Worldwide Collaborative Research and Learning Environments: The Access Grid Experience

Leslie Southern, Ohio Supercomputer Center, USA
Mary Bea Walker, National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign, USA

An Introduction to the Access Grid

The Access Grid (AG) is a collaboration technology designed to support wide-area group-to-group scientific collaboration. The Access Grid was developed by the Futures Laboratory [1] in the Mathematics and Computer Science Division at Argonne National Laboratory, with support from the US Department of Energy. Since its initial deployment in the summer of 1999, the number of people in the AG user community has grown into the hundreds worldwide.

The physical infrastructure of the AG consists of a grid [2] of geographically distributed room-based spaces, called Nodes. Each Node houses a standard set of hardware designed to enable group-to-group interactions. Nodes are instrumented with commodity equipment such as PCs, cameras, microphones, and projectors. With high-quality yet affordable audio and video and large-format displays, the AG goes beyond traditional desktop teleconferencing, creating compelling workspaces for group work.

Access Grid software is peer-to-peer, allowing multiple geograhpically dispersed collaborative communities to form across the grid. The core of the software is its Virtual Venues server, around which a virtual community of like-minded collaborators can form. A Virtual Venue can be viewed as a conference room on the grid that Nodes can occupy together, allowing people from different physical locations to be co-located in a virtual space. A Virtual Venues server hosts multiple Virtual Venues, so that simultaneous events can take place in different virtual spaces.

The Access Grid supports multiple interaction modalities. It has been used for ad hoc meetings, as well as for more formal interactions such as site reviews, technology demonstrations, conferences, and training sessions. The Partners for Advanced Computational Services (PACS) and a number of individual institutions in the 50+ National Science FoundationNational Computational Science Alliance (the Alliance) have offered a number of high performance computing (HPC) training events using the Access Grid to reach a wide audience with similar interests.

Training the HPC Community Using the AG

HPC workshops cover a range of topics, including parallel programming environments and performance analysis for applications that use HPC systems. Two PACS-sponsored HPC workshops are discussed here giving special attention to the planning and preparation required for successful Access Grid events. One workshop covered Message Passing Interface (MPI) and the other workshop covered performance tuning. The MPI workshop was a multi-day event and included hands-on exercises. The performance tuning workshop was 3 hours long and consisted of lecture only. In both cases, while the complexities of the planning and preparation varied, the process itself was the same and differed little from locally delivered training events. The planning and preparation process for global training events, those using the AG, is compared to local training events and further described. Additional information is provided about the workshops, and conclusions based on our experiences are provided.

Planning and Preparation

Planning a training event over the AG is much like planning a local training event except for adding a technical component and providing multiple site participation.

To plan local events, we

- identify an event (i.e., determine instructor, topic, format, and length)
- schedule the event
- prepare and deliver an announcement
- receive registrations
- prepare and make available materials
- offer workshop
- summarize evaluations

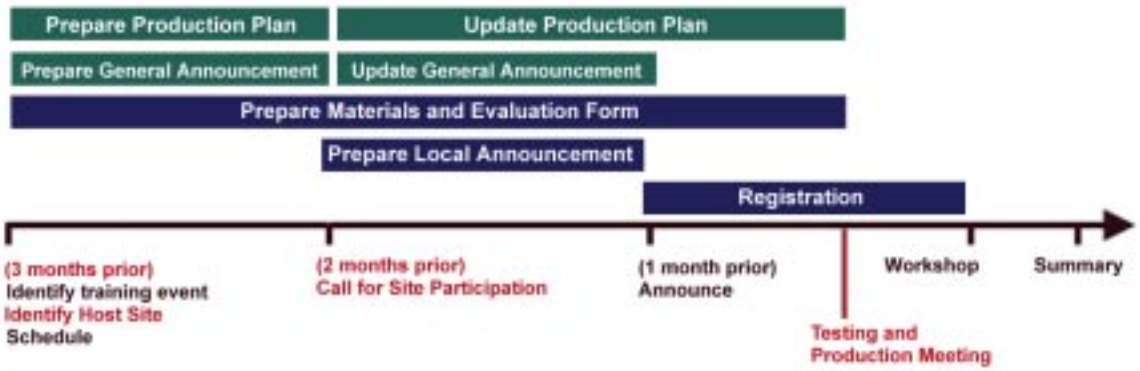
For global events offered over the AG, we follow the steps above and

- identify a host site
- create a general announcement for all sites
- call for site participation
- develop a production plan
- schedule testing time

We allow approximately two months for planning, preparing, and offering a local training event and approximately three months for global events. The diagrams below illustrate the steps taken in planning and preparing a local training event and a global training event. Distinct time-dependent activities are denoted below the timeline, and planning and preparations are shown above the timeline.

