bw|HPC

Baden-Württemberg
MINISTERIUM FÜR WISSENSCHAFT, FORSCHUNG UND KUNST

DFG Deutsche Forschungsgemeinschaft

NEMO bwForCluster

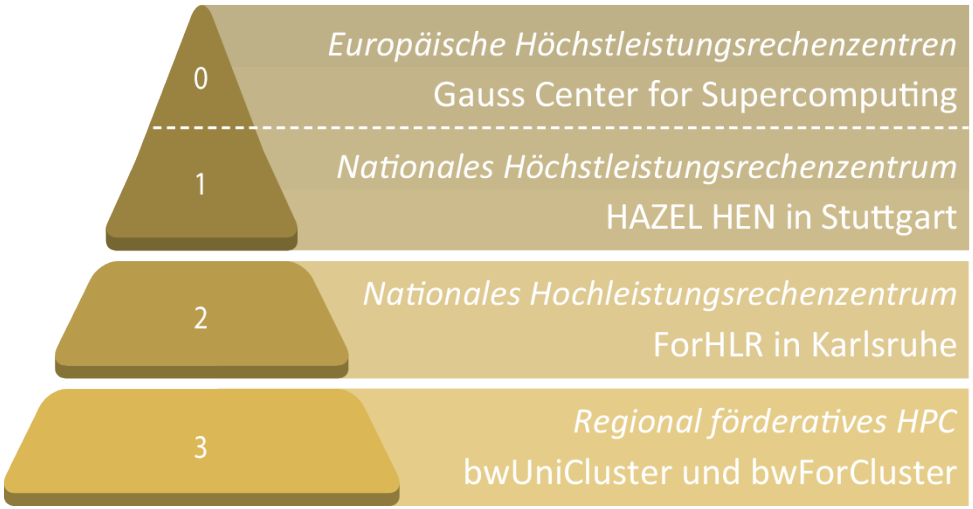
Journées SUCCES

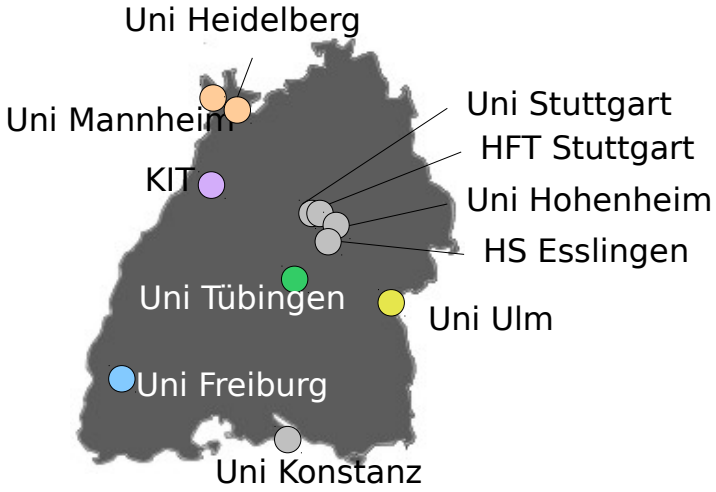
24/11/2016, Paris

When HPC meets CLOUD

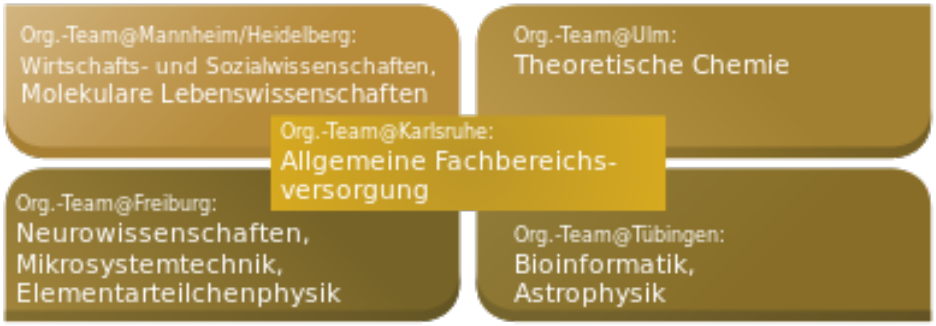
Bernd Wiebelt – eScience Department - University of Freiburg, Germany

Introducing bwHPC





User Support



(*) https://mwk.baden-wuerttemberg.de/fileadmin/redaktion/m-mwk/intern/dateien/pdf/Forschung/Umsetzungskonzept_bwHPC.pdf

The Workstation is not enough

- Scientific workloads transcend the capabilities and capacities of traditional workstations and desktop computers
- This is an irreversible ongoing progress, future scientists need to be prepared to look beyond the desktop: Think distributed, think parallel
- Computational resources becoming increasingly accessible for all scientific users
  - High Performance Computing Resources
  - Compute Grids
  - Compute Clouds

Question: What is the **best** resource?

