

## Questions à Choix Multiples / Multiple Choice Questions

### Question 1:

Quelle est une caractéristique principale des GPU par rapport aux CPU ? / What is a primary feature of GPUs compared to CPUs?

- a) Ils minimisent la latence pour un ou deux threads. / They minimize latency for one or two threads.
- b) Ils maximisent le débit des programmes parallèles. / They maximize throughput of parallel programs.
- c) Ils sont optimisés pour des tâches séquentielles complexes. / They are optimized for complex sequential tasks.
- d) Ils ont moins de cœurs mais sont plus puissants. / They have fewer cores but are more powerful.

Réponse/Answer : b) Ils maximisent le débit des programmes parallèles. / b) They maximize throughput of parallel programs.

### Question 2:

Que classe la taxonomie de Flynn ? / What does Flynn's Taxonomy classify?

- a) Seulement les architectures GPU. / GPU architectures only.
- b) Différents niveaux de parallélisme. / Different levels of parallelism.
- c) Le flux d'instructions et de données dans les architectures. / Instruction and data flow in architectures.
- d) Les métriques de performance en informatique. / Performance metrics in computing.

Réponse/Answer : c) Le flux d'instructions et de données dans les architectures. / c) Instruction and data flow in architectures.

### Question 3:

Lequel N'EST PAS un type de parallélisme dans les applications ? / Which is NOT a type of parallelism in applications?

- a) Parallélisme des tâches / Task parallelism
- b) Parallélisme des données / Data parallelism
- c) Parallélisme des instructions / Instruction parallelism
- d) Parallélisme des threads / Thread parallelism

Réponse/Answer : d) Parallélisme des threads. / d) Thread parallelism.

### Question 4:

Quelle est une métrique clé pour mesurer le débit ? / What is a key metric to measure throughput?

- a) Microsecondes / Microseconds
- b) Gflops / Gflops
- c) Vitesse PCI-Express / PCI-Express speed
- d) Threads par bloc / Threads per block

Réponse/Answer : b) Gflops / b) Gflops.

#### **Question 5:**

Quel est un exemple de parallélisme des données ? / What is an example of data parallelism?

- a) Plusieurs fonctions s'exécutant indépendamment. / Multiple functions running independently.
- b) Un seul élément de données traité par un thread. / A single data item processed by one thread.
- c) De nombreux éléments de données traités par plusieurs threads. / Many data items processed by multiple threads.
- d) Exécution séquentielle d'un programme. / Sequential execution of a program.

Réponse/Answer : c) De nombreux éléments de données traités par plusieurs threads. / c) Many data items processed by multiple threads.

#### **Question 6:**

Quel type d'organisation mémoire utilise une mémoire partagée ? / What type of memory organization uses shared memory?

- a) Multi-nœud / Multi-node
- b) Multiprocesseur / Multiprocessor
- c) Mémoire distribuée / Distributed memory
- d) Monocœur / Single-core

Réponse/Answer : b) Multiprocesseur. / b) Multiprocessor.

#### **Question 7:**

Quel est le rôle de l'hôte dans CUDA ? / What is the role of the host in CUDA?

- a) Exécuter des noyaux parallèles de données. / Execute data-parallel kernels.
- b) Gérer l'environnement GPU. / Manage the GPU environment.
- c) Remplacer le GPU dans les calculs. / Replace the GPU in computations.
- d) Optimiser l'utilisation de la bande passante. / Optimize bandwidth utilization.

Réponse/Answer : b) Gérer l'environnement GPU. / b) Manage the GPU environment.

#### **Question 8:**

Quelle architecture gère un flux d'instructions et plusieurs flux de données ? / Which architecture handles one instruction stream and multiple data streams?

- a) SISD / SISD
- b) SIMD / SIMD
- c) MISD / MISD
- d) MIMD / MIMD

Réponse/Answer : b) SIMD / b) SIMD.

#### **Question 9:**

Quel niveau d'API CUDA offre plus de contrôle mais est plus difficile à programmer ? / What API level in CUDA provides more control but is harder to program?

- a) API pilote / Driver API
- b) API runtime / Runtime API
- c) API hôte / Host API
- d) API périphérique / Device API

Réponse/Answer : a) API pilote. / a) Driver API.

#### **Question 10:**

Quelle est une caractéristique des threads GPU ? / What is a defining characteristic of GPU threads?

- a) Ils sont lourds. / They are heavyweight.
- b) Ils impliquent des commutations de contexte lentes. / They involve slow context switches.
- c) Ils sont légers. / They are lightweight.
- d) Ils privilégient la réduction de la latence. / They prioritize latency reduction.

Réponse/Answer : c) Ils sont légers. / c) They are lightweight.

## Questions à Choix Multiples avec Calculs et Codes en C / Multiple Choice Questions with Calculations and C Codes

### Question 1:

Quelle fonction CUDA est utilisée pour allouer de la mémoire sur l'appareil ? / Which CUDA function is used to allocate memory on the device?

- a) malloc / malloc
- b) cudaMalloc / cudaMalloc
- c) free / free
- d) cudaMemcpy / cudaMemcpy

Réponse/Answer : b) cudaMalloc / b) cudaMalloc

### Question 2:

Quel mot-clé est utilisé pour spécifier une fonction de noyau CUDA ? / Which keyword is used to specify a CUDA kernel function?

- a) \_\_global\_\_ / \_\_global\_\_
- b) \_\_device\_\_ / \_\_device\_\_
- c) \_\_host\_\_ / \_\_host\_\_
- d) \_\_shared\_\_ / \_\_shared\_\_

Réponse/Answer : a) \_\_global\_\_ / a) \_\_global\_\_

### Question 3:

Quelle est la portée de la mémoire partagée dans CUDA ? / What is the scope of shared memory in CUDA?

- a) À l'ensemble du GPU / To the entire GPU
- b) Aux threads individuels uniquement / To individual threads only
- c) Au bloc de threads / To the thread block
- d) Aux grilles de threads / To the thread grids

Réponse/Answer : c) Au bloc de threads / c) To the thread block

### Question 4:

Quelle est la formule pour calculer l'index global dans une matrice 2D CUDA ? / What is the formula to calculate the global index in a 2D CUDA matrix?

- a) ix = threadIdx.x + blockIdx.x \* blockDim.x
- b) iy = threadIdx.y + blockIdx.y \* blockDim.y

- c)  $\text{idx} = \text{iy} * \text{nx} + \text{ix}$
- d) Toutes les réponses ci-dessus / All of the above

Réponse/Answer : d) Toutes les réponses ci-dessus / d) All of the above

### Question 5:

Analysez le code suivant :

```
__global__ void matrixAdd(float *A, float *B, float *C, int nx, int ny) {  
    int ix = threadIdx.x + blockIdx.x * blockDim.x;  
    int iy = threadIdx.y + blockIdx.y * blockDim.y;  
    int idx = iy * nx + ix;  
    if (ix < nx && iy < ny) {  
        C[idx] = A[idx] + B[idx];  
    }  
}
```

Que fait ce code CUDA ? / What does this CUDA code do?

- a) Multiplie deux matrices / Multiplies two matrices
- b) Ajoute deux matrices élément par élément / Adds two matrices element-wise
- c) Effectue une somme vectorielle / Performs a vector sum
- d) Alloue de la mémoire pour une matrice / Allocates memory for a matrix

Réponse/Answer : b) Ajoute deux matrices élément par élément / b) Adds two matrices element-wise

### Question 6:

Quelle est la syntaxe correcte pour invoquer un noyau CUDA avec 32 threads dans un bloc ? /  
What is the correct syntax to invoke a CUDA kernel with 32 threads in a block?

