Example Gateway: LEAD – Linked Environment for Atmospheric Discovery

Suresh Marru
TeraGrid Grid Infrastructure Group
Indiana University
Session Overview

• Introduction to LEAD Science Gateway.
• Demo/Hands-on with LEAD Portal
• GridChem Advanced Support
• OGCE Hands-on (Afternoon Session)
  – Construct, Execute Workflows
  – Audience Interactive Session: Register Applications, Install services, Compose Workflows...
Acknowledgements

• LEAD Slides & Discussion courtesy of Dennis Gannon, Beth Plale & the LEAD Team.

• GridChem Slides & Discussion courtesy of Sudhakar Pamidighantam, Rion Dooley, Vikram Gazula & the GridChem Team.
Linked Environments for Atmospheric Discovery (LEAD)

- LEAD through an integrated framework empowers meteorology community (Virtual Organization) to mine observational and model weather data and execute linear and ensembles of customized meteorology workflows while capturing provenance.
The LEAD Vision

Revolutionize the ability of scientists, students, and operational practitioners to observe, analyze, predict, understand, and respond to intense local weather by interacting with it dynamically and adaptively in real time.
Example: “Optimal” Weather Prediction Using Dynamic Adaptivity

- Streaming Observations
- Data Mining
- Storms Forming
- Forecast Model

On-Demand Grid Computing

Instrument Steering
Refine forecast grid
WELCOME TO THE LEAD PORTAL

Linked Environments for Atmospheric Discovery (LEAD) makes meteorological data, forecast models, and analysis and visualization tools available to anyone who wants to interactively explore the weather as it evolves. The LEAD Portal brings together all the necessary resources in a convenient access point... read more

FEATURES FOR ANYONE INTERESTED IN THE WEATHER

- Research & Education
- Educate Students in high school, or middle schools
- Visitors "Newcomers" or the curious

POPULAR TOOLS

- Visualize Weather Data
- Integrated Data Viewer
- Access a Forecast or Analysis
- Geographic Region Search

QUICK LINKS

- Live Weather
- LEAD Grid
- LEAD Blog
- Glossary

THE LEAD TEAM

- Colorado State University
- Howard University
- Indiana University
- Millersville University
- UAH
- UCAF
- Illinois
- Oregon State University
- Utah

Analyze & Predict

Research & Reproducibility

Discover & Visualize

Education & Outreach
Students can access a suite of learning materials, particularly LEAD-to-LEARN modules, case studies that provide guided inquiry into complex atmospheric phenomena with the options to explore these conditions using distributed resources.

Search 7 data sets

Store, retrieve, and share experiment results/model output

Build, launch, and monitor experiment
The Realization in Software

Workflow graph

Application services

Compute Engine

User Portal

Portal server

Workflow Engine

Fault Tolerance & scheduler

Event Notification Bus

Data Catalog service

MyLEAD Agent service

MyLEAD User Metadata catalog

Providence Collection service

Data Management Service

Data Storage

TERAGRID Exotic Data Scale Facility

UIUC PSC ORNL NCSA IU TACC

Data Storage
LEAD Portal Demo
Application Services

• Workflows are built by composing web services
  – Fortran applications are “wrapped” by a Application Factory which generates a web service for the app.
    • Instances of the service are dynamically created using Globus
    • Registers WSDL for the service with a registry
  – Each service generates a stream of notifications that log the service actions back to the XMC Cat Metadata Catalog.
Workflow Composition, Execution & Monitoring

Xbaya enables users to construct, share, execute and monitor sequence of tasks executing on their local workstations to high-end grid-enabled compute resources.
Service Monitoring via Events

- The service output is a stream of events
  - I am running your request
  - I have started to move your input files.
  - I have all the files
  - I am running your application.
  - The application is finished
  - I am moving the output to you file space
  - I am done.

- These are automatically generated by the service using a distributed event system (WS-Eventing / WS-Notification)
  - Topic based pub-sub system with a well known “channel”.

