

# Laboratorio 1 Introducción al paralelismo - hilos

Integrantes:

Luisa Fernanda Bermúdez Girón Karol Daniela Ladino Ladino

Squad:

Inside Out

Profesor:

Javier Iván Toquica Barrera

Curso:

ARSW-1

Fecha De Entrega:

25-08-2023

### I. Introducción

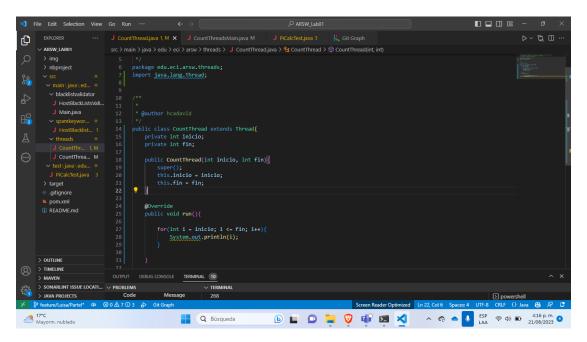
El paralelismo y los hilos son herramientas esenciales para optimizar la eficiencia y el rendimiento de los sistemas informáticos modernos. Al comprender cómo funcionan y cómo aplicarlos de manera efectiva, los desarrolladores pueden crear aplicaciones más rápidas y capaces de manejar tareas más complejas.

En este laboratorio se busca tener una comprensión práctica y profunda del paralelismo y el uso de hilos en Java, así como su impacto en el rendimiento de las aplicaciones. También nos permitirá explorar cómo diseñar implementaciones eficientes y optimizadas.

### II. Desarrollo del laboratorio

#### Parte I Hilos Java

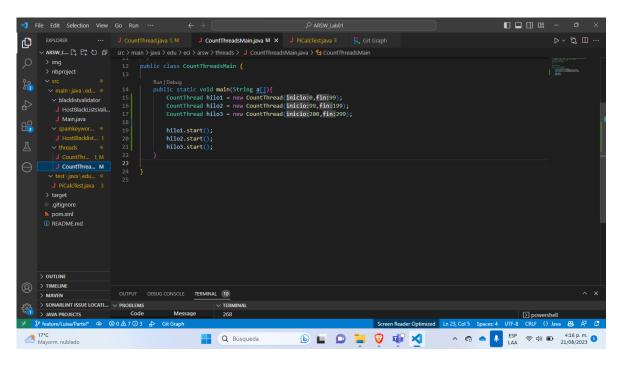
 De acuerdo con lo revisado en las lecturas, complete las clases CountThread, para que las mismas definan el ciclo de vida de un hilo que imprima por pantalla los números entre A y B.



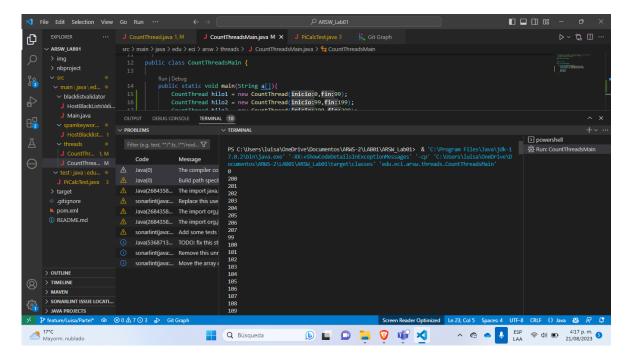
- 2. Complete el método **main** de la clase CountMainThreads para que:
  - I. Cree 3 hilos de tipo CountThread, asignándole al primero el intervalo [0..99], al segundo [99..199], y al tercero [200..299].

