ATLAS EventIndex monitoring system using the Kibana analytics and visualization platform

D. Barberis[1], S.E. Cárdenas Zárate[2*], A. Favareto[3], A. Fernandez Casani[4], E.J. Gallas[5], C. Garcia Montoro[4], S. Gonzalez de la Hoz[4], J. Hrivnac[6], D. Malon[7], F. Prokoshin[2], J. Salt[4], J. Sanchez[4], R. Toebbicke[8], R. Yuan[6]**
on behalf of the ATLAS Collaboration

[1] Università di Genova and INFN, Genova, Italy
[2] Universidad Técnica Federico Santa María, Valparaíso, Chile
[3] Università degli Studi and INFN, Genova, Italy
[4] Instituto de Física Corpuscular, Valencia, Spain
[5] Department of Physics, Oxford University, Oxford, United Kingdom
[6] Laboratoire de l'Accélérateur Linéaire, Orsay, France
[7] Argonne National Laboratory, Argonne, USA
[8] CERN, Europe

*Presenter, primary author
**Corresponding author: Fedor.Prokoshin@cern.ch

Abstract. The ATLAS EventIndex is a data catalogue system that stores event-related metadata for all (real and simulated) ATLAS events, on all processing stages. As it consists of different components that depend on other applications (such as distributed storage, and different sources of information) we need to monitor the conditions of many heterogeneous subsystems, to make sure everything is working correctly. This paper describes how we gather information about the EventIndex components and related subsystems: the Producer-Consumer architecture for data collection, health parameters from the servers that run EventIndex components, EventIndex web interface status, and the Hadoop infrastructure that stores EventIndex data. This information is collected, processed, and then displayed using CERN service monitoring software based on the Kibana analytic and visualization package, provided by CERN IT Department. EventIndex monitoring is used both by the EventIndex team and ATLAS Distributed Computing shifts crew.

1. Introduction

The EventIndex [1] is a catalogue of the event-related information for the ATLAS experiment [2]. It is organized in a number of relatively independent modules or subsystems, collecting the event data from the datasets on the CERN central server (Tier-0) and on the Grid, sending this information to CERN, processing and storing it to the Hadoop [3] file system (HDFS) and, finally, exposing an interface for the queries by the users and production system. A smooth operation and interaction of these components is necessary for the effective work of the EventIndex and the success of the analyses depending of its data. We developed a system for automatically monitoring states, critical parameters

and performance numbers of the EventIndex components. It provides information to the EventIndex operation team through the Kibana [4] visualization package. In addition, global statuses of the critical components are displayed on the ATLAS central system monitoring page for ATLAS computing shifters. The system is in operation since late 2014.

2. Monitoring of the EventIndex components

The EventIndex consists of number of components of different type distributed between computers at CERN and on the Grid. Each of these components requires a different approach for data collecting and processing. Monitoring information is collected by jobs running on one of the EventIndex computers automatically at specified time managed by *acron* scheduler. The gathered information is organised into *xml* files and then pushed to the ATLAS computing services monitoring server running Kibana.

Kibana is a data visualisation software on top of the ElasticSearch engine. It provides a number of tools to display information, like dashboards, queries, filters and time ranges, so information stored in an ElasticSearch cluster can be visualized as interactive plots such as pie charts, histograms, text, maps and other.

The general scheme of the EventIndex monitoring is presented in Figure 1. Below we shortly describe its components.

