CoreGRID and Clouds - Future Perspectives

Ramin Yahyapour

Technische Universität Dortmund,
IT and Media Center,
44227 Dortmund, Germany

Abstract. Since 2004 the research interests of the CoreGRID community has evolved from distributed large scale computing to service-based computing and Clouds. The adoption of the SOA paradigm and virtualization has resulted in an unprecedented flexibility in creating distributed applications. Old and new research challenges need to be mastered to exploit fully the potential of cloud infrastructures. In this article we present outstanding cloud-related research questions that need to be addressed. It is proposed that a pan-European research community is needed to bridge existing knowledge gaps.

Keywords: Cloud Computing, HPC, Grid Computing.

1 Introduction

One key requirement for the CoreGRID network is dynamic adaption to changes in the scientific landscape. New research challenges have arisen which need to be addressed. Lately, the advent of clouds has caused disruptive changes in the IT infrastructure world. However, there is a significant overlap between the scientific questions related to Grid and Cloud research. Naturally, part of the Grid community is also active in Cloud research. Current and future cloud research challenges are considered by means of selected examples. It is argued that these outstanding challenges need to be addressed at a European scale.

Grids and clouds have many similarities as they both address questions concerning access to resources in a large-scale distributed environment. Thus, there is significant overlap between the two areas in the ways that infrastructures may evolve. The CoreGRID programme of work has focused mostly on Grids with particular questions in application engineering and middleware management for Grids. The underlying use-case was typically resource sharing between different administrative domains for collaborated problem solving in virtual organizations. Due to the size of the research network, CoreGRID has provided a wide variety of results e.g. component models, schedulers, SLA management, or workflow management. With the advent of Cloud Computing, it is reasonable to reconsider the future research questions and to compare them between the different application fields.

M.R. Guarracino et al. (Eds.): Euro-Par 2010 Workshops, LNCS 6586, pp. 257–262, 2011.

2 Research Challenges in Cloud Computing and Grids

Cloud computing is very successful in creating a layered architecture that separates the infrastructure access from applications. Infrastructure as a Service can be utilized to run arbitrary applications. Similarly, applications can be broken down to several software services which run on such virtual infrastructures. Grids target a similar space by combining resources from different providers in a networked infrastructure. Grids also require substantial middleware efforts to form the basis of core services. Thus, there is a significant overlap between Grid and Cloud research challenges. The following is - a quite subjective - selection of main research themes that we are currently facing.

2.1 Scalability

The HPC world has significant experience in exploiting large-scale systems. The use of 1.000s of processors is challenging but well understood for many application scenarios. However, we are still in the early days of many-core infrastructures; we will see a dramatic increase in processing cores in the next years. We are already at the verge of deploying Exa-scale systems with millions of processor cores. Similarly we will also see federations of such systems in large-scale distributed infrastructures. Currently, there is doubt as to whether science has the right scalability answers. Existing approaches are not sufficient to cope with this challenge. We will need new models for supporting future Cloud infrastructures. This is not only limited to application design and parallelization, but also to aspects of infrastructure management. A multi-layered architecture will need suitable solutions to cope with the size of such infrastructures.

2.2 Improve Efficiency

High-performance computing is heavily focused on optimization. Considerable research effort has been expended to decrease response-time for relevant problems or to increase the throughput of infrastructures. This included optimization for efficient utilization of machines. As such, efficiency was and remains a crucial aspect for managing Grids and Clouds for the future. However, efficiency may well extend to more areas that considered in the past. We already live in a time in which energy consumption and Green IT became major aspects in running IT infrastructures. Most Cloud computing models are based on a clear business model and so costs in general will become a crucial aspect in optimizing service executions. As such, our assumption is that Grid and Cloud management will require novel solutions supporting multi-criteria optimization.