Diagram 1: Local Training Event Timeline

Diagram 2: Global Training Event Timeline



In the global training diagram above, additional steps are indicated in red and additional requirements are indicated in green. The major differences between a global training event and a local training event are the addition of the host site, call for site participation, a testing and production meeting, a general announcement, and a production plan.

Host Site

The host site leads the event production and technical direction. The host site is responsible for developing and maintaining the general workshop announcement and production plan. The host site oversees the

planning and preparations and keeps an open channel of communications with all necessary parties at participating sites. The host site is typically predetermined by the location of the primary instructor.

Call for Site Participation

The call for site participation solicits interest from sites that have Access Grid Nodes at their locations. The event is scheduled in a virtual venue and announced to the `ag-event-announce@mcs.anl.gov` mailing alias. Included in the description is the call for site participation. The call for site participation includes a pointer to the general announcement that describes the workshop, clearly identifies local arrangement responsibilities, and specifies call for participation deadlines.

Whether a site chooses to participate often depends on the interests of whoever receives the call for participation. Because the mailing list is growing and not a significant number of potentially interested parties may be part of the existing list, often the call is widely distributed to individuals with Access Grid Nodes at their locations. In the event of a wide distribution, care must be taken to distinguish the call for *site* participation from the announcement for *individual* participation (i.e., a user who wants to register to attend).

Testing and Production Meeting

The testing and production meeting is held at least one week before the workshop. This meeting is an opportunity for the participating sites to meet each other's local arrangements staff and node operators as well as the instructor(s); test audio levels, video streams, telephone backups, and collaborative tools; discuss logistics about audience interactions (e.g., how and when questions are asked and identified), camera views, and display wall layouts; go over agenda; and resolve any issues and answer any questions. The testing and production meeting is also used as a brief dry run for the instructors. Depending on the number and experience of the participating sites and complexity of the workshop, the testing and production meeting lasts between 30 minutes and 2 hours. It is also used to determine whether another testing and production meeting is necessary.

General Announcement and Local Announcements

The general announcement is a web-based mechanism for disseminating information about the workshop and registering users to attend. The general announcement provides detailed information about the workshop, such as an abstract, agenda, objectives, target audience, prerequisites, and instructor. It also includes workshop dates and times for a given time zone. It points to site-specific local announcements.

Local announcements provide an overview of the workshop and point to the general announcement for the workshop details. Each local announcement provides specifics about registration, the facility, and location. Each local announcement also provides local times. In addition to web-based versions of the local announcements, each participating site often sends electronic versions to their local communities through email.

Many users can learn about AG workshops through various avenues. Participating sites are encouraged to think globally. People can attend where it most convenient for them – perhaps at their home site or at another location, if they are traveling.

Production Plan

The web-based production plan provides contact information of each participating site, and the technical details needed for the workshop as well as those needed for the testing and production meeting.

In most cases, the role of arranging the workshop and the role of operating the AG Node at a given location are the responsibilities of separate individuals. For a smooth and successful production, these individuals need to be aware of one another's activities. The production plan is intended to support both roles. It includes contact information for both roles. It also includes detailed agendas and materials that both roles

require. For those arranging the workshop, the production plan points to supplemental materials, such as evaluation forms. For those operating the node, the production plan includes server information for needed collaborative tools (e.g., distributed PowerPoint and/or desktop sharing) and backup telephone numbers.

Workshops

The Alliance PACS have offered a number of HPC training events using the Access Grid. Two are identified here because of their differing formats and, therefore, differing planning and preparation complexities. Our 2-day MPI workshop included a hands-on component.. The other workshop covered performance tuning, was lecture only, and lasted 3 hours. In each case, the Ohio Supercomputer Center (OSC) provided the instruction and was designated as the host site.

