AMS data production facilities at science operations center at CERN

V Choutko¹, A Egorov¹, A Eline¹ and B Shan²

¹Massachusetts Institute of Technology, Cambridge, MA
²Beihang University, Beijing, China

E-mail: aegorov@cern.ch

Abstract. The Alpha Magnetic Spectrometer (AMS) is a high energy physics experiment on the board of the International Space Station (ISS). This paper presents the hardware and software facilities of Science Operation Center (SOC) at CERN. Data Production is built around production server - a scalable distributed service which links together a set of different programming modules for science data transformation and reconstruction. The server has the capacity to manage 1000 paralleled job producers, i.e. up to 32K logical processors. Monitoring and management tool with Production GUI is also described.

1. Introduction
The Alpha Magnetic Spectrometer (AMS) is a high energy physics experiment on the board of the International Space Station (ISS). The AMS Science Operation Center (SOC) at CERN receives and stores all AMS science and housekeeping data, ensures full science data reconstruction, calibration and alignment, keeps data available for further analysis at CERN and AMS Regional Centers. Data Production undertaken by SOC computing infrastructure, is split into three stages [1]:

• Pre-production - ~1 minute science data frames are de-framed by checking raw events for consistency and indexing, events are reassembled into ~23 minutes raw files, one raw file for each run.
• First production - raw events are reconstructed into AMS physics data format based on CERN ROOT [2] package. Data summary files are produced as well as event summary files for fast event selection.
• Second production - produces analysis ready datasets.

Physical data reconstruction during first and second production requires event processing done in parallel to shorten elapsed time to get physics data soon after the readiness of raw data.

2. SOC computing infrastructure
Figure 1 shows SOC computing infrastructure, including SOC own computing resources as well as dedicated CERN computing resources, rented by AMS experiment.

Its own computing resources consist of 21 hosts (or nodes) powered by dual or quad Intel Xeon processors, and run under SLC5/6, CentOS7 operating systems. It provides computing power for pre-production and first production running 24/7 on freshly arrived data, as well as for development of Data Production software.
Dedicated CERN computing resources consist of 390 virtual hosts with 16 logical processors each, and run SLC6 operating system. They are used primarily for physics analysis as well as for second production and MC-simulation. Second production runs every 3-6 months incrementally.

![Figure 1. SOC computing infrastructure.](image)

The main amount of disk space is also provided by CERN (see Table 1 for details). CERN EOS [3] file system has been chosen as main storage to keep raw data and resulting CERN ROOT files. Data access rate was measured about 100 MB/sec. To provide data safety all files are archived to CERN CASTOR [4] tape system. CERN CVMFS [5] read-only file system is used to deploy Data Production repositories and conditional databases to run data processing applications at CERN and Remote Computing Centers.

| File system | Capacity | Destination | Access |
|-------------|----------|-------------|--------|
| EOS         | 4 PB     | Main storage to keep raw data and resulting CERN ROOT files | Distributed: CERN, Remote Computing Centers |
| CASTOR      | -        | Tape archiver for all raw data and resulting CERN ROOT files | Distributed: CERN, Remote Computing Centers |
| AFS [6]     | 25 TB    | Conditional database with Time Dependent Values, logs of execution of Data Production modules, Data Production repositories and software development framework | Distributed: CERN, Remote Computing Centers |
| CVMFS       | 25 TB    | Fast read-only mirror of AFS | Distributed: CERN, Remote Computing Centers |
| SOC disk array | 300 TB | Redundant disk space | Only inside SOC computing subnetwork |

Developed Data Production software has the ability of operator assistant or automated redirection of IO streams in case of losing disk access, its performance degradation or by other technical reason.
3. Software facilities of AMS Data Production

As it was mentioned before, Data Production is split into pre-production, first and second production. It is built around Oracle relational AMS database providing complete catalogization of raw data and produced ROOT data files.

The software used on pre-production stage consists of single-threaded application called Frame Decoder to reassemble ~1 minute science data frames into ~23 minutes raw files for further use on first and second production. It does not require computing power and may run on SOC own front-ends.