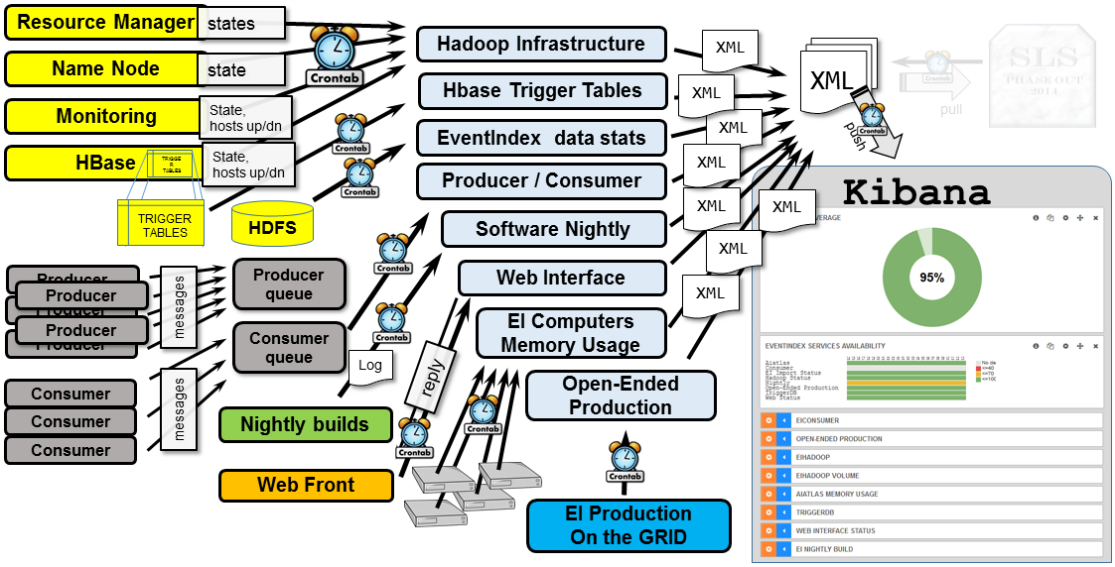

Figure 1: EventIndex monitoring components and data flow

The EventIndex components (consumers, importing and validation, grid production management and monitoring itself, web services) run on a number of servers at both CERN and the Wigner research center. We use some of these machines also for development. Currently we monitor accessibility and memory usage on these servers (see Figure 2) by running system information tools. Low available memory and aggressive swap usage may lead to performance degradation of the services they host, so If swap usage on some of these machines exceeds the limit, an alert for the EventIndex operation team is raised.

The EventIndex gathers information from the "producer" tasks running on the ATLAS central production servers (Tier-0) and on the Grid. The machinery that imports event data on the Grid is called "Open-ended production". It receives from the Atlas Metadata Interface [5] (AMI) information on datasets that were produced recently and supplies this information to ProdSys2 [6], the second-generation ATLAS distributed production system, which actually runs the tasks extracting event information. In the corresponding part of the monitoring we display information on the number of

datasets of different types found and fed to the processing by the EventIndex jobs, and on any errors found in this process. The information is gathered from the log files.

