Towards the Autonomic Business Grid

Carsten Franke, Wolfgang Theilmann, Yi Zhang, Roy Sterritt

SAP Research CEC Belfast
TEIC Building, University of Ulster
Jordanstown Campus, Northern Ireland
carsten.franke@sap.com
wolfgang.theilmann@sap.com
yi01.zhang@sap.com

University of Ulster
Computer Science Research Institute,
Jordanstown Campus, Northern Ireland
R.Sterritt@ulster.ac.uk

Abstract

Business Grids are envisioned to become the adaptive service-oriented utility infrastructure for arbitrary business application satisfying the needs for agility and cost efficiency in today's and tomorrow's ever changing business environment. This industrial report paper discusses the main areas that contribute to this vision, namely Grid computing, autonomic computing and virtualization technologies. Furthermore, it shows how they relate to each other.

Keywords: business grids, autonomic systems, autonomous systems, virtualization

1. Introduction

The emerging reality of today's and tomorrow's global business environment results in a need to cope with frequently changing business conditions, such as changing strategies, models, processes, workloads, workflows, customers, and competitors. Information and Communication Technology (ICT) infrastructures must supply a means to meet these dynamic realities. Business Grids aim to provide ICT infrastructures "as a utility" so that the software and hardware resources such as applications, components, systems, and services seamlessly adapt to the dynamic business requirements [1][2].

The motivation for achieving Business Grids is they promise to provide the potential to significantly extend the flexibility of service-oriented architectures from the architectural level through to the infrastructural level [1].

The concept of Autonomic Computing (self-managing systems) [3][4][5] represents a paradigm to address a massive ever growing complexity and cost burden within today's systems, as well as a means to provide a self-managing infrastructure to enable future pervasive, ubiquitous, ambient, and service-oriented computation and communications [6][7][8].

This industrial report paper briefly examines the research landscape and challenges to achieve the self-managing Business Grid. First, the paper briefly discusses the background to Business Grids followed by a short review of traditional Grid technologies. It then looks at the Autonomic & Autonomous Systems initiative and its relevance to Business Grids. Another corner stone technique to achieving the autonomic Business Grid, virtualization, is then discussed. The paper ends with a brief conclusion.

2. Business Grids

The main driver for Business Grids is the fact that businesses in future service-oriented economies need to act in a more agile fashion than ever before. This will be possible only if high-level business requirements can be translated into lower level ICT requirements with a high level of automation, so that ultimately the ICT environment adapts automatically to changing business needs.

Consequently, Business Grids are envisioned to become the general adaptive and service-oriented utility infrastructure for arbitrary business application that supports this need for agility, of course, in a cost-efficient way.

Different business and business areas have different focuses on what they require from their Business Grids, which influences the specific research challenges. However, many research issues are valid for a wide range of business applications. In most cases, the individual preferences on the various functionalities vary. Ultimately, all these views would form a seamless part of the autonomic Business Grid. Some of these focuses are:

Data Centre Automation

This is the traditional setup of large scale enterprise customers who operate large ICT installations but still lack an overall and consistent picture of their operational management. Business Grids should support this scenario by automating and self-managing the whole range of operational tasks, especially the management of cross-installation issues (e.g. resource sharing between different applications of different users). Specific issues are to learn about conceptual limitations and drawbacks of various approaches.

Mass Hosting

Mass hosting is highly relevant for small and medium sized enterprises (SMEs) seeking to avail of the ICT benefits large organizations have. The costs for operating an ICT system are currently prohibitive for such companies. Hosting is the main option that promises to reduce operation costs to an acceptable level for SME.

Business Grids should support the utmost automation and self-management for mass hosting scenarios, especially reducing the need for manual interaction for single customers. Specific issues are isolation between the hosted parties (security, confidentiality), Service Level Agreement (SLA) provisioning, and Quality of Service (QoS) monitoring.

Dynamic Outsourcing

Following the strategic demands from businesses that ICT allows customers highly flexible business operations a specific aspect herein is on outsourcing/insourcing. Customers with smaller businesses, who originally started with a hosted solution, might at some point in time decide to insource that solution, e.g. to save costs or for security reasons. Other customers operating ICT solutions on their own might decide to outsource some solutions which are of no strategic importance but just became a general purpose service.

Business Grids should support the dynamic in- and outsourcing by combining features of the data centre and the mass hosting scenario. Additional specific issues are on-demand negotiation of requirements and licensing issues.

