Panorama 360: End-to-End Online Performance Data Capture and Analysis of Scientific Workflows

George Papadimitriou[1], Rafael Ferreira da Silva[1], Karan Vahi[1], Rajiv Mayani[1], Mats Rynge[1], Ewa Deelman[1], Anirban Mandal[2], Cong Wang[2], Mariam Kiran[3], Vickie Lynch[4], Jeffrey Vetter[4]

[1]University of Southern California – Information Sciences Institute [2]Renaissance Computing Institute [3]Lawrence Berkeley National Laboratory [4]Oak Ridge National Laboratory

PANORAMA 360

Overview of the Project Description

The Panorama 360 aims to further the understanding of the behavior of scientific workflows as they are executing in heterogeneous environments. Panorama 360 collects and correlates workflow performance data into a comprehensive view, that can characterize the **end-to-end** workflow performance on today's systems and drive the design of the future generation systems.

The Panorama 360 architecture collects data from these individual data sources: the ***Pegasus WMS***, the ***Globus Online*** service, the *TCP STatistics and Analysis Tool (**Tstat**)* and the ***DARSHAN*** HPC I/O Characterization Tool.

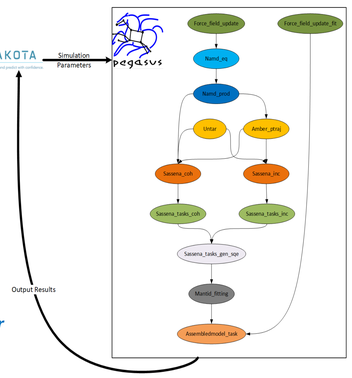
Our approach for correlating the real time application and infrastructure monitoring data can be used to verify application behavior, perform anomaly detection and diagnosis, and support adaptivity during workflow execution, in an **online** manner. This can lead to improved performance and stability of scientific workflows and benefit the DOE-relevant applications.

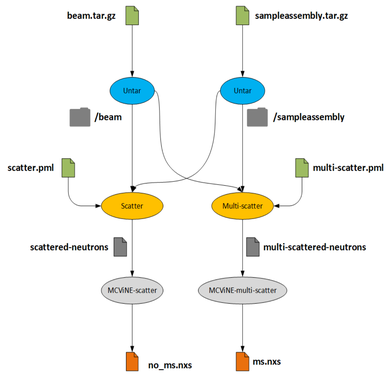
Ultimately, by having all these data and analysis tools in our disposal, we envision the creation of a workflow performance data repository and a collection of tools that will be publicly available and can drive science forward.

IMPACT ON DOE SCIENCE

Nanodiamond and MCViNE Workflows

- We have extended the SNS-Nanodiamond workflow with (a) parameter sweep, and (b) feedback loop capability, using ***Dakota*** (https://dakota.sandia.gov).
- Dakota based on an internal model generates the ***Simulation Parameters*** and triggers a ***Pegasus*** analysis driver that initiates a new workflow.
- Dakota can instantiate multiple simulation workflows in parallel and by using the ***Pegasus Ensemble Manager*** we can have more control over their execution.





- ***MCViNE*** (http://www.mcvine.org) is a Monte Carlo neutron ray-tracing program for computer modeling and simulations that mirror real neutron scattering experiments.
- MCViNE is used to study data from many SNS instruments, and some of them are ARCS and SEQUOIA.
- We collaborate closely with SNS scientists and explore ways to support MCViNE workflows with Pegasus.

