Grid computing has emerged as an important new field focusing on large-scale resource sharing and high-performance orientation. The astronomy community has an abundance of imaging datasets at its disposal which are essentially the “crown jewels” for the astronomy community. However, these astronomy datasets are generally terabytes in size and contain hundreds of millions of objects separated into millions of files—factors that make many analyses impractical to perform on small computers. The key question we answer is: “How can we leverage Grid resources to make the analysis of large astronomy datasets a reality for the astronomy community?” Our answer is “AstroPortal,” a gateway to grid resources tailored for the astronomy community. To address this question, we have developed a collection of Web Services-based systems that use grid computing to federate large computing and storage resources for dynamic analysis of large datasets. Building on the Globus Toolkit 4, we have built a prototype consisting of various systems (AstroPortal, DYRE – Dynamic Resource pool Engine, 3DcacheGrid – Dynamic Distributed Data cache for Grid applications, and CompuStore – Computational Scheduler) to enable a “stacking” analysis; the analysis sums multiple regions of the sky, a function that can help both identify variable sources and detect faint objects. We have deployed AstroPortal and the related systems on the TeraGrid distributed infrastructure and applied the stacking function to the Sloan Digital Sky Survey (SDSS), DR5, which comprises more than 320 million objects dispersed over 1.5 million files, a total of 3.5 terabytes of compressed data, with promising results. AstroPortal gives the astronomy community a new tool to advance their research and to open new doors to opportunities never before possible on such a large scale.
The “AstroPortal” is a gateway to grid resources, around the AstroPortal, we have developed a collection of Web Services-based systems that use grid computing to facilitate large computing and storage resources for dynamic analysis of large datasets. Building on the Globus Toolkit 4, we have built a prototype consisting of various systems (AstroPortal, DYRE – Dynamic Resource pool Engine, 3DeacheGrid – Dynamic Distributed Data cache for Grid applications, and ComputStore – Computational Scheduler) to enable a "stacking" analysis; the analysis sums multiple regions of the sky, a function that can help both identify variable sources and detect faint objects.

We propose three different systems that can interoperate with each other in order to offer a complete storage and resource management solution to enable the efficient dynamic analysis of large datasets.

- **DYRE**, Dynamic Resource pool Engine, is an AstroPortal specific implementation of dynamic resource provisioning. DYRE essentially handles all the necessary tasks associated with state monitoring, resource allocation based on observed state, resource deallocation based on observed state, and exposing relevant information to other systems.

- **3DeacheGrid**, Dynamic Distributed Data cache for Grid application, allows applications to achieve a good separation of concerns between their business logic and the complicated data management task of large data sets. We utilize a tiered hierarchy of storage resources to cache data in faster but smaller data storage tiers.

- **ComputStore**, a work scheduler that uses both the storage and compute resource management systems 3DeacheGrid and DYRE; the goal of ComputStore is to make the best scheduling decisions given some work, the available compute resources (DYRE), and the available data caches (3DeacheGrid) which are stored on the compute resources.

**Performance Results**

---

For more information, please visit [http://people.cs.uchicago.edu/~iraiciu/research/](http://people.cs.uchicago.edu/~iraiciu/research/)

This work was funded by the NSF TeraGrid grant from June 2005 - September 2006; the work is currently being funded by the NASA ABC GSRT from October 2006 - September 2007.