Evolution of Monitoring System for AMS Science Operation Centre

V Choutko\textsuperscript{1}, O Demakov\textsuperscript{1}, A Egorov\textsuperscript{1}, A Eline\textsuperscript{1}, B S Shan\textsuperscript{2,4} and R Shi\textsuperscript{3}

\textsuperscript{1} Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge, MA 02139, USA
\textsuperscript{2} Beihang University, 37 Xueyuan Road, Haidian Qu, Beijing 100191, China
\textsuperscript{3} Southeast University, 2 Sipailou, Xuanwu Qu, Nanjing, Jiangsu 210018, China

E-mail: baosong.shan@cern.ch

Abstract. The Alpha Magnetic Spectrometer [1] (AMS) has collected over 95 billion cosmic ray events since it was installed on the International Space Station (ISS) on May 19, 2011. The original science data collected by the detector are transferred to the Payload Operations Control Centre at CERN, and shared to the Science Operation Centre (SOC). SOC takes responsibilities to process the science data, including the format transformation, flight data reconstruction, and Monte-Carlo [2] (MC) simulation. The data produced by SOC are ready for detector performance evaluation and physics analysis. Different types of science data are processed by different procedures, some of the processing are done on SOC own computing facilities, and others by CERN-managed services, so it is important that all the processing procedures as well as the computing facilities and services are monitored. This paper presents the SOC monitoring tools and the integration with CERN monitoring infrastructure.

1. Data processing flow of SOC

Figure 1 shows that the data collected by the AMS detector are transferred by the Tracking and Data Relay Satellites (TDRS) to White Sands, and then transmitted to the POCC by the ground system. The original data, in the form of one-minute frame files, arrive at POCC nearly in real-time, and POCC shares the data to SOC via NFS [3].

1.1. Data preproduction

To facilitate detector calibrations and alignments, data of each quarter of the ISS orbit (between either pole and the equator), typically about 23 minutes, are labeled as one run. Data from the same run are decoded from frame files, and repacked into one RAW file. Each RAW file is copied to EOS [4] after completing the consistency check, and the metadata is inserted into an Oracle parallel database [5].

1.2. Data production

Once the metadata for the RAW files are recorded in the database, production jobs can then be requested and submitted to run on SOC own computing facilities or CERN LSF [6] hosts. Production jobs produce ROOT [7] files and Time Dependent Variable (TDV) files for the AMS

\textsuperscript{4} Corresponding author
Figure 1. Data collected by AMS are transferred by the Tracking and Data Relay Satellites (TDRS) to White Sands, NM, USA, and then transmitted to AMS Payload Operations Control Centre (POCC) at CERN by the ground transmission system.

conditional database. ROOT files are validated and uploaded to EOS, and the metadata are also recorded in the database.

AMS data production includes two stages: first production and second production. The first production runs in a fully automated manner and produces the data summary files for quick sub-detector performance evaluations. Usually the reconstructed data are available in two to three hours after the raw data arrive. The second production uses all the available calibrations, alignments and ancillary data from the ISS, as well as monitoring values (temperatures, pressures, and voltages), to produce the datasets ready for physics analysis. The second production usually runs every 6 months incrementally. However, in case of major updates of the reconstruction software, a full reproduction may be needed. Figure 2 shows the data flow of the first and second production.

Figure 2. Data flow of the first and second production.
1.3. Monte-Carlo production
The MC production reads simulation parameters and produces both the RAW and ROOT files for data analysis. SOC manages the MC production at CERN (running on LSF hosts), and the validation of MC files returned from the computing centres.

2. SOC shifts and monitoring tools
Daytime and evening shifts are scheduled to monitor the SOC activities and facilities, and various tools were developed to help the shift takers to check the status and to notify experts in case of issues.

2.1. Frame monitor
Figure 3 shows a plot produced by the frame monitor. Due to the incomplete coverage of TDRS satellites, the frame data does not arrive on a continuous basis. Typically the shift taker will notify the experts if no frame has arrived for more than one hour.

