Working with Workflows:
Highlights from 5 Years Building Scientific Workflows

T. Critchlow, I. Altintas, G. Chin, D. Crawl, H. Iyer, A. Khan S. Klasky, S. Koehler, B. Ludaescher, P. Mouallem, M. Nagappan, N. Podhorszki, A. Shoshani, C. Silva, R. Tchoua, M. Vouk

Introduction

In 2006, the SciDAC Scientific Data Management (SDM) Center proposed to continue its work deploying leading-edge data management and analysis capabilities to scientific applications. One of three thrust areas within the proposed center was focused on Scientific Process Automation using workflow technology. As a founding member of the Kepler consortium [LAB+09], the SDM Center team was well positioned to begin deploying workflows immediately. We were also keenly aware of some of the deficiencies in Kepler when applied to high performance computing workflows, which allowed us to focus our research and development efforts on critical new capabilities that were ultimately integrated into the Kepler open source distribution, benefiting the entire community.

Significant work was required to ensure that Kepler was capable of supporting large-scale production runs for SciDAC applications. Our work on *generic actors and templates* have improved the portability of workflows across machines and provided a higher level of abstraction for workflow developers. *Fault tolerance and provenance tracking* were obvious areas for improvement within Kepler, given the longevity and complexity of our target workflows. To monitor workflow execution, we developed and deployed a web-based *dashboard*, initially targeted to a few specific applications. We then generalized this interface and released it so it could be deployed at other locations. Outreach has always been a primary focus of our work, and we had many successful deployments across a number of scientific domains while continually publishing and presenting our work. This short paper describes our most significant accomplishments over the past five years. Additional information about the SDM Center can be found in the companion paper: "The Scientific Data Management Center: Available Technologies and Highlights."

Workflow Abstractions

In Kepler, workflows are composed of a linked set of components, called actors. Generic actors [CSC+11] provide specific computational or workflow capabilities by abstracting or encapsulating several functions or protocols into a more general version. A goal of generic actors is to hide as much of the implementation details as possible from the user and to provide an interface through the actor's ports and parameters that are natural and relevant solely to the function or task at hand. The first generic actor developed and deployed was the *GenericFileTransfer* actor, which supports the copying of files from one location to another using a variety of transfer protocols regardless of whether the source or destination locations are remote or local. We have also developed the *genericSSH* actor, which can establish connection with SSH servers using a password, phassphrase, passcode, or an SSH certificate, and the *genericJobLaunch* actor, which encapsulates all the tasks involved in job launching into a single actor.

Development of new workflows can be time consuming and can require considerable workflow design expertise. To reduce the overhead and complexity of that task we explored a template-based approach to generation of Kepler workflows. A prototype of the proposed framework was developed and assessed in the context of monitoring workflows. We have also prototyped and explored both actor- and parameter-binding workflow templates that dynamically bind concrete actors and parameters to abstract ones at execution time. Indications are that templates can reduce design complexity as much as 50%.

Fault Tolerance

We developed and deployed multiple workflow reliability and fault tolerance approaches, ranging from managing single actor failures to recovery solutions for the whole workflow.

Caching and Smart Checkpointing: Caching stores actor input data of a completed stateless actor invocation with the associated output data. Upon re-execution with identical input data, the actual computation can be skipped and the corresponding output data retrieved from the cache and reused. Smart checkpointing extends the caching approach by utilizing Kepler recorded provenance to restore a workflow execution after a failure [MCA+10, KMR+11], allowing for fast recovery of workflows.

Contingency actor: When re-execution cannot avoid a failure, other mechanisms are required. This situation led to development of a special contingency actor [CA08]. This actor detects faults in the enclosed subworkflow and allows the user to choose different ways to handle these failures: the subworkflows can be re-executed a given number of times with input data retrieved from provenance, a user-defined alternative for the subworkflow can be executed instead, or a graceful failure can be initiated.

Framework: To manage external middleware layer failures observed in the field, we integrated the above techniques into a framework. This integration required addition of an external lightweight Error-State Handling (ESH) layer that monitors workflow components, middleware, and the overall health of the workflow execution environment. The ESH layer communicates with the contingency actor(s) and other recovery mechanisms in the workflow to choose appropriate fault tolerance mechanisms.

