



PROYECTO 4



Elaborado por: M. en C. Ukranio Coronilla

El presente proyecto consiste en una aplicación distribuida que permite realizar búsquedas sobre un conjunto de textos en inglés con ayuda de dos servidores de procesamiento que utilizan los tipos de CPU **más económicos** que soportan las instancias de GCP. Toda la aplicación debe estar programada con Java y utilizar HTTP como medio de comunicación.

Los textos en inglés a procesar pueden descargarse de:

<https://www.gutenberg.org/>

Todos los textos en inglés deben encontrarse almacenados en **Google Cloud Storage (GCS)** y los dos servidores de procesamiento tendrán que descargarlos de GCS mediante una API para poder procesarlos. Para acceder al Cloud Storage se puede utilizar cualquier API como la Google Cloud Storage API, Google Cloud Storage JSON API, Google Drive API o Storage Transfer API.

Dado que las instancias pueden fallar en cualquier momento (tratar a las instancias como ganado y no como mascotas), al crearse una nueva instancia tendría que descargar los textos de la GCS.

Aunque en este proyecto no se va a considerar que una instancia pueda fallar, si se considerará en un futuro.

Las dos instancias conforman el backend y su trabajo consiste en encontrar **todos** los palíndromos (véase <https://es.wikipedia.org/wiki/Pal%C3%A1ndromo>) de 1, 2, 3, 4 y hasta 5 palabras que existan en todos los textos que se encuentren almacenados en el **GCS** en ese momento. Cómo se muestra en la explicación de Wikipedia, no debe importar si son mayúsculas o minúsculas, los espacios, ni los puntos, comas ni saltos de línea.

El frontend consiste en un programa cliente que utilice la librería de Java **Lanterna**, el cual se deberá ejecutar en una terminal UNIX dentro de su computadora local y mostrará los resultados de las búsquedas en tiempo real imprimiendo cada palíndromo y el identificador del texto donde se encontró. También debe imprimir los títulos de los textos almacenados en GCS con su identificador respectivo y por último imprime un menú con dos opciones, para iniciar la búsqueda y para salir de la aplicación. Ejemplo:

```
ukranio@HUAWEIR5: ~/SD_2( x + v

PALINDROMOS ENCONTRADOS Y TEXTO DONDE SE ENCUENTRAN:
UNA PALABRA: level-3, madam-3, racecar-6, civic-4, deified-2, refer-1
DOS PALABRAS: too hot to hoot-6, live evil-3, step on no pets-5
TRES PALABRAS: dog sees god-6, top spot pot-1, evil olive-2
CUATRO PALABRAS: was it a car or a cat I saw-2, never odd or even-5, no devil lived on-1
CINCO PALABRAS:

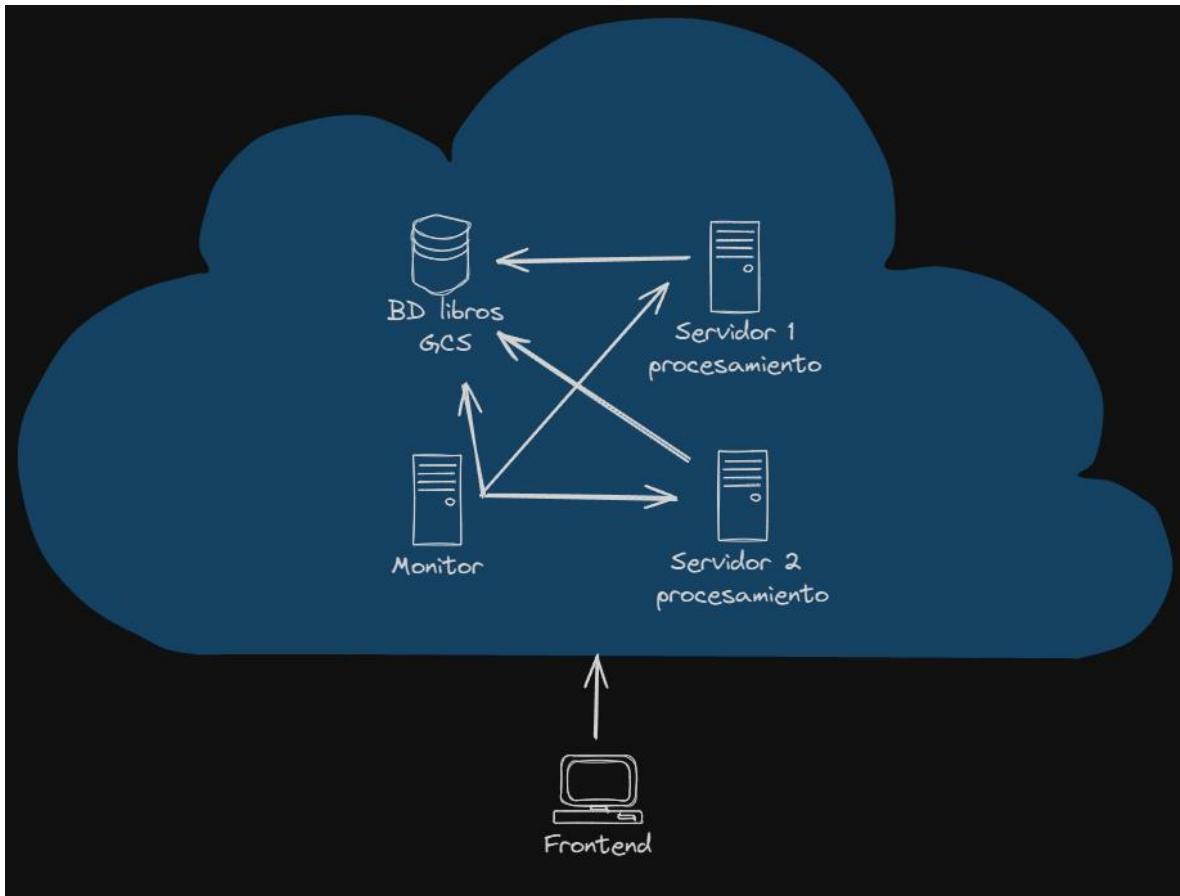
TEXTOS EN GCS:
1 Frankenstein; Or, The Modern Prometheus by Mary Wollstonecraft Shelley
2 Moby Dick; Or, The Whale by Herman Melville
3 Romeo and Juliet by William Shakespeare
4 Pride and Prejudice by Jane Austen
5 Alice's Adventures in Wonderland by Lewis Carroll
6 The Strange Case of Dr. Jekyll and Mr. Hyde by Robert Louis Stevenson

1.- Iniciar la búsqueda
2.- Salir del programa
Ingresa una opción y presiona Enter: |
```