![Figure 3. Frame monitor plots the frame data arriving rate by log scale.](image)

2.2. List of data files
As shown in the following, the monitoring webpage lists the detailed information of each run, including run number (decimal and hexadecimal), run tag, path, timestamp, number of events, type, size, and the frame range. Shift takers can check the latency of the preproduction and the first production from this webpage.

```
1489481014 58c7ad36 f128 /afs/cern.ch/ams/Offline/RunsDir/SCI/1489000000/1489481014 14/03/2017 09:43:33 1032007 SCI 5 19 4 OK 2132 SCIBPB/RT/2766/438 SCIBPB/RT/2766/462 /eosams/Data/AMS02/2014/ISS.B1070/std/1489481014.00000001.root
```

2.3. Production monitor
The production monitor is designed to monitor and manage the first production. The production monitor uses the CORBA [8] protocol to communicate with the producer processes to get the running status and processing progress.
The production monitor also communicates with the production server so that an administrator can configure and manage the first production. For example, the administrator can modify the participating host list, and the numbers of threads that the producer should start with on different hosts.

Figure 4 is a screenshot of the production monitor. There are 5 pages which provide information of the producer processes, producer input/output, service status, control parameters, and production logs.

Figure 4. Producer input/output page of the production monitor. The upper part shows the production job information (input), including server number, run number, job identity, starting time, numbers of first and last events, job priority, job type, executing host, job status and job status history. The lower part shows the information of produced ROOT files (output), including server number, run number, ROOT file time, numbers of first and last events of the ROOT file, file path, checksum, size, and status.

2.4. Monte-Carlo production monitoring site

Unlike the “centralised” first production, Monte-Carlo production runs at various AMS computing centres. To identify different simulation tasks, MC production is organised in datasets, and within the same dataset the simulation jobs have the same particle type and the same version of simulation software. A dataset consists of one or more simulation job templates, and each job template has the same simulating parameters. A website is designed to show the progress of the production datasets, the estimation of the time to finish, and the contribution from different centres.

Figure 5 is the home page of the site, listing the overall status of active production datasets. By clicking on one of the datasets, or input a dataset name in the search box on top of this page, the dataset completing status page will be opened, as shown in Figure 6. Figure 7 shows the progress page of a job template, which plots the stacked completing ratio by date and computing centre (one can choose accumulative display or not). The parameters (data cards) of the job template, the produced file list and run list are also shown as links at the bottom of this page. Figure 8 shows the monthly CPU contribution of computing centres.
2.5. Hardware monitoring

SOC has 21 hosts (264 cores) and 300 TB storage for production and data storage. NetMonitor is a monitoring tool to poll the status of the SOC facilities (production servers/hosts and NFS servers) and the services running on them (NFS client, frame decoder, RAW validator, ROOT validator, SlowControlDataBase generator, etc.). In case of issues, NetMonitor will send out warning messages (by SMS and E-mail) to the system administrators.

3. Integration with the CERN monitoring infrastructure

Since we are increasing our use of the resources/services provided by CERN, the availability status of those resources/services are needed be monitored. For example, the availability and quota status of AFS [9] and EOS, and the availability of the LSF [6] and FTS3 [10] services. The CERN monitoring infrastructure provides various tools to manage the monitoring metrics and to virtualise them.
Figure 7. Job template completing status (by date/centre) page of the MC production monitoring site.

3.1. SOC monitoring metrics
The first production is used for detector performance evaluations, so monitoring the first production is of the highest priority. The key criteria for the first production is delay: the delay of frame files indicates if flight data are transferred to POCC correctly; RAW delay shows if production server, producers, and ROOT validator are working normally; fresh run delay (the time passed since the validation time of the most recent run) can be used to check if the first production jobs are requested timely.

3.2. Implementation
To describe the delays, a metric class named “amssoc.proddelay” is defined in “Lemon Metrics Manager” [11], and 4 integers are used to indicate the 4 delay time (in second). A metric instance with a unique ID and a sensor are defined based on the metric class, and the metric is registered on a Puppet [12] based virtual host. On this host, the “lemon-agent” service is installed, which runs “lemon-sensor” Python script periodically. Currently the sensor script is executed every 15 minutes, retrieving the 4 delay values, and sending the values into the Elasticsearch [13]. Once the values reach Elasticsearch, they can be queried and visualised. Figure 9 shows a Kibana [14] page, which is created with the metric ID as the querying filed to visualise the delay data.