Virtual Organizations

A virtual organization comprises a set of (legally) independent organizations that temporarily provide resources and skills to achieve a defined goal.

Business Grids should support the flexible generation of virtual organizations as well as the flexible collaboration within a virtual organization in a way that different infrastructures, operated by different organizations, can be linked together and provide resources in a flexible but still reliable and secure way.

Service-oriented Knowledge Utilities

Service-oriented Knowledge Utilities (SOKU) envisage "a flexible, powerful and cost-efficient way of building, operating and evolving IT intensive solutions for use by businesses, science and society" [16]. This is to be achieved by strictly following the paradigms of service-orientation, knowledge assistance (semantic descriptions) and utilities (immediately usable).

Business Grids should support the infrastructure underlying a SOKU in a flexible and self-managed way (following the vision of a utility). The technical challenges are similar to the previously described scenarios with the distinction that the environment is much more dynamic and heterogeneous. Stakeholders may join or leave a SOKU, new services are introduced, non-functional properties such as security, confidentiality, SLAs or QoS have to be managed in a multi-party context. Furthermore, a high demand for interoperability is observed.

3. Grid Computing

In general, a grid infrastructure promises seamless access to computational and storage resources, and offers the possibility of cheap, ubiquitous distributed computing. Grid technology has a fundamental impact on the economy by creating new areas, such as e-Government and e-Health, new business opportunities, such as computational and data storage services, and changing business models, such as greater organizational and service devolution [9][10]. The Grid is a very active area of research and development; with the number of academic grids jumping six fold in a single year [11].

Historically, the Grid arose out of a need to perform massive computation. The current direction demonstrates the potential to change the structure of service provision and create a new grid service economy. The success of the grid will be founded on the development of new grid-enabled software systems and the evolution of legacy systems to grid-enabled systems. However, this evolution is not limited to the legacy systems but mainly includes the ability of upcoming grids to execute legacy software without further modifications. There are several middleware frameworks for distributed computing, many modeling techniques for software artifacts, and many development processes for controlling the creation of

Fourth IEEE International Workshop on Engineering of Autonomic and Autonomous Systems (EASe'07)
0-7695-2809-0/07 $20.00 © 2007 IEEE

new software systems and managing the evolution of existing software systems. However, other approaches like XtreemOS [17] aim to modify the operating system only. Thus, in such an environment an additional middleware is not necessary. Furthermore, legacy software can be executed without further modifications.

Within scientific grids, a fundamental challenge is creating correct, robust, flexible and cost-effective grid-enabled software [12]. In economic environments, the focus is more on adapting existing grid approaches such that existing business services can be executed without modifications.

In general, the Grid aims to be self-configuring, self-tuning and self-healing, similar to the goals of Autonomic Computing [15]. Its aim to fulfill the vision of Corbato's Multics [14] – like a utility company, a massive resource to which a customer gives his or her computational or storage needs [13]. As such, the self-management vision of Autonomic Computing and related initiatives are required to provide some of the answers to achieve this vision [15].

4. Autonomicity and Autonomy

The Autonomous and Autonomic Systems (AAS) initiatives address research issues concerned with creating self-directing and self-managing systems (selfware or self-* properties). The overarching vision of AAS is the creation of self-directing and self managing systems in accordance with high-level guidance from humans to address today's concerns of complexity and total cost of ownership while meeting tomorrow's needs for pervasive and ubiquitous computation and communication [17].

This vision overlaps to some degree with many other research initiatives; pervasive, ubiquitous, utility, invisible, ambient; computing and communications, while focuses on specifically addressing self-* aspects.

Autonomic Systems work through creating a cooperative environment where elements, nodes and components are each assigned an autonomic manager (AM) (Figure 1) [4][5].

These autonomic managers provide the self-awareness (self-monitoring and self-adjusting of the managed component) and environment-awareness (monitoring and reacting to the dynamic conditions of the environment). The autonomic manager to autonomic manager communications (AM⇔AM in Figure 1) includes several dynamic loops of control—for instance a fast loop provide reflex reactions and a slower loop providing coordinated event telemetry. These loops will not only trigger autonomic/self-management activity but also feed up into higher autonomy (self-directed application) planes [6][8].