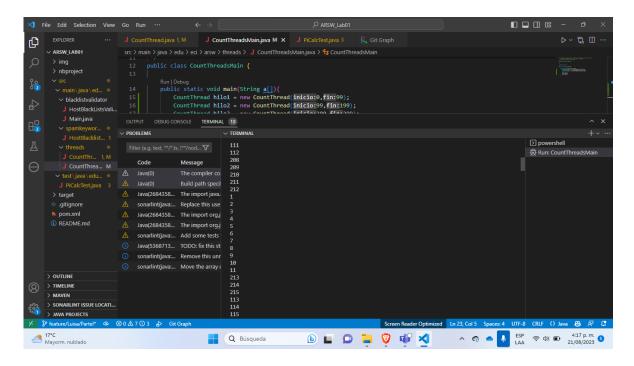
```
Run | Debug
public static void main(String a[]){
    CountThread hilo1 = new CountThread(inicio:0,fin:99);
    CountThread hilo2 = new CountThread(inicio:99,fin:199);
    CountThread hilo3 = new CountThread(inicio:200,fin:299);
```

II. Inicie los tres hilos con 'start()'.



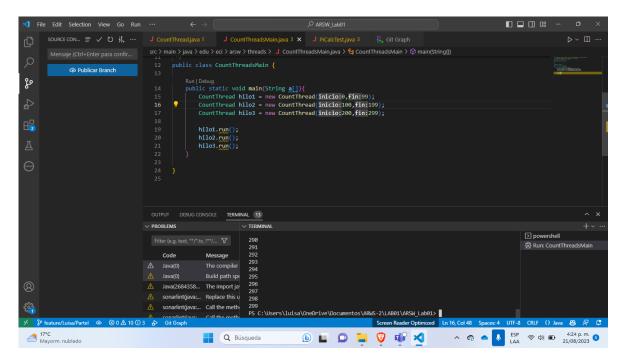
III. Ejecute y revise la salida por pantalla.



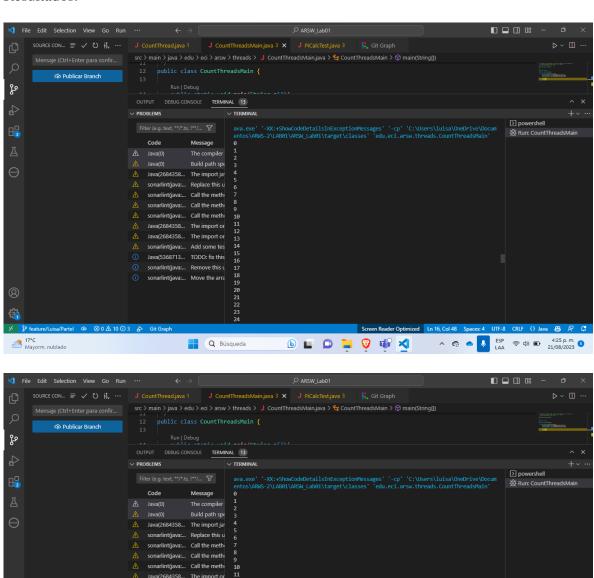


IV. Cambie el incio con 'start()' por 'run()'. ¿Cómo cambia la salida?, ¿por qué?.

Para poder tomar todos los números de 0 al 299, modificamos el rango del segundo hilo de [99..199] a [100...199]



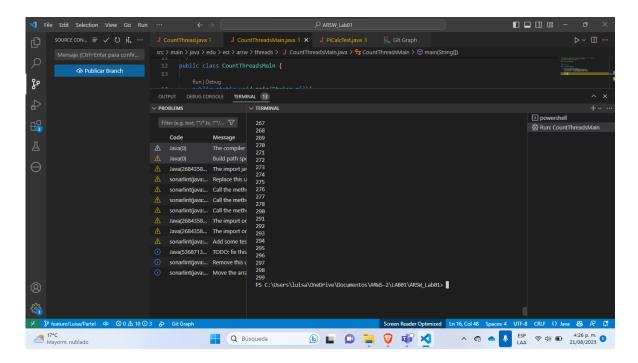
#### Resultados:



△ Java(2684358... The import or
△ sonarlint(java:... Add some tes
⊙ Java(5368713... TODO: fix this
⊙ sonarlint(java:... Remove this u
⊙ sonarlint(java:... Move the arra

