Integrated End-to-end Performance Prediction and Diagnosis for Extreme Scientific Workflows

Final Technical Report

for DOE Award No. DE-SC0012630

Submitted by: Ilkay Altintas, PI
University of California, San Diego

Executive Summary

This report details the accomplishments from the ASCR funded project "Integrated End-to-end Performance Prediction and Diagnosis for Extreme Scientific Workflows" under the award numbers FWP-66406 and DE-SC0012630, with a focus on the UC San Diego (Award No. DE-SC0012630) part of the accomplishments. We refer to the project as IPPD.

The main activities of IPPD were centered on the development and integration of provenance information to capture empirically workflow information and identify the sources of bottlenecks as well as variability, and modeling and simulation to provide insights and predictively explore multiple scenarios for development of advanced techniques for optimization of resources and workflow execution. After a Phase I of the project, a Phase II effort focused on three major aspects: a) observe how data is generated, distributed, and used; b) analyze how data is (repeatedly) consumed with a focus both on repeated patterns and anomalies; and c) explore how to optimize data motion. The project leveraged and extended our existing tools with new research and demonstrated our work on the Belle II workflow suite as well as on workflows from NSLS-II.

The highlights outlined in this report as a part of the collaborative goals above are as follows:

Provenance for Workflows. Provenance is used to provide information enabling quality control, re-run computational workflows, and reproduce results. IPPD has built a scalable provenance management system that enables the capture of provenance from the high-level workflow through all relevant system levels in one integrated environment. Leveraging this work, our efforts have included using provenance as an enabling technique.

Modeling Performance and Faults with Deep Learning. Deep learning has become a popular method for characterizing complex systems. Resource contention can be difficult to characterize analytically. UC San Diego team explored the potential of deep learning methods for predicting performance and workflow faults using a provenance dataset that represents two years of production Belle II data.

Dynamic Bottleneck Alleviation. Exploiting our provenance, analysis, and modeling efforts, IPPD explored and developed several techniques for dynamically detecting and alleviating bottlenecks in data movement. UC San Diego team developed PPoDS, a new methodology and set of tools for team science and intelligent end-to-end workflow development. It includes the SmartFlows Toolkit for real-time data collection.

Accelerating Workflows. Finally, we have participated in considerable effort demonstrating our techniques on production-like workflow configurations.

Specific UC San Diego Accomplishments

In Phase I of IPPD, the UCSD team primarily worked on workflow performance prediction using the Kepler workflow system and its provenance framework. Another focus was the work on extending its provenance data model to include performance related variables, and performed a gap analysis between existing schemas and requirements for predictive analytics to achieve higher scheduling optimization. Significant UCSD contributions have included a framework for performance prediction of arbitrarily nested workflows that run on distributed platforms. Our technique views a workflow as a collection of sub-modules running on specific resources and performs localized learning for each resource site. It utilizes instruction set characterization, machine configuration and system workload information to predict overall workflow performance metrics. Specifically, the following was achieved:

1. Identify essential characteristics of hardware resources, program instructions and system load to make accurate predictions of performance metrics of workflow execution instances.
2. Develop a modular ML-based model that trains resource-specific agents to learn the behavior of modular building blocks of a large-scale workflow.
3. Demonstrate that this modular technique scales to large workflows involving arbitrary levels of nested tasks and complex dataflow patterns.
4. Empirically show that resource-node level predictors deliver a scalable solution for wide range of workflows from a relatively small training sample.

In Phase II, we built on this work to develop a modular framework that leverages Machine Learning for creating precise performance predictions of a workflow. We built an approach to leveraging Deep Learning algorithms to discover solutions to unique problems that arise in a system with computational infrastructure that is spread over a wide area. We presented an approach to make the execution of Scientific Workflows more reliable, robust and efficient by predicting if they are likely to fail.

