

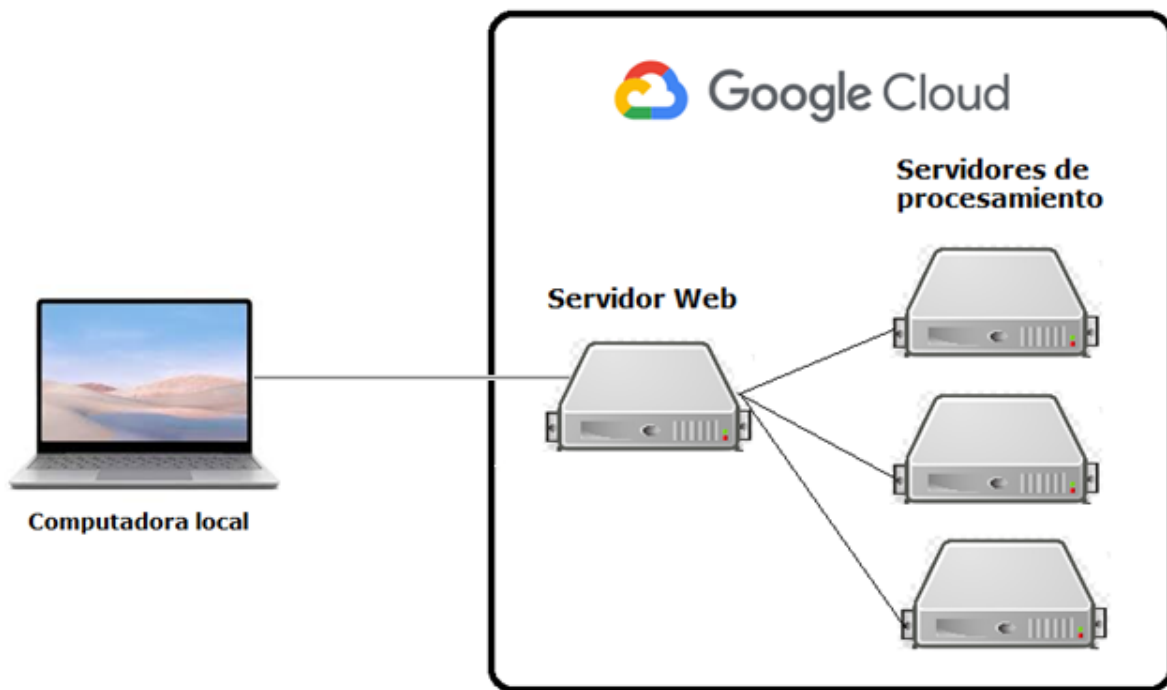
Desarrollo de Sistemas Distribuidos

PROYECTO FINAL

Para el presente proyecto se programará un sistema distribuido en la nube que realice procesamiento para minería de texto. La explicación del algoritmo que se debe programar se presenta como “TERCER PROPUESTA DE SOLUCIÓN” en el siguiente video:

<https://youtu.be/Xab-ZgofNbQ>

Posteriormente deberá desarrollar de manera colaborativa con el equipo de trabajo que le tocó, un sistema distribuido que realiza la búsqueda de frases en un conjunto de textos utilizando tf-idf. El sistema debe contar con una interfaz web como la vista en la clase 38, y con tres servidores de procesamiento adicionales al servidor web en la nube como se muestra en la imagen.



En la interfaz web el usuario debe poder introducir una frase, y al dar click para enviar, el sistema distribuido deberá hacer la minería de texto para posteriormente enviar **en orden de relevancia** una lista con todos los textos contenidos en la base de datos.