Q Búsqueda

(b) 🔲 🔘 📜 🦁 💞 🔀

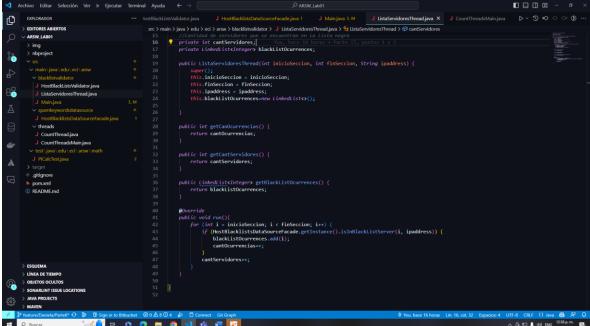


Como podemos observar cuando iniciamos los hilos con start() arranca los 3 hilos al tiempo, es por eso que en la ejecución se van mostrando intercalados los resultados, esto se debe a que los hilos son asincrónicos, por lo tanto, no tienen que esperar que uno termine pare que el otro empiece, mientras que cuando iniciamos los hilos con run() este va a ejecutar los hilos uno por uno, en este caso, hasta que no termine la ejecución del hilo 1, no va a iniciar la del hilo 2 y así sucesivamente.

### Parte II Hilos Java

Para 'refactorizar' este código, y hacer que explote la capacidad multi-núcleo de la CPU del equipo, realice lo siguiente:

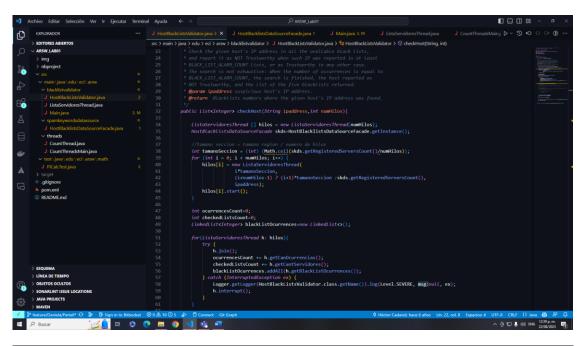
1. Cree una clase de tipo Thread que represente el ciclo de vida de un hilo que haga la búsqueda de un segmento del conjunto de servidores disponibles. Agregue a dicha clase un método que permita 'preguntarle' a las instancias del mismo (los hilos) cuantas ocurrencias de servidores maliciosos ha encontrado o encontró.

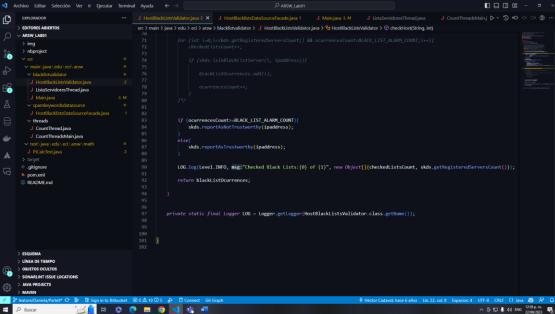


2. Agregue al método 'checkHost' un parámetro entero N, correspondiente al número de hilos entre los que se va a realizar la búsqueda (recuerde tener en cuenta si N es par o impar!). Modifique el código de este método para que divida el espacio de búsqueda entre las N partes indicadas, y paralelice la búsqueda a través de N hilos. Haga que dicha función espere hasta que los N hilos terminen de resolver su respectivo subproblema, agregue las ocurrencias encontradas por cada hilo a la lista que retorna el método, y entonces calcule (sumando el total de ocurrencuas encontradas por cada hilo) si el número de ocurrencias es mayor o igual a BLACK\_LIST\_ALARM\_COUNT. Si se da este caso, al final se DEBE reportar el host como confiable o no confiable, y mostrar

el listado con los números de las listas negras respectivas. Para lograr este comportamiento de 'espera' revise el método join del API de concurrencia de Java. Tenga también en cuenta:

- Dentro del método checkHost Se debe mantener el LOG que informa, antes de retornar el resultado, el número de listas negras revisadas VS. el número de listas negras total (línea 60). Se debe garantizar que dicha información sea verídica bajo el nuevo esquema de procesamiento en paralelo planteado.
- Se sabe que el HOST 202.24.34.55 está reportado en listas negras de una forma más dispersa, y que el host 212.24.24.55 NO está en ninguna lista negra.





