

# “Cloud computing from a HPC provider perspective”

Lutz Schubert HLRS Stuttgart, Germany  
schubert@hlrs.de

## Abstract:

Cloud computing is considered a major commercial innovation of the recent years and the prognosis for further developments in this area indicates increasing interest from both commercial and academic side. However, there is little agreement over what actually makes a cloud and interpretations range from "the internet" via "shared resources" to "specialised efficient execution platforms". This divergence is originating from the broad scope of commercial offerings claiming to provide "cloud" capabilities, without specifying details.

Recently, many working groups have therefore tried to specify the term "cloud" with respect to its relevance in the according domains more clearly. Whilst this does not necessarily reduce the scope of interpretations, it fixes and binds them more clearly to the usage area. Notably, there are strong overlaps regarding specific capabilities which crystallize to be the core cloud features, such as elasticity and availability and which shall be presented in more detail. Taken by themselves, none of these features do actually pose new technological requirements or do justify the claim that clouds offer innovation over previous systems. Instead, the main innovation to be associated with clouds consists in the commercial sale of the added resource capabilities to external customers, as opposed to internal purposes only. This signifies an important step in "utility computing" as envisaged by John McCarthy back in 1961.

Accordingly, a comparison to grid computing as a means to grant access to remote resources as if they would be locally available, comes to mind. Along that line, first speculations of high performance computing (HPC) clouds have already been undertaken, which thus should grant easy access to strong computational resources in a similar fashion to EGEE. Mostly, this is associated with the types of machines exposed via Amazon EC2 which offer multi-core and thus parallel systems. It must be noted though that even machines with 8 (virtual!) cores à 2.5-3.0 GHz machines are far from what is considered a "High Performance Computer" and the implications from applying cloud concepts to this domain differ strongly from those of server farms, such as the EC2. In particular the main interests in cloud systems, such as elasticity and availability, do not compare to the main interests in HPC systems, which focus on reliability and performance. There is however a strong conceptual overlap between the concepts of elasticity (cloud) and scalability (HPC), as well as reliability (HPC) and availability (cloud) which typically do not relate to the same technical aspects. Nonetheless, the some cloud specific capabilities bring in interesting new challenges to HPC provisioning, in particular related to ease of access and dynamicity, respectively adaptability, which reflect future customer needs and demands. The main issue therefore consists in distinguishing essential cloud versus intrinsic HPC features and examining their feasibility, respectively technical implications. The presentation will elaborate further what makes a cloud, basing on current understandings and how (and in how far) this relates to the specifics of high performance computing provisioning, respectively why the typical cloud concepts do not easily allow for the realisation of what is commonly called a "HPC cloud". Along this line, new developments and trends will be discussed, that bring cloud like capabilities to HPC provisioning.

The High Performance Computing Centre (HLRS) is a research and service institution affiliated to the University of Stuttgart (USTUTT). It is one of the three national supercomputing centres in Germany and one of the three members of the GAUSS centre for supercomputing and is offering services to academic users and industry. USTUTT-HLRS research and operation focus is oriented towards the operation of supercomputers owned by hww GmbH (a public-private partnership consortium comprising among others T-Systems SfR and the Porsche AG), services and consultancy for scientific and industrial users, research in the area of supercomputer architectures, simulation software, software engineering and Grid computing, teaching and training in distributed systems, software engineering and programming models, as well as cooperation with international partners from industry and research. The High Performance Computing Center Stuttgart (HLRS) of the University of Stuttgart supports researchers and industry with leading edge supercomputing technology. Services are supplied in collaboration with scientific and industrial partners through hkz-bw and hww GmbH (T-Systems, T-Systems SfR GmbH and Porsche AG). In European, national, and industrial projects HLRS conducts basic and applied research in HPC together with partners from research and industry. Collaborative research with automotive industry goes through the Automotive Simulation Center Stuttgart (ASCS). Recently HLRS has investigated more into the specifics of compute resource provisioning over "clouds". In particular the AMT (Applications, Models & Tools) department of Stefan Wesner assessed the possibility to grant access to HPC resources in a fashion similar to cloud providers; the ISIS (Intelligent Service Infrastructures) department of Lutz Schubert has been investigating cloud-like capabilities across horizontal stacks for some time now [3][4].

Recent projects at HLRS have examined the integration of public cloud providers, such as Amazon EC2 to dynamically extend computation resources (BREIN [5][6]); investigate into cloud-like capabilities for reducing the carbon footprint (GAMES [7]); and design new operating systems for highly-scalable and dynamic environments (S(o)OS [1][2][8]). Lutz Schubert was also invited expert and rapporteur to the Cloud Computing expert working group initiated by the European Commission [9].

1. L. Schubert, A. Kipp, and S. Wesner, "Above the Clouds: From Grids to Service-oriented Operating Systems", Towards the Future Internet – A European Research Perspective, G. Tselentis, J. Domingue, A. Galis, A. Gavras, D. Hausheer, S. Krco, V. Lotz, and T. Zahariadis, Amsterdam: IOS Press, 2009, pp. 238-249.
2. L. Schubert, S. Wesner, A. Kipp, and A. Arenas, "Self-Managed Microkernels: From Clouds Towards Resource Fabrics", In Proceedings of the First International Conference on Cloud Computing, Munich: 2009.
3. A. Kipp, L. Schubert, and C. Geuer-Pollmann, "Dynamic Service Encapsulation", Proceedings of the First International Conference on Cloud Computing, Munich: 2009.
4. L. Schubert and A. Kipp, "Principles of Service Oriented Operating Systems," Networks for Grid Applications, Second International Conference, GridNets 2008, P. Vicat-Blanc Primet, T. Kudoh, and J. Mambretti, Springer, 2009, pp. 56-69.
5. S. Taylor, G. Laria, P. Ritovato, and L. Schubert, "Business Collaborations in Grids: The BREIN Architectural Principals and VO Model", Grid Economics and Business Models, J. Altmann, R. Buyya, and O.F. Rana, Berlin / Heidelberg: Springer, 2009, pp. 171-181.
6. BREIN: Business objective driven Reliable and Intelligent grids for real business, <http://www.eu-brein.com/>
7. GAMES: Green Active Management of Energy in IT Service centres
8. S(o)OS: Service-oriented Operating Systems – <http://www.soos-project.eu>
9. L. Schubert, K. Jeffery, and B. Neidecker-Lutz (eds), "The Future of Cloud Computing" – available at <http://cordis.europa.eu/fp7/ict/ssai/docs/cloud-report-final.pdf>