A Workflow Approach to Designed Reservoir Study

Gabrielle Allen, Promita Chakraborty, Dayong Huang, Zhou Lei,
John Lewis, Xin Li, Christopher D White, Xiaoxi Xu, Chongjie Zhang
Center for Computation & Technology
Louisiana State University
Baton Rouge, LA 70803, USA
gallen@cct.lsu.edu, promita@cct.lsu.edu, dayong@cct.lsu.edu,
zlei@cct.lsu.edu, jlewis@cct.lsu.edu, xli@cct.lsu.edu,
cdwhite@cct.lsu.edu, theresaxu@cct.lsu.edu, czhang@cct.lsu.edu

ABSTRACT

Reservoir simulations are commonly used to predict the performance of oil and gas reservoirs, taking into account a myriad of uncertainties in the geophysical structure of the reservoir as well as operational factors such as well location. Designed reservoir study provides a robust tool to quantify the impact of uncertainties in model input variables, and can be used to simulate, analyze, and optimize reservoir development. However, such studies are computationally challenging, involving massive (terabyte or petabyte) geographically distributed datasets and requiring hundreds or tens of thousands of simulation runs. Providing petroleum engineers with integrated workflow through a secure and easy-to-use user interface will enable new advanced reservoir studies. This paper describes the workflow solution and user interface designed and implemented for reservoir uncertainty analysis in the UCoMS project (Ubiquitous Computing and Monitoring System for discovery and management of energy resources).

Categories and Subject Descriptors

J.2 [**Computer Applications**]: Physical Sciences And Engineering; I.6 [**Computing Methodologies**]: Simulation and Modeling

General Terms

Design, Experimentation, Measurement

1. INTRODUCTION

Designed reservoir study is based on the numerical flow simulation of many representative models of possible scenarios for the reservoir, through combining the attributes that characterize the reservoir with the uncertainty in the value of the attributes. The advantages of such studies are clear: designed suites of models give lower estimation errors compared with conventional one-at-a-time sensitivity studies; the main effects of uncertainty factors and their interactions can be assessed; response models resulting from designed simulation can be efficiently used as a proxy model for a reservoir simulator; hypotheses can be tested; models can be discriminated between. Thus, designed reservoir study is regarded as an effective tool to improve the process of reservoir study. However, reservoir simulations are notorious for their computational cost and their volume and variety of output. The major technical challenges in designed reservoir study [12] can be summarized as followed:

1. The need for dataset management. Large-scale (terabyte or petabyte) geographically distributed datasets are involved in the generation of the basic reservoir model. The model-related datasets include geological & geophysical data and well-logging data;
2. The need to rapidly and repeatedly perform many hundreds or tens of thousands of time-consuming simulations with different reservoir models to quantify the impacts of different uncertainty factors. A single high performance computing facility cannot satisfy the requirements of massive reservoir simulations runs;
3. The need for an integrated, secure, easy-to-use user interface for uncertainty analysis. Reservoir engineers now manually handle all stages of uncertainty analysis, including provisioning, staging, result retrieval, and post processing.

The UCoMS project [10] (Ubiquitous Computing and Monitoring System) is researching and developing new grid computing and sensor network technologies for the management of energy resources. One goal of this project is to support computation-intensive, fine-grained simulations and enable a huge amount of measured data storage and real-time processing, while providing safety monitoring on well platforms. As the core grid-aware toolkit in the UCoMS project, the ResGrid [13] addresses the needs for large-scale data management and execution support for reservoir uncertainty analysis. The ResGrid uses the Grid Application Toolkit (GAT) [1] to access a wide range of grid services. Through the GAT API, data grid tools, such as metadata, replica and transfer services, are used to meet the massive size and geographically distributed characteristics of reservoir study data. Workflow, task farming, and grid resource allocation are used to support large-scale computation. Contemporary reservoir simulation tools include stochastic simulation, flow

WORKS'07, June 25, 2007, Monterey, California, USA.

modeling, and reservoir simulations, which together form a workflow. Load balancing strategies are adopted to allocate grid resources. A task farming method dispatches flow simulations with different reservoir models and configurations on various allocated compute resources. The ResGrid portal provides petroleum researchers with a central and intuitive Web-based interface to create, submit and manage workflows for reservoir simulations and track associated data files.

