How much memory per CPU core is requested?

***Alois Schlögl*, Waleed Khalid, Stefano Elefante, and Stephan Stadlbauer**

Institute of Science and Technology Austria (ISTA), Klosterneuburg, Austria

Investments in HPC infrastructure are expensive, and it is important to specify the technical requirements according to the actual need. The amount of available memory per CPU core [1] is one of these crucial parameters. Usage data from the past has been used to compare requested with currently available resources.

The usage data was extracted from the schedulers database (Slurm) for a 10-month time span. During that period, the ISTA scientific compute cluster has been used by 242 users from 47 research groups and about 7.6 Mio Jobs were processed, using 19 million (virtual) core hours .

The requested memory per core was evaluated for each job, and used to compute the cumulative distribution of the workload (i.e. core-hours) of each job. Moreover, the available memory per (virtual) core of each machine was extracted, and the cumulative distribution of the memory (total 76 TB) is obtained. The cumulative distribution for these two metrics are shown in Fig. 1, and the corresponding quantile values are extracted in Table 1.

Table 1: Quantile values of available and used memory [MB] per core.

quantile	available	used
0.10	2584	250
0.25	3815	1000
0.50	3815	3000
0.75	7629	6250
0.90	15625	11000

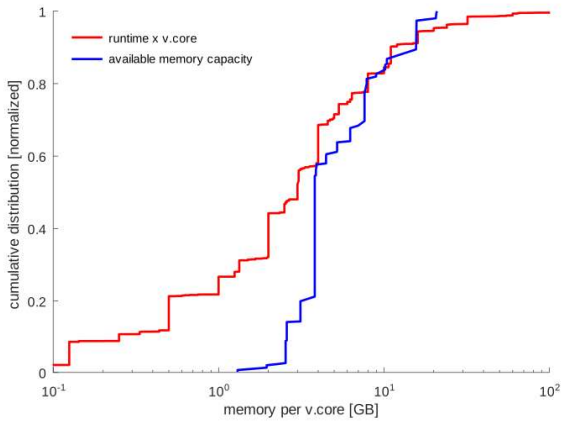

Fig. 1: Cumulative distribution of memory per core.

This HPC cluster is rather heterogeneous with nodes ranging from 1.3 to 20 GB per core. The median of requested memory-per-core is 3 GB per core, with an inter-quartile range between 1 and 6.25 GB. About 5% of the workload would benefit from more memory, and about 25 % of the workload requires 1 GB per core or even less. Overall, the available hardware does match reasonably well to the workload distribution, and will be used for deciding on the technical specification of future extensions.

The actual use and the installed capacity in terms of memory per core have been compared. These results were useful for specifying the requirements of the next extension of the HPC cluster at ISTA.

References

[1] Darko Zivanovic, et al. (2017) Main Memory in HPC: Do We Need More or Could We Live with Less? ACM Transactions on Architecture and Code Optimization, Vol 14(3) pp.1–26.