

Tarea 1 - Opción 1 – SSS-Index (Sistemas Distribuidos)

Esta tarea consiste en crear un índice, denominado SSS-Index utilizando el modelo de programación distribuida MPI. Este índice es utilizado para indexar elementos multimedia, de tal forma de procesar consultas de manera eficiente realizando descarte de elementos.

Este índice consiste en una tabla de distancias entre pivotes y el resto de los elementos de la base de datos. Luego, este índice es capaz de realizar un descarte de elementos utilizando estas distancias almacenadas. La creación de este índice sigue los siguientes pasos (de acuerdo a la figura 1):

- 1) El primer pivote (P_0 en la figura 1) corresponde a un elemento de la base de datos que es elegido aleatoriamente. Para efectos de esta tarea P_0 será el primer vector de la lista que se le proporcionará como base de datos.
- 2) El segundo pivote es cualquier elemento de la base de datos que este al menos a distancia $M \cdot \alpha$ de P_0 .
- 3) Desde el tercer pivote en adelante, un nuevo pivote es aquel elemento que este al menos a distancia $M \cdot \alpha$ de todos los pivotes previos.
- 4) Luego de que no es posible obtener más pivotes, se debe calcular la distancia entre todos los elementos pivotes y todos los elementos restantes.

M = Distancia máxima entre dos elementos.

α = Constante real.

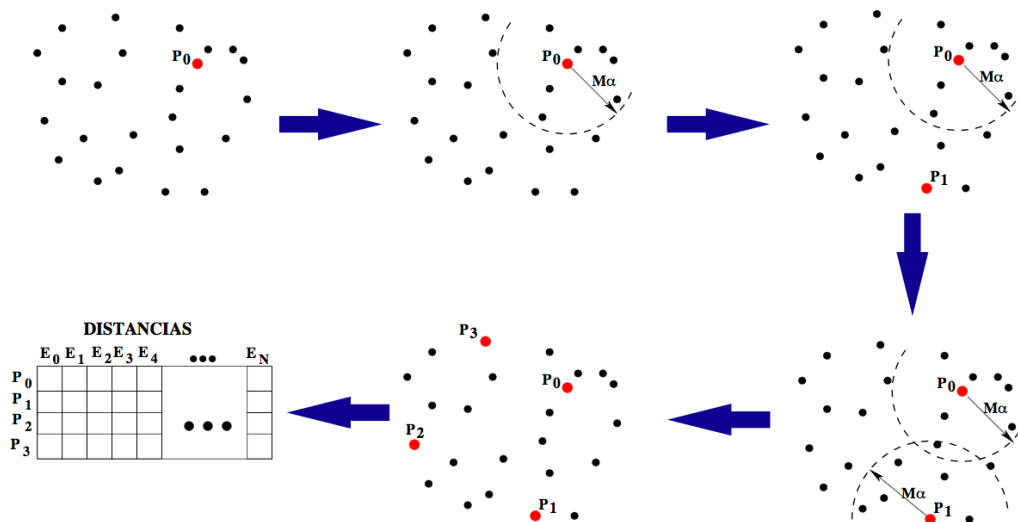


Figura 1.

Esta tarea consiste solamente en la creación del índice SSS. Para esto, usted deberá realizar los cálculos de distancia entre elementos de manera distribuida entre los nodos de su programa MPI. Usted decida la forma de distribuir el trabajo del cálculo de distancias.

Cada elemento de la base de datos es un vector de dimensión 20. Utilice `#define DIM 20` para indicar la dimensión de los elementos en su programa en C. Le será entregado una base de datos de prueba junto con el valor de M . Use un valor de $\alpha=0,4$. Debe usar la distancia euclidiana como función de distancia entre elementos.

Su tarea será revisada utilizando archivos de prueba con el siguiente formato:

test.txt:

```
5
20.4
0.137 -0.03 0.09 0.10 0.2 -0.1 -0.05 0.08 0.7 0.6 0.01 0.2 0.4 0.03 0.4 0.6 0.2 0.5 0.6 0.9
-0.3 0.9 0.18 0.4 0.1 0.03 0.77 -0.2 1.3 0.1 0.3 0.1 0.02 0.6 0.7 0.1 -0.8 0.8 0.7 0.083
-0.0 0.2 -0.031 0.24 0.1 0.5 0.7 0.8 -0.1 0.5 -0.0 -0.2 0.1 0.5 0.02 0.06 0.96 0.56 0.16 0.30
0.9 0.5 0.26 0.1 -0.9 0.003 -0.8 0.6 0.03 0.7 0.1 -0.9 0.02 0.3 0.3 0.0 0.7 0.65 0.028 -0.044
0.4 -0.1 0.0 0.1 0.5 -0.0 0.8 0.8 -0.6 0.4 0.27 0.4 0.34 0.5 -0.2 0.82 -0.1 0.3 0.043 0.246
```

La primera fila indica la cantidad de elementos (vectores) de la base de datos. La segunda fila el valor de M. Desde la tercera fila en adelante están los vectores que conforman la base de datos. Las coordenadas entre sí están separadas por un espacio. Hay un elemento de la base de datos por fila.

De esta forma, su programa podrá ser ejecutado redirigiendo la entrada desde el teclado al archivo de prueba, de la siguiente manera:

```
mpirun -np X a.out < test.txt
```

Donde X puede ser cualquier valor entre 3 y 24.

La salida de su programa debe imprimir por pantalla solamente los elementos de la matriz de distancias (ver Figura 1). No debe imprimir ningún otro dato.

Para calificar su tarea, se tomará en cuenta los siguientes ítems:

- 1) Correcta lectura de los datos.
- 2) Adecuada transferencia de información entre nodos.
- 3) Correcta distribución de los cálculos de distancia entre elementos.
- 4) Correcta impresión por pantalla como salida de su programa.

Notas :

- Plazo de entrega: Por LMS a más tardar el **26 de Mayo del 2022**.
- El trabajo es de 2 personas.
- Debe enviar un email indicando con quién trabajará y cuál opción de Tarea abordará, al email ricardo.j.barrientos@gmail.com a más tardar el Lunes 16 de Mayo.