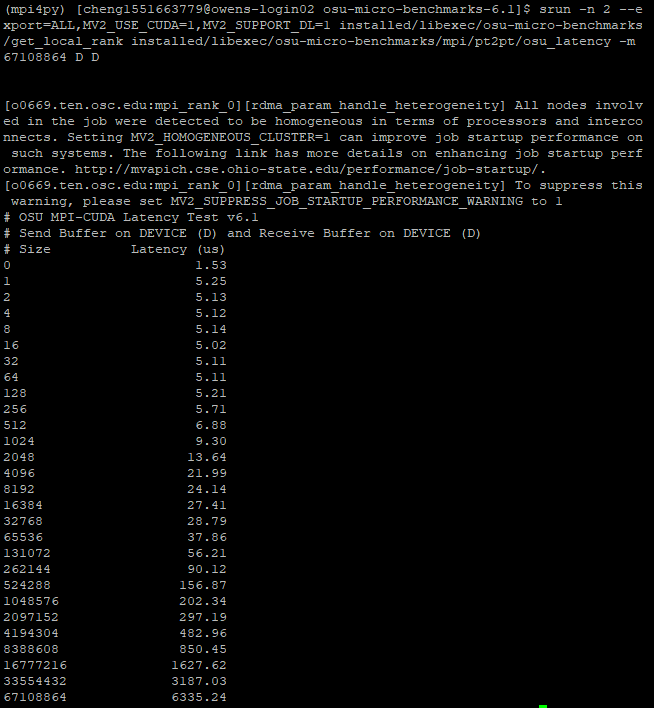
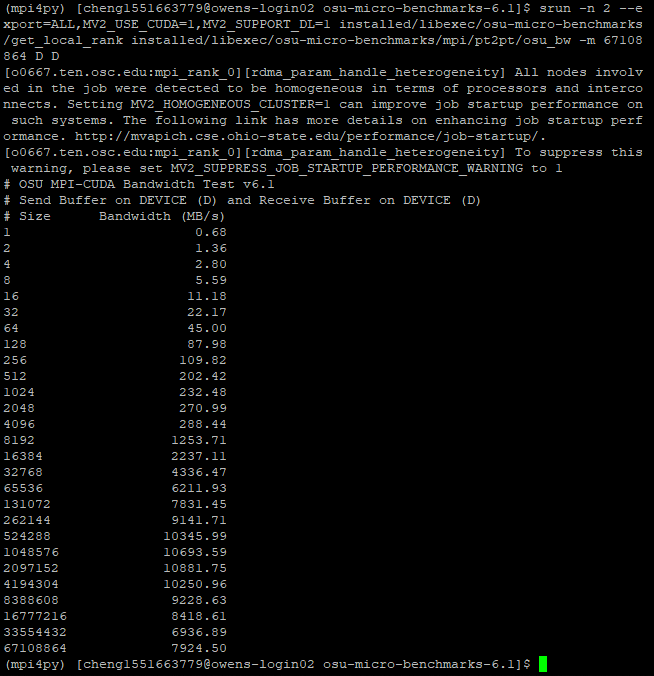
Allocate 2 nodes and run OMB

