Services Grids in Industry – On-Demand Provisioning and Allocation of Grid-Based Business Services

In recent years, Grid technology has found its way into the industry domain and offers opportunities for new on-demand business services. This article outlines the current expansion of Grid computing in industry. It also discusses the significant challenges that still need to be solved to further promote the adoption in the business domain.

DOI 10.1007/s12599-008-0009-0

The Authors

Dipl.-Inf. Jens Vykoukal
Dipl.-Wirtsch.-Inf. Martin Wolf
Chair of Business Administration, esp. Information Systems
Goethe Universität Frankfurt
Grüneburgplatz 1
60323 Frankfurt
Germany
{jvykoukal | mwolf}@wiwi.uni-frankfurt.de

Prof. Dr. Roman Beck
Chair of Business Administration, esp. E-Finance and Services Science
Goethe Universität Frankfurt
Grüneburgplatz 1
60323 Frankfurt
Germany
rbeck@wiwi.uni-frankfurt.de

Received: 2008-05-02
Accepted: 2008-09-02
Accepted after two revisions by Prof. Dr. Buhl.

This article is also available in German in print and via http://www.wirtschaftsinformatik.de: Vykoukal J, Wolf M, Beck R (2008) Service-Grids in der Industrie – On-Demand-Bereitstellung und Nutzung von Grid-basierten Services in Unternehmen. WIRTSCHAFTSINFORMATIK. doi: 10.1007/11576-008-0114-1.

1 Introduction

Over the last few years, Grid computing (Berman et al. 2003; Foster and Kesselman 1999) has evolved into a well-understood technology that provides users and applications with immediate access to a large pool of IT resources, such as supercomputers, storage systems, databases and services, which can then be used as a unified resource.

Following the widely accepted definition by Foster (2002), a Grid is a system that coordinates IT resources that are not subject to centralized control, uses standards, open protocols and interfaces, and delivers non-trivial qualities of service. Regardless of their operating characteristics, Grid computing enables heterogeneous and geographically dispersed IT resources to be virtually shared and accessed across an industry, enterprise, or workgroup.

However, significant challenges still exist at all levels which have to be addressed, including the specification and development of Grid-adapted applications, IT infrastructure, resource management, as well as networking and security issues (Parashar and Lee 2005, p. 479).

A relatively new Grid-related research field is the analysis of the economic principles needed to establish markets for Grid resources. The vision is that Grid technology is evolving towards a new business platform offering resources or services on-demand and creating new market opportunities (Caracas and Altmann 2007).

Despite these scientific approaches, the benefits of Grid adoption for the industry need to be intensively examined and promoted in order to accelerate the penetration of Grid technology into the business domain. Thanos et al. (2007) state that prior to solving complex technological issues, the economic benefits resulting from the adoption of Grid technology need to be analyzed in detail, including the analysis of Grid business models.

In contrast to this, Franke et al. (2007) claim that Grid business models can only be addressed if the technical Grid framework is mature enough to evaluate and validate such business models. Further, they state that the direct application of existing Grid technologies in the industry is not possible since the business requirements of Grid architectures are not yet covered by the current academic Grid approaches.

This paper focuses on both the economic aspects and business requirements of Grid technology in the industry. Furthermore, the paper depicts the current trend of enterprises to purchase Grid services offered by third-party utility providers on a use-on-demand, pay-per-use basis. Although currently there are only a few service providers who commercially offer Grid resources on-demand, the emerging cloud computing systems such as Amazon's Elastic Compute Cloud have great potential for enabling the creation of market-makers that further virtualize compute clouds from different providers (Buyya and Sulistio 2008). "Computing in the cloud" provides organizations with the opportunity to access a bulk of IT capabilities directly from external service providers over the Internet as needed.

The remainder of the paper is organized as follows: section 2 deals with the benefits of Grid technology for both academia and industry. In section 3, the emerging trend of enterprises to source Grid resources on-demand from external service providers is depicted, whereas section 4 presents current academic approaches to appropriate market mechanisms required to establish such markets for trading Grid resources

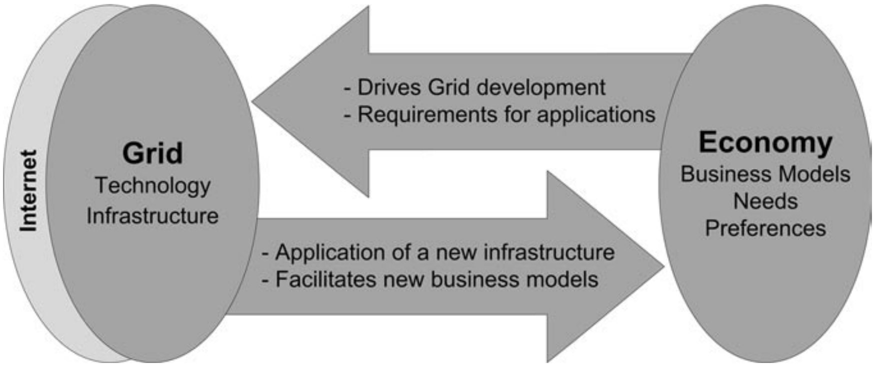

Fig. 1 Relation between Grid technology and Economy

or services. The requirements needed to be fulfilled in order to facilitate the development of Services Grids to a mature technology are presented in section 5. In section 6, the emerging expansion of Grid technology to the industry domain is illustrated. Section 7 concludes the paper with final remarks and points out further challenges and future directions of Grid computing in the industry.

