bwHPC Governance of the ENM community

Bernd Wiebelt, Dirk von Suchodoletz, and Michael Janczyk

Rechenzentrum, Universität Freiburg

The bwForCluster NEMO is an entry-level (Tier-3) HPC resource for researchers in the state of Baden-Württemberg. It was designed to offer high performance compute resources to the scientific domains of Elementary Particle Physics, Neuroscience and Microsystems Engineering (ENM). The expectations of these heterogeneous user communities have to be respected and tended to. Usage profiles vary from work group to work group. For some users, this is their first encounter with computation beyond the desktop. Other users are more experienced, coming from (smaller) HPC systems of their own. Furthermore, some scientists have extended NEMO with a financial contributions of their own and thus have become shareholders. For the smooth, fair and harmonic operation of the system it is imperative to balance the interests of the various scientific communities and shareholders. This is accomplished by offering the opportunity to participate on all levels of the decision making processes and to jointly moderate and develop the operating model. To this end, two consulting governance bodies have been established. On the one hand, there is the *NEMO Nutzerversammlung* (general user assembly), meeting at least once per year to gather feedback from all users. On the other hand, there is the *NEMO Cluster-Beirat* (advisory board) composed of representatives from the scientific communities and the shareholders. The advisory board meets more often and can therefore discuss and advise more timely on current topics.

1 Managing a shared resource

The bwForCluster NEMO and the corresponding HPC competence center ENM are located in Freiburg, offering their services to all scientists in Baden-Württemberg working in the scientific fields of Elementary Particle Physics, Neuroscience and Microsystems Engineering (ENM). NEMO was financed by the *Ministry of Science, Research and the Arts Baden-Württemberg* and the *German Research Foundation*. NEMO was procured with the intention of providing researchers from these scientific fields an entry-level (Tier-3) HPC resource which gives them the opportunity to start progressing beyond the limits of desktop-bound computing. The acquisition and deployment of the NEMO hardware was accompanied by the installation of the HPC competence center ENM in the framework of the bwHPC-C5 project [1]. The competence center ENM offers service and support to further the scientists' efforts in adapting their applications to HPC environments.

During the application phase, scientists from the ENM communities have supplemented NEMO with financial resources of their own, consequently becoming shareholders of the final system. Serving three scientific communities and the shareholders, the user community of the bwForCluster NEMO has become large and heterogeneous. This heterogeneous community does not necessarily have coherent objectives with respect to the initial setup of an HPC system. Therefore, during the procurement phase, there were already lots of interactions and consultations between all involved parties to balance these interests and find a joint mode of operation. This cooperation has continued since NEMO became operational. The governance structures outlined herein aim to formalize those already established and well tested ways for acting cooperatively and towards a common goal.

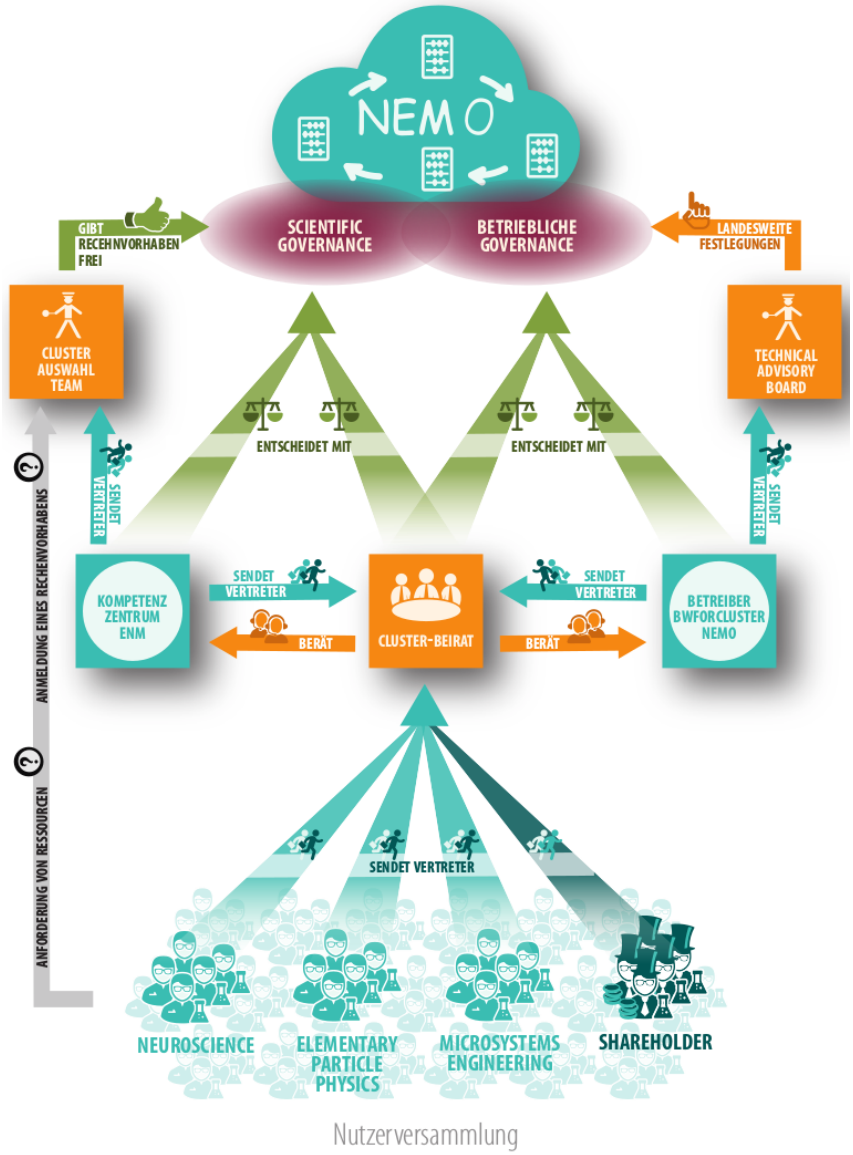

Figure 1: bwForCluster NEMO governance [3].

