Modern Education Society's
Wadia College of Engineering, Pune

NAME OF STUDENT:	**CLASS:**
SEMESTER/YEAR:	**ROLL NO:**
DATE OF PERFORMANCE:	**DATE OF SUBMISSION:**
EXAMINED BY:	**EXPERIMENT NO: HPC- 03**

TITLE: Parallel Reduction.

AIM: Implement Parallel Reduction using Min, Max, Sum and Average operations.

OBJECTIVES:

Students should be able to perform CUDA Program.

PRE-REQUISITES:

1. Knowledge of CUDA programming.
2. Knowledge of parallel programming.
3. Knowledge of operations on large vectors.

What is CUDA:-

CUDA (Compute Unified Device Architecture) is a parallel computing platform and programming model developed by NVIDIA. It allows developers to use the power of NVIDIA graphics processing units (GPUs) to accelerate computation tasks in various applications, including scientific computing, machine learning, and computer vision. CUDA provides a set of programming APIs, libraries, and tools that enable developers to write and execute parallel code on NVIDIA GPUs. It supports popular programming languages like C,C++,and Python, and provides a simple programming model that abstracts away much of the low level details of GPU architecture. Using CUDA, developers can exploit the massive parallel is and high computational power of GPUs to accelerate computationally intensive tasks, such as matrix operations, image processing, and deep learning. CUDA has become an important tool for scientific research and is widely used in fields like physics, chemistry, biology, and engineering.

Execution of Program over CUDA Environment

Here are the steps to run a CUDA program for adding two large vectors:

1. Install CUDA Toolkit: First, you need to install the CUDA Toolkit on your system. You can download the CUDA Toolkit from the NVIDIA website and follow the installation instructions provided.
2. Set up CUDA environment: Once the CUDA Toolkit is installed, you need to set up the CUDA environment on your system. This involves setting the PATH and LD_LIBRARY_PATH environment variables to the appropriate directories.
3. Write the CUDA program: You need to write a CUDA program that performs the addition of two large vectors. You can use a text editor to write the program and save it with a .cu extension.
4. Compile the CUDA program: You need to compile the CUDA program using the nvcc compiler that comes with the CUDA Toolkit. The command to compile the program is:

```
nvcc -o program_name program_name.cu
```

5. This will generate an executable program named program_name.

Run the CUDA program: Finally, you can run the CUDA program by executing the executable file generated in the previous step. The command to run the program is:

```
./program_name
```

This will execute the program and perform the addition of two large vectors.

QUESTIONS FOR REVIEW:

1. How is CUDA programming useful to study parallel algorithms?
2. Discuss the importance of parallel reduction operations.
3. Discuss the operations on vectors and the approach for parallel algorithm design for the same.
4. How is parallelism achieved in CUDA?
5. Explain Grid, Block and Thread structure in relation with parallel reduction.