Provenance Tracking

Kepler provides a provenance framework [CA08] that keeps a record of chain of custody for data and process products within workflow design and execution. Provenance recording is an important feature of scientific workflow systems as it facilitates tracking the origin of scientific end products, supporting validation and reproducibility of the processes used to derive these scientific products. The Kepler Provenance Recorder (KPR) collects information about workflow structure and actor executions enabling tracking of the resulting data. KPR has plug-in interfaces for new data models, metadata formats, and storage destinations all designed to serve the multidisciplinary requirements of a broad user community.

Three types of provenance information are collected: the workflow structure (actors and parameters), workflow evolution (how parameters change over time), and workflow execution (the data products read/written during execution). The provenance information collected by the recorder can be stored to multiple data models including an SQL schema.

Additionally, a Query API implemented to retrieve provenance information from this schema is used by the Kepler Reporting System [LAB+09b] and the dashboard.

Dashboard

The eSiMon dashboard was created to help scientists monitor, manage, and collaborate efficiently with teams of researchers working on large high-performance computing (HPC) machines [BKM+09]. eSiMon was designed to be used efficiently across all browsers and platforms and uses Adobe Flash for the front-end of the system. The five main forms of data that are presented to the user are (1) a list of the variables created during the run, (2) extra metadata, such as input files, used during the run, (3) movies and images of the variables, which are constantly updated during and after the run, (4) postprocessing data, along with the provenance information for this data, and (5) vector data. Another key feature of the dashboard is the use of the Kepler provenance tracking system to allow scientists to analyze data, without their knowledge of where the data is on the file, or tape, storage system. The eSiMon dashboard can be a stand-alone tool or can be integrated into a more complex system. It was released as open source so that users can maintain their personal dashboards, and it is currently available for download.
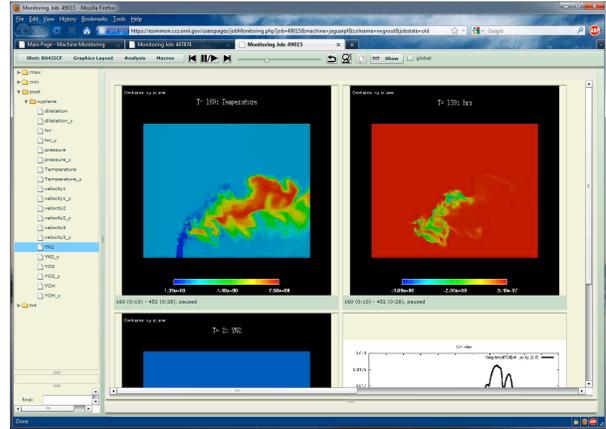
Deployment to Applications

A critical component of our activities is transfer of workflow technology to science teams. We have deployed workflows to a number of science teams to help introduce Kepler into their domains and obtain insights that we can incorporate back into Kepler.

Center for Plasma Edge Simulation (CPES): This fusion project has been a main user of our technologies, providing both requirements and feedback. As the first project using Kepler for monitoring and postprocessing of HPC applications, our actors for SSH and job management found immediate use. eSiMon has been used as the front-end to fusion scientists, and the content was created by Kepler workflows. Provenance tracking was first used in the fusion monitoring workflow and by fusion users of eSiMon. The above tools make up EFFIS, the End-to-end Framework for Fusion Integrated Simulation [CKP+10]. The integrated simulation is driven by a Kepler "coupling" workflow. The plasma state in the simulation by the XGC0 code on a supercomputer is constantly monitored, which involves a data conversion step using another fusion code (M3D) and a parameter-study using yet another code (Elite) on an analysis cluster. If the plasma state is found to be unstable, the XGC0 simulation is stopped and a combined XGC0 and M3D magneto-hydrodynamic simulation is started to go through the turbulent phase of the fusion reaction. After that, XGC0 is continued again alone. Additionally, the workflow creates plots and 2D visualizations from each code's output at every timestep for eSiMon, which can be used for on-line monitoring of the coupled simulation as well as for postprocessing runs.
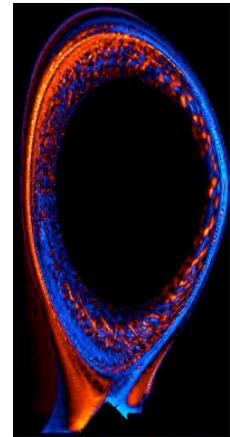
ITER: The SDM Center collaborated with the ITER European Integrated Tokomak Modeling project team at the Institute of Fusion Research, France. In 2006, this group has selected the Kepler workflow system for their workflow development [IJH+10]. In 2007, the SDM Center team hosted a group of visitors from the ITER organization for a month and organized a two-day meeting at SDSC with participants from ITER, SPA, and CPES during their stay. The initial visit was followed by yearly, month-long visits by members of the ITER ITM group in 2008, 2009, and 2010 where we shared development of various components of interest to the ITER teams, including the provenance and fault-tolerance frameworks, simulation code coupling, distributed execution, and actor and module development in Kepler.

