

# ARTÍCULO DE INVESTIGACIÓN

Tecnologías de programación paralela y concurrente



Gerardo Daniel Naranjo Gallegos, A01209499.

A01209499@itesm.mx

Tecnológico de Monterrey, Campus Querétaro.

Materia: Multiprocesadores.

Profesor: Pedro Oscar Pérez Murueta.

pperezm@tec.mx

1 DE DICIEMBRE DE 2019

# Contenido

Resumen	2
Introducción	3
Desarrollo	4
Librería cheader.h y cppheader.h	4
Secuencial en C	4
Secuencial en C++	5
Secuencial en Java	6
Paralelo en OpenMP	6
Paralelo TBB	7
Paralelo en CUDA	8
Paralelo con <i>Threads</i> en Java	8
Paralelo en Java con Fork/Join	9
Análisis de resultados	10
Comparativa entre tecnologías secuenciales	10
Comparativa entre tecnologías paralelas	11
Comparativa entre secuenciales y paralelas de Java	12
Comparativa entre secuencial y paralela de C y C++	13
Comparativa de todas las tecnologías implementadas:	14
Conclusión	15
Referencias	16
Anexos	18
Librería cheader.h	18
Librería cppheader.h	20

Implementación en C	22
Implementación en C++	25
Implementación en Java	28
Implementación en OpenMP	31
Implementación en TBB	34
Implementación en CUDA	37
Implementación Threads	41
Implementación Fork/Join	44

#### Resumen

En este artículo de investigación se desarrollarán algoritmos que resuelvan un problema de programación, en distintas herramientas que los puedan hacer paralelizados y de manera secuencial, con el objetivo de comparar las técnicas de programación paralela y concurrente.

El problema por resolver será operaciones con vectores. Se utilizará el servidor otorgado por el profesor, con el objetivo de realizar una comparación equitativa.

#### Introducción

Este es un artículo de investigación, llevado a cabo para la materia de Multiprocesadores con el profesor Pedro Oscar Pérez Murueta durante el semestre agosto-diciembre del 2019, que presenta un problema de programación resuelto con las técnicas vistas en clase: threads en Java, Fork/Join en Java, OpenMP en C, Intel Threading Building Blocks (TBB) en C++ y CUDA.

Este artículo pretende desarrollar las habilidades de: dominio del idioma inglés, pensamiento crítico, visión del entorno internacional, comunicación escrita, y uso eficiente de la informática y las telecomunicaciones.

La importancia de una buena programación, de la paralelización de los códigos y de la optimización de los recursos de la computadora ha sido tratado a lo largo del curso y se puede obtener información interesante de artículos como *The Free Lunch Is Over: A Fundamental Turn Toward Concurrency in Software* y *Welcome to the Jungle: Or, A Heterogeneous Supercomputer in Every Pocket*, ambos escritos por Sutter's Mill. (Herb Sutter, 2011) (Herb Sutter, 2005)

El problema seleccionado para este artículo de investigación es la suma y resta de dos vectores, utilizando todas las técnicas de paralelización vistas, debido a que el objetivo de este artículo es la comparación de las distintas herramientas vistas en clase y no el desarrollo complejo de un algoritmo de programación.

La suma y resta de dos vectores consiste en la suma o resta de cada uno de sus componentes, dando como resultado un tercer vector. Por ejemplo, si tenemos un arreglo A=(3,5,7) y un arreglo B=(2,4,6) el resultado de la sumatoria sería un tercer vector:  $C=\left((3+2),(5+4),(7+6)\right)=(5,9,13)$  y el resultado de la resta para los mismos vectores sería el vector: D=((3-2)+(5-4)+(7-6))=(1,1,1).

#### Desarrollo

Para desarrollar las distintas técnicas utilizadas es recomendable iniciar resolviendo el problema de forma secuencial, es decir, en C, C++ y Java. Posteriormente, desarrollar la parte paralela (*threads*, *fork/join*, OpenMP, TBB y CUDA).

Para el desarrollo, de manera opcional, es posible utilizar software de entorno de desarrollo integrado (IDE) o algún editor de código fuente, con el objetivo de que resulte más fácil. En este caso, se optó por utilizar Atom. (Atom, 2019)

Todos los códigos elaborados se encontrarán en la sección de <u>Anexos</u>, al final del documento y anexos a la carpeta del mismo. En todas las tecnologías utilizadas, se implementaron vectores de 100 millones de datos cada uno; siendo el primero llenado de manera aleatoria y el segundo con números naturales ascendentes (1, 2, 3, 4...).

Finalmente, cabe mencionar que todas las pruebas fueron realizadas en el servidor del profesor, con el objetivo de obtener datos más certeros y tener constancia del poder de cómputo.

#### Librería cheader.h y cppheader.h

Todos los códigos desarrollados utilizan la librería "cheader.h" o "cppheader.h", la cual ha sido otorgada por el profesor y sirve para obtener el tiempo de ejecución de un programa, rellenar arreglos con números consecutivos, rellenar arreglos con números aleatorios y mostrar arreglos en la terminal.

Ambas librerías pueden ser encontradas en la sección de Anexos, al final del documento, junto con los códigos desarrollados.

#### Secuencial en C

El desarrollo de forma secuencial en C representa la base para su posterior implementación paralelizada en tecnologías como OpenMP o CUDA. Presenta la ventaja de ser un lenguaje orientado a Unix en sus inicios y de ser de fácil implementación. (Brian Kernighan, 1991)

A continuación, se adjunta la Imagen 1, en donde se puede observar una terminal en el servidor compilando el código y ejecutándolo, comprobando que se crean ambos vectores correctamente y que se efectúan la suma y resta de manera correcta.

Finalmente, se puede observar que el tiempo de ejecución del algoritmo es de 371.78950 milisegundos.



Imagen 1. Algoritmo en C.