LEARN MORE

Website

https://panorama360.github.io

GitHub Repository

https://github.com/Panorama360

Pegasus Office Hours

https://pegasus.isi.edu/blog/online-pegasus-office-hours

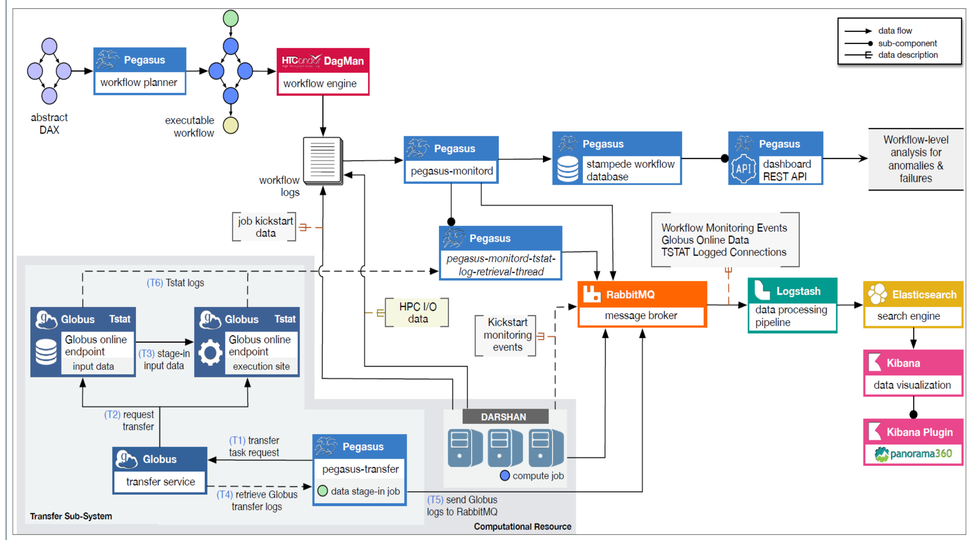
Panorama 360 is funded by the US Department of Energy under Grant #DE-SC0012636M

PERFORMANCE DATA CAPTURE

Overview of the Data Capture Architecture

Panorama 360's data capture architecture can be divided into **5 entities:**

- The workflow management system (Pegasus)
- The data sources (Pegasus, Globus, Tstat, Darshan)
- The search (ElasticSearch) and visualization (Kibana) engines
- The message broker (RabbitMQ)
- The data processing pipeline (Logstash)

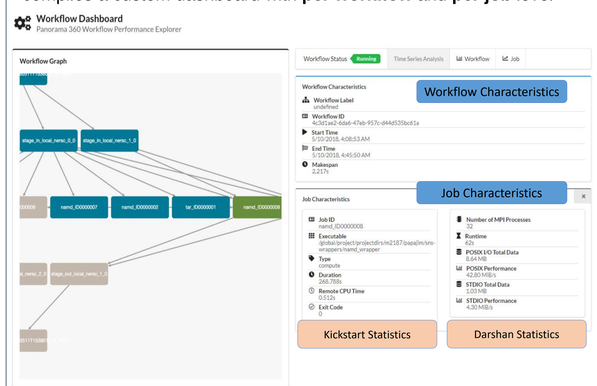


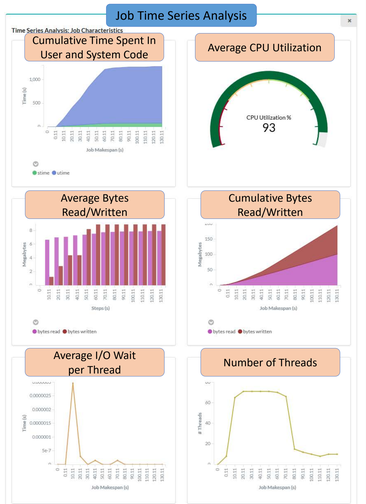
- Pegasus, apart from executing the workflow, is the entity that orchestrates the data movement between the data sources and the analysis framework.
- Pegasus-monitord reports events about the status of the workflow, and extracts Darshan reports from the job's stdout.
- Pegasus-kickstart wraps the execution of jobs and provides aggregated execution statistics (duration, I/O, memory usage), after job completion. Additionally it has an **online** monitoring feature that allows us to collect refined traces with an interval, as low as 1 second.
- Darshan provides POSIX and MPI I/O file access statistics. It tracks the application's file access patterns and the performance of the underlying file system.
- Globus reports back overall transfer statistics and information about each individual transfer (throughput, files transferred, errors etc.)
- Tstat captures low level TCP statistics and apart from revealing network issues, it can help us understand the reason behind an underperforming network connection.

