A Generic Task-Farming Framework for Reservoir Analysis in a Grid Environment

Santiago Pena, Dayong Huang, Zhou Lei, Gabrielle Allen
Center for Computation & Technology, Louisiana State University, U.S.A.
Xin Li, Chris D. White
Department of Petroleum Engineering, Louisiana State University, U.S.A.

Abstract

The vast majority of research projects involving Grid Computing have focused on the development and standardization of the middleware that allows Grids to function. Despite the maturity of the middleware as the Grid core, the actual application of Grids to solving concrete problems is far from being a common practice, much less a standard one. In this context, we present ResGrid, a Grid-aware toolkit for reservoir uncertainty analysis. With the help of ResGrid, a reservoir engineer can transparently take advantage of Grid resources and services for compute-intensive and data-intensive uncertainty analysis while achieving a better understanding of multiphase reservoir modeling. ResGrid improves the efficiency of reservoir analysis, and its flexible architecture allows its utilization by other application areas such as coastal modeling and climate simulations. ResGrid is our first step towards developing generic Grid-aware toolkits that could be applied to any science in a cost- and time-efficient manner.

1. Introduction

Optimal reservoir management requires accurate and realistic models of the reservoirs, preferably captured as quickly and economically as possible. Reservoir simulation is a tool utilized to obtain accurate models of gas and oil reservoirs. This type of simulation combines geology, geophysics, numerical methods, and high performance computing. A key tool in predicting reservoir performance is uncertainty analysis, which requires massive reservoir simulations with different models. In order to perform the simulations, scientists must use computers to assist them in building dynamic models that predict the movement of oil and gas flowing under the surface of the earth. Due to the complexity and the large set of variables that must be taken into account in this process (such as permeability, density porosity, saturation, etc.), reservoir modeling is considered a highly computation-intensive problem. Ideally, one would like to perform an exhaustive evaluation of reservoir simulations with uncertainty factors to characterize the impacts, but such an endeavor is computationally infeasible.

A very promising technology that has the potential to address the computational challenges presented in resource-intensive simulations is Grid Computing [8]. Grid Computing is a distributed system or environment formed by several domains or entities, in which each entity can act as a provider and client indistinctly, and that allows its members to access and coordinate high-end computational capabilities across different administrative domains. Around the world, there have been many efforts to research and develop the core middleware to support Grid communities. Technologies such as the Globus Toolkit [11, 1], Grid Security Infrastructure (GSI) [12], GridSphere [13], and Condor [14], are considered to be the Grid Computing standards. However there are only a few Grid-enabled applications that fully exploit the potential of Grid environments. This is due to the constant change, complexity, and evolution of the Grid. There is a clear need for a high-level application that allows programmers to deal with the different Grid resources without requiring a change in their preferred software applications.

Our work focuses on the design and development of an integrated problem-solving environment for reservoir uncertainty analysis called ResGrid. ResGrid is a Grid-aware toolkit for reservoir uncertainty analysis. This toolkit improves the efficiency and accuracy of reservoir analysis in a cost-efficient manner. ResGrid has been designed with a highly flexible and adaptable architecture that allows its utilization by other application areas, such as coastal modeling and climate simulations, with only minor modifications to the code. In this context, ResGrid is our first step towards developing generic Grid-aware toolkits that could be applied to any science in a cost- and time-efficient manner.

This paper is organized as follows. In Section 2, we provide an introduction to reservoir uncertainty analysis and describe a typical use-case scenario along with its current limitations. In Section 3, we provide an overview of the architecture and implementation of the ResGrid software. Section 4 discusses related work, Section 5 summarizes the results obtained at the time of this writing, and finally Section 6 provides conclusions and details of future work.

2. Reservoir Uncertainty Analysis

2.1. Problem Description

Reservoir simulation [9] is the main approach for characterizing a reservoir in the planning and evaluation of sequential development phases. A reservoir can be represented by a mathematical model; however, to obtain an analytical solution of a reservoir, numerical simulation is required. The number of simulation runs involved in these kinds of factorial designs is very large, particularly if many uncertainty factors are considered. This has motivated attempts to improve the computation technologies and optimization studies utilized.

Uncertainty analysis and sensitivity studies play a key role in reservoir performance prediction. In addition, experimental design and response surface methodology [15] 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 forecasts.

Prior to investments, petroleum exploration and production engineers must be able to identify the reservoir characteristics and various uncertainty factors, and then quantify and analyze these uncertainty factors in the data acquisition program. A reservoir simulation involves numerous fields of study and consists of the following steps:

1. Geologists build a 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 base 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.