### Parte II.I Para discutir el Martes (NO para implementar aún)

La estrategia de paralelismo antes implementada es ineficiente en ciertos casos, pues la búsqueda se sigue realizando aún cuando los N hilos (en su conjunto) ya hayan encontrado el número mínimo de ocurrencias requeridas para reportar al servidor como malicioso. ¿Cómo se podría modificar la implementación para minimizar el número de consultas en estos casos?, qué elemento nuevo traería esto al problema?

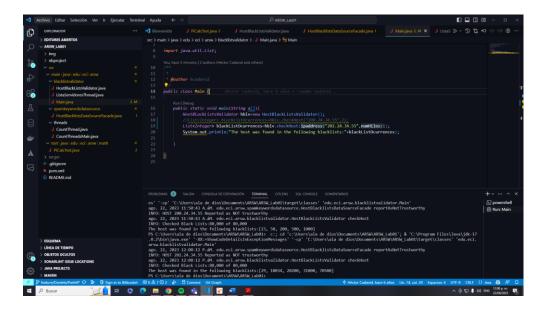
Para minimizar el número de consultas podemos implementar una medida de sincronización de los hilos en la cual se busca que los hilos tengan un estado en común para saber cuántas veces se encontró dicha ocurrencia, para esto se usara una clase que va a funcionar como un "Semáforo" para saber si el hilo puede continuar con la búsqueda cuando su estado sea verde o debe parar cuando su estado sea rojo.

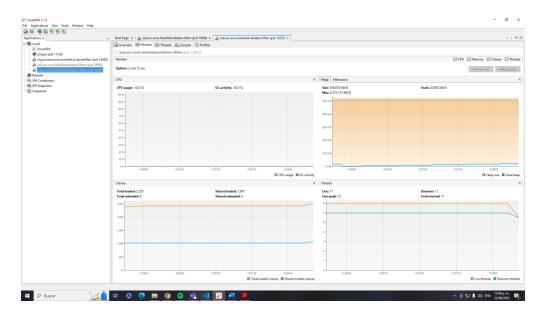
En este caso se modificaría el código del método run(), para que antes de hacer cualquier cosa los hilos pregunten si el semáforo está en verde y realicen el proceso, para esto se utilizara un contador que se inicializara en cero y cada vez que el hilo encuentre una instancia en el servidor se vaya incrementando la cantidad de ocurrencias en este, de esta forma cuando el contador llegue a la cantidad requerida va a bloquearse, es decir ahora el estado del semáforo va a cambiar a rojo llevando así a que cuando los hilos vuelvan a preguntar por el estado terminen su ejecución rompiendo cualquier clase de iteración.

## Parte III Evaluación de Desempeño

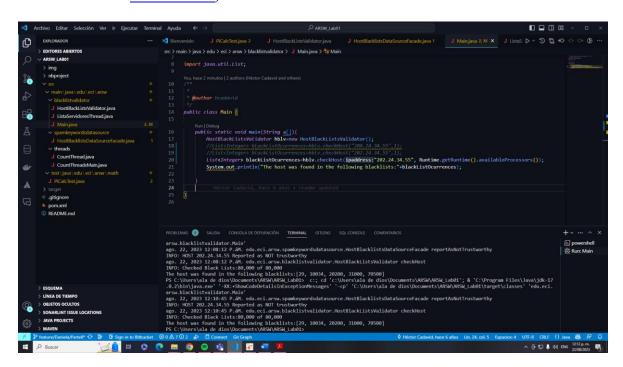
A partir de lo anterior, implemente la siguiente secuencia de experimentos para realizar las validaciones de direcciones IP dispersas (por ejemplo 202.24.34.55), tomando los tiempos de ejecución de los mismos (asegúrese de hacerlos en la misma máquina):