- a) kernel\_name<<<1, 32>>>()
- b) kernel\_name<<<32, 1>>>()
- c) kernel\_name<<<4, 8>>>()
- d) kernel\_name<<<32, 32>>>()

Réponse/Answer : a) kernel\_name<<<1, 32>>>() / a) kernel\_name<<<1, 32>>>()

### Question 7:

Quelle fonction CUDA est utilisée pour synchroniser les threads dans un bloc ? / Which CUDA function is used to synchronize threads in a block?

- a) cudaMemcpy / cudaMemcpy
- b) \_\_syncthreads() / \_\_syncthreads()
- c) atomicAdd / atomicAdd
- d) cudaMalloc / cudaMalloc

Réponse/Answer : b) `_syncthreads()` / b) `_syncthreads()`

#### Question 8:

Quelle est la différence entre `gridDim` et `blockDim` ? / What is the difference between `gridDim` and `blockDim`?

- a) `gridDim` représente les dimensions des blocs dans une grille / `gridDim` represents the dimensions of blocks in a grid
- b) `blockDim` représente les dimensions des threads dans un bloc / `blockDim` represents the dimensions of threads in a block
- c) Les deux / Both
- d) Aucune des réponses ci-dessus / None of the above

Réponse/Answer : c) Les deux / c) Both

#### Question 9:

Quel est le rôle de `atomicAdd` en CUDA ? / What is the role of `atomicAdd` in CUDA?

- a) Ajouter deux variables de manière atomique / Add two variables atomically
- b) Synchroniser les threads / Synchronize threads
- c) Gérer les collisions dans la mémoire globale / Handle collisions in global memory
- d) Les réponses a et c / Both a and c

Réponse/Answer : d) Les réponses a et c / d) Both a and c

## 20 Questions - CUDA and GPU Architectures

### Question 1

Quel est le rôle principal du Streaming Multiprocessor (SM) dans une architecture GPU ? /  
What is the primary role of the Streaming Multiprocessor (SM) in a GPU architecture?

- a) Gérer les opérations de transfert de mémoire / Handle memory transfer operations
- b) Exécuter les threads en parallèle / Execute threads in parallel
- c) Gérer les communications CPU-GPU / Manage CPU-GPU communications
- d) Optimiser les instructions SIMD / Optimize SIMD instructions

Réponse / Answer: b) Exécuter les threads en parallèle / Execute threads in parallel

### Question 2

Combien de threads composent un warp dans CUDA ? / How many threads are in a warp in CUDA?

- a) 16
- b) 32
- c) 64
- d) 128

Réponse / Answer: b) 32

### Question 3

Quelle est la différence clé entre SIMD et SIMT ? / What is the key difference between SIMD and SIMT?

- a) SIMD exécute des instructions différentes pour chaque thread / SIMD executes different instructions for each thread
- b) SIMT permet aux threads d'avoir des chemins d'exécution indépendants / SIMT allows threads to have independent execution paths
- c) SIMD et SIMT sont identiques / SIMD and SIMT are identical
- d) SIMD gère plusieurs threads dans un seul registre / SIMD handles multiple threads in a single register

Réponse / Answer: b) SIMT permet aux threads d'avoir des chemins d'exécution indépendants / SIMT allows threads to have independent execution paths

#### **Question 4**

Dans CUDA, quelle commande est utilisée pour synchroniser tous les threads d'un bloc ? /  
In CUDA, which command is used to synchronize all threads in a block?

- a) \_\_global\_\_
- b) \_\_syncthreads()
- c) cudaMemcpy
- d) cudaDeviceSynchronize()

Réponse / Answer: b) \_\_syncthreads()

#### **Question 5**

Que signifie le terme 'warp divergence' ? / What does the term 'warp divergence' mean?

- a) Une divergence dans les données partagées entre threads / A divergence in shared data between threads
- b) Des threads d'un même warp suivant des chemins d'exécution différents / Threads in the same warp following different execution paths
- c) Des threads s'exécutant sur plusieurs SM en parallèle / Threads executing on multiple SMs in parallel
- d) Une interruption dans le flux d'instructions d'un thread / An interruption in the instruction flow of a thread

Réponse / Answer: b) Des threads d'un même warp suivant des chemins d'exécution différents / Threads in the same warp following different execution paths

#### **Question 1**

Quel est le rôle principal du Streaming Multiprocessor (SM) dans une architecture GPU ? /  
What is the primary role of the Streaming Multiprocessor (SM) in a GPU architecture?

- a) Gérer les opérations de transfert de mémoire / Handle memory transfer operations
- b) Exécuter les threads en parallèle / Execute threads in parallel
- c) Gérer les communications CPU-GPU / Manage CPU-GPU communications
- d) Optimiser les instructions SIMD / Optimize SIMD instructions

Réponse / Answer: b) Exécuter les threads en parallèle / Execute threads in parallel

#### **Question 2**

Combien de threads composent un warp dans CUDA ? / How many threads are in a warp in CUDA?

- a) 16
- b) 32
- c) 64
- d) 128

Réponse / Answer: b) 32

### Question 3

Quelle est la différence clé entre SIMD et SIMT ? / What is the key difference between SIMD and SIMT?

- a) SIMD exécute des instructions différentes pour chaque thread / SIMD executes different instructions for each thread
- b) SIMT permet aux threads d'avoir des chemins d'exécution indépendants / SIMT allows threads to have independent execution paths
- c) SIMD et SIMT sont identiques / SIMD and SIMT are identical
- d) SIMD gère plusieurs threads dans un seul registre / SIMD handles multiple threads in a single register

Réponse / Answer: b) SIMT permet aux threads d'avoir des chemins d'exécution indépendants / SIMT allows threads to have independent execution paths

### Question 4

Dans CUDA, quelle commande est utilisée pour synchroniser tous les threads d'un bloc ? / In CUDA, which command is used to synchronize all threads in a block?

- a) \_\_global\_\_
- b) \_\_syncthreads()
- c) cudaMemcpy
- d) cudaDeviceSynchronize()

Réponse / Answer: b) \_\_syncthreads()

### Question 5

Que signifie le terme 'warp divergence' ? / What does the term 'warp divergence' mean?

- a) Une divergence dans les données partagées entre threads / A divergence in shared data between threads
- b) Des threads d'un même warp suivant des chemins d'exécution différents / Threads in the same warp following different execution paths

- c) Des threads s'exécutant sur plusieurs SM en parallèle / Threads executing on multiple SMs in parallel
- d) Une interruption dans le flux d'instructions d'un thread / An interruption in the instruction flow of a thread

Réponse / Answer: b) Des threads d'un même warp suivant des chemins d'exécution différents / Threads in the same warp following different execution paths

### Question 6

Quelle architecture GPU a introduit les Tensor Cores ? / Which GPU architecture introduced Tensor Cores?

- a) Fermi
- b) Kepler
- c) Volta
- d) Pascal

Réponse / Answer: c) Volta

### Question 7

Combien de cycles d'horloge en moyenne faut-il pour accéder à la mémoire globale dans CUDA ? / How many clock cycles on average are needed to access global memory in CUDA?

- a) 10-20
- b) 50-100
- c) 400-800
- d) 1000-2000

Réponse / Answer: c) 400-800

### Question 8

Quel outil de profilage est utilisé pour analyser les performances des kernels CUDA ? / Which profiling tool is used to analyze CUDA kernel performance?

- a) nvcc
- b) cudaMemcpy
- c) nvprof
- d) gprof

Réponse / Answer: c) nvprof

### **Question 9**

Quelle unité exécute les instructions pour les threads dans un warp ? / Which unit executes instructions for threads in a warp?

- a) CUDA Cores
- b) Tensor Cores
- c) Warp Scheduler
- d) Shared Memory

Réponse / Answer: a) CUDA Cores

### **Question 10**

Quel est le rôle principal de la mémoire partagée (shared memory) dans CUDA ? / What is the primary role of shared memory in CUDA?

