**Problem Statement:** Matrix Multiplication using CUDA C

**Code:**

**Command to install nvidia gpu on google colab:**  !nvidia-smi

import numpy as np

import cupy as cp

import time

# Size of matrices

N = 1024

# Create random matrices on CPU

A\_cpu = np.random.rand(N, N).astype(np.float32)

B\_cpu = np.random.rand(N, N).astype(np.float32)

# ---------- CPU ----------

start\_cpu = time.time()

C\_cpu = np.matmul(A\_cpu, B\_cpu)

end\_cpu = time.time()

print(f"CPU Matrix Multiplication Time: {end\_cpu - start\_cpu:.4f} seconds")

# ---------- GPU ----------

A\_gpu = cp.asarray(A\_cpu)

B\_gpu = cp.asarray(B\_cpu)

start\_gpu = time.time()

C\_gpu = cp.matmul(A\_gpu, B\_gpu)

cp.cuda.Device(0).synchronize()

end\_gpu = time.time()

print(f"GPU Matrix Multiplication Time: {end\_gpu - start\_gpu:.4f} seconds")

**Command to run the shells in google colab:**

Shift enter

**Output:**  
CPU Matrix Multiplication Time: 0.0516 seconds

GPU Matrix Multiplication Time: 0.6664 seconds