1. MOTIVATION

It is increasingly difficult to design, analyze, and optimize large-scale workflows for scientific computing, especially in situations where time-critical decisions should be taken. Workflows are often designed to execute on a loosely connected set of distributed and heterogeneous computational resources. Each computational resource may have vastly different capabilities, ranging from sensors to high performance clusters. Frequently, workflows are composite applications built from loosely connected parts, and often workflow tasks communicate via files sent over general-purpose networks. As a result of this complex software and execution space, large-scale scientific workflows exhibit extreme performance variability, often due to blocking for data and contending for data-movement resources. Therefore, it is critical to understand clearly the factors that influence performance, data-movement bottlenecks and to develop techniques to minimize them, enabling more science via improved response times and throughput.

IPPD worked towards an integrated approach to the prediction and diagnosis of extreme scale workflows for scientific computing that enables exploration, prediction and optimization of performance of a workflow and its components. Underlying this is a multi-scale view that enables fine-grained component analysis using simulation-based tools that allow in-depth analysis and, through suitable abstractions, end-to-end workflow performance analysis and prediction using analytical prediction. Provenance information collected through instrumentation and monitoring of actual workflow execution, provides empirical bounds on the expected performance as well as identifying priority areas and scenarios for further performance modeling and simulation.

2. PROVENANCE for WORKFLOWS

In IPPD, workflow provenance was used to carry out empirical studies of workflow performance variations and their causes. Results of this work, produced by the collaborative partnership of PNNL, BNL and UC San Diego resulted in techniques for collection and storage of provenance information in a form that enables insights into workflow behavior that directed IPPD's modeling and simulation efforts towards high value targets.

UC San Diego team contributed to IPPD's efforts on provenance metrics hybridization. Realizing that in order to have a more complete understanding of a description of a workflow application, it becomes necessary to combine or merge disclosed provenance with the observed system metrics, we designed and built a data model as a critical link between the workflow application and its execution environment. This model supported ACME climate simulations study reconstruction and workflow performance analysis, and high-energy physics Belle2 scheduling strategies, as well as reproducibility of workflows in biology and thermal modeling.

A key research question emerged, '*How can provenance assertions be made by relating historical evidence to interrelated time-series metric events?*'. Provenance metrics are used to provide this link by defining translations of provenance data into time-series measurement representation for storage in a Metrics Store, and thus, enabling the alignment of provenance and the system metrics data. Conducting performance analysis of this data, we were able to successfully use metrics on PNNL's Seapearl cluster (instrumented for Power and Thermal activities) to predict or detect a compute node based on its power/thermal signature.

3. ANALYSIS: MODELING PERFORMANCE AND FAULTS WITH DEEP LEARNING

3.1. Performance Prediction Using Machine Learning

We have developed a modular framework that leverages Machine Learning for creating precise performance predictions of a workflow. A diagram is shown in Figure 1. The central idea is to partition a workflow in such a way that makes the task of forecasting each atomic unit manageable and gives us a way to combine the individual predictions efficiently. We recognize a combination of an executable and a specific physical resource as a single module. This gives us a handle to characterize workload and machine power as a single unit of prediction. The modular approach of the presented framework allows it to adapt to highly complex nested workflows and scale to new scenarios. We present performance estimation results of independent workflow modules executed on the XSEDE SDSC Comet cluster using various Machine Learning algorithms. The results provide insights into the behavior and effectiveness of different algorithms in the context of scientific workflow performance prediction.

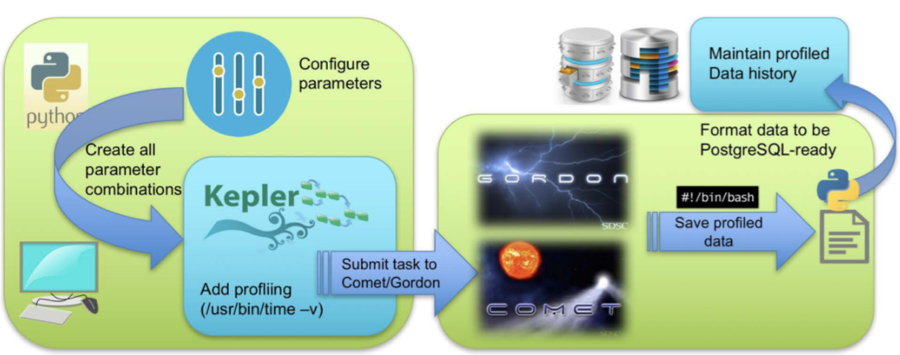

Figure 1. Profiling Framework is designed with scalability at its core. The Kepler workflow system interfaces with the computing resource (such as Comet). A local Python scripts automates the job of generating all possible profiling experiment combinations and invokes Kepler to submit tasks to the distributed infrastructure.

