uber / **horovod**

New issue

horovod performance decrease dramatically when run on multiple server #221

Closed **scotthuang1989** opened this issue on Mar 23 · 8 comments

scotthuang1989 commented on Mar 23

I have 2 server, each have 4 GPUs, if I run horovod on single server, 1 epoch takes 10 seconds, but If I run it on 2 serer, it takes 60 seconds. I am not familiar with mpi. So I can only debug it by observing system resources. when I run it on multiple server.

1. cpu utilization is around (60%), which is a bit higher then run it with single server (40%)
2. network trans/receive is about 20M/s, I have a 10G network card, so it should not be the bottleneck
3. for most time, GPU utilization is almost 0.

Is there any tool or documentation to debug this issue?

alsrgv commented on Mar 24 Collaborator

@scotthuang1989, what model are you training and are you using NCCL?

alsrgv added the question label on Mar 24

scotthuang1989 commented on Mar 24

just a simple LSTM model which is ok when run locally. And one of my college told me maybe is because network cable. But it need administrator to go to server room to check it, I will update the status when they have confirmation.

alsrgv commented on Mar 25 Collaborator

@scotthuang1989, how large is the checkpoint of your model?

scotthuang1989 commented on Mar 25

around 16M. I have 2 server, each have 4 GPUS.

alsrgv commented on Mar 25 • edited Collaborator

I see. You may have trouble scaling this small model over regular 10GbE because of latency. You may need RoCE or InfiniBand low latency network. Training bigger model may help with scaling, but then you may become bandwidth constrained.

Rule of thumb: 25GbE for 1080TI, 50GbE for P100, 100GbE for V100.

scotthuang1989 commented on Mar 25

My GPU is 1080Ti. sound like a long way to go...

Assignees

No one assigned

Labels

question

Projects

None yet

Milestone

No milestone

Notifications

2 participants

alsrgv commented on Mar 25 Collaborator

Yeah, network is quite demanding for distributed deep learning. How many epochs are you training? I wish my epochs took 10 seconds to run.

scotthuang1989 commented on Mar 26

In this example, 10 seconds if I run my model only on local GPUs, If I go distributed, it takes 60 seconds.

⊘ **scotthuang1989** closed this on Mar 26