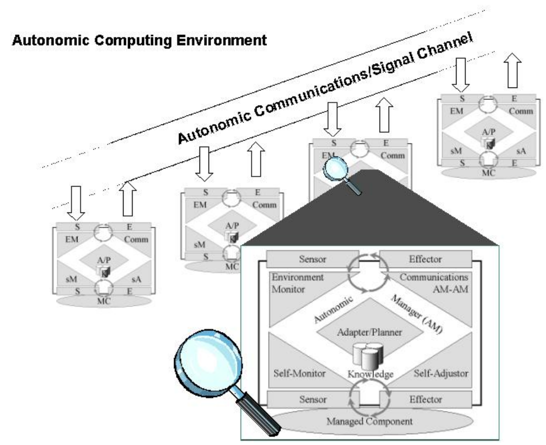

Figure 1. Autonomic Environment & Autonomic Elements

The achievement of the vision for Business Grids will require Autonomocity and Autonomy (A&A) as an intrinsic part of the infrastructure providing *self* functionality.

Some of the high level areas that A&A will span are:

Engineering: the system and software engineering (of applications, components, and middleware) integrates and perhaps even *self*-capture Business Grids related requirements.

Landscape design: Landscape design and in particular sizing is the determination of the amount of hardware needed to host a specific customer solution. Typically, sizing is a joint activity of customers and consultants combining knowledge about the customer business and the relevant software solutions. At the moment, sizing can be supported by some tools that map solution requirements onto hardware demand. Business Grids should support such a *self* sizing procedure which relates model information about the customized business solution with the respective resource demands.

Resource Allocation: The allocation of hardware resources to host specific software components must be *self*-managed. Furthermore, it would be beneficial, if the allocated hardware resources are even *self*-sized accordingly to the provider's objectives.

Deployment: The deployment of solutions on existing hardware resources is to be *self*-directed. This may include the deployment of operating systems, application servers, and business components.

Resource Adaptation: Hardware resources are dynamically *self*-configured; re-sized and re-allocated during run-time according to changes in the demand.

Fourth IEEE International Workshop on Engineering of Autonomic and Autonomous Systems (EASe'07)
0-7695-2809-0/07 $20.00 © 2007 IEEE

This includes the migration of already deployed solutions if necessary.

Operation: The full system operation and administration (e.g. software logistics) are automated as part of the *self*-management as much as possible.

Management: Furthermore, the systems must have the standard autonomic computing (self-managing) properties, namely, self-configuring, self-healing, self-optimization, self-protecting and self-adjusting.

5. Virtualization

Like the A&A systems research, virtualization is another corner stone for achieving the promise of Business Grids. It will provide the critical key facility of abstraction from the real hardware and operating system while enabling flexible service deployment, resource allocation and resource adaptation.

At the same time, virtualization technologies also add new complexity to deploy enterprise application. As such virtualization also requires the self-facilities promised from the autonomic systems research initiatives to take advantages of the abstraction while hiding the complexity from administrators and end users.

Virtualization also raises research challenges, for instance, performance impact. Since virtualization technologies add another layer between the applications and hardware this extra layer brings a performance overhead to enterprise applications, which must be managed.

Another research area to be addressed is how to efficiently manage virtualized and physical resources in virtualization environments for enterprise applications. As highlighted, using virtualization, applications are no longer running directly on physical resources. Therefore, new problems have to be faced such as how to schedule applications over virtual machines efficiently or how to schedule virtual machines over physical machines efficiently. To address these problems, the overheads of virtualization processes have to be established, as well as the exterminating the capacity of virtual and physical resources under the virtualization environments.

Research results have demonstrated that the performance of IO communication is worse in virtualization environments compared with the application on real hardware [18]. To run enterprise applications efficiently in virtualization environment, an investigation of efficient communication possibilities between virtual machines and between virtual machines and other physical resources is also needed. Furthermore, how to coordinate IO operations of applications under virtualization environments also needs addressed.

Virtualization technologies also present challenges to software development. As has been highlighted, one of the main advantage of virtualization being to hide the physical resources from operating systems, applications and/or end users. As such operating systems and applications do not need to be modified in order to run on virtualization resources. However, this leads to several new questions, like what might happen if we let operating systems and applications know about the virtualization resources. Furthermore, will the performance of applications be improved? Will the application have better fault tolerance?

To take full advantages of virtualization technologies, we need to develop autonomic and adaptive software which can adjust itself to the virtualization environments at runtime. The following questions have to be addressed to develop such software for virtualization environments: How should applications support the virtualization? How to coordinate applications and operating systems within virtualized environment? [18]

Apart from the challenges highlighted above, virtualization also brings more opportunities and challenges to system/application administration. In comparison to physical machines and devices, virtualization adds virtual machines and devices to system administrator's duty, something which the autonomic computing vision would hope to provide as self-management.