The bwHPC initiative has dedicated superordinate governance bodies that specify the basic

policies that bwHPC resources – such as the bwForCluster NEMO – have to adhere to [2]. However, avoiding micro-management, these policies are deliberately not chosen to encompass each and every parameter of operation. Thus, there are decisions which have to be taken by the operators of a bwHPC resource in accordance with their respective user communities.

For the bwForCluster NEMO, a twofold approach was chosen [3] based on the experience acquired during the procurement phase of NEMO. Two governance bodies have been created, with either taking a designated role for the gathering of community feedback and getting recommendations for taking scientific and technical decisions:

- The NEMO general user assembly ("NEMO Nutzerversammlung") is taking place biannually. The event is widely announced on various websites and mailing lists and all current and prospective NEMO users are invited to participate. The aim is to report on the current status, gather broad feedback coming from the actual user base and inform on current and future developments with respect to NEMO. Common ideas and problems from users can be addressed and discussed at this meeting.
- The NEMO advisory board ("NEMO Cluster-Beirat") is a committee composed of representatives of NEMO's scientific communities and its shareholders. It is complementary to the general user assembly in the sense that the board's meetings happen more often and can be convoked on demand. It is thus able to give advice if immediate or short-term decisions have to be taken. Additionally, the board members serve as contact partners for their respective communities. In this way, common issues can be consolidated and addressed in a concerted fashion, instead of dealing with each individually. Last but not least, it is the task of the advisory board to reconcile the diverging interests of the affected scientific communities and shareholders.

2 Fair-share and Quality of Service

A supercomputer such as the bwForCluster NEMO enforces a time sharing policy, meaning there is no individual user who has exclusive and unlimited access to all provided resources at all times. Instead, the user specifies the required resources (CPU cores, RAM, etc.) in a job description file which also states for how long the said resources will be needed. A scheduling software monitors the available computing resources and executes the job as soon as the requirements can be met.

In the default mode of operation, each user has an equal chance to allocate resources for his jobs. Once resources get scarce, the decision for resource allocation is based on a fair-share heuristic, which takes the usage history into account. In principle, this means that chances to allocate resources are higher for users that have previously not used the system as intensely as other users. The fair-share mechanism thus balances the concurrent demands of all users based on previous usage.

Compute resources (i.e. processor clock cycles) are highly volatile. At any given time, they are either used or not used. Unused compute resources are lost. Therefore, incrementation of fair-share priority ("saving of compute resources") can only be allowed up to a certain threshold. Otherwise, users who have not been using the system for an extended period of time would be able to monopolize and block the system for other users at a later time. To prevent this, fair share priorities will be reduced by a decay function over time.

Beyond the default fair-share mode of operation, it is possible to provide more fine grained usage rights, going so far as to give de facto exclusive rights to selected users on parts of the

HPC resource. This can be accomplished in multiple technical ways described by quality of service policies. In general, shareholders would be given a fixed percentage of compute resources representing their share. The interesting question is what happens with shareholder compute cycles that are not used. The following strategies are possible, including mixes and variations:

- Do not offer unused shareholder compute cycles to other users. Unfortunately, this is also the least efficient policy, wasting precious resources.
- Define a quality of service policy with guaranteed availability windows via rollback reservations. For example, a shareholder QoS policy could guarantee availability of 10% of the shareholder part in 1 hour, 50% of the share in 4 hours and 100% of the share in 24 hours.
- Guarantee 100% of the shareholder resources immediately. However, the resources are part of a preemptive scheduler queue, which can be used by other users with a zero QoS guarantee, meaning that jobs in this queue are not guaranteed to finish at all. This would still be useful for backfilling with smaller jobs, provided that experienced users can cope with prematurely terminated jobs by having a solid checkpointing strategy.

In the last two cases, the shareholder will be compensated by the right to use more resources than his share allows at times when there are a free compute cycles available.

3 Conclusions

The NEMO governance structures follow common practices established in research institutions. The first general user assembly was held on August 14th 2016 and the first NEMO advisory board meeting took place on December 14th 2016. Both events provided valuable feedback and resulted in several operational adjustments. Fair-share and quality of service have been approved as technical instruments to balance the competing interests of the three scientific communities and shareholders.

References

[1] http://www.bwhpc-c5.de

[2] Wesner, S., Walter, T., Wiebelt, B., von Suchodoletz, D., Schneider, G. “Strukturen und Gremien einer bwHPC-Governance – Momentaufnahmen und Perspektiven.” in Kooperation von Rechenzentren Governance und Steuerung – Organisation, Rechtsgrundlagen, Politik, de Gruyter (2016): 315–329.

[3] Wiebelt, B., Janczyk, M., von Suchodoletz, D., Aertsen, A., Rotter, S., Schuhmacher, M., Greiner, A., Quast, G. “Strukturvorschlag für eine bwHPC-Governance der ENM-Community.” in Kooperation von Rechenzentren Governance und Steuerung – Organisation, Rechtsgrundlagen, Politik, de Gruyter (2016): 343–354.