1.gpu

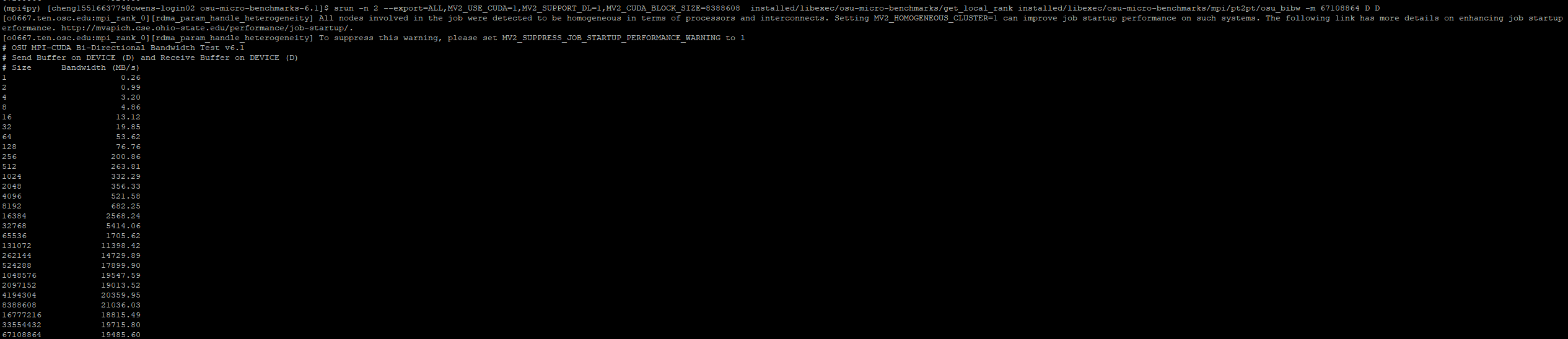
Latency



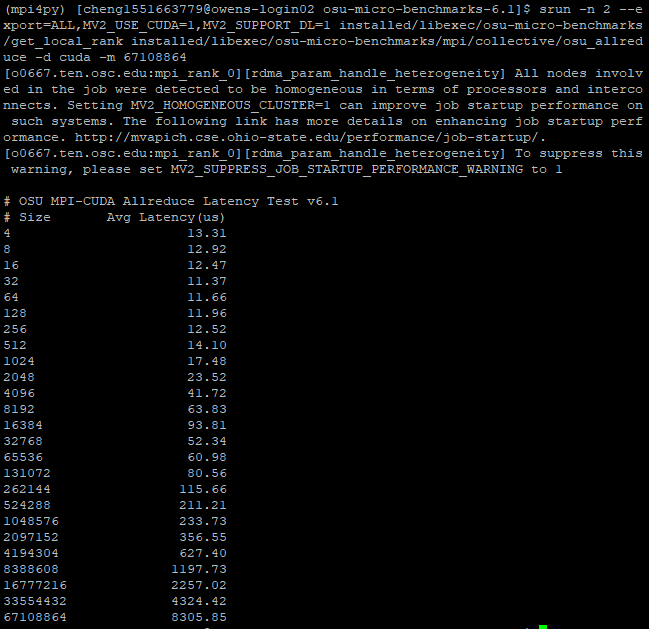
Bw



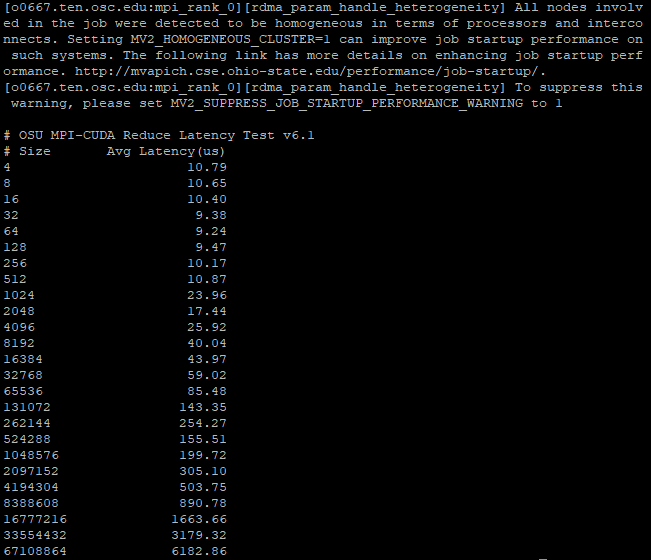
Bibw



Allreduce

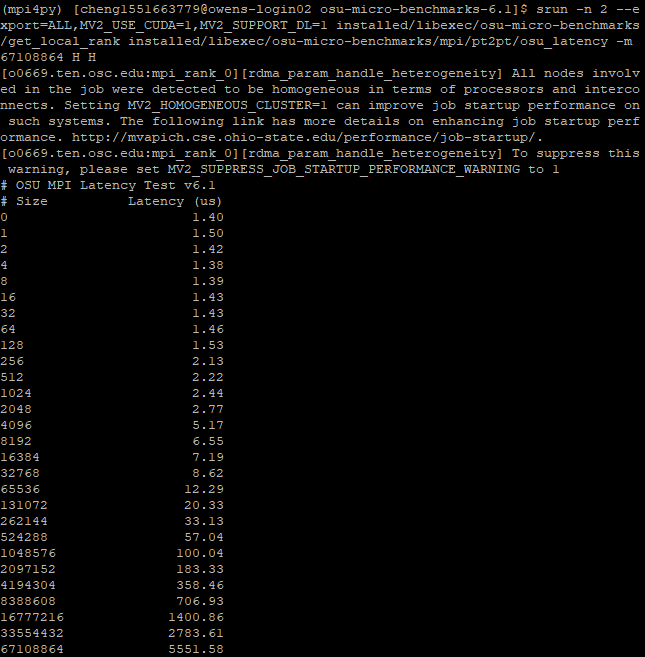


Reduce

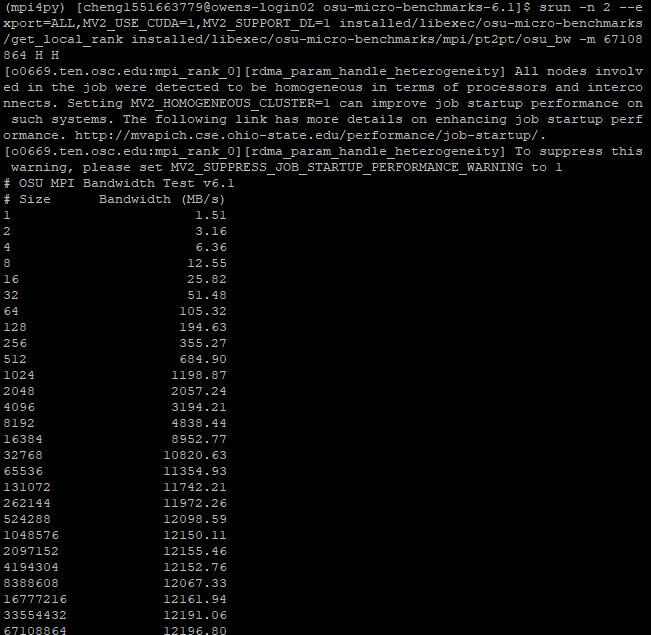


Cpu

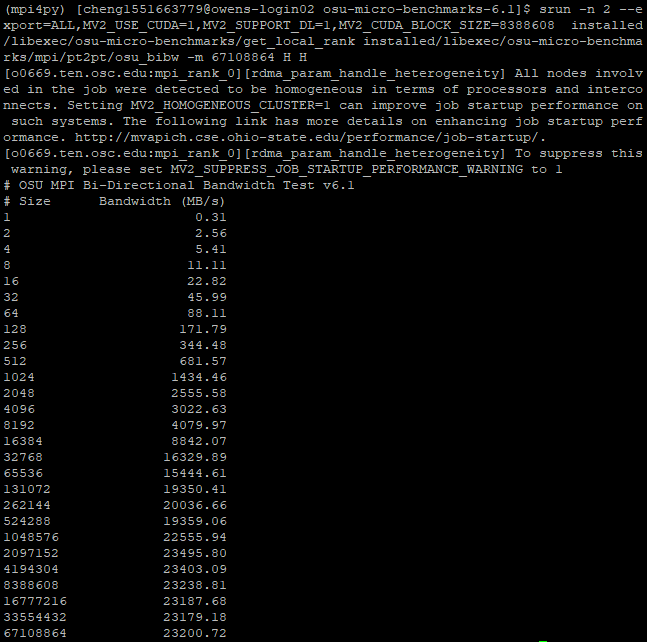
Latency



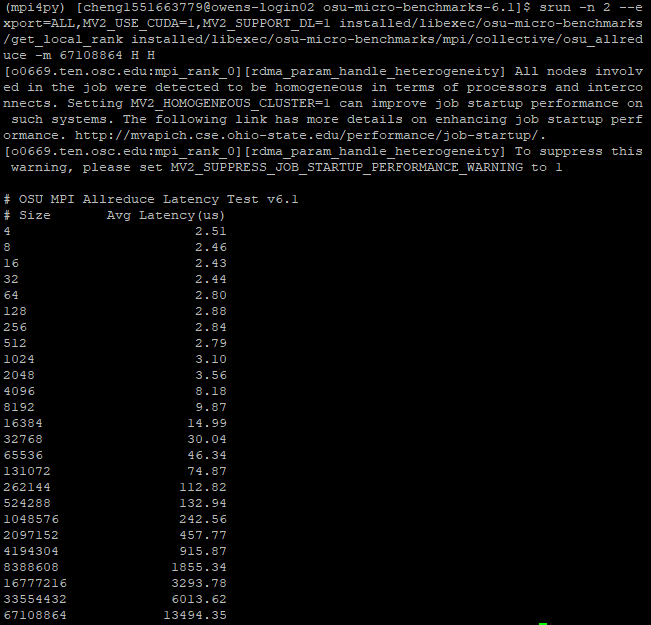
Bw



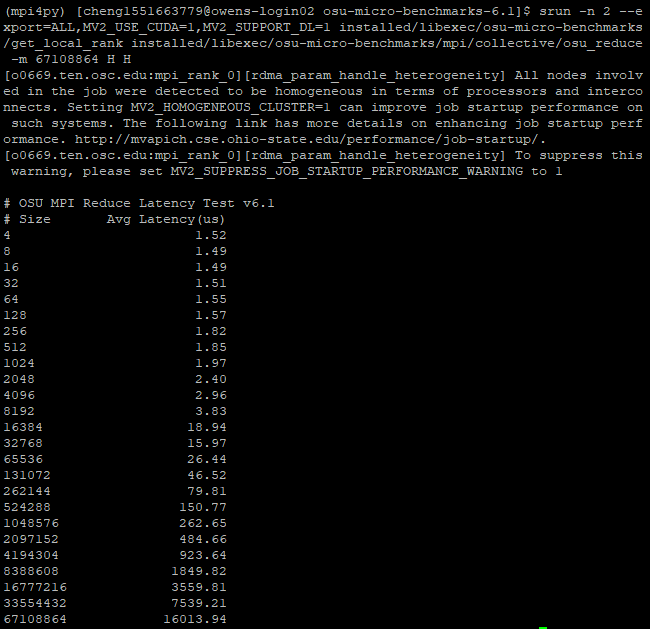
Bibw



Allreduce



