

## 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) Dépassement de capacité de la mémoire globale / Global memory overflow

**Réponse / Answer:**

b) 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) 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) 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) 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 téra-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) 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) Une utilisation déséquilibrée des partitions mémoire / An uneven usage of memory partitions

b) Un alignement mémoire incorrect / An incorrect memory alignment

c) Un dépassement des limites de la mémoire locale / An overflow of local memory

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) Réduire les transactions mémoire / To reduce memory transactions

b) Augmenter la latence mémoire / To increase memory latency

c) Réduire la capacité mémoire / To reduce memory capacity

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) Mémoire globale et locale / Global and Local Memory

b) Mémoire partagée et constante / Shared and Constant Memory

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