

## Esenciales de CUDA

Algunas **definiciones**:

- Un **kernel** es una *subrutina* que contiene instrucciones específicas para la GPU.
- Un **CUDA core** es el análogo a un *núcleo* o *procesador* de un CPU.

## Jerarquía de la memoria

La GPU tiene **mallas** con un número definido de **bloques**, y estos a su vez tienen un número fijo de **hilos**.

- Existen **1024** hilos por bloque.
- Los *bloques* y los *hilos* tienen **dimensionalidad 3**:  $x$ ,  $y$  y  $z$ .
- Existen  $2^{31} - 1$  bloques por *mall*a para la dimensión  $x$  y 65535 para las otras dos.<sup>1</sup>

## Uso de kernels

La sintaxis para usar *kernels* es

```
call nombre_kernel<<<numero_bloques, hilos_por_bloque>>>(argumentos, ...)
```

Los *kernels* pueden tener cualquiera de los siguientes **atributos**:

- **host**: Solamente se puede llamar desde el CPU y **no** se ejecuta en GPU.
- **global**: Puede ser llamado en CPU o en GPU, y **si** se ejecuta en GPU.
- **device**: Solamente puede ser llamado en GPU, y **si** se ejecuta en GPU.

---

<sup>1</sup>Tabla de referencia de especificaciones técnicas.