This paper describes the design and implementation of the workflow solution for reservoir uncertainty analysis, which integrates data management and computation support into a unified problem solving environment. The environment provides an easy-to-use, portlet-based workflow management application.

The remainder of this paper is organized as follows. Section 2 outlines some major workflow systems. Section 3 introduces the reservoir uncertainty analysis workflow. Section 4 describes the detail of the workflow design and implementation. Section 5 shows a portlet-based workflow management application. Section 6 presents the conclusions and looks towards possible future developments.

2. WORKFLOW SYSTEMS

There exist many workflow management systems for Grids.

Triana [17] is a visual workflow composition system where the workflow components are service-oriented. It consists of an intuitive graphical user interface and an underlying subsystem, which allows integration with multiple services and interfaces, including GAT, GAP [16], Globus, etc.

Pegasus [4] is a flexible framework that enables the plugging in of a variety of components from informaiton services and catalogs to resource and data selection algorithms.

Taverna [15] is a domain-specific system, which workflows are limited to the specification and execution of ad hoc *in silico* experiments using bioinformatics resources.

Kepler [2] is another system for composing and executing scietific workflows, in which a workflow is composed of independent actors communicating through well-defined interfaces and an actor represents parameterized operations that act on an input to produce an output.

The Chimera Virtual Data System (VDS) [5] is a system for deriving data rather than generating them explicitly form a workflow. It combines a virtual data catalog, for representing data derivation procedures and derived data, with a virtual data language interpreter that translates user requests into data definition and query operations on the database.

Motivated by various workflow management projects, we implemented a Grid-aware workflow for reservoir studies. It combines seismic inversion, reservoir modeling, flow numerical simulations, as well as massive data managment to perform reservoir studies (e.g., quantifying the impacts of uncertain factors).

3. DESIGNED RESERVOIR STUDY WORKFLOW

During the development phase for an oil and gas field, the limited information provided by sparsely distributed wells in the field is not enough to provide accurate and validated 3D geological models. The older techniques of evaluating risks and uncertainties by running sensitivities on various input parameters on the un-validated model on a one-at-a-time basis is considered inadequate since interactions between parameters, which could have significant impact on the forecast ultimate recovery, cannot be captured. Therefore, petroleum engineers are moving to adopt Experimental Design (ED) techniques to systematically quantify the impact of the model input variable uncertainties on the ultimate oil or gas recovery. As shown in Figure 1, the reservoir uncertainty analysis process involves four major sequential steps [19, 20]:

1. Reservoir Characterization — Static (hard and soft) data, such as geological, geophysical, and well log/core data are incorporated into geological models through conditional geostatistical simulation.
2. Reservoir Simulation Model Construction — Upscaled geological models and engineering data are used to build complex flow models to forecast the ultimate recovery. Reservoir performance predictions commonly consider many scenarios, cases, and realizations.
3. Reservoir Simulations — This consists of the following steps: geostatistical realizations are generated; other parameters like fluid flow, well locations and factors are used to form the models; each model is simulated to obtain production profiles and recovery factors; economic performance indicators, like ROI (Return on Investment) and NPV (Net Present Value) are calculated; further models are generated as the product of the base model and one combination of uncertainty factors (with different levels) generated from the simulation; thus the number of runs of reservoir simulation is directly proportional to the number of uncertainty factors. Experimental design [18] is adopted to identify the optimal settings for all the factors of interest.
4. Post Processing — A response surface model is constructed, and then using Monte Carlo Simulation the differences between the response surfaces can be analyzed with statistical methodologies, e.g. a χ^2-likelihood test.

