

Latency (unit:ms)

0.2

0.1

0.0

1

2

4

8

Number of GPUs



In Memory



Out Memory