While the multi-day format of the MPI workshop introduced some challenge, the hands-on component posed the most challenges. In addition to local node operation and local workshop arrangements, we needed local MPI technical support and local system access. Most of the participating sites had either one or the other capability but not both. Given the varying capabilities at the participating sites, we decided to schedule the lab exercises at the end of each day for a couple of hours. Attendees were encouraged to go to a nearby lab, if one was available, or go back to their desktops. We chose to use a simple chat tool and email as a means to communicate with the instructor and/or MPI technical support people. Supplements were available on the general announcement for the lab procedures and exercises. Most of the local sites provided accounts to their HPC systems running a version of MPI and were responsible for disseminating site-specific information about access and usage.

For the MPI workshop, 8 sites from around the United States participated. They included OSC, the Albuquerque High Performance Computing Center, Boston University, Dartmouth College, the University of Kansas, National Center for Supercomputing Applications (NCSA)/University of Illinois at Urbana-Champaign, North Dakota State University, and the University of Kentucky. Attendance at each site varied. A total of 96 attendees participated. The overall evaluation score was 4.64 out of 5.00 [3].

The performance workshop attracted a larger number of sites (12), including two from European countries. Coordinated sites that participated were OSC, Albuquerque High Performance Computing Center, Boston University, CINECA (interuniversity consortium consisting of 15 Italian universities), Dartmouth, National Center for Atmospheric Research, NCSA, Pittsburgh Supercomputing Center, San Diego Supercomputing Center, University of Kansas, University of Kentucky, and University of Manchester. In addition to the sites known before the workshop, four additional sites joined at the time of the workshop. These sites had limited audiences and minimal interactions with other workshop participants. A total of 102 known attendees participated in the performance tuning workshop. The overall evaluation score was 4.29 out of 5.00.

Planning and Preparation Differences between Local and Global Events

The main differences between local training events and global training events that take advantage of the Access Grid capabilities are the addition of a technical component and multiple site participation. To offer a smooth and successful global training event requires additional time to coordinate and plan. The complexities of the training event, such as format, correspond to the complexities of the planning and preparations. It was also noted that larger audiences and, therefore, a larger pool of experiences were an unexpected added value for the workshops. The workshops were more interactive; the knowledge shared was excellent and extremely helpful for all participants.

When offering global events, whether the events are informal events like ad hoc meetings or more formal events like training sessions, the involved staff– those making the arrangements and those operating the AG Nodes – should have an understanding of the Access Grid infrastructure. Mailing lists, online documentation, instructor-led tutorials, and Web-based AG courses are some of the avenues used to obtain this knowledge. More information about these can be found at http://www.accessgrid.org/.

Teaching People About the AG: Building Web-Based Courses

Motivation

Since June 1999, the staff of the Futures Laboratory [1] at Argonne National Laboratory has held face-to-face tutorials to teach new users how to build and use AG infrastructure. These tutorials are hands-on and so the number of attendees at a given tutorial is necessarily limited. In addition, the format of the tutorials requires students and instructors to share the same physical space.

At the end of January 2001, an author conducted a formal evaluation of a 60-participant, hands-on tutorial conducted by the Futures Lab. This is what participants hoped to gain by attending the live AG tutorial:

- "A better understanding of exactly what the AG is and its capabilities
- An understanding of some issues surrounding creation & operation of an AG node
- Understanding of building an AG node
- A basic understanding of what goes into building an AG
- The basics of designing, operating and maintaining an AG
- Concepts on fundamental design & implementation of the AG
- Ideas on what is a good layout of an AG node and what is needed for an AG node
- See an actual working AG node
- I need to build a node. I also need to create applications.
- Need documentation to train the trainer at home
- More on how it works beyond Web information; issues, problems
- Want info on group-to-group, one-on-one collaboration; human-human interaction
- Current hardware/software overview and future updates & improvements
- Get a close-up of total requirements to run an AG Node"

At the end of the tutorial, the participants were asked if they met their goals. Most responded, "Yes." Others said they wanted more documentation, videos, and longer, more extensive tutorials. More than two-thirds of respondents said that the tutorials increased their knowledge and ability for their professional work. Seventy-two percent of respondents wanted to attend another tutorial. Many mentioned items they would like to see covered that were not in the tutorials owing to time limitations.

2001 was a busy year in the AG community. In particular, the 2001 SC Conference was to hold its first global technical conference over the Access Grid. The project, called SG Global, was to bring together over 39 participating AG sites worldwide to discuss ideas relating to high-end computing. Many of the participants in SC Global were new to the AG and required elementary training. Many SC Global participants were unable to attend a hands-on tutorial. This created a problem: *How to stretch a few experts' time to fulfill the needs of a growing community that requires immediate and up-to-date AG training?* The potential solution: Internet-enabled e-training.