Adicionalmente deberá existir una terminal SSH de una tercera instancia encargada de monitorear el uso de CPU de las dos instancias de procesamiento en tiempo real y con su respectiva IP, así como de imprimir los títulos de los textos que se encuentran almacenados en GCS en todo momento, para lo cual deberá apoyarse en la librería Lanterna. Ejemplo:

Por último, si se agregan nuevos textos a GCS y se inicia de nuevo la búsqueda, automáticamente se realizarán búsquedas también en los textos que se acaban de agregar además de los que ya existían.

La arquitectura general se muestra a continuación:



Para crear el video demostrativo de su proyecto deberá mostrar y en este orden:

1. La ubicación y nombre de al menos diez textos en el Google Cloud Storage.
2. La ejecución de los servidores de procesamiento 1 y 2 los cuales imprimen el título del libro sobre el que están trabajando, cuando les hagan una solicitud y el número de palabras del palíndromo que buscan.
3. La ejecución del monitor que muestra los archivos almacenados y el uso de CPU de los servidores de procesamiento.
4. La ejecución del cliente en una terminal de su computadora realizando la búsqueda de palíndromos, dejando ver en todo momento el monitor que muestra sólo el uso del CPU de ambos servidores de procesamiento, los cuales tendrán que reflejar que ambos servidores están trabajando y no sólo un CPU (hay distribución de carga). Al terminar la búsqueda de palíndromos hacer una captura de pantalla de los resultados.
5. Agregar al menos tres archivos nuevos al Google Cloud Storage y realizar la búsqueda de palíndromos de nuevo, pero ahora mostrando las impresiones de los servidores de procesamiento. Al terminar la búsqueda comparar los resultados obtenidos con los que están en la captura haciendo notar que no son los mismos resultados.

Importante: Este proyecto es individual. Suba el código fuente de cada clase en un archivo separado, y cada archivo de código que suba debe contener al inicio como comentario el número

de proyecto, su nombre completo y el grupo al que pertenece, de no hacerlo así se le descontará un punto de la calificación. No suba archivos class, archivos comprimidos ni ligas a sitios web externos pues no le será tomado en cuenta el proyecto. Asimismo, deberá subir un video breve mostrando como se ejecuta su proyecto y los cinco puntos que se indican en orden. Se recomienda utilizar OBS Studio con baja resolución para no subir archivos muy pesados y no es necesario que hable en el video.