Leveraging Grid Technologies For Reservoir Uncertainty Analysis

Zhou Lei, Dayong Huang, Archit Kulshrestha, Santiago Pena, Gabrielle Allen
Center for Computation and Technology, Louisiana State University, Baton Rouge, LA 70803, USA
Xin Li, Richard Duff, Subhash Kalla, Chris D. White, John R. Smith
Department of Petroleum Engineering, Louisiana State University, Baton Rouge, LA 70803, USA
{*zlei, xli, dayong, duff, architk, santiago, gallen*}*@cct.lsu.edu*
{*kalla2, cdwhite, jsmith5*}*@lsu.edu*

Abstract

Reservoir uncertainty analysis is targeted at obtaining assessments and predictions of reservoir performance, for the purpose of guiding development and operational decisions. However, accurately analyzing various reservoir uncertainty factors is a challenging issue due to the associated large-scale data manipulation and massive reservoir simulations which cannot be easily handled with the typical resources of a single institution. Security issues hinder effective collaborations between researchers interested in reservoir studies. We leverage Grid computing technologies to address these concerns. A data replication tool has been implemented for manipulating raw geological&geophysical (G&G) data, well logging data, and simulation results. A task farming framework has been developed for massive reservoir simulation executions. GSI (Grid Security Infrastructure) has been employed for security. This paper describes the design and implementation on these solutions. The case studies are introduced to verify our contributions. Our efforts also provide Grid solutions for other computing-intensive and data-intensive uncertainty analysis, such as coastal modeling.

1. Introduction

Although technological advances have improved significantly in contemporary petroleum exploration and development, risk has not been reduced in all cases. For instance, the high costs associated with platform design and well construction in deepwater projects lead to large initial capital investments being made with only limited knowledge of reservoir architecture and geology. Proir to investments, petroleum exploration and production engineers need to be able to identify the characteristics and various uncertainty factors of a reservoir, and then quantify and analyze these uncertainty factors in the data acquisition. Reservoir simulation [1] is the main approach for characterizing a reservoir in the planning and evaluation of sequential development phases. A reservoir engineer adopts uncertainty analysis/sensitivity study [2] to predict reservoir performance. In an uncertainty analysis process, various combinations of uncertainty factors are assessed to construct diverse models for reservoir simulations and simulation results are analyzed to estimate the sensitivity issues.

Response surface and experimental design methods are frequently used for uncertainty study of complex reservoir systems, which are computing-intensive and data-intensive processes. Let us take an example. These methods are applied to a single-well water-drive gas reservoir with a radial geometry [3]. Fourteen factors are considered for this study. There are eleven geologic factors (initial pressure, horizontal permeability, connate water saturation, critical gas saturation, gas end point, water end point water Corey exponent, gas Corey exponent, non-Darcy coefficient, aquifer size, anisotropy ratio) and three engineering factors (completion length ratio, tubing head pressure, tubing diameter). The simulation runs for full factorial design would be $4^6 \times 3^8 = 26,873,856$ if six factors have four levels and eight factors have three levels.Conservatively assumed a single simulation run with a common grid size of 50 feet for a middle scale reservoir consumes 6 minutes CPU time, the total execution time would be 2,687,386 hours (or over 100 days on a 1024 processor cluster). Meanwhile, large-scale data are involved in such a study. Geological&geophysical (G&G) data and well logging data are geographically distributed, which size scale is terabytes, even petabytes. The average result dataset of one single simulation reaches up to 50 Megabytes. Massive simulations lead to storage needs which cannot easily accommodated with a typical storage resource.

To conduct uncertainty analysis, a reservoir engineer has to minimize the number of uncertainty factors and factor levels, which may often lead to incorrect conclusions.

Grid computing technologies [4] provide tools for coordinated resource sharing to support distributed, dynamic, and heterogeneous virtual organizations. Grid computing is an active area of research, which holds great potential promise for large-scale science and engineering applications. Our work focuses on leveraging Grid computing technologies for large-scale reservoir uncertainty analysis. We provide data manipulation mechanism to handle reservoir modeling related data (i.e., G&G datasets and well logging datasets) and task management strategy to execute massive simulations with different reservoir models. We also put the efforts on security consideration and result analysis.