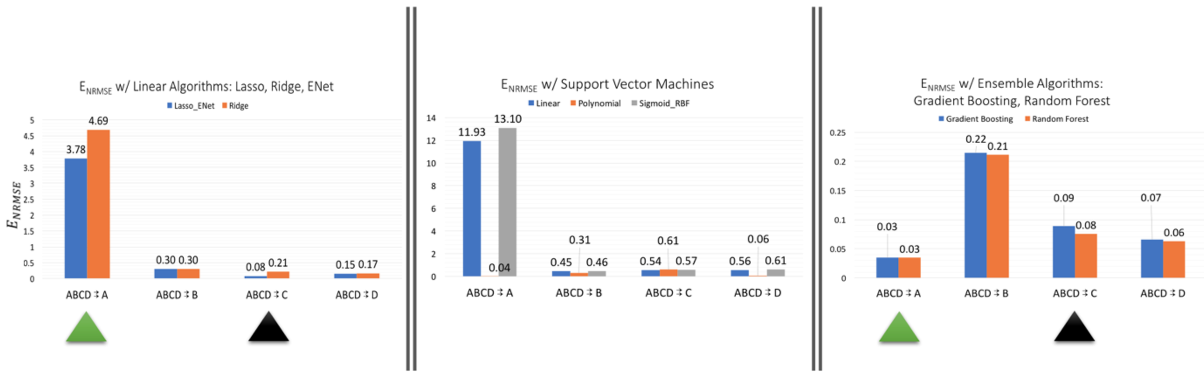

Figure 2. Summary of different performance models. Simple models can give accurate results in unexpected domains – evaluate before abandoning the linear models. Use heavy weight complex models wisely – verify and deploy as necessary.

Figure 2 summarizes our results evaluating several different modeling techniques. When Linear models get samples from all modules, they perform excellent on some modules, but do not give reliable results for all modules. Support Vector Machine based models perform sporadically even when given training samples from all modules. Polynomial Kernel shows fairly good performance. Linear still outperforms in 2 / 4 cases. Support Vector Machines can give precise results for some modules. Ensemble methods give outstanding performance across all modules, when given chance to learn about each module. In this graph, we see drastic improvement on module A (low error), a module on which all models struggled when under data deprivation.

3.2. Mining Facility Logs for Performance and Faults

Belle II, an important workflow in the High Energy Physics (HEP) community, is a good example of a distributed workflow. Raw data is generated by the Belle II detector located at Japan's High Energy Accelerator Research Organization (KEK). Physicists around the world generate additional data through Monte Carlo simulations and user analysis.

Using a provenance dataset that represents two years of production Belle II data, we have applied deep learning to characterize performance and faults.

Deep Learning on Operational Facility Data Related to Large-Scale Distributed Area Scientific Workflows. We have outlined an approach to leveraging Deep Learning algorithms to discover solutions to unique problems that arise in a system with computational infrastructure that is spread over a wide area [30]. The presented vision, motivated by a real scientific use case from Belle II experiments, is to develop multi-layer neural networks to tackle forecasting, anomaly detection and optimization challenges in a complex and distributed data movement environment. Through this vision based on Deep Learning principles, we achieved reduced congestion events, faster file transfer rates, and enhanced site reliability.

Deep Learning for Enhancing Fault Tolerant Capabilities of Scientific Workflows. We developed an approach to make the execution of Scientific Workflows more reliable, robust and efficient. We focused on the question of whether it is possible to increase performance by forecasting job failures in distributed workflows. Although some jobs fail 'quickly' (e.g., crash soon after launch), others fail by 'slowly' (e.g., hang or crash long after launch). If we can predict soon after job launch whether a job is likely to fail 'slowly', the management system can kill and re-execute the job, recovering much of the wasted time.