DATA VISUALIZATION

Panorama 360 Kibana Plugin

In order to visualize the collected data in a *meaningful* and *compact* way, we have created a **Kibana Visualization Plugin** that correlates the information provided by pegasus-monitord with the monitoring data and compiles a custom dashboard with **per workflow** and **per job** level



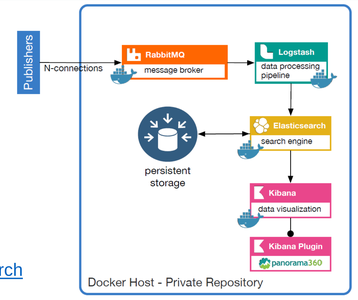


LOCAL DEPLOYMENT

Deploying the Data Collection Back-End Using Containers

- Configuring the monitoring back-end, even though it relies on well established tools, can turn out to be a *cumbersome* and *challenging* process.
- We have **automated** the deployment of the data collection backend using Docker and Docker-Compose.
- We provide pre-configured container recipes that can *collect* and *store* persistently **all** the data the Panorama 360 architecture produces.

https://github.com/Panorama360/data-collection-arch



OPEN ACCESS DATA

Overview of the Open Access Data Endpoints

- We have created an open access repository containing data from over **40 workflow executions** of a data-intensive (1000Genome[1]) and a compute-intensive (SNS[2]) workflow.
- *Globus* facilitated the transfers, while *ExoGENI*[3] and *Cori* at NERSC[4] were used to execute the computational tasks.

ElasticSearch Index	# Records	Description
panorama_transfer	1,150	Logs retrieved from the Globus transfer service
panorama_kickstart	78,600	Resource utilization traces collected by Pegasus-Kickstart
panorama_stampede	34,400	Workflow execution events and Darshan logs

elasticsearch — Can be accessed programmatically at: https://data.panorama.isi.edu

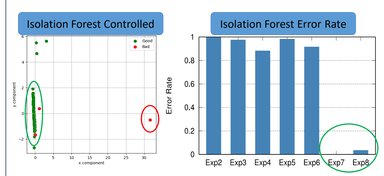
kibana — Can be accessed via a web browser at: https://kibana.panorama.isi.edu

[1] 1000Genome: https://github.com/pegasus-isi/1000genome-workflow
[2] SNS: https://github.com/pegasus-isi/SNS-Workflow
[3] ExoGENI: http://www.exogeni.net/
[4] NERSC: https://www.nersc.gov/

MACHINE LEARNING EXPLORATIONS

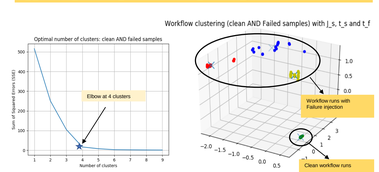
Workflow Level Analysis and Infrastructure's Network Performance Analysis

- **Detecting anomalous TCP** file transfers using *Isolation Forest.* We were able to find transfers *experiencing packet reordering* (Exp.7-8) with very good accuracy.
- But, we got poor performance on *packet loss* (Exp.2-4) and *packet duplication* (Exp.5-6).



- **Workflow-level analysis** using *Unsupervised Clustering* with unlabeled data. Using feature vectors such as:

J_s = #job_instances_succeeded/#job_instances_done
J_f = #job_instances_failed/#job_instances_done
t_s = Sum(local_duration(successful_job_instances))/#job_instances_succeeded
t_f = Sum(local_duration(failed_job_instances))/#job_instances_failed
o_i_s = #job_instances_succeeded/#total_workflow_jobs



COLLABORATIONS

RAMSES — We collaborate with the DOE *RAMSES* project in collecting xfer statistics from Globus transfers and discussing ideas about network research on workflows.

OAK RIDGE National Laboratory / summit — We are working closely with the *Advanced Data and Workflows* group at OLCF, to provide an easy way of accessing **Summit** and **RHEA** with Pegasus using *Kubernetes* and containers.

PanDA — We are working with the DOE *BigPanda* project to allow Pegasus workflows to be executed using BigPanda's pilot job submission mechanisms. http://news.pandawms.org/bigpanda.html

X-SWAP PROJECT — In collaboration with the DOE *X-SWAP* project we are working on a **benchmark workflow** for current and next generation systems, based on the ZTF astrophysics pipeline. https://www.ztf.caltech.edu/