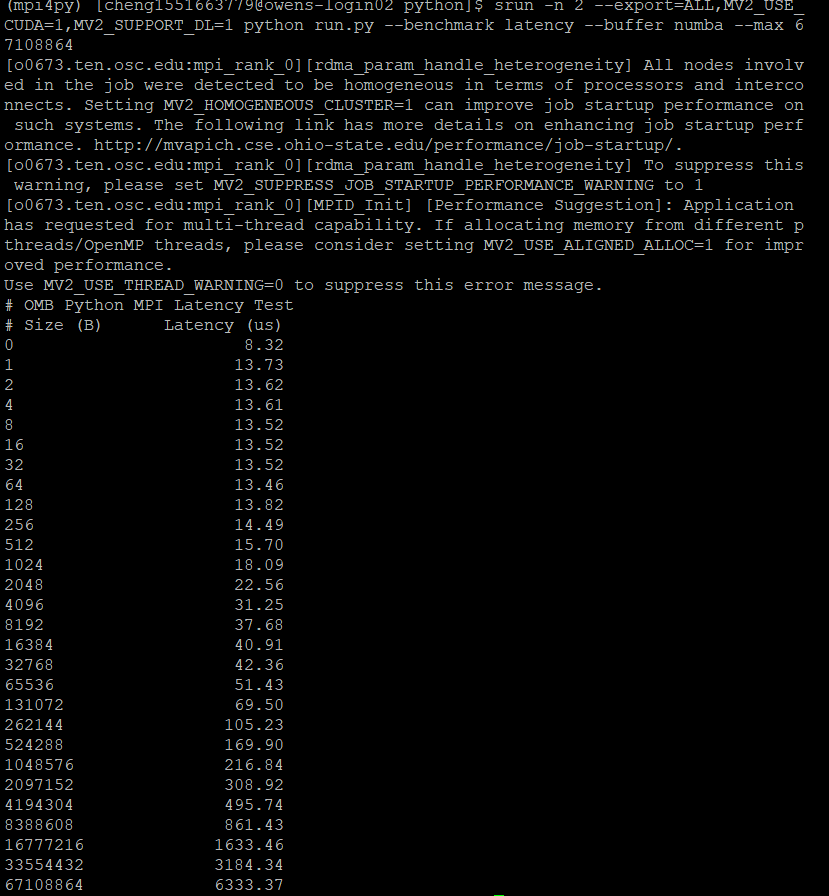
Reduce



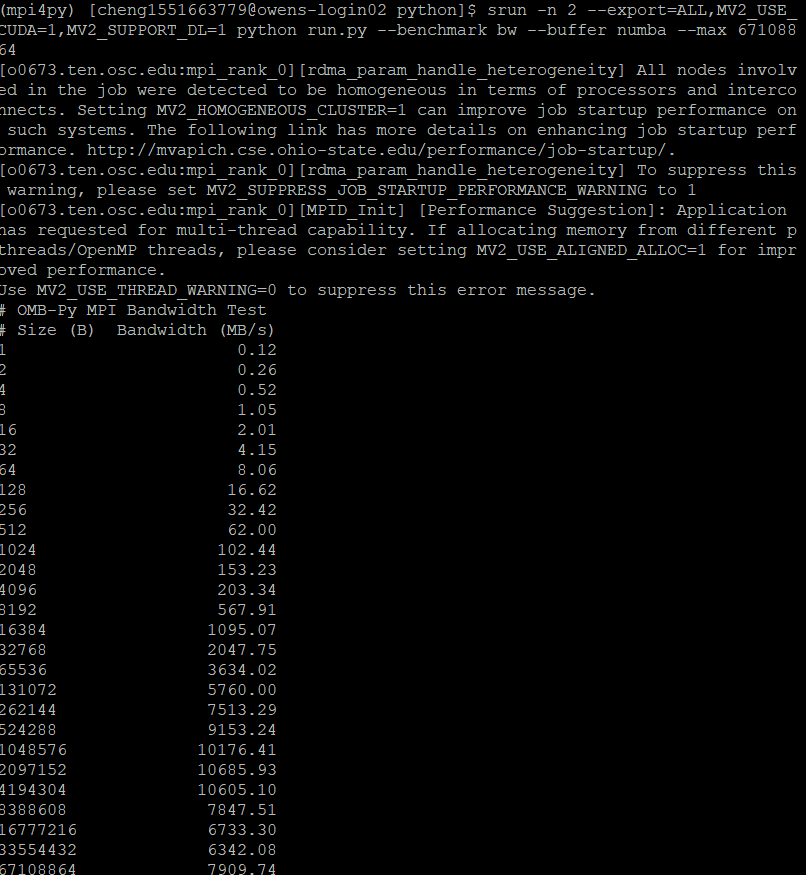
Allocate 2 nodes and run OMB-PY

1.numba

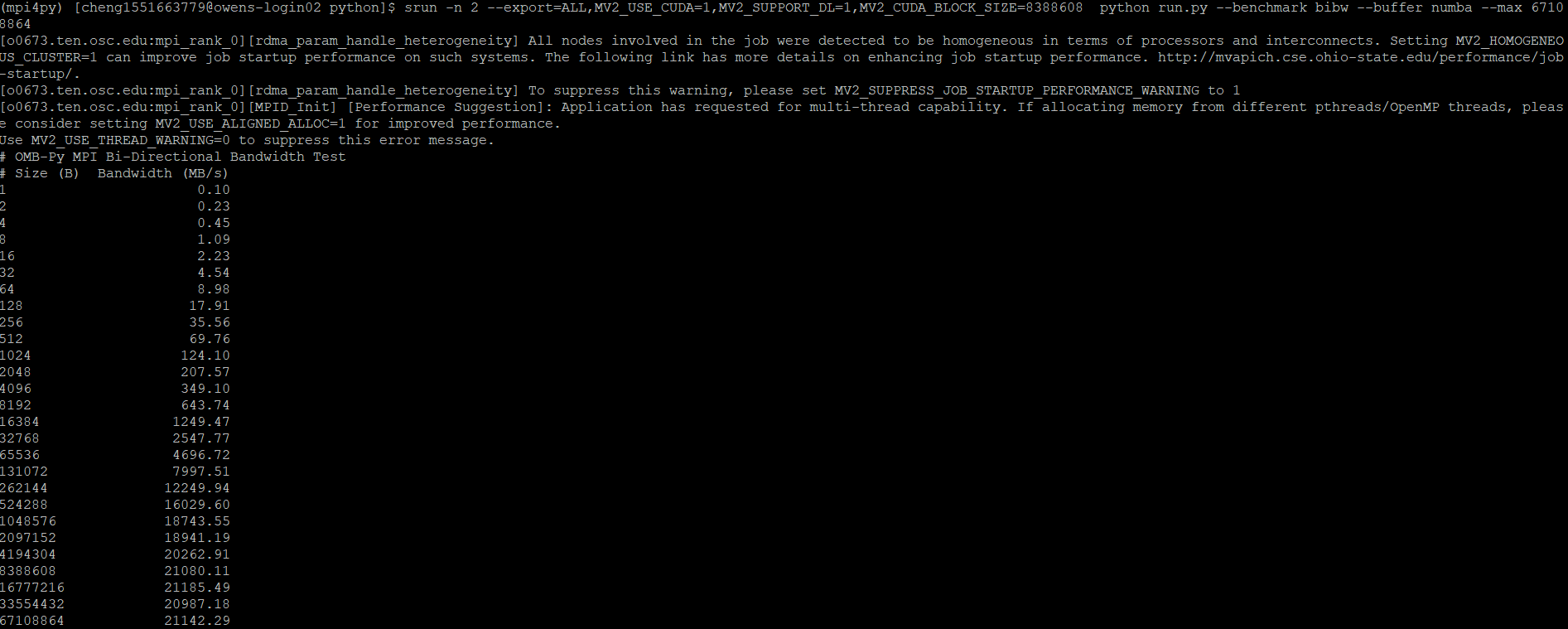
latency



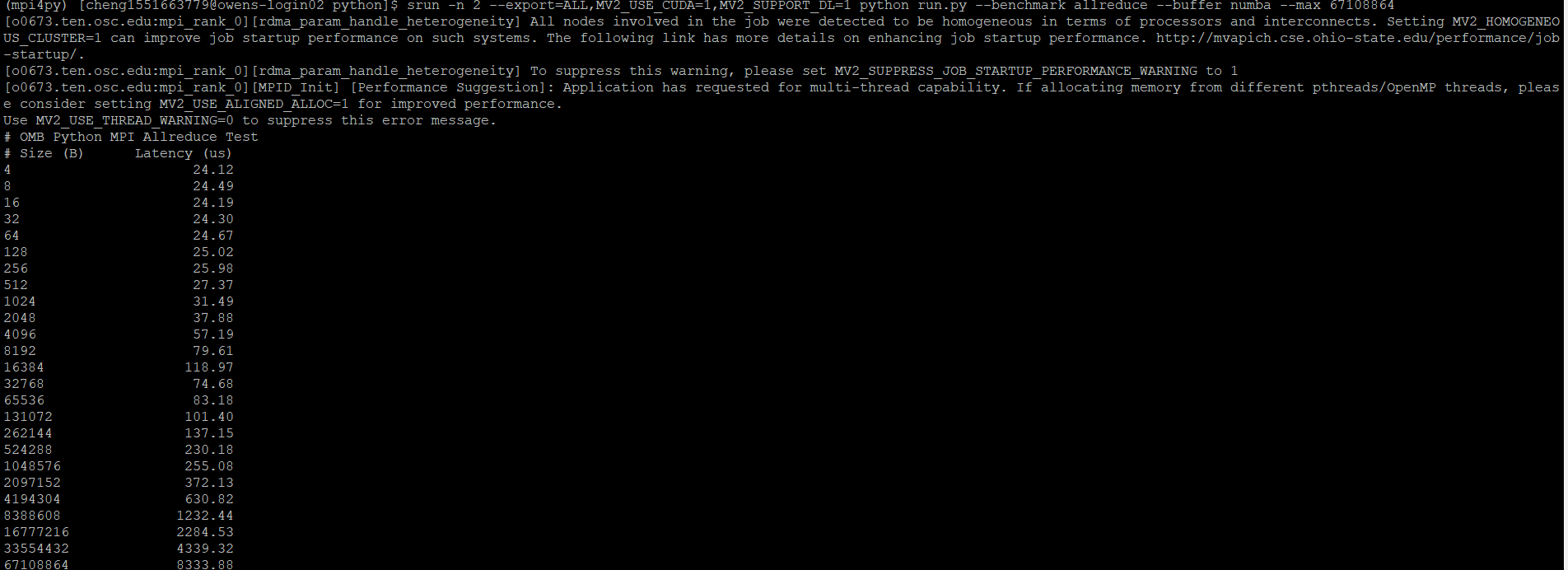
Bw



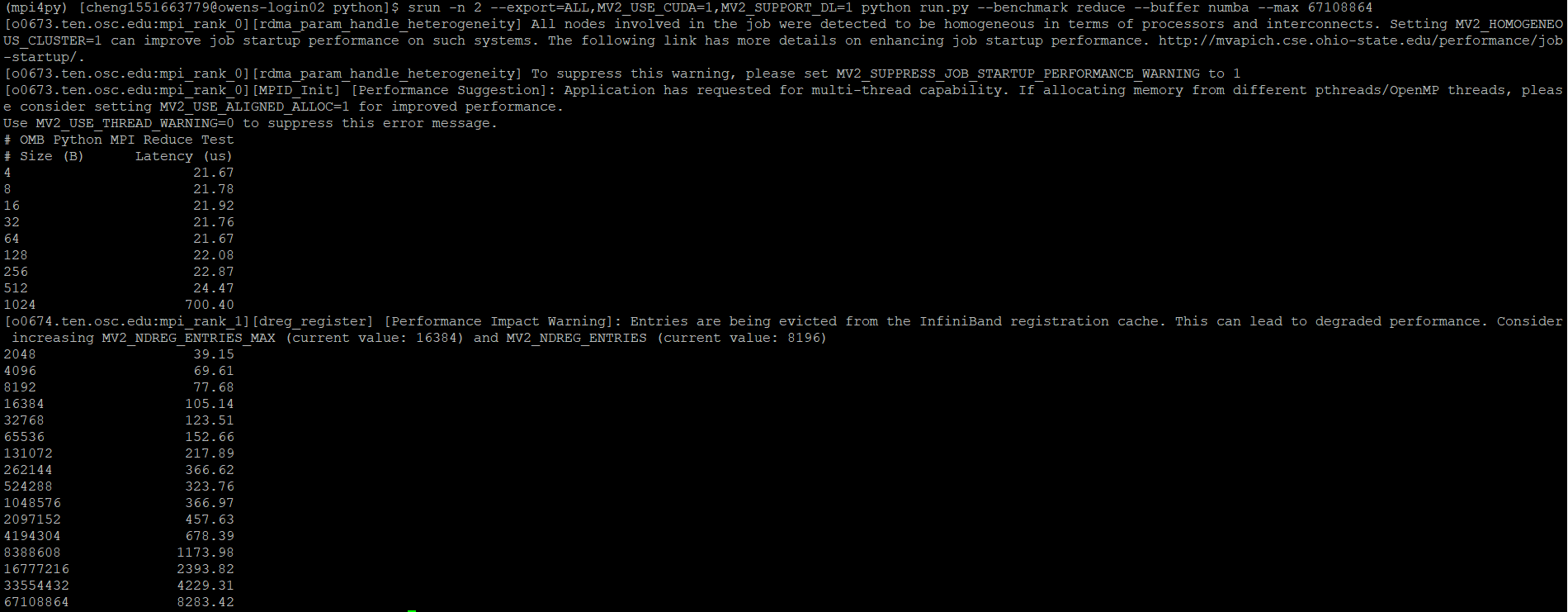
Bibw



Allreduce

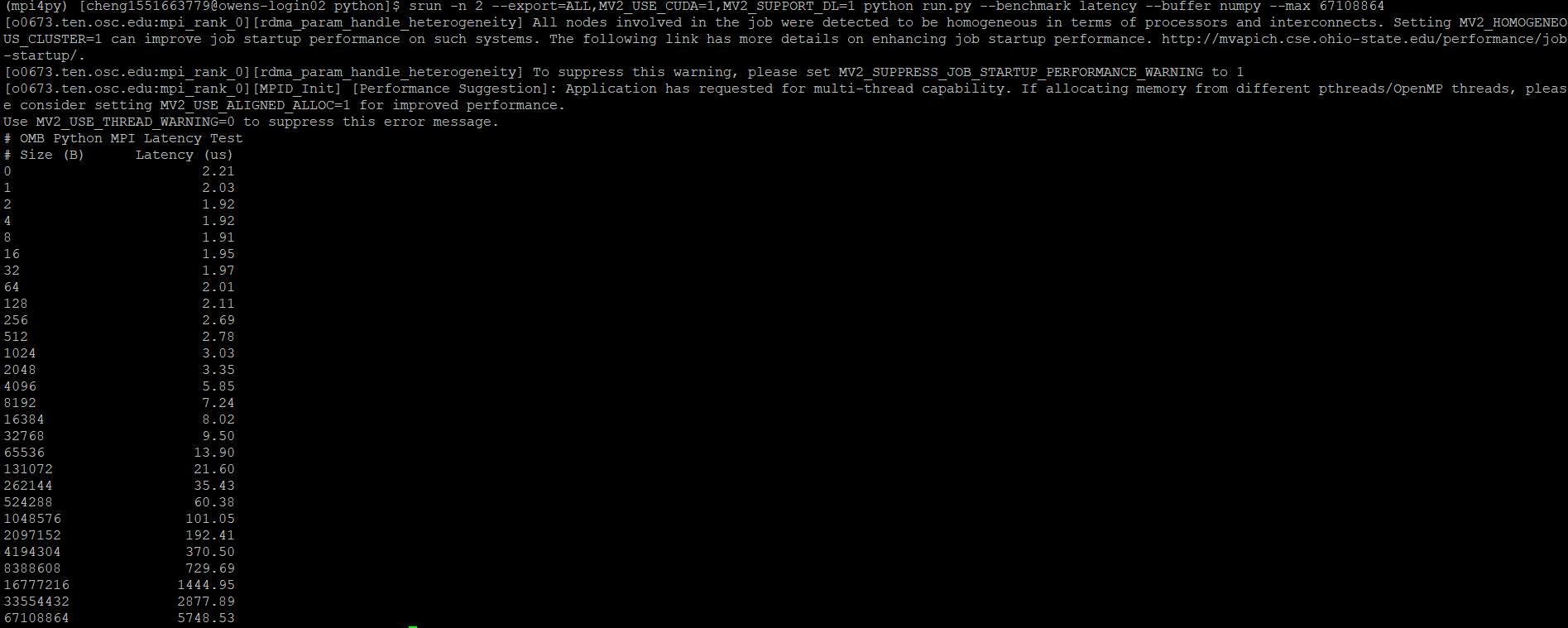


Reduce

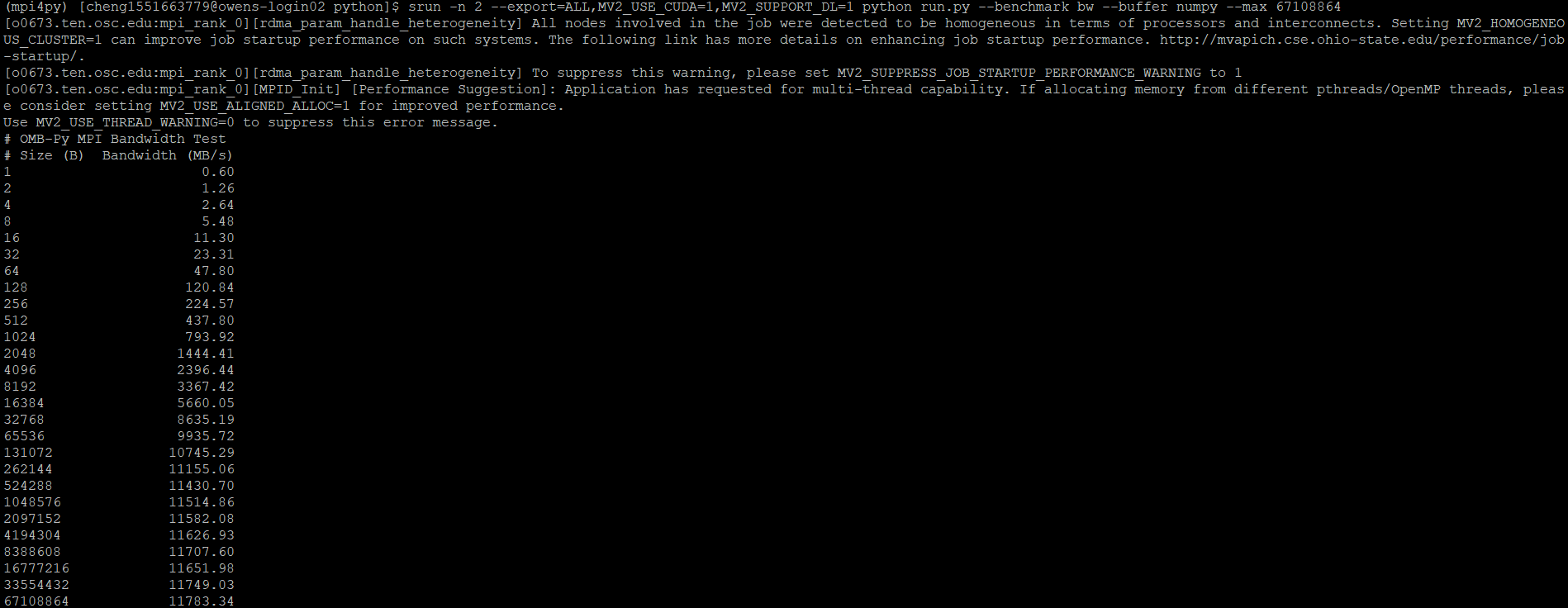


2.numpy

