The Step from Grid Research to Successful Grid Businesses

Jörn Altmann
School of Information Technology
International University in Germany
76646 Bruchsal, Germany
jorn.altmann@acm.org

Abstract

Resent research generated many valuable solutions for technical issues of Grid computing. The Grid research comprised scientific applications, middleware architectures, resource optimization tools, as well as management tools. However, until now, there are only very few businesses built around these Grid solutions. The existing business models only address very specific niches. Within this paper, we discuss the technology components that are missing to turn Grid computing from a research topic into a successful service for everybody. The goal is to aggregate spare resources from anybody on the Internet and sell it to somebody on the Internet in the desired size. Only if this is possible, Grid computing will be affordable and accessible to everybody in the future. Within this paper, we point out some of the different tools that are necessary to make this happen and discuss the business incentives for using Grid computing.

Keywords

Resource Allocation, Grid, Business Models, Internet Data Centers, Economics

1. Introduction

The huge investments of many countries in Grid computing research have resulted in a comprehensive set of technical solutions. The technical solutions range from managing a large number of computers, Grid middleware architectures, resource optimization tools, and modifications to scientific applications so that they can be executed on the Grid. Therefore, it is possible today to execute applications distributed over different sites, provide additional processing power on demand, use spare computational resources of computers worldwide owned by different organizations/people, or simply reduce the cost for managing the IT infrastructure of a company. Those features are essential to Grid computing.

However, none two of these features can be achieved for the same application at the same time. All available solutions just address one of these features. Therefore, Grid computing is restricted to a certain set of applications. Such applications can be divided into the following four groups:

- Applications to lower the risk of shortage of processing power,
- Applications requiring altruistic people,
- Scientific applications, and
- Applications to lower the risk of IT management failure in companies.

These application areas describe precisely the limited range of business models that exist for Grid computing nowadays. A more detailed description of the business models of these application areas is given in the next section. Future Grid business models are discussed after that.

2. Current Grid Business Models

2.1 Aggregation of Computational Resources

The past research has shown that the provisioning of large computational resources to scientific applications provides an opportunity to answer scientific questions, which could not be answered beforehand. The aggregation of computational resources resolves the issue of inefficient processing power of supercomputers. Through the combination of the processing power of several supercomputers, it becomes possible to execute more complex algorithms and solve models in greater detail. However, this is only possible within an administrative domain. The administrative domain could be just an organization or a group of organizations that agreed on sharing their computational resources.

Worldwide there are several national research projects, which focus on the design and execution of scientific algorithms for the Grid (e.g. DataGrid [7], EGEE [8]).

2.2 Aggregation of Spare Computational Resources

The aggregation of spare computational resources is very popular for applications that can be executed on any computer. Here, the owners of the computers agree to have applications being executed that run independently. This requires trust between the owner of the application and the owner of the computer. This approach only works if there are many altruistic people, which agreed to donate spare computational resources to the owner of the application.

Some very successful examples of this kind of Grid applications are the cancer research project and the SETI (Search for Extraterrestrial Intelligence) project [3][4]. Both projects require the download of a screensaver onto the computer of the participants, which will get executed if the computer is idle.

2.3 Reduced Cost of Managing Resources

For small and medium size enterprises, Grid computing could be a mean to reduce the cost of managing their IT infrastructure. These enterprises would benefit from outsourcing their IT resources. The cost savings come from lowering the risk of losing company-specific IT knowledge when one of two (or three) IT employees leaves the company. In such a small company, the loss of knowledge is difficult to compensate and threatens the operation of the entire company. In order to lower this risk, small or medium-size companies tend to outsource their IT operations to large IT management companies. If this happens, the location of the computers at the small company stays the same and the IT management company becomes the new employer of the IT managing personnel. While the computational resources are technically independent of each other, the management of the IT resources is centralized.

2.4 Reduced Risk of Resource Shortage

For some companies, it is important to have sufficient resources available at any time for performing their business critical calculations. Since the demand for these calculations can increase because of unpredictable events, there is a risk that the demand is higher than the available resources. If this happens, a huge revenue loss could be the consequence for the company. In order to reduce this risk of resource shortage, companies are willing to rent computational power (e.g. processor cycle) on a per-usage basis. The provider of the resource charges a price per processor cycle that reflects the cost of the hardware as well as the lowered risk to the company.

An example for this kind of Grid application can be found in the banking industry. There are banks that are willing to pay for the computer usage instead of purchasing a computer. Companies such as Hewlett-Packard and IBM are offering servers and pricing models that address this need of companies [5][6].

3. Future Grid Business Models

In order to enable new business models for Grid computing, it is necessary to develop technology that allows extending the application area of Grid computing. Currently, each of the business models described in the previous section can only be achieved under a certain requirement, which has to be one of the following:

- Single administrative domain (i.e. only a certain group of people is allowed to access the resource),
- Resource concentration in a single location (i.e. the computational resource is located at the consumer’s site because of security concerns),
- Altruistic behavior (i.e. people contribute resources and access to their computer since it might benefit the society at large),

- Resource reservation for a single application (i.e. scheduling issue are suppressed since only one single application can access the resource).

