

Ciencia de Datos

- Módulo 4

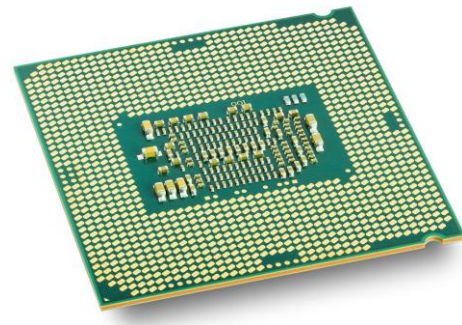
CPU, GPU y TPU



Unidad central de procesamiento

La Unidad central de procesamiento (CPU), constituye el cerebro del dispositivo informático. Su función es procesar secuencias de instrucciones, que obtienen a través de varios equipos periféricos, incluidos dispositivos de entrada / salida y unidades de almacenamiento auxiliares. Estas secuencias de instrucciones son las que realizan los programas instalados en el dispositivo.

En las computadoras modernas, la CPU está contenida en un chip de circuito integrado llamado microprocesador.



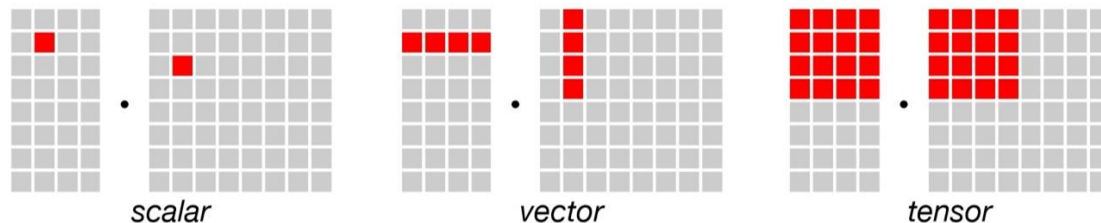
CPU, GPU y TPU

La **Unidad de procesamiento central (CPU)**, la **Unidad de procesamiento de gráficos (GPU)** y la **Unidad de procesamiento de tensores (TPU)** son procesadores con un propósito y una arquitectura especializados.

Siempre y cuando contemos con el compilador correcto, todas son capaces de realizar las mismas instrucciones, pero cada una posee características que los hace más eficientes en la resolución de determinado tipo de problema.



Unidad mínima de procesamiento



La principal diferencia entre las arquitecturas es su unidad mínima de procesamiento.

- CPU: Unidad de datos 1 X 1 (escalar).
Puede manejar decenas de operaciones por ciclo.
- GPU: Unidad de datos 1 X N (vector).
Puede manejar decenas de miles de operaciones por ciclo.
- TPU: Unidad de datos N X N (tensor).
Puede manejar hasta 128000 operaciones por ciclo.



Propósito

CPU: un procesador diseñado para resolver cada problema computacional de manera general. El diseño de memoria caché y memoria está diseñado para ser óptimo para cualquier problema de programación general.

GPU: un procesador diseñado para acelerar la representación de gráficos.

TPU: un coprocesador diseñado para acelerar el desarrollo de tareas de aprendizaje profundo utilizando TensorFlow (un marco de programación); No se han desarrollado compiladores para TPU que podrían usarse para programación de propósito general; por lo tanto, **requiere un esfuerzo significativo para realizar una programación general en TPU**



Uso

CPU: problemas de programación de propósito general.

GPU: representación de gráficos, entrenamiento e inferencia del modelo de aprendizaje automático, problemas de programación con alcance de paralelización, problema de programación de propósito general.

TPU: limitado al entrenamiento e inferencia del modelo de aprendizaje automático (solo en el modelo de TensorFlow).