2 Benefits of Grid computing

Following the publication of Foster and Kesselman's seminal work in (Foster and Kesselman 1999), interest in Grid computing has grown rapidly, attracting considerable attention from both academia and industry. This interest has been facilitated by the availability of powerful computers and high-speed networks as low-cost commodity components and ultimately by the growth of the Internet (Parashar and Lee 2005, p. 480) that made it very easy to discover and manage computing resources distributed across a Grid.

Grid systems have a set of latent properties that extend those of traditional computers, bringing several benefits (Al-Khannak and Bitzer 2008; Buyya and Sulistio 2008; Gentzsch 2005; Strong 2005):

- On-demand provisioning of geographically dispersed, heterogeneous resources
- Coordinated resource sharing and problem solving through virtual organizations
- Seamless computing power achieved by exploiting under-utilized or unused resources to solve compute-intensive problems
- Resource allocation and load balancing based on Service Level Agreements (SLAs) to meet Quality of Service (QoS) requirements
- Reduced administration effort with integration of resources as compared to managing multiple standalone systems
- A more reliable, resilient, and highly available infrastructure with autonomic management capabilities and on-demand aggregation of resources from multiple sites to meet unforeseen demand

Besides these rather technical benefits of Grid technology, there are several economic benefits enterprises can expect from the adoption of Grid solutions (Hwang and Park 2007; Thanos et al. 2007; Vykoukal et al. 2008):

- Increased productivity due to reduced processing time
- Cost reduction by means of higher resource utilization, lower IT operating costs, and economies of scale and scope
- Increased business agility, flexibility, and scalability to meet variable business demands
- Reduced time-to-market of new products leading to increased competitiveness in the market
- Increased inter-operability between different applications

This paper mainly focuses on the properties which are important for the adoption of Grid computing in a business environment.

3 Move towards utility computing

In the early days of Grid computing a global Grid as a real extension of the Internet was envisioned. However, over the years it turned out that enterprises were building up their own Grid infrastructures by transforming their vertically integrated silos, each responsible for a distinct enterprise function or application, into horizontally integrated, service-oriented architectures using different implementations, standards, and protocols (Foster and Tuecke 2005, p. 27; Gentzsch 2005; Stockinger 2006b). An overview and classification scheme for different types of Grids is provided by Kurdi et al. (2008). For the business domain, the Services Grid is the most promising Grid type since it defines a dynamic, reliable, ubiquitous and scalable infrastructure (Weissman and Lee 2001, p. 95) that reflects the recent evolution towards a Grid architecture based on Web services concepts and technologies. With the help of standard protocols and interfaces a Services Grid provides and shares resources and applications as services (Zou et al. 2005). Services advertise themselves so that they can be discovered by consumers based on the service type and other attributes they exhibit (Sobolewski and Kolonay 2006).

Currently, Services Grids are still in relatively early stages of development, and standards are still under development. One such specification is the Open Grid Services Architecture (OGSA), a standard for Grid computing environments based on Web services. OGSA supports the idea of encapsulating any form of functionality made available through a service and allows the creation of loosely coupled, platform-independent, distributed applications based on established Web service standards supporting transparency, adoption, and virtualization (Foster et al. 2003). Each service's functionality is exposed and available through standard interface definitions

Parallel to the emergence of Services Grid standards, within the last few years the prices for data and compute center services have decreased significantly, encouraging the industry to outsource their computing needs (Buyya and Sulistio 2008). As a consequence, the paradigm of utility computing is envisioned as the next generation of computational resources, infrastructure, applications, and business processes delivery in a security-rich, shared, scalable and standards-based computer environment over the Internet on a use-on-demand, pay-per-use basis (Rappa 2004; Ross and Westerman 2004).

Hence, utility computing denotes both a separation between a Grid service provider and a consumer of services (storage, CPUs, applications, etc.) and the ability to negotiate a desired QoS from the provider. In this context, the service provider can be an enterprise's IT department or an external utility provider (Foster and Tuecke 2005, p. 29). In the latter case, Grid utility computing allows enterprises to utilize advanced technologies without the need to build a dedicated infrastructure (Gayek et al. 2004, p. 43) which has a significant impact on the management of IT resources due to the business model shift from "buy a computer" to "computing on demand" (Altmann and Routzounis 2006).

Fig. 1 which is derived from Weishäupl (2005, p. 4) depicts the relation between Grid technology and the economy. The growth of the Internet and the achievements in Grid computing have led to a well understood infrastructure for Grid applications that can drive and support new business models and opportunities for enterprises. On the other hand, existing business models and economic principles further facilitate the development and improvement of Grid technology by providing what is required from the business perspective.