- a) Stocker les données pour un accès rapide entre threads d'un même bloc / Store data for fast access among threads in the same block
- b) Synchroniser les threads entre différents blocs / Synchronize threads across different blocks
- c) Charger les données de la mémoire globale pour l'utiliser plus tard / Load data from global memory for later use
- d) Remplacer les registres en cas de dépassement / Replace registers in case of overflow

Réponse / Answer: a) Stocker les données pour un accès rapide entre threads d'un même bloc / Store data for fast access among threads in the same block

### **Question 11**

Quelle architecture CUDA a introduit le parallélisme dynamique ? / Which CUDA architecture introduced dynamic parallelism?

- a) Fermi
- b) Kepler
- c) Volta
- d) Ampere

Réponse / Answer: b) Kepler

### **Question 12**

Quelle commande CUDA est utilisée pour allouer de la mémoire sur le GPU ? / Which CUDA command is used to allocate memory on the GPU?

- a) cudaMalloc
- b) cudaMemcpy
- c) cudaFree
- d) cudaSync

Réponse / Answer: a) cudaMalloc

### **Question 13**

Que signifie SIMT dans le contexte de CUDA ? / What does SIMT mean in the context of CUDA?

- a) Single Instruction Multiple Threads
- b) Single Instruction Multiple Tasks
- c) Scalable Independent Multiple Threads
- d) Synchronized Independent Multiple Tasks

Réponse / Answer: a) Single Instruction Multiple Threads

### **Question 14**

Quelle est la latence typique de la mémoire partagée par rapport à la mémoire globale ? / What is the typical latency of shared memory compared to global memory?

- a) Plus rapide de 10 à 20x / 10 to 20x faster
- b) Plus rapide de 100 à 150x / 100 to 150x faster
- c) Plus lente de 2x / 2x slower
- d) Identique / Identical

Réponse / Answer: a) Plus rapide de 10 à 20x / 10 to 20x faster

### **Question 15**

Dans CUDA, que fait la fonction cudaMemcpy ? / What does the cudaMemcpy function do in CUDA?

- a) Alloue de la mémoire sur le GPU / Allocates memory on the GPU
- b) Libère la mémoire sur le GPU / Frees memory on the GPU

- c) Copie les données entre l'hôte et le périphérique / Copies data between host and device
- d) Synchronise les threads sur le GPU / Synchronizes threads on the GPU

Réponse / Answer: c) Copie les données entre l'hôte et le périphérique / Copies data between host and device

### Question 16

Comment les threads sont-ils regroupés dans CUDA ? / How are threads grouped in CUDA?

- a) En blocs et grilles / In blocks and grids
- b) En matrices et vecteurs / In matrices and vectors
- c) En pages et clusters / In pages and clusters
- d) En bancs et warps / In banks and warps

Réponse / Answer: a) En blocs et grilles / In blocks and grids

### Question 17

Quel est le rôle principal de la mémoire constante dans CUDA ? / What is the primary role of constant memory in CUDA?

- a) Stocker les instructions du programme / Store program instructions
- b) Fournir un accès rapide à des données immuables pour tous les threads / Provide fast access to immutable data for all threads
- c) Synchroniser les threads / Synchronize threads
- d) Remplacer la mémoire partagée / Replace shared memory

Réponse / Answer: b) Fournir un accès rapide à des données immuables pour tous les threads / Provide fast access to immutable data for all threads

### Question 18

Quelle est la taille maximale d'un warp dans CUDA ? / What is the maximum size of a warp in CUDA?

- a) 16
- b) 32
- c) 64
- d) 128

Réponse / Answer: b) 32

### **Question 19**

Comment un kernel CUDA est-il défini ? / How is a CUDA kernel defined?

- a) Avec `_device_`
- b) Avec `_global_`
- c) Avec `_shared_`
- d) Avec `_host_`

Réponse / Answer: b) Avec `_global_`

### **Question 20**

Qu'est-ce qu'un warp scheduler ? / What is a warp scheduler?

- a) Une unité qui exécute plusieurs threads en parallèle / A unit that executes multiple threads in parallel
- b) Une unité qui ordonne l'exécution des warps / A unit that schedules the execution of warps
- c) Une unité qui gère la mémoire partagée / A unit that manages shared memory
- d) Une unité qui synchronise les threads / A unit that synchronizes threads

Réponse / Answer: b) Une unité qui ordonne l'exécution des warps / A unit that schedules the execution of warps

### **Question 21**

Dans CUDA, que signifie le terme 'global memory coalescing' ? / What does 'global memory coalescing' mean in CUDA?

- a) Une réduction de la latence des threads / A reduction in thread latency
- b) L'accès simultané efficace à la mémoire globale / Efficient simultaneous access to global memory
- c) La synchronisation des threads dans un warp / Thread synchronization within a warp
- d) L'optimisation des instructions dans un warp / Optimization of instructions in a warp

Réponse / Answer: b) L'accès simultané efficace à la mémoire globale / Efficient simultaneous access to global memory

## CUDA Kernel Execution Configuration Questions / Questions de Configuration d'Exécution des Noyaux CUDA

### Question 1:

Given a matrix of size  $nx = 1024$  and  $ny = 1024$ , and a 2D grid with block dimensions  $dimx = 16$ ,  $dimy = 16$ :

1. What will be the grid dimensions  $grid.x$  and  $grid.y$ ?
2. How many total threads will be launched in the kernel?

\*\*Answer / Réponse:\*\*

$grid.x = 64$ ,  $grid.y = 64$ , Total threads = 1048576

\*\*Explanation / Explication:\*\*

The grid dimensions are calculated as:

$grid.x = ceil(nx / dimx) = ceil(1024 / 16) = 64$

$grid.y = ceil(ny / dimy) = ceil(1024 / 16) = 64$

Total threads =  $grid.x * grid.y * block.x * block.y = 64 * 64 * 16 * 16 = 1048576$ .

### Question 2:

Given a matrix of size  $nx = 2048$  and  $ny = 2048$ , and a 2D grid with block dimensions  $dimx = 32$ ,  $dimy = 32$ :

1. What will be the grid dimensions  $grid.x$  and  $grid.y$ ?
2. How many total threads will be launched in the kernel?

\*\*Answer / Réponse:\*\*

$grid.x = 64$ ,  $grid.y = 64$ , Total threads = 4194304

\*\*Explanation / Explication:\*\*

The grid dimensions are calculated as:

$grid.x = ceil(nx / dimx) = ceil(2048 / 32) = 64$

$grid.y = ceil(ny / dimy) = ceil(2048 / 32) = 64$

Total threads =  $grid.x * grid.y * block.x * block.y = 64 * 64 * 32 * 32 = 4194304$ .

### Question 3:

Given a matrix of size  $nx = 8192$  and  $ny = 8192$ , and a 2D grid with block dimensions  $dimx = 64$ ,  $dimy = 64$ :

1. What will be the grid dimensions  $grid.x$  and  $grid.y$ ?
2. How many total threads will be launched in the kernel?

\*\*Answer / Réponse:\*\*

$grid.x = 128$ ,  $grid.y = 128$ , Total threads = 67108864

\*\*Explanation / Explication:\*\*

The grid dimensions are calculated as:

$$\begin{aligned}\text{grid.x} &= \text{ceil}(nx / \text{dimx}) = \text{ceil}(8192 / 64) = 128 \\ \text{grid.y} &= \text{ceil}(ny / \text{dimy}) = \text{ceil}(8192 / 64) = 128\end{aligned}$$

Total threads =  $\text{grid.x} * \text{grid.y} * \text{block.x} * \text{block.y} = 128 * 128 * 64 * 64 = 67108864$ .