The goal of research in Grid computing has to be the loosening of these requirements. It should be possible to run Grid technology across different domains. The security of the program and the data to be executed as well as the hosting computer should be guaranteed. Incentive structures should be in place to compensate non-altruistic people for contributing their computational resources to the Grid. Finally, each application should be able to express its preferences to the Grid, so that the resource allocation mechanism can resolve resource conflicts of competing applications.

If this can be achieved, Grid computing will be successful. Processing power and software will be affordable for everybody. Under this model, there is no need to own any computers or software in order to run a business. All services required for running the business can be purchased on a per-usage basis from the Grid. Consequently, it will be possible to start or run a business with very little upfront investment. There is no cost of ownership. Since this will enable one-person companies to provide niche services to a large base of customers, the opportunity is given that many new small enterprises will start to exist, generating jobs.

4. The Missing Tools and Models for the Grid

In order to build this open and flexible system of computers, tools have to be available that support the operators of a Grid subsystem (e.g. data center) to make economically efficient decisions. However, only little research in this research area exists [1][2]. The existing research only proposes ways of how different companies can trade Grid resources and discusses the kind of trading infrastructure that has to be in place.

Therefore, these tools need to be developed. They must be able to assess the economical efficiency of resource purchases on the Grid. Using these tools, operators of Grid subsystems can make informed decisions about:

- Maintenance strategies for their data centers (i.e. questions such as when should faulty components be replaced or what kind of recovery mechanisms should be implemented can be answered),
- Purchases of computers for their data centers, and
- Outsourcing of computing jobs to other data centers using the Grid.

An example illustrates how such a tool can support the operator of a Grid subsystem. Figure 1 shows the number of servers per customer (A, B, C, D) allocated within a data center and the Grid over the course of a day. Between 6am and 3pm, the demand for computational resources is higher than the data center can provide. During the rest of the day, the data center has spare capacity available.

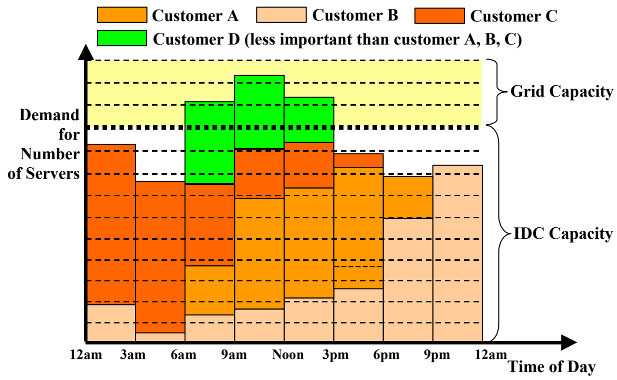

Figure 1. Example of the demand for servers within a data center

At the time when the demand is higher than the available resources, the operator of the data center has to decide whether to buy the service from the Grid or pay the penalty for not fulfilling the contract with customer D. The decision can be as simple as:

$$\text{Revenue}_{\text{customer D}} - \text{Cost}_{\text{grid service}} > \text{Cost}_{\text{penalty for not fulfilling contract}}$$

The Grid service will be purchased if the revenue from customer D minus the cost of the Grid service is greater than the penalty for not fulfilling the contract.

In addition to this, models are necessary to understand the stability of such a Grid system and the interaction between the different players, which compete with each other and, at the same time, can buy services from each other. The result of those models could provide information about:

- The optimal size of a data centers, and
- The kind of data that can be exchanged between players, without revealing important business information.

The overall outcome of using these tools and models would be knowledge about new methods on operational decisions for Grid subsystems, sustainable business models, and guidelines for policy design for guaranteeing reliable and resilient utility computing for everybody.

5. Conclusion

We analyzed the existing business models for the Grid. All of them are based on one of four models: aggregation of computational resources, aggregation of spare computational resources, reduced cost of managing resources, and reduced risk of resource shortage. Furthermore, we indicated the set of tools that are missing for establishing more sophisticated business models on the Grid.

6. References

[1] J. Altmann, "A Model for Resource Sharing for Internet Data Center Providers within the Grid," *IEEE GECON2004, International Workshop on Grid Economics and Business Models*, Seoul, South-Korea, April 2004.

[2] Chris Kenyon, Giorgos Cheliotis, "Grid Resource Commercialization: Economic Engineering and Delivery Scenarios," in: Grid Resource Management: State of the Art and Research Issues, Editors: J. Nabrzyski, J. Schopf and J. Weglarz, Kluwer, 2003.

[3] Brian S. McConnell, "Beyond Contact: A Guide to SETI and Communicating with Alien Civilizations," ISBN:0596000375, O'Reilly, March 2001.

[4] United Devices Cancer Research Project, http://www.grid.org/projects/cancer/ .

[5] Hewlett-Packard, "Pay Per Use for Servers," http://www.hp.com/large/ /infrastructure/utilitycomputing/images/PayperUse_util_pricing.pdf , September 2001.

[6] IBM, "Industry Insights," http://www-1.ibm.com/services/ondemand /thinking.html .

[7] Wolfgang Hoschek, Javier Jaen-Martinez, Asad Samar, Heinz Stockinger, and Kurt Stockinger, "Data Management in an International Data Grid Project," IEEE/ACM International Workshop on Grid Computing, Grid2000, Bangalore, India, December 2000.

[8] EGEE - Enabling Grids for E-Science, http://egee-intranet.web.cern.ch/egee-intranet/index.html , Februray 2005.