Welcome to the MATRIX

Use Case	HPC	Cloud	Grid
Massively parallel Simulation	✔	✘	✘
Embarassingly parallel Analysis	✔	✔	✔
Heavy Input/Output	✔	✘	?
Large Storage Needs	?	?	✔
Flexible Software Environment	✘	✔	✘
Opportunistic Resource	✘	✔	?
Special Hardware (GPGPU, MIC)	✔	?	✘
Easy accessible for beginners	✘	✔	✘

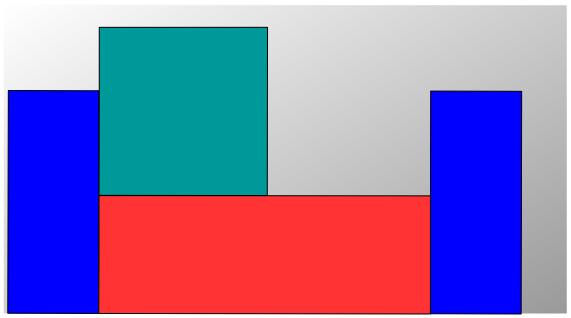
Look, a Supercomputer!

Look, a Cloud!

Look, a Grid!

Capability versus Capacity

Capability Computing (HPC)	Capacity Computing (GRID, CLOUD)
Solve single large problems	Solve many small problems
Massively parallel	Embarrassingly parallel
Optimize for performance (TFlops, Memory, I/O)	Optimize for throughput (#Jobs done/sec)
Special Hardware: GPGPU, MIC, ExaScale	Commodity (Server) Hardware: General Purpose CPU
Motto: „Scale up“	Motto: „Scale out“



Capa*-Computing on NEMO

de.NBI
BioInformatics
Docker Containers
Galaxy Frontend

LHC/CMS
Users
Virtual Research
Environment

Theoretical
Physics
High Memory
per Core

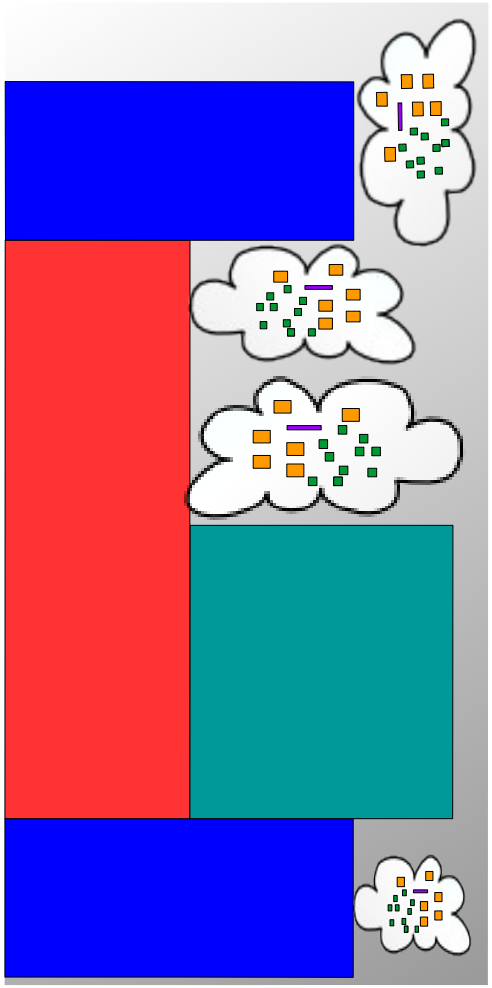
LHC/ATLAS
Users
Complex Software
Environment