![Diagram showing event monitoring process with service instance, notification channel, and pub-sub system]
“A significant need exists in many disciplines for long-term, distributed, and stable data and metadata repositories”
- NSF Blue-Ribbon Advisory Panel on Cyberinfrastructure

“Metadata is key to being able to share results”
- UK e-Science Core Programme Study

Slide Source: Scott Jensen, Indiana University
XMC Cat exploits characteristics of scientific metadata schemas for a generic solution that:

- Adapts to varied schemas
- Exhibits fast XML responses
- Enhances scalability
- Search GUI adapts to schema

Metadata "Shredded" to Relational Tables

Complex Search

Build Response From Concepts

Concepts Stored as XML

Query Result Based on Community XML Schema

Slide Source: Scott Jensen, Indiana University
Automated Metadata Gathering

After the data is registered with metadata catalog, extraction shims opens files and extracts metadata and augments already populated metadata

Example: Detailed view of procedures within the WRF output post-processing shim

Slide Source: Yiming Sun, Indiana University
Hasthi
A Scalable, Distributed, Robust, Recouping Management Framework

Meta-model is a model of the system created inside managers/coordinate (used for decisions/distributed across managers)

User Defined Rules (e.g.)

```java
rule "CreateAlternativeForRegistry"
when
not exists( ManagedService( state == "UpState", type == "Registry"));
then
system.execute( new CreateServiceAction("Registry"));
end
```

Managable Resources
Can manage resources that supports WSDM specification. Resources should preserve required amount of state across failures/changes

Salient Features:

- User Defined Rules which enforce global assertions about the system.
- Highly scalable – tested to manage 80k to 100k resources.
- System can auto-recoup from failures of the Resources as well as the management components – coordinator, manager.

More Info: Srinath Perera hperera@cs.indiana.edu
Architecture of Hasthi

Managers, and bootstrap nodes are placed in a P2P network, and a Coordinator is elected among managers using P2P broadcast.

Coordinator provides global control

Summerized Meta-model is a model of the managed system created inside the coordinator.

System

Decision Unit

Coordinator

Manager Control

Global Actions

Resource Layer

Partial Meta-model is a model of assigned resources created in each Manager (used for decisions)

Partial meta-model

Manager

Managers control resources assigned to them by the Coordinator

Resource Monitoring Data

Local Management Actions

Join Messages

Manageble Resources resources that supports WSDM Specification.

Administrators

Management GUI

Join Messages

Forward Join

Bootstrap nodes
Lessons Learned From LEAD

• Large number of users create a surges in load exposing previously unknown problems.
• Troubleshooting large scale distributed infrastructure needs coordinated debugging involving multiple experts.
• Scientists need Flexibility but too much Flexibility is called Confusion.
Intelligent Workflow automation

- Most Grid workflows transform one data product into another.
- They are composed of chains of operations that are composed based on the semantics of input & output data products.
- If the semantics of discoverable services are richly defined it is possible to automatically derive a basic workflow to produce a desired result from available dataproducts.
- “Build self-assembling, ontologically described grids!” – marlon pierce
More Information

Wednesday December 10\textsuperscript{th} 6pm to 8pm

• Posters:
  – Srinath Perera: Managing E-Science Cyber-Infrastructures: A Case Study
  – Yiming Sun: Limits of Automated Curation of e-Science Data

• Demonstrations:
  – Open Grid Computing Environment’s Workflow Suite for E-Science Projects
GridChem Advanced Support

• Help with Workflows
• Software Repositories
• Benchmarking
• Scheduling
• Community Account Fair Share Policy
• GridChem is a molecular Chemistry grid serving chemistry researchers in running chemistry applications on Grid Resources.
Coupled Chemistry Workflows

- Challenges:
  - Some apps have rich Client Gui’s, a challenge with asynchronous long running workflows
  - Workflow Verification Service
Parametric Workflows

- **Challenges:**
  - Parametric sweep scheduling, monitoring iteration steps, graphical composition
Human Interaction Workflows

• Challenges:
  – Need to step into workflow execution.
  – Optimize and suggest changes to user.
OGCE DEMO Portal

- Google for OGCE, ogce web site -> Tutorials
- URL:  
  http://ogceportal.iu.teragrid.org:8080/gridsphere/gridsphere
- User name: train01 to train30
- Password:
Workflow Suite Architecture

- **Application Developers (A2)**
  - Service Toolkit
  - Web UI

- **Workflow Provider (B1, B3)**
  - **Xbaya (B2, C2)**

- **End Users (C1)**
  - **(C8)**

- **Notification Bus**

- **GridFTP/SFTP/HTTP (C7)**

- **XRegistry (A3)**

- **Service Factory (C3)**

- **Application Services (C4)**

- **Application (C6)**

- **Compute Nodes (C5)**

- **End Users (C7)**

- **Input Data (C7)**
Hands On

• Configure and Run pre constructed workflow
• Compose new workflow, launch it and monitor progress.
• Register new application services on TG Resources.
• Any other interactive tasks ???