Currently, there are only a few third-party utility providers that commercially offer Grid services on a use-on-demand, pay-per-use basis, such as Sun Microsystems offering a utility computing service at $1 per CPU-hour (Sun Microsystems 2008). This is an all-inclusive price, encompassing job submission, result storage and retrieval, as well as some workflow support. This allows organizations to purchase computing power as needed without the long-term life-cycle costs associated with capital acquisition, systems management, and operational expense. The Sun Grid offers a pool of pre-installed applications and alternative options for deploying a customer's own applications on Sun's resources. Although Sun's Grid resources are fast in execution and the pricing scheme is easy to understand, it does not necessarily yield a cost reduction for the consuming organization. According to Blau et al. (2008), $1 per CPU-hour equals the break-even if the organization's data centers are little utilized. If the resource utilization of the organization is constantly high, the adoption of Grid technology and the development of one's own Grid infrastructure can lead to total costs lower than $1 per CPU-hour (Opitz et al. 2008). Otherwise, without consistently high levels of utilization, the costs per CPU and hour are higher than $1. In the latter case, Sun's offering is less cost-intensive for the organization than building its own Grid infrastructure. Evidently, the pricing scheme offered by Sun is not flexible since no price differentiation is applied.

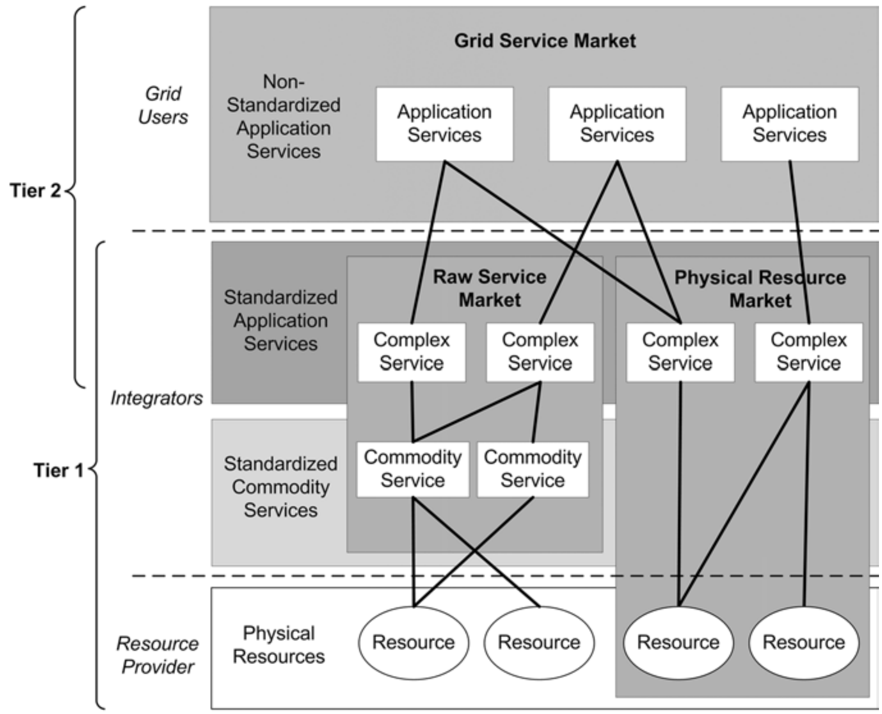

Fig. 2 Layers of different services types in a two-tier Grid service market

In contrast, Amazon's Elastic Compute Cloud (EC2) offers on-demand computing resources in the form of virtual machines accessible over the Internet (Amazon 2008). The EC2 provides three different types of virtual hardware that the user can instantiate, starting at $0.10 per instance-hour for a small instance and going up to $0.80 per instance-hour (for an instance with more memory, multiple cores, etc.). Although Amazon's pricing scheme incorporates the pay-per-use principle much better than Sun's pricing, both pricing models are too simplistic and too inflexible.

The answer to the question of whether the use of utility computing is economically reasonable for an organization very much depends on its resource needs and can therefore not be answered easily. Accordingly, the two alternatives (building your own Grid infrastructure versus sourcing resources from a third-party provider) have to be analyzed anew for each organization (Opitz et al. 2008). Nevertheless, especially for small companies without the financial opportunity to invest in their own Grid infrastructure, the on-demand use of utility computing resources can open up new business models (Schneider 2006) due to the lower total cost of ownership (Altmann and Routzounis 2006).

A market survey by Schikuta et al. (2005), analyzing mainly the middle-European IT market and focusing on how Grid technologies are integrated into a commercial IT environment, revealed that many companies are using distributed computing technologies but are not yet ready to adopt Grid technology. One of the major issues that have to be overcome in this context is the lack of viable business models for the appropriate accounting and pricing of consumable Grid resources or services (Neumann et al. 2007, pp. 1–4) as elaborated on in the following section.

4 Grid service markets

Due to its scientific roots, early Grid applications mainly focused on the non-profit sharing of resources for computationally demanding tasks between scientific institutions (Stockinger 2006a, p. 3; Németh and Sunderam 2003, p. 10). With the emergence of the commercial utilization of Services Grids, a transition in the Grid infrastructure requirements occurred

(Jiménez-Peris et al. 2007; Neumann et al. 2006, pp. 206–207; Kielmann 2006, p. 133). One of the major challenges in a commercial setting is the implementation of an adequate accounting and pricing mechanism for consumable Grid resources or services (Neumann et al. 2007, pp. 1–4). As far as services are concerned, Neumann et al. (2007) distinguish between three different kinds of services delivered through a Grid: resources, raw services and complex services, each exhibiting unique requirements when it comes to an appropriate market negotiation mechanism for pricing.

