Astronomical Data Analysis Software and Systems XXVI
ASP Conference Series, Vol. 521
Marco Molinaro, Keith Shortridge, and Fabio Pasian, eds.

Accelerated Distributed Visualisation in The Theoretical Astrophysical Observatory (TAO)

Elliott Ayling,[1] Amr Hassan,[2] Tim Dykes,[1] Christopher Fluke,[2]
Claudio Gheller,[3] and Mel Krokos[1]

[1]*University of Portsmouth, Portsmouth, Hampshire, United Kingdom;*
elliott.ayling@port.ac.uk,tim.dykes@port.ac.uk,mel.krokos@port.ac.uk

[2]*Swinburne University of Technology, Melbourne, Victoria, Australia;*
ahassan@swin.edu.au,cfluke@swin.edu.au

[3]*CSCS-ETHZ, Lugano, Ticino, Switzerland;* cgheller@cscs.ch

Abstract. The Theoretical Astrophysical Observatory (TAO) is a virtual laboratory that provides pipelines to build mock observations of galaxy survey data and other derived data products based on different cosmological dark matter simulations and semi-analytic galaxy formation models. We present recent work towards web-based interactive visual exploration of large datasets in TAO, using the high performance volume rendering technique Splotch - a ray-casting approach for visualising complex and multivariate large-scale point clouds. We outline how Splotch was adapted to accommodate specific TAO requirements and discuss work to expose functionalities through a web interface giving non technical TAO users the possibility to deploy Splotch in an HPC environment for easy access to state-of-the-art simulated data for a wide user base of astronomers.

1. Introduction

Turning massive raw datasets into useful scientific outcomes is a challenge that astronomers have to face in their day-to-day activities. Providing astronomers with easy-to-use tools to access, analyse, visualise, and process this raw data and its derived data products is a vital step towards a better and more efficient utilisation of modern datasets. Cloud-based data processing and analysing pipelines are thought to be a good aid in addressing such usability and accessibility challenges. They can provide on-demand access to large computational resources and data repositories, efficient tools for data processing, and customised data analysis pipelines via easy-to-use web interfaces (or thin-clients). Visualisation with interactive human feedback can be a really effective means of data inspection, and is becoming an increasingly reliable analytical tool for scientists over the recent years. However, due to the inherent difficulty of implementing efficient in-situ visualisation functionalites in cloud-based environments, many pipelines depend on completely automated processes with no feedback mechanisms.

The Theoretical Astrophysical Observatory (TAO) is part of the All-Sky Virtual Observatory (ASVO) project which constitutes the first effort to federate astronomical datasets from different astronomical facilities in Australia. TAO is an e-Research flagship project as it has been highlighted within the decadal plan for Australian As-

tronomy (2016-2025). TAO hosts different cosmological dark matter simulations and hydrodynamic simulations and provides the scientists with pipelines on-demand to extract derived data products out of these simulations to serve a range of highly specialised scientific needs. The main advantage of TAO is its intuitive web interface which allows non-technical users to exploit HPC architectures and query massive datasets easily.

The work reported in this paper is a step forward towards integrating high performance enabled 3D visualisation with cloud-based data processing pipelines. We have integrated the HPC-enabled volume rendering algorithm Splotch (Dolag et al. 2008) within the Theoretical Astrophysical Observatory (TAO) (Bernyk et al. 2016) to provide end-users with an efficient and easy-to-use mechanism for visualising catalogues produced via the TAO science modules. We expect this new functionality to enable users to perform further exploration and discovery activities once the main areas of interest, relevant characteristics, or intriguing correlations have been identified.

