Linked Environments for Atmospheric Discovery (LEAD):
Transforming the sensing and numerical prediction of high impact local weather through dynamic adaptation

Mohan Ramamurthy
Unidata

Plus

Many LEAD PIs, Students, and Developers

CISL Seminar, 2 March 2007
At least 20 killed as twisters slam several states
Eight teens die inside Ala. school; tornado also tears through Ga. hospital

ENTERPRISE, Ala. - A violent storm system that ripped apart an Alabama high school as students hunkered inside later tore through Georgia, hitting a hospital and raising the death toll to at least 20 across the Midwest and Southeast.

Eight students were killed when a tornado struck Enterprise High School, blowing out the
Mesoscale Weather Impact

- Each year, mesoscale weather – floods, tornadoes, hail, strong winds, lightning, and winter storms – causes hundreds of deaths, routinely disrupts transportation and commerce, and results in annual economic losses > $13B.
- According to one estimate, a quarter of the U. S. GDP is weather and climate sensitive.
What Today’s Weather Observation and Prediction Infrastructure Does…

Virtually Nothing!!!
Traditional NWP Methodology

**Static Observations**
- Radar Data
- Mobile Mesonets
- Surface Observations
- Upper-Air Balloons
- Commercial Aircraft
- Geostationary and Polar Orbiting Satellites
- Wind Profilers
- GPS Satellites

**Analysis/Assimilation**
- Quality Control
- Retrieval of Unobserved Quantities
- Creation of Gridded Fields

**Prediction/Detection**
- PCs to Teraflop Systems

**Product Generation, Display, Dissemination**

The Process is Entirely Prescheduled and Serial; It Does NOT Respond to the Weather!
State-of-the-Art

- Sophisticated mesoscale numerical models and other tools abound (ARPS, MM5, WRF, COAMPS)
- All components are complex even if used individually
- Process control infrastructures are not widely available and are unwieldy (e.g., 50K line Perl scripts for ARPS)
- The NET RESULT
  - **Huge learning curve**, especially for students (70-30 rule)
  - Limited sophistication of experiments
  - No easily portable
  - Cannot run in grid environments
  - Disincentive for use – requires substantial human capital and physical infrastructure
  - Do not assimilate observations, especially local ones!!
Motivation #1 for LEAD

• Weather technologies are very hard to use in sophisticated ways because they’re so complicated and linked together using cumbersome scripts - creating a HUGE DIVIDE BETWEEN THE HAVES AND THE HAVE NOTS
Example: Run a Fine-Grid Forecast Experiment Over a Large Area Using Real Doppler Radar Data

It took an MS student a few months, working with a research scientist, to learn how to run this simple case.

This case is VERY TYPICAL!
Motivation #2 for LEAD

• High impact local weather is **VERY DYNAMIC** while our meteorological tools, cyber environments and learning modalities are almost entirely **STATIC**
Two Overarching Goals

• **Goal #1**: Lowering the barrier for using complex end-to-end weather prediction systems
  - **Democratize** the availability of advanced weather technologies for research and education
  - **Empower** application in a grid context
  - **Facilitate** rapid understanding, experiment design and execution

• **Goal #2**: Dynamic Adaptivity
  - Models and hazardous weather detection systems **responding to** observations and their own output
  - Models and hazardous weather detection systems **driving the collection of** observations
  - **IT infrastructures** providing on-demand, fault tolerant services
Dynamic Model Response

STATIC OBSERVATIONS
- Radar Data
- Mobile Mesonets
- Surface Observations
- Upper-Air Balloons
- Commercial Aircraft
- Geostationary and Polar Orbiting Satellite
- Wind Profilers
- GPS Satellites

Analysis/Assimilation
- Quality Control
- Retrieval of Unobserved Quantities
- Creation of Gridded Fields

Prediction/Detection

Prediction/Detection

Product Generation, Display, Dissemination

Models Responding to Observations And Their Own Output

End Users
- NWS
- Private Companies
- Students
Observations Collected as Models and Detection Systems Dictate

DYNAMIC OBSERVATIONS

Analysis/Assimilation
- Quality Control
- Retrieval of Unobserved Quantities
- Creation of Gridded Fields

Prediction/Detection
- PCs to Teraflop Systems

Product Generation, Display, Dissemination

Models and Algorithms Driving Sensors

Linked Environments for Atmospheric Discovery

End Users
- NWS
- Private Companies
- Students
The LEAD Team

Linked Environments for Atmospheric Discovery
LEAD Grid & TeraGrid

- Variety of resources and configurations
- Capabilities differ per site; common core requirements for Linux and Globus

Software Stack: Mostly NMI
- Includes Globus, MyProxy, OGCE, NWS and many other middleware components
Unidata LEAD Testbed