2.2. Typical Uncertainty Analysis Scenario

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. However, there is no integrated, secure, and easy-to-use problem-solving environment available for use, although some efforts [16] have been made. A reservoir engineer is required to manually make these toolkits communicate and work together.

Figure 1 demonstrates a typical reservoir uncertainty analysis scenario. First, a reservoir engineer specifies the uncertainty factors and these factor levels. Once the reservoir models associated with the combinations of uncertainty factors and levels are constructed, the user submits them to a high performance computing facility (e.g., a cluster) for multiple simulation runs. A batch file or the job scheduler at the facility invokes multiple executions of geostatistics algorithms and a reservoir simulator. Simulation results are then moved to a data storage device. Post-processing software is used to view and analyze the results. This requires that the user locate the data files containing the appropriate results in the data storage device, and move the files to a local machine using data transfer tools, such as FTP. Typically, analysis and visualization of results make use of Windows-based commercial software packages.

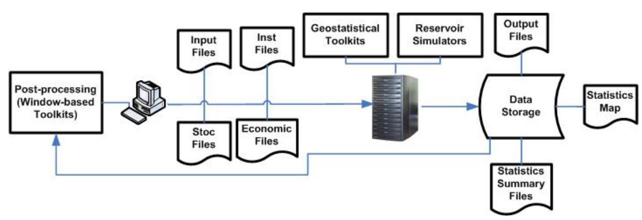

Figure 1. Workflow involved in a typical reservoir uncertainty analysis scenario

2.3. Limitations

During the uncertainty analysis process, a reservoir engineer needs to minimize the number of uncertainty factors and the factor levels considered. This common practice causes a loss of accuracy in the predictions produced by the model. The reason for this reduction in the number of variables is the fact that running a full-scale simulation may take up to 100 days in a 1024 processor cluster, and the output of such run could reach the order of Petabytes. Therefore, it is clear that a single high-performance computing facility cannot satisfy such requirements. In addition, there is no easy-to-use integrated environment for uncertainty analysis.

The lack of an integrated solving environment is another issue that limits studies of reservoir uncertainty analysis. A reservoir engineer needs to manually handle all stages of the process, including provisioning, staging, visualization, result retrieval, and sensitivity analysis.

Security issues hinder effective collaborations between researchers interested in reservoir studies due to the potential commercial benefits of their results.

3. ResGrid Toolkit for Reservoir Uncertainty Analysis

ResGrid is a Grid-aware experimental toolkit designed and implemented with the objective of facilitating and enhancing reservoir modeling. Moreover, ResGrid's design would allow us to adapt this toolkit to offer solutions to other areas of research with minor modifications to the original code. It is our goal to develop a generic toolkit for massive simulations that makes efficient use of the existing Grid tools.

3.1. Use-Case

The ResGrid provides a reservoir engineer with a problem-solving environment for uncertainty studies. It allows a user to conveniently collect geological and geophysical data, specify the uncertainty parameter space, invoke numerical reservoir simulations across the Grid, and analyze and visualize simulation results.

The ResGrid use case scenario is illustrated in Figure 2. Typically, there are ten steps:

1. A user logs into the ResGrid portal and retrieves a GSI certificate from a proxy server. The certificate authorizes the user to access the Grid resources and implement secure data transfer.
2. The user specifies the uncertainty factor parameter space and the size of the reservoir grid block, which will be used for reservoir model construction and result analysis.
3. By clicking on the "Launch" button, the user invokes the execution of the ResGrid services.
4. The first service is reservoir modeling. This triggers a data-archiving tool and analyzes the uncertainty factor parameter space specified by the user in Step 2.
5. The modeling service constructs reservoir models and starts the Grid resource-brokering service.
6. The resource-brokering service captures the dynamic information from the information service provided by the Grid, makes a decision on the appropriate resource for each single simulation run with the help of load balancing strategies, and then calls the massive simulation execution service.
7. The simulation executions are invoked to all the resources available on the Grid.
8. Once all the simulation runs have been completed, the sensitivity analysis service is activated to analyze the simulation results.
9. The visualization service visualizes the simulation results from the sensitivity analysis.
10. The user views on the ResGrid portal the results generated by the visualization service.

This scenario illustrates that the workload of a reservoir engineer is reduced, because all the user needs to do is interact with a Web-based Grid portal designed for uncertainty analysis. The ResGrid services take care of security, data acquisition, resource management, result analysis, and visualization. The reservoir engineer is no longer required to manually manage any of these activities when using ResGrid for uncertainty analysis.