2. Preparing Splotch for integration within TAO

Splotch is a volumetric ray-casting algorithm intended for visualising complex, large-scale particle-based simulations. Splotch is optimised in terms of memory usage and can exploit modern HPC architectures through the use of OpenMP, MPI, CUDA (Rivi et al. 2014), GPI, and recently Xeon Phi (Dykes et al. 2016). Splotch maintains a high level of performance with good scalability on parallel architectures with no dependencies for external libraries except when needed for parallelism or for handling external file formats. Multiple file formats and simulation codes are supported such as pure binaries and HDF5 and GADGET, RAMSES and ENZO respectively. To produce an image, particles are first rasterised by applying geometric transformations and colour assignments. Rendering is realised via rays cast through lines of sight where contributions of each individual particle are calculated by solving a radiative transfer equation.

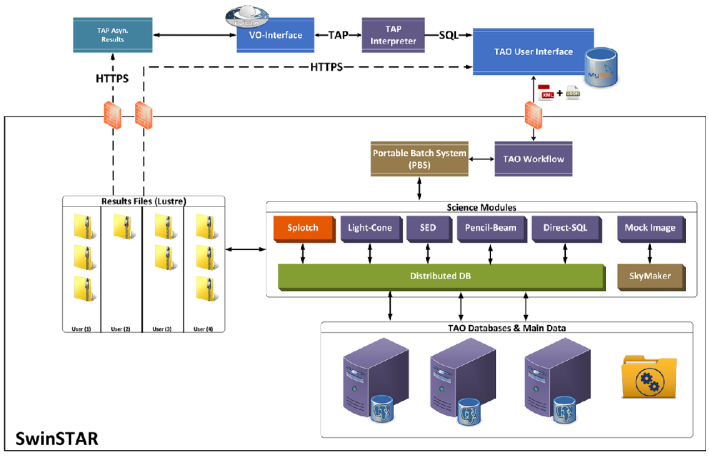

Figure 1. Schematic diagram of the TAO software modules.

The TAO architecture is shown in figure 1 (for details please see Bernyk et al. (2016)). TAO is made up of a number of interlinked software components to provide end users with seamless access to the underlying HPC infrastructure and big-data

services via web-interfaces. We have created a Splotch science module within TAO, providing it with direct access to the TAO HPC infrastructure and the various data products that can be generated via TAO. For this, we used the Swinburne HPC platform (gSTAR), which is a cluster of fifty nodes, each node containing two 6-core X5650 CPUs, 48GB RAM and two Tesla C2070 GPUs. The process of running Splotch on this platform is directly controlled by the standard TAO workflow and the relevant PBS modules. The TAO workflow module turns user interactions with the user interface (UI) components into suitable parameters for the Splotch science module, and having started the execution (to be scheduled and managed by the PBS module), it monitors the execution progress until the final visualisation results are generated and communicated back to end users.

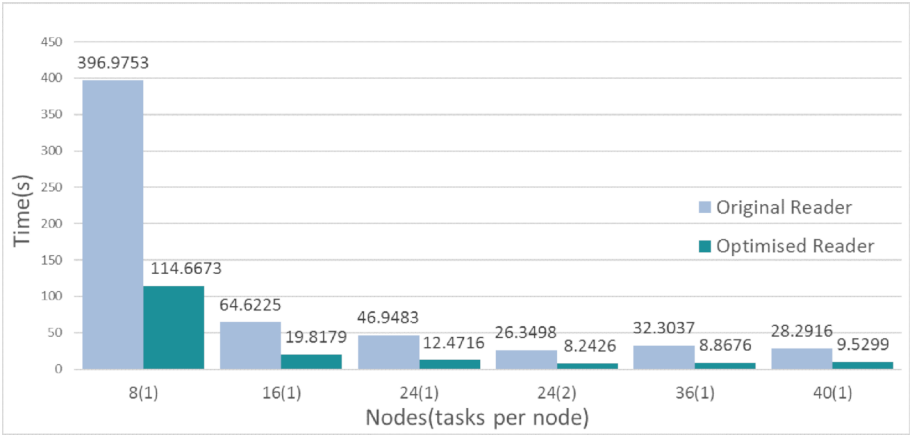

Figure 2. Total walltime results of using Splotch to produce an output with the modified HDF5 reader, across various configurations.

