

System	WikiText-2		CIFAR-10	
	Runtime (hrs)	GPU Utilization (%)	Runtime (hrs)	GPU Utilization (%)
PyTorch Distributed	60.48	7	17.96	6
DeepSpeed w/o Parallelism	62.99	6	19.21	7
DeepSpeed Model/ Data Parallel	35.16	21	10.94	23
FlexFlow	33.78	22	10.09	26
DeepSpeed Model/ Task Parallel	32.87	26	9.70	24
Pipeline Parallelism	12.4	68	3.62	52
Hydra	8.1	82	2.50	81