Scaling Across GPU architectures r = 1, n = 358403.5 Base Base-sh 3.0 CAT Millan Topa Cagigas 0.5 H100 A100

(Ampere)

(Hopper)

(Volta)