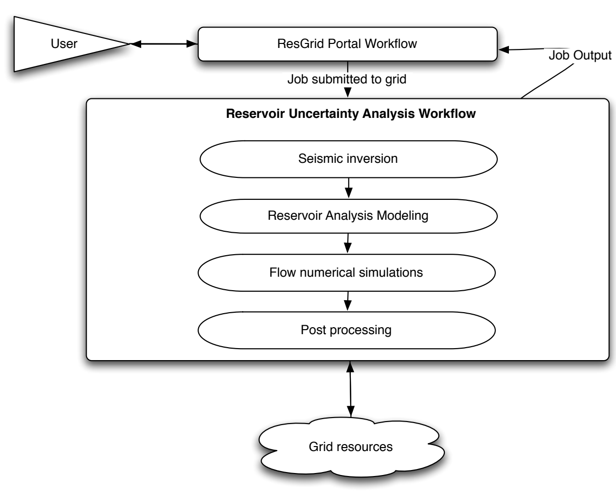

Figure 1: Reservoir uncertainty analysis workflow.

4. WORKFLOW IMPLEMENTATION

As the availability of computational power increases, scientists and engineers are able to run more and more complex jobs. As these jobs get more intricate and distributed, it is becoming increasingly difficult to keep track of the flow within the grids, clusters, data storage systems and archives. A good example is reservoir uncertainty analysis in the UCoMS project which involves handling of complex jobs to analyze reservoir uncertain factors to improve performance prediction. An efficient workflow description can smoothly keep track of the information and data flow within the system.

Figure 2 shows the implementation diagram of the reservoir uncertainty analysis workflow. The first step is to input initial parameters and generate reservoir models. Multiple models will be used by simulations. The second step is to select a reservoir simulator and one (or more) geoscience simulation algorithms. The third step is the deployment of a large number of reservoir simulations across grid resources. Finally, data archiving and analysis are conducted.

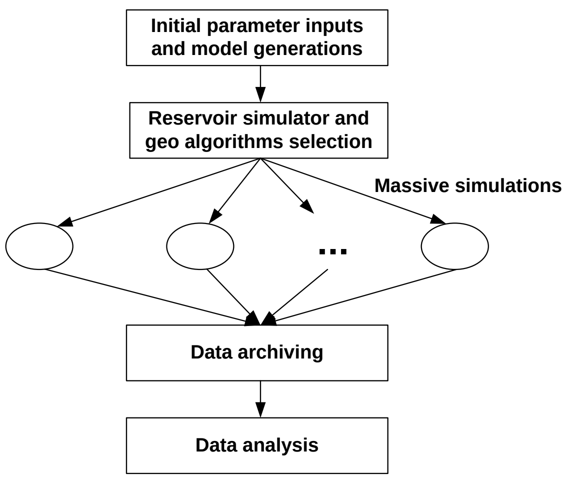

Figure 2: Implementation diagram for reservoir uncertainty analysis workflow.

4.1 Task Farming

A task farming framework is used to take reservoir models as inputs, check a resource broker for resource allocation, and invoke simulations. Each single simulation integrates geostatistics algorithms with reservoir simulation. Data conversion is provided between the chosen geostatistics algorithm and reservoir simulation execution. The specific configuration of such a computational workflow is left open to allow a user to specify his/her own computational model without need to change other components.

This task farming based framework has four modules for: resource brokering, staging in/out, invocation, and status monitoring.