2.3 Reliability

Due to the size of future infrastructures and the dynamic composition of applications from many different services, we will increasingly face reliability questions. This will require better understanding of software design and novel programming paradigms for such infrastructures. Moreover, managing quality-of-service will play a major role to handle these large-scale software and infrastructure landscapes. The multiple layers in large-scale distributed infrastructures will require a suitable abstraction of service quality. Models on creating redundancy, fault-tolerance, and automatic adaption will be crucial. For instance, service-level agreements are already an industry standard in ITIL-compliant IT management. However, there is not yet sufficient support for automatically managing large infrastructures by SLAs. This will have to be taken into account in future systems.

2.4 Reducing Complexity

Most aspects mentioned above relate to the overall challenge of mastering system complexity. These systems are large-scale, very dynamic, and span multiple administrative domains. Such systems will require a high degree of automatic and autonomous management. Current approaches try to tackle this. however, it seems necessary to completely revise the way that infrastructures are managed with the transition from thousands to millions of cores or software components. Our systems are still too complicated. As a consequence, many potential user groups cannot fully exploit the technology. There is a clear need to lower the entry barrier and to also target non-experts as users. This can only be achieved by hiding the complexity and by providing easy-to-use tools, portals, or programming environments. Suitable user support systems will be essential for the broad proliferation of Clouds.

2.5 Data Management

The handling of data in distributed environments was one of the main questions in Grid systems. Data Grids have been one of the first usage scenarios. After several years of research in data management, the challenge seems larger than ever. Managing huge amounts of data seems more of an issue than managing processing power. As of now, we still have no clear understanding on how to handle data on a global scale. Map-Reduce and Hadoop became common infrastructures for many usage scenarios. But this does not yet bridge from globally distributed data repositories to individual data stores. There remains a significant gap in efficient and automatic data management. Again, it seems necessary to revisit this data challenge and to come up with fresh ideas for the future of Clouds and Grids.

2.6 Trust and Security

Grid environments already have quite sophisticated security management, e.g. through virtual organizations, certificate management, or support for secure communication channels. But it is also obvious that many of these approaches do not scale well for main stream adoption and future infrastructures. Clouds currently have even less well established security methods than Grids. It is obvious that trust and security will remain high up on the scientific agenda. It needs to be easy-to-use, non-intrusive, but at the same time reach an even higher security levels than before. Industrial adoption of cloud-based services may require support for ISO/IEC 27001, SAS 70, or SOX compliance which is typically not found in Grid infrastructures.

3 HPC and Cloud Computing

A significant part of the CoreGRID research was linked to HPC-related Grid computing. Today Grids offer common production facilities in e-Science [4]. Many scientific communities rely on Grids for resource sharing and collaboration in virtual organizations. It is not foreseeable that this will change in the near future. On the contrary, we see that more scientific communities are likely to require Grid infrastructures [2].

There is still a vivid discussion on the relation between Grid computing and Cloud computing [7]. Our understanding is that Grids and Clouds and Clouds are different. Both approaches address similar research areas but with different use-cases in mind (HPC/e-Science vs. Commercial operations) and with different technologies (common access to different infrastructure resources vs. virtualization abstraction). Both environments successfully co-exist and will have their share in their respective application realm.

However, there is no strong trend for Grids to be adopted as main stream technology for the service economy beyond e-Science and HPC. Instead, Cloud computing is taking this role. It changes the way applications are executed and infrastructure are managed. Almost all large data centers have adopted virtualization as a core technology to increase flexibility and resource utilization. Similarly, we see more applications being run as cloud-based services. There is a trend to adopt private public or hybrid clouds as an operational model for many organizations.

HPC remains an important application area, but we see more differences in the user communities requesting access to such resources. On one hand, we have the top end of HPC resources in the renowned global super-computing centers. Users of such resources are typically experts who are able to adapt to the available HPC resources. A very good understanding of the underlying hardware is required (e.g. on the specific cache structure, the interconnection network, processor capabilities). This user group is able to extract very high performance from such systems. On the other hand, we see many users who also need access to HPC resources but who are unable or unwilling to adapt to the specific hardware infrastructures. For those users it would be inefficient to access the top end HPC super computers. Instead they need easy and fast access to resources which are similar to the infrastructures which they know.