3.3. Notification and trouble shooting
To provide automatic notification, thresholds are defined in an exception metric class, and when any of the delay values exceeds its threshold, then the trouble shooting program will send alerting
Figure 8. Monthly CPU contribution page of the MC production monitoring site.

Figure 9. Kibana page to monitor the delay data. Purple dots show the latency of frame files, green dots show the latency of RAW files. The blue and red dots show the elapsed time since the validation of the last validated run and the latest run (with the biggest run number) respectively.

A set of rules describing the trouble shooting procedures are defined in a database, which is called a rule base. When the trouble shooting program is executed by the delay exception, it goes step by step according to the rules defined in the rule base to diagnose the problem(s),
and notifications are sent to the experts about the production status as well as the potential reason(s) according to the diagnostics results. After each manual intervention, rule base can be updated by experts.

4. Conclusion
The AMS SOC monitoring tools are designed to check and report the status of SOC facilities and services. Based on the monitoring infrastructure provided by CERN, a monitoring system for SOC is designed, which provides a glance view of the first production status. To assist trouble shooting, the records of past actions of the experts are extracted into rules and stored into database. In case of issues, automatic trouble shooting attempts are carried out, and notifications with diagnostic results are sent to experts.

In future, we plan to add more sensors for SOC services, like RAW/ROOT validators, job requestor, data movers, etc., and to integrate their data to the monitoring page. Further, the resource availability data of CERN services (EOS/CASTOR [15]/AFS/CVMFS [16]) will also be integrated into the monitoring system.

References

[1] Ting S 2013 The alpha magnetic spectrometer on the international space station Nuclear Physics B-Proceedings Supplements 243 12–24
[2] Binder K 1987 Monte carlo methods Quantum Monte Carlo Methods 241
[3] Shepler S, Eisler M, Robinson D, Callaghan B, Thurlow R, Noveck D and Beame C 2003 Network file system (NFS) version 4 protocol Network
[4] Peters A J and Janyst L 2011 Exabyte scale storage at CERN Journal of Physics: Conference Series vol 331 (IOP Publishing) p 052015
[5] DeWitt D and Gray J 1992 Parallel database systems: the future of high performance database systems Communications of the ACM 35 85–98
[6] Zhou S 1992 LSF: Load sharing in large heterogeneous distributed systems I Workshop on Cluster Computing vol 136
[7] Antcheva I, Ballintijn M, Bellenot B, Biskup M, Brun R, Buncic N, Canal P, Casadei D, Couet O, Fine V et al. 2011 ROOT: A c++ framework for petabyte data storage, statistical analysis and visualization Computer Physics Communications 182 1384–1385
[8] Siegel J and Frantz D 2000 CORBA 3 fundamentals and programming vol 2 (John Wiley & Sons New York, NY, USA)
[9] Satyanarayanan M 1993 Distributed file systems Distributed Systems. Addison-Wesley and ACM Press, 821 145–154
[10] Ayllon A, Salichos M, Simon M and Keeble O 2014 FTS3: New data movement service for WLCG Journal of Physics: Conference Series vol 513 (IOP Publishing) p 032081
[11] Hogan J E, Lemon K N and Rust R T 2002 Customer equity management: Charting new directions for the future of marketing Journal of service Research 5 4–12
[12] Loope J 2011 Managing Infrastructure with Puppet: Configuration Management at Scale (“O’Reilly Media, Inc.”)
[13] Gormley C and Tong Z 2015 Elasticsearch: The Definitive Guide (“O’Reilly Media, Inc.”)
[14] Andrade P, Fiorini B, Murphy S, Pigueiras L and Santos M 2015 Monitoring evolution at CERN Journal of Physics: Conference Series vol 664 (IOP Publishing) p 052002
[15] Presti G L, Barrong O, Earl A, Rioja R M G, Fonce S, Taurelli G, Waldron D and Dos Santos M C 2007 CASTOR: A distributed storage resource facility for high performance data processing at CERN. MSST vol 7 (Citeseer) pp 275–280
[16] Aguado Sanchez C, Bloomer J, Buncic P, Franco L, Klemmer S and Mato P 2008 CVMFS - a file system for the CernVM virtual appliance Proceedings of XII Advanced Computing and Analysis Techniques in Physics Research vol 1 p 52