The Resource Brokering module manages the grid resources and shares loads across the grid. It accesses external information services and extracts resource information of interest into a list. Each item of this list represents a resource available in a grid. Most functions of execution management, such as load balancing and staging in/out, rely on this resource list. The Staging In/Out module uploads model datasets and executables to and downloads simulation results from a particular resource. To upload executables, this module needs to check the Resource Brokering module to obtain the type of operating system on a particular remote resource. By this way, the module decides which kind of executable binaries are needed. Retrieving load balancing calculation of the Resource Brokering module, the Staging In/Out module is aware of how many and which simulation models should be run on the resource. This module also needs to figure out work directory used on the resource. Once obtaining the required information, this module transfers the datasets with security. The staging out procedure is similar to the staging in procedure. It downloads the simulation results from remote sites. The Invocation module needs to handle remote execution. This module is to communicate with various local resource management systems (LRMS) on remote resources and invoke simulation executions on the corresponding LRMS queues. The Status Monitoring module is in charge of the communication with LRMSs. There are two levels of queues for status monitoring: resource queue on submission machine and LRMS job queue on each remote resource. Each resource that is running simulations has an entry in the resource queue. On a particular resource, the job queue of LRMS is checked periodically. Once all the simulations dispatched to the resource have been accomplished, the corresponding resource entry in the resource queue is removed.

4.2 Archive

A number of reservoir simulations are involved in a typical uncertainty analysis workflow, which generate large-scale simulation result datasets. The result dataset of a single simulation depends on the configuration of the simulator. An average size reaches up to 50 Megabytes or so. Massive simulations lead to storage needs which cannot easily accommodated with a typical storage resource. Therefore, to facilitate future comparison and analysis among different runs and between different users, a data archive is being developed to store simulation results.

The archive system is implemented using a client-server model in a grid environment. The archive clients are deployed on the supercomputers and clusters where the simulations are conducted. Once the simulation is complete, the simulation results are transferred to an archive server asynchronously using the transfer protocol of choice with a set of meta data. The transfer protocol is coded into the archive client using GAT. Application coded with GAT enables a user (or the client code) to choose the middleware to use at run time. The archive server provides data integrity in the grid environment. It uses an atomic transaction mechanism to wrap each dataset transfer. The transaction control messages are transferred using SOAP messages. The physical data transfer call also uses the GAT which provides flexibility to use different transfer protocols.

5. WORKFLOW MANAGEMENT

We use XML to describe the UCoMS workflow. Since the workflow is specialized for reservoir uncertainty analysis, the workflow XML description is relatively simple, containing a parameter description and required environment variables for reservior simulations. To provide an easy inferface for users (mainly, petroleum engineers), a portlet-based application for workflow management was developed, including capabilities for workflow creation, deletion, tracking and reuse.

5.1 Portlet Development

The workflow description created through the portlets is submitted to a remote machine for execution. Grid portlet services are used to simplify authentication and job submission. There are a number of grid portal frameworks and toolkits, including GridSphere [14], GridPort [3], OGCE [9], and Java COG [11]. Based on our comparative analysis of GridPortlets and OGCE [21], we choose GridSphere and GridPortlets as our main framework and toolkit to speed up the process of developing and deploying the ResGrid portal.

GridSphere is an open-source portal framework, developed by the European GridLab project [7]. It provides a well documented set of functionality, including portlet management, user management, layout management, and role-based access control. Its portlet-based architecture offers flexibility and extensibility for portal development and facilitates software component sharing and code reuse. GridSphere is compliant with the *JSR-168* portlet specification [8] which allows portlets to be developed independently of a specific portal framework. GridSphere's portlet service model provides developers with a way to encapsulate reusable business logic into the services that may be shared between many portlets.

The advantages of using GridSphere come not only from its core functionalities, but also from its associated grid portal toolkit, *GridPortlets*. GridPortlets abstracts the details of underlying grid technologies and offers a consistent and uniform high-level service API, enabling developers to easily create customized grid portal web-applications. The GridPortlets services provide functionalities to manage proxy credentials, resources, jobs, and remote files, and supports persistent information about credentials, resources, and jobs submitted by users. GridPortlets delivers five well-designed and easy-to-use portlets, i.e., resource registry, resource browser, credential management, job submission, and file management.

5.2 Workflow Creation

Due to the complexity of the workflow involved, efforts have been made to design the creation process by breaking it into four major steps:

1. General Info – Specify general information on execution, including simulation name and computational resources.
2. Realization – Specify services or algorithms, problem scales and initial parameters to generate input files.
3. Factors – Specify parameters for uncertainty factors, which basically instruct the task farming to generate simulation runs.
4. Wells – Specify parameters for wells, which control the scale of each simulation run.