This paper is organized as follows. In Section 2, we describe the background of our research. Section 3 describes our efforts on leveraging Grid computing technologies for reservoir uncertainty analysis. We provide the case studies in Section 4. Related work is shown in Section 5. Finally, Section 6 provides the conclusions and details of future work.

2. Background

The objective of reservoir uncertainty analysis is obtain assessments and predictions of reservoir performance, for the purpose of guiding development and operational decisions. Through reservoir studies, engineers strive to forecast the results and consequences of different development and production scenarios.

A reservoir can be represented as a mathematical model by applying the mass conservative law (i.e., Darcy's law), relative permeability and capillary pressure relationship in a differential equation [1]:

$$\nabla \cdot (\rho_m K \lambda_m \nabla P_m) - q_m = \frac{\partial(\phi \rho_m S_m)}{\partial t}$$

where $m =$ oil, water, or gas; ρ_m = density; $K =$ permeability; $\lambda_m =$ mobility; $P_m =$ pressure; $q_m =$ production rate; $\phi =$ porosity; $S_m =$ saturation; and $t =$ time. To obtain an analytical solution of a reservoir, numerical simulation is required. A reservoir simulation consists of the following steps: 1) Geologists build the most representative geological model using seismic, well logging and other geological data. 2) Geostatistical realizations are generated to sample the uncertainty of geological parameters. 3) Reservoir engineers combine geology, fluid and flow parameters, along with well locations and other engineering factors to constitute a base model. 4) This model is simulated to obtain production profiles and recovery factors for a chosen recovery process. 5) Economic performance indicators, such as ROI (Return on Investment) and NPV (Net Present Value), are calculated.

Uncertainty analysis and sensitivity studies with reservoir simulations are critical for forecasting reservoir performance. Experimental design and response surface methodology [5] provide mechanisms to assess uncertainty by providing inference with a number of reservoir simulations, as well as to quantify the influences on production and economic forecast. A *design* is a set of factor-value (varied parameters) combinations for which responses are modeled. More than two levels (not just low value and high value) of each factor must be considered for a non-linear oil and gas reservoir response. A response surface model associated with a combination of uncertainty factors is an empirical fit of reservoir simulation results as follows:

$$\begin{aligned}\widehat{y}_j(\vec{x}) &= \widehat{\beta}_{j,1} + \sum_{i=1}^{k} \widehat{\beta}_{j,i+1} x_i + \widehat{\beta}_{j,i+k+1} x_1 x_2 + \widehat{\beta}_{j,k+2} x_1 x_3 \\ &\quad + \ldots + \widehat{\beta}_{j,1+k(k+1)/2} x_{k-1} x_k + \sum_{i=1}^{k} \widehat{\beta}_{j,k+1+k(k+1)/2} x_i^2\end{aligned}$$

where $y =$ responses; $x =$ uncertainty factors; $k =$ the number of uncertainty factors; and $\beta =$ regressors. A large number of reservoir simulation runs are involved in this kind of factorial designs if many uncertainty factors are considered, which motivates the improvement of both computation technologies and optimization studies.

Reservoir engineers have adopted various hardware platforms and software packages to perform uncertainty analysis with the experimental design and response surface methodology for reservoir studies. Used hardware platforms ranges from personal computers to high performance clusters. Software includes diverse open source or commercial reservoir simulators, geostatistics toolkits, visualization tools, etc. Using reservoir modeling and simulation runs with different combinations of uncertainty factors and factor levels, the sensitivity of each uncertainty factor can be identified.

There is no integrated, secure, and ease-to-use problem solving environment available for use although some efforts [6] have been made. A reservoir engineer needs to manually make these toolkits and platforms work together.

