Keywords: Grid computing, SOA, Web services, Case studies, SME.

Abstract: This paper describes practical experiences from a project to couple Grid and SOA technologies in smaller environments. Web services have been applied in two structurally different case studies to solve tasks with a Grid that is integrated into a SOA and vice versa. The case studies have revealed important insight on how and when to couple SOA and Grid technologies including monitoring aspects. Some interesting general rules are derived on what has to be observed when combining SOA and Grid in smaller environments. Performance and software technical analysis have been used in validating the results. They also clearly showed the benefits gained by employing SOA and Grid concepts form both a performance as well as an architectural perspective.

1 INTRODUCTION

1.1 Motivation

Today’s IT systems are often comprised of quite different technical parts. A heterogeneous communication infrastructure is thus used to connect them. To exchange information between such systems XML has established itself as some lingua franca. Based on the XML fundament more specialized languages became popular for service description and communication between systems. A well known example is the combination of service interface description based on WSDL and abstracted service communication with SOAP, in this paper jointly called SOAP/Web Services (SOAP/WS).

No wonder, that SOAP/WS is examined for communication nowadays within enterprise software architectures as well. Especially enterprise IT tries to utilize SOA to become more flexible and to allow for better control. At the same time distributed, parallel problem processing across heterogeneous hardware became more and more popular recently; Grid technology, which couples such systems to become virtual super computers is here the means of choice. Since in both cases (SOA and Grid) communication is required, a common usage of XML-based SOAP/WS ((W3C, 2007), (Conrad et al., 2005), (Erl, 2005)), SOA ((Conrad et al., 2005), (Erl, 2005), (Krafzig, Banke and Slama, 2005)) and Grid ((Foster and Kesselman, 2003), (Conrad et al., 2005)) seems interesting for further investigation. Moreover this combination can be seen as an advanced project experience for students.

This experience report thus describes two case studies from student projects in the SOA/Grid space and presents experiences gained within them. While the first case study focuses on Grid computing technology for parallel problem solving with some smaller SOA part, the second study uses an SOA combined with a simulated data Grid. Both case studies commonly use SOAP/WS as communication base. Another requirement was the usage of Open Source Software only. For example the Grid technology Globus Toolkit (Foster, 2006) was used and the Nagios (Nagios, 2007) monitoring tool was extended to allow for initial SOA-Grid monitoring. Some performance and software technical analysis was performed as well.

1.2 Related Work

Related work comes from different areas: Coupling of SOA and Grid in very large scale IT landscapes is examined in (Holmes, Johnson and Miller, 2003) and (Chen et al., 2006). There is even a widely accepted proposed architecture (OGSA; (Foster et al., 2005)) for these kind of systems. It is however a
very general architecture concept (similar in (Huang, 2003)) not focused on a specific problem context.

More specific work just occurs sporadically in some kind of individual solution, e.g. for astronomy (Wang et al., 2006), bioinformatics (Xu et al., 2006), or geosciences ((Fraser, Rankine and Woodcock, 2007), (Patra and Das, 2005)). This paper thus adds more specific work that especially addresses smaller IT environments in the combined GRID/SOA space. The fact that many of the subject related articles are quite recent, supports the claim that there is still much work needed in this area.

For distributed search in Grids a data structure is proposed by (Tadepalli, 2006), not combined with SOA however. Prototypes for easier WS development in Grids are shown in ((Bocchi et al., 2005), (Kwon, Choi and Cho, 2006)), but they assume an SOA to communicate among the Grid nodes only.

2 SOA, GRID AND WEB SERVICES

To utilize the benefits of SOA and Grid within one single IT system jointly, a concept is required to couple both technologies. Two options are obvious: Option one utilizes Web services as communication technology between Grid nodes; option two is an SOA, which uses a Grid as some super service for calculation or data storage.

Within option one the nodes of a Grid can be seen as services within an SOA. Each node can typically be used via SOAP/WS and together the services form a Grid. All communication within the Grid is thus based on Web services. Part of the case study for this option is to check, whether the flexibility gained by the usage of Web services help significantly or whether the overall Grid performance is influenced negatively instead. Possibly the decision about the practical relevance of this option is even depending on the specific application and the type of Grid (computing Grid, data Grid, etc).

In the second option the whole Grid as one (service) element within an SOA is examined. The Grid offers its service to other components within the SOA; again SOAP/WS is used, but only to communicate with the node which controls the Grid but not necessarily within the Grid itself. Internally the Grid could instead use any kind of communication protocol. This architectural option seems well suited if an existing computing or data

Grid shall be utilized within an SOA. From an SOA viewpoint the Grid is just another service like other services. Grid specifics are not used.