We applied deep learning techniques to develop a mechanism that forecasts the final state (success or failure) of a dynamic job in a large-scale particle physics experiment, with minimal data gathering, and as early as possible in job's life cycle. The key advantage of having a predictive mechanism to identify and anticipate failure-prone jobs is the potential for designing intelligent Fault Tolerance mechanisms to handle anomalous events. We achieved a 14 percent improvement in computational resources utilization, and an overall classification accuracy of 85 percent on real tasks executed in a High Energy Physics Computing workflow. To the best of our knowledge, this is the most exhaustive and first of its kind study of neural network architectures in context of a real-dataset profiled from a large-scale scientific workflow.

4. PPoDS and SmartFlows

Today's computing has diverse workload characteristics spanning high-performance computing, high-throughput computing and big data analytics. The traditional supercomputing applications are stronger than ever on their way to embrace exascale computing capacity. As our ability to collect data in real-time from internet-of-things has improved, the demand to process such data at scale has increased and requires big data processing capabilities. We observe a growing number of applications, including smart cities, precision medicine, energy management and smart manufacturing, that require a combination of advanced data analytics with traditional modeling and simulations. In addition, thanks to the advances in new computer architectures, most scientific codes are ported for special environments, e.g., GPUs. There is also an increasing demand for computing from scientific disciplines like social sciences which were not traditionally seen as supercomputing disciplines. In fact, every domain of science and engineering today can take advantage of big data and computing. A challenge for today's computing architectures is the ability to respond to such heterogeneous needs and lowering the barriers to computing for long tail researchers as well as supporting the most cutting-edge computing applications.

On the software side, we observe many new ways to manage big data and high-performance storage as well as new forms of data integrity technologies, e.g., blockchain. Use of analytical and big data frameworks, e.g., Spark and Keras (keras.io), are common in individual machine learning applications and as a part of integrated data-driven scientific simulations. Such heterogeneous capability in computing and software brings with it the need for software systems that can coordinate applications across different scales of computing, data and networking needs. A number of software innovations like cluster virtualization and container technologies, e.g., Docker (docker.com) and Singularity (singularity.lbl.gov), increased the portability of these software frameworks and environments, making it possible to turn any executable to run as a service on multiple platforms. Kubernetes (kubernetes.io) has emerged as a dynamic container and resource management platform that can automate the configuration and orchestration of computing resources for varying workloads. Gateways , Jupyter notebooks (jupyter.org) and similar enabling web and mobile interface have lowered the barriers for many more to access data on the fly and take advantage of computing.

All these make workflows even more needed at the converged application level to enable communications with data and computing middleware, while optimizing resources and dynamically adapting to the changes during the execution of integrated applications. Workflows provide an ideal programming model for deployment of computational and data science applications on all scales of computing and provide a platform for system integration of data, modeling tools and computing while making the applications reusable and reproducible. They make it possible to manage dynamic-data driven applications and decision support using advances in big data platforms and on-demand computing systems, e.g., dynamic data-driven fire behavior modeling in real-time (e.g., wifire.ucsd.edu).

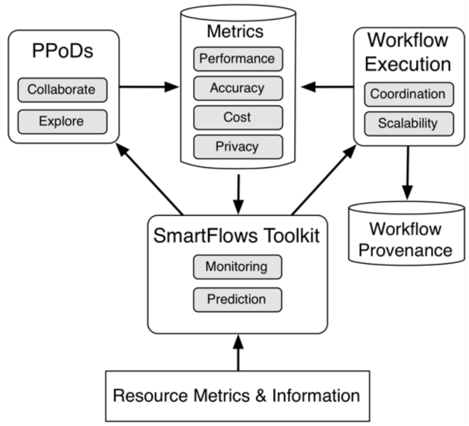

Fig. 3. The high-level PPoDs architecture showing the dependencies between the presented collaborative workflow-driven science tools.

