### 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 DARIFMS RPC/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.

### 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 data processing pipeline (Logstash)
- The message broker (RabbitMQ)

### Performance Anomaly Detection

**Workflow Performance**

Complexity of end-to-end workflow enactment brings increased potential for performance problems that need to be detected and mitigated.

**Infrastructure - Network Performance**

In controlled experiments there are clear normal/anomalous clusters.

- None found in 1000 Genome workflow runs
- Found most influential variables (Reducing 150 to 28)
- FITT, RTT, window details, fragment size
- In 1000 Genome workflow, RTT was much higher due to the change in topology, and that was most likely the cause of no clusters.

---

**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-monitor with the monitoring data and compiles a custom dashboard with per workflow and per job level information, that is updated automatically during the workflow execution.