The following chapter shows a case study for both options. Based on the gained experienced both architecture options are conclusively valued.

3 CASE STUDIES

For a practical evaluation of the combined Grid and SOA concepts in smaller environments (smaller with respect to hardware in use), two prototypes where developed as case studies from Q3/2006 till mid 2007. Both prototypes used open source Grid- and SOA technology (Globus IV-Toolkit (Foster, 2006), Celtix 1.x (Celtix, 2007) or Apache CXF Enterprise Service Bus (CXF, 2007)). The development was executed by 11 bachelor students in their final study year, supervised by 2 professors. Effort was 1 day per week per student. Both prototypes were based on medium size Linux PC workstations as well as 2 multi processor Linux servers.

3.1 Focus Grid: “RSA Key Challenge”

3.1.1 Description

The idea behind the first case study was to utilize a computing Grid to break (reasonable easy) RSA keys based on factorization of prim numbers. Motivation for this prototype was the "RSA Key Challenge" (RSA, 2007). Within this case study a simple “brute force” factorization algorithm using division by prim numbers was used. The focus was on getting started with Grid and SOA technology with a reasonable easy problem domain.

Figure 1 shows the architecture of the first case study. Monitors mean end users, pc symbols mean system components possibly distributed across different computing nodes. SOAP/WS is used for communication between nodes.

Starting point in figure 1 is the presentation server component. It is a servlet, which takes user-initiated tasks for factorization and passes them on to the management server. In return it gets back results and presents them to the end user. Core component is the management server WS. It distributes sub tasks for prim factorization to different Grid nodes.

For some more technical detail, listing 1 in appendix A shows an excerpt of the management server’s WSDL. It contains operations to start and delete (Grid node) jobs, get a list a of current jobs etc. Note the relation to resource descriptions and
stateful services, which occasionally occur in Grid service architectures.

Each Grid node is a Web service itself, which takes a number to be factorized and an interval to be tested. If the factorization is successful, the result is returned, otherwise an error. The Grid nodes process their number intervals in parallel. Using the monitoring tool Nagios runtime supervision of the overall SOA-Grid is performed and visualized.

![Figure 1: Case study 1 – “RSA Key Challenge”](image)

### 3.1.2 Experiences

Development of the software for this case study was performed just as a typical standard application would have been. No major problems were encountered and the students had no problems to develop software for this specific area. Even the new technologies such as web services and Grid computing did not pose any major challenges apart from the standard setup time required to be familiar with a new technology.

The project team of students was divided into sub teams which were used as domain experts during the course of the project. Domain experts were concerned with tackling the following work packages: project management, detailed specification of the scenario, Grid technology, enterprise service bus (ESB) technology, hardware and system software as well as monitoring.

Some minor problems have been encountered in the Grid software team. These were due to the somewhat unhandy web services layer of the Grid software used (Globus toolkit 4). Consequently this layer has not been used directly in the case study but rather the more basic GridFTP has been used for data transfer on the Globus side. Since the case study’s goal was to examine web services in a Grid context the team developed its own web service wrapper for the GridFTP base technology. This wrapper has been based on another open source product, namely the Celtix ESB implementation.

With this case study focusing on using a Grid to solve a computationally complex problem it should also be evaluated how and to what extent a Grid will help to solve the problem faster than a single machine could. Our lab environment for performing this quantitative analysis consisted of up to 4 completely identical Grid nodes. Standard PCs with exactly the same hardware and software components have been used. Since only relative times are relevant in this study the simple hardware is sufficient as long as all nodes use the same.

<table>
<thead>
<tr>
<th>Number of Nodes</th>
<th>Very small numbers</th>
<th>Small numbers</th>
<th>Large numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4.5 s</td>
<td>4.4 s</td>
<td>1254 s</td>
</tr>
<tr>
<td>2</td>
<td>4.4 s</td>
<td>4.5 s</td>
<td>618 s</td>
</tr>
<tr>
<td>3</td>
<td>4.4 s</td>
<td>4.5 s</td>
<td>419 s</td>
</tr>
<tr>
<td>4</td>
<td>4.7 s</td>
<td>3.6 s</td>
<td>120 s</td>
</tr>
</tbody>
</table>

As shown in Table 1 we can see that for very small numbers only the Grid setup time dominates and this using multiple Grid nodes is not reasonable. With a little bit larger numbers we can detect the first significant speedup, but still using a Grid does not gain enough. This situation changes completely with using really large numbers (in our study these numbers consisted of 23 digits, but this is definitely dependent on the hardware used). Here we can observe a significant speedup by using a Grid. Since the setup time of the Grid is negligible here (compare the 4s for very small numbers to the absolute time required now), even two Grid nodes already lead to a near linear speedup. In the case of 4 nodes we even observed super-linear speedup which is due to some specifics of our implementation. In general at least relatively linear speedup can be expected. Note, that we did not test speedup which goes widely beyond 4 nodes, e.g. to 100 or 1000 nodes. However, due to the “linear design” of our RSA scenario, we still expect at least very reasonable speedup for larger scale environments as well. This should hold at least as long as we do not hit network speed or latency boundaries significantly.