Moreover, there is a new opportunity here for workflows to become even more useful and more aligned with the way teams of scientists collaborate and develop integrated applications. Starting with the question "can there be a methodology to make workflows a systematic part of the collaborative scientific process?" and tackling the problem of "what would a toolkit look like for optimizing workflow effectivity from multiple perspectives within a team?", we built a new methodology and set of tools for team science and intelligent end-to-end workflow development. Figure 3 depicts key relationships. Specifically, we contributions were:

(1) an introduction to the conceptual PPoDS (Process for the Practice of Data Science) methodology for collaborative metric-based workflow design,

(2) a framework design for measuring and testing exploratory workflows using the PPoDS metrics,

(3) a design for capturing data during exploratory workflow development to make intelligent scalability and steering possible,

(4) an introduction to the SmartFlows Toolkit for real-time data collection, benchmarks and intelligence for smart workflow execution, and

(5) a collaboration-centered reference architecture using contributions 1-4 to extend workflow systems with dynamic, predictable and programmable interfaces to teams, systems and scalable infrastructure while bridging the exploratory and scalable activities in the scientific process.

5. Publications Resulting from this Research

- Jianwu Wang, Daniel Crawl, Shweta Purawat, Mai Nguyen and Ilkay Altintas. 2015. "Big data provenance: Challenges, state of the art and opportunities," 2015 IEEE International Conference on Big Data (Big Data), 2015, pp. 2509-2516, doi: 10.1109/BigData.2015.7364047.
- Singh, A., Stephan, E., Elsethagen, T., MacDuff, M., Raju, B., Schram, M., Kleese van Dam, K., J Kerbyson, D., Altintas I. 2016. Leveraging Large Sensor Streams for Robust Cloud Control, In Proceedings of the Big Data for Cloud Operations Management: Problems, Approaches, Tools, and Best Practices Workshop at IEEE International Conference on Big Data (BigData 2016).
- Elsethagen, T., E. Stephan, B. Raju, M. Schram, M. MacDuff, D. Kerbyson, K. K. van Dam, A. Singh, and I. Altintas. 2016. "Data Provenance Hybridization Supporting Extreme-Scale Scientific Workflow Applications." In 2016 New York Scientific Data Summit (NYSDS), 1–10. doi:10.1109/NYSDS.2016.7747819
- Alok Singh, Arvind Rao, Shweta Purawat, and Ilkay Altintas. 2017. A Machine Learning Approach for Modular Workflow Performance Prediction. In Proceedings of the 12th Workshop on Workflows in Support of Large-Scale Science (WORKS '17). ACM,NewYork,NY,USA,7:1–7:11. https://doi.org/10.1145/3150994.3150998

- Alok Singh, Eric Stephan, Malachi Schram, and Ilkay Altintas. 2017. Deep Learning on Operational Facility Data Related to Large-Scale Distributed Area Scientific Workflows. In 2017 IEEE 13th International Conference on e-Science (e-Science). IEEE,586–591. https://doi.org/10.1109/eScience.2017.94
- Alok Singh, Ilkay Altintas, Malachi Schram, and Nathan Tallent. 2018. Deep Learning for Enhancing Fault Tolerant Capabilities of Scientific Workflows. In Second IEEE Intl. Workshop on Benchmarking, Performance Tuning and Optimization for Big Data Applications (Proc. Of the IEEE Intl. Conf. on Big Data). 3905–3914. https://doi.org/10.1109/BigData.2018.8622509
- Ilkay Altintas, Shweta Purawat, Daniel Crawl, Alok Singh, and Kyle Marcus. 2019. Towards A Methodology and Framework for Workflow-Driven Team Science. IEEE Computing in Science and Engineering (2019), 38–49. https: //doi.org/10.1109/MCSE.2019.2919688
- Shweta Purawat, Cathie Olschanowsky, Laura E. Condon, Reed Maxwell, Ilkay Altintas. 2020. Scalable Workflow-Driven Hydrologic Analysis in HydroFrame. In: Krzhizhanovskaya V. et al. (eds) Computational Science – ICCS 2020. ICCS 2020. Lecture Notes in Computer Science, vol 12137. Springer, Cham. https://doi.org/10.1007/978-3-030-50371-0_20
- Alok Singh, Shweta Purawat, Arvind Rao, Ilkay Altintas. 2021. Modular performance prediction for scientific workflows using Machine Learning, Future Generation Computer Systems, Volume 114, 2021, Pages 1-14, ISSN 0167-739X.