21st International Conference on Computing in High Energy and Nuclear Physics **CHEP2015** Okinawa Japan: April 13 - 17, 2015

Contribution ID: **388** Type: **poster presentation**

Development of site-oriented Analytics for Grid computing centres

The field of analytics, the process of analysing data to visualise meaningful patterns and trends, has become increasingly important to a wide range of scientific applications as the volume and variety of accessible data available to process (so called Big Data) has significantly increased. There are a number of scalable analytic platforms and services which have risen in prominence (such as Elasticsearch) which enable unstructured data from numerous sources to be gathered, curated and visualised through a single extensible interface. There is ongoing work in the HEP community evaluating these tools, for example in the augmentation of system management at regional computing centres. In this context the provisioning of analytic solutions for computing sites pledging resources to the Worldwide LHC Computing Grid (WLCG) is an area of considerable interest.

Each Grid computing centre generates a wealth of monitoring data from multiple sources as part of their ongoing operations. These include system logging, Grid middleware services, LRMS scheduling information, network and storage utilisation and workload performance. This rich set of data is available for exploitation using analytics tools to enable post-facto diagnostics and a more comprehensive understanding of site systems, extending existing work on site monitoring. A site-oriented analytics portal would allow administrators to more easily leverage their available logging and monitoring data to determine the causes in variations in workload performance that may be unclear from a single data source.

In this study we will explore the components necessary for a WLCG site-oriented analytics platform. A corpus of relevant time-series based monitoring and logging data collected at two UK Grid computing centres (ECDF and Glasgow) will be categorised and stored. We will then explore the use of this data as part of a distributed analytics system. A necessary part of this work will be an examination of the appropriate level of visibility for different categories of site data. Furthermore, we will explore the extent to which machine learning techniques could be harnessed to provide predictive capability in error detection by using curated site data as a continuous training set.

This model is being developed with a particular focus on providing a solution for Grid computing centres rather than attempting to cover all data sources generated by a Virtual Organisation (VO). Such an approach is intended to complement analytic development from larger VOs (such as the LHC experiments) whilst benefiting smaller VOs who may not have the resources available to develop these types of tools. Based on this work we will look towards areas of best practice in developing systems of this nature.

Primary authors: WASHBROOK, Andrew John (University of Edinburgh (GB)); CROOKS, David (University of Glasgow (GB))

Co-authors: Prof. BRITTON, David (University of Glasgow (GB)); QIN, Gang (University of Glasgow (GB)); ROY, Gareth Douglas (University of Glasgow (GB)); Dr SKIPSEY, Samuel Cadellin

Presenters: WASHBROOK, Andrew John (University of Edinburgh (GB)); CROOKS, David (University of Glasgow (GB))

Track Classification: Track7: Clouds and virtualization