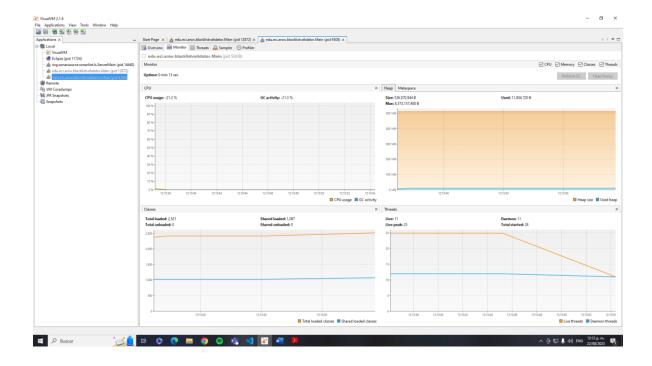
1. Un solo hilo.



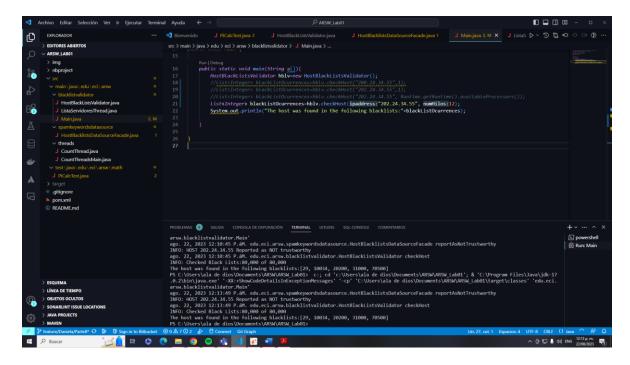


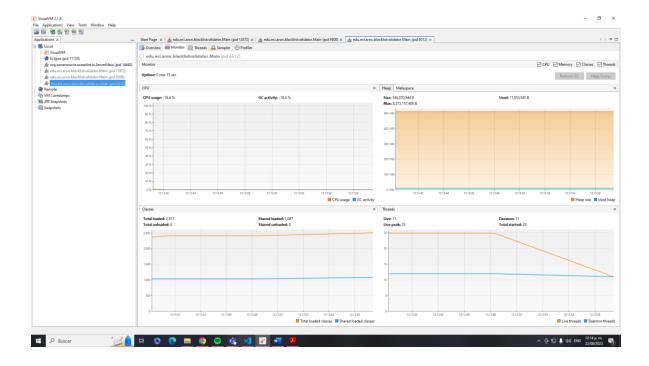
2. Tantos hilos como núcleos de procesamiento (haga que el programa determine esto haciendo uso del <u>API Runtime</u>).