3. Leveraging Grid Technologies for Reservoir Uncertainty Analysis

In this section, we introduce the motivations of our efforts, and then describe the solutions on large-scale data management, massive reservoir simulations, and security consideration leveraging Grid computing technologies.

3.1. Motivations

Limitations restrain advanced reservoir uncertainty analysis. A single high performance computing facility cannot satisfy the requirement of massive reservoir simulation runs. Large-scale data management is required for both modeling-related data and simulation results. Security issues hinder effective collaborations between researchers interested in reservoir studies. To conduct a reservoir uncertainty analysis process, a reservoir engineer needs to minimize the number of uncertainty factors and factor levels, which may often cause the loss of correct conclusions.

People pursue two different approaches to address and improve reservoir uncertainty analysis. One is to develop optimization algorithms to minimize the search space for the “most plausible” sets of model parameters [7]. The other is to push the limits of the latest computational technologies to provide large-scale data and computing capabilities for massive simulation executions. We concentrate our efforts on the latter approach.

We integrate Grid computing technologies with reservoir uncertainty analysis, pursuing more precise reservoir performance prediction and less risky decision making in the planning and evaluation of sequential development phases. In reservoir uncertainty analysis, thousands of simulations associated with different combinations of uncertainty factors and factor levels can be executed across multiple computing resources. Large-scale data storage and manipulation can be achieved with the help of data Grid technologies. In addition, security issues in a virtual organization have been addressed by Grid computing technologies, which enables reservoir modeling-related data to be integrated with high security.

3.2. Data Management

There are three data management activities involved in a reservoir uncertainty analysis process: (i) acquiring distributed modeling-related data; (ii) constructing reservoir models, and (iii) archiving massive simulation results. A data replication tool has been designed and implemented to manipulate large-scale raw G&G data and simulation results. With the help of this tool, the reservoir modeling mechanism has been accomplished.

A. Data Replication Tool

This tool is designed to replicate data sets on data Grids, which provides a mechanism for fast, efficient, robust, and secure replication of data. Reservoir uncertainty analysis process employs it for acquiring distributed modeling-related data and archiving massive simulation results.

This tool has three modules: metadata service, replica location service, and high performance data transfer service. They are lightweight, providing the interfaces to query the external services given by a Grid. The interaction among these services is shown in Figure 1. Provided the information describing the required data (Step 1 in the figure), metadata service retrieves logical filenames (Step 2), replica location service obtains the physical file locations which map to the logical filenames (Step 3), and then the physical files are downloaded via high performance data transfer service (Step 4). Once the data transfer is completed, this tool updates metadata server and replica location server database (Step 5 and 5’) due to this data transfer transaction.

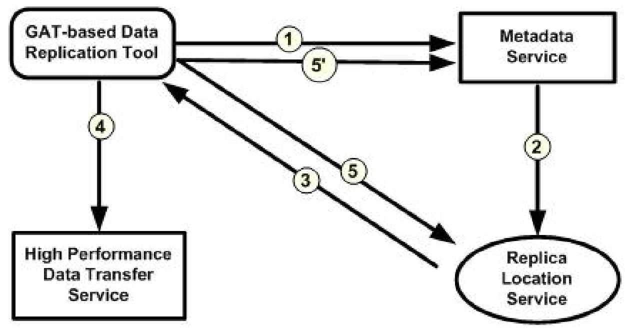

Figure 1. Data replication scenario

This tool was implemented on top of the Grid Application Toolkit (GAT) [8]. The GAT is a high level application programming toolkit. It is a unified simple programming interface for the Grid infrastructure. The GAT provides various high-level Grid functionality abstractions by GAT-CPIs (Capability Provider Interface). There are three kinds of CPIs involved in this tool: metadata CPI, logical file CPI, and file transfer CPI. Correspondingly, three GAT adaptors were developed to implement these CPIs: MCAT [9] adaptor for metadata service, Globus [10] RLS adaptor for the mappings from logical filename to physical filenames, and GridFTP adaptor for high performance data transfer.

