Results of discussions in 4 rounds „World-Cafe“ about the topic

„Quo vadis climate research? Capacity vs. Capability Computing“ (DKRZ User Workshop / natESM Workshop, 12.10.2022, DKRZ, Hamburg)

Definitions:

“Supercomputers generally aim for the maximum in capability computing rather than capacity computing. Capability computing is typically thought of as using the maximum computing power to solve a single large problem in the shortest amount of time. Often a capability system is able to solve a problem of a size or complexity that no other computer can, e.g. a very complex weather simulation application.” [1] *“The use of the most powerful supercomputers to solve the largest and most demanding problems, in contrast to capacity computing. The main figure of merit in capability computing is time to solution. In capability computing, a system is often dedicated to running one problem.”* [2]	*“Capacity computing, in contrast, is typically thought of as using efficient cost-effective computing power to solve a few somewhat large problems or many small problems. Architectures that lend themselves to supporting many users for routine everyday tasks may have a lot of capacity but are not typically considered supercomputers, given that they do not solve a single very complex problem.”* [1] *“The use of smaller and less expensive high-performance systems to run parallel problems with more modest computational requirements, in contrast to capability computing. The main figure of merit in capacity computing is the cost/performance ratio.”* [2]

Sources: [1] https://en.wikipedia.org/wiki/Supercomputer#Capability_versus_capacity
[2] https://itlaw.fandom.com/wiki/

- All 4 groups agreed on: Both, capacity **and** capability computing are important for future climate research. Not all open scientific questions can be answered by either or. In particular, not all scientific questions can be answered by the envisaged “parameterization-free climate model(s)”.
- The demands of different communities (their various codes!) within the “climate community” (who is this?) are not homogeneous!
- Participants did not agree on the question: what does DKRZ currently provide? Is it a facility for capacity or for capability computing. In a way, both!
- Agreement on: “capability hardware” can also be used for “capacity requirements”, but not necessarily vice versa. For example, one can always use only a part of a “capability machine” to run ensemble simulations in parallel, but if a “capacity machine” does not provide a fast interconnect between **all** nodes (but for instance only within islands), scaling might be limited for “capability” codes, although parallel ensemble simulations are still possible.

- There was no consensus about what is more expensive: a “capability” machine, or a “capacity” machine. A “capability” machine will probably require a faster (more expensive) inter-connect between all nodes, but a “capacity” machine will probably require more (and faster) nodes.
- For costs (e.g. in preparation of new climate research computers), the full picture needs to be considered/discussed, with potential trade-offs:
    - Hardware:
        - compute nodes vs.
        - (fast) network facility vs.
        - working memory per node vs.
        - data storage (disk & tape), vs.
    - staff to support software development towards efficient usage for a variety of codes vs.
    - software tools (e.g. compilers for automatic code optimization and parallelization, etc.); Could/should the climate community “invest” in the development of better tools? Beyond DSLs?
- The focus must not only be on “computing” alone, but also on “data”. Here, the same concept of “capability” vs. “capacity” applies.
- Independent of “capability” or “capacity”: codes must be GPU-enabled, which is a major effort, and there is not sufficient funding for refactoring of codes or the developments of new codes.
- The effort to optimize (develop) codes for capability is way larger (and therefore more expensive) than for capacity!
- Not many groups (codes!) can currently **efficiently** use GPUs, thus there is much capacity but very low capability in the community.
- Moreover, Exa-Scale is much discussed about capability, not so much about capacity. This might underrate the potential of “capacity” solutions (to answer scientific questions!).
- Other technologies might come up: what comes after GPUs? Is there anything on the horizon, such as ARM, FPGA, etc.? Does the climate community focus too much on GPUs?

(Patrick Jöckel, DLR)