For HPC experts, the adoption of virtualization and cloud computing would be counter-productive. Application performance would suffer, while top end super computers would not be well utilized. However, for the second group clouds may be a viable alternative. This group might not require the fine grain optimization of their application to gain the final percent in performance. Instead, a fast and easy transition from existing computers to a larger resource set would suffice. Here, a Cloud infrastructure might provide on demand the required compute resources on which an arbitrary number of virtual machines with the necessary applications can run. Thus, we might see that Cloud computing will also become an operational model in the HPC eco-system in the low- and mid-range market. These are typically local or regional data centers with HPC resources for their user community. Similarly, we already see HPC services provided by commercial providers such as Amazon [3].

Obviously, this will not apply to all applications: Software that is tightly coupled with high communication demand will not be well suited for a Cloud. Today cloud computing is usually agnostic of the underlying hardware and network infrastructure. For HPC applications, we will need specific support to realize acceptable network performance. This combination of Cloud with HPC is an interesting scientific research subject.

4 Outlook

While CoreGRID gained significant international visibility in the Grid realm, there is no similar impact on the Cloud community. However, many (not all) of its members are very active in service-based computing and Clouds. This is quite natural as the research challenges mentioned above are shared in Grids and Clouds. Unfortunately, we see again a large fragmentation in the European research landscape.

Several good attempts have been made to provide a joint research agenda, e.g. by the NESSI European technology platform [1] or the S-Cube network of excellence [5]. However, none of these serves as a pan-European Cloud research center focusing on basic scientific research questions. These research challenges mentioned above are major and need to be addressed. Due to the size of these challenges, it is unlikely that single research groups or companies can solve those. It will require again a joint effort by many scientists to overcome these obstacles.

CoreGRID can play an important role to support such efforts. However, this will require a significant evolution in its research structure and membership. Despite the overlap in research questions, the scientific community for cloud research is not identical. It will require a gathering of experts from different disciplines to achieve a similar successful network as CoreGRID constitutes for Grids.

References

1. Nessi strategic research agenda, vol. 3 (May 2009)
2. Bird, I., Jones, B., Kee, K.F.: The organization and management of grid infrastructures. Computer 42, 36–46 (2009)
3. Evangelinos, C., Hill, C.N.: Cloud computing for parallel scientific hpc applications: Feasibility of running coupled atmosphere-ocean climate models on amazon's ec2. Cloud Computing and Its Applications (October 2008)

4. Jones, B.: The use of grids for fusion applications in europe and future directions. In: GMAC 2009: Proceedings of the 6th International Conference Industry Session on Grids Meets Autonomic Computing, pp. 39–40. ACM, New York (2009)
5. Metzger, A., Pohl, K.: Towards the next generation of service-based systems: The S-cube research framework. In: van Eck, P., Gordijn, J., Wieringa, R. (eds.) CAiSE 2009. LNCS, vol. 5565, pp. 11–16. Springer, Heidelberg (2009)
6. Meyer, N., Talia, D., Yahyapour, R.: Grid and Services Evolution. Springer Publishing Company, Heidelberg (2008) (incorporated)
7. Schwiegelshohn, U., Badia, R.M., Bubak, M., Danelutto, M., Dustdar, S., Gagliardi, F., Geiger, A., Hluchy, L., Kranzlmüller, D., Laure, E., Priol, T., Reinefeld, A., Resch, M., Reuter, A., Rienhoff, O., Rüter, T., Sloot, P., Talia, D., Ullmann, K., Yahyapour, R., von Voigt, G.: Perspectives on grid computing. Future Gener. Comput. Syst. 26(8), 1104–1115 (2010)