Mem. S.A.It. Suppl. Vol. 19, 384

Memorie della
Supplementi

AVES: A Computer Cluster System approach for INTEGRAL Scientific Analysis

M. Federici[1] B.L. Martino[2] L. Natalucci[1] P. Ubertini[1]

[1] INAF-IASF Istituto di Astrofisica Spaziale e Fisica Cosmica Roma – Via fosso del cavaliere 100, - 00133 Roma, Italy. e-mail: memmo.federici@iasf-roma.inaf.it

[2] CNR-IASI Istituto di Analisi dei Sistemi ed Informatica, Viale Manzoni 30, - 00185 Roma Italy.

Abstract. The AVES computing system, based on an Cluster architecture is a fully integrated, low cost computing facility dedicated to the archiving and analysis of the INTEGRAL data. AVES is a modular system that uses the software resource manager (SLURM) and allows almost unlimited expandibility (65,536 nodes and hundreds of thousands of processors); actually is composed by 30 Personal Computers with Quad-Cores CPU able to reach the computing power of 300 Giga Flops (300×10^9 Floating point Operations Per Second), with 120 GB of RAM and 7.5 Tera Bytes (TB) of storage memory in UFS configuration plus 6 TB for users area. AVES was designed and built to solve growing problems raised from the analysis of the large data amount accumulated by the INTEGRAL mission (actually about 9 TB) and due to increase every year. The used analysis software is the OSA package, distributed by the ISDC in Geneva. This is a very complex package consisting of dozens of programs that can not be converted to parallel computing. To overcome this limitation we developed a series of programs to distribute the workload analysis on the various nodes making AVES automatically divide the analysis in N jobs sent to N cores. This solution thus produces a result similar to that obtained by the parallel computing configuration. In support of this we have developed tools that allow a flexible use of the scientific software and quality control of on-line data storing. The AVES software package is constituted by about 50 specific programs. Thus the whole computing time, compared to that provided by a Personal Computer with single processor, has been enhanced up to a factor 70.

Key words. aves: cluster - slurm: resources manager- osa: scientific software package

1. Introduction

The INTEGRAL mission (Winkler et al. 2003), operational since 2002, has already generated an archive of pre-processed data of about 9 TB size up to August 2010. Given the current data rate it is foreseen that the archive storage needs to be incremented by about 1.5 TB/year, at least until 2012 following the approved mission extension or 2014 in the case of a further required extension. In the meanwhile, more scientific analysis applications require to span over time periods of years with thousands of Science Windows (SCW, is a single pointing lasting 2000-3000 s as part of a

Send offprint requests to: M. Federici

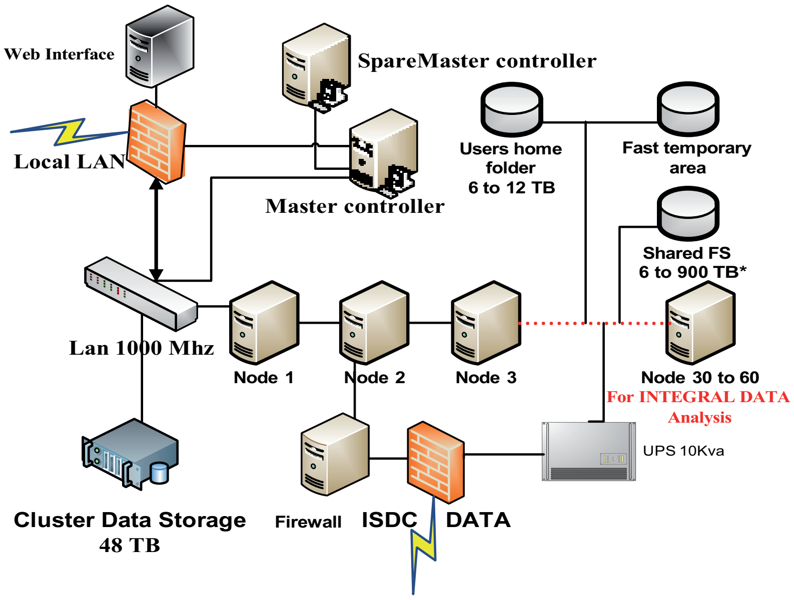

Fig. 1. block diagram of the hardware structure of AVES. ∗: 480 TB of UFS capability for 30 nodes configuration.

