**HW 9 Short Response and Screenshots**

Sylvia Xiaoshi Yang

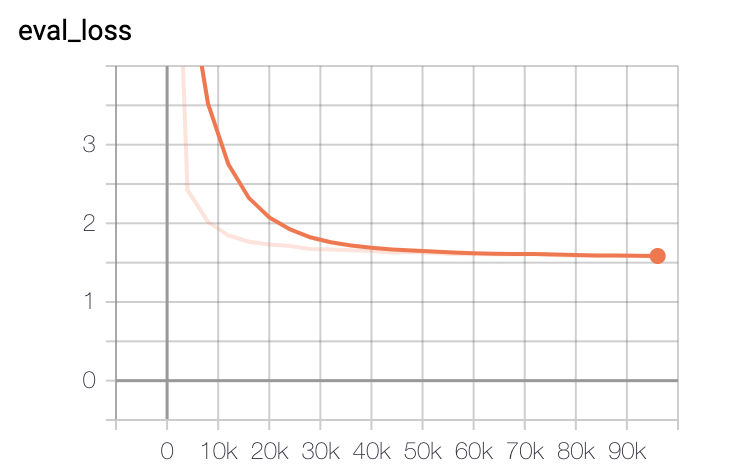
* **How long does it take to complete the training run? (hint: this session is on distributed training, so it *will* take a while)**

It took me 47+ hours for the 100k steps. Should be taking 6+ days in theory to complete the 300k steps.

* **Do you think your model is fully trained? How can you tell?**

Not yet as both the training and evaluation loss are decreasing.

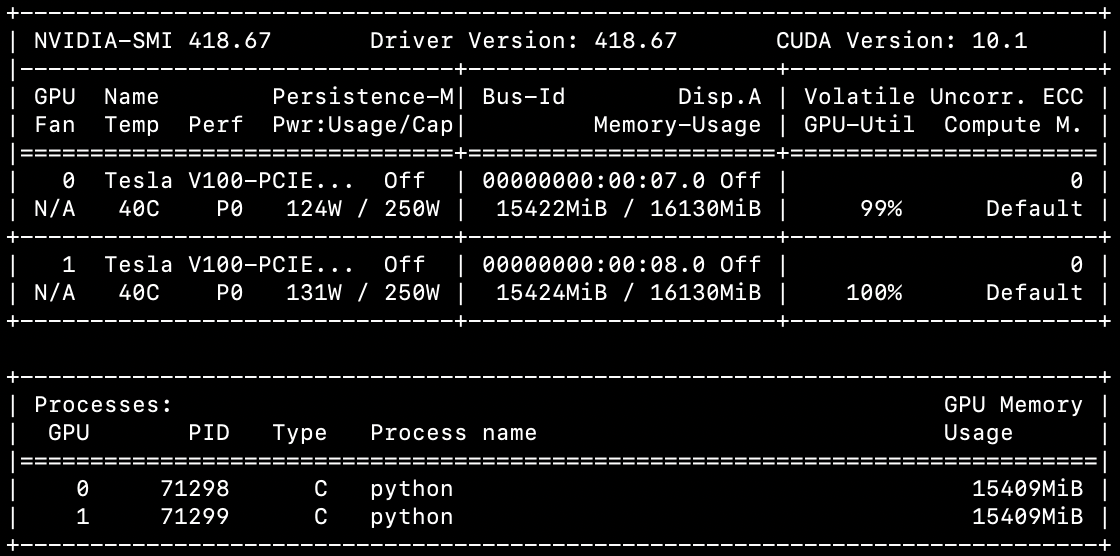
0.395 
O. 385 
0.375 
0.365 
0.355 
0.345 
o 
IOk 20k 30k 40k 50k 60k 70k 80k gok 



* **Were you overfitting?**

No. Eval loss value is around 1.6 and does not go up. Training loss value is still decreasing.

* **Were your GPUs fully utilized?**



Yes. It is fully utilized at 100%.

* **Did you monitor network traffic (hint: apt install nmon ) ? Was network the bottleneck?**

No. The outsize is around 100-200 MB, and the connection is 1000 MB, so it is not fully utilized.

nmon—14g 
Network 
1/F Name 
10 
e.e 
e.e 
e.e 
e.e 
ethl 
7.5 
1.1 
dockere 
e.e 
e.a 
e.e 
e.e 
e.a 
e.e 
ethe 
88.1 
76.3 
98.3 
[H for 
1/0 
Recv=KB/s Trans=KB/s packin packout 
Refresh= 2secs 
insize 
—22 : 34.11 
Trans 
98 
6 
outsize 
186.4 
184.2 
1181.5 
1181.5 
e.a 
12.4 
e.a 
3881614 . 5 
e.e 
92.3 
3675695.8 
.8 
8 

* **Take a look at the plot of the learning rate and then check the config file. Can you explain this setting?**

爿 00 L 相 日 
当 爿 0 レ : ℃ こ 
e 」 一 旧 」 e 

Learning rate takes a ramp up up until 8000 steps, then decreases gradually.

As the policy indicates, the learning rate has a “noam” learning rate decreasing scheme.

* **How big was your training set (mb)? How many training lines did it contain?**

Training set had 798 MB with 5278534 lines.

* **What are the files that a TF checkpoint is comprised of?**
* A .index file (checkpoint)
* A .data file (variable values)
* A .meta file (graphics)
* **How big is your resulting model checkpoint (mb)?**

The model checkpoint is 829 MB

* **Remember the definition of a "step". How long did an average step take?**

~1.7 seconds

* **How does that correlate with the observed network utilization between nodes?**

I think the network utilization between nodes should not have a significant impact on the GPU.