. Deben estar presentes todos los integrantes del grupo.

La entrega se compone de:

- En un archivo .zip código fuente de los programas que conforman el sistema y un archivo Readme donde indique cómo ejecutarlo.
- Se debe complementar la documentación de la primera entrega
- Un video de máximo 10 minutos donde muestra la topología implementada y explicar los siguientes aspectos de su proyecto:
 - a. Distribución de componentes en máquinas.
 - b. Librerías y patrones usados
 - c. Tratamiento de la falla del gestor.
 - d. Algoritmo de búsqueda en la DHT
- Un informe de máximo 4 páginas donde explican los experimentos realizados y resultados obtenidos. Aparte de las especificaciones que se encuentran en el informe, se debe agregar: especificaciones de hw y sw donde se realizaron las medidas y herramientas de medición utilizadas. Debe mostrar en tablas y gráficos los resultados obtenidos, así como un análisis de dichos resultados.

Equipos de Trabajo.

El proyecto se realizará en grupos de trabajo de máximo tres personas

No puede existir replicación de documentos ni de código fuente entre grupos, lo cual se consideraría plagio.