dither pattern for the basic observation) to be performed in a single run. In fact, using conventional PC systems this results in an overburden especially in terms of computing time that can be of the order of months. We have then approached a strategy aimed to increase the computing power while keeping the system costs to a relatively low range. Due to the large amount of files that compose the data package, scientific analysis has some additional complexities determined by the validation of data in the database. In fact performing the analysis on the conventional systems used before AVES, the user may experience random crashes during the run caused by various types of data files often missing or defective. The worst case occurs when the job is close to end (after running for several days) and undergoes crash. This event (unfortunately frequent) makes unusable the entire analysis and it is therefore necessary to re-raise the run after correcting the problem. The occurrence of this phenomenon causes a significant slowdown of scientific production. AVES software is designed to solve this problem brilliantly. With the ability to produce a breakdown of scientific results for each run, the user is enabled to use only the partitions "slices" deemed valid and meaningful. In particular, the validation of data structures has been made a tool easy to use, which operates a comprehensive quality check of files involved in the analysis, automatically correcting the list file and generating a number of files used for statistics. For the final solution to the problem of "quality check "of the entire data archive (currently about 9 Tera Bytes) is being considered an automatic global validation procedure able to build a dynamic database through which the user can validate the files list using simple "Queries". Access to the database will allow users to further accelerate the validation of their input list, increasing the global performance of AVES.

2. Overview

The priorities driving the AVES designing were low cost and high performance computing capability. To achieve this goal it was decided to use only conventional and readily available components. The mechanical structure that houses the individual components of the cluster is composed of a common, very robust and yet extremely economical shelving. This is suitable to accommodate up to 60 units/nodes (240 CPU). AVES is housed in an air conditioned (redundant) environment that guarantees the correct working temperature. The individual nodes that compose the cluster are from commercial personal computer, then low-cost hardware with good quality. The diagram in Fig. 1 shows the current configuration of Aves. It is composed of 30 nodes, each equipped with Intel Quad-Core processor, 4 GB of DDR2 RAM with access to 1066 Mhz, and 250 GB of Hard Disk. Overall Aves develops with its 120 GB of RAM and 120 CPU computing power close to 300 Giga Flops. The cluster is managed by two redundant Master unit, and on both is installed the resource manager SLURM (Simple Linux Utility for Resource Management) (Yoo et al. 2003). The scientific data reside on a server with NAS DATA-RAID-6 storage capacity of 48 TB and with redundancy built by an identical unit. AVES has a further memory resource. This space is achieved through a UFS file system (Union File System) that allows the union and uses free space on hard disk of each Node. The expandability of the AVES USF in the current configuration consists of 30 nodes, using conventional low cost (about 90 Euro for 1 TB) hard disk may reach over 480 Tera Bytes. The result of scientific analysis is stored on a NAS RAID-5 of the current capacity of 6TB. Access to AVES is via a secure remote connection in the ”ssh” protocol which allows users to export graphical windows. It is currently under study a system for user access through a web interface. AVES has an individually controlled power supply (UPS) controlled via WEB. The power of 10kVA can guarantee, in the absence of the main power, an endurance of several hours for a load of up to 60 nodes.

3. Software strategy

Actually AVES is a calculation tool designed to perform only the scientific analysis of data obtained from the INTEGRAL satellite, but we can customize the features of AVES for other projects. Its versatility allows easy use for other computing applications in particular, thanks to the great expansion of memory UFS AVES can be successfully used where a large storage space is needed. Scientific tests are carried out using the package developed from ISDC: OSA (Goldwurm et al. 2003). OSA does not seem suitable to perform multiple runs unless one subdivides a single run into multiple instances of calculation in succession, so that it can emulate a split run ”parallel”. A series of graphical I/F (GUI) has been produced to efficiently use the features of the calculation in clustering and to facilitate their use by all users.

3.1. Software Flow Charts

The AVES handling software is composed of about 50 scripts making use of the bash shell language. Actually it is possible to group the different tasks performed by these programs into 4 procedures two of which are described below for clarity.

3.2. AvesLogin

The diagram in Fig. 2 shows how the procedure manages the access of users, the setting of needed environment variables and provides information about available software and objects already created or analyzed. Then it activates graphical interfaces to manage the entry of parameters. Through scripts that make this procedure the user can run ”N” instances of the calculation for how many resources have been allocated. Then he can remotely control the status of the various running jobs choosing between all the remaining active sessions (which are therefore in preparation) using any client that allows a session to ”ssh”. This feature is essential because it allows the user to close his login session without closing the job in progress.

