

## Технические характеристики топовых ускорителей Nvidia

	A100 (80 ГБ)	A100 (40 ГБ)	V100
Ядер FP32 CUDA	6912	6912	5120
Макс. частота	1,41 ГГц	1,41 ГГц	1,53 ГГц
Частота памяти	3,2 Гбита/с HBM2e	2,4 Гбита/с HBM2	1,75 Гбита/с HBM2
Шина памяти	5120 бит	5120 бит	4096 бит
Полоса пропускания памяти	2,0 ТБ/с	1,6 ТБ/с	0,9 ТБ/с
VRAM	80 ГБ	40 ГБ	16 ГБ/32 ГБ
Числа одинарной точности	19,5 TFLOPs	19,5 TFLOPs	15,7 TFLOPs
Числа двойной точности	9,7 TFLOPs (1/2 FP32 rate)	9,7 TFLOPs (1/2 FP32 rate)	7,8 TFLOPs (1/2 FP32 rate)
Тензоры INT8	624 TOPs	624 TOPs	N/A
Тензоры FP16	312 TFLOPs	312 TFLOPs	125 TFLOPs
Тензоры TF32	156 TFLOPs	156 TFLOPs	N/A
Интерконнект	NVLink 3 12 Links (600GB/sec)	NVLink 3 12 Links (600GB/sec)	NVLink 2 6 Links (300GB/sec)
GPU	GA100 (826 мм <sup>2</sup> )	GA100 (826 мм <sup>2</sup> )	GV100 (815 мм <sup>2</sup> )
Транзисторов	54,2 млрд	54,2 млрд	21,1 млрд
TDP	400 Вт	400 Вт	300 Вт/350 Вт
Процесс производства	TSMC 7N	TSMC 7N	TSMC 12nm FFN
Интерфейс	SXM4	SXM4	SXM2/SXM3
Архитектура	Ampere	Ampere	Volta