#### Secuencial en C++

Este lenguaje está basado en C, optimizado para la orientación a objetos. Servirá de base para la implementación de las tecnologías en paralelo. (Balagurusamy, E., 2007)

A continuación, en la Imagen 2, se puede observar la compilación del código y su ejecución. Su tiempo de procesamiento es de 450.623 milisegundos, que es bastante lento por comparación con C, pero este algoritmo presenta una característica de compilación, una optimización de nivel 3, que reduce el tiempo de ejecución a 102.704 milisegundos, haciéndolo más de 3 veces más rápido que el algoritmo en C, que no presenta optimización.

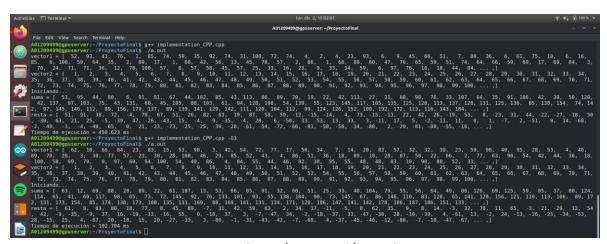


Imagen 2. Implementación en C++.

#### Secuencial en Java

Java se programa de forma diferente a C o C++; es un lenguaje orientado a objetos. Servirá de base para la implementación paralela mediante *Fork/Join* y *Threads*. (J. Steven Perry, 2012)

En la Imagen 3, se puede observar cómo se compila el código y cómo se ejecuta. El tiempo de procesamiento es de 100.29 milisegundos. Los algoritmos en Java no cuentan con optimización (como C++), no obstante, resulta ser el más rápido de ejecución en forma secuencial.



Imagen 3. Implementación en Java.

#### Paralelo en OpenMP

El primer algoritmo de ejecución en paralelo es el favorito de muchos, OpenMP. Este algoritmo resulta ser muy fácil de paralelizar, una vez elaborado el código secuencial. El lenguaje OpenMP resulta ser una API enfocada a la programación multiproceso, con memoria compartida. (Leonardo Dagum, 1998) (OpenMP, 2019)

El algoritmo desarrollado es en gran parte el código secuencial. En la Imagen 4 se puede observar cómo se compila y ejecuta, obteniendo un tiempo de 98.2668 milisegundos, representando un avance respecto a los secuenciales; pero la verdadera mejora es al realizarle una optimización de nivel 3, obteniendo 74.8306 milisegundos.

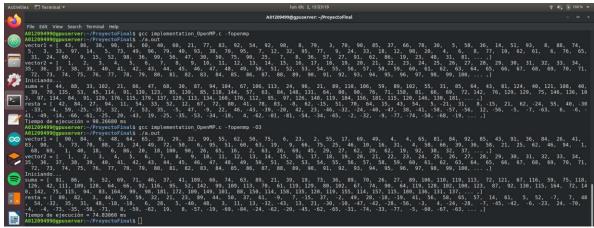


Imagen 4. Implementación en OpenMP.

#### Paralelo TBB

Intel *Threading Building Blocks* (TBB) es una biblioteca actualizada constantemente que se puede encontrar en GitHub, desarrollada por Intel específicamente para potenciar el paralelismo de códigos escritos en C++; una alternativa al uso de *Threads*. (Intel Corporation, 2019)

La Imagen 5 contiene su compilación y ejecución, dando un tiempo de ejecución de 119.506 milisegundos, comparablemente mejor que el tiempo en C++. Además, como C++, también cuenta con una optimización de hasta nivel 3, reduciendo su tiempo hasta 73.1463 milisegundos.



Imagen 5. Implementación en TBB.

#### Paralelo en CUDA

CUDA es una tecnología muy particular. Es una plataforma desarrollada por NVIDIA para otorgar herramientas de desarrollo, para ser usada en su software, ya que es enfocado al procesamiento de algoritmos en paralelo en una GPU y no en CPU, como el resto de las tecnologías, lo que resulta en un procesamiento increíblemente rápido, sobre todo en aplicaciones de gráficos (como procesamiento de imágenes). (NVIDIA Corporation, 2019)

En la Imagen 6, se puede observar su compilación y ejecución, dando un increíble tiempo de ejecución de 0.00410 milisegundos, es decir, 4.1 microsegundos; siendo, por mucho, la tecnología más rápida.

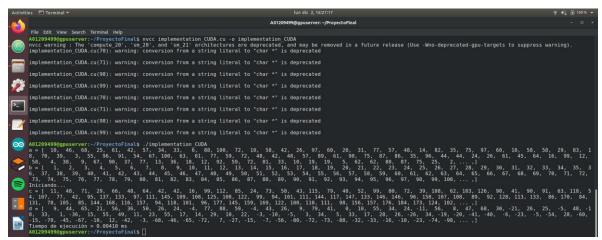


Imagen 6. Implementación en CUDA.

#### Paralelo con *Threads* en Java

Java presenta dos formas de paralelización, mediante el uso de *Threads* y mediante el uso de *Fork/Join*. La primera es una tecnología que permite a Java crear objetos (*Threads*) con el aforo de correr simultáneamente al método principal. Una forma de verlo es como contar con dos variables *program counter* en un mismo algoritmo; es decir, el mismo algoritmo corre en diferentes núcleos del procesador. (González, 2010)

Dentro de la Imagen 7 se puede observar cómo se compila el algoritmo y cómo se ejecuta, obteniendo un tiempo de ejecución de 65.3 milisegundos. Resultando ser más rápido que Java. Ya que es basado en Java, no existe una optimización como en C++.



Imagen 7. Implementación con Threads en Java.

#### Paralelo en Java con Fork/Join

Fork/Join es un *framework* de Java que permite realizar algoritmos a ejecutar en múltiples núcleos del procesador al máximo, en donde los subprocesos a paralelizar sean independientes entre sí; su manera de funcionar es distribuyendo el procesamiento en un *pool* de *Threads*. (Oracle, 2019)

En la Imagen 8, se observa el tiempo de ejecución en 0.21 milisegundos, siendo muchísimo mejor que Java y que Java con *Threads*. Tampoco cuenta con optimización, al ser basado en Java y no en C++; no obstante, la mejora es realmente significativa.