Unlike pre-production, first and second production use computing power of ~80 and few thousand processors correspondingly. Developed software implements clients-server approach where physical data reconstruction done by paralleled job producers (one run for each job) is managed by scalable service called production server (the server). Other software modules are involved as server clients:

- WEB engine – prepares job scripts, submits reconstruction jobs to the server by schedule or by operator request.
- Job validator – validates finished jobs as well as ready ROOT data files, provides file transfer to disk storage. The list of finished jobs and ready data files is taken from the server.
- Oracle Berkeley DB [7] engine – is run-time mirror of the internal tables of the server, provides automated recovery of first and second production in case of hardware or software failures.
- Production GUI - graphical tool to manage the server and to monitor execution of remote job producers.

The server links all software modules together by CORBA [8] interface for object-oriented remote procedure calls. Used ORBIT2 API is extended with orbitcpp [9] bindings and is fully compatible with client software modules written with C++, Python and Perl programming languages. Each running instance of the server has the capacity to manage around 1000 paralleled producers, i.e. up to 32K logical processors. We use 2 different servers running first and second production.

4. AMS production server

Is used to manage parallel execution of job producers on SOC computing infrastructure. The predefined configuration of the server includes a set of AMS own computing nodes and (or) a set of dedicated CERN virtual computing nodes. The first ones have higher priority to start new producers.

The node to start an instance of the producer on the set of AMS own computing nodes is chosen by the server using the criteria (1):

$$\max_{i=1,n} (p_i - b_i) \times c_i, \ n < 21$$

where:

- $p_i$ - is predefined amount of logical processors per node,
- $b_i$ - is number of busy logical processors per node,
- $c_i$ - is predefined CPU ratio per node,
- $n$ - is number of predefined nodes.

The node to start an instance of the producer on the set of dedicated CERN virtual computing nodes is chosen automatically by CERN Batch Service [10]. For both kinds of nodes the number of logical processors to start a paralleled producer is chosen by the server basing on its configuration (usually 8-12 logical processors), and is acknowledged by the producer during initialization.

On start the producer requests from the server calibration constants and slow control corrections. The state of job processing is periodically sent to the server (once per ~20 sec). On finish the producer uploads the server with event tags for fast event selection (event summary files). Ready ROOT data files are validated and transferred to disk storage by job validator, the list of finished jobs and ready data files is taken from the server.

The server also implements:

- Monitoring of job execution status: waiting for submission, submitted, running, finished, pended, failed, killed, canceled.
• Automated job restarting on job failure (the number of retries is predefined) or by operator request.
• Automated job killing on execution timeout or by operator request.
• Logging of the server-producer conversation.

Production GUI provides the server with the follows features:
• Server configuration tools.
• Manual job management such as job canceling, restarting or removing from production.
• Job status monitoring (see Figure 2) and viewing job logging information.
• Server health monitoring.

5. Conclusions
The current status of SOC computing infrastructure is shown. The operation mode of Data Production is highly automated and normally does not require human intervention. Production GUI simplifies monitoring of Data Production as well as the diagnostics of job run results.

![Figure 2. Job status monitoring in Production GUI.](image_url)

![Figure 3. Typical monthly performance of first production.](image_url)
From the beginning of operations of AMS on the board of ISS, more than 120000 run files (~88 millions of physical events) have been processed by Data Production. The average delay from data arrival to SOC and data readiness for physics analysis was measured around 4.2 hours (see Figure 3).

References
[1] Choutko V, Shan B, Egorov A, Eline A. April 13, 2015. Computing Strategy of the AMS-02 Experiment. CHEP 2015, Okinawa
[2] Agostinelli S, Allison J, Amako K, Apostolakis J, Araujo H, Arce P, Asai M, Axen D, Banerjee S, Barrand G et al. 2003. Nuclear instruments and methods in physics research section A: Accelerators, Spectrometers, Detectors and Associated Equipment 506 250-303
[3] Large Disk Storage at CERN. URL: http://eos.web.cern.ch/
[4] The CERN Advanced STORage manager. URL: http://castor.web.cern.ch/
[5] CernVM File System. URL: https://cernvm.cern.ch/portal/filesystem
[6] OpenAFS Website. URL: http://www.openafs.org/
[7] Oracle Berkeley DB. The Industry Leading Embedded Database. URL: https://www.oracle.com/database/berkeley-db/db.html
[8] Bolton F. 2002. Pure CORBA. Sams Publishing, ISBN 0-672-321812-1
[9] C++ bindings for the ORBit CORBA ORB. URL: http://orbitcpp.sourceforge.net/
[10] CERN Batch Service URL: http://information-technology.web.cern.ch/services/batch