Universidad del Valle de Guatemala
Computación Paralela y Distribuida
Docente: Sebastián Galindo
Semestre 2, 2024

UVG
UNIVERSIDAD DEL VALLE DE GUATEMALA

Hoja de Trabajo # 3

Fecha de Entrega: 10 de noviembre, 2023.

Descripción: De forma individual, realice los siguientes ejercicios y prácticas para comenzar a familiarizarnos con CUDA y programación en GPUs. Para cada inciso, incluya evidencia de su procedimiento o salida (capturas de pantalla, etc., en un PDF). No olvide adjuntar su código. Dependiendo de si usa su computador o el Lab, puede que algunos comandos cambien levemente, por lo que deben estar atentos en dado caso para indagar y utilizar el comando adecuado.

Entregables: Deberá entregar un documento con las respuestas a las preguntas planteadas en cada ejercicio (incluyendo diagramas o screenshots si es necesario), junto con todos los archivos de código que programe debidamente comentados e identificados. La entrega de la hoja es individual.

Materiales: necesitará una máquina con GPU Nvidia. Puede utilizar Google Collab para esta actividad.

Contenido

Ejercicio 1 (50 puntos)

El programa hello.cu ilustra la forma básica del modelo de ejecución para CUDA. Realice las siguientes acciones para comprender el efecto de la configuración del kernel y su relación con el Compute Capability de una tarjeta.

1. Compile el programa (ignore la advertencia sobre código deprecado en caso le salga):

   $ nvcc hello.cu -o hello

2. Ejecute el programa. Observe cuántas veces se imprime el mensaje y su conexión con la configuración de la llamada al kernel − hello<<<g,b>>>():

   $ ./hello

3. Modifique el programa para correr 2 bloques de 1024 hilos. Modificarlo también para que imprima su nombre y carnet. Busque en el despliegue de consola el mensaje del último hilo de la serie (1023).

   CAPTURA DE PANTALLA DE LA EJECUCIÓN CON 2 BLOQUES DE 1024 HILOS

4. Busque en el sitio de Nvidia el Compute Capability de la tarjeta que poseen las máquinas del Laboratorio (o de la computadora que está utilizando). Escriba acá el valor de CC y busque la tabla resumen con las características técnicas del CC:

   Compute Capability:

5. Modifique el programa para correr 1 bloque de 2048 hilos.

   CAPTURA DE PANTALLA DE LA EJECUCIÓN CON 1 BLOQUE DE 2048 HILOS
   EXPLIQUE EN POCAS PALABRAS EL RESULTADO

6. Busque en la tabla de CC los siguientes datos de la GPU que está utilizando:
    i. Warp size
    ii. Maximum number of threads per block
    iii. Maximum dimensionality of a grid of thread blocks
    iv. Maximum size per grid dimension
    v. Maximum dimensionality of a thread block
    vi. Maximum size per block dimension

Ejercicio 2 (50 puntos)

El programa hello2.cu ilustra la forma para calcular un identificador global al momento de usar hilos que pertenecen a bloques diferentes. Realice las siguientes acciones para comprender el efecto de la configuración del kernel y su relación con la forma de calcular el ID único de los hilos.

1. Descargue, compile y ejecute hello2.cu. Observe la relación de la configuración de la llamada al kernel con la geometría de los hilos y el resultado. Escriba la respuesta a los dos enunciados:
    i. Máximo ID de los hilos:
    ii. Ejecución de los hilos en orden:

2. Observe que la fórmula genérica para cálculo del ID global está en los comentarios. Modifique el programa para que imprima también su nombre y carné. Luego, realice la siguiente modificación al programa (al inicio del main) y use la fórmula genérica para derivar el nuevo cálculo de ID:

```
dim3 g (4,2);
dim3 b (32,16);
hello <<<g, b>>>();
```

FÓRMULA PARA CALCULAR EL ID GLOBAL Y SALIDA DE PANTALLA

CAPTURA DE PANTALLA DE LA NUEVA CONFIGURACIÓN (buscar el mensaje impreso por el hilo con el máximo ID global)

3. Revise nuevamente la información del Compute Capability respecto a las dimensiones máximas de hilos-bloque en x, y, & z para una grilla. Cree una configuración para lanzar exitosamente el kernel para procesar 100,000 datos. (Sugerencia: busque una configuración que lance como mínimo 100,000 hilos. Modifique el kernel para que imprima el mensaje únicamente si es el ID global máximo)

CAPTURA DE PANTALLA CORRIENDO EL CODIGO CON LA NUEVA CONFIGURACIÓN