Imagen 8. Implementación en Java con Fork/Join

#### Análisis de resultados

#### Comparativa entre tecnologías secuenciales

El principal objetivo de este artículo de investigación es el análisis de los resultados obtenidos, mediante una comparativa de los mismos. Se obtendrá el *SpeedUp*, que es la división del tiempo de ejecución en secuencial sobre el tiempo de ejecución en paralelo:

$$SpeedUp = \frac{Tiempo\ Secuencial}{Tiempo\ Paralelo}$$

A continuación, se presenta una tabla comparativa de las tecnologías secuenciales y sus tiempos de ejecución:

Comparativa de Tecnologías Secuenciales		
Tecnología Tiempo de ejecución Tiempo con optimizac		Tiempo con optimización
С	371.78950	-
C++	450.62300	102.70400
Java	100.29000	-

Tabla 1: Comparativa de tecnologías secuenciales.

Por lo tanto, podemos determinar que Java tiene un procesamiento de hasta 4 veces más rápido que C y C++, en promedio. Sin embargo, con la optimización de C++ se logra una gran mejora, quedando casi a la par que Java.

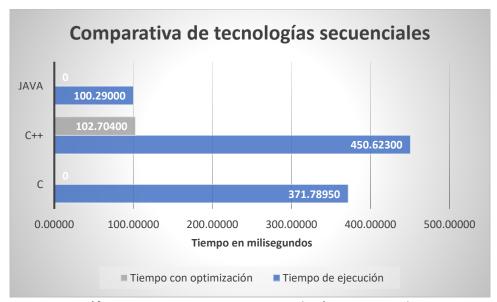


Gráfico 1: Comparativa entre tecnologías secuenciales.

#### Comparativa entre tecnologías paralelas

Continuando, se realizará la misma comparativa para las tecnologías paralelas:

Comparativa de Tecnologías Paralelas		
Tecnología	Tiempo de ejecución	Tiempo con optimización
OpenMP	98.26680	74.83060
TBB	119.50600	73.14630
CUDA	0.00390	-
Threads	65.30000	-
Fork/Join	0.21000	-

Tabla 2: Comparativa de tecnologías Paralelas.

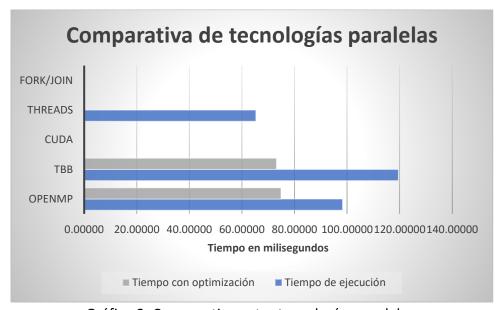


Gráfico 2: Comparativa entre tecnologías paralelas.

Se puede apreciar que la tecnología que menor tiempo de procesamiento tiene es CUDA y la más próxima a ella es Java con Fork/Join, que sigue estando muy alejada.

CUDA sobre Fork/Join representa una diferencia de velocidad de  $\frac{0.21}{0.0039} = 53.846$ .

La razón de la gran rapidez de CUDA es que utiliza el procesamiento en GPU y no en CPU. Para fines de la comparativa, podemos verlo como que CUDA tiene muchísimos núcleos de procesamiento (tantos como tenga la tarjeta de NVIDIA), mientras que el CPU se limita a 8 o 16, generalmente, de Intel.

#### Comparativa entre secuenciales y paralelas de Java

Comenzando con la tecnología de Java, se realizará la comparativa entre los tiempos de ejecución de forma secuencial y paralela:

Comparativa entre Secuencial y Paralelo de Java		
Tecnología Tiempo de ejecución SpeedUp		
Java	100.29000	-
Threads	65.30000	1.53583
Fork/Join	0.21000	477.57143

Tabla 3: Comparativa entre tiempos de ejecución en secuencial y paralelo de Java.



Gráfico 3: Comparativa entre tiempos de ejecución en secuencial y paralelo de Java.

En esta comparativa resulta evidente la ventaja en el uso de la tecnología en paralelo de *Fork/Join* para la ejecución del algoritmo, obteniendo un SpeedUp increíble de 477.57.

La razón de esto es que la tecnología *Fork/Join* utiliza un *pool* de *Threads* para resolver el algoritmo paralelizable, creando una aceleración impresionante respecto a Java.

#### Comparativa entre secuencial y paralela de C y C++

Las tecnologías C y C++ dan paso a las tecnologías de OpenMP, TBB y CUDA, por esta razón se eligió realizar una comparativa entre todas las anteriores y no solo algunas de ellas. Resultarán interesantes los resultados presentados a continuación:

Comparativa tecnologías secuenciales y paralelas de C y C++			
Tecnología	Tiempo de ejecución	SpeedUp respecto a C	SpeedUp respecto a C++
С	371.78950	-	-
C++	450.62300	-	-
OpenMP	98.26680	3.783470104	4.585709517
ТВВ	119.50600	3.111053002	3.770714441
CUDA	0.00390	95330.64103	115544.359

Tabla 4.1: Comparativa tecnologías secuenciales y paralelas de C y C++

Comparativa tecnologías secuenciales y paralelas de C++ optimizadas		
Tecnología	Tiempo con optimización	SpeedUp respecto a C++
C++	102.70400	-
OpenMP	74.83060	1.372486657
TBB	73.14630	1.404090159
CUDA	0.00390	26334.35897

Tabla 4.2: Comparativa tecnologías secuenciales y paralelas de C++ optimizadas

Sin la necesidad de presentar un gráfico, es evidente que CUDA resulta ser, por mucho, la tecnología con menor tiempo de procesamiento, sin importar que se le compare con o sin optimización a las demás tecnologías. El motivo de esto es explicado en la sección <a href="Comparativa entre tecnologías paralelas">Comparativa entre tecnologías paralelas</a>, de este documento.

