2016ASPC..505..220A

Astronomical Surveys and Big Data
ASP Conference Series, Vol. 505
A. M. Mickaelian, A. Lawrence, & T. Yu. Magakian, eds.

Computational Astrophysics Towards Exascale Computing and Big Data

Hrachya V. Astsatryan,[1] Aram V. Knyazyan,[1] and Areg M. Mickaelian[2]

[1]*NAS RA Institute for Informatics and Automation Problems, Yerevan, Armenia;* `aram.knyazyan@ipia.sci.am,hrach@sci.am`

[2]*NAS RA V. Ambartsumian Byurakan Astrophysical Observatory, Byurakan 0213, Aragatzotn Province, Armenia;* `aregmick@yahoo.com`

Abstract. Traditionally, Armenia has a leading position both within the computer science and Information Technology and Astronomy and Astrophysics sectors in the South Caucasus region and beyond. For instance recent years Information Technology (IT) became one of the fastest growing industries of the Armenian economy (EIF 2013). The main objective of this article is to highlight the key activities that will spur Armenia to strengthen its computational astrophysics capacity thanks to the analysis made of the current trends of e-Infrastructures worldwide.

1. Introduction

Modeling and numerical simulations are considered to be the third pillar of the science after theory and experimentation being the heart of multiple domains, such as scientific, economic, or financial. Since modeling and simulations are essential for multiple scientific advances, the control of all the aspects of high performance computing (HPC) and the capacity to exploit the masses of data to tackle the solution of these complex models is inescapable. In most areas, the HPC challenges are related to the modeling of complex systems for a better understanding of larger and more complex problems. Coupling codes from different research areas allows for new perspectives and enables us to tackle complex multi-physics phenomena by taking advantage of the knowledge of each discipline involved. Over the past few years resilience has became a major issue for HPC systems, especially for large petascale systems and future exascale systems (Dongarra et al. 2011). "Big Data" is now considered to be the fourth pillar of science (Hey et al. 2009), and - just as HPC is - understanding and overcoming "Big Data" challenges are crucial for science and industrial advancements and competitiveness. The volume and the complexity of data do not stop growing.

Recent years the earth science community, such as astronomy and astrophysics, depends on “Big Data” and HPC, because it is based on large data archives which contain both observed and measured data, as well as simulated one (Reed and Dongarra 2015). It involves the exploration, analysis and reprocessing of high volumes of data as well as the modeling and simulation of complex coupled systems on multiple scales, and the exhaustive evaluation of multi-dimensional parameter spaces.

The main objective of this article is to highlight key activities that will spur Armenia to strengthen its computational astrophysics capacity thanks to the analysis made of the current trends of computational astrophysics worldwide. The current State of e-infrastructures and Computational Astrophysics Infrastructure in Armenia is presented in Section 2. The further activities to strengthen the computational astrophysics capacity in Armenia can be found in Section 3. Finally, the conclusion and directives for future research are drawn in the conclusion section.

2. Computational astrophysics infrastructure in Armenia

The infrastructure collects and stores results of astrophysical simulations and provides data analysis resources to researchers of Byurakan Astrophysical Observatory of the National Academy of Sciences of Armenia (BAO; http://www.bao.am) and beyond. The infrastructure consists of the data resources, computing resources, data management resources and services.

2.1. Data resources

The core of the data resources is the Digitized First Byurakan Survey (Mickaelian et al. 2007), which is the largest and the first systematic objective prism survey of the extragalactic sky. It covers 17,000 sq.deg. in the Northern sky together with a high galactic latitudes region in the Southern sky. The DFBS has been carried out in 1965-1980 using the Schmidt telescope (1.5 deg. Prism). Each FBS plate contains low-dispersion spectra of some 15,000-20,000 objects; the whole survey consists of about 20,000,000 objects. The objects selection can be made by their color, broad emission or absorption lines, SED in order to discover, classify and investigate them. The observations received from Byurakan 1m, 0.5m, and 0.2m Schmidt telescopes (photographic plates), classical 2.6m telescope (photographic plates and films) and smaller old telescopes are constituted in the electronic observational journals. All these telescopes now have all observational data in digital form: names, coordinates, and magnitudes of the observed objects, equipment, receiver, emulsion, filters, date, time, and exposure of observations, sky and weather conditions, and observers. The Byurakan electronic plate database contains data for 35,000 photographic plates and film exposures. Search by different parameters is available (observing programs, telescopes, observing mode, dates, emulsions, observers, etc.).

2.2. Computing resources

The computing resources of the Armenian e-infrastructure are included in a complex national IT infrastructure consisting of both communication and distributed computing infrastructures (Astsatryan et al. 2015a). The backbone of the Academic Scientific Research Computer Network of Armenia (ASNET-AM) is used to have access both computational resources and services (Fig. 1).