This tool demonstrates high flexibility and portability in diverse Grid environments. Benefiting from the design of the GAT, a user can replace different adaptors without any change of this data replication tool. These replacements totally depend on the Grid core services available and the performance requirements. For instance, a user can easily replace the MCAT adaptor by Globus RLS based advert adaptor for metadata management, GridFTP adaptor by the CURL [11] adaptor for data transfer, or RLS logical file adaptor by the SRB [12] adaptor.

B. Reservoir Modeling

Reservoir modeling aims to build models for massive simulations. The procedure consists of constructing the base model, creating uncertainty parameter space, and generating reservoir models for each combination of uncertainty factors with different factor levels.

Figure 2 illustrates the scenario of reservoir modeling. Modeling-related data include G&G data, exploration well data, production well data, etc. These datasets are geographically distributed with the size of terabytes. With the help of the data replication tool, the base model is generated by extracting the modeling-related data. A reservoir engineer provides uncertainty factors and factor levels, which create the uncertainty parameter space. Based on this base model and the parameter space, massive reservoir models are constructed, each of which is associated with one combination of uncertainty factors and different factor levels. The number of models depends on the parameter space. Typically, the number reaches up to multiple thousands. These models are the inputs of missive reservoir simulation runs.

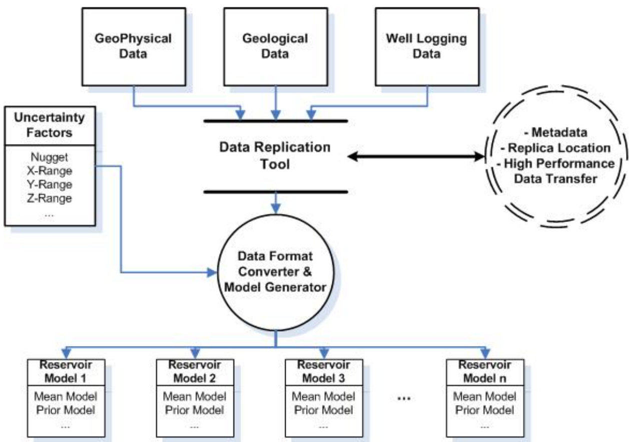

Figure 2. Reservoir modeling scenario

3.3. Massive Reservoir Simulations

The management of massive reservoir simulations across a Grid includes workflow definition of a single simulation, resource allocation, and simulation invocation.

Figure 3 shows the scenario of massive reservoir simulations. Task farming is engaged as the framework that takes reservoir models as inputs, checks a resource broker for resource allocation, and invokes massive simulations. Post process includes result analysis and visualization. Large-scale computation capability is involved in this scenario.

The workflow of a single simulation run integrates geostatistics algorithms with one execution of a reservoir simulator. Data conversion mechanism is developed between geostatistics algorithms and reservoir simulation execution. The definition of such a computational workflow is open to allow a user to specify his/her own computational model without change on any other component.

A resource broker was employed to manage Grid resources to share loads across a Grid. It captures the resource information and uses load balancing strategies to dispatch the simulation runs. Two major factors of a resource are considered: computational capability and architecture. There is a matrix adopted to describe the features