Dejando de lado a CUDA, se puede concluir que los mejores *SpeedUp* son aquellos respecto a C++ sin optimización, pero una vez que se optimizan, el *SpeedUp* disminuye. No obstante, se puede concluir, al igual que en la comparativa de Java, que las tecnologías paralelas ayudan enormemente a aprovechar los recursos de un multiprocesador.

#### Comparativa de todas las tecnologías implementadas:

Finalmente, a manera de resumen, una comparativa de todas las tecnologías implementadas en este artículo de investigación:

Comparativa de Tecnologías		
Tecnología	Tiempo de ejecución	Tiempo con optimización
С	371.78950	-
C++	450.62300	102.70400
Java	100.29000	-
OpenMP	98.26680	74.83060
TBB	119.50600	73.14630
CUDA	0.00390	-
Threads	65.30000	-
Fork/Join	0.21000	-

Tabla 5: Comparativa de todas las tecnologías implementadas.



Gráfico 5: Comparativa de todas las tecnologías implementadas.

#### Conclusión

Con base en los resultados obtenidos en este articulo de investigación, podemos concluir que la solución de algoritmos en paralelo representa un aprovechamiento de los recursos disponibles del dispositivo, específicamente del procesador o GPU, dando como resultado la resolución del problema en un tiempo extremadamente inferior a la forma secuencial.

En este articulo se realizó la prueba con un algoritmo "sencillo", de operaciones con vectores, pero cuando hablamos de otro algoritmo más complejo, el *SpeedUp* obtenido será mucho más apreciable.

Me gustaría recalcar que la tecnología de CUDA fue superior la resto, logrando un excepcional tiempo de ejecución. Sin olvidar que el procesamiento es realizado en el GPU, por lo que el primer paso es contar con uno... y no es el caso de todas las personas. Además, he de señalar que CUDA, al ser resuelto en GPU, trabajará mejor para problemas asociados a vectores, matrices y todo lo relacionado al procesamiento de imágenes, entre otros. He de recordar también que CUDA entiende el cómputo heterogéneo, es decir, CPU y GPU combinados para obtener lo mejor de cada uno.

Consecutivamente, he de marcar que la solución del algoritmo mediante la técnica de *Fork/Join* en Java resultó la mejor opción (sin tomar en cuenta CUDA ni GPU). Logrando un increíble SpeedUp entre el mismo Java secuencial y Java con *Threads*.

Finalmente, en cuanto a las soluciones en forma secuencial, Java también resultó ganador.

Para concluir, si he de optar por tomar un lenguaje en particular, me quedo con Java, porque tanto en la parte secuencial como paralela resultó ser el mejor; con excepción de CUDA, pero CUDA requiere de un GPU.

#### Referencias

- Atom. (2019). Atom. Obtenido de Atom: https://atom.io/
- Balagurusamy, E. (2007). Programación orientada a objetos con C++. Madrid: McGraw-Hill.
- Brian Kernighan, D. R. (1991). El lenguaje de programación C. Pearson Educación.
- González, A. (Agosto de 2010). *Hilos (Threads) en Java*. Obtenido de Departamento de Electrónica | Universidad Técnica Federico Santa María: http://profesores.elo.utfsm.cl/~agv/elo330/2s10/lectures/Java/threads/JavaThreads.html
- Herb Sutter. (30 de marzo de 2005). *The Free Lunch Is Over: A Fundamental Turn Toward Concurrency in Software*. Obtenido de http://www.gotw.ca/publications/concurrency-ddj.htm
- Herb Sutter. (2011). *Welcome to the Jungle*. Obtenido de Herb Sutter: https://herbsutter.com/welcome-to-the-jungle/
- Intel Corporation. (2019). *Intel®Threading Building Blocks*. Obtenido de Intel: https://software.intel.com/en-us/tbb
- J. Steven Perry. (03 de diciembre de 2012). Conceptos básicos del lenguaje Java. Obtenido de IBM.com: https://www.ibm.com/developerworks/ssa/java/tutorials/jintrotojava1/index.html
- Leonardo Dagum, R. M. (1998). *OpenMP: An IndustryStandard API for SharedMemory Programming.* IEEE COMPUTATIONAL SCIENCE & ENGINEERING.
- NVIDIA Corporation. (2019). *CUDA Zone*. Obtenido de NVIDIA: https://developer.nvidia.com/cuda-zone
- OpenMP. (2019). *OpenMP Books*. Obtenido de OpenMP: https://www.openmp.org/resources/openmp-books/

Oracle. (2019). Fork/Join. Obtenido de The Java Tutorials: https://docs.oracle.com/javase/tutorial/essential/concurrency/forkjoin.html

#### Anexos

#### Librería cheader.h

```
#ifndef TIMER H
#define TIMER H
#include <stdlib.h>
#include <time.h>
#include <sys/time.h>
#include <sys/types.h>
#define N
                         10
#define DISPLAY
                         100
#define MAX VALUE 10000
struct timeval startTime, stopTime;
int started = 0;
void start timer() {
     started = 1;
     gettimeofday(&startTime, NULL);
}
double stop timer() {
     long seconds, useconds;
     double duration = -1;
     if (started) {
          gettimeofday(&stopTime, NULL);
          seconds = stopTime.tv sec - startTime.tv sec;
          useconds = stopTime.tv usec - startTime.tv usec;
          duration = (seconds * 1000.0) + (useconds /
1000.0);
          started = 0;
     }
     return duration;
}
void random array(int *array, int size) {
     int i;
     srand(time(0));
     for (i = 0; i < size; i++) {
         array[i] = (rand() % 100) + 1;
     }
}
```

```
void fill_array(int *array, int size) {
    int i;
    for (i = 0; i < size; i++) {
        array[i] = (i % MAX_VALUE) + 1;
    }
}

void display_array(char *text, int *array) {
    int i;
    printf("%s = [%4i", text, array[0]);
    for (i = 1; i < DISPLAY; i++) {
        printf(",%4i", array[i]);
    }
    printf(", ..., ]\n");
}

#endif /* TIMER_H_ */</pre>
```