The authors, with support from the National Center for Supercomputing Applications, from Lisa Childers and others from Argonne National Laboratory and from personnel from the Ohio Supercomputer Center, made it their goal to build a set of elementary AG tutorials that would be accessible via the World Wide Web. This would allow independent study without requiring students and instructors to travel. To develop the online training course, the authors assembled existing AG documentation [4-11], which, though abundant, did not represent a cohesive set of standalone AG training materials. From this material and using feedback from the tutorial evaluation, the authors created a master outline for the web-based tutorial about the AG.

Web-based Tutorials

Based on the master outline and existing documentation, the tutorials were created using WebCT™. This product is one of the leading distance learning technologies used in higher education. WebCT™ provides

mechanisms for building and navigating through web-based materials and includes features such as an online glossary, self tests, and tracking. Initially, three tutorials were outlined; however, given the amount of content, the target audience, and the goal of each tutorial, a fourth one was also developed. The tutorials include

1. ***How to Build and Install an Access Grid Node (AGN): An Elementary Guide for Technical Users***
2. ***How to Operate an Access Grid Node (AGN): An Elementary Guide for Technical Users***
3. ***How to Produce an Access Grid Event: An Elementary Guide for Technical Users***
4. ***How to Set Up an Access Grid Meeting: An Elementary Guide for Non-Technical Users***

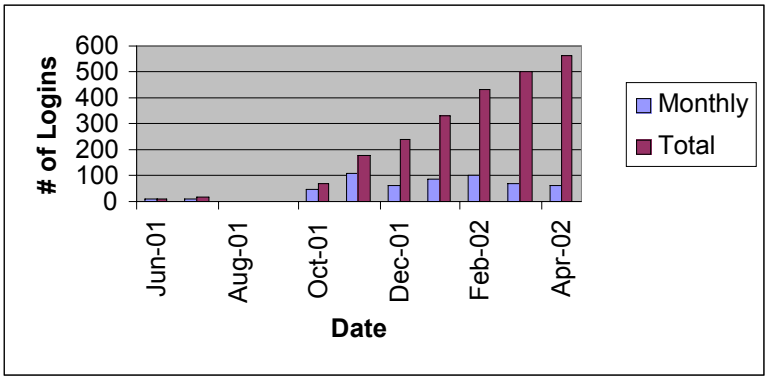
The tutorials were created in six months and reviewed in approximately 30 days. Key staff and the AG user community reviewed the tutorials. The newer sites that participated in SC Global were part of the review team. The four elementary tutorials were formally released to the public 30 October 2002 -- a few weeks before the SC Global event. This was ample time to assist student workers in understanding the AG and how to operate an AG node.

A login is required to access WebCT™ materials. While a guest login is provided, several users prefer creating their own. WebCT™ provides a discussion space, so logins help the community identify one another. They are also a means of tracking data concerning the tutorials. To date, 602 users have created logins for the AGiB tutorials. Because personal information about the user is optional, locations for just 235 of these logins are known. As the table below illustrates, users are not only located around the USA but also around the world.

US States (Top 15 in Descending Order)		**Countries (Top 15 in Descending Order)**	
US State	**Number of Logins**	**Country**	**Number of Logins**
Illinois	48	USA	159
California	32	UK	25
Florida	12	Australia	12
Indiana	11	Japan	9
Ohio	11	Canada	8
New Jersey	8	Italy	8
New York	7	Korea	8
Mississippi	6	China	7
Louisiana	5	India	7
Massachusetts	5	Taiwan	7
North Carolina	5	Germany	6
Washington	5	Spain	6
Arizona	4	Brazil	4
Washington DC	4	France	4
Montana	4	Mexico	4

Based on domain names (i.e., .edu, .com, .gov and .mil, and .org), institutions were categorized into four groups. Because internet service providers do not necessarily reflect an accurate institution type of the user, they were excluded. Of known types, 73% were academic, 14% corporate, 11% government, and 1% nonprofit.

Since the tutorials were publicly announced, an average of 80 additional logins are created each month. The chart below illustrates the number of logins per month along with the cumulative data.



The timeline begins the month the files were placed on the WebCT server (http://webct.ncsa.uiuc.edu:8900/public/AGIB/) at the National Center for Supercomputing Applications (NCSA). The review period occurred during the month of October and public access began in November.