Figure 2 illustrates that the good performance results by adding more Grid nodes are not dependent on the existence of a solution to the factorization problem (red line) or non-existence (green line). In both cases similar speedup can be achieved. This test has been using large numbers only.
3.2 Focus SOA: Specialized Web Search Engine – “IT Web Indexer”

3.2.1 Description

Figure 3 illustrates the second case study. It shows an “IT” specialized Web search engine, an “IT Web Indexer”. The components work jointly in a SOA, which uses a simulated data Grid as a storage service. For time reasons, this data Grid was implemented as a distributed, replicated database.

The IT Web Indexer works as follows: a Web crawler WS manages a list of URLs to be examined in the data Grid. Each loaded Web page is reduced to its text content only and passed to an analysis WS, which examines the page for “IT relevance” by means of a pre-defined IT glossary. If a certain score is exceeded, the page is seen as relevant and stored using the data Grid service. The value is passed to an index server WS, which indexes the word from the Web page, eliminates duplicates etc.

End users access the results by means of a Web page, which interacts with the user (input gathering and validation, result preparation) as well as with the request server WS. The latter separates the user query into IDs, which are prepared as queries for the data Grid storage service. The query result is a list of page IDs, from which the user can pick, like in popular Web search engines.

3.2.2 Experiences

As expected using a SOA with Web Services for communication lead to an extremely flexible system architecture. On one hand the separation of the whole software system into several functional components induces a very well structured software system. Single components (where a component can be identified by the functionality it provides to the system as a whole) can be replaced by different implementations as desired. On the other hand the different software components within the case study could be distributed arbitrarily among the available resources. This leads to a well improved usage of resources which can even be adjusted dynamically depending on the current need of the system. Thus it was possible to achieve a pretty good system throughput (measured in terms of web sites scanned and hits found per time unit) even with the very limited resources available in a student project.

As an example for the good flexibility, which the SOA-based design provides, one could exchange the front end easily with a self-developed application, which just calls the appropriate Web service. Other options would be specialized front-ends for mobile devices like PDAs or highly interactive Web GUIs based on Web 2.0 technologies like Ajax. Similarly, the crawler Web service could be implemented in different versions, e.g. to search other data sources like local documents rather than the Web. Figure 4 shows such SOA-based architectural variations of the “IT Web Indexer”.

Figure 2: “RSA Key Challenge” – Speedup with/without solution (large numbers).

Figure 3: Case study 2 – “IT Web Indexer”.

Figure 4: SOA-based architectural variations of the “IT Web Indexer”.
The data Grid providing database storage and access capabilities for the search engine has been included into the SOA as a “super service”, meaning that communication with the data Grid is performed just as communication with any service in the system (thus making it a service); on the other hand the data Grid provides much more functionality and its internal complexity is much larger compared to a classical service (making it a super service). Nevertheless viewing the data Grid as a service made it possible to include it smoothly into the system: There was e.g. no need for any specific interaction methods.

In total, the service-based design of the system resulted in a highly flexible architecture, as promised by SOA.

Similar to our study only limited resources will be available in a realistic setting where Grid and SOA are to be employed in smaller environments, e.g. small companies. Our second case study shows that even then a combination of the two technologies can lead to a considerable performance boost due to the flexible architecture and improved resource usage.

As in the first case study a specifically extended version of the open source tool Nagios provided integrated monitoring capabilities of both SOA and Grid. Such integrated capabilities proofed to be essential for both the development as well as the operation phase. During the development phase monitoring is important for locating errors originating from the distributed nature of the application. During system operation, monitoring is required for availability control as well as for system tuning issues. The following Figure 5 shows a screenshot of the running monitoring system, which monitors the SOA-Grid. As can be seen, all nodes run fine except node visogrid04 which is currently down.

In our experience without proper monitoring of the different nodes it is quite difficult to automatically detect available resources which may be employed by a different component for optimized system performance. Without monitoring, the overall SOA-Grid would be much harder to develop and almost impossible to manage even in small environments like ours.

