# An introduction to Nsight Systems

Mike Giles

## 1 Overview and client installation

NVIDIA's Nsight Systems profiling tools are usually used in two stages.

First, the user uses a command such as

#### nsys profile application\_code

to run their code and produce an output file with a name of the form filename.nsys-rep.

Then, the user runs an interactive graphics program to display the results of the profiling, using a command such as

### nsys-ui filename.nsys-rep

The first command has to be used on a GPU node which is capable of executing your CUDA code. In this case, both nsys and nsys-ui are very likely to have been installed already.

However, the benefit of the two-stage process is that you can copy the report back to your desktop or laptop system to run nsys-ui locally to work with it interactively. In this case you may need to install nsys-ui on your local system even though you don't have an NVIDIA GPU.

The software and installation details are available from NVIDIA here.

On my Ubuntu system at home, for which I have sudo rights, I downloaded the appropriate \*.deb file, and installed it using the command:

sudo apt install nsight-systems-2023.4.1\_2023.4.1.97-1\_amd64.deb

NOTE: this was correct on 6/12/2023 but may change to a newer version in the future.

For desktops within the university it might be better to ask your IT support team to install it for general use, or you could perhaps install it in your own user space just for your use.

### 2 Using the interactive viewer

Starting the viewer with a command of the form

```
nsys-ui filename.nsys-rep
```

produces a display which looks like

Yerw Icols Help			
rt1 nsys-rep ×			
Timeline View -			📾 🔍 🖳 tx 🔥 <u>3 warnings, 11 me</u>
• 0a	1a 2a 3a 4a 5a 6a 7a 8a 9a 10a	11s 12s 13s 14s	15a 16a 17a
CUDA HW (0000.02:00.0 - NVIDIA	· · · · · · · · · · · · · · · · · · ·		
Threads (7)			
(38434) laplace3d			
OS runtime libraries CUDA API	Restorm. Resta		
Profiler overhead	L NU		
(38444) cuda-EvtHandlr			
OS runtime libraries			
Profiler overhead			
5 threads hidden+			
na View 👻			
			Name *
		Descrip	ution:
	Right-click a timeline row and select "Show in Events View" to see events here		
	regrit calck a simesine row and select, and in Events view, to see events nee view.		

Clicking on CUDA API, and then right-clicking on "Show in Events View" leads to the bottom Events View pane showing a list of the kernels and CUDA operations carried out on the GPU (including how long they each took):

	Systems 2023.2.1													
w Iools H	Help													
ysirep X														
line View	*											📼 Q 🖓 🛒	1x 🛆	3 warnings, 11 m
	• Cs	18	38	ls 5s	69 71	8 89	98 108	118 12	s	13s	149	159	168	178
HW (0000.0	202:00.0 - NVIDIA	· · · · · ·												
is (7)														
34) laplace3	*3d													
runtime libr	Ibraries													
JDA API		CudaEven.	4											
ofiler overhe	head CUDAp	2												1
i runtime libr	libraries													
S runtime libr	Ibraries head													
8444] cuda-Evt 26 runtime libr Profiler overhes hreads hidden	Ibraries head													
S runtime libr	Ibraries head len = +											1		
IS runtime libr trofiler overhei hreads hidden	Ibraties head len+									1 1	1	1	1 I	
S runtime libr rofiler overhes reads hidden	Ibraries head len = +									1				
G runtime libr rofiler overhei rreads hidden	Ibraties head len+									1		Name		
S runtime libr ofiler overhes reads hidden	Ibraries head len								Duration	TID	Description:	1 1		1 1
S runtime libr ofiler overhes reads hidden Yew	Ibraries head len							I I I I			1	1 1		
runtime libr ofiler overhei eads hidden iew * Na	Ibranies head ien							Start	Duration	וו	1	1 1		
vads hidden www.	Itrailes head ifin+  Name cuModul6GetLootingMode cuUbraryLoadData cudebeatCreate							Start 0.81867s 0.825095c 0.820794c	Duration 2:125 ps 196.004 ps 517.002 ms	TID 38434 38434 38434	1	1 1		
ender overheiten und der overheiten verheiten und der overheiten verheiten v	ktranies head im							Sart 0.318475 0.3205956 0.3205956 0.3205956 0.3205956 0.3205956	Duration 2.125 µs 196,004 µs 517,002 ms 3.468 µs	TID 38434 38434 38434 38434	1	1 1		
ender verheiten ihre einer verheiten einer ver	Itraries head en							9art 0.819956 0.829756 1.33786 1.33784	Duration 2.125 ps 196.04 ps 517.02 ms 3.468 ps 1.470 ms	TID 38434 38434 38434 38434 38434 38434	1	1 1		
ir untime libri ofiler overhese eads hidden iew * Na	Itrailes head inn							Start 0.31 8675 0.207958 0.207958 1.32764 1.32764 1.32764 1.329523	Duration 2.125 µs 196.004 µs 517.002 ms 3.468 µs 1.470 ms 1.334 ms	TID 38434 38434 38434 38434 38434 38434 38434	1	1 1		
ir untime libri ofiler overhese eads hidden iew * Na	htraines head tim+ Name cutAduideEett.ootingMode cutLivary LoadOtta cutLivary LoadOtta cutLivary LoadOtta cutLivary LoadOtta cutLivary LoadOtta							Bart 0.81847 0.81847 0.81847 0.820595 0.8205900000000000000000000000000000000000	Duration 2.125 ps 196.004 ps 517.002 ms 3.468 ps 1.470 ms 1.334 ms 2.808 ms	TID 38434 38434 38434 38434 38434 38434 38434 38434	1	1 1		
s runtime libr ofiler overhes resids hidden litew * Na	kkanise head fmn							50at 0.8116/76 0.82095/50 0.82095/50 0.82095/50 1.3376/4 1.3376/4 1.3376/4 1.3376/4 1.3376/4 1.3376/2 1.3376/4 1.3376/2	Duration 2.125 ps 196.004 ps 517.002 ms 3.468 ps 1.470 ms 1.334 ms 2.808 ms 64.689 ms	TID 38434 38434 38434 38434 38434 38434 38434 38434 38434	1	1 1		
s runtime libr ofiler overhes resids hidden litew * Na	htraines head tim+ Name cutAduideEett.ootingMode cutLivary LoadOtta cutLivary LoadOtta cutLivary LoadOtta cutLivary LoadOtta cutLivary LoadOtta							Bart 0.81847 0.81847 0.81847 0.820595 0.8205900000000000000000000000000000000000	Duration 2.125 ps 196.004 ps 517.002 ms 3.468 ps 1.470 ms 1.334 ms 2.808 ms	TID 38434 38434 38434 38434 38434 38434 38434 38434	1	1 1		