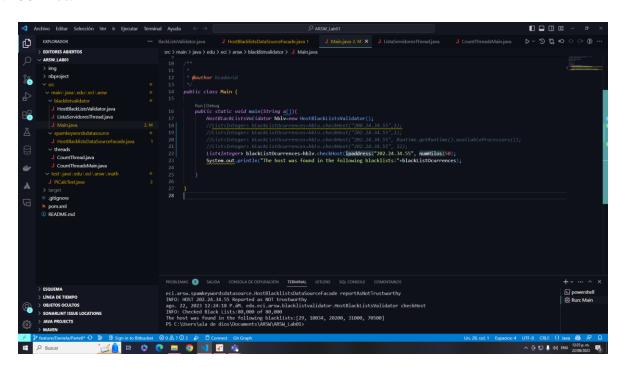


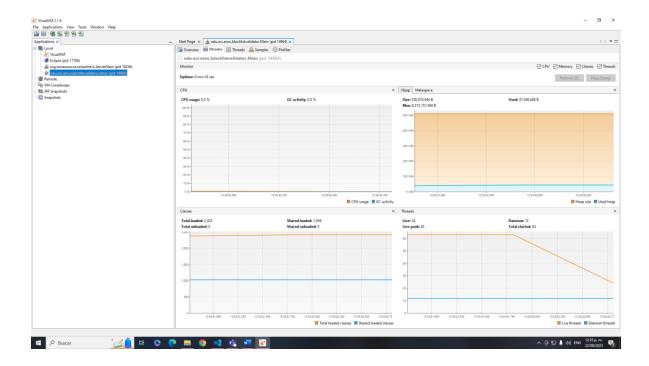
3. Tantos hilos como el doble de núcleos de procesamiento.



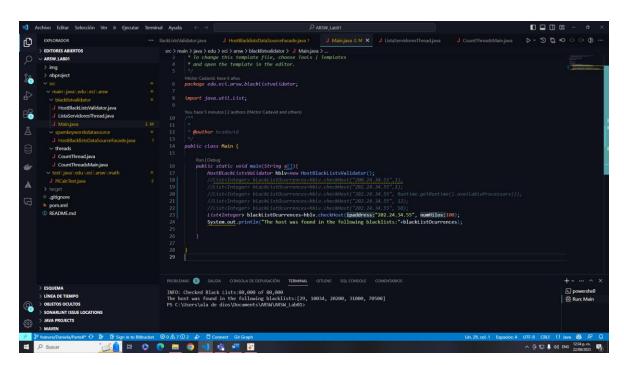


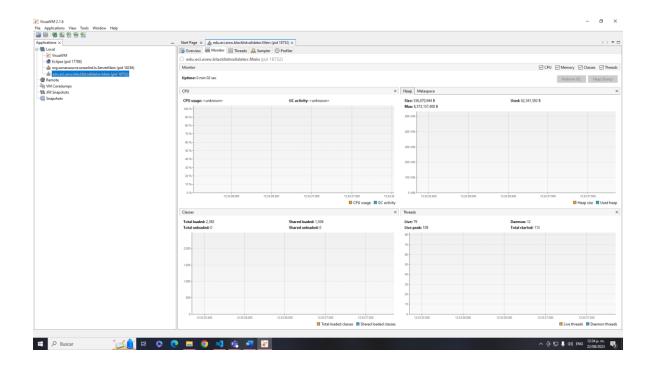
### 4. 50 hilos.





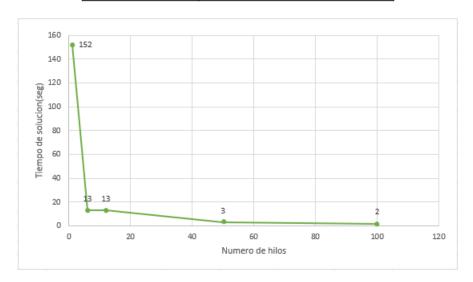
### 5. 100 hilos.

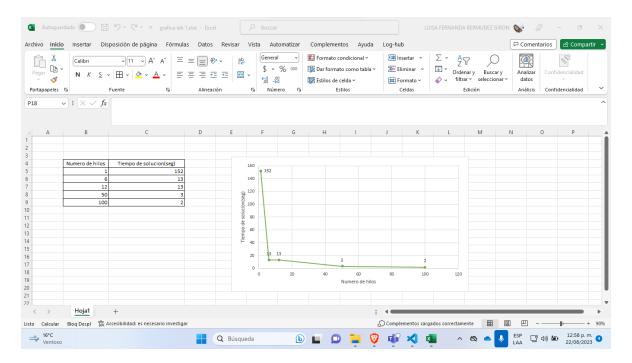




Con lo anterior, y con los tiempos de ejecución dados, haga una gráfica de tiempo de solución vs. número de hilos.