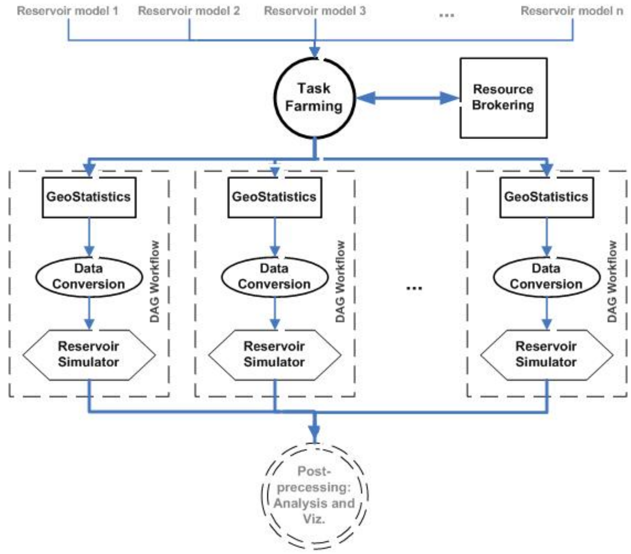

Figure 3. Massive reservoir simulations scenario

of a resource, which includes CPU number, CPU speed, CPU load averages, network bandwidth, memory size, local resource management system load, etc. Based on these features, a measurement can be taken to calculate the computational capability of a resource. The architecture factor is used to decide which type of binaries of geostatistics algorithms and reservoir simulators should be staged. In our current usecases, CPU number and CPU speed of a resource are critical because our adopted reservoir simulator and geostatistics algorithms are sequential programs. The computational capability of a resource C_i is measured as follows:

$$C_i = CPUNumber \times CPUSpeed$$

The load balancing strategy aims to dispatch certain number of simulations to a resource according to its computational capability. The following equation is used to decide the number of simulations N_i on a resource i:

$$N_i = T \times \frac{C_i}{\sum_{k=1}^{n} C_k}$$

where $n =$ the number of all available resources; $C_k =$ the computational capability of a resource k; $T =$ the total simulation number of uncertainty analysis.

Task farming is the framework to invoke massive simulation runs. Uncertainty analysis requires a number of nearly identical simulation runs with different models to produce the meaningful results. Task farming across a Grid is such a mechanism to utilize multiple resources to meet such a requirement.

This scenario was accomplished for use. The workflow of a single simulation is described by a perl script. The resource broker keeps track of which machines are available to run jobs, how the machines should be utilized, and when a machine is no longer available. A load balancing strategy was implemented to share massive reservoir simulation runs across a Grid. It assigns a value to each resource as its weight. Using the weight, the resource broker decides how many reservoir simulation runs should be dispatched on the resource. Condor-G [13] and Globus GRAM are employed to invoke the executions on remote resources. Condor-G lets one submit jobs into a queue and have a log detailing the life cycle of the jobs along with everything else expected from a job queuing system. Considering data transfer performance and large-scale storage requirement of simulation results, GridFTP is used instead of the input and output management provided by Condor-G (i.e., Globus GASS). Globus GRAM provides underlying software to utilize Grid resources, such as authentication and remote program execution.

3.4. Security

Data security is the major concern in reservoir studies. Reservoir modeling-related data, such as G&G data, well logging data, result data, etc. are very sensitive to petroleum engineering because of the potential commercial profits.

Our security solution is based on Grid Security Infrastructure (GSI) [14], the de facto Grid security standard. The GSI provides robust security mechanisms. It includes an OpenSSL implementation. It also provides a single sign-on mechanism, so that once a user is authenticated, a proxy certificate is created. With this certificate, a reservoir engineer can perform data operations within the Grid securely. The operations include pre-staging, modeling, lauching simulations, analyzing results, data archiving, etc. Sensitive data cannot be accessed without authorization and authentication.

4. Case Studies

Our design is being developed in close coordination with researchers from Petroleum Engineering Department at LSU and the first application is to compare three different stochastic simulation algorithms according to their flow reponses. It is a common way to create permeability fields by stochastic simulation algorithms, which honors the available information at the wells and reproduces the pattern of spatial variability between wells. The stochastic simulation can be categorized into direct (LU matrix Gaussian Simulation) and sequential approaches (Sequential Gaussian Simlation). LU matrix Gaussian Simulation (LUSIM) is rigorous but slow. Sequential Gaussian Simulation (SGSIM) is quicker but inaccurate. We create a hybrid simulation (HYBRID) to take the advantages of the direct and sequential approaches. The flow response is expensive but more important and precise to find differences among these simulatoins. In our experimental design and response surface model, four geological factors (e.g. nugget effect, x-range, y-range and z-range) have four levels to cover the all feasible factor values. Four-level full factorial design requires 256 simulation runs for each algorithm. Five realizations are created for each geostatistical parameter combination. The total simulation run is $3840 (= 256 \times 5 \times 3)$. All the simulation runs at CCT Grid testbeds, including two Linux clusters, helix (256 nodes) and supermike (1024 nodes). Sweep efficiency, break through time and upscaled permeability are extracted as responses from the summary files. Multiple linear regressions fits response surface models for the four factors and three responses in three directions. Main effects, interacting effects and quadratic effects are obtained (14 regression coefficients). Several points are concluded after the study:

- LUSIM permeability fields give the best prediction for all the responses, but the difference between LUSIM and HYBRID is rather small. LUSIM is much more time-consuming than HYBRID.
- All factors are significant for at least one response.
- Most interaction terms are insignificant.
- All factors have significant quadratic terms for at least one response.
- All algorithms are most sensitive to along the correlation, rather than diagonal and cross the correlation.

The solutions adopting Grid computing technologies are also being used for coastal modeling in the SCOOP project [15]. Our study shows that coastal modeling scenario has many similarities to reservoir uncertainty analysis, which needs to archive large-scale datasets and execute massive simulations.

5. Related Work

An autonomic reservoir framework [16] has been studied by W. Bangerth, H. Klie, etc. A prototype application was designed and developed to use P2P interactions between applications and services on the Grid to enable the automonic optimization of an oil reservoir. It optimized the placement and operation of oil wells to maximize overall revenue. The application consisted of instances of distributed multi-model, multi-block reservoir simulation components provided by IPARS, simulated annealing based optimization services provided by VFSA, economic modeling services, and experts connected via pervasive collaorative portals. This framework emphasizes the optimization and the integration of high level services of reservoir management, such as well placement and economical influence. Our efforts focus on reservoir performance prediction and uncertainty analysis based on the G&G characteristics of a reservoir.

COUGAR [17] is an industrial contribution on reservoir studies. It is a reservoir uncertainty analysis tool with the ability to make use of Grid resources to run a number of reservoir simulations and achieve the reduction in the individual result turnaround time. COUGAR takes as an input the ECLIPSE® [18] reservoir software model, and then uses experimental design to select a set of reservoir simulations runs to encompass the uncertain domain. However, it does not address large-scale G&G data integration, and its framework is tied to commercial packages, such as LSF® [19] for execution invocation, ECLIPSE®for reservoir simulator, and security issues are not considered. Our work integrates G&G data and well logging data, and provides an open, generic task farming framework to solve reservoir uncertainty analysis with open source software packages and a tight security consideration.

6. Conclusions and Future Work

Our work aims to leverage Grid computing technologies for reservoir uncertainty analysis. Grid solutions for large-scale data management and massive reservoir simulations have been presented and implemented.

A data replication tool was implemented using the GAT. A reservoir engineer can archive different G&G datasets and well logging data, which are required for construction of reservoir simulation base model. Based on this tool and the uncertainty factor parameter space generation mechanism, a reservoir modeling mechanism for uncertainty analysis was developed.

A Grid resource broker and a task farming framework have been developed. The resource broker captures the Grid resource information and uses load balancing strategies to dispatch the reservoir simulations across a Grid. Using task framing, the reservoir simulation runs have been combined with geostatistics algorithms and resource management on a Grid.

Reservoir researchers from Petroleum Engineering Department, LSU have adopted these Grid solutions for their uncertainty studies. Coastal modeling researchers from the SCOOP project are using these solutions in their development of an integrated infrastructure for ocean observing and prediction.

However, there remains much work ahead on Grid-enable uncertainty analysis. Our long-term objective is to provide an integrated and easy-to-use problem solving environment for uncertainty analysis. Currently, a Grid portal is under development, which will provide an easy-to-use interface for a user. Efforts are underway to provide monitoring and steering capabilities at runtime during the execution of a given simulation run, to provide the possibility to check job status and terminate the job if an error occurs. Another challenging issue in our future work is how to couple the latest optimization algorithms to reduce the problem scale without the risk of losing correct conclusions.