#### Question 4:

Given a matrix of size  $nx = 16384$  and  $ny = 8192$ , and a 2D grid with block dimensions  $\text{dimx} = 32$ ,  $\text{dimy} = 32$ :

1. What will be the grid dimensions  $\text{grid.x}$  and  $\text{grid.y}$ ?
2. How many total threads will be launched in the kernel?

\*\*Answer / Réponse:\*\*

$$\text{grid.x} = 512, \text{grid.y} = 256, \text{Total threads} = 134217728$$

\*\*Explanation / Explication:\*\*

The grid dimensions are calculated as:

$$\begin{aligned}\text{grid.x} &= \text{ceil}(nx / \text{dimx}) = \text{ceil}(16384 / 32) = 512 \\ \text{grid.y} &= \text{ceil}(ny / \text{dimy}) = \text{ceil}(8192 / 32) = 256\end{aligned}$$

Total threads =  $\text{grid.x} * \text{grid.y} * \text{block.x} * \text{block.y} = 512 * 256 * 32 * 32 = 134217728$ .

#### Question 5:

Given a matrix of size  $nx = 512$  and  $ny = 512$ , and a 2D grid with block dimensions  $\text{dimx} = 8$ ,  $\text{dimy} = 8$ :

1. What will be the grid dimensions  $\text{grid.x}$  and  $\text{grid.y}$ ?
2. How many total threads will be launched in the kernel?

\*\*Answer / Réponse:\*\*

$$\text{grid.x} = 64, \text{grid.y} = 64, \text{Total threads} = 262144$$

\*\*Explanation / Explication:\*\*

The grid dimensions are calculated as:

$$\begin{aligned}\text{grid.x} &= \text{ceil}(nx / \text{dimx}) = \text{ceil}(512 / 8) = 64 \\ \text{grid.y} &= \text{ceil}(ny / \text{dimy}) = \text{ceil}(512 / 8) = 64\end{aligned}$$

Total threads =  $\text{grid.x} * \text{grid.y} * \text{block.x} * \text{block.y} = 64 * 64 * 8 * 8 = 262144$ .

## Questions à Choix Multiples / Multiple Choice Questions

### Question 1:

Quelle sera la valeur du tableau a[] après l'exécution du noyau suivant avec configuration <<<3, 4>>> ? / What will be the value of array a[] after executing the following kernel with configuration <<<3, 4>>>?

```
__global__ void kernel(int *a) {
    int i = threadIdx.x + blockIdx.x * blockDim.x;
    a[i] = blockDim.x;
}
```

\*\*Options / Choix:\*\*

- a) [4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4]
- b) [0, 1, 2, 3, 0, 1, 2, 3, 0, 1, 2, 3]
- c) [0, 0, 0, 0, 1, 1, 1, 2, 2, 2, 2]
- d) [0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11]

\*\*Réponse / Answer:\*\* a) [4, 4, 4, 4, 4, 4, 4, 4, 4, 4]

### Question 2:

Quelle sera la valeur du tableau a[] après l'exécution du noyau suivant avec configuration <<<3, 4>>> ? / What will be the value of array a[] after executing the following kernel with configuration <<<3, 4>>>?

```
__global__ void kernel(int *a) {
    int i = threadIdx.x + blockIdx.x * blockDim.x;
    a[i] = threadIdx.x;
}
```

\*\*Options / Choix:\*\*

- a) [4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4]
- b) [0, 1, 2, 3, 0, 1, 2, 3, 0, 1, 2, 3]
- c) [0, 0, 0, 0, 1, 1, 1, 2, 2, 2, 2]

d) [0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11]

\*\*Réponse / Answer:\*\* b) [0, 1, 2, 3, 0, 1, 2, 3, 0, 1, 2, 3]

### Question 3:

Quelle sera la valeur du tableau a[] après l'exécution du noyau suivant avec configuration <<<3, 4>>> ? / What will be the value of array a[] after executing the following kernel with configuration <<<3, 4>>>?

```
__global__ void kernel(int *a) {
    int i = threadIdx.x + blockIdx.x * blockDim.x;
    a[i] = blockIdx.x;
}
```

\*\*Options / Choix:\*\*

a) [4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4]

b) [0, 1, 2, 3, 0, 1, 2, 3, 0, 1, 2, 3]

c) [0, 0, 0, 0, 1, 1, 1, 2, 2, 2, 2]

d) [0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11]

\*\*Réponse / Answer:\*\* c) [0, 0, 0, 0, 1, 1, 1, 2, 2, 2]

### Question 4:

Quelle sera la valeur du tableau a[] après l'exécution du noyau suivant avec configuration <<<3, 4>>> ? / What will be the value of array a[] after executing the following kernel with configuration <<<3, 4>>>?

```
__global__ void kernel(int *a) {
    int i = threadIdx.x + blockIdx.x * blockDim.x;
    a[i] = i;
}
```

\*\*Options / Choix:\*\*

a) [4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4]

b) [0, 1, 2, 3, 0, 1, 2, 3, 0, 1, 2, 3]

c) [0, 0, 0, 0, 1, 1, 1, 2, 2, 2, 2]

d) [0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11]

**\*\*Réponse / Answer:\*\*** d) [0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11]

## Questions à Choix Multiples / Multiple Choice Questions

### Question 1:

Une matrice de taille  $nx = 1024$  et  $ny = 1024$  utilise une grille 2D avec des blocs de dimensions  $dimx = 16$ ,  $dimy = 16$ . /

A matrix of size  $nx = 1024$  and  $ny = 1024$  uses a 2D grid with block dimensions  $dimx = 16$ ,  $dimy = 16$ :

1. Quelles sont les dimensions de la grille (grid.x, grid.y) ? / What are the grid dimensions (grid.x, grid.y)?

2. Combien de threads au total sont lancés ? / How many total threads are launched?

\*\*Options / Choix:\*\*

a) grid.x = 64, grid.y = 64, Total threads = 65536

b) grid.x = 32, grid.y = 32, Total threads = 32768

c) grid.x = 16, grid.y = 16, Total threads = 8192

d) grid.x = 128, grid.y = 128, Total threads = 16384

\*\*Réponse / Answer:\*\* a) grid.x = 64, grid.y = 64, Total threads = 65536

### Question 2:

Une matrice de taille  $nx = 2048$  et  $ny = 1024$  utilise une grille 2D avec des blocs de dimensions  $dimx = 32$ ,  $dimy = 16$ . /

A matrix of size  $nx = 2048$  and  $ny = 1024$  uses a 2D grid with block dimensions  $dimx = 32$ ,  $dimy = 16$ :

1. Quelles sont les dimensions de la grille (grid.x, grid.y) ? / What are the grid dimensions (grid.x, grid.y)?

2. Combien de threads au total sont lancés ? / How many total threads are launched?

\*\*Options / Choix:\*\*

a) grid.x = 64, grid.y = 32, Total threads = 65536

b) grid.x = 32, grid.y = 16, Total threads = 32768

c) grid.x = 64, grid.y = 64, Total threads = 131072

d) grid.x = 128, grid.y = 64, Total threads = 262144

\*\*Réponse / Answer:\*\* c) grid.x = 64, grid.y = 64, Total threads = 131072

### Question 3:

Une matrice de taille nx = 8192 et ny = 8192 utilise une grille 2D avec des blocs de dimensions dimx = 64, dimy = 64. /

A matrix of size nx = 8192 and ny = 8192 uses a 2D grid with block dimensions dimx = 64, dimy = 64:

1. Quelles sont les dimensions de la grille (grid.x, grid.y) ? / What are the grid dimensions (grid.x, grid.y)?

2. Combien de threads au total sont lancés ? / How many total threads are launched?

\*\*Options / Choix:\*\*

a) grid.x = 128, grid.y = 128, Total threads = 134217728

b) grid.x = 64, grid.y = 64, Total threads = 16777216

c) grid.x = 256, grid.y = 256, Total threads = 67108864

d) grid.x = 512, grid.y = 512, Total threads = 268435456

\*\*Réponse / Answer:\*\* a) grid.x = 128, grid.y = 128, Total threads = 134217728

### Question 4:

Une matrice de taille nx = 16384 et ny = 16384 utilise une grille 2D avec des blocs de dimensions dimx = 32, dimy = 32. /

A matrix of size nx = 16384 and ny = 16384 uses a 2D grid with block dimensions dimx = 32, dimy = 32:

1. Quelles sont les dimensions de la grille (grid.x, grid.y) ? / What are the grid dimensions (grid.x, grid.y)?

2. Combien de threads au total sont lancés ? / How many total threads are launched?

\*\*Options / Choix:\*\*

a) grid.x = 512, grid.y = 512, Total threads = 268435456

b) grid.x = 256, grid.y = 256, Total threads = 67108864

c) grid.x = 128, grid.y = 128, Total threads = 16777216

d) grid.x = 64, grid.y = 64, Total threads = 4194304

\*\*Réponse / Answer:\*\* a) grid.x = 512, grid.y = 512, Total threads = 268435456

## **Questions à Choix Multiples : Mémoire CUDA / Multiple-Choice Questions: CUDA Memory**

### **Question 1:**

En français : Quelle est la mémoire la plus rapide dans un GPU ? / What is the fastest memory in a GPU? / In English: Quelle est la mémoire la plus rapide dans un GPU ? / What is the fastest memory in a GPU?

- a) a) Mémoire partagée / Shared Memory
- b) b) Registres / Registers
- c) c) Mémoire constante / Constant Memory
- d) d) Mémoire globale / Global Memory

### **Réponse / Answer:**

- b) Registres / Registers

### **Question 2:**

En français : Quelle mémoire est optimisée pour une localité spatiale en 2D ? / Which memory is optimized for 2D spatial locality? / In English: Quelle mémoire est optimisée pour une localité spatiale en 2D ? / Which memory is optimized for 2D spatial locality?

- a) a) Mémoire partagée / Shared Memory
- b) b) Mémoire constante / Constant Memory
- c) c) Mémoire texture / Texture Memory
- d) d) Mémoire globale / Global Memory

### **Réponse / Answer:**

- c) Mémoire texture / Texture Memory

### **Question 3:**

En français : Quel type de mémoire est utilisé par tous les threads d'un bloc ? / Which type of memory is used by all threads in a block? / In English: Quel type de mémoire est utilisé par tous les threads d'un bloc ? / Which type of memory is used by all threads in a block?

- a) a) Mémoire locale / Local Memory
- b) b) Mémoire partagée / Shared Memory
- c) c) Mémoire constante / Constant Memory

d) d) Registres / Registers

**Réponse / Answer:**

b) Mémoire partagée / Shared Memory

**Question 4:**

En français : Quel est le principal avantage de la mémoire unifiée ? / What is the main advantage of Unified Memory? / In English: Quel est le principal avantage de la mémoire unifiée ? / What is the main advantage of Unified Memory?

a) a) Accès plus rapide / Faster access

b) b) Migration automatique des données / Automatic data migration

c) c) Plus grande capacité / Larger capacity

d) d) Moins coûteux / Less expensive

**Réponse / Answer:**

b) Migration automatique des données / Automatic data migration

**Question 5:**

En français : Qu'est-ce que l'accès mémoire coalescé ? / What is coalesced memory access? / In English: Qu'est-ce que l'accès mémoire coalescé ? / What is coalesced memory access?

a) a) Lorsque tous les threads accèdent à la même adresse / When all threads access the same address

b) b) Lorsque les threads accèdent à des adresses alignées et contiguës / When threads access aligned and contiguous addresses

c) c) Lorsque les données sont stockées dans la mémoire partagée / When data is stored in shared memory

d) d) Lorsque les données sont copiées entre l'hôte et le dispositif / When data is copied between host and device

**Réponse / Answer:**

b) Lorsque les threads accèdent à des adresses alignées et contiguës / When threads access aligned and contiguous addresses

**Question 6:**

En français : Qu'est-ce que le spilling des registres ? / What is register spilling? / In English: Qu'est-ce que le spilling des registres ? / What is register spilling?

a) a) Transfert de données vers une mémoire plus rapide / Data transfer to faster memory

b) b) Transfert de données des registres vers la mémoire locale / Data transfer from registers to local memory

c) c) Déclaration incorrecte de variables / Incorrect variable declaration

d) d) Dépassemement de capacité de la mémoire globale / Global memory overflow

**Réponse / Answer:**

b) Transfert de données des registres vers la mémoire locale / Data transfer from registers to local memory

**Question 1:**

En français : Quelle est la mémoire la plus rapide dans un GPU ? / What is the fastest memory in a GPU? / In English: Quelle est la mémoire la plus rapide dans un GPU ? / What is the fastest memory in a GPU?

a) a) Mémoire partagée / Shared Memory

b) b) Registres / Registers

c) c) Mémoire constante / Constant Memory

d) d) Mémoire globale / Global Memory

**Réponse / Answer:**

b) Registres / Registers

**Question 2:**

En français : Quelle mémoire est optimisée pour une localité spatiale en 2D ? / Which memory is optimized for 2D spatial locality? / In English: Quelle mémoire est optimisée pour une localité spatiale en 2D ? / Which memory is optimized for 2D spatial locality?

a) a) Mémoire partagée / Shared Memory

b) b) Mémoire constante / Constant Memory

c) c) Mémoire texture / Texture Memory

d) d) Mémoire globale / Global Memory

**Réponse / Answer:**

c) Mémoire texture / Texture Memory

**Question 3:**

En français : Quel type de mémoire est utilisé par tous les threads d'un bloc ? / Which type of memory is used by all threads in a block? / In English: Quel type de mémoire est utilisé par tous les threads d'un bloc ? / Which type of memory is used by all threads in a block?

- a) a) Mémoire locale / Local Memory
- b) b) Mémoire partagée / Shared Memory
- c) c) Mémoire constante / Constant Memory
- d) d) Registres / Registers

**Réponse / Answer:**

b) Mémoire partagée / Shared Memory

**Question 4:**

En français : Quel est le principal avantage de la mémoire unifiée ? / What is the main advantage of Unified Memory? / In English: Quel est le principal avantage de la mémoire unifiée ? / What is the main advantage of Unified Memory?

- a) a) Accès plus rapide / Faster access
- b) b) Migration automatique des données / Automatic data migration
- c) c) Plus grande capacité / Larger capacity
- d) d) Moins coûteux / Less expensive

**Réponse / Answer:**

b) Migration automatique des données / Automatic data migration

**Question 5:**

En français : Qu'est-ce que l'accès mémoire coalescé ? / What is coalesced memory access? / In English: Qu'est-ce que l'accès mémoire coalescé ? / What is coalesced memory access?

- a) a) Lorsque tous les threads accèdent à la même adresse / When all threads access the same address
- b) b) Lorsque les threads accèdent à des adresses alignées et contiguës / When threads access aligned and contiguous addresses
- c) c) Lorsque les données sont stockées dans la mémoire partagée / When data is stored in shared memory
- d) d) Lorsque les données sont copiées entre l'hôte et le dispositif / When data is copied between host and device

**Réponse / Answer:**

b) Lorsque les threads accèdent à des adresses alignées et contiguës / When threads access aligned and contiguous addresses

### **Question 6:**

En français : Qu'est-ce que le spilling des registres ? / What is register spilling? / In English: Qu'est-ce que le spilling des registres ? / What is register spilling?

