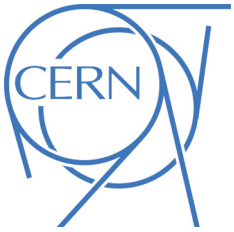
ATLAS NOTE

August 29, 2013



A Data Mining Application for ATLAS Detector Controls Data

R. Balkin[a]

[a]*CERN, Geneva, Switzerland*

Abstract

Modern experiments in high energy physics conducted in the Large Hadron Collider in CERN are incredibly complex. Measurements in these experiments require many different and advanced systems, integrated in complicated machines called detectors. ATLAS is a general purpose detector at the LHC. Such an elaborate machine requires a sophisticated control system to allow users to control and monitor its performance. This system, called the ATLAS Detector Control System (DCS), regularly logs information regarding the various subsystems, such as system information coming from DCS servers and software tools. It also notifies users about any malfunctions in the detectors software or hardware using alarms. Accessing log information and alarms efficiently may help detect problem more easily, increase our understanding of the various systems and even help predict and prevent future problems. DCS Data miner (DDM) is an application created using a tool called Splunk, in order to make data mining of log and alarm information easy and efficient.

1 Motivation

Log information is stored in order to enable DCS developers to follow in detail the status of the various control system of the detectors in case the system behaves abnormally. Alarms, on the other hand, are used to inform detector users about software and hardware malfunction. Both Log and Alarm information is stored in databases. Presently, the ATLAS[1] DCS[2] logs large amount of information regarding everyday functionality of its various components. While these logs are a bit redundant, they hold significant details about how the system works. Exploring this sea of information is not easy. Looking past redundancies and identifying patterns is not a trivial task. Even after this computationally difficult task is achieved, displaying this information in a user-friendly and informative way presents its own difficulties. Unlike log information, alarms are used to inform the users of various software and hardware problems. When viewing them textually, it may be hard to filter the more urgent alarms from others or even figuring out which alarms are currently active. Currently the alarms are organized in a single table, in a UI in the ATLAS control room. Such an interface serves well the basic needs of the online monitor, but theres also a clear need for an offline tool to mine into this data in case it is needed. Displaying this information in a visual and informative way can simplify the users work and optimize the data analysis process.

2 Technical Aspects of the Application

Splunk[3] is a commercial tool that can be used with a free license to a certain extent. It enables you to index large amount of information. Indexing is a process in which Splunk get raw text information, log information for example, and breaks it down to Events, usually according to timestamps. An event is described by a unique timestamp and various other attributes, called fields, that can be automatically or manually extracted during the indexing progress. For example: text lines in the from ¡timestamp¿ COMMENT=... VALUE=... will be indexed as events with distinct timestamps, and two fields: comment and value. These events can be easily filtered and displayed graphically in various ways by Splunk. Splunk is modular framework that supports applications which perform specific tasks. One of these applications, 'Splunk DB Connect' [4], enables Splunk to directly connect to a remote database and via SQL requests, parse log information which immediately starts getting indexed. After the information is parsed and indexed, it can be easily manipulated and presented. One useful functionality is called Drill down, and it allows you to dig deeper in your search results by clicking on fields or words of interest, narrowing the search and getting results that are more enlightening. Data mining can be done easily by clicking a part of a chart and drilling down, without any complicated filtering commands. For example, results that may initially display a wide range of sub detector systems can by drilled down by clicking on a specific sub detector in the table or chart. The data mining is complemented by graphics that may highlight areas of interest, for example an hour of the day where extensive log activity took place. Our knowledge of the systems allows us to preconfigure forms to display useful information in a way that was not possible in the past. Forms are pages that displayed preconfigured search results and charts, according to input given by the user like a time range for example. Some useful forms that DDM offers to DCS users:

- A graphic display of all the currently running alarms, sorted by duration. Choosing a specific alarm priority is done simply by selecting the desired priority from a drop-down menu

- One class of log messages contains information regarding disconnections between different subsystems. A distribution of these disconnections in respect to the target systems is presented both in a chart and in a table. Trends can be easily detected in this manner

- The Alarms Duration Timeline page (Figure 1) visualizes the duration of alarms in a line chart. This chart can be helpful in detecting periodicity of alarms. Such periodicities might help in finding deeper underlying problems in the system

A free search screen is also available to allow users with some Splunk knowledge to take advantage of what this powerful tool has to offer. This screen allows the user to get direct access to all events and using a search command entered in the search bar, filtering and displayed can be done according the users needs. Scheduled searches can be easily configured and performed using Splunk. This searches can be organized in a report and sent to a list of email recipients. Another use for a schedule search it to produce a notification in case of undesired events, for example an excess amount of alarms from a certain type.

3 Conclusion

Log information is presently a huge untapped source of information. Though being stored, it is not analyzed regularly. One of Splunks biggest advantages is its ability to facilitate the work of data analysts in making sense of large amounts of information. Alarms are hard to interpret due the fact that the interface is mostly text-based and not graphical. DCS Data Miner, using the Splunk application, has the power to filter DCS information efficiently and display the same information in an informative way, making the decision making process of the user much easier. Furthermore, Using Splunks built-in functionality saves time developing tools for any current or future analysis. Lastly, it is an easily expandable tool, since user-written custom scripts can be incorporated into it, enhancing its functionality even more.

References

[1] The ATLAS Collaboration, JINST **3**, S08003 (2008).

[2] S. Schlenker *et al.* ICALEPCS 2011, Conf.Proc.MOBAUST02 (2011).

[3] Splunks website `www.splunk.com`

[4] Splunk DB Connect `http://apps.splunk.com/app/958`

Figure 1: The Alarms Duration Timeline page