6. Conclusion

Business Grids promise to be an ICT infrastructure that can meet today's and tomorrow's dynamic business environment with its frequently changing business strategies, models, processes, workloads, workflows, customers, and competitors.

The adaptability required in the ICT infrastructure raises many research challenges. Grid technologies, Virtualization, and Autonomic & Autonomous Systems research initiatives are corner stones for achieving the vision of the Business Grid.

These initiatives are also interdependent, since for instance, both grid technologies and virtualization create more complexity in the enterprise landscape and require self-management to achieve the required vision of adaptability.

The industrial report paper had briefly discussed this research landscape and some of the challenges being addressed in the research communities.

Acknowledgements

The development of this paper was supported at the University of Ulster by the Computer Science Research Institute (CSRI) and the Centre for Software Process Technologies (CSPT), funded by Invest NI through the Centres of Excellence Programme, under the EU Peace II initiative.

References

[1] S. Dawson, C. Franke, W. Gerteis, J. Gheel, A. Jackson, W. Theilmann, M. Wuesthoff, Y. Zhang, Vision Paper: Business Grids", SAP White Paper, Sept. 2006.
[2] NESSI-Grid, "Grid Vision and Strategic Research Agenda", Deliverables D.1.1 & D.1.3 of NESSI-Grid project, http://www.nessi-europe.com, November 2006
[3] P. Horn, "Autonomic computing: IBM perspective on the state of information technology," IBM T.J. Watson Labs, NY, 15th October 2001.
[4] R. Sterritt, "Towards Autonomic Computing: Effective Event Management", Proc. IEEE/NASA SEW, Greenbelt, MD, Dec. 2002
[5] R. Sterritt, D.W. Bustard, "Autonomic Computing: a Means of Achieving Dependability?" Proc. of 10th IEEE International Conference on the Engineering of Computer Based Systems (ECBS '03), Huntsville, Alabama, USA, April 7-11, IEEE CS Press, pp 247-251, 2003
[6] R. Sterritt, "Autonomic Computing", Innovations Syst Softw Eng, 2005 1(1): 79-88,
[7] R. Sterritt, M. Parashar, H. Tianfield, R. Unland, "A concise introduction to autonomic computing", Advanced Engineering Informatics, Vol. 19, No. 3, Elsevier, Pages 181-187, Jul 2005,
[8] M.G. Hinchey, R. Sterritt, "Self-managing software" Computer, 39(2), pp. 107- 109, Feb. 2006,
[9] D. De Roure, N. Jennings, N. Shadbolt, "A Future e-Science Infastructure" aka Research Agenda for the Semantic Grid, EPSRC/DTI Core e-Science Programme, December 2001.
[10] I. Foster, C. Kesselman, J.M. Nick, S. Tuecke, "The Physiology of the Grid – An Open Grid Service Architecture for Distributed Systems Integration", http://www.globus.org/research/papers/ogsa.pdf June 2002.
[11] O. Malik, "Ian Foster = Grid Computing", Grid Today, Oct. 2002.
[12] M. Atkinson, J. Crowcroft, C. Goble, J. Gurd, T. Rodden, N. Shadbolt, M. Sloman, I. Sommerville, T. Storey, "Computer Challenges to emerge from e-Science " aka UK e-Science Research Agenda, National e-Science Centre : http://umbriel.dcs.gla.ac.uk/nesc/general/news/uk_e-Science _agenda.html
[13] J. Ledlie, J. Shneidman, M. Seltzer, J. Huth, "Scooped, Again" in Proceedings of IPTPS 2003 conference. Berkeley, CA. February 2003.
[14] F.J. Corbato, V.A. Vyssotsky, "Introduction and Overview of Multics System", Proceedings of AFIPS FJCC, 1965.
[15] R. Sterritt, "Pulse monitoring: extending the health-check for the autonomic grid," Industrial Informatics, 2003. INDIN 2003. Proc. IEEE International Conference on , vol., no.pp. 433- 440, 21-24 Aug. 2003
[16] NGG3 Expert Group Report, "Strategic Future for European Grids: Next Generation GRIDs based on SOKU - A new paradigm for service delivery and software infrastructure". December 2005
[17] IEEE Task Force on Autonomous and Autonomic Systems, 2005, http://tab.computer.org/aas/
[18] SAP Virtex Project, 2006.
[19] XtreemOS, http://www.xtreemos.org