

IRIS-HEP Undergraduate Fellowship - Suhaib Shaikh

Project: Proactive Site Monitoring

Mentor: Dr. Derek Weitzel - UNL

Researchers using the OSG rely on the stability of resources in order to accomplish their science. The OSG's GRACC accounting service collects usage information for all sites contributing to and all jobs that run on the OSG. The accounting service is a large source of information on the OSG. The accounting data is stored within an Elasticsearch database at UNL. Monitoring using this accounting service exists that will alert when a site completely fails, but there is no alerting on a decrease in functionality of a site.

My project would be to develop **proactive site monitoring** to alert on site issues detected from the GRACC accounting system. The alert would run periodically on OSG resources.

The **proactive site monitoring** project will be broken into Learning, development and feedback sections. Feedback will be solicited from the OSG's operations team on the format of the alert.

Weeks 1-2 - Learning:

OSG's GRACC accounting service data is complex with many attributes for each site and job. The attributes used to detect issues must be researched. Two attributes that may show a troubled site is core hours and CPU efficiency. A decrease in core hours may indicate that the site is failing to run jobs. A decrease in CPU efficiency at a site may indicate an I/O issue at the site, or even a data cache failure that caused a failover to a distant cache. A significant increase in either of these attributes could indicate a change at the site that may necessitate action by the OSG operations team.

Outlier detection using machine learning can detect variances of attributes outside of the regular operation. I will need to study outlier detection techniques to understand the benefits of each and to find the technique that will work best with the GRACC dataset.

Weeks 3-6 - Development:

The development of the alert will require querying GRACC's Elasticsearch cluster, writing machine learning framework for outlier detection, and finally the email of the alert to the operations team.

Week 7 - Feedback:

The reports of troubled sites will be emailed to the operations team in the OSG. The sensitivity of the outlier detection will need to be refined to minimize false positives. Further, the email

must give context of the outlier by showing images of the differences between normal and an outlier. Feedback from the OSG operations team will be solicited to refine the alerts.

Weeks 8-9 - Development:

Refinements of the reports from the feedback of the OSG operations team.

Week 10 - Production:

Refinement of the reports in order to put the alerting into production.