One of the benefits of this approach is that a reservoir engineer could start a massive simulation from his web-enabled phone using newly acquired data obtained from the drilling sensors, and can monitor the status and even visualize the results with his/her phone. In this scenario, the engineer does not need to leave the site in order to perform new simulations or steer existent ones based on the new data.

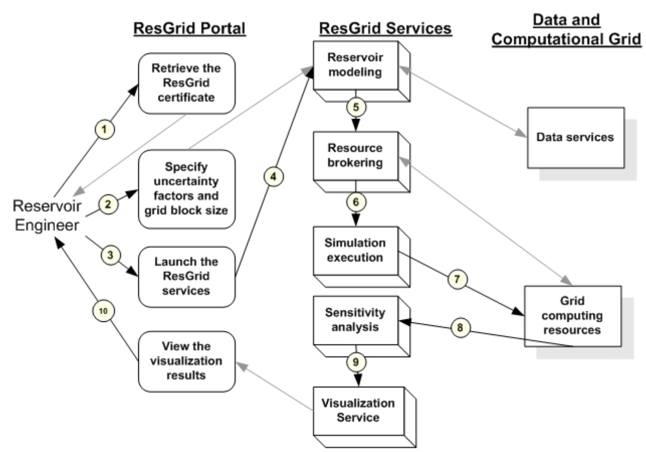

Figure 2. ResGrid usage scenario for a reservoir engineer

3.2. Architecture

The ten steps previously described in Section 3.1 are accomplished by four general components: *ResGrid Portal*, *Data Preparation*, *Massive Simulation* and *Post-Processing Routines*. Figure 3 depicts the interaction of these four components.

Figure 3. Components of ResGrid

ResGrid Portal

The ResGrid portal (based on GridSphere [13]) provides the entry point to the ResGrid, unifying Grid portal technologies and providing a web-based user interface. First, the portal asks for credentials from the user to establish his/her identity. A GSI certificate is retrieved to provide authentication to access Grid resources and secure data transfer across the Grid environment. Second, the ResGrid portal provides interfaces to specify the uncertainty factors and the problem-solving scales (e.g., the grid block size of a reservoir). Third, a user can view and visualize results via this Grid portal.

As mentioned earlier, one of the goals of our work is to design a highly flexible toolkit that could be applied to other areas. To achieve this goal, special attention is paid to the development of a portal that can be quickly adapted to provide new features for reservoir simulation, or that can serve as an interface to new projects. The current architecture of the portal is depicted in Figure 4. This architecture design is still a work in progress. Further research is needed in this area to guarantee the flexibility and adaptability features of the portal.

Data Preparation

This component archives distributed modeling-related data, creates the uncertainty parameter space, and constructs the reservoir models. Figure 5 shows the structure of the Reservoir Modeling component.

A data archiving tool was designed for data acquisition. There are three modules in this tool: metadata service, replica location service, and high performance data transfer service. The mechanism employed by this tool is as follows: Given the information describing the required data, the metadata service retrieves logical file names, the replica location service locates the physical files which map to the logical filenames, and then these physical files are relocated via high-performance data transfer. These modules are implemented as interfaces to other Grid middleware services.

Modeling-related data include geological and geophysical data, exploration well data, production well data, etc. These datasets, often Terabytes and even Petabytes in size, are geographically distributed. With the help of the data archiving tool, a base model is generated by extracting the modeling-related data. Uncertainty factors and factor levels are provided and the uncertainty parameter space is created. Utilizing 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, it is up to multiple thousands). These models are stored in separate files that are used as the input of massive reservoir simulation runs.