4 CONCLUSION AND FUTURE WORK

4.1 Lessons Learned from Case Studies

Both case studies showed that the combination of SOA and Grid technologies is possible without any problems. We can also conclude that both architectural variants may be employed for implementation. Which of the two variants should be preferred is strongly application dependent.

The first architecture where a computing Grid uses web services for internal communication has been proven to provide an easy implementation of a computing Grid. Without using web services (and by using a proprietary communication protocol instead) the system would have been much less flexible. i.e. using different hardware and operating system on some of the Grid nodes would not have been possible without additional development effort. This architecture is particularly beneficial, if the messages exchanged among the Grid nodes are diverse and the hardware, operating system and/or software on the Grid nodes varies. Even though web services incur a certain communication overhead we observed a significant speedup of the Grid application and the improved flexibility outweighs
the overhead in this context. The performance analysis, which we performed, showed the significant speedups, which are possible especially even for small SOA-Grids like ours.  

The architecture in the second case study, where a database Grid was integrated into a distributed application as a super web service, has also proven to be efficient for this kind of application. In this case the flexibility of the SOA has been used in favor of implementation of the distributed application. The advantage of flexible resource allocation based on a possibly heterogeneous computing infrastructure facilitates an efficient implementation of this complex application. Communication within the data Grid is proprietary in this case, since the messages to be exchanged are fixed and well-known. This holds since they are defined by the database system used for implementing the data Grid. Nevertheless Web services are used to access the data Grid as a whole in order to be able to use the data Grid as flexible and wide spread as possible. The highly increased flexibility shows, that SOA usage makes sense, even for comparable small applications. The easy integration of a Grid as a super-service, as well as flexible exchange or additions of services within the second case study pointed this clearly.

In both studies we observed that a loose coupling of the two concepts Grid and SOA proved beneficial. The Grids have been integrated into the distributed application by means of services in both cases. This makes flexible usage of the Grids possible. Within the Grids services have been used for the computing Grid but not for the data Grid. As a general rule we can state that services within the Grid provide an advantage as long as many diverse services are used within the Grid and a heterogeneous computing infrastructure is used. If that is not the case proprietary communication should be used within the Grid.

Distributed applications running on a heterogeneous infrastructure require advanced monitoring capabilities of the system as a whole. In our studies the tool Nagios with specific extensions developed within the scope of the studies proved to be extremely valuable. It is possible to monitor hardware as well as software status of the system. Monitoring for both Grid and SOA could be nicely integrated and a good overview of the system as a whole is achieved.

All the software developed within the case studies has been based on open source products. The fact that the applications have been successfully implemented based on this kind of software with acceptable effort show that even in such complex application scenarios open source software is an alternative. Especially in the case of SMEs where financial investments in IT have to be quite limited it is important to be able to use open source software. Cheap software complements the potential to use a heterogeneous computing infrastructure and flexible resource allocation very well. Jointly it is possible to implement a complex and highly productive software system at comparable low cost. This is especially important for smaller environments in SMEs.

Eventually, the results showed as well, that such relatively complex integration projects are quite feasible for teaching purposes in (advanced) student courses respectively projects.

4.2 Future Work

The quantitative evaluations of the distributed applications developed in the case studies should be extended further. Especially many more different hardware and software foundations on the Grid nodes should be examined. This could potentially lead to more insight into the influence of heterogeneity for the Grid and SOA applications. Of course interesting as well would be to add significantly more Grid nodes, to validate the scalability results. We plan to explore this in the future. In this case care must be taken however, that our targeted “smaller environments” are still addressed here primarily.

If the customized extensions to Nagios would be extended somewhat further, a tool for general Web Service and Grid management and monitoring could be obtained. Such a tool would be of great benefit to many different SOA-based software systems and has many potential use cases.

The combination of SOA and Grid should also be employed for implementation of real-world applications used within a SME. Practical experiences gained from this kind of applications and from the integration of the different technologies into an actual IT landscape of a SME could reveal further perception of the applicability of the architectural variants. Finally more benefits of the combination of Grid and SOA could be deduced and the necessity for improvements could be detected.

ACKNOWLEDGEMENTS

We would like to thank our students from the ViSoGrid project team for their highly productive and enthusiastic work in this project. Special thanks goes to the team members E. Friedrich, B. Hellmann, A. Reich, M. Schaaf, and J. Salzwedel.
REFERENCES


APPENDIX A

The following listing 1 shows an excerpt of the WSDL for the management server from the RSA key challenge case study.

36
Listing 1: WSDL excerpt for RSA key challenge.