

**King Saud University**  
**College of Computer and Information Sciences**  
**Department of Computer Science**  
**CSC453 – Parallel Processing – Tutorial No 3 – Fall 2021**

## Question

1. What GPGPU stands for and what does it mean.

General Purpose GPU, a GPU that has the ability to perform calculations that are usually dedicated for CPU

2. Why CUDA is said Heterogeneous computing.

2- Processing is handled by two different processors  
the low latency code performed by CPU in a Serial way  
the high latency code performed by GPU in a Parallel way

3. Give the definition of the following terms:

a. Device: Refers to the GPU and its memory

b. Kernel. A function that runs on the device. One kernel executed at a time and Many threads execute each kernel.

c. Grid of thread blocks. The kernel is executed by a grid of thread blocks. Each Grid has a collection of blocks and each block has a collection of threads.

d. Warp.

Group of 32 threads of the same block

4. Explain the parallel programming model of CUDA.

One kernel is executed at a time. Kernel executed by a grid of thread blocks and threads of the same Warp they execute the same instruction at the same time

5. Enumerate and explain the different types of memory adopted by CUDA.

6. Explain Why the Constant memory is cached, while the Global memory is not.

Constant Memory is read only. caching has no overhead because it doesn't has cache coherency problem  
Global Memory is read/write. It has cache coherency problem and the overhead to maintain it will be very high we have thousands of threads running.

cached == read only

1- Registers: per thread, 32bit, on chip

2- local memory: per thread, relative large, in DRAM

3- shared memory: per block, 16KB, on chip

4- Global memory: per grid, non cached, in DRAM

5- Constant memory: per grid, cached, in DRAM

6- Texture memory: per grid, cached, in DRAM

Registers, shared: On chip

rest are in DRAM

**King Saud University**  
**College of Computer and Information Sciences**  
**Department of Computer Science**  
**CSC453 – Parallel Processing – Tutorial No 3 – Fall 2021**