In the case of resources (processing time, data storage etc.) effective pricing can be realized by double auction algorithms (Friedman 1993, pp. 3–26). Raw services encompass standardized services which represent an abstraction of a designated set of underlying resources. Finally, complex services are tailored to a specific business requirement and thus only address a small target group. In order to provide a complex service it might be necessary to orchestrate several different underlying raw services. Hence, pricing is the minor issue in this case while effective service discovery and negotiation are required here (Bell and Ludwig 2005, p. 265–267; Neumann et al. 2007, p. 8).

One approach to facilitating effective service discovery is the enrichment of Grid services with additional semantic and logical annotations based on ontology artifacts like, for example, Web Ontology Language (OWL) or Job Submission Description Language (JSDL) descriptions and the implementation of respective search algorithms for service discovery (Bell and Ludwig 2005, pp. 265–267, Neumann et al. 2007, p. 3).

The above-mentioned heterogeneous requirements of the potential Grid market finally lead to a two-tier market approach which contains a resource market for allocating resources and raw services and a service market for complex services (Streitberger et al. 2008). **Fig. 2** illustrates the aforementioned development in more detail, based on research conducted by Blau and Schnizler (2008), as well as Neumann et al. (2007). On the highest level of abstraction, Grid users demand business-specific application services which can be traded on a Service Grid market (Tier 2) as utilities. Integrators compose and offer standardized complex services which can be part of the business-tailored application services or can be provided as autonomous services. Complex services represent a mix of commodity services and resources and can thus be traded either on service markets or on resource markets (Tier 2, Tier 1). Finally, resources delivered by resource providers can be exchanged on resource markets (Tier 1).

The establishment of accepted market allocation mechanisms such as auctions is of key importance for setting appropriate incentives for market participants. While in small scale settings bids and offers on resources or services might be processed by a central resource broker (centralized market), this approach underperforms or even fails in larger scenarios (Streitberger et al. 2008; Buss et al. 2008). Therefore, decentralized market approaches based on the "Catallaxy" market paradigm proposed by Hayek (1945) establish a matching of supply and demand in a Peer-2-Peer (P2P) manner and thus overcome the bottleneck found in a centralized market. A Catallaxy represents a self-organized coordination approach to implementing a free market for electronic services brokerage. In this concept autonomous agents maximize their own utility despite incomplete information by constantly processing and exchanging price information with agents in their proximity (Eymann et al. 2003, pp. 491–495; Streitberger et al. 2008).

Beck et al. (2008) assess the cost differences in three scenarios (product pricing, unit pricing, market pricing) each reflecting a real-world scenario. Product pricing represents the status quo in the industry, where the aggregated IT costs are split between the different departments which obtain IT resources. Unit pricing comprises the on-demand pricing of IT services and resources obtained, which reflects the service provision in a Grid system. Finally, market pricing encompasses on-demand pricing based on a price negotiated on a market platform. The simulation results reveal that there is a cost saving potential of up to 43% if market pricing is applied.

5 Requirements of Services Grids

Besides the development of an appropriate market mechanism for trading Grid resources and services, several other issues need to be addressed in order to facilitate the evolution of Services Grids to a mature technology applicable to real-world enterprise settings (Jiménez-Peris et al. 2007; Strong 2005; Plaszczak and Wellner 2006, pp. 42–63), as described in the following.

5.1 Support for service level agreements

An important issue that has to be assessed in the context of Services Grids is the effective enforcement and monitoring of SLAs. An SLA creates a contract between a client and a service provider ensuring a specific level of QoS. The QoS provided might be measured by different key indicators such as availability (uptime), response time or throughput. One conflict that can occur is that different customers simultaneously demand the resources or services assured leading to a violation of the SLAs. This conflict can be overcome by the implementation of resource virtualization mechanisms where physical resources are split into multiple isolated execution environments that can be assigned to different clients.

Most of the current literature regards resource virtualization leading to higher utilization as the major aim of Grid computing (Strong 2005; Foster and Kesselman 2004). Since virtualization technology primarily aims at putting into effect cost savings, it is quite astonishing that the quantification of costs incurred in Grid solutions has been badly neglected by scientific research so far. To the best of our knowledge, only Gray (2003) and Opitz et al. (2008) explicitly assess the costs of Grid systems. While Gray (2003) chooses an approach which is based on rather rough estimates to point out that Grid applications have to be very compute-intensive with respect to a minimum of input data in order to operate in an economical manner, Opitz et al. (2008) elaborate on expressing the total costs of ownership (TCO) of Grid through an explicit formula. Therefore, several cost categories (hardware, business premises including electricity, software, personnel, data communication) are taken into account. Moreover, the authors apply their cost calculation in order to compare several working Grid systems (EGEE I, EGEE II, Novartis Enterprise Grid) with their commercial counterparts represented by the current offers of Sun Microsystems and Amazon (see section 3). The authors distinguish between already utilized systems where idle time is used (cycle stealing) and designated Grid systems with no other utilization. In all cases the costs of the commercial offers equal or undercut the costs of the relevant self-developed systems (Opitz et al. 2008).

