The Panorama project aims to further the understanding of the behavior of scientific workflows as they are executing in heterogeneous environments. Panorama's approach to modeling and diagnosing the runtime performance of complex scientific workflows is to integrate extreme-scale systems testbed experimentation, structured analytical modeling and parallel systems simulation into a comprehensive workflow framework that can characterize the end-to-end workflow performance on today's and future generation architectures, which can be used to improve the overall workflow performance and reliability. The Panorama architecture includes the individual components: the Aspen analytical application modeling software, the ROSS simulation framework, the Pegasus workflow management system, and how they are used to model the behavior of DOE-relevant applications. By having a coupled model of the application and execution environment, decisions can be made about resource provisioning, application task scheduling, data management within the application, etc. Our approach for correlating the real time application and infrastructure monitoring data can be used to verify application behavior, perform anomaly detection and diagnosis, and support adaptivity during workflow execution.

**IMPACT ON DOE SCIENCE**

Diamonds that deliver!

Panorama enabled cutting-edge domain science research and development that has the potential to solve some of the challenges associated with drug discovery and delivery:

- The motions of a tRNA (or transfer RNA) model system can be enhanced when coupled with nanodiamonds, or diamond nanoparticles approximately 5 to 10 nanometers in size.
- We have developed an SNS Pegasus workflow to confirm that nanodiamonds enhance the dynamics of tRNA when in the presence of water. The workflow calculates the time-averaged properties (e.g., torsion angles) of the folded tRNA and the potential to solve some of the challenges for drug discovery and delivery.
- Apache Airflow uses the Apache Beam framework, diagnosing the potential to solve some of the challenges for drug discovery and delivery.

**MODELING AND SIMULATION**

Scalable Workload Generation for Application Performance Modeling and Simulation

- We have created a new technique for generating scalable workloads from real applications, and implemented a prototype, called Durango, using a performance modeling toolkit.
- We demonstrate the efficacy of Durango’s direct integration approach, which links Aspen into CODES as part of the running network simulation model. Here, Aspen generates the application-level computation timing events, which in turn drives the start of a network communication phase.

**BURST BUFFERS**

On the use of Burst Buffers for Accelerating Scientific Workflow Executions

Burst Buffers (BB) have emerged as a non-volatile storage solution that is positioned between the processors’ memory and the PFS, buffering the large volume of data produced by the application at an higher rate than the PFS, while seamlessly draining the data to the PFS asynchronously.

We explored the impact of Burst Buffers (BB) in scientific workflow applications. Using a software stack including Pegasus-WMS and HT-Condor, we ran a workflow on the Cori system at NERSC which included provisioning and releasing remote-shared BB nodes. Our application wrote and read about 550 GB of data.

**Major Findings:**

- I/O write performance was improved by a factor of 9, and I/O read performance by a factor of 15.
- Performance decreased slightly at node counts above 64 (potential I/O ceiling).
- I/O performance must be balanced with parallel efficiency when using burst buffers with highly parallel applications.
- I/O contention may limit the broad applicability of burst buffers for all application types (e.g., in situ processing).

**NETWORK PROVISIONING**

Data Flow Prioritization for Scientific Workflows Using a Virtual SDX

We developed mechanisms to arbitrate and prioritize data flows from competing workflows by leveraging advanced network provisioning technologies like a virtual Software Defined Exchange (SDX).

- Software Defined Exchanges (SDX) – meeting point of networks to exchange traffic, securely and with QoS, using SDN protocols
- Virtual SDX – virtual overlay acting as SDX without persistent physical location
- ExoGENI virtual SDX can modify compute, network, storage to support changing demands of SDX

**Prioritized Data Flows**

- Virtual SDX transparently arbitrates workflow data flows communicated by Pegasus

---

**LEARN MORE**

Panorama Website
http://sites.google.com/site/panoramaworkflows/

Panorama is funded by the US Department of Energy under Grant DOE-SC0012636