#### Librería cppheader.h

```
#ifndef TIMER H
#define TIMER H
#include <ctime>
#include <cstdio>
#include <cstdlib>
#include <sys/time.h>
#include <sys/types.h>
const int N = 10;
const int DISPLAY = 100;
const int MAX VALUE = 10000;
class Timer {
private:
    timeval startTime;
    bool started;
public:
    Timer() :started(false) {}
    void start() {
    started = true;
        gettimeofday(&startTime, NULL);
    double stop() {
        timeval endTime;
        long seconds, useconds;
        double duration = -1;
        if (started) {
               gettimeofday(&endTime, NULL);
               seconds = endTime.tv sec - startTime.tv sec;
               useconds = endTime.tv usec -
startTime.tv usec;
               duration = (seconds * 1000.0) + (useconds /
1000.0);
               started = false;
          return duration;
    }
};
void random array(int *array, int size) {
     int i;
```

```
srand(time(0));
     for (i = 0; i < size; i++) {
          array[i] = (rand() % 100) + 1;
}
void fill array(int *array, int size) {
     int i;
     for (i = 0; i < size; i++) {
         array[i] = (i % MAX VALUE) + 1;
     }
}
void display array(const char *text, int *array) {
     int i;
     printf("%s = [%4i", text, array[0]);
     for (i = 1; i < DISPLAY; i++) {
         printf(",%4i", array[i]);
     printf(", ...,]\n");
}
#endif
```

#### Implementación en C

```
/*----
-----
* Multiprocesadores: Proyecto final.
* Fecha: 01 de diciembre del 2019.
* Autor: Gerardo Daniel Naranjo Gallegos, A01209499.
* Implementación: C.
_____*/
// -----
----//
//
          LIBRERÍAS
// -----
----//
#include <stdio.h>
#include <stdlib.h>
#include <math.h>
#include "utils/cheader.h"
// -----
----//
//
                          DEFINIR
VARIABLES QUE NO CAMBIAN
// -----
----//
#define SIZE 100000000
// -----
----//
```

```
//
                 FUNCIÓN
----//
void calculate vectors (int* suma, int* resta, int* vector1,
int* vector2) {
 for (int i = 0; i < SIZE; i++) {
   suma[i] = vector1[i] + vector2[i];
   resta[i] = vector1[i] - vector2[i];
 }
}
// -----
----//
//
                     MAIN
                                                   //
// -----
----//
int main(int argc, char const *argv[]) {
 // Definir variables a utilizar:
 int *vector1, *vector2, *suma, *resta;
 double ms = 0.0;
 // Inicializar variables:
 vector1 = (int *) malloc(sizeof(int) * SIZE);
    vector2 = (int *) malloc(sizeof(int) * SIZE);
 suma = (int *) malloc(sizeof(int) * SIZE);
 resta = (int *) malloc(sizeof(int) * SIZE);
 // Rellenar vectores utilizados:
 random array(vector1, SIZE);
    fill array(vector2, SIZE);
 // Mostrar ambos vectores:
 display array("vector1", vector1);
    display array("vector2", vector2);
 // Imprimir mensaje de inicio:
 printf("Iniciando... \n");
 // Método FOR para ejecutar la función:
 for (int i = 0; i < N; i++) {
   start timer();
```

```
calculate vectors(suma, resta, vector1, vector2);
   ms += stop_timer();
  }
  // Imprimir/Mostrar en la terminal los vectores
resultantes:
  display array("suma", suma);
  display array("resta", resta);
  // Imprimir/Mostrar en terminal el tiempo de ejecución del
programa:
  printf("Tiempo de ejecución = %.51f ms \n", (ms / N));
  // Liberar la memoria utilizada por las variables:
  free (vector1);
  free (vector2);
  free (suma);
  free(resta);
  // Fin del programa:
  return 0;
}
```

```
Implementación en C++
/*-----
* Multiprocesadores: Proyecto final.
* Fecha: 01 de diciembre del 2019.
* Autor: Gerardo Daniel Naranjo Gallegos, A01209499.
* Implementación: C++.
*-----
_____*/
// -----
----//
//
         LIBRERÍAS
// -----
----//
#include <iostream>
#include "utils/cppheader.h"
// -----
----//
//
      DEFINIR NAMESPACE
                    //
// -----
----//
using namespace std;
// -----
----//
```

```
DEFINIR
VARIABLES QUE NO CAMBIAN
----//
const int SIZE = 100000000;
// -----
_____//
//
              FUNCIÓN
// -----
----//
class calculate vectors {
private:
   int *vector1, *vector2, *suma, *resta, size;
public:
   calculate vectors(int *arrayD, int *arrayC, int *arrayA,
int *arrayB, int s)
      : resta(arrayD), suma(arrayC), vector1(arrayA),
vector2(arrayB), size(s) {}
   void calculate() {
       for (int i = 0; i < size; i++) {
          suma[i] = vector1[i] + vector2[i];
          resta[i] = vector1[i] - vector2[i];
       }
   }
};
// -----
----//
//
                 MAIN
                                          //
// -----
_____//
int main(int argc, char const *argv[]) {
   // Definir variables a utilizar:
   Timer t:
   double ms = 0.0;
```

```
// Inicializar variables:
     int *vector1 = new int[SIZE];
     int *vector2 = new int[SIZE];
     int *suma = new int[SIZE];
     int *resta = new int[SIZE];
     // Rellenar vectores utilizados:
     random array(vector1, SIZE);
     fill array (vector2, SIZE);
     // Mostrar ambos vectores:
     display_array("vector1", vector1);
     display array("vector2", vector2);
     // Inicializar a la función:
     calculate vectors cv(resta, suma, vector1, vector2,
SIZE);
     // Imprimir mensaje de inicio:
     cout << "Iniciando..." << endl;</pre>
     // Método FOR para ejecutar la función:
     for (int i = 0; i < N; i++) {
          t.start();
          cv.calculate();
         ms += t.stop();
     }
     // Imprimir/Mostrar en la terminal los vectores
resultantes:
     display array("suma", suma);
     display array("resta", resta);
     // Imprimir/Mostrar en terminal el tiempo de ejecución
del programa:
     cout << "Tiempo de ejecución = " << (ms/N) << " ms" <<
endl;
     // Liberar la memoria utilizada por las variables:
     delete [] vector1;
     delete [] vector2;
     delete [] suma;
     delete [] resta;
     // Fin del programa:
     return 0;
}
```