Combustion S3D monitoring workflow: Given an input deck, these workflows submit a simulation run on the ORNL or NERSC supercomputer machines, transfer data to an analysis cluster, and perform several data analysis steps on the results. A second workflow has been developed that prepares large runs after code modifications; that is, a simulation code is checked out, compiled and tested on a smaller scale, in order to ensure that large runs do not fail as a result of preventable problems with the new code. This interaction resulted in important feedback from the scientists, which includes part of our work on provenance and fault-tolerance.

ScalaBLAST: This demonstration workflow, called from a web interface, performs a comparison of N submitted genomes against M genomes stored in a library. This requires generating the series of NxM individual comparisons, submitting the results on a cluster, and aggregating the results.

Groundwater Modeling and Analysis: A set of Kepler-based scientific workflows have been constructed to support subsurface flow and transport modeling using the STOMP (Subsurface Transport Over Multiple Phases) simulation. The high-level groundwater modeling workflow involved specific computational tasks including clustering, multivariate interpolation, subsurface flow and transport simulation, and data visualization. Additional low-level workflows were developed to support data staging and simulation job submission. Furthermore, an iterative workflow was designed to collect several input variable ranges and perform a parameter study using fixed code and combinations of input variables values.

The Atmospheric Radiation Measurement (ARM) Program: The ARM program deploys multiple "Value Added Products" (VAPs) that derive scientifically meaningful information from the original raw data sets using a complex combination of data transformations and scientific models. Unfortunately, most of these VAPs are defined through scripts, with no provenance tracking and limited fault tolerance. We developed a demonstration workflow for one of their most complex VAPs, which not only provided an improved execution platform but enabled tracking of data provenance for the first time [CGS+09]. After completion of this demonstration infrastructure, responsibility for the maintenance of this workflow and additional workflow development was transferred to the ARM development team. Over the past year, they have continued to explore how provenance information could be effectively utilized to meet their programmatic requirements [SHC+10].

Conclusions

Over the past five years, our activities have both established Kepler as a viable scientific workflow environment and demonstrated its value across multiple science applications. We

have published over 70 peer-reviewed papers on the technologies highlighted in this short paper and have given Kepler tutorials at SC06, SC07, SC08, and SciDAC 2007. Our outreach activities have allowed scientists to learn best practices and better utilize Kepler to address their individual workflow problems.

While the SciDAC 2 program ends this year, Kepler has a vibrant open source community behind it and is poised to remain one of the few successful scientific workflow environments. Looking towards the SciDAC 3 program, new and exciting developments in workflow technology are anticipated as the program looks towards exascale [CC10].

References

[BKM+09] Roselyne Barreto, Scott Klasky, Pierre Mouallem, Norbert Podhorszki, Mladen Vouk, Collaboration Portal for Petascale Studies, International Symposium on Collaborative Technologies and Systems, 2009. CTS '09. pp. 384-393, May 18-22, 2009,.

