Elasticsearch – An Open Source Log Analysis Tool

Content:
Logging to application and system information is crucial to running any service. Whether it is looking for trends in use, investigating specific problems or analysing security incidents, logs are an essential tool for any service manager or system administrator. However, many applications produce large volumes of logging information from a number of disparate, but related, daemons running on an increasing number of machines. In this scenario, log analysis with traditional command line tools such as awk and grep is hugely inefficient. In addition, presenting the results of analysis using applications like Excel™ is a time consuming process for a busy administrator. At STFC, the Scientific Computing Department has adopted Elasticsearch (www.elasticsearch.org) for collating of logs from the mass storage system. Elasticsearch is built from a number of open source components; LogStash, Kibana and Lucene. It is currently gathering information from about 300 machines at the Rutherford Appleton Laboratory covering both the WLCG Tier 1 and local facilities storage systems. In this presentation we discuss the evolution of Elasticsearch within STFC and present specific use cases where it has clear advantages over traditional methods. We also compare it with the previous logging system developed at CERN (DLF) and show that Elasticsearch is a more generically useful tool. Finally we demonstrate its usefulness for looking at both trends and specific event analysis.

Primary authors: Mr. APPLEYARD, Rob (STFC)
Co-authors: Mr. ADAMS, James (STFC)
Presenter: Mr. APPLEYARD, Rob (STFC)

Session classification: Big Data II
Track classification: Big Data
Type: Oral