**Использование инструментов разработки CUDA-программ**

1. Откомпилируйте программу с уровнем оптимизации O2 и запустите профилировщик Nsight Compute (команда sudo nv-nsight-cu).
2. Создайте новый проект, укажите откомпилированный файл «Application Executable» и в разделе Activity → Profile выходной файл «Output File» и запустите профилировку «Launch»
3. Найдите и запишите в отчет:
   1. Время работы kernel-функции, конфигурацию пространства потоков, количество используемых одним потоком регистров
   2. Соотношение между вычислительными операциями и инструкциями доступа к памяти. Является ли ваша программа арифметически интенсивной?
   3. Практическую и теоретическую загруженность (Кнопка Occupancy calculator) мультипроцессора. Происходит ли у вас объединение нескольких блоков параллельных потоков для выполнения на одном мультипроцессоре? Сколько блоков объединяется?
   4. В соответствии с параметрами вашей видеокарты, указанными в Occupancy calculator, оцените сколько блоков параллельных потоков будет объединяться на вашем мультипроцессоре, если вы сделаете размер блока 32? Запишите в отчет ваши вычисления
4. Инструкции. Найдите и укажите в отчете проценты:
   1. Соотношение между «Issue Active» и «Inst Executed» (executed - среднее число инструкций, выполняемых варпом за 1 такт работы процессора, без учета повторных инструкций, выполняемых из-за отсутствия данных в кэше или конфликтов обращения к разделяемой или константной памяти, или спиллинга регистров)
   2. Соотношение между инструкциями доступа к данным (Lsu), арифметикой (Fma, Alu) и инструкциями ветвления (Adu)
5. Память. Найдите и укажите в отчете процент времени, когда менеджер памяти был загружен (DRAM: Cycles Active)
6. Подумайте над значениями показателей. Соответствуют ли они вашим ожиданиям? Можно ли еще ускорить программу?
7. Если значения показателей говорят, что программу ускорить можно более чем в 2 раза, сделайте это. Прикрепите к ответу старый и новый вариант программы и напишите во сколько раз она ускорилась.