Since a dedicated archive system is running for simulation results, users do not need to specify archive information in the workflow. The portlet interface is used for workflow creation, where the wizard design eases the process. Detailed validation is performed on the user input at each step. When a user creates a workflow, the data is saved into a database and an XML file is generated, which acts as the workflow description. Then the portlet will submit the workflow via Globus GRAM [6] to the remote workflow engine to execute.

The workflow execution starts with model generation. Various models are created, based on seismic inversion and parameter specification from the portlet interface. The next step of the workflow is to conduct flow numerical simulation for each generated model, which are assigned to remote computational resources. Data archive component is invoked to collect the simulation outputs across the grid, followed by post processing.

5.3 Workflow Tracking

After submitting the workflow, the user can track the status of the workflow execution. Currently, the information about the execution status is gathered by Globus GRAM and it is still relatively limited. As ongoing work is completed, more complex workflow tracking mechanism will be implemented to gather detailed status information for each simulation runs.

Once the workflow execution is finished, simulation results are pushed into the archive system. An archive interface has been developed for users to effectively retrieve simulation results from the archive. This interface supports the searches based on two kinds of metadata. The first set of metadata is the archive system dataset ID, which is associated with the time when the simulation finished, the name of the user who submitted the simulation, the host name of the server where the job was executed, and so on. There is a unique identifier for the dataset in the archive system. The second set of meta data is generated from simulation input parameter files and is closely coupled to the application.

For example, a petroleum engineer wants to query the archive system about a run conducted within the last 30 days. The user specifies a date range in the portal and submits the query to the portal server. The portal queries the archive system, and returns all the data set entries corresponding to that time frame. If this user is interested only in which algorithm the results are generated by, he/she can restrict the searching criteria to a particular simulation algorithm. The end result would then show all the simulation results from the specified algorithm, which occurred within the last 30 days.

5.4 Workflow Reuse

One important feature in our implementation is the ability for reusing workflows. Creating workflows for the ResGrid application requires setting many numeric parameter inputs for different services and the process can be tedious. One observation is that many workflow instances share very similar parameters, and our implementation allows users to create workflow templates. When the user creates a workflow and selects a template, the template values are set as default inputs for the workflow, enhancing usability and making workflow creation quicker. Templates can also be generated from existing submitted workflows. Provisions have been made to edit and delete existing templates.

6. CONCLUSIONS AND FUTURE WORK

We have presented and implemented a workflow solution for designed reservoir study. Contemorary reservoir simulation tools include stochastic simulation, flow modeling, and reservoir simulations, which together form a workflow. Our workflow solution integrates data management and computation support into a unified problem solving environment for large-scale reservoir uncertainty analysis. The GridSphere based portal provides petroleum researchers with a central and intuitive Web-based interface to create, submit and manage workflows of reservoir simulations and track associated data files.

The current workflow implementation is the first step to providing a user environment for grid enabled reservoir uncertainty analysis. Future work is currently focused on adding visualization, monitoring, notification, and collaborative technologies to the workflow solution.

Firstly, the portal will integrate a ResGrid visualization component, which is under development separately. Visualization is used to present the results of potentially huge numbers of simulations and assist further analysis. With the help of this visualization component, a user can obtain easy-to-understand images via a Web browser. Secondly, efforts are underway to provide sophisticated workflow execution monitoring and steering capabilities at runtime during the execution of a given simulation run, to provide the possibility to check workflow status and to terminate the job if an error occurs. Finally, we are also expecting the workflow solution to provide notification service, which means that a user can receive the updated information via email or any instant messenger (e.g., AIM) when a reservoir uncertainty analysis is running across the grid. Work is ongoing to incorporate *public as well as private* templates, expanding the current *private only* workflow template feature, to enable users to share their templates amongst each other.