Numero de hilos	Tiempo de solucion(seg)
1	152
6	13
12	13
50	3
100	2





Analice y plantee hipótesis con su compañero para las siguientes preguntas (puede tener en cuenta lo reportado por jVisualVM):

#### 1. Según la <u>ley de Amdahls</u>:

 $S(n) = \frac{1}{(1-P) + \frac{P}{n}}$ , donde S(n) es el mejoramiento teórico del desempeño, P la fracción paralelizable del algoritmo, y n el número de hilos, a mayor n, mayor debería ser dicha mejora. Por qué el mejor desempeño no se logra con los 500 hilos?, cómo se compara este desempeño cuando se usan 200?.

Al comparar la ejecución con 200 y 500 hilos, se puede inferir que el mejor rendimiento no se logra con 500 hilos. Esto se debe a una consideración clave en la Ley de Amdahl, donde la capacidad de paralelización tiene un límite máximo teniendo en cuenta el proceso o sistema. Después de cierto punto, aumentar el número de hilos no resulta en un aumento proporcional del rendimiento, ya que la porción secuencial del algoritmo se vuelve más significativa.

En otras palabras, la ley de Amdahl sugiere que a medida que se incrementa el número de hilos, el beneficio en términos de mejora de desempeño se atenúa y eventualmente se estabiliza. Por lo tanto, es más efectivo utilizar un número menor de núcleos que operen a una frecuencia más alta en lugar de aumentar significativamente la cantidad de núcleos a costa de una frecuencia de operación más baja.

Esta observación se alinea con la necesidad de encontrar un equilibrio entre la cantidad de hilos y la capacidad del hardware, considerando que más hilos no siempre se traducen en un mejor rendimiento debido a las limitaciones físicas y a la sobrecarga asociada con la administración de múltiples hilos.

- 2. Cómo se comporta la solución usando tantos hilos de procesamiento como núcleos comparado con el resultado de usar el doble de éste?.
  - En este caso al comparar la solución usando tantos hilos de procesamiento como núcleos con la solución de usar el doble de este, podemos observar que se comportan de la misma forma.
- 3. De acuerdo con lo anterior, si para este problema en lugar de 100 hilos en una sola CPU se pudiera usar 1 hilo en cada una de 100 máquinas hipotéticas, la ley de Amdahls se aplicaría mejor?. Si en lugar de esto se usaran c hilos en 100/c máquinas distribuidas (siendo c es el número de núcleos de dichas máquinas), se mejoraría?. Explique su respuesta.

Teniendo en cuenta la situación hipotética y lo aprendido anteriormente, podemos inferir que la ley de Amdahls se aplicaría mejor en situaciones en las que se busque y se tenga un aprovechamiento de la porción paralelizable en cada una de las c maquinas hipotéticas que fueran cuello de botella, ya que esta es una de las limitaciones de la ley de Amdahls, porque si el proceso no está siendo limitado por la CPU, después de una cantidad determinada de núcleos se perderá la ganancia de rendimiento.

### III. Conclusiones

- La ejecución de múltiples hilos en Java proporciona paralelismo, lo que permite que los hilos se ejecuten simultáneamente y sin una secuencia estricta. Para esto pudimos observar que al iniciar los hilos con start(), se ejecutan de manera asincrónica, lo que resulta en una salida intercalada de resultados, mientras que al cambiar el inicio de los hilos a run(), los hilos se ejecutan de manera secuencial, uno después del otro, lo que afecta la eficiencia del paralelismo.
- Se debe tener en cuenta tanto la naturaleza específica del problema que se está abordando como las capacidades fundamentales del hardware en el que se está aplicando la estrategia de paralelización.
- La Ley de Amdahl se presenta como una herramienta útil para comprender las restricciones de rendimiento que surgen al incrementar la cantidad de hilos utilizados.
- Se observa cómo los beneficios obtenidos a través del paralelismo pueden llegar a un punto en el que su impacto positivo disminuye considerablemente.