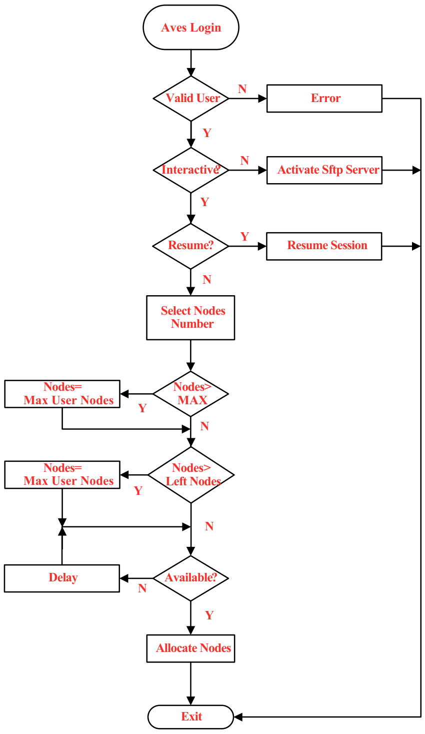

Fig. 2. Login procedure.

3.3. ClustStorageInit

The ClustStorageInit procedure makes the merge in a single folder all free space available from the mass memory of each node of the cluster and allows to use it as a shared UFS storage space(Union File System). Then, in case of success, it enable the essential services for the communication between nodes and the resource manager (slurm).

4. Graphical User Interface (GUI)

AVES use is facilitated by a series of graphical interfaces that allow easy selection of the parameters for analyzing and saving the settings used for future reuse. The "GUI" interfaces consists of dozens of graphical windows which are input parameters for analysis. The AVES GUI allow the users to overcome the difficulties of placing parameters of the analysis programs "OSA".

5. Scientific Analysis Example

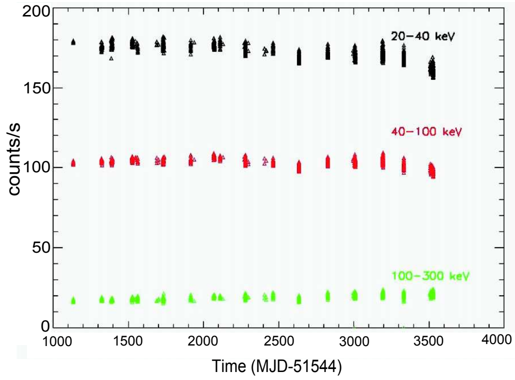

Fig. 3. Long term Crab light curves.

From raw data files it is possible to create higher level science products such as images and spectra integrated on each science window. The processing includes different stages as individual pixel threshold and gain corrections, rise time and energy computation for each event, filtering for good time intervals, etc. Sky images are then reconstructed by iterative fitting of the corrected and histogrammed data with models computed using an initial cross-correlation and expected source positions from catalogs. Once the images are extracted fluxes and positions are given in output and these data in turn can be used e.g. to build long term light curves or study spectral variability for periods of up to several years. The example given in Fig. 3 is a plot obtained from the analysis of 1415 pointings with the Crab Nebula in the Field of View of the INTEGRAL/IBIS (Ubertini et al. 2003) telescope. The observations span a period of 6.5 years from 2003 to 2009. The total exposure time is approximately 3 x 10^6 s. The production and analysis of 5725 images (1415 pointings times 5 energy bands) required a computing time of about 7.3 hours in the AVES cluster with 20 core CPUs. The same analysis performed with single core PC, requires a time of the order of 10 days.

6. Conclusions

Aves is a cluster computer system designed for scientific analysis of the INTEGRAL observatory data. The AVES computing performance, are a factor between 40 and 70, better if compared to a system of analysis done on workstations with single core. Due to the large amount of data produced so far from the satellite (about 9 TB) it is no longer possible to make a massive analysis with desktop computer systems. AVES was an economic response to this problem.

Acknowledgements. MF is grateful to Giuliano Sabatino purchasing manager of IASF-Rome for its valuable work. This system has been produced with the ASI contract. The IASF authors acknowledge the ASI financial support via grant ASI-INAF I/008/07/0/.

References

Winkler, C., et al. 2003, Astronomy and Astrophysics, 411, L1

Yoo, A., Jette, M. and Grondona, M. 2003, Job Scheduling Strategies for Parallel Processing, volume 2862 of Lecture Notes in Computer Science, pages 44-60

Goldwurm, A., et al. 2003, Astronomy and Astrophysics, 411, L223

Ubertini, P., et al. 2003 Astronomy and Astrophysics, 411, L131