7. Acknowledgements

We acknowledge the Grid testbeds provided by CCT, LSU. This research is supported by DOE DE-FG02-04ER46136 and the CCT.

References

[1] K. Aziz, A. Settari. "Petroleum Reservoir Simulation". *Applied Science Publishers Ltd.*, London, 1979.

[2] C.D.White, B.J. Willis, et al. "Identifying and Estimating Significant Geologic Parameters With Experimental Design". *SPE Annual Technical Conference and Exhibition, Dallas, Texas, U.S.A.*, October 1-4, 2000.

[3] S.Kalla, C.D.White, "Efficient Design of Reservoir Simulation Studies for Development and Optimization". *SPE Annual Technical Conference and Exhibition, Dallas, Texas, U.S.A.*, October 9-12, 2005.

[4] I. Foster, C. Kesselman, S. Tuecke. "The anatomy of the Grid: enabling scalable virtual organizations". *International Journal of High Performance Computing Applications, 15(3):200-222*, 2001.

[5] C.D. White, S.A. Royer. "Experimental Design as a Framework for Reservoir Studies". *2003 SPE Reservoir Simulation Symposium, Houston, Texas, U.S.A.*, February 3-5, 2003.

[6] J. Zhang, M. Delshad, et al. "An Efficient Reservoir Simulation Approach to Design and Optimize Improved Oil Recovery Processes with Distributed Computing". *SPE Latin American and Caribbean Petroleum Engineering Conference, Rio de Janeiro, Brazil*, June 20-23, 2005.

[7] R.W. Schulze-Reigert, J.K. Axmann, et at. "Evolutionary Algorithms Applied to History Matching of Complex Reservoirs". *2001 SPE Reservoir Simulation Symposium, Houston, Texas, U.S.A.*, February 11-14, 2001.

[8] G. Allen, K. Davis, et al. "The Grid Application Toolkit: Towards Generic and Easy Application Programming Interfaces for the Grid". *Proceedings of the IEEE, Vol. 93, No. 3, pp. 534-550*, March 2005.

[9] MCAT. *http://www.npaci.edu/DICE/SRB/mcat.html*.

[10] I. Foster, C. Kesselman. "Globus: A toolkit-based Grid architecture". *The Grid: Blueprint for a New Computing Infrastructure,* Morgan Kaufmann Publishers, San Francisco, CA, pp. 259-278, 1999.

[11] CURL and LIBCURL. *http://curl.haxx.se/*.

[12] A. Rajasekar, M. Wan, et al. "Storage Resource Broker - Managing Distributed Data in a Grid". *Computer Society of*

India Journal, Special Issue on SAN, Vol. 33, No. 4, pp. 42-54, Oct 2003.

[13] D. Thain, T. Tannenbaum, M. Livny. "Condor and the Grid", *Grid Computing: Making the Global Infrastructure a Reality*. John Wiley, 2003. ISBN: 0-470-85319-0.

[14] I. Foster, C. Kesselman, et al. "A security architecture for computational grids". *Proceedings of the 5th ACM Conference on Computer and Communications Security, San Francisco, CA, pp. 83-92*, 1998.

[15] SURA Coastal Ocean Observing and Prediction Program (SCOOP). *http://www1.sura.org/3000/3300Coa stal.html*.

[16] W. Bangerth, H. Klie, et al. "An Autonomic Reservoir Framework for the Stochastic Optimization of Well Placement". *Cluster Computing: The Journal of Networks, Software Tools, and Applications, Special Issue on Challenges of Large Applications in Distributed Environments*, Kluwer Academic Publishers, Vol. 8, No. 4, pp. 255-269, 2005.

[17] COUGAR. *http://www.slb.com/media/services/software/reseng/ressim_ondemand.pdf*.

[18] ECLIPSE. *http://www.slb.com/content/services/software/reseng/eclipse_simulators*.

[19] LSF. *http://www.platform.com/Products/Platform.LSF.Family*.