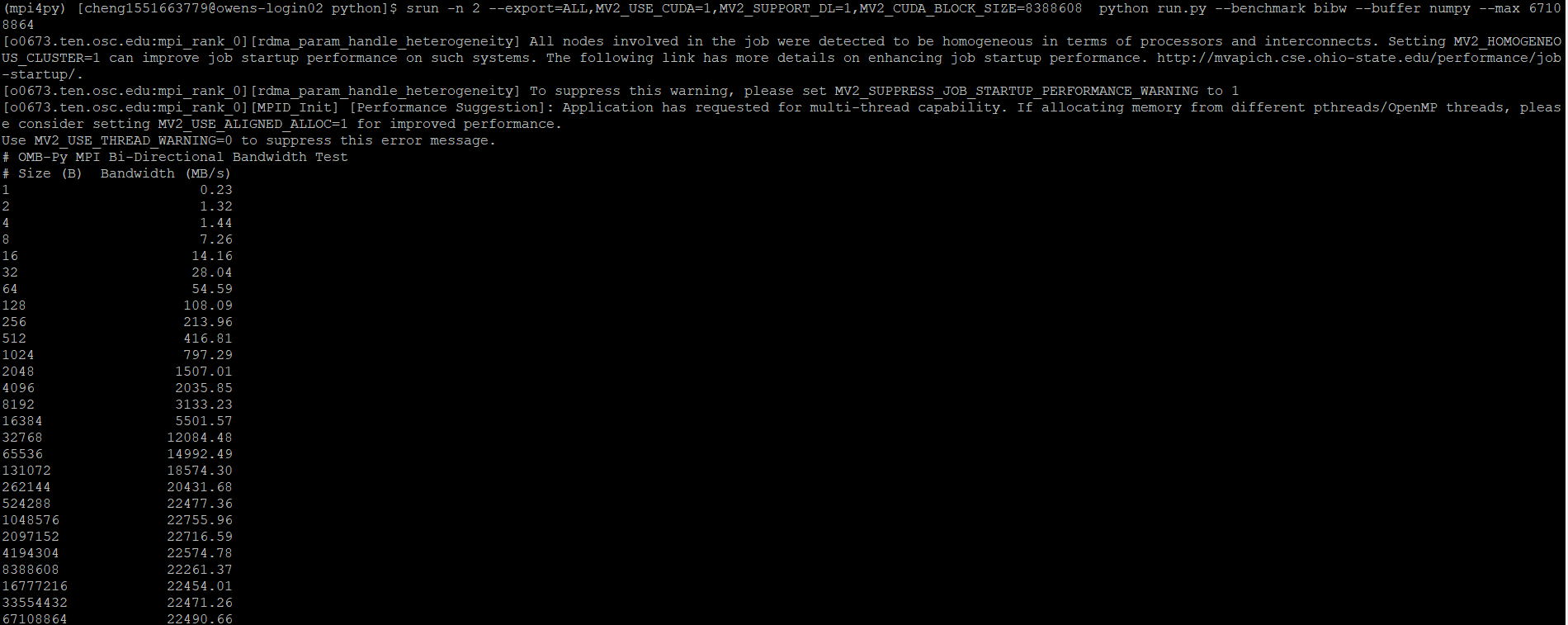
Latency



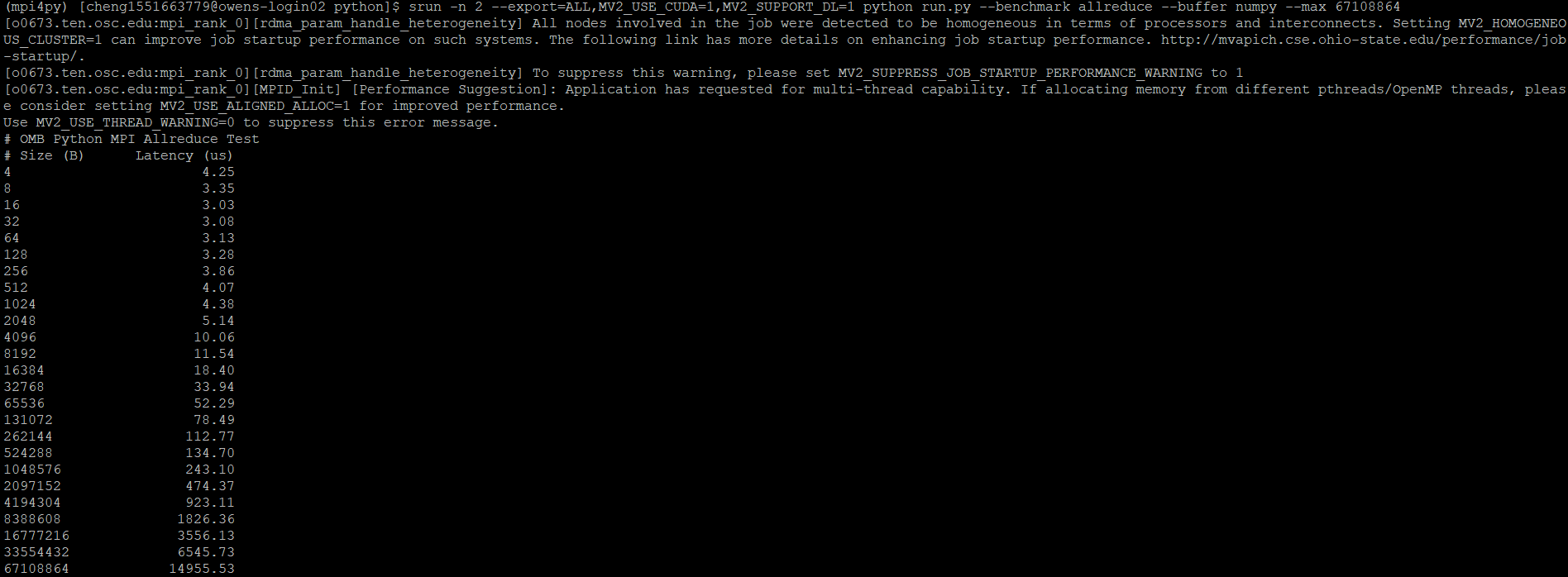
Bw



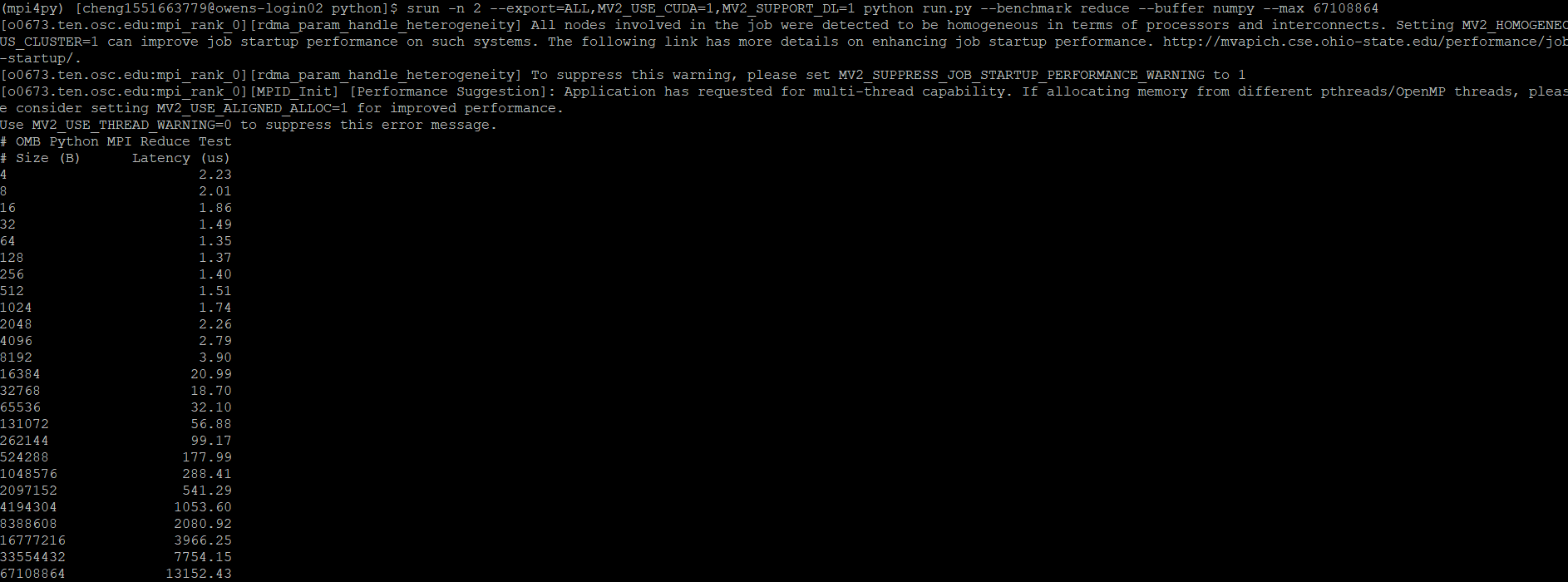
Bibw



Allreduce

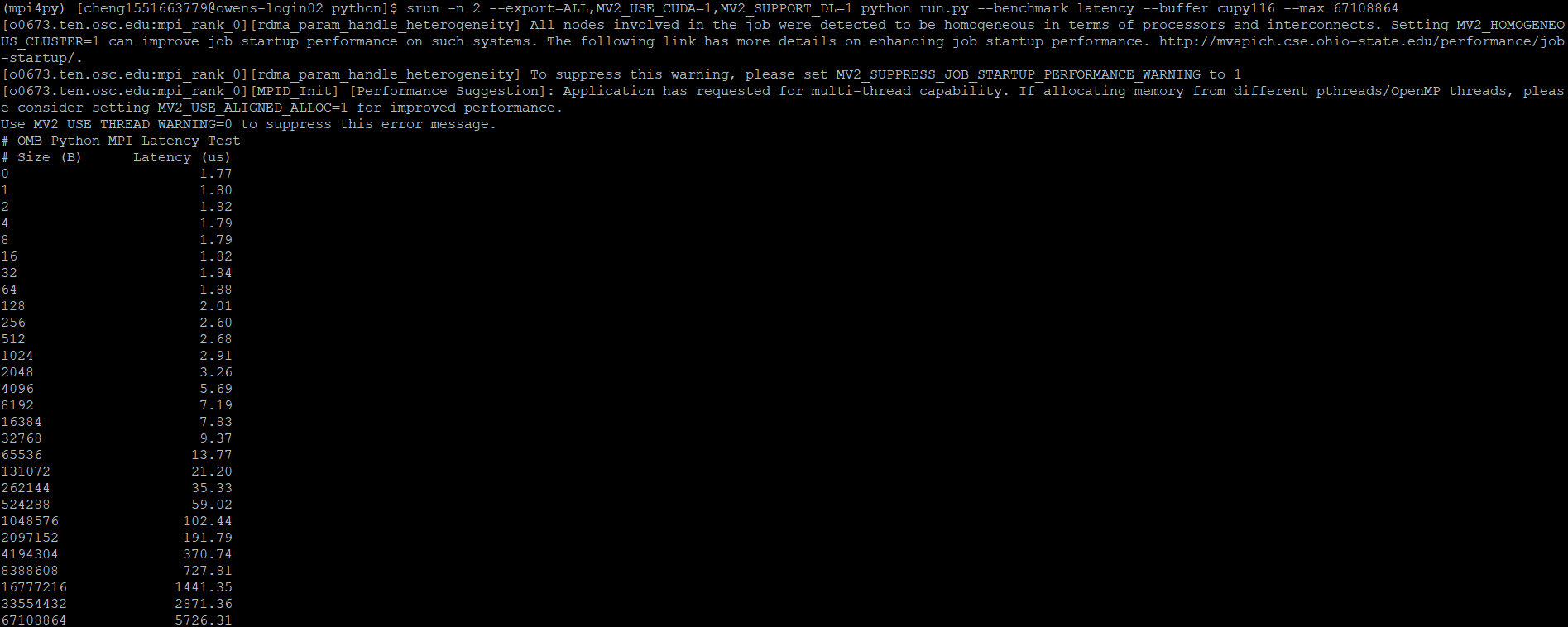


Reduce

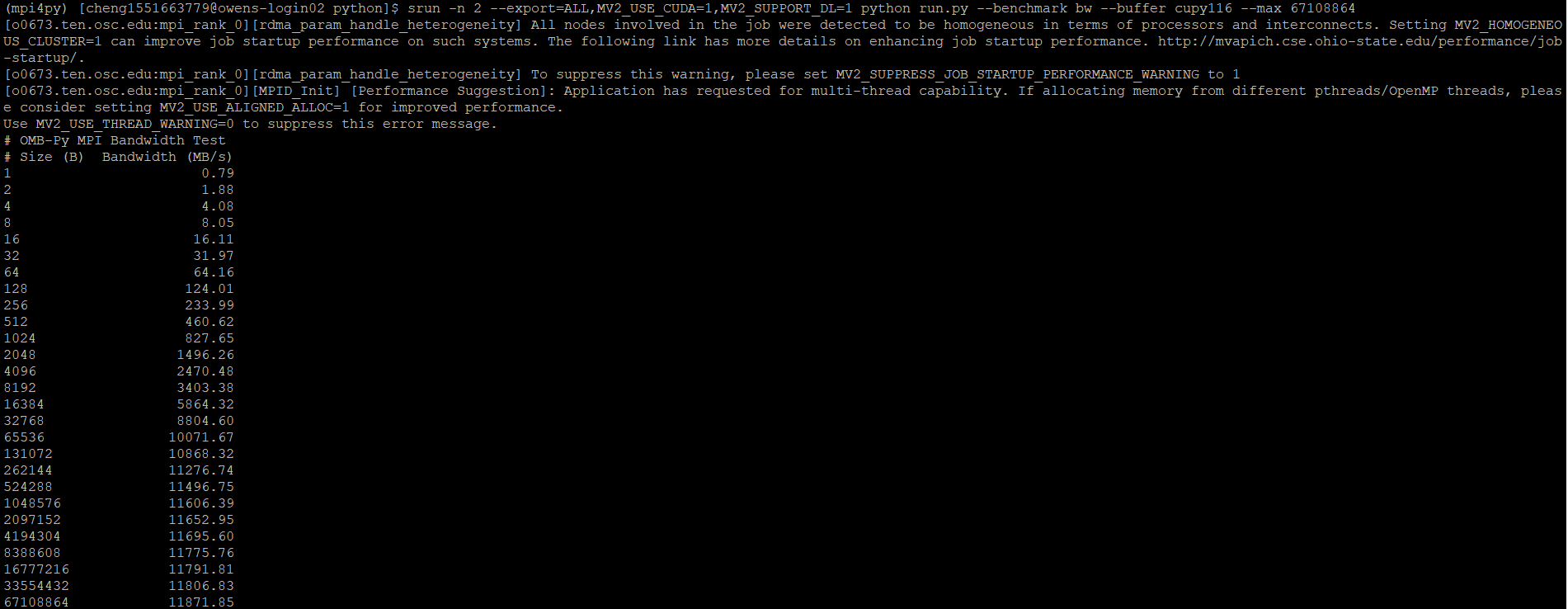


3.cupy

Latency



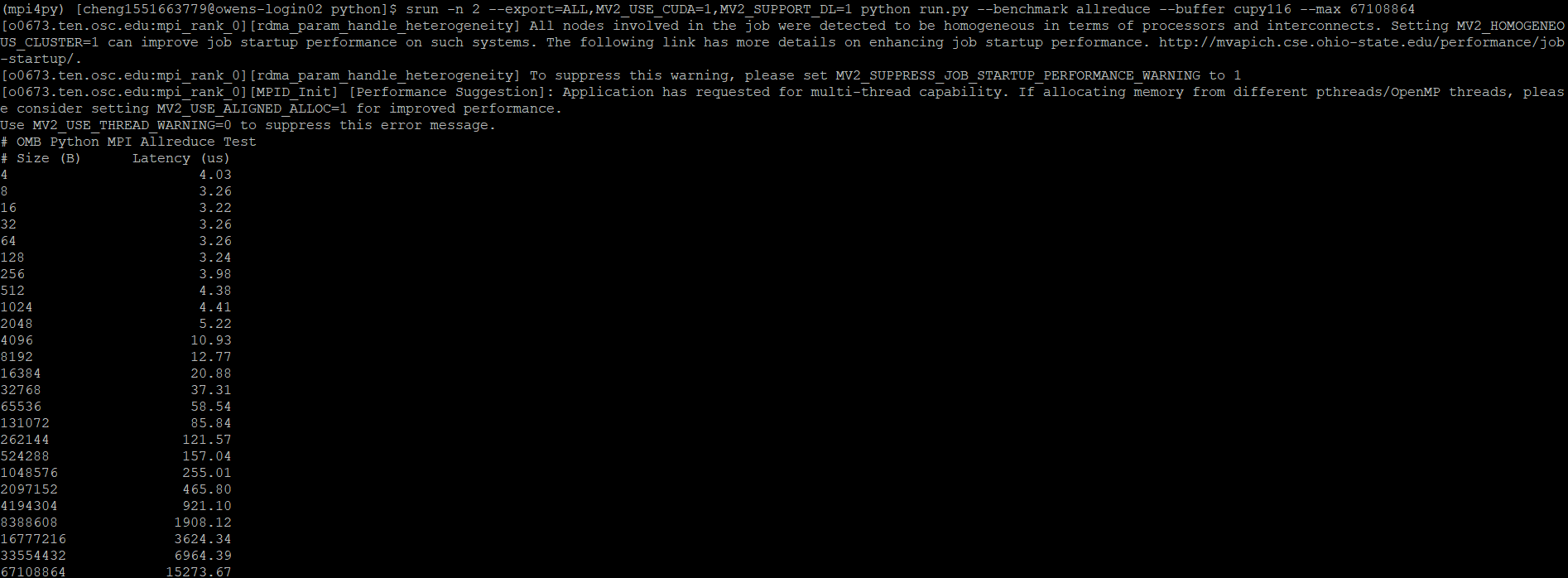
Bw



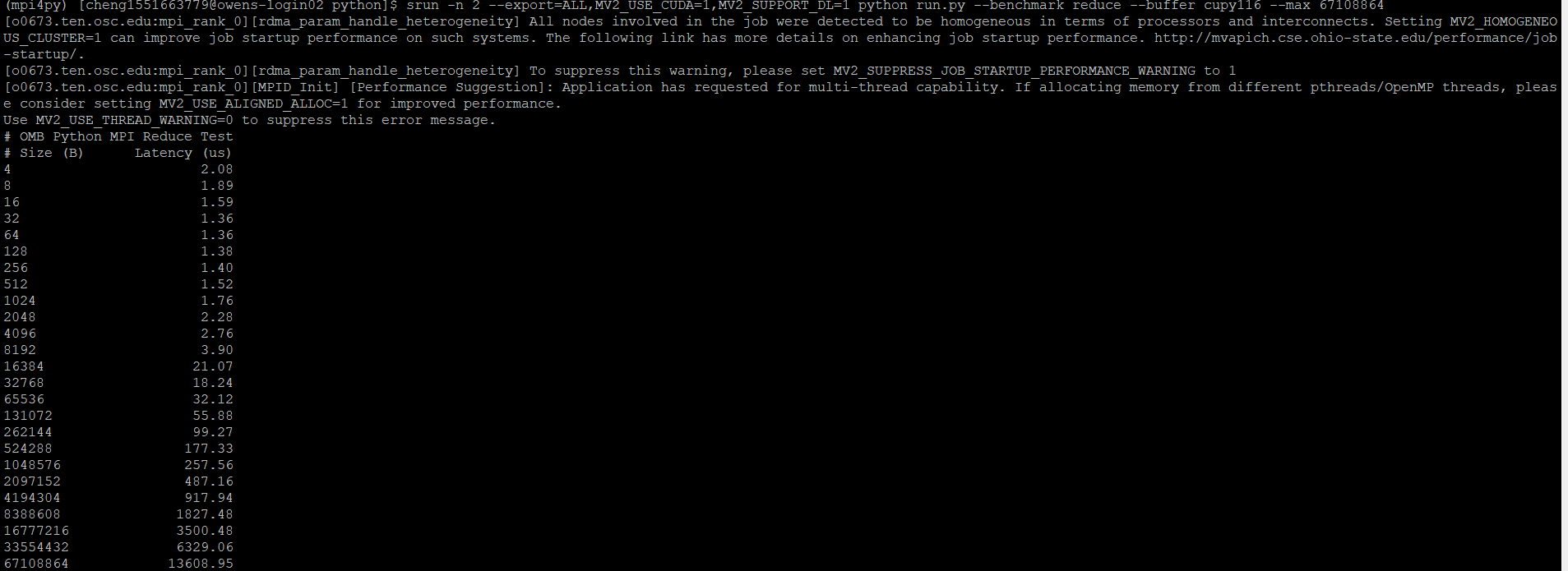
Bibw



Allreduce



Reduce



