Extended Abstract: The Renaissance Simulations Laboratory

Kacper Kowalik,* NCSA at UIUC; Michael Norman, SDSC at UCSD; Britton Smith, SDSC at UCSD; Matthew Turk, iSchool at UIUC
*Corresponding author address: kowalikk@illinois.edu

Abstract: *The Renaissance Simulations are a set of astrophysical calculations run on the Blue Waters supercomputer, designed to study the way that the Universe was reionized. The simulations are cutting edge, featuring radiation transport, chemical processes, hydrodynamics, gravity and star formation prescriptions; they have been run on the Blue Waters supercomputer using the ENZO simulation platform. We present the Renaissance Simulations Laboratory (RSL), a web-based mechanism of accessing, analyzing, visualizing and then studying the outputs from the Renaissance Simulations, taking a hybrid approach of utilizing both general-purpose systems such as the Jupyter Notebook and bespoke, constructed widgets such as SQL queries around halo catalogs.*

1. Introduction

The increasing size and information richness of modern astrophysical simulations act in concert to ensure that the biggest bottleneck to comprehensive analysis of these simulations is a lack of human time. Oftentimes, the simulations are too big to move between institutions, so even if individuals choose to share data, this is done without a plan for making that data actionable.

The Renaissance Simulations [1] are a suite of calculations studying a set of regions in the early universe. The initial simulation covered a region of the universe 28.4 Mpc/h on a side using the WMAP7 best-fit cosmology and had a 512^3 root grid resolution and three levels of static nested grids. Initial conditions were generated at z = 99 using MUSIC [2], and a low-resolution simulation was run to z = 6 to find regions suitable for re-simulation using ENZO [3]. The volume was then smoothed on a physical scale of 5 comoving Mpc, and regions of high, average and low mean density were chosen for re-simulation. These subvolumes (see Figure.1), designated the *"Rare peak," "Normal,"* and *"Void"* regions, with comoving volumes of 133.6, 220.5, and 220.5 Mpc^3, were re-simulated with an effective initial resolution of 4096^3 grid cells and particles in the region of interest, giving a dark matter mass resolution of 2.9 × 104 $M_\odot$.

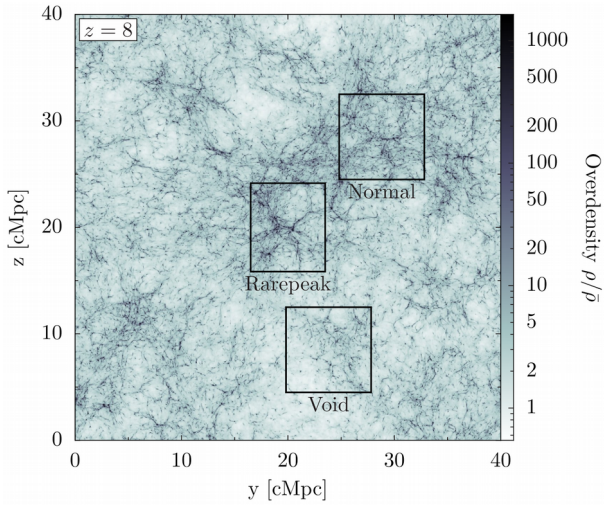


Figure 1. Projection through the computational domain of the initial Renaissance Simulation showing color coded overdensity of the dark matter. The subvolumes of the *Rare peak*, *Normal*, and *Void* simulations are marked with black rectangles.

2. Implementation

We have decided to address this issue -- of shared, but not usefully shared data -- by developing and making available an intuitive interface to the data that allows for parallel analysis, detailed studies of the full, unprocessed data, and both general purpose and specially-

Presented at Gateways 2016, San Diego Supercomputer Center, La Jolla, CA, November 2-3, 2016.
https://gateways2016.figshare.com/

constructed interfaces to the data.

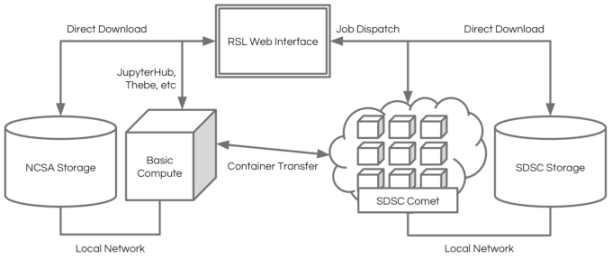


Figure 2. Architectural diagram of the RSL, including initial flow between cyberinfrastructure components spanning over NCSA and SDSC computational resources.

Presenting only a general-purpose interface, in this case the Jupyter Notebook [4] with an installation of yt (http://yt-project.org/ ; [5]) is both beneficial in the flexibility it provides and a detraction, as it restricts our audience to the simulation community. The most successful astrophysical simulation data sharing interface, the Millennium Database, presents a unique and tuned user interface that exposes SQL queries. This restricts the ability of individuals to conduct freeform analysis, but it also enables a lower barrier to entry that opens it up to a significantly larger audience. By presenting both the freeform analysis available in the Jupyter Notebook -- where individuals can bring their own tools and software as well as use the pre-installed yt package -- and the restricted interface, the RSL will enable both domain experts, such as those who ran the simulations initially, as well as other non-specialists from other subdomains of astrophysics to mine these simulations for information and conduct specialized inquiry.

In contrast to many -- but certainly not all -- Science Gateways, we have taken the approach with the RSL of exposing user interfaces that are familiar to users that would otherwise analyze data independent of a gateway as well as those that are tuned to the specific simulation, with a restricted set of functionality. Jupyter Notebooks, running inside Docker containers, are launched next to the full dataset on resources provided by SDSC and NCSA through the National Data Service. The next phase of the RSL will also include launching these jobs on SDSC Comet and other XSEDE resources, where the computations can be executed in parallel, allowing researchers to scale up their computations to take advantage of the full size and scope of the Renaissance Simulation datasets. A diagram, indicating our initial deployment strategy is included in Figure 2.

3. Conclusions

The Science Gateways Institute provides an enormous opportunity to collaborate on developing, deploying, and enhancing the RSL. In particular, we look forward to learning from the other projects, particularly with respect to cyberinfrastructure and usability, and utilizing CI from the overall Science Gateways Initiative. As RSL utilizes other projects supported by the NSF SI2 ecosystem (such as yt) we believe this to be an ideal collaboration.

4. References

[1] J. H. Wise, M. J. Turk, M. L. Norman, and T. Abel. The Birth of a Galaxy: Primordial Metal 17 Enrichment and Stellar Populations. ApJ, 745:50, January 2012.

[2] Hahn, O., & Abel, T. 2011, Monthly Notices of the Royal Astronomical Society, 415, 2101

[3] Bryan, G.L., Norman, M.L., O'Shea, B.W., et al. 2014, Astrophysical Journal Supplements Series, 211, 19

[4] Kluyver, Thomas, et al. "Jupyter Notebooks—a publishing format for reproducible computational workflows." Positioning and Power in Academic Publishing: Players, Agents and Agendas: Proceedings of the 20th International Conference on Electronic Publishing. IOS Press, 2016.

[5] Turk, Matthew J., et al. "yt: A multi-code analysis toolkit for astrophysical simulation data." The Astrophysical Journal Supplement Series 192.1 (2010): 9.