

Bitfusion Flexdirect vs Network Relationship Experiments (Mnist)

Physical host at 10.110.125.166 (with Nvidia gtx 1050 Ti gpu, ESXi)

Flexdirect Server at 10.110.125.113 (with Nvidia gtx 1050 Ti gpu pass through)

Training script Mnist with 20 Epoch

Flexdirect Client	Description	Training Time per epoch (s)	Max GPU utilization	Avg GPU utilization	Min GPU utilization	GPU memory usage
10.110.125.13(vm)	1. Vm that contains the Flexdirect server 2. Vm that builds on the physical host (10.110.125.166) that contains the GPU 3. Vm Uses GPU resources directly (ESXi pass through)	7.85s(batch_size = 1) 0.6s(batch_size = 1024)	24%(batch_size = 1) 37(batch_size = 1024)	23%(batch_size = 1) 35(batch_size = 1024)	18%(batch_size = 1) 33(batch_size = 1024)	93%
10.110.125.13(vm)	1. Vm that contains the flexdirect server 2. Vm that builds on the physical host (10.110.125.166) that contains the GPU 3. Vm uses Flexdirect Client to obtain the GPU resources	8.45s(batch_size = 1) 0.55s(batch_size = 1024)	23%(batch_size = 1) 35(batch_size = 1024)	22%(batch_size = 1) 33(batch_size = 1024)	21%(batch_size = 1) 33(batch_size = 1024)	91%
10.110.125.80(vm)	1. Vm that builds on the physical host (10.110.125.166) that contains the GPU 2. Using Flex Direct Client to obtain the GPU resources	23.6s(batch_size = 1) 1.2s(batch_size = 1024)	9%(batch_size = 1) 17%(batch_size = 1024)	5%(batch_size = 1) 16%(batch_size = 1024)	8%(batch_size = 1) 14%(batch_size = 1024)	91%
10.110.124.237(vm)	1. Vm that resides in the same LAN as the Flexdirect server (10.110.125.113) 2. Using Flex Direct Client to obtain the GPU resources	856s(batch_size = 1) 23.45s(batch_size = 1024)	1%(batch_size = 1) 2%(batch_size = 1024)	0%(batch_size = 1) 1% (batch_size = 1)	0%(batch_size = 1) 0%(batch_size = 1024)	91%
10.184.75.87(vm)	1. Vm that is deployed on Nimbus 2. Using Flex Direct Client to obtain the GPU resources	FAIL (connection time out)	FAIL (connection time out)	FAIL (connection time out)	FAIL (connection time out)	FAIL (connection time out)