- a) a) Transfert de données vers une mémoire plus rapide / Data transfer to faster memory
- b) b) Transfert de données des registres vers la mémoire locale / Data transfer from registers to local memory
- c) c) Déclaration incorrecte de variables / Incorrect variable declaration
- d) d) Dépassement de capacité de la mémoire globale / Global memory overflow

### **Réponse / Answer:**

b) Transfert de données des registres vers la mémoire locale / Data transfer from registers to local memory

### **Question 7:**

En français : Quelle est la taille d'un segment de cache L1 ? / What is the size of an L1 cache segment? / In English: Quelle est la taille d'un segment de cache L1 ? / What is the size of an L1 cache segment?

- a) a) 32 octets / 32 bytes
- b) b) 64 octets / 64 bytes
- c) c) 128 octets / 128 bytes
- d) d) 256 octets / 256 bytes

### **Réponse / Answer:**

c) 128 octets / 128 bytes

### **Question 8:**

En français : Quelle mémoire est idéale pour les données lues par tous les threads d'un warp ? / Which memory is ideal for data read by all threads in a warp? / In English: Quelle mémoire est idéale pour les données lues par tous les threads d'un warp ? / Which memory is ideal for data read by all threads in a warp?

- a) a) Mémoire globale / Global Memory
- b) b) Mémoire locale / Local Memory
- c) c) Mémoire constante / Constant Memory
- d) d) Mémoire partagée / Shared Memory

**Réponse / Answer:**

c) Mémoire constante / Constant Memory

**Question 9:**

En français : Quelle fonction CUDA est utilisée pour transférer des données entre l'hôte et le dispositif ? / Which CUDA function is used to transfer data between host and device? / In English: Quelle fonction CUDA est utilisée pour transférer des données entre l'hôte et le dispositif ? / Which CUDA function is used to transfer data between host and device?

- a) a) cudaMalloc / cudaMalloc
- b) b) cudaMemcpy / cudaMemcpy
- c) c) cudaDeviceSynchronize / cudaDeviceSynchronize
- d) d) cudaFree / cudaFree

**Réponse / Answer:**

b) cudaMemcpy / cudaMemcpy

**Question 10:**

En français : Quel est le principal défi de la mémoire globale ? / What is the main challenge of global memory? / In English: Quel est le principal défi de la mémoire globale ? / What is the main challenge of global memory?

- a) a) Sa lenteur d'accès / Its slow access speed
- b) b) Sa petite taille / Its small size
- c) c) Son coût élevé / Its high cost
- d) d) Sa portée limitée / Its limited scope

**Réponse / Answer:**

a) Sa lenteur d'accès / Its slow access speed

**Question 11:**

En français : Quelle mémoire est automatiquement migrée entre l'hôte et le dispositif dans CUDA 6.0+ ? / Which memory is automatically migrated between host and device in CUDA 6.0+? / In English: Quelle mémoire est automatiquement migrée entre l'hôte et le dispositif dans CUDA 6.0+ ? / Which memory is automatically migrated between host and device in CUDA 6.0+?

- a) a) Mémoire locale / Local Memory
- b) b) Mémoire unifiée / Unified Memory
- c) c) Mémoire partagée / Shared Memory

d) d) Mémoire constante / Constant Memory

**Réponse / Answer:**

b) Mémoire unifiée / Unified Memory

**Question 12:**

En français : Que signifie l'accès à la mémoire non coalescé ? / What does uncoalesced memory access mean? / In English: Que signifie l'accès à la mémoire non coalescé ? / What does uncoalesced memory access mean?

- a) a) Les threads accèdent à des adresses dispersées / Threads access scattered addresses
- b) b) Les threads accèdent à des adresses contiguës / Threads access contiguous addresses
- c) c) Les threads partagent les mêmes adresses / Threads share the same addresses
- d) d) Les threads utilisent des registres / Threads use registers

**Réponse / Answer:**

a) Les threads accèdent à des adresses dispersées / Threads access scattered addresses

**Question 13:**

En français : Quelle est la capacité typique de la mémoire partagée ? / What is the typical capacity of shared memory? / In English: Quelle est la capacité typique de la mémoire partagée ? / What is the typical capacity of shared memory?

- a) a) Quelques kilo-octets / A few kilobytes
- b) b) Plusieurs méga-octets / Several megabytes
- c) c) Plusieurs giga-octets / Several gigabytes
- d) d) Quelques tera-octets / A few terabytes

**Réponse / Answer:**

a) Quelques kilo-octets / A few kilobytes

**Question 14:**

En français : Pourquoi la mémoire unifiée simplifie-t-elle la programmation CUDA ? / Why does Unified Memory simplify CUDA programming? / In English: Pourquoi la mémoire unifiée simplifie-t-elle la programmation CUDA ? / Why does Unified Memory simplify CUDA programming?

- a) a) Elle augmente la vitesse d'accès / It increases access speed
- b) b) Elle réduit la gestion des pointeurs / It reduces pointer management
- c) c) Elle réduit la latence mémoire / It reduces memory latency

d) d) Elle améliore les performances GPU / It improves GPU performance

**Réponse / Answer:**

b) Elle réduit la gestion des pointeurs / It reduces pointer management

**Question 15:**

En français : Qu'est-ce que le camping de partitions ? / What is partition camping? / In English: Qu'est-ce que le camping de partitions ? / What is partition camping?

- a) a) Une utilisation déséquilibrée des partitions mémoire / An uneven usage of memory partitions
- b) b) Un alignement mémoire incorrect / An incorrect memory alignment
- c) c) Un dépassement des limites de la mémoire locale / An overflow of local memory
- d) d) Une surcharge des registres / An overload of registers

**Réponse / Answer:**

a) Une utilisation déséquilibrée des partitions mémoire / An uneven usage of memory partitions

**Question 16:**

En français : Quel est le principal objectif de l'accès mémoire coalescé ? / What is the main goal of coalesced memory access? / In English: Quel est le principal objectif de l'accès mémoire coalescé ? / What is the main goal of coalesced memory access?

- a) a) Réduire les transactions mémoire / To reduce memory transactions
- b) b) Augmenter la latence mémoire / To increase memory latency
- c) c) Réduire la capacité mémoire / To reduce memory capacity
- d) d) Augmenter la taille de la mémoire locale / To increase local memory size

**Réponse / Answer:**

a) Réduire les transactions mémoire / To reduce memory transactions

**Question 17:**

En français : Quels types de mémoire sont cachés dans CUDA ? / Which memory types are cached in CUDA? / In English: Quels types de mémoire sont cachés dans CUDA ? / Which memory types are cached in CUDA?

- a) a) Mémoire globale et locale / Global and Local Memory
- b) b) Mémoire partagée et constante / Shared and Constant Memory
- c) c) Mémoire locale et constante / Local and Constant Memory

d) d) Mémoire globale et constante / Global and Constant Memory

**Réponse / Answer:**

d) Mémoire globale et constante / Global and Constant Memory

**Question 18:**

En français : Qu'est-ce que la mémoire zero-copy ? / What is zero-copy memory? / In English: Qu'est-ce que la mémoire zero-copy ? / What is zero-copy memory?

- a) a) Mémoire qui ne nécessite pas de transfert entre l'hôte et le GPU / Memory that requires no transfer between host and GPU
- b) b) Mémoire qui n'a pas besoin de synchronisation / Memory that needs no synchronization
- c) c) Mémoire à faible latence / Low-latency memory
- d) d) Mémoire partagée par tous les threads d'un GPU / Memory shared by all GPU threads

**Réponse / Answer:**

a) Mémoire qui ne nécessite pas de transfert entre l'hôte et le GPU / Memory that requires no transfer between host and GPU

**Question 19:**

En français : Quels sont les deux principes de localité en mémoire ? / What are the two principles of memory locality? / In English: Quels sont les deux principes de localité en mémoire ? / What are the two principles of memory locality?

