LEAD at Unidata

Status Update, October 2009

Mohan Ramamurthy

The official period of performance for the LEAD project ended on 30 September 2009 after a one-year the no-cost extension period. The six year ITR project had pioneered a new approach for integrating complex weather data, assimilation, modeling, mining, and cyberinfrastructure systems in innovative ways to empower researchers and students with capabilities heretofore available at only a few major universities and research or operational centers around the world. Containing virtually all elements of modern cyberinfrastructure – from adaptive sensors and high-performance Grid computing and networking to huge data sets, human decision making and complex virtual organizations – LEAD functionality also was integrated with the TeraGrid as one the most successful TeraGrid Science Gateway projects and the project continues to serve as an *avant-garde* research system for the meteorological and computer science communities. LEAD became a principal application driver for helping TeraGrid identify and solve some of its most important challenges and prepare for the next generation XD environment.

Even though the project has officially ended, the technologies developed LEAD, including the LEAD portal and system, continue to exist and will be available for users as long as feasible. The UPC continues provide data, software and support for the project and valuable assistance in the testing and deployment of and end-user support for LEAD systems in the atmospheric sciences community. Recently, with the UPC made a major upgrade tothe Unidata LEAD test bed since it continues to be a primary resource of data for LEAD workflows and the LEAD system at the University of Indiana..

Workshop on Cloud Computing and Collaborative Technologies in the Geosciences

The LEAD project, in partnership with Indiana University Pervasive Technologies Institute Data to Insight Center and with funding from NSF, organized a workshop on Cloud Computing and Collaborative Technologies in the Geosciences on 17-18 September in Indianapolis. Cloud computing promises unlimited resources for carrying out research or conducting operational computing tasks. But how well suited is it to advancing geosciences research and operational activities? How does one take advantage of it? Collaborative sharing and social networking technologies, such as Facebook and Twitter, are a second and in some ways complementary technology that promises better science outcomes. In what way can collaborative technologies contribute to scientific or operational advancements in the geosciences?

The goals of the workshop were to:

1. Identify the key benefits and risks to adopting collaborative and cloud computing technologies in operational and research settings.
2. Present current best practices to making effective utilization of these resources.

3. Identify the computer science research challenges that must be overcome to advance cloud computing and collaborative technologies in the geosciences.

The workshop attracted considerable interest from a wide range of researchers, educators and practioners in computer science, informatics, and geosciences, as well as officials from NSF, NOAA and NASA. In all, 56 participants attended the workshop and they engaged in a lively discussion on the workshop topics.

Upgrade of the LEAD Testbed at the UPC

The Unidata LEAD Testbed was recently upgraded, using funds remaining in the project during the no-cost extension period, to meet the ongoing and future needs of the project. Upgrades of the LEAD systems will improve performance, reliability and efficiency. Jobs previously distributed across several systems are now consolidated on new, powerful Intel Xeon E5430 servers with 64 GB RAM for caching and the latest Fedora 11 operating system. These servers now directly access their storage over a direct connect SAS bus as opposed to the previous SATA/NFS configuration. The storage itself has been upgraded to faster, higher-capacity drives for a total of 72 TB of storage and all data products now reside on a mirrored configuration for better redundancy and uptime. We continue to store data products from anywhere from 3 weeks to a year depending on demand.

	Dataset	# Days
LEAD	NAM 40 km CONUS	300
	ADAS 10 km CONUS	45
	NEXRAD Level II	183
	Surface Observations	600
	Profiler Data	600
	Upper Air Observations	600
	UPC Steered Regional	130
Broader Community	All NAM datasets	300
	DGEX	220
	RUC	220
	NEXRAD Level III	200
	HRS	160
	Satellite	160
	ECMWF	130
	CMC	130
	UKMET	130
	GFS 0.5 Deg	120
	RUC2	105

Figure. Data now available, both historical (on disk) and in real time, within the LEAD and broader community data catalogs.

As stated in previous status reports, the Unidata LEAD test bed continues to be a primary resource of data for LEAD workflows. In addition providing real-time access to data from a number of different datastreams, it also hosts a rolling archive of data for varying durations for case-study or event-oriented research. The LEAD testbed archive (above) is not only used by the LEAD project but it has also been made available to the broader Unidata community for the past five years.

LEAD Beyond the ITR Phase

Even the ITR LEAD project has concluded, the LEAD-developed technologies and the infrastructure will continue to operate with other funding, at least for the foreseeable future. The PIs continue to strategize on how best to maintain and operate the LEAD system and continue engaging users in meteorology and related disciplines. Unidata will continue to provide real-time and case study data sets for LEAD workflows, along with end user tools like the IDV.