1

Observations on WoCo9

William D. Gropp

Mathematics and Computer Science Division
Argonne National Laboratory

The Ninth Working Conference on Grid-Based Problem Solving Environments brought together researchers and practitioners interested in the use of numerical software in a distributed setting. This note summarizes some of the successes and challenges in working with numerical software in a Grid environment. This is not a comprehensive list of the contributions in the meeting; rather, it provides one person's impressions of the common themes and issues. In addition, some items were noticeable by their absence, and a number of these are discussed here. Opportunities for future work also are described.

1.1 Successes

First and foremost, this meeting illustrated the many successes of numerical computing on the Grid. Grids are in everyday use for scientific applications. Examples described in this workshop included MediGrid, NAREGEI Nano, LEAD, GridSolve, and St Germain. Many of these are data-centric, bringing together data and other resources that are distributed around the world.

One interesting use of the Grid is to address the dynamic and real-time access to resources for urgent computations, such as disasters, and for fluctuating demand, such as occasional real-time data acquisition and processing.

Talks at this meeting also described the development of both superstructures and infrastructures that have been developed to ease the creation of effective Grid tools. Partly because of this wealth of tools, many Grid-based services are being provided.

Another positive note was the absence of debate about the relative advantages of the Web and the Grid—this subject was not seen as important to the users, who simply want to get their work done. In addition, there was no confusion about the differences between a specification and an implementation, and hence the discussion remained focused on appropriate issues.

1.2 Challenges

The talks emphasized that many challenges remain, including the following:

Security (either the lack of it or the complexity and inconvenience of providing and dealing with it). This remains an unsolved problem in the sense that the available approaches are too cumbersome and too fragile.

Fragmentation and premature standardization. In some cases, there are clearly too many different projects (fragmentation). In others, such as the Grid RPC used in the GridSolve project, premature standardization has taken place, with the standard getting in front of thorough understanding of practical needs.

Precise syntax, semantics, and effective descriptions. These are often still missing, as systems try to "give users what they want." This is the same trap that has often made complex programming in shell languages so difficult.

Scalability of solutions. Some systems work well only because they have not been widely adopted. Part of the problem is in the implementation, but part also may be in the design; the HPC community has learned that scalability requires careful, deliberate design. This point was summed up in one comment made during the meeting: "Scalability doesn't happen by accident."

1.3 What Wasn't Discussed

Also interesting was what received little or no discussion. Like the dog that did not bark in the famous Sherlock Holmes story, the lack of discussion may indicate an unsolved problem.

In particular, numerical properties and the interactions between components received little attention. It may be important to preserve conservation laws and avoid numerical instabilities caused by exciting parasitic modes. Yet only Norris's presentation on computational quality of service explicitly addressed some of these issues; there was little discussion of the deeper mathematical issues (though some came up in Friday's presentations). The session on dynamic data-driven application systems did touch on this topic, but in terms of dynamic control of a computation, rather than the establishment of "contracts" of numerical properties between components.

Also nearly absent were quantified measures of comparison between approaches. The lack of such measures makes it hard to evaluate progress or compare different approaches. This situation is a striking contrast to that in high-performance computing, where there are many (sometimes controversial) measures of comparison. Until there is more quantification, the development of Grid- or Web-based systems will remain more of an art than a science.

1.4 Opportunities

The meeting clearly identified a number of opportunities, including the following:

1. **Develop performance models and measures of success.** How can a researcher compare different approaches? A partial ordering, perhaps a multidimensional ordering, would be helpful. This would enable researchers to quantify design decisions and address real performance issues, such as scalability, that arise when Grid applications become successful.
2. **Develop community standards and best practices.** Vouk's talk on workflow tools described a large number of current tools. This area appears to be ripe for discussing community standards that would allow the next generation of workflow tools to build on current results. Many of the other talks discussed aspects of mathematical software that could benefit from even informal community standards or best practices in software engineering.
3. **Explore collaborations in developer tools.** For example, the GAT/SAGA, NAREGI-PSE, and Java CoG project have all developed their own sets of tools for building Grid- or Web-based applications; other projects have developed their own tools based on lower-level tools, such as using Perl scripts or ssh. While concurrent development is valuable for exploring different approaches and gaining experience, we should be moving to the next generation of tools, building on a common basis. More collaborations in these tools would help in this effort.
4. **Use the Grid to develop communities.** Can we use the Grid to discover scientists who are trying to use the Grid or Web in a similar way—for example, using data-mining techniques applied to project Web pages and publications to identify people who should at least check out one another's Web sites.

As this brief summary shows, the working conference provided a stimulating and fruitful experience for the participants.