The computational resources of the Armenian National Grid Initiative (ArmNGI) are used by enabling to pursue a variety of scientific users in utilizing the Grid for their applications. The computational resources (about 500 cores) of Armenian Grid infrastructure distributed among our leading research (National Academy of Sciences, Yerevan Physics Institute) and academic (Yerevan State University, State Engineering University of Armenia) organizations are located in the cities of Yerevan and Ashtarak (Fig. 2).

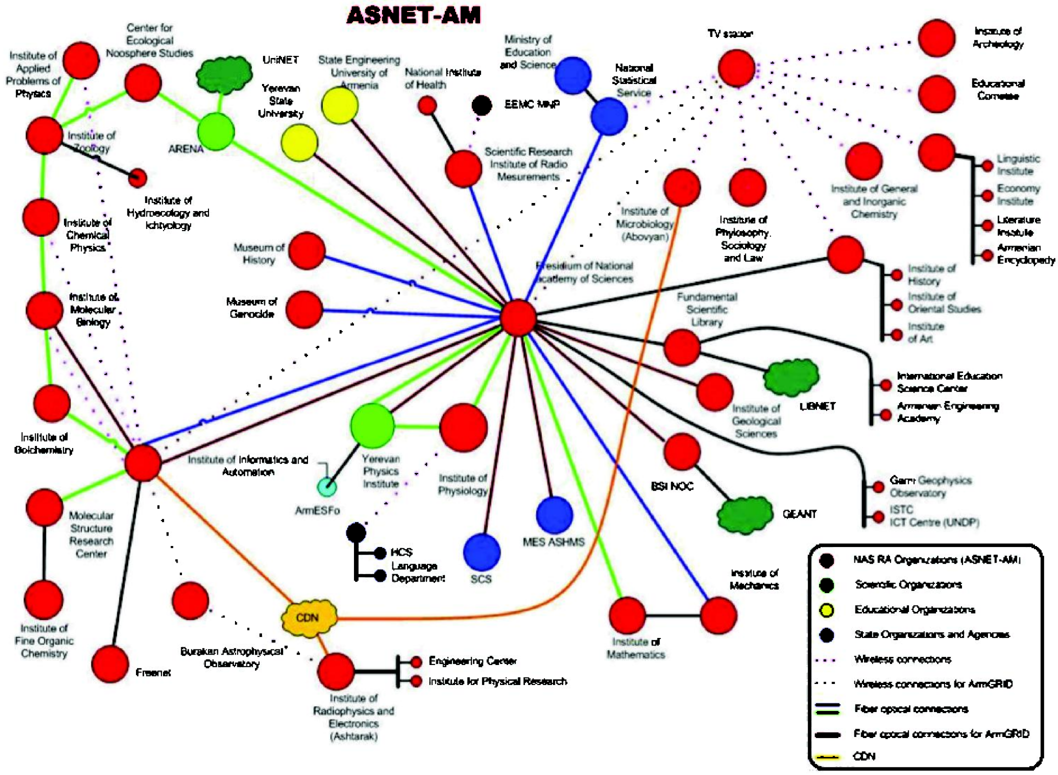

Figure 1. Topology of ASNET-AM.

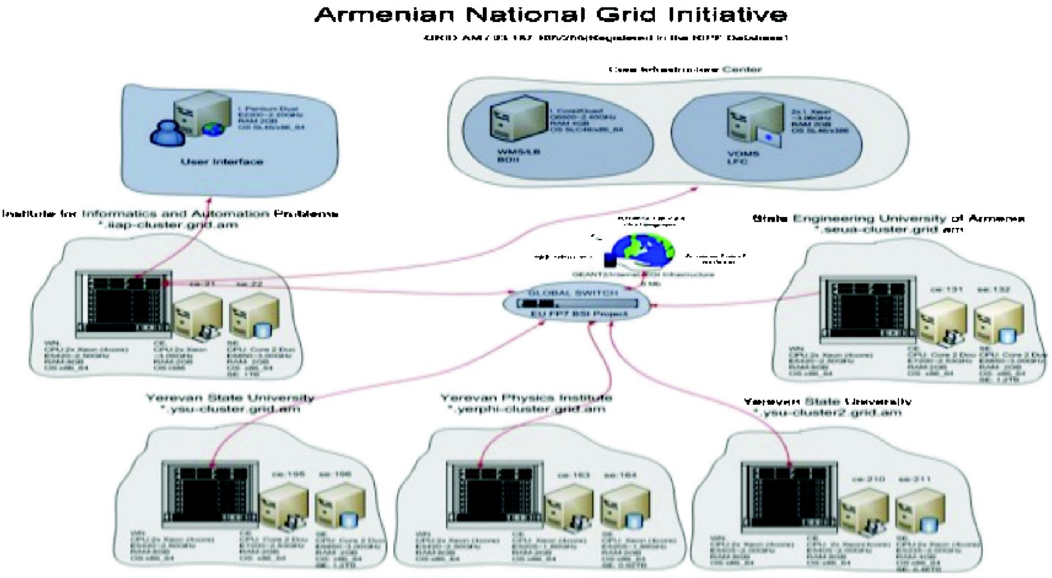

Figure 2. Topology of ArmNGI infrastructure.