Para distribuir el trabajo existen dos soluciones principalmente, la primera es distribuir los textos entre los tres servidores, y la segunda es que todos los servidores tengan todos los textos, y asignar a cada servidor los textos en los que hará la búsqueda. **Se utilizará la segunda opción debido a su escalabilidad.**

Los textos donde se realizarán las búsquedas se podrán descargar desde la siguiente liga:

https://drive.google.com/file/d/1BEdeEo-ypzTQcbm7_t1am7QKVKdyWYtXD/view?usp=sharing

El sistema distribuido deberá ser tolerante a fallos, de tal manera que, si uno o dos de los servidores de procesamiento se les mata exactamente después de haberse enviado la consulta, el usuario seguirá recibiendo la respuesta correcta (para esta parte no es necesario usar ZooKeeper pero es posible su uso si se desea). Para mas detalles ver la Rúbrica de evaluación:

Rúbrica

Implementación	Puntos
Minería de texto con: "PRIMER PROPUESTA DE SOLUCIÓN"	1.5
Minería de texto con: "SEGUNDA PROPUESTA DE SOLUCIÓN"	3
Minería de texto con: "TERCER PROPUESTA DE SOLUCIÓN"	5
Sistema similar al de la clase 38 (con front end apropiado) funcionando con alguna propuesta de solución de minería de texto.	2
Sistema de tolerancia a fallos básico: Si se mata uno o dos servidores de procesamiento en el momento que no se está realizando ninguna consulta el sistema devuelve la respuesta correcta.	1.5
Sistema de tolerancia a fallos deseable: Si se mata uno o dos servidores de procesamiento mientras se está haciendo la consulta, el sistema devuelve la respuesta correcta.	3

El proyecto será subido **únicamente por uno de los integrantes del equipo** (líder del equipo) a la plataforma Teams. Se debe de subir únicamente el código (sin archivos class) a la plataforma, con cada clase en un archivo separado, y archivo Makefile o pom.xml para compilar. Cada archivo debe contener al inicio como comentario el nombre de los integrantes y el grupo al que pertenece cada integrante.

Se recomienda ir subiendo cada vez la última versión que se tenga por si se cierra el sistema o se va la luz o el internet. Si no se subió ningún archivo en automático la calificación del proyecto final será de cero para ese equipo.

Es responsabilidad del líder del equipo notificar si uno de los miembros no trabajó o merece un porcentaje menor de la calificación en el proyecto.

Importante: Al igual que en otros proyectos, en ninguna parte del proyecto debe existir código descargado de internet o que sea copiado del código de otros equipos. En cualquier caso, si existe plagio al proyecto se le asignará la calificación de cero aún después de haber sido evaluado.

Revisión

El proyecto será revisado y calificado el día lunes 26 de junio comenzando a las 7:00 AM, en el salón 2110 según una lista donde se irán anotando los equipos, y hasta el momento en que no haya más equipos por revisar.

Calificaciones finales

Las calificaciones finales en archivos Excel se publicarán en la plataforma teams el día martes 27 de junio a las 10:00 AM. Si existiera alguna aclaración del tercer parcial que implique cambios que puedan modificar la calificación final, favor de hacérmelo saber el mismo 27 de junio en un horario entre las 10:00 AM y 11:59 AM pasando conmigo a mi cubículo ubicado en la sala de profesores 2108.

Examen extraordinario

Como se especificó al inicio del curso, el promedio de todos los proyectos en el semestre corresponde al 40% de la calificación final del extraordinario. El otro 60% corresponderá al desarrollo de una práctica con lo visto durante el semestre, que se tendrá que resolver en el aula por lo que se requiere traer su LAP para programarla. El examen extraordinario se realizará el viernes 30 de junio en un horario de 8:00AM a 10:00AM en el salón 2110.

ETS

Con objeto de cubrir más temas prácticos tomé la decisión de sacrificar las clases de teoría del tercer parcial. Si se va a presentar ETS será necesario también estudiar los siguientes temas correspondientes al tercer parcial:

LIBRO	TEMAS
<i>"Distrubuted Systems" de George Coulouris</i>	<i>Temas: 18, 18.1, 18.2 y 18.2.1 contenidos entre las páginas 765 y 771. Temas: 9, 9.1, 9.2, 9.2.1 contenidos entre las páginas 381 y 393. Temas: 9.6, 9.7, 9.7.1, 9.7.2, 9.7.3 contenidos entre las páginas 411- 419.</i>
<i>"SISTEMAS DISTIBUIDOS Principios y Paradigmas" de Andrew S. Tanenbaum</i>	<i>Temas: 7.4, 7.4.1, 7.4.2 contenidos entre las páginas 296 - 302.</i>