Clicking on one of those, and then zooming in by ctrl+mouse wheel (which zooms in time relative to the current mouse location, a bit like Google Map zooms) gives the following:

IDIA Nsight Sys Yew Iools Hel											
View Tools Hei	ib.										
									-		
Timeline View										Q 🖓 👖 1x 🛆 3 🗰	
	2s • 2ms	+92.02ms +92.0	04ms +92.05ms	+92.08ms +92.1ms +92.12ms +92	2.14ms +92.16ms +92.18ms	+92.2ms +92.22ms	+92.24ms	+92.26m	s +92.28ms	+92.3ms +92.32ms	
DA HW (0000.02)	00.0 - NV 1										
eads (7)											
18434) laplace3d											
OS runtime librar	ties		÷								
CUDA API		GPUJaplace3d	GPU_Japlace3d	0PU. 0P. 0P. 0P. 0. 0P. 0PU.Japlace3d 0P. 00.			cudaEventSynchronik				
Profiler overhead											
8444] cuda-EvtH	and a										
05 runtime librar					pol						
Profiler overhead					por						
threads hidden											
	4										
s View 👻											
										Name •	
* Nam						Start	Duration	TID	<ul> <li>Description:</li> </ul>		
	cudaEventSynchronize					2.091895	68.633 µs	38434	Call to GPU_laplace3d		
	cudaEventRecord					2.091995	2.504 µs	38434	Kernel launcher Begins: 2.09205s		
	GPU_laplace3d					2.092s	46.829 µs	38434	Ends: 2.09208s (+29.120 us)	/	
	GPU_laplace3d					2.09205s	29.120 µs	38434	Return value: 0 GPU: 0000.02:00.0 - NVIDIA	GeForce RTX 2080 Ti	
	GPU_laplace3d					2.09208s	5.057 µs	38434	Stream: 7	An once mine above 11	
	GPU_leplace3d					2.09208s	4.481 µs	38434	Latency: 4.548 ms→ Correlation ID: 476		
	GPU_laplace3d					2.09209s	3.961 µs	38434			
	GPU_laplace3d					2.09209s	3.791 µs	38434			
	GPU_laplace3d					2.09215	3.687 µs	38434			
	GPU_laplace3d					2.0921s	3.867 µs	38434			

Clicking on the keyboard symbol in the top-right gives other ways of navigating the primary Timeline View pane. There is also a slider bar which controls the vertical spacing.

For more examples of the use of nsys and nsys-ui see these slides from ISC 2023.