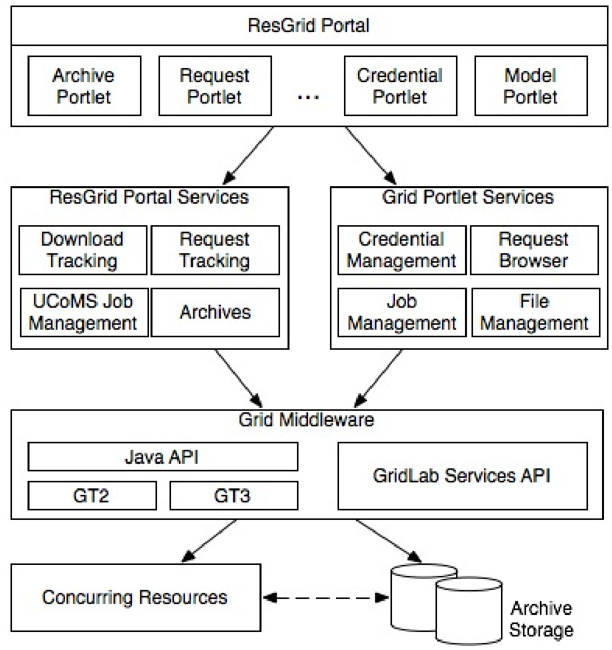

Figure 4. ResGrid Portal Architecture

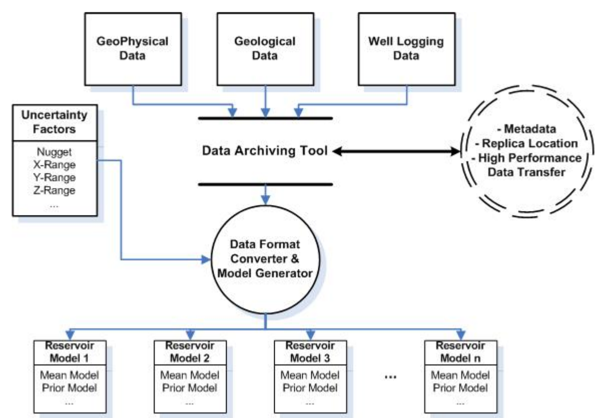

Figure 5. Components of Reservoir Modeling

Massive Simulation

The Massive Reservoir Simulation component is in charge of massive simulations management, which includes workflow determination for a single reservoir simulation, resource allocation, and massive simulation invocation.

Task farming is the framework used to invoke massive simulation runs. As previously mentioned, the basic idea of uncertainty analysis is to run massive reservoir simulations with different models, which requires a number of nearly identical runs to produce the meaningful results. In order to perform task farming over the Grid, we are using Condor-G [2] to distribute different reservoir models among various Grid resources.

Condor-G provides resource selection by the use of matchmaking with ClassAds [3]. However, the matchmaking service does not provide a sophisticated load balancing system that can implement high-level user policies. In our case, we developed a set of scripts that are executed before the submission of jobs to Condor-G that can apply different load-balancing policies. In this way, we can decide what percentage of the total jobs will run on each available resource or assign priorities to different resources. More complex policies need to be implemented in order to add a maximum of flexibility to our simulations.

Visualization

Visualization is used to graphically demonstrate uncertainty analysis results and assist decision-making for further analysis. A reservoir engineer interacts with the visualization component via the ResGrid portal to obtain the visualization images dynamically and according to the user's configuration.

3.3. Implementation Status

The core components (Data Preparation and Massive Simulation) have been fully implemented. These components have been used by reservoir researchers and petroleum engineers at Louisiana State University (LSU) to perform their simulations. The toolkit was used with a command-line tool without the portal interface. The LSU experience provided us with valuable feedback that is allowing us to improve ResGrid to the point that it can be easily used not only with the portal, but also in command-line mode. In addition, we are working on the implementation of load-balancing policies that would allow us to improve efficiency, increase flexibility and give the user better control to decide what, where, and when to run a simulation.

The ResGrid portal interface is still under development. GridSphere provides credential retrieval, job submission, and file browser portlets [13] that can be customized to fit our application. The biggest challenge for the portal implementation is to design it in such a way that it is flexible and can be easily adapted to other areas of research.

The Visualization component of the ResGrid is still under development due to its tight coupling with the portal. HDF5 [4] is adopted for data format and OpenDX [5] to display the results.

The security of all the ResGrid services is based on GSI. GSI provides robust security mechanisms including 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 operations within the Grid securely.

4. Related Work

An autonomic reservoir framework [10] has been studied by Bangerth et. al. 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 collaborative 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 characteristics of a reservoir.

COUGAR [6] is an industry-developed package for 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. However, it does not address large-scale data integration, and its framework is tied to commercial packages, such as LSF® for execution invocation, ECLIPSE® for reservoir simulation, and security issues are not considered. In addition, the concept of Grid used in COUGAR is limited to IBM's Grid, which in turn is a high-performance computing facility. Therefore, it can be argued that COUGAR is not really a Grid application, but just task farming in IBM's facility. As a consequence, scalability could become a factor since the computational power available depends on what IBM can offer, which may be a large amount, but does not compare to the unbounded computational power of the real Grid. When fully deployed, ResGrid will provide an open generic framework to solve reservoir uncertainty analysis with open source software packages and security considerations. Additionally, the portal interface would allow scientists to easily manage and start simulations and even review the results from their web-enabled phones without leaving the actual drilling site.