- a) a) Temporelle et spatiale / Temporal and Spatial
- b) b) Horizontale et verticale / Horizontal and Vertical
- c) c) Séquentielle et parallèle / Sequential and Parallel
- d) d) Interne et externe / Internal and External

**Réponse / Answer:**

a) Temporelle et spatiale / Temporal and Spatial

**Question 20:**

En français : Quelle commande permet d'activer le cache L1 pour la mémoire globale ? / Which command enables L1 caching for global memory? / In English: Quelle commande permet d'activer le cache L1 pour la mémoire globale ? / Which command enables L1 caching for global memory?

- a) a) -Xptxas -dlcm=ca / -Xptxas -dlcm=ca
- b) b) -Xptxas -dlcm=cg / -Xptxas -dlcm=cg

- c) c) -arch=sm\_61 / -arch=sm\_61
- d) d) -use\_fast\_math / -use\_fast\_math

**Réponse / Answer:**

- a) -Xptxas -dlcm=ca / -Xptxas -dlcm=ca

**Question 21:**

En français : Quelle est la mémoire la plus lente dans un GPU ? / What is the slowest memory in a GPU? / In English: Quelle est la mémoire la plus lente dans un GPU ? / What is the slowest memory in a GPU?

- a) a) Registres / Registers
- b) b) Mémoire locale / Local Memory
- c) c) Mémoire globale / Global Memory
- d) d) Mémoire texture / Texture Memory

**Réponse / Answer:**

- c) Mémoire globale / Global Memory

## 20 Questions - CUDA Parallel Programming

### Question 1

Quelle est la dimension de la grille et le nombre total de threads pour une configuration avec block.x = 32, block.y = 16 et une matrice de 10240 x 2560 ? / What is the grid dimension and the total number of threads for a configuration with block.x = 32, block.y = 16 and a matrix of 10240 x 2560?

- a) grid.x = 320, grid.y = 160, Total threads = 26214400
- b) grid.x = 160, grid.y = 80, Total threads = 6553600
- c) grid.x = 640, grid.y = 320, Total threads = 104857600
- d) grid.x = 320, grid.y = 160, Total threads = 13107200

Réponse / Answer: a) grid.x = 320, grid.y = 160, Total threads = 26214400

### Question 2

Quelle est la dimension de la grille et le nombre total de threads pour une configuration avec block.x = 64, block.y = 64 et une matrice de 12288 x 12288 ? / What is the grid dimension and the total number of threads for a configuration with block.x = 64, block.y = 64 and a matrix of 12288 x 12288?

- a) grid.x = 192, grid.y = 192, Total threads = 150994944
- b) grid.x = 96, grid.y = 96, Total threads = 37748736
- c) grid.x = 384, grid.y = 384, Total threads = 603979776
- d) grid.x = 192, grid.y = 192, Total threads = 75497472

Réponse / Answer: a) grid.x = 192, grid.y = 192, Total threads = 150994944

### Question 3

Quelle est la dimension de la grille et le nombre total de threads pour une configuration avec block.x = 16, block.y = 64 et une matrice de 4096 x 20480 ? / What is the grid dimension and the total number of threads for a configuration with block.x = 16, block.y = 64 and a matrix of 4096 x 20480?

- a) grid.x = 256, grid.y = 320, Total threads = 83886080
- b) grid.x = 128, grid.y = 160, Total threads = 20971520
- c) grid.x = 512, grid.y = 640, Total threads = 335544320
- d) grid.x = 256, grid.y = 320, Total threads = 41943040

Réponse / Answer: a) grid.x = 256, grid.y = 320, Total threads = 83886080

#### Question 4

Quelle est la dimension de la grille et le nombre total de threads pour une configuration avec block.x = 16, block.y = 8 et une matrice de 3584 x 1536 ? / What is the grid dimension and the total number of threads for a configuration with block.x = 16, block.y = 8 and a matrix of 3584 x 1536?

- a) grid.x = 224, grid.y = 192, Total threads = 5505024
- b) grid.x = 112, grid.y = 96, Total threads = 1376256
- c) grid.x = 448, grid.y = 384, Total threads = 22020096
- d) grid.x = 224, grid.y = 192, Total threads = 2752512

Réponse / Answer: a) grid.x = 224, grid.y = 192, Total threads = 5505024

#### Question 5

Quelle est la dimension de la grille et le nombre total de threads pour une configuration avec block.x = 8, block.y = 64 et une matrice de 1280 x 14336 ? / What is the grid dimension and the total number of threads for a configuration with block.x = 8, block.y = 64 and a matrix of 1280 x 14336?

- a) grid.x = 160, grid.y = 224, Total threads = 18350080
- b) grid.x = 80, grid.y = 112, Total threads = 4587520
- c) grid.x = 320, grid.y = 448, Total threads = 73400320
- d) grid.x = 160, grid.y = 224, Total threads = 9175040

Réponse / Answer: a) grid.x = 160, grid.y = 224, Total threads = 18350080

#### Question 6

Quelle est la dimension de la grille et le nombre total de threads pour une configuration avec block.x = 16, block.y = 8 et une matrice de 4608 x 1536 ? / What is the grid dimension and the total number of threads for a configuration with block.x = 16, block.y = 8 and a matrix of 4608 x 1536?

- a) grid.x = 288, grid.y = 192, Total threads = 7077888
- b) grid.x = 144, grid.y = 96, Total threads = 1769472
- c) grid.x = 576, grid.y = 384, Total threads = 28311552
- d) grid.x = 288, grid.y = 192, Total threads = 3538944

Réponse / Answer: a) grid.x = 288, grid.y = 192, Total threads = 7077888

### Question 7

Quelle est la dimension de la grille et le nombre total de threads pour une configuration avec block.x = 16, block.y = 8 et une matrice de 4096 x 1792 ? / What is the grid dimension and the total number of threads for a configuration with block.x = 16, block.y = 8 and a matrix of 4096 x 1792?

- a) grid.x = 256, grid.y = 224, Total threads = 7340032
- b) grid.x = 128, grid.y = 112, Total threads = 1835008
- c) grid.x = 512, grid.y = 448, Total threads = 29360128
- d) grid.x = 256, grid.y = 224, Total threads = 3670016

Réponse / Answer: a) grid.x = 256, grid.y = 224, Total threads = 7340032

### Question 8

Quelle est la dimension de la grille et le nombre total de threads pour une configuration avec block.x = 64, block.y = 64 et une matrice de 6144 x 8192 ? / What is the grid dimension and the total number of threads for a configuration with block.x = 64, block.y = 64 and a matrix of 6144 x 8192?

- a) grid.x = 96, grid.y = 128, Total threads = 50331648
- b) grid.x = 48, grid.y = 64, Total threads = 12582912
- c) grid.x = 192, grid.y = 256, Total threads = 201326592
- d) grid.x = 96, grid.y = 128, Total threads = 25165824

Réponse / Answer: a) grid.x = 96, grid.y = 128, Total threads = 50331648

### Question 9

Quelle est la dimension de la grille et le nombre total de threads pour une configuration avec block.x = 64, block.y = 16 et une matrice de 12288 x 4608 ? / What is the grid dimension and the total number of threads for a configuration with block.x = 64, block.y = 16 and a matrix of 12288 x 4608?