7. ACKNOWLEDGMENTS

We offer special thanks to Dr. Ian Taylor, Center for Computation and Technology (CCT), Dr. John Smith and Mr. Richard Duff, Department of Petroleum Engineering, Louisiana State University (LSU), for their thoughtful review and comments. This work is a part of the UCoMS project which is sponsored by the U.S. Department of Energy (DOE) under Award Number DE-FG02-04ER46136 and the Board of Regents, State of Louisiana, under Contract No. DOE/LEQSF(2004-07). Additional support was provided by the CCT.

8. REFERENCES

[1] G. Allen, K. Davis, and et al. The gridlab grid application toolkit: Towards generic and easy application programming interfaces for the grid. *Proceedings of the IEEE*, 93:534–550, March 2005. No.3.

[2] I. Altintas, C. Berkley, and et al. Kepler: Towards a grid-enabled system for scientific workflows. In *Workflow in Grid Systems Workshop at the Global Grid Forum (GGF10)*, Berlin, Germany, March 2004.

[3] M. Dahan, M. Thomas, and et al. Grid portal toolkit 3.0 (gridport). In *13th IEEE International Symposium on High Performance Distributed Computing*, Honolulu, Hawaii, June 2004.

[4] E. Deelman, J. Blythe, and et al. Pegasus : Mapping scientific workflows onto the grid. In *2nd European Across Grids Conference*, Nicosia, Cyprus, January 2004.

[5] I. Foster, J. Voeckler, and et al. Chimera: A virtual data system for representing, querying, and automating data derivation. In *14th International Conference on Scientific and Statistical Database Management (SSDBM02)*, Edinburgh, Scotland, July 2002.

[6] Globus Project Home Page.

[7] GridLab: A Grid Application Toolkit and Testbed Project.

[8] Java Community Process. JSR 168: Portlet Specification v1.0.

[9] Open Grid Computing Environments Collaboratory.

[10] UCoMS Project.

[11] G. Laszewski, I. Foster, and et al. A java commodity grid kit. *Concurrency and Computation: Practice and Experience*, 13:643–662, 2001. No. 8-9.

[12] Z. Lei, D. Huang, and et al. Leveraging grid technologies for reservoir uncertainty analysis. In *High Performance Computing Symposium (HPC06)*, Huntsville, Alabama, April 2006.

[13] Z. Lei, D. Huang, and et al. Resgrid: A grid-aware toolkit for reservoir uncertainty analysis. In *IEEE International Symposium on Cluster Computing and the Grid (CCGrid06)*, Singapore, May 2006.

[14] J. Novotny, M. Russell, and O. Wehrens. Gridsphere: A portal framework for building collaborations. In *1st International Workshop on Middleware for Grid Computing*, Rio de Janeiro, September 2003.

[15] T. Oinn, M. Addis, and et al. Taverna: A tool for the composition and enactment of bioinformatics workflows. *Bioinformatics*, 20(17):3045–3054, November 2004.

[16] I. Taylor, M. Shields, and et al. Triana applications within grid computing and peer to peer environments. *Journal of Grid Computing*, 1(2):199–217, 2003.

[17] I. Taylor, M. Shields, and et al. Visual grid workflow in triana. *Journal of Grid Computing*, 3(3-4):153–169, September 2005.

[18] C. White and S. Royer. Experimental design as a framework for reservoir studies. In *2003 SPE Reservoir Simulation Symposium*, Houston, Texas, Feburary 2003.

[19] C. White, B. Willis, and et al. Identifying and estimating significant geologic parameters with experimental design. *SPE Journal (SPE 74140)*, pages 311–324, 2001.

[20] B. Willis and C. White. Quantitative outcrop data for flow simulation. *Journal of Sedimentary Research*, 70, No. 4:788–802, July 2000.

[21] C. Zhang, I. Kelley, and G. Allen. Grid portal solutions: A comparison of gridportlets and ogce. In *Special Issue GCE05 of Concurrency and Computation: Practice and Experience*, 2006.