## The platform

OS: Ubuntu 18.04.5 LTS

RAM: 188.9 GB

Graphic: NV94

CPU: Intel® Xeon(R) CPU E5-2660 v2 @ 2.20GHz × 40

Package manager and Environment: Anaconda Navigator 2.1.4

Deep learning framework: Tensorflow 2.8.0

Coding language: Python 3.9.13

## A simplified version for using DMP

#go to the DeepMetagenome directory

!cd ../DeepMetagenome

#manually create two folder, namely resultsDMP and resultsnum2, for saving the DMP and Transformer models files.

#check the current work directory by typing ‘os.getcwd()’

#there are three variables subjected to change, MinLen or MaxLen, CutOffValue. New variables can also be added and modified the modules. For MT, we use MaxLen = 200 and CutOffValue = 0.000001

#call all the functions

#Here, the model frameworks based on LSTM and Transformer are compared respectively, Main and Main\_2 correspond to LSTM framework and Transformer framework respectively. the variable num\_heads can be modified in Deep\_Model\_ts\_2. For MT, we compared num\_head2 =2 and num\_heads = 4.

$from Main import \*

#$from Main\_2 import \*

#create a folder saving all the prediction dataset. Training dataset should be in the current work directory, or elsewhere and use absolute address

$preprocessing\_for\_multi\_FASTA("FILEPATH")

#or preprocessing\_for\_single\_FASTA(file). (Updated on 2024/5/12)

$training\_dataset\_preprocess("../MTtraningbinaryDatabase.csv") #here we take MTs as an example

#start to build model and prediction

$my\_model()

$model\_pred()

#an output file can be seen in the work directory. The file is a .csv file with three features of ‘index’, ‘probability’ and ‘sequence’.