Portal Servers for Web, OpenDAP, THREDDS, OSCAR, Grid and LDM Services

~30 GFLOP Processing Cluster
Beowulf Cluster with Eight Nodes each with Two Athlon 2400+ CPUs Connected by a Gigabit Fibre Network

48 TB Storage Cluster
In LEAD, Most Everything is a (Web) Service
[But in a Grid Framework]

Service A (ADAS)
Service B (WRF)
Service C (NEXRAD Stream)
Service D (MyLEAD)
Service E (Cataloging)
Service F (Visualization)
Service G (Monitoring)
Service H (Scheduling)
Service I (ESML)
Service J (Repository)
Service K (Ontology)
Service L (Decoder)

Many others…
Start by Building Simple Prototypes to Establish the Services/Other Capabilities…

Prototype X

Service C (NEXRAD Stream)
Service F (IDV)
Service L (Decoder)
Service E (VO Catalog)
Service D (MyLEAD)
...and then Solve General Problems by Linking them Together in Workflows

These web services can be used as stand-alone capabilities, independent of the LEAD infrastructure.

Event-driven workflows will be dynamic – i.e., can change with time automatically.
Why A Service-Oriented Architecture?

- Flexible and malleable
- Platform independent (emphasis on protocols, not platforms)
- Invokable over a network
- Loose integration via modularity
- Evolvable and re-usable (e.g. Java)
- Interoperable by use of standards → robustness
What’s a service anyway?

- A “web server” that runs an application for you.
  - You send it requests (XML documents) and it processes the information and send replies (notifications) when it is done.

![Diagram showing the process of a service request, running WRF, and publishing notifications.]
When it receives two messages (source and destination URL) it decodes the file and invokes GridFTP to move it.
  - Note: each message also identifies the user and experiment name.

When the move is complete it sends a message indicating that it is done and where to find the file.

It also sends a “notification event” with the same information.
Linked Environments for Atmospheric Discovery (LEAD)

NCSA, IU, TACC

User Orchestrates Web Services to Create Regional Forecast

Data Service

Prediction Service

Product Generation & Data Mining Service

User running local analysis and display tools

Decoder Service

Orchestration Service

User running local analysis and display tools

Unidata

Unidata

IU, UAH

OU

OU
LEAD Portal

LEAD is a TeraGrid Science Gateway Project
For now, one community account for all users

Portal Framework: GridSphere;
Portlet API: JSR 168 compliant for services

portal.leadproject.org
Graphical Data Search

LEAD Data Search

Data Products:
- All
- METAR
- Rawinsondes
- NEXRAD Level II
- NEXRAD Level III
- GOES
- NCEP etc.
- ADAS
- WRF

Data Categories:
- Any
- Temperature
- Wind
- Pressure
- Moisture
- Clouds
- Precipitation
- Terrain
- Land Cover
- Soil
- Surface Water
- Sub-Surface Water
- Electrification

Spatial Range

Temporal Range
- Any temporal range
- Begin Time: 2006-10-10T00:00
- End Time: 2006-10-12T24:00

Search Area
Drag balloons to adjust the search area.

Found 1 match(es) ...

Title: Weather and Research Forecast Model/Steered WRF
Publisher: UCAR/UNidata
Dataformat: NetCDF
Spatial: west: -152.856, east: -48.354989998599
north: 57.28, south: 12.18
Temporal: Begin 20060827 00000000, End 20060828 12000000
Abstract: WRF model data experiment at UPC Steered by precipitation algorithm implemented in GEMPAK and described at:
http://www.unidata.ucar.edu/software/gempak/ftmodel/ NOTE: as described on the page, the algorithm is also used to steer the WRF NMM model being run at UPC. It had also been used to steer WS-ETA. The URLs shown have a form:
http://lead.unidata.ucar.edu/cgi-bin/mem-dods/data/pub/other/case_studies/WRF/PCSteered/wrfout_d01_{date}.nc where [date] has a form: YYYY-MM-DD-HH_MM_SS and indicates the start time of the regional model run.

Click a data product to visualize it.
http://lead4.unidata.ucar.edu:8080/thredds/dodsC/model/UCAR/UNIDATA/WRF/STEERED/wrfout_d01_2006-08-27_00_00_00.nc
http://lead4.unidata.ucar.edu:8080/thredds/fileServer/model/UCAR/UNIDATA/WRF/STEERED/wrfout_d01_2006-08-27_00_00_00.nc
Domain Selection & Experiment Builder

Graphical Domain Selection

Drop down menu for selection of runs
Automatic Workflow Generation

Pre-configured workflows automatically generated, which can be modified by users via a composition tool.