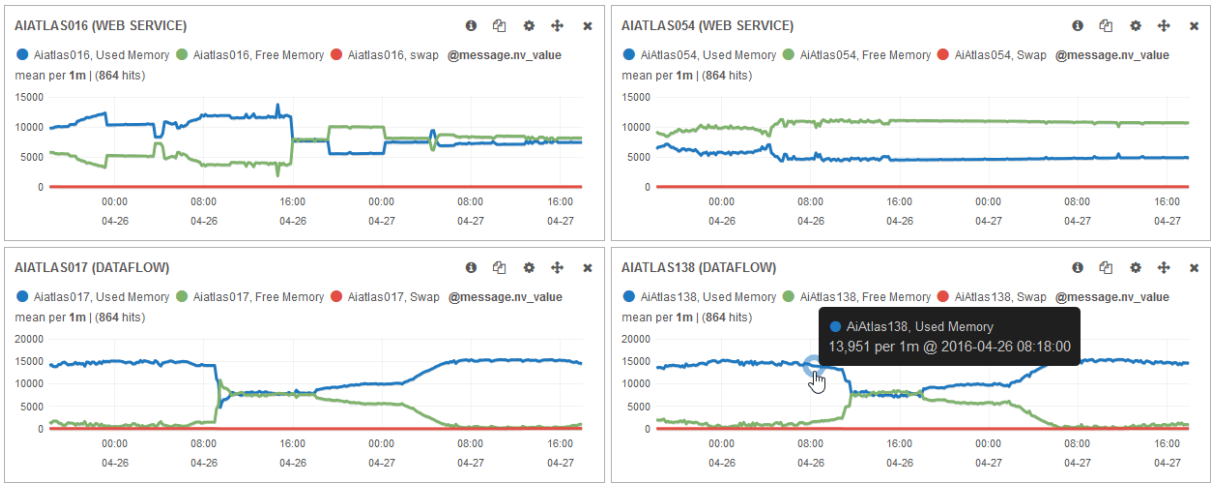

Figure 2: EventIndex servers memory usage monitoring

We track the performance of the producers that gather event information [7], consumers that receive this information from the producers and the messaging system that provides the transport of data between them by processing monitoring messages that are sent to special queues by the consumers. Figure 3 shows typical rates of the messages and amounts of data transfers. If no messages come for a long time or corrupted data appear, an alert is raised for both EventIndex operations and ATLAS computing shifters.

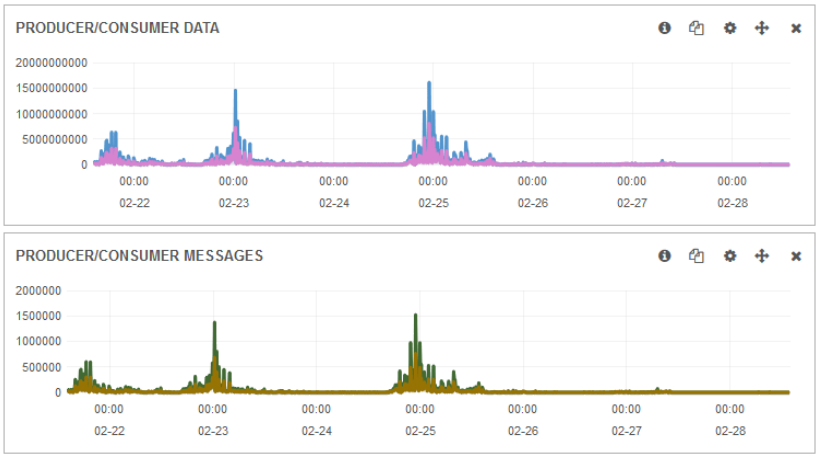

Figure 3: EventIndex Consumers monitoring

The EventIndex data are stored and managed by the CERN Hadoop cluster “lxhadoop”. Although Hadoop is designed so that it can automatically fix or bypass most of the problems, still there are critical components that cannot be fixed automatically. Also, experience obtained in the process of the EventIndex commissioning shows that simultaneous failures of few less significant components like one of 24 Hadoop hosts or 22 HBase region servers may still lead to significant degradation of the system performance. So we monitor the Hadoop and HBase master state and the number of inactive hosts and region servers through the corresponding webpages. Significant degradations of this system are reported.

The Hadoop storage system has a storage capacity around 1.5 PB and host most of the EventIndex data. Its size grew to more than 100 TB including transient files created by the consumers, with information on real data from Run 2 reaching 40 TB in half a year. We gather data volume information using HDFS tools. Figure 4 shows the data growth as presented by the monitoring display.