Moreover, the flexible design of ResGrid allows scientists to easily add functionality to perform customized simulations.

5. Results

We constructed 27 RSMs (Response Surface Model) for three responses of three geostatistical algorithms (Sequential Gaussian Simulations, LU Decomposition, and Hybrid) in three directions totaling 5120 runs. The average execution time of each run is around 2 hours, which would yield a total execution time of 10240 hours if the jobs were run sequentially. With ResGrid the 5120 runs were executed in a time period of 480 hours, producing a dramatic increase in throughput.

The analysis of the results obtained escapes the scope of this work. However, it is worth mentioning that our results indicate that the sequential simulation model may overestimate stochastic fluctuations, compared with the more rigorous LU-decomposition methods.

6. Conclusions and Future Work

ResGrid provides an integrated, flexible, scalable and easy-to-use problem solving environment for reservoir uncertainty analysis. The current command-line implementation of ResGrid has been adopted by the reservoir researchers from the Petroleum Engineering Department at LSU. With this new software, the scientists can perform a simulation in one week, whereas in the past, the same simulation took 160 days. Moreover, the data manipulation and scripting skills required to run jobs with ResGrid have been reduced to a minimum. In this way, the researchers can concentrate on the science aspect of the reservoir simulation instead of wasting time and resources on implementation details. In addition, ResGrid's flexible design has allowed researchers from the SCOOP project [7] to use parts of the toolkit in their development of an integrated infrastructure for coastal modeling.

The development of the portal and visualization components of ResGrid constitute the main focus for our efforts in the near future. In order to design a portal application that could be extended, modified, and used with minimal effort, we will develop job submission and job monitoring portlet-templates which can be easily customized to be used by other applications. Additionally, a better understanding of the the user interface requirements needs to be gained. In this context, the feedback obtained from the researchers using the current command-line implementation of ResGrid is invaluable.

In the long term, we plan to test the flexibility aspect of our toolkit by applying it to other research areas that need massive simulations such as Physics, Coastal Modeling or Climate Simulation. In doing so, we believe that we will gain the experience required to produce a generic toolkit for massive simulations in the Grid.

7. Acknowledgements

We acknowledge the use of Grid testbeds provided by the Center for Computation & Technology (CCT). This research is supported by the UCOMS grant (DOE DE-FG02-04ER46136) and the CCT.

References

[1] Globus Toolkit `http://www.globus.org`.
[2] Condor-G `http://www.cs.wisc.edu/condor/condorg/`.
[3] Condor Matchmaking `http://www.cs.wisc.edu/condor/manual/v6.2/2_3Condor_Matchmaking.html`.
[4] HDF5 `http://hdf.ncsa.uiuc.edu/HDF5/`.
[5] OpenDX `http://www.opendx.org`.
[6] COUGAR `http://www.slb.com/media/services/software/reseng/ressim_ondemand.pdf`.
[7] SURA Coastal Ocean Observing and Prediction Program `http://scoop.sura.org`.
[8] G. Allen and K. Davis. "Enabling applications on the Grid: A GridLab overview'. Special issue on Grid Computing: Infrastructure and Applications, August 2003, 2003.
[9] K. Aziz and A. Settari. "Petroleum Reservoir Simulation". *Applied Science Publishers Ltd.*, 1979.
[10] W. Bangerth and H. Klie. "An Autonomic Reservoir Framework for the Stochastic Optimization of Well Placement". *The Journal of Networks, Software Tools, and Applications, Special Issue on Challenges of Large Applications in Distributed Environments*, 8(4):255–269, 2005.
[11] I. Foster and C. Kessellman. *The Grid: Blueprint for a New Computing Infrastructure*, chapter "Globus: A toolkit-based Grid architecture". Morgan Kaufmann. Second Edition.
[12] I. Foster and C. Kessellman. "A security architecture for computational grids". 1998.
[13] J. Novotny, M. Russell, and O. Wehrens. "Gridsphere, A Portal Framwork for Building Collaborations". *IEEE Concurrency and Practice*, 2004.
[14] D. Thain, T. Tannenbaum, and M. Livny. *Grid Computing: Making the Global Infrastructure a Reality*, chapter "Condor and the Grid".
[15] C. White and S. Royer. "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, 2005.
[16] J. Zhang and M. Delshad. "Experimental Design as a Framework for Reservoir Studies". 2003 SPE Reservoir Simulation Symposium, February 3-5, 2003.