2.3. Data management resources

The data management service provides a way to manipulate tables of astronomical data in a uniform way. A data management service allows authorized users to upload the files remotely via Web interface or via special scripts from the access node. All real data is stored in the dedicated storage element of the Armenian Grid infrastructure and the metadata in the database. The data management portal provides a data search engine to find images by coordinates, search-area radius and equinox. The result of search engine includes some basic parameters of a requested image like observation date, coordinates, image size, etc.. Then image can be viewed through the web more detailed, or user can download Flexible Image Transport System (FITS) file of image for more detailed research. User can make a gallery of featured images and download it as one zipped file (Astsatryan et al. 2010).

API service for astronomical data sharing is based on IVOA Simple Image Access Protocol standard by giving a possibility to make a search request through the http, and as a response get astronomical metadata in VOTable format. Different required and optional parameters for search criteria can be passed through http request, such as POS (RA and DEC coordinates), SIZE (search-area radius), image format and size, naxis, etc.

2.4. Cross-correlation Service

The cross-correlation service (Knyazyan et al. 2011a,b) enables to compare, i.e., cross-correlate the astronomical catalogs resulting from different observations of the same part of the space. Cross-correlation of two catalogs is a comparison of coordinates line by line. The catalog's each line coordinate is correlated to the second catalog's all lines coordinates and if the difference of the values of the coordinates is smaller than a prior determined certain value, then it can be concluded that the same object is present in both the first and the second catalogs. As for those lines, which don't have conjunction, deserve serious exploration. In case there is no conjunction, it means that the object is a new one, a proper motion star or an asteroid. However, it isn't always applicable, since there is a possibility that the mentioned object emerges just as a "random signal", which can be an error of the extraction making program or an error arising from the digitization of the plate, etc. The selection of the comparison radius value is also of a very high significance.

3. Further activities to strengthen the computational astrophysics capacity in Armenia

One of the great opportunities in Armenia is that it is still possible to coordinate the development of the major IT infrastructures to tackle scientific and societal challenges, which has not always been the case in other countries, leading to occasional inefficient investments. The first step to strengthening the computational astrophysics capacity of Armenia was the joint project of BAO and the Institute for Informatics and Automation Problems (IIAP; http://iiap.sci.am) aimed at the deployment of a modern virtual digital environment to tackle data management challenges.

The further activities to strengthen the computational astrophysics capacity in Armenia include not only to extend the digitized material, but also to provide new services and possibilities to the user communities.

3.1. Armenian Virtual Observatory

The International Virtual Observatory (IVOA) and its European implementation are designing a set of international open standards and tools for the discovery, access and processing of astronomical data. VO is a platform for launching astronomical investigations by providing an access to huge data banks, software systems with user-friendly interfaces for data processing, analysis and visualization. The Armenian Virtual Observatory project (ArVO) is being developed since 2005 (Knyazyan 2011). The DFBS is one of the key contribution of ArVO into the International Virtual Observatory Alliance. It is planned to build and implement SOA set of services to tackle authentication and authorization (GSI based), data discovery, task execution, access to storage systems (such as VOSpace).

3.2. Implementation of standards

It is planned to implement widely the international open standards for the discovery, access and processing of astronomical data. Based on IVOA standards we plan to built-in metadata services, replication and distributed storage, which will provide access, search, retrieval and in situ processing for this data using the protocols and standards. These standards are already in use at astronomical facilities across the globe, and implementing them enables us to provide advanced data retrieval services to end-users. It will be incorporated with the OGC (Open Geospatial Consortium) standards to offer various services to the earth science community (Astsatryan et al. 2015b,c,d).