Neuroscience
Simulations + Data Analysis
High Performance
Filesystem / Network

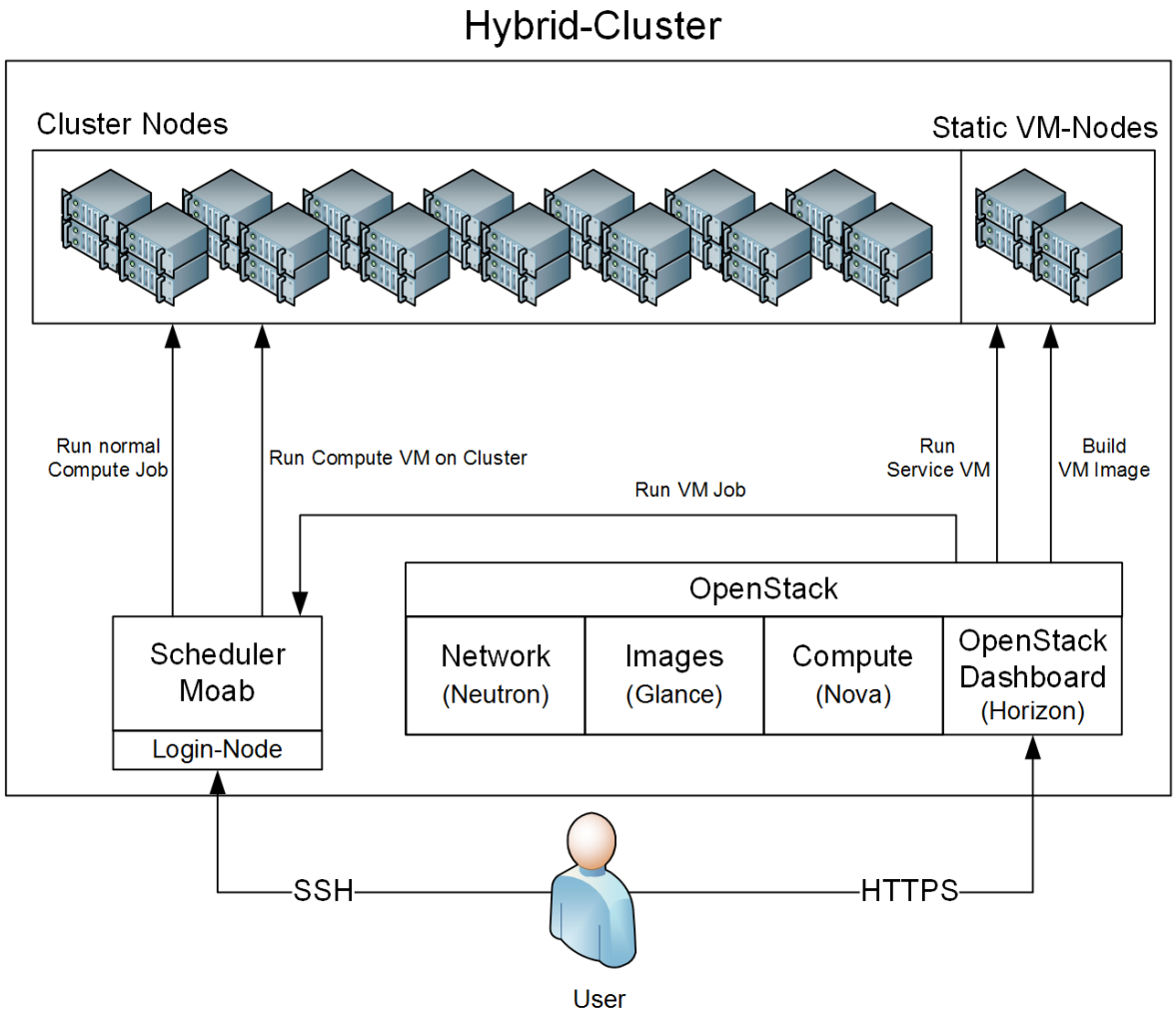
Microsystems Engineering
MPI Performance (e.g. LAMMPS)

Capa*-Computing on NEMO

- Capacity Computing uses HPC building blocks
- High Performance Network is useful for both
  - MPI for massively parallel communication
  - Fast parallel storage for large number of small jobs
- MPI is sometimes just a convenient scheduler
  - For synchronization
  - For set-up/clean-up phase
  - MPI-tasks might be embarrassingly parallel



Capa*-Computing on NEMO



Source: Konrad Meier, eScience, Uni Freiburg

Virtual Research Environments

- Tailored to the scientific application
- Customizable by the users:
  - VRE per scientist
  - VRE per scientific workgroup
  - VRE per scientific field
- Enables reproducibility of results
- Abstraction from underlying hardware
- Enables versioning of research environments
- Allows citation and referencing of software methods
- Requires Open Data
- Requires Workflow Management
- Requires Data Management Plan

Conclusion & Outlook

- Cloud and Grid workloads add value to HPC systems
- HPC, Cloud and Grid should be partners, not competitors
- Virtual Research Environments are a key technology to establish sustainable scientific workflows
- Data Management Plans are obvious companions to Virtual Research Environments

Capa*-Computing
Connecting standard components in
(not yet) standard ways