Workflow: GPEL
Based on BPEL4WS

Linked Environments for Atmospheric Discovery
Basic Gateway SOA

Grid Portal Server

Gateway Services
- Proxy Certificate Server / vault
- Application Events
- Resource Broker
- App. Resource catalogs
- Workflow engine
- User Metadata Catalog MyLEAD

Core Grid Services
- Security Services
- Information Services
- Self Management
- Resource Management
- Execution Management
- Data Services

Physical Resource Layer
OGSA-like Layer

Linked Environments for Atmospheric Discovery
LEAD Metadata Schema

-- Federal Geographic Data Committee (FGDC) compliant
-- LEAD profile extends FGDC schema
-- metadata subsystem crosswalks from THREDDS catalog into LEAD Resource Catalog and myLEAD Catalog
Other Supporting Services

- Seige client and Ensemble Broker engine
  - Supports high level orchestration across grid platforms
  - State persisted in database
  - Execution service
    - Front ends GRAM (now), MOAB (in progress)
    - Monitors running jobs through callback

Queries for NAM data, from THREDDS Data Service

Obtains domain center from Integrated Data Viewer

(developed for Unidata Workshop, July 13 2006)
Siege-Unidata IDV Integration

Handed correct NAM data set as selected in Siege

Use Field selector in IDV to create display
LEAD Deployment

- LEAD community deployment began at the Unidata Users Workshop, held July 10-14, 2006, in Boulder, CO
  - Theme: Expanding the Use of Models as Educational Tools in the Atmospheric & Related Sciences
- Participants used the LEAD system to make on-demand weather prediction, in their chosen region of interest, using the Weather Research and Prediction model.
- The system used LEAD orchestration tools, real-time data, and leveraged TeraGrid resources for computations.
LEAD at the Unidata Workshop

- Workshop User Experience
  - Exposure to Education Thrust developments and capabilities
  - 2 sets of 25 users selected the modeling region and submitted 2 WRF model jobs each
  - Browse/Search/Visualization of input North American Mesoscale (NAM) model and Assimilated (ADAS) near real-time data
  - Browse of experiment metadata and Visualization of model results
  - Visual comparison of model results with RADAR data
LEAD at the Unidata Workshop

• Workshop participant’s responses
  – “Sharing data/models, democratization of data and models were the highlight of the workshop.”
  – “I liked that it was a step by step process. The steps were well broken down into easily understood components.”
  – “Easy to use.”
  – I liked the “ease of running WRF and fact that it might use someone else’s computer to do the computing”
  – “I will implement it in my introductory level Met. and Climate class and workshops for 6-12 teachers”

• LEAD Deployment Going Forward
  – LEAD Beta Users Program underway
Supercell / tornado study

- **Where we are now**
  - The Siege/ensemble broker system has been used to study this problem
  - 105 simulations with WRF were run, pre- and post-processed through Siege
  - 2 TB of output
  - Run on tungsten (Teragrid Xeon cluster)
  - Only variable: initial storm placement

- **Credit:** Brian Jewett (next 2 slides)

- **Presented at 23rd Conference on Severe Local Storms, St. Louis, November 2006**
  - P11.4 "Numerical modeling of cell interaction"
Linked Environments for Atmospheric Discovery

120 minutes
THREDDS Data Repository

- THREDDS Data Repository (TDR) will
  - Find and manage storage space
    - Implement space management policies
    - May involve various storage devices
    - Could include replication
  - Move the data into the repository, return a unique ID
    - http, gridftp, scp
  - Incorporate existing metadata
  - Generate additional metadata where possible
  - Integrate with a THREDDS Data Server
Way Forward – A Scenario

• The LEAD PIs are discussing the next phase of LEAD – life after the ITR phase
• Proviso: LEAD goals are realized and the community and sponsors are willing to support it.
• In one scenario, LEAD will have a home in UCAR/Unidata and be supported as a community facility
• There are potential ties to CISL – an ideal project to link the Unidata community with CISL
• LEAD can possibly link to the NCAR TeraGrid node at the new Data Center
• The democratization envisioned in LEAD presents a unique opportunity for realizing the promise of the TeraGrid, a $250-500 million investment by NSF
• LEAD also fits nicely with the current plans for the Petascale Collaboratory for the Geosciences effort – LEAD and Unidata are already collaboratories
• CISL and Unidata’s unique relationship as sister programs in UCAR presents unique partnership opportunities to benefit the community.
Future

• Bridging Communities with range of computing capabilities and needs should be a long term goal.

• Many technological, cultural and systemic challenges remain in connecting PI, departmental and campus computing systems with high-end Grid environments like the TeraGrid.

• Specifically, we need to build bridges between the traditional Unidata community and High-end Computing programs like CISL and Petascale Collaboratory for the Geosciences.
Contact Information

• LEAD PI: Prof. Kelvin Droegemeier, kkd@ou.edu

• My email: mohan@ucar.edu

http://portal.leadproject.org/