#### Implementación en Java

```
/*-----
* Multiprocesadores: Proyecto final.
* Fecha: 01 de diciembre del 2019.
* Autor: Gerardo Daniel Naranjo Gallegos, A01209499.
* Implementación: JAVA.
*-----
_____*/
// -----
_____//
//
              FUNCIÓN
// -----
 ----//
public class calculate vectors {
   private int vector1[], vector2[], suma[], resta[];
   public calculate vectors(int resta[], int suma[], int
vector1[], int vector2[]) {
       this.resta = resta;
       this.suma = suma;
       this.vector1 = vector1;
       this.vector2 = vector2;
   }
   public void calculate() {
       for (int i = 0; i < suma.length; <math>i++) {
           suma[i] = vector1[i] + vector2[i];
           resta[i] = vector1[i] - vector2[i];
   }
```

```
// -----
----//
                      MAIN
                                                    //
// -----
_____//
    public static void main(String args[]) {
        // Definir variables a utilizar:
        final int SIZE = 100 000 000;
        long startTime = 0;
        long stopTime = 0;
        double acum = 0.0;
        // Inicializar variables:
        int vector1[] = new int[SIZE];
        int vector2[] = new int[SIZE];
        int suma[] = new int[SIZE];
        int resta[] = new int[SIZE];
        // Rellenar vectores utilizados:
        Utils.randomArray(vector1);
        Utils.fillArray(vector2);
        // Mostrar ambos vectores:
        Utils.displayArray("vector1", vector1);
        Utils.displayArray("vector2", vector2);
        // Llamada a la función:
        calculate vectors e = new calculate vectors(resta,
suma, vector1, vector2);
        // Método FOR para ejecutar la función:
        for (int i = 0; i < Utils.N; i++) {
             startTime = System.currentTimeMillis();
             e.calculate();
             stopTime = System.currentTimeMillis();
             acum += (stopTime - startTime);
         }
        // Imprimir/Mostrar en la terminal los vectores
resultantes:
        Utils.displayArray("suma", suma);
        Utils.displayArray("resta", resta);
```

# Implementación en OpenMP /\*----\* Multiprocesadores: Proyecto final. \* Fecha: 01 de diciembre del 2019. \* Autor: Gerardo Daniel Naranjo Gallegos, A01209499. \* Implementación: OpenMP. \*-----\_\_\_\_\_\*/ // ---------// // LIBRERÍAS // ---------// #include <stdio.h> #include <stdlib.h> #include <math.h> #include "utils/cheader.h" // ---------// // DEFINIR VARIABLES QUE NO CAMBIAN -----// ---------// #define SIZE 100000000 // ---------//

```
//
                  FUNCIÓN
----//
void calculate vectors(int *resta, int *suma, int *vector1,
int *vector2, int size) {
    #pragma omp parallel for shared(resta, suma, vector1,
vector2, size)
    for (int i = 0; i < size; i++) {
         suma[i] = vector1[i] + vector2[i];
         resta[i] = vector1[i] - vector2[i];
}
____//
//
                       MAIN
                                                       //
// -----
----//
int main(int argc, char const *argv[]) {
    // Definir variables a utilizar:
    int *vector1, *vector2, *suma, *resta;
    double ms = 0.0, result = 0.0;
    // Inicializar variables:
    vector1 = (int *) malloc(sizeof(int) * SIZE);
    vector2 = (int *) malloc(sizeof(int) * SIZE);
    suma = (int *) malloc(sizeof(int) * SIZE);
    resta = (int *) malloc(sizeof(int) * SIZE);
    // Rellenar vectores utilizados:
    random array(vector1, SIZE);
    fill array (vector2, SIZE);
    // Mostrar ambos vectores:
    display array("vector1", vector1);
    display array("vector2", vector2);
    // Imprimir mensaje de inicio:
    printf("Iniciando...\n");
    // Método FOR para ejecutar la parte en paralelo:
```

```
for (int j = 0; j < N; j++) {
          start timer();
          calculate vectors (resta, suma, vector1, vector2,
SIZE);
         ms += stop timer();
     }
     // Imprimir/Mostrar en la terminal los vectores
resultantes:
     display array("suma", suma);
     display array("resta", resta);
     // Imprimir/Mostrar en terminal el tiempo de ejecución
del programa:
     printf("Tiempo de ejecución = %.51f ms\n", (ms / N));
     // Liberar la memoria utilizada por las variables:
     free (vector1);
     free (vector2);
     free(suma);
     free (resta);
     // Fin del programa:
     return 0;
}
```

```
Implementación en TBB
/*----
* Multiprocesadores: Proyecto final.
* Fecha: 01 de diciembre del 2019.
* Autor: Gerardo Daniel Naranjo Gallegos, A01209499.
* Implementación: TBB.
*-----
_____*/
// -----
----//
//
          LIBRERÍAS
                             //
// -----
----//
#include <iostream>
#include <tbb/task scheduler init.h>
#include <tbb/parallel for.h>
#include <tbb/blocked_range.h>
#include "utils/cppheader.h"
// -----
----//
//
       DEFINIR NAMESPACE
                       //
// -----
----//
using namespace std;
using namespace tbb;
// -----
----//
```

```
DEFINIR
VARIABLES QUE NO CAMBIAN
----//
const int SIZE = 100000000;
const int GRAIN = 10000;
// -----
_____//
//
              FUNCIÓN
// -----
 ----//
class calculate vectors {
private:
   int *vector1, *vector2, *suma, *resta;
public:
   calculate vectors(int *arrayD, int *arrayC, int *arrayA,
       : resta(arrayD), suma(arrayC), vector1(arrayA),
vector2(arrayB) {}
   void operator() (const blocked range<int> &r) const {
       for (int i = r.begin(); i != r.end(); i++) {
          suma[i] = vector1[i] + vector2[i];
          resta[i] = vector1[i] - vector2[i];
   }
};
// -----
----//
//
                  MAIN
                                          //
// -----
_____//
int main(int argc, char const *argv[]) {
   // Definir variables a utilizar:
   Timer t;
   double ms = 0.0;
```

```
// Inicializar variables:
     int *vector1 = new int[SIZE];
     int *vector2 = new int[SIZE];
     int *suma = new int[SIZE];
     int *resta = new int[SIZE];
     // Rellenar vectores utilizados:
     random array(vector1, SIZE);
     fill array(vector2, SIZE);
     // Mostrar ambos vectores:
     display array("vector1", vector1);
     display array("vector2", vector2);
     // Imprimir mensaje de inicio:
     cout << "Iniciando..." << endl;</pre>
     // Método FOR para ejecutar la parte en paralelo:
     for (int i = 0; i < N; i++) {
          t.start();
          parallel for (blocked range<int>(0, SIZE, GRAIN),
            calculate vectors (resta, suma, vector1,
vector2));
         ms += t.stop();
     }
     // Imprimir/Mostrar en la terminal los vectores
resultantes:
     display array("suma", suma);
     display array("resta", resta);
     // Imprimir/Mostrar en terminal el tiempo de ejecución
del programa:
     cout << "Tiempo de ejecución = " << (ms/N) << " ms" <<
endl;
     // Liberar la memoria utilizada por las variables:
     delete [] vector1;
     delete [] vector2;
     delete [] suma;
     delete [] resta;
     // Fin del programa:
     return 0;
}
```