Web-based Tutorials Under Development

Two additional sets of materials are under development for the suite of AGiB tutorials. They include an overview for decision makers and a troubleshooting guide. The need for these materials is based on user input from the AG community and those professionals interested in return on investment (ROI). Such comments as these drove the creation of the "Overview for Decision Makers".

- "I want to explain the benefits and costs of the AG to upper management and attempt to get funding and support. If successful, I want to set up a node."

- "I want to sell my admin people on the idea of funding an AG Node for our campus. I got the 35,000 ft. overview, but I won't be able to answer questions about costs for training, maintenance, full time equivalent people to dedicate to this project. I will now flounder around trying to answer these questions."

An interactive web-based format may appeal to some decision makers, but not necessarily all, so a supplement was added to the overview. This supplement is a ROI brief that can easily be printed and handed to higher management.

Planning is underway for the troubleshooting tutorial. Much of the content for it must be gathered from multiple sources, such as node operators, event producers, event technical directors, audiences, and mail archives.

There are many opportunities throughout the tutorials to add web-based media that enhance the learning experience. Possible media alternatives include video with audio and graphic animations. Both media provide visual queues and spatial references. Each requires time and resources to develop.

An AG Node Overview video is near completion. Work on this 3-minute video paralleled the work on the 4 tutorials. It was scripted in the spring, shot in the summer, and edited in the fall. Hyperlinks to the tutorial will appear in many locations throughout the tutorials.

A 3-dimensional graphic model of an AG Node is under development. It will be used throughout the tutorials as a common reference point for a number of components discussed.

Acknowledgments

The authors thank Rick Stevens and the Futures Laboratory staff, especially Lisa Childers, in the Mathematics and Computer Science Division of Argonne National Laboratory; Sandie Kappes, Jackie Kern, and Karen Green at the National Center for Supercomputing Applications; Jennifer Teig Von Hoffman and Russ Wolf at Boston University; and Angie Galipault and Barbara Woodall at the Ohio Supercomputer Center.

References

[1] Disz, T.L., Evard, R., Henderson, M.W., Nickless, W., Olson, R., Papka, M.E. and Stevens, R. Designing the Future of Collaborative Science – Argonne's Futures Laboratory. *Ieee Parallel & Distributed Technology*, *3* (2). 14-21.

[2] Foster, I. and Kesselman, C. (eds.). *The Grid: Blueprint for a New Computing Infrastructure*. Morgan Kaufmann, 1999.

[3] Leslie Southern, OSC, "Report on March 28 & 29, 2001 MPI Workshop over the Access Grid", April, 2001, http://alliance.osc.edu/mpi/report.pdf.

[4] B. Nickless, "Access Grid Networking: Chautauqua 2K: Minimizing Risk," (March, 2000) PPT slides on AG networking).
http://www.mcs.anl.gov/fl/accessgrid/ag-tutorial-mar00.htm

[5] Robert Olson , Argonne National Lab, "Access Grid Hardware Specification" (ANL, 2001) http://www.accessgrid.org/agdp/guide/spec.html

[6] Gary Refka, INSORS, "Building an Access Grid Node", Version 2.4.1, 2001
http://www.accessgrid.org/agdp/guide/building-an-access-grid-node.html

[7] Lisa Childers, "The Access Grid Node: An Operator's Manual" PPT slides (ANL, March, 2000). http://www.mcs.anl.gov/fl/accessgrid/ag-tutorial-mar00.htm

[8] Lisa Childers, "The Access Grid Node: Chautauqua 2K: Minimizing Risk," PPT slides (March, 2000) http://www.mcs.anl.gov/fl/accessgrid/ag-tutorial-mar00.htm

[9] Jennifer Teig Von Hoffman, Boston University, "Producing an Access Grid Event: Lessons Learned: Alliance Chautauqua;" PPT slides (Boston U, 1999). From
http://www.mcs.anl.gov/fl/accessgrid/ag-tutorial-mar00.htm

[10] Russ Wolf, Boston University, "If I Had Only Known: Technical Experiences, " PPT slides (March, 2000) http://www.mcs.anl.gov/fl/accessgrid/ag-tutorial-mar00.htm

[11] Mary Fritsch, Argonne National Laboratory, How to Set Up an Access Grid Meeting: A Guide for Non-technical Users, March, 2001.