3.3. Services

We plan to extend the portal functionality by providing new services to the user communities. As for data discovery tools, we plan to implement the Aladin (Bonnarel et al. 1996) and Open SkyQuery (http://openskyquery.net/Sky/SkySite). The Aladin desktop/light will be customized addressing challenges such as locating data of interest; accessing and exploring distributed datasets, visualizing multi-wavelength data. Compliance with existing or emerging VO standards, interconnection with other visualization or analysis tools, ability to easily compare heterogeneous data are key topics allowing Aladin to be a powerful data exploration and integration tool as well as a science enabler. The Open SkyQuery allows users to cross-match astronomical catalogs and select subsets of catalogs with a general and powerful query language. Users can also import a personal catalog of objects and cross-match it against selected databases. As analyzes tools we will use SPLAT (Draper 2013) and Specview. SPLAT is a graphical tool for displaying, comparing, modifying and analyzing astronomical spectra stored in NDF, FITS and TEXT files. SPLAT is part of the STARJAVA collection. Specview is a tool for 1-D spectral visualization and analysis of astronomical spectrograms. It is written in Java thus can be run anywhere Java is supported. Specview is capable of reading all the Hubble Space Telescope spectral data formats, as well as data from several other instruments (such as IUE, FUSE, ISO, FORS and SDSS), preview spectra from MAST, and data from generic FITS and ASCII tables.

VisIVO (Becciani et al. 2010) is an integrated suite of tools and services specifically designed for the Virtual Observatory. This suite constitutes a software framework for effective visual discovery in currently available (and next-generation) very large-scale astrophysical data sets. VisIVO consists of VisiVO Desktop, a stand alone application for interactive visualization on standard PCs; VisIVO Server, a grid-enabled platform for high performance visualization; and VisIVO Web, a custom designed web portal supporting services based on the VisIVO Server functionality.

As alternative to clusters, grids, and supercomputers, we plan to use Cloud computing resources, as elastic computing model where users can lease computing and

2016ASPC..505..220A

storage resources on demand from a remote infrastructure. In order to use effectively the distributed resources, the data and computing will be parallelized and distributed trough MapReduce (Dean and Ghemawat 2010) or other parallelization technologies on machine learning platforms (e.g., Hadoop, Graphlab, Spark, etc.). Hadoop (http://hadoop.apache.org) is an open-source software framework for storing data and running applications on clusters of commodity hardware. It provides massive storage for any kind of data, enormous processing power and the ability to handle virtually limitless concurrent tasks or jobs. Apache Spark using Hadoop is a fast and general engine for large-scale data processing.

4. Conclusion

The further activities will enable to extend the portal functionalities using the international standards, well-known protocols and software tools. It will increase the visibility of Armenian community integrating it in known astronomical search engines.

References

Astsatryan H., Sahakyan V., Shoukourian Yu, Cros P.H., Dayde M., Dongarra J., Oster P. 2015a, IEEE Proc. 14th RoEduNet International Conference - Networking in Education and Research (NER'2015), Craiova, Romania, September 24-26 2015, p. 28-33

Astsatryan H., Hayrapetyan A., Narsisian W., et al. 2015b, Elsevier Computer Standards & Interfaces 31(40), 79

Astsatryan H., Narsisian W., Hayrapetyan A., et al. 2015c, Springer Earth Science Informatics 8, 453

Astsatryan H., Petrosyan Z., Abrahamyan R., et al. 2015d, 11th IEEE eScience Conference, Munich, Germany, 31 Aug - 04 Sep 2015, p. 207

Astsatryan H., Knyazyan A., Mickaelian A., Sargsyan L. 2010, Parallel computing and control problems, Moscow, p. 109

Bonnarel F., Ziaeepour H., Bartlett J., et al. 1996, IAU Proc. Series, Vol. 179, p 469

Becciani U., Costa A., Antonuccio-Delogu V., et al. 2010, PASP 122, 119

"Information and Telecommunication Technologies Sector in Armenia" 2013, Enterprise Incubator Foundation, Yerevan, Armenia

Dongarra J., Beckman P., et. al. 2011, International J. High Performance Computing 25, 3

Dean J., Ghemawat S. 2010, MapReduce: A Flexible Data Processing Tool - Communications of the ACM 53(1), 72-77

Draper P. W. 2013, Starlink SPLAT-VO, [Online], Available: http://star-www.dur.ac.uk/pdraper/splat/splat-vo/

Hey T., Tansley S., Tolle K. 2009, The Fourth Paradigm: Data-Intensive Scientific Discovery," vol. 1. Redmond, WA: Microsoft Research

Knyazyan A., Mickaelian A., Astsatryan H. 2011a, Proc. CSIT-2011, Yerevan, p. 395-397

Knyazyan A., Mickaelian A., Astsatryan H. 2011b, Inform. Theories and Applications 18, 243

Knyazyan A. 2011, Astrophysical Virtual Observatories and Armenian Virtual Observatory Development – Vestnik, vol. 14, Yerevan, Armenia, p. 111-119

Mickaelian A., Nesci R., Rossi C., et al. 2007, A&A 464, 1177

Reed D., Dongarra J. 2015, Exascale Computing and Big Data Communications of the ACM – Vol. 58, No. 7, p. 56-68