[CA08] Daniel Crawl, Ilkay Altintas, A Provenance-Based Fault Tolerance Mechanism for Scientific Workflows, in Proceedings of International Provenance and Annotation Workshop (IPAW 2008), Salt Lake City, UT, pp. 152-159, 2008.

[CGS+09] J. Chase, I. Gorton, C. Sivaramakrishnan, J. Almquist, A. Wynne, G. Chin, T. Critchlow, Kepler + MeDICi –Service-Oriented Scientific Workflow Applications, in Proceedings of International Conference on Web Services (ICWS 2009), Los Angeles, CA, July 2009.

[CC10] T. Critchlow, G. Chin Jr., Supercomputing and Scientific Workflows Gaps and Requirements, A short paper in The Fifth IEEE International Workshop on Scientific Workflows (SWF 2011) published in the Proceedings of the Seventh IEEE World Conference on Services (Services 2011). Washington, DC, July 2011.

[CKP+10] J. Cummings, S. Klasky, N. Podhorszki, R. Barreto, J. Lofstead, K. Schwan, C. Docan, M. Parashar, A. Sim, A. Shoshani, EFFIS: and End-to-end Framework for Fusion Integrated Simulation, The 18th Euromicro International Conference on Parallel, Distributed and Network-Based Computing (PDP) 2010, pp. 428-434, 2010, doi=10.1109/PDP.2010.97.

[CSC+11] G. Chin Jr., C. Sivaramakrishnan, T. Critchlow, K. Schuchardt, A.H.H. Ngu, Scientist-Centered Workflow Abstractions via Generic Actors, Workflow Templates, and Context-Awareness for Groundwater Modeling and Analysis, The Fifth IEEE International Workshop on Scientific Workflows (SWF 2011) published within the Proceedings of the Seventh IEEE World Conference on Services (Services 2011). Washington, DC, July 2011.

[IJH+10] F. Imbeaux, J.B. Lister, G.T.A. Huysmans, W. Zwingmann, M. Airaj, L. Appel, V. Basiuk, D. Coster, L.-G. Eriksson, B. Guillerminet, D. Kalupin, C. Konz, G. Manduchi, M. Ottaviani, G. Pereverzev, Y. Peysson, O. Sauter, J. Signoret, P. Strand and ITM-TF work programme contributors, A Generic Data Structure for Integrated Modelling of Tokamak Physics and Subsystems, Computer Physics Communications, 181, no. 6, June 2010, pp. 987-998, ISSN 0010-4655, DOI: 10.1016/j.cpc.2010.02.001.

[KMR+11] Sven Koehler, Timothy McPhillips, Sean Riddle, Daniel Zinn, Bertram Ludaescher. Improving Workflow Fault Tolerance through Provenance-based Recovery, SSDBM 2011.

[LAB+09] B. Ludaescher, I. Altintas, S. Bowers, J. Cummings, T. Critchlow, E. Deelman, D. D. Roure, J. Freire, C. Goble, M. Jones, S. Klasky, T. McPhillips, N. Podhorszki, C. Silva, I. Taylor, M. Vouk. In A. Shoshani and D. Rotem, editors, Scientific Process Automation and Workflow Management, Scientific Data Management: Challenges, Existing Technology, and Deployment, Computational Science Series, chapter 13. Chapman & Hall/CRC, 2009.

[LAB+09b] Ben Leinfelder, Ilkay Altintas, Derik Barseghian, Daniel Crawl, Matthew B. Jones, Aaron Schultz, Debi Staggs, An Integrated Approach to Managing Workflow Runs and Generating Reports in Kepler, in Eighth Biennial Ptolemy Miniconference. April 2009.

[MCA+10] P. Mouallem, D. Crawl, I. Altintas, M. Vouk, U. Yildiz, A Fault-Tolerance Architecture for Kepler-based Distributed Scientific Workflows, SSDBM 2010, LNCS 6187, pp. 452-460, 2010.

[SHC+10] E Stephan, T Halter, T Critchlow, P Pinheiro Da Silva, L Salayandia, Using Domain Requirements to Achieve Science-Oriented Provenance, a short paper in the 3rd International Provenance and Annotation Workshop (IPAW'2010), June 2010.