# Implementación en CUDA /\*-----\* Multiprocesadores: Proyecto final. \* Fecha: 01 de diciembre del 2019. \* Autor: Gerardo Daniel Naranjo Gallegos, A01209499. \* Implementación: CUDA. \*-----\_\_\_\_\_\*/ // ---------// // LIBRERÍAS // ---------// #include <stdio.h> #include <stdlib.h> #include <math.h> #include "utils/cheader.h" // -----\_\_\_\_\_// // DEFINIR VARIABLES QUE NO CAMBIAN -----// ---------// #define MIN(vector1, vector2) (vector1<vector2?vector1:vector2)</pre> #define SIZE 100000000 #define THREADS #define BLOCKS MIN(32, (SIZE + THREADS - 1) / THREADS)

```
----//
                  FUNCIÓN
----//
global void calculate(int *resta, int *suma, int
*vector1, int *vector2) {
    int i = threadIdx.x + (blockIdx.x * blockDim.x);
    if (i < SIZE) {
         suma[i] = vector1[i] + vector2[i];
         resta[i] = vector1[i] - vector2[i];
    }
}
//
                      MAIN
// -----
int main(int argc, char const *argv[]) {
    // Definir variables a utilizar:
    int *vector1, *vector2, *suma, *resta;
    int *device vector1, *device vector2, *device suma,
*device resta;
    double ms = 0.0;
    // Inicializar variables:
    vector1 = (int *) malloc(sizeof(int) * SIZE);
    vector2 = (int *) malloc(sizeof(int) * SIZE);
    suma = (int *) malloc(sizeof(int) * SIZE);
    resta = (int *) malloc(sizeof(int) * SIZE);
    // Rellenar vectores utilizados:
    random array(vector1, SIZE);
    fill array(vector2, SIZE);
    // Mostrar ambos vectores:
    display array("a", vector1);
    display array("b", vector2);
```

```
// Asignar memoria a las variables en el dispositivo
(GPU):
     cudaMalloc((void**) &device vector1, SIZE *
sizeof(int));
     cudaMalloc((void**) &device vector2, SIZE *
sizeof(int));
     cudaMalloc((void**) &device suma, SIZE * sizeof(int));
     cudaMalloc((void**) &device resta, SIZE * sizeof(int));
     // Copiar vectores a vectores en el dispositivo (GPU):
     cudaMemcpy(device vector1, vector1, SIZE * sizeof(int),
cudaMemcpyHostToDevice);
     cudaMemcpy(device vector2, vector2, SIZE * sizeof(int),
cudaMemcpyHostToDevice);
     // Imprimir mensaje de inicio:
     printf("Iniciando...\n");
     // Método FOR para ejecutar la parte en paralelo en la
GPU:
     for (int j = 0; j < N; j++) {
          start timer();
          calculate <<< BLOCKS, THREADS>>> (device resta,
device suma, device vector1, device vector2);
         ms += stop timer();
     }
     // Copiar vectores resultantes del GPU al CPU:
     cudaMemcpy(suma, device suma, SIZE * sizeof(int),
cudaMemcpyDeviceToHost);
     cudaMemcpy(resta, device resta, SIZE * sizeof(int),
cudaMemcpyDeviceToHost);
     // Imprimir/Mostrar en la terminal los vectores
resultantes:
     display array("c", suma);
     display array("d", resta);
     // Imprimir/Mostrar en terminal el tiempo de ejecución
del programa:
    printf("Tiempo de ejecución = %.51f ms\n", (ms / N));
     // Liberar la memoria utilizada en el GPU por las
variables:
     cudaFree(device vector1);
     cudaFree(device vector2);
```

```
cudaFree(device_suma);
cudaFree(device_resta);

// Liberar la memoria del CPU utilizada por las
variables:
    free(vector1);
    free(vector2);
    free(suma);
    free(resta);

// Fin del programa:
    return 0;
}
```