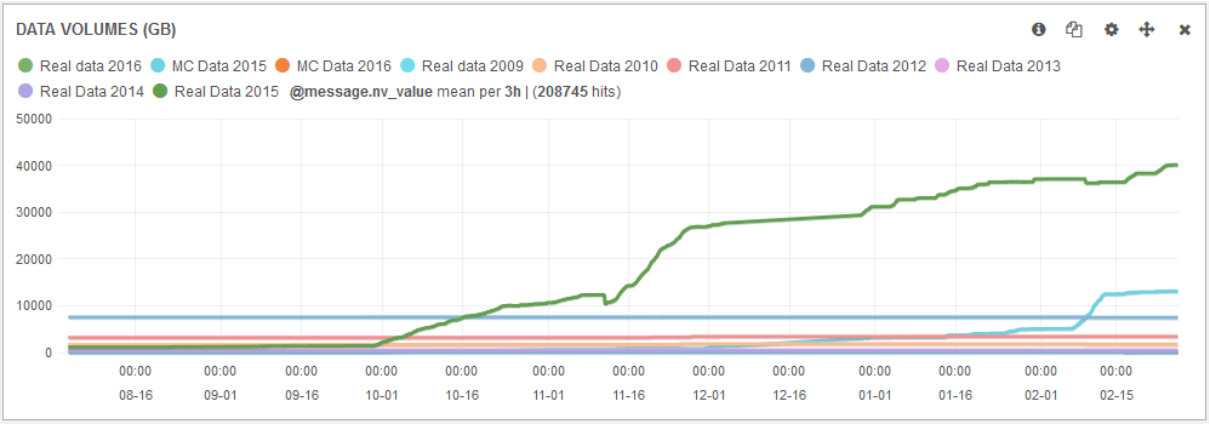

Figure 4: EventIndex data volume monitoring

The trigger decision information of each event appears as a bitmask that is decoded using trigger tables replicated from the ATLAS condition metadata database [8] (COMA). We monitor the status of replication by processing logs and plot the number of trigger tables replicated and the ranges of run numbers and trigger keys for which the trigger information is available.

The EventIndex can be accessed through the command line interface from any Linux computer running CVMFS (CERN network file system [9] optimized to deliver experiment software in a fast, scalable, and reliable way) or using a client package. However, the main way of accessing it is through the web interface and its GUI, or providing plain text commands suitable for the ATLAS distributed analysis system [10][11] (PanDA). We monitor the web interface by periodically sending simple *curl* queries.

3. Conclusions and outlook

System level monitoring in the new Kibana environment at CERN is ready and running, providing information for the EventIndex operation team and for ATLAS shifters. As the EventIndex system evolves and operating experience is accumulated, changes and additions are applied to the monitoring system in order to insure its reliability and ease of use.

References

[1] Barberis D *et al.* 2015 The ATLAS EventIndex: architecture, design choices, deployment and first operation experience, *J. Phys.: Conf. Ser.* **664** 042003 doi:10.1088/1742-6596/664/4/042003.
[2] The ATLAS Collaboration 2008 The ATLAS Experiment at the CERN Large Hadron Collider, *JINST 3 S08003* doi:10.1088/1748-0221/3/08/S08003
[3] Hadoop: http://hadoop.apache.org
[4] Kibana: https://www.elastic.co/products/kibana
[5] Solveig Albrand *et al.* 2010 The ATLAS metadata interface *J. Phys.: Conf. Ser.* **219** 042030 doi:10.1088/1742-6596/219/4/042030
[6] M Borodin *et al*. 2015 Scaling up ATLAS production system for the LHC Run 2 and beyond: project ProdSys2, *J. Phys.: Conf. Ser.* **664** 062005 doi:10.1088/1742-6596/664/6/062005
[7] J Sánchez *et al*. 2015 Distributed Data Collection for the ATLAS EventIndex, *J. Phys.: Conf. Ser.* **664** 042046 doi:10.1088/1742-6596/664/4/042046
[8] E J Gallas *et al.* 2012 Conditions and configuration metadata for the ATLAS experiment, *J. Phys. Conf. Ser.*, **396** 052033 doi:10.1088/1742-6596/396/5/052033
[9] A De Salvo *et al* 2012 Software installation and condition data distribution via CernVM File System in

ATLAS *J. Phys.: Conf. Ser.* **396** 032030 doi:10.1088/1742-6596/396/3/032030
[10] A Klimentov *et al*. 2015 Next Generation Workload Management System For Big Data on Heterogeneous Distributed Computing, *J. Phys.: Conf. Ser.* **608** 012040 doi:10.1088/1742-6596/608/1/012040
[11] De K *et al.* 2015 The future of PanDA in ATLAS distributed computing, *J. Phys.: Conf. Ser.* **664** 062035 doi:10.1088/1742-6596/664/6/062035