- a) grid.x = 192, grid.y = 288, Total threads = 56623104
- b) grid.x = 96, grid.y = 144, Total threads = 14155776
- c) grid.x = 384, grid.y = 576, Total threads = 226492416
- d) grid.x = 192, grid.y = 288, Total threads = 28311552

Réponse / Answer: a) grid.x = 192, grid.y = 288, Total threads = 56623104

### **Question 10**

Quelle est la dimension de la grille et le nombre total de threads pour une configuration avec block.x = 16, block.y = 64 et une matrice de 4608 x 20480 ? / What is the grid dimension and the total number of threads for a configuration with block.x = 16, block.y = 64 and a matrix of 4608 x 20480?

- a) grid.x = 288, grid.y = 320, Total threads = 94371840
- b) grid.x = 144, grid.y = 160, Total threads = 23592960
- c) grid.x = 576, grid.y = 640, Total threads = 377487360
- d) grid.x = 288, grid.y = 320, Total threads = 47185920

Réponse / Answer: a) grid.x = 288, grid.y = 320, Total threads = 94371840

### **Question 11**

Quelle est la dimension de la grille et le nombre total de threads pour une configuration avec block.x = 64, block.y = 8 et une matrice de 16384 x 1792 ? / What is the grid dimension and the total number of threads for a configuration with block.x = 64, block.y = 8 and a matrix of 16384 x 1792?

- a) grid.x = 256, grid.y = 224, Total threads = 29360128
- b) grid.x = 128, grid.y = 112, Total threads = 7340032
- c) grid.x = 512, grid.y = 448, Total threads = 117440512
- d) grid.x = 256, grid.y = 224, Total threads = 14680064

Réponse / Answer: a) grid.x = 256, grid.y = 224, Total threads = 29360128

### **Question 12**

Quelle est la dimension de la grille et le nombre total de threads pour une configuration avec block.x = 64, block.y = 8 et une matrice de 18432 x 512 ? / What is the grid dimension and the total number of threads for a configuration with block.x = 64, block.y = 8 and a matrix of 18432 x 512?

- a) grid.x = 288, grid.y = 64, Total threads = 9437184
- b) grid.x = 144, grid.y = 32, Total threads = 2359296
- c) grid.x = 576, grid.y = 128, Total threads = 37748736
- d) grid.x = 288, grid.y = 64, Total threads = 4718592

Réponse / Answer: a) grid.x = 288, grid.y = 64, Total threads = 9437184

### **Question 13**

Quelle est la dimension de la grille et le nombre total de threads pour une configuration avec block.x = 64, block.y = 8 et une matrice de 6144 x 1792 ? / What is the grid dimension and the total number of threads for a configuration with block.x = 64, block.y = 8 and a matrix of 6144 x 1792?

- a) grid.x = 96, grid.y = 224, Total threads = 11010048
- b) grid.x = 48, grid.y = 112, Total threads = 2752512
- c) grid.x = 192, grid.y = 448, Total threads = 44040192
- d) grid.x = 96, grid.y = 224, Total threads = 5505024

Réponse / Answer: a) grid.x = 96, grid.y = 224, Total threads = 11010048

### **Question 14**

Quelle est la dimension de la grille et le nombre total de threads pour une configuration avec block.x = 32, block.y = 16 et une matrice de 5120 x 1536 ? / What is the grid dimension and the total number of threads for a configuration with block.x = 32, block.y = 16 and a matrix of 5120 x 1536?

- a) grid.x = 160, grid.y = 96, Total threads = 7864320
- b) grid.x = 80, grid.y = 48, Total threads = 1966080
- c) grid.x = 320, grid.y = 192, Total threads = 31457280
- d) grid.x = 160, grid.y = 96, Total threads = 3932160

Réponse / Answer: a) grid.x = 160, grid.y = 96, Total threads = 7864320

### **Question 15**

Quelle est la dimension de la grille et le nombre total de threads pour une configuration avec block.x = 16, block.y = 32 et une matrice de 2048 x 8192 ? / What is the grid dimension and the total number of threads for a configuration with block.x = 16, block.y = 32 and a matrix of 2048 x 8192?

- a) grid.x = 128, grid.y = 256, Total threads = 16777216
- b) grid.x = 64, grid.y = 128, Total threads = 4194304
- c) grid.x = 256, grid.y = 512, Total threads = 67108864
- d) grid.x = 128, grid.y = 256, Total threads = 8388608

Réponse / Answer: a) grid.x = 128, grid.y = 256, Total threads = 16777216

### **Question 16**

Quelle est la dimension de la grille et le nombre total de threads pour une configuration avec block.x = 16, block.y = 64 et une matrice de 4608 x 10240 ? / What is the grid dimension and the total number of threads for a configuration with block.x = 16, block.y = 64 and a matrix of 4608 x 10240?

- a) grid.x = 288, grid.y = 160, Total threads = 47185920
- b) grid.x = 144, grid.y = 80, Total threads = 11796480
- c) grid.x = 576, grid.y = 320, Total threads = 188743680
- d) grid.x = 288, grid.y = 160, Total threads = 23592960

Réponse / Answer: a) grid.x = 288, grid.y = 160, Total threads = 47185920

### **Question 17**

Quelle est la dimension de la grille et le nombre total de threads pour une configuration avec block.x = 16, block.y = 64 et une matrice de 5120 x 16384 ? / What is the grid dimension and the total number of threads for a configuration with block.x = 16, block.y = 64 and a matrix of 5120 x 16384?

- a) grid.x = 320, grid.y = 256, Total threads = 83886080
- b) grid.x = 160, grid.y = 128, Total threads = 20971520
- c) grid.x = 640, grid.y = 512, Total threads = 335544320
- d) grid.x = 320, grid.y = 256, Total threads = 41943040

Réponse / Answer: a) grid.x = 320, grid.y = 256, Total threads = 83886080

### **Question 18**

Quelle est la dimension de la grille et le nombre total de threads pour une configuration avec block.x = 32, block.y = 16 et une matrice de 10240 x 3584 ? / What is the grid dimension and the total number of threads for a configuration with block.x = 32, block.y = 16 and a matrix of 10240 x 3584?

- a) grid.x = 320, grid.y = 224, Total threads = 36700160
- b) grid.x = 160, grid.y = 112, Total threads = 9175040
- c) grid.x = 640, grid.y = 448, Total threads = 146800640
- d) grid.x = 320, grid.y = 224, Total threads = 18350080

Réponse / Answer: a) grid.x = 320, grid.y = 224, Total threads = 36700160

### **Question 19**

Quelle est la dimension de la grille et le nombre total de threads pour une configuration avec block.x = 16, block.y = 32 et une matrice de 5120 x 5120 ? / What is the grid dimension and the total number of threads for a configuration with block.x = 16, block.y = 32 and a matrix of 5120 x 5120?

- a) grid.x = 320, grid.y = 160, Total threads = 26214400
- b) grid.x = 160, grid.y = 80, Total threads = 6553600
- c) grid.x = 640, grid.y = 320, Total threads = 104857600
- d) grid.x = 320, grid.y = 160, Total threads = 13107200

Réponse / Answer: a) grid.x = 320, grid.y = 160, Total threads = 26214400

### **Question 20**

Quelle est la dimension de la grille et le nombre total de threads pour une configuration avec block.x = 16, block.y = 16 et une matrice de 1536 x 2560 ? / What is the grid dimension and the total number of threads for a configuration with block.x = 16, block.y = 16 and a matrix of 1536 x 2560?

- a) grid.x = 96, grid.y = 160, Total threads = 3932160
- b) grid.x = 48, grid.y = 80, Total threads = 983040
- c) grid.x = 192, grid.y = 320, Total threads = 15728640
- d) grid.x = 96, grid.y = 160, Total threads = 1966080

Réponse / Answer: a) grid.x = 96, grid.y = 160, Total threads = 3932160