# Implementación Threads /\*----- \* \* Multiprocesadores: Proyecto final. \* Fecha: 01 de diciembre del 2019. \* Autor: Gerardo Daniel Naranjo Gallegos, A01209499. \* Implementación: THREADS. \*

```
____//
//
                  FUNCIÓN
// -----
----//
public class calculate vectors JavaThreads extends Thread {
    private int vector1[], vector2[], suma[], resta[],
inicio, fin;
    public calculate vectors JavaThreads(int resta[], int
suma[], int vector1[], int vector2[], int inicio, int fin) {
         this.vector1 = vector1;
         this.vector2 = vector2;
         this.suma = suma;
         this.resta = resta;
         this.inicio = inicio;
        this.fin = fin;
    }
    public void run() {
         for (int i = inicio; i < fin; i++) {</pre>
             suma[i] = vector1[i] + vector2[i];
             resta[i] = vector1[i] - vector2[i];
```

// -----

```
}
    }
----//
//
                       MAIN
                                                       //
// -----
----//
    public static void main(String args[]) {
         // Definir variables a utilizar:
         final int SIZE = 100 000 000;
         calculate vectors JavaThreads threads[];
         int block;
         long inicioTiempo, finTiempo;
         double acum = 0.0;
         // Inicializar variables:
         int vector1[] = new int[SIZE];
         int vector2[] = new int[SIZE];
         int suma[] = new int[SIZE];
         int resta[] = new int[SIZE];
         // Rellenar vectores utilizados:
         Utils.randomArray(vector1);
         Utils.fillArray(vector2);
         // Mostrar ambos vectores:
         Utils.displayArray("vector1", vector1);
         Utils.displayArray("vector2", vector2);
         // Asignación del tamaño del bloque, según el
número de threads disponibles en el dispositivo:
         block = SIZE / Utils.MAXTHREADS;
         threads = new
calculate vectors JavaThreads[Utils.MAXTHREADS];
         // Método FOR para ejecutar la parte en paralelo:
         for (int j = 1; j \le Utils.N; j++) {
              for (int m = 0; m < threads.length; m++) {</pre>
                   if (m != threads.length - 1) {
                       threads[m] = new
calculate vectors JavaThreads(resta, suma, vector1, vector2,
(m * block), ((m + 1) * block));
                  } else {
```

```
threads[m] = new
calculate_vectors_JavaThreads(resta, suma, vector1, vector2,
(m * block), Utils.N);
               inicioTiempo = System.currentTimeMillis();
               for (int k = 0; k < threads.length; k++) {
                    threads[k].start();
               for (int l = 0; l < threads.length; <math>l++) {
                    try {
                         threads[1].join();
                    } catch (InterruptedException e) {
                         e.printStackTrace();
                    }
               finTiempo = System.currentTimeMillis();
               acum += (finTiempo - inicioTiempo);
          }
          // Imprimir/Mostrar en la terminal los vectores
resultantes:
          Utils.displayArray("suma", suma);
          Utils.displayArray("resta", resta);
          // Imprimir/Mostrar en terminal el tiempo de
ejecución del programa :
          System.out.printf("Tiempo de ejecución = %.5f ms
\n", (acum / Utils.N));
     }
}
```

```
Implementación Fork/Join
/*----
* Multiprocesadores: Proyecto final.
* Fecha: 01 de diciembre del 2019.
* Autor: Gerardo Daniel Naranjo Gallegos, A01209499.
* Implementación: FORK/JOIN.
*-----
_____*/
// -----
----//
//
            LIBRERÍAS
// -----
----//
import java.util.concurrent.RecursiveAction;
import java.util.concurrent.ForkJoinPool;
// -----
----//
//
            FUNCIÓN
// -----
----//
public class calculate vectors FORKJOIN extends
RecursiveAction {
   // Variables que no cambian:
   private static final int SIZE = 100 000 000;
   private static final int MIN = 100 \ \overline{000};
```

```
private int vector1[], vector2[], suma[], resta[],
inicio, fin;
    public calculate vectors FORKJOIN(int inicio, int
fin,int resta[], int suma[], int vector1[], int vector2[]) {
         this.inicio = inicio;
         this.fin = fin;
         this.resta = resta;
         this.suma = suma;
         this.vector1 = vector1;
         this.vector2 = vector2;
    }
    protected void computeDirectly() {
         for (int i = inicio; i < fin; i++) {
              suma[i] = vector1[i] + vector2[i];
              resta[i] = vector1[i] - vector2[i];
         }
    }
    @Override
    protected void compute() {
         if ( (fin - inicio) <= MIN ) {
              computeDirectly();
         } else {
              int mitad = inicio + ( (fin - inicio) / 2 );
              invokeAll(
                  new calculate_vectors FORKJOIN(inicio,
mitad, resta, suma, vector1, vector2),
                  new calculate vectors FORKJOIN (mitad,
fin, resta, suma, vector1, vector2)
              );
         }
    }
_____//
//
                       MAIN
                                                        //
// -----
_____//
    public static void main(String args[]) {
         // Definir variables a utilizar:
         ForkJoinPool pool;
         long inicioTiempo, finTiempo;
```

```
double acum = 0.0;
          // Inicializar variables:
          int vector1[] = new int[SIZE];
          int vector2[] = new int[SIZE];
          int suma[] = new int[SIZE];
          int resta[] = new int[SIZE];
          // Rellenar vectores utilizados:
          Utils.randomArray(vector1);
          Utils.fillArray(vector2);
          // Mostrar ambos vectores:
          Utils.displayArray("vector1", vector1);
          Utils.displayArray("vector2", vector2);
          // Método FOR para ejecutar la parte en paralelo:
          for (int j = 0; j < Utils.N; j++) {
               inicioTiempo = System.currentTimeMillis();
               pool = new ForkJoinPool(Utils.MAXTHREADS);
               pool.invoke(new calculate vectors FORKJOIN(0,
Utils.N, resta, suma, vector1, vector2));
               finTiempo = System.currentTimeMillis();
               acum += (finTiempo - inicioTiempo);
          }
          // Imprimir/Mostrar en la terminal los vectores
resultantes:
          Utils.displayArray("suma", suma);
          Utils.displayArray("resta", resta);
          // Imprimir/Mostrar en terminal el tiempo de
ejecución del programa:
          System.out.printf("Tiempo de ejecución = %.5f ms
\n", (acum / Utils.N));
}
```