To prepare Splotch for this integration, we have made modifications to facilitate its interface with the TAO workflow, as well as to support TAO data products. In particular, adjustments were made to the HDF5 reader to support the compound datatype - a structured datatype containing one or more member datasets. This reader was also optimised by reducing memory use and avoiding unnecessary data reads, through modifying the underlying buffer structure to only contain and read the fields which have been specified by end users. These optimisations resulted in significant speed gains when tested on a snapshot (containing 16,042,057 particles) out of a 181GB dataset. Figure 2 displays the results of these tests and demonstrates that Splotch has shown good scalability when taking advantage of the underlying HPC architecture of TAO - through the use of parallelism via MPI, as well as exploiting GPUs via CUDA.

Further to this, we have modified Splotch to have the option to provide calculated values for specific visualisation parameters used in its pipeline within TAO. If specified, Splotch will calculate a suitable camera position encapsulating a bounding box of the particles within view of the camera parameters. This provides a mechanism for an initial rendering result, giving the user a good overview of the dataset under inspection. However, a camera position can still be manually specified if required. The smoothing length (which is used during the rendering stage of the Splotch pipeline) can also be specified for automated calculation of a good starting value which can then be further tweaked by a user as required.

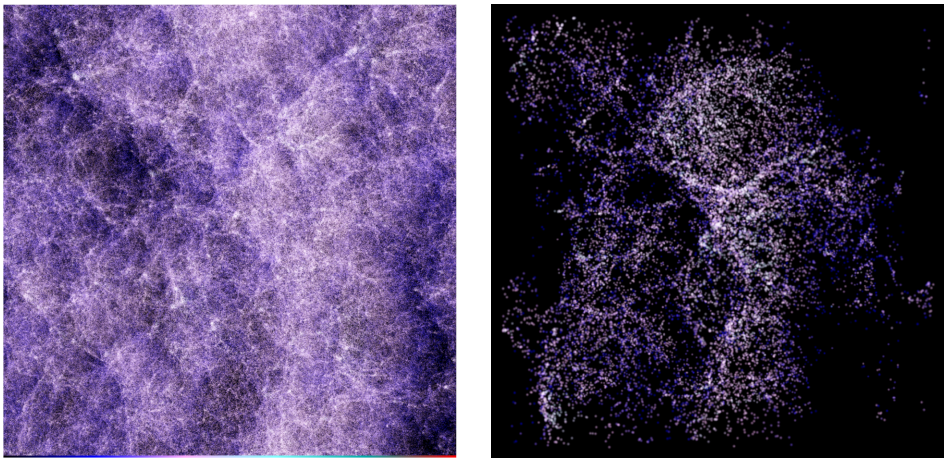

Figure 3. Two Splotch rendered images of HDF5 compound datasets, on the gSTAR HPC system.

3. Further Work and Conclusions

We have presented a first attempt of integrating the HPC visualisation code Splotch within TAO. We have shown how Splotch has been adapted and modified so as to be ready to be fully integrated into TAO's web interface, as well as showing how the code has been optimised to fully exploit the underlying HPC architecture. This work will serve as the basis for how TAO interacts with Splotch through future development of an easy-to-use web interface based on Django[1], a high-level Python web framework, which will allow non-technical TAO users to access state-of-the-art simulations. Work is underway to also offer the complete functionality of Splotch on a HPC system from within a browser.

Acknowledgments. This work was carried out as part of the final year dissertation of Elliott Ayling by using the swinSTAR supercomputer at Swinburne University of Technology.

References

Bernyk, M., et al. 2016, The Astrophysical Journal Supplement Series, 223, 9
Dolag, K., et al. 2008, New Journal of Physics, 10
Dykes, T., et al. 2016, International Journal of High Performance Computing Applications
Rivi, M., et al. 2014, Astronomy and Computing, 5, 9

[1] `https://www.djangoproject.com`