3. In the point to point, the latency test on CPU is better than GPU, because that CPU is good at process the big data. The bw and bibw test on CPU is better than GPU, the bandwidth of CPU is higher than the bandwidth of GPU. It shows that the CPU could deal with the large data transfer.

In the collective benchmarks on CPU is better than GPU at the small size, as the size becomes larger, the GPU performs better than CPU. Because that GPU could process the collective benchmarks into small pieces, the GPU is good at process the many small parts.

4. In the point to point, the latency, bw, and bibw test on cupy is the best among these three. The numpy would be the second one. The numba would be the wrost.

In the collective benchmarks, the result is totally different with the point to point tests. The numba performs better than other two. The numpy would be the second one. The cupy performance is the wrost.

5.By watching the test between OMB and OMB-PY, I find that the performance on python is relative steady.

In some tests in OMB, the CPU would perfrom better than GPU's. GPU would perfrom better than CPU is some tests. The performance on OMB-py would be about the average of CPU and GPU on OMB. The reason why the performance is about average would be that python could choose one of the CPU and GPU which would perform better.

6.Both allreduce and reduce operations for OMB would be the least latency on GPU. The latency on CPU would be the biggest latency in most cases. The OMB-py would be the middle of these two. The reason is that GPU could easily handle large tests which is consisted of many small cases. The CPU could not handle these tasks easily. OMB-PY would choose the better one in python environment.