5.2 Integration of risk management

Besides the support and enforcement of SLAs, risk management for Grid services is also important in a business context. Since taking a specific level of risk is part of most businesses, adequate identification and treatment of risk is pivotal (Djemame et al. 2006). Risk in this context subsumes the probability of violating a SLA accidentally or on purpose. The situation is made more hazardous by the asymmetrical distribution of information between service providers and customers, leading to a principal agent problem (Eymann et al. 2008). In the case of a principal agent problem, an agent may try to behave in an opportunistic manner by exploiting its superior access to information. As far as Grid services are concerned, a service provider (agent) may have access to more information about the contractual details than the customer (principal) or the customer may be unable to effectively monitor the success of the provider's actions. Based on this, Eymann et al. (2008) propose a trust and reputation framework for different kinds of Grid coordination forms in order to overcome this situation. Following Pratt and Zeckhauser (1991) the construction of a reciprocal relationship, in particular a reputation mechanism reflecting past experience, is an adequate means of solving the negative influence of a unidirectional principal-agent relationship. Depending on the level of reputation, a risk deduction is applied to the price of the acquired service, thus creating an incentive for service providers to develop a good reputation.

A different approach is proposed by Djemame et al. (2006) who utilize a mediating software broker which evaluates a potential service or resource provider with respect to its past SLA fulfillment record. For this purpose, historical data about the results of completed SLAs and the risk self-assessment of the provider is archived and evaluated by a confidence service. Based on this information the software broker decides whether a provider systematically tends to overestimate or underestimate the risk of failing to fulfill the agreed SLA and additionally provides a corrected confidence measure. More than that, the broker may serve as a contractor for a whole workflow consisting of several single jobs. In this case, the broker autonomously negotiates with several different resource providers in order to provide a single homogeneous workflow for the customer which is finally covered by one single SLA.

5.3 Provision of a secure infrastructure

As far as the Grid infrastructure is concerned, it has to be ensured that as well as secure communication, in terms of integrity, confidentiality and non-repudiation, appropriate means for authentication, authorization, delegation, auditing, and recovery are implemented (Smith et al. 2006). Many of these issues are directly addressed by, e. g., the Grid Security Infrastructure (GSI) of the Globus Toolkit (GT) middleware. Since the Grid headnode must be accessible from outside, it is particularly exposed to multiple attack aimed at weaknesses in implementation (Schmidt et al. 2007). Moreover, several problems arise from the fact that the standard setup of the GT assumes that the Grid headnode simultaneously represents the cluster headnode. This might lead to a situation where local data of the cluster nodes is compromised due to exploits of the Grid middleware. Since Services Grids encompass the decentralized and concurrent execution of unknown code, additional measures have to be taken in order to ensure that no malicious code compromizes the Grid from the inside. Otherwise it may become feasible for an attacker to utilize Grid nodes, e. g., as a mail server for outgoing junk or in order to start denial-of-service attacks. Even if it can be assumed that no external attacks are taking place, there must be guarantees that no service provider is able to alter the (billing) data of the utilized resources acquired from a resource provider and vice versa. The same applies for information kept in databases of the service provider which must not be accessible for the resource provider.

5.4 Support for transactional data

Due to the fact that most business applications are data-intensive, two of the major requirements of many business applications are a consistent view of the processed data and furthermore, preserving the ACID (Atomicity, Consistency, Isolation, Durability) properties of transactions (Elmagarmid 1992, pp. 1–23). While the preservation of the consistency of a local database can currently be achieved, keeping track of the consistency of all data replicas in a distributed database environment is a much more demanding task. In this case, the need for synchronization goes alongside a significantly growing protocol overhead. This has lead to the dominance of stateless applications with read-only access to persistent data in Grid computing. Some work has been done on the provision of a homogeneous view of heterogeneous data containing similar information (data integration). One current trend addressed by the OGSA-DAI middleware is the encapsulation of data access via a standardized web service interface (OGSA-DAI 2008).

5.5 Autonomic support

An additional goal to be aimed at is autonomic management which comprises the self-healing, self-provisioning, self-optimizing and self-configuring capabilities of the Services Grid (Jiménez-Peris et al. 2007; Kurdi et al. 2008, pp. 10–11). Self-healing deals with the area of tension between replication and the demand for a high level of availability since replication based on replicas does not inherently lead to a high availability. In case of failures, the workload may be redirected to working replicas but some time is still needed to perform the activation of a replica. Self-provisioning aims at the ability of the Services Grid to dynamically discover and allocate new resources if the workload increases. Dynamic adaptation of performance relevant parameters depending on the current workload without the need for a designated system administrator is an essential aspect of self-optimization in order to achieve the best possible system performance. The balanced distribution of load among different sites is addressed by the self-configuring capabilities.

6 Grid computing in selected industries

The emerging expansion of Grid technology to the industry domain may be clearly seen in the initiatives of software vendors to offer a commercial platform for Grid computing such as Sun's N1 strategy, IBMs Grid and Grow, Microsoft's Dynamic Systems Initiative, and so forth (Strong

2005, p. 55). To date, many promising Grid systems have been launched in a variety of industries or are about to be. In the following, several examples are presented in order to provide a first impression of the impact and results of Grid adoption in selected industries. Due to the scope of this article, the examples aim simply to provide an overview of Services Grid adoption in selected industries rather than a complete survey.

6.1 IT industry

The software development company Synopsys changed from several distributed server farms for regression testing and code integration of their electronic design automation software to a Grid infrastructure (DesignSphere) as a single point of providing IT resources (Plaszczak and Wellner 2006, pp. 125–131). The new Grid system finally encompassed 15,000 CPUs worldwide. The transition to a Grid infrastructure led to a shorter processing time of regression tests and a more stable and significantly higher utilization (above 80%) of the underlying resources. As a major drawback of the Grid solutions, extra overhead is caused by the fact that input and output data has to be transferred to and from the Grid system before and after execution. Additionally, users are no longer able to access single computing nodes for dedicated computing tasks as was the case before.

6.2 Aviation industry

Another example of a Services Grid is provided by the DAME project, which is a joint-venture of Rolls-Royce and Science Applications International Corporation (SAIC) (Plaszczak and Wellner 2006, pp. 141–149; Jackson et al. 2005). The newly founded corporation Data Systems & Solutions (DSS) introduced the DAME system in 2002 in order to offer a sound Grid-based Equipment Health Monitoring (EHM) solution for third party customers. EHM encompasses the monitoring of physical assets such as aircraft turbines to predict necessary maintenance work and failure beforehand. This is done by evaluating data covering operational statistics, fuel burn, temperatures and vibration recordings in almost real-time. Sophisticated probability models are utilized to provide a good approximation of the necessary maintenance intervals and expected lifetime. One business model that is derived by EHM is Rolls-Royce's "Power by the Hour" which offers the all-inclusive rental of aircraft turbines on an hourly basis. Since the amount of turbine status data exceeds the limit of available bandwidth in the air, part of the data has to be passed to the DAME system on the ground. The service offered, which is targeted on aircraft turbines in this case, can also be applied to other physical assets and vehicles such as container ships.

6.3 Financial services industry

One of the most promising application domains of Grid systems is the financial services industry with its information-driven business processes and its high computational demands (Hackenbroch and Henneberger 2007; Schwind et al. 2007). Software products in the financial service industry are traditionally segmented by product, process or geographic concern which makes the inter silo re-use of provided services challenging (Bell and Ludwig 2005, p. 266). Hinz et al. (2007) assess the implications of a Grid-based approach to portfolio performance measurement in a large German bank. Since the actual return for a specific time period represents a key performance indicator (KPI) as steering tool for the management, the simultaneous retrieval of performance data is a challenging but important task for the IT infrastructure. At present the evaluation of all customer portfolios (corporate and retail) of the bank takes several weeks and thus is only performed once a year. Apparently, a more frequent provision of performance data would facilitate sustainable management decisions. The calculation of the time-weighted return (TWR) may be split up into several TWR calculations of sub-periods which are then geometrically linked together and thus applicable for parallelization on a Grid system leading to a significantly shorter process execution time.

Other Grid applications in the financial services industry illustrate the value of Grid-based solutions for risk analysis (as deployed at Bank One, BNP Paribas) and portfolio rebalancing (Charles Schwab) (Schmerken 2003; Harris 2004).

7 Conclusion

This paper has depicted the current trend of enterprises to increase their flexibility by purchasing computing resources from external service providers. This option allows enterprises to source computing power on-demand without the long-term life-cycle costs associated with capital acquisition, systems management, and operational expense, which leads to a faster IT response to changing business needs.

However, there are only a few computing utility providers such as Sun Microsystems and Amazon that provide computing resources on an on-demand, pay-per-use basis, but the offered pricing schemes are too inflexible since no appropriate price differentiation is applied. Arising from this lack of viable business models for the appropriate accounting and pricing of consumable Grid resources or services, this paper has presented current scientific approaches in the development of market mechanisms for trading Grid-based resources and services. Such Grid markets have a high potential for facilitating new business opportunities within enterprises. Although the benefits of on-demand computing marketplaces are obvious, significant additional research has to be conducted in order to establish incentives for Grid market participation by means of appropriate pricing and accounting mechanisms for Grid markets. The goal has to be a shift from isolated Grid solutions toward fully linked, interoperable ad-hoc Grid marketplaces (Broberg et al. 2008) in order to virtualize computing resources and services from different providers.

Besides the possibility of purchasing IT resources or services from third-party utility providers, many enterprises are already running their own Services Grids benefiting from Grid computing, by, e. g., increased productivity, higher resource utilization, reduced administrative effort, and increased business agility and flexibility. However, there are still significant challenges to be resolved in order to facilitate the evolution of Services Grids to a mature technology. These challenges include further improvements in the stability of Grid systems (Stockinger 2006a, p. 3), reduction of Grid coordination protocol overhead, the development of holistic approaches to virtualization and SLAs (Jiménez-Peris et al. 2007, p. 291), as well as improvements in terms of security issues (Suri and Singh 2007, p. 86), trust and reputation mechanisms (Eymann et al. 2008), and risk management (Djemame et al. 2006).

It is expected that the horizontal integration of IT resources in enterprises will continue in the future (Foster and Tuecke 2005, p. 31) and that more and more industry domains will move towards the use of Grid and utility computing in order to benefit from the properties of Services Grids.

References

Al-Khannak R, Bitzer B (2008) Modifying modern power systems quality by integrating grid computing technology. Proceedings of the international conference on renewable energies and power quality, Santander

Altmann J, Routzounis S (2006) Economic modeling of grid services. Proceedings of the e-Challenges 2006 conference, Barcelona

Amazon (2008) Amazon elastic compute cloud. http://aws.amazon.com/ec2/. Abruf am 2008-09-08

Beck R, Schwind M, Hinz O (2008) Grid economics in departmentalized enterprises. J Grid Computing 6(3):277–290

Bell D, Ludwig S (2005) Grid service discovery in the financial markets sector. J Computing and Information Technology 13(4):265–270

Berman F, Fox G, Hey AJG (2003) Grid computing: making the global infrastructure a reality. Wiley, New York

Blau B, Neumann D, Weinhardt C, Michalk W (2008) Provisioning of service mashup topologies. Proceedings of the 16th European conference on information systems (ECIS'08), Galway

Blau B, Schnizler B (2008) Description languages and mechanisms for trading service objects in grid markets. Proceedings of the Multikonferenz Wirtschaftsinformatik (MKWI'08), Munich

Buss G, Streitberger W, Eymann T, Veit D (2008) Markets and negotiations in the grid economy: how decentralized and centralized economic approaches perform. In: Bichler M et al (Hrsg) Multikonferenz Wirtschaftsinformatik 2008, Berlin

Buyya R, Sulistio A (2008) Service and utility oriented distributed computing systems: challenges and opportunities for modeling and simulation communities. Proceedings of the 41th annual simulation symposium, Ottawa

Broberg J, Venugopal S, Buyya R (2008) Market-oriented grids and utility computing: the state-of-the-art and future directions. J Grid Computing 6(3):255–276

Caracas A, Altmann J (2007) A pricing information service for grid computing. Proceedings of the 5th international workshop on middleware for grid computing, Newport Beach

Djemame K, Gourlay I, Padgett J, Birkenheuer G, Hovestadt OK, Voss K (2006) Introducing risk management into the grid. Proceedings of the second IEEE international conference on e-science and grid computing (e-Science'06), Amsterdam

Elmagarmid A (1992) Database transaction model for advanced applications. Morgan Kaufmann, San Francisco

Eymann T, König S, Matros R (2008) A framework for trust and reputation in grid environments. J Grid Computing 6(3):225–237

Eymann T, Sackmann S, Müller G (2003) Hayeks Katallaxie – Ein zukunftsweisendes Konzept für die Wirtschaftsinformatik? WIRTSCHAFTSINFORMATIK 45(5):491–496

Foster I (2002) What is the grid? A three point checklist. GRIDtoday 1(6)

Foster I, Kesselman C (1999) The grid: blueprint for a new computing infrastructure. Morgan Kaufmann, San Francisco

Foster I, Kesselman C (2004) The grid 2: blueprint for a new computing infrastructure. Morgan Kaufmann, San Francisco

Foster I, Kesselman C, Nick JM, Tuecke S (2003) The physiology of the grid: an open grid services architecture for distributed systems integration. In: Berman F, Fox G, Hey AJG (Hrsg) Grid computing: making the global infrastructure a reality. Wiley, New York

Foster I, Tuecke S (2005) Describing the elephant: the different faces of IT as service. ACM Queue 3(6):26–34

Franke C, Hohl A, Robinson P, Scheuermann B (2007) On business grid demands and approaches. Proceedings of the 4th international workshop on grid economics and business models (GECON 2007), Rennes

Friedman D (1993) The double auction market institution: a survey. In: Friedman D, Rust J (eds) The double auction market: institutions, theories, and evidence. Addison-Wesley, Reading, S 3–26

Gayek PW, Nesbitt R, Pearthree H, Shaikh A, Snitzer B (2004) A web content serving utility. IBM Systems Journal 43(1):43–63

Gentzsch W (2005) The grid: defining the future of the net. MCNC Grid Computing and Network Services

Gray J (2003) Distributed computing economics. Technical Report, Microsoft Research, Redmond

Hackenbroch W, Henneberger M (2007) Grid computing infrastructures and their value for risk management. Proceedings of the 40th annual Hawaii international conference on system sciences (HICSS'07), Waikoloa

Harris D (2004) Will financial services drive grid adoption? GRIDtoday 3(40)

Hayek FA von (1945) The use of knowledge in society. American Economic Review XXXV(4):519–530

Hinz O, Setzer M, Vykoukal J, Beck R (2007) Portfolio performance measurement based on service-oriented grid computing. Proceedings of the 3rd international workshop on enterprise applications and services in the finance industry (FinanceCom 2007), Montreal

Hwang J, Park J (2007) Decision factors of enterprises for adopting grid computing. Proceed-

Abstract

Jens Vykoukal, Martin Wolf, Roman Beck

Services Grids in Industry – On-Demand Provisioning and Allocation of Grid-Based Business Services

Over the last few years Grid computing has attracted considerable attention from the industry, because it offers opportunities for new on-demand business services for enterprises. In this article, we depict the current trend of enterprises to source Grid services offered by third-party utility providers on a use-on-demand, pay-per-use basis, thus leading to faster IT response to changing business needs. We outline the emerging expansion of Grid computing in industry and discuss the significant challenges still needing to be solved to further promote the adoption of Grid technology in the business domain.

Keywords: Grid computing, Utility computing, Services Grid, Grid economy, Grid market

ings of the 4th international workshop on grid economics and business models (GECON 2007), Rennes
Jackson T, Austin J, Fletcher M et al (2005) Distributed health monitoring for aero-engines on the GRID: DAME. Proceedings of the IEEE aerospace conference, Big Sky
Jiménez-Peris R, Patiňo-Martínez M, Kemme B (2007) Enterprise grids: challenges ahead. J Grid Computing 5(3):283–294
Kielmann T (2006) Grid applications: from early adopters to mainstream users. J Grid Computing 4(2):133–134
Kurdi H, Li M, Al-Raweshidy H (2008) Grid computing: a classification of emerging and traditional grid systems. IEEE Distributed Systems 9(3):1–13
Németh Z, Sunderam V (2003) Characterizing grids: attributes, definitions, and formalisms. J Grid Computing 1(1):9–23
Neumann D, Holtmann C, Orwat C (2006) Grid-economics. WIRTSCHAFTSINFORMATIK 48(3):206–209
Neumann D, Stoesser J, Weinhardt C (2007) Bridging the grid adoption gap: developing a roadmap for trading grids. Proceedings of the 20th Bled eConference – merging and emerging technologies, processes, and institutions. Bled
OGSA-DAI (2008) Open grid service architecture data access and integration. http://www.ogsadai.org.uk/. Abruf am 2008-09-08
Opitz A, König H, Szalewska S (2008) What does grid computing cost? J Grid Computing 6(4):385–397
Parashar M, Lee CA (2005) Grid computing: introduction and overview. Proceedings of the IEEE, special issue on grid computing 93(3):479–484
Plaszczak P, Wellner R (2006) Grid computing: the savvy manager's guide. Morgan Kaufmann, San Francisco
Pratt JW, Zeckhauser RJ (1991) Principals and agents. Harvard Business School Press, Boston
Rappa MA (2004) The utility business model and the future of computing services. IBM Systems Journal 43(1):32–42
Ross JW, Westerman G (2004) Preparing for utility computing: the role of IT architecture and relationship management. IBM Systems Journal 43(1):5–19
Schikuta E, Donno F, Stockinger H, Wanek H, Weishäupl T, Vinek E, Witzany C (2005) Business in the grid: project results. Proceedings of the 1st Austrian grid symposium, Hagenberg
Schmerken I (2003) Girding for grid. Wall Street and Technology 21(4):16–18
Schmidt M, Smith M, Fallenbeck N, Picht H, Freisleben B (2007) Building a demilitarized zone with data encryption for grid environments. In: GridNets. Lyon
Schneider M (2006) Grid-computing – science oder fiction? WIRTSCHAFTSINFORMATIK 48(1):75–76
Schwind M, Hinz O, Beck R (2007) A cost-based multi-resource auction for service-oriented grid computing. Proceeding of the 8th IEEE/ACM international conference on grid computing (Grid 2007), Austin
Smith M, Friese T, Engel M, Freisleben B (2006) Countering security threats in service-oriented on-demand grid computing using sandboxing and trusted computing techniques. J Parallel Distributed Computing 66(2006):1189–1204
Sobolewski M, Kolonay RM (2006) Federated grid computing with interactive service-oriented programming. Concurrent Engineering 14(1):55–66
Stockinger H (2006a) Grid computing: a critical discussion on business applicability. IEEE Distributed Systems 7(6):1–8
Stockinger Heinz (2006b) Grid computing in physics and life sciences. Proceedings of the 18th international conference on computer, information and systems science and engineering (CISE'06), World Enformatika Society, Vienna, 16.–18. Dez. 2006
Streitberger W, Hudert S, Eymann T, Schnizler B, Zini F, Catalano M (2008) On the simulation of grid market coordination approaches. J Grid Computing 6(3):349–366
Strong P (2005) Enterprise grid computing. ACM Queue 3(6):50–59
Sun Microsystems (2008) Sun grid compute utility. http://www.sun.com/service/sungrid/. Abruf am 2008-09-08
Suri PK, Singh M (2007) Analysis of services, challenges and performance of a grid. Int J Computer Science and Network Security 7(7):84–88
Thanos G, Courcoubetis C, Stamoulis G (2007) Adopting the grid for business purposes: the main objectives and the associated economic issues. Proceedings of the 4th international workshop on grid economics and business models (GECON 2007), Rennes
Vykoukal J, Setzer M, Beck R (2008) Grid architecture for risk management: a case study in a financial institution. Proceedings of the 12th Pacific Asia conference on information systems (PACIS 2008), Suzhou
Weishäupl T (2005) Business and the grid – economic and transparent utilization of virtual resources. PhD thesis, Faculty of Computer Science, University of Vienna, Vienna
Weissman JB, Lee B-D (2001) The service grid: supporting scalable heterogeneous services in wide-area networks. Proceedings of the 2001 symposium on applications and the internet (SAINT 2001), San Diego
Zou D, Qiang W, Shi X (2005) A formal general framework and service access model for service grid